INTRODUÇÃO AO ORANGE

João Ricardo Arcoverde
LIVIA
UERJ

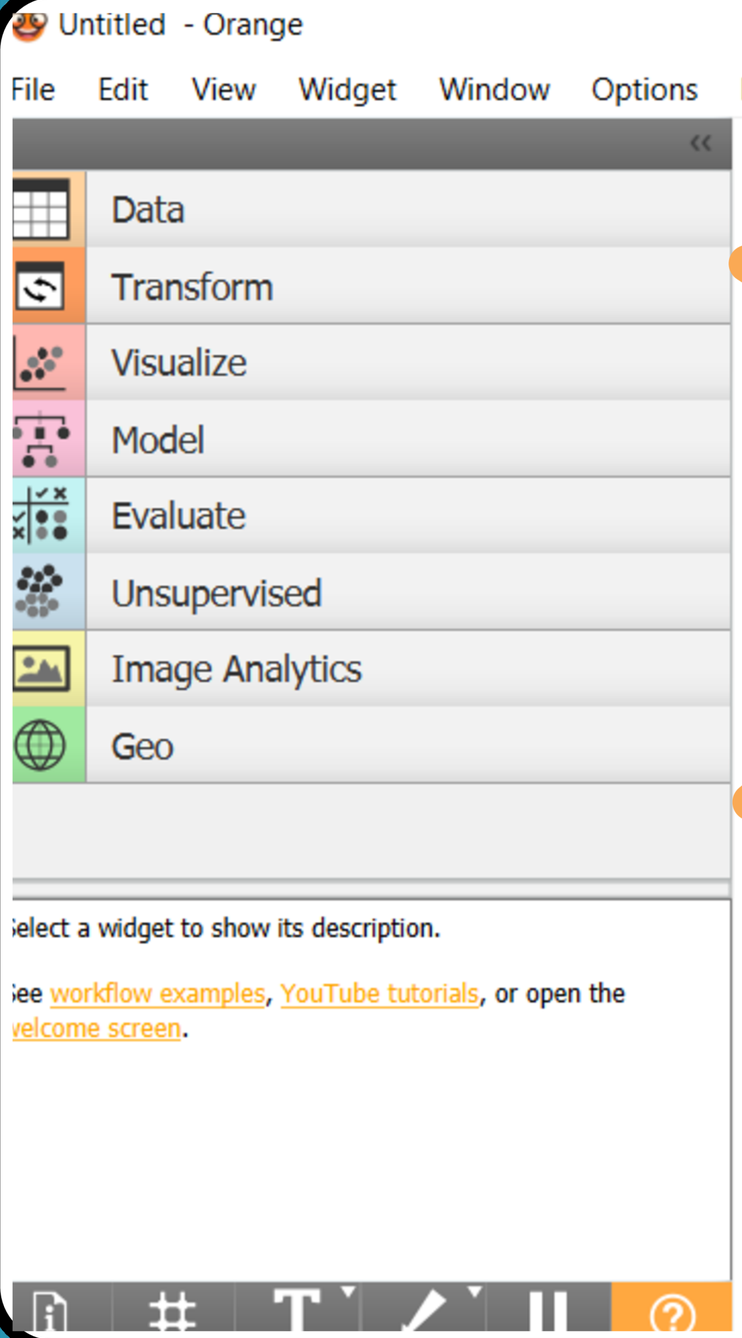
O QUE É O ORANGE?

É uma plataforma de código aberto, principalmente usada para visualização e análise de dados.

- Fácil uso
- Alta flexibilidade
- Análise Exploratória de Dados
- Modelagem Preditiva/Machine Learning
- Visualização de Dados Interativa

