

# **Análise da mobilidade ferroviária e suas possíveis interações na região metropolitana da cidade do Rio de Janeiro**



## **Resumo**

Este relatório apresenta uma investigação detalhada sobre os dados relacionados ao fluxo mensal total nas vias ferroviárias operadas pela concessionária Supervia na cidade do Rio de Janeiro e eventuais efeitos na Região metropolitana do Rio de Janeiro, no período desde 2019 a maio de 2024.

O estudo também avalia a dinâmica do desenvolvimento da capital do Rio de Janeiro bem como busca possíveis interações com as regiões residenciais e atividades econômicas em sua área de abrangência.

## **Abstract**

This report presents a detailed investigation into data related to the total monthly flow on railway routes operated by the Supervia through Concession contract in the city of Rio de Janeiro and possible effects in the metropolitan region of Rio de Janeiro, in the period from 2019 to May 2024.

The study also evaluates the dynamics of development in the capital of Rio de Janeiro as well as looking for possible interactions with residential regions and economic activities in its area of coverage.

## Introdução temática / teórica

Atualmente, a SuperVia atende a região metropolitana do Rio de Janeiro (Rio de Janeiro, Duque de Caxias, Nova Iguaçu, Nilópolis, Mesquita, Queimados, São João de Meriti, Belford Roxo, Japeri, Magé, Paracambi e Guapimirim), através de uma malha ferroviária de 270 quilômetros dividida em cinco ramais e três extensões.

Segundo o estudo da CNT (2016) **Transporte metroferroviário de passageiros**<sup>1</sup>, o sistema de trens metropolitanos conta com 103 estações (15 estações são acessíveis), sendo 86 locais, seis integradas e 11 terminais. A integração com ônibus e táxis é feita em 84 estações, com o metrô em cinco, com o BRT em seis e com o teleférico do Morro do Alemão em duas estações. Há ainda integração tarifária com o metrô (Bilhete Único), ônibus intermunicipal (Bilhete Único), ônibus municipal (Bilhete Único Carioca) e BRT (Bilhete Único Carioca).

A empresa transporta atualmente por meio de 201 trens em operação, em média, 300.000 passageiros, o que representa uma queda de 50% dos passageiros transportados em 2018. Em 17/08/2016, o recorde de 735 mil passageiros foi alcançado.

Fazem parte do material rodante 786 carros. As composições são formadas por seis carros em cada trem, sendo o intervalo entre trens de 300 segundos (cinco minutos). Trabalham, na SuperVia, 2.464 empregados e 2.324 funcionários terceirizados. Predomina, na tração do sistema, a do tipo elétrica por catenária autocompensada ou flexível (147 km). Há ainda 18 km com tração elétrica por catenária rígida e 55 km com tração por locomotiva a diesel. O regime de contratação de energia é cativo, sendo, em 2015, o consumo energético de tração de 249.860 MWh e, de auxiliares, de 13.151 MWh.

| Dados operacionais                  | 2011    | 2012    | 2013    | 2014    | 2015    |
|-------------------------------------|---------|---------|---------|---------|---------|
| Extensão operacional (km)           | 220     | 220     | 220     | 220     | 220     |
| Número de estações                  | 98      | 100     | 102     | 102     | 102     |
| Número de Linhas (ramais)           | 7       | 7       | 7       | 7       | 7       |
| Total de carros                     | 613     | 663     | 663     | 768     | 786     |
| Entrada de passageiros/ano (mil)    | 142.292 | 143.592 | 152.352 | 164.002 | 178.079 |
| Passageiros transportados/ano (mil) |         | 150.772 | 159.970 | 180.733 | 203.366 |

<sup>1</sup> <http://repositorio.itl.org.br/jspui/handle/123456789/186>

A nível estadual, a entidade responsável pela regulação dos sistemas metroferroviários é a Agência Reguladora dos Serviços Públicos Concedidos de Transportes Aquaviários, Ferroviários e Metroviários e de Rodovias do Estado do Rio de Janeiro – AGETRANSP.

As origens da malha ferroviária em operação remontam ao ano de 1858. O sistema ferroviário do Rio de Janeiro, durante algumas décadas, se restringiu somente ao transporte de carga de longo percurso. Porém, a partir de 1930, o transporte ferroviário urbano se expandiu continuamente. Em 1962 foi registrado o maior fluxo de passageiros, cerca de 262.700.000, transportados em um ano nos trens urbanos do Rio de Janeiro. No entanto, já na década de 50, o sistema de trens de passageiros começou a sofrer competições bastante desvantajosas do sistema rodoviário e, em menos de 20 anos, por razões econômicas, ocorreu a desativação de grande parte da rede.

Em um contexto mais recente e na sequência do já referido processo de estadualização dos sistemas da CBTU, o sistema de trens metropolitanos do Rio de Janeiro foi entregue, em 1994, à Companhia Fluminense de Trens Urbanos – Flumitrens. Em 1997, os trens de passageiros e o metrô – foram incluídos no Programa Estadual de Desestatização. Em 1998, a operação e a manutenção do serviço de trens foram concedidas, por meio de licitação, ao consórcio Bolsa 2000, por um período de 25 anos renováveis por mais 25 anos. Ainda em 1998, a empresa SuperVia, criada pelo consórcio, assumiu essa operação na região metropolitana. Em 2001, a Companhia Estadual de Engenharia de Transportes e Logística – Central assumiu a operação da linha de trens de Niterói a Itaboraí e do sistema de bondes que ligam o centro da cidade do Rio de Janeiro ao bairro de Santa Teresa – malhas que não fizeram parte da licitação de 1998. Em 2010, a concessão da SuperVia foi estendida até o ano de 2018 e, em 2011, a SuperVia passa a integrar o grupo a Odebrecht TransPort empresa com empreendimentos urbanos, sistemas integrados a aeroportos e logística.

Em 07 de junho de 2021, o Grupo SuperVia apresentou pedido de recuperação judicial, com o intuito de adequar sua situação financeira à nova conjuntura da economia do país e permitir sua reorganização. Constituída em 02 de setembro de 1998, a Supervia, empresa que opera o principal serviço do Grupo SuperVia, é detentora de direitos e obrigações decorrentes da concessão para a prestação de serviços de transporte ferroviário de passageiros na região metropolitana do Rio de Janeiro e para

execução dos investimentos contemplados no Contrato de Concessão e aditamentos posteriores, em caráter exclusivo, inicialmente pelo prazo de 25 anos e atualmente prorrogado, através do 8º Termo Aditivo ao Contrato de Concessão, até o ano de 2048.

## Metodologia

No artigo ***Geração de embarques nas estações de trem da cidade do Rio de Janeiro***<sup>2</sup>, Rafael Lara, Rosana Brandão e Licínio Portugal, da COPPE/UFRJ, destacam que “O transporte sobre trilhos, particularmente o metroferroviário, abrange não somente questões de cunho funcional das áreas urbanas, mas compreendem variáveis de diversas naturezas, especialmente aquelas que dizem respeito ao perfil socioeconômico da região. Com base nisto se justifica a necessidade do estudo dos fatores que influenciam o uso dos modais trem e metrô nas metrópoles do Brasil.”

Utilizando a regressão simples na qual cada modelo contempla apenas a interferência de uma variável explicativa, sendo a variável dependente a taxa de embarque. Como resultado, houve uma boa aderência entre o número de imóveis comerciais com as taxas de embarques, baixa correlação com o Índice de Desenvolvimento Humano (IDH) das regiões próximas às estações de trens e razoável correlação com nível de empregos, particularmente no Comércio.

Este estudo baseia-se na análise de dados provenientes da base de dados disponíveis da **Supervia**, em atendimento à solicitação de informações através do excelente sistema RJDigital:

Seu pedido de informação de protocolo 20240606815058 para o órgão AGETRANSP foi concluído. em 15:34 - 25/06/2024.

**Protocolo:** 20240606815058

**Status:** Concluído

**Descrição:**

Segue em anexo Resposta Ouvidoria da Agetransp e planilhas solicitadas.

