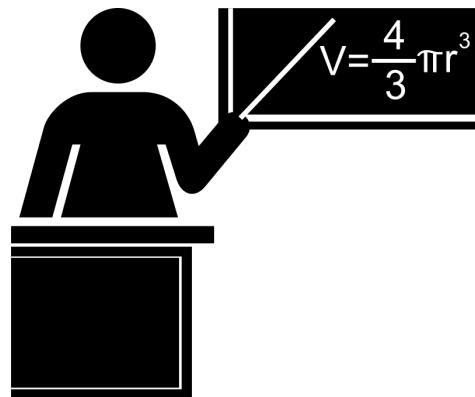
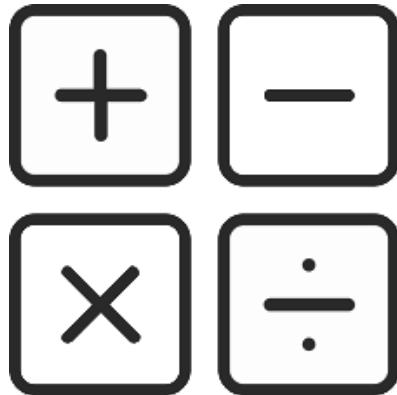


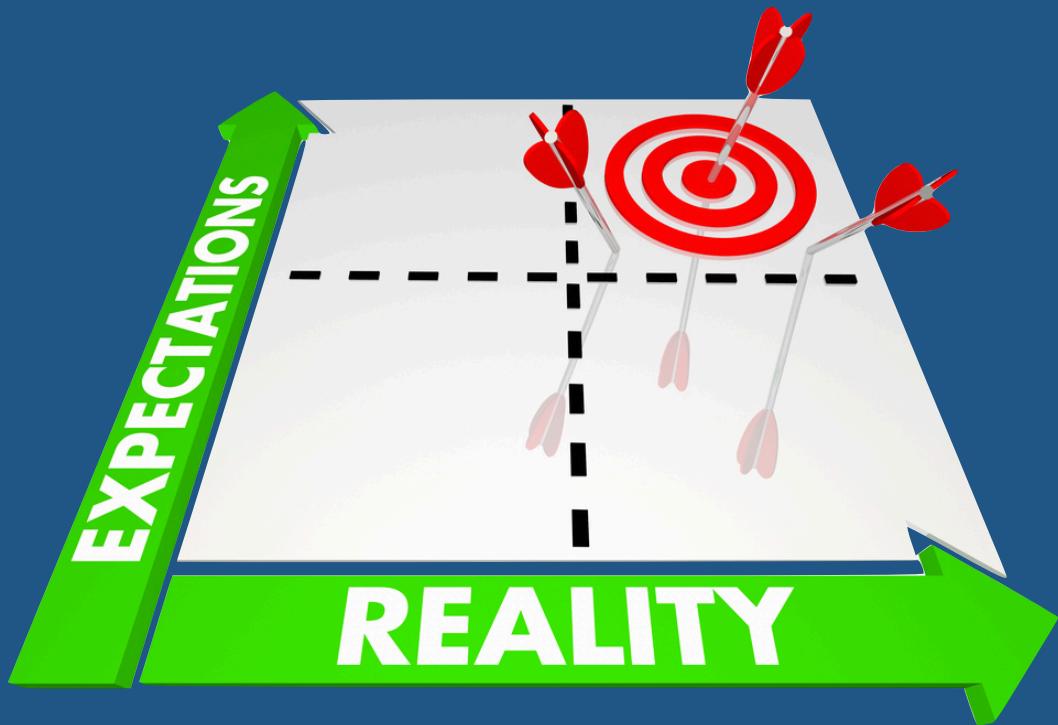
PROF. EDUARDO PÉCORA

DATA SCIENCE PPGOLD

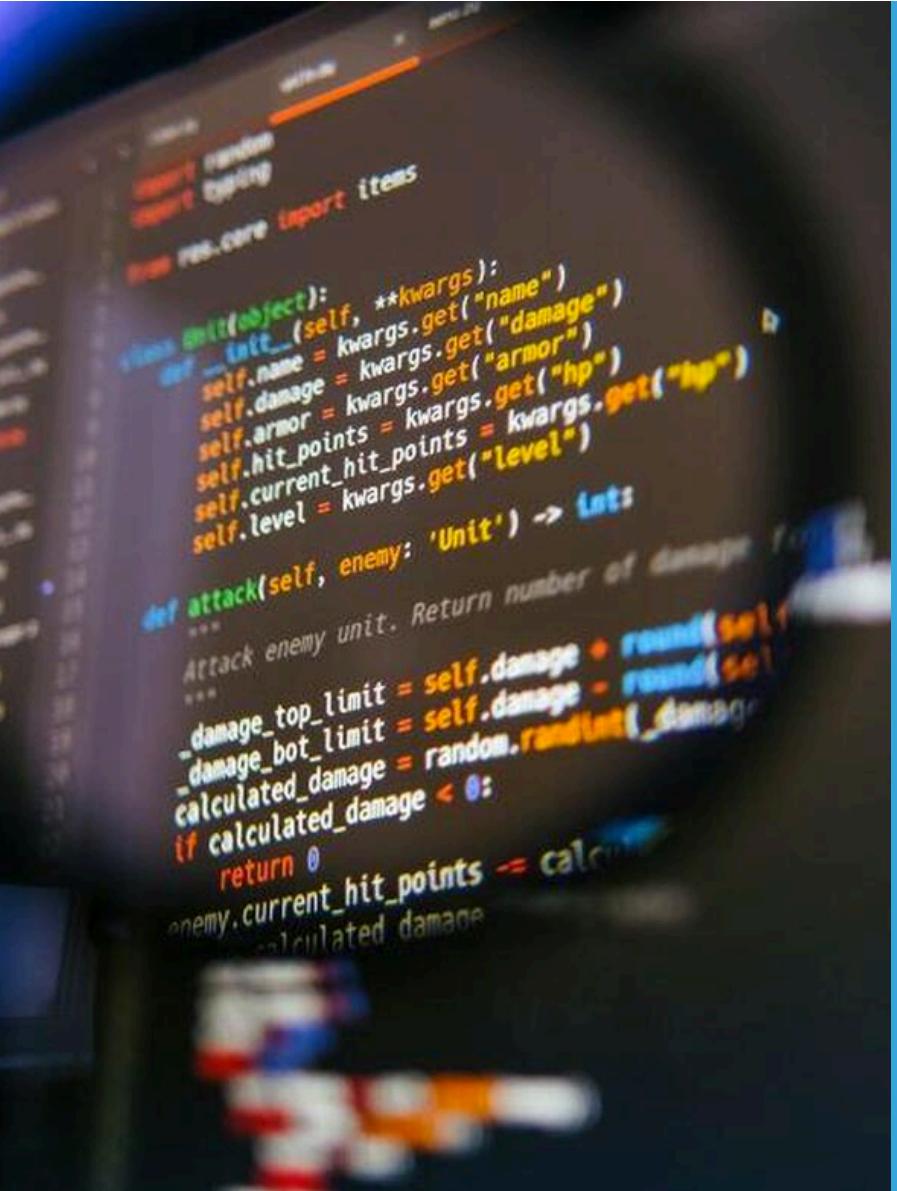


JUST LITTLE BIT ABOUT MYSELF





DATA SCIENCE
GOLD7037



ESTA NÃO É UMA DISCIPLINA
DE PYTHON, MAS ...

