## **MOTOR VEHICLE COLLISION- SQL QUERIES for VISUALIZATION**

1. How many accidents occurred in NYC, Austin, and Chicago?

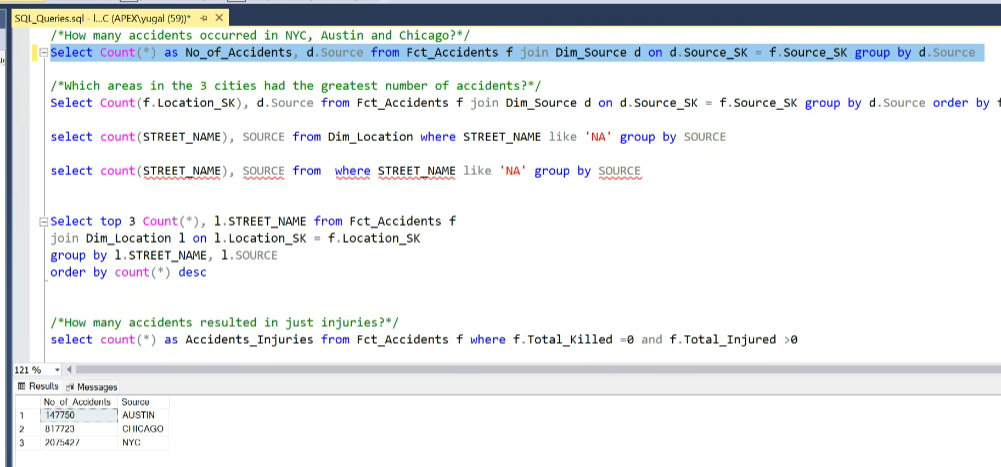
SQL Query:-

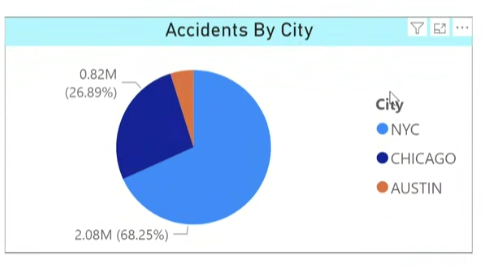
SELECT COUNT (\*) AS Number\_of\_Accidents, d.Source

FROM Fct\_Accidents f

Join Dim\_Source d ON d.Source\_SK = f.Source\_SK

GROUP BY d.Source;





1. Which areas in the 3 cities had the greatest number of accidents? (Top 3 areas in each city).

SQL Query:-

WITH RankedAccidents AS (

SELECT Dim\_Location.STREET\_NAME, Dim\_Location.SOURCE,

COUNT (Fct\_Accidents.Accidents\_FCT\_SK) AS Total Accidents,

ROW\_NUMBER () OVER (PARTITION BY Dim\_Location. SOURCE ORDER BY

COUNT (Fct\_Accidents.Accidents\_FCT\_SK) DESC) AS Rank

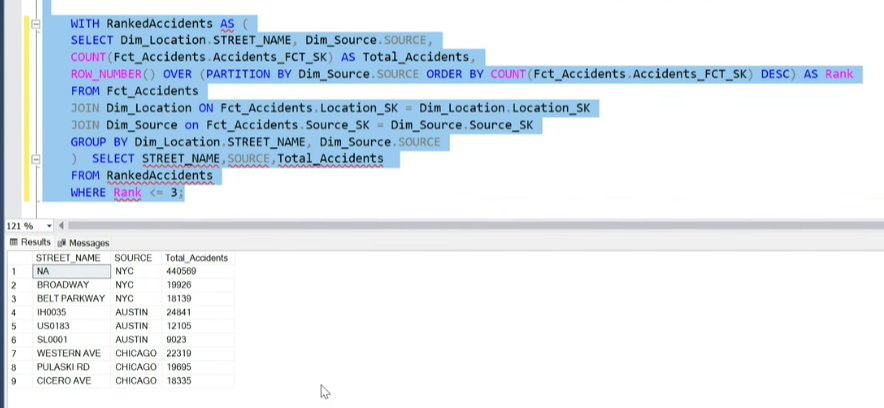
FROM Fct\_Accidents

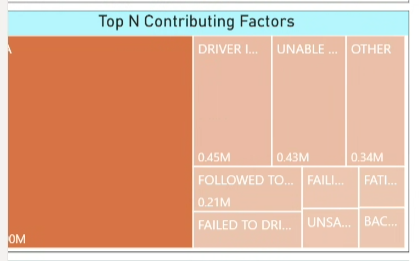
JOIN Dim\_Location ON Fct\_Accidents.Location\_SK = Dim\_Location.Location\_SK

