Crime Prediction and Analysis in South Africa

Thendo Manthada

Table of Content

Executive Summary

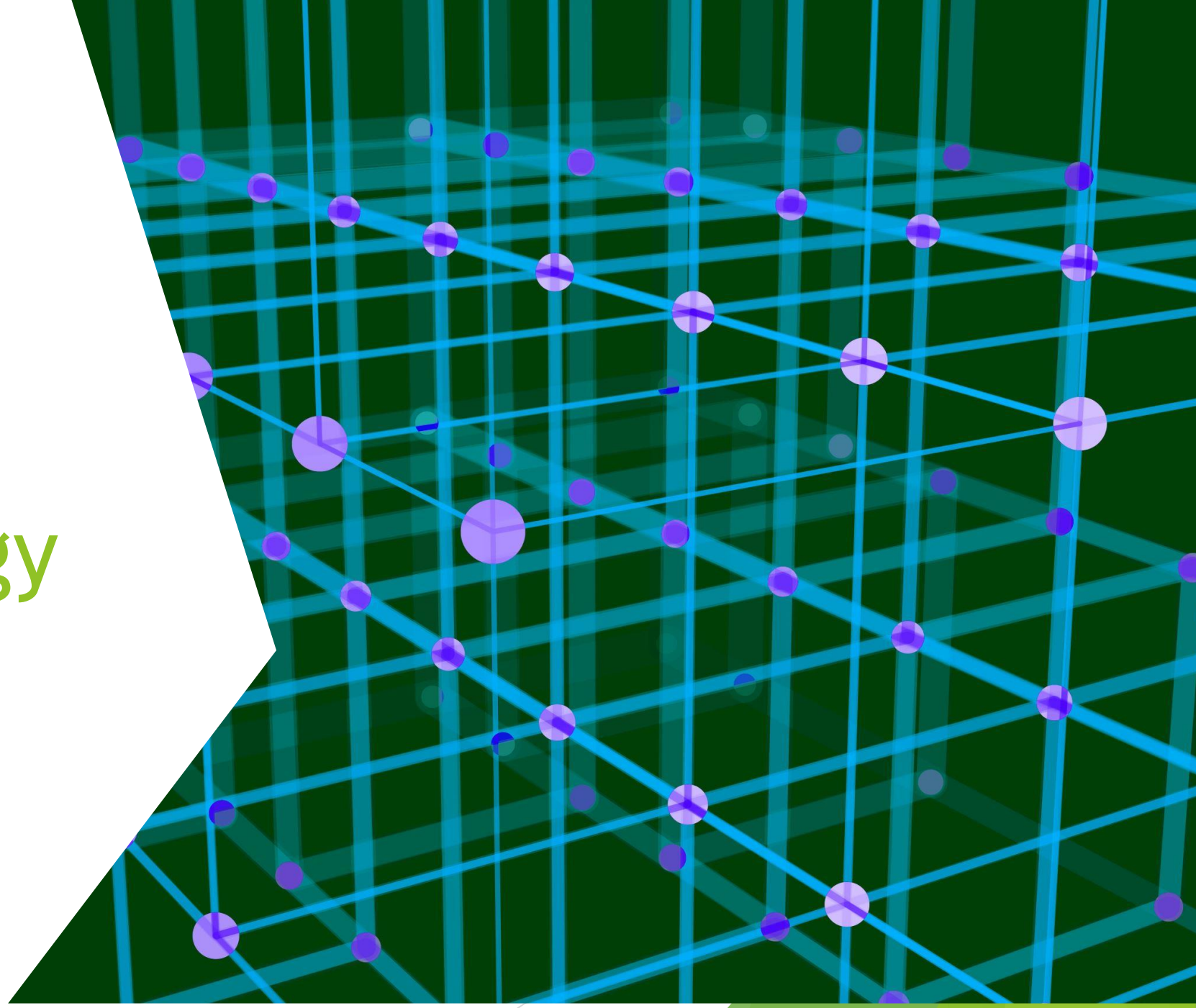
- The aim of this project is to make crime prediction using the features present in the dataset. The dataset is extracted from the official sites. With the help of machine learning, using python as core we can predict the type of crime which will occur in a particular area.
- The objective would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. Building the model will be done using better algorithm depending upon the accuracy. The K-Nearest Neighbor (KNN) classification and other algorithm will be used for crime prediction.
- Visualization of dataset is done to analyze the crimes which may have occurred in the country. This work will help the law enforcement agencies to predict and detect crimes in South Africa with improved accuracy and thus reduces the crime rate.

