Investigate_a_Soccer Dataset

September 15, 2021

1 Project: Soccer Data Analysis

1.1 Table of Contents

Introduction

Data Wrangling

Exploratory Data Analysis

Conclusions

Introduction

This soccer database comes from Kaggle. It contains data for soccer matches, players, and teams from several European countries from 2008 to 2016. The database covers 24,637 matches and over 10,000 players. Detailed match events like goal types, possession, corner, cross, fouls, cards are also recorded in the database. Moreover, this dataset integrated players and teams attribute ratings sourced from EA Sports' FIFA video game series.

We're most interested in figuring out if a player's potential for assisting goal is high, his potential for a goal is also high, and which club team improved the most over the period, and each related ranking associated with Tottenham Hotspur.

Data Wrangling

The dataset is a SQL database file containing seven tables of "Country," "League," "Match," "Player," "Play-er_Attribute," "Team," "Team_Attributes." We downloaded the SQL file from the Kaggle repository and extracted the CSV files from the DB browser application to get the data source. For the research, we used four tables formatted as CSV: `Match, Team, Player,` and `Player_Attributes.`

First, we sorted the `Match` table with required columns, extracted data, and the names of players who recorded goals or assists and then merged it with each of the other CSV files after renaming the ID columns. We add new columns for the winner of the match and goal point by net goals and finally get 7173 lines of data with no null values to clear the questions.

1.1.1 General Properties

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
```

```
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)

%matplotlib inline
```

[2]:
```
# see the entire columns

df = pd.read_csv('Match.csv')
pd.set_option('display.max_columns', None)
```

[3]:
```
df.head()
```

[3]:

	id	country_id	league_id	season	stage	date	match_api_id	home_team_api_id	away_team_api_id	home_team_goal	away_team_goal	home_player_X1	home_player_X2	home_player_X3	home_player_X4	home_player_X5	home_player_X6	home_player_X7	home_player_X8	home_player_X9	home_player_X10	home_player_X11
0	1	1	1	2008/2009	1	2008-08-17 00:00:00	492473	9987	9993	1	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2	1	1	2008/2009	1	2008-08-16 00:00:00	492474	10000	9994	0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	3	1	1	2008/2009	1	2008-08-16 00:00:00	492475	9984	8635	0	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	4	1	1	2008/2009	1	2008-08-17 00:00:00	492476	9991	9998	5	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	5	1	1	2008/2009	1	2008-08-16 00:00:00	492477	7947	9985	1	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
   away_player_X1  away_player_X2  away_player_X3  away_player_X4  \
0             NaN             NaN             NaN             NaN
1             NaN             NaN             NaN             NaN
2             NaN             NaN             NaN             NaN
3             NaN             NaN             NaN             NaN
4             NaN             NaN             NaN             NaN

   away_player_X5  away_player_X6  away_player_X7  away_player_X8  \
0             NaN             NaN             NaN             NaN
1             NaN             NaN             NaN             NaN
2             NaN             NaN             NaN             NaN
3             NaN             NaN             NaN             NaN
4             NaN             NaN             NaN             NaN

   away_player_X9  away_player_X10  away_player_X11  home_player_Y1  \
0             NaN              NaN              NaN             NaN
1             NaN              NaN              NaN             NaN
2             NaN              NaN              NaN             NaN
3             NaN              NaN              NaN             NaN
4             NaN              NaN              NaN             NaN

   home_player_Y2  home_player_Y3  home_player_Y4  home_player_Y5  \
0             NaN             NaN             NaN             NaN
1             NaN             NaN             NaN             NaN
2             NaN             NaN             NaN             NaN
3             NaN             NaN             NaN             NaN
4             NaN             NaN             NaN             NaN

   home_player_Y6  home_player_Y7  home_player_Y8  home_player_Y9  \
0             NaN             NaN             NaN             NaN
1             NaN             NaN             NaN             NaN
2             NaN             NaN             NaN             NaN
3             NaN             NaN             NaN             NaN
4             NaN             NaN             NaN             NaN

   home_player_Y10  home_player_Y11  away_player_Y1  away_player_Y2  \
0              NaN              NaN             NaN             NaN
1              NaN              NaN             NaN             NaN
2              NaN              NaN             NaN             NaN
3              NaN              NaN             NaN             NaN
4              NaN              NaN             NaN             NaN

   away_player_Y3  away_player_Y4  away_player_Y5  away_player_Y6  \
0             NaN             NaN             NaN             NaN
1             NaN             NaN             NaN             NaN
2             NaN             NaN             NaN             NaN
3             NaN             NaN             NaN             NaN
```

```
4             NaN             NaN             NaN             NaN

   away_player_Y7  away_player_Y8  away_player_Y9  away_player_Y10  \
0             NaN             NaN             NaN              NaN
1             NaN             NaN             NaN              NaN
2             NaN             NaN             NaN              NaN
3             NaN             NaN             NaN              NaN
4             NaN             NaN             NaN              NaN

   away_player_Y11  home_player_1  home_player_2  home_player_3  \
0              NaN            NaN            NaN            NaN
1              NaN            NaN            NaN            NaN
2              NaN            NaN            NaN            NaN
3              NaN            NaN            NaN            NaN
4              NaN            NaN            NaN            NaN

   home_player_4  home_player_5  home_player_6  home_player_7  home_player_8  \
0            NaN            NaN            NaN            NaN            NaN
1            NaN            NaN            NaN            NaN            NaN
2            NaN            NaN            NaN            NaN            NaN
3            NaN            NaN            NaN            NaN            NaN
4            NaN            NaN            NaN            NaN            NaN

   home_player_9  home_player_10  home_player_11  away_player_1  \
0            NaN             NaN             NaN            NaN
1            NaN             NaN             NaN            NaN
2            NaN             NaN             NaN            NaN
3            NaN             NaN             NaN            NaN
4            NaN             NaN             NaN            NaN

   away_player_2  away_player_3  away_player_4  away_player_5  away_player_6  \
0            NaN            NaN            NaN            NaN            NaN
1            NaN            NaN            NaN            NaN            NaN
2            NaN            NaN            NaN            NaN            NaN
3            NaN            NaN            NaN            NaN            NaN
4            NaN            NaN            NaN            NaN            NaN

   away_player_7  away_player_8  away_player_9  away_player_10  \
0            NaN            NaN            NaN             NaN
1            NaN            NaN            NaN             NaN
2            NaN            NaN            NaN             NaN
3            NaN            NaN            NaN             NaN
4            NaN            NaN            NaN             NaN

   away_player_11 goal shoton shotoff foulcommit card cross corner possession  \
0             NaN  NaN    NaN     NaN         NaN  NaN   NaN    NaN         NaN
1             NaN  NaN    NaN     NaN         NaN  NaN   NaN    NaN         NaN
```

```
2              NaN  NaN    NaN     NaN        NaN  NaN   NaN    NaN        NaN
3              NaN  NaN    NaN     NaN        NaN  NaN   NaN    NaN        NaN
4              NaN  NaN    NaN     NaN        NaN  NaN   NaN    NaN        NaN

   B365H  B365D  B365A   BWH   BWD   BWA   IWH  IWD  IWA   LBH  LBD   LBA  \
0   1.73   3.40   5.00  1.75  3.35  4.20  1.85  3.2  3.5  1.80  3.3  3.75
1   1.95   3.20   3.60  1.80  3.30  3.95  1.90  3.2  3.5  1.90  3.2  3.50
2   2.38   3.30   2.75  2.40  3.30  2.55  2.60  3.1  2.3  2.50  3.2  2.50
3   1.44   3.75   7.50  1.40  4.00  6.80  1.40  3.9  6.0  1.44  3.6  6.50
4   5.00   3.50   1.65  5.00  3.50  1.60  4.00  3.3  1.7  4.00  3.4  1.72

   PSH  PSD  PSA   WHH   WHD   WHA   SJH  SJD   SJA   VCH   VCD   VCA   GBH  \
0  NaN  NaN  NaN  1.70  3.30  4.33  1.90  3.3  4.00  1.65  3.40  4.50  1.78
1  NaN  NaN  NaN  1.83  3.30  3.60  1.95  3.3  3.80  2.00  3.25  3.25  1.85
2  NaN  NaN  NaN  2.50  3.25  2.40  2.63  3.3  2.50  2.35  3.25  2.65  2.50
3  NaN  NaN  NaN  1.44  3.75  6.00  1.44  4.0  7.50  1.45  3.75  6.50  1.50
4  NaN  NaN  NaN  4.20  3.40  1.70  4.50  3.5  1.73  4.50  3.40  1.65  4.50

    GBD   GBA   BSH   BSD   BSA
0  3.25  4.00  1.73  3.40  4.20
1  3.25  3.75  1.91  3.25  3.60
2  3.20  2.50  2.30  3.20  2.75
3  3.75  5.50  1.44  3.75  6.50
4  3.50  1.65  4.75  3.30  1.67
```

1.1.2 Data Cleaning

Based on the first data loaded in the previous cell, I will merge it with the other two data, which contains the player's attribute, birthday, height, weight, player's overall rating, potential, and team names. According to the initial interest, I will then extract age and names for players who recorded a goal or assist from the merged data. Lastly, I will calculate the goal point, which is home_team_goal - away_team_goal.

[4]:
```
df.shape
```

[4]:
```
(25979, 115)
```

[5]:
```
df = df[[
    'season', 'date', 'home_team_api_id', 'away_team_api_id', 'home_team_goal',
    'away_team_goal', 'home_player_1', 'home_player_2', 'home_player_3',
    'home_player_4', 'home_player_5', 'home_player_6', 'home_player_7',
    'home_player_8', 'home_player_9', 'home_player_10', 'home_player_11',
    'away_player_1', 'away_player_2', 'away_player_3', 'away_player_4',
    'away_player_5', 'away_player_6', 'away_player_7', 'away_player_8',
    'away_player_9', 'away_player_10', 'away_player_11', 'goal'
]]
```

[6]:
```
df.head()
```

[6]:

	season	date	home_team_api_id	away_team_api_id	\
0	2008/2009	2008-08-17 00:00:00	9987	9993	
1	2008/2009	2008-08-16 00:00:00	10000	9994	
2	2008/2009	2008-08-16 00:00:00	9984	8635	
3	2008/2009	2008-08-17 00:00:00	9991	9998	
4	2008/2009	2008-08-16 00:00:00	7947	9985	

	home_team_goal	away_team_goal	home_player_1	home_player_2	\
0	1	1	NaN	NaN	
1	0	0	NaN	NaN	
2	0	3	NaN	NaN	
3	5	0	NaN	NaN	
4	1	3	NaN	NaN	

	home_player_3	home_player_4	home_player_5	home_player_6	home_player_7	\
0	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	

	home_player_8	home_player_9	home_player_10	home_player_11	\
0	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	

	away_player_1	away_player_2	away_player_3	away_player_4	away_player_5	\
0	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	

	away_player_6	away_player_7	away_player_8	away_player_9	away_player_10	\
0	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	

	away_player_11	goal
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN

```
4             NaN  NaN
```

[7]:
```
df.dropna(inplace = True)
```

[8]:
```
df.shape
```

[8]:
```
(13325, 29)
```

[9]:
```
df = df.sort_values('date').reset_index(drop = True)
```

[10]:
```
df.head()
```

[10]:
```
      season                 date  home_team_api_id  away_team_api_id  \
0  2008/2009  2008-08-15 00:00:00              9823              9790
1  2008/2009  2008-08-16 00:00:00              9825              8659
2  2008/2009  2008-08-16 00:00:00              8472              8650
3  2008/2009  2008-08-16 00:00:00              8654              8528
4  2008/2009  2008-08-16 00:00:00              8668              8655

   home_team_goal  away_team_goal  home_player_1  home_player_2  \
0               2               2        27284.0        35988.0
1               1               0        23686.0        26111.0
2               0               1        32562.0        38836.0
3               2               1        36374.0        30966.0
4               2               3        31465.0        30371.0

   home_player_3  home_player_4  home_player_5  home_player_6  home_player_7  \
0        39774.0        33085.0        30894.0        38244.0        30872.0
1        38835.0        30986.0        31291.0        31013.0        30935.0
2        24446.0        24408.0        36786.0        38802.0        24655.0
3        23818.0        37277.0        30687.0        36394.0        37169.0
4        24004.0        33086.0        30857.0        24011.0       109058.0

   home_player_8  home_player_9  home_player_10  home_player_11  \
0        38843.0        95078.0         30638.0         32118.0
1        39297.0        26181.0         30960.0         36410.0
2        17866.0        30352.0         23927.0         24410.0
3        24223.0        24773.0         34543.0         23139.0
4        23268.0        24846.0         24006.0         24160.0

   away_player_1  away_player_2  away_player_3  away_player_4  away_player_5  \
0        25524.0        36183.0        27293.0        37787.0        37156.0
1        36373.0        36832.0        23115.0        37280.0        24728.0
2        30660.0        37442.0        30617.0        24134.0       414792.0
3        34421.0        34987.0        35472.0       111865.0        25005.0
4        30622.0        37764.0        19020.0        23921.0        24136.0
```

```
   away_player_6  away_player_7  away_player_8  away_player_9  away_player_10  \
0        30749.0        30598.0        39106.0        38216.0         33101.0
1        24664.0        31088.0        23257.0        24171.0         25922.0
2        37139.0        30618.0        40701.0        24800.0         24635.0
3        35327.0        25150.0        97988.0        41877.0        127857.0
4        30342.0        23889.0        23916.0        23922.0         34176.0

   away_player_11                                               goal
0         30764.0  <goal><value><comment>n</comment><stats><goals…
1         27267.0  <goal><value><comment>n</comment><stats><goals…
2         30853.0  <goal><value><comment>n</comment><stats><goals…
3         34466.0  <goal><value><comment>n</comment><stats><goals…
4         30646.0  <goal><value><comment>n</comment><stats><goals…
```

[11]:
```
# extract player_id who recorded goal or assist

df['goal_player_id'] = df.goal.str.extract('(1>[0-9]{5,6})', expand = True)
df['goal_player_id'] = df['goal_player_id'].astype(str).str[2:]
df['goal_player_id'] = df['goal_player_id'].replace('n', np.NaN)
df['goal_player_id'] = pd.to_numeric(df['goal_player_id'])

df['assist_player_id'] = df.goal.str.extract('(2>[0-9]{5,6})', expand = True)
df['assist_player_id'] = df['assist_player_id'].astype(str).str[2:]
df['assist_player_id'] = df['assist_player_id'].replace('n', np.NaN)
df['assist_player_id'] = pd.to_numeric(df['assist_player_id'])

df.drop(columns = ['goal'], inplace = True)
```

[12]:
```
df.head()
```

[12]:
```
      season                 date  home_team_api_id  away_team_api_id  \
0  2008/2009  2008-08-15 00:00:00              9823              9790
1  2008/2009  2008-08-16 00:00:00              9825              8659
2  2008/2009  2008-08-16 00:00:00              8472              8650
3  2008/2009  2008-08-16 00:00:00              8654              8528
4  2008/2009  2008-08-16 00:00:00              8668              8655

   home_team_goal  away_team_goal  home_player_1  home_player_2  \
0               2               2        27284.0        35988.0
1               1               0        23686.0        26111.0
2               0               1        32562.0        38836.0
3               2               1        36374.0        30966.0
4               2               3        31465.0        30371.0

   home_player_3  home_player_4  home_player_5  home_player_6  home_player_7  \
0        39774.0        33085.0        30894.0        38244.0        30872.0
1        38835.0        30986.0        31291.0        31013.0        30935.0
```

```
2        24446.0        24408.0        36786.0        38802.0        24655.0
3        23818.0        37277.0        30687.0        36394.0        37169.0
4        24004.0        33086.0        30857.0        24011.0       109058.0

   home_player_8  home_player_9  home_player_10  home_player_11  \
0        38843.0        95078.0         30638.0         32118.0
1        39297.0        26181.0         30960.0         36410.0
2        17866.0        30352.0         23927.0         24410.0
3        24223.0        24773.0         34543.0         23139.0
4        23268.0        24846.0         24006.0         24160.0

   away_player_1  away_player_2  away_player_3  away_player_4  away_player_5  \
0        25524.0        36183.0        27293.0        37787.0        37156.0
1        36373.0        36832.0        23115.0        37280.0        24728.0
2        30660.0        37442.0        30617.0        24134.0       414792.0
3        34421.0        34987.0        35472.0       111865.0        25005.0
4        30622.0        37764.0        19020.0        23921.0        24136.0

   away_player_6  away_player_7  away_player_8  away_player_9  away_player_10  \
0        30749.0        30598.0        39106.0        38216.0         33101.0
1        24664.0        31088.0        23257.0        24171.0         25922.0
2        37139.0        30618.0        40701.0        24800.0         24635.0
3        35327.0        25150.0        97988.0        41877.0        127857.0
4        30342.0        23889.0        23916.0        23922.0         34176.0

   away_player_11  goal_player_id  assist_player_id
0         30764.0         30872.0               NaN
1         27267.0         26181.0           39297.0
2         30853.0         30853.0           30889.0
3         34466.0         23139.0           36394.0
4         30646.0         30342.0           23916.0
```

[13]:
```
# merge team_long_name for home and away team each

df_team = pd.read_csv('Team.csv')
df_team.drop(columns = ['id', 'team_fifa_api_id', 'team_short_name'], inplace =␣
↪True)
df_team.head()
```

[13]:
```
   team_api_id     team_long_name
0         9987           KRC Genk
1         9993       Beerschot AC
2        10000   SV Zulte-Waregem
3         9994   Sporting Lokeren
4         9984  KSV Cercle Brugge
```

[14]:
```
df_team.nunique()
```

[14]:
```
team_api_id       299
team_long_name    296
dtype: int64
```

[15]:
```
df_home = df_team.copy()
df_away = df_team.copy()
```

[16]:
```
# remane column names for merging

df_home = df_home.rename(columns = {'team_api_id' : 'home_team_api_id',␣
↪'team_long_name' : 'home_team_name'})
df_home.head()
```

[16]:

	home_team_api_id	home_team_name
0	9987	KRC Genk
1	9993	Beerschot AC
2	10000	SV Zulte-Waregem
3	9994	Sporting Lokeren
4	9984	KSV Cercle Brugge