University of Ljubljana

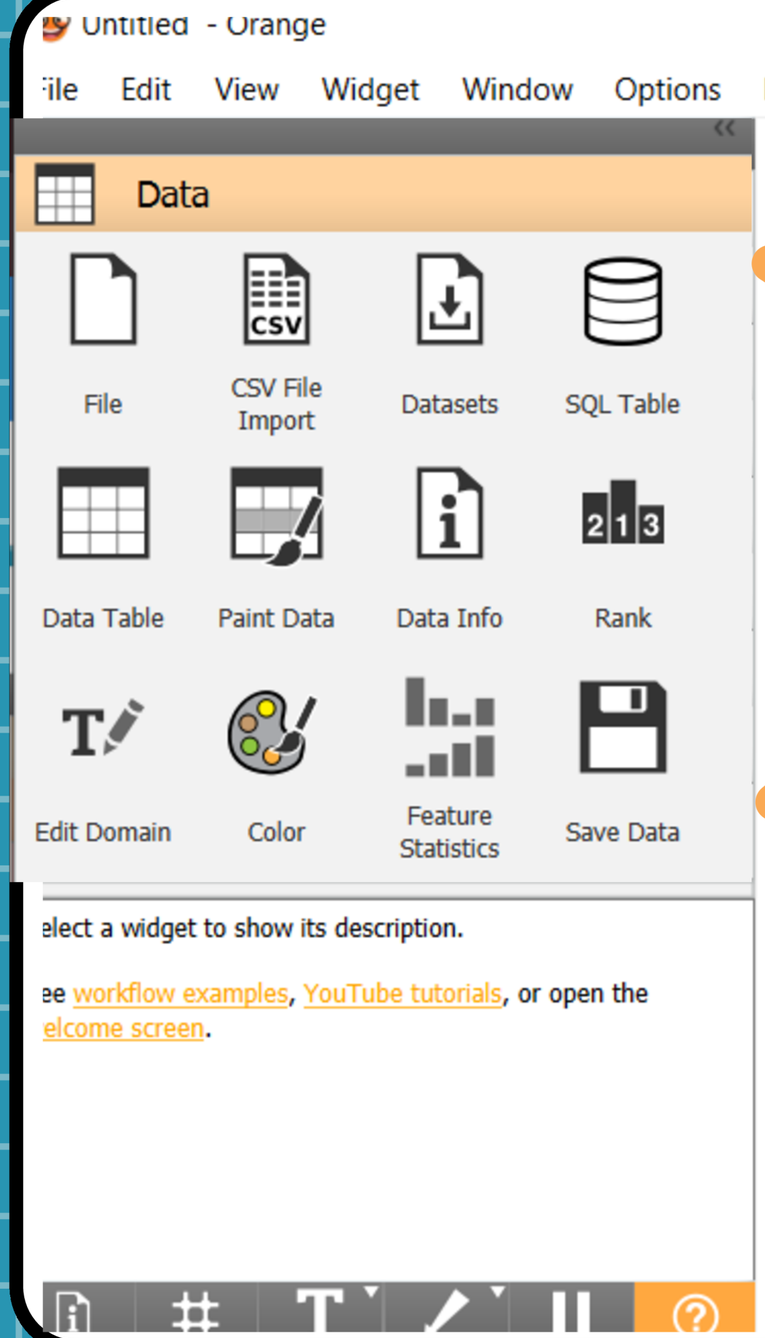
INTERFACE



Painel

Área de Trabalho

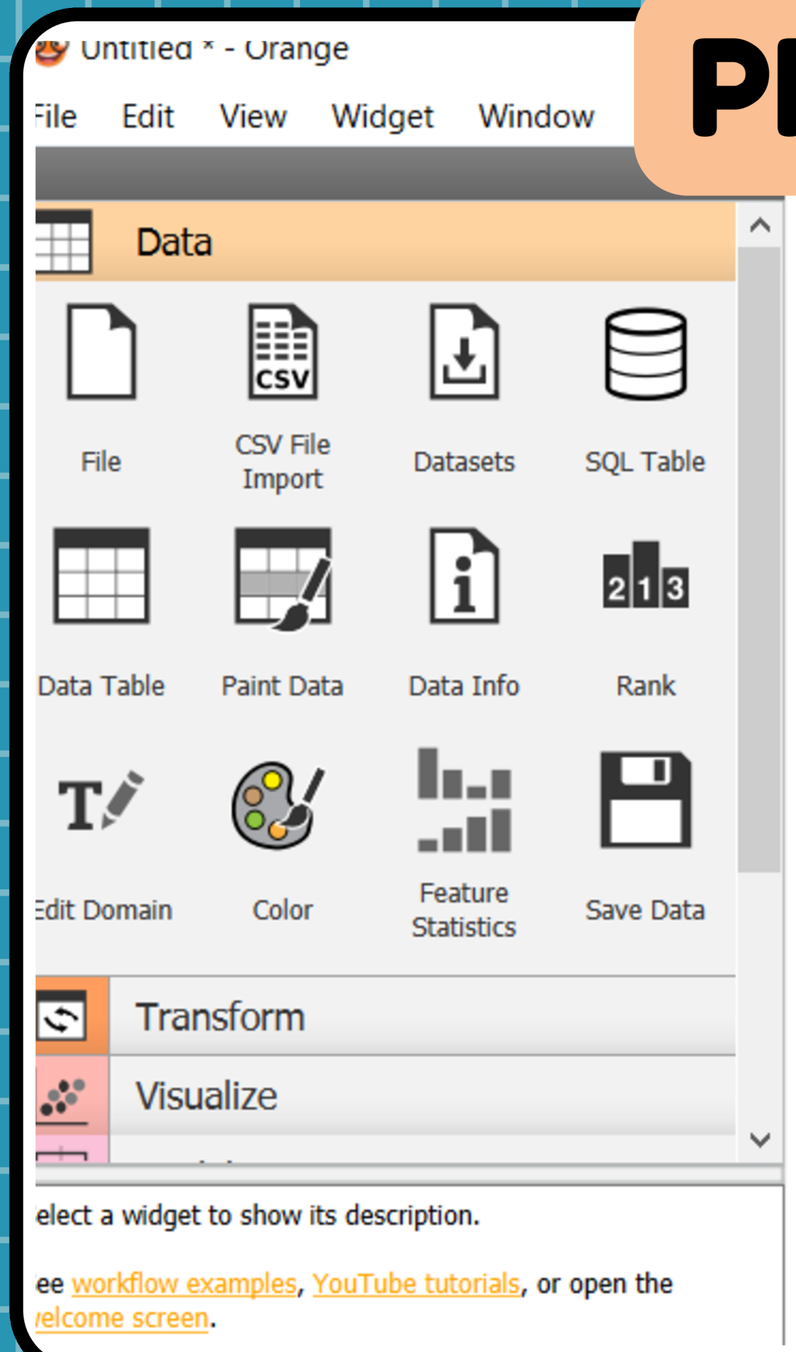
INTERFACE



Ferramentas

Área de Trabalho

PRINCIPAIS FERRAMENTAS



Arquivo

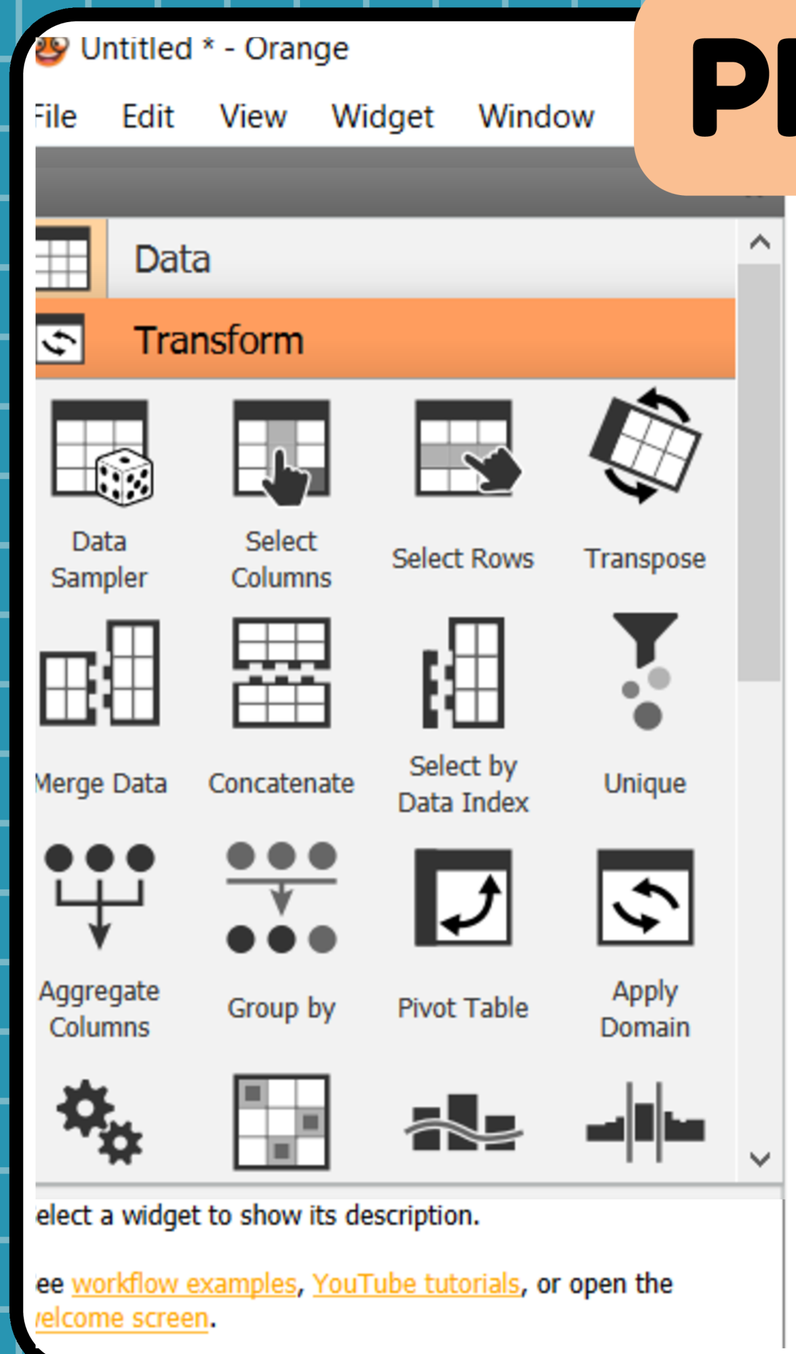
Tabela

Informações

Editar domínio

Estatística dos Atributos

PRINCIPAIS FERRAMENTAS

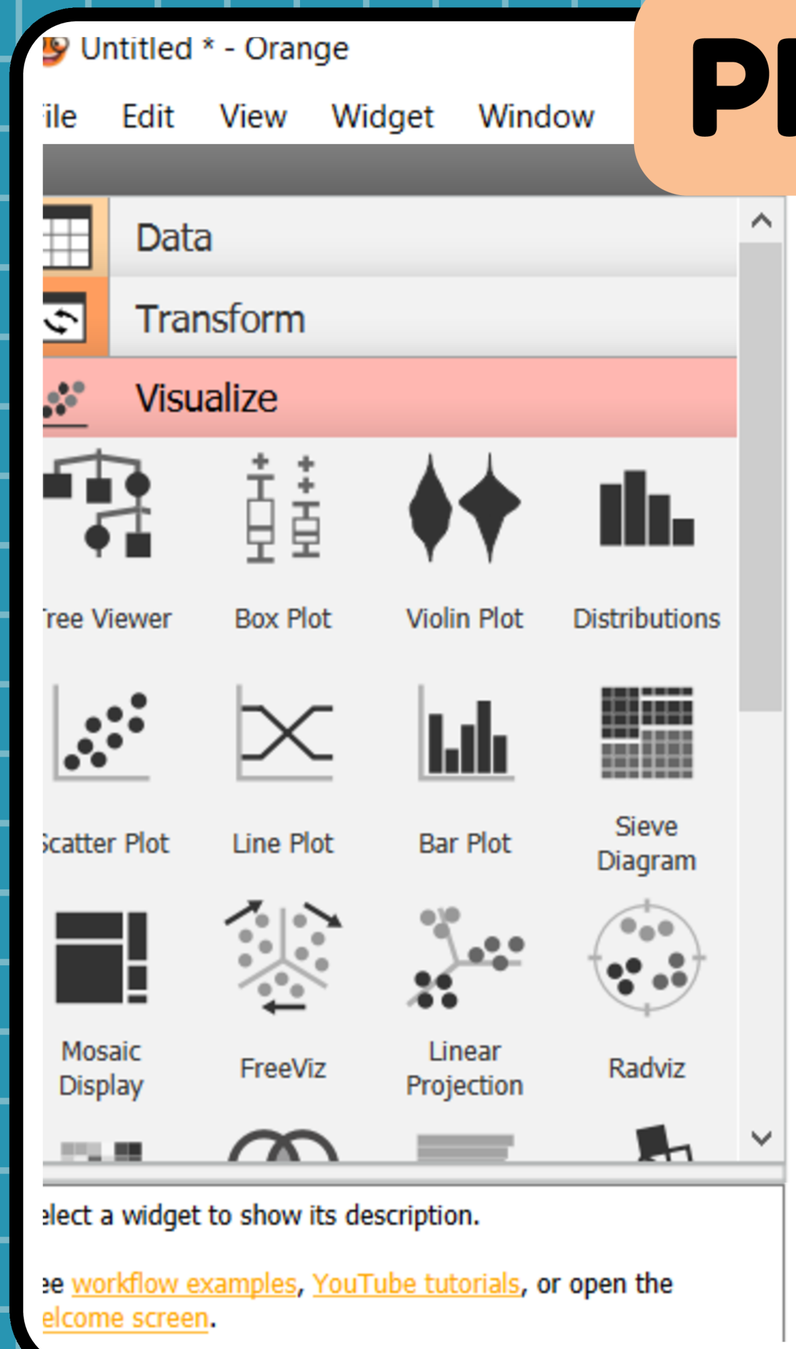


Amostragem dos Dados

Selecionar colunas

Padronização

PRINCIPAIS FERRAMENTAS

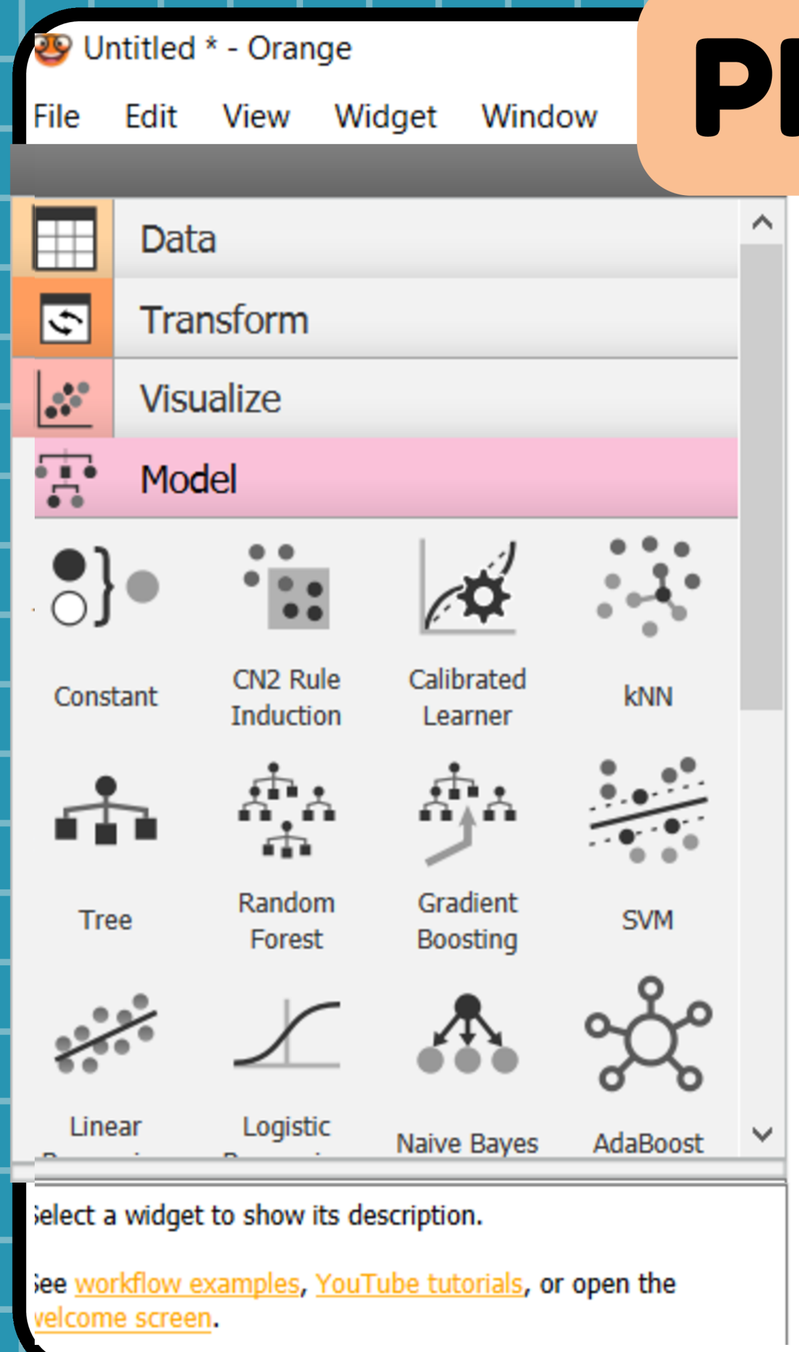


