



Data Science Abordagens e Práticas

PROF. CLÁUDIO PINHEIRO



DADO É O NOVO RECURSO NATURAL



Um **mar de dados**
é ao mesmo tempo
o combustível e
o resultado da
transformação digital.

2.5B

gigabytes de novos dados
são gerados **todos os dias**

4/5 deles são não-estruturados

80%
dados não
estruturados

EM 2020

1,7 MB

de nova informação será criada por segundo para cada ser humano do planeta

Dobramos a cada dois anos

2010

2015

2018

2020

44 Zettabytes

Gap

Sensores

Imagens

Textos

Atuais

Traditional

Atualmente os clientes esperam ao menos uma experiência personalizada baseada em valor...

- **Antecipar e atender minhas necessidades financeiras**
- **Tornar as interações comigo convenientes e fáceis**
- **Estar onde eu precisar, em tempo real, proativamente**
- **Lembrar de todas as minhas interações**

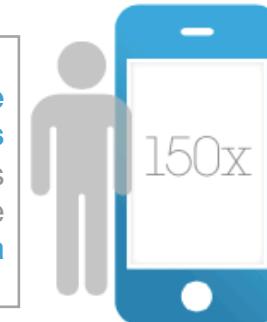
...e estão dispostos a trocar informações com o banco

80%

dos indivíduos estão dispostos a **compartilhar** suas **informações pessoais** em troca de ofertas personalizadas

Usuários de smartphones

checam seus aparelhos aproximadamente **150 vezes ao dia**



84%

dos **millennials** e

70%

dos **baby boomers** dizem que os conteúdos gerados por usuários e nas mídias sociais influenciam na compra



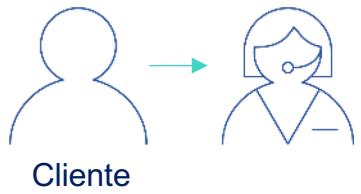
5min...



é o **tempo de resposta** que as pessoas esperam quando entram em contato com uma marca nas mídias sociais

