

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

FACULDADE DE CIÊNCIAS - CAMPUS BAURU

DEPARTAMENTO DE COMPUTAÇÃO

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

GUILHERME LIMA ZANIN

**COLHENDO *INSIGHTS*: ANÁLISE DE DADOS NA PRODUÇÃO DE
CANA-DE-AÇÚCAR**

BAURU

Novembro/2025

GUILHERME LIMA ZANIN

**COLHENDO *INSIGHTS*: ANÁLISE DE DADOS NA PRODUÇÃO DE
CANA-DE-AÇÚCAR**

Trabalho de Conclusão de Curso do Curso
de Ciência da Computação da Universidade
Estadual Paulista “Júlio de Mesquita Filho”,
Faculdade de Ciências, Campus Bauru.
Orientador: Prof. Dr. Clayton Reginaldo Pereira

BAURU
Novembro/2025

Z31c Zanin, Guilherme Lima
 Colhendo Insights: Análise de Dados na produção de
 cana-de-açúcar / Guilherme Lima Zanin. -- Bauru, 2025
 134 p. : il., tabs., fotos

 Trabalho de conclusão de curso (Bacharelado - Ciência
 da Computação) - Universidade Estadual Paulista
 (UNESP), Faculdade de Ciências, Bauru
 Orientador: Clayton Reginaldo Pereira

 1. Ciência de Dados. 2. Cana-de-Açúcar. 3. Nutrientes
 do Solo. 4. Modelos Preditivos. 5. Agricultura de Precisão.
 I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados
fornecidos pelo autor(a).

Guilherme Lima Zanin

Colhendo *Insights*: Análise de Dados na Produção de Cana-de-Açúcar

Trabalho de Conclusão de Curso do Curso de Ciência da Computação da Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências, Campus Bauru.

Banca Examinadora

Prof. Dr. Clayton Reginaldo Pereira

Orientador

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Profa. Dra. Simone das Graças Domingues Prado

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Prof. Dr. Leandro Aparecido Passos

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Bauru, 13 de Novembro de 2025.

Agradecimentos

Dedico este trabalho, primeiramente, à minha família — meus pais, Val e Vera, e meu irmão, Lucas — que, com amor incondicional, dedicação e esforço incansável, sempre acreditaram em mim e fizeram de tudo para que eu pudesse trilhar meu próprio caminho. Foram eles que, com exemplos diários de honestidade, coragem e perseverança, me ensinaram que os sonhos só se realizam com trabalho, fé e determinação. Cada conquista minha é, na verdade, reflexo do apoio e das renúncias que vocês fizeram por mim. Este trabalho é, antes de tudo, de vocês.

Dedico também este trabalho à memória da minha querida avó, cuja presença, mesmo não estando mais fisicamente entre nós, continua viva em cada lembrança, gesto e valor que me transmitiu. Sua força, sabedoria e carinho permanecem como um farol que orienta meus passos e me inspira a ser alguém melhor a cada dia. Tenho a certeza de que sua luz acompanha esta conquista, tornando-a ainda mais especial.

Aos meus amigos, deixo um agradecimento sincero por estarem sempre presentes nos momentos mais importantes — nas horas de alegria, de cansaço e de incerteza. A amizade, o companheirismo e as risadas de vocês tornaram essa caminhada mais leve e inesquecível. Cada conversa, incentivo e gesto de apoio fez toda a diferença para que eu chegasse até aqui.

Expresso minha mais profunda gratidão ao meu orientador, Dr. Clayton Reginaldo Pereira, por todo o conhecimento compartilhado, pela paciência, pelo apoio constante e por acreditar neste projeto desde o início. Mais do que um orientador, foi um verdadeiro mentor e amigo, cuja dedicação e incentivo foram fundamentais para o desenvolvimento e conclusão deste trabalho.

Por fim, agradeço a todos que, de alguma forma, contribuíram para esta jornada — colegas, professores, amigos e familiares —, pois cada palavra de incentivo, cada gesto de apoio e cada momento compartilhado tornaram este projeto mais significativo e completo.

A persistência é o caminho do êxito.

Charles Chaplin

Resumo

A cana-de-açúcar ocupa posição central na economia brasileira, consolidando o país como o maior produtor mundial dessa cultura. Apesar de sua relevância, a produtividade ainda enfrenta entraves relacionados ao manejo nutricional do solo, principalmente pela ausência de controle preciso sobre as quantidades ideais de nutrientes necessárias ao desenvolvimento da lavoura. Essa falta de precisão faz com que muitos produtores deixem de alcançar o potencial produtivo de suas áreas, resultando em menor rentabilidade e uso ineficiente de insumos. Nesse cenário, a ciência de dados surge como um recurso estratégico para apoiar a gestão agrícola, permitindo identificar padrões ocultos, analisar correlações entre nutrientes e fornecer recomendações mais assertivas para o manejo da fertilização. Este trabalho tem como objetivo aplicar técnicas de análise de dados para estudar as relações entre os nutrientes presentes no solo e a produtividade da cana-de-açúcar. A pesquisa concentrou-se na coleta, organização e tratamento de dados nutricionais, seguida de análises estatísticas e exploratórias. Em seguida, foram desenvolvidos e avaliados modelos preditivos supervisionados, com destaque para a regressão, visando estimar a influência de nutrientes específicos e oferecer insights que apoiem práticas de fertilização mais eficientes. Os resultados esperados incluem a definição mais precisa das quantidades de nutrientes necessárias, a maximização da produtividade agrícola, a redução de desperdícios e a promoção de práticas mais sustentáveis. Ao integrar ciência de dados ao manejo nutricional da cana-de-açúcar, este estudo busca oferecer subsídios técnicos que auxiliem a tomada de decisão dos produtores rurais, fomentem a agricultura de precisão e impulsionem a competitividade do setor sucroenergético no Brasil.

Palavras-chave: Ciência de Dados; Cana-de-Açúcar; Nutrientes do Solo; Modelos Preditivos; Agricultura de Precisão.

Abstract

Sugarcane plays a central role in the Brazilian economy, consolidating the country as the world's largest producer of this crop. Despite its relevance, productivity still faces challenges related to soil nutrient management, mainly due to the lack of precise control over the ideal amounts of nutrients required for crop development. This lack of accuracy prevents many producers from reaching the full productive potential of their fields, resulting in lower profitability and inefficient use of inputs. In this context, data science emerges as a strategic tool to support agricultural management, enabling the identification of hidden patterns, the analysis of correlations between nutrients, and the generation of more assertive recommendations for fertilization management. This study aims to apply data analysis techniques to investigate the relationships between soil nutrients and sugarcane productivity. The research focused on the collection, organization, and processing of nutritional data, followed by statistical and exploratory analyses. Subsequently, supervised predictive models, with emphasis on regression, were developed and evaluated to estimate the influence of specific nutrients and provide insights that support more efficient fertilization practices. The expected results include a more accurate definition of the necessary nutrient quantities, maximization of agricultural productivity, reduction of waste, and promotion of more sustainable practices. By integrating data science into the nutritional management of sugarcane, this study seeks to provide technical support for farmers' decision-making, foster precision agriculture, and enhance the competitiveness of Brazil's sugar-energy sector.

Keywords: Data Science; Sugarcane; Soil Nutrients; Predictive Models; Precision Agriculture.

Lista de figuras

Figura 1 – Morfologia da Cana-de-Açúcar.	21
Figura 2 – Sistema Radicular da Cana-de-Açúcar.	22
Figura 3 – Fases da Fenologia da Cana-de-Açúcar.	23
Figura 4 – Diferenças entre os tipos de plantio da cana-de-açúcar: cana-planta e cana-soca.	26
Figura 5 – Distribuição das principais regiões produtoras de cana-de-açúcar no Brasil.	32
Figura 6 – Exemplo de Boxplot	36
Figura 7 – Arquitetura cliente-servidor empregada no AgroApp.	45
Figura 8 – Tela de Login do AgroApp.	46
Figura 9 – Tela inicial (Home) do AgroApp com cards de gestão e análise.	47
Figura 10 – Exemplos de telas dos módulos de Clima, Produção e Análise do Solo.	48
Figura 11 – Exemplo de cards de análise e relatórios no AgroApp.	50
Figura 12 – Arquitetura MVC adotada no AgroAnalysis.	52
Figura 13 – Tela de Login do AgroAnalysis.	52
Figura 14 – Tela de Análise de Nutrientes.	53
Figura 15 – Tela de Análise do Solo.	54
Figura 16 – Tela de Predição da Produção.	55
Figura 17 – Tela de Análise Estatística Avançada.	55
Figura 18 – Exemplos de análises estatísticas apresentadas nos cards do sistema.	56
Figura 19 – Amostra do <i>dataframe</i> da fazenda 26, contendo dados de amostras de solo, com informações sobre nutrientes e variáveis de manejo.	58
Figura 20 – Amostra do <i>dataframe</i> de produção, contendo dados sobre produtividade e qualidade da cana-de-açúcar.	59
Figura 21 – Exemplo do <i>dataframe</i> original antes do processo de Limpeza dos Dados.	60
Figura 22 – Amostra do <i>dataframe</i> de produção, exibindo as amostras com dados nulos antes do processo de Limpeza dos Dados.	61
Figura 23 – Aplicação do filtro e <i>dataframe</i> correspondente à profundidade de 20 cm.	61
Figura 24 – Aplicação do filtro e <i>dataframe</i> correspondente à profundidade de 40 cm.	62
Figura 25 – Histogramas e boxplots dos nutrientes P, pH e Ca para amostras coletadas a 20 cm de profundidade.	79
Figura 26 – Exemplos representativos dos gráficos comparativos de médias de nutrientes por profundidade (20 cm em azul e 40 cm em laranja) gerados para cada lote e quadra.	86
Figura 27 – Frequências dos nutrientes ideais a 20 cm e 40 cm de profundidade.	88
Figura 28 – Box-plot para o nutriente P (Variância Homogênea).	94
Figura 29 – Box-plot para o nutriente M.O (Variância Heterogênea).	95

Figura 30 – Variância Explicada Acumulada - PCA 20 cm	99
Figura 31 – Biplot - PCA 20 cm	100
Figura 32 – Variância Explicada Acumulada - PCA 40 cm	101
Figura 33 – Biplot - PCA 40 cm	102
Figura 34 – Método do Cotovelo - 20 cm	103
Figura 35 – Método do Cotovelo - 40 cm	103
Figura 36 – Silhouette Score - 20 cm	104
Figura 37 – Silhouette Score - 40 cm	104
Figura 38 – Visualização dos <i>clusters</i> via PCA - 20 cm	105
Figura 39 – Visualização dos <i>clusters</i> via PCA - 40 cm	105
Figura 40 – Distribuição do <i>Cluster</i> a 20 cm por Quadra	107
Figura 41 – Distribuição do <i>Cluster</i> a 40 cm por Quadra	108
Figura 42 – Importância dos Nutrientes na Clusterização (20 cm)	109
Figura 43 – Importância dos Nutrientes na Clusterização (40 cm)	110
Figura 44 – Matriz de Correlação de Pearson entre as variáveis quantitativas produtivas.	115
Figura 45 – Boxplot de Produtividade para Intervalo entre Cortes.	121
Figura 46 – Importância dos nutrientes a 20cm e 40cm de profundidade para a produ- tividade.	126
Figura 47 – Gráficos de Dispersão para Previsões a 20cm de Profundidade	127
Figura 48 – Gráficos de Dispersão para Previsões a 40cm de Profundidade	127

Lista de quadros

Quadro 1 – Comparação entre os tipos de plantio da cana-de-açúcar	27
Quadro 2 – Bibliotecas utilizadas e suas respectivas finalidades	42
Quadro 3 – Faixas ideais de nutrientes e atributos químicos do solo para o cultivo de cana-de-açúcar.	66
Quadro 4 – Classificação de normalidade e presença de <i>outliers</i> para os nutrientes a 20 cm de profundidade.	80
Quadro 5 – Classificação de normalidade e presença de <i>outliers</i> para os nutrientes a 40 cm de profundidade.	81
Quadro 6 – Medidas descritivas dos nutrientes do solo a 20 cm de profundidade. . . .	82
Quadro 7 – Medidas descritivas dos nutrientes do solo a 40 cm de profundidade. . . .	82
Quadro 8 – Transformações aplicadas aos nutrientes a 20 cm de profundidade e respectivos <i>p-values</i> pós-transformação.	83
Quadro 9 – Transformações aplicadas aos nutrientes a 40 cm de profundidade e respectivos <i>p-values</i> pós-transformação.	84
Quadro 10 – Síntese dos nutrientes dentro dos níveis ideais em cada lote/quadra e profundidade.	87
Quadro 11 – Resultados do teste <i>t</i> pareado para comparação dos nutrientes entre as profundidades de 20 cm e 40 cm.	90
Quadro 12 – Resultados do teste de homogeneidade de variâncias para os nutrientes analisados.	93
Quadro 13 – Correlação entre os nutrientes a 20cm de profundidade.	96
Quadro 14 – Correlação entre os nutrientes a 40cm de profundidade.	97
Quadro 15 – Classificação de normalidade para os dados de produção.	113
Quadro 16 – Resultados do teste de normalidade e transformações aplicadas aos dados de produção, com os respectivos <i>p-values</i> após as transformações.	114
Quadro 17 – Comparações múltiplas de médias entre variedades de cana-de-açúcar (teste de Tukey HSD)	117
Quadro 18 – Comparações múltiplas de médias entre estágios de corte (teste de Tukey HSD)	118

Lista de abreviaturas e siglas

AED	Análise Exploratória dos Dados
Al	Alumínio
ANN	Redes Neurais Artificiais
Ar	Açúcares redutores
Art	Açúcares redutores totais
ATR	Açúcar Total Recuperável
B	Boro
Brix	Teor de sólidos solúveis totais
Ca	Cálcio
CSV	Comma-Separated Values
CTC	Capacidade de Troca de Cátions
Cu	Cobre
Est Corte	Estágio do Corte
Esp	Espécie
Fe	Ferro
K	Potássio
Ls	Litros de sacarose por tonelada
Mg	Magnésio
Mn	Manganês
Mo	Molibdênio
M.O	Matéria Orgânica
N	Nitrogênio
Na	Sódio

P	Fósforo
Pbu	Peso do Bagaço Úmido
PCA	Análise de Componentes Principais
pH	Potencial do Hidrogênio
Pol%Caldo	Percentual de sacarose aparente no caldo extraído
Pol%Cana	Percentual de sacarose aparente na cana
Prod. Real	Produção Real
S	Enxofre
S.B	Soma das Bases
Sat.	Saturação por Alumínio
Tch Est	Toneladas de cana por hectare estimada
Tch Real	Toneladas de cana por hectare real
V	Saturação por Bases
Zn	Zinco

Sumário

1	INTRODUÇÃO	18
1.1	Problemática	19
1.2	Justificativa	19
1.3	Objetivos	19
1.3.1	Objetivo Geral	19
1.3.2	Objetivos Específicos	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Aspectos Biológicos da Cana-de-Açúcar	21
2.1.1	Morfologia e Taxonomia	21
2.1.2	Sistema Radicular	22
2.1.3	Fenologia	23
2.1.4	Tipos de Plantio	23
2.2	Nutrição Mineral e Fertilidade do Solo	27
2.2.1	Principais Nutrientes e Características Químicas do Solo na Produção da Cana-de-Açúcar	27
2.2.2	Faixas Ideais de Nutrientes para Cana-de-Açúcar	30
2.2.3	Distribuição de Nutrientes no Perfil do Solo	30
2.2.4	Fatores Ambientais	31
2.2.5	Manejo da Fertilidade do Solo no Brasil	32
2.3	Ciência de Dados	33
2.3.1	Dados Tabulares: <i>Comma-Separated Values (CSV)</i>	34
2.3.2	Análise Exploratória dos Dados (AED)	34
2.3.3	Fundamentos Estatísticos e Matemáticos	35
2.3.3.1	Medidas de Tendência Central e Dispersão	35
2.3.3.2	Boxplots	36
2.3.3.3	Correlação e Estatística Inferencial	37
2.3.4	Integração entre Agricultura e Ciência de Dados	39
2.3.5	Modelagem Estatística e Aprendizado de Máquina	39
3	METODOLOGIA	42
3.1	Ferramentas Utilizadas	42
3.2	Bibliotecas Empregadas	42
3.3	Abordagem Estatística	43
3.4	Modelagem Preditiva	43
3.5	Desenvolvimento de Sistemas	44

3.5.1	AgroApp: Aplicativo Mobile	44
3.5.1.1	Arquitetura do Sistema	44
3.5.1.2	Interfaces e Funcionalidades	45
3.5.2	AgroAnalysis: Sistema Web	50
3.5.2.1	Arquitetura do Sistema	51
3.5.2.2	Interfaces e Funcionalidades	52
3.5.3	Considerações sobre o Desenvolvimento	57
3.6	Conjunto de Dados	57
3.6.1	Limpeza de Dados	59
3.6.2	Tratamento de Dados	61
3.7	Procedimentos de Análise Exploratória dos Dados	62
3.7.1	Procedimentos para Caracterização dos Nutrientes e Dados de Produção	62
3.7.2	Procedimentos para Normalização dos Dados	63
3.8	Procedimentos para Análise da Distribuição Nutricional por Lote e Quadra	65
3.8.1	Definição de Faixas Ideais de Nutrientes	65
3.8.2	Cálculo de Médias por Lote e Quadra	66
3.8.3	Classificação dos Nutrientes	66
3.8.4	Geração de Visualizações Comparativas	67
3.8.5	Tabulação dos Resultados	67
3.8.6	Análise de Frequências	67
3.9	Procedimentos para Análise de Correlação entre Nutrientes	68
3.9.1	Método de Correlação Utilizado	68
3.10	Procedimentos para o PCA	69
3.11	Procedimentos para Análise de Clusterização	70
3.12	Procedimentos para Testes Estatísticos Inferenciais	72
3.12.1	Teste t Pareado para Comparação entre Profundidades	72
3.12.2	Teste de Homogeneidade de Variâncias (Levene)	72
3.12.3	Análise de Variância (ANOVA)	73
3.12.4	Teste de Kruskal-Wallis	73
3.12.5	Correlação de Spearman	73
3.12.6	Interpretação dos Resultados	74
3.13	Procedimentos para Análise da Relação entre Nutrientes e Produção	74
3.13.1	Correlação entre Nutrientes e Variáveis de Produção	74
3.13.2	Análise de Importância de Nutrientes para a Produtividade	75
3.13.3	Comparação entre Profundidades	75
3.14	Procedimentos para Construção dos Modelos Preditivos	76
4	RESULTADOS	78
4.1	Análises sobre os Nutrientes	78

4.1.1	Caracterização Inicial dos Dados	78
4.1.1.1	Análise de Normalidade e Identificação de <i>Outliers</i>	78
4.1.1.2	Estatísticas Descritivas	81
4.1.1.3	Normalização dos Dados	83
4.1.2	Distribuição Nutricional por Lote e Quadra	85
4.1.2.1	Variação Vertical dos Nutrientes	85
4.1.2.2	Síntese dos Nutrientes por Lote e Quadra	87
4.1.2.3	Padrões de Distribuição por Profundidade	88
4.1.3	Teste de Comparação de Profundidade	89
4.1.3.1	Nutrientes com Diferenças Altamente Significativas	90
4.1.3.2	Nutrientes com Diferenças Moderadas	91
4.1.3.3	Nutrientes sem Diferença Significativa	91
4.1.3.4	Implicações Agronômicas	92
4.1.4	Teste de Homogeneidade de Variâncias	92
4.1.4.1	Interpretação dos Resultados	92
4.1.4.2	Visualizações dos Dados	93
4.1.5	Correlação entre os Nutrientes	95
4.1.5.1	Correlação entre Nutrientes a 20 cm de Profundidade	96
4.1.5.2	Correlação entre Nutrientes a 40 cm de Profundidade	96
4.1.5.3	Análise Conjunta das Correlações	97
4.1.5.4	Implicações para a Fertilidade do Solo	97
4.1.5.5	Conclusão	98
4.1.6	Análise de Componentes Principais	98
4.1.6.1	Variância Explicada Acumulada - PCA 20 cm	98
4.1.6.2	Biplot - PCA 20 cm	99
4.1.6.3	Variância Explicada Acumulada - PCA 40 cm	100
4.1.6.4	Biplot - PCA 40 cm	101
4.1.6.5	Interpretação Geral da PCA	102
4.1.6.6	Conclusões	102
4.1.7	Análise de Clusterização	103
4.1.7.1	Método do Cotovelo	103
4.1.7.2	Silhouette Score	104
4.1.7.3	Visualização dos <i>Clusters</i> via PCA	104
4.1.7.4	Conclusões	105
4.1.8	Análise dos Perfis de Solo nas Quadras dos Lotes	105
4.1.8.1	Distribuição dos Perfis de Solo a 20 cm de Profundidade	106
4.1.8.2	Distribuição dos Perfis de Solo a 40 cm de Profundidade	106
4.1.8.3	Interpretação e Implicações	106
4.1.8.4	Visualização Gráfica	107

4.1.9	Análise de Sensibilidade Nutricional	108
4.1.9.1	Importância dos Nutrientes na Clusterização a 20 cm	108
4.1.9.2	Importância dos Nutrientes na Clusterização a 40 cm	109
4.1.9.3	Conclusões	110
4.1.10	Análise de Variância (ANOVA) para Comparação entre os <i>Clusters</i>	110
4.1.10.1	ANOVA para 20 cm	110
4.1.10.2	ANOVA para 40 cm	111
4.1.10.3	Conclusões	111
4.2	Análises sobre a Produção	112
4.2.1	Caracterização Inicial dos Dados de Produção	112
4.2.2	Normalização dos Dados de Produção	113
4.2.3	Correlação entre as variáveis produtivas	114
4.2.4	Conclusão sobre as Correlações Produtivas	116
4.2.5	Diferença de Produtividade entre Variedades de Cana-de-Açúcar	116
4.2.5.1	Interpretação dos Resultados	117
4.2.5.2	Conclusão	117
4.2.6	Diferença de Produtividade entre Estágios de Corte	118
4.2.6.1	Interpretação dos Resultados	118
4.2.6.2	Conclusão	119
4.2.7	Efeito do Histórico de Cortes	119
4.2.7.1	Correlação Spearman entre Intervalo de Cortes e Produtividade	119
4.2.7.2	Testes de Normalidade e Homogeneidade de Variâncias	119
4.2.7.3	Teste de Kruskal-Wallis	120
4.2.7.4	Boxplots de Produtividade por Intervalo de Corte	120
4.2.7.5	Conclusão	122
4.3	Nutrientes e Produção	122
4.3.1	Correlação entre Nutrientes e Variáveis de Produção	122
4.3.1.1	Melhores Correlações (Mais Fortes)	122
4.3.1.2	Piores Correlações (Mais Negativas)	123
4.3.1.3	Conclusão	124
4.3.2	Importância dos Nutrientes para a Produtividade	125
4.3.3	Construção dos Modelos para Previsão de Produção e Produtividade	126
4.3.3.1	Visualizações dos Resultados	126
4.3.3.2	Acurácia dos Modelos Obtidos	127
4.3.3.3	Conclusão	128
5	CONCLUSÃO	129
5.1	Principais Resultados	129
5.2	Sistemas Desenvolvidos	130
5.3	Implicações Práticas	130

5.4	Considerações Finais	131
5.5	Trabalhos Futuros	131
	REFERÊNCIAS	133

1 Introdução

A cana-de-açúcar (*Saccharum spp.*) ocupa posição estratégica na economia brasileira, consolidando o país como o maior produtor mundial desde a década de 1980. Segundo a Embrapa ([Empresa Brasileira de Pesquisa Agropecuária, 2023](#)), o setor sucroenergético brasileiro representa cerca de 40% da produção global, concentrando-se principalmente nas regiões Sudeste e Centro-Oeste, que juntas respondem por aproximadamente 90% da produção nacional. Na safra 2023/2024, a União da Indústria de Cana-de-Açúcar ([União da Indústria de Cana-de-Açúcar, 2024](#)) registrou o processamento recorde de 654,43 milhões de toneladas na região Centro-Sul, crescimento de 19,29% em relação ao ciclo anterior. Esses números evidenciam a relevância econômica do setor, que movimenta as cadeias produtivas de açúcar, etanol e bioenergia, contribuindo significativamente para o Produto Interno Bruto (PIB) agrícola nacional.

O sucesso da produção de cana-de-açúcar no Brasil deve-se a uma combinação de fatores, incluindo condições edafoclimáticas favoráveis – com temperaturas médias entre 20°C e 30°C e precipitações entre 1.200 mm e 1.500 mm ([Empresa Brasileira de Pesquisa Agropecuária, 2023](#)) – e a incorporação de avanços tecnológicos ao longo das últimas décadas. Apesar dessas condições, ainda existem desafios importantes relacionados à eficiência produtiva, especialmente no manejo nutricional do solo. A aplicação inadequada de insumos, seja por deficiência ou excesso, compromete o potencial produtivo das lavouras, aumenta os custos operacionais e pode gerar impactos ambientais negativos ([SILVA et al., 2023](#); [OLIVEIRA et al., 2010](#)).

Nesse cenário, a ciência de dados surge como ferramenta transformadora, capaz de extrair padrões e gerar insights a partir de grandes volumes de dados heterogêneos. Técnicas de aprendizado de máquina e análise estatística permitem não apenas descrever fenômenos complexos, mas também prever comportamentos futuros com base em informações históricas ([PROVOST; FAWCETT, 2013](#)). Na agricultura, esses métodos têm sido aplicados para apoiar decisões de manejo de insumos e otimizar a produtividade, alinhando-se aos princípios da agricultura de precisão ([MOLIN; AMARAL; COLAÇO, 2015](#)).

Estudos recentes demonstram o potencial dessas técnicas no setor sucroenergético. Souza *et al.* (2022) utilizaram Redes Neurais Artificiais (ANN) para prever a produção de açúcar e etanol com base em variáveis históricas das macrorregiões brasileiras, enquanto ([AONO et al., 2025](#)) aplicaram algoritmos de aprendizado de máquina para prever características agrônomicas a partir de dados genotípicos e fenotípicos. Modelos de regressão têm se mostrado eficazes na quantificação da influência de nutrientes específicos sobre a produtividade da cana-de-açúcar ([SUGAWARA; et al., 2007](#)), fornecendo bases científicas para a otimização da adubação.

1.1 Problemática

Apesar do crescente uso de tecnologias e modelos preditivos, a eficiência produtiva da cana-de-açúcar ainda enfrenta limitações. A falta de controle preciso sobre a aplicação de nutrientes essenciais gera subutilização do potencial produtivo das áreas cultivadas, aumento de custos e impactos ambientais. Além disso, há lacuna na integração de dados nutricionais e agronômicos para orientar a adubação de forma localizada, ajustada a diferentes estágios fenológicos da cultura.

1.2 Justificativa

A aplicação de técnicas de ciência de dados ao manejo nutricional da cana-de-açúcar justifica-se pela necessidade de aumentar a produtividade de forma sustentável. Modelos preditivos bem estruturados podem fornecer informações precisas sobre a influência de nutrientes específicos, permitindo decisões mais eficientes na adubação e manejo do solo. Dessa forma, espera-se reduzir custos, minimizar impactos ambientais e fortalecer a competitividade do setor sucroenergético brasileiro.

1.3 Objetivos

1.3.1 Objetivo Geral

Analisar e modelar a relação entre os nutrientes do solo e a produtividade da cana-de-açúcar utilizando técnicas de ciência de dados, com foco nas regiões Sudeste e Centro-Oeste do Brasil.

1.3.2 Objetivos Específicos

- Coletar e harmonizar dados agronômicos históricos de diferentes propriedades e regiões;
- Realizar análises exploratórias para identificar padrões, tendências e possíveis correlações entre nutrientes e produtividade;
- Realizar análises inferenciais para quantificar a influência dos nutrientes e características do solo na produção, testando hipóteses e relações estatísticas;
- Desenvolver modelos de regressão capazes de estimar a influência de diferentes nutrientes sobre o rendimento da cultura;
- Avaliar a precisão e robustez dos modelos preditivos, fornecendo subsídios técnicos para otimização da adubação;

- Gerar recomendações prescritivas para aplicação de nutrientes no solo, adaptadas às condições específicas de cada área, visando a melhora da produtividade;
- Construir um dashboard interativo que permita aos usuários visualizar e explorar os resultados das análises de forma intuitiva e prática;
- Desenvolver um aplicativo mobile que facilite o controle das fazendas pelo produtor rural, permitindo acesso rápido às análises, recomendações e informações sobre a produtividade e manejo nutricional;
- Fornecer suporte técnico para decisões estratégicas de manejo do solo, promovendo a agricultura de precisão e a sustentabilidade do setor sucroenergético.

Ao integrar ciência de dados ao manejo nutricional da cana-de-açúcar, este trabalho busca fornecer insights técnicos e operacionais que auxiliem produtores rurais na tomada de decisão, promovam a agricultura de precisão e fortaleçam a sustentabilidade e a competitividade do setor sucroenergético nacional.

2 Fundamentação Teórica

2.1 Aspectos Biológicos da Cana-de-Açúcar

A cana-de-açúcar (*Saccharum spp.*) é uma gramínea tropical da família *Poaceae*, cultivada principalmente para a produção de açúcar, etanol e bioenergia. No Brasil, a cultura ocupa posição estratégica na economia agrícola, sendo o país o maior produtor mundial. Compreender os aspectos biológicos dessa espécie é fundamental para o manejo adequado, otimização da produtividade e aplicação eficiente de insumos.

2.1.1 Morfologia e Taxonomia

A cana-de-açúcar apresenta características morfológicas típicas das gramíneas, com **caule ereto, cilíndrico e segmentado por nós, folhas lineares e flores em panículas**. O colmo armazena sacarose e atua como estrutura de sustentação, enquanto os nós contêm gemas que dão origem a novos brotos. A planta pode atingir de 2 a 6 metros de altura, dependendo da variedade, das condições climáticas e do manejo adotado.

A taxonomia é complexa devido à ocorrência de hibridizações entre *S. officinarum*, *S. spontaneum* e *S. robustum*, resultando em cultivares adaptadas a diferentes regiões e condições de cultivo. Como mostrado na figura 1, a cana-de-açúcar apresenta caule ereto, folhas lineares e flores em panículas, com colmo que armazena sacarose e nós que dão origem a novos brotos.

Figura 1 – Morfologia da Cana-de-Açúcar.



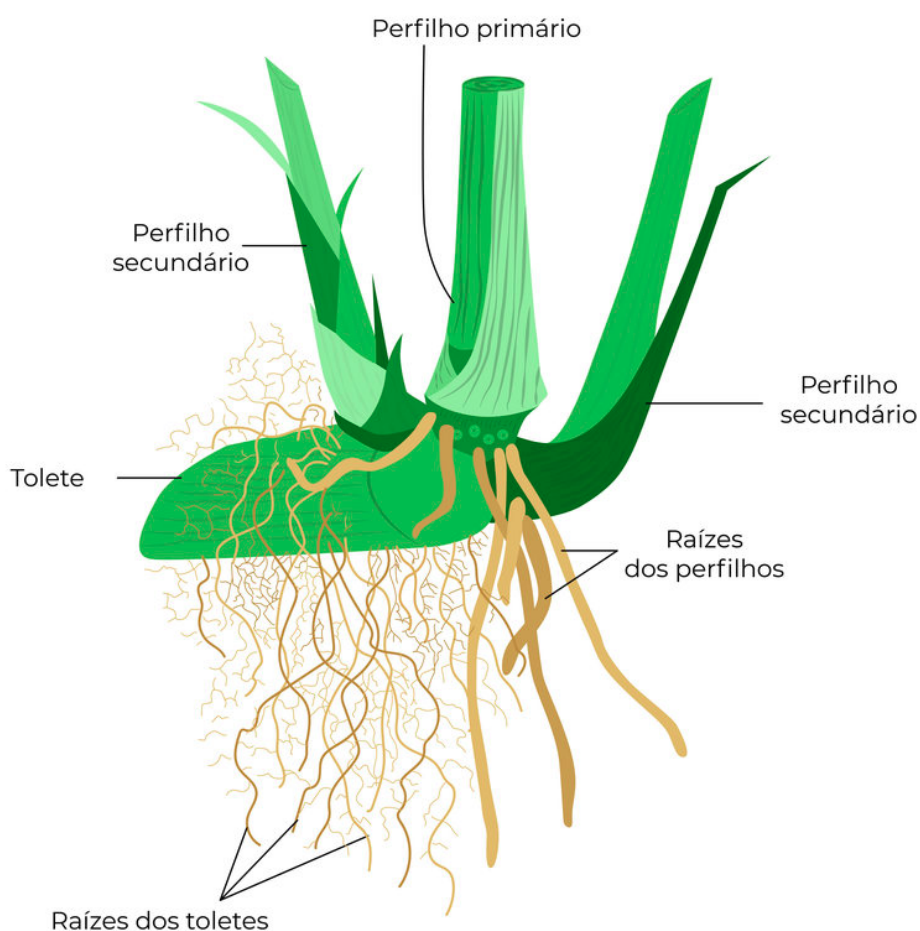
Fonte: Adaptada de Sliderpla, (2022).

2.1.2 Sistema Radicular

O sistema radicular é composto por raízes adventícias que se desenvolvem a partir dos nós do colmo enterrado, desempenhando papel fundamental na absorção de água e nutrientes. Como ilustrado na figura 2, as raízes podem ser superficiais ou profundas, e conhecer sua distribuição é essencial para a aplicação eficiente de fertilizantes:

- **Raízes superficiais:** nos primeiros 30–40 cm do solo, absorvem rapidamente nutrientes disponíveis e auxiliam na sustentação.
- **Raízes profundas:** podem atingir mais de 1,5 m, garantindo acesso a água e nutrientes em camadas profundas.
- **Importância para fertilização:** conhecer a distribuição radicular permite aplicar fertilizantes de forma eficiente e reduzir perdas por lixiviação.

Figura 2 – Sistema Radicular da Cana-de-Açúcar.



Fonte: Elevagro - Sistema Radicular da Cana-de-Açúcar.

2.1.3 Fenologia

O ciclo da cana-de-açúcar pode ser dividido em cinco fases principais: **brotamento**, **perfilhamento**, **crescimento**, **maturação** e **floração**. Cada fase possui exigências específicas de água, nutrientes e condições ambientais. A figura 3 ilustra essas fases e suas principais características.

- **Brotamento:** emissão dos brotos a partir das gemas, dependente de temperatura, umidade e profundidade de plantio.
- **Perfilhamento:** formação de novos colmos laterais, aumentando densidade e biomassa.
- **Crescimento:** alongamento de colmos e folhas, intensa demanda por nitrogênio, fósforo e potássio.
- **Maturação:** desaceleração do crescimento vegetativo e acúmulo de sacarose nos colmos.
- **Floração:** indica maturidade fisiológica, podendo competir com o acúmulo de sacarose; influenciada por fotoperíodo e temperatura.

Figura 3 – Fases da Fenologia da Cana-de-Açúcar.



Fonte: Embrapa, (2022).

2.1.4 Tipos de Plantio

A produção de cana-de-açúcar ocorre por meio de dois tipos principais de plantio: **cana-planta** e **cana-soca**. Cada modalidade apresenta características específicas que impactam diretamente o manejo, a produtividade e a sustentabilidade econômica da lavoura.

O plantio de cana-planta é realizado em áreas recém-preparadas ou após a colheita de outra cultura, utilizando colmos de alta qualidade, selecionados pela sanidade e pelo vigor. Esse cuidado inicial assegura o estabelecimento uniforme da lavoura e contribui para o sucesso do ciclo produtivo.

O ciclo da cana-planta é considerado longo, variando entre 12 e 18 meses até a primeira colheita. Durante esse período, a cultura necessita de condições adequadas de solo, disponibilidade hídrica e manejo fitossanitário eficiente para expressar seu potencial produtivo.

Em termos de investimento, o sistema de cana-planta exige maior aporte inicial, incluindo preparo intensivo do solo, aquisição de insumos e maior atenção aos cuidados sanitários. Embora o custo seja mais elevado, esse investimento é fundamental para garantir o bom desenvolvimento da lavoura.

A produtividade inicial costuma ser alta devido ao vigor do material propagado. Como os colmos utilizados são selecionados e a área é recém-implantada, a cultura encontra condições favoráveis para expressar sua melhor capacidade produtiva nesse primeiro corte.

Entre as principais vantagens da cana-planta estão a menor incidência de doenças e pragas, resultado do plantio em área nova e da qualidade dos colmos utilizados. Além disso, o produtor pode escolher cultivares mais adaptadas às condições locais, aumentando a eficiência da implantação.

Por outro lado, uma desvantagem significativa é que o retorno econômico ocorre apenas após a primeira colheita, o que demanda planejamento financeiro adequado por parte do produtor. O ciclo mais longo e o investimento inicial elevado fazem com que o fluxo de caixa dependa de uma boa gestão até que a lavoura comece a gerar receitas.

O plantio de cana-soca utiliza as brotações que permanecem no campo após a colheita da cana-planta, permitindo dar continuidade ao cultivo sem a necessidade de replantio completa da área. Essa estratégia reduz significativamente os custos de produção e acelera o retorno econômico, já que elimina etapas de preparo intensivo do solo.

O ciclo da cana-soca é mais curto em comparação ao da cana-planta, variando geralmente entre 10 e 14 meses até a colheita. Essa característica torna o processo produtivo mais dinâmico e permite maior frequência de cortes dentro do mesmo período de cultivo.

Em termos de investimento, a cana-soca demanda custos iniciais menores, porque o solo não precisa ser preparado desde o início e muitos dos insumos aplicados anteriormente — como corretivos e fertilizantes — continuam presentes no perfil do solo. Isso resulta em uma economia significativa para o produtor, que consegue manter a área produtiva sem novos gastos elevados.

A produtividade da cana-soca costuma ser moderada e tende a diminuir gradualmente após três ou quatro cortes consecutivos. Essa queda ocorre devido ao desgaste do sistema

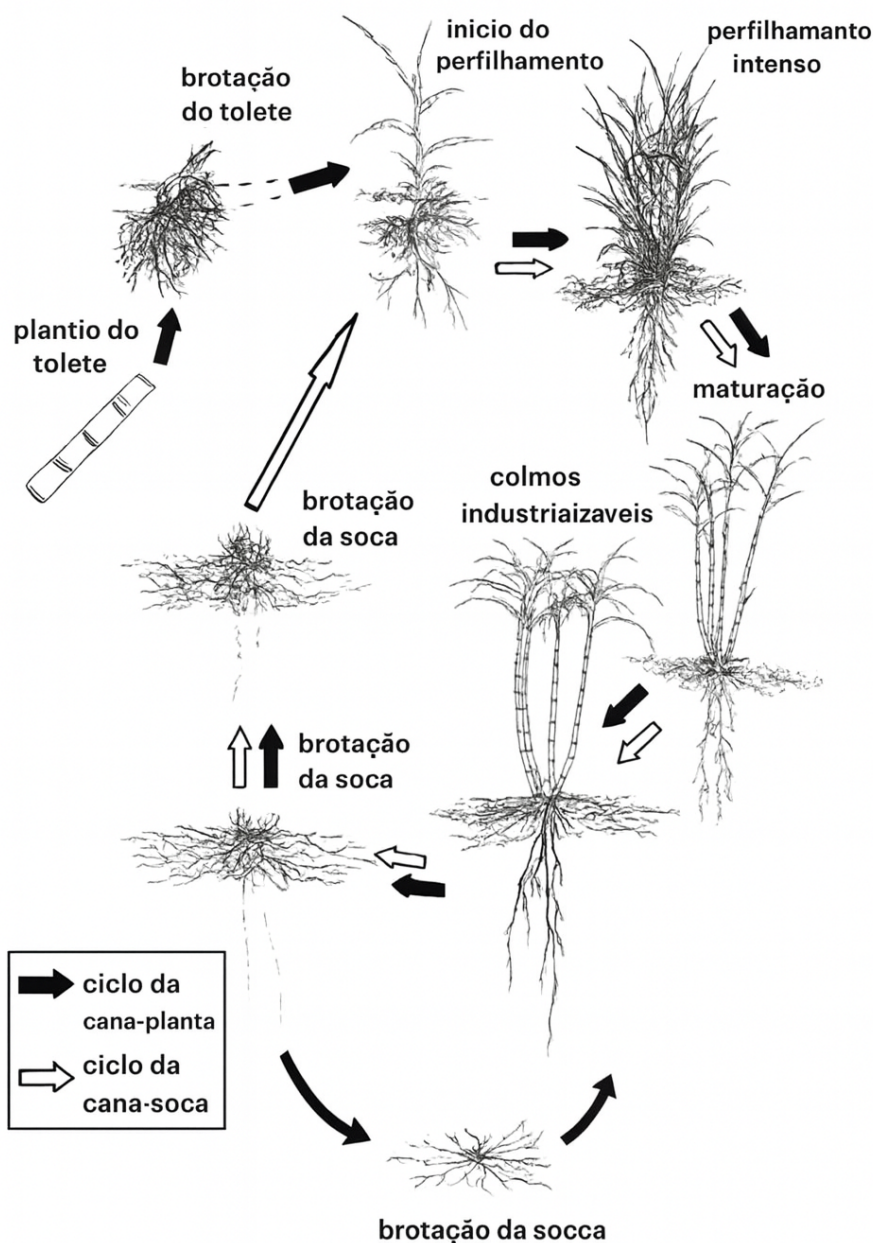
radicular, ao acúmulo de resíduos e à maior incidência de fatores bióticos e abióticos que afetam a vitalidade da cultura ao longo do tempo.

Entre as vantagens desse sistema, destaca-se o retorno econômico mais rápido, já que não há necessidade de uma nova implantação. Além disso, há um aproveitamento eficiente dos nutrientes remanescentes no solo, o que contribui para reduzir custos com adubação nos primeiros ciclos.

Por outro lado, a cana-soca apresenta desvantagens importantes. A cultura torna-se mais suscetível a doenças e pragas, devido ao acúmulo de material vegetal e ao envelhecimento do canavial. Também é comum observar menor uniformidade inicial nas brotações, o que pode afetar o desenvolvimento da lavoura e, conseqüentemente, sua produtividade.

