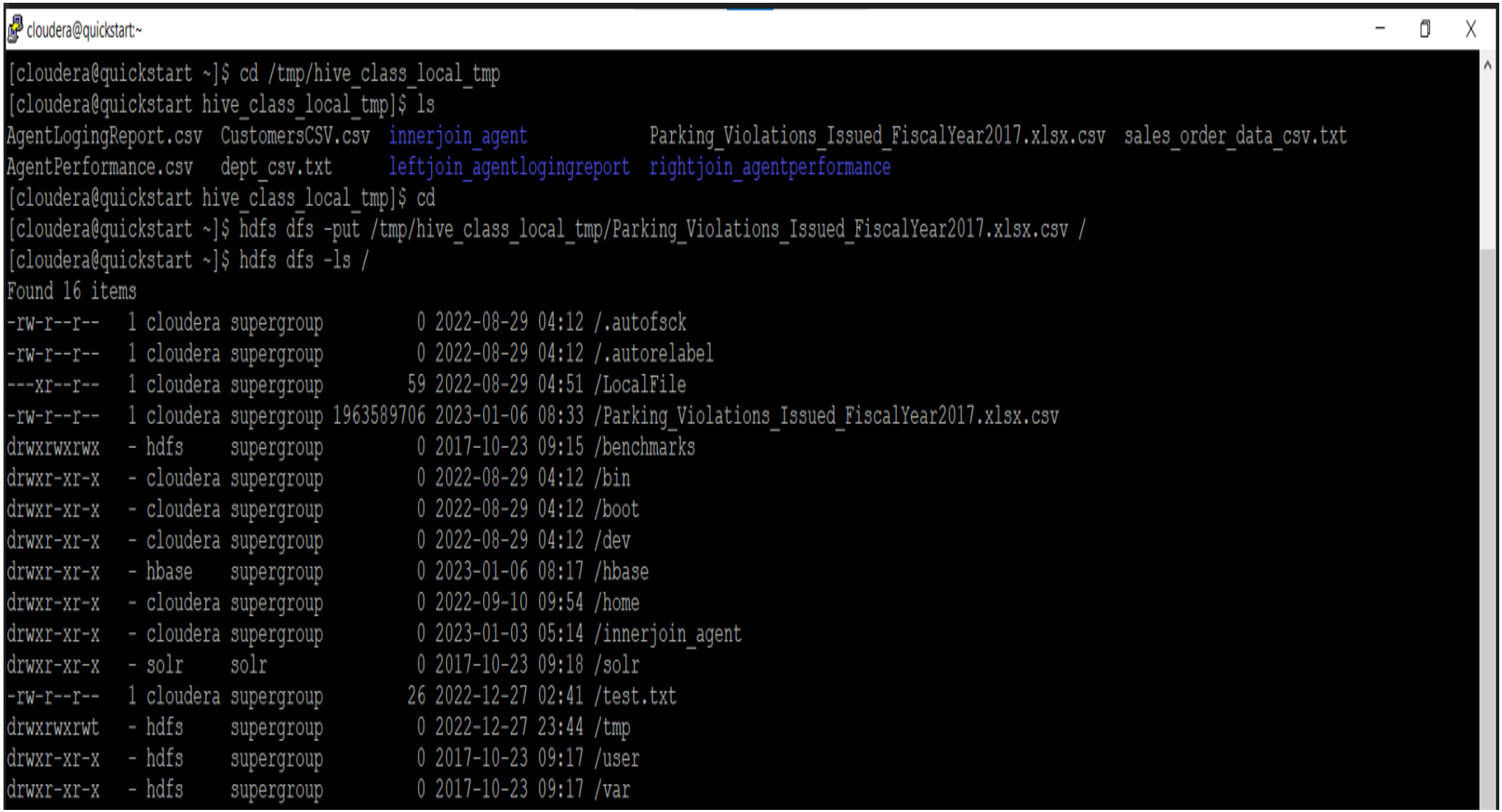
Download Dataset - https://data.cityofnewyork.us/browse?q=parking+tickets

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



cloudera@quickstart:~

```
hive> create table Parking_Violations_Issued_FiscalYear2017
    >     (
    >     summons_number bigint,
    >     plate_id varchar(50),
    >     registration_state varchar(50),
    >     plate_type varchar(50),
    >     issue_date varchar(50),
    >     violation_code int,
    >     vehicle_body_type varchar(50),
    >     vehicle_make varchar(50),
    >     issuing_agency varchar(50),
    >     street_code1 int,
    >     street_code2 int,
    >     street_code3 int,
    >     vehicle_expiration_date int,
    >     violation_location int,
    >     violation_precinct int,
    >     issuer_precinct int,
    >     issuer_code int,
    >     issuer_command varchar(50),
    >     issuer_squad varchar(50),
    >     violation_time timestamp,
    >     time_first_observed timestamp,
    >     violation_country varchar(50),
    >     violation_in_front_of_or_opposite varchar(50),
    >     house_number varchar(50),
    >     street_name varchar(50),
    >     intersecting_street varchar(50),
    >     date_first_observed int,
    >     law_section int,
    >     sub_division varchar(50),
    >     violation_legal_code varchar(50),
    >     days_parking_in_effect varchar(50),
    >     from_hours_in_effect timestamp,
    >     to_hours_in_effect timestamp,
    >     vehicle_color varchar(50),
    >     unregistered_vehicle int,
    >     vehicle_year int,
    >     meter_number varchar(50),
    >     feet_from_curb int,
    >     violation_post_code varchar(50),
    >     violation_description varchar(50),
    >     no_standing_or_stopping_violation varchar(1),
    >     hydrant_violation varchar(1),
    >     double_parking_violation varchar(1)
```

