



```
colab:search.google.com/drive/1UcaY0uJH4nEq/a5Y00QyWp1dF5roll9-AUR0IM0MBLJB

Untitled8.ipynb
File Edit View Insert Runtime Tools Help
0 Commands 4 Code 4 Text

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler, LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error

# Load data
df = pd.read_csv('HousePricePrediction.xlsx')
print(df.head())
print(df.describe())

# Check for missing values
print(df.isnull().sum())

# Correlation matrix to understand feature relationships
correlation_matrix = df.select_dtypes(include='number').corr()

sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()

# Preprocessing: Selecting features and target variable
X = df[['bedrooms', 'bathrooms', 'year_built', 'living_area_sq_ft', 'lot_size_sq_ft', 'pool', 'fireplace', 'garage', 'basement']]
y = df['price']

# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

