

Por onde começar a estudar Ciência de Dados?



Organização:

Grupo de Estudos em Data Science PyLadies-SP

Grupo de Estudos em Data Science

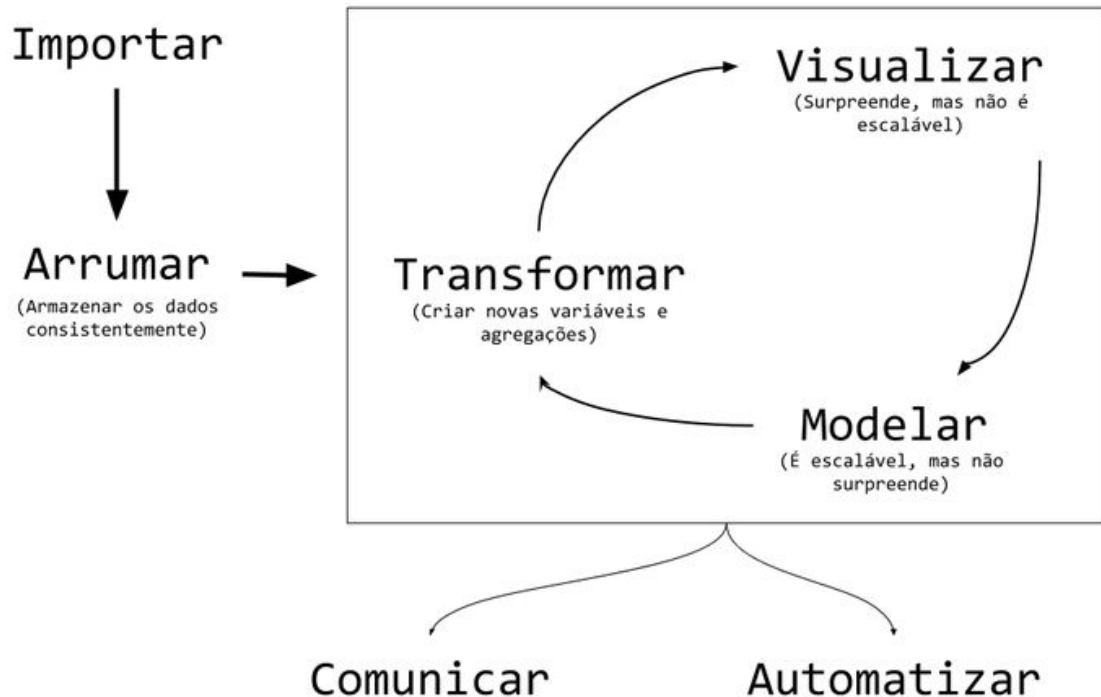
Após a reunião de planejamento das PyLadies-São Paulo em julho/2018, um grupo de mulheres se reuniu para formar o Grupos de Estudos em Data Science. O grupo tem como objetivos estudar Python e Ciência de Dados e retribuir para a comunidade o conhecimento adquirido através de eventos e produções de materiais.



Ciências de Dados



Ciência de Dados



Por onde começar?

Estatística?



Programação?

Área de domínio?



Etapas da Ciência de Dados

COLETA DE DADOS

Obtenção dos dados a partir de várias fontes

- Arquivos CSV, Excel
- Bancos de dados
- Redes Sociais
- Página web



- API Request (Facebook, Twitter)
- Pandas (CSV, Excel)
- Beautiful Soup
- mysql.connector

TRATAMENTO DOS DADOS

Organização dos dados coletados

- Remover dados desnecessários
- Tratar dados faltantes
- Tratar inconsistências



- Python Nativo
- Pandas
- Numpy



Etapas da Ciência de Dados

ANÁLISE EXPLORATÓRIA

Extrair informações dos dados

- Identificar outliers
- Identificar padrões
- Identificar variáveis importantes
- Plotar gráficos

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Scipy



MODELAGEM

Construções de modelos

- Realizar projeções em séries temporais
- Realizar previsões
- Reconhecimento de imagens
- Detecção de fraudes
- Recomendação de produtos e serviços