---

<sup>2</sup> [http://redpgv.coppe.ufrj.br/arquivos/Rafael\\_Txs\\_Embarque\\_Trem\\_RiodeTransp\\_31.0.pdf](http://redpgv.coppe.ufrj.br/arquivos/Rafael_Txs_Embarque_Trem_RiodeTransp_31.0.pdf)

## Código Python para mostrar a tabela de estações e ramais dos arquivos Geojson:

```
import geopandas as gpd

# Carregar o arquivo GeoJSON com as estações
estacoes =
gpd.read_file('/content/drive/MyDrive/Python/TrensRJ/supervia_stations.geojson')

# Exibir todas as colunas do arquivo de estações
print("Colunas disponíveis no arquivo 'supervia_stations.geojson':")
print(estacoes.columns)

# Carregar o arquivo GeoJSON com as linhas de trem
trens_rj =
gpd.read_file('/content/drive/MyDrive/Python/TrensRJ/Trajeto_Trem.geojson.json')

# Exibir todas as colunas do arquivo de trajetos
print("\nColunas disponíveis no arquivo 'Trajeto_Trem.geojson.json':")
print(trens_rj.columns)

# Criar a tabela de estações com as colunas disponíveis
tabela_estacoes = estacoes[['estacao', 'ramal', 'latitude', 'longitude', 'cep', 'geometry']]

# Exibir a tabela de estações
print("\nTabela de estações:")
print(tabela_estacoes)

# Criar uma tabela com as colunas que podem ser relevantes dos trajetos
# Neste caso, não temos 'objectid' ou 'name', mas podemos usar 'ramal'
# Se houver alguma relação entre estações e ramais nos trajetos
tabela_trens = trens_rj[['ramal', 'geometry']]

# Exibir a tabela de trajetos
print("\nTabela de trajetos:")
print(tabela_trens)
```

```
Colunas disponíveis no arquivo 'supervia_stations.geojson':
Index(['estacao', 'sigla', 'ramal', 'latitude', 'longitude', 'nomeantigo',
      'bicicletario', 'km', 'endereco', 'cep', 'geometry'],
      dtype='object')
```

```
Colunas disponíveis no arquivo 'Trajeto_Trem.geojson.json':
Index(['objectid', 'ramal', 'flg_ativa', 'status', 'data_inc',
      'st_lengthshape', 'geometry'],
      dtype='object')
```

Tabela de estações:

|     | estacao             | ramal          | latitude     | longitude    | \ |
|-----|---------------------|----------------|--------------|--------------|---|
| 0   | Adeus               | Teleférico A   | -22.86529824 | -43.26119648 |   |
| 1   | Agostinho Porto     | Belford Roxo   | -22.78860158 | -43.38142338 |   |
| 2   | Alemão              | Teleférico A   | -22.85825628 | -43.27096508 |   |
| 3   | Anchieta            | Japeri         | -22.82538659 | -43.40239292 |   |
| 4   | Augusto Vasconcelos | Santa Cruz     | -22.89284855 | -43.53715455 |   |
| ..  | ...                 | ...            | ...          | ...          |   |
| 100 | Triagem             | Saracuruna     | -22.89677514 | -43.24413659 |   |
| 101 | Vigário Geral       | Saracuruna     | -22.80565025 | -43.30542913 |   |
| 102 | Vila Inhomirim      | Vila Inhomirim | -22.57145329 | -43.18484303 |   |
| 103 | Vila Militar        | Santa Cruz     | -22.86273998 | -43.40220459 |   |
| 104 | Vila Rosali         | Belford Roxo   | -22.79406425 | -43.37530791 |   |

|     | cep       | geometry                    |
|-----|-----------|-----------------------------|
| 0   | 21061-640 | POINT (-43.2612 -22.8653)   |
| 1   | 25550-660 | POINT (-43.38142 -22.7886)  |
| 2   | 21061-240 | POINT (-43.27097 -22.85826) |
| 3   | 21645-010 | POINT (-43.40239 -22.82539) |
| 4   | 23013-440 | POINT (-43.53715 -22.89285) |
| ..  | ...       | ...                         |
| 100 | 20911-270 | POINT (-43.24414 -22.89678) |
| 101 | 21215-370 | POINT (-43.30543 -22.80565) |
| 102 | 25900-000 | POINT (-43.18484 -22.57145) |
| 103 | 21615-310 | POINT (-43.4022 -22.86274)  |
| 104 | 25510-220 | POINT (-43.37531 -22.79406) |

[105 rows x 6 columns]

Tabela de trajetos:

|   | ramal          | geometry                                          |
|---|----------------|---------------------------------------------------|
| 0 | Guapimirim     | LINESTRING (-43.24686 -22.67769, -43.24592 -22... |
| 1 | Vila Inhomirim | LINESTRING (-43.24708 -22.67778, -43.2441 -22.... |
| 2 | Santa Cruz     | MULTILINESTRING ((-43.38349 -22.85502, -43.384... |
| 3 | Japeri         | MULTILINESTRING ((-43.38349 -22.85502, -43.384... |
| 4 | Belford Roxo   | LINESTRING (-43.19186 -22.90354, -43.19479 -22... |
| 5 | Saracuruna     | LINESTRING (-43.19186 -22.90354, -43.19479 -22... |
| 6 | Paracambi      | LINESTRING (-43.65298 -22.64429, -43.65482 -22... |
| 7 | Deodoro        | LINESTRING (-43.19186 -22.90354, -43.19479 -22... |

**Código Python para cálculos estatísticos da quantidade de passageiros por estações ao longo de 2019 à 2024:**

```
import pandas as pd

# Caminho do arquivo
file_path = '/content/drive/MyDrive/Python/TrensRJ/Dados_Supervia.xlsx'

# Carregar a planilha "Supervia2019_2024"
df = pd.read_excel(file_path, sheet_name='Supervia2019_2024')
```

```
# Converter o cabeçalho (que contém as datas no formato DD/MM/AAAA) para
datetime, ignorando as colunas não-data
colunas = []
for col in df.columns:
    try:
        colunas.append(pd.to_datetime(col, format='%d/%m/%Y'))
    except (ValueError, TypeError):
        colunas.append(col) # Se não for uma data, manter como está

df.columns = colunas

# Definir intervalo de datas para filtrar (janeiro/2019 a março/2024)
start_date = pd.to_datetime('2019-01-01')
end_date = pd.to_datetime('2024-03-31')

# Filtrar as colunas que estão no intervalo de datas desejado
colunas_data = [col for col in df.columns if isinstance(col, pd.Timestamp)
and start_date <= col <= end_date]

# Criar um DataFrame para armazenar as médias trimestrais
media_trimestral = pd.DataFrame()

# Percorrer cada estação e calcular a média trimestral
for station in df['Estação'].unique(): # Supondo que há uma coluna
'Estação' no DataFrame
    dados_estacao = df[df['Estação'] == station][colunas_data]

    # Converter todos os dados para numérico, valores inválidos serão
convertidos para NaN
    dados_estacao = dados_estacao.apply(pd.to_numeric, errors='coerce')

    # Transpor os dados para que cada coluna seja um trimestre
    dados_estacao_transposta = dados_estacao.T
    dados_estacao_transposta.index =
pd.to_datetime(dados_estacao_transposta.index)

    # Agrupar os dados por trimestres e calcular a média
    media_trimestre = dados_estacao_transposta.resample('Q').mean()

    # Transpor para ter as médias trimestrais em colunas e adicionar a
estação como índice
    media_trimestre_transposta = media_trimestre.T
    media_trimestre_transposta['Estação'] = station

    # Concatenar com o DataFrame de médias trimestrais
    media_trimestral = pd.concat([media_trimestral,
media_trimestre_transposta.set_index('Estação')])
```

```
# Resetar o índice para exibir a tabela corretamente e ordenar as colunas
por data
media_trimestral = media_trimestral.sort_index(axis=1)

# Exibir a tabela completa com as médias trimestrais por estação
print(media_trimestral)

# Opcional: salvar o resultado em um arquivo Excel
media_trimestral.to_excel('/content/drive/MyDrive/Python/TrensRJ/média_trim
estral_supervia.xlsx')
```

|                     | Média         | Mediana  | Desvio Padrão | Variância    |
|---------------------|---------------|----------|---------------|--------------|
| Augusto Vasconcelos | 47973.539683  | 43813.0  | 11048.337029  | 1.220658e+08 |
| Agostinho Porto     | 9082.393443   | 6953.0   | 4965.053653   | 2.465176e+07 |
| Anchieta            | 46564.158730  | 43772.0  | 11902.728426  | 1.416749e+08 |
| Austin              | 148651.206349 | 141290.0 | 38725.108598  | 1.499634e+09 |
| Benjamin do Monte   | 36796.555556  | 30611.0  | 12873.138726  | 1.657177e+08 |
| ...                 | ...           | ...      | ...           | ...          |
| Triagem             | 53950.571429  | 47989.0  | 18078.784780  | 3.268425e+08 |
| Vigário Geral       | 18292.412698  | 15327.0  | 7612.376818   | 5.794828e+07 |
| Vila Inhomirim      | 3295.000000   | 3059.0   | 1976.763608   | 3.907594e+06 |
| Vila Militar        | 17397.746032  | 16320.0  | 4802.122130   | 2.306038e+07 |
| Vila Rosali         | 1902.344262   | 1640.0   | 707.505521    | 5.005641e+05 |

[85 rows x 4 columns]

## Código Python para gerar gráfico BoxPlot por Ramais ao longo de 2019 à 2024:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# 1. Carregar os dados e transformar para formato longo
file_path = '/content/drive/MyDrive/Python/TrensRJ/Dados_Supervia.xlsx'
df = pd.read_excel(file_path, sheet_name='Supervia2019_2024')
df_long = df.melt(id_vars=['Estação', 'Ramal'], var_name='Data',
value_name='Valor')
df_long['Data'] = pd.to_datetime(df_long['Data'])

# 2. Converter coluna 'Valor' para numérica
df_long['Valor'] = pd.to_numeric(df_long['Valor'], errors='coerce')

# 3. Remover valores ausentes
df_long = df_long.dropna(subset=['Valor'])

# 4. Filtrar período desejado
start_date = '2019-01-01'
```

```
end_date = '2024-03-31'
df_filtered = df_long[(df_long['Data'] >= start_date) & (df_long['Data'] <=
end_date)]

