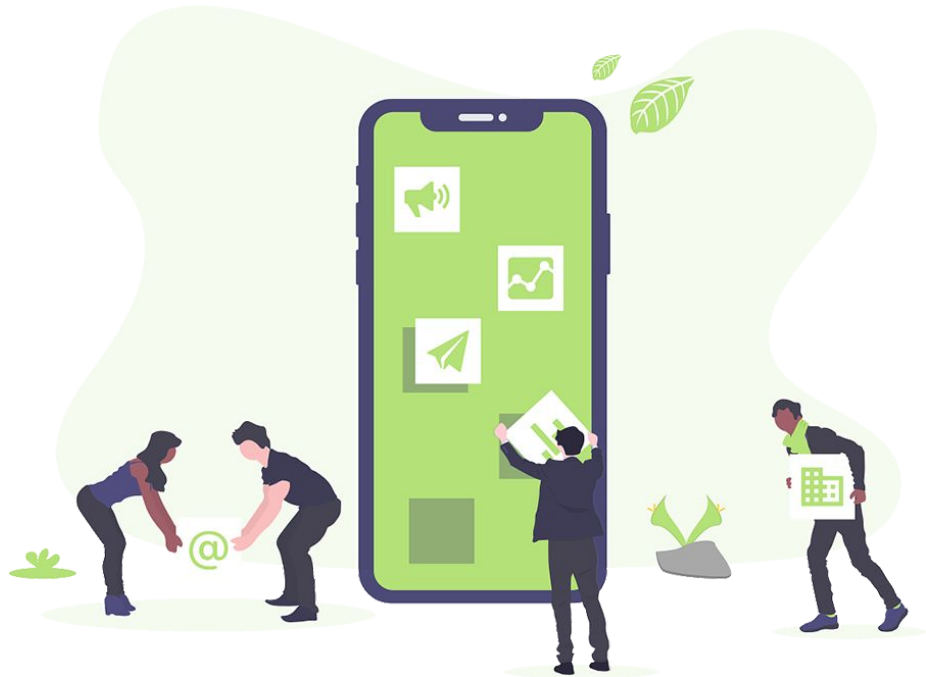
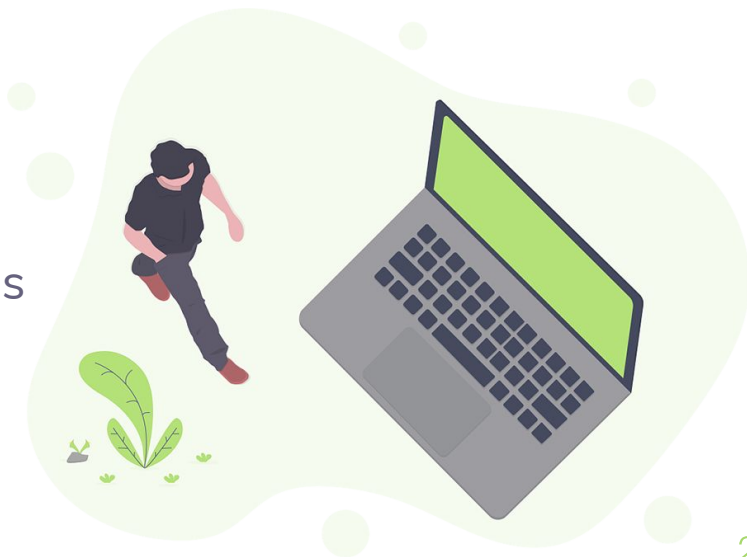


# Data Mining



# O que é data mining?

- Explorar grandes quantidades de dados
- Busca por anomalias;
- Busca de padrões consistentes e relacionamentos sistemáticos entre as variáveis;
- Tipicamente relacionado a negócios, mercados ou pesquisas científicas.



“ ... o processo não trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e compreensíveis.



# Objetivos

Transformar o dado em  
conhecimento utilizável

Explicativo - Confirmativo - Exploratório



# Passos do data mining

Limpeza - Seleção - Data mining - Conhecimento



# Hierarquia do processo

- Limpeza
- Seleção
- Data mining
- Avaliação/Visualização



# Vantagens

Aspectos Lógicos x Empresariais



## Aspectos Lógicos

- Análise de grandes bases de dados;
- Descoberta de informações inesperadas;
- Modelos precisos.

## Aspectos Empresariais

- Empresa inteirada sobre assuntos do seu domínio;
- Obtenção de análises precisas;
- Menor trabalho para organização dos dados.





