

Big Data Analytics

MSc. Wesley Lima

Roteiro

- Por que Big Data Analytics?
-
- O que é Big Data Analytics?
-
- Big Data Analytics Estratégias
-
- Alguns termos importantes
-
- Ferramentas, Linguagens e Bibliotecas



Por que Big Data Analytics?

1. Tornar organizações mais inteligentes e eficientes.





Por que Big Data Analytics?

2. Otimizar operações comerciais de forma personalizada para o cliente.





Por que Big Data Analytics?

3. Redução de Custos.



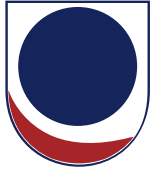
Parkland Hospital uses analytics and predictive modelling to identify high-risk patients and predict likely outcomes once patients are sent home. As a result, Parkland reduced 30-day readmissions for patients with heart failure, by 31 percent, saving \$500,000 annually.





O que é Big Data Analytics?





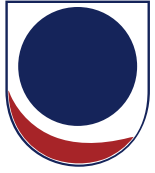
O que é Big Data Analytics?

Analytics:

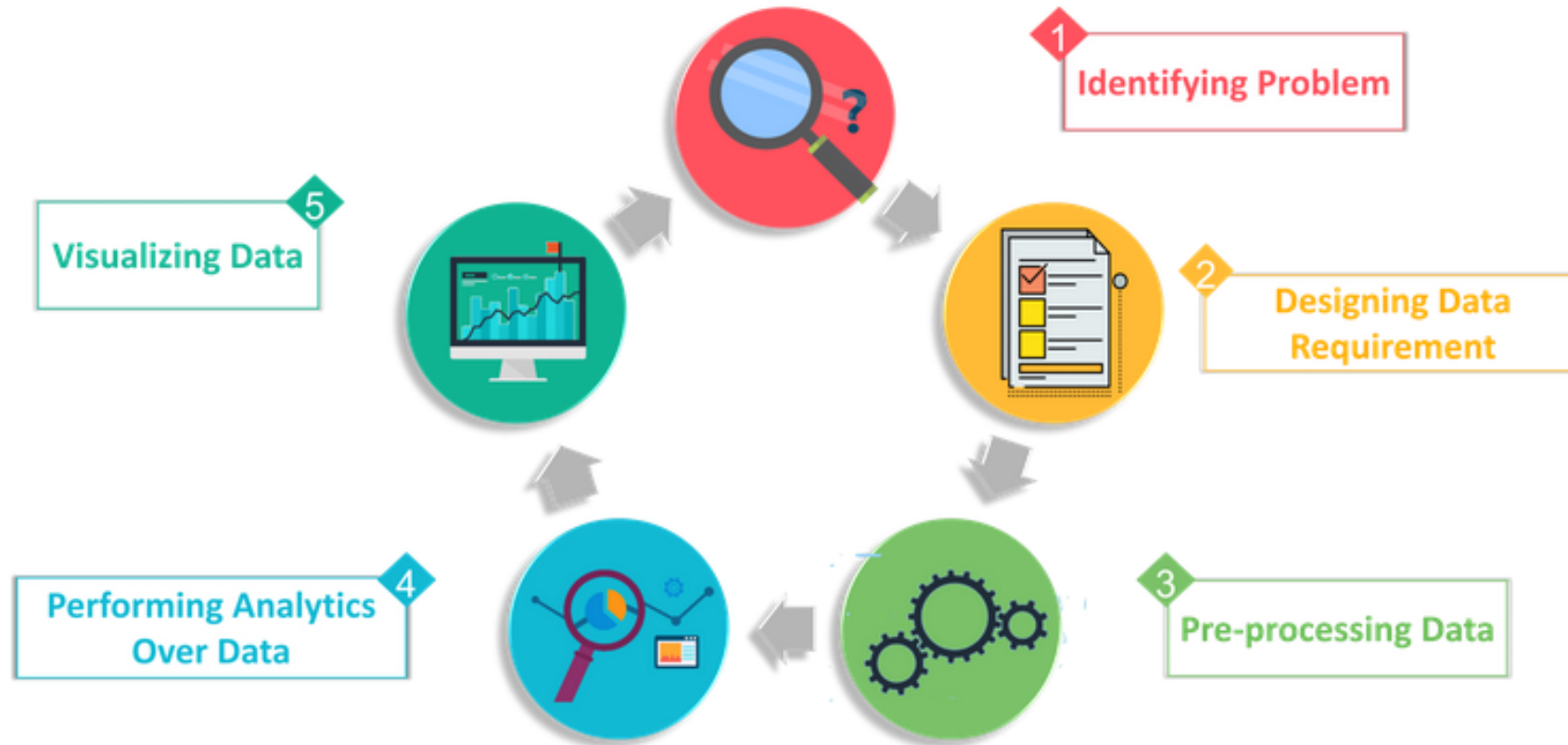
Habilidade de utilizar dados, realizar análises e utilizar raciocínio sistemático para conduzir a um processo de tomada de decisão mais eficiente;