[17]:
```
df_away = df_team.rename(columns = {'team_api_id' : 'away_team_api_id',␣
↪'team_long_name' : 'away_team_name'})
df_away.head()
```

[17]:

	away_team_api_id	away_team_name
0	9987	KRC Genk
1	9993	Beerschot AC
2	10000	SV Zulte-Waregem
3	9994	Sporting Lokeren
4	9984	KSV Cercle Brugge

[18]:
```
df = pd.merge(df, df_home, how = 'inner')
df = pd.merge(df, df_away, how = 'inner')

# pd.merge(df, df_copy, left_on = 'home_team_api_id', right_on = 'team_api_id')
```

[19]:
```
df.shape
```

[19]:
```
(13325, 32)
```

[20]:
```
# merge goal_player's name and physical record
# merge assist_player's name and physical record

df_player = pd.read_csv('Player.csv')
df_player.drop(columns = ['id', 'player_fifa_api_id'], inplace = True)
```

[21]:
```
df_player.head()
```

[21]:
```
   player_api_id         player_name             birthday  height  weight
0         505942  Aaron Appindangoye  1992-02-29 00:00:00  182.88     187
1         155782     Aaron Cresswell  1989-12-15 00:00:00  170.18     146
2         162549         Aaron Doran  1991-05-13 00:00:00  170.18     163
3          30572       Aaron Galindo  1982-05-08 00:00:00  182.88     198
4          23780        Aaron Hughes  1979-11-08 00:00:00  182.88     154
```

[22]:
```
df_goal = df_player.copy()
df_assist = df_player.copy()
```

[23]:
```
df_goal = df_goal.rename(columns = {'player_api_id' : 'goal_player_id',
                                    'player_name':'goal_player_name',
                                    'birthday':'goal_birthday',
                                    'height':'goal_height',
                                    'weight':'goal_weight'})
df_goal.head()
```

[23]:
```
   goal_player_id    goal_player_name        goal_birthday  goal_height  \
0          505942  Aaron Appindangoye  1992-02-29 00:00:00       182.88
1          155782     Aaron Cresswell  1989-12-15 00:00:00       170.18
2          162549         Aaron Doran  1991-05-13 00:00:00       170.18
3           30572       Aaron Galindo  1982-05-08 00:00:00       182.88
4           23780        Aaron Hughes  1979-11-08 00:00:00       182.88

   goal_weight
0          187
1          146
2          163
3          198
4          154
```

[24]:
```
df_assist = df_assist.rename(columns = {'player_api_id' : 'assist_player_id',
                                        'player_name':'assist_player_name',
                                        'birthday':'assist_birthday',
                                        'height':'assist_height',
                                        'weight':'assist_weight'})
df_assist.head()
```

[24]:
```
   assist_player_id  assist_player_name      assist_birthday  assist_height  \
0            505942  Aaron Appindangoye  1992-02-29 00:00:00         182.88
1            155782     Aaron Cresswell  1989-12-15 00:00:00         170.18
2            162549         Aaron Doran  1991-05-13 00:00:00         170.18
3             30572       Aaron Galindo  1982-05-08 00:00:00         182.88
4             23780        Aaron Hughes  1979-11-08 00:00:00         182.88

   assist_weight
0            187
```

```
1            146
2            163
3            198
4            154
```

[25]:
```
df.shape
```

[25]:
```
(13325, 32)
```

[26]:
```
df = pd.merge(df, df_goal, how = 'inner')
df = pd.merge(df, df_assist, how = 'inner')
```

[27]:
```
df.shape
```

[27]:
```
(7173, 40)
```

[28]:
```
df.isnull().sum()
```

[28]:
```
season               0
date                 0
home_team_api_id     0
away_team_api_id     0
home_team_goal       0
away_team_goal       0
home_player_1        0
home_player_2        0
home_player_3        0
home_player_4        0
home_player_5        0
home_player_6        0
home_player_7        0
home_player_8        0
home_player_9        0
home_player_10       0
home_player_11       0
away_player_1        0
away_player_2        0
away_player_3        0
away_player_4        0
away_player_5        0
away_player_6        0
away_player_7        0
away_player_8        0
away_player_9        0
away_player_10       0
away_player_11       0
goal_player_id       0
```

```
assist_player_id      0
home_team_name        0
away_team_name        0
goal_player_name      0
goal_birthday         0
goal_height           0
goal_weight           0
assist_player_name    0
assist_birthday       0
assist_height         0
assist_weight         0
dtype: int64
```

[29]:
```
# merge goal_player's overall rating and potential
# merge assist_player's overall rating and potential

df_att = pd.read_csv('Player_Attributes.csv')
```

[30]:
```
df_att = df_att[['player_api_id', 'overall_rating', 'potential']]
df_att.sample(10)
```

[30]:

	player_api_id	overall_rating	potential
83400	22929	72.0	75.0
54490	165709	74.0	84.0
96959	302696	64.0	74.0
92080	45245	64.0	66.0
87563	56952	73.0	73.0
121742	111865	76.0	77.0
58768	27663	74.0	74.0
56602	23986	69.0	70.0
38416	17600	65.0	70.0
101008	248150	57.0	67.0

[31]:
```
df_att_goal = df_att.copy()
df_att_assist = df_att.copy()
```

[32]:
```
df_att_goal = df_att_goal.rename(columns = {'player_api_id' : 'goal_player_id',
                                            'overall_rating':
 ↪'goal_overall_rating',
                                            'potential':'goal_potential'})
df_att_goal.head()
```

[32]:

	goal_player_id	goal_overall_rating	goal_potential
0	505942	67.0	71.0
1	505942	67.0	71.0
2	505942	62.0	66.0
3	505942	61.0	65.0

```
4          505942                 61.0            65.0
```

[33]:
```
df_att_assist = df_att_assist.rename(columns = {'player_api_id' :␣
↪'assist_player_id',
                                                'overall_rating':
↪'assist_overall_rating',
                                                'potential':'assist_potential'})
df_att_assist.head()
```

[33]:

	assist_player_id	assist_overall_rating	assist_potential
0	505942	67.0	71.0
1	505942	67.0	71.0
2	505942	62.0	66.0
3	505942	61.0	65.0
4	505942	61.0	65.0

[34]:
```
df_att_goal = df_att_goal.groupby('goal_player_id').mean().reset_index()
df_att_assist = df_att_assist.groupby('assist_player_id').mean().reset_index()
```

[35]:
```
df.shape
```

[35]:
```
(7173, 40)
```

[36]:
```
df = pd.merge(df, df_att_goal, how = 'inner')
df = pd.merge(df, df_att_assist, how = 'inner')
```

[37]:
```
df.shape
```

[37]:
```
(7173, 44)
```

[38]:
```
df['goal_point'] = np.abs(df['home_team_goal'] - df['away_team_goal'])
```

[39]:
```
# Calculate win or loss and then put the result into new column

df.loc[df['home_team_goal'] - df['away_team_goal'] == 0, 'match_winner_team'] =␣
↪'draw'
df.loc[df['home_team_goal'] - df['away_team_goal'] > 0, 'match_winner_team'] =␣
↪df['home_team_name']
df.loc[df['home_team_goal'] - df['away_team_goal'] < 0, 'match_winner_team'] =␣
↪df['away_team_name']

df.loc[df['home_team_goal'] - df['away_team_goal'] == 0, 'win_location'] =␣
↪'draw'
df.loc[df['home_team_goal'] - df['away_team_goal'] > 0, 'win_location'] = 'home␣
↪win'
df.loc[df['home_team_goal'] - df['away_team_goal'] < 0, 'win_location'] = 'away␣
↪win'
```

[40]:
```
# Calculate goal or assist player's ages and then put the result into new column

df['goal_player_age'] = pd.to_datetime(df.date).dt.year - pd.to_datetime(df.
 ↪goal_birthday).dt.year
df['assist_player_age'] = pd.to_datetime(df.date).dt.year - pd.to_datetime(df.
 ↪assist_birthday).dt.year
```

[41]:
```
df.columns.values
```

[41]:
```
array(['season', 'date', 'home_team_api_id', 'away_team_api_id',
       'home_team_goal', 'away_team_goal', 'home_player_1',
       'home_player_2', 'home_player_3', 'home_player_4', 'home_player_5',
       'home_player_6', 'home_player_7', 'home_player_8', 'home_player_9',
       'home_player_10', 'home_player_11', 'away_player_1',
       'away_player_2', 'away_player_3', 'away_player_4', 'away_player_5',
       'away_player_6', 'away_player_7', 'away_player_8', 'away_player_9',
       'away_player_10', 'away_player_11', 'goal_player_id',
       'assist_player_id', 'home_team_name', 'away_team_name',
       'goal_player_name', 'goal_birthday', 'goal_height', 'goal_weight',
       'assist_player_name', 'assist_birthday', 'assist_height',
       'assist_weight', 'goal_overall_rating', 'goal_potential',
       'assist_overall_rating', 'assist_potential', 'goal_point',
       'match_winner_team', 'win_location', 'goal_player_age',
       'assist_player_age'], dtype=object)
```

[42]:
```
df = df[['season','home_team_name', 'away_team_name','home_team_goal',␣
 ↪'away_team_goal','goal_point','match_winner_team', 'win_location',
    'goal_player_name', 'goal_player_age', 'goal_height', 'goal_weight',␣
 ↪'goal_overall_rating', 'goal_potential',
    'assist_player_name', 'assist_player_age', 'assist_height',␣
 ↪'assist_weight', 'assist_overall_rating', 'assist_potential','home_player_1',
       'home_player_2', 'home_player_3', 'home_player_4', 'home_player_5',
       'home_player_6', 'home_player_7', 'home_player_8', 'home_player_9',
       'home_player_10', 'home_player_11', 'away_player_1',
       'away_player_2', 'away_player_3', 'away_player_4', 'away_player_5',
       'away_player_6', 'away_player_7', 'away_player_8', 'away_player_9',
       'away_player_10', 'away_player_11']]
```

[43]:
```
df = df.sort_values(['season']).reset_index(drop = True)
```

[44]:
```
df.shape
```

[44]:
```
(7173, 42)
```

Exploratory Data Analysis

[45]:
```
df.describe()
```

[45]:

	home_team_goal	away_team_goal	goal_point	goal_player_age
count	7173.000000	7173.000000	7173.000000	7173.000000
mean	1.771086	1.341140	1.551094	26.713091
std	1.303291	1.179405	1.250573	3.790556
min	0.000000	0.000000	0.000000	17.000000
25%	1.000000	0.000000	1.000000	24.000000
50%	2.000000	1.000000	1.000000	27.000000
75%	2.000000	2.000000	2.000000	29.000000
max	10.000000	9.000000	9.000000	40.000000

	goal_height	goal_weight	goal_overall_rating	goal_potential
count	7173.000000	7173.000000	7173.000000	7173.000000
mean	181.648775	169.230308	75.347249	79.325768
std	6.573613	15.283542	5.870668	5.232353
min	160.020000	126.000000	58.000000	63.214286
25%	177.800000	159.000000	71.285714	75.826087
50%	182.880000	168.000000	75.066667	79.130435
75%	185.420000	179.000000	79.275862	82.687500
max	203.200000	225.000000	92.192308	95.230769

	assist_player_age	assist_height	assist_weight	assist_overall_rating
count	7173.000000	7173.000000	7173.000000	7173.000000
mean	26.611878	179.963940	165.549003	74.967821
std	3.806668	6.346866	14.681495	5.823612
min	17.000000	160.020000	126.000000	53.000000
25%	24.000000	175.260000	154.000000	71.045455
50%	26.000000	180.340000	165.000000	74.666667
75%	29.000000	182.880000	174.000000	79.166667
max	41.000000	203.200000	225.000000	92.192308

	assist_potential	home_player_1	home_player_2	home_player_3
count	7173.000000	7173.000000	7173.000000	7173.000000
mean	79.012432	80507.206887	112152.088805	89655.592639
std	5.165845	94808.256128	122876.877059	107452.580930
min	59.500000	2984.000000	9079.000000	2752.000000
25%	75.555556	30657.000000	31290.000000	27660.000000
50%	78.857143	37193.000000	41165.000000	37879.000000
75%	82.409091	103428.000000	160627.000000	113311.000000
max	95.230769	698273.000000	748432.000000	696443.000000

	home_player_4	home_player_5	home_player_6	home_player_7
count	7173.000000	7173.000000	7173.000000	7173.000000
mean	92539.452949	113395.209815	108555.468423	100380.783633
std	103329.961959	122512.198766	121251.533732	117151.589044
min	2752.000000	2752.000000	2802.000000	2802.000000
25%	29471.000000	31291.000000	30895.000000	30618.000000
50%	39027.000000	42008.000000	40945.000000	40015.000000

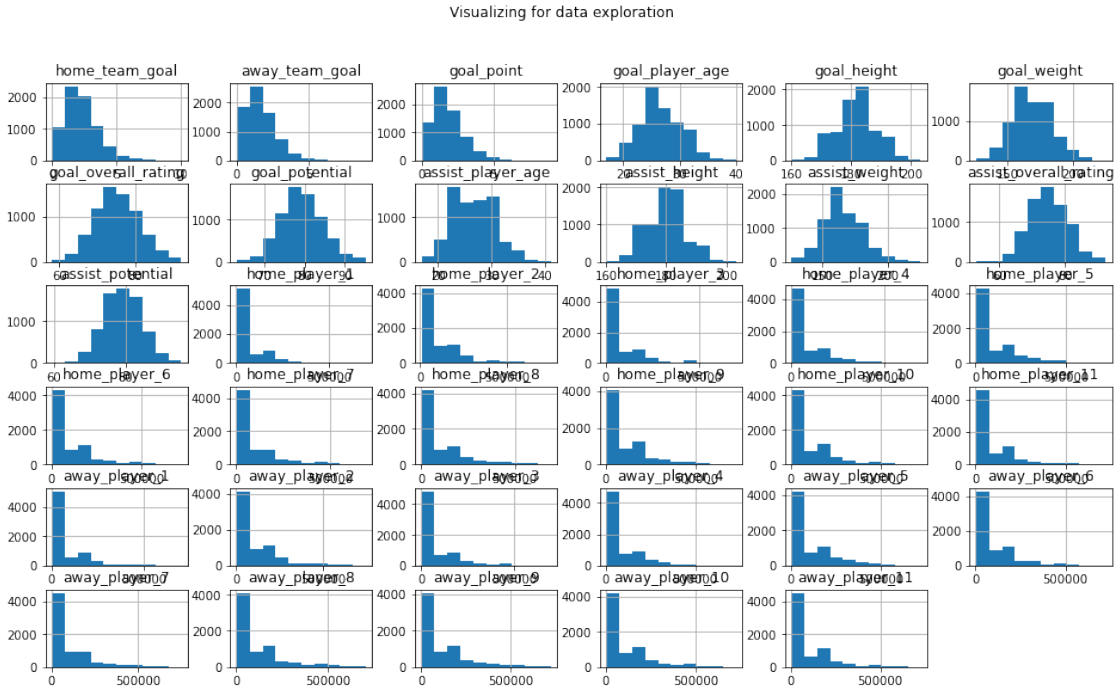
```
75%    128037.000000  163200.000000  160448.000000  143795.000000
max    696443.000000  720738.000000  722766.000000  692984.000000

       home_player_8  home_player_9  home_player_10  home_player_11  \
count    7173.000000    7173.000000     7173.000000     7173.000000
mean   112470.811376  117457.202705   108205.342535   100410.349784
std    124972.078512  127227.024306   120568.801816   113762.569969
min      2802.000000    2802.000000     2802.000000     2802.000000
25%     30905.000000   31244.000000    30881.000000    30853.000000
50%     41694.000000   41887.000000    40636.000000    40197.000000
75%    166509.000000  169217.000000   161291.000000   150832.000000
max    693171.000000  722766.000000   722766.000000   717557.000000

       away_player_1  away_player_2  away_player_3  away_player_4  \
count    7173.000000    7173.000000    7173.000000    7173.000000
mean    82355.182908  114751.936010   89077.280078   94329.548167
std     96676.791215  124546.520082  105351.586865  106591.764010
min      2984.000000    2790.000000    2752.000000    2752.000000
25%     30657.000000   31306.000000   27660.000000   27721.000000
50%     37233.000000   41666.000000   37604.000000   38899.000000
75%    104986.000000  167065.000000  113270.000000  133201.000000
max    698273.000000  748432.000000  696443.000000  728414.000000

       away_player_5  away_player_6  away_player_7  away_player_8  \
count    7173.000000    7173.000000    7173.000000    7173.000000
mean   113877.421163  108579.981179  103824.482783  116681.808309
std    121969.632701  121157.025817  121700.941017  129678.748268
min      2790.000000    2802.000000    2802.000000    2802.000000
25%     32569.000000   30894.000000   30682.000000   30955.000000
50%     43061.000000   41232.000000   40501.000000   42479.000000
75%    163971.000000  160243.000000  148830.000000  173408.000000
max    720738.000000  722766.000000  750435.000000  710807.000000

       away_player_9  away_player_10  away_player_11
count    7173.000000     7173.000000     7173.000000
mean   118114.735118   109654.600446   104726.394256
std    128829.093368   121612.462611   118533.839217
min      2802.000000     2802.000000     2802.000000
25%     31235.000000    30881.000000    30854.000000
50%     41890.000000    41061.000000    40792.000000
75%    171091.000000   161420.000000   161328.000000
max    722766.000000   722766.000000   725718.000000
```