NÓS USAREMOS PYTHON NESTA
DISCIPLINA

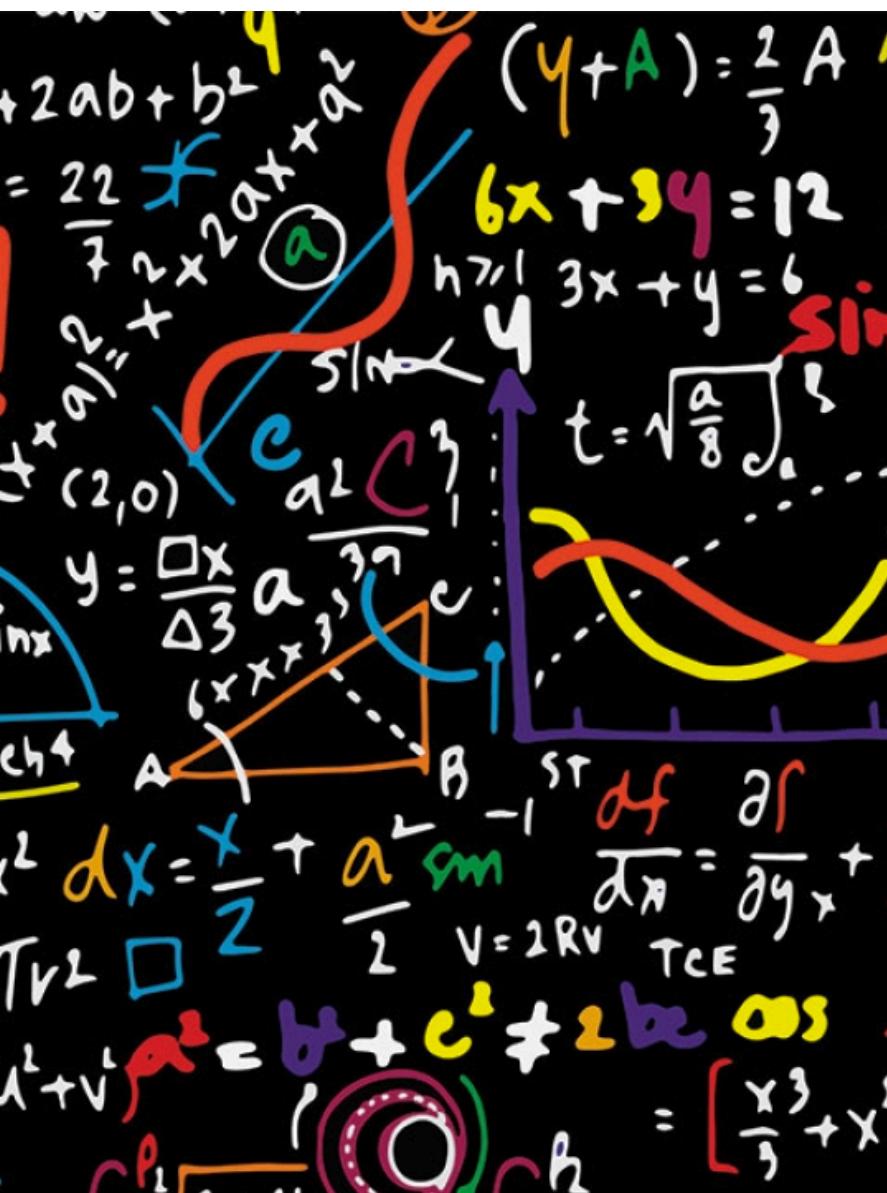
E PRA MIM ISSO SÃO DUAS
COISAS BEM DISTINTAS



O FOCO DESTA
DISCIPLINA É ENSINAR
A BASE DO DATA
SCIENCE



O FOCO DESTA
DISCIPLINA É NA
TOMADA DE DECISÃO
EMBASADA EM DADOS



NÓS USAREMOS
ESTATÍSTICA,
MATEMÁTICA E
COMPUTAÇÃO AQUI

DATA SCIENCE

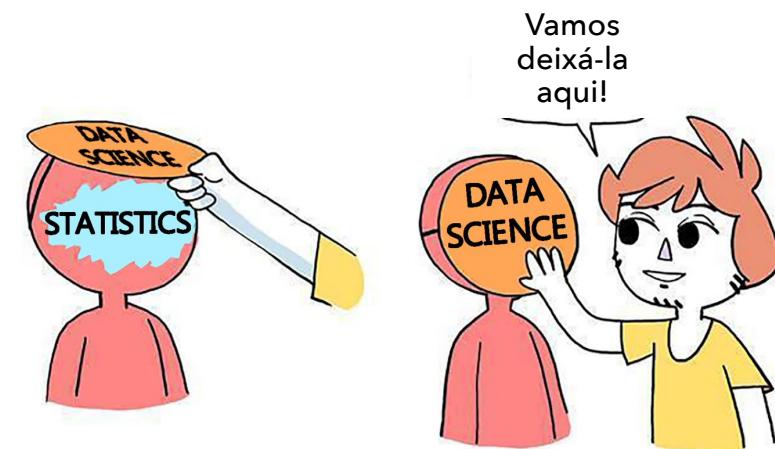
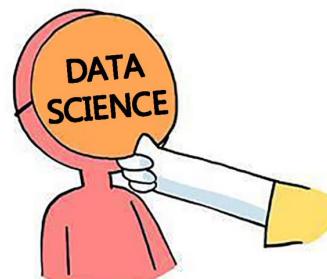
INTRODUÇÃO



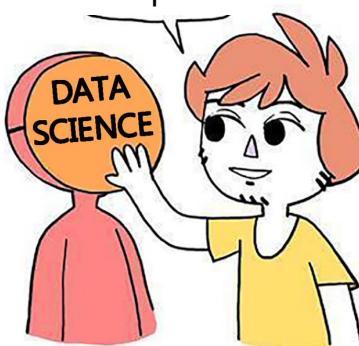
Oi Data Science porque
você sempre usa esta
máscara?

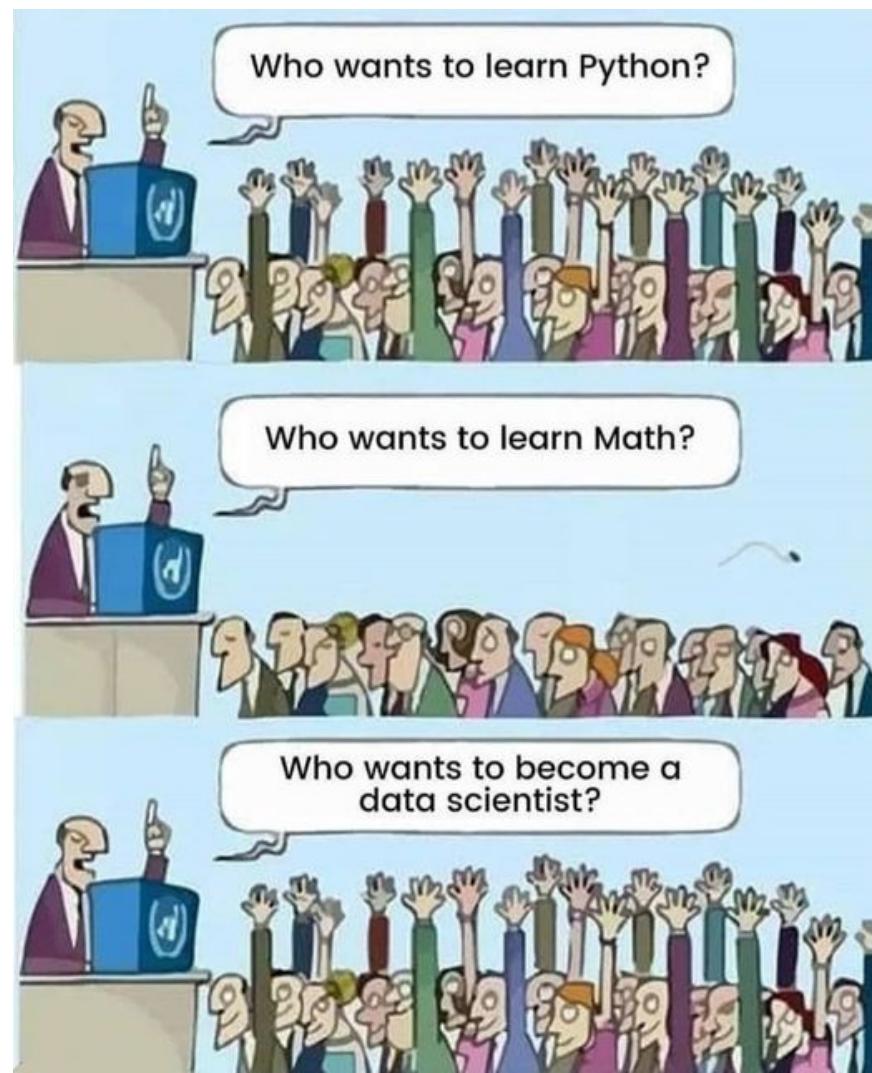


@statsanddatasci



Vamos
deixá-la
aqui!





```
if awake:  
    code()  
elif tired:  
    drink_coffee()
```

Data

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and D.J. Patil

From the Magazine (October 2012)

