



**CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY**

| | |
|--|------------------------------------|
| Name(s) Anish Neervannan | Project Number 35863 |
| Project Title Comparing Crime Rates in Various Communities Using K-Means | |
| Abstract Objectives/Goals The purpose of this experiment was to compare the extent to which various factors affected the crime rates in communities throughout the country, using the K-Means clustering algorithm. Methods/Materials The experiment used a combination of the K-Means clustering algorithm and a regression analysis to account for all outliers. A sample code for K-Means, found online, was edited and compiled to suit the needs of the analysis. A data set was used for this experiment, where only the variables required were used and the communities with unknown data points were eliminated. The data was formatted to an excel file, the program was run on that data and the cluster centroids results were graphed. The best type (polynomial/logarithmic) line of fit was chosen depending on the data set trend. Results It was found that low number of police units per 100,000 people was the biggest contributor to crime rates, followed by high poverty rates, low high school graduation rates, high homelessness rates, and high unemployment rates. The relationship between the number of police units per 100,000 people and the crime rate was that the crime rate went down as the number of police units per 100,000 people went up. However, the relationship between the unemployment rate and the crime rate showed that the crime rate only went down marginally as the unemployment rate went down. Conclusions/Discussion The hypothesis was partially supported; the order in which the factors affected crime rates was only partially right. For example, unemployment rates were not the second largest contributor to high crime rates; they were the smallest contributors. A possible explanation for this was that the use of the K-means clustering algorithm produced results different from those of previous experiments. | |
| Summary Statement This project compared the extent to which various factors affected the crime rates in communities throughout the country, using the K-Means clustering algorithm. | |
| Help Received None | |