[46]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7173 entries, 0 to 7172
Data columns (total 42 columns):
```

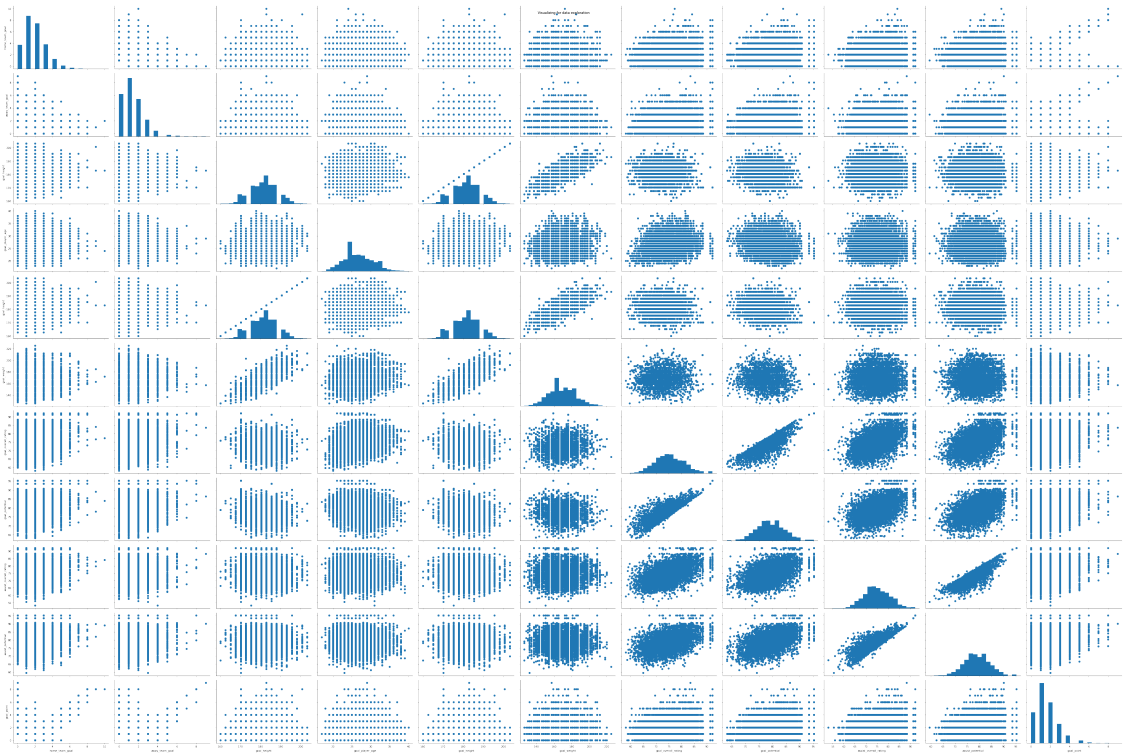
```
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   season                7173 non-null   object
 1   home_team_name        7173 non-null   object
 2   away_team_name        7173 non-null   object
 3   home_team_goal        7173 non-null   int64
 4   away_team_goal        7173 non-null   int64
 5   goal_point            7173 non-null   int64
 6   match_winner_team     7173 non-null   object
 7   win_location          7173 non-null   object
 8   goal_player_name      7173 non-null   object
 9   goal_player_age       7173 non-null   int64
 10  goal_height           7173 non-null   float64
 11  goal_weight           7173 non-null   int64
 12  goal_overall_rating   7173 non-null   float64
 13  goal_potential        7173 non-null   float64
 14  assist_player_name    7173 non-null   object
 15  assist_player_age     7173 non-null   int64
 16  assist_height         7173 non-null   float64
 17  assist_weight         7173 non-null   int64
 18  assist_overall_rating 7173 non-null   float64
 19  assist_potential      7173 non-null   float64
 20  home_player_1         7173 non-null   float64
 21  home_player_2         7173 non-null   float64
 22  home_player_3         7173 non-null   float64
 23  home_player_4         7173 non-null   float64
 24  home_player_5         7173 non-null   float64
 25  home_player_6         7173 non-null   float64
 26  home_player_7         7173 non-null   float64
 27  home_player_8         7173 non-null   float64
 28  home_player_9         7173 non-null   float64
 29  home_player_10        7173 non-null   float64
 30  home_player_11        7173 non-null   float64
 31  away_player_1         7173 non-null   float64
 32  away_player_2         7173 non-null   float64
 33  away_player_3         7173 non-null   float64
 34  away_player_4         7173 non-null   float64
 35  away_player_5         7173 non-null   float64
 36  away_player_6         7173 non-null   float64
 37  away_player_7         7173 non-null   float64
 38  away_player_8         7173 non-null   float64
 39  away_player_9         7173 non-null   float64
 40  away_player_10        7173 non-null   float64
 41  away_player_11        7173 non-null   float64
dtypes: float64(28), int64(7), object(7)
memory usage: 2.3+ MB
```

[47]:
```
import pylab as pl
df.hist(figsize = (16, 9));
pl.suptitle('Visualizing for data exploration'); # Not for the slide deck
```



[48]:
```
# variable categorizing

numeric_vars = ['home_team_goal', 'away_team_goal', 'goal_height',␣
 ↪'goal_player_age', 'goal_height', 'goal_weight', 'goal_overall_rating',␣
 ↪'goal_potential', 'assist_overall_rating', 'assist_potential', 'goal_point']
```

[49]:
```
g = sb.PairGrid(data = df, vars = numeric_vars, height = 3.5, aspect = 1.5);
g = g.map_diag(plt.hist, bins = 20);
g.map_offdiag(plt.scatter);
pl.suptitle('Visualizing for data exploration'); # Not for the slide deck
```



[50]:
```
df.corr()
```

[50]:

	home_team_goal	away_team_goal	goal_point	\
home_team_goal	1.000000	-0.226308	0.429955	
away_team_goal	-0.226308	1.000000	0.044003	
goal_point	0.429955	0.044003	1.000000	
goal_player_age	0.017355	-0.023576	-0.004172	
goal_height	-0.005330	0.018827	0.000825	
goal_weight	0.004341	0.012457	0.009130	
goal_overall_rating	0.071515	0.022713	0.189790	
goal_potential	0.061785	0.030799	0.201377	
assist_player_age	0.000526	-0.037462	-0.001281	
assist_height	-0.013662	-0.013565	-0.017747	
assist_weight	0.012508	-0.027929	-0.000427	
assist_overall_rating	0.065087	0.001484	0.195009	
assist_potential	0.061609	0.015981	0.206672	
home_player_1	-0.020029	0.019742	-0.017732	
home_player_2	-0.038896	0.011963	-0.009858	
home_player_3	-0.025490	0.016694	0.000172	
home_player_4	-0.019752	0.018546	-0.001408	
home_player_5	-0.041358	0.041180	-0.019944	
home_player_6	-0.033357	0.025636	-0.006978	
home_player_7	-0.022243	0.000362	-0.001494	

home_player_8	-0.027645	0.025273	-0.013167
home_player_9	-0.028230	0.015147	-0.000378
home_player_10	-0.046937	0.016841	-0.022538
home_player_11	-0.033622	0.036463	-0.010261
away_player_1	0.021333	-0.002769	0.003571
away_player_2	0.014689	-0.016356	-0.001456
away_player_3	0.024932	-0.006091	0.010085
away_player_4	0.017158	-0.009734	0.005991
away_player_5	0.007520	-0.019160	0.000635
away_player_6	-0.020176	-0.003553	-0.020417
away_player_7	0.001423	0.015763	0.013852
away_player_8	-0.008890	-0.002055	-0.003246
away_player_9	0.000440	-0.008301	-0.001057
away_player_10	0.022364	-0.022865	0.015938
away_player_11	0.022787	-0.011000	0.015876

	goal_player_age	goal_height	goal_weight	\
home_team_goal	0.017355	-0.005330	0.004341	
away_team_goal	-0.023576	0.018827	0.012457	
goal_point	-0.004172	0.000825	0.009130	
goal_player_age	1.000000	0.089946	0.162916	
goal_height	0.089946	1.000000	0.744458	
goal_weight	0.162916	0.744458	1.000000	
goal_overall_rating	0.219615	-0.057127	0.051867	
goal_potential	-0.154636	-0.100543	-0.036667	
assist_player_age	0.104633	0.000630	0.019163	
assist_height	-0.004279	0.030341	0.028235	
assist_weight	0.015103	0.024879	0.046022	
assist_overall_rating	0.038964	-0.019593	0.018801	
assist_potential	-0.008391	-0.029306	-0.007433	
home_player_1	-0.022107	-0.013669	-0.039265	
home_player_2	-0.095927	-0.017087	-0.069225	
home_player_3	-0.077036	0.000794	-0.041621	
home_player_4	-0.050701	-0.002291	-0.025053	
home_player_5	-0.075340	-0.029785	-0.061201	
home_player_6	-0.075944	-0.010456	-0.059610	
home_player_7	-0.083366	-0.026022	-0.062719	
home_player_8	-0.087186	-0.019881	-0.062425	
home_player_9	-0.104616	-0.004352	-0.051712	
home_player_10	-0.115418	0.000787	-0.052271	
home_player_11	-0.147429	-0.012844	-0.062005	
away_player_1	-0.063654	-0.026555	-0.059970	
away_player_2	-0.081441	0.004480	-0.050136	
away_player_3	-0.095043	-0.017828	-0.064380	
away_player_4	-0.059936	-0.029161	-0.050935	
away_player_5	-0.080475	-0.023691	-0.057460	
away_player_6	-0.085122	-0.025166	-0.070444	

```
away_player_7               -0.072761    -0.022661    -0.052291
away_player_8               -0.066529     0.002343    -0.042975
away_player_9               -0.111144    -0.024218    -0.066088
away_player_10              -0.107010    -0.039644    -0.083153
away_player_11              -0.111035    -0.051495    -0.088288

                    goal_overall_rating  goal_potential  assist_player_age  \
home_team_goal                 0.071515        0.061785           0.000526
away_team_goal                 0.022713        0.030799          -0.037462
goal_point                     0.189790        0.201377          -0.001281
goal_player_age                0.219615       -0.154636           0.104633
goal_height                   -0.057127       -0.100543           0.000630
goal_weight                    0.051867       -0.036667           0.019163
goal_overall_rating            1.000000        0.859870           0.044004
goal_potential                 0.859870        1.000000           0.003660
assist_player_age              0.044004        0.003660           1.000000
assist_height                 -0.033882       -0.037990           0.063249
assist_weight                 -0.002198       -0.018917           0.151466
assist_overall_rating          0.565428        0.516260           0.230318
assist_potential               0.517010        0.532517          -0.140899
home_player_1                 -0.186411       -0.115281          -0.059050
home_player_2                 -0.245199       -0.115114          -0.092195
home_player_3                 -0.193962       -0.091084          -0.087357
home_player_4                 -0.186886       -0.092333          -0.087812
home_player_5                 -0.230524       -0.119138          -0.084528
home_player_6                 -0.224594       -0.106014          -0.086321
home_player_7                 -0.224019       -0.113904          -0.088034
home_player_8                 -0.244463       -0.126888          -0.094869
home_player_9                 -0.238509       -0.102984          -0.105106
home_player_10                -0.263979       -0.122360          -0.099693
home_player_11                -0.256560       -0.102205          -0.115678
away_player_1                 -0.189464       -0.100822          -0.063313
away_player_2                 -0.215382       -0.084084          -0.068917
away_player_3                 -0.211058       -0.096849          -0.084252
away_player_4                 -0.175000       -0.071077          -0.058501
away_player_5                 -0.232474       -0.114588          -0.079860
away_player_6                 -0.219191       -0.095122          -0.069419
away_player_7                 -0.204537       -0.099673          -0.068333
away_player_8                 -0.233704       -0.126664          -0.087714
away_player_9                 -0.231926       -0.088718          -0.097388
away_player_10                -0.245103       -0.107779          -0.092607
away_player_11                -0.260758       -0.124122          -0.085948

                    assist_height  assist_weight  assist_overall_rating  \
home_team_goal          -0.013662       0.012508               0.065087
away_team_goal          -0.013565      -0.027929               0.001484
goal_point              -0.017747      -0.000427               0.195009
```

```
goal_player_age           -0.004279      0.015103             0.038964
goal_height                0.030341      0.024879            -0.019593
goal_weight                0.028235      0.046022             0.018801
goal_overall_rating       -0.033882     -0.002198             0.565428
goal_potential            -0.037990     -0.018917             0.516260
assist_player_age          0.063249      0.151466             0.230318
assist_height              1.000000      0.743710            -0.071608
assist_weight              0.743710      1.000000             0.025265
assist_overall_rating     -0.071608      0.025265             1.000000
assist_potential          -0.092771     -0.050713             0.860976
home_player_1             -0.024608     -0.055204            -0.205480
home_player_2             -0.015914     -0.067264            -0.257719
home_player_3             -0.020282     -0.051899            -0.216402
home_player_4             -0.015823     -0.051960            -0.207464
home_player_5             -0.014679     -0.044574            -0.246605
home_player_6             -0.010450     -0.054465            -0.249277
home_player_7             -0.022409     -0.062207            -0.245453
home_player_8             -0.014367     -0.054930            -0.253758
home_player_9             -0.018349     -0.054634            -0.249215
home_player_10            -0.013386     -0.046416            -0.254926
home_player_11            -0.010086     -0.042233            -0.260606
away_player_1             -0.011157     -0.039983            -0.203019
away_player_2             -0.032038     -0.069434            -0.219625
away_player_3             -0.023383     -0.054584            -0.211317
away_player_4             -0.017946     -0.048509            -0.183889
away_player_5             -0.031581     -0.050704            -0.234406
away_player_6              0.002847     -0.036791            -0.237087
away_player_7             -0.045001     -0.068596            -0.227843
away_player_8              0.001542     -0.042822            -0.247408
away_player_9              0.014114     -0.038884            -0.250100
away_player_10            -0.017622     -0.040850            -0.242517
away_player_11            -0.028180     -0.055447            -0.259085

                    assist_potential  home_player_1  home_player_2  \
home_team_goal              0.061609      -0.020029      -0.038896
away_team_goal              0.015981       0.019742       0.011963
goal_point                  0.206672      -0.017732      -0.009858
goal_player_age            -0.008391      -0.022107      -0.095927
goal_height                -0.029306      -0.013669      -0.017087
goal_weight                -0.007433      -0.039265      -0.069225
goal_overall_rating         0.517010      -0.186411      -0.245199
goal_potential              0.532517      -0.115281      -0.115114
assist_player_age          -0.140899      -0.059050      -0.092195
assist_height              -0.092771      -0.024608      -0.015914
assist_weight              -0.050713      -0.055204      -0.067264
assist_overall_rating       0.860976      -0.205480      -0.257719
assist_potential            1.000000      -0.114334      -0.128986
```

```
home_player_1                -0.114334       1.000000       0.253155
home_player_2                -0.128986       0.253155       1.000000
home_player_3                -0.095586       0.298651       0.296802
home_player_4                -0.092833       0.332877       0.302199
home_player_5                -0.125527       0.295679       0.301374
home_player_6                -0.126431       0.236311       0.303254
home_player_7                -0.131293       0.275992       0.310836
home_player_8                -0.122957       0.257887       0.313798
home_player_9                -0.105057       0.254012       0.303582
home_player_10               -0.124423       0.279069       0.345115
home_player_11               -0.117938       0.260045       0.322696
away_player_1                -0.107707       0.177097       0.253357
away_player_2                -0.093879       0.221141       0.303192
away_player_3                -0.098311       0.196732       0.273357
away_player_4                -0.082799       0.202302       0.264040
away_player_5                -0.116330       0.223647       0.258373
away_player_6                -0.121547       0.233400       0.266870
away_player_7                -0.115224       0.231044       0.272525
away_player_8                -0.127467       0.230568       0.287283
away_player_9                -0.113922       0.273482       0.280462
away_player_10               -0.118771       0.234928       0.312538
away_player_11               -0.133961       0.238972       0.324702

                     home_player_3  home_player_4  home_player_5  \
home_team_goal           -0.025490      -0.019752      -0.041358
away_team_goal            0.016694       0.018546       0.041180
goal_point                0.000172      -0.001408      -0.019944
goal_player_age          -0.077036      -0.050701      -0.075340
goal_height               0.000794      -0.002291      -0.029785
goal_weight              -0.041621      -0.025053      -0.061201
goal_overall_rating      -0.193962      -0.186886      -0.230524
goal_potential           -0.091084      -0.092333      -0.119138
assist_player_age        -0.087357      -0.087812      -0.084528
assist_height            -0.020282      -0.015823      -0.014679
assist_weight            -0.051899      -0.051960      -0.044574
assist_overall_rating    -0.216402      -0.207464      -0.246605
assist_potential         -0.095586      -0.092833      -0.125527
home_player_1             0.298651       0.332877       0.295679
home_player_2             0.296802       0.302199       0.301374
home_player_3             1.000000       0.272699       0.273954
home_player_4             0.272699       1.000000       0.290747
home_player_5             0.273954       0.290747       1.000000
home_player_6             0.289916       0.290778       0.268251
home_player_7             0.278417       0.276301       0.266025
home_player_8             0.247196       0.291862       0.264175
home_player_9             0.304253       0.325409       0.312711
home_player_10            0.286130       0.320921       0.308503
```

home_player_11	0.286359	0.301066	0.310965
away_player_1	0.218223	0.201758	0.212492
away_player_2	0.238936	0.240837	0.244441
away_player_3	0.229340	0.222983	0.243354
away_player_4	0.232823	0.241022	0.235089
away_player_5	0.243536	0.220727	0.234535
away_player_6	0.252511	0.240544	0.251876
away_player_7	0.241504	0.248507	0.263325
away_player_8	0.281438	0.265652	0.265157
away_player_9	0.293540	0.267441	0.261036
away_player_10	0.259141	0.239421	0.261753
away_player_11	0.249471	0.270112	0.261691

	home_player_6	home_player_7	home_player_8	\
home_team_goal	-0.033357	-0.022243	-0.027645	
away_team_goal	0.025636	0.000362	0.025273	
goal_point	-0.006978	-0.001494	-0.013167	
goal_player_age	-0.075944	-0.083366	-0.087186	
goal_height	-0.010456	-0.026022	-0.019881	
goal_weight	-0.059610	-0.062719	-0.062425	
goal_overall_rating	-0.224594	-0.224019	-0.244463	
goal_potential	-0.106014	-0.113904	-0.126888	
assist_player_age	-0.086321	-0.088034	-0.094869	
assist_height	-0.010450	-0.022409	-0.014367	
assist_weight	-0.054465	-0.062207	-0.054930	
assist_overall_rating	-0.249277	-0.245453	-0.253758	
assist_potential	-0.126431	-0.131293	-0.122957	
home_player_1	0.236311	0.275992	0.257887	
home_player_2	0.303254	0.310836	0.313798	
home_player_3	0.289916	0.278417	0.247196	
home_player_4	0.290778	0.276301	0.291862	
home_player_5	0.268251	0.266025	0.264175	
home_player_6	1.000000	0.283973	0.293661	
home_player_7	0.283973	1.000000	0.290090	
home_player_8	0.293661	0.290090	1.000000	
home_player_9	0.295471	0.277625	0.316816	
home_player_10	0.300690	0.320483	0.354263	
home_player_11	0.289535	0.284822	0.310463	
away_player_1	0.237843	0.214036	0.231898	
away_player_2	0.264519	0.274312	0.285759	
away_player_3	0.228169	0.255671	0.258458	
away_player_4	0.258961	0.235312	0.238143	
away_player_5	0.250594	0.249260	0.250428	
away_player_6	0.259976	0.255480	0.274264	
away_player_7	0.270169	0.229152	0.247106	
away_player_8	0.283253	0.251477	0.263745	
away_player_9	0.289305	0.255822	0.290582	

