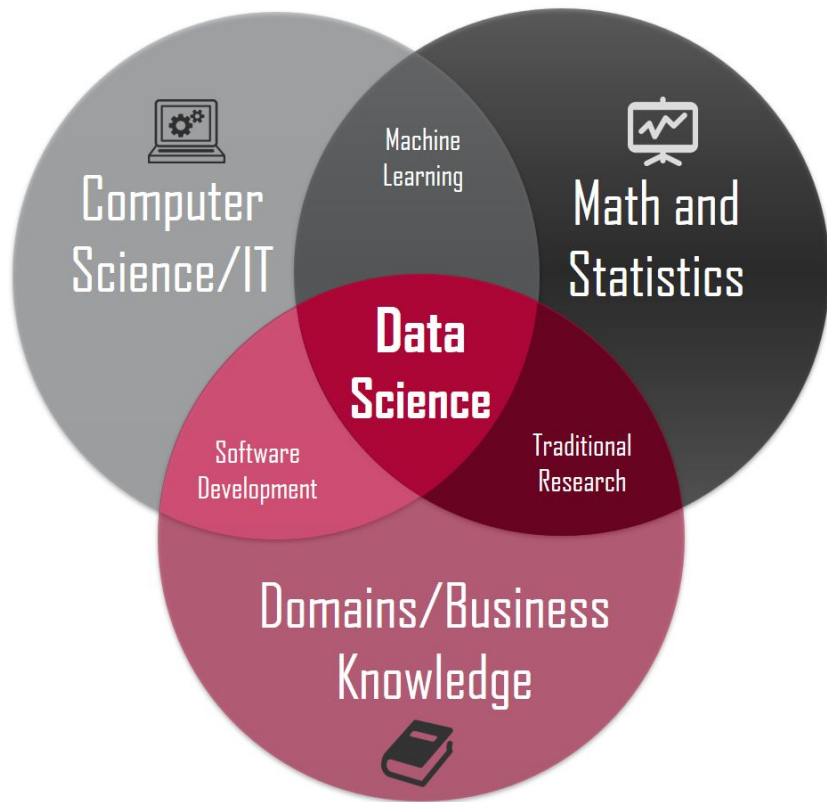
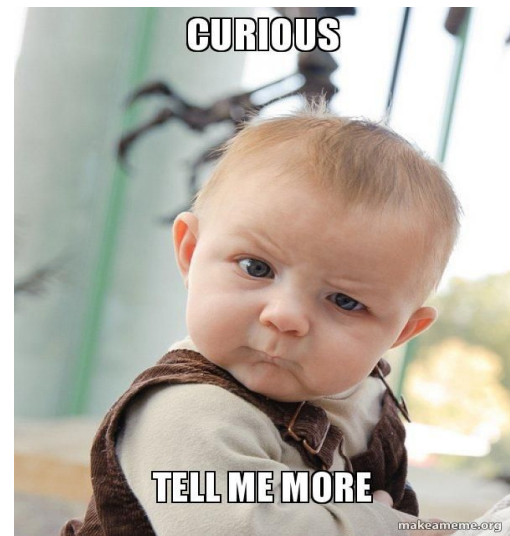


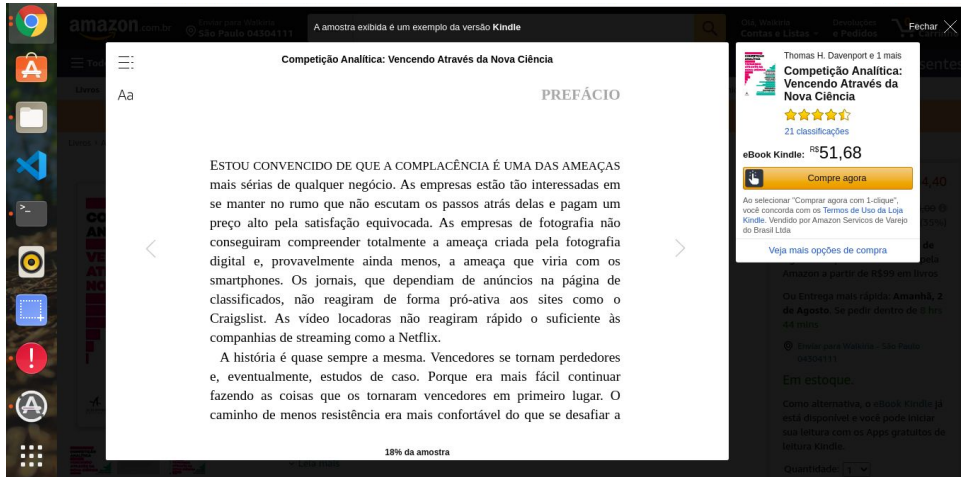
Ciência de Dados como
alavanca nos negócios



**Ser curioso e mergulhar
no problema!!!**



Intuição vs análises



Ciência

- Ciência refere-se a qualquer conhecimento ou prática sistemática
- A ciência de dados é uma parte que estuda dados dentro de um contexto da ciência
- **Método científico**

Ciência

- Ciência refere-se a qualquer conhecimento ou prática sistemática
- A ciência de dados é uma parte que estuda dados dentro de um contexto da ciência
- **Método científico: conjunto de regras para se realizar um experimento e adquirir novos conhecimentos**



Ciência



1º - Observação: É a etapa em que o pesquisador observa uma determinada matéria ou fenômeno.

