



CIÊNCIA DE DADOS

O BOOTCAMP

por Thais Viana

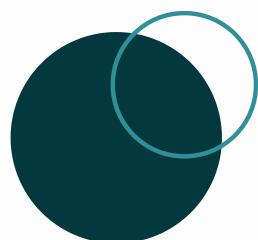
Thais Viana

A INSTRUTORA

VOCÊS

O TIME DE DESENVOLVIMENTO

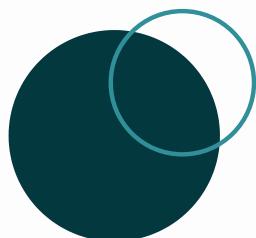
NOSSA AGENDA



HOJE

O QUE É O CURSO
MOTIVAÇÃO
BIBLIOGRAFIA
RECURSOS DISPONÍVEIS
COMPUTAÇÃO ÉTICA
LEI DOS DADOS
OPEN DATA
NUMPY

NOSSA AGENDA



TERÇA & QUARTA

TUTORIAL DE PANDAS
TUTORIAL DE #1 SCIKIT LEARN

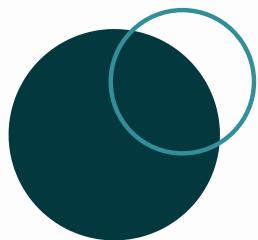
