In [4]:
```
import sqlite3
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


db_path = "/home/satvik/Downloads/database.sqlite"
conn = sqlite3.connect(db_path)
cursor = conn.cursor()


def get_table_names():
    cursor.execute("SELECT name FROM sqlite_master WHERE type='table';")
    return [table[0] for table in cursor.fetchall() if table[0] != "sqlit

tables = get_table_names()
print("Tables in Database:", tables)


def get_table_info(table_name):
    cursor.execute(f"PRAGMA table_info({table_name});")
    return [(col[1], col[2]) for col in cursor.fetchall()]

table_info = {table: get_table_info(table) for table in tables}
for table, columns in table_info.items():
    print(f"\nTable: {table}")
    for col_name, col_type in columns:
        print(f"  {col_name} ({col_type})")


def check_data_quality(table_name):
    df = pd.read_sql_query(f"SELECT * FROM {table_name} LIMIT 1000", conn
    missing_values = df.isnull().sum().sum()
    duplicate_rows = df.duplicated().sum()
    return {"Missing Values": missing_values, "Duplicate Rows": duplicate

data_quality = {table: check_data_quality(table) for table in tables}
print("\nData Quality Report:")
for table, report in data_quality.items():
    print(f"{table}: {report}")



match_df = pd.read_sql_query("SELECT home_team_goal, away_team_goal FROM
match_df['goal_diff'] = match_df['home_team_goal'] - match_df['away_team_

plt.figure(figsize=(8,5))
sns.histplot(match_df['goal_diff'], bins=20, kde=True)
plt.title("Distribution of Home Team Goal Difference")
plt.xlabel("Goal Difference (Home - Away)")
plt.ylabel("Match Count")
plt.show()


