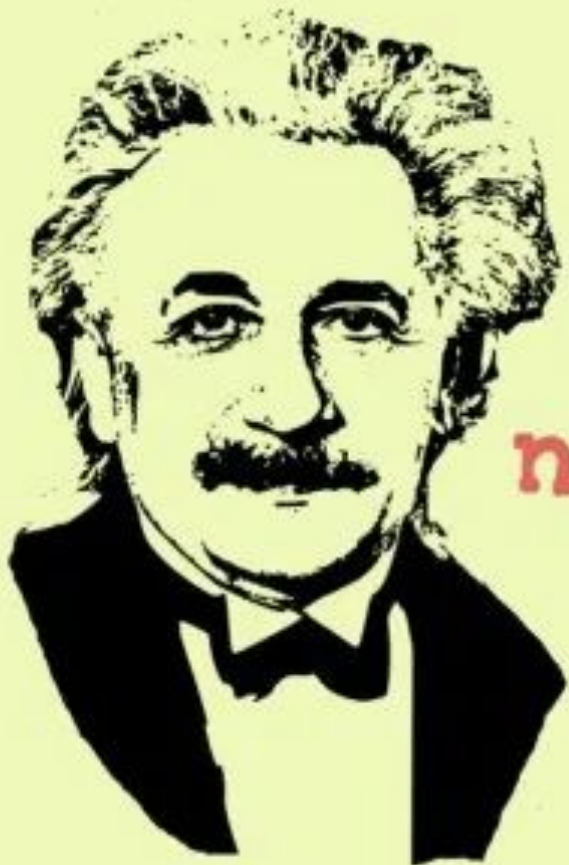


INTRODUÇÃO À CIÊNCIA DE DADOS

Introdução à Ciência de Dados

CIÊNCIA DE DADOS



The important thing is
never to stop questioning.
Be curious
- Albert Einstein

science freek

CIÊNCIA DE DADOS

A expressão "Data Science" tem origem nos anos 1960, mas a "Ciência de Dados" que lidamos atualmente é algo novo, ainda em transformação e muitas vezes controversa.

Necessário compreender a diferença entre "Senso Comum" e "Ciência", visto que a disciplina é **Ciência** de Dados.

SENSO COMUM (Primário)

SENSO COMUM: É o conhecimento do dia-a-dia, primário ou imediato. O que vem dos nossos sentidos, das tradições, nosso contato diário com o ambiente.

Senso comum responde a grande parte das questões corriqueiras, frequentes e diárias.

CONHECIMENTO APROFUNDADO (Secundário)

- Conhecimento além das sensações.
- **Demanda observação sistemática** (mais detalhada, mais dimensões, busca por características incomuns).
- **Não precisa ser definitivo**, deve ser questionado, busca evolução e aprimoramento.
- O conhecimento exige provas. Gera argumentos. Coloca questões. Busca negar o que era certo e comprovar o que não era aceito. **Eterna busca!**