Introduction

Crime is defined as any harmful act against the public which the State wishes to prevent and which, upon conviction, is punishable by fine, imprisonment, and/or death. Historically solving crimes has been the privilege of legal enforcement specialists. But lately computer analysts have aided the legal enforcement in solving crimes by discovering crime patterns. Crime rate is affected by number of Economic factors such as Income level, unemployment rate and Gross Domestic Product (GDP), Consumer Price Index (CPI) etc.

An economic crisis is a situation where the economy of a country experiences a sudden downturn as a result of financial crisis. Whether there is economic crisis or not, the economic indicators like unemployment have a direct impact on the South African population and thus affect crime rates. Due to financial crisis, there is an increase in unemployment and large number of population suffers sudden reduction in their income.

Methodology



Methodology

- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data collection & Data wrangling

Data Collection

- Crime dataset from South African statistic (StatsSA) is used in CSV format.
- Crime dataset from South African Police services is used in CSV format indicating the number of crimes reported across the country in various police stations
- Crime dataset from South African Police services is used in CSV format indicating the number of crimes categories across the country.

Data wrangling

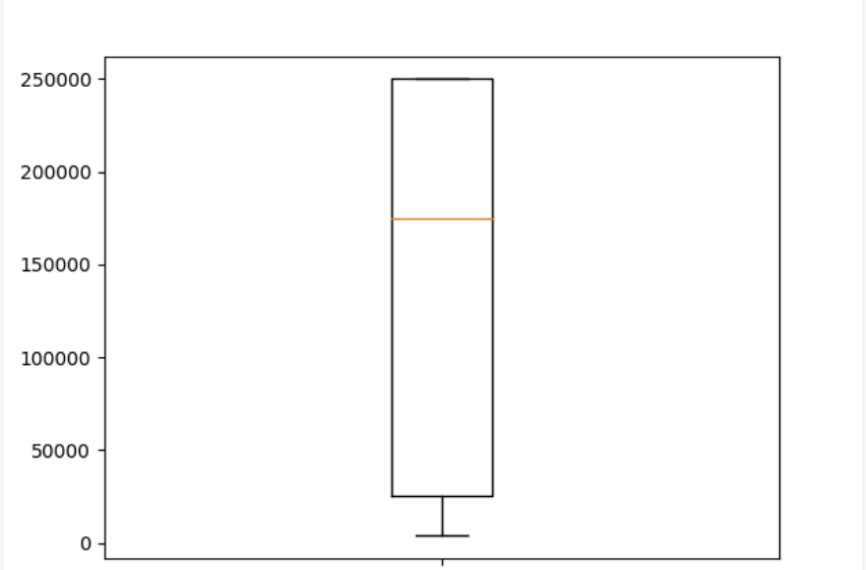
- Data was 1st captured into an excel spreadsheet and then converted to csv files.
- Removed any missing values present in the dataset
- The dataset consists of a 3 Data frames with multiple columns .
- The final dataset contains 32,9 & 30 rows which speaks to 9Provinces.

Exploratory data analysis (EDA) using visualization

EDA with Data Visualization

```
In [4]: print(df)
```

	Province	Total Crime count
0	Gauteng	42335
1	Eastern Cape	163432
2	Free State	87656
3	Limpopo	94218
4	Kwazulu-Natal	251245
5	North-west	88837
6	Mpumalanga	944878
7	Western Cape	426289
8	Northern Cape	317529