Distribuições

Gráfico de Pontos

Gráfico de Barras

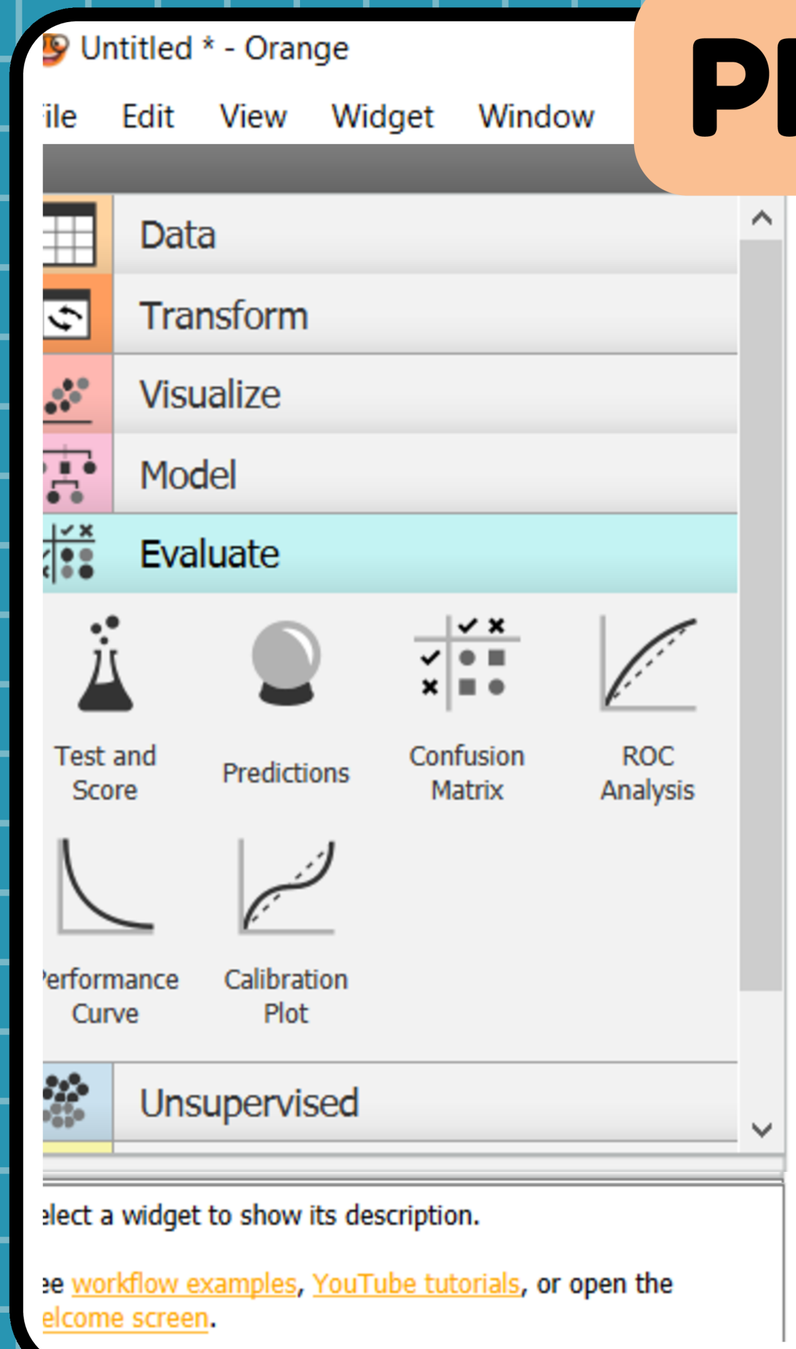
PRINCIPAIS FERRAMENTAS



Regressão Linear

Regressão Logística

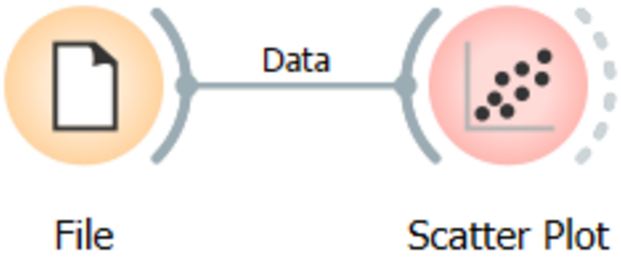
PRINCIPAIS FERRAMENTAS



Previsões

Matriz de Confusão

ORANGE VS. PYTHON

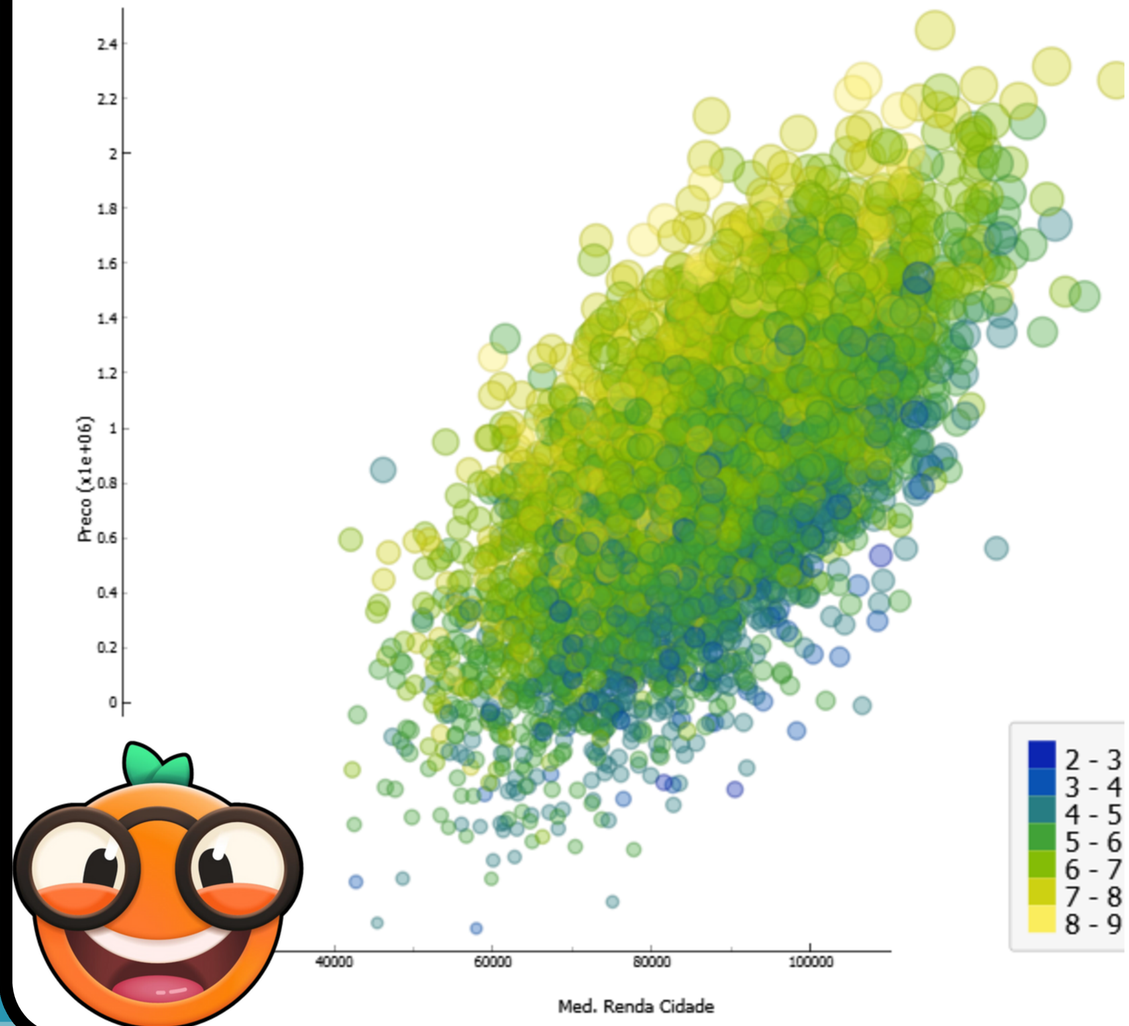


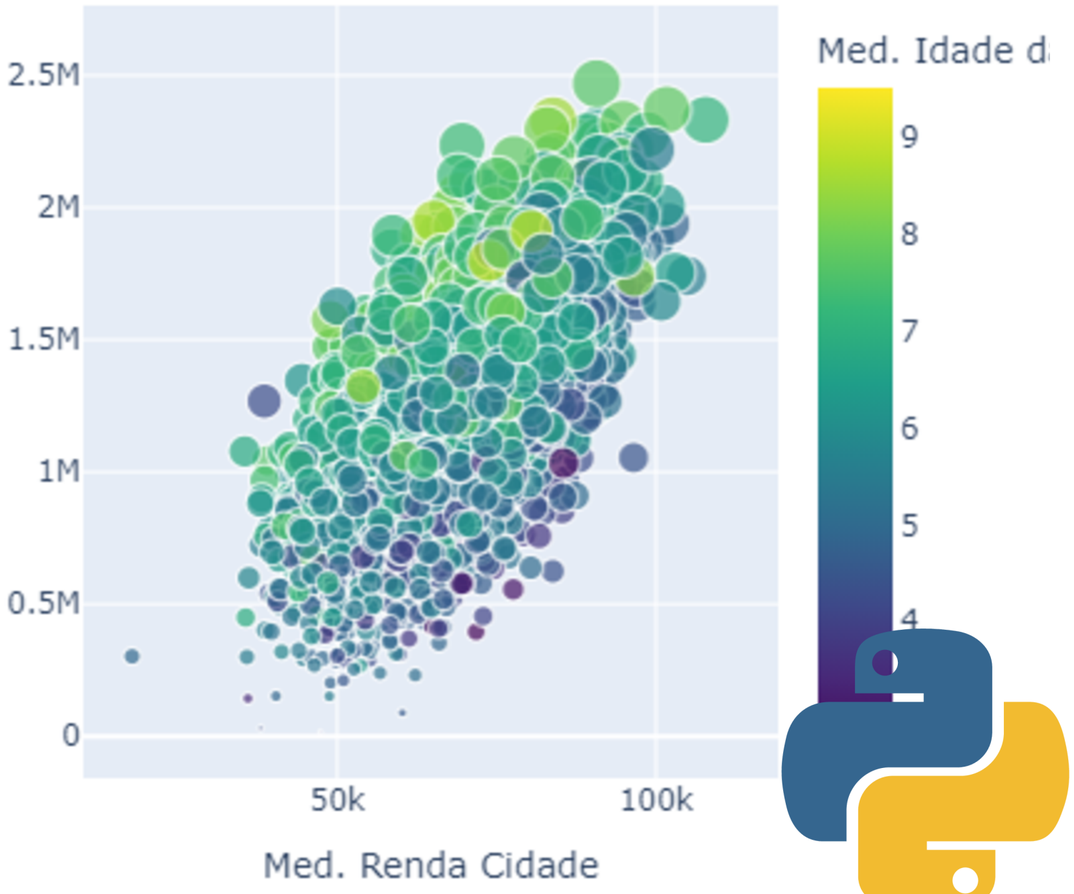

```python
import pandas as pd
import plotly.express as px

df = pd.read_csv('USA_housing.csv')

fig = px.scatter(df,
                 x="Med. Renda Cidade",
