

Análise fluxo passageiros L1 Jabaquara

Esse estudo mostra o desenvolvimento da lógica passo a passo para verificar o fluxo de pessoas durante 3 anos de uma estação da linha azul do metrô.

Importando as bibliotecas

```
import pandas as pd
import os
import matplotlib.pyplot as plt
```

Verificando o diretório

```
notebook_dir = os.getcwd()
```

```
notebook_dir
```

```
 '/home/notebook'
```

Os arquivos são exibidos em uma lista

```
arquivos_no_diretorio = os.listdir(notebook_dir)
# Exibe os arquivosprint(arquivos_no_diretorio)
```

```
['.ipynb_checkpoints', 'L1', 'L15', 'L2', 'L3', 'Untitled.ipynb']
```

```
pasta_alvo = 'L1'caminho_pasta_alvo = os.path.join(notebook_dir, pasta_alvo)
```

```
os.chdir(caminho_pasta_alvo)
```

```
Ler_pasta = os.getcwd()
```

```
Ler_pasta
```

```
 '/home/notebook/L1'
```

Meu erro foi atribuir o nome do arquivo na variavel 'arquivo_2021' invés de atribuir como lista

```
arquivo_2021 = 'limpo_metro_2021 - L15.csv'
```

```
arquivo_2021 = arquivos_no_diretorio[0]
```

```
dados_2021 = pd.read_csv(arquivo_2021, nrows=5)
```

Com isso consegui ler a tabela L1

dados_2021

	LINHA 1-AZUL	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6
0	Estação	Jan*	Fev*	Mar*	Abr*	Mai*	Jun*
1	Jabaquara	46	48	40	39	45	46
2	Conceição	15	16	12	12	14	15
3	São Judas	9	9	7	7	8	9
4	Saúde	16	17	13	13	15	16

Com isso consegui ler a tabela L2

```
Ler_pasta = os.getcwd()  
Ler_pasta
```

```
'/home/notebook/L1'
```

```
arquivos_no_diretorio = os.listdir(Ler_pasta)  
# Exibe os arquivosprint(arquivos_no_diretorio)
```

```
['limpo_metro_2021 - L1.csv', 'limpo_metro_2022 - L1.csv', 'limpo_metro_2023 - L1.csv']
```

```
arquivo_2022 = arquivos_no_diretorio[1]  
arquivo_2022
```

```
'limpo_metro_2022 - L1.csv'
```

```
dados_2022 = pd.read_csv(arquivo_2022, nrows=5)
```

dados_2022

	LINHA 1-AZUL	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6
0	Estação	Jan*	Fev*	Mar*	Abr*	Mai*	Jun*
1	Jabaquara	53	58	63	66	66	64
2	Conceição	16	20	22	24	23	22
3	São Judas	10	11	13	14	13	13
4	Saúde	18	22	24	25	25	24

Com isso é possível ler a tabela L3

```
arquivos_no_diretorio = os.listdir(Ler_pasta)
# Exibe os arquivosprint(arquivos_no_diretorio)
```

```
['limpo_metro_2021 - L1.csv', 'limpo_metro_2022 - L1.csv', 'limpo_metro_2023 - L1.csv']
```

```
arquivo_2023 = arquivos_no_diretorio[2]
arquivo_2023
```

```
'limpo_metro_2023 - L1.csv'
```

```
dados_2023 = pd.read_csv(arquivo_2023, nrows=5)
```

```
dados_2023
```

	Estação	Jan	Fev	Mar	Abr	Mai	Jun
0	Jabaquara	63	66	70	72	70	68
1	Conceição	20	24	26	26	26	25
2	São Judas	12	14	15	15	15	15
3	Saúde	21	26	27	27	27	27
4	Praça da Árvore	12	15	15	15	16	15

Para remover o cabeçalho

Filtrando

os dados para a estação Vila Prudente de 2021 e 2022

```
dados_2021_sem_cabecalho = pd.read_csv(arquivo_2021, skiprows=1, nrows=5)
dados_2022_sem_cabecalho = pd.read_csv(arquivo_2022, skiprows=1, nrows=5)
dados_2023_sem_cabecalho = pd.read_csv(arquivo_2023, nrows=5)
dados_2021_sem_cabecalho, dados_2022_sem_cabecalho, dados_2023_sem_cabecalho
```