player_attr_df = pd.read_sql_query("SELECT overall_rating FROM Player_Att

plt.figure(figsize=(8,5))
sns.histplot(player_attr_df['overall_rating'], bins=20, kde=True)
```

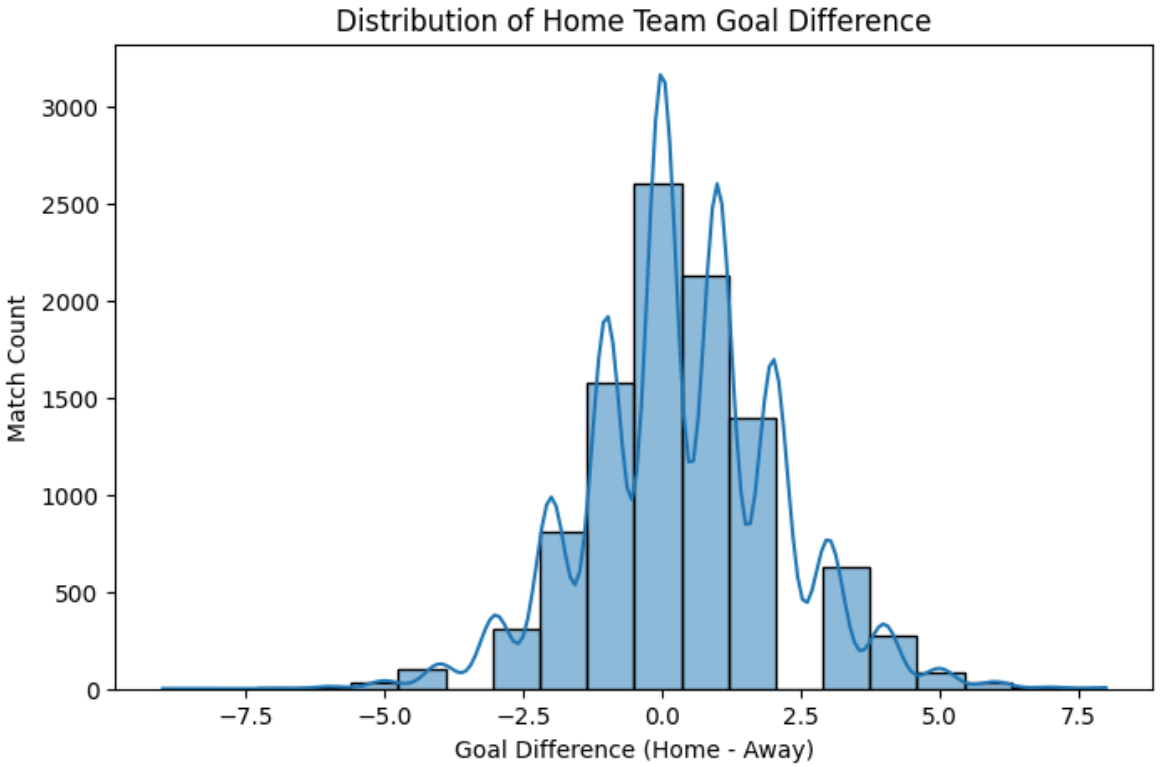
```
plt.title("Distribution of Player Overall Ratings")
plt.xlabel("Overall Rating")
plt.ylabel("Frequency")
plt.show()


team_performance_df = pd.read_sql_query("""
    SELECT season, AVG(home_team_goal + away_team_goal) as avg_goals
    FROM Match GROUP BY season
""", conn)

plt.figure(figsize=(10,5))
sns.lineplot(x=team_performance_df['season'], y=team_performance_df['avg_
plt.title("Average Goals Per Match Over Seasons")
plt.xlabel("Season")
plt.ylabel("Average Goals Per Match")
plt.xticks(rotation=45)
plt.show()


conn.close()
```

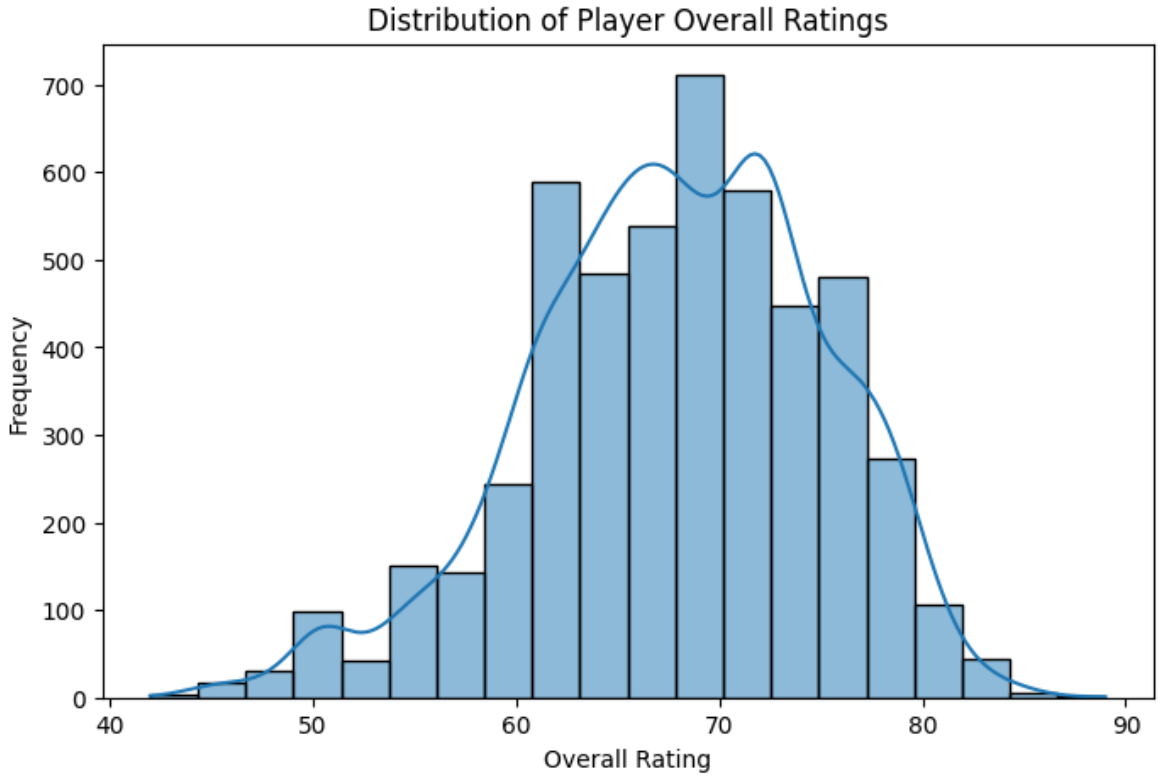
```
Tables in Database: ['Player_Attributes', 'Player', 'Match', 'League', 'Co
untry', 'Team', 'Team_Attributes']

Table: Player_Attributes
  id (INTEGER)
  player_fifa_api_id (INTEGER)
  player_api_id (INTEGER)
  date (TEXT)
  overall_rating (INTEGER)
  potential (INTEGER)
  preferred_foot (TEXT)
  attacking_work_rate (TEXT)
  defensive_work_rate (TEXT)
  crossing (INTEGER)
  finishing (INTEGER)
  heading_accuracy (INTEGER)
  short_passing (INTEGER)
  volleys (INTEGER)
  dribbling (INTEGER)
  curve (INTEGER)
  free_kick_accuracy (INTEGER)
  long_passing (INTEGER)
  ball_control (INTEGER)
  acceleration (INTEGER)
  sprint_speed (INTEGER)
  agility (INTEGER)
  reactions (INTEGER)
  balance (INTEGER)
  shot_power (INTEGER)
  jumping (INTEGER)
  stamina (INTEGER)
  strength (INTEGER)
  long_shots (INTEGER)
  aggression (INTEGER)
  interceptions (INTEGER)
  positioning (INTEGER)