                 y = "Preco",
                 color = "Med. Idade das Casas da Cidade"
                 size = "Preco",
                 color_continuous_scale=px.colors.sequent
                )
fig.show()
```

ORANGE VS. PYTHON





LIMITAÇÕES DO ORANGE

Restrito a Ferramentas	Você cria suas funções
Customização Limitada	Bibliotecas Próprias
Dificuldade com Big Data	Bibliotecas para Big Data
Comunidade Menor	Comunidade Maior

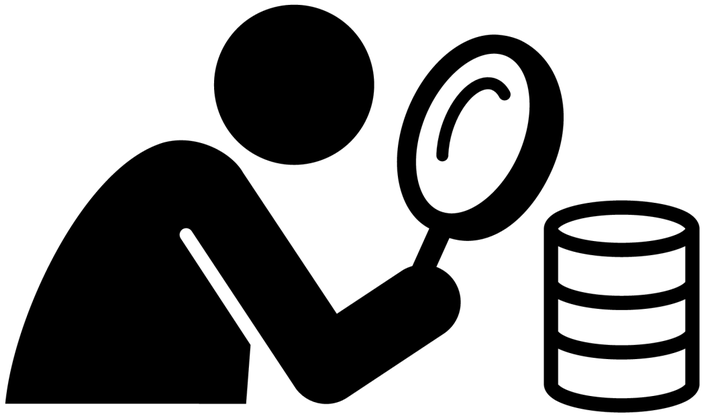
CONCEITOS BÁSICOS

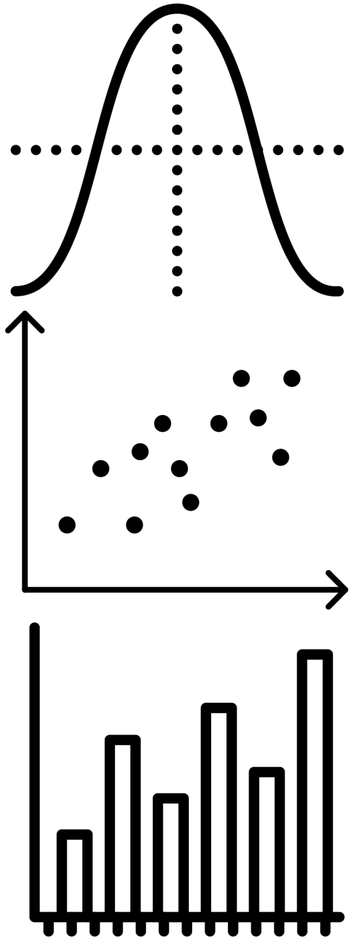
- Colunas
- Linhas
- Variáveis Numéricas
- Variáveis Categóricas

	nome	cidade	idade	nota
10	Pedro	São Paulo	34	83.0
11	Maria	São Paulo	23	59.0
12	Janaina	Rio de Janeiro	32	86.0
13	Wong	Brasília	43	89.0
14	Roberto	Salvador	38	98.0
15	Marco	Curitiba	31	61.0
16	Paula	Belo Horizonte	34	44.0

ANÁLISE EXPLORATÓRIA DE DADOS (EDA)

- É um conjunto de técnicas gráficas e estatísticas, com o objetivo de examinar e resumir as características principais de um conjunto de dados, por meio da identificações de padrões e anomalias.