# 5. Ajustes personalizados por ramal
ramais = df_filtered['Ramal'].unique()
for ramal in ramais:
    plt.figure(figsize=(12, 6))

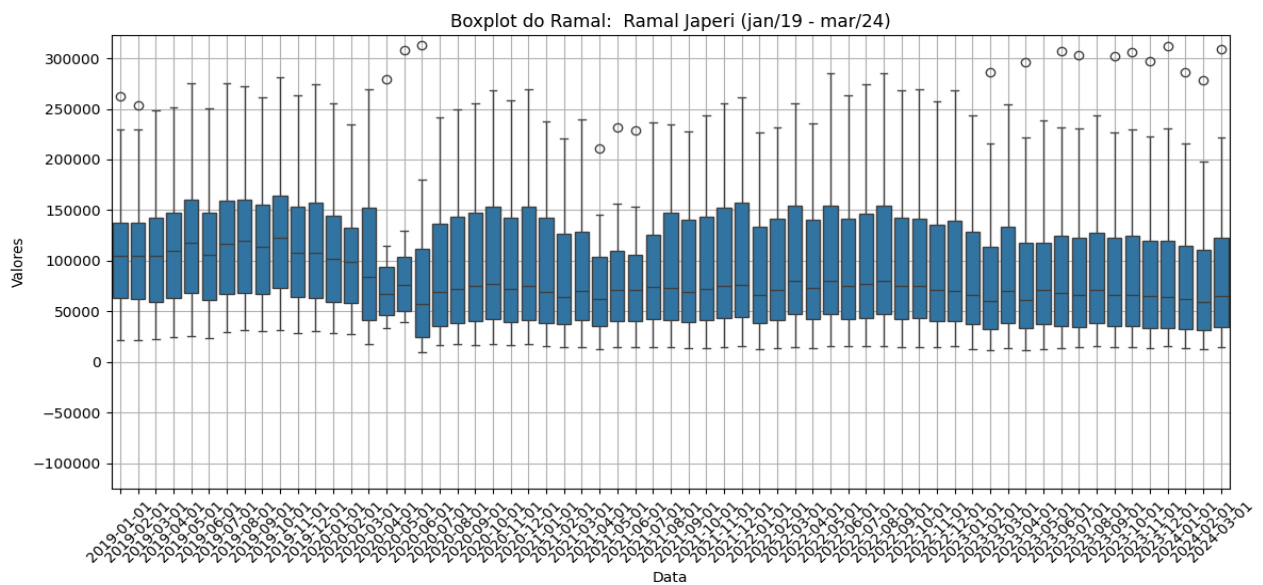
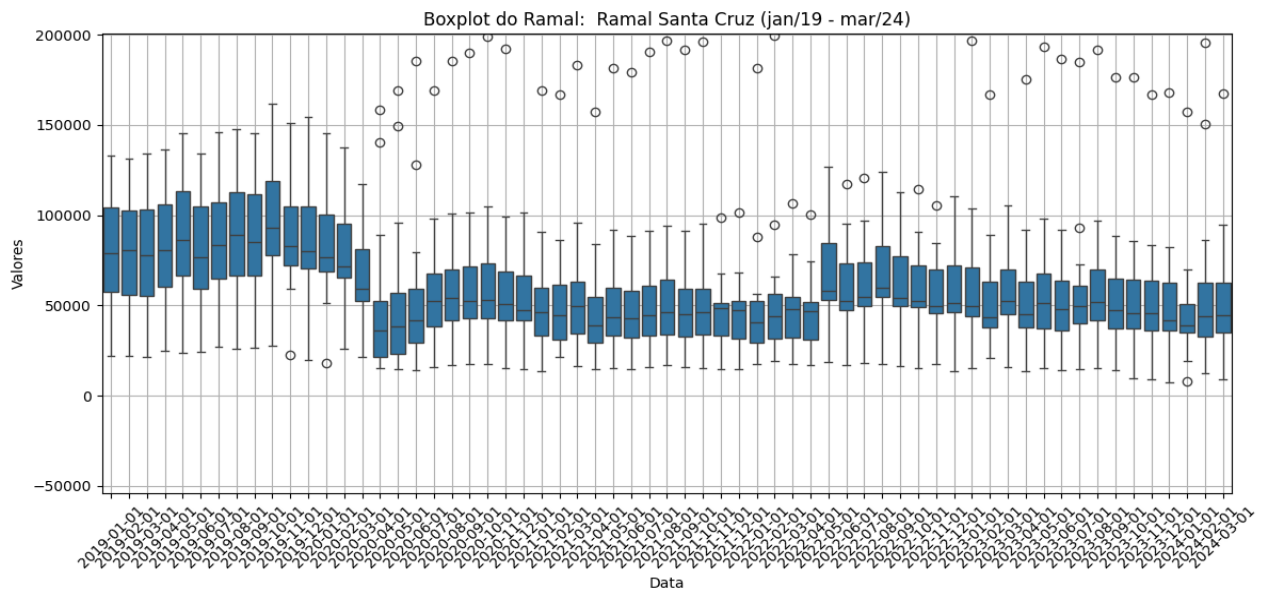
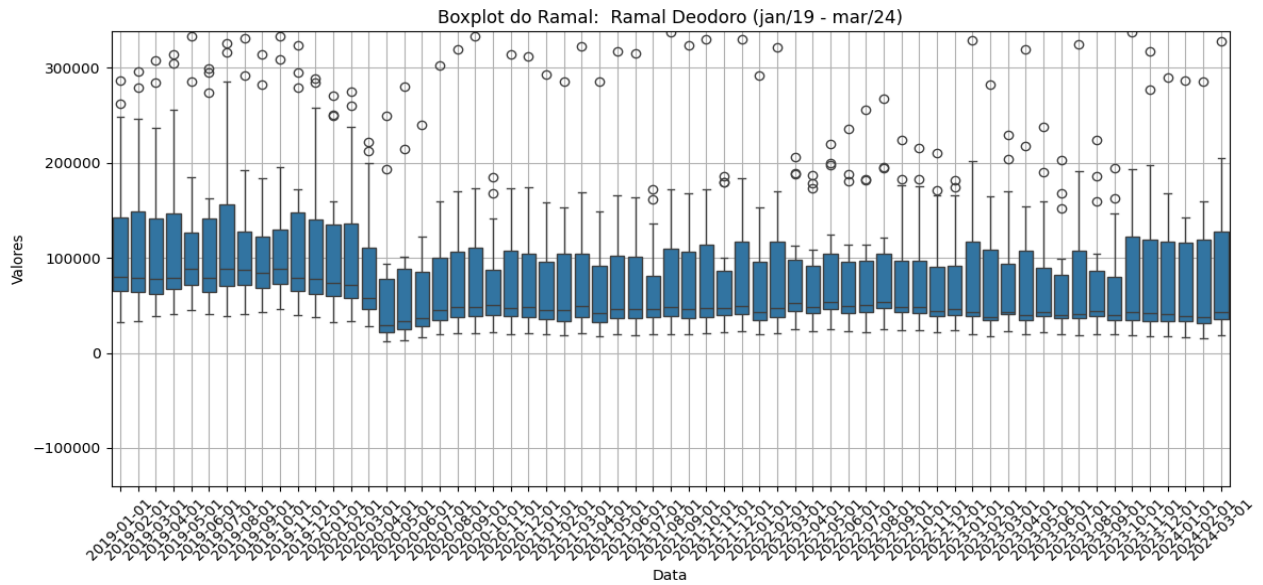
    # Filtrar os dados do ramal específico
    data_ramal = df_filtered[df_filtered['Ramal'] == ramal]