Como ilustrado na figura 4, essas diferenças entre os tipos de plantio impactam diretamente o manejo, a produtividade e o planejamento econômico da lavoura.

Figura 4 – Diferenças entre os tipos de plantio da cana-de-açúcar: cana-planta e cana-soca.



Fonte: Segato, *et al*, (2006)

O quadro 1 apresenta uma comparação direta entre os principais aspectos da cana-planta e da cana-soca, facilitando a visualização das diferenças mais relevantes entre os dois tipos de plantio. Essa comparação complementa a discussão anterior, permitindo compreender como características como ciclo produtivo, investimento, produtividade e manejo influenciam a escolha do sistema de cultivo.

Quadro 1 – Comparação entre os tipos de plantio da cana-de-açúcar

Aspecto	Cana-Planta	Cana-Soca
Ciclo até primeira colheita	12–18 meses	10–14 meses
Investimento inicial	Alto	Baixo
Produtividade inicial	Alta	Moderada
Susceptibilidade a doenças	Baixa	Alta
Uniformidade da lavoura	Alta	Média
Número de cortes	1º corte mais produtivo	Safras subsequentes mais rápidas
Manejo necessário	Planejado e intensivo	Monitoramento constante

Fonte: Elaborado pelo autor.

2.2 Nutrição Mineral e Fertilidade do Solo

A produtividade da cana-de-açúcar está intimamente ligada à disponibilidade e ao equilíbrio de nutrientes no solo, bem como às propriedades químicas que regulam a absorção desses elementos pelas raízes. Devido ao elevado crescimento vegetativo e acúmulo de biomassa, a cultura é considerada altamente exigente em nutrientes, demandando manejo criterioso de fertilização. Estudos recentes destacam que a adoção de fontes organominerais e multinutrientes pode contribuir para maior eficiência na disponibilização de P, K, Ca, Mg e S, inclusive em profundidade, favorecendo o desenvolvimento do sistema radicular e a produtividade de longo prazo (CRUSCIOL et al., 2020; HERRERA; FERNEY, 2019; OLIVEIRA et al., 2018).

2.2.1 Principais Nutrientes e Características Químicas do Solo na Produção da Cana-de-Açúcar

A cana-de-açúcar exige tanto macronutrientes quanto micronutrientes para crescimento, desenvolvimento e produção de colmos. Além das funções individuais, observa-se interação entre eles e influência de estresses ambientais. Os principais nutrientes e atributos químicos do solo considerados neste estudo são descritos a seguir:

- **P** – Essencial no metabolismo energético (ATP/ADP), no desenvolvimento inicial e no crescimento radicular. A baixa mobilidade do P nos solos tropicais, somada à fixação por óxidos de Fe e Al, torna sua disponibilidade um dos principais gargalos para a cultura. A aplicação localizada, o uso de fontes solúveis e de organominerais têm demonstrado eficiência em aumentar a absorção e prolongar os efeitos do fósforo no perfil do solo (CRUSCIOL et al., 2020).

- **M.O** – Atua como condicionador natural do solo, influenciando a retenção de água, a disponibilidade de nutrientes e a atividade microbiana. A mineralização da M.O. libera nitrogênio, fósforo e enxofre de forma gradual, além de reduzir a toxidez de alumínio e aumentar a capacidade de troca catiônica (CTC). Em sistemas de cana crua, a manutenção da palhada após a colheita contribui significativamente para a reposição de M.O. (OLIVEIRA et al., 2018).
- **pH do Solo** – Regula a disponibilidade de nutrientes e a atividade enzimática da raiz. Valores abaixo de 5,5 intensificam a solubilização de Al e Mn em formas tóxicas, enquanto valores acima de 7 podem reduzir a absorção de micronutrientes como Zn, Cu e Mn. O manejo da acidez via calagem e gessagem é fundamental para manter o pH entre 5,5 e 6,5, faixa considerada ideal para a cana (OLIVEIRA et al., 2018).
- **K e %K/T** – Nutriente mais extraído pela cultura, desempenha funções na regulação osmótica, abertura estomática e transporte de açúcares dos colmos para as folhas. O índice K/T reflete a saturação relativa do K na CTC, servindo como indicador de disponibilidade. O uso de fontes multinutrientes e organominerais reduz perdas por lixiviação e aumenta a eficiência da adubação (CRUSCIOL et al., 2020; HERRERA; FERNEY, 2019).
- **Ca e %Ca/T** – Associado ao crescimento de novas raízes e à neutralização do alumínio trocável. Também contribui para a integridade das membranas celulares. O índice Ca/T permite avaliar o equilíbrio entre Ca, Mg e K na CTC, sendo desejável que o cálcio predomine para maior estabilidade estrutural do solo (OLIVEIRA et al., 2018).
- **Mg e %Mg/T** – Constituinte central da clorofila, essencial para a fotossíntese e para a ativação de diversas enzimas. Deficiências provocam clorose internerval em folhas mais velhas. O índice Mg/T avalia a saturação de magnésio na CTC. Fontes como a polyhalita podem fornecer Mg de liberação lenta, favorecendo o equilíbrio com Ca e K (HERRERA; FERNEY, 2019).
- **Al** – Na forma trocável (Al^{3+}), presente em solos ácidos, exerce efeito tóxico sobre as raízes, inibindo seu crescimento e limitando a absorção de nutrientes. A calagem neutraliza esse efeito, aumentando a disponibilidade de Ca e Mg e diminuindo a toxidez (OLIVEIRA et al., 2018).
- **H+Al** – Representa a acidez potencial do solo e é utilizado no cálculo da necessidade de calagem. Valores elevados indicam solos com maior risco de toxidez e baixa saturação por bases, necessitando correções químicas (OLIVEIRA et al., 2018).
- **S** – Participa da síntese de aminoácidos sulfurados (cisteína e metionina), proteínas, vitaminas e coenzimas. Também desempenha papel na eficiência do uso de nitrogênio. A deficiência manifesta-se como clorose uniforme em folhas jovens. Fertilizantes como a

polyhalita oferecem enxofre em formas de liberação gradual, garantindo sua disponibilidade durante o ciclo da cana (HERRERA; FERNEY, 2019).

- **S.B** – Corresponde ao somatório de Ca, Mg, K e Na trocáveis. Valores elevados estão associados a solos de maior fertilidade, com predominância de cátions benéficos na CTC, o que resulta em melhor crescimento radicular e absorção de nutrientes (OLIVEIRA et al., 2018).
- **CTC** – Define o potencial do solo em reter nutrientes catiônicos. Solos com baixa CTC (arenosos) apresentam maior risco de lixiviação, enquanto solos argilosos e ricos em M.O. possuem maior resiliência nutricional. Esse parâmetro é fundamental para o planejamento de doses e fontes de fertilizantes (OLIVEIRA et al., 2018).
- **V** – Reflete a porcentagem da CTC ocupada por bases (Ca, Mg, K e Na). Para a cana, valores entre 60 e 80% são considerados ideais, proporcionando maior equilíbrio nutricional e maior eficiência no desenvolvimento radicular (OLIVEIRA et al., 2018).
- **Sat.** – Refere-se à proporção da CTC ocupada por Al trocável. Valores elevados indicam risco de toxidez e baixa disponibilidade de bases, sendo necessária a aplicação de corretivos de acidez (OLIVEIRA et al., 2018).
- **Micronutrientes:** Zn, B, Cu, Mn, Fe e Mo são vitais em pequenas concentrações. Deficiências podem afetar crescimento, qualidade do colmo e eficiência do N e P, principalmente sob estresses de calor ou seca (AZEVEDO et al., 2011).

Além desses nutrientes e atributos químicos, é importante mencionar, ainda que de forma sucinta, o papel dos **micronutrientes**, que não foram analisados diretamente neste estudo, mas exercem funções vitais na fisiologia da cana. Entre eles destacam-se: **Zn**, atuante na síntese de auxinas e no alongamento celular; **Cu**, essencial para processos enzimáticos de oxirredução; **B**, fundamental na integridade da parede celular e transporte de açúcares; **Mn**, ativador de diversas enzimas e participante da fotólise da água na fotossíntese; **Fe**, indispensável para a formação da clorofila e respiração celular; e **Mo**, essencial no metabolismo do nitrogênio. Deficiências desses elementos, mesmo em pequenas concentrações, podem comprometer significativamente a produtividade e a qualidade da cultura (OLIVEIRA et al., 2018).

Em síntese, a interação entre esses atributos determina não apenas a disponibilidade imediata de nutrientes, mas também a sustentabilidade do sistema de produção da cana-de-açúcar em longo prazo. Evidências mostram que fontes organominerais (CRUSCIOL et al., 2020), adubos multinutrientes como a polyhalita (HERRERA; FERNEY, 2019) e práticas de manejo corretivo de acidez (OLIVEIRA et al., 2018) representam estratégias promissoras para a manutenção da fertilidade e para o incremento da produtividade.

2.2.2 Faixas Ideais de Nutrientes para Cana-de-Açúcar

A Embrapa estabelece faixas de referência consideradas ideais para os principais nutrientes e atributos químicos do solo voltados à cultura da cana-de-açúcar (Embrapa, 2024). Essas faixas representam os níveis necessários para que a planta consiga expressar todo o seu potencial produtivo. A matéria orgânica entre 15 e 25 g/dm³ contribui para melhorar a estrutura do solo, aumentar a retenção de água e fornecer nutrientes de forma gradual. O pH, ideal entre 5,1 e 5,5, garante um ambiente levemente ácido que favorece a absorção dos elementos essenciais, prevenindo tanto a indisponibilidade quanto possíveis toxicidades.

O fósforo, adequado na faixa de 13 a 30 mg/dm³, é indispensável para o desenvolvimento radicular e para os processos energéticos da planta. Já o cálcio, recomendado entre 4 e 7 mmolc/dm³, participa da formação das paredes celulares e ajuda a equilibrar a química do solo. O potássio, situado entre 1,6 e 3,0 mmolc/dm³, exerce papel essencial na regulação hídrica, na tolerância a estresses e na qualidade dos colmos. O magnésio, com valores ideais entre 5 e 8 mmolc/dm³, é elemento central da clorofila e fundamental para a fotossíntese.

A capacidade de troca de cátions (CTC), que deve estar entre 80 e 150 mmolc/dm³, indica a capacidade do solo de reter e disponibilizar nutrientes, sendo determinante para a fertilidade. Por fim, a saturação por bases (V%), com valores entre 51 e 70%, assegura um equilíbrio adequado entre os cátions nutritivos e a acidez do solo, criando um ambiente químico propício ao crescimento vigoroso e eficiente da cana-de-açúcar.

Essas faixas representam os intervalos nos quais os nutrientes promovem condições adequadas de fertilidade do solo e desenvolvimento das plantas. Valores abaixo dos limites inferiores indicam deficiências nutricionais que podem comprometer a produtividade, enquanto concentrações acima dos limites superiores podem resultar em desequilíbrios nutricionais, antagonismos entre elementos ou, em casos extremos, toxidez. O monitoramento periódico desses parâmetros através de análises de solo e a correção quando necessário são práticas essenciais para o manejo eficiente da cultura e para a manutenção da sustentabilidade produtiva ao longo dos ciclos de cultivo.

2.2.3 Distribuição de Nutrientes no Perfil do Solo

A disponibilidade de nutrientes ao longo do perfil do solo depende de sua mobilidade química, da textura e estrutura do solo e da capacidade da cana-de-açúcar em explorar diferentes camadas através de seu sistema radicular. Em geral, nutrientes de baixa mobilidade, como o **P** e o **K**, concentram-se nas camadas superficiais (0–20 cm), devido à baixa lixiviação e à interação com coloides do solo. Por outro lado, elementos mais móveis, como o **N** e o **S**, podem percolar para camadas mais profundas, sendo encontrados até 40 cm ou mais (ROSSETTO et al., 2004).

A distribuição do sistema radicular acompanha essa dinâmica. Azevedo et al. (2011)

observaram que cerca de 70% da densidade de raízes da cana-de-açúcar está concentrada até os 20 cm de profundidade, refletindo a maior disponibilidade de nutrientes e de matéria orgânica nessa camada (AZEVEDO et al., 2011). Confirmaram que a adubação nitrogenada influencia diretamente essa distribuição, aumentando a densidade radicular em camadas superficiais e a eficiência na absorção de nutrientes (OTTO et al., 2009).

Apesar dessa predominância superficial, as raízes em profundidade (20–40 cm) desempenham papel essencial em situações de estresse hídrico, garantindo a absorção de água e de nutrientes residuais em camadas subsuperficiais. Nesse contexto, práticas como a aplicação de **gesso agrícola** são estratégicas, pois favorecem a mobilidade de cálcio e enxofre para maiores profundidades, estimulando o crescimento radicular e aumentando a resiliência da cultura frente a variações climáticas.

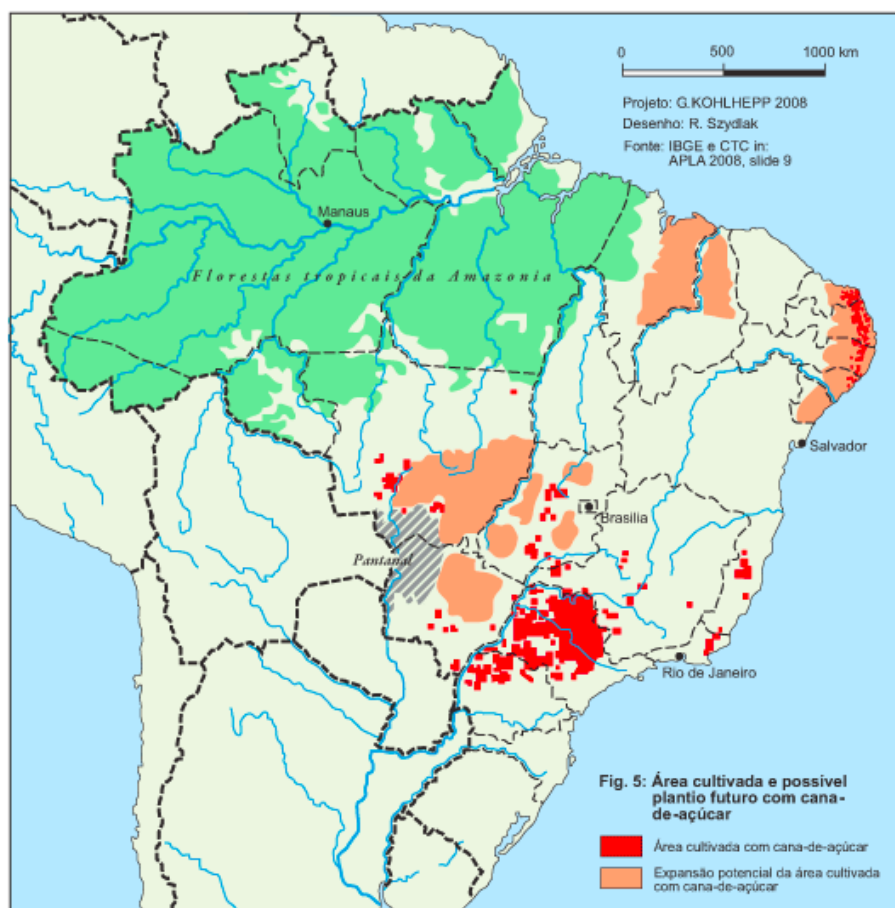
Portanto, a avaliação do solo em diferentes profundidades é indispensável para o manejo da fertilidade em cana-de-açúcar. Enquanto a camada de 0–20 cm reflete a disponibilidade imediata de nutrientes e a resposta direta à adubação, a camada de 20–40 cm garante a sustentabilidade do sistema em longo prazo, especialmente sob condições de déficit hídrico e elevada demanda nutricional.

2.2.4 Fatores Ambientais

Os fatores ambientais exercem influência direta sobre o crescimento e a produtividade da cana-de-açúcar. A temperatura é um dos elementos mais determinantes: valores entre 20 °C e 30 °C são considerados ideais para o desenvolvimento da cultura, enquanto temperaturas abaixo de 15 °C tendem a retardar o crescimento e aquelas acima de 35 °C podem gerar estresse térmico, comprometendo o vigor da planta (Empresa Brasileira de Pesquisa Agropecuária, 2023). A luz e o fotoperíodo também desempenham papel fundamental, já que a cana depende de uma intensidade luminosa adequada e apresenta sensibilidade à duração do dia, fatores que influenciam processos fisiológicos essenciais, como a fotossíntese e a formação dos colmos. A disponibilidade hídrica, por sua vez, precisa estar em equilíbrio: a falta de água reduz o crescimento, prejudica o acúmulo de biomassa e diminui a produtividade, enquanto o excesso pode causar encharcamento, reduzir a oxigenação do solo e favorecer a lixiviação de fertilizantes, diminuindo sua eficiência. Além disso, o clima geral da região onde a cultura é implantada tem grande impacto no desempenho da lavoura. Regiões tropicais e subtropicais normalmente oferecem condições mais favoráveis, mas fatores como vento, umidade relativa e a ocorrência de eventos climáticos extremos também interferem diretamente no ciclo da cana, podendo limitar ou potencializar sua produtividade.

A figura 5 apresenta o mapa do Brasil destacando as regiões com as condições mais favoráveis para o manejo e cultivo da cana-de-açúcar.

Figura 5 – Distribuição das principais regiões produtoras de cana-de-açúcar no Brasil.



Fonte: [SciELO \(2010\)](#)

2.2.5 Manejo da Fertilidade do Solo no Brasil

Os solos tropicais brasileiros, onde a maior parte da cana-de-açúcar é cultivada, apresentam características químicas que impõem desafios ao manejo da fertilidade, como elevada acidez, baixa saturação por bases e limitada disponibilidade de fósforo ([AZEVEDO et al.,](#)). Para superar essas limitações, diferentes práticas agronômicas vêm sendo amplamente utilizadas:

- **Calagem:** prática fundamental para a correção da acidez do solo, aumento da saturação por bases e melhoria da disponibilidade de cálcio e magnésio. Também reduz a toxicidade do alumínio, favorecendo o crescimento radicular em camadas superficiais e subsuperficiais ([CAIRES, 2007](#)).
- **Gessagem:** aplicação de gesso agrícola promove o deslocamento de cálcio e enxofre para camadas mais profundas (20–40 cm), estimulando o desenvolvimento radicular e aumentando a resiliência da planta em condições de estresse hídrico ([FERREZ et al., 2015](#)).

- **Adubação mineral:** a adição de nutrientes, sobretudo N, P e K, em doses adequadas e parceladas ao longo do ciclo, é indispensável para suprir a alta demanda nutricional da cana-de-açúcar. A recomendação deve ser baseada em análises de solo e, quando possível, complementada com análises foliares ([FREIRE et al., 2024](#)).
- **Adubação orgânica:** o uso de resíduos agroindustriais, como a **vinhaça**, e de biofertilizantes tem sido incorporado ao manejo, promovendo reciclagem de nutrientes, aumento da matéria orgânica e melhoria da capacidade de retenção de água no solo ([OLIVEIRA et al., 2009](#)).
- **Rotação de culturas e plantas de cobertura:** contribuem para a melhoria da estrutura do solo, aumento da atividade biológica e reposição de nutrientes. Leguminosas, em especial, podem fixar nitrogênio atmosférico, reduzindo a necessidade de fertilizantes sintéticos.

Essas práticas integradas são determinantes para o aumento da produtividade e sustentabilidade do setor. Estudos apontam que, em solos devidamente corrigidos e adubados, a produtividade pode superar **100–120 t ha⁻¹** de colmos, enquanto em áreas mal manejadas dificilmente ultrapassa **60–70 t ha⁻¹** ([Embrapa, 2024](#)). Assim, o manejo adequado da fertilidade representa um dos principais diferenciais competitivos da produção brasileira de cana-de-açúcar.

2.3 Ciência de Dados

A Ciência de Dados é um campo interdisciplinar que surgiu da necessidade de extrair conhecimento útil e informações relevantes a partir de grandes volumes de dados, muitas vezes complexos e heterogêneos ([PROVOST; FAWCETT, 2013](#)). Combinando técnicas estatísticas, matemáticas e computacionais, a ciência de dados permite identificar **padrões, relações e tendências** que não seriam perceptíveis apenas por meio de observação direta ou análise manual.

O desenvolvimento da Ciência de Dados está intimamente ligado à evolução da tecnologia da informação e à crescente disponibilidade de dados digitais. Inicialmente, métodos estatísticos tradicionais eram utilizados de forma isolada para análise de pequenas bases de dados. No entanto, o aumento exponencial do volume de dados coletados em diferentes contextos — como pesquisas agrícolas, monitoramento de colheitas, análises de fertilidade do solo e registros de produtividade — exigiu metodologias integradas capazes de processar, organizar e interpretar essas informações de maneira eficiente e confiável.

No contexto deste trabalho, a Ciência de Dados é essencial para a análise da produção de cana-de-açúcar. Através dela, é possível estudar o impacto de práticas de manejo, correção de fertilidade, adubação e condições ambientais sobre a produtividade, identificar relações entre nutrientes do solo e crescimento da planta, além de avaliar a eficácia de diferentes

estratégias de manejo agrônomo. Sem a aplicação de princípios de ciência de dados, seria inviável organizar, sintetizar e interpretar adequadamente a complexidade das informações coletadas, comprometendo a geração de insights confiáveis que possam orientar decisões técnicas e estratégicas no cultivo da cana-de-açúcar.

Dessa forma, a Ciência de Dados fornece o arcabouço metodológico necessário para transformar dados brutos em conhecimento aplicado, permitindo que os resultados das análises contribuam para o aumento da produtividade, sustentabilidade e eficiência do sistema de produção de cana-de-açúcar.

2.3.1 Dados Tabulares: *Comma-Separated Values (CSV)*

Para realizar análises quantitativas na produção de cana-de-açúcar, os dados são frequentemente organizados em formatos tabulares, sendo o CSV (*Comma-Separated Values*) um dos mais utilizados. Este formato consiste em arquivos de texto onde cada linha representa uma observação (ou registro) e cada coluna representa uma variável (ou atributo), separados por vírgulas, ponto e vírgula ou outro delimitador.

O uso de arquivos CSV permite integração com ferramentas de análise de dados, como Python (*pandas*, *numpy*, *matplotlib*) e R, facilitando a manipulação, limpeza, exploração e visualização de dados. Em pesquisas agrícolas, os CSVs podem conter informações sobre produtividade, tipos de plantio, adubação, nutrientes do solo e outras variáveis que fundamentam análises estatísticas.

2.3.2 Análise Exploratória dos Dados (AED)

A Análise Exploratória dos Dados é uma etapa inicial e fundamental dentro da ciência de dados, com o objetivo de compreender a **estrutura**, **características** e **padrões** dos dados antes da aplicação de modelos ou técnicas inferenciais (TUKEY; *et al.*, 1977).

No contexto da produção de cana-de-açúcar, a Análise Exploratória de Dados (AED) permite examinar informações sobre produtividade, características do solo, tipos de plantio e distribuição de nutrientes. Essa abordagem é essencial para detectar padrões e tendências nos dados, como a forma pela qual a aplicação de adubos ou a correção do solo influenciam a produtividade da lavoura.

Além disso, a AED possibilita identificar *outliers*, ou seja, colheitas ou áreas com produtividade atípica que podem indicar problemas de manejo, falhas no registro de dados ou condições ambientais excepcionais. Compreender a distribuição dos dados também é fundamental, permitindo verificar normalidade, assimetria e curtose, informações que orientam a escolha de testes estatísticos e garantem a validade das análises.

A análise exploratória também é útil para examinar correlações entre diferentes variáveis, possibilitando explorar relações entre fatores de manejo agrícola, propriedades do solo e

produtividade. Essa compreensão contribui para decisões mais embasadas e estratégicas no planejamento das atividades agrícolas.

Entre as principais técnicas e ferramentas utilizadas na AED, destacam-se as visualizações gráficas, como histogramas, boxplots, scatterplots, pairplots e mapas de calor (*heatmaps*) de correlação, que permitem uma interpretação visual imediata dos padrões presentes nos dados. Somados a isso, os sumários estatísticos, incluindo médias, medianas, quartis, desvios padrão, intervalos interquartis, assimetria e curtose, fornecem medidas quantitativas que complementam a análise visual. Por fim, a identificação de valores ausentes ou inconsistentes é um passo essencial para garantir a qualidade dos dados e a confiabilidade das conclusões obtidas.

2.3.3 Fundamentos Estatísticos e Matemáticos

2.3.3.1 Medidas de Tendência Central e Dispersão

As medidas de tendência central e dispersão permitem compreender a distribuição dos dados e sua variabilidade, essenciais antes de aplicar modelos inferenciais.

- **Média:** A média aritmética é calculada conforme a Equação 2.1.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad (2.1)$$

onde:

- x_i : valor da i -ésima observação
- n : número total de observações
- $\bar{x}$: média dos valores, representando o valor típico da amostra,
- **Mediana:** valor central dos dados ordenados, menos sensível a valores extremos, útil quando há *outliers*.
- **Quartis:** Q_1, Q_2, Q_3 dividem os dados em quatro partes iguais:
 - Q_1 : 25% dos dados abaixo
 - Q_2 : mediana, 50% dos dados abaixo
 - Q_3 : 75% dos dados abaixo
- **Intervalo Interquartil (IQR):** O IQR é calculado conforme a Equação 2.2.

$$IQR = Q_3 - Q_1, \quad (2.2)$$

onde:

- IQR : mede a dispersão central (50% dos dados)

- **Desvio-padrão:** O desvio-padrão é calculado conforme a Equação 2.3.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}, \quad (2.3)$$

onde:

- s : quantifica a variação média em torno da média, em mesmas unidades da variável

- **Coeficiente de variação:** O coeficiente de variação é calculado conforme a Equação 2.4.

$$CV = \frac{s}{\bar{x}} \times 100\%, \quad (2.4)$$

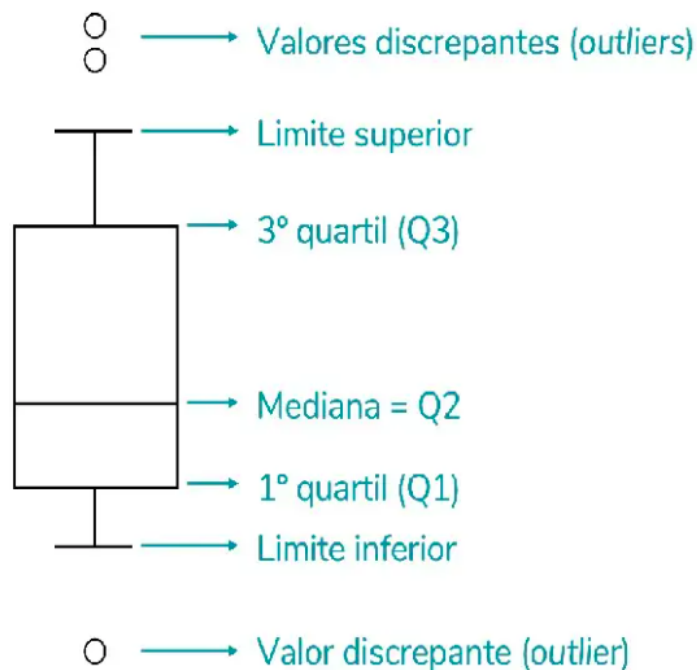
onde:

- CV : dispersão relativa em porcentagem, permitindo comparação entre variáveis com diferentes magnitudes

2.3.3.2 Boxplots

O **boxplot** facilita a visualização de quartis e *outliers*, oferecendo uma interpretação gráfica da dispersão dos dados. Como ilustrado na figura 6, esse tipo de gráfico permite identificar rapidamente valores atípicos e compreender a distribuição de um conjunto de dados.

Figura 6 – Exemplo de Boxplot



Fonte: SiteWare - Boxplot

- Caixa: Q_1 a Q_3 , mostrando o núcleo central dos dados

- Linha central: mediana (Q_2)
- Bigodes: estendem-se até $1,5 \times IQR$; valores fora desse intervalo são potenciais *outliers*
- *Outliers*: pontos extremos que podem indicar erros de medição ou condições excepcionais de produtividade

2.3.3.3 Correlação e Estatística Inferencial

- **Correlação de Pearson:** A correlação de Pearson é calculada conforme a Equação 2.5.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad -1 \leq r \leq 1, \quad (2.5)$$

onde:

- x : valores da primeira variável,
 - y : valores da segunda variável,
 - $\bar{x}$: média dos valores de x ,
 - $\bar{y}$: média dos valores de y ,
 - r : intensidade e direção da relação linear entre duas variáveis,
 - $r = 1$: correlação perfeita positiva; $r = -1$: correlação perfeita negativa; $r = 0$: ausência de relação linear.
- **Shapiro-Wilk:** O teste de Shapiro-Wilk avalia a normalidade dos dados por meio do estatístico W , definido por:

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(n+1-i)} - a_{n+1-i} x_{(i)} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (2.6)$$

onde:

- $x_{(i)}$ são os valores da amostra ordenados de forma crescente;
- $\bar{x}$ é a média amostral;
- a_i são coeficientes derivados dos valores esperados dos *order statistics* de uma distribuição normal padrão.

O numerador mede o grau de alinhamento entre os valores observados e aqueles esperados sob normalidade, enquanto o denominador representa a variabilidade total da amostra. Valores de W próximos de 1 indicam forte aderência a normalidade. A interpretação final é feita com base no p-value do teste.

- **Teste t para duas amostras independentes** A estatística do teste t é calculada conforme as Equações 2.7 e 2.8.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (2.7)$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}, \quad (2.8)$$

onde:

- $\bar{x}_1, \bar{x}_2$: médias das amostras.
 - s_1, s_2 : desvios-padrão das amostras.
 - n_1, n_2 : tamanhos das amostras.
 - s_p : desvio-padrão combinado.
 - t : estatística do teste t, comparada à tabela t para inferência.
- **Variância entre e dentro dos grupos** As variâncias entre os grupos e dentro dos grupos são definidas matematicamente pela Equação 2.9:

$$S_B^2 = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2}{k - 1}, \quad S_W^2 = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{N - k}, \quad (2.9)$$

onde:

- S_B^2 : variância entre os grupos,
 - S_W^2 : variância dentro dos grupos,
 - k : número de grupos,
 - n_j : tamanho do j -ésimo grupo,
 - x_{ij} : valor da i -ésima observação do j -ésimo grupo,
 - $\bar{x}_j$: média do j -ésimo grupo,
 - $\bar{x}$: média geral de todas as observações,
 - N : número total de observações ($N = \sum_{j=1}^k n_j$).
- **Intervalos de Confiança** O intervalo de confiança da média é dado pela Equação 2.10.

$$\bar{x} \pm z \frac{s}{\sqrt{n}}, \quad (2.10)$$

onde:

- $\bar{x}$: média da amostra.

- s : desvio-padrão da amostra.
- n : tamanho da amostra.
- z : valor crítico da distribuição normal.

- **ANOVA (Análise de Variância)** A estatística F da ANOVA é definida pela Equação 2.11.

$$F = \frac{\text{variância entre grupos}}{\text{variância dentro dos grupos}}, \quad (2.11)$$

onde:

- F : estatística que indica se há diferença significativa entre médias de três ou mais grupos.
- Variância entre grupos: diferenças entre as médias dos grupos.
- Variância dentro dos grupos: variação interna dentro de cada grupo.

2.3.4 Integração entre Agricultura e Ciência de Dados

O avanço tecnológico no setor agrícola tem impulsionado o conceito de **agricultura digital**, caracterizada pelo uso intensivo de dados e ferramentas computacionais para apoiar a tomada de decisão no campo. Dentro desse contexto, a **agricultura de precisão** surge como uma prática que busca otimizar o uso de insumos (água, fertilizantes, defensivos) e maximizar a produtividade, reduzindo custos e impactos ambientais (MOLIN; AMARAL; COLAÇO, 2015).

O emprego de *big data* no agronegócio possibilita a integração de diferentes fontes de informação, como dados climáticos, características do solo, manejo nutricional e histórico de safras. Essa integração permite tanto o monitoramento em tempo real quanto a previsão de tendências, possibilitando decisões mais assertivas no planejamento agrícola (WOLFERT et al., 2017).

Além disso, a incorporação de técnicas de **Inteligência Artificial (IA)** e **Aprendizado de Máquina** no setor agrícola amplia a capacidade de análise, permitindo a construção de modelos preditivos e sistemas de recomendação. Essas abordagens possibilitam desde a detecção precoce de doenças em plantas até a previsão da produtividade em função de variáveis ambientais e de manejo (LIAKOS et al., 2018).

Assim, a integração entre agricultura e ciência de dados estabelece uma base sólida para o desenvolvimento de sistemas computacionais que apoiem os produtores rurais na gestão eficiente e sustentável de suas áreas de cultivo.

2.3.5 Modelagem Estatística e Aprendizado de Máquina

No âmbito da análise de dados aplicada à agricultura, a utilização de modelos estatísticos e de aprendizado de máquina é fundamental para transformar informações brutas em

conhecimento útil. Entre as técnicas mais tradicionais está a **Regressão Linear Múltipla**, que busca modelar a relação entre uma variável dependente (como a produtividade da cana-de-açúcar) e múltiplas variáveis independentes (como as quantidades de nutrientes presentes no solo).

O modelo geral da regressão linear múltipla é expresso conforme a Equação 2.12.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon, \quad (2.12)$$

onde:

- Y : variável dependente (no caso deste trabalho, a produtividade da cana-de-açúcar);
- β_0 : intercepto do modelo, ou seja, o valor esperado de Y quando todas as variáveis independentes são iguais a zero;
- β_i ($i = 1, 2, \dots, p$): coeficientes de regressão que representam o efeito médio de cada variável independente sobre Y , mantendo as demais constantes;
- X_i ($i = 1, 2, \dots, p$): variáveis independentes, que neste trabalho correspondem às quantidades de nutrientes disponíveis no solo;
- ε : termo de erro aleatório, que captura a variação de Y não explicada pelo modelo.

A estimativa dos coeficientes β_i é geralmente realizada pelo método dos **Mínimos Quadrados Ordinários (MQO)**, que busca minimizar a soma dos quadrados dos resíduos, conforme a Equação 2.13.

$$\text{SQRes} = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (2.13)$$

onde:

- y_i : valor observado da variável dependente para a i -ésima observação;
- $\hat{y}_i$: valor estimado pelo modelo para a i -ésima observação;
- n : número total de observações.

Esse tipo de modelagem permite não apenas estimar a influência individual de cada nutriente sobre a produtividade, mas também prever os valores esperados de Y a partir de diferentes cenários de manejo nutricional. Dessa forma, fornece suporte para estratégias de adubação mais precisas e eficientes (BUSSAB; MORETTIN, 2010).

Além das técnicas estatísticas clássicas, modelos de aprendizado de máquina mais complexos, como **Random Forest** e **Redes Neurais**, têm sido aplicados na agricultura para previsão de safras e análise de produtividade, permitindo modelagens mais robustas em cenários com grande volume de dados ou relações não lineares entre variáveis.

3 Metodologia

A metodologia adotada neste trabalho foi estruturada em etapas que integraram análise estatística, análise inferencial, modelagem preditiva e desenvolvimento de sistemas computacionais para aplicação prática dos resultados.

3.1 Ferramentas Utilizadas

As análises estatísticas, inferenciais e de modelagem foram conduzidas em *Jupyter Notebook*, ambiente que permitiu a execução interativa de códigos em *Python* e a integração com diferentes bibliotecas. Esse ambiente possibilitou não apenas a manipulação e visualização dos dados, mas também a reprodutibilidade dos experimentos realizados.

3.2 Bibliotecas Empregadas

Diversas bibliotecas em Python foram utilizadas para dar suporte às etapas de análise, modelagem e visualização dos resultados. O quadro 2 apresenta as principais bibliotecas e suas finalidades.

Quadro 2 – Bibliotecas utilizadas e suas respectivas finalidades

Biblioteca	Finalidade
<i>pandas, numpy</i>	Manipulação, tratamento e organização dos dados.
<i>matplotlib, seaborn</i>	Construção de gráficos e visualização exploratória.
<i>scipy.stats, statsmodels</i>	Aplicação de testes estatísticos (ex.: correlação de Spearman, ANOVA, teste t) e análise inferencial.
<i>scikit-learn</i>	Implementação de algoritmos de aprendizado de máquina, incluindo regressão linear múltipla, padronização de variáveis (StandardScaler) e redução de dimensionalidade (PCA).
<i>RandomForestRegressor, RandomForestClassifier</i>	Modelos auxiliares para análise de importância de variáveis.
<i>shap</i>	Interpretação dos modelos, identificando a contribuição individual de cada variável nas previsões.
<i>semopy</i>	Modelagem de equações estruturais exploratórias.

Fonte: Elaborado pelo autor.

3.3 Abordagem Estatística

O trabalho foi fundamentado em dois ramos principais da Estatística:

- **Estatística Descritiva:** utilizada para organizar, resumir e explorar os dados, fornecendo uma visão inicial sobre a distribuição e variabilidade das variáveis observadas (GUEDES et al., 2005).
- **Estatística Inferencial:** empregada para generalizar os resultados obtidos a partir da amostra para a população, permitindo avaliar relações entre variáveis e validar hipóteses por meio de técnicas como testes t e ANOVA (BUSSAB; MORETTIN, 2010).

Essa combinação forneceu suporte tanto para a compreensão inicial do conjunto de dados quanto para a avaliação da significância estatística das relações observadas.

3.4 Modelagem Preditiva

Para a etapa de modelagem, foi adotado o modelo de **Regressão Linear Múltipla**, escolhido por sua capacidade de avaliar simultaneamente a influência de múltiplas variáveis independentes, como nutrientes e fatores ambientais, sobre a variável dependente, que neste caso é a produtividade da cana-de-açúcar. Essa abordagem permite compreender como cada fator contribui individualmente e em conjunto para os resultados observados na lavoura.

A utilização da regressão linear múltipla possibilitou estimar coeficientes que indicam a contribuição de cada variável explicativa, fornecendo uma visão detalhada da importância relativa dos diferentes fatores sobre a produtividade. Além disso, o modelo permite quantificar o impacto individual e combinado desses fatores, oferecendo informações valiosas para otimizar o manejo agrícola e as intervenções no solo.

Outra vantagem do modelo é a possibilidade de realizar previsões de valores futuros de produção, com base nos dados observados e nas relações identificadas entre as variáveis. Essa capacidade de previsão contribui para o planejamento estratégico da lavoura, permitindo antecipar resultados e ajustar práticas de manejo de forma mais eficiente.

A implementação foi realizada com o auxílio das bibliotecas `scikit-learn` e `statsmodels`, que permitiram tanto o ajuste do modelo quanto a avaliação de sua significância estatística por meio de métricas como R^2 , erro quadrático médio (MSE) e testes de hipóteses para os coeficientes.

3.5 Desenvolvimento de Sistemas

Com o objetivo de aplicar os resultados das análises em um contexto prático e acessível aos produtores, foram desenvolvidas duas soluções tecnológicas:

3.5.1 AgroApp: Aplicativo Mobile

O AgroApp foi desenvolvido e projetado para oferecer ao produtor rural acesso rápido e intuitivo às principais análises estatísticas realizadas sobre a lavoura. A ferramenta permite prever a produção com base em dados históricos e variáveis de manejo, fornecendo informações estratégicas que ajudam na tomada de decisão e no planejamento das atividades agrícolas.

Além disso, o aplicativo possibilita acompanhar o histórico de vendas e o desempenho financeiro da fazenda, permitindo ao produtor ter uma visão consolidada das receitas, despesas e resultados econômicos da propriedade. O AgroApp também oferece funcionalidades para controlar o estoque de fertilizantes, defensivos, sementes e equipamentos, garantindo que os insumos estejam sempre disponíveis quando necessários e evitando desperdícios ou faltas.

De forma geral, a plataforma auxilia no gerenciamento integral da propriedade, centralizando informações e simplificando o acompanhamento de diversas atividades. Com isso, o produtor pode tomar decisões mais embasadas, otimizar recursos e aumentar a eficiência operacional da lavoura.

A aplicação foi desenvolvida em React Native, um framework baseado em JavaScript que permite criar aplicativos nativos para Android e iOS a partir de um único código-fonte. Essa abordagem proporciona maior eficiência no desenvolvimento e na manutenção, evitando a necessidade de duplicar código para diferentes plataformas e garantindo uma base única e consistente.

Além disso, o framework facilita a integração com APIs externas, permitindo o consumo de dados de forma simplificada e eficiente. O uso de bibliotecas modernas, como Redux para gerenciamento de estado e Axios para requisições HTTP, contribui para a organização do código e garante uma comunicação confiável entre o aplicativo e o servidor.

O React Native também possibilita a integração com bibliotecas gráficas e componentes nativos, o que permite a criação de visualizações de dados interativas e interfaces ricas para o usuário. Combinado ao JavaScript, o framework permite aplicar conceitos de programação funcional, reatividade e modularidade diretamente no desenvolvimento, facilitando futuras expansões e atualizações do aplicativo.

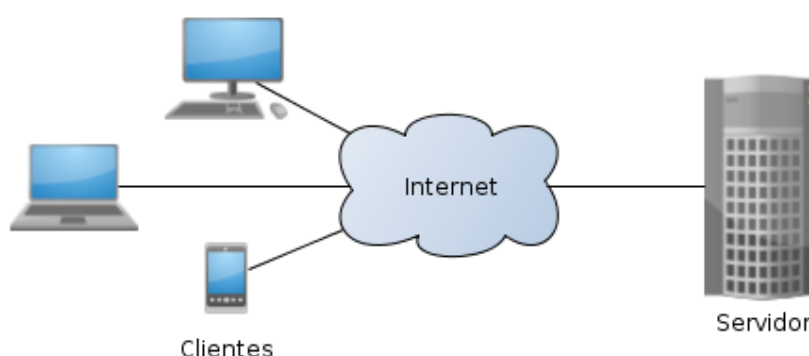
3.5.1.1 Arquitetura do Sistema

O AgroApp segue a **arquitetura cliente-servidor**, onde o *cliente* (aplicativo no celular) solicita informações e serviços ao *servidor* (backend que processa dados e envia respostas).

