

```
# Import necessary libraries
import os
import numpy as np
import pandas as pd
import librosa
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error
import joblib
```

```
# Set paths to your dataset directories
TRAIN_AUDIO_DIR = 'dataset\audios_train'
TEST_AUDIO_DIR = 'dataset\audios_test'
```

```
# Load CSV files
train_df = pd.read_csv('C:/Users/dubey/Downloads/shl-intern-hiring-
assessment (1)/dataset/train.csv')
test_df = pd.read_csv('C:/Users/dubey/Downloads/shl-intern-hiring-
assessment (1)/dataset/test.csv')
print("Training samples:", train_df.shape[0])
print("Test samples:", test_df.shape[0])
```

```
Training samples: 444
Test samples: 195
```

```
# Display the first few rows of the training data
print("First 5 rows of training data:")
print(train_df.head())
```

```
First 5 rows of training data:
```

	filename
0	audio_706.wav
1	audio_800.wav
2	audio_68.wav
3	audio_1267.wav
4	audio_683.wav

```
# Basic statistics about the training data
print("\nTraining Data Summary:")
print(train_df.describe())
```

```
Training Data Summary:
```

	filename
count	195
unique	195
top	audio_706.wav
freq	1

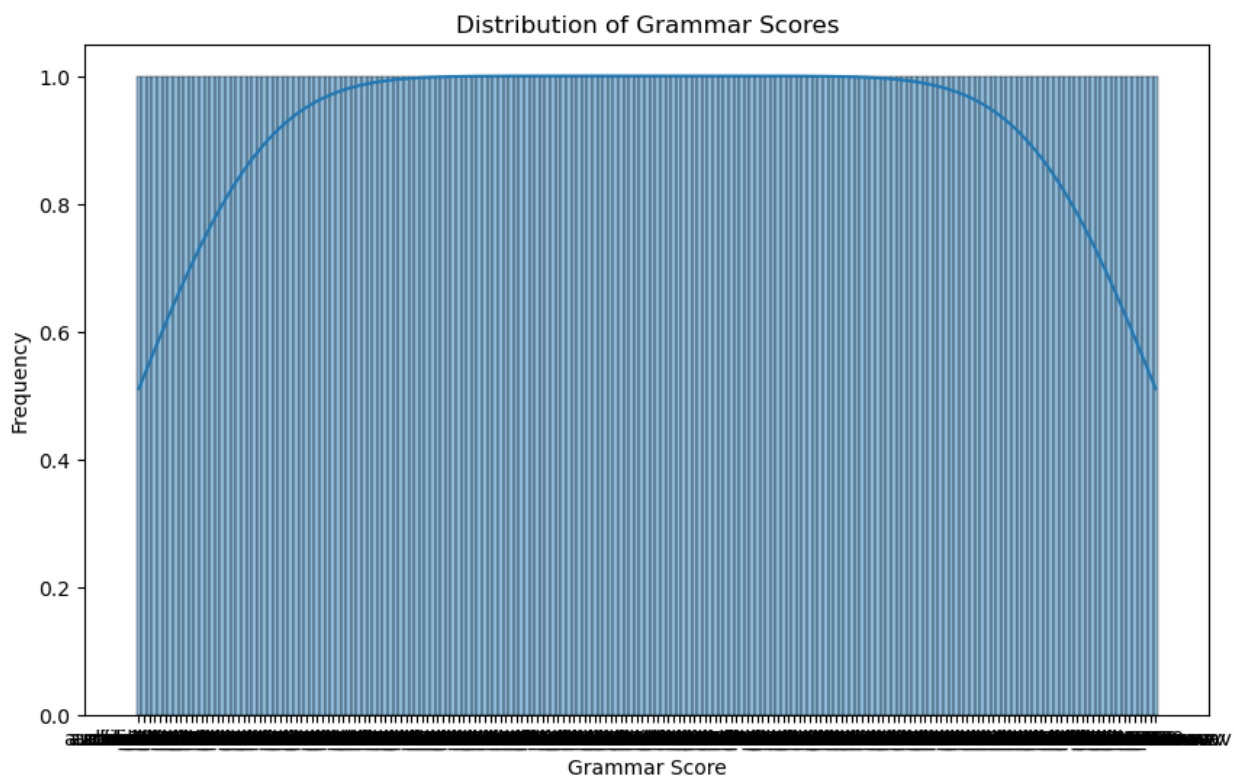
```

# Check for missing values
print("\nMissing values in training data:")
print(train_df.isnull().sum())

Missing values in training data:
filename      0
dtype: int64

# Plot distribution of Grammar Scores
plt.figure(figsize=(10, 6))
sns.histplot(train_df['filename'], bins=10, kde=True)
plt.title('Distribution of Grammar Scores')
plt.xlabel('Grammar Score')
plt.ylabel('Frequency')
plt.show()

