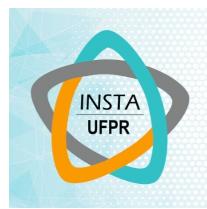
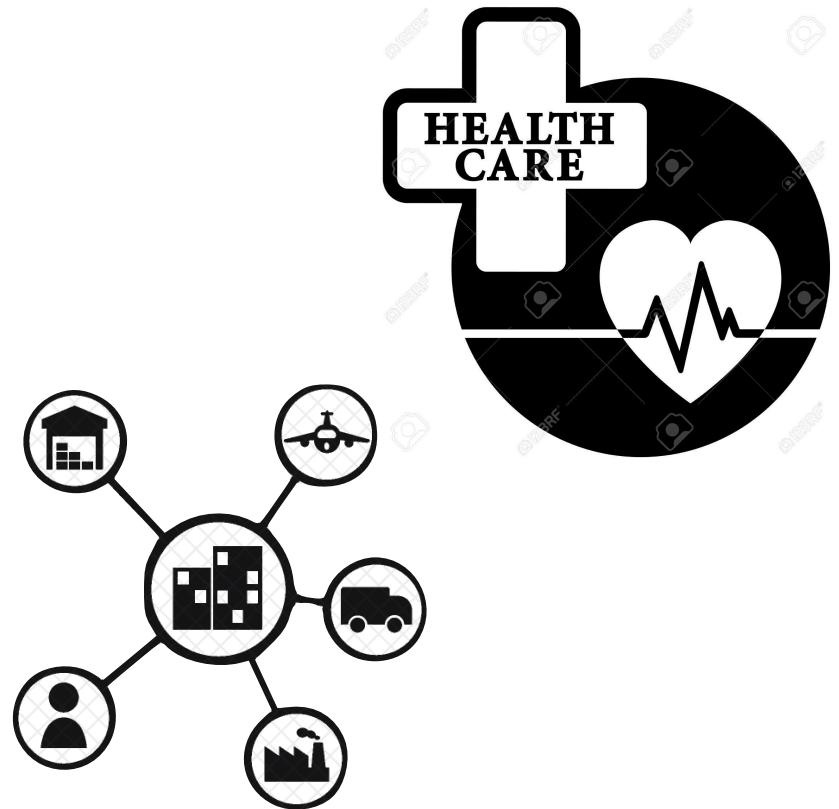
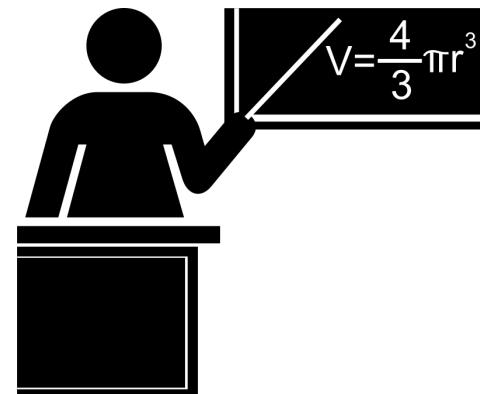
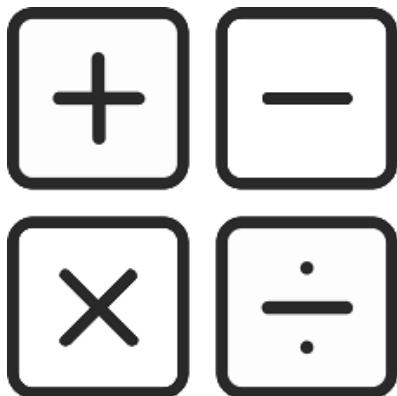


PROF. EDUARDO PÉCORA

SUPPLY 4.0 DATA SCIENCE

Just little bit about myself

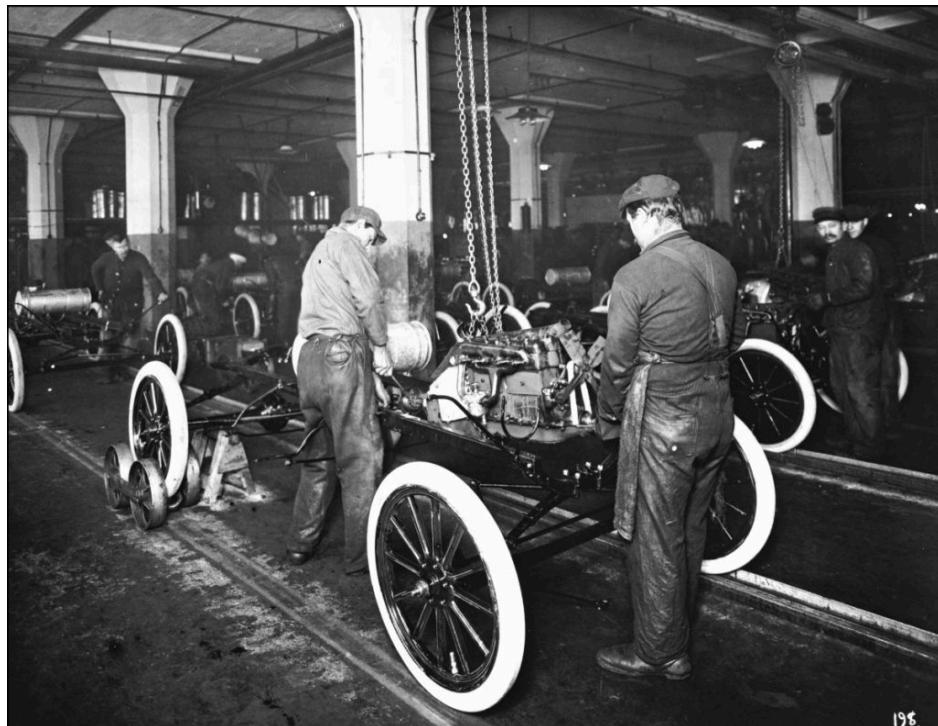


A black and white aerial photograph of a massive concrete dam. The dam is a thick, curved wall that tapers towards the top. A narrow walkway runs along the top edge of the dam, supported by a metal railing. A single person stands on the walkway, appearing very small against the scale of the structure. The water of the reservoir behind the dam is dark and calm.

PROF. EDUARDO PÉCORA

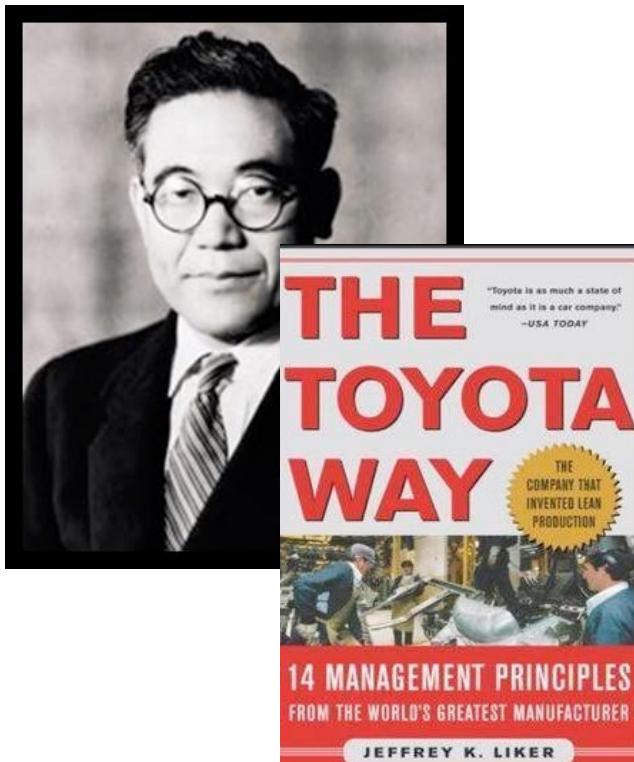
SUPPLY 4.0

HISTÓRIA DO SUPPLY-CHAIN



**Foco Dominante:
Capacidade de
Produção**

HISTÓRIA DO SUPPLY-CHAIN



Foco Dominante: Eficiência Interna



HISTÓRIA DO SUPPLY-CHAIN



Foco Dominante: Eficiência Externa



Primeiros
conceitos de
Logística



HISTÓRIA DO SUPPLY-CHAIN

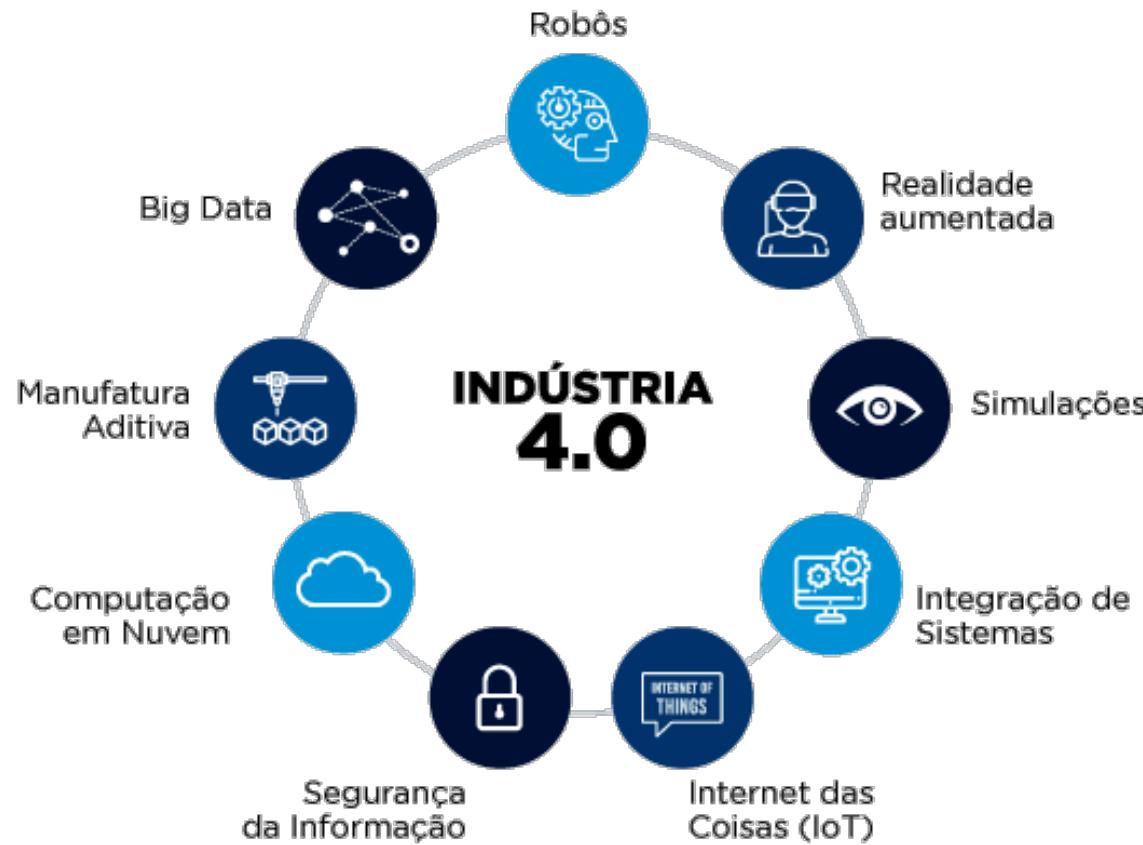
Foco Dominante: Eficiência Externa



Primeiros conceitos de
Cadeia de
Suprimentos



INDUSTRIA 4.0



Foco Dominante:
Transformação
digital da
INDÚSTRIA



SUPER CASE - INSTA

SUPPLY CHAIN 4.0



Foco Dominante:
Transformação digital
da CADEIA DE
SUPRIMENTOS



SUPER CASE - INSTA

MAIS SOBRE SUPPLY 4.0

Operations and Supply Chain Strategy in the Industry 4.0 Era

Concepts and Implementation

Guilherme Frederico



Prof. Eduardo Pécora INSTA / UFPR

Supply Chain 4.0: concepts, maturity and research agenda

Guilherme F. Frederico

School of Management, Federal University of Paraná (UFPR), Curitiba, Brazil

Jose Arturo Garza-Reyes and Anthony Anosike

Centre for Supply Chain Improvement, the University of Derby, Derby, UK, and

Vikas Kumar

Bristol Business School, University of the West of England, Bristol, UK

Supply Chain Management: An International Journal

Volume 25 · Number 2 · 2020 · 262–282

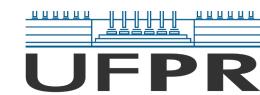


Supply Chain Management: An International Journal

25/2 (2020) 262–282

© Emerald Publishing Limited [ISSN 1359-8546]

[DOI [10.1108/SCM-09-2018-0339](https://doi.org/10.1108/SCM-09-2018-0339)]



A black and white aerial photograph of a massive concrete dam. The dam is a thick, curved wall that slopes down to a wide base. A paved walkway runs along the top of the dam, with a metal railing. Two small figures of people are standing on the walkway, one near the center and another further down towards the right. The water of the reservoir behind the dam is dark and calm.