NAIVE BAYES

KNN

KMEANS

WEB SCRAPING

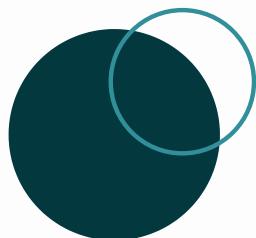
NOSSA AGENDA



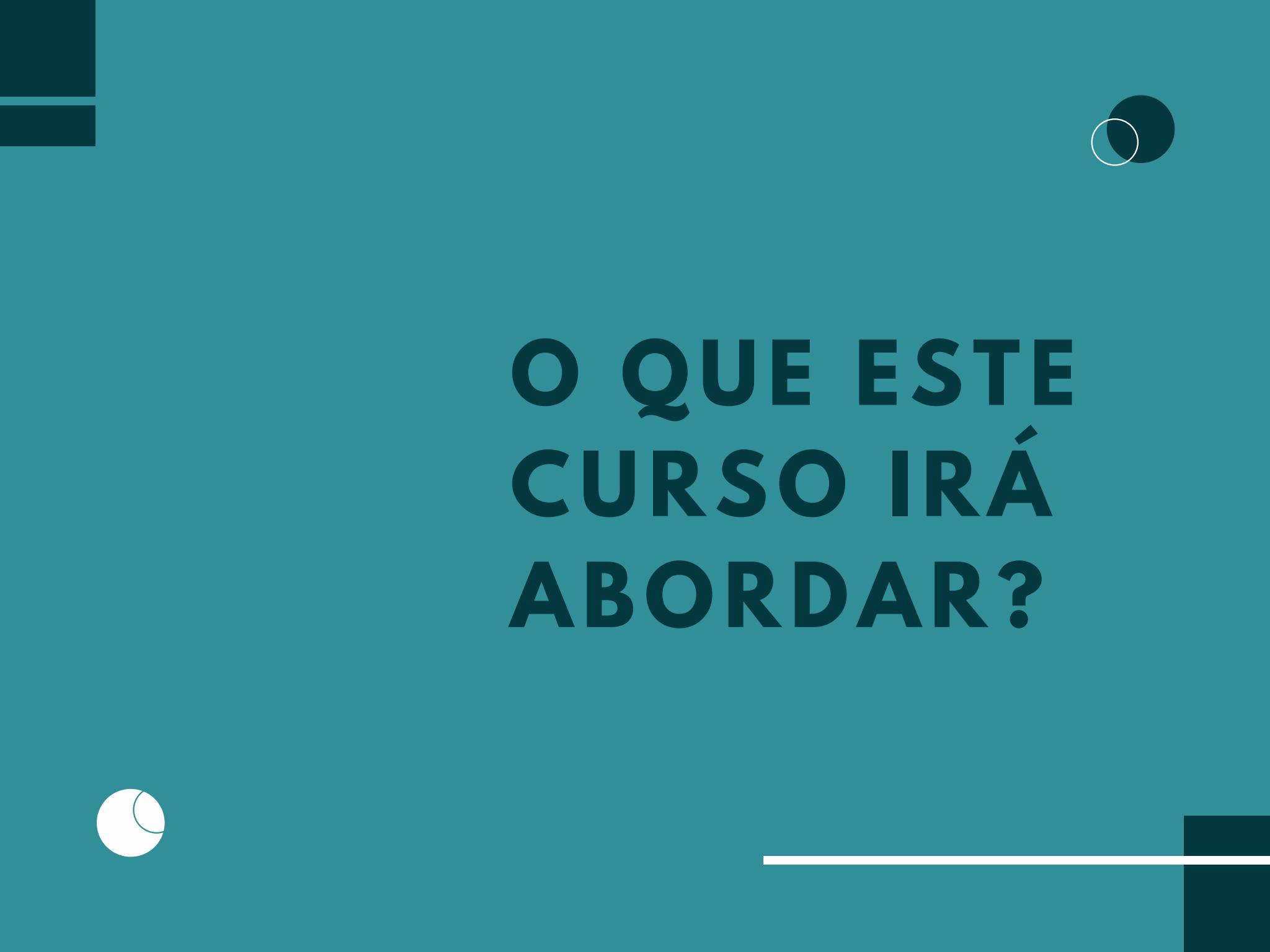
QUINTA MANHÃ

TUTORIAL DE #2 SCIKIT LEARN
RANDOM FOREST REGRESSOR
DECISION TREE CLASSIFIER

NOSSA AGENDA

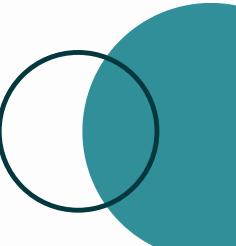


QUINTA TARDE & SEXTA
PROJETO

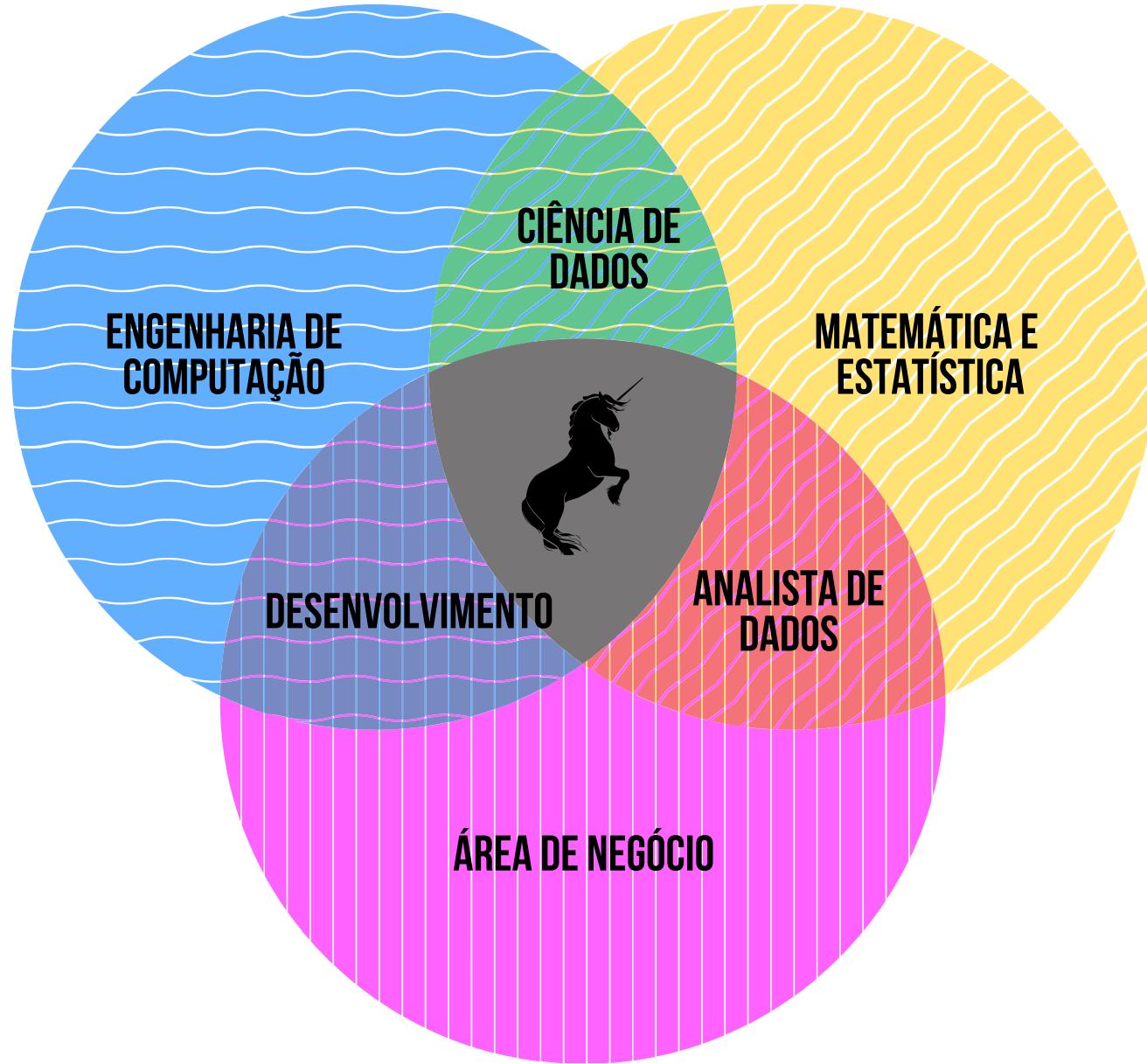


O QUE ESTE CURSO IRÁ ABORDAR?

e qual a
diferença?