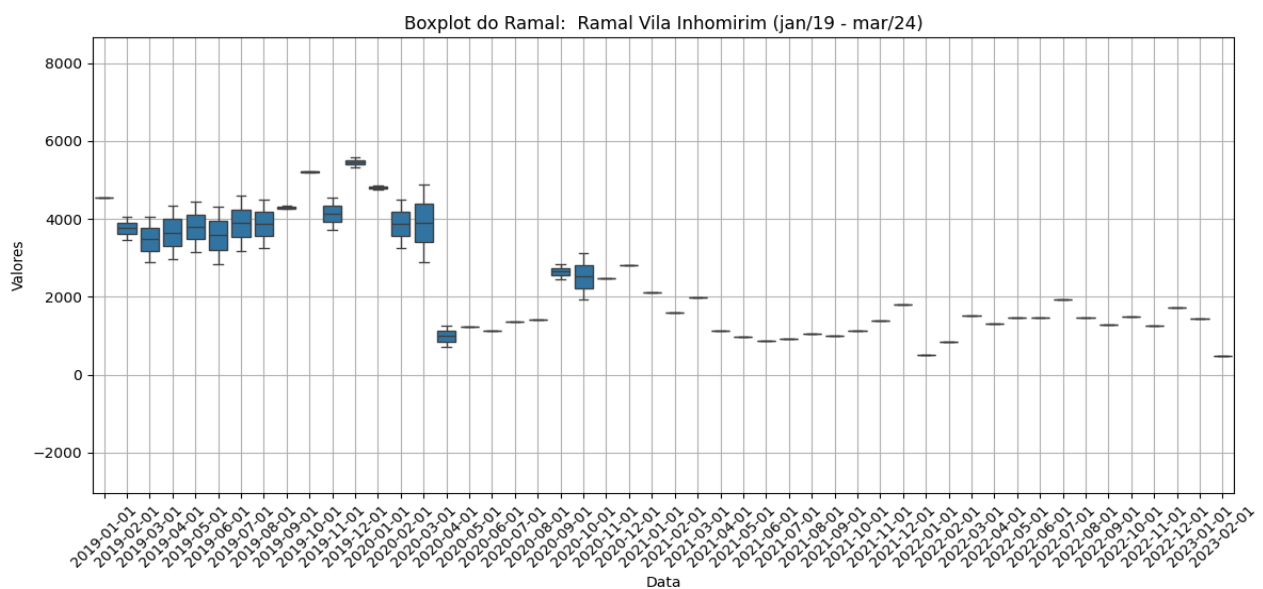
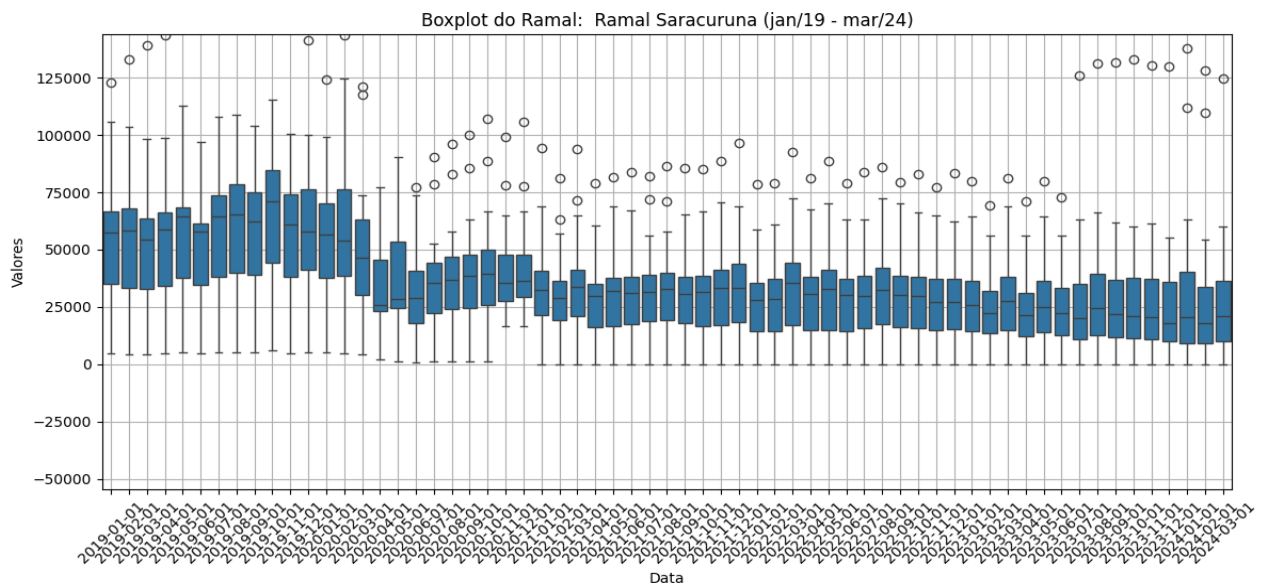
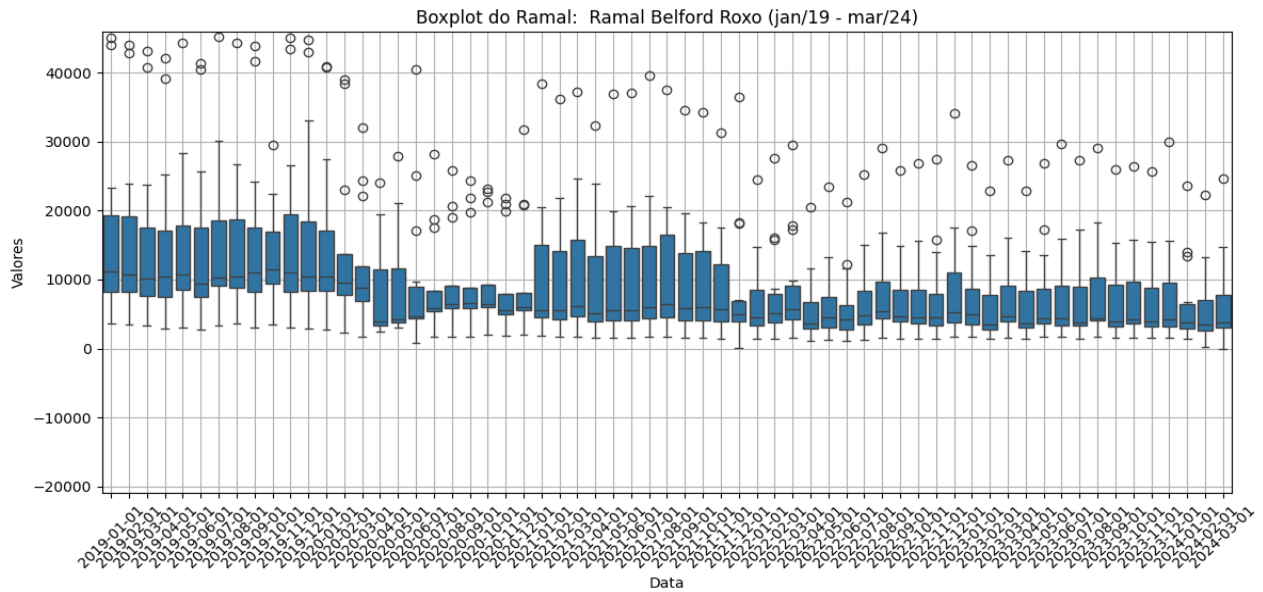
    # Remover outliers se necessário (opcional, ajustável por percentil)
    q1 = data_ramal['Valor'].quantile(0.25)
    q3 = data_ramal['Valor'].quantile(0.75)
    iqr = q3 - q1
    lower_bound = q1 - 1.5 * iqr
    upper_bound = q3 + 1.5 * iqr
    data_no_outliers = data_ramal[(data_ramal['Valor'] >= lower_bound) &
(data_ramal['Valor'] <= upper_bound)]

    # Gráfico com ajuste de outliers e escala apropriada
    sns.boxplot(x='Data', y='Valor', data=data_no_outliers)
    plt.title(f'Boxplot do Ramal: {ramal} (jan/19 - mar/24)')
    plt.xlabel('Data')
    plt.ylabel('Valores')
    plt.xticks(rotation=45)
    plt.grid(True)

    # Ajustar limite do eixo y para melhorar a visualização (opcional)
    plt.ylim(lower_bound, upper_bound)

    plt.tight_layout()
    plt.show()
```





## Código Python para gerar mapa da média por estações ao longo de 2019 à 2024:

### 1. Carregar os dados de passageiros

```
python
Copiar código
file_path = '/content/drive/MyDrive/Python/TrensRJ/Dados_Supervia.xlsx'
df = pd.read_excel(file_path, sheet_name='Supervia2019_2024')
df_long = df.melt(id_vars=['Estação', 'Ramal'], var_name='Data',
value_name='Valor')
df_long['Data'] = pd.to_datetime(df_long['Data'])
```

- **file\_path:** Caminho do arquivo Excel que contém os dados de passageiros.
  - **pd.read\_excel:** Lê os dados da planilha Supervia2019\_2024.
  - **df.melt:** Converte o DataFrame para o formato longo, reorganizando colunas como datas em uma única coluna chamada Data.
  - **pd.to\_datetime:** Converte a coluna Data para o tipo datetime, permitindo manipulações temporais.
- 

### 2. Converter a coluna 'Valor' para numérico

```
python
Copiar código
df_long['Valor'] = pd.to_numeric(df_long['Valor'], errors='coerce')
```

- **pd.to\_numeric:** Converte os valores da coluna Valor para números, substituindo valores inválidos por NaN (errors='coerce').
- 

### 3. Validar e relatar dados descartados

```
python
Copiar código
descartados = df_long[df_long['Valor'].isna()]
if not descartados.empty:
    print("Linhas descartadas devido a valores inválidos ou ausentes:")
    print(descartados[['Estação', 'Ramal', 'Data', 'Valor']])

df_long = df_long.dropna(subset=['Valor'])
```

- Identifica linhas com valores ausentes ou inválidos em Valor.
  - Exibe as linhas descartadas, caso existam.
  - Remove essas linhas do DataFrame para evitar problemas na análise.
- 

### 4. Calcular a média de passageiros por estação

```
python
```

Copiar código

```
media_por_estacao = df_long.groupby('Estação')['Valor'].mean().reset_index()
media_por_estacao.rename(columns={'Valor': 'media'}, inplace=True)
```

- **groupby:** Agrupa os dados por Estação.
  - **mean:** Calcula a média de passageiros para cada estação.
  - **rename:** Renomeia a coluna Valor para media.
- 

## 5. Carregar os dados geográficos

python

Copiar código

