**Customer Service Requests Analysis Project**

**Summary Report**

Problem Statement:

Exploratory Data Analysis:

Data analysis is required to be performed of service request (311) calls from New York City. The data wrangling techniques to be utilized to understand the pattern in the data and visualize the major types of complaints.

**Filename:** 311\_Service\_Requests\_from\_2010\_to\_Present.csv file

**Solution:**

1. **Understand the dataset:**
   1. Identify the shape of the dataset

The dataset had 364558 records and 53 features

* 1. Identify variables with null values

There were NULL values is many features involved.

1. **Perform basic data exploratory analysis:**
   1. Basic EDA

Features such as “Agency”, “Complaint Type”, “Descriptor” was carried out.

* “Agency” was NYPD for all the records
* Major complaint types were:
  + Blocked Driveway 100881
  + Illegal Parking 92679
  + Noise - Street/Sidewalk 51692
  + Noise - Commercial 44109
* Major description of the complaint were:
  + No Access 75888
  + Loud Music/Party 69708
  + Posted Parking Sign Violation 27200
* Different type of description was analysed for various types of complaints.
  1. Utilize missing value treatment.

In case of column ‘Facility Type’, there were 2389 missing values. Since the facility type is 'Precinct' for all the remaining cases, the missing values are utilised by updating the missing values by 'Precinct'.

* 1. Analyze the date column and remove the entries if it has an incorrect timeline

since the dates are within one year period, there are no outliers observed in the dataset. The same was also confirmed by checking the response time. The outliers were also checked using the boxplot graph. Since there are not many outliers in response time, it can be concluded that there are no outliers in the dates.

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

* 1. Draw a frequency plot for city-wise complaints

Graphical user interface, application

Description automatically generated

The top 3 cities having complaints was Brooklyn, New York and Bronx.

* 1. Draw scatter and hexbin plots for complaint concentration across Brooklyn

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generated

1. **Find major types of complaints:**
2. Plot a bar graph of count vs. complaint types

Graphical user interface, application, Word

Description automatically generated

1. Find the top 10 types of complaints

Graphical user interface, text, application

Description automatically generated

1. Display the types of complaints in each city in a separate dataset
2. **Visualize the major types of complaints in each city**
3. **Check if the average response time across various types of complaints is different**

The response time (in hours) is different for various complaint types. Since there are more than 3 types of complaints, I have considered ANOVA test to validate the average response time for top 5 complaint types.

Since some complaints have not closed, the response time in hours is blank for those cases. The same has been removed using dropna() method.

Under the ANOVA test:

H0: the average response time is the same across different complaint types

Ha: the average response time is different

Since the P value is less than Alpha (p < alpha), we shall reject the H0. i.e. the average response time across types of complaints is different.

1. **Identify significant variables by performing a statistical analysis using p-values and chi-square values (Optional)**

The Chi-square test is carried out to check the relationship between two variables. The relationship between 'city' and 'complaint types' was compared. Considering that there are many cities, the relationship was compared between top 5 cities w.r.t. number of complaints.

To validate using chi-square test, the relationship between cities and complaint types

# h0: there is no relationship between city and complaint type

# ha: there is a relationship between city and complaint type

Since the P-value 0.0 is lower than Alpha 0.05, we reject the null hypothesis H0. Therefore, there is a relationship between city and complaint types.