cloudera@quickstart:~

```
    >     street_code3 int,
    >     vehicle_expiration_date int,
    >     violation_location int,
    >     violation_precinct int,
    >     issuer_precinct int,
    >     issuer_code int,
    >     issuer_command varchar(50),
    >     issuer_squad varchar(50),
    >     violation_time timestamp,
    >     time_first_observed timestamp,
    >     violation_country varchar(50),
    >     violation_in_front_of_or_opposite varchar(50),
    >     house_number varchar(50),
    >     street_name varchar(50),
    >     intersecting_street varchar(50),
    >     date_first_observed int,
    >     law_section int,
    >     sub_division varchar(50),
    >     violation_legal_code varchar(50),
    >     days_parking_in_effect varchar(50),
    >     from_hours_in_effect timestamp,
    >     to_hours_in_effect timestamp,
    >     vehicle_color varchar(50),
    >     unregistered_vehicle int,
    >     vehicle_year int,
    >     meter_number varchar(50),
    >     feet_from_curb int,
    >     violation_post_code varchar(50),
    >     violation_description varchar(50),
    >     no_standing_or_stopping_violation varchar(1),
    >     hydrant_violation varchar(1),
    >     double_parking_violation varchar(1)
    >     )
    >     row format serde'org.apache.hadoop.hive.serde2.OpenCSVSerde'
    >     tblproperties ("skip.header.line.count" = "1");
OK
Time taken: 1.456 seconds
hive> load data inpath 'Parking_Violations_Issued_FiscalYear2017.xlsx.csv' into table Parking_Violations_Issued_FiscalYear2017;
FAILED: SemanticException Line 1:17 Invalid path ''Parking_Violations_Issued_FiscalYear2017.xlsx.csv'': No files matching path hdfs://quickstart.cloudera:8020
/user/cloudera/Parking_Violations_Issued_FiscalYear2017.xlsx.csv
hive> load data inpath '/Parking_Violations_Issued_FiscalYear2017.xlsx.csv' into table Parking_Violations_Issued_FiscalYear2017;
Loading data to table default.parking_violations_issued_fiscalyear2017
Table default.parking_violations_issued_fiscalyear2017 stats: [numFiles=1, totalSize=1963589706]
OK
Time taken: 1.535 seconds
```

cloudera@quickstart:~

```
hive> create table Parking_orc
    >     (
    >     summons_number bigint,
    >     plate_id varchar(50),
    >     registration_state varchar(50),
    >     plate_type varchar(50),
    >     issue_date varchar(50),
    >     violation_code int,
    >     vehicle_body_type varchar(50),
    >     vehicle_make varchar(50),
    >     issuing_agency varchar(50),
    >     street_code1 int,
    >     street_code2 int,
    >     street_code3 int,
    >     vehicle_expiration_date int,
    >     violation_location int,
    >     violation_precinct int,
    >     issuer_precinct int,
    >     issuer_code int,
    >     issuer_command varchar(50),
    >     issuer_squad varchar(50),
    >     violation_time timestamp,
    >     time_first_observed timestamp,
    >     violation_country varchar(50),
    >     violation_in_front_of_or_opposite varchar(50),
    >     house_number varchar(50),
    >     street_name varchar(50),
    >     intersecting_street varchar(50),
    >     date_first_observed int,
    >     law_section int,
    >     sub_division varchar(50),
    >     violation_legal_code varchar(50),
    >     days_parking_in_effect varchar(50),
    >     from_hours_in_effect timestamp,
    >     to_hours_in_effect timestamp,
    >     vehicle_color varchar(50),
    >     unregistered_vehicle int,
    >     vehicle_year int,
    >     meter_number varchar(50),
    >     feet_from_curb int,
    >     violation_post_code varchar(50),
    >     violation_description varchar(50),
    >     no_standing_or_stopping_violation varchar(1),
    >     hydrant_violation varchar(1),
    >     double_parking_violation varchar(1)
```

cloudera@quickstart:~

```
    >     street_code3 int,
    >     vehicle_expiration_date int,
    >     violation_location int,
    >     violation_precinct int,
    >     issuer_precinct int,
    >     issuer_code int,
    >     issuer_command varchar(50),
    >     issuer_squad varchar(50),
    >     violation_time timestamp,
    >     time_first_observed timestamp,
    >     violation_country varchar(50),
    >     violation_in_front_of_or_opposite varchar(50),
    >     house_number varchar(50),
    >     street_name varchar(50),
    >     intersecting_street varchar(50),
    >     date_first_observed int,
    >     law_section int,
    >     sub_division varchar(50),
    >     violation_legal_code varchar(50),
    >     days_parking_in_effect varchar(50),
    >     from_hours_in_effect timestamp,
    >     to_hours_in_effect timestamp,
    >     vehicle_color varchar(50),
    >     unregistered_vehicle int,
    >     vehicle_year int,
    >     meter_number varchar(50),
    >     feet_from_curb int,
    >     violation_post_code varchar(50),
    >     violation_description varchar(50),
    >     no_standing_or_stopping_violation varchar(1),
    >     hydrant_violation varchar(1),
    >     double_parking_violation varchar(1)
    >     )
    > stored as orc;
OK
Time taken: 0.442 seconds
hive> insert overwrite table Parking_orc select * from Parking_Violations_Issued_FiscalYear2017 where year(to_date(from_unixtime(unix_timestamp(issue_date, 'M
M/dd/yyyy')))) = '2017';
Query ID = cloudera_20230106092727_4ffc37a8-7014-48cd-8f05-4f9916c5a098
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1673021823550_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673021823550_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673021823550_0003
Hadoop job information for Stage-1: number of mappers: 8; number of reducers: 0
```