O engajamento é cada vez mais digital

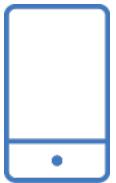
ANTES



AGORA



Atendimento
ao cliente



Apps Móveis



Canais de
Mensagem

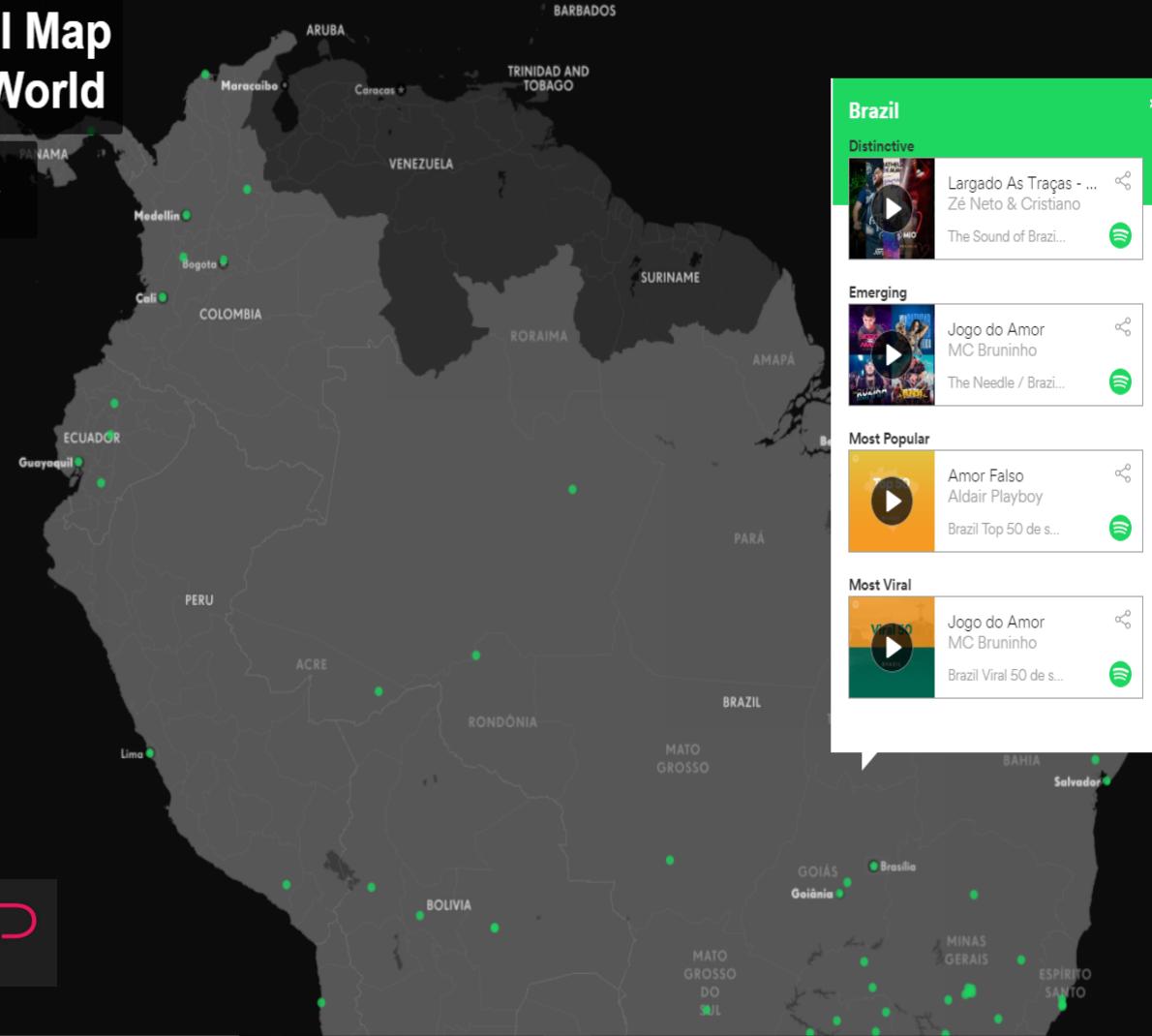
FIAP



To learn how this map works,
see insights.spotify.com.

Musical Map of the World

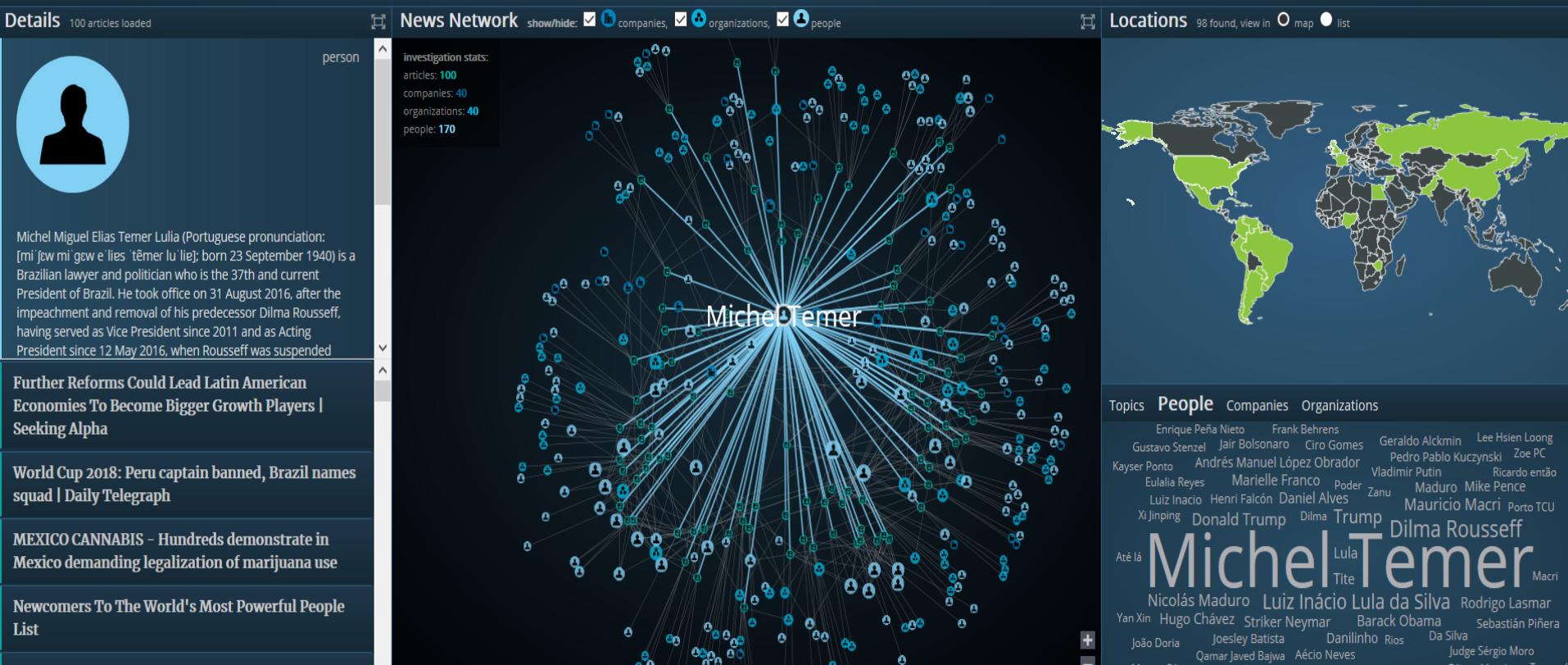
Zoom with the +/- symbols to the top-left, drag to move.



