Final Project Paper

Spatial Database Design and Administration

Name: Deepika Mallika Tamilselvan

Notice: Dr. David Haynes

Data: 02/05/2023

Topic: Traffic Accidents analysis through Spatial Analysis

1. **Introduction:**

Traffic accidents pose a severe threat to human life and a major problem in many urban areas around the world, leading to fatalities, injuries, and property damage. Since traffic crashes also cause economic loss, it affects the Gross Domestic Product of a country. Hence, efforts should be taken to minimize accidents. Hotspots, also known as “black spots”. A spatial analysis of traffic accidents can provide valuable insights into the spatial distribution and underlying causes of accidents, which can inform strategies for improving traffic safety. Through this project, I propose to create a spatial database that stores information on traffic accidents, analyses the spatial distribution of traffic accidents, and identifies areas with high numbers of accidents.

2. Motivation:

Here are some key factors that generate my interest in Accident analysis. The report says road traffic injuries are the leading cause of death for children and young adults aged 5-29 years. Approximately 1.3 million people die each year because of road traffic crashes. More than half of all road traffic deaths are among vulnerable road users: pedestrians, cyclists, and motorcyclists. 93% of the world's fatalities on the roads occur in low- and middle-income countries, even though these countries have approximately 60% of the world's vehicles. Road traffic crashes cost most countries 3% of their gross domestic product. So the United Nations General Assembly has set an ambitious target of halving the global number of deaths and injuries from road traffic crashes by 2030. I believe that spatial database analysis will lower the risk of road accidents. This project is to identify areas with high numbers of traffic accidents and investigate potential causes of these accidents. By utilizing spatial data, we can better understand the relationship between traffic accidents and various spatial factors, such as road geometry, traffic flow, and land use. The results of this analysis can inform the development of traffic safety strategies and interventions to reduce the number of traffic accidents. This study will focus only on Minnesota.

Age Group	Male Drivers in Fatal Crashes	Female Drivers in Fatal Crashes	Unk Gender Drivers in Fatal Crashes	Total in Fatal Crashes	Male Drivers in All Crashes	Female Drivers in All Crashes	Unk Gender Drivers in All Crashes	Total in All Crashes
<15	0	0	0	0	47	17	1	65
15-19	30	14	1	45	5,435	4,292	58	9,785
20-24	32	17	0	49	6,815	4,990	124	11,929
25-29	45	13	1	59	6,317	4,175	97	10,589
30-34	39	13	0	52	5,822	3,666	77	9,565
35-39	28	12	1	41	5,186	3,343	45	8,574
40-44	31	10	0	41	4,374	2,700	35	7,109
45-49	33	21	0	54	3,741	2,273	28	6,042
50-54	34	7	0	41	3,837	2,133	19	5,989
55-59	46	6	0	52	3,829	2,140	16	5,985
60-64	32	8	0	40	3,195	1,764	12	4,971
65-69	23	4	0	27	2,247	1,302	6	3,555
70-74	13	4	0	17	1,503	952	7	2,462
75-79	14	1	0	15	968	597	1	1,566
80-84	8	2	0	10	555	438	4	997
85+	7	1	0	8	399	256	7	662
Unk	0	0	0	0	21	6	31	58
Total	**415**	**133**	**3**	**551**	**54,291**	**35,044**	**568**	**89,903**