cloudera@quickstart:~

```
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/parking_orc/.hive-staging_hive_2023-01-06_09-27-04_692_6530322159604742300-1/-ext-10000
Loading data to table default.parking_orc
Table default.parking_orc stats: [numFiles=8, numRows=4984543, totalSize=151927047, rawDataSize=10677518377]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 8   Cumulative CPU: 3237.02 sec   HDFS Read: 1963760778 HDFS Write: 151927742 SUCCESS
Total MapReduce CPU Time Spent: 53 minutes 57 seconds 20 msec
OK
Time taken: 1469.61 seconds
hive> SELECT issue_date FROM Parking_orc LIMIT 5;
OK
03/09/2017
01/18/2017
03/02/2017
03/06/2017
05/04/2017
Time taken: 1.098 seconds, Fetched: 5 row(s)
hive> set hive.cli.print.header = true;
hive> select * from Parking_orc LIMIT 5;
OK
parking_orc.summons_number      parking_orc.plate_id    parking_orc.registration_state  parking_orc.plate_type  parking_orc.issue_date  parking_orc.violation_
code    parking_orc.vehicle_body_type   parking_orc.vehicle_make        parking_orc.issuing_agency      parking_orc.street_code1        parking_orc.street_cod
e2      parking_orc.street_code3        parking_orc.vehicle_expiration_date     parking_orc.violation_location  parking_orc.violation_precinct  parking_orc.is
suer_precinct   parking_orc.issuer_code parking_orc.issuer_command      parking_orc.issuer_squad        parking_orc.violation_time       parking_orc.time_first
_observed       parking_orc.violation_country   parking_orc.violation_in_front_of_or_opposite   parking_orc.house_number        parking_orc.street_name parkin
g_orc.intersecting_street       parking_orc.date_first_observed parking_orc.law_section parking_orc.sub_division        parking_orc.violation_legal_code     p
arking_orc.days_parking_in_effect       parking_orc.from_hours_in_effect        parking_orc.to_hours_in_effect  parking_orc.vehicle_color       parking_orc.un
registered_vehicle      parking_orc.vehicle_year        parking_orc.meter_number        parking_orc.feet_from_curb      parking_orc.violation_post_code parkin
g_orc.violation_description     parking_orc.no_standing_or_stopping_violation   parking_orc.hydrant_violation   parking_orc.double_parking_violation
4631633384      AVM7975 NY      PAS     03/09/2017      36      SUBN    GMC     V       0       0       0       0       NULL    0       0       0            N
ULL     NULL    BK                      WB LINDEN BLVD @ LIN    COLN AVE        0       NULL    B       T               NULL    NULL    GY      NULL    2010 0
PHTO SCHOOL ZN SPEED VIOLATION
8196557280      GWB7054 NY      PAS     01/18/2017      70      SUBN    TOYOT   T       NULL    NULL    NULL    20170105        109     109     109     NULL T
401     J       NULL    NULL    Q       F       35-11   Prince St               0       408     j3              YYYYYYY NULL    NULL    BL      NULL    2015 0
05      70A-Reg. Sticker Expired (NYS)
4631184358      EXZ9820 NY      PAS     03/02/2017      36      4DSD    HONDA   V       0       0       0       0       NULL    0       0       0            N
ULL     NULL    BK                      WB FLATLANDS AVE @ E    100 ST  0       NULL    B       T               NULL    NULL    GR      NULL    1997         0
PHTO SCHOOL ZN SPEED VIOLATION
4007039033      GZE1511 NY      PAS     03/06/2017      5       4DSD    TOYOT   V       0       0       0       0       NULL    0       0       0            N
ULL     NULL    BK                      NB UTICA AVE @ CHURC    H AVE   0       NULL    C       T               NULL    NULL    WH      NULL    2001         0
BUS LANE VIOLATION
8539360652      GEH9367 NY      PAS     05/04/2017      70      4DSD    DODGE   T       NULL    NULL    NULL    20170325        78      78      78      NULL T
301     E       NULL    NULL    K       F       433     Dean St         0       408     j3              YYYYYYY NULL    NULL    GY      NULL    2011         0
19      70A-Reg. Sticker Expired (NYS)
Time taken: 0.247 seconds, Fetched: 5 row(s)
hive>
```

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

```
hive> SELECT COUNT(summons_number) AS Total_tickets FROM parking_orc;
Query ID = cloudera_20230107021111_fa4c5d04-aa8f-4cfa-a009-5886413cd1ea
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_1673082023969_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-07 02:11:42,984 Stage-1 map = 0%,  reduce = 0%
2023-01-07 02:11:51,324 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.26 sec
2023-01-07 02:11:58,648 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.17 sec
MapReduce Total cumulative CPU time: 5 seconds 170 msec
Ended Job = job_1673082023969_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.17 sec   HDFS Read: 21419303 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 170 msec
OK
total_tickets
4984543
Time taken: 26.325 seconds, Fetched: 1 row(s)
hive>
```

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

cloudera@quickstart:~

```
hive> SELECT registration_state,COUNT(summons_number) AS Total_tickets FROM parking_orc GROUP BY registration_state ORDER BY Total_tickets DESC;
Query ID = cloudera_20230107021414_d51c171f-30b2-4c87-b9f7-687a4c2fe6a1
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_1673082023969_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-07 02:14:37,571 Stage-1 map = 0%,  reduce = 0%
2023-01-07 02:14:48,018 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.36 sec
2023-01-07 02:14:55,316 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.18 sec
MapReduce Total cumulative CPU time: 7 seconds 180 msec
Ended Job = job_1673082023969_0005
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_1673082023969_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-07 02:15:05,366 Stage-2 map = 0%,  reduce = 0%
2023-01-07 02:15:11,650 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.28 sec
2023-01-07 02:15:18,945 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 3.15 sec
MapReduce Total cumulative CPU time: 3 seconds 150 msec
Ended Job = job_1673082023969_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.18 sec   HDFS Read: 23162458 HDFS Write: 1575 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.15 sec   HDFS Read: 6739 HDFS Write: 508 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 330 msec
OK
registration_state      total_tickets
NY      3922010
NJ      436618
PA      128911
CT      64656
```

cloudera@quickstart:~

```
CT      64656
FL      63767
IN      41742
MA      35850
VA      31431
MD      27662
NC      24856
TX      17290
IL      17179
GA      16068
99      14650
AZ      11321
OH      11279
CA      11128
ME      9932
SC      9527
MN      9271
OK      8350
TN      7838
DE      7263
MI      6606
RI      5326
NH      3762
VT      3379
AL      2903
WA      2778
OR      2381
MO      2272
ON      2245
WI      1954
QB      1833
DC      1779
IA      1779
CO      1694
DP      1657
KY      1644
LA      1563
MS      1447
WV      1156
AR      909
SD      807
NM      724
ID      691
NV      663
KS      640
```