GROUP BY Dim\_Location. STREET\_NAME, Dim\_Location. SOURCE

) SELECT STREET\_NAME, SOURCE, Total Accidents FROM RankedAccidents

WHERE Rank <= 3;





1. How many accidents resulted in just injuries?

* Overall

SQL Query:-

SELECT COUNT (\*) AS Accidents\_Injuries

FROM Fct\_Accidents f

WHERE f.Total\_Killed =0 and f.Total\_Injured >0

* By City

SQL Query:-

SELECT d.Source, COUNT (\*) AS Accidents

FROM Fct\_Accidents f

JOIN Dim\_Source d ON d.Source\_SK = f.Source\_SK

WHERE f.Total\_Killed =0 and f.Total\_Injured >0

GROUP BY d.Source

**OR**

SELECT Source, City\_Accidents\_Injuries, Overall\_Accidents\_Injuries

FROM (

SELECT d.Source,

COUNT (\*) AS City\_Accidents\_Injuries,

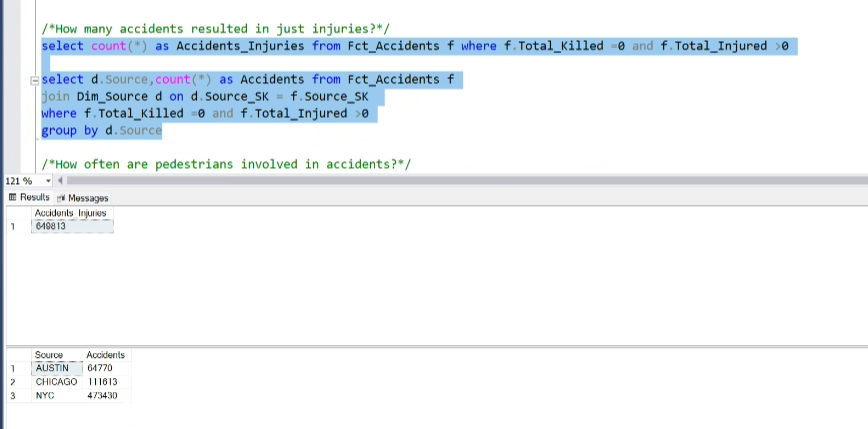
SUM (COUNT (\*)) OVER () AS Overall\_Accidents\_Injuries

FROM Fct\_Accidents f

JOIN Dim\_Source d ON f.Source\_SK = d.Source\_SK

WHERE f.Total\_Killed = 0 AND f.Total\_Injured > 0

GROUP BY d.Source) AS Sub;





1. How often are pedestrians involved in accidents?

* Overall

SQL Query:-

SELECT COUNT (\*) AS Accidents\_Pedestrians FROM Fct\_Accidents f WHERE f.IS\_PEDESTRIAN = 'Y';

* By City

SQL Query:-

SELECT d.Source, COUNT (\*) AS Accidents\_Pedestrians

FROM Fct\_Accidents f

JOIN Dim\_Source d ON d.Source\_SK = f.Source\_SK

WHERE f.IS\_PEDESTRIAN = 'Y'

GROUP BY d.Source

**OR**

SELECT d.Source,

COUNT (\*) AS Accidents\_Pedestrians,

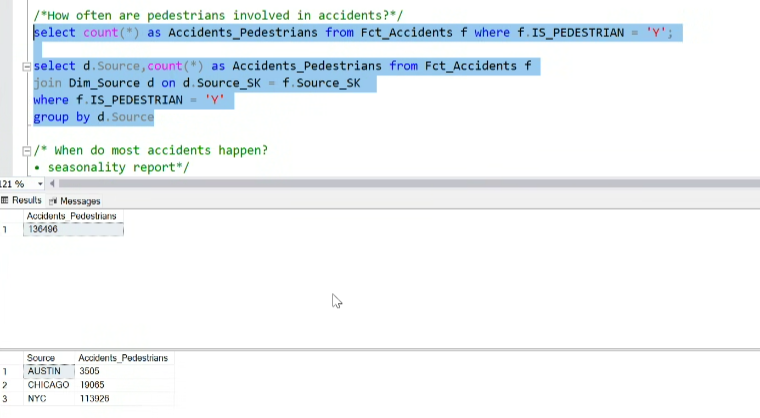
SUM (COUNT (\*)) OVER () AS Total\_Accidents\_Pedestrians