Exploratory data analysis (EDA) using visualization

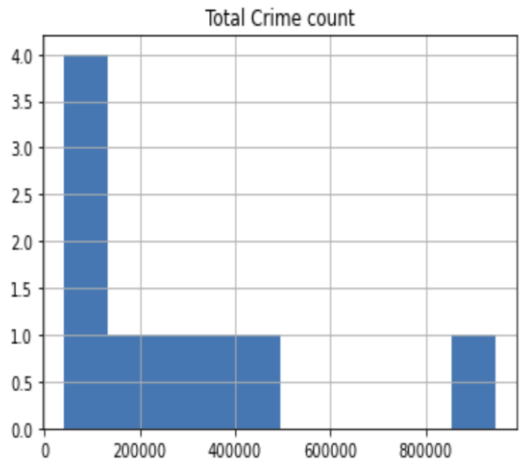
EDA with Data Visualization

- The below graph show the total number of crime committed ,a python code was used in order to plot the graphs accordingly

```
In [17]: df = pd.DataFrame(df, columns = ['Province', 'Total Crime count'] )
```

```
In [18]: # create histogram for numeric data
         df.hist()

         # show plot
         plt.show()
```

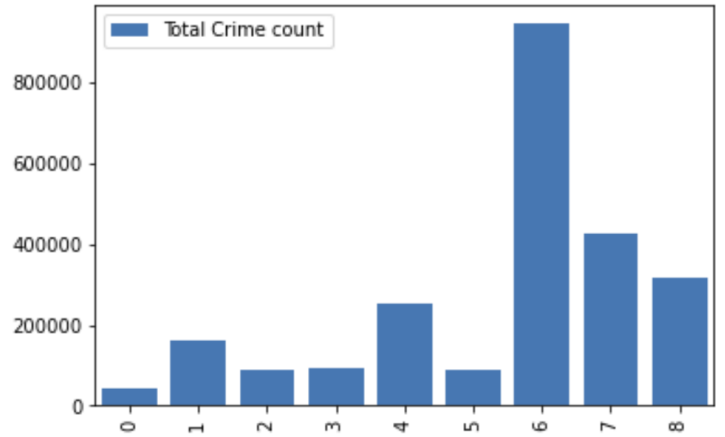



```
# Dataframe of previous code is used here

# Plot the bar chart for numeric values
# a comparison will be shown between

df.plot.bar()

# plot between 2 attributes
plt.bar(df['Province'], df['Total Crime count'])
plt.show()
```



Exploratory data analysis (EDA) using SQL

Summary of SQL Analysis performed

- Total number of crimes per province across the country
- Total number of crimes reported in different police stations across years
- Total number of crimes per categories across various years
- Provinces (Areas)with the highest crimes

Exploratory data analysis (EDA) using SQL

- The Sql table below show the total number of crimes per province across the country

select * from crimestats

Run time: **0.024 s**

Result set 1

Find

PROVINCE	TOTAL_CRIME_COUNT
Gauteng	42335
Eastern Cape	163432
Free State	87656
Limpopo	94218
Kwazulu-Natal	251245

Result set is truncated, only the first 9 rows have been loaded. Select "View all loaded data" on the right top of the result to view all loaded rows.

More

Exploratory data analysis (EDA) using SQL

- The Sql table below show the total number of crimes reported in different police stations across many years

STATION	PROVINCE	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
MITCHELLS PLAIN	WESTERN CAPE	5 850	6 310	6 044	4 768	4 609	4 914
DURBAN CENTRAL	KWAZULU-NATAL	2 304	2 336	2 528	2 477	1 724	1 969
KLEINVLEI	WESTERN CAPE	2 049	2 149	1 874	1 561	1 541	1 720
BISHOP LAVIS	WESTERN CAPE	2 753	2 977	2 577	2 738	2 472	2 898
ATLANTIS	WESTERN CAPE	1 267	1 424	1 669	1 853	2 078	2 638
PHOENIX	KWAZULU-NATAL	1 452	1 785	2 043	2 323	2 653	3 224
POLOKWANE	LIMPOPO	208	321	414	511	508	720
DELFT	WESTERN CAPE	2 770	3 381	2 953	3 035	3 439	2 926
CAPE TOWN CENTRAL	WESTERN CAPE	1 832	1 963	2 149	2 360	2 712	2 796
LENTEGEUR	WESTERN CAPE	0	0	1 126	1 933	2 050	2 444
STEENBERG	WESTERN CAPE	1 352	1 630	1 273	1 456	1 831	1 698
KRAAIFONTEIN	WESTERN CAPE	2 017	2 340	2 515	3 357	3 332	4 502

