

Hive-Mini-Project

Objective - The assignment is meant for you to apply learnings of the module on Hive on a real-life dataset. One of the major objectives of this assignment is gaining familiarity with how an analysis works in Hive and how you can gain insights from large datasets.

Problem Statement - New York City is a thriving metropolis and just like most other cities of similar size, one of the biggest problems its residents face is parking. The classic combination of a huge number of cars and a cramped geography is the exact recipe that leads to a large number of parking tickets.

In an attempt to scientifically analyse this phenomenon, the NYC Police Department regularly collects data related to parking tickets. This data is made available by NYC Open Data portal. We will try and perform some analysis on this data.

Download Dataset - <https://data.cityofnewyork.us/browse?q=parking+tickets>

Note: Consider only the year 2017 for analysis and not the Fiscal year for year 2017 only/

```
hive> create table parking_violations_tickets_issued
```

```
> (
> Summons_Number bigint,
> Plate_ID string,
> Registration_State string,
> Plate_Type string,
> Issue_Date date,
> Violation_Code int,
> Vehicle_Body_Type string,
> Vehicle_Make string,
> Issuing_Agency string,
> Street_Code1 int,
> Street_Code2 int,
> Street_Code3 int,
```

> Vehicle_Expiration Date,
> Violation_Location int,
> Violation_Precinct int,
> Issuer_Precinct int,
> Issuer_Code int,
> Issuer_Command string,
> Issuer_Squad string,
> Violation_Time string,
> Time_First_Observed string,
> Violation_County string,
> Violation_In_Front_Of_Opposite string,
> House_Number string,
> Street_Name string,
> Intersecting_Street string,
> Date_First_Observed int,
> Law_Section int,
> Sub_Division string,
> Violation_Legal_Code string,
> Days_Parking_In_Effect string,
> From_Hours_In_Effect string,
> To_Hours_In_Effect string,
> Vehicle_Color string,
> Unregistered_Vehicle int,
> Vehicle_Year string,
> Meter_Number string,

```

> Feet_From_Curb int,
> Violation_Post_Code string,
> Violation_Description string,
> No_Standing_or_Stopping_Violation string,
> Hydrant_Violation string,
> Double_Parking_Violation string)
> row format delimited
> fields terminated by ','
> tblproperties ("skip.header.line.count" = "1");

```

OK

Time taken: 0.224 seconds

```

hive> create table parking_violations_tickets_issued
  > (
  > Summons_Number bigint,
  > Plate_ID string,
  > Registration_State string,
  > Plate_Type string,
  > Issue_Date date,
  > Violation_Code int,
  > Vehicle_Body_Type string,
  > Vehicle_Make string,
  > Issuing_Agency string,
  > Street_Code1 int,
  > Street_Code2 int,
  > Street_Code3 int,
  > Vehicle_Expiration_Date,
  > Violation_Location int,
  > Violation_Precinct int,
  > Issuer_Precinct int,
  > Issuer_Code int,
  > Issuer_Command string,
  > Issuer_Squad string,
  > Violation_Time string,
  > Time_First_Observed string,
  > Violation_County string,
  > Violation_In_Front_of_or_Opposite string,
  > House_Number string,
  > Street_Name string,
  > Intersecting_Street string,
  > Date_First_Observed int,
  > Law_Section int,
  > Sub_Division string,
  > Violation_Legal_Code string,
  > Days_Parking_In_Effect string,
  > From_Hours_In_Effect string,
  > To_Hours_In_Effect string,
  > Vehicle_Color string,
  > Unregistered_Vehicle int,
  > Vehicle_Year string,
  > Meter_Number string,
  > Feet_From_Curb int,
  > Violation_Post_Code string,
  > Violation_Description string,
  > No_Standing_or_Stopping_Violation string,
  > Hydrant_Violation string,
  > Double_Parking_Violation string)
> row format delimited
> fields terminated by ','
> tblproperties ("skip.header.line.count" = "1");

```

OK

Time taken: 0.224 seconds

```
hive> load data local inpath 'file:///tmp/hive_class_1/Parking_Violations_Issued_Fiscal_Year_2017.csv'  
into table parking_violations_tickets_issued;
```

Loading data to table parking_project.parking_violations_tickets_issued

Table parking_project.parkingViolations_ticketsIssued stats: [numFiles=1, totalSize=2086913576]

OK

```
hive> load data local inpath 'file:///tmp/hive_class_1/Parking_Violations_Issued_Fiscal_Year_2017.csv' into table parking_violations_tickets_issued;
>Loading data to table parking.project.parking_violations_tickets issued
Table parking.project.parking_violations_tickets issued stats: [numFiles=1, totalSize=2086913576]
OK
```

Creating Partitions and buckets based on country and code.

Creating Tables:-

```
create table parking_violations_2017
```

(

Summons_Number bigint,

Plate_ID string,

Registration_State string,

Plate_Type string,

Issue Date date,

Violation_Code in

Vehicle_Body_Type string

Vehicle_Make string,

Issuing_Agency string

Street_Code1 int,

Street_Code2 int,

Street_Code3 int,

Vehicle_Expiration

Violation_Location int,
Violation_Precinct int,
Issuer_Precinct int,
Issuer_Code int,
Issuer_Command string,
Issuer_Squad string,
Violation_Time string,
Time_First_Observed string,
Violation_In_Front_of_or_Opposite string,
House_Number string,
Street_Name string,
Intersecting_Street string,
Date_First_Observed int,
Law_Section int,
Sub_Division string,
Violation_Legal_Code string,
Days_Parking_In_Effect string,
From_Hours_In_Effect string,
To_Hours_In_Effect string,
Vehicle_Color string,
Unregistered_Vehicle int,
Vehicle_Year string,
Meter_Number string,
Feet_From_Curb int,
Violation_Post_Code string,
Violation_Description string,
No_Standing_or_Stopping_Violation string,

```

Hydrant_Violation string,
Double_Parking_Violation string)
partitioned by (Violation_County string)
CLUSTERED BY (Violation_Code) sorted by (Violation_Code) INTO 8 BUCKETS
row format delimited
fields terminated by ','

tblproperties ("skip.header.line.count" = "1");

```

```

hive> create table parking_violations_2017
  > (
  > Summons_Number bigint,
  > Plate_Number string,
  > Registration_State string,
  > Plate_Type string,
  > Issue_Date string,
  > Violation_Code int,
  > Vehicle_Body_Type string,
  > Vehicle_Make string,
  > Issuing_Agency string,
  > Street_Code1 int,
  > Street_Code2 int,
  > Street_Code3 int,
  > Vehicle_Expiration_Date,
  > Violation_Location int,
  > Violation_Precinct int,
  > Issuer_Precinct int,
  > Issuer_Code int,
  > Issuer_Command string,
  > Violation_Time string,
  > Time_First_Observed string,
  > Violation_In_Front_of_or_Opposite string,
  > House_Number string,
  > Street_Name string,
  > Intersecting_Street string,
  > Date_First_Observed int,
  > Lat_Direction string,
  > Sub_Direction string,
  > Violation_Legal_Code string,
  > Days_Parking_In_Effect string,
  > From_Hours_In_Effect string,
  > To_Hours_In_Effect string,
  > Vehicle_Color string,
  > Unregistered_Vehicle int,
  > Vehicle_Fuel string,
  > Meter_Number string,
  > Feet_From_Curb int,
  > Violation_Post_Code string,
  > Violation_Description string,
  > No_Standing_or_Stopping_Violation string,
  > Hydrant_Violation string,
  > Double_Parking_Violation string)
  > partitioned by (Violation_County string)
  > CLUSTERED BY (Violation_Code) sorted by (Violation_Code) INTO 8 BUCKETS
  > row format delimited
  > fields terminated by ','
  > tblproperties ("skip.header.line.count" = "1");
OK
Time taken: 0.261 seconds

