Hadoop Project HQL Commands

```
[root@sandbox ~]# hdfs dfs -get /Project_1/filtered_data.csv
[root@sandbox ~]# ls -l
total 134728
-rw-------  2 root root      2439 Jun  2  2016 anaconda-ks.cfg
-rw-r--r--  1 root root    373937 Oct 25  2016 blueprint.json
-rw-r--r--  1 root root        20 Oct 25  2016 build.out
-rw-r--r--  1 root root  75719611 Nov  2 20:21 filtered_data.csv
drwxr-xr-x  2 root root      4096 Oct 25  2016 hdp
-rw-r--r--  2 root root      7243 Jun  2  2016 install.log
-rw-r--r--  2 root root      1680 Jun  2  2016 install.log.syslog
-rw-r--r--  1 root root  61829279 Nov  2 16:04 KCDP_Final_Crime_Data_Original_1.csv
-rw-r--r--  1 root root       284 Oct 25  2016 sandbox.info
lrwxrwxrwx  1 root root        48 Oct 25  2016 start_ambari.sh -> /usr/lib/hue/tools/start_scripts/start_ambari.sh
lrwxrwxrwx  1 root root        47 Oct 25  2016 start_hbase.sh -> /usr/lib/hue/tools/start_scripts/start_hbase.sh
[root@sandbox ~]#
```

```
hive> create database Crime_Database;
OK
Time taken: 0.157 seconds
hive> use Crime_Database;
OK
Time taken: 0.27 seconds
hive>
```

```
hive> create table crime_data(Report_No string,Reported_year int,Reported_month int,Reported_day int,Reported_hour int,Reported_minute int,From_year int,From_month int,From_day int,From_hour int,From_minute int,offense string,IBRS string,Description string,Beat float,Address string,City string,Zip_code float,Rep_Dist string,Area string,DVFlag boolean,Involvement string,Race string,Sex string,Age float,Firearm_Used_Flag boolean,Age_Group string,Age_Range string,Age_Group2 string,Reported_Hour_f float,Month_Year string)
    > row format delimited fields terminated by ',';
OK
Time taken: 0.631 seconds
hive>
```

```
hive> load data local inpath 'filtered_data.csv' into table crime_data;
Loading data to table crime_database.crime_data
Table crime_database.crime_data stats: [numFiles=1, numRows=0, totalSize=75719611, rawDataSize=0]
OK
Time taken: 1.298 seconds
hive>
```

Verify data

```
hive> select * from crime_data limit 5;
OK
Report_No      NULL   NULL   NULL   NULL   NULL   NULL   NULL   NULL   NULL   NULL   Offense IBRS   Description      NULL   Addr
ess    City   NULL   Rep_Dist      Area   NULL   Involvement      Race   Sex    NULL   NULL   Age_Group      Age Range      Age
Group  NULL   Month_Year
KC20017500     2020   3      8      19     24     2020   3      8      19     0      Stealing – Shoplift      90J    Trespass of
Real Property  345.0  11600  E US 40 HWY      KANSAS CITY      64133.0 PJ3601 EPD    false  VIC    U      U      38.0   false  35-4
4      30-40  30-39  19.4   3-2020
KC20009823     2020   2      7      18     30     2020   2      7      18     1      Robbery (Armed Street)   120    Robbery 542.
0      8100  E BANNISTER RD    KANSAS CITY      64134.0 PJ6554 SPD    false  SUS    B      M      38.0   true   35-44  30-40  30-3
9      18.5   2-2020
KC20011191     2020   2      13     12     15     2020   2      13     12     15     Stealing – Shoplift      23C    Shoplifting3
45.0   11600  E US 40 HWY      KANSAS CITY      64133.0 PJ3601 EPD    false  ARR CHA SUS    B      M      37.0   false  35-44  30-4
0      30-39  12.25  2-2020
KC20002142     2020   1      9      11     29     2020   1      8      13     50     Stealing – Other         23H    All Other La
rceny  414.0  5600  NW 78TH TER    KANSAS CITY      64151.0 PP0447 NPD    false  CMP VIC W      M      54.0   false  45-59  50-6
0      50-59  11.483334       1-2020
Time taken: 0.172 seconds, Fetched: 5 row(s)
hive>
```

select count(*) from crime_data;

```
hive> select count(*) from crime_data;
Query ID = root_20241102202730_66c6b4bb-57fb-4619-9361-e012e3f79b97
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0010)

--------------------------------------------------------------------------------
        VERTICES      STATUS   TOTAL   COMPLETED   RUNNING   PENDING   FAILED   KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED    5          5          0         0        0        0
Reducer 2 ......    SUCCEEDED    1          1          0         0        0        0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 13.61 s
--------------------------------------------------------------------------------
OK
349354
Time taken: 15.456 seconds, Fetched: 1 row(s)
hive>
```

