Dataset Selection

Make a git account:
1. Create a repo called 'Machine Learning Final Project -- Last, First'
   a. Where last, first is your name

Select a dataset:
1. https://data.nasdaq.com/search
2. https://www.kaggle.com/datasets
3. https://data.gov/
4. https://medium.com/analytics-vidhya/top-100-open-source-datasets-for-data-science-cd5a8d67cc3d

Requirements:
1. Pick a target column in the data that would make sense to predict from a marketing perspective. Could be pricing, volume, distributions, a category, and so on.
2. The dataset must be at least 2000 rows. More is better
3. The dataset must have at least 10 columns, with at least two string columns. More is better

Perform initial analysis:
- Look at each of the columns, if available look at a data dictionary. Decide what columns look important and begin looking at statistical distributions, missing data, relevance, pair plots and so on.
- Analyze the relationship between your target column and all other columns, if necessary switch target columns or datasets
- Find three relationships in the dataset, at least one aggregate relationship and plot them. Describe the plots.
- With 3-5 sentences sum up your findings and articulate your plan for prediction.

Document / Submission:
- Put a link to your dataset, all your plots, and your analysis code in the github.
- In the readme put your all of your findings with your plots included as inline attachments
- Turn in your link via canvas, ensure read access is allowed / open to public