

A **análise exploratória de dados (AED)** — ou *Exploratory Data Analysis (EDA)* em inglês — é a etapa inicial da análise de dados, em que o objetivo é **conhecer, entender e resumir** os dados antes de aplicar modelos estatísticos ou algoritmos de aprendizado de máquina.

👉 Em outras palavras, é o momento de investigar o conjunto de dados para descobrir padrões, identificar erros, entender relações entre variáveis e levantar hipóteses.

⌚ Objetivos principais da Análise Exploratória

1. Compreender a estrutura dos dados:

- Quantas colunas e linhas existem?
- Quais são os tipos de dados (numérico, texto, data etc.)?

2. Detectar valores ausentes e inconsistências:

- Existem valores nulos ou fora do padrão?

3. Identificar valores extremos (*outliers*):

- Há registros muito altos ou baixos que destoam dos demais?

4. Observar distribuições:

- Como as variáveis numéricas estão distribuídas (média, mediana, desvio padrão)?
- Existem assimetrias ou concentrações?

5. Ver relações entre variáveis:

- Como uma variável se relaciona com outra (ex: vendas × preço, idade × renda)?
-

🧠 Ferramentas e técnicas usadas

- **Bibliotecas (em Python):**

- **pandas** → manipulação e resumo dos dados
- **matplotlib e seaborn** → visualização de gráficos
- **numpy** → cálculos numéricos

- **Funções comuns:**

```
df.head()          # primeiras linhas  
df.info()         # tipos de dados e nulos  
df.describe()     # estatísticas descritivas  
df.isnull().sum()# contagem de valores nulos  
df.corr()         # correlação entre variáveis
```

- **Gráficos típicos:**

- Histograma → distribuição de uma variável
- Boxplot → detecção de outliers
- Gráfico de dispersão → relação entre duas variáveis
- Gráfico de barras → comparação de categorias



Exemplo prático simples

```
import pandas as pd  
import matplotlib.pyplot as plt  
  
df = pd.read_csv("vendas.csv")  
  
print(df.describe())  
print(df.isnull().sum())  
  
plt.hist(df['total_vendas'])  
plt.title("Distribuição de Vendas")  
plt.show()
```

Nesse exemplo, o analista pode perceber se há produtos com vendas muito altas (outliers), se há dados ausentes ou se as vendas seguem uma distribuição normal.



Em resumo:

A Análise Exploratória de Dados é como uma “investigação inicial” para entender o que os dados têm a dizer antes de tomar decisões ou criar modelos.

Aqui vai um roteiro passo a passo de uma Análise Exploratória de Dados (AED) completa em Python, com **exemplos práticos**, explicações e os **gráficos mais usados**.



1. Importar bibliotecas e carregar os dados

```
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
# Exemplo: arquivo de vendas  
Df = pd.read_csv("vendas.csv")
```

Dica: Use `encoding='utf-8'` se o arquivo tiver acentos, e `sep=';'` se o separador for ponto e vírgula.



2. Conhecer a estrutura dos dados

```
df.head()          # Mostra as 5 primeiras linhas  
df.info()         # Mostra tipos de dados e valores nulos  
df.shape          # Retorna número de linhas e colunas  
df.columns        # Lista o nome das colunas
```

- ⌚ Objetivo: entender o que há no dataset — variáveis, tipos (numérico, texto, data) e se há dados faltando.
-



3. Estatísticas descritivas

```
df.describe().round(2)
```

- `count`: número de valores válidos
- `mean`: média
- `std`: desvio padrão (dispersão dos dados)
- `min` e `max`: valores extremos

💡 Serve para identificar dados fora da faixa esperada (ex: idade = 200 anos, preço = -50).



4. Identificar valores ausentes e duplicados

```
df.isnull().sum()      # Conta valores ausentes por coluna  
df.duplicated().sum() # Conta linhas duplicadas
```

Para limpar:

```
df.drop_duplicates(inplace=True)  
df.fillna(0, inplace=True) # ou outro valor apropriado
```



5. Entender variáveis categóricas e numéricas

```
df.select_dtypes(include='object').columns    # categóricas  
df.select_dtypes(include=np.number).columns    # numéricas
```

Salve o código abaixo com vendas.csv

loja	categoria	produto	quantidade	preco	total_vendas
A	Eletrônicos	TV	5	1800	9000
B	Roupas	Camiseta	10	50	500
A	Eletrônicos	Notebook	2	3500	7000
C	Alimentos	Arroz	30	25	750
B	Eletrônicos	Fone	8	200	1600
A	Roupas	Calça	5	120	600
C	Alimentos	Feijão	20	10	200
B	Alimentos	Refrigerante	15	8	120