hive> select count(distinct offense) from crime_data;

```
hive> select count(distinct offense) from crime_data;
Query ID = root_20241102202859_b173f9c4-70c3-46ba-bafe-f564ba12b11f
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0010)


----------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED    5        5         0        0        0        0
Reducer 2 ......    SUCCEEDED    1        1         0        0        0        0
Reducer 3 ......    SUCCEEDED    1        1         0        0        0        0
----------------------------------------------------------------------------------
VERTICES: 03/03  [============================>>] 100%  ELAPSED TIME: 13.25 s
----------------------------------------------------------------------------------
OK
156
Time taken: 14.936 seconds, Fetched: 1 row(s)
hive>
```

```
hive> select distinct offense from crime_data;
Query ID = root_20241102202954_e177c751-c7c9-4c58-9c69-35a676427bff
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0010)


----------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED    5        5         0        0        0        0
Reducer 2 ......    SUCCEEDED    1        1         0        0        0        0
----------------------------------------------------------------------------------
VERTICES: 02/02  [============================>>] 100%  ELAPSED TIME: 11.74 s
----------------------------------------------------------------------------------
OK
Abandonment of a Child
Abuse of a Child
Adult Entertainment Violation
Alcohol Influence Report
Animal Abuse/Cruelty
Animal Bite
Arson
Assault (Aggravated)
Assault (Aggravated) on Department Member/Outside Law Enforcement Officer
Assault (Non-Aggravated)
Assault (Non-Aggravated) on Department Member/Outside Law Enforcement Officer
```

```
Stolen Auto
Suicide
Suicide - Attempted
Tampering
Tampering with Physical Evidence
Tavern/Nightclub Response Report
Terroristic Threats
Tobacco Law Violation
Tow-In Report/Authorization Not to Tow
Trafficking in Identifications
Trespass of Real Property
Unfounded
Unlawful Endangerment of Another
Unregistered Sex Offender
Vehicular - Fatality
Vehicular - Injury
Vehicular - Injury Hit and Run
Vehicular - Non-Injury
Vehicular - Non-Injury Hit and Run
Vehicular - Fatality
Vehicular - Injury
Vehicular - Injury Hit and Run
Vehicular - Non-Injury
Vehicular - Non-Injury Hit and Run
Violation of Ex-Parte Order of Protection
Violation of Full Order of Protection
Time taken: 12.344 seconds, Fetched: 156 row(s)
hive>
```

Yearly Trends

```
hive> select Reported_year,count(*) as total_crimes
    > from crime_data
    > group by Reported_year
    > order by Reported_year;
```

```
hive> select Reported_year,count(*) as total_crimes
    > from crime_data
    > group by Reported_year
    > order by Reported_year;
Query ID = root_20241102205829_608ea929-3e09-4d9a-b2d7-f96187a499a7
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1730562569604_0011)

----------------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------
Map 1 ..........     SUCCEEDED    5      5        0        0        0       0
Reducer 2 ......     SUCCEEDED    1      1        0        0        0       0
Reducer 3 ......     SUCCEEDED    1      1        0        0        0       0
----------------------------------------------------------------------------------------
VERTICES: 03/03   [==========================>>] 100%   ELAPSED TIME: 12.76 s
----------------------------------------------------------------------------------------
```

```
OK
NULL        1
2020        53703
2021        67425
2022        75414
2023        82141
2024        70670
Time taken: 18.13 seconds, Fetched: 6 row(s)
```

Monthly Trends

```
hive> select Reported_month,count(*) as total_crimes
    > from crime_data
    > group by Reported_month
    > order by Reported_month;
```

```
hive> select Reported_month,count(*) as total_crimes
    > from crime_data
    > group by Reported_month
    > order by Reported_month;
Query ID = root_20241102210108_64869534-5c3e-40de-9ce1-5af062d9a192
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0011)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 12.43 s
--------------------------------------------------------------------------------
```

```
OK
NULL     1
1        28181
2        25768
3        28173
4        28681
5        31411
6        32557
7        34625
8        34885
9        33141
10       29380
11       22644
12       19907
Time taken: 13.119 seconds, Fetched: 13 row(s)
hive>
```

Top 5 offenses

```
hive> select offense,count(*) as offense_count
    > from crime_data
    > group by offense
    > order by offense_count desc
    > limit 5;
```

```
hive> select offense,count(*) as offense_count
    > from crime_data
    > group by offense
    > order by offense_count desc
    > limit 5;
Query ID = root_20241102210321_3a9a52a3-bd5e-459b-9c25-48bc97f49c40
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0011)

----------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED     5        5        0        0       0       0
Reducer 2 ......   SUCCEEDED     1        1        0        0       0       0
Reducer 3 ......   SUCCEEDED     1        1        0        0       0       0
----------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 11.29 s
----------------------------------------------------------------------------
OK
Stolen Auto     36138
Domestic Violence Assault (Non-Aggravated)      29800
Property Damage 26083
Stealing from Auto (Theft from Auto)     23561
Assault (Aggravated)     23430
Time taken: 12.079 seconds, Fetched: 5 row(s)
hive>
```

Offense Distribution by gender

```
hive> select Sex,offense,count(*) as offense_count
    > from crime_data
    > group by Sex,offense
    > order by Sex,offense_count desc;
```

```
hive> select Sex,offense,count(*) as offense_count
    > from crime_data
    > group by Sex,offense
    > order by Sex,offense_count desc;
Query ID = root_20241102210806_e81f1f65-f165-4ec7-b620-33db6093a8b4
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0011)

----------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED    5        5        0        0        0       0
Reducer 2 ......    SUCCEEDED    1        1        0        0        0       0
Reducer 3 ......    SUCCEEDED    1        1        0        0        0       0
----------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 13.06 s
----------------------------------------------------------------------------
OK
B      Endangering the Welfare of a Child      457
B      Violation of Full Order of Protection   138
B      Violation of Ex-Parte Order of Protection      97
B      Domestic Violence Assault (Non-Aggravated)     63
B      Child Custody Violation 62
B      Abuse of a Child        41
B      Domestic Violence Assault (Aggravated)  34
```

