

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”  
FACULDADE DE ENGENHARIA  
CAMPUS DE ILHA SOLTEIRA**

**Flávio Menegueço Bezerra**

**Estudo comparativo entre o número de pessoas transportadas pela  
SPTrans e o número de casos de Covid 19 no município de São Paulo no  
ano de 2020**

**CURSO DE GRADUAÇÃO EM ENGENHARIA CIVIL**

**Flávio Menegueço Bezerra**

**Estudo comparativo entre o número de pessoas transportadas pela SPTrans e o número de casos de Covid 19 no município de São Paulo no ano de 2020**

Trabalho de Conclusão de Curso,  
apresentada à Faculdade de  
Engenharia de Ilha Solteira – UNESP  
como parte dos requisitos para  
obtenção do título de Bacharel em  
Engenharia Civil.

Profa. Dra. Luzenira Alves  
Brasileiro  
**Orientadora**

FICHA CATALOGRÁFICA  
Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

B574e Bezerra, Flavio Menegueço.  
Estudo comparativo entre o número de pessoas transportadas pela SPTrans e o número de casos de Covid 19 no município de São Paulo no ano de 2020 / Flavio Menegueço Bezerra. -- Ilha Solteira: [s.n.], 2021  
61 f. : il.

Trabalho de conclusão de curso (Graduação em Engenharia Civil) -  
Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira, 2021

Orientador: Luzenira Alves Brasileiro  
Inclui bibliografia

1. SPTrans. 2. Covid 19. 3. Modelos regressivos. 4. Série histórica.

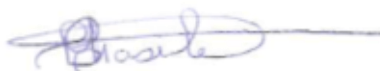
  
Raiane da Silva Santos  
Supervisora Técnica de Seção  
Seção Técnica de Referência, Atendimento ao usuário e Documentação  
Diretoria Técnica de Biblioteca e Documentação  
CRB/8 - 9999

## CERTIFICADO DE APROVAÇÃO

TÍTULO: Estudo comparativo entre o número de pessoas transportadas pela SPTrans e o número de casos de Covid 19 no município de São Paulo no ano de 2020

**AUTOR:** Flávio Menegueço Bezerra

**ORIENTADORA:** Profa. Dra. Luzenira Alves Brasileiro



Profª. Drª. Luzenira Alves Brasileiro  
UNESP – Campus de Ilha Solteira (Orientadora)



Prof. Dra. Maria da Consolação Fonseca de Albuquerque  
UNESP – Campus de Ilha Solteira



Prof. Dr. Marco Antônio de Moraes Alcântara  
UNESP – Campus de Ilha Solteira

**Ilha Solteira**  
**09/12/2021**

## AGRADECIMENTOS

Agradeço a UNESP e a todo o corpo docente pela oportunidade de estudar em uma das grandes faculdades brasileiras e por ter me preparado como um profissional com excelente bagagem técnica e teórica.

Agradeço a minha família por todo o suporte durante o período da graduação. Em especial pela minha mãe, que me ajudou e me deu suporte, mesmo em momentos que eu estava errado.

Agradeço também ao meu avô "*in memoriam*" pelo seu esforço. Sem ele, muito provavelmente eu não seria o primeiro da família a me graduar em uma das faculdades mais importantes do Brasil.

E por fim, mas não menos importante, agradeço a todos os moradores da Eterna República Moita por todo o convívio durante o período da graduação. Levarei as histórias para toda a minha vida. Foi lá que tive os meus maiores fracassos e as minhas maiores vitórias, mas que foi por meio delas que me tornei a pessoa que sou hoje.

“Estude muito o que mais lhe interessa da maneira mais indisciplinada, irreverente e original possível”

**Richard Feynman**

## RESUMO

No ano de 2020, as sociedades ao redor do mundo vivenciaram um desafio que será lembrado pelas gerações futuras como a pandemia de COVID 19. Independentemente do tipo de estrutura social, ficou evidente o papel da gestão pública como um elemento relevante na desaceleração do crescimento exponencial de contágio.

De forma que, os principais países com melhores performances na redução da proliferação do vírus, foram conseqüentemente aqueles que mais se dispuseram a tomar medidas impopulares, como distanciamento social e fechamento de locais com características de alta taxa de aglomerações.

Entretanto, mesmo com o poder público disposto a tomar medidas mais drásticas, em muitos casos se mostrou necessário que a população estivesse também de acordo com tais medidas, com propensão a aderir a essas práticas públicas.

Desta forma, este trabalho tem como objetivo propor análises de tomada de decisões baseadas em ferramentas estatísticas, como testes de hipóteses de séries históricas e regressões lineares, sobre o transporte urbano no município de São Paulo realizado pela SPTrans e o número de casos de Covid 19 no ano de 2020.

Esta análise é importante a fim de qualificar as ações públicas dispostas a reduzir a curva de contágio, com uma certa flexibilização, ou seja, sem desativar totalmente as principais frotas de ônibus circulantes. Desta forma, a gestão, além de ser mais técnica, ganha um elemento mais transparente sobre a população, que terá mais propensão a aderir a medidas protetivas.

**Palavras-chave:** SPTrans, Covid 19, modelos regressivos, série histórica.

## **ABSTRACT**

In the year of 2020, societies all around the world experienced a challenge which will be remembered by many future generations, the COVID 19 pandemic. Regardless of the social structure, the role of public management as a relevant element in the slowdown of the exponential growth of the disease became evident.

As a result, the first countries with the best reduction performances were consequently those that were most willing to take unpopular measures, such as social distancing and closing down locations that have high agglomeration rate.

However, even with the government willing to take more drastic measures, in many cases it proved necessary that the population also agreed with such measures, with a tendency to adhere to these public practices.

Thus, this work aims to propose analysis of decision-making based on statistical tools, such as hypothesis tests on time series and linear regressions, about urban transport in the city of São Paulo carried out by SPTrans and the number of Covid 19 cases in 2020.

This analysis is important in order to qualify public actions willing to reduce the contagion curve, with a certain flexibility, that is, without completely deactivating the main fleets of circulating buses. In this way, the management, in addition to being more technical, gains more transparency about the population, who will be more likely to adhere to protective measures.

**Palavras-chave:** SPTrans, Covid 19, linear regressions, time series.



## Índice de Figuras:

FIGURA 1.1 - EFEITO DAS DIVERSAS VARIÁVEIS QUANDO APLICADAS A MODELOS ESTATÍSTICOS .....	2
FIGURA 1.2 – MAPA DE ACESSIBILIDADE DE TRANSPORTES PÚBLICOS NO MUNICÍPIO DE SÃO PAULO.....	6
FIGURA 1.3 – PROPORÇÃO DE UTILIZAÇÃO DE DIFERENTES MODAIS NO MUNICÍPIO DE SÃO PAULO.....	6
FIGURA 1.4 – MAPA COM A DISPOSIÇÃO DAS DIFERENTES LINHAS DE METRÔ NO MUNICÍPIO DE SÃO PAULO .....	7
FIGURA 1.5 – MAPA COM A DISPOSIÇÃO DE RENDA SOBRE O MUNICÍPIO DE SÃO PAULO .....	8
FIGURA 1.6 - ALTERAÇÃO NA ESCOLHA DO TRANSPORTE COM O DECORRER DO TEMPO .....	9
FIGURA 1.7 - DESCRITIVO DO PORQUÊ HOUE UMA MAIOR FREQUÊNCIA NA UTILIZAÇÃO DO AUTOMOVEM EM 2020 .....	10
FIGURA 1.8 - DESCRITIVO DO PORQUÊ HOUE UMA MENOR FREQUÊNCIA NA UTILIZAÇÃO DO AUTOMOVEM EM 2020 .....	10
FIGURA 3.1 - FORMATO DE PROCESSOS ESTOCASTICOS AO LONGO DO TEMPO .....	12
FIGURA 4.1 – REPRESENTAÇÃO DA DIVISÃO FEITA PELA SPTRANS DA CIDADE DE SÃO PAULO EM DISTRITOS .....	34

## **Índice de Tabelas:**

TABELA 4.1 – RESULTADO DO TESTE ADF PARA OS DADOS BRUTOS DOS CASOS DE COVID .....	26
TABELA 4.2 - RESULTADO DO TESTE ADF PARA OS DADOS DE CASOS DE COVID APÓS O PROCESSO DE DIFERENCIAÇÃO.....	27
TABELA 4.3 - RESULTADO DO TESTE ADF PARA OS DADOS BRUTOS DE PASSAGEIROS.....	28
TABELA 4.4- RESULTADO DO TESTE ADF PARA OS DADOS DE PASSAGEIROS APÓS O PROCESSO DE DIFERENCIAÇÃO.....	29
TABELA 4.5 – RESULTADOS DO TESTE DE GRANGER PARA DIFERENTES INTERVALOS .....	31
TABELA 4.6 - DISTRIBUIÇÃO DOS PASSAGEIROS NO ANO DE 2020.....	35
TABELA 4.7 – RESULTADOS PARA OS DADOS BRUTOS.....	40
TABELA 4.8 – RESULTADOS DOS COEFICIENTES DO MODELO PARA OS DADOS BRUTOS .....	41
TABELA 4.9 - RESULTADOS PARA OS DADOS COM FILTRO DE TENDÊNCIA.....	43
TABELA 4.10 - RESULTADOS DOS COEFICIENTES DO MODELO PARA OS DADOS COM FILTRO DE TENDÊNCIA .....	43

## Índice de Gráficos:

GRÁFICO 4.1 - COMPARATIVO ENTRE A EVOLUÇÃO DE CASOS DE COVID COM O NÚMERO DE PASSAGEIROS SOBRE TODO O PERÍODO DE 2020 .....	21
GRÁFICO 4.2 - SÉRIE DO NÚMERO DE PASSAGEIROS COM DADOS BRUTOS.....	21
GRÁFICO 4.3 - DECOMPOSIÇÃO DA SÉRIE DO NÚMERO DE PASSAGEIROS .....	22
GRÁFICO 4.4 - COMPARAÇÃO ENTRE A SÉRIE COM DADOS BRUTOS E COM FILTRO DE TENDÊNCIA DO NÚMERO DE PASSAGEIROS .....	23
GRÁFICO 4.5 - SÉRIE COM DADOS BRUTOS DOS CASOS DE COVID .....	23
GRÁFICO 4.6 - DECOMPOSIÇÃO DA SÉRIE DE CASOS DE COVID.....	25
GRÁFICO 4.7 - COMPARAÇÃO ENTRE A SÉRIE COM DADOS BRUTOS E COM FILTRO DE TENDÊNCIA DOS CASOS DE COVID .....	25
GRÁFICO 4.8 – SÉRIE DO NÚMERO DE CASOS DE COVID DECORRENTE DE UMA DIFERENCIAÇÃO.....	27
GRÁFICO 4.9 – SÉRIE DO NÚMERO DE CASOS DE COVID DECORRENTE DE UMA DIFERENCIAÇÃO.....	28
GRÁFICO 4.10 - RESULTADO DO FEEDBACK CONSIDERANDO O VALOR DE P-VALUE.....	32
GRÁFICO 4.11 - DISTRIBUIÇÃO DOS PASSAGEIROS NO ANO DE 2020 .....	35
GRÁFICO 4.12 - DISTRIBUIÇÃO DOS PASSAGEIROS AO LONGO DO ANO DE 2020 .....	36
GRÁFICO 4.14 – SÉRIE DE DADOS FILTRADOS PELA TENDÊNCIA DE CASOS DE COVID.....	38
GRÁFICO 4.13–SÉRIE DE DADOS BRUTOS DE CASOS DE COVID.....	38
GRÁFICO 4.15 – SÉRIE DE DADOS BRUTOS DE PASSAGEIROS POR DISTRITO .....	39
GRÁFICO 4.16 – SÉRIE DE DADOS FILTRADOS PELA TENDÊNCIA DE PASSAGEIROS POR DISTRITO .....	39
GRÁFICO 4.17- CORRELAÇÃO ENTRE O VALOR REAL COM O VALOR AJUSTADO PARA OS DADOS BRUTOS .....	41
GRÁFICO 4.18 – RESÍDUOS PARA OS DADOS BRUTOS.....	42
GRÁFICO 4.19 - CORRELAÇÃO ENTRE O VALOR REAL COM O VALOR AJUSTADO PARA OS DADOS COM FILTRO DE TENDÊNCIA.....	43
GRÁFICO 4.20 - RESÍDUOS PARA OS DADOS COM FILTRO DE TENDÊNCIA.....	44

## Sumário

<b>1. INTRODUÇÃO</b> .....	<b>1</b>
1.1. A IMPORTÂNCIA DA ESTRUTURA URBANA.....	3
1.2. A DESIGUALDADE SOCIOESPACIAL VISTA NO TRANSPORTE PÚBLICO.....	4
1.2.1. <i>Cenário da mobilidade na região metropolitana paulista</i> .....	6
<b>2. OBJETIVOS</b> .....	<b>11</b>
<b>3. REVISÃO DE LITERATURA</b> .....	<b>12</b>
3.1. SÉRIES HISTÓRICAS.....	12
3.2. SÉRIES ESTACIONÁRIAS.....	14
3.2.1. <i>Metodologia via Raiz Unitária</i> .....	14
3.2.2. <i>Teste de Hipótese Augmented Dickey-Fuller</i> .....	16
3.2.3. <i>Método via Diferenciação</i> .....	16
3.3. SÉRIES COM CAUSALIDADE.....	17
3.3.1. <i>Teste de Causalidade (Granger)</i> .....	17
3.4. MODELOS ESTATÍSTICOS.....	18
3.4.1. <i>Regressão linear</i> .....	18
<b>4. METODOLOGIA E DESENVOLVIMENTO</b> .....	<b>19</b>
4.1. DADOS UTILIZADOS.....	19
4.1.1. <i>Dados dos passageiros de ônibus</i> .....	19
4.1.2. <i>Dados de contágios pela Covid 19</i> .....	20
4.2. MANIPULAÇÃO E FILTROS.....	20
4.2.1. <i>Análise dos dados de passageiros</i> .....	21
4.2.2. <i>Análise dos dados de Covid 19</i> .....	23
4.3. TESTES DE HIPÓTESES.....	26
4.3.1. <i>Teste ADF com Diferenciação</i> .....	26
4.3.2. <i>Teste de Granger</i> .....	30
4.4. REGRESSÃO LINEAR.....	34
4.4.1. <i>Consideração e estudo dos dados de transporte</i> .....	34
4.4.2. <i>Estudo sobre as variações semanais</i> .....	37
4.4.3. <i>Resultados dos modelos</i> .....	40
<b>5. CONCLUSÃO</b> .....	<b>47</b>
<b>6. REFERÊNCIAS</b> .....	<b>48</b>

## 1. INTRODUÇÃO

No ano de 2020, o mundo vivenciou a pandemia de SARS-CoV-2, causadora de um grande impacto em todo o globo terrestre. Ela modificou o comportamento social e econômico, fazendo com que governos adotassem medidas até então nunca vistas em ambiente global (SHEU, MCKAY e ERIC A. STORCH, 2020).

Os primeiros sintomas da doença e suas vias transmissoras eram desconhecidas num primeiro momento, sendo considerada a princípio como uma situação endêmica restrita ao território da China.

E com o tempo e com a adoção de registros de casos, a Covid 19 (O nome Covid é a junção de letras que se referem a *(co)rona (vi)rus (d)isease*) apresentou uma alta taxa de transmissibilidade, pois em um curto espaço de tempo o vírus se alastrou para outras regiões do planeta.

Contudo, em pouco tempo também pôde ser observado pelos pesquisadores da área de saúde que as principais formas de transmissão da Covid 19 se deram pelas vias aéreas, em locais com pouca circulação de ar e associado com a concentração de pessoas (SHEU, MCKAY e ERIC A. STORCH, 2020).

Diante desse fato, os governantes utilizaram medidas impopulares para tentar desacelerar a proliferação do vírus, tais como, impedir a aglomeração de pessoas, uso de profilaxias (higienização das mãos e uso de máscaras), culminando com a retratação da curva de contágio (LAVIE, BATTAT, *et al.*, 2021).

**Figura 1.1 - Efeito das diversas variáveis quando aplicadas a modelos estatísticos**



**Fonte: (BRAUNER, MINDERMANN, et al., 2021)**

No começo do cenário pandêmico, mesmo sem muitos estudos para auxiliar a gestão sanitária, muitas cidades orientais adotaram medidas restritivas de circulação da população, em que a longo prazo mostrou-se como um excelente bloqueador sobre o crescimento de contágios (GUERCHE-SÉBLAIN, CHAKIR, *et al.*, 2021).

Mesmo assim, com todos os atritos entre diferentes setores e nações internacionais, com relação a doença, se comprovou que a forma mais eficaz de se controlar o vírus é o distanciamento social (SHEN, PENG, *et al.*, 2020), que também se fez necessário entender como a cultura e a forma com que as cidades se distribuem implicariam com a dificuldade de se implementar as decisões para impedir aglomerações.

## 1.1. A importância da estrutura urbana

Considerando que o principal facilitador da transmissão da Covid é a aglomeração (SHEN, PENG, *et al.*, 2020), a estrutura das cidades, assim como as características urbanas no que diz respeito a moradia e a mobilidade se tornam fatores fundamentais no estudo da doença.

Neste sentido, na visão mais clássica, pode-se segregar as cidades, no que diz respeito a sua estrutura, em duas grandes classificações: cidades horizontais e cidades espalhadas (PRADHAN, 2017). No entanto, como o crescimento urbano decorre de uma maneira muitas vezes orgânica, adicionado ao fato de uma possível rotatividade de gestores públicos, com maior ou menor posicionamento sobre o meio urbano, as cidades modernas acabam tendo um misto entre essas duas classificações.

Um exemplo disso é o perímetro urbano de São Paulo, em que o centro possui característica estruturada horizontalmente. Isto significa que existe uma elevada concentração de prédios, com uma estrutura habitacional (rede de abastecimento, saneamento e serviços) muito bem desenvolvida e com uma média de deslocamento entre trabalho – casa menor do que comparado a outras regiões.

Este é um acontecimento que pode ser explicado pelo fato do centro acabar sendo também a região mais rica, com uma gama de possibilidades público/privada, como linhas de metrô e ônibus mais interligadas, abrigando uma porcentagem extremamente elevada da população com maior poder aquisitivo na região.

Tais características fazem com que seja automático correlacionar o centro de São Paulo com outras cidades com idêntica estrutura, como é o caso de Nova York. Isto, pois no caso novaiorquino, o trem tem um elevado papel como opção de transporte, visto que o custo de se trafegar com um veículo individual é extremamente elevado considerando o próprio custo de se manter o automóvel, assim como o preço da zona de estacionamentos (HARRIS, 2020).

Entretanto, como já explicitado anteriormente, grande parte das cidades não apresentam apenas uma estrutura, mas sim um conjunto de ambos. Este é o caso do entorno da região de São Paulo, quando se analisa o trajeto da população que reside na periferia metropolitana, e que se desloca diariamente para o trabalho, onde muitas vezes não é no mesmo bairro, tem-se uma média mais elevada de tempo de trajeto (LEIVA, REIS e FILHO, 2020).

Dessa forma, define-se este tipo de estrutura como cidades espraiadas, com a existência de regiões dormitórios. Isso pois, a região em que as pessoas residem não acaba tendo uma devida importância cultural como um espaço público, pois sua função social é apenas de moradia (PRADHAN, 2017).