FIAP



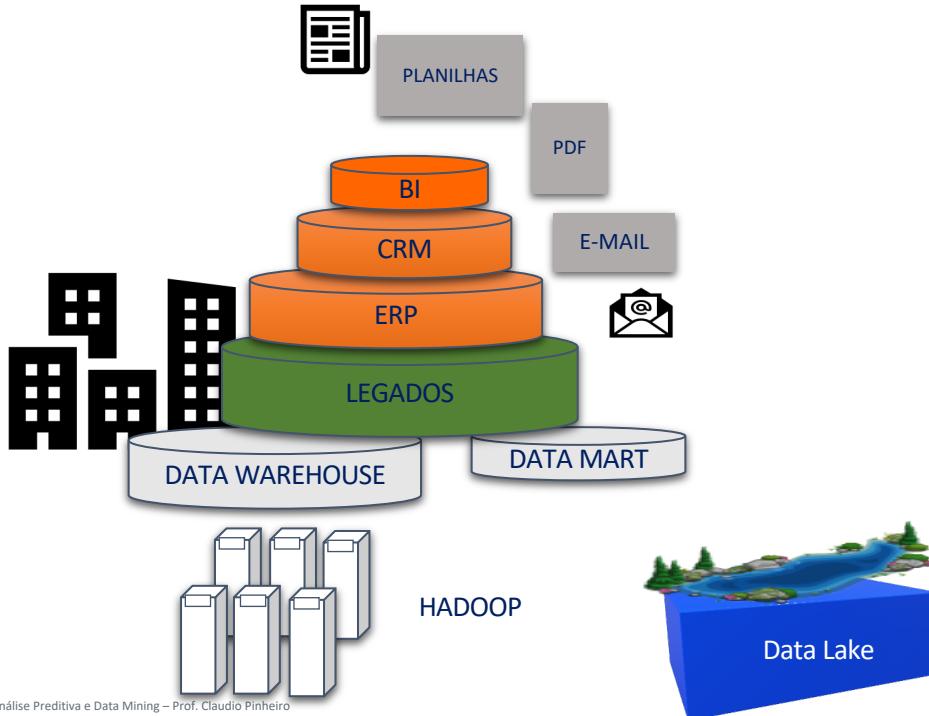
CARTO attribution





A cada minuto:

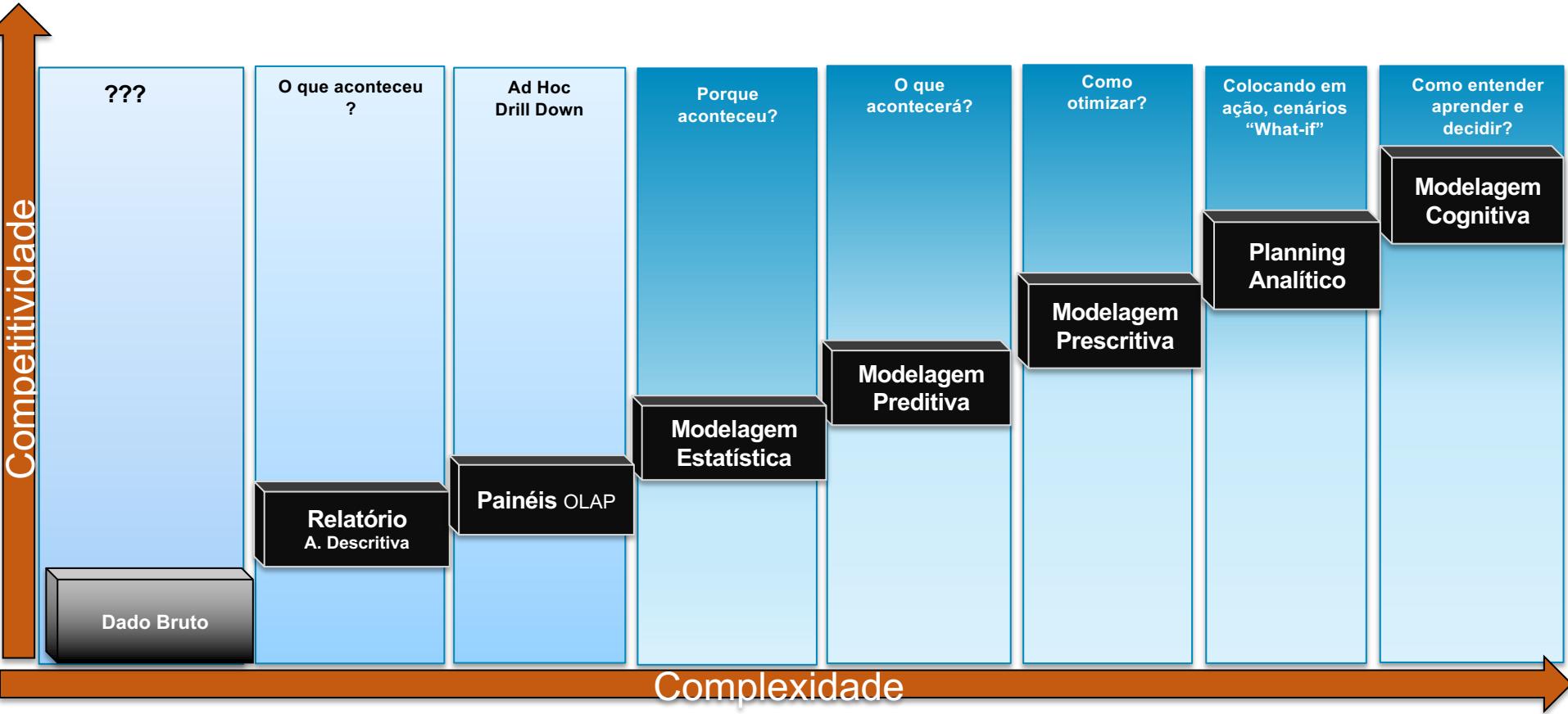
- 1.8 milhões de likes no facebook
- 204 milhões de e-mails
- 200 mil uploads de fotos
- 2,78 milhões de videos assistidos youtube
- 72 horas de videos - uploaded