Top 5 Offenses by ZIP Code

```
hive> select Zip_code,count(*) as offenses_by_zip_code
    > from crime_data
    > group by Zip_code
    > order by offenses_by_zip_code desc
    > limit 5;
```

```
hive> select Zip_code,count(*) as offenses_by_zip_code
    > from crime_data
    > group by Zip_code
    > order by offenses_by_zip_code desc
    > limit 5;
Query ID = root_20241102211330_7485e585-31c8-417f-8aa1-27059ea6ab38
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0011)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 12.47 s
--------------------------------------------------------------------------------
OK
64111.0 26783
64130.0 22125
64127.0 19784
64108.0 18491
64132.0 17013
Time taken: 13.022 seconds, Fetched: 5 row(s)
```

Average Age of Offenders by Offense Type

```
hive> select offense,avg(Age) as Avg_age_offenders
    > from crime_data
    > group by offense
    > order by Avg_age_offenders desc;
```

```
hive> select offense,avg(Age) as Avg_age_offenders
    > from crime_data
    > group by offense
    > order by Avg_age_offenders desc;
Query ID = root_20241102211805_6e1e6675-673e-48a9-9107-c1bfd0dc0bba
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0011)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 13.36 s
--------------------------------------------------------------------------------
OK
Possession of Gambling Device or Records      71.0
Financial Exploitation of the Elderly   64.11428571428571
Elder Abuse     54.734939759036145
Curfew Violation        48.4
Animal Bite     47.0
Identity Theft  44.062875309078066
Domestic Violence Lethality Screen for First Responders (LAP)   44.0
```

```
Abandonment of a Child  34.529411764705884
Rape - Statutory        34.493506493506494
Kidnapping       34.46712802768166
Invasion of Privacy      34.1564417177914
Murder  34.14713656387665
Eluding / Resisting a Lawful Stop        34.01764057331863
Adult Entertainment Violation    34.0
Domestic Violence Burglary (Residential)        33.857290589451914
False Imprisonment       33.6
Domestic Violence Robbery (Strong-Armed)        33.536986301369865
Liquor Law Violation     33.526315789473685
Cold Case Sex Offense    33.25
Rape – Statutory         33.11904761904762
Suicide – Attempted      32.75
Police Vehicle Damage (Form 154 P.D.)    32.4
Human Trafficking/Commercial Sex Acts    32.25581395348837
Outside Correspondence   32.0
Escape from Custody/Confinement 32.0
Officer Involved Shooting - Fatal        31.695652173913043
Tobacco Law Violation    31.666666666666668
Tavern/Nightclub Response Report         31.166666666666668
Interdiction     30.75
Possession of Illegal Firearm    30.37074829931973
Vehicular – Fatality     30.0
Suicide 27.0
Offense NULL
Time taken: 13.993 seconds, Fetched: 156 row(s)
hive>
```

Firearm count

```
hive> select count(case when Firearm_Used_Flag = True then 1
end) as Firemarm_Used,
    > count(case when Firearm_Used_Flag = False then 1 end) as
Firearm_not_used
    > from crime_data;
```

```
hive> select count(case when Firearm_Used_Flag = True then 1 end) as Firemarm_Used,
    > count(case when Firearm_Used_Flag = False then 1 end) as Firearm_not_used
    > from crime_data;
Query ID = root_20241102214929_77dfeb0d-0694-405a-91fc-6391facede2b
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED      5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 10.81 s
--------------------------------------------------------------------------------
OK
31599   315841
Time taken: 15.535 seconds, Fetched: 1 row(s)
hive>
```

DvFlag Count

```
hive> select count(case when DVFlag = True then 1 end) as
Domestic_violence,
    > count(case when DVFlag = False then 1 end) as
Non_Domestic_violence
    > from crime_data;
```

```
hive> select count(case when DVFlag = True then 1 end) as Domestic_violence,
    > count(case when DVFlag = False then 1 end) as Non_Domestic_violence
    > from crime_data;
Query ID = root_20241102215402_e184be19-e377-4c79-aee1-a1d47334c63a
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 11.87 s
--------------------------------------------------------------------------------
OK
31580   315860
Time taken: 12.409 seconds, Fetched: 1 row(s)
hive>
```

Firearm Usage by Offense Type

```
hive> select offense,count(*) as Firearm_Usage
    > from crime_data
    > where Firearm_Used_Flag = True
    > group by offense
    > order by Firearm_Usage desc;
```

```
hive> select offense,count(*) as Firearm_Usage
    > from crime_data
    > where Firearm_Used_Flag = True
    > group by offense
    > order by Firearm_Usage desc;
Query ID = root_20241102215645_3c24bdce-0f0c-4c31-9e2d-8e5cb7c1134f
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 10.81 s
--------------------------------------------------------------------------------
OK
Assault (Aggravated)    15133
Robbery (Armed Street)  4388
Domestic Violence Assault (Aggravated)  3361
Murder  1993
Robbery (Business)      1506
City Weapons Offense    1270
```

