```
In [3]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.feature_selection import SelectKBest, f_regression
# Load the dataset
df = pd.read_csv('nba_team_stats_playoffs_00_to_21.csv')
# Display the first few rows of the dataset and column names
print(df.head())
print(df.columns)
# Check for missing values
print(df.isnull().sum())
# Inspect the column names to identify non-numeric columns
print(df.columns)
# Ensure column names are correctly identified
columns_to_drop = ['team']
for col in df.columns:
    if 'season' in col.lower():
         columns_to_drop.append(col)
print(f"Columns to drop: {columns_to_drop}")
# Drop non-numeric columns if they exist (adjust as necessary based on actual column names)
df = df.drop(columns=columns_to_drop)
# If there are missing values, handle them (e.g., fill with mean, median, or drop)
df = df.fillna(df.mean())
# Verify all columns are numeric now
print(df.dtypes)
\# Define the features (X) and the target variable (y)
X = df.drop(columns=['win_percentage', 'wins', 'losses', 'plus_minus', 'games_played', 'minutes']) # Adjust based on your dataset's column names
y = df['win_percentage']
# Split the data 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)
# Normalize/scale the data
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
# Feature selection using correlation
correlation_matrix = df.corr()
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
# Feature selection using SelectKBest
selector = SelectKBest(score_func=f_regression, k='all')
selector.fit(X_train_scaled, y_train)
feature_scores = pd.DataFrame({'Feature': X.columns, 'Score': selector.scores_})
feature_scores = feature_scores.sort_values(by='Score', ascending=False)
print(feature_scores)
# Fit a linear regression model
lr = LinearRegression()
lr.fit(X_train_scaled, y_train)
y_pred_lr = lr.predict(X_test_scaled)
# Evaluate the linear regression model
print('Linear Regression RMSE:', np.sqrt(mean_squared_error(y_test, y_pred_lr)))
print('Linear Regression R^2:', r2_score(y_test, y_pred_lr))
# Fit a random forest regressor model
rf = RandomForestRegressor(n_estimators=100, random_state=42)
rf.fit(X_train_scaled, y_train)
y_pred_rf = rf.predict(X_test_scaled)
# Evaluate the random forest model
print('Random Forest RMSE:', np.sqrt(mean_squared_error(y_test, y_pred_rf)))
print('Random Forest R^2:', r2_score(y_test, y_pred_rf))
# Feature importance from random forest
feature_importances = pd.DataFrame({'Feature': X.columns, 'Importance': rf.feature_importances_})
feature_importances = feature_importances.sort_values(by='Importance', ascending=False)
print(feature_importances)
# Plot feature importance
plt.figure(figsize=(12, 8))
sns.barplot(x='Importance', y='Feature', data=feature_importances)
plt.title('Feature Importance from Random Forest')
plt.show()
   teamstatspk
                                 team games_played wins losses \
0
              0
                     Milwaukee Bucks
                                                  23
                                                         16
                                                                   7
1
              1
                        Phoenix Suns
                                                  22
                                                         14
                                                                   8
2
                       Brooklyn Nets
                                                  12
                                                          7
                                                                   5
              3 Philadelphia 76ers
                                                  12
                                                          7
                                                                   5
3
                       Atlanta Hawks
                                                  18
                                                         10
4
   win_percentage minutes points field_goals_made field_goals_attempted \
0
             0.696
                        48.4
                               110.3
                                                     42.1
                                                                              91.1
1
             0.636
                        48.0
                               109.0
                                                     40.9
                                                                              85.0
                                                     40.7
2
             0.583
                        48.4
                              112.5
                                                                              86.2
3
             0.583
                        48.0
                              116.3
                                                     42.5
                                                                              85.8
4
                                                                              86.8
             0.556
                              106.3
                                                     38.9
                        48.0
        rebounds assists turnovers steals blocks blocks_attempted \
0
             49.0
                       22.8
                                   12.7