```
away_player_10              0.265458       0.275924       0.282440
away_player_11              0.281969       0.274087       0.287464

                      home_player_9  home_player_10  home_player_11  \
home_team_goal            -0.028230      -0.046937       -0.033622
away_team_goal             0.015147       0.016841        0.036463
goal_point                -0.000378      -0.022538       -0.010261
goal_player_age           -0.104616      -0.115418       -0.147429
goal_height               -0.004352       0.000787       -0.012844
goal_weight               -0.051712      -0.052271       -0.062005
goal_overall_rating       -0.238509      -0.263979       -0.256560
goal_potential            -0.102984      -0.122360       -0.102205
assist_player_age         -0.105106      -0.099693       -0.115678
assist_height             -0.018349      -0.013386       -0.010086
assist_weight             -0.054634      -0.046416       -0.042233
assist_overall_rating     -0.249215      -0.254926       -0.260606
assist_potential          -0.105057      -0.124423       -0.117938
home_player_1              0.254012       0.279069        0.260045
home_player_2              0.303582       0.345115        0.322696
home_player_3              0.304253       0.286130        0.286359
home_player_4              0.325409       0.320921        0.301066
home_player_5              0.312711       0.308503        0.310965
home_player_6              0.295471       0.300690        0.289535
home_player_7              0.277625       0.320483        0.284822
home_player_8              0.316816       0.354263        0.310463
home_player_9              1.000000       0.343589        0.320157
home_player_10             0.343589       1.000000        0.333278
home_player_11             0.320157       0.333278        1.000000
away_player_1              0.244532       0.239147        0.239989
away_player_2              0.302011       0.305971        0.293901
away_player_3              0.269123       0.258709        0.264454
away_player_4              0.265686       0.257812        0.240601
away_player_5              0.253619       0.258153        0.249043
away_player_6              0.284301       0.287452        0.286912
away_player_7              0.293231       0.268112        0.275850
away_player_8              0.304059       0.294932        0.275227
away_player_9              0.309623       0.302778        0.288046
away_player_10             0.269229       0.289443        0.280880
away_player_11             0.290601       0.287559        0.284921

                      away_player_1  away_player_2  away_player_3  \
home_team_goal             0.021333       0.014689       0.024932
away_team_goal            -0.002769      -0.016356      -0.006091
goal_point                 0.003571      -0.001456       0.010085
goal_player_age           -0.063654      -0.081441      -0.095043
goal_height               -0.026555       0.004480      -0.017828
goal_weight               -0.059970      -0.050136      -0.064380
```

```
goal_overall_rating        -0.189464      -0.215382      -0.211058
goal_potential             -0.100822      -0.084084      -0.096849
assist_player_age          -0.063313      -0.068917      -0.084252
assist_height              -0.011157      -0.032038      -0.023383
assist_weight              -0.039983      -0.069434      -0.054584
assist_overall_rating      -0.203019      -0.219625      -0.211317
assist_potential           -0.107707      -0.093879      -0.098311
home_player_1               0.177097       0.221141       0.196732
home_player_2               0.253357       0.303192       0.273357
home_player_3               0.218223       0.238936       0.229340
home_player_4               0.201758       0.240837       0.222983
home_player_5               0.212492       0.244441       0.243354
home_player_6               0.237843       0.264519       0.228169
home_player_7               0.214036       0.274312       0.255671
home_player_8               0.231898       0.285759       0.258458
home_player_9               0.244532       0.302011       0.269123
home_player_10              0.239147       0.305971       0.258709
home_player_11              0.239989       0.293901       0.264454
away_player_1               1.000000       0.247893       0.292041
away_player_2               0.247893       1.000000       0.297057
away_player_3               0.292041       0.297057       1.000000
away_player_4               0.345781       0.283692       0.276638
away_player_5               0.301377       0.301176       0.246548
away_player_6               0.262596       0.289100       0.309439
away_player_7               0.273936       0.326957       0.273748
away_player_8               0.266284       0.288534       0.280165
away_player_9               0.261929       0.306466       0.299425
away_player_10              0.282096       0.331557       0.295483
away_player_11              0.265624       0.331709       0.301211

                       away_player_4  away_player_5  away_player_6  \
home_team_goal              0.017158       0.007520      -0.020176
away_team_goal             -0.009734      -0.019160      -0.003553
goal_point                  0.005991       0.000635      -0.020417
goal_player_age            -0.059936      -0.080475      -0.085122
goal_height                -0.029161      -0.023691      -0.025166
goal_weight                -0.050935      -0.057460      -0.070444
goal_overall_rating        -0.175000      -0.232474      -0.219191
goal_potential             -0.071077      -0.114588      -0.095122
assist_player_age          -0.058501      -0.079860      -0.069419
assist_height              -0.017946      -0.031581       0.002847
assist_weight              -0.048509      -0.050704      -0.036791
assist_overall_rating      -0.183889      -0.234406      -0.237087
assist_potential           -0.082799      -0.116330      -0.121547
home_player_1               0.202302       0.223647       0.233400
home_player_2               0.264040       0.258373       0.266870
home_player_3               0.232823       0.243536       0.252511
```

home_player_4	0.241022	0.220727	0.240544
home_player_5	0.235089	0.234535	0.251876
home_player_6	0.258961	0.250594	0.259976
home_player_7	0.235312	0.249260	0.255480
home_player_8	0.238143	0.250428	0.274264
home_player_9	0.265686	0.253619	0.284301
home_player_10	0.257812	0.258153	0.287452
home_player_11	0.240601	0.249043	0.286912
away_player_1	0.345781	0.301377	0.262596
away_player_2	0.283692	0.301176	0.289100
away_player_3	0.276638	0.246548	0.309439
away_player_4	1.000000	0.284636	0.277087
away_player_5	0.284636	1.000000	0.272777
away_player_6	0.277087	0.272777	1.000000
away_player_7	0.290738	0.280308	0.293884
away_player_8	0.316196	0.274982	0.295468
away_player_9	0.310470	0.309906	0.307199
away_player_10	0.310931	0.313899	0.294923
away_player_11	0.301410	0.326596	0.299487

	away_player_7	away_player_8	away_player_9	\
home_team_goal	0.001423	-0.008890	0.000440	
away_team_goal	0.015763	-0.002055	-0.008301	
goal_point	0.013852	-0.003246	-0.001057	
goal_player_age	-0.072761	-0.066529	-0.111144	
goal_height	-0.022661	0.002343	-0.024218	
goal_weight	-0.052291	-0.042975	-0.066088	
goal_overall_rating	-0.204537	-0.233704	-0.231926	
goal_potential	-0.099673	-0.126664	-0.088718	
assist_player_age	-0.068333	-0.087714	-0.097388	
assist_height	-0.045001	0.001542	0.014114	
assist_weight	-0.068596	-0.042822	-0.038884	
assist_overall_rating	-0.227843	-0.247408	-0.250100	
assist_potential	-0.115224	-0.127467	-0.113922	
home_player_1	0.231044	0.230568	0.273482	
home_player_2	0.272525	0.287283	0.280462	
home_player_3	0.241504	0.281438	0.293540	
home_player_4	0.248507	0.265652	0.267441	
home_player_5	0.263325	0.265157	0.261036	
home_player_6	0.270169	0.283253	0.289305	
home_player_7	0.229152	0.251477	0.255822	
home_player_8	0.247106	0.263745	0.290582	
home_player_9	0.293231	0.304059	0.309623	
home_player_10	0.268112	0.294932	0.302778	
home_player_11	0.275850	0.275227	0.288046	
away_player_1	0.273936	0.266284	0.261929	
away_player_2	0.326957	0.288534	0.306466	

away_player_3	0.273748	0.280165	0.299425
away_player_4	0.290738	0.316196	0.310470
away_player_5	0.280308	0.274982	0.309906
away_player_6	0.293884	0.295468	0.307199
away_player_7	1.000000	0.308380	0.273439
away_player_8	0.308380	1.000000	0.308757
away_player_9	0.273439	0.308757	1.000000
away_player_10	0.333986	0.347668	0.361167
away_player_11	0.301159	0.320117	0.344572

	away_player_10	away_player_11
home_team_goal	0.022364	0.022787
away_team_goal	-0.022865	-0.011000
goal_point	0.015938	0.015876
goal_player_age	-0.107010	-0.111035
goal_height	-0.039644	-0.051495
goal_weight	-0.083153	-0.088288
goal_overall_rating	-0.245103	-0.260758
goal_potential	-0.107779	-0.124122
assist_player_age	-0.092607	-0.085948
assist_height	-0.017622	-0.028180
assist_weight	-0.040850	-0.055447
assist_overall_rating	-0.242517	-0.259085
assist_potential	-0.118771	-0.133961
home_player_1	0.234928	0.238972
home_player_2	0.312538	0.324702
home_player_3	0.259141	0.249471
home_player_4	0.239421	0.270112
home_player_5	0.261753	0.261691
home_player_6	0.265458	0.281969
home_player_7	0.275924	0.274087
home_player_8	0.282440	0.287464
home_player_9	0.269229	0.290601
home_player_10	0.289443	0.287559
home_player_11	0.280880	0.284921
away_player_1	0.282096	0.265624
away_player_2	0.331557	0.331709
away_player_3	0.295483	0.301211
away_player_4	0.310931	0.301410
away_player_5	0.313899	0.326596
away_player_6	0.294923	0.299487
away_player_7	0.333986	0.301159
away_player_8	0.347668	0.320117
away_player_9	0.361167	0.344572
away_player_10	1.000000	0.323004
away_player_11	0.323004	1.000000

1.1.3 Research Question 1

Soccer is one of the most popular sports in the world, even in Asia. Because soccer was known to Asia much later than the other continents, Asian teams hardly won the game with Europe or South America. When they had a landslide loss, I frequently came across an article or comments from audiences making excuses for the failure in the game with the player's physical benefit, age, overall rating, and potentials. To clarify this question, I would like to see a correlation between a player's overall rating/potential/height/weight/age and recording goals.

We made two groups. Each group has 30 players per the number of records from the top and bottom, respectively, because comparing those with no goal record with the top players is difficult to determine a valid comparison value. We get all the index numbers of each group and then calculate the average of the player's overall ratings and find the cumulated average for each group. We figured out that the top 30 players who recorded goals had 12 more overall ratings and 9 points in potentials. However, they were three years older and 31bs heavier than the other group. Considering the range of player's overall ratings and potential are 34 and 32 each, the gap is significant between those two groups. Therefore, it can be evidence that the overall rating and potentials with accumulated experience in the league positively affect making a goal. To make sure if the height and weight affect overall ratings and potential ratings, we implemented a regression analysis by plotting a scatter chart putting height and weight at X-axis and Potential ratings and overall ratings at Y-axis for the entire players recorded more than a goal. We also selected colors per the held location to check those three variables in the same plot. From the implementation, we could not find a significant correlation between them. In these findings, on top of everything, I can genuinely rely on the EA's rating algorism and do not prefer those to excuse the failure due to physical benefit only, and I can mention that top players' age mean experience.

Attribute description for top 30th and bottom 30th goal or assist players are as follows.

[51]:
```
df.head()
```

[51]:

	season	home_team_name	away_team_name	\
0	2008/2009	RC Deportivo de La Coruña	Athletic Club de Bilbao	
1	2008/2009	Portsmouth	Everton	
2	2008/2009	Borussia Dortmund	FC Schalke 04	
3	2008/2009	VfB Stuttgart	FC Schalke 04	
4	2008/2009	Fiorentina	Napoli	

	home_team_goal	away_team_goal	goal_point	match_winner_team	\
0	3	1	2	RC Deportivo de La Coruña	
1	2	1	1	Portsmouth	
2	3	3	0	draw	
3	2	0	2	VfB Stuttgart	
4	2	1	1	Fiorentina	

	win_location	goal_player_name	goal_player_age	goal_height	\

```
0     home win              Ze Castro               25       182.88
1     home win        Leighton Baines               25       170.18
2         draw       Jefferson Farfan               24       177.80
3     home win       Jefferson Farfan               24       177.80
4     home win  Mario Alberto Santana               28       177.80

   goal_weight  goal_overall_rating  goal_potential assist_player_name  \
0          181            73.105263       75.263158         Joan Verdu
1          154            81.031250       82.906250       Glen Johnson
2          185            82.333333       83.925926   Heiko Westermann
3          185            82.333333       83.925926        Pavel Pardo
4          157            75.137931       76.103448      Massimo Gobbi

   assist_player_age  assist_height  assist_weight  assist_overall_rating  \
0                 25         177.80            157              76.772727
1                 25         182.88            154              77.909091
2                 25         190.50            192              77.193548
3                 32         175.26            150              75.454545
4                 29         182.88            172              72.869565

   assist_potential  home_player_1  home_player_2  home_player_3  \
0         79.090909        33018.0        33756.0        34104.0
1         79.484848        36286.0        26108.0        34418.0
2         78.548387        27358.0        79737.0        71399.0
3         79.272727        30648.0        34113.0        36146.0
4         74.608696        24503.0        39621.0        39729.0

   home_player_4  home_player_5  home_player_6  home_player_7  home_player_8  \
0       150328.0        41167.0        37793.0        37449.0        38573.0
1        24216.0        23953.0        34036.0        24653.0        24747.0
2        36388.0       414794.0        43319.0        36132.0        30707.0
3        37401.0        38842.0        30628.0        34112.0        42346.0
4        24504.0        33816.0        24502.0        33888.0        24507.0

   home_player_9  home_player_10  home_player_11  away_player_1  \
0        37441.0         31045.0         37452.0        33764.0
1        38820.0         23190.0         30830.0        31465.0
2        27363.0         31718.0         35990.0        49586.0
3        34114.0         27326.0         40203.0        27299.0
4        31725.0         30881.0         92666.0        41322.0

   away_player_2  away_player_3  away_player_4  away_player_5  away_player_6  \
0        33025.0        33993.0        37604.0       154938.0        30287.0
1        26256.0        23268.0        33086.0        24846.0        24006.0
2        27303.0        30704.0        27301.0        30250.0        27483.0
3        27303.0        27461.0        27301.0        30250.0        27483.0
4        32769.0        39569.0        39509.0        39535.0        41887.0
```

	away_player_7	away_player_8	away_player_9	away_player_10	away_player_11
0	37410.0	96619.0	37605.0	33866.0	33030.0
1	39618.0	30371.0	36012.0	30735.0	40792.0
2	30251.0	27461.0	30702.0	30249.0	26437.0
3	30251.0	38913.0	26437.0	35997.0	30249.0
4	24546.0	41309.0	41350.0	24624.0	18925.0

[52]:
```
# Goal_player: top 30 vs bottom 30

top_goalplayer_index = df.goal_player_name.value_counts()[:30].index.values
bottom_goalplayer_index = df.goal_player_name.value_counts()[-30:].index.values
```

Overall Rating

[53]:
```
# cumulated average for each group

topagemean = []

for var in top_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_overall_rating.mean()
    topagemean.append(mean)

np.mean(topagemean)
```

[53]: 83.04045724099339

[54]:
```
df.goal_overall_rating.max()
```

[54]: 92.1923076923077

The average overall rating of players' recorded goals from the top 30th is 82.8, and the maximum is 92.2.

[55]:
```
# cumulated average for each group

bottomagemean = []

for var in bottom_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_overall_rating.mean()
    bottomagemean.append(mean)

np.mean(bottomagemean)
```

[55]: 70.17985178843222

The average overall rating of players' recorded goals from the top 30th is **12 more** than the bottom 30th goal players'.

Potential

[56]:
```
# cumulated average for each group

topagemean = []

for var in top_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_potential.mean()
    topagemean.append(mean)

np.mean(topagemean)
```

[56]: 85.04135613171688

[57]:
```
# cumulated average for each group

bottomagemean = []

for var in bottom_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_potential.mean()
    bottomagemean.append(mean)

np.mean(bottomagemean)
```

[57]: 74.16554237618911

The average overall rating of players' recorded assists from the top 30th is **9 more** than the bottom 30th goal players'.

age

[58]:
```
topagemean = []

for var in top_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_player_age.mean()
    topagemean.append(mean)

np.mean(topagemean)
```

[58]: 27.628260927925112

[59]:
```
bottomagemean = []

for var in bottom_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_player_age.mean()
    bottomagemean.append(mean)

np.mean(bottomagemean)
```

[59]: 26.5

The average age of players' recorded goals from the top 30th is **3 older** than the bottom 30th goal players'.

height

[60]:
```
topheightmean = []

for var in top_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_height.mean()
    topheightmean.append(mean)

np.mean(topheightmean)
```

[60]:
```
182.7953333333334
```

[61]:
```
bottomheightmean = []

for var in bottom_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_height.mean()
    bottomheightmean.append(mean)

np.mean(bottomheightmean)
```

[61]:
```
181.44066666666671
```

There is **no significant difference** in between.

weight

[62]:
```
topheightmean = []

for var in top_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_weight.mean()
    topheightmean.append(mean)