```

Beforing loading data we need to set some properties for hive.

```

hive> set hive.exec.dynamic.partition=true;
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> set hive.enforce.bucketing = true;

```

Inserting Data into parking_violations_2017

```

hive> insert into parking_violations_2017
  > partition(Violation_County) Select
  > Summons_Number,

```

> Plate_ID,
> Registration_State,
> Plate_Type,
> Issue_Date,
> Violation_Code,
> Vehicle_Body_Type,
> Vehicle_Make,
> Issuing_Agency,
> Street_Code1,
> Street_Code2,
> Street_Code3,
> Vehicle_Expiration,
> Violation_Location,
> Violation_Precinct,
> Issuer_Precinct,
> Issuer_Code,
> Issuer_Command,
> Issuer_Squad,
> Violation_Time,
> Time_First_Observed,
> Violation_In_Front_Of_Or_Opposite,
> House_Number,
> Street_Name,
> Intersecting_Street,
> Date_First_Observed,
> Law_Section,
> Sub_Division,

```

> Violation_Legal_Code,
> Days_Parking_In_Effect,
> From_Hours_In_Effect,
> To_Hours_In_Effect,
> Vehicle_Color,
> Unregistered_Vehicle,
> Vehicle_Year,
> Meter_Number,
> Feet_From_Curb,
> Violation_Post_Code,
> Violation_Description,
> No_Standing_or_Stopping_Violation,
> Hydrant_Violation,
> Double_Parking_Violation,
> Violation_County
> from parking_violations_tickets_issued
> where year(Issue_Date) = '2017';

```

```

Time taken for adding to write entity : 13
Partition parking_project.parking_violations_2017[violation_county=BR] stats: [numFiles=8, numRows=38406, totalSize=6244359, rawDataSize=6605953]
Partition parking_project.parking_violations_2017[violation_county=BX] stats: [numFiles=8, numRows=6222, totalSize=19735108, rawDataSize=12666816]
Partition parking_project.parking_violations_2017[violation_county=K1] stats: [numFiles=8, numRows=12049, totalSize=20957070, rawDataSize=20845021]
Partition parking_project.parking_violations_2017[violation_county=K2] stats: [numFiles=8, numRows=14617, totalSize=7200000, rawDataSize=14617]
Partition parking_project.parking_violations_2017[violation_county=NY] stats: [numFiles=8, numRows=178636, totalSize=33500941, rawDataSize=33232305]
Partition parking_project.parking_violations_2017[violation_county=Q1] stats: [numFiles=8, numRows=92949, totalSize=1790777, rawDataSize=17897828]
Partition parking_project.parking_violations_2017[violation_county=QN] stats: [numFiles=8, numRows=31123, totalSize=5450245, rawDataSize=5419122]
Partition parking_project.parking_violations_2017[violation_county=QN1] stats: [numFiles=8, numRows=1, totalSize=167, rawDataSize=166]
Partition parking_project.parking_violations_2017[violation_county=R] stats: [numFiles=8, numRows=6262, totalSize=121075, rawDataSize=1203813]
Partition parking_project.parking_violations_2017[violation_county=S1] stats: [numFiles=8, numRows=6530, totalSize=145206, rawDataSize=138676]
Partition parking_project.parking_violations_2017[violation_county=_HIVE_DEFAULT_PARTITION_] stats: [numFiles=8, numRows=1697, totalSize=277281, rawDataSize=275584]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 0 Cumulative CTFU: 46.72 sec HDFS Read: 202089431 HDFS Write: 100588846 SUCCESS
Total MapReduce CTFU Time Spent: 46 seconds 720 msec
OK
Time taken: 67.783 seconds

```

The analysis can be divided into two parts:

Part-I: Examine the data

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

```
hive> select count(distinct summons_number) as Number_of_Tickets,year(issue_date) as issue_year
from parking_violations_2017 group by year(issue_date);
```

```

hive> select count(distinct summons_number) as Number_of_Tickets,year(issue_date) as issue_year from parking_violations_2017 group by year(issue_date);
Query ID = cloudera_20221020230000_69d39bb6-a8f0-418f-9924-07c2f59cd64
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0174, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0174/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0174
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-20 23:00:09,795 Stage-1 map = 0%, reduce = 0%
2022-10-20 23:00:18,355 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 1.25 sec
2022-10-20 23:00:19,452 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.94 sec
2022-10-20 23:00:26,223 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.69 sec
MapReduce Total cumulative CPU time: 10 seconds 690 msec
Ended Job = job_1662892505239_0174
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 10.69 sec HDFS Read: 100627859 HDFS Write: 12 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 690 msec
OK
539901 2017
Time taken: 26.315 seconds, Fetched: 1 row(s)

```

Total number of tickets for the year 2017 are 539901.

2.) Find out how many unique states the cars which got parking tickets came from.

```
hive> select count(distinct Registration_State) as Registration_States from parking_violations_2017;
```

```

hive> select count(distinct Registration_State) as Registration_States from parking_violations_2017;
Query ID = cloudera_20221021182121_00f9c4a0-blcd-4f64-aaba-ebc3bc2092ed
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0177, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0177/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0177
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 18:21:41,049 Stage-1 map = 0%, reduce = 0%
2022-10-21 18:21:47,623 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 1.35 sec
2022-10-21 18:21:48,657 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.39 sec
2022-10-21 18:21:54,881 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.55 sec
MapReduce Total cumulative CPU time: 5 seconds 550 msec
Ended Job = job_1662892505239_0177
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 5.55 sec HDFS Read: 100626648 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 550 msec
OK
63
Time taken: 22.648 seconds, Fetched: 1 row(s)

```

Total number of Unique States are : 63 .

3.) Some parking tickets don't have addresses on them, which is cause for concern. Find out how many such tickets there are(i.e. tickets where either "Street Code 1" or "Street Code 2" or "Street Code 3" is empty).

```
hive> select count(distinct summons_number) as Tickets_without_address from parking_violations_2017 where Street_code1=0 or Street_code2=0 or Street_code3=0;
```

```

hive> select count(distinct summons_number) as tickets without address from parking_violations_2017 where street_code1=0 or street_code2=0 or street_code3=0;
Query ID = cloudera_20221021182828_5b5cfbd0-8ebc-4686-afe8-e41f8090054a
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0179, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0179/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0179
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 18:28:49,390 Stage-1 map 0%, reduce = 0%
2022-10-21 18:28:55,603 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 1.41 sec
2022-10-21 18:28:56,633 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.75 sec
2022-10-21 18:29:02,811 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.19 sec
MapReduce Total cumulative CPU time: 8 seconds 190 msec
Ended Job = job_1662892505239_0179
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 8.19 sec HDFS Read: 100628473 HDFS Write: 7 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 190 msec
OK
180473
Time taken: 20.601 seconds, Fetched: 1 row(s)

```

Number of Tickets without Address:180473.