Exploratory data analysis (EDA) using SQL

- The Sql table below show the total number of crimes per categories across various years

Result set 1

Find

CRIME_CATEGORY	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
Murder	15 554	16 213	17 023	17 805	18 673	19 016
Sexual Offences	60 539	60 888	56 680	53 617	51 895	49 660
Attempted murder	14 730	16 236	16 989	17 537	18 127	18 205
Assault with the intent to inflict grievous bodily harm	191 612	185 050	182 333	182 556	182 933	170 616
Common assault	180 165	171 653	166 081	161 486	164 958	156 450
Common robbery	52 566	53 196	53 505	54 927	54 110	53 418
Robbery with aggravating circumstances	100 769	105 488	118 963	129 045	132 527	140 956
Rape	47 069	48 408	45 349	43 195	41 503	39 828
Sexual Assault	7 194	6 967	6 597	6 087	6 212	6 271
Attempted Sexual Offences	3 535	3 293	2 913	2 641	2 573	2 073
Contact Sexual Offences	2 741	2 220	1 821	1 694	1 607	1 488
Carjacking	9 417	9 931	11 180	12 773	14 602	16 717

Exploratory data analysis (EDA) using SQL

- The Sql table below shows provinces (Areas)with the highest crimes

Result set 1

PROVINCE	TOTAL_CRIME_COUNT
Mpumalanga	944878
Western Cape	426289
Northern Cape	317529
Kwazulu-Natal	251245
Eastern Cape	163432
Limpopo	94218
North-west	88837
Free State	87656
Gauteng	42335

Interactive visual analytics using Folium and Plotly Dash

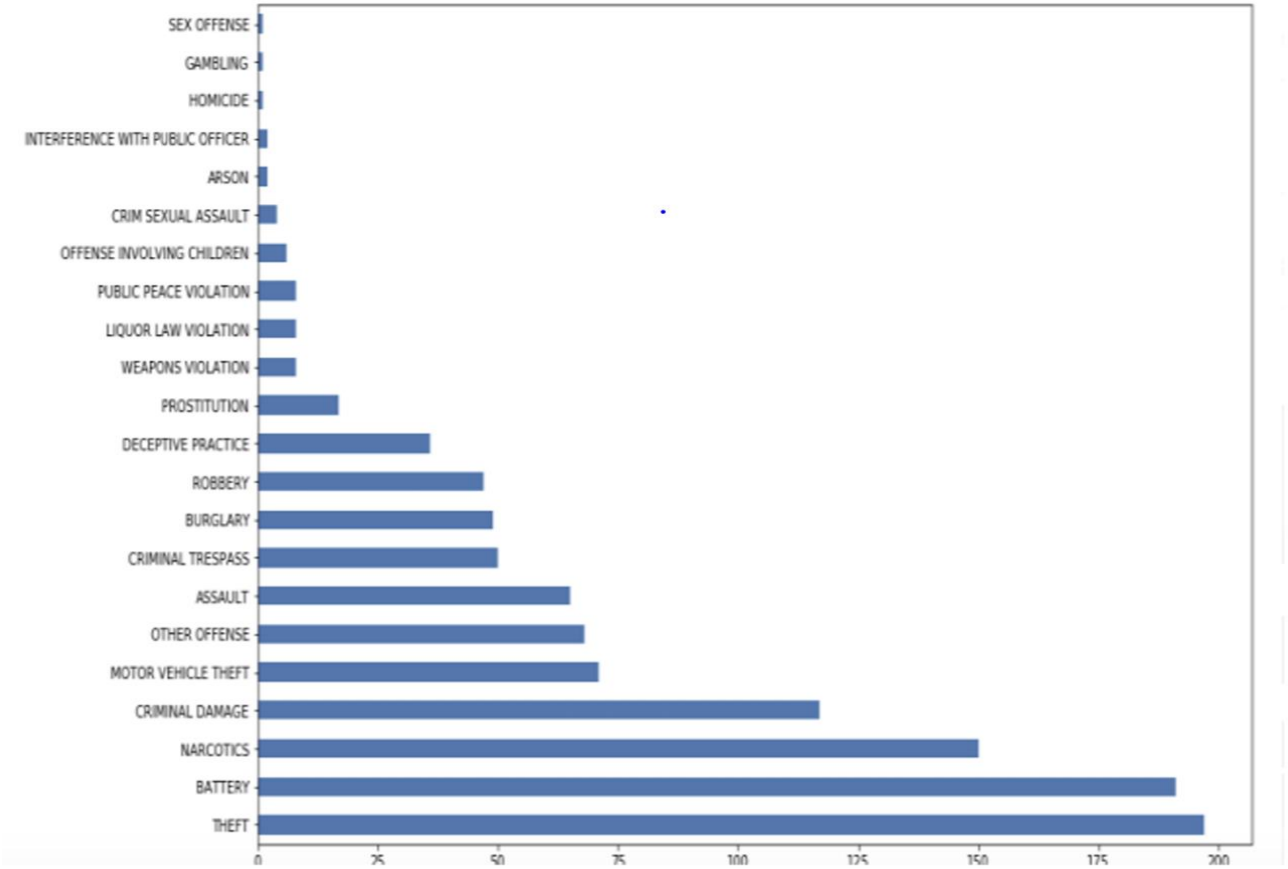
This part deals with the analysis done on the dataset and plotting them into various graphs like bar, scatter, Folium and plotly.

