



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Faculdade de Engenharia e Ciências de Guaratinguetá

ISABELA MANSUR CAETANO

Análise de dados para correlacionar o nível de satisfação do cliente com a qualidade de rede: um ensaio técnico

Guaratinguetá - SP
2023

Isabela Mansur Caetano

Análise de dados para correlacionar o nível de satisfação do cliente com a qualidade de rede: um ensaio técnico

Trabalho de Graduação apresentado ao Conselho de Curso de Graduação em Engenharia Elétrica da Faculdade de Engenharia e Ciências do Campus de Guaratinguetá, Universidade Estadual Paulista, como parte dos requisitos para obtenção do diploma de Graduação em Engenharia Elétrica.

Orientador: Prof. Dr. Leonardo Mesquita

Guaratinguetá - SP
2023

C128a	Caetano, Isabela Mansur Análise de dados para correlacionar o nível de satisfação do cliente com a qualidade de rede: um ensaio técnico / Isabela Mansur Caetano - Guaratinguetá, 2023. 66 f : il. Bibliografia: f. 64-66 Trabalho de Graduação em Engenharia Elétrica – Universidade Estadual Paulista, Faculdade de Engenharia e Ciências de Guaratinguetá, 2023. Orientador: Prof. Dr. Leonardo Mesquita 1. Sistemas de comunicação móvel. 2. Sistemas de comunicação sem fio. 3. Satisfação do consumidor. I. Título.
	CDU 621.395

ISABELA MANSUR CAETANO

ESTE TRABALHO DE GRADUAÇÃO FOI JULGADO ADEQUADO COMO
PARTE DO REQUISITO PARA A OBTENÇÃO DO DIPLOMA DE
“GRADUADO EM ENGENHARIA ELÉTRICA”

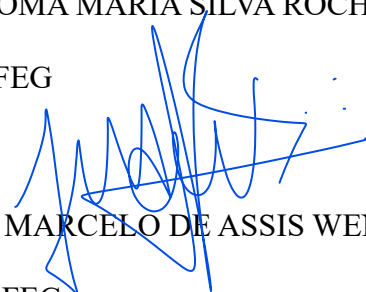
APROVADO EM SUA FORMA FINAL PELO CONSELHO DE CURSO DE
GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Prof. Dr. DANIEL JULIEN BARROS DA SILVA SAMPAIO
Coordenador

BANCA EXAMINADORA:


Prof. Dr. LEONARDO MESQUITA
Orientador/UNESP-FEG

Prof.^a Dra. PALOMA MARIA SILVA ROCHA
RIZOL
UNESP-FEG


Prof. Me. JOSÉ MARCELO DE ASSIS WENDLING
JÚNIOR
UNESP-FEG

DADOS CURRICULARES

ISABELA MANSUR CAETANO

NASCIMENTO	03-06-1999 – Volta Redonda / R.J.
FILIAÇÃO	Carlos Gleison Caetano Barbosa Roberta Mansur Caetano
2017/2023	Formação acadêmica (Engenharia Elétrica) UNESP - Guaratinguetá

AGRADECIMENTOS

Desde o ingresso a esse curso até a conclusão desse trabalho, tenho de agradecer a minha família. Aos meus pais *Roberta Mansur Caetano* e *Carlos Gleison Caetano Barbosa*, agradeço por todo o suporte que me deram nos momentos mais difíceis dessa graduação, que independentemente do quão longe estivessem de mim, sempre se mantiveram presentes. Sem o apoio de vocês, eu jamais teria chegado tão longe;

Agradeço ao meu irmão, *Ciro Mansur Caetano*, e à minha prima, *Mariana Mansur Barroso*, por sempre terem sido exemplos na minha vida pessoal e acadêmica;

Agradeço a minha professora e tutora de intercâmbio, *Professora Dra. Paloma Rizol*, por todo incentivo e ajuda nos meus anos na França;

Agradeço ao meu professor e orientador, *Professor Dr. Leonardo Mesquita*, por toda ajuda nesse trabalho e, anos antes, por despertar minha curiosidade a esse setor da engenharia;

Agradeço aos meus colegas de trabalho *Dyego Fonseca* e *Vinicius Santos*, por estarem sempre dispostos a tirar minhas dúvidas para o desenvolvimento desse estudo, ajudando-me a enxergar minhas próprias capacidades em momentos que eu mesma duvidei;

Agradeço ao meu diretor *Fernando Rodini*, por todos os conhecimentos que me passou e por acreditar no meu trabalho;

Agradeço aos professores da Faculdade de Engenharia do Campus de Guaratinguetá, pela dedicação em promover o conhecimento. E aos funcionários, em criar um ambiente acolhedor.

“O sol estava quase nascendo, aqueles poucos minutos em que a noite e o dia são uma coisa só. E eu percebi que nunca tinha dito de verdade “eu te amo”. Então eu resolvi começar.”

Rafael Gomes

RESUMO

As comunicações móveis sem fio foram revolucionárias para a sociedade contemporânea, elas evoluíram rapidamente nas últimas décadas e hoje fornecem diversos serviços amplamente difundidos. A garantia da qualidade desses serviços é verificada constantemente pelas operadoras telefônicas por meio de medições realizadas periodicamente pelos dispositivos celulares. Tem-se como objetivo deste trabalho, propor estratégias para permanência de usuários por meio de análises de dados. A partir das informações fornecidos pela operadora sobre condições de rádio e pesquisa de satisfação dos usuários, foram realizadas análises de forma a delimitar as condições de rede geradoras de insatisfação. Esse padrão seria aplicado para os demais usuários da operadora de forma a identificar usuários de alto poder aquisitivo com grandes possibilidades de deixar a operadora. Buscava-se assim, criar estratégias de ação que garantissem a permanência desses clientes.

PALAVRAS-CHAVE: Comunicações móveis; análise de dados; perfil de satisfação.

ABSTRACT

Wireless mobile communications were revolutionary for contemporary society, they have evolved rapidly in recent decades and today provide a wide range of services. The guarantee of the quality of these services is constantly verified by the telephone operators through measurements carried out periodically by the cellular devices. The objective of this work is to propose strategies for the permanence of users through data analysis. Based on the information provided by the operator on radio conditions and the user satisfaction survey, analyzes were carried out to delimit the network conditions that generate dissatisfaction. This standard would be applied to the operator's other users to identify users with high purchasing power who are likely to leave the operator. Thus, the aim was to create action strategies that would guarantee the permanence of these clients.

KEYWORDS: Mobile communication; data analysis; satisfaction profiling.

LISTA DE ILUSTRAÇÕES

Figura 1 - Crescimento de aparelhos móveis em relação a telefones residenciais e o aumento paralelo do uso da internet no período de 1998 a 2009	17
Figura 2 – Diagrama temporal para a acusação de <i>Radio Link Failure</i>	20
Figura 3 – Imagem explicativa da transição entre <i>idle</i> e <i>connected mode</i>	20
Figura 4 – Fluxograma das etapas do projeto	23
Figura 5 – Diagrama com <i>inputs</i> e <i>outputs</i> da etapa 1. Os dados utilizados nessa etapa foram estabelecidos	24
Figura 6 – Diagrama com os bancos de dados utilizados na etapa 2	26
Figura 7 – Diagrama com os dados obtidos na tabela de satisfação que geraram um dos <i>dataframe output</i> da etapa 2	26
Figura 8 – Diagrama com os dados obtidos na tabela identificadora que geraram um dos <i>dataframe output</i> da etapa 2	27
Figura 9 – Diagrama com os dados obtidos nas tabelas Arieso que geraram um dos <i>dataframe output</i> da etapa 2	28
Figura 10 – Diagrama com os dados obtidos inicialmente na tabela Arieso que geraram um dos <i>dataframe output</i> da etapa 2	28
Figura 11 – Diagrama com os dados obtidos inicialmente na tabela Arieso que geraram um dos <i>dataframe output</i> da etapa 2	29
Figura 12 – Diagrama com os dados obtidos na tabela Arieso 3G que geraram um dos <i>dataframe output</i> da etapa 2	29
Figura 13 – Imagem explicativa <i>K-means clustering</i>	30
Figura 14 – Imagem explicativa sobre <i>Elbow Technique</i> na determinação do número ideal de clusters	31
Figura 15 – Imagens geradas em python com marcadores de intensidade de sinal sobre a região do Chile em diferentes escalas de zoom	34
Figura 16 – Imagem gerada em python por meio de base de dados corrigidas exibindo o posicionamento correto dos pares de coordenadas	35
Figura 17 – Gráfico para análise comportamental dos diferentes níveis de satisfação para diversos parâmetros analisados dois a dois	36
Figura 18 – Gráfico com linhas comportamentais para os diferentes níveis de satisfação em termos de RSRP e RSRQ	37

Figura 19 – Gráficos de contadores de usuários por cenários de RSRP para cada um dos níveis de satisfação	38
Figura 20 – <i>Output</i> obtido em python com a quantidade de usuários em cada um dos níveis de satisfação	39
Figura 21 - Gráficos com densidade de probabilidade para diferentes cenários de RSRP em cada um dos níveis de satisfação	40
Figura 22 - Gráfico com a densidade de probabilidade para diferentes cenários de RSRP para a amostra completa de usuários	41
Figura 23 – Gráfico de contadores de usuários para cenários de RSRP utilizando a amostra completa com identificação de cor por nível de satisfação	42
Figura 24 – Gráfico de contadores de usuários para diferentes taxas de <i>downlink throughput</i> utilizando a amostra completa com identificação de cor por nível de satisfação	42
Figura 25 – Gráfico de contadores de usuários para cenários de RSRQ utilizando a amostra completa com identificação de cor por nível de satisfação	43
Figura 26 – Gráfico de contadores de usuários para diferentes valores de <i>releases</i> utilizando a amostra completa com identificação de cor por nível de satisfação	44
Figura 27 – Gráfico com linhas comportamentais para os diferentes níveis de satisfação em termos de RSRP e <i>releases</i>	44
Figura 28 - Gráficos com densidade de probabilidade para diferentes valores de <i>releases</i> em cada um dos níveis de satisfação	45
Figura 29 - Gráficos com densidade de probabilidade para diferentes valores de <i>drops</i> em cada um dos níveis de satisfação	46
Figura 30 – Amostra do <i>dataframe</i> usado para a clusterização com os parâmetros considerados pelo sistema	46
Figura 31 – Gráfico gerado por meio do <i>Elbow Technique</i> para definir o número de agrupamentos do sistema	47
Figura 32 - Gráfico para análise comportamental dos diferentes agrupamentos realizados pelo <i>k-means clustering</i> para diversos parâmetros analisados dois a dois	48
Figura 33 – Gráfico gerado por meio do <i>Elbow Technique</i> para definir o número de agrupamentos do sistema	49
Figura 34 – Divisão computadorizada feita do sistema para volumes de <i>downlink</i> e <i>uplink</i>	49
Figura 35 – Posicionamento gráfico dos grupos de mais baixa satisfação na pesquisa em relação aos volumes <i>downlink</i> e <i>uplink</i>	50

Figura 36 – Posicionamento gráfico dos grupos de mais alta satisfação na pesquisa em relação aos volumes <i>downlink</i> e <i>uplink</i>	51
Figura 37 – Fragmento do código exibindo a quantidade total de linhas do <i>dataframe</i> no <i>output</i>	51
Figura 38 – Fragmento do código exibindo a quantidade de linhas do <i>dataframe</i> para o cluster 1 no <i>output</i>	52
Figura 39 – Gráfico gerado por meio do <i>Elbow Technique</i> para definir o número de agrupamentos do sistema	53
Figura 40 - Divisão computadorizada feita do sistema para <i>downlink throughput</i> e minutos de uso	53
Figura 41 – Fragmento do código exibindo a quantidade de linhas do <i>dataframe</i> para o cluster 2 no <i>output</i>	54
Figura 42 – Posicionamento gráfico dos grupos de mais baixa satisfação na pesquisa em relação ao <i>downlink throughput</i> e minutos de uso	55
Figura 43 – Fragmento do código exibindo a criação do <i>dataframe</i> com os clientes de maior insatisfação (níveis 1, 2, 3 e 4)	55
Figura 44 – Fragmento do código exibindo a quantidade total de usuários no grande grupo dos insatisfeitos	56
Figura 45 – Fragmento do código exibindo a quantidade de usuários no grande grupo dos insatisfeitos que foram classificados como pertencentes ao agrupamento 2	56
Figura 46 – Gráfico gerado por meio do <i>Elbow Technique</i> para definir o número de agrupamentos do sistema	56
Figura 47 - Divisão computadorizada feita do sistema para <i>downlink throughput</i> , minutos de uso e RSRP	57
Figura 48 – Fragmento do código exibindo a quantidade de usuários do sistema que foram classificados como pertencentes ao agrupamento 1	58
Figura 49 – Posicionamento gráfico dos grupos de mais baixa satisfação na pesquisa em relação ao <i>downlink throughput</i> , minutos de uso e RSRP	59
Figura 50 – Fragmento do código exibindo a quantidade total de usuários no grande grupo dos insatisfeitos	60
Figura 51 – Fragmento do código exibindo a quantidade de usuários no grande grupo dos insatisfeitos que foram classificados como pertencentes ao agrupamento 1	60
Figura 52 - Gráfico de contadores de usuários por níveis de satisfação para todos os usuários que estiveram conectados no 3G	61

Figura 53 - Gráfico de contadores de usuários por níveis de satisfação para todos os usuários que estiveram conectados tanto no 3G quanto no 4G 61

LISTA DE TABELAS

Tabela 1 - Código de cores utilizado pela operadora, implementado no trabalho para a ilustração das condições de sinal no mapa.....	19
Tabela 2 – Condições das células de acordo com RSRP e RSRQ	33

LISTA DE ABREVIATURAS E SIGLAS

3GPP	3rd Generation Partnership Project
HTML	HyperText Markup Language
KPI	Key Performance Indicator
LTE	Long Term Evolution
QoS	Quality of service
RLF	Radio Link Failure
RRC	Radio Resource Control
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Reference Quality
UE	User Equipment
UId _{operadora}	User Identification
UTM	Universal Transverse Mercator
SMS	Short Message Service

LISTA DE SÍMBOLOS

dB	decibel
dBm	decibel milliwatts
Mbps	megabits per second
W	watts

SUMÁRIO

1	INTRODUÇÃO	16
1.1	OBJETIVOS DO TRABALHO	16
2	REVISÃO DA LITERATURA	17
2.1	DEFINIÇÃO DOS KPIS	18
3	MATERIAIS E MÉTODOS	21
3.1	MATERIAIS	21
3.2	MÉTODOS	22
3.2.1	Etapa 1: Agrupar as coordenadas em bins.....	23
3.2.2	Etapa 2: Relacionar a pesquisa de satisfação com os KPIS dos usuários	25
3.2.3	Etapa 3: Criação dos perfis de satisfação	30
3.2.4	Etapa 4: Definir possíveis bins geradora de descontentamento	31
3.2.5	Etapa 5: Definir usuários de alto perfil e seus KPIS.....	32
4	RESULTADOS E DISCUSSÕES	34
5	CONSIDERAÇÕES FINAIS	65
	REFERÊNCIAS	66
	BIBLIOGRAFIA CONSULTADA	68

1 INTRODUÇÃO

A comunicação móvel sem fio surgiu há algumas décadas, mas em tão pouco tempo já tomou um papel muito importante na vida de seus usuários. A cada nova geração dos sistemas móveis de comunicação, inovações foram implementadas. Todas essas inovações fazem parte, hoje, da dinâmica social. Dentre os serviços oferecidos estão chamadas de voz, mensagens de texto, uso da banda larga e transferência de dados via internet.

Para garantir o bom funcionamento desses serviços, medições internas dos aparelhos celulares são constantemente verificadas pelas operadoras, porém essa riqueza de informações disponíveis sobre as condições da rede acaba por gerar um enorme volume de dados a serem armazenados, os quais muitas vezes não são utilizados.

Em média, 52% dos dados armazenados por empresas não são utilizados. Muito além dos impactos financeiros pela compra de armazenamento, essa centena de zetabytes inutilizados gera um grande impacto no meio ambiente. Cerca de 5,8 milhões de toneladas de dióxido de carbono são lançados na atmosfera por ano como resultado da energia utilizada pelos sistemas de armazenamento dessa *dark data* (Ramos, 2018; Tan, 2020).

Dessa forma, tornou-se imprescindível a identificação dos dados úteis e seu tratamento. Uma vez tratados, esses dados podem se transformar em informações que auxiliam em ações de negócios.

As empresas de telefonia hoje fazem parte desta realidade. Elas detêm uma grande riqueza de informações que não são exploradas, as quais, caso analisadas, podem trazer tanto benefícios para a própria empresa quanto para o consumidor final.

1.1 Objetivos do trabalho

Este projeto tem como objetivo principal a análise e cruzamento de dados de diversas fontes para decisão estratégica de planos de ação de engenharia. Mais especificamente, busca-se realizar análise da base de dados fornecida pela operadora telefônica, traçar mapas de insatisfação dos clientes e definir cenários de qualidade de rede que levam um cliente a mudar de operadora. Assim, definir priorização de investimentos e atividades de expansão para maior impacto na percepção do cliente.

2 REVISÃO DA LITERATURA

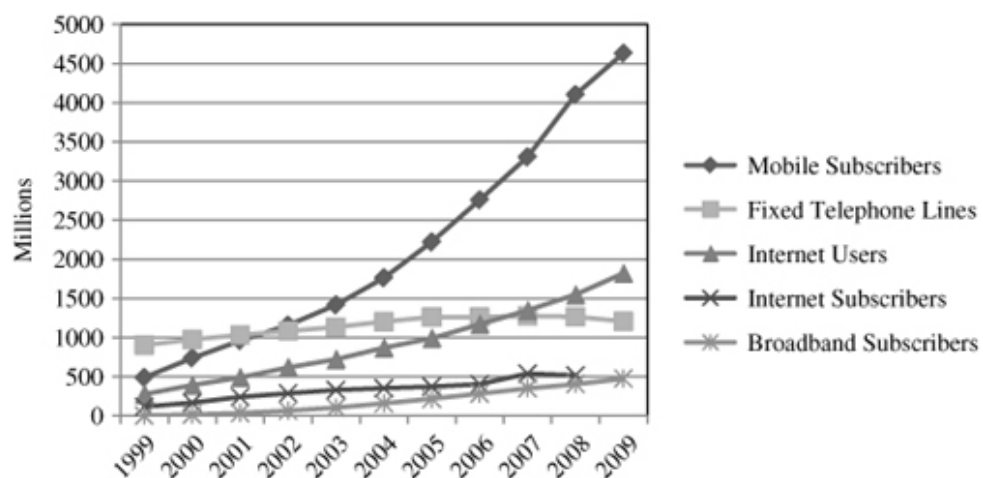
As tecnologias de comunicações móveis sem fio começaram a surgir no início dos anos 1980. A história dessas tecnologias é marcada por avanços, o surgimento de uma nova geração dos sistemas de comunicação móvel, que ocorrem com intervalos de cerca de 10 anos (Cox, 2021).

Sua primeira geração 1G, possibilitou o tráfego de voz. Esse acesso era realizado por meio de celulares analógicos. Já a segunda geração 2G, teve seu início no fim da década de 1980. Sua grande inovação foi a utilização de modulação digital, o que ampliou a capacidade e qualidade do som. Além disso, novos serviços como o SMS (*short message service*) se tornaram realidade (Agiwal; Roy; Saxena, 2016; Stroski, 2017; Batalha et al., 2022).

