

Project Proposal

Padma Danturty

The research question I will explore is, what groups of people are more likely to be arrested and not charged in San Jose? I will also investigate, for each of these groups, what are the most common incidents that lead to an arrest without a charge?

I will be using a the San Jose Police Arrests dataset from Kaggle which has information about from 2015 to 2022 about arrests, incidents, and charges. My plan is to join the datasets and investigate the differences between the rows (arrests) that do and do not have a corresponding charge. I want to look at the difference in statistics between race, ethnicity, gender, and age. I can perform some permutation testing to see if the differences are statistically significant.

I also plan to look at the incidents (most arrests will have a corresponding incident unless it was an arrest from a warrant). I want to see which incidents are most common for the arrests without charges, and also which indicents are most common among each group.

If I have time, I will go a comprehensive search using regex tools to group together the incidents (combining BURGLARY VEHICLE<\$950 and BURGLARY VEHICLE<\$400 for example). I would also like to investigate the interactions between the groups (for example race and gender) for which groups are more likely to be arrested without a charge.