

Exploração de Dados

Prof. Esp. Victor Venites



School of AI – São Paulo
Aula 6

Até Aqui

- **Aula 1 –**
 - Introdução a Deep Learning – Uma abordagem visual
 - Data Science Aplicada na Mobilidade Urbana: Da Modelagem ao Machine Learning
 - Aplicação de Data Science para construção de um modelo de previsão de vendas de uma cadeia de varejo
- **Aula 2 –**
 - Algoritmos Clássicos em IA - Recomendação e Árvore de Decisão
 - Data Science & Advanced Analytics Aplicada na Mobilidade Urbana: Da Modelagem ao Machine Learning
- **Aula 3 –**
 - Diga olá para seu Hardware
 - Uma aplicação de visão computacional aplicada ao varejo
 - Machine Translation para previsão de vendas
- **Aula 4 –**
 - Back to Basics
 - Entendendo Regressão
- **Aula 5 –**
 - Estatística Básica
 - Entendendo Regressão 2



Objetivo

- Passar experiência com bases de dados
 - Deixar o aluno apto para saber visualizar e manipular uma base
 - Ter noções de como fazer isso
 - Mostrar um resumo do que envolve a exploração de dados
 - Partir do viés da computação
 - Levantar questões... E responder a maioria!
-
- **Material:** GitHub
 - **Video:** Youtube