No início dos anos 2000, a terceira geração 3G de comunicações móveis já estava sendo lançada mundialmente. Ela permitiu transferências de dados por meio da banda larga. Assim, novos serviços se tornaram possíveis junto ao acesso à Internet (Batalha et al., 2022). Essa nova tecnologia permitia a transferência de dados em uma velocidade que alcançava 2 Mbps (megabits por segundo) (Stroski, 2017).

O gráfico abaixo (Figura 1) mostra o crescimento da comunicação sem fio em comparação com os telefones residenciais dos anos de 1998 a 2009, além do aumento do uso da internet (Ghosh et al., 2010).

Figura 1 - Crescimento de aparelhos móveis em relação a telefones residenciais e o aumento paralelo do uso da internet no período de 1998 a 2009



Fonte: Ghosh et al. (2010)

O 4G, quarta geração do sistema de comunicação móvel, foi desenvolvido no fim dos anos 2000. Essa nova tecnologia também é conhecida como LTE (*long term evolution*) pois deu continuidade aos planos criados no 3GPP (*3rd Generation Partnership Project*), conjunto de associações de diversos países que definem normas e especificações para o sistema móvel. O 3GPP LTE permitiu ao usuário altas taxas de transferências de dados ao mesmo tempo que uma baixa latência (Furht, 2009). Essa tecnologia pode alcançar velocidades de até 300 Mbps (Braga, 2018).

Com os benefícios introduzidos às redes de comunicações móveis por meio do 4G, novas possibilidades de serviços online surgiram, novos negócios se viabilizaram. O que pode ser exemplificado ao observar o crescimento da utilização dessas redes: o tráfego de dados móveis cresceu cerca de 70% em 2014 (Agiwal; Roy; Saxena, 2016).

Hoje vive-se a transição da tecnologia 4G para 5G. A implementação dessa quinta geração de tecnologia no sistema de comunicação móvel é dada de forma desigual ao redor do mundo. Para sua implementação, grandes investimentos em infraestrutura são necessários. Assim, países desenvolvidos são líderes nessas mudanças, garantindo seu pioneirismo no desenvolvimento de tecnologias inovadoras (Grijpink et al., 2020; Crandall, 2021).

Atualmente, o acesso a redes de dados móveis é amplamente difundido. Estima-se que no fim de 2021, ao redor do mundo, cerca de 8.65 bilhões de números telefônicos tinham acesso à internet (Statista, 2022).

Para garantir a qualidade dos serviços de rede oferecidos pela operadora aos seus usuários, utiliza-se nesse trabalho análises de medições internas da rede realizadas pelos dispositivos celulares. Dentre as medidas reportadas pelo UE (*user equipment*) sobre a rede LTE, serão dadas aqui as definições de RSRP e RSRQ (Afroz et al., 2015).

2.1 Definição dos KPIs

O *Reference Signal Received Power* (RSRP) é uma medição realizada pelo UE sobre a intensidade do sinal de uma determinada célula. Por meio deste parâmetro é estabelecida qual a melhor célula para acesso ou decisões de transferência de célula, *handover*. O RSRP geralmente é expresso em dBm e compreende valores de -140 a -44 dBm com intervalos de -1 dBm (Afroz et al., 2015; Rodrigues, 2012).

Pela perspectiva dos usuários, o RSRP está relacionado também com o QoS (*Quality of service*), porém os intervalos de RSRP não estão bem definidos em relação a qualidade da experiência do usuário (Rodrigues, 2012).

Assim como o RSRP, o *Reference Signal Reference Quality* (RSRQ) também é uma medição realizada pelo UE com o papel principal de auxiliar na determinação da melhor célula para acesso, seja nos procedimentos de *handover* quando o UE está em modo *connected*, seja para nova seleção da célula quando o UE está em modo inativo, *idle*. O RSRQ é uma métrica de qualidade de sinal de uma determinada célula e seus valores podem variar de -19.5 a -3 dB com intervalos de 0.5 dB (Rodrigues, 2012; Afroz et al., 2015).

A leitura combinada dos dois KPIs, *key performance indicator*, aqui apresentados pode indicar problemas na cobertura ou de interferência de sinal, sendo ambos, portanto, bons indicadores da qualidade do serviço (Rodrigues, 2012). Como exemplificado na Tabela 1.

Tabela 1 – Condições das células de acordo com RSRP e RSRQ

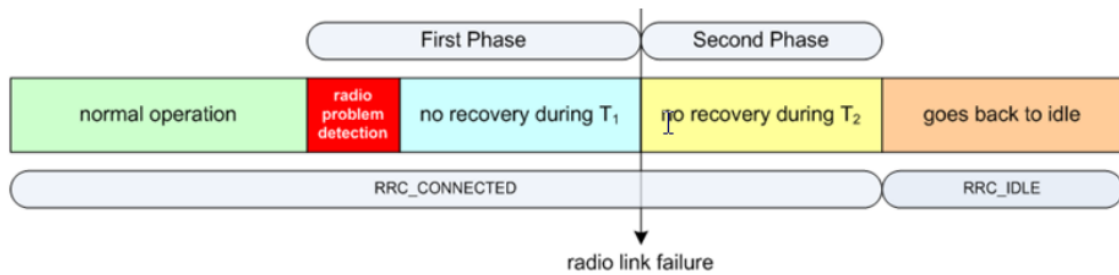
<i>RF Quality</i>	<i>RSRP(dBm)</i>	<i>RSRQ(dB)</i>
Excellent	≥ -80	≥ -10
Good	-80 to -90	-10 to -15
‘Mid Cell’	-90 to -100	-15 to -20
‘Cell Edge’	< -100	< -20

Fonte: GOOGLE (2023)

Boas condições de RSRP e RSRQ contribuem para o aumento do *throughput*, sendo essa uma outra métrica de QoS fortemente percebida pelos usuários. O *throughput* é a taxa de transmissão de dados, ela diz respeito ao envio de dados do UE ao site no caso do *uplink*, ou recebimento de dados do site pelo dispositivo, *downlink* (Silva; Chamon, 2004; Yue et al., 2018; Batalha et al., 2022). Dadas essas definições, o volume de *throughput* se refere a quantidade de dados enviados nessas transmissões.

Outro parâmetro utilizado neste estudo foi o número de *drops*. Os *drops* estão relacionados com o *Radio Link Failure* (RLF), problema que ocorre quando o UE perde a conexão com o site e apresenta problemas para restabelecê-la dentro de um limite de tempo pré-determinado, detalhado na Figura 2. O RLF é um grande desafio para a comunicação móvel pois além de aumentar a latência, ele diminui a confiabilidade da rede, o que impacta diretamente o usuário (Boutiba; Bagaa; Ksentini, 2021).

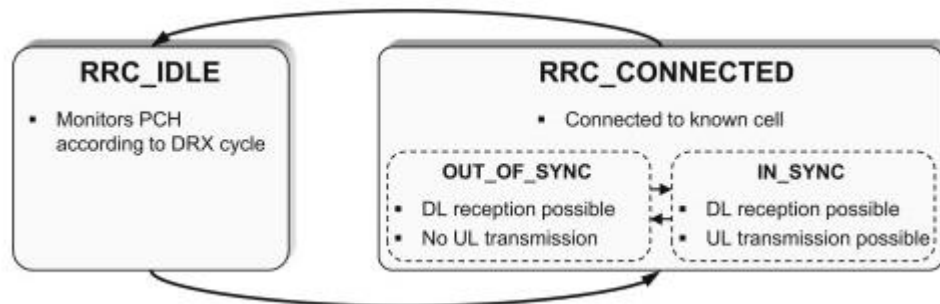
Figura 2 – Diagrama temporal para a acusação de *Radio Link Failure*



Fonte: Puttonen et al. (2010)

De acordo com as medições da rede realizadas pelo UE, o dispositivo pode sair de seu modo conectado e passar ao seu modo inativo, *idle*, onde não há transferência de dados. Apesar disso, o UE se mantém acampado na célula, sendo necessário apenas restabelecer o controle de conexão. Nessa passagem para o modo *idle*, ocorre a liberação da conexão ao RRC (*Radio Resource Control*), *release*. Este foi mais um parâmetro usado no estudo (Agrawal1, 2016). A seguir, uma imagem explicativa da transição para *idle* Figura 3.

Figura 3 – Imagem explicativa da transição entre *idle* e *connected mode*



Fonte: Perez (2017)

3 MATERIAIS E MÉTODOS

3.1 Materiais

Para o desenvolvimento desse estudo contou-se com dados fornecidos por uma empresa que atua no setor de telecomunicação. Esses dados são referentes a usuários no território chileno. É essencial neste trabalho que a identidade dos usuários se mantenha anônima. Sendo assim, para manter a confidencialidade, foram aqui apresentados apenas as análises das condições de rede dos usuários enquanto seus dados identificadores foram omitidos.

A amostra fornecida para a realização deste estudo conta com três diferentes formatos: a tabela de satisfação dos usuários, a tabela com informações de rádio da rede e a tabela identificadora. Esses três diferentes formatos de tabelas serão explicados em detalhes a seguir.

Vale salientar que essa é apenas uma amostra representativa dentro de toda riqueza de dados da operadora. A amostra não é capaz de entregar a real vivência dos usuários, mas pode retratar alguns cenários aos quais os clientes foram expostos, permitindo algumas análises dos serviços prestados. Todos os dados aqui citados foram manipulados dentro do próprio servidor da operadora, não podendo ser acessados localmente na máquina.

A tabela de satisfação dos usuários é baseada em uma pesquisa de satisfação realizada com os clientes da operadora. Essa pesquisa é realizada mensalmente. A amostra fornecida conta com uma tabela referente a uma pesquisa de satisfação conduzida em julho de 2022.

Nesta pesquisa foram entrevistados 755 usuários, cada um deles representa uma linha da tabela. Ela conta também com 101 colunas, com as seguintes informações:

- dia, horário e local da realização da pesquisa;
- dados dos usuários como idade, gênero e cidade de residência;
- satisfação sobre o atendimento nos *call-centers*;
- satisfação sobre o funcionamento da operadora;
- indicação dos usuários por meio de identificadores que podem ser correlacionados entre fontes de informação;
- tipo de conta (pré-pago, controle ou pós-pago), status do serviço e fidelidade

A maioria das perguntas de satisfação são objetivas, entretanto, algumas são discursivas, como por exemplo os motivos para recomendar a operadora. Essas perguntas objetivas sobre a satisfação do cliente quanto ao funcionamento da operadora que foram levadas em consideração para o estudo.

As tabelas com informações de rádio da rede fornecem dados sobre redes 3G e 4G, sendo as tabelas de ambas as tecnologias bem similares. A fonte dessas tabelas é a Arieso, empresa que fornece dados de geolocalização para otimização de redes de operadoras, porém houve um pré-processamento desses dados por parte da própria operadora, reduzindo seu volume.

As medições de rede encontradas nessas tabelas datam de 23 a 25 de setembro de 2022. A amostra apresenta uma tabela por hora para cada um desses três dias, com exceção das 23h, totalizando, portanto, 23 tabelas por dia. Cada uma dessas tabelas apresenta medições realizadas em quatro diferentes horários (hh:00, hh:15, hh:30, hh:45), visto que as medições consideram um intervalo de 15 minutos.

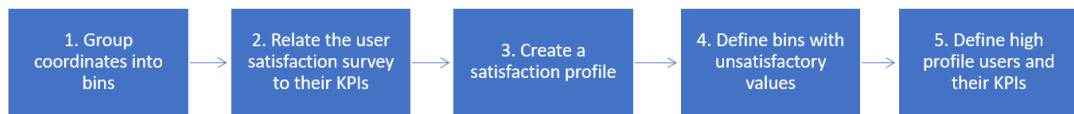
Dentre os diversos dados encontrados nessas tabelas, tem-se: dados de *handover*, minutos de uso *indoor* e *outdoor*, volume de *throughput uplink*, volume de *throughput downlink*, média de *throughput uplink*, média de *throughput downlink*, minutos de uso, minutos de uso de voz, minutos de uso estacionário, minutos de uso andando, minutos de uso em um veículo, total de *drops* na conexão, total *releases* na conexão, média de RSRP e RSRQ, coordenadas de onde o dispositivo estava sendo usado em UTM e identificadores internos dos usuários.

Por fim, uma última tabela foi fornecida na amostra. A tabela identificadora informa todos os usuários da operadora, apresentando, portanto, mais de 15 milhões de usuários. Teve-se acesso neste estudo a uma versão reduzida desta tabela que conta apenas com os usuários presentes na pesquisa de satisfação. Dentre suas 34 colunas com diferentes informações, encontra-se os dados de identificação interna usados pela operadora, que será chamado aqui de $UId_{operadora}$ (*user identification*), e os identificadores usados na pesquisa de satisfação, permitindo assim que as duas tabelas citadas anteriormente se relacionem.

3.2 Métodos

O método escolhido para o desenvolvimento deste projeto levou em consideração que as tabelas detalhadas anteriormente seriam os *inputs* do estudo. A partir desses *inputs*, foi criado um plano de cinco etapas partindo desde a limpeza e estruturação dos dados até a criação de uma lista de usuários de alto perfil com grande probabilidade de trocar de operadora. Essas etapas foram melhor descritas na Figura 4.

Figura 4 – Fluxograma das etapas do projeto



Fonte: Autoria própria

Todas as etapas foram realizadas por meio do software de código aberto Anaconda. Foi escolhida a linguagem Python. O subaplicativo Jupyter foi utilizado como ambiente de codificação. Em seu notebook foram inseridas diversas bibliotecas como: pandas e numpy para manipulação da base de dados e cálculos científicos, matplotlib e seaborn para a criação de representações gráficas, math para o cálculo de logaritmos, sklearn.cluster e sklearn.preprocessing para o agrupamento em clusters, openpyxl para manipulações de tabela em excel, folium para a interação com mapas, banca.colormap para a definições de cores, utm para a conversão dos pares de coordenadas e zipfile para a leitura de arquivos compactados.

3.2.1 Etapa 1: Agrupar as coordenadas em bins

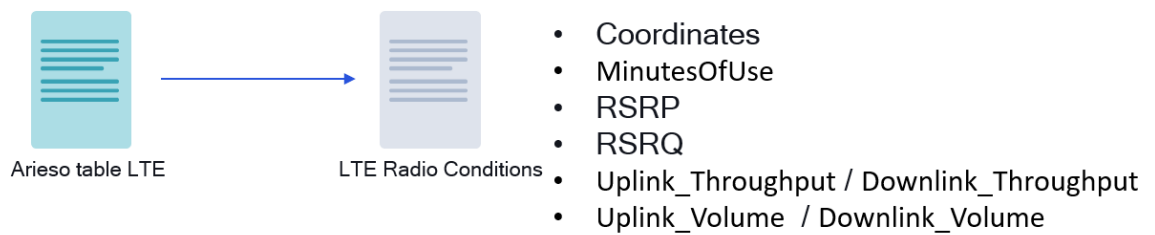
A primeira etapa do estudo busca dividir o território chileno em áreas de cerca de 50x50m, este valor foi escolhido pois permite observar as condições de rádio dos sites sem grande perda de precisão. Utilizando os arquivos com dados da rede, para cada uma dessas áreas foram encontrados os valores de KPIs que posteriormente foram plotados sobre o mapa do Chile seguindo um código de cores.

Assim, nesta primeira etapa, um novo *dataframe* foi criado. Cada novo par de coordenadas encontrado nos arquivos originais foi adicionado ao novo *dataframe*. Uma vez encontrado o mesmo par de coordenadas uma segunda vez, uma média ponderada dos KPIs foi realizada.

Usando essa lógica, os valores de KPIs de uma região são dados pelas experiências que diferentes usuários em diferentes momentos foram expostos nesse mesmo site. Estes diferentes cenários devem ser levados em consideração com pesos diferentes. Estes pesos devem considerar as atividades realizadas por cada um desses usuários, para assim criarmos uma representação mais condizente de cada uma dessas áreas.

O novo *dataframe* conta com as seguintes informações advindas das tabelas originais: par de coordenadas, minutos de uso, RSRP médio, RSRQ médio, *uplink throughput*, *downlink throughput*, volume de *uplink throughput*, volume de *downlink throughput*. O que pode ser observado no seguinte diagrama Figura 5.

Figura 5 – Diagrama com *inputs* e *outputs* da etapa 1. Os dados utilizados nessa etapa foram estabelecidos



Fonte: Autoria própria

O cálculo da média ponderada dos KPIs, no caso dos valores de RSRP e RSRQ, foi realizado em função dos minutos de uso, já no caso do *uplink throughput* e *downlink throughput* em função dos seus respectivos volumes. Vale lembrar que os valores de RSRP e RSRQ presentes nas tabelas originais se encontram em dBm e dB, sendo necessária a conversão para valores lineares previamente ao cálculo.

Os cálculos foram realizados da seguinte maneira. Inicialmente os valores de RSRP foram convertidos de dBm para W usando a expressão abaixo.

$$RSRP_W = 10^{\frac{RSRP_{dBm}}{10}} / 1000$$

Com todos os valores de RSRP convertidos, a média ponderada foi realizada em função dos minutos de uso.

$$RSRP_{Wtotal} = \frac{(RSRP_{Wtotal} * MinOfUse_{total}) + (RSRP_{Witem} * MinOfUse_{item})}{MinOfUse_{total} + MinOfUse_{item}}$$

Nessa expressão, $RSRP_{item}$ diz respeito aos novos valores encontrados um a um nas tabelas originais. Esses valores são acrescentados ao novo *dataframe* por meio da média ponderada. Os valores encontrados, portanto, no novo *dataframe* $RSRP_{total}$ dizem respeito a um conjunto de valores já lido pelo código.

Ao fim do cálculo, os valores de RSRP devem ser retornados a dBm, unidade essa tipicamente usada em análises de telecomunicações.

$$RSRP_{dBm} = 10 * \log_{10}(RSRP_w * 1000)$$

Pode-se afirmar ainda que o efeito prático dessas médias ponderadas é descrito pela seguinte fórmula.

$$RSRP = \frac{\sum(RSRP * MinOfUse)}{\sum MinOfUse}$$

Cálculo semelhante foi realizado com os valores de RSRQ.

Já as operações matemáticas utilizando os dados de *throughput*, podem ser exemplificadas da forma apresentada abaixo. Nelas, os pesos usados são os volumes de dados enviado ou recebidos entre o UE e a rede.

$$tput_{total} = \frac{(tput_{total} * vol_{total}) + (tput_{item} * vol_{item})}{vol_{total} + vol_{item}}$$

Ao fim de cada um dos cálculos de média ponderada, os valores de minutos de uso e volume de *throughput* são também atualizados no novo *dataframe*. Os valores de cada item são somados aos valores totais.

Uma vez encontrados esses KPIs para cada um dos pares de coordenadas, esses dados foram plotados sobre o mapa do Chile seguindo o código de cores da própria operadora quanto ao RSRP. Isso nos permite encontrar regiões com uma potência de sinal mais fraca.

3.2.2 Etapa 2: Relacionar a pesquisa de satisfação com os KPIs dos usuários

A segunda etapa do projeto consistiu em relacionar os usuários da pesquisa de satisfação às condições de rádio a qual eles foram expostos. Todos os três formatos de tabelas encontrados na amostra foram usados.

Inicialmente extraiu-se os dados das tabelas que foram considerados pertinentes ao estudo, em seguida eles foram relacionados. Assim, foram encontrados os valores de KPIs médios que cada um dos usuários da tabela de satisfação foi exposto além, claro, das notas de qualidade da operadora fornecidas pelos usuários na pesquisa. Novamente o conceito de média ponderada mostrado anteriormente foi utilizado.