np.mean(topheightmean)
```

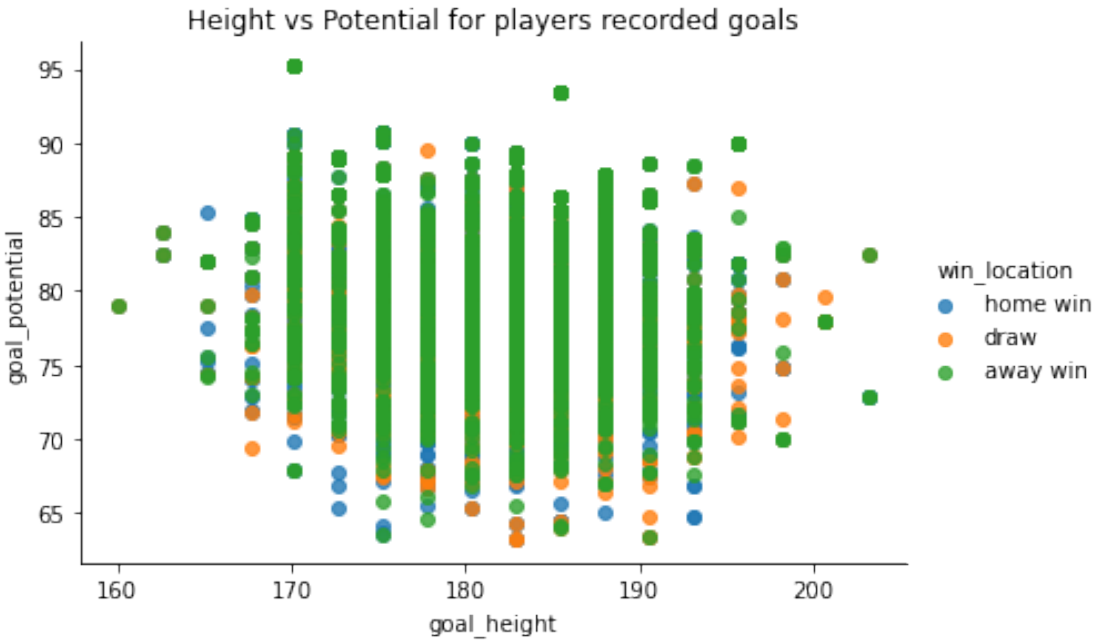
[62]:
```
171.56666666666666
```

[63]:
```
bottomheightmean = []

for var in bottom_goalplayer_index:
    mean = df.loc[df['goal_player_name'] == var].goal_weight.mean()
    bottomheightmean.append(mean)

np.mean(bottomheightmean)
```

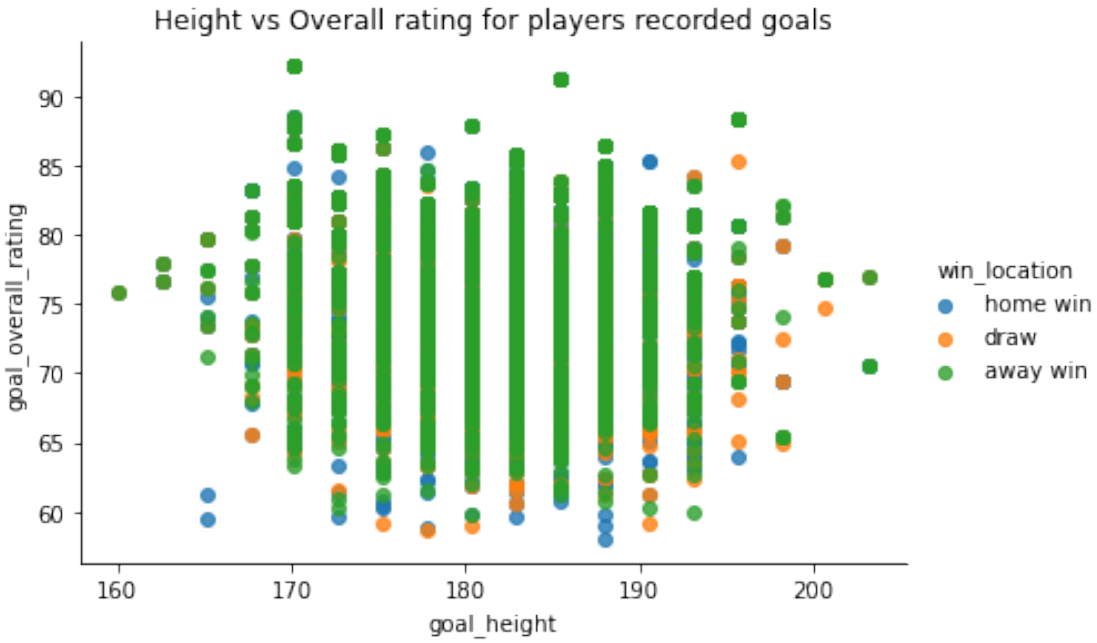
[63]:
```
167.53333333333333
```

The average weight of players' recorded goals from the top 30th is **3 lbs heavier** than the bottom 30th goal players'.

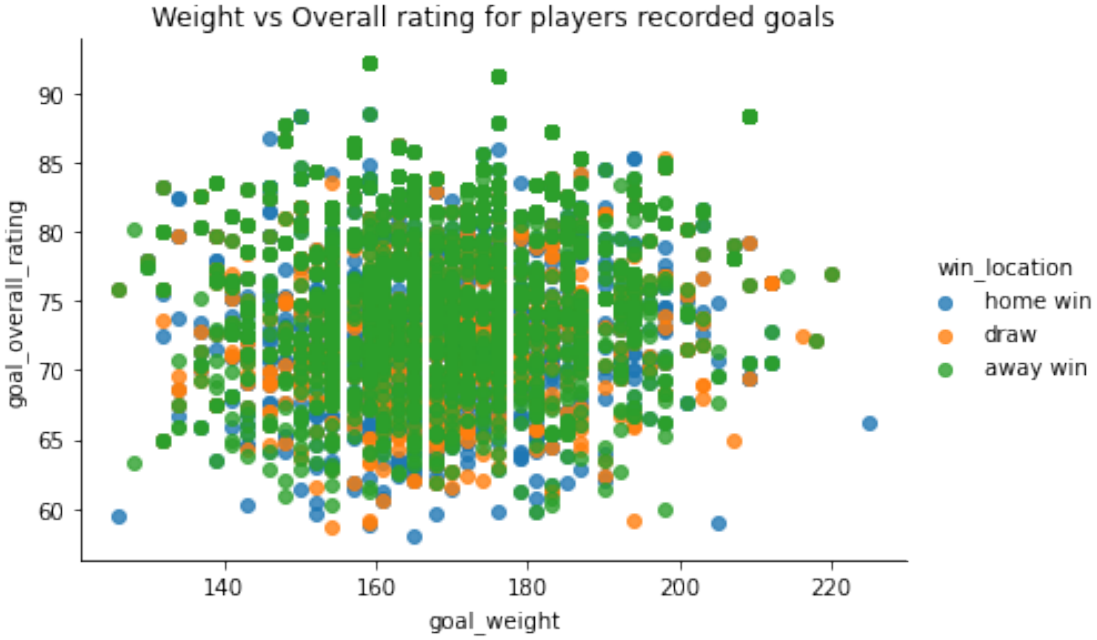
[64]:
```
g = sb.FacetGrid(data = df, hue = 'win_location', height = 4, aspect = 1.5);
g.map(sb.regplot, 'goal_height', 'goal_potential', fit_reg = False);
g.add_legend();
plt.title('Height vs Potential for players recorded goals');
plt.show();
```



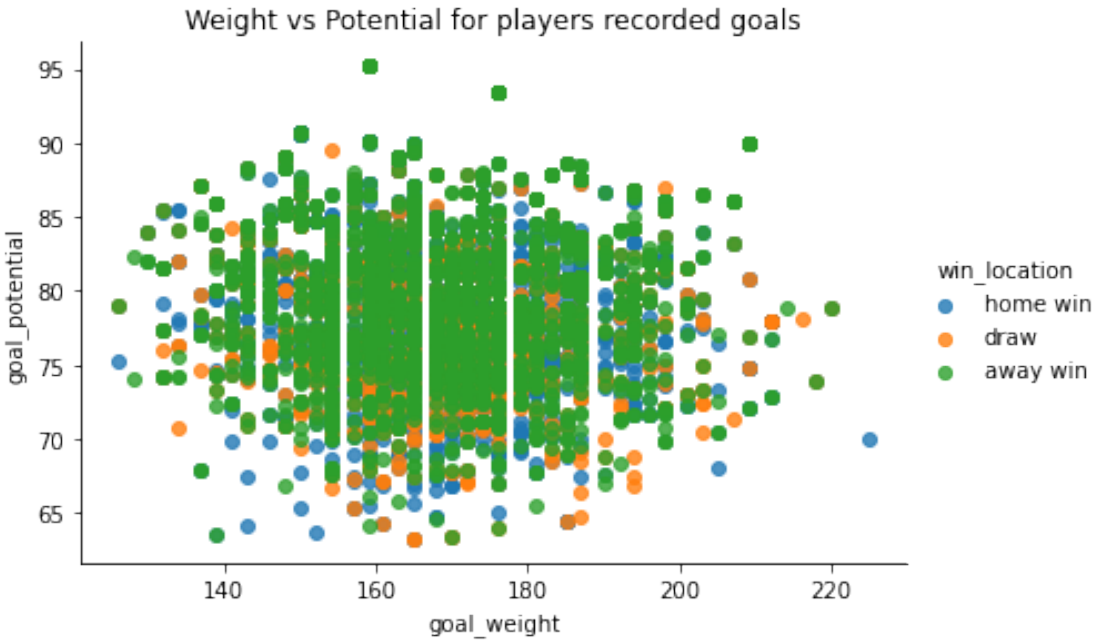
[65]:
```
g = sb.FacetGrid(data = df, hue = 'win_location', height = 4, aspect = 1.5)
g.map(sb.regplot, 'goal_height', 'goal_overall_rating', fit_reg = False);
g.add_legend();
plt.title('Height vs Overall rating for players recorded goals');
plt.show();
```



[66]:
```
g = sb.FacetGrid(data = df, hue = 'win_location', height = 4, aspect = 1.5)
g.map(sb.regplot, 'goal_weight', 'goal_overall_rating', fit_reg = False);
g.add_legend();
plt.title('Weight vs Overall rating for players recorded goals');
plt.show();
```



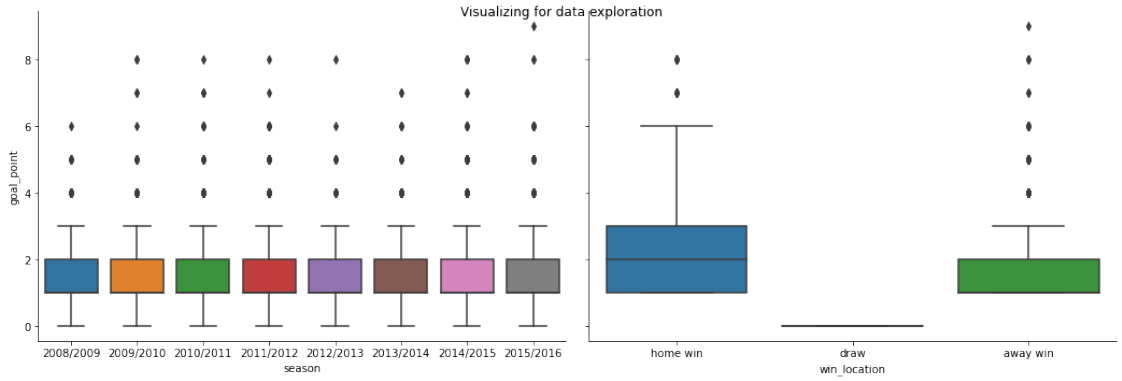
[67]:
```
g = sb.FacetGrid(data = df, hue = 'win_location', height = 4, aspect = 1.5)
g.map(sb.regplot, 'goal_weight', 'goal_potential', fit_reg = False);
g.add_legend();
plt.title('Weight vs Potential for players recorded goals');

plt.show();
```



[68]:
```
#variable categorizing

numeric_vars = ['goal_point']
categoric_vars = ['season', 'win_location']
```

variable categorizing

[69]:
```
def boxgrid(x, y, **kwargs):
    sb.boxplot(x, y)

g = sb.PairGrid(data = df, y_vars = numeric_vars, x_vars = categoric_vars,␣
↪height = 5, aspect = 1.5);
g.map(boxgrid);
pl.suptitle('Visualizing for data exploration'); # Not for the slide deck
plt.show();
```



[70]:
```
df.groupby('win_location').goal_point.mean()
```

[70]:
```
win_location
away win    1.782358
draw        0.000000
home win    1.998031
Name: goal_point, dtype: float64
```

1.1.4 Research Question 2

We know that the best play for the season is not always the best assister. I doubt that the attributes in assisting goal are possibly only ability to pass the ball well on the correct times but nothing to do with making goal. Therefore, I would like to know if a player's potential for assisting a goal is high, his potential for making a goal is also high.

First, we plotted a regression graph by putting the potential ratings of players recorded assists at X-axis and their overall ratings at Y-axis so as to check it surely related before implementing the intended test, and the graph shows closely co-related pattern. Then, we plotted a regression graph by putting assist-potential at X-axis and goal-potential at Y-axis. We also selected colors per the held location to check those three variables in the same plot. The scatter plot shows a regressive pattern heading straight upward, but it was not tight as much as the former analysis was; therefore, to make sure, we also conducted a Logistic Regression model to check the coefficient and P-value. Although the results show a low R-squared value, considering the coefficient;0.52 and P-values; less than 0.0001, we could find evidence that there is a remarkable correlation between those two attributes. From the analysis, I would recommend improving the passing skill first and then developing tactical moves to make goals for the soccer dreamers.

[71]:
```
df.head()
```

[71]:
```
      season             home_team_name           away_team_name  \
0  2008/2009  RC Deportivo de La Coruña  Athletic Club de Bilbao
```

```
1  2008/2009                 Portsmouth                 Everton
2  2008/2009          Borussia Dortmund           FC Schalke 04
3  2008/2009              VfB Stuttgart           FC Schalke 04
4  2008/2009                 Fiorentina                  Napoli

   home_team_goal  away_team_goal  goal_point          match_winner_team  \
0               3               1           2  RC Deportivo de La Coruña
1               2               1           1                 Portsmouth
2               3               3           0                       draw
3               2               0           2              VfB Stuttgart
4               2               1           1                 Fiorentina

  win_location       goal_player_name  goal_player_age  goal_height  \
0     home win             Ze Castro               25       182.88
1     home win        Leighton Baines              25       170.18
2         draw       Jefferson Farfan              24       177.80
3     home win       Jefferson Farfan              24       177.80
4     home win  Mario Alberto Santana              28       177.80

   goal_weight  goal_overall_rating  goal_potential assist_player_name  \
0          181            73.105263       75.263158         Joan Verdu
1          154            81.031250       82.906250       Glen Johnson
2          185            82.333333       83.925926   Heiko Westermann
3          185            82.333333       83.925926        Pavel Pardo
4          157            75.137931       76.103448      Massimo Gobbi

   assist_player_age  assist_height  assist_weight  assist_overall_rating  \
0                 25         177.80            157              76.772727
1                 25         182.88            154              77.909091
2                 25         190.50            192              77.193548
3                 32         175.26            150              75.454545
4                 29         182.88            172              72.869565

   assist_potential  home_player_1  home_player_2  home_player_3  \
0         79.090909        33018.0        33756.0        34104.0
1         79.484848        36286.0        26108.0        34418.0
2         78.548387        27358.0        79737.0        71399.0
3         79.272727        30648.0        34113.0        36146.0
4         74.608696        24503.0        39621.0        39729.0