# Desvantagens

Alto custo para implantação.

Novidade e complexidade dentro das empresas

Forte interação com analistas humanos para refinar e conduzir o processo.

# Aplicações Reais

Walmart

Bank of America







**Bank of America<sup>®</sup>**



# Projeto Final

Apresentação da base de dados, algumas informações e definições



# BD: European Soccer Database

	Relacionamentos	Atributos	Tuplas
Country	2	2	11
League	2	2	11
Match	4	89	25979
Player	2	7	11060
Player Attributes	1	42	183978
Team	2	5	299
Team Attributes	1	23	1458

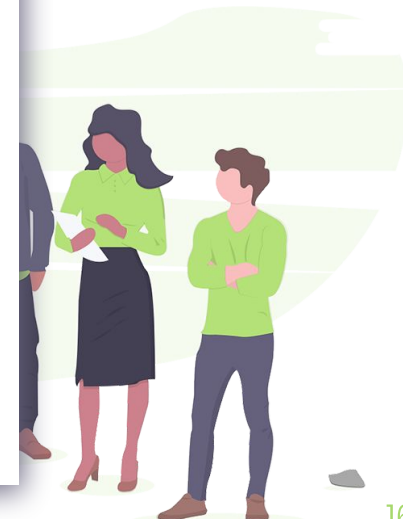
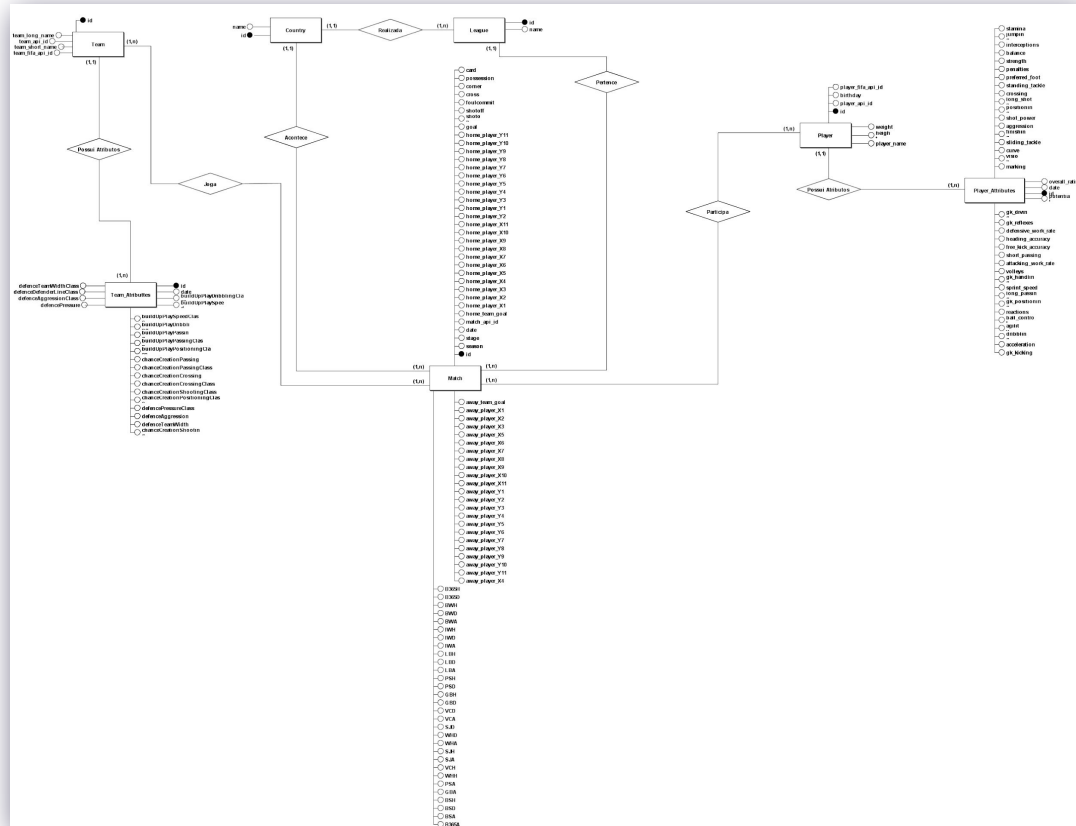


# 222.796

Número total de registros para  
serem explorados



# Diagrama Entidade-Relacionamento





# Ferramenta de Data Mining

**Weka**



# Obrigado!

Alguma questão?