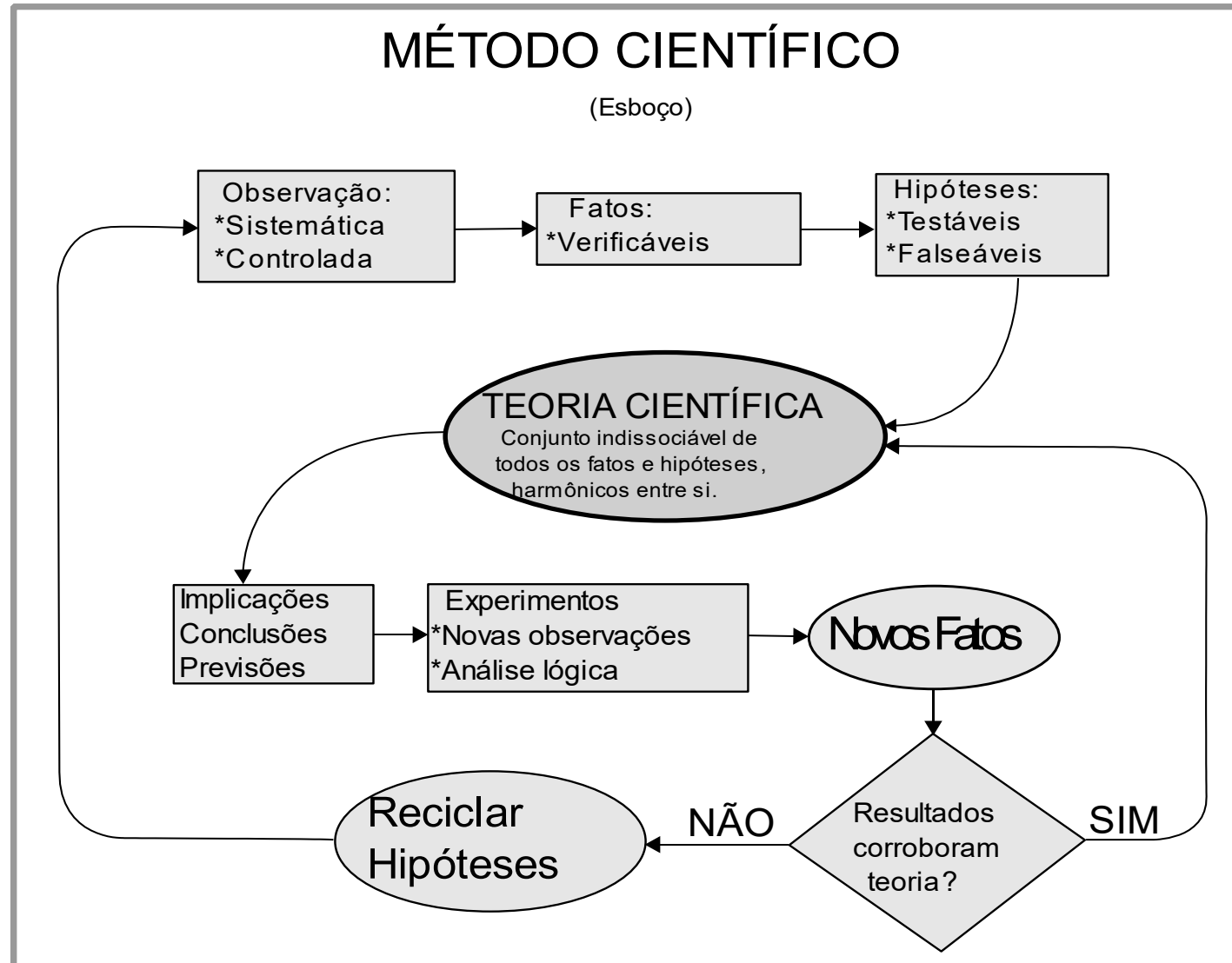
Termos Populares da Atualidade

- Manipulação de Dados
- Visualização de Dados
- Tratamento de Dados
- Bancos de Dados
- Scraping
- Data Mining
- Data Analysis
- Missing Values
- Normalização

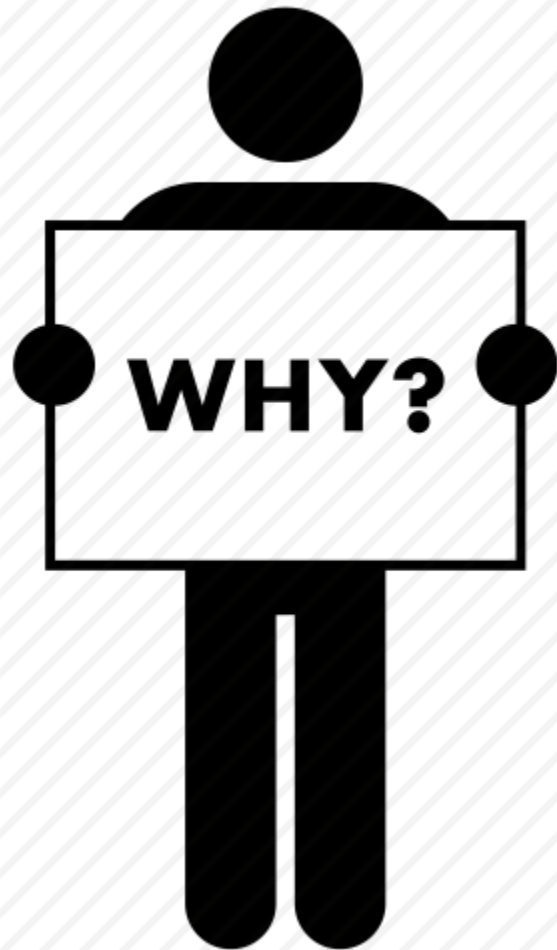


- *Wikipedia*

- *Oi?*



Por quê?



- Para buscar respostas/hipóteses na resolução de problemas informados.
- Uma vez que você tem sua base de dados, antes de fazer sua Análise de Dados deve explorar os dados.
- Visando entender suas Propriedades e Composições.
- Tornando a base apropriada para cálculos matemáticos(e usar o Numpy), em forma de matrizes e números.

Como?

- Utilizando das noções básicas de banco de dados
- Com o Python como ferramenta
- Técnicas comuns
- Perguntas frequentes e as perguntas certas
- Processo orgânico, como na vida



O que?

- Dados/Escalas desconhecidos - > idade, peso, textos
- Valores perdidos (espaços em branco)
- Fontes duvidosas
- Muitas colunas/linhas
- Precisa de mais colunas/linhas
- Juntar com outras bases
- Dados Paramétricos/Não-Paramétricos
- Variáveis Dependentes/Independentes



Banco de Dados SQL?