Vale salientar que só foi possível relacionar todos os dados por meio da tabela identificadora, como pode ser observado na imagem abaixo Figura 6.

Figura 6 – Diagrama com os bancos de dados utilizados na etapa 2



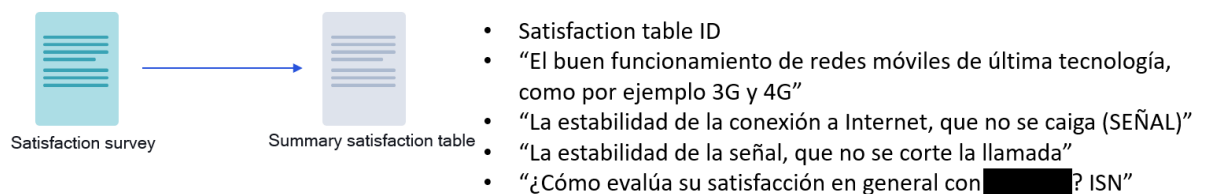
Fonte: Autoria própria

Da tabela de satisfação, os dados retirados foram o identificador dos usuários nela estabelecido sendo este o parâmetro utilizado para futuras correlações, no caso dessa fonte de informações, e as respostas de 0 a 10 dadas as seguintes perguntas:

- Sobre o funcionamento das redes móveis de última tecnologia, como por exemplo o 3G e 4G (*“El buen funcionamiento de redes móviles de última tecnología, como por ejemplo 3G y 4G”*)
- A estabilidade da conexão à internet, se a conexão cai, referente ao sinal (*“La estabilidad de la conexión a Internet, que no se caiga (SEÑAL)”*)
- A estabilidade do sinal, se há cortes nas chamadas (*“La estabilidad de la señal, que no se corte la llamada”*)
- Como você avalia sua satisfação em geral com a operadora (*“¿Cómo evalúa su satisfacción en general con ***** ? ISN”*)

As informações extraídas da tabela de satisfação são exibidas na Figura 7.

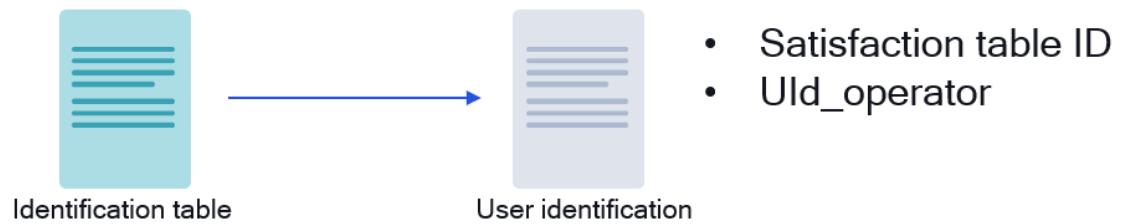
Figura 7 – Diagrama com os dados obtidos na tabela de satisfação que geraram um dos *dataframe output* da etapa 2



Fonte: Autoria própria

Da tabela identificadora, foram extraídos apenas os identificadores da tabela de satisfação e das tabelas de condições de rede, $UId_{operadora}$, como mostrado na Figura 8.

Figura 8 – Diagrama com os dados obtidos na tabela identificadora que geraram um dos *dataframe output* da etapa 2



Fonte: Autoria própria

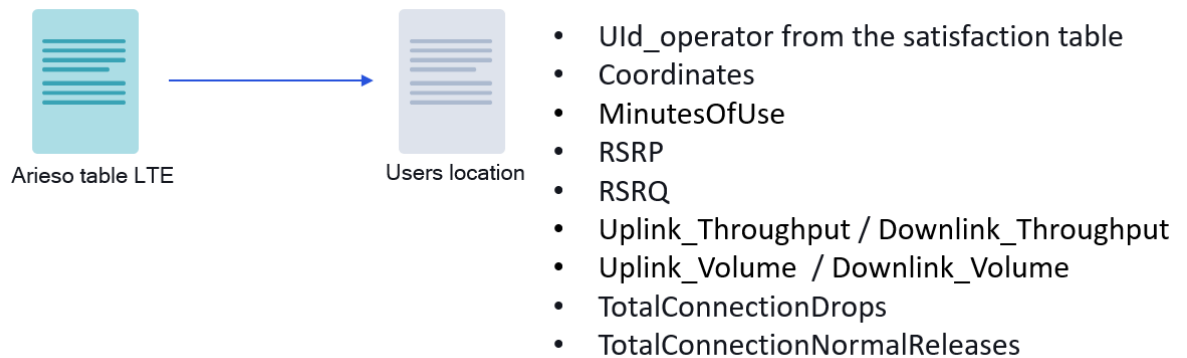
Por fim, foram retirados os dados das tabelas com informações de rede LTE. Nessa etapa, três novos *dataframes* foram criados. Todos eles apresentavam os mesmos parâmetros, porém com diferentes agrupamentos. Os parâmetros usados foram: $UId_{operadora}$ dos usuários presentes na tabela de satisfação, par de coordenadas, minutos de uso, RSRP médio, RSRQ médio, *uplink throughput*, *downlink throughput*, volume de *uplink throughput*, volume de *downlink throughput*, total de *drops* na conexão e total de *releases* na conexão. Os parâmetros extraídos podem ser vistos na Figura 9 abaixo.

O primeiro *dataframe* busca indicar todas as coordenadas que cada um dos usuários da tabela de satisfação esteve. Um mesmo usuário pode ser encontrado diversas vezes nesse banco de dados. Cada uma das linhas de mesmo $UId_{operadora}$ deve conter diferentes pares de coordenadas.

A lógica aqui usada considerou três diferentes cenários. Interagindo linha por linha com as tabelas Arieso previamente filtradas com apenas os usuários que responderam à pesquisa de satisfação:

1. Ao encontrar um novo $UId_{operadora}$ o programa deve adicioná-lo ao novo *dataframe* junto com os KPIs já presentes na tabela Arieso e as coordenadas do local onde estes usuários estavam ativos;
2. Ao encontrar um $UId_{operadora}$ já presente no novo *dataframe*, porém em uma nova coordenada, deve-se novamente adicionar os valores ao novo *dataframe*.
3. Ao encontrar um $UId_{operadora}$ já presente no novo *dataframe* em um par de coordenadas que este já esteve ativo anteriormente, os cálculos de média ponderada devem ser realizados aos valores de KPIs para atualização dos parâmetros do novo banco de dados.

Figura 9 – Diagrama com os dados obtidos nas tabelas Arieso que geraram um dos *dataframe output* da etapa 2

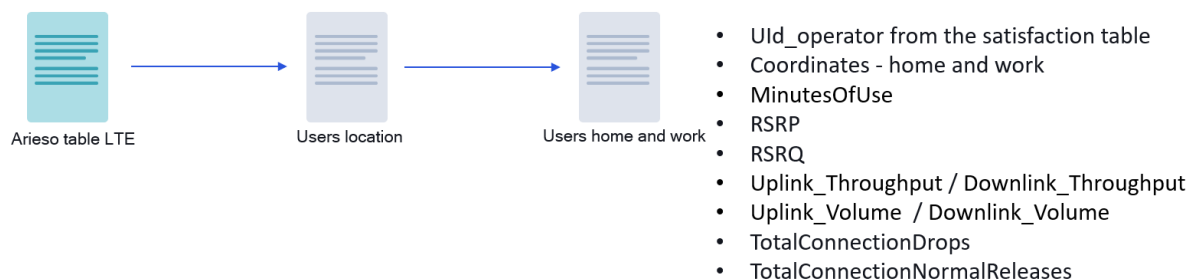


Fonte: Autoria própria

Uma vez esses novos dados sendo filtrados para cada um dos valores de Uid_{operadora} individualmente e ordenando-os em ordem decrescente de minutos de uso, encontra-se os locais onde cada usuário esteve ativo por maior período de tempo.

Para o segundo *dataframe* aqui gerado, foi considerado que os dois locais onde os usuários estiveram por mais tempo são referentes as suas casas e locais de trabalho. Sendo assim, no segundo *dataframe*, cada Uid_{operadora} pode ser encontrado duas vezes, sendo estes valores referentes aos dois diferentes locais onde o usuário passou mais tempo ativo. Os dados encontrados nesse *dataframe* podem ser visto na Figura 10.

Figura 10 – Diagrama com os dados obtidos inicialmente na tabela Arieso que geraram um dos *dataframe output* da etapa 2

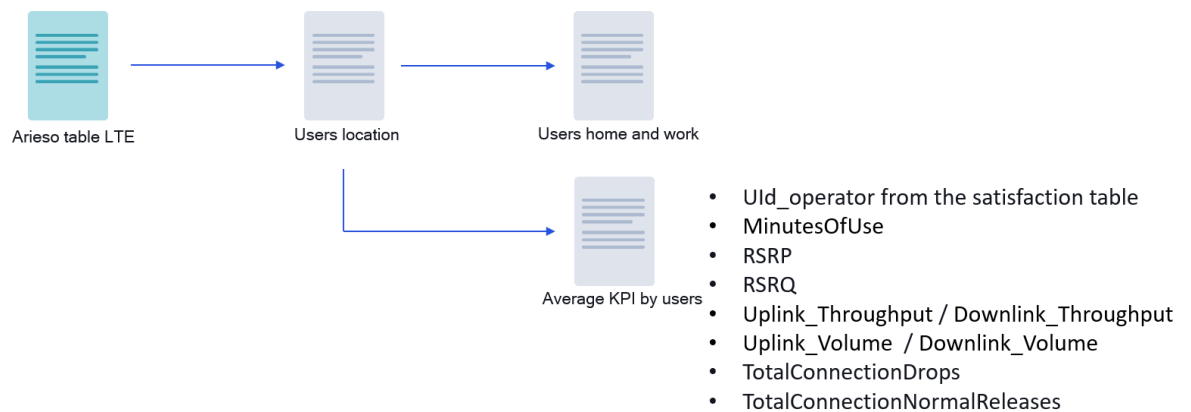


Fonte: Autoria própria

O último *dataframe* criado a partir deste banco de dados retoma o arquivo que chamamos aqui de *Users Location* e agrupa de maneira geral os KPIs a que os usuários foram expostos. Neste último arquivo, cada usuário é encontrado uma única vez na tabela. Médias ponderadas

são realizadas para agrupar os dados das diversas localizações. O diagrama deste último arquivo é mostrado na Figura 11.

Figura 11 – Diagrama com os dados obtidos inicialmente na tabela Arieso que geraram um dos *dataframe output* da etapa 2

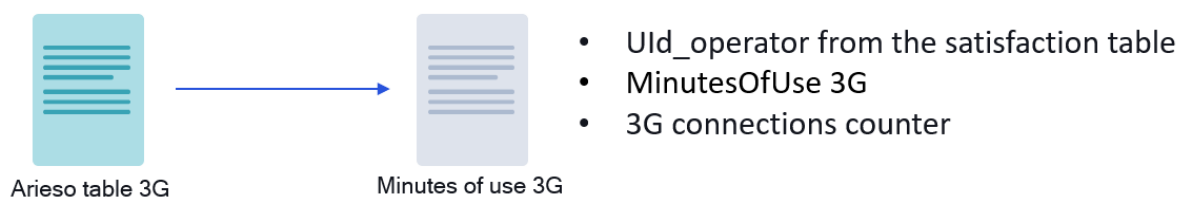


Fonte: Autoria própria

Essa diversidade de banco de dados gerados proporciona maior quantidade de análises. Ao analisar por exemplo a tabela *Users home and work*, está sendo assumida a hipótese de que a experiência do usuário nos locais onde ele permanece a maior parte do seu dia é predominante no seu nível de satisfação. Porém ao considerar o arquivo *Average KPI by users*, assume-se que problemas de sinal que ocorrem no período de deslocamento do usuário também influenciam em seu contentamento.

Os dados Arieso sobre a rede 3G também foram analisados, porém só se considerou interessante para o estudo dois parâmetros: os minutos de uso que os usuários presentes na tabela de satisfação estiveram conectados ao 3G e a quantidade de vezes que estes usuários realizaram a conexão ao 3G, como exibido no diagrama da Figura 12.

Figura 12 – Diagrama com os dados obtidos na tabela Arieso 3G que geraram um dos *dataframe output* da etapa 2



Fonte: Autoria própria

Uma vez que todos esses dados passaram pelo *data mining* e os parâmetros importantes para o estudo foram extraídos, inicia-se a terceira etapa do projeto.

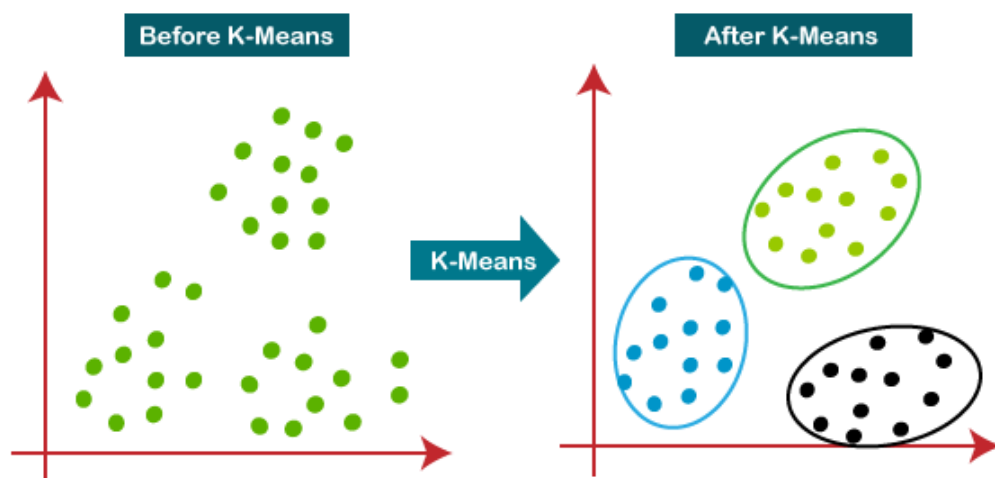
3.2.3 Etapa 3: Criação dos perfis de satisfação

Busca-se na terceira etapa a criação de perfis de satisfação. Aqui foram analisados os dados encontrados na etapa 2. Trabalhando com as diferentes bases de dados em conjunto, apura-se os limites dos KPIs da rede para os diferentes níveis de satisfação. Neste passo, procura-se encontrar um padrão de comportamento dos usuários, de forma que possa ser assumido, por exemplo, que usuários expostos a um determinado cenário tem grandes chances de estarem satisfeitos.

Esta etapa tenta responder qual o limiar dos valores de KPI que geram clientes satisfeitos e insatisfeitos. Para isso, foram feitas análises tanto para os níveis de satisfação de forma particular quanto para grupos generalizados. Análises gráficas foram realizadas assim como a aplicação do *k-means clustering*.

O *k-means clustering* é um método que consiste no posicionamento de centroides para a divisão dos grupos de acordo com a menor distância entre os pontos e esses centroides, como mostra a Figura 13. O processo de otimização dos clusters acontece ao diminuir a soma entre as distâncias desses pontos e o centroide mais próximo (Likas; Vlassis; Verbeek, 2003).

Figura 13 – Imagem explicativa *K-means clustering*

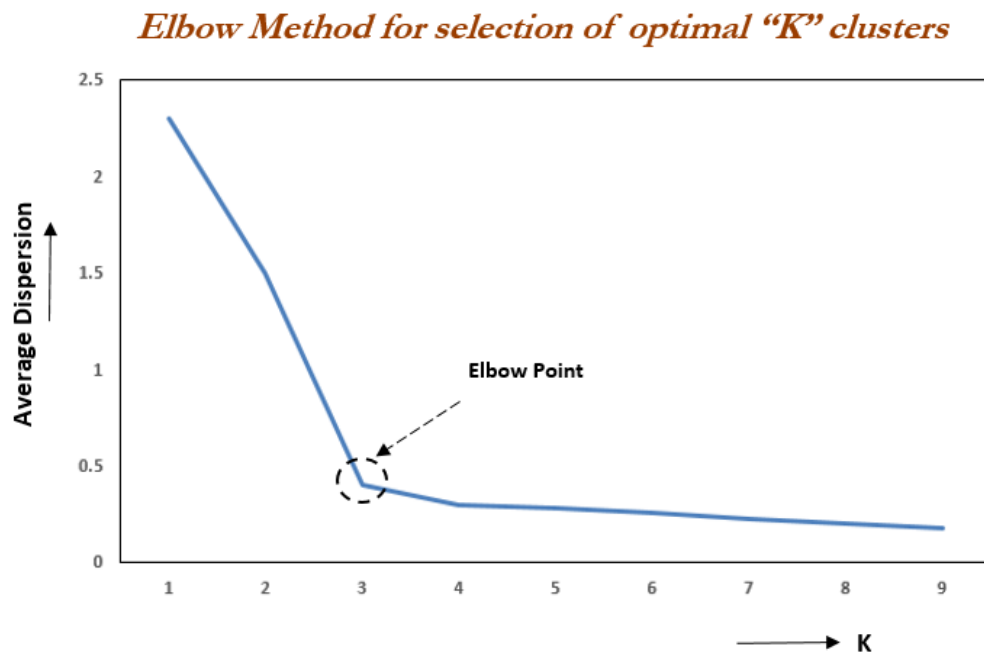


Fonte: Sharma (2021)

A solução ideal nesse caso seria o aumento do número de centroides para o mesmo valor de quantidade de pontos, porém se isso ocorrer, o agrupamento não será feito. Assim, é utilizado o método da soma dos erros quadráticos para encontrar o número ótimo de clusters (Likas; Vlassis; Verbeek, 2003; Syakur et al., 2018).

Esse processo de busca do número ideal de agrupamentos é resolvido de forma computadorizada. Por meio do próprio código em python, ao usar o chamado *Elbow Technique*, pode-se encontrar este valor ideal, como exemplificado na Figura 14.

Figura 14 – Imagem explicativa sobre *Elbow Technique* na determinação do número ideal de clusters



Fonte: Dangeti (2019)

3.2.4 Etapa 4: Definir possíveis bins geradora de descontentamento

Os perfis criados na etapa anterior devem ser aplicados na etapa 4 para os dados encontrados no mapa do Chile, desenvolvido na etapa 1. Dessa maneira é possível encontrar as regiões do mapa que tem maior probabilidade de gerar usuários insatisfeitos.

Ao identificar essas regiões no mapa, pode-se organizar estratégias de negócio para priorizar investimentos nessas regiões.

3.2.5 Etapa 5: Definir usuários de alto perfil e seus KPIs

Já na etapa 5, as tabelas Arieso devem ser filtradas de acordo com os valores de $UID_{operator}$. O $UID_{operator}$, além de ser um número único que identifica para a operadora o dispositivo celular usado, também apresenta informações sobre o modelo de celular utilizado. Uma vez as tabelas Arieso sendo filtradas para apenas usuários com dispositivos celulares de alto valor aquisitivo, são estabelecidos os usuários de alto perfil.

Ao aplicar o perfil de satisfação encontrado na etapa 3 a essa lista de usuários de alto perfil, pode-se identificar os usuários de poder aquisitivo teoricamente mais elevado que estão sendo expostos a condições de rede que geram grandes chances de insatisfação.

Com esse resultado, estratégias de negócio como distribuição de pacote de bônus de dados podem ser realizadas aos possíveis *churn users* de alto perfil, usuários que possuem aparelhos celulares *flagships* com possibilidade de deixar a operadora.