Top 5 Domestic Violence Flag by offense

```
hive> select offense,count(*) as DV_offenses
    > from crime_data
    > where DVFlag = True
    > group by offense
    > order by DV_offenses desc
    > limit 5;
```

```
hive> select offense,count(*) as DV_offenses
    > from crime_data
    > where DVFlag = True
    > group by offense
    > order by DV_offenses desc
    > limit 5;
Query ID = root_20241102220217_11d2ce4c-b838-461e-997c-7673a7d398ec
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

----------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1          1        0        0       0       0
----------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 9.30 s
----------------------------------------------------------------------------------
OK
Domestic Violence Assault (Non-Aggravated)      17434
Domestic Violence Assault (Aggravated)  5736
Domestic Violence Property Damage       1719
Harassment / Intimidation - Domestic Violence   1345
Violation of Full Order of Protection   618
Time taken: 9.77 seconds, Fetched: 5 row(s)
```

Offenses by Reported Year and Month (Year-Month Breakdown)

```
hive> select Reported_year,Reported_month,count(*) as
offenses_by_year_month
    > from crime_data
    > group by Reported_year,Reported_month
    > order by Reported_year,Reported_month;
```

```
hive> select Reported_year,Reported_month,count(*) as offenses_by_year_month
    > from crime_data
    > group by Reported_year,Reported_month
    > order by Reported_year,Reported_month;
Query ID = root_20241102220456_0933edb6-f756-420f-886d-fdadacc1b4ce
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

--------------------------------------------------------------------------------
        VERTICES      STATUS   TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5        5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1        1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1        1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 10.52 s
--------------------------------------------------------------------------------
OK
NULL    NULL    1
2020    1       4000
2020    2       3955
2020    3       3804
2020    4       3447
2020    5       4134
2020    6       4873
```

2020	5	4154
2020	6	4873
2020	7	4852
2020	8	5618
2020	9	5159
2020	10	5091
2020	11	4988
2020	12	3782
2021	1	5287
2021	2	4587
2021	3	5110
2021	4	5068
2021	5	5770
2021	6	6117
2021	7	6337
2021	8	6293
2021	9	6536
2021	10	6167
2021	11	5422
2021	12	4731
2022	1	5954
2022	2	5232
2022	3	6177
2022	4	6514
2022	5	6852
2022	6	6306
2022	7	6990
2022	8	6806
2022	9	6697
2022	10	6875

Race and Gender Distribution for Domestic Violence Cases

```
hive> select Race,Sex,count(*) as DV_by_race_gender
    > from crime_data
    > where DVFlag = True
    > group by Race,Sex
    > order by DV_by_race_gender desc;
```

```
hive> select Race,Sex,count(*) as DV_by_race_gender
    > from crime_data
    > where DVFlag = True
    > group by Race,Sex
    > order by DV_by_race_gender desc;
Query ID = root_20241102220801_d4f44ae1-b374-4159-980c-ffd26032ed3a
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

----------------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------
Map 1 ..........     SUCCEEDED     5        5        0        0        0        0
Reducer 2 ......     SUCCEEDED     1        1        0        0        0        0
Reducer 3 ......     SUCCEEDED     1        1        0        0        0        0
----------------------------------------------------------------------------------------
VERTICES: 03/03  [============================>>] 100%  ELAPSED TIME: 9.28 s
----------------------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------------------
OK
B        M        10000
B        F        9008
W        F        5997
W        M        5367
U        U        415
U        F        363
U        M        362
I        M        37
I        F        28
B        U        2
W        U        1
Time taken: 9.818 seconds, Fetched: 11 row(s)
hive>
```

Count of Unique Offense Descriptions

```
hive> select count(distinct Description) as unique_desc
    > from crime_data;
```

```
hive> select count(distinct Description) as unique_desc
    > from crime_data;
Query ID = root_20241102221106_6382fe3c-922b-4803-801f-328ba889b284
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 9.40 s
--------------------------------------------------------------------------------
OK
61
Time taken: 10.045 seconds, Fetched: 1 row(s)
hive>
```

Top 5 Most Common Offense Description

```
hive> select Description,count(*) as desc_count
    > from crime_data
    > group by Description
    > order by desc_count desc
    > limit 5;
```

```
hive> select Description,count(*) as desc_count
    > from crime_data
    > group by Description
    > order by desc_count desc
    > limit 5;
Query ID = root_20241102221314_c8cbb279-14fb-4370-91c5-9310c7519667
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

--------------------------------------------------------------------------------
        VERTICES      STATUS   TOTAL   COMPLETED   RUNNING   PENDING   FAILED   KILLED
--------------------------------------------------------------------------------
Map 1 ..........     SUCCEEDED     5         5         0         0         0         0
Reducer 2 ......     SUCCEEDED     1         1         0         0         0         0
Reducer 3 ......     SUCCEEDED     1         1         0         0         0         0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==============================>>] 100%   ELAPSED TIME: 9.52 s
--------------------------------------------------------------------------------
OK
Simple Assault   48457
Motor Vehicle Theft      36256
Aggravated Assault       34955
Vandalism/Destruction of Property       34054
Theft From Motor Vehicle        23169
Time taken: 10.026 seconds, Fetched: 5 row(s)
```