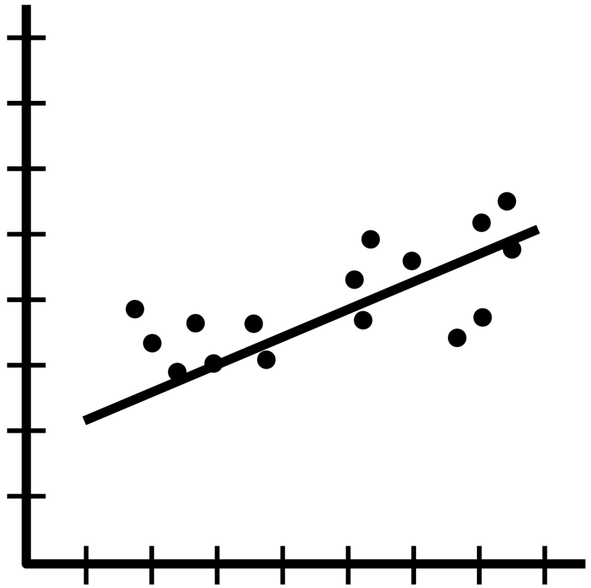




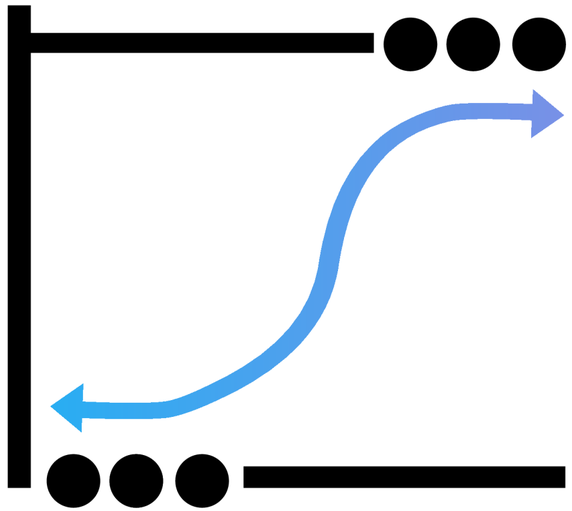
MODELAGEM

- **Tarefa responsável por representar dados através de funções matemáticas, buscando entender as principais influências sobre uma variável de interesse e prever comportamentos futuros.**