Esse processo de expansão horizontal sem o devido estudo da função do local sobre a sociedade é marcante na história das cidades Latino-americanas, principalmente pela necessidade da população mais carente financeiramente buscar moradia, aliado também a necessidade de manter contato com os grandes polos comerciais ou industriais (LEIVA, REIS e FILHO, 2020).

E o interessante é que, analisando diferentes cidades ao longo da América Latina, por mais que existam processos históricos distintos, o processo urbano acabou acontecendo de maneira semelhante em muitas cidades: O centro com a concentração da economia local e a população trabalhadora morando em seu entorno.

Entretanto, é nesse momento que a região metropolitana acaba ganhando características diferenciadas. Isto pela história e importância do ônibus sobre a cidade durante a sua expansão horizontal, que se garantiu a estrutura pendular na rotina da população mais carente que trabalha no centro e reside em cidades no entorno da capital.

## **1.2. A desigualdade socioespacial vista no transporte público**

Como já descrito anteriormente, a cidade de São Paulo é caracterizada por sua complexidade urbana. Isso decorre pela sua diversidade cultural, assim como por ter uma alta desigualdade socioeconômica.

O interessante ao analisar a estrutura da cidade de São Paulo, observado historicamente, é que se caracterizou pela descentralização espacial da população mais carente para regiões mais afastadas e com pouca infraestrutura no local, deixando o centro, direcionado para os interesses monetários (indústria e comércio) ou para a moradia de uma população com maior poder aquisitivo.

Relevante a este processo é que, analisando os diferentes investimentos urbanos na região, a centralidade de investimentos públicos acaba também favorecendo as regiões centrais com melhor suporte urbano.

E, vale também ressaltar que é uma característica a importância do investimento público brasileiro no ambiente urbano. Isso, pois o ambiente privado tem baixa capilaridade de investimento no setor urbanístico da cidade, tendo seu esforço



mais direcionado em ambientes privados (regiões de condomínio privados ou ambientes direcionados ao comércio como shoppings centers).

Desta forma, considerando também a maior concentração do esforço público em regiões mais centrais com um aporte relevante de infraestrutura e uma ausência da presença privada no processo de urbanização, o cenário desigual com relação a estruturação de serviços como energia elétrica, água encanada, esgoto e vias de transporte, acabam sendo desiguais com relação ao território metropolitano de São Paulo.

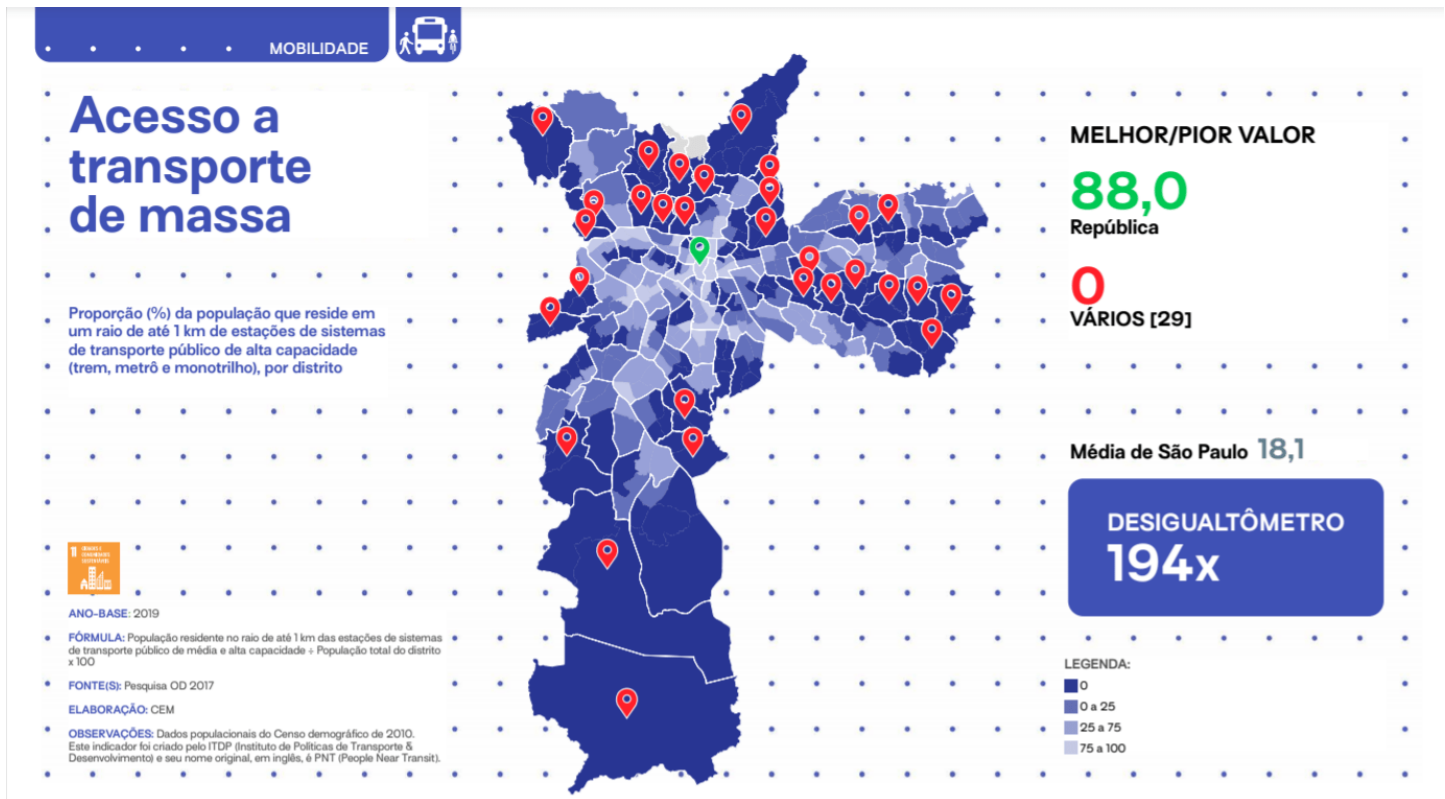
E, com relação ao transporte, a desigualdade acaba também aparecendo quando o centro concentra os melhores serviços públicos. Ou seja, a população que teria maior dependência e necessidade de um transporte para ir ao trabalho, em que o transporte individual tem um maior custo, fica prejudicada.

Além disso, o fato da concentração de investimento no centro acaba também indicando porque o ônibus tem maior utilização do que comparados a outros modais, flexibilidade de implementação.

### 1.2.1. Cenário da mobilidade na região metropolitana paulista

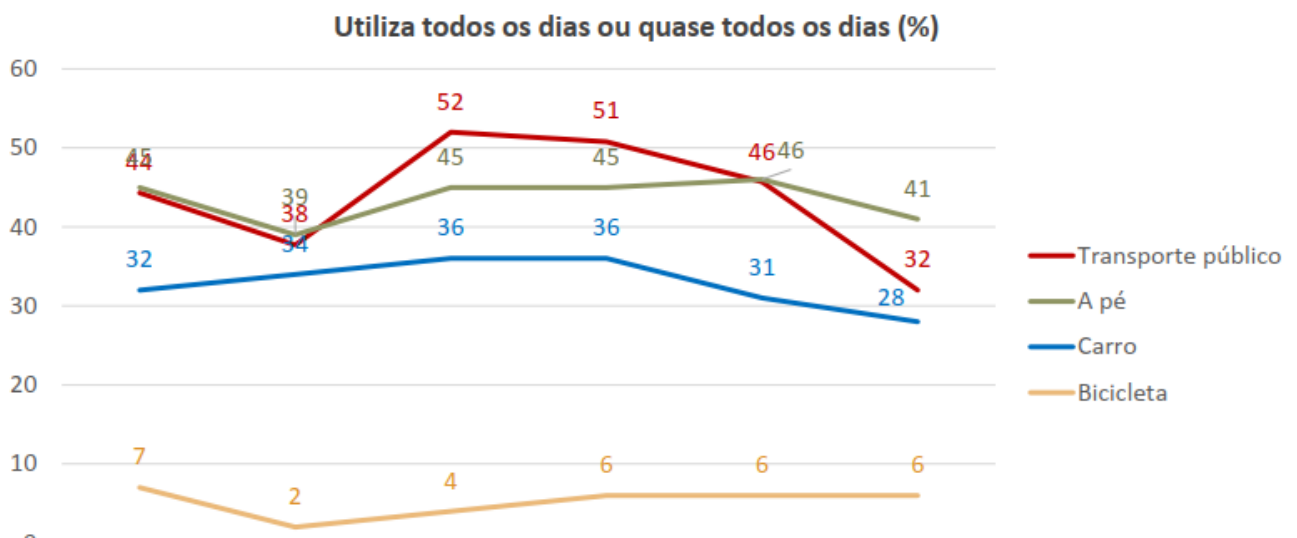
Analisando a realidade paulistana, existem três principais formas de se deslocar: o metrô/trem, o ônibus, o automóvel, a bicicleta e a pé.

**Figura 1.2 – mapa de acessibilidade de transportes públicos no município de São Paulo**



Fonte: (REDE NOVA SÃO PAULO, 2020)

**Figura 1.3 – Proporção de utilização de diferentes modais no município de São Paulo**

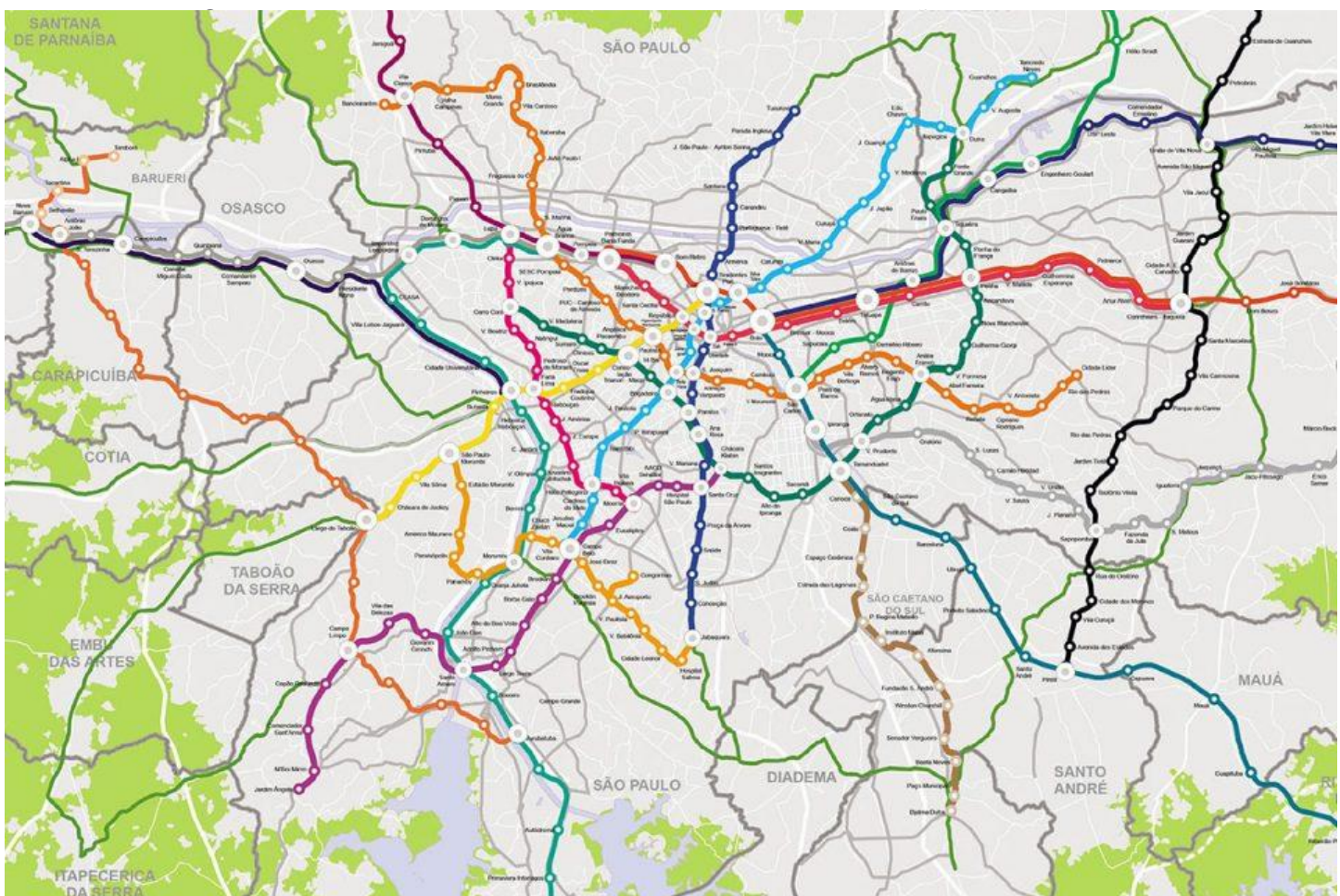


Fonte: (REDE NOSSA SÃO PAULO, 2020)

Cada um tem a sua vantagem e desvantagem. Entretanto, analisando de forma mais analítica, pode-se entender que, para fins diários de trabalho, o ônibus e o automóvel acabam sendo as principais formas de transporte.

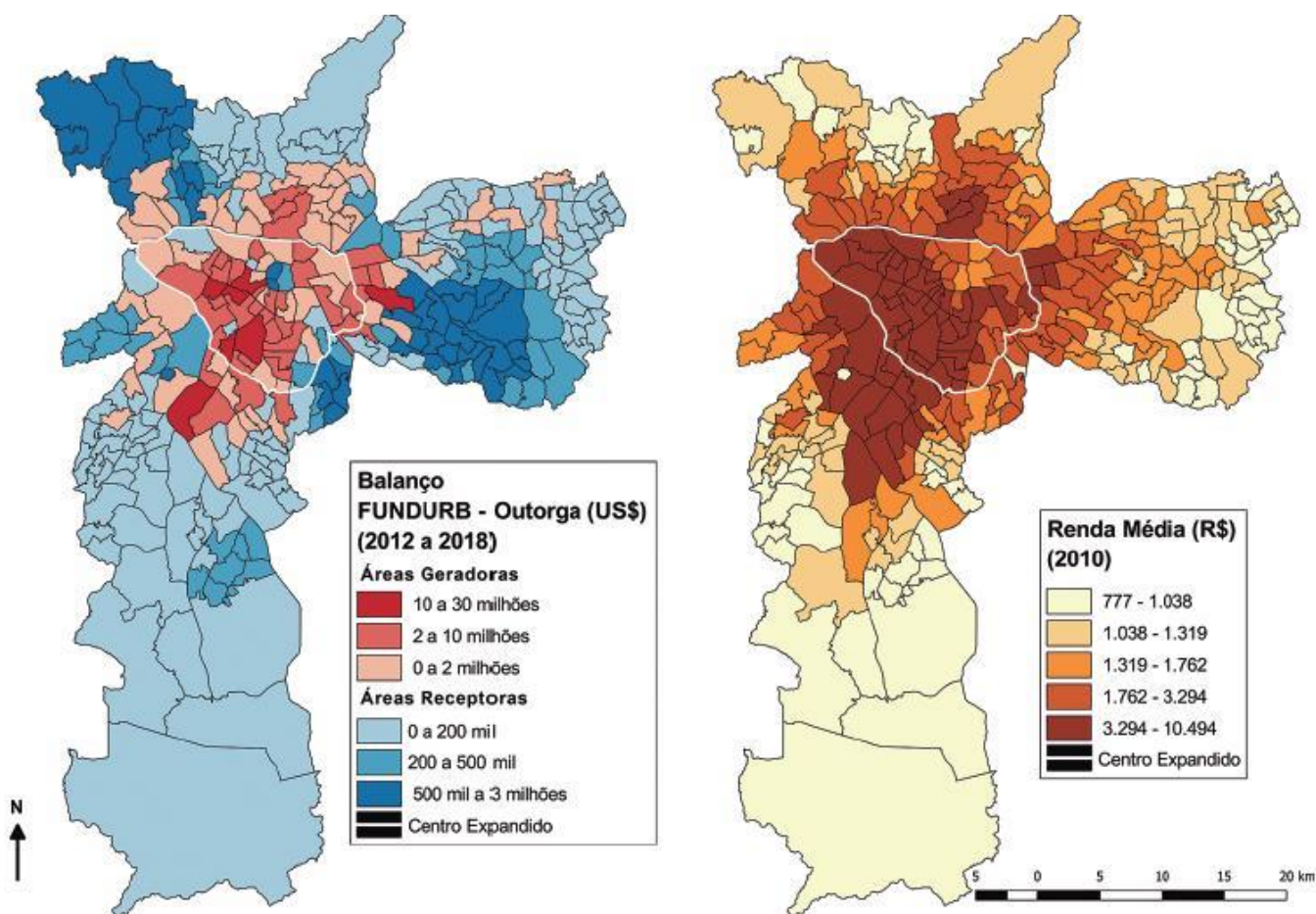
Isso pelo fato do metrô e do trem necessitarem de alta complexidade estrutural para sua implementação, assim como para sua utilização. E, analisando o cenário de investimento urbanístico, fica claro que pela necessidade de estrutura, este modal se concentra na região central da cidade.

**Figura 1.4 – Mapa com a disposição das diferentes linhas de metrô no município de São Paulo**



Fonte: (METRO CPTM, 2019)

**Figura 1.5 – Mapa com a disposição de renda sobre o município de São Paulo**



**Fonte: (LEITE, ACOSTA, et al., 2019)**

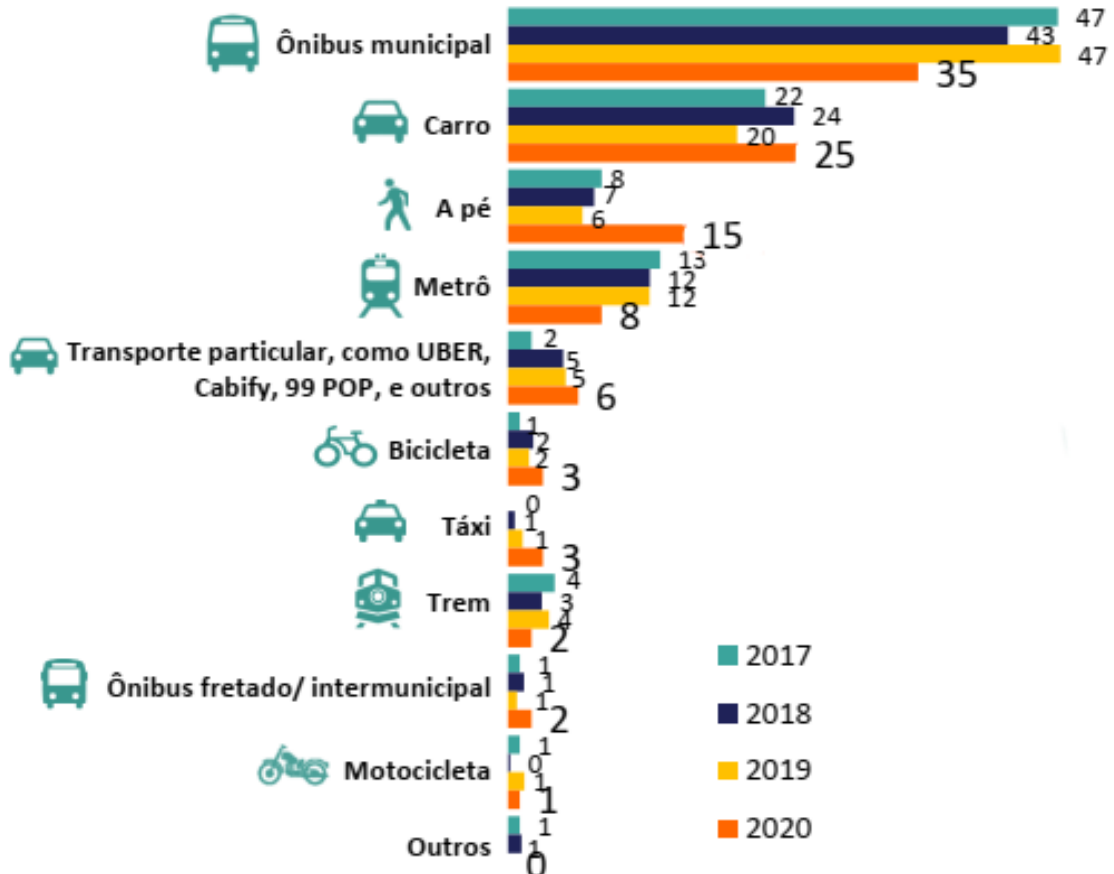
Portanto, o transporte acaba se restringindo a basicamente duas possibilidades: o transporte individual ou o ônibus.

