NYC311 Customer Service Analysis R

Malay Raj

2023-06-12

## NYC 311 CUSTOMER SERVICE REPORT ANALYSIS.

Problem Objective :

Perform a service request data analysis of New York City 311 calls. You will focus on the data wrangling techniques to understand the pattern in the data and also visualize the major complaint types.

For this analysis I haven’t included the null values.

Domain: Customer Service

## Loading Required Packages.

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.2.3

## Warning: package 'ggplot2' was built under R version 4.2.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.1.8  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

library(janitor)

## Warning: package 'janitor' was built under R version 4.2.3

##   
## Attaching package: 'janitor'  
##   
## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

library(dplyr)  
library(ggplot2)  
library(skimr)

## Warning: package 'skimr' was built under R version 4.2.3

library(plyr)

## ------------------------------------------------------------------------------  
## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)  
## ------------------------------------------------------------------------------  
##   
## Attaching package: 'plyr'  
##   
## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize  
##   
## The following object is masked from 'package:purrr':  
##   
## compact

library(lubridate)  
library(scales)

##   
## Attaching package: 'scales'  
##   
## The following object is masked from 'package:purrr':  
##   
## discard  
##   
## The following object is masked from 'package:readr':  
##   
## col\_factor

## Displaying my current working directory.

getwd()

## [1] "C:/Users/KIIT/OneDrive/Desktop/Bellabeat/NYC Customer Service Requests Analysis"

## Collecting the data.

NYC = read.csv("C:/Users/KIIT/OneDrive/Desktop/Bellabeat/NYC Customer Service Requests Analysis/NYC311data.csv")

## Displaying the first six rows of the dataset.

head(NYC)

## Unique.Key Created.Date Closed.Date Agency  
## 1 32310363 12/31/2015 11:59:45 PM 01-01-16 0:55 NYPD  
## 2 32309934 12/31/2015 11:59:44 PM 01-01-16 1:26 NYPD  
## 3 32309159 12/31/2015 11:59:29 PM 01-01-16 4:51 NYPD  
## 4 32305098 12/31/2015 11:57:46 PM 01-01-16 7:43 NYPD  
## 5 32306529 12/31/2015 11:56:58 PM 01-01-16 3:24 NYPD  
## 6 32306554 12/31/2015 11:56:30 PM 01-01-16 1:50 NYPD  
## Agency.Name Complaint.Type  
## 1 New York City Police Department Noise - Street/Sidewalk  
## 2 New York City Police Department Blocked Driveway  
## 3 New York City Police Department Blocked Driveway  
## 4 New York City Police Department Illegal Parking  
## 5 New York City Police Department Illegal Parking  
## 6 New York City Police Department Illegal Parking  
## Descriptor Location.Type Incident.Zip  
## 1 Loud Music/Party Street/Sidewalk 10034  
## 2 No Access Street/Sidewalk 11105  
## 3 No Access Street/Sidewalk 10458  
## 4 Commercial Overnight Parking Street/Sidewalk 10461  
## 5 Blocked Sidewalk Street/Sidewalk 11373  
## 6 Posted Parking Sign Violation Street/Sidewalk 11215  
## Incident.Address Street.Name Cross.Street.1 Cross.Street.2  
## 1 71 VERMILYEA AVENUE VERMILYEA AVENUE ACADEMY STREET WEST 204 STREET  
## 2 27-07 23 AVENUE 23 AVENUE 27 STREET 28 STREET  
## 3 2897 VALENTINE AVENUE VALENTINE AVENUE EAST 198 STREET EAST 199 STREET  
## 4 2940 BAISLEY AVENUE BAISLEY AVENUE EDISON AVENUE B STREET  
## 5 87-14 57 ROAD 57 ROAD SEABURY STREET HOFFMAN DRIVE  
## 6 260 21 STREET 21 STREET 5 AVENUE 6 AVENUE  
## Intersection.Street.1 Intersection.Street.2 Address.Type City Landmark  
## 1 ADDRESS NEW YORK   
## 2 ADDRESS ASTORIA   
## 3 ADDRESS BRONX   
## 4 ADDRESS BRONX   
## 5 ADDRESS ELMHURST   
## 6 ADDRESS BROOKLYN   
## Facility.Type Status Due.Date  
## 1 Precinct Closed 01-01-16 7:59  
## 2 Precinct Closed 01-01-16 7:59  
## 3 Precinct Closed 01-01-16 7:59  
## 4 Precinct Closed 01-01-16 7:57  
## 5 Precinct Closed 01-01-16 7:56  
## 6 Precinct Closed 01-01-16 7:56  
## Resolution.Description  
## 1 The Police Department responded and upon arrival those responsible for the condition were gone.  
## 2 The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  
## 3 The Police Department responded and upon arrival those responsible for the condition were gone.  
## 4 The Police Department responded to the complaint and took action to fix the condition.  
## 5 The Police Department responded and upon arrival those responsible for the condition were gone.  
## 6 The Police Department responded and upon arrival those responsible for the condition were gone.  
## Resolution.Action.Updated.Date Community.Board Borough  
## 1 01-01-16 0:55 12 MANHATTAN MANHATTAN  
## 2 01-01-16 1:26 01 QUEENS QUEENS  
## 3 01-01-16 4:51 07 BRONX BRONX  
## 4 01-01-16 7:43 10 BRONX BRONX  
## 5 01-01-16 3:24 04 QUEENS QUEENS  
## 6 01-01-16 1:50 07 BROOKLYN BROOKLYN  
## X.Coordinate..State.Plane. Y.Coordinate..State.Plane. Park.Facility.Name  
## 1 1005409 254678 Unspecified  
## 2 1007766 221986 Unspecified  
## 3 1015081 256380 Unspecified  
## 4 1031740 243899 Unspecified  
## 5 1019123 206375 Unspecified  
## 6 986312 180032 Unspecified  
## Park.Borough School.Name School.Number School.Region School.Code  
## 1 MANHATTAN Unspecified Unspecified Unspecified Unspecified  
## 2 QUEENS Unspecified Unspecified Unspecified Unspecified  
## 3 BRONX Unspecified Unspecified Unspecified Unspecified  
## 4 BRONX Unspecified Unspecified Unspecified Unspecified  
## 5 QUEENS Unspecified Unspecified Unspecified Unspecified  
## 6 BROOKLYN Unspecified Unspecified Unspecified Unspecified  
## School.Phone.Number School.Address School.City School.State School.Zip  
## 1 Unspecified Unspecified Unspecified Unspecified Unspecified  
## 2 Unspecified Unspecified Unspecified Unspecified Unspecified  
## 3 Unspecified Unspecified Unspecified Unspecified Unspecified  
## 4 Unspecified Unspecified Unspecified Unspecified Unspecified  
## 5 Unspecified Unspecified Unspecified Unspecified Unspecified  
## 6 Unspecified Unspecified Unspecified Unspecified Unspecified  
## School.Not.Found School.or.Citywide.Complaint Vehicle.Type  
## 1 N NA NA  
## 2 N NA NA  
## 3 N NA NA  
## 4 N NA NA  
## 5 N NA NA  
## 6 N NA NA  
## Taxi.Company.Borough Taxi.Pick.Up.Location Bridge.Highway.Name  
## 1 NA NA   
## 2 NA NA   
## 3 NA NA   
## 4 NA NA   
## 5 NA NA   
## 6 NA NA   
## Bridge.Highway.Direction Road.Ramp Bridge.Highway.Segment Garage.Lot.Name  
## 1 NA  
## 2 NA  
## 3 NA  
## 4 NA  
## 5 NA  
## 6 NA  
## Ferry.Direction Ferry.Terminal.Name Latitude Longitude  
## 1 40.86568 -73.92350  
## 2 40.77595 -73.91509  
## 3 40.87032 -73.88852  
## 4 40.83599 -73.82838  
## 5 40.73306 -73.87417  
## 6 40.66082 -73.99257  
## Location  
## 1 (40.86568153633767, -73.92350095571744)  
## 2 (40.775945312321085, -73.91509393898605)  
## 3 (40.870324522111424, -73.88852464418646)  
## 4 (40.83599404683083, -73.82837939584206)  
## 5 (40.733059618956815, -73.87416975810375)  
## 6 (40.66082272389114, -73.99256786342693)

