



Análise fluxo de passageiros Metrô

A análise de dados é um processo abrangente que envolve várias fases, desde a extração inicial dos dados até a análise e modelagem final. Utilizei o chat GPT 4.0 para simular 10 perguntas relacionadas ao meu sistema para eu responder.

Extração de dados

Essa fase envolve coletar os dados necessários, juntá-los para assim iniciar a preparação.

Demandas | Portal da Transparência
<https://transparencia.metrosp.com.br/dataset/demandas>

Limpeza e preparação de dados

Uma vez extraídos, os dados precisam ser limpos e preparados para análise. Isso pode incluir remoção de dados duplicados ou irrelevantes, tratamentos de valores ausentes, correção de erros, conversão de dados para o formato correto. Isso garante a qualidade dos dados.

Primeira tentativa de organizar os meses em ordem crescente

Abaixo criei uma nova coluna diretamente do PowerBi que relaciona o Mês com o número do mês, e utilizei a função switch para concatenar os valores.

The screenshot shows the Power BI Data Editor interface. At the top, there is a code editor window containing DAX code. Below it is a table view showing data from a query named 'tp_dados.metro_concatenados'. The table has columns: Estação, Ano, Mês, Qtd, and n_mes. The n_mes column contains numerical values corresponding to the month names in the Mês column. A tooltip or preview area at the bottom shows the first few rows of the DAX code.

Estação	Ano	Mês	Qtd	n_mes
Corinthians-Itaquera	2021	Abr	45	4
Corinthians-Itaquera	2021	Mai	52	5
Corinthians-Itaquera	2021	Jun	53	6
Corinthians-Itaquera	2021	Jul	53	7
Corinthians-Itaquera	2021	Ago	58	8
Corinthians-Itaquera	2021	Set	60	9
Corinthians-Itaquera	2021	Out	62	10
Corinthians-Itaquera	2021	Nov	66	11
Corinthians-Itaquera	2021	Dez	67	12
Artur Alvim	2021	Jan	37	1

```
n_mes =  
SWITCH(  
    tp_dados.metro_concatenados[Mês],  
    "Jan", 1,  
    "Fev", 2,  
    "Mar", 3,  
    "Abr", 4,  
    "Mai", 5,  
    "Jun", 6,  
    "Jul", 7,
```

```

    "Ago", 8,
    "Set", 9,
    "Out", 10,
    "Nov", 11,
    "Dez", 12,
    BLANK() -- Lidar com valores nulos ou desconhecidos, se
)

```

Porém continuou com esse visual:

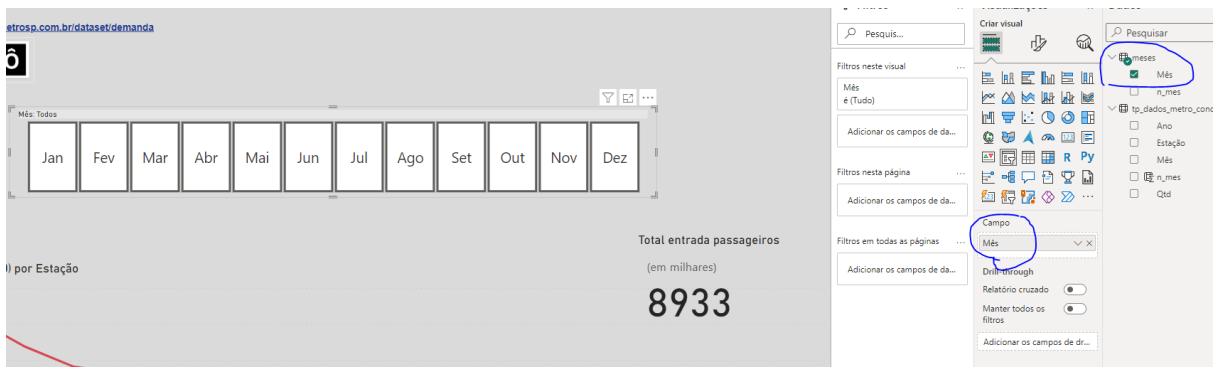
Mês: Seleções múltiplas			
Abr	Fev	Jun	Nov
Ago	Jan	Mai	Out
Dez	Jul	Mar	Set

Abaixo o relacionamento:



Correção do ordenamento

Eu estava com dificuldade para estabelecer que o mês Jan = 1, Fev = 2, etc. Então criei uma outra tabela chamado meses e criei o relacionamento entre meses, na segmentação de dados, eu utilizei esta nova tabela meses para organizar.



Abaixo uma amostra da tabela principal que contém as informações das 3 linhas de metrô.

