UFO SIGHTINGS

AVISTAMENTOS DE OBJETOS VOADORES NÃO IDENTIFICADOS (OVNIS) DE 1969-2019

ANA ELISA E RODRIGO

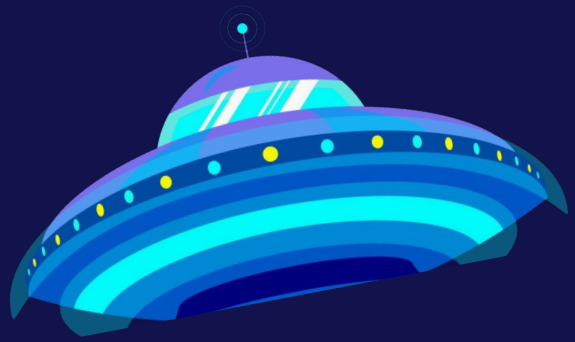
ORGANIZAÇÃO

01	ESCOLHA DA BASE DE DADOS	• Bases de dados confiáveis recomendadas:
		  ○ Our World In Data
		  ○ Sidra Ibge
		  ○ **Kaggle***

02	ESCOLHA DO TEMA	• Data science salaries 2023 • Pharmaceutical prices in pakistan • Aviation accidents aviation • **UFO sightings ***
03	REUNIÃO - INICIAL	• Estudo das Plataformas • Definição de Plataformas/Linguagens para análise • Limpeza de Dados
04	REUNIÃO - ANÁLISE	• Definição das perguntas com base na análise dos dados • Criação de Gráficos
05	REUNIÃO - AJUSTES FINAIS	• Apresentação

BASE DE DADOS

- 65.000 relatórios de avistamentos de OVNIs;
- Filtro US;



1. ‘datetime’ - Ano/Mês/Dia/Horário
2. ‘city’ - Cidade do US
3. ‘state’ - Estado do US
4. ‘shape’ - Formato do OVNI
5. ‘duration (seconds)’ - Duração do avistamento em segundos
6. ‘duration (hours/min)’ - Duração do avistamento em horas e minutos
7. ‘summary’ (editado como ‘comments’) - Descrição do fato
8. ‘latitude’ - Latitude
9. ‘longitude’ - Longitude
10. ‘dia_da_semana’ - Dia da semana do avistamento

LIMPEZA DE DADOS

python

```python
import pandas as pd

# Ler o arquivo UFO.csv e relacioná-lo ao dataframe 'df'
df = pd.read_csv(r'C:\Users\Usuário\Desktop\ideais\arquivos\UFO.csv')

# Definir os valores específicos para cada coluna
df['datetime'] = pd.to_datetime(df['datetime'], errors='coerce')
df['city'] = df['city'].str.capitalize()
df['state'].fillna('Unknown', inplace=True)
df['country'] = df['country'].str.upper()
df['shape'] = df['shape'].str.strip().str.capitalize()
df['duration (seconds)'] = pd.to_numeric(df['duration (seconds)'], errors='coerce')
df['comments'].fillna('No comments', inplace=True)
df['date posted'] = pd.to_datetime(df['date posted'], format='%m/%d/%Y', errors='coerce')
df['latitude'] = pd.to_numeric(df['latitude'], errors='coerce')
df['longitude '] = pd.to_numeric(df['longitude '], errors='coerce')

# Devido ao grande volume de dados e inconsistência nos dados que não sejam no território dos EUA, foi necessário filtrar tais dados para apresentar apenas os dados dos Estados Unidos
df = df[df['country'] == 'US'].copy()
```

LIMPEZA DE DADOS

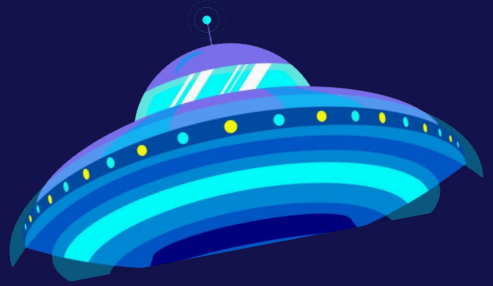
python

```
21  # Ao calcular a média com relação à duração, foi visto uma variação gigantesca com relação aos 5% dos maiores valores, representando um aumento de mais de 1500%, então foram retirados tais dados para obter uma média mais confiável
22  cutoff = df['duration (seconds)'].quantile(0.95)
23
24  # Definir a coluna apenas com os valores sem os 5% maiores
25  df = df[df['duration (seconds)'] <= cutoff]
26
27  # Converte a coluna duration seconds para int, pois o powerbi detecta o .0 (float) como um zero adicional, então é enviado como int
28  df['duration (seconds)'] = df['duration (seconds)'].astype(int)
29
30  # Definir o formato da coluna datetime para ano mês dia - hora minuto segundo, isso se viu necessário para um reconhecimento melhor para a próxima ação
31  df['datetime'] = pd.to_datetime(df['datetime'], format='%Y-%m-%d %H:%M:%S')
32
33  # Necessário definir o formato dos horários para ter menor chance de erros na hora de separar as colunas datetime em data e tempo, para obter analises de tempo e data separadas
34  df['data'] = df['datetime'].dt.date
35  df['hora'] = df['datetime'].dt.time
36
```