- Analysis done are

1. Types of crimes committed over Time (Month).
2. Types of crimes committed over Time (Hour).
3. No of crimes of all types of crime over various location.

Interactive visual analytics using Folium and Plotly Dash

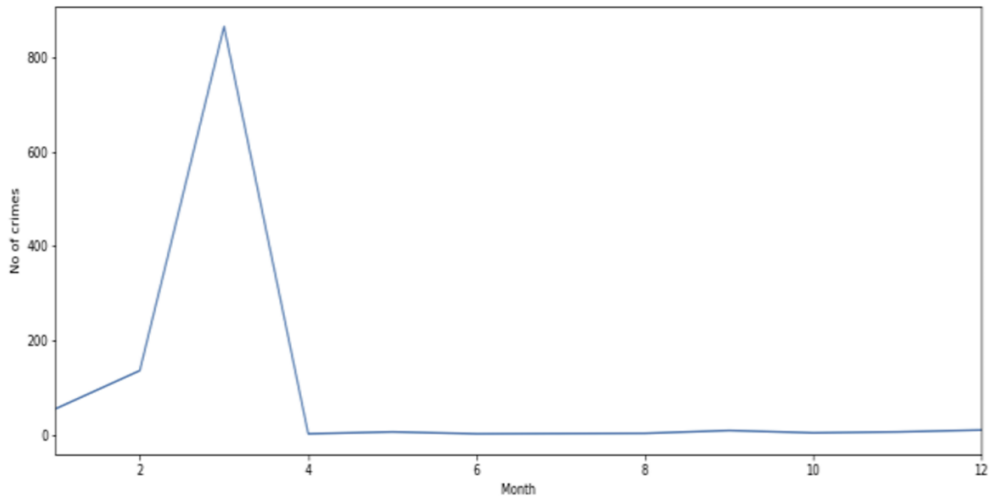
This graph shows which crimes have occurred most in the city. The y coordinate denotes the Types of crimes committed and x coordinate denotes the no of crimes committed. The highest being Theft.