4 RESULTADOS E DISCUSSÕES

Para o desenvolvimento prático do projeto, foi imprescindível o respeito à ordem das etapas mostradas anteriormente. As etapas são complementares umas às outras, a ordem exposta acima garante o fluxo de trabalho correto do estudo.

Na primeira etapa, durante a divisão do território chileno em áreas de igual tamanho considerando a influência dos sites, foi observado que a grande maioria dos pares de coordenada em UTM eram terminados em zero. Assim, percebeu-se que um pré-processamento das coordenadas era realizado pela operadora. Essas coordenadas presentes na tabela já seguiam o padrão das áreas de 50x50m.

Uma vez assumido que o pré-processamento da operadora já centralizou as posições dos usuários de forma linear, o agrupamento das condições de rádio em cada uma dessas coordenadas pôde ser realizado diretamente.

Seguiu-se a lógica de agrupamento apresentada anteriormente, assim como os cálculos teóricos das médias ponderadas.

Uma vez obtido o arquivo final que agrupa todas as condições de rádio que ocorreram em um mesmo ponto do mapa, as coordenadas foram convertidas de UTM para latitude e longitude. Aqui foi assumido para a conversão que todas as coordenadas em UTM dizem respeito a região 19H.

Um mapa interativo foi gerado. Este mapa conta com diversos marcadores de 50m de diâmetro, todos os marcadores foram coloridos de acordo com seu valor de RSRP. O mapa segue o código de cores da Tabela 2, padrão da operadora.

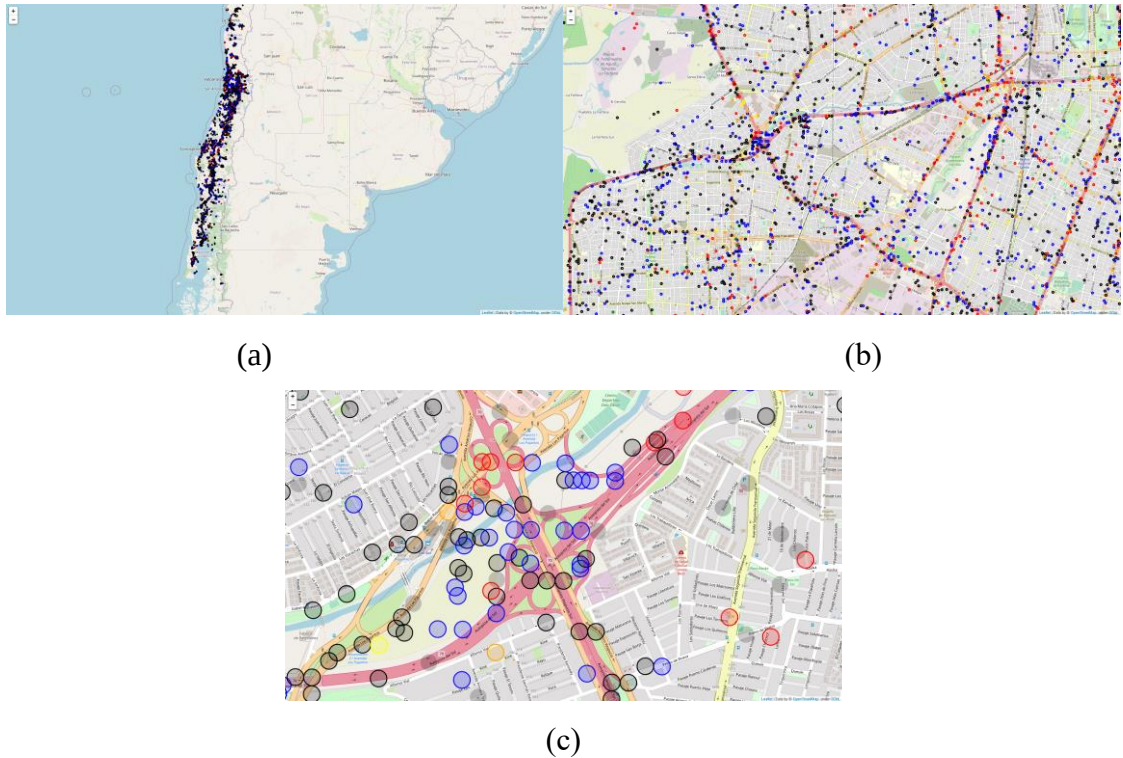
Tabela 2 - Código de cores utilizado pela operadora, implementado no trabalho para a ilustração das condições de sinal no mapa

<i>Cor</i>	<i>Valores de RSRP</i>
Preto	-120 a -110
Azul escuro	-110 a -100
Vermelho	-100 a -90
Laranja	-90 a -80
Amarelo	-80 a -75
Verde	-75 a -65
Azul claro	>-65

Fonte: Autoria própria

Este mapa foi criado no formato HTML, nele o usuário é capaz de mover a região observada e utilizar ferramentas de zoom. Na Figura 15, algumas imagens encontradas estão sendo exibidas.

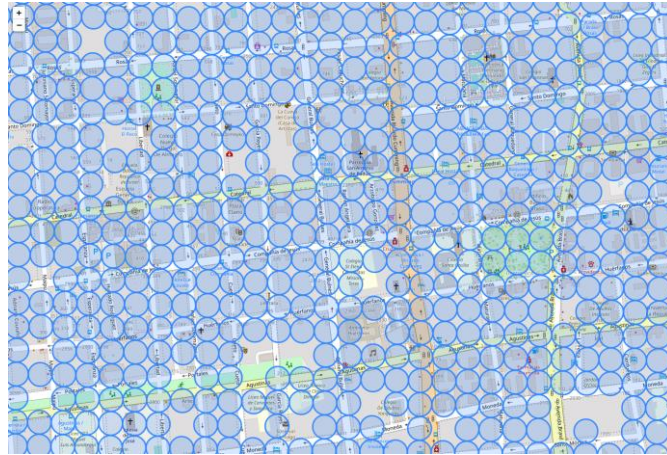
Figura 15 – Imagens geradas em python com marcadores de intensidade de sinal sobre a região do Chile em diferentes escalas de zoom



Fonte: Autoria própria

Como pode-se observar na figura 15c, as circunferências que deveriam sempre se tangenciar, por vezes se sobrepõem. Isso indica um erro no pré-processamentos de dados. Este erro foi informado ao grupo responsável e está sendo corrigido para as próximas amostras de dados, como pode ser visto no exemplo da Figura 16 a seguir.

Figura 16 – Imagem gerada em python por meio de base de dados corrigidas exibindo o posicionamento correto dos pares de coordenadas



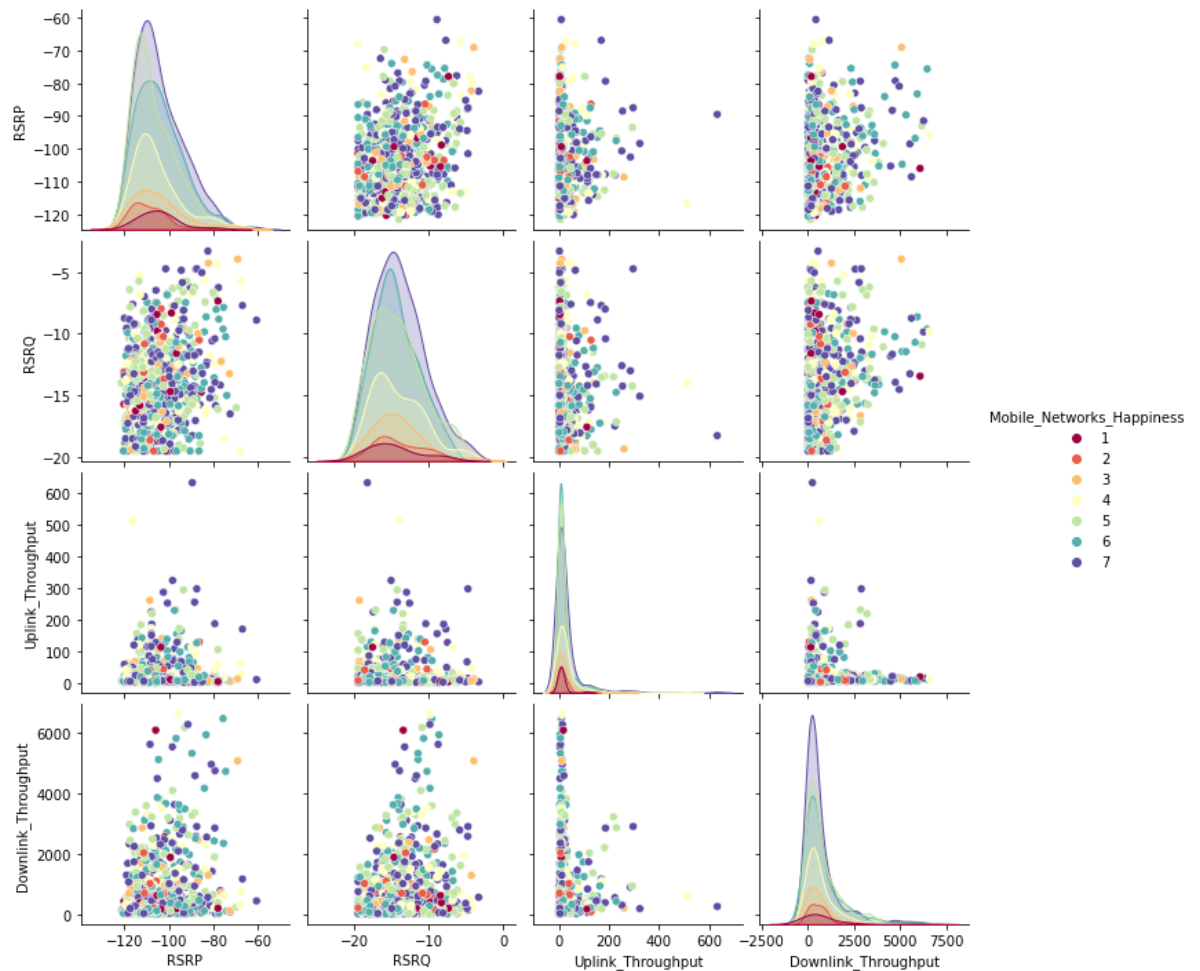
Fonte: Autoria própria

Prosseguindo para a segunda etapa do projeto, foi realizada a limpeza de dados e extração de parâmetros pertinentes ao estudo como detalhado anteriormente. Dentro do próprio código, usando funções da biblioteca pandas como o *.merge*, os três bancos de dados foram reunidos e correlacionados no mesmo *dataframe*.

Tendo agora um único *dataframe* que concentra todos os dados relevantes ao estudo iniciou-se a busca pelos padrões de comportamento para a criação dos perfis de satisfação.

Os principais KPIs foram analisados de forma gráfica para os diferentes níveis de satisfação. Nessa primeira análise foi utilizada a satisfação quanto a qualidade dos dados móveis da operadora de forma geral. Neste primeiro momento os parâmetros de RSRP, RSRQ, *uplink throughput* e *downlink throughput* foram examinados dois a dois. Como pode ser visto na imagem abaixo Figura 17, os diferentes níveis de satisfação são representados pelas diferentes cores, sendo a menor satisfação o nível 1.

Figura 17 – Gráfico para análise comportamental dos diferentes níveis de satisfação para diversos parâmetros analisados dois a dois

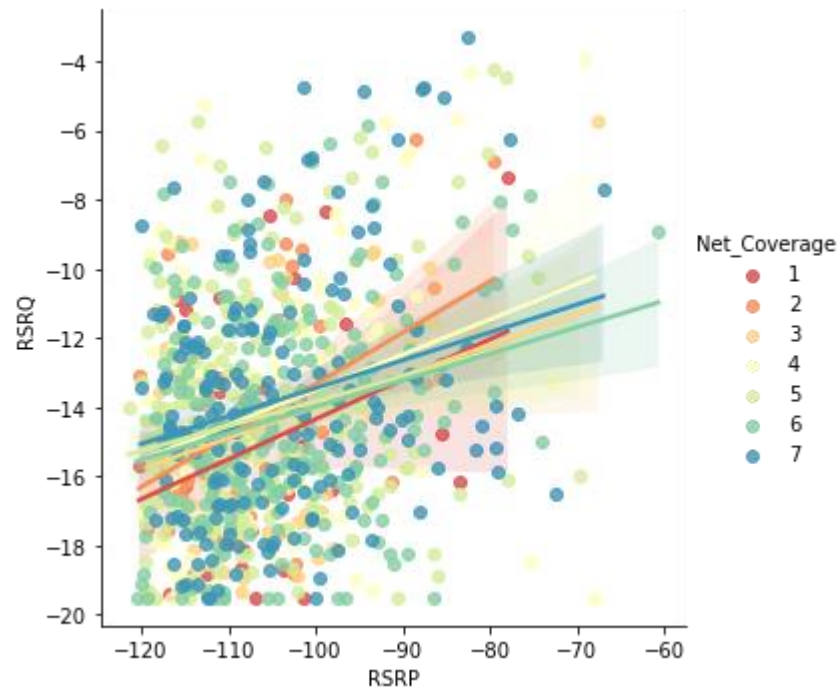


Fonte: Autoria própria

Nesta primeira análise, porém, pode-se perceber que não há um grupo de cores em uma região bem definida para nenhum dos parâmetros. Sendo assim, em um primeiro momento, não foi encontrado um cenário comum de exposição para os grupos de satisfação.

Ainda buscando encontrar um comportamento bem definido dos grupos de satisfação, dois parâmetros que teoricamente estariam relacionados foram observados: a satisfação quanto o sinal de telefone e os dados de RSRP e RSRQ. Nesta análise, linhas são traçadas para representar o comportamento de cada um dos grupos, porém como pode ser visto na Figura 18 a seguir, as linhas de satisfações se sobrepõem.

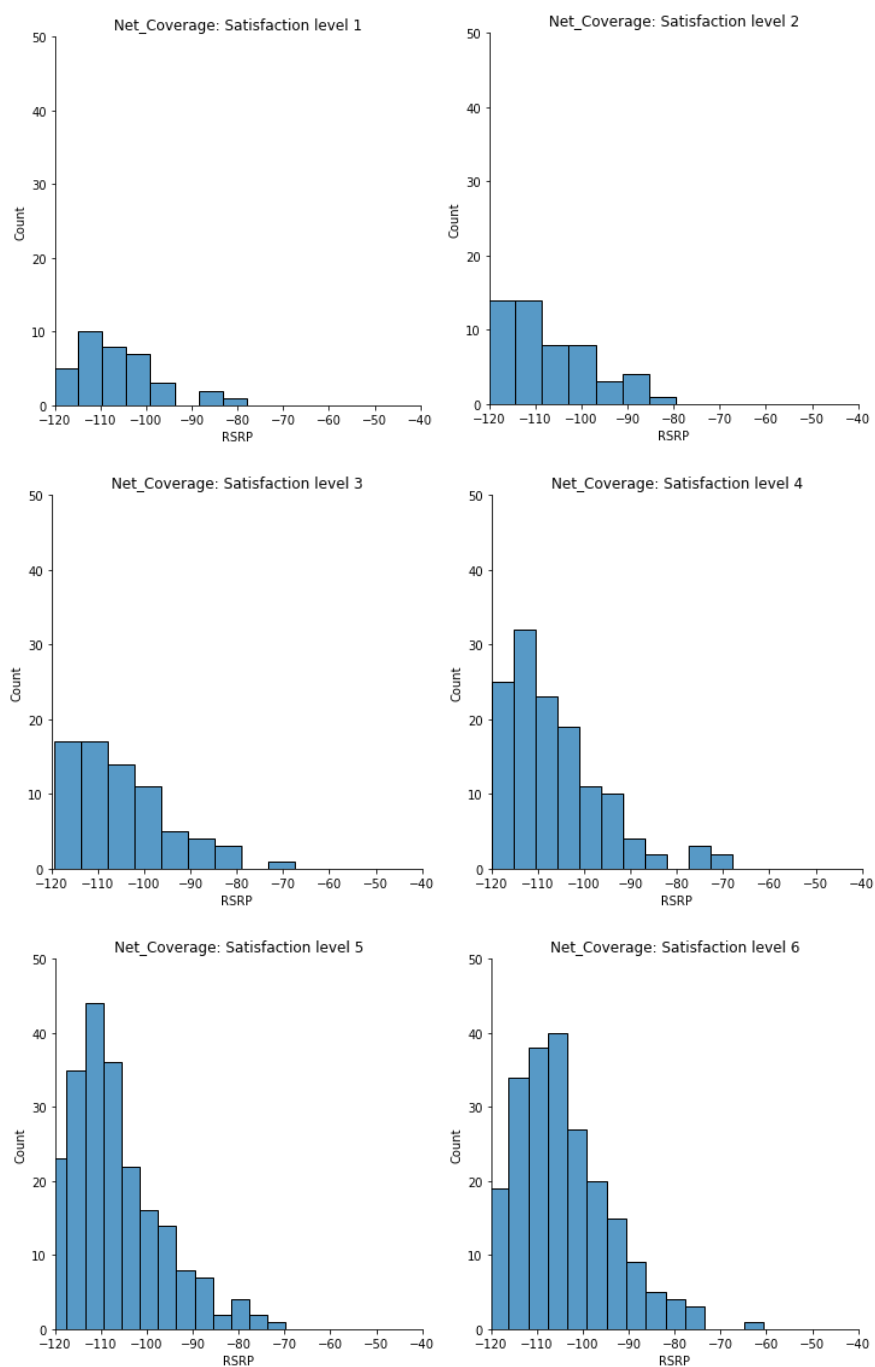
Figura 18 – Gráfico com linhas comportamentais para os diferentes níveis de satisfação em termos de RSRP e RSRQ

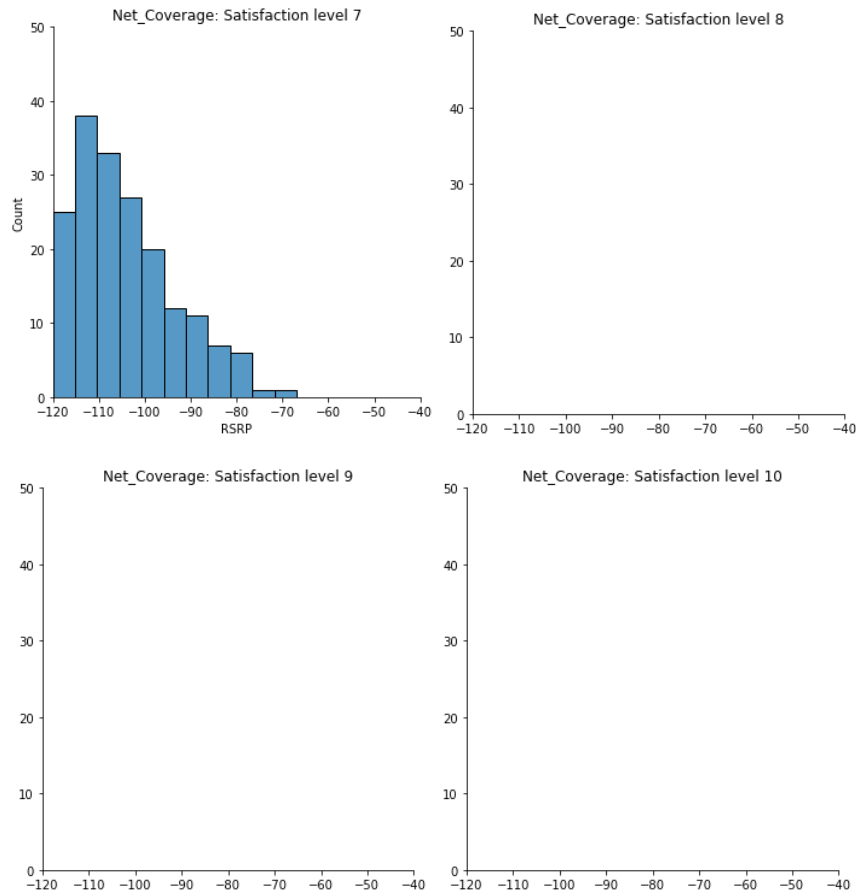


Fonte: Autoria própria

O resultado encontrado acima foi surpreendente, era imaginado que os dois parâmetros usados na análise anterior tivessem uma correlação direta com a qualidade do sinal percebida pelo usuário. Para uma observação mais detalhada do comportamento descrito acima, cada um dos grupos de satisfação foi analisado individualmente na Figura 19. Os seguintes gráficos mostram a quantidade de usuários e seus valores de RSRP para cada um dos níveis de satisfação.

