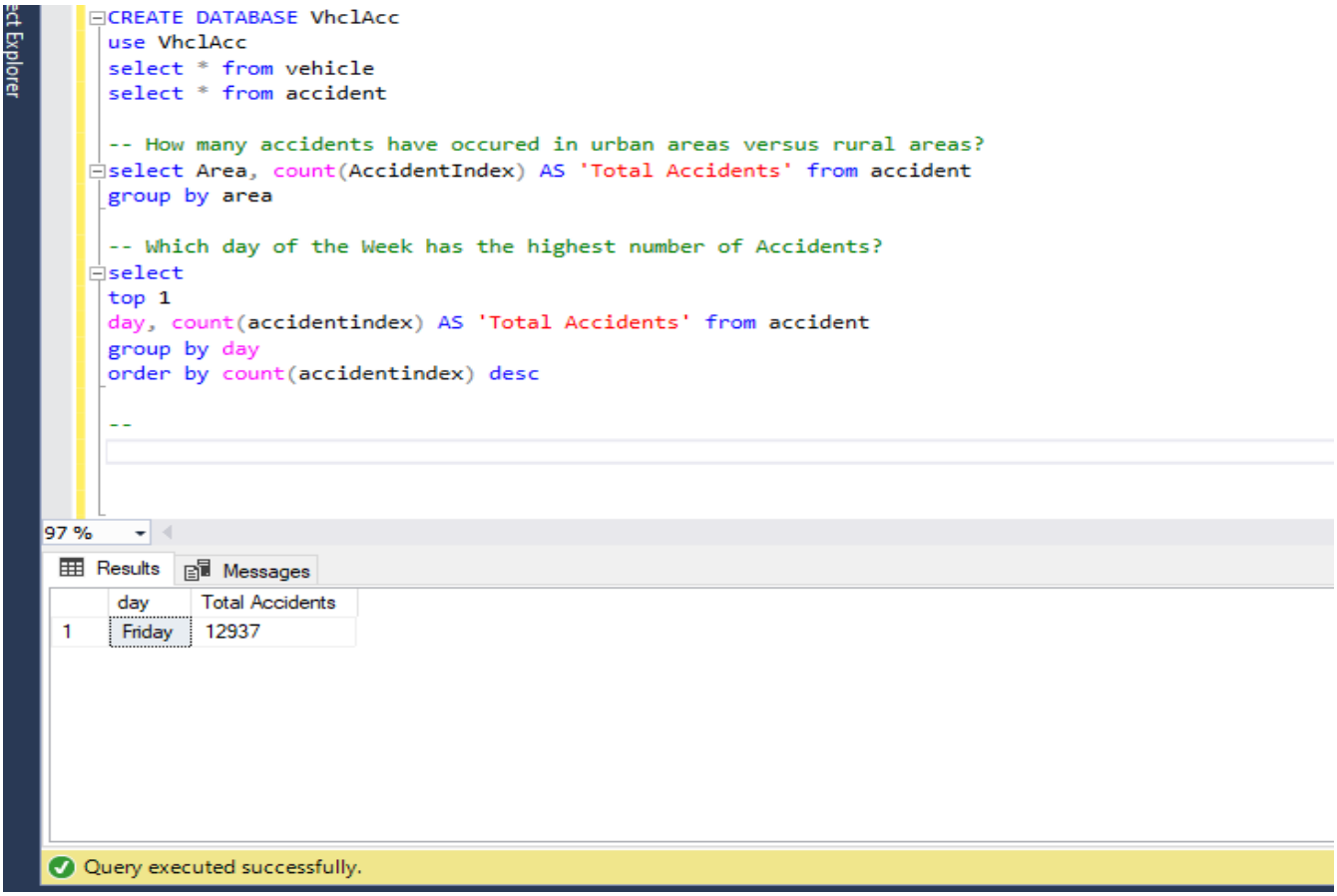
In this project I walk through SQL queries to run EXLORATORY data analysis about the accidents. I used MS SQL Server to run SQL queries to have better insights and have better image about the data.

First I create schema (database). Create a database and I call it "VhclAcc".

