

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



Relatório de avistamento de Objetos Voadores Não Identificados.

Sprint 1 → Coleta de dados

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, julho de 2021

Sumário

1. Objetivos	3
2. Descrição do problema	4
3. Desenvolvimento	5
3.1 Código implementado	6
4. Considerações Finais	9
Referências	10

1. Objetivos

O objetivo desta primeira sprint é:

- Extrair dados sobre avistamento de objetos voadores não identificados, usando linguagem de programação python.

2. Descrição do problema

São disponibilizados relatórios de Ovnis na The National UFO Reporting Center que é uma organização dos EUA (Estados Unidos da América) que investiga avistamento de objetos voadores não identificados que surgem no céu. Tais dados são disponibilizados desde da década 1974, quando foi criada. Segundo a Wikipedia, a The National UFO Reporting Center já identificou 90.000 OVNIS avistados durante esse período. A mesma disponibiliza os dados gráficos e estatísticas para quaisquer pessoas que se interessem possam ver. Além disso, a organização fornece um número telefônico para que pessoas que avistarem quaisquer objeto voador, possam ligar diretamente ao centro e avisar. A partir do conhecimento da Nuforc e do que ela disponibiliza.

3. Desenvolvimento

Como já informado nos objetivos desta sprint, a linguagem utilizada pelo grupo é Python por ter várias bibliotecas, esta linguagem auxilia o desenvolvedor a fazer requisições e retornar o status code, a biblioteca utilizada é a Pandas.

3.1 Código implementado

- 3.1.1 - Ano de 1997

Para começar foi realizado os imports das bibliotecas e extraído dados do ano de 1997:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
td_data = list()
th_data = list()
#1997
print('1997')
print('Setembro')
URL = f'http://www.nuforc.org/webreports/ndxe199709.html'
headers={
    'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36'
}
page = requests.get(URL,headers=headers)
soup = BeautifulSoup(page.content, 'html.parser')
table = soup.findAll('table')[0]
tr = table.findAll(['tr'])
try:
    for cell in tr:
        th = cell.find_all('th')
        th_data.append([col.text.strip('\n') for col in th])
        td = cell.find_all('td')
        row = [i.text.replace('\n', '') for i in td]
        td_data.append(row)
finally:
    print('Completado')
```

- 3.1.2 - De 1998 a 1999

Na segunda parte do código pode-se ver que foi realizado um for i in range para percorrer e atualizar as listas.

```
#1998 de 1999 entre mês de janeiro até Setembro
for i in range(8,10):
```

```

print('199{}'.format(i))
for p in range(1,10):
    print('Mês ',p)
    URL = f'http://www.nuforc.org/webreports/ndxe199{i}0{p}.html'
    headers={
        'User-Agent':'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36'
    }
    page = requests.get(URL,headers=headers)
    soup = BeautifulSoup(page.content,'html.parser')
    table = soup.findAll('table')[0]
    tr = table.findAll(['tr'])
    try:
        for cell in tr:
            th = cell.find_all('th')
            th_data.append([col.text.strip('\n') for col in th])
            td = cell.find_all('td')
            row = [i.text.replace('\n','') for i in td]
            td_data.append(row)
    finally:
        print('Completado')

```

- 3.1.3 - De 1997 a 1999 entre mês outubro até dezembro

Aqui, repetimos o que foi visto anteriormente. O código é praticamente o mesmo, o que muda é o valor que armazenamos nas variáveis para trazer o resultado especificado, que no caso foi entre os anos de 1997 a 1999 entre os meses de outubro a dezembro.

```

# 1997 de 1999, entre mês outubro até dezembro
for i in range(7,10):
    print('199{}'.format(i))
    for p in range(0,3):
        print(' 1{}'.format(p))
        URL = f'http://www.nuforc.org/webreports/ndxe199{i}1{p}.html'
        headers={
            'User-Agent':'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36'
        }
        page = requests.get(URL,headers=headers)

```

```

soup = BeautifulSoup(page.content, 'html.parser')
table = soup.findAll('table')[0]
tr = table.findAll(['tr'])
try:
    for cell in tr:
        td = cell.find_all('td')
        row = [i.text.replace('\n', '') for i in td]
        td_data.append(row)
finally:
    print('Completado')

```

- 3.1.4 - de 2000 a 2009

Similarmente ao que ocorreu anteriormente, temos o mesmo código, atualizando somente os valores armazenados nas variáveis, desta vez para trazer os resultados entre os anos de 2000 a 2009

