NYC Parking Tickets: An Exploratory Analysis

Examining the data

Ques 1. Find the total number of tickets for the year.

Ans: There are 5431918 Tickets Issues in the year 2017

Ques 2. Find out the number of unique states from where the cars that got parking tickets came

Ans: There are 64 Unique states in the Dataset.

Aggregation tasks

Ques. 1 How often does each violation code occur? Display the frequency of the top five violation codes.

```
+--------------+------+
|Violation_Code| count|
+--------------+------+
|            21|768087|
|            36|662765|
|            38|542079|
|            14|476664|
|            20|319646|
+--------------+------+
```

Ques. 2 How often does each 'vehicle body type' get a parking ticket? How about the 'vehicle make'?

```
+----------------+-------+
|Vehicle_BodyType|  count|
+----------------+-------+
|            SUBN|1883954|
|            4DSD|1547312|
|             VAN| 724029|
|            DELV| 358984|
|             SDN| 194197|
+----------------+-------+
```

Ques. 3 Find the (5 highest) frequencies of tickets for each of the following:
--> 'Violation Precinct': Using this, can you draw any insights for parking violations in any specific areas of the city?

```
+--------------+------+
|Violation_Prec| count|
+--------------+------+
|             0|925596|
|            19|274445|
|            14|203553|
|             1|174702|
|            18|169131|
|           114|147444|
+--------------+------+
```

--> 'Issuer Precinct' : Here, you would have noticed that the dataframe has the'Violating Precinct' or 'Issuing Precinct' as '0'. These are erroneous entries. Hence, you need to provide the records for five correct precincts.

```
+---------------+-------+
|Issuer_Precinct|  count|
+---------------+-------+
|              0|1078406|
|             19| 266961|
|             14| 200495|
|              1| 168740|
|             18| 162994|
|            114| 144054|
+---------------+-------+
```

Ques. 4 Find the violation code frequencies for three precincts that have issued the most number of tickets. Do these precinct zones have an exceptionally high frequency of certain violation codes? Are these codes common across precincts?

```
+---------------+--------------+-----+
|Issuer_Precinct|Violation_Code|count|
+---------------+--------------+-----+
|             19|            46|48445|
|             14|            14|45036|
|              1|            14|38354|
|             19|            38|36386|
|             19|            37|36056|
|             14|            69|30464|
|             19|            14|29797|
|             19|            21|28415|
|             14|            31|22555|
|              1|            16|19081|
|             14|            47|18364|
|              1|            20|15408|
|             19|            20|14629|
|              1|            46|12745|
|             19|            40|11416|
|             14|            42|10027|
|             19|            16| 9926|
|              1|            38| 8535|
|             14|            46| 7679|
|              1|            17| 7526|
+---------------+--------------+-----+
only showing top 20 rows
```

Listing out Frequencies of Voilation Codes for the top three Issuer Precinct

we can see there are few Voilation code featuring more often than others,

Yes some precinct zones have an exceptionally high frequency of certain violation codes and they are common as well.

Here are some of the Voilation Codes:

Voilation Code No. 46.

Issuer_Precinct: 1 Count: 12745

Issuer_Precinct: 14 Count: 7679

Issuer_Precinct: 19 Count: 48445

Voilation Code No. 14.

Issuer_Precinct: 1 Count:12745

Issuer_Precinct: 14 Count: 45036

Issuer_Precinct: 19 Count: 48445

Also Some other Voilation Codes with high count,

Voilation Code No. 38.

Issuer_Precinct: 19 Count: 36386

Issuer_Precinct: 1 Count: 8535

Ques. 5 Find out the properties of parking violations across different times of the day:

- *Find a way to deal with missing values, if any. (Hint: Check for the null values using 'isNull' under the SQL. Also, to remove the null values, check the 'dropna' command in the API documentation.)*

Although we have already checked for Null values again checking for Same, and we do not find a Null Values in the DataFrame.

- *The Violation Time field is specified in a strange format. Find a way to make this a time attribute that you can use to divide into groups.*

Handled the time according to different scenarios.

Ques. 7 Find The fines collected from all the instances of parking violation constitute a source of revenue for the NYC Police Department. Let’s take an example of estimating this for the three most commonly occurring codes:

Find the total occurrences of the three most common violation codes.

The top 3 Violation codes are 21, 36 and 38

Then, visit the website:

http://www1.nyc.gov/site/finance/vehicles/services-violation-codes.page

It lists the fines associated with different violation codes. They're divided into two categories: one for the highest-density locations in the city and the other for the rest of the city. For the sake of simplicity, take the average of the two.

Code: 21, Fine: $55

Code: 36, Fine: $50

Code: 38, Fine: $50

Using this information, find the total amount collected for the three violation codes with the maximum tickets. State the code that has the highest total collection.

What can you intuitively infer from these findings?

Violation_Code	count	Total_Penalty
21	768087	42244785
36	662765	33138250
38	542079	27103950