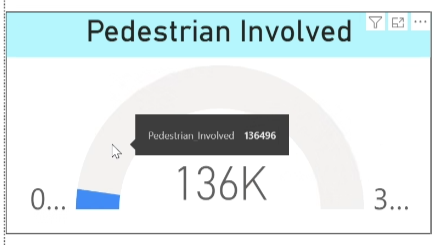
FROM Fct\_Accidents f

JOIN Dim\_Source d ON f.Source\_SK = d.Source\_SK

WHERE f.IS\_PEDESTRIAN = 'Y'

GROUP BY d.Source;





1. When do most accidents happen? (Seasonality Report)

SQL Query:-

SELECT d.Season, COUNT (\*) AS No\_of\_Accidents FROM Fct\_Accidents f

JOIN Dim\_Date d ON f.Date\_SK = d.Date\_SK

GROUP BY d.Season

ORDER BY COUNT (\*) DESC;

SELECT Season, No\_of\_Accidents,

SUM (No\_of\_Accidents) OVER (ORDER BY No\_of\_Accidents DESC) AS Cumulative\_Total,

ROUND (100.0 \* No\_of\_Accidents / SUM (No\_of\_Accidents) OVER (), 2) AS Percentage\_of\_Total

FROM (

SELECT d.Season, COUNT (\*) AS No\_of\_Accidents

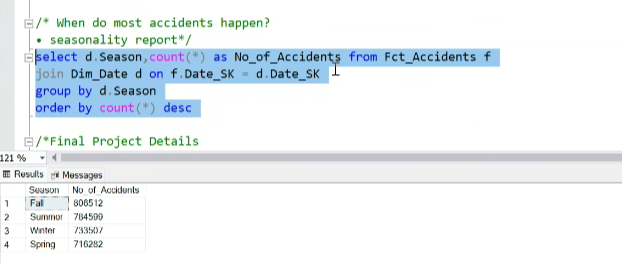
FROM Fct\_Accidents f

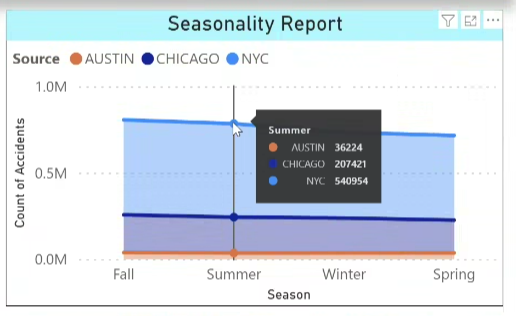
JOIN Dim\_Date d ON f.Date\_SK = d.Date\_SK

GROUP BY d.Season

) AS SeasonData

ORDER BY No\_of\_Accidents DESC;





1. How many motorists are injured or killed in accidents?

SQL Query:-

SELECT

SUM (Motorist\_Injured) AS Total\_Motorist\_Injuries,

SUM (Motorist\_Killed) AS Total\_Motorist\_Fatalities

FROM Fct\_Accidents;

SELECT s.Source,

SUM (Motorist\_Injured) AS Total\_Motorist\_Injuries,

SUM (Motorist\_Killed) AS Total\_Motorist\_Fatalities

FROM Fct\_Accidents f

JOIN Dim\_Source s ON f.Source\_SK = s.Source\_SK

GROUP BY S.Source; /\* Chicago doesn't have bifurcation for motorists\*/

SELECT s.Source,

SUM (f.Motorist\_Injured) AS Total\_Motorist\_Injuries,

SUM (f.Motorist\_Killed) AS Total\_Motorist\_Fatalities,

SUM (SUM (f.Motorist\_Injured)) OVER () AS Overall\_Motorist\_Injuries,

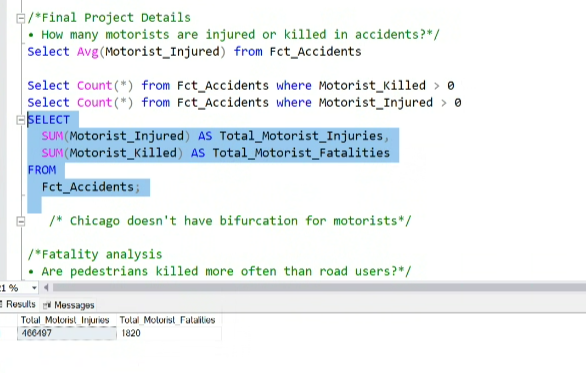
SUM (SUM (f.Motorist\_Killed)) OVER () AS Overall\_Motorist\_Fatalities

