

Relatório - Mineração de Preços de Casas

Códigos e Explicações

1-

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.preprocessing import StandardScaler
```

```
dados = pd.read_csv("/content/PRECO_S_CASAS.csv")
```

Explicação:

Esses códigos importam bibliotecas essenciais, como Pandas para manipulação de dados e Scikit-learn para aprendizado de máquina. Em seguida, eles carregam um conjunto de dados sobre preços de casas de um arquivo CSV.

2-

```
dados.isnull().sum()
dados.dropna(inplace=True)
```

Explicação: Esses comandos estão verificando se há valores nulos em nosso conjunto de dados e removendo as linhas que os contêm, garantindo que não haja dados ausentes.

Resposta:

```
Casa      0
Preco     0
Area      0
Quartos   0
Banheiros 0
Ofertas   0
Tijolo    0
Bairro    0
```

3-

```
scaler = StandardScaler()
```

```
dados["Area"] = scaler.fit_transform(dados["Area"].values.reshape(-1, 1))
dados["Quartos"] = scaler.fit_transform(dados["Quartos"].values.reshape(-1, 1))
dados["Banheiros"] = scaler.fit_transform(dados["Banheiros"].values.reshape(-1, 1))
```

Explicação: Esses trechos estão padronizando as variáveis 'Area', 'Quartos' e 'Banheiros'

usando StandardScaler. Isso é importante para garantir que todas as variáveis estejam na mesma escala.

4-

```
X = dados[["Area", "Quartos", "Banheiros"]]  
y = dados["Preco"]
```

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

Explicação: Aqui estamos dividindo nossos dados em conjuntos de treinamento e teste para avaliar o desempenho do modelo.

—

5-

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

Explicação: Estamos criando e treinando um modelo de regressão linear para prever os preços das casas com base nos atributos fornecidos.

6-

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

```
rmse = mean_squared_error(y_test, y_pred)  
r2 = r2_score(y_test, y_pred)
```

```
print(f"RMSE: {rmse:.2f}")  
print(f"R²: {r2:.2f}")
```

Explicação: Aqui estamos avaliando o desempenho do modelo usando as métricas de erro quadrático médio (RMSE) e coeficiente de determinação (R^2).

Resposta:

RMSE: 320149938.23

R²: 0.46

7-