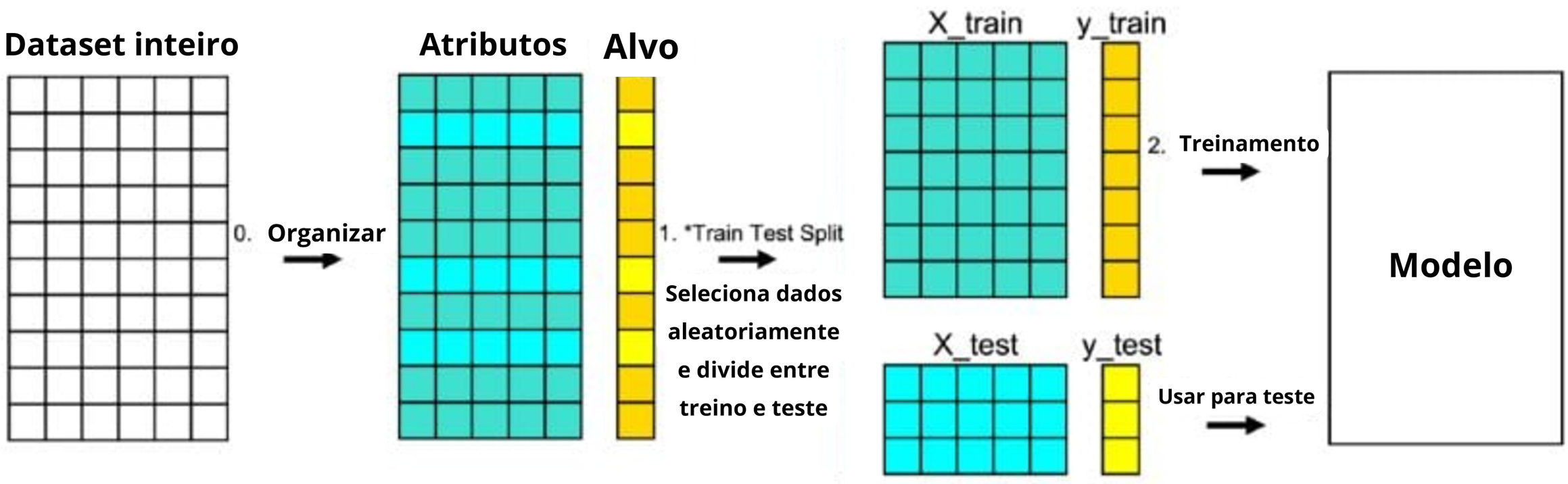
Regressão Linear



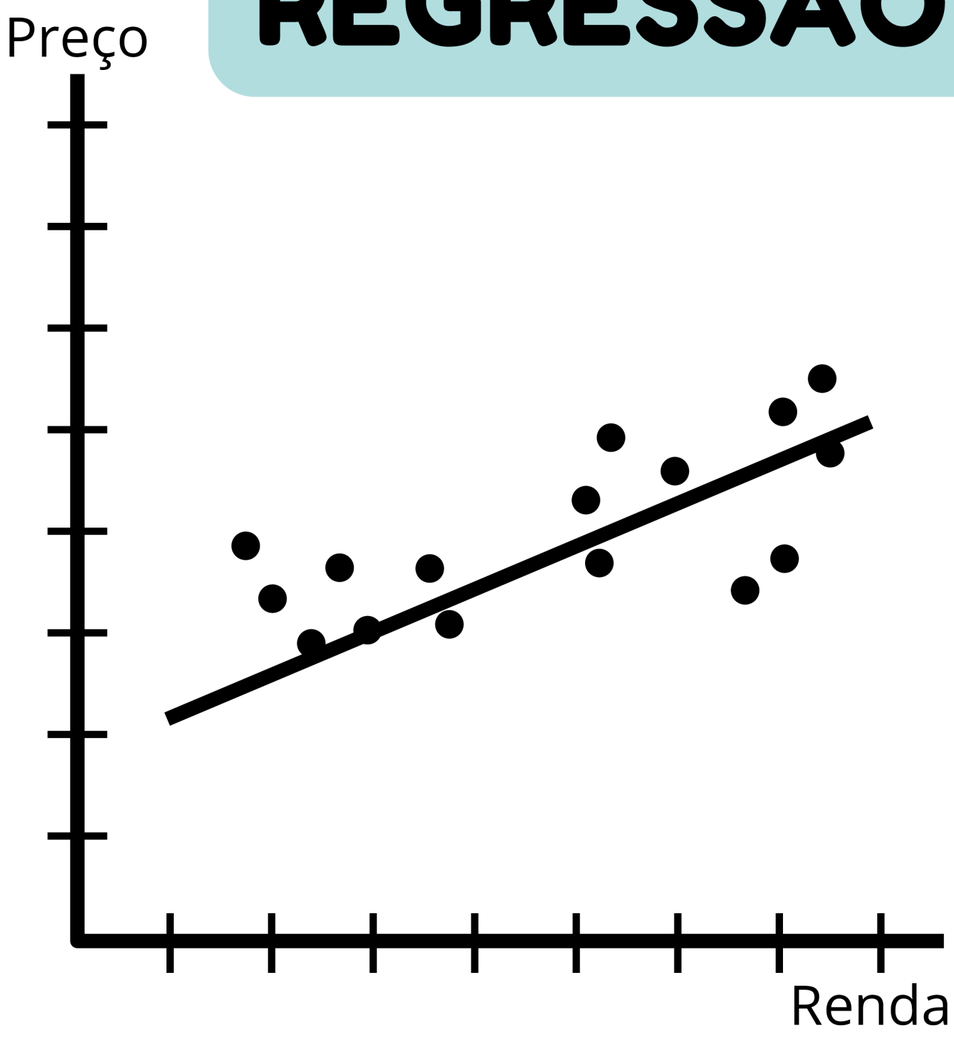
Regressão Logística



TREINAMENTO E VALIDAÇÃO



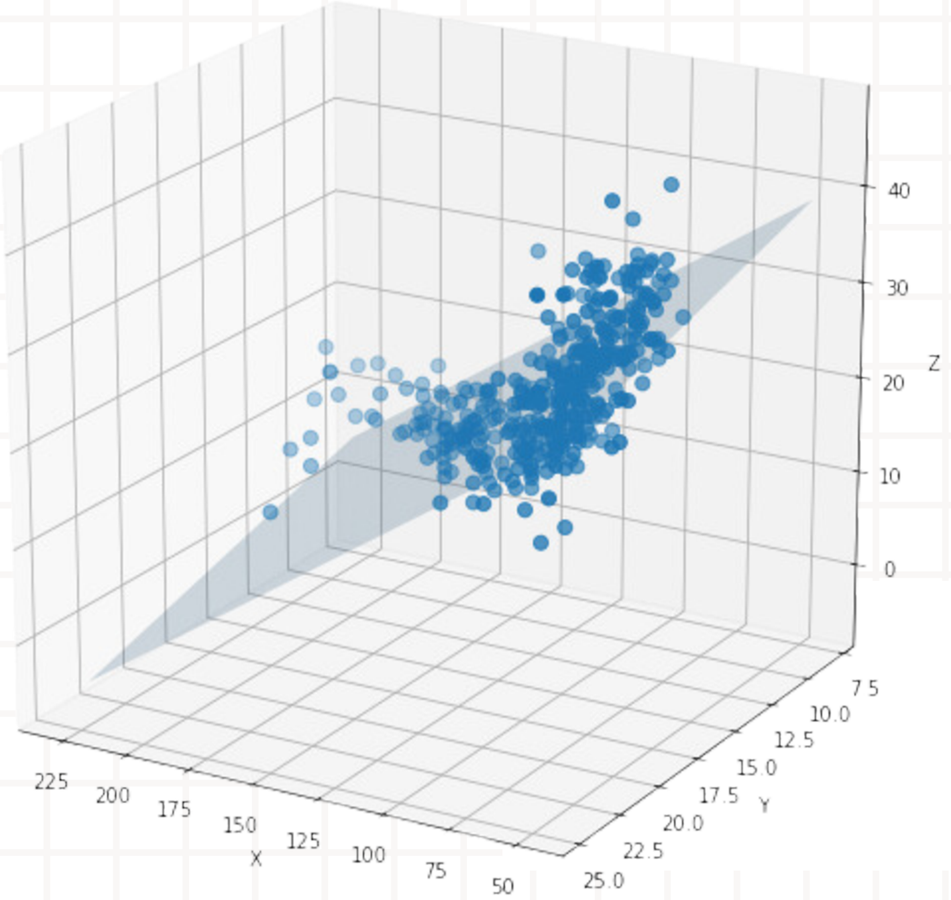
REGRESSÃO LINEAR SIMPLES



$$Y_i = \beta_0 + \beta_1 X_i$$

- **Previsão de Valor Numérico**
- **Ajuste com Mínimos Quadrados**
- **Variável Dependente (Y)**
- **Variável Independente (X)**
- **Interseção (B0)**
- **Inclinação (B1)**
- **MSE, RMSE, MAE e R^2**
- **Exemplo: Preço e Renda**

