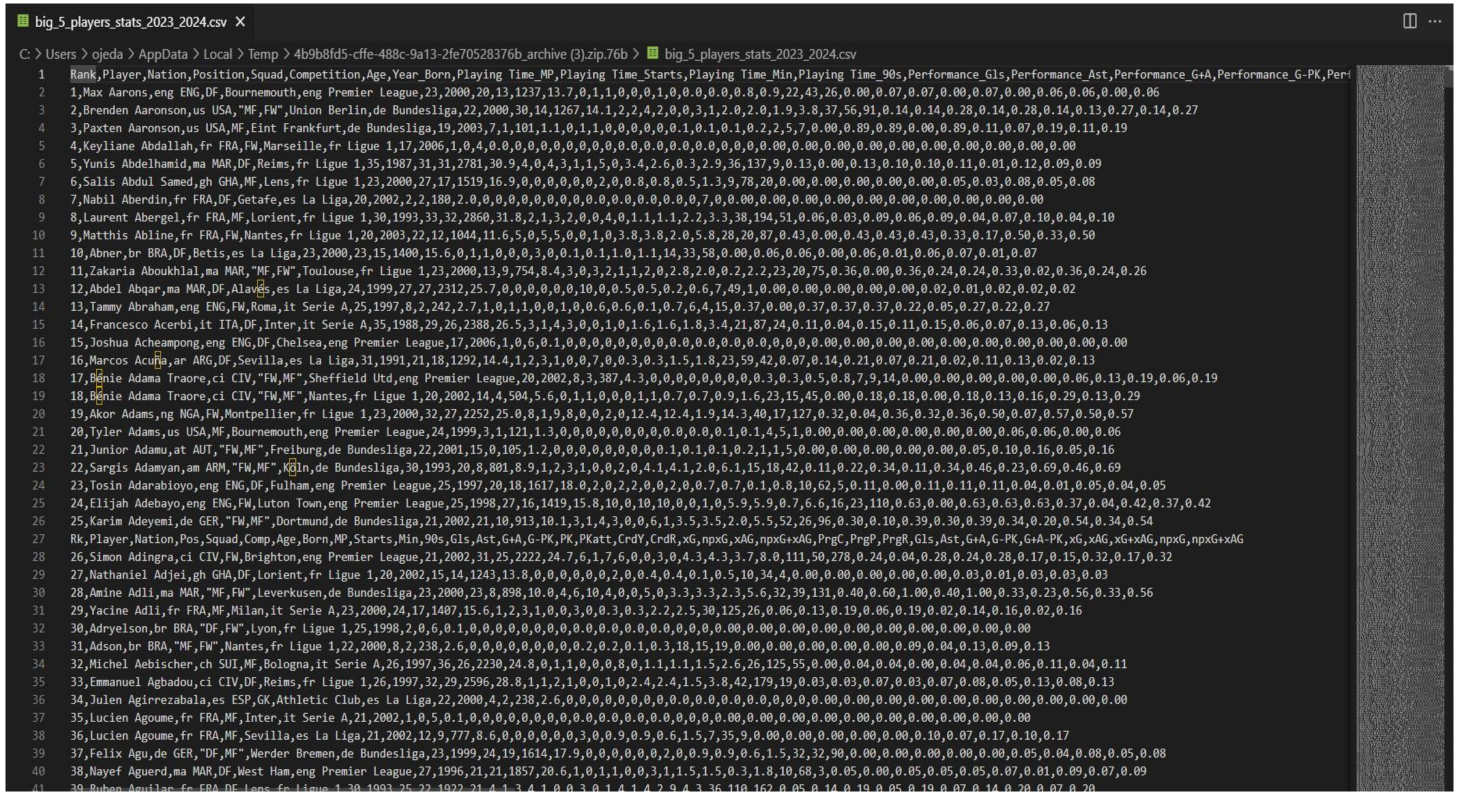
Using Statistics for Data Analysis

David Carbajal

The Data

Dataset: “big_5_players_stats_2023_2024.csv” – Kaggle

- Player Statistics from the Top 5 Leagues of Soccer in Europe from the 2023-2024 league season.
- Statistics include player name, league, position, team, goals, assists, G+A, minutes played, expected goals/assist, age, nationality, progression, etc.
- Across 2,959 players