```



```

import os
import librosa

def get_audio_duration(file_path, sr=22050):
    try:
        y, sr = librosa.load(file_path, sr=sr)
        return librosa.get_duration(y=y, sr=sr)
    except:
        pass

```

```

except Exception as e:
    print(f"Error loading {file_path}: {e}")
    return None

# Compute durations for each audio file in the training set
durations = []
for file in train_df['filename']:
    path = os.path.join('dataset/audios_train', file)
    duration = get_audio_duration(path)
    durations.append(duration)

# Add durations as a new column in the training dataframe
train_df['duration'] = durations

# Plot the distribution of audio durations
plt.figure(figsize=(10, 6))
sns.histplot(train_df['duration'], bins=20, kde=True, color='teal')
plt.title("Distribution of Audio Durations")
plt.xlabel("Duration (seconds)")
plt.ylabel("Frequency")
plt.show()

C:\Users\dubey\AppData\Local\Temp\ipykernel_4668\982126251.py:6:
UserWarning: PySoundFile failed. Trying audioread instead.
    y, sr = librosa.load(file_path, sr=sr)
c:\Users\dubey\anaconda3\Lib\site-packages\librosa\core\audio.py:184:
FutureWarning: librosa.core.audio.__audioread_load
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)

Error loading dataset/audios_train\audio_706.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_706.wav'
Error loading dataset/audios_train\audio_800.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_800.wav'
Error loading dataset/audios_train\audio_68.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_68.wav'
Error loading dataset/audios_train\audio_1267.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_1267.wav'
Error loading dataset/audios_train\audio_683.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_683.wav'
Error loading dataset/audios_train\audio_1242.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_1242.wav'
Error loading dataset/audios_train\audio_908.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_908.wav'
Error loading dataset/audios_train\audio_888.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_888.wav'
Error loading dataset/audios_train\audio_137.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\audio_137.wav'
Error loading dataset/audios_train\audio_770.wav: [Errno 2] No such

```

file or directory: 'dataset/audios_train\\audio_770.wav'
Error loading dataset/audios_train\\audio_735.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_735.wav'
Error loading dataset/audios_train\\audio_1026.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1026.wav'
Error loading dataset/audios_train\\audio_1214.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1214.wav'
Error loading dataset/audios_train\\audio_1122.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1122.wav'
Error loading dataset/audios_train\\audio_1022.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1022.wav'
Error loading dataset/audios_train\\audio_726.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_726.wav'
Error loading dataset/audios_train\\audio_1205.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1205.wav'
Error loading dataset/audios_train\\audio_1116.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1116.wav'
Error loading dataset/audios_train\\audio_1240.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1240.wav'
Error loading dataset/audios_train\\audio_151.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_151.wav'
Error loading dataset/audios_train\\audio_437.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_437.wav'
Error loading dataset/audios_train\\audio_1217.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1217.wav'
Error loading dataset/audios_train\\audio_831.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_831.wav'
Error loading dataset/audios_train\\audio_1315.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1315.wav'
Error loading dataset/audios_train\\audio_1323.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1323.wav'
Error loading dataset/audios_train\\audio_1256.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1256.wav'
Error loading dataset/audios_train\\audio_1033.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1033.wav'
Error loading dataset/audios_train\\audio_858.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_858.wav'
Error loading dataset/audios_train\\audio_274.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_274.wav'
Error loading dataset/audios_train\\audio_196.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_196.wav'
Error loading dataset/audios_train\\audio_1193.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1193.wav'
Error loading dataset/audios_train\\audio_138.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_138.wav'
Error loading dataset/audios_train\\audio_644.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_644.wav'
Error loading dataset/audios_train\\audio_1012.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1012.wav'