                                             7.8
                                                      4.2
                                                                         3.8
1
             42.7
                       23.2
                                   11.9
                                             6.5
                                                      4.2
                                                                         3.6
  . . .
2
             42.6
                       22.6
                                   11.6
                                             7.1
                                                      4.8
                                                                         6.2
  . . .
             44.7
                       24.9
                                   12.8
                                             8.0
                                                      6.2
                                                                          3.8
  . . .
                       20.2
                                   12.4
                                             6.7
                                                                          4.3
             42.4
  . . .
   personal_fouls personal_fouls_drawn plus_minus
                                                           season
0
              18.0
                                      20.1
                                                     5.1 2020-21
1
              19.7
                                      18.5
                                                     4.5 2020-21
2
              21.2
                                      18.6
                                                     6.3 2020-21
3
              23.0
                                      23.8
                                                    7.5 2020-21
4
              21.2
                                      20.3
                                                    -1.5 2020-21
[5 rows x 29 columns]
Index(['teamstatspk', 'team', 'games_played', 'wins', 'losses',
        'win_percentage', 'minutes', 'points', 'field_goals_made',
        'field_goals_attempted', 'field_goal_percentage', 'three_pointers_made',
        'three_pointers_attempted', 'three_point_percentage',
       'free_throws_made', 'free_throw_attempted', 'free_throw_percentage', 'offensive_rebounds', 'defensive_rebounds', 'rebounds', 'assists', 'turnovers', 'steals', 'blocks', 'blocks_attempted', 'personal_fouls',
        'personal_fouls_drawn', 'plus_minus', 'season'],
      dtype='object')
teamstatspk
                               0
                               0
team
                               0
games_played
wins
losses
                              0
win_percentage
                              0
                              0
minutes
                              0
points
field_goals_made
field_goals_attempted
field_goal_percentage
three_pointers_made
three_pointers_attempted
three_point_percentage
free_throws_made
free_throw_attempted
                               0
free_throw_percentage
offensive_rebounds
defensive_rebounds
rebounds
                               0
assists
turnovers
steals
blocks
blocks_attempted
personal_fouls
personal_fouls_drawn
plus_minus
season
dtype: int64
'field_goals_attempted', 'field_goal_percentage', 'three_pointers_made',
        'three_pointers_attempted', 'three_point_percentage',
       'free_throws_made', 'free_throw_attempted', 'free_throw_percentage', 'offensive_rebounds', 'defensive_rebounds', 'rebounds', 'assists',
        'turnovers', 'steals', 'blocks', 'blocks_attempted', 'personal_fouls',
        'personal_fouls_drawn', 'plus_minus', 'season'],
      dtype='object')
Columns to drop: ['team', 'season']
teamstatspk
games_played
                                 int64
                                 int64
wins
losses
                                 int64
                               float64
win_percentage
minutes
                               float64
                               float64
points
                               float64
field_goals_made
                               float64
field_goals_attempted
field_goal_percentage
                               float64
three_pointers_made
                               float64
                               float64
three_pointers_attempted
                               float64
three_point_percentage
free_throws_made
                               float64
                               float64
free_throw_attempted
                               float64
free_throw_percentage
                               float64
offensive_rebounds
defensive_rebounds
                               float64
                               float64
rebounds
                               float64
assists
turnovers
                               float64
steals
                               float64
blocks
                               float64
blocks_attempted
                               float64
personal_fouls
                               float64
personal_fouls_drawn
                               float64
plus_minus
                               float64
dtype: object
                                                                      Correlation Matrix