LIMPEZA DE DADOS


```
33  # Necessário definir o formato dos horários para ter menor chance de erros na hora de separar as colunas datetime em
    data e tempo, para obter analises de tempo e data separadas
34  df['data'] = df['datetime'].dt.date
35  df['hora'] = df['datetime'].dt.time
36
37  # Criar uma coluna com o dia da semana a partir da coluna datetime
38  df['dia_da_semana'] = df['datetime'].dt.strftime('%A')
39
40  # Criar outro arquivo com todas as informações atualizadas
41  df.to_csv(r'C:\Users\Usuário\Desktop\ideais\arquivos\UFO_PowerBi.csv', index=False)
```

ANÁLISE DE DADOS



Power BI

pandas

seaborn

python

matplotlib

ANÁLISE DE DADOS

1. Quais dias da semana possuem mais relatos de aparecimentos de OVNIs?

2. Qual período (manhã,tarde,noite,madrugada) possui mais relatos de aparecimentos de OVNIs?

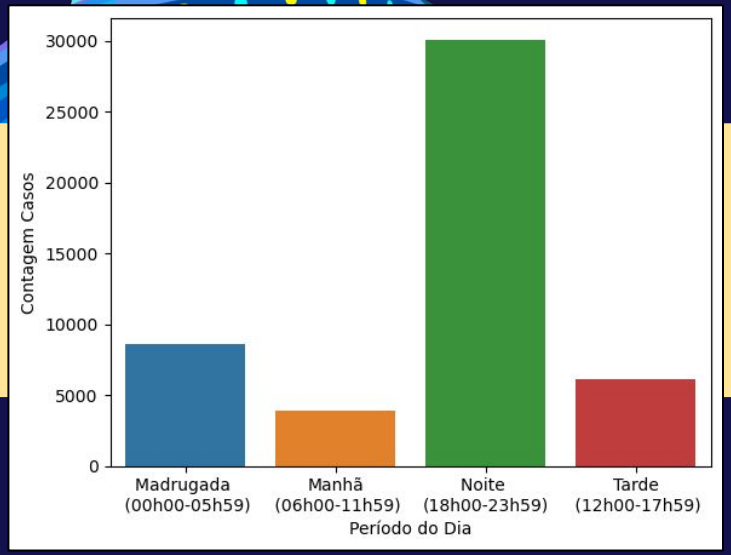
3. Qual hora média das ocorrências?

4. Qual a mediana da duração (em segundos) dos OVNIs relatados?

5. Quais cidades têm mais ocorrências?

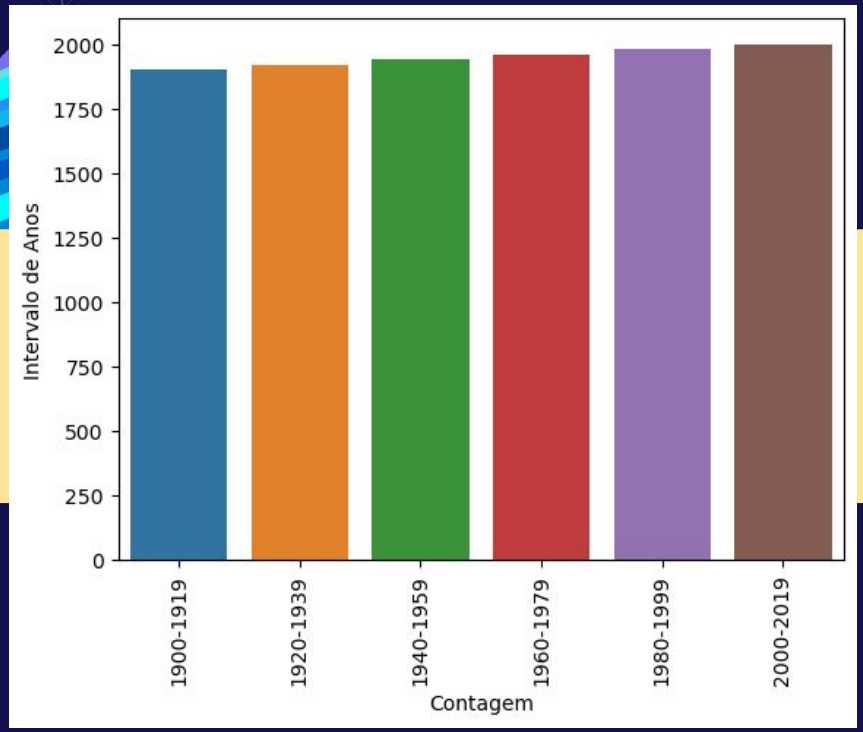
6. Qual o formato de maior ocorrência?