No caso do automóvel, que é o transporte com maior custo individual, pela indivisibilidade dos custos ou impossibilidade de auxílio direto governamental, o ônibus acaba desempenhando um papel fundamental na vida e no funcionamento da cidade de São Paulo para a população mais afastada do centro.

E o interessante é que, mesmo com o ambiente pandêmico, e a redução das linhas de ônibus, a utilização deste modal continuou muito relevante em questões percentuais. Isso não significa que inexistiu uma diminuição com relação a proporção de pessoas que utilizaram o ônibus no período pandêmico.

Muito deste dado advém do fato da adequação por diversas empresas pelo *home office*, em que muitos trajetos antes feitos por carros individuais foram extintos, diminuindo assim o trânsito nas grandes vias.

**Figura 1.6 - Alteração na escolha do transporte com o decorrer do tempo**



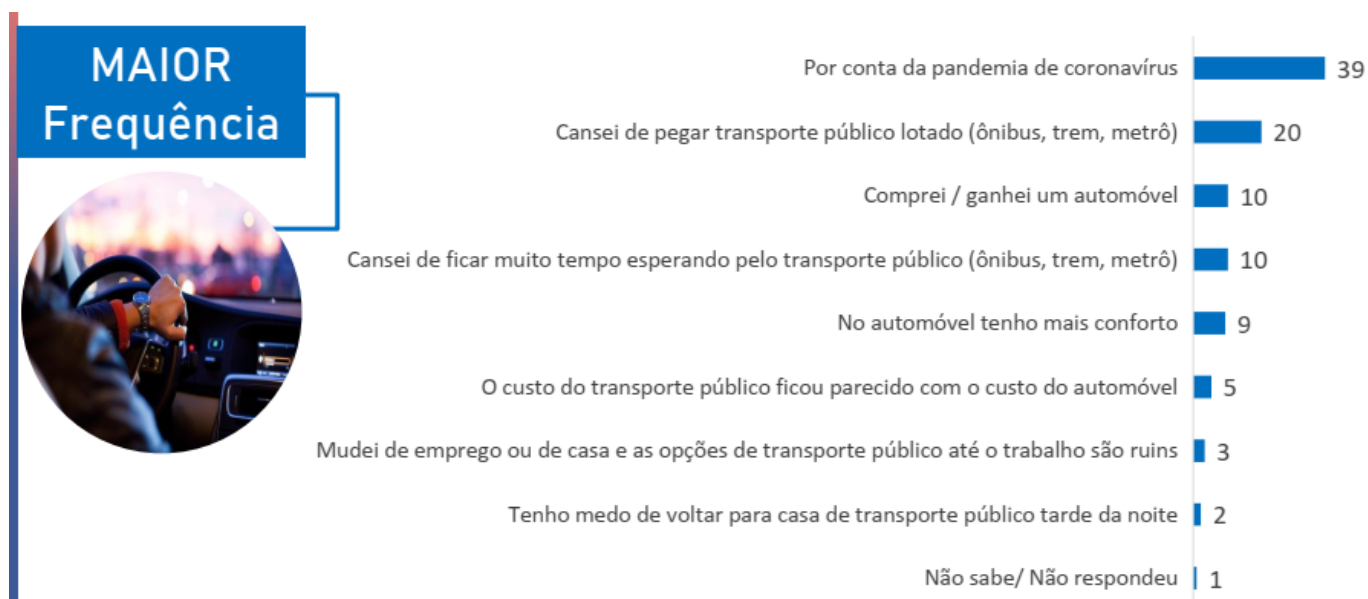
**Fonte: (REDE NOSSA SÃO PAULO, 2020)**

Agora, analisando apenas as variáveis sobre o cenário pandêmico, o automóvel individual é a melhor opção possível, visando apenas questões sanitárias.

E, por conta do ônibus ter a característica de aglomeração adicionado ao fato de que ele continuou sendo utilizado ao longo da pandemia por uma elevada parcela da população, culminou com a possibilidade do ônibus ter sido um impacto extremamente relevante para o desenvolvimento das curvas de contaminação ao longo do tempo.

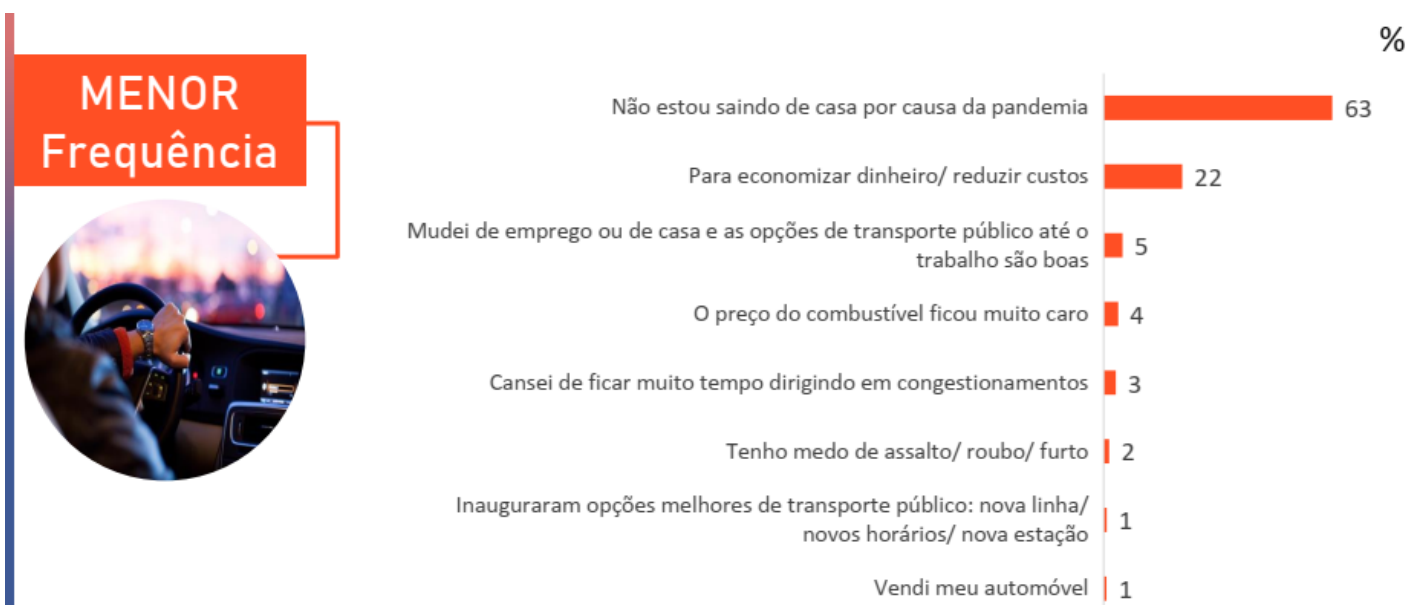
E este fato é comprovado também quando se compara as diferentes medidas governamentais ao redor do mundo, indicando o quão importante é a restrição não apenas de locomoção individual, mas também da restrição do transporte público para a desaceleração da curva de contaminação. Um exemplo claro disso foi Nova York, com relação ao metro (HARRIS, 2020).

**Figura 1.7 - Descritivo do porquê houve uma maior frequência na utilização do automóvel em 2020**



**Fonte: (REDE NOSSA SÃO PAULO, 2020)**

**Figura 1.8 - Descritivo do porquê houve uma menor frequência na utilização do automóvel em 2020**



**Fonte: (REDE NOSSA SÃO PAULO, 2020)**

## 2. OBJETIVOS

O trabalho terá como objetivo comparar os dados do número de pessoas que transitaram de ônibus pela empresa SPTrans com o número de casos de Covid 19, durante todo o ano de 2020 no município de São Paulo.

Para isso, será utilizado testes de hipóteses estatísticos para qualificar os dados de casos de Covid e os dados de número de passageiros, determinando possíveis efeitos de causalidade, assim como o *feedback* de resposta.

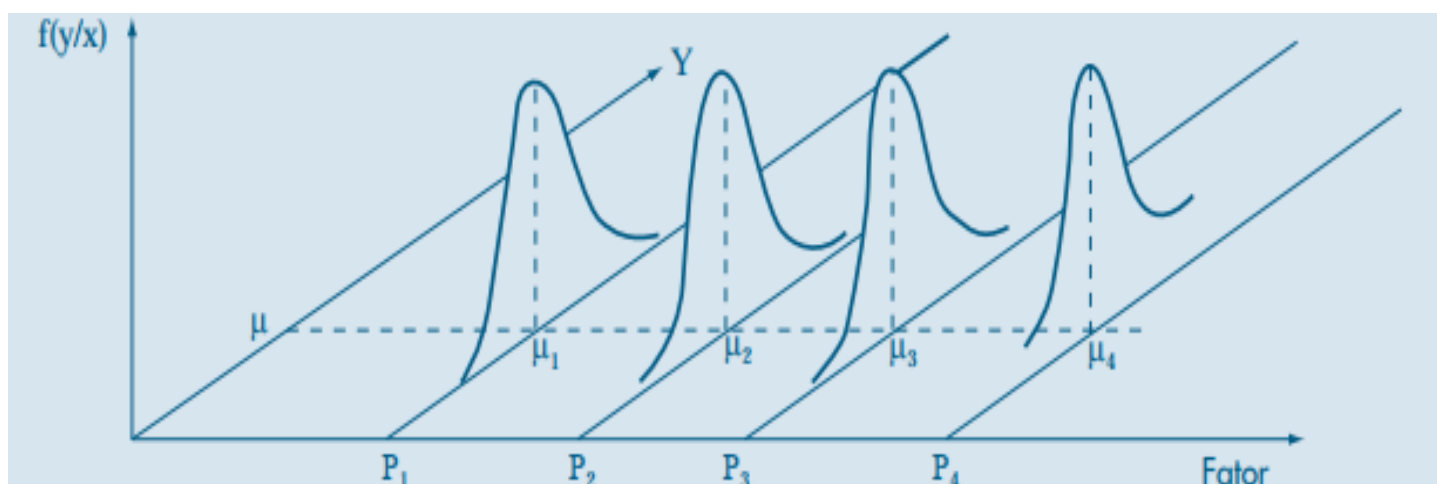
Além disso, será proposto a utilização da regressão linear como método preditivo, a fim de que se consiga, a partir do resultado de feedback, previsões quantitativas decorrentes do número de passageiros que irão ou não aumentar as taxas de contaminação, e desta forma o aumento de casos na região metropolitana de São Paulo, tanto para os dados brutos quanto para os dados filtrados pela sua tendência.

### 3. REVISÃO DE LITERATURA

#### 3.1. Séries Históricas

Séries históricas são conjuntos de dados que são estruturados com os seus respectivos valores de datas. Em estatística matemática existem diversas técnicas para se analisar uma série histórica (VISHWAS e PATEL, 2020).

**Figura 3.1 - Formato de processos estocásticos ao longo do tempo**



*Fonte: (BUSSAB e MORETTIN, 2010)*

O mais interessante, que destoa bastante de analisar recortes temporais é que a fotografia do momento importa. Ou seja, uma mesma grandeza quantitativa com registro e estrutura temporal pode ser dependente de diversas outras variáveis, e assim ter um comportamento extremamente destoante quando analisado em dois momentos distintos.

Desta forma, a análise temporal se torna um método imprescindível quando se quer analisar a evolução no tempo de uma determinada variável. Também será possível determinar correlações entre outras variáveis, tomando como fundamento possíveis efeitos de causalidade entre distintas possibilidades.

E, as análises se tornam cada vez mais complexas por um fato muito simples, importa a ordem de como é estruturado os dados. Entretanto, a ordem não implica uma dependência e correlação proporcional entre a respectiva posição dos dados com o seu respectivo valor.

Com mais detalhes, suponha que um conjunto de dados tenha um registro anual com valores de vendas. Quando se analisa os dados, vê que existe um período



do ano que as vendas aumentam vertiginosamente, enquanto no decorrer dos outros meses ela se mantem com um comportamento padrão.

Agora, sem analisar a causa raiz dessa problemática, basta que se analise e faça a pergunta: Se os dados não fossem sequenciados de maneira cronológica, seria possível teorizar e analisar possíveis causas e consequências dos dados?

E a resposta é não. Sendo impossível condicionar prováveis sazonalidades ou casualidades com eventos externos. Apenas seria possível analisar máximos, mínimos e médias do conjunto de dados.

Aí que entra o conceito de que a ordem de como os dados são estruturados em uma série histórica importa. Entretanto, sua posição não implica em necessariamente uma proporcionalidade.

Quando se coloca esta proporcionalidade, isso significa que, não é devido ao fato de que, por exemplo, o mês de fevereiro é o mês subsequente a janeiro que o valor de vendas deste mês será maior ou menor do que o mês de anterior (janeiro).

Parece um pouco óbvio essa relação, mas quando se analisam sazonalidades ou comportamentos padrões de séries históricas, este tipo de conceito se torna extremamente relevante para a análise.

Novamente com um exemplo, suponha que um determinado produto tenha um comportamento de vendas no verão. Ou seja, nos meses iniciais do ano, o faturamento deste produto é muito mais elevado do que quando comparado com outros meses do ano.

Analisando a situação, mas com um olhar mais focado na estruturação dos dados, é relevante a qualificação dos meses, pois existe um comportamento sazonal sobre a variável, assim como a sua ordem, para que consiga analisar como que esta variável muda ao redor do ano.

Desta forma, volta a problemática inicial, pois a estruturação dos dados históricos importa com relação a qualificação de eventos sazonais, assim como a sua ordem para conseguir inferir evoluções em variáveis. Entretanto, nesta ordem, não necessariamente existe uma correlação de proporcionalidade entre o evolutivo do tempo com a variável a ser estudada.

## 3.2. Séries Estacionárias

Quando se diz que uma série é estacionária, isso não significa que ela tem o mesmo valor sobre todos os registros, mas sim que ela tem características estatísticas semelhantes para quaisquer momentos distintos (VISHWAS e PATEL, 2020).

Ou seja, se retirar duas partes de uma série histórica estacionária, ambas as partes terão características estatísticas semelhantes.

Um exemplo disso, que pode ser utilizado como um teste preliminar para determinar se uma série histórica é estacionária, é segregar uma série em diferentes recortes de períodos e aplicar a média aritmética. Se ela for estacionária, então todas as partes terão a média muito parecida numericamente (mesma ordem de grandeza).

Dentro das séries estacionárias existem quatro grupos distintos:

- **Estacionário estrito ou forte:** é uma distribuição de probabilidade conjunta incondicional que não muda quando deslocado no tempo. Isso significa que a distribuição é a mesma ao longo do tempo.

- **Estacionário de primeira ordem:** significa que nunca muda com o tempo. Outras estatísticas, as propriedades podem mudar, como a variação.

- **Estacionário fraco (segunda ordem):** tem dados de média, variância e covariância constantes ao longo da série temporal. Também é conhecido como covariância / estacionariedade média.

- **Tendência estacionária:** esta série varia em torno das tendências (a média é variada). Essas tendências podem ser lineares ou quadráticas.

- **Diferença estacionária:** este tipo de série tem um ou mais diferenças quando os dados estão em uma série temporal.

Uma das formas mais prática e precisa de se comprovar que uma determinada série histórica é estacionária é por meio de teste de hipóteses, em que via uma hipótese nula e uma hipótese alternativa adjunto com a interpretação do *p-value*, constata-se se a série histórica é estacionária ou não.

### 3.2.1. Metodologia via Raiz Unitária

Teste estatísticos são uma metodologia muito forte para qualificar séries históricas. Isso vem do fato de que a estrutura lógica é a mesma quando se comparado com o método científico, com algumas peculiaridades.

Isso pois, utiliza basicamente duas possibilidades: uma hipótese nula e uma hipótese alternativa. Com tais preceitos com lógica booleana, tenta classificar se uma determinada série é ou não estacionária.

Entretanto, mesmo com a idêntica lógica, existem diversas metodologias para se determinar a existência ou não de estacionariedade. Existindo até testes com uma mesma estrutura, utilizando raízes unitárias, apenas variando a lógica ou combinando diferentes testes em um único.

Mas, antes de tudo, é interessante representar o que significa o método via raiz unitária. O conceito básico se fundamenta que as propriedades de uma série histórica não são constantes. Desta forma, ao analisar esta série histórica, via teste de hipótese, pode ser considerada estacionária ou não, dentro de um certo grau de incerteza predefinido.

Matematicamente, pode descrever qualquer ponto dentro de uma série histórica como:

$$Y_t = \alpha \times Y_{t-1} + E_t$$

$$Y_{t-1} = \alpha \times Y_{t-2} + E_{t-1}$$

Em que:

$Y_t$  : Valor em um tempo t

$Y_{t-1}$ : Valor anterior ao tempo t

$\alpha$ : coeficiente de ajuste

$E_t$ : erro observado no tempo t

Desta maneira, utilizando uma lógica de agrupamento dentro de uma série histórica com n registros, tem-se que:

$$Y_t = \alpha^n \times Y_{t-n} + E_{t-1} \times \alpha^i$$

Com esta equação, pode-se influir duas análises importantes:

- Se  $\alpha$  convergir para um, isso significa que o valor de  $Y_t$  é igual  $Y_{t-n}$  mais a soma de todos os erros dos registros anteriores. Desta forma, isso também significa que a variância aumenta com o decorrer do tempo, sendo está a definição de raiz unitária.
- Em suma, pela definição que uma série histórica não pode ter a variância dependente do tempo, analisa se  $\alpha$  é igual a 1 a fim de se saber a existência de uma raiz unitária.

Um ponto relevante é que, não necessariamente não ter todos os  $\alpha$  monitorados iguais a 1 não implica que a série não será estacionária. Para isso, utiliza-se o princípio de incerteza, em que como monitoramento utiliza-se o *p-value*.

Dentre os diversos testes de hipóteses, tem-se a se destacar:

- **Dickey-Fuller (DF):** baseado em regressão linear, em que autocorrelação é um grande problema para este método;
- **Augmented Dickey-Fuller (ADF):** Este método resolve o problema de autocorrelação do método DF e lida com grandes e complexos modelos;
- **Schmidt-Phillips:** Analisa-se os coeficientes das variáveis determinísticas como hipótese nula e alternativa. Substitutos deste teste são o ro-test e tau-test.
- **Phillips-Perron (PP):** É um melhoramento do teste DF, resolvendo o problema de autocorrelação e heterocedasticidade nos erros;
- **Kwiatkowski-Phillips-Schmidt-Shin (KPSS):** A lógica deste teste é a reversa da aplicada pelo ADF. Ou seja, ele considera como hipótese nula que a série contém tendência, e como hipótese alternativa que existe raiz unitária (série estacionária).