Figura 19 – Gráficos de contadores de usuários por cenários de RSRP para cada um dos níveis de satisfação





Fonte: Autoria própria

Os contadores dos gráficos, nesse exemplo, mostram valores bem distintos dentre os diferentes grupos. Um indicativo, porém, que justifica os diferentes comportamentos entre os grupos são as quantidades desiguais de usuários presentes em cada um deles, o que foi confirmado abaixo na Figura 20.

Figura 20 – *Output* obtido em python com a quantidade de usuários em cada um dos níveis de satisfação

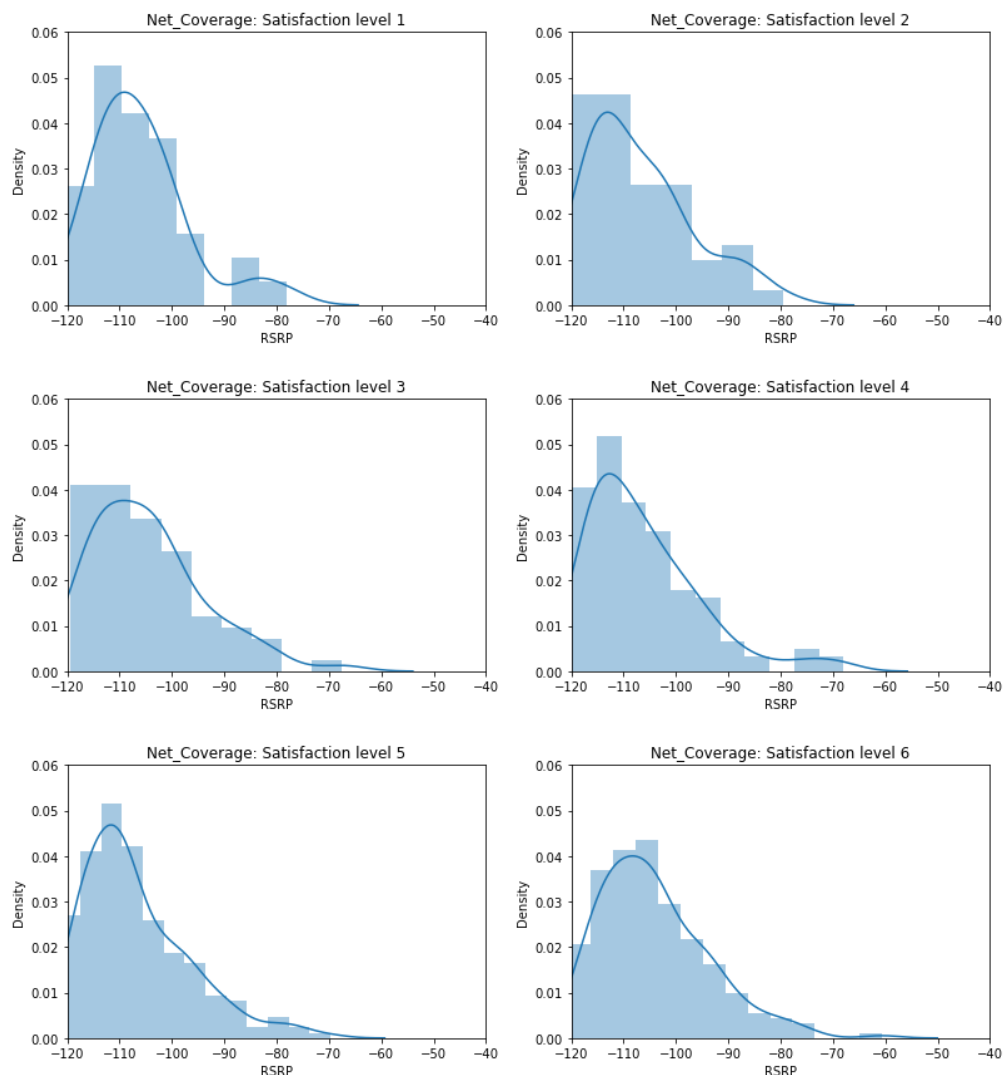
```
Total users with satisfaction 1 in 'Net_Coverage': 36
Total users with satisfaction 2 in 'Net_Coverage': 52
Total users with satisfaction 3 in 'Net_Coverage': 72
Total users with satisfaction 4 in 'Net_Coverage': 131
Total users with satisfaction 5 in 'Net_Coverage': 214
Total users with satisfaction 6 in 'Net_Coverage': 215
Total users with satisfaction 7 in 'Net_Coverage': 181
Total users with satisfaction 8 in 'Net_Coverage': 0
Total users with satisfaction 9 in 'Net_Coverage': 0
Total users with satisfaction 10 in 'Net_Coverage': 0
```

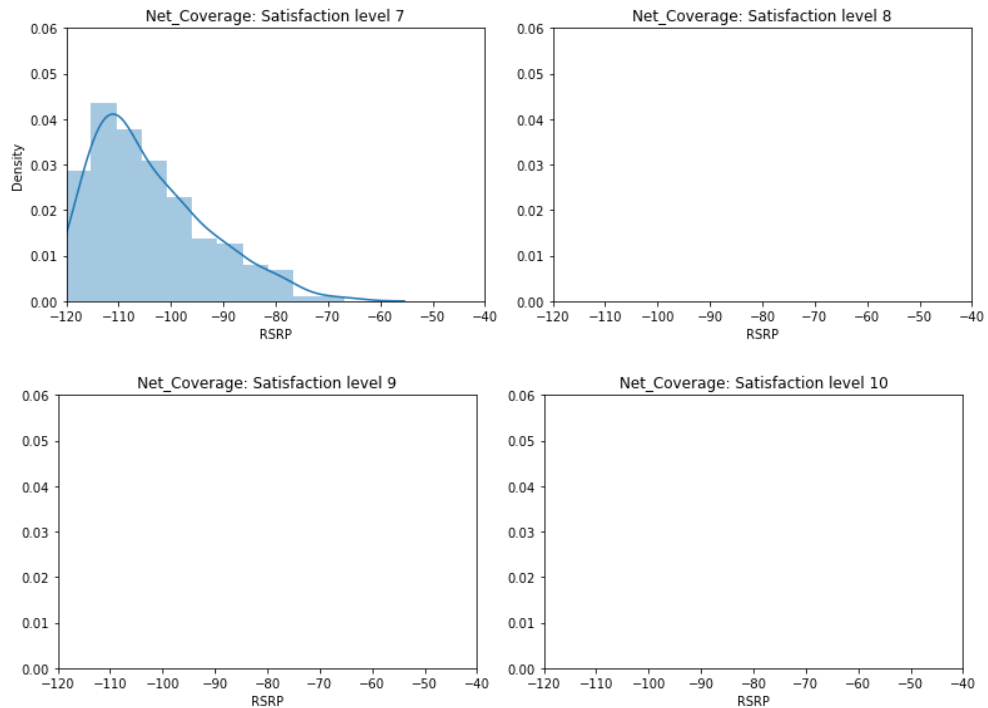
Fonte: Autoria própria

Uma vez comprovado que o número de usuários das amostras separadas pelo nível de satisfação é diferente, vê-se a necessidade de normalizar as curvas. Há grupos de maior identificação dos usuários como o nível 6 que conta com 215 usuários, enquanto tem-se o grupo 1 de menor identificação com apenas 36 usuários além de outros grupos com os quais os usuários não se identificam.

Assim, pode-se observar abaixo as análises gráficas que consideram novamente a quantidade de usuário por valor de RSRP para cada uma das satisfações, porém agora os gráficos mostram a densidade de probabilidade para cada uma dessas amostras. Uma vez que os gráficos foram todos normalizados na Figura 21, a comparação entre eles é de mais fácil realização.

Figura 21 - Gráficos com densidade de probabilidade para diferentes cenários de RSRP em cada um dos níveis de satisfação

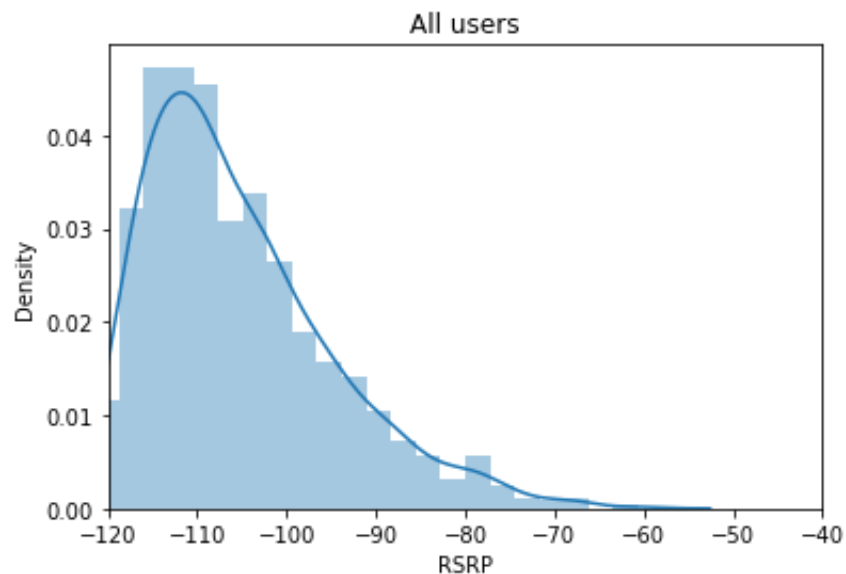




Fonte: Autoria própria

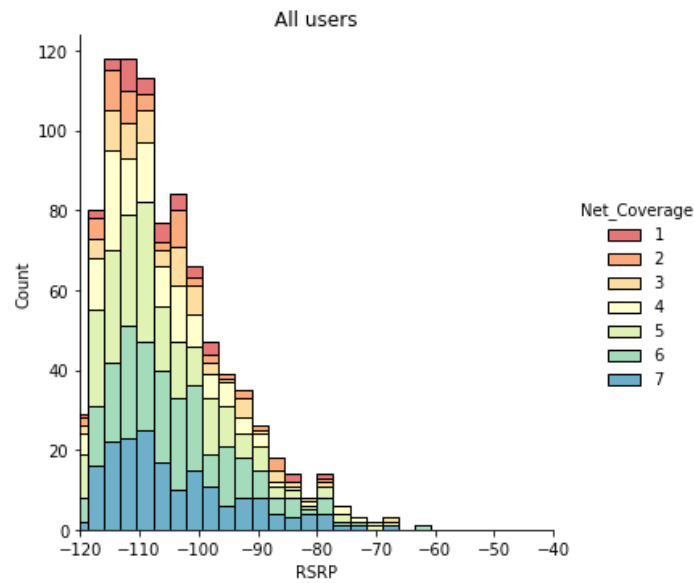
Ao utilizar o método acima, pode-se observar de maneira geral que a curva de densidade de probabilidade se repete para todos os níveis de satisfação. Em todos os casos exibidos acima, a maioria dos usuários foi exposta a valores de RSRP entre -115 e -105 dBm. O que é coerente com a distribuição considerando todos os níveis observada nas Figuras 22 e 23.

Figura 22 - Gráfico com a densidade de probabilidade para diferentes cenários de RSRP para a amostra completa de usuários



Fonte: Autoria própria

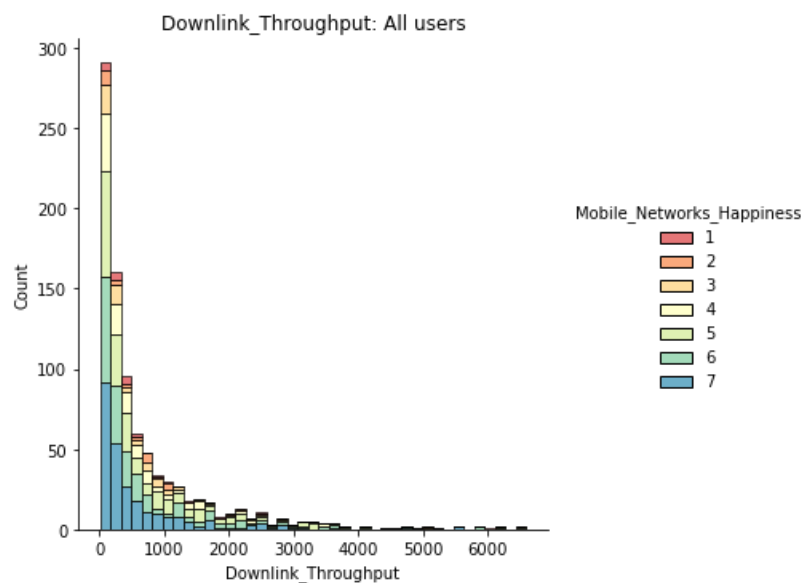
Figura 23 – Gráfico de contadores de usuários para cenários de RSRP utilizando a amostra completa com identificação de cor por nível de satisfação



Fonte: Autoria própria

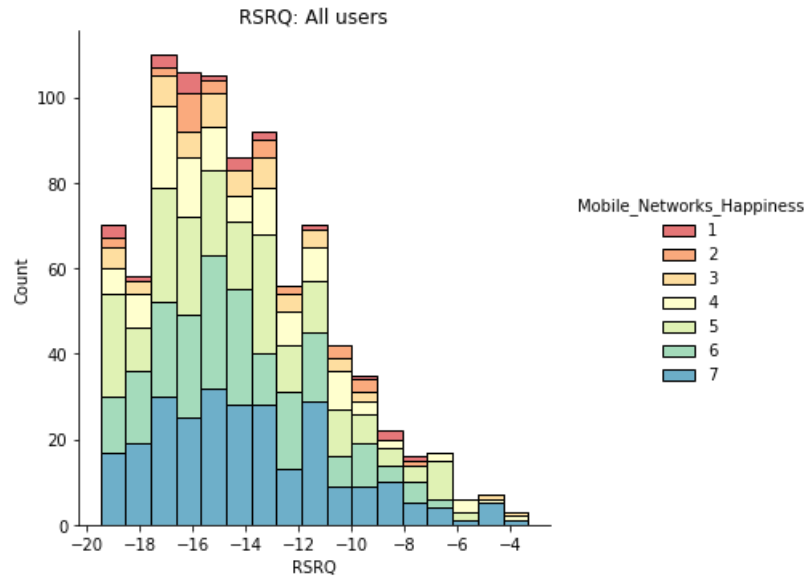
O mesmo estudo foi realizado para outros parâmetros considerando a satisfação geral quanto a operadora. As imagens abaixo, Figuras 24 e 25, mostram os estudos quanto aos valores de RSRQ e de *downlink throughput*.

Figura 24 – Gráfico de contadores de usuários para diferentes taxas de *downlink throughput* utilizando a amostra completa com identificação de cor por nível de satisfação



Fonte: Autoria própria

Figura 25 – Gráfico de contadores de usuários para cenários de RSRQ utilizando a amostra completa com identificação de cor por nível de satisfação



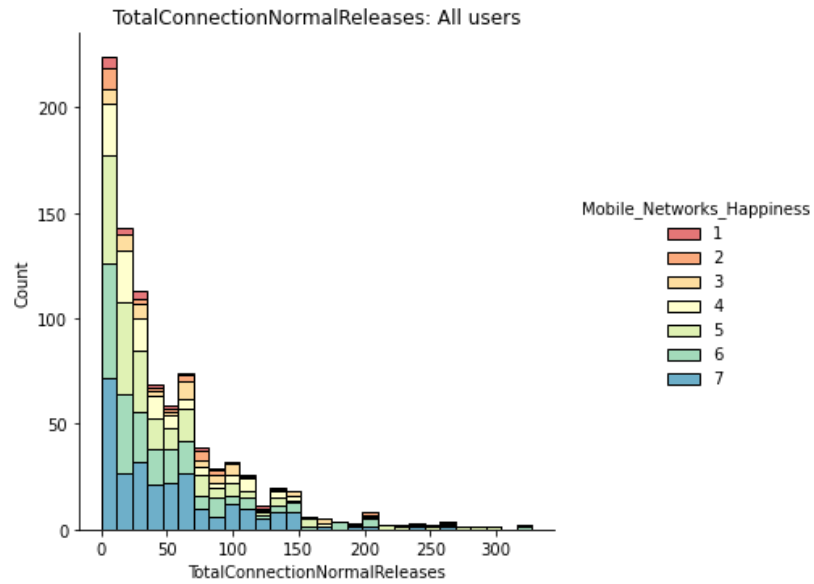
Fonte: Autoria própria

Em ambos os estudos, percebe-se o mesmo comportamento. Os picos das curvas individuais ocorrem nas mesmas regiões, sendo essas regiões os picos referentes a todos os usuários.

O mesmo estudo foi repetido agora para os parâmetros de total de *releases* na conexão e total de *drops* na conexão. Porém foram novamente inconclusivos.

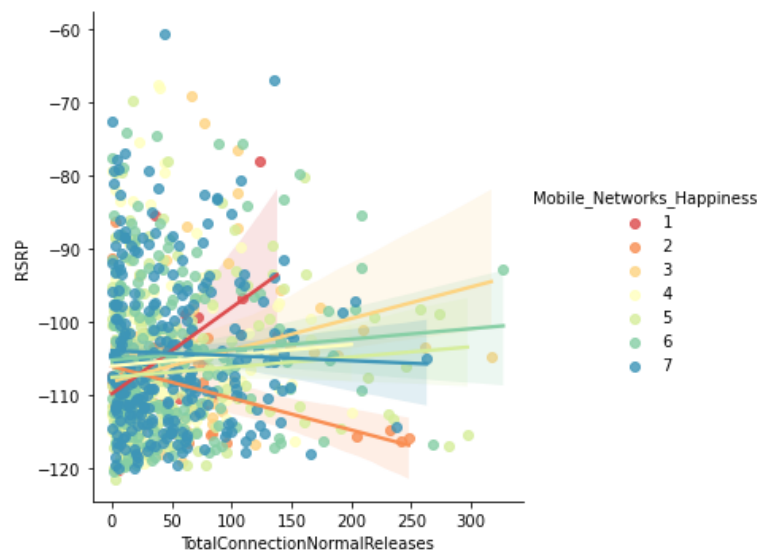
Abaixo encontra-se o estudo quanto ao total de *releases* na conexão, Figuras 26, 27 e 28.