(Estação	Jan*	Fev*	Mar*	Abr*	Mai*	Jun*	Jul*	Ago*	Set*	\
0	Jabaquara	46	48	40	39	45	46	48	51	55	
1	Conceição	15	16	12	12	14	15	15	16	18	
2	São Judas	9	9	7	7	8	9	9	10	10	
3	Saúde	16	17	13	13	15	16	16	18	19	
4	Praça da Árvore	9	10	8	8	9	9	10	11	11	
	Out*	Nov*	Dez*	Média							
0	57	61	59	50							
1	19	20	19	16							
2	11	12	11	9							
3	20	21	20	17							
4	12	13	12	10	,						
	Estação	Jan*	Fev*	Mar*	Abr*	Mai*	Jun*	Jul*	Ago*	Set*	\
0	Jabaquara	53	58	63	66	66	64	64	66	67	
1	Conceição	16	20	22	24	23	22	21	23	24	
2	São Judas	10	11	13	14	13	13	13	13	14	
3	Saúde	18	22	24	25	25	24	23	25	26	
4	Praça da Árvore	10	12	14	14	14	14	13	15	14	

	Out*	Nov*	Dez*	Média											
0	69	69	64	64											
1	24	25	21	22											
2	15	14	13	13											
3	26	27	23	24											
4	15	15	13	14	,										
			Estação	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	\
0			Jabaquara	63	66	70	72	70	68	66	69	72	70	73	
1			Conceição	20	24	26	26	26	25	23	26	28	28	28	
2			São Judas	12	14	15	15	15	15	14	16	16	16	16	
3			Saúde	21	26	27	27	27	27	24	27	27	27	28	
4	Praça	da	Árvore	12	15	15	15	16	15	14	16	16	16	16	
	Dez	Média													
0	68	69													
1	25	25													
2	14	15													
3	25	26													
4	14	15)												

Existe uma incompatibilidade das linhas, onde o mês tem asterisco (*) e o ano 2023 não tem

```
dados_2021_asterisco = pd.read_csv(arquivo_2021, skiprows=1).rename(columns=lambda x: x.replace('*', ''))
dados_2022_asterisco = pd.read_csv(arquivo_2022, skiprows=1).rename(columns=lambda x: x.replace('*', ''))
dados_2023_asterisco = pd.read_csv(arquivo_2023)
```

`dados_2023_asterisco`

	Estação	Jan	Fev	Mar	Abr	Mai	Jun
0	Jabaquara	63	66.000	70.000	72.000	70.000	68.00
1	Conceição	20	24.000	26.000	26.000	26.000	25.00
2	São Judas	12	14.000	15.000	15.000	15.000	15.00
3	Saúde	21	26.000	27.000	27.000	27.000	27.00
4	Praça da Árvore	12	15.000	15.000	15.000	16.000	15.00
5	Santa Cruz	76	88.000	91.000	91.000	91.000	90.00
6	Vila Mariana	13	19.000	21.000	21.000	20.000	19.00
7	Ana Rosa ¹	59	70.000	75.000	75.000	75.000	73.00
8	Paraíso ¹	73	87.000	94.000	94.000	94.000	90.00
9	Vergueiro	21	25.000	26.000	26.000	26.000	25.00
10	São Joaquim	17	24.000	33.000	34.000	32.000	27.00
11	Japão-Liberdade	19	20.000	22.000	22.000	21.000	21.00
12	Sé2	144	164.000	172.000	171.000	170.000	167.00
13	São Bento	48	51.000	51.000	51.000	51.000	52.00
14	Luz	108	122.000	127.000	130.000	132.000	128.00

15	Tiradentes	10	15.000	16.000	16.000	16.000	15.00
16	Armênia	17	20.000	21.000	21.000	21.000	20.00
17	Portuguesa-Tietê	43	43.000	44.000	45.000	44.000	45.00
18	Carandiru	8	10.000	11.000	11.000	11.000	11.00
19	Santana	41	47.000	50.000	50.000	49.000	49.00
20	Jardim São Paulo-Ayrton Senna	8	11.000	11.000	11.000	11.000	11.00
21	Parada Inglesa	9	11.000	12.000	12.000	12.000	12.00
22	Tucuruvi	47	54.000	56.000	56.000	56.000	54.00
23	Total	891	1.025	1.087	1.091	1.086	1.06

```
dados_Jabaquara_2021 = dados_2021_asterisco[dados_2021_asterisco['Estação'] == 'Jabaquara']
dados_Jabaquara_2022 = dados_2022_asterisco[dados_2022_asterisco['Estação'] == 'Jabaquara']
dados_Jabaquara_2023 = dados_2023_asterisco[dados_2023_asterisco['Estação'] == 'Jabaquara']
```

```
dados_Jabaquara_2023
```