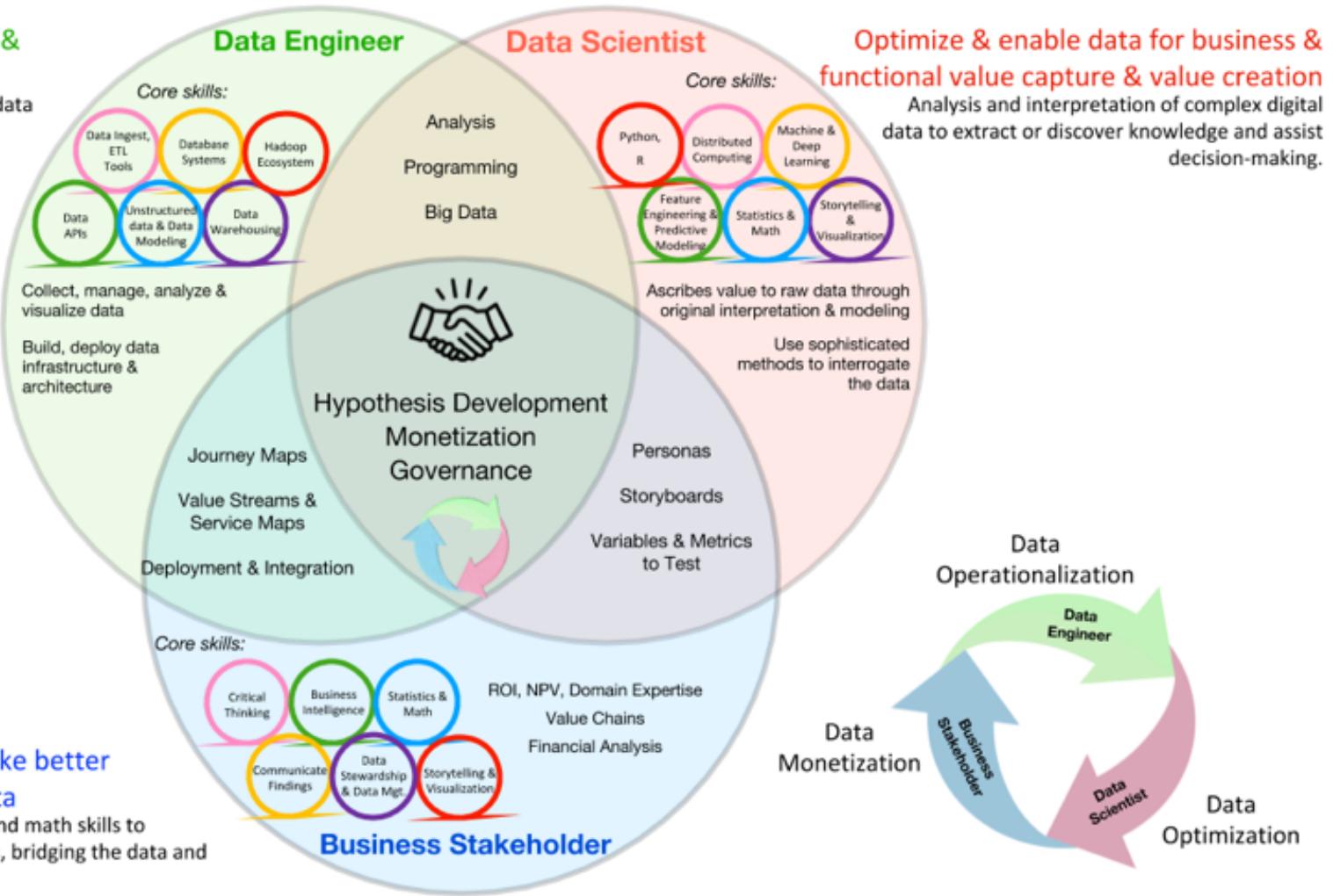
Carreiras em Ciência de Dados



Data Science Roles & How They Interact

Enable data access & utilization & enable value capture

Builds and supports the infrastructure or 'data pipe' and all associated SW engineering infrastructure tasks.



Help the business make better decisions through data

Blend of business, analytic and math skills to explore and solve challenges, bridging the data and business communities.



Josh Wills
@josh_wills

 Follow

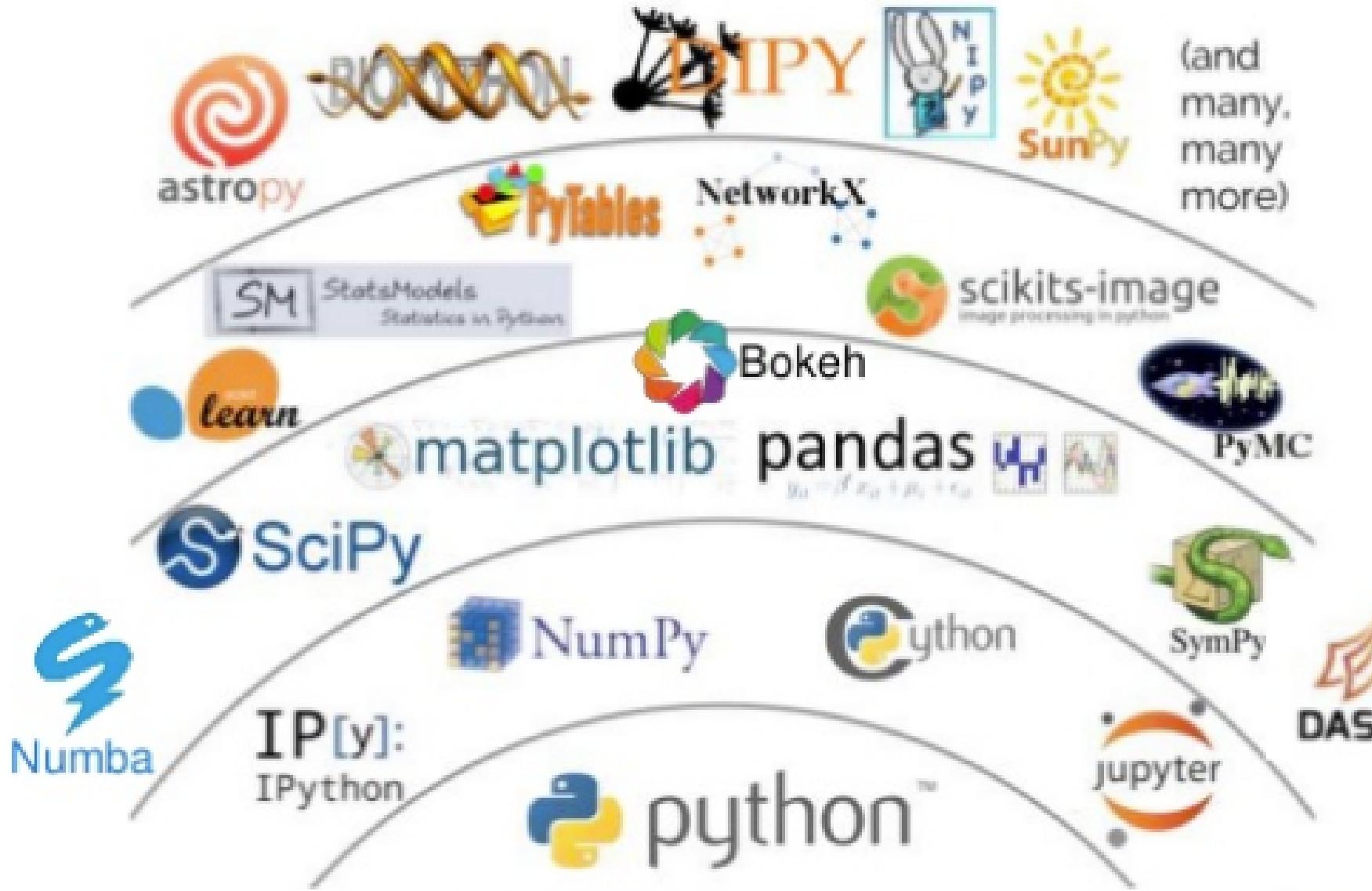
Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

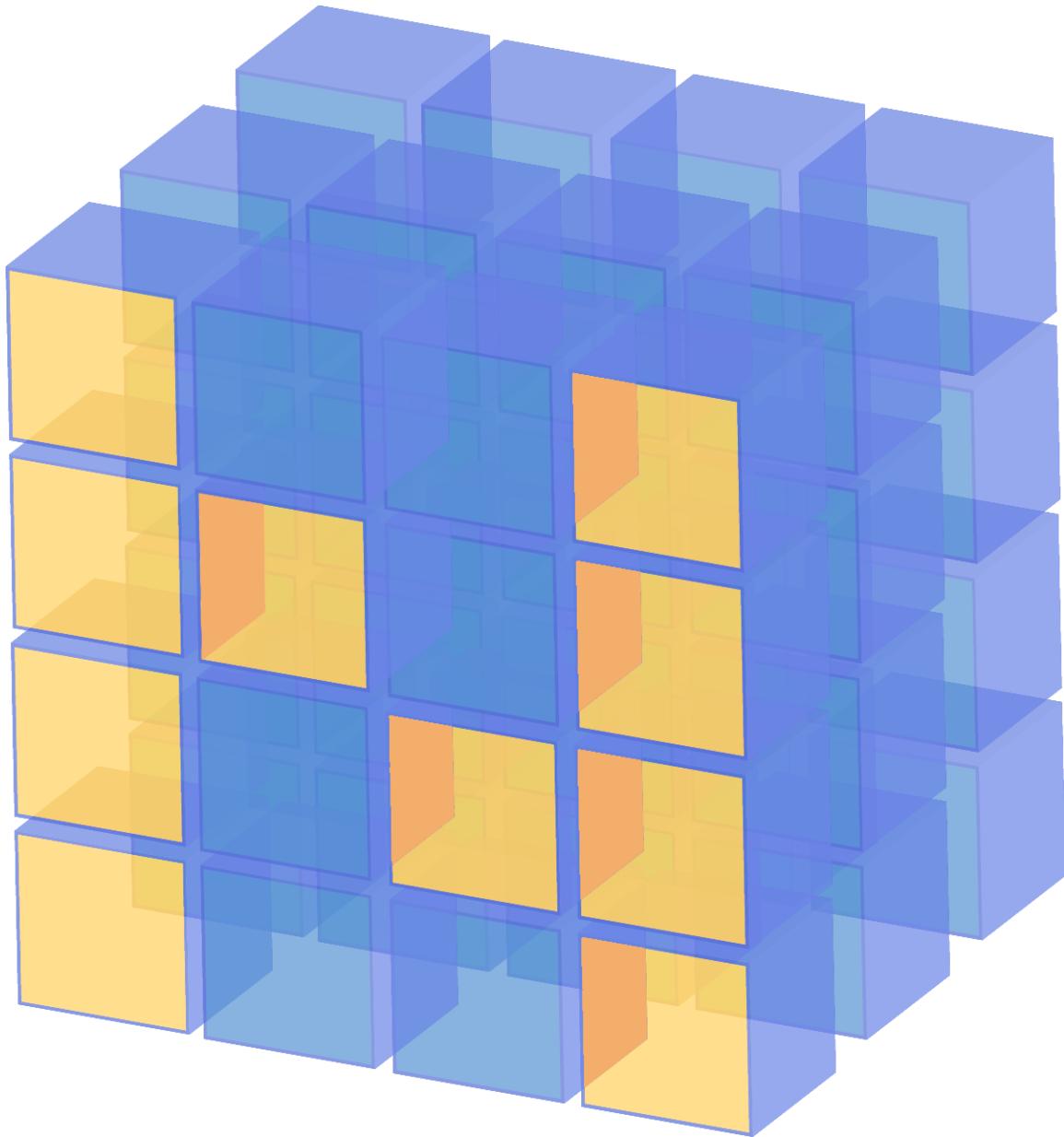
 Reply  Retweet  Favorite  More

9:55 AM - 3 May 12



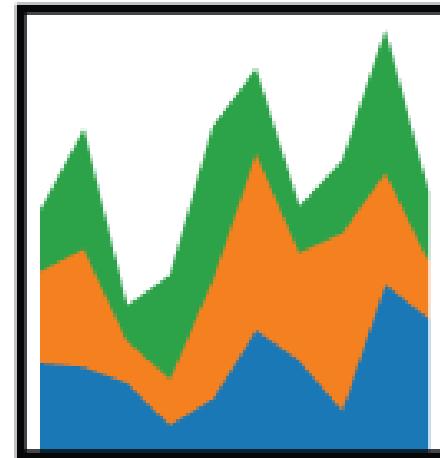
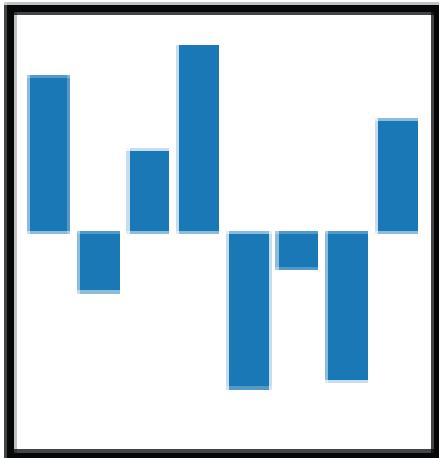
python™





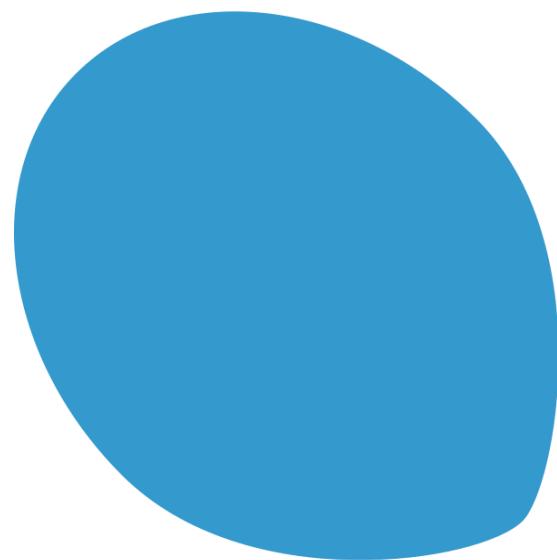
pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



matplotlib



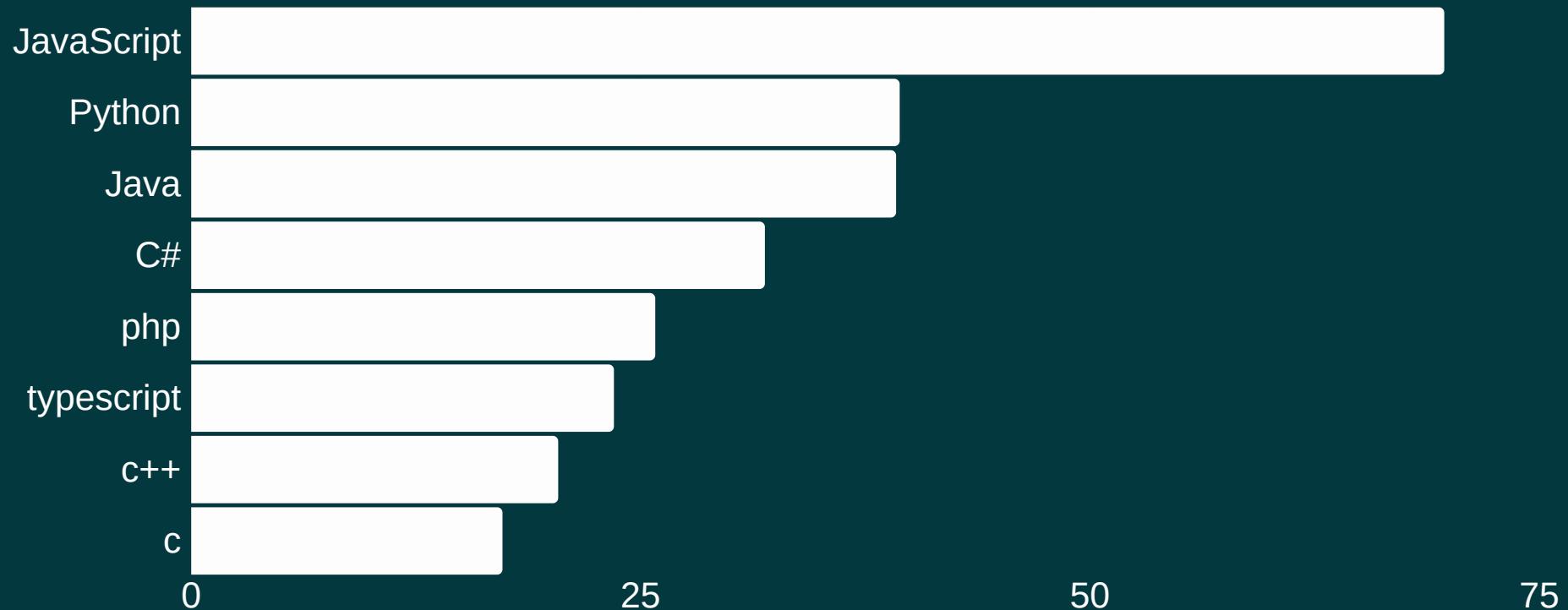




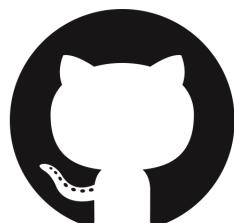
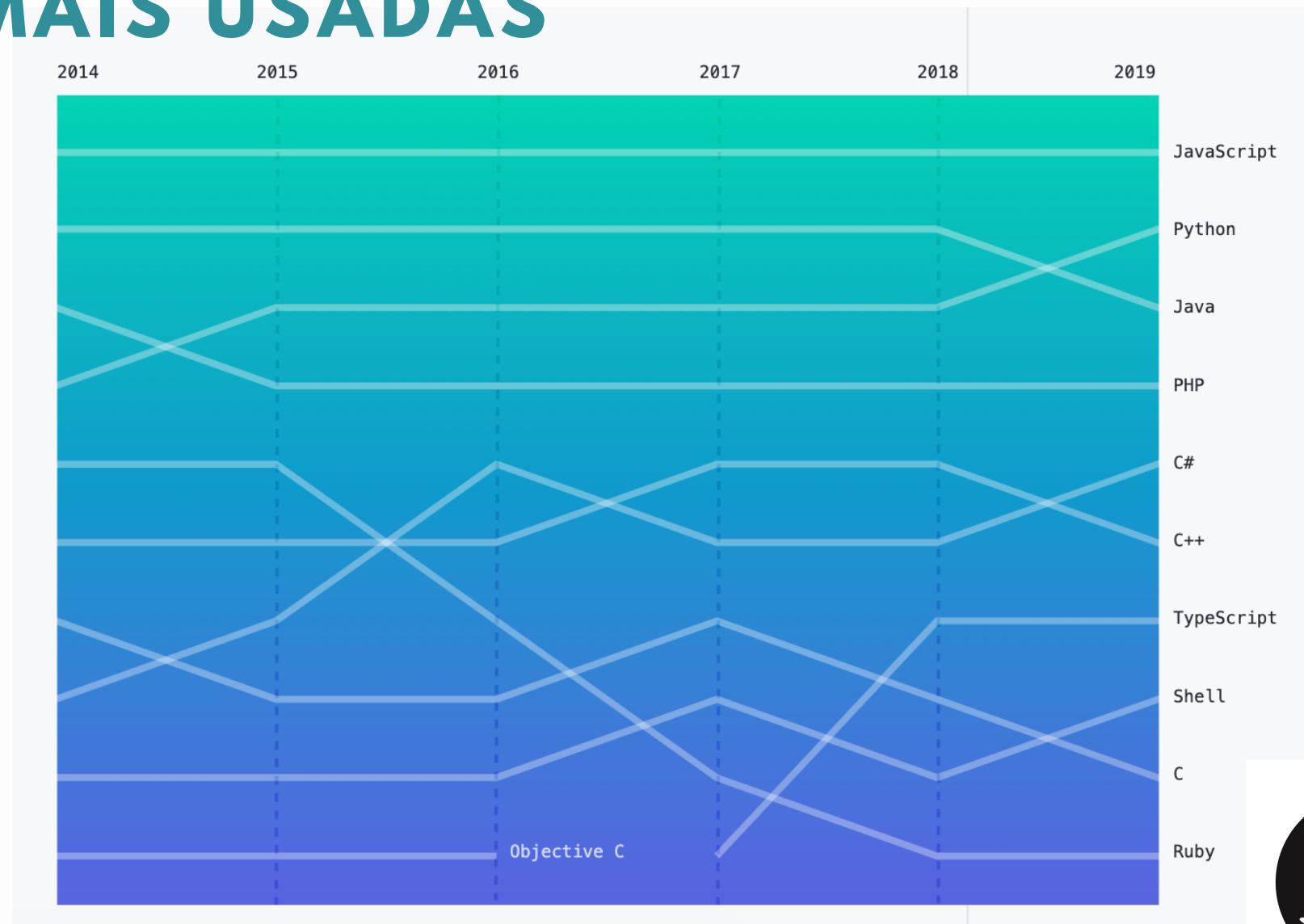
PYTHON CENTRADO

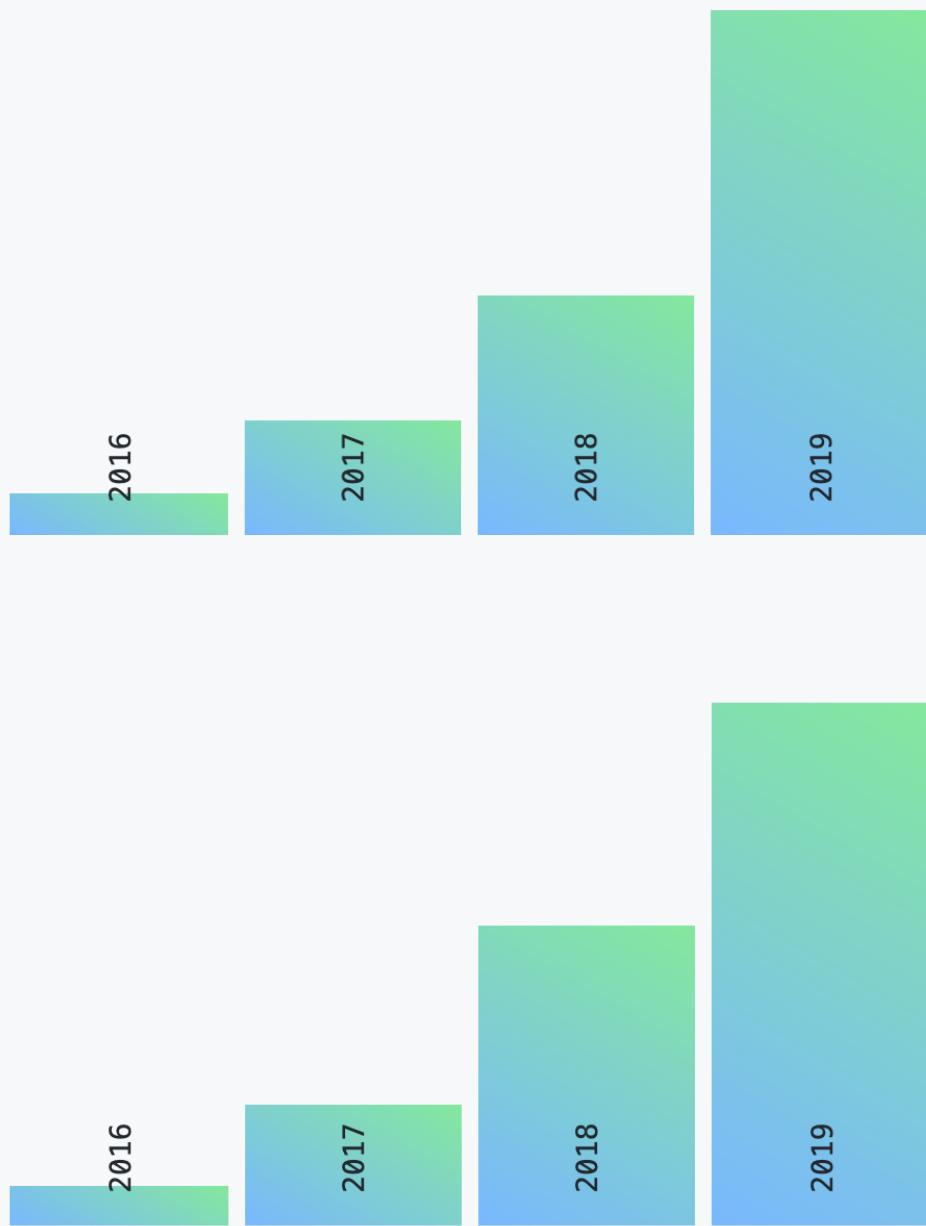
PORQUE?

LINGUAGENS DE PROGRAMAÇÃO MAIS USADAS EM 2019.