Top 5 Most Frequent Reporting District

```
hive> select Rep_Dist,count(*) as offenses_by_Repdist
    > from crime_data
    > group by Rep_Dist
    > order by offenses_by_Repdist desc
    > limit 5;
```

```
hive> select Rep_Dist,count(*) as offenses_by_Repdist
    > from crime_data
    > group by Rep_Dist
    > order by offenses_by_Repdist desc
    > limit 5;
Query ID = root_20241102221510_34dff49c-1f72-458d-babb-4b846d562a75
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)

--------------------------------------------------------------------------------
        VERTICES      STATUS    TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........     SUCCEEDED      5          5        0        0       0       0
Reducer 2 ......     SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......     SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03   [==========================>>] 100%   ELAPSED TIME: 9.07 s
--------------------------------------------------------------------------------
OK
PJ3601  7788
PP0321  4292
PJ4990  2507
PJ2650  2374
PJ1938  1887
Time taken: 9.522 seconds, Fetched: 5 row(s)
```

Top 5 Streets with Most Reported Offenses

```
hive> select Address,count(*) as offense_by_address
    > from crime_data
    > group by Address
    > order by offense_by_address desc
    > limit 5;
```

```
hive> select Address,count(*) as offense_by_address
    > from crime_data
    > group by Address
    > order by offense_by_address desc
    > limit 5;
Query ID = root_20241102221807_0e91bb15-415d-4e5e-aba7-1b348b651dd9
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)


----------------------------------------------------------------------------
        VERTICES      STATUS   TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------
Map 1 ..........     SUCCEEDED     5         5        0        0       0       0
Reducer 2 ......     SUCCEEDED     1         1        0        0       0       0
Reducer 3 ......     SUCCEEDED     1         1        0        0       0       0
----------------------------------------------------------------------------
VERTICES: 03/03  [===========================>>] 100%  ELAPSED TIME: 10.44 s
----------------------------------------------------------------------------
OK
11600  E US 40 HWY      5658
8500  N BOARDWALK AVE   3531
100  E LINWOOD BLVD     2294
2300  HOLMES ST 1782
2300  E MEYER BLVD      1679
Time taken: 10.963 seconds, Fetched: 5 row(s)
```

Offenses by Reported Day of the Week

```
hive> select from_unixtime(unix_timestamp(concat(Reported_year,
'-', Reported_month, '-', Reported_day), 'yyyy-MM-dd'), 'EEEE')
AS day_of_week,
    > COUNT(*) AS offenses_by_dayofweek
    > from crime_data
    > group by from_unixtime(unix_timestamp(concat(Reported_year,
'-', Reported_month, '-', Reported_day), 'yyyy-MM-dd'), 'EEEE')
    > order by offenses_by_dayofweek DESC;
```

```
hive> select from_unixtime(unix_timestamp(concat(Reported_year, '-', Reported_month, '-', Reported_day), 'yyyy-MM-dd'), 'EEEE') AS day_of_we
ek,
    > COUNT(*) AS offenses_by_dayofweek
    > from crime_data
    > group by from_unixtime(unix_timestamp(concat(Reported_year, '-', Reported_month, '-', Reported_day), 'yyyy-MM-dd'), 'EEEE')
    > order by offenses_by_dayofweek DESC;
Query ID = root_20241102222456_d25025aa-3c72-4d6b-934e-c547237f0289
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0012)


----------------------------------------------------------------------------
        VERTICES      STATUS   TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------
Map 1 ..........     SUCCEEDED     5         5        0        0       0       0
Reducer 2 ......     SUCCEEDED     1         1        0        0       0       0
Reducer 3 ......     SUCCEEDED     1         1        0        0       0       0
----------------------------------------------------------------------------
VERTICES: 03/03  [===========================>>] 100%  ELAPSED TIME: 11.82 s
----------------------------------------------------------------------------
```

```
OK
Monday   53007
Friday   50580
Tuesday  50367
Wednesday        49985
Thursday         48491
Saturday         48482
Sunday   48441
NULL     1
Time taken: 12.441 seconds, Fetched: 8 row(s)
hive>
```

Percentage of Offenses Involving Firearms

```
hive> SELECT (COUNT(IF(Firearm_Used_Flag = TRUE, 1, NULL)) *
100.0) / COUNT(*) AS Firearm_Usage_per FROM crime_data;
```

```
hive> SELECT (COUNT(IF(Firearm_Used_Flag = TRUE, 1, NULL)) * 100.0) / COUNT(*) AS Firearm_Usage_per FROM crime_data;
Query ID = root_20241102223836_60de4d53-0519-46a1-8134-b75237ba86c6
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1730562569604_0013)

----------------------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      5          5        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 12.00 s
----------------------------------------------------------------------------------------------
OK
9.044980163387281
Time taken: 17.14 seconds, Fetched: 1 row(s)
hive>
```

Offense Count by IBRS Code

```
hive> select IBRS,count(*) as offenses_by_IBRS
    > from crime_data
    > group by IBRS
    > order by offenses_by_IBRS desc;
```

```
hive> select IBRS,count(*) as offenses_by_IBRS
    > from crime_data
    > group by IBRS
    > order by offenses_by_IBRS desc;
Query ID = root_20241102224107_7c1118a8-cf11-432c-ad78-abd9c6f2b890
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0013)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 8.46 s
--------------------------------------------------------------------------------
OK
13B     48457
240     36256
13A     34955
290     34054
23F     23169
23C     23062
```

```
23F     23169
23C     23062
220     17351
90J     16119
90Z     14157
23H     13447
120     12379
23G     10907
23D     8966
13C     8557
35A     5095
90C     4674
90D     4530
280     3898
26A     3863
520     2913
26B     2874
26F     2605
250     2359
11A     1999
90F     1913
09A     1662
35B     1461
270     1217
26E     1210
11B     1181
200     951
11D     780
100     378
36B     304
```