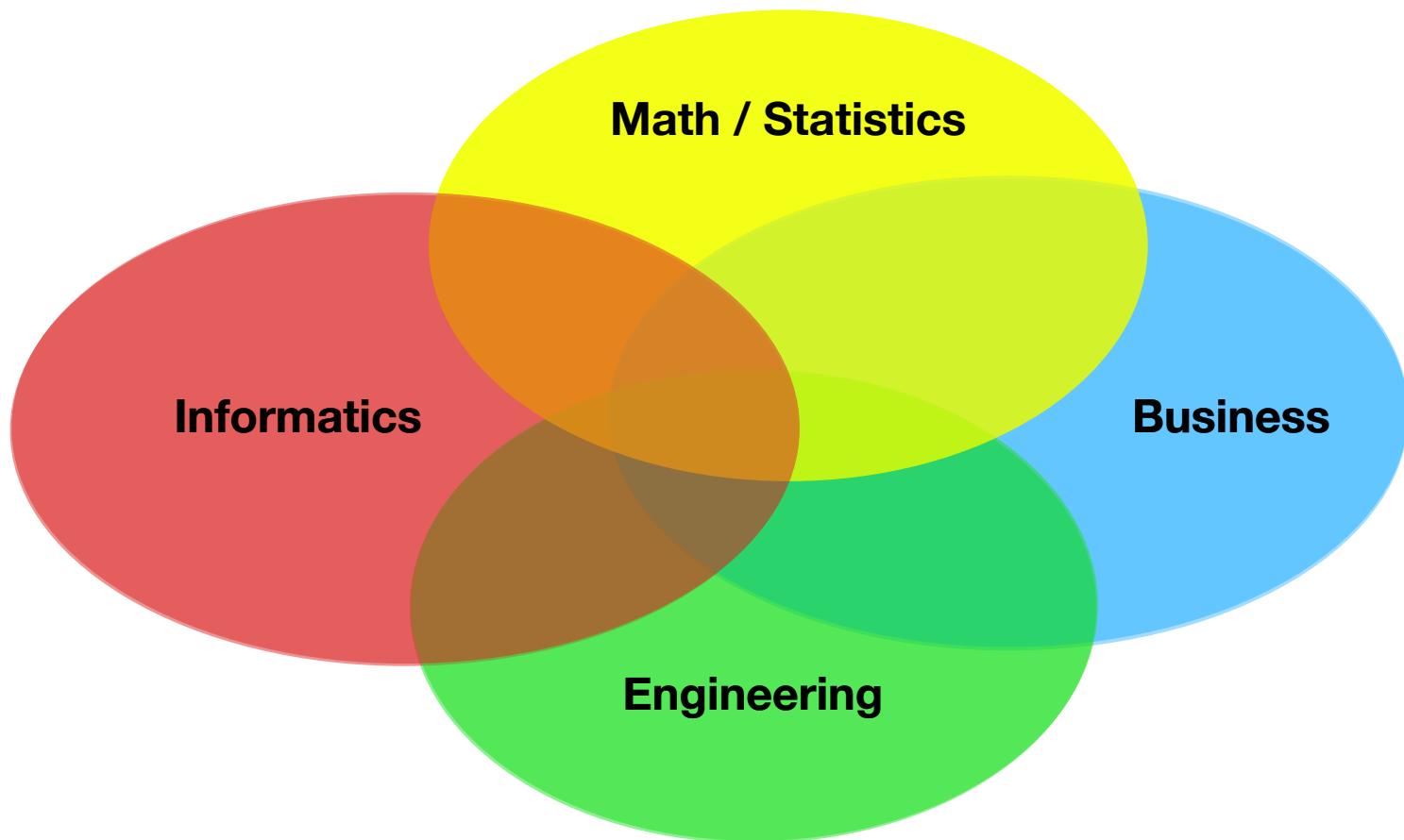


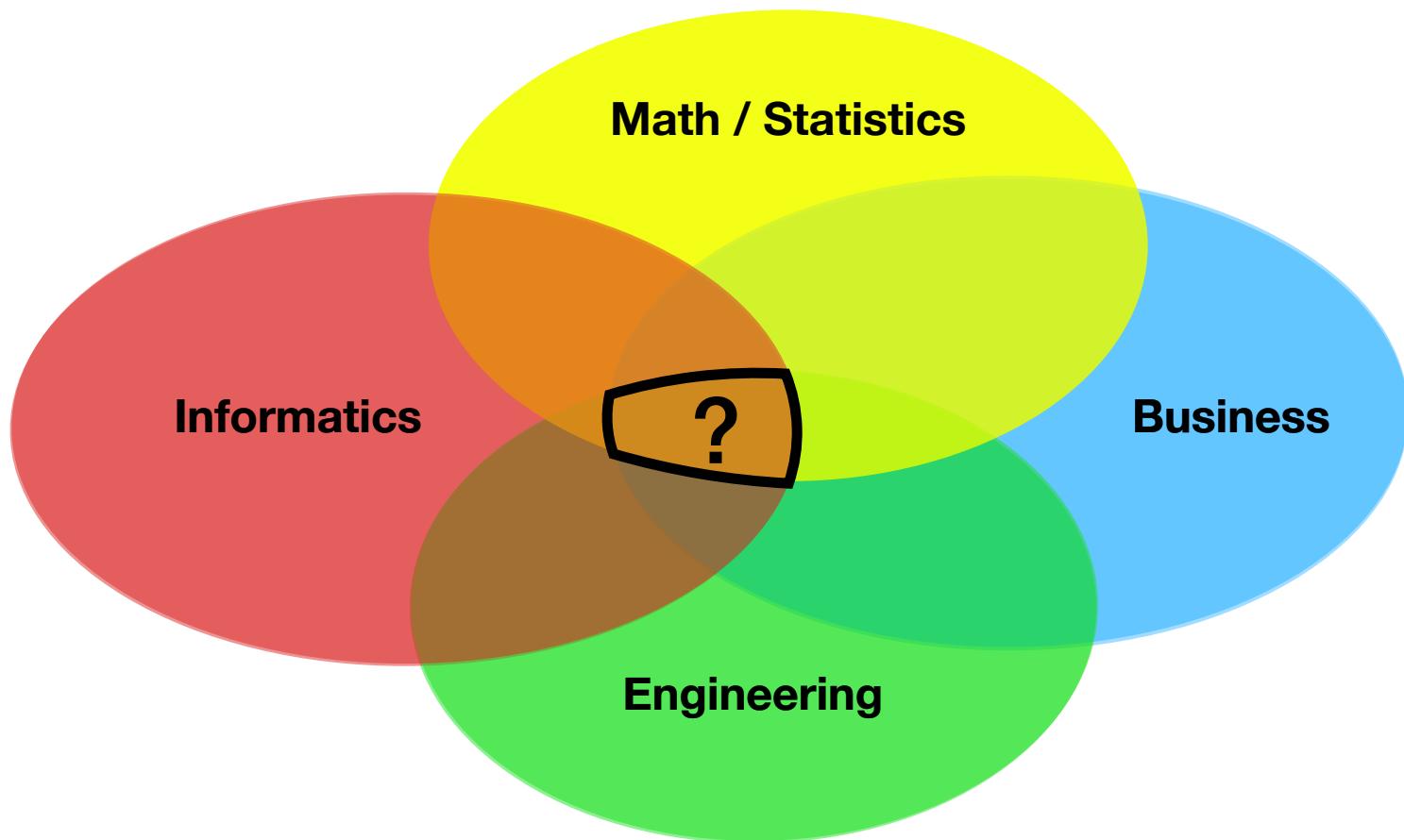
Artwork: Tamar Cohen, Andrew J Buboltz, 2011, silk screen on a page from a high school yearbook, 8.5" x 12"

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

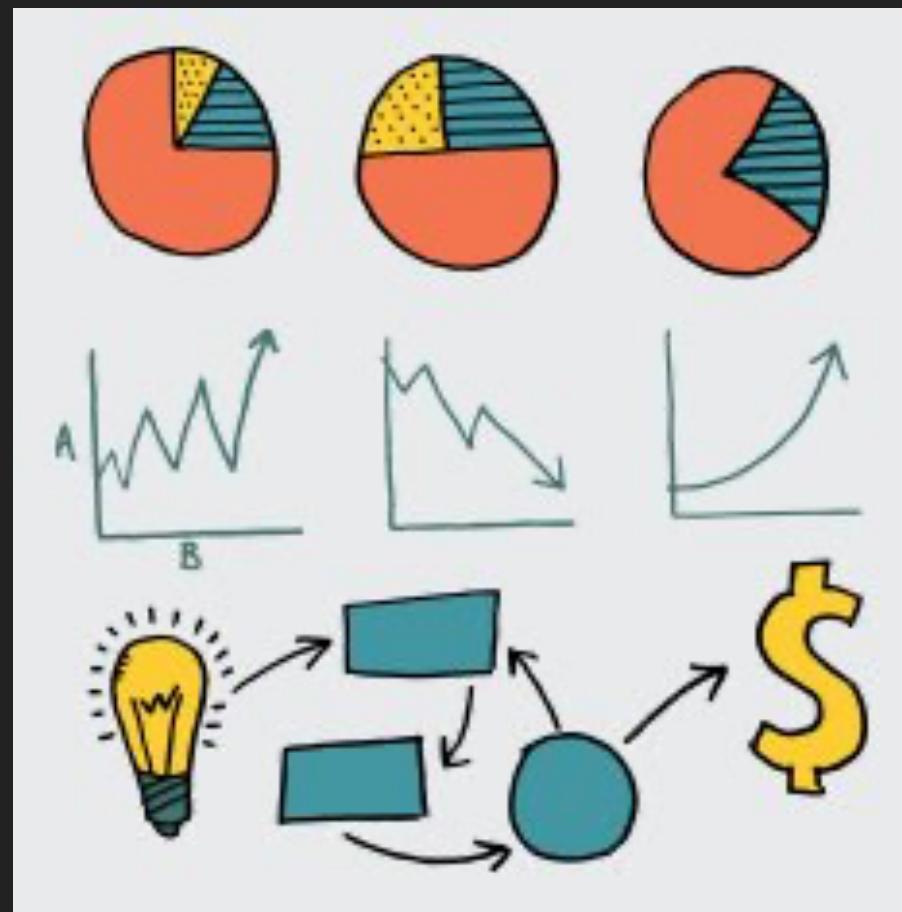
POSICIONANDO O DATA SCIENCE







OBJETIVO DO DATA SCIENCE



MAS EU POSSO SER SÓ UM USUÁRIO DO
DATA SCIENCE?