```
print(f"Coeficientes: {modelo.coef_}")
```

Explicação: Estamos imprimindo os coeficientes aprendidos pelo modelo de regressão linear para cada variável.

Resposta: Coeficientes: [8734.83761539 7432.13195417 6708.54485934]

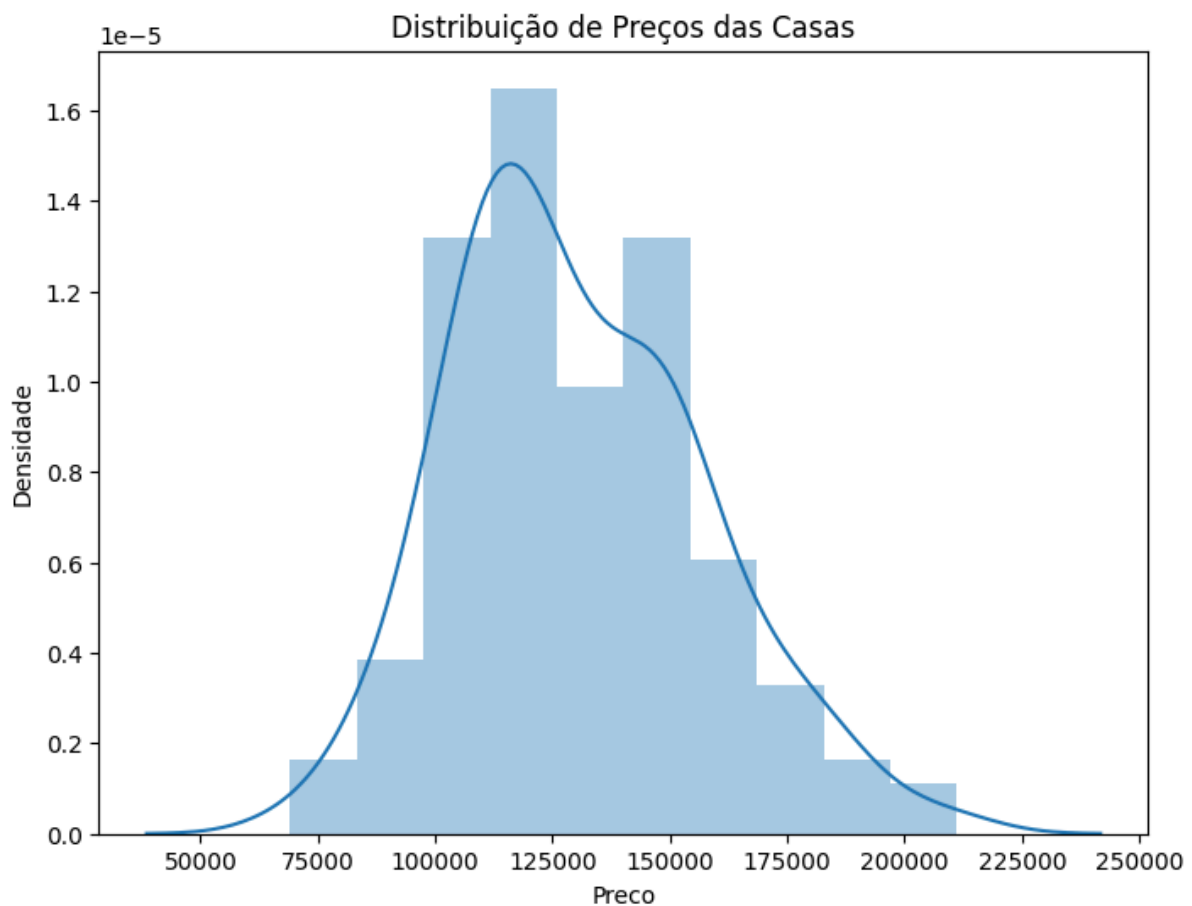
8-

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
plt.figure(figsize=(8, 6))
sns.distplot(dados["Preco"])
plt.xlabel("Preco")
plt.ylabel("Densidade")
plt.title("Distribuição de Preços das Casas")
plt.show()
```

Explicação: Aqui estamos visualizando a distribuição dos preços das casas usando um gráfico de distribuição.

