**Atividade de Ciências de Dados**

**Aluno:** Heberth Vinícius Amarante Severo

**Descrição do problema:**

Essa é a quinta parte da nossa atividade de Web Scraping, nela iremos modificar o arquivo csv da atividade anterior e gerar um novo arquivo csv. Iremos criar novas colunas e consequentemente inserir os dados necessários nelas, por fim iremos criar um novo arquivo csv. Todas as etapas dessa atividade serão descritas e explicadas a seguir.

**1ª Etapa**

Para o início do arquivo no Colab é necessário fazer a importação do pandas, como mostra a Figura 1.

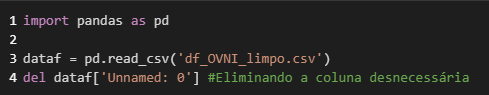


Figura 1 Importação do pandas e leitura do csv

Nesse primeiro trecho de código há apenas três linhas, a primeira é para a importação do pandas, a linha de número três estamos lendo o nosso arquivo csv, e por fim, na linha de número quatro estamos deletando uma linha que não tem utilidade para nossa atividade.

**2ª Etapa**

Nessa etapa iremos criar duas colunas, uma para a data e uma para a hora, nessa etapa também iremos deletar a coluna ‘Date / Time’, que já existe no nosso arquivo csv.

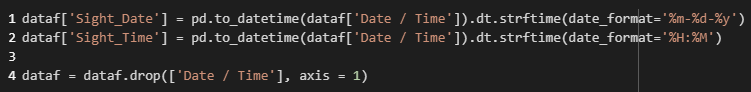


Figura 2 Criando as colunas data e hora, deletando Date / Time

Na primeira linha desse trecho de código estamos criando a coluna *Sight\_Date* para isolar a data, o formato escolhido para essa coluna foi o padrão norte americano, mês/dia/ano.

A segunda linha é a criação da coluna para a hora, chamada de *Sight\_Time*, estamos passando o formato tradicional para hora, que é hora:minuto.

A quarta linha é para deletar a coluna *Date / Time* do nosso csv.

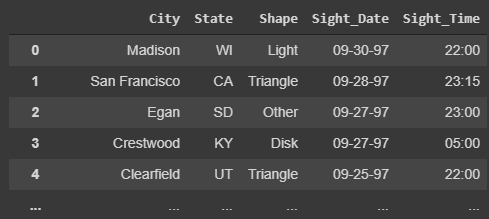


Figura 3 Csv com as novas colunas para data e hora

A Figura 3 mostra como está o nosso arquivo até o momento, após a inserção das colunas da data e da hora, e agora sem a coluna *Date / Time*.

**3ª Etapa**

Nessa etapa iremos criar mais uma coluna no data frame, esse coluna se chamará Sight\_Weekday e nela será mostrado os dias da semana que foram feitos os relatos.



Figura Criação da coluna Sight\_Weekday

Podemos observar na Figura 4, basicamente, duas coisas, no início da linha estamos dando o nome desejado para a nossa nova coluna, e no final da linha estamos passando o que queremos colocar nela, no nosso caso queremos o nome do dia da semana, por isso passamos *dt.weekday\_name*, que serve para mostrar o nome do dia.

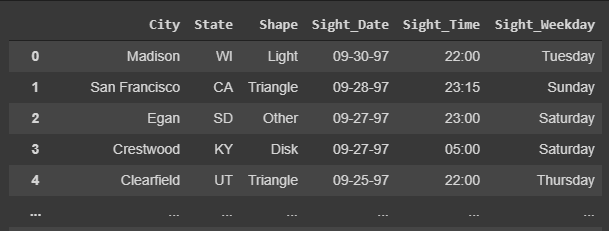


Figura Resultado após a inserção da coluna Sight\_Weekday

A Figura 5 mostra como está o nosso data frame até o momento, após a inserção da coluna *Sight\_Weekday*.

**4ª Etapa**

Essa etapa é a criação de mais duas colunas que se referem a data, criaremos uma coluna para o dia do relato e uma coluna para o mês do relato, essa colunas se chamarão *Sight\_Day* e *Sight\_Month*, como mostra a Figura 6.



Figura Criação das colunas Sight\_Day e Sight\_Month

Após a execução do bloco de código anterior, o nosso data frame terá duas novas colunas. A Figura 7 mostra o arquivo atualizado.



Figura Data frame após a inserção das novas colunas

**5ª Etapa**

Essa é a última etapa dessa atividade, nela iremos criar um arquivo csv, o nome desse arquivo será *df\_OVNI\_preparado*, como mostra a Figura 8.



Figura Criação do arquivo csv

**Observação**

Este relatório estará no GitHub, cujo link está a seguir: <https://github.com/heberth-hvas/icd_5>

**Referências**

**Google Colab:** <https://colab.research.google.com/notebooks/welcome.ipynb>

**Documentação Pandas:** <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.append.html>

**Pandas dt.weekday:** <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.dt.weekday.html>

**GeekforGeeks pandas timestamp:** <https://www.geeksforgeeks.org/python-pandas-timestamp-weekday/>

**Pandas to\_datetime:** <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.to_datetime.html>

**GeekforGeeks pandas to\_datetime:** <https://www.geeksforgeeks.org/python-pandas-to_datetime/>