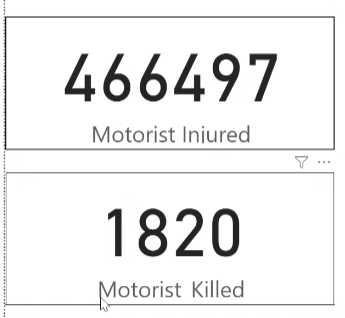
FROM Fct\_Accidents f

JOIN Dim\_Source s ON f.Source\_SK = s.Source\_SK

GROUP BY s.Source

ORDER BY s.Source;





1. Which top 5 areas in 3 cities have the most fatal accidents?

SQL Query:-

SELECT top 5 COUNT (\*), l.STREET\_NAME, l.SOURCE FROM Fct\_Accidents f

JOIN Dim\_LocatiON l ON l.LocatiON\_SK = f.LocatiON\_SK

GROUP BY l.STREET\_NAME, l.SOURCE

HAVING l.STREET\_NAME NOT LIKE 'NA'

ORDER BY COUNT (\*) DESC /\*Given NYC's NA COUNT of street name is ignored\*/

SQL Query:-

WITH RankedAccidents AS (

SELECT Dim\_LocatiON.STREET\_NAME, Dim\_Source.SOURCE,

COUNT (Fct\_Accidents.Accidents\_FCT\_SK) AS Total\_Accidents,

ROW\_NUMBER () OVER (PARTITION BY Dim\_Source.SOURCE

ORDER BY COUNT (Fct\_Accidents.Accidents\_FCT\_SK) DESC) AS Rank

FROM Fct\_Accidents

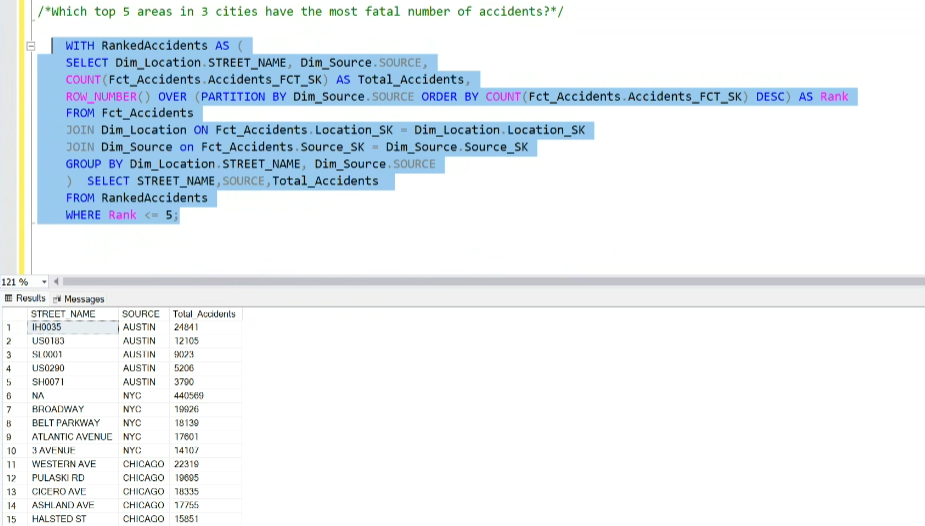
JOIN Dim\_LocatiON ON Fct\_Accidents.LocatiON\_SK = Dim\_LocatiON.LocatiON\_SK

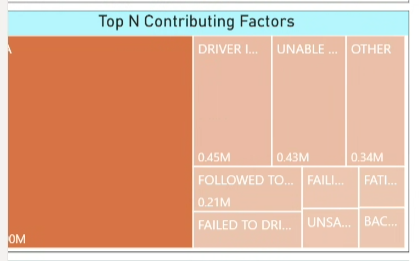
JOIN Dim\_Source ON Fct\_Accidents.Source\_SK = Dim\_Source.Source\_SK

GROUP BY Dim\_LocatiON.STREET\_NAME, Dim\_Source.SOURCE

) SELECT STREET\_NAME, SOURCE, Total\_Accidents

FROM RankedAccidents WHERE Rank <= 5;





1. Time-based analysis of accidents (Time of the day, day of the week, weekdays or weekends).

* Day of the Week

SELECT d.Day, COUNT (\*) AS No\_of\_Accidents

FROM Fct\_Accidents f

JOIN Dim\_Date d ON f.Date\_SK = d.Date\_SK

GROUP BY d.Day

ORDER BY No\_of\_Accidents;