Error loading dataset/audios_train\audio_1101.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1101.wav'
Error loading dataset/audios_train\audio_805.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_805.wav'
Error loading dataset/audios_train\audio_709.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_709.wav'
Error loading dataset/audios_train\audio_1293.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1293.wav'
Error loading dataset/audios_train\audio_1123.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1123.wav'
Error loading dataset/audios_train\audio_1048.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1048.wav'
Error loading dataset/audios_train\audio_820.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_820.wav'
Error loading dataset/audios_train\audio_767.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_767.wav'
Error loading dataset/audios_train\audio_841.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_841.wav'
Error loading dataset/audios_train\audio_1286.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1286.wav'
Error loading dataset/audios_train\audio_665.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_665.wav'
Error loading dataset/audios_train\audio_1311.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1311.wav'
Error loading dataset/audios_train\audio_29.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_29.wav'
Error loading dataset/audios_train\audio_1289.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1289.wav'
Error loading dataset/audios_train\audio_177.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_177.wav'
Error loading dataset/audios_train\audio_599.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_599.wav'
Error loading dataset/audios_train\audio_920.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_920.wav'
Error loading dataset/audios_train\audio_971.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_971.wav'
Error loading dataset/audios_train\audio_1275.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1275.wav'
Error loading dataset/audios_train\audio_1291.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1291.wav'
Error loading dataset/audios_train\audio_320.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_320.wav'
Error loading dataset/audios_train\audio_545.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_545.wav'
Error loading dataset/audios_train\audio_287.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_287.wav'
Error loading dataset/audios_train\audio_172.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_172.wav'
Error loading dataset/audios_train\audio_322.wav: [Errno 2] No such

file or directory: 'dataset/audios_train\\audio_322.wav'
Error loading dataset/audios_train\\audio_541.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_541.wav'
Error loading dataset/audios_train\\audio_148.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_148.wav'
Error loading dataset/audios_train\\audio_1317.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1317.wav'
Error loading dataset/audios_train\\audio_1243.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1243.wav'
Error loading dataset/audios_train\\audio_348.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_348.wav'
Error loading dataset/audios_train\\audio_662.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_662.wav'
Error loading dataset/audios_train\\audio_540.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_540.wav'
Error loading dataset/audios_train\\audio_733.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_733.wav'
Error loading dataset/audios_train\\audio_1297.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1297.wav'
Error loading dataset/audios_train\\audio_949.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_949.wav'
Error loading dataset/audios_train\\audio_1035.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1035.wav'
Error loading dataset/audios_train\\audio_882.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_882.wav'
Error loading dataset/audios_train\\audio_401.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_401.wav'
Error loading dataset/audios_train\\audio_113.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_113.wav'
Error loading dataset/audios_train\\audio_435.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_435.wav'
Error loading dataset/audios_train\\audio_719.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_719.wav'
Error loading dataset/audios_train\\audio_698.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_698.wav'
Error loading dataset/audios_train\\audio_959.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_959.wav'
Error loading dataset/audios_train\\audio_1166.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1166.wav'
Error loading dataset/audios_train\\audio_543.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_543.wav'
Error loading dataset/audios_train\\audio_89.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_89.wav'
Error loading dataset/audios_train\\audio_884.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_884.wav'
Error loading dataset/audios_train\\audio_19.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_19.wav'
Error loading dataset/audios_train\\audio_641.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_641.wav'

