

# O que é predição?

Predição é usar **dados históricos** para **estimar valores futuros** ou **classificar** algo.  
Existem dois grandes tipos:

## 1) Regressão

Quando você quer prever um *número contínuo*

Ex.:

- Prever vendas do próximo mês
- Prever preço de uma casa

## 2) Classificação

Quando você quer prever *categorias*

Ex.:

- Detectar se o cliente vai comprar (sim/não)
- Classificar se um aluno vai passar ou reprovar

# Exemplo completo: Prever vendas (Regressão)

## 1. Criar um dataset de exemplo

```
import pandas as pd
```

```
df = pd.DataFrame({  
    "Mes": [1,2,3,4,5,6,7,8,9,10,11,12],  
    "Investimento":  
[500,700,200,1000,400,900,850,100,1200,500,450,1100],  
    "Vendas": [50,65,30,100,40,95,85,20,120,55,48,130]  
})  
  
df
```

## 2. Separar variáveis explicativas (X) e alvo (y)

```
X = df[["Investimento"]] # variável usada para prever
y = df["Vendas"]         # variável alvo
```

## 3. Dividir em treino e teste

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

## 4. Treinar um modelo de Regressão Linear

```
from sklearn.linear_model import LinearRegression

modelo = LinearRegression()
modelo.fit(X_train, y_train)
```

## 5. Fazer previsões

```
predicoes = modelo.predict(X_test)
predicoes
```

## 6. Comparar valores reais x previstos

```
resultado = pd.DataFrame({
    "Real": y_test,
    "Previsto": predicoes
})

resultado
```

## 7. Visualizar o modelo

```
import matplotlib.pyplot as plt

plt.scatter(df["Investimento"], df["Vendas"])
plt.plot(df["Investimento"], modelo.predict(df[["Investimento"]]))
plt.xlabel("Investimento")
plt.ylabel("Vendas")
plt.title("Regressão Linear para Predição de Vendas")
plt.show()
```

## Interpretação

- A **linha azul** representa a melhor relação entre investimento → vendas.
- Conforme o **Investimento aumenta**, a **predição de vendas** também sobe.
- Se você colocar um valor novo, como:

```
modelo.predict([[1500]])
```

Você obtém a predição das vendas com investimento de R\$1500.

## Exemplo rápido de Classificação

Dataset artificial:

```
df = pd.DataFrame({
    "Idade": [18, 22, 25, 30, 40, 50],
    "Renda": [1000, 2000, 3000, 4000, 6000, 8000],
    "Comprou": [0, 0, 1, 1, 1, 1] # 0 = não, 1 = sim
})
```

Treinando:

```
from sklearn.tree import DecisionTreeClassifier
```

```
X = df[["Idade", "Renda"]]  
y = df["Comprou"]  
  
modelo = DecisionTreeClassifier()  
modelo.fit(X, y)
```

Predição:

```
modelo.predict([[28, 3500]]) # idade, renda
```

Retorna **1** → **provável comprador**