

DATA SCIENCE & STATISTICAL COMPUTING

ANÁLISE EXPLORATÓRIA E ESTATÍSTICA DESCRITIVA

Prof. MEng. Daniel de Souza Carvalho
profdaniel.carvalho@fiap.com.br
@danielscarvalho

LAB: Pandas + Numpy

- Processo de
- Ciências dos Dados



THE DATA SCIENCE PROCESS



Data Engineers

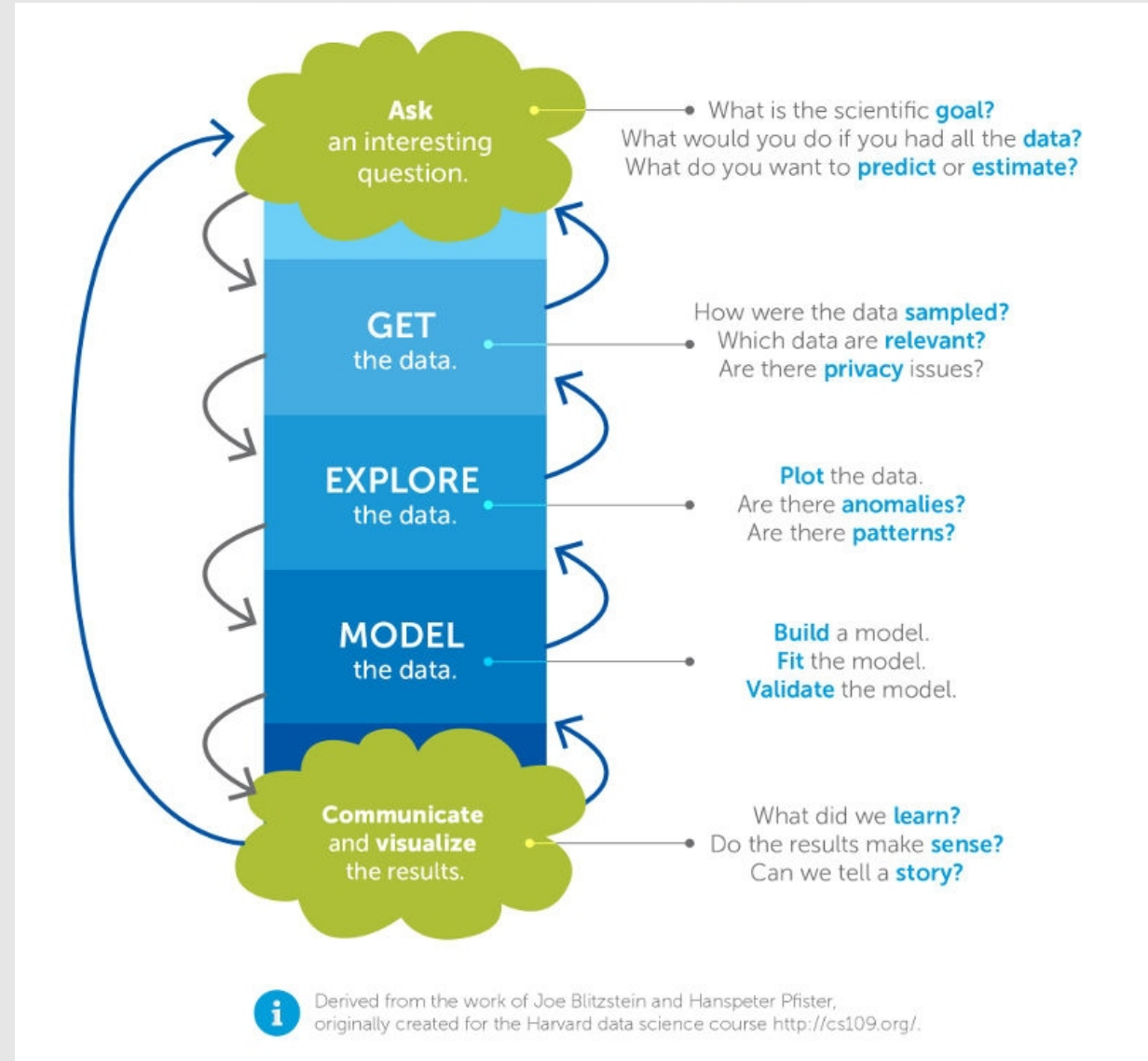
Data Analysts

Machine Learning Engineers

Data Scientists

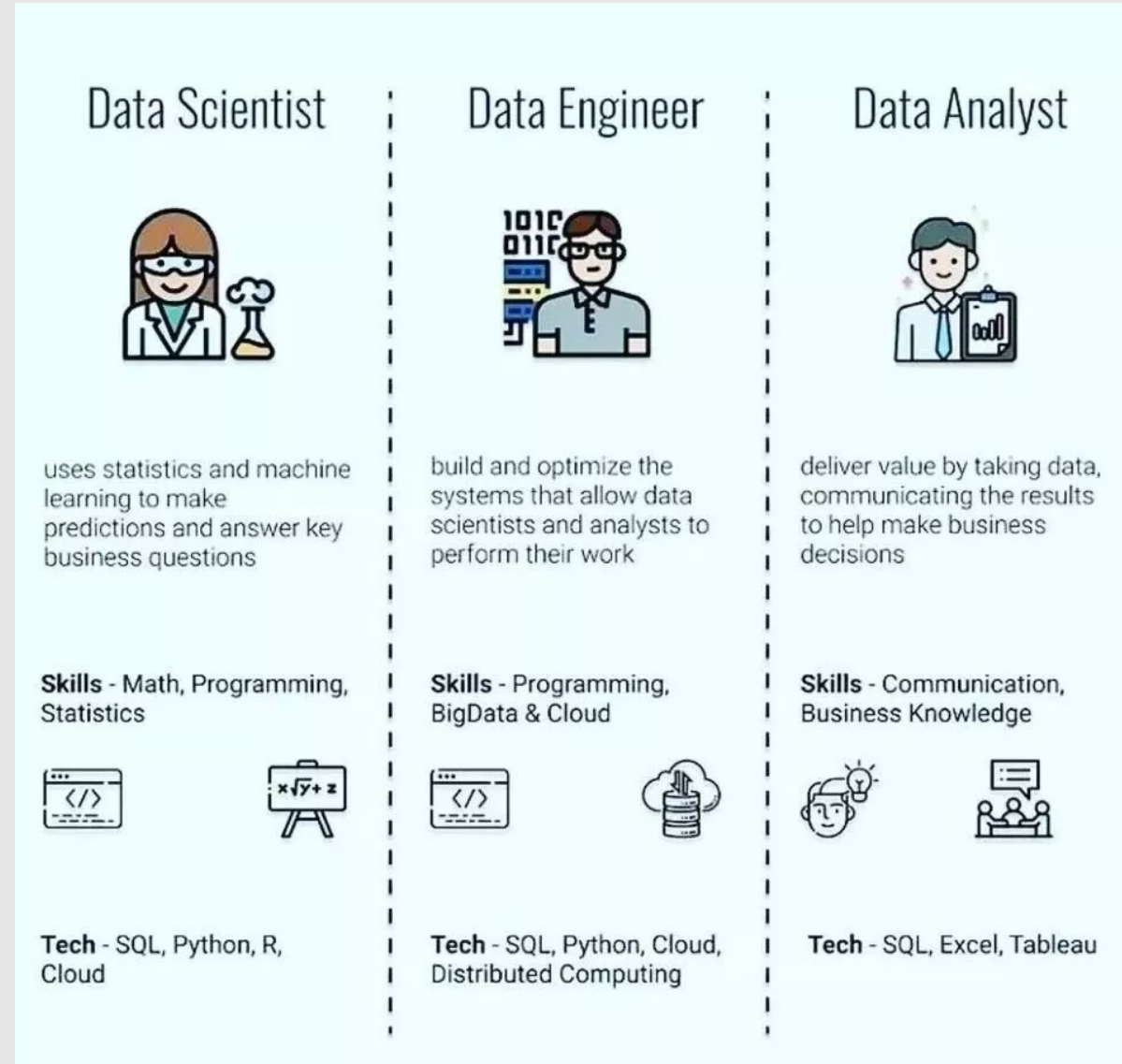
LAB: Pandas + Numpy

- **Processo de**
- **Ciências dos Dados**



LAB: Pandas + Numpy

- **Data Scientist**
- **Data Engineer**
- **Data Analyst**



LAB: Pandas + Numpy

- **Parte 1**
- Documentação online oficial do Pandas (Inglês)
- “10 minutes to pandas”
- https://pandas.pydata.org/docs/user_guide/10min.html
- Vamos trabalhar juntos no ambiente Jupyter

LAB: Pandas + Numpy

- **Parte 2**
- Documentação online oficial do Pandas (Inglês)
- “Cookbook” (Livro de receitas)
- https://pandas.pydata.org/docs/user_guide/cookbook.html#cookbook
- Vamos trabalhar juntos no ambiente Jupyter

LAB: Pandas + Numpy

- **Parte 3**

- Atividade em duplas (*paring – pair programming*)
- Escolher um conjunto de dados do Kaggle
- Importar e explorar os dados com Pandas no Jupyter
- Mostrar e compartilhar com os colegas suas descobertas
- Documente com MARKDOWN, explique suas observações
- Publicar o notebook no GitHub

1. Ask
2. Get
3. Explore
4. Model
5. Communicate

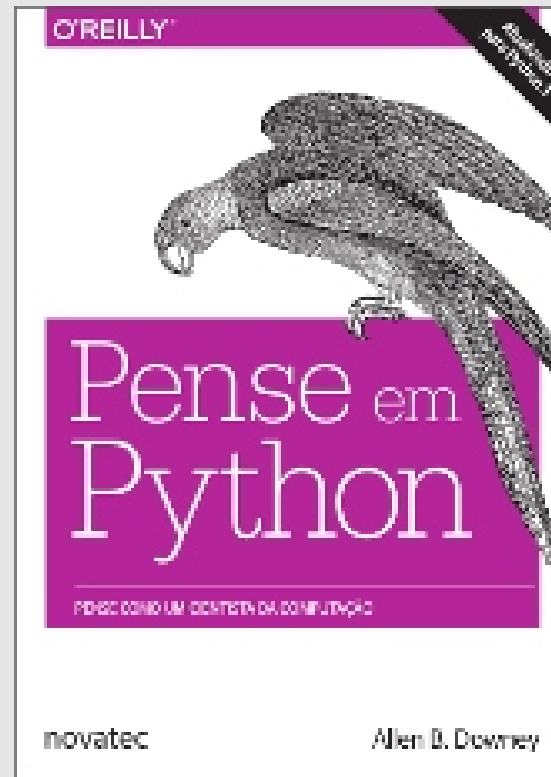
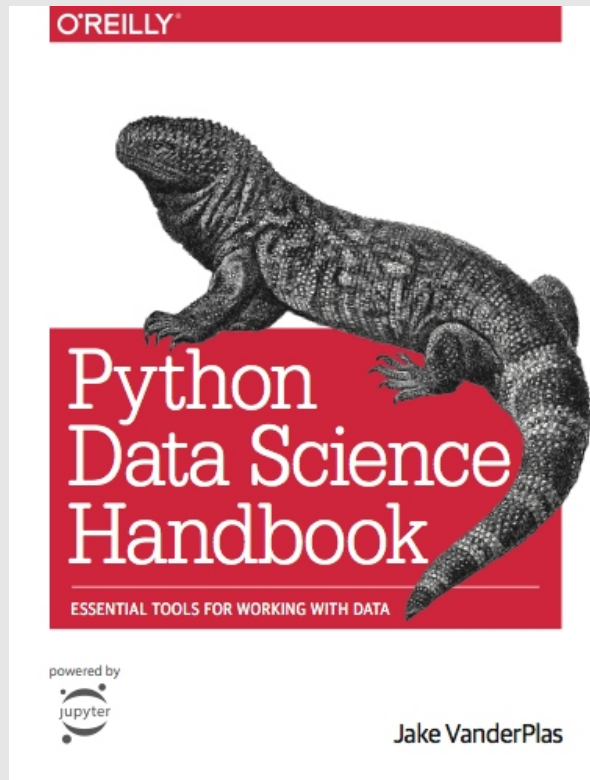
- <https://www.kaggle.com/>



kaggle

LAB: Pandas + Numpy

- **Livros online** (PDF ou Site HTML)



<https://jakevdp.github.io/PythonDataScienceHandbook/>

<https://penseallen.github.io/PensePython2e/>

<https://books.goalkicker.com/PythonBook/>

LAB: Pandas + Numpy

