Stocks

List of stocks:

* S&P 500
* Amazon
* Coca Cola
* Starbucks
* Toyota
* Bank of America

Holidays:

* Valentine’s Day
* 4th of July
* Thanksgiving/Black Friday
* Christmas
* +/- 2 business days (include the holiday if applicable, to make it 5 days)

Date range:

* Test with small data set (1 week) and extend from there
* Test with a longer time frame after (5 years?)
  + 2015-2020

What to look for:

* Average Daily Percentage change of the data
* Trends (code?)
  + Compare p-values ANOVA test (and other tests)
    - (Sukhjit)
  + Graph
    - Date vs. Closing Price for whole time period (Paul)
      * Line?
      * X is one column (date) vs. Price (y-value)
    - Q1 (x-axis) vs. Average Daily Percent change (Paul)
      * Bar graph
        + Adding error bars (st dev) – possible extra addition
        + Coloring conditionally?
      * Box and whiskers plot
        + 25th, median, 75th, outliers, average are shown
      * Other plots? Dot plot?
    - Holidays (x-axis) vs. Average Daily Percent change (Sukhjit)
      * See how to select only the 10 day range
      * Bar graph
      * Line graph zoomed into 10 day range around holiday
    - (Average percentage change) ~ (whether it’s a holiday or not) + (which quarter we are in)
      * Challenge: Will this predict something?
      * Compare the regressions on both variables
    - Function:
      * User inputs file from website
      * And, then it analyzes that file
      * Spits out table and average daily percent change
      * Tiem vs. price for x period

Outline:

1. Read in from a file, data
2. Keep the useful information (date, closing) + add column (daily percent change for each row)
3. Quarterly division and graphing
4. Holiday division and graphing
5. ANOVA analysis – applied to the data after grouping
6. Goal: Code is ideally done by Sunday, move on to thinking of conclusions and presenting what we did

Testing:

* ANOVA – check if the average percentage growth is different between the 4 quarters, and possibly the X holidays

Other trends to check:

* Analyze what happens on holidays
  + Thanksgiving, Black Friday, Christmas
    - Spendy holidays: might expect Amazon, eCommerce type to do well
  + Near holidays (+/- the holiday if it’s closed)
* Seasonal differences
  + Are stocks more likely to go up in one season vs. the other?

Data Processing:

* Combining the data sets or find a way to call the columns of each stock

Analysis:

* (1) Is there a significant difference in stock performance during holidays vs. the rest of the year?
* (2) Is there a significant difference in stock performance between different quarters?
* (3) Is there a significant correlation between the stock price changes within the groups, and the overall market (S&P 500)?
  + Are the general same or different compared to the S&P 500?
    - Visual – comparing the bar graphs
    - Statistical tests – comparing the averages
  + If this true, maybe conclude that the overall market indicates the trends
  + If this is not true, maybe conclude that individual industries behave differently
* (Extra?) Are the changes in stock price correlated across industries (tech vs. cars)?

Time:

4 weeks

Week 1- get the proposal in, figure out what we want to do

Week 2- start writing the code (finish the code if possible)

Week 3- Analysis of the results

Week 4- Reports, recording, final touches.

Potential ideas:

* Stocks
* Parking tickets
  + Prediction of parking tickets based on vehicle characteristics, categories, etc.
* Basketball
* Snake game (challenging)
* Sudoku solver
* tetris ???? (very challenging)
* Simulating enderman placing blocks within a certain boundary (haha)
* <https://www.kaggle.com/datasets/satayjit/student-performance-bd>
  + Predictors of student performance
  + Parents’ education, city/area, gender, parent’s job
  + Linear regression? Correlation?

Ask about:

* What data can we get from tic tac toe to analyze/create analysis (visualization)
* What kind of hypothesis are you looking for in TTT?

Gro