- 10 grandes comandos do T-SQL:
 - **Acesso:** Grant, Deny, Revoke
 - **Tabela:** Create, Drop, Alter
 - **Dados:** Insert, Select, Update, Delete
 - Extras: Join, GroupBy, OrderBy

Passo-a-Passo

- 1 -> Importar a base
- 2 -> Entender as dimensões
- 3 -> Valores nulos, vazios, incompletos
- 4 -> Filtrar linhas/colunas
- 5 -> Juntar tabelas
- ...
- X - > Dominar o mundo



Onde adquirir bases de Dados?

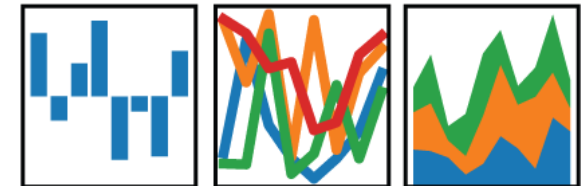
- Sites conhecidos:
 - **Kaggle** - <https://www.kaggle.com/datasets>
 - **UCI – Machine Learning Repository** - <https://archive.ics.uci.edu/ml/datasets.html>
 - **Portal da Transparência** - <http://www.portaltransparencia.gov.br/download-de-dados>
 - **Europa – Eurostat** - <https://ec.europa.eu/eurostat/en/data/database>
 - **Google Dataset Search** - <https://toolbox.google.com/datasetsearch>

Hands-On



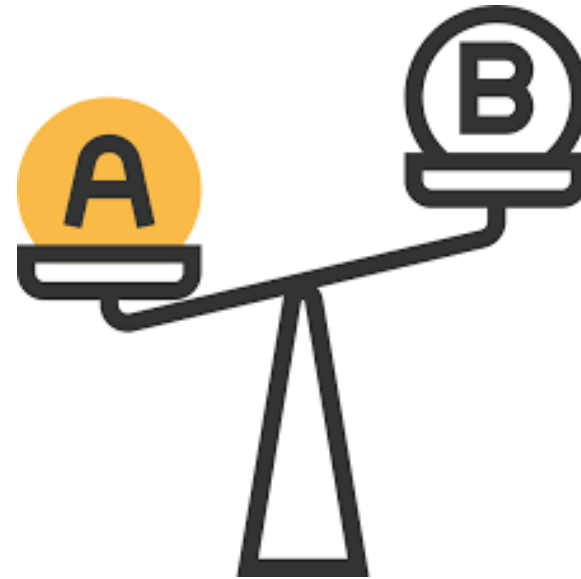
- Jupyter Notebook + Pandas
- “Codando”