Part-II: Aggregation tasks

1.) How often does each violation code occur? (frequency of violation codes - find the top 5)

```

hive> select count(Violation_Code) as frequency_ofViolation_code, Violation_Code from
parkingViolations_2017 group by Violation_Code order by frequency_ofViolation_code desc limit 5;

```

```

hive> select count(Violation_Code) as frequency_ofViolation_code, Violation_Code from parkingViolations_2017 group by Violation_Code order by frequency_ofViolation_code desc limit 5;
Query ID = cloudera_20221021183636_lc3d3397-1dae-49a4-9836-f11330df1ff
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0180, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0180/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0180
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 18:36:34,264 Stage-1 map = 0%, reduce = 0%
2022-10-21 18:36:40,493 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 0.08 sec
2022-10-21 18:36:41,159 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.15 sec
2022-10-21 18:36:47,737 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.89 sec
MapReduce Total cumulative CPU time: 6 seconds 890 msec
Ended Job = job_1662892505239_0180
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0181, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0181/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0181
Hadoop job information for Stage-2: number of mappers: 17; number of reducers: 1
2022-10-21 18:36:54,446 Stage-2 map = 0%, reduce = 0%
2022-10-21 18:36:58,700 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.75 sec
2022-10-21 18:37:04,700 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.01 sec
MapReduce Total cumulative CPU time: 2 seconds 10 msec
Ended Job = job_1662892505239_0181
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 6.89 sec HDFS Read: 100625976 HDFS Write: 1915 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.01 sec HDFS Read: 6845 HDFS Write: 45 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 900 msec
OK
76292 21
66094 36
53782 38
47179 14
31858 20
Time taken: 37.05 seconds, Fetched: 5 rows
hive>
```

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

```

hive> select vehicleBodyType, count(summons_number) as frequency_ofVehicleBody from
parkingViolations_2017 group by vehicleBodyType order by frequency_ofVehicleBody desc limit 5;

```

```

hive> select Vehicle_Body_Type,count(summons_number)as frequency_of_Vehicle_Body from parkingViolations_2017 group by vehicle_body_type order by frequency_of_vehicle_body desc limit 5;
Query ID = cloudera_20221021184141_162138de-704f-4902-b22d-fdc9607614a2
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0182, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0182/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0182
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 18:41:44,754 Stage-1 map = 0%, reduce = 0%
2022-10-21 18:41:44,754 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.58 sec
2022-10-21 18:41:57,252 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.69 sec
MapReduce Total cumulative CPU time: 4 seconds 690 msec
Ended Job = job_1662892505239_0182
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0183, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0183/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0183
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 18:42:03,923 Stage-2 map = 0%, reduce = 0%
2022-10-21 18:42:03,923 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.84 sec
2022-10-21 18:42:16,158 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.96 sec
MapReduce Total cumulative CPU time: 1 seconds 960 msec
Ended Job = job_1662892505239_0183
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 4.69 sec HDFS Read: 100625793 HDFS Write: 8610 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.96 sec HDFS Read: 13761 HDFS Write: 55 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 650 msec
OK
FBRN 197286
ASBD 153957
VAN 71767
DELV 35602
SDN 19238
Time taken: 37.422 seconds, Fetched: 5 row(s)
hive> 
```

hive> select Vehicle_make,count(summons_number) as frequency_of_vehicle_make from parkingViolations_2017 group by vehicle_make order by frequency_of_vehicle_make desc limit 5;

```

hive> select Vehicle_make,count(summons_number) as frequency_of_vehicle_make from parkingViolations_2017 group by vehicle_make order by frequency_of_vehicle_make desc limit 5;
Query ID = cloudera_20221021184545_b49d1856-2f6b-4944-b0b6-3fcc5b231a5c
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0184, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0184/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0184
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 18:46:03,570 Stage-1 map = 0%, reduce = 0%
2022-10-21 18:46:10,810 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.48 sec
2022-10-21 18:46:17,040 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.65 sec
MapReduce Total cumulative CPU time: 4 seconds 650 msec
Ended Job = job_1662892505239_0184
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0185, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0185/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0185
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 18:46:23,771 Stage-2 map = 0%, reduce = 0%
2022-10-21 18:46:28,958 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.88 sec
2022-10-21 18:46:34,123 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.01 sec
MapReduce Total cumulative CPU time: 2 seconds 10 msec
Ended Job = job_1662892505239_0185
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 4.65 sec HDFS Read: 100625778 HDFS Write: 17061 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.01 sec HDFS Read: 22192 HDFS Write: 59 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 660 msec
OK
FORD 63183
TOYOT 59913
HONDA 53972
NISSA 45341
CHEVR 35305
Time taken: 37.271 seconds, Fetched: 5 row(s)
hive> [cloudera@quickstart ~]$ 
```

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

a.) Violating Precincts (this is the precinct of the zone where the violation occurred)

hive> select Violation_Precinct,count(Violation_Precinct) as frequency_violation_precinct from parkingViolations_2017 group by Violation_Precinct order by frequency_violation_precinct desc limit 5;

```

hive> select Violation_Precinct,count(Violation_Precinct) as frequencyViolation_Precinct from parkingViolations_2017 group by Violation_Precinct order by frequencyViolation_Precinct desc limit 5;
Query ID = cloudera_20221021185151_a2de5c20-7b6e-4925-8c2d-f49825d1db86
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0186, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0186/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0186
Hadoop job information for Stage-1: number of mappers: 27 number of reducers: 1
2022-10-21 18:51:43,650 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.4 sec
2022-10-21 18:51:49,363 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 3.38 sec
2022-10-21 18:51:50,.392 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.38 sec
2022-10-21 18:51:55,.582 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.42 sec
MapReduce Total cumulative CPU time: 4 seconds 420 msec
Ended Job = job_1662892505239_0186
Launching Job 2 out of 2
Number of reduce tasks not specified. Estimated at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0187, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0187/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0187
Hadoop job information for Stage-2: number of mappers: 17 number of reducers: 1
2022-10-21 18:52:02,.478 Stage-2 map = 0%, reduce = 0%
2022-10-21 18:52:08,.699 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.75 sec
2022-10-21 18:52:14,.042 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.82 sec
MapReduce Total cumulative CPU time: 1 seconds 820 msec
Ended Job = job_1662892505239_0187
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 4.42 sec HDFS Read: 100625828 HDFS Write: 2600 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.82 sec HDFS Read: 7742 HDFS Write: 43 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 240 msec
OK
0 92182
19 27131
14 20021
1 17465
18 16929
Time taken: 40.132 seconds, Fetched: 5 row(s)
hive> 
```

b.) Issuer Precincts (this is the precinct that issued the ticket)

```

hive> select Issuer_Precinct,count(Issuer_Precinct) as frequencyViolation_Issues from parkingViolations_2017 group by Issuer_Precinct order by frequencyViolation_Issues desc limit 5;
```

```

hive> select Issuer_Precinct,count(Issuer_Precinct) as frequencyViolation_Issues from parkingViolations_2017 group by Issuer_Precinct order by frequencyViolation_Issues desc limit 5;
Query ID = cloudera_20221021185454_56e71221_648d-4e45-aecb-8cf9394a8b0d
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0188, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0188/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0188
Hadoop job information for Stage-1: number of mappers: 27 number of reducers: 1
2022-10-21 18:54:47,145 Stage-1 map = 0%, reduce = 0%
2022-10-21 18:54:53,422 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.1 sec
2022-10-21 18:54:58,738 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.32 sec
MapReduce Total cumulative CPU time: 5 seconds 320 msec
Ended Job = job_1662892505239_0188
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0189, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0189/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0189
Hadoop job information for Stage-2: number of mappers: 17 number of reducers: 1
2022-10-21 18:55:11,.583 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.78 sec
2022-10-21 18:55:16,.738 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.96 sec
MapReduce Total cumulative CPU time: 1 seconds 960 msec
Ended Job = job_1662892505239_0189
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 5.32 sec HDFS Read: 100625811 HDFS Write: 5402 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.96 sec HDFS Read: 10531 HDFS Write: 44 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 280 msec
OK
0 107287
19 26391
14 19970
1 16873
18 16330
Time taken: 37.086 seconds, Fetched: 5 row(s)
hive> 
```