ANÁLISE DE DADOS



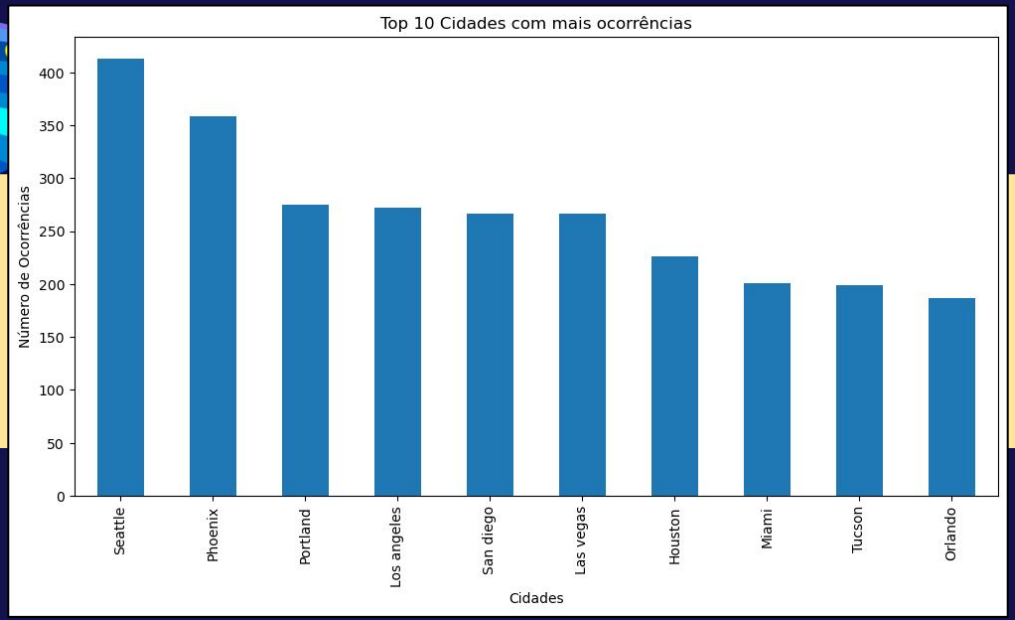
- **Padronização coluna dia/mês/ano/horário**
- **Separação Horários Definidos**
- **Contagem de casos**

