**CREATE TABLE :**

Create table if not exists TrafficViolation

(Date\_Of\_Stop STRING,Time\_of\_Stop STRING,Agency STRING,SubAgency STRING,Description STRING,loca STRING,Lattitude int,Longitude int,accident STRING,Belts STRING,Personal\_Injury STRING,Property\_Damage STRING,Fatal STRING,Commercial\_License STRING,HAZMAT STRING,Commercial\_Vehicle STRING,Alchol STRING,Workzone STRING,State STRING,Vehicle\_Type STRING,Man\_Year STRING,Vehicle\_Name STRING,Model STRING,Color STRING,

Violation STRING,Charge STRING,Article STRING,Contribution STRING,Race STRING,Gender STRING,

Driver\_city STRING,Driver\_State STRING,DL\_State STRING,Arrest\_Type STRING,Geo\_location STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

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

**LOAD DATA INTO:**

LOAD DATA INPATH '\HdiSamples\Traffic\_Violations.txt' OVERWRITE INTO TABLE TrafficViolation;

**SELECT STATE QUERY:**

Select DL\_State,COUNT(DL\_State) AS NO\_OF\_INCIDENTS\_OCCURED FROM TrafficViolation

GROUP BY DL\_State ORDER BY NO\_OF\_INCIDENTS\_OCCURED DESC LIMIT 20;

**SELECT DRIVER\_CITY QUERY:**

Select Driver\_city,COUNT(Driver\_city) AS NO\_OF\_INCIDENTS\_OCCURED FROM TrafficViolation

GROUP BY Driver\_city ORDER BY NO\_OF\_INCIDENTS\_OCCURED DESC LIMIT 20;

**GENDER QUERY:**

Select Gender,COUNT(Gender) AS NO\_OF\_INCIDENTS\_OCCURED FROM TrafficViolation

GROUP BY Gender ORDER BY NO\_OF\_INCIDENTS\_OCCURED DESC LIMIT 20;

**QUERY FOR ARREST TYPE:**

Select Arrest\_Type,COUNT(Arrest\_Type) AS NO\_OF\_INCIDENTS\_OCCURED FROM TrafficViolation

GROUP BY Arrest\_Type ORDER BY NO\_OF\_INCIDENTS\_OCCURED DESC LIMIT 20;

**QUERY FOR LOCATION VALUES:**

select location\_values.latval, location\_values.longval from location\_values where location\_values.distance between 0 and 400

**QUERY FOR VIOLATIONS AT DIFFERENT TIMES:**

Select TIME,COUNT(TIME) AS INCEDENTS\_RECORDED FROM location\_values GROUP BY TIME ORDER BY INCEDENTS\_RECORDED DESC LIMIT 10

**QUERY FOR VIOLATIONS IN A YEAR:**

Select YEAR,COUNT(YEAR) AS INCEDENTS\_RECORDED FROM location\_values GROUP BY YEAR ORDER BY INCEDENTS\_RECORDED DESC LIMIT 10;

**CREATE TABLE LOCATION\_VALUES:**

create table location\_values(Date\_Of\_Stop STRING,Time\_of\_Stop STRING,GENDER STRING,RACE STRING,latVal double,longVal double, YEAR STRING,TIME STRING,distance double);

**INSERT OVERWRITE QUERY:**

INSERT OVERWRITE TABLE location\_values Select Date\_Of\_Stop,Time\_of\_Stop,GENDER,RACE,Lattitude,Longitude,SUBSTR(Date\_Of\_Stop,6,4),SUBSTR(Time\_of\_Stop,0,2),

2 \* asin(

sqrt(

cos(radians(39.2196))\*

cos(radians(Lattitude)) \*

pow(sin(radians(-77.05772- Longitude)/2),2)

+

pow(sin(radians(39.2196- Lattitude)/2),2)

)) \*3956

from TrafficViolation;

**QUERY FOR VIOLATIONS OF MEN:**

select location\_values.distance as score, count(\*) as occurences

from (

select case

when distance between 0 and 5 then ' 0- 5'

when distance between 5 and 10 then '5-10'

when distance between 10 and 15 then '10-15'

when distance between 15 and 20 then '15-20'

when distance between 20 and 25 then '20-25'

when distance between 25 and 30 then '25-30'

when distance between 30 and 35 then '30-35'

when distance between 35 and 40 then '35-40'

when distance between 40 and 50 then '40-50'

else '>50'

end as distance

from location\_values WHERE GENDER='M') location\_values

group by location\_values.distance

order by occurences;

**QUERY FOR VIOLATIONS BY RACE:**

Select Race,COUNT(Race) AS NO\_OF\_INCIDENTS\_OCCURED FROM TrafficViolation

GROUP BY Race ORDER BY NO\_OF\_INCIDENTS\_OCCURED DESC LIMIT 20;

**QUERY FOR VIOLATIONS BY NATIVE AMERICANS:**

select location\_values.distance as score, count(\*) as occurences

from (

select case

when distance between 0 and 5 then ' 0- 5'

when distance between 5 and 10 then '5-10'

when distance between 10 and 15 then '10-15'

when distance between 15 and 20 then '15-20'

when distance between 20 and 25 then '20-25'

when distance between 25 and 30 then '25-30'

when distance between 30 and 35 then '30-35'

when distance between 35 and 40 then '35-40'

when distance between 40 and 50 then '40-50'

else '>50'

end as distance

from location\_values WHERE RACE='NATIVE\_AMERICAN') location\_values

group by location\_values.distance

order by occurences;

**QUERY FOR VIOLATIONS IN THE YEAR 2015:**

select location\_values.distance as score, count(\*) as occurences

from (

select case

when distance between 0 and 5 then ' 0- 5'

when distance between 5 and 10 then '5-10'

when distance between 10 and 15 then '10-15'

when distance between 15 and 20 then '15-20'

when distance between 20 and 25 then '20-25'

when distance between 25 and 30 then '25-30'

when distance between 30 and 35 then '30-35'

when distance between 35 and 40 then '35-40'

when distance between 40 and 50 then '40-50'

else '>50'

end as distance

from location\_values WHERE YEAR='2015') location\_values

group by location\_values.distance

order by occurences;

**QUERY FOR VIOLATIONS AT TIME 22:00:**

select location\_values.distance as score, count(\*) as occurences

from (

select case

when distance between 0 and 5 then ' 0- 5'

when distance between 5 and 10 then '5-10'

when distance between 10 and 15 then '10-15'

when distance between 15 and 20 then '15-20'

when distance between 20 and 25 then '20-25'

when distance between 25 and 30 then '25-30'

when distance between 30 and 35 then '30-35'

when distance between 35 and 40 then '35-40'

when distance between 40 and 50 then '40-50'

else '>50'

end as distance

from location\_values WHERE TIME='22') location\_values

group by location\_values.distance

order by occurences;