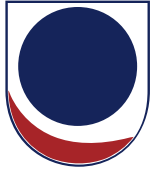
Big Data Analytics:

Habilidade de examinar grandes e diferentes tipos de dados para descobrir padrões ocultos, correlações e outros insights;



Big Data Analytics Estratégias





Definições

O que é um modelo?

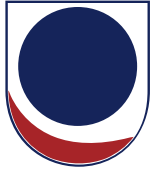
O que é um modelo matemático / estatístico?

O que é um modelo preditivo?

O que é um modelo descritivo?

O que é correlação?





Definições

O que é um modelo?

É uma representação de um objeto, sistema ou ideia em uma forma simplificada.

O que é um modelo matemático / estatístico?

É uma representação ou interpretação simplificada da realidade, ou uma interpretação de um fragmento de um sistema, segundo uma estrutura de conceitos mentais ou experimentais.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$



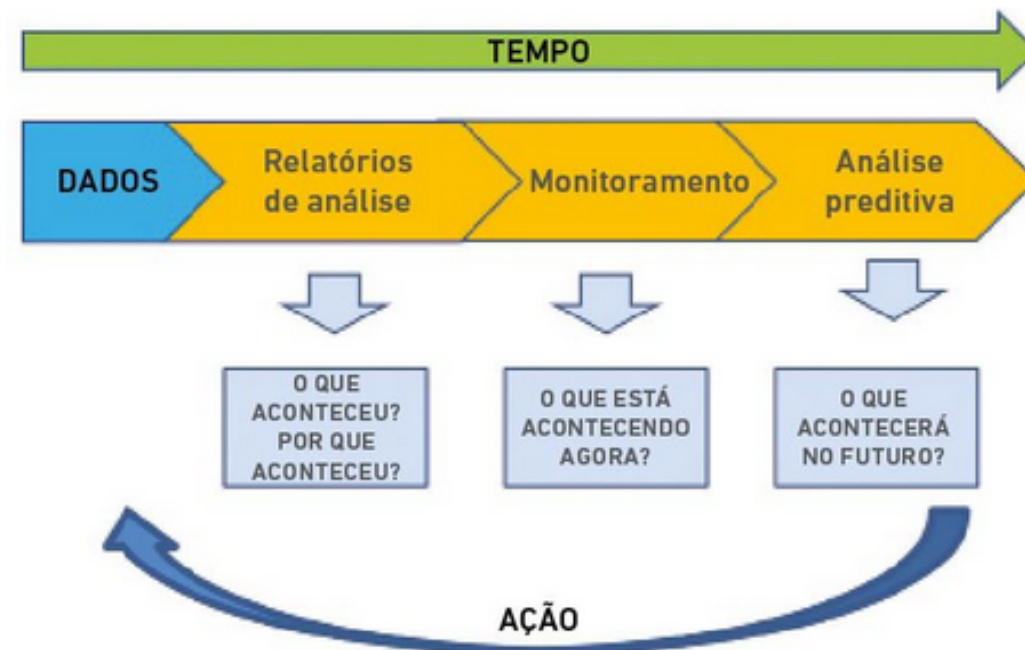
Definições

O modelo Preditivo

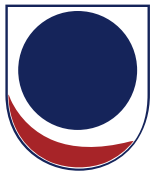
Um modelo preditivo é uma “função matemática” que, aplicada a uma massa de dados, consegue identificar padrões ocultos e prever o que poderá ocorrer;

O kernel da análise preditiva é baseado na captura das relações entre as variáveis explicativas e a variável resposta.

