Los Angeles City Crime Analysis

Analysis and Visaualization Of Los Angeles city crime data

Naga Jyothi Kunaparaju
Saint Peters University, Data science
nkunaparaju@saintpeters.com — 1 (347) 341 3307

Abstract

Crime analysis helps to identify patterns and trends in crime. Analyzing these patterns can help law enforcement agencies to develop crime prevention strategies. By utilizing modern technics and tools, thorough investigation is possible. Developing methods to do crime analysis, helps in identify patterns quickly and efficiently. Aim of this project is to develop dynamic and interactive tools based on Los-Angeles city crime data by using Data mining and Data Visualization technics.

Introduction

Data analysis in crime prevention had been used since many years. COMSTAT is one such kind tool set used by many police department to show spikes in crimes using statistics available for comparison. Today with advancement in technology, many software tools and technic are available to analyze heavy volumes of data. Complicated patterns could be revealed through this software with ease.

*D*ata mining plays an important role in terms of prediction and analysis, for example clustering will provide more insight into data to build models. Clustering makes data divide into groups. The data within the group is very similar and very dissimilar when compared to other groups

*C*orrelation is another way to find trends. Some economic variables are highly correlated with crime statistics. These variables could be used to predict future crime rates based on past trends. Researching various regression models to find influential factors effecting crime also another way of doing analysis. This project is also aims in creating interactive interface using NLP.

*G*oal of this research is to either to find new patterns or advancement or existing ones by using all above methods and presenting them user-friendly format. Below fig compares Los Angeles with US .Statistics collected from http://www.city-data.com/crime/crime-Los-Angeles-California.html

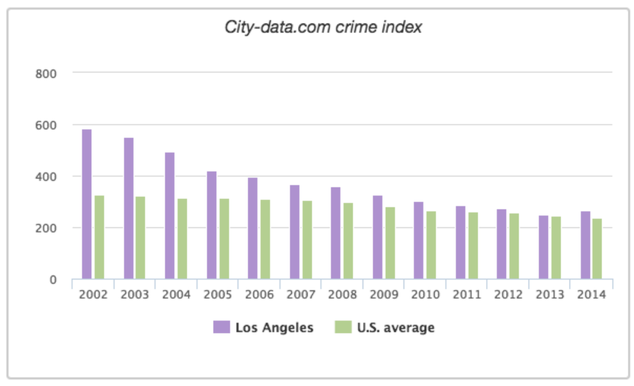

Figure 1: Crimes per 100,000 (Los Angeles vs US)

Data

Data set for this project collected from https://catalog.data.gov/dataset/crimes-2012-2015. According to https://www.baruch.cuny.edu/nycdata/public-safety/crime-select cities.htm, Los-angles is one of the higher crime rate city in USA. Crime rate and types of crime vary from one area to other. Various social and economic factors effect considerably crime rate in different geography. Data analysis of one city will be different than other. For this project we choose Los-Angles crime data set for analysis.

*M*ain tasks related data are loading, cleaning (Including missing value interpretation) transforming and formatting to apply statistical methods. As many would say data cleaning and preparing is 80 percent of data analysis. Data set consists of 760,000 rows and 15 Attributes, including: Timeocc, Area, CrmCd, CrmCdDesc, Status, Location, Cross Street, Lat, Long, DateRptd, Dateocc. This data is divided into multiple data sets, for example:

- Time variant such as Daily, Monthly, Hourly
- Geographical information split into area and location.
- List of Violent crimes with crime codes 1 to 8 are separated into a different data set for analysis.

Method

After data preparation main task is to choose method. As technology growing there are many advanced technics are available. Choosing right method which fit the data is very important task in analysis. Finding correlation between various attributes useful to find patterns. Visualizing Time-series might also reveal certain patterns in data. Geo coding tools are very helpful visualizing data in a way which will yield good results to find trends for future prevention of crime.

- Pandas and Pyplot Libraries are used to create plots
- In raw data set, Lat long combined into single column. For Geographical representation, it has been split into two columns.
- Data set is also loaded into Excel, Tableau for cleansing and formatting.
- Crime Data set is maintained by LAPD.

Results

Time series analysis might reveal pattern of crime like particular time of the day, day in a week or time period like holiday season. Results of the Logistic regression may be finding influential factor such Geographical or economic factors. We can also find correlation between factors. Maps also helpful in identifying areas and cross-sections which are more vulnerable to crime.

