Assignment 3

Question 1

Design a function using the Pandas package to calculate the average profit of all companies belonging to each state in the United States. The profit of each stock can be calculated as follows:

$$r_t = p_t - p_{t-1}$$

r_t: Profit at time t

p_t: Adjusted price at time t

p_{t-1}: Adjusted price at time t-1

If the data for a stock first appears after 2018, consider the stock's initial appearance as the starting point (t-1). If the data for a stock ends before December 31, 2020, consider the end date of the stock's data as (t-1).

The information about which state each stock (tradingitemid) belongs to can be found by referring to the 'state' column in the info.txt file. Exclude companies whose headquarters are not in the United States.

The adjusted price can be obtained by multiplying the 'priceclose' column in close_price.csv with the 'divadjfactor' column in dividend_adjustment.csv.

Note: To correctly multiply the adjustment factor, you should use the price corresponding to the pricingdate in close_price.csv and the divadjfactor that falls between the fromdate and todate in dividend_adjustment.csv.

Please be aware that each stock may have different t-1 and t time points.

Question 2

Using the Plotly and Pandas packages, propose a function that generates a dynamic graph of the adjusted prices for a given tradingitemid, based on the stock's price and dividend adjustment factor.

The considerations for the adjusted prices are the same as in Q1.
The dynamic graph should include the following elements:

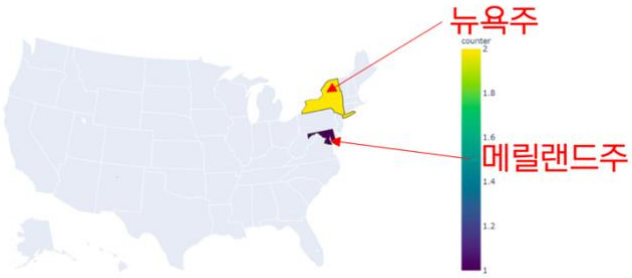
- Date slider (at the bottom)
- Date selection button (top left corner)
- Graph title
- Axis labels
- Axis color: 'Black'
- Grid color: '#F5F5F5'

Question 3

Propose a function that takes a list of tradingitemids as input and generates a Choropleth map representing the U.S. states where the corresponding companies of the tradingitems belong.

For example, if the input tradingitemids are [2631520, 2631609, 49031561], where 2631520 and 2631609 belong to New York and 49031561 belongs to Maryland, the graph should be generated as shown below. New York should be represented with the color corresponding to a value of 2, and Maryland with the color corresponding to a value of 1.

Note: The information about which U.S. state each tradingitemid corresponds to can be found in the 'state' column of the info.txt file. Companies with headquarters outside the U.S. should not be included.



Question 4

The 'games.csv' file contains the game information of teams belonging to the National Basketball Association (NBA). Based on this data, propose a function that takes a team name as input and finds the worst opponent, i.e., the team with the lowest away game winning percentage against the input team. The function should return the average score when playing against that team.

Note: Use data from 2010 onwards.
Column Information:

- Team Name: TEAM_NAME
- Game Date: GAME_DATE
- Win/Loss: WL
- Opponent: MATCHUP
- "@" or "vs." indicates whether it was an away game or a home game.
- Team's score: PTS

Question 5

The 'player_info.csv' file contains information about players belonging to the National Basketball Association (NBA). Based on this data, propose a function that takes a school as input and calculates the average career length of currently active players and players who are no longer active but entered the NBA from the input school.

Column Information:
- School: school
- Career Length: season_exp
- Activity Status: rosterstatus