* Weekend vs. Weekday Accidents

-- Weekend Accidents

SELECT COUNT (\*) AS Weekend\_Accidents

FROM Fct\_Accidents f

JOIN Dim\_Date d ON f.Date\_SK = d.Date\_SK

WHERE d.Day > 5;

-- Weekday Accidents

SELECT COUNT (\*) AS Weekday\_Accidents

FROM Fct\_Accidents f

JOIN Dim\_Date d ON f.Date\_SK = d.Date\_SK

WHERE d.Day <= 5;

* Hour of the Day

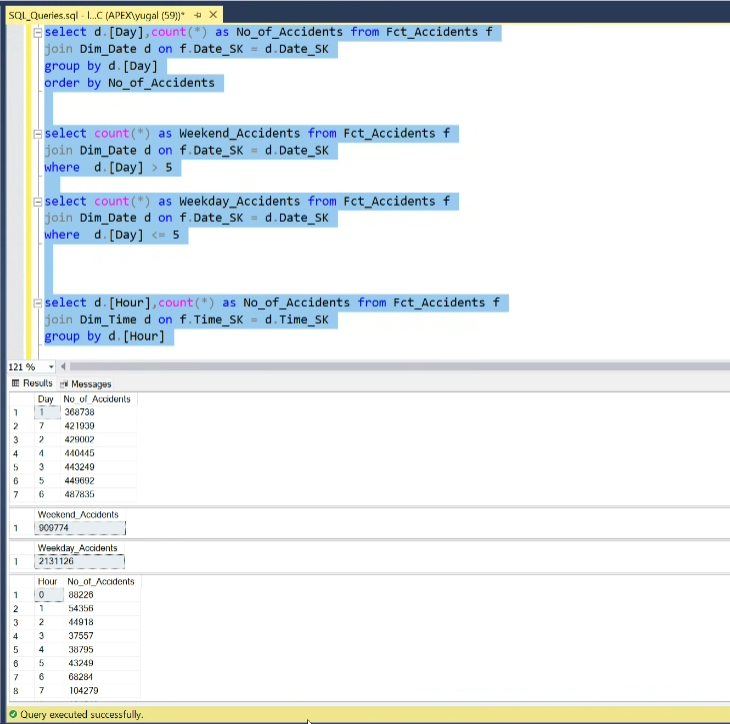
SELECT d.Hour, COUNT (\*) AS No\_of\_Accidents

FROM Fct\_Accidents f

JOIN Dim\_Time d ON f.Time\_SK = d.Time\_SK

GROUP BY d.Hour

ORDER BY d.Hour;



USE FILTERS ON VISUALIZATION

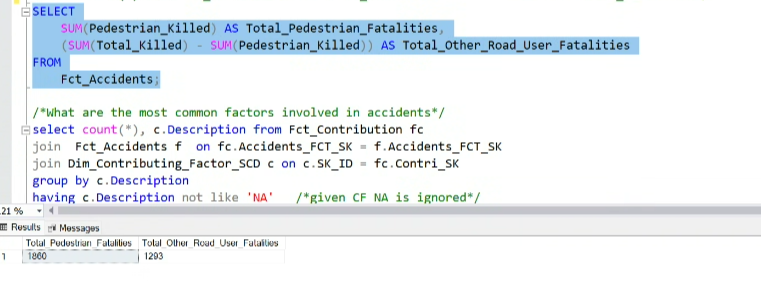
1. Are pedestrians killed more often than road users? (Fatality analysis)

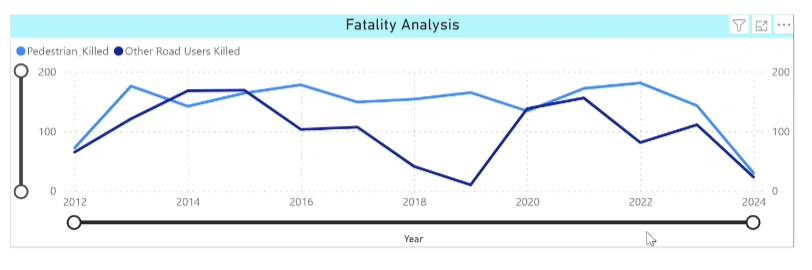
SELECT

SUM (Pedestrian\_Killed) AS Total\_Pedestrian\_Fatalities,

(SUM (Total\_Killed) – SUM (Pedestrian\_Killed)) AS Total\_Other\_Road\_User\_Fatalities