Age Distribution for Top 3 Offenses

```
hive> SELECT Offense, Age, COUNT(*) AS Age_Distribution
    > from crime_data
    > where offense in ('Stolen Auto','Domestic Violence Assault
(Non-Aggravated)','Property Damage')
    > group by offense,Age
    > order by offense,Age;
```

```
hive> SELECT Offense, Age, COUNT(*) AS Age_Distribution
    > from crime_data
    > where offense in ('Stolen Auto','Domestic Violence Assault (Non-Aggravated)','Property Damage')
    > group by offense,Age
    > order by offense,Age;
Query ID = root_20241102225051_6f597b5c-a886-46e6-82ec-6503593c1d91
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0013)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED     1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED     1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 10.96 s
--------------------------------------------------------------------------------
OK
Domestic Violence Assault (Non-Aggravated)       NULL    94
Domestic Violence Assault (Non-Aggravated)       18.0    486
Domestic Violence Assault (Non-Aggravated)       19.0    596
Domestic Violence Assault (Non-Aggravated)       20.0    723
Domestic Violence Assault (Non-Aggravated)       21.0    777
Domestic Violence Assault (Non-Aggravated)       22.0    858
```

```
Domestic Violence Assault (Non-Aggravated)        83.0    4
Domestic Violence Assault (Non-Aggravated)        84.0    2
Domestic Violence Assault (Non-Aggravated)        85.0    7
Domestic Violence Assault (Non-Aggravated)        86.0    5
Domestic Violence Assault (Non-Aggravated)        87.0    5
Domestic Violence Assault (Non-Aggravated)        88.0    2
Domestic Violence Assault (Non-Aggravated)        89.0    1
Domestic Violence Assault (Non-Aggravated)        91.0    1
Domestic Violence Assault (Non-Aggravated)        92.0    4
Domestic Violence Assault (Non-Aggravated)        94.0    1
Domestic Violence Assault (Non-Aggravated)        97.0    1
Property Damage NULL    4
Property Damage 18.0    152
Property Damage 19.0    211
Property Damage 20.0    315
Property Damage 21.0    406
Property Damage 22.0    437
Property Damage 23.0    555
Property Damage 24.0    507
Property Damage 25.0    618
Property Damage 26.0    569
Property Damage 27.0    552
Property Damage 28.0    562
Property Damage 29.0    543
Property Damage 30.0    681
Property Damage 31.0    565
Property Damage 32.0    487
Property Damage 33.0    544
Property Damage 34.0    475
```

Get the Top 5 Areas with the Most Offenses Per Month

```
hive> SELECT Reported_month, Area, offenses_bymonth_area
    > from (
    > SELECT Reported_month, Area, COUNT(*) AS
offenses_bymonth_area,ROW_NUMBER() OVER (PARTITION BY
Reported_month ORDER BY COUNT(*) DESC)
AS rank
    > from crime_data
    > GROUP BY Reported_month, Area
    > ) ranked_data
    > where rank <=5
    > ORDER BY Reported_month, offenses_bymonth_area DESC;
```

```
hive> SELECT Reported_month, Area, offenses_bymonth_area
    > from (
    > SELECT Reported_month, Area, COUNT(*) AS offenses_bymonth_area,ROW_NUMBER() OVER (PARTITION BY Reported_month ORDER BY COUNT(*) DESC)
AS rank
    > from crime_data
    > GROUP BY Reported_month, Area
    > ) ranked_data
    > where rank <=5
    > ORDER BY Reported_month, offenses_bymonth_area DESC;
Query ID = root_20241102230025_d1e5721a-665c-4b76-9cc1-d5f2827aee20
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0013)

--------------------------------------------------------------------------------
        VERTICES      STATUS   TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      5          5        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......   SUCCEEDED      1          1        0        0       0       0
Reducer 4 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 04/04  [==========================>>] 100%  ELAPSED TIME: 10.70 s
--------------------------------------------------------------------------------
```

```
1          CPD      7836
1          EPD      7073
1          MPD      5300
1          SPD      3013
1          NPD      2415
2          CPD      7271
2          EPD      6542
2          MPD      4891
2          SPD      2706
2          NPD      2113
3          CPD      7803
3          EPD      6886
3          MPD      5511
3          SPD      3135
3          NPD      2371
4          CPD      7592
4          EPD      7511
4          MPD      5681
4          SPD      3121
4          NPD      2340
5          CPD      8817
5          EPD      7851
5          MPD      6120
5          SPD      3271
5          NPD      2676
6          CPD      9061
6          EPD      8432
6          MPD      6678
6          SPD      2956
```

Average Reporting Time for Each Offense Type

```
hive> select offense,avg(Reported_hour) as
Avg_Hour,avg(Reported_minute) as Avg_Min
   > from crime_data
   > group by offense
   > order by Avg_Hour;
```

```
hive> select offense,avg(Reported_hour) as Avg_Hour,avg(Reported_minute) as Avg_Min
    > from crime_data
    > group by offense
    > order by Avg_Hour;
Query ID = root_20241102230424_54d825a0-2884-41a0-ab0a-9491dc6cb734
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0013)

--------------------------------------------------------------------------------
        VERTICES      STATUS    TOTAL   COMPLETED   RUNNING   PENDING   FAILED   KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED      5          5         0         0        0        0
Reducer 2 ......    SUCCEEDED      1          1         0         0        0        0
Reducer 3 ......    SUCCEEDED      1          1         0         0        0        0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 9.47 s
--------------------------------------------------------------------------------
OK
Offense NULL     NULL
Adult Entertainment Violation   0.0     50.0
Officer Involved Shooting - Fatal       4.565217391304348       15.391304347826088
Police Vehicle Damage (Form 154 P.D.)   9.0     17.0
Service of a Full Order of Protection   9.066666666666666       26.933333333333334
Burglary (Non-Residential)      9.229497176660393       29.202339338531864
Trafficking in Identifications  9.8125  30.96875
Stealing from Coin Operated Device      9.925925925925926       32.05555555555556
```

```
Trafficking in Identifications  9.8125   30.96875
Stealing from Coin Operated Device      9.925925925925926       32.05555555555556
False Imprisonment      10.0    27.0
Alcohol Influence Report        10.254798761609907      29.66160990712074
Interdiction    10.464285714285714      30.678571428571427
Tavern/Nightclub Response Report        10.833333333333334      34.458333333333336
Arson   10.85995085995086       30.848894348894348
Possession of Gambling Device or Records        11.0    47.0
Animal Bite     11.0    17.0
Missing Person  11.0    23.0
Unregistered Sex Offender       11.0    38.0
Cold Case Sex Offense   11.0    37.0
Burning or Exploding    11.320388349514563      27.16504854368932
Tampering       11.339823008849558      28.398230088495577
Domestic Violence Burglary (Residential)        11.781799379524301      29.7952430196483
Vehicular – Injury Hit and Run  11.924528301886792      32.15094339622642
Abandonment of a Child  11.977272727272727      27.25
Discharge of Firearm (LEO Only) 12.0    12.2
Stealing from Auto (Theft from Auto)    12.002801239336192      29.39484741734222
Rape    12.04137039431157       28.770523594053007
Stolen Auto     12.07847694947147      29.311943106978802
Vehicular – Non-Injury  12.132142857142858      30.182142857142857
Tow-In Report/Authorization Not to Tow  12.196078431372548      32.01960784313726
Property Damage 12.280182494344976      29.554844151362957
Unfounded       12.336842105263157      28.178947368421053
Vehicular - Injury      12.396648044692737      29.90223463687151
Miscellaneous City Ordinance Violation  12.42741935483871       28.67008797653959
Trespass of Real Property       12.434009249377446      29.092564923514765
Sodomy  12.438502673796792      30.087344028520498
```

Top 5 Most Common Involvement Type

```
hive> select Involvement,count(*) as involvement_count
    > from crime_data
    > group by Involvement
    > order by involvement_count desc
    > limit 5;
```

```
hive> select Involvement,count(*) as involvement_count
    > from crime_data
    > group by Involvement
    > order by involvement_count desc
    > limit 5;
Query ID = root_20241102230949_9185b705-19dd-4ffb-9a67-f4fc963a5a78
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0013)

--------------------------------------------------------------------------------
        VERTICES      STATUS   TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5        5        0        0        0       0
Reducer 2 ......    SUCCEEDED     1        1        0        0        0       0
Reducer 3 ......    SUCCEEDED     1        1        0        0        0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 8.97 s
--------------------------------------------------------------------------------
OK
VIC     174647
SUS     70689
ARR CHA SUS     40325
CMP VIC 28113
VIC SUS 6096
Time taken: 9.459 seconds, Fetched: 5 row(s)
```

Top 5 Crimes by Specific Address and Offense Type

```
hive> select offense,count(*) as offense_count
    > from crime_data
    > where Address = '11600  E US 40 HWY'
    > group by offense
    > order by offense_count desc
    > limit 5;
```

```
hive> select offense,count(*) as offense_count
    > from crime_data
    > where Address = '11600  E US 40 HWY'
    > group by offense
    > order by offense_count desc
    > limit 5;
Query ID = root_20241102233143_3dbc3572-df03-4d77-8b25-2fb71bc535f8
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1730562569604_0014)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED      5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 12.73 s
--------------------------------------------------------------------------------
OK
Stealing - Shoplift     3583
Stealing – Shoplift     561
Trespass of Real Property      424
Possession/Sale/Distribution of a Controlled Substance  105
```

Top 5 Areas with Highest Average Offender Age

```
hive> select Area,avg(Age) as Avg_Age_offenders
    > from crime_data
    > group by Area
    > order by Avg_Age_offenders desc
    > limit 5;
```

```
hive> select Area,avg(Age) as Avg_Age_offenders
    > from crime_data
    > group by Area
    > order by Avg_Age_offenders desc
    > limit 5;
Query ID = root_20241102233447_ccbc3eb2-8ce6-480d-b9ab-76b6c602f4de
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0014)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED      5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [============================>>] 100%  ELAPSED TIME: 11.16 s
--------------------------------------------------------------------------------
OK
SCP     39.220682036845666
EPD     38.93589787545285
MPD     38.82174466404319
NPD     38.76328372280962
SPD     38.621529108932585
Time taken: 11.668 seconds, Fetched: 5 row(s)
```

Gender and Age Group Interaction for Firearm-Related Offenses

```
hive> select Sex,Age_Group,count(*) as Firearm_offenses
    > from crime_data
    > where Firearm_Used_Flag = True
    > Group by Sex,Age_Group
    > order by Firearm_offenses desc;
```

```
hive> select Sex,Age_Group,count(*) as Firearm_offenses
    > from crime_data
    > where Firearm_Used_Flag = True
    > Group by Sex,Age_Group
    > order by Firearm_offenses desc;
Query ID = root_20241102233705_7b27f12e-59fc-48f8-9bf4-8f6acfd915ca
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0014)

----------------------------------------------------------------------------
        VERTICES        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED     5        5         0        0       0       0
Reducer 2 ......    SUCCEEDED     1        1         0        0       0       0
Reducer 3 ......    SUCCEEDED     1        1         0        0       0       0
----------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 13.06 s
----------------------------------------------------------------------------
OK
M       35-44   8566
M       25-34   4574
M       18-24   3717
F       35-44   3206
F       25-34   2510
M       45-59   2304
F       18-24   2028
```

```
OK
M       35-44   8566
M       25-34   4574
M       18-24   3717
F       35-44   3206
F       25-34   2510
M       45-59   2304
F       18-24   2028
U       35-44   1277
F       45-59   1184
M       60+     800
M       0-17    601
F       60+     558
F       0-17    240
U       25-34   14
U       18-24   12
U       0-17    3
U       45-59   3
U       60+     2
Time taken: 13.592 seconds, Fetched: 18 row(s)
hive>
```

Top 5 Offense Counts by Area and Gender

```
hive> SELECT Area, Sex, offense_count
    > from (
    > SELECT Area, Sex, COUNT(*) AS offense_count
    > from crime_data
    > GROUP BY Area, Sex
    > ) as counts
    > ORDER BY offense_count DESC
    > limit 5;
```

```
hive> SELECT Area, Sex, offense_count
    > from (
    > SELECT Area, Sex, COUNT(*) AS offense_count
    > from crime_data
    > GROUP BY Area, Sex
    > ) as counts
    > ORDER BY offense_count DESC
    > limit 5;
Query ID = root_20241102235237_ab216932-54a0-4af5-9fe7-7f3dc88b00bb
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1730562569604_0015)

----------------------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 ..........    SUCCEEDED      5          5        0        0       0       0
Reducer 2 ......    SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......    SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 13.16 s
----------------------------------------------------------------------------------------------
OK
```

```
OK
CPD     M       51663
EPD     M       46001
MPD     M       33482
CPD     F       30474
EPD     F       29685
Time taken: 18.617 seconds, Fetched: 5 row(s)
hive>
```

Number of Offenses Involving Firearms by Year

```
hive> select Reported_year,count(*) Firearm_offense_cnt
    > from crime_data
    > where Firearm_Used_Flag = 1
    > group by Reported_year
    > order by Reported_year;
```

```
hive> select Reported_year,count(*) Firearm_offense_cnt
    > from crime_data
    > where Firearm_Used_Flag = 1
    > group by Reported_year
    > order by Reported_year;
Query ID = root_20241102235654_2fc3a0cd-ff85-4cf7-a79f-a9284f2cf5c9
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0015)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      5          5        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 12.76 s
--------------------------------------------------------------------------------
OK
2020    5154
2021    5942
2022    6713
2023    7124
2024    6666
Time taken: 13.33 seconds, Fetched: 5 row(s)
```

Race Distribution of Domestic Violence Offenders

```
hive> select Race,count(*) DV_cnt
    > from crime_data
    > where DVFlag = 1
    > group by Race
    > order by DV_cnt desc;
```

```
hive> select Race,count(*) DV_cnt
    > from crime_data
    > where DVFlag = 1
    > group by Race
    > order by DV_cnt desc;
Query ID = root_20241102235906_b09b6d6b-5a79-4d61-a0b0-0ab37df3fe79
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1730562569604_0015)

--------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------
Map 1 ..........     SUCCEEDED    5        5        0        0        0       0
Reducer 2 ......     SUCCEEDED    1        1        0        0        0       0
Reducer 3 ......     SUCCEEDED    1        1        0        0        0       0
--------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 12.76 s
--------------------------------------------------------------------------
OK
B       19010
W       11365
U       1140
I       65
Time taken: 13.373 seconds, Fetched: 4 row(s)
hive>
```

Trend of Offenses by Month Over the Years

```
hive> select concat(Reported_month,'-',Reported_year) as
Month_year
    > from crime_data
    > group by Reported_month,Reported_year
    > order by Month_year;
```

```
hive> select concat(Reported_month,'-',Reported_year) as Month_year
    > from crime_data
    > group by Reported_month,Reported_year
    > order by Month_year;
Query ID = root_20241103000407_8494b0e0-3609-4795-948e-84501cb093ba
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1730562569604_0015)

-------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-------------------------------------------------------------------
Map 1 ..........    SUCCEEDED    5        5        0        0        0      0
Reducer 2 ......    SUCCEEDED    1        1        0        0        0      0
Reducer 3 ......    SUCCEEDED    1        1        0        0        0      0
-------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 12.41 s
-------------------------------------------------------------------
OK
NULL
1-2020
1-2021
1-2022
1-2023
1-2024
10-2020
10-2021
```

```
4-2024
5-2020
5-2021
5-2022
5-2023
5-2024
6-2020
6-2021
6-2022
6-2023
6-2024
7-2020
7-2021
7-2022
7-2023
7-2024
8-2020
8-2021
8-2022
8-2023
8-2024
9-2020
9-2021
9-2022
9-2023
9-2024
Time taken: 13.001 seconds, Fetched: 59 row(s)
```