```
estacoes =
gpd.read_file('/content/drive/MyDrive/Python/TrensRJ/supervia_stations.geojson')
trens_rj =
gpd.read_file('/content/drive/MyDrive/Python/TrensRJ/Trajetos_Trem.geojson.json')
```

- Carrega dois arquivos GeoJSON:
    - **estacoes:** Contém a localização geográfica das estações.
    - **trens\_rj:** Contém os trajetos das linhas de trem.
- 

## 6. Combinar a média de passageiros com as estações geográficas

python

Copiar código

```
gdf_estacoes = estacoes.merge(media_por_estacao, left_on='estacao',
right_on='Estação', how='left')
```

- **merge:** Combina o DataFrame geográfico das estações com a média de passageiros calculada, baseando-se no nome das estações.
- 

## 7. Ajustar tamanhos proporcionalmente às médias

python

Copiar código

```
min_size, max_size = 50, 500
gdf_estacoes['size'] = gdf_estacoes['media'].fillna(0)
gdf_estacoes['size'] = (
    (gdf_estacoes['size'] - gdf_estacoes['size'].min()) /
    (gdf_estacoes['size'].max() - gdf_estacoes['size'].min())
) * (max_size - min_size) + min_size
```

- **fillna(0):** Substitui valores NaN em media por 0.
- **Escalonamento:** Ajusta os tamanhos dos pontos no mapa proporcionalmente às médias de passageiros, entre os tamanhos mínimo (50) e máximo (500).

## 8. Criar o mapa base e adicionar as camadas

```
python
Copiar código
fig, ax = plt.subplots(figsize=(16, 12))
trens_rj.plot(ax=ax, color='blue', linewidth=1, alpha=0.7, label='Linhas de Trem')
gdf_estacoes.plot(ax=ax, column='media', cmap='viridis', legend=True,
                  markersize=gdf_estacoes['size'], alpha=0.7, edgecolor='k')
```

- Cria um mapa:
    - **trens\_rj.plot**: Plota as linhas de trem em azul.
    - **gdf\_estacoes.plot**: Plota as estações, coloridas por média de passageiros (column='media') e com tamanhos ajustados (markersize).
- 

## 9. Ajustar os limites do mapa

```
python
Copiar código
ax.set_xlim(gdf_estacoes.geometry.x.min() - 0.01, gdf_estacoes.geometry.x.max() + 0.01)
ax.set_ylim(gdf_estacoes.geometry.y.min() - 0.01, gdf_estacoes.geometry.y.max() + 0.01)
```

- Define os limites do mapa com uma margem em torno das coordenadas mínimas e máximas.
- 

## 10. Adicionar um fundo de mapa

```
python
Copiar código
ctx.add_basemap(ax, source=ctx.providers.CartoDB.Positron,
               crs=gdf_estacoes.crs.to_string())
```

- **add\_basemap**: Adiciona um fundo de mapa com design limpo (projeto CartoDB.Positron).
- 

## 11. Finalizar o mapa

```
python
Copiar código
plt.title('Mapa das Estações com Média de Passageiros (jan/19 - mar/24)',
         fontsize=16)
plt.xlabel('Longitude')
plt.ylabel('Latitude')
```

