

Analysis of The CRSS

By Jacob Haslett

To start out we'll look at some simple facts. As we progress we'll dive deeper into the different elements and topics of the Crash Report Sampling System (CRSS). Now remember, not all accidents are reported. Estimates are that more than six million accidents occur across the United States every year and most are never reported to local law enforcement for various and unknown reasons.

To begin, let's look at the count of accidents that have been reported per year.

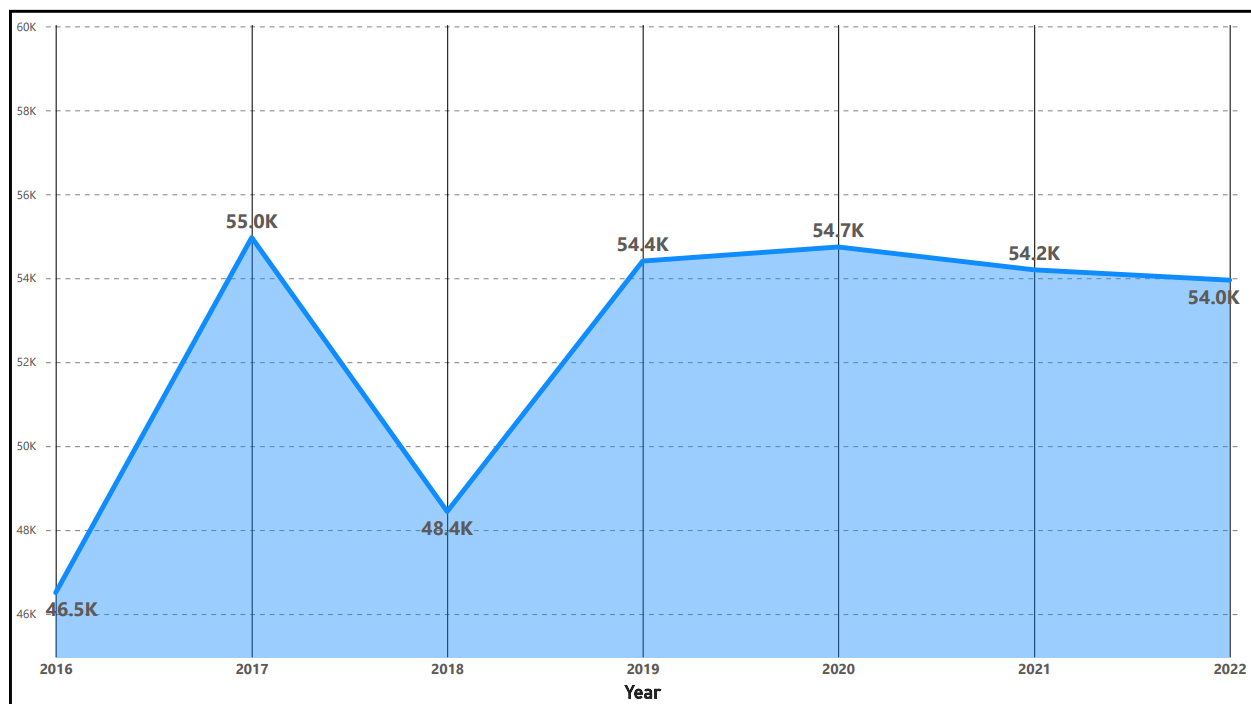


Image Acc-01

What we see here is that 2016 had the least amount of accidents with a total of 46.5 thousand. The most recorded accidents took place in 2017 with 55 thousand, an increase of 8.5 thousand, or 18.279%. Calculating the average, we see approximately 52.5 thousand accidents per year. I anticipated to see 2020 as the year of least accidents, taking into consideration the “work from home” movement and quarantine policies. Instead 2020 came out as the year with the second highest count for accidents with a count of 54.7 thousand, trailing 2017 by roughly 300 accidents nationwide.

This is just a count of reported accidents. Later on we'll look into finer details such as number of vehicles, people in each vehicle, or not in a vehicle at all. But, for now let's break down this data a little more. No where near the estimated six million, but, for me these numbers were still surprising. I started to wonder how these numbers would look broken up into a finer

granularity. Perhaps there's a month or even a season with an increase in accidents. Let's find out...

Accident count per quarter:

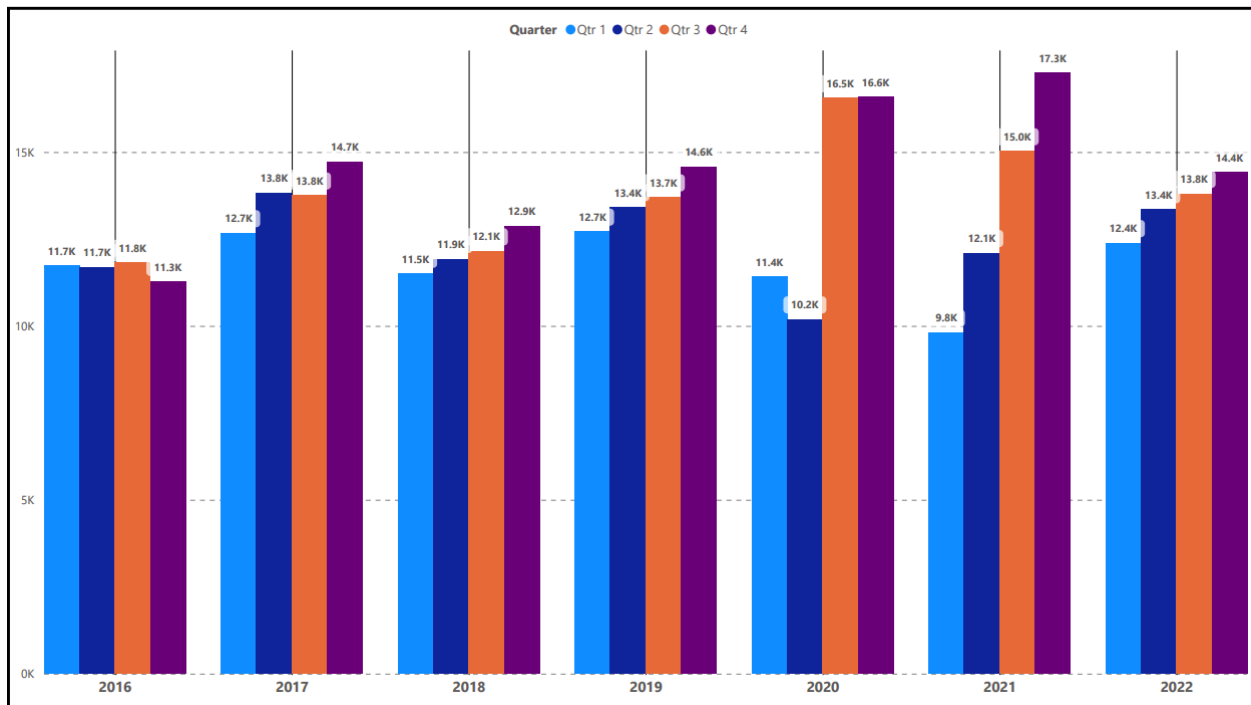


Image Acc-02

With the exception of 2016, it appears the fourth quarter of each year has the highest count of accidents. I assumed there would be more accidents during the winter months. However, the first quarter of each year has less accidents. Considerably less during 2020 and 2021. Let's see what kind of numbers we'll see if we drill further into a monthly breakdown.

ADDITIONAL FACTS:

<u>Year</u>	<u>Min Qtr</u>	<u>Min Count</u>	<u>Max Qtr</u>	<u>Max Count</u>	<u>Difference Count</u>	<u>Difference %</u>	<u>Avg per Qtr</u>
2016	4th	11,267	3rd	11,815	548	4.86%	11,627.75
2017	1st	12,668	4th	14,718	2,050	16.18%	13,742.25
2018	1st	11,500	4th	12,870	1,370	11.91%	12,110.75
2019	1st	12,713	4th	14,583	1,870	14.71%	13,602.25
2020	2nd	10,189	4th	16,587	6,398	62.79%	13,686.25
2021	1st	9,808	4th	17,272	7,464	76.1%	13,550.00
2022	1st	12,383	4th	14,416	2,033	16.42%	13,488.75

Month with Most Accidents:

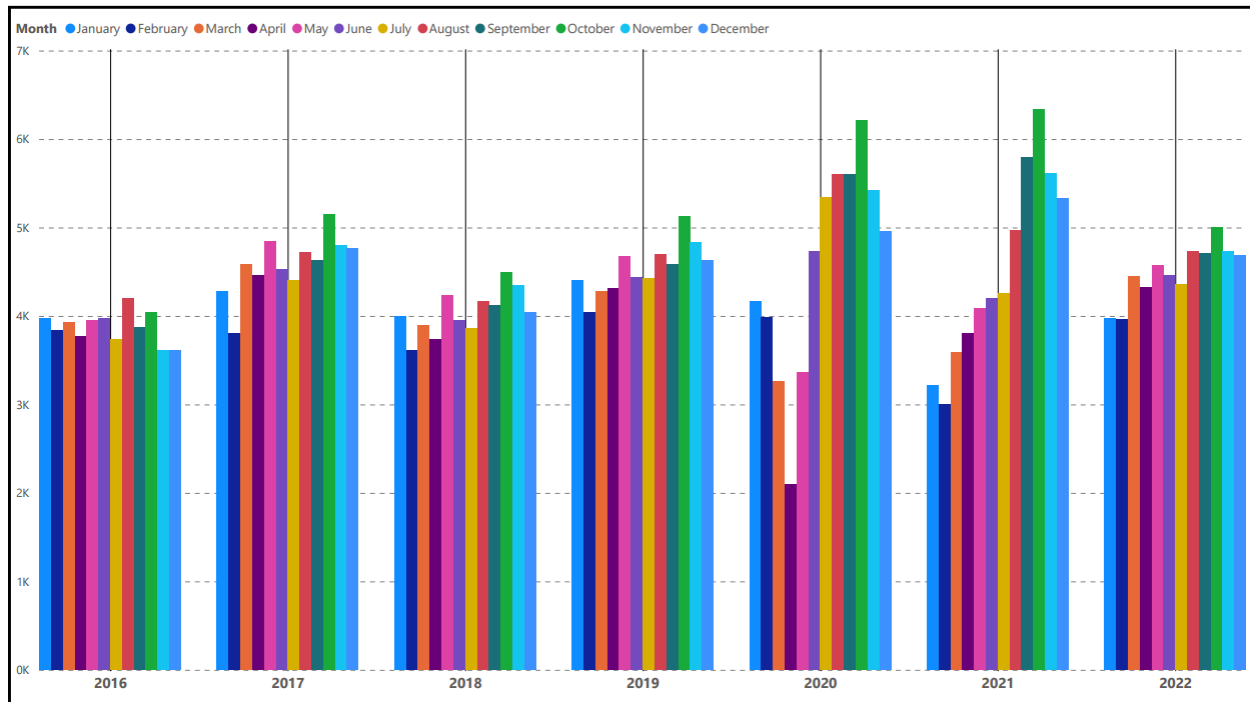


Image Acc-03

Well apparently October has the most accidents, with the exception of 2016 which saw the most accidents in August. Living in New England (known for freezing winter weather) I had always thought the most accidents would happen in the winter time as a result of the snowy and icy conditions. This brings up the question of what conditions see the most accidents?

Weather Conditions During Accidents:

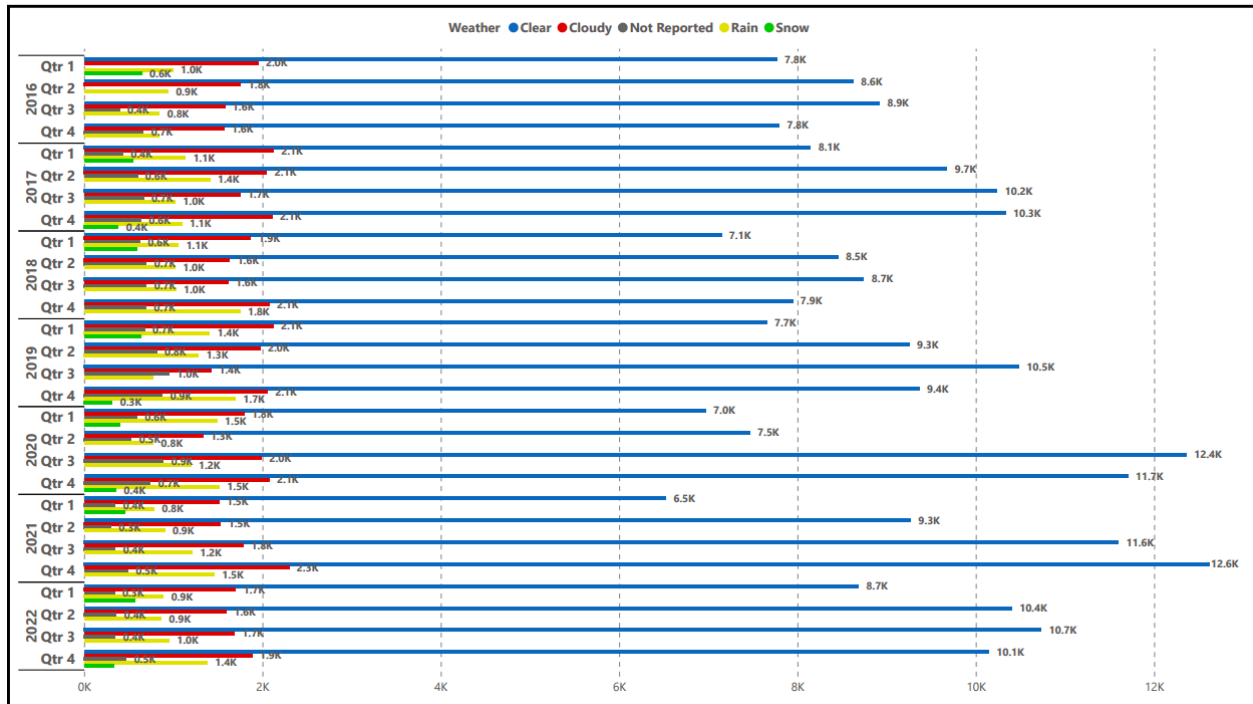


Image Acc-04

Well it seems weather conditions have very little influence on accidents. We can see that accidents happen in clear weather a significant amount more than other conditions. I applied a filter to only display counts of more than 300 for each condition to only show the top five.

Weather Condition Percentages:

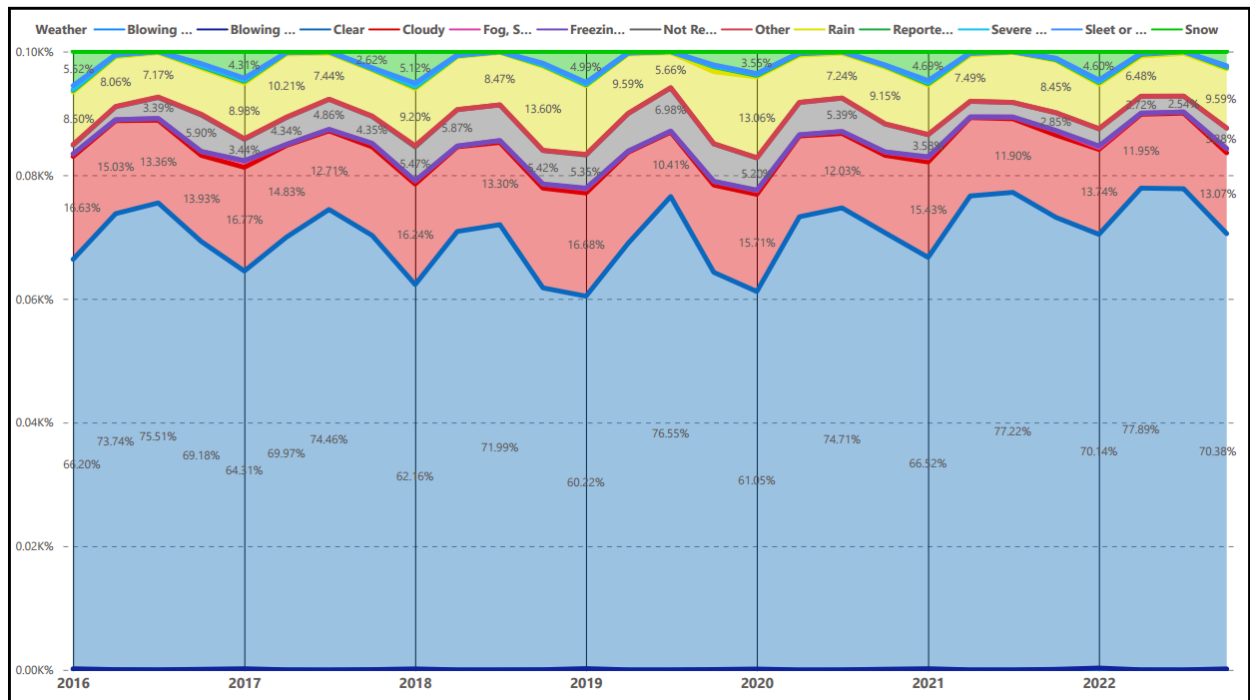


Image Acc-05

Removing the filter and looking at percentages, we can see that accidents happen during clear skies in a range of 60.22% to 77.89%. On average clear skies make up 69.78%, with cloudy coming in at 13.7%, third place goes to rain at 8.56% and continuing on with snow 2.48%, fog/smog/smoke 0.4%, sleet/hail 0.21%, blowing snow 0.14%, freezing rain 0.09%, severe crosswinds 0.09%, and blowing sand/soil/dirt 0.03%. Additionally 0.07% was reported as “other”, 0.24% reported as “unknown”, and 4.22% is not reported at all.

Accidents During Snow by Region:

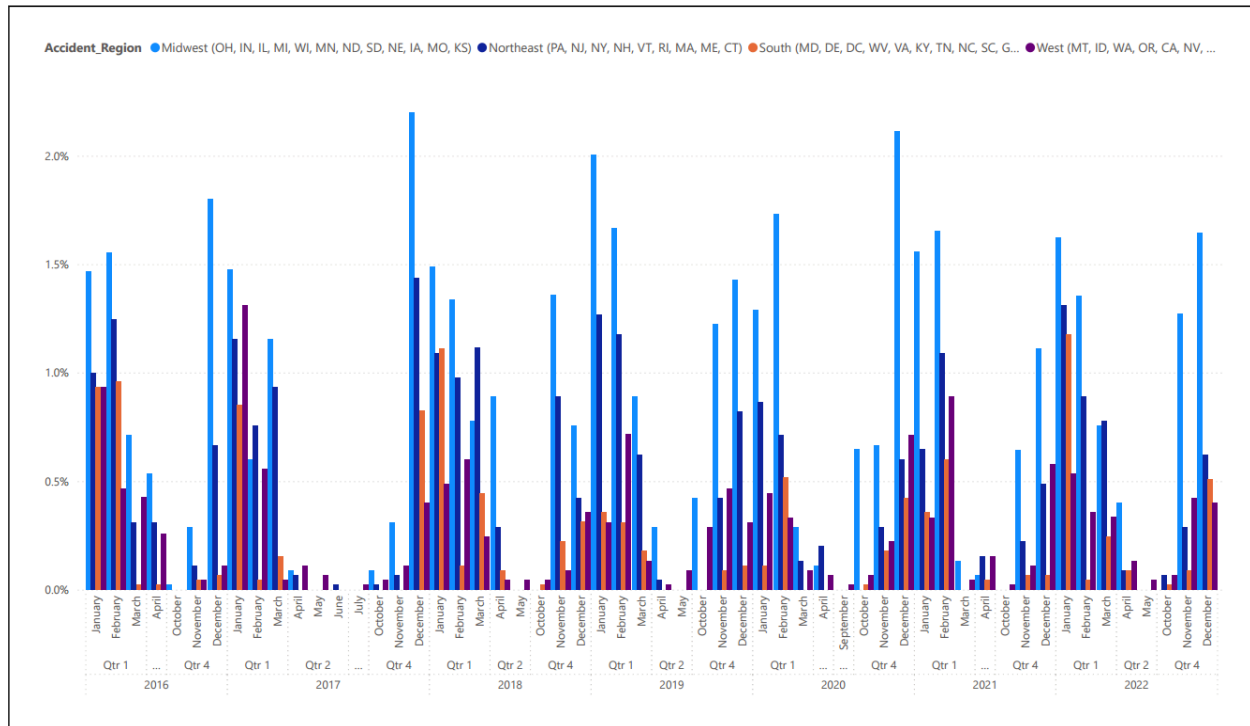


Image Acc-06

Here the percentage of accidents during snow are broken up into regions (Northeast, South, Midwest, and West). We can see that even in these regions the effect of snow was very little, reaching just over 2% a couple of times when dates are broken down into quarters. If weather doesn't have much of an influence on accidents, maybe we'll find a trend in the days of the week.

Accident Count by Day(2016-2019):

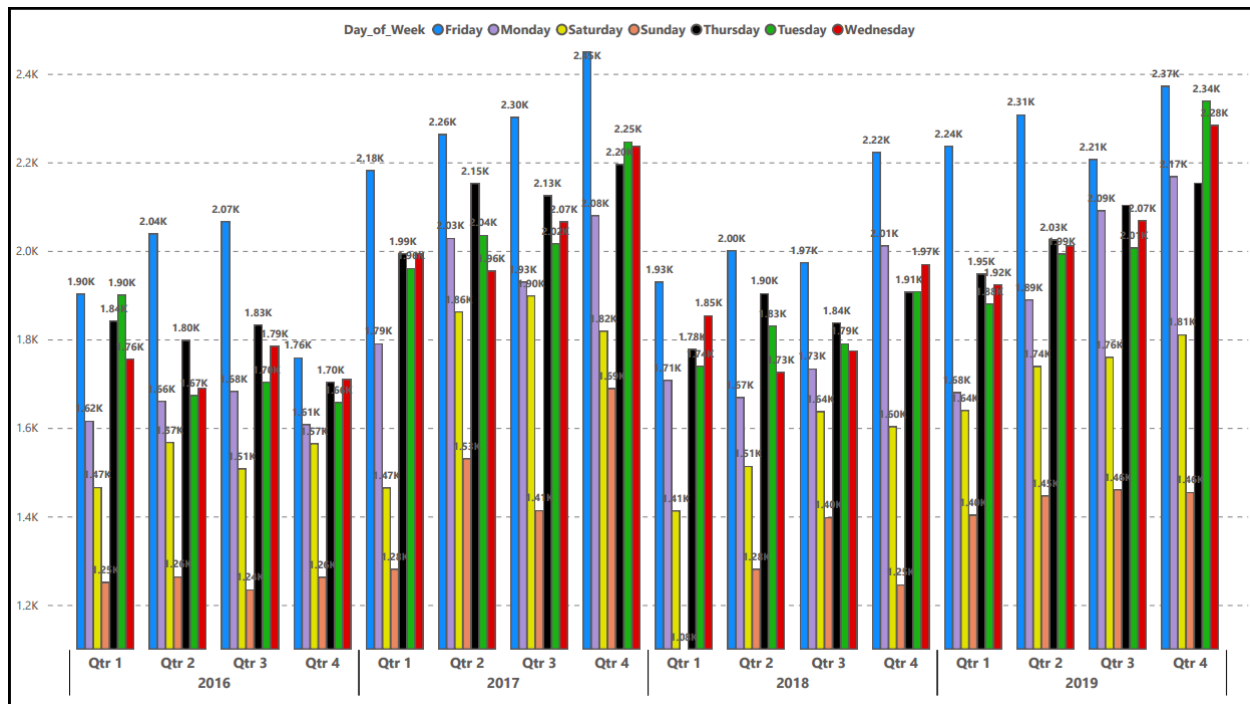


Image Acc-07

Accident Count by Day(2020-2022):

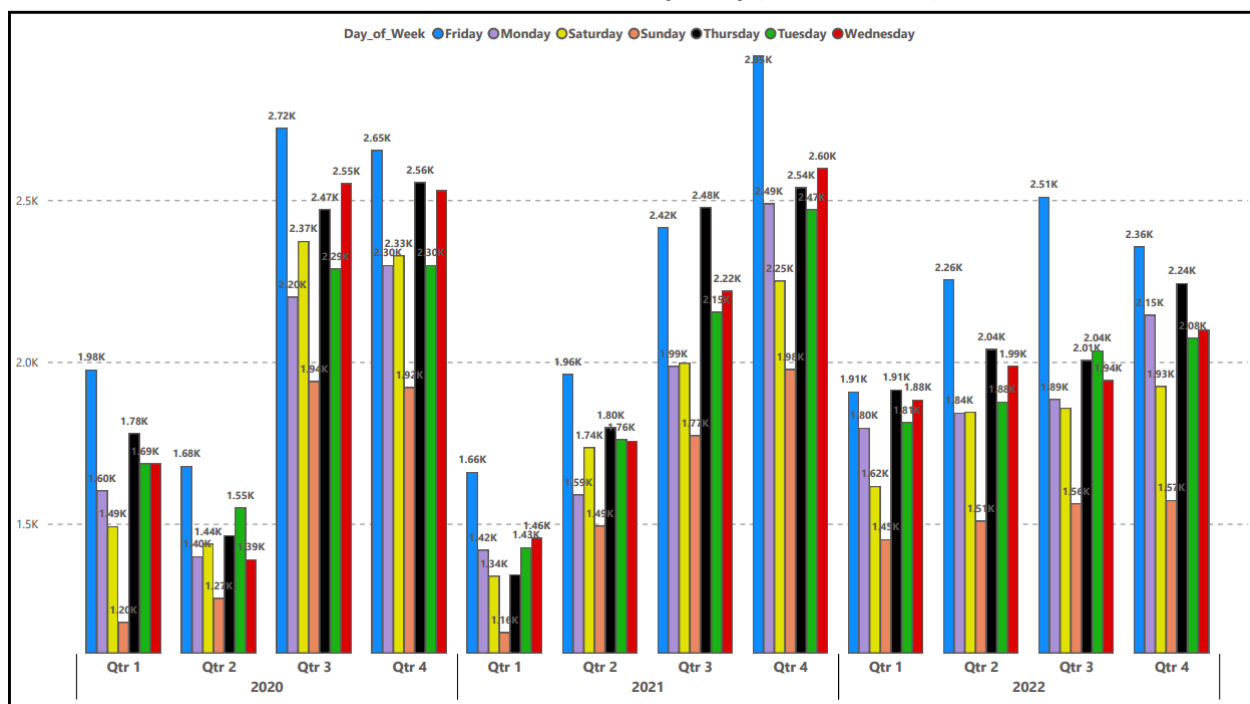


Image Acc-08

We can quickly see that Friday consistently sees more accidents than any other day, with the exception of quarter 3 in 2021 when Thursday saw 62 more accidents than Friday. My first thought, with Friday being the end of the work week for most, is people just rushing home in

anticipation for the weekend. But, to know for sure we would need to account for the time that accidents occur. So let's do that.

% of Accidents During Rush Hour:

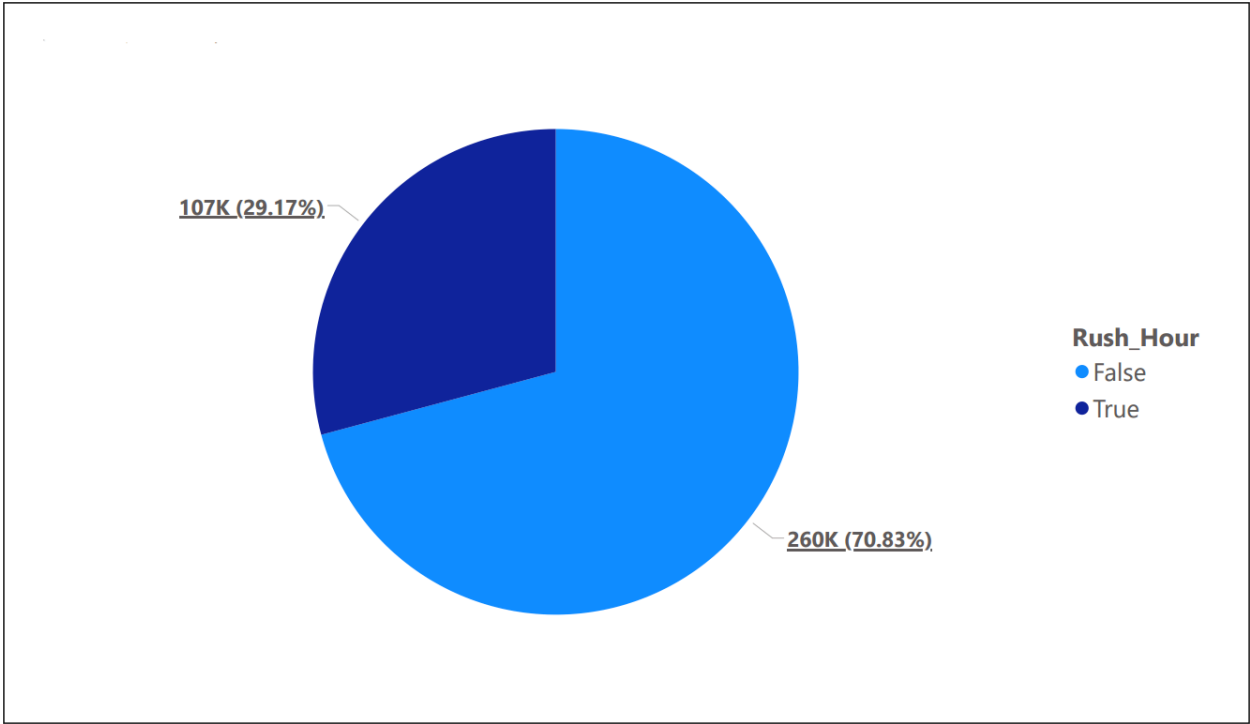


Image Acc-09

Accidents During Rush Hour by Day:

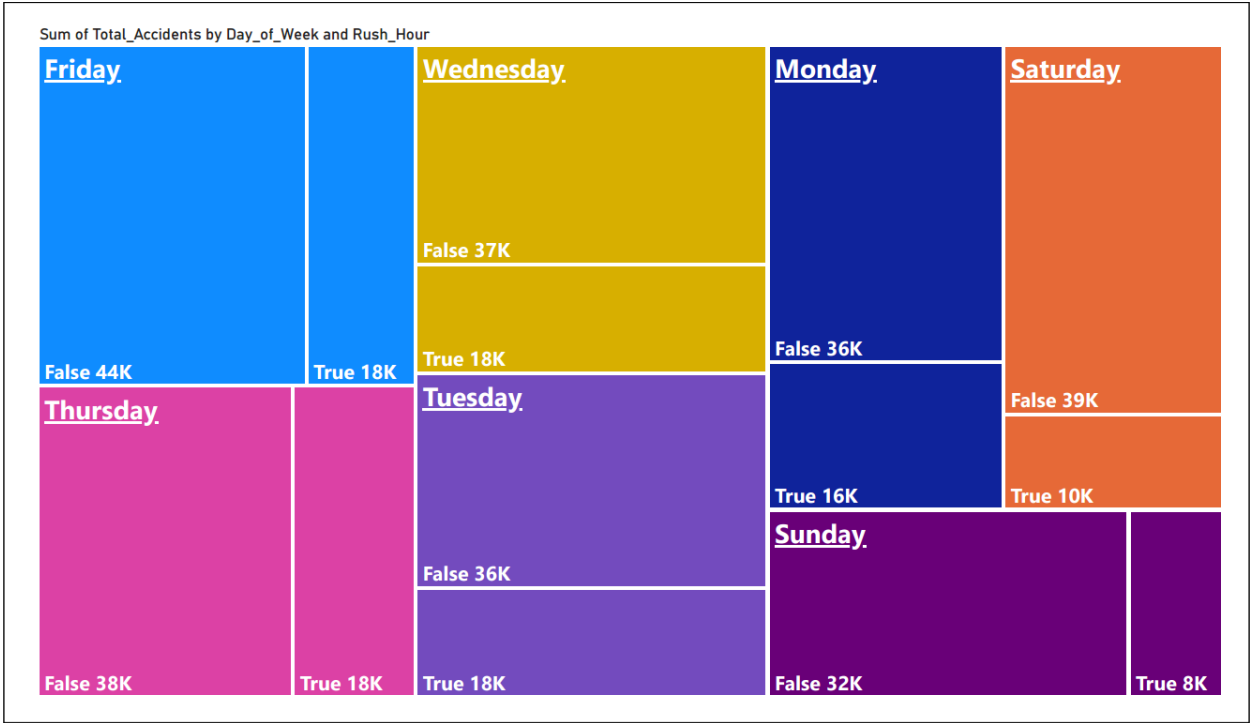


Image Acc-10

Accidents During Rush Hour by Hour:

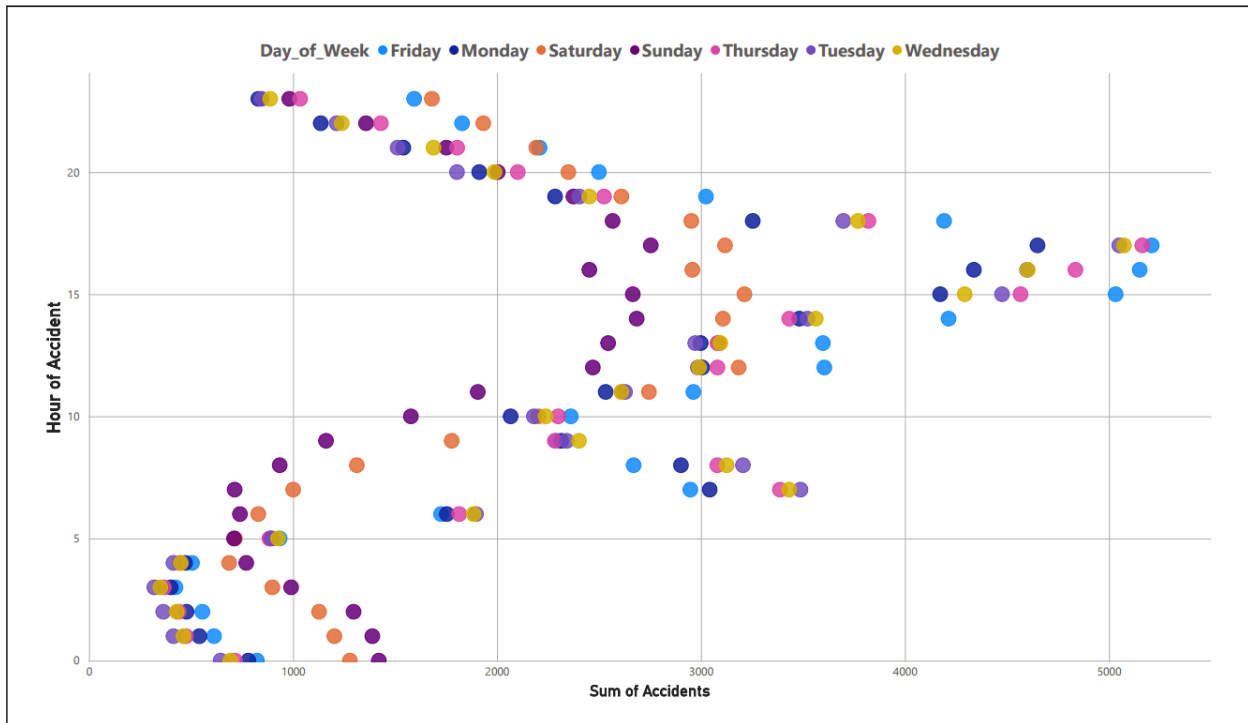


Image Acc-11

Count of Vehicles Involved:

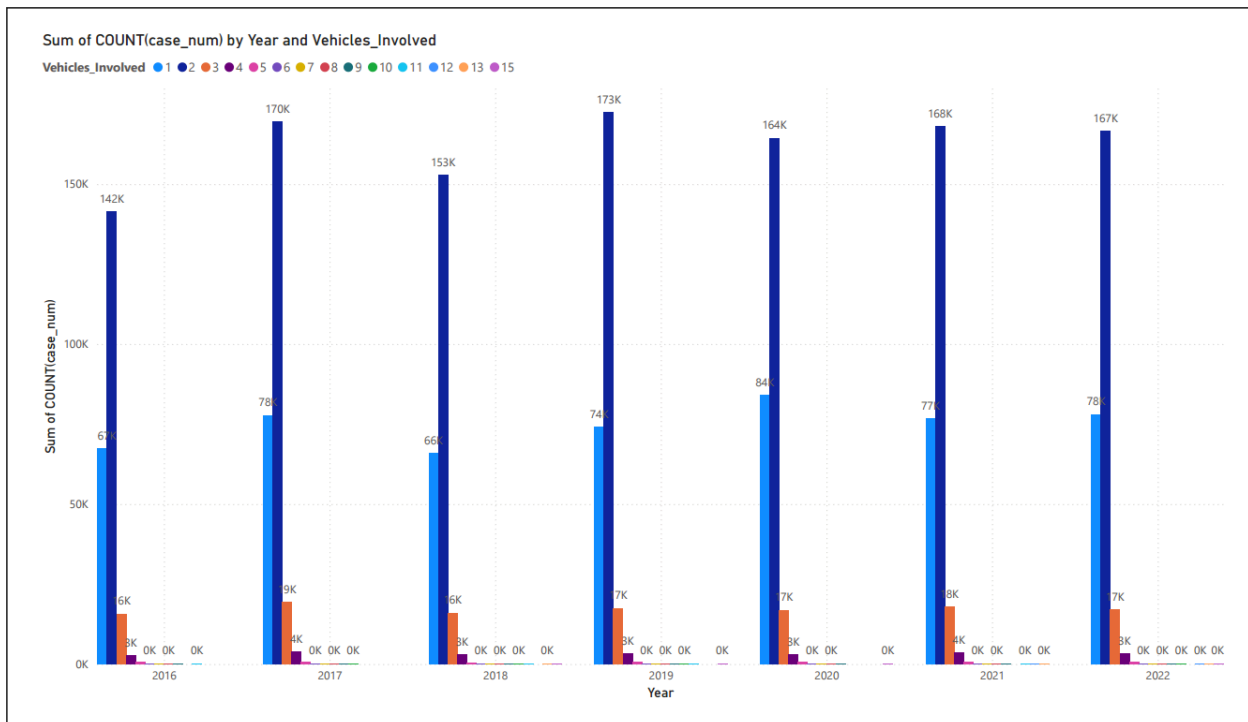


Image Acc-12

Road Relation & Junctions:

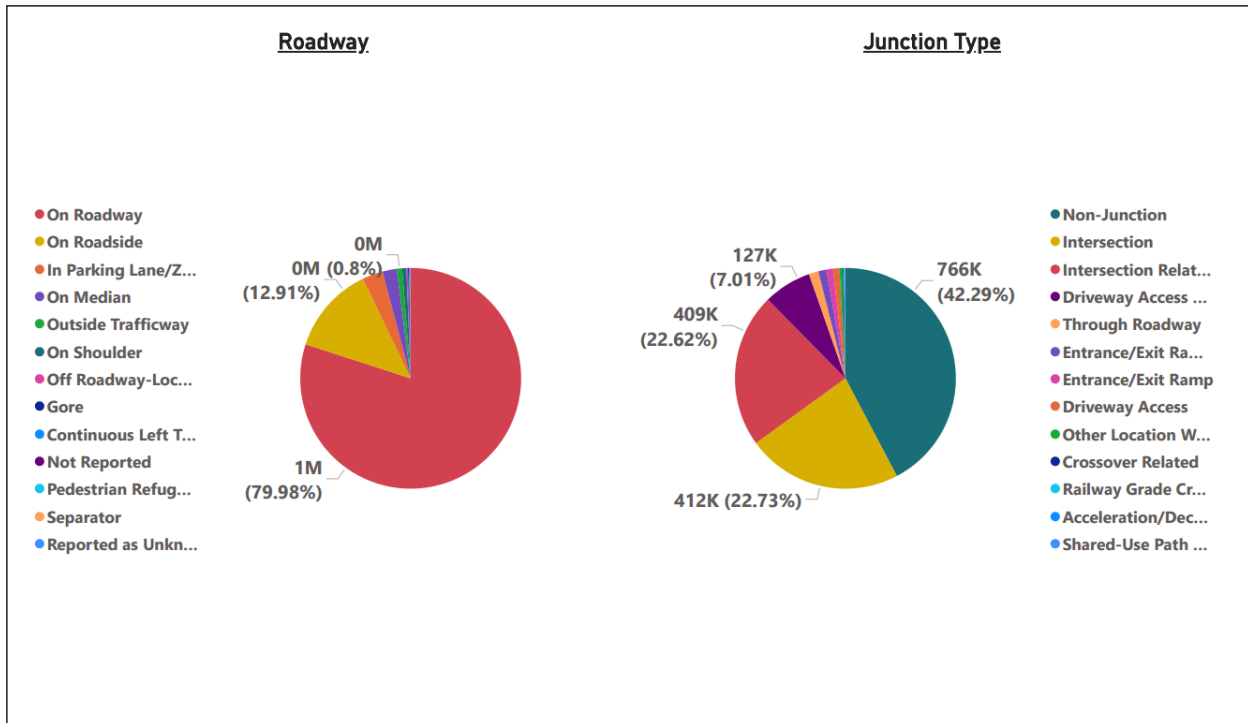


Image Acc-13

People Injured per Accident % Most Severe Injuries:

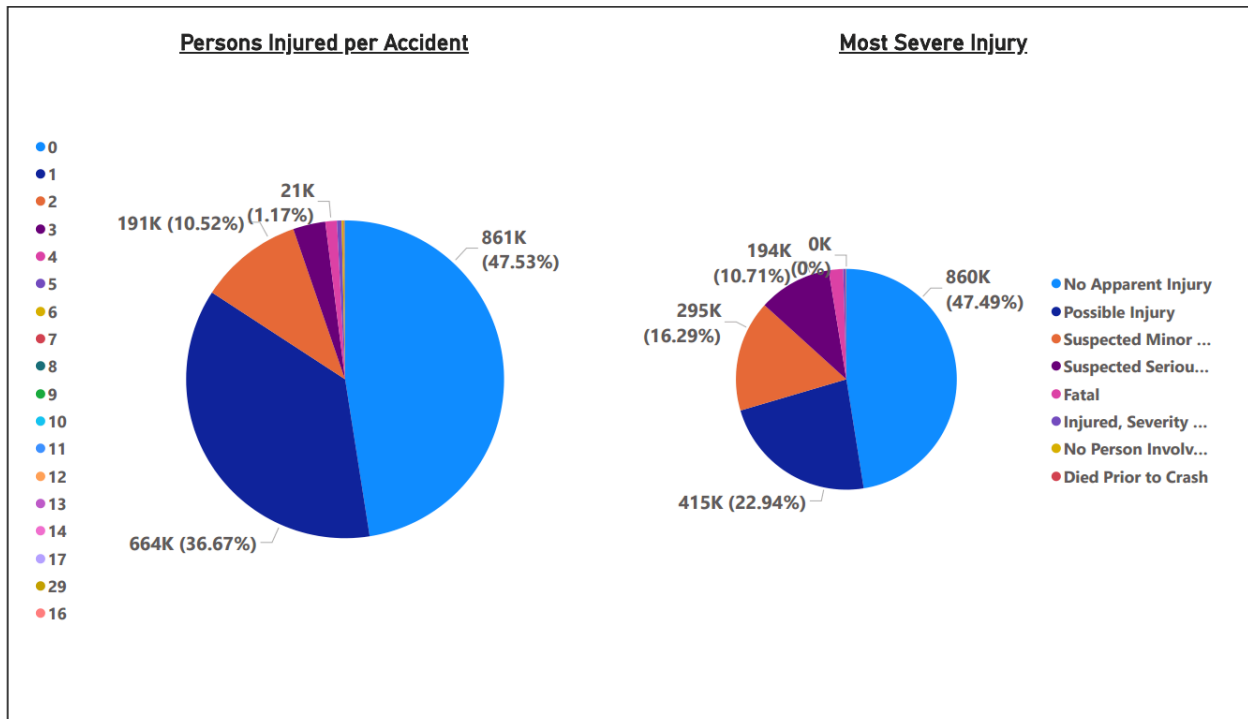


Image Acc-14

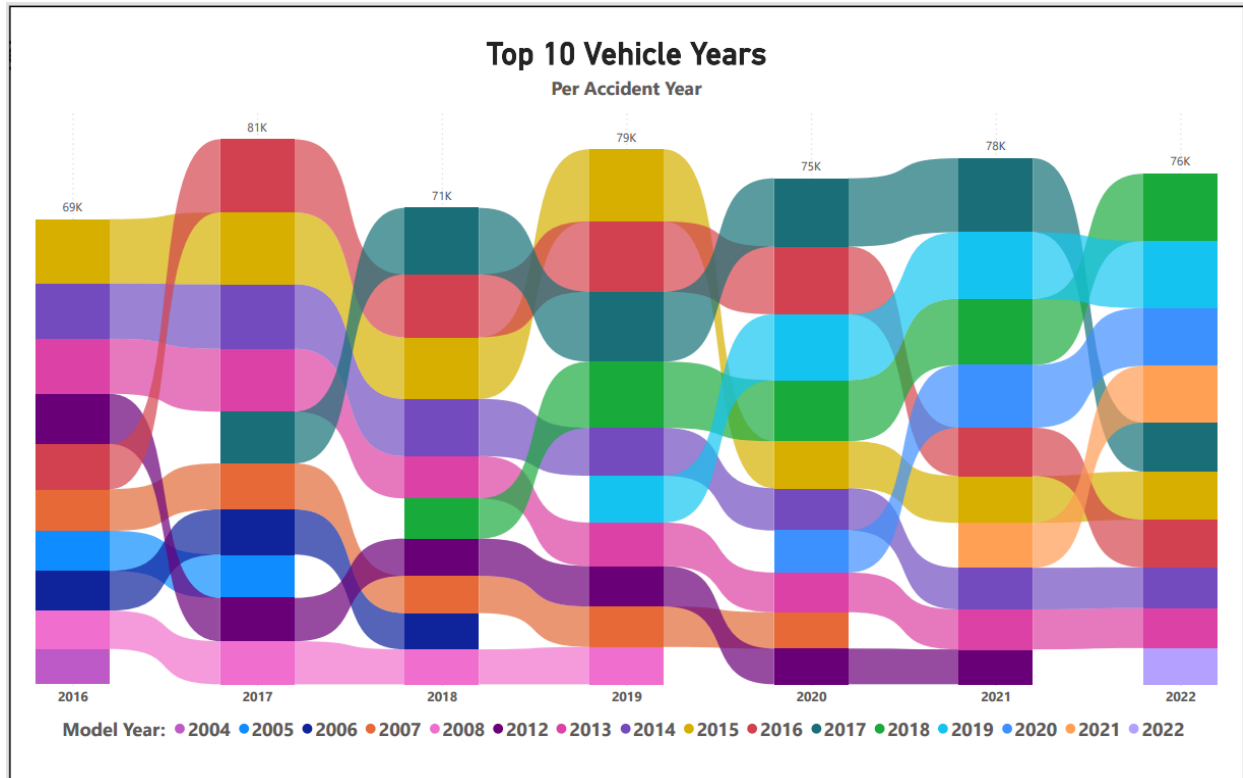


Image Veh-01

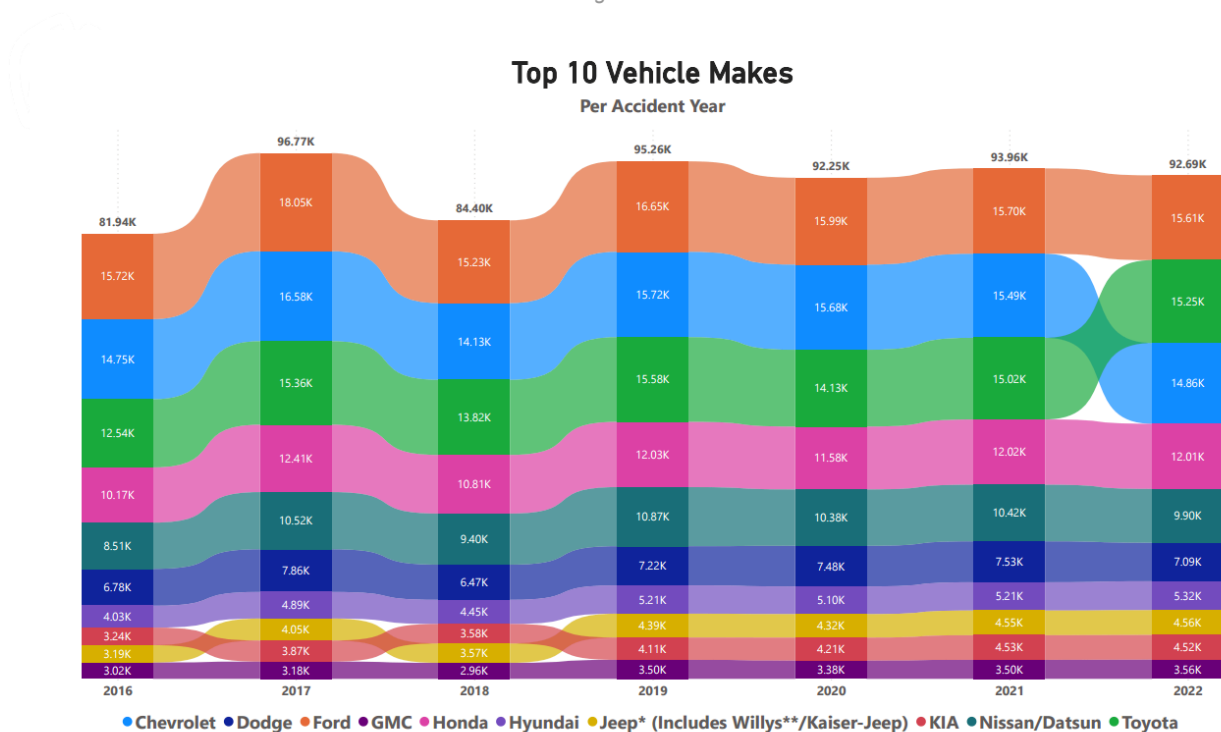


Image Veh-02

10

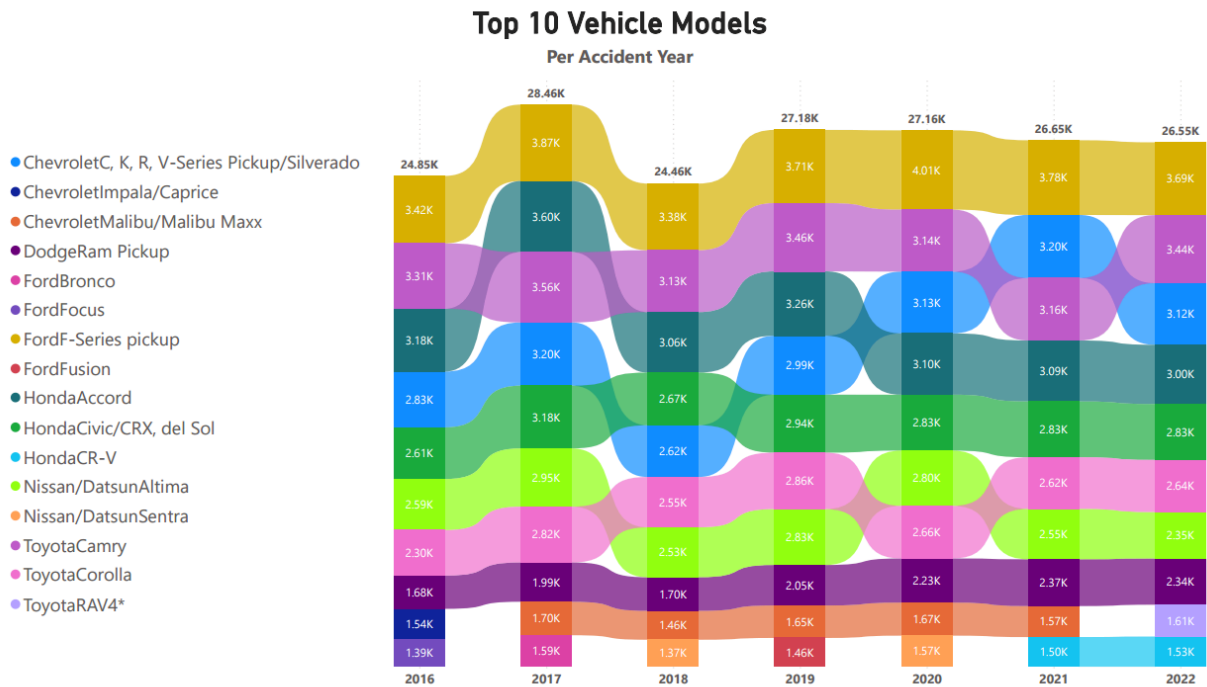


Image Veh-03

I modified the query a bit to filter out Vehicle Make ID 99, indicating the vehicle was reported as "Unknown".



Top 10 Year Make Model

Per Accident Year

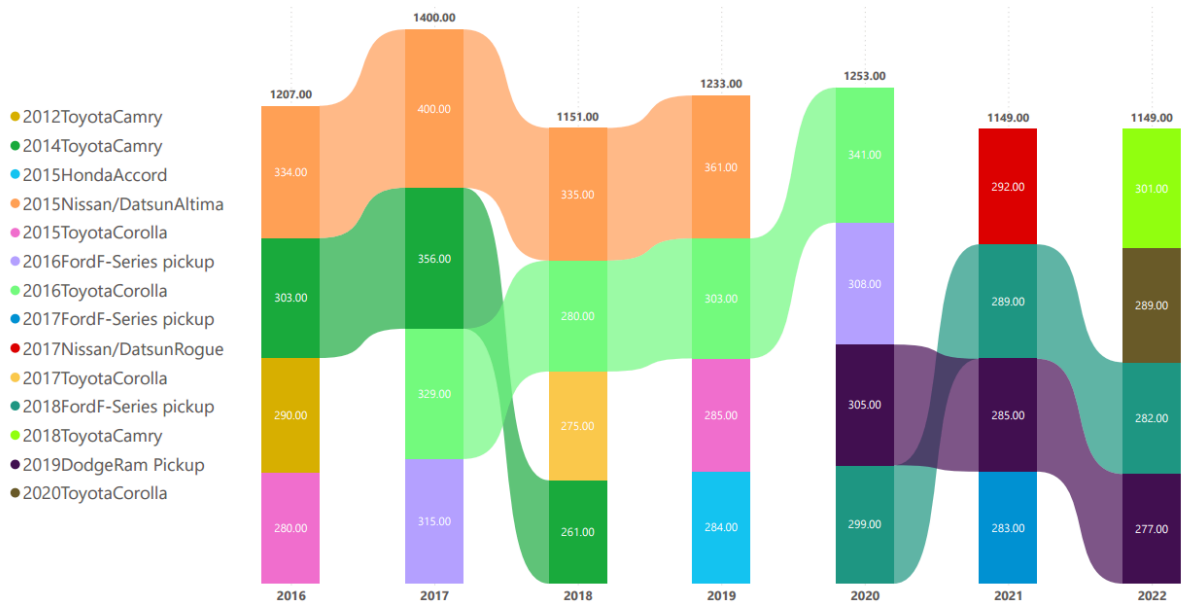


Image Veh-04

