**Mini Project Report**

**“Crime Analysis of UK”**



**Submitted by**

SHREYANSH DANGE (T194003)

SHUBHAM CHANDGUDE (T194013))

SANTANU CHOUBEY (T194097)

**Under Supervision of**

Prof. Jayvant H. Devare

Miss. Shobha Mourya



**School of Computer Engineering and Technology**

MIT ACADEMY OF ENGINEERING

November 2019





**Acknowledgments**



We would like to express our gratitude to the School of Computer Engineering and Technology

of MIT Academy of Engineering, who encouraged us to do this Data Analytics project: Crime Analysis in UK.

We wish to express our profound thanks to our project mentors, Mrs. Shobha Mourya and Mr. Jaywant Devare for his full cooperation and guidance and advice for the betterment of our project.

We also want to thank Dr. Yogesh Bhalerao for providing us with the basic infrastructure and

other facilities and for including such interesting courses in the curriculum.

November 2019

Shreyansh Dange (T194003)

Shubham Chandgude (T194012)

Santanu Choubey (T194097)





**Contents**

**ACKNOWLEDGEMENTS** **2**

**CONTENTS** **3**

**1. INTRODUCTION** **3**

OVERVIEW 3

BACKGROUND AND MOTIVATION 3

OBJECTIVE 4

METHODOLOGY 4

**2. DATA SET** **5**

DESCRIPTION 6

NUMBER OF ATTRIBUTES USED 6

**3. SYSTEM WORKING** **10**

BLOCK DIAGRAM **10**

SAMPLE CODE 10

SCREENSHOT RESULT 6

**4. CONCLUSION** **12**

**5. FUTURE WORK**  **14**

**6. REFERENCES** **16**



**INTRODUCTION**

**OVERVIEW**

Crime analysis involves systematic analysis for identifying and analyzing patterns and trends in crime and disorder. Information on patterns can help law enforcement agencies deploy resources in a more effective manner, and assist detectives in identifying and apprehending suspects. Crime analysis also plays a role in devising solutions to crime problems, and formulating crime prevention strategies. Quantitative social science data analysis methods are part of the crime analysis process, though qualitative methods such as examining police report narratives also play a role.

**Crime Analysis in the United Kingdom**

Analysts support policing through the provision and support of four key 'Intelligence Products', these being the Strategic Assessment, the Tactical Assessment, the Problem Profile and the Subject Profile (formerly referred to as the 'Target Profile') and ten 'analytical tools & techniques'. The key skills of an analyst within UK law enforcement must to be identify patterns and trends, make inferences in relation to these patterns, provide recommendations to support action and provide products and briefings that deliver this information and interpretation clearly and in an appropriate format for the audience.

**BACKGROUND AND MOTIVATION**

In a country like the UK the crime incident become serious issue. Day by day the crimes committed by the people is increasing whether it is in the form of theft, rape, robbery, harassment, chain snatching, mob lynching and murder. People are in the more risk now, so some actions should be taken from the UK police. So, to do that, an analysis is necessary which tells the snapshots of crime and the more crowded location of the robbery and other crimes which helps the UK police department to track the location and catch the thieves.

**OBJECTIVES**

1. To track the location of places where the crimes are happening more often.
2. To make inference which type of crimes are frequent
3. To analyze the pattern of crime so that the authorities can prevent such crimes in future
4. To detect the time of the year, which is more prone to incidents of crimes in different areas of the UK.
5. So, using the reports generated by analyzing crime patterns, the authorities become more efficient and smarter in tackling the problem of preventing crime, and make our society a better place to live.

**METHODOLOGY**

**Packages Used:**

1. dpylr- Data Manipulation library
2. leaflet- Interactive Web Maps with the JavaScript 'Leaflet'

**Libraries Used:**

1. tidyverse- coherent system of packages for data manipulation, exploration and visualization
2. rgdal- Bindings for the 'Geospatial' Data Abstraction Library
3. RColorBrewer- Library for color schemes in Maps
4. ClassInt- Univariate Class Intervals library
5. tidyr- Messy data manipulation library
6. lubridate- Time and Date Handling library

**Steps**:

1. Installation and Loading of necessary Libraries
2. Conversion of Dataset “crime\_data.csv” to R suitable format
3. Analysis of whole dataset i.e. “crime.csv” using standard R functionalities
4. Statistical Analysis on Data Frames extracted from the dataset
5. Data Visualization

**DATA SET**

**DESCRIPTION**

Here we use the data of the crimes recorded by Greater Manchester Police between January and December 2015 which were downloaded from the website data.police.UK This is the site for open data about crime and policing in England, Wales and Northern Ireland. We can download street-level crime, outcome, and stop and search data in clear and simple CSV format and explore the API containing detailed crime data and information about individual police forces and neighborhood teams. We use a .csv file which given us in detail information of all crimes that were recorded. Name of the file is crime\_data.csv

**NUMBER OF ATTRIBUTES USED**

It contains various relevant information such as-

1.**date** - It shows the date on which the crime was wrongfully committed. It is stored in “yyyy-mm-dd” format.

2. **Location -** It gives the details of the area, such as the name of the street where the crime was recorded

**3. Borough-** It depicts a town or district in which the area of crime fall administratively

**4. lsoa-** A **Lower Layer Super Output Area (LSOA)** is a GEOGRAPHIC AREA designed to improve the reporting of small area statistics in England and Wales

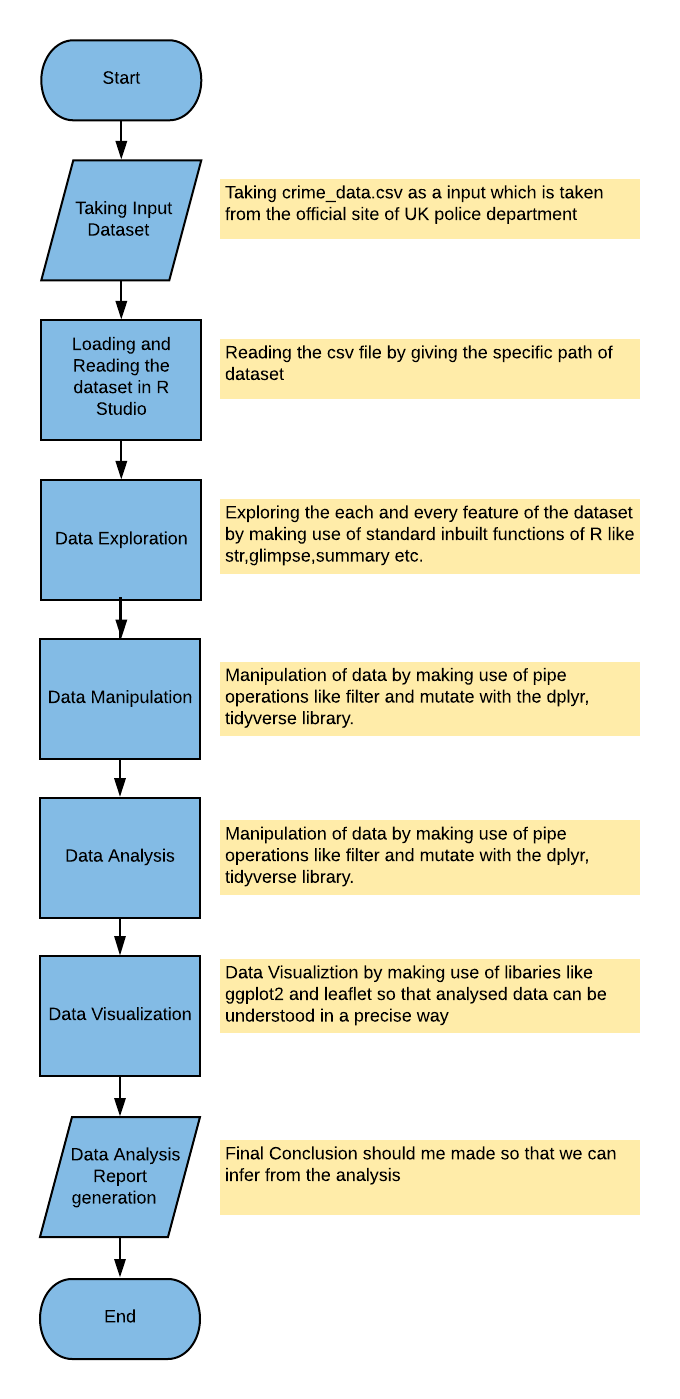
**5. category -** it illustrate the type of crime recorded for the ease of segregate, analyse and solve the crime

**6. long -** It gives the longitude coordinate of the location of crime in the map of the UK. It can be used to plot the map which helps in greater visualisation compared to normal table and charts.

**7 . lat-** It gives the latitude coordinate of the location of crime in the map of UK

**SYSTEM WORKING**

**BLOCK DIAGRAM**



**SAMPLE CODE**

1. Statistical Analysis On dataset “crime\_data.csv”

## Summary statistics ##

# Load the necessary packages

library(tidyverse)

# Set your working directory to where the crime data are stored

setwd("../")

# Read the data

crimes <- read\_csv("crime\_data.csv")

# Frequency of crime by borough (in descending order)

count(crimes, borough, sort = TRUE)

# Frequency of crime category (in descending order)

count(crimes, category, sort = TRUE)

# Frequency and percent of crime category (in descending order)

count(crimes, category, sort = TRUE) %>%

mutate(percent = round(n/sum(n)\*100, 1))

# Frequency of crime category by borough (in descending order)

crimes %>%

group\_by(category, borough) %>%

summarise(n = n()) %>%

ungroup() %>%

arrange(desc(n))

# Mean frequency of crime category per borough

crimes %>%

group\_by(borough, category) %>%

summarise(total = n()) %>%

group\_by(category) %>%

summarise(average = round(mean(total, na.rm=TRUE), 0))

# Frequency and percent of Vehicle crime by borough (in descending order)

crimes %>%

filter(category == "Vehicle crime") %>%

group\_by(borough) %>%

summarise(n = n()) %>%

ungroup() %>%

arrange(desc(n)) %>%

mutate(percent = round(n/sum(n)\*100, 1))

1. Geographical / Web plot for locations where the Robbery happened repeatedly

## Repeat locations ##

# Load the necessary packages

library(tidyverse) ; library(leaflet)

# Set your working directory to where the crime data are stored

setwd("../")

# Read the data

crimes <- read.csv("crime\_data.csv", header = T)

# Identify repeat locations for Robbery offences

crimes %>%

filter(category == "Robbery") %>%

group\_by(long, lat, borough) %>%

summarise(n = n()) %>%

ungroup() %>%

arrange(desc(n))

# Plot the top 10 repeat locations using the leaflet package

rpt\_locs <- crimes %>%

filter(category == "Robbery") %>%

group\_by(long, lat, location, borough) %>%

summarise(n = n()) %>%

ungroup() %>%

arrange(desc(n)) %>%

slice(1:10)

popup <- paste0("<strong>Frequency: </strong>", rpt\_locs$n,

"<br><strong>Location: </strong>", rpt\_locs$location,

"<br><strong>Borough: </strong>", rpt\_locs$borough)

leaflet() %>%

addProviderTiles("CartoDB.Positron") %>%

addCircleMarkers(data = rpt\_locs,

~long, ~lat,

fillColor = "white", color = "red",

radius = ~n, # this may need to be controlled e.g. radius = ~n\*0.1

popup = ~popup)

1. Pdf Generation for 80-20 rule

## The 80-20 rule ##

# A small proportion of places, offenders, victims, and property account for most of the crime.

# Load the necessary packages

library(tidyverse)

# Set your working directory to where the crime data are stored

setwd("../")

# Read the data

crimes <- read.csv("crime\_data.csv", header = T)

# Calculate cumulative frequencies and percentages for Robbery

tbl <- crimes %>%

filter(category == "Robbery") %>%

group\_by(borough) %>%

summarise(n = n()) %>%

ungroup() %>%

arrange(desc(n)) %>%

mutate(percent.crimes = round(n/sum(n)\*100, 1),

cum.percent.crimes = round(cumsum(percent.crimes), 1),

percent.boroughs = 1/n()\*100,

cum.percent.boroughs = round(cumsum(percent.boroughs), 1)) %>%

select(Borough = borough,

`No. crimes` = n,

`% crimes` = percent.crimes,

`Cum. % crimes` = cum.percent.crimes,

`Cum. % boroughs` = cum.percent.boroughs)

# Create a simple table and save as a pdf

library(gridExtra)

pdf("80-20\_rule.pdf", height=11, width=8.5)

grid.table(tbl, rows = NULL)

dev.off()

1. Visualization of Borough level as a proportion of total robbery offences (12/2015)

## Waffle charts ##

# Load the necessary packages

library(dplyr) ; library(tidyr)

# Set your working directory to where the crime data are stored

setwd("../")

# Read the data

crimes <- read.csv("crime\_data.csv", header = T)

# Frequency of crime by borough (in descending order)

df <- crimes %>%

filter(date == "2015-12-01" & category == "Robbery") %>%

count(borough, sort = TRUE) %>%

mutate(percent = round(n/sum(n)\*100, 1)) %>%

select(borough, percent) %>%

spread(borough, percent)

df <- df[, order(-df[which(rownames(df) == '1'), ]) ]

# Use waffle

library(waffle)

waffle(df, rows = 4, size = 2,

colors=(RColorBrewer::brewer.pal(n=10,"Set3")),

title="Borough level Robbery as a proportion of total Robbery offences, 12/2015",

legend\_pos = "bottom")

ggsave("waffle.png", scale=1.5, dpi=300)

1. Crime level in manchester by month

## Heatmap ##

# Load the necessary packages

library(dplyr)

library(tidyr)

# Set your working directory to where the crime data are stored

setwd("../")

# Read the data

crimes <- read.csv("crime\_data.csv", header = T)

## Using ggplot2

library(ggplot2)

df <- crimes %>%

filter(borough == "Manchester") %>%

group\_by(category, date) %>%

summarise(n = n())

ggplot(df, aes(x=category, y=date, fill=n)) +

geom\_tile(aes(fill=n)) +

geom\_text(aes(label=n), size=4, color="black") +

scale\_x\_discrete("", expand = c(0,0)) +

scale\_y\_discrete("", expand = c(0,-2)) +

scale\_fill\_gradient("Frequency", low = "white", high = "steelblue") +

theme\_bw() +

theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +

theme(axis.text=element\_text(size=12)) +

theme(legend.position="none") +

ggtitle("Crime levels in Manchester by month\n") +

theme(plot.title = element\_text(face="bold", size="20"))

ggsave("calendar\_heatmap.png", scale = 1, dpi = 300)

## Using d3heatmap

library(d3heatmap)

df <- crimes %>%

filter(borough == "Manchester") %>%

group\_by(category, date) %>%

summarise(n = n()) %>%

spread(date, n, fill = 0)

row.names(df) <- df$category

df$category <- NULL

d3heatmap(df, scale="column", dendrogram = "none",

color = scales::col\_quantile("Blues", NULL, 5),

# color = scales::col\_bin("Blues", NULL, bins = 7, pretty = TRUE, na.color = "#808080"),

xaxis\_font\_size = 10, yaxis\_font\_size = 10)

1. Category wise and location wise crime using Bar chart

#bar Chart

## Bar charts app ##

library(dplyr) ; library(ggplot2)

crimes <- read.csv("C:/Users/spchandgude/Downloads/crime\_analysis-master/sample\_data/crime\_data.csv", header = T) %>%

group\_by(borough, category)

crimes

ggplot(crimes,aes(x=factor(borough),fill=factor(category ))) + geom\_bar()

1. Percentage of Crime Rates using BoxPlot

#Box Plot

crimes <- read.csv("C:/Users/spchandgude/Downloads/crime\_analysis-master/sample\_data/crime\_data.csv", header = T)

c<-count(crimes, category, sort = TRUE) %>%

mutate(percent = round(n/sum(n)\*100, 1))

c

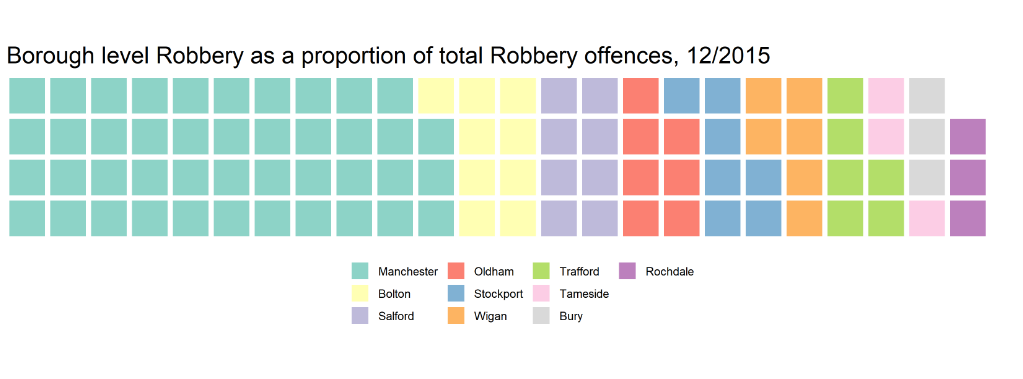
ggplot(c,aes(x=n,y=percent))+geom\_boxplot()

**SCREENSHOT RESULT**

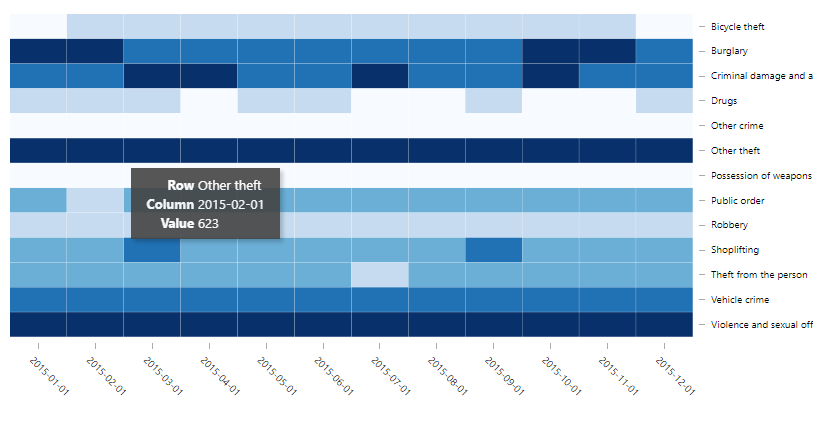
1. **Location For Repeated Crimes**
2. **20 -80 Rule**

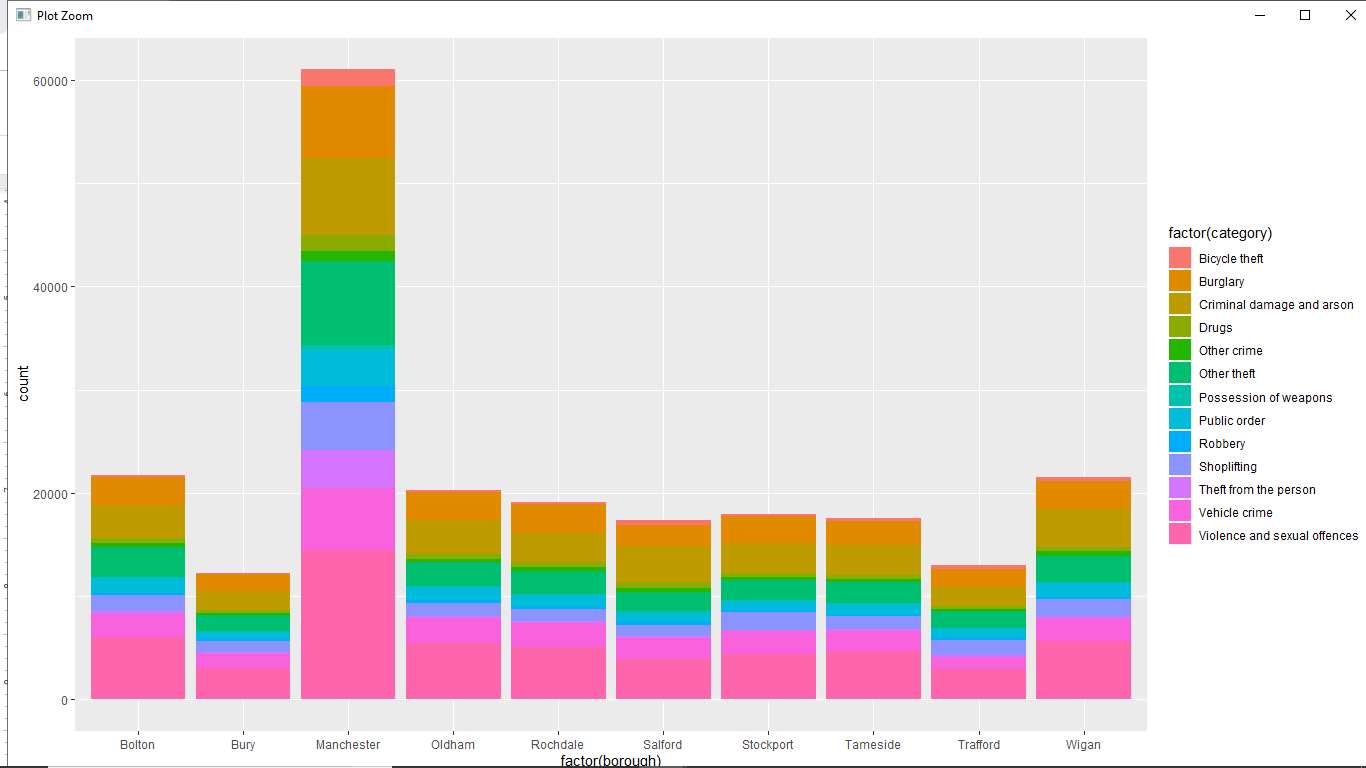


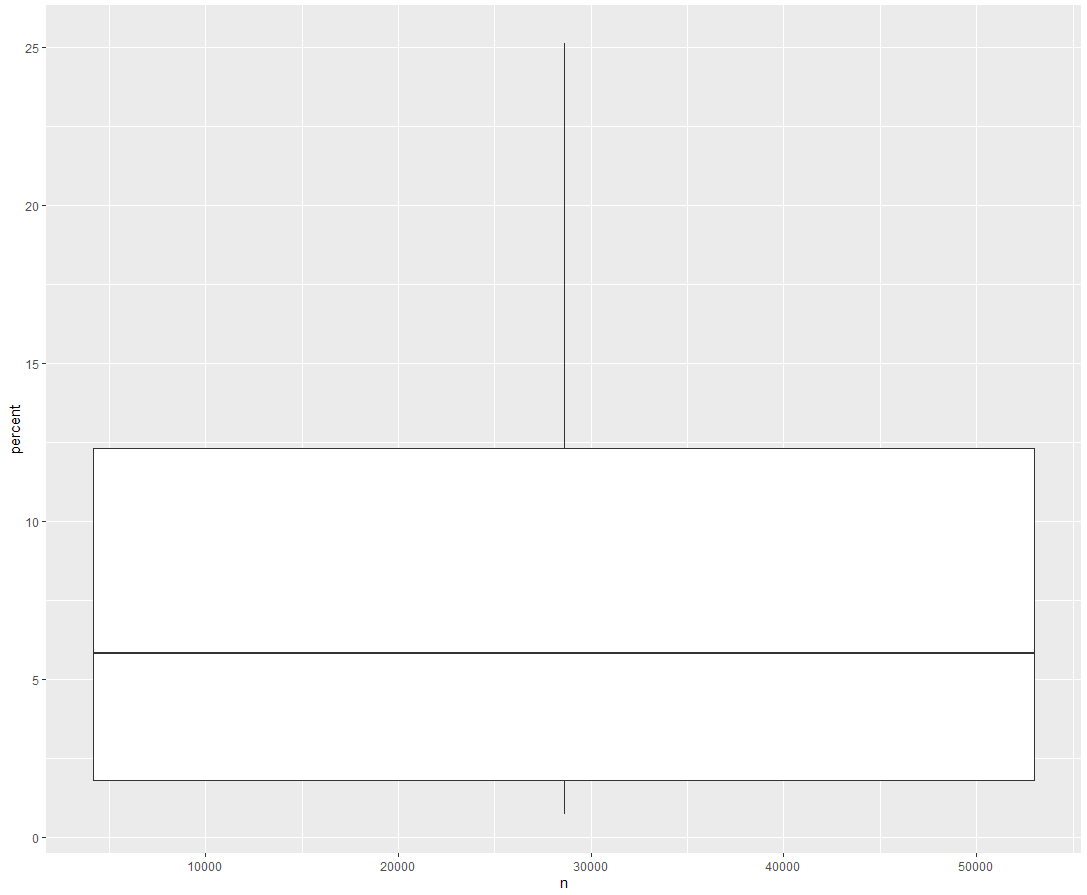
1. **Waffle Chart**



1. **Heat map for Crime level in Manchester by month**



1. **Category wise and location wise crime using Bar chart**
2. **Percentage of Crime Rates using Box plot**



**CONCLUSION**

In conclusion, crime analysis is gradually making its way up the ladder as a very effective method of crime prevention, reduction and apprehension. Crime analysts on the other hand are having a harder time with this process and although there are various ways to make their journey easier, more ways need to be looked into and implemented to ensure that their importance and the need for the roles they play in crime prevention, reduction and apprehension is relayed to police departments and agencies. More training and education is required for both police officers and crime analysts on how to combine both methods of working to create one very strong and effective system. It has already been established that crime analysis is very important and useful to police departments and agencies. It needs to be established that without crime analysts, there would be no crime analysis. Therefore, crime analysts are equally as important, if not more, than crime analysis itself.

**FUTURE WORK**

Machine Learning algorithms can be applied for predicting the future crime rate in UK.

How much Police force is required in which area of the country can be estimated. Ultimately crime rate can be reduced by knowing the prediction of the crime rate.

**REFERENCES**

* https://uk.sagepub.com/en-gb/eur/an-introduction-to-r-for-spatial-analysis-and-mapping/book241031
* https://popcenter.asu.edu/library/reading/pdfs/55stepsUK.pdf
* https://github.com/Robinlovelace/Creating-maps-in-R
* https://data.gov.uk/