



INSTITUTO FEDERAL DE BRASÍLIA

CAMPUS BRASÍLIA

CURSO SUPERIOR DE TECNOLOGIA EM SISTEMAS PARA INTERNET

Gabriel Santos

Marcos Vasconcellos

Pedro L. B. G. de Araujo

**RELATÓRIO DE PRÁTICA INTEGRADA DE CIÊNCIA DE DADOS E INTELIGÊNCIA
ARTIFICIAL: TAREFA 5.4 – EXPLORAÇÃO DE DADOS COM SQL**

Brasília

2020

SUMÁRIO

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

1. Objetivos

Partindo pelo ponto da grande importância dos campos da Inteligência Artificial e da Ciência de Dados na área da Tecnologia da Informação, possuir no mínimo um breve conhecimento e contato com tais tecnologias, mostra-se imprescindível para um verdadeiro profissional de TI.

Sobre o campo da Inteligência Artificial, de acordo com Stuart Russell; Peter Norvig (2013), a IA define-se por um comportamento que se relaciona a processos de pensamento e raciocínio. E para seu estudo e trabalho, devemos considerar quatro definições principais sobre a Inteligência Artificial, sendo estas: pensar como um humano; agir como humano; pensar racionalmente, e agir racionalmente.

Passando para a área da Ciência de Dados, Joel Grus, (2016), diz que a quantidade de dados espalhadas por todos os lados em nosso cotidiano, é extratossférica. E um cientista de dados é alguém que trabalha com dados dos mais diversos tipos e tamanhos, geralmente desorganizados, extraindo-os com o objetivo de transformá-los em conhecimento.

Apresentada uma abordagem geral sobre conceitos importantes relativos à ambas às áreas temas deste projeto, o objetivo deste documento é introduzir gradativamente o leitor, ao desafio a nós proposto pelo Instituto Federal de Brasília. Tal desafio, consiste na aplicação de diversas tarefas menores, gradativamente aplicadas ao longo do projeto, visando um objetivo final. Para alcançarmos este objetivo, utilizaremos conhecimentos e ferramentas ligados à IA e Ciência de Dados, e metodologias ágéis como o Scrum.

2. Descrição do Problema

Para a primeira etapa do projeto, precisaremos realizar a coleta de dados relacionados à OVNIS, de um período total de vinte anos. Tal coleta terá como base o *website* <http://www.nuforc.org/Nuforc>, e o nosso objetivo, é abordar fatos interessantes extraídos desses dados. Além disso, é importante ressaltar que todos os dados utilizados, restringem-se à estados dos Estados Unidos.

A extração e exibição dos dados coletados, deverá ser feita de forma tabular, gerados em um arquivo .CSV, facilitando a leitura e análise dos mesmos.

3. Desenvolvimento

A nossa atuação na resolução do problema proposto, inicia-se com a utilização da linguagem Python e algumas de suas bibliotecas. Segundo OLIVEIRA, Marcos (2019) “As bibliotecas e pacotes Python são um conjunto de módulos e funções úteis que minimizam o uso de código em nossa vida cotidiana. Essas bibliotecas e pacotes destinam-se a uma variedade de soluções modernas.”.

Algumas bibliotecas específicas foram utilizadas como base para esta etapa inicial do projeto. Estas, respectivamente denominam-se e possuem como objetivo:

- **Numpy**: É o mais popular pacote de processamento de Arrays do Python. Não apenas oferece arrays e matrizes, também as gerencia.
- **Requests**: Uma grande biblioteca HTTP que funciona sob a licença Apache 2.0. Seu objetivo é tornar as solicitações HTTP mais responsáveis e fáceis de usar.
- **BeautifulSoup**: Consiste em uma biblioteca para extração e análise de documentos HTML e XML.
- **Pandas**: Uma das diversas bibliotecas Python voltadas para a área da Ciência de Dados. É um pacote de software Python voltado para a manipulação de estruturas de dados de forma intuitiva.

