**Big Data Project Kickoff Meeting:**

**Attendees**: Sowmya, Jason, Alec, Tom

**Potential Data Sources:**

* Crime Data from 2020 to Present in Los Angeles
  + <https://catalog.data.gov/dataset/crime-data-from-2020-to-present>
  + Analyzing crime data, especially from a big city such as Los Angeles, can allow us to look into the financial impact of crime since 2020 and see what policies can help with reducing that impact. This dataset contains more than 400,000 rows of 10+ variables giving us plenty of observations to work with.
* Income inequality and Wealth Distribution
  + <https://www.kaggle.com/datasets/sujaykapadnis/world-inequality-dataset>
* Amazon Customer Review Dataset
  + <https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/#complete-data>
  + This dataset can give us information on where Amazon gets most of its positive reviews and analysis of this could help Amazon financially grow for the future. We can also see the impact of price changes in customer’s purchasing patterns. This dataset has over 233.1 million reviews recorded. We may be able to see patterns of more consumption of essential products over luxury products in times of economic downturn. Understanding these seasonal trends can help businesses with inventory management, marketing and more.
* New York Car Statistics Dataset
  + <https://www.kaggle.com/datasets/ahmettalhabektas/new-york-cars-big-data-2023>
  + This dataset provides two separate sets one being a csv of car ratings and the other being a large dataset of cars in NYC. There is potential in this dataset to predict prices and brands of cars and look at cars with common accident records as well. We could also join the two data sets to compare ratings of the cars. This dataset is 175490 rows × 24 columns.

**Assign Group Roles:**

* Data Analyst/Cleaning: Sowmya and Alec
  + Clean up big dataset and convert into dataframe, write code to run analysis on dataset to answer research question
  + These members are chosen for this role as they both have Python experience and have done past projects with analyzing big datasets
  + This role is crucial for the project as the dataset has to undergo both cleaning and analysis in order for us to draw useful conclusions related to our hypothesis
* Writer/ Editor/ Presenter: Tom
  + Concisely summarize not only the process of analyzing and cleaning data, but also the purpose and the findings of this project
  + This role is crucial to effectively deliver and showcase the observations from this study to any interested individual. For this role, it is vital to understand everyone’s role throughout the project as much as it is to comprehend the nature and the procedures of this study.
  + This member was chosen due to past experience in compiling reports and comfortably presenting in front of a big audience.
* Project Manager / Research: Jason
  + Find possible data set that our group will probably use to generate our final project.
  + This role is crucial to keep everything in progress, make sure we got every on time and find all problems we encountered and try to overcome them with group members together.
  + This member was chosen because of experience about research and managing team project.

**Project Calendar:**

* Meeting schedule: We will meet bi-weekly to discuss where we are at in our project. We will also try to meet with the instructor once or twice throughout the project in order to check in on specifications and other questions.
  + If we have any delays in our work, we will still meet to discuss the delays and if possible work through any hurdles together in our meetings.
* We will start cleaning the dataset right away, once our final topic is chosen and also do some exploratory data analysis.
* In the following week, we will continue to analyze the data and start compiling results and draft the report.
* After data is fully analyzed, we will start working on the final draft of the paper as well as any presentational aspects. All team members will be involved in revising both the paper and any presentational parts.

**Progress Report Template:**

* If an issue or an internal disagreement emerges to surface, we will discuss the matter as a group. If mediation is still needed, we will bring this matter to the attention of Professor Friedman.
* To keep everyone on the same page, we will post updates in our group chats to summarize our week’s overall progress. This includes a list of completed tasks matched with the responsible individual and the relevant findings. Furthermore, the group chat will vocalize the plan for the coming week and the responsibilities assigned to each individual along with its deadlines.