John Doe

MAS ATÉ PARA ISSO NÓS PRECISAMOS
DE UMA ESTRUTURA, QUE SEJA
REPLICÁVEL E INCREMENTAL.

Eduardo Pécora



METODOLOGIA DO DATA SCIENCE

DE “A” À “Z”

WHAT IS A METHODOLOGY?

methodology noun

 Save Word

meth·od·ol·o·gy | \,me-thə-'dä-lə-jē \

plural **methodologies**

Definition of *methodology*

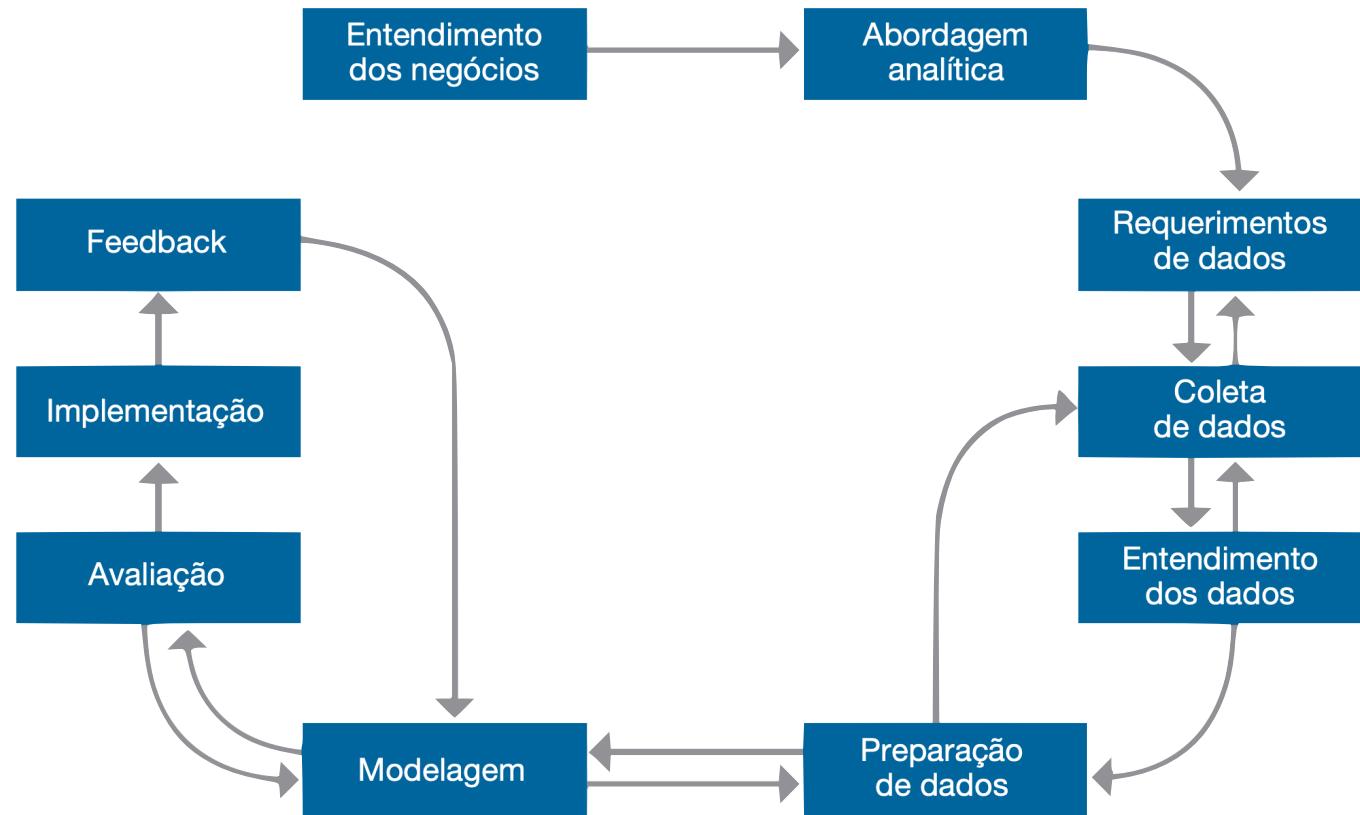
- 1 : a body of methods, rules, and postulates employed by a discipline : a particular procedure or set of procedures

Source www.m-w.com

WHY SHOULD I CARE ABOUT METHODOLOGY?

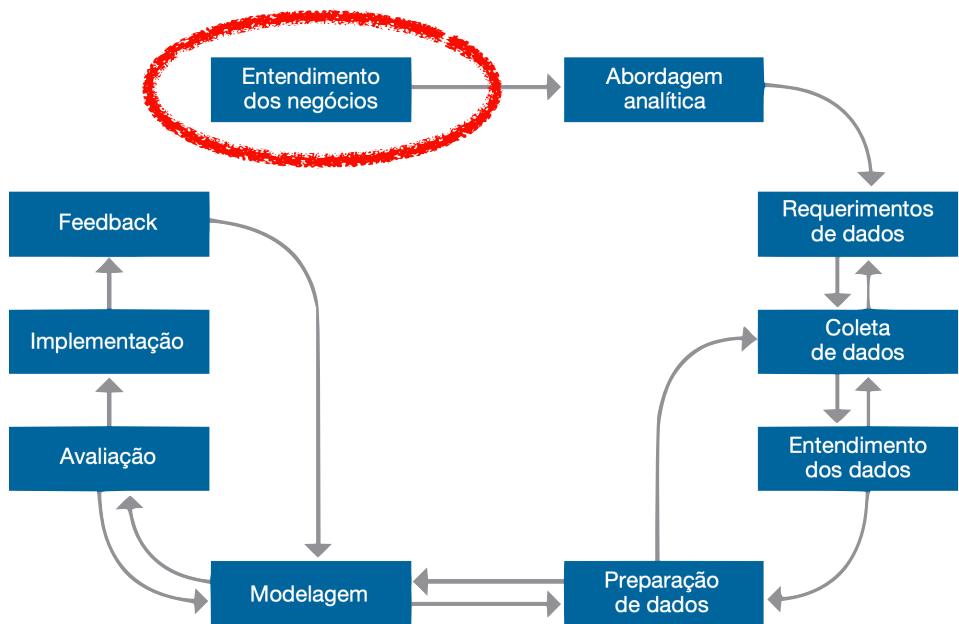
- ▶ Methodology is a strategy to drive the processes
 - ▶ Do you care about strategy in your company?
 - ▶ Do you care about processes in your job?
- ▶ Concerning Data Science, methodology is a step by step process from the identification of your problem to the final implementation of the solution.

DE A A Z



Source: Rollings John B., Foundational methodology for data science. IBM white paper
<https://www.ibm.com/downloads/cas/B1WQ0GM2>

ENTENDIMENTO DOS NEGÓCIOS



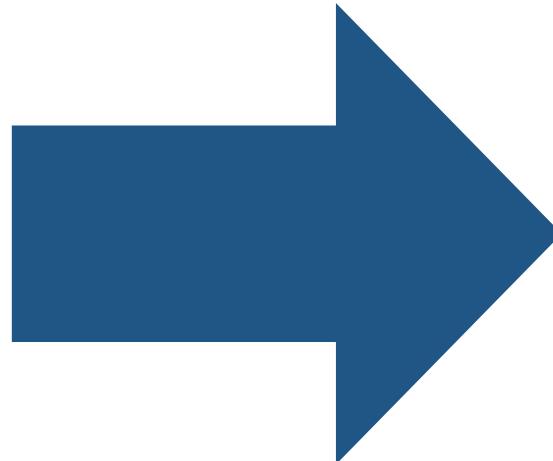
Qual é o problema
que estou tentando
resolver?

Source: Rollings John B., Foundational methodology for data science. IBM white paper
<https://www.ibm.com/downloads/cas/B1WQ0GM2>

ANIMAL SHELTER



Qual é o problema que estou tentando resolver?