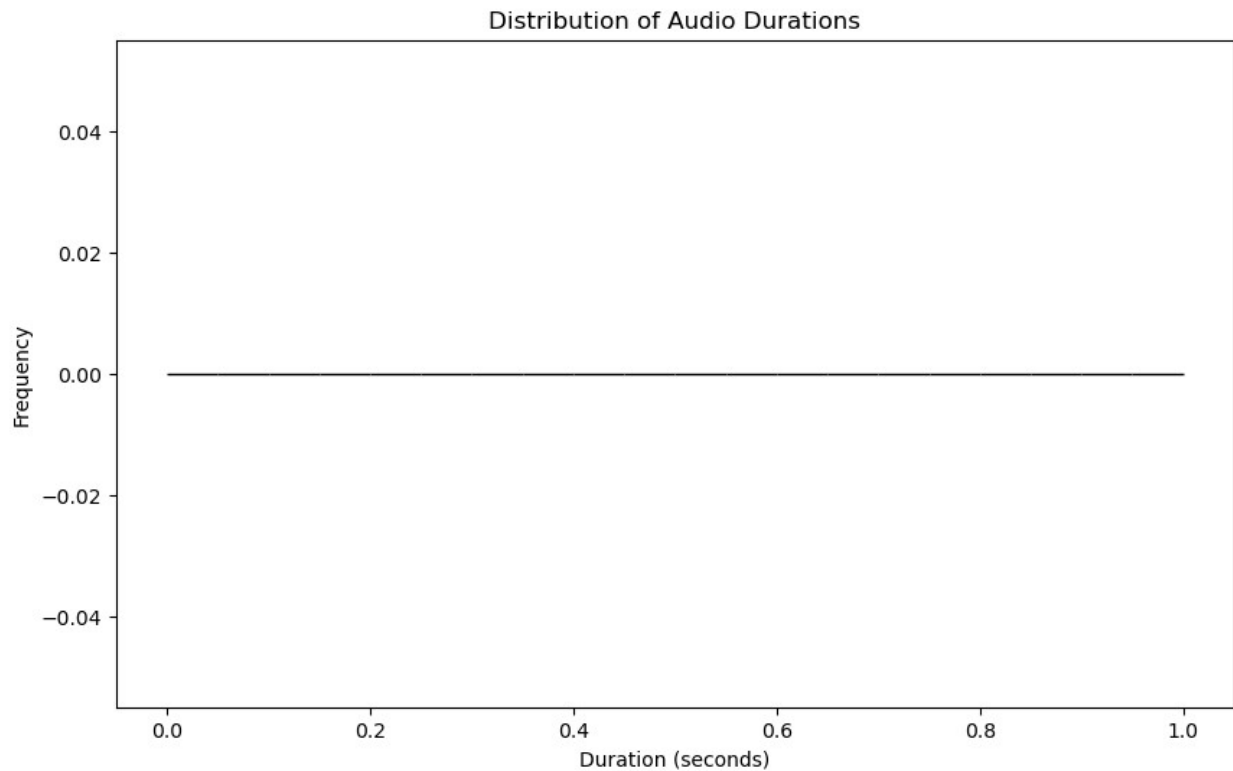
Error loading dataset/audios_train\audio_1115.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1115.wav'
Error loading dataset/audios_train\audio_811.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_811.wav'
Error loading dataset/audios_train\audio_75.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_75.wav'
Error loading dataset/audios_train\audio_1138.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1138.wav'
Error loading dataset/audios_train\audio_66.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_66.wav'
Error loading dataset/audios_train\audio_500.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_500.wav'
Error loading dataset/audios_train\audio_922.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_922.wav'
Error loading dataset/audios_train\audio_656.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_656.wav'
Error loading dataset/audios_train\audio_218.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_218.wav'
Error loading dataset/audios_train\audio_1278.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1278.wav'
Error loading dataset/audios_train\audio_525.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_525.wav'
Error loading dataset/audios_train\audio_290.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_290.wav'
Error loading dataset/audios_train\audio_1195.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1195.wav'
Error loading dataset/audios_train\audio_572.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_572.wav'
Error loading dataset/audios_train\audio_1089.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1089.wav'
Error loading dataset/audios_train\audio_488.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_488.wav'
Error loading dataset/audios_train\audio_759.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_759.wav'
Error loading dataset/audios_train\audio_103.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_103.wav'
Error loading dataset/audios_train\audio_1280.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1280.wav'
Error loading dataset/audios_train\audio_153.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_153.wav'
Error loading dataset/audios_train\audio_300.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_300.wav'
Error loading dataset/audios_train\audio_20.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_20.wav'
Error loading dataset/audios_train\audio_10.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_10.wav'
Error loading dataset/audios_train\audio_1292.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1292.wav'
Error loading dataset/audios_train\audio_281.wav: [Errno 2] No such

file or directory: 'dataset/audios_train\\audio_281.wav'
Error loading dataset/audios_train\\audio_261.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_261.wav'
Error loading dataset/audios_train\\audio_1081.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1081.wav'
Error loading dataset/audios_train\\audio_276.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_276.wav'
Error loading dataset/audios_train\\audio_159.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_159.wav'
Error loading dataset/audios_train\\audio_1124.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1124.wav'
Error loading dataset/audios_train\\audio_690.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_690.wav'
Error loading dataset/audios_train\\audio_1061.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1061.wav'
Error loading dataset/audios_train\\audio_521.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_521.wav'
Error loading dataset/audios_train\\audio_263.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_263.wav'
Error loading dataset/audios_train\\audio_1068.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1068.wav'
Error loading dataset/audios_train\\audio_4.wav: [Errno 2] No such file
or directory: 'dataset/audios_train\\audio_4.wav'
Error loading dataset/audios_train\\audio_676.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_676.wav'
Error loading dataset/audios_train\\audio_165.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_165.wav'
Error loading dataset/audios_train\\audio_282.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_282.wav'
Error loading dataset/audios_train\\audio_1179.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1179.wav'
Error loading dataset/audios_train\\audio_217.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_217.wav'
Error loading dataset/audios_train\\audio_388.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_388.wav'
Error loading dataset/audios_train\\audio_550.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_550.wav'
Error loading dataset/audios_train\\audio_499.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_499.wav'
Error loading dataset/audios_train\\audio_1190.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1190.wav'
Error loading dataset/audios_train\\audio_433.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_433.wav'
Error loading dataset/audios_train\\audio_885.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_885.wav'
Error loading dataset/audios_train\\audio_48.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_48.wav'
Error loading dataset/audios_train\\audio_308.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_308.wav'