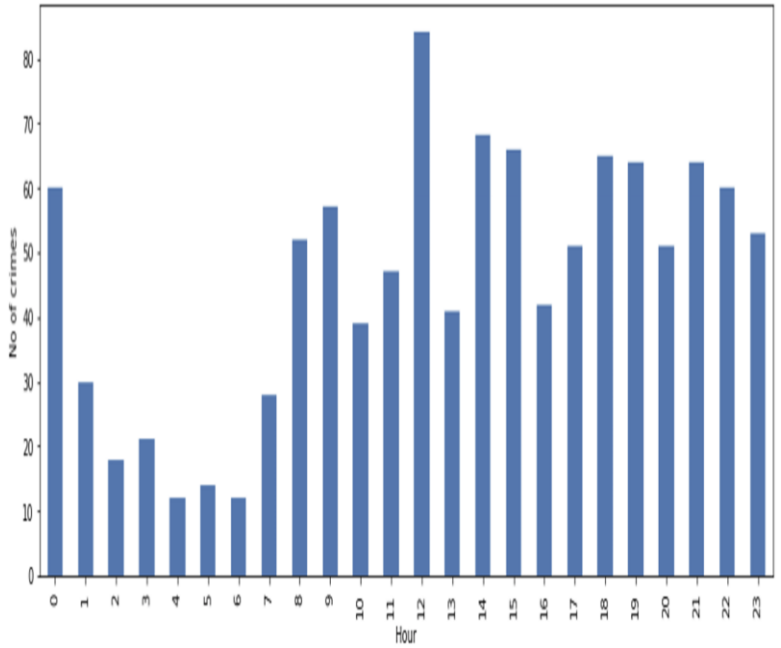


Interactive visual analytics using Folium and Plotly Dash

The graph below tells us about in which month occurrence of crimes is highest. As we can see the month of march is peak where rate of occurrence is high. The x coordinate denotes the month and y coordinated denotes the crime rate.

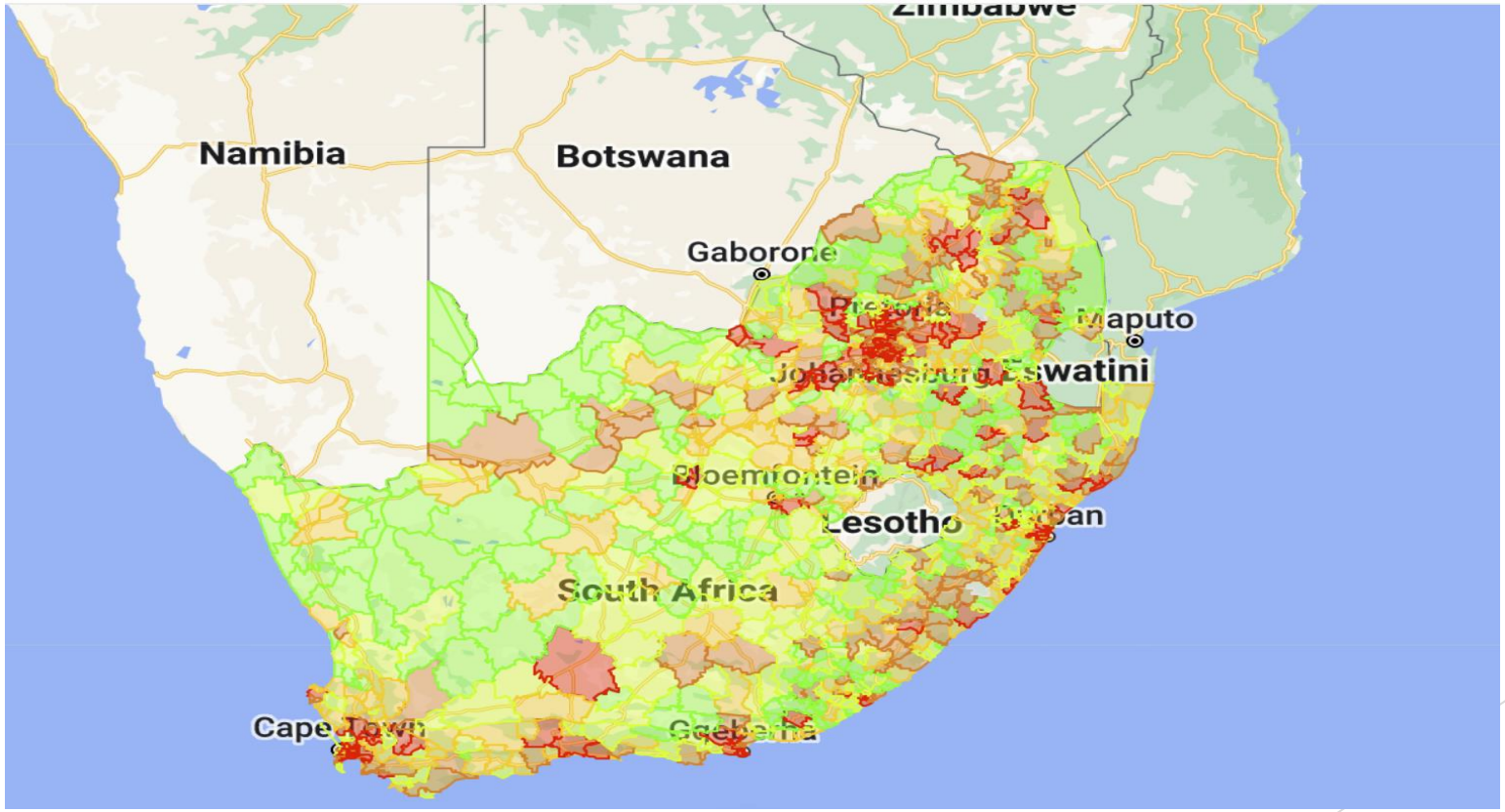
And also shows crime occurrence over particular hour. The x axis is hour and y axis is rate of crimes