pandas

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


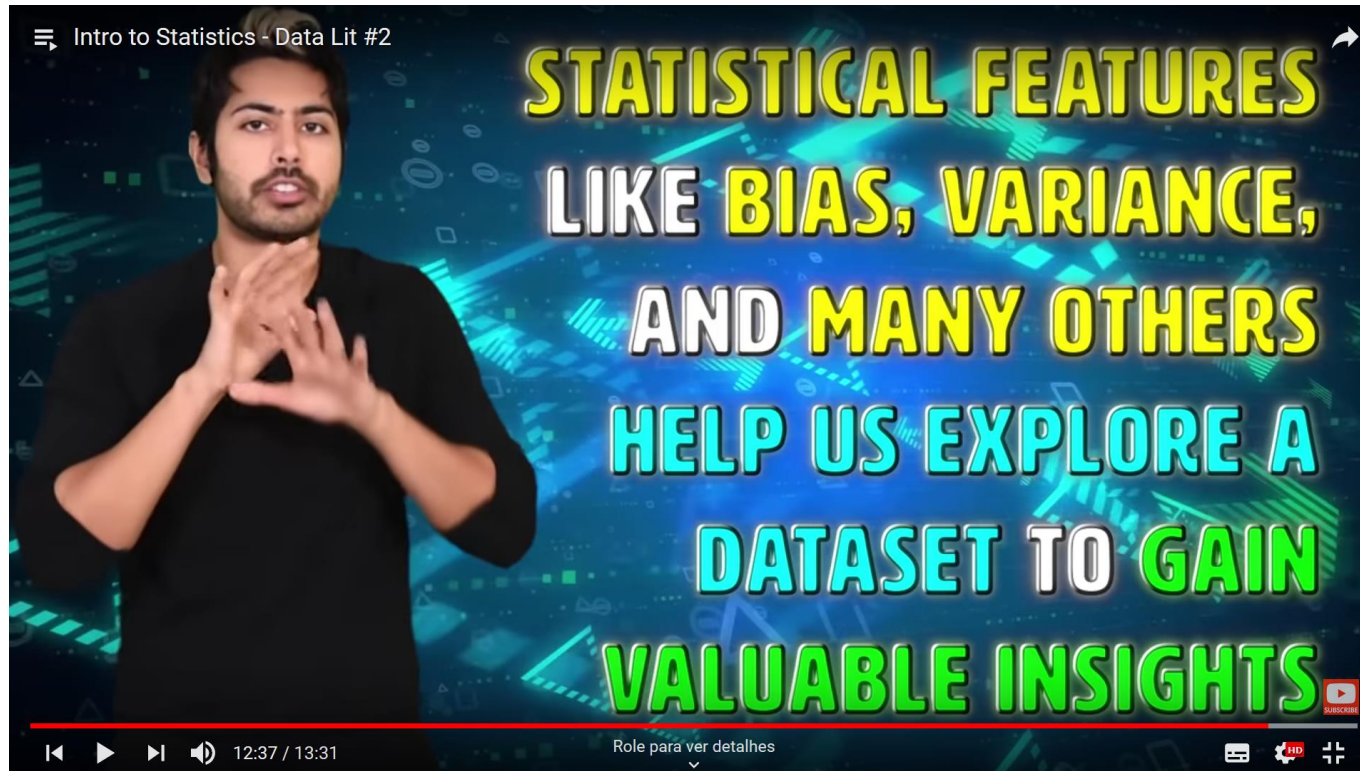
Técnicas Pertinentes

- PCA: Principal Componente Analysis
- T-Student: e também o Teste-Z
- ANOVA: Análise de Variância
- MANOVA: ANOVA Multivariada
- Mann-Whitney
- Teste de Wilcoxon



Dados Paramétricos/Não-Paramétricos

Ponto importante



Parte 2:
- Veremos mais
sobre estatística

Intro to Statistics - Data Lit #2

<https://www.youtube.com/watch?v=MdHtK7CWpCQ&index=3&list=PL2-dafEMk2A5ET-lsVVabwN-cFk5r8TFe>

Revisão

- Mapa Mental
- Aplicações
- Dúvidas
- Feedback...
 - O que achou da aula?
 - Como foi sua experiencia?



Referências Bibliográficas - Livros



- **Big Data Analytics With R** – Simon Walkowiak (2016), ISBN 978-1-78646-645-7
- **Comece Pelo Porquê** – Simon Sinek(2018), ISBN 978-85-431-0663-2
- **Introdução à Ciência de Dados** – Fernando Amaral (2016), ISBN 978-85-7608-934-6
- **Multivariate Statistical Quality Control Using R** – Edgar Santos-Fernández (2012), ISBN 978-1-4614-5452-6

Referências Bibliográficas - Vídeos

- **Exploring data in R –**
 - <https://www.youtube.com/watch?v=mgsIMeYwNmM>
- **School of AI São Paulo –**
 - <https://www.youtube.com/channel/UCcQgGC19k35ayQNsspyyBhQ>



Obrigado!

Att,

Victor Venites.

LinkedIn: <http://victorvenites.com/>

E-mail: contato@victorvenites.com