- **Cliente:** exibe os dados, recebe interações do usuário e envia requisições ao servidor.
- **Servidor:** processa as requisições, realiza cálculos e consultas, e retorna os resultados ao cliente.

Essa abordagem mantém o aplicativo leve, concentra o processamento no servidor e facilita atualizações e manutenção do sistema. Como mostrado na figura 7, a arquitetura distribui claramente as responsabilidades entre cliente e servidor, permitindo um desempenho mais eficiente e melhor gerenciamento do sistema.

Figura 7 – Arquitetura cliente-servidor empregada no AgroApp.



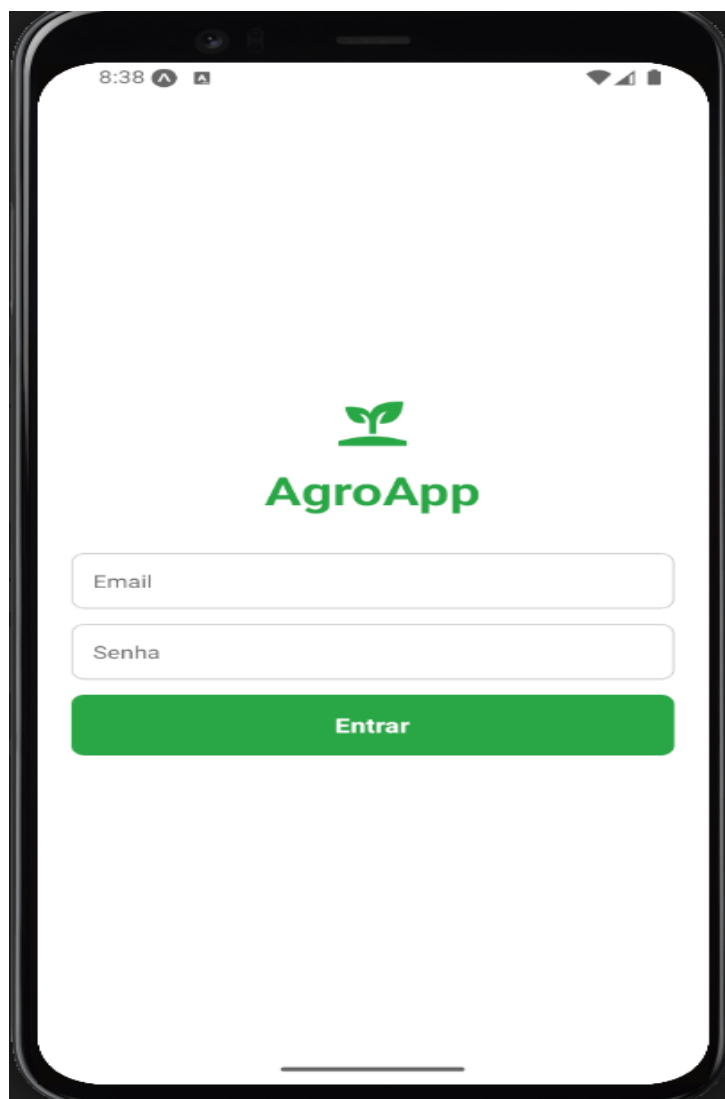
Fonte: [DevTech - Arquitetura Cliente-Servidor](#)

3.5.1.2 Interfaces e Funcionalidades

O AgroApp foi desenvolvido com foco em usabilidade, clareza visual e acesso rápido às principais informações do sistema. As telas do aplicativo foram organizadas de forma a facilitar o uso em campo, com interface intuitiva e cards informativos que sintetizam indicadores essenciais para o produtor.

A tela de login (figura 8) é a porta de entrada do sistema, garantindo a autenticação segura dos usuários. Ela permite o acesso individualizado de cada produtor, preservando a privacidade e os dados operacionais da fazenda.

Figura 8 – Tela de Login do AgroApp.



Fonte: Elaborada pelo autor.

A tela inicial do aplicativo (figura 9) é composta por **cards interativos**, cada um representando uma funcionalidade específica voltada à gestão da fazenda. Esses cards foram projetados para oferecer ao usuário uma visão panorâmica do negócio, permitindo acesso direto a cada módulo do sistema. Entre os principais cards, destacam-se:

- **Gestão de Estoque:** controla os insumos disponíveis, como sementes, fertilizantes, defensivos e ferramentas, auxiliando na organização do almoxarifado da fazenda;
- **Vendas e Clientes:** exibe o histórico de vendas realizadas, os volumes comercializados e o número de clientes atendidos, permitindo acompanhar o desempenho comercial;
- **Clima:** apresenta a previsão meteorológica diária e semanal, incluindo informações sobre chuva, temperatura e velocidade do vento, fundamentais para o planejamento de

atividades agrícolas;

- **Produção:** possibilita a predição da produtividade com base nas quantidades de nutrientes informadas pelo usuário, oferecendo uma estimativa do rendimento esperado;
- **Análise do Solo:** permite a inserção de dados nutricionais e indica, automaticamente, quais nutrientes estão acima, abaixo ou dentro do ideal, recomendando produtos de correção conforme a necessidade;
- **Relatórios:** reúne diversos cards de análise de dados, onde o usuário pode visualizar gráficos e estatísticas consolidadas da fazenda.

Figura 9 – Tela inicial (Home) do AgroApp com cards de gestão e análise.

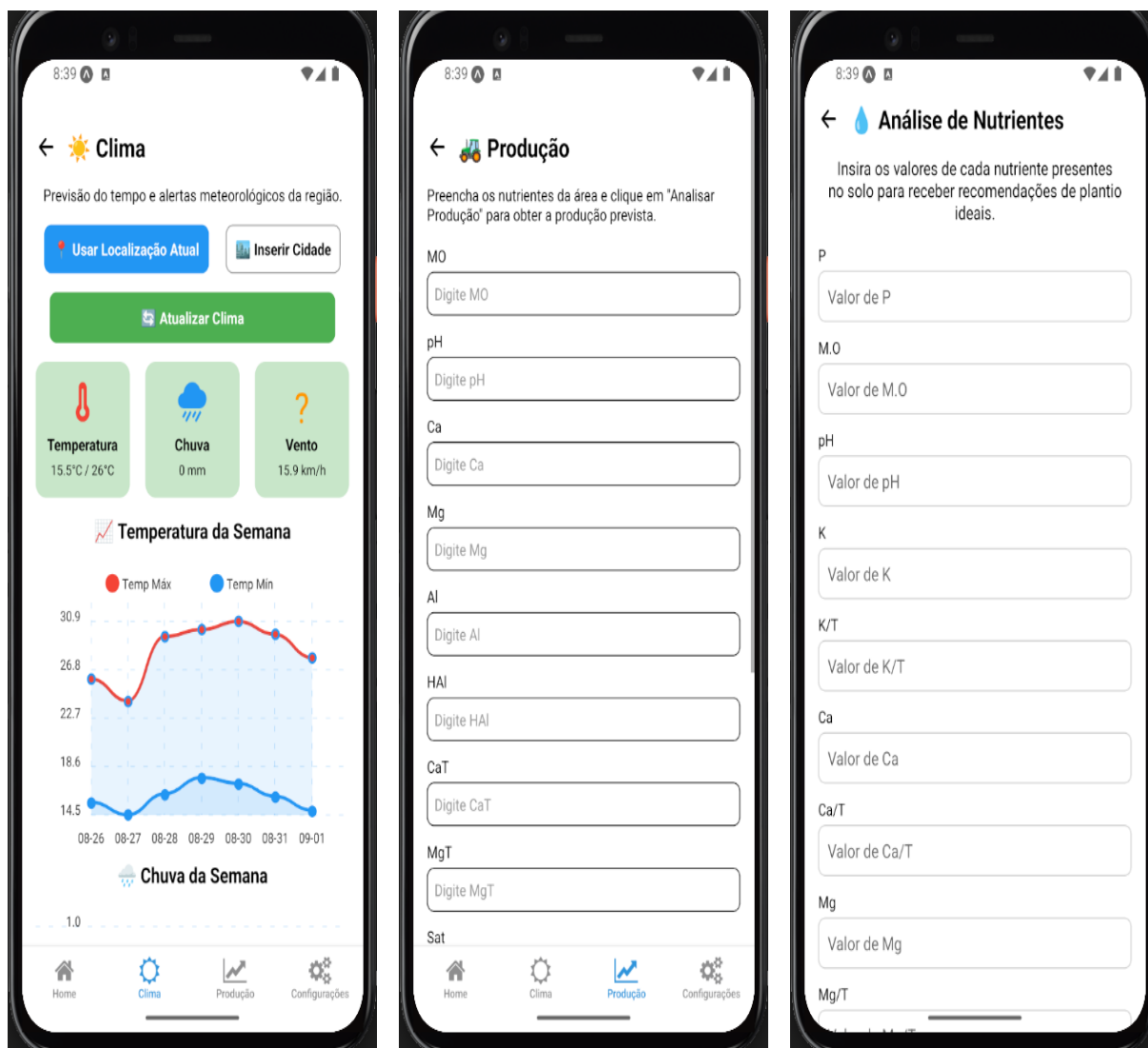


Fonte: Elaborada pela autor.

O módulo de **Clima**, figura 10 (a), disponibiliza dados meteorológicos integrados e constantemente atualizados via API, fornecendo informações essenciais para o planejamento de

atividades agrícolas, como irrigação e colheita. O módulo de **Produção**, figura 10 (b), permite estimar a produtividade esperada a partir dos níveis de nutrientes informados pelo usuário, contribuindo para o acompanhamento do potencial produtivo da fazenda. Por fim, o módulo de **Análise do Solo**, figura 10 (c), avalia o estado nutricional do solo, identificando quais nutrientes estão acima, abaixo ou dentro do intervalo ideal, e recomenda produtos corretivos ou fertilizantes conforme a necessidade. Esses módulos podem ser visualizados na figura 10.

Figura 10 – Exemplos de telas dos módulos de Clima, Produção e Análise do Solo.



(a) Módulo de Clima.

(b) Módulo de Predição.

(c) Módulo de Análise do Solo.

Fonte: Elaborada pelo autor.

A seção de **Relatórios** do AgroApp (figura 11) é composta por múltiplos cards de análise, nos quais são apresentadas informações consolidadas sobre o desempenho das fazendas. Esses cards reúnem os principais resultados obtidos pelas análises estatísticas e predições, permitindo ao produtor compreender de forma prática a situação atual de sua propriedade.

As análises foram organizadas em diferentes grupos temáticos, cada um voltado para uma dimensão específica da gestão agrícola. O grupo **Nutrientes** apresenta indicadores individuais de macro e micronutrientes do solo, permitindo identificar quais elementos estão dentro, acima ou abaixo do ideal. Essa visão detalhada auxilia na correção de deficiências e no uso racional de insumos, contribuindo para a otimização da fertilidade do solo.

O grupo **Fertilidade** reúne variáveis que influenciam diretamente o potencial produtivo do solo, como pH, M.O, CTC e S.B. Essas informações permitem avaliar a qualidade e o equilíbrio químico do solo, fornecendo subsídios importantes para o planejamento das práticas de manejo.

O grupo **Resumo e Indicadores** sintetiza as informações mais relevantes em painéis comparativos, oferecendo uma visão geral do desempenho da fazenda e dos resultados obtidos com as práticas de manejo. Já o grupo Relações entre Nutrientes apresenta correlações e interações entre diferentes nutrientes, identificando padrões de dependência que podem influenciar a produtividade e a eficiência da adubação.

O grupo **Comparações Espaciais** permite analisar diferenças entre lotes ou quadras, evidenciando regiões com maior ou menor desempenho produtivo. Por sua vez, o grupo Análises Temporais possibilita acompanhar a evolução das variáveis ao longo do tempo, destacando tendências e efeitos de práticas de manejo ou condições climáticas sobre os resultados agrícolas.

Essa estrutura modular torna os relatórios mais dinâmicos e interpretáveis, permitindo que o produtor selecione e visualize rapidamente o tipo de análise desejada. Dessa forma, o AgroApp não apenas disponibiliza dados, mas transforma-os em conhecimento aplicável à tomada de decisão, favorecendo uma agricultura mais precisa e sustentável.

Figura 11 – Exemplo de cards de análise e relatórios no AgroApp.



Fonte: Elaborada pelo autor.

3.5.2 AgroAnalysis: Sistema Web

O AgroAnalysis é um sistema web desenvolvido e projetado para oferecer acesso multiusuário mediante autenticação segura. A plataforma permite que os usuários acessem de forma detalhada todas as análises realizadas no estudo, abrangendo desde estatística descritiva e inferencial até previsões de produção, proporcionando uma visão completa do desempenho da lavoura.

O sistema disponibiliza gráficos interativos e relatórios consolidados que resumem os resultados das análises, facilitando a interpretação das informações e permitindo uma avaliação rápida das variáveis críticas. Além disso, oferece ferramentas para filtragem, comparação e exportação de dados, de modo a apoiar decisões estratégicas e a implementação de práticas agrícolas mais eficientes.

O AgroAnalysis também inclui funcionalidades de gestão de usuários e permissões, garantindo que o acesso às informações seja controlado de acordo com o perfil de cada usuário. Ademais, o sistema integra-se ao AgroApp, permitindo a sincronização de dados da fazenda e a atualização em tempo real das análises, promovendo consistência e agilidade na gestão da informação.

A aplicação foi construída utilizando *React*, que permite a criação de interfaces reativas e componentes reutilizáveis, garantindo organização, modularidade e facilidade de manutenção do código. Para o processo de *build*, adotou-se o *Vite*, uma ferramenta moderna que oferece carregamento extremamente rápido durante o desenvolvimento, permitindo atualizações instantâneas. Além disso, o *Vite* realiza otimizações avançadas no código final destinado à produção, resultando em aplicações web mais leves, eficientes e com desempenho superior.

A arquitetura do sistema integra-se facilmente a diversas bibliotecas do ecossistema *JavaScript*, o que possibilita o uso de ferramentas como *Recharts* para visualização gráfica, *Fetch* para consumo de *APIs* e soluções de gerenciamento de estados, fundamentais para manter a sincronização dos dados entre os componentes da interface.

Graças à combinação de *React* com *Vite*, o AgroAnalysis consegue oferecer uma experiência fluida, responsiva e dinâmica, suportando a visualização de análises estatísticas complexas, relatórios detalhados e previsões de produção de maneira intuitiva, rápida e acessível ao usuário.

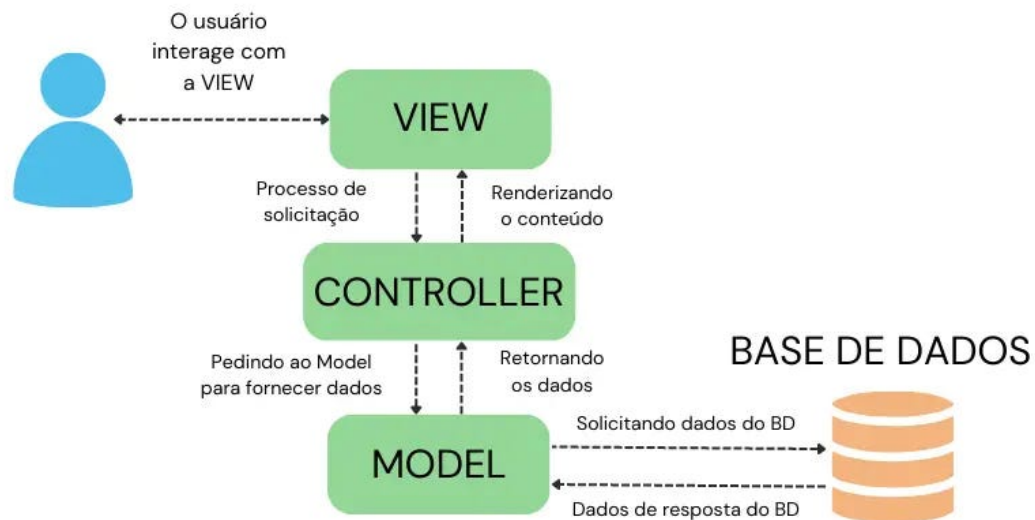
3.5.2.1 Arquitetura do Sistema

O AgroAnalysis utiliza a **arquitetura MVC (Model-View-Controller)**, que separa a aplicação em três componentes principais:

- **Model (Modelo):** gerenciamento dos dados e lógica de negócio, incluindo análises estatísticas, usuários e relatórios;
- **View (Visão):** interface com o usuário, construída em React, exibindo gráficos e relatórios interativos;
- **Controller (Controlador):** intermediário entre Model e View, processando requisições e atualizando a interface conforme necessário.

Essa arquitetura organiza a aplicação em componentes com responsabilidades bem definidas, permitindo maior organização do código, separação de responsabilidades e facilitando a manutenção do sistema, como ilustrado na figura 12.

Figura 12 – Arquitetura MVC adotada no AgroAnalysis.



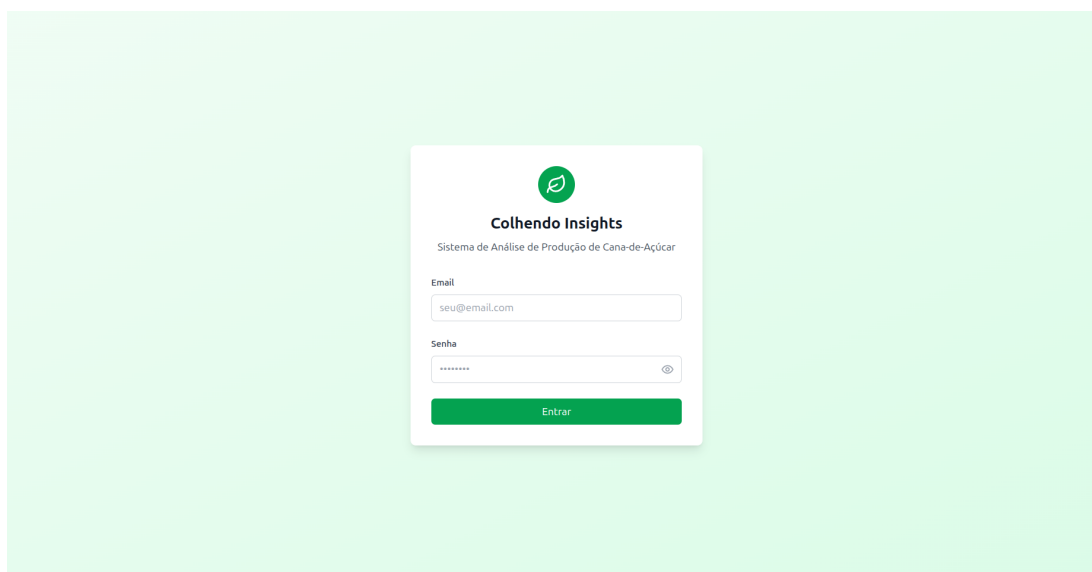
Fonte: [Medium - Arquitetura MVC](#)

3.5.2.2 Interfaces e Funcionalidades

A interface do AgroAnalysis foi projetada para oferecer uma navegação intuitiva e uma experiência de uso fluida. As principais telas estão descritas a seguir.

A tela inicial do sistema (figura 13) apresenta o formulário de autenticação de usuários. Ela garante o controle de acesso e a segurança das informações, permitindo que cada usuário acesse apenas os dados correspondentes à sua fazenda.

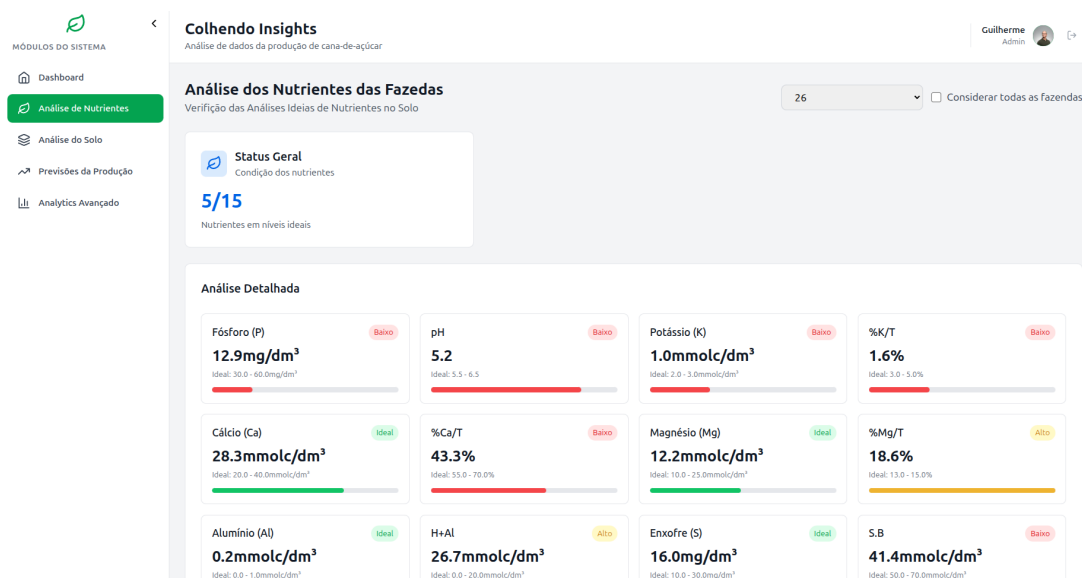
Figura 13 – Tela de Login do AgroAnalysis.



Fonte: Elaborada pelo autor.

A tela de análise de nutrientes (figura 14) tem como objetivo identificar quais fazendas apresentam médias dos nutrientes acima ou abaixo dos valores ideais. Além disso, a interface fornece recomendações de produtos a serem aplicados, de acordo com o estado de cada nutriente — seja para correção de deficiências ou controle de excessos. Os gráficos e indicadores auxiliam o produtor a compreender visualmente o equilíbrio nutricional do solo.

Figura 14 – Tela de Análise de Nutrientes.



Fonte: Elaborada pelo autor.

A tela de análise do solo (figura 15) permite que o usuário insira valores de nutrientes específicos e visualize instantaneamente quais estão dentro ou fora dos limites ideais. De forma complementar, o sistema sugere produtos para correção de cada nutriente, conforme sua situação. Essa funcionalidade transforma dados técnicos em orientações práticas de manejo.

Figura 15 – Tela de Análise do Solo.

Colhendo Insights
Análise de dados da produção de cana-de-açúcar

Verificação de Nutrientes no Solo
Informe os valores dos nutrientes e clique no botão para analisar

Fósforo (P) (mg/dm³) Informe Fósforo (P)	Matéria Orgânica (M.O) (g/dm³) Informe Matéria Orgânica (M.O)	pH Informe pH
Potássio (K) (mmolc/dm³) Informe Potássio (K)	%K/T (%) Informe %K/T	Cálcio (Ca) (mmolc/dm³) Informe Cálcio (Ca)
%Ca/T (%) Informe %Ca/T	Magnésio (Mg) (mmolc/dm³) Informe Magnésio (Mg)	%Mg/T (%) Informe %Mg/T
Enxofre (S) (mg/dm³) Informe Enxofre (S)	CTC (mmolc/dm³) Informe CTC	Saturação por Bases (V) (%) Informe Saturação por Bases (V)
Alumínio (Al) (mmolc/dm³) Informe Alumínio (Al)	H+Al (mmolc/dm³) Informe H+Al	S.B (mmolc/dm³) Informe S.B
Saturação por Al (%) Informe Saturação por Al		

Analisar Nutrientes

Fonte: Elaborada pelo autor.

A tela de predição da produção (figura 16) foi desenvolvida para que o usuário insira as quantidades de nutrientes e obtenha a estimativa da produção agrícola. O sistema também realiza uma análise dos nutrientes informados, indicando quais estão dentro dos limites ideais e quais necessitam de ajuste, fornecendo suporte à decisão com base em dados.

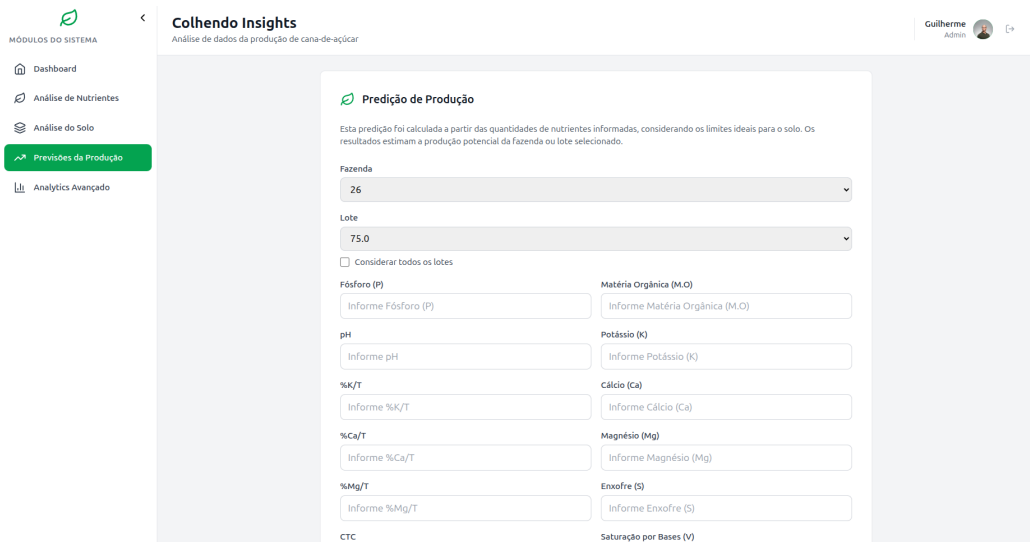
Para gerar essas estimativas, o sistema utiliza os modelos de regressão linear múltipla previamente treinados. A produtividade predita para cada profundidade é calculada por meio da Equação (3.1), que representa o modelo de regressão linear múltipla. Em seguida, a predição final é obtida pela média aritmética das predições das duas profundidades, conforme a Equação (3.2).

$$\hat{y} = \beta_0 + \sum_{j=1}^p \beta_j x_j. \quad (3.1)$$

$$\hat{y}_{\text{final}} = \frac{\hat{y}_{20\text{cm}} + \hat{y}_{40\text{cm}}}{2}. \quad (3.2)$$

Essa integração das camadas superficial e subsuperficial proporciona uma estimativa mais equilibrada e robusta da produtividade.

Figura 16 – Tela de Predição da Produção.



Fonte: Elaborada pelo autor.

A tela de análise estatística avançada (figura 17) reúne uma série de análises de dados, como distribuições, correlações e agrupamentos. Seu objetivo é permitir que o usuário compreenda o comportamento das variáveis do solo e o desempenho das fazendas de maneira mais profunda. Essa funcionalidade é essencial para produtores e técnicos que desejam utilizar a estatística como ferramenta estratégica de gestão.

Figura 17 – Tela de Análise Estatística Avançada.



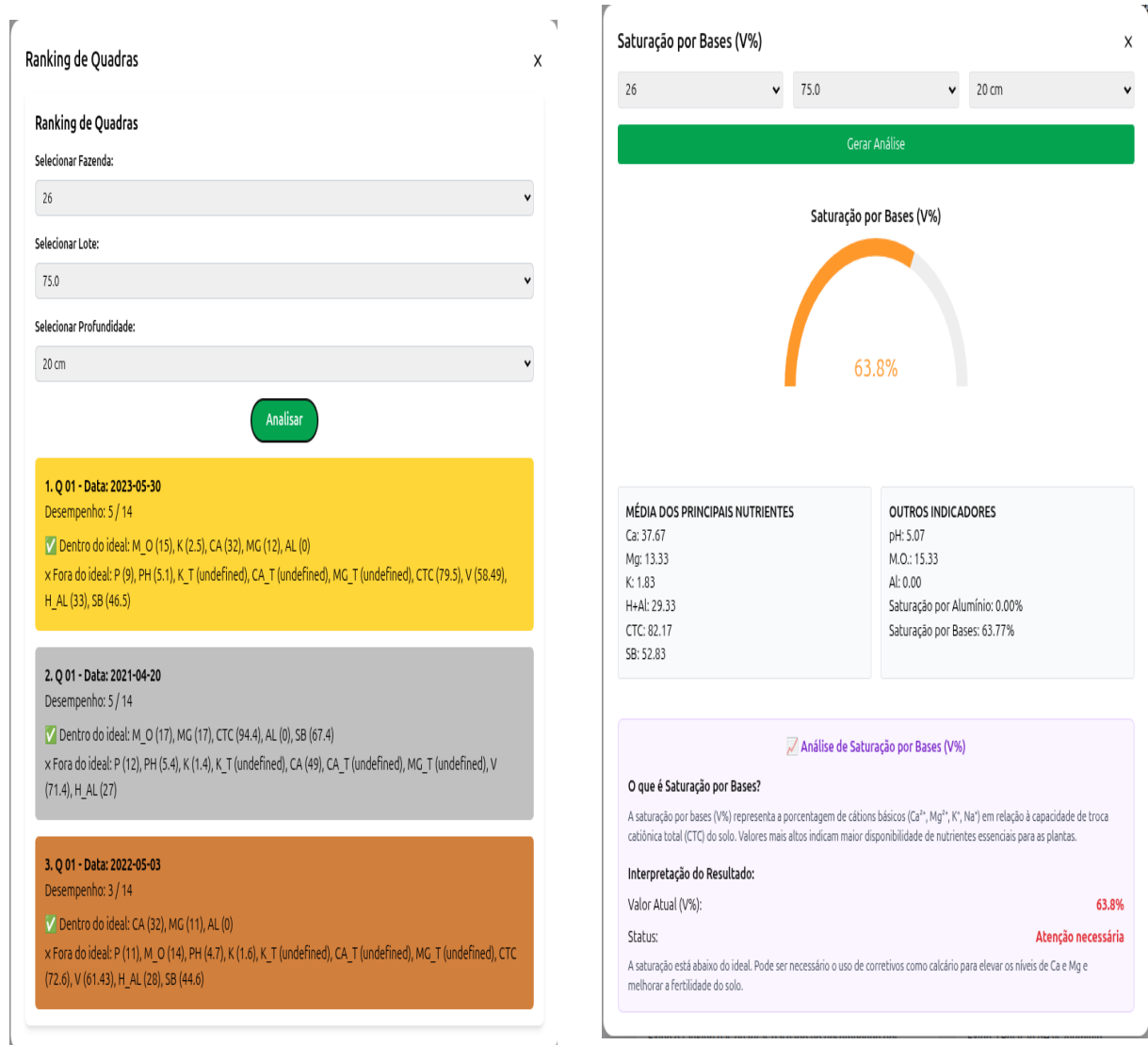
Fonte: Elaborada pelo autor.

Cada **card** apresentado nessa seção representa uma análise estatística específica, desenvolvida para explorar diferentes perspectivas sobre os dados da fazenda. Esses cards podem incluir, por exemplo, correlação entre variáveis, análise de variância, identificação de

outliers, (PCA) e agrupamento de quadras com comportamento similar. Essa estrutura modular permite que o usuário visualize e compare múltiplas análises de forma simultânea e intuitiva.

Além da tela principal, o sistema apresenta exemplos visuais que ilustram como os resultados são organizados em cada card. Na figura 18, são exibidos dois exemplos de análises disponíveis: à esquerda, o card de *ranking de quadras*, que ordena as áreas de produção de acordo com o desempenho médio; à direita, a *PCA*, que permite identificar padrões e relações entre os nutrientes do solo de maneira visual e sintética.

Figura 18 – Exemplos de análises estatísticas apresentadas nos cards do sistema.



(a) Card de ranking de quadras.

(b) Análise de componentes principais dos nutrientes.

Fonte: Elaborada pelo autor.

3.5.3 Considerações sobre o Desenvolvimento

Ambos os sistemas foram projetados com foco na usabilidade, eficiência e integração entre frontend e backend. Essa combinação garante que os resultados das análises estatísticas sejam apresentados em tempo real, apoiando a tomada de decisão na produção agrícola e permitindo que os produtores realizem simulações de manejo nutricional de forma prática e intuitiva.

3.6 Conjunto de Dados

Nesta seção, são apresentados os conjuntos de dados utilizados neste estudo, descrevendo suas características, estrutura e a forma de obtenção.

Os dados foram fornecidos por meio de uma parceria com o agrônomo **Thiago Ramalho**, que disponibilizou informações referentes à fazenda 26, sob sua gestão, localizada na região de Lençóis Paulista.

Os conjuntos de dados disponibilizados contêm informações detalhadas sobre cada amostra de solo coletada. Entre as variáveis presentes, estão a data da amostra, o lote, a quadra e a profundidade de coleta, além dos principais atributos químicos utilizados na análise deste estudo, como P, M.O, pH, K e os percentuais de K/T, Ca/T e Mg/T. Também estão incluídos Ca, Mg, Al, H+Al, S, S.B, CTC, V e Sat.

As amostras de solo são compostas de forma que, para cada fazenda, existe uma mesma amostragem correspondente a um determinado lote e quadra, coletada em duas profundidades distintas: 20 cm e 40 cm. Essa estrutura permite avaliar a variação dos nutrientes em diferentes camadas do solo, possibilitando análises comparativas mais precisas sobre a disponibilidade de elementos químicos em profundidades distintas.

A figura 19 apresenta um exemplo do *dataframe* da fazenda 26, mostrando as colunas referentes aos nutrientes e variáveis de manejo.

Figura 19 – Amostra do *dataframe* da fazenda 26, contendo dados de amostras de solo, com informações sobre nutrientes e variáveis de manejo.

	Data	Lote	Quadra	Prof.	P	M.O.	pH	K	%K/T	Ca	%Ca/T	Mg	%Mg/T	Al	H+Al	S	S.B.	CTC	V	Sat.
0	2016-09-02 00:00:00	56.0	Q 05/01	20	69	14	4.7	1.4	3.153153	16	36.036036	5	11.261261	3	22	0	22.4	44.4	50.45045	11.811024
1	2016-09-02 00:00:00	56.0	Q 05/01	40	51	9	4.5	0.8	1.639344	9	18.442623	3	6.147541	5	36	5	12.8	48.8	26.229508	28.089888
2	2016-09-02 00:00:00	56.0	Q 05/02	20	11	13	4.2	1.1	2.676399	10	24.330900	4	9.732360	8	26	0	15.1	41.1	36.739659	34.632035
3	2016-09-02 00:00:00	56.0	Q 05/02	40	5	12	4	0.6	1.408451	5	11.737089	2	4.694836	10	35	5	7.6	42.6	17.840376	56.818182
4	2016-09-02 00:00:00	56.0	Q 05/03	20	47	13	4.5	2.1	4.365904	15	31.185031	4	8.316008	6	27	0	21.1	48.1	43.866944	22.140221
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
535	2023-10-04 00:00:00	51.0	Q 02 e 03	40	3	9	4.9	0.6	1.520455	13	32.943182	4	10.136364	0	27	8	17.6	44.6	39.461883	0
536	2023-09-20 00:00:00	64.0	Q 01 e 02	20	11	14	5.3	0.9	1.338395	24	35.690544	10	14.871060	0	17	0	34.9	51.9	67.244701	0
537	2023-09-20 00:00:00	64.0	Q 01 e 02	40	3	11	5.2	0.5	1.045455	10	20.909091	6	12.545455	0	18	18	16.5	34.5	47.826087	0
538	2023-09-20 00:00:00	65.0	Q 01,02 e 03	20	11	15	5.7	0.9	1.157603	39	50.162791	16	20.579606	0	16	0	55.9	71.9	77.746871	0
539	2023-09-20 00:00:00	65.0	Q 01,02 e 03	40	8	11	5.2	0.4	0.658503	20	32.925170	9	14.816327	0	19	13	29.4	48.4	60.743802	0

Fonte: Elaborada pelo autor.

Além dos dados de solo, também foi obtido um conjunto de dados relacionado à produção. Esse conjunto foi estruturado em um *dataframe* específico, contendo diversas informações sobre a produtividade e a qualidade da cana-de-açúcar. As colunas presentes nesse conjunto incluem o nome do parceiro, setor, fazenda, lote e quadra, além do estágio de corte, variedade cultivada, espaçamentos (Esp 1 e Esp 2), data de plantio, corte anterior, último corte e o tipo de ambiente. Também estão registrados dados referentes à área colhida, produção real, TCH real e estimado, bem como características tecnológicas da cana, como Brix, Ls, Pbu, ATR, fibra, Ar, Art, Pol

Essas variáveis compõem o conjunto de atributos necessários para analisar o desempenho produtivo, permitindo relacionar características agrônômicas, ambientais e tecnológicas ao rendimento final da cultura.

A figura 20 ilustra o *dataframe* de produção, apresentando as colunas associadas à produtividade e qualidade da cana-de-açúcar. Esse conjunto de dados é essencial para a análise de como a produção de cana está relacionada aos parâmetros de manejo e às características do solo.

Figura 20 – Amostra do *dataframe* de produção, contendo dados sobre produtividade e qualidade da cana-de-açúcar.

	Nome Parceiro	Setor	Fazenda	Lote	Quadra	Estag Corte	Variedade	Esp 1	Esp 2	Data plantio	...	Brix	Ls	Pbu	ATR(kg/t)	Fibra(%)	Ar	Art	Pol
35	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	58	1	AM REF	RB975201	1,5	0,9	2022- 02-01	...	15.431149	51.368574	137.309957	109.53	11.86	0.71	11.97	
36	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	58	2	AM REF	RB975201	1,5	0,9	2022- 02-01	...	15.431273	51.369114	137.310348	109.53	11.86	0.71	11.97	
37	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	6	1	AM REF	AGRUPADAS	1,5	0,9	2022- 03-17	...	16.862021	57.671098	139.805546	120.61	12.06	0.66	13.18	
38	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	6	2	AM REF	AGRUPADAS	1,5	0,9	2022- 03-17	...	18.290852	64.947999	142.747063	133.07	12.30	0.58	14.54	
39	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	6	3	AM REF	AGRUPADAS	1,5	0,9	2022- 03-17	...	17.857940	62.743239	141.856373	129.31	12.22	0.60	14.13	

Fonte: Elaborada pelo autor.

3.6.1 Limpeza de Dados

Após a obtenção dos dados brutos, foi realizado um processo de limpeza dos dados com o objetivo de assegurar a qualidade e a consistência das informações utilizadas nas análises. O principal procedimento consistiu na filtragem de valores nulos, removendo amostras incompletas ou com ausência de informações em variáveis essenciais, como nutrientes e parâmetros do solo.

Além disso, foi necessário excluir as duas primeiras linhas do *dataframe* da fazenda, uma vez que a primeira continha informações sobre possíveis produtos aplicados a cada nutriente e a segunda apresentava as respectivas unidades de medida. Esses valores eram inconsistentes com a estrutura esperada do conjunto de dados e poderiam comprometer a correta interpretação das variáveis durante a análise.

A figura 21 apresenta um exemplo do *dataframe* original antes da etapa de limpeza, evidenciando as linhas iniciais que precisaram ser removidas. O processo de limpeza foi aplicado aos *dataframe* da fazenda 26, garantindo a padronização dos dados e a confiabilidade dos resultados obtidos nas etapas subsequentes de análise.

Figura 21 – Exemplo do *dataframe* original antes do processo de Limpeza dos Dados.

	Data	Lote	Quadra	Prof.	P	M.O.	pH	K	%K/T	Ca	%Ca/T	Mg	%Mg/T	Al	H+Al
0	NaN	NaN	NaN	NaN	Resina	S. Sulfurosa	Sol. CaCl2	Resina	NaN	Resina	NaN	Resina	NaN	KCl	Tampão SMP
1	Coleta	NaN	NaN	cm	mg/dm3	g/dm3	NaN	mmolc/dm3	NaN	mmolc/dm3	NaN	mmolc/dm3	NaN	mmolc/dm3	Fosf.Cál
2	2016-09-02 00:00:00	56.0	Q 05/01	20	69	14	4.7	1.4	3.153153	16	36.036036	5	11.261261	3	22
3	2016-09-02 00:00:00	56.0	Q 05/01	40	51	9	4.5	0.8	1.639344	9	18.442623	3	6.147541	5	36
4	2016-09-02 00:00:00	56.0	Q 05/02	20	11	13	4.2	1.1	2.676399	10	24.330900	4	9.732360	8	26
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
537	2023-10-04 00:00:00	51.0	Q 02 e 03	40	3	9	4.9	0.6	1.520455	13	32.943182	4	10.136364	0	27
538	2023-09-20 00:00:00	64.0	Q 01 e 02	20	11	14	5.3	0.9	1.338395	24	35.690544	10	14.871060	0	17
539	2023-09-20 00:00:00	64.0	Q 01 e 02	40	3	11	5.2	0.5	1.045455	10	20.909091	6	12.545455	0	18
540	2023-09-20 00:00:00	65.0	Q 01,02 e 03	20	11	15	5.7	0.9	1.157603	39	50.162791	16	20.579606	0	16
541	2023-09-20 00:00:00	65.0	Q 01,02 e 03	40	8	11	5.2	0.4	0.658503	20	32.925170	9	14.816327	0	19

Fonte: Elaborada pelo autor.

No caso do *dataframe* de produção, que contém informações sobre a produtividade da cana-de-açúcar, foi identificado que algumas amostras apresentavam valores inconsistentes. Esses dados inconsistentes poderiam afetar a análise e interpretação dos resultados. Para garantir a qualidade das informações, foi adotada a estratégia de exclusão dessas amostras inconsistentes, removendo apenas os registros com dados problemáticos. Esse processo assegurou que o *dataframe* de produção fosse mantido livre de erros, pronto para ser utilizado nas análises subsequentes. A figura 22 apresenta o *dataframe* de produção após a limpeza, exibindo amostras com dados nulos que foram removidos.

Figura 22 – Amostra do *dataframe* de produção, exibindo as amostras com dados nulos antes do processo de Limpeza dos Dados.

```
prod[prod.isnull().any(axis=1)].head()
```

	Nome Parceiro	Setor	Fazenda	Lote	Quadra	Estag Corte	Variedade	Esp 1	Esp 2	Data plantio	...	Brix	Ls	Pbu	ATR(kg/t)	Fibra(%)	Ar	Art	Pol(%)
35	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	58	1	AM REF	RB975201	1,5	0,9	2022- 02-01	...	15.431149	51.368574	137.309957	109.53	11.86	0.71	11.97	1
36	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	58	2	AM REF	RB975201	1,5	0,9	2022- 02-01	...	15.431273	51.369114	137.310348	109.53	11.86	0.71	11.97	1
37	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	6	1	AM REF	AGRUPADAS	1,5	0,9	2022- 03-17	...	16.862021	57.671098	139.805546	120.61	12.06	0.66	13.18	1
38	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	6	2	AM REF	AGRUPADAS	1,5	0,9	2022- 03-17	...	18.290852	64.947999	142.747063	133.07	12.30	0.58	14.54	1
39	JOSE ROBERTO BATTISTELLA E OUTROS	92	25	6	3	AM REF	AGRUPADAS	1,5	0,9	2022- 03-17	...	17.857940	62.743239	141.856373	129.31	12.22	0.60	14.13	1

5 rows × 27 columns

Fonte: Elaborada pelo autor.

3.6.2 Tratamento de Dados

Após a etapa de limpeza, os dados foram organizados e tratados para possibilitar análises mais direcionadas. Um dos principais procedimentos consistiu na separação das amostras de solo conforme a profundidade de coleta, criando subconjuntos específicos para 20 cm e 40 cm.

Essa divisão permitiu analisar de forma independente a distribuição e o comportamento dos nutrientes em diferentes camadas do solo, possibilitando identificar variações químicas em função da profundidade. Dessa maneira, foi possível compreender melhor como fatores nutricionais e físico-químicos se comportam nas diferentes regiões analisadas de cada fazenda.

As figuras 23 e 24 apresentam a aplicação do filtro sobre o conjunto de dados, resultando na criação de *dataframes* específicos para as profundidades de 20 cm e 40 cm, respectivamente. Essa separação foi essencial para garantir análises consistentes e comparativas entre as camadas do solo.

Figura 23 – Aplicação do filtro e *dataframe* correspondente à profundidade de 20 cm.

```
: nutrientes_prof_20 = nutrientes.query("Prof == '20'")
nutrientes_prof_20.head()
```

	Data	Lote	Quadra	Prof	P	M.O	pH	K	K/T	Ca	Ca/T	Mg	Mg/T	Al	H+Al	S	S.B	CTC	V	Sat.
1	2021-04-20	75.0	Q 01	20	12	17	5.4	1.4	1.483051	49	51.906780	17	18.008475	0	27	0	67.4	94.4	71.398305	0
2	2021-04-29	56.0	Q 07	20	11	17	4.8	1.5	1.960784	29	37.908497	11	14.379085	0	35	0	41.5	76.5	54.248366	0
4	2021-04-29	56.0	Q Expansão	20	28	22	4.8	3.1	3.294368	39	41.445271	16	17.003188	0	36	0	58.1	94.1	61.742827	0
7	2021-08-09	56.0	Q 08 e 12	20	16	15	5.1	1.8	2.004454	39	43.429844	17	18.930958	0	32	0	57.8	89.8	64.365256	0
8	2021-08-09	50.0	Q 01 e 02	20	10	12	5.1	1.4	1.933702	32	44.198895	12	16.574586	0	27	0	45.4	72.4	62.707182	0

Fonte: Elaborada pelo autor.

Figura 24 – Aplicação do filtro e *dataframe* correspondente à profundidade de 40 cm.

```

: nutrientes_prof_40 = nutrientes.query("Prof == '40'")
: nutrientes_prof_40.head()

```

	Data	Lote	Quadra	Prof	P	M.O	pH	K	K/T	Ca	Ca/T	Mg	Mg/T	Al	H+Al	S	S.B	CTC	V	Sat.
0	2021-04-20	75.0	Q 01	40	7	14	5.2	1	1.204819	40	48.192771	14	16.867470	0	28	58	55	83	66.26506	0
3	2021-04-29	56.0	Q 07	40	9	15	4.5	0.8	1.162791	20	29.069767	8	11.627907	2	40	46	28.8	68.8	41.860465	6.493506
5	2021-04-29	56.0	Q Expansão	40	13	16	4.9	1.7	1.960784	30	34.602076	12	13.840830	0	43	7	43.7	86.7	50.403691	0
6	2021-08-09	6.0	Q 01	40	12	11	5.1	0.6	0.849858	32	45.325779	12	16.997167	0	26	17	44.6	70.6	63.172805	0
10	2021-08-09	56.0	Q 08 e 12	40	12	14	5.2	1.3	1.472254	38	43.035108	16	18.120045	0	33	24	55.3	88.3	62.627407	0

Fonte: Elaborada pelo autor.

Em contraste, para os dados relativos à produção de cana-de-açúcar, não foi necessário realizar um tratamento específico, pois os dados de produção não são influenciados pela profundidade do solo. A produção de cana-de-açúcar é determinada por uma série de fatores como a variedade da planta, o manejo agrícola e as condições climáticas, mas a profundidade do solo, no contexto desse estudo, não exerce um efeito direto sobre os dados de produção. Dessa forma, o *dataframe* de produção foi mantido intacto e utilizado sem a necessidade de separação por profundidade.

3.7 Procedimentos de Análise Exploratória dos Dados

A etapa de análise exploratória dos dados teve como objetivo compreender as características e comportamentos presentes nos conjuntos de dados das amostras de solo e produção. Essa fase inicial buscou identificar padrões, tendências e possíveis relações entre as variáveis, além de calcular estatísticas descritivas que fornecem uma visão geral sobre a distribuição dos nutrientes, parâmetros físico-químicos do solo e variáveis de produção.

Por meio dessa análise preliminar, foi possível obter um entendimento mais profundo sobre os dados coletados, identificar inconsistências ou valores atípicos e direcionar as etapas seguintes do estudo de forma mais assertiva.

3.7.1 Procedimentos para Caracterização dos Nutrientes e Dados de Produção

Para cada nutriente analisado nas profundidades de 20 cm e 40 cm, bem como para as variáveis de produção, inicialmente foram calculadas estatísticas descritivas, incluindo média, mediana, desvio-padrão e variância. Esses indicadores permitiram caracterizar a tendência central e a dispersão dos dados, fornecendo uma compreensão quantitativa preliminar do comportamento das variáveis.

Em seguida, foram elaboradas diferentes visualizações gráficas com o objetivo de facilitar a interpretação das distribuições. Os histogramas foram utilizados para avaliar a forma da distribuição, permitindo a identificação de assimetrias ou possíveis padrões não lineares. Já os

boxplots possibilitaram a observação da dispersão dos dados por meio dos quartis, além de facilitar a detecção de valores extremos (*outliers*) com base na estrutura gráfica da distribuição.

Também foi aplicado o teste de normalidade de Shapiro-Wilk, implementado por meio da função `stats.shapiro()` da biblioteca `scipy.stats`, com o propósito de verificar se os dados seguiam uma distribuição normal. Esse teste forneceu evidências estatísticas essenciais para orientar a escolha dos métodos inferenciais adequados nas etapas posteriores da análise.

Por fim, a identificação de *outliers* foi realizada tanto de forma visual, por meio dos boxplots, quanto por um procedimento sistemático baseado no intervalo interquartil (IQR). Esse critério permitiu detectar valores atípicos que poderiam distorcer medidas estatísticas ou influenciar negativamente os resultados das análises subsequentes, garantindo maior robustez aos métodos adotados.

Esses procedimentos foram aplicados de forma consistente tanto aos dados de solo (profundidades de 20 cm e 40 cm) quanto aos dados de produção, permitindo uma caracterização completa e comparável de todas as variáveis do estudo.

Os resultados dessas análises, incluindo as estatísticas descritivas detalhadas, os gráficos ilustrativos, os resultados dos testes de normalidade para cada variável e as interpretações decorrentes, são apresentados de forma completa no Capítulo 4.

3.7.2 Procedimentos para Normalização dos Dados

Considerando que grande parte dos nutrientes e variáveis de produção apresentaram distribuições não normais, e visando possibilitar a aplicação de métodos estatísticos paramétricos com maior robustez, foi implementada uma etapa sistemática de normalização dos dados.

O processo de normalização seguiu os seguintes passos metodológicos:

1. **Identificação de variáveis não normais:** seleção das variáveis que apresentaram $p \leq 0,05$ no teste de Shapiro-Wilk, indicando rejeição da hipótese de normalidade;
2. **Seleção e aplicação de transformações:** para cada variável identificada, foram testadas diferentes transformações estatísticas, selecionando aquela que melhor aproximou os dados à distribuição normal. As transformações consideradas foram:

- **Transformação logarítmica:** aplicada através da Equação (3.3), onde c é uma constante de ajuste utilizada quando necessário para evitar valores negativos ou zeros.

$$y' = \log(y + c). \quad (3.3)$$

- **Transformação Box-Cox:** definida matematicamente pela Equação (3.4), que descreve a transformação para diferentes valores do parâmetro λ . Em seguida, o

parâmetro ótimo é estimado por máxima verossimilhança, conforme a Equação (3.5).

$$y'(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \\ \log(y), & \text{se } \lambda = 0 \end{cases} \quad (3.4)$$

$$L(\lambda) = -\frac{n}{2} \ln \left(\frac{1}{n} \sum_{i=1}^n [y'_i(\lambda) - \bar{y}'(\lambda)]^2 \right) + (\lambda - 1) \sum_{i=1}^n \ln(y_i). \quad (3.5)$$

- **Transformação Box–Cox com deslocamento:** variante da transformação Box–Cox utilizada quando os dados possuem valores negativos ou iguais a zero. A forma matemática da transformação é apresentada na Equação (3.6).

$$y' = \frac{(y + c)^\lambda - 1}{\lambda}. \quad (3.6)$$

- **Transformação Johnson:** família de transformações que busca aproximar os dados a uma distribuição normal. Para a família S_U (dados não limitados), utiliza-se a Equação (3.7), onde os parâmetros γ , δ , ξ e λ são estimados para que a variável transformada siga uma distribuição normal padrão.

$$z = \gamma + \delta \sinh^{-1} \left(\frac{y - \xi}{\lambda} \right). \quad (3.7)$$

- **Transformação Rankit:** método que substitui os valores observados pelos quantis correspondentes da distribuição normal padrão. A Equação (3.8) mostra como cada observação ordenada $x_{(i)}$ é transformada.

$$y_i = \Phi^{-1} \left(\frac{i - 0.5}{n} \right). \quad (3.8)$$

- **Tratamento de outliers:** em casos específicos onde valores extremos comprometiam significativamente a distribuição, foi aplicada a substituição desses outliers pela média da variável, seguida de reavaliação da normalidade.
3. **Validação da normalização:** após cada transformação, foi aplicado novamente o teste de Shapiro-Wilk para verificar se a normalidade foi atingida ($p > 0,05$);
 4. **Critério de seleção:** O critério de seleção da transformação final aplicada a cada variável considerou simultaneamente os resultados estatísticos e a inspeção gráfica. A escolha foi guiada pelo p -value obtido no teste de normalidade de Shapiro-Wilk, priorizando-se transformações que apresentassem valores de $p > 0,05$, indicando a não rejeição da hipótese de normalidade. Além disso, foram analisados visualmente os histogramas, buscando distribuições com forma aproximada à normal, e os boxplots, a fim de verificar a presença de possíveis distorções ou valores extremos. Outro aspecto importante foi a preservação da interpretabilidade dos dados sempre que possível, evitando transformações

que dificultassem a compreensão prática dos resultados. Por fim, também se buscou minimizar distorções excessivas nos valores originais, garantindo que a transformação fosse estatisticamente adequada sem comprometer a coerência dos dados.

Este procedimento de normalização foi aplicado de forma independente a três conjuntos de dados distintos. Para os nutrientes avaliados a 20 cm de profundidade, a transformação foi empregada somente naqueles que apresentaram distribuição não normal. O mesmo critério foi utilizado para os nutrientes medidos a 40 cm de profundidade, garantindo consistência metodológica entre as camadas analisadas. Por fim, as variáveis de produção que não seguiram distribuição normal também foram transformadas, assegurando que todas as análises subsequentes fossem conduzidas com dados mais adequados aos pressupostos estatísticos.

É importante ressaltar que, diferentemente dos nutrientes do solo, os dados de produção não foram separados por profundidade, uma vez que a produtividade da cana-de-açúcar é determinada por fatores como variedade da planta, manejo agrícola e condições climáticas, não sendo diretamente influenciada pelas camadas do solo analisadas.

Essa etapa de normalização foi fundamental para viabilizar a aplicação de análises estatísticas paramétricas mais robustas e confiáveis nas etapas subsequentes do estudo, garantindo que as premissas dos testes estatísticos fossem adequadamente atendidas. Os resultados detalhados da normalização são apresentados no Capítulo 4.

3.8 Procedimentos para Análise da Distribuição Nutricional por Lote e Quadra

Para avaliar as condições nutricionais específicas de cada área produtiva, foi desenvolvido um procedimento metodológico estruturado para análise da distribuição de nutrientes por lote e quadra, considerando as duas profundidades de amostragem (20 cm e 40 cm).

3.8.1 Definição de Faixas Ideais de Nutrientes

Com base em referências técnicas da literatura agrônômica ([Embrapa, 2024](#)) e em recomendações estabelecidas em colaboração com o agrônomo **Thiago Ramalho**, responsável pelo manejo das áreas estudadas, foram definidas faixas de concentração consideradas ideais para cada nutriente analisado. Essas faixas representam os intervalos nos quais os nutrientes promovem condições adequadas de fertilidade do solo e desenvolvimento das plantas, considerando as características edafoclimáticas específicas da região de estudo e as particularidades do manejo adotado.

O quadro 3 apresenta as faixas ideais estabelecidas para os principais nutrientes e atributos químicos do solo analisados neste estudo.

Quadro 3 – Faixas ideais de nutrientes e atributos químicos do solo para o cultivo de cana-de-açúcar.

Nutriente/Atributo	Faixa Ideal
Fósforo (P)	> 30 mg/dm ³
Matéria Orgânica (M.O.)	15–25 g/dm ³
pH	5,5–6,5
Potássio (K)	> 2 mmol _c /dm ³
K/T	3–5%
Cálcio (Ca)	20–40 mmol _c /dm ³
Ca/T	55–70%
Magnésio (Mg)	10–25 mmol _c /dm ³
Mg/T	13–15%
Enxofre (S)	> 10 mg/dm ³
Capacidade de Troca Catiônica (CTC)	80–200 mmol _c /dm ³
Saturação por Bases (V)	> 80%
Alumínio (Al)	< 1 mmol _c /dm ³
H+Al	0–20 mmol _c /dm ³
Soma de Bases (S.B.)	50–70 mmol _c /dm ³
Saturação por Alumínio (Sat.)	< 10%

Fonte: Elaborado pelo autor.

Essas faixas foram utilizadas como referência para a classificação dos nutrientes em cada lote e quadra analisados, permitindo identificar áreas com deficiências, adequação ou excessos nutricionais. Valores abaixo dos limites inferiores indicam condições de deficiência que podem comprometer a produtividade, enquanto concentrações acima dos limites superiores podem resultar em desequilíbrios nutricionais, antagonismos entre elementos ou, em casos específicos, toxidez. O monitoramento periódico desses parâmetros através de análises de solo e a correção quando necessário são práticas essenciais para o manejo eficiente da cultura e para a manutenção da sustentabilidade produtiva ao longo dos ciclos de cultivo.

3.8.2 Cálculo de Médias por Lote e Quadra

Para cada combinação de lote e quadra, foram calculadas as médias das concentrações de cada nutriente nas profundidades de 20 cm e 40 cm. Este procedimento permitiu caracterizar o estado nutricional médio de cada área específica, eliminando a variabilidade intra-parcela e fornecendo valores representativos para comparação com as faixas ideais estabelecidas.

3.8.3 Classificação dos Nutrientes

Após o cálculo das médias, cada nutriente foi classificado em uma das três categorias definidas a partir das faixas de recomendação. Valores cuja concentração média ficou abaixo do limite inferior foram classificados como abaixo do ideal, indicando deficiência nutricional.

Quando a média se situou dentro da faixa recomendada, o nutriente foi categorizado como dentro do ideal, representando condições adequadas do solo. Por fim, quando a concentração média ultrapassou o limite superior da faixa ideal, o nutriente foi classificado como acima do ideal, sugerindo excesso que pode comprometer o equilíbrio químico ou a absorção de outros elementos.

3.8.4 Geração de Visualizações Comparativas

Para facilitar a interpretação dos resultados, foram gerados gráficos de barras comparativos para cada lote e quadra, nos quais são apresentadas as concentrações médias de todos os nutrientes analisados nas profundidades de 20 cm (em azul) e 40 cm (em laranja). Esses gráficos permitem visualizar rapidamente a variação vertical dos nutrientes entre as duas camadas do solo, além de possibilitar a identificação visual dos elementos que se encontram fora das faixas consideradas ideais. Também permitem realizar comparações diretas entre diferentes áreas de produção, oferecendo uma visão integrada e intuitiva das condições nutricionais do solo ao longo da propriedade.

3.8.5 Tabulação dos Resultados

Os nutrientes classificados como “dentro do ideal” em cada lote e quadra foram organizados em quadros síntese separados por profundidade, permitindo uma visualização estruturada e comparativa das condições nutricionais. Essa tabulação possibilita identificar rapidamente quais áreas apresentam melhores níveis de equilíbrio químico no solo, além de evidenciar quais nutrientes são mais frequentemente encontrados dentro das faixas recomendadas. Também torna possível observar padrões de distribuição nutricional entre as diferentes áreas estudadas, auxiliando na tomada de decisão e no planejamento de práticas de manejo mais eficientes.

3.8.6 Análise de Frequências

Para complementar a análise individual por lote e quadra, foi realizada uma avaliação de frequências agregadas, contabilizando quantas vezes cada nutriente foi classificado como “ideal” em todas as áreas avaliadas, de maneira separada para as profundidades de 20 cm e 40 cm. Essa abordagem possibilita identificar quais nutrientes apresentam maior estabilidade e adequação geral ao longo das áreas estudadas, bem como aqueles que demonstram maior variabilidade e, portanto, exigem maior atenção no manejo nutricional. Além disso, essa análise evidencia diferenças entre as camadas superficial e subsuperficial do solo, permitindo compreender como a distribuição dos nutrientes varia verticalmente e como isso pode influenciar práticas de correção e adubação.

Os resultados obtidos através destes procedimentos metodológicos são apresentados detalhadamente no Capítulo 4, fornecendo subsídios para a tomada de decisão quanto ao

manejo nutricional específico de cada área produtiva.

3.9 Procedimentos para Análise de Correlação entre Nutrientes

A análise de correlação entre os nutrientes do solo foi conduzida para identificar relações lineares entre as variáveis químicas e físico-químicas nas profundidades de 20 cm e 40 cm. Esta análise permite compreender como os nutrientes interagem entre si e como essas interações podem influenciar a disponibilidade e o equilíbrio nutricional do solo.

3.9.1 Método de Correlação Utilizado

Foi empregado o coeficiente de correlação de Pearson, utilizado para medir a força e a direção da relação linear entre duas variáveis contínuas. Esse coeficiente varia de -1 a $+1$. Valores próximos de $+1$ indicam uma forte correlação positiva, o que significa que ambas as variáveis tendem a aumentar simultaneamente. Por outro lado, valores próximos de -1 representam uma forte correlação negativa, isto é, quando uma variável aumenta, a outra tende a diminuir. Já valores próximos de 0 indicam ausência de relação linear entre as variáveis, sugerindo que não há tendência consistente de aumento ou redução conjunta.

A análise de correlação foi realizada separadamente para as profundidades de 20 cm e 40 cm, o que permitiu identificar se as relações entre os nutrientes se mantêm consistentes em diferentes camadas do solo. Além disso, essa abordagem possibilitou detectar possíveis diferenças na dinâmica nutricional entre as regiões superficial e subsuperficial. Também foi possível avaliar o comportamento específico de cada profundidade no que diz respeito às interações entre os nutrientes, oferecendo uma compreensão mais detalhada das variações estruturais presentes no perfil do solo.

Os resultados das análises de correlação foram organizados em matrizes de correlação e apresentados tanto em formato tabular quanto por meio de mapas de calor (*heatmaps*). Essa forma de apresentação facilita a identificação visual dos pares de nutrientes que exibem correlações fortes, sejam elas positivas ou negativas. Além disso, os *heatmaps* permitem observar padrões de associação entre grupos de nutrientes, evidenciando possíveis interações relevantes para o manejo agrícola. Também se tornam mais claras as diferenças nas estruturas de correlação entre as duas profundidades analisadas, contribuindo para uma compreensão mais ampla do comportamento nutricional no perfil do solo.

Para cada correlação significativa identificada, foram examinadas as possíveis explicações agronômicas e pedológicas que poderiam justificar a relação observada. Entre os fatores considerados, incluíram-se os processos biogeoquímicos que regem as interações entre nutrientes, como casos de antagonismo ou sinergismo que afetam sua disponibilidade no solo. Também foram avaliados os efeitos de práticas de manejo, como calagem e adubação, que podem alterar simultaneamente a concentração de múltiplos elementos. Além disso, levou-se em

conta a influência das características físico-químicas do solo, as quais podem condicionar o comportamento conjunto de diversos nutrientes. Outro aspecto considerado diz respeito aos processos de mobilidade e lixiviação ao longo do perfil do solo, que podem contribuir para padrões específicos de correlação entre diferentes profundidades. Os resultados detalhados dessa análise, incluindo as matrizes de correlação completas e suas respectivas interpretações, são apresentados no Capítulo 4.

3.10 Procedimentos para o PCA

O PCA foi aplicado para reduzir a dimensionalidade do conjunto de dados de nutrientes do solo, identificando os principais padrões de variação e as relações subjacentes entre as variáveis.

Antes da aplicação da Análise de Componentes Principais (PCA), os dados foram padronizados utilizando o `StandardScaler` da biblioteca `scikit-learn`, de modo a transformar todas as variáveis para média zero e variância unitária. Essa etapa é fundamental para eliminar o efeito das diferentes unidades de medida e escalas, garantindo que variáveis com magnitudes maiores não acabem dominando a análise. Além disso, a padronização permite uma comparação adequada entre os nutrientes, mesmo quando apresentam faixas de concentração distintas, assegurando que todos os fatores contribuam de forma equilibrada para a extração dos componentes principais.

O PCA foi conduzido de forma independente para as profundidades de 20 cm e 40 cm, o que permitiu identificar se os padrões de variação nutricional diferem entre as camadas do solo. Essa abordagem possibilitou avaliar quais nutrientes exercem maior influência em cada profundidade, evidenciando os elementos que mais contribuem para a variabilidade observada. Além disso, tornou-se possível comparar a estrutura de variabilidade entre as camadas superficial e subsuperficial, oferecendo uma visão detalhada das diferenças na distribuição e interação dos nutrientes ao longo do perfil do solo.

O número adequado de componentes principais a serem retidos foi determinado por meio de diferentes critérios complementares. Primeiramente, foi realizada a análise da variância explicada acumulada, avaliando quantos componentes seriam necessários para explicar pelo menos 90% da variância total dos dados. Em seguida, analisou-se o *scree plot*, identificando o ponto de inflexão ou “cotovelo” no gráfico de variância explicada por componente, que indica a quantidade de componentes mais informativos. Por fim, considerou-se o critério de Kaiser, retendo apenas os componentes com autovalores superiores a 1, garantindo que cada componente contribua de maneira significativa para a representação da variabilidade dos dados.

Para cada componente principal retido, realizou-se a interpretação com base em múltiplos procedimentos complementares. Primeiramente, foi feita a análise dos *loadings*, identificando quais nutrientes apresentam maiores cargas (ou pesos) em cada componente, o que indica

sua contribuição para a formação daquele eixo de variação. Com base nesses resultados e nas interpretações agronômicas, cada componente recebeu um rótulo descritivo, permitindo nomeá-lo de forma representativa, como por exemplo “Acidez do Solo” ou “Fertilidade Básica”. Além disso, foram gerados *biplots*, gráficos que combinam as projeções das amostras (*scores*) com os vetores dos nutrientes (*loadings*), possibilitando a visualização simultânea de amostras e variáveis no espaço reduzido e facilitando a interpretação conjunta das relações entre elas.

A adequação do PCA foi avaliada com base em diversos critérios. Primeiramente, considerou-se o percentual de variância explicada pelos componentes principais selecionados, garantindo que a maior parte da variabilidade dos dados fosse representada. Em seguida, analisou-se a distribuição das amostras no espaço definido pelos componentes principais, permitindo verificar padrões e agrupamentos entre os dados. Por fim, foi realizada a verificação da interpretabilidade agronômica dos componentes obtidos, assegurando que cada eixo de variação tivesse significado prático em termos de manejo e características do solo. Os resultados completos da análise PCA, incluindo os gráficos de variância explicada, *biplots* e interpretações detalhadas dos componentes, são apresentados no Capítulo 4.

3.11 Procedimentos para Análise de Clusterização

A análise de clusterização foi aplicada com o objetivo de identificar grupos de amostras de solo com características nutricionais similares, permitindo segmentar as áreas estudadas em perfis homogêneos de fertilidade.

Para a análise de agrupamento, foi utilizado o algoritmo K-Means, implementado por meio da biblioteca *scikit-learn*, que organiza as amostras de modo a minimizar a variância intra-*cluster*. A escolha do K-Means se justifica por sua eficiência computacional, sendo adequado para conjuntos de dados de médio porte, além da facilidade de interpretação dos clusters formados. Ademais, o método é especialmente apropriado para dados numéricos contínuos, como os nutrientes e variáveis de produção analisados neste estudo, permitindo uma segmentação clara e significativa das amostras.

Antes da clusterização, os dados foram padronizados utilizando o *StandardScaler*, seguindo os mesmos procedimentos descritos para o PCA. Esta padronização é essencial para evitar que variáveis com maiores escalas dominem o cálculo das distâncias no algoritmo de clusterização.

O número ideal de *clusters* (k) foi determinado utilizando dois métodos complementares. O primeiro, denominado Método do Cotovelo (*Elbow Method*), baseia-se na avaliação da inércia, que corresponde à soma das distâncias quadráticas das amostras aos seus centróides, para diferentes valores de k . A partir dessa análise, identifica-se o ponto de inflexão em que o ganho marginal na redução da inércia diminui significativamente, o que é visualizado por meio de gráficos que relacionam a inércia ao número de *clusters*.

O segundo método utilizado foi o *Silhouette Score*, que calcula o coeficiente de silhueta para diferentes valores de k . Esse coeficiente mede o quão semelhante uma amostra é ao seu próprio *cluster* em comparação aos demais *clusters*. Valores de silhueta variam de -1 a $+1$, sendo que valores próximos a $+1$ indicam uma boa separação entre os *clusters*. O número de *clusters* selecionado corresponde àquele que maximiza o coeficiente médio de silhueta, garantindo uma segmentação clara e consistente das amostras.

A clusterização foi realizada de forma independente para as profundidades de 20 cm e 40 cm, permitindo avaliar se os perfis de solo se mantêm consistentes entre as diferentes camadas. Essa abordagem possibilitou identificar diferenças na estruturação dos grupos nutricionais em cada profundidade, evidenciando variações específicas na distribuição de nutrientes. Além disso, a segmentação independente facilita a elaboração de recomendações de manejo mais precisas e direcionadas, adaptadas às condições particulares de cada camada do solo.

Os *clusters* formados foram visualizados utilizando diferentes abordagens complementares. Primeiramente, as amostras foram projetadas no espaço definido pelos dois primeiros componentes principais do PCA, sendo os pontos coloridos de acordo com o *cluster* atribuído. Essa projeção permitiu avaliar visualmente a separação e a consistência dos grupos formados.

Além disso, foram elaborados gráficos de perfis para cada *cluster*, comparando as médias dos nutrientes por meio de gráficos de radar ou de barras. Esses gráficos evidenciam as características distintivas de cada grupo, facilitando a interpretação agrônômica e a identificação de padrões nutricionais relevantes para o manejo do solo.

Para cada *cluster* identificado, foi realizada uma caracterização detalhada com base em múltiplos critérios. Primeiramente, calcularam-se as médias e os desvios-padrão de cada nutriente dentro de cada *cluster*, permitindo uma visão quantitativa das características nutricionais predominantes. Em seguida, identificaram-se as quadras e lotes pertencentes a cada grupo, possibilitando associar os clusters às áreas físicas da fazenda. Também foi analisada a distribuição espacial dos *clusters* nas áreas estudadas, verificando padrões geográficos e potenciais relações com o manejo do solo. Por fim, realizou-se a interpretação agrônômica das características nutricionais de cada grupo, fornecendo informações relevantes para a tomada de decisão e recomendações de manejo específicas.

Para identificar quais nutrientes foram mais determinantes na formação dos *clusters*, utilizou-se o `RandomForestClassifier` da biblioteca `scikit-learn`. Inicialmente, foi treinado um modelo de classificação com o objetivo de prever o *cluster* atribuído a cada amostra com base nos nutrientes disponíveis.

Em seguida, foram extraídas as importâncias das variáveis (*feature importances*) fornecidas pelo modelo, permitindo identificar quais nutrientes mais influenciaram a separação entre os grupos. Essas importâncias foram então ordenadas e visualizadas para facilitar a interpretação. Por fim, realizou-se a interpretação agrônômica da relevância de cada nutriente,

permitindo compreender como cada elemento contribui para a segmentação das áreas e fornecendo subsídios para recomendações de manejo direcionadas. Os resultados completos da clusterização, incluindo os gráficos de determinação do número de *clusters*, visualizações das projeções, caracterização detalhada de cada *cluster* e análise de importância das variáveis, são apresentados no Capítulo 4.

3.12 Procedimentos para Testes Estatísticos Inferenciais

Para avaliar a significância das diferenças observadas entre grupos e profundidades, foram aplicados diversos testes estatísticos inferenciais, selecionados de acordo com as características dos dados e as hipóteses a serem testadas.

3.12.1 Teste t Pareado para Comparação entre Profundidades

Para verificar se as concentrações de nutrientes diferem significativamente entre as profundidades de 20 cm e 40 cm, foi aplicado o teste t pareado para cada nutriente individualmente.

O teste t pareado foi escolhido por ser apropriado para comparar amostras dependentes, uma vez que os dados das profundidades de 20 cm e 40 cm foram coletados nos mesmos locais, considerando os mesmos lotes e quadras. Para a realização do teste, adotou-se um nível de significância de $\alpha = 0,05$.

As hipóteses testadas foram definidas da seguinte forma: a hipótese nula (H_0) estabelece que não há diferença significativa nas médias entre as duas profundidades, enquanto a hipótese alternativa (H_1) indica que existe uma diferença significativa nas médias entre as profundidades de 20 cm e 40 cm.

3.12.2 Teste de Homogeneidade de Variâncias (Levene)

Antes da aplicação de testes paramétricos como ANOVA, foi verificada a homogeneidade das variâncias entre grupos utilizando o teste de Levene.

O procedimento adotado iniciou-se com a aplicação do teste de Levene para cada nutriente, com o objetivo de comparar as variâncias entre as profundidades ou entre os *clusters* formados. Para todos os testes, foi utilizado um nível de significância de $\alpha = 0,05$.

As hipóteses testadas foram definidas da seguinte forma: a hipótese nula (H_0) estabelece que as variâncias são homogêneas entre os grupos, enquanto a hipótese alternativa (H_1) indica que as variâncias não são homogêneas. Os resultados obtidos a partir do teste de Levene foram utilizados para orientar a escolha entre testes estatísticos paramétricos ou não-paramétricos subsequentes, garantindo a validade das análises comparativas.

3.12.3 Análise de Variância (ANOVA)

Para comparar as médias de nutrientes entre múltiplos grupos, como *clusters*, variedades ou estágios de corte, foi aplicada a Análise de Variância (ANOVA).

No caso da ANOVA One-Way, este teste foi utilizado para comparar médias entre mais de dois grupos independentes. O nível de significância adotado foi $\alpha = 0,05$. As hipóteses testadas foram definidas da seguinte forma: a hipótese nula (H_0) estabelece que todas as médias dos grupos são iguais, enquanto a hipótese alternativa (H_1) indica que pelo menos uma média difere das demais.

O teste post-hoc de Tukey HSD foi aplicado nos casos em que a ANOVA indicou diferenças significativas ($p < 0,05$). Este teste permite identificar quais pares de grupos apresentam diferenças estatisticamente significativas entre suas médias. Além disso, o Tukey HSD controla a taxa de erro tipo I (*family-wise error rate*, FWER) em 0,05, utilizando ajuste de Bonferroni, e fornece intervalos de confiança para as diferenças entre as médias, permitindo uma interpretação mais precisa das comparações múltiplas.

3.12.4 Teste de Kruskal-Wallis

Para situações em que as premissas de normalidade ou homogeneidade de variâncias não foram atendidas, foi aplicado o teste não-paramétrico de Kruskal-Wallis. Este teste constitui uma alternativa não-paramétrica à ANOVA e permite comparar as distribuições de três ou mais grupos independentes.

O Kruskal-Wallis baseia-se na comparação de postos (*ranks*) ao invés de médias, proporcionando uma análise robusta quando os dados não seguem distribuição normal ou apresentam variâncias heterogêneas. O nível de significância adotado foi $\alpha = 0,05$. As hipóteses testadas são definidas da seguinte forma: a hipótese nula (H_0) estabelece que as distribuições dos grupos são iguais, enquanto a hipótese alternativa (H_1) indica que pelo menos uma distribuição difere das demais.

3.12.5 Correlação de Spearman

Além da correlação de Pearson, descrita anteriormente, foi utilizada a correlação de Spearman para avaliar relações monotônicas entre variáveis, especialmente quando as premissas de linearidade não foram satisfeitas.

A correlação de Spearman mede a associação entre variáveis com base em postos (*ranks*), sendo adequada para dados ordinais ou quando a relação entre as variáveis não apresenta caráter linear. O coeficiente resultante varia de -1 a +1, e sua interpretação é similar à do coeficiente de Pearson, indicando força e direção da relação entre as variáveis. Para todas as análises, adotou-se um nível de significância $\alpha = 0,05$ para testar a significância das

correlações observadas.

3.12.6 Interpretação dos Resultados

Para todos os testes estatísticos aplicados, os resultados foram interpretados considerando o *p-valor* obtido em relação ao nível de significância estabelecido ($\alpha = 0,05$). Além disso, a magnitude do efeito observado, como diferenças entre médias ou força das correlações, foi analisada para fornecer uma visão mais completa dos resultados.

A interpretação também levou em conta a relevância prática e agronômica das diferenças estatísticas detectadas, garantindo que os achados tivessem significado para o manejo do solo e da produtividade da cana-de-açúcar. Por fim, a consistência dos resultados foi avaliada à luz do conhecimento teórico e empírico disponível sobre fertilidade do solo, permitindo conclusões confiáveis e fundamentadas.

Os resultados detalhados de todos os testes estatísticos, incluindo quadros com valores de estatísticas de teste, *p-valores* e interpretações, são apresentados no Capítulo 4.

3.13 Procedimentos para Análise da Relação entre Nutrientes e Produção

Para investigar como os nutrientes do solo influenciam a produtividade da cana-de-açúcar, foram conduzidas análises específicas de correlação e importância de variáveis relacionando os dados de solo com os dados de produção.

3.13.1 Correlação entre Nutrientes e Variáveis de Produção

Foi realizada análise de correlação de Pearson entre os nutrientes do solo, nas profundidades de 20 cm e 40 cm, e as variáveis de produção, como Produtividade Real, TCH Real, Brix, ATR, Fibra, Pureza, entre outras. Essa análise permitiu avaliar a intensidade e a direção das relações lineares entre os fatores nutricionais e os indicadores produtivos da cana-de-açúcar.

O procedimento consistiu inicialmente no cálculo da matriz de correlação completa entre todas as variáveis de solo e de produção. Em seguida, foram identificadas as correlações estatisticamente significativas, considerando um nível de significância de $p < 0,05$. As correlações foram classificadas por magnitude em fracas, moderadas ou fortes, além de serem separadas em positivas, indicando relação direta, e negativas, indicando relação inversa.

A análise foi realizada de forma independente para as profundidades de 20 cm e 40 cm, permitindo identificar em qual camada os nutrientes exercem maior influência sobre a produção, bem como observar possíveis diferenças na dinâmica nutricional entre o solo superficial e o subsuperficial.

Na etapa de seleção das correlações mais relevantes, foram identificadas as 10 correlações positivas mais fortes, destacando as relações diretas mais consistentes entre nutrientes e variáveis de produção. De forma complementar, também foram selecionadas as 10 correlações negativas mais intensas, evidenciando as relações inversas mais significativas.

O foco dessa análise foi direcionado às correlações que apresentaram maior relevância agronômica, permitindo priorizar aquelas com potencial aplicação prática no manejo do solo e na otimização da produtividade da cana-de-açúcar.

3.13.2 Análise de Importância de Nutrientes para a Produtividade

Para quantificar a contribuição individual de cada nutriente na predição da produtividade, foi utilizado o algoritmo `RandomForestRegressor` da biblioteca `scikit-learn`. Inicialmente, os dados foram preparados selecionando os nutrientes como variáveis independentes (*features*) e as variáveis de produção como variáveis dependentes (*targets*). Nessa etapa, também foram tratados valores ausentes e realizadas padronizações quando necessário, garantindo consistência para o treinamento do modelo.

O treinamento do modelo consistiu no ajuste de modelos Random Forest para prever a Prod. Real e o Tch Real, sendo realizados separadamente para as profundidades de 20 cm e 40 cm. A validação cruzada foi aplicada para assegurar a robustez e confiabilidade dos resultados obtidos.

Após o treinamento, as importâncias das variáveis (*feature importances*) foram extraídas, permitindo ordenar os nutrientes por importância decrescente. Gráficos de barras foram gerados para destacar visualmente os nutrientes mais influentes na predição da produtividade.

Finalmente, a interpretação dos resultados possibilitou identificar os nutrientes com maior poder preditivo, comparar as importâncias entre as diferentes profundidades do solo e realizar uma análise agronômica dos nutrientes prioritários, fornecendo subsídios para decisões de manejo voltadas ao aumento da produtividade da cana-de-açúcar.

3.13.3 Comparação entre Profundidades

Para avaliar qual profundidade de amostragem apresenta nutrientes com maior influência sobre a produção, foram comparados os coeficientes de correlação entre os nutrientes e as variáveis de produção nas duas profundidades analisadas. Além disso, as importâncias dos nutrientes extraídas dos modelos Random Forest foram utilizadas como critério complementar, permitindo verificar quais nutrientes possuem maior poder preditivo em cada camada do solo.

A consistência dos resultados entre as diferentes abordagens, tanto a correlação quanto a importância de variáveis, foi analisada para identificar padrões robustos e confiáveis. Essa comparação possibilitou determinar em qual profundidade os nutrientes exercem maior impacto

sobre a produtividade, oferecendo informações relevantes para o manejo agrônomo e a tomada de decisão.

Os resultados detalhados dessas análises, incluindo as matrizes de correlação, gráficos de importância e interpretações agrônômicas, são apresentados no Capítulo 4.

3.14 Procedimentos para Construção dos Modelos Preditivos

Para possibilitar a predição da produção e produtividade da cana-de-açúcar com base nas características nutricionais do solo, foram desenvolvidos modelos de regressão linear múltipla.

Foram construídos dois modelos independentes, um para a profundidade de 20 cm e outro para 40 cm, devido ao comportamento diferenciado do solo. As camadas superficiais e subsuperficiais apresentam dinâmicas nutricionais distintas, com diferentes concentrações e mobilidades de nutrientes, o que justifica a análise separada.

Além disso, a absorção de nutrientes pelas raízes varia conforme a profundidade, sendo alguns elementos mais relevantes em determinadas camadas. Modelos específicos permitem capturar essas particularidades, melhorando a precisão das predições.

Por fim, a combinação dos resultados dos dois modelos, por meio da média das estimativas, proporciona uma previsão final mais equilibrada e robusta da produtividade total, integrando as informações das diferentes camadas do solo.

As variáveis independentes (preditoras) incluíram todos os nutrientes analisados: P, M.O, pH, K, K/T, Ca, Ca/T, Mg, Mg/T, Al, H+Al, S, S.B, CTC, V e Sat., sendo que todos os dados foram previamente normalizados conforme os procedimentos descritos na Seção 3.7.2. As variáveis dependentes (respostas) consideradas foram a produção real (**Prod. Real**, em toneladas) e as toneladas de cana por hectare (**Tch Real**).

Os dados foram divididos em conjuntos de treinamento e teste, sendo 80% destinados ao treinamento dos modelos e 20% reservados para avaliação da capacidade preditiva. A divisão foi realizada de forma aleatória, garantindo a representatividade de ambos os conjuntos.

Foram ajustados modelos de regressão linear múltipla utilizando a classe `LinearRegression` da biblioteca `scikit-learn`. Durante o treinamento, os coeficientes de regressão foram estimados pelo método dos mínimos quadrados ordinários, determinando os parâmetros β que minimizam a soma dos resíduos quadrados. Os modelos foram treinados separadamente para as profundidades de 20 cm e 40 cm. Na etapa de validação, os modelos treinados foram aplicados ao conjunto de teste para calcular predições de Prod. Real e Tch Real, permitindo a comparação com os valores observados.

A qualidade e a capacidade preditiva dos modelos foram avaliadas pelo coeficiente de determinação (R^2) e pelo erro quadrático médio (RMSE). O R^2 mede a proporção da variância da variável resposta explicada pelo modelo, variando de 0 a 1, com valores próximos

a 1 indicativos de melhor ajuste, calculado pela equação $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$. O RMSE quantifica a magnitude média dos erros de predição, expresso nas mesmas unidades da variável resposta, sendo calculado por $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$, onde valores menores indicam predições mais precisas.

Para facilitar a interpretação, foram gerados gráficos de dispersão comparando valores reais versus valores preditos. Cada ponto representa uma observação do conjunto de teste, e linhas de referência indicam a predição perfeita, enquanto linhas de tendência mostram possíveis desvios sistemáticos. Foram produzidos gráficos separados para cada modelo (20 cm e 40 cm) e para cada variável resposta (Prod. Real e Tch Real).

Os coeficientes de regressão foram analisados para identificar os nutrientes com maior influência positiva ou negativa sobre a produção, a magnitude do efeito de cada nutriente (quantificada pelo valor absoluto de β_j), a consistência entre os modelos de 20 cm e 40 cm e as implicações práticas para manejo nutricional. A significância de cada coeficiente foi avaliada por testes t, verificando $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$, utilizando a estatística:

$$t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, \quad (3.9)$$

onde $SE(\hat{\beta}_j)$ é o erro padrão do coeficiente estimado.

Para obter a estimativa final de produção e produtividade, foram realizadas predições separadas utilizando os modelos de 20 cm ($\hat{y}_{20cm}$) e 40 cm ($\hat{y}_{40cm}$). A predição final foi calculada pela média aritmética das duas estimativas:

$$\hat{y}_{final} = \frac{\hat{y}_{20cm} + \hat{y}_{40cm}}{2}, \quad (3.10)$$

onde:

- $\hat{y}_{final}$ é a predição final de produção ou produtividade, obtida integrando as contribuições das duas profundidades;
- $\hat{y}_{20cm}$ é a predição realizada com base no modelo ajustado para a profundidade de 20 cm;
- $\hat{y}_{40cm}$ é a predição realizada com base no modelo ajustado para a profundidade de 40 cm.

Integrando as contribuições das camadas superficial e subsuperficial do solo e proporcionando uma previsão mais equilibrada e robusta.

Os resultados detalhados, incluindo valores de R^2 e RMSE, gráficos de dispersão, análise dos coeficientes e interpretações agronômicas, são apresentados no Capítulo 4.

4 Resultados

Este capítulo apresenta os resultados obtidos a partir das análises estatísticas, inferenciais e de modelagem descritas no Capítulo 3. Os resultados estão organizados em três seções principais, que abordam diferentes aspectos do estudo: análises sobre os nutrientes do solo; análises sobre a produção; e as relações entre nutrientes e produção, incluindo a construção de modelos preditivos.

4.1 Análises sobre os Nutrientes

Esta seção apresenta os resultados das análises realizadas sobre os nutrientes do solo, abrangendo desde a caracterização inicial dos dados até análises estatísticas mais complexas, como testes de comparação entre profundidades, correlações, PCA e clusterização.

4.1.1 Caracterização Inicial dos Dados

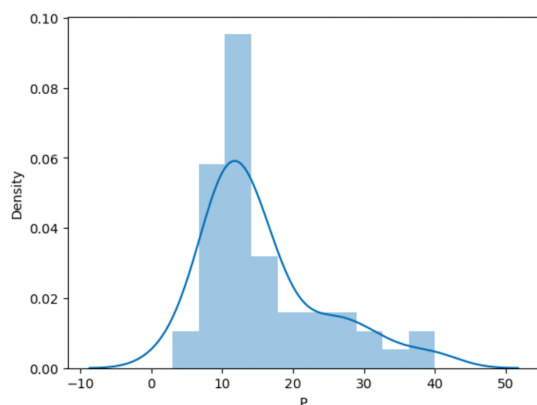
Antes de apresentar as análises inferenciais e de modelagem, esta seção caracteriza inicialmente os dados de solo, incluindo estatísticas descritivas, análise de normalidade e identificação de *outliers*.

4.1.1.1 Análise de Normalidade e Identificação de *Outliers*

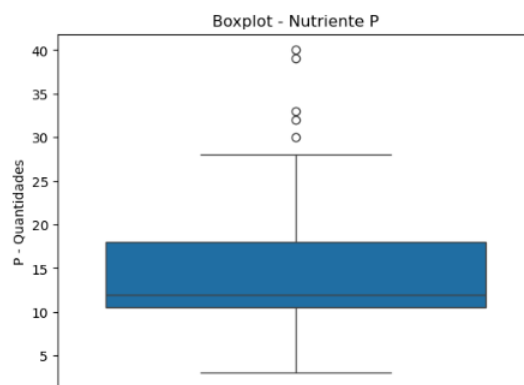
As análises realizadas para as profundidades de 20 cm e 40 cm permitiram caracterizar a distribuição dos principais nutrientes e parâmetros químicos do solo nessas camadas. Para cada nutriente, foram calculadas as estatísticas descritivas e aplicado o teste de Shapiro-Wilk para verificar a normalidade das distribuições.

A figura 25 apresenta, para a profundidade de 20 cm, os histogramas e boxplots dos nutrientes P, pH e Ca, organizados lado a lado para facilitar a análise conjunta. Essa disposição permite observar a distribuição das amostras, a dispersão dos valores e a presença de assimetrias ou possíveis *outliers*. Para cada nutriente, o histograma mostra a frequência dos valores registrados, enquanto o boxplot evidencia a tendência central e a variabilidade, oferecendo uma visão clara do comportamento estatístico das amostras e auxiliando na avaliação da normalidade.

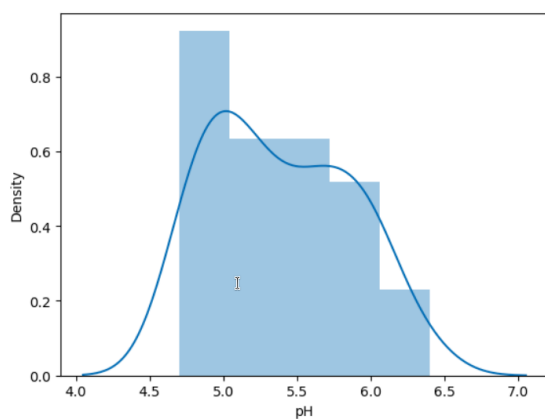
Figura 25 – Histogramas e boxplots dos nutrientes P, pH e Ca para amostras coletadas a 20 cm de profundidade.



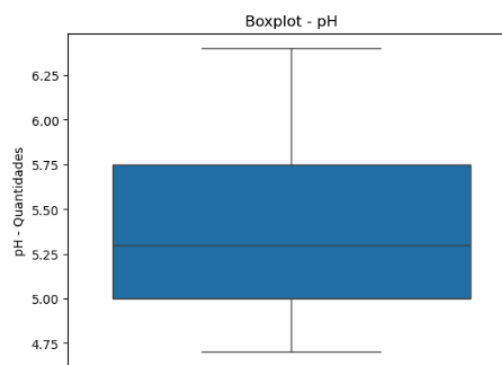
(a) Histograma do nutriente P.



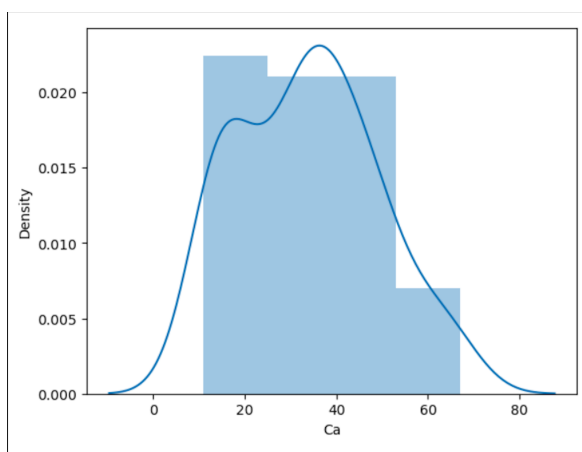
(b) Boxplot do nutriente P.



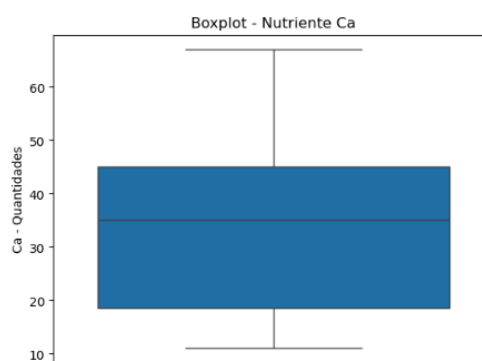
(c) Histograma do pH.



(d) Boxplot do pH.



(e) Histograma do nutriente Ca.



(f) Boxplot do nutriente Ca.

Fonte: Elaborada pelo autor.

Após a visualização, observa-se que o nutriente P apresenta um histograma com assimetria à direita, indicando maior concentração de valores baixos e ocorrência de valores

altos menos frequentes. Seu boxplot confirma esse comportamento ao evidenciar a presença de *outliers*. Para o pH, o histograma apresenta formato mais próximo de uma distribuição normal, comportamento reforçado pelo boxplot, que não mostra *outliers*. O mesmo padrão é observado para o Ca, cujo histograma também se aproxima da normalidade e cujo boxplot não indica valores discrepantes, sugerindo uma distribuição mais homogênea para esse nutriente.

Os quadros 4 e 5 apresentam, respectivamente, os resultados consolidados do teste de Shapiro-Wilk e a identificação de *outliers* para as profundidades de 20 cm e 40 cm.

Quadro 4 – Classificação de normalidade e presença de *outliers* para os nutrientes a 20 cm de profundidade.

Nutriente	Distribuição	P-Value	Outliers
<i>P</i>	Não Normal	6.58×10^{-5}	Sim
<i>M.O</i>	Não Normal	0.0021	Sim
<i>pH</i>	Não Normal	0.0194	Não
<i>K</i>	Não Normal	4.84×10^{-5}	Sim
<i>K/T</i>	Não Normal	1.49×10^{-5}	Sim
<i>Ca</i>	Não Normal	0.0283	Não
<i>Ca/T</i>	Normal	0.0754	Não
<i>Mg</i>	Não Normal	0.0403	Não
<i>Mg/T</i>	Normal	0.4965	Sim
<i>Al</i>	Não Normal	7.72×10^{-16}	Sim
<i>H+Al</i>	Normal	0.2537	Não
<i>S</i>	Normal	1.0	Não
<i>S.B</i>	Não Normal	0.2222	Não
<i>CTC</i>	Normal	0.0716	Não
<i>V</i>	Não Normal	0.0042	Não
<i>Sat.</i>	Não Normal	7.72×10^{-16}	Sim

Fonte: Elaborado pelo autor.

Quadro 5 – Classificação de normalidade e presença de *outliers* para os nutrientes a 40 cm de profundidade.

Nutriente	Distribuição	P-Value	Outliers
<i>P</i>	Não Normal	9.95×10^{-5}	Sim
<i>M.O</i>	Não Normal	0.0212	Sim
<i>pH</i>	Normal	0.2271	Não
<i>K</i>	Não Normal	8.12×10^{-6}	Sim
<i>K/T</i>	Não Normal	1.22×10^{-6}	Sim
<i>Ca</i>	Não Normal	0.0022	Não
<i>Ca/T</i>	Normal	0.6462	Não
<i>Mg</i>	Não Normal	0.0035	Não
<i>Mg/T</i>	Normal	0.2171	Não
<i>Al</i>	Não Normal	1.09×10^{-13}	Sim
<i>H+Al</i>	Normal	0.1933	Sim
<i>S</i>	Não Normal	3.98×10^{-5}	Não
<i>S.B</i>	Não Normal	0.0011	Não
<i>CTC</i>	Não Normal	0.0265	Não
<i>V</i>	Normal	0.4118	Não
<i>Sat.</i>	Não Normal	1.01×10^{-13}	Sim

Fonte: Elaborado pelo autor.

A análise dos resultados do teste de Shapiro-Wilk revelou que a maioria dos nutrientes apresentou distribuições não normais em ambas as profundidades. Para a profundidade de 20 cm, apenas cinco nutrientes (*Ca/T*, *Mg/T*, *H+Al*, *S* e *CTC*) apresentaram distribuição normal ($p > 0,05$), representando aproximadamente 31% do total de variáveis analisadas. Para a profundidade de 40 cm, a situação foi similar, com cinco nutrientes (*pH*, *Ca/T*, *Mg/T*, *H+Al* e *V*) apresentando normalidade.

De modo geral, observou-se que poucos nutrientes apresentaram distribuição normal, especialmente nas camadas mais superficiais. A presença de *outliers* foi identificada em aproximadamente 38% dos nutrientes na profundidade de 20 cm e 31% na profundidade de 40 cm. A alta variabilidade de elementos como P, K e Ca, evidenciada tanto pelos valores de p-value extremamente baixos quanto pela presença de *outliers*, indica forte heterogeneidade espacial, possivelmente decorrente de variações nas práticas de manejo e fertilização entre as diferentes quadras e lotes analisados.

4.1.1.2 Estatísticas Descritivas

Para complementar a análise de normalidade, os quadros 6 e 7 apresentam as medidas descritivas para ambas as profundidades, permitindo uma comparação quantitativa da variabilidade dos nutrientes entre as camadas do solo.

Quadro 6 – Medidas descritivas dos nutrientes do solo a 20 cm de profundidade.

Nutriente	Média	Mediana	Desvio-Padrão	Variância
<i>P</i>	15.70	12.00	8.54	73.01
<i>M.O</i>	15.16	15.00	3.74	13.97
<i>pH</i>	5.39	5.30	0.48	0.23
<i>K</i>	1.24	1.00	0.60	0.36
<i>K/T</i>	1.93	1.55	1.04	1.09
<i>Ca</i>	33.78	35.00	15.11	228.25
<i>Ca/T</i>	47.24	48.16	11.44	130.96
<i>Mg</i>	14.02	16.00	6.82	46.54
<i>Mg/T</i>	19.50	20.39	6.18	38.26
<i>Al</i>	0.03	0.00	0.28	0.08
<i>H+Al</i>	23.92	24.00	6.79	46.07
<i>S</i>	0.00	0.00	0.00	0.00
<i>S.B</i>	49.05	52.00	21.88	478.99
<i>CTC</i>	72.97	72.40	20.79	432.36
<i>V</i>	64.34	67.24	14.43	208.29
<i>Sat.</i>	0.19	0.00	1.34	1.79

Fonte: Elaborado pelo autor.

Quadro 7 – Medidas descritivas dos nutrientes do solo a 40 cm de profundidade.

Nutriente	Média	Mediana	Desvio-Padrão	Variância
<i>P</i>	10.15	9.00	6.69	44.73
<i>M.O</i>	12.21	13.00	3.93	15.49
<i>pH</i>	5.11	5.10	0.33	0.11
<i>K</i>	0.69	0.60	0.33	0.11
<i>K/T</i>	1.33	1.18	0.77	0.59
<i>Ca</i>	22.84	21.00	10.13	102.73
<i>Ca/T</i>	39.38	38.99	9.41	88.63
<i>Mg</i>	10.31	9.00	5.20	27.09
<i>Mg/T</i>	17.63	17.39	5.64	31.89
<i>Al</i>	0.31	0.00	0.99	0.98
<i>H+Al</i>	29.54	30.00	7.86	61.81
<i>S</i>	31.92	21.00	20.66	426.87
<i>S.B</i>	33.84	30.50	15.24	232.52
<i>CTC</i>	63.39	62.40	16.55	273.90
<i>V</i>	51.60	54.30	13.23	175.24
<i>Sat.</i>	1.23	0.00	3.94	15.53

Fonte: Elaborado pelo autor.

A avaliação conjunta das medidas descritivas revelou padrões consistentes entre as duas profundidades analisadas. A camada de 20 cm apresentou, em geral, maiores valores médios e maior variância para a maioria dos nutrientes, especialmente para *P* (média de 15,70 mg/dm³

contra 10,15 mg/dm³ a 40 cm), *K* (1,24 mmol_c/dm³ contra 0,69 mmol_c/dm³) e *Ca* (33,78 mmol_c/dm³ contra 22,84 mmol_c/dm³). Esse comportamento pode estar relacionado à maior influência das práticas de manejo e adubação na camada superficial, onde ocorre a aplicação direta de fertilizantes e corretivos.

Já na profundidade de 40 cm, observou-se menor variabilidade para a maioria das variáveis, evidenciada pelos menores valores de desvio-padrão e variância, indicando maior estabilidade química e menor interferência antrópica nesta camada. Destaca-se, entretanto, o comportamento do *H+Al* (acidez potencial), que apresentou valores médios superiores a 40 cm (29,54 mmol_c/dm³) em comparação com 20 cm (23,92 mmol_c/dm³), sugerindo maior acidez em profundidade, possivelmente devido à menor eficiência da calagem nesta camada.

Um aspecto notável foi observado para o nutriente *S* (enxofre), que apresentou valor médio de 0,00 a 20 cm (sem variabilidade) e 31,92 mg/dm³ a 40 cm, com alta variância (426,87), indicando forte heterogeneidade na distribuição deste elemento em profundidade e possivelmente refletindo processos de lixiviação ou mobilidade diferenciada no perfil do solo.

4.1.1.3 Normalização dos Dados

Considerando que a maioria dos nutrientes apresentou distribuições não normais, conforme demonstrado nas análises anteriores, e seguindo os procedimentos metodológicos descritos na Seção 3.7.2, procedeu-se à etapa de normalização dos dados.

O quadro 8 apresenta as transformações aplicadas aos nutrientes coletados a 20 cm de profundidade, juntamente com os respectivos *p-values* obtidos após cada transformação pelo teste de Shapiro-Wilk.

Quadro 8 – Transformações aplicadas aos nutrientes a 20 cm de profundidade e respectivos *p-values* pós-transformação.

Nutriente	Transformação Aplicada	P-Value Pós-Transformação
P	Box-Cox com Deslocamento	0,0955
M.O	Logarítmica	0,1095
pH	Rankit	0,6299
K	Box-Cox com Deslocamento	0,0546
K/T	Logarítmica	0,1631
Ca	Johnson Transformation	0,9997
Mg	Johnson Transformation	0,9928
Al	Média sobre os <i>outliers</i>	1,0000
S.B	Johnson Transformation	0,9904
V	Johnson Transformation	0,3551
Sat.	Média sobre os <i>outliers</i>	1,0000

Fonte: Elaborado pelo autor.

Observa-se que todas as variáveis analisadas atingiram aproximação satisfatória à

normalidade após as transformações aplicadas, com todos os *p-values* superiores a 0,05. As transformações mais utilizadas foram a Johnson Transformation, aplicada em 36% dos casos (Ca, Mg, S.B e V), e as transformações Box-Cox com deslocamento e logarítmica, cada uma aplicada em 18% dos casos. Para os nutrientes Al e Sat., que apresentavam *outliers* extremos, optou-se pela substituição desses valores pela média, resultando em *p-values* perfeitos (1,0000).

De forma análoga, o quadro 9 apresenta as transformações aplicadas aos nutrientes coletados a 40 cm de profundidade.

Quadro 9 – Transformações aplicadas aos nutrientes a 40 cm de profundidade e respectivos *p-values* pós-transformação.

Nutriente	Transformação Aplicada	P-Value Pós-Transformação
<i>P</i>	Box-Cox com Deslocamento	0,2030
<i>M.O</i>	Logarítmica	0,1395
<i>K</i>	Box-Cox com Deslocamento	0,1475
<i>K/T</i>	Box-Cox	0,7325
<i>Ca</i>	Johnson Transformation	0,0710
<i>Mg</i>	Johnson Transformation	0,0549
Al	Média sobre os <i>outliers</i>	1,0000
S	Johnson Transformation	0,1188
S.B	Johnson Transformation	0,1183
CTC	Johnson Transformation	0,7930
Sat.	Média sobre os <i>outliers</i>	1,0000

Fonte: Elaborado pelo autor.

Para a profundidade de 40 cm, todas as variáveis também atingiram normalidade após as transformações, com *p-values* superiores a 0,05. Nota-se que alguns nutrientes, como Ca e Mg, apresentaram *p-values* ligeiramente acima do limiar (0,0710 e 0,0549, respectivamente), mas ainda dentro dos critérios de aceitação para normalidade. A Johnson Transformation foi a técnica mais empregada nesta profundidade, aplicada em 45% dos casos, refletindo a natureza mais complexa das distribuições dos nutrientes em camadas mais profundas.

Apesar de algumas diferenças nas transformações aplicadas entre as duas profundidades, o objetivo foi o mesmo: garantir aproximação à normalidade, reduzir o impacto de valores extremos e permitir comparações estatísticas confiáveis entre as camadas do solo. A escolha de transformações distintas para o mesmo nutriente em profundidades diferentes (como K/T, que recebeu transformação logarítmica a 20 cm e Box-Cox a 40 cm) reflete as particularidades de cada distribuição e reforça a importância de uma abordagem individualizada para cada conjunto de dados.

Em síntese, a etapa de normalização foi essencial para viabilizar a aplicação das análises estatísticas paramétricas apresentadas nas seções subsequentes, garantindo que as premissas dos testes fossem adequadamente atendidas e que os resultados obtidos fossem estatisticamente

válidos e interpretáveis.

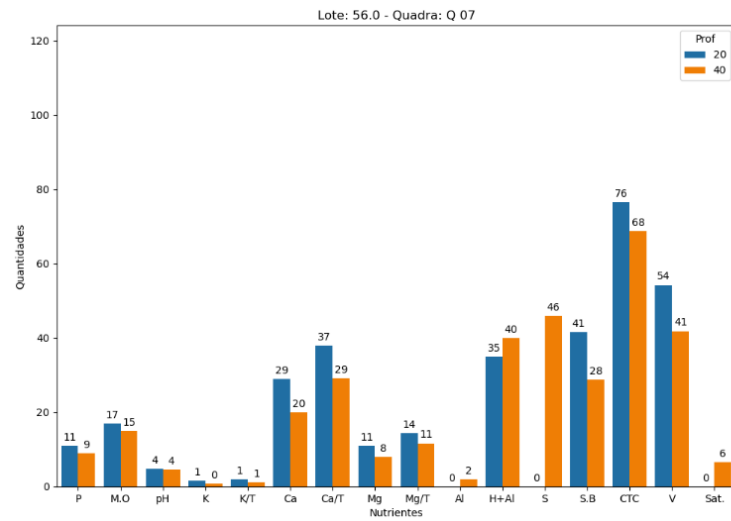
4.1.2 Distribuição Nutricional por Lote e Quadra

A partir dos dados normalizados e validados, procedeu-se à análise das concentrações médias dos nutrientes por lote e quadra, considerando as profundidades de 20 cm e 40 cm. Essa abordagem estratificada permitiu identificar com precisão quais nutrientes se encontram dentro ou fora dos níveis considerados ideais, auxiliando na compreensão da variabilidade espacial da fertilidade do solo e no diagnóstico de possíveis deficiências ou excessos nutricionais.

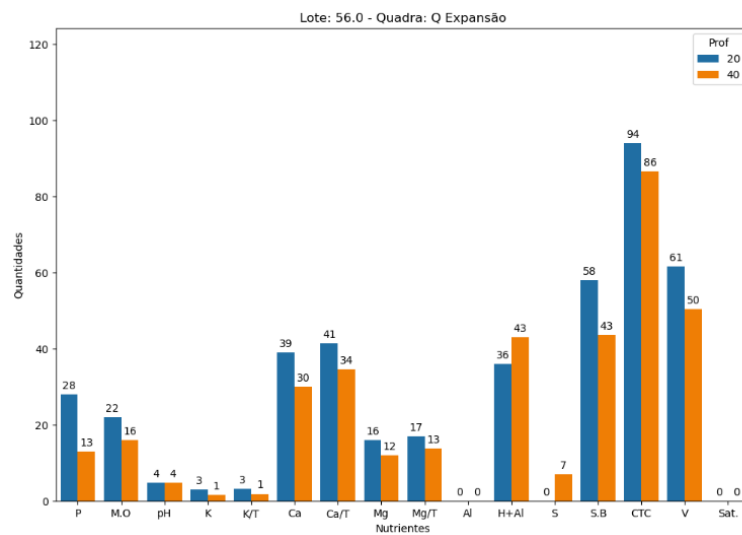
4.1.2.1 Variação Vertical dos Nutrientes

A figura 26 apresenta três exemplos representativos dos gráficos comparativos gerados para todos os lotes e quadras do estudo, referentes aos lotes 56 e 75. Em cada gráfico, as barras azuis correspondem às concentrações médias a 20 cm e as laranjas às médias a 40 cm, permitindo uma visualização clara das diferenças entre as profundidades analisadas.

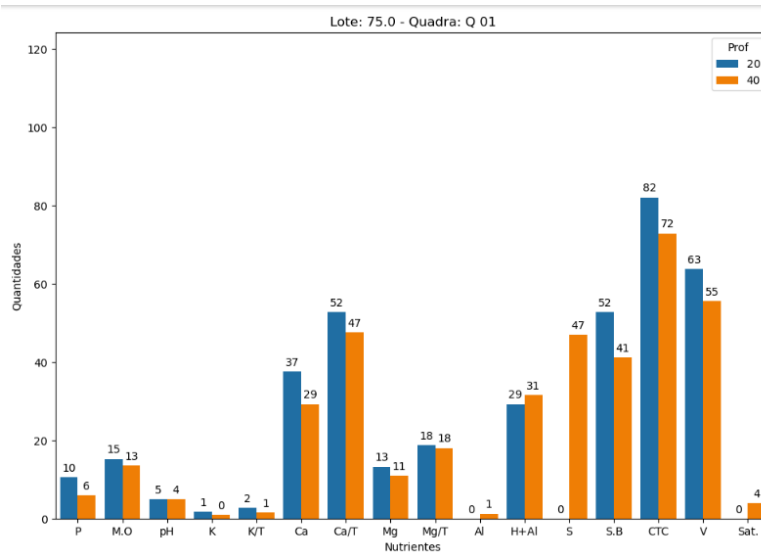
Figura 26 – Exemplos representativos dos gráficos comparativos de médias de nutrientes por profundidade (20 cm em azul e 40 cm em laranja) gerados para cada lote e quadra.



(a) Lote 56 — Quadra 07.



(b) Lote 56 — Quadra Exp.



(c) Lote 75 — Quadra 01.

Os gráficos evidenciam que, em diversas situações, há redução nas concentrações em camadas mais profundas. Nutrientes como *P* e *K* apresentaram maior variabilidade entre as camadas, sugerindo mobilidade e possível lixiviação em determinadas regiões. Essa variação vertical é um indicativo importante da dinâmica de movimentação de nutrientes no perfil do solo.

4.1.2.2 Síntese dos Nutrientes por Lote e Quadra

O quadro 10 apresenta uma compilação detalhada dos nutrientes que se mantiveram dentro dos níveis ideais nas duas profundidades para cada lote e quadra analisados. Esta síntese permite identificar rapidamente as condições nutricionais específicas de cada área.

Quadro 10 – Síntese dos nutrientes dentro dos níveis ideais em cada lote/quadra e profundidade.

Lote	Quadra	Nutrientes ideais (20 cm)	Nutrientes ideais (40 cm)
75	Q 01	M.O, Ca, Mg, Al, S.B, CTC, Sat.	Ca, Mg, S, Sat.
56	Q 07	M.O, Ca, Mg, Mg/T, Al, Sat.	M.O, Ca, S, Sat.
56	Q Expansão	M.O, K, K/T, Ca, Mg, Al, S.B, CTC, Sat.	M.O, Ca, Mg, Mg/T, Al, CTC, Sat.
56	Q 08 e Q 12	M.O, Ca, Mg, Al, S.B, CTC, Sat.	Ca, Mg, Al, S, S.B, CTC, Sat.
56	Q 01 e Q02	M.O, pH, Ca/T, Mg, Al, H+Al, CTC, V, Sat.	Ca, Ca/T, Mg, Al, S, S.B, CTC, Sat.
56	Q 01, Q02, Q05 e Q06	M.O, pH, K/T, Mg/T, Al, Sat.	M.O, Al, S, Sat.
56	Q 07-08-12	M.O, Ca, Mg, Al, Sat.	S
56	Q 03, Q04 e Q09	M.O, pH, Ca/T, Al, H+Al, S.B, CTC, V, Sat.	M.O, Mg, Al, S, Sat.
56	Q 09, Q10 e Q11	M.O, pH, Ca/T, Al, H+Al, S.B, CTC, V, Sat.	M.O, pH, Al, H+Al, S, S.B, CTC, V, Sat.
56	Q 07, Q 10 e Q 11	M.O, pH, Ca/T, Mg, Al, S.B, CTC, V, Sat.	Ca, Mg, Al, S, CTC, Sat.
56	Q 03, Q 04, Q 05, Q 08 e Q 12	P, M.O, Ca/T, Al, S.B, CTC, Sat.	Ca, Ca/T, Mg, Al, S, CTC, Sat.
06	Q 01	Mg, Al, S.B, CTC, V, Sat.	Ca, Mg, Al, S, Sat.
06	Q 01 e Q 02	pH, Mg, Al, H+Al, S.B, CTC, V, Sat.	Ca, Mg, Al, S, V
06	Q 01 a Q 03	pH, Ca, Mg, Al, H+Al, S.B, V, Sat.	Mg, Al, H+Al, S, Sat.
50	Q 01 e Q 02	Ca, Al, Sat.	Al, S, Sat.
50	Q 03	Ca, Mg/T, Al, Sat.	Mg/T, Al, S, Sat.
50	Q 04	P, M.O, pH, Al, H+Al, CTC, V, Sat.	P, M.O, pH, Mg, Al, S.B, CTC, V, Sat.
69	Q 01	M.O, Ca, Mg, Al, S.B, CTC, Sat.	M.O, Ca, Mg, Al, S, Sat.
70	Q 01	P, M.O, K/T, Ca, Mg, Al, S.B, CTC, Sat.	M.O, Ca, Mg, Al, S, Sat.
55	Q 01 e Q 02	M.O, pH, Ca, Mg, Al, H+Al, V, Sat.	Ca, Mg/T, Al, S, Sat.
54	Q 01 a Q 03	pH, Mg, Al, H+Al, S.B, V, Sat.	Ca, Mg, Al, S, Sat.
67	Q 01, Q 02 e Q 03	M.O, K/T, Al, Sat.	K/T, Al, S, Sat.
73	Q 01 e Q 02	P, M.O, K/T, Al, Sat.	M.O, Mg/T, S
74	Q 01 e Q 02	M.O, Sat.	M.O, K/T, S
05	Q 01 e Q 02	M.O, Al, Sat.	M.O, Ca, Mg, Al, S, Sat.
12	Q 01 e Q 02	pH, Mg, Al, H+Al, S.B, V, Sat.	pH, Mg, Al, H+Al, S, V, Sat.
65	Q 01, Q 02 e Q 03	pH, Mg, Al, H+Al, S.B, CTC, V, Sat.	Ca, Mg, Al, S, Sat.
64	Q 01 e Q 02	Ca, Mg, Al, Sat.	Mg/T, Al, S, Sat.
63	Q 01 e Q 02	pH, Mg, Al, S.B, CTC, V, Sat.	Ca, Mg, Al, S, CTC, Sat.
61	Q 01 e Q 02	pH, Ca/T, Mg, Al, H+Al, S.B, CTC, V, Sat.	pH, Mg, Al, S, S.B, CTC, V, Sat.
51	Q 01, Q 04 e Q 05	Al, Sat.	Al, S, Sat.
51	Q 02 e Q 03	Mg/T, Al, Sat.	Al, Sat.

Fonte: Elaborado pelo autor.

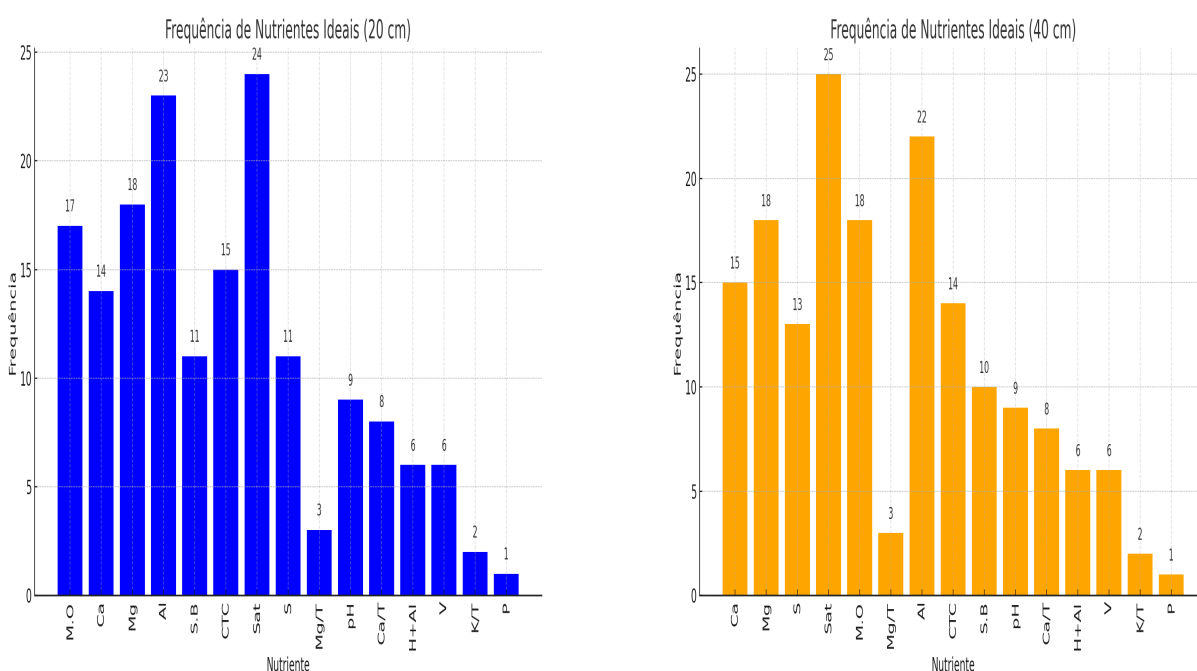
A análise do quadro 10 revela que nutrientes como *M.O*, *Ca*, *Mg*, *Al* e *a* (*Sat.*)

apresentaram maior frequência de valores dentro da faixa ideal, indicando uma condição relativamente equilibrada para esses elementos na maioria das áreas estudadas.

4.1.2.3 Padrões de Distribuição por Profundidade

Para uma compreensão mais ampla da distribuição vertical dos nutrientes, a figura 27 apresenta as frequências agregadas de nutrientes ideais para as profundidades de 20 cm e 40 cm, considerando todos os lotes e quadras analisados.

Figura 27 – Frequências dos nutrientes ideais a 20 cm e 40 cm de profundidade.



(a) Frequência de Nutrientes Ideais (20 cm)

(b) Frequência de Nutrientes Ideais (40 cm)

Fonte: Elaborada pelo autor.

- **Camada superficial (20 cm):** Na profundidade de 20 cm, os nutrientes mais frequentes em níveis ideais são **M.O.**, **Ca**, **Mg**, **Al** e **Saturação**. Essa maior diversidade de nutrientes reflete a intensa atividade biológica e a dinâmica de ciclagem de nutrientes característica da camada superficial. A presença destacada de **CTC** (Capacidade de Troca Catiônica) e **S.B** (Soma de Bases) indica boa capacidade de retenção e disponibilização de nutrientes, condição favorável ao desenvolvimento inicial das plantas.
- **Camada subsuperficial (40 cm):** Na profundidade de 40 cm, destacam-se **Ca**, **Mg**, **S** (Enxofre) e **Saturação**. A manutenção de **Ca** e **Mg** em níveis adequados nesta camada é fundamental para o desenvolvimento radicular em profundidade e para a nutrição das plantas. A presença significativa de **S** é particularmente relevante, uma vez que este nutriente é essencial para a síntese de proteínas vegetais e sua disponibilidade em profundidade contribui para o crescimento robusto das plantas.

A comparação entre as profundidades evidencia que a camada superficial apresenta maior diversidade de nutrientes em níveis ideais, reflexo da concentração de processos biológicos e do aporte de matéria orgânica. Por outro lado, a camada de 40 cm mantém nutrientes estruturais importantes (**Ca** e **Mg**), essenciais para a exploração radicular em profundidade e para a resiliência das culturas em períodos de estresse hídrico. A redução na diversidade de nutrientes em profundidade, combinada com a presença marcante de **S**, sugere processos de lixiviação seletiva e demonstra a importância de estratégias de manejo que considerem a dinâmica vertical dos nutrientes no perfil do solo.

4.1.3 Teste de Comparação de Profundidade

Para verificar se as diferenças observadas nas concentrações de nutrientes entre as profundidades de 20 cm e 40 cm são estatisticamente significativas, foi aplicado o teste *t* pareado para cada nutriente, conforme os procedimentos descritos no Capítulo 3. Este teste permite avaliar se as médias das duas profundidades diferem de forma significativa, considerando o nível de significância de $\alpha = 0,05$.

O quadro 11 apresenta os resultados do teste *t* pareado para todos os nutrientes analisados, evidenciando o grau de significância das diferenças encontradas entre as duas camadas do solo.

Quadro 11 – Resultados do teste *t* pareado para comparação dos nutrientes entre as profundidades de 20 cm e 40 cm.

Nutriente	<i>P-value</i> (Teste <i>t</i>)
<i>P</i>	$1,93 \times 10^{-70}$
<i>M.O</i>	$2,55 \times 10^{-14}$
<i>pH</i>	0,9568
<i>K</i>	$1,18 \times 10^{-21}$
<i>K/T</i>	$9,70 \times 10^{-08}$
<i>Ca</i>	$3,61 \times 10^{-22}$
<i>Ca/T</i>	$1,86 \times 10^{-08}$
<i>Mg</i>	0,0005
<i>Mg/T</i>	0,0012
<i>Al</i>	1,0000
<i>H+Al</i>	$2,41 \times 10^{-11}$
<i>S</i>	$1,41 \times 10^{-18}$
<i>S.B</i>	$5,49 \times 10^{-05}$
<i>CTC</i>	0,0022
<i>V</i>	$1,26 \times 10^{-07}$
<i>Sat.</i>	1,0000

Fonte: Elaborado pelo autor.

4.1.3.1 Nutrientes com Diferenças Altamente Significativas

A maioria dos nutrientes apresentou *p-values* extremamente baixos, indicando diferenças altamente significativas entre as profundidades. Destacam-se:

- **P:** Com $p = 1,93 \times 10^{-70}$, apresentou a diferença mais significativa entre todas as variáveis analisadas. Este resultado é esperado, uma vez que o fósforo possui baixa mobilidade no solo e tende a se acumular nas camadas superficiais, onde ocorre maior atividade biológica e aplicação de fertilizantes.
- **K e suas relações:** O potássio apresentou $p = 1,18 \times 10^{-21}$, enquanto a relação K/T mostrou $p = 9,70 \times 10^{-08}$. Ambos os resultados confirmam diferenças significativas, refletindo a maior concentração deste nutriente na camada superficial, embora sua mobilidade no perfil seja maior que a do fósforo.
- **Ca:** O cálcio obteve $p = 3,61 \times 10^{-22}$. Esse valor indica distribuição significativamente diferente entre as profundidades, o que pode estar relacionado tanto à aplicação superficial de calcário quanto à dinâmica de lixiviação de bases no perfil do solo.

- **Enxofre (S):** Com $p = 1,41 \times 10^{-18}$, o enxofre demonstrou diferença expressiva entre as camadas, possivelmente associada à sua forma aniônica (sulfato), que apresenta maior mobilidade no solo e pode ser lixiviado para camadas mais profundas.
- **Matéria Orgânica (M.O):** Apresentou $p = 2,55 \times 10^{-14}$, confirmando que a concentração de matéria orgânica é significativamente maior na camada superficial, resultado direto do acúmulo de resíduos vegetais e da atividade biológica mais intensa nesta zona.

4.1.3.2 Nutrientes com Diferenças Moderadas

Alguns nutrientes apresentaram *p-values* baixos, mas não tão extremos quanto os anteriores:

- **Magnésio (Mg) e Mg/T:** Com $p = 0,0005$ e $p = 0,0012$, respectivamente, o magnésio mostrou diferença estatisticamente significativa entre as profundidades, embora menos pronunciada que o cálcio. Isso pode indicar uma distribuição mais homogênea deste elemento no perfil do solo.
- **CTC (Capacidade de Troca Catiônica):** O valor de $p = 0,0022$ indica diferença significativa, refletindo a maior capacidade de retenção de cátions na camada superficial, onde há maior teor de matéria orgânica e argila ativa.
- **Soma de Bases (S.B):** Com $p = 5,49 \times 10^{-05}$, este parâmetro apresentou diferença significativa, o que é coerente com a maior concentração dos cátions básicos (Ca, Mg e K) na camada superficial.

4.1.3.3 Nutrientes sem Diferença Significativa

Três variáveis não apresentaram diferenças estatisticamente significativas entre as profundidades:

- **pH:** Com $p = 0,9568$, o pH apresentou distribuição estatisticamente igual nas duas profundidades. Este resultado sugere que as práticas de calagem foram eficientes em promover a correção da acidez em profundidade, ou que o solo naturalmente apresenta pH relativamente uniforme no perfil.
- **Al:** O valor de $p = 1,0000$ indica que não há diferença entre as concentrações de alumínio nas duas profundidades. Como o alumínio está intimamente relacionado ao pH do solo e ambos não apresentaram diferenças significativas, este resultado é coerente e sugere condições de acidez similares nas duas camadas.
- **(Sat.):** Com $p = 1,0000$, a saturação por bases não diferiu entre as profundidades. Este resultado aparentemente contraditório — considerando que os nutrientes individuais

mostraram diferenças — pode ser explicado pela relação proporcional entre as bases e a CTC em ambas as camadas, mantendo índices de saturação similares mesmo com concentrações absolutas diferentes.

4.1.3.4 Implicações Agronômicas

Os resultados do teste estatístico confirmam que existe uma estratificação nutricional marcante no perfil do solo estudado, com exceções importantes para pH, alumínio e saturação por bases. A predominância de diferenças significativas reforça a necessidade de considerar a profundidade de amostragem e análise do solo como fator determinante nas recomendações de adubação e correção.

Além disso, é fundamental implementar estratégias de manejo que promovam a distribuição vertical mais uniforme de nutrientes, especialmente aqueles que apresentaram alta significância estatística nas diferenças, como P, K, Ca e S.

Práticas que favoreçam o desenvolvimento radicular em profundidade também são valorizadas, uma vez que a camada de 40 cm apresenta menores concentrações da maioria dos nutrientes, embora mantenha condições adequadas de pH e saturação.