3.1 Código implementado

Abaixo, serão apresentados as estruturas de códigos com a utilização das bibliotecas citadas, que nos possibilitaram obter a extração e realizar a análise dos dados relacionados à OVNIS.

A Figura 1, exibe à seguir a importação de todas as bibliotecas utilizadas nesta etapa do projeto.

```
from bs4 import BeautifulSoup
import requests
import pandas as pd
import numpy as np
```

Figura 1 - Importação das Bibliotecas Python Utilizadas

A imagem a seguir, conta com o trecho do código responsável pelo armazenamento das variáveis de acordo com os períodos especificados.

```
years_months = [199709,199710,199711,199712,
199801,199802,199803,199804,199805,199806,199807,199808,199809,199810,199811,199812,
199901,199902,199903,199904,199905,199906,199907,199908,199909,199910,199911,199912,
200001,200002,200003,200004,200005,200006,200007,200008,200009,200010,200011,200012,
200101,200102,200103,200104,200105,200106,200107,200108,200109,200110,200111,200112,
200201,200202,200203,200204,200205,200206,200207,200208,200209,200210,200211,200212,
200301,200302,200303,200304,200305,200306,200307,200308,200309,200310,200311,200312,
200401,200402,200403,200404,200405,200406,200407,200408,200409,200410,200411,200412,
200501,200502,200503,200504,200505,200506,200507,200508,200509,200510,200511,200512,
200601,200602,200603,200604,200605,200606,200607,200608,200609,200610,200611,200612,
200701,200702,200703,200704,200705,200706,200707,200708,200709,200710,200711,200712,
200801,200802,200803,200804,200805,200806,200807,200808,200809,200810,200811,200812,
200901,200902,200903,200904,200905,200906,200907,200908,200909,200910,200911,200912,
201001,201002,201003,201004,201005,201006,201007,201008,201009,201010,201011,201012,
201101,201102,201103,201104,201105,201106,201107,201108,201109,201110,201111,201112,
201201,201202,201203,201204,201205,201206,201207,201208,201209,201210,201211,201212,
201301,201302,201303,201304,201305,201306,201307,201308,201309,201310,201311,201312,
201401,201402,201403,201404,201405,201406,201407,201408,201409,201410,201411,201412,
201501,201502,201503,201504,201505,201506,201507,201508,201509,201510,201511,201512,
201601,201602,201603,201604,201605,201606,201607,201608,201609,201610,201611,201612,
201701,201702,201703,201704,201705,201706,201707,201708]
```

Figura 2 - Armazenamento das Variáveis de acordo com Períodos Específicos

Abaixo, na Figura 3, temos o código responsável por combinar as urls com os períodos.

```
str = 'http://www.nuforc.org/webreports/ndxe{}.html'

urls = []
for year_month in years_months:
    urls.append(str.format(year_month))
```

Figura 3 – Combinação das urls com Períodos

Logo após, para facilitar a visualização e análise dos dados, as tabelas são adicionadas em uma lista.

```
dfo = []
for url in urls:
    df = pd.read_html(url, header = 0)
    dfo.append(df[0])
```

Figura 4 – Adicionando as Tabelas em uma Lista

Adiante, organizamos as tabelas em uma estrutura de DataFrame (matriz).

```
ovnis = []
ovnis = pd.DataFrame(dfo[0])
for ovni in dfo:
    if (ovni.equals(dfo[0])):
        print()
    else:
        ovnis = ovnis.append(ovni).reset_index(drop = True)
```