## Checking the number of rows and dimension of the dataset.

nrow(NYC)

## [1] 300698

dim(NYC)

## [1] 300698 53

## Renaming few columns which we further need in analysis.

NYC=plyr::rename(NYC, replace=c("Unique.Key" = "Unique\_key",  
 "Created.Date" = "Created\_date",  
 "Closed.Date" = "Closed\_date",  
 "Location.Type" = "Location\_type",  
 "Incident.Zip" = "Incident\_zip",  
 "Complaint.Type" = "Complaint\_type",  
 "Resolution.Action.Updated.Date" = "Updated\_closed\_date",  
 "Incident.Address" = "Incident\_address",  
 "Address.Type" = "Address\_type",   
 "Due.Date" = "Due\_date",  
 "Community.Board" = "Community\_board",  
 "X.Coordinate..State.Plane." = "X\_coordinate",   
 "Y.Coordinate..State.Plane." = "Y\_coordinate"))

## Viewing the new column names.

colnames(NYC)

## [1] "Unique\_key" "Created\_date"   
## [3] "Closed\_date" "Agency"   
## [5] "Agency.Name" "Complaint\_type"   
## [7] "Descriptor" "Location\_type"   
## [9] "Incident\_zip" "Incident\_address"   
## [11] "Street.Name" "Cross.Street.1"   
## [13] "Cross.Street.2" "Intersection.Street.1"   
## [15] "Intersection.Street.2" "Address\_type"   
## [17] "City" "Landmark"   
## [19] "Facility.Type" "Status"   
## [21] "Due\_date" "Resolution.Description"   
## [23] "Updated\_closed\_date" "Community\_board"   
## [25] "Borough" "X\_coordinate"   
## [27] "Y\_coordinate" "Park.Facility.Name"   
## [29] "Park.Borough" "School.Name"   
## [31] "School.Number" "School.Region"   
## [33] "School.Code" "School.Phone.Number"   
## [35] "School.Address" "School.City"   
## [37] "School.State" "School.Zip"   
## [39] "School.Not.Found" "School.or.Citywide.Complaint"  
## [41] "Vehicle.Type" "Taxi.Company.Borough"   
## [43] "Taxi.Pick.Up.Location" "Bridge.Highway.Name"   
## [45] "Bridge.Highway.Direction" "Road.Ramp"   
## [47] "Bridge.Highway.Segment" "Garage.Lot.Name"   
## [49] "Ferry.Direction" "Ferry.Terminal.Name"   
## [51] "Latitude" "Longitude"   
## [53] "Location"