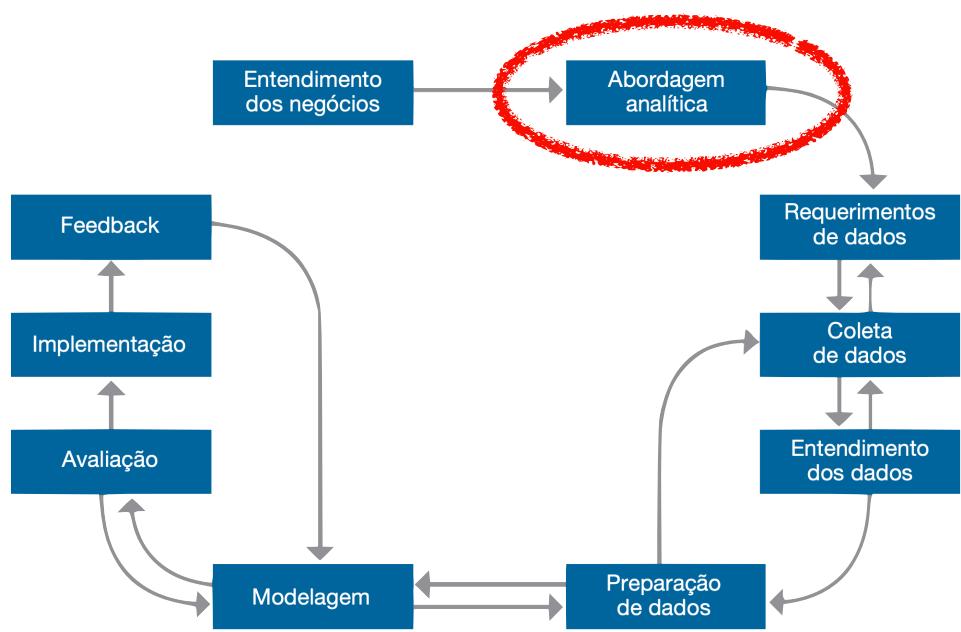


OutcomeType
Return_to_owner
Euthanasia
Adoption
Transfer
Died

kaggle

Source: <https://www.kaggle.com/c/shelter-animal-outcomes>

ABORDAGEM ANALÍTICA



Como eu posso usar dados para resolver esse problema?

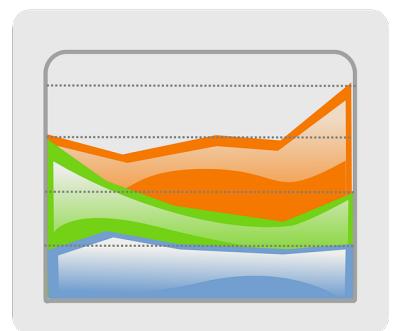
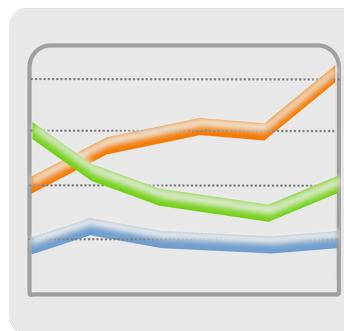
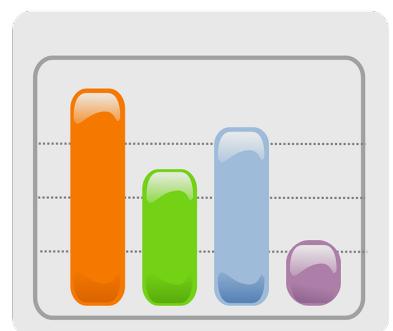
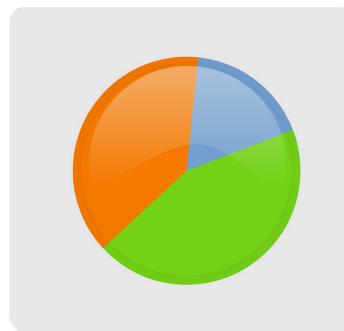


A resposta para esta pergunta depende do entendimento do negócio.

Source: Rollings John B., Foundational methodology for data science. IBM white paper
<https://www.ibm.com/downloads/cas/B1WQ0GM2>

ABORDAGENS ANALÍTICAS

- ▶ Descritiva / Diagnóstica
 - ▶ Status Quo
 - ▶ O que aconteceu?
 - ▶ Por que isso está acontecendo?
- ▶ Preditiva (Forecasting)
 - ▶ Tendências
 - ▶ Futuro
- ▶ Prescritiva
 - ▶ Como eu resolvo esse problema?



DESCRIPTIVE AND PREDICTIVE ANALYSIS

- ▶ Descriptive analytics can tell you what has happened or what is happening now.
- ▶ When there is a need to get a comprehensive picture of what is going on in the organization, descriptive analysis is applied.
- ▶ It describes the past and summarizes basic data that describes analysis different aspects of the business. Reports provide information about sales, customers, operations, finance, and correlations between the data. Questions about previous business dealings, details of product users, the status of certain outlets, and sales indicators of supplies may be answered.

Source: Zhang, Arthur. Data Analytics: Practical Guide to Leveraging the Power of Algorithms, Data Science, Data Mining, Statistics, Big Data, and Predictive Analysis to Improve Business, Work, and Life (pp. 150-151). Kindle Edition.

DESCRIPTIVE AND PREDICTIVE ANALYSIS

- ▶ Predictive analysis asks questions about the future.
- ▶ Predictive analytics is used when information about forthcoming commercial activity is required. It provides organizations with actionable insights based on data, and estimates future outcomes. It predicts customer behavior in sales and marketing, demands for operation, or determines risk factors for finance.
- ▶ This type of analytic forecasting constructs investigative outlines on the various stages of industry. These forecasts may comprise of intelligence regarding a consumer or commodity that will make business decisions clear. Using information such as customer and employer behavior, the highs and lows of markets, and other such related data, accurate judgements can be made for prospective business direction.

Source: Zhang, Arthur. Data Analytics: Practical Guide to Leveraging the Power of Algorithms, Data Science, Data Mining, Statistics, Big Data, and Predictive Analysis to Improve Business, Work, and Life (pp. 150-151). Kindle Edition.

QUANDO A PERGUNTA É ...

Porque isso está acontecendo?



OU

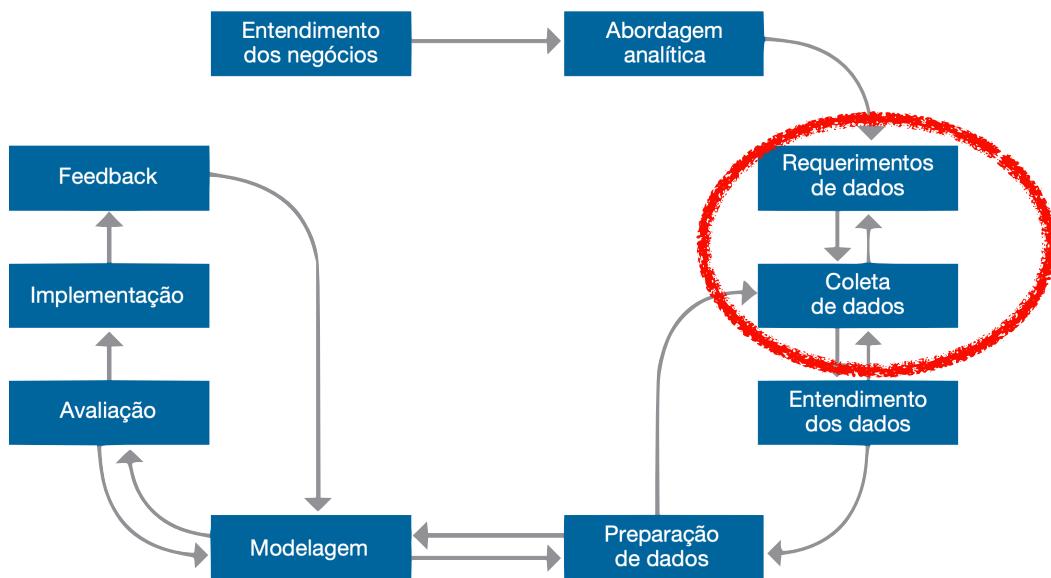
Determinada parte
do processo afeta ou
não a satisfação?

DE QUAL ABORDAGEM
ESTAMOS FALANDO?

NO ANIMAL SHELTER PROBLEM, QUAL
ABORDAGEM VOCÊ USARIA?



NECESSIDADE E COLETA DE DADOS



Quais dados eu preciso?
Como eu vou coletar esses
dados?

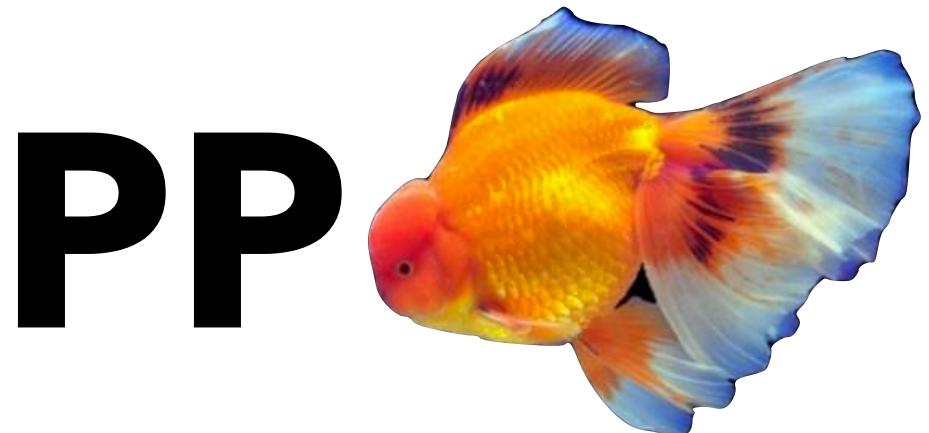
Source: Rollings John B., Foundational methodology for data science. IBM white paper

NO ANIMAL SHELTER PROBLEM, QUAIS
DADOS VOCÊ COLETARIA?