2º - Elaboração do problema (fase do questionamento): Nessa etapa, o cientista ou pesquisador elabora perguntas sobre o fenômeno ou material analisado, tais como: Por que esse fenômeno ocorre?; Como esse fenômeno ocorre?; Quais são os fatores que originaram esse fenômeno?";

3º - Hipóteses: Ponto de partida da experimentação

4º - Experimento: O objetivo é encontrar a resposta para cada um dos questionamentos que foram elaborados.

5º - Análise: o pesquisador analisa cada um dos resultados para verificar se eles são suficientes para explicar cada um dos problemas levantados e também se estão de acordo com as hipóteses.

6º - Conclusão: elaboração de uma teoria

Ciência

- Ciência refere-se a qualquer conhecimento ou prática sistemática
- A ciência de dados é uma parte que estuda dados dentro de um contexto da ciência
- **Método científico: conjunto de regras para se realizar um experimento e adquirir novos conhecimentos**

reprodutibilidade



**I DON'T KNOW
WHAT IS DATA SCIENCE**

**AND AT THIS POINT
I'M TOO AFRAID TO ASK**

Ciência de dados

Área interdisciplinar que segundo o livro, Data Science para Negócios, pode ser definida como:

“um conjunto de princípios fundamentais que norteiam a extração de conhecimento a partir de dados.”

Ciência de dados

“A ciência de dados é uma abordagem multidisciplinar para extrair insights dos grandes e crescentes volumes de dados coletados e criados pelas organizações atuais. A ciência de dados abrange a preparação de dados para análise e processamento, realizando análises avançadas de dados e apresentando os resultados para revelar padrões e permitir que as partes interessadas tirem conclusões informadas.” IBM

Ciência de dados

“A ciência de dados é uma combinação de análise de dados, desenvolvimento algorítmico e tecnologia para resolver problemas analíticos. O objetivo principal é o uso de dados para gerar valor de negócios.” <https://www.analyticsvidhya.com/blog/>