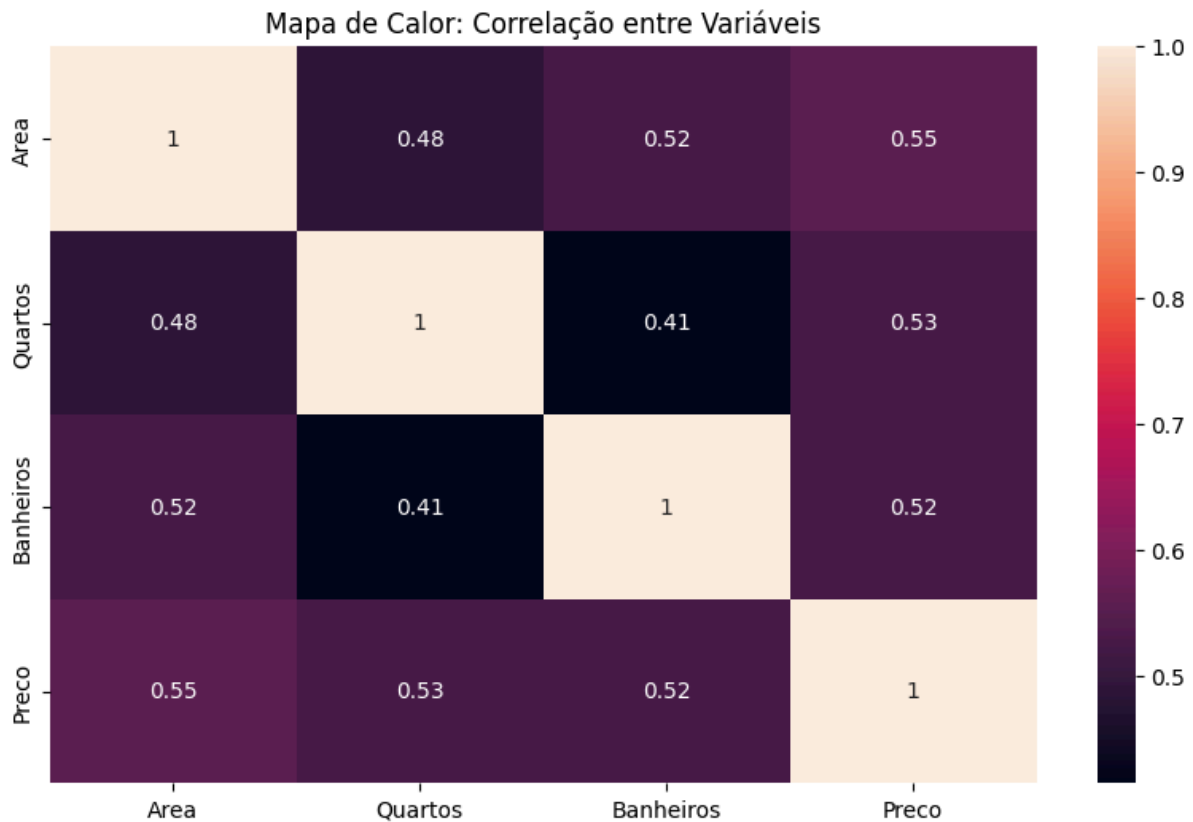
Resposta:



9-

```
plt.figure(figsize=(10, 6))
sns.heatmap(dados[["Area", "Quartos", "Banheiros", "Preco"]].corr(), annot=True)
plt.title("Mapa de Calor: Correlação entre Variáveis")
plt.show()
```

Explicação: Estamos visualizando a correlação entre as variáveis usando um mapa de valor, que mostra a relação entre 'Area', 'Quartos', 'Banheiros' e 'Preco'.



Conclusão do Relatório

Importância da Ciência de Dados para uma Empresa:

1. A Ciência de Dados é fundamental para as empresas pois permite extrair insights valiosos dos dados, gerando vantagem competitiva. Com análises avançadas, as empresas podem tomar decisões mais informadas, identificar padrões de mercado, entender o comportamento do cliente e otimizar processos internos, resultando em eficiência operacional e melhor desempenho financeiro.

Objetivo do Aprendizado de Máquina na Ciência de Dados:

2. O objetivo do Aprendizado de Máquina na Ciência de Dados é desenvolver algoritmos e modelos que possam aprender com os dados e fazer previsões ou tomar decisões sem serem explicitamente programados. Isso permite automatizar tarefas, como classificação, regressão, clusterização, entre outras, e descobrir padrões nos dados.

Objetivo e Mecanismos da Mineração de Dados:

3. O objetivo da Mineração de Dados é descobrir padrões, tendências e relações nos dados que possam ser úteis para tomar decisões ou realizar previsões. Isso é feito através de técnicas como regressão, classificação, clusterização, associação, entre outras, que ajudam a extrair conhecimento dos dados.

Web Scraping e sua Importância no Contexto de Big Data:

4. O Web Scraping é a técnica de extrair dados de websites de forma automatizada. No contexto de Big Data, onde lidamos com grandes volumes de dados de diversas fontes, o Web Scraping pode ser uma ferramenta poderosa para coletar dados relevantes de fontes na web que não estão disponíveis em conjuntos de dados convencionais. Isso amplia o escopo dos dados que podem ser analisados e utilizados na ciência de dados.