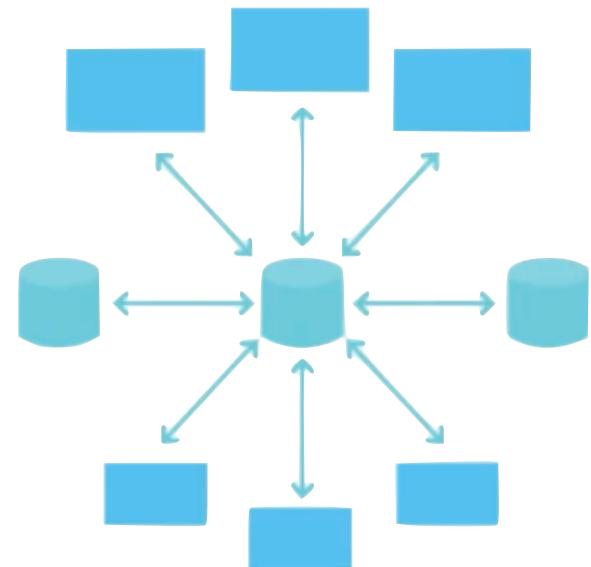
DEFINIR E SELECIONAR O ESCOPO DOS DADOS (CORTE)

- ▶ Todos os tipo e raças de animais? Até o PPGOLD Fish?
- ▶ Corte temporal - somente animais que chegaram após a reestruturação ?
- ▶ Vamos considerar todas as saídas ?
 - ▶ Qual as consequências disso?
- ▶ Existe algum critério de exclusão?



FONTES DE DADOS

- ▶ Portal de vendas
- ▶ Warehouse Management System (WMS)
- ▶ ERP
- ▶ Formulários

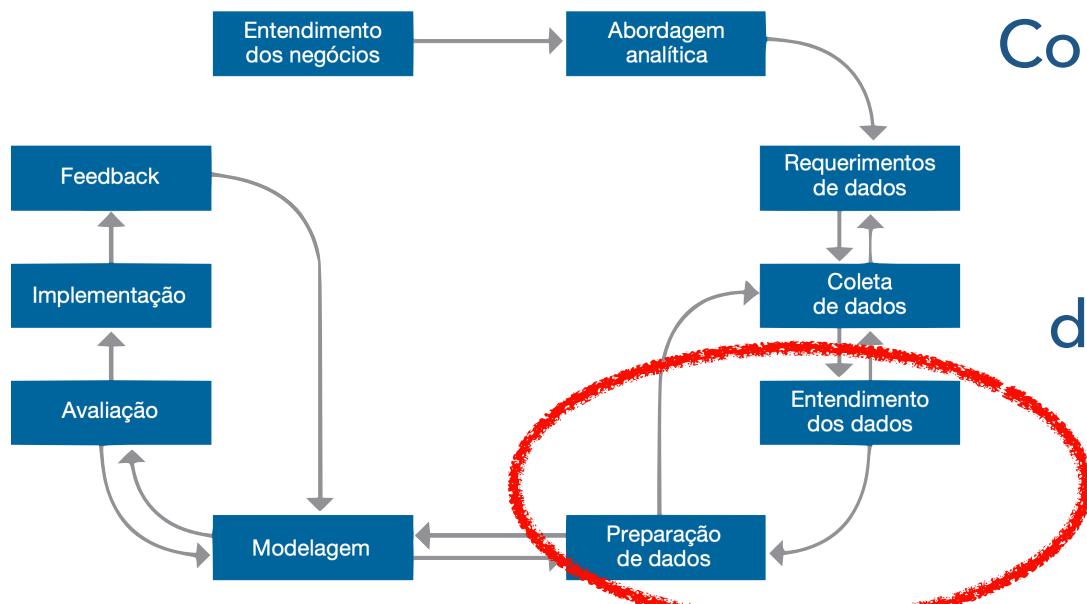


EXISTEM OUTROS DADOS QUE VCS ACHAM PERTINENTES?



> DATA NOT
AVAILABLE

DO ENTENDIMENTO A PREPARAÇÃO



Como assim “limpar” os dados?

Ao preparar e manipular os dados não introduzo um viés?

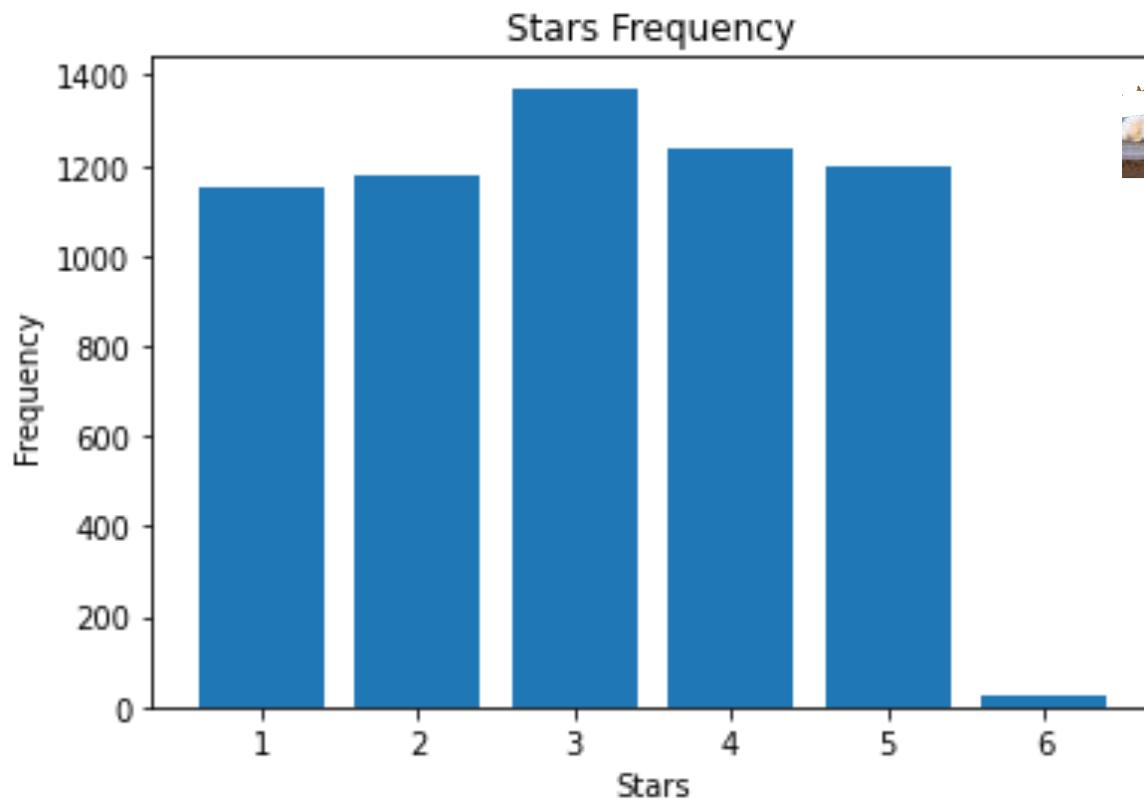
Source: Rollings John B., Foundational methodology for data science. IBM white paper

BD MERGING



Só quem é da NOVA GERAÇÃO vai entender essa piada!

EXEMPLO – AVALIAÇÃO NO PROCESSO DE ADOÇÃO



NOTA DE 6 ESTRELAS,
UAU!



