CURSO ANÁLISE DE MERCADO NO FUTEBOL

Analytics

Eduardo Cecconi

Realização: FMF

Curadoria e execução: FOOTURE

Parceiro social: Observatório da Discriminação Racial no Futebol

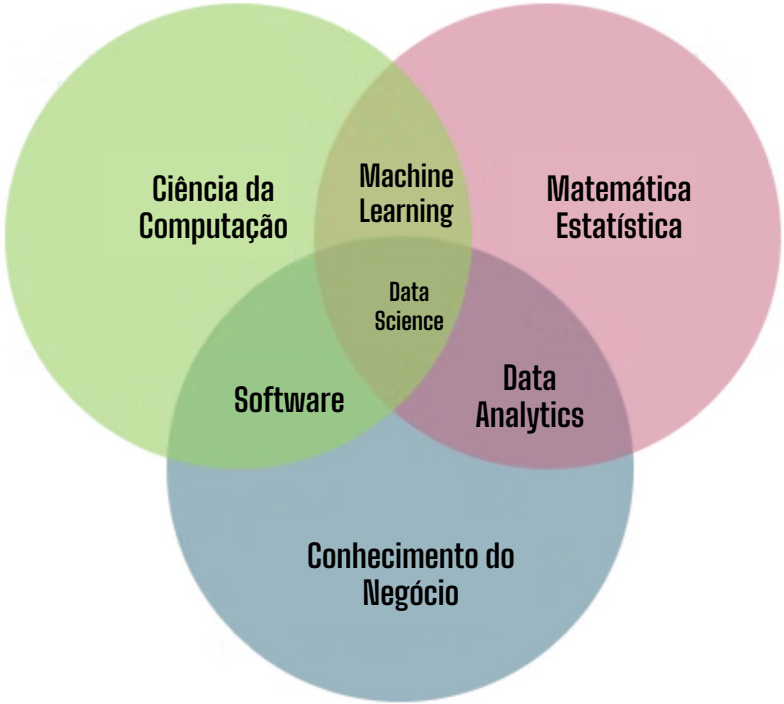
Analytics

Data Science: conjunto de técnicas de programação destinadas a:

- Coletar
- Tratar
- Manipular
- Organizar
- Analisar
- Extrair
- **e Apresentar dados**, entre outros, para subsidiar a tomada de decisão.

Analytics

Data Analytics



- É uma ramificação da Data Science destinada a **responder às perguntas que orientam a pesquisa** realizada
- Requer **conhecimento do negócio** e de **matemática/estatística**

Analytics

Esqueçam o Big Data – o que utilizamos é SMART DATA

Smart Data é o Big Data processado, ou seja, são **selecionados/filtrados** os dados que realmente importam para a modelagem.

Analytics

Smart Data

- **Reduz o volume** de dados e, consequentemente, a **demanda computacional** (armazenamento e processamento)
- **Aumenta** a **velocidade** e a **agilidade** do processo
- **Tende a aumentar** também a **precisão** e o **valor** da informação, porque vai “direto ao ponto”
- **Data Wrangling:** técnicas de manipulação para organizar o banco de dados

Analytics

Objetivos

- Oferecer **Insights**
- Amparar a **Tomada de Decisão**
- Auxiliar na **Resolução de Problemas**

Analytics

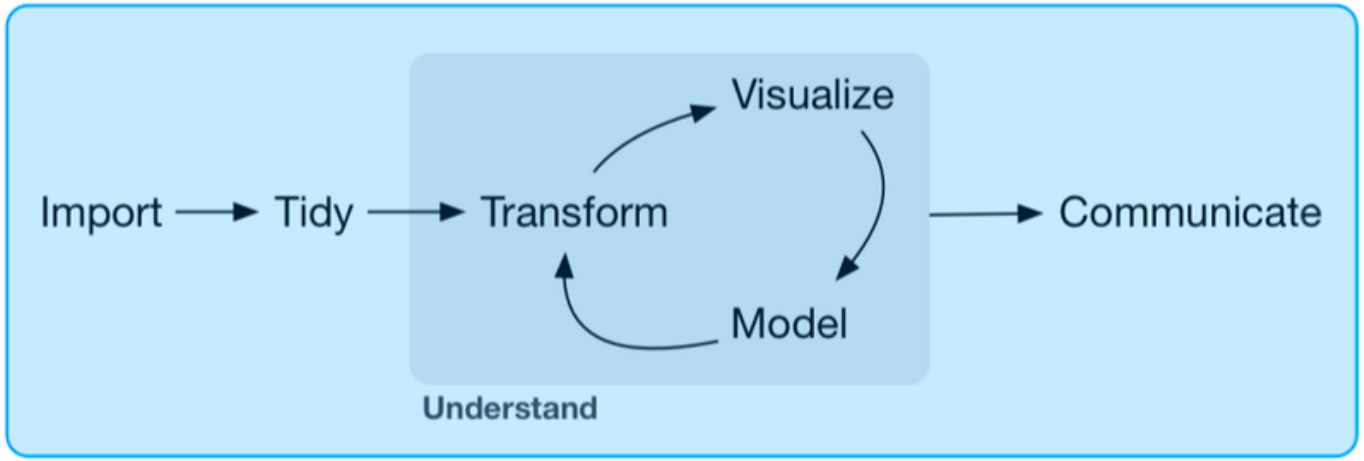
Tipos de Análises

- **Descritiva** o que aconteceu?
- **Preditiva** o que pode acontecer?
- **Prescritiva** o que devemos fazer?