	Estação	Jan	Fev	Mar	Abr	Mai	Jun
0	Jabaquara	63	66.0	70.0	72.0	70.0	68.0

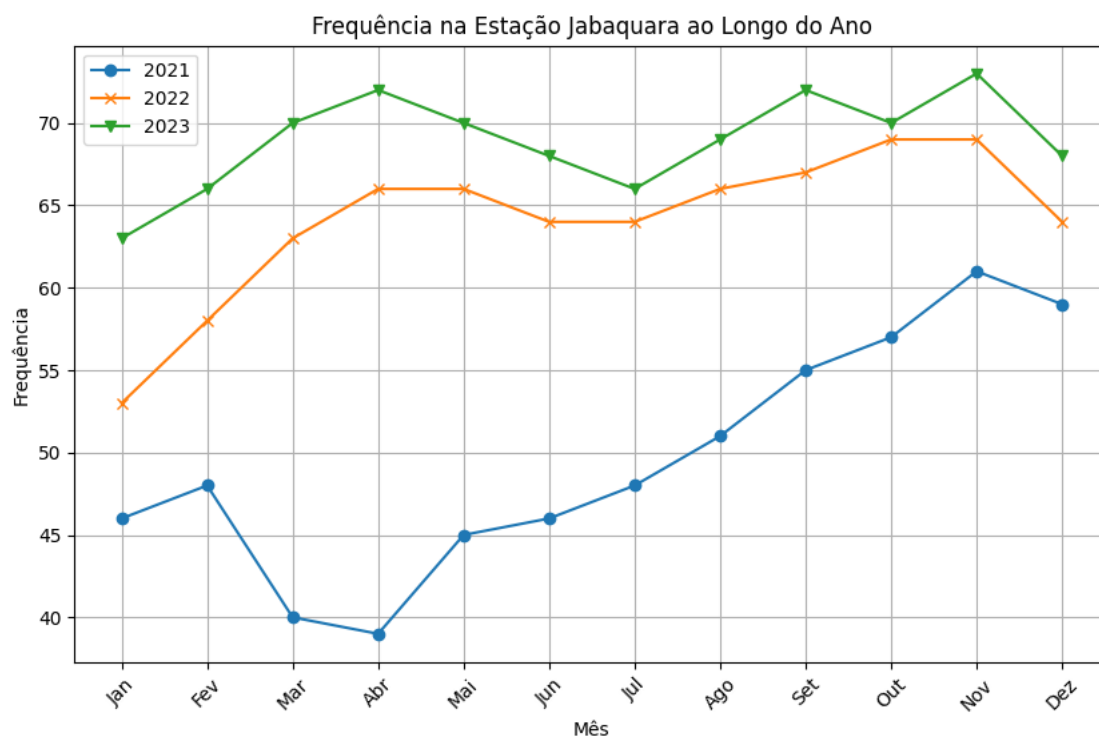
```
meses_sem_asterisco = ['Jan', 'Fev', 'Mar', 'Abr', 'Mai', 'Jun', 'Jul', 'Ago', 'Set', 'Out', 'Nov', 'Dez']
```

Unindo os dados dos dois anos para uma comparação

```
meses = ['Jan', 'Fev', 'Mar', 'Abr', 'Mai', 'Jun', 'Jul', 'Ago', 'Set', 'Out', 'Nov', 'Dez']
valores_2021 = dados_Jabaquara_2021[meses].values.flatten()
valores_2022 = dados_Jabaquara_2022[meses].values.flatten()
valores_2023 = dados_Jabaquara_2023[meses].values.flatten()
```

Criando o gráfico

```
plt.figure(figsize=(10, 6))
plt.plot(meses, valores_2021, label='2021', marker='o')
plt.plot(meses, valores_2022, label='2022', marker='x')
plt.plot(meses, valores_2023, label='2023', marker='v')
plt.title('Frequência na Estação Jabaquara ao Longo do Ano')
plt.xlabel('Mês')
plt.ylabel('Frequência')
plt.xticks(rotation=45)
plt.legend()
plt.grid(True)
plt.show()
```



O gráfico no eixo X é o mês, no eixo Y a quantidade de pessoas em milhares

Extraindo informação do gráfico

É possível analisar que a cada ano que passou, aumentou o fluxo de pessoas, isso devido o fator da pandemia em 2020, que o governo de São Paulo implantou medidas de contenção para evitar o fluxo de passageiros.

Ao longo do ano de 2021, as pessoas voltaram a trabalhar gradativamente, conforme novas orientações de segurança foram implementadas.

No anos seguintes(2022, 2023) o gráfico tende a se comportar da mesma maneira ao longo dos meses, com a diferença da quantidade de pessoas que foram aderindo o meio de transporte.