big_5_players_stats_2023_2024.csv
C: > Users > ojeda > AppData > Local > Temp > 4b9b8fd5-cffe-488c-9a13-2fe70528376b_archive (3).zip.76b > big_5_players_stats_2023_2024.csv

```
Rank,Player,Nation,Position,Squad,Competition,Age,Year_Born,Playing Time_MP,Playing Time_Starts,Playing Time_Min,Playing Time_90s,Performance_Gls,Performance_Ast,Performance_G+A,Performance_G-PK,Perf
1,Max Aarons,eng ENG,DF,Bournemouth,eng Premier League,23,2000,20,13,1237,13.7,0,1,1,0,0,0,1,0,0.0,0.0,0.8,0.9,22,43,26,0.00,0.07,0.07,0.00,0.07,0.00,0.06,0.06,0.00,0.06
2,Brenden Aaronson,us USA,"MF,FW",Union Berlin,de Bundesliga,22,2000,30,14,1267,14.1,2,2,4,2,0,0,3,1,2.0,2.0,1.9,3.8,37,56,91,0.14,0.14,0.28,0.14,0.28,0.14,0.13,0.27,0.14,0.27
3,Paxten Aaronson,us USA,MF,Eint Frankfurt,de Bundesliga,19,2003,7,1,101,1.1,0,1,1,0,0,0,0,0,0.1,0.1,0.1,0.2,2,5,7,0.00,0.89,0.89,0.00,0.89,0.11,0.07,0.19,0.11,0.19
4,Keyliane Abdallah,fr FRA,FW,Marseille,fr Ligue 1,17,2006,1,0,4,0.0,0,0,0,0,0,0,0,0,0.0,0.0,0.0,0.0,0,0,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00
5,Yunis Abdelhamid,ma MAR,DF,Reims,fr Ligue 1,35,1987,31,31,2781,30.9,4,0,4,3,1,1,5,0,3.4,2.6,0.3,2.9,36,137,9,0.13,0.00,0.13,0.10,0.10,0.11,0.01,0.12,0.09,0.09
6,Salis Abdul Samed,gh GHA,MF,Lens,fr Ligue 1,23,2000,27,17,1519,16.9,0,0,0,0,0,0,2,0,0.8,0.8,0.5,1.3,9,78,20,0.00,0.00,0.00,0.00,0.00,0.05,0.03,0.08,0.05,0.08
7,Nabil Aberdin,fr FRA,DF,Getafe,es La Liga,20,2002,2,2,180,2.0,0,0,0,0,0,0,0,0,0.0,0.0,0.0,0.0,0,7,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00
8,Laurent Abergel,fr FRA,MF,Lorient,fr Ligue 1,30,1993,33,32,2860,31.8,2,1,3,2,0,0,4,0,1.1,1.1,2.2,3.3,38,194,51,0.06,0.03,0.09,0.06,0.09,0.04,0.07,0.10,0.04,0.10
9,Matthis Abline,fr FRA,FW,Nantes,fr Ligue 1,20,2003,22,12,1044,11.6,5,0,5,5,0,0,1,0,3.8,3.8,2.0,5.8,28,20,87,0.43,0.00,0.43,0.43,0.43,0.33,0.17,0.50,0.33,0.50
10,Abner,br BRA,DF,Betis,es La Liga,23,2000,23,15,1400,15.6,0,1,1,0,0,0,3,0,0.1,0.1,1.0,1.1,14,33,58,0.00,0.06,0.06,0.00,0.06,0.01,0.06,0.07,0.01,0.07
11,Zakaria Aboukhlal,ma MAR,"MF,FW",Toulouse,fr Ligue 1,23,2000,13,9,754,8.4,3,0,3,2,1,1,2,0,2.8,2.0,0.2,2.2,23,20,75,0.36,0.00,0.36,0.24,0.24,0.33,0.02,0.36,0.24,0.26
12,Abdel Abqar,ma MAR,DF,Alavés,es La Liga,24,1999,27,27,2312,25.7,0,0,0,0,0,0,10,0,0.5,0.5,0.2,0.6,7,49,1,0.00,0.00,0.00,0.00,0.00,0.02,0.01,0.02,0.02,0.02
13,Tammy Abraham,eng ENG,FW,Roma,it Serie A,25,1997,8,2,242,2.7,1,0,1,1,0,0,1,0,0.6,0.6,0.1,0.7,6,4,15,0.37,0.00,0.37,0.37,0.37,0.22,0.05,0.27,0.22,0.27
14,Francesco Acerbi,it ITA,DF,Inter,it Serie A,35,1988,29,26,2388,26.5,3,1,4,3,0,0,1,0,1.6,1.6,1.8,3.4,21,87,24,0.11,0.04,0.15,0.11,0.15,0.06,0.07,0.13,0.06,0.13
15,Joshua Acheampong,eng ENG,DF,Chelsea,eng Premier League,17,2006,1,0,6,0.1,0,0,0,0,0,0,0,0,0.0,0.0,0.0,0.0,0,0,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00
16,Marcos Acuña,ar ARG,DF,Sevilla,es La Liga,31,1991,21,18,1292,14.4,1,2,3,1,0,0,7,0,0.3,0.3,1.5,1.8,23,59,42,0.07,0.14,0.21,0.07,0.21,0.02,0.11,0.13,0.02,0.13
17,Bénie Adama Traore,ci CIV,"FW,MF",Sheffield Utd,eng Premier League,20,2002,8,3,387,4.3,0,0,0,0,0,0,0,0,0.3,0.3,0.5,0.8,7,9,14,0.00,0.00,0.00,0.00,0.00,0.06,0.13,0.19,0.06,0.19
18,Bénie Adama Traore,ci CIV,"FW,MF",Nantes,fr Ligue 1,20,2002,14,4,504,5.6,0,1,1,0,0,0,1,1,0.7,0.7,0.9,1.6,23,15,45,0.00,0.18,0.18,0.00,0.18,0.13,0.16,0.29,0.13,0.29
19,Akor Adams,ng NGA,FW,Montpellier,fr Ligue 1,23,2000,32,27,2252,25.0,8,1,9,8,0,0,2,0,12.4,12.4,1.9,14.3,40,17,127,0.32,0.04,0.36,0.32,0.36,0.50,0.07,0.57,0.50,0.57
20,Tyler Adams,us USA,MF,Bournemouth,eng Premier League,24,1999,3,1,121,1.3,0,0,0,0,0,0,0,0,0.0,0.0,0.1,0.1,4,5,1,0.00,0.00,0.00,0.00,0.00,0.00,0.06,0.06,0.00,0.06
21,Junior Adamu,at AUT,"FW,MF",Freiburg,de Bundesliga,22,2001,15,0,105,1.2,0,0,0,0,0,0,0,0,0.1,0.1,0.1,0.2,1,1,5,0.00,0.00,0.00,0.00,0.00,0.05,0.10,0.16,0.05,0.16
22,Sargis Adamyan,am ARM,"FW,MF",Köln,de Bundesliga,30,1993,20,8,801,8.9,1,2,3,1,0,0,2,0,4.1,4.1,2.0,6.1,15,18,42,0.11,0.22,0.34,0.11,0.34,0.46,0.23,0.69,0.46,0.69
23,Tosin Adarabioyo,eng ENG,DF,Fulham,eng Premier League,25,1997,20,18,1617,18.0,2,0,2,2,0,0,2,0,0.7,0.7,0.1,0.8,10,62,5,0.11,0.00,0.11,0.11,0.11,0.04,0.01,0.05,0.04,0.05
24,Elijah Adebayo,eng ENG,FW,Luton Town,eng Premier League,25,1998,27,16,1419,15.8,10,0,10,10,0,0,1,0,5.9,5.9,0.7,6.6,16,23,110,0.63,0.00,0.63,0.63,0.63,0.37,0.04,0.42,0.37,0.42
25,Karim Adeyemi,de GER,"FW,MF",Dortmund,de Bundesliga,21,2002,21,10,913,10.1,3,1,4,3,0,0,6,1,3.5,3.5,2.0,5.5,52,26,96,0.30,0.10,0.39,0.30,0.39,0.34,0.20,0.54,0.34,0.54
Rk,Player,Nation,Pos,Squad,Comp,Age,Born,MP,Starts,Min,90s,Gls,Ast,G+A,G-PK,PK,PKatt,CrdY,CrdR,xG,npxG,xAG,npxG+xAG,PrgC,PrgP,PrgR,Gls,Ast,G+A,G-PK,G+A-PK,xG,xAG,xG+xAG,npxG,npxG+xAG
26,Simon Adingra,ci CIV,FW,Brighton,eng Premier League,21,2002,31,25,2222,24.7,6,1,7,6,0,0,3,0,4.3,4.3,3.7,8.0,111,50,278,0.24,0.04,0.28,0.24,0.28,0.17,0.15,0.32,0.17,0.32
27,Nathaniel Adjei,gh GHA,DF,Lorient,fr Ligue 1,20,2002,15,14,1243,13.8,0,0,0,0,0,0,2,0,0.4,0.4,0.1,0.5,10,34,4,0.00,0.00,0.00,0.00,0.00,0.03,0.01,0.03,0.03,0.03
28,Amine Adli,ma MAR,"MF,FW",Leverkusen,de Bundesliga,23,2000,23,8,898,10.0,4,6,10,4,0,0,5,0,3.3,3.3,2.3,5.6,32,39,131,0.40,0.60,1.00,0.40,1.00,0.33,0.23,0.56,0.33,0.56
29,Yacine Adli,fr FRA,MF,Milan,it Serie A,23,2000,24,17,1407,15.6,1,2,3,1,0,0,3,0,0.3,0.3,2.2,2.5,30,125,26,0.06,0.13,0.19,0.06,0.19,0.02,0.14,0.16,0.02,0.16
30,Adryelson,br BRA,"DF,FW",Lyon,fr Ligue 1,25,1998,2,0,6,0.1,0,0,0,0,0,0,0,0,0.0,0.0,0.0,0.0,0,0,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00
31,Adson,br BRA,"MF,FW",Nantes,fr Ligue 1,22,2000,8,2,238,2.6,0,0,0,0,0,0,0,0,0.2,0.2,0.1,0.3,18,15,19,0.00,0.00,0.00,0.00,0.00,0.09,0.04,0.13,0.09,0.13
32,Michel Aebischer,ch SUI,MF,Bologna,it Serie A,26,1997,36,26,2230,24.8,0,1,1,0,0,0,8,0,1.1,1.1,1.5,2.6,26,125,55,0.00,0.04,0.04,0.00,0.04,0.04,0.06,0.11,0.04,0.11
33,Emmanuel Agbadou,ci CIV,DF,Reims,fr Ligue 1,26,1997,32,29,2596,28.8,1,1,2,1,0,0,1,0,2.4,2.4,1.5,3.8,42,179,19,0.03,0.03,0.07,0.03,0.07,0.08,0.05,0.13,0.08,0.13
34,Julen Agirrezabala,es ESP,GK,Athletic Club,es La Liga,22,2000,4,2,238,2.6,0,0,0,0,0,0,0,0,0.0,0.0,0.0,0.0,0,0,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00
35,Lucien Agoume,fr FRA,MF,Inter,it Serie A,21,2002,1,0,5,0.1,0,0,0,0,0,0,0,0,0.0,0.0,0.0,0.0,0,0,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00
36,Lucien Agoume,fr FRA,MF,Sevilla,es La Liga,21,2002,12,9,777,8.6,0,0,0,0,0,0,3,0,0.9,0.9,0.6,1.5,7,35,9,0.00,0.00,0.00,0.00,0.00,0.10,0.07,0.17,0.10,0.17
37,Felix Agu,de GER,"DF,MF",Werder Bremen,de Bundesliga,23,1999,24,19,1614,17.9,0,0,0,0,0,0,2,0,0.9,0.9,0.6,1.5,32,32,90,0.00,0.00,0.00,0.00,0.00,0.05,0.04,0.08,0.05,0.08
38,Nayef Aguerd,ma MAR,DF,West Ham,eng Premier League,27,1996,21,21,1857,20.6,1,0,1,1,0,0,3,1,1.5,1.5,0.3,1.8,10,68,3,0.05,0.00,0.05,0.05,0.05,0.07,0.01,0.09,0.07,0.09
39,Ruben Aguilar,fr FRA,DF,Lens,fr Ligue 1,30,1993,25,22,1922,21.4,1,3,4,1,0,0,3,0,1.4,1.4,2.9,4.3,36,110,162,0.05,0.14,0.19,0.05,0.19,0.07,0.14,0.20,0.07,0.20
```

