**# EXAMINING THE DATA #**

- The data set contains 10803028 rows and 10 columns

# Total number of Parking Tickets in 2017 is 5431918, which is the same as the number of columns in 2017 dataset.

Data in each row corresponds to one Parking Ticket issued.

# 65 Unique states, before cleaning the data. (remove '99')

64 Unique states after replacing 99 with 'NY'

**# AGGREGATION TASKS #**

# The below are the the top 5 Violation Codes that occur frequently (descending order)

21, 36, 38, 14 & 20.

# The below are the the top 5 Vehicle body types that recieve most number of tickets (descending order)

SUBN, 4DSD, VAN, DELV & SDN.

# The below are the the top 5 Vehicle Make that recieve most number of tickets (descending order)

FORD, TOYOT, HONDA, NISSA & CHEVR.

# The below are the the top 5 Violation Precinct where the most no. of violation has occured in 2017 (descending order)

19, 14, 1, 18 & 114.

The area with precinct '19' has the most no. of violations (274445), that contributes more than 5% of the total tickets in 2017 in NYC area.

# The below are the the top 5 Issuer Precinct where the most no. of tickets have been issued in 2017 (descending order)

19, 14, 1, 18 & 114.

The top 5 Violation Precinct and Issuer Precicnt are the same. But frequency does not match, meaning some tickets have been issued by other precincts.

# Violation Code '14' has the highest frequency among the top 3 Issuer Percinct (19, 14 & 1).

This code is commong across the percincts and also the 4th highest in top 5 Violation Codes.

# No null values in "Violation Time" column in the data set for 2017.

# The below are my assumptions of splitting 24hrs into 6 bins (4hrs each):

00:00 - 04:00 = Late Night

04:00 - 08:00 = Early Morning

08:00 - 12:00 = Morning

12:00 - 16:00 = Afternoon

16:00 - 20:00 = Evening

20:00 - 24:00 = Night

# The below are the 3 most commonly occuring Violation codes in the bins as defined above:

00:00 - 04:00 = Late Night - 21, 40, 14.

04:00 - 08:00 = Early Morning - 14, 40, 21.

08:00 - 12:00 = Morning - 21, 36, 38.

12:00 - 16:00 = Afternoon - 38, 36, 37.

16:00 - 20:00 = Evening - 38, 14, 37.

20:00 - 24:00 = Night - 7, 40, 14.

# Three most commonly occurring Violation Codes are 21, 36 & 38. The below are the most common time of the day:

Violation Code 21 - Morning with 598062 violations.

Violation Code 36 - Morning with 348165 violations.

Violation Code 38 - Afternoon with 184829 violations.

# The data set as been categorized into 4 seasons (winter, spring, summer & fall) based on Issue Date, as below:

Month 12, 1 & 2 - Winter

Month 3, 4 & 5 - Spring

Month 6, 7 & 8 - Summer

Month 9, 10 & 11 - Fall

# The below are the 3 most commonly occurring Violation codes in each season (as defined above):

Winter - 21, 36, 38.

Spring - 21, 36, 38.

Summer - 21, 36, 38.

Fall - 46, 21, 40.

# The 3 most commonly occurring Violation codes for each season are the same as the 3 most commonly occurring Violation codes for the entire year

for Winter, Spring & Summer (21, 36, 38). Fall season has very less Violations when compared to others and have 46, 21 & 40 as the most occurring violation codes.

# The total no. of occurrence of Violation Codes 21, 36 & 38 is 1972931, which is more than 36% of the total number of tickets in 2017.

# The below are the Avg. fees for the 3 most common Violation Codes:

Violation Code 21 - $55

Violation Code 36 - $50

Violation Code 38 - $50

# The below are the Total Amount collected for the 3 most common Violation Codes:

Violation Code 21 - $42244785

Violation Code 36 - $36452075

Violation Code 38 - $29814345

# Violation Code 21 has the highest total collection. Most of the collection comes from "No Parking" violations which is Violation Code 21.