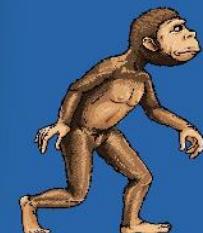
Analista de negócios

- *Conhecer o cliente*
- *Aumentar as vendas*
- *Reducir despesas*
- *Entender o que aconteceu*
- *Melhorar as margens*
- *Prever o que acontecerá*
- *Tomar as melhores decisões*
- *Melhorar a participação de mercado*
- *Bater metas de produtividade*
- *Melhorar a experiência do cliente*
- *Diminuir os riscos*
- *Ser mais eficiente*
- *Inovar*

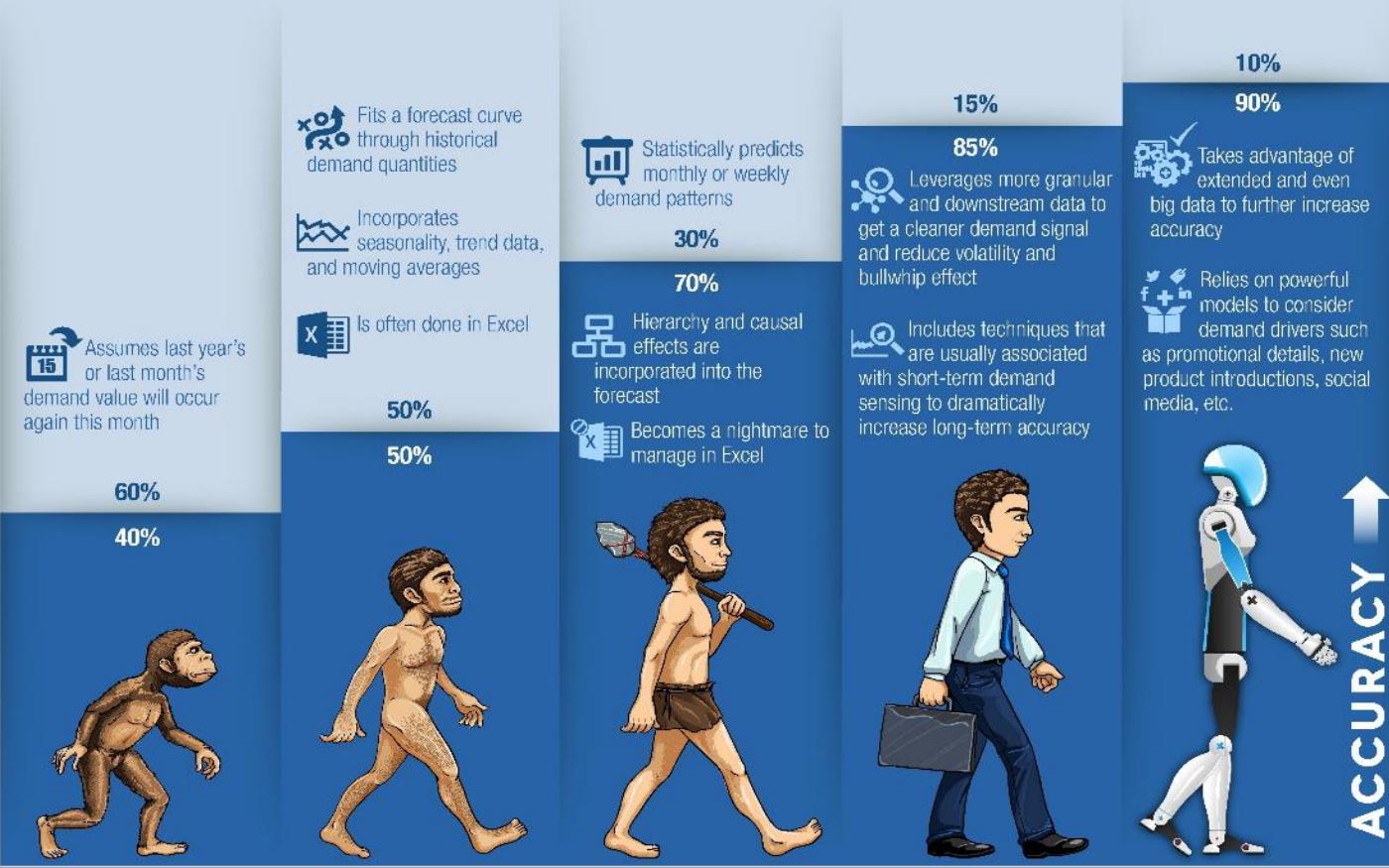
Evolução Analítica



↓ ERROR



ACCURACY ↑



Quanto iremos vender mensalmente deste produto na região sul em 2019?



