

Relatório EP1: Processando dados da pesquisa Origem Destino

Alexia Carolina Scheffer da Silva — 10724100

Felipe Erich Appel Romero — 8519204

Felipe Ribeiro de Campos Marques — 10856716

Gustavo Jyun Hayashida — 10816902

Julio Silva Ladeira — 10872350

Mitiko Celene de Araújo Hashizume — 10828986

Repositório: <https://github.com/FelipeRCMarques/EP-AED2/tree/master/EP1>

Linguagem e bibliotecas

A linguagem escolhida foi **Python** por causa da facilidade dos integrantes do grupo com ela e pela quantidade de bibliotecas disponíveis que ajuda no manejo de dados.

Foi utilizado três bibliotecas:

pandas: para a leitura do .csv

```
pip install pandas
```

Com o pandas, pegamos as coordenadas:

CO_DOM_X e **CO_DOM_Y**: coordenadas do domicílio

CO_ESC_X e **CO_ESC_Y**: coordenadas da escola

CO_TR1_X e **CO_TR1_Y**: coordenadas do trabalho 3

CO_TR2_X e **CO_TR2_Y**: coordenadas do trabalho 2

CO_O_X e **CO_O_Y**: coordenadas da origem

CO_D_X e **CO_D_Y**: coordenadas do destino

CO_T1_X e **CO_T1_Y**: coordenadas da transferência 1

CO_T2_X e **CO_T2_Y**: coordenadas da transferência 2

CO_T3_X e **CO_T3_Y**: coordenadas da transferência 3

ID_PESS: ID do entrevistado

Todas essas coordenadas foram selecionadas, pois em todas ela pode haver contato com outras pessoas. Mesmo no próprio domicílio, se alguma pessoa sai de casa, ela pode trazer o vírus para dentro, por exemplo. Então todos esses dados são importantes.

matplotlib.pyplot: para gerar o histograma

```
pip install matplotlib
```

Com o pyplot, fizemos o histograma:



Nele, o eixo x é os números de frequentadores em unidades e o eixo y em é a quantidade de lugares em escala de log.

time: para a contagem do tempo de execução

Análise do tempo de processamento

Como o tempo de processamento depende de diversas situações no computador, como a quantidade de CPU e a quantidade de processos em execução, por exemplo. Todos os programas não essenciais foram fechados e o programa foi rodado 6 vezes:

1º: 48.29 segundos

2º: 47.78 segundos

3º: 48.32 segundos

4º: 48.41 segundos

5º: 48.83 segundos

6º: 48.47 segundos

Média: 48.35 segundos

E usar uma estrutura, como uma lista, que estivesse ordenada por número de frequentadores, melhoraria o algoritmo, pois não precisaria de um loop.