## Inspecting the dataframe and look for inconguencies.

str(NYC)

## 'data.frame': 300698 obs. of 53 variables:  
## $ Unique\_key : int 32310363 32309934 32309159 32305098 32306529 32306554 32306559 32307009 32308581 32308391 ...  
## $ Created\_date : chr "12/31/2015 11:59:45 PM" "12/31/2015 11:59:44 PM" "12/31/2015 11:59:29 PM" "12/31/2015 11:57:46 PM" ...  
## $ Closed\_date : chr "01-01-16 0:55" "01-01-16 1:26" "01-01-16 4:51" "01-01-16 7:43" ...  
## $ Agency : chr "NYPD" "NYPD" "NYPD" "NYPD" ...  
## $ Agency.Name : chr "New York City Police Department" "New York City Police Department" "New York City Police Department" "New York City Police Department" ...  
## $ Complaint\_type : chr "Noise - Street/Sidewalk" "Blocked Driveway" "Blocked Driveway" "Illegal Parking" ...  
## $ Descriptor : chr "Loud Music/Party" "No Access" "No Access" "Commercial Overnight Parking" ...  
## $ Location\_type : chr "Street/Sidewalk" "Street/Sidewalk" "Street/Sidewalk" "Street/Sidewalk" ...  
## $ Incident\_zip : chr "10034" "11105" "10458" "10461" ...  
## $ Incident\_address : chr "71 VERMILYEA AVENUE" "27-07 23 AVENUE" "2897 VALENTINE AVENUE" "2940 BAISLEY AVENUE" ...  
## $ Street.Name : chr "VERMILYEA AVENUE" "23 AVENUE" "VALENTINE AVENUE" "BAISLEY AVENUE" ...  
## $ Cross.Street.1 : chr "ACADEMY STREET" "27 STREET" "EAST 198 STREET" "EDISON AVENUE" ...  
## $ Cross.Street.2 : chr "WEST 204 STREET" "28 STREET" "EAST 199 STREET" "B STREET" ...  
## $ Intersection.Street.1 : chr "" "" "" "" ...  
## $ Intersection.Street.2 : chr "" "" "" "" ...  
## $ Address\_type : chr "ADDRESS" "ADDRESS" "ADDRESS" "ADDRESS" ...  
## $ City : chr "NEW YORK" "ASTORIA" "BRONX" "BRONX" ...  
## $ Landmark : chr "" "" "" "" ...  
## $ Facility.Type : chr "Precinct" "Precinct" "Precinct" "Precinct" ...  
## $ Status : chr "Closed" "Closed" "Closed" "Closed" ...  
## $ Due\_date : chr "01-01-16 7:59" "01-01-16 7:59" "01-01-16 7:59" "01-01-16 7:57" ...  
## $ Resolution.Description : chr "The Police Department responded and upon arrival those responsible for the condition were gone." "The Police Department responded to the complaint and with the information available observed no evidence of the"| \_\_truncated\_\_ "The Police Department responded and upon arrival those responsible for the condition were gone." "The Police Department responded to the complaint and took action to fix the condition." ...  
## $ Updated\_closed\_date : chr "01-01-16 0:55" "01-01-16 1:26" "01-01-16 4:51" "01-01-16 7:43" ...  
## $ Community\_board : chr "12 MANHATTAN" "01 QUEENS" "07 BRONX" "10 BRONX" ...  
## $ Borough : chr "MANHATTAN" "QUEENS" "BRONX" "BRONX" ...  
## $ X\_coordinate : int 1005409 1007766 1015081 1031740 1019123 986312 1001578 1011117 1030662 984378 ...  
## $ Y\_coordinate : int 254678 221986 256380 243899 206375 180032 245627 244417 196163 166541 ...  
## $ Park.Facility.Name : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ Park.Borough : chr "MANHATTAN" "QUEENS" "BRONX" "BRONX" ...  
## $ School.Name : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Number : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Region : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Code : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Phone.Number : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Address : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.City : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.State : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Zip : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Not.Found : chr "N" "N" "N" "N" ...  
## $ School.or.Citywide.Complaint: logi NA NA NA NA NA NA ...  
## $ Vehicle.Type : logi NA NA NA NA NA NA ...  
## $ Taxi.Company.Borough : logi NA NA NA NA NA NA ...  
## $ Taxi.Pick.Up.Location : logi NA NA NA NA NA NA ...  
## $ Bridge.Highway.Name : chr "" "" "" "" ...  
## $ Bridge.Highway.Direction : chr "" "" "" "" ...  
## $ Road.Ramp : chr "" "" "" "" ...  
## $ Bridge.Highway.Segment : chr "" "" "" "" ...  
## $ Garage.Lot.Name : logi NA NA NA NA NA NA ...  
## $ Ferry.Direction : chr "" "" "" "" ...  
## $ Ferry.Terminal.Name : chr "" "" "" "" ...  
## $ Latitude : num 40.9 40.8 40.9 40.8 40.7 ...  
## $ Longitude : num -73.9 -73.9 -73.9 -73.8 -73.9 ...  
## $ Location : chr "(40.86568153633767, -73.92350095571744)" "(40.775945312321085, -73.91509393898605)" "(40.870324522111424, -73.88852464418646)" "(40.83599404683083, -73.82837939584206)" ...

## Changing datatypes of Created\_date, Closed\_date, Updated\_closed\_date and Due\_date. All in one single format.

NYC$Created\_date <- as.POSIXct(NYC$Created\_date, format = "%m/%d/%Y %I:%M:%S %p")  
NYC$Closed\_date <- as.POSIXct(NYC$Closed\_date, format = "%m-%d-%y %H:%M")  
NYC$Updated\_closed\_date <- as.POSIXct(NYC$Updated\_closed\_date, format = "%m-%d-%y %H:%M")  
NYC$Due\_date <- as.POSIXct(NYC$Due\_date, format = "%m-%d-%y %H:%M")

## Checking if they are changed or not.

str(NYC)

## 'data.frame': 300698 obs. of 53 variables:  
## $ Unique\_key : int 32310363 32309934 32309159 32305098 32306529 32306554 32306559 32307009 32308581 32308391 ...  
## $ Created\_date : POSIXct, format: "2015-12-31 23:59:45" "2015-12-31 23:59:44" ...  
## $ Closed\_date : POSIXct, format: "2016-01-01 00:55:00" "2016-01-01 01:26:00" ...  
## $ Agency : chr "NYPD" "NYPD" "NYPD" "NYPD" ...  
## $ Agency.Name : chr "New York City Police Department" "New York City Police Department" "New York City Police Department" "New York City Police Department" ...  
## $ Complaint\_type : chr "Noise - Street/Sidewalk" "Blocked Driveway" "Blocked Driveway" "Illegal Parking" ...  
## $ Descriptor : chr "Loud Music/Party" "No Access" "No Access" "Commercial Overnight Parking" ...  
## $ Location\_type : chr "Street/Sidewalk" "Street/Sidewalk" "Street/Sidewalk" "Street/Sidewalk" ...  
## $ Incident\_zip : chr "10034" "11105" "10458" "10461" ...  
## $ Incident\_address : chr "71 VERMILYEA AVENUE" "27-07 23 AVENUE" "2897 VALENTINE AVENUE" "2940 BAISLEY AVENUE" ...  
## $ Street.Name : chr "VERMILYEA AVENUE" "23 AVENUE" "VALENTINE AVENUE" "BAISLEY AVENUE" ...  
## $ Cross.Street.1 : chr "ACADEMY STREET" "27 STREET" "EAST 198 STREET" "EDISON AVENUE" ...  
## $ Cross.Street.2 : chr "WEST 204 STREET" "28 STREET" "EAST 199 STREET" "B STREET" ...  
## $ Intersection.Street.1 : chr "" "" "" "" ...  
## $ Intersection.Street.2 : chr "" "" "" "" ...  
## $ Address\_type : chr "ADDRESS" "ADDRESS" "ADDRESS" "ADDRESS" ...  
## $ City : chr "NEW YORK" "ASTORIA" "BRONX" "BRONX" ...  
## $ Landmark : chr "" "" "" "" ...  
## $ Facility.Type : chr "Precinct" "Precinct" "Precinct" "Precinct" ...  
## $ Status : chr "Closed" "Closed" "Closed" "Closed" ...  
## $ Due\_date : POSIXct, format: "2016-01-01 07:59:00" "2016-01-01 07:59:00" ...  
## $ Resolution.Description : chr "The Police Department responded and upon arrival those responsible for the condition were gone." "The Police Department responded to the complaint and with the information available observed no evidence of the"| \_\_truncated\_\_ "The Police Department responded and upon arrival those responsible for the condition were gone." "The Police Department responded to the complaint and took action to fix the condition." ...  
## $ Updated\_closed\_date : POSIXct, format: "2016-01-01 00:55:00" "2016-01-01 01:26:00" ...  
## $ Community\_board : chr "12 MANHATTAN" "01 QUEENS" "07 BRONX" "10 BRONX" ...  
## $ Borough : chr "MANHATTAN" "QUEENS" "BRONX" "BRONX" ...  
## $ X\_coordinate : int 1005409 1007766 1015081 1031740 1019123 986312 1001578 1011117 1030662 984378 ...  
## $ Y\_coordinate : int 254678 221986 256380 243899 206375 180032 245627 244417 196163 166541 ...  
## $ Park.Facility.Name : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ Park.Borough : chr "MANHATTAN" "QUEENS" "BRONX" "BRONX" ...  
## $ School.Name : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Number : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Region : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Code : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Phone.Number : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Address : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.City : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.State : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Zip : chr "Unspecified" "Unspecified" "Unspecified" "Unspecified" ...  
## $ School.Not.Found : chr "N" "N" "N" "N" ...  
## $ School.or.Citywide.Complaint: logi NA NA NA NA NA NA ...  
## $ Vehicle.Type : logi NA NA NA NA NA NA ...  
## $ Taxi.Company.Borough : logi NA NA NA NA NA NA ...  
## $ Taxi.Pick.Up.Location : logi NA NA NA NA NA NA ...  
## $ Bridge.Highway.Name : chr "" "" "" "" ...  
## $ Bridge.Highway.Direction : chr "" "" "" "" ...  
## $ Road.Ramp : chr "" "" "" "" ...  
## $ Bridge.Highway.Segment : chr "" "" "" "" ...  
## $ Garage.Lot.Name : logi NA NA NA NA NA NA ...  
## $ Ferry.Direction : chr "" "" "" "" ...  
## $ Ferry.Terminal.Name : chr "" "" "" "" ...  
## $ Latitude : num 40.9 40.8 40.9 40.8 40.7 ...  
## $ Longitude : num -73.9 -73.9 -73.9 -73.8 -73.9 ...  
## $ Location : chr "(40.86568153633767, -73.92350095571744)" "(40.775945312321085, -73.91509393898605)" "(40.870324522111424, -73.88852464418646)" "(40.83599404683083, -73.82837939584206)" ...

## Dropping few columns because we don’t need them in our further analysis.

NYC = subset(NYC, select = -c(Agency, Closed\_date, Agency.Name, Street.Name, Cross.Street.1, Cross.Street.2, Intersection.Street.1, Intersection.Street.2, Landmark, Facility.Type, Status, Resolution.Description, Park.Borough, Park.Facility.Name, School.Name, Incident\_zip, Incident\_address, School.Number, School.State, School.Region, School.Code, School.Phone.Number, School.Address, School.City, School.Zip, School.Not.Found, School.or.Citywide.Complaint, Vehicle.Type, Taxi.Company.Borough, Taxi.Pick.Up.Location, Bridge.Highway.Name, Bridge.Highway.Direction, Road.Ramp, Bridge.Highway.Segment, Garage.Lot.Name, Ferry.Direction, Ferry.Terminal.Name, Location))

## Checking the column names which are left for analysis.

colnames(NYC)

## [1] "Unique\_key" "Created\_date" "Complaint\_type"   
## [4] "Descriptor" "Location\_type" "Address\_type"   
## [7] "City" "Due\_date" "Updated\_closed\_date"  
## [10] "Community\_board" "Borough" "X\_coordinate"   
## [13] "Y\_coordinate" "Latitude" "Longitude"

## Creating a new column request\_closing\_time with the difference of Updated\_closed\_date and Created\_date.

NYC$request\_closing\_time <- difftime(NYC$Updated\_closed\_date,NYC$Created\_date)

## Changing the datatype of the new column ‘request\_closing\_time’ to numeric.

NYC$request\_closing\_time <- as.numeric(NYC$request\_closing\_time)

## Checking the final dimensions which we need in analysis.

dim(NYC)

## [1] 300698 16

## Checking for duplicate values in our dataframe.

anyDuplicated(NYC)

## [1] 0

## Checking for null values in our dataset.

sum(is.na(NYC))

## [1] 798328

## Dropping all the null values.

NYC <- NYC[complete.cases(NYC), ]

## Checking the current dimensions after dropping all the null values.

dim(NYC)

## [1] 1842 16

## Seeing how many observations fall under each of these columns.

table(NYC$Complaint\_type)

##   
## Animal Abuse Bike/Roller/Skate Chronic Blocked Driveway   
## 36 3 523   
## Derelict Vehicle Disorderly Youth Drinking   
## 80 5 6   
## Graffiti Homeless Encampment Illegal Fireworks   
## 2 14 2   
## Illegal Parking Noise - Commercial Noise - House of Worship   
## 445 272 3   
## Noise - Park Noise - Street/Sidewalk Noise - Vehicle   
## 15 303 91   
## Posting Advertisement Traffic Urinating in Public   
## 15 15 3   
## Vending   
## 9

table(NYC$Descriptor)

##   
## Banging/Pounding Blocked Hydrant   
## 18 110   
## Blocked Sidewalk Car/Truck Horn   
## 51 16   
## Car/Truck Music Chained   
## 59 2   
## Commercial Overnight Parking Congestion/Gridlock   
## 138 8   
## Detached Trailer Double Parked Blocking Traffic   
## 1 22   
## Double Parked Blocking Vehicle Drag Racing   
## 9 3   
## Engine Idling In Car   
## 20 1   
## In Prohibited Area In Public   
## 5 5   
## Loud Music/Party Loud Talking   
## 427 144   
## N/A Neglected   
## 22 11   
## No Access No Shelter   
## 371 3   
## Nuisance/Truant Other (complaint details)   
## 1 13   
## Overnight Commercial Storage Partial Access   
## 16 152   
## Playing in Unsuitable Place Police Report Requested   
## 4 2   
## Posted Parking Sign Violation Tortured   
## 91 6   
## Truck Route Violation Unauthorized Bus Layover   
## 4 7   
## Underage - Licensed Est Unlicensed   
## 1 4   
## Vehicle With License Plate   
## 15 80

table(NYC$Location\_type)

##   
## Club/Bar/Restaurant Commercial   
## 155 1   
## Highway House and Store   
## 1 2   
## House of Worship Park/Playground   
## 3 17   
## Residential Building/House Store/Commercial   
## 33 124   
## Street/Sidewalk Subway Station   
## 1505 1

table(NYC$Address\_type)

##   
## ADDRESS BLOCKFACE INTERSECTION LATLONG   
## 1 1573 34 207 27

table(NYC$Community\_board)

##   
## 01 BRONX 01 BROOKLYN 01 MANHATTAN 01 QUEENS   
## 8 76 14 72   
## 01 STATEN ISLAND 02 BRONX 02 BROOKLYN 02 MANHATTAN   
## 23 7 14 29   
## 02 QUEENS 02 STATEN ISLAND 03 BRONX 03 BROOKLYN   
## 40 22 17 28   
## 03 MANHATTAN 03 QUEENS 03 STATEN ISLAND 04 BRONX   
## 44 30 42 36   
## 04 BROOKLYN 04 MANHATTAN 04 QUEENS 05 BRONX   
## 9 31 51 34   
## 05 BROOKLYN 05 MANHATTAN 05 QUEENS 06 BRONX   
## 23 12 86 20   
## 06 BROOKLYN 06 MANHATTAN 06 QUEENS 07 BRONX   
## 17 35 27 41   
## 07 BROOKLYN 07 MANHATTAN 07 QUEENS 08 BRONX   
## 31 19 30 10   
## 08 BROOKLYN 08 MANHATTAN 08 QUEENS 09 BRONX   
## 26 12 16 29   
## 09 BROOKLYN 09 MANHATTAN 09 QUEENS 10 BRONX   
## 23 18 61 48   
## 10 BROOKLYN 10 MANHATTAN 10 QUEENS 11 BRONX   
## 36 37 28 35   
## 11 BROOKLYN 11 MANHATTAN 11 QUEENS 12 BRONX   
## 27 11 10 35   
## 12 BROOKLYN 12 MANHATTAN 12 QUEENS 13 BROOKLYN   
## 51 84 39 14   
## 13 QUEENS 14 BROOKLYN 14 QUEENS 15 BROOKLYN   
## 45 39 14 30   
## 16 BROOKLYN 17 BROOKLYN 18 BROOKLYN 81 QUEENS   
## 5 13 77 1

table(NYC$Borough)

##   
## BRONX BROOKLYN MANHATTAN QUEENS STATEN ISLAND   
## 320 539 346 550 87

## CONDUCTNG DESCRIPTIVE ANALYSIS:

Descriptive analysis on whole NYC dataframe at once.

summary(NYC)

## Unique\_key Created\_date Complaint\_type   
## Min. :30298181 Min. :2015-03-31 16:21:59.00 Length:1842   
## 1st Qu.:30741718 1st Qu.:2015-05-31 23:44:54.50 Class :character   
## Median :31210056 Median :2015-07-31 23:34:47.50 Mode :character   
## Mean :31296594 Mean :2015-08-14 00:31:56.71   
## 3rd Qu.:31881600 3rd Qu.:2015-10-31 20:11:31.25   
## Max. :32310640 Max. :2015-12-31 23:59:45.00   
## Descriptor Location\_type Address\_type City   
## Length:1842 Length:1842 Length:1842 Length:1842   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Due\_date Updated\_closed\_date   
## Min. :2015-04-01 00:21:00.00 Min. :2015-04-01 00:00:00.00   
## 1st Qu.:2015-06-01 07:44:30.00 1st Qu.:2015-06-01 07:45:45.00   
## Median :2015-08-01 07:34:00.00 Median :2015-08-01 04:54:30.00   
## Mean :2015-08-14 08:31:27.75 Mean :2015-08-14 07:01:19.08   
## 3rd Qu.:2015-11-01 04:10:45.00 3rd Qu.:2015-11-01 00:52:30.00   
## Max. :2016-01-01 07:59:00.00 Max. :2016-01-03 16:22:00.00   
## Community\_board Borough X\_coordinate Y\_coordinate   
## Length:1842 Length:1842 Min. : 917482 Min. :122688   
## Class :character Class :character 1st Qu.: 993216 1st Qu.:181806   
## Mode :character Mode :character Median :1005457 Median :203228   
## Mean :1005545 Mean :204960   
## 3rd Qu.:1020089 3rd Qu.:231581   
## Max. :1065122 Max. :270185   
## Latitude Longitude request\_closing\_time  
## Min. :40.50 Min. :-74.24 Min. : 11.23   
## 1st Qu.:40.67 1st Qu.:-73.97 1st Qu.: 168.15   
## Median :40.72 Median :-73.92 Median : 307.60   
## Mean :40.73 Mean :-73.92 Mean : 389.37   
## 3rd Qu.:40.80 3rd Qu.:-73.87 3rd Qu.: 491.18   
## Max. :40.91 Max. :-73.71 Max. :3939.63

## SUMMARY STATISTICS.

skimr::skim\_without\_charts(NYC)