4.) Find the violation code frequency across 3 precincts which have issued the most number of tickets - do these precinct zones have an exceptionally high frequency of certain violation codes?

```

hive> select Issuer_Precinct, Violation_Code, count(*) as Violation_Code_Frequency from parkingViolations_2017 group by Issuer_Precinct, Violation_Code order by Violation_Code_Frequency desc limit 10;
```

```

hive> select Issuer_Precinct, Violation_Code, count(*) as Violation_Code_Frequency from parkingViolations_2017 group by Issuer_Precinct, Violation_Code order by Violation_Code_Frequency desc limit 10;

```

```

Query ID = cloudera_20221021190404_cb06a58f-e636-430a-836b-bf0ea2b5e3f3
Total Jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0190, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0190/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0190
Hadoop job information for Stage-1: number of mappers: 27 number of reducers: 1
2022-10-21 19:04:35,277 Stage-1 map = 0%    reduce = 0%
2022-10-21 19:04:40,498 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 1.02 sec
2022-10-21 19:04:42,534 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.79 sec
2022-10-21 19:04:49,791 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.9 sec
MapReduce Total cumulative CPU time: 5 seconds 300 msec
Ended Job = job_1662892505239_0190
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0191, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0191/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0191
Hadoop job information for Stage-2: number of mappers: 17 number of reducers: 1
2022-10-21 19:05:00,744 Stage-2 map = 0%, reduce = 0%, Cumulative CPU 0.0 sec
2022-10-21 19:05:01,719 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.46 sec
2022-10-21 19:05:06,963 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.68 sec
MapReduce Total cumulative CPU time: 2 seconds 680 msec
Ended Job = job_1662892505239_0191
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1  Cumulative CPU: 5.3 sec  HDFS Read: 100626171 HDFS Write: 82616 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1  Cumulative CPU: 2.68 sec  HDFS Read: 88043 HDFS Write: 106 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 960 msec
OK
issuer_precinct violation_code violation_code_frequency
0      36      66094
0       7      20897
0      21     12440
18      14      5936
0       5      4725
19      46     4707
14      14     4395
1       14      3821
19      37      3649
19      38      3574
Time taken: 38.964 seconds, Fetched: 10 row(s)

```

We are not considering 0 and 1,Therefore 18,19,14 are the three issue precincts which have the maximum number of violations .

For Issue Precincts=18:-

```

hive> select issuer_precinct,Violation_Code,count(*) as Violation_Code_Frequency from
parking_violations_2017 where issuer_precinct=18 group by issuer_precinct,Violation_Code order by
Violation_Code_Frequency desc limit 10;

```

```

hive> select issuer_precinct,Violation_Code,count(*) as Violation_Code_Frequency from parking_violations_2017 where issuer_precinct=18 group by issuer_precinct,Violation_Code order by Violation_Code_Frequency desc limit 10;

```

```

Query ID = cloudera_20221021191212_704032e5-felb-40d4-87e9-3e2b1201f7e4
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0192, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0192/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0192
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 19:12:15,377 Stage-1 map = 0%,  reduce = 0%
2022-10-21 19:12:22,729 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.89 sec
2022-10-21 19:12:27,900 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.22 sec
MapReduce Total cumulative CPU time: 5 seconds 220 msec
Ended Job = job_1662892505239_0192
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0193, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0193/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0193
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 19:12:34,031 Stage-2 map = 0%,  reduce = 0%
2022-10-21 19:12:39,194 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.79 sec
2022-10-21 19:12:44,395 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.92 sec
MapReduce Total cumulative CPU time: 1 seconds 920 msec
Ended Job = job_1662892505239_0193
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1  Cumulative CPU: 5.22 sec  HDFS Read: 100627723 HDFS Write: 1328 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 1.92 sec  HDFS Read: 6756 HDFS Write: 104 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 140 msec
OK
issuer_precinct violation_code  violation_code_frequency
18      14      5073
18      69      2027
18      47      1366
18      31      1224
18      46      810
18      42      619
18      38      612
18      84      497
18      19      468
18      20      385
Time taken: 35.159 seconds, Fetched: 10 row(s)

```

For Issue Precincts=19:-

```

hive> select issuer_precinct,Violation_Code,count(*) as Violation_Code_Frequency from
parking_violations_2017 where issuer_precinct=19 group by issuer_precinct,Violation_Code order by
Violation_Code_Frequency desc limit 10;

```

```

hive> select issuer_precinct,Violation_Code,count(*) as Violation_Code_Frequency from parking_violations_2017 where issuer_precinct=19 group by issuer_precinct,Violation_Code order by Violation_Code_Frequency desc limit 10;

```

```

Query ID = cloudera_20221021191515_870bd7f1-8731-4400-a4a5-328d6c5942d3
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0194, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0194/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0194
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 19:15:37,206 Stage-1 map = 0%,  reduce = 0%
2022-10-21 19:15:44,496 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.05 sec
2022-10-21 19:15:49,661 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.14 sec
MapReduce Total cumulative CPU time: 5 seconds 140 msec
Ended Job = job_1662892505239_0194
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0195, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0195/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0195
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 19:15:55,776 Stage-2 map = 0%,  reduce = 0%
2022-10-21 19:16:00,934 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.78 sec
2022-10-21 19:16:07,127 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.88 sec
MapReduce Total cumulative CPU time: 1 seconds 880 msec
Ended Job = job_1662892505239_0195
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1  Cumulative CPU: 5.14 sec  HDFS Read: 100627723 HDFS Write: 1231 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 1.88 sec  HDFS Read: 6659 HDFS Write: 107 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 20 msec
OK
issuer_precinct violation_code  violation_code_frequency
19      46      4707
19      37      3649
19      38      3574
19      14      2888
19      21      2842
19      20      1479
19      40      1070
19      16      982
19      71      752
19      19      686
Time taken: 36.18 seconds, Fetched: 10 row(s)

```

For Issue Precincts=14:-

```

hive> select issuer_precinct,Violation_Code,count(*) as Violation_Code_Frequency from
parking_violations_2017 where issuer_precinct=14 group by issuer_precinct,Violation_Code order by
Violation_Code_Frequency desc limit 10;

```

```

hive> select issuer_precinct,Violation_Code,count(*) as Violation_Code_Frequency from parking_violations_2017 where issuer_precinct=14 group by issuer_precinct,Violation_Code order by Violation_Code_Frequency desc limit 10;

```

```

Query ID = cloudera_20221021191717_e8eb5fd5-aa44-4650-a7ab-a622aac88464
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0196, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0196/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0196
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 19:17:55,157 Stage-1 map = 0%,  reduce = 0%
2022-10-21 19:18:01,433 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 2.57 sec
2022-10-21 19:18:02,461 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.73 sec
2022-10-21 19:18:07,650 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.83 sec
MapReduce Total cumulative CPU time: 4 seconds 830 msec
Ended Job = job_1662892505239_0196
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0197, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0197/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0197
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 19:18:14,728 Stage-2 map = 0%,  reduce = 0%
2022-10-21 19:18:18,875 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.73 sec
2022-10-21 19:18:25,081 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.87 sec
MapReduce Total cumulative CPU time: 1 seconds 870 msec
Ended Job = job_1662892505239_0197
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1  Cumulative CPU: 4.83 sec  HDFS Read: 100627723 HDFS Write: 1395 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 1.87 sec  HDFS Read: 6823 HDFS Write: 104 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 700 msec
OK
issuer_precinct violation_code  violation_code_frequency
14      14      4395
14      69      2997
14      31      2279
14      47      1855
14      42      986
14      46      734
14      19      702
14      84      683
14      82      533
14      40      391
Time taken: 37.635 seconds, Fetched: 10 row(s)