  vision (INTEGER)
  penalties (INTEGER)
  marking (INTEGER)
  standing_tackle (INTEGER)
  sliding_tackle (INTEGER)
  gk_diving (INTEGER)
  gk_handling (INTEGER)
  gk_kicking (INTEGER)
  gk_positioning (INTEGER)
  gk_reflexes (INTEGER)

Table: Player
  id (INTEGER)
  player_api_id (INTEGER)
  player_name (TEXT)
  player_fifa_api_id (INTEGER)
  birthday (TEXT)
  height (INTEGER)
  weight (INTEGER)

Table: Match
  id (INTEGER)
  country_id (INTEGER)
  league_id (INTEGER)
```

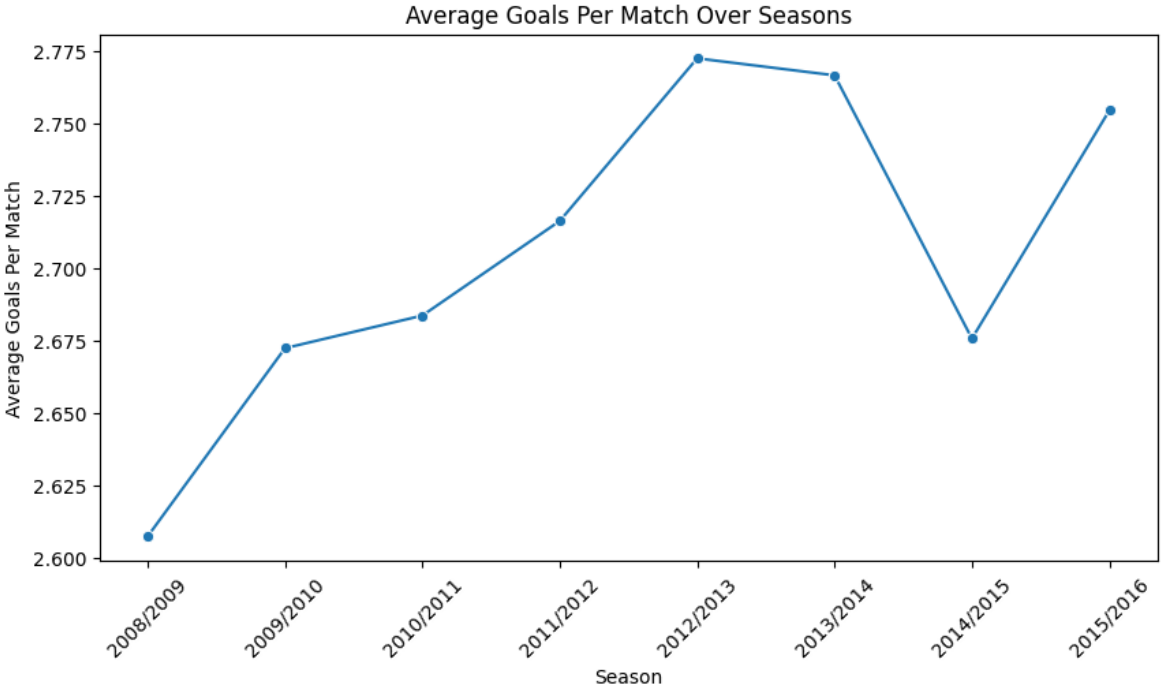
```
season (TEXT)
stage (INTEGER)
date (TEXT)
match_api_id (INTEGER)
home_team_api_id (INTEGER)
away_team_api_id (INTEGER)
home_team_goal (INTEGER)
away_team_goal (INTEGER)
home_player_X1 (INTEGER)
home_player_X2 (INTEGER)
home_player_X3 (INTEGER)
home_player_X4 (INTEGER)
home_player_X5 (INTEGER)
home_player_X6 (INTEGER)
home_player_X7 (INTEGER)
home_player_X8 (INTEGER)
home_player_X9 (INTEGER)
home_player_X10 (INTEGER)
home_player_X11 (INTEGER)
away_player_X1 (INTEGER)
away_player_X2 (INTEGER)
away_player_X3 (INTEGER)
away_player_X4 (INTEGER)
away_player_X5 (INTEGER)
away_player_X6 (INTEGER)
away_player_X7 (INTEGER)
away_player_X8 (INTEGER)
away_player_X9 (INTEGER)
away_player_X10 (INTEGER)
away_player_X11 (INTEGER)
home_player_Y1 (INTEGER)
home_player_Y2 (INTEGER)
home_player_Y3 (INTEGER)
home_player_Y4 (INTEGER)
home_player_Y5 (INTEGER)
home_player_Y6 (INTEGER)
home_player_Y7 (INTEGER)
home_player_Y8 (INTEGER)
home_player_Y9 (INTEGER)
home_player_Y10 (INTEGER)
home_player_Y11 (INTEGER)
away_player_Y1 (INTEGER)
away_player_Y2 (INTEGER)
away_player_Y3 (INTEGER)
away_player_Y4 (INTEGER)
away_player_Y5 (INTEGER)