Data summary

|  |  |
| --- | --- |
| Name | NYC |
| Number of rows | 1842 |
| Number of columns | 16 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 6 |
| POSIXct | 3 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Complaint\_type | 0 | 1 | 7 | 25 | 0 | 19 | 0 |
| Descriptor | 0 | 1 | 3 | 30 | 0 | 36 | 0 |
| Location\_type | 0 | 1 | 7 | 26 | 0 | 10 | 0 |
| Address\_type | 0 | 1 | 0 | 12 | 1 | 5 | 0 |
| City | 0 | 1 | 5 | 19 | 0 | 47 | 0 |
| Community\_board | 0 | 1 | 8 | 16 | 0 | 60 | 0 |
| Borough | 0 | 1 | 5 | 13 | 0 | 5 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Unique\_key | 0 | 1 | 31296594.35 | 594961.54 | 30298181.00 | 30741717.50 | 31210056.00 | 31881599.50 | 32310640.00 |
| X\_coordinate | 0 | 1 | 1005545.41 | 22724.18 | 917482.00 | 993215.50 | 1005456.50 | 1020088.75 | 1065122.00 |
| Y\_coordinate | 0 | 1 | 204959.56 | 31527.91 | 122688.00 | 181805.50 | 203227.50 | 231580.50 | 270185.00 |
| Latitude | 0 | 1 | 40.73 | 0.09 | 40.50 | 40.67 | 40.72 | 40.80 | 40.91 |
| Longitude | 0 | 1 | -73.92 | 0.08 | -74.24 | -73.97 | -73.92 | -73.87 | -73.71 |
| request\_closing\_time | 0 | 1 | 389.37 | 347.87 | 11.23 | 168.15 | 307.60 | 491.17 | 3939.63 |

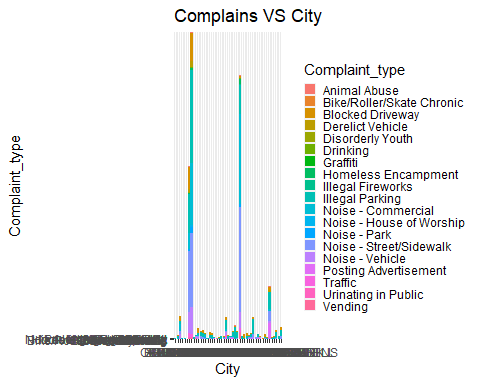
**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| Created\_date | 0 | 1 | 2015-03-31 16:21:59 | 2015-12-31 23:59:45 | 2015-07-31 23:34:47 | 1831 |
| Due\_date | 0 | 1 | 2015-04-01 00:21:00 | 2016-01-01 07:59:00 | 2015-08-01 07:34:00 | 1384 |
| Updated\_closed\_date | 0 | 1 | 2015-04-01 00:00:00 | 2016-01-03 16:22:00 | 2015-08-01 04:54:30 | 1213 |

## DATA VISUALIZATION:

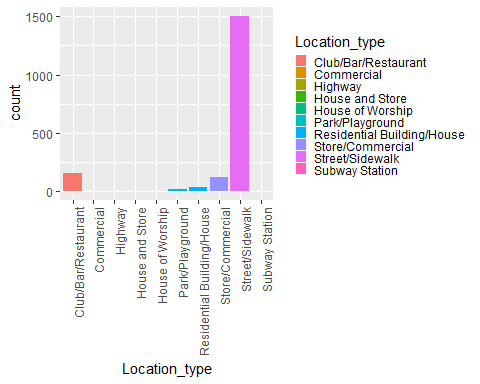
Checking the complaint types of each city.

ggplot(NYC) +   
 geom\_col(mapping = aes(x=City, y=Complaint\_type, fill=Complaint\_type)) +  
 theme(legend.key.size = unit(0.3, "cm")) +  
 labs(title="Complains VS City")



Checking for the location type where most of the complains come.

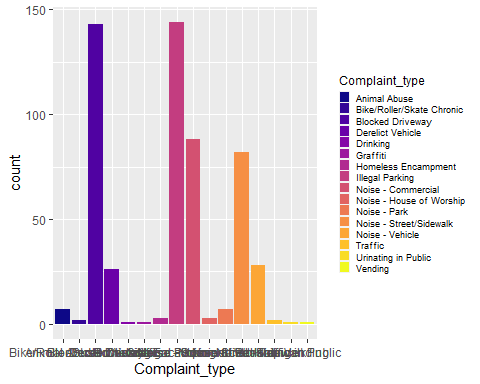
ggplot(NYC, aes(Location\_type, fill = Location\_type)) +  
 geom\_bar() +  
 theme(legend.key.size = unit(0.3, "cm")) +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



## As we saw from the first visualization the most of the complains are from ‘Brooklyn’ city.

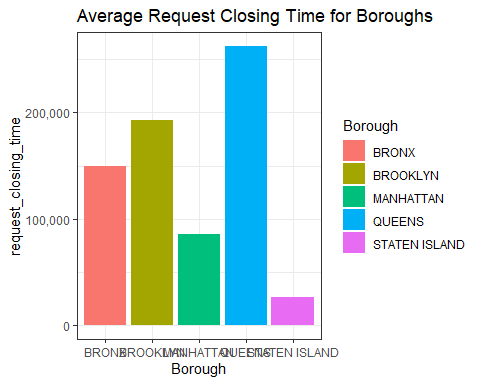
Let’s check which type of complains comes most from that city.