Figura 26 – Gráfico de contadores de usuários para diferentes valores de *releases* utilizando a amostra completa com identificação de cor por nível de satisfação



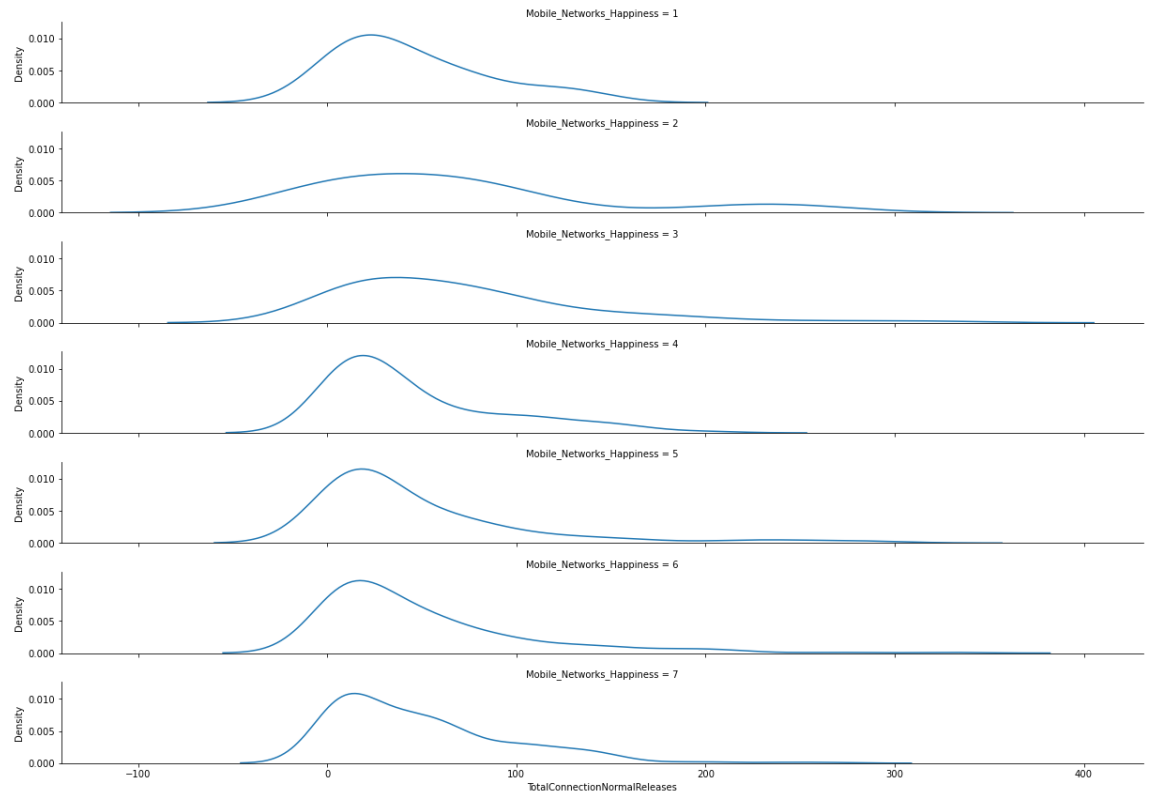
Fonte: Autoria própria

Figura 27 – Gráfico com linhas comportamentais para os diferentes níveis de satisfação em termos de RSRP e *releases*



Fonte: Autoria própria

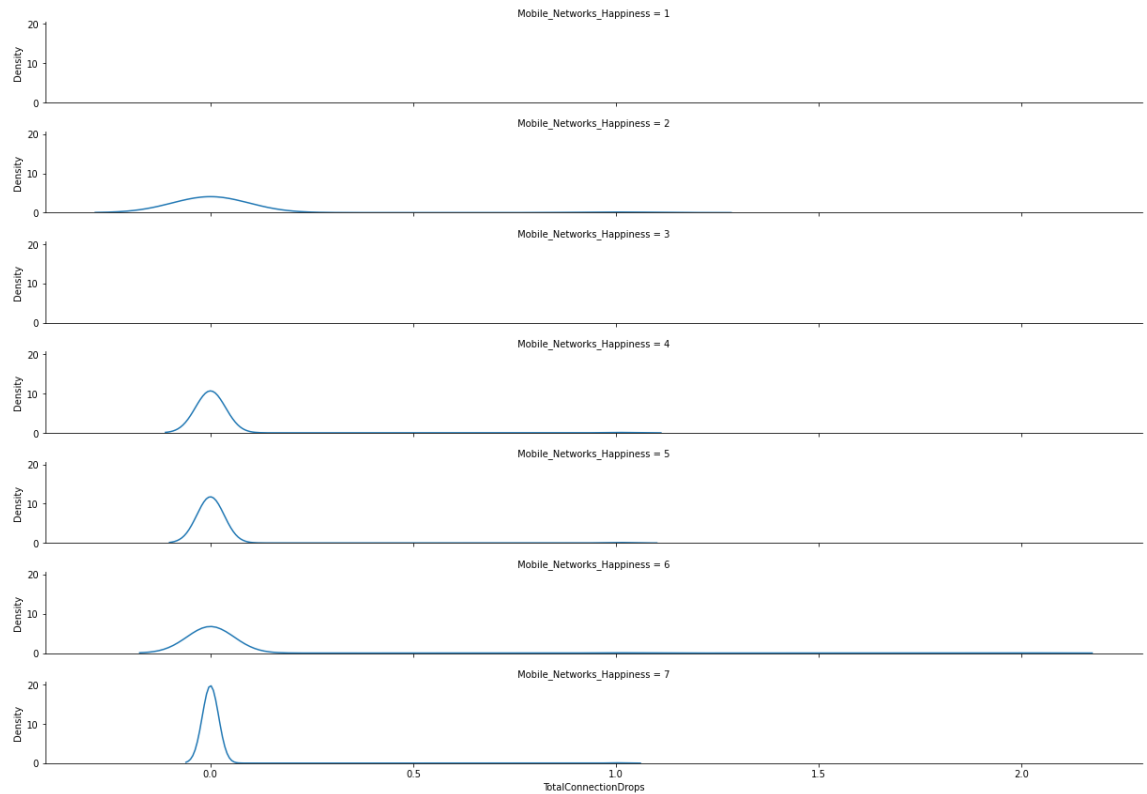
Figura 28 - Gráficos com densidade de probabilidade para diferentes valores de *releases* em cada um dos níveis de satisfação



Fonte: Autoria própria

A seguir são exibidos os estudos quanto ao total de *drops* na conexão, Figura 29. Neste caso ainda se encontrou o problema de ausência de dados.

Figura 29 - Gráficos com densidade de probabilidade para diferentes valores de *drops* em cada um dos níveis de satisfação



Fonte: Autoria própria

Uma vez que as comparações gráficas não foram suficientes para se encontrar um comportamento comum e bem definido das condições de rádio, o processo de clusterização foi implementado usando o sistema da Figura 30.

Figura 30 – Amostra do *dataframe* usado para a clusterização com os parâmetros considerados pelo sistema

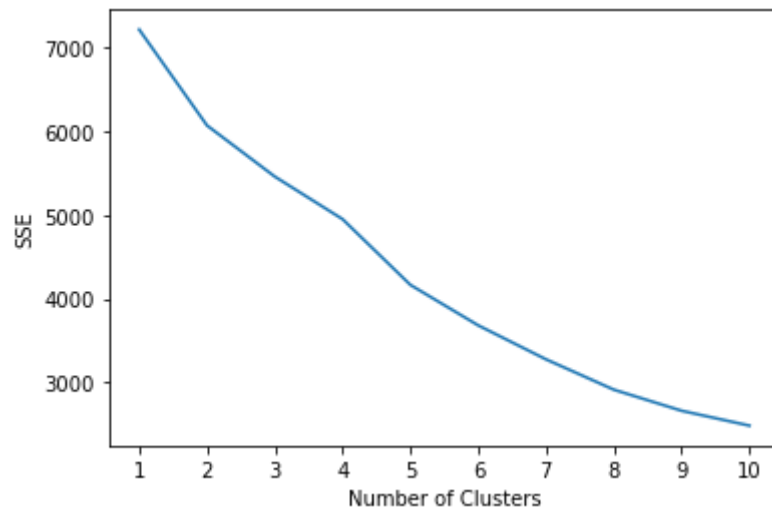
	RSRP	RSRQ	Uplink_Throughput	Downlink_Throughput	Uplink_Volume	Downlink_Volume	TotalConnectionDrops	TotalConnectionNormalReleases
0	-102.500000	-13.750000	0.595000	24.040000	12	10	0	3
1	-113.833333	-13.916667	1.938333	16.973333	11	10	0	4
2	-112.750000	-15.940000	8.115000	91.085000	54	47	0	0
3	-112.500000	-13.750000	47.100000	474.520000	191	270	0	0
4	-98.383333	-11.076667	12.166667	4082.263333	3604	192309	0	12

Fonte: Autoria própria

A primeira análise usando o método de *k-means clustering* utilizou todos os parâmetros disponíveis nos *dataframes* que dizem respeito as condições de rede como *inputs* do sistema.

Ao utilizar métodos computadorizados para encontrar o número ideal de agrupamentos, foi utilizado o *Elbow Technique* e o gráfico da Figura 31 foi encontrado.

Figura 31 – Gráfico gerado por meio do *Elbow Technique* para definir o número de agrupamentos do sistema

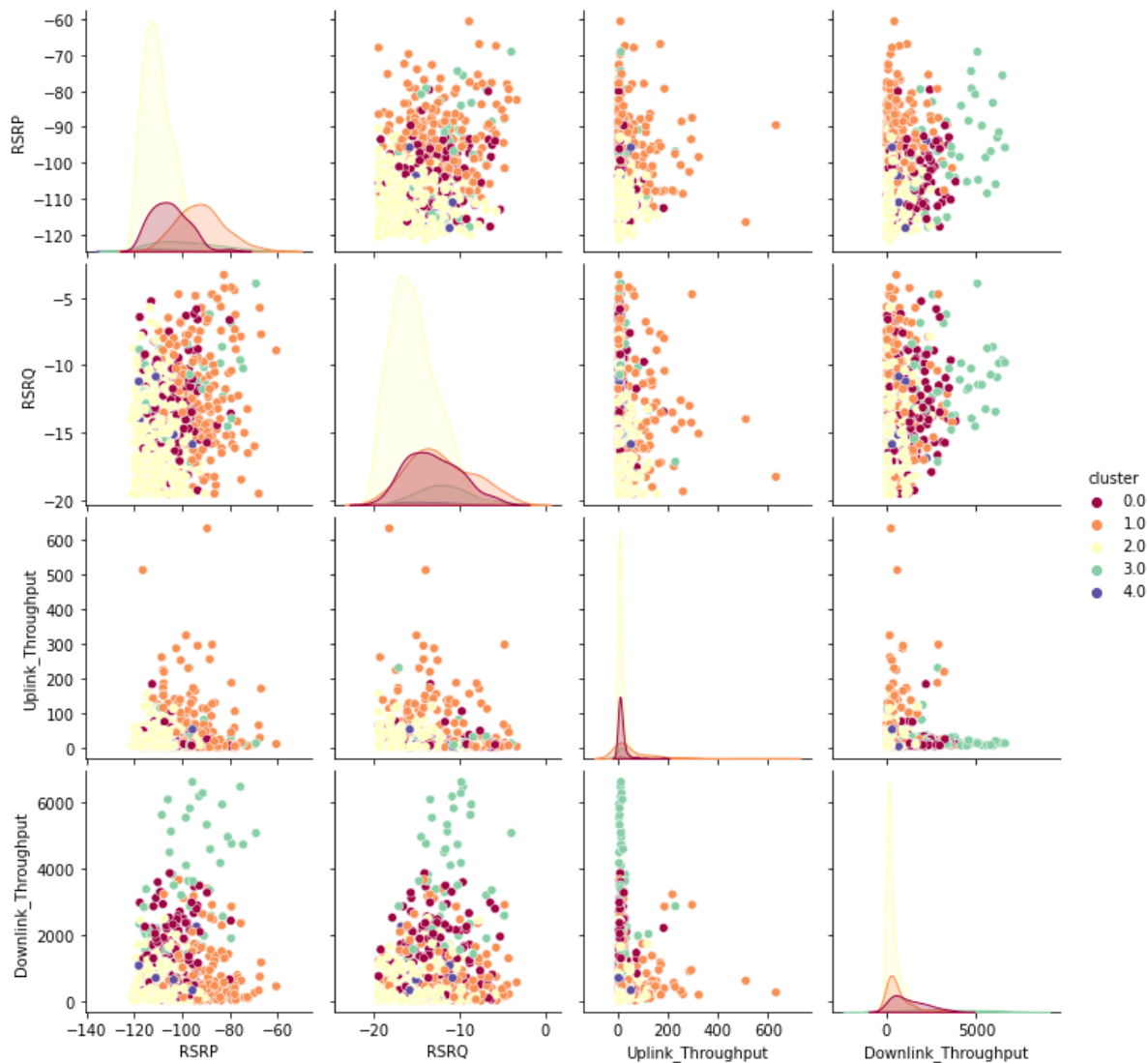


Fonte: Autoria própria

A partir da interpretação do gráfico acima, foi introduzido ao programa o valor de 5 clusters. Assim, os usuários da amostra foram separados em 5 grupos diferentes por critérios que o computador julgou ideal. Cabe agora na análise interpretar a divisão realizada pelo computador de acordo com os parâmetros estabelecidos e encontrar embasamento teórico para a separação realizada. O principal objetivo é encontrar um grupo bem estabelecido com critérios teóricos que justifiquem a insatisfação dos usuários.

Para isso, foi plotado o gráfico da Figura 32 o qual cada um dos principais parâmetros do sistema foi analisado dois a dois. Desta vez, as diferentes cores encontradas no gráfico consistem nos diferentes grupos criados pelo método de clusterização computadorizada.

Figura 32 - Gráfico para análise comportamental dos diferentes agrupamentos realizados pelo *k-means clustering* para diversos parâmetros analisados dois a dois

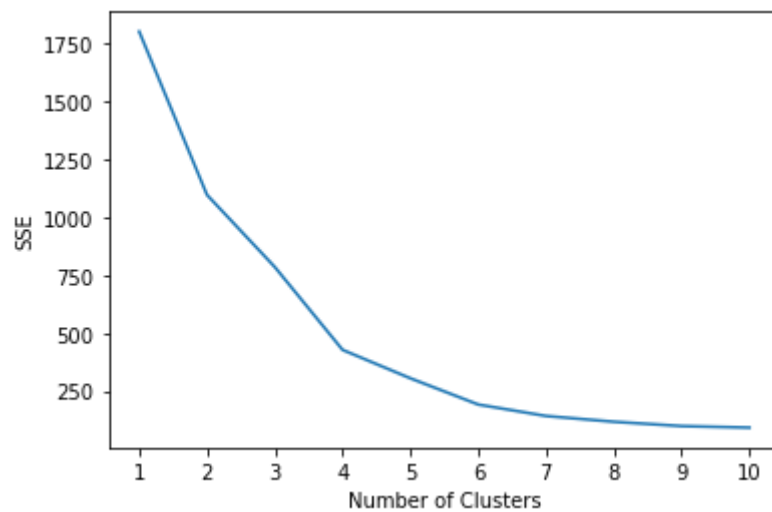


Fonte: Autoria própria

Nenhum grupo foi encontrado de forma a ratificar o pensamento anterior. Adotou-se, portanto, uma diferente narrativa para a busca dos *churn users*. As análises que foram realizadas a seguir consideraram um sistema menos complexo, com menos parâmetros. Uma vez o método de clusterização sendo aplicado, é mais fácil encontrar as justificativas teóricas para a mais baixa satisfação do usuário. Uma vez essas hipóteses sendo criadas, é observada a representatividade das hipóteses do agrupamento insatisfeito dos clusters em relação à pesquisa de satisfação. Em outras palavras, ao encontrar a partir da divisão pelos clusters um grupo que teoricamente esteja insatisfeito com o serviço da operadora, será comparado quanto desse grupo também se autodeclara insatisfeito. Uma vez encontrando uma alta correlação entre os dois agrupamentos, os limites dos perfis de satisfação podem ser delimitados.

A primeira análise realizada utilizou os parâmetros volume de *uplink throughput* e volume de *downlink throughput*. Ao utilizar o *Elbow Technique*, o gráfico da Figura 33 foi encontrado.

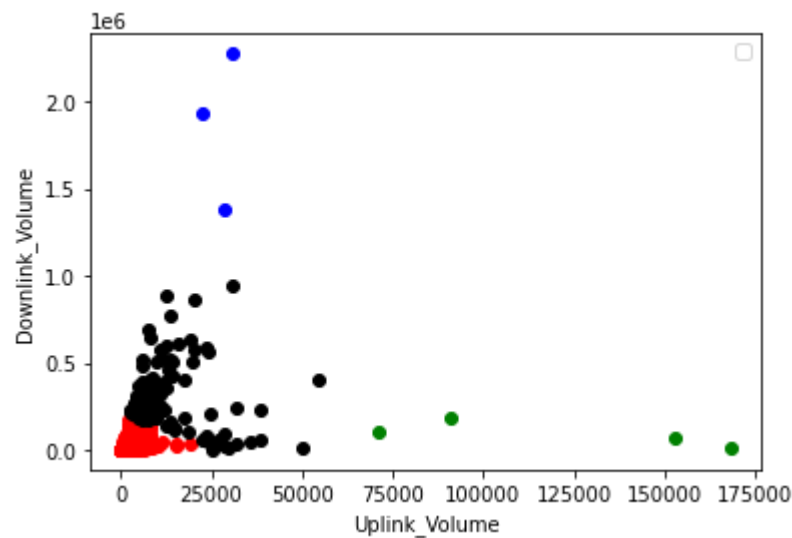
Figura 33 – Gráfico gerado por meio do *Elbow Technique* para definir o número de agrupamentos do sistema



Fonte: Autoria própria

O agrupamento nesse sistema utilizou 4 clusters. A divisão computadorizada nesses grupos pode ser observada no gráfico a seguir da Figura 34.

Figura 34 – Divisão computadorizada feita do sistema para volumes de *downlink* e *uplink*



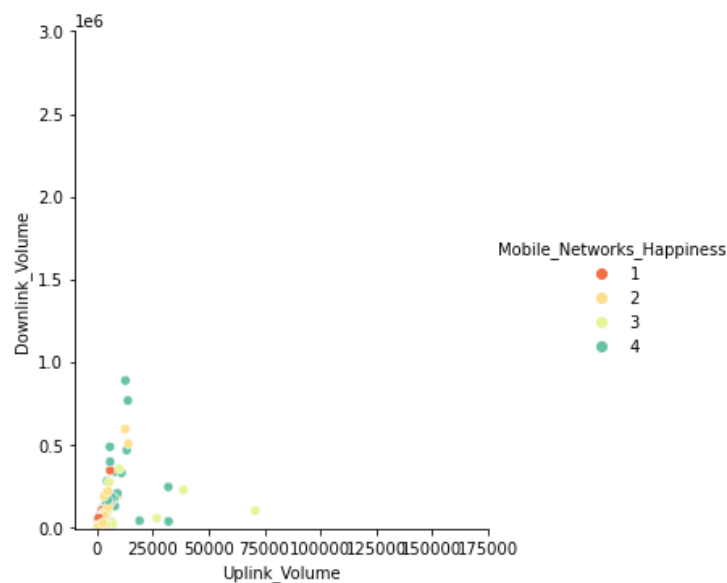
Fonte: Autoria própria

Ao observar o gráfico, criou-se a hipótese de que o grupo insatisfeito neste sistema é o representado pela cor vermelha uma vez que ele não foi capaz de enviar um grande volume de dados nem de recebê-los.