Estação	Ano	Linha	Mês	Qtd
Jabaquara	2021	L1 - Azul	Jan	46
Jabaquara	2021	L1 - Azul	Fev	48
Jabaquara	2021	L1 - Azul	Mar	40
Jabaquara	2021	L1 - Azul	Abr	39
Jabaquara	2021	L1 - Azul	Mai	45
Jabaquara	2021	L1 - Azul	Jun	46
Jabaquara	2021	L1 - Azul	Jul	48
Jabaquara	2021	L1 - Azul	Ago	51
Jabaquara	2021	L1 - Azul	Set	55
Jabaquara	2021	L1 - Azul	Out	57
Jabaquara	2021	L1 - Azul	Nov	61
Jabaquara	2021	L1 - Azul	Dez	59
Conceição	2021	L1 - Azul	Jan	15
Conceição	2021	L1 - Azul	Fev	16
Conceição	2021	L1 - Azul	Mar	12
Conceição	2021	L1 - Azul	Abr	12
Conceição	2021	L1 - Azul	Mai	14
Conceição	2021	L1 - Azul	Jun	15
Conceição	2021	L1 - Azul	Jul	15
Conceição	2021	L1 - Azul	Ago	16
Conceição	2021	L1 - Azul	Set	18
Conceição	2021	L1 - Azul	Out	19
Conceição	2021	L1 - Azul	Nov	20
Conceição	2021	L1 - Azul	Dez	19
São Judas	2021	L1 - Azul	Jan	9
São Judas	2021	L1 - Azul	Fev	9
São Judas	2021	L1 - Azul	Mar	7

Abaixo fiz o relacionamento atribuindo o mês com o seu número ordenado.

Mês	n_mes
Jan	1
Fev	2
Mar	3
Abr	4
Mai	5
Jun	6
Jul	7
Ago	8
Set	9
Out	10
Nov	11
Dez	12

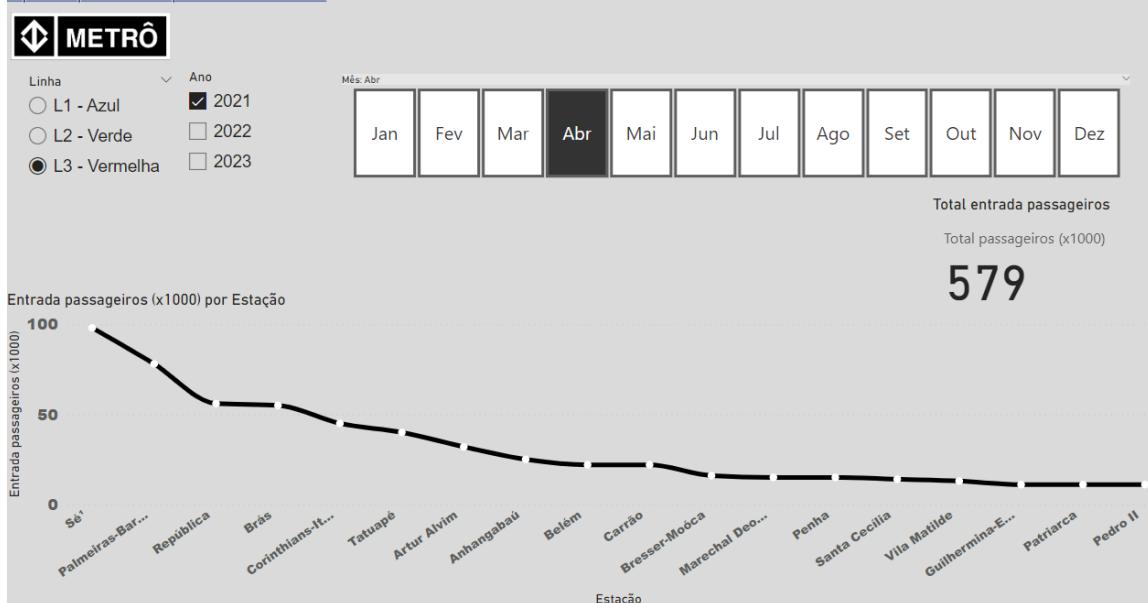
Exploração de Dados (EDA - Exploratory Data Analysis)

Ferramentas como gráficos, são usadas nesta fase.