Figure 2.1 Statistical report on age and gender

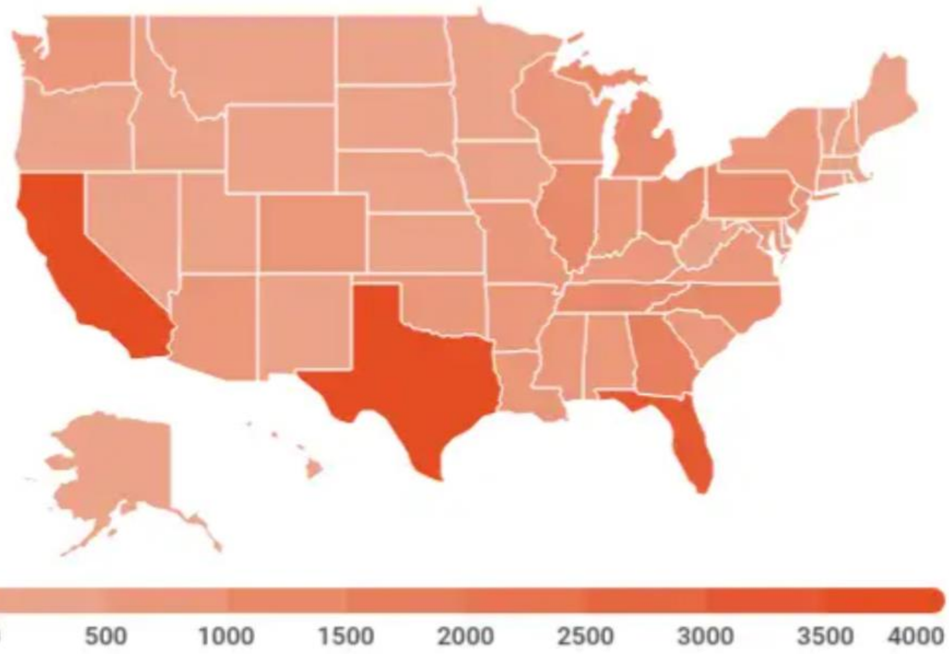


Figure 2.1 US Traffic accident

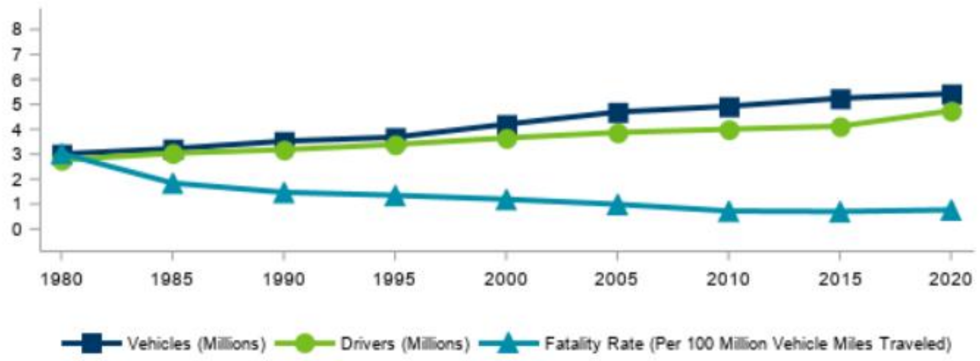
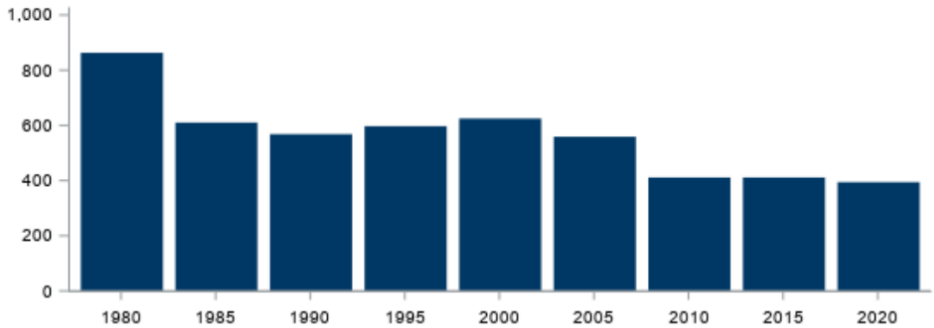


Figure 2.3 Vehicles, Drivers, and Fatality Rate



Department of Public Safety, Office of Traffic Safety - Minnesota Motor Vehicle Crash Facts, 2020