3.)

cloudera@quickstart:~

```
DE      7263
MI      6606
RI      5326
NH      3762
VT      3379
AL      2903
WA      2778
OR      2381
MO      2272
ON      2245
WI      1954
QB      1833
DC      1779
IA      1779
CO      1694
DP      1657
KY      1644
LA      1563
MS      1447
WV      1156
AR      909
SD      807
NM      724
ID      691
NV      663
KS      640
NE      633
UT      508
MT      456
GV      315
NS      308
AK      273
ND      239
WY      175
HI      145
AB      72
PE      59
NB      53
BC      50
PR      37
MB      15
SK      7
FO      7
Time taken: 49.789 seconds, Fetched: 65 row(s)
hive>
```

4.) 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(summons_number) AS Total_tickets FROM parking_orc WHERE street_code1 = 0 OR street_code2 = 0 OR street_code3 = 0;
Query ID = cloudera_20230107021919_7fef53ca-13b3-4b29-90d0-c3b46b265c8f
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_1673082023969_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-07 02:19:24,492 Stage-1 map = 0%,  reduce = 0%
2023-01-07 02:19:34,914 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.49 sec
2023-01-07 02:19:42,249 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.56 sec
MapReduce Total cumulative CPU time: 6 seconds 560 msec
Ended Job = job_1673082023969_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.56 sec   HDFS Read: 23136621 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 560 msec
OK
total_tickets
1667289
Time taken: 27.177 seconds, Fetched: 1 row(s)
hive>
```

Part-II: Aggregation tasks

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

cloudera@quickstart:~

```
hive> SELECT violation_code,COUNT(violation_code) AS Violation_frequency FROM parking_orc GROUP BY violation_code ORDER BY Violation_frequency DESC LIMIT 5;
Query ID = cloudera_20230107022222_077042ac-5bf6-456a-8837-fc83c29db408
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_1673082023969_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-07 02:22:50,436 Stage-1 map = 0%,  reduce = 0%
2023-01-07 02:22:57,796 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.51 sec
2023-01-07 02:23:06,135 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.55 sec
MapReduce Total cumulative CPU time: 5 seconds 550 msec
Ended Job = job_1673082023969_0008
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_1673082023969_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0009
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-07 02:23:15,006 Stage-2 map = 0%,  reduce = 0%
2023-01-07 02:23:21,281 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.3 sec
2023-01-07 02:23:28,579 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 3.22 sec
MapReduce Total cumulative CPU time: 3 seconds 220 msec
Ended Job = job_1673082023969_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.55 sec   HDFS Read: 3419509 HDFS Write: 2170 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.22 sec   HDFS Read: 7279 HDFS Write: 50 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 770 msec
OK
violation_code  violation_frequency
21      704742
36      608664
38      497732
14      437184
```

```
violation_code  violation_frequency
21      704742
36      608664
38      497732
14      437184
20      293405
Time taken: 47.895 seconds, Fetched: 5 row(s)
hive>
```

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

cloudera@quickstart:~

```
hive> SELECT vehicle_body_type,vehicle_make,COUNT(summons_number) AS Total_tickets FROM parking_orc GROUP BY vehicle_body_type,vehicle_make ORDER BY Total_tic
kets DESC LIMIT 5;
Query ID = cloudera_20230107022727_8e44813e-4a0a-44e9-a4a1-a1a90f76dafa
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_1673082023969_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-07 02:27:15,017 Stage-1 map = 0%,  reduce = 0%
2023-01-07 02:27:27,632 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 8.57 sec
2023-01-07 02:27:35,996 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 12.39 sec
MapReduce Total cumulative CPU time: 12 seconds 390 msec
Ended Job = job_1673082023969_0010
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_1673082023969_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0011
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-07 02:27:44,489 Stage-2 map = 0%,  reduce = 0%
2023-01-07 02:27:51,798 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 3.25 sec
2023-01-07 02:27:59,120 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 5.55 sec
MapReduce Total cumulative CPU time: 5 seconds 550 msec
Ended Job = job_1673082023969_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 12.39 sec   HDFS Read: 29248788 HDFS Write: 275940 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 5.55 sec   HDFS Read: 281625 HDFS Write: 88 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 940 msec
OK
vehicle_body_type       vehicle_make    total_tickets
4DSD    TOYOT   250340
VAN     FORD    245754
SUBN    TOYOT   213774
```

```
vehicle_body_type       vehicle_make    total_tickets
4DSD    TOYOT   250340
VAN     FORD    245754
SUBN    TOYOT   213774
4DSD    HONDA   211910
SUBN    HONDA   207985
Time taken: 53.044 seconds, Fetched: 5 row(s)
hive>
```

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)

cloudera@quickstart:~