ANÁLISE DE DADOS

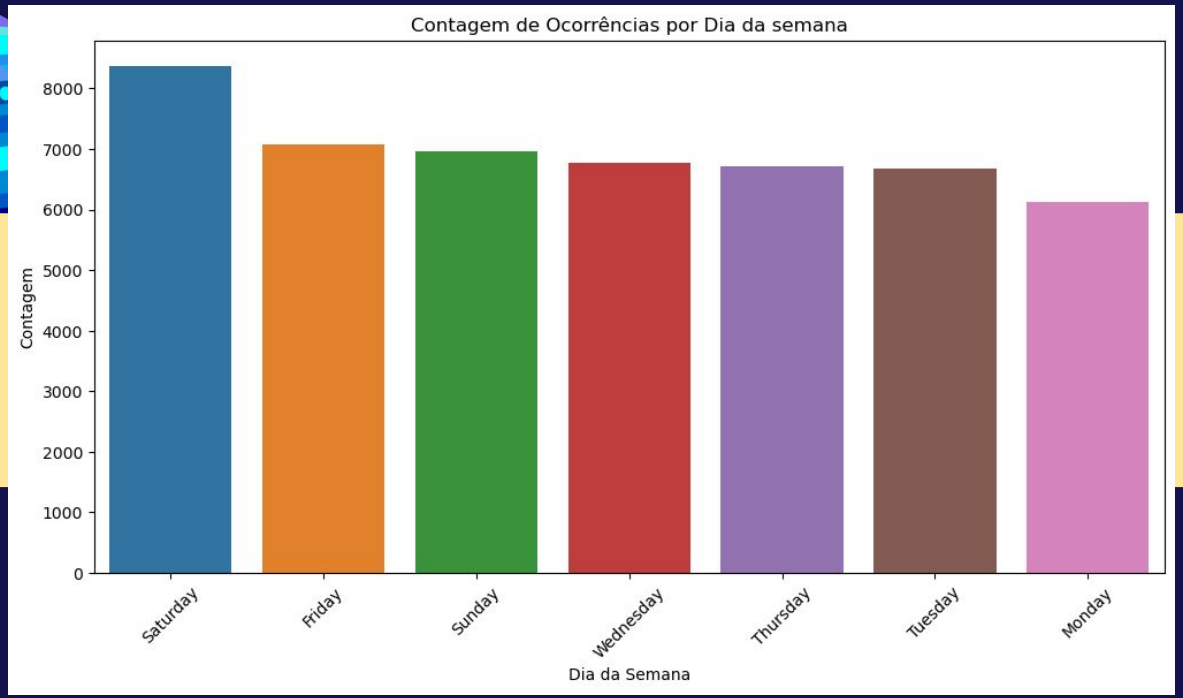


- **20-20 anos**
- **Contagem de casos de aparecimento nos anos**

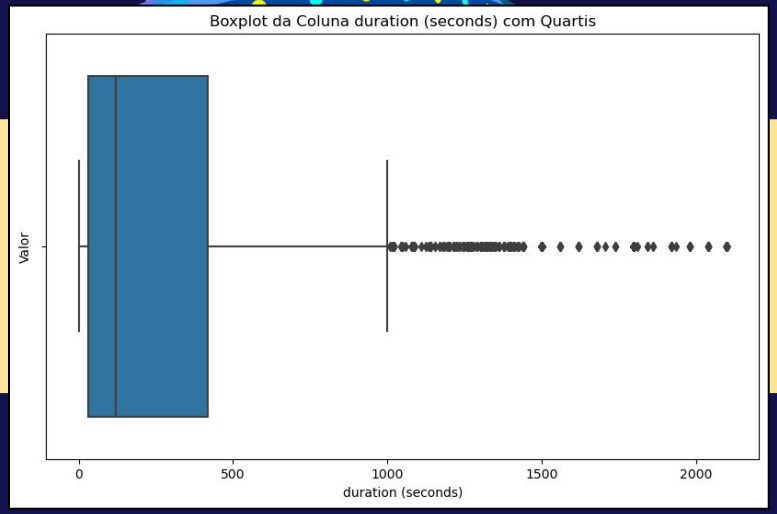
ANÁLISE DE DADOS

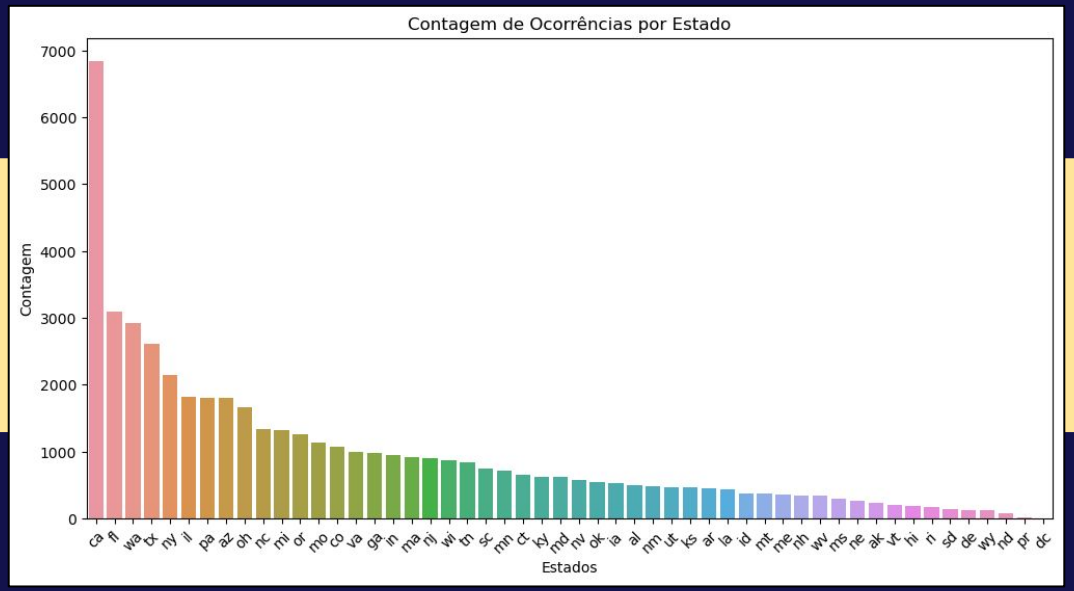


ANÁLISE DE DADOS



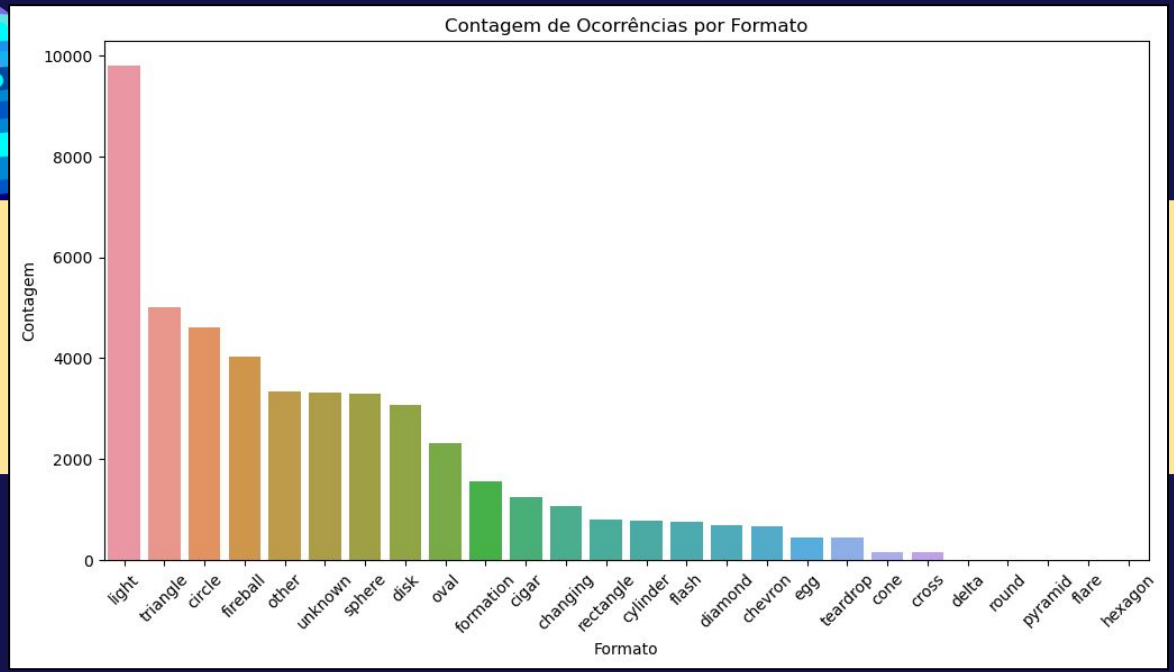
ANÁLISE DE DADOS



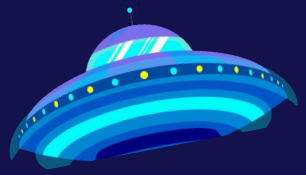