Predição vs Previsão



Fluxo do modelo Preditivo



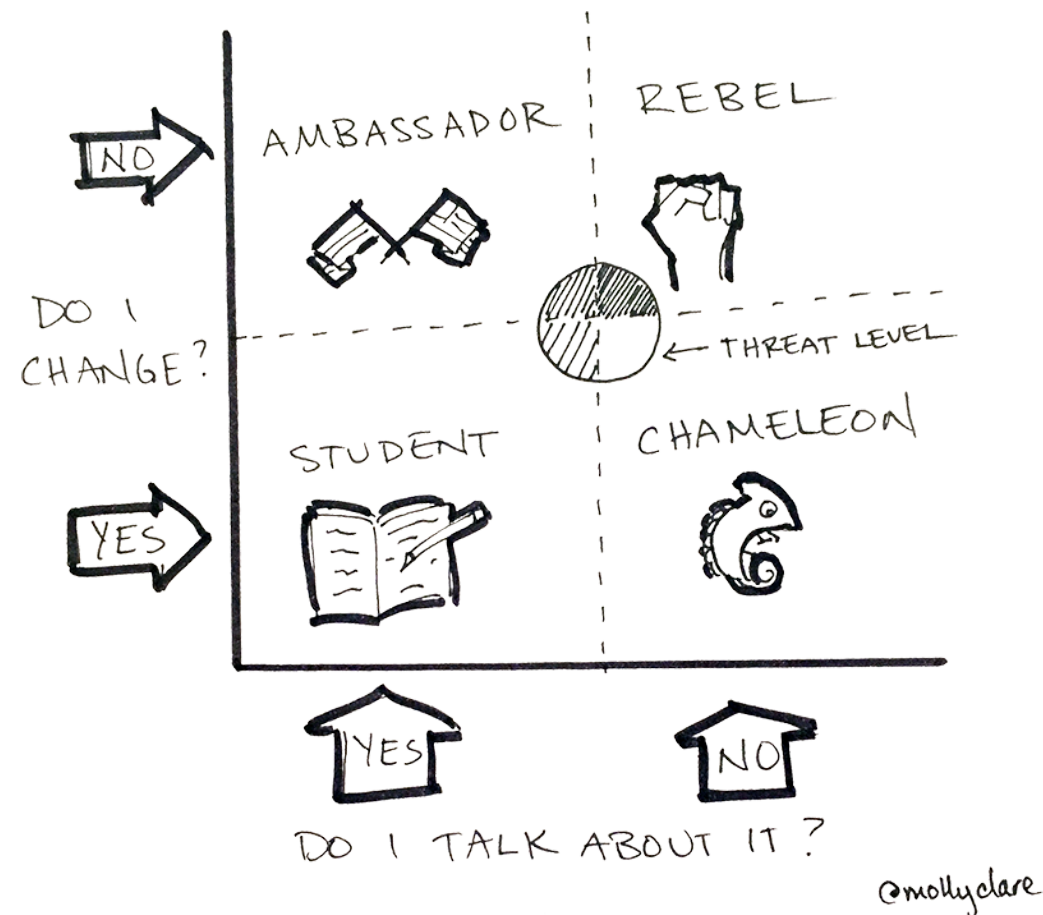
Definições

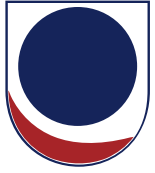
O modelo Descritivo

Um modelo descritivo quantifica as relações nos dados;

Um modelo descritivo responde: o que são os dados, a que se referem, o que nos dizem e sobre o que estão baseados;

Os modelos descritivos identificam relações entre clientes e produtos.





Definições

O modelo de Decisão

Descrevem a relação entre todos os elementos de uma decisão;

Geralmente são usados para desenvolver lógica de decisão ou um conjunto de regras comerciais ou organizações que produzirão a ação desejada para cada cliente ou circunstância;





