

Python Open Data
Science Stack



Grupo de Usuários Python da Baixada Santista

O que é o Py013?

- É o grupo de usuários Python da baixada santista que tem como propósito compartilhar conhecimentos na linguagem Python na região.
- Criado em Janeiro de 2019.
- Estamos presentes no site da Python Brasil.

<https://python.org.br/comunidades-locais/>



Quem sou eu?

João Carlos Nascimento Santos

- Formado em Sistemas de informação - 2006 - Unisantos
- Pós-Graduação em Business Intelligence
- Trabalho com BI desde 2009 como consultor SAP BI.
- A Inteligência Artificial irá mudar a saúde!!



Contexto

Data Sets

Open data day

Python

Inteligência Artificial

Machine learning

Baixada Santista

Meetup

Jupyter notebook

Python Open Data Science Stack

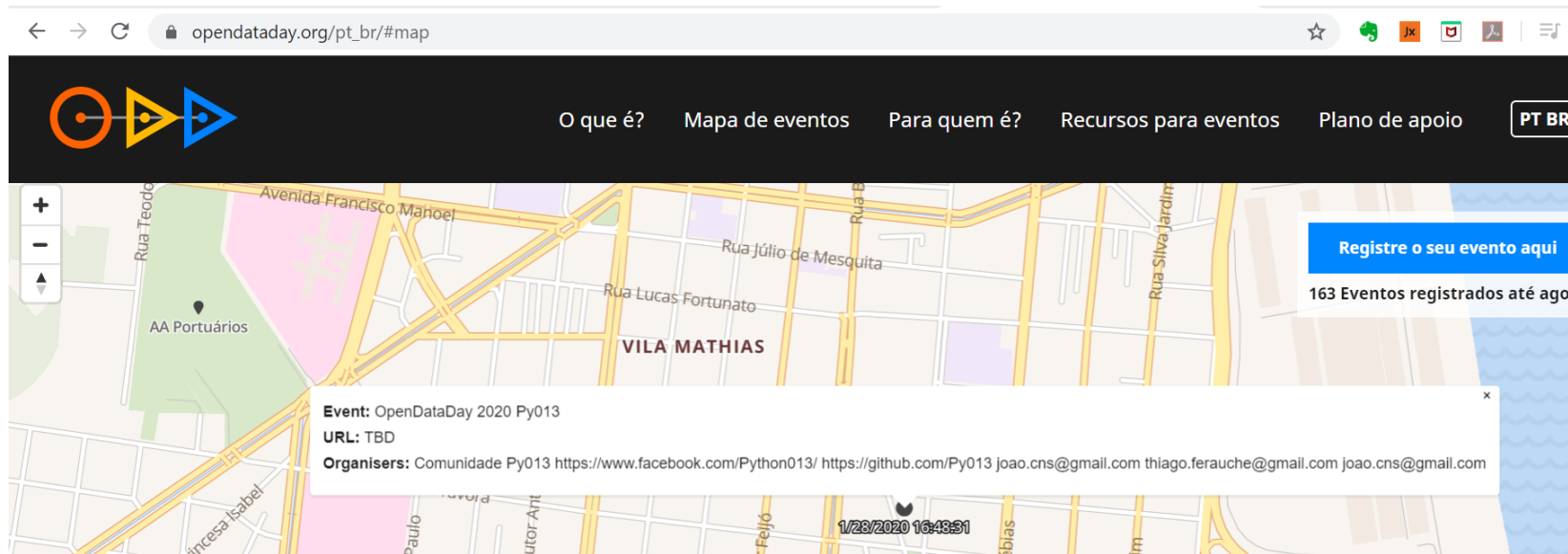
Análise Exploratória

Py013



Open Data Day

- O Dia dos Dados Abertos é uma celebração anual dos dados abertos em todo o mundo.
- É uma oportunidade para mostrar os benefícios dos dados abertos e encorajar a adoção de políticas de dados abertos no governo, empresas e na sociedade civil.
- Iniciamos na baixada santista em 2019.



Open Data Day – 07/03/2020



Meetup

- Contextualizar o open data day - 2020
- Organizar o conteúdo que apresentaremos no dia 07/03/2020.
- Encontro para nos reunir e promover a comunidade Py013!.



Python

- Porque python e open data day?.



KDNUGGETS ANNUAL POLL 2019

1.800 cientistas de dados

“Quais softwares ou linguagem você usou
para analisar dados nos últimos 12
meses?”



46,6%



python™ 65,8%.



python™

Linguagem de programação

Muito fácil fazer alterações nas análises
(1 linha de código)

Facilita análises colaborativas.
Garante reprodutibilidade dos resultados em novas
amostras.

Gratuito

Comunidade ativa de programadores

Linguagem de programação geral

Py013



Python Open Data Science Stack

- Pydata.org: Comunidade de apoio as bibliotecas de análise de dados do python.