   home_player_4  home_player_5  home_player_6  home_player_7  home_player_8  \
0       150328.0        41167.0        37793.0        37449.0        38573.0
1        24216.0        23953.0        34036.0        24653.0        24747.0
2        36388.0       414794.0        43319.0        36132.0        30707.0
3        37401.0        38842.0        30628.0        34112.0        42346.0
4        24504.0        33816.0        24502.0        33888.0        24507.0
```

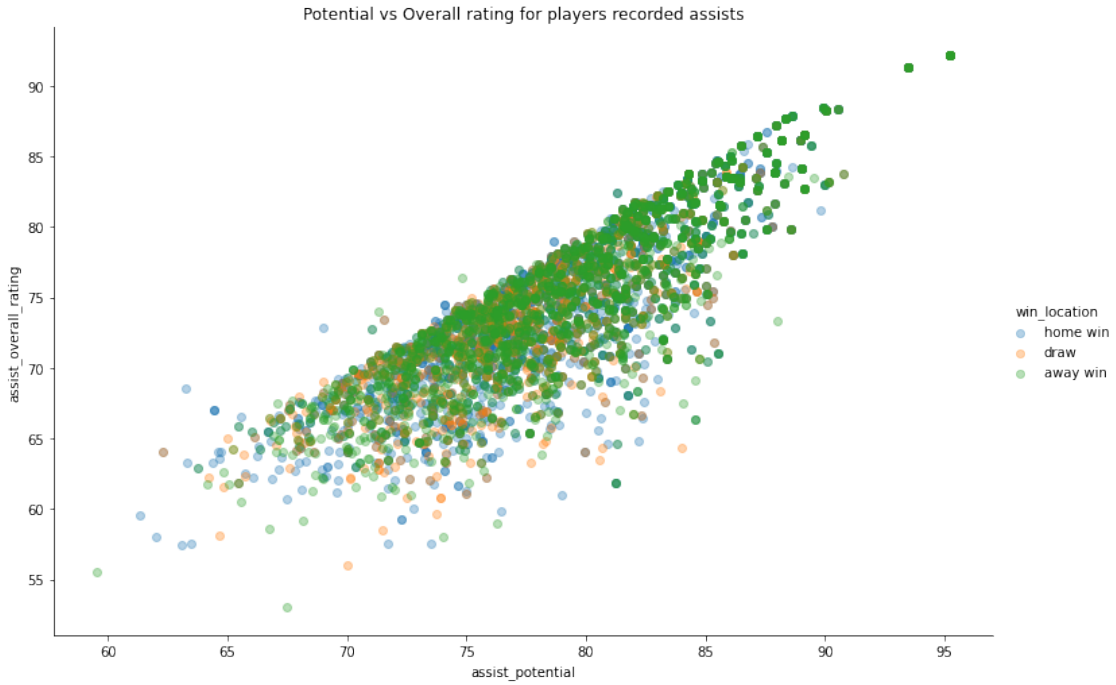
	home_player_9	home_player_10	home_player_11	away_player_1	\
0	37441.0	31045.0	37452.0	33764.0	
1	38820.0	23190.0	30830.0	31465.0	
2	27363.0	31718.0	35990.0	49586.0	
3	34114.0	27326.0	40203.0	27299.0	
4	31725.0	30881.0	92666.0	41322.0	

	away_player_2	away_player_3	away_player_4	away_player_5	away_player_6	\
0	33025.0	33993.0	37604.0	154938.0	30287.0	
1	26256.0	23268.0	33086.0	24846.0	24006.0	
2	27303.0	30704.0	27301.0	30250.0	27483.0	
3	27303.0	27461.0	27301.0	30250.0	27483.0	
4	32769.0	39569.0	39509.0	39535.0	41887.0	

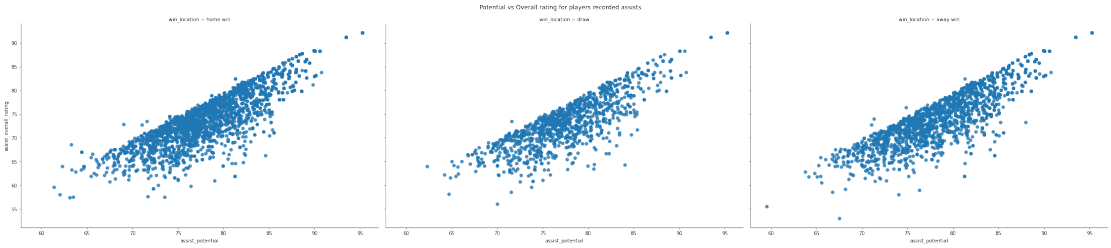
	away_player_7	away_player_8	away_player_9	away_player_10	away_player_11
0	37410.0	96619.0	37605.0	33866.0	33030.0
1	39618.0	30371.0	36012.0	30735.0	40792.0
2	30251.0	27461.0	30702.0	30249.0	26437.0
3	30251.0	38913.0	26437.0	35997.0	30249.0
4	24546.0	41309.0	41350.0	24624.0	18925.0

*explores at least three variables in relation to the primary question

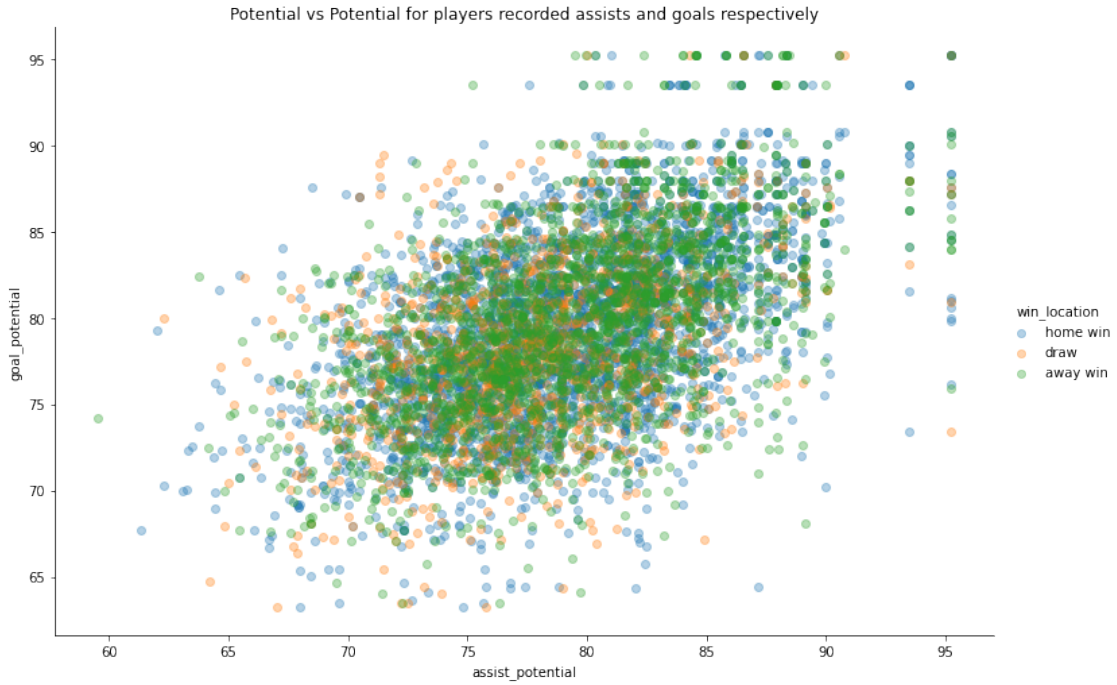
[119]:
```
g = sb.FacetGrid(data = df, hue = 'win_location', height = 7, aspect = 1.5)
g.map(sb.regplot, 'assist_potential', 'assist_overall_rating',  scatter_kws =␣
 ↪{'alpha': 1/3}, fit_reg = False);
g.add_legend();
plt.title('Potential vs Overall rating for players recorded assists'); # Not␣
 ↪for the slide deck

plt.show();
```



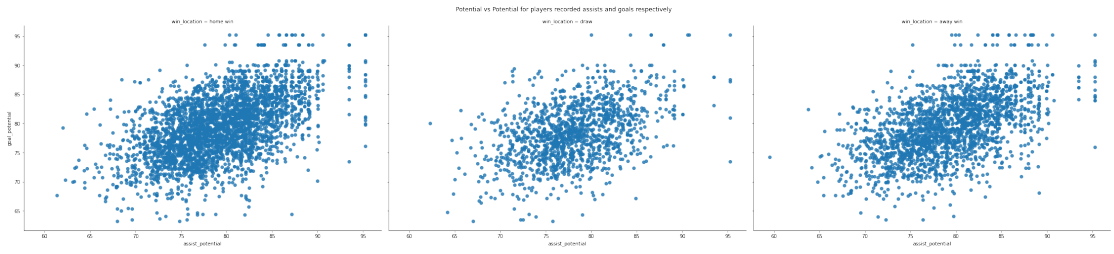
[128]:
```
g = sb.FacetGrid(data = df, col = 'win_location', height = 7, aspect = 1.5,␣
↪col_wrap = 3)
g.fig.suptitle('Potential vs Overall rating for players recorded assists');
g.map(sb.regplot, 'assist_potential', 'assist_overall_rating', fit_reg = False);
g.add_legend();
plt.show();
```



[120]:
```
g = sb.FacetGrid(data = df, hue = 'win_location', height = 7, aspect = 1.5)
g.map(sb.regplot, 'assist_potential', 'goal_potential', scatter_kws = {'alpha':␣
↪1/3}, fit_reg = False);
g.add_legend();
plt.title('Potential vs Potential for players recorded assists and goals␣
↪respectively');
plt.show();
```



[127]:
```
g = sb.FacetGrid(data = df, col = 'win_location', height = 7, aspect = 1.5,␣
 ↪col_wrap = 3)
g.fig.suptitle('Potential vs Potential for players recorded assists and goals␣
 ↪respectively');
g.map(sb.regplot, 'assist_potential', 'goal_potential', fit_reg = False);
g.add_legend();
plt.show();
```



[74]:
```
import statsmodels.api as sm

df['intercept'] = 1
lm = sm.OLS(df['assist_potential'], df[['intercept', 'goal_potential']])
result = lm.fit()
result.summary()
```

[74]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:       assist_potential   R-squared:                       0.284
Model:                           OLS   Adj. R-squared:                  0.283
Method:                Least Squares   F-statistic:                     2838.
Date:               Wed, 15 Sep 2021   Prob (F-statistic):               0.00
Time:                       14:36:34   Log-Likelihood:                -20760.
No. Observations:               7173   AIC:                         4.152e+04
Df Residuals:                   7171   BIC:                         4.154e+04
Df Model:                          1
Covariance Type:            nonrobust
==============================================================================
==
                     coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
--
intercept         37.3070      0.785     47.555      0.000      35.769
38.845
goal_potential     0.5257      0.010     53.277      0.000       0.506
0.545
==============================================================================
Omnibus:                       34.263   Durbin-Watson:                   1.667
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               47.008
Skew:                           0.035   Prob(JB):                     6.20e-11
Kurtosis:                       3.390   Cond. No.                     1.21e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.21e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

1.1.5 Research Question 3

As a Tottenham Hotspur lover, I never skip watching the game. Although the recent two seasons were not so great as they did, there is no doubt that they have been a prestigious team for many years without a tremendous downturn per the stadium's location. So, I want to find if they won more when they played at their home stadium and the gap is remarkable.

To answer this and the next questions, we melted the initial DataFrame by relocating the entire players' IDs and numbers. And then, selected the team winner as Tottenham Hotspur and categorized each location per season. We chose a count plot as setting X-axis with the location of the game and added different

colors on the bars per season. According to the shown chart, except for the 2013/2014 season, the team had more winning records when they played in the home stadium than in the opponents' place. The team gradually increased the number of wins played in away stadiums after touching the lowest winning record during the 2012/2013 season. The result is interesting enough to know that how the home benefit affects the game and why they have two matches per team when in the Europa championship league in each location.

[75]:
```
idv = ['season', 'home_team_name', 'away_team_name', 'home_team_goal',
       'away_team_goal', 'goal_point', 'match_winner_team',
       'win_location', 'goal_player_name', 'goal_player_age',
       'goal_height', 'goal_weight', 'goal_overall_rating',
       'goal_potential', 'assist_player_name', 'assist_player_age',
       'assist_height', 'assist_weight', 'assist_overall_rating',
       'assist_potential']
```

[76]:
```
vlv = ['home_player_1', 'home_player_2',
       'home_player_3', 'home_player_4', 'home_player_5', 'home_player_6',
       'home_player_7', 'home_player_8', 'home_player_9',
       'home_player_10', 'home_player_11', 'away_player_1',
       'away_player_2', 'away_player_3', 'away_player_4', 'away_player_5',
       'away_player_6', 'away_player_7', 'away_player_8', 'away_player_9',
       'away_player_10', 'away_player_11']
```

[77]:
```
dfp = df.melt(id_vars = idv, value_vars = vlv, var_name = 'player_number',␣
↪value_name = 'player_id')
```

[78]:
```
dfp.head()
```

[78]:

	season	home_team_name	away_team_name	home_team_goal	away_team_goal	goal_point	match_winner_team	win_location	goal_player_name	goal_player_age	goal_height
0	2008/2009	RC Deportivo de La Coruña	Athletic Club de Bilbao	3	1	2	RC Deportivo de La Coruña	home win	Ze Castro	25	182.88
1	2008/2009	Portsmouth	Everton	2	1	1	Portsmouth	home win	Leighton Baines	25	170.18
2	2008/2009	Borussia Dortmund	FC Schalke 04	3	3	0	draw	draw	Jefferson Farfan	24	177.80
3	2008/2009	VfB Stuttgart	FC Schalke 04	2	0	2	VfB Stuttgart	home win	Jefferson Farfan	24	177.80
4	2008/2009	Fiorentina	Napoli	2	1	1	Fiorentina				

```
4     home win  Mario Alberto Santana               28       177.80

   goal_weight  goal_overall_rating  goal_potential assist_player_name  \
0          181            73.105263       75.263158         Joan Verdu
1          154            81.031250       82.906250       Glen Johnson
2          185            82.333333       83.925926   Heiko Westermann
3          185            82.333333       83.925926        Pavel Pardo
4          157            75.137931       76.103448      Massimo Gobbi

   assist_player_age  assist_height  assist_weight  assist_overall_rating  \
0                 25         177.80            157              76.772727
1                 25         182.88            154              77.909091
2                 25         190.50            192              77.193548
3                 32         175.26            150              75.454545
4                 29         182.88            172              72.869565

   assist_potential  player_number  player_id
0         79.090909  home_player_1    33018.0
1         79.484848  home_player_1    36286.0
2         78.548387  home_player_1    27358.0
3         79.272727  home_player_1    30648.0
4         74.608696  home_player_1    24503.0
```

[79]:
```
df_pid = pd.read_csv('Player.csv')
df_pid.drop(columns = ['id', 'player_fifa_api_id', 'birthday', 'height',␣
↪'weight'], inplace = True)
df_pid = df_pid.rename(columns = {'player_api_id' : 'player_id'})
df_pid.head()
```

[79]:
```
   player_id         player_name
0     505942  Aaron Appindangoye
1     155782     Aaron Cresswell
2     162549         Aaron Doran
3      30572       Aaron Galindo
4      23780        Aaron Hughes
```

[80]:
```
dfp = pd.merge(dfp, df_pid, how = 'inner')
dfp.head()
```

[80]:
```
       season             home_team_name           away_team_name  \
0  2008/2009  RC Deportivo de La Coruña  Athletic Club de Bilbao
1  2008/2009  RC Deportivo de La Coruña          Real Madrid CF
2  2008/2009  RC Deportivo de La Coruña              Sevilla FC
3  2008/2009  RC Deportivo de La Coruña         Atlético Madrid
4  2008/2009  RC Deportivo de La Coruña         Real Valladolid

   home_team_goal  away_team_goal  goal_point         match_winner_team  \
```

```
0                3                1           2  RC Deportivo de La Coruña
1                2                1           1  RC Deportivo de La Coruña
2                1                3           2                 Sevilla FC
3                1                2           1           Atlético Madrid
4                1                0           1  RC Deportivo de La Coruña

  win_location     goal_player_name  goal_player_age  goal_height  \
0     home win            Ze Castro               25       182.88
1     home win                Mista               30       182.88
2     away win  Rodolfo Bodipo Diaz               32       182.88
3     away win        Sergio Aguero               21       172.72
4     home win       Lassad Nouioui               23       187.96

   goal_weight  goal_overall_rating  goal_potential   assist_player_name  \
0          181            73.105263       75.263158          Joan Verdu
1          163            74.800000       79.600000      Andres Guardado
2          174            67.133333       71.266667          Joan Verdu
3          163            86.114286       88.971429         Diego Forlan
4          172            69.000000       74.133333  Rodolfo Bodipo Diaz

   assist_player_age  assist_height  assist_weight  assist_overall_rating  \
0                 25         177.80            157              76.772727
1                 22         170.18            148              79.176471
2                 26         177.80            157              76.772727
3                 30         180.34            172              81.636364
4                 32         182.88            174              67.133333

   assist_potential  player_number  player_id       player_name
0         79.090909  home_player_1    33018.0  Daniel Aranzubia
1         82.470588  home_player_1    33018.0  Daniel Aranzubia
2         79.090909  home_player_1    33018.0  Daniel Aranzubia
3         82.409091  home_player_1    33018.0  Daniel Aranzubia
4         71.266667  home_player_1    33018.0  Daniel Aranzubia
```

[81]:
```
dfp.shape
```

[81]:
```
(157806, 23)
```

[82]:
```
ttm_win = dfp[dfp.match_winner_team == 'Tottenham Hotspur']
ttm_win
```

[82]:
```
           season     home_team_name       away_team_name  home_team_goal  \
55      2009/2010          Portsmouth    Tottenham Hotspur               1
84      2009/2010  Tottenham Hotspur           Portsmouth               2
373     2012/2013  Tottenham Hotspur  Queens Park Rangers               2
603     2009/2010          Hull City    Tottenham Hotspur               1
730     2009/2010  Tottenham Hotspur              Arsenal               2
```

```
…           …                …                    …                …
157105  2015/2016  Tottenham Hotspur    Manchester United                3
157182  2015/2016  Tottenham Hotspur            Sunderland               4
157247  2015/2016  Tottenham Hotspur    Manchester United                3
157494  2015/2016  Tottenham Hotspur          Norwich City               3
157564  2009/2010  Tottenham Hotspur      Birmingham City                2

        away_team_goal  goal_point  match_winner_team win_location  \
55                   2           1  Tottenham Hotspur     away win
84                   0           2  Tottenham Hotspur     home win
373                  1           1  Tottenham Hotspur     home win
603                  5           4  Tottenham Hotspur     away win
730                  1           1  Tottenham Hotspur     home win
…                  …           …                  …           …
157105               0           3  Tottenham Hotspur     home win
157182               1           3  Tottenham Hotspur     home win
157247               0           3  Tottenham Hotspur     home win
157494               0           3  Tottenham Hotspur     home win
157564               1           1  Tottenham Hotspur     home win

           goal_player_name  goal_player_age  goal_height  goal_weight  \
55              Ledley King               29       187.96          190
84             Peter Crouch               29       200.66          185
373            Bobby Zamora               31       185.42          192
603           Jermain Defoe               27       170.18          154
730              Danny Rose               20       172.72          159
…                       …                …            …           …
157105            Dele Alli               20       185.42          168
157182  Patrick van Aanholt               26       175.26          148
157247            Dele Alli               20       185.42          168
157494           Harry Kane               22       187.96          143
157564         Peter Crouch               28       200.66          185

        goal_overall_rating  goal_potential assist_player_name  \
55                 80.777778       83.444444      Niko Kranjcar
84                 76.727273       78.000000        Gareth Bale
373                74.115385       75.769231  Alejandro Faurlin
603                81.000000       81.714286    Tom Huddlestone
730                71.965517       78.551724      Jermain Defoe
…                        …               …                  …
157105             61.886364       81.250000  Christian Eriksen
157182             70.677419       76.516129       Adam Johnson
157247             61.886364       81.250000  Christian Eriksen
157494             72.413043       80.913043          Dele Alli
157564             76.727273       78.000000    Tom Huddlestone

        assist_player_age  assist_height  assist_weight  \
```

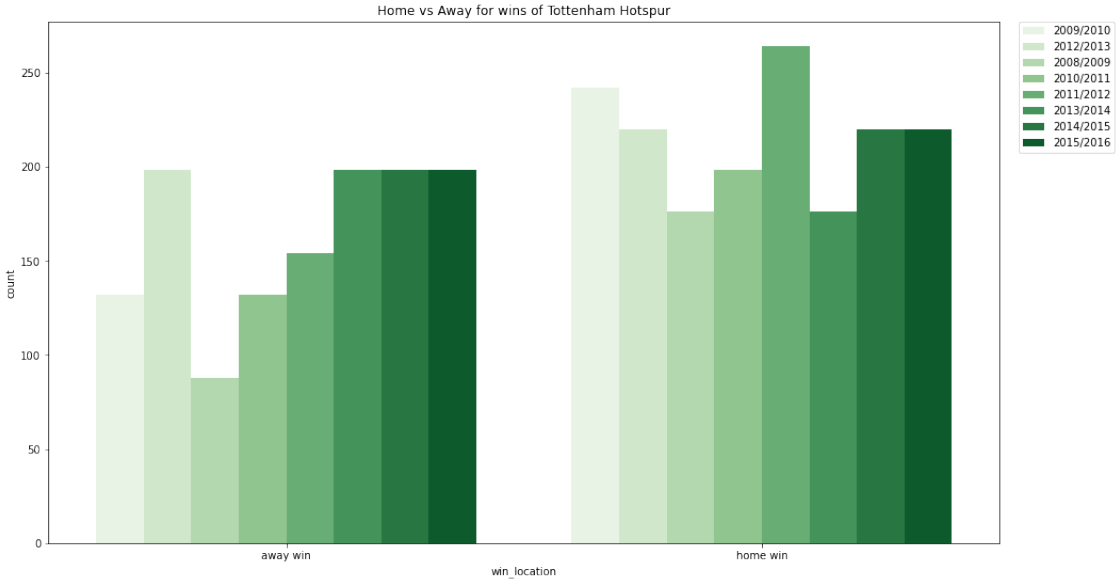
```
55                    25         185.42            165
84                    21         182.88            163
373                   26         185.42            174
603                   23         187.96            176
730                   28         170.18            154
…                    …           …                 …
157105                24         177.80            157
157182                29         175.26            139
157247                24         177.80            157
157494                19         185.42            168
157564                23         187.96            176

        assist_overall_rating  assist_potential   player_number  player_id  \
55                  77.142857         78.476190   home_player_1    36286.0
84                  84.096774         89.000000   away_player_1    36286.0
373                 72.238095         76.952381   away_player_1    30989.0
603                 76.851852         80.518519   home_player_1    23021.0
730                 81.000000         81.714286   away_player_1    23686.0
…                         …                 …               …        …
157105              79.818182         87.545455  away_player_11   696365.0
157182              76.500000         80.083333   away_player_1   303919.0
157247              79.818182         87.545455   away_player_2   639058.0
157494              61.886364         81.250000   away_player_9    38366.0
157564              76.851852         80.518519   away_player_8    24005.0

                player_name
55              David James
84              David James
373             Julio Cesar
603             Boaz Myhill
730          Manuel Almunia
…                       …
157105      Marcus Rashford
157182      Jordan Pickford
157247  Timothy Fosu-Mensah
157494   Vadis Odjidja-Ofoe
157564          Lee Carsley

