**Final Update Description**

DJ Bryan

Bryan Pham JaeEun Lee David Gigremosa

The Project Description

This project is comparing salaries of professional sports players in three major fields, the National Basketball Association, the National Football League, and Major League Baseball. This code displays a bar chart to visualize salaries and can be stacked over and over to compare them together in an easier way. The user can continuously search up player names, granted they are contained within the player database we use, in any order to compare against one another. The code also checks to see if the player is in the database and informs the user if they are not or if their name or affiliation was spelled incorrectly.

Our project code finds the highest 10 earning players from all three professional leagues and shows in the bar chart. This bar chart is displayed as soon as the program is run and the user can choose to close this chart in order to create their own individualized bar chart with players of their choice. If users want to compare to more players, they enter the name and affiliation as instructed by the prompts. Then it searches correct data and stacks in the chart as it looks comparable with each other.

How To Run The Code

***Important*: This project uses scipy, pandas, numpy and pandas**

1. In Terminal, type ‘python or python3 FinalProject’ and hit enter, or just click the green triangle button on the right top. If you don’t have python-scipy use the follow commands to download and install scipy  
     
    sudo apt-get install python-scipy

pip install scipy

Source: <https://stackoverflow.com/questions/24808043/importerror-no-module-named-scipy>

1. Top 10 bar-chart will be generated and close the chart after checking it.
   1. The five number summaries are also generated in the terminal
2. Enter the player name of your choice in the terminal as follow the instruction.
3. Enter the player’s respective affiliation of the typed player above.
4. All of the things that the user typed should have no error with spells or empty space.
5. The new bar-chart will be generated and close the window after check.
6. Type ‘quit’ to stop the program or just press any key to continuer.
7. The bar chart can indefinitely add new player and their salary info

How to interpret the data

first plot

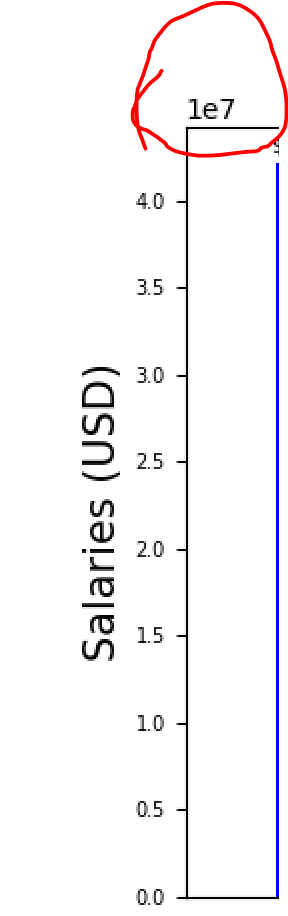
- a graph that shows the top 10 highest players of all the analyzed leagues.

- each of the bars has a label that shows how much the person earns

- the x-axis shows the players names

- the y-axis shows the salary of players in US Dollars

Once the player salaries get into the millions, there will be a 1e6 or 1e7 which represents the number of the bar chart times 10^6 or 10^7



- There is also a legend that shows what color each affiliation has, with red representing the NBA players, green representing the NFL players, and blue representing the MLB players

Finally, in the terminal output there are five number summaries for

MLB

NBA

NFL

All leagues

The second plot that appears has the same y and x-axis as the first plot

Only differences are that the chart also displays the of the player based on their salary and has no 5 number summaries

Bibliography

<https://stackoverflow.com/questions/50649853/trying-to-merge-2-dataframes-but-get-valueerror>

Referenced this to convert the entire column to a desired data type

nba\_df['Salary'] = nba\_df['Salary'].astype(float)

mlb\_df['Salary'] = mlb\_df['Salary'].astype(float)

nfl\_df['Salary'] = nfl\_df['Salary'].astype(float)

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.sort_values.html>

Referenced this to sort data in a data frame in descending order

merge\_data = merge\_data.sort\_values(by = ['Salary'], ascending = False)

<https://pandas.pydata.org/pandas-docs/stable/user_guide/text.html>

Clean NFL data of commas and dollar signs

df['Salary'] = df['Salary'].str.replace('$', '')

df['Salary'] = df['Salary'].str.replace(',', '')

df['Salary'] = df['Salary'].str.strip()

<https://stackoverflow.com/questions/7837722/what-is-the-most-efficient-way-to-loop-through-dataframes-with-pandas>

Referenced this to loop through the "Names" column in our data

for index, row in df.iterrows():

<https://stackoverflow.com/questions/21800169/python-pandas-get-index-of-rows-which-column-matches-certain-value>

Find an index of a found value

row = find\_player.index[find\_player['Name'] == player]

https://matplotlib.org/3.1.0/api/\_as\_gen/matplotlib.pyplot.bar.html

Make a bar chart using matplotlib

plt.figure(figsize=(13, 5))

labels = plt.bar(x, y, color = colors)

https://stackoverflow.com/questions/34001751/python-how-to-increase-reduce-the-fontsize-of-x-and-y-tick-labels

Adjusts the size of x and y ticks

plt.tick\_params(axis="x", labelsize= 7)

plt.tick\_params(axis="y", labelsize= 7)

<https://stackoverflow.com/questions/712082/barchart-sizing-of-text-barwidth-with-matplotlib-python>

Adjust the font-size of font labels

plt.ylabel('Salaries (USD)', fontsize = 15)

plt.xlabel('Player Names', fontsize = 15)

plt.suptitle('Ten highest base salaries in the NBA, NFL, and MLB', fontsize = 20)

<https://matplotlib.org/3.1.0/tutorials/intermediate/legend_guide.html>

Creates a legend that shows a color for each of the affiliations

NBA\_patch = mpatches.Patch(color='red', label='NBA')

NFL\_patch = mpatches.Patch(color='green', label='NFL')

MLB\_patch = mpatches.Patch(color='blue', label='MLB')

autolabel(), autolabel\_1()

Entire function from: <https://matplotlib.org/3.1.1/gallery/lines_bars_and_markers/barchart.html>

Labels each bar with a dollar value and percentile

<https://www.geeksforgeeks.org/scipy-stats-zscore-function-python/>

Calculates the z-score from the input data

z\_score = (salary - collective\_mean[0]) / collective\_stdev[0]

percentile = st.norm.cdf(z\_score)