**HIVE TABLES CREATION:**

hive

show databases;

use database\_name;

set hive.cli.print.header=true;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

create external table homicide

(Address string,

Date\_homicide string,

Time\_homicide string,

Location\_homicide string,

Neighborhood string,

Age string,

Gender string,

Race string,

Name string,

Cause string,

case\_number string) row format delimited fields terminated by ',' location '/user/ubuntu/homicide\_data';

create table homicide\_internal like homicide;

insert overwrite table homicide\_internal select \* from homicide where Date\_homicide != 'Date';

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

drop table if exists crime\_homicide\_allData;

create table if not exists crime\_homicide\_allData as

select trim(c.case\_number) case\_number,

trim(c.district) district,

h.Address,

h.Date\_homicide,

h.Time\_homicide,

h.Location\_homicide,

h.Neighborhood,

h.Age,

h.Gender,

h.Race,

h.Name,

h.Cause

from crime\_internal c

join homicide\_internal h

on c.case\_number=h.case\_number

where trim(primary\_type) in ('HOMICIDE');

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

drop table if exists crime\_homicide\_queryData;

create table if not exists crime\_homicide\_queryData as

select trim(c.case\_number) case\_number,

trim(c.district) district,

d.district\_name,

d.zipcode,

h.Address,

h.Date\_homicide,

h.Time\_homicide,

h.Location\_homicide,

h.Neighborhood,

h.Age,

h.Gender,

h.Race,

h.Name,

h.Cause,

c.arrest,

c.latitude,

c.longitude

from crime\_internal c

join homicide\_internal h

on c.case\_number=h.case\_number

join districtdata d

on d.district=c.district

where trim(primary\_type) in ('HOMICIDE');

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

create external table crime(

ID\_crime string,

Case\_Number string,

Date\_crime string,

Block string,

IUCR string,

Primary\_Type string,

Description string,

Location\_Description string,

Arrest string,

Domestic string,

Beat string,

District string,

Ward string,

Community\_Area string,

FBI\_Code string,

X\_Coordinate string,

Y\_Coordinate string,

Year\_crime string,

Updated\_On string,

Latitude string,

Longitude string

) row format delimited fields terminated by ',' location '/user/ubuntu/crime\_data/';

create table crime\_internal like crime;

insert overwrite table crime\_internal select \* from crime where Date\_crime != 'Date';

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

create external table districtdata

(district string,

district\_name string,

zipcode string) row format delimited fields terminated by ',' location '/user/ubuntu/DistrictData';

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

drop table if exists crime\_homicide\_internal;

create table if not exists crime\_homicide\_internal as

select trim(c.case\_number) case\_number,trim(c.district) district ,trim(c.domestic) domestic,

trim(h.gender) gender,case when length(trim(race))=0 then 'Other' else trim(h.race) end as race

from crime\_internal c join homicide\_internal h on c.case\_number=h.case\_number where trim(primary\_type) in ('HOMICIDE');

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

create table crime\_homicide\_binary as

select case when district in ('001') then 1 else 0 end as district\_1,

case when district in ('002') then 1 else 0 end as district\_2,

case when district in ('003') then 1 else 0 end as district\_3,

case when district in ('004') then 1 else 0 end as district\_4,

case when district in ('005') then 1 else 0 end as district\_5,

case when district in ('006') then 1 else 0 end as district\_6,

case when district in ('007') then 1 else 0 end as district\_7,

case when district in ('008') then 1 else 0 end as district\_8,

case when district in ('009') then 1 else 0 end as district\_9,

case when district in ('010') then 1 else 0 end as district\_10,

case when district in ('011') then 1 else 0 end as district\_11,

case when district in ('012') then 1 else 0 end as district\_12,

case when district in ('013') then 1 else 0 end as district\_13,

case when district in ('014') then 1 else 0 end as district\_14,

case when district in ('015') then 1 else 0 end as district\_15,

case when district in ('016') then 1 else 0 end as district\_16,

case when district in ('017') then 1 else 0 end as district\_17,

case when district in ('018') then 1 else 0 end as district\_18,

case when district in ('019') then 1 else 0 end as district\_19,

case when district in ('020') then 1 else 0 end as district\_20,

case when district in ('021') then 1 else 0 end as district\_21,

case when district in ('022') then 1 else 0 end as district\_22,

case when district in ('023') then 1 else 0 end as district\_23,

case when district in ('024') then 1 else 0 end as district\_24,

case when district in ('025') then 1 else 0 end as district\_25,

case when race in ('Asian') then 1 else 0 end as race\_asian,

case when race in ('Black') then 1 else 0 end as race\_black,

case when race in ('White') then 1 else 0 end as race\_white,

case when race in ('Other') then 1 else 0 end as race\_other,

case when domestic in ('true') then 1 else 0 end as domestic,

case when gender in ('Male') then 1 else 0 end as gender

from crime\_homicide\_internal;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

create table crime\_binary as

select

case when substr(date\_crime,1,2) in ('01') then 1 else 0 end as month\_1,

case when substr(date\_crime,1,2) in ('02') then 1 else 0 end as month\_2,

case when substr(date\_crime,1,2) in ('03') then 1 else 0 end as month\_3,

case when substr(date\_crime,1,2) in ('04') then 1 else 0 end as month\_4,

case when substr(date\_crime,1,2) in ('05') then 1 else 0 end as month\_5,

case when substr(date\_crime,1,2) in ('06') then 1 else 0 end as month\_6,

case when substr(date\_crime,1,2) in ('07') then 1 else 0 end as month\_7,

case when substr(date\_crime,1,2) in ('08') then 1 else 0 end as month\_8,

case when substr(date\_crime,1,2) in ('09') then 1 else 0 end as month\_9,

case when substr(date\_crime,1,2) in ('10') then 1 else 0 end as month\_10,

case when substr(date\_crime,1,2) in ('11') then 1 else 0 end as month\_11,

case when substr(date\_crime,1,2) in ('12') then 1 else 0 end as month\_12,

case when primary\_type in ('BURGLARY') then 1 else 0 end as BURGLARY,

case when primary\_type in ('CRIM SEXUAL ASSAULT') then 1 else 0 end as CRIM\_SEXUAL\_ASSAULT,

case when primary\_type in ('CRIMINAL DAMAGE') then 1 else 0 end as CRIMINAL\_DAMAGE,

case when primary\_type in ('CRIMINAL TRESPASS') then 1 else 0 end as CRIMINAL\_TRESPASS,

case when primary\_type in ('DECEPTIVE PRACTICE') then 1 else 0 end as DECEPTIVE\_PRACTICE,

case when primary\_type in ('HOMICIDE') then 1 else 0 end as HOMICIDE,

case when primary\_type in ('MOTOR VEHICLE THEFT') then 1 else 0 end as MOTOR\_VEHICLE\_THEFT,

case when primary\_type in ('NARCOTICS') then 1 else 0 end as NARCOTICS,

case when primary\_type in ("NON-CRIMINAL (SUBJECT SPECIFIED)") then 1 else 0 end as NON\_CRIMINAL\_SUBJECT\_SPECIFIED,

case when primary\_type in ('OBSCENITY') then 1 else 0 end as OBSCENITY,

case when primary\_type in ('OTHER NARCOTIC VIOLATION') then 1 else 0 end as OTHER\_NARCOTIC\_VIOLATION,

case when primary\_type in ('OTHER OFFENSE') then 1 else 0 end as OTHER\_OFFENSE,

case when primary\_type in ('PROSTITUTION') then 1 else 0 end as PROSTITUTION,

case when primary\_type in ('PUBLIC PEACE VIOLATION') then 1 else 0 end as PUBLIC\_PEACE\_VIOLATION,

case when primary\_type in ('RITUALISM') then 1 else 0 end as RITUALISM,

case when primary\_type in ('SEX OFFENSE') then 1 else 0 end as SEX\_OFFENSE,

case when primary\_type in ('WEAPONS VIOLATION') then 1 else 0 end as WEAPONS\_VIOLATION,

case when primary\_type in ('ARSON') then 1 else 0 end as ARSON,

case when primary\_type in ('ASSAULT') then 1 else 0 end as ASSAULT,

case when primary\_type in ('BATTERY') then 1 else 0 end as BATTERY,

case when primary\_type in ('CONCEALED CARRY LICENSE VIOLATION') then 1 else 0 end as CONCEALED\_CARRY\_LICENSE\_VIOLATION,

case when primary\_type in ('DOMESTIC VIOLENCE') then 1 else 0 end as DOMESTIC\_VIOLENCE,

