CRIME STATISTICS ON THE TOP 100 MOST POPULOUS CITIES IN THE UNITED STATES FOR BUSINESS DECISION MAKING

INTRODUCTION

Starting a new business is a daunting task. The business leadership has to be focused on many criteria when determining a new location. Financial incentives, workforce issues, access to transportation, and infrastructure are important considerations. Locating to the most populous areas often allows for many of these considerations to be handled. However, with the increase in population, there may be the factor of crime, especially violent crime, introduced into the location decision. This data review will evaluate crime data and provide perspective on the 100 most populous cities in the United States. Along with the other factors the potential business reviews in their decision-making, this crime data will be helpful in their review.

DATA ACQUISITION AND CLEANING

The Federal Bureau of Investigations (FBI) compiles crime data in the United States. The data to be reviewed in this will be the Federal Bureau of Investigations Uniform Crime Reports. The statistics are from the year 2017 for the 100 most populous cities in America. The statistics were calculated as a rate per 100,000 people. They were tabulated and published in a Wikipedia table.

Data was scraped from this table and a pandas data frame was created. The crimes, as noted, are reported at a rate of 100,000 people per year. The table provided information for Violent and Property crimes with four subgroups under each heading.

For purposes of presentation, total violent crime as well its subgroups along with the total for property crimes were the most important for evaluation. Thus, the subgroups under property crimes were deleted. The column names for Violent Crime 1, Violent Crime 2, Violent Crime 3, Violent Crime 4, and Violent Crime were renamed Murder, Rape, Robbery, and Violent Crime Total, respectively.

Many of the City names from the Wikipedia table had footnote numbers or precinct designations attached. Each of these was replaced with the proper City name. Washington, DC was a special issue since it is not a City, State combination and had to be addressed separately.

All of the datatypes in the Wikipedia table were objects. Population was changed to datatype integer. Violent Crime Total, Murder, Rape, Robbery, Assault, and Property Crime were changed to datatype floats.

The States were changed to their respective two letter postal code abbreviations. These were then concatenated with the city and placed under a new column Location. The City and State Columns were deleted.

There were eight missing values listed as NaN. Averages were calculated for the values under specific crime columns and NaN replaced. The missing values in the Total Violent Crime column were calculated and NaN replaced. A spurious last line without any valuable information was deleted leaving 100 cities with all of there values in a Pandas dataframe.

Several manipulations of this final dataframe were accomplished for cross checking the changes made to this point.

To prepare for the actual data analysis, two final dataframes, df8 and df9, were created and sorted with the index reset at 1. Df8 was based on the Violent Crime Total and Df9 was based on Murder.

DATA ANALYSIS

Maximum, minimums, and mean were calculated for the columns Violent Crime Total and Murder. For the total Violent Crimes, the maximum was 2082.29/100,000, the minimum was 61.21/100,000 and the mean was 731.27/100,000. For Murder the maximum was 66.07/100,000, the minimum was 0.72/100,000, and the mean was 11.68/100,000.

Comprehensive data is presented in df8 (Graph 1) and df9 (Graph 2) in descending order indexed from 1 for Total Violent Crime and Murder, respectively. The data for these are presented in two horizontal bar graphs. These two bar graphs are dramatic visuals of the range of both Total Violent Crime and Murder over the 100 most populous cities in the United States.

The top twenty cities in the United States for Murder and Total Violent Crime are presented in Graphs 3 and 4. St. Louis, Baltimore, and Detroit take the top three spots for murder in the United States. For Total Violent Crime, it is unfortunately still the same three just switching Detroit for Baltimore in spots 2 and 3.

Graph 5 shows the Total Property Crimes in the top twenty cities for Violent Crime while Graph 6 shows the Total Property Crimes in the top twenty cities for Murder. These graphs confirm that there is a high level of crime in all of these communities.

The twenty United State cities with the lowest Total Violent Crime are shown in Graph 7. Graph 8 shows the twenty United State cities with the lowest murder rate.

The safest city for both murder and total violent crime is Irving, CA. Three of the four safest cities for murder and total violent crime are located in California. There is a seven-fold difference in the total violent crime rate from the Number 1 city to Number 20. Interestingly, there is only a 3.5 fold difference from Number 1 to Number 20 in the murder rate.

The final graph, Number 9, represents the Property Crime in the twenty safest cities regarding murder. All of the cities are tightly grouped from approximately 1500 to 3000 property crimes but Spokane, WA is dramatically different at just over 7000 property crimes.

CONCLUSION

This study presents the information for Violent Crime and its subgroups and Total Property Crime for the one hundred most populous cities in the United States. These rates were obtained from an FBI database published in Wikipedia. There was only a small amount of missing data that was corrected via averages. This can enter some error in the calculations. Also with data that has been translated from an original source to a secondary source, some errors in translation can occur. In the end, business developers can utilize this data along with all of their other criteria to make the decision of where the business can be located.

https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/72037ecd-81d9-4914-a9d1-3dfcfc113ca5/view?access\_token=1280eff636719b9216e4dcd0e8c0920961fb653f483cbfe3feab1370fc71b1fb