Figure 2.4 Minnesota Traffic Fatalities

	2016	2017	2018	2019	2020	Record High	
Fatal Crashes	**357**	**341**	**349**	**333**	**369**	**878**	**1973**
Injury Crashes	**21,734**	**21,272**	**20,244**	**19,902**	**15,071**	**33,868**	**1978**
Serious	1,702	1,561	1,341	1,297	1,310	5,109	1984
Minor	8,642	8,199	7,327	7,260	5,940	12,326	1985
Possible	11,390	11,512	11,576	11,345	7,821	18,578	1996
PDO Crashes	**56,978**	**56,852**	**58,622**	**60,401**	**41,687**	**94,810**	**1975**
Total Crashes	**79,069**	**78,465**	**79,215**	**80,636**	**57,127**	**123,106**	**1975**
Total Injuries	**29,825**	**29,412**	**27,877**	**27,260**	**20,529**	**50,332**	**1978**
Serious	1,992	1,849	1,660	1,520	1,569	6,573	1984
Minor	11,097	10,539	9,429	9,346	7,656	17,670	1985
Possible	16,736	17,024	16,788	16,394	11,304	28,631	1996
Total Fatalities	**392**	**358**	**381**	**364**	**394**	**1,060**	**1968**
MN Fatality Rate	0.67	0.63	0.63	0.60	0.76	24	1934
U.S. Fatality Rate	1.18	1.25	1.25	1.1	1.37	18	1925
MN Economic Loss (billions)	$1.87	$1.79	$1.79	$1.87	$1.87	$1.87	2016

Figure 2.5 Minnesota Traffic Trends

3. Datasets

There were four tables and 2 shape files used in this project for accidents different CSV data has been collected from different sources like mn.gov and NHTSA CRSS Accident records and Google API can be used to geocode the accident locations. Minnesota shapefile was downloaded from mn geospatial Commons.

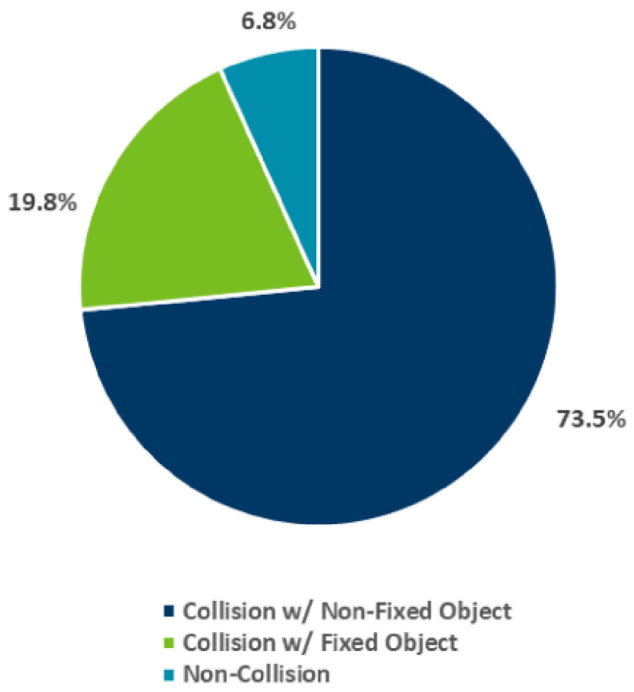


Figure 3.1 Minnesota Traffic Crash Type

4. ER Diagram

An entity-relationship (ER) diagram is a graphical representation of entities and the relationships between the tables and how the tables are connected to one another here in this project accident_id is the primary key and the location_id is the foreign key. There are 5 tables used here they are: Accident_table, Accident_factors, Factors, Accident_locations and mn_counties.