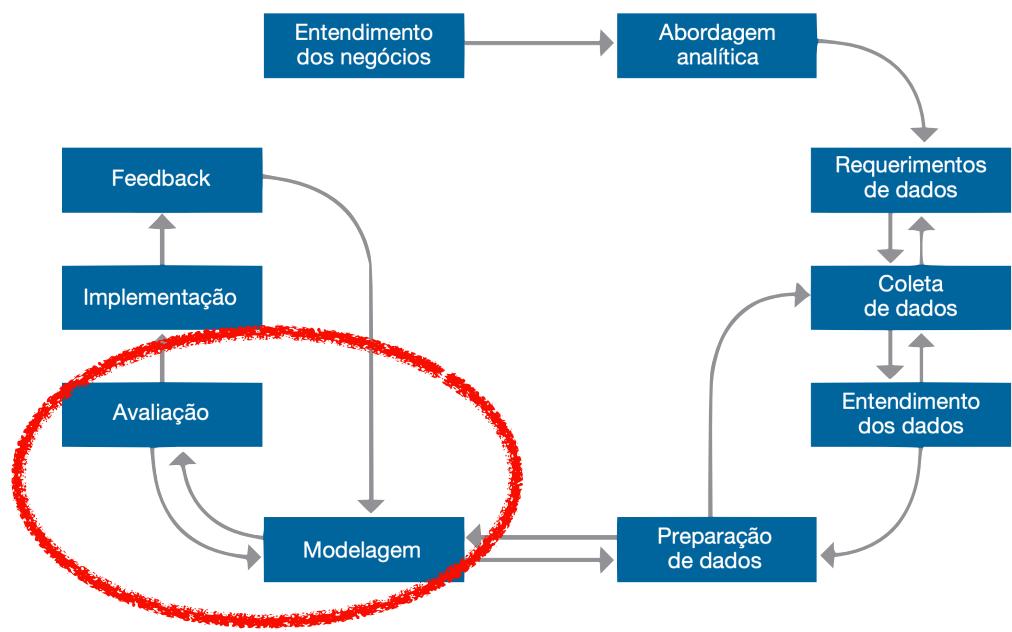
NÃO PODE!
ESTÁ ERRADO!

PRECISA LIMPAR!

LIMPEZA DOS DADOS



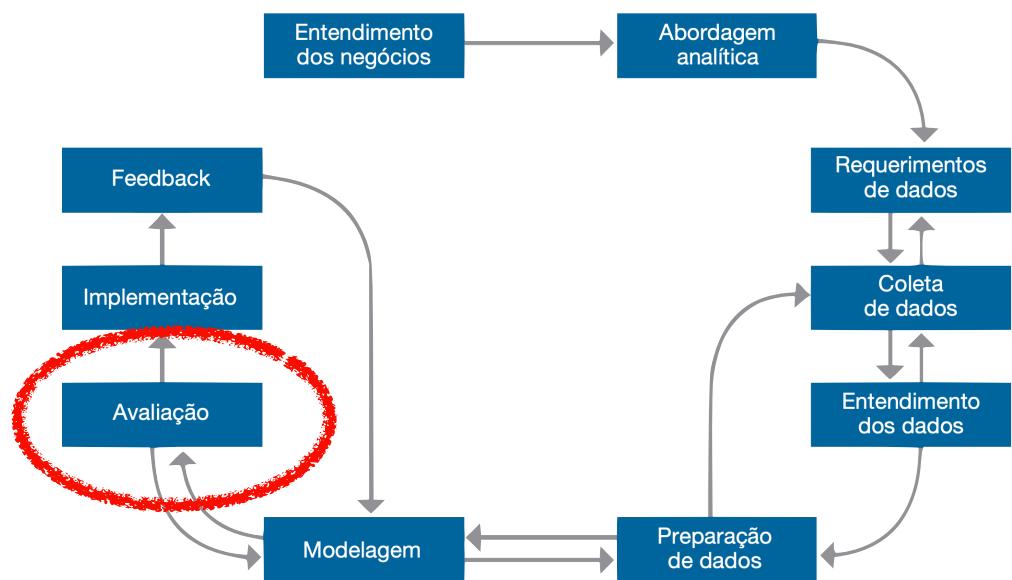
O MODELO (FINALMENTE)



Como eu posso visualizar os resultados de maneira rápida?

Source: Rollings John B., Foundational methodology for data science. IBM white paper

O MODELO (FINALMENTE)

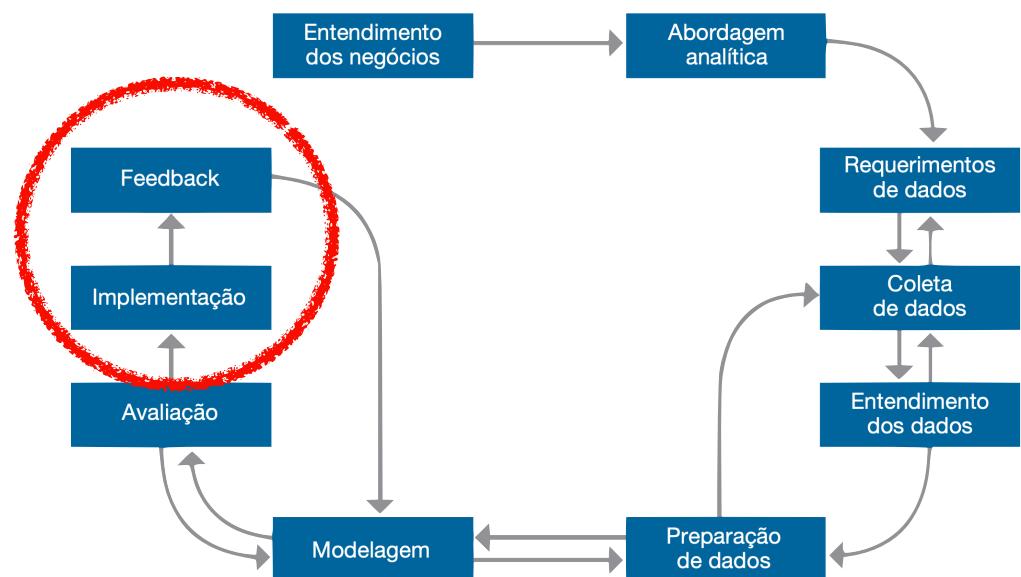


O modelo escolhido realmente responde o quer quero saber?

Tem como melhorar esse modelo?

Source: Rollings John B., Foundational methodology for data science. IBM white paper

O MODELO (FINALMENTE)



O modelo é aplicável no dia-a-dia?

Você pode obter um feedback construtivo para responder à pergunta?

Source: Rollings John B., Foundational methodology for data science. IBM white paper



JUPYTER NOTEBOOKS



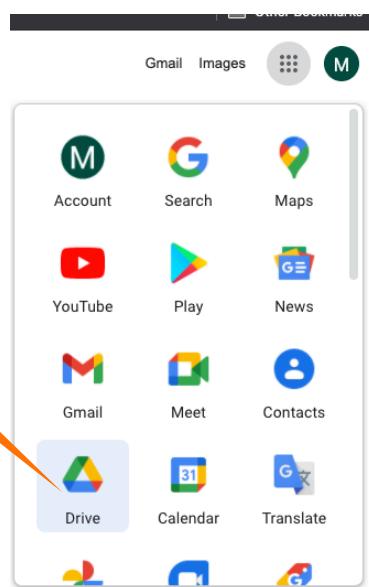
ANACONDA[®]