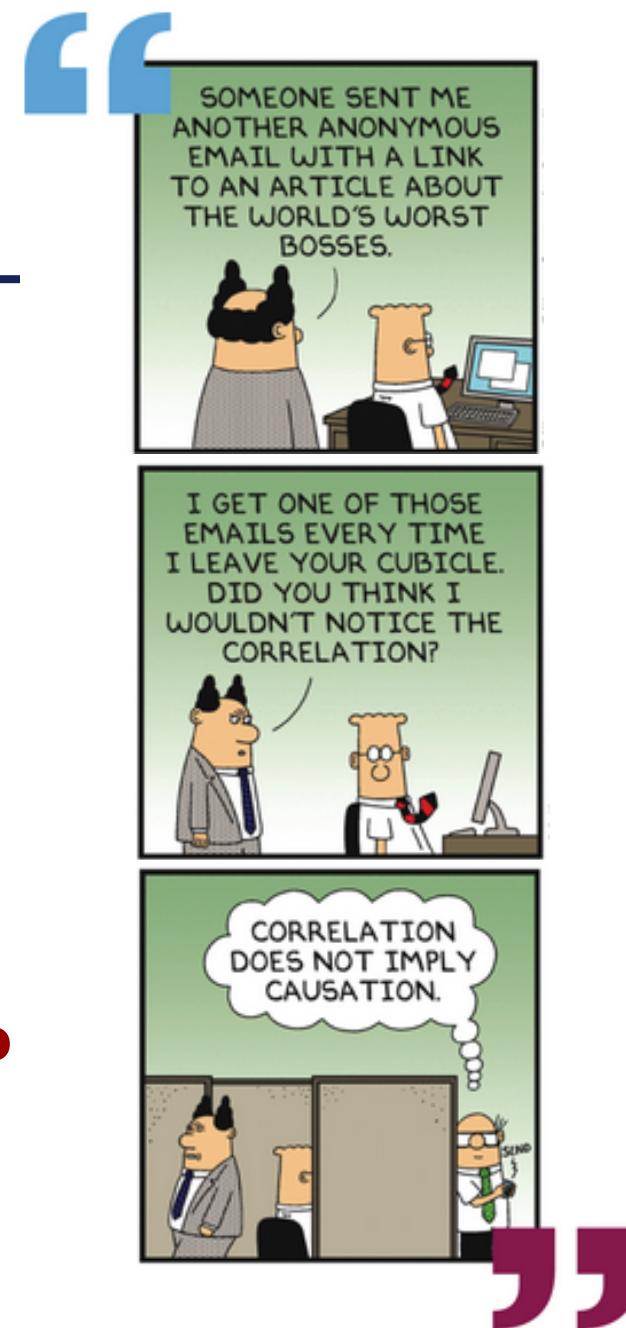
Definições

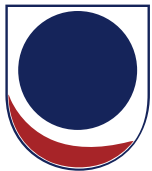
O que é correlação?

Correlação é uma medida da associação entre duas variáveis;

O que é casualidade?

Causalidade é a relação entre um evento (a causa) e um segundo evento (o efeito), em que o segundo acontecimento é entendido como uma consequência do primeiro.





Definições



Em uma universidade fictícia as pessoas verificaram que **80% dos alunos desistentes** de um determinado curso possuíam notas abaixo da média.