Figura 5 – Transformação das Tabelas em DataFrame

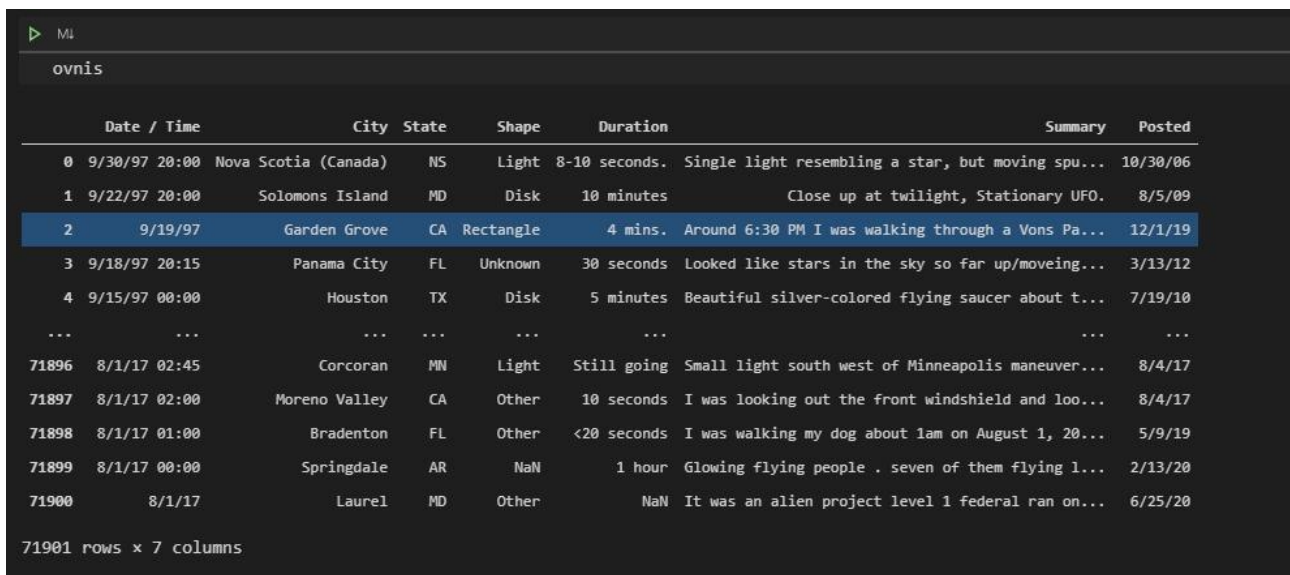
Por fim, o resultado da extração, organização e análise dos dados, foi enviada para um arquivo do tipo CSV, criado através do código abaixo.

```
ovnis.to_csv('OVNIS.csv', index=False)
```

Figura 6 – CSV com o Resultado da Extração e Análise dos Dados

Como resultado da primeira etapa do projeto, o arquivo CSV com a extração e estruturação de todos os dados coletados foi gerado. À seguir, temos alguns exemplos desse resultado. E como pode ser visto, a tabela separa suas informações em colunas, como por exemplo data e hora dos relatos, estado onde ocorreu, duração, descrição, entre outros.

Na Figura 7, temos a apresentação do resultado através da IDE. Já na Figura 8, temos o resultado final gerado sendo exibido através do arquivo final OVNIS.csv gerado, aberto no Excel.



The screenshot shows a Python IDE window titled 'ovnis' displaying a table with 7 columns: Date / Time, City, State, Shape, Duration, Summary, and Posted. The table contains 71901 rows of data. The first few rows are highlighted in blue.

	Date / Time	City	State	Shape	Duration	Summary	Posted
0	9/30/97 20:00	Nova Scotia (Canada)	NS	Light	8-10 seconds.	Single light resembling a star, but moving spu...	10/30/06
1	9/22/97 20:00	Solomons Island	MD	Disk	10 minutes	Close up at twilight, Stationary UFO.	8/5/09
2	9/19/97	Garden Grove	CA	Rectangle	4 mins.	Around 6:30 PM I was walking through a Vons Pa...	12/1/19
3	9/18/97 20:15	Panama City	FL	Unknown	30 seconds	Looked like stars in the sky so far up/moveing...	3/13/12
4	9/15/97 00:00	Houston	TX	Disk	5 minutes	Beautiful silver-colored flying saucer about t...	7/19/10
...
71896	8/1/17 02:45	Corcoran	MN	Light	Still going	Small light south west of Minneapolis maneuver...	8/4/17
71897	8/1/17 02:00	Moreno Valley	CA	Other	10 seconds	I was looking out the front windshield and loo...	8/4/17
71898	8/1/17 01:00	Bradenton	FL	Other	<20 seconds	I was walking my dog about 1am on August 1, 20...	5/9/19
71899	8/1/17 00:00	Springdale	AR	NaN	1 hour	Glowing flying people . seven of them flying l...	2/13/20
71900	8/1/17	Laurel	MD	Other	NaN	It was an alien project level 1 federal ran on...	6/25/20