*Sem remoção de outliers

ANÁLISE DE DADOS



ANÁLISE DE DADOS



UFO Sightings Analysis

Total 65.000 lines of data

Light

Formato com Maior Incidencia

Seattle

Cidade com Maior Incidencia

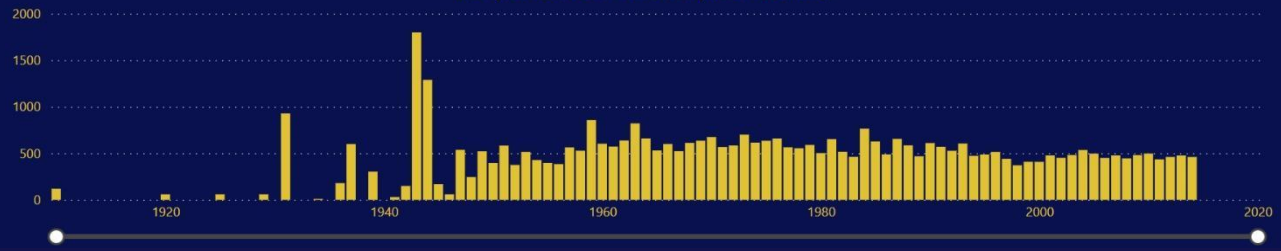
22:00

Hora Média

Saturday

Dia com Maior Incidencia

Duração dos avistamentos ao passar dos anos



Obrigado