Homework 2

Due: April 21, 2020, 6:00 AM (in the Morning)

For this homework analyze the given datasets in python to graphically answer the questions listed below. You should use Plotly for plotting unless otherwise specified. Your submission should include the Jupyter notebook of your code with respective report and all the plots saved as .html in plots folder. The submission should be one .zip file named NAME_SURNAME_HW_2.zip (for example Poghos_Poghosyan_HW_2.zip).

Problem 1:

The data in hw2_women_nba.csv is statistics calculated from women NBA player for a given year. Most of the features are explained here: https://en.wikipedia.org/wiki/Basketball_statistics?fbclid=IwAR3q38bugupLrsRVwY6Lq4vuXt47Xfda3JAMbYhBx6wvSPz-pUEd0189agg.

a) Plot the desirable position in each team. Where desirability is measured by percentage of free throws made.

b) For each team calculate the number of games played per player in that team. Plot the result.

c) For each team calculate the average experience. Plot the result using bar chart. The teams with the most and least experience should be highlighted by green and red colors respectively. The rest of the teams should have blue color.

d) Plot each team's average for free throw made and attempted as a stacked bar chart.

e) For each player calculate their efficiency calculated as https://en.wikipedia.org/wiki/Efficiency_(basketball). Plot the best 5 and the worst 5 players, sorted from the best to the worst. The 5 best players should be colored green and the worst 5 players colored red.

f) Using the results from point e) calculate the average efficiency of team. Plot the result as bar chart. Sort the bars by the average age of the player per team from youngest to the oldest.

g) For each player calculate the following statistic: number of years playing divided by their age. Create a scatterplot using the calculated statistic and efficiency calculated in point e).

Problem 2:

Chose two stocks different from each other and different from the one chosen in problem 2. Using pandas_datareader library download daily stock price data from yahoo finance for 1 year for the chosen stocks. Calculate the daily return for each of the stocks.

a) Plot the daily return for the 2 stocks in one plot

b) Plot the daily return for the 2 stocks in 2 plots (subplots)

Assume you have bought 300 USD worth of stock 1 and 700 USD worth of stock 2, and sold your holdings one year later. Calculate the value of your holdings (in USD) for each stock over time. (For example the day you bought the stocks your holding for stock 1 was 300, and the value for stock 2 was 700. Assume the price of stock 1 went up by 5% and the price of stock 2 went down by 10%, then your holding for stock 1 will be 315, and the holding for stock 2 will be 670.)

c) Plot the value of your holdings over time in 2 plots (subplots) as area chart.

d) Plot the value of your holdings over time in 1 plot as stacked area chart.

e) Calculate what percentage of your total holding each stock represented over time. For example the day you bought the stocks your total holding was 300 + 700 = 1000 USD, Stock 1 represented 30% of your total holding (100 * 300/1000), and stock 2 represented 70% of your total holding (100 * 70/1000). Plot the result as percentage stacked area chart.

Problem 3:

The data in hw2_worldbank_gdp.csv file is yearly GPD in USD for each country. The data spans from 1960 to 2017. The data is downloaded from https://data.worldbank.org. Take data for years 2009 and 2010. Keep only the countries that have no missing data, discard the rest. Calculate the GPD percentage change for each country from 2009 to 2010. Using the GDP percentage change create a Choropleth map. The color map should be centered at 0 (white), all positive numbers should have a shade of red (the higher the absolute value the more should the color intensity be), and all negative numbers should have a shade of blue (the higher the absolute value the more should the color intensity be)