                                                                                                                                                 1.0
              teamstatspk - 1-0.0603.0601.0705.0401.02-0.560.510.4-0.220.730.770.120.20.270.150.340.48-0.2-0.340.190.0007.0910.040.410.60.014
            games played -0.06 1 0.980.820.820.080.250.2-D.070.390.160.09 8.280.1 D.090.070.0420.2 0.150.220.130.160.240.260.10.05 3.71
                      wins -0.060.98 1 0.680.870.0680.270.230.0740.420.170.0940.3 0.120.110.058.040.240.180.250.110.180.250.260.180.040.76
                                                                                                                                                - 0.8
                    losses -0.070.820.68 1 0.480.0940.140.1-10.046.21 0.10.0670.140.06050302.0960.03050507.0265.0980.160.0920.150.180.102.0760.41
           win_percentage -0.040.820.870.48 1 0.11 0.3 0.250.08 6.450.170.07 20.370.180.140.150.03 0.280.220.240.09 9.170.280.240.16.03 0.86
                  minutes 9.0240.080.0680940.11 1 0.140.140.160.020900494001600104120.110.03080330.180.170.0509.02060150.130.0870.165.0045096
                                                                                                                                                - 0.6
                    points -0.5 (0.250.270.14 0.3 0.14 1 0.9 0.66 0.6 0.7 30.67 0.4 30.2 40.1 60.2 30.0 5 (3.4 90.3 80.5 8 -0.2 0.0 7 (3.0 6 9.0 80) 0.2 (7.3 10.3 6 9.0 80)
         field goals made -0.510.210.230.110.250.14 0.9 1 0.730.670.580.52 0.40.09 0.120.090.02 0.440.360.62 0.280.080.064.060207 0.240.31
    field_goals_attempted -0.40.070L07040405.088.160.660.73 1-0.0210.50.510.0140.190.130.280.460.570.490.250.170.0620.240.070.120.051
                                                                                                                                                - 0.4
     field_goal_percentage -0.220.390.420.210.450.0290.60.6-0.0211 0.30.150.560.0102030.02-6.340.13-0.10.360.1-0.00060270.3-0.026.230.51
     three_pointers_made -0.7 0.160.17 0.1 0.107.00 0.730.58 0.5 0.3 1 0.960.470.130.210.210.350.49 0.2 0.460.0906020.06 0.150.260.350.18
three_pointers_attempted -0.70.090.0904.0607.070.010.670.520.550.150.96 1 0.220.12-0.2 0.2 -0.30.510.25 0.40.0701030.0770.1-0.270.380.067
                                                                                                                                                - 0.2
   three_point_percentage -0.120.28 0.3 0.140.87000 D443 0.40.0180.560.470.22 1 0.080.10.0940.290.150.060.320.090203020320.190.080.0350.41
        free_throws_made -0.2 0.110.120.0650.180.120.240.0920.140.0120.130.120.08 1 0.920.260.240.00603140.086.120.0220.120.0240.0340.180
    free_throw_attempted -0.270.0930.110.0320.140.110.160.120.19.0340.21-0.2-0.120.92 1 -0.150.280.030.150.0870.170.0610.140.0140.3190.0220.16
                                                                                                                                                - 0.0
    free_throw_percentage -0.16.0703.05020950.110.0380.230.090.130.026.210.20.0940.260.15 1 0.10.0749.0101.01-10.1-10.0903.040.709050340.096
       offensive_rebounds -0.340.0402.0401.0305.0301.0305.0508.0201.280.340.35-0.3-0.290.240.280.12 1 -0.120.530.05060690.120.230.290.190.26.012
      defensive_rebounds -0.48 0.2 0.24 0.57 0.28 0.18 0.49 0.44 0.46 0.13 0.49 0.51 0.16 0.06 0.07 0.12 1 0.78 0.4 0.10 0.02 0.17 0.05 0.27 0.17 0.29
                                                                                                                                                - -0.2
                 rebounds --0.20.150.180.0260.220.170.380.360.57-0.1 0.2 0.250.060.140.150.010.530.78 1 0.310.130.0530.290.130.1-0.024.26
                    assists -0.340.220.250.0980.240.0590.580.620.490.360.46 0.4 0.320.0806.08070101.0560.4 0.31 1 0.0830.20.056.0304.04060570.28
                 turnovers -0.190.130.110.1-15.09090260.2-0.280.250.1-40.0906.0701.0912.130.170.101.0690.10.1-30.08 1 0.112.0707.0210.240.150.18
                    steals 9.0070.160.180.0920.170.0105.0720.080.170.0606020.0303.030202020601.0909.120.02060550.2 0.12 1 0.140.060.0720.130.2
                                                                                                                                                 -0.4
                    blocks 9.09 D.240.250.150.280.130.0600.0604.0602.0207.0601.07070320.120.140.0407.230.170.290.0505.0770.14 1 0.040500402.0907.31
        blocks_attempted -0.040.260.260.180.260.0807.0809.0602.260.370.150.1-0.10.0204.010400307.240.0507.1-30.03040210.060.045110.0709.07-30.27
            personal_fouls -0.410.170.180.120.180.150.0207.0705.0702.02-6.260.270.080.340.340.096.190.270.1-10.046.240.07020040207911 0.260.14
                                                                                                                                                 -0.6
     personal_fouls_drawn -0.60.0530.040.0766.030200405310.240.120.230.350.380.0305.0304.0202.0340.260.170.02040570.150.170.0907.07-30.26 1 0.018
               plus_minus-0.010.710.760.410.860.0960.360.3-10.0500.510.180.0670.410.180.160.09060170.290.260.280.180.20.310.270.160.018 1
                                                                                          offensive_rebounds
                                                                       three_pointers_attempted