```
plt.legend()
plt.tight_layout()
plt.show()
```



## Desenvolvimento e resultados

A Análise Fatorial por Componentes Principais (PCA) configura-se em técnica exploratória que lida com variáveis MÉTRICAS que possuem, entre si, consideráveis valores de correlação, a fim de se estabelecer nova(s) variável(is) que capture(m) o comportamento conjunto das variáveis originais. Essas variáveis têm o nome de Fator. Logo, temos 2 condições para se usar o PCA:

1. As variáveis devem ser métricas;
2. Elas devem ter um alto grau de correlação entre si.

Pesquisadores da Universidade de Leeds, Inglaterra, realizaram uma meta-análise com 93 estudos sobre o impacto do transporte sobre empregos com título *Does transport help people to gain employment? A systematic review and meta-analysis of the empirical evidence*<sup>3</sup>, encontraram 17 estudos que associaram acesso ao transporte público e probabilidade de empregabilidade.

*A meta-regressão realizada no estudo encontrou forte correlação em possuir um automóvel e empregabilidade, especialmente for a dos EUA. O acesso mais próximo de estações de transporte público resultou em maiores vantagens competitivas para jovens e mulheres.*

A tabela abaixo demonstra as variáveis por bairros da cidade do Rio de Janeiro obtidos no site DATA.RIO: rendas domiciliares por família e per capita obtida (Tabela 875 do IBGE), os estabelecimentos comerciais e empregos gerados pelas 5 principais atividades econômicas que geram mais empregos como comércio, serviços privados e serviços públicos (Tabela 2843 do IBGE):

| Bairro            | Região Administrativa | Renda2024capita | Renda2024     | Massa salarial 2021 | CNPJ2016 | Empregos2016 |
|-------------------|-----------------------|-----------------|---------------|---------------------|----------|--------------|
| Abolição          | Méier                 | R\$ 2.268,18    | R\$ 6.511,61  | R\$ 4.037.911,96    |          |              |
| Acari             | Pavuna                | R\$ 945,06      | R\$ 2.970,25  | R\$ 3.336.519,28    |          |              |
| Alto da Boa Vista | Tijuca                | R\$ 5.044,05    | R\$ 13.288,92 | R\$ 4.509.415,39    |          |              |
| Anchieta          | Anchieta              | R\$ 1.128,52    | R\$ 3.451,24  | R\$ 13.931.474,32   | 188      | 1490         |
| Andaraí           | Vila Isabel           | R\$ 3.878,05    | R\$ 10.432,50 | R\$ 20.091.529,51   |          |              |
| Anil              | Jacarepaguá           | R\$ 2.123,44    | R\$ 6.308,21  | R\$ 7.651.343,18    |          |              |
| Bancários         | Ilha do Governador    | R\$ 2.242,26    | R\$ 6.594,79  | R\$ 613.403,30      |          |              |
| Bangu             | Bangu                 | R\$ 1.118,10    | R\$ 3.473,78  | R\$ 46.688.015,94   | 1894     | 19262        |
| Barra da Tijuca   | Barra da Tijuca       | R\$ 6.284,67    | R\$ 17.600,10 | R\$ 572.263.763,26  |          |              |
| Benfica           | São Cristovão         | R\$ 1.231,19    | R\$ 3.729,85  | R\$ 26.864.349,28   | 601      | 12073        |
| Bento Ribeiro     | Madureira             | R\$ 1.453,91    | R\$ 4.339,74  | R\$ 4.604.070,31    | 241      | 1740         |
| Bonsucesso        | Ramos                 | R\$ 1.495,11    | R\$ 4.450,52  | R\$ 83.360.810,97   | 1432     | 29791        |
| Botafogo          | Botafogo              | R\$ 6.373,24    | R\$ 14.667,09 | R\$ 584.513.879,16  |          |              |
| Bras de Pina      | Penha                 | R\$ 1.361,62    | R\$ 4.245,07  | R\$ 4.045.621,60    | 229      | 1478         |
| Cachambi          | Méier                 | R\$ 2.268,18    | R\$ 6.511,61  | R\$ 6.604.794,35    |          |              |
| Cacuia            | Ilha do Governador    | R\$ 2.242,26    | R\$ 6.594,79  | R\$ 4.886.318,42    |          |              |
| Caju              | Portuária             | R\$ 845,85      | R\$ 2.597,67  | R\$ 28.059.057,13   |          |              |
| Campinho          | Madureira             | R\$ 1.453,91    | R\$ 4.339,74  | R\$ 7.252.678,37    |          |              |

<sup>3</sup> [Does transport help people to gain employment? A systematic review and meta-analysis of the empirical evidence - White Rose Research Online](#)

|                                |                    |              |               |                      |       |        |
|--------------------------------|--------------------|--------------|---------------|----------------------|-------|--------|
| Campo Grande                   | Campo Grande       | R\$ 1.252,90 | R\$ 3.948,69  | R\$ 157.108.887,56   | 4037  | 44483  |
| Cascadura                      | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 17.605.484,46    | 593   | 5502   |
| Catete                         | Botafogo           | R\$ 6.373,24 | R\$ 14.667,09 | R\$ 18.429.618,18    |       |        |
| Catumbi                        | Rio Comprido       | R\$ 1.972,81 | R\$ 5.895,91  | R\$ 1.808.488,51     |       |        |
| Centro                         | Centro             | R\$ 2.572,35 | R\$ 5.340,28  | R\$ 3.007.140.816,61 | 14492 | 391986 |
| Cidade de Deus                 | Cidade de Deus     | R\$ 855,52   | R\$ 2.733,85  | R\$ 1.200.326,67     |       |        |
| Cidade Universitaria           | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 176.678.742,37   |       |        |
| Coelho Neto                    | Pavuna             | R\$ 945,06   | R\$ 2.970,25  | R\$ 19.917.876,48    | 573   | 9869   |
| Colegio                        | Irajá              | R\$ 1.750,23 | R\$ 5.133,82  | R\$ 3.903.979,24     | 97    | 406    |
| Copacabana                     | Copacabana         | R\$ 6.046,06 | R\$ 13.187,18 | R\$ 117.559.121,32   |       |        |
| Cordovil                       | Vigário Geral      | R\$ 1.051,04 | R\$ 3.354,07  | R\$ 14.373.177,70    |       |        |
| Cosmos                         | Campo Grande       | R\$ 1.252,90 | R\$ 3.948,69  | R\$ 3.903.979,24     | 162   | 1221   |
| Curicica                       | Jacarepaguá        | R\$ 2.123,44 | R\$ 6.308,21  | R\$ 55.333.839,80    |       |        |
| Del Castilho                   | Inhaúma            | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 17.781.615,92    | 324   | 5747   |
| Deodoro                        | Realengo           | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 3.045.935,21     | 43    | 301    |
| Encantado                      | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 1.812.120,62     |       |        |
| Engenho da Rainha              | Inhaúma            | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 2.446.915,38     |       |        |
| Engenho de Dentro              | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 22.740.511,81    | 473   | 5370   |
| Engenho Novo                   | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 10.014.256,24    |       |        |
| Estacio                        | Rio Comprido       | R\$ 1.972,81 | R\$ 5.895,91  | R\$ 15.726.952,90    |       |        |
| Flamengo                       | Botafogo           | R\$ 6.373,24 | R\$ 14.667,09 | R\$ 66.166.437,47    |       |        |
| Freguesia - Jacarepagua        | Jacarepaguá        | R\$ 2.123,44 | R\$ 6.308,21  | R\$ 47.642.071,44    |       |        |
| Freguesia - Ilha do Governador | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 10.450.307,38    |       |        |
| Galeao                         | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 34.121.227,02    |       |        |
| Gavea                          | Lagoa              | R\$ 9.513,25 | R\$ 22.906,12 | R\$ 73.562.690,49    |       |        |
| Gloria                         | Centro             | R\$ 2.572,35 | R\$ 5.340,28  | R\$ 73.546.244,72    |       |        |
| Grajau                         | Vila Isabel        | R\$ 3.878,05 | R\$ 10.432,50 | R\$ 8.622.823,14     |       |        |
| Guadalupe                      | Anchieta           | R\$ 1.128,52 | R\$ 3.451,24  | R\$ 6.188.920,74     |       |        |
| Guaratiba                      | Guaratiba          | R\$ 923,06   | R\$ 3.006,19  | R\$ 20.530.561,13    |       |        |
| Higienopolis                   | Inhaúma            | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 6.795.499,14     |       |        |
| Honorio Gurgel                 | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 3.731.484,56     | 41    | 320    |
| Humaita                        | Botafogo           | R\$ 6.373,24 | R\$ 14.667,09 | R\$ 22.054.424,99    | 394   | 4499   |
| Inhauma                        | Inhaúma            | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 76.797.292,10    |       |        |
| Inhoaiba                       | Campo Grande       | R\$ 1.252,90 | R\$ 3.948,69  | R\$ 3.312.415,75     | 130   | 976    |
| Ipanema                        | Lagoa              | R\$ 9.513,25 | R\$ 22.906,12 | R\$ 83.199.203,12    |       |        |
| Iraja                          | Irajá              | R\$ 1.750,23 | R\$ 5.133,82  | R\$ 44.824.574,12    |       |        |
| Itanhangá                      | Barra da Tijuca    | R\$ 6.284,67 | R\$ 17.600,10 | R\$ 4.487.254,96     |       |        |
| Jacare                         | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 13.085.017,57    |       |        |
| Jacarepagua                    | Jacarepaguá        | R\$ 2.123,44 | R\$ 6.308,21  | R\$ 340.619.619,68   |       |        |
| Jardim Botanico                | Lagoa              | R\$ 9.513,25 | R\$ 22.906,12 | R\$ 74.746.115,00    |       |        |
| Jardim Carioca                 | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 6.059.900,14     |       |        |

|                          |                    |              |               |                    |      |       |
|--------------------------|--------------------|--------------|---------------|--------------------|------|-------|
| Jardim Guanabara         | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 14.300.706,36  |      |       |
| Jardim Sulacap           | Realengo           | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 4.925.312,25   |      |       |
| Lagoa                    | Lagoa              | R\$ 9.513,25 | R\$ 22.906,12 | R\$ 342.649.501,68 |      |       |
| Laranjeiras              | Botafogo           | R\$ 6.373,24 | R\$ 14.667,09 | R\$ 57.113.319,84  |      |       |
| Leblon                   | Lagoa              | R\$ 9.513,25 | R\$ 22.906,12 | R\$ 90.080.577,07  |      |       |
| Lins de Vasconcelos      | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 29.504.363,39  |      |       |
| Madureira                | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 168.366.863,65 | 1956 | 18893 |
| Magalhaes Bastos         | Realengo           | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 1.253.161,89   | 72   | 449   |
| Mangueira                | São Cristovão      | R\$ 1.231,19 | R\$ 3.729,85  | R\$ 1.037.055,00   | 28   | 348   |
| Manguinhos               | Ramos              | R\$ 1.495,11 | R\$ 4.450,52  | R\$ 96.497.067,14  | 289  | 11000 |
| Maracana                 | Vila Isabel        | R\$ 3.878,05 | R\$ 10.432,50 | R\$ 223.253.515,18 | 726  | 8764  |
| Marechal Hermes          | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 39.293.354,10  | 253  | 2053  |
| Maria da Graça           | Inhaúma            | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 4.454.073,50   |      |       |
| Meier                    | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 196.668.148,59 | 1981 | 13762 |
| Monero                   | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 2.601.730,85   |      |       |
| Olaria                   | Ramos              | R\$ 1.495,11 | R\$ 4.450,52  | R\$ 21.608.035,50  | 581  | 6467  |
| Oswaldo Cruz             | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 2.385.257,42   | 153  | 1726  |
| Paciencia                | Santa Cruz         | R\$ 853,72   | R\$ 2.783,35  | R\$ 6.594.019,13   | 279  | 2024  |
| Padre Miguel             | Bangu              | R\$ 1.118,10 | R\$ 3.473,78  | R\$ 5.850.338,20   | 315  | 2885  |
| Parada de Lucas          | Vigário Geral      | R\$ 1.051,04 | R\$ 3.354,07  | R\$ 9.073.114,76   | 138  | 1564  |
| Pavuna                   | Pavuna             | R\$ 945,06   | R\$ 2.970,25  | R\$ 70.253.138,03  |      |       |
| Pechincha                | Jacarepaguá        | R\$ 2.123,44 | R\$ 6.308,21  | R\$ 29.176.979,46  |      |       |
| Pedra de Guaratiba       | Guaratiba          | R\$ 923,06   | R\$ 3.006,19  | R\$ 1.981.367,71   |      |       |
| Penha                    | Penha              | R\$ 1.361,62 | R\$ 4.245,07  | R\$ 117.099.548,96 |      |       |
| Penha Circular           | Penha              | R\$ 1.361,62 | R\$ 4.245,07  | R\$ 36.352.589,44  | 773  | 14516 |
| Piedade                  | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 7.081.916,99   | 322  | 3961  |
| Pilares                  | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 22.846.154,96  |      |       |
| Portuguesa               | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 6.928.501,75   |      |       |
| Quintino Bocaiuva        | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 39.303.585,49  | 223  | 8595  |
| Ramos                    | Ramos              | R\$ 1.495,11 | R\$ 4.450,52  | R\$ 235.306.533,61 | 801  | 11621 |
| Realengo                 | Realengo           | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 36.292.856,70  | 826  | 9915  |
| Recreio dos Bandeirantes | Barra da Tijuca    | R\$ 6.284,67 | R\$ 17.600,10 | R\$ 30.273.664,96  |      |       |
| Riachuelo                | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 6.066.144,14   | 129  | 1220  |
| Ribeira                  | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 7.610.023,76   |      |       |
| Rio Comprido             | Rio Comprido       | R\$ 1.972,81 | R\$ 5.895,91  | R\$ 635.678.983,42 |      |       |
| Rocha Miranda            | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 5.319.240,77   | 333  | 2543  |
| Santa Cruz               | Santa Cruz         | R\$ 853,72   | R\$ 2.783,35  | R\$ 127.351.397,03 | 1013 | 15177 |
| Santissimo               | Campo Grande       | R\$ 1.252,90 | R\$ 3.948,69  | R\$ 2.546.880,02   | 127  | 811   |

|                     |                    |              |               |                    |     |      |
|---------------------|--------------------|--------------|---------------|--------------------|-----|------|
| Santo Cristo        | Portuária          | R\$ 845,85   | R\$ 2.597,67  | R\$ 265.376.117,02 |     |      |
| Sao Conrado         | Lagoa              | R\$ 9.513,25 | R\$ 22.906,12 | R\$ 8.292.270,68   |     |      |
| Sao Cristovao       | São Cristóvão      | R\$ 1.231,19 | R\$ 3.729,85  | R\$ 267.646.222,43 | 127 | 811  |
| Senador Camara      | Bangu              | R\$ 1.118,10 | R\$ 3.473,78  | R\$ 5.892.211,23   | 194 | 2344 |
| Senador Vasconcelos | Campo Grande       | R\$ 1.252,90 | R\$ 3.948,69  | R\$ 2.482.454,42   | 137 | 1058 |
| Sepetiba            | Santa Cruz         | R\$ 853,72   | R\$ 2.783,35  | R\$ 2.558.707,85   |     |      |
| Tanque              | Jacarepaguá        | R\$ 2.123,44 | R\$ 6.308,21  | R\$ 8.701.230,23   |     |      |
| Taquara             | Jacarepaguá        | R\$ 2.123,44 | R\$ 6.308,21  | R\$ 54.427.063,37  |     |      |
| Taua                | Ilha do Governador | R\$ 2.242,26 | R\$ 6.594,79  | R\$ 2.634.399,71   |     |      |
| Tijuca              | Tijuca             | R\$ 5.044,05 | R\$ 13.288,92 | R\$ 123.605.395,46 |     |      |
| Todos os Santos     | Méier              | R\$ 2.268,18 | R\$ 6.511,61  | R\$ 3.931.092,93   |     |      |
| Tomas Coelho        | Inhaúma            | R\$ 1.419,47 | R\$ 4.235,99  | R\$ 1.116.655,03   |     |      |
| Turiacu             | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 9.455.921,27   |     |      |
| Urca                | Botafogo           | R\$ 6.373,24 | R\$ 14.667,09 | R\$ 1.873.923,01   |     |      |
| Vargem Grande       | Barra da Tijuca    | R\$ 6.284,67 | R\$ 17.600,10 | R\$ 5.713.477,01   |     |      |
| Vaz Lobo            | Madureira          | R\$ 1.453,91 | R\$ 4.339,74  | R\$ 3.134.191,15   |     |      |
| Vicente de Carvalho | Irajá              | R\$ 1.750,23 | R\$ 5.133,82  | R\$ 4.916.095,29   | 240 | 2985 |
| Vigario Geral       | Vigário Geral      | R\$ 1.051,04 | R\$ 3.354,07  | R\$ 50.336.087,76  | 230 | 6952 |
| Vila da Penha       | Irajá              | R\$ 1.750,23 | R\$ 5.133,82  | R\$ 14.472.436,97  |     |      |
| Vila Isabel         | Vila Isabel        | R\$ 3.878,05 | R\$ 10.432,50 | R\$ 271.909.105,27 |     |      |
| Vila Valqueire      | Jacarepaguá        | R\$ 2.123,44 | R\$ 6.308,21  | R\$ 15.168.040,28  |     |      |
| Vista Alegre        | Irajá              | R\$ 1.750,23 | R\$ 5.133,82  | R\$ 3.757.322,10   |     |      |
| Praca da Bandeira   | Tijuca             | R\$ 5.044,05 | R\$ 13.288,92 | R\$ 17.094.193,35  |     |      |
| Praca Seca          | Jacarepaguá        | R\$ 2.123,44 | R\$ 6.308,21  | R\$ 5.655.016,03   |     |      |

A Análise PCA é um método matemático usado para simplificar conjuntos de dados complexos. Ele encontra padrões identificando direções (chamadas de componentes principais) que capturam a maior variação nos dados.

Imagine que você tem uma tabela onde as linhas são genes e as colunas são amostras (uma típica matriz de dados RNA-seq). O PCA compacta esses dados de alta dimensão em menos variáveis Componentes Principais (PCs) para que possamos visualizar e compreender melhor os relacionamentos.

O resultado? Um gráfico PCA. Cada ponto representa uma amostra, e quanto mais próximos dois pontos estiverem, mais semelhantes serão seus perfis de expressão. Mas aqui está o problema: o gráfico PCA pode mudar com base em como você pré-processa seus dados.

## Código Python para análise fatorial por Componentes Principais (PCA):

