Our group consists of Justin, Brevan and myself. Our study looks at the relationship between specific crime rates and crime inducing attributes in California Counties.

(Next Slixe)

It’s been well documented that factors like racial diversity, economic condition, education and population density correlate with crime rate. However, even though the correlation between these attributes and crime have been established, some cities don’t see any improvement in crime reduction. Take Sacremento for example, the city spent 150 million on local police in 2019 but still stands have one of the highest crime rates in America.

(Next Slide)

We suspect that resources are being inefficiently allocated due to the fact that the police doesn’t know which crimes they should be watching out for.

(Next Slide)

Thus, our goal for this study is to find a classification model that can accurately predict specific crime rates in California counties. In the process, we want to discover correlations between attributes of California counties to determine which attributes contribute to higher crime rates and which crimes are being influenced.

(Next Slide - Brevan)

We first ran K means Clustering on our data to get a better understanding of the possible correlations between subsets of the data as well as identify any potential outliers.

For clustering, we did a range transformation on our data, shifting them to the range 0 to 1, so that attributes are equally weighted but still maintain the same distribution. The number of clusters are chosen by minimizing intra-cluster distance.

(Next Slide)

This is a graph of the average intra-cluster distance for different number of clusters. Looking at the graph, it is evident that the decrease in average distance within centroid starts to level off after cluster 7 so we chose 7 as our most optimal number of clusters.

(Next Slide)

Here are the results of our cluster analysis. Based on the characteristics of each cluster, Clusters 2 and 6 looks to be potential outliers while cluster 4 shows a strong correlation between Total population, African population and robbery.

(Next Slide)

Heres a heat map visualizer for the clusters. From the map, it is quite obvious that cluster 6 is an outlier as one of its attributes is 900% the average. Clusters 2 and 4 shows strong correlation between robbery, Asian population and African population.

(Next Slide – Brevan)