3 estágios e 6 passos

Adaptado de: Dados Demais, Thomas Davenport
Por: Walkiria Resende

1 Formulação do Problema



Reconhecimento do Problema



Revisão das descobertas anteriores



2 Solução do Problema



Modelagem



Dados



Análise dos dados

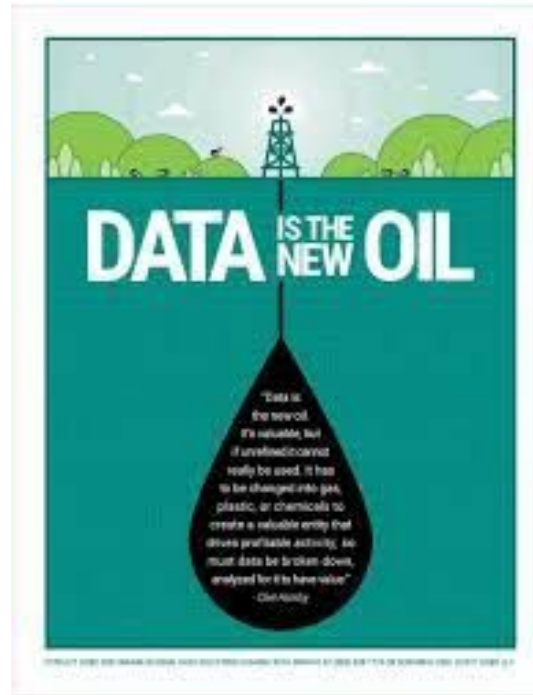


3 Resultados: Comunicado da ação

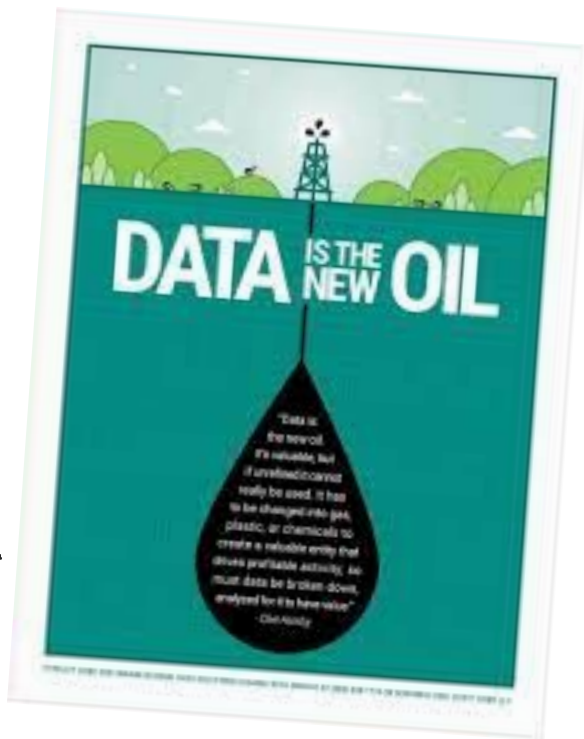


Apresentação dos resultados e adoção de providências

Porque estudar ciência de dados?



Porque estudar ciência de dados?



CASES DE SUCESSO



amazon

Walmart

CASES DE SUCESSO



Recomendação;
Predição de séries de
sucesso;

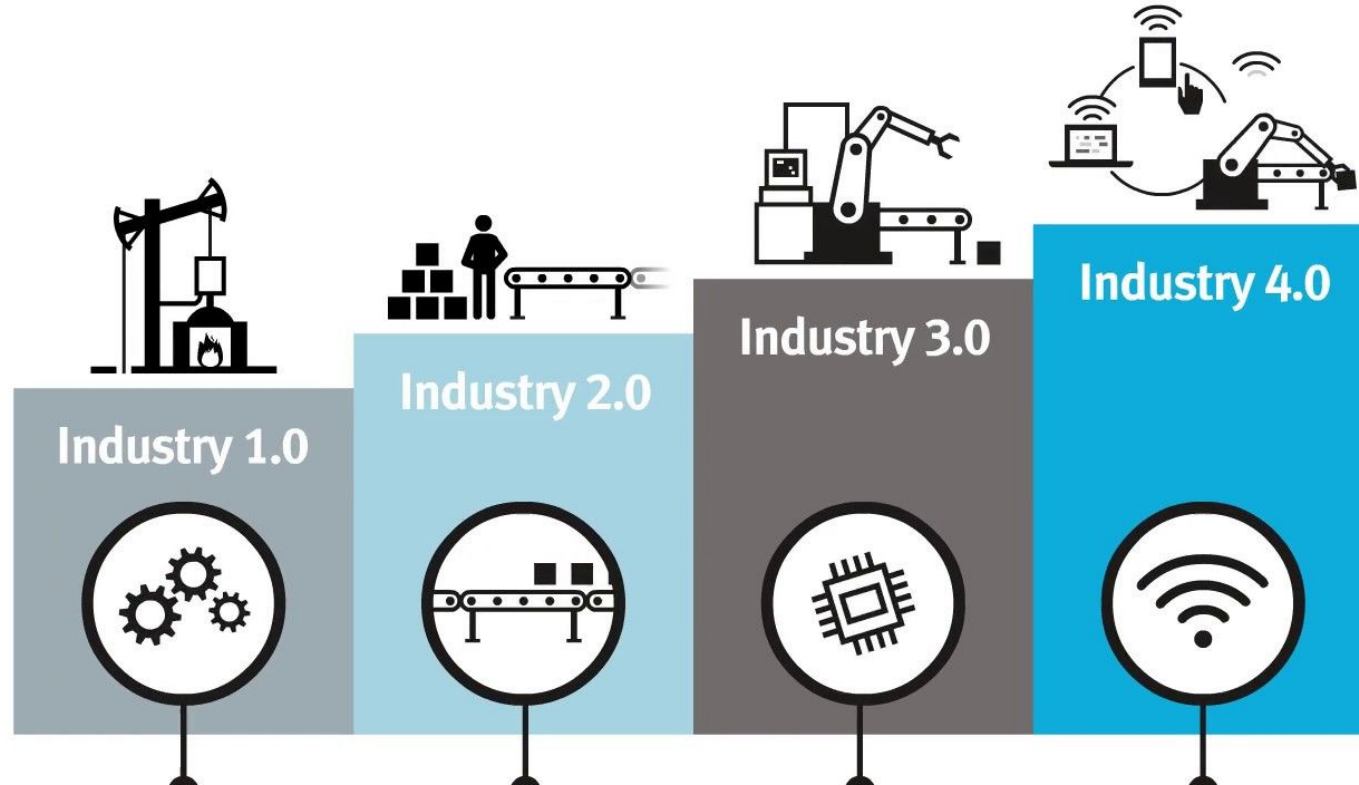
Recomendação;
Cadeia de suprimentos
(software proprietário);
Envio antecipado de
pacotes;



Precificação dinâmica;
Produtos comprados
antes de um furacão



De onde vem os dados?



