*Violent Crime Rates in California*

**1. Abstract**

*For this project, I analyzed the violent crime rates in California by creating research analysis questions, including trends over time, geographic crime rates, seasonal patterns, and predictive modeling. By using 3 different models, such as diagnostic, linear regression, and logistic regression, I was able to find trends for specific time periods, which counties had the lowest and greatest crime rates, seasonal patterns, if the total crime rate decreased or increased from 2000 to 2013, and high crime rates based on a county's population. My findings showed stable trends, no seasonal patterns, and a low prediction for high crime rates based on population size.*

**2. Introduction**

*Throughout the years, California's crime rates have risen and gotten worse. High crime rates can affect the safety of communities. By analyzing California crime rate patterns and predictors, I can provide law enforcement with valuable resources. My project primarily focuses on the crime rates in California and focuses on the analytical question I came up with. The questions are, “Are there significant increases or decreases during specific time periods?”, “Which counties or cities in California have the highest and lowest violent crime rates?”, “Did the Total Violent Crime rate increase or decrease in California from 2000 to 2013?”, “Are there seasonal patterns in violent crime rates?”, and “Can we predict whether a county has a high crime rate based on its population size?”. This project aims to find important insights regarding crime trends and patterns that will help with decision-making.*

**3. Dataset**

*The dataset that I chose and used is violent crime data based in California. This dataset shows the violent crime trends, geographic rates, and the relationship between crime rates and population. This data has multiple variables. I got this data from Data.gov.*

**4. Exploratory Data Analysis (EDA)**

*file:///C:/Users/Chike/AppData/Local/Temp/RtmpuE5qXg/preview-6de0118c1be9.htm*

**Statistical summary of the data.**

*Mean\_Rate is 5.1354, Median\_Rate is 5.2535, SD\_Rate is 0.758, Min\_Rate is 3.959, Max\_Rate is 6.2175, Q1 is 4.4834, and Q3 is 5.7024.*

**Visualization plots**

***Question 1:*** *I used a line plot. The line plot shows a decrease over time*

***Question 2:*** *I used a bar plot. The bar plot shows the difference in crime rates in each county. Los Angeles has the highest crime rate and Marin has the lowest crime rate.*

***Question 3:*** *I used a line plot. The line plot shows a downward trend of the total crimes from 2000 to 2013.*

***Question 4:*** *I used a box plot. The box plot shows the crime rates for each year*

***Question 5:*** *I used a scatter plot. The scatter plot shows the prediction based on crime rate and population.*

**2 Statistical Tests**

***Question 1:*** *I used a statistical test. The statistical test showed that the crime rate decreases by 0.179 per year.*

***Question 4:*** *I used a statistical test. The statistical test showed that the Diagnostic model test was good for making predictions about crime rates.*

***Question 5:*** *I used a statistical test. The statistical test shows that you can't predict high crime rates based on population.*

**5. Models**

***Question 1:*** *For question 1 I used a linear regression model to display the relationship between time and crime rates. The regression line has a negative slope of -0.17896 which means that the crime rate has decreased. The model performs well and has a significant coefficient.*

***Question 4:*** *For question 4 I used a linear regression/diagnostic model to display the violent crime rates for different years. The model showed a negative trend meaning that violent crime rates have decreased. The diagnostic confirms the assumptions are met and performs well.*

***Question 5:*** *For question 5 I used a logistic regression model to perform a prediction on if a county has a high crime rate based on its population. Since the predicted\_prob is close to 0.5 which is near the threshold for classification this makes the model unsure. The model shows it has a weak predictive power making it known that population is not a good predictor for high crimes.*

**6. Summary of learning**

What I learned was the decreasing trends of crime rates over the years. I also learned which counties have the highest and lowest crime rates. My hypothesis was that the bigger the population the higher the crime rates are. With my findings and models, it shows that is not the case. One big challenge was having to review past assignments and practices and see how we used certain plots and models to implement them in my findings. One approach I would've taken is to go into the Excel file and get rid of null and duplicate columns and rows.