

Modelagem Linear Para Aprendizado de Máquina

Sprint 02

1CCR

Integrantes

Leticia Giordani RM: 564028

Lívia Dos Santos RM: 562967

Luize Martinês RM: 564016

Sarah Sayako RM: 563841

Professor: Roberto Gutierrez

Parte 1 – Compreensão Inicial do Problema

As colunas mais relevantes são:

Solar_Potential_kwh_m2 : Mostra quanta energia pode ser gerada.

Installation_Area_m2 : Indica o espaço disponível para instalar os painéis solares.

Energy_Demand_kwh : mostra o quanto de energia o local consome.

Optimal_Solar_Utilization_% : mostra quão bem o sistema pode aproveitar a energia solar.

Payback_Period_Years : mostra em quantos anos o sistema se pagaria.

Carbon_Reduction_% : mostra em quantos anos o sistema se pagaria.

Justificativa : essas colunas estão diretamente ligadas a eficiência, viabilidade financeira e ambiental da instalação.

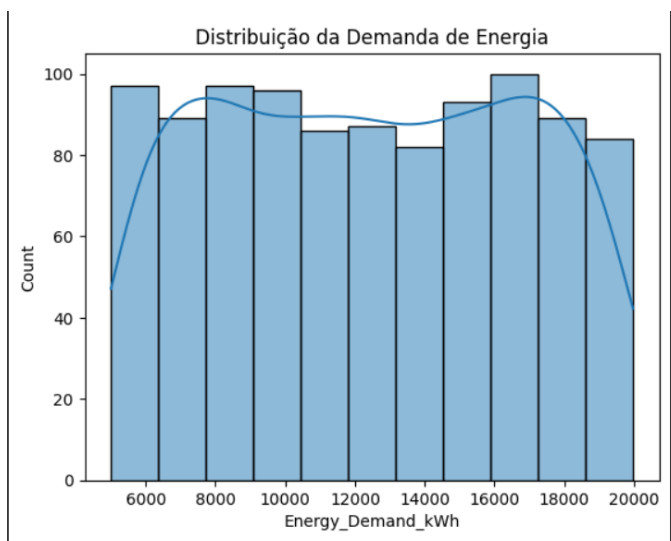
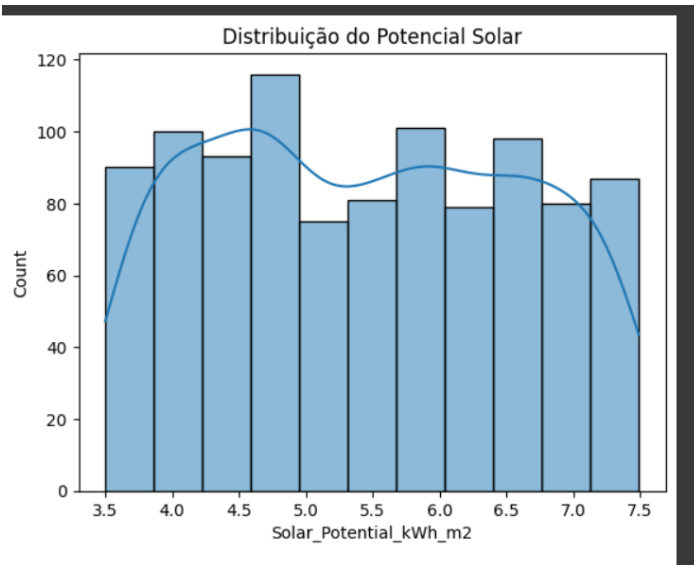
Parte 2 – Visualização dos Dados

2.1 Histogramas de duas variáveis numéricas

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

sns.histplot(df['Solar_Potential_kWh_m2'], kde=True)
plt.title('Distribuição do Potencial Solar')
plt.show()

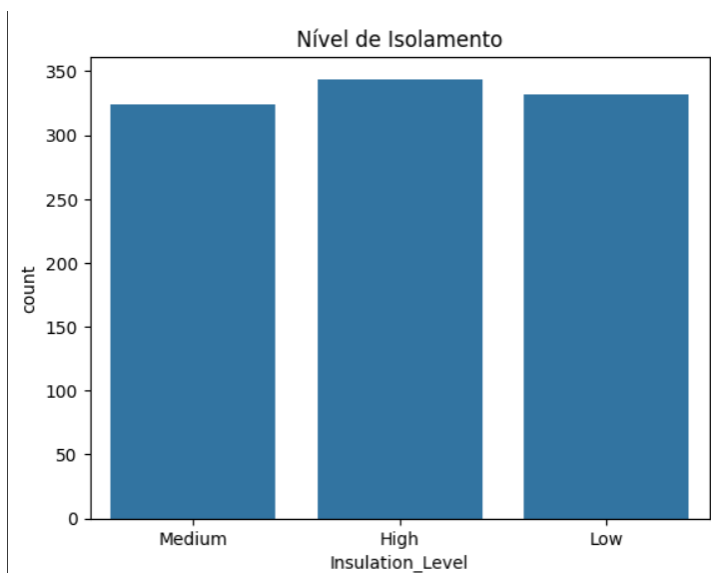
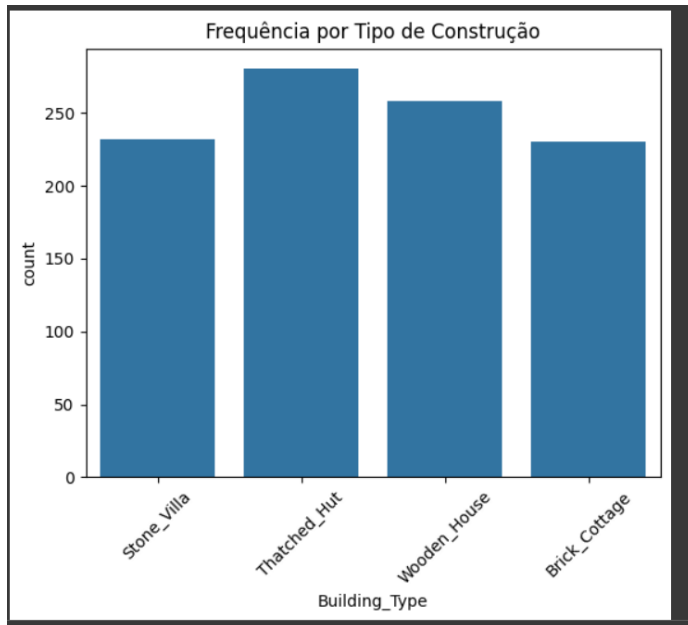
sns.histplot(df['Energy_Demand_kWh'], kde=True)
plt.title('Distribuição da Demanda de Energia')
plt.show()
```



2.2 Gráficos de barras de duas variáveis categóricas

```
sns.countplot(x='Building_Type', data=df)
plt.title('Frequência por Tipo de Construção')
plt.xticks(rotation=45)
plt.show()

sns.countplot(x='Insulation_Level', data=df)
plt.title('Nível de Isolamento')
plt.show()
```



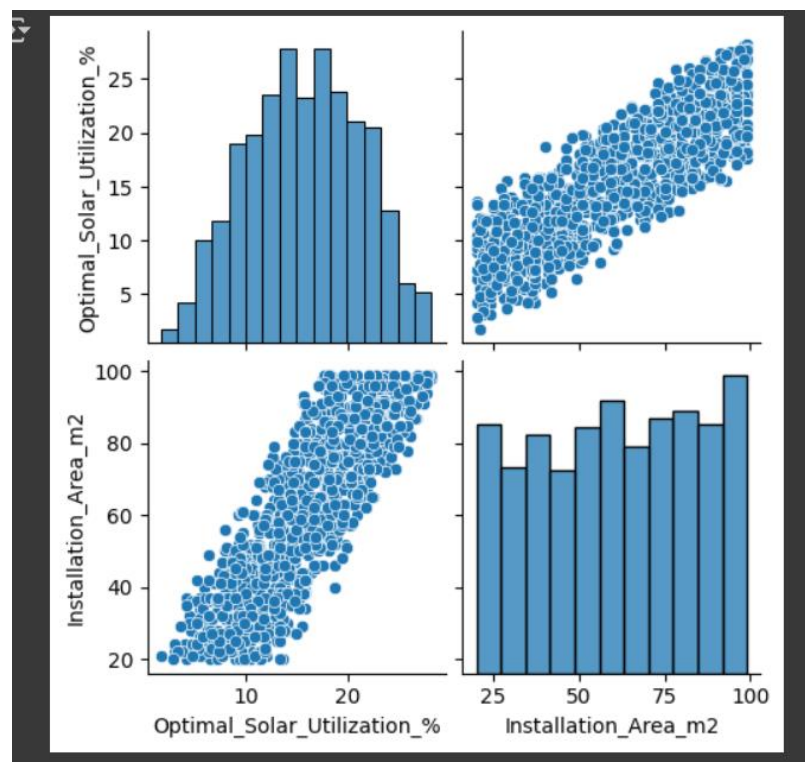
Interpretação dos gráficos: Os histogramas mostram distribuições próximas da normal, principalmente no potencial solar. Nos gráficos de barras, há variação entre as categorias, com algumas mais frequentes. Isso sugere um leve desbalanceamento, comum em dados reais, reforçando o **realismo** do conjunto de dados

Parte 3 – Explorando Relações entre colunas com Pair Plot

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

cols = ['Optimal_Solar_Utilization_', 'Installation_Area_m2']
sns.pairplot(df[cols])

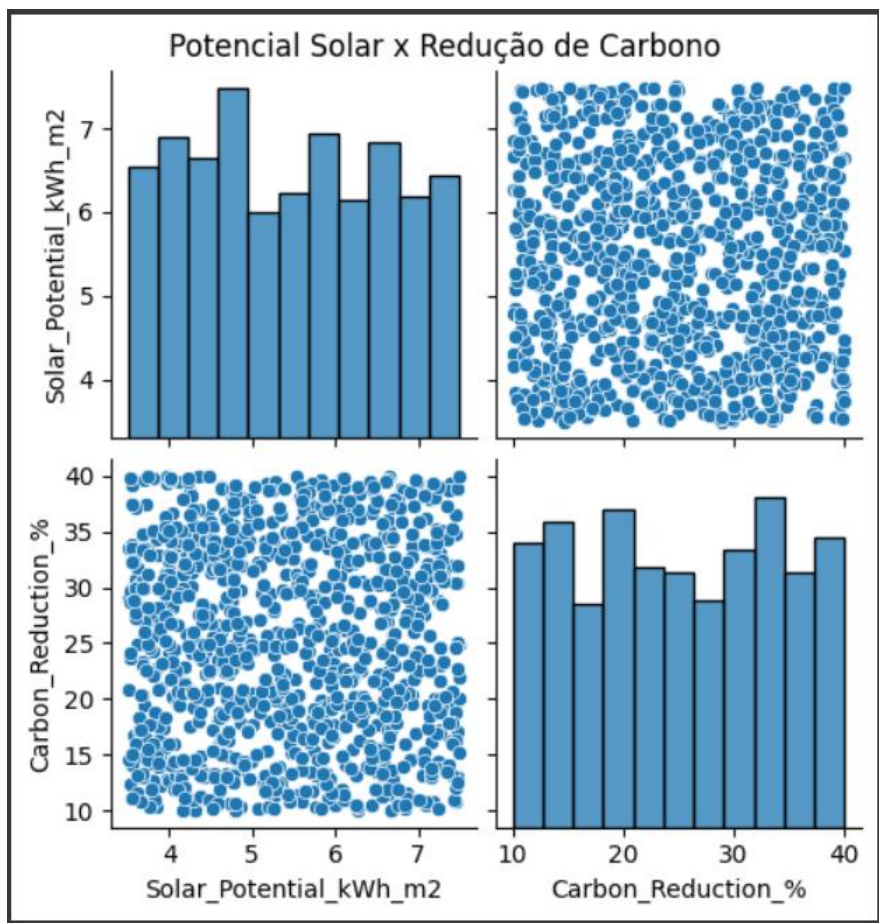
plt.show()
```



Existe uma relação positiva: quanto maior a área de instalação, maior tende a ser a utilização de energia solar. Isso faz sentido, pois mais espaço permite colocar mais painéis solares.

Parte 4 – Explorando relações entre outras colunas

```
▶ sns.pairplot(df, vars=['Solar_Potential_kWh_m2', 'Carbon_Reduction_%'])  
plt.suptitle('Potencial Solar x Redução de Carbono', y=1.02)  
plt.show()
```



há uma relação aparente: locais com maior potencial solar tendem a contribuir mais com a redução de carbono. Isso confirma a expectativa de que o uso eficiente da energia solar ajuda na sustentabilidade.

5. Reflexão final sobre o dataset

O dataset utilizado nesta atividade é chamado de Green Energy Dataset e tem como objetivo simular informações relacionadas ao aproveitamento de energia solar em diferentes tipos de construções. Ele apresenta dados como área disponível para instalação de painéis solares, tipo de construção, nível de isolamento térmico, potencial solar do local, entre outros. Essas informações permitem analisar a viabilidade da instalação de sistemas de energia solar em residências ou empresas, identificando os fatores que mais influenciam no aproveitamento da energia solar. Embora provavelmente seja um conjunto de dados fictício, ele é útil para fins educacionais e para a aplicação de técnicas de análise de dados e modelagem linear em problemas reais relacionados a sustentabilidade.

.