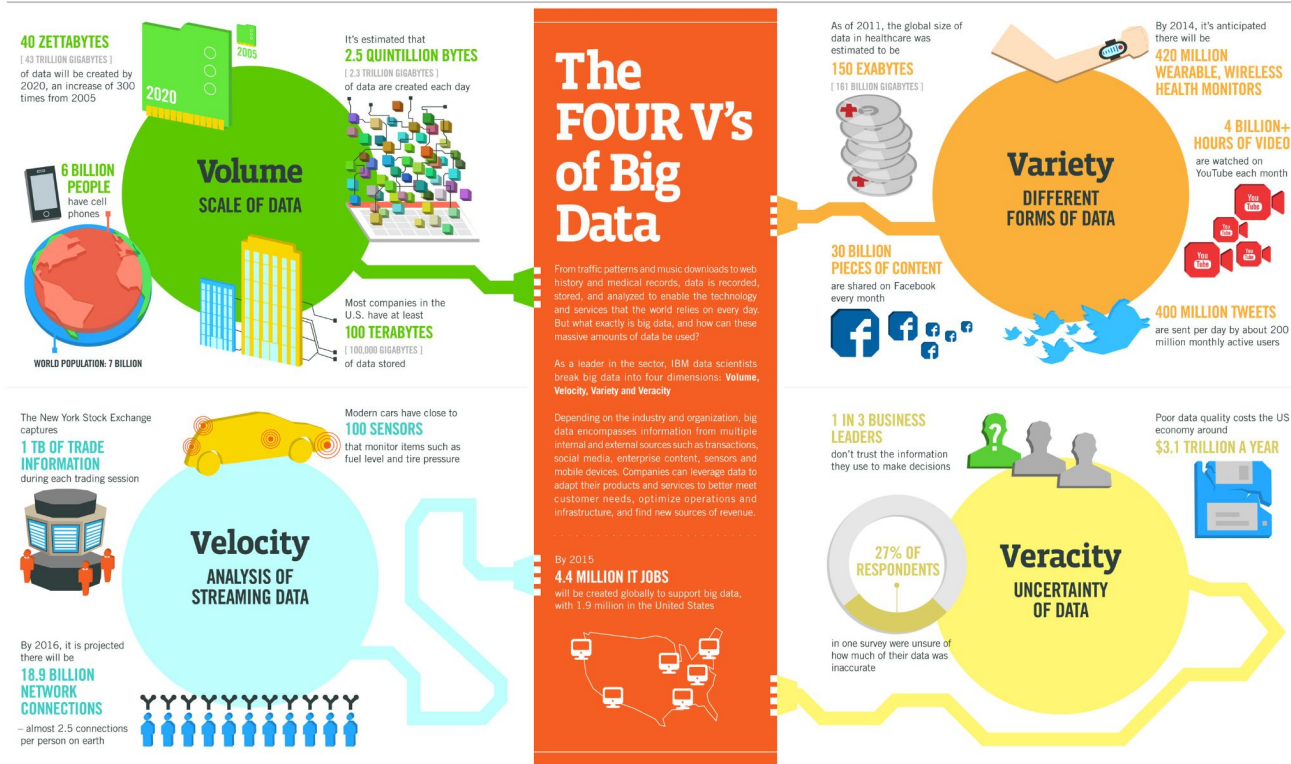
De onde vem os dados?



