

Big data: a organização do dia a dia em um disco rígido

Pint of Science 2017

Marcus Nunes

marcus@marcusnunes.me

Natal, 17 de maio de 2017

Departamento de Estatística
Universidade Federal do Rio Grande do Norte

O Que é Big Data?

O Que é Big Data?

- Não existe consenso a respeito de uma definição sobre o que realmente é big data
- A área ainda é nova; não houve tempo para o conhecimento sedimentar
- Em geral, diz respeito a áreas do conhecimento onde as ferramentas de análise de dados tradicionais não são a melhor escolha possível

O Que é Big Data?

- Big Data são os dados que possuem 3 V:
- Volume
- Velocidade
- Variedade

Que Problemas Podem ser Resolvidos com Big
Data?

Que Problemas Podem ser Resolvidos com Big Data?

- Considere três tamanhos diferentes de loja:
 - Loja Pequena (comércio de bairro)
 - Loja Média (Nordestão)
 - Loja Grande (Amazon)
- Qual é o tipo de relacionamento que cada uma destas lojas possui com seus clientes?

Que Problemas Podem ser Resolvidos com Big Data?

- Loja Pequena: o cliente recorrente recebe tratamento personalizado, diferente de um cliente eventual
- Loja Média: o cliente recorrente de uma loja média recebe tratamento padrão, idêntico ao de um cliente eventual
- Loja Grande: o cliente recorrente de uma loja grande recebe tratamento personalizado, diferente de um cliente eventual

Quem Trabalha com Big Data?

Quem Trabalha com Big Data?

- Competências de um profissional 100% capacitado para trabalhar com Big Data:
 - Estatística
 - Programação
 - Negócios
 - Conhecer bem a área de atuação (internet, marketing, área financeira etc)

Quem Trabalha com Big Data?

- Que tipo de profissionais temos no momento?
 - Bons estatísticos e matemáticos que escrevem códigos sem otimização
 - Bons cientistas da computação que entendem um pouco de estatística e matemática
 - Bons cientistas da computação que entendem um pouco de negócios, depois de muita experiência na área
 - Especialistas em alguma área de atuação
 - Gerentes que sabem fazer estas pessoas trabalharem juntas

Quem Trabalha com Big Data?

- Cientista de Dados (*Data Scientist*) é o novo nome para Estatístico
- Alguns dizem que o Cientista de Dados é um Estatístico que mora em São Francisco e usa um Mac
- No fundo, ambos são a mesma coisa, embora uma destas profissões trabalhe melhor seu marketing pessoal

Exemplos

Target

- Como a Target descobriu que uma adolescente estava grávida antes do pai dela descobrir



Fonte: <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father>

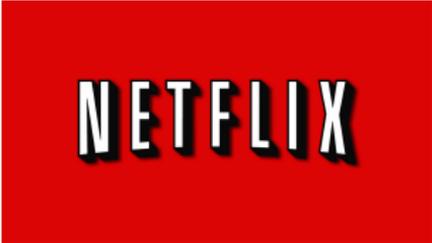
- Compartilhamos dados com empresas sempre que fazemos compras
- As empresas usam estes dados para traçar o perfil dos clientes
- Com isto, podem enviar ofertas direcionadas para cada segmento de público

- Pai furioso foi a uma loja da Target em Minneapolis, exigindo falar com o gerente
- Dizia que a filha recebeu, pelo correio, ofertas de roupas de bebê e berços
- Detalhe: a filha ainda era aluna do Ensino Médio

- Na lógica do pai, isto estaria encorajando a filha a engravidar
- O gerente da Target se desculpou e disse que ia averiguar o que tinha acontecido
- Nada de estranho foi descoberto; mesmo assim, o gerente ligou para o pai uns dias depois e se desculpou novamente

- Exemplo de perfil traçado pela Target:
- Imagine uma cliente chamada Jenny Ward, 23 anos, moradora de Atlanta
- Se em março ela comprou
 - hidratante de manteiga de cacau
 - uma bolsa grande
 - suplementos de zinco e magnésio
 - um tapete azul claro
- Segundo o modelo da loja, há 87% de chance de ela estar grávida e dar a luz no fim de agosto

- No caso específico da Target, o modelo deles é capaz de enviar correspondência para as clientes grávidas em momentos bastante específicos da gestação
- Além disso, serve para indicar como organizar fisicamente as lojas (por exemplo, colocar fraldas ao lado de cerveja)
- É possível descobrir que, quando casam, as pessoas comprar tipos diferentes de cereal

The Netflix logo is centered in the upper half of the slide. It consists of the word "NETFLIX" in a bold, white, sans-serif font with a black drop shadow, set against a solid red rectangular background.

- Como a Netflix fez engenharia reversa em Hollywood
- Para entender o que as pessoas procuram em filmes, a empresa criou 76.897 micro-gêneros de filmes

Fonte:

<http://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/>

- Gêneros de filmes em locadoras? Ação, Aventura, Comédia, Drama etc.
- Gêneros de filmes na Netflix? African-American Crime Documentaries, Scary Cult Movies from the 1980s, Feel-good Romantic Spanish-Language TV Shows, Visually Striking Latin American Comedies e mais
- Assim, é mais fácil sugerir filmes que podem agradar o usuário

- Por exemplo, O Tesouro de Sierra Madre e Os Vingadores - A Era de Ultron são ambos filmes de aventura



- É possível que ambos os filmes estivessem numa mesma prateleira de uma locadora
- Na Netflix, um dos filmes poderia ser classificado como “Aventura Dramática Passada no México na Década de 1920” e o outro como “Aventura com Grupo de Super-Heróis e Grande Orçamento”
- Notem como é possível especificar o gênero do filme tanto quanto quisermos

Obrigado!