```
hive> SELECT violation_precinct,COUNT(summons_number) AS Total_tickets FROM parking_orc GROUP BY violation_precinct ORDER BY Total_tickets DESC LIMIT 5;
Query ID = cloudera_20230107033434_76864e56-1655-42ec-b802-bf026960bc22
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_1673082023969_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-07 03:34:30,990 Stage-1 map = 0%,  reduce = 0%
2023-01-07 03:34:42,926 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.26 sec
2023-01-07 03:34:51,300 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.2 sec
MapReduce Total cumulative CPU time: 7 seconds 200 msec
Ended Job = job_1673082023969_0018
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_1673082023969_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0019
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-07 03:35:03,569 Stage-2 map = 0%,  reduce = 0%
2023-01-07 03:35:11,149 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.37 sec
2023-01-07 03:35:20,542 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.31 sec
MapReduce Total cumulative CPU time: 4 seconds 310 msec
Ended Job = job_1673082023969_0019
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.2 sec   HDFS Read: 25265656 HDFS Write: 3532 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.31 sec   HDFS Read: 8645 HDFS Write: 48 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 510 msec
OK
violation_precinct      total_tickets
0       849858
19      251820
14      186905
1       160269
```

```
OK
violation_precinct      total_tickets
0       849858
19      251820
14      186905
1       160269
18      155280
Time taken: 63.996 seconds, Fetched: 5 row(s)
hive>
```

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

cloudera@quickstart:~

```
hive> SELECT issuer_precinct,COUNT(summons_number) AS Total_tickets FROM parking_orc GROUP BY issuer_precinct ORDER BY Total_tickets DESC LIMIT 5;
Query ID = cloudera_20230107033838_f09f2a3e-e18b-4b10-b08e-84858d286119
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_1673082023969_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-07 03:38:14,735 Stage-1 map = 0%,  reduce = 0%
2023-01-07 03:38:25,732 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.31 sec
2023-01-07 03:38:35,125 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.33 sec
MapReduce Total cumulative CPU time: 7 seconds 330 msec
Ended Job = job_1673082023969_0020
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_1673082023969_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0021
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-07 03:38:46,123 Stage-2 map = 0%,  reduce = 0%
2023-01-07 03:38:53,514 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.38 sec
2023-01-07 03:39:02,000 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.51 sec
MapReduce Total cumulative CPU time: 4 seconds 510 msec
Ended Job = job_1673082023969_0021
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.33 sec   HDFS Read: 25805348 HDFS Write: 10450 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.51 sec   HDFS Read: 15551 HDFS Write: 48 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 840 msec
OK
issuer_precinct total_tickets
0       989729
19      244973
14      184141
1       154808
```

```
OK
issuer_precinct total_tickets
0       989729
19      244973
14      184141
1       154808
18      149655
Time taken: 59.193 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?

cloudera@quickstart:~

```
hive> SELECT violation_precinct,issuer_precinct,violation_code,COUNT(summons_number) AS Total_tickets FROM parking_orc GROUP BY violation_precinct,issuer_prec
inct,violation_code ORDER BY Total_tickets DESC LIMIT 5;
Query ID = cloudera_20230107034545_e253451c-2479-42ba-a39a-4e76bb041e3a
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_1673082023969_0022, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0022/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0022
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-07 03:45:22,078 Stage-1 map = 0%,  reduce = 0%
2023-01-07 03:45:35,048 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.87 sec
2023-01-07 03:45:43,408 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.62 sec
MapReduce Total cumulative CPU time: 10 seconds 620 msec
Ended Job = job_1673082023969_0022
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_1673082023969_0023, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673082023969_0023/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673082023969_0023
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-07 03:45:55,922 Stage-2 map = 0%,  reduce = 0%
2023-01-07 03:46:05,412 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 4.8 sec
2023-01-07 03:46:14,814 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 7.28 sec
MapReduce Total cumulative CPU time: 7 seconds 280 msec
Ended Job = job_1673082023969_0023
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.62 sec   HDFS Read: 32913133 HDFS Write: 324680 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 7.28 sec   HDFS Read: 330523 HDFS Write: 69 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 900 msec
OK
violation_precinct      issuer_precinct violation_code  total_tickets
0       0       36      608664
0       0       7       192861
18      18      14      46049
```

```
violation_precinct      issuer_precinct violation_code  total_tickets
0       0       36      608664
0       0       7       192861
18      18      14      46049
19      19      46      44381
0       0       5       44073
Time taken: 64.916 seconds, Fetched: 5 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))

```
hive> create view Parking_season as Select summons_number,violation_code, issue_date, case
    > when substring(issue_Date,1,2) in ('12','01','02') then 'winter'
    > when substring(issue_Date,1,2) in ('03','04','05') then 'spring'
    > when substring(issue_Date,1,2) in ('06','07','08') then 'summer'
    > when substring(issue_Date,1,2) in ('09','10','11') then 'fall'
    > else 'not_defined' end as seasons
    > from parking_orc;
OK
summons_number  violation_code  issue_date      seasons
Time taken: 0.507 seconds
hive>
```

cloudera@quickstart:~

```
hive> SELECT seasons,COUNT(summons_number) as Frequencie_tickets FROM Parking_season GROUP BY seasons ORDER BY Frequencie_tickets DESC;
Query ID = cloudera_20230110024444_24979de8-8f6a-4746-8b24-3d1ded4b281e
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_1673330787972_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-10 02:45:09,925 Stage-1 map = 0%,  reduce = 0%
2023-01-10 02:45:41,987 Stage-1 map = 27%,  reduce = 0%, Cumulative CPU 18.69 sec
2023-01-10 02:45:47,494 Stage-1 map = 46%,  reduce = 0%, Cumulative CPU 24.89 sec
2023-01-10 02:45:51,903 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 28.86 sec
2023-01-10 02:46:06,444 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 33.01 sec
MapReduce Total cumulative CPU time: 33 seconds 10 msec
Ended Job = job_1673330787972_0001
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_1673330787972_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0002
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-10 02:46:26,265 Stage-2 map = 0%,  reduce = 0%
2023-01-10 02:46:41,308 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 5.67 sec
2023-01-10 02:46:55,580 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 9.89 sec
MapReduce Total cumulative CPU time: 9 seconds 890 msec
Ended Job = job_1673330787972_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 33.01 sec   HDFS Read: 26045899 HDFS Write: 180 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 9.89 sec   HDFS Read: 5156 HDFS Write: 44 SUCCESS
Total MapReduce CPU Time Spent: 42 seconds 900 msec
OK
seasons frequencie_tickets
spring  2637380
winter  1564476
```