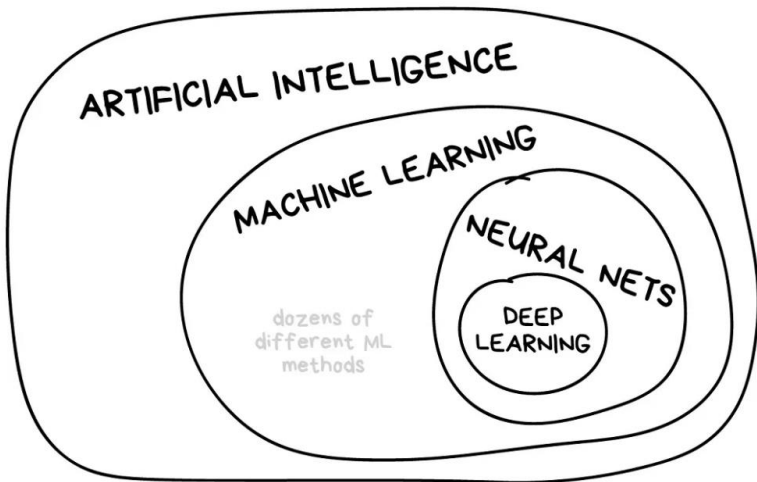
Py013

Inteligência Artificial

- Existem várias definições:

Possibilidade: “Capacidade de tornar a melhor decisão possível dada a informação disponível, com a capacidade de se adaptar a novas situações”.

Segundo essa definição, Inteligência é um problema de análise de dados.



Machine learning: Maquinas aprendendo sozinhas!

Tomada de decisão via identificação de padrões complexos nos dados.

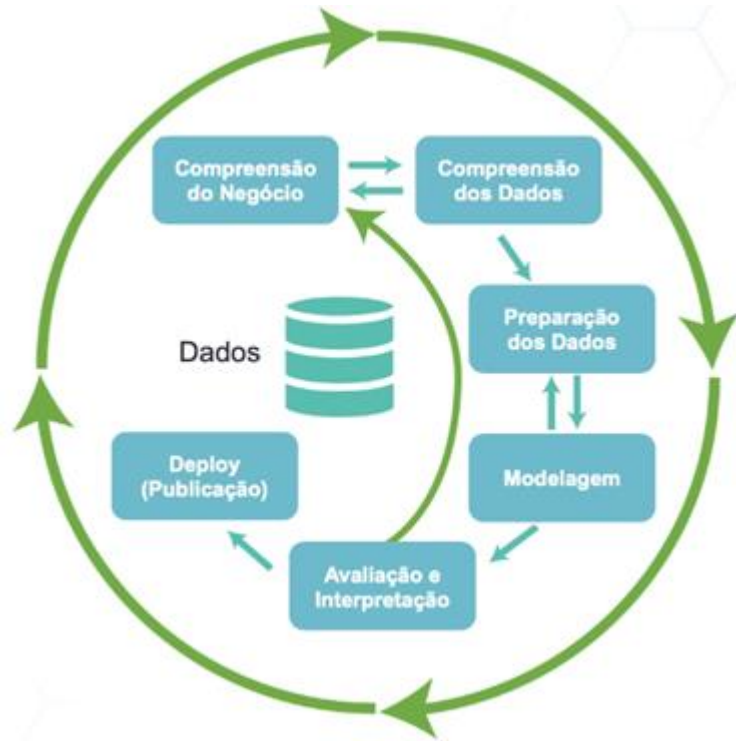


Inteligência Artificial

- O lab de IA e que estamos é uma grande iniciativa no nossa região!
- O open data day pode ser uma ótima oportunidade para explorarmos os dados abertos, usar o potencial de análise de dados das bibliotecas do python e assim possamos criar uma solução usando IA.



Analise Exploratória



Data Science Road Map

- Dentro de um processo de data science é fase onde conhecemos os dados, em busca de padrões e respostas para compreender o contexto desse dados.
- Iremos somente até essa fase!.
Compreensão dos dados.

DataSets

- São as base de dados onde faremos nossas análises exploratórias.
- Seguindo o open data day, são os repositórios públicos que iremos explorar.

<http://www.dados.gov.br/> - Porta Brasileiro de dados abertos

<http://www.portaltransparencia.gov.br/download-de-dados> - Portal das transparências

<https://data.humdata.org/> - Dados Humanitários

<https://data.worldbank.org/> - World Bank

http://dados.prefeitura.sp.gov.br/pt_PT/ - Dados Abertos da prefeitura de São Paulo

<https://index.okfn.org/> - Global Open Data Index

<http://abre.ai/aNtC> - Os 8 Princípios dos dados abertos



Jupyter Notebook

- Iniciativa open-source via web para criar e compartilhar códigos, visualizações e documentação, etc...
- Podemos fazer, limpeza nos dados, transformação, modelagem estatística, machine learning etc..



Referências:

Open Data Day

- <https://opendataday.org/>

Inteligência Artificial em saúde: o uso de machine learning

- https://www.youtube.com/playlist?list=PLAudUnJeNg4tvUFZ8tXQDoAkFAASQzOHm&disable_polymer=true
- **Machine Learning para todos, de forma simples e com exemplo**
<http://www.datageeks.com.br/machine-learning/>

