Data Analysis on Crime Rates in Metropolitan Areas

ISM 6405 - Business Intelligence
Group 14

Submitted by -
Himanshu Vaishnav (9817-1321)
Nishad Shinde (6619-9763)
Piyush Singh (0633-9460)
Rahul Roy (1181-1339)
Sahil Raut (8109-7575)

Executive Summary:

Crime is a part of every society. Thinking that it can only happen to someone else and doing nothing to prepare yourself or not taking the necessary precautionary measures makes you the perfect soft victim for the criminals.

FBI UCR (https://ucr.fbi.gov/) Uniform Crime Reports compiles official data on crime in the United States, published by FBI. UCR Program has been the starting place for law enforcement executives, students of criminal justice, researchers, members of the media, and the public at large seeking information on crime in the nation. The program was conceived in 1929 by the International Association of Chiefs of Police to meet the need for reliable uniform crime statistics for the nation. In 1930, the FBI was tasked with collecting, publishing, and archiving those statistics.

Today, four annual publications, Crime in the United States, National Incident-Based Reporting System, Law Enforcement Officers Killed and Assaulted, and Hate Crime Statistics are produced from data received from over 18,000 city, university/college, county, state, tribal, and federal law enforcement agencies voluntarily participating in the program. The crime data are submitted either through a state UCR Program or directly to the FBI's UCR Program.

Crime rate is a count of crimes complied to assess the effectiveness of a crime control policy, and the impact of the policy on the risk of crime victimization. We will be analyzing different metropolitan cities on murder, rape, robbery, aggravated assault, property crime, burglary, larceny theft, and motor vehicle theft.

Project Motivation:

As students with background in information technology and engineering, the idea of understanding crime and seeing a trend in their happenings excited the whole group. With the help of this project, we can understand and determine the crime rate in three consecutive years in metropolitan cities in USA. We will use our class learnings into our project which will give us hands on experience to new statistical methods, technologies and tools.

Tools and Dataset Description:

Tools: Tableau and MS Excel

Dataset: The Crime data shared with us were organized into two files with the properties provided below.

Data Set	*Number of rows*	*Number of Columns*
Principal Cities	1500	27
Crimes	1534	27

Variables used in Analysis:

Variable	*Description*
Principal Cities	Number of principal cities
Population	Population in each city
Aggravated Assault	Sum of Aggravated Assault for each city.
Burglary	Sum of Burglary for each city.
Larceny-theft	Sum of Larceny-theft for each city.
Motor vehicle theft	Sum of Motor vehicle theft for each city.
Murder	Sum of Murder for each city.
Property Crime	Sum of Property Crime for each city.
Rape	Sum of Rape for each city.

Robbery	Sum of Robbery for each city.
Violent Crime	Sum of Violent Crime for each city.

Results and Interpretation:

Data analytics is the technique and process to quantitatively and qualitatively identify the business challenges and make informed decisions to enhance the productivity. It is a step by step process that involves understanding the data, organizing and manipulating it into groups that delivers meaningful insights.

For our project, after understanding the data set, the data was preprocessed i.e. data integrity problems were detected and resolved using omission and imputation techniques as appropriate. We also lead into problems where the data was noisy, and values were missing. Further, the obvious outliers were identified and eliminated for the current analysis, thus completing the data cleaning process.

For better understanding, some examples are explained below:

Data Integrity/ Noisy Data:

The dataset has a variable “Principal Cities” that indicates the Major Cities having various crimes. This variable is an integral part in deriving the state/area where the crimes are the most. The column contained incomplete or missing data as some crimes had null values. For example, some records had ‘0’ and some had white spaces. This problem was resolved by filtering the data and changing the data to one standardized format.

Further, for the same variable, some of the records had ambiguous data. For example, certain crimes for same cities but from different states. In certain cases, imputation method was adopted by using another variable from our data set, “Principal City”, to find the same records. In cases where even the “CITY” attribute contained ambiguity and depending on the number of records, omission method was adopted to resolve the issue.

Trends in Metropolitan Areas for Crime (Tableau Analysis):

I. Burglary:

DataSet*:* This data is the aggregated data of Burglary in Metropolitan areas. This data is a small part of the whole dataset as the majority of the dataset is crime specific data where different crimes have reported in Metropolitan cities.

Analysis: As per the tableau analysis, Houston, TX has most burglary crime rate. However, this crime rate has reduced 14.52% in last 3 years. Although, in Los Angeles, burglary has increased by 5% in last 3 years. There are a few Metropolitan cities, where burglary has reduced by a great percentage. For Example, in Memphis burglary has gone down by 20% and in New York it has reduced by 24% in last 3 years.

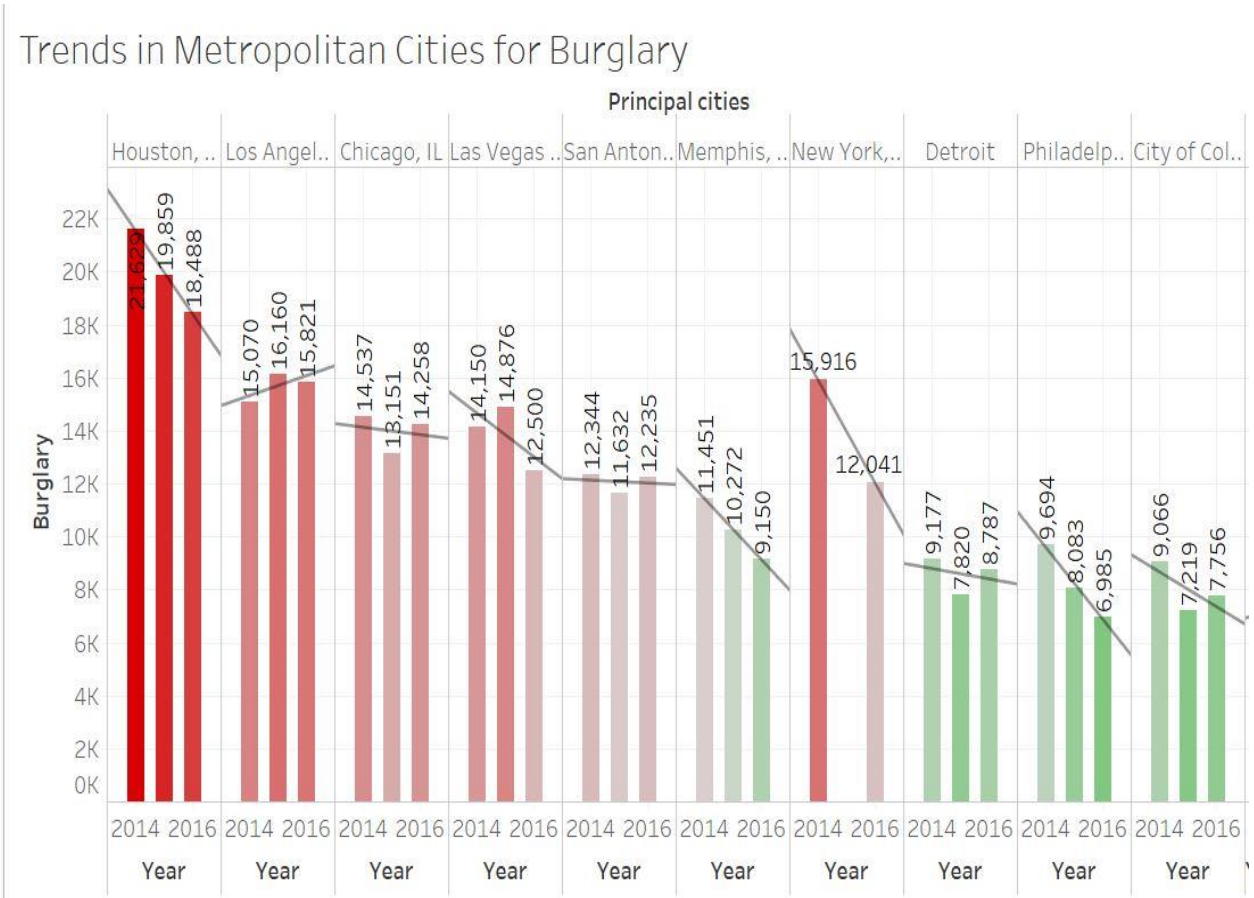


Fig 1. Trends in metropolitan areas for burglary

II. Aggravated Assault:

DataSet: This particular data is the aggregated data of Aggravated Assault in Metropolitan areas. This data is a small part of the whole dataset as the majority of the dataset is crime specific data where different crimes have reported in Metropolitan cities.

Analysis: As per the tableau analysis, New York has most aggravated assault crime rate. However, this crime rate has reduced 2% in last 3 years, which is not a significant reduction in crime rate. Although, in Chicago, aggravated assault rate has increased by 26% and in Los Angeles, it has drastically increased by 61% in last 3 years.

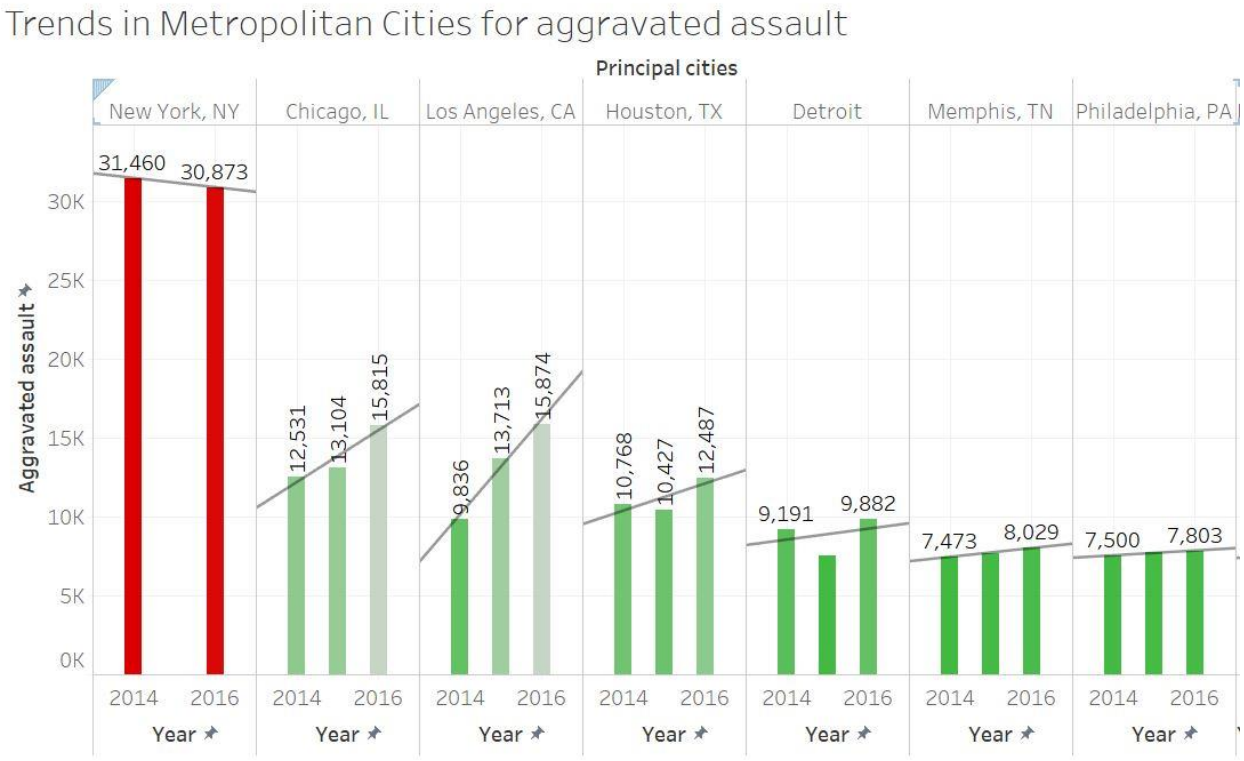


Fig 2. Trends in metropolitan areas for aggravated assault

III. Murder:

Dataset: This dataset is the aggregated data of murder crimes in metropolitan cities of USA. The data displayed above was extracted by cleaning the bigger dataset and ranking it according to maximum number of murders.

Analysis: According to the data, Chicago, Il has the maximum murder rate (1654), which is the aggregate of three years (2014-2016). Detroit and Baltimore follow Chicago and have almost similar murder rates. Among the top 20 cities, City of Louisville has the least number of murders, which is 202.

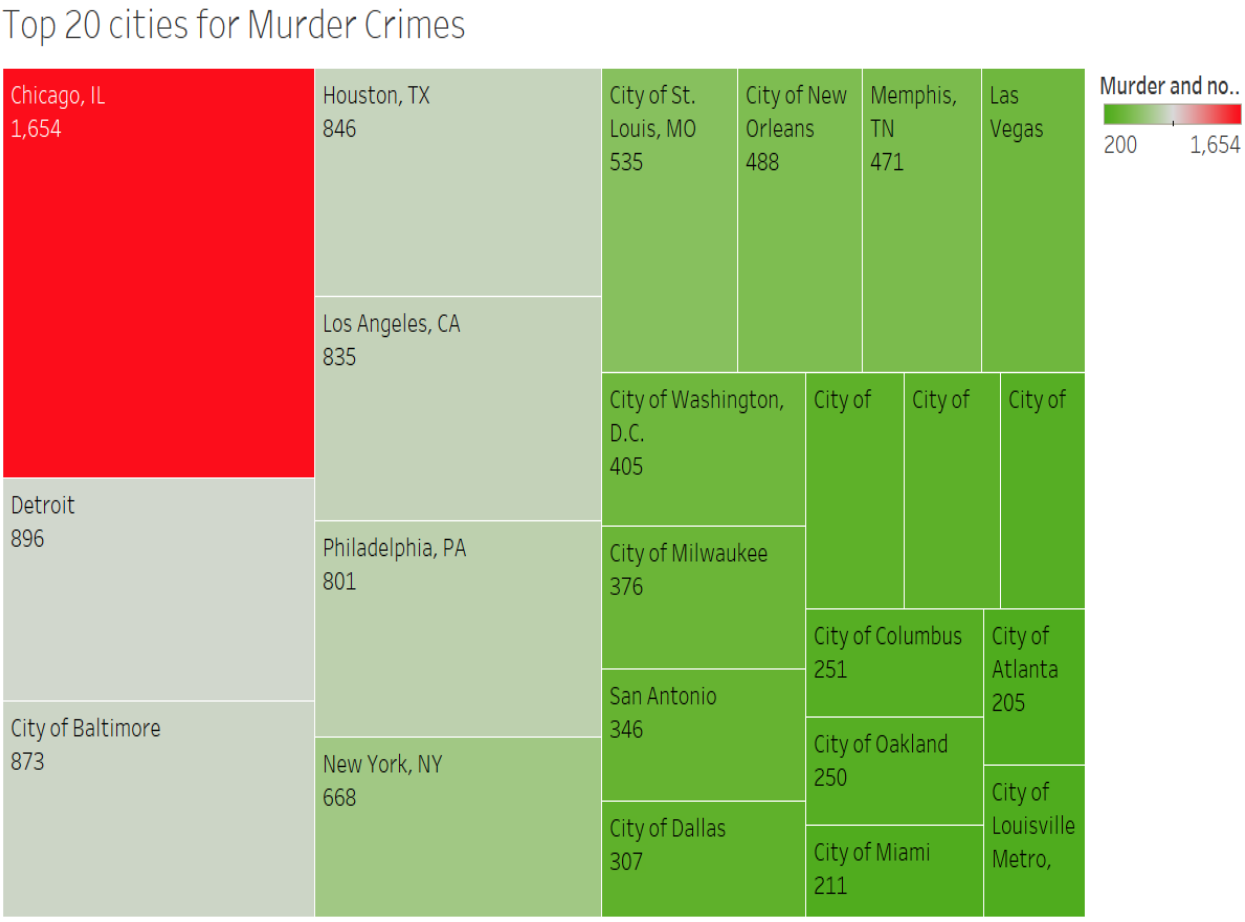


Fig 3. Top 20 cities for murders

IV. Property Crime:

Dataset: This dataset is the aggregated data of property crimes in metropolitan cities of USA. The above data was extracted from all the crime data of the metropolitan cities of the States from year 2014 to 2016 and was ranked in descending order of the property crimes.

Analysis: As per the analysis conducted on tableau, Houston, Texas was found with the maximum total property rate crimes from the year 2014 to 2016 followed by Los Angeles (275,793), New York (261,025) and Chicago (252,516). The above cities are the Top 100 cities which had the most number of property crimes in the last 3 years.

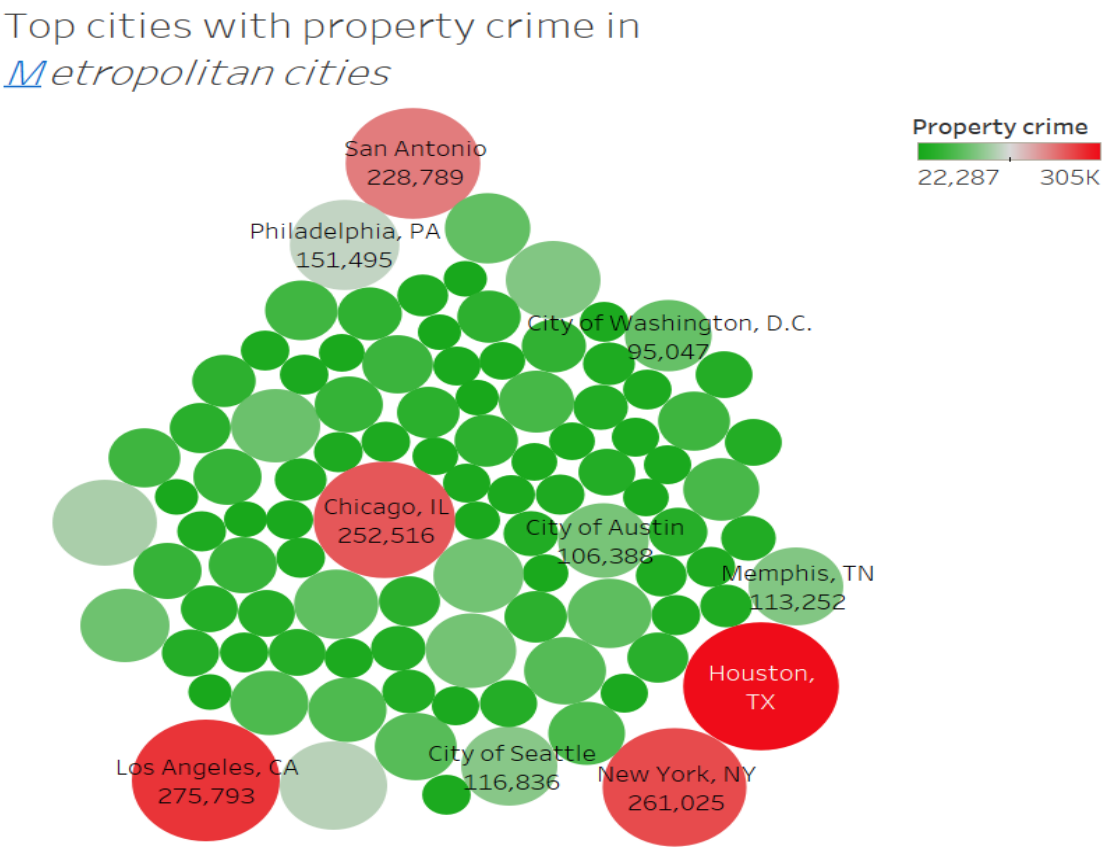


Fig 4. Top cities with property crime

Maximum and Minimum crime rates in Metropolitan areas (Tableau Analysis):

I. Minimum crime rates in Metropolitan areas:

Analysis: As per tableau analysis, we have found top 10 safe cities with least crime rate in Metropolitan cities. The city of Marco Island has the least crime rate compared to other metropolitan cities. The city has total of 192 combine crimes including larceny-theft and burglary with '0' murders and '0' robberies in 2016. The city of North Mankato also has less crime rate with count of '0' rapes, murders, and robberies.

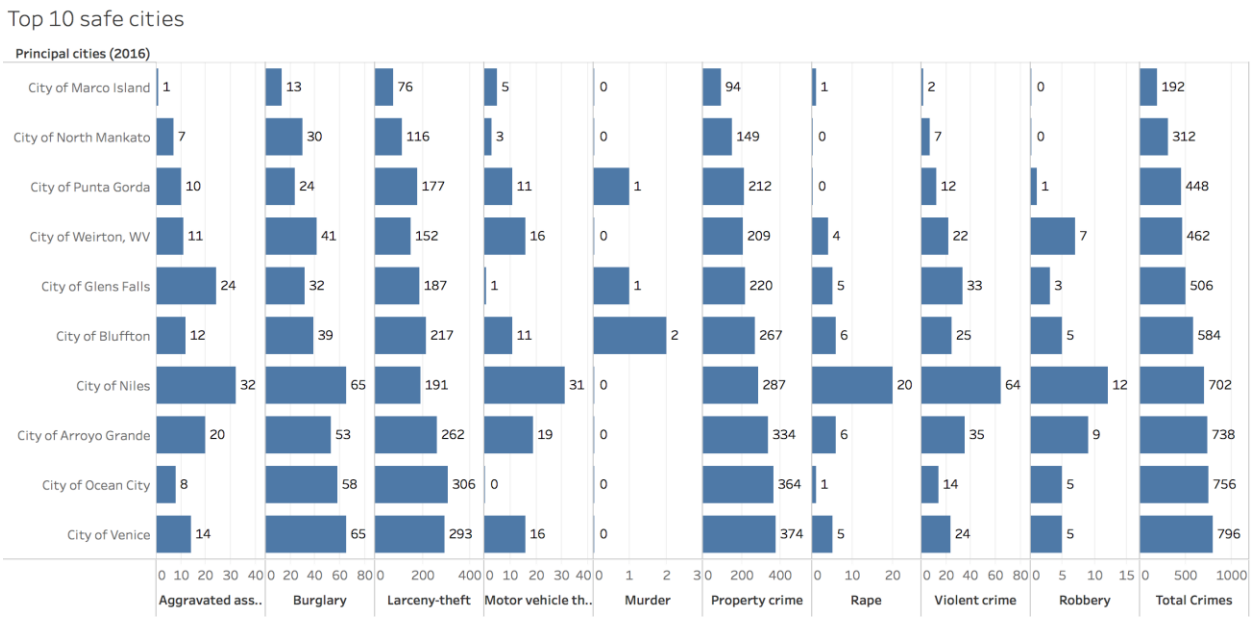


Fig 5. Top 10 safe cities

II. Maximum crime rates in Metropolitan areas:

Analysis: As per tableau analysis, we have found top 10 dangerous cities with highest crime rate in Metropolitan cities. New York has the leading crime rate compared to other metropolitan cities. The city has total of 16.9% combine crime rates with inflated rate of 25% in aggravated assault and 23% in violent crimes in 2016. The city of Jacksonville is second highest crime rate with 17% in burglary and 15% in property crime rate.