Por fim, a homogeneidade do pH e da saturação por bases nas duas profundidades representa um aspecto positivo do manejo atual, indicando que a correção da acidez em profundidade tem sido eficiente.

Estes resultados estatísticos complementam as análises descritivas anteriores e fornecem base quantitativa para a tomada de decisões no manejo da fertilidade do solo nas áreas estudadas.

4.1.4 Teste de Homogeneidade de Variâncias

Seguindo os procedimentos metodológicos descritos no Capítulo 3, o teste de homogeneidade de variâncias foi aplicado para cada um dos 16 nutrientes analisados, com o objetivo de verificar se as variâncias dos dados eram homogêneas entre os grupos. O teste de Levene foi utilizado para avaliar as variâncias de cada nutriente individualmente.

Os resultados indicam que a maioria dos nutrientes apresentou variâncias homogêneas, com exceção de alguns nutrientes que apresentaram variâncias heterogêneas. O quadro 12 mostra o valor de p obtido para cada nutriente, bem como o resultado do teste de homogeneidade de variâncias.

4.1.4.1 Interpretação dos Resultados

- **Nutrientes com variâncias homogêneas:** Os nutrientes que apresentaram **p-valor maior que 0,05** indicam que suas variâncias são homogêneas, ou seja, a variabilidade

Quadro 12 – Resultados do teste de homogeneidade de variâncias para os nutrientes analisados.

Nutriente	P-Value	Resultado
<i>P</i>	0.4684	Homogênea
<i>M.O</i>	0.0485	Heterogênea
<i>pH</i>	0.9634	Homogênea
<i>K</i>	0.0516	Homogênea
<i>K/T</i>	0.3191	Homogênea
<i>Ca</i>	0.0088	Heterogênea
<i>Ca/T</i>	0.1379	Homogênea
<i>Mg</i>	0.0357	Heterogênea
<i>Mg/T</i>	0.6339	Homogênea
<i>Al</i>	1.0000	Homogênea
<i>H+Al</i>	0.4188	Homogênea
<i>S</i>	0.0000	Heterogênea
<i>S.B</i>	0.0019	Heterogênea
<i>CTC</i>	0.0167	Heterogênea
<i>V</i>	0.0195	Heterogênea
<i>Sat.</i>	1.0000	Homogênea

Fonte: Elaborado pelo autor.

dos dados é semelhante entre os grupos. Esses nutrientes incluem *P*, *pH*, *K*, *K/T*, *Ca/T*, *Al*, *H+Al*, *Mg/T* e *Sat.*.

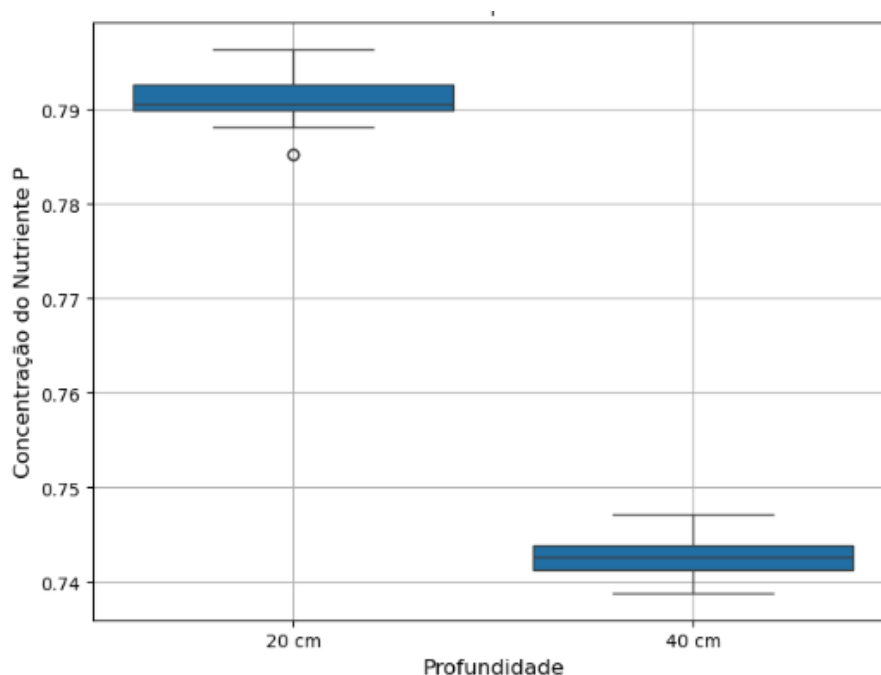
- **Nutrientes com variâncias heterogêneas:** Para os nutrientes com **p-valor menor ou igual a 0,05**, as variâncias são heterogêneas, ou seja, a dispersão dos dados entre os grupos é significativamente diferente. Esses nutrientes incluem *M.O*, *Ca*, *Mg*, *S*, *S.B*, *CTC* e *V*.

Esses resultados são importantes porque indicam que, ao escolher os testes estatísticos subsequentes, devemos considerar a homogeneidade das variâncias. Para os nutrientes com variâncias heterogêneas, foi necessário utilizar testes estatísticos que não assumem homogeneidade de variâncias, como a análise de variância não paramétrica, para garantir a validade dos testes.

4.1.4.2 Visualizações dos Dados

A seguir, são apresentadas representações gráficas dos box-plots para ilustrar as distribuições dos nutrientes analisados. As figuras 28 e 29 apresentam os boxplots de alguns dos nutrientes.

Figura 28 – Box-plot para o nutriente P (Variância Homogênea).



Fonte: Elaborada pelo autor.

A figura 28 apresenta o boxplot do nutriente P, classificado como possuindo variância homogênea segundo o teste de Levene. Visualmente, observa-se uma distribuição mais estável entre os grupos, com baixa dispersão relativa e mínima presença de outliers, o que reforça o resultado estatístico. A estrutura do boxplot é compacta, com quartis bem definidos e sem alongamentos extremos, indicando que os valores do nutriente apresentam comportamento consistente e variação semelhante entre os grupos analisados.

4.1.5.1 Correlação entre Nutrientes a 20 cm de Profundidade

Na profundidade de 20 cm, os resultados mostram algumas correlações notáveis entre os nutrientes. A correlação mais forte encontrada foi entre **Ca/T** e **Mg/T (0.91)**, o que sugere que esses dois nutrientes estão fortemente relacionados, possivelmente devido à presença de cátions trocáveis que influenciam a dinâmica do solo. Além disso, **Ca/T** também apresentou correlações moderadas com **K (0.27)** e **CTC (0.86)**, sugerindo que o cálcio trocável está associado a outros nutrientes e à capacidade de troca catiônica do solo.

Outros pares de nutrientes como **K/T** e **pH (-0.60)** indicam uma relação inversa, com a acidez do solo tendo uma correlação negativa com a relação **K/T**, o que pode ser explicado pelo efeito do pH sobre a solubilidade dos cátions no solo. Também se observou uma correlação negativa entre **S** e **H+Al (-0.68)**, indicando que o aumento da acidez do solo pode reduzir a disponibilidade de enxofre, um nutriente essencial para as plantas.

A correlação entre **S.B** e **CTC (0.16)** é baixa, o que sugere que, apesar de ambos estarem relacionados com a capacidade de troca catiônica, a relação entre eles pode ser mais complexa ou influenciada por outros fatores não capturados nesta análise.

O quadro 13 apresenta as correlações entre os nutrientes na profundidade de 20 cm.

Quadro 13 – Correlação entre os nutrientes a 20cm de profundidade.

Nutriente	P	M.O	pH	K	K/T	Ca	Ca/T	Mg	Mg/T	Al	H+Al	S	S.B	CTC	V	Sat.
P	1.00	0.38	0.25	0.37	0.10	0.095	0.38	-0.03	0.39	0.00	0.14	0.00	-0.03	0.47	-0.19	0.00
M.O	0.38	1.00	-0.07	0.36	0.35	0.12	0.21	0.19	0.37	0.00	0.16	0.00	0.19	0.16	-0.34	0.00
pH	0.25	-0.07	1.00	-0.25	-0.60	0.16	0.58	0.21	0.50	0.00	-0.68	0.00	0.21	0.46	0.09	0.00
K	0.37	0.36	-0.25	1.00	0.78	0.00	0.27	-0.04	0.31	0.00	0.51	0.00	-0.04	0.43	-0.11	0.00
K/T	0.10	0.35	-0.60	0.78	1.00	0.04	-0.19	-0.15	-0.09	0.00	0.66	0.00	-0.15	-0.15	-0.16	0.00
Ca	0.09	0.12	0.16	0.00	0.04	1.00	0.04	0.18	0.01	0.00	0.03	0.00	0.18	0.02	-0.02	0.00
Ca/T	0.38	0.21	0.58	0.27	-0.19	0.04	1.00	0.22	0.91	0.00	-0.22	0.00	0.22	0.86	-0.05	0.00
Mg	-0.02	0.19	0.21	-0.04	-0.15	0.18	0.22	1.00	0.26	0.00	-0.12	0.00	1.0	0.16	-0.03	0.00
Mg/T	0.39	0.37	0.50	0.31	-0.09	0.01	0.91	0.26	1.00	0.00	-0.23	0.00	0.26	0.79	-0.17	0.00
Al	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
H+Al	0.14	0.16	-0.68	0.51	0.66	0.03	-0.22	-0.12	-0.23	0.00	1.00	0.00	-0.12	0.00	-0.08	0.00
S	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
S.B	-0.02	0.19	0.21	-0.04	-0.15	0.18	0.22	1.00	0.26	0.00	-0.12	0.00	1.00	0.16	0.04	0.00
CTC	0.47	0.16	0.46	0.43	-0.15	0.02	0.86	0.16	0.79	0.00	0.00	0.00	0.16	1.00	0.00	0.00
V	-0.19	-0.34	0.09	-0.11	-0.16	-0.02	-0.05	0.03	-0.17	0.00	-0.08	0.00	0.04	0.00	1.00	0.00
Sat.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Fonte: Elaborado pelo autor.

4.1.5.2 Correlação entre Nutrientes a 40 cm de Profundidade

Na profundidade de 40 cm, os resultados mostraram padrões de correlação interessantes. A correlação mais forte observada foi entre **Ca/T** e **Mg/T (0.85)**, refletindo novamente a estreita interação entre esses dois nutrientes. A correlação entre **K/T** e **pH (-0.52)** também foi notável, sugerindo que o pH do solo está negativamente relacionado com a relação entre **K** e **T**, influenciando a solubilidade de potássio. Observa-se também uma correlação moderada entre **S** e **H+Al (0.25)**, indicando que o aumento da acidez do solo pode promover a mobilização do enxofre.

O quadro 14 apresenta as correlações entre os nutrientes na profundidade de 40 cm.

Quadro 14 – Correlação entre os nutrientes a 40cm de profundidade.

Nutriente	P	M.O	pH	K	K/T	Ca	Ca/T	Mg	Mg/T	Al	H+Al	S	S.B	CTC	V	Sat.
P	1.00	0.45	0.39	0.33	0.02	0.16	0.61	0.16	0.62	0.00	0.17	-0.27	-0.05	0.20	0.42	0.00
M.O	0.45	1.00	0.16	0.24	0.21	0.22	0.37	0.22	0.44	0.00	-0.03	-0.21	-0.11	0.21	0.20	0.00
pH	0.39	0.16	1.00	-0.19	-0.52	0.09	0.42	0.09	0.56	0.00	-0.57	-0.03	0.25	0.05	0.79	0.00
K	0.33	0.24	-0.19	1.00	0.76	0.00	0.39	0.00	0.24	0.00	0.62	0.05	-0.25	0.00	-0.05	0.00
K/T	0.02	0.21	-0.52	0.76	1.00	0.04	0.12	0.04	0.02	0.00	0.64	-0.05	-0.14	0.03	-0.57	0.00
Ca	0.16	0.22	0.09	0.00	0.04	1.00	0.16	1.0	0.14	0.00	0.05	-0.15	0.14	0.13	0.00	0.00
Ca/T	0.61	0.37	0.42	0.39	0.12	0.16	1.00	0.16	0.85	0.00	0.11	0.05	0.16	0.00	0.56	0.00
Mg	0.16	0.22	0.09	0.00	0.04	1.0	0.16	1.00	0.14	0.00	0.05	-0.15	0.14	0.13	0.00	0.00
Mg/T	0.62	0.44	0.56	0.24	0.02	0.14	0.85	0.14	1.00	0.00	-0.09	-0.16	0.1	0.05	0.67	0.00
Al	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
H+Al	0.17	-0.03	-0.57	0.62	0.64	0.05	0.11	0.05	-0.09	0.00	1.00	0.02	-0.15	-0.04	-0.55	0.00
S	-0.27	-0.21	-0.03	0.05	-0.05	-0.15	0.05	-0.15	-0.16	0.00	0.02	1.00	-0.07	-0.32	0.05	0.00
S.B	-0.05	-0.11	0.25	-0.25	-0.14	0.14	0.16	0.14	0.1	0.00	-0.15	-0.07	1.00	0.04	0.05	0.00
CTC	0.20	0.21	0.05	0.00	0.03	0.13	0.00	0.13	0.05	0.00	-0.04	-0.32	0.04	1.00	0.02	0.00
V	0.42	0.20	0.79	-0.05	-0.57	0.00	0.56	0.00	0.67	0.00	-0.55	0.05	0.05	0.02	1.00	0.00
Sat.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Fonte: Elaborado pelo autor.

4.1.5.3 Análise Conjunta das Correlações

Ao comparar as correlações observadas nas profundidades de 20 cm e 40 cm, destacam-se algumas diferenças importantes. Na profundidade de 20 cm, a correlação mais forte foi observada entre **Ca/T** e **Mg/T** (**0.91**), indicando uma interação estreita entre esses dois nutrientes, possivelmente devido à presença de cátions trocáveis no solo que afetam tanto o cálcio quanto o magnésio. Em contrapartida, na profundidade de 40 cm, a correlação mais forte foi entre **Ca/T** e **Mg/T** (**0.85**), mas a maior interação foi observada entre **K/T** e **pH** (**-0.52**), sugerindo uma relação inversa entre a acidez do solo e a relação entre potássio e cátions trocáveis.

Essa diferença nas correlações pode ser atribuída à variação nas propriedades físicas e químicas do solo nas diferentes profundidades. O efeito do pH sobre a solubilidade dos nutrientes pode ser mais pronunciado em profundidades maiores, onde a mobilidade de cátions e a disponibilidade de nutrientes podem ser mais afetadas pela acidez do solo.

Essas correlações revelam como os nutrientes interagem de maneira complexa nas diferentes camadas do solo. Por exemplo, enquanto na camada de 20 cm, a interação entre **Ca/T** e **Mg/T** é mais forte, na camada de 40 cm, a relação entre o pH e a disponibilidade de potássio ganha destaque, o que sugere que a gestão de nutrientes deve considerar as variações de profundidade para otimizar o uso e a disponibilidade dos nutrientes.

4.1.5.4 Implicações para a Fertilidade do Solo

As correlações observadas nas profundidades de 20 cm e 40 cm oferecem insights valiosos sobre como a interação entre os nutrientes influencia a fertilidade do solo. A alta correlação entre **Ca/T** e **Mg/T** em ambas as profundidades sugere que esses dois nutrientes

estão intimamente relacionados e que a gestão de um pode impactar a disponibilidade do outro. Portanto, estratégias de fertilização e manejo de solo devem considerar o equilíbrio entre cálcio e magnésio para garantir a adequada disponibilidade desses nutrientes para as plantas.

Além disso, as correlações negativas entre **pH** e **K/T**, tanto na profundidade de 20 cm quanto na de 40 cm, indicam que ajustes no pH do solo têm um impacto significativo na disponibilidade de potássio. O pH influencia diretamente a solubilidade dos cátions no solo, e a correlação negativa sugere que, em solos mais ácidos, a disponibilidade de potássio pode ser limitada.

Da mesma forma, a correlação entre **S** e **H+Al** (**-0.68** a 20 cm e **0.25** a 40 cm) indica que a acidez do solo pode afetar a disponibilidade de enxofre, um nutriente essencial para as plantas. Isso reforça a necessidade de monitorar e ajustar os níveis de acidez do solo para otimizar a disponibilidade de nutrientes essenciais.

4.1.5.5 Conclusão

As análises de correlação realizadas nas profundidades de 20 cm e 40 cm revelaram padrões consistentes de interação entre os nutrientes, com destaque para a forte relação entre **Ca/T** e **Mg/T** e a interação do pH com **K/T**. A compreensão dessas correlações é essencial para a formulação de estratégias de manejo do solo e da fertilização. Aprofundar-se nas interações entre os nutrientes permite otimizar o uso dos recursos disponíveis, promover a saúde do solo e garantir um ambiente favorável para o crescimento das plantas. Com base nas correlações observadas, podemos concluir que a gestão da acidez do solo, juntamente com o controle da disponibilidade de cátions trocáveis, é crucial para manter a fertilidade do solo e maximizar o rendimento das culturas.

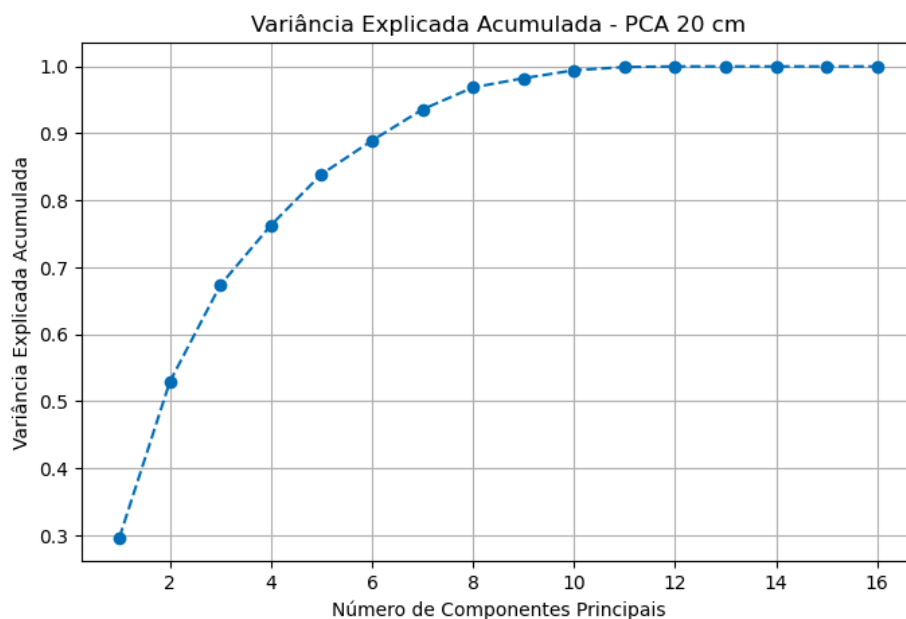
4.1.6 Análise de Componentes Principais

Seguindo os procedimentos metodológicos descritos no Capítulo 3, o PCA foi aplicado para identificar padrões e reduzir a dimensionalidade dos dados dos nutrientes no solo, de forma a entender as relações entre as variáveis e identificar os principais fatores responsáveis pela variação nos dados. O PCA foi realizado para duas profundidades de amostragem: 20 cm e 40 cm, sendo as variáveis analisadas os nutrientes do solo.

4.1.6.1 Variância Explicada Acumulada - PCA 20 cm

A primeira análise de PCA foi realizada para os dados coletados a uma profundidade de 20 cm, e a variância explicada acumulada foi representada na figura 30. Observa-se que os primeiros componentes principais explicam uma grande parte da variação dos dados. Em torno de 2 a 3 componentes principais são suficientes para explicar mais de 90% da variância, o que indica que os dados podem ser bem representados com um número reduzido de componentes.

Figura 30 – Variância Explicada Acumulada - PCA 20 cm



Fonte: Elaborada pelo autor.

O PCA revelou as contribuições dos principais nutrientes para os componentes principais (PCs) nos dados de 20 cm de profundidade. Abaixo estão os nutrientes mais influentes para cada componente principal identificado:

- **PC1** (Acidez): Ca/T (0.46), Mg/T (0.46), CTC (0.42), pH (0.36) e P (0.25) apresentaram as maiores influências, sendo relacionados com a eficiência do cálcio e magnésio, capacidade de troca catiônica e acidez do solo.
- **PC2** (Acidez potencial): K (0.50), K/T (0.49), H:Al (0.43), Ca/T (-0.24) e pH (-0.29) apresentaram contribuições consideráveis, com destaque para a fertilidade básica e acidez potencial.
- **PC3** (Fertilidade básica): S.B. (0.62), Mg (0.59), Ca (0.59) e S (0.44) estiveram entre os mais influentes, com destaque para a fertilidade básica e os fatores que afetam a troca catiônica do solo.

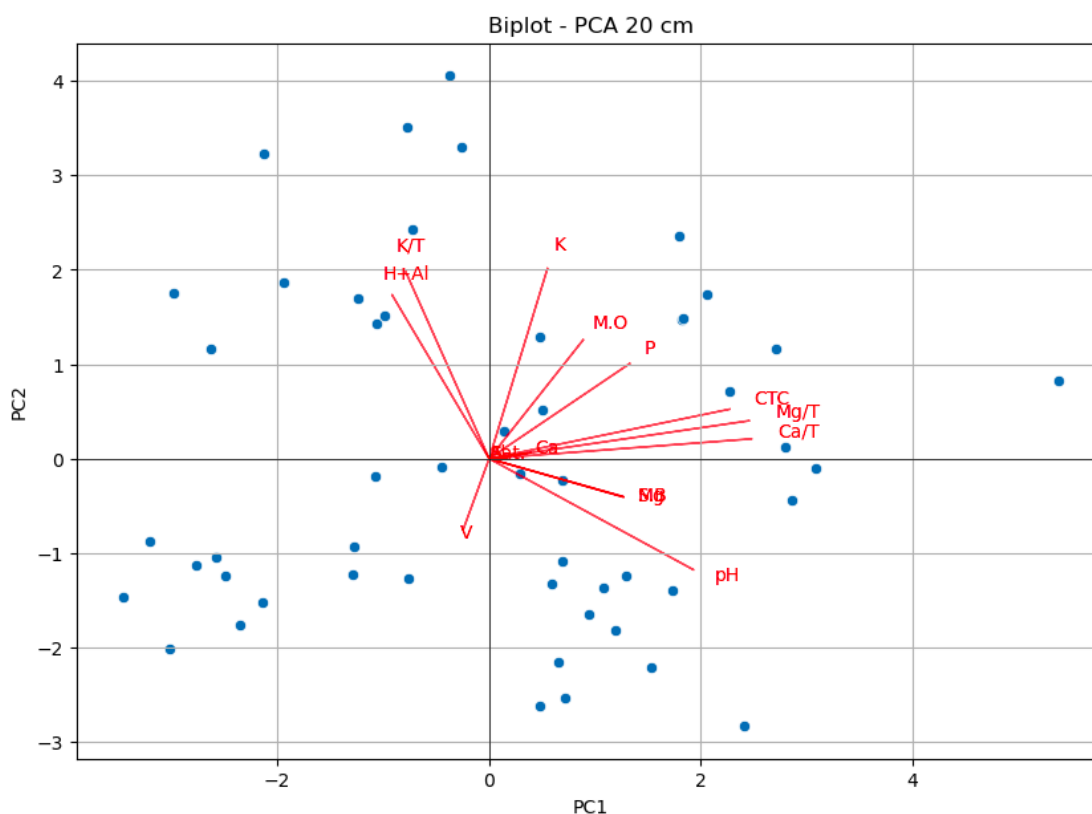
A interpretação geral dos componentes revelou que o primeiro componente (PC1) está mais relacionado com a acidez do solo, enquanto os componentes seguintes (PC2 e PC3) refletem principalmente a fertilidade básica e a capacidade de troca catiônica.

4.1.6.2 Biplot - PCA 20 cm

O biplot da PCA para os dados de 20 cm (figura 31) mostra a projeção dos dados nos dois primeiros componentes principais (PC1 e PC2). As setas representam os nutrientes

que mais influenciam cada componente. Nutrientes como Ca/T, Mg/T, K/T e H:Al possuem grandes influências, apontando para uma forte correlação entre fertilidade básica e a acidez do solo. Os pontos azuis no gráfico representam as amostras do solo, enquanto as setas vermelhas indicam as variáveis mais influentes. Essa visualização permite compreender melhor como os nutrientes interagem e como os diferentes solos se agrupam em relação aos fatores analisados.

Figura 31 – Biplot - PCA 20 cm

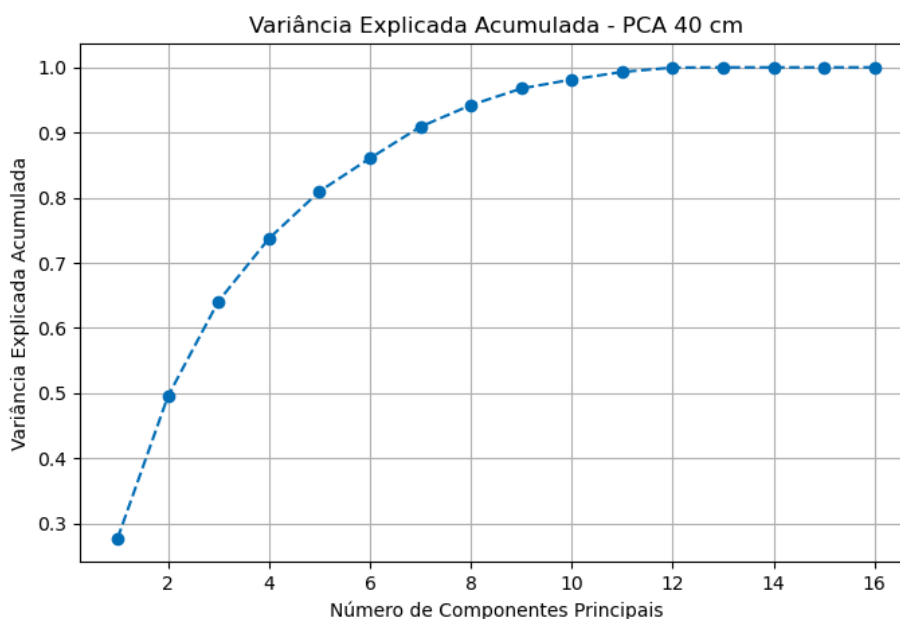


Fonte: Elaborada pelo autor.

4.1.6.3 Variância Explicada Acumulada - PCA 40 cm

A análise de PCA para a profundidade de 40 cm (figura 32) apresentou um padrão semelhante ao PCA de 20 cm, com uma explicação acumulada de variância superior a 90% após a incorporação dos primeiros três componentes principais. A principal diferença observada é que a profundidade de 40 cm apresentou um pouco mais de variação explicada pelos primeiros componentes, o que pode indicar que os nutrientes a essa profundidade têm um papel mais uniforme e são mais interdependentes.

Figura 32 – Variância Explicada Acumulada - PCA 40 cm



Fonte: Elaborada pelo autor.

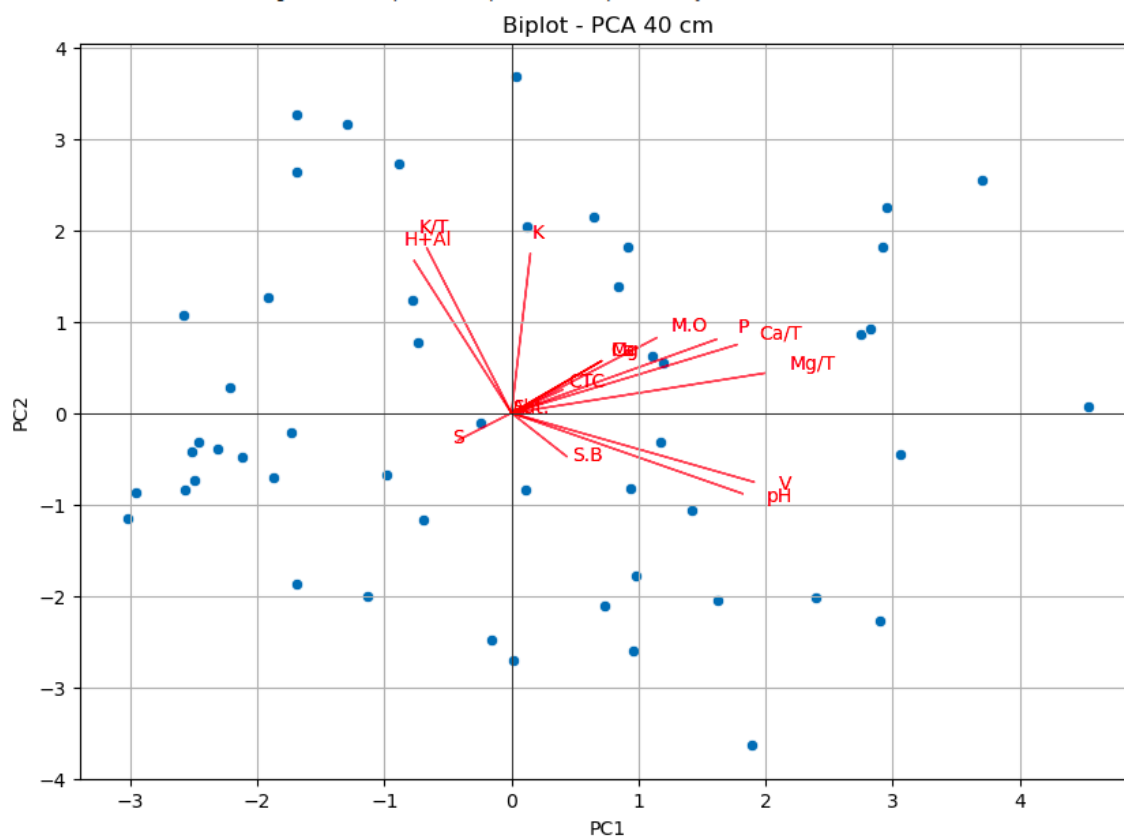
Para os dados de 40 cm de profundidade, as contribuições dos nutrientes para os componentes principais foram as seguintes:

- **PC1** (Acidez): Mg/T (0.44), V (0.42), Mg (0.40) e Ca/T (0.39) foram os nutrientes com maior contribuição para o primeiro componente, novamente associados à eficiência do magnésio e ao controle da acidez do solo.
- **PC2** (Acidez potencial): K/T (0.49), K (0.47), H:Al (-0.45) e M.O. (0.22) mostraram ser os principais influentes, refletindo a fertilidade básica e a acidez potencial.
- **PC3** (Fertilidade básica): Ca (0.59), Mg (0.56) e S (0.44) novamente se destacaram, como foi observado nos dados de 20 cm.

4.1.6.4 Biplot - PCA 40 cm

O biplot da PCA para os dados de 40 cm (figura 33) apresentou uma configuração semelhante ao de 20 cm, com as setas representando os nutrientes que mais influenciam os componentes principais. As amostras de solo, representadas pelos pontos azuis, estão distribuídas de forma similar às aquelas observadas na profundidade de 20 cm, com uma predominância dos fatores relacionados à fertilidade básica e acidez do solo.

Figura 33 – Biplot - PCA 40 cm



Fonte: Elaborada pelo autor.

4.1.6.5 Interpretação Geral da PCA

A análise geral dos componentes principais indica que a estrutura do solo requer um número maior de componentes principais para ser bem representada. Isso sugere que a variabilidade dos nutrientes nos dois perfis de solo (20 cm e 40 cm) é complexa e não pode ser completamente explicada por apenas dois ou três componentes. A presença de múltiplos nutrientes inter-relacionados e sua influência sobre as propriedades do solo, como a acidez e a fertilidade básica, reforça a necessidade de considerar um conjunto maior de componentes para obter uma análise precisa.

4.1.6.6 Conclusões

A PCA revelou que a acidez do solo e a fertilidade básica são as variáveis dominantes que explicam a maior parte da variância dos dados de nutrientes, tanto para a profundidade de 20 cm quanto para a profundidade de 40 cm. O modelo de PCA permitiu identificar os nutrientes mais influentes para cada componente e forneceu uma visão clara da interação entre as variáveis, facilitando a compreensão dos fatores que afetam a qualidade do solo. A análise sugere que o solo deve ser monitorado em diferentes profundidades, levando em conta

os componentes principais identificados, para uma gestão eficiente dos nutrientes.

4.1.7 Análise de Clusterização

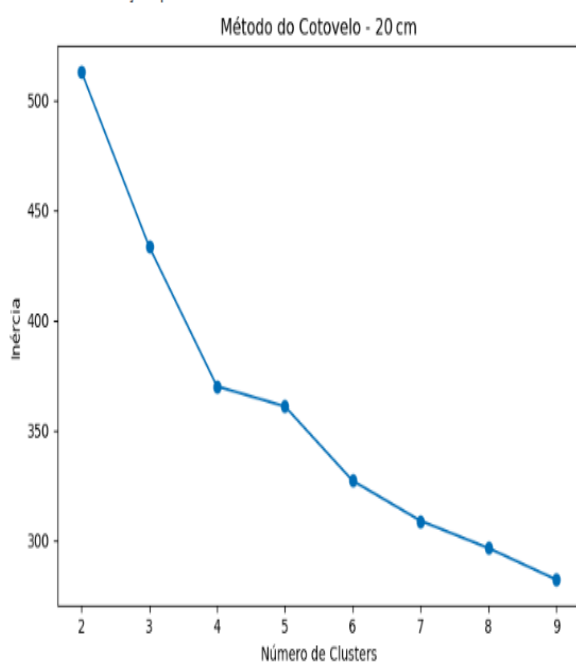
Seguindo os procedimentos metodológicos descritos no Capítulo 3, nesta seção apresentamos a análise de clusterização dos dados utilizando os métodos do Cotovelo e do Silhouette Score. A análise foi realizada para dois conjuntos de dados com diferentes profundidades: 20 cm e 40 cm. A seguir, discutimos os resultados obtidos a partir de cada uma das técnicas utilizadas para determinar o número ótimo de *clusters*, buscando uma análise mais precisa e robusta.

4.1.7.1 Método do Cotovelo

O gráfico do Método do Cotovelo apresenta a inércia para diferentes números de *clusters*. Observa-se que a inércia diminui de forma acentuada até o ponto de 4 *clusters*, após o qual a redução torna-se menos significativa. Esse comportamento sugere que o número ideal de *clusters* para os dados de 20 cm seria entre 3 e 4. O mesmo padrão é observado nos dados de 40 cm, onde a inércia também diminui de forma marcante até 4 *clusters*, estabilizando após esse ponto.

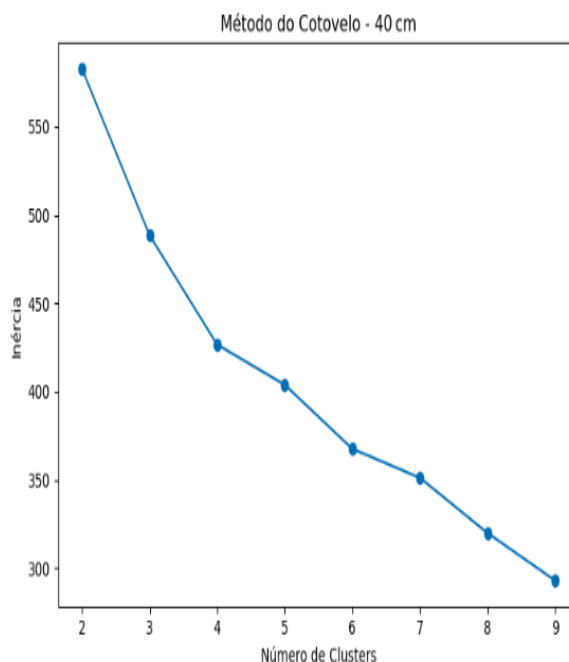
A figura 34 ilustra a análise para o conjunto de dados de 20 cm, enquanto a figura 35 apresenta os resultados para o conjunto de dados de 40 cm.

Figura 34 – Método do Cotovelo - 20 cm



Fonte: Elaborada pelo autor.

Figura 35 – Método do Cotovelo - 40 cm



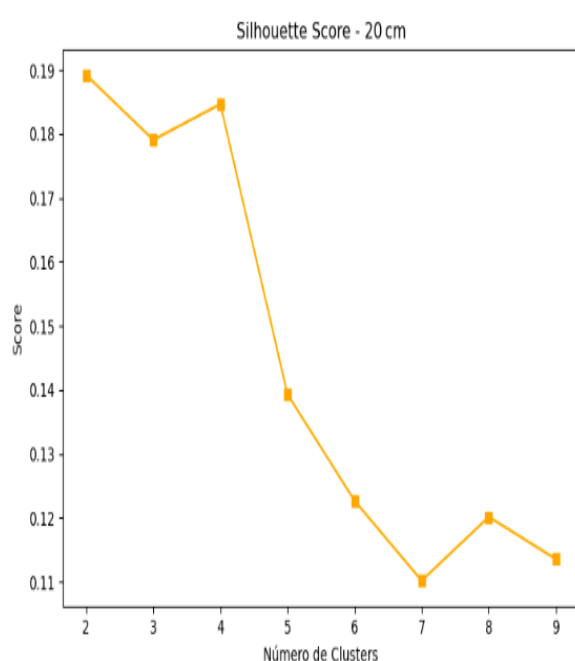
Fonte: Elaborada pelo autor.

4.1.7.2 Silhouette Score

O gráfico do Silhouette Score foi utilizado para avaliar a qualidade da separação entre os *clusters* formados. Para os dados de 20 cm, o maior valor de Silhouette Score ocorre com 2 *clusters*, mas esse valor diminui com o aumento do número de *clusters*, indicando que a coesão e a separação não são significativamente melhoradas ao adicionar mais *clusters*. Para os dados de 40 cm, a maior pontuação é observada com 4 *clusters*, mas, como no conjunto de dados de 20 cm, os valores diminuem ao se aumentar o número de *clusters* além desse ponto.

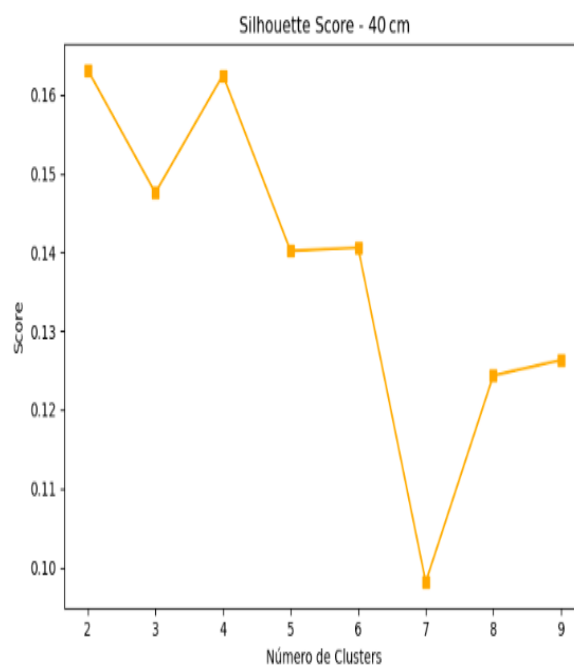
A figura 36 apresenta o Silhouette Score para o conjunto de dados de 20 cm, enquanto a figura 37 exibe os resultados para o conjunto de dados de 40 cm.

Figura 36 – Silhouette Score - 20 cm



Fonte: Elaborada pelo autor.

Figura 37 – Silhouette Score - 40 cm



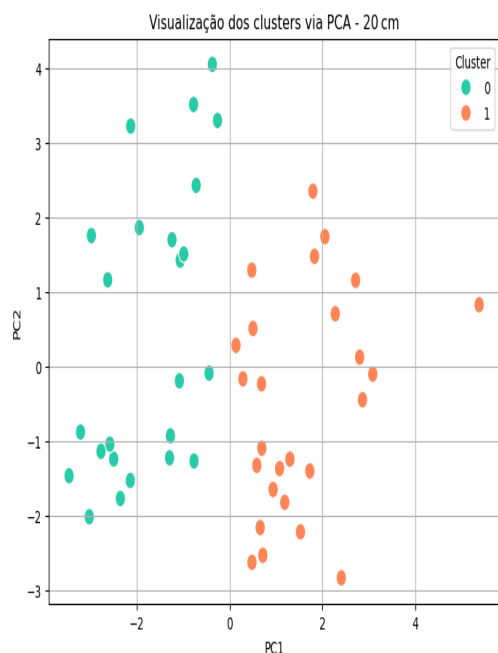
Fonte: Elaborada pelo autor.

4.1.7.3 Visualização dos *Clusters* via PCA

A visualização dos *clusters* foi realizada por meio da análise de componentes principais (PCA), que possibilitou observar a separação dos *clusters* em duas dimensões. Para os dados de 20 cm, a separação entre os *clusters* é claramente visível, com dois grupos bem definidos. Da mesma forma, para os dados de 40 cm, a visualização mostra uma boa separação entre os *clusters*, confirmando a eficiência do processo de clusterização.

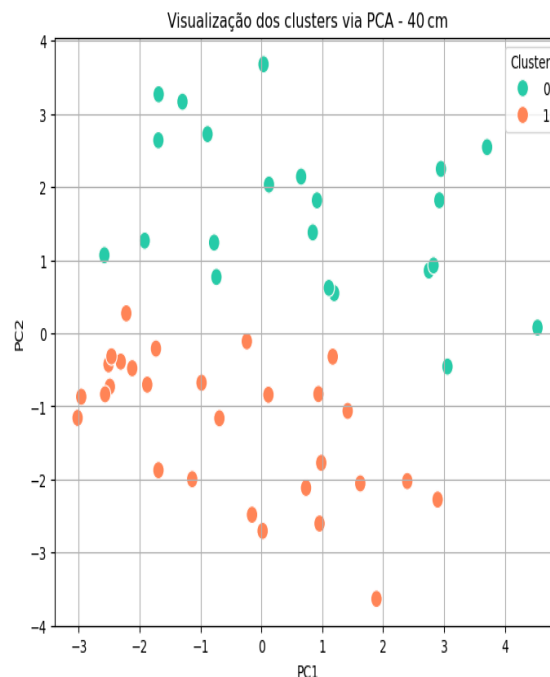
A figura 38 apresenta a visualização para o conjunto de dados de 20 cm, enquanto a figura 39 mostra a visualização dos *clusters* para o conjunto de dados de 40 cm.