```

Common Codes:-

```

hive> select issuer_precinct,Violation_Code,count(*) as Violation_Code_Frequency from
parking_violations_2017 where issuer_precinct in(14,19,18) group by issuer_precinct,Violation_Code
order by Violation_Code_Frequency desc limit 10;

```

```

hive> select issuer_precinct,Violation_Code,count(*) as Violation_Code_Frequency from parking_violations_2017 where issuer_precinct in(14,19,18) group by issuer_precinct,Violation_Code order
by Violation_Code_Frequency desc limit 10;

```

```

Query ID = cloudera_20221021192121_f0d468a5-cb5a-4853-9c65-2a86dc15bf3b
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0198, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0198/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0198
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 19:21:07,345 Stage-1 map = 0%,  reduce = 0%
2022-10-21 19:21:13,672 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.08 sec
2022-10-21 19:21:18,956 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.25 sec
MapReduce Total cumulative CPU time: 5 seconds 250 msec
Ended Job = job_1662892505239_0198
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0199, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0199/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0199
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 19:21:26,057 Stage-2 map = 0%,  reduce = 0%
2022-10-21 19:21:31,277 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.8 sec
2022-10-21 19:21:37,487 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.1 sec
MapReduce Total cumulative CPU time: 2 seconds 100 msec
Ended Job = job_1662892505239_0199
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1  Cumulative CPU: 5.25 sec  HDFS Read: 100624827 HDFS Write: 3782 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 2.1 sec  HDFS Read: 9201 HDFS Write: 110 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 350 msec
OK
issuer_precinct violation_code violation_code_frequency
18      14      5073
19      46      4707
14      14      4395
19      37      3649
19      38      3574
14      69      2997
19      14      2888
19      21      2842
14      31      2279
18      69      2027
Time taken: 37.349 seconds, Fetched: 10 row(s)
hive> 
```

5.) Find out the properties 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.

```

hive> select from_unixtime(unix_timestamp(concat(violation_time,'M'),'HHmmSS'),"HH:mm:ss") as
data_for_dates from parkingViolations_2017 limit 10;
```

```

hive> select from_unixtime(unix_timestamp(concat(violation_time,'M'),'HHmmSS'),"HH:mm:ss") as data_for_dates from parkingViolations_2017 limit 10;
OK
data_for_dates
12:51PM
02:56AM
10:24AM
11:46AM
11:30AM
08:35AM
10:02AM
12:16PM
09:44AM
01:30AM
Time taken: 0.125 seconds, Fetched: 10 row(s)
hive> 
```

6.) 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

>First we need to divide 24 hours into 6 equal parts by creating view on partition table

```
create view vw_parkingViolations_partitioned_2017
```

```
partitioned on (Violation_Code)
```

as

SELECT

Summons_Number,

Violation_Time,

Issuer_Precinct,

case

when substring(Violation_Time,1,2) in ('00','01','02','03','12') and upper(substring(Violation_Time,-1))='A' then 1

when substring(Violation_Time,1,2) in ('04','05','06','07') and upper(substring(Violation_Time,-1))='A' then 2

when substring(Violation_Time,1,2) in ('08','09','10','11') and upper(substring(Violation_Time,-1))='A' then 3

when substring(Violation_Time,1,2) in ('12','00','01','02','03') and upper(substring(Violation_Time,-1))='P' then 4

when substring(Violation_Time,1,2) in ('04','05','06','07') and upper(substring(Violation_Time,-1))='P' then 5

when substring(Violation_Time,1,2) in ('08','09','10','11') and upper(substring(Violation_Time,-1))='P' then 6

else null

end as Violation_Time_bin,

Violation_Code

from parking_violations_2017

where

Violation_Time is not null

or (length(Violation_Time)=5 and upper(substring(Violation_Time,-1))in ('A','P'))

and substring(Violation_Time,1,2) in ('00','01','02','03','04','05','06','07','08','09','10','11','12'));

```

hive> create view vw_parking_violations_partitioned_2017
> partitioned on (Violation_Code)
> as
> SELECT
> Summons_Number,
> Violation_Time,
> Issuer_Precinct,
> case
> when substring(Violation_Time,1,2) in ('00','01','02','03','12') and upper(substring(Violation_Time,-1))='A' then 1
> when substring(Violation_Time,1,2) in ('04','05','06','07') and upper(substring(Violation_Time,-1))='A' then 2
> when substring(Violation_Time,1,2) in ('08','09','10','11') and upper(substring(Violation_Time,-1))='A' then 3
> when substring(Violation_Time,1,2) in ('12','00','01','02','03') and upper(substring(Violation_Time,-1))='P' then 4
> when substring(Violation_Time,1,2) in ('04','05','06','07') and upper(substring(Violation_Time,-1))='P' then 5
> when substring(Violation_Time,1,2) in ('08','09','10','11') and upper(substring(Violation_Time,-1))='P' then 6
> else null
> end as Violation_Time_bin,
> Violation_Code
> from parking_violations_2017
> where
> Violation_Time is not null
> or (length(Violation_Time)=5 and upper(substring(Violation_Time,-1)) in ('A','P'))
> and substring(Violation_Time,1,2) in ('00','01','02','03','04','05','06','07','08','09','10','11','12'));
OK
summons_number violation_time issuer_precinct violation_time_bin      violation_code
Time taken: 0.267 seconds
hive> [REDACTED]

```

```

hive> select violation_code,violation_time_bin,count(*) frequency_violation from
vw_parking_violations_partitioned_2017 group by violation_code,violation_time_bin limit 7;

```

```

hive> select violation_code,violation_time_bin,count(*) frequency_violation from vw_parking_violations_partitioned_2017 group by violation_code,violation_time_bin limit 7;
Query ID = cloudera_20221021195050_e6cbbbba-6803-49c8-acb0-2440cded19de
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0200, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0200/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0200
Hadoop job information for Stage-1: number of mappers: 27 number of reducers: 1
2022-10-21 19:50:51,669 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 3.26 sec
2022-10-21 19:50:52,719 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.67 sec
2022-10-21 19:50:57,895 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 11.08 sec
MapReduce Total cumulative CPU time: 11 seconds 80 msec
Ended Job = job_1662892505239_0200
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 11.08 sec   HDFS Read: 100633871 HDFS Write: 46 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 80 msec
OK
violation_code  violation_time_bin      frequency_violation
0              3                  5
0              4                  13
0              5                  3
1              2                  2
1              3                  22
1              4                  21
1              5                  13
Time taken: 22.55 seconds, Fetched: 7 row(s)
hive> [REDACTED]

```

for bin-1:-

```

select Violation_Code,count(*) frequency_violation from vw_parking_violations_partitioned_2017 where
Violation_Time_bin =1 group by Violation_Code order by frequency_violation desc limit 7;

