**Relatório de exploração dos dados coletados do site *nuforc.org***

Aluno: Daniel Barros Candeira

1. **Introdução**

O relatório tem por objetivo demonstrar as etapas percorridas para visualizar os 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**

Após a realização da exploração dos dados, nessa etapa será realizada a visualização de dados, tendo como atividades, criar gráfico um gráfico de barras agrupadas e um gráfico de barras empilhadas contendo os 4 estados e formatos mais populares, criar um mapa que indique a quantidade de relatos separados por cidades estadunidenses, e um mapa do estado da Califórnia indicando as cidades que possuem 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 bibliotecas **Requests, BeautifulSoup** e **Pandas**, **Gmaps**, **Codezips**, **Matplotlib** e **Seaborn** que servem respectivamente para realizar requisições HTTP[[1]](#footnote-1) que retornam o código HTML que compõe a página, possibilitar a navegação dentro das tags HTML, e para lidar com dados de forma estatística e com base/arquivos de dados volumosos, plotar mapas geográficos, e os dois últimos são utilizados para plotar gráficos diversos.

**1º Passo:**

Antes da escrita dos códigos, é realizada a importação das bibliotecas na **Fig 1** e a importação a base de dados (arquivo CSV) que está hospedado no Github com a utilização da biblioteca **pandas** e o **pandasql** na **Fig 2**, como resultado da importação da base é retornado um dataframe que contém todas a linhas e colunas salvas no arquivo CSV, no final da linha de importação é utilizada a função ***drop*** a fim de remover uma coluna irrelevante do dataframe.

Uma imagem contendo interior

Descrição gerada automaticamente

Fig 1. Importando bibliotecas.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 2. Importando a base de dados.

Após a importação dos dados, utilizamos a função a biblioteca **requests** e **beautifulsoup** na **Fig 3,** que retorna todas as variáveis presentes no dataframe, assim como a quantidade de dados não nulos entre outras informações.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 3. Selecionando os estados dos EUA.

Depois selecionamos e renomeamos as colunas que contém o nome e a abreviação dos estados **Fig 4**, que retorna o dataframe **Fig 5.**

Uma imagem contendo objeto

Descrição gerada automaticamente

Fig 4. Renomeando e selecionando as colunas com nome e abreviação dos estados.

Uma imagem contendo captura de tela, parede

Descrição gerada automaticamente

Fig 5. DataFrame retornado pela Fig 4.

Após isso são realizadas quatro consultas SQL que criam respectivamente, um dataframe que contém apenas posts realizados nos EUA **Fig 6**, um dataframe com os formatos mais relatados **Fig 7**, um com os estados mais relatados **Fig 8** e o ultimo contém os 4 estados que possuem mais relatos e os 4 formatos mais populares nos Estados Unidos separados por esses estados **Fig 9**.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 6. Visualizando apenas os dados dos Estados Unidos.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 7. Visualizando os formatos mais relatados.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 8. Visualizando as cidades mais relatadas.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 9. Selecionando os 4 formatos mais relatados nos 4 estados mais relatados.

Após a realização das consultas SQL podemos criar os gráficos utilizando os dados retornados no dataframe **Fig 9** e **Fig 10**, que utilizam as bibliotecas **matplotlib** e **seaborn**, no gráfico de barras empilhada usamos a função **pivot** do pandas para rearranjar os dados e conseguir realizar o plot**.**

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 9. Plotando gráfico de barras agrupadas.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 10. Plotando gráfico de barras empilhadas.

**2º Passo:**

Para realizar a segunda etapa são utilizadas as bibliotecas **zipcodes** e **gmaps**, a biblioteca zipcodes retorna os dados de longitude latitude e outras informações sobre os estados e cidades, e a gmaps cria pot de mapas.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 11. Utilizando a biblioteca zipcodes para coletar os dados de todos os estados pertencentes aos Estados Unidos.

Selecionamos então as colunas *long, lat, city e state* do dataframe us\_zipcodes **Fig 12**, e excluímos as linhas que não possuem valores de longitude e latitude.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 12. Utilizando a biblioteca zipcodes para coletar os dados de todos os estados pertencentes aos Estados Unidos.

Após isso, realizamos uma consulta SQL para juntas os dataframes que contém os dados de latitude e longitude com os dados do dataset óvni com o objetivo de agrupar os relatos postados pelas cidades **Fig 13**, criamos e exportamos o mapa no formato html **Fig 14**.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 13. Agrupando os dados dos datasets.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 14. Coletando a latitude, a longitude e o número de relatos por estado.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 14. Criando e exportando mapa para formato html.

Uma imagem contendo árvore

Descrição gerada automaticamenteUma imagem contendo texto, mapa

Descrição gerada automaticamente

Fig 15. Mapas gerados.

**3º Passo:**

Da mesma forma que o passo anterior, realizamos uma consulta SQL, neste passo, contudo apenas selecionamos as cidades pertencentes ao estado da Califórnia.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 16. Selecionando os estados da Califórnia.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 17. Realizando consulta, que retorna latitude, longitude e relatos das cidades do estado da Califórnia.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 18. Criando e exportando mapa da Califórnia.

Uma imagem contendo texto, mapa

Descrição gerada automaticamenteUma imagem contendo ao ar livre

Descrição gerada automaticamente

Uma imagem contendo ao ar livre

Descrição gerada automaticamente

Fig 19. Mapas gerados.

**4º Passo:**

Onde na Califórnia está localizada a maior quantidade de visualizações de objetos voadores não identificados? E qual será a razão?

**Resposta:**

A maior quantidade de visualizações se concentra nas cidades de Los Angeles e San Diego. Em 1942, foram reportadas um grande número de objetos não identificados em Pearl Harbour em Los Angeles o que pode estar relacionado a esse maior número de visualizações, já em 2017 foi vazado um vídeo de OVNIs em San Diego, o que também pode estar associado a uma maior número de relatos na cidade.

**Github**:<https://github.com/dan-candeira/Nuforc-Project/tree/master/5.4-Exploracao>

**Bibliografia**

<https://jupyter-gmaps.readthedocs.io/en/latest/tutorial.html#basic-concepts>

<https://pypi.org/project/zipcodes/>

<https://www.vooo.pro/insights/pivot-table-em-pandas-explicado/>

<https://losangeleno.com/strange-days/battle-of-los-angeles/>

<https://pt.wikipedia.org/wiki/Incidente_do_USS_Nimitiz>

1. Protocolo de Transferência de Hipertexto. [↑](#footnote-ref-1)