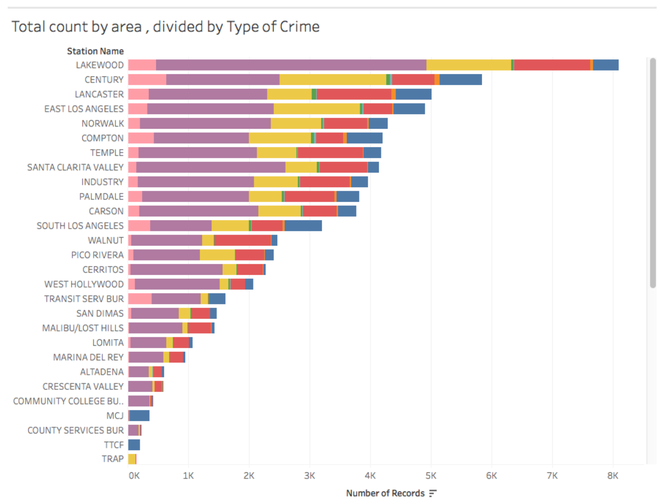

Figure 2: Total crime count By Area

By observing Hourly plot, Peak is around 12 pm. Afternoon, there are more number crimes compared to early hours of the day

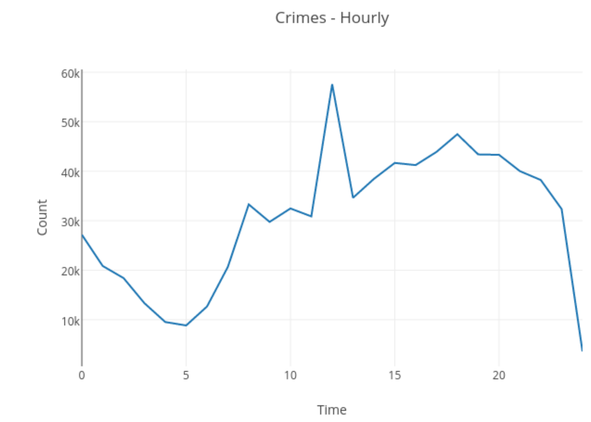

Figure 3: Hourly total

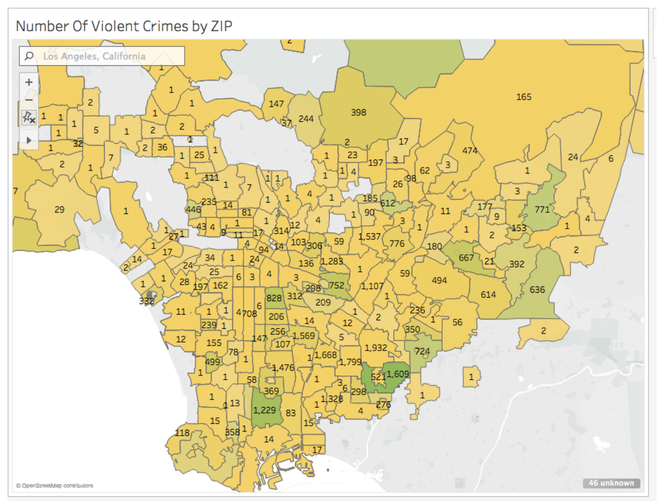

Figure 4: Violent Crimes by Zip

Figure 5: Violent Crimes week day

Conclusions

- – Daily : Every day on average around 100 to150 crimes happen in LA.
  - – Hourly : By observing Hourly plot, Peak is around 12 pm. Afternoon, there are more number crimes compared to early hours of the day
  - – Monthly : Monthly plot shows crime rate higher in August
  - – Heat map by weekday: Heat map revealed most interesting pattern Friday is max crimes happening day even without traffic related crimes.
  - – Area : Lake wood area is most dangerous among all .
  - – Crimes by Zip code : We can visualize how crimes are scattered overall in LA. For example if you know Zip code you can find out number of crimes with in that Zip code so that we will have an option to choose whether this area is safer or not for living.
  - – Crime Type : Among all crime types Traffic related issues are higher

References

[1] *Http://www.city-data.com/crime/crime-Los-Angeles-California.html, Https://data.lacity.org/A-Safe-City/LAPD-Crime-and-Collision-Raw-Data-for-2015/ttiz-7an8/data.*