CS 241

**12 Prove Milestone - Data Analysis**

**PART I**

*Jose Israel Carmona Morales*

To initialize the libraries and load the data:

import pandas as pd #The library that manipulates our data

import seaborn as sns #Used to plotting and graphing

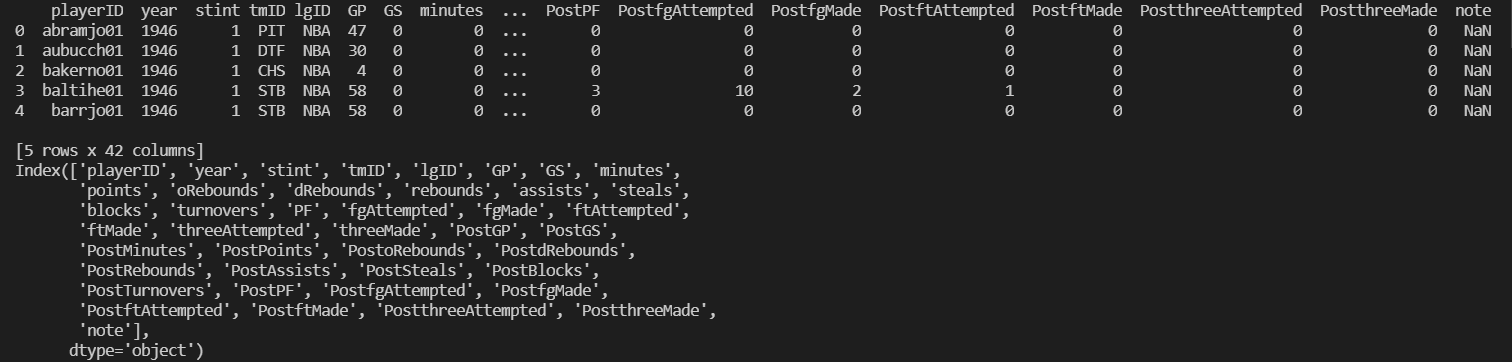
import matplotlib.pyplot as plt #If we need any low level methods

bestb\_players = pd.read\_csv("basketball\_players.csv")

print(bestb\_players.head())

print(bestb\_players.columns)

**OUTPUT**

****

**REQUIREMENT 01.**

It finds the mean and the median of numbers of points scored.

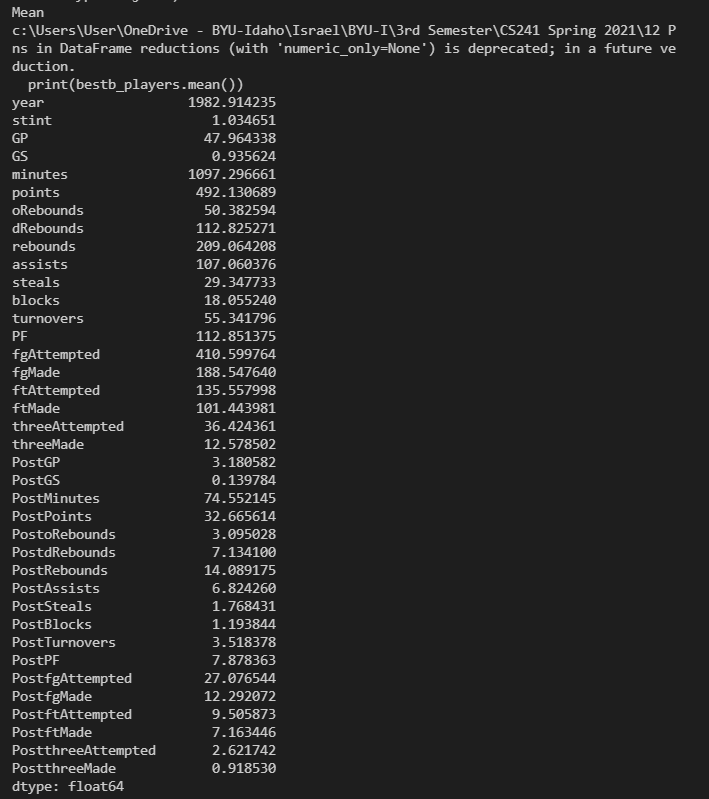
print("Mean")

print(bestb\_players.mean())

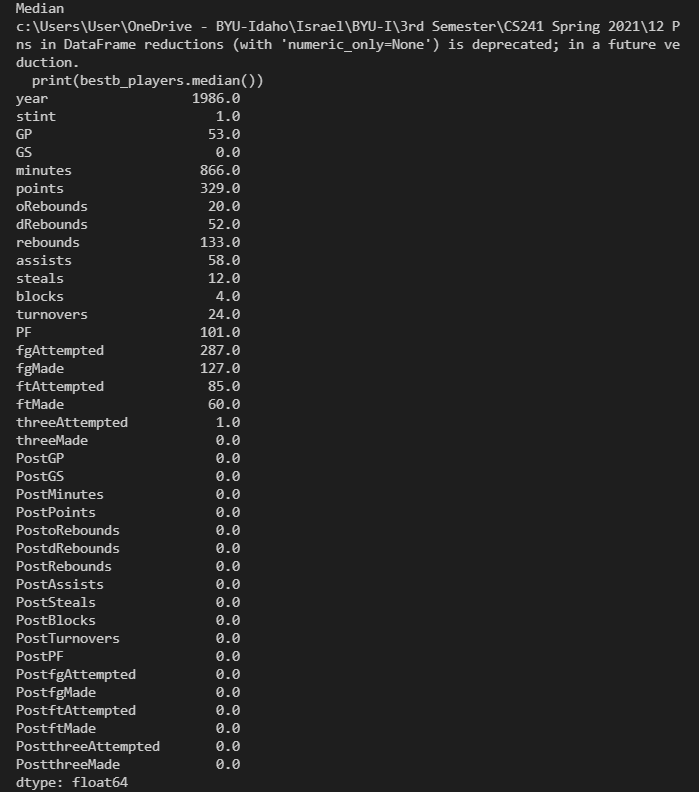
print("Median")

print(bestb\_players.median())

**MEAN**



**MEDIAN**

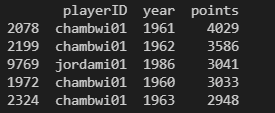


**REQUIREMENT 02.**

It finds the highest number of points per season, sorted the data by points, year and shows the highest 5.

print(bestb\_players[["playerID", "year", "points"]].sort\_values("points", ascending = False).head(5))

**OUTPUT**



**REQUIREMENT 03.**

Produces a boxplot that shows the distribution of total points, total assists, and total rebounds.

sns.boxplot(data = bestb\_players[["rebounds", "points", "assists"]])

plt.show()

**OUTPUT**

**Chart, box and whisker chart

Description automatically generated**

**REQUIREMENT 04.**

Produces a boxplot that shows the distribution of total points, total assists, and total rebounds.

year\_nbagroup = bestb\_players[["points", "year"]].groupby("year").median()

print(year\_nbagroup.head())

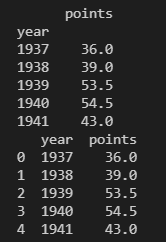
year\_nbagroup = year\_nbagroup.reset\_index()

print(year\_nbagroup.head())

sns.scatterplot(data = year\_nbagroup, x = "year", y = "points")

plt.show()

**OUTPUT**

 **Chart, scatter chart

Description automatically generated**