Error loading dataset/audios_train\audio_746.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_746.wav'
Error loading dataset/audios_train\audio_1019.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1019.wav'
Error loading dataset/audios_train\audio_487.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_487.wav'
Error loading dataset/audios_train\audio_198.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_198.wav'
Error loading dataset/audios_train\audio_958.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_958.wav'
Error loading dataset/audios_train\audio_72.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_72.wav'
Error loading dataset/audios_train\audio_1321.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1321.wav'
Error loading dataset/audios_train\audio_1159.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1159.wav'
Error loading dataset/audios_train\audio_428.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_428.wav'
Error loading dataset/audios_train\audio_1058.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1058.wav'
Error loading dataset/audios_train\audio_408.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_408.wav'
Error loading dataset/audios_train\audio_857.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_857.wav'
Error loading dataset/audios_train\audio_692.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_692.wav'
Error loading dataset/audios_train\audio_633.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_633.wav'
Error loading dataset/audios_train\audio_1173.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1173.wav'
Error loading dataset/audios_train\audio_554.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_554.wav'
Error loading dataset/audios_train\audio_180.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_180.wav'
Error loading dataset/audios_train\audio_1013.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1013.wav'
Error loading dataset/audios_train\audio_386.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_386.wav'
Error loading dataset/audios_train\audio_519.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_519.wav'
Error loading dataset/audios_train\audio_564.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_564.wav'
Error loading dataset/audios_train\audio_1054.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1054.wav'
Error loading dataset/audios_train\audio_556.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_556.wav'
Error loading dataset/audios_train\audio_1183.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1183.wav'
Error loading dataset/audios_train\audio_306.wav: [Errno 2] No such

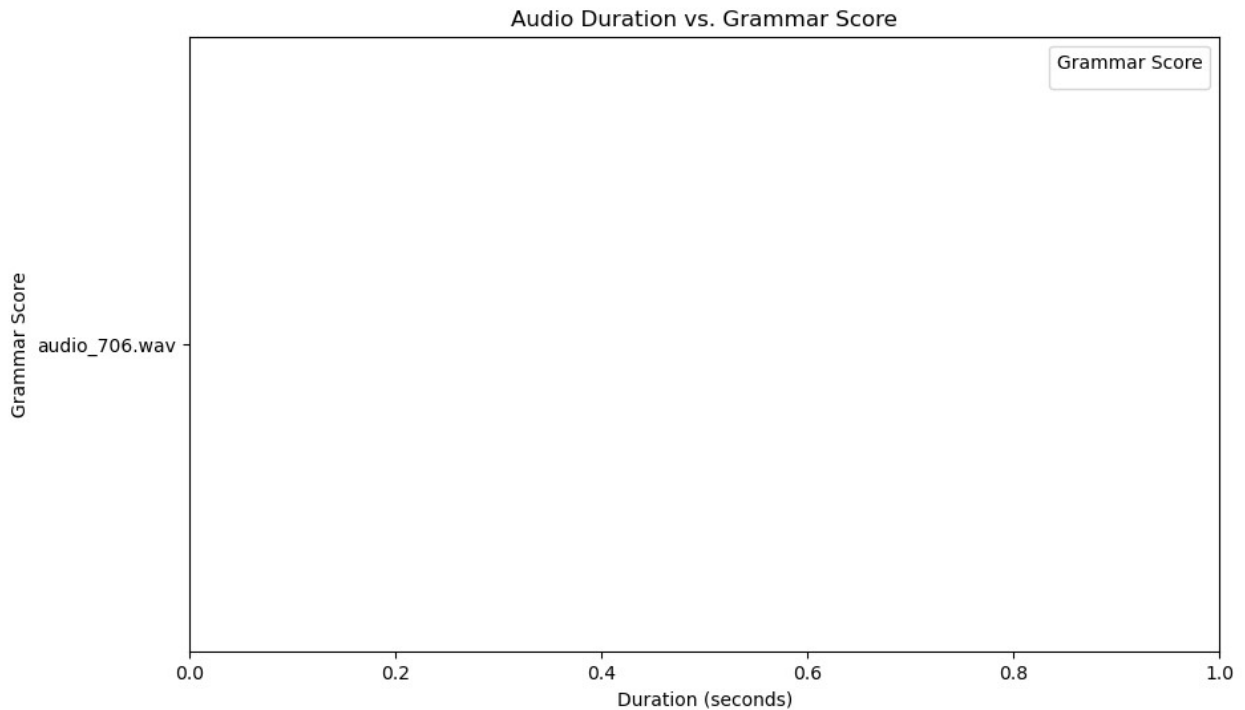
file or directory: 'dataset/audios_train\\audio_306.wav'
Error loading dataset/audios_train\\audio_713.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_713.wav'
Error loading dataset/audios_train\\audio_225.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_225.wav'
Error loading dataset/audios_train\\audio_1091.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1091.wav'
Error loading dataset/audios_train\\audio_107.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_107.wav'
Error loading dataset/audios_train\\audio_286.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_286.wav'
Error loading dataset/audios_train\\audio_546.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_546.wav'
Error loading dataset/audios_train\\audio_580.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_580.wav'
Error loading dataset/audios_train\\audio_235.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_235.wav'
Error loading dataset/audios_train\\audio_109.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_109.wav'
Error loading dataset/audios_train\\audio_932.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_932.wav'
Error loading dataset/audios_train\\audio_330.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_330.wav'
Error loading dataset/audios_train\\audio_569.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_569.wav'
Error loading dataset/audios_train\\audio_158.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_158.wav'
Error loading dataset/audios_train\\audio_321.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_321.wav'
Error loading dataset/audios_train\\audio_360.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_360.wav'
Error loading dataset/audios_train\\audio_391.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_391.wav'
Error loading dataset/audios_train\\audio_448.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_448.wav'
Error loading dataset/audios_train\\audio_1215.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1215.wav'
Error loading dataset/audios_train\\audio_422.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_422.wav'
Error loading dataset/audios_train\\audio_998.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_998.wav'
Error loading dataset/audios_train\\audio_702.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_702.wav'
Error loading dataset/audios_train\\audio_897.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_897.wav'
Error loading dataset/audios_train\\audio_394.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_394.wav'
Error loading dataset/audios_train\\audio_1169.wav: [Errno 2] No such
file or directory: 'dataset/audios_train\\audio_1169.wav'