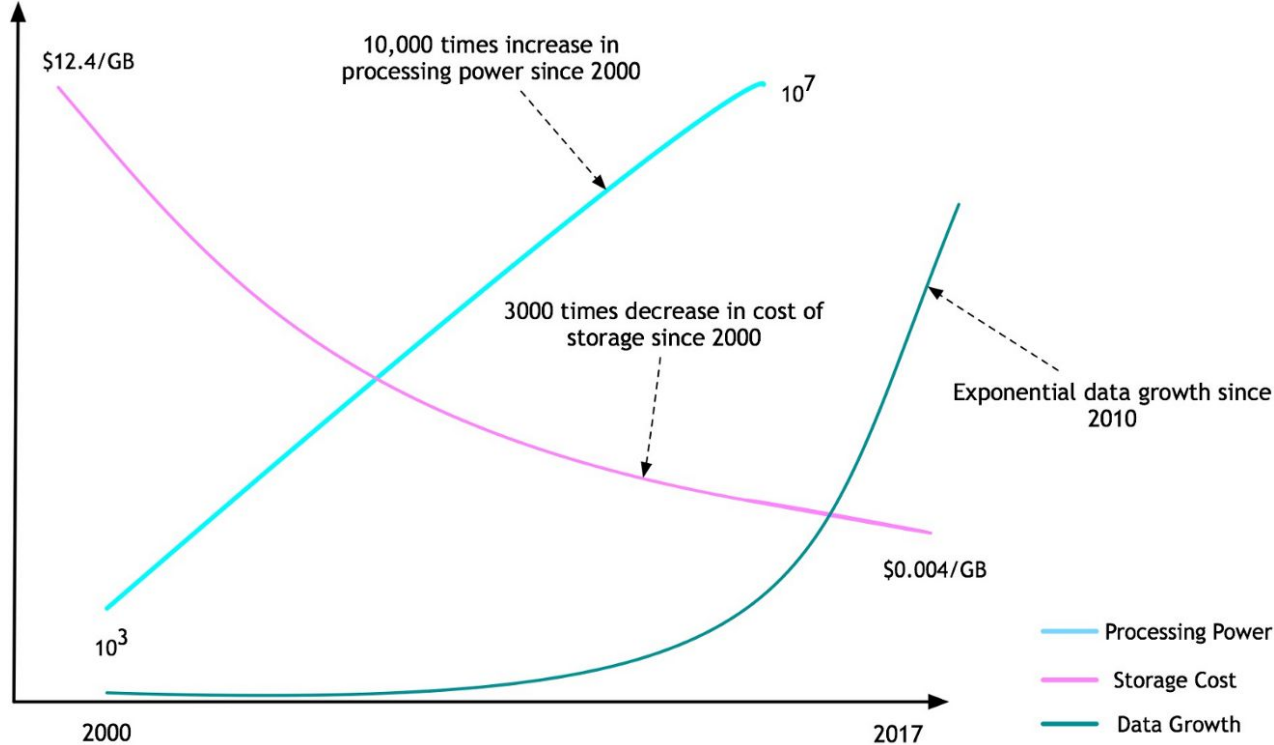
De onde vem os dados?



De onde vem os dados?



Poder computacional



Poder computacional - porque isso importa?

Algorithm	Classification/Regression	Training	Prediction
Decision Tree	C+R	$O(n^2p)$	$O(p)$
Random Forest	C+R	$O(n^2pn_{trees})$	$O(pn_{trees})$
Random Forest	R Breiman implementation	$O(n^2pn_{trees})$	$O(pn_{trees})$
Random Forest	C Breiman implementation	$O(n^2\sqrt{p}n_{trees})$	$O(pn_{trees})$
Extremely Random Trees	C+R	$O(npn_{trees})$	$O(npn_{trees})$
Gradient Boosting (n_{trees})	C+R	$O(npn_{trees})$	$O(pn_{trees})$
Linear Regression	R	$O(p^2n + p^3)$	$O(p)$
SVM (Kernel)	C+R	$O(n^2p + n^3)$	$O(n_{sv}p)$
k-Nearest Neighbours (naive)	C+R	—	$O(np)$
Nearest centroid	C	$O(np)$	$O(p)$
Neural Network	C+R	?	$O(pn_{l_1} + n_{l_1}n_{l_2} + \dots)$
Naive Bayes	C	$O(np)$	$O(p)$

n: número de amostras
p: número de características

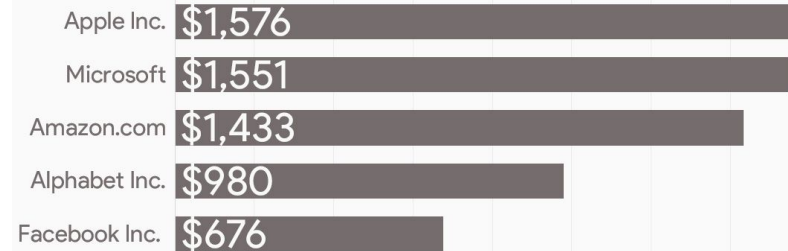
Alavanca

The five largest companies

in 2010



vs 2020

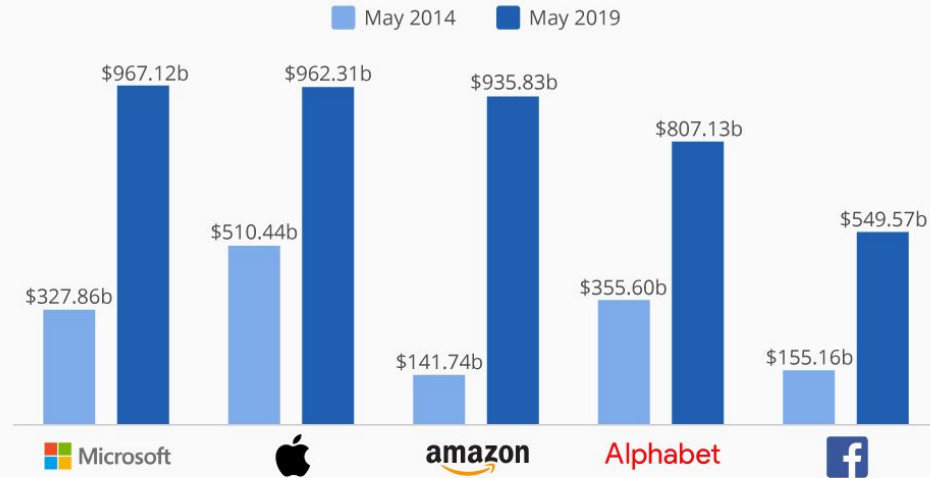


Largest companies in terms of market capitalization.
All figures in 2020 billion US dollars; 2020 companies as of July 1st.
Source: ycharts.com

