

INTRODUÇÃO AO DATA SCIENCE COM PYTHON





OLÁ!

Caio Jordão Carvalho

Labrasoft

<https://github.com/cjlcarvalho>

DATA SCIENCE

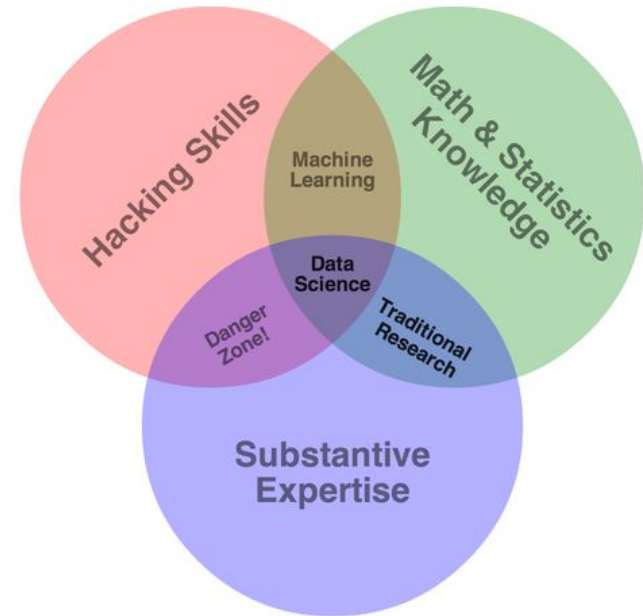
Entendendo alguns conceitos...

“

*Data Scientist: The Sexiest
Job of the 21st Century.*

Principais características

- Hacking skills
- Math & statistical knowledge
- Substantive expertise



Programação vs Ferramentas



E por que Python?

- Gratuito
- Simples de codificar
- Várias bibliotecas para DS fáceis de usar

Numpy

Numerical Python

Numpy

- Principal biblioteca em Python para trabalhar com programação científica.
- Escrita em C.
- Numpy Arrays!



E como usar?

```
>>> import numpy as np  
>>> np.array([1, 2, 3])  
array([1, 2, 3])  
>>> □
```

Pandas

Python Data Analysis Library

Que “Pandas” são esses?



Pandas

- Evolução do Numpy.
- Provê estruturas de fácil manipulação, além de ferramentas de análises de dados.
- Series!
- DataFrames!

Usando Series

```
>>> import pandas as pd
>>> df = pd.Series([3, 5, 0, -8], index=[1, 2, 3, 4])
>>> df
1      3
2      5
3      0
4     -8
dtype: int64
>>> df[1]
3
```

DataFrames!

```
>>> import pandas as pd
>>> data = pd.DataFrame([1, 2, 3, 4], index=['a', 'b', 'c', 'd'], columns=['y'])
>>> data
   y
a  1
b  2
c  3
d  4
>>> 
```

Coletando dados da web

Dados, dados e mais dados

Certo, onde conseguir esses dados?

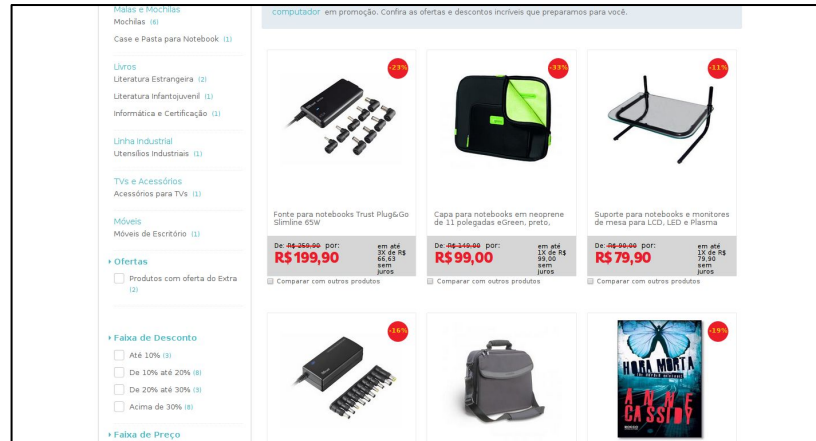
- Big Data
- Small Data
- Open Data

Crawlers, crawlers, crawlers!

- Web Crawling vs Web Scraping
- Crawlers + Scrapers
- Scrapy, Requests, BeautifulSoup...

Como um app de busca de ofertas atua?

- E se eu quisesse procurar um produto em um site de uma loja de departamentos? Qual é o raciocínio por trás da coleta dessas informações?



Reconheçam os padrões da web!

- Vamos procurar um produto! Hora de codificar um simples scraper!



PERGUNTAS?