NYC %>%  
 filter(City == 'BROOKLYN') %>%  
 ggplot(aes(x = Complaint\_type, fill = Complaint\_type)) +  
 geom\_bar() +  
 scale\_fill\_viridis\_d(option = "plasma", guide = "legend") +  
 theme(legend.text = element\_text(size = 7),  
 legend.title = element\_text(size = 9),  
 legend.key.size = unit(0.3, "cm"))



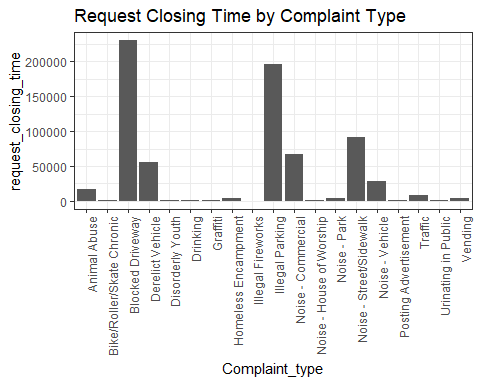
Checking if average of request closing time is related to Boroughs or not.

ggplot(NYC, aes(x = Borough, y = request\_closing\_time, fill = Borough)) +  
 geom\_bar(stat = 'identity') +  
 labs(title = 'Average Request Closing Time for Boroughs') +  
 theme\_bw() +   
 scale\_y\_continuous(labels = scales::comma)



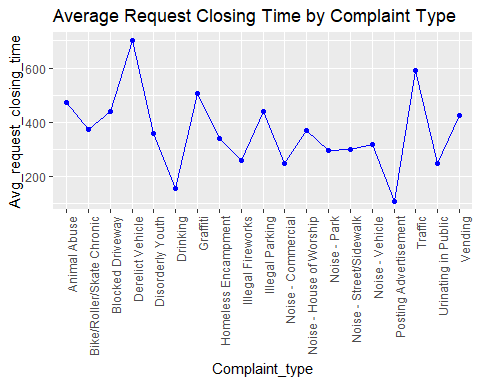
## Checking whether the response time across complaint types is similar or not.

ggplot(NYC, aes(x = Complaint\_type, y = request\_closing\_time)) +  
 geom\_bar(stat = 'identity') +  
 labs(title = 'Request Closing Time by Complaint Type') +  
 theme\_bw() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



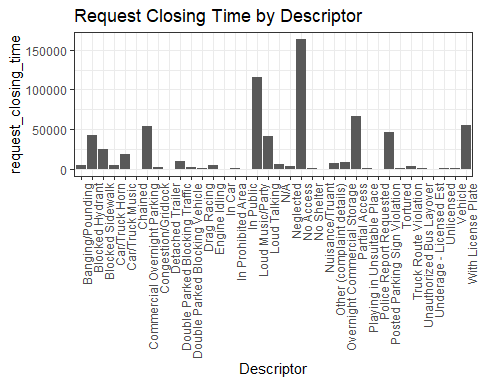
Also checking for the average of response time across complaint types is similar or not.

avg\_request\_closing\_time <- aggregate(request\_closing\_time ~ Complaint\_type, data = NYC, FUN = mean)  
ggplot(avg\_request\_closing\_time, aes(x = Complaint\_type, y = request\_closing\_time)) +  
 geom\_point(color = "blue") +  
 geom\_line(color = "blue", group = 1) +  
 labs(title = 'Average Request Closing Time by Complaint Type', x = 'Complaint\_type', y = 'Avg\_request\_closing\_time') +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



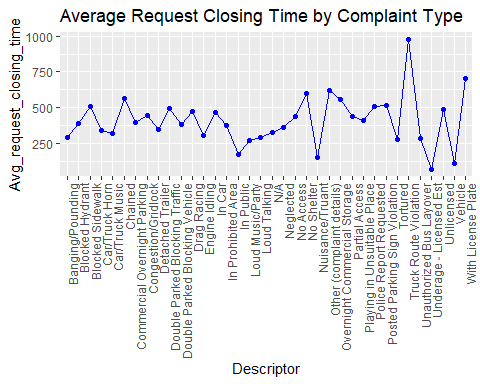
## Checking whether the average response time across descriptor is similar or not.

ggplot(NYC, aes(x = Descriptor, y = request\_closing\_time)) +  
 geom\_bar(stat = 'identity') +   
 labs(title = 'Request Closing Time by Descriptor') +  
 theme\_bw() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



Also checking for the average of response time across descriptor is similar or not.

avg\_request\_closing\_time <- aggregate(request\_closing\_time ~ Descriptor, data = NYC, FUN = mean)  
ggplot(avg\_request\_closing\_time, aes(x = Descriptor, y = request\_closing\_time)) +  
 geom\_point(color = "blue") +  
 geom\_line(color = "blue", group = 1) +  
 labs(title = 'Average Request Closing Time by Complaint Type', x = 'Descriptor', y = 'Avg\_request\_closing\_time') +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



## Checking the relation of average request\_closing\_time with different locations.

avg\_request\_time <- aggregate(request\_closing\_time ~ Location\_type, data = NYC, FUN = mean)  
ggplot(avg\_request\_time, aes(x = Location\_type, y = request\_closing\_time)) +  
 geom\_point(color = "blue") +  
 geom\_line(color = "blue", group = 1) +  
 geom\_text(aes(label = round(request\_closing\_time)), vjust = -0.5, color = "black") +  
 labs(title = "Average Request Closing Time by Location Type",  
 x = "Location\_type", y = "Avg\_request\_closing\_time") +  
 theme\_bw() +  
 theme(plot.title = element\_text(size = 14),  
 axis.text.x = element\_text(angle = 90, hjust = 1))