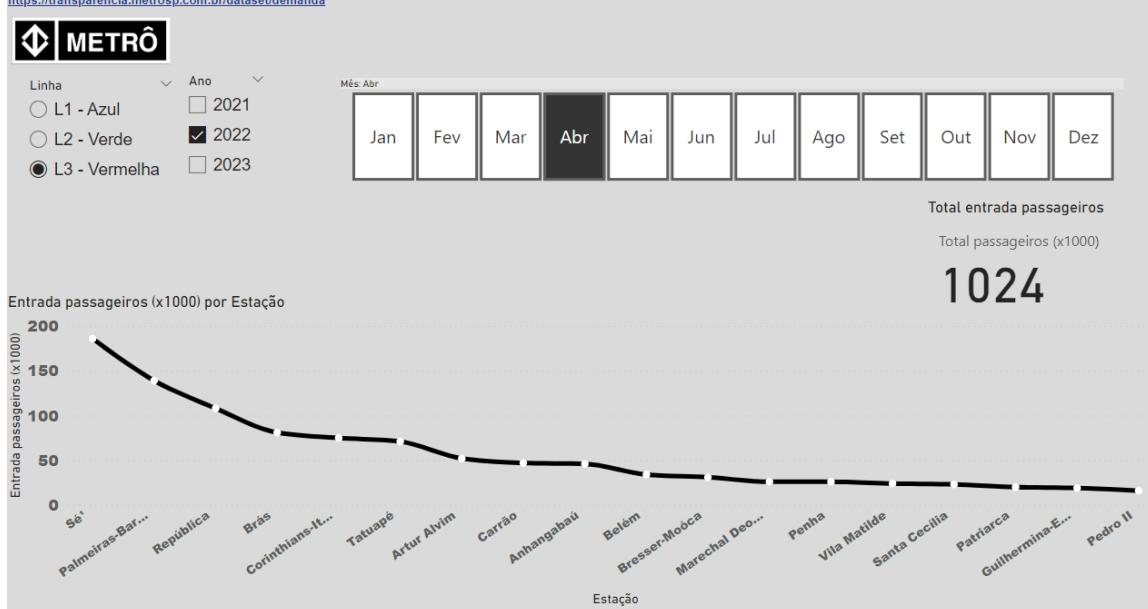
Definição de objetivos

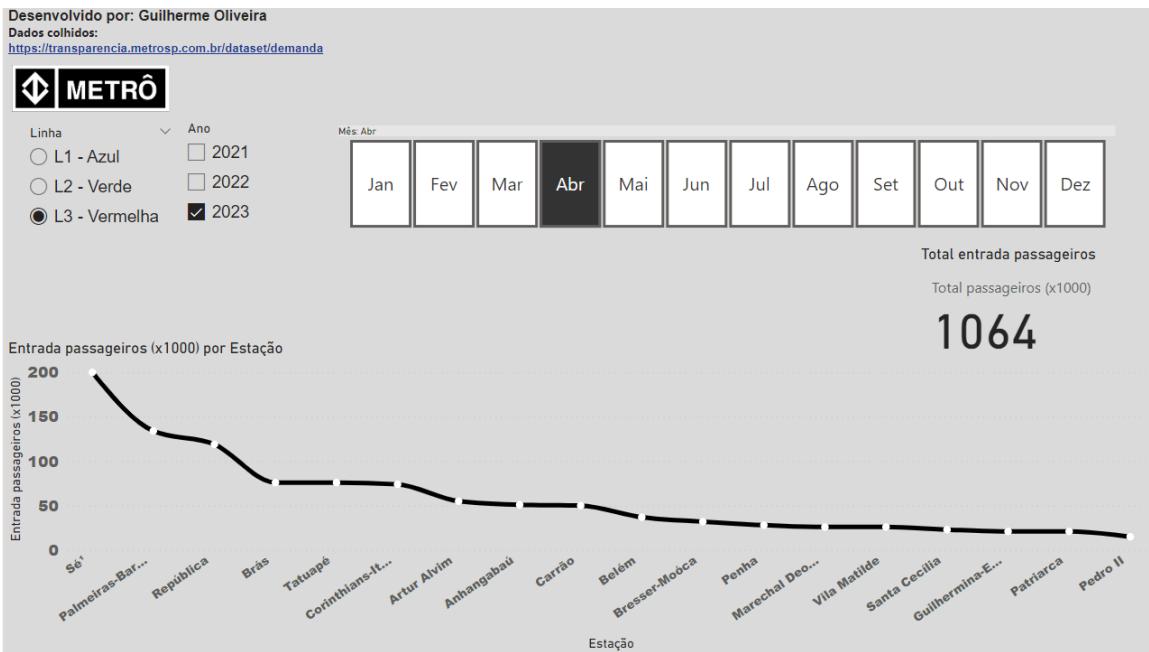
1. **Tendências Anuais:** Como a utilização das estações de metrô mudou ao longo dos anos 2021, 2022 e 2023? Existem estações específicas que mostram um aumento ou diminuição significativa na utilização ao longo desse período?
 - a. Devido a pandemia da COVID-19 no ano de 2020, todo fluxo de transporte foi afetado de maneira a evitar mais contato social, consequentemente, menos passageiros na malha metroviária. Cada ano que passou ficou mais flexível as políticas de saúde, e assim, aumentou o fluxo gradativamente.
 - b. O mês de abril houve uma grande diminuição comparado com os outros anos.

Desenvolvido por: Guilherme Oliveira
Dados colhidos:
<https://transparencia.metrosp.com.br/dataset/demanda>