Para observar a representatividade do sistema em relação a autodeclaração dos usuários na pesquisa de satisfação, os níveis de contentamento dos usuários foram divididos de forma mais abrangente. Sendo assim, foi considerado no grande grupo dos insatisfeitos as respostas de 1 a 4 da pesquisa de satisfação. Já o grande grupo dos satisfeitos abrangeu as respostas de 5 a 7, visto que os demais níveis não foram selecionados pelos usuários.

Ao observar o posicionamento gráfico dos usuários insatisfeitos, Figura 35, percebe-se que eles estão realmente em sua grande maioria posicionados na condição de rádio considerada como pior na hipótese.

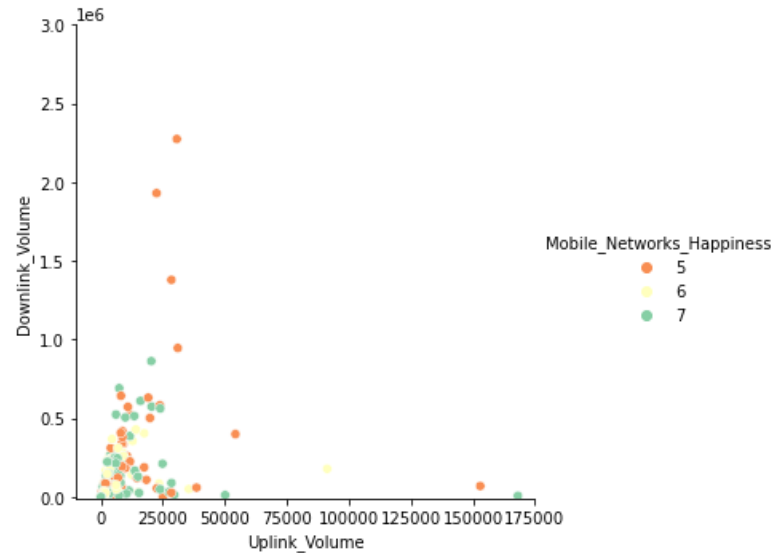
Figura 35 – Posicionamento gráfico dos grupos de mais baixa satisfação na pesquisa em relação aos volumes *downlink* e *uplink*



Fonte: Autoria própria

Ao observar, porém, o grande grupo dos usuários satisfeitos, Figura 36, percebe-se que eles também se encontram em sua grande maioria na região tida como pior na hipótese criada, região vermelha.

Figura 36 – Posicionamento gráfico dos grupos de mais alta satisfação na pesquisa em relação aos volumes *downlink* e *uplink*



Fonte: Autoria própria

Uma vez que os dois grupos, tanto de usuários autodeclarados satisfeitos quanto os insatisfeitos, se encontram expostos a essa mesma condição de rádio, a hipótese criada não foi validada no estudo. Isso pode ser comprovado ao observar em detalhes a divisão realizada pelo cluster, Figuras 37 e 38.

Figura 37 – Fragmento do código exibindo a quantidade total de linhas do *dataframe* no *output*

```
In [275]: #append cluster assingments to original DataFrame
df_cluster['cluster'] = kmeans.labels_
print(df_cluster)
```

	Uplink_Volume	Downlink_Volume	cluster
0	12	10	1
1	11	10	1
2	54	47	1
3	191	270	1
4	3604	192309	1
..	...	...	...
896	39	67	1
897	1464	36613	1
898	1813	26786	1
899	136	182	1
900	52	66	1

[901 rows x 3 columns]

Fonte: Autoria própria

Figura 38 – Fragmento do código exibindo a quantidade de linhas do *dataframe* para o cluster 1 no *output*

```
In [281]: df_cluster[df_cluster.cluster==1]
```

Out[281]:

	Uplink_Volume	Downlink_Volume	cluster
0	12	10	1
1	11	10	1
2	54	47	1
3	191	270	1
4	3604	192309	1
...	...	...	...
896	39	67	1
897	1464	36613	1
898	1813	26786	1
899	136	182	1
900	52	66	1

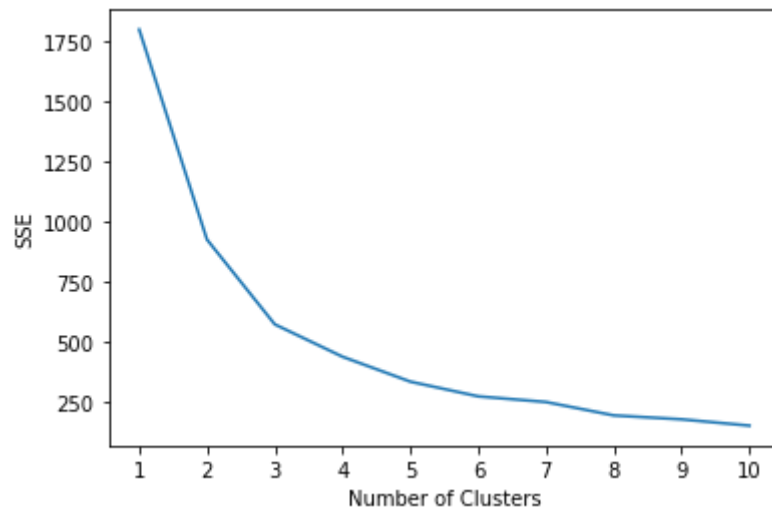
785 rows × 3 columns

Fonte: Autoria própria

Pelas imagens exibidas acima, percebe-se que o sistema apresenta um total de 901 usuários, sendo que 785 deles encontram-se na região vermelha do gráfico tida como geradora de insatisfação. Sendo assim, 87,12% dos usuários estariam inseridos na hipótese criada, derogando, portanto, sua pertinência ao estudo.

Para a criação da próxima hipótese, foi analisado um sistema que contava com os parâmetros de minutos de uso e *downlink throughput*. Novamente, a *Elbow Technique* foi utilizada e como pode ser observada na Figura 39, o número ideal de clusters encontrados foi de 3.

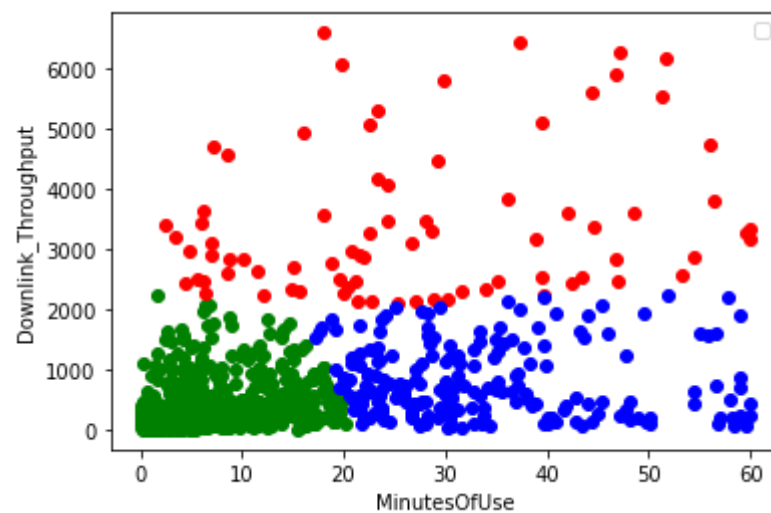
Figura 39 – Gráfico gerado por meio do *Elbow Technique* para definir o número de agrupamentos do sistema



Fonte: Autoria própria

A divisão computadorizada para este sistema é ilustrada pelo gráfico da Figura 40.

Figura 40 - Divisão computadorizada feita do sistema para *downlink throughput* e minutos de uso



Fonte: Autoria própria

Ao realizar a interpretação teórica deste gráfico, assume-se que os usuários localizados na região de cor azul do gráfico são os tidos como insatisfeitos uma vez que mesmo estes usuários se mantendo conectados por um longo período de tempo, eles não foram capazes de alcançar um alto valor de *throughput*.

A região azul do gráfico utilizada na hipótese diz respeito a 21,98% do total de usuários. Portanto, ela representa apenas um pequeno recorte da amostra, podendo ser usada na criação dos perfis de satisfação caso validada, como observado na Figura 41.

Figura 41 – Fragmento do código exibindo a quantidade de linhas do *dataframe* para o cluster 2 no *output*

```
In [247]: df_cluster[df_cluster.cluster==2]
```

Out[247]:

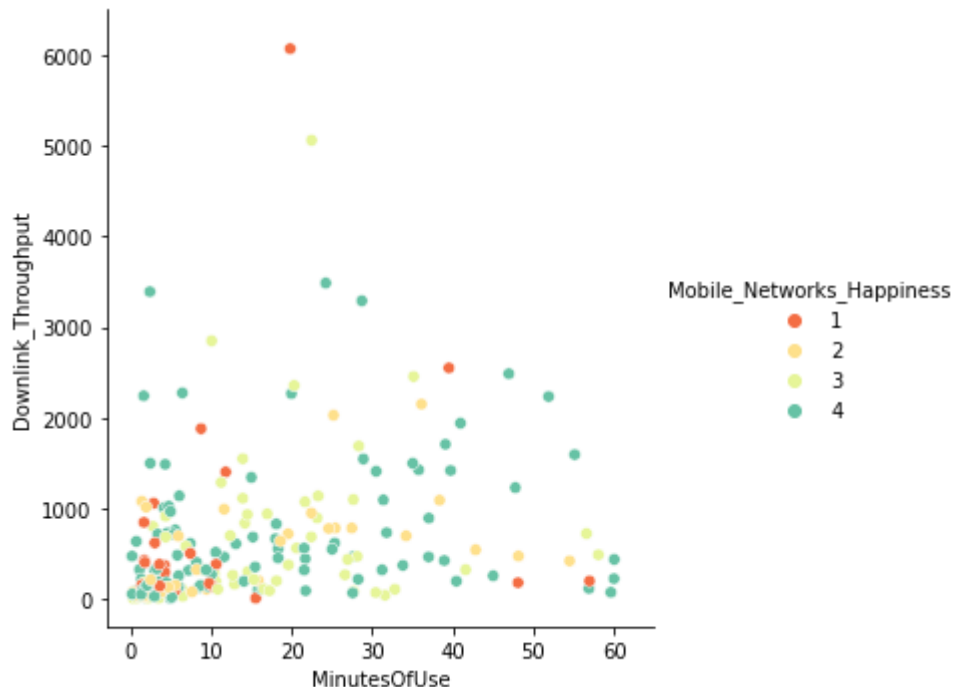
	MinutesOfUse	Downlink_Throughput	cluster
9	57.888	2200.457500	2
15	45.888	1587.842917	2
19	17.763	1712.429205	2
34	58.484	78.034167	2
35	50.163	106.380000	2
...	...	...	...
850	55.806	1556.615625	2
869	35.138	1288.025625	2
870	22.847	267.630545	2
877	25.181	2028.946250	2
879	33.363	1638.997917	2

198 rows × 3 columns

Fonte: Autoria própria

Ao observar, porém, o posicionamento do grande grupo dos insatisfeitos em relação aos mesmos parâmetros, percebe-se que esses dois grupos não se sobrepõem, Figura 42.

Figura 42 – Posicionamento gráfico dos grupos de mais baixa satisfação na pesquisa em relação ao *downlink throughput* e minutos de uso



Fonte: Autoria própria

Isso indica que há usuários da operadora que foram expostos as condições presentes na hipótese acima, porém esses usuários não se autodeclararam insatisfeitos. Como pode ser observado nas imagens abaixo, Figuras 43, 44 e 45, o grande grupo dos insatisfeitos conta com 220 usuários, porém apenas 63 deles estão presentes na região apontada. Desse modo, apenas 28,64% dos usuários insatisfeitos estão incluídos na hipótese aqui criada, o que a invalida.

Figura 43 – Fragmento do código exibindo a criação do *dataframe* com os clientes de maior insatisfação (níveis 1, 2, 3 e 4)

```
In [312]: bad_data = df_4g
bad_data['cluster'] = kmeans.labels_
bad_data = bad_data[(bad_data['Mobile_Networks_Happiness']==1')+
                    (bad_data['Mobile_Networks_Happiness']==2')+
                    (bad_data['Mobile_Networks_Happiness']==3')+
                    (bad_data['Mobile_Networks_Happiness']==4')]
bad_data.reset_index(drop=True, inplace=True)
bad_data.head()
```

Fonte: Autoria própria

Figura 44 – Fragmento do código exibindo a quantidade total de usuários no grande grupo dos insatisfeitos

```
In [298]: bad_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 220 entries, 0 to 219
```

Fonte: Autoria própria

Figura 45 – Fragmento do código exibindo a quantidade de usuários no grande grupo dos insatisfeitos que foram classificados como pertencentes ao agrupamento 2

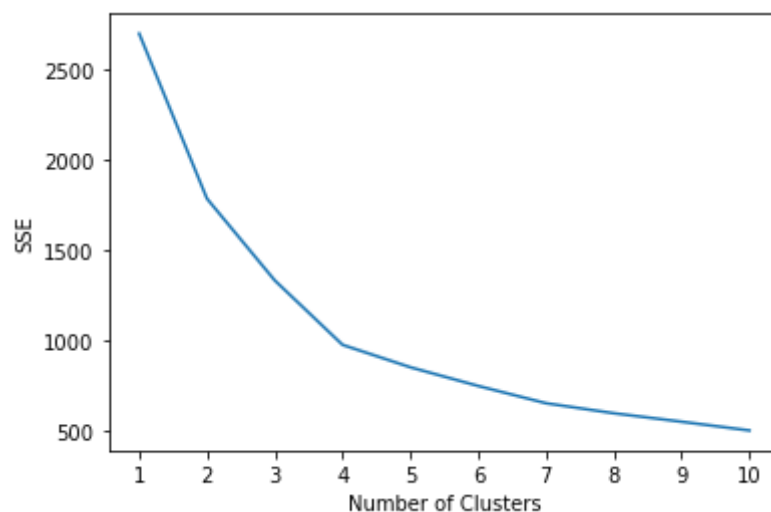
```
In [324]: bad_data[bad_data.cluster==2]

Out[324]: 63 rows x 20 columns
```

Fonte: Autoria própria

Para a próxima análise, os parâmetros de minutos de uso, *downlink throughput* e RSRP foram usados no sistema. Buscava-se aqui, uma redução do grupo de insatisfeitos por meio do acréscimo de mais um terceiro parâmetro para assim aumentar a correlação do estudo com a pesquisa de satisfação e validar a hipótese. Foi encontrado, neste caso, que o número ideal de clusters é 4, como pode ser observado no gráfico da Figura 46.

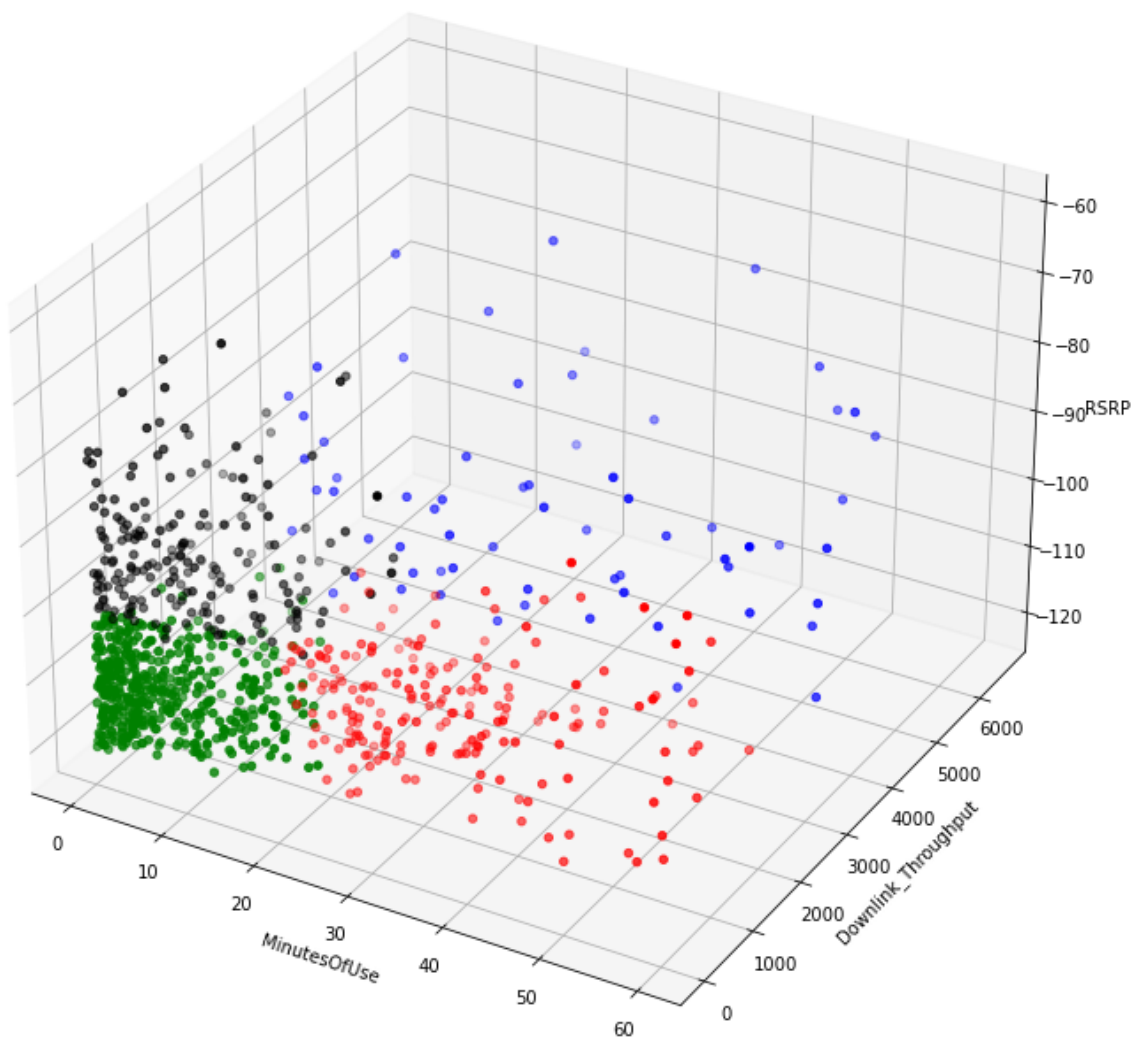
Figura 46 – Gráfico gerado por meio do *Elbow Technique* para definir o número de agrupamentos do sistema



Fonte: Autoria própria

Assim como mostrado anteriormente, o sistema foi dividido de forma gráfica pelo próprio programa, o resultado obtido pode ser observado a seguir, Figura 47.

Figura 47 - Divisão computadorizada feita do sistema para *downlink throughput*, minutos de uso e RSRP



Fonte: Autoria própria

Neste caso, o grupo teoricamente mais insatisfeito é o vermelho pois este foi exposto há um cenário de RSRP baixo e apesar de longo período de tempo conectado, não foi capaz de atingir valores elevados de *downlink throughput*.

As mesmas análises foram realizadas para a validação dessa hipótese. Assim, os usuários expostos as condições de rádio do grupo vermelho representam 21,75% da amostra, como observado na Figura 48. Este valor, diferente do que era imaginado, não diminuiu ao acrescentar um terceiro parâmetro na análise.

