

mODULE 1 r PRACTICE

ALY 6010- WEEK 1



jESEEKA sHAH

NUID-002134289

**Introduction**

The analysis that I have carried out is based on the field interrogation and observation that got recorded while field interaction of the Boston police department and individuals. I have taken the data from the Boston Police Department for the year 2020 and this data set was updated last in April 2021. This is to dataset that shows us the interrogation between the Boston police department and the private individuals that were caught based on different circumstances in state of Massachusetts. This will help us to check which place near, in and around Boston has encounter different circumstances for interrogation and which state has the most frequent field interactions. The dataset consists of the 5717 records and 22 features. Following are the features that are mainly considered for circumstances considered in this analysis.

|  |  |
| --- | --- |
| **Features** | **Values** |
| Circumstance | Encountered, Stopped and Observed |
| Basis | Encounter, Reasonable Suspicion, Intel, Probable Cause |
| Key Situations | Homeless, Drug, Juvenile, Shorts Fired, Arrest Warrant, Search Warrant, Victim shot, Victim stabbed |
| Vehicle type | Passenger car/truck, SUV (sport utility vehicle), Passenger Van, Other Vehicle, Pickup truck |
| Vehicle state | MA |

**Analysis performed on the data**

* After reading the dataset and storing the values in a variable using read.csv. The dataset gets cleaned, spaces get removed and Na present in the dataset. Figure 2 shows the summary of the features that are used from the dataset.

A screenshot of a computer

Description automatically generated with low confidence

***Figure1: showing the removal of the NA***  A picture containing diagram

Description automatically generated

***Figure 2: the summary of the columns***

* Using head and tail function to show the first and last few rows that are present in the dataset or data frame. Syntax-***head*** *(data frame name, number of rows) ,* ***tail****(data frame name, number of rows).* Graphical user interface, text, application, email

  Description automatically generated

***Figure 3: showing the top five rows of that dataframe***

* Using ***class(*)** to determine the type of that object.
* After making a clear data frame without any missing values. The next step should be starting with proper exploratory data analysis.
* **Filter**() used to add condition as ***state***==”MA” and **stop duration** is than or equal to 10.
* **Group**() used for grouping the elements based on certain parameter that will retrieve the data as desired for analysis.
* **Arrange()**  used to give the order to the data that was retrieved.

**Case 1- filtering based on the state =”MA” and then grouping based on Basis**

Chart, bar chart

Description automatically generated

***Figure 4: Bar plot showing the interrogation by BPD based on Basis***

The bar plot here indicate that based on Probable Cause and Reasonable Suspicion the interrogation done by Boston police department are more frequent as compared to direct encounter or intel.

**Case 2- Filtering based on Intel and cities**

Chart, pie chart

Description automatically generated

***Fig 5: interaction between BPD and private individuals based on the intel and cities***

The pie chart shows that interrogation conducted by the Boston police department in the state on “MA” based on intel was more in **Roxbury** followed by Boston and Dorchester.

**Case 3- Filtering based on Probable cause and cities**

Chart, bar chart

Description automatically generated

***Figure 6: Plot showing interrogation bases on the probable cause***

The figure 6, shows that based on the probable cause the interaction of Boston police department is more in the cities of **Dorchester** and **Roxbury** compared to South Boston and Hyde Park.

**Case 4- Interrogation over years**

Chart, line chart

Description automatically generated

***Figure 7: showing the occurrence based on the field interrogation***

The above graph shows that the interrogation done by the Boston police has been increasing from the year 2000 and has reached the peak in the year 2020 based on the data considered.

**Conclusion**

The analysis performed on the dataset pointed out that based on Probable Cause and Reasonable Suspicion the interrogation is conducted more in the state of “**MA”** and has increased over the time from **2000** to **2020**. After filtering the dataset, based on the **intel** the interrogation was more in the city of **Roxbury**, based on probable cause the city **Dorchester and Roxbury** was interrogated more.

**References**

* GeeksforGeeks. (2021, June 25). *Filter DataFrame columns in R by given condition*. <https://www.geeksforgeeks.org/filter-dataframe-columns-in-r-by-given-condition/>
* *R Error in plot.window(. . .) : need finite “xlim” values (Example) | How to Fix*. (2021, June 30). Statistics Globe. <https://statisticsglobe.com/plot-error-need-finite-xlim-values-in-r>
* *BPD Field Interrogation and Observation (FIO) - Analyze Boston*. (n.d.). Boston Police Department. <https://data.boston.gov/dataset/boston-police-department-fio>