```

```

hive> select Violation_Code,count(*) frequencyViolation from vw_parking_violations_partitioned_2017 where Violation_Time_bin =1 group by Violation_Code order by frequencyViolation desc limit 7;
Query ID = cloudera_20221021200101_79908ae7-a153-4765-8ce4-5efb5e330186
Total Jobs: 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0207, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0207/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0207
Hadoop job information for Stage-1: number of mappers: 2/ number of reducers: 1
2022-10-21 20:01:27,158 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:01:27,158 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.38 sec
2022-10-21 20:01:25,534 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.29 sec
2022-10-21 20:01:39,693 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.93 sec
MapReduce Total cumulative CPU time: 7 seconds 930 msec
Ended Job = job_1662892505239_0207
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0208, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0208/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0208
Hadoop job information for Stage-2: number of mappers: 1/ number of reducers: 1
2022-10-21 20:01:45,843 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:01:51,025 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.83 sec
2022-10-21 20:01:57,323 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.08 sec
MapReduce Total cumulative CPU time: 2 seconds 80 msec
Ended Job = job_1662892505239_0208
MapReduce Tasks: Map: 2 Reduce: 1 Cumulative CPU: 7.93 sec HDFS Read: 100633429 HDFS Write: 1379 SUCCESS
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.08 sec HDFS Read: 6431 HDFS Write: 54 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 10 msec
OK
violation_code frequencyViolation
21      3659
40      2583
14      1547
78      1532
20      1161
7       1022
46      653
Time taken: 36.189 seconds, Fetched: 7 row(s)
hive> 
```

```

OK
violation_code  frequencyViolation
21      3659
40      2583
14      1547
78      1532
20      1161
7       1022
46      653
Time taken: 36.189 seconds, Fetched: 7 row(s)

```

for bin-2:-

```

select Violation_Code,count(*) frequency_Violation from vw_parking_violations_partitioned_2017 where
Violation_Time_bin =2 group by Violation_Code order by frequency_Violation desc limit 7;

```

```

hive> select Violation_Code,count(*) frequencyViolation from vw_parking_violations_partitioned_2017 where Violation_Time_Bin = 2 group by Violation_Code order by frequencyViolation desc limit 7;
Query ID = cloudera_20221021200202_74a3412e-5649-43d7-9513-00b28f3d713f
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducer.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0209, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0209/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0209
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 20:03:04,973 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:03:11,396 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 1.54 sec
2022-10-21 20:03:13,448 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.93 sec
2022-10-21 20:03:18,608 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.66 sec
MapReduce Total cumulative CPU time: 8 seconds 660 msec
Ended Job = job_1662892505239_0209
Job ID: job_1662892505239_0209
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducer.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0210, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0210/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0210
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 20:03:25,360 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:03:29,519 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.81 sec
2022-10-21 20:03:35,718 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.93 sec
MapReduce Total cumulative CPU time: 1 seconds 930 msec
Ended Job = job_1662892505239_0210
MapReduce Jobs Launched
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 0.66 sec HDFS Read: 100633429 HDFS Write: 1574 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.93 sec HDFS Read: 6626 HDFS Write: 55 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 590 msec
OK
violation_code frequencyViolation
14      7250
40      6102
21      5669
20      4320
71      2242
7       1907
46      1622
Time taken: 36.961 seconds, Fetched: 7 row(s)
hive> [REDACTED]

```

```

OK
violation_code frequencyViolation
14      7250
40      6102
21      5669
20      4320
71      2242
7       1907
46      1622
Time taken: 36.961 seconds, Fetched: 7 row(s)

```

for bin-3:-

```

select Violation_Code,count(*) frequency_Violation from vw_parking_violations_partitioned_2017 where
Violation_Time_Bin = 3 group by Violation_Code order by frequency_Violation desc limit 7;

```

```

hive> select Violation_Code,count(*) frequencyViolation from vw_parking_violations_partitioned_2017 where Violation_Time_bin =3 group by Violation_Code order by frequencyViolation desc limit 7;
2022-10-21 20:04:14,952 ID = cloudera_20221021200404_7cb42994-57ba-4875-8e15-6383e1d80134
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0211, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0211/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0211
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 20:04:14,952 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:04:14,952 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.09 sec
2022-10-21 20:04:23,492 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.65 sec
2022-10-21 20:04:30,637 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.5 sec
MapReduce Total cumulative CPU time: 10 seconds 500 msec
Ended Job = job_1662892505239_0211
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0212, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0212/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0212
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-10-21 20:04:37,929 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:04:43,052 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.8 sec
2022-10-21 20:04:47,172 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.95 sec
MapReduce Total cumulative CPU time: 1 seconds 950 msec
Ended Job = job_1662892505239_0212
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1   Cumulative CPU: 10.5 sec  HDFS Read: 100633429 HDFS Write: 1695 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1   Cumulative CPU: 1.95 sec  HDFS Read: 6747 HDFS Write: 61 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 450 msec
OK
violation_code  frequencyViolation
21      59464
36      34768
38      17588
14      14708
46      10893
71      9547
20      8980
Time taken: 39.195 seconds, Fetched: 7 row(s)
hive> 
```

```

OK
violation_code  frequencyViolation
21      59464
36      34768
38      17588
14      14708
46      10893
71      9547
20      8980
Time taken: 39.195 seconds, Fetched: 7 row(s)

```

for bin-4:-

```

select Violation_Code,count(*) frequency_Violation from vw_parking_violations_partitioned_2017 where
Violation_Time_bin =4 group by Violation_Code order by frequency_Violation desc limit 7;
```

```

hive> select Violation_Code,count(*) frequencyViolation from vw_parking_violations_partitioned_2017 where Violation_Time_bin =4 group by Violation_Code order by frequencyViolation desc limit 7;
Query ID = cloudera_20221021200505 d2813ba5-b979-458f-a91d-d161d256a60c
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducer.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0213, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0213/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0213
Hadoop job information for Stage-1: number of mappers: 27 number of reducers: 1
2022-10-21 20:05:31,652 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.38 sec
2022-10-21 20:05:37,659 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 6.6 sec
2022-10-21 20:05:39,728 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.43 sec
2022-10-21 20:05:43,868 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.43 sec
MapReduce Total cumulative CPU time: 8 seconds 430 msec
Ended Job = job_1662892505239_0213
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducer.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0214, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0214/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0214
Hadoop job information for Stage-2: number of mappers: 17 number of reducers: 1
2022-10-21 20:05:50,910 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:05:56,390 Stage-2 map = 100%, reduce = 0%
2022-10-21 20:06:02,658 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.04 sec
MapReduce Total cumulative CPU time: 2 seconds 40 msec
Ended Job = job_1662892505239_0214
MapReduce Jobs Completed: 2
Time taken: 38.487 seconds, Fetched: 7 row(s)
hive> []

```

```

Total MapReduce jobs done: 2  Seconds: 38.487
OK
+-----+
| violation_code | frequencyViolation |
+-----+
|       36       |        28600      |
|       38       |        23877      |
|       37       |        16779      |
|       14       |        13974      |
|       46       |        12372      |
|       20       |        11591      |
|       71       |        10171      |
+-----+
Time taken: 38.487 seconds, Fetched: 7 row(s)
hive> []

```

for bin-5:-

```

select Violation_Code,count(*) frequency_Violation from vw_parking_violations_partitioned_2017 where
Violation_Time_bin =5 group by Violation_Code order by frequency_Violation desc limit 7;

