**Relatório de Adição de variáveis aos dados coletados do site *nuforc.org***

Aluno: Daniel Barros Candeira

1. **Introdução**

O relatório tem por objetivo demonstrar as etapas percorridas para adicionar variáveis aos dados coletados do site *nuforc.org* entre 07 de setembro de 1997 e 07 de agosto de 2017, site que contém relatos de “visualização” de OVNIs – Objetos Voadores Não Identificados, a fim de identificar informações relevantes.

1. **Problema**

Nessa etapa será realizada a adição de variáveis aos dados, tendo como atividades, carregar o arquivo df\_OVNI\_limpo.csv em um dataframe, dividir o conteúdo da coluna *Date / Time* em duas novas colunas no mesmo dataframe e deletar a coluna *Date / Time*, criar a coluna com o dia da semana em que ocorreram os relatos, e criar a coluna com o número do mês em que ocorreram os relatos.

1. **Resolução**

A fim de realizar os passos identificados na seção 2, foi desenvolvido um script utilizando a linguagem de programação **Python**, e a biblioteca **Pandas**,que server para lidar com dados de forma estatística e com base/arquivos de dados volumosos.

**1º Passo:**

Antes de iniciar a escrita do código, é realizada a importação da biblioteca a ser utilizadas e carregamos o dataframe, como é demonstrado na **Fig 1**.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 1. Importando a biblioteca e carregando o dataframe.

**2º Passo:**

Após carregarmos o dataframe, separamos a coluna *Date / Time* em duas novas colunas, *Sight\_Date* e *Sight\_Time* como o demonstrado na **Fig 2.**

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 2. Separando a coluna Date / Time.

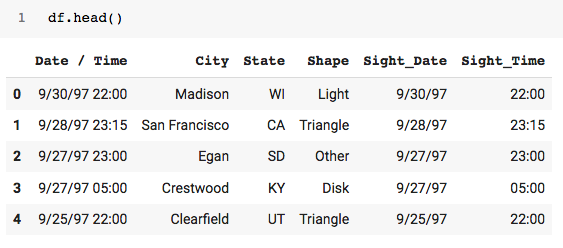


Fig 3. Dataframe resultante.

Para realizar a separação se utiliza a função *str.split* do **pandas** que separa os dados em duas novas colunas que são armazenadas na variável datetime e depois são inseridas no dataframe principal, tendo como resultado o demonstrado na **Fig 4**.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 4. Resultado da separação da coluna Date / Time.

**3º Passo:**

O terceiro passo é remover a coluna *Date / Time* do dataframe (**Fig 5)**, para isso, utilizamos a função *del* para deletar a coluna do dataframe**.**

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 5. Removendo a coluna Date / Time.

**4º Passo:**

Após removermos a coluna *Date / Time* do dataframe, criamos a coluna com o dia da semana (**Fig 6)**, para realizar esse passo, primeiro criamos um dicionário com os dias da semana, depois usamos a função *to\_datetime* para converter a coluna *Sight\_Date* de string para o tipo datetime e a função *dt.dayofweek* para identificar o dia da semana, que é retornada na forma de índice (0 - segunda-feira, 1 - terça-feira etc.), por fim usamos a função *map* para substituir o índice pelo dia da semana, tendo como resultado o dataframe na **Fig 7**.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 6. Criando coluna com dias da semana.

Uma imagem contendo parede

Descrição gerada automaticamente

Fig 7. Dataframe resultante da adição da coluna com dias da semana.

**5º Passo:**

Da mesma forma que separamos a coluna *Date / Time*, nessa etapa utilizamos a função *str.split* (**Fig 8)** para separar a data em dia e mês, após isso adicionamos as colunas *Sight\_Day* e *Sight\_Month* ao nosso dataframe principal (**Fig 9**).

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 8. Criando as colunas com dia e mês dos relatos.



Fig 8. Resultado da criação das colunas com dia e mês dos relatos.

**6º Passo:**

Por fim, convertemos o dataframe em um arquivo CSV com o uso da função *to\_csv* do **pandas**.



Fig 8. Criando arquivo CSV.

**Github**:<https://github.com/dan-candeira/Nuforc-Project/tree/master/5.7-Preparacao>