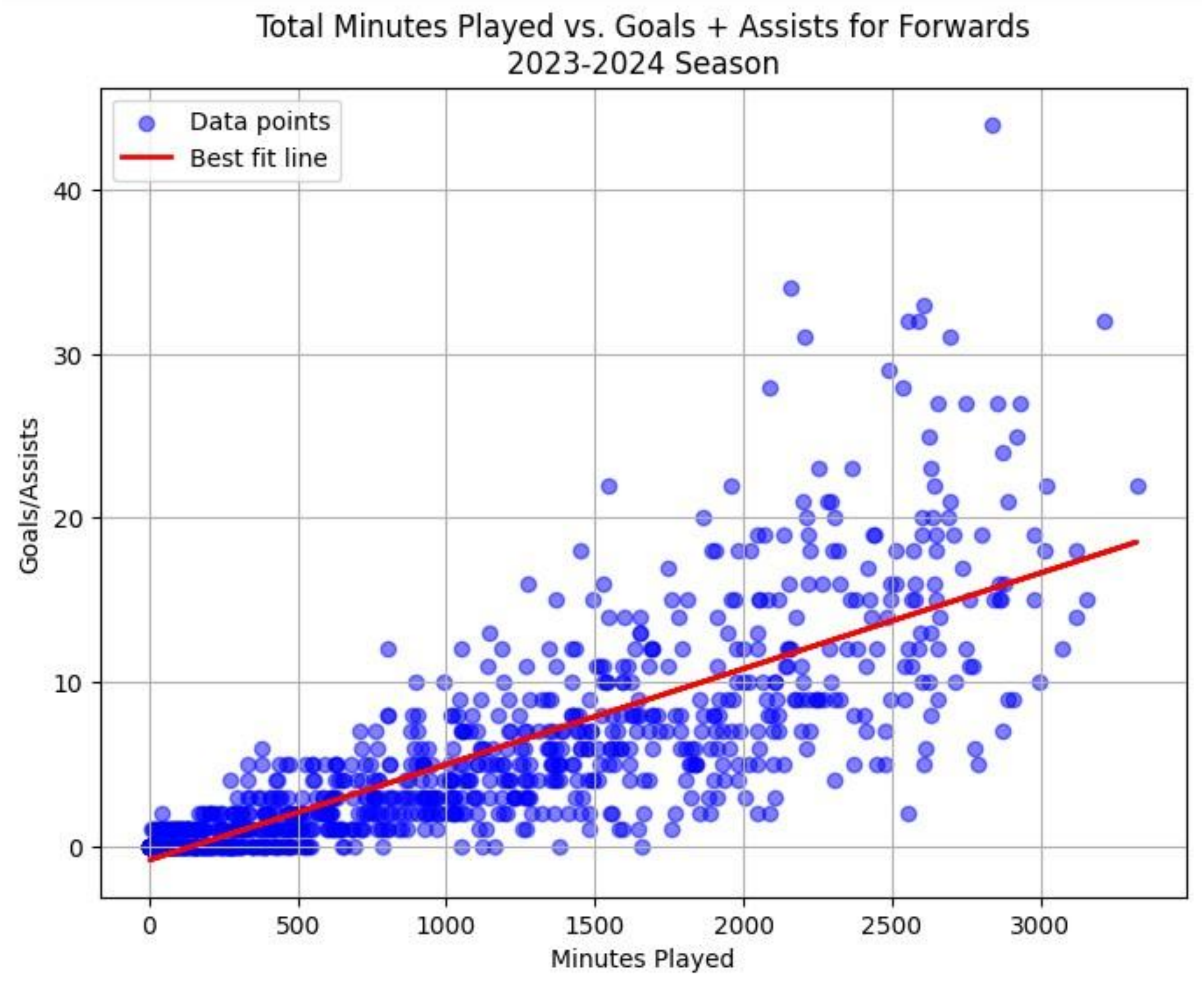
Case #1:

- Is there a correlation between total amount of minutes played and goals + assists (Strictly considering players playing in the forward positions).
- Fields Relevant: Player Name, Position, Minutes Played, Goals, and Assists

Model

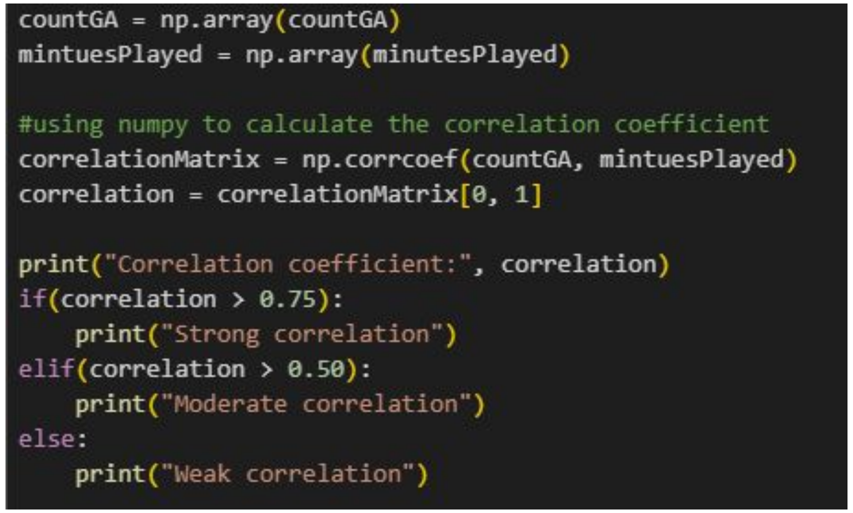
- Calculated the correlation coefficient between the minutes players and goals + assists by forwards.
- Included a scatter Plot with a graphed best fit line.
- Demonstrates the relationship between minutes played and goals/assists for forwards, where the regression line indicates the general trend.

```
Correlation coefficient: 0.8101015524402119
Strong correlation
```