```
OK
seasons frequencie_tickets
spring  2637380
winter  1564476
summer  782687
Time taken: 153.984 seconds, Fetched: 3 row(s)
hive>
```

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

cloudera@quickstart:~

```
hive> SELECT violation_code,COUNT(summons_number) AS Total_Violations FROM Parking_season WHERE seasons = 'winter' GROUP BY violation_code ORDER BY Total_Viol
ations DESC LIMIT 3;
Query ID = cloudera_20230110025656_eed9ad48-45f8-4bcd-bcba-afac76f84fa1
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_1673330787972_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-10 02:56:28,461 Stage-1 map = 0%,  reduce = 0%
2023-01-10 02:56:59,390 Stage-1 map = 36%,  reduce = 0%, Cumulative CPU 19.47 sec
2023-01-10 02:57:05,964 Stage-1 map = 55%,  reduce = 0%, Cumulative CPU 25.65 sec
2023-01-10 02:57:07,046 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 27.14 sec
2023-01-10 02:57:21,294 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 31.44 sec
MapReduce Total cumulative CPU time: 31 seconds 440 msec
Ended Job = job_1673330787972_0003
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_1673330787972_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0004
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-10 02:57:38,833 Stage-2 map = 0%,  reduce = 0%
2023-01-10 02:57:51,027 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 3.6 sec
2023-01-10 02:58:03,144 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 7.84 sec
MapReduce Total cumulative CPU time: 7 seconds 840 msec
Ended Job = job_1673330787972_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 31.44 sec   HDFS Read: 29307498 HDFS Write: 2089 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 7.84 sec   HDFS Read: 7194 HDFS Write: 30 SUCCESS
Total MapReduce CPU Time Spent: 39 seconds 280 msec
OK
violation_code  total_violations
21      218685
```

```
violation_code  total_violations
21      218685
36      203226
38      172146
Time taken: 111.505 seconds, Fetched: 3 row(s)
hive>
```

cloudera@quickstart:~

```
hive> SELECT violation_code,COUNT(summons_number) AS Total_Violations FROM Parking_season WHERE seasons = 'spring' GROUP BY violation_code ORDER BY Total_Viol
ations DESC LIMIT 3;
Query ID = cloudera_20230110025959_a6b52d91-7b09-479a-ac72-f5b388092983
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_1673330787972_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-10 02:59:44,893 Stage-1 map = 0%,  reduce = 0%
2023-01-10 03:00:11,974 Stage-1 map = 27%,  reduce = 0%, Cumulative CPU 20.26 sec
2023-01-10 03:00:18,540 Stage-1 map = 46%,  reduce = 0%, Cumulative CPU 26.51 sec
2023-01-10 03:00:21,828 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 30.39 sec
2023-01-10 03:00:37,160 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 34.61 sec
MapReduce Total cumulative CPU time: 34 seconds 610 msec
Ended Job = job_1673330787972_0005
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_1673330787972_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-10 03:00:52,797 Stage-2 map = 0%,  reduce = 0%
2023-01-10 03:01:03,686 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.64 sec
2023-01-10 03:01:16,987 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 6.94 sec
MapReduce Total cumulative CPU time: 6 seconds 940 msec
Ended Job = job_1673330787972_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 34.61 sec   HDFS Read: 29307498 HDFS Write: 2159 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 6.94 sec   HDFS Read: 7264 HDFS Write: 30 SUCCESS
Total MapReduce CPU Time Spent: 41 seconds 550 msec
OK
violation_code  total_violations
21      369279
```

```
violation_code  total_violations
21      369279
36      316643
38      248857
Time taken: 108.447 seconds, Fetched: 3 row(s)
hive>
```

cloudera@quickstart:~

```
hive> SELECT violation_code,COUNT(summons_number) AS Total_Violations FROM Parking_season WHERE seasons = 'summer' GROUP BY violation_code ORDER BY Total_Viol
ations DESC LIMIT 3;
Query ID = cloudera_20230110030303_0c81b464-fa8e-495b-9c06-93e8bf40aa50
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_1673330787972_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-10 03:03:19,007 Stage-1 map = 0%,  reduce = 0%
2023-01-10 03:03:46,336 Stage-1 map = 18%,  reduce = 0%, Cumulative CPU 26.59 sec
2023-01-10 03:03:51,987 Stage-1 map = 46%,  reduce = 0%, Cumulative CPU 32.79 sec
2023-01-10 03:03:57,441 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 37.6 sec
2023-01-10 03:04:11,722 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 42.16 sec
MapReduce Total cumulative CPU time: 42 seconds 160 msec
Ended Job = job_1673330787972_0007
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_1673330787972_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0008
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-10 03:04:27,435 Stage-2 map = 0%,  reduce = 0%
2023-01-10 03:04:39,413 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.67 sec
2023-01-10 03:04:52,750 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 6.81 sec
MapReduce Total cumulative CPU time: 6 seconds 810 msec
Ended Job = job_1673330787972_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 42.16 sec   HDFS Read: 29307498 HDFS Write: 1941 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 6.81 sec   HDFS Read: 7046 HDFS Write: 28 SUCCESS
Total MapReduce CPU Time Spent: 48 seconds 970 msec
OK
violation_code  total_violations
21      116778
```

```
violation_code  total_violations
21      116778
36      88795
38      76729
Time taken: 110.424 seconds, Fetched: 3 row(s)
hive>
```

cloudera@quickstart:~