--Next, I import the Vehicle and Accidents .CSV files to this database as tables. (Right click on VhclAcc > tasks > import flat file > Refresh

--Now, I am going to answer to some questions with SQL to conduct exploratory data analysis. When possible I will try to use give more than one answer for each question as required:



```sql
-- What is the AVG age of Vehicles involved in accidents based on their type? First of all, I want to see the total number of accidents by Vehicle Type. Also because
 -- The agevehicle column has some NULL Values, so I will need to add a filter
select VehicleType, COUNT(accidentindex) AS 'Total Accidents'
 from vehicle
 where AgeVehicle IS NOT NULL
 group by VehicleType
-- Now as per question requirements I have to calculate the AVG of the Age Vehicle column. To do so, first thing first
 -- I had to change the data type of the agevehicle column from VARCHAR to INTEGER
alter table vehicle
 alter column agevehicle integer
 -- Final complete query
select VehicleType, COUNT(accidentindex) AS 'Total Accidents', AVG(agevehicle) AS 'AVG Age Vehicle'
 from vehicle
 where AgeVehicle IS NOT NULL
 group by VehicleType
 order by 'Total Accidents' DESC
```

97 %

Results | Messages

	VehicleType	Total Accidents	AVG Age Vehicle
1	Car	137379	8
2	Van / Goods 3.5 tonnes mgw or under	9803	6
3	Motorcycle 125cc and under	6669	6
4	Motorcycle over 500cc	5604	10
5	Taxi/Private hire car	4228	6
6	Bus or coach (17 or more pass seats)	4174	7
7	Goods 7.5 tonnes mgw and over	2967	5
8	Motorcycle 50cc and under	1631	6
9	Motorcycle over 125cc and up to 500cc	1545	10
10	Goods over 3.5t. and under 7.5t	763	6
11	Other vehicle	373	7

```sql
select * from accident
 select * from vehicle
select agevehicle from vehicle
 where agevehicle IS NOT NULL
 order by agevehicle DESC

-- Can we identify any trends in accidents based on the age of vehicles involved? (the toughest question)
 -- First answer based off agevehicle range
select vehicletype,
 COUNT(CASE WHEN ageVehicle between 1 and 35 THEN accidentindex ELSE NULL END) AS 'No of Accidents for Vehicles age 1-35',
 COUNT(CASE WHEN ageVehicle between 36 and 70 THEN accidentindex ELSE NULL END) AS 'No of Accidents for Vehicles age 36-70',
 COUNT(CASE WHEN ageVehicle between 71 and 95 THEN accidentindex ELSE NULL END) AS 'No of Accidents for Vehicles age 71-95',
 COUNT(CASE WHEN ageVehicle between 96 and 105 THEN accidentindex ELSE NULL END) AS 'No of Accidents for Vehicles age 96-105'
 from vehicle
 where agevehicle IS NOT NULL
 group by vehicletype
```

97 %

Results | Messages

	vehicletype	No of Accidents for Vehicles age 1-35	No of Accidents for Vehicles age 36-70	No of Accidents for Vehicles age 71-95	No of Accidents for Vehicles age 96-105
1	Agricultural vehicle	299	5	0	0
2	Motorcycle - unknown cc	119	1	0	0
3	Motorcycle 125cc and under	6643	26	0	0
4	Taxi/Private hire car	4228	0	0	0
5	Bus or coach (17 or more pass seats)	4173	1	0	0
6	Motorcycle 50cc and under	1627	4	0	0
7	Mobility scooter	6	0	0	0

```sql
-- Second solution: I want to find the trend between the age of the vehicle and the number of accidents. I will use the IF condition to label the data
-- and after that I want analyse the data. First of all I want to select the COUNT of Accidentindex, then I want to call the AVERAGE age of group. To define the variable
-- called AGEGROUP I will use a Subquery
SELECT Agegroup, COUNT(accidentindex) AS 'Total Number of Accidents', AVG(agevehicle) AS 'Average Year' FROM(
SELECT accidentindex, agevehicle,
    CASE
    WHEN agevehicle BETWEEN 0 AND 5 THEN 'New'
    WHEN agevehicle BETWEEN 6 AND 10 THEN 'Regular'
    ELSE 'Old'
    END AS 'AgeGroup'
FROM vehicle
)AS Subquery
GROUP BY Agegroup
```