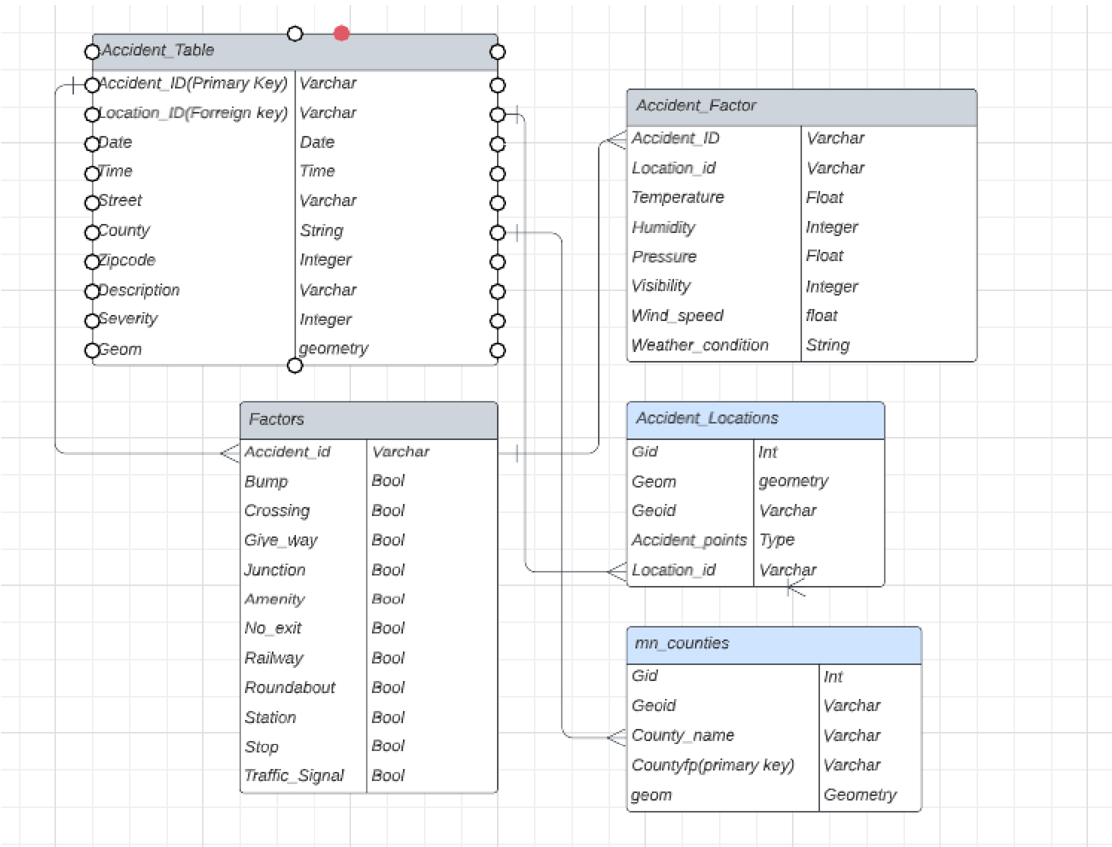


Figure 4.1 ER-Diagram

5. Methodology

Data Collection: The first step is collecting data on traffic accidents from reliable sources such as mn.gov and NHTSA CRSS Accident Records. The data has the location of the accident, the time of the accident, the type of accident, description, wind speed, and pressure. To enhance the location data, Google API can be used to geocode the accident locations. Data Compilation: Once the data has been collected, it needs to be compiled into a single dataset. The data should be cleaned to remove any duplicates, incomplete or incorrect data. It is important to ensure that the data is consistent and properly formatted for use in the spatial analysis. Spatial Analysis Software: This project uses specialized spatial analysis software such as ArcGIS Pro to visualize the accident point feature. The software should be able to handle large datasets and perform a variety of spatial analysis techniques such as density mapping, clustering analysis, and spatial regression analysis. To identify potential contributing factors to the high number of accidents in the hotspots.

6. Results

Based on my analysis of traffic accidents in the state of Minnesota, I found that there were a total of 31185 accidents reported in the dataset from 2016 to 2021. The counties with the highest number of accidents were Hennepin County, Ramsey County, and Dakota County, with 9156, 3997, and 1792 accidents respectively. Also found that the primary cause of accidents happened on the junctions there were 3946 accidents happening here. Next to that traffic signal took 1234 accidents. Streets like I-94 E and I-94 W have the highest number of accidents. Finally, analyzed the monthly trends in traffic accidents and found that Jan and Feb had the lowest number of accidents. Summer months like Jun, July, August, and September had a comparatively high number of accidents. The winter months of October, November, and December had the highest number of accidents, likely due to weather conditions.

```sql
WITH accident_counts AS (
  SELECT
    counties.name,
    COUNT(accidents.*) AS num_accidents
  FROM
    accident_locations AS accidents
    JOIN mn_counties AS counties
    ON ST_Intersects(accidents.geom, counties.geom)
  GROUP BY counties.name
),
top_counties AS (
  SELECT
    name,
    num_accidents,
    ROW_NUMBER() OVER (ORDER BY num_accidents DESC) AS rank
  FROM accident_counts
),
top_severity AS (
  SELECT
    c.name AS county_name,
    AVG(a.severity) as avg_severity
  FROM
    accident_table a
    JOIN accident_locations l ON a.location_id = a.location_id
    JOIN mn_counties c ON ST_Intersects(l.geom, c.geom) AND ST_Intersects(c.geom, l.geom)
  GROUP BY c.name
),
overall_avg_severity AS (
  SELECT AVG(a.severity) AS avg_severity FROM accident_table a
)
```

```sql
  GROUP BY c.name
),
overall_avg_severity AS (
  SELECT AVG(a.severity) AS avg_severity FROM accident_table a
)
SELECT
  top_counties.name AS county_name,
  top_counties.num_accidents,
  top_severity.avg_severity AS county_avg_severity,
  overall_avg_severity.avg_severity AS overall_avg_severity
FROM
  top_counties
  JOIN mn_counties AS counties ON top_counties.name = counties.name
  LEFT JOIN top_severity ON top_counties.name = top_severity.county_name
  CROSS JOIN overall_avg_severity
ORDER BY top_counties.rank;
```

mn_counties 1

WITH accident_counts AS (SELECT counties.name, C(Enter a SQL expression to filter results (use Ctrl-

Grid | Text | Record

county_name	num_accidents	county_avg_severity	overall_avg_severity
Hennepin	9156	2.0283790283790284	2.0283790283790284
Ramsey	3998	2.0283790283790284	2.0283790283790284
Dakota	1792	2.0283790283790284	2.0283790283790284
Washington	1320	2.0283790283790284	2.0283790283790284
Anoka	1175	2.0283790283790284	2.0283790283790284
St. Louis	920	2.0283790283790284	2.0283790283790284
Wright	853	2.0283790283790284	2.0283790283790284
Stearns	813	2.0283790283790284	2.0283790283790284
Olmsted	569	2.0283790283790284	2.0283790283790284
Sherburne	514	2.0283790283790284	2.0283790283790284
Scott	495	2.0283790283790284	2.0283790283790284
Rice	414	2.0283790283790284	2.0283790283790284
Pine	413	2.0283790283790284	2.0283790283790284
Chisago	393	2.0283790283790284	2.0283790283790284
Douglas	364	2.0283790283790284	2.0283790283790284
Goodhue	353	2.0283790283790284	2.0283790283790284
Blue Earth	349	2.0283790283790284	2.0283790283790284
Carlton	325	2.0283790283790284	2.0283790283790284
Winona	312	2.0283790283790284	2.0283790283790284
Carver	307	2.0283790283790284	2.0283790283790284
Crow Wing	299	2.0283790283790284	2.0283790283790284
Benton	269	2.0283790283790284	2.0283790283790284
Morrison	267	2.0283790283790284	2.0283790283790284
Freeborn	266	2.0283790283790284	2.0283790283790284
Nicollet	261	2.0283790283790284	2.0283790283790284
Otter Tail	257	2.0283790283790284	2.0283790283790284
Itasca	248	2.0283790283790284	2.0283790283790284
Steele	233	2.0283790283790284	2.0283790283790284
Cass	220	2.0283790283790284	2.0283790283790284
Kandiyohi	190	2.0283790283790284	2.0283790283790284
Todd	185	2.0283790283790284	2.0283790283790284
Mille Lacs	181	2.0283790283790284	2.0283790283790284
Beltrami	172	2.0283790283790284	2.0283790283790284
Becker	170	2.0283790283790284	2.0283790283790284
McLeod	156	2.0283790283790284	2.0283790283790284

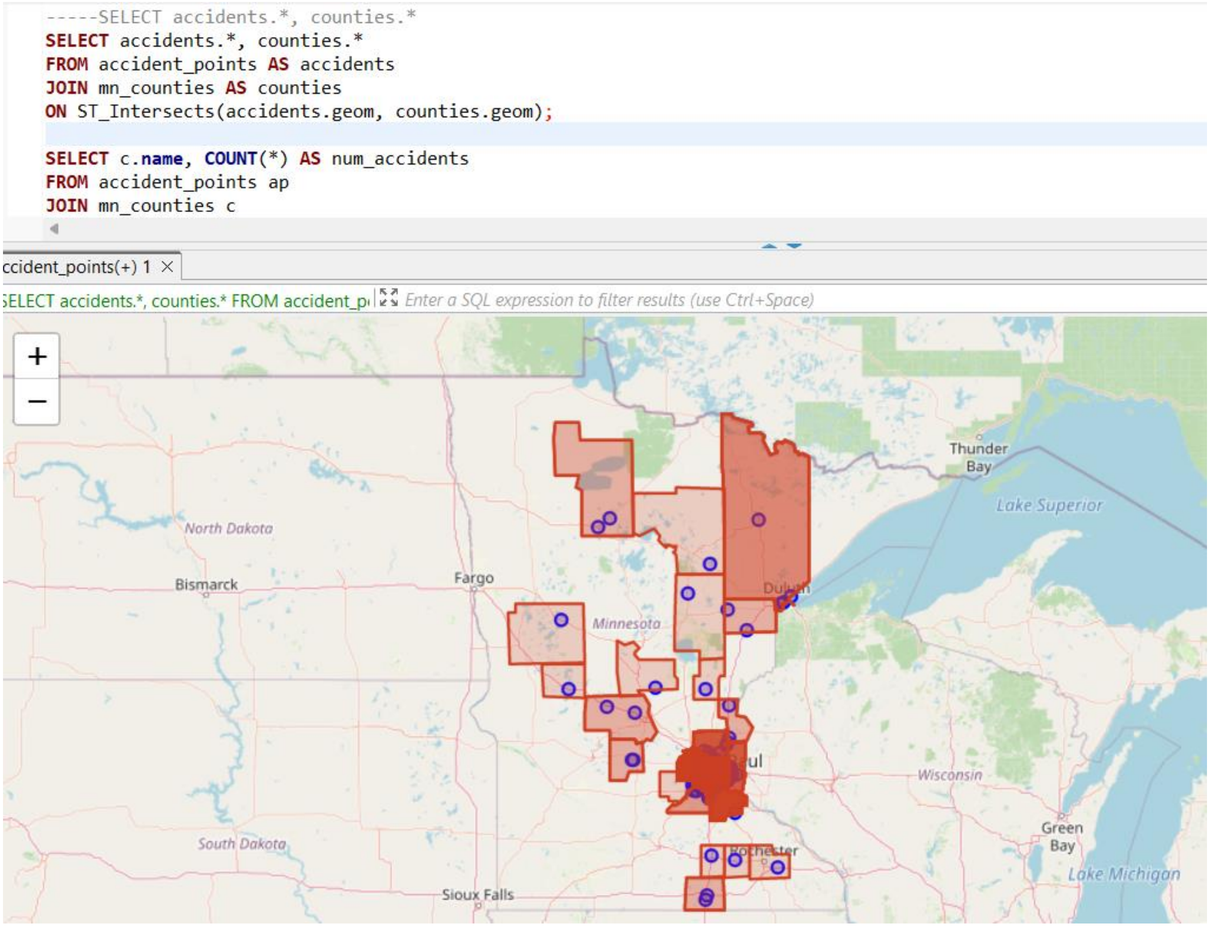
Month	Fatal Crashes	Injury Crashes	PDO Crashes	Total Crashes	Killed	Injured
January	16	1,475	5,813	7,304	18	2,019
February	20	1,276	4,451	5,747	22	1,698
March	24	842	2,464	3,330	24	1,132
April	19	660	1,862	2,541	20	883
May	25	963	2,151	3,139	27	1,315
June	34	1,413	2,951	4,398	37	1,959
July	50	1,562	3,222	4,834	54	2,193
August	46	1,434	3,214	4,694	47	1,969
September	45	1,452	3,315	4,812	47	2,006
October	28	1,519	4,355	5,902	35	2,044
November	33	1,220	3,486	4,739	34	1,611
December	29	1,255	4,403	5,687	29	1,700
Total	**369**	**15,071**	**41,687**	**57,127**	**394**	**20,529**

```sql
SELECT EXTRACT(MONTH FROM date) AS month, COUNT(*) AS num_accidents
FROM accident_table
GROUP BY month
ORDER BY month ASC;
```

Results 1

SELECT EXTRACT(MONTH FROM date) AS montl | Enter a SQL expression to filter results (use Ctrl+Space)

```
month|num_accidents|
-----+-------------+
    1|          481|
    2|          324|
    3|         1570|
    4|         1509|
    5|         2053|
    6|         2942|
    7|         2405|
    8|         2454|
    9|         2441|
   10|         2734|
   11|         3599|
   12|         8673|
```



7. Limitations

Since this was a huge dataset maintenance cleaning took more time than I anticipated Collecting the dataset difficult some dataset that I used previously for this stugy doesn't have all the data which I need for this study. If the accident data is incomplete, the analysis may not be accurate or representative of the actual accident rates. For example, if some accidents are not reported, the results may not capture the full extent of the problem. Some of the data is not accurate or consistent, the analysis may be affected. For example, if there are errors in the location or date of accidents, it may be difficult to accurately determine accident rates. The accuracy and detail of the results may be limited by the spatial and temporal resolution of the data. For example, if the accident data is only available at a county level, it may not be possible to identify accident hotspots in specific neighborhoods or streets.

Reference:
1. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. https://carto.com/blog/predicting-traffic-accident-hotspots-with-spatial-data-science
3. https://ejmcm.com/article_4548_078ada5ba8f7220898cdf51c294f413f.pdf
4. https://www.ijeat.org/wp-content/uploads/papers/v9i2/B3848129219.pdf
5. https://www.policygenius.com/auto-insurance/car-accidents-by-state/
6. https://dps.mn.gov/divisions/ots/reports-statistics/Documents/2020-minnesota-motor-vehicle-crash-facts.pdf

Other Queries:

```
-----SELECT accidents.*, counties.*
SELECT accidents.*, counties.*
FROM accident_points AS accidents
JOIN mn_counties AS counties
ON ST_Intersects(accidents.geom, counties.geom);

-----Calculate the percentage of accidents that occur during rush hour (defined as 7-9am and 4-6pm on weekdays)
WITH rush_hour_accidents AS (
SELECT
Location_ID,
Date,
CASE
WHEN DATE_PART('hour', Time) >= 7 AND DATE_PART('hour', Time) <= 9 AND DATE_PART('dow', Date) >= 1 AND DATE_PART('dow', Date) <= 5
OR DATE_PART('hour', Time) >= 16 AND DATE_PART('hour', Time) <= 18 AND DATE_PART('dow', Date) >= 1 AND DATE_PART('dow', Date) <= 5
THEN 1
ELSE 0
END AS rush_hour
FROM Accident_table
)
SELECT
Date,
COUNT(*) AS total_accidents,
SUM(rush_hour) AS rush_hour_accidents,
100.0 * SUM(rush_hour) / COUNT(*) AS percentage_rush_hour_accidents
FROM rush_hour_accidents
GROUP BY Date
ORDER BY Date;
```