Quais produtos devemos recomendar a um cliente que está comprando um determinado produto?



Sumarização

Previsão

Segmentação

Tarefas Analíticas

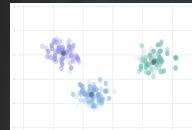
Recomendação

Classificação

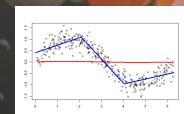
Estimação

Quantos produtos vendemos?
Qual o tempo médio de atendimento?
Quanto dura o ciclo de venda?

Quantos grupos diferentes de clientes temos?



Qual é a classe deste novo cliente?



Quanto devemos gastar em mídia para vendermos 20% a mais?

Tarefas Analíticas

- **Tarefas** (o que fazer)

- Sumarização e visualização
- Associação
- Classificação
- Segmentação
- Estimação
- Predição

- **Técnicas** (com o quê e como fazer)

- Estatística Básica
- Análise de Regressão
- Análise de cluster
- Análise Discriminante
- Séries temporais
- Árvores de Decisão
- Redes Neurais
- Mineração de Textos...

Tarefas Analíticas

- Sumarização e visualização

- Consiste em descrever o conjunto de dados
- Objetiva encontrar descrições compactas dos dados, facilitar a interpretação dos mesmos,
- Permite visualizar os aspectos de conjunto de dados extensos
- Mesmo com suas limitações permite alguns insights

Tarefas Analíticas

- Sumarização e visualização
- **Regras de Associação**

Consiste em determinar relações entre campos e / ou entidades de um banco de dados.

Por exemplo, pode-se identificar produtos em um supermercado que usualmente são vendidos juntos.

Tarefas Analíticas

- Sumarização e visualização

- Regras de Associação

Classificação

Consiste em examinar as características de uma entidade para associá-la a uma ou mais classes existentes (encontrar um conjunto de modelos ou funções que descrevem e distinguem as classes de dados com o propósito de predizer a classe da nova observação);

As classes são previamente definidas e bem caracterizadas.

Por exemplo, uma pessoa candidata a um empréstimo bancário pode ser classificada como baixo, médio ou alto risco de ser inadimplente.

A construção do modelo é feita com os dados de treinamento, isto é, objetos cujas classes são conhecidas.

Tarefas Analíticas

- Sumarização e visualização

- Regras de Associação

- Classificação

- Segmentação

Consiste em separar as entidades de um conjunto de dados em subgrupos, de forma que as entidades dentro de cada subgrupo sejam mais semelhantes entre si em relação às entidades dos outros subgrupos.

Os critérios de semelhança são definidos por meio das variáveis aleatórias que descrevem as entidades.

Por exemplo, pode-se segmentar uma cartela de clientes por diferentes perfis de consumo.

A idéia é dividir um conjunto em segmentos menores os quais formam grupos mais inter-relacionados

Tarefas Analíticas

- Sumarização e visualização
- Regras de Associação
- Classificação
- Segmentação
- **Estimação**

Consiste em estimar valores presentes de uma variável a partir da observação de uma ou muitas outras variáveis que têm relação com a primeira.

Por exemplo, a renda de uma pessoa pode ser estimada a partir da observação de variáveis que descrevem o seus hábitos de consumo.

Tarefas Analíticas

- Sumarização e visualização
- Regras de Associação
- Classificação
- Segmentação
- Estimação
- Predição**

Consiste em estimar valores futuros de uma (ou mais) variável a partir da observação da própria variável no passado e no presente.

Por exemplo, o volume de vendas de um produto para os próximos meses pode ser estimado a partir das vendas observadas nos últimos anos.

**Precisamos de uma
plataforma de dados
para habilitar nossas
estratégias em
escala com
tranquilidade?**