Figura 48 – Fragmento do código exibindo a quantidade de usuários do sistema que foram classificados como pertencentes ao agrupamento 1

```
In [256]: df_cluster[df_cluster.cluster==1]
```

Out[256]:

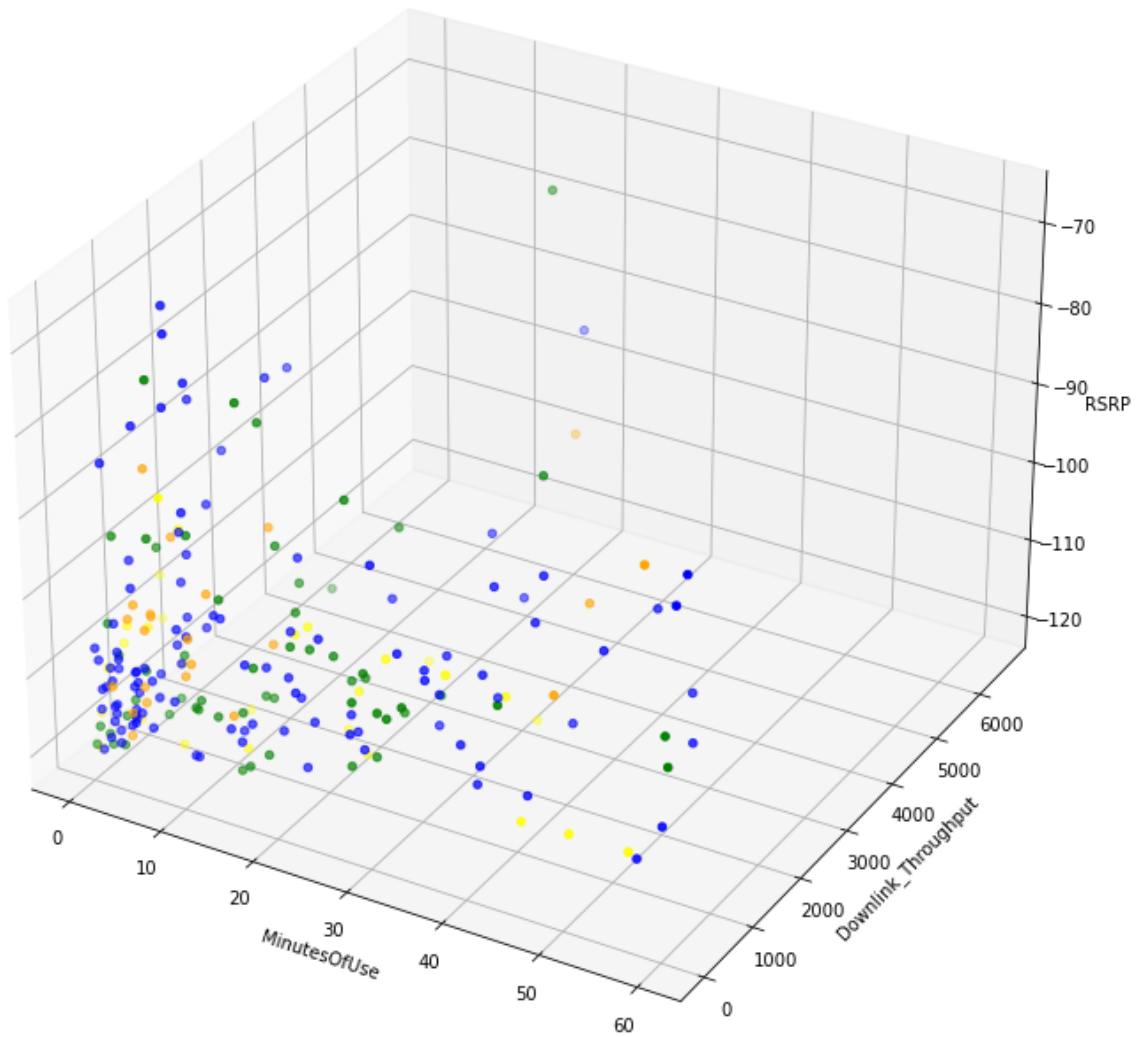
	MinutesOfUse	RSRP	Downlink_Throughput	cluster
9	57.888	-110.866250	2200.457500	1
15	45.888	-106.250000	1587.842917	1
18	18.066	-109.291667	1033.292373	1
19	17.763	-103.826667	1712.429205	1
34	58.484	-89.145833	78.034167	1
...	...	...	...	...
849	59.021	-108.236250	886.193750	1
850	55.806	-106.746458	1556.615625	1
869	35.138	-112.080417	1288.025625	1
877	25.181	-111.381250	2028.946250	1
879	33.363	-105.476250	1638.997917	1

196 rows x 4 columns

Fonte: Autoria própria

Novamente, ao observar a composição dos usuários dos mais baixos níveis de satisfação, Figura 49, eles foram expostos a cenários diferentes dos adotados na hipótese.

Figura 49 – Posicionamento gráfico dos grupos de mais baixa satisfação na pesquisa em relação ao *downlink throughput*, minutos de uso e RSRP



Fonte: Autoria própria

O percentual de sobreposição destes dois grupos, o grupo criado pelo cluster e o grupo autodeclaração insatisfeito na pesquisa, segue apresentando um baixo resultado, de apenas 29,09% nesse caso, como observado nas Figuras 50 e 51.

Figura 50 – Fragmento do código exibindo a quantidade total de usuários no grande grupo dos insatisfeitos

```
In [298]: bad_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 220 entries, 0 to 219
```

Fonte: Autoria própria

Figura 51 – Fragmento do código exibindo a quantidade de usuários no grande grupo dos insatisfeitos que foram classificados como pertencentes ao agrupamento 1

```
In [311]: bad_data[bad_data.cluster==1]

Out[311]: 64 rows x 20 columns
```

Fonte: Autoria própria

Como nenhuma das hipóteses criadas por meio dos estudos usando os dados de LTE foi validada, optou-se por analisar os dados de 3G para a criação dos perfis de satisfação.

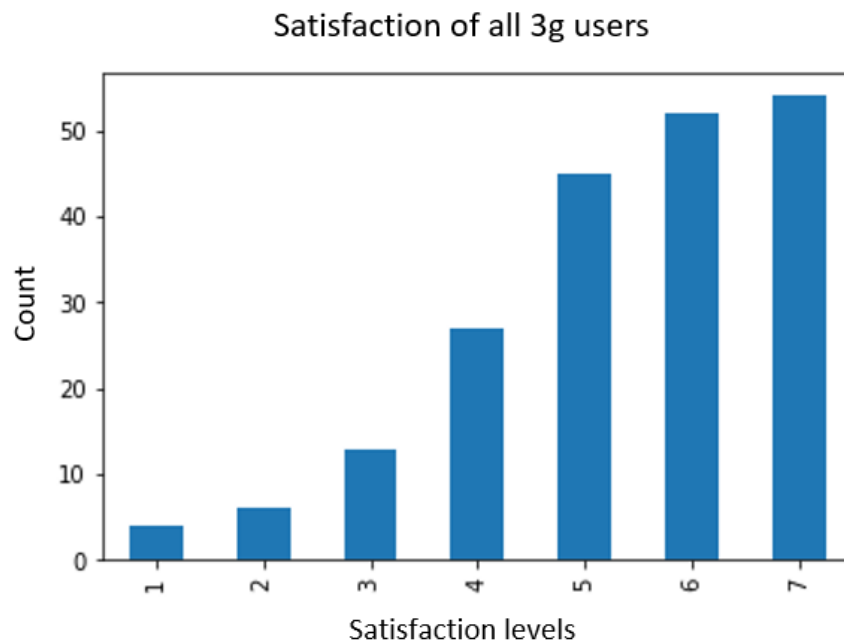
Dessa forma, as tabelas de 3G foram analisadas utilizando os dados de minutos de uso de 3G e quantidade de conexões a esse tipo de tecnologia. Neste estudo foram considerados apenas os registros dos usuários com valores diferentes de zero para volume de *throughput*, pois busca-se aqui os dados dos usuários que realmente estiveram ativos utilizando este tipo de tecnologia, assim esses usuários são realmente impactados pela troca de tecnologia na conexão.

O esperado neste estudo é que os usuários que mais utilizaram o 3G sejam os mais descontentes, visto que esta é uma tecnologia mais antiga que permite atingir velocidades menores no envio e recebimento de dados.

Dos bancos de dados de 3G, 181 usuários que responderam à pesquisa de satisfação foram encontrados, 20 deles usavam tanto 3G quanto 4G e os demais utilizam apenas a tecnologia 3G.

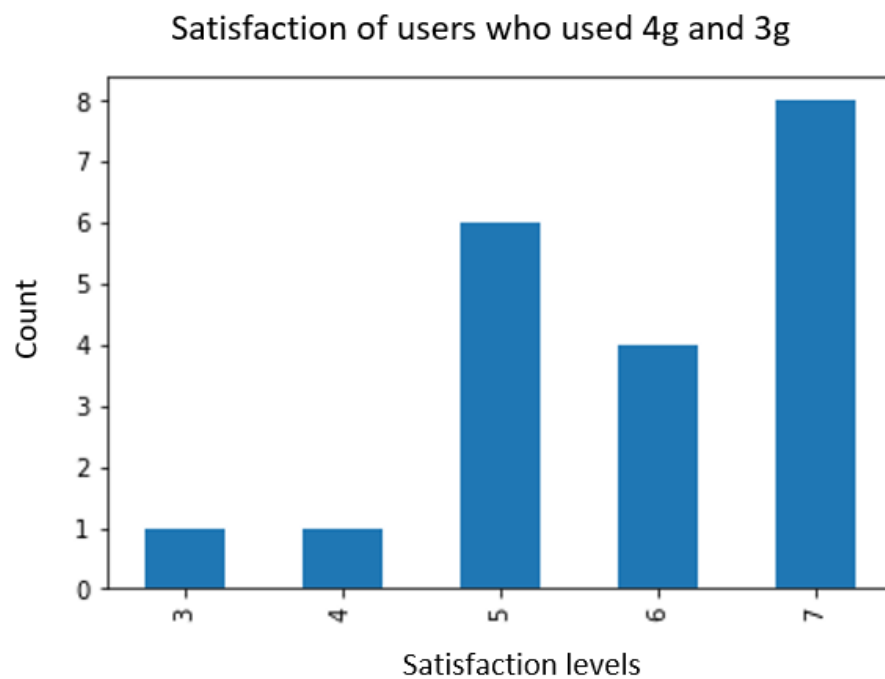
Os gráficos abaixo, Figuras 52 e 53, apresentam os números desses usuários em cada um dos níveis de satisfação.

Figura 52 - Gráfico de contadores de usuários por níveis de satisfação para todos os usuários que estiveram conectados no 3G



Fonte: Autoria própria

Figura 53 - Gráfico de contadores de usuários por níveis de satisfação para todos os usuários que estiveram conectados tanto no 3G quanto no 4G



Fonte: Autoria própria

Como pode ser observado, os usuários que utilizaram o 3G apresentam bons níveis de satisfação quanto aos serviços da operadora, invalidando novamente a hipótese criada.

Uma vez que ao analisar os dados, não foi possível encontrar um padrão comportamental entre os usuários que se autodeclararam insatisfeitos com os serviços da operadora, não puderam ser estabelecidos os limites das condições de rádio que dividem o contentamento do usuário. Não havendo limites claramente estabelecidos, não foi possível a criação dos perfis de satisfação assim como os demais passos do estudo.

5 CONSIDERAÇÕES FINAIS

A identificação dos *churn users* não pôde ser realizada, uma vez que o projeto foi interrompido na etapa de criação dos perfis de satisfação. A dificuldade de encontrar padrões de rádio que definissem os níveis de satisfação abre precedente para algumas afirmações.

É questionada a relevância das pesquisas de satisfação realizadas pelas operadoras de telefonia. Não se sabe a real pertinência dos dados lá encontrados pois estes são autodeclarações subjetivas dos próprios clientes sobre suas experiências como usuários. Não se sabe, além disso, a seriedade dos usuários ao responder essas pesquisas ou o que os motivou a isso. Para uma pesquisa de satisfação mais sólida é importante a adoção de perguntas mais objetivas e opções de respostas que façam retratos mais próximos da experiência dos usuários.

Outro possível cenário é a baixa pertinência dos parâmetros analisados neste estudo na experiência real do usuário. O UE dentro da área de cobertura, questiona-se até que ponto a intensidade do sinal por ele recebido influencia na satisfação do usuário durante suas atividades cotidianas com as redes móveis. Sendo assim, é imprescindível que demais estudos sejam realizados a partir de diferentes KPIs.

Por fim, percebe-se que os usuários com maior acesso a tecnologias anteriores, como o 3G, que em algumas regiões ainda tem uma maior cobertura, apresentam altos níveis de satisfação. Muito disso deve-se a utilização menos expressiva dos serviços de rede por parte desses usuários, que apesar de terem acesso a planos de voz e dados, utilizam mais as chamadas. Nos serviços de voz, quando comparadas as tecnologias 3G e 4G, encontram-se performances parecidas, o que justifica os níveis de satisfação encontrados.

REFERÊNCIAS

- AFROZ, F. et al. SINR, RSRP, RSSI and RSRQ Measurements in Long Term Evolution Networks. **International journal of wireless and mobile networks**, v. 7, n. 4, p. 113–123, 2015.
- AGIWAL, M.; ROY, A.; SAXENA, N. Next generation 5G wireless networks: A comprehensive survey. **IEEE Communications Surveys & Tutorials**, v. 18, n. 3, p. 1617–1655, 2016.
- BATALHA, I. et al. Análise de Throughput para 5G NSA com Variação dos Parâmetros RSRP e SINR. *In*: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS, 40., 2022, Santa Rita do Sapucaí. **Anais [...]**. Santa Rita do Sapucaí: Sociedade Brasileira de Telecomunicações, 2022. Disponível em: <https://biblioteca.sbrrt.org.br/articles/3552>. Acesso em: 2 jan. 2023.
- BOUTIBA, K.; BAGAA, M.; KSENTINI, A. Radio link failure prediction in 5G networks. *In*: IEEE GLOBAL COMMUNICATIONS CONFERENCE (GLOBECOM) 1., 2021, Madrid. **Anais [...]**. Madrid: Global Communications Conference, 2021. Disponível em: <https://www.eurecom.fr/en/publication/6672>. Acesso em: 2 jan. 2023.
- BRAGA, L. **4G/LTE: saiba como o 4G funciona**. Disponível em: <https://tecnoblog.net/responde/lte-4g-como-funciona/>. Acesso em: 3 jan. 2023.
- GRIJPKIN, F. et al. **Connected world: An evolution in connectivity beyond the 5G revolution**. 2020. Disponível em: <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/connected-world-an-evolution-in-connectivity-beyond-the-5g-revolution>. Acesso em: 4 jan. 2023.
- COX, C. **An introduction to 5G - the new radio, 5G network and beyond an introduction to 5G - the new radio, 5G network and beyond**: The new radio, 5G network and beyond. Hoboken, NJ, USA: Wiley-Blackwell, 2021.
- CRANDALL, R. E. **5G implementation around the world**, 2021. Disponível em: <https://www.ascm.org/ascm-insights/5g-implementation-around-the-world/>. Acesso em: 5 jan. 2023.
- DANGETI, P. **Statistics for Machine Learning -**: Essential statistical concepts for exploring predictive analytics and machine learning using Python and R. 2. ed. Birmingham, England: Packt Publishing, 2019.
- FURHT, B.; AHSON, S. A. (EDS.). **Long term evolution: 3GPP LTE radio and cellular technology**. Londres, England: CRC Press, 2019.
- GHOSH, A.; ZHANG, J. **Fundamentals of Lte**. [s.l.] Prentice Hall, 2011.
- LIKAS, A.; VLASSIS, N.; J. VERBEEK, J. The global k-means clustering algorithm. **Pattern recognition**, v. 36, n. 2, p. 451–461, 2003.

MOBILE COMMUNICATIONS - STATISTICS & FACTS. Disponível em: <https://www.statista.com/topics/1147/mobile-communications/>. Acesso em: 17 jan. 2023.

PEREZ, A. Mobile network signaling protocols. *In: Implementing IP and Ethernet on the 4G Mobile Network*. [s.l.] Elsevier, 2017. p. 15–44.

PUTTONEN, J. et al. Coverage optimization for Minimization of Drive Tests in LTE with extended RLF reporting. *In: ANNUAL IEEE INTERNATIONAL SYMPOSIUM ON PERSONAL, INDOOR AND MOBILE RADIO COMMUNICATIONS*, 21., 2010, Istanbul. **Anais [...]**. Istanbul: IEEE, 2010. Disponível em: <https://ieeexplore.ieee.org/document/5671907>. Acesso em: 2 jan. 2023.

RAMOS, D. **Cómo convertir los datos oscuros en oro para tu negocio. Emprendedores, mide mejor tu negocio**. Espanha, 2018. Disponível em: <https://www.pressreader.com/spain/emprendedores/20180223/284807876553146>. Acesso em: 9 ago. 2022.

RODRIGUES, C. F. R. **Avaliação da Qualidade e Desempenho de Serviços em Redes LTE**. Lisboa: Instituto Politécnico de Lisboa, nov. 2012.

SHARMA, P. **K means clustering simplified in python**. Disponível em: <https://www.analyticsvidhya.com/blog/2021/04/k-means-clustering-simplified-in-python/>. Acesso em: 8 jan. 2023.

SILVA, L. B. C.; CHAMON, M. A. **Cálculo de enlace para sistemas de comunicação digital via satélite**. São José dos Campos, SP: INPE, 2004. 49 p.

STROSKI, P. N. **Gerações de comunicações móveis**, 2017. Disponível em: <https://www.electricalibrary.com/2017/09/12/geracoes-de-comunicacoes-moveis/>. Acesso em: 3 jan. 2023.

SYAKUR, M. A. et al. Integration K-means clustering method and elbow method for identification of the best customer profile cluster. **IOP conference series. Materials science and engineering**, v. 336, p. 012017, 2018.

TAN, A. **Dark data is contributing to carbon emissions**. Disponível em: <https://www.computerweekly.com/news/252482121/Dark-data-is-contributing-to-carbon-emissions>. Acesso em: 8 ago. 2022.

TWILIO DOCS. Disponível em: www.twilio.com/docs/iot/supersim/how-determine-good-cellular-signal-strength. Acesso em: 2 jan. 2023.

YUE, C. et al. LinkForecast: Cellular link bandwidth prediction in LTE networks. **IEEE transactions on mobile computing**, v. 17, n. 7, p. 1582–1594, 2018.

BIBLIOGRAFIA CONSULTADA

AGRAWAL, J.; MOR, P.; KELLER, J. M.; DUBEY, P. **LTE: User Equipment (UE) States**. IARJSET International Advanced Research Journal in Science, Engineering and Technology v. 3, 2016

Dalla prima alla quinta generazione, il passato e il futuro degli standard di telecomunicazione. Disponível em: <<https://www.sicomtesting.com/pt/blog/dal-1g-al-5g-il-passato-e-il-futuro-degli-standard-gsm-umts-hspa-ed-lte/>>. Acesso em: 9 jan. 2023.

HSPA e WiMax Móvel I: Tecnologias. Disponível em: <https://www.teleco.com.br/tutoriais/tutorialhspawimax1/pagina_2.asp>. Acesso em: 10 jan. 2023.

K-Means Clustering – What it is and How it Works – Learn by Marketing. Disponível em: <<https://www.learnbymarketing.com/methods/k-means-clustering/>>. Acesso em: 10 jan. 2023.

Radio Link Failure and Out of Synch. Disponível em: <<https://gotechnicalnow.com/radio-link-failure-and-out-of-synch/>>. Acesso em: 9 jan. 2023.

Tudo sobre dark data e seus riscos para a empresa. Disponível em: <<https://www.infonova.com.br/>>. Acesso em: 8 ago. 2022.