```

```

hive> select Violation_Code,count(*) frequencyViolation from vw_parking_violations_partitioned_2017 where Violation_Time_bin =5 group by Violation_Code order by frequencyViolation desc limit 7;
Query ID = cloudera_20221021200606_27f2c641-5053-49b2-8ca4-476621bf45cb
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducer.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0215, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0215/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0215
Hadoop job information for Stage-1: number of mappers: 27 number of reducers: 1
2022-10-21 20:07:10.883 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 2.99 sec
2022-10-21 20:07:13.967 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.81 sec
2022-10-21 20:07:18.090 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.56 sec
MapReduce Total cumulative CPU time: 10 seconds 560 msec
Ended Job = job_1662892505239_0215
Launching Job 2 out of 2
Number of reducers not specified. Determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducer.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0216, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0216/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0216
Hadoop job information for Stage-2: number of mappers: 17 number of reducers: 1
2022-10-21 20:07:25.859 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:07:32.054 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.8 sec
2022-10-21 20:07:37.210 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.0 sec
MapReduce Total cumulative CPU time: 2 seconds 0 msec
Ended Job = job_1662892505239_0216
MapReduce Jobs Localized on Node:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 10.56 sec HDFS Read: 100633429 HDFS Write: 1677 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.0 sec HDFS Read: 6729 HDFS Write: 56 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 560 msec
OK
violation_code  frequencyViolation
38          10147
14          7609
37          6943
7           5342
20          4311
46          4230
71          3393
Time taken: 40.639 seconds, Fetched: 7 row(s)
hive> █
```

```

OK
violation_code  frequencyViolation
38          10147
14          7609
37          6943
7           5342
20          4311
46          4230
71          3393
Time taken: 40.639 seconds, Fetched: 7 row(s)
hive> █
```

for bin-6:-

```

select Violation_Code,count(*) frequency_Violation from vw_parking_violations_partitioned_2017 where
Violation_Time_bin =6 group by Violation_Code order by frequency_Violation desc limit 7;
```

```

hive> select Violation_Code,count(*) frequencyViolation from vw_parking_violations_partitioned_2017 where Violation_Time_bin =6 group by Violation_Code order by frequencyViolation desc limit 7;
Query ID = cloudera_20221021200808 9cea64ad-f13b-4da6-ha5c-04d8b6a8ce13
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducer.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0217, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0217/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0217
Hadoop job information for Stage-1: number of mappers: 27; number of reducers: 1
2022-10-21 20:08:43,670 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 2.6 sec
2022-10-21 20:08:45,797 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.56 sec
2022-10-21 20:08:50,988 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.06 sec
MapReduce Total cumulative CPU time: 10 seconds 60 msec
Ended Job = job_1662892505239_0217
Launching Job 2 out of 2
Number of reducers determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducer.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0218, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0218/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0218
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 20:08:57,708 Stage-2 map = 0%, reduce = 0%, Cumulative CPU 0.67 sec
2022-10-21 20:09:02,848 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.67 sec
2022-10-21 20:09:09,109 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.69 sec
MapReduce Total cumulative CPU time: 1 seconds 690 msec
Ended Job = job_1662892505239_0218
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 10.06 sec HDFS Read: 100633429 HDFS Write: 1532 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.69 sec HDFS Read: 6584 HDFS Write: 54 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 750 msec
OK
violation_code frequencyViolation
7      2601
40     2159
14     2091
38     2034
20     1495
46     1303
78     936
Time taken: 41.465 seconds, Fetched: 7 row(s)
hive> 
```

```

OK
violation_code  frequencyViolation
7      2601
40     2159
14     2091
38     2034
20     1495
46     1303
78     936
Time taken: 41.465 seconds, Fetched: 7 row(s)
hive> 
```

7.) 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)

```

select Violation_Time_bin, count(*) frequency_times from vw_parking_violations_partitioned_2017
where Violation_Code in (21, 38,36) group by Violation_Time_bin order by frequency_times desc limit 6;
```

```

hive> select Violation_Time_bin, count(*) frequency_times from vw_parking_violations_partitioned_2017 where Violation_Code in (21, 38,36) group by Violation_Time_bin order by frequency_times
desc limit 6;
Query ID = cloudera_20221021201616_a9320127-07fd-4143-9497-97d6da5fb6fa
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job 1: job_1662892505239_0223, Tracking URL: http://quickstart.cloudera:8088/proxy/application_1662892505239_0223/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0223
Hadoop job information for Stage-1: number of mappers: 27 number of reducers: 1
2022-10-21 20:17:05.164 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:17:12.598 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 2.97 sec
2022-10-21 20:17:14.697 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.54 sec
2022-10-21 20:17:18.819 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.69 sec
MapReduce Total cumulative CPU time: 8 seconds 690 msec
Ended Job : job_1662892505239_0223
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job : job_1662892505239_0224, Tracking URL: http://quickstart.cloudera:8088/proxy/application_1662892505239_0224/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0224
Hadoop job information for Stage-2: number of mappers: 17 number of reducers: 1
2022-10-21 20:17:25.617 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:17:33.537 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.83 sec
2022-10-21 20:17:35.023 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.54 sec
MapReduce Total cumulative CPU time: 1 seconds 940 msec
Ended Job : job_1662892505239_0224
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1   Cumulative CPU: 8.69 sec  HDFS Read: 100632195 HDFS Write: 241 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1   Cumulative CPU: 1.54 sec  HDFS Read: 5301 HDFS Write: 46 SUCCESS
Total MapReduce CPU time Spent: 10 seconds 630 msec
OK
violation_time_bin      frequency_times
3           111820
4           59924
5           11492
2           7187
1           3688
6           2051
Time taken: 36.255 seconds, Fetched: 6 row(s)
hive> 
```

```

OK
violation_time_bin      frequency_times
3           111820
4           59924
5           11492
2           7187
1           3688
6           2051
Time taken: 36.255 seconds, Fetched: 6 row(s)
hive> 
```

8.) Let's try and find some seasonality in this data

a.) First, divide the year into some number of seasons, and find frequencies of tickets for each season. (Hint: A quick Google search reveals the following seasons in NYC: Spring(March, April, March); Summer(June, July, August); Fall(September, October, November); Winter(December, January, February))

Seasons -->Month interval

spring -->March, April, May

summer -->June, July, August

autumn -->September, October, November

winter -->December, January, February

=====

```
hive>create view vw_tickets_issued_2017
```

partitioned on (Violation_Code)

as

```

select Issuer_Precinct,
case
when MONTH(Issue_Date) between 03 and 05 then 'spring'
when MONTH(Issue_Date) between 06 and 08 then 'summer'
when MONTH(Issue_Date) between 09 and 11 then 'autumn'
when MONTH(Issue_Date) in (1,2,12) then 'winter'
else 'unknown'
end as season,
Violation_Code
from parking_violations_2017;
```

```

hive> create view vw_tickets_issued_2017
> partitioned on (Violation_Code)
> as
> select Issuer_Precinct,
> case
> when MONTH(Issue_Date) between 03 and 05 then 'spring'
> when MONTH(Issue_Date) between 06 and 08 then 'summer'
> when MONTH(Issue_Date) between 09 and 11 then 'autumn'
> when MONTH(Issue_Date) in (1,2,12) then 'winter'
> else 'unknown'
> end as season,
> Violation_Code
> from parking_violations_2017;
OK
issuer_precinct season violation_code
Time taken: 0.155 seconds
hive> █
```

```

hive> select season, count(*) as frequency from vw_tickets_issued_2017 group by season order by
frequency desc;
```

```

hive> select season, count(*) as frequency from vw_tickets_issued_2017 group by season order by frequency desc;
Query ID = cloudera_20221021203636_5efa5242-3abc-4cac-9977-b3ec69fcfc3
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0225, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0225/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0225
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 20:36:13,220 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:36:21,664 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 2.32 sec
2022-10-21 20:36:22,707 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.22 sec
2022-10-21 20:36:27,862 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.34 sec
MapReduce Total cumulative CPU time: 9 seconds 340 msec
Ended Job = job_1662892505239_0225
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0226, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0226/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0226
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 20:36:33,918 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:36:39,072 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.74 sec
2022-10-21 20:36:45,263 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.85 sec
MapReduce Total cumulative CPU time: 1 seconds 850 msec
Ended Job = job_1662892505239_0226
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 9.34 sec HDFS Read: 100628262 HDFS Write: 180 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.85 sec HDFS Read: 5074 HDFS Write: 41 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 190 msec
OK
season  frequency
spring   285877
winter   169464
summer   84560
Time taken: 38.373 seconds, Fetched: 3 row(s)
hive> 

```

```

OK
season  frequency
spring   285877
winter   169464
summer   84560
Time taken: 38.373 seconds, Fetched: 3 row(s)
hive> 

```

b.) Then, find the 3 most common violations for each of these seasons.