case when primary\_type in ('GAMBLING') then 1 else 0 end as GAMBLING,

case when primary\_type in ('HUMAN TRAFFICKING') then 1 else 0 end as HUMAN\_TRAFFICKING,

case when primary\_type in ('INTERFERENCE WITH PUBLIC OFFICER') then 1 else 0 end as INTERFERENCE\_WITH\_PUBLIC\_OFFICER,

case when primary\_type in ('INTIMIDATION') then 1 else 0 end as INTIMIDATION,

case when primary\_type in ('KIDNAPPING') then 1 else 0 end as KIDNAPPING,

case when primary\_type in ('LIQUOR LAW VIOLATION') then 1 else 0 end as LIQUOR\_LAW\_VIOLATION,

case when primary\_type in ('NON - CRIMINAL','NON-CRIMINAL') then 1 else 0 end as NONCRIMINAL,

case when primary\_type in ('OFFENSE INVOLVING CHILDREN') then 1 else 0 end as OFFENSE\_INVOLVING\_CHILDREN,

case when primary\_type in ('PUBLIC INDECENCY') then 1 else 0 end as PUBLIC\_INDECENCY,

case when primary\_type in ('ROBBERY') then 1 else 0 end as ROBBERY,

case when primary\_type in ('STALKING') then 1 else 0 end as STALKING,

case when primary\_type in ('THEFT') then 1 else 0 end as THEFT,

case when arrest in ('TRUE') then 1 else 0 end as arrest

from crime\_internal;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**HIVE QUERIES:**

**Crimes by Crime Category:**

create table query1 as

select Primary\_Type,Description,count(\*) as CrimeCount

from crime\_internal

group by Primary\_Type,Description order by Primary\_Type,Description;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Year wise crime data:**

create table query2 as

select year\_crime,primary\_type,count(\*) as CrimeCount

from crime\_internal

group by year\_crime,primary\_type

order by year\_crime,primary\_type;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Districts with the crime category it is most vulnerable to:**

create table query3 as

select d.district,e.district\_name,e.zipcode,d.primary\_type,d.count\_crimetype\_district from

(select c.district,b.primary\_type,c.count\_crimetype\_district from

(select district,max(CrimeCount) count\_crimetype\_district

from

(select district,primary\_type,count(\*) as CrimeCount

from crime\_internal group by district,primary\_type)a group by district)c

join

(select district,primary\_type,count(\*) as CrimeCount

from crime\_internal

group by district,primary\_type)b

on b.district=c.district and c.count\_crimetype\_district=b.CrimeCount) d

join districtdata\_internal e

on d.district=e.district;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Year wise crime and arrest data:**

create table query4 as

select a.year\_crime,CrimeCount,ArrestCount from

(select year\_crime,count(\*) as CrimeCount

from crime\_internal

group by year\_crime)a

join

(select year\_crime,count(\*) as ArrestCount

from crime\_internal

where arrest='TRUE'

group by year\_crime)b

on a.year\_crime=b.year\_crime

order by year\_crime;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Districts with the crime category having poor arrest rate:**

drop table if exists query5;

create table query5 as

select a.district,y.district\_name,y.zipcode,(ArrestCount/CrimeCount)\*100 as TotalArrestPercentage,e.primary\_type,min\_percentage\_district\_crimetype as Primary\_Type\_Min\_Arrest\_Rate

from

(select district,count(\*) as CrimeCount from crime\_internal group by district)a

join

(select district,count(\*) as ArrestCount from crime\_internal where arrest='TRUE' group by district)b on a.district=b.district

join

(select district,min(percentage\_district\_crimetype) min\_percentage\_district\_crimetype from

(

-- This query will give you the percentage of arrest versus total crime for each district ,primary\_type combination.

select a.district,a.primary\_type,(arrrest\_count\_crimetype/total\_count\_crimetype)\*100 percentage\_district\_crimetype

from

(select district,primary\_type,count(\*) as arrrest\_count\_crimetype

from crime\_internal where arrest='TRUE'

group by district,primary\_type)a

join

(select district,primary\_type,count(\*) as total\_count\_crimetype

from crime\_internal group by district,primary\_type)b on a.district=b.district and a.primary\_type=b.primary\_type)c group by district)d on a.district=d.district join

(

-- This query will give you the percentage of arrest versus total crime for each district ,primary\_type combination.