                                                                               free_throws_made
                                                                                   free_throw_attempted
                                                                                      free_throw_percentage
                                            win_percentage
                                                           field_goals_attempted
                                                               field_goal_percentage
                                                                                               defensive_rebounds
                                                                   three_pointers_made
                                                                                                                              personal_fouls_drawn
                             teamstatspk
                                games_played
                                                                           three_point_percentage
                                                                                                          turnovers
                                                                                                                      blocks_attempted
                                                                                                                          personal_fouls
                                                                                                                                  plus_minus
                       Feature
       field_goal_percentage 77.738009
4
7
      three_point_percentage 47.724431
1
                        points 31.978178
12
           defensive_rebounds 27.939550
17
                        blocks 22.805380
2
             field_goals_made 19.982412
14
                       assists 19.566528
                      rebounds 17.939062
13
18
             blocks_attempted 13.420696
19
               personal_fouls 12.217235
5
          three_pointers_made 11.393331
8
             free_throws_made
                                 9.368876
                                  9.128651
16
                        steals
9
         free_throw_attempted
                                  5.200534
15
                     turnovers
                                  4.841836
10
       free_throw_percentage
                                 4.468435
6
    three_pointers_attempted
                                 3.036432
0
                                 1.193557
                  teamstatspk
       field_goals_attempted
3
                                 1.075042
20
        personal_fouls_drawn 0.441941
11
          offensive_rebounds
                                 0.172470
Linear Regression RMSE: 0.15658768740262988
Linear Regression R^2: 0.35049088915467974
Random Forest RMSE: 0.15493914202793602
Random Forest R^2: 0.36409488176389615
                       Feature Importance
4
       field_goal_percentage
                                   0.191815
7
      three_point_percentage
                                   0.094311
16
                                   0.080896
                        steals
17
                                   0.077771
                        blocks
                      rebounds
13
                                   0.067683
12
           defensive_rebounds
                                   0.059681
               personal_fouls
19
                                   0.050066
       field_goals_attempted
3
                                   0.043242
       free_throw_percentage
10
                                   0.040166
                     turnovers
15
                                   0.038041
8
             free_throws_made
                                   0.032832
           offensive_rebounds
                                   0.029105
11
18
            blocks_attempted
                                   0.028215
1
                        points
                                   0.027338
         personal_fouls_drawn
20
                                   0.025479
14
                       assists
                                   0.025396
9
         free_throw_attempted
                                   0.025227
2
             field_goals_made
                                   0.019342
0
                  teamstatspk
                                   0.017644
    three_pointers_attempted
6
                                   0.013633
          three_pointers_made
5
                                   0.012119
                                                                       Feature Importance from Random Forest
       field_goal_percentage
     three_point_percentage
                       steals
                      blocks
                    rebounds
         defensive_rebounds
               personal_fouls
       field_goals_attempted
      free_throw_percentage
                   turnovers
 Feature
           free_throws_made
```

offensive_rebounds

blocks_attempted

personal_fouls_drawn

free throw attempted

three_pointers_attempted

three_pointers_made

field_goals_made

teamstatspk

points

assists

0.000

0.025

0.050

0.075

0.100 Importance 0.125

0.150

0.175

0.200