97 %

Results | Messages

	Agegroup	Total Number of Accidents	Average Year
1	Old	137141	13
2	Regular	59046	8
3	New	61658	2

```sql
-- Are there any specific weather conditions that contribute to severe accidents? What is the relationship between the severity and the weather condition
select distinct(severity) from accident
select distinct(weatherconditions) from accident
-- FIRST ANSWER in order to reproduce a Pivot Table with severity on rows and No of Accidents for each weather condtions on columns
select severity, count(accidentindex) AS 'Total No of Accidents',
COUNT(CASE WHEN weatherconditions = 'Fog or Mist' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Fog or Mist',
COUNT(CASE WHEN weatherconditions = 'Fine no high winds' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Fine no high winds',
COUNT(CASE WHEN weatherconditions = 'Fine + high winds' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Fine + high winds',
COUNT(CASE WHEN weatherconditions = 'Raining + high winds' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Raining + high winds',
COUNT(CASE WHEN weatherconditions = 'Snowing no high winds' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Snowing no high winds',
COUNT(CASE WHEN weatherconditions = 'Unknown' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Unknown',
COUNT(CASE WHEN weatherconditions = 'Snowing + high winds' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Snowing + high winds',
COUNT(CASE WHEN weatherconditions = 'Raining no high winds' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Raining no high winds',
COUNT(CASE WHEN weatherconditions = 'Other' THEN accidentindex ELSE NULL END) AS 'Number of Accidents Other'
from accident
group by severity
order by count(accidentindex) DESC
```

97 %

Results | Messages

	Total No of Accidents	weatherconditions	severity
1	57141	Fine no high winds	Slight
2	8706	Fine no high winds	Serious
3	7511	Raining no high winds	Slight
4	1164	Unknown	Slight
5	1050	Raining no high winds	Serious
6	1020	Raining + high winds	Slight
7	924	Other	Slight
8	884	Fine + high winds	Slight
9	668	Fine no high winds	Fatal
10	257	Snowing no high winds	Slight

```sql
-- Are there any specific weather conditions that contribute to severe accidents? SECOND ANSWER with a shorter query using aggregation function, GROUP BY and ORDER BY
select count(accidentindex) AS 'Total No of Accidents', weatherconditions, severity from accident
group by weatherconditions, severity
order by count(accidentindex) DESC
```

97 %

Results | Messages

	Total No of Accidents	weatherconditions	severity
1	57141	Fine no high winds	Slight
2	8706	Fine no high winds	Serious
3	7511	Raining no high winds	Slight
4	1164	Unknown	Slight
5	1050	Raining no high winds	Serious
6	1020	Raining + high winds	Slight
7	924	Other	Slight
8	884	Fine + high winds	Slight
9	668	Fine no high winds	Fatal
10	257	Snowing no high winds	Slight
11	251	Fog or mist	Slight

```sql
-- Are there any specific weather conditions that contribute to severe accidents? THIRD ANSWER with an aggregation of Total Accidents
-- Based off all the Severities regardless their type
select count(severity) AS 'Total No of Accidents for All Severity', weatherconditions from accident
group by weatherconditions
order by count(severity) DESC
-- By doing so I did not define any severity I.E. slight, serious or fatal, so If I want to define any specific severity I can add a FILTER. To do so, I am going to
-- DECLARE the VARIABLE severity and because severity is a string I want to put a VARCHAR of 100. In this case I chose to filter by Fatal Accidents
DECLARE @severity VARCHAR(100)
SET @severity = 'Fatal'

select count(severity) AS 'Total No of Accidents', weatherconditions from accident
WHERE severity = @severity
group by weatherconditions
order by count(severity) DESC
```

97 %

Results | Messages

	Total No of Accidents	weatherconditions
1	668	Fine no high winds
2	84	Raining no high winds
3	18	Fine + high winds
4	17	Raining + high winds
5	17	Unknown
6	7	Other
7	5	Fog or mist
8	1	Snowing no high winds