-->**For Spring:-**

```

select Violation_Code, count(*) as frequency from vw_tickets_issued_2017 where season = 'spring'
group by Violation_Code order by frequency desc limit 3;

```

```

hive> select Violation_Code, count(*) as frequency from vw_tickets_issued_2017 where
    > season = 'spring' group by Violation_Code order by frequency desc limit 3;
Query ID = cloudera_20221021203939_c8ebff4c-218d-4e54-b66c-b926e1c9f186
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0227, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0227/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0227
Hadoop job information for Stage-1: number of mappers: 27 number of reducers: 1
2022-10-21 20:39:25,434 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:39:35,130 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 4.85 sec
2022-10-21 20:39:36,180 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 11.7 sec
2022-10-21 20:39:43,602 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 13.89 sec
MapReduce Total cumulative CPU time: 13 seconds 890 msec
Ended Job = job_1662892505239_0227
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0228, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0228/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0228
Hadoop job information for Stage-2: number of mappers: 17 number of reducers: 1
2022-10-21 20:39:50,573 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:39:56,761 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.82 sec
2022-10-21 20:40:01,910 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.97 sec
MapReduce Total cumulative CPU time: 1 seconds 970 msec
Ended Job = job_1662892505239_0228
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1   Cumulative CPU: 13.89 sec   HDFS Read: 100629972 HDFS Write: 1865 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1   Cumulative CPU: 1.97 sec   HDFS Read: 6887 HDFS Write: 27 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 860 msec
OK
violation_code  frequency
21      40043
36      34354
38      27000
Time taken: 44.029 seconds, Fetched: 3 row(s)
hive> █
```

```

OK
violation_code  frequency
21      40043
36      34354
38      27000
Time taken: 44.029 seconds, Fetched: 3 row(s)
hive> █
```

-->*For Winter:-*

```

select Violation_Code, count(*) as frequency from vw_tickets_issued_2017 where season = 'winter'
group by Violation_Code order by frequency desc limit 3;
```

```

hive> select Violation_Code, count(*) as frequency from vw_tickets_issued_2017 where
    > season = 'winter' group by Violation_Code order by frequency desc limit 3;
Query ID = cloudera_20221021204242_3dd3ce28-500d-4462-8895-a5430e92ff9f
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0229, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0229/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0229
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 20:42:14,849 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:42:22,337 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 2.51 sec
2022-10-21 20:42:24,404 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2022-10-21 20:42:27,505 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.62 sec
MapReduce Total cumulative CPU time: 10 seconds 620 msec
Ended Job = job_1662892505239_0229
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0230, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0230/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0230
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 20:42:34,416 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:42:39,567 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.72 sec
2022-10-21 20:42:45,719 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.86 sec
MapReduce Total cumulative CPU time: 1 seconds 860 msec
Ended Job = job_1662892505239_0230
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 10.62 sec HDFS Read: 100629973 HDFS Write: 1812 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.86 sec HDFS Read: 6843 HDFS Write: 27 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 480 msec
OK
violation_code  frequency
21      23683
36      22084
38      18451
Time taken: 38.476 seconds, Fetched: 3 row(s)
hive> 
```

```

OK
violation_code  frequency
21      23683
36      22084
38      18451
Time taken: 38.476 seconds, Fetched: 3 row(s)
hive> 
```

-->*For summer:-*

```

select Violation_Code, count(*) as frequency from vw_tickets_issued_2017 where season = 'summer'
group by Violation_Code order by frequency desc limit 3;
```

```

hive> select Violation_Code, count(*) as frequency from vw_tickets_issued_2017 where
    > season = 'summer' group by Violation_Code order by frequency desc limit 3;
Query ID = cloudera_20221021204444_79e556b3-a729-43e4-b241-6d8d7c21f9cb
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0231, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0231/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0231
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 20:45:04,384 Stage-1 map = 0%,  reduce = 0%
2022-10-21 20:45:13,834 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 3.2 sec
2022-10-21 20:45:14,975 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.2 sec
2022-10-21 20:45:20,063 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.41 sec
MapReduce Total cumulative CPU time: 11 seconds 410 msec
Ended Job = job_1662892505239_0231
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0232, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0232/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0232
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 20:45:27,276 Stage-2 map = 0%,  reduce = 0%
2022-10-21 20:45:32,429 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.78 sec
2022-10-21 20:45:37,655 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.88 sec
MapReduce Total cumulative CPU time: 1 seconds 880 msec
Ended Job = job_1662892505239_0232
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1  Cumulative CPU: 11.41 sec  HDFS Read: 100629973 HDFS Write: 1645 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 1.88 sec  HDFS Read: 6676 HDFS Write: 25 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 290 msec
OK
violation_code  frequency
21      12566
36      9656
38      8331
Time taken: 39.068 seconds, Fetched: 3 row(s)
hive> 
```

```

OK
violation_code  frequency
21      12566
36      9656
38      8331
Time taken: 39.068 seconds, Fetched: 3 row(s)
hive> 
```

-->*For autumn:-*

```

select Violation_Code, count(*) as frequency from vw_tickets_issued_2017 where season = 'autumn'
group by Violation_Code order by frequency desc limit 3;
```

```
hive> select Violation_Code, count(*) as frequency from vw_tickets_issued_2017 where
    > season = 'autumn' group by Violation_Code order by frequency desc limit 3;
Query ID = cloudera_20221021204848_fd21dc7e-419c-4d8a-9c2e-6300dbcfcba5
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0233, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0233/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0233
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-10-21 20:48:35,206 Stage-1 map = 0%, reduce = 0%
2022-10-21 20:48:42,473 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 1.4 sec
2022-10-21 20:48:43,511 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.75 sec
2022-10-21 20:48:46,620 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.9 sec
MapReduce Total cumulative CPU time: 6 seconds 900 msec
Ended Job = job_1662892505239_0233
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1662892505239_0234, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1662892505239_0234/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1662892505239_0234
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-21 20:48:54,065 Stage-2 map = 0%, reduce = 0%
2022-10-21 20:49:59,288 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.82 sec
2022-10-21 20:49:05,469 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 1.73 sec
MapReduce Total cumulative CPU time: 1 seconds 730 msec
Ended Job = job_1662892505239_0234
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 6.9 sec HDFS Read: 100629973 HDFS Write: 96 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.73 sec HDFS Read: 5127 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 630 msec
OK
violation_code frequency
Time taken: 37.067 seconds
hive> █
```

```
OK
violation_code frequency
Time taken: 37.067 seconds
hive> █
```