**Introduction**

The Crown Prosecution Service (CPS) publishes a report monthly on outcome of CPS proceedings in magistrates' courts and in the crown court by principal offence category. The data set contains XXXXXX monthly analysis data for the twenty four months period from July 2015 after new the new data assurance regime was introduced [1]. The report contains number and percentages of convictions and unsuccessful conviction by defendant basis. One court case can represent several observations in the data set according to the number of defendants [1]. Principal offence is deciding at the time of case finalization they can be one of homicide, offences against the person, sexual offences, burglary, robbery, theft and handling, fraud and forgery, criminal damage, drugs offences, public order, motoring, and all other offences excluding motoring. The data set also contains column for administrative finalisations, the cases which could not proceed due to an administrative issue such as unexecuted warrant for the arrest of the defendant, or summons have not served by the police because they were unable to trace the defendant, or the defendant has died or is unfit to plead. These cases are summarized without categorising by into principal offences [1].

**Data Cleaning**

The column names had long names with spaces which make it hard to deal in R. Therefore, all the long column names renamed to their abbreviations.

The values of the columns with percentage data are character type. First, percentage sign removed, and then values converted to numeric values. Because numeric values, support more analysis method than character values. Which then enable to draw more insight to the data.

Then after round of data visualizing, noticed the data set contained row for the national values of the principal offences court outcomes. As shown in Figure 1.1. national values which are much larger than all the other values, hinders the data visualizing and identifying patterns and prominent features. Hence the national data has separated from the data set and copied to another data frame.

Then another data frame was created combining all the data sets to create a time series data set. As a preparation do the integration, a new variable *Date* was introduced to all the data sets, even the data set only had the month and year, date was added as first day of the respective month for the simplicity.

Column *percentage of L motoring offences unsuccessful* has removed from the data set as this column has values only 100 and NA. Hence the column has no data variation resulting it not being valuable to analysis for pattern recognition or prediction.

**Descriptive analysis**

National

|  |  |
| --- | --- |
| Chart, bar chart  Description automatically generated | Chart, bar chart  Description automatically generated |
| Chart, bar chart  Description automatically generated | Chart, bar chart  Description automatically generated |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Number of Offences Against the Person Convictions

Chart, histogram, scatter chart

Description automatically generated

Figure 1.1

The data analysed with box plots to identify any outliers. The national data was identified as an outlier and separated from the dataset as described in **section 1.1(Data cleaning).** Then the data visualized with a box plot and *Metropolitan and City, West Midlands and GreaterManchester* was outliers for most of the *number of principal offences variables*. However, those three counties have three highest population numbers in the UK. Therefore, it is justifiable that the counties have higher number of court cases.

Chart, scatter chart

Description automatically generated

Then the percentage of the successful and unsuccessful cases were analysed with bar graphs.

A picture containing diagram

Description automatically generated

Percentage of Homicide Convictions

A picture containing table

Description automatically generated

Chart, histogram

Description automatically generated

Chart, line chart, histogram

Description automatically generated