Alavanca

GAFAM Market Cap Jumps \$2.7 Trillion Over Past Five Years

Market capitalization of selected tech companies in May 2019 and May 2014*



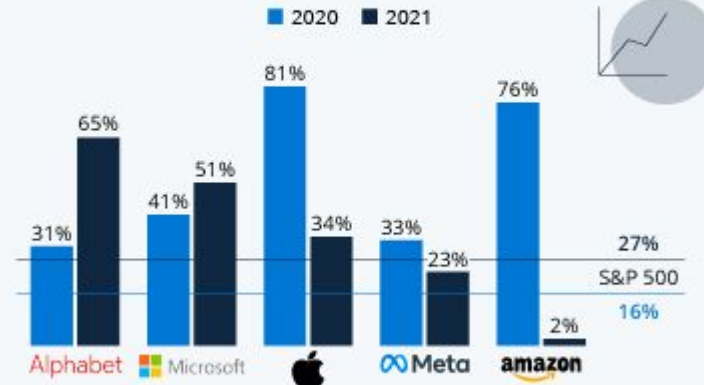
* data as of May 2 of the respective year
Source: YCharts

© StatistaCharts

statista

Tech Giants Largely Rally Through Another Pandemic Year

Stock price change of the world's largest tech companies in 2020 and 2021



Source: Yahoo! Finance

©

statista

Competidor Análítico

“uma organização que utiliza extensiva e sistematicamente a análise para pensar melhor e ter um desempenho melhor do que a concorrência”

Four Pillars of Analytic Competition

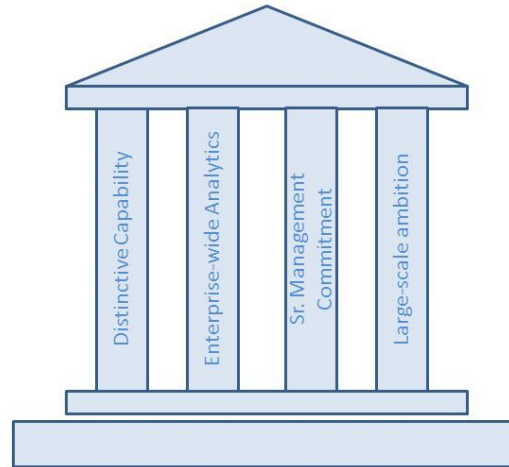


Figure 2-1 : Davenport and Harris

Marketing

- Precificação
- Localização de Lojas
- Customização de sites

Recursos Humanos


- Quais pessoas contratar
- Quais pessoas tendem a deixar a empresa
- Qual o nível de remuneração


Cadeia de Fornecimento


- Que níveis de estoque manter
- Onde instalar centros de distribuição
- Carregamento de caminhões

Inúmeras possibilidades

Mais casos interessantes: o Furacão Frances

 **BRASIL.com**
Ajuda | Página simplificada

 70 ANOS
1938 - 2008

 This page has been archived and is no longer updated.
[Find out more about page archiving.](#)

Atualizado às: 02 de setembro, 2004 - 04h39 GMT (01h39 Brasília)

[Envie por e-mail](#) [Versão para impressão](#)

Furacão Frances pode chegar aos EUA em 36 horas
Rafael Gomez de Miami

O Centro Nacional de Furacões dos Estados Unidos emitiu nesta quarta-feira um alerta para a costa leste da Flórida, devido à aproximação do furacão Frances.

O alerta significa que, de acordo com os meteorologistas, a região, que inclui as populosas cidades de Miami, Fort Lauderdale e West Palm Beach, pode ser atingida pelo furacão dentro de 36 horas.

O Frances, um furacão de categoria 4 na escala de força que vai até 5, estava na noite desta quarta-feira ao norte/noroeste da República Dominicana, no Caribe, com ventos que superam 220 km/h e chuvas torrenciais.

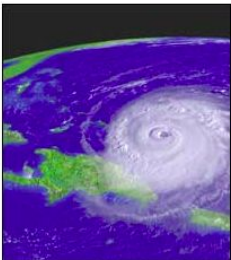


Foto de satélite mostra o furacão Frances ao norte da República Dominicana nesta quarta-feira.