Para este trabalho se utilizará como teste o Augmented Dickey-Fuller (ADF).

### 3.2.2. Teste de Hipótese Augmented Dickey-Fuller

Neste teste, fundamentalmente, analisa-se duas hipóteses:

- **Hipótese nula ( $H_0$ ):** não existe raiz unitária sobre a série histórica analisada, e assim a série não pode ser considerada como estacionária;
- **Hipótese alternativa ( $H_a$ ):** A série histórica é estacionária.

Para facilitar a análise, utiliza-se o *p-value*. É muito comum utilizar um grau de incerteza de 95%. Desta forma, na prática, fazendo o teste de hipótese, se o *p-value* for menor do que 5%, a série pode ser considerada estacionária, caso contrário ela será não estacionária.

### 3.2.3. Método via Diferenciação

Considerando que após realizado os testes de hipótese de estacionariedade, e sendo condicionado que dentro da incerteza de 95% a série não é estacionária, desta forma ainda é possível fazer algumas tratativas a fim de transformá-la em estacionária.

Uma metodologia muito simples é por meio da diferenciação. Ou seja, fundamentalmente subtrai um elemento da série por seu antecessor. Desta forma, as consequências decorrem de uma alteração no comportamento da média sobre a série.

Assim, a média com este processo se torna mais próxima quando analisado em recortes. Por conseguinte, características como tendência e sazonalidade também são controladas. E por conseguinte, uma série histórica não estacionária, com este procedimento muito simplório, pode se tornar uma série estacionária.

### 3.3. Séries com Causalidade

Em vários casos, é muito mais interessante analisar o desenvolvimento de uma série histórica, do que simplesmente analisar a somatória de todo o período. Isso pois com esta abordagem o problema se torna muito mais simples de se explicar, além do fato das hipóteses serem também mais simples de se propor quando analisado apenas as tendências de crescimento e de decrescimento.

Entretanto, com o aumento da interatividade do mundo, é muito difícil que uma variável não tenha alguma dependência com outra variável. Ou seja, é muito complicado que em uma análise de uma série histórica se tenha um processo estocástico puro.

No caso, matematicamente, quando se coloca que uma série tem um efeito de causalidade em outra série histórica, na realidade está se implicando que a variância de uma série tem dependência com a variância de outra série.

#### 3.3.1. Teste de Causalidade (Granger)

Assim como apresentado para teste de estacionariedade, para determinar que duas séries históricas tenham um efeito de causalidade, é recomendado que se faça alguns testes de hipóteses, sendo o mais conhecido sendo o “*direct Granger procedure*”.

Neste teste, basicamente, conclui-se que o futuro não pode alterar o presente ou o passado, assim como o passado não necessariamente afeta o futuro. Um exemplo clássico desta proposição é que o meteorologista, com sua bagagem de conhecimento, não é causador da chuva, mesmo sua previsão sendo feita antes do fato (PRIESTLEY, 1993).

Agora, de uma forma mais matemática, o teste implica que se uma série histórica tem efeito de causalidade sobre outra série histórica estacionária, então é possível prever os valores da série afetada por valores defasados da série histórica defasado.

E uma das formas de se analisar este acontecimento é via relações lineares, como é utilizado pelo método “*direct Granger procedure*”. Entretanto, de maneira simplificada, pode-se definir o teste da seguinte maneira: considerando uma defasagem predefinida, uma série histórica terá um efeito de causalidade se for possível prever outra série com seus valores retroativos (PRIESTLEY, 1993).

Desta forma, este será o teste utilizado neste trabalho para avaliar causalidade, dado um grau de incerteza de 95%. Ou seja, se o *p-value* do teste de Granger for menor do que 5%, haverá causalidade.

### 3.4. Modelos estatísticos

Quando se tem uma situação ou evento em que as consequências dele podem acarretar problemas ou situações muito adversas, é extremamente relevante se saber prever a sua ocorrência.

Para isso, utiliza-se modelos estatísticos, onde estabelece uma hipótese com dependência entre variáveis a fim de que se possa afirmar a possibilidade de eventos adversos e como quantificar seus danos.

#### 3.4.1. Regressão linear

Neste modelo, considera-se que as variáveis estabelecem entre si uma dependência linear. Ou seja, pode combinar variáveis através da combinação entre elas com um fator de ajuste (BUSSAB e MORETTIN, 2010);

Matematicamente, pode-se escrever este modelo como:

$$y_i = \sum_{n=1}^i \alpha_i \times x_i + \beta + \varepsilon$$

Em que:

*y<sub>i</sub>: variável dependente*

*x<sub>i</sub>: variável independente*

*α<sub>i</sub>: coeficiente angular*

*β: coeficiente linear*

*ε: resíduos*

O modelo de regressão linear é um dos mais utilizado em estudo pela sua facilidade de implementação. Outro fato que corrobora para sua utilização é a facilidade de entendimento, pelo ajuste ser uma reta, além do fato do processamento computacional ser menor do que comparado a modelos não lineares.

## **4. METODOLOGIA E DESENVOLVIMENTO**

### **4.1. Dados Utilizados**

#### **4.1.1. Dados dos passageiros de ônibus**

Os dados referentes ao número de passageiros que utilizaram o ônibus como meio de transporte foram retirados do site oficial da companhia SPTrans (SECRETARIA MUNICIPAL E TRÂNSITO, 2020), sendo sua divulgação imposta por lei federal (BRASIL, 2011).

Agora, no que diz respeito a característica dos dados, a SPTrans disponibiliza em seu portal duas opções: utilizar os dados diários ou utilizar o acumulado do mês. No caso dos dados mensais, a grande vantagem é que as linhas e as empresas já foram acumulados e filtrados corretamente, sendo que o trabalho de compilar os dados do ano de 2020 é um facilitador por se utilizar apenas 12 planilhas de Excel.

Entretanto, mesmo o trabalho sendo facilitado, pelo fato de se ter somado a informação dos meses, o número de dados se torna muito restrito, que no caso ter-se-ia apenas 12 valores a serem estudados.

Desta forma, considerando que a análise se pautará em análises não só qualitativas, mas também análises preditivas, se é necessário utilizar o maior número de dados possíveis, para que qualquer tipo de variação na série histórica (outliers) seja vista individualmente, e não de forma conglomerada (somado com outros dados).

O fato da necessidade de se filtrar manualmente os dados pelo bom senso se dá pelo fato do número de passageiros assim como a quantidade de ônibus disponíveis durante o período de 2020, foi totalmente correlacionado com políticas públicas a fim de se diminuir o contágio do vírus. Ou seja, pode-se existir enviesamento nos dados.

Portanto, optou-se por utilizar as informações diárias, para que se tenha um maior controle dos dados, assim como uma maior observância das variâncias de comportamento do modal. No entanto, como já descrito anteriormente, existe o problema laboral sobre os dados.

E isto se dá pelo fato de que o número de ônibus nas planilhas de Excel varia dependendo do dia. Desta forma, uma simples concatenação não resolveria o problema, sendo necessário primeiramente compilar todos os ônibus registrados diariamente e filtrar as duplicatas, a fim de montar uma base de registro.

Novamente, não é apenas o nome da linha que indicará a base de registro. Outro problema que permeará o trabalho é que uma mesma linha de ônibus é feita por mais de uma empresa terceirizada. Desta forma, as variáveis de registro se pautarão em duas variáveis: o nome da linha e o nome da empresa responsável por ela.

Desta forma, compilou-se todos os registros de empresas e linhas em uma planilha e retirou todas as duplicatas, tendo assim o catálogo total de ônibus que funcionaram no período de 2020.

Agora, com a base neste registro, compilou-se todos os dados diários em uma única tabela, utilizando como chave o código da linha e a empresa responsável.

Desta forma, foi compilado todos os dados de 2020 em uma única tabela, tendo o registro total de passageiros em todo o ano de 2020 estratificado por linha e por empresa responsável pelo serviço.

#### **4.1.2. Dados de contágios pela Covid 19**

Para estes dados, foi utilizado as informações do SEADE (Sistema Estadual de Análise de Dados) contidas no Github em formato .csv (SEADE, 2020), em que se disponibiliza os dados de Covid para todo o estado de São Paulo.

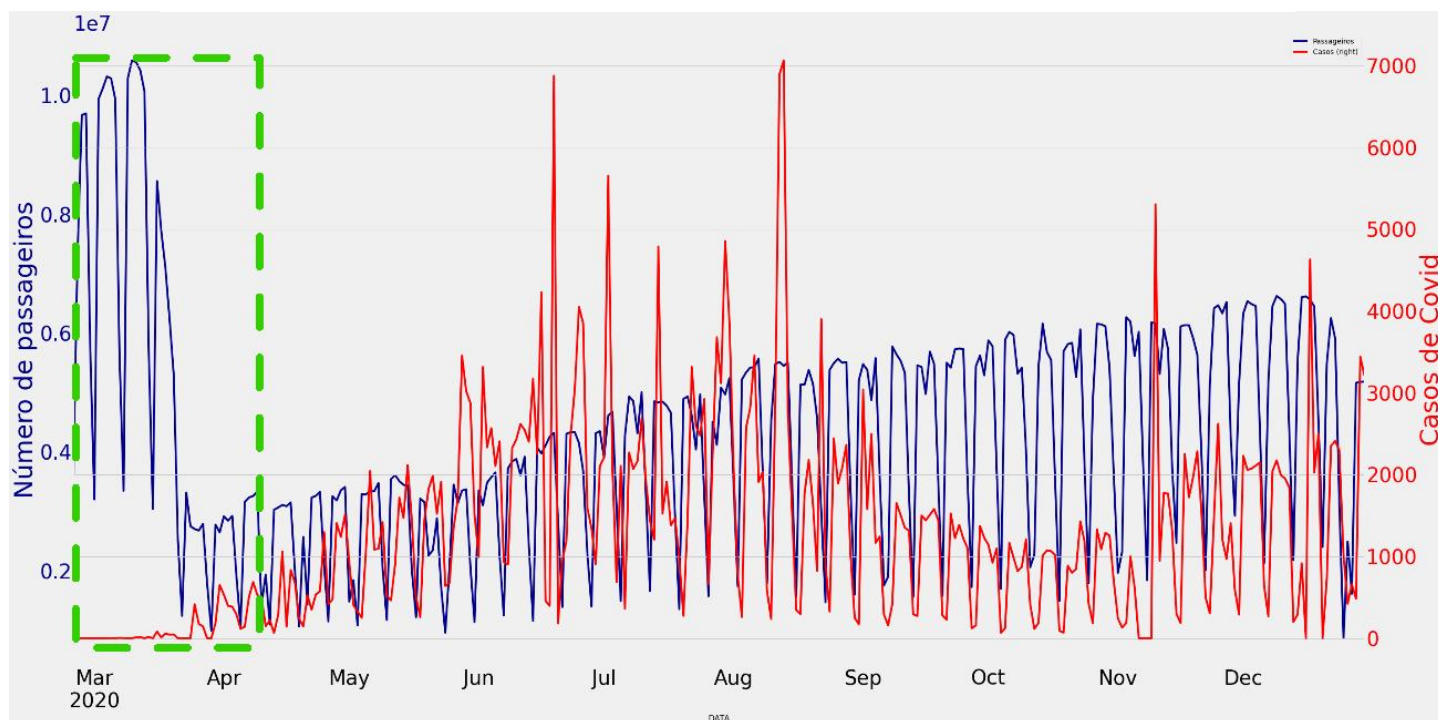
Para este conjunto de informações, foi mais fácil segregar os dados pois já estavam devidamente estruturados, bastando apenas filtrar a região de São Paulo e selecionar as colunas referentes as datas e ao número de contaminados por dia.

### **4.2. Manipulação e filtros**

Um elemento que apareceu durante a análise dos dados foi que do momento em que registrou o primeiro caso até as medidas de contenção com a redução de frota e restrição de passageiros, houve um período de adaptação do sistema. Desta forma, não se considerou os dados anteriores a abril de 2020, a fim de se excluir este efeito de feedback.



**Gráfico 4.1 - Comparativo entre a evolução de casos de covid com o número de passageiros sobre todo o período de 2020**

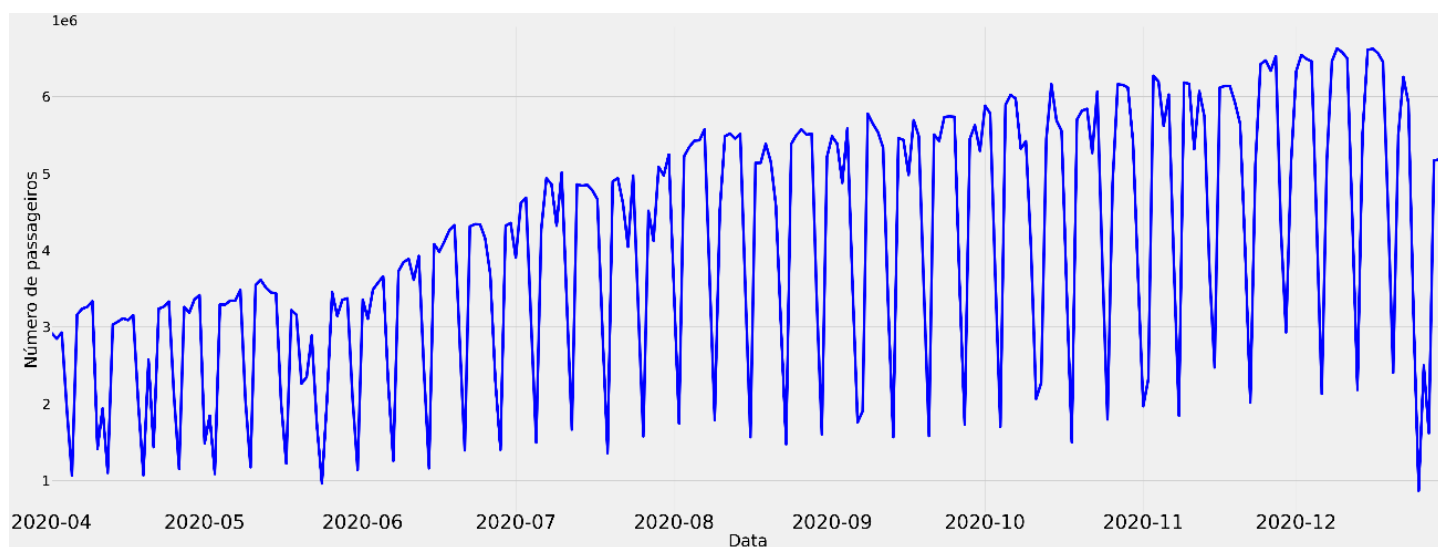


*Fonte: Elaborado pelo autor*

#### 4.2.1. Análise dos dados de passageiros

Ao se compilar os dados de passageiros transportados pela empresa SPTtrans durante o período de abril de 2020 até dezembro de 2020, obteve o comportamento representado pelo Gráfico 4.2.

**Gráfico 4.2 - Série do número de passageiros com dados brutos**



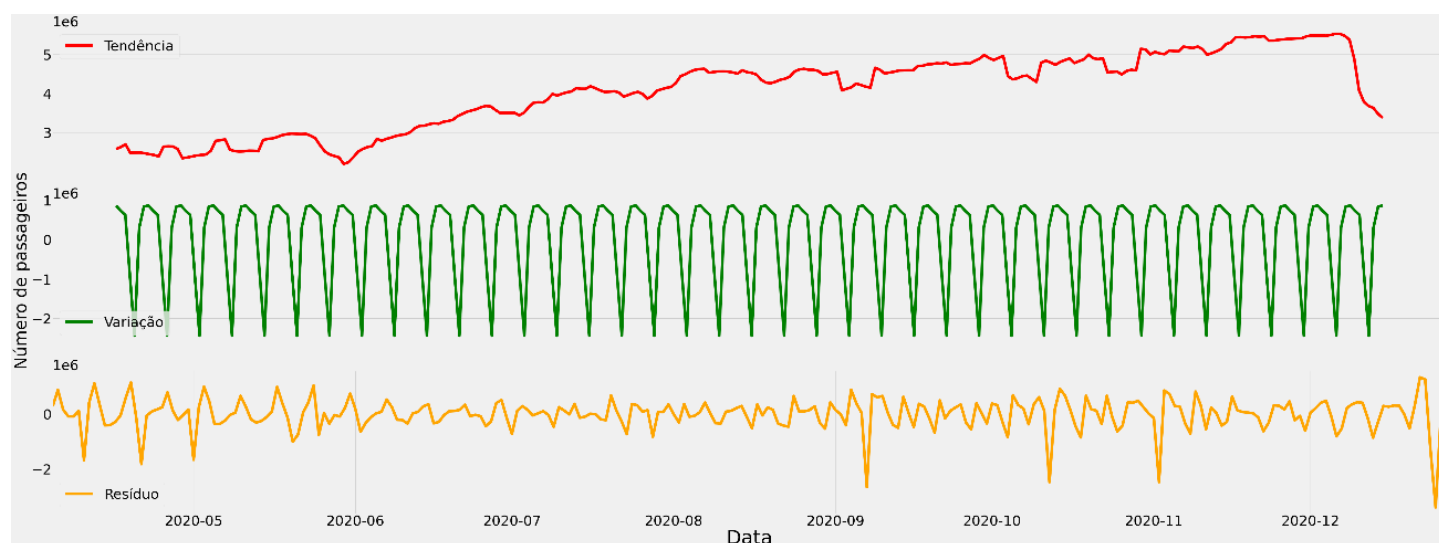
*Fonte: Elaborado pelo autor*

Um fator importante analisado, é que existe um comportamento de certa forma sazonal. Na realidade, este comportamento é explicado pela função social que o ônibus tem na região metropolitana de São Paulo. Isto pois, a principal necessidade dos passageiros utilizando o ônibus é de se deslocar para o trabalho. Desta maneira, considerando os turnos de segunda a sexta, nos finais de semana o número de passageiros é menor.

Entretanto, para a análise, fica complicado de analisar este tipo de variação, por mais que seja relevante, visto que dentro da análise, isso pode mascarar a tendência de crescimento e decréscimo dos passageiros.

Desta forma, para facilitar a análise, foi utilizado a função de série histórica da biblioteca do *python* a fim de se filtrar os valores correspondentes a tendência, a sazonalidade e a ruído, sendo o resultado apresentado pelo Gráfico 4.3.

