Documentação do preparo dos dados – André Yuri Marques dos Santos

1. Baixei os arquivos enviados por e-mail e em seguida fui conferi-los.

2. Assim que abri os arquivos, me deparei com a ausência de formatação do arquivo CSV, então formatei da forma que achei adequado para dar sequência a resolução do teste. Os dados contidos no arquivo não se encontravam separados por colunas, ajustei da seguinte maneira: no excel, selecionei a coluna única e em seguida, Dados > Texto para Colunas > Delimitado > Avançar e marquei a caixa Virgula como delimitador. Feito isso, ajustei a largura das colunas em Células > Formatar > Ajustar Largura das Colunas.

3. Após formatar as colunas dentro da tabela, resolvi fazer o tratamento que achei adequado nos dados através do Google Colab.

4. Reorganizei os dados através de novas colunas na tabela, sendo essas colunas *year_month* e *year*, onde fiz um agrupamento dos dados por cada mês de seus respectivos anos e depois um agrupamento por cada ano em específico.

As etapas para esse tratamento que fiz nos dados utilizando o Google Colab foram as seguintes:

```
# Utilizei a biblioteca Pandas, então importei a mesma através do comando:

Import pandas as pd

# Usei para ler o arquivo Excel (tive que específicar o engine openpyxl pois sem ele não estava conseguindo visualizar a tabela aqui no colab)

df = pd.read_excel('audiencia_jpgnews.xlsx', engine='openpyxl')

# Coloquei 4000 pois como padrão o df.head só exibe as 5 primeiras linhas dos dados

df.head(4000)

# Usei para acessar as informações de colunas e linhas da tabela

df.info()

# Usei o groupby para agrupar os dados por mês\ano e dispositivo (desktop ou mobile) - [numeric_columns] utilizado para calcular a soma e a contagem das colunas pageviews e videoviews

df.groupby(['year_month', 'device'])[numeric_columns].sum()
```

```
# Usei "meses = pd.DataFrame" para criar uma nova variável chamada meses e criar um novo dataframe com o groupby dos meses e anos

meses = pd.DataFrame(df.groupby(['year_month', 'device'])[numeric_columns].sum())


# Assim como anteriormente, usei o groupby para agrupar os dados por ano e dispositivo (desktop ou mobile) - [numeric_columns] utilizado para calcular a soma e a contagem das colunas pageviews e videoviews

df.groupby(['year', 'device'])[numeric_columns].sum()


# Usei "anos = pd.DataFrame" para criar uma nova variável chamada anos e criar um novo dataframe com o groupby dos anos

anos = pd.DataFrame(df.groupby(['year', 'device'])[numeric_columns].sum())


# Comandos utilizados para criar/baixar no Colab as novas tabelas no formato xlsx com as colunas de meses e anos filtrados

meses.to_excel("Dados_Mes.xlsx")

anos.to_excel("Dados_Anos.xlsx")
```

5. Após tratar os dados no Google Colab, retornei ao Excel para poder gerar os gráficos e tabelas necessários para a análise dos dados fornecidos.

Fontes utilizadas:

Python+Excel: Tratamento de dados de uma tabela no Python/Pandas *https://www.youtube.com/watch?v=f3z2vL1LiDo&ab_channel=ThalesVeloso*

Apostila de Visualização de Dados professor Giancarlo Costa, UERJ.

https://drive.google.com/file/d/1vNY9muuqG8gSSMpns1aMSb3ONmANQSct/view?usp=drive_link