71901 rows x 7 columns

Figura 7 – Exibição dos Dados Extraídos e Análises via IDE

90	4/15/98 20:00,Richmond,CA,Disk,10 minutes,"Disk over Richmond, CA",3/4/08	
91	4/15/98 13:00,San Antonio,TX,Cylinder,3 mintues,bright yellow lights over san antonio texas,1/10/09	
92	4/15/98 08:20,Baraboo,WI,Sphere,30 sec,Stationary silver sphere over town.,5/2/11	
93	4/15/98 07:45,Catawba,OH,Triangle,15 sec,My friend and I spotted a craft just before going to school.,11/28/07	
94	4/12/98 17:00,Level Cross,NC,Unknown,7 mins,Two rings and a sphere shape craft,2/14/08	
95	4/9/98 23:00,Golden Valley,MN,Cylinder,3 minutes,UFO sightings in Golden Valley and Robbinsdale,9/19/19	
96	4/5/98 11:50,Soquel,CA,Other,1 min 20sec,Greyish green wing shape travelling upthe coast Santa Cruz California,8/7/11	
97	4/4/98 23:00,Northern California,CA,Light,4 Minutes,The light made no sound then split into two lights.,7/16/06	
98	4/1/98 21:00,Henry County,OH,Triangle,2 hours,"Bright light, fireball, and triangular craft, lasted a few hours.",10/14/13	

Figura 8 – CSV com Exemplos de Exibição dos Dados Extraídos e Análises

4. Considerações Finais

Com a conclusão de todas as etapas de trabalho, o resultado obtido nos gerou um arquivo do tipo CSV, apresentando todas as informações extraídas do *website Nuforc.com*, em forma de tabela. De acordo com o que foi abordado até o momento, sabemos que os dados reunidos, estavam ligados à relatos ligados à aparições de OVNIS, com foco em estados dos Estados Unidos.

Para tanto, o uso da linguagem Python e suas bibliotecas foi essencial. Através da integração de ambos, foi possível gerar a tabela e organizar as informações coletadas com o uso de colunas, separando as diferentes informações por data e hora, tipo, descrição do ocorrido, entre outros. Com isso, é facilmente notada a importância do campo da Ciência de Dados, que cumpre exatamente com o que promete: coletar dados espalhados em grandes quantidades, e gerar informações úteis com estes dados. E isso, em uma visão generalizada, consegue mudar a forma com que o mundo e suas informações são vistas. Trazendo uma maior facilidade na visualização e entendimento dos mais diversos dados, assim como um maior valor para os mesmos.

Projetos completos e de real valor para o mercado como este que se inicia (que por enquanto se conclui), tendem apenas à agregar positivamente no conhecimento e experiência prática dos estudantes dos cursos de Tecnologia da Informação do Instituto Federal de Brasília. Tais conhecimentos práticos, servem tanto para dar uma base de como um profissional atua no mercado, com os mais diversos tipos de projetos e demandas, assim como para prepará-los para o tal mercado, quiçá para a montagem de um portfólio.

5. Referências

RUSSELL Stuart; NORVIG Peter, **Inteligência Artificial, 3ª edição**. Rio de Janeiro: Elsevier, 2013.

GRUS, Joel. ***Data Science do Zero: Primeiras Regras com Python***. Rio de Janeiro: Altabooks, 2016.

OLIVEIRA, Marcos. TERMINAL ROOT. **As 30 Melhores bibliotecas e pacotes Python para Iniciantes**. 2019. Disponível em: <https://terminalroot.com.br/2019/12/as-30-melhores-bibliotecas-e-pacotes-python-para-iniciantes.html>. Acesso em: 18 de Setembro de 2020.