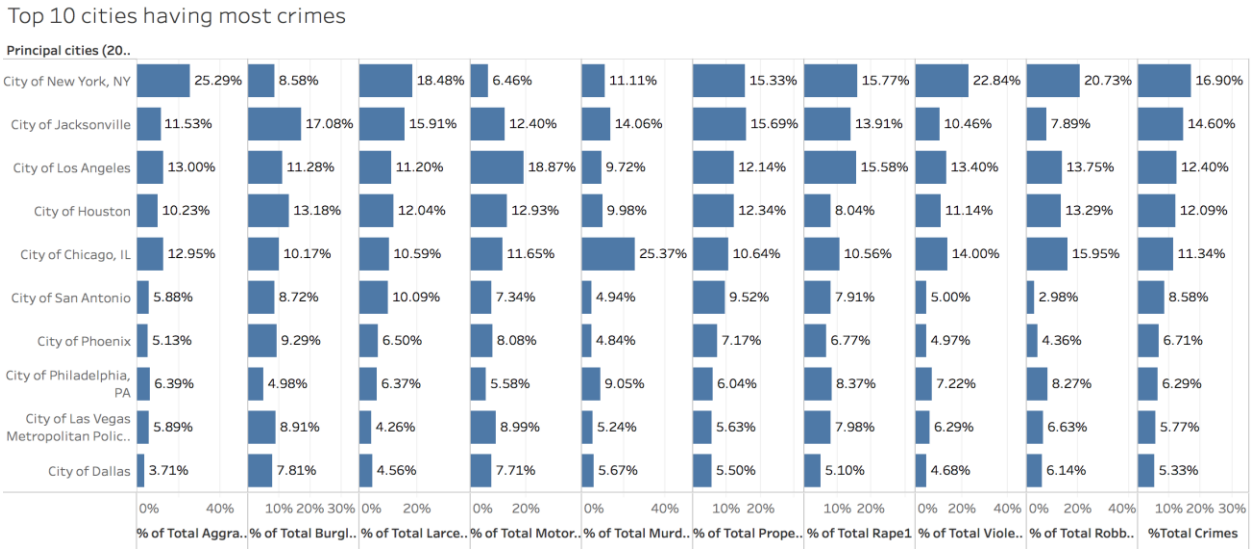


Fig 6. Top 10 cities with highest crime

Comparison between multiple crimes:

Analysis: As per tableau analysis, we noticed that few particular cities have a very significant number of crimes occurring together. In our example we have considered the crime (robbery and rape) together to identify the cities (Chicago, Los Angeles, Houston, and Jacksonville) to have a very high number of reported cases as compared to other cities in the dataset.

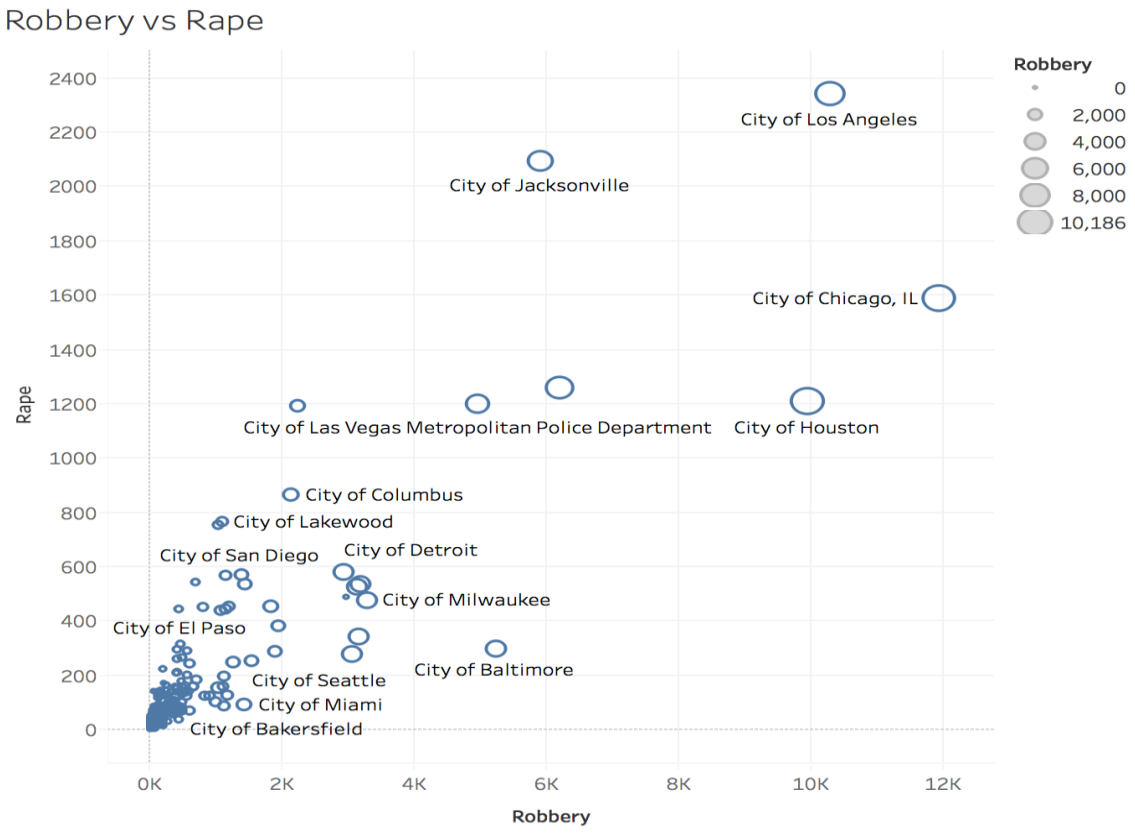


Fig 7. Comparison of robbery and rape in different cities

Correlation between Crimes:

We have used correlation to show whether and how strongly pairs of crimes are related in metropolitan cities. Crimes in cities are related. Sometimes, one crime could lead to another crime. In figure, for example violent crimes (VC) are highly correlated to robbery (RO) and aggravated assault (AA). Similarly, property crime (PC) is strongly correlated to larceny-theft (LT).