**Gráfico 4.3 - Decomposição da série do número de passageiros**



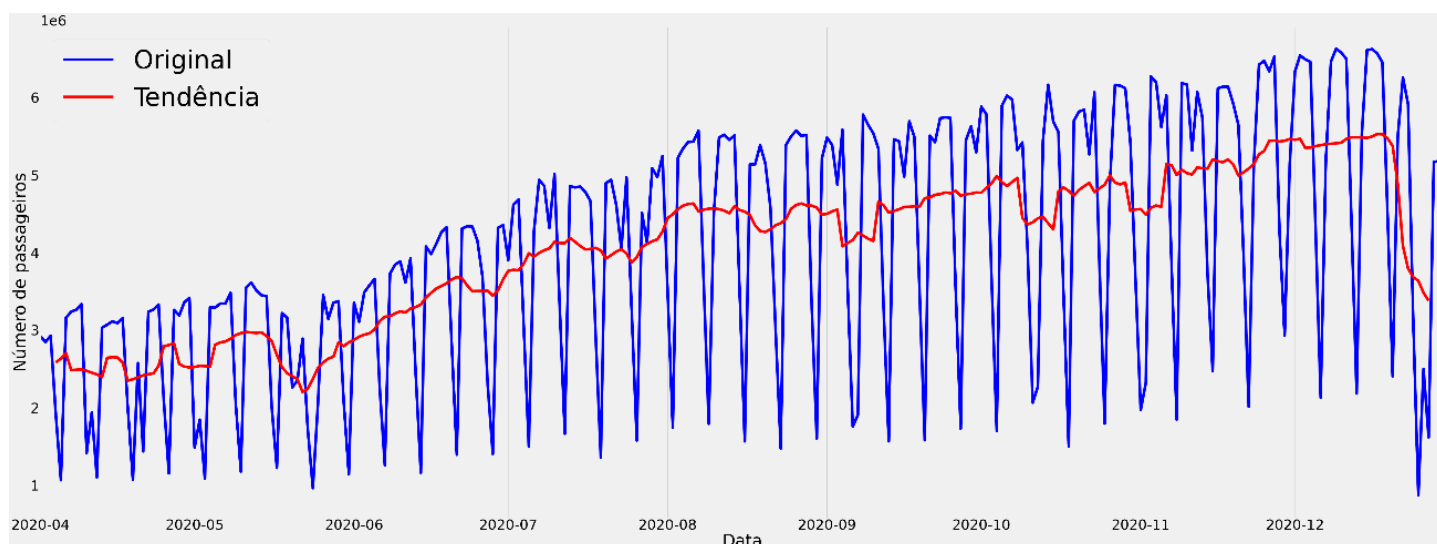
**Fonte: Elaborado pelo autor**

Ou seja, o objetivo deste trabalho se resumirá a analisar não apenas o valor absoluto de passageiros, mas também a tendência de passageiros. Vale ressaltar que este processo resultará numa maior diferença entre o valor real do valor esperado, visto que não será trabalhado valores reais, mas sim valores filtrados.

Entretanto, como os valores referentes aos passageiros são dependentes de diversas variáveis, como eventos, em que, por exemplo aumenta-se o número de passageiros no final de semana, o filtro de tendência, de uma certa maneira, deixará possivelmente os dados mais coerentes com a realidade, diminuindo a influência dos outliers sobre a análise.

Desta forma, os dados finais após as manipulações estão demonstrados no Gráfico 4.4, em que se compara o valor real com os valores de tendência filtrados.

**Gráfico 4.4 - Comparação entre a série com dados brutos e com filtro de tendência do número de passageiros**

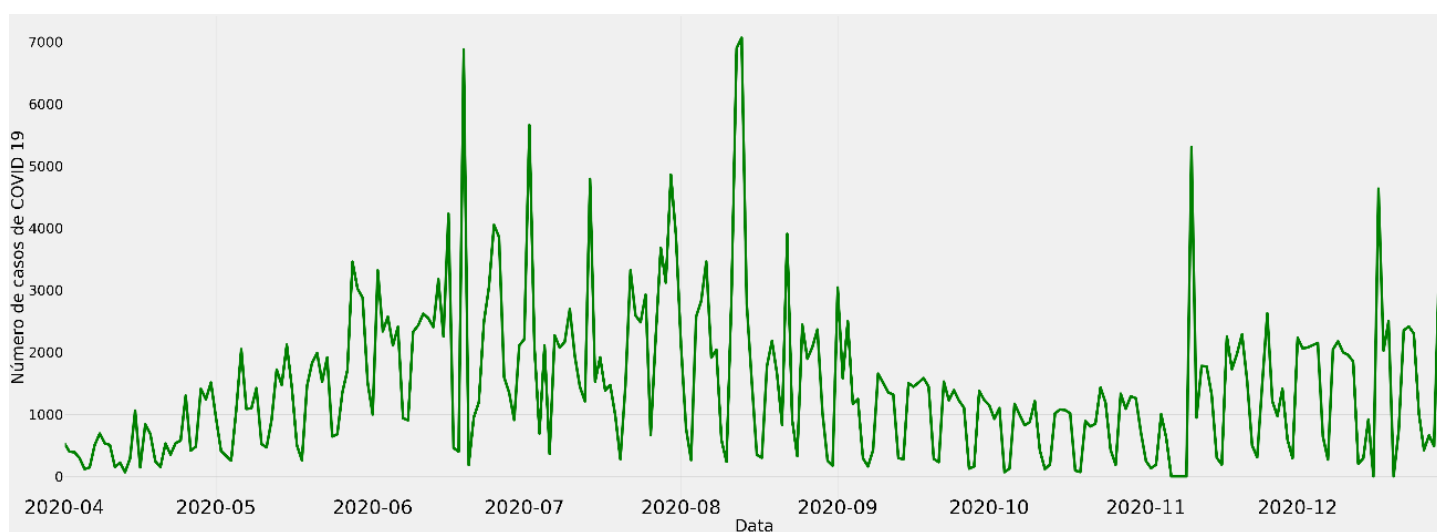


*Fonte: Elaborado pelo autor*

#### 4.2.2. Análise dos dados de Covid 19

Para a análise dos dados de Covid, a tratativa foi muito semelhante ao utilizada pelo de passageiros. E, muito interessante que o comportamento foi também semelhante ao observado pelos dados de ônibus com relação a variação semanal como demonstrado pelo Gráfico 4.5.

**Gráfico 4.5 - Série com dados brutos dos casos de Covid**



*Fonte: Elaborado pelo autor*

Entretanto, mesmo com o idêntico comportamento, a origem é extremamente distinta. Isso pois, todos os registros de covid são feitos manualmente. E, o que aconteceu durante todo o ano foi o acúmulo dos dados do final de semana transportados para a segunda-feira.

Desta forma, diferente do caso do transporte, em que os finais de semana são menores por conta da dinâmica metropolitana, existe uma variação sazonalidade para os casos de Covid por uma questão mais de processo de registro do que por um fator orgânico.

Este fato se evidencia pelos picos de casos logo após um período de baixa. Isso demonstra que os casos dos finais de semana estão todos contidos sobre as segundas-feiras. Tal hipótese não causaria problema se fosse analisado apenas os anos de 2020 como um todo. Sendo certo que os dias seriam todos somados.

Contudo, esta característica afetarà as análises pois não será observado apenas o valor absoluto, mas também se fará um comparativo entre a evolução e uma correlação entre as duas séries históricas.

Uma possibilidade seria estudar quando seria o *lead time* do registro entre os finais de semana e retirar uma parcela dos dados da segunda feira, e redimensioná-los sobre os sábados e os domingos proporcionalmente.

Entretanto, como existem diversas fontes de hospitais que trabalharam com estes registros, ficaria inviável fazer um estudo sobre este comportamento a fim de se concertar os atrasos de registros.

Dessa forma, será feito o mesmo processo de filtros pela biblioteca do *python*, visando se segregar os valores referentes a tendência, a sazonalidade e a resíduos, sendo o resultado apresentado pelo Gráfico 4.6.

**Gráfico 4.6 - Decomposição da série de casos de Covid**

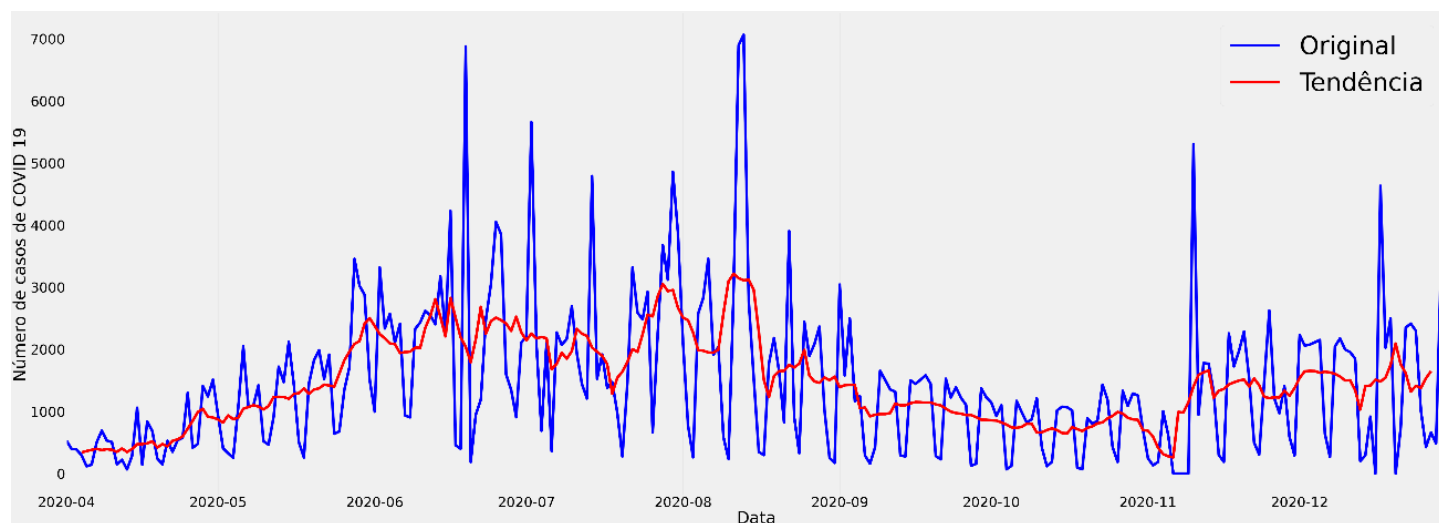


**Fonte: Elaborado pelo autor**

Um ponto importante é que o gráfico de resíduos de ambos (valor absoluto menos o valor de tendência e o valor de sazonalidade) ficaram em torno do valor zero. O que é um bom resultado, sendo os picos possíveis valores de outliers, tanto de concentração de registros para os casos de Covid, como eventos culturais para os dados de passageiros.

Desta forma, o resultado dos filtros e a comparação entre valor real contra o valor de tendência estão apresentados pelo Gráfico 4.7.

**Gráfico 4.7 - Comparação entre a série com dados brutos e com filtro de tendência dos casos de Covid**



**Fonte: Elaborado pelo autor**

### 4.3. Testes de Hipóteses

Dadas as duas séries históricas, fez-se dois testes de hipóteses em ambas: teste de ADF com a correção via diferenciação, para depois fazer o teste de Granger entre as duas, a fim de se estabelecer um grau de causalidade entre as duas séries.

Vale ressaltar que, como ambos os testes de hipóteses trabalham com a variabilidade das séries, foi utilizado os dados brutos de ambas. Ou seja, não se trabalhou com os dados de tendência filtrados.

Portanto, os testes estão mais suscetíveis a outliers. Entretanto, a utilização de filtros inviabilizaria a análise via teste de hipótese, pelo fato das variações entre dias com a aplicação dos filtros, tornariam as variações mais *flat*, do que comparado com os valores originais.

#### 4.3.1. Teste ADF com Diferenciação

##### 4.3.1.1. Teste dos dados de Covid 19

Para a análise, separou-se os dados, aplicando o teste ADF, sendo os resultados iniciais apresentados pela Tabela 4.1.

**Tabela 4.1 – Resultado do teste ADF para os dados brutos dos casos de Covid**

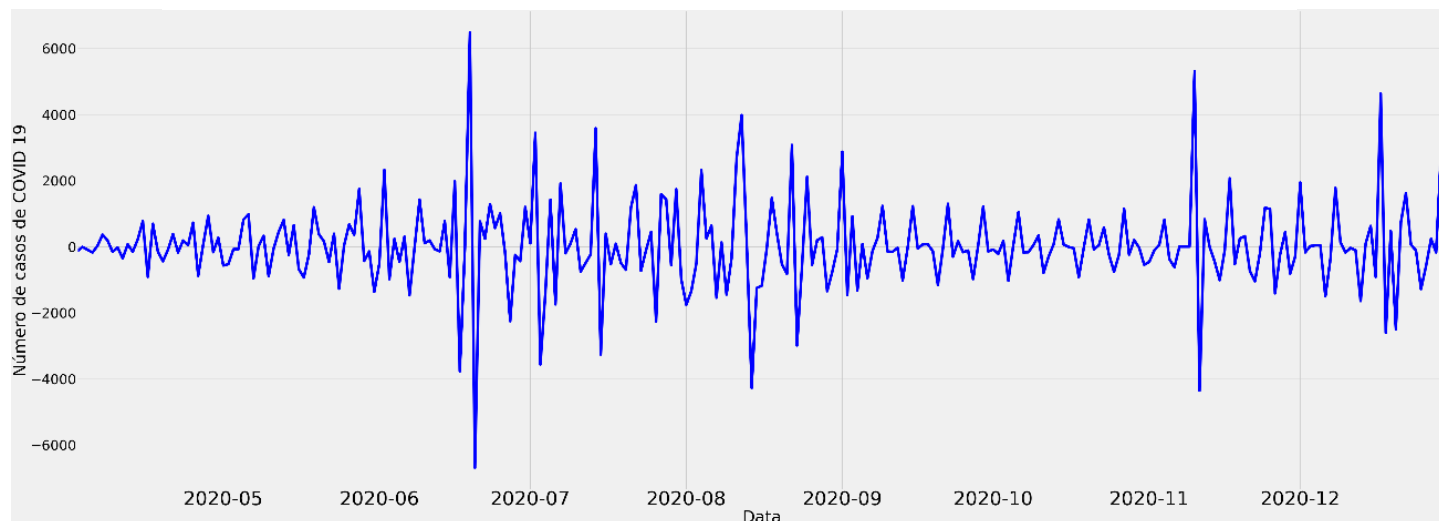
<b>ADF Statistic:</b>	-1,919722
<b>p-value:</b>	0,322841
<b>Critical Values:</b>	
<b>1%</b>	-3,456
<b>5%</b>	-2,873
<b>10%</b>	-2,573

*Fonte: Elaborado pelo autor*

Desta forma, considerando um nível de incerteza de 95%, não se pode considerar que os dados de Covid têm características de estacionariedade. Para isso, será utilizado o processo de diferenciação. O resultado está apresentado pelo Gráfico 4.8.

Pelo comportamento decorrente do processo de diferenciação, já dá para se falar que a série já pode ter um comportamento estacionário com apenas uma diferenciação. Mesmo assim, foi-se feito o teste ADF novamente nesta série para garantir numericamente este comportamento, como apresentado pela Tabela 4.2

**Gráfico 4.8 – Série do número de casos de Covid decorrente de uma diferenciação**



*Fonte: Elaborado pelo autor*

**Tabela 4.2 - Resultado do teste ADF para os dados de casos de Covid após o processo de diferenciação**

<b>ADF Statistic</b>	<b>-9,297289</b>
<b>p-value:</b>	<b>0,000000</b>
<b>Critical Values:</b>	
<b>1%</b>	<b>-3,456</b>
<b>5%</b>	<b>-2,873</b>
<b>10%</b>	<b>-2,573</b>

*Fonte: Elaborado pelo autor*

Desta forma, como o *p-value* está extremamente baixo, pode-se afirmar que a série de casos de Covid com um processo de diferenciação é estacionária.

#### 4.3.1.2. Teste dos dados de passageiros

Assim como foi feito com a série de Covid, os dados de passageiros foram analisados via teste ADF para se saber, se a série pode ser considerado uma série histórica estacionária, sendo os resultados do teste apresentados pela Tabela 4.3.

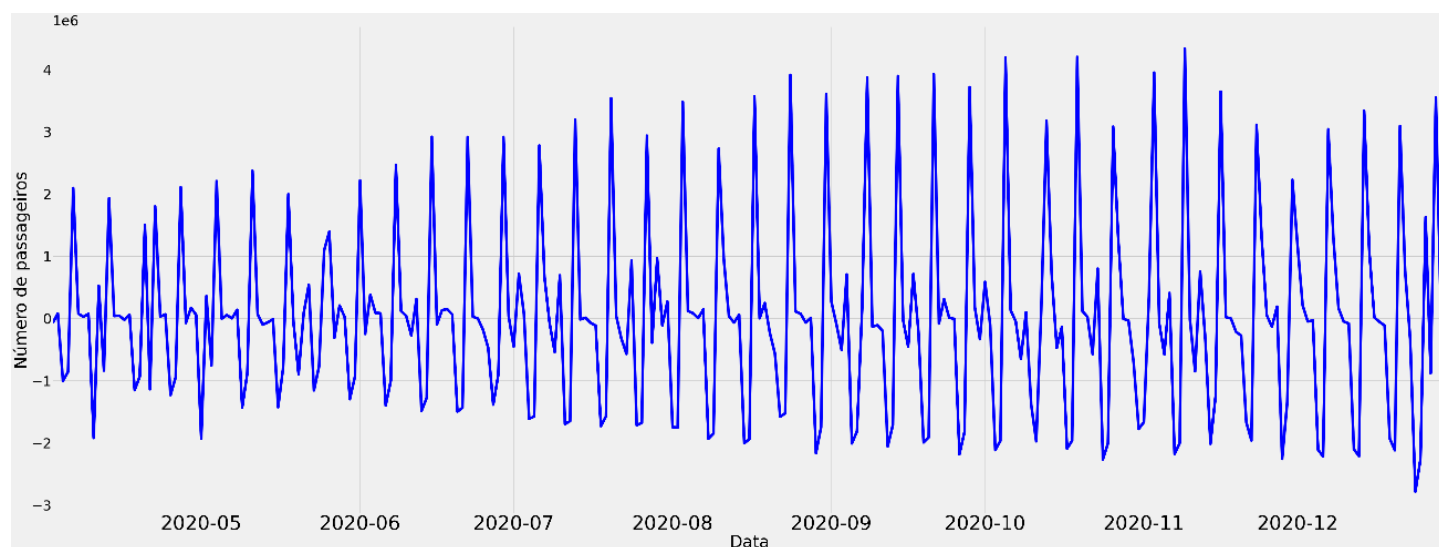
**Tabela 4.3 - Resultado do teste ADF para os dados brutos de passageiros**

<b>ADF Statistic</b>	-1,733127
<b>p-value</b>	0,414130
<b>Critical Values</b>	
<b>1%</b>	-3,456
<b>5%</b>	-2,873
<b>10%</b>	-2,573

*Fonte: Elaborado pelo autor*

Desta forma, assim como o que aconteceu com os dados de Covid, a série de passageiros também não podem ser considerados como uma série estacionária. Assim, foi feito o processo de diferenciação, sendo o resultado apresentado pelo Gráfico 4.9.

**Gráfico 4.9 – Série do número de casos de Covid decorrente de uma diferenciação**



*Fonte: Elaborado pelo autor*



Assim como o de Covid, o formato do gráfico com apenas um processo de diferenciação já deixou os dados com um comportamento de estacionariedade. E da mesma forma, se refez o teste ADF para garantir numericamente o comportamento descrito.

**Tabela 4.4- Resultado do teste ADF para os dados de passageiros após o processo de diferenciação**

<b>ADF Statistic</b>	-5,643939
<b>p-value</b>	0,000001
<b>Critical Values</b>	
<b>1%</b>	-3,456
<b>5%</b>	-2,873
<b>10%</b>	-2,573

**Fonte: Elaborado pelo autor**

Desta forma, as duas séries históricas com um processo de diferenciação tornaram-se estacionárias. Assim, pode-se analisar o processo de causalidade para diferentes dias de intervalo de resposta via teste de Granger.

