

Projeto de prática Integrada de ciência de dados, inteligência artificial e machine learning



Relatório de avistamento de Objetos Voadores Não Identificados.

Sprint 2 → Acréscimo de Variáveis

Curso: Tecnologia em sistemas para internet

Estudantes:

Brenda Lopes Miranda Teixeira
Mateus Gomes da Silva Fonteles
Rickson Queiroz Marques de Souza
Samuel Araújo Lopes

Professores

Fábio Henrique
Diego Queiroz
Ana Régia

Brasília, agosto de 2021

Sumário

1. Objetivos	3
2. Descrição do problema	4
3. Desenvolvimento	5
3.1. Código implementado	6
4. Considerações Finais	8
Referências	9

1. Objetivos

Nesta etapa do trabalho, com nosso arquivo `OVNIS_limpo.csv`, será executada a seguinte sequência de tarefas:

1. Dividir o conteúdo da coluna `Date/Time` em duas novas colunas
2. Deletar a coluna `Date/Time`
3. Dividir a tabela `Date` para extrair colunas com os dias da semana, mês e dias.
4. Salvar o *dataframe* resultante em um arquivo de nome `'df_OVNI_preparado.csv'`

2. Descrição do problema

Após havermos procedido com a limpeza do nosso dataset, vamos dividir as colunas de data e hora, para melhor realizar operações com as mesmas, assim como separar a data em diversos níveis de granularidades, que permitirão realizar análises mais detalhadas dos relatos, podendo separá-los por dias da semana ou do mês. por exemplo, e saber em que dias da semana ou períodos do mês os avistamentos são mais frequentes.

3. Desenvolvimento

Este trabalho está sendo desenvolvido usando um Script Python por ser uma linguagem orientada a objetos é bastante maleável, o grupo está utilizando a plataforma Google Colaboratory, assim todos podem modificar e acrescentar o código quando necessário.

3.1. Código implementado

- Importação de bibliotecas e carregamento de arquivo:
- Para esta etapa do projeto somente será necessária a importação do Pandas.

```
#Importando Bibliotecas
import pandas as pd

#Carregando o arquivo csv
df_OVNI_limpo = pd.read_csv('df_OVNI_limpo.csv')
df_OVNI_limpo
```

- Preparação da tabela:

```
#Gerando a coluna Date/time para o formato datetime
df_OVNI_limpo['Date / Time'] = pd.to_datetime(df_OVNIS_limpo['Date /
Time'])
df_OVNI_limpo

#Chamando o tipo
df_OVNI_limpo.dtypes
```

- Criando as colunas:

```
#Dia
df_OVNI_limpo['Sight_day'] =df_OVNI_limpo['Date / Time'].dt.day
df_OVNI_limpo

#Mes
df_OVNI_limpo['Sight_month'] =df_OVNI_limpo['Date / Time'].dt.month
df_OVNI_limpo

#Tempo
df_OVNI_limpo['Sight_time'] =df_OVNI_limpo['Date / Time'].dt.time
df_OVNI_limpo

#Data
df_OVNI_limpo['Sight_date'] =df_OVNI_limpo['Date / Time'].dt.date
df_OVNI_limpo

#Fim de semana
```

```

df_OVNI_limpo['Sight_weekday'] =df_OVNI_limpo['Date / Time'].dt.weekday
df_OVNI_limpo

#Dias da semana
    dayOfWeek={0:'Segunda-feira',    1:'Terça-feira',    2:'Quarta-feira',
3:'Quinta-feira', 4:'Sexta-feira', 5:'Sábado', 6:'Domingo'}
df_OVNI_limpo['Sight_weekday']          =df_OVNI_limpo['Date /
Time'].dt.dayofweek.map(dayOfWeek)
df_OVNI_limpo

df_OVNI_limpo['Sight_weekday'].value_counts()
df_OVNI_limpo.drop(['Date / Time'],axis=1,inplace=True)
df_OVNI_limpo

```

- Armazenando as modificações no arquivo 'df_OVNI_preparado.csv'.

```

#Armazenando no novo CSV 'df_OVNIS_preparado'

df_OVNI_limpo.to_csv('df_OVNI_preparado.csv',index=False)
df_OVNI_preparado= pd.read_csv('df_OVNI_preparado.csv')
df_OVNI_preparado

```

4. Considerações Finais

Partimos de uma coluna data com a qual não se podia trabalhar adequadamente ou realizar cálculos, por conter informações de mais de um tipo, para uma definição bem específica e segmentada do tempo dos avistamentos, com a possibilidade de aplicar agora mais filtros e observar os dados pelo ponto de vista do tempo de forma mais clara.

Referências

- Dividir data (dia, mês, ano) em novas colunas - DataFrame Pandas. 2021. Disponível em:
 - <https://pt.stackoverflow.com/questions/428236/dividir-data-dia-m%C3%AAs-ano-em-novas-colunas-dataframe-pandas>