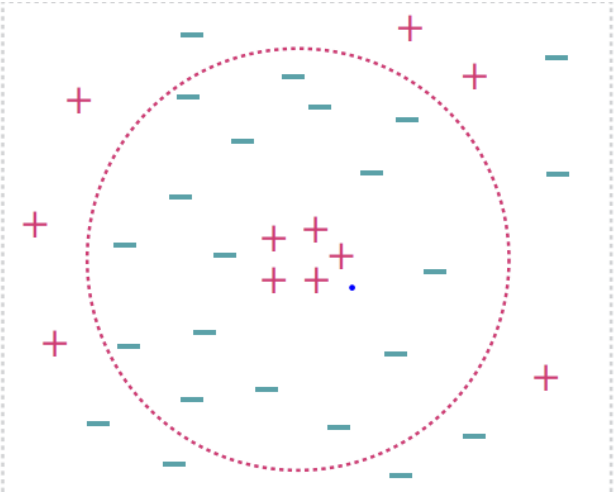
Interactive visual analytics using Folium

- This graph shows which crimes have occurred most in the city. The y coordinate denotes the Types of crimes committed and x coordinate denotes the no of crimes committed.



Predictive Analysis using classification models

- The k-NN k- Nearest Neighbor classifier (k-NN) is a classification algorithm which was used in this assessment ,modeling procedure of the training data until it is necessary to be labelled and classified.
- Data points that a located far from the neighbourhood
- Accuracy was calculated so as to assist in choosing the optimal model by means of the largest value and here the final value k=9 was selected for the model.



k	Kappa	Accuracy
5	0.8543771	0.9330784
7	0.8545504	0.9330784
9	0.8623644	0.9366185

Results

This machine-learning-based crime analysis involves the collection of data, data classification, identification of patterns, prediction, and visualization. K-nearest neighbor (KNN) and boosted decision tree algorithms were also implemented to analyze the crime dataset. In the study, a total of 322,000 crime datasets between 2011 and 2021 were analyzed, and crime prediction with an accuracy of between 85% and 93% was obtained by predicting the crime using ML algorithms.

Conclusion

- The whole analysis assessment indicated the areas of high crime and the stations with the most reported crime across the country, which can easily show the police which area need more attention in terms of more police work and resources needed for the police to do their work, including education on crime and the consequences there off.
- ML predictions show the crimes that are most likely to happen which can help in preventing or minimizing such crime.

Reference

- Crime Stats SA
- Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention | Visual Computing for Industry, Biomedicine, and Art | Full Text (springeropen.com)
- extension://elhekieabhbkpmcefcoobjddigjcaadp/https://www.ajhtl.com/uploads/7/1/6/3/7163688/article_25_9_3_372-381.pdf
- The impact of crime on inbound tourism to South Africa: An application of the bounds test: African Security Review: Vol 22, No 1 (tandfonline.com)
- The Impact of Crime on International Tourist Numbers to Cape Town | SpringerLink
- Latest crime statistics: Murder, kidnapping and commercial crimes increase - The Mail & Guardian (mg.co.za)