Big Data Analytics for Forecasting of Crime Data

Anusha K[1], Dr M.K Jayanti Kannan[2]

[1]Anusha k, M.Tech, Department of CSE, FET, Jain University, Karnataka, India.
[2]Dr M.K Jayanti Kannan, Professor, Department of CSE, FET, Jain University, Karnataka, India.

Abstract -*The main aim of this project is to explore the data and then draw out the distinct questions that come to mind at first glance of the data and try to answer those questions using the machine learning based crime prediction. Some interesting facts and patterns are discovered from criminal data which is been taken from the Kaggle it contains the data set of crime in Vancouver. Which includes the data from 2003-01-01 to 2017-07-13 it contains 530,652records. The Predictive model shows that the prophet model handles outliers well and also it is robust to missing data and shifts in the trend. These outcomes are going to benefit the police departments to better understand crime issues and provide insights that will enable them to track activities, predict the likelihood of incidents, effectively deploy resources and optimize the decision-making process*.

Key Words: **Data visualization, Crime Prediction, Machine Learning, Data analysis, Big data analytics(BDA).**

1. INTRODUCTION

Day by day the crime rate is increasing so much that it has become difficult to predict the crime because the crime is neither systematic nor random. Since there is a more hi-tech methods and also due to modern technologies all this help the criminals to achieve their misdeeds. According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder, sex abuse, gang rape etc have been increased. Maybe we cannot predict who all may be the victims of crime but surely, we can predict the place that has the probability for the occurrence of the crime [1].

Criminal activities are present mostly in every region of the world because of this they are affecting the quality of life and also the socio-economic development. Previous researches in crime prediction have predicted that there are various factors that affects the crime rate may be like education, poverty, employment, and climate [2]. Dealing with the crime is a major concern for many government organizations and also, they are using different advanced technology to tackle such issues. Crime Analysis, is a sub branch of criminology, that studies the behavioral pattern of criminal activities and tries to identify the indicators of such events.

Big data analytics (BDA) is a systematic approach for analyzing and identifying different patterns, relations, and trends within a large volume of data.BDA can be applied to a criminal data where data analysis is conducted for visualization and trends prediction. Several the state-of-the-art data mining and deep learning techniques are used and also the predictive results show that the Prophet model and Keras stateful LSTM perform better than neural network models [3].

In recent years, Big Data Analytics (BDA) has become an emerging approach for analyzing data and extracting information and their relations in a wide range of application areas [4].Analysis of such big data help us to effectively keep track of occurred events, identify similarities from incidents, deploy resources and make quick decisions accordingly [5]. This can also help further our understanding of both historical issues and current situations, ultimately ensuring improved safety/security and quality of life, as well as increased cultural and economic growth.

One of the main key technologies is the Machine learning which is an innovative, interdisciplinary, and growing research area, which can build paradigms and techniques across various fields for deducing useful information.

Machine learning is a technique in which the computers make decisions without any human intervention. Machine learning have been applied to many fields like self-driving cars, speech recognition, web search, and an improved understanding of the human genome. Machine-learning-based crime analysis usually involves data collection, classification, pattern identification, prediction, and visualization. Traditional data mining techniques association analysis, classification and prediction, cluster analysis, and outlier analysis identify patterns in structured data while newer techniques identify patterns from both structured and unstructured data [6].

The data was collected from the Kaggle it contains the dataset of crime in Vancouver. Which includes the data from 2003-01-01 to 2017-07-13 it contains 530,652 records.

The approach to this project was to explore the data firstly, then draw out the distinct questions that come to mind a first glance of the data:

- What's the distribution of crimes per day?
- Which days have the highest and lowest average number crimes?
- Is crime decreasing or increasing?
- Is this trend the same for all categories?
- Is there any trend within the year?
- What are the various types of crime in Vancouver?

- Which are the safest and dangerous area in Vancouver?
- Which hour has the maximum crime rate over the given period of time?

All these above questions are been answered in this project and also the prophet model is being used.

2. LITERATURE SURVEY

Combating the criminal activity is always been a priority for the governments around the world, many researches has been done to effectively find counter measures and indicators of crime prior to happening. Criminologists have been pursuing to identify hotspots that need major attention from law enforcement agencies. Various researches have proposed different crime prediction algorithms. The accuracy of prediction mainly depends on the attributes selected and the dataset used as a reference.

BDA has become a prominent approach for analyzing and extracting the information. Due to continuous urbanization and growing populations, cities play important central roles in our society. However, such developments have also been accompanied by an increase in violent crimes and accidents. To tackle such problems, sociologists, analysts, and safety institutions have devoted much effort towards mining potential patterns and factors [7]. In relation to public policy however, there are many challenges in dealing with large amounts of available data.

As a result, new methods and technologies need to be devised in order to analyze this heterogeneous and multi-sourced data [8]. Analysis of such big data enables us to effectively keep track of occurred events, identify similarities from incidents, deploy resources and make quick decisions accordingly [9].

Data mining is useful in not only the discovery of new knowledge or phenomena but also for enhancing our understanding of known ones. With the support of such techniques, BDA can help us easily identify crime patterns which occur in a particular area and how they are related with time. The implications of machine learning and statistical techniques on crime or other big data applications such as traffic accidents or time series data, will enable the analysis, extraction and understanding of associated patterns and trends, ultimately assisting in crime prevention and management.

The Location-Based Social Networks (LBSN) using that we can collect a vast range of information which can help us to understand the regional dynamics (i.e. human mobility) across an entire city. LBSN provides unprecedented opportunities to tackle various social problems. While crime event prediction has been investigated widely due to its social importance, its success rate is far from satisfactory. The existing studies rely on relatively static features such as regional characteristics, demographic information and the topics obtained from tweets but very few studies focus on exploring human mobility through social media.

Traditionally, crime event prediction uses the historical patterns of crime events, the information collected from geographic information systems (GIS) and demographic variables, e.g. sex, income, age, and race and so on. But these variables are almost constant or only change slowly over time. Therefore, they do not capture the short-term variations of the factors which are relevant to the occurrence of crime events[18].With the widespread use of social media such as Twitter and Foursquare in the last decade, large volumes of information have been generated which provide unprecedented opportunities to capture the city dynamics, i.e. human mobility across a city [19].

By using Geographical Analysis, we can predict that there are various techniques to map hotspots, but among them, the choropleth mapping is widely used to describe the geographic information of crime incidents [24]. Choropleth map represents the proportion of statistical measurements or its density with shaded colors. This makes it easy to recognize where crime incidents are more condensed, which gives insights into criminal behavior. Geographic Information System (GIS) has been used as a powerful analytical tool for crime mapping. It shows the locations of crime series with various geographic information on one map, which helps police officers to make decisions for operational and tactical purposes [25].

3. METHODOLOGY

This chapter deals with various methodologies adopted during this project.

3.1 Data Collection

In data collection step usually, we collect data from different web sites like news sites, blogs, social media, RSS feeds etc. In this the data for analysis is collected from Kaggle which is openly available. The data set which is been used is the Crime in Vancouver. It was extracted on 2017-07-18 and it contains 530,652 records from 2003-01-01 to 2017-07-13.

3.2 Featured Attributes

For each entry of crime incidents in the datasets, the following 12 featured attributes are included:

- Type
- Year
- Month
- Day
- Hour
- Minute

- Hundred_Block
- Neighborhood
- X
- Y
- Latitude
- Longitude

3.3 Data Preprocessing

Before implementing any algorithms on our datasets, a series of preprocessing steps are performed for data conditioning as presented below:

- Defaulting records by filling the blank data for "HOUR" column to "00".
- Blank records for "NEIGHBOURHOOD" as "N/A".
- Blank records for "HUNDRED_BLOCK" as "N/A".
- Deleting "MINUTE" column as predicting information to the actual minute is not necessary here.
- Adding "NeighbourhoodID" column as a category key ID for Neighbourhood.
- Adding "CrimeTypeID" column as a category key ID for "TYPE" as in type of crime.
- Adding "Incident" column as a row count to keep track of incident totals per crime type, etc.
- Combining date fields and adding a column "Date" format.
- Using "Date" to get weekday names.

3.4 Narrative Visualization

Considering the geographic nature of the crime incidents, an interactive map based on Google map was used for data visualization, where crime incidents are clustered according to their latitude/longitude information.

3.5 Prediction Models

In order to tackle the problem of crime trends forecasting I have explored several state-of-the-art machine learning and deep learning algorithms and time series models.

3.6 Time series model

A time series is a sequence of numerical data points successively indexed or listed/graphed in the time order. Usually, the successive data points within a time series are equally spaced in time, hence these data are discrete in time.

3.7 Prophet model

The Prophet model is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and/or daily seasonality, plus holiday effects.

It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

The Prophet model is designed to handle complex features in time series, it also designed to have intuitive parameters that can be adjusted without knowing the details of the underlying model. The Prophet model decomposes time series into three main components, i.e. the trend, seasonality, and holidays. They are combined in the following equation:

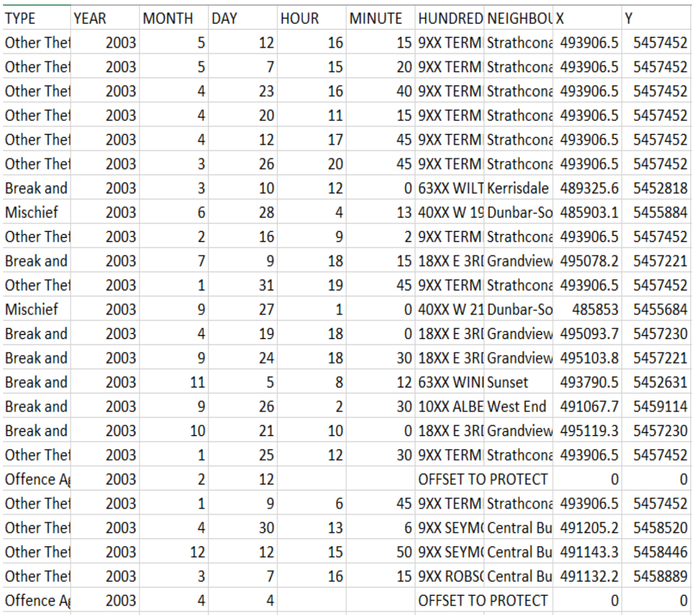
$$y(t) = g(t) + s(t) + h(t) + \varepsilon t \quad (1)$$

where g(t) is the trend of any non-periodic changes in the time series, s(t) represents periodic changes (e.g., weekly and yearly seasonality), and h(t) represents the holiday effects of any potentially irregular schedules over one or more days. The error term εt represents any random effects which are not accommodated by the model.

4. RESULTS

4.1 Data Preprocessing

The figure 4.1 shows the original dataset which has incomplete data like empty cells, unnecessary columns and several irrelevant features.

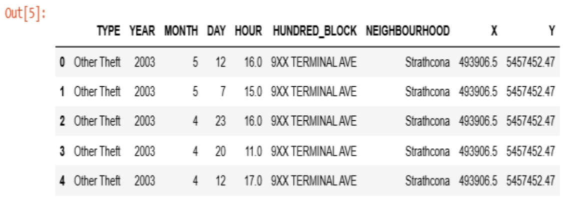


TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED	NEIGHBOU	X	Y
Other Thef	2003	5	12	16	15	9XX TERMI	Strathcona	493906.5	5457452
Other Thef	2003	5	7	15	20	9XX TERMI	Strathcona	493906.5	5457452
Other Thef	2003	4	23	16	40	9XX TERMI	Strathcona	493906.5	5457452
Other Thef	2003	4	20	11	15	9XX TERMI	Strathcona	493906.5	5457452
Other Thef	2003	4	12	17	45	9XX TERMI	Strathcona	493906.5	5457452
Other Thef	2003	3	26	20	45	9XX TERMI	Strathcona	493906.5	5457452
Break and	2003	3	10	12	0	63XX WILT	Kerrisdale	489325.6	5452818
Mischief	2003	6	28	4	13	40XX W 19	Dunbar-So	485903.1	5455884
Other Thef	2003	2	16	9	2	9XX TERMI	Strathcona	493906.5	5457452
Break and	2003	7	9	18	15	18XX E 3RI	Grandview	495078.2	5457221
Other Thef	2003	1	31	19	45	9XX TERMI	Strathcona	493906.5	5457452
Mischief	2003	9	27	1	0	40XX W 21	Dunbar-So	485853	5455684
Break and	2003	4	19	18	0	18XX E 3RI	Grandview	495093.7	5457230
Break and	2003	9	24	18	30	18XX E 3RI	Grandview	495103.8	5457221
Break and	2003	11	5	8	12	63XX WINI	Sunset	493790.5	5452631
Break and	2003	9	26	2	30	10XX ALBE	West End	491067.7	5459114
Break and	2003	10	21	10	0	18XX E 3RI	Grandview	495119.3	5457230
Other Thef	2003	1	25	12	30	9XX TERMI	Strathcona	493906.5	5457452
Offence Ag	2003	2	12			OFFSET TO PROTECT		0	0
Other Thef	2003	1	9	6	45	9XX TERMI	Strathcona	493906.5	5457452
Other Thef	2003	4	30	13	6	9XX SEYMO	Central Bu	491205.2	5458520
Other Thef	2003	12	12	15	50	9XX SEYMO	Central Bu	491143.3	5458446
Other Thef	2003	3	7	16	15	9XX ROBSO	Central Bu	491132.2	5458889
Offence Ag	2003	4	4			OFFSET TO PROTECT		0	0

Figure 4.1 Original Dataset

4.2 Remove unnecessary columns

Column MINUTE can be deleted as it's of no need in this level. In figure 4.2 we can see the output where the MINUTE is been removed.



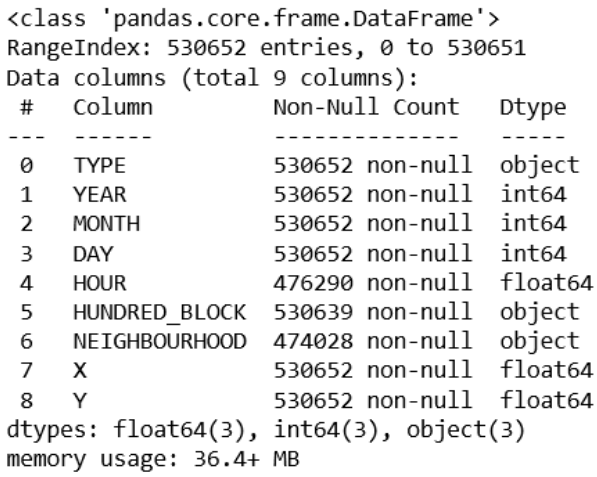
Out[5]:

	TYPE	YEAR	MONTH	DAY	HOUR	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y
0	Other Theft	2003	5	12	16.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47
1	Other Theft	2003	5	7	15.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47
2	Other Theft	2003	4	23	16.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47
3	Other Theft	2003	4	20	11.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47
4	Other Theft	2003	4	12	17.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47

Figure 4.2

4.3 Check for Missing values

In Figure 4.3 We can see that we have 530,652 entries, but some columns (HOUR, HUNDRED_BLOCK and NEIGHBOURHOOD) have less, which means that there are missing values.



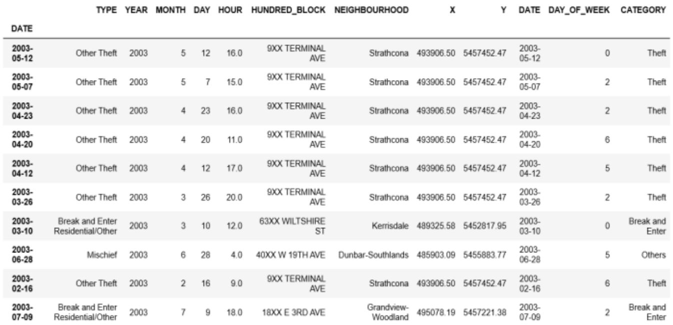
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 530652 entries, 0 to 530651
Data columns (total 9 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   TYPE           530652 non-null  object
 1   YEAR           530652 non-null  int64
 2   MONTH          530652 non-null  int64
 3   DAY            530652 non-null  int64
 4   HOUR           476290 non-null  float64
 5   HUNDRED_BLOCK  530639 non-null  object
 6   NEIGHBOURHOOD  474028 non-null  object
 7   X              530652 non-null  float64
 8   Y              530652 non-null  float64
dtypes: float64(3), int64(3), object(3)
memory usage: 36.4+ MB
```

Figure 4.3 Missing Values

4.4 Filling the Missing Values

In Figure 4.4 we can see the data set which is been preprocessed the following steps are carried out in order to have a preprocessed dataset.

- Defaulting records by filling the blank data for "HOUR" column to "00".
- Blank records for "NEIGHBOURHOOD" as "N/A".
- Blank records for "HUNDRED_BLOCK" as "N/A".
- Deleting "MINUTE" column as predicting information to the actual minute is not necessary here.
- Adding "NeighbourhoodID" column as a category key ID for Neighbourhood.
- Adding "CrimeTypeID" column as a category key ID for "TYPE" as in type of crime.
- Adding "Incident" column as a row count to keep track of incident totals per crime type, etc.
- Combining date fields and adding a column "Date" format.
- Using "Date" to get weekday names.



	TYPE	YEAR	MONTH	DAY	HOUR	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y	DATE	DAY_OF_WEEK	CATEGORY
DATE												
2003-05-12	Other Theft	2003	5	12	16.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	2003-05-12	0	Theft
2003-05-07	Other Theft	2003	5	7	15.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	2003-05-07	2	Theft
2003-04-23	Other Theft	2003	4	23	16.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	2003-04-23	2	Theft
2003-04-20	Other Theft	2003	4	20	11.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	2003-04-20	6	Theft
2003-04-12	Other Theft	2003	4	12	17.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	2003-04-12	5	Theft
2003-03-26	Other Theft	2003	3	26	20.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	2003-03-26	2	Theft
2003-03-10	Break and Enter Residential/Other	2003	3	10	12.0	63XX WILTSHIRE ST	Kerrisdale	489325.58	5452817.95	2003-03-10	0	Break and Enter
2003-06-28	Mischief	2003	6	28	4.0	40XX W 19TH AVE	Dunbar-Southlands	485903.09	5455883.77	2003-06-28	5	Others
2003-02-16	Other Theft	2003	2	16	9.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	2003-02-16	6	Theft
2003-07-09	Break and Enter Residential/Other	2003	7	9	18.0	18XX E 3RD AVE	Grandview-Woodland	495078.19	5457221.38	2003-07-09	2	Break and Enter

Figure 4.4 preprocessed datasets.

4.5 What's the distribution of crimes per day?

In the Figure 4.5 we can see that the distribution looks like a *normal distribution* with a mean around *95 crimes per day*. There is one outlier over 600. Let's find out which day it is.

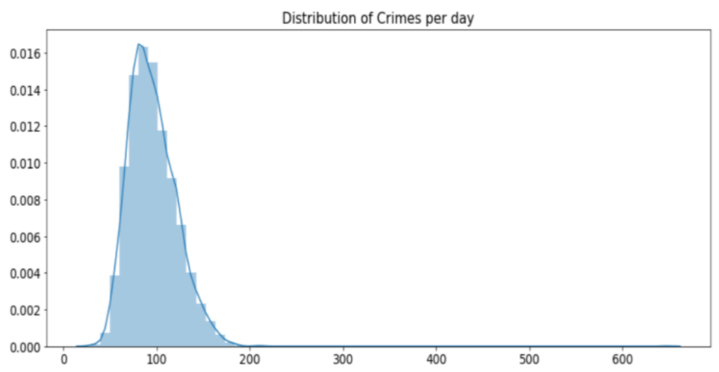
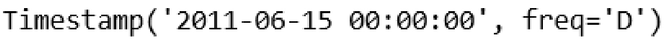


Figure 4.5 Distribution of crime per day

4.6 Outlier

From the above observation we have found out that there is an outlier over 600 in the Figure 4.6 we can find out the day which had that outlier. The day was 2011-06-15.



```
Timestamp('2011-06-15 00:00:00', freq='D')
```

Figure 4.6 Outlier

4.7 Time series graph

The Figure 4.7 shows the time series graph per day from that we can understand that some days over the Control Limits, indicating *signals*. Also, the period of 2003 to 2008 is above the average.

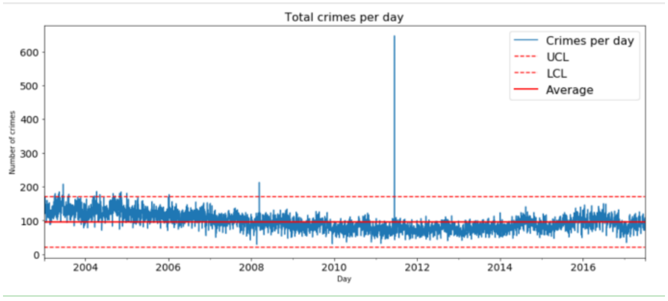


Figure 4.7 Total crimes per day

4.8 Which days have the highest and lowest average number crimes?

The Figure 4.8 shows a heat map where in which it states that Blue means good days, Red bad days, White average days. By the figure 4.8 we can analyze that the calmest day of crime is Christmas Day, the worst day is New Year's Day, January 1. The first day of the month is a busy day for all months. Halloween (October 30,31 and November 1) are also dangerous days. The second week of summer months are usually the most dangerous.

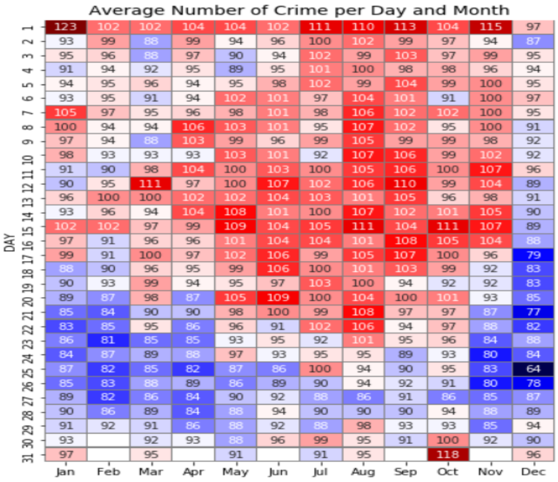


Figure 4.8 Crime Heat map per day

4.9 Is crime decreasing or increasing?

Figure 4.9 shows the crimes per month where Average number of crimes per month decreased from 4,000 crimes per month to around 2,400 in the period of 2003 to 2011. From 2011 to 2014, the moving average was around the same. But from 2014, the average has increased and 2016 reached similar levels of 2008.

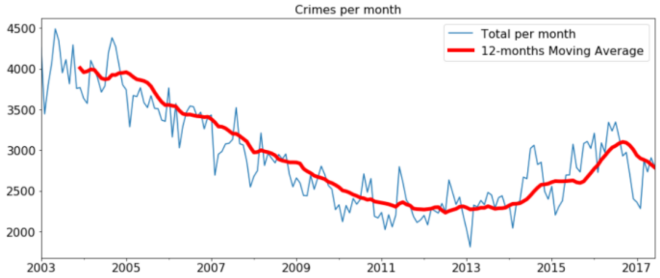


Figure 4.9 Crimes per month

4.10 What are the various types of crime?

The below figure 4.10 shows the way in which the crime has been categorized.

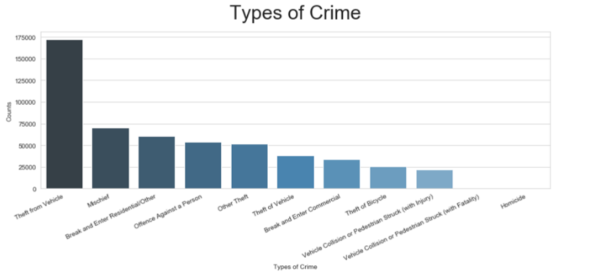


Figure 4.10 Types of crime

4.11 Trends over the years of the crime in Vancouver

The figure 4.11 shows that the year 2003 had the highest crime and gradually it decreased from 2005 to 2009 around 2010 to 2013 it was stable and it gradually increased from 2014 to 2016 and there was a decrease in the year 2017.

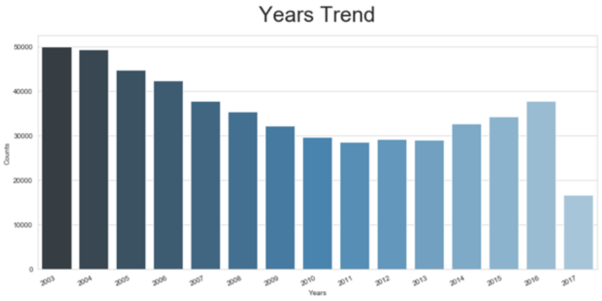


Figure 4.11 Years Trend

4.12 Heat map with months and categories

From figure 4.12 we can make out that the Break and Enter has most incidents in January. Theft and Others are more frequent during summer months. Maybe because people go out more and there is an increased number of tourists? December is not a "good month" for Theft and Others.

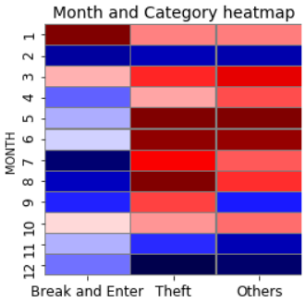
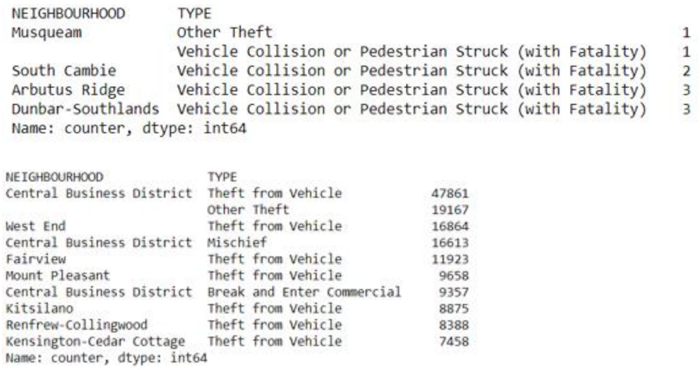


Figure 4.12 Heat map with months and categories

4.13 Which are the Safest and Dangerous area in Vancouver?

The safest area is Musqueam, south cambia, Arbutus Ridge etc are some of the safest area where the West End, Fair view etc are the dangerous area the figure 4.13 gives the detail information about the safest and dangerous area.



```
NEIGHBOURHOOD      TYPE
Musqueam           Other Theft                                               1
                   Vehicle Collision or Pedestrian Struck (with Fatality)    1
South Cambie       Vehicle Collision or Pedestrian Struck (with Fatality)    2
Arbutus Ridge      Vehicle Collision or Pedestrian Struck (with Fatality)    3
Dunbar-Southlands  Vehicle Collision or Pedestrian Struck (with Fatality)    3
Name: counter, dtype: int64
```

```
NEIGHBOURHOOD              TYPE
Central Business District  Theft from Vehicle            47861
                           Other Theft                   19167
West End                   Theft from Vehicle            16864
Central Business District  Mischief                      16613
Fairview                   Theft from Vehicle            11923
Mount Pleasant             Theft from Vehicle             9658
Central Business District  Break and Enter Commercial     9357
Kitsilano                  Theft from Vehicle             8875
Renfrew-Collingwood        Theft from Vehicle             8388
Kensington-Cedar Cottage   Theft from Vehicle             7458
Name: counter, dtype: int64
```

Figure 4.13 Safest and Dangerous area

4.14 What hours do crimes happen?

Most crimes happen between 17:00 and 20:00 and the Break and Enter has some activity between 3 and 5 am. Theft doesn't occur much in those hours between 3 and 5 as expected, because most people are sleeping.

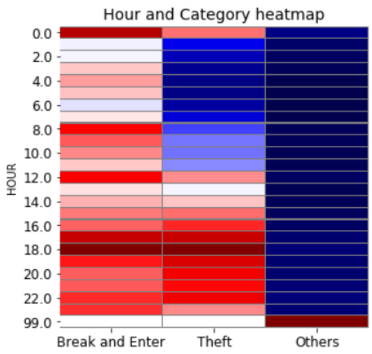


Figure 4.14 Hours and Category Heatmap

4.15 Exploring Category Theft

We saw that before that most crimes are in the category "Theft". Let's understand it in a better way.

```
<Figure size 432x288 with 0 Axes>
```

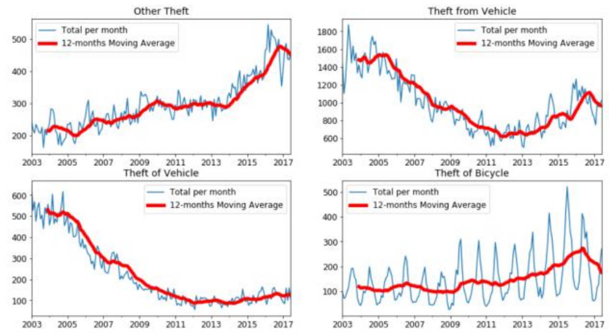


Figure 4.15 Exploring category Theft

Theft of Vehicle: It had a major decrease, from an average of around 520 crimes per month in 2003 to around 100 in 2012. Although the average has been increasing in the past years, it's way below 2003.

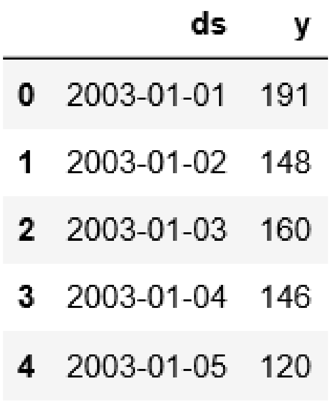
Other Theft: It has been in increasing, from around 200 to almost 500 crimes per month.

Theft from Vehicle: It decreased along with "Theft of Vehicle" until 2012, but then it increased significantly.

Theft of Bicycle: It has peak during summer month.

4.16 Prophet Model

In order to work with prophet model, we have to first process the dataset accordingly the Figure 4.16 shows the dataset which is been processed according to the prophet model.



	ds	y
0	2003-01-01	191
1	2003-01-02	148
2	2003-01-03	160
3	2003-01-04	146
4	2003-01-05	120

Figure 4.16 Processed Dataset

4.17 Without removing Outliers (Plain Model)

The Figure 4.17 states that the black dots are the actual values and the blue line is the predicted value. The light blue lines are the lower and upper intervals.

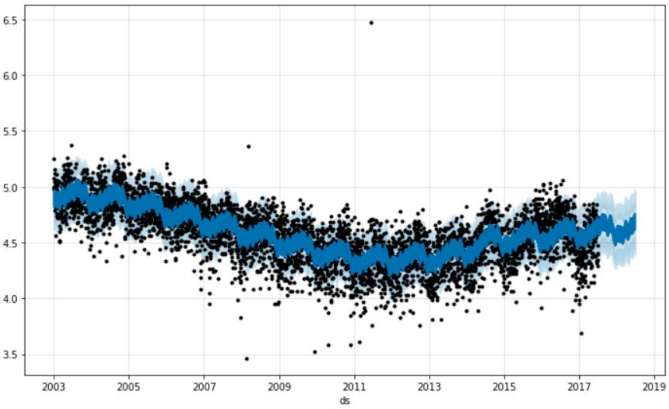


Figure 4.17 Without Removing Outliers

4.18 Forecast components

Prophet displays the general, weekly and monthly trends. In Figure 18 it shows that in the first block the number of crimes decreased until 2011 and then start increasing. The weekly subplot shows that Friday and Saturday have more crimes. The yearly subplot shows that summers months have higher number of crimes.

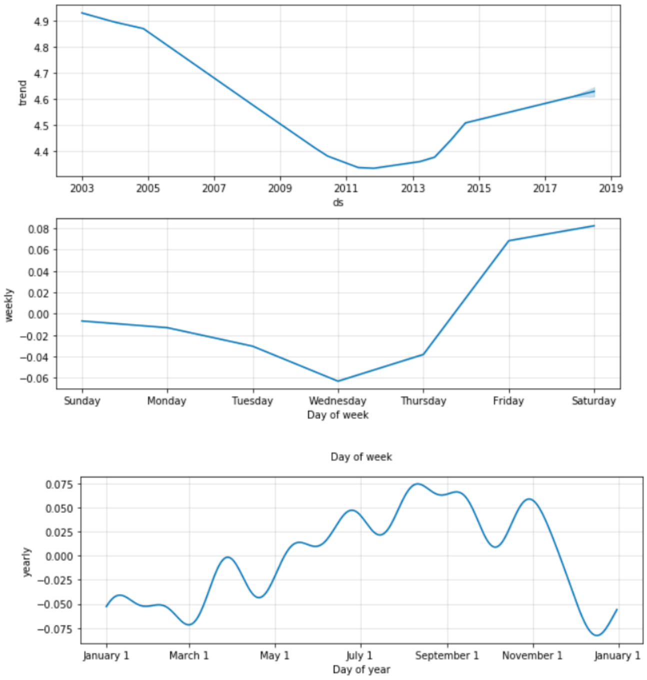


Figure 4.18 Forecast components

4.19 Measure the error

To measure the error for this model the Mean Absolute Percentage Error is used (MAPE) the error is 12.64.

4.19.1 Removing Outliers

By removing the outliers, the model performed only a little better (0.01%) than the plain model. This is probably because the outliers that were considered were only a few data points (18 of 5295). They were not a sequence of data points that compromised the trend, so in this case, they didn't make much difference. Prophet handled them well in the first plain model. The error is 12.63

4.19.2 Including Holidays

Three sets of holidays have been considered.

One with no added window: Mother's Day, Victoria Day, Canada Day, Labor Day, Remembrance Day, Christmas.

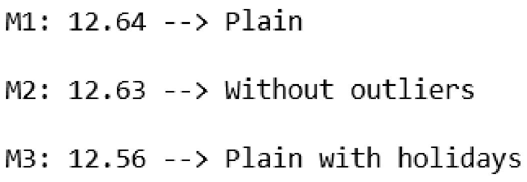
Another with a -1 lower and 1 upper window: Halloween, New Year's

And another with a -2 lower and a 1 upper window

The error measure is 12.56 which indicates the holidays have a better result.

4.19.3 Comparing Models

The Figure 4.19.3 shows the comparison of all the above stated models. We could analyze that the holidays have the better result.



```
M1: 12.64 --> Plain

M2: 12.63 --> Without outliers

M3: 12.56 --> Plain with holidays
```

Figure 4.19.3 Models comparison

4.20 Holiday Plot

In figure 4.20 we can see that there is a a peak every year that model 1 didn't have. This is the result of adding a holiday that had a relevant impact. This is going to be reflected in the forecast.

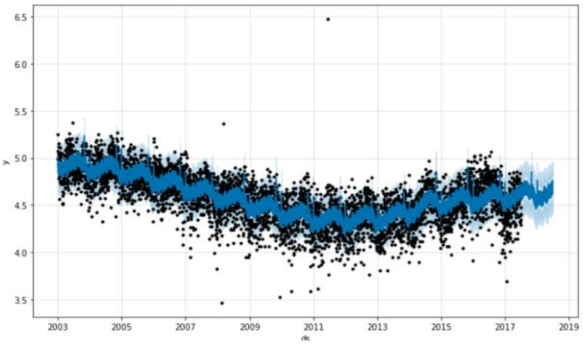
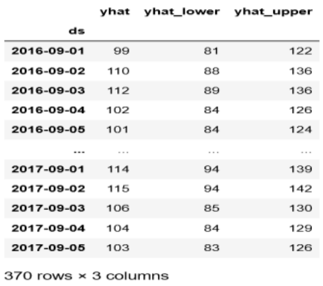


Figure 4.20 Holiday Plot

4.21 Forecast

To see the forecasted numbers, we can check the forecast data frame for example, the forecasted number of crimes for 2016-09-01 is 99, with an interval of 81 - 122.



ds	yhat	yhat_lower	yhat_upper
2016-09-01	99	81	122
2016-09-02	110	88	136
2016-09-03	112	89	136
2016-09-04	102	84	126
2016-09-05	101	84	124
...	...	...	...
2017-09-01	114	94	139
2017-09-02	115	94	142
2017-09-03	106	85	130
2017-09-04	104	84	129
2017-09-05	103	83	126

370 rows × 3 columns

Figure 4.21 Forecast

4.22 Data Visualization

The goal of this is to show the data on a geographical map of Vancouver. These are the following things which are considered.

- where are car theft happening overall?
- when and where would be the worst place to park?

4.22.1Where is car theft happening overall?

The Figure 4.22.1 gives a geographical representation

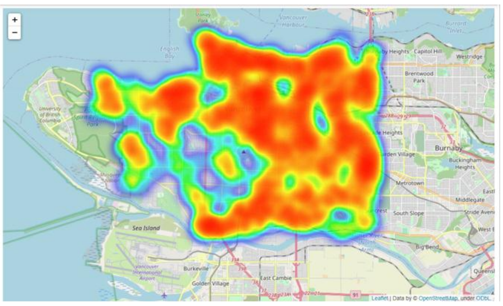
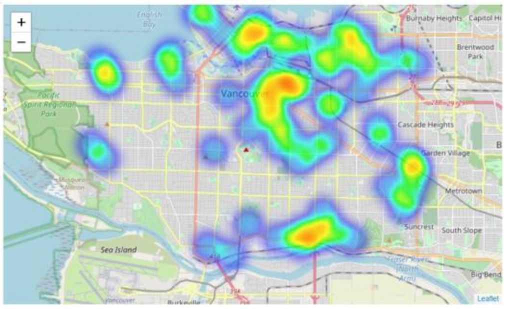


Figure 4.22.1 Geographical representation of car theft

4.22.2 When and where would be the worst place to park?



5. CONCLUSIONS

In this the Vancouver crime data for last 14 years was used which had around 530,652 records and the approach for this project was to explore the data and then draw out the distinct question that come to at the first glance of the data. It was been able to answer all the questions which were raised and also the prophet model was used which handled well with the outliers, missing data and also dramatic changes in the time series. The results provide the new insights into crime trends and will assist both police departments and law enforcement agencies in their decision making. In future the more importance can be given to the multivariant visualization and also for spatial analysis to uncover more potential patterns and trends within these datasets.

REFERENCES

[1] S. Sathyadevan, D. M. S and S. G. S., "Crime analysis and prediction using data mining," 2017 First International Conference on Networks & SoftComputing (ICNSC2014), Guntur,2017, pp.406-412.

[2] H. Adel, M. Salheen, and R. Mahmoud, "Crime in relation to urban design. Case study: the greater Cairo region," Ain Shams Eng. J., vol. 7, no. 3, pp. 925-938, 2016.

[3] M. Feng et al., "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data," in IEEE Access, vol. 7, pp. 106111-106123,2019.

[4] A. Gandomi and M. Haider, ''Beyond the hype: Big data concepts, methods, and analytics,'' Int. J. Inf. Manage., vol. 35, no. 2, pp. 137–144, Apr. 2015.

[5] A. Agresti, An Introduction to Categorical Data Analysis, 3rd ed. Hoboken, NJ, USA: Wiley, 2018.

[6] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," IEEE Computer, vol.37, no. 4, pp. 50-56, Apr. 2004.

[7] H. Hassani, X. Huang, M. Ghodsi, and E. S. Silva, "A review of datamining applications in crime," Stat. Anal. Data Mining, ASA Data Sci. J., vol. 9, no. 3, pp. 139–154, Apr. 2016.

[8] Z. Jia, C. Shen, Y. Chen, T. Yu, X. Guan, and X. Yi, "Big-data analysis of multi-source logs for anomaly detection on network-based system," in Proc. 13th IEEE Conf. Autom. Sci. Eng. (CASE), Xi'an, China, Aug. 2017, pp. 1136–1141.

[9] A. Agresti, An Introduction to Categorical Data Analysis, 3rd ed. Hoboken, NJ, USA: Wiley, 2018.

[10] M. Huda, A. Maseleno, M. Siregar, R. Ahmad, K. A. Jasmi, N. H. N. Muhamad, and P. Atmotiyoso, "Big data emerging technology: Insights into innovative environment for online learning resources," Int.J. Emerg. Technol. Learn., vol. 13, no. 1, pp. 23–36, Jan. 2018.

[11] J. Zakir and T. Seymour, "Big data analytics," Issues Inf. Syst., vol. 16, no. 2, pp. 81–90, 2015.

[12] U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns," in Proc. 2nd Asian Conf. Defence Technol., Chiang Mai, Thailand, 2016, pp. 123–128.

[13] L. Stopar, P. Skraba, D. Mladenic, and M. Grobelnik, "StreamStory: Exploring multivariate time series on multiple scales," IEEE Trans. Vis.Comput. Graphics, vol. 25, no. 4, pp. 1788–1802, Apr. 2019.

[14] Y. Gao, Y. Xia, J. Qiao, and S. Wu, "Solution to gang crime based on graph theory and analytical hierarchy process," Neurocomputing, vol. 140, pp. 121–127, Sep. 2014.

[15] L. Guo, X. Cai, F. Hao, D. Mu, C. Fang, and L. Yang, "Exploiting finegrained co-authorship for personalized citation recommendation," IEEE Access, vol. 5, pp. 12714–12725, 2017.

[16] Malathi, A., & Baboo, S. S. (2011). An enhanced algorithm to predict a future crime using data mining.

[17] Corcoran, J. J., Wilson, I. D., & Ware, J. A. (2003). Predicting the geo-temporal variations of crime and disorder. International Journal of Forecasting, 19(4), 623-634.

[18] Cohn EG (1990) Weather and crime. Br J Criminol 30(1):51–64

[19] Zheng Y, Capra L, Wolfson O, Yang H (2014) Urban computing: concepts, methodologies, and applications. ACM Trans Intell Syst Technol 5(3):38

[20] Bowers K, Johnson S (2014) Crime mapping as a tool for security and crime prevention. In: Gill M (ed) The handbook of security. Springer, Berlin, pp 566–587

[21] Steenbeek W, Weisburd D (2015) Where the action is in crime? An examination of variability of crime across different spatial units in the Hague, 2001–2009. J Quant Criminol 32(3):449–469. https ://doi.org/10.1007/ s1094 0-015-9276-3

[22] Deadman D (2003) Forecasting residential Burglary. Int J Forecast 19(4):567–578

[23] Weisburd D, Cave B, Piquero AR (2016) How do criminologists interpret statistical explanation of crime? A review of quantitative modeling in published studies. In: Piquero AR (ed) The handbook of criminological theory. Springer, Berlin, pp 395–414

[24] M. V. Barnadas, Machine learning applied to crime prediction, Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, Sep. 2016.

[25] S. Chainey and J. Ratcliffe, GIS and crime mapping, John Wiley & Sons, Incorporated, 2015.

[26] Arunima S. Kumar, Raju K. Gopal, Data Mining Based CrimeInvestigation System: Taxonomy and Relevance, IEEE, 2015, ISBN: 978-1-4799-8553-1/15