away_player_Y6 (INTEGER)
away_player_Y7 (INTEGER)
away_player_Y8 (INTEGER)
away_player_Y9 (INTEGER)
away_player_Y10 (INTEGER)
away_player_Y11 (INTEGER)
home_player_1 (INTEGER)
home_player_2 (INTEGER)
home_player_3 (INTEGER)
home_player_4 (INTEGER)
home_player_5 (INTEGER)
home_player_6 (INTEGER)
home_player_7 (INTEGER)
home_player_8 (INTEGER)
```

```
  home_player_9 (INTEGER)
  home_player_10 (INTEGER)
  home_player_11 (INTEGER)
  away_player_1 (INTEGER)
  away_player_2 (INTEGER)
  away_player_3 (INTEGER)
  away_player_4 (INTEGER)
  away_player_5 (INTEGER)
  away_player_6 (INTEGER)
  away_player_7 (INTEGER)
  away_player_8 (INTEGER)
  away_player_9 (INTEGER)
  away_player_10 (INTEGER)
  away_player_11 (INTEGER)
  goal (TEXT)
  shoton (TEXT)
  shotoff (TEXT)
  foulcommit (TEXT)
  card (TEXT)
  cross (TEXT)
  corner (TEXT)
  possession (TEXT)
  B365H (NUMERIC)
  B365D (NUMERIC)
  B365A (NUMERIC)
  BWH (NUMERIC)
  BWD (NUMERIC)
  BWA (NUMERIC)
  IWH (NUMERIC)
  IWD (NUMERIC)
  IWA (NUMERIC)
  LBH (NUMERIC)
  LBD (NUMERIC)
  LBA (NUMERIC)
  PSH (NUMERIC)
  PSD (NUMERIC)
  PSA (NUMERIC)
  WHH (NUMERIC)
  WHD (NUMERIC)
  WHA (NUMERIC)
  SJH (NUMERIC)
  SJD (NUMERIC)
  SJA (NUMERIC)
  VCH (NUMERIC)
  VCD (NUMERIC)
  VCA (NUMERIC)
  GBH (NUMERIC)
  GBD (NUMERIC)
  GBA (NUMERIC)
  BSH (NUMERIC)
  BSD (NUMERIC)
  BSA (NUMERIC)

Table: League
  id (INTEGER)
  country_id (INTEGER)
  name (TEXT)

Table: Country
  id (INTEGER)
```

```
  name (TEXT)

Table: Team
  id (INTEGER)
  team_api_id (INTEGER)
  team_fifa_api_id (INTEGER)
  team_long_name (TEXT)
  team_short_name (TEXT)

Table: Team_Attributes
  id (INTEGER)