Analytics

Programação

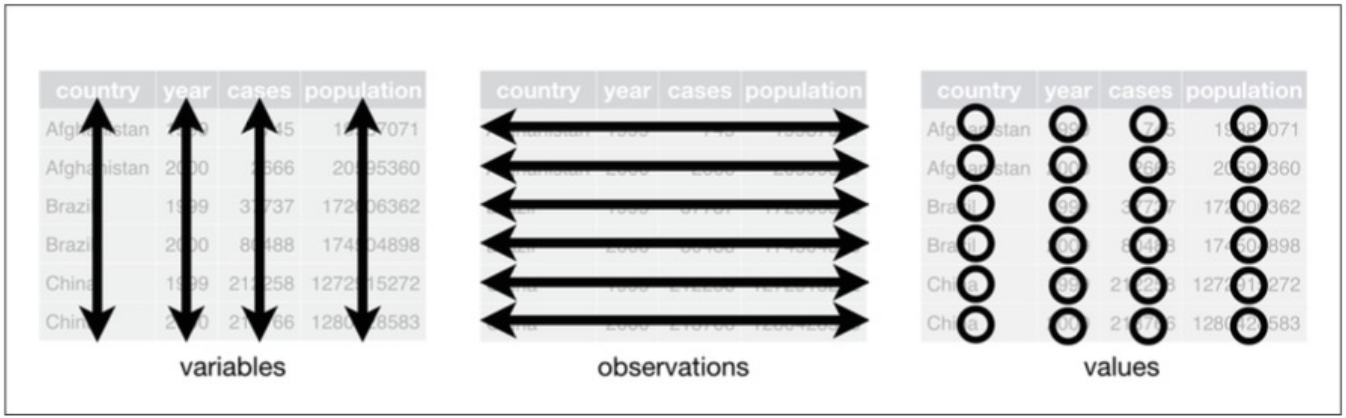
Tidy = Arrumar



Organizar o Banco de Dados é fundamental

Analytics

Banco de Dados



Variáveis nas colunas;

Observações nas linhas;

Intersecções = **Valores.**

Variável

Característica de dada amostra ou população, que **pode ser medida, contada ou categorizada**.
Dividem-se em 2 tipos de escalas de mensuração:

- **Qualitativas / Categóricas** não podem ser medidas, apenas categorizadas ou contadas. Não permitem cálculo de estatísticas descritivas.

- **Quantitativas / Métricas** dividem-se em **contínuas** (com casa decimal) e **discretas** (números inteiros). Podem ser mensuradas e permitem cálculos de estatísticas descritivas, como média, por exemplo.

Banco de Dados

Atenção aos detalhes:

- **Grafia das variáveis** (caracteres especiais, espaços, maiúsculas/minúsculas, etc)
- **Tipos de variáveis** numéricas (números inteiros ou fracionados), caracteres/strings (letras), lógicas (falso/verdadeiro)
- **Escalas de medidas** (aplicar funções matemáticas que permitem equiparar variáveis em escalas diferentes)
- **Diferentes fontes de dados** que “não se conversam”

Três passos para iniciar a análise

- **Qual o tipo de variável** (quantitativa ou qualitativa)
- **Qual o propósito da análise** (comparação, relação, previsão)
- **Qual o método estatístico apropriado**

Estatísticas Descritivas

Permitem extrair informação das variáveis

- **Tabelas de Frequência** quantidade de ocorrências (contagem ou percentual), aplicadas principalmente para variáveis categóricas
- **Medidas de Posição** média, mediana, moda, percentis, etc
- **Medidas de Dispersão** amplitude, variância, desvio padrão, etc

Técnicas Multivariadas Exploratórias

Estudam a relação entre variáveis para:

- **Redução ou simplificação** dos dados
- **Classificação/agrupamento** de variáveis categóricas
- Extração de **Fatores**
- Elaboração de **rankings de desempenho**
- **Diagnóstico**

Não há intenção de se fazer previsões/inferências

Modelos de Regressão

- Servem para **prever comportamentos** com base na associação entre **uma variável dependente** e **uma ou mais variáveis explicativas**
- Utiliza a **correlação entre as variáveis**

Podem ser utilizados para:

- Determinar **quais variáveis estão relacionadas** à variável dependente
- **Entender o relacionamento** entre as variáveis dependente e explicativas
- **Prever valores** desconhecidos da variável dependente

Machine Learning

- **Aprendizagem de Máquina**
- Automatizar tarefas com base na **descoberta sistemática de padrões** nos conjuntos de dados disponíveis
- **Modelos Supervisionados:** regressão, classificação, árvores de decisão, redes neurais, etc
- **Modelos Não-Supervisionados:** agrupamento (cluster), componentes principais (fatorial), redução de dimensionalidade, etc

Analytics

Ferramentas

Excel explorar e analisar dados

Tableau/Power BI visualizar dados em dashboards

Python/R construir modelos estatísticos

SQL interação entre bases de dados