[3014 rows x 23 columns]
```

[83]:
```
sum(ttm_win.win_location == 'home win') / ttm_win.win_location.shape[0]
```

[83]:
```
0.5693430656934306
```

[84]:
```
plt.figure(figsize = [16, 9])
```

```
sb.countplot(data = ttm_win, x = 'win_location', hue = 'season', palette =␣
↪'Greens')
plt.legend(bbox_to_anchor=(1.02,1),loc = 'upper left', borderaxespad=0)
plt.title('Home vs Away for wins of Tottenham Hotspur');

plt.show()
```



1.1.6 Research Question 4

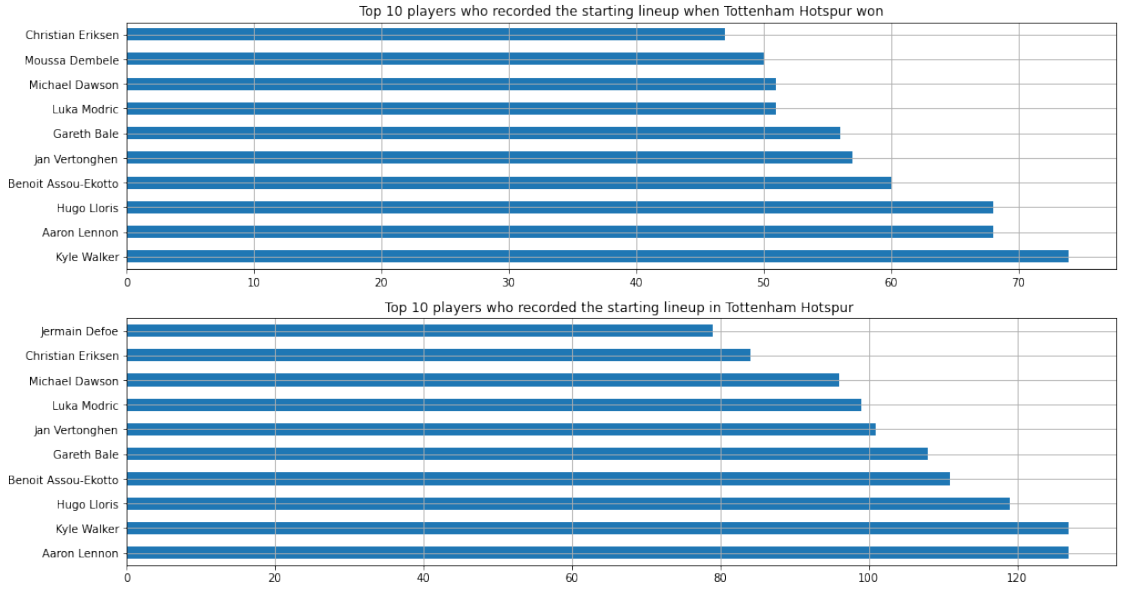
As a Tottenham Hotspur lover, it is also important to know even about single details. Since I have limited knowledge about who recorded the most goal, assist, and starting lineup of the game in Tottenham Hotspur over the recent decade, I want to get the information from this question.

I grouped two cases for this research; one is when the team won, and the other is else case. We plotted a bar chart for each of the top 10 players. The top player who recorded in the starting lineup was 'Aaron Lennon' and 'Kyle Walker' respectively. And the top player who recorded goals was 'Gareth Bale' and 'Harry Kane' respectively. Lastly, the top player who recorded assists was 'Aaron Lennon' in both cases. I did not know that Bale was so great when he was here. Indeed, he was in his prime during the period with the team.

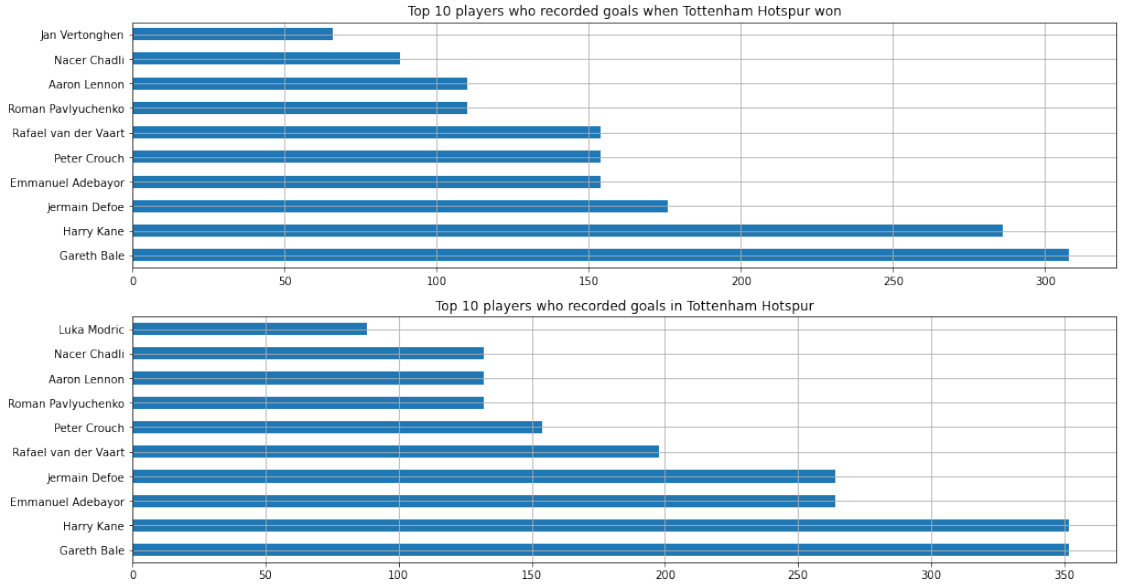
[85]:
```
ttm_all = dfp.query('home_team_name == "Tottenham Hotspur" | away_team_name ==␣
↪"Tottenham Hotspur"')
```

[86]:
```
plt.subplot(2, 1, 1);
ttm_win.player_name.value_counts()[:10].plot(kind='barh', figsize = (16,9),␣
↪grid = True);
```

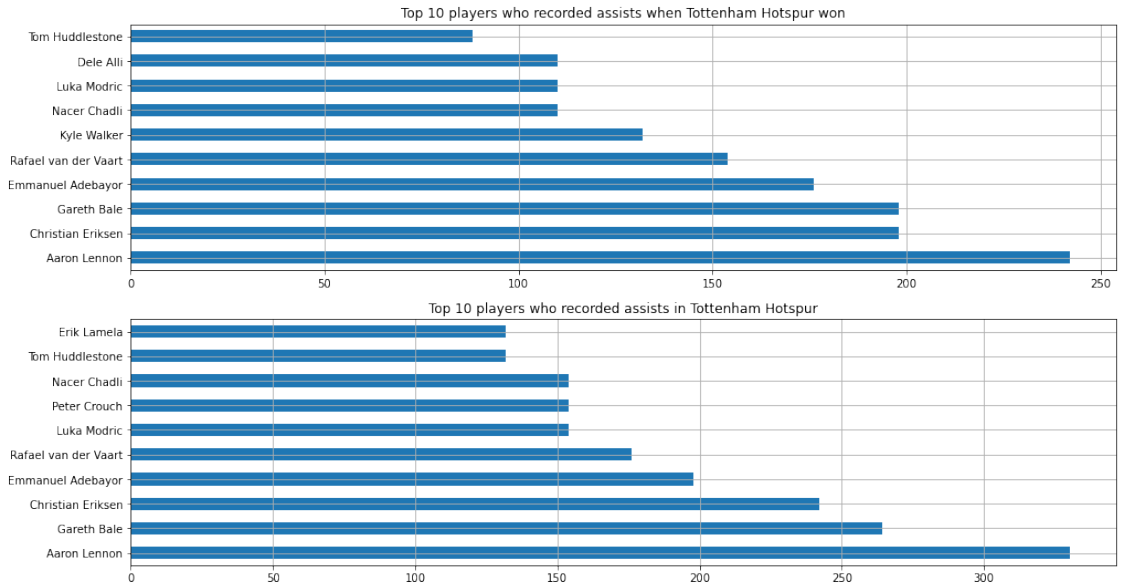
```
plt.title('Top 10 players who recorded the starting lineup when Tottenham␣
 ↪Hotspur won');

plt.subplot(2, 1, 2);
ttm_all.player_name.value_counts()[:10].plot(kind='barh', figsize = (16,9),␣
 ↪grid = True);
plt.title('Top 10 players who recorded the starting lineup in Tottenham Hotspur␣
 ↪');
```



[87]:
```
plt.subplot(2, 1, 1);
ttm_win.goal_player_name.value_counts()[:10].plot(kind='barh', figsize =␣
 ↪(16,9), grid = True);
plt.title('Top 10 players who recorded goals when Tottenham Hotspur won');

plt.subplot(2, 1, 2);
ttm_all.goal_player_name.value_counts()[:10].plot(kind='barh', figsize =␣
 ↪(16,9), grid = True);
plt.title('Top 10 players who recorded goals in Tottenham Hotspur');
```



[88]:
```
plt.subplot(2, 1, 1);
ttm_win.assist_player_name.value_counts()[:10].plot(kind='barh', figsize =␣
 ↪(16,9), grid = True);
plt.title('Top 10 players who recorded assists when Tottenham Hotspur won');

plt.subplot(2, 1, 2);
ttm_all.assist_player_name.value_counts()[:10].plot(kind='barh', figsize =␣
 ↪(16,9), grid = True);
plt.title('Top 10 players who recorded assists in Tottenham Hotspur');
```



1.1.7 Research Question 5

Because European soccer aged more than a hundred years, I believe a new team has advanced to the primary league from the secondary league or vice versa during history. With this curiosity, I want to check which club team improved the most over the period and each related ranking?

To find the answer, we sorted each team per the number of winning for each season and combined the entire season data to get a simplified DataFrame. From the new DataFrame, we sorted out the top and bottom ten teams in recording wins and compared each season. To select the most improved team, we applied conditions that the team's overall performance should not have a sudden downturn over the period. From the bar chart plotted with bivariable, the most improved team during the period was FC Bayern Munich. The team only recorded five wins in the first season. They maintained the standings without plummetting and finally recorded five times as many wins in the last season. There should be a remarkable difference between the first and last season. I realized there is no champion team last for long in the league. And every team can become the champion under the proper training because there is not a high difference in the level of each team. After all, the team's attributions are more likely standardized.

[89]:
```
# get sorted DataFrame for this research.

season0809 = pd.DataFrame(df[df['season'] == '2008/2009'].match_winner_team.
 ↪value_counts()).reset_index()
season0809 = season0809.rename(columns = {'match_winner_team' : '2008-2009'})

season0910 = pd.DataFrame(df[df['season'] == '2009/2010'].match_winner_team.
 ↪value_counts()).reset_index()
season0910 = season0910.rename(columns = {'match_winner_team' : '2009-2010'})

season1011 = pd.DataFrame(df[df['season'] == '2010/2011'].match_winner_team.
 ↪value_counts()).reset_index()
season1011 = season1011.rename(columns = {'match_winner_team' : '2010-2011'})

season1112 = pd.DataFrame(df[df['season'] == '2011/2012'].match_winner_team.
 ↪value_counts()).reset_index()
season1112 = season1112.rename(columns = {'match_winner_team' : '2011-2012'})

season1213 = pd.DataFrame(df[df['season'] == '2012/2013'].match_winner_team.
 ↪value_counts()).reset_index()
season1213 = season1213.rename(columns = {'match_winner_team' : '2012-2013'})

season1314 = pd.DataFrame(df[df['season'] == '2013/2014'].match_winner_team.
 ↪value_counts()).reset_index()
season1314 = season1314.rename(columns = {'match_winner_team' : '2013-2014'})
```

```
season1415 = pd.DataFrame(df[df['season'] == '2014/2015'].match_winner_team.
 ↪value_counts()).reset_index()
season1415 = season1415.rename(columns = {'match_winner_team' : '2014-2015'})

season1516 = pd.DataFrame(df[df['season'] == '2015/2016'].match_winner_team.
 ↪value_counts()).reset_index()
season1516 = season1516.rename(columns = {'match_winner_team' : '2015-2016'})

season0816 = pd.merge(season0809, season0910, how = 'outer')
season0816 = pd.merge(season0816, season1011, how = 'outer')
season0816 = pd.merge(season0816, season1112, how = 'outer')
season0816 = pd.merge(season0816, season1213, how = 'outer')
season0816 = pd.merge(season0816, season1314, how = 'outer')
season0816 = pd.merge(season0816, season1415, how = 'outer')
season0816 = pd.merge(season0816, season1516, how = 'outer')
```

[90]:
```
season = season0816.loc[1:, :]
season
```

[90]:

	index	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016
1	Manchester United	24.0	25.0	21.0	26.0	26.0	18.0	17.0	14.0
2	Liverpool	21.0	16.0	16.0	14.0	16.0	25.0	15.0	14.0
3	Chelsea	19.0	27.0	20.0	18.0	22.0	22.0	24.0	12.0
4	Arsenal	18.0	23.0	19.0	16.0	17.0	23.0	19.0	19.0
5	FC Barcelona	16.0	16.0	26.0	25.0	NaN	9.0	27.0	26.0
..	…	…	…	…	…	…	…	…	…
165	Frosinone	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5.0
166	Carpi	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5.0
167	Roda JC Kerkrade	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0
168	ES Troyes AC	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.0
169	De Graafschap	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.0

[169 rows x 9 columns]

[91]:
```
season.groupby('index')['2008-2009'].mean().sort_values(ascending = False).
↪head(10)
```

[91]:
```
index
Manchester United    24.0
Liverpool            21.0
Chelsea              19.0
Arsenal              18.0
FC Barcelona         16.0
Everton              14.0
Inter                13.0
Fulham               13.0
Tottenham Hotspur    12.0
Sevilla FC           12.0
Name: 2008-2009, dtype: float64
```

[92]:
```
season.groupby('index')['2015-2016'].mean().sort_values(ascending = False).
↪head(10)
```

[92]:
```
index
Paris Saint-Germain    28.0
Real Madrid CF         28.0
FC Barcelona           26.0
FC Bayern Munich       25.0
PSV                    25.0
Juventus               25.0
Borussia Dortmund      24.0
Atlético Madrid        24.0
Ajax                   23.0
Napoli                 22.0
Name: 2015-2016, dtype: float64
```

[93]:
```
season[season['index'] == 'FC Bayern Munich']
```

[93]:
```
                index  2008-2009  2009-2010  2010-2011  2011-2012  2012-2013  \
32  FC Bayern Munich        5.0       17.0       12.0       15.0        NaN

    2013-2014  2014-2015  2015-2016
32       13.0       23.0       25.0
```

melting

[94]:
```
season.columns.values
```

[94]:
```
array(['index', '2008-2009', '2009-2010', '2010-2011', '2011-2012',
       '2012-2013', '2013-2014', '2014-2015', '2015-2016'], dtype=object)
```

[95]:
```
season_vars = ['2008-2009', '2009-2010', '2010-2011', '2011-2012', '2012-2013',␣
↪'2013-2014', '2014-2015', '2015-2016']
```

[96]:
```
season.shape
```

[96]:
```
(169, 9)
```

[97]:
```
season_m = season.melt(id_vars='index',
                       value_vars=season_vars,
                       var_name='season',
                       value_name='wins')
```

[98]:
```
season_m
```

[98]:
```
                  index     season  wins
0     Manchester United  2008-2009  24.0
1             Liverpool  2008-2009  21.0
2               Chelsea  2008-2009  19.0
3               Arsenal  2008-2009  18.0
4          FC Barcelona  2008-2009  16.0
…                    …        …    …
1347          Frosinone  2015-2016   5.0
1348              Carpi  2015-2016   5.0
1349   Roda JC Kerkrade  2015-2016   4.0
1350       ES Troyes AC  2015-2016   3.0
1351      De Graafschap  2015-2016   2.0