  team_fifa_api_id (INTEGER)
  team_api_id (INTEGER)
  date (TEXT)
  buildUpPlaySpeed (INTEGER)
  buildUpPlaySpeedClass (TEXT)
  buildUpPlayDribbling (INTEGER)
  buildUpPlayDribblingClass (TEXT)
  buildUpPlayPassing (INTEGER)
  buildUpPlayPassingClass (TEXT)
  buildUpPlayPositioningClass (TEXT)
  chanceCreationPassing (INTEGER)
  chanceCreationPassingClass (TEXT)
  chanceCreationCrossing (INTEGER)
  chanceCreationCrossingClass (TEXT)
  chanceCreationShooting (INTEGER)
  chanceCreationShootingClass (TEXT)
  chanceCreationPositioningClass (TEXT)
  defencePressure (INTEGER)
  defencePressureClass (TEXT)
  defenceAggression (INTEGER)
  defenceAggressionClass (TEXT)
  defenceTeamWidth (INTEGER)
  defenceTeamWidthClass (TEXT)
  defenceDefenderLineClass (TEXT)

Data Quality Report:
Player_Attributes: {'Missing Values': 216, 'Duplicate Rows': 0}
Player: {'Missing Values': 0, 'Duplicate Rows': 0}
Match: {'Missing Values': 25348, 'Duplicate Rows': 0}
League: {'Missing Values': 0, 'Duplicate Rows': 0}
Country: {'Missing Values': 0, 'Duplicate Rows': 0}
Team: {'Missing Values': 11, 'Duplicate Rows': 0}
Team_Attributes: {'Missing Values': 658, 'Duplicate Rows': 0}
```


Distribution of Home Team Goal Difference
Match Count
Goal Difference (Home - Away)
3000
2500
2000
1500
1000
500
0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5


Distribution of Player Overall Ratings
Frequency
Overall Rating
700
600
500
400
300
200
100
0
40
50
60
70
80
90


Average Goals Per Match Over Seasons
Average Goals Per Match
2.775
2.750
2.725
2.700
2.675
2.650
2.625
2.600
2008/2009
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
Season

In [ ]:

In [ ]: