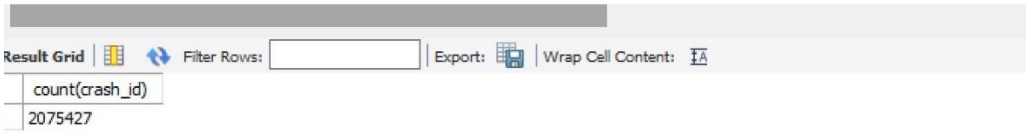
SQL Execution

1) How many accidents happened in NYC , Austin , Chicago

By City

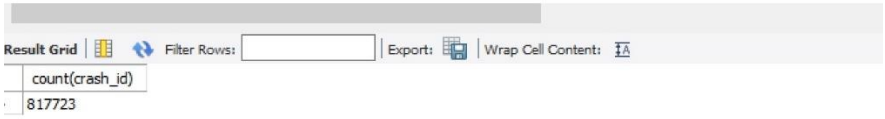
select count(crash_id) from fact_crash where source_system = 'New_York'

```
1    select count(crash_id) from fact_crash  where source_system = 'New_York'
```



select count(crash_id) from fact_crash where source_system = 'Chicago'

```
1    select count(crash_id) from fact_crash  where source_system = 'Chicago'
```



select count(crash_id) from fact_crash where source_system = 'Austin'

```
1   select count(crash_id) from fact_crash  where source_system = 'Austin'
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(crash_id)
147750

OverAll

Select count(crash_id) from fact_crash

```
1   select count(crash_id) from fact_crash
2
```

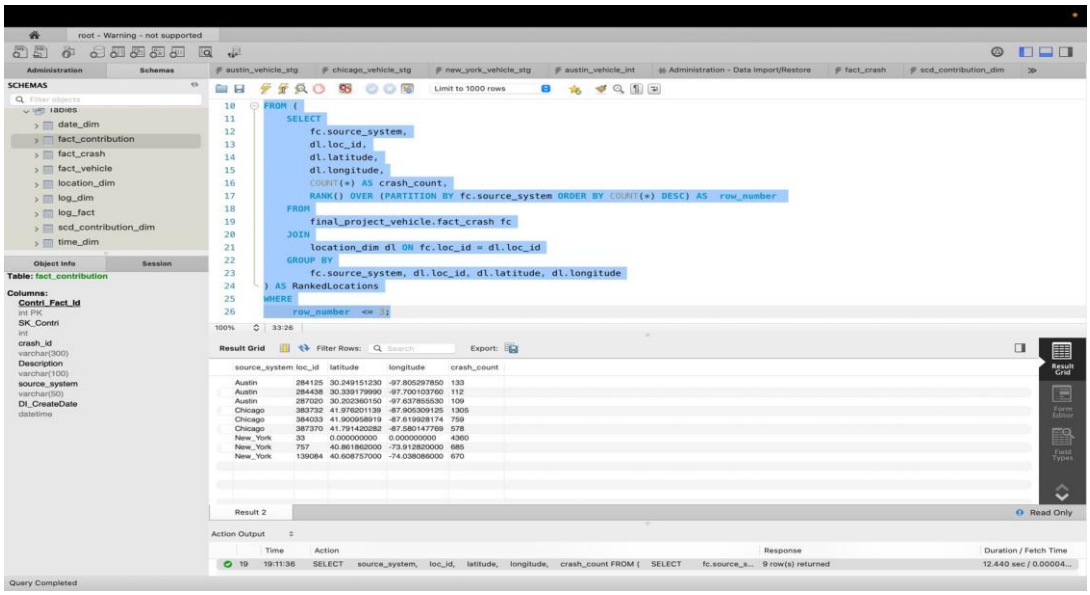
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(crash_id)
3040900

2) Which areas in three cities has greatest number of accidents?

```
SELECT
   source_system,
   loc_id,
   latitude,
   longitude,
   crash_count
FROM (
```

```
    SELECT
        fc.source_system,
        dl.loc_id,
        dl.latitude,
        dl.longitude,
        COUNT(*) AS crash_count,
        RANK() OVER (PARTITION BY fc.source_system ORDER BY COUNT(*) DESC) AS
☐ row_number ☐
    FROM
        final_project_vehicle.fact_crash fc
    JOIN
        location_dim dl ON fc.loc_id = dl.loc_id
    GROUP BY
        fc.source_system, dl.loc_id, dl.latitude, dl.longitude
) AS RankedLocations
WHERE
    ☐ row_number ☐ <= 3;
```



3) How many accidents resulted in just injuries, at two levels 1) Overall 2) By City?

OverAll

```
select Just_Injured_Fl, count(*) from fact_crash group by Just_Injured_Fl
```

```
13
14      select Just_Injured_Fl, count(*) from fact_crash group by Just_Injured_Fl
15
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Just_Injured_Fl	count(*)
1	650027
0	2390873

By City

select source_system, Just_Injured_Fl , count(*) from fact_crash group by source_system,Just_Injured_Fl

```
15
16      select source_system, Just_Injured_Fl , count(*) from fact_crash group by source_system,Just_Injured_Fl
17
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

source_system	Just_Injured_Fl	count(*)
New_York	1	473644
New_York	0	1601783
Austin	0	82980
Austin	1	64770
Chicago	1	111613
Chicago	0	706110

4) How often are the pedestrians involved in accidents at two levels 1) Overall 2) By City?

OverAll

```
19    select PEDESTRIAN_INVOLVED, count(*) from vehicle_int.fact_crash group by  PEDESTRIAN_INVOLVED
20
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

PEDESTRIAN_INVOLVED	count(*)
0	2108316
1	114861
NULL	817723

By City

```
15
16    select source_system, Just_Injured_Fl , count(*) from fact_crash group by source_system,Just_Injured_Fl
17
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

source_system	PEDESTRIAN_INVOLVED	count(*)
New_York	0	1961501
New_York	1	113926
Austin	0	146815
Austin	1	935
Chicago	NULL	817723

5) When do most accidents happen?

select t.Time_Of_Day,count(CRASH_ID) from fact_crash fc INNER JOIN time_dim t on fc.time_id = t.time_id group by t.Time_Of_Day

```
24
25    select t.Time_Of_Day,count(CRASH_ID) from fact_crash fc INNER JOIN time_dim t on fc.time_id = t.time_id group by t.Time_Of
26
```

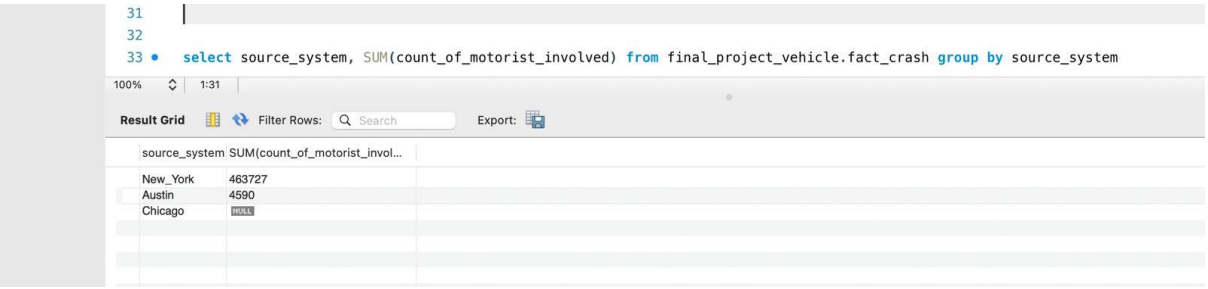
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Time_Of_Day	count(CRASH_ID)
Night	1060310
Morning	790917
Afternoon	1189673

Result Grid | Form Editor

6) How many motorist are injured or killed in accidents

```
select source_system, SUM(count_of_motorist_involved) from final_project_vehicle.fact_crash group by source_system
```



7) Which top 5 areas in three cities have most fatal number of accidents?

```
WITH CityDeaths AS (
    SELECT
        L.loc_id,
        L.latitude,
        L.longitude,
        F.source_system AS City,
        SUM(F.total_death_count) AS TotalDeaths
    FROM final_project_vehicle.fact_crash F
    JOIN Location_dim L ON F.loc_id = L.loc_id
    WHERE F.source_system IN ('Austin', 'New_York', 'Chicago')
    GROUP BY L.loc_id, L.latitude, L.longitude, F.source_system
```

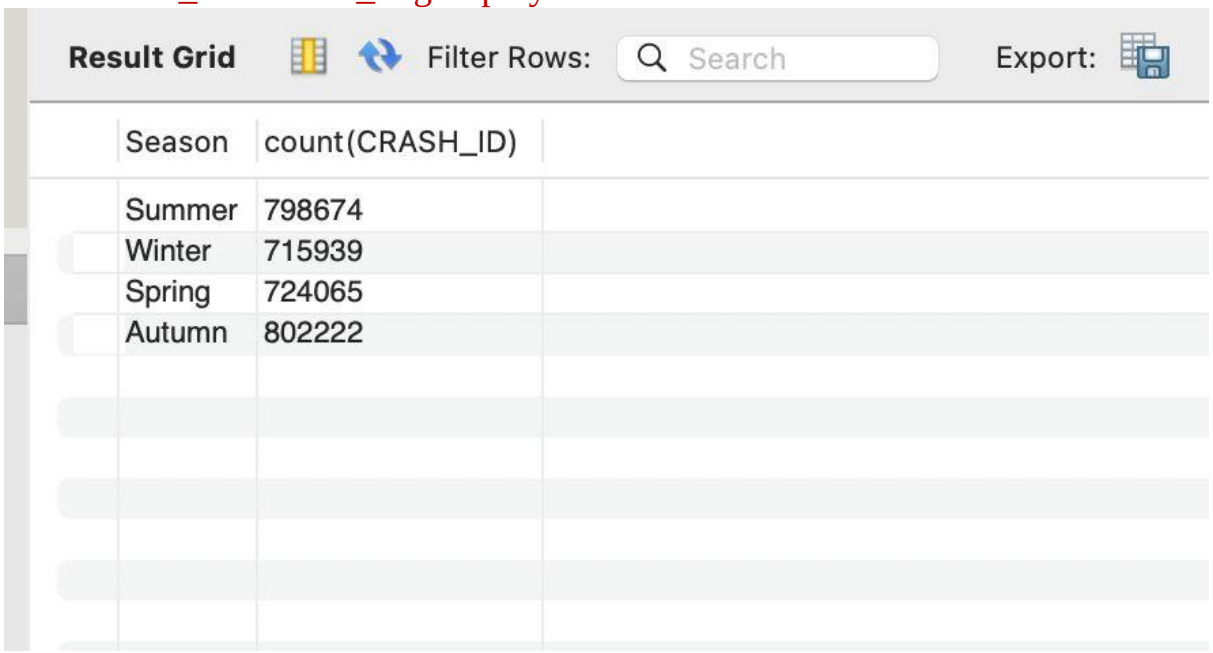
```
)

SELECT loc_id, latitude, longitude, City, TotalDeaths

FROM (

    SELECT *,

        ROW_NUMBER() OVER (PARTITION BY City ORDER BY TotalDeaths DESC) AS CityRank

    FROM CityDeaths

) Ranked

WHERE CityRank <= 5;
```

loc_id	latitude	longitude	City	TotalDeaths
374941	30.207165457	-97.757662720	Austin	4
299089	30.389762041	-97.733765919	Austin	4
297203	30.343924991	-97.696998172	Austin	3
320792	30.384494554	-97.742496239	Austin	3
356594	30.214906265	-97.732462482	Austin	3
390662	41.692742149	-87.610015536	Chicago	4
383518	41.786485236	-87.676665443	Chicago	4
609122	41.880986206	-87.693309786	Chicago	4
675075	41.873821995	-87.683504205	Chicago	3
504302	41.735258527	-87.704631511	Chicago	3
33	0.000000000	0.000000000	New_Y...	22
13469	40.729046000	-74.010730000	New_Y...	16
3489	40.656160000	-73.767360000	New_Y...	14
43043	40.697204500	-73.813334100	New_Y...	10
73289	40.748398000	-73.728790000	New_Y...	8

8) Time based analysis of accident?

```
select d.Season, count(CRASH_ID) from final_project_vehicle.fact_crash fc
INNER JOIN date_dim
```

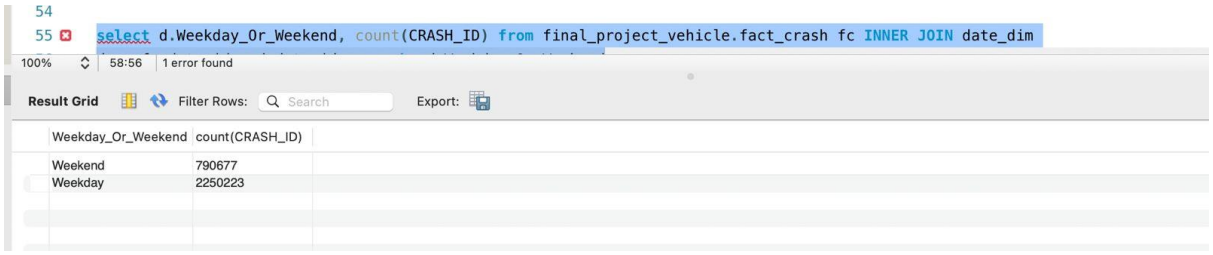
d on fc.date_id = d.date_id group by d.Season



Season	count(CRASH_ID)
Summer	798674
Winter	715939
Spring	724065
Autumn	802222

select d.Weekday_Or_Weekend, count(CRASH_ID) from final_project_vehicle.fact_crash fc INNER JOIN date_dim

d on fc.date_id = d.date_id group by d.Weekday_Or_Weekend



Weekday_Or_Weekend	count(CRASH_ID)
Weekend	790677
Weekday	2250223

select d.Day_of_Week,count(CRASH_ID) from fact_crash fc INNER JOIN date_dim d on fc.date_id = d.date_id group by d.Day_of_Week

Day_of_Week	count(CRASH_ID)
Saturday	421939
Wednesday	440445
Tuesday	443249
Monday	429002
Sunday	368738
Thursday	449692
Friday	487835

9) Fatality Analysis – Are pedestrians killed more often than road users

```
select SUM(Pedestrian_death_count) as Total_Pedestrian_death_count ,
SUM(Road_Users_death_count) as total_Road_Users_death_count from
final_project_vehicle.fact_crash
```

Total_Pedestrian_death_co...	total_Road_Users_death_count
326	5191

10)What are most common factors involved in accidents?

```
select s.Description, count(s.Description) from vehicle_curation.fact_contribution fco INNER JOIN vehicle_curation.scd_contribution_dim s on fco.SK_Contri = s.SK_Contri group by s.Description order by count(s.Description) desc
```

```
4 • select s.Description, count(s.Description) from crashinspection_int.fact_contribution fco INNER JOIN crashinspection_int.scd
5
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

Description	count(s.Description)
UNABLE TO DETERMINE	425938
other	338755
FTC	214866
FAILING TO YIELD RIGHT-OF-WAY	104173
Unsafe Speed	98381
Backed without Safety	82558
Passed in No Passing Lane	64838
Passed on Shoulder	57127
Turned Improperly - Cut Corner on Left	56937

Result Grid

Form Editor

11) Using Austin and NYC datasets, Create a visualization to show number of incidents that involved more than 2 vehicles. Show this data as a comparision between these 2 cities.

```
SELECT Source_System, COUNT(*) AS NumberOfIncidents
FROM (
  SELECT Source_System, Crash_Id
  FROM final_project_vehicle.fact_vehicle AS CombinedData
  GROUP BY Source_System, Crash_Id
  HAVING COUNT(*) > 2
) AS IncidentsMoreThanTwoVehicles
GROUP BY Source_System;
```

```sql
SELECT Source_System, COUNT(*) AS NumberOfIncidents
FROM (
  SELECT Source_System, Crash_Id
  FROM final_project_vehicle.fact_vehicle AS CombinedData
  GROUP BY Source_System, Crash_Id
  HAVING COUNT(*) > 2
) AS IncidentsMoreThanTwoVehicles
GROUP BY Source_System;
```

100% 1:57 1 error found

Result Grid Filter Rows: Search Export:

Source_System	NumberOfIncide...
Austin	25552
New_York	1065137