Figura 38 – Visualização dos *clusters* via PCA
- 20 cm



Fonte: Elaborada pelo autor.

Figura 39 – Visualização dos *clusters* via PCA
- 40 cm



Fonte: Elaborada pelo autor.

4.1.7.4 Conclusões

Com base nas análises realizadas, podemos concluir que o número ideal de *clusters* para ambos os conjuntos de dados, 20 cm e 40 cm, está entre 3 e 4 *clusters*. A análise do Método do Cotovelo e o Silhouette Score mostram que a adição de mais *clusters* após esse ponto não resulta em melhorias substanciais na qualidade da clusterização. A visualização dos *clusters* via PCA confirma a separação clara entre os grupos, indicando que a técnica de clusterização foi bem-sucedida.

Em resumo, a utilização combinada dos métodos do Cotovelo, do Silhouette Score e da PCA proporcionou uma análise robusta e eficiente, com bons resultados em ambos os conjuntos de dados.

4.1.8 Análise dos Perfis de Solo nas Quadras dos Lotes

A análise dos perfis de solo foi realizada em duas profundidades, 20 cm e 40 cm, com o objetivo de identificar padrões de concentração de tipos de solo por quadra. Este estudo visa compreender a variação do solo em diferentes áreas do terreno, levando em consideração a distribuição dos *clusters* de solo por profundidade e quadra, bem como a predominância de cada tipo de solo em cada lote.

4.1.8.1 Distribuição dos Perfis de Solo a 20 cm de Profundidade

O quadro de distribuição dos perfis de solo a 20 cm revela que, em diversas quadras, o *cluster* predominante é o tipo de solo 1 (indicado pela coluna de valores em 1.0). No lote 5, por exemplo, as quadras Q 01 e Q 02 apresentam 2 amostras, com o *cluster* predominante sendo o tipo 0, sugerindo uma maior variação do solo nesta região. Padrões semelhantes podem ser observados em outros lotes, como os lotes 6 e 7, onde a distribuição das amostras e *clusters* é semelhante, com 2 amostras totais e 0 como *cluster* predominante.

Por outro lado, em quadras como a **Q Expansão** (lote 14), o *cluster* predominante é 1.0, com 2 amostras, indicando uma área com solo mais homogêneo. A presença de *clusters* específicos em outras quadras, como em **Q 01, Q 02 e Q 03**, também demonstra um solo mais uniforme, com o cluster predominante sendo o tipo 1.0, reforçando a estabilidade do solo em algumas áreas.

4.1.8.2 Distribuição dos Perfis de Solo a 40 cm de Profundidade

O quadro de 40 cm revela uma tendência semelhante, com o *cluster* predominante sendo o tipo 1.0 nas quadras. Nos lotes 5, 6 e 7, observamos que a distribuição de solo continua a ser majoritariamente do tipo 1.0, com a predominância deste tipo em quadras como Q 01 e Q 02, onde o número de amostras varia entre 2 e 3 amostras totais.

Além disso, as quadras localizadas em áreas mais amplas, como a **Q Expansão** (lote 14) e **Q 09, Q 10 e Q 11** (lote 13), mantêm uma consistência na distribuição das amostras, com a predominância do tipo de solo 1.0, sugerindo uma maior homogeneidade do solo nessas regiões. No entanto, quadras como **Q 01, Q 02, e Q 03**, que apresentaram uma maior diversidade a 20 cm, continuam a mostrar o mesmo padrão de distribuição, embora com uma leve mudança na predominância do tipo de solo. Esse padrão pode indicar uma transição gradual na composição do solo ao longo da profundidade, com um aumento da homogeneidade à medida que se observa a maior profundidade.

4.1.8.3 Interpretação e Implicações

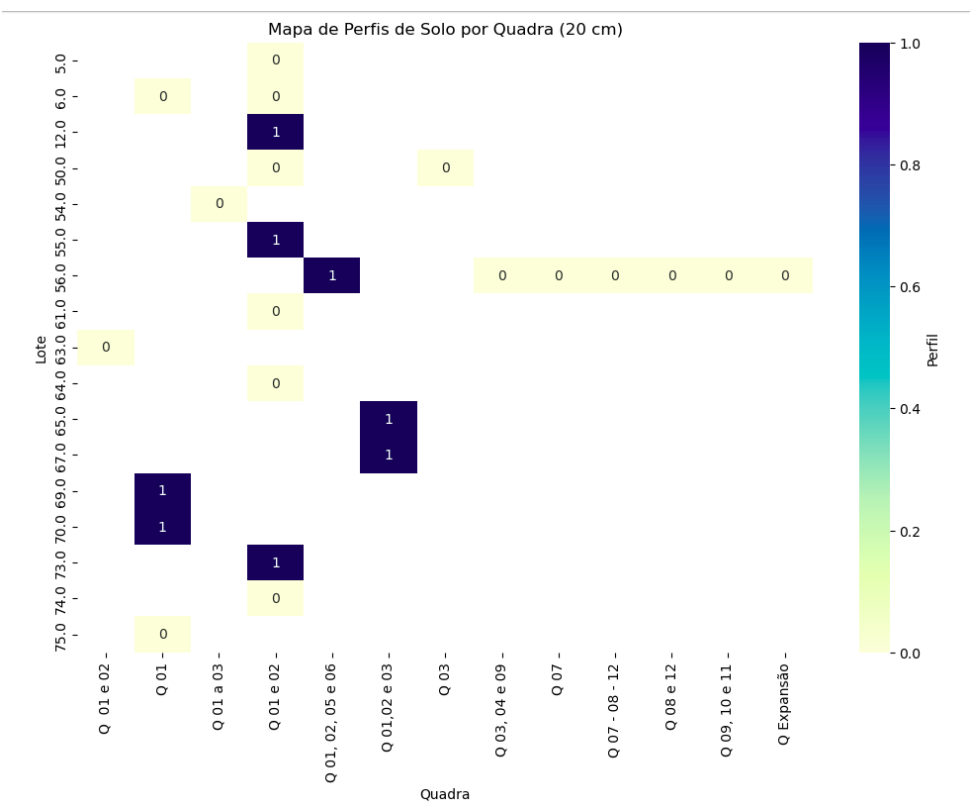
- **Cluster Predominante:** Em várias quadras, o tipo de solo predominante é o 1.0, o que sugere uma presença de solo mais uniforme. Essa homogeneidade pode indicar áreas mais estáveis, que são menos suscetíveis a variações climáticas ou mudanças no uso do solo, tanto a 20 cm quanto a 40 cm de profundidade.
- **Variações entre Quadras:** Apesar da predominância do *cluster* 1.0, algumas quadras, como a **Q 01, Q 02 e Q 03**, exibem maior diversidade nos tipos de solo, especialmente nas profundidades mais rasas (20 cm). Isso pode refletir variações na composição do solo, sendo um fator relevante para a análise do comportamento do solo em diferentes profundidades.

- **Número de Amostras:** A quantidade de amostras por quadra também é um fator importante. Quadras com mais amostras, como a **Q 01 e Q 02**, apresentam uma maior confiabilidade nas conclusões sobre o tipo de solo. Em contraste, quadras com menos amostras podem ser mais suscetíveis a erros ou variações nos resultados.

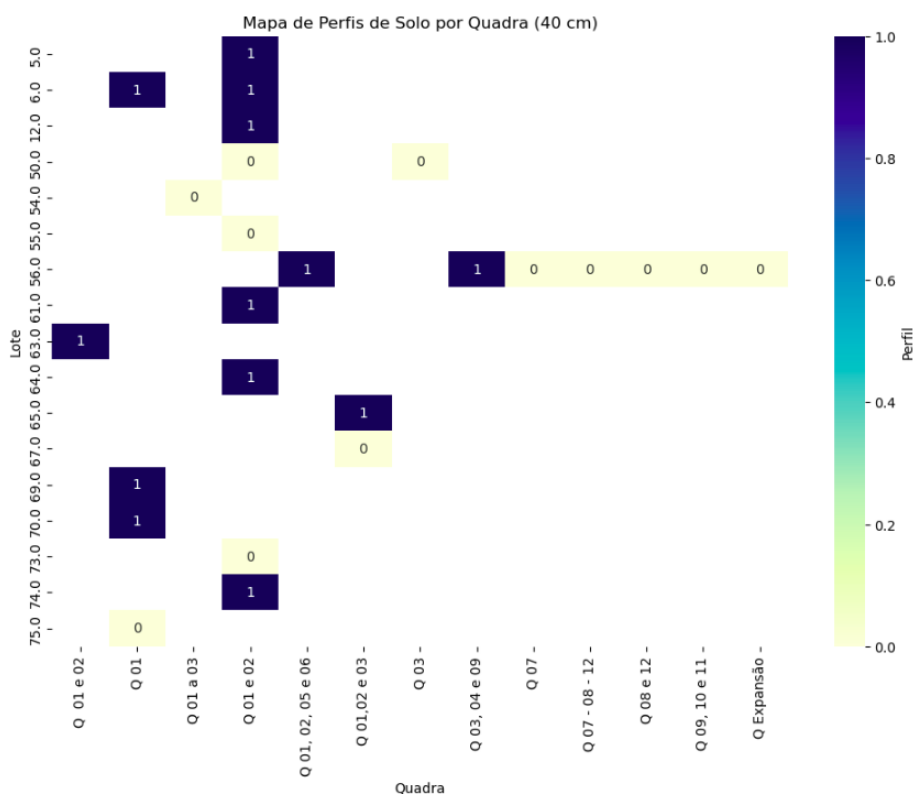
4.1.8.4 Visualização Gráfica

Para facilitar a compreensão dos dados, foram gerados gráficos ilustrando a distribuição dos perfis de solo por quadra, conforme os quadros descritos. A seguir, são apresentados dois gráficos, um para a profundidade de 20 cm (figura 40) e outro para 40 cm (figura 41), que permitem uma visualização clara dos padrões de distribuição dos tipos de solo.

Figura 40 – Distribuição do *Cluster* a 20 cm por Quadra



Fonte: Elaborada pelo autor.

Figura 41 – Distribuição do *Cluster* a 40 cm por Quadra

Fonte: Elaborada pelo autor.

Esses gráficos ajudam a visualizar de forma clara as áreas de maior variação no tipo de solo, bem como as áreas mais homogêneas, permitindo uma análise mais profunda sobre a variação do solo e suas implicações para a caracterização do terreno para o desenvolvimento dos lotes.

4.1.9 Análise de Sensibilidade Nutricional

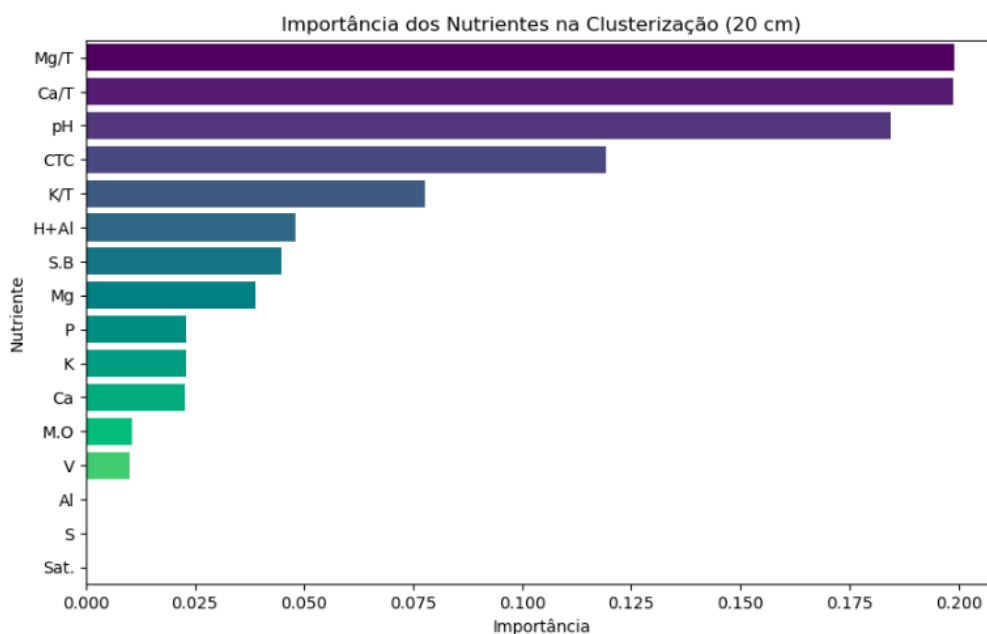
Nesta seção, são apresentados os resultados da análise de sensibilidade nutricional dos nutrientes com base nos dados de clusterização, para as profundidades de 20 cm e 40 cm. Os gráficos a seguir ilustram a importância relativa dos nutrientes para a formação de *clusters* em ambas as profundidades.

4.1.9.1 Importância dos Nutrientes na Clusterização a 20 cm

O gráfico a seguir (figura 42) mostra a importância dos nutrientes na clusterização do solo a 20 cm de profundidade. Como observado, o nutriente **Mg/T** apresenta a maior importância, com um valor significativo em torno de 0.200. Esse dado sugere que a relação magnésio/cálcio e o pH do solo são fundamentais para a formação de *clusters* na camada superficial. Em seguida, **Ca/T** e **pH** também se destacam, com valores próximos a 0.175.

Nutrientes como **CTC** (Capacidade de Troca Catiônica) e **K/T** também desempenham papéis importantes, com valores ao redor de 0.125, enquanto outros nutrientes, como **H+Al** e **S.B**, têm uma importância menor.

Figura 42 – Importância dos Nutrientes na Clusterização (20 cm)

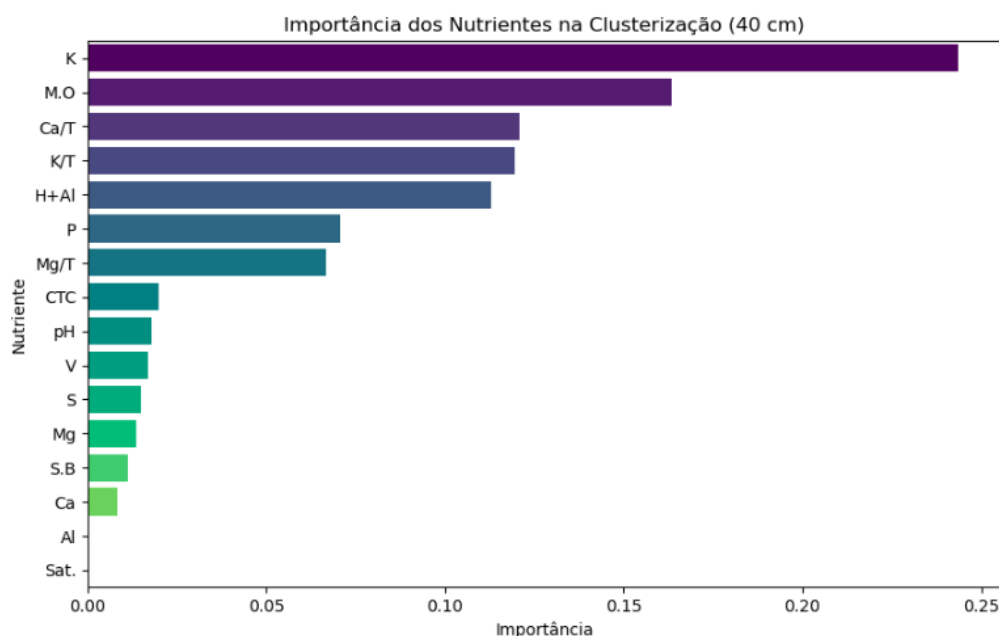


Fonte: Elaborada pelo autor.

4.1.9.2 Importância dos Nutrientes na Clusterização a 40 cm

O gráfico a seguir (figura 43) mostra a importância dos nutrientes na clusterização do solo a 40 cm de profundidade. Ao contrário da profundidade de 20 cm, o nutriente **K** assume uma importância ainda mais expressiva, com um valor superior a 0.25. Isso sugere que o potássio tem um papel crucial na dinâmica de trocas iônicas e no crescimento das plantas, especialmente nas camadas mais profundas do solo. A **M.O.** e o **Ca/T** também têm uma importância significativa, com valores superiores a 0.10. Já o **K/T** e o **H+Al** mantêm sua relevância, enquanto o impacto do **Mg/T** diminui em relação à camada superficial. Nutrientes como **P** e **CTC** continuam a ser importantes, mas com um impacto relativamente menor.

Figura 43 – Importância dos Nutrientes na Clusterização (40 cm)



Fonte: Elaborada pelo autor.

4.1.9.3 Conclusões

A análise revelou que a importância dos nutrientes varia com a profundidade do solo. Para a profundidade de 20 cm, a relação magnésio/cálcio e o pH do solo são os fatores mais determinantes. Já para os 40 cm, o potássio e a matéria orgânica ganham maior relevância. Esses resultados destacam as diferenças nas dinâmicas de nutrientes em camadas diferentes do solo, o que deve ser considerado no manejo nutricional e na formulação de estratégias de fertilização adequadas às necessidades das plantas em cada profundidade do solo.

Esses achados são cruciais para otimizar práticas agrícolas, como a aplicação de fertilizantes, e podem ser usados para ajustar recomendações específicas de manejo nutricional.

4.1.10 Análise de Variância (ANOVA) para Comparação entre os *Clusters*

Seguindo os procedimentos metodológicos descritos no Capítulo 3, a análise de variância (ANOVA) foi realizada para comparar as médias dos nutrientes entre os dois *clusters*, tanto para a profundidade de 20 cm quanto para a de 40 cm. A seguir, apresentamos os resultados dessa análise para cada profundidade, com ênfase nos nutrientes que apresentaram diferenças estatisticamente significativas entre os *clusters*, conforme indicado pelo valor de *p*.

4.1.10.1 ANOVA para 20 cm

Os resultados da ANOVA para a profundidade de 20 cm indicam que diversos nutrientes apresentaram diferenças significativas entre os *clusters*. Entre os mais relevantes destacam-se o

Ca/T, cuja média foi de 38,34 no *Cluster* 0 e 55,17 no *Cluster* 1, com estatística $t = -7,80$ e p -valor de $3,94 \times 10^{-10}$. O Mg/T também apresentou diferença significativa, com médias de 14,85 e 23,63 e $t = -7,18$ ($p = 3,55 \times 10^{-9}$). O pH variou significativamente entre os grupos, com $t = -6,76$ e $p = 1,57 \times 10^{-8}$, assim como a CTC, cujas médias foram 57,85 e 86,41 para os *clusters* 0 e 1, respectivamente ($t = -6,73$, $p = 1,79 \times 10^{-8}$).

Outros nutrientes que também apresentaram diferenças estatisticamente significativas incluem K/T (0,67 vs. 0,28; $t = 4,02$; $p = 2,03 \times 10^{-4}$), H+Al ($t = 3,63$; $p = 6,79 \times 10^{-4}$), Mg (10,78 vs. 15,84; $t = -3,15$; $p = 2,79 \times 10^{-3}$) e S.B ($t = -3,15$; $p = 2,82 \times 10^{-3}$).

Por outro lado, alguns nutrientes não apresentaram diferenças significativas entre os *clusters*. O P manteve médias iguais de 0,79 para ambos os grupos ($t = -1,73$; $p = 0,09$), enquanto o Ca apresentou médias de 29,71 e 35,50 ($t = -1,59$; $p = 0,12$). A M.O teve diferença mínima ($t = -0,64$; $p = 0,53$), e o V apresentou médias muito próximas, 66,07 e 65,67 ($t = 0,14$; $p = 0,89$), assim como o K, com médias de 0,44 em ambos os *clusters* ($t = 0,14$; $p = 0,89$). Os nutrientes Al, S e a Sat. não apresentaram valores disponíveis para comparação entre os *clusters*.

4.1.10.2 ANOVA para 40 cm

A ANOVA para a profundidade de 40 cm revelou resultados semelhantes aos observados em 20 cm, com algumas diferenças notáveis. Entre os nutrientes que apresentaram diferenças estatisticamente significativas entre os *clusters* destacam-se o K, com média de 0,35 no *Cluster* 0 e 0,27 no *Cluster* 1, apresentando $t = 8,14$ e $p = 1,59 \times 10^{-10}$, indicando uma diferença substancial. O K/T também se mostrou relevante, com médias de 0,47 e -0,14 e $t = 5,80$ ($p = 5,51 \times 10^{-7}$), assim como a M.O, cujas médias foram 2,68 e 2,28, com $t = 5,44$ e $p = 1,92 \times 10^{-6}$. O P apresentou diferença significativa, com $t = 4,94$ e $p = 1,01 \times 10^{-5}$, e o H+Al também mostrou relevância estatística ($t = 4,50$; $p = 4,96 \times 10^{-5}$). Outros nutrientes com diferenças significativas incluem Ca/T (44,98 vs. 35,14; $t = 4,18$; $p = 1,56 \times 10^{-4}$) e Mg/T ($t = 3,68$; $p = 6,54 \times 10^{-4}$).

Por outro lado, diversos nutrientes não apresentaram diferenças significativas entre os *clusters*. O Mg apresentou médias de 10,60 e 9,11 ($t = 1,14$; $p = 0,26$), enquanto o S teve médias de 26,79 e 31,86 ($t = -1,13$; $p = 0,26$). O Ca manteve médias de 23,47 e 20,52 ($t = 1,13$; $p = 0,26$), e o S.B apresentou diferença mínima ($t = -0,87$; $p = 0,39$). O pH apresentou médias de -0,08 e 0,06 ($t = -0,47$; $p = 0,64$), a CTC de 62,59 e 61,93 ($t = 0,14$; $p = 0,89$), e o V de 51,47 e 51,72 ($t = -0,06$; $p = 0,95$). Os nutrientes Al e Sat. não apresentaram valores disponíveis para comparação entre os *clusters*.

4.1.10.3 Conclusões

A análise de variância (ANOVA) para as profundidades de 20 cm e 40 cm revelou diferenças significativas em vários nutrientes entre os *clusters*. Para a profundidade de 20

cm, nutrientes como Ca/T , Mg/T , pH e CTC mostraram diferenças significativas entre os *clusters*, sugerindo que esses fatores são cruciais para a formação dos grupos de solo na camada superficial.

Para a profundidade de 40 cm, os nutrientes K, K/T, M.O e P foram os que mais se destacaram, com diferenças significativas entre os *clusters*, indicando que a dinâmica nutricional nas camadas mais profundas do solo é diferente da camada superficial.

Esses resultados são importantes para o manejo nutricional do solo, pois indicam quais nutrientes devem ser monitorados mais de perto para melhorar a fertilidade e a produtividade agrícola em diferentes profundidades do solo.

4.2 Análises sobre a Produção

Esta seção apresenta os resultados das análises realizadas sobre as variáveis de produção da cana-de-açúcar, incluindo caracterização inicial dos dados, testes de normalidade, normalização, correlações entre variáveis produtivas e análises inferenciais comparando variedades, estágios de corte e efeito do histórico de cortes.

4.2.1 Caracterização Inicial dos Dados de Produção

Além da análise de normalidade para os nutrientes, também foi realizada uma análise semelhante para os dados de produção, como *Área Colhida*, *Prod Real*, *Tch Real*, entre outros. O quadro 15 apresenta os resultados do teste de Shapiro-Wilk para os dados de produção. Todos os parâmetros da produção mostraram distribuições não normais, com p-values muito baixos, indicando que esses dados não seguem uma distribuição normal.

É importante notar que, ao contrário dos nutrientes do solo, os dados de produção não são influenciados pela profundidade do solo, uma vez que a produtividade é determinada por fatores como a variedade da cana, o manejo agrícola e as condições climáticas, não sendo diretamente afetada pelas camadas do solo. Portanto, os dados de produção foram tratados separadamente dos dados de solo e não necessitaram de ajustes relacionados à profundidade.

Quadro 15 – Classificação de normalidade para os dados de produção.

Nutriente	Distribuição	P-Value
<i>Área Colhida</i>	Não Normal	3.13e-13
<i>Prod Real</i>	Não Normal	1.80e-21
<i>Tch Real</i>	Não Normal	3.35e-20
<i>Tch Est</i>	Não Normal	1.49e-21
<i>Brix</i>	Não Normal	2.64e-10
<i>Ls</i>	Não Normal	5.86e-11
<i>Pbu</i>	Não Normal	6.93e-07
<i>ATR</i>	Não Normal	2.08e-11
<i>Fibra</i>	Não Normal	6.90e-07
<i>Ar</i>	Não Normal	1.85e-20
<i>Art</i>	Não Normal	1.99e-11
<i>Pol%Cana</i>	Não Normal	1.17e-11
<i>Pol%Caldo</i>	Não Normal	2.79e-11
<i>Pureza</i>	Não Normal	3.52e-20

Fonte: Elaborado pelo autor.

4.2.2 Normalização dos Dados de Produção

Considerando os resultados da análise de normalidade e seguindo os procedimentos metodológicos descritos na Seção 3.7.2, foi realizada a normalização dos dados de produção. O quadro 16 apresenta os resultados do teste de normalidade e as transformações aplicadas aos dados de produção, como *Área Colhida*, *Prod Real*, *Tch Real*, entre outros.

Observa-se que todos os parâmetros de produção foram transformados com a técnica de Rankit ou Johnson Transformation, atingindo uma aproximação satisfatória à normalidade, com p-values elevados (próximos de 1,0), conforme evidenciado no quadro.

Quadro 16 – Resultados do teste de normalidade e transformações aplicadas aos dados de produção, com os respectivos *p-values* após as transformações.

Nutriente	Transformação Aplicada	P-Value Pós-Transformação
Área Colhida	Rankit	1.0000
Prod Real	Rankit	1.0000
Tch Real	Rankit	1.0000
Tch Est	Rankit	0.8912
Brix	Rankit	1.0000
Ls	Rankit	1.0000
Pbu	Johnson Transformation	0.4008
ATR	Rankit	0.9406
Fibra	Johnson Transformation	0.3815
Ar	Rankit	0.8970
Art	Rankit	0.8970
Pol%Cana	Rankit	0.8947
Pol%Caldo	Rankit	0.9414
Pureza	Rankit	0.9917

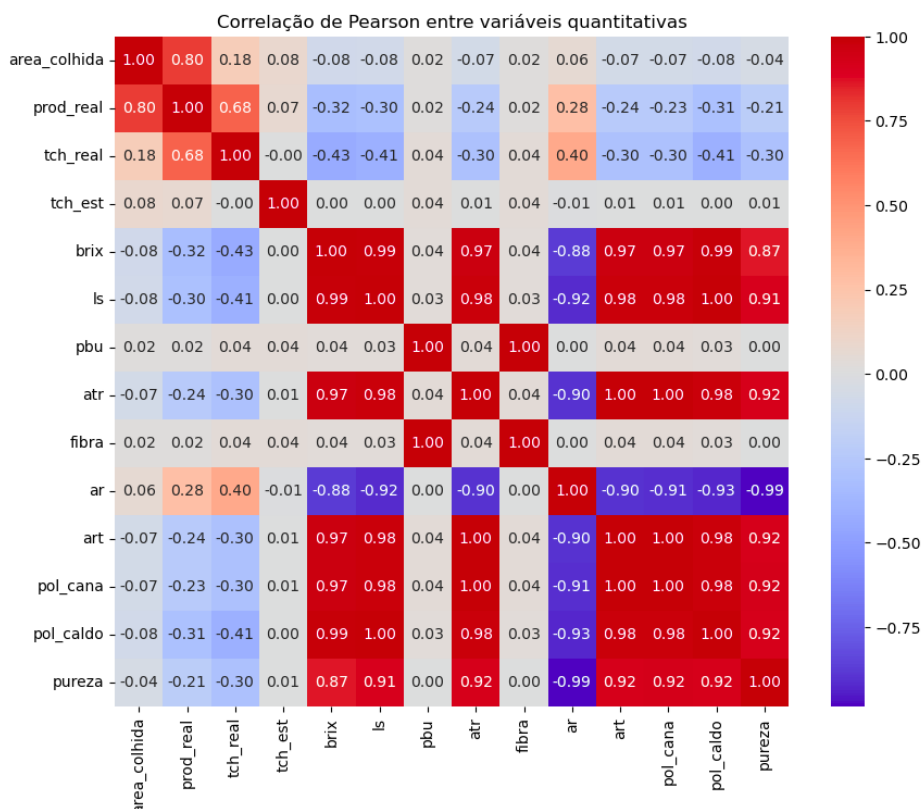
Fonte: Elaborado pelo autor.

4.2.3 Correlação entre as variáveis produtivas

Compreender as correlações entre as variáveis de produção de cana-de-açúcar é essencial para otimizar o processo produtivo e melhorar a eficiência operacional. Ao identificar como variáveis como a área colhida, a produção real, o teor de fibra, o brix e o teor de polpa estão interligadas, é possível tomar decisões mais informadas sobre o manejo, a colheita e o processamento da cana. Esse conhecimento permite não apenas maximizar a produção, mas também garantir que as práticas agrícolas e os processos de fabricação sejam ajustados de maneira a obter o melhor rendimento e qualidade possíveis.

A matriz de correlação entre as variáveis quantitativas do conjunto de dados, apresentada na figura 44, revela as interações entre diferentes fatores relacionados à produção. A seguir, são destacadas as principais observações extraídas da análise.

Figura 44 – Matriz de Correlação de Pearson entre as variáveis quantitativas produtivas.



Fonte: Elaborada pelo autor.

- **Área Colhida e Produção Real:** A variável `area_colhida` apresenta uma forte correlação positiva com a `prod_real` (0.80). Isso indica que, conforme a área colhida aumenta, a produção real tende a aumentar de forma proporcional, o que é esperado em cenários agrícolas, onde mais área cultivada resulta em maior produção.
- **Produção Real e Tecnologia Real:** A correlação entre `prod_real` e `tch_real` é moderada (0.68), sugerindo que a produção real está positivamente associada ao `tch_real`, mas a relação não é perfeita. O valor mostra que fatores externos, como variações no processo de produção, podem influenciar a produção de forma independente da tecnologia utilizada.
- **Tecnologia Estimada:** A variável `tch_est` tem correlações muito baixas com as demais variáveis, com exceção de algumas relações com `brix` (0.99) e `ls` (0.99). Essa alta correlação sugere que, quando a estimativa de produtividade é realizada com base no `brix` ou nos valores de `LS`, a estimativa da tecnologia real é bastante precisa.
- **Brix e ls (Sólidos Solúveis):** As variáveis `brix` e `ls` estão fortemente correlacionadas entre si (0.99), refletindo que essas variáveis estão intimamente relacionadas no processo de produção, provavelmente devido à dependência de características físico-químicas das plantas ou da matéria-prima utilizada.

- **Relações com Fibra, Açúcar e Polpa:** Variáveis como fibra, ar, pol_cana, e pol_caldo apresentam correlações negativas significativas com outras variáveis, como brix e ls. A correlação negativa de ar com brix (-0.88) e com fibra (-0.92) sugere que o aumento na quantidade de ar ou de fibra pode reduzir a concentração de sólidos solúveis e brix, o que é uma observação relevante para o controle de qualidade no processo de produção.
- **Pureza:** A variável pureza apresenta correlações moderadas com a maioria das variáveis, com destaque para a relação com fibra (0.87) e pol_cana (0.91). A alta correlação com pol_cana sugere que a pureza da cana está fortemente relacionada ao conteúdo de pol na cana, o que pode refletir a qualidade da matéria-prima utilizada na produção.
- **Ar e Pureza:** A correlação negativa entre ar e pureza (-0.99) indica que a presença de ar afeta negativamente a pureza da cana, um dado importante para ajustes na produção visando maior rendimento e qualidade.

4.2.4 Conclusão sobre as Correlações Produtivas

A análise de correlação evidenciou as inter-relações entre as variáveis-chave do processo de produção, como área colhida, produtividade, e características físico-químicas da matéria-prima, como fibra e polpa. A forte correlação entre a área colhida e a produção real reflete a lógica agrícola, enquanto as relações entre variáveis como brix, ls, e pureza indicam a importância do controle das variáveis de processo para otimizar a produção e a qualidade. A correlação negativa entre algumas variáveis também sugere a necessidade de ajustes finos no processo produtivo para melhorar a eficiência e maximizar a pureza do produto final.

Esta análise fornece uma base sólida para aprofundar o estudo dos fatores que influenciam a produção e pode ser fundamental para ajustar práticas agrícolas e de processamento, visando à melhoria da produção.

4.2.5 Diferença de Produtividade entre Variedades de Cana-de-Açúcar

Seguindo os procedimentos metodológicos descritos no Capítulo 3, a análise de variância (ANOVA) foi conduzida para verificar se há diferenças significativas na produtividade (Tch Real) entre as variedades de cana-de-açúcar. O teste ANOVA revelou um valor de $F = 18.2213$ com um $p\text{-valor} = 0.0000$, indicando que existem diferenças significativas entre as médias de produtividade das variedades testadas. Este resultado sugere que ao menos uma das variedades apresenta uma produtividade significativamente diferente das demais.

Para investigar quais variedades apresentaram diferenças significativas, foi realizado o teste de comparações múltiplas de Tukey HSD (Honest Significant Difference), com controle da taxa de erro tipo I (FWER) em 0.05. O teste de Tukey permite identificar quais pares de variedades possuem diferenças na produtividade de forma estatisticamente significativa.

Os resultados do teste de Tukey estão apresentados no quadro 17. A partir desse quadro, é possível observar quais comparações entre as variedades apresentaram diferenças significativas, considerando o valor ajustado de p-valor.

Quadro 17 – Comparações múltiplas de médias entre variedades de cana-de-açúcar (teste de Tukey HSD)

Grupo 1	Grupo 2	Diff. Média	p-valor Ajustado	Limite Inferior	Limite Superior	Rejeição
AGRUPADAS	CT96-3346	-23.0065	0.6131	-56.4683	10.4552	Não
AGRUPADAS	CTC1007	19.8313	0.1955	-3.1547	42.8173	Não
AGRUPADAS	CTC20	-14.551	0.9956	-50.2125	21.1106	Não
AGRUPADAS	CTC4	10.5622	0.9969	-16.0894	37.2138	Não
AGRUPADAS	CTC9001	-16.1181	0.0232	-31.2532	-0.983	Sim
AGRUPADAS	CTC9002	-21.9167	0.0393	-43.3824	-0.4509	Sim
AGRUPADAS	CTC9003	-23.1557	0.9064	-65.0881	18.7767	Não
AGRUPADAS	CTC9006	32.5099	0.2238	-5.8892	70.9089	Não
AGRUPADAS	CV7870	-16.3914	0.0012	-29.2157	-3.5671	Sim
AGRUPADAS	IACSP97-4039	-9.0514	0.9993	-34.8103	16.7075	Não
AGRUPADAS	RB855156	26.762	0.9628	-27.0309	80.555	Não
AGRUPADAS	RB867515	-21.3024	0.4709	-50.1017	7.4969	Não
AGRUPADAS	RB92579	-16.5527	1.0	-82.2252	49.1198	Não
AGRUPADAS	RB966928	-7.0015	0.83	-18.7134	4.7105	Não
AGRUPADAS	RB975201	34.6398	0.0	15.3871	53.8926	Sim
AGRUPADAS	RB975242	35.6178	0.0	19.8438	51.3918	Sim
AGRUPADAS	RB975375	43.1227	0.0	22.4946	63.7508	Sim
AGRUPADAS	SP83-2847	-36.1231	0.0086	-67.7684	-4.4779	Sim

Fonte: Elaborado pelo autor.

4.2.5.1 Interpretação dos Resultados

A partir dos resultados do teste de Tukey, observamos que as seguintes comparações apresentaram diferenças significativas na produtividade ($T_{ch\ Real}$) entre as variedades:

- "*CTC9001*" e "*CTC9002*" mostraram diferenças significativas em relação a várias outras variedades, com valores de p-valor ajustado abaixo de 0.05.
- "*CV7870*" e "*SP83-2847*" também apresentaram diferenças significativas, com valores ajustados de p-valor de 0.0012 e 0.0086, respectivamente.
- Por outro lado, variedades como "*CT96-3346*", "*CTC1007*", "*CTC20*", "*CTC4*", "*IACSP97-4039*", e "*RB855156*" não apresentaram diferenças estatísticas significativas entre si, o que sugere que essas variedades têm uma produtividade similar.

4.2.5.2 Conclusão

A análise de ANOVA e o teste de Tukey HSD permitiram identificar que, embora algumas variedades apresentem diferenças significativas na produtividade, outras não apresentam variações estatísticas significativas. Isso é importante para os produtores de cana-de-açúcar, pois pode ajudar na escolha de variedades mais produtivas e, portanto, mais vantajosas para

o cultivo. A utilização de variedades que se destacam por sua produtividade pode otimizar a produção e contribuir para a melhoria da rentabilidade do cultivo de cana.

4.2.6 Diferença de Produtividade entre Estágios de Corte

Seguindo os procedimentos metodológicos descritos no Capítulo 3, a análise de variância (ANOVA) foi realizada para verificar se existe diferença significativa na produtividade (Tch Real) entre os diferentes estágios de corte da cana-de-açúcar. O resultado do teste ANOVA foi $F = 161.4236$ com um $p\text{-valor} = 0.0000$, indicando que há uma diferença estatisticamente significativa entre os estágios de corte. Esse resultado sugere que a produtividade da cana varia significativamente conforme o estágio de corte, sendo necessário investigar mais profundamente essas diferenças.

Para identificar quais estágios de corte apresentaram diferenças significativas, foi utilizado o teste de comparações múltiplas de Tukey HSD (Honest Significant Difference), com controle da taxa de erro tipo I (FWER) em 0.05. O teste de Tukey permite realizar comparações entre as médias dos estágios de corte, determinando quais pares de estágios têm produtividade significativamente diferente.

Os resultados do teste de Tukey estão apresentados no quadro 18. A partir desse quadro, observamos as comparações entre os grupos e as diferenças médias de produtividade, com os respectivos valores ajustados de $p\text{-valor}$.

Quadro 18 – Comparações múltiplas de médias entre estágios de corte (teste de Tukey HSD)

Grupo 1	Grupo 2	Diff. Média	p-valor Ajustado	Limite Inferior	Limite Superior	Rejeição
2 VERÕES	AM REF	-6.2522	0.9877	-23.1362	10.6318	Não
2 VERÕES	ANO - INVERNO REF	-19.1954	0.0773	-39.3018	0.9109	Não
2 VERÕES	BIS ANO INVERNO REF	-56.2622	0.043	-111.6915	-0.8328	Sim
2 VERÕES	SOCA-2	-49.6996	0.0	-66.5104	-32.8888	Sim
2 VERÕES	SOCA-3	-72.4133	0.0	-89.3247	-55.5019	Sim
2 VERÕES	SOCA-4	-75.6622	0.0	-92.613	-58.7115	Sim
2 VERÕES	SOCA-5	-79.4963	0.0	-96.8804	-62.1123	Sim
2 VERÕES	SOCA-6	-80.8207	0.0	-97.8257	-63.8158	Sim
2 VERÕES	SOCA-7	-77.1028	0.0	-96.1909	-58.0147	Sim
2 VERÕES	SOCA-8	-72.9613	0.0	-98.6201	-47.3025	Sim
2 VERÕES	SOCA-9	-89.7437	0.0	-145.173	-34.3143	Sim

Fonte: Elaborado pelo autor.

4.2.6.1 Interpretação dos Resultados

A análise do teste de Tukey revelou que as comparações entre os estágios de corte apresentaram várias diferenças significativas na produtividade (Tch Real). Entre as comparações mais relevantes, destacam-se:

- "2 VERÕES" apresentou diferenças significativas em relação a vários outros estágios de corte, como "SOCA-2", "SOCA-3", "SOCA-4", "SOCA-5", "SOCA-6", "SOCA-7",

"SOCA-8", e "SOCA-9", com p-valor ajustados abaixo de 0.05. Essas diferenças indicam que a produtividade nos estágios de corte mais avançados, como "SOCA-5" e "SOCA-6", é significativamente menor quando comparada com "2 VERÕES".

- "BIS ANO INVERNO REF" não apresentou diferenças significativas em relação aos outros estágios de corte, com p-valor ajustados superiores a 0.05. No entanto, comparado a "2 VERÕES", "BIS ANO INVERNO REF" teve uma diferença significativa de produtividade, com um p-valor ajustado de 0.043.

4.2.6.2 Conclusão

A análise de ANOVA e o teste de Tukey HSD permitiram concluir que existem diferenças significativas na produtividade da cana-de-açúcar conforme o estágio de corte. Os estágios de corte mais avançados, como "SOCA-6", "SOCA-7", "SOCA-8" e "SOCA-9", apresentaram uma produtividade consideravelmente menor em comparação com os estágios iniciais, como "2 VERÕES". Essas informações são valiosas para os produtores, pois indicam quais estágios de corte têm o maior impacto na produtividade, auxiliando na escolha dos momentos ideais para o corte e no planejamento da colheita. O conhecimento sobre a diferença de produtividade entre os estágios pode ser utilizado para melhorar a eficiência do processo produtivo e otimizar a colheita da cana-de-açúcar.

4.2.7 Efeito do Histórico de Cortes

A análise do efeito do histórico de cortes na produtividade (Tch Real) foi conduzida por meio de correlação e testes estatísticos para verificar se o intervalo entre os cortes (curto, médio e longo) influencia de maneira significativa a produtividade da cana-de-açúcar.

4.2.7.1 Correlação Spearman entre Intervalo de Cortes e Produtividade

Primeiramente, foi realizada uma análise de correlação de Spearman entre o intervalo de cortes e a produtividade. O coeficiente de correlação foi de 0.1889^{**} , com um p-valor = 0.0000, indicando que existe uma correlação fraca, porém estatisticamente significativa, entre o intervalo entre os cortes e a produtividade. Embora o valor de correlação não seja alto, ele sugere que há uma relação entre o intervalo entre os cortes e a produtividade da cana, o que pode implicar que intervalos mais curtos entre os cortes estão, de alguma forma, relacionados a uma maior produtividade.