```
hive> SELECT violation_code,COUNT(summons_number) AS Total_Violations FROM Parking_season WHERE seasons = 'fall' GROUP BY violation_code ORDER BY Total_Violat
ions DESC LIMIT 3;
Query ID = cloudera_20230110030606_45754a11-2683-4042-9da0-d18d3325b35a
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_1673330787972_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-10 03:06:16,231 Stage-1 map = 0%,  reduce = 0%
2023-01-10 03:06:36,385 Stage-1 map = 18%,  reduce = 0%, Cumulative CPU 15.52 sec
2023-01-10 03:06:42,930 Stage-1 map = 46%,  reduce = 0%, Cumulative CPU 21.76 sec
2023-01-10 03:06:47,287 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 25.94 sec
2023-01-10 03:07:01,558 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 30.22 sec
MapReduce Total cumulative CPU time: 30 seconds 220 msec
Ended Job = job_1673330787972_0009
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_1673330787972_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0010
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-10 03:07:18,257 Stage-2 map = 0%,  reduce = 0%
2023-01-10 03:07:29,285 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.64 sec
2023-01-10 03:07:42,399 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 5.82 sec
MapReduce Total cumulative CPU time: 5 seconds 820 msec
Ended Job = job_1673330787972_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 30.22 sec   HDFS Read: 29307494 HDFS Write: 96 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 5.82 sec   HDFS Read: 5201 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 36 seconds 40 msec
OK
violation_code  total_violations
Time taken: 101.545 seconds
```

```
Total MapReduce CPU Time Spent: 50 seconds 40 msec
OK
violation_code  total_violations
Time taken: 101.545 seconds
hive>
```

PARTITIONING

cloudera@quickstart:~

```
hive> set hive.exec.dynamic.partition.mode = nonstrict;
hive> CREATE TABLE ParkingSeason_Partition
    > (
    > summons_numbers BIGINT,
    > violation_code INT,
    > issue_date VARCHAR(50) )
    > PARTITIONED BY (seasons VARCHAR(50))
    > STORED AS ORC
    > ;
OK
Time taken: 0.167 seconds
hive> INSERT OVERWRITE TABLE ParkingSeason_Partition partition (seasons) SELECT * FROM Parking_Season;
Query ID = cloudera_20230110034444_22506868-725c-4626-a5d2-d1d4867ea92c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1673330787972_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1673330787972_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1673330787972_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-01-10 03:44:50,506 Stage-1 map = 0%,  reduce = 0%
2023-01-10 03:45:15,912 Stage-1 map = 14%,  reduce = 0%, Cumulative CPU 14.31 sec
2023-01-10 03:45:22,436 Stage-1 map = 27%,  reduce = 0%, Cumulative CPU 20.85 sec
2023-01-10 03:45:28,983 Stage-1 map = 41%,  reduce = 0%, Cumulative CPU 27.16 sec
2023-01-10 03:45:34,433 Stage-1 map = 55%,  reduce = 0%, Cumulative CPU 33.32 sec
2023-01-10 03:45:40,979 Stage-1 map = 82%,  reduce = 0%, Cumulative CPU 39.42 sec
2023-01-10 03:45:46,571 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 46.72 sec
MapReduce Total cumulative CPU time: 47 seconds 940 msec
Ended Job = job_1673330787972_0011
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/hive_class_b1.db/parkingseason_partition/.hive-staging_hive_2023-01-10_03-44-34_717_387786
2982639072067-1/-ext-10000
Loading data to table hive_class_b1.parkingseason_partition partition (seasons=null)
         Time taken for load dynamic partitions : 1450
        Loading partition {seasons=spring}
        Loading partition {seasons=summer}
        Loading partition {seasons=winter}
         Time taken for adding to write entity : 11
Partition hive_class_b1.parkingseason_partition{seasons=spring} stats: [numFiles=1, numRows=2637380, totalSize=15004423, rawDataSize=279562280]
Partition hive_class_b1.parkingseason_partition{seasons=summer} stats: [numFiles=1, numRows=782687, totalSize=4268315, rawDataSize=82964822]
Partition hive_class_b1.parkingseason_partition{seasons=winter} stats: [numFiles=1, numRows=1564476, totalSize=8790743, rawDataSize=165834456]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 47.94 sec   HDFS Read: 29300877 HDFS Write: 28063765 SUCCESS
Total MapReduce CPU Time Spent: 47 seconds 940 msec
```

```
Stage-Stage-1: Map: 1   Cumulative CPU: 47.94 sec   HDFS Read: 29300877 HDFS Write: 28063765 SUCCESS
Total MapReduce CPU Time Spent: 47 seconds 940 msec
OK
parking_season.summons_number   parking_season.violation_code   parking_season.issue_date      parking_season.seasons
Time taken: 78.059 seconds
hive>
```

```
hive> SELECT * FROM ParkingSeason_Partition LIMIT 20;
OK
parkingseason_partition.summons_numbers parkingseason_partition.violation_code  parkingseason_partition.issue_date     parkingseason_partition.seasons
4631633384      36      03/09/2017      spring
4631184358      36      03/02/2017      spring
4007039033      5       03/06/2017      spring
8539360652      70      05/04/2017      spring
8525962235      21      04/20/2017      spring
8520357982      21      05/04/2017      spring
4634755634      36      05/31/2017      spring
8483877818      47      04/20/2017      spring
8517065839      46      05/12/2017      spring
8533365240      46      04/07/2017      spring
8564954023      14      05/31/2017      spring
8490425115      21      04/04/2017      spring
4633847594      36      05/15/2017      spring
8523152696      71      03/07/2017      spring
8091522538      21      03/31/2017      spring
1420174393      40      04/17/2017      spring
8527308990      70      03/23/2017      spring
8525320055      21      05/19/2017      spring
8521616776      20      05/08/2017      spring
8509673860      14      05/21/2017      spring
Time taken: 0.253 seconds, Fetched: 20 row(s)
hive>
```

cloudera@quickstart:~