Error loading dataset/audios_train\audio_95.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_95.wav'
Error loading dataset/audios_train\audio_604.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_604.wav'
Error loading dataset/audios_train\audio_221.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_221.wav'
Error loading dataset/audios_train\audio_34.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_34.wav'
Error loading dataset/audios_train\audio_179.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_179.wav'
Error loading dataset/audios_train\audio_21.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_21.wav'
Error loading dataset/audios_train\audio_1176.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1176.wav'
Error loading dataset/audios_train\audio_285.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_285.wav'
Error loading dataset/audios_train\audio_1178.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_1178.wav'
Error loading dataset/audios_train\audio_135.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_135.wav'
Error loading dataset/audios_train\audio_512.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_512.wav'
Error loading dataset/audios_train\audio_529.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_529.wav'
Error loading dataset/audios_train\audio_762.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_762.wav'
Error loading dataset/audios_train\audio_379.wav: [Errno 2] No such file or directory: 'dataset/audios_train\\audio_379.wav'



```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=train_df, x='duration', y='filename',
               hue='filename', palette='coolwarm')
plt.title("Audio Duration vs. Grammar Score")
plt.xlabel("Duration (seconds)")
plt.ylabel("Grammar Score")
plt.legend(title='Grammar Score')
plt.show()
```

No artists with labels found to put in legend. Note that artists whose label start with an underscore are ignored when legend() is called with no argument.



```
# If there's any categorical feature or file names you want to
# examine,
# you can also check their unique values or counts. For example:
print("\nUnique file names in training data:")
print(train_df['filename'].nunique())
```

```
Unique file names in training data:
195
```

3. Model Training and Evaluation

Split the training data into training and validation sets. Train a regression model (e.g., RandomForestRegressor) to predict scores. Then evaluate using metrics such as MSE and MAE.

```
import os

TRAIN_AUDIO_DIR = 'C:/Users/dubey/Downloads/shl-intern-hiring-
assessment (1)/dataset/audios_train' # Adjust the path as needed

# List files in the directory to confirm they exist
files = os.listdir(TRAIN_AUDIO_DIR)
print("Files in the training directory:", files[:5]) # Print the
first 5 files

Files in the training directory: ['audio_1008.wav', 'audio_1017.wav',
'audio_102.wav', 'audio_1024.wav', 'audio_1025.wav']
```

```
# For example, if the audios_train folder is in the current directory:
TRAIN_AUDIO_DIR = 'audios_train'
```

```
file_path = os.path.join(TRAIN_AUDIO_DIR, row['filename'])
```

```
import os
import librosa
```

```
TRAIN_AUDIO_DIR = 'C:/Users/dubey/Downloads/shl-intern-hiring-
assessment (1)/dataset/audios_train'
```

```
# List a few audio files
```

```
audio_files = [f for f in os.listdir(TRAIN_AUDIO_DIR) if
f.endswith('.wav')]
print("Sample files:", audio_files[:5])
```

```
# Load a sample audio
```

```
file_path = os.path.join(TRAIN_AUDIO_DIR, audio_files[0])
y, sr = librosa.load(file_path, sr=None)
print(f"Loaded: {file_path} | Duration: {len(y)/sr:.2f} sec | Sample
Rate: {sr}")
```

```
Sample files: ['audio_1008.wav', 'audio_1017.wav', 'audio_102.wav',
'audio_1024.wav', 'audio_1025.wav']
```

```
Loaded: C:/Users/dubey/Downloads/shl-intern-hiring-assessment
(1)/dataset/audios_train/audio_1008.wav | Duration: 60.01 sec | Sample
Rate: 16000
```

```
import pandas as pd
```

```
data = []
```

```
for file in audio_files:
    path = os.path.join(TRAIN_AUDIO_DIR, file)
    try:
        y, sr = librosa.load(path, sr=None)
        duration = len(y) / sr
        data.append({'filename': file, 'duration': duration,
'sample_rate': sr})
    except Exception as e:
        print(f"Error loading {file}: {e}")
```

```
df = pd.DataFrame(data)
```

```
print(df.head())
```

	filename	duration	sample_rate
0	audio_1008.wav	60.010688	16000
1	audio_1017.wav	60.074688	16000
2	audio_102.wav	60.074688	16000
3	audio_1024.wav	60.074688	16000
4	audio_1025.wav	61.034687	16000

```

import numpy as np
import librosa

def extract_logmel(file_path, sr=16000, n_mels=128):
    y, _ = librosa.load(file_path, sr=sr)
    mel = librosa.feature.melspectrogram(y=y, sr=sr, n_mels=n_mels)
    logmel = librosa.power_to_db(mel)
    return logmel

# Example
logmel = extract_logmel(file_path)
print(logmel.shape)

(128, 1876)

for idx, row in train_df.iterrows():
    file_path = os.path.join(TRAIN_AUDIO_DIR, row['filename'])
    if not os.path.exists(file_path):
        print(f"File not found: {file_path}")
        continue # Skip to the next file

    features = extract_features(file_path)
    if features is not None:
        features_list.append(features)
        labels.append(row['label'])

import os
import numpy as np
import pandas as pd
import librosa

# Load CSV
train_df = pd.read_csv("C:/Users/dubey/Downloads/shl-intern-hiring-
assessment (1)/dataset/train.csv")

# Audio directory
TRAIN_AUDIO_DIR = 'C:/Users/dubey/Downloads/shl-intern-hiring-
assessment (1)/dataset/audios_train'

# Feature extractor function
def extract_features(file_path):
    try:
        y, sr = librosa.load(file_path, sr=None)
        mfccs = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=20)
        mfccs_mean = np.mean(mfccs.T, axis=0)
        return mfccs_mean
    except Exception as e:
        print(f"Error processing {file_path}: {e}")
        return None

```

```

# Extract features for all audio files
features_list = []
labels = []

for idx, row in train_df.iterrows():
    file_name = row["filename"]
    label = row["label"]
    file_path = os.path.join(TRAIN_AUDIO_DIR, file_name)  # ☐ FIXED
    LINE

    if os.path.exists(file_path):
        features = extract_features(file_path)
        if features is not None:
            features_list.append(features)
            labels.append(label)
        else:
            print(f"⚠ Failed to extract features from: {file_path}")
    else:
        print(f"☐ File not found: {file_path}")

# Convert to NumPy arrays
X = np.array(features_list)
y = np.array(labels)

print("☐ Feature extraction complete.")
print("Extracted features shape:", X.shape)
print("Labels shape:", y.shape)

☐ Feature extraction complete.
Extracted features shape: (444, 20)
Labels shape: (444,)

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error
import matplotlib.pyplot as plt
import joblib

# Train-validation split
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2,
random_state=42)

# Train model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Predict
y_pred = model.predict(X_val)

# Evaluate

```



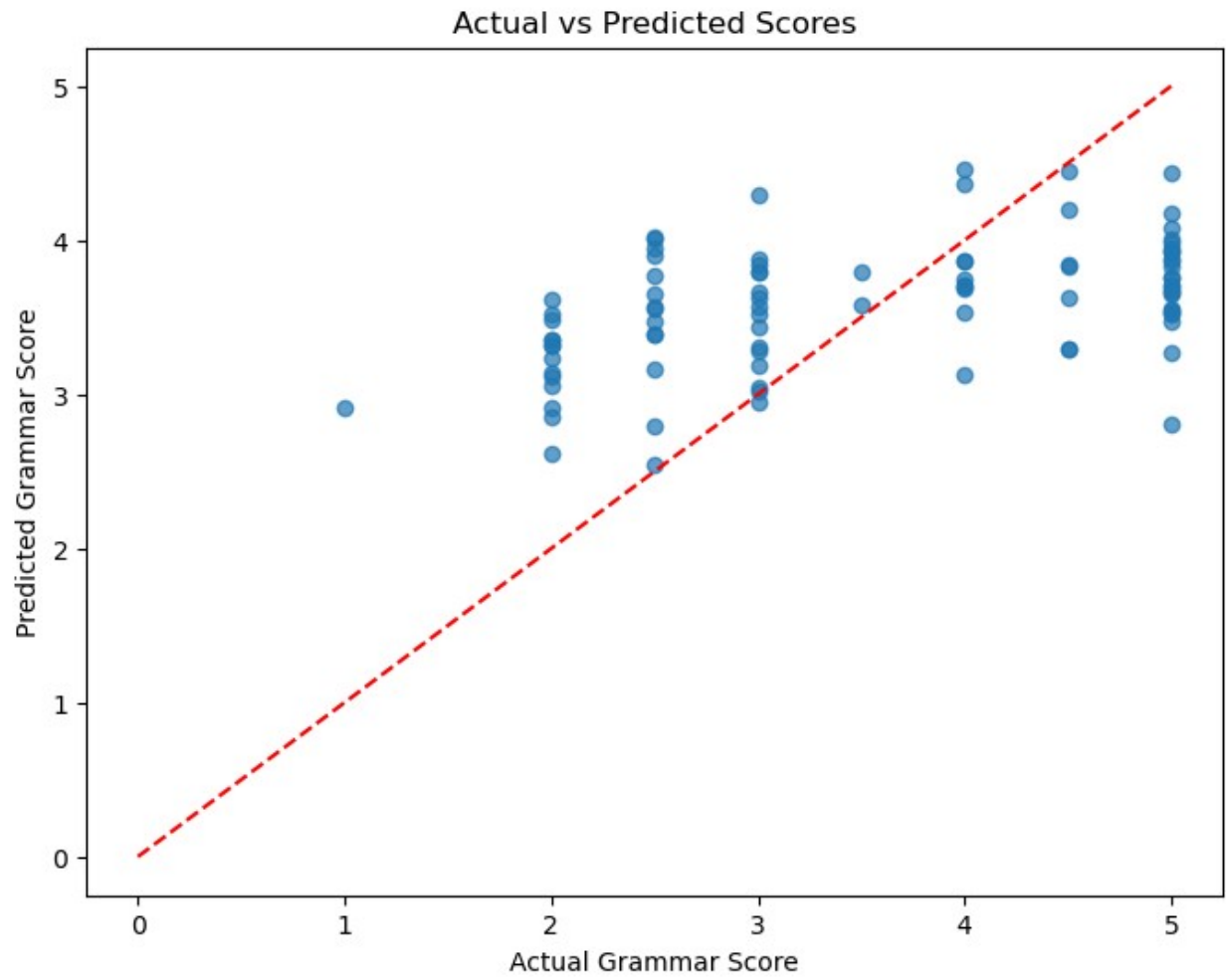
```
mse = mean_squared_error(y_val, y_pred)
mae = mean_absolute_error(y_val, y_pred)

print("Validation MSE:", mse)
print("Validation MAE:", mae)

# Visualization
plt.figure(figsize=(8, 6))
plt.scatter(y_val, y_pred, alpha=0.7)
plt.xlabel("Actual Grammar Score")
plt.ylabel("Predicted Grammar Score")
plt.title("Actual vs Predicted Scores")
plt.plot([0, 5], [0, 5], 'r--')
plt.show()

# Save model
joblib.dump(model, "grammar_scoring_model.pkl")
print("✅ Model saved as grammar_scoring_model.pkl")

Validation MSE: 1.0830837078651685
Validation MAE: 0.9166292134831462
```



□ Model saved as grammar_scoring_model.pkl