FROM Fct\_Accidents;





1. What are the most common factors involved in accidents?

SELECT

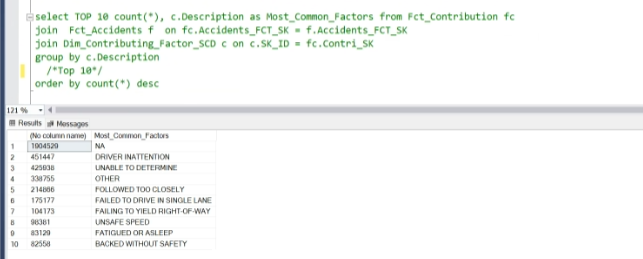
COUNT (\*) Number, c.DescriptiON AS Most\_Common\_Factors FROM Fct\_Contribution fc

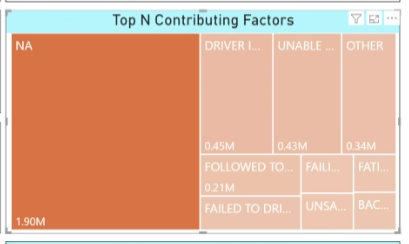
JOIN Fct\_Accidents f ON fc.Accidents\_FCT\_SK = f.Accidents\_FCT\_SK

JOIN Dim\_Contributing\_Factor\_SCD c ON c.SK\_ID = fc.Contri\_SK

GROUP BY c.Description

ORDER BY COUNT (\*) DESC;





/\* ADDITIONAL REQURIREMENT\*/

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

SELECT

Count (Vehicle\_Fct\_Sk) AS Number\_of\_Vehicle, s.Source

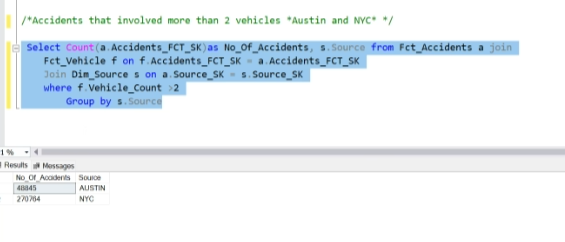
FROM Fct\_Accidents a

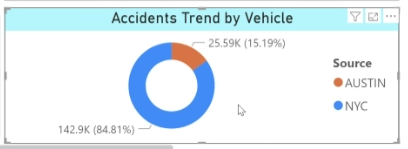
JOIN Fct\_Vehicle f ON f.Accidents\_FCT\_SK = a.Accidents\_FCT\_SK

JOIN Dim\_Source s ON a.Source\_SK = s.Source\_SK

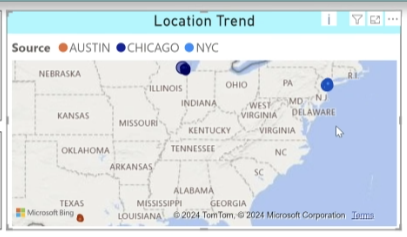
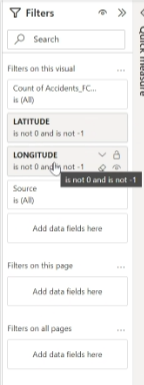
WHERE f.Vehicle\_Count >2

GROUP BY s.Source

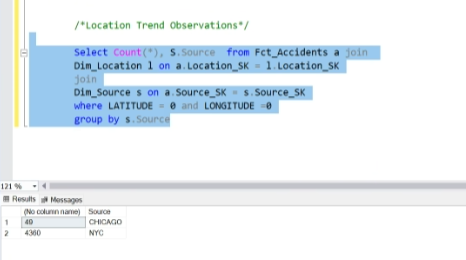


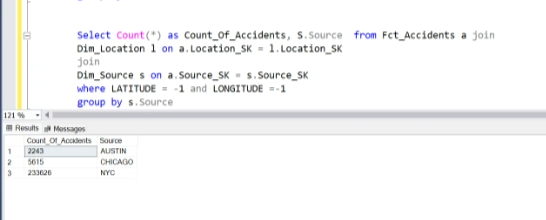


/\* LOCATION TREND\*/

* Ignoring 0 and 1 values for latitude longitude as they are outlier.
* Latitude longitude combination (0,0) is present in New York.
* Latitude longitude combination (-1,-1) is to replace missing values.





## **PowerBI VISUALIZATION**

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

## **Tableau VISUALIZATION**

A screenshot of a computer screen

Description automatically generated

