Instituto Federal de Brasília

Nome: Vanessa de Andrade Formiga

**Acréscimo de Variáveis**

**Definição do Problema:**

Na penúltima parte do trabalho usando web scraping, teve como problema a ser solucionado nesta etapa, a divisão da variável Date/Time, para em seguida criar as colunas Sight\_Date e Sight\_Time, e depois excluir a coluna Date/Time. Na próxima etapa precisava encontrar a fim de buscar respostas para a problema proposta, que era saber a existência de um dia da semana que tinha mais relatos de ocorrências de Ovnis relatos. Para responder esse questionamento precisou gerar uma nova coluna denominada Sight\_Weekdays, que tem como finalidade ter a informação sobre o dia da semana que possui mais ocorrências de relatos, Como iniciativa de detalha mais as buscas, criou-se duas variáveis através da coluna Sight\_Date, foram criadas a partir disso as colunas Sight\_Month e Sight\_Day. E a atividade que finalizou o trabalho tratou-se de ser o salvamento de um novo arquivo, que recebeu o nome de df\_Ovni\_preparado.

**Solução:**

As etapas desse trabalho dividiu-se em partes menores. Na primeira etapa precisou carregar o arquivo gerado na etapa anterior em um DataFrame. Para carregar esse arquivo, teve que criar uma variável chamada df, que recebe o apelido dado no momento que ocorreu a importação da biblioteca pandas, é também recebe como argumento o read\_csv, que tem como finalidade a leitura do arquivo. E dentro dos parênteses passa o nome do arquivo e a sua extensão. A resolução dessa etapa, e mostrada na figura 1 anexada a seguir.

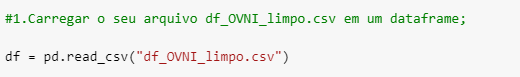


figura 1 – carregar o arquivo em um dataframe.

O resultado dessa etapa está anexada na figura 2 exibida a seguir.

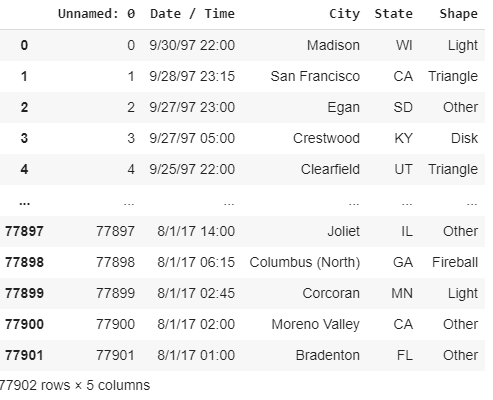


figura 2 – exibe o resultado da etapa.

Nessa segunda etapa precisava resolver vários problemas mostrados na figura 3, que era preciso dividir a coluna Date / Time em duas novas colunas, uma com o nome de Sight\_Date e a outra com o nome de Sight\_Time. Primeiro teve que transformar a coluna Date/Time em datetime, para depois criar a coluna Sight\_Date, que recebe o novo nome, o nome anterior e como essa data será apresentada em mês, dia e ano. E na coluna Sight\_Time passa o novo que é Sight\_Time, o novo nome anterior que é Date/Time e como ela será apresentada em hora, minuto e segundo. A tarefa que finalizou ocorreu um drop nas colunas Date/Time e Unnamed: 0.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

figura 3 – transforma o Date/Time em datetime, separa essa coluna e criar as colunas Sight\_Date e Sight\_Time e exclui a coluna Date/Time.

O resultado obtido nessa segunda etapa está sendo mostrados nas figuras 4,5 6.



figura 4 – exibe o resultado.



figura 5 – exibe o resultado.



figura 6 – exibe o resultado.

Na terceira etapa precisou ter a informação de qual o dia da semana que possui mais ocorrências de relatos registrados. Para conseguir esse dado, passo o nome da coluna Sight\_Weekdays, e depois transformou a variável Sight\_Date em datetime, e passou o day\_name, que tem como finalidade mostrar o dia da semana. Como ocorreu isso, essa detalhado na figura 7 mostrado a seguir.