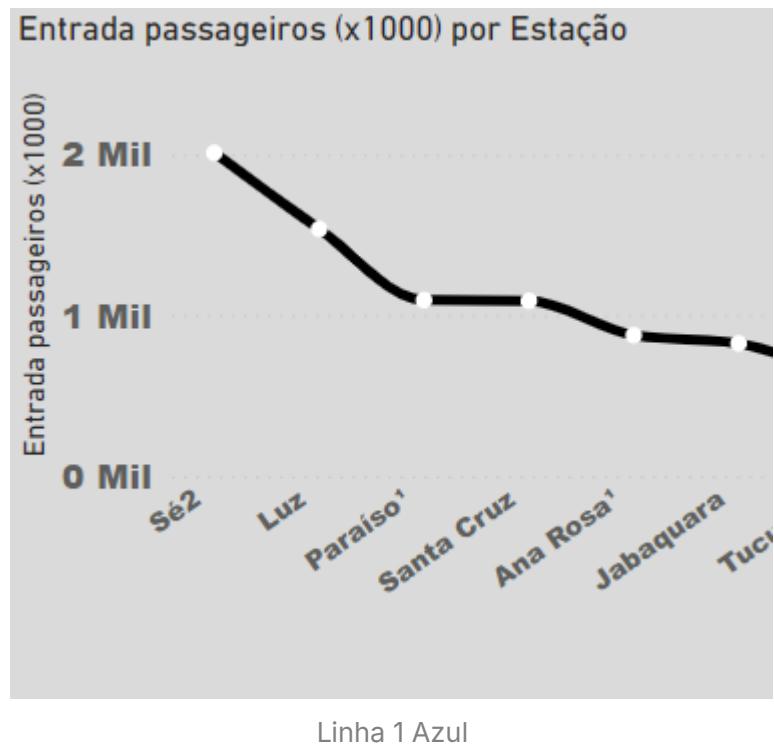


Desenvolvido por: Guilherme Oliveira
Dados colhidos:
<https://transparencia.metrosp.com.br/dataset/demanda>





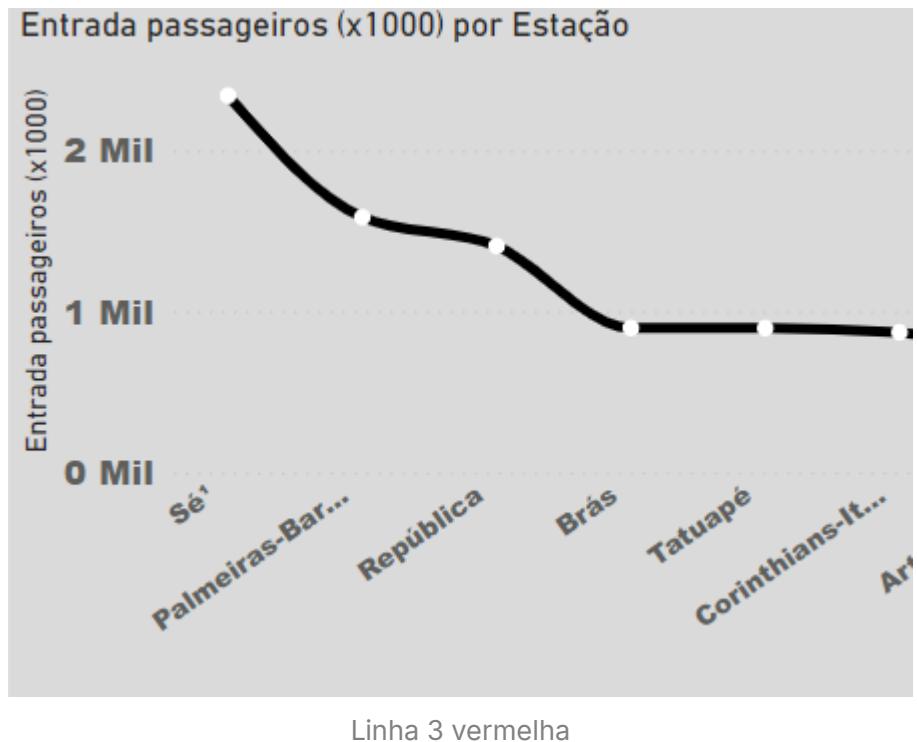
2. **Padrões Sazonais:** Existem padrões sazonais evidentes na utilização das estações de metrô? Por exemplo, há meses específicos que consistentemente mostram maior ou menor utilização ao longo dos anos?
3. **Estações Mais e Menos Utilizadas:** Quais foram as estações mais e menos utilizadas em cada ano? Essas posições mudam ao longo dos anos ou são consistentes?
 - a. As 6 estações com mais movimentações possuem baldeação com outras linhas de trem/metrô.



