

TECNOLOGIA EM SISTEMAS PARA INTERNET

**Kenia Duarte
Lucas dos Santos
Ana Amélia da Conceição**

**RELATÓRIO DE PRÁTICA INTEGRADA
DE
CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL**

Brasília - DF

03/10/2020

Sumário

1. Objetivos	3
2. Descrição do problema	4
3. Desenvolvimento	5
3.1 Código implementado	5
4. Considerações Finais	6
Referências	7

1. Objetivos

Este trabalho tem por objetivo investigar dados de vinte anos, entre os anos 1997 e 2017. Coletar esses dados entre os meses de setembro de 1997 e setembro de 2017. Os dados serão salvos em um DataFrame e exportado para em um arquivo .CSV.

2. Descrição do problema

Para esta coleta temos que descobrir fatos interessantes das observações de OVINIS, que são os objetos voadores não identificados. Objeto ou luz visto no céu, que não consegue ser identificado. Essa coleta será feita dentro de um período de vinte anos, de setembro de 1997 a agosto de 2017. O desafio é fazer a extração de dados sendo de forma tabular, porque para analisar bem os dados eles precisam se tornar tabelas.

Existe um tipo de OVNI ("Shape") mais popular, chamado de "Light". Essa categoria pode ser definida para o projeto. Será que não se trata apenas de um erro de visualização ou impressão errônea por parte do expectador? Esse tipo de insight será muito útil no momento de desenvolver a predição.

3. Desenvolvimento

Nesse passo do projeto será feita a limpeza de dados.

Sequência de tarefas

1. Carregar o seu arquivo OVNIS.csv em um dataframe;
2. Remover registros que tenham valores vazios (None, Unknown, ...) para City, State e Shape;
3. Manter somente os registros referentes aos 51 estados dos Estados Unidos;
4. Remover variáveis irrelevantes para a análise (Duration, Summary e Posted);
5. Manter somente os registros de Shapes mais populares (com mais de 1000 ocorrências);
6. Salvar o dataframe final em um arquivo CSV com o nome "df_OVNI_limpo".

3.1 Código implementado

4. Considerações Finais

Os dados que foram analisados são muito interessantes. Uma maneira melhor de analisar e compreender cada um dos dados foi a limpeza dos dados pesquisados.

Referências

Vooco – Insights. Data Science. Python. Gestão. **Guia para iniciantes de Web Scraping em Python (usando BeautifulSoup).**

<https://www.vooco.pro/insights/guia-para-iniciantes-de-web-scraping-em-python-usando-beautifulsoup/> - acessado em 11/09/2020.

5.7 - Limpeza dos dados.

https://canvas.instructure.com/courses/1839379/assignments/13790564?module_item_id=28164089