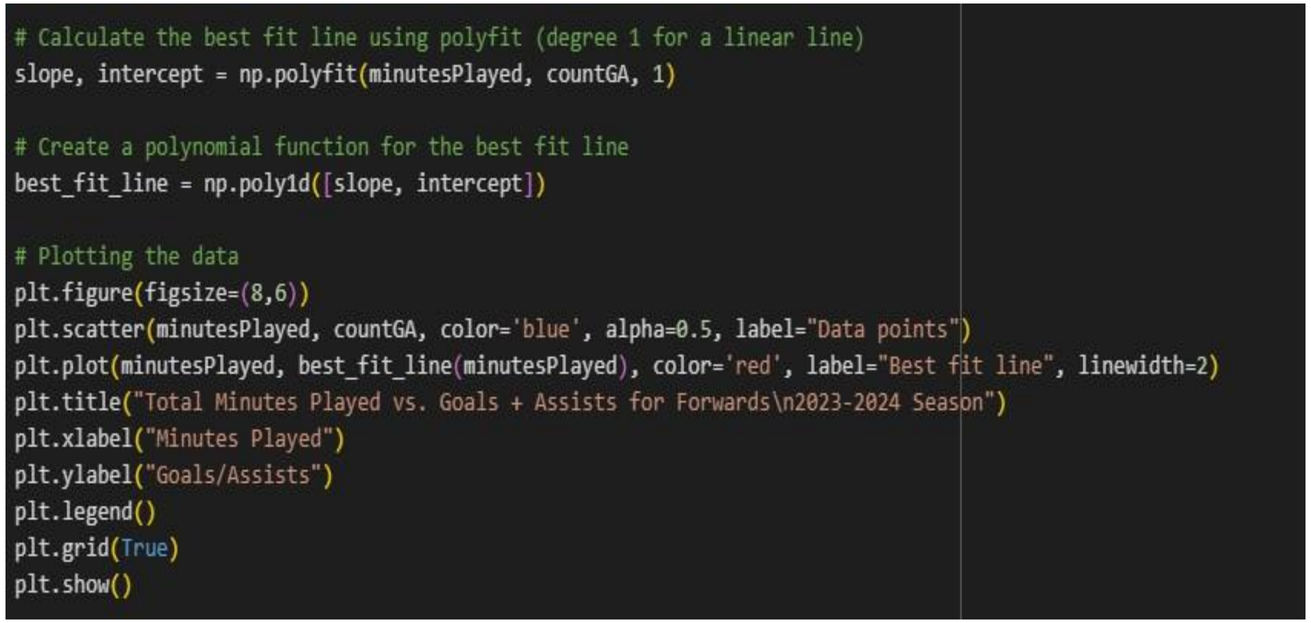


Implementation

- Python tools and libraries such as numpy, matplotlib, and csv were utilized.
- Read and uploaded data from the kaggle csv file.
- In order to compute the correlation calculation, np.corrcoef() was used.
- Visualization created through the use of matplotlib.



```
countGA = np.array(countGA)
mintuesPlayed = np.array(minutesPlayed)

#using numpy to calculate the correlation coefficient
correlationMatrix = np.corrcoef(countGA, mintuesPlayed)
correlation = correlationMatrix[0, 1]

print("Correlation coefficient:", correlation)
if(correlation > 0.75):
    print("Strong correlation")
elif(correlation > 0.50):
    print("Moderate correlation")
else:
    print("Weak correlation")
```



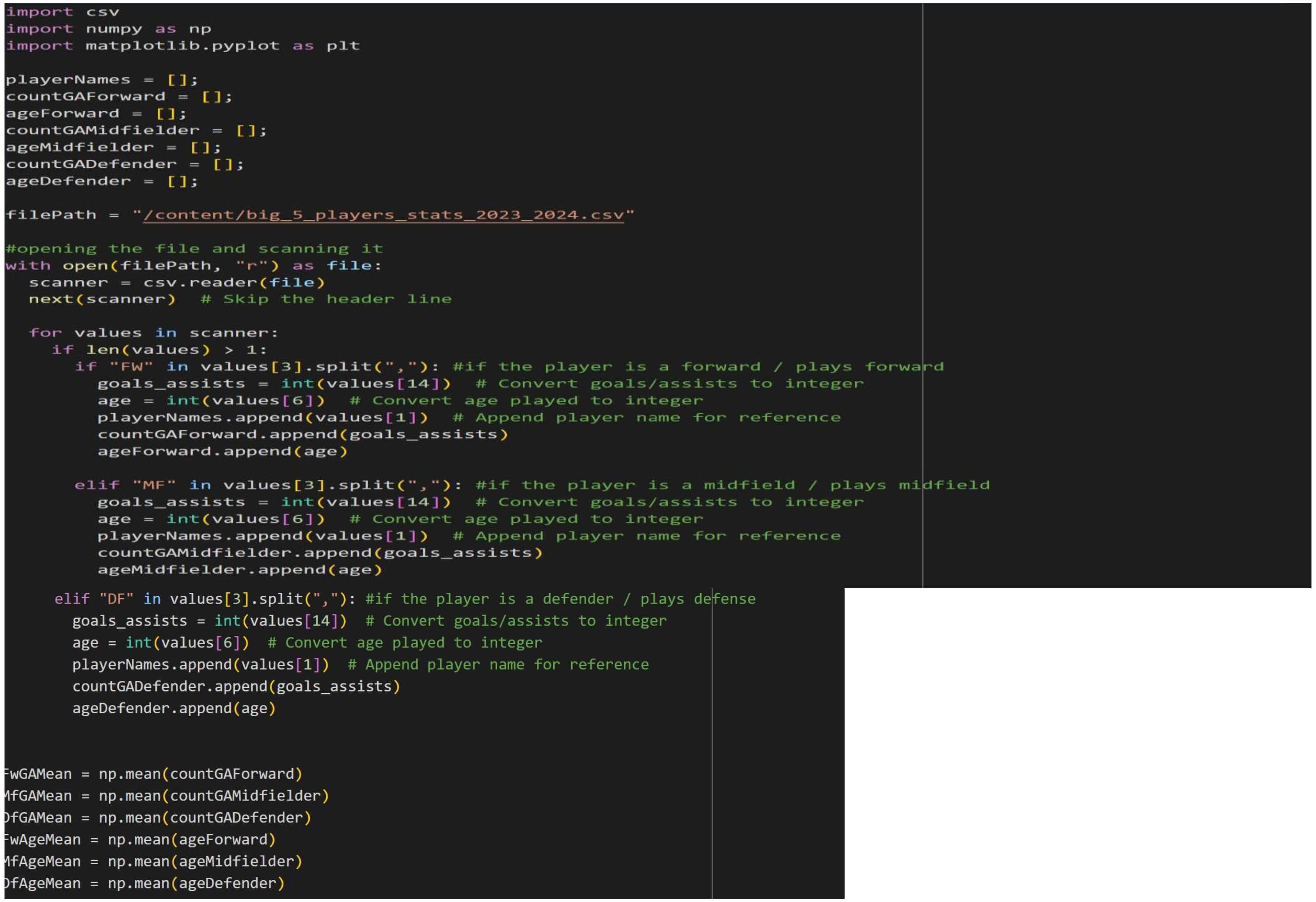
```
# Calculate the best fit line using polyfit (degree 1 for a linear line)
slope, intercept = np.polyfit(minutesPlayed, countGA, 1)

# Create a polynomial function for the best fit line
best_fit_line = np.poly1d([slope, intercept])

# Plotting the data
plt.figure(figsize=(8,6))
plt.scatter(minutesPlayed, countGA, color='blue', alpha=0.5, label="Data points")
plt.plot(minutesPlayed, best_fit_line(minutesPlayed), color='red', label="Best fit line", linewidth=2)
plt.title("Total Minutes Played vs. Goals + Assists for Forwards\n2023-2024 Season")
plt.xlabel("Minutes Played")
plt.ylabel("Goals/Assists")
plt.legend()
plt.grid(True)
plt.show()
```