Linha 1 Azul



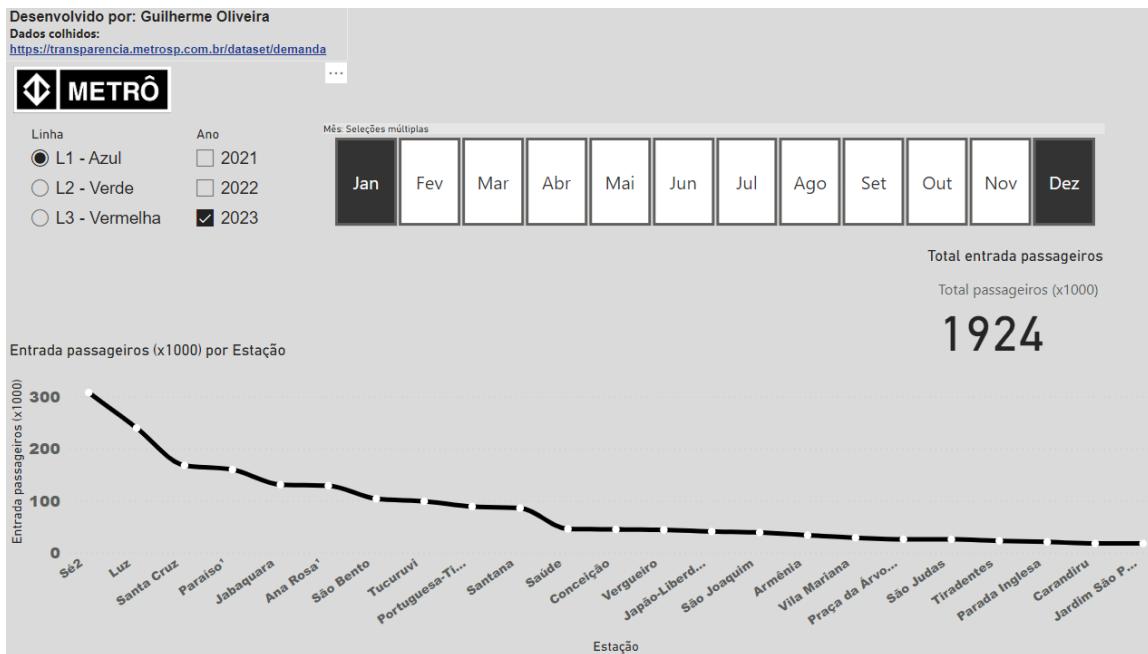
Linha 2 verde



Linha 3 vermelha

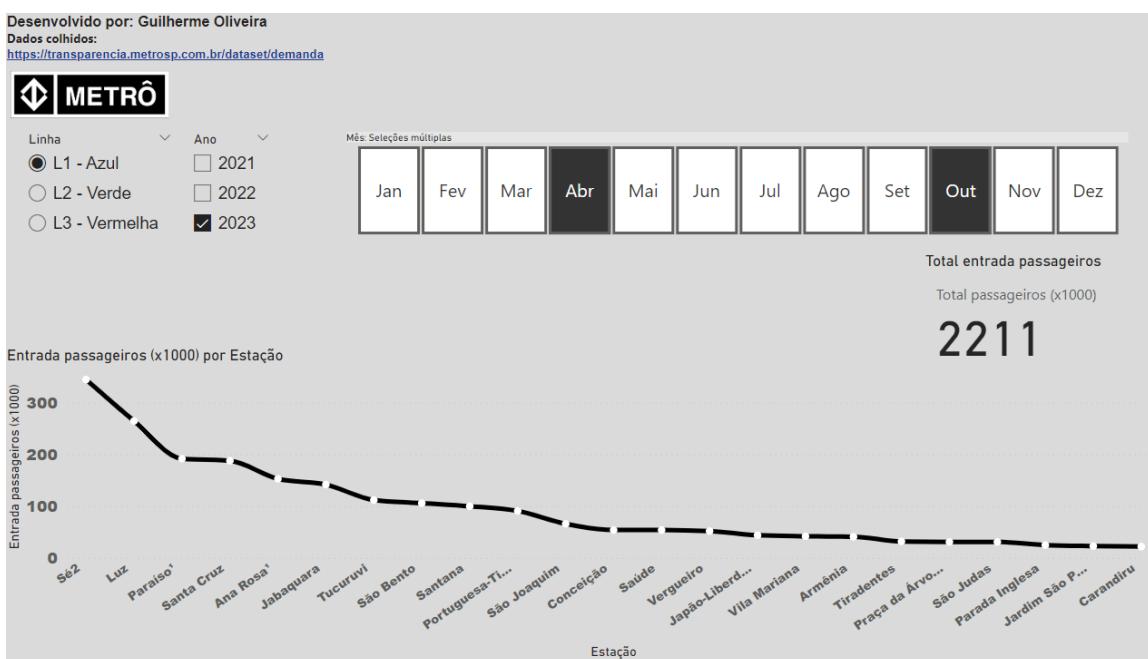
4. Impacto de Eventos Específicos: Pode-se identificar o impacto de eventos específicos (como feriados, eventos esportivos ou culturais) na utilização das estações de metrô? Isso requereria uma análise mais detalhada ou dados adicionais sobre eventos locais.

- a. Meses como Janeiro, Dezembro possuem menor fluxo de pessoas pois são comemorados inicio e final de ano. Época de férias escolares e recesso.



Ano 2023, Soma de Jan + Dez totalizando 1924

Utilizando 2 meses aleatórios fora de época possui maior fluxo.

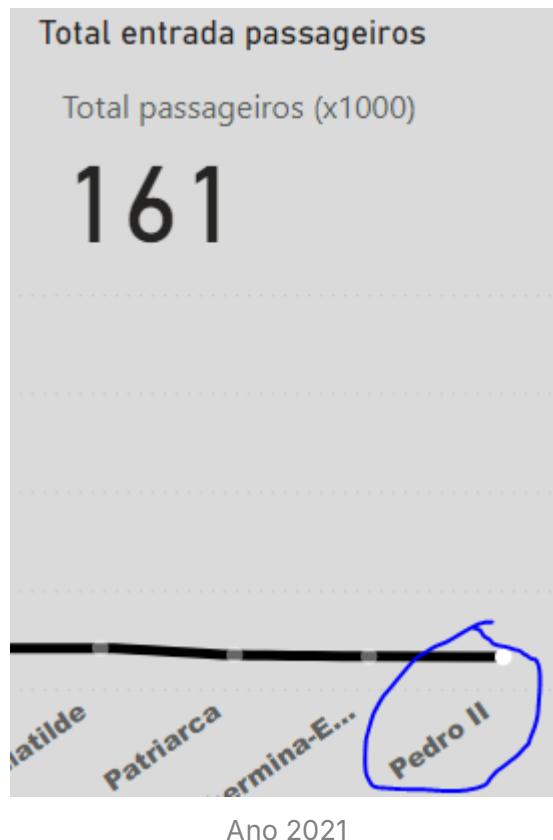


Ano 2023, Soma de Abr + Out totalizando 2211

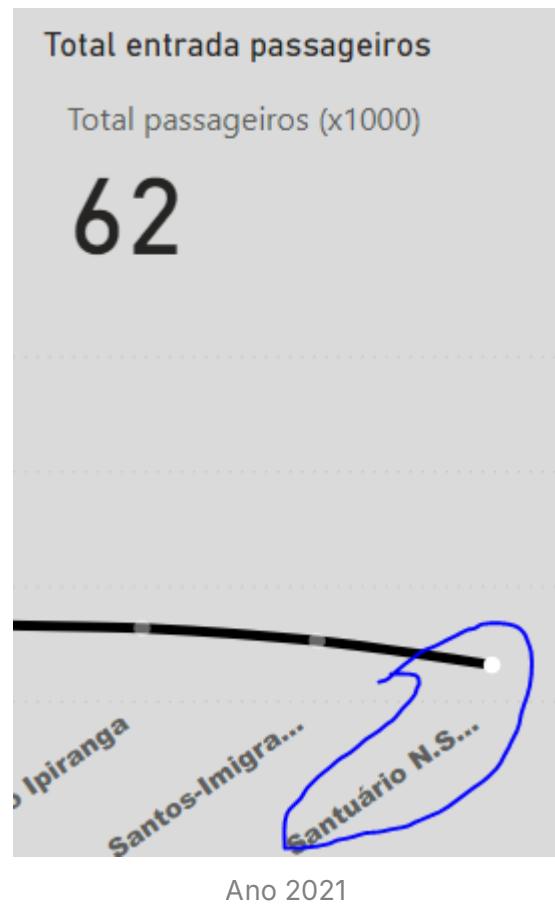
5. Comparação entre Estações: Como a utilização das estações se compara dentro de regiões específicas da cidade? Existem diferenças significativas na utilização de estações que são geograficamente próximas?

- É possível analisar que na linha 3 vermelha a estação com menor fluxo é Pedro II, ela possui um menor número de residência na região e fica

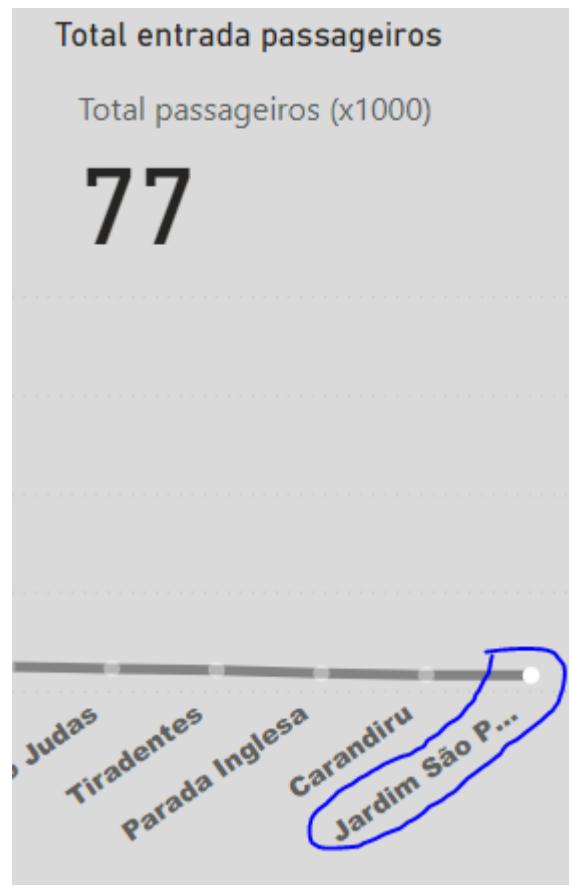
próximo da estação Sé. Isto pode influenciar os passageiros a migrar para outra estação.



- b. Na linha 2 verde a estação com menor fluxo é a Santuário N.Sra de Fátima-Sumaré, é uma região mais calma e próxima das outras estações.

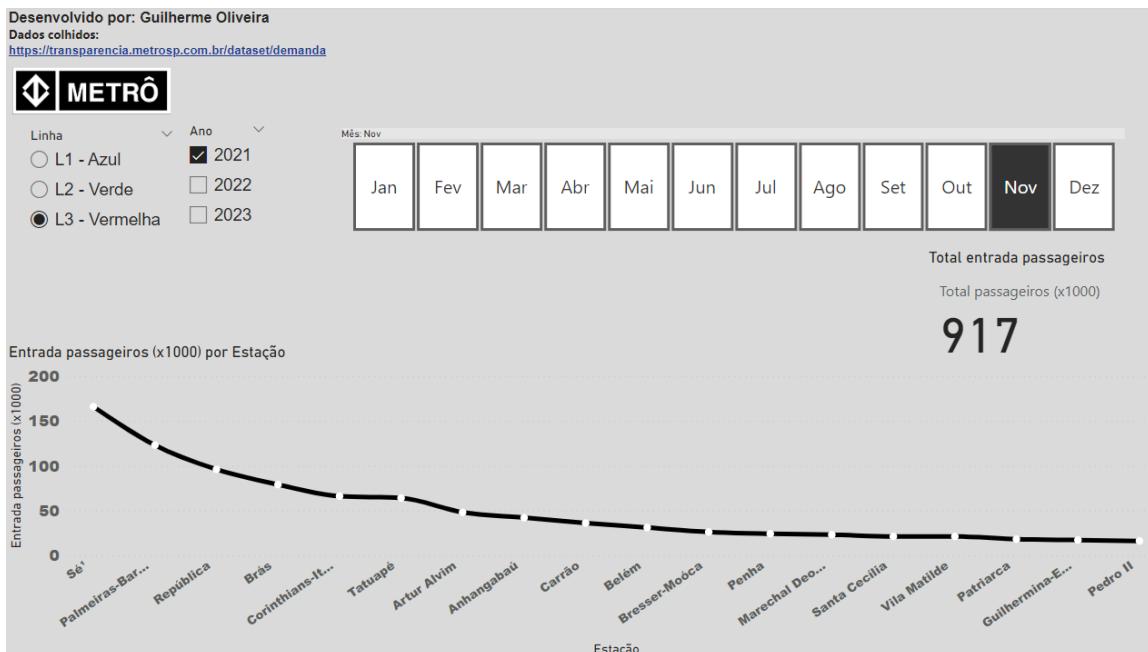


- c. Na linha 1 azul a estação com menor fluxo é Jardim São Paulo. Lá é uma zona residencial.

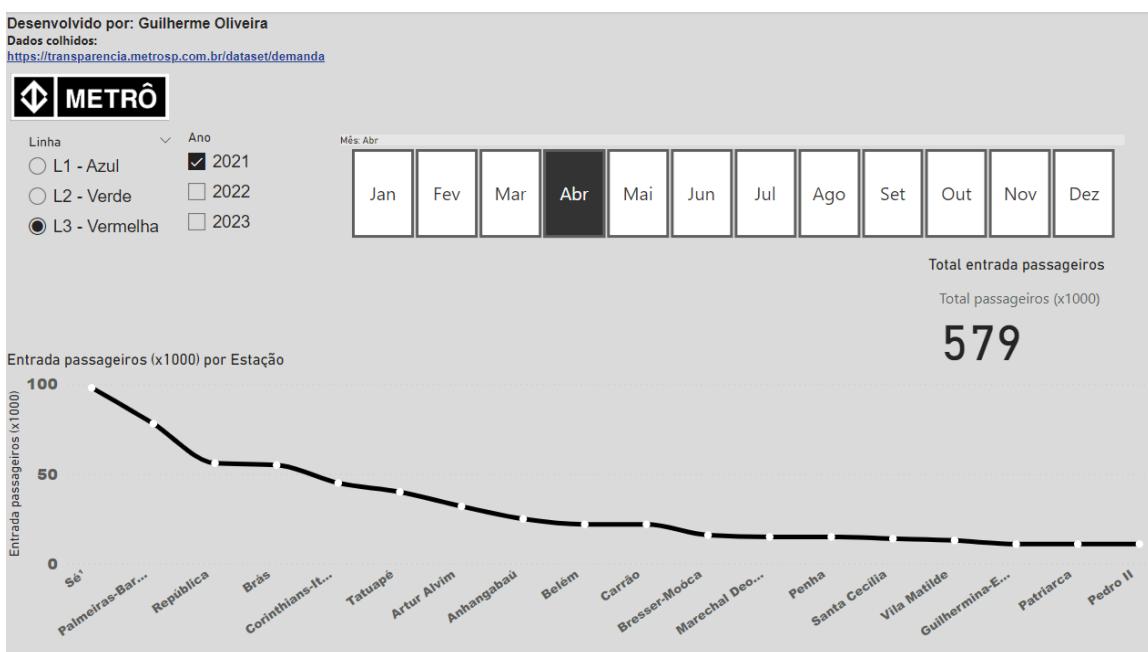


6. **Variação Mensal:** Qual mês apresentou a maior variação na utilização do metrô entre as estações? E qual estação teve a maior variação de utilização ao longo do ano?

- a. Usando o ano de 2021, o mês de abril teve o pior desempenho do ano, enquanto novembro teve o melhor.



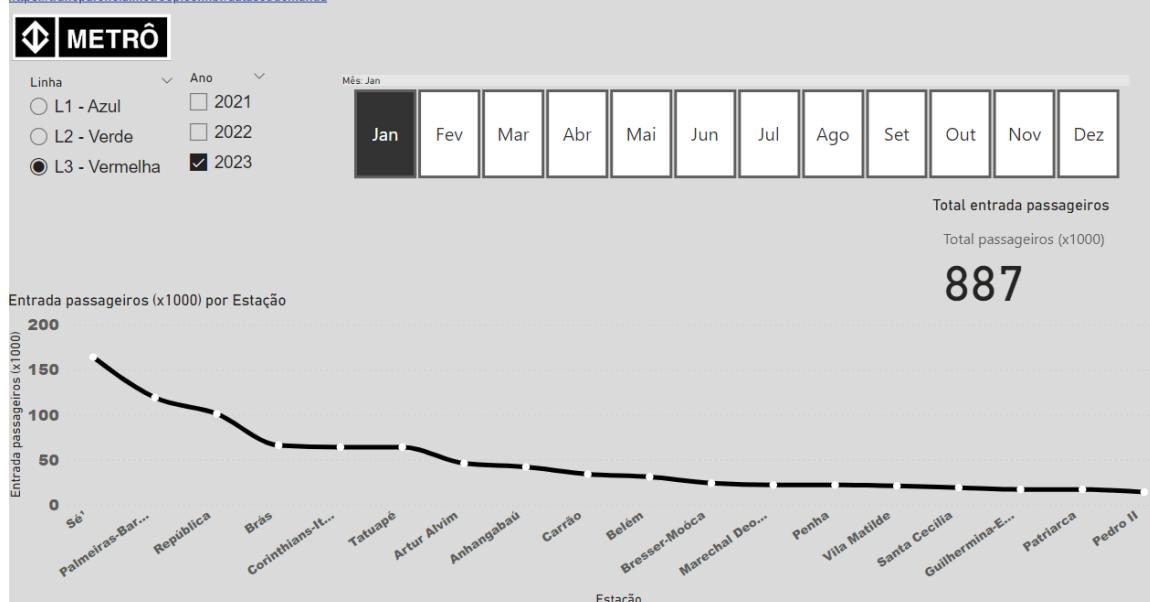
Ano 2021, mês Novembro



Ano 2021, mês Abril

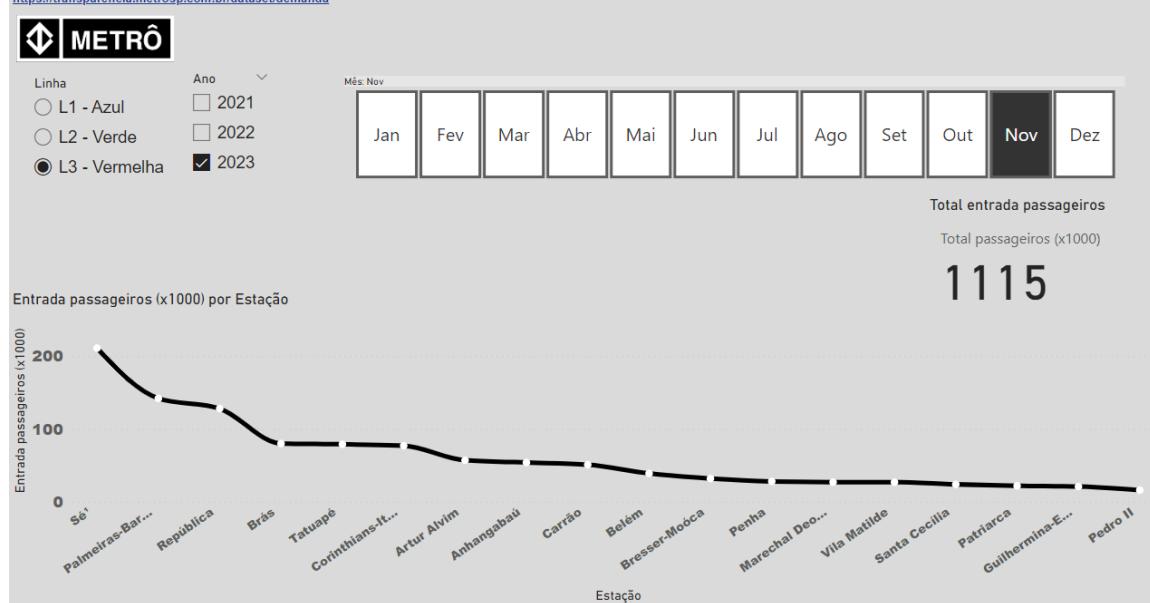
Já analisando o ano 2023, Janeiro teve o menor número de passageiros enquanto o mês de Novembro teve a maior alta.

Desenvolvido por: Guilherme Oliveira
 Dados colhidos:
<https://transparencia.metrosp.com.br/dataset/demand>



Ano 2023, mês Janeiro

Desenvolvido por: Guilherme Oliveira
 Dados colhidos:
<https://transparencia.metrosp.com.br/dataset/demand>



Ano 2023, Mês Novembro

Melhorias

As perguntas abaixo necessitam de uma análise gráfica mais robusta para responder as questões.

7. **Análise de Médias:** Como a média de utilização de cada estação se compara com a média geral do sistema de metrô? Existem estações que estão significativamente acima ou abaixo dessa média?
8. **Correlações:** Existe alguma correlação entre a utilização de estações adjacentes ou dentro de linhas específicas de metrô? Isso pode indicar padrões de deslocamento ou a influência de fatores locais.
9. **Análise de Crescimento:** Quais estações mostraram o maior crescimento percentual na utilização de um ano para outro? Isso pode indicar áreas de desenvolvimento urbano, mudanças demográficas ou melhorias na infraestrutura.
10. **Previsões Futuras:** Com base nos dados disponíveis, pode-se tentar prever a utilização das estações para o próximo ano? Quais métodos ou modelos poderiam ser usados para essa previsão?