CIÊNCIA

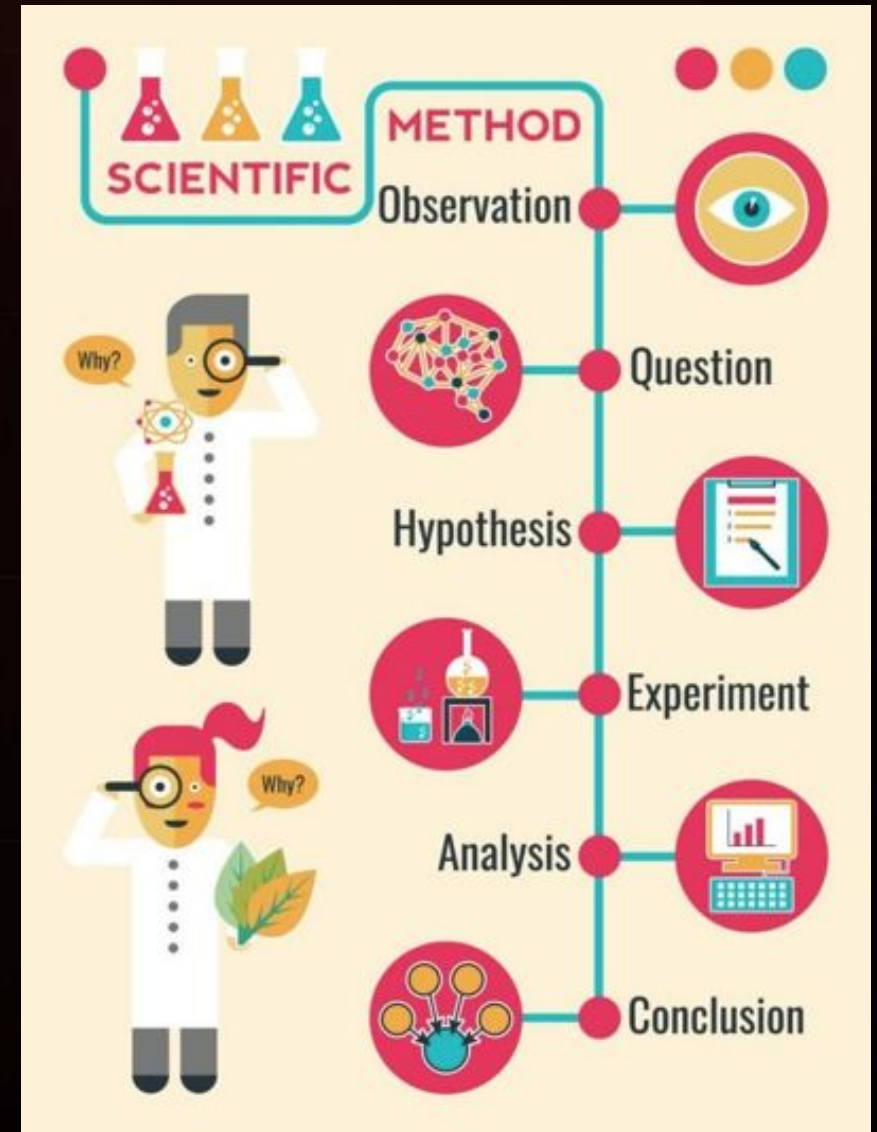
- A palavra "**ciência**" vem do latim "scientia", que significa "conhecimento".
- Ciência refere-se a **qualquer conhecimento ou prática sistemáticos**.
- Em sentido estrito, ciência refere-se ao **sistema de adquirir conhecimento baseado no método científico** bem como ao corpo organizado de conhecimento conseguido através de tais pesquisas.

CIÊNCIA

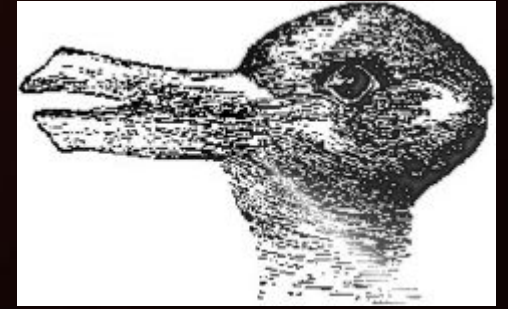
- A ciência é aquele tipo de conhecimento que **busca compreender verdades ou leis naturais** para explicar o funcionamento das coisas e do universo em geral.
- É por isso que cientistas fazem observações, verificações, medições, análises e classificações, procurando **entender os fatos e traduzi-los para uma linguagem estatística**.
- E é aí que entra o método científico.

MÉTODO CIENTÍFICO

- O **método científico** é, basicamente, um conjunto de regras para se realizar uma experiência, com o objetivo de produzir um novo conhecimento, além de corrigir conhecimentos preexistentes.
- Essas regras são necessárias justamente para coibir a subjetividade, direcionando a pesquisa para a produção de conhecimentos válidos — em suma, científicos.



REVOLUÇÕES CIENTÍFICAS



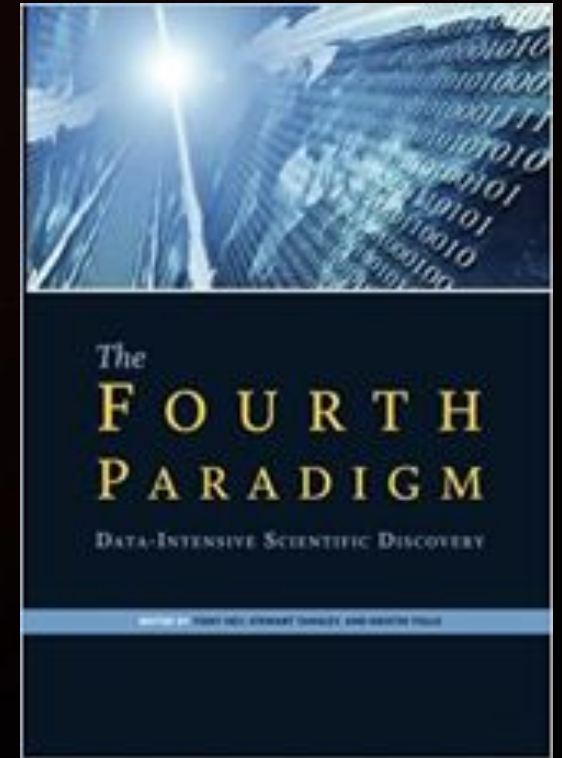
- Revoluções científicas são **mudanças de paradigmas** que alteram o modelo de fazer ciência.
- Thomas Kuhn, publicou em 1962 a “estrutura das revoluções científicas”.
- Kuhn garante que:
"Quando um número suficiente de anomalias significativas se acumula contra um paradigma atual, a disciplina científica é lançada em um **estado de crise**. Durante a crise, novas ideias, talvez anteriormente descartadas, são experimentadas."