NOTÍCIAS RELACIONADAS

- ▶ **Tempestade tropical avança ao norte do Japão**
01 de setembro, 2004 | Notícias
- ▶ **Furacão na Flórida deve custar US\$ 20 bi, dizem seguradoras**
17 de agosto, 2004 | Notícias
- ▶ **Bush visita região atingida por furacão na Flórida**
15 de agosto, 2004 | Notícias
- ▶ **Furacão Charley atinge costa da Flórida**
14 de agosto, 2004 | Notícias
- ▶ **Furacão Charley causa ao menos 15 mortes na Flórida**
14 de agosto, 2004 | Notícias
- ▶ **Furacão Charley chega a Cuba e se dirige a Havana**
13 de agosto, 2004 | Notícias
- ▶ **Furacão Charley e tempestade Bonnie se dirigem aos EUA**
12 de agosto, 2004 | Notícias

O furacão Frances estava a caminho, avançando pelo Caribe, ameaçando atingir a costa atlântica da Flórida. Os residentes se mudaram para terrenos mais elevados, porém distantes, em Bentonville, Arkansas. **Executivos das lojas Walmart decidiram que a situação oferecia uma grande oportunidade para uma de suas mais recentes armas orientadas em dados: a tecnologia preditiva.** Uma semana antes de a tempestade atingir a costa, Linda M. Dillman, **diretora executiva** de informação, pressionou sua equipe para trabalhar em previsões baseadas no que havia acontecido quando o furacão Charley apareceu, várias semanas antes. Com o apoio dos trilhões de bytes de histórico de compras contidos no banco de dados do Walmart, ela sentiu que a empresa poderia “começar a prever o que aconteceria, em vez de esperar que acontecesse”. (Hays, 2004)

Mais casos interessantes: o Furacão Frances

- O que podemos prever?
 - Água? DVDs?

Mais casos interessantes: o Furacão Frances

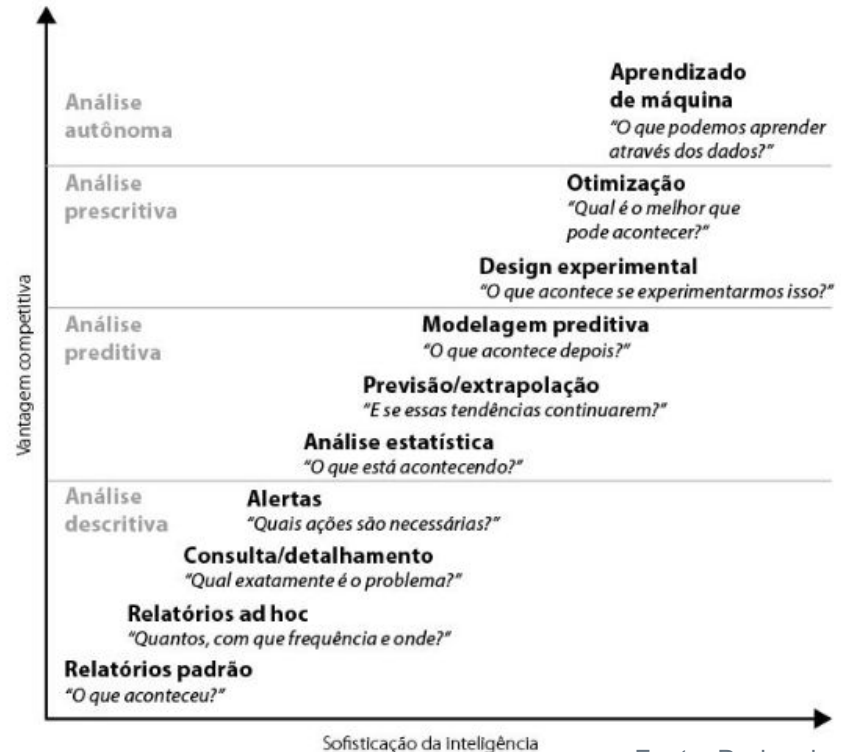
- O que podemos prever?
 - Água? DVDs?

O New York Times (Hays, 2004) relatou que:

“... especialistas exploraram os dados e descobriram que as lojas realmente precisariam de certos produtos — e não apenas das habituais lanternas. ‘Não sabíamos, no passado, que havia tido um aumento nas vendas de Pop-Tarts de morango, sete vezes acima do normal, antes de um furacão’, disse a Sra. Dillman em uma entrevista recente. ‘E o principal produto pré-furacão mais vendido era a cerveja.’”

Vantagem competitiva

- Aumento contínuo da quantidade de dados
- Tecnologias mais inovadoras
- Estratégia



Vantagem competitiva

- Vamos pensar em um exemplo!



Data Driven

A orientação por dados é sobre a construção de ferramentas, habilidades e, mais importante, uma cultura que atua nos dados.

