COLLEGE OF ENGINEERING
COMPUTER SCIENCE
VIRGINIA TECH

Information Visualization

CS 5764

Homework # 1

In this assignment, you will learn how to connect to Yahoo stock API using Python Pandas package & explore the dataset by calculating some statistics explained in class. [display all calculated numbers with 2-digit decimal precision].

Write the program for this assignment for a general case that works if the number of features is increased or if more companies are added.

For the table questions, use either the 'tabulate' or 'prettytable' package.

1. Using the pandas_datareader package connect to yahoo database and load the stock value for the following giant companies. [6 pts]

   ```
   stocks = ['AAPL','ORCL', 'TSLA', 'IBM','YELP', 'MSFT']
   ```

   You will need the following package to be able to connect to yahoo API. Make sure to use the updated version of the pandas and pandas' data_reader (You can use the "pip install --upgrade pandas" and "pip install --upgrade pandas-datareader"). Pick the start date as '2013-01-01' and the end date is May 22nd , 2024.

   ```
   from pandas_datareader import data
   import yfinance as yf
   yf.pdr_override()
   df = data.get_data_yahoo('AAPL', start='', end='')
   ```

2. The database contains the stock values of 6 major giant companies. Each company dataset contains 6 features: "High", "Low", "Open", "Close", "Volume", "Adj Close" in USD($). Load the data set and create a table as shown below for the mean of each attribute. Display the table of the console. There are multiple ways to create a table in python. Pick a method of your choice. [16 pts]

Mean Value comparison						
Feature / Name	High ($)	Low ($)	Open ($)	Close ($)	Volume	Adj Close ($)
'AAPL'						
'ORCL'						
'TSLA'						
'IBM'						
'YELP'						
'MSFT'						
Maximum Value						
Minimum Value						
Maximum company name						
Minimum company name						

3. Repeat question 2 for the variance. [16 pts]

Variance comparison						
Feature / Name	High($)	Low($)	Open($)	Close($)	Volume	Adj Close($)
'AAPL'						
'ORCL'						
'TSLA'						
'IBM'						
'YELP'						
'MSFT'						
Maximum Value						
Minimum Value						
Maximum company name						
Minimum company name						

4. Repeat question 2 for the std. [16 pts]

Standard Deviation Value comparison						
Feature / Name	High($)	Low($)	Open($)	Close($)	Volume	Adj Close($)
'AAPL'						
'ORCL'						
'TSLA'						
'IBM'						
'YELP'						
'MSFT'						
Maximum Value						
Maximum Value						
Maximum company name						
Minimum company name						

5. Repeat question 2 for the median. [16 pts]

Median Value comparison						
Feature / Name	High($)	Low($)	Open($)	Close($)	Volume	Adj Close($)
'AAPL'						
'ORCL'						
'TSLA'						
'IBM'						
'YELP'						
'MSFT'						
Maximum Value						
Minimum Value						
Maximum company name						
Minimum company name						

6. Calculate the correlation matrix for the Apple company with all the given features. Display the correlation matrix on the console. Hint. You may use .corr() for the calculation of correlation matrix. Write down your observation about the correlation matrix. [4 pts]

7. Repeat question 6 for `'ORCL', 'TSLA', 'IBM','YELP', 'MSFT'.` [20 pts]

8. If you are a financial advisor and asked to recommend a safe investment based on stock volatility [standard deviation] to an investor, then which stock of company will you recommend for investment based on this dataset. Explain your answer and mention the name of the company that is riskier for investment and the name of the company that is safer for investment by creating a comparison table. [6 pts]

Upload a formal **report (as a single pdf**) plus **the .py file** through BB by the due date.