4.2.7.2 Testes de Normalidade e Homogeneidade de Variâncias

O ChatGPT disse:

A normalidade das distribuições de produtividade nos diferentes intervalos de corte (curto, médio e longo) foi avaliada por meio de um teste de normalidade. Os resultados

indicaram que nenhuma das distribuições segue uma distribuição normal, já que todos os testes apresentaram resultados estatisticamente significativos.

Para o intervalo de corte curto, a estatística do teste foi 0,9087; para o intervalo médio, a estatística foi 0,9403 e para o intervalo longo, a estatística observada foi 0,8442. Esses resultados evidenciam que as distribuições de produtividade diferem significativamente de uma distribuição normal em todos os intervalos analisados.

Esses resultados indicam que as distribuições de produtividade não são normais para qualquer intervalo entre os cortes, o que implica que as suposições para testes paramétricos, como a ANOVA, não são atendidas. Esse fato torna o uso de testes não paramétricos mais apropriado.

A homogeneidade das variâncias também foi avaliada, com um valor de estatística de 3.0102 e $p\text{-valor} = 0.0502$. Embora o $p\text{-valor}$ seja ligeiramente superior ao limiar de 0.05, ele ainda sugere que as variâncias podem ser consideradas homogêneas, permitindo a utilização do teste de Kruskal-Wallis para comparações entre os grupos.

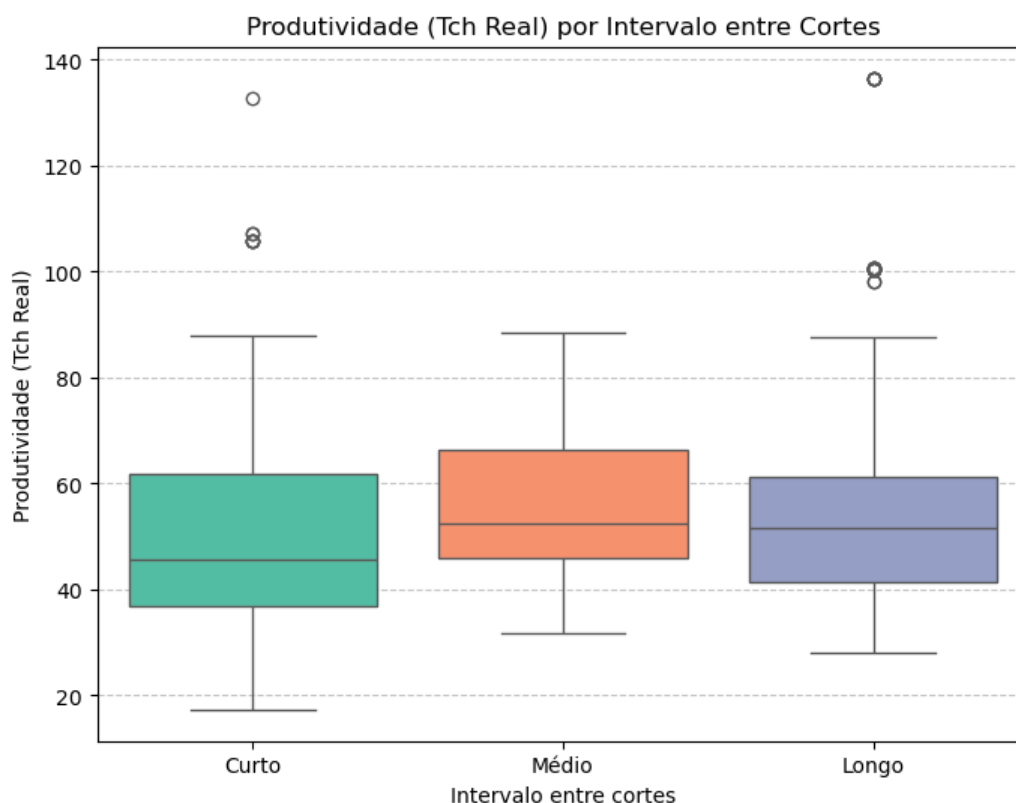
4.2.7.3 Teste de Kruskal-Wallis

Como as distribuições não são normais, foi utilizado o teste de Kruskal-Wallis para avaliar se há diferenças significativas na produtividade (Tch_{Real}) entre os diferentes intervalos de corte. O resultado do teste foi $H = 15.4457$, com um $p\text{-valor} = 0.0004$, indicando que existe uma diferença estatisticamente significativa na produtividade entre os intervalos de corte. Isso sugere que o histórico de cortes (intervalo entre os cortes) tem um efeito sobre a produtividade da cana, com intervalos diferentes apresentando produtividades distintas.

4.2.7.4 Boxplots de Produtividade por Intervalo de Corte

A figura 45 apresenta os boxplots da produtividade para os três grupos de intervalo de corte (curto, médio e longo), a fim de visualizar as distribuições e as possíveis diferenças entre os grupos.

Figura 45 – Boxplot de Produtividade para Intervalo entre Cortes.



Fonte: Elaborada pelo autor.

Os boxplots apresentados permitem observar de forma detalhada o comportamento da produtividade em função dos diferentes intervalos entre cortes. No caso do intervalo curto, percebe-se a presença de alguns outliers, indicando que há registros de produtividade que se distanciam significativamente do comportamento central do grupo. Isso sugere maior variabilidade e possíveis oscilações associadas a cortes mais frequentes.

Para o intervalo médio, observa-se um boxplot sem a presença de outliers, evidenciando uma distribuição mais estável e homogênea. A dispersão dos valores é mais equilibrada, indicando que esse intervalo tende a apresentar produtividades mais consistentes e com menor ocorrência de valores extremos.

Já no intervalo longo, novamente surgem alguns outliers, o que demonstra que, apesar de a maior parte dos valores estar concentrada em uma faixa relativamente definida, existem casos de produtividade atípica. Essa condição pode refletir variações ambientais, manejo diferenciado ou condições específicas que impactam lavouras submetidas a intervalos maiores entre os cortes.

De modo geral, a comparação visual evidencia que o intervalo médio apresenta a distribuição mais estável, enquanto os intervalos curto e longo carregam maior presença de valores extremos. Isso sugere que a frequência intermediária de cortes pode estar associada a

uma produtividade mais regular ao longo do tempo.

4.2.7.5 Conclusão

Os resultados obtidos indicam que o intervalo entre os cortes tem um efeito significativo na produtividade da cana-de-açúcar. Embora a correlação entre o intervalo de corte e a produtividade seja fraca, o teste de Kruskal-Wallis revelou que há diferenças estatísticas entre os intervalos, com os intervalos curtos apresentando uma produtividade mais alta. A análise de normalidade e homogeneidade das variâncias mostrou que os dados não seguem distribuições normais, o que justificou o uso do teste não paramétrico.

A compreensão desse efeito pode ser útil para os produtores, pois sugere que o ajuste no intervalo entre os cortes pode influenciar a produtividade da cana, ajudando na definição de práticas de manejo que otimizem o rendimento das lavouras.

4.3 Nutrientes e Produção

Esta seção apresenta os resultados das análises que relacionam os nutrientes do solo com as variáveis de produção da cana-de-açúcar, incluindo correlações, análise de importância de variáveis e construção de modelos preditivos.

4.3.1 Correlação entre Nutrientes e Variáveis de Produção

Seguindo os procedimentos metodológicos descritos no Capítulo 3, a análise das correlações entre os nutrientes presentes no solo e as variáveis de produção forneceu informações importantes sobre como a composição nutricional do solo pode influenciar a produtividade e a qualidade da cana-de-açúcar. Foram encontradas um total de 224 correlações, das quais 17 foram estatisticamente significativas ($p < 0.05$). A seguir, são apresentadas as correlações mais fortes, tanto positivas quanto negativas, destacando os nutrientes que mais influenciam as variáveis de produção.

4.3.1.1 Melhores Correlações (Mais Fortes)

As correlações mais fortes, com valores de r positivos, indicam uma relação direta entre os nutrientes e as variáveis de produção. As melhores correlações encontradas foram:

- **Matéria Orgânica (M.O) a 40 cm e Brix:** $r = 0.409$ (positiva) com $p < 0.05$, indicando que uma maior concentração de matéria orgânica no solo está positivamente relacionada ao teor de açúcar na cana.

- **Matéria Orgânica (M.O) a 40 cm e Polpa de Cana (Pol% Cana):** $r = 0.401$ (positiva) com $p < 0.05$, sugerindo que solos com maior matéria orgânica favorecem o aumento do teor de polpa da cana.
- **Matéria Orgânica (M.O) a 40 cm e ATR (kg/t):** $r = 0.401$ (positiva) com $p < 0.05$, evidenciando uma relação positiva entre a matéria orgânica e a quantidade de açúcar recuperado por tonelada de cana.
- **Matéria Orgânica (M.O) a 40 cm e Pureza:** $r = 0.388$ (positiva) com $p < 0.05$, indicando que maiores concentrações de matéria orgânica no solo estão associadas a uma maior pureza na cana.
- **Matéria Orgânica (M.O) a 20 cm e Brix:** $r = 0.352$ (positiva) com $p < 0.05$, sugerindo uma relação positiva, embora mais fraca, entre o teor de matéria orgânica a 20 cm de profundidade e o teor de açúcar na cana.
- **Matéria Orgânica (M.O) a 20 cm e Polpa de Cana (Pol% Cana):** $r = 0.348$ (positiva) com $p < 0.05$, refletindo uma leve tendência de aumento do teor de polpa com maiores concentrações de matéria orgânica.
- **Matéria Orgânica (M.O) a 20 cm e ATR (kg/t):** $r = 0.348$ (positiva) com $p < 0.05$, mostrando que mais matéria orgânica no solo também pode impactar positivamente a quantidade de açúcar recuperado.
- **Matéria Orgânica (M.O) a 40 cm e Fibra:** $r = 0.342$ (positiva) com $p < 0.05$, indicando que uma maior quantidade de matéria orgânica está associada a um aumento no teor de fibra da cana.
- **Matéria Orgânica (M.O) a 20 cm e Pureza:** $r = 0.337$ (positiva) com $p < 0.05$, sugerindo que uma maior concentração de matéria orgânica pode melhorar a pureza da cana.
- **Potássio (K) a 40 cm e Fibra:** $r = 0.335$ (positiva) com $p < 0.05$, indicando uma correlação positiva entre o teor de potássio no solo e o teor de fibra na cana.

Essas correlações positivas sugerem que os nutrientes, especialmente a matéria orgânica, têm um papel crucial no aumento da produtividade e na melhoria das características da cana-de-açúcar, como o teor de açúcar (Brix), a pureza, a fibra e a quantidade de polpa.

4.3.1.2 Piores Correlações (Mais Negativas)

Por outro lado, também foram observadas correlações negativas entre alguns nutrientes e as variáveis de produção, indicando que um aumento em certos nutrientes pode estar

relacionado a uma queda em outras características da cana-de-açúcar. As piores correlações encontradas foram:

- **Matéria Orgânica (M.O) a 20 cm e Tch Real:** $r = -0.375$ (negativa) com $p < 0.05$, mostrando que um aumento na matéria orgânica a 20 cm de profundidade pode estar negativamente associado ao rendimento de produção (Tch Real).
- **Matéria Orgânica (M.O) a 40 cm e Tch Real:** $r = -0.374$ (negativa) com $p < 0.05$, indicando uma relação negativa entre a matéria orgânica a 40 cm e a produtividade total.
- **Enxofre (S) a 40 cm e ATR (kg/t):** $r = -0.344$ (negativa) com $p < 0.05$, sugerindo que maiores concentrações de enxofre no solo podem reduzir a quantidade de açúcar recuperado por tonelada de cana.
- **Enxofre (S) a 40 cm e Polpa de Cana (Pol% Cana):** $r = -0.343$ (negativa) com $p < 0.05$, refletindo uma tendência negativa entre o teor de enxofre e o teor de polpa da cana.
- **Enxofre (S) a 40 cm e Brix:** $r = -0.339$ (negativa) com $p < 0.05$, indicando que o aumento do enxofre no solo pode estar negativamente relacionado ao teor de açúcar na cana.

Essas correlações negativas sugerem que o excesso de certos nutrientes, como enxofre e matéria orgânica em determinadas profundidades, pode prejudicar a produtividade e a qualidade da cana-de-açúcar, especialmente no que diz respeito à produção de açúcar e ao teor de polpa.

4.3.1.3 Conclusão

A análise das correlações entre os nutrientes e as variáveis de produção revelou tanto relações positivas quanto negativas entre os nutrientes do solo e as características da cana-de-açúcar. As correlações positivas, especialmente com a matéria orgânica, indicam que o aumento desses nutrientes pode favorecer a produtividade, a pureza e o teor de fibra da cana. No entanto, as correlações negativas sugerem que o excesso de alguns nutrientes, como enxofre e matéria orgânica em determinadas profundidades, pode prejudicar o rendimento de produção e a qualidade do produto final.

Esses resultados fornecem informações importantes para o manejo do solo, indicando que é necessário equilibrar os nutrientes para otimizar a produção e a qualidade da cana-de-açúcar. O conhecimento dessas correlações pode auxiliar os produtores na formulação de estratégias para melhorar a eficiência e maximizar a produtividade das lavouras.

4.3.2 Importância dos Nutrientes para a Produtividade

Seguindo os procedimentos metodológicos descritos no Capítulo 3, foi realizada a análise de importância dos nutrientes para a produtividade utilizando o algoritmo RandomForestRegressor. A produtividade dos sistemas agrícolas, especialmente aqueles voltados para a fertilidade do solo, está intimamente relacionada ao equilíbrio adequado dos nutrientes.

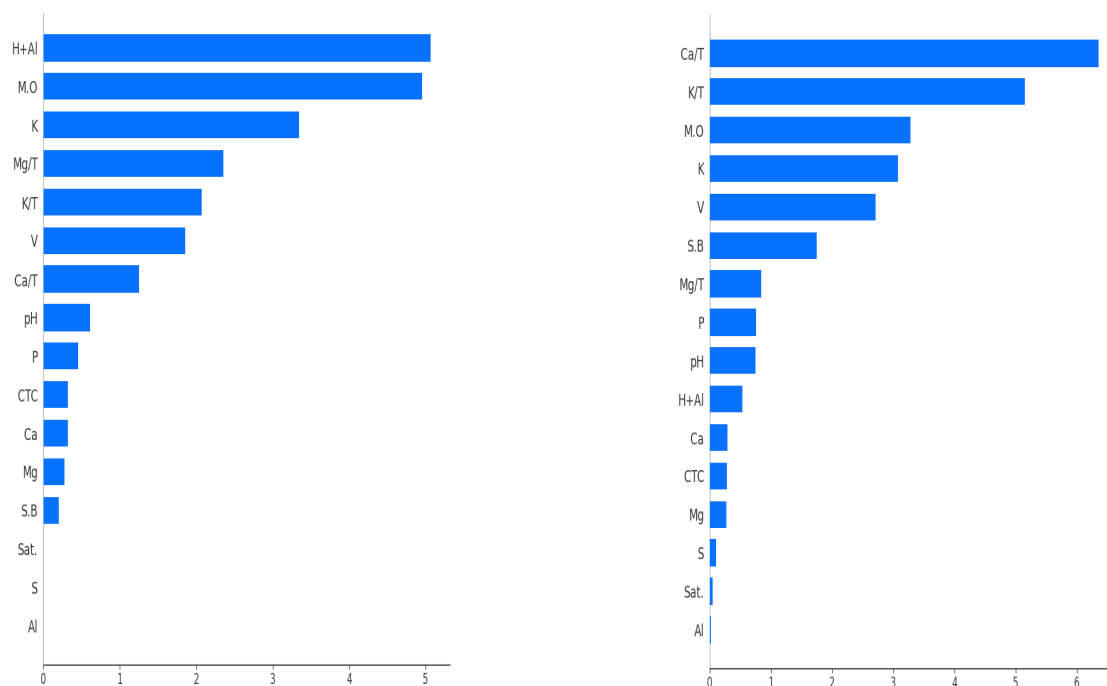
Os dados mostram que **H+Al** e **M.O** são os fatores mais influentes para 20cm e 40cm de profundidade, respectivamente. Esses nutrientes têm um impacto significativo na qualidade geral do solo, afetando a disponibilidade de outros nutrientes e influenciando a acidez e basicidade do solo. Níveis adequados de matéria orgânica contribuem para a estrutura do solo, retenção de água e atividade microbiana, todos essenciais para o crescimento das culturas.

O conteúdo de **K** no solo é outro elemento importante, influenciando a fotossíntese, a síntese de proteínas e a vitalidade geral das plantas. Da mesma forma, **Mg/T** e **K/T** também são críticos, com o magnésio contribuindo para a produção de clorofila e o potássio aumentando a resistência da planta a doenças e estresses ambientais.

Outros elementos, como **Ca/T** e **V**, aparecem logo abaixo em importância. O cálcio auxilia no desenvolvimento das raízes e melhora o equilíbrio do pH do solo, enquanto a saturação de bases é um indicador da capacidade do solo de reter nutrientes essenciais.

A figura 46 exibe a distribuição das importâncias dos nutrientes em duas profundidades do solo: a 20 cm e a 40 cm. Esses dados são importantes para entender a dinâmica dos nutrientes em diferentes camadas do solo e como isso pode impactar a produtividade da cana-de-açúcar.

Figura 46 – Importância dos nutrientes a 20cm e 40cm de profundidade para a produtividade.



(a) Nutrientes a 20 cm de profundidade.

(b) Nutrientes a 40 cm de profundidade.

Fonte: Elaborada pelo autor.

Em conclusão, garantir o equilíbrio adequado desses nutrientes no solo é crucial para maximizar a produtividade. Ajustes nos níveis de nutrientes com base nessa análise podem levar a um crescimento mais vigoroso das plantas, melhor resistência a doenças e, consequentemente, maiores rendimentos, contribuindo para a sustentabilidade e o sucesso dos sistemas agrícolas.

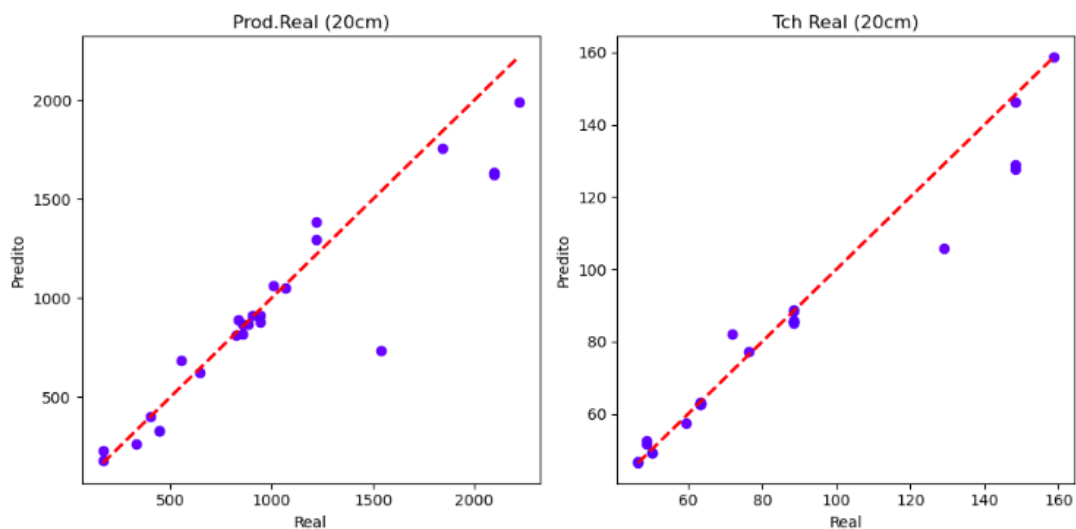
4.3.3 Construção dos Modelos para Previsão de Produção e Produtividade

Seguindo os procedimentos metodológicos descritos no Capítulo 3, foram construídos dois modelos de regressão linear múltipla, um para a profundidade de 20cm e outro para a profundidade de 40cm. O objetivo principal é capturar as diferentes influências que as variáveis nutricionais podem exercer sobre a produção (*Prod.Real*) e a produtividade (*Tch Real*) em diferentes camadas do solo.

4.3.3.1 Visualizações dos Resultados

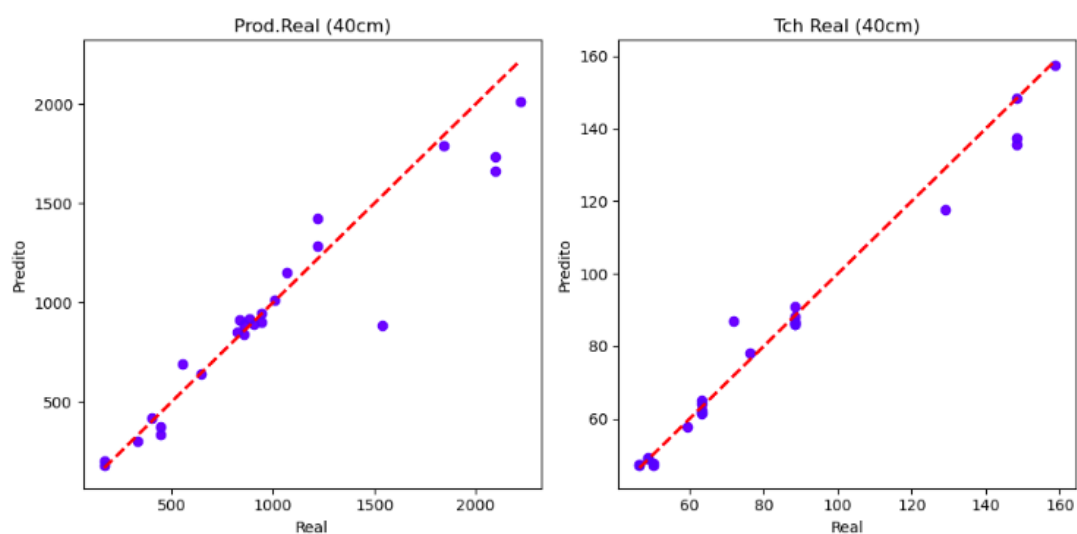
Para facilitar a interpretação da qualidade das previsões, foram gerados gráficos de dispersão comparando valores reais versus valores preditos. A figura 47 apresenta os gráficos para as previsões a 20cm de profundidade, enquanto a figura 48 mostra os resultados para as previsões a 40cm de profundidade.

Figura 47 – Gráficos de Dispersão para Previsões a 20cm de Profundidade



Fonte: Elaborada pelo autor.

Figura 48 – Gráficos de Dispersão para Previsões a 40cm de Profundidade



Fonte: Elaborada pelo autor.

4.3.3.2 Acurácia dos Modelos Obtidos

A seguir, são apresentadas as métricas de desempenho obtidas pelos modelos de regressão linear desenvolvidos para as profundidades de 20 cm e 40 cm. As métricas utilizadas para avaliação foram o coeficiente de determinação (R^2) e o erro quadrático médio (RMSE), que permitem avaliar, respectivamente, o grau de explicação do modelo sobre a variabilidade dos dados e o erro médio de predição em relação aos valores observados.

▪ Modelagem Regressão Linear – 20 cm:

- **Prod. Real:** $R^2 = 0.848$, RMSE = 224.278
- **TCH Real:** $R^2 = 0.950$, RMSE = 7.812

▪ **Modelagem Regressão Linear – 40 cm:**

- **Prod. Real:** $R^2 = 0.890$, RMSE = 190.356
- **TCH Real:** $R^2 = 0.977$, RMSE = 5.288

Observa-se que, em ambas as profundidades analisadas, os modelos apresentaram um bom ajuste, com valores de R^2 superiores a 0.84, indicando que mais de 84% da variação observada nos dados é explicada pelas variáveis independentes incluídas nas regressões.

Além disso, nota-se uma melhora nos resultados para a profundidade de 40 cm, onde o R^2 atingiu 0.890 para a produtividade real e 0.977 para o TCH real, acompanhada de uma redução significativa no RMSE. Esses resultados sugerem que os atributos do solo coletados a 40 cm de profundidade possuem uma relação mais consistente com a variável resposta, o que pode indicar maior estabilidade e representatividade dessa camada para a modelagem da produtividade.

Em síntese, os modelos obtidos demonstram boa capacidade preditiva e coerência estatística, validando a utilização da regressão linear como ferramenta eficaz para estimar a produtividade agrícola com base nas características físico-químicas do solo em diferentes profundidades.

4.3.3.3 Conclusão

A construção dos dois modelos, um para cada profundidade (20cm e 40cm), permite uma análise detalhada das variações de produção e produtividade em função das camadas do solo. A previsão final, obtida a partir da média dos resultados dos dois modelos, proporciona uma estimativa mais equilibrada e precisa, refletindo as contribuições de cada camada na produtividade total. Com isso, é possível entender melhor como as diferentes profundidades do solo influenciam o desempenho das culturas, possibilitando um planejamento mais assertivo e eficiente na agricultura.

5 Conclusão

Este trabalho investigou as relações entre os nutrientes do solo em diferentes profundidades e a produtividade da cana-de-açúcar, desenvolvendo ferramentas de análise estatística e modelos preditivos para apoiar a tomada de decisão no manejo agrícola. Por meio de uma abordagem que integrou estatística descritiva, análises inferenciais, técnicas de aprendizado de máquina e desenvolvimento de sistemas computacionais, foi possível atingir resultados significativos que contribuem tanto para o conhecimento científico quanto para a aplicação prática na agricultura de precisão.

5.1 Principais Resultados

A análise exploratória revelou que a maioria dos nutrientes (69%) apresentou distribuições não normais, com presença de *outliers* em aproximadamente 38% dos nutrientes a 20 cm e 31% a 40 cm, evidenciando forte heterogeneidade espacial. A aplicação sistemática de transformações estatísticas (Box-Cox, Johnson, logarítmica e Rankit) permitiu normalizar os dados, viabilizando testes paramétricos robustos.

O teste t pareado confirmou diferenças altamente significativas entre as profundidades para a maioria dos nutrientes, destacando-se Fósforo ($p = 1,93 \times 10^{-70}$), Potássio ($p = 1,18 \times 10^{-21}$) e Cálcio ($p = 3,61 \times 10^{-22}$). Notavelmente, pH, Alumínio e Saturação por Bases não apresentaram diferenças significativas, sugerindo eficiência das práticas de calagem em promover correção em profundidade. A camada de 20 cm apresentou maiores concentrações e maior variabilidade, enquanto 40 cm demonstrou maior estabilidade química.

A análise de correlação revelou interações complexas, com destaque para a forte relação entre Ca/T e Mg/T (0,91 a 20 cm e 0,85 a 40 cm) e a correlação negativa entre K/T e pH (-0,60 a 20 cm), evidenciando o efeito da acidez sobre a disponibilidade de potássio. A Análise de Componentes Principais identificou três eixos principais de variação: acidez (PC1), acidez potencial (PC2) e fertilidade básica (PC3), com os três primeiros componentes explicando mais de 90% da variância total.

A clusterização identificou entre 3 e 4 perfis distintos de fertilidade, validados pelos métodos do Cotovelo e Silhouette Score. A análise de sensibilidade revelou que Mg/T, Ca/T e pH foram mais determinantes a 20 cm, enquanto K, M.O e Ca/T apresentaram maior relevância a 40 cm, reforçando a necessidade de estratégias de manejo específicas para cada camada.

A análise de variância entre variedades revelou diferenças altamente significativas ($F = 18,22$, $p < 0,0001$), com RB975375, RB975242 e RB975201 apresentando desempenho superior. O estágio de corte também mostrou efeito significativo ($F = 161,42$, $p < 0,0001$),

com declínio progressivo da produtividade nos cortes subsequentes. O histórico de cortes confirmou relação significativa com a produtividade ($H = 15,45$, $p = 0,0004$), indicando que intervalos mais curtos tendem a proporcionar maior produtividade.

Das 224 correlações avaliadas entre nutrientes e variáveis produtivas, 17 foram estatisticamente significativas. As correlações positivas mais fortes foram observadas entre M.O e variáveis de qualidade a 40 cm (Brix: $r = 0,409$; Pol%Cana: $r = 0,401$; ATR: $r = 0,401$), evidenciando o papel fundamental da matéria orgânica na qualidade tecnológica da cana. A análise de importância revelou que H+Al e M.O foram os fatores mais influentes para a produtividade a 20 cm e 40 cm, respectivamente.

Os modelos de regressão linear múltipla apresentaram desempenho satisfatório, com capacidade preditiva superior a 40 cm: Prod. Real ($R^2 = 0,890$, $RMSE = 190,36$) e Tch Real ($R^2 = 0,977$, $RMSE = 5,29$). Os valores elevados de R^2 indicam que mais de 84% da variação na produtividade pode ser explicada pelos nutrientes do solo, validando a abordagem de manejo baseada em análise química.

5.2 Sistemas Desenvolvidos

A materialização dos resultados em sistemas computacionais representa contribuição significativa. O AgroApp integra funcionalidades de predição de produtividade, análise de solo, gestão de estoque e monitoramento climático em interface intuitiva. O AgroAnalysis complementa com análises estatísticas avançadas, incluindo visualizações de correlações, PCA, clusterização e predições. Ambos os sistemas foram projetados considerando usabilidade e acessibilidade, fundamentais para adoção no campo.

5.3 Implicações Práticas

Os resultados fornecem subsídios concretos para otimização do manejo:

1. **Amostragem estratificada:** A significância das diferenças entre profundidades reforça a necessidade de amostragem em múltiplas camadas;
2. **Correção em profundidade:** A maior acidez a 40 cm indica necessidade de estratégias para incorporação de corretivos em profundidade;
3. **Manejo por zona:** A identificação de *clusters* viabiliza agricultura de precisão com aplicação de insumos em taxa variável;
4. **Seleção varietal:** As diferenças significativas entre variedades indicam oportunidade de otimização através de escolha criteriosa;

5. **Renovação planejada:** O declínio da produtividade com o avanço dos cortes fornece base para decisões sobre momento ideal de renovação;
6. **Equilíbrio nutricional:** As correlações identificadas orientam manutenção de proporções adequadas entre nutrientes, evitando antagonismos.

5.4 Considerações Finais

Este trabalho demonstrou que a integração de análises estatísticas rigorosas, técnicas de aprendizado de máquina e desenvolvimento de sistemas aplicados pode gerar conhecimento acionável para o setor agrícola. Os modelos preditivos desenvolvidos apresentaram capacidade de explicar mais de 84% da variação na produtividade com base nos nutrientes do solo, e sua materialização em sistemas computacionais representa passo importante na democratização da agricultura de precisão.

Os resultados reforçam que o manejo nutricional eficiente da cana-de-açúcar demanda abordagem holística, considerando múltiplos nutrientes, suas interações, a estratificação vertical no perfil do solo e fatores relacionados à variedade e ao histórico de manejo. A identificação de perfis nutricionais distintos abre possibilidade de manejo diferenciado por zona, com potencial de otimização do uso de insumos e aumento de eficiência produtiva.

A forte relação entre matéria orgânica e qualidade tecnológica, evidenciada pelas correlações significativas com Brix, ATR e Pureza, destaca a importância de práticas que favoreçam acúmulo de carbono no solo, contribuindo para sustentabilidade do sistema. Este trabalho ilustra como a ciência de dados pode ser aplicada à agricultura, transformando dados em informação, informação em conhecimento, e conhecimento em ação.

Os sistemas desenvolvidos representam ponte entre pesquisa científica e prática agrícola, com potencial de impactar positivamente a produtividade, lucratividade e sustentabilidade da produção de cana-de-açúcar. A continuidade desta linha de pesquisa pode contribuir significativamente para consolidação da agricultura de precisão como paradigma na produção agrícola brasileira.

5.5 Trabalhos Futuros

A partir dos resultados alcançados neste estudo e das limitações identificadas, diversas possibilidades de aprofundamento e expansão são sugeridas para trabalhos futuros. Um primeiro caminho consiste na ampliação da base de dados utilizada. Embora o conjunto atual tenha permitido análises robustas, a inclusão de novas fazendas, diferentes ambientes de produção e séries históricas mais longas possibilitaria aprimorar a generalização dos modelos, além de permitir análises temporais mais refinadas sobre tendências e impactos do manejo ao longo dos anos.

Outra direção importante envolve o desenvolvimento e teste de modelos preditivos mais avançados. Embora a regressão linear múltipla tenha apresentado excelente desempenho — explicando mais de 84% da variação da produtividade — algoritmos não lineares, como Random Forest, Gradient Boosting, XGBoost e Redes Neurais Artificiais, podem captar interações complexas entre nutrientes e efeitos de saturação que modelos lineares não conseguem representar plenamente. Estudos comparativos entre diferentes abordagens podem identificar modelos ainda mais precisos e robustos para apoiar decisões agronômicas.

No âmbito das análises estatísticas, a incorporação de métodos espaciais e geoestatísticos representa uma oportunidade relevante, especialmente considerando a heterogeneidade evidenciada nas amostras. O uso de *krigagem*, variogramas e mapas interpolados permitiria visualizar a variabilidade espacial dos nutrientes e gerar recomendações de manejo mais precisas, alinhadas à lógica da agricultura de precisão. Da mesma forma, modelos mistos (mixed models) poderiam explorar a variação entre quadras, lotes e fazendas, oferecendo compreensão mais detalhada da estrutura hierárquica dos dados.

Outra possibilidade é a integração de variáveis climáticas e informações remotas. Dados de precipitação, temperatura, *NDVI*, imagens de satélite ou sensores embarcados em drones podem complementar os dados de solo e enriquecer as análises preditivas, ampliando o entendimento do efeito combinado entre ambiente, manejo e fertilidade. A inclusão dessas variáveis tem potencial de aumentar substancialmente a precisão das previsões de produtividade e qualidade da cana.

Quanto aos sistemas desenvolvidos, há espaço para evolução tanto funcional quanto tecnológica. No AgroApp, a incorporação de módulos adicionais — como recomendações automáticas de calagem e adubação com base nos resultados das análises, geração de relatórios personalizados e integração com sistemas de gestão agrícola (ERPs) — ampliaria o valor prático da plataforma. Para o AgroAnalysis, a implementação de dashboards interativos, análises espaciais e versões online acessíveis em nuvem poderia democratizar ainda mais o acesso às informações e favorecer a adoção por produtores e consultores.

Por fim, recomenda-se a realização de experimentos de campo que validem empiricamente as previsões dos modelos e as recomendações sugeridas pelos sistemas. Ensaios controlados que testem diferentes doses de nutrientes, combinações de manejo e estratégias de correção permitirão comparar os resultados reais com os valores estimados, fortalecendo a confiabilidade das ferramentas e contribuindo para evoluir o conhecimento agronômico.

Em conjunto, essas possibilidades apontam para uma linha contínua de pesquisa que une ciência de dados, agronomia e desenvolvimento de sistemas, fortalecendo a agricultura de precisão e promovendo maior eficiência, sustentabilidade e competitividade na produção de cana-de-açúcar.

Referências

- AONO, A. H.; MIRANDA, M. C. d. C.; PINHEIRO, J. B.; FAGUNDES, T. G. Uma abordagem inovadora baseada em imagens para fenotipagem de sementes de soja usando técnicas de aprendizado de máquina. *Agronomy Journal*, v. 117, n. 3, p. 1234–1245, 2025. Acesso em: 16 set. 2025. Disponível em: <https://www.researchgate.net/publication/364576252_A_novel_image-based_approach_for_soybean_seed_phenotyping_using_machine_learning_techniques>.
- AZEVEDO, M. C. B. d.; CHOPART, J. L.; MEDINA; *et al.* Sugarcane root length density and distribution from root intersection counting on a trench-profile. *Scientia Agricola*, SciELO Brasil, v. 68, p. 94–101, 2011.
- AZEVEDO, N. S. de; MARTINEZ, Y. P.; VEIGA, A. K. da C.; BATISTA, A. H.; SANTOS, W. O.; MARTINS, A. G.; WENDLING, B.; PEREIRA, M. G. Manejo do solo em sistemas integrados de produção.
- BUSSAB, W. d. O.; MORETTIN, P. A. Estatística básica. In: *Estatística básica*. [S.l.: s.n.], 2010. p. xvi–540.
- CAIRES, E. F. Calagem no sistema de plantio direto para correção da acidez e suprimento de ca e mg como nutrientes. *Simpósio sobre Informações Recentes para Otimização da Produção Agrícola, International Plant Nutrition Institute IPNI, Piracicaba, SP, 28p*, 2007.
- CRUSCIOL, C. A. C.; CAMPOS, M. d.; MARTELLO, J. M.; ALVES, C. J.; NASCIMENTO, C. A. C.; PEREIRA, J. C. d. R.; CANTARELLA, H. Organomineral fertilizer as source of p and k for sugarcane. *Scientific Reports*, Nature Publishing Group UK London, v. 10, n. 1, p. 5398, 2020.
- Embrapa. *Análise de solo*. 2024. <<https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/cana-de-acucar/producao/correcao-e-adubacao/diagnose-das-necessidades-nutricionais/analise-de-solo>>. Acessado em: 12 out. 2025.
- Empresa Brasileira de Pesquisa Agropecuária. *Cana-de-Açúcar*. 2023. <<https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/cana>>. Acessado em 28 de maio de 2024.
- FERREZ, R. L. d. S.; BARBOSA, M.; BATISTA; *et al.* Gessagem em cana-de-açúcar: efeitos no solo e produção. *InterfacEHS – Saúde, Meio Ambiente e Sustentabilidade*, v. 10, n. 1, p. 1–15, jun 2015. Acesso em: 20 set. 2025. Disponível em: <https://www.researchgate.net/publication/291830867_Gessagem_em_cana_de_acucar_efeitos_no_solo_e_producao>.
- FREIRE, F.; REZENDE, J.; SILVA, F. da; ANDRADE, C. de; REIN, T.; MARCHIORI, L.; FERNANDO, J. F. *et al.* Fertilidade do solo e adubação em cana-de-açúcar. 2024.
- GUEDES, T. A.; MARTINS, A. B. T.; ACORSI, C. R. L.; JANEIRO, V. Estatística descritiva. *Projeto de ensino aprender fazendo estatística*, sn, v. 1, p. 1–49, 2005.
- HERRERA, B.; FERNEY, W. *The potential of polyhalite as a multi-nutrient fertilizer for sugarcane*. Tese (Doutorado) — Universidade de São Paulo, 2019.

LIAKOS, K. G.; BUSATO, P.; MOSHOU, D.; PEARSON, S.; BOCHTIS, D. Machine learning in agriculture: A review. *Sensors*, Mdpi, v. 18, n. 8, p. 2674, 2018.

MOLIN, J. P.; AMARAL, L. R. do; COLAÇO, A. *Agricultura de precisão*. [S.l.]: Oficina de textos, 2015.

OLIVEIRA, A. C. S. d.; SOUZA, A. A. d.; LACERDA, W. S.; GONÇALVES, L. R. Aplicação de redes neurais artificiais na previsão da produção de álcool. *Ciência e Agrotecnologia*, v. 34, n. 2, p. 279–284, 2010. Acesso em: 16 set. 2025. Disponível em: <<https://www.scielo.br/j/cagro/a/HWFRGpHsMrQkShH8WDXggbM/?lang=pt>>.

OLIVEIRA, E. L. d.; ANDRADE, L. A. d. B.; FARIA, M. A. d.; EVANGELISTA, A. W. P.; MORAIS, A. R. d. Uso de vinhaça de alambique e nitrogênio em cana-de-açúcar irrigada e não irrigada. *Pesquisa agropecuária brasileira*, SciELO Brasil, v. 44, p. 1398–1403, 2009.

OLIVEIRA, M. W. de; MACÊDO, G. A. R.; MARTINS, J. A.; SILVA, V. S. G. da; OLIVEIRA, A. B. de. Mineral nutrition and fertilization of sugarcane. *Sugarcane. Technology and Research*, IntechOpen, v. 1, p. 169–191, 2018.

OTTO, R.; TRIVELIN, P. C. O.; FRANCO, H. C. J.; FARONI, C. E.; VITTI, A. C. Root system distribution of sugar cane as related to nitrogen fertilization, evaluated by two methods: monolith and probes. *Revista Brasileira de Ciência do Solo*, SciELO Brasil, v. 33, p. 601–611, 2009.

PROVOST, F.; FAWCETT, T. *Data Science for Business: What you need to know about data mining and data-analytic thinking*. [S.l.]: "O'Reilly Media, Inc.", 2013.

ROSSETTO, R.; SPIRONELLO, A.; CANTARELLA, H.; QUAGGIO, J. A. Calagem para a cana-de-açúcar e sua interação coma adubação potássica. *Bragantia*, SciELO Brasil, v. 63, p. 105–119, 2004.

SILVA, J. R.; OLIVEIRA, A. F.; SANTOS, R. M.; *et al.* Fertilidade do solo e absorção de nutrientes em cana-de-açúcar. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 27, n. 6, p. 396–402, 2023. Acesso em: 16 set. 2025. Disponível em: <<https://www.scielo.br/j/rbeaa/a/sgZNsQtFvRSx587NstrLGQh/?lang=pt>>.

SUGAWARA, L. M.; *et al.* Estimativa de produtividade de cana-de-açúcar (*Saccharum officinarum* L.) por meio de técnica de análise de regressão linear múltipla. In: *Simpósio Brasileiro de Sensoriamento Remoto, 13., 2007, Florianópolis*. Florianópolis: INPE, 2007.

TUKEY, J. W.; *et al.* *Exploratory data analysis*. [S.l.]: Springer, 1977. v. 2.

União da Indústria de Cana-de-Açúcar. *Safra 2023/2024 termina como a maior da história*. 2024. <<https://unica.com.br/noticias/safra-2023-2024-termina-como-a-maior-da-historia>>. Acessado em 26 de maio de 2024.

WOLFERT, S.; GE, L.; VERDOUW, C.; BOGAARDT, M. *Big data in smart farming—a review, agricultural systems*. 153: 69–80. [S.l.]: ISSN-0308-521X, <https://doi.org/10.1016/j.agry.2017.01.023>, 2017.