[1352 rows x 3 columns]
```

[99]:
```
Munich = season_m[season_m['index'] == "FC Bayern Munich"]
Munich
```

[99]:
```
                  index     season  wins
31    FC Bayern Munich  2008-2009   5.0
200   FC Bayern Munich  2009-2010  17.0
369   FC Bayern Munich  2010-2011  12.0
538   FC Bayern Munich  2011-2012  15.0
707   FC Bayern Munich  2012-2013   NaN
876   FC Bayern Munich  2013-2014  13.0
1045  FC Bayern Munich  2014-2015  23.0
1214  FC Bayern Munich  2015-2016  25.0
```

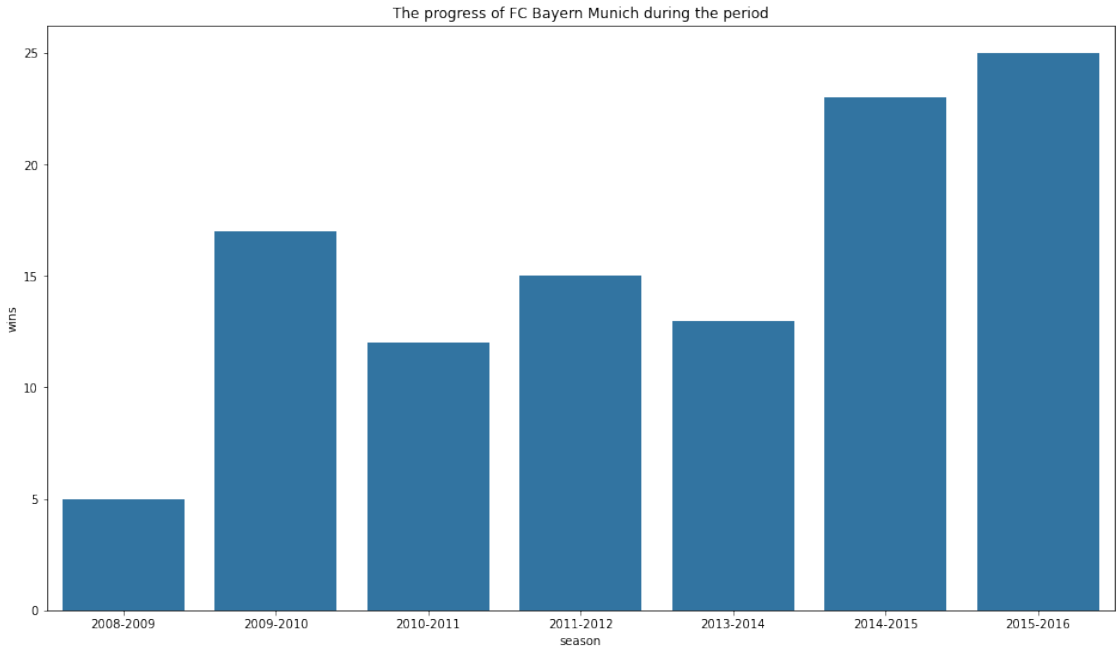
[100]:
```
Munich = Munich.dropna()
```

[101]:
```
Munich
```

[101]:
```
                  index     season  wins
31    FC Bayern Munich  2008-2009   5.0
```

```
200   FC Bayern Munich  2009-2010  17.0
369   FC Bayern Munich  2010-2011  12.0
538   FC Bayern Munich  2011-2012  15.0
876   FC Bayern Munich  2013-2014  13.0
1045  FC Bayern Munich  2014-2015  23.0
1214  FC Bayern Munich  2015-2016  25.0
```

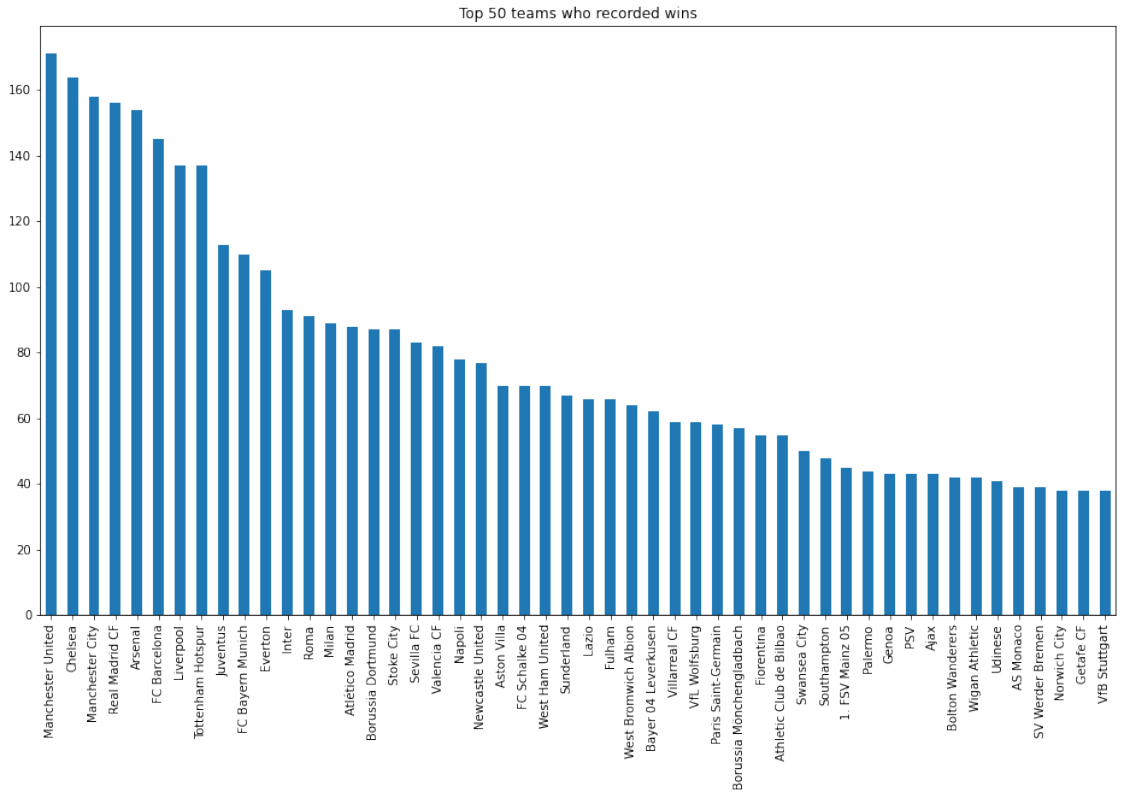
[102]:
```
plt.figure(figsize = [16,9]);
base_color = sb.color_palette()[0];

sb.barplot(data = Munich, x = 'season', y = 'wins', color = base_color);
plt.title('The progress of FC Bayern Munich during the period');
```



1.1.8 Additional finding

In the entire season, Manchester United won most; the top player recorded goals and assists are Cristiano Ronaldo and Lionel Messi, respectively.

[103]:
```
df.match_winner_team.value_counts()[1:51].plot(kind = 'bar', figsize = (16,9));
plt.title('Top 50 teams who recorded wins');
```



Manchester United won most.

[104]:
```
df.goal_player_name.nunique()
```

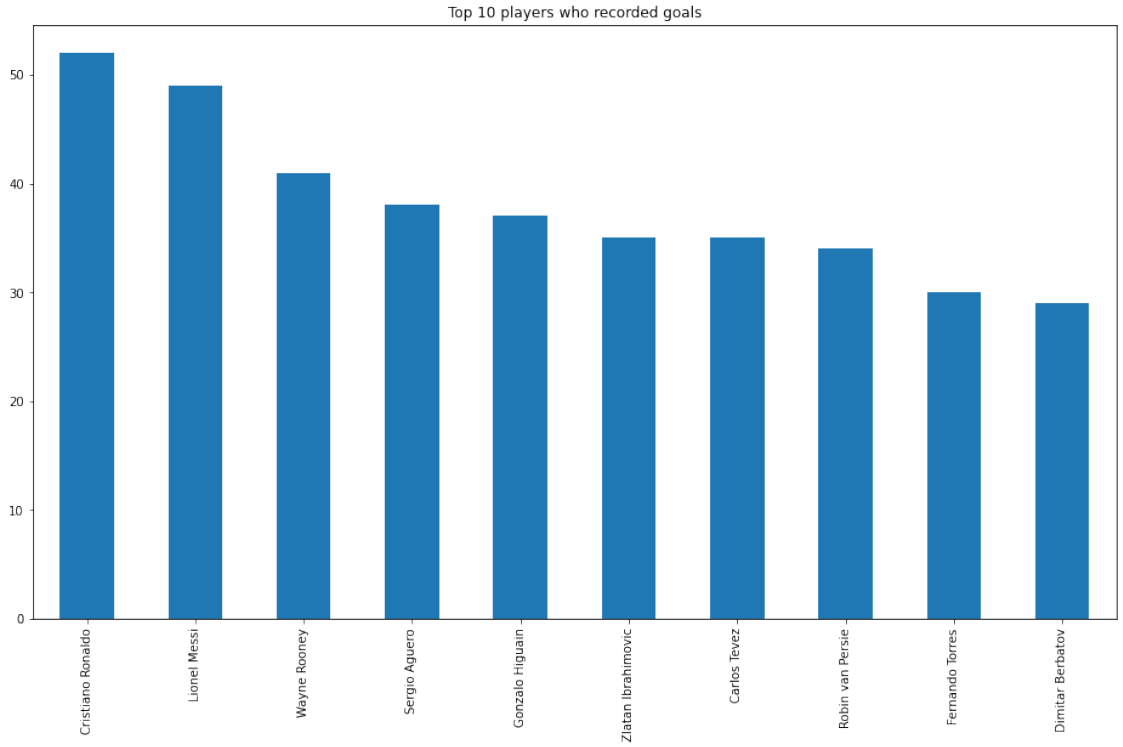
[104]:
```
2186
```

[105]:
```
df.goal_player_name.value_counts()[:25]
```

[105]:
```
Cristiano Ronaldo     52
Lionel Messi          49
Wayne Rooney          41
Sergio Aguero         38
Gonzalo Higuain       37
Zlatan Ibrahimovic    35
Carlos Tevez          35
Robin van Persie      34
Fernando Torres       30
Dimitar Berbatov      29
Aritz Aduriz          28
Darren Bent           26
Luis Suarez           25
Karim Benzema         25
```

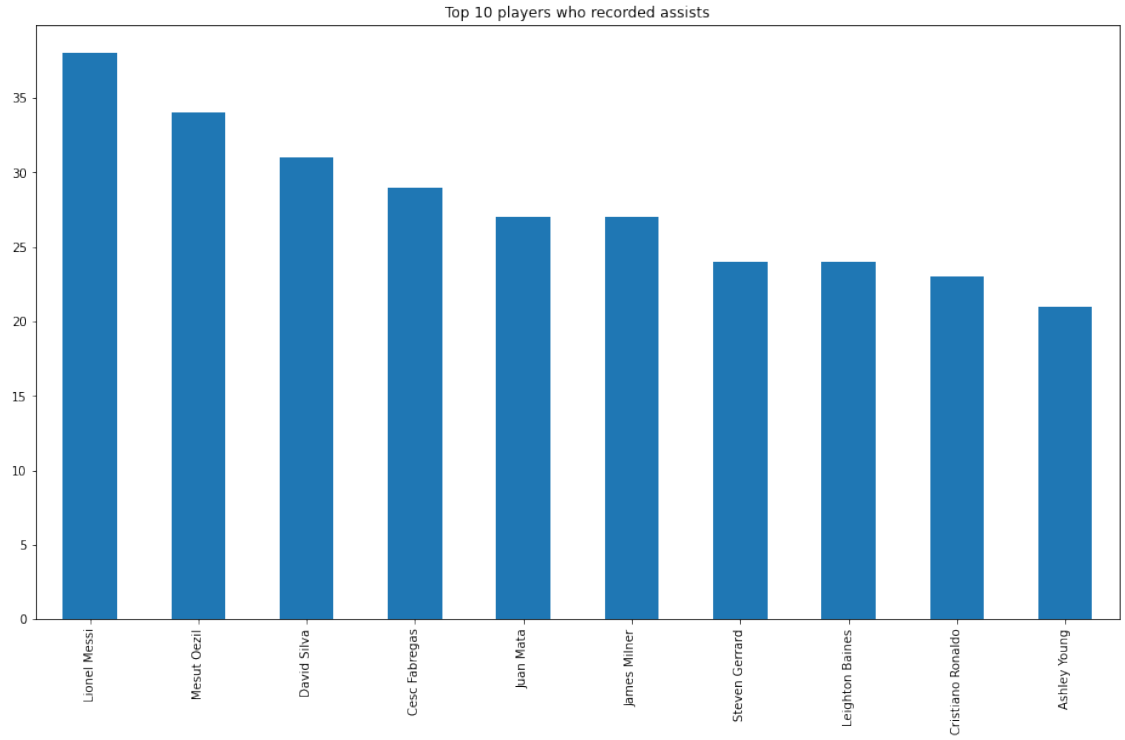
```
Edin Dzeko             24
Edinson Cavani         24
Gareth Bale            24
Emmanuel Adebayor      23
Peter Crouch           22
Gabriel Agbonlahor     20
Antonio Di Natale      20
Olivier Giroud         20
Robert Lewandowski     20
Thomas Mueller         20
Frank Lampard          19
Name: goal_player_name, dtype: int64
```

[106]:
```
df.goal_player_name.value_counts()[:10].plot(kind = 'bar', figsize = (16,9));
plt.title('Top 10 players who recorded goals');
```



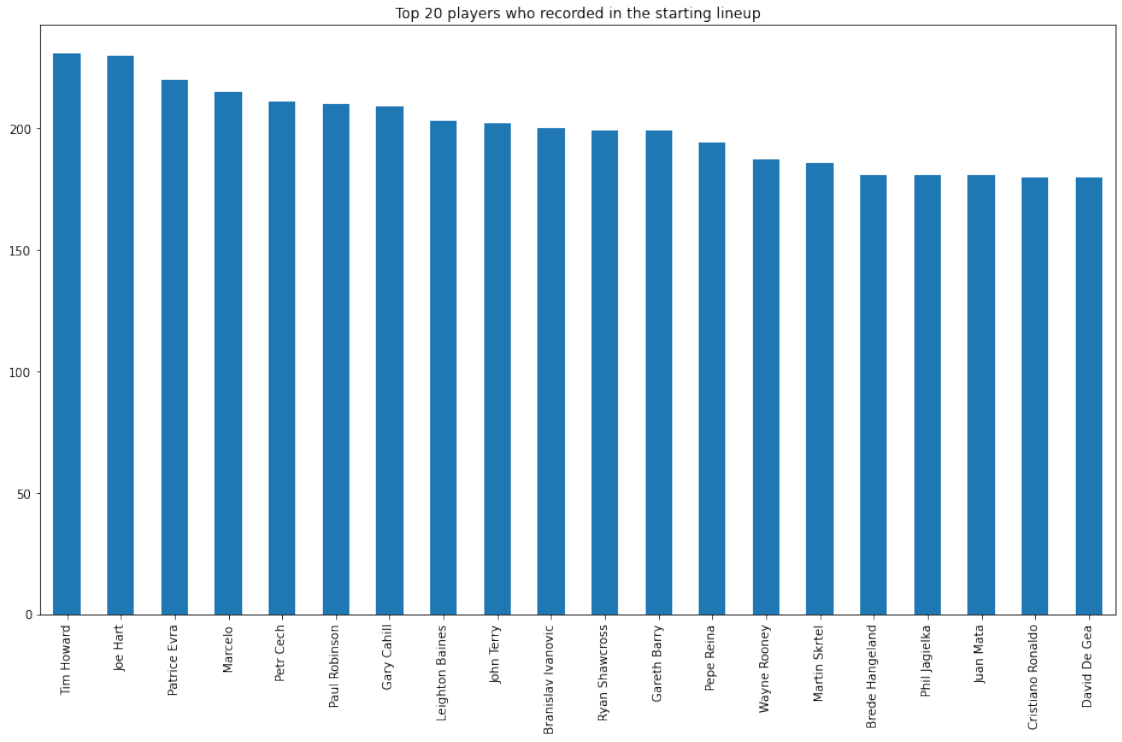
Cristiano Ronaldo recorded the most goals.

[107]:
```
df.assist_player_name.value_counts()[:10].plot(kind = 'bar', figsize = (16,9));
plt.title('Top 10 players who recorded assists');
```



Lionel Messi recorded the most assist.

[108]:
```
dfp.player_name.value_counts()[:20].plot(kind='bar', figsize = (16,9));
plt.title('Top 20 players who recorded in the starting lineup');
```



Tim Howard recorded the most starting lineup.

Conclusions

We were most interested in figuring out if a player's potential for assisting goals is high, his potential for recording goals is also high, which club team improved the most over the period, and each related ranking associated with Tottenham Hotspur. Before we write a summary, we would like to highlight that we analyzed data having 7173 lines for nine seasons of 11 countries' leagues for these questions. As we all know, the number of data matters in finding more precise answers. And also, given that a private company updated the attribute of players, it may not contain objective ratings according to the players' actual performance, as an instance, some players could suffer an injury and end the season earlier.

We conducted the multidirectional statistical approach and learned that there is a remarkable correlation between a player's assist potential and goal potential. Except for the 2013/2014 season, Tottenham Hotspur had a more winning record when they played in the home stadium than in the opponents' place. The top player who recorded in the starting lineup in Tottenham Hotspur was 'Kyle Walker', the top player who recorded goals was 'Harry Kane, and the top player who recorded assists was 'Aaron Lennon.' Besides, we also figured out FC Bayern Munich improved the most over the period. In the entire season, Manchester United won most; the top player recorded goals and assists are Cristiano Ronaldo and Lionel Messi, respectively. Additionally, we found out that the top 30 players who

recorded goals had 12 more points in overall ratings and 9 points in potentials; however, they were three years older and 3lbs heavier than the bottom 30 players who recorded goals.

In these findings, we can reduce our prejudice about the superficial aspects such as physical benefits and age in winning a game, passing skill presumably the fundamental factor to become a successful forwarder, and a home stadium benefit exists. Finally, each team has an open chance to become a champion.

2 Useful website for additional research

https://www.kaggle.com/hugomathien/soccer