LINGUAGENS DE PROGRAMAÇÃO MAIS USADAS





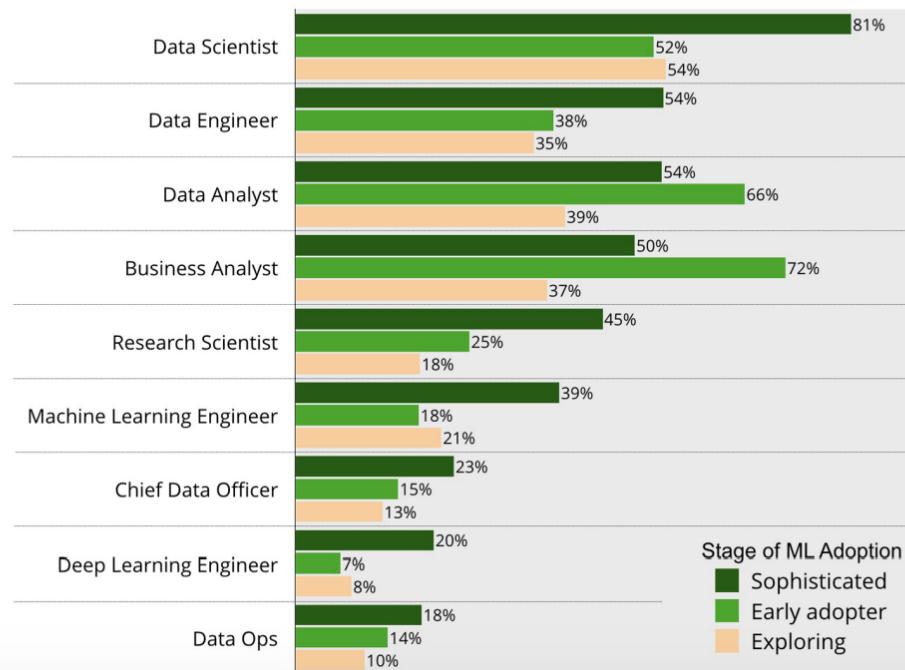
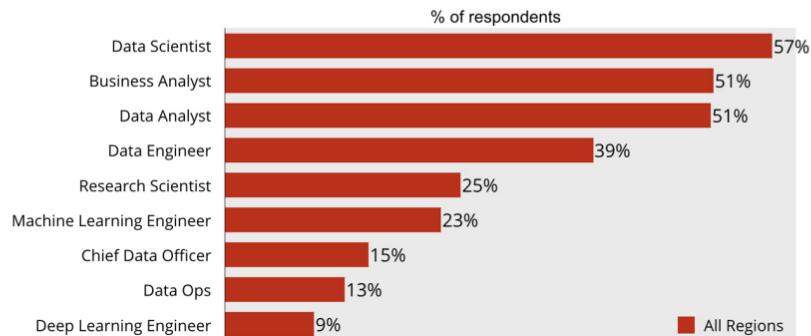
Growth of Jupyter Notebooks, 2016-2019

How else can we tell data science is growing on GitHub? The use of [Jupyter Notebooks](#) (by number of repositories with Jupyter as their primary language) has seen more than 100% growth year-over-year for the last three years.*

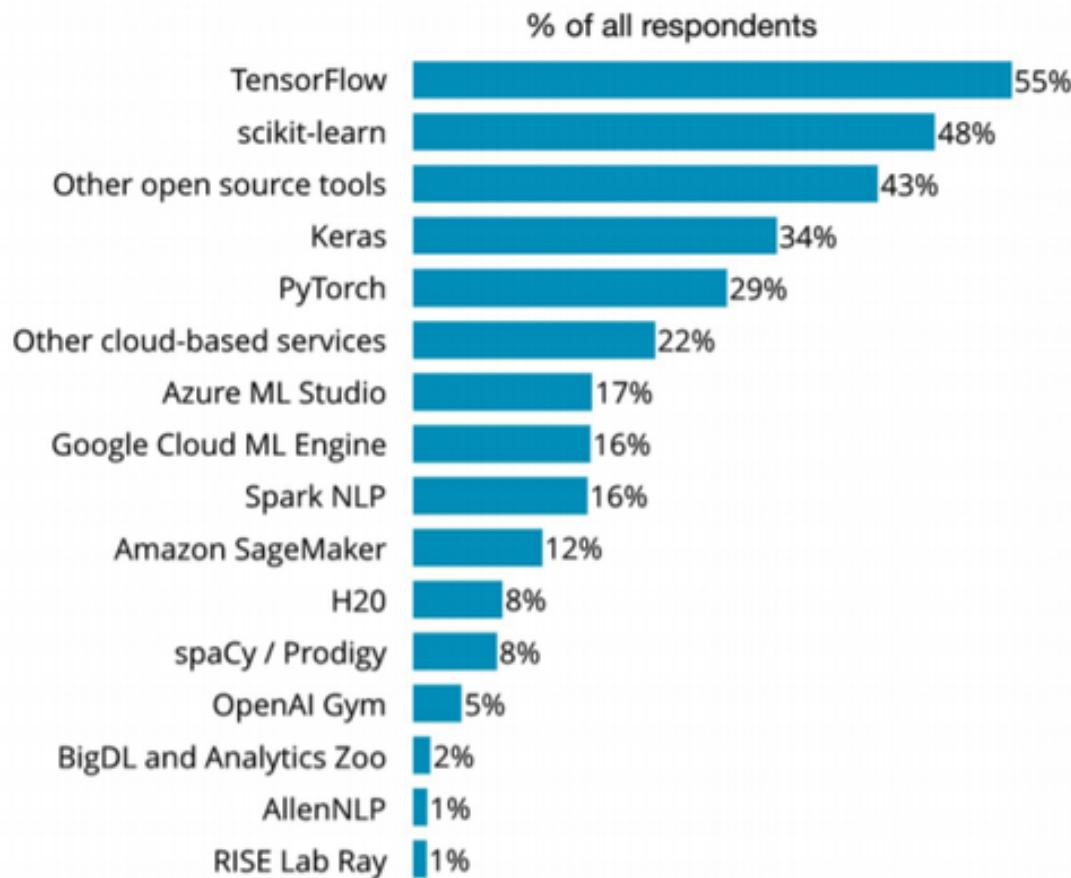
Growth of natural language processing repositories, 2016-2019

Natural language processing (NLP) is also picking up steam on GitHub, as packages like [NLTK](#) lower the barrier to entry for NLP work.*

Which of these titles are used in the organization where you work?



