Problem Statement:

New York City is a thriving metropolis and like most other cities of its size, one of the biggest problems faced by its residents is the lack of parking space. The classic combination of a huge number of cars and cramped geography is the exact recipe that leads to a large number of parking tickets.

In an attempt to scientifically analyze this phenomenon, the NYC Police Department regularly collects data related to parking tickets. This data is made available by the NYC Open Data portal. Your job is to try and perform some analysis on this data, in order to answer the questions that follow.

You need to copy this data to your own bucket and then run the required queries

Consider only the data for the year 2017 itself for your analysis, not the current year.

The data dictionary is available on this page.

The analysis can be divided into two parts:

Part-I: Examine the data

1. Find the total number of tickets for the year.
2. Find out the total number of states to which the cars with tickets belong. The count of states is mandatory here, providing the exact list of states is optional.
3. Some parking tickets don't have addresses on them, which is a cause for concern. Find out the number of such tickets, which have no addresses. (i.e. tickets where one of the Street Codes, i.e. "Street Code 1" or "Street Code 2" or "Street Code 3" is empty)

Part-II: Aggregation tasks

1. What are the top 5 most frequently occurring violation codes? (Note that frequency means the number of occurrences over a time period. The list should be in descending order)
2. How often does each vehicle body type get a parking ticket? How about the vehicle? (List the top 5 for both)
3. A precinct is a police station that has a certain zone of the city under its command. You will find two further classifications of precincts:
   a. Violating Precincts - These are precincts where the violations have occurred.
   b. Issuer Precincts - These are precincts that issued the tickets.
4. Find the top 5 Violating Precincts and Issuer Precincts by frequency.
5. Find the violation code frequency across the top 3 precincts which have issued the highest number of tickets. Do these precinct zones have an exceptionally high frequency of certain violation codes? If yes, list them.

a. Find out the frequency of parking violations across different times of the day: The Violation Time field is specified in a strange format. Find a way to make this into a time attribute that you can use to divide into groups.
b. Divide 24 hours into 6 equal discrete bins of time. The intervals you choose are at your discretion. For each of these groups, find the 3 most commonly occurring violations.
c. Now, try another direction. For the 3 most commonly occurring violation codes, find the most common times of day (in terms of the bins from the previous part).

Let's try and find some seasonality in this data:
a. First, divide the year into seasons, and find frequencies of tickets for each season. (Hint: A quick Google search reveals the following seasons in NYC: Spring(March, April, May); Summer(June, July, August); Fall(September, October, November); Winter(December, January, February))
b. Then, find the 3 most common violations for each of these seasons.