```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

# Caminho do arquivo
file_path = "/content/drive/MyDrive/Python/TrensRJ/Supervia_PCA.xlsx"

# Carregar as planilhas
df_posto_empregos = pd.read_excel(file_path,
sheet_name='Posto_Empregos2021')
df_estat_supervia = pd.read_excel(file_path, sheet_name='EstatSupervia')

# Selecionar as colunas relevantes
df_posto_empregos_selected = df_posto_empregos[['Renda2024', 'CNPJ2016',
'Soma de Empregos2016']]
df_estat_supervia_selected = df_estat_supervia.filter(like='Média Mensal')

# Concatenar os dados para a análise
data_combined = pd.concat([df_posto_empregos_selected,
df_estat_supervia_selected], axis=1)

# Tratar valores ausentes
data_combined = data_combined.dropna()

# Padronizar os dados
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data_combined)

# Aplicar PCA
pca = PCA()
pca_results = pca.fit_transform(data_scaled)

# Criar DataFrame com os resultados
pca_df = pd.DataFrame(
    data=pca_results,
    columns=[f'PC{i+1}' for i in range(pca_results.shape[1])]
)

# Explicação da variância
explained_variance = pca.explained_variance_ratio_

# Exibir resultados
print("Variância explicada por componente principal:")
for i, var in enumerate(explained_variance, start=1):
    print(f"PC{i}: {var:.2%}")
```

```
print("\nCargas dos Componentes Principais:")
components_df = pd.DataFrame(
    pca.components_,
    columns=data_combined.columns,
    index=[f'PC{i+1}' for i in range(pca.components_.shape[0])]
)
print(components_df)

# Salvar os resultados em um arquivo Excel
output_file = "/content/drive/MyDrive/Python/TrensRJ/PCA_Results.xlsx"
with pd.ExcelWriter(output_file) as writer:
    pca_df.to_excel(writer, sheet_name='PCA Results', index=False)
    components_df.to_excel(writer, sheet_name='PCA Loadings')

print(f"Resultados salvos em: {output_file}")
```

|                                               |
|-----------------------------------------------|
| Variância explicada por componente principal: |
|-----------------------------------------------|

|             |
|-------------|
| PC1: 66.12% |
|-------------|

|             |
|-------------|
| PC2: 33.17% |
|-------------|

|            |
|------------|
| PC3: 0.71% |
|------------|

|                                    |  |  |
|------------------------------------|--|--|
| Cargas dos Componentes Principais: |  |  |
|------------------------------------|--|--|

|  | Renda2024 | CNPJ2016 | Soma de Empregos2016 |
|--|-----------|----------|----------------------|
|--|-----------|----------|----------------------|

|     |          |          |          |
|-----|----------|----------|----------|
| PC1 | 0.070718 | 0.705437 | 0.705236 |
|-----|----------|----------|----------|

|     |          |           |           |
|-----|----------|-----------|-----------|
| PC2 | 0.997488 | -0.047148 | -0.052863 |
|-----|----------|-----------|-----------|

|     |           |          |           |
|-----|-----------|----------|-----------|
| PC3 | -0.004041 | 0.707203 | -0.706999 |
|-----|-----------|----------|-----------|

## 1. Carregamento dos Dados

- Os dados são carregados de duas planilhas:
  - **Posto\_Empregos2021**: Contém informações como renda, número de CNPJs e empregos.
  - **EstatSupervia**: Contém médias mensais relacionadas à Supervia.
- Dados relevantes são selecionados, resultando em variáveis métricas (quantitativas e contínuas).

## 2. Concatenação e Tratamento

- As colunas selecionadas de ambas as tabelas são combinadas, formando um único DataFrame.
- Linhas com valores ausentes (NaN) são removidas para evitar erros no cálculo.

### 3. Padronização dos Dados

- **StandardScaler**: Os dados são escalonados para média 0 e desvio padrão 1.
- Padronizar é essencial na PCA porque variáveis com magnitudes diferentes poderiam influenciar mais os resultados.

### 4. Aplicação da PCA

- A PCA é aplicada ao conjunto de dados padronizados.
- Os resultados incluem:
  - **Componentes principais (PCs)**: Novas variáveis ortogonais que explicam a variância nos dados.
  - **Variância explicada**: Percentual de variância capturado por cada componente.
  - **Cargas dos componentes**: Contribuição de cada variável original em cada componente principal.

### 5. Exportação dos Resultados

- Os resultados são salvos em um arquivo Excel, contendo:
  - **Resultados da PCA** (valores de cada observação para os PCs).
  - **Cargas dos componentes principais** (relação entre variáveis originais e PCs).

---

## Resultados e Interpretação

### 1. Variância Explicada por Componente Principal

- O código imprime a proporção da variância explicada por cada componente (ex.: PC1, PC2, etc.).
- Se o **PC1** explicar a maior parte da variância (ex.: > 50%), ele resume bem os dados originais.

- A soma acumulada da variância explicada indica quantos componentes são necessários para capturar uma parcela significativa (ex.: 90%) da variância total.

### Exemplo de interpretação:

yaml

Copiar código

Variância explicada por componente principal:

PC1: 62.5%

PC2: 18.3%

PC3: 10.1%

...

- **PC1** captura 62,5% da variância, mostrando que ele sintetiza grande parte da informação.
- **PC2** e **PC3** explicam variâncias menores, mas ainda podem ser relevantes.

---

## 2. Cargas dos Componentes Principais

- As cargas indicam a contribuição de cada variável original em cada componente.
- Valores altos (positivos ou negativos) mostram variáveis mais importantes para o respectivo PC.

### Exemplo de saída das cargas:

markdown

Copiar código

Cargas dos Componentes Principais:

|     | Renda2024 | CNPJ2016 | Soma de Empregos2016 | Média Mensal 1 | ... |
|-----|-----------|----------|----------------------|----------------|-----|
| PC1 | 0.65      | 0.72     | 0.18                 | 0.10           | ... |
| PC2 | -0.45     | 0.30     | -0.78                | 0.12           | ... |
| PC3 | 0.12      | -0.10    | 0.10                 | 0.95           | ... |

### Interpretação:

- Para o **PC1**, variáveis como Renda2024 e CNPJ2016 têm altas cargas positivas, indicando forte contribuição.
  - Para o **PC2**, a variável Soma de Empregos2016 contribui negativamente, enquanto outras têm menor influência.
  - Para o **PC3**, Média Mensal 1 domina, sugerindo que captura aspectos independentes.
- 

### 3. Correlação entre as Variáveis

- A PCA identifica correlações entre variáveis:
  - Variáveis que têm cargas semelhantes (e sinal igual) em um componente estão **positivamente correlacionadas**.
  - Variáveis com cargas de sinais opostos estão **negativamente correlacionadas**.
  - Componentes principais são ortogonais (não correlacionados entre si).

#### Exemplo de conclusão:

- **Renda2024** e **CNPJ2016** têm cargas altas e similares no **PC1**, indicando forte correlação positiva.
  - **Soma de Empregos2016** apresenta relação negativa no **PC2**, sugerindo que captura variabilidade distinta.
- 

## Conclusão

### 1. São variáveis métricas?

- Sim, as variáveis escolhidas (Renda2024, CNPJ2016, etc.) são quantitativas e contínuas, adequadas para PCA.

### 2. Correlação entre variáveis?

- Sim, a análise PCA indica correlação entre variáveis (representada pelas cargas nos componentes principais).

- Variáveis com cargas altas e semelhantes em um componente principal estão correlacionadas.

### 3. Redução de Dimensionalidade:

- A PCA reduz a dimensionalidade, mantendo a maior parte da variância. Isso permite simplificar a análise sem perder muita informação.

**Código Python para plotar um mapa de bairros com empregos que geram passageiros:**

```
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
import contextily as ctx

# 1. Carregar os dados da planilha POSTOS_EMPREGOS2021
file_path = "/content/drive/MyDrive/Python/TrensRJ/Supervia_PCA.xlsx"
df_postos = pd.read_excel(file_path, sheet_name='Posto_Empregos2021')

# Filtrar dados da coluna 'Soma de Empregos2016'
df_postos = df_postos.dropna(subset=['Soma de Empregos2016']) # Remover
linhas com valores ausentes
df_postos['Soma de Empregos2016'] = pd.to_numeric(df_postos['Soma de
Empregos2016'], errors='coerce') # Converter para numérico
df_postos = df_postos.dropna(subset=['Soma de Empregos2016']) # Remover
valores inválidos

# 2. Carregar os dados geográficos dos bairros
bairros =
gpd.read_file('/content/drive/MyDrive/Python/Limite_de_Bairros.geojson')

# Verificar colunas disponíveis
print("Colunas em bairros:", bairros.columns)
print("Colunas em df_postos:", df_postos.columns)

# 3. Combinar os dados de empregos com os dados geográficos
gdf_bairros = bairros.merge(df_postos, left_on='nome', right_on='Bairro',
how='inner')

# 4. Criar o mapa base
fig, ax = plt.subplots(figsize=(16, 12))

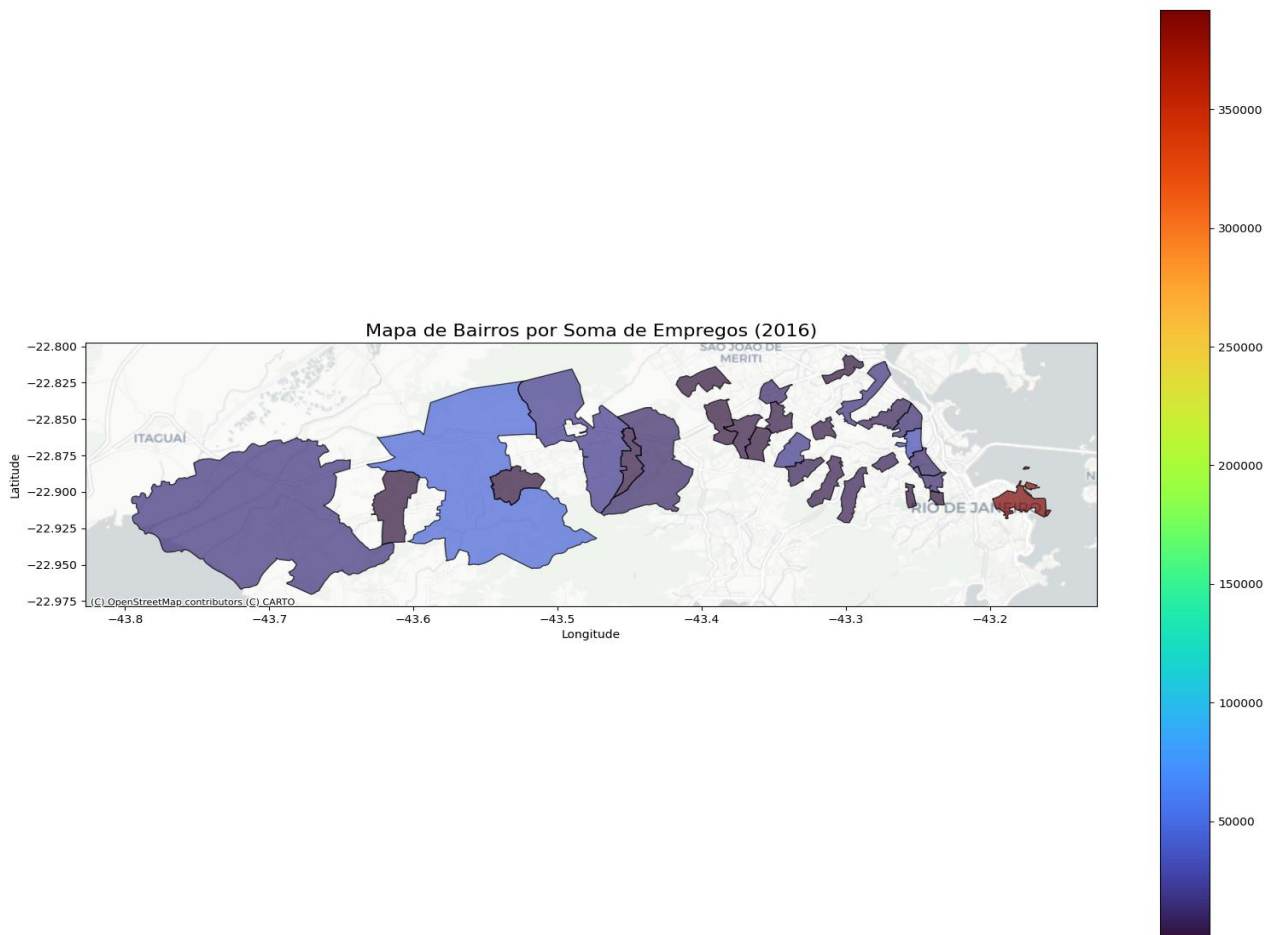
# Plotar os bairros com escala de cores baseada na Soma de Empregos2016
```

```
gdf_bairros.plot(ax=ax, column='Soma de Empregos2016', cmap='turbo',
legend=True, edgecolor='black', alpha=0.7)

# Adicionar um fundo de mapa com Contextily
ctx.add_basemap(ax, source=ctx.providers.CartoDB.Positron,
crs=bairros.crs.to_string())

# Adicionar título e legendas
plt.title('Mapa de Bairros por Soma de Empregos (2016)', fontsize=16)
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.tight_layout()

# Exibir o mapa
plt.show()
```



## **Conclusões:**

É notável observar que os efeitos das mudanças do mercado de trabalho após a Pandemia de Covid-19 afetaram a utilização dos transportes coletivos nas grandes cidades do mundo, incluindo a região metropolitana do Rio de Janeiro. A queda da frequência de passageiros no Sistema ferroviário foi de 50% em relação aos níveis de utilização observados em 2016;

Os estudos acadêmicos indicam uma grande necessidade de melhor integração com os demais modais de transporte coletivos como ônibus e metrô;

Ao longo das décadas, as intervenções com dinheiro público dos gestores urbanos trouxeram maior conforto e segurança operacional nos trens urbanos da cidade do Rio de Janeiro, mas contribuíram para o déficit crescente nas contas do estado do Rio de Janeiro;

As médias observadas nos Ramais Belford Roxo e Vila Inhomirim indicam a necessidade de substituição do Modal de transporte para outro mais econômico e com um modelo de negócios atraente para a iniciativa privada;

A nossa proposta sobre a substituição dos trens e estações do Ramal Belford Roxo busca uma melhor segurança operacional, aumentar o número de passageiros atendidos, auxiliar na recuperação judicial da Concessionária de trens e uma alternativa viária para a ligação do Centro da cidade com regiões empobrecidas para promover sua revitalização;

Faz-se necessário um trabalho conjunto entre governo estadual, AGETRANSP e ANTT para avaliar alternativas e soluções de integração em diferentes modais de transporte na Região metropolitana do Rio de Janeiro.

Outra proposta que poderia ser estudada seria a duplicação da passagem férrea sobre a Avenida Francisco Bicalho para possibilitar maior fluxo de trens com segurança e menor tempo de viagens.