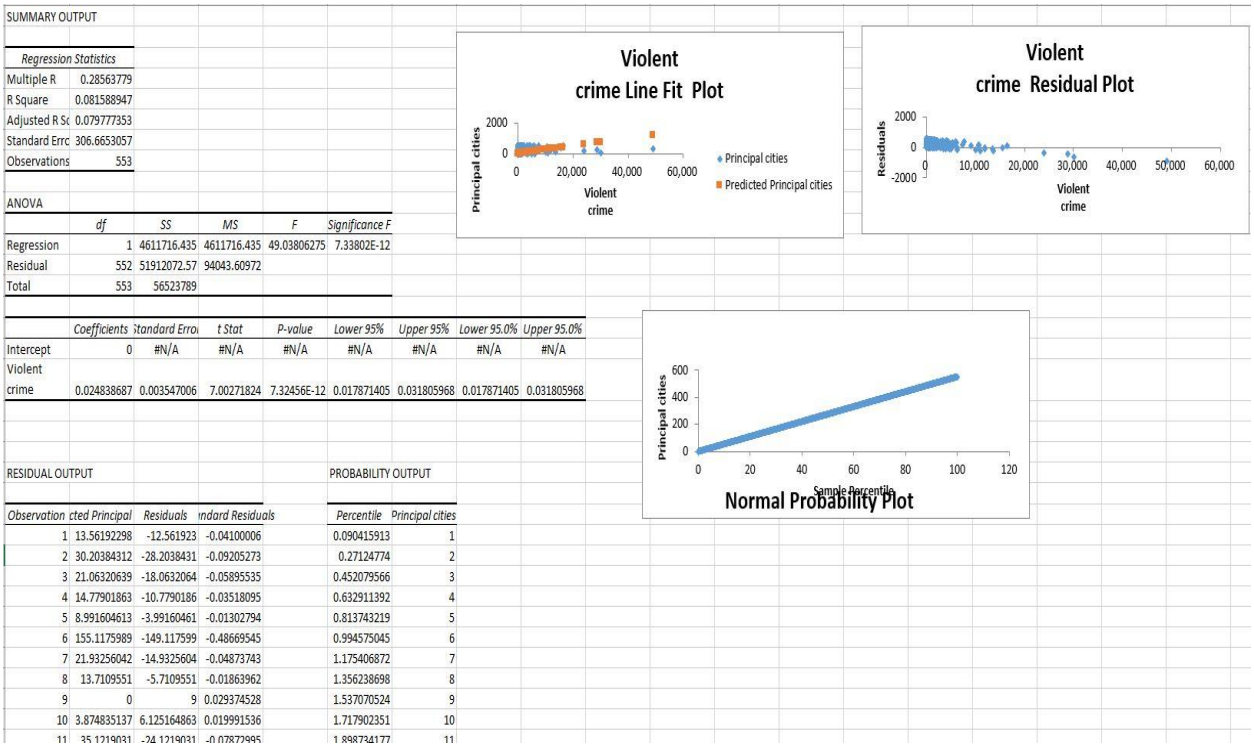
	VC	MNM	RA	RO	AA	PC	B	LT	MVT
VC	1								
MNM	0.871860719	1							
RA	0.927156618	0.804825332	1						
RO	0.983365297	0.892762746	0.906756523	1					
AA	0.992682258	0.838110992	0.907678638	0.955736412	1				
PC	0.927976821	0.818380251	0.949491785	0.91718039	0.908079122	1			
B	0.862283225	0.824954038	0.889867102	0.857421926	0.838116793	0.936886853	1		
LT	0.921041766	0.784295425	0.938098369	0.905509329	0.905437664	0.989502007	0.883786258	1	
MVT	0.822491797	0.797717071	0.85611958	0.832859255	0.789184988	0.890199891	0.921816088	0.823144152	1

VC- Violent Crime, RA-Rape, RO- Robbery, AA- Aggravated Assault, PC- Property Crime, LT- Larceny Theft, MVT- Motor Vehicle Theft, MNM- Murder and nonnegligent manslaughter

Fig 8. Correlation between crimes

Regression Analysis:

To find relationship between our variables we have performed regression analysis using MS Excel. In our analysis we examined the linear relationship of one of the dependent variables(burglary) against other independent variables (other crimes). We have used the multiple regression model with 8 independent variables. We notice a r-square with a value of 0.081 which is very close to 0 than 1, thus we can say that the independent variable does not significantly influence dependent variable. The adjusted r-square 0.797 suggests that 7.97% of the variation in burglary crimes is explained by the variation in other crimes.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.28563779
R Square	0.081588947
Adjusted R Sc	0.079777353
Standard Erro	306.6653057
Observations	553

ANOVA

	df	SS	MS	F	Significance F
Regression	1	4611716.435	4611716.435	49.03806275	7.33802E-12
Residual	552	51912072.57	94043.60972		
Total	553	56523789			

	Coefficients	tandard Erro	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Violent crime	0.024838687	0.003547006	7.00271824	7.32456E-12	0.017871405	0.031805968	0.017871405	0.031805968

RESIDUAL OUTPUT

Observation	cted Principal	Residuals	ndard Residuals
1	13.56192298	-12.561923	-0.04100006
2	30.20384312	-28.2038431	-0.09205273
3	21.06320639	-18.0632064	-0.05895535
4	14.77901863	-10.7790186	-0.03518095
5	8.991604613	-3.99160461	-0.01302794
6	155.1175989	-149.117599	-0.48669545
7	21.93256042	-14.9325604	-0.04873743
8	13.7109551	-5.7109551	-0.01863962
9	0	9	0.029374528
10	3.874835137	6.125164863	0.019991536
11	35.1219031	-24.1219031	-0.07872995

PROBABILITY OUTPUT

Percentile	Principal cities
0.090415913	1
0.27124774	2
0.452079566	3
0.632911392	4
0.813743219	5
0.994575045	6
1.175406872	7
1.356238698	8
1.537070524	9
1.717902351	10
1.898734177	11

Fig 9. Regression analysis

Cluster Analysis:

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters.

We have incorporated cluster analysis for our data set to identify clusters of cities with similar crime rates. To identify our clusters, we have used non-hierarchical algorithm for our large data set. We chose '5' clusters of data to identify records closest to the centroid. K-means clustering helped us understand and identify cluster of cities with high or low crimes for the large data set.

Variables									
# Selected Variables	9								
Selected Variables	Violent cri	Murder ar	Rape1	Robbery	Aggravate	Property c	Burglary	Larceny- t	Motor vehicle theft

Cluster Centers

Cluster	iolent crim	nnegligent	Rape1	Robbery	ravated ass	operty crim	Burglary	arceny- the	or vehicle theft
Cluster-1	639.5723	8.72327	62.75472	176.7799	391.3145	4301.862	786.8491	3080.27	434.7421
Cluster-2	2292.525	35	195.9672	686.4262	1375.131	12364.25	2274.934	8665.754	1423.557
Cluster-3	192.8696	2.06689	22.50836	46.2107	122.7592	1393.856	256.8294	1063.278	99.66221
Cluster-4	26545.83	350	1734.167	9346	15115.67	98402.83	13680.33	74165.83	10556.67
Cluster-5	7296.143	113.1071	548.6071	2506.429	4128	33670.46	6471.536	22249.21	4949.714

Distance Between Centers	Cluster-1	Cluster-2	Cluster-3	Cluster-4	Cluster-5
Cluster-1	0	10167.11	3634.327	123092.5	36696.2
Cluster-2	10167.11	0	13798.21	112964	26543.7
Cluster-3	3634.327	13798.21	0	126700.7	40318.84
Cluster-4	123092.5	112964	126700.7	0	86651.83
Cluster-5	36696.2	26543.7	40318.84	86651.83	0

Data Summary

Cluster	#Obs	Avg. Dist
Cluster-1	159	1443.527
Cluster-2	61	3843.064
Cluster-3	299	847.093
Cluster-4	6	23283.07
Cluster-5	28	11085.39
Overall	553	2110.882

Fig 10. K-means cluster analysis

Future Prospects:

In an attempt to reduce crimes in United States. Certain areas will be under surveillance for safety purpose. Analyzing this data might also give answers to many related questions like which cities need more screening to reduce its crime rate.

In the future, law enforcements can even expect intelligent suggestions from Analysis of Crime rates regarding what skills and surveillance they need to provide to the specific city. Suggestions will also include the crime patterns for the cities with maximum crime occurrences.
In order to reduce the crime rate FBI should use this analysis, to track the crime pattern and deploy more investigations in affected regions.

Conclusion:

The analysis of the Crime rate dataset helped us to find safe and cities with maximum crimes. We were able to clean the data by using different techniques and were able to determine which cities have increasing crimes and needs more law enforcements. We also got skillful hands on experience of how tableau can be used to visualize and analyze data.

References:

https://ucr.fbi.gov, FBI's Uniform crime reporting program

www.coursera.org, Data Visualization and Communication with Tableau

http://elearning.ufl.edu/, Business Intelligence