select a.district,a.primary\_type,(arrrest\_count\_crimetype/total\_count\_crimetype)\*100 percentage\_district\_crimetype

from

(select district,primary\_type,count(\*) as arrrest\_count\_crimetype

from crime\_internal where arrest='TRUE'

group by district,primary\_type)a

join

(select district,primary\_type,count(\*) as total\_count\_crimetype

from crime\_internal group by district,primary\_type)b on a.district=b.district and a.primary\_type=b.primary\_type)e on d.district=e.district and d.min\_percentage\_district\_crimetype=e.percentage\_district\_crimetype

join districtdata\_internal y on y.district=a.district;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Violent Crimes involving gun :**

create table query6 as

select Year\_crime, Count (case\_number) as TotalCrimes,

SUM ( if ( (DESCRIPTION LIKE '%FIREARM%' OR

DESCRIPTION LIKE '%GUN%'), 1 , 0) ) AS TotalWithGuns,

SUM ( if (

(primary\_type = 'ASSAULT' AND (DESCRIPTION LIKE '%FIREARM%' OR DESCRIPTION LIKE '%GUN%') AND (NOT DESCRIPTION LIKE '%SALE%' OR NOT DESCRIPTION LIKE '%POS%'))

OR

(primary\_type = 'CRIM SEXUAL ASSAULT' AND (DESCRIPTION LIKE '%FIREARM%' OR DESCRIPTION LIKE '%GUN%') AND (NOT DESCRIPTION LIKE '%SALE%' OR NOT DESCRIPTION LIKE '%POS%'))

OR

(primary\_type = 'OFFENSE INVOLVING CHILDREN' AND (DESCRIPTION LIKE '%FIREARM%' OR DESCRIPTION LIKE '%GUN%') AND (NOT DESCRIPTION LIKE '%SALE%' OR NOT DESCRIPTION LIKE '%POS%'))

OR

(primary\_type = 'OTHER OFFENSE' AND (DESCRIPTION LIKE '%FIREARM%' OR DESCRIPTION LIKE '%GUN%') AND (NOT DESCRIPTION LIKE '%SALE%' OR NOT DESCRIPTION LIKE '%POS%'))

OR

(primary\_type = 'RITUALISM' AND (DESCRIPTION LIKE '%FIREARM%' OR DESCRIPTION LIKE '%GUN%') AND (NOT DESCRIPTION LIKE '%SALE%' OR NOT DESCRIPTION LIKE '%POS%'))

OR

(primary\_type = 'ROBBERY' AND (DESCRIPTION LIKE '%FIREARM%' OR DESCRIPTION LIKE '%GUN%') AND (NOT DESCRIPTION LIKE '%SALE%' OR NOT DESCRIPTION LIKE '%POS%'))

OR

(primary\_type = 'WEAPSONS VIOLATION' AND (DESCRIPTION LIKE '%FIREARM%' OR DESCRIPTION LIKE '%GUN%') AND (NOT DESCRIPTION LIKE '%SALE%' OR NOT DESCRIPTION LIKE '%POS%'))

, 1 , 0)) AS TotalViolentWIthGuns

from crime\_internal

group by Year\_crime

order by Year\_crime desc ;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Homicides by Ethnicity and Gender :**

create table query\_7 as

select race,gender,count(\*) as Homicide\_Count

from crime\_homicide\_allData

group by race,gender;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Homicides by Gender and Cause of Homicide:**

create table query\_8 as

select gender,cause,count(\*) as Homicide\_Count

from crime\_homicide\_allData

group by gender,cause;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Homicides criminal arrest rate by ethnicity and gender:**

create table query\_9 as

select a.race,a.gender,(Homicide\_Criminal\_Arrest\_Count/Homicide\_Count)\*100 Homicide\_Criminal\_Arrest\_Percentage from

(select race,gender,count(\*) as Homicide\_Count

from crime\_homicide\_queryData

group by race,gender) a

join

(select race,gender,count(\*) as Homicide\_Criminal\_Arrest\_Count

from crime\_homicide\_queryData

where arrest='TRUE'

group by race,gender) b

on a.race=b.race and a.gender=b.gender;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Homicides by Geo Location:**

create table query\_10 as

select a.District\_Name,a.latitude,a.longitude,count(\*) Homicide\_Count

from crime\_homicide\_queryData a

group by a.District\_Name,a.latitude,a.longitude;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*