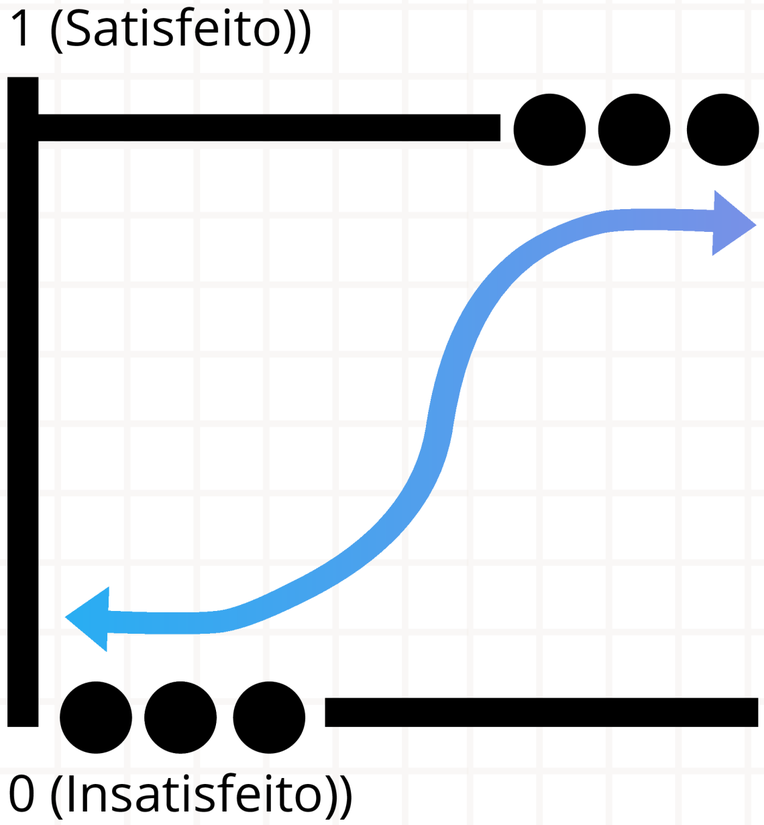
REGRESSÃO LINEAR MÚLTIPLA



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots$$

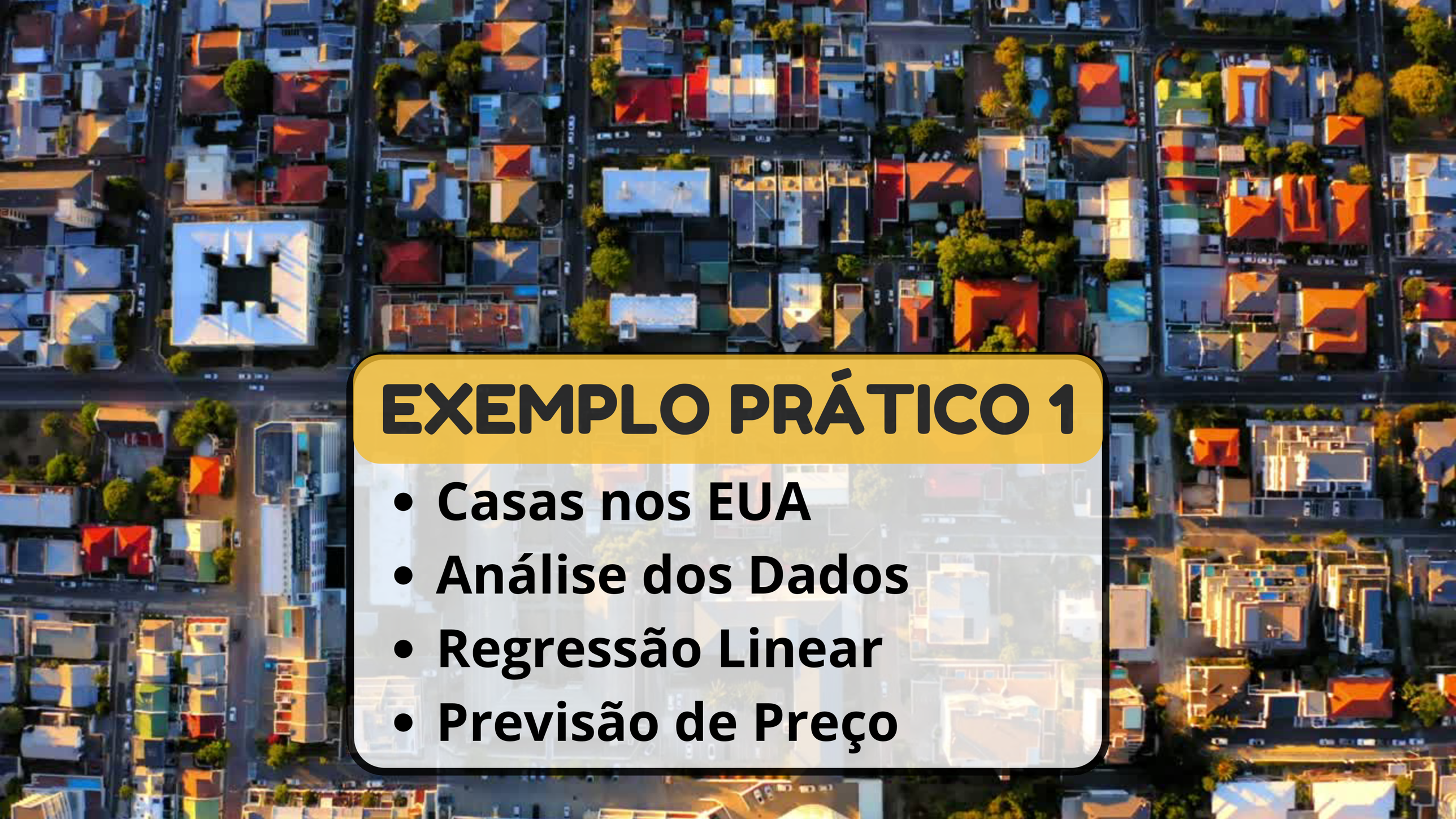
- **Previsão de Valor Numérico**
- **Ajuste com Mínimos Quadrados**
- **Variável Dependente (Y)**
- **Variáveis Independentes (Xi)**
- **Interseção (B0)**
- **Contribuição (B1)**
- **MSE, RMSE, MAE e R^2**
- **Exemplo: Preço, Renda e População**

REGRESSÃO LOGÍSTICA



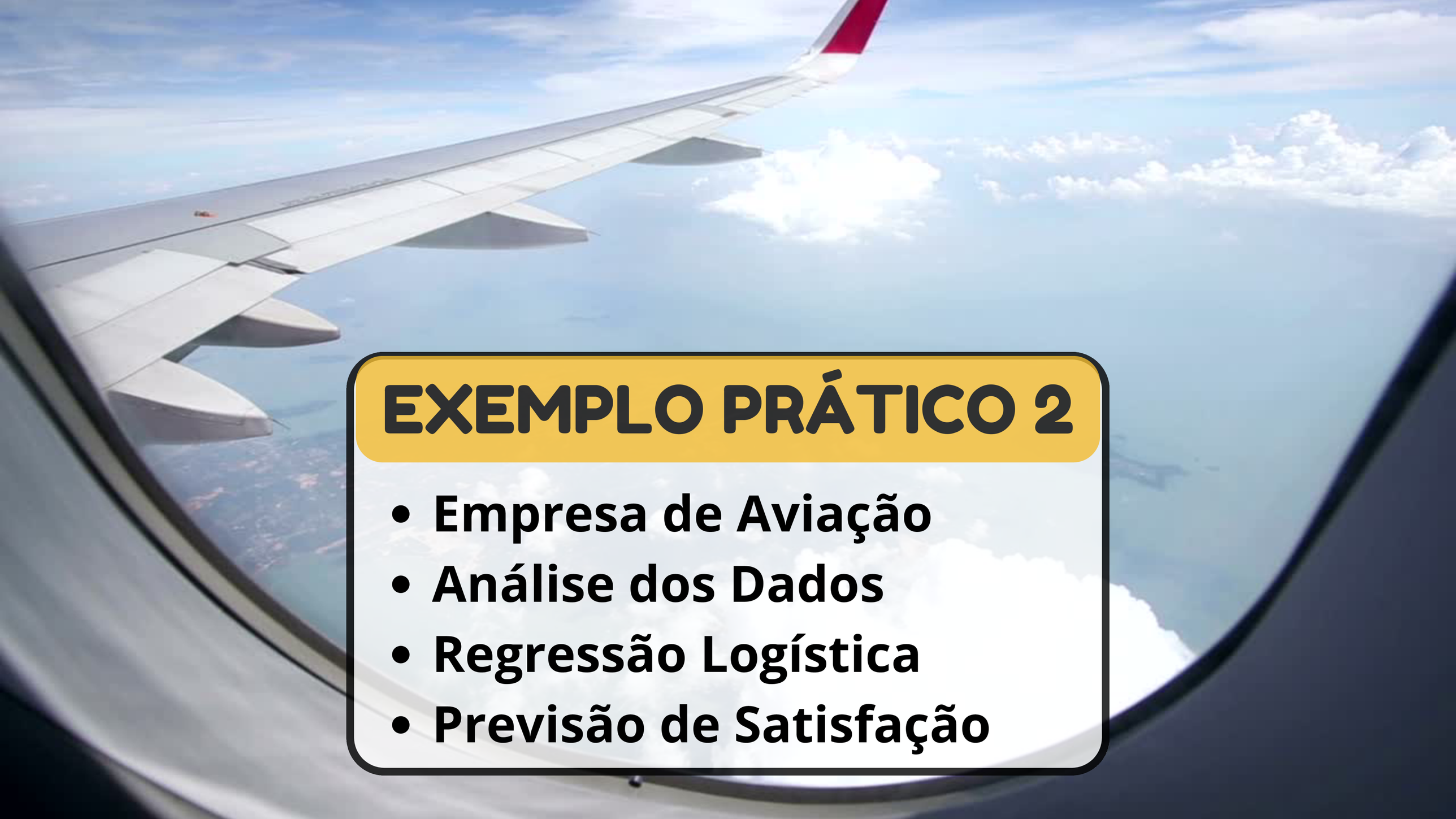
$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

- **Problemas de Classificação**
- **Função Sigmoide**
- **Probabilidade (p)**
- **Interseção (B0)**
- **Inclinação (B1)**
- **AUC, CA, Prec e MCC**



EXEMPLO PRÁTICO 1

- Casas nos EUA
- Análise dos Dados
- Regressão Linear
- Previsão de Preço



EXEMPLO PRÁTICO 2

- Empresa de Aviação
- Análise dos Dados
- Regressão Logística
- Previsão de Satisfação

ALGUMA DÚVIDA?

Lembre de estudar e aplicar esses ensinamentos!