Data Driven

A orientação por dados é sobre a construção de ferramentas, habilidades e, mais importante, uma cultura que atua nos dados.

Alguns pré-requisitos:

- **Coletar dados:** dados são o ingrediente chave, mas não qualquer dado. O dado precisa ser relevante para o problema em questão. Também precisa ser também deve ser oportuno, preciso, limpo, imparcial; e talvez o mais importante, tem que ser confiável.



Data Driven

- **Acesso aos dados:** Os dados devem ser acessíveis e consultáveis. É preciso armazenar os dados de forma que possam ser usados, isso quer dizer, usar ferramentas como bancos de dados sql, NOSql, Hadoop de acordo com as necessidades.

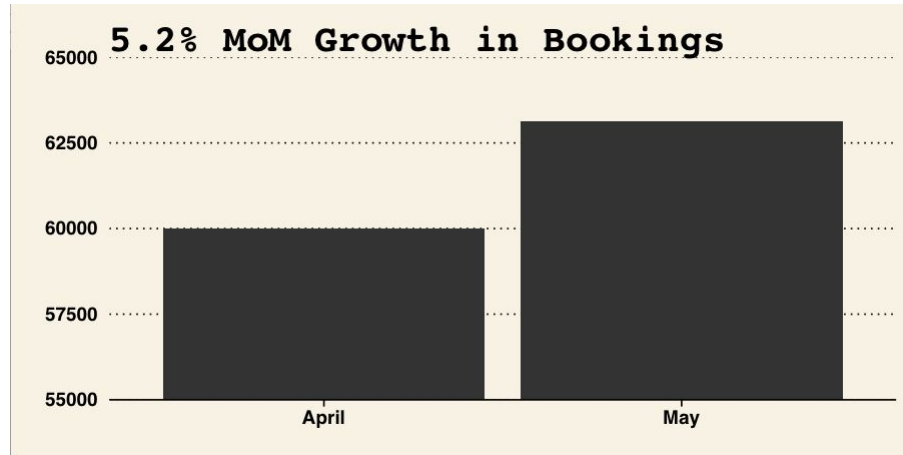
Outro ponto importante é ter uma cultura de compartilhamento de dados, para que eles possam oferecer uma experiência completa ao usuário

E por fim, usar ferramentas para extrair informação desses dados.



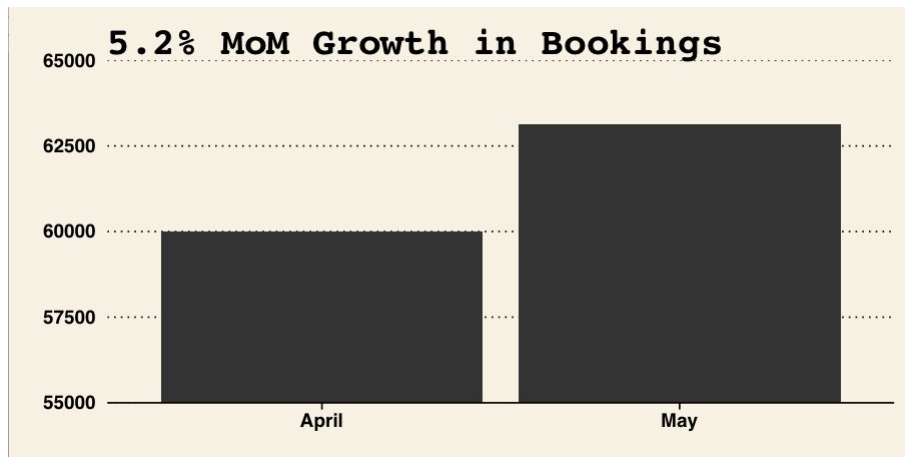
Data Driven

- Reporte



Data Driven

- Reporte



Vamos refletir sobre o que pode acarretar esse valor

Data Driven

- Alerta

“Ding, ding, ding! Web-app server #14 has just averaged more than 98% CPU utilization for the last five minutes.”

Data Driven

- Alerta e Reporte para Análises

Reporte

O processo de organizar dados em resumos informativos para monitorar o desempenho de diferentes áreas de uma empresa

Análise

Transformando ativos de dados em insights competitivos que conduzirão decisões e ações de negócios usando pessoas, processos e tecnologias



Dicas!

Se você deseja ser um **cientista de dados, mão na massa**, é interessante estudar bastante teoria e ir já começar a testar, por exemplo, participar de uma competição no Kaggle.

Para cada tópico de DS, procure os algoritmos e métodos principais, e entenda o que eles fazem, como fazem, parâmetros, aplicações

Agora, se você vai trabalhar com dados, mas não será o responsável por implementações (**usuário, cliente, líder**), é interessante conhecer os métodos que existem e entender de maneira geral a aplicação de cada um

