In this alternative way, we can check for **all the** indexes that is equal to 24 and replace these indexes of the commute_times with 18.

Project Problem

Share your broad areas of interest, and some specific research questions of interest in these areas with me. Investigate the datasets available in your area(s) of interest. I have shared data resources on Moodle Share the potential datasets with me.

During my research of different datasets, I found multiple dataset areas that interested me. In this part, I will share these areas of interest, a short description of why this area interested me, some sample datasets I found in these areas and my potential research questions.

https://openpolicing.stanford.edu/data/
data approved!

Crime Data Analysis

This is the first area I would like to work on. I think that there can be many different research questions on this topic. The datasets I found so far are mostly Year to Date so I have to look for more datasets for more long-range research questions.

Datasets Available

Los Angeles Crime Data (2020-Present)

NY Complaint Data (Year to date)

NYPD Arrest Data (Year to date)

Possible Research Questions

- Which neighborhood in each city has the highest crime rate?
- Are certain types of crimes clustered in the same areas of the city?
- Does the crime rate / complaints have a timely pattern? (Specific days of the year/day of the week, etc.)
- COVID-19 restrictions and the crime rates.

Car Crash Data

Similar to the crime data, this topic has the potential to be very interesting depending on the offerings of the dataset we use. There are many research questions that can be used.

Datasets Available

NY Car Crash Data

IL, Chicago Car Crash Data

Possible Research Questions

- How do crash rates fluctuate in relation to time?
- Does major events affect the car-crash rates?
- Where are the top crash locations?
- Investigating if the crashes are more because of driver-related reasons or external reasons

---

Traffic Stops Data

This is another area I would like to work on during the semester. In my opinion this is the most fun topic as there is immense amount of data offerings.

If I choose this project, at the later phases, we can even associate data from this project with the car crash data to have interesting visual elements.

Datasets Available

Traffic stop data in Connecticut

Stanford Open Policing Project (Traffic stops data from all around the USA)

Washington DC traffic stops data

Possible Research Questions

- How do traffic stop rates vary by driver demographic characteristics?
- How do traffic stop frequencies and outcomes vary by the time factor?
- Which types of violations are most likely to result in more severe outcomes (arrest, vehicle searches, warrants, etc.)?
- Comparing the traffic stop statistics between multiple states.

Amazon Product Reviews

In this last category, I have extremely large datasets of Amazon Reviews including many important information about the reviewed products. I think that it can give important insights about the product categories, public opinion about a general category, etc.

Datasets Available

2013 Amazon Product Reviews (34.69M Reviews in Total)

2014 Amazon Product Reviews (82.83M Reviews in Total)

2018 Amazon Product Reviews (233.1M Reviews in Total)

2023 Amazon Product Reviews (571.54M Reviews in Total)

All these four datasets start from May 1996

Amazon Customer Review (Specific Products)

Possible Research Questions

- How have average ratings changed over time?
- Relationship between the big events (2008 crisis, COVID-19) and the review frequency, score, etc.
- Which product categories has the most ratings / highest ratings / highest variance in ratings?

As I mentioned, among all these areas, the traffic stops data interested me the most. I think that the Stanford Police Dataset is a very strong data source I can use during my semester-long project. The dataset includes many interesting information about specific attributes of a traffic (or pedestrian) stop including the date&time of the stop, the demographics of the person stopped, the reason and the outcome of the stop. Depending on how extended I would like to make the project, I can compare two counties in the same state or in different states. I think that this dataset has a huge potential of telling many stories depending on my research questions.

I wanted to include these other areas interested me as a backup plan if my project does not go as planned due to restrictions in the dataset, etc.