Which of the following AI tools are you using? (Select all that apply.)



O'REILLY®



**PARA ALÉM
DO PYTHON**

Platform



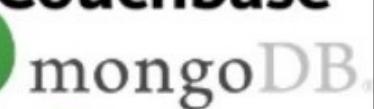
Framework



MLlib



Tools



Languages



python™



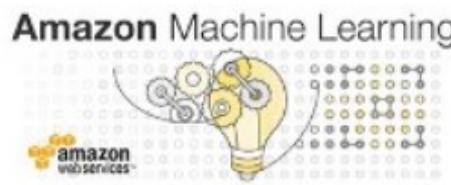
SQL



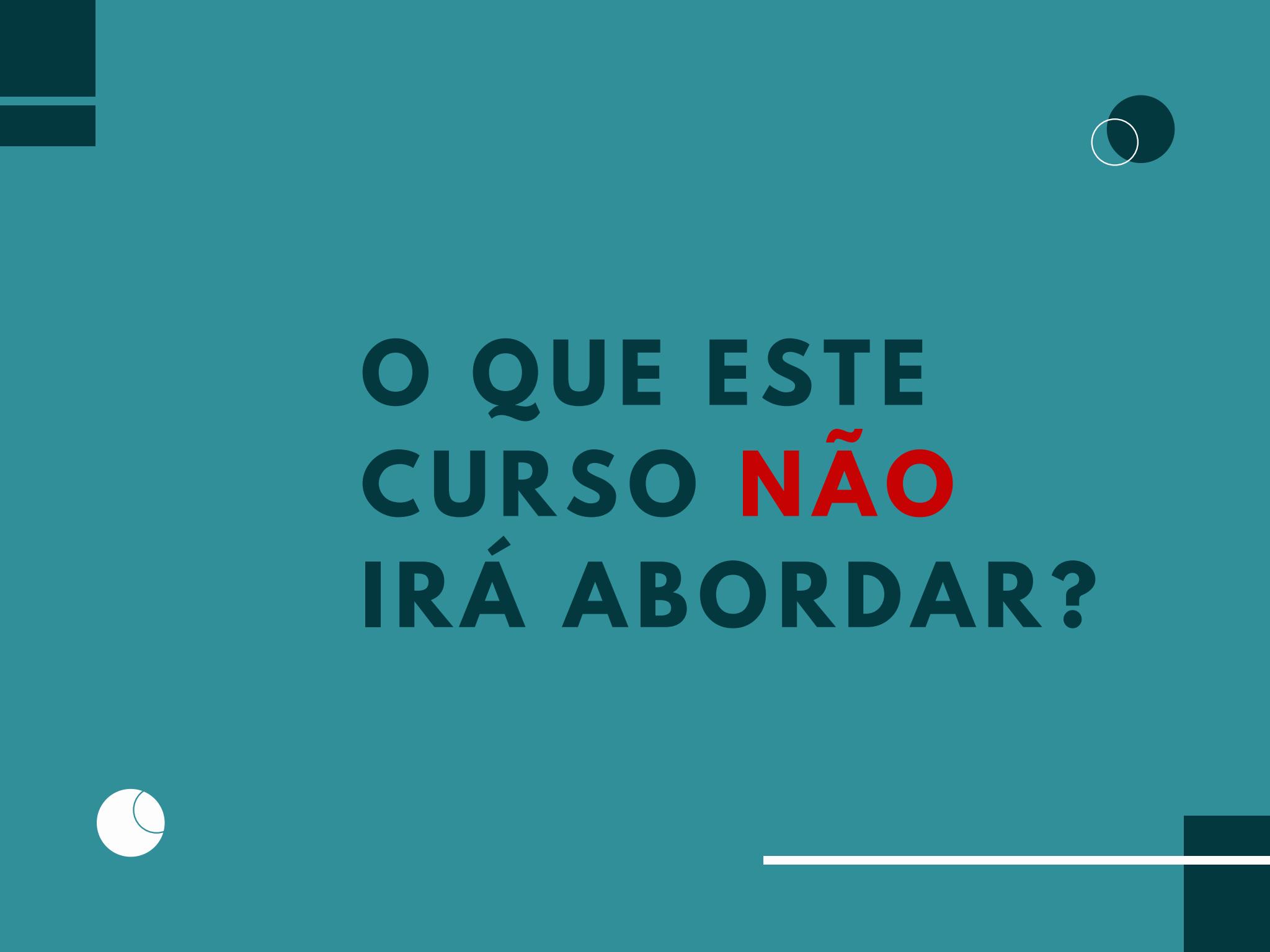
SaaS



Google
Machine Learning



Azure Machine Learning



O QUE ESTE
CURSO NÃO
IRÁ ABORDAR?

DATA Engineer

Develops, constructs, tests, and maintains architectures. Such as databases and large-scale processing systems.



DataCamp
Learn Data Science By Doing

DATA Scientist

Cleans, massages and organizes (big) data. Performs descriptive statistics and analysis to develop insights, build models and solve a business need.



Responsibilities...

ON DAILY BASIS

Develop, construct, test, and maintain architectures (such as databases and large-scale processing systems)



Ensure architecture will support the requirements of the business



Discover opportunities for **data acquisition**



Develop data set processes for data modeling, mining and production



Employ a variety of languages and tools (e.g. scripting languages) to **marry systems together**



Recommend ways to **improve data** reliability, efficiency and quality



Conduct research to **answer industry and business questions**



Leverage large volumes of data from internal and external sources to answer that business



Employ sophisticated analytics programs, machine learning and statistical methods to **prepare data for use in predictive and prescriptive modeling**



Explore and examine data to **find hidden patterns**



Automate work through the use of predictive and prescriptive analytics



Tell stories to key stakeholders based on their analysis

Data Science Workflow

Business Understanding



Data Understanding

Data Importing



Data Cleaning & Manipulation



Statistical Modeling & Machine Learning



Reporting & Visualization

Languages, Tools & Software