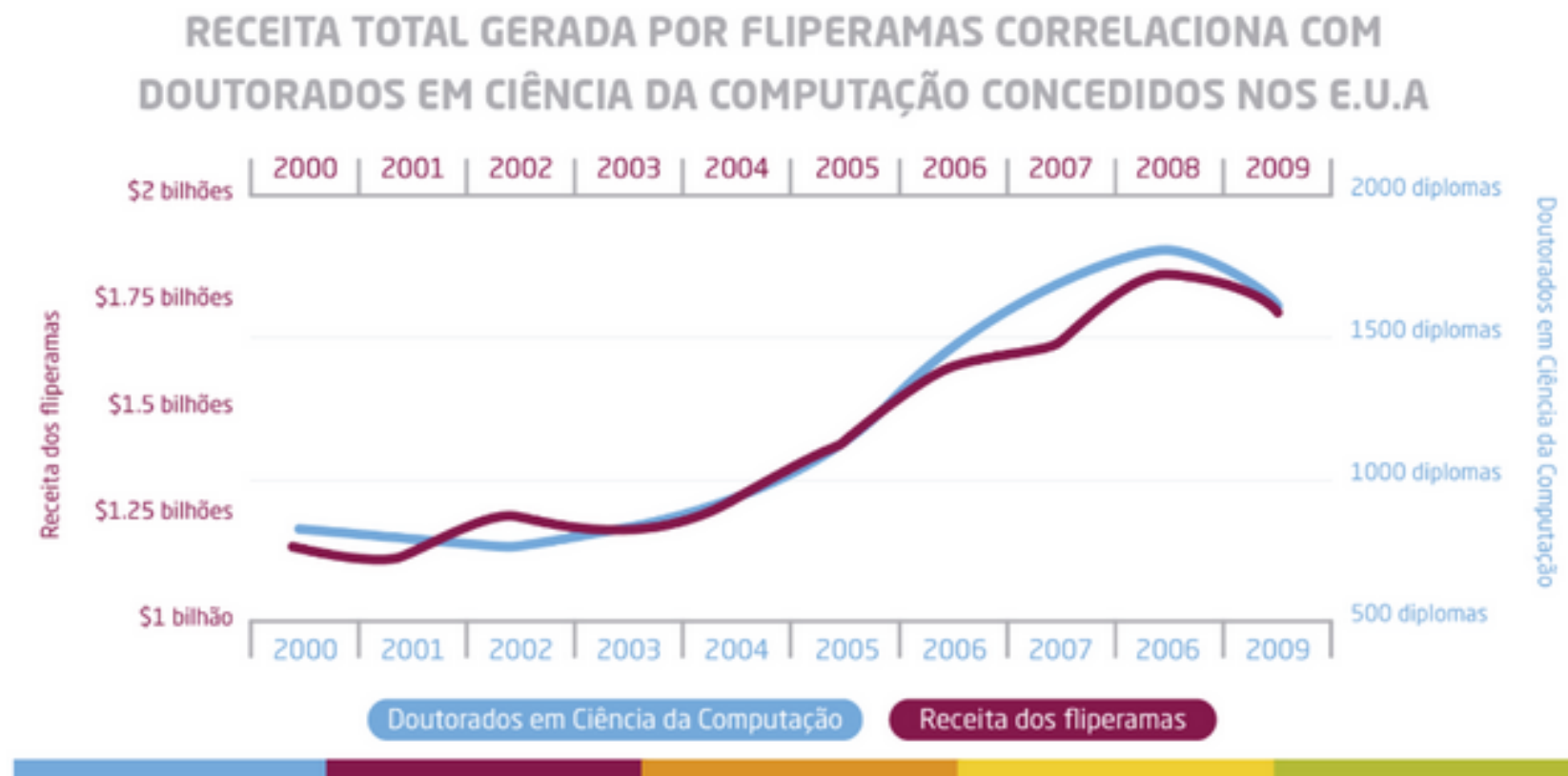
Logo, concluíram que os alunos desistiam do curso porque estavam com notas baixas. Nesse caso, notamos que existe uma **correlação** entre notas baixas e desistência, mas, não necessariamente, uma **causalidade**.

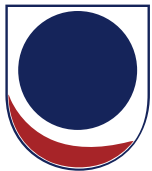


Essa falsa causalidade ocorre porque nossos cérebros buscam automaticamente por padrões. **Este processo de busca é falho**, e resulta na detecção de padrões que não existem de verdade, o que chamamos de **falácias lógicas**.



Definições





Correlação - Amazon

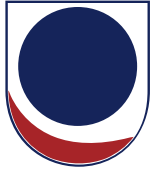
Como era feita a recomendação de livros?

Fazer recomendações com base no histórico de compras dos clientes;

Posteriormente: Comparar os produtos e as associações entre eles, correlação.

Como a Netflix recomenda filmes?





Ferramentas, Linguagens e Bibliotecas

Ferramentas

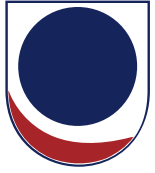
- **Hadoop;**
- **Pig;**
- **Hive;**
- **Spark;**
- **MapReduce**

Linguagens

- **Python;**
- **R;**
- **Java;**
- **SQL;**
- **SCALA.**

Bibliotecas

- **Scikit;**
- **TensorFlow;**
- **Pytorch;**
- **Keras;**
- **Caffe.**



Questões

- 1. Como uma empresa deve se preparar para trabalhar com Big Data Analytics?**
- 2. Apresente e descreva ferramentas, linguagens e bibliotecas que são mais utilizadas em Big Data?**
- 3. Um modelo preditivo deve ser criado a partir dos dados brutos?**
- 4. Como encontrar correlações em dados não estruturados, dados textuais de redes sociais, por exemplo?**



Obrigado
Até a próxima aula...