

NYC PARKING TICKETS – ASSIGNMENT

Examine the data:

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

```
+-----+
|Total Count Of Tickets(2017)|
+-----+
|                               5431918|
+-----+
```

2. Find out the number of unique states from where the cars that got parking tickets came. (Hint: Use the column 'Registration State'.)

There is a numeric entry '99' in the column, which should be corrected. Replace it with the state having the maximum entries. Provide the number of unique states again.

```
+-----+
|unique_states_count|
+-----+
|                               64|
+-----+
```

Aggregation tasks:

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

```
+-----+-----+
|Violation Code|Frequency|
+-----+-----+
|           21|    768087|
|           36|    662765|
|           38|    542079|
|           14|    476664|
|           20|    319646|
+-----+-----+
```

2. How often does each 'vehicle body type' get a parking ticket? How about the 'vehicle make'? (Hint: Find the top 5 for both.)

Parking Ticket Frequency per Vehicle Body Type:

```
+-----+-----+
|Vehicle Body Type|Frequency|
+-----+-----+
|           SUBN|    1883954|
|           4DSD|    1547312|
|           VAN|     724029|
|           DELV|    358984|
|           SDN|     194197|
+-----+-----+
```

Parking Ticket Frequency per Vehicle Make:

| Vehicle Make | Frequency |
|--------------|-----------|
| FORD | 636844 |
| TOYOT | 605291 |
| HONDA | 538884 |
| NISSA | 462017 |
| CHEVR | 356032 |

3. A precinct is a police station that has a certain zone of the city under its command. Find the (5 highest) frequencies of tickets for each of the following:

- 'Violation Precinct' (This is the precinct of the zone where the violation occurred). Using this, can you draw any insights for parking violations in any specific areas of the city?

| Violation Precinct | Frequency |
|--------------------|-----------|
| 0 | 925596 |
| 19 | 274445 |
| 14 | 203553 |
| 1 | 174702 |
| 18 | 169131 |
| 114 | 147444 |

- Precinct 19 tops as the highest number of parking tickets for a Violation Precinct.
- Address for Precinct 19 is 153 E 67th St, New York, NY 10065, USA which means there is a higher number of parking violations happening around the 67th Street in NYC i.e. this area has severe problem of parking space.

- 'Issuer Precinct' (This is the precinct that issued the ticket.)

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. (Hint: Print the top six entries after sorting.)

| Issuer Precinct | Frequency |
|-----------------|-----------|
| 0 | 1078406 |
| 19 | 266961 |
| 14 | 200495 |
| 1 | 168740 |
| 18 | 162994 |
| 114 | 144054 |

- Precinct 19 also tops as the highest number of parking tickets for an Issue Precinct

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? (Hint: In the SQL view, use the 'where' attribute to filter among three precincts.)
- Highest number of Violation codes for Top 3 Precincts with highest number of parking tickets:

| Precinct | Violation Code | Frequency of Violation Codes |
|----------|----------------|------------------------------|
| 19 | 46 | 50785 |
| 14 | 14 | 45885 |
| 1 | 14 | 40226 |

Violation Code 14 is common across precinct 1 & 14 with inside each precinct having the highest frequency. Interestingly, Precinct 19 also have high count of violation code 14, which is 30376 and no violation of code 1.

As for Precinct 19, Violation Code 46 has the highest frequency. However, the count of violation code 46 for precinct 1 & 14 compared to precinct 19 is not very high, which is 13534 & 8411 respectively.

5. Find out the properties of parking violations across different times of the day:
- (1) 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.)
 - (2) 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.
 - (3) Divide 24 hours into six equal discrete bins of time. Choose the intervals as you see fit. For each of these groups, find the three most commonly occurring violations.
(Hint: Use the CASE-WHEN in SQL view to segregate into bins. To find the most commonly occurring violations, you can use an approach similar to the one mentioned in the hint for question 4.)
 - (4) Now, try another direction. For the three most commonly occurring violation codes, find the most common time of the day (in terms of the bins from the previous part).
6. Let's try and find some seasonality in this data:
- i) First, divide the year into a certain number of seasons, and find the frequencies of tickets for each season. (Hint: Use Issue Date to segregate into seasons.)
 - ii) Then, find the three most common violations for each of these seasons. (Hint: You can use an approach similar to the one mentioned in the hint for question 4.)
7. 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:
- (1) Find the total occurrences of the three most common violation codes.

- (2) 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.

- (3) 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.
- (4) What can you intuitively infer from these findings?