PROF. EDUARDO PÉCORA

DATA SCIENCE

MATERIAL DO SUPER CASSE – SUPPLY 4.0

Arquivos:

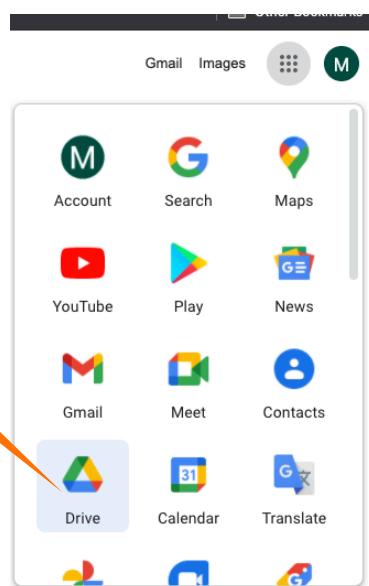
[SuperCase_01_2.ipynb](#)
[SuperCase_data_6000_3.csv](#)



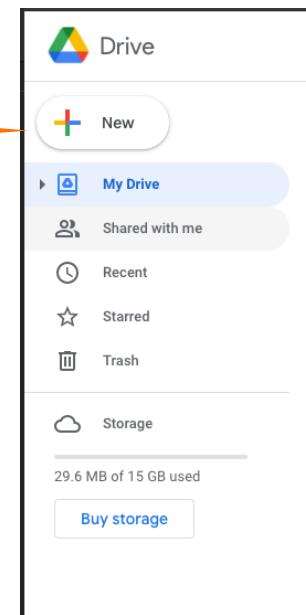
GOOGLE COLAB
JUPITER NOTEBOOKS

Você irá precisar de uma conta GOOGLE

FAÇA LOGIN E CLIQUE
NO GOOGLE DRIVE



CLIQUE EM NEW



SUPER CASE - INSTA



SUPER CASE - INSTA

Google Workspace Marketplace

Search: 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
```

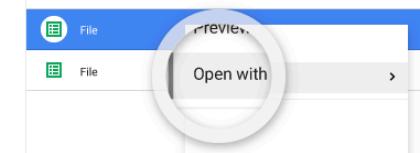
To execute the code in the above cell, select it with a click and then either press the play button to the left of the code, or use the keyboard shortcut "Command/Ctrl+Enter". To edit the code, just click the cell and start editing.

Overview

Colaboratory has been installed!

Where to find Colaboratory

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



PRONTO!



SUPER CASE - INSTA

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 message about storage usage: '29.6 MB of 15 GB used' and a 'Buy storage' button. A callout bubble points from the word 'AQUI' in the text below to the 'Google Colaboratory' option in the 'More' menu. The 'More' menu also includes 'Google Drawings', 'Google My Maps', 'Google Sites', 'Google Apps Script', 'Google Jamboard', and a 'Connect more apps' option.

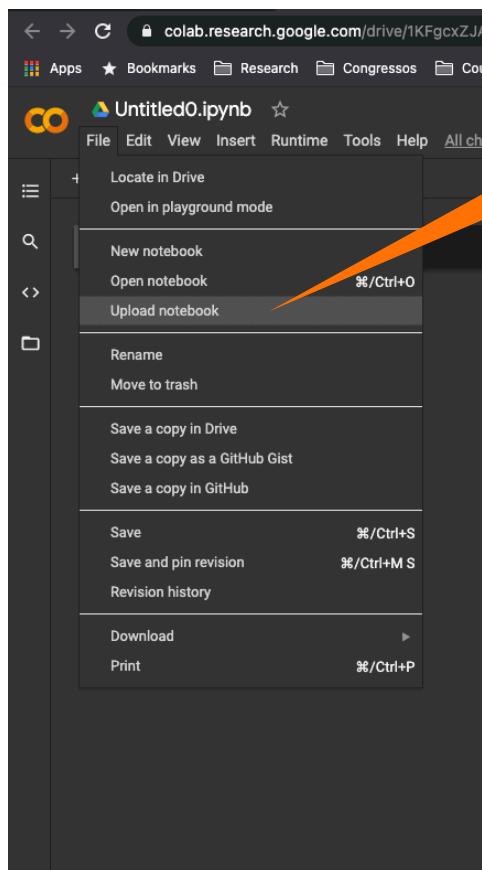
ELE IRÁ APARECER AQUI

- Google Drawings
- Google My Maps
- Google Sites
- Google Apps Script
- Google Colaboratory
- Google Jamboard

+ Connect more apps



SUPER CASE - INSTA



CLIQUE EM
FILE -> UPLOAD NOTEBOOK



ESCOLHA O ARQUIVO!
VOCÊ ESTÁ PRONTO PARA
COMEÇAR O SUPER CASE

MAIS DETALHES ?

ESTE TUTORIAL DA ANNE BONNER É ÓTIMO:

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

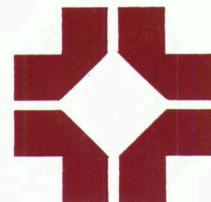


PROJETOS ATUAIS



CIRRELT

Centre interuniversitaire
de recherche
sur les réseaux d'entreprise,
la logistique et le transport



HOSPITAL DE CLÍNICAS
UNIVERSIDADE FEDERAL DO PARANÁ



Patient Prioritization

Prof. J.E.Pécora

Doc. Ana Lopes

Doc. Amanda Macedo

Doc. Débora Lemos



Previsão de Vendas com Deep Learning com variação da periodicidade

**Prof. J.E. Pécora
M.Sc. Lucas Marchioro**



CIRRELT

Centre interuniversitaire
de recherche
sur les réseaux d'entreprise,
la logistique et le transport



Roteamento para entrega de rações

**Prof. J.E. Pécora
M.Sc. Moisé K. Tokarski**



CIRRELT

Centre interuniversitaire
de recherche
sur les réseaux d'entreprise,
la logistique et le transport



Framework para implementação de projetos da indústria 4.0

**Prof. J.E. Pécora
M.Sc. Letícia H. Radi Mira**



CIRRELT

Centre interuniversitaire
de recherche
sur les réseaux d'entreprise,
la logistique et le transport



Algoritmo do tipo grey wolf optimizer para o problema de transporte com designação e carregamento

**Prof. J.E. Pécora
M.Sc. Felipe Kauai Pereira**



Intra Hospital Patient Transportation

Prof. J.E. Pécora

Prof. Angel Ruiz

Prof. Nathalia Ortiz

Doc. Vinicius Ton

M.Sc. Lorena Carvalho



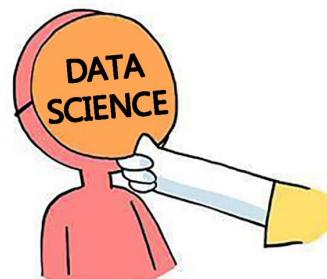
DATA SCIENCE

INTRODUÇÃO

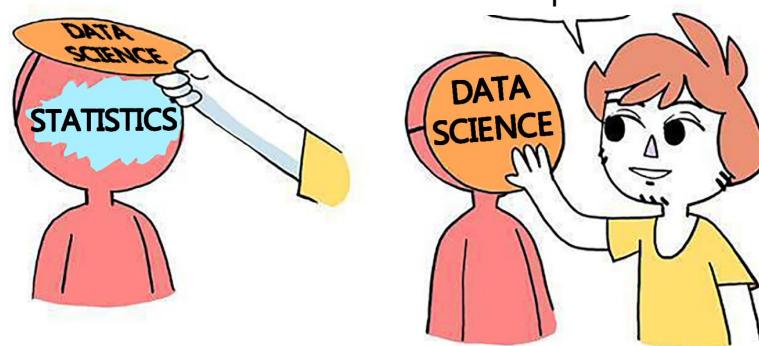
Oi Data Science porque
você sempre usa esta
máscara?

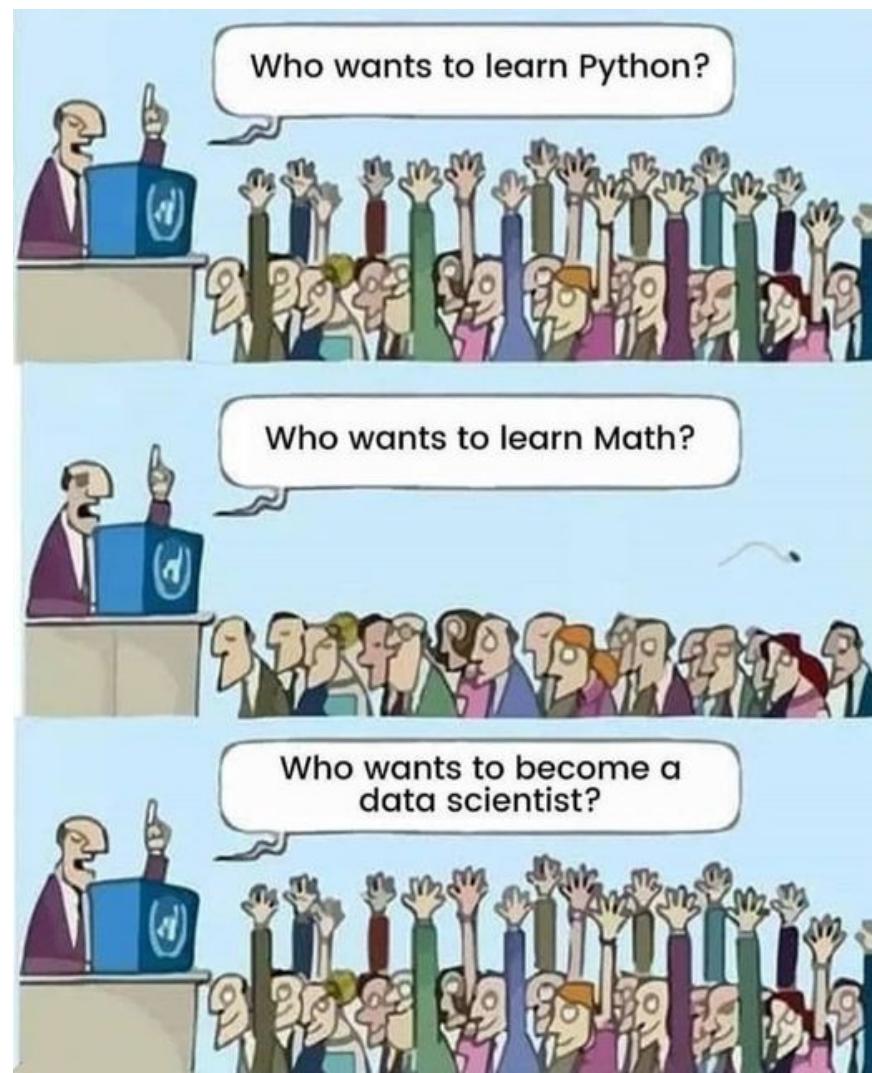


@statsanddatasci



Vamos
deixá-la
aqui!



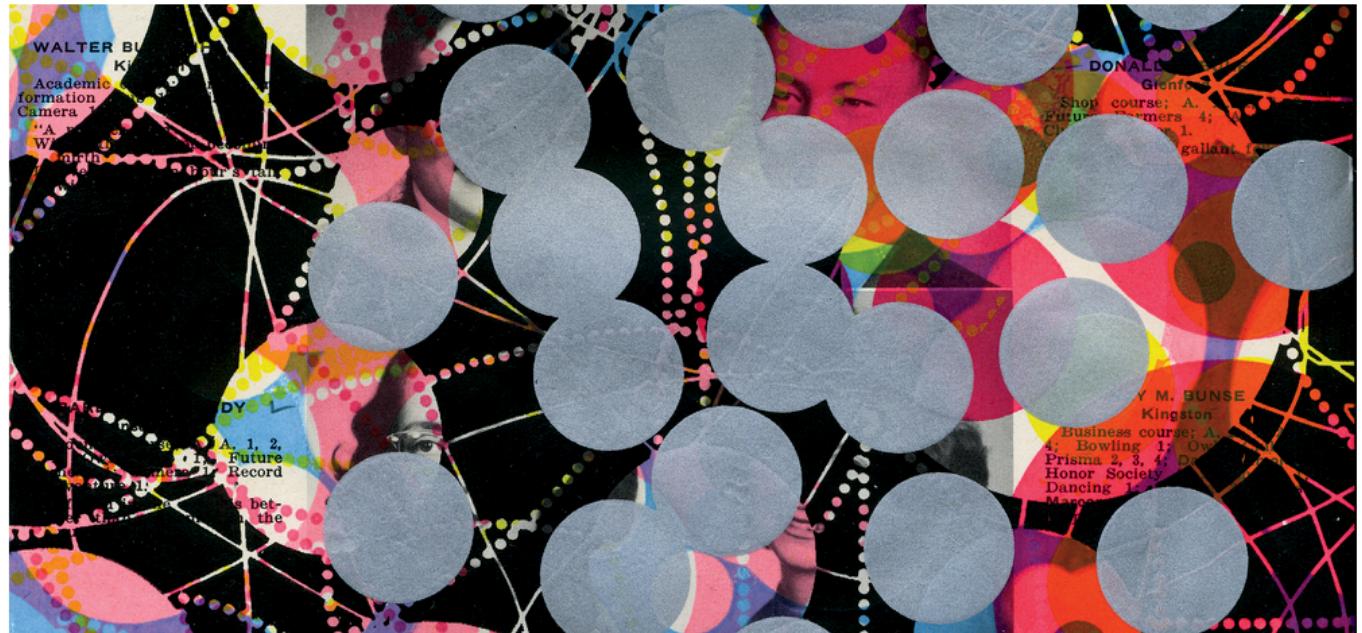


Data

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)



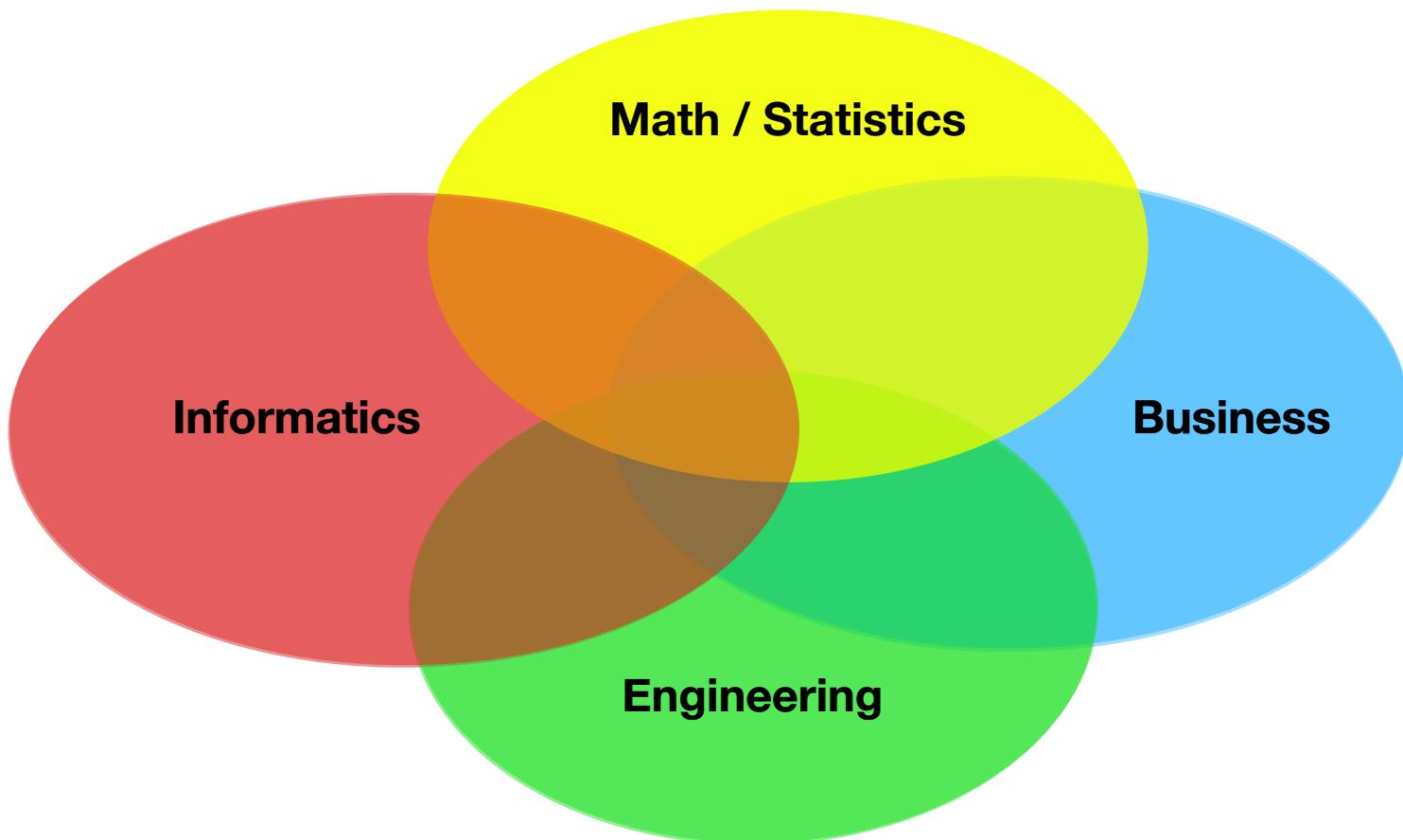
Artwork: Tamar Cohen, Andrew J Buboltz, 2011, silk screen on a page from a high school yearbook, 8.5" x 12"

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

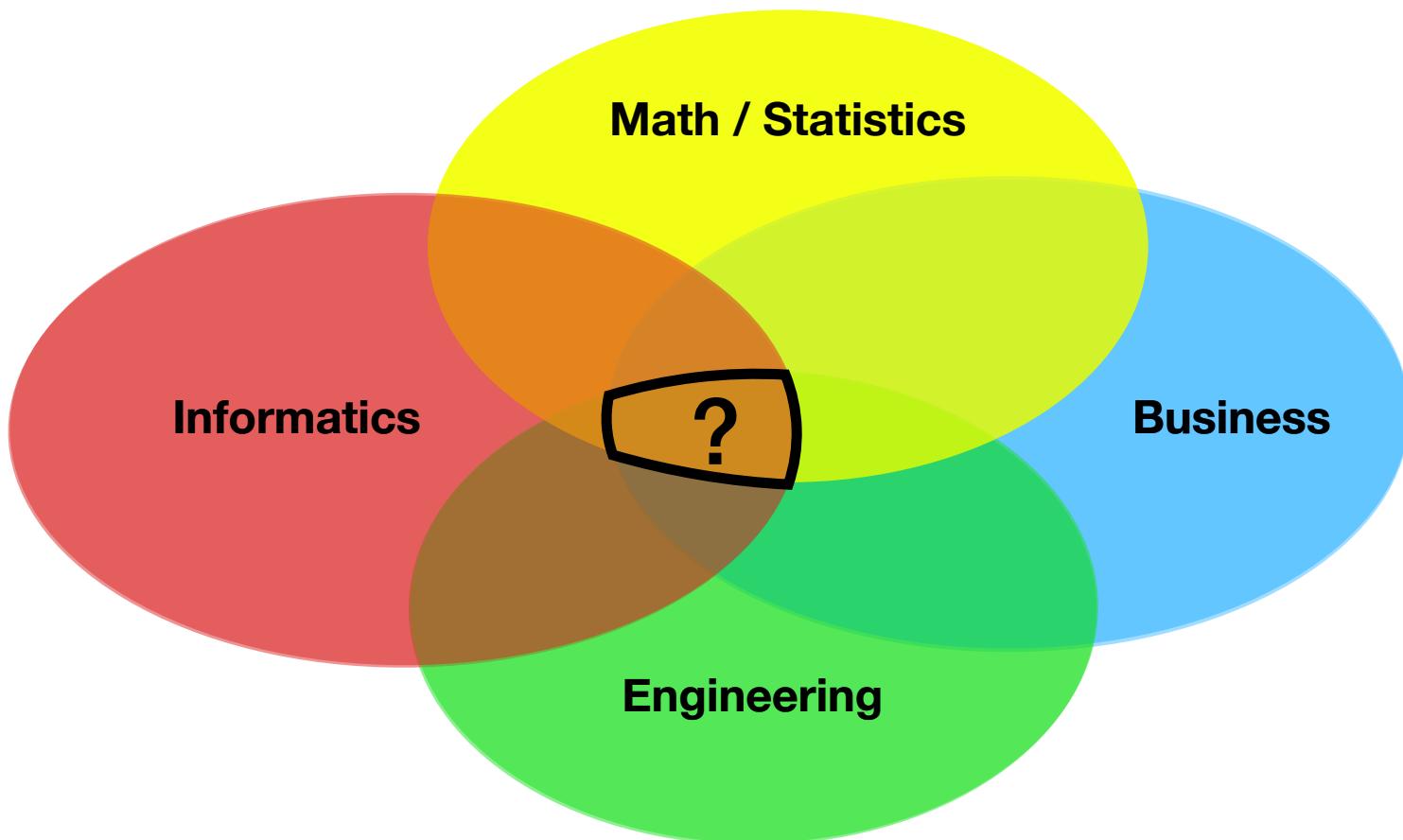
Posicionando o Data Science



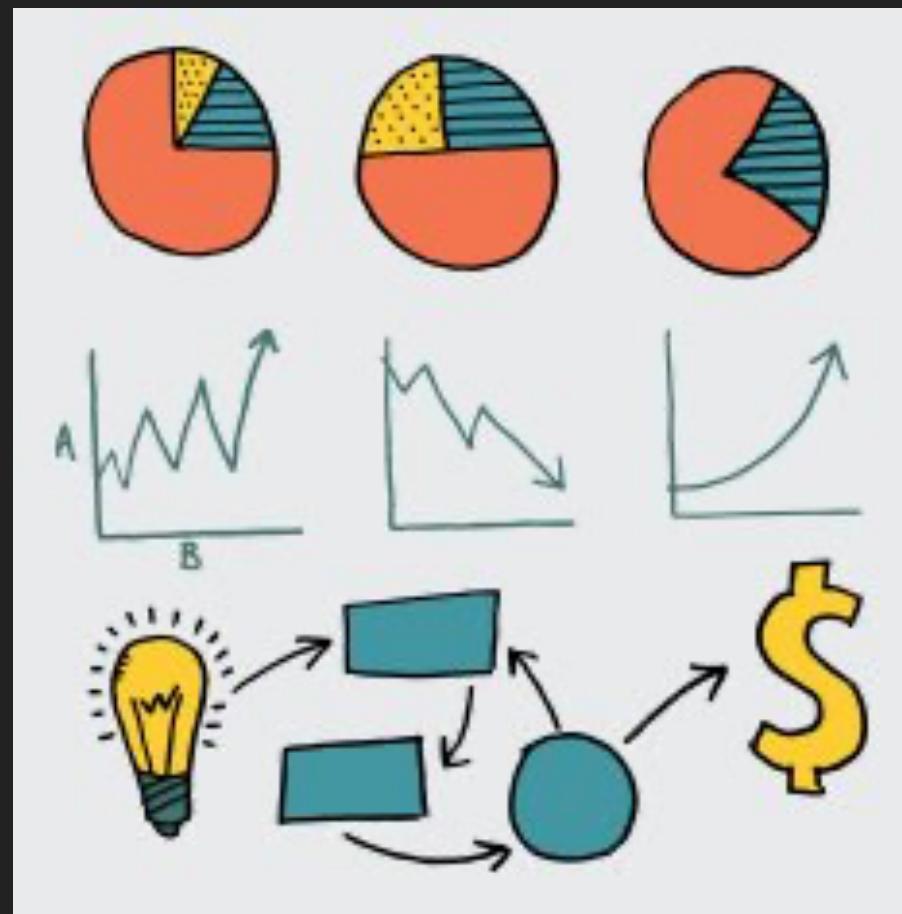
Data Science?



Data Science?



OBJETIVO DO DATA SCIENCE



TEXT

MAS EU POSSO SER SÓ UM USUÁRIO DO
DATA SCIENCE?

John Doe

TEXT

MAS 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

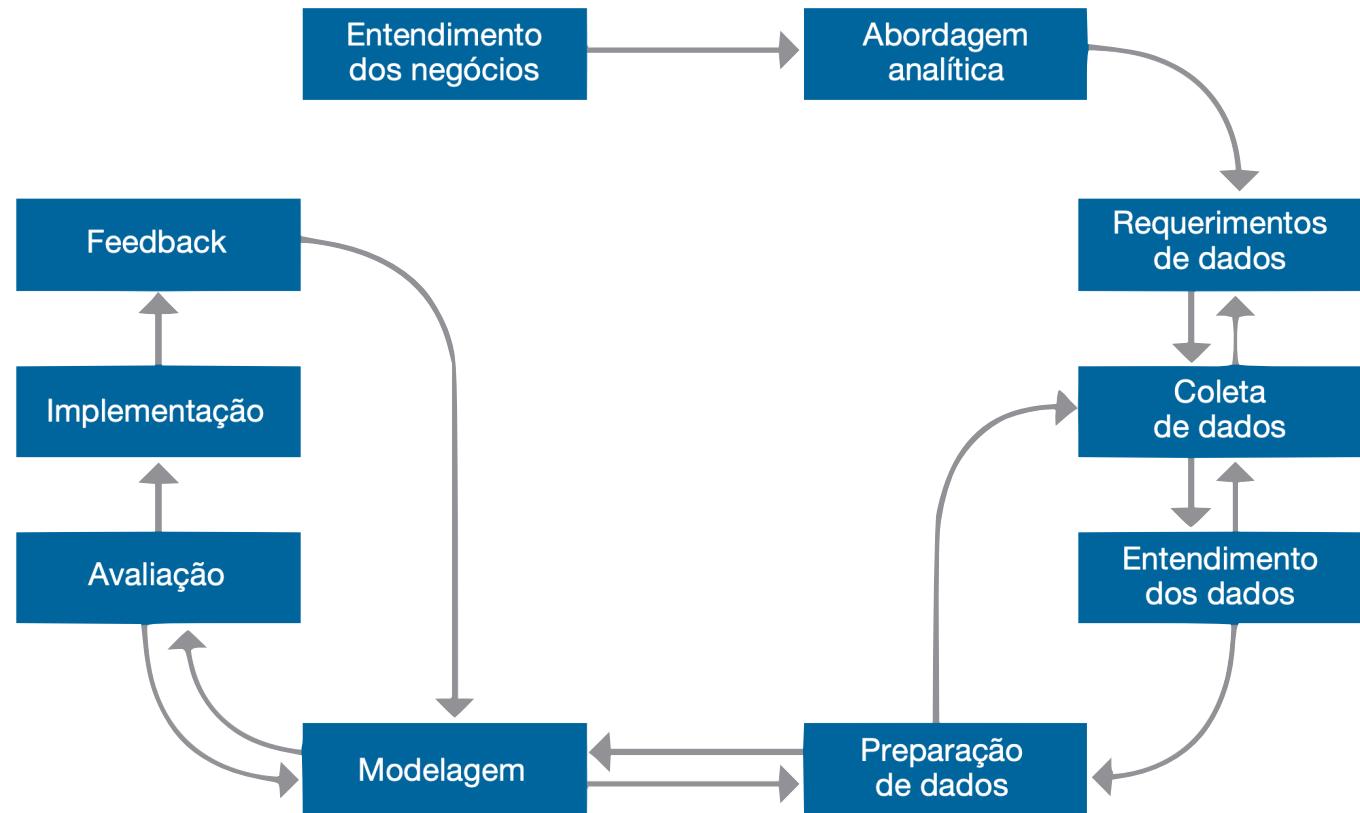
Source www.m-w.com

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.

SUPER CASE - INSTA

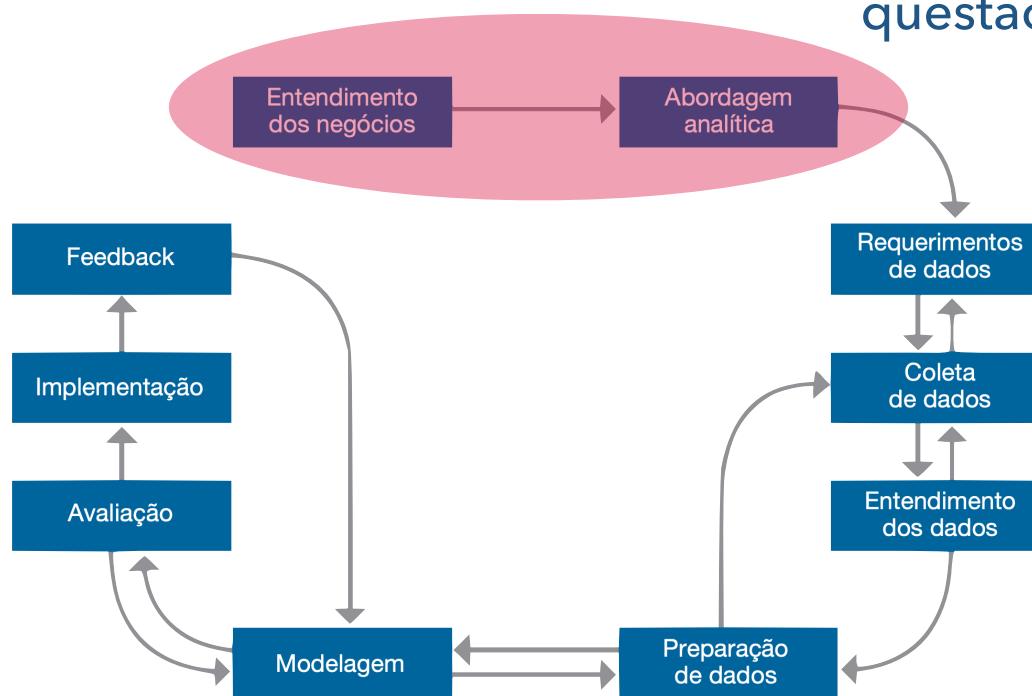
DE A A Z



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

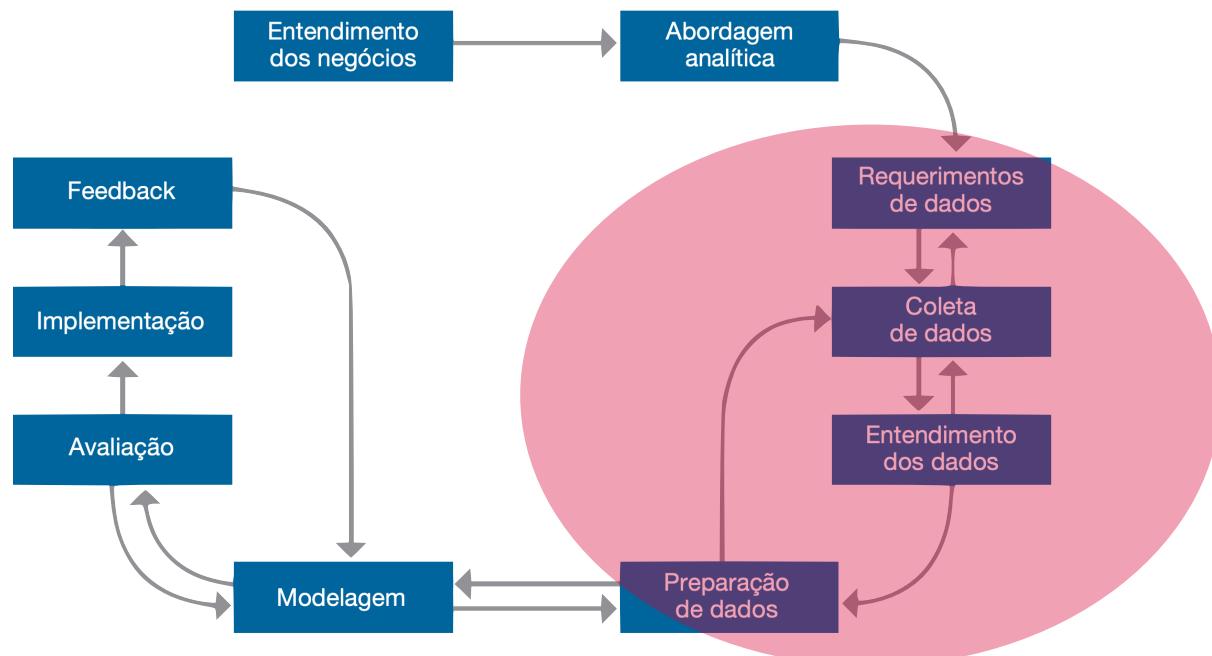
AS 10 QUESTÕES-CHAVE

- 1) Qual o problema que eu quero resolver?
- 2) Como eu posso usar dados para resolver esta questão?



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

AS 10 QUESTÕES-CHAVE



Trabalhando com os dados:

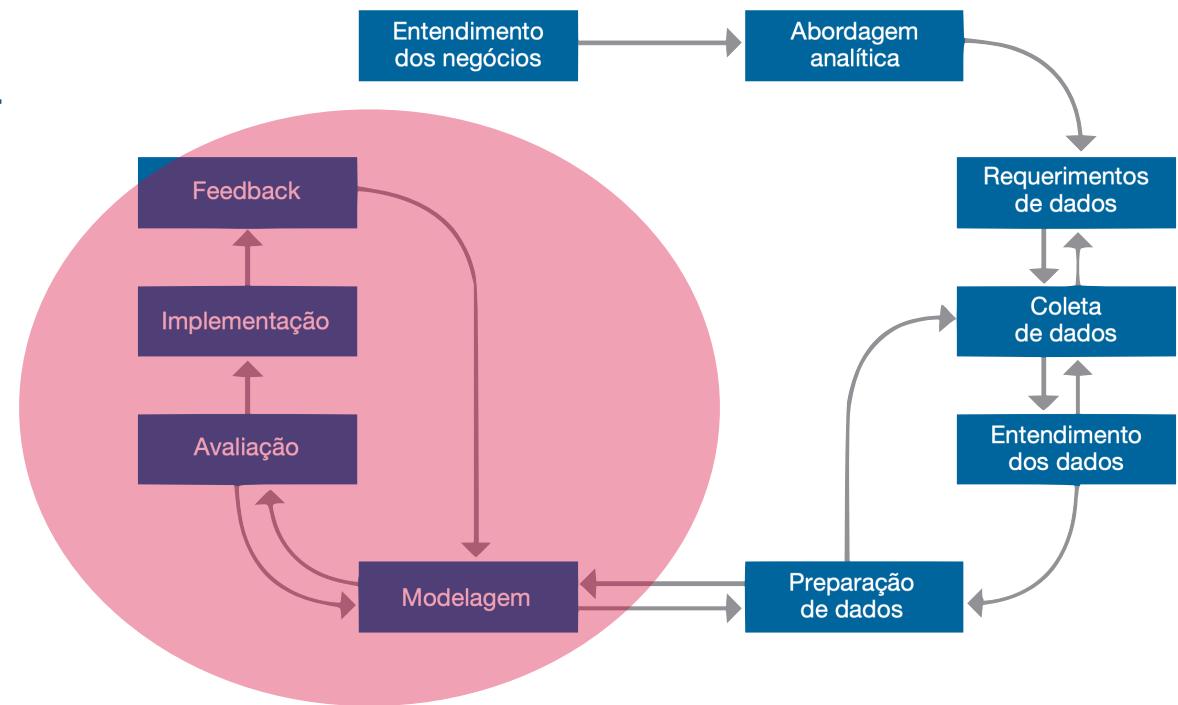
- 3) Qual dado será necessário?
- 4) De onde virão e como eu terei acesso?
- 5) Os dados que coletados são representativos?
- 6) Será necessário algum trabalho de manipulação adicional?

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

AS 10 QUESTÕES-CHAVE

Chegando na resposta:

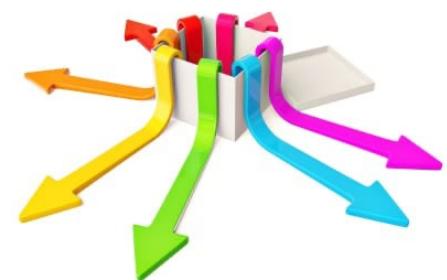
- 7) Como eu posso visualizar melhor estes dados para responder a pergunta?
- 8) O modelo escolhido realmente responde o quer quero saber?
- 9) O modelo é aplicável no dia-a-dia?
- 10) Você pode obter um feedback construtivo para responder à pergunta?



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

CASE DISTRIBUIÇÃO DE VENDAS ONLINE

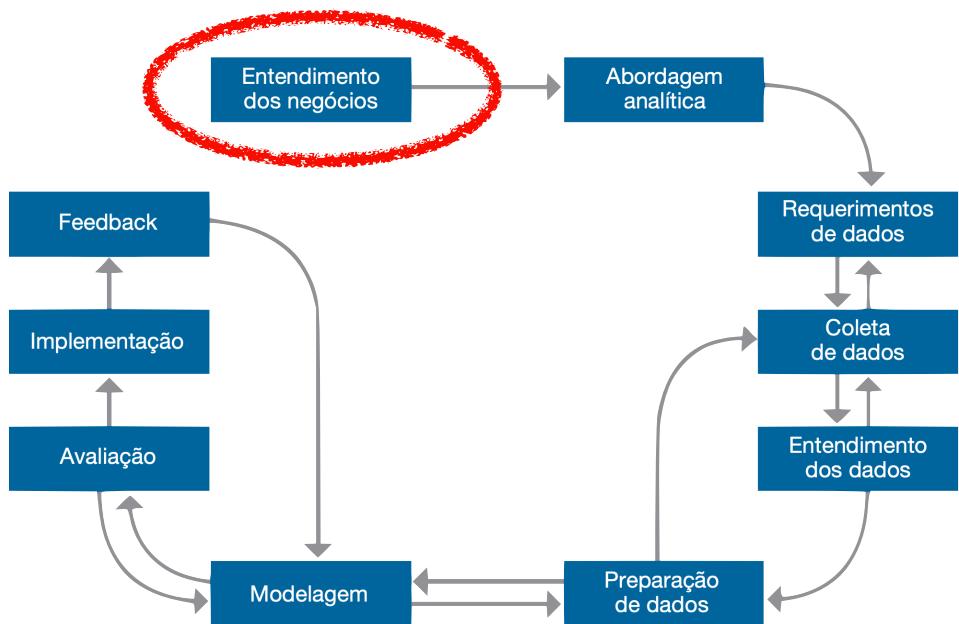
Henrique é o gerente de distribuição de uma empresa produz e vende eletrônicos. Devido ao crescimento das compras online, a empresa foi obrigada a fazer uma reestruturação e modernização da sua rede logística com a inclusão de novos CDs e novas rotas de entrega. Após essa reestruturação Henrique notou que a classificação da experiência de compra online para venda no varejo, que pode ser de uma a cinco estrelas, caiu consideravelmente.



Henrique está sob pressão do CEO para identificar os porquês desta queda de satisfação. Pois devido a esta queda, o projeto de reestruturação logística parou até que se identifique os principais fatores que geram a insatisfação dos clientes de varejo.



ENTENDIMENTO DOS NEGÓCIOS



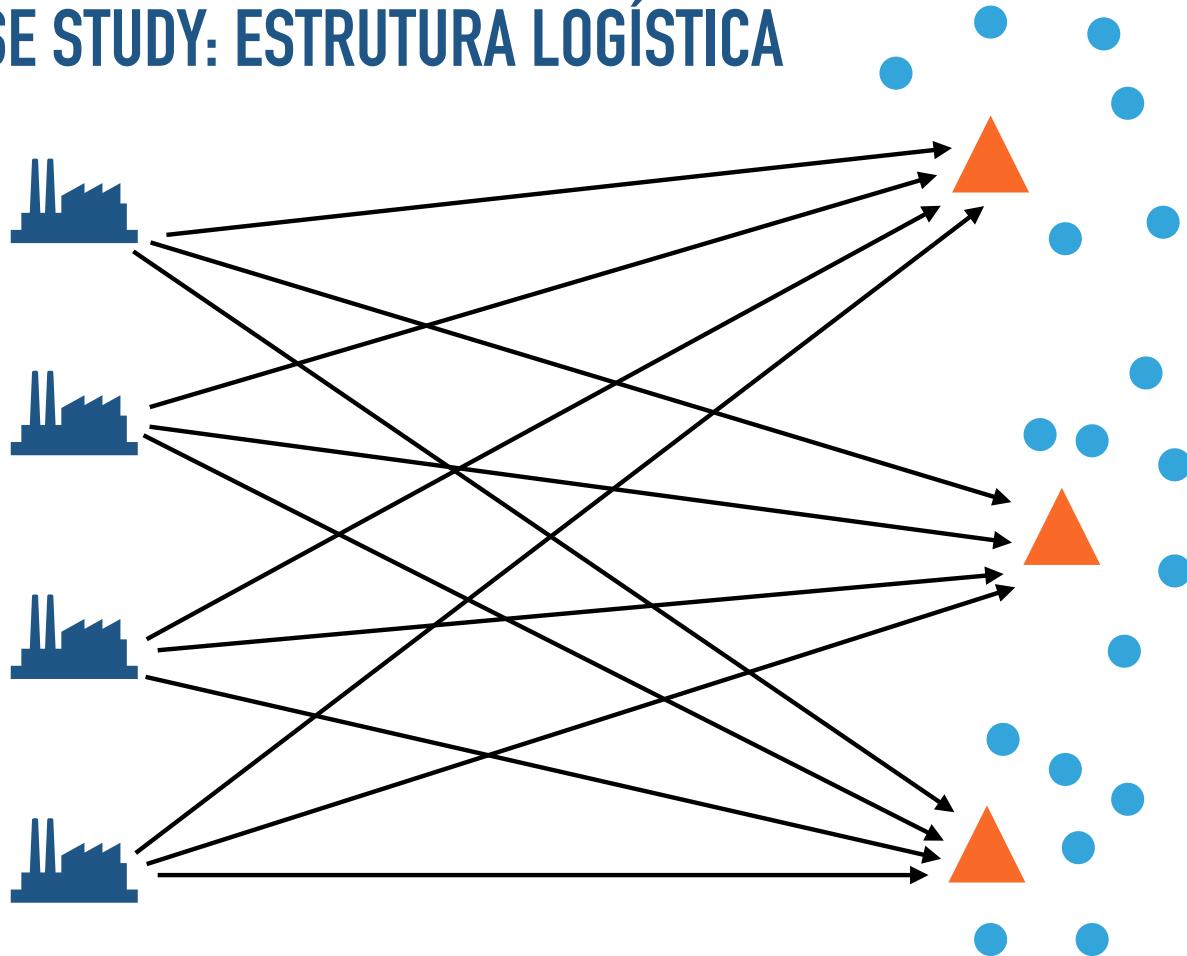
Qual é o problema
que estou tentando
resolver?

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

CASE STUDY - ENTENDIMENTO DO NEGÓCIO

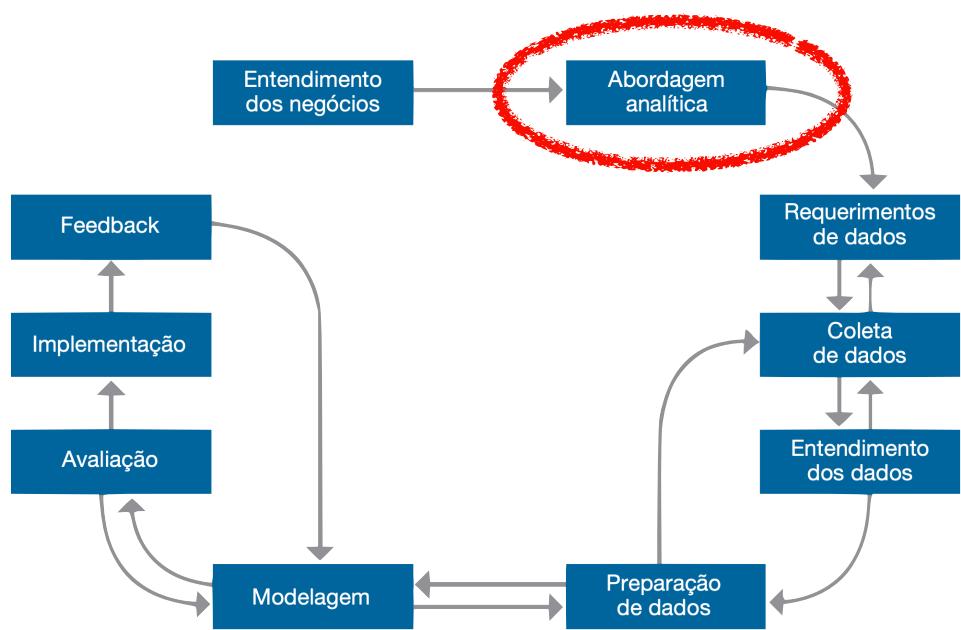
- ▶ Dentro do portfólio de produtos Henrique selecionou 15 para analisar.
- ▶ A empresa conta com 4 fábricas que podem produzir qualquer um destes 15 produtos.
- ▶ A rede logística tem 3 CDs
- ▶ De acordo com a estratégia de operações da empresa somente os CDs podem atender os clientes finais que estejam dentro do seu raio de ação.

CASE STUDY: ESTRUTURA LOGÍSTICA



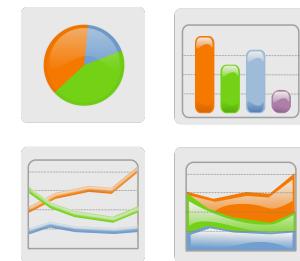
- ▶ O primeiro elo da rede logística requer uma reestruturação, pois hoje todas as fábricas podem servir todos os CDs.
- ▶ O segundo elo já está segmentado, ou seja, cada CD é responsável pela distribuição para uma região.

ABORDAGEM ANALÍTICA



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

Como eu posso usar dados para resolver esse problema?



A resposta para esta pergunta depende do entendimento do negócio.

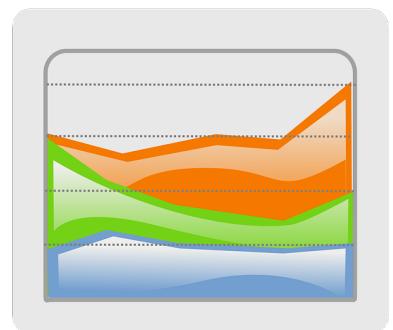
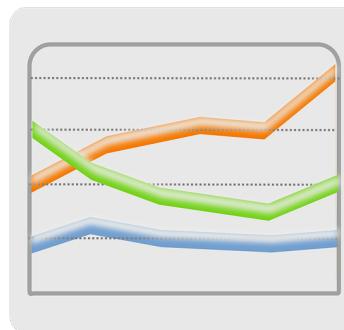
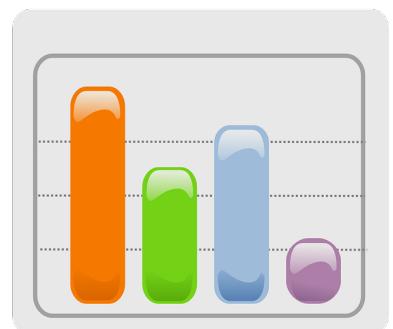
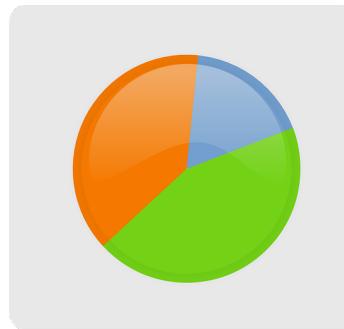
CASE STUDY



- ▶ O portfólio de produtos é o mesmo de antes e após a restruturação logística.
- ▶ Mas após a reestruturação a satisfação do cliente caiu.
- ▶ Compreender de forma clara qual foi a sequência de eventos que levou ao aumento da insatisfação dos clientes.

ABORDAGENS ANALÍTICAS

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



ABORDAGENS DIAGNÓSTICAS

► Objetivos

- ▶ Identificar anomalias
- ▶ Descobertas
- ▶ Determinar relações de causa e efeito

► Técnicas

- ▶ Classificação
- ▶ Busca de padrões
- ▶ Correlações
- ▶ Data Mining
- ▶ Process Mining



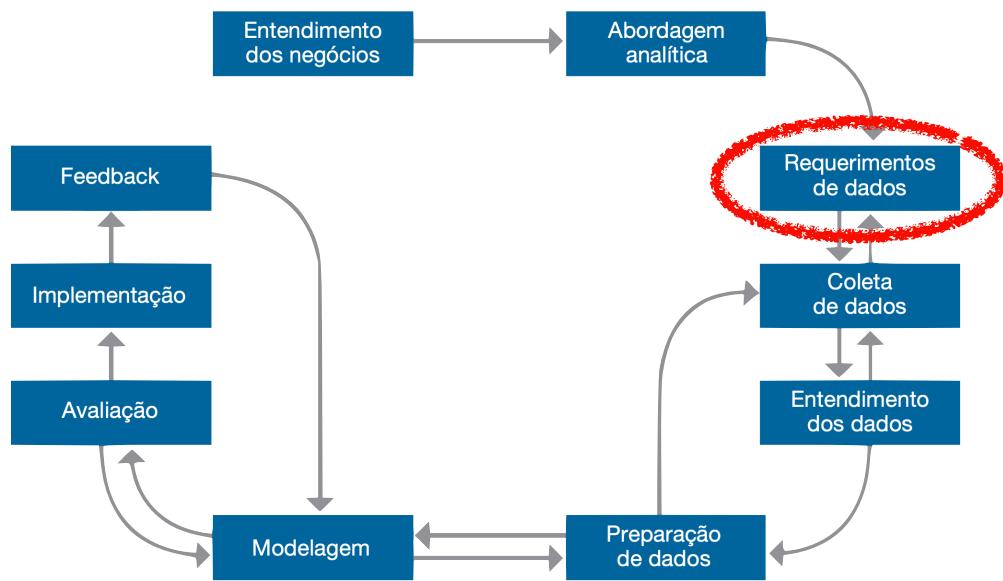
CASE STUDY



Porque isso está
acontecendo?

Determinada parte
do processo afeta
ou não a
satisfação?

ABORDAGEM ANALÍTICA



Como eu posso usar dados para resolver esse problema?

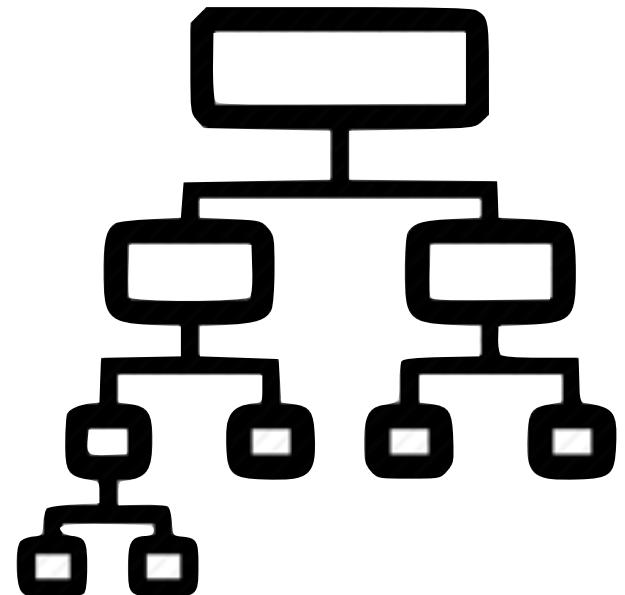


A resposta a esta pergunta depende do entendimento do negócio.

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

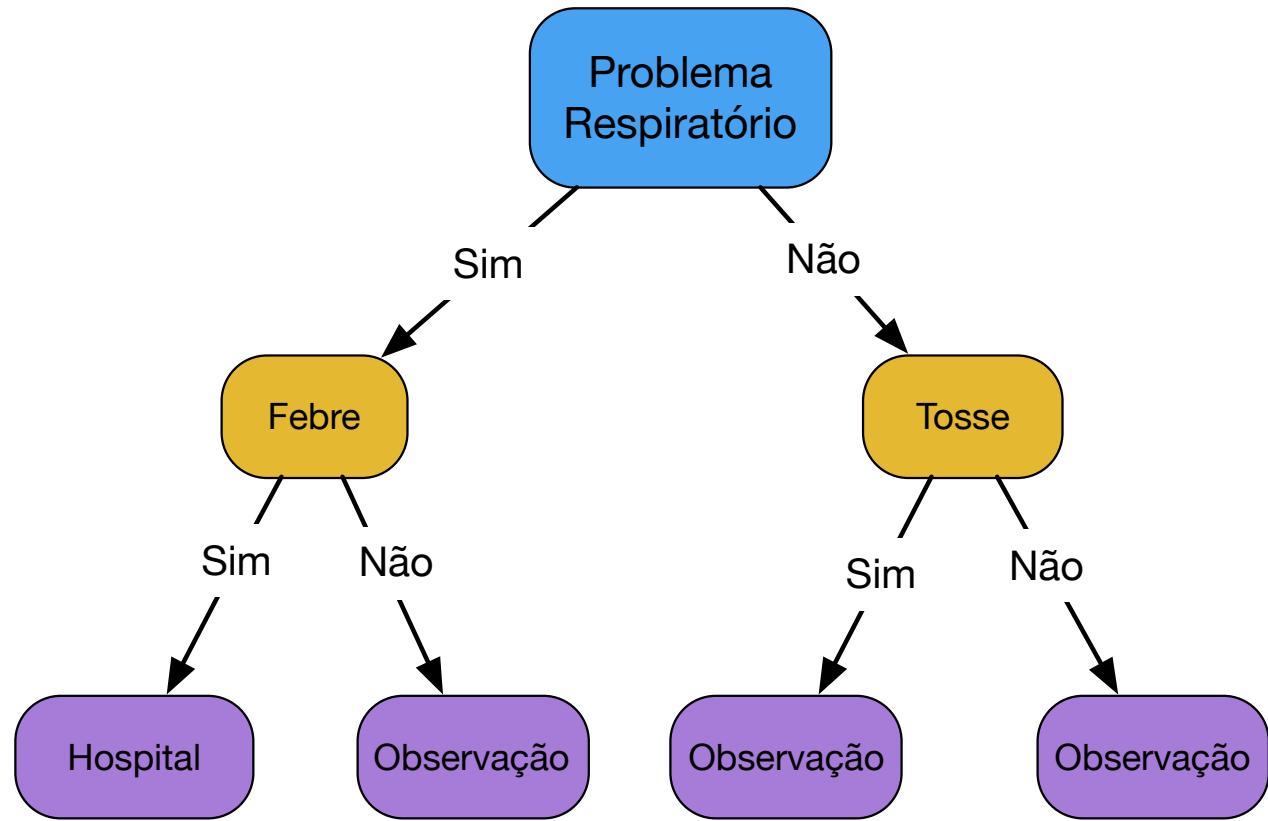
CASE STUDY: DECISION TREE

- ▶ Saída em classes (categorias)
- ▶ Tem um “Caminho decisório” explícito demonstrando as condições que levam a baixa satisfação
- ▶ Probabilidade do resultado
- ▶ Fácil de entender e aplicar

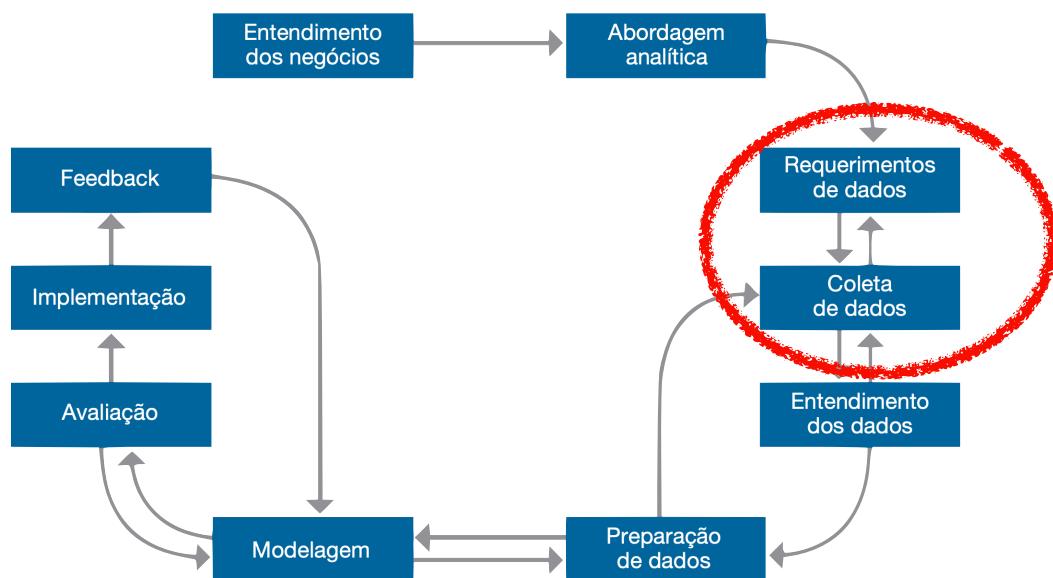


SUPER CASE - INSTA

EXEMPLO



NECESSIDADE E COLETA DE DADOS



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



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

TEXT

QUAIS DADOS VOCÊ COLETARIA?

Eduardo Pécora

DEFINIR E SELECIONAR O ESCOPO DOS DADOS (CORTE)

- ▶ Vendas online com retorno de satisfação
- ▶ Corte temporal - somente vendas após a reestruturação ?
- ▶ Vamos considerar as entregas em atraso junto com as outras ?
 - ▶ Qual as consequências disso?
- ▶ Vamos considerar todas as 4 fábricas e todos os 3 CDs?
- ▶ Existe algum critério de exclusão?

MINHAS DECISÕES

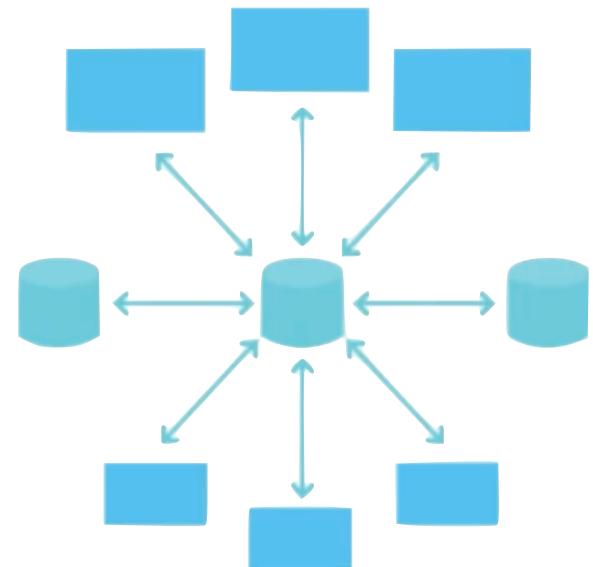
- ▶ Vendas online com retorno de satisfação (**SIM**)
- ▶ Corte temporal - somente vendas após a reestruturação ? (**SIM**)
- ▶ Vamos considerar as entregas em atraso junto com as outras ? (**NÃO**)
 - ▶ Qual as consequências disso? (**Causa viés**)
- ▶ Vamos considerar todas as 4 fábricas e todos os 3 CDs? (**SIM, mas podemos reconsiderar**)
- ▶ Existe algum critério de exclusão? (**dados faltantes, dados errados, ...**)

CASE STUDY: ESTRUTURA DOS DADOS

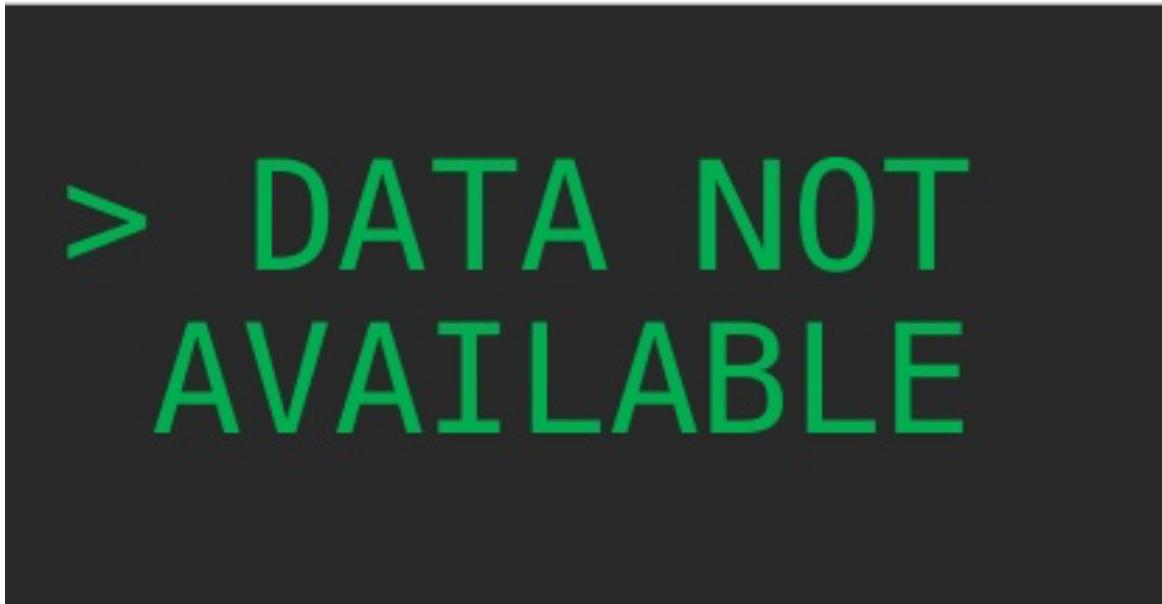
- ▶ Conteúdos e formatos de acordo com as entradas necessárias para uma árvore de decisão
- ▶ Uma linha por venda executada, com as variáveis de decisão e as variáveis independentes
- ▶ Dados contendo o máximo de informações do processo logístico da entrega

CASE STUDY: FONTES DE DADOS

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



EXISTEM OUTROS DADOS QUE VCS ACHAM PERTINENTES?



> DATA NOT
AVAILABLE

SUPER CASE - INSTA

BD MERGING



Só quem é da NOVA GERAÇÃO vai entender essa piada!

SUPER CASE - INSTA

SUPERCASE_DATA_6000_3.CSV



SUPER CASE - INSTA

DATA TABLE

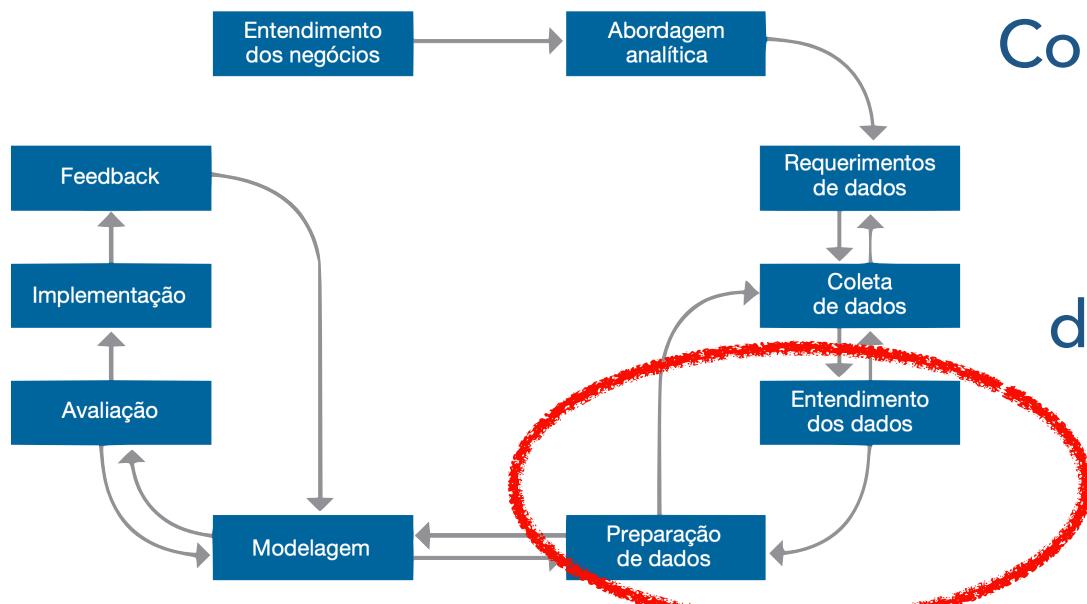
Linha do Data Table	Descrição	Classe
Stars	Satisfação com a venda	Numérica
Product	Produto	Categórica
Origin	Fábrica de Origem	Categórica
CD_name	CD utilizado	Categórica
Dest	Cliente Final	Categórica
DistanceTravelled_Fac_CD	Distancia Fab - CD	Numérica
TruckType	Caminhão Fab - CD	Categórica
DriverF_CD	Motorista Fab - CD	Categórica
DaysStocked	Dias estocados	Numérica
FloorCD	Andar do porta palete	Numérica
DistanceTravelled_CD_Client	Distancia CD - Cliente	Numérica
TruckType2	Caminhão CD - Cliente	Categórica
DriverCD_C	Motorista CD - Cliente	Categórica
RoutDeliveryOrder	Ordem de entrega	Numérica



pythonTM

Seção: Importar Arquivo de dados

DO ENTENDIMENTO A PREPARAÇÃO



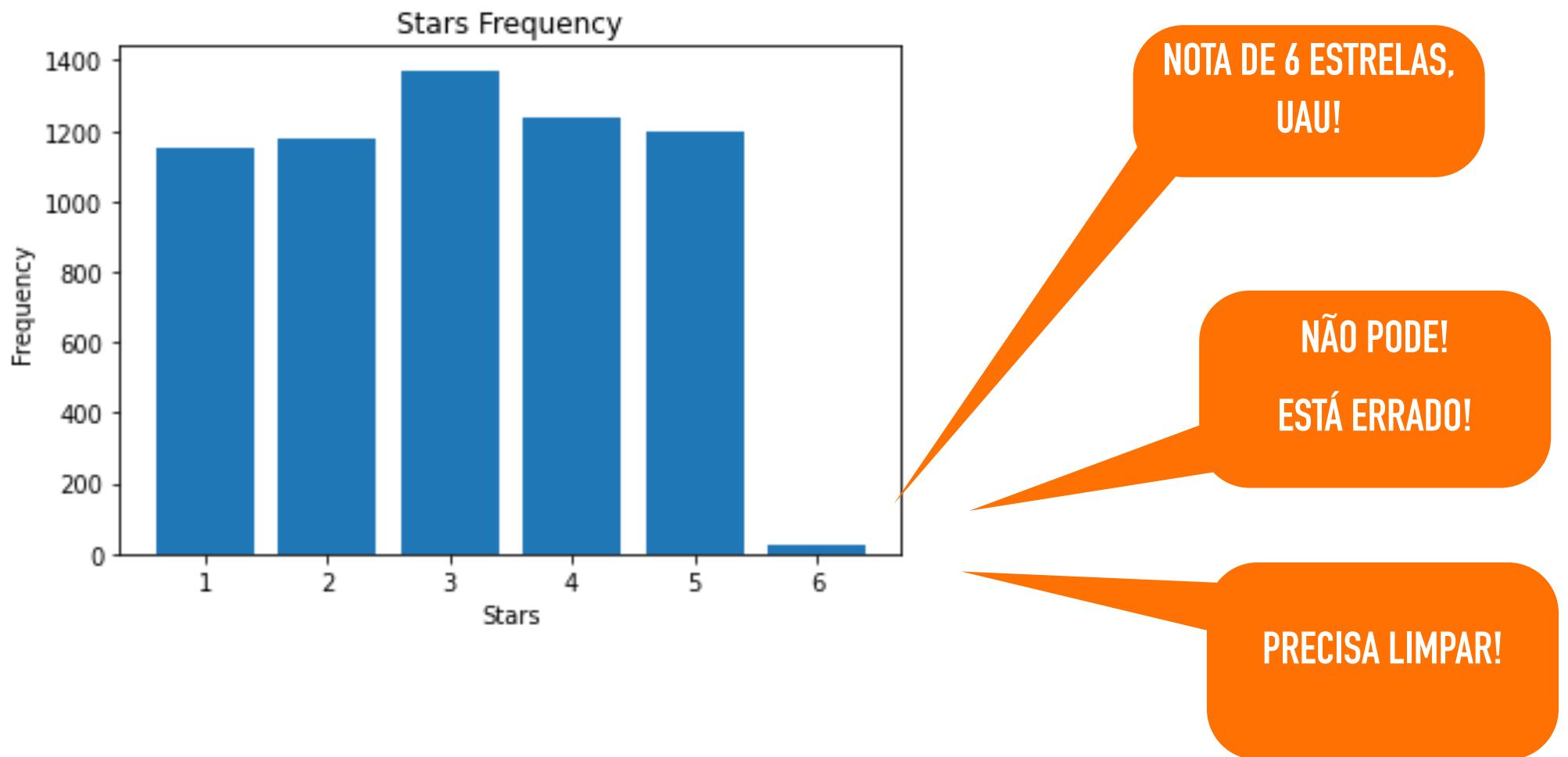
Como assim “limpar” os dados?

Ao preparar e manipular os dados não introduzo um viés?

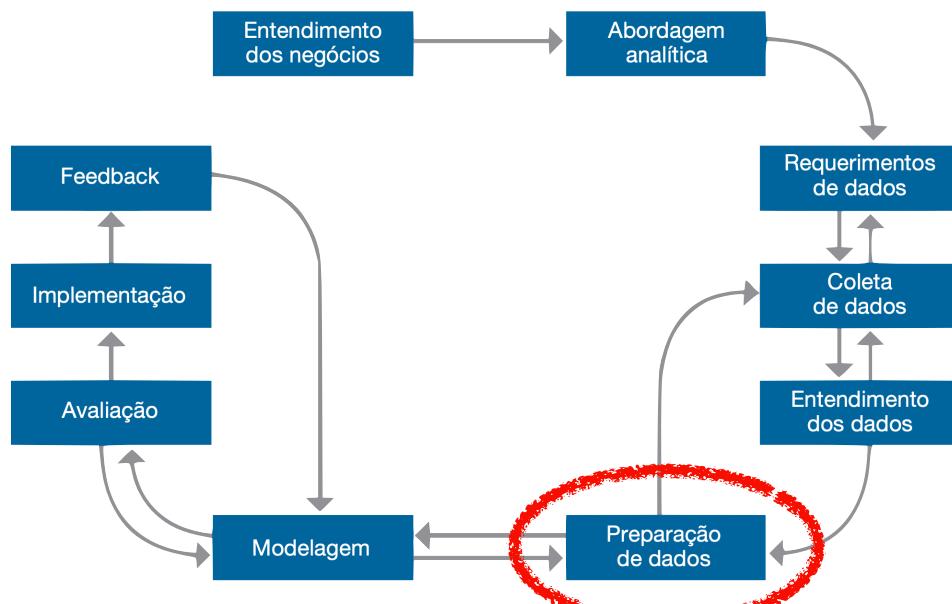
Source: Rollings John B., Foundational methodology for data science. IBM white paper

LIMPEZA DOS DADOS





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

DECISION TREE CLASSIFIER



CASE STUDY: VARIÁVEIS

Quem são as variáveis dependentes (resultados) e as independentes (condições) neste caso?



pythonTM

Seção:

Separar o DataFrame em variáveis dependentes e independentes

SUPER CASE - INSTA

DATA TABLE

Linha do Data Table	Descrição	Classe
Stars	Satisfação com a venda	Numérica
Product	Produto	Categórica
Origin	Fábrica de Origem	Categórica
CD_name	CD utilizado	Categórica
Dest	Cliente Final	Categórica
DistanceTravelled_Fac_CD	Distancia Fab - CD	Numérica
TruckType	Caminhão Fab - CD	Categórica
DriverF_CD	Motorista Fab - CD	Categórica
DaysStocked	Dias estocados	Numérica
FloorCD	Andar do porta palete	Numérica
DistanceTravelled_CD_Client	Distancia CD - Cliente	Numérica
TruckType2	Caminhão CD - Cliente	Categórica
DriverCD_C	Motorista CD - Cliente	Categórica
RoutDeliveryOrder	Ordem de entrega	Numérica

PRODUTO

{ P01, P02, P03, ..., P014, P15 }

- ▶ Uma única coluna “Producut” com 15 valores (texto) distintos
- ▶ Sem uma ordenação clara (pense como um computador)
- ▶ Isso complica a vida do algoritmo

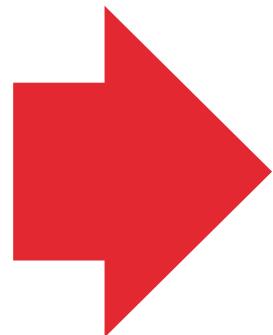
SUPER CASE - INSTA

PRODUTO

{ P01, P02, P03, ..., P014, P15 }

- ▶ Solução: criar uma coluna para cada valor, ou seja, 15 novas colunas

Product
P01
P02
P02
P03



	P01	P02	P03
P01	1	0	0
P02	0	1	0
P02	0	1	0
P03	0	0	1

SUPER CASE - INSTA

PRODUTO

{ P01, P02, P03, ..., P014, P15 }

LINGUAGEM HUMANA

- ▶ Solução: criar uma coluna para cada valor, ou seja, 15 novas colunas

Product
P01
P02
P02
P03

ONE
HOT
ENCODING

P01	P02	P03
1	0	0
0	1	0
0	1	0
0	0	1

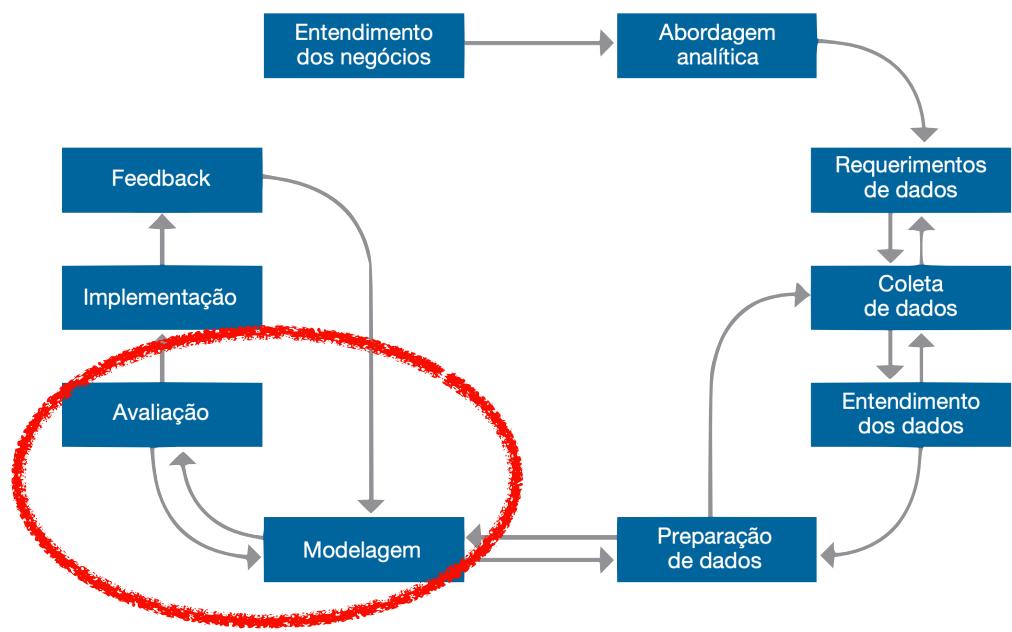
LINGUAGEM BINÁRIA



Seção:

Adequação das variáveis categóricas em colunas binárias -
OneHotEncoder

O MODELO (FINALMENTE)



Como eu posso visualizar os
dados de maneira rápida?

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

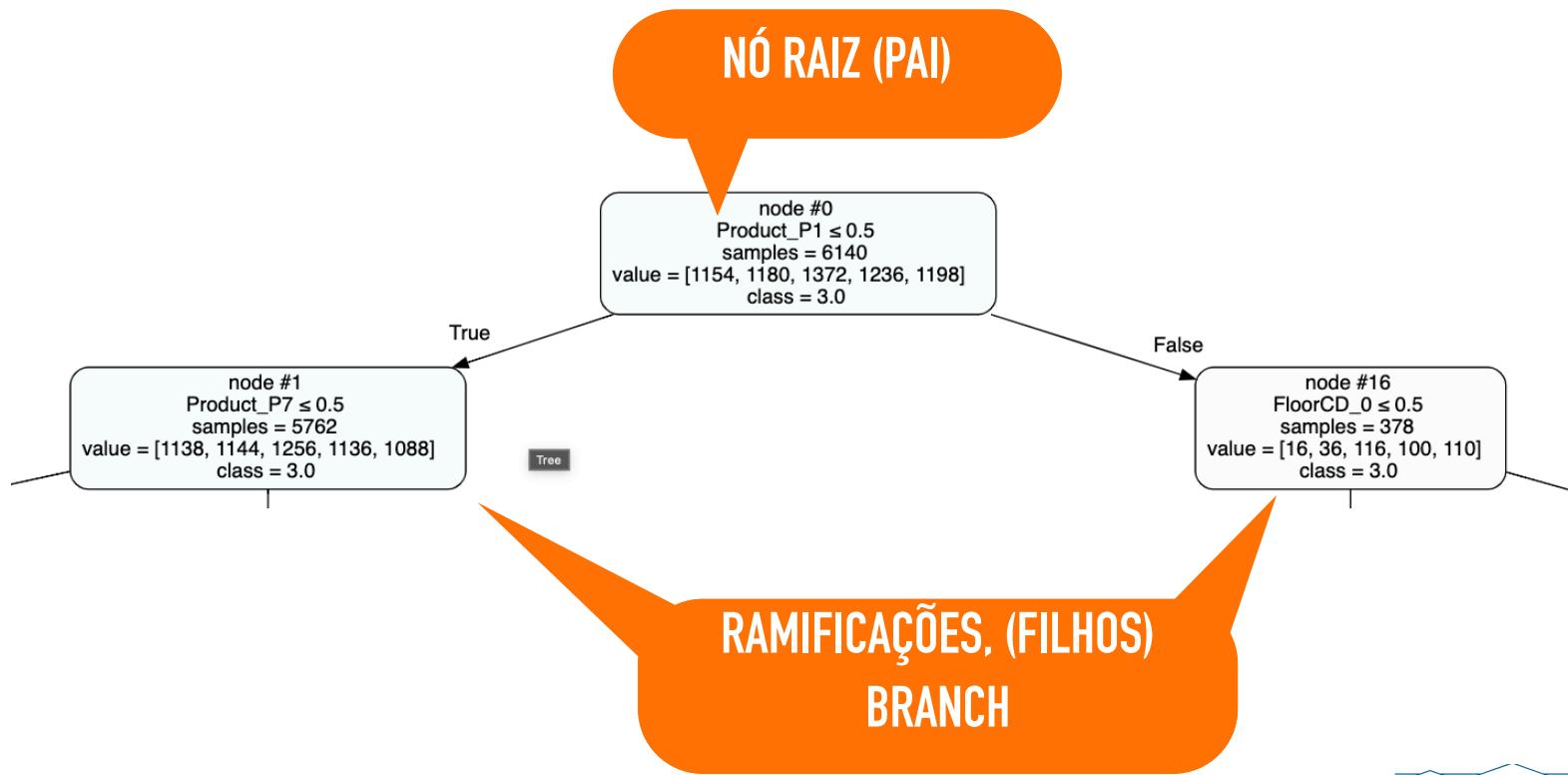


pythonTM

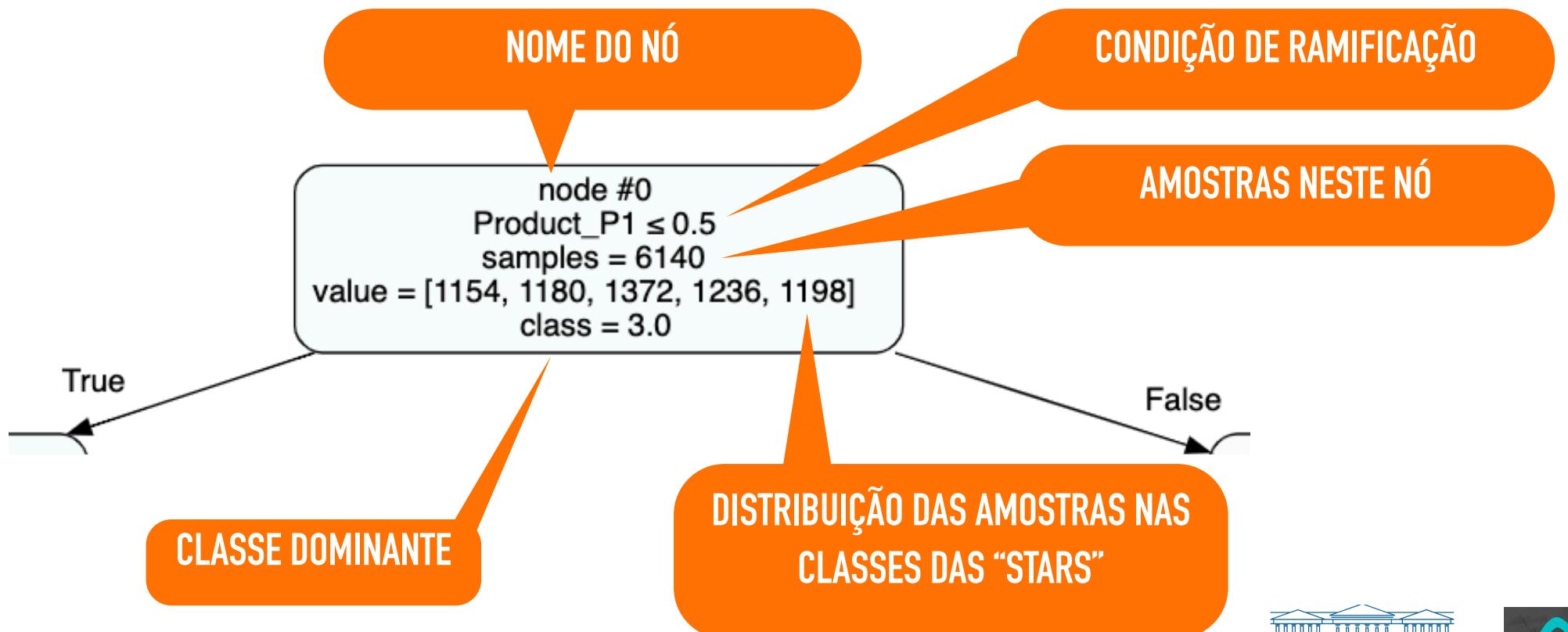
Seção:

DecisionTreeClassifier e os seus Parâmetros

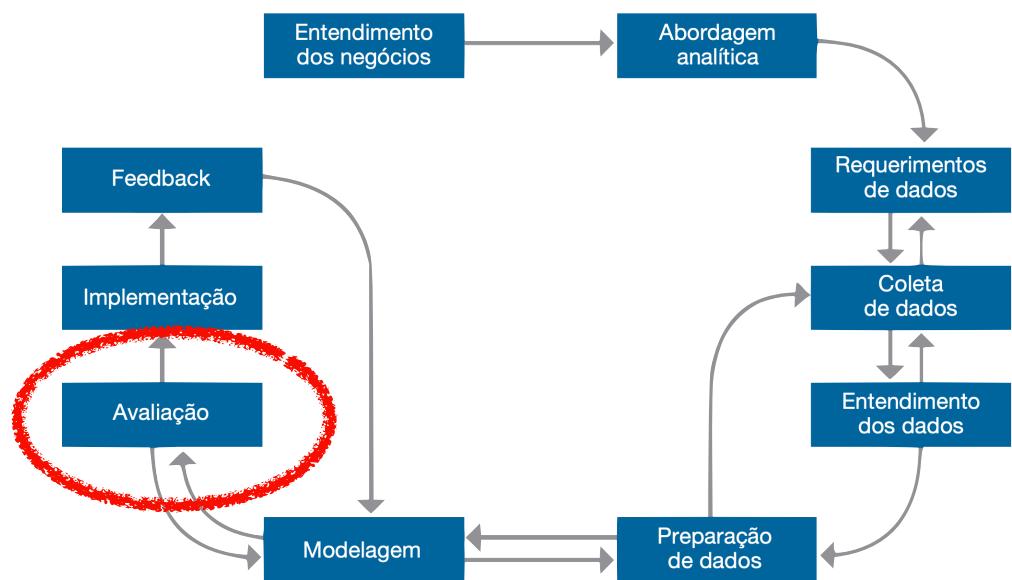
INTERPRETANDO A SAÍDA EM ÁRVORE



INTERPRETANDO A SAÍDA EM ÁRVORE



O MODELO (FINALMENTE)



Tem como melhorar esse modelo?

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

LEMBRA DO OVERFITTING?

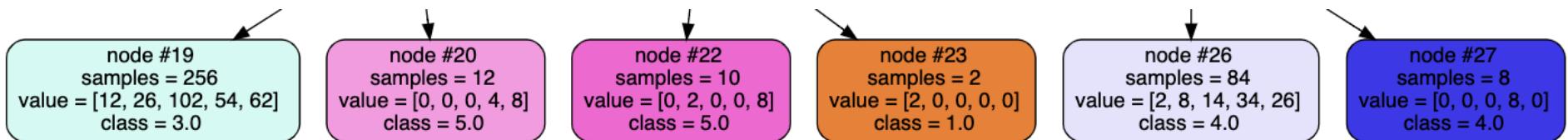
node #23
samples = 2
value = [2, 0, 0, 0, 0]
class = 1.0



Somente 2 registros de
mais de 6000
originais?



PRECISO DE TODAS ESSAS 5 CLASSES DE SAÍDA?



OU UMA AVALIAÇÃO “RUIM”, “MEDIANA” E “BOA” JÁ SERIA SUFICIENTE PARA OS MEUS PROPÓSITOS?



DELETE O SIMBOLO # DAS LINHAS ABAIXO E RODE O CÓDIGO NOVAMENTE

- Separar o DataFrame em variáveis dependentes e independentes

```
▶ # split the DataFrame  
stars = starsData["stars"] # Só a coluna "stars"  
stars_class = stars[:]  
variables = starsData.iloc[:,1:] # Todas as outras colunas com excessão da "stars"  
  
# Deixe essa parte do código comentado poor enquanto  
  
stars_class.loc[stars <= 2] = "bad"  
stars_class.loc[stars == 3] = "average"  
stars_class.loc[stars >= 4] = "good"  
  
stars_class.head()
```

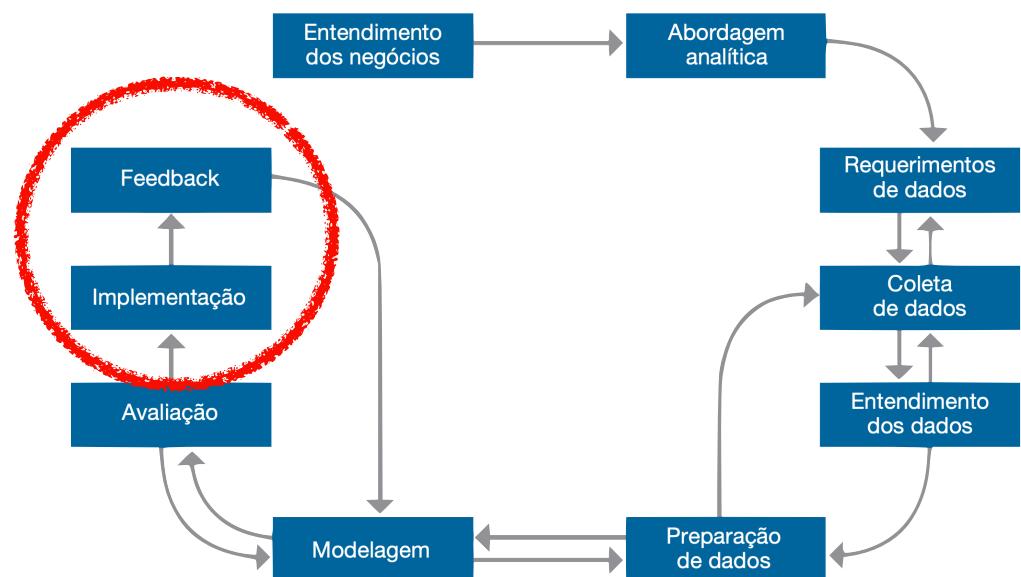
LEMBRE-SE DE INSERIR O ARQUIVO DE DADOS NOVAMENTE ;)

TEXT

O QUE MODIFICOU NOS RESULTADOS?

Eduardo Pécora

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



PPGOLD

Gestão de Organizações, Liderança e Decisão
Mestrado Profissional

MUITO OBRIGADO!

EDUARDO PÉCORA
PECORA@UFPR.BR