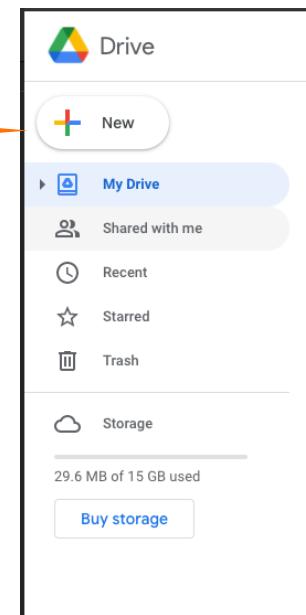
GOOGLE COLAB
JUPYTER NOTEBOOKS

VOCÊ IRÁ PRECISAR DE UMA CONTA GOOGLE

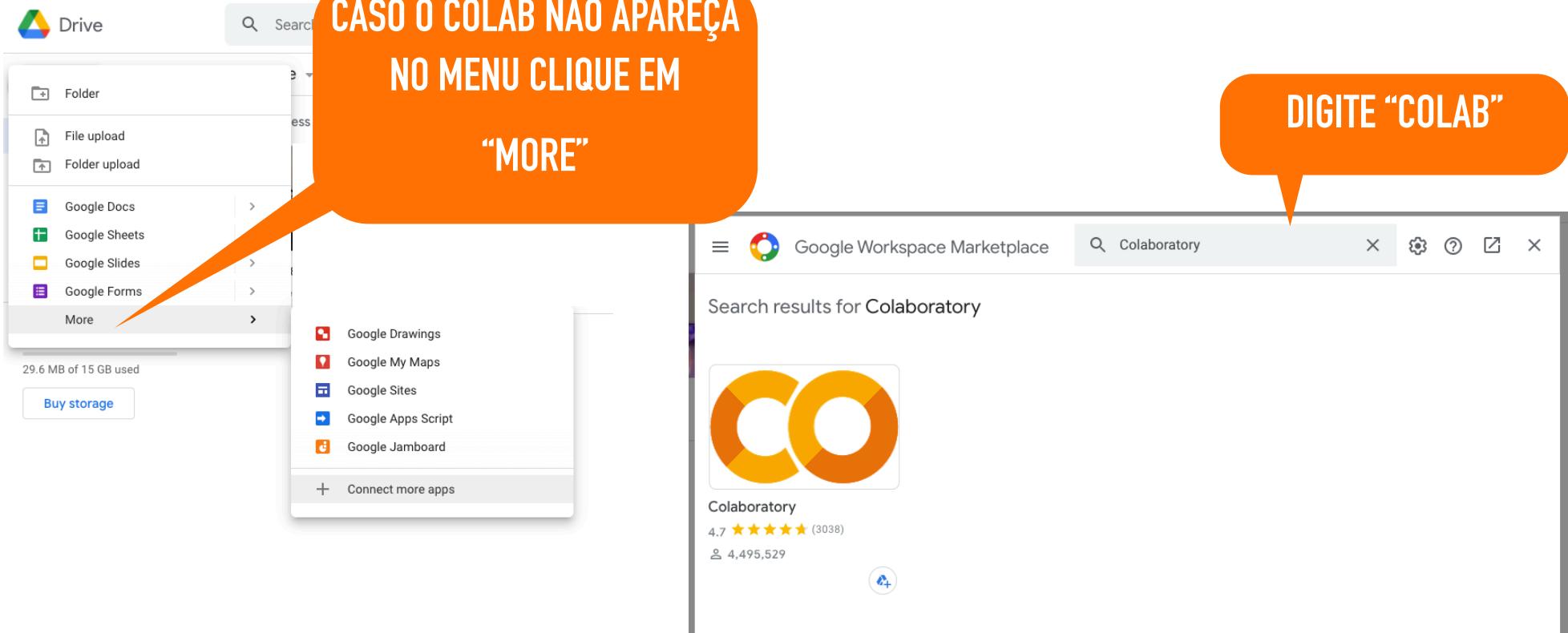
FAÇA LOGIN E CLIQUE
NO GOOGLE DRIVE



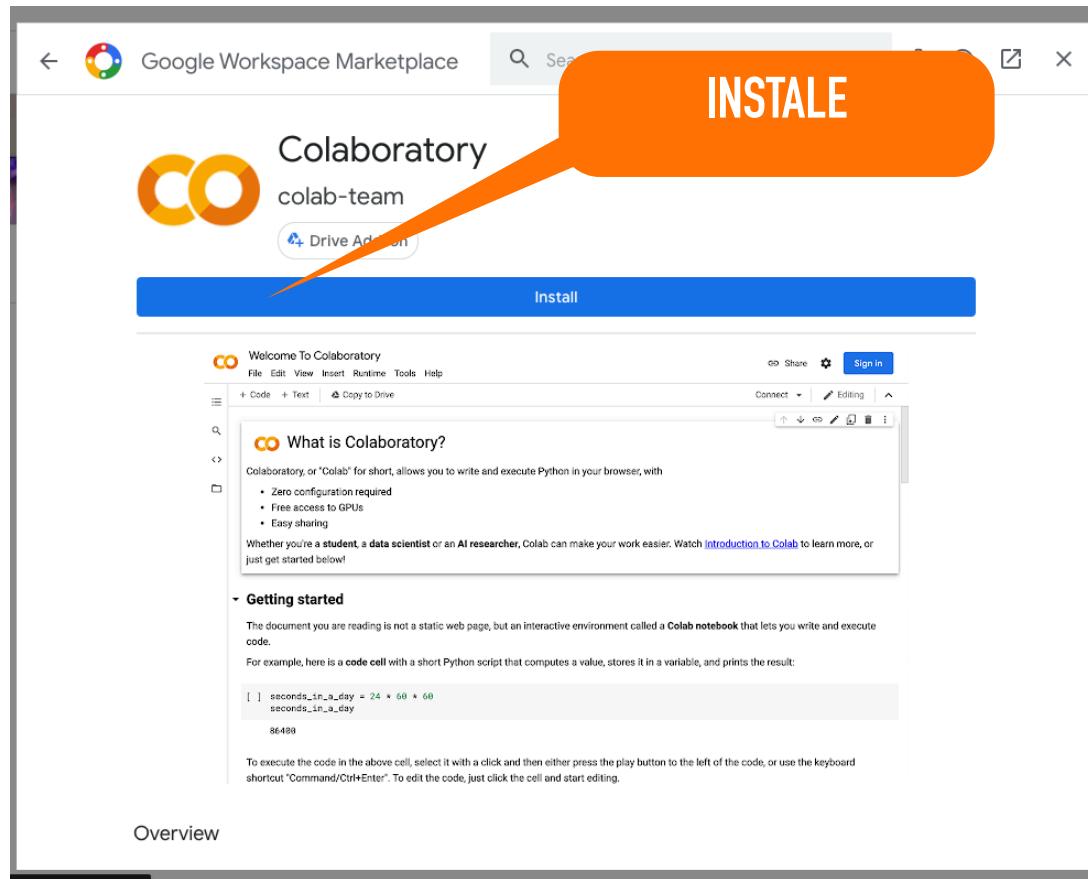
CLIQUE EM NEW



DATA SCIENCE - PPGOLD



DATA SCIENCE - PPGOLD



Google Workspace Marketplace

Search

INSTALE

Colaboratory
colab-team

Drive Add-on

Install

Welcome To Colaboratory

File Edit View Insert Runtime Tools Help

Share Sign in

What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with:

- Zero configuration required
- Free access to GPUs
- Easy sharing

Whether you're a student, a data scientist or an AI researcher, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

Getting started

The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

```
[ ] seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day
86400
```

To execute the code in the above cell, select it with a click and then either press the play button to the left of the code, or use the keyboard shortcut "Command/Control+Enter". To edit the code, just click the cell and start editing.

Overview

Colaboratory has been installed!

Where to find Colaboratory

You can find Colaboratory by choosing "Open with" in Drive:



review...

Open with

DONE

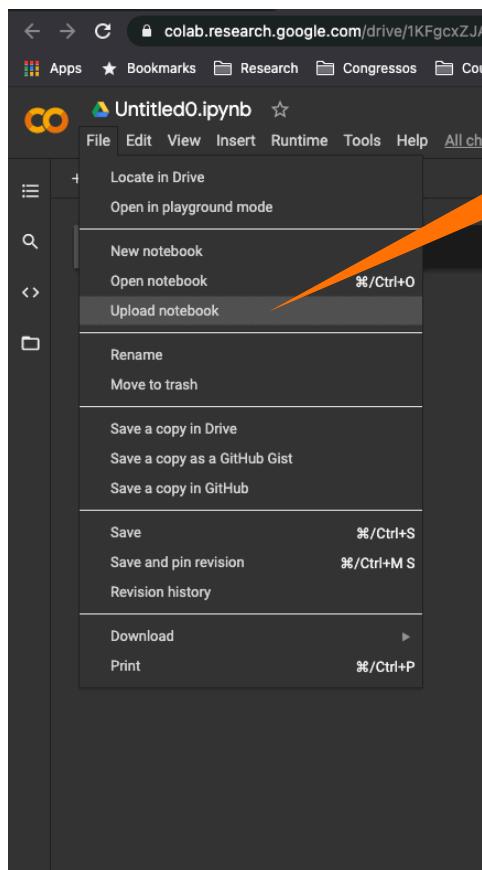
PRONTO!

DATA SCIENCE - PPGOLD

The screenshot shows the Google Drive interface. On the left, there's a sidebar with options like 'Folder', 'File upload', 'Folder upload', and a list of Google apps: Google Docs, Google Sheets, Google Slides, Google Forms, and 'More'. Below this is a storage status message: '29.6 MB of 15 GB used' and a 'Buy storage' button. A large orange callout bubble points to the 'More' menu, which is expanded to show 'Google Drawings', 'Google My Maps', 'Google Sites', 'Google Apps Script', 'Google Colaboratory' (which is highlighted with a gray background), 'Google Jamboard', and a 'Connect more apps' option.

ELE IRÁ APARECER AQUI

DATA SCIENCE - PPGOLD



CLIQUE EM
FILE -> UPLOAD NOTEBOOK

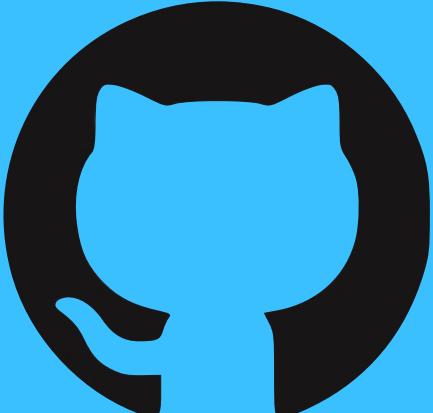


ESCOLHA O ARQUIVO!
VOCÊ ESTÁ PRONTO PARA
COMEÇAR

MAIS DETALHES ?

ESTE TUTORIAL DA ANNE BONNER É ÓTIMO:

<https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c>



Github

VERSION CONTROL SYSTEM

- ▶ Permite que você acompanhe as mudanças em seus documentos.
- ▶ Facilita a recuperação de versões mais antigas de seu documento se você cometer um erro
- ▶ Facilita muito a colaboração com outras pessoas.
- ▶ Muito usado para programação, mas pode ser usado para documentação também



[HTTPS://TRY.GITHUB.IO/](https://try.github.io/)