Case 2:

- Finding the average goals + assists and age of the players by position. (Forward, Midfielder, and Defense)

- Finding whether or not there is a correlation between players ages, and their goals and assists. This was done by positions (Forward, Midfielder, and Defense)

Implementation of mean



```
import csv
import numpy as np
import matplotlib.pyplot as plt

playerNames = [];
countGAForward = [];
ageForward = [];
countGAMidfielder = [];
ageMidfielder = [];
countGADefender = [];
ageDefender = [];

filePath = "/content/big_5_players_stats_2023_2024.csv"

#opening the file and scanning it
with open(filePath, "r") as file:
  scanner = csv.reader(file)
  next(scanner)  # Skip the header line

  for values in scanner:
    if len(values) > 1:
      if "FW" in values[3].split(","): #if the player is a forward / plays forward
        goals_assists = int(values[14])  # Convert goals/assists to integer
        age = int(values[6])  # Convert age played to integer
        playerNames.append(values[1])  # Append player name for reference
        countGAForward.append(goals_assists)
        ageForward.append(age)

      elif "MF" in values[3].split(","): #if the player is a midfield / plays midfield
        goals_assists = int(values[14])  # Convert goals/assists to integer
        age = int(values[6])  # Convert age played to integer
        playerNames.append(values[1])  # Append player name for reference
        countGAMidfielder.append(goals_assists)
        ageMidfielder.append(age)

      elif "DF" in values[3].split(","): #if the player is a defender / plays defense
        goals_assists = int(values[14])  # Convert goals/assists to integer
        age = int(values[6])  # Convert age played to integer
        playerNames.append(values[1])  # Append player name for reference
        countGADefender.append(goals_assists)
        ageDefender.append(age)


FwGAMean = np.mean(countGAForward)
MfGAMean = np.mean(countGAMidfielder)
DfGAMean = np.mean(countGADefender)
FwAgeMean = np.mean(ageForward)
MfAgeMean = np.mean(ageMidfielder)
DfAgeMean = np.mean(ageDefender)
```

- Imported python tools and libraries such as numpy, matplotlib, and csv.
- Read data and created lists using data from the csv file.
- Calculated mean using np.mean(list).

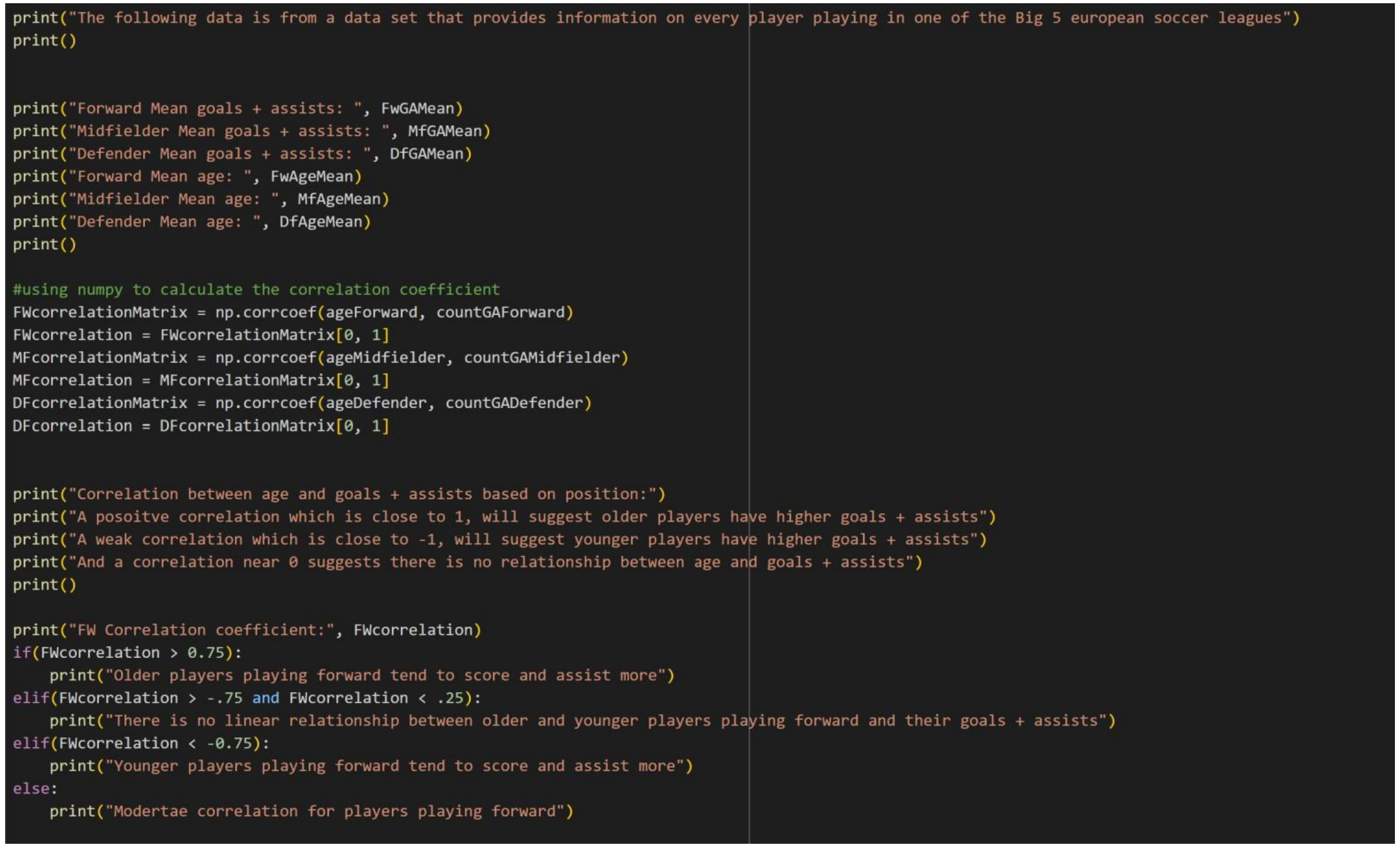
Implementation cont.

Using python libraries and previously created lists, I found the correlation between ages and goals + assists using np.corrcoef

Depending on the result, I stated whether correlation was visible or not.