- Scikit-learn
- NLTK
- TensorFlow



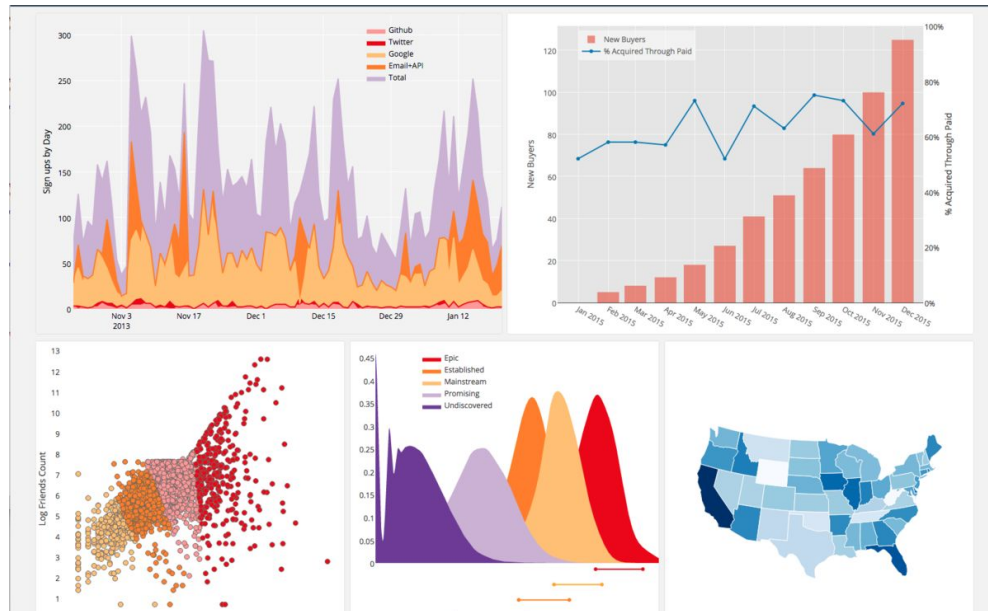
Etapas da Ciência de Dados

COMUNICAÇÃO

Comunicar os resultados obtidos

- Plotar gráficos
- Divulgar métricas de avaliação
- Conclusões

- Matplotlib
- Seaborn
- Plotly
- Tableau



NÃO SE DESESPERE! Escolha um tema por vez para estudar, isso ajuda você a progredir nos estudos.



Sugestão de trilha estudos programação

Python

- Estruturas da linguagem
- Funções
- Laços (for, while, etc)

Pandas

- Importar bases
- Manipular colunas
- Joins

Numpy

- Manipular vetores e matrizes

Matplotlib e Seaborn

- Plotar gráficos

Scikit-learn

- Treinar modelos

Sugestão de trilha estudos estatística

Estatística Descritiva

- Teorema do limite central
- Medidas de dispersão
- Distribuição de frequências

Probabilidade e Testes de Significância

- Probabilidades: Básica e Condicional
- Teorema de Bayes
- Variáveis Aleatórias: Contínuas e Discretas
- Testes de Hipótese
- P-valor
- Intervalo de Confiança

Estimadores e Regressões

- Modelos de predição e classificação

Dicas

- Não foque apenas em programação;
- Aprenda conceitos básicos de estatística e matemática;
- Aprenda a comunicar suas análises e resultados;
- Nunca deixe de praticar;
- Aprenda um conceito de cada vez;
- Defina metas;
- Estude com amigas, isso torna o processo mais divertido;
- Compartilhe o aprendizado, isso mostra o quanto você evoluiu;
- Lembre-se: o valor do Cientista de Dados não é saber desenvolvimento ou estatística, mas sim ter insights de seus dados!



Tenha consciência que o aprendizado é difícil, demorado e requer dedicação. Mas seu esforço será recompensado!

Por onde eu estudo?

- Livro Guia Mangá de Estatística - Shin Takahashi
- Plataforma Kaggle - <https://www.kaggle.com/>
- Podcast Pizza de Dados - <https://pizzadedados.com/>
- Documentação Pandas -
<https://pandas.pydata.org/pandas-docs/stable/index.html>
- Udacity - <https://www.udacity.com/>
- Canal EstaTiDados
<https://www.youtube.com/channel/UC4jROkPjTvnXRkuo2GAwKXw>
- Minerando Dados - <http://minerandodados.com.br/>

Por onde eu estudo?

- Cientista de Dados com GIFs - <https://paulovasconcellos.com.br/>
- Data Hackers <https://datahackers.com.br/>
- Estatística Básica - P. A. Bussab, W. de O. Moretin -
<https://edisciplinas.usp.br/mod/resource/view.php?id=2425203>
- Git Hub PyLadies São Paulo:
<https://github.com/PyLadiesSP/data-science>
- Git Hub Letícia Portela -
<https://github.com/leportella/datascience-pizza>
- Data Science para Leigos - [Lillian Pierson](#)

E por hoje é só pessoal!

 PyLadiesSP

 PyLadiesSãoPaulo

 PyLadiesSP

 @PyLadiesSP

 PyLadiesSP

 saopaulo@pyladies.com



Mulheres que
amam programar
e ensinar Python