Information Visualisation Project (Assignment-1)

Akash Prasad RA2011003010336

February 6, 2023

1 Introduction

New York City has made data on service request of the public citizens and wants to get it analyzed and help to answer a few questions that will help the citizens as well as the different department of the city to improve response time and efficiency.

Using data wrangling techniques get the information from the dataset and try to find different hidden patterns and come to conclusions which will be presented by using data visualization techniques.

2 Data Schema:

2.1 Data Source

Download the 311-service-requests-nyc.zip file using the link given in the Customer Service Requests Analysis project problem statement and extract the $311_Service_Requests_from_2010_to_Present.csv file$

This dataset contains the data from 2010 to present all the complaints and service requests by the citizens.Download the dataset and extract the csv file from the zip file and then apply the data wrangling techniques to make the dataset in working data format.

2.2 Dataset Description

Dataset Description :

Unique Key (Plain text) - Unique identifier for the complaints
Created Date (Date and Time) - The date and time on which the complaint is raised
Closed Date (Date and Time) - The date and time on which the complaint is closed
Agency (Plain text) - Agency code
Agency Name (Plain text) - Name of the agency
Complaint Type (Plain text) - Type of the complaint
Descriptor (Plain text) - Complaint type label (Heating - Heat, Traffic Signal Condition - Controller)
Location Type (Plain text) - Type of the location (Residential, Restaurant, Bakery, etc)
Incident Zip (Plain text) - Zip code for the location
Incident Address (Plain text) - Address of the location
Street Name (Plain text) - Name of the street
Cross Street 1 (Plain text) - Detail of cross street
Cross Street 2 (Plain text) - Detail of another cross street
Intersection Street 1 (Plain text) - Detail of intersection street if any
Intersection Street 2 (Plain text) - Detail of another intersection street if any
Address Type (Plain text) - Categorical (Address or Intersection)
City (Plain text) - City for the location
Landmark (Plain text) - Empty field
Facility Type (Plain text) - N/A
Status (Plain text) - Categorical (Closed or Pending)
Due Date (Date and Time) - Date and time for the pending complaints
Resolution Action Updated Date (Date and Time) - Date and time when the resolution was provided
Community Board (Plain text) - Categorical field (specifies the community board with its code)

Borough (Plain text) - Categorical field (specifies the community board)
X Coordinate (State Plane) (Number)
Y Coordinate (State Plane) (Number)
Park Facility Name (Plain text) - Unspecified
Park Borough (Plain text) - Categorical (Unspecified, Queens, Brooklyn etc)
School Name (Plain text) - Unspecified
School Number (Plain text) - Unspecified
School Region (Plain text) - Unspecified
School Code (Plain text) - Unspecified
School Phone Number (Plain text) - Unspecified
School Address (Plain text) - Unspecified
School City (Plain text) - Unspecified
School State (Plain text) - Unspecified
School Zip (Plain text) - Unspecified
School Not Found (Plain text) - Empty Field
School or Citywide Complaint (Plain text) - Empty Field
Vehicle Type (Plain text) - Empty Field
Taxi Company Borough (Plain text) - Empty Field
Taxi Pick Up Location (Plain text) - Empty Field
Bridge Highway Name (Plain text) - Empty Field
Bridge Highway Direction (Plain text) - Empty Field
Road Ramp (Plain text) - Empty Field
Bridge Highway Segment (Plain text) - Empty Field
Garage Lot Name (Plain text) - Empty Field
Ferry Direction (Plain text) - Empty Field
Ferry Terminal Name (Plain text) - Empty Field
Latitude (Number) - Latitude of the location
Longitude (Number) - Longitude of the location
Location (Location) - Coordinates (Latitude, Longitude)

3 Tasks to perform

There are a few stages of tasks that is to be performed :

3.1 Data Overview

Take an overview of the dataset by performing the following steps:

1. Import the dataset
2. Visualise the dataset and explore the columns
3. Shape of the dataset
4. Data types of the columns

3.2 EDA (Exploratory Data Analysis)

EDA or Exploratory Data Analysis is very important to be done on the dataset before any further steps.
Tasks performed in EDA:
1. Frequency Plot of Null values
2. Identify the Null values
3. Null values treatment
4. Outliers Identification
5. Outlier Treatment

3.3 Analysis of all columns

All columns are to be analyzed for the null values treatment, outliers, data types, and normal distribution of the values. Categorical Columns are analyzed and visualized for class imbalances and treated accordingly.
After all the analysis of the numerical and categorical columns, the date-Time column is to analyze with the following tasks:

1. Calculate the Time elapsed between the closed date and the creation date
2. Convert the calculated time elapsed into seconds or minutes for better representation
3. View the descriptive statistics of the time elapsed newly made column

3.4 Questions That are to be answered

Questions that are to be answered are as follows:

1. Find Major Types of Complaints 2. Check if the average response time across various types of complaints varies or not
3. Identify significant variables affecting response time
4. Analysis of response time across all cities with same type of complaints