If the correlation was near –1, this meant that younger players proved to score + assist more

If the correlation was near 1, this proved older players proved to score + assist more



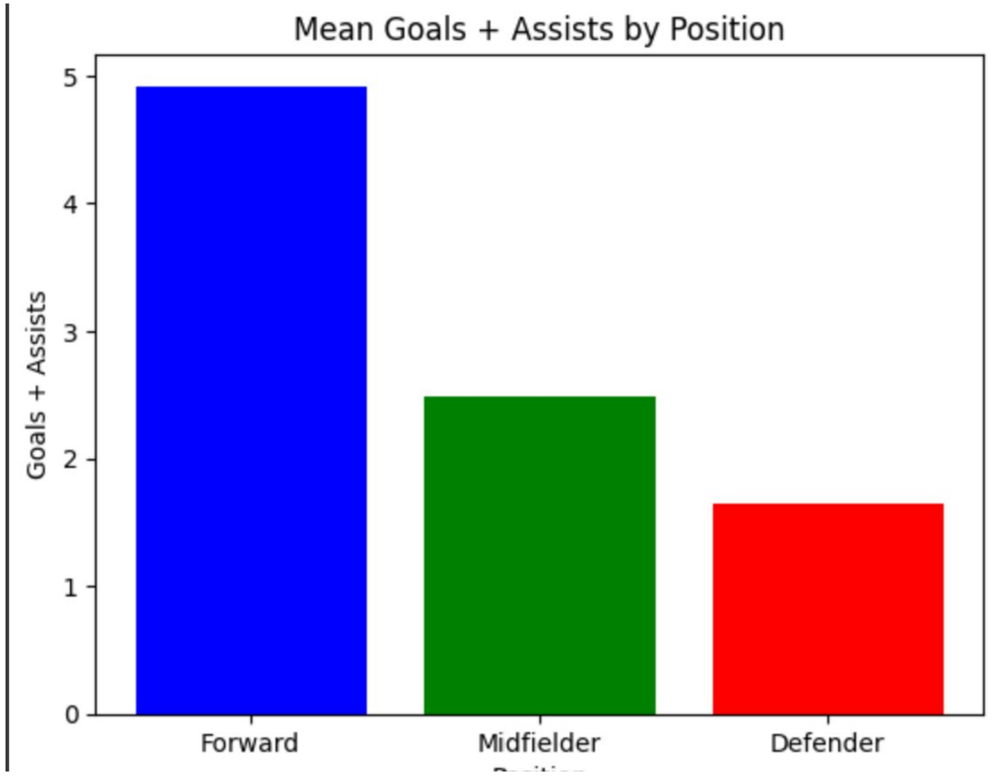
```
print("The following data is from a data set that provides information on every player playing in one of the Big 5 european soccer leagues")
print()


print("Forward Mean goals + assists: ", FwGAMean)
print("Midfielder Mean goals + assists: ", MfGAMean)
print("Defender Mean goals + assists: ", DfGAMean)
print("Forward Mean age: ", FwAgeMean)
print("Midfielder Mean age: ", MfAgeMean)
print("Defender Mean age: ", DfAgeMean)
print()

#using numpy to calculate the correlation coefficient
FWcorrelationMatrix = np.corrcoef(ageForward, countGAForward)
FWcorrelation = FWcorrelationMatrix[0, 1]
MFcorrelationMatrix = np.corrcoef(ageMidfielder, countGAMidfielder)
MFcorrelation = MFcorrelationMatrix[0, 1]
DFcorrelationMatrix = np.corrcoef(ageDefender, countGADefender)
DFcorrelation = DFcorrelationMatrix[0, 1]


print("Correlation between age and goals + assists based on position:")
print("A posoitve correlation which is close to 1, will suggest older players have higher goals + assists")
print("A weak correlation which is close to -1, will suggest younger players have higher goals + assists")
print("And a correlation near 0 suggests there is no relationship between age and goals + assists")
print()

print("FW Correlation coefficient:", FWcorrelation)
if(FWcorrelation > 0.75):
    print("Older players playing forward tend to score and assist more")
elif(FWcorrelation > -.75 and FWcorrelation < .25):
    print("There is no linear relationship between older and younger players playing forward and their goals + assists")
elif(FWcorrelation < -0.75):
    print("Younger players playing forward tend to score and assist more")
else:
    print("Modertae correlation for players playing forward")
```

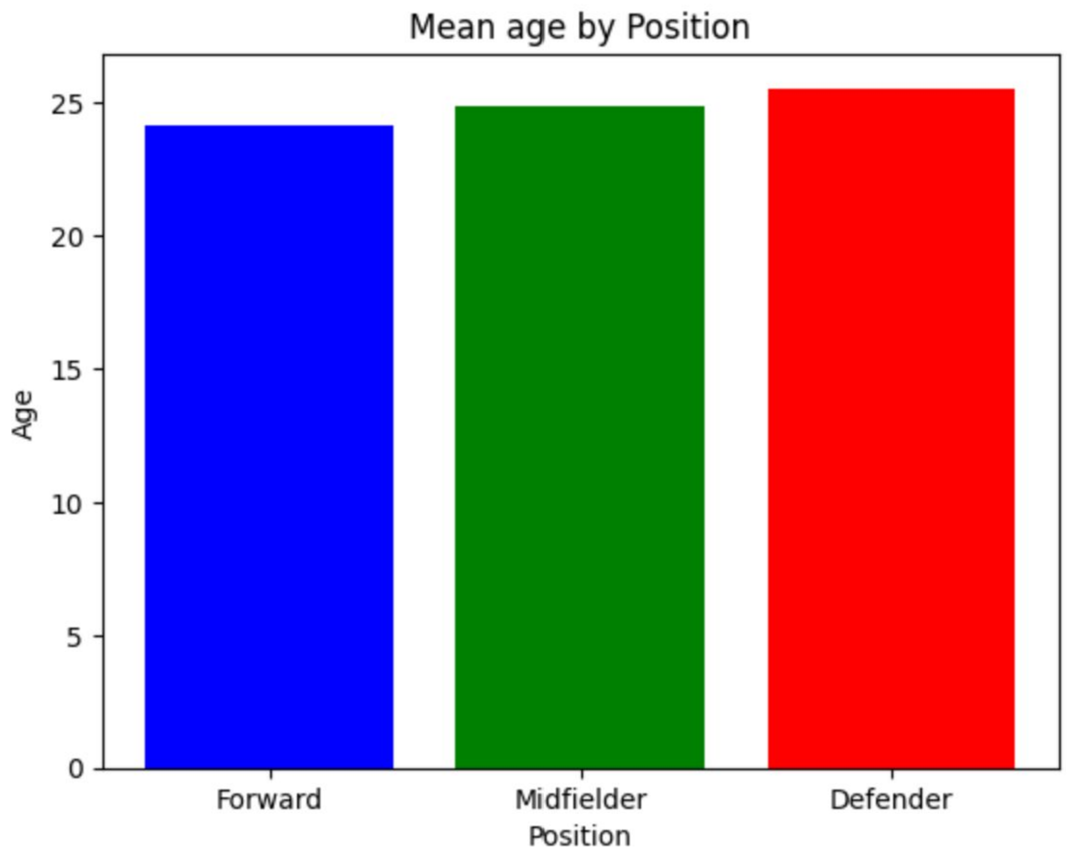
Results:

- Forward Mean goals + assists: 4.917307692307692
- Midfielder Mean goals + assists: 2.4788359788359786
- Defender Mean goals + assists: 1.6477541371158393
- Forward Mean age: 24.139423076923077
- Midfielder Mean age: 24.8505291005291
- Defender Mean age: 25.541371158392433

Models

- Mean goals + assist and age models by position





Results Cont.

- No linear relationship was found between players age and their goals + Assists.

- A correlation near 0 suggests there is no relationship between age and goals + assists

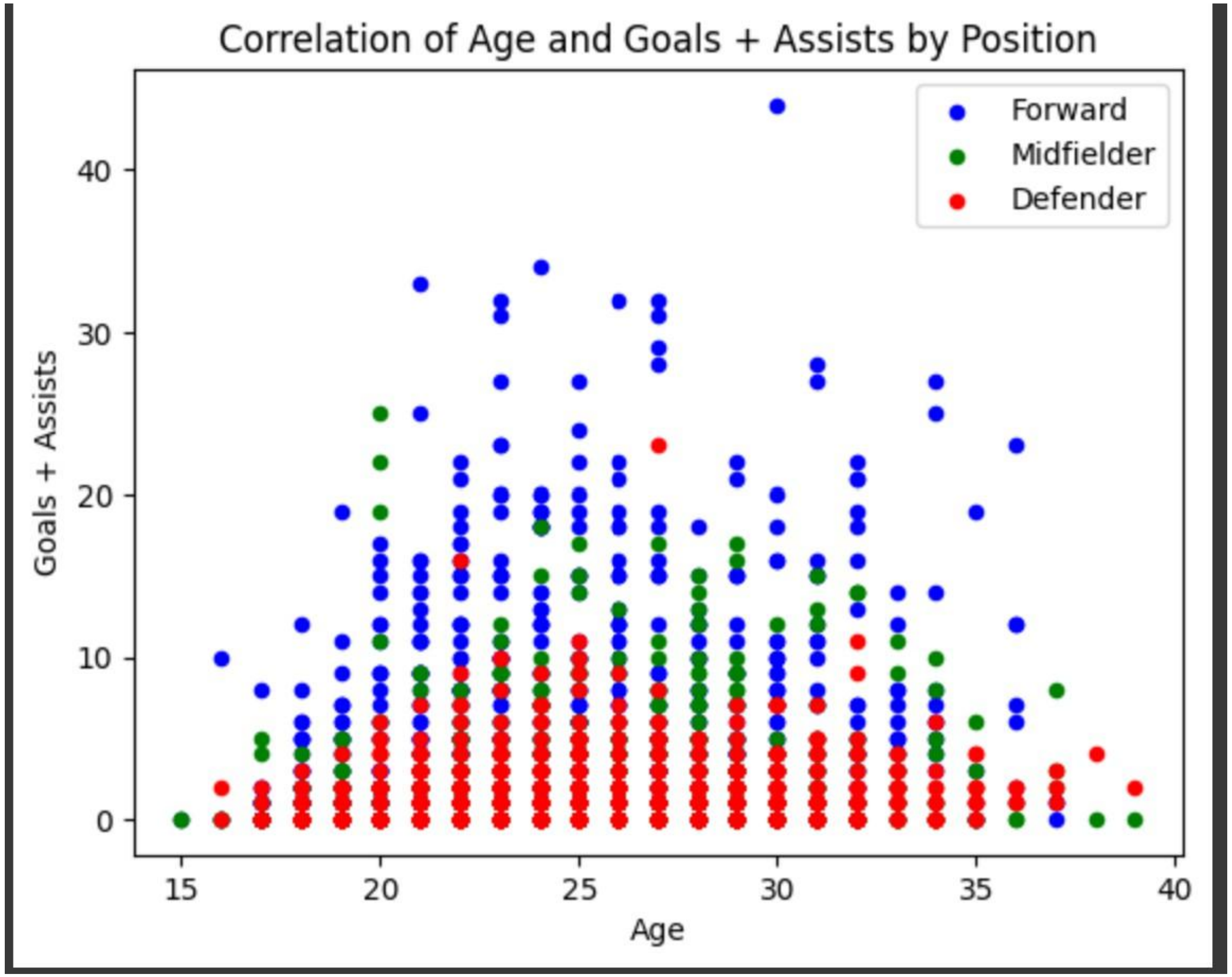
- Forward Correlation coefficient: 0.22912766798821843
- There is no linear relationship between older and younger players playing forward and their goals + assists.

- Midfielder Correlation coefficient: 0.1756854272595472
- There is no linear relationship between older and younger players playing midfield and their goals + assists.

- Defender Correlation coefficient: 0.07301425197406965
- There is no linear relationship between older and younger players playing defense and their goals + assists.

Model

Demonstrates no correlation between ages and goals + assists


Correlation of Age and Goals + Assists by Position
Forward
Midfielder
Defender
Goals + Assists
Age