EXEMPLOS DE REVOLUÇÕES

- Ptolomeu – Geocentrismo (apoiado pela Igreja), baseado no senso comum de que a Terra está parada e tudo gira em torno dela.
- Copérnico (1473-1543) – Heliocentrismo (Sol é o centro de tudo), apresenta a hipótese.
- Galileu Galilei (1564-1642)– Com telescópio ele confirma (mais de 50 anos depois de Copérnico) o Heliocentrismo.
 - Foi condenado pela Santa Inquisição
 - Suas descobertas apresentam uma revolução
 - Algumas teorias de Newton nascem a partir disso.
- Muitos ainda dizem que **estudar dados é uma grande besteira!**

PARADIGMAS DA CIÊNCIA

- Jim Gray, cientista da computação, defendia a teoria de que a ciência vive seu **quarto paradigma**.
- Os paradigmas são:
 - Ciência Empírica (Baseada em observação)
 - Ciência Teórica (Baseada em teorias, modelos e estudos matemáticos)
 - Ciência Computacional (Baseada em simulações computacionais)
 - E-Science (Baseada em dados, evidências)



O QUE É CIÊNCIA DE DADOS?

Há uma infinidade de definições, dentre elas:

Ciência de dados é uma **área interdisciplinar** voltada para o estudo e a análise de dados, que visa a **extração de conhecimento**, detecção de padrões e/ou obtenção de insights para possíveis tomadas de decisão.

A Ciência de Dados estuda o dado desde a produção até o descarte!

O QUE É CIÊNCIA DE DADOS?

Podemos pensar também:

Se partimos do princípio de que é uma Ciência (Science), a ideia é que você **gere hipóteses a partir de uma observação** e então **utilize os dados** para compreender se a **hipótese se confirmará ou não**, utilizando-a de forma estratégica posteriormente.

Sempre pense em tomada de decisão orientada em dados!

SOBRE CIÊNCIA DE DADOS

Também chamada de Data Science, usaremos daqui para frente a sigla **DS** para fazer referência ao tema. Tecnicamente DS é a interseção que ocorre entre:

- **Habilidades de programação** e de uso de ferramentas tecnológicas.
- Muito **conhecimento de estatística e matemática** (análise de modelos, compreensão de técnicas de visualização, etc...)

SOBRE CIÊNCIA DE DADOS

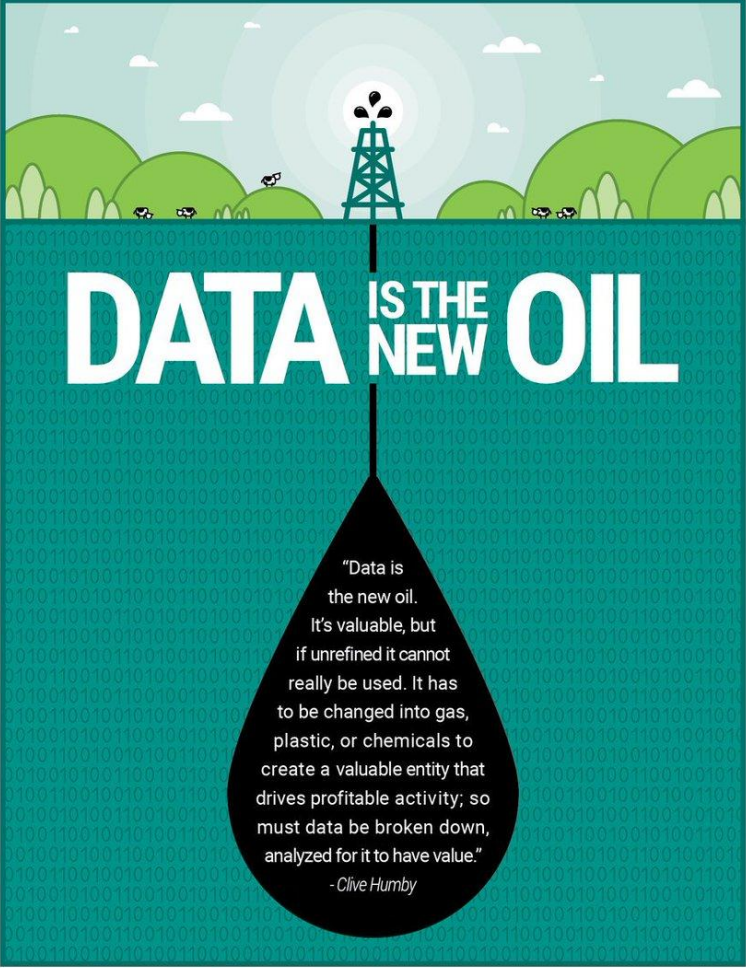
- Competência significativa em uma **área do conhecimento**, com capacidade de analisar informações e resolver problemas.
- Cá entre nós, de profissionais que saibam lidar com dados **em suas mais variadas vertentes**.