```
[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/warehouse/hive_class_b1.db
Found 37 items
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 01:24 /user/hive/warehouse/hive_class_b1.db/agentlogingreport
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 09:03 /user/hive/warehouse/hive_class_b1.db/agentlogingreport_dynamic_part
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 01:29 /user/hive/warehouse/hive_class_b1.db/agentperformance
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 09:16 /user/hive/warehouse/hive_class_b1.db/agentperformance_dynamic_part
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 09:07 /user/hive/warehouse/hive_class_b1.db/buck_agentlogingreport_dynamic_part
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 09:18 /user/hive/warehouse/hive_class_b1.db/buck_agentperformance_dynamic_part
drwxrwxrwx   - cloudera supergroup          0 2022-09-20 00:23 /user/hive/warehouse/hive_class_b1.db/buck_locations
drwxrwxrwx   - cloudera supergroup          0 2022-09-20 00:20 /user/hive/warehouse/hive_class_b1.db/buck_users
drwxrwxrwx   - cloudera supergroup          0 2022-09-15 02:23 /user/hive/warehouse/hive_class_b1.db/covid_full_grouped
drwxrwxrwx   - cloudera supergroup          0 2022-09-15 02:40 /user/hive/warehouse/hive_class_b1.db/covid_full_grouped_orc
drwxrwxrwx   - cloudera supergroup          0 2022-09-15 03:10 /user/hive/warehouse/hive_class_b1.db/covidfullgrouped_dynamicpart
drwxrwxrwx   - cloudera supergroup          0 2022-09-13 00:34 /user/hive/warehouse/hive_class_b1.db/csv_table
drwxrwxrwx   - cloudera supergroup          0 2022-12-29 03:22 /user/hive/warehouse/hive_class_b1.db/customer_bucketing
drwxrwxrwx   - cloudera supergroup          0 2022-12-29 02:17 /user/hive/warehouse/hive_class_b1.db/customer_dynamic_part
drwxrwxrwx   - cloudera supergroup          0 2022-12-29 01:41 /user/hive/warehouse/hive_class_b1.db/customer_parquet
drwxrwxrwx   - cloudera supergroup          0 2022-12-29 03:39 /user/hive/warehouse/hive_class_b1.db/customer_partition_bucketing
drwxrwxrwx   - cloudera supergroup          0 2022-12-29 02:03 /user/hive/warehouse/hive_class_b1.db/customer_static_part
drwxrwxrwx   - cloudera supergroup          0 2022-12-29 01:33 /user/hive/warehouse/hive_class_b1.db/customers
drwxrwxrwx   - cloudera supergroup          0 2022-09-06 03:16 /user/hive/warehouse/hive_class_b1.db/department_data
drwxrwxrwx   - cloudera supergroup          0 2022-09-08 02:52 /user/hive/warehouse/hive_class_b1.db/department_data_external
drwxrwxrwx   - cloudera supergroup          0 2022-09-06 04:08 /user/hive/warehouse/hive_class_b1.db/department_data_hdfs
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 04:48 /user/hive/warehouse/hive_class_b1.db/innerjoin_agent
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 04:54 /user/hive/warehouse/hive_class_b1.db/leftjoin_agentlogingreport
drwxrwxrwx   - cloudera supergroup          0 2022-09-20 00:11 /user/hive/warehouse/hive_class_b1.db/locations
drwxrwxrwx   - cloudera supergroup          0 2023-01-07 02:09 /user/hive/warehouse/hive_class_b1.db/parking_orc
drwxrwxrwx   - cloudera supergroup          0 2023-01-07 05:38 /user/hive/warehouse/hive_class_b1.db/parking_orc1
drwxrwxrwx   - cloudera supergroup          0 2023-01-05 04:57 /user/hive/warehouse/hive_class_b1.db/parking_violations_issued_fiscalyear2017
drwxrwxrwx   - cloudera supergroup          0 2023-01-10 03:45 /user/hive/warehouse/hive_class_b1.db/parkingseason_partition
drwxrwxrwx   - cloudera supergroup          0 2023-01-03 05:06 /user/hive/warehouse/hive_class_b1.db/rightjoin_agentperformance
drwxrwxrwx   - cloudera supergroup          0 2022-09-12 23:38 /user/hive/warehouse/hive_class_b1.db/sales_data
drwxrwxrwx   - cloudera supergroup          0 2022-09-12 23:41 /user/hive/warehouse/hive_class_b1.db/sales_data_bkup
drwxrwxrwx   - cloudera supergroup          0 2022-09-12 23:47 /user/hive/warehouse/hive_class_b1.db/sales_data_parquet
drwxrwxrwx   - cloudera supergroup          0 2023-01-01 02:08 /user/hive/warehouse/hive_class_b1.db/sales_order_csv
drwxrwxrwx   - cloudera supergroup          0 2023-01-01 02:22 /user/hive/warehouse/hive_class_b1.db/sales_order_orc
drwxrwxrwx   - cloudera supergroup          0 2022-09-10 09:46 /user/hive/warehouse/hive_class_b1.db/test
drwxrwxrwx   - cloudera supergroup          0 2022-09-10 09:49 /user/hive/warehouse/hive_class_b1.db/test_external
drwxrwxrwx   - cloudera supergroup          0 2022-09-20 00:09 /user/hive/warehouse/hive_class_b1.db/users
[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/warehouse/hive_class_b1.db/parkingseason_partition
Found 3 items
drwxrwxrwx   - cloudera supergroup          0 2023-01-10 03:45 /user/hive/warehouse/hive_class_b1.db/parkingseason_partition/seasons=spring
drwxrwxrwx   - cloudera supergroup          0 2023-01-10 03:45 /user/hive/warehouse/hive_class_b1.db/parkingseason_partition/seasons=summer
drwxrwxrwx   - cloudera supergroup          0 2023-01-10 03:45 /user/hive/warehouse/hive_class_b1.db/parkingseason_partition/seasons=winter
[cloudera@quickstart ~]$
```