Uma imagem contendo parede

Descrição gerada automaticamente

figura 7 – criar a coluna Sight\_Weekdays.

O resultado está sendo imprimido na figura 8.



figura 8 – exibe o resultado.

Depois de ter o dado sobre o dia da semana sobre a maior quantidade de ocorrências relatadas, e com isso respondeu à questão anexada na figura 9 exibida a seguir.



figura 9 – mostra a pergunta e a responda que foi pedida para ser solucionada.

Na quarta etapa mostrada na figura 10 anexada a seguir, precisou fazer um detalhamento dos dados, para isso pegou a coluna a coluna Sight\_Date, e separou essa coluna em Sight\_Month e Sight\_Day. Para ter mais informações sobre mês separadamente, foi passado o nome da nova variável, o nome da antiga, é também passou o strftime junto com a porcentagem mais a letra m para conseguir os dados do mês, e na coluna para o dia, informou o nome da coluna que será chamada Sight\_Day, depois transformou a coluna em Sight\_Date em datetime e por último passou o strtftime mais a porcentagem e a letra d.

Uma imagem contendo objeto

Descrição gerada automaticamente

figura 10 – criar as colunas Sight\_Month e Sight\_Day.

As figuras 11 e 12 mostra o resultado dessa etapa.

Uma imagem contendo texto

Descrição gerada automaticamente

figura 11 – mostra o resultado.

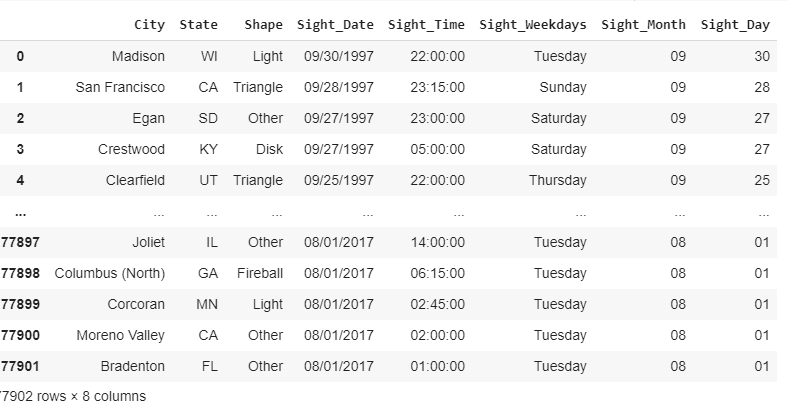


figura 12 – mostra o resultado.

Na quinta e última etapa informado na figura 13 mostrada a seguir, nesse trabalho a atividade que precisava ser feita, era salvar o arquivo em formato .csv com um novo nome, Com intenção disso, informou o df mais o to\_csv que serve para escrever o csv, juntamente com nome df\_OVNI\_preparado e a sua extensão.



figura 13 – salvar um novo arquivo chamado df\_OVNI\_preparado.

O resultado da última etapa está sendo mostrado a seguir na figura 14.



figura 14 – mostra o resultado.

**Será que existe um dia da semana com mais ocorrências de relatórios para OVNIs ?**

Sim. O dia da semana que possui mais ocorrências de relatos é a terça feira.

**Link Para o GitHub:**

<https://github.com/vanessaformiga/Nuforc/tree/master/acr%C3%A9scimo%20de%20vari%C3%A1veis>

**Ferramenta:**

As ferramentas utilizadas para desenvolver essa parte do trabalho foram o Colab e o Excel.

**Bibliografia:**

<https://colab.research.google.com/notebooks/intro.ipynb#recent=true>

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.to_datetime.html>

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.dt.day_name.html#pandas.Series.dt.day_name>

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.dt.strftime.html?highlight=dt%20strftime#pandas.Series.dt.strftime>