Não há como negar, DS é a Estatística Moderna (com todas as diferenças que o mundo proporcionou)!

POR QUE ESTUDAR DS?

- Porque aumentou demais a **geração e produção de dados** (internet, dispositivos móveis, sensors....)
- **Big Data** é uma realidade;
- Custo baixo (mas ainda alto) para **armazenar dados**;
- Aumento da **capacidade de processamento**;
- Evolução das tecnologias;
- Pela certeza de que a análise de dados é capaz de gerar conhecimento e **vantagens em tomadas de decisões** nas mais variadas áreas.

POR QUE ESTUDAR DS?



DATA IS THE NEW OIL

"Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, or chemicals to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value."

-Clive Humby

CONTACT GOBO FOR ORGANIZATIONAL DATA SOLUTIONS GUSHING WITH INSIGHT AT (888) 938-7779 OR GOBOINFO.COM. ©2017 GOBO LLC.



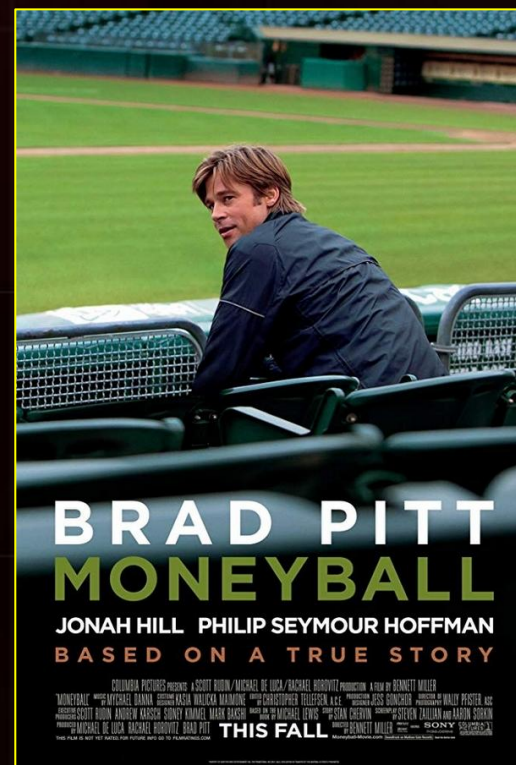
Clive Humby, um matemático londrino especializado em ciência de dados.

Google, Facebook, Instagram, Microsoft, Apple já sabem disso faz algum tempo!

ALÉM DE ESTUDAR

- Descobrir conhecimento em conjuntos de dados não é uma tarefa de programador (falaremos disso adiante).
- É necessário um conjunto de habilidades, mas principalmente é necessário abrir a mente para compreender os benefícios e as possibilidades que os dados oferecem.
- Intercale estudo com outros materiais para que possa estimular o cérebro a ser um bom analista de dados.
- Recomendações:

CONTEÚDO PERTINENTE!



VIDA REAL (MUITOS EXEMPLOS)

- Futebol: LIVERPOOL (Campeão de tudo em 2019)
 - Ian Graham (doutor em física teórica) que trabalha para o Liverpool elaborou o conceito de “domínio de campo”, que agrega diversas estatísticas coletadas no monitoramento de partidas.

(<https://www.liverpool.com/liverpool-fc-news/features/liverpool-transfer-news-jurgen-klopp-17569689>)

Negócios

Fleury cria Saúde iD, um 'one-stop shop' da saúde

1 de Setembro de 2020

FINALIZANDO

- DS é sobre você **descobrir conhecimento a partir de dados**.
- **DS é uma ciência**, lembre-se sempre disso em seus projetos!
- Dados são o **novo petróleo** (vamos falar mais sobre isso).
- Nem todo excelente programador (inclusive em python ou R) é um bom analista de dados (vamos falar mais sobre isso).
 - **Essa teoria toda é tão importante? Eu quero programar!!**
- Estudar Ciência de Dados tem muito a ver com sua **capacidade de abstração** sobre um conjunto de dados, é preciso **ampliar o olhar**, **abrir a mente!**

INTRODUÇÃO À CIÊNCIA DE DADOS

Introdução à Ciência de Dados