Hive Project-2

Name: - Trijit Adhikary

```
$ head -1 Parking_Violations_Issued_-_Fiscal_Year_2017.csv
Summons Number,Plate ID,Registration State,Plate Type,Issue Date,Violation Code,Vehicle Body Type,Vehicle Make,Issuing Age
ncy,Street Code1,Street Code2,Street Code3,Vehicle Expiration Date,Violation Location,Violation Precinct,Issuer Precinct,I
ssuer Code,Issuer Command,Issuer Squad,Violation Time,Time First Observed,Violation County,Violation In Front Of Or Opposi
te,House Number,Street Name,Intersecting Street,Date First Observed,Law Section,Sub Division,Violation Legal Code,Days Par
king In Effect    ,From Hours In Effect,To Hours In Effect,Vehicle Color,Unregistered Vehicle?,Vehicle Year,Meter Number,F
eet From Curb,Violation Post Code,Violation Description,No Standing or Stopping Violation,Hydrant Violation,Double Parking
 Violation
```

```
data = "Summons Number,Plate ID,Registration State,Plate Type,Issue Date,Violation Code,Vehicle Body Type,Vehicle Make,Issuing Age
data = data.lower().replace(' ','_').split(',')
for i in data:
    print(i+',')
```

Table creating: -

```
create table violation_all_data(
    summons_number int,
    plate_id int,
    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 string,
    violation_location string,
    violation_precinct int,
    vehicle_color string,
    unregistered_vehicle string,
    vehicle_year int,
    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");
```

```
create table violations(
    violation_code int,
    state string,
    body_type string,
    make string,
    streetc1 int,
    streetc2 int,
    streetc3 int,
    vpolice int,
    ipolice int,
    violation_time string,
    issue_date string
)
row format delimited
stored as orc;
```

Find the total number of tickets for the year: -

```
hive (project2)> select count(*) from violations where year(from_unixtime(unix_timestamp(issue_date,'MM/dd/yyyy'),'yyyy-MM-dd')) = 2017
               > ;
```

```
OK
_c0
5431903
```

Filtering 2017 data: -

```
hive (project2)> create table violations_17 as
               > select * from violations
               > where year(from_unixtime(unix_timestamp(issue_date,'MM/dd/yyyy'),'yyyy-MM-dd')) =2017;
Query ID = cloudera_20220922185454_a430d6bd-e8e8-4877-83d0-bebe69c1e701
Total jobs = 3
```

Find out how many unique states the cars which got parking tickets came from: -

```
hive (project2)> select count(distinct state) from violations_17;
```

```
_c0
65
```

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 (project2)> select count(*) from violations_17 where streetc1 is null or streetc2 is null or streetc3 is null;
```

```
OK
_c0
4
```

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

```
hive (project2)> select violation_code, count(*) as vcount from violations_17 group by violation_code order by vcount desc limit 5;
```

```
violation_code  vcount
21      768082
36      662765
38      542079
14      476660
20      319646
```

How often does each vehicle body type get a parking ticket? How about the vehicle make?

```
hive (project2)> select body_type, count(*) as vcount from violations_17 group by body_type order by vcount desc limit 5;
```

```
body_type       vcount
SUBN    1883953
4DSD    1547307
VAN     724025
DELV    358982
SDN     194197
```

```
hive (project2)> select make, count(*) as vcount from violations_17 group by make order by vcount desc limit 5;
```

```
make    vcount
FORD    636842
TOYOT   605290
HONDA   538884
NISSA   462017
CHEVR   356032
Time taken: 133.416 seconds, Fetched: 5 row(s)
```

Violating Precincts (this is the precinct of the zone where the violation occurred): -

```
hive (project2)> select vpolice, count(*) as vcount from violations_17 group by vpolice order by vcount desc limit 5;
```

```
vpolice vcount
0       925596
19      274443
14      203552
1       174702
18      169131
Time taken: 127.707 seconds, Fetched: 5 row(s)
```

Issuer Precincts (this is the precinct that issued the ticket): -

```
hive (project2)> select ipolice, count(*) as vcount from violations_17 group by ipolice order by vcount desc limit 5;
```

```
ipolice vcount
0       1078403
19      266959
14      200494
1       168740
18      162994
Time taken: 126.802 seconds, Fetched: 5 row(s)
```

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 (project2)> select ipolice,violation_code, count(*) vcount from violations_17 where ipolice in (0,19,14,1) group by ipolice,violat
ion_code order by ipolice aesc,vcount desc;
```

ipolice	violation_code	vcount
0	36	662765
0	7	210175
0	21	126050
0	5	48076
0	66	5258
0	14	4222
0	78	3564
0	20	2801
0	17	1688
0	40	1687
0	46	1439
0	85	1380
0	19	913
0	41	832
0	71	814
0	38	660
0	70	559
0	80	524
0	16	498
0	74	435
0	51	385
0	24	351
0	31	340
0	98	257
0	61	228
0	10	227
0	37	220
0	0	191
0	94	184
0	67	164

1	14	38354
1	16	19081
1	20	15408
1	46	12745
1	38	8535
1	17	7526
1	37	6470
1	31	5853
1	69	5672
1	19	5375
1	10	4712
1	40	4592
1	21	4055
1	71	3581
1	84	3310
1	42	2708
1	51	2223
1	9	2206
1	70	2183
1	48	1907
1	53	1737
1	50	1374
1	13	1367
1	24	1193
1	74	1135
1	82	775
1	4	461
1	60	438
1	23	421
1	78	406
1	66	368
1	26	290
1	68	282
1	18	254
1	89	206

14	14	45036
14	69	30464
14	31	22555
14	47	18364
14	42	10027
14	46	7679
14	19	7030
14	84	6743
14	82	5052
14	40	3582
14	17	3534
14	38	3269
14	9	2874
14	20	2761
14	71	2757
14	13	2701
14	48	2439
14	89	1960
14	50	1824
14	11	1745
14	79	1495
14	70	1461
14	10	1319
14	37	1256
14	64	1070
14	23	1044
14	21	1029
14	53	953
14	24	946
14	16	940
14	74	768
14	35	675
14	8	588
14	51	559
14	52	549

19	46	48444
19	38	36386
19	37	36056
19	14	29797
19	21	28414
19	20	14629
19	40	11416
19	16	9926
19	71	7493
19	19	6856
19	10	5643
19	84	4910
19	70	4459
19	18	3148
19	69	2910
19	31	2080
19	53	1736
19	50	1483
19	17	1464
19	48	1460
19	74	1329
19	24	1029
19	42	903
19	82	888
19	47	702
19	51	539
19	9	480
19	13	445
19	64	389
19	45	241
19	23	207
19	78	189