```

#2000 de 2009 entre mês de janeiro até setembro
for i in range(0,10):
    print('{}'.format(i))
    for p in range(1,10):
        print('{}'.format(p))
        URL = f'http://www.nuforc.org/webreports/ndxe200{i}0{p}.html'
        headers={
            'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.103
Safari/537.36'
        }
        page = requests.get(URL, headers=headers)
        soup = BeautifulSoup(page.content, 'html.parser')
        table = soup.findAll('table')[0]
        tr = table.findAll(['tr'])
        try:
            for cell in tr:
                td = cell.find_all('td')
                row = [i.text.replace('\n', '') for i in td]
                td_data.append(row)
        finally:
            print('Completado')

```

- 3.1.5 - de 2010 a 2016

Em continuação, temos o mesmo código, porém, com especificações nas variáveis para os anos de 2010 a 2016 entre os meses de janeiro a setembro

```
#2010 de 2016, entre o mês de janeiro até setembro
for i in range(0,7):
    print('201{}'.format(i))
    for p in range(1,10):
        print('{}'.format(p))
        URL = f'http://www.nuforc.org/webreports/ndxe201{i}0{p}.html'
        headers={
            'User-Agent':'Mozilla/5.0 (X11; Linux x86_64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.103
Safari/537.36'
        }
        page = requests.get(URL,headers=headers)
        soup = BeautifulSoup(page.content, 'html.parser')
        table = soup.findAll('table')[0]
        tr = table.findAll(['tr'])
        try:
            for cell in tr:
                td = cell.find_all('td')
                row = [i.text.replace('\n','') for i in td]
                td_data.append(row)
        finally:
            print('Completado')
```

- 3.1.6 - 2017

Já aqui, as variáveis indicam o ano de 2017 no mês de agosto. O print também mostra a parte final do código, onde ele deve imprimir, ou seja, exibir uma saída (no caso o nosso resultado) de acordo com os valores indicados nas variáveis.

```
#2010 até 2016, entre o mês de outubro até dezembro
for i in range(0,7):
    print('201{}'.format(i))
    for p in range(0,3):
        print('1{}'.format(p))
        URL = f'http://www.nuforc.org/webreports/ndxe201{i}1{p}.html'
        headers={
            'User-Agent':'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36'
        }
        page = requests.get(URL,headers=headers)
        soup = BeautifulSoup(page.content, 'html.parser')
        table = soup.findAll('table')[0]
        tr = table.findAll(['tr'])
        try:
            for cell in tr:
                td = cell.find_all('td')
                row = [i.text.replace('\n','') for i in td]
                td_data.append(row)
        finally:
            print('Completado')
```

```

page = requests.get(URL,headers=headers)
soup = BeautifulSoup(page.content, 'html.parser')
table = soup.findAll('table')[0]
tr = table.findAll(['tr'])
try:
    for cell in tr:
        td = cell.find_all('td')
        row = [i.text.replace('\n','') for i in td]
        td_data.append(row)
finally:
    print('Completado')
#2017 mês de agosto
print('Ano 2017')
for p in range(1,9):
    print(' ',p)
    URL = f'http://www.nuforc.org/webreports/ndxe20170{p}.html'
    headers={
        'User-Agent':'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36'
    }
    page = requests.get(URL,headers=headers)
    soup = BeautifulSoup(page.content, 'html.parser')
    table = soup.findAll('table')[0]
    tr = table.findAll(['tr'])
    try:
        for cell in tr:
            td = cell.find_all('td')
            row = [i.text.replace('\n','') for i in td]
            td_data.append(row)
    finally:
        print('Completado')
del th_data[1:-1]
th_data.pop(-1)
th_data
df = pd.DataFrame(td_data)
df1 = df.dropna()
column = []
for i in th_data:
    column = i
df1.columns = column
df1

```

```
df1.to_csv('OVNIS.csv',index=False)
total = pd.read_csv('OVNIS.csv')
total
```

4. Considerações Finais

Ao concluir o trabalho, conseguimos identificar o quanto a programação pode ser útil, com diversas finalidades e aplicações. Aqui excedemos o básico, desenvolvendo um código capaz de acessar dados de um site externo e armazenar estas informações em um arquivo .csv de nossa autoria, tudo de forma automatizada, possibilitando informar parâmetros personalizados para uma busca e resultado mais específicos.

Referências

- The national UFO reporting center. Dedicated to the Collection and Dissemination of Objective UFO Data. Nuforc, EUA. Disponível em: <<http://www.nuforc.org>>. Acesso em 27 de Julho de 2021.