### 4.3.2. Teste de Granger

Considerando que os dados de passageiros e os de casos de Covid 19 já foram tratados com relação a estacionariedade, desta maneira já se pode aplicar o teste de causalidade.

Um fator importante é que, qual será a abordagem de espaço de tempo que será avaliado na comparação. Ou seja, qual será o intervalo de tempo em que possivelmente o número de passageiros afetará os dados de Covid (tempo de *feedback*).

Para esta problemática, existem diversas formas matemáticas visando definir este intervalo, entretanto não existe uma regra (PRIESTLEY, 1993). Desta forma, considerando a literatura e os conhecimentos sobre a doença, sabe-se que a maioria dos pacientes começam a relatar sintomas a partir do quinto dia após a exposição (OMS, 2019).

Com este conhecimento, aplicará um intervalo de sete dias. Ou seja, o teste de hipótese fará correlações de variabilidade entre os dados de Covid com o relativo dado de número passageiros defasados de um a até sete dias.

Um elemento relevante é que em nenhum momento se fez a análise de covariância entre as duas séries históricas. Desta forma, para a análise do teste, será utilizado os valores de  $\chi^2$ .

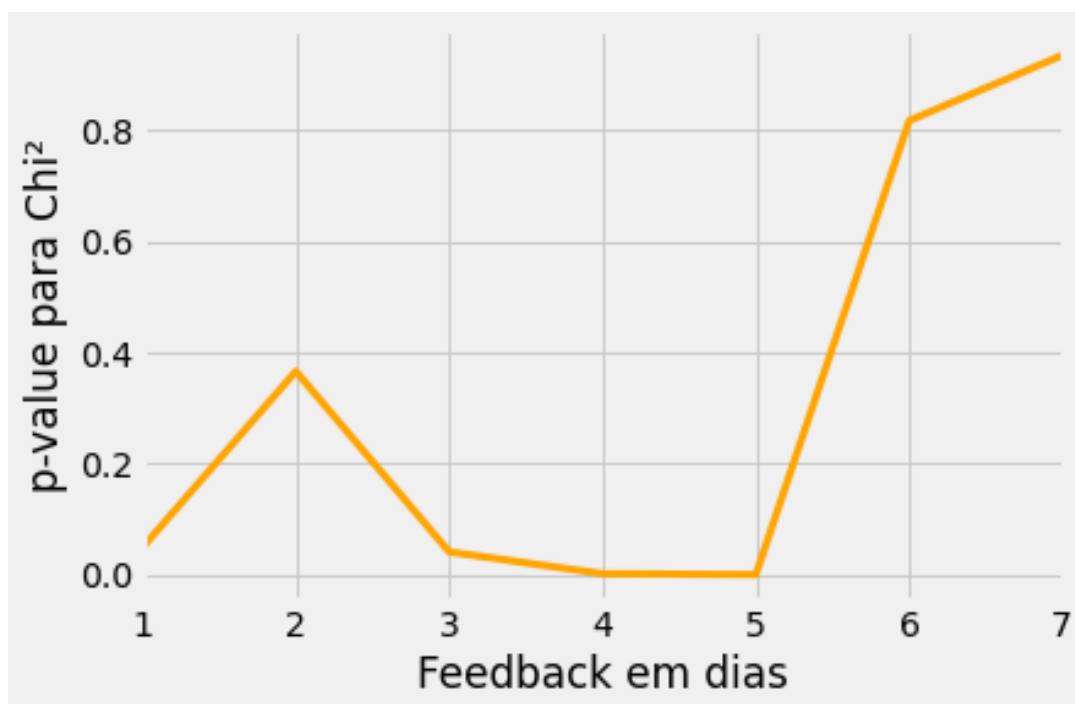
Portanto os resultados do teste são apresentados pela Tabela 4.5 e pelo Gráfico 4.10.

**Tabela 4.5 – Resultados do teste de Granger para diferentes intervalos**

Granger Causality				p
number of lags (no zero) 1	ssr based F test	F	38,309	0,051
	ssr based chi2 test	chi2	38,736	0,049
	likelihood ratio test	chi2	38,463	0,050
	parameter F test	F	38,309	0,051
number of lags (no zero) 2	ssr based F test	F	0,986	0,374
	ssr based chi2 test	chi2	20,096	0,366
	likelihood ratio test	chi2	20,022	0,368
	parameter F test	F	0,986	0,374
number of lags (no zero) 3	ssr based F test	F	26,818	0,047
	ssr based chi2 test	chi2	82,594	0,041
	likelihood ratio test	chi2	81,356	0,043
	parameter F test	F	26,818	0,047
number of lags (no zero) 4	ssr based F test	F	42,217	0,003
	ssr based chi2 test	chi2	174,715	0,002
	likelihood ratio test	chi2	169,275	0,002
	parameter F test	F	42,217	0,003
number of lags (no zero) 5	ssr based F test	F	87,134	0,000
	ssr based chi2 test	chi2	454,316	0,000
	likelihood ratio test	chi2	419,672	0,000
	parameter F test	F	87,134	0,000
number of lags (no zero) 6	ssr based F test	F	0,465	0,834
	ssr based chi2 test	chi2	29,335	0,817
	likelihood ratio test	chi2	29,175	0,819
	parameter F test	F	0,465	0,834
number of lags (no zero) 7	ssr based F test	F	0,322	0,944
	ssr based chi2 test	chi2	23,890	0,935
	likelihood ratio test	chi2	23,784	0,936
	parameter F test	F	0,322	0,944

**Fonte: Elaborado pelo autor**

**Gráfico 4.10 - Resultado do feedback considerando o valor de  $p$ -value**



**Fonte: Elaborado pelo autor**

Com o resultado, tem-se que para se existir causalidade entre o número de passageiros com relação ao número de casos de Covid no município de São Paulo, considerando um  $p$ -value de 5%, deve considerar um feedback de 5 dias. E mais do que isso, o número da faixa está em torno ao apresentado pela pesquisadora em que na maioria dos pacientes relatam que os sintomas aparecem nesta faixa de tempo.

Pode parecer pouco, mas com este conhecimento, pode ser analisado a ocorrência de novos casos de Covid numa folga de 5 dias de predição, apenas com os dados de passageiros. Mesmo assim, um ponto muito relevante sobre as próximas análises é que existem inconsistências nos dados de Covid, decorrente do processo de subnotificação.

Isto pois muitas pesquisas confrontaram os valores de óbitos por questões respiratórias com os dados de Covid. Ou seja, a conta inversa, considerando que a taxa de mortalidade é constante entre diferentes povos giraria em torno de 1%, não sendo muito dependente das condições da estrutura urbana e das oportunidades de saúde para a população.

E na maioria dos estudos, é reportado que o Brasil teve um processo de subnotificação tanto de casos de Covid quanto de óbitos devido a complicações da doença. Ou seja, mesmo que esteja comprovado uma certa causalidade e que exista

uma possibilidade de se analisar predições via modelos estatísticos, o resultado sempre será abaixo do observado pela realidade.

Isto acontece pelo fato de que todas as análises as quais foram feitas utilizando como parâmetro os dados oficiais, em que como relatado, tiveram problemas com relação a subnotificação e atraso no registro. Mesmo com isso, as possibilidades de *inputs* com a análise destes dados ajudarão a entender o processo pandêmico, no caso do município de São Paulo com relação a seu principal modal urbano, podendo auxiliar em outros processos sanitários.

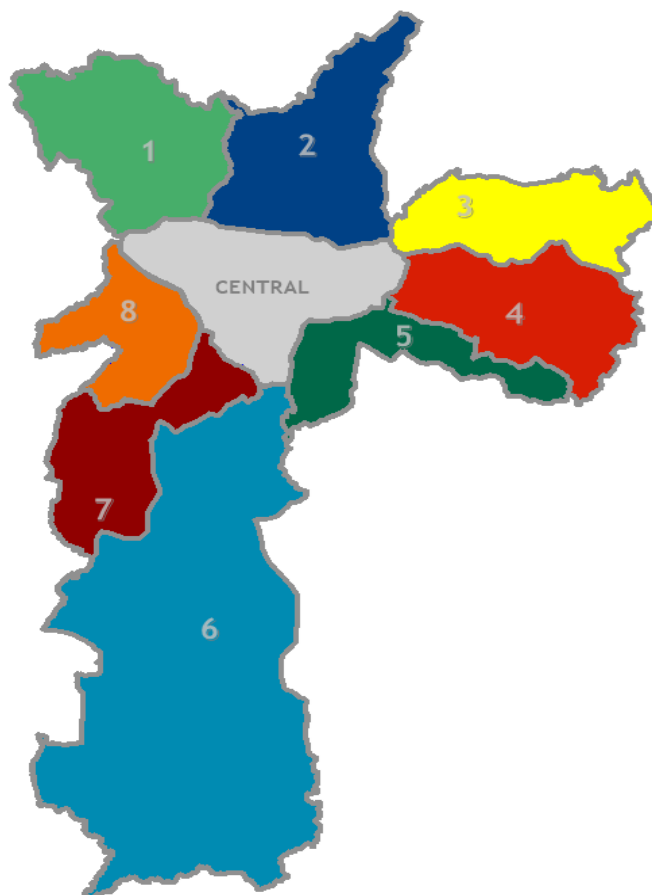
## 4.4. Regressão Linear

### 4.4.1. Consideração e estudo dos dados de transporte

Dentre os diversos modelos existentes, foi escolhido utilizar o modelo de regressão linear. A escolha por este modelo foi simples, como a equação trabalha com princípios básicos da matemática, como multiplicação e soma, este fator faz com que possa aplicar os resultados em planilhas eletrônicas. Isso entra um pouco no objetivo do trabalho, de se conseguir aplicar os conhecimentos sobre a gestão pública.

Colocado isso, outro fator importante é como é distribuído o comportamento da população sobre a utilização do ônibus no município de São Paulo, como explicitado na seção 1.2.1. Desta forma, é interessante fazer a análise estratificado em regiões. A própria SPTrans faz uma distinção geográfica sobre as linhas como indicado pela Figura 4.1.

**Figura 4.1 – Representação da divisão feita pela SPTrans da cidade de São Paulo em distritos**



**Fonte: (SPTRANS, 2020)**

Entretanto, mesmo havendo uma divisão de nove setores, nos dados cedidos pela empresa, existem o registro apenas da região 1 até a 8. Ou seja, para este trabalho se suprimirá as características da região nove (central).

É um fator importante é que mesmo com as oito regiões, a soma sobre as regiões quando analisados sobre o número de passageiros para a data estudada (a partir de abril), é bastante diferente. Este comportamento decorre pelo fato da desigualdade na utilização do transporte.

Ou seja, as regiões analisadas têm características bem estabelecidas quando são comparadas entre si pelo poder aquisitivo das pessoas. Este tema foi trabalho na seção 1.2.1.

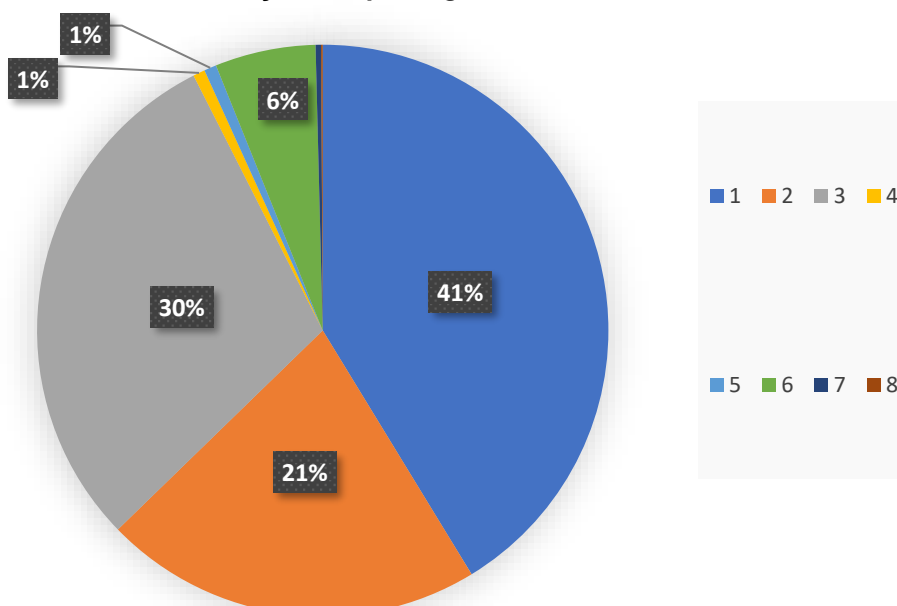
Estas características são materializadas pela Tabela 4.6 e pelo Gráfico 4.11.

**Tabela 4.6 - Distribuição dos passageiros no ano de 2020**

ÁREA	NÚMERO DE PASSAGEIROS	PASSAGEIROS EM %	PASSAGEIROS EM % ACUM.	MÉDIA DIÁRIA DE PASSAGEIROS
1	548.446.234	41,28%	41,28%	1.769.181
2	285.059.337	21,45%	62,73%	919.546
3	396.036.476	29,81%	92,53%	1.277.537
4	8.950.610	0,67%	93,21%	28.873
5	9.262.718	0,70%	93,90%	29.880
6	5.691.200	5,70%	99,60%	244.165
7	4.085.021	0,31%	99,91%	13.177
8	1.218.594	0,09%	100,00%	3.931
<b>Total Geral</b>	<b>1.328.750.190</b>	<b>100,00%</b>		<b>535.786</b>

Fonte: Elaborado pelo autor

**Gráfico 4.11 - Distribuição dos passageiros no ano de 2020**

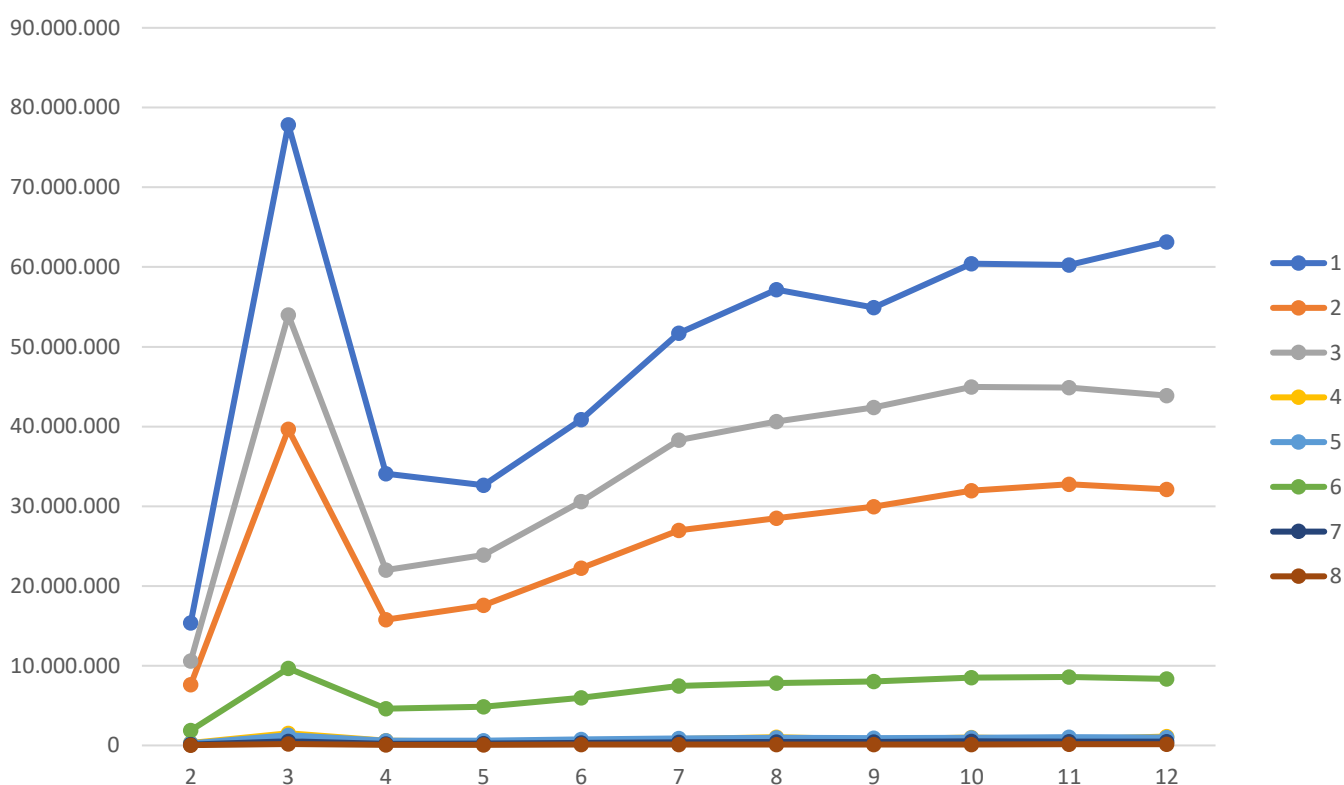


Fonte: Elaborado pelo autor

Agora, um fator que poderá alterar o resultado do estudo é se as regiões mantiveram o comportamento ao longo de 2020, considerando o cenário pandêmico.

Para este estudo, de forma preliminar, tem-se os Gráfico 4.12, demonstrando que, mesmo com a evolução do número de passageiros e a flexibilização das medidas sanitárias, o número de passageiros são proporcionais entre si. Ou seja, mesmo com o crescimento do número de passageiros, o comportamento dentre os distritos se mantiveram semelhantes ao longo do ano de 2020.

**Gráfico 4.12 - Distribuição dos passageiros ao longo do ano de 2020**



**Fonte: Elaborado pelo autor**

Portanto, observando a estratificação entre distritos assim como considerando que o comportamento ao longo de 2020 foi proporcional entre os distritos, para este trabalho será feito a seguinte tratativa: se considerará como variáveis independentes o número de passageiros por distrito e como variável dependente os casos de Covid depois de 5 dias, respeitando o valor de feedback calculado na seção 4.3.2.



Assim, no final do trabalho teremos a seguinte equação:

$$C_{5 \text{ dias}} = \sum_{n=1}^8 \alpha_i \times P_i$$

Em que:

$C_{5 \text{ dias}}$ : casos de Covid em 5 dias  
 $P_i$ : número de passageiros por região  
 $\alpha_i$ : Coeficiente angular da regressão linear

Com isso, será possível ao final do trabalho estipular quantos casos será computado com os dados de passageiros por região do dia analisado.

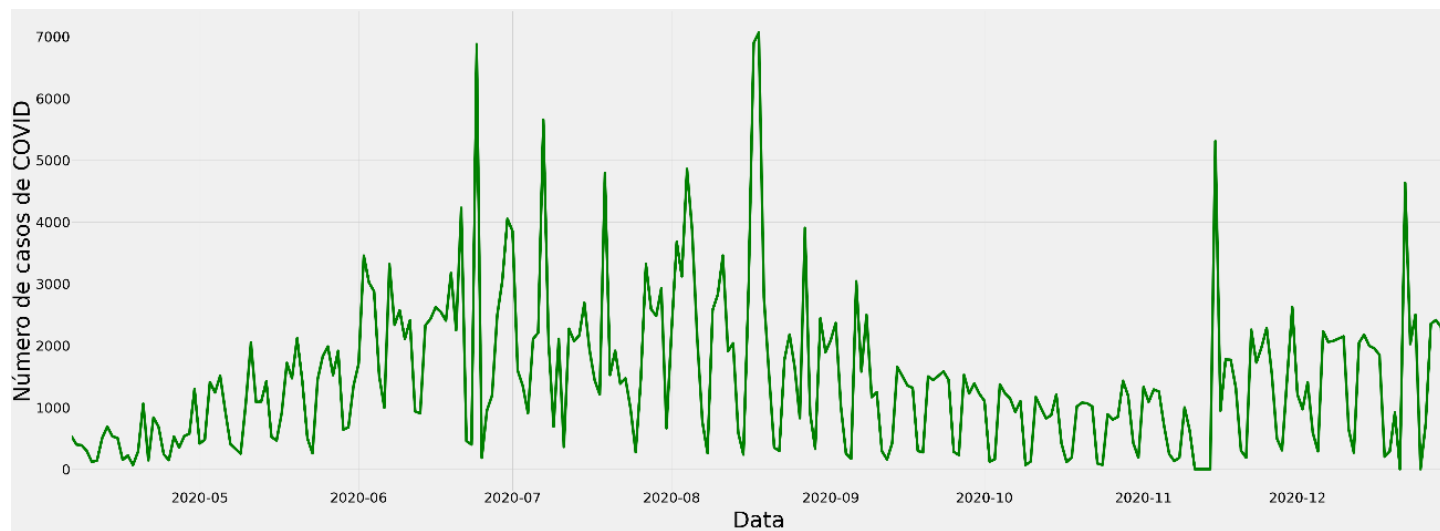
#### 4.4.2. Estudo sobre as variações semanais

Como descrito na seção 4.2, os dados tanto de passageiros como de Covid apresentam sazonalidades semanais. Para o caso de Covid, este comportamento decorre do acúmulo de registros nas segundas feira, e no caso do ônibus, por conta deste modal ter maiores utilizações nos dias úteis.

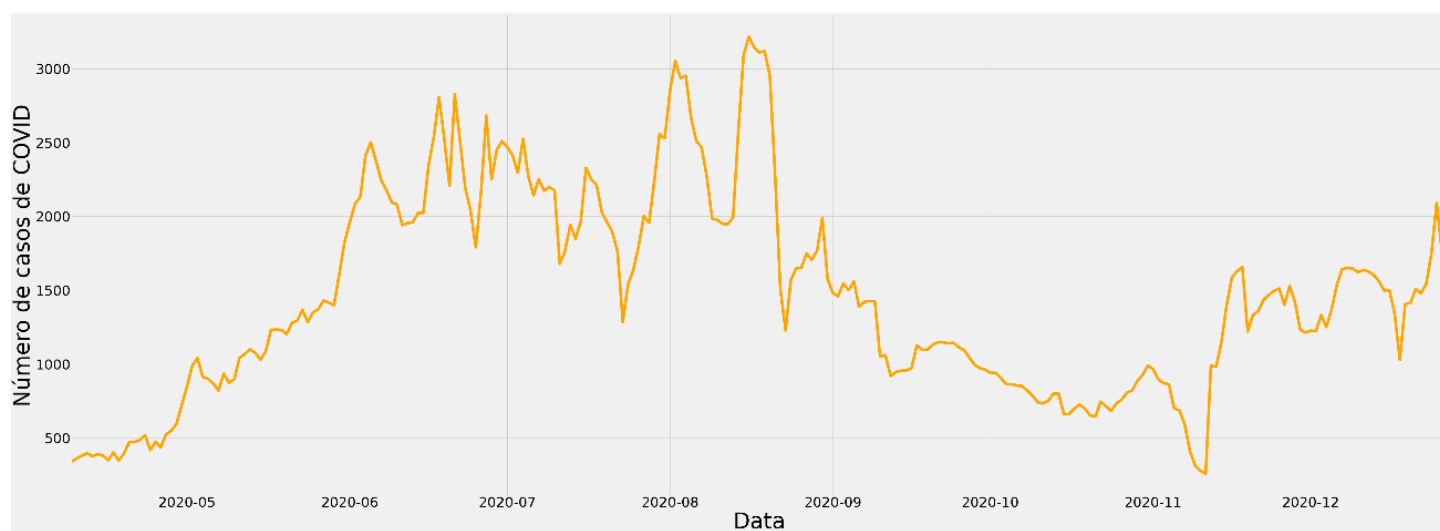
Dessa forma, assim como foi feito anteriormente, se plicará um filtro para diminuir a influência de tais características. Com isso, resultará em dois conjuntos de dados: Número de passageiros por distrito e casos de Covid com filtro de tendência e sem filtro de tendência.

Ambos serão aplicados regressão linear a fim de se saber qual conjunto apresentará melhores resultados, e será utilizado como métrica de comparação o *p-value* segregado por área e o valor que Chi<sup>2</sup>.

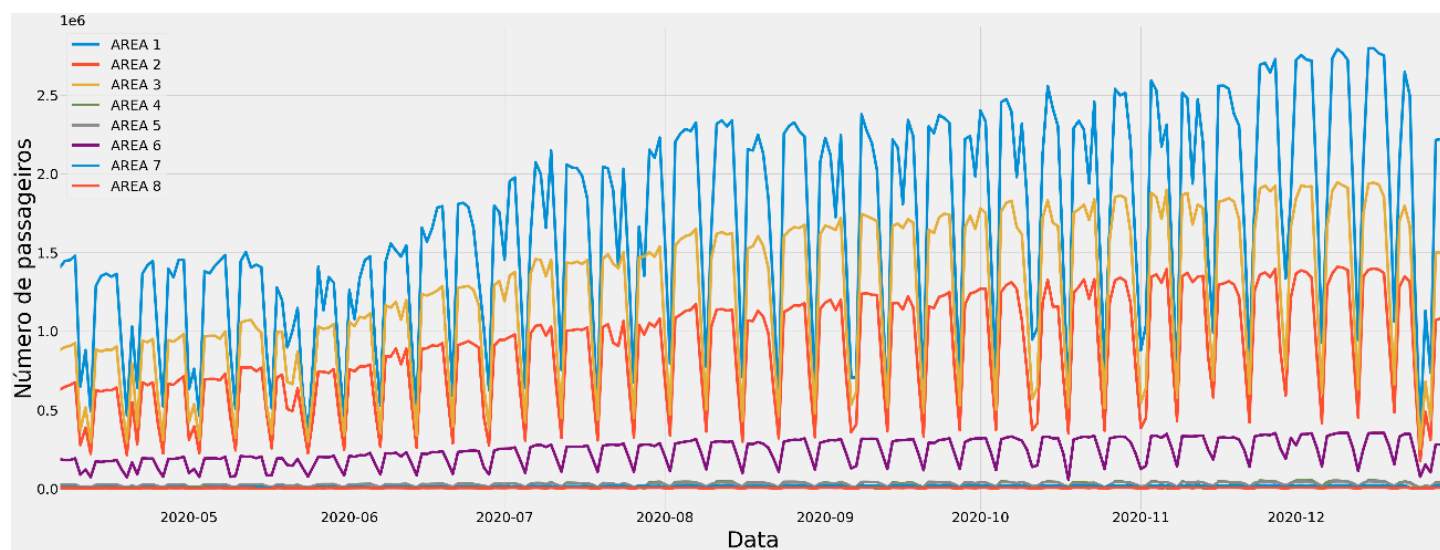
Os dados com filtro assim como os valores absolutos estão apresentados pelo Gráfico 4.13, Gráfico 4.14 , Gráfico 4.15 e pelo Gráfico 4.16.

**Gráfico 4.14 – Série de dados brutos de casos de Covid**

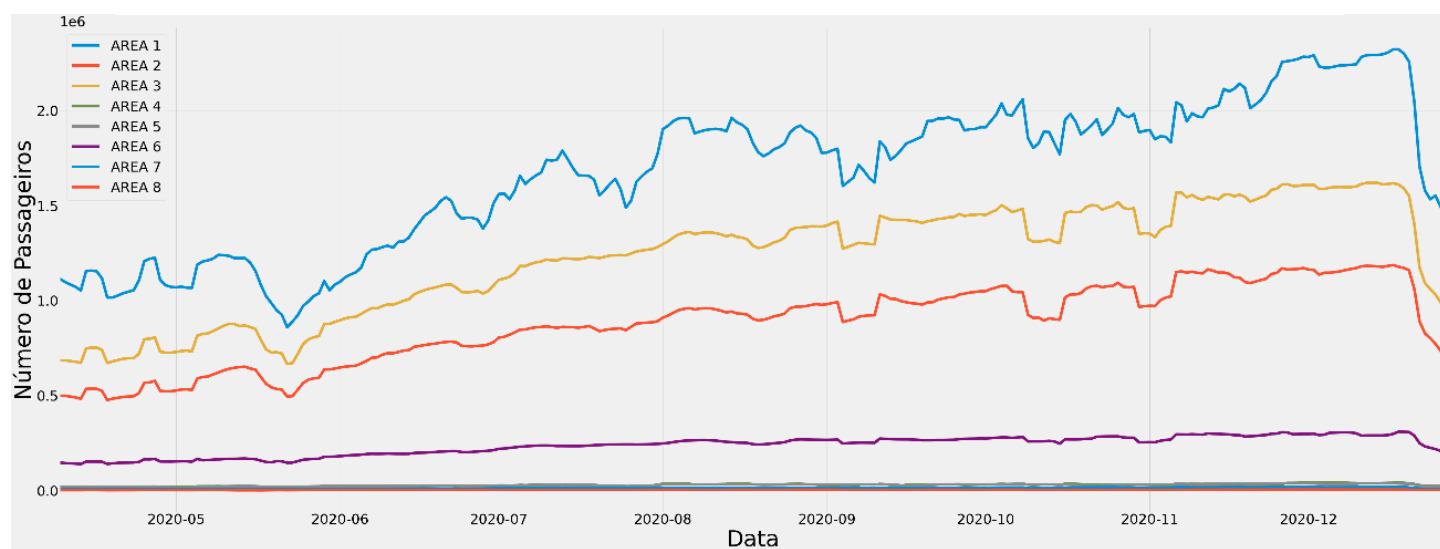
**Fonte: Elaborado pelo autor**

**Gráfico 4.13 – Série de dados filtrados pela tendência de casos de Covid**

**Fonte: Elaborado pelo autor**

**Gráfico 4.15 – Série de dados brutos de passageiros por distrito**

Fonte: Elaborado pelo autor

**Gráfico 4.16 – Série de dados filtrados pela tendência de passageiros por distrito**

Fonte: Elaborado pelo autor

### 4.4.3. Resultados dos modelos

Considerando que será utilizado dois grupos de dados com tratativas distintas (com e sem filtro) e se fará teste de qualidade para cada conjunto de dado, será particionado ambos em dois grupos com as seguintes proporções e finalidades:

- 80% dos dados serão ajustados no modelo de regressão linear e será por meio deles que se ajustará os coeficientes da reta;
- 20% dos dados serão utilizados como base de comparação. Ou seja, após os coeficientes ajustados, se calculará o valor de casos pela equação do modelo, e assim se comparará o valor de predição com o valor real, montando assim métricas como o  $\chi^2$ .

Vale ressaltar, que não existe uma regra para determinar a divisão entre dados de modelagem e dados de teste. Desta forma, caso fosse escolhido outro tipo de proporção, mesmo assim estaria correto. No entanto, é importante escolher uma proporção que os dados de modelagem sejam suficientes para que os coeficientes tenham valores sólidos.

#### 4.4.3.1. Sem filtro de tendência

Definido as variáveis, o resultado da comparação de passageiros por distrito com o número de covid passados 5 dias está apresentado pela Tabela 4.7

**Tabela 4.7 – Resultados para os dados brutos**

<b>Dep. Variable:</b>	C_T	<b>R-squared (uncentered):</b>	0,595
<b>Model:</b>	OLS	<b>Adj. R-squared (uncentered):</b>	0,582
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	47,90
<b>Date:</b>	Sun, 14 Nov 2021	<b>Prob(F-statistic):</b>	5.12e-47
<b>Time:</b>	13:12:09	<b>Log-Likelihood:</b>	-2289,0
<b>No. Observations:</b>	269	<b>AIC:</b>	4594
<b>Df Residuals:</b>	261	<b>BIC:</b>	4623
<b>Df Model:</b>	8	<b>Covariance Type:</b>	nonrobust

*Fonte: Elaborado pelo autor*

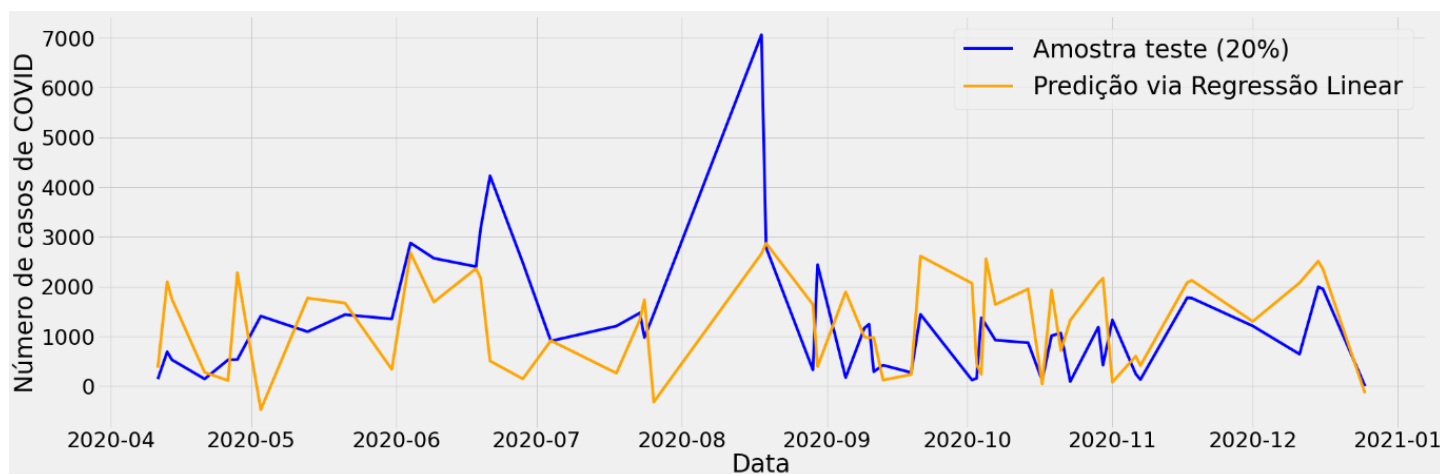
**Tabela 4.8 – Resultados dos coeficientes do modelo para os dados brutos**

ÁREA	Coef	std err	t	P> t	[0.025	0.975]
1	-0,0029	0,001	-0,29920	0,003	-0,005	-0,001
2	0,0003	0,003	0,108	0,914	-0,005	0,006
3	0,0001	0,003	0,039	0,969	-0,005	0,005
4	0,0441	0,013	3,374	0,001	0,018	0,07
5	0,044	0,027	1,635	0,103	-0,009	0,097
6	0,0191	0,005	3,902	0,000	0,009	0,029
7	-0,1887	0,064	-2,929	0,004	-0,316	-0,062
8	0,3598	0,135	2,664	0,008	0,094	0,626

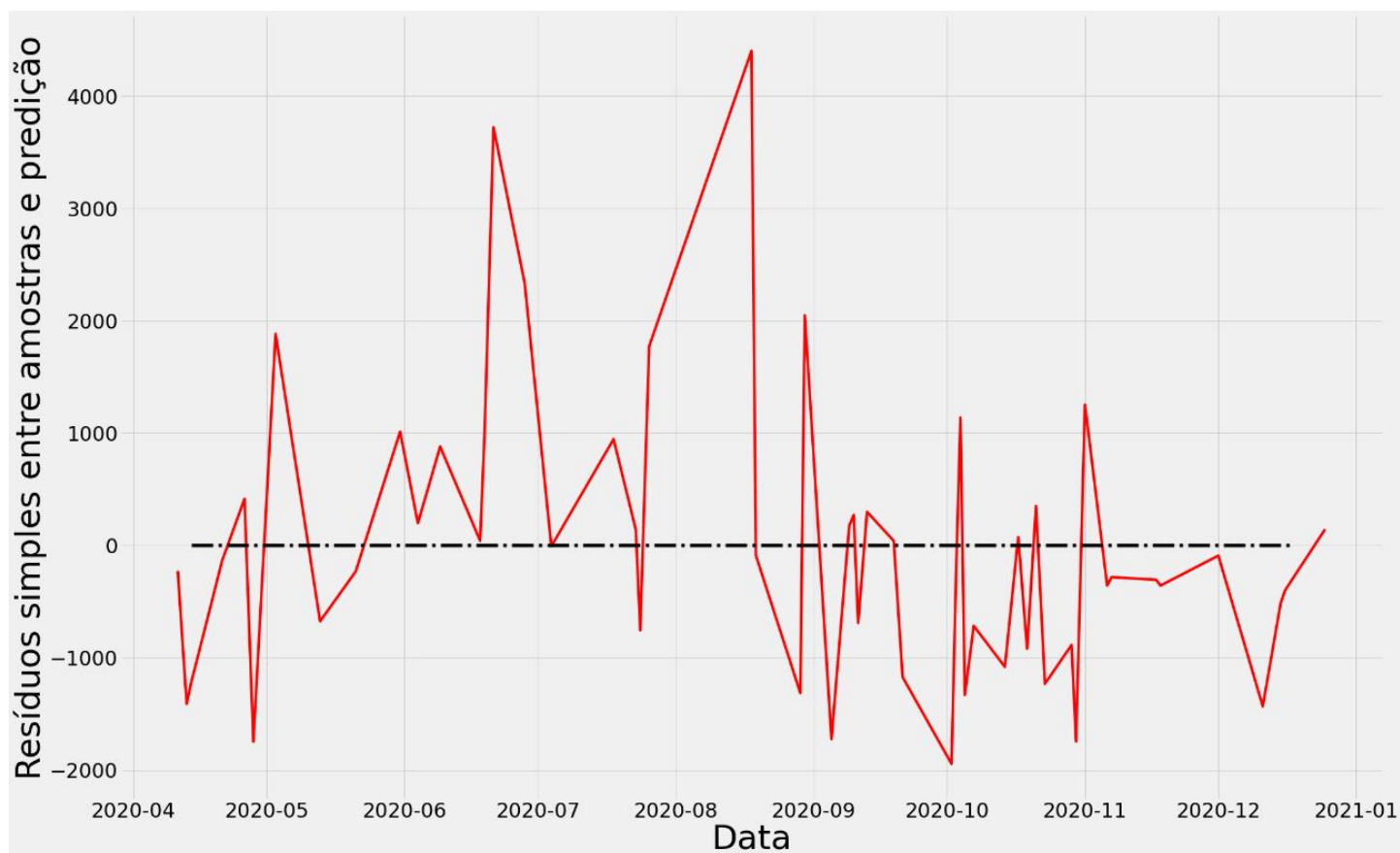
*Fonte: Elaborado pelo autor*

Considerando como métrica o  $\chi^2$  ajustado (visto que está se utilizando mais de uma variável independente), tem-se que o modelo de regressão linear não tem um excelente ajuste. Mesmo assim, quando visto o mesmo cenário graficamente, a análise é um pouco mais distinta, como indicado pelo Gráfico 4.17 e o Gráfico 4.18.

**Gráfico 4.17- Correlação entre o valor real com o valor ajustado para os dados brutos**



*Fonte: Elaborado pelo autor*

**Gráfico 4.18 – Resíduos para os dados brutos**

*Fonte: Elaborado pelo autor*

Dessa forma, mesmo com um valor de  $\text{Chi}^2$  alto, os resíduos na realidade têm um comportamento sazonal. Isso quando analisado os resultados do período de maio até setembro, em que o valor ajustado está mais distante do real. Esta característica pode ser decorrente de uma mudança de comportamento, ou medidas restritivas. Mesmo assim, como uma ferramenta preditiva, associada com a facilidade de aplicação, o ajuste torna-se uma importante ferramenta.

#### **4.4.3.2. Com filtro de tendência**

Assim como feito com os dados brutos, a mesma tratativa foi feita com os dados com filtro de tendência, sendo os resultados apresentados pela Tabela 4.9 e a Tabela 4.10.

**Tabela 4.9 - Resultados para os dados com filtro de tendência**

Dep. Variable:	C_T	R-squared (uncentered):	0,873
Model:	OLS	Adj. R-squared (uncentered):	0,869
Method:	Least Squares	F-statistic:	219.8
Date:	Sun, 14 Nov 2021	Prob (F-statistic):	9,21E-110
Time:	13:12:09	Log-Likelihood:	-2043,4
No. Observations:	263	AIC:	4103
Df Residuals:	255	BIC:	4131
Df Model:	8	Covariance Type:	nonrobust

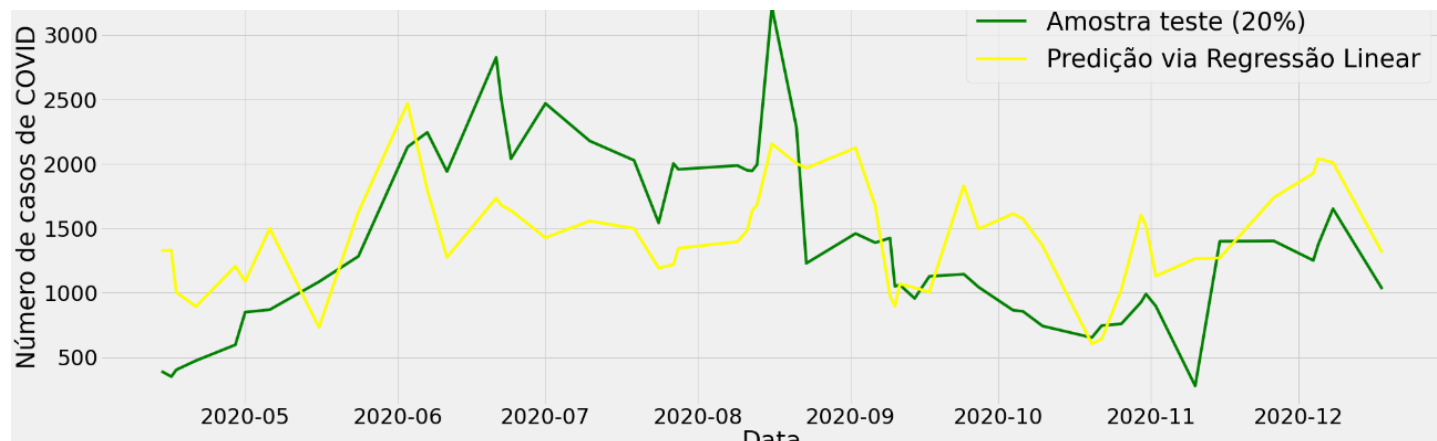
Fonte: Elaborado pelo autor

**Tabela 4.10 - Resultados dos coeficientes do modelo para os dados com filtro de tendência**

ÁREA	coef	std err	t	P> t	[0.025	0.975]
1	-0,0035	0,0010	-3,5500	0,0000	-0,0050	-0,0020
2	-0,0027	0,0020	-1,2480	0,2130	-0,0070	0,0020
3	0,0050	0,0030	1,9500	0,0520	0,0000	0,0100
4	0,0488	0,0130	3,7030	0,0000	0,0230	0,0750
5	0,1019	0,0200	4,9900	0,0000	0,0620	0,1420
6	0,0043	0,0070	0,6230	0,5340	-0,0090	0,0180
7	-0,3238	0,0740	-4,3550	0,0000	-0,4700	-0,1770
8	0,6696	0,1280	5,2230	0,0000	0,4170	0,9220

Fonte: Elaborado pelo autor

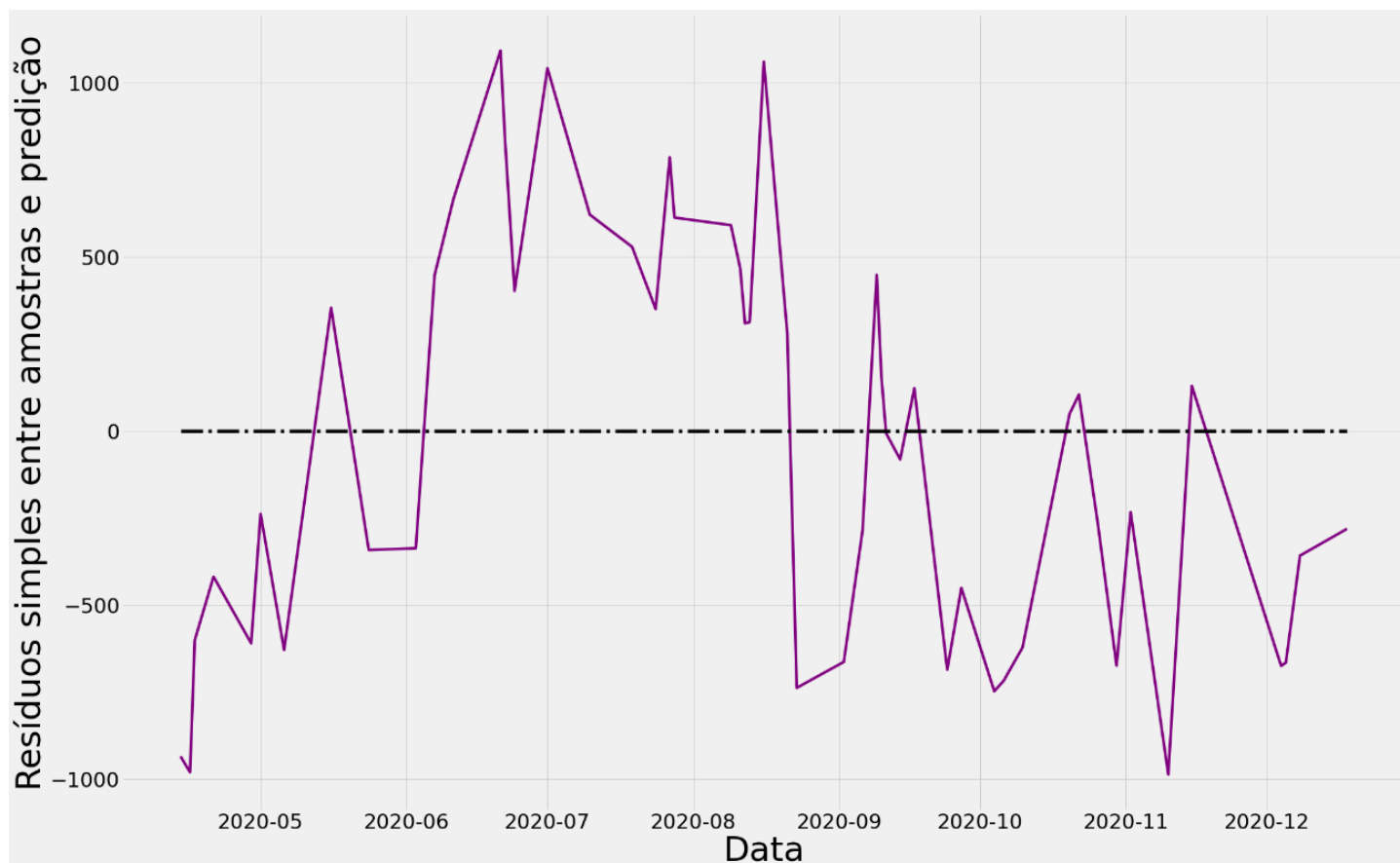
Para se analisar o comportamento sobre os dados ajustados com o real, tem-se o Gráfico 4.19 e o Gráfico 4.20.

**Gráfico 4.19 - Correlação entre o valor real com o valor ajustado para os dados com filtro de tendência**

Fonte: Elaborado pelo autor

O interessante é que, assim como ocorrido com os dados sem filtro, no período de maio até setembro, os resíduos são os mais altos. O que indica que mesmo com filtro, essa região apresenta um comportamento que é distinto das outras.

**Gráfico 4.20 - Resíduos para os dados com filtro de tendência**



*Fonte: Elaborado pelo autor*

#### 4.4.3.3. Comparação entra os dois modelos

Analisando o  $\chi^2$ , vê-se que o modelo com os dados sem filtro tem uma aproximação com o real muito maior do que comparado com o com filtro (com um  $\chi^2$  ajustado de 0,582 contra 0,869 do com filtro de tendência).

Um elemento também bastante importante é que as regiões tiveram coeficientes bastante semelhantes entre os dois ajustes. Isso quando se compara as proporções entre os coeficientes de ajuste. Entretanto mais especificamente, o comportamento dos distritos sobre cada ajuste foi ligeiramente distinto.



Este fato advém quando se analisa os valores de *p-value* de cada variável. No modelo sem filtro os distritos 2 e 3 foram aqueles com maior valor de *p-value*. Desta maneira, eles são aqueles que mais propiciam um maior resíduo quando comparado com o valor de predição com o valor real. Já para o modelo com filtro, os distritos com maior *p-value* são o distrito 2 e 6.

No presente trabalho, não será desenvolvido o motivo desta possível diferença. Contudo, já se esperava uma diferença entre distintos, principalmente devido a diferença socioeconômica apresentado pelas distintas regiões. Ou seja, por conta de que para cada estratificação social apresenta um comportamento no que diz respeito a escolha de transporte. Ou seja, era esperado um desvio quando compilados todas as regiões paulistanas em um mesmo modelo estatístico.

Considerando os resultados dos dois testes, para este trabalho será considerado como modelo aquele em que não se fez o filtro de tendência. E tal escolha é fundamentada em dois grandes princípios: facilidade de aplicabilidade e melhor ajuste quando comparado o valor real com o valor de predição.

No caso da aplicabilidade, como para a execução deste modelo, utilizou-se os dados brutos, é muito mais fácil para uma pessoa que utiliza apenas uma planilha como ferramenta de cálculo conseguir calcular a relação entre passageiros com o valor futuro de casos de Covid na região do município de São Paulo.

No caso do ajuste, como já comparado, o valor de  $\text{Chi}^2$  foi menor do que comparado com o valor de quando filtrado. Desta forma, com os dados não filtrados os resíduos são muito menores. Contudo, este comportamento não é representado no gráfico de resíduos, visto que o valor máximo é de 4000 para os dados não filtrado, enquanto para o filtrado é de 1000.

Isto ocorre pelo fato de que, como já indicado anteriormente, os meses de maio até setembro tiveram um comportamento distinto quando comparado com outras regiões do ano. Ou seja, pode-se considerar os dados desta região como outliers. E muito provavelmente, devido a estes dados, o valor de  $\text{Chi}^2$  foi mais elevado nos dados filtrados, visto que quando se analisa a aplicabilidade de filtros de tendência, outliers tendem a espalhar seu comportamento em maiores períodos.

Ou seja, quando visto os outliers nos dados brutos, os comportamentos distintos se concentram em regiões muito específicas a um nível diário. Já no caso da aplicação do filtro, que se utilizou como método a tendência semanal (período de 7

dias), estes valores pontuais se espalharam para outros períodos aumentando, desta forma, a média semanal dos dados como um todo.

Em suma, escolhido o modelo e a estruturação dos dados (sem filtro de tendência), a equação que descreve o comportamento é descrita como:

$$C_{5 \text{ dias}} = -0,0029 \times P_{A_1} + 0,0003 \times P_{A_2} + 0,0001 \times P_{A_3} - 0,0441 \times P_{A_4} + 0,044 \times P_{A_5} \\ + 0,0191 \times P_{A_6} - 0,1887 \times P_{A_7} + 0,3598 \times P_{A_8}$$

Com esta equação, considerando os dados registrados a nível dia, é possível saber se haverá ou não um pico de casos de Covid nos próximos 5 dias sabendo o número de passageiros divididos por distritos, sem a necessidade de se aplicar nenhum filtro de tendência.

## 5. CONCLUSÃO

Considerando que objetivo principal do trabalho era encontrar uma correlação entre o número de passageiros transportados pela SPTrans com o número de casos de Covid no município de São Paulo, concluindo-se de pronto, ele foi atingido.

Vale apontar algumas ressalvas. A principal, que permeará durante a utilização do modelo, é o fato de que os dados oficiais dos casos de Covid no Brasil são considerados, por muitos pesquisadores, menores do que quando comparado com dados oficiais de outros países.

Ou seja, por conta da subnotificação de casos de Covid, por mais que as variáveis estatísticas indiquem que o ajuste do modelo está aceitável para a proposição, os valores de saída do modelo serão sempre menores do que os apresentados na realidade.

Mesmo assim, vale ressaltar que a aplicabilidade do trabalho é extremamente alta. Para conseguir avaliar um possível aumento de casos decorrentes do número de passageiros divididos por distritos, basta utilizar uma planilha eletrônica, sem a necessidade do conhecimento de programação ou de modelagem estatística.

Com isso, qualquer gestor ou servidor público consegue ter tomadas de decisão mais assertivas e fundamentadas com um viés estatístico. Facilitando com isso, a tomada de decisão das medidas mais flexíveis sobre o transporte, acompanhando e redirecionando as linhas de ônibus sobre a região metropolitana de São Paulo.

Pode-se até mesmo utilizar diferentes modais, ou flexibilizar horários, com o objetivo de controlar as curvas epidemiológicas, sem a paralização total das linhas de ônibus.

## 6. REFERÊNCIAS

BRASIL. [S.l.]: Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso xxxiii do art. 5º, no inciso ii do § 3º do art. 37 e no § 2º do art. 216 da constituição federal; altera a lei nº 8.112, de 11 de dezembro de 1990; revoga a lei nº, 2011.

BRAUNER, J. M. et al. Inferring the effectiveness of government interventions against COVID-19. **Science**, 2021. Disponível em: <<https://www.science.org/doi/10.1126/science.abd9338>>.

BURTON, E.; WILLIAMS, K.; JENKS, M. **The Compact City: A Sustainable Urban Form?** 1ª. ed. [S.l.]: Taylor & Francis Ltd, 1996. ISBN 978-0419213000.

BUSSAB, W. D. O.; MORETTIN, P. A. **Estatística Básica**. 6. ed. [S.l.]: Saraiva, 2010. ISBN 978-85-02-08177-2.

GUERCHE-SÉBLAIN, C. E. et al. Experience from five Asia-Pacific countries during the first wave of the COVID-19 pandemic: Mitigation strategies and epidemiology outcomes. **Travel Medicine and Infectious Disease**, 2021. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S147789392100212X>>.

HARRIS, J. E. The Subways Seeded the Massive Coronavirus Epidemic in New York City. **NBER**, 2020. Disponível em: <<https://www.nber.org/papers/w27021>>.

LAVIE, G. et al. Change in hospitalizations and 30-day mortality of patients with acute myocardial infarction during the first COVID-19 lockdown – A pure social isolation effect? **Cardiovascular Revascularization Medicine**, 2021. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S155383892100631X>>.

LEITE, C. et al. Indicadores de desigualdade para financiamento urbano de cidades saudáveis. **USP Cidades globais**, 2019. Disponível em: <<https://www.scielo.br/j/ea/a/CGzJ3yXFy8tVyzpnNjf7n8r/?lang=pt>>.

LEIVA, G. D. C.; REIS, D. S. D.; FILHO, R. D. O. Estrutura urbana e mobilidade populacional: implicações para o distanciamento social e disseminação da Covid-19. **Rebep**, 2020. Disponível em: <<https://www.scielo.br/j/rbepop/a/xn4pKxkvHTtSFX65HC5L3zp/?lang=pt>>.

LEONARD MLODINOW. **O Arco-Iris De Feynman**. [S.l.]: Sextante, 2005.

METRO CPTM. **Em mapa discreto, Metrô dá pistas sobre futuras linhas e expansões**, 2019. Disponível em: <<https://www.metrocptm.com.br/em-mapa-discreto-metro-da-pistas-sobre-futuras-linhas-e-expansoes/>>. Acesso em: dez. 2020.

OMS. **Coronavirus disease (COVID-19) pandemic**, 2019. Disponível em: <<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>>. Acesso em: abr. 2020.

PRADHAN, B. **Spatial Modeling and Assessment of Urban Form**. [S.l.]: Springer, v. I, 2017. ISBN 978-3-319-54216-4.

PRIESTLEY, M. B. **Developments in Time Series Analysis**. [S.l.]: Springer, 1993. ISBN 978-1-4899-4515-0.

REDE NOSSA SÃO PAULO. **Mobilidade Urbana**, 2020. Disponível em: <[https://www.nossasaopaulo.org.br/wp-content/uploads/2019/09/Apresentacao\\_Pesquisa\\_ViverEmSP\\_MobilidadeUrbana\\_2019.pdf](https://www.nossasaopaulo.org.br/wp-content/uploads/2019/09/Apresentacao_Pesquisa_ViverEmSP_MobilidadeUrbana_2019.pdf)>. Acesso em: dez. 2020.

REDE NOVA SÃO PAULO. **Mapa da Desigualdade 2020 revela diferenças entre os distritos da capital paulista**, 2020. Disponível em: <<https://www.nossasaopaulo.org.br/2020/10/29/mapa-da-desigualdade-2020-revela-diferencas-entre-os-distritos-da-capital-paulista/>>. Acesso em: dez. 2020.

SEADE. **Repositório de dados sobre casos e óbitos decorrentes do COVID-19 nos municípios do Estado de São Paulo e sobre leitos e internações por Departamento Regional de Saúde**, 2020. Disponível em: <<https://github.com/seade-R/dados-covid-sp>>. Acesso em: jan. 2020.

SECRETARIA MUNICIPAL E TRÂNSITO. **Passageiros Transportados - 2020**, 2020. Disponível em: <[https://www.prefeitura.sp.gov.br/cidade/secretarias/mobilidade/institucional/sptrans/ acesso\\_a\\_informacao/agenda/index.php?p=292723](https://www.prefeitura.sp.gov.br/cidade/secretarias/mobilidade/institucional/sptrans/ acesso_a_informacao/agenda/index.php?p=292723)>.

SHEN, M. et al. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. **Biorxiv**, 2020. Disponível em: <<https://www.biorxiv.org/content/10.1101/2020.01.23.916726v1>>.

SHEU, J. C.; MCKAY, D.; ERIC A. STORCH. COVID-19 and OCD: Potential impact of exposure and response prevention therapy. **Journal of Anxiety Disorders**, v. Volume 76, 2020. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0887618520301286>>.

SPTRANS. **Relatório Íntegro da Administração 2020**, 2020. Disponível em: <<https://www.sptrans.com.br/relatorio-integrado-da-administracao-2020/>>.

VISHWAS, B. V.; PATEL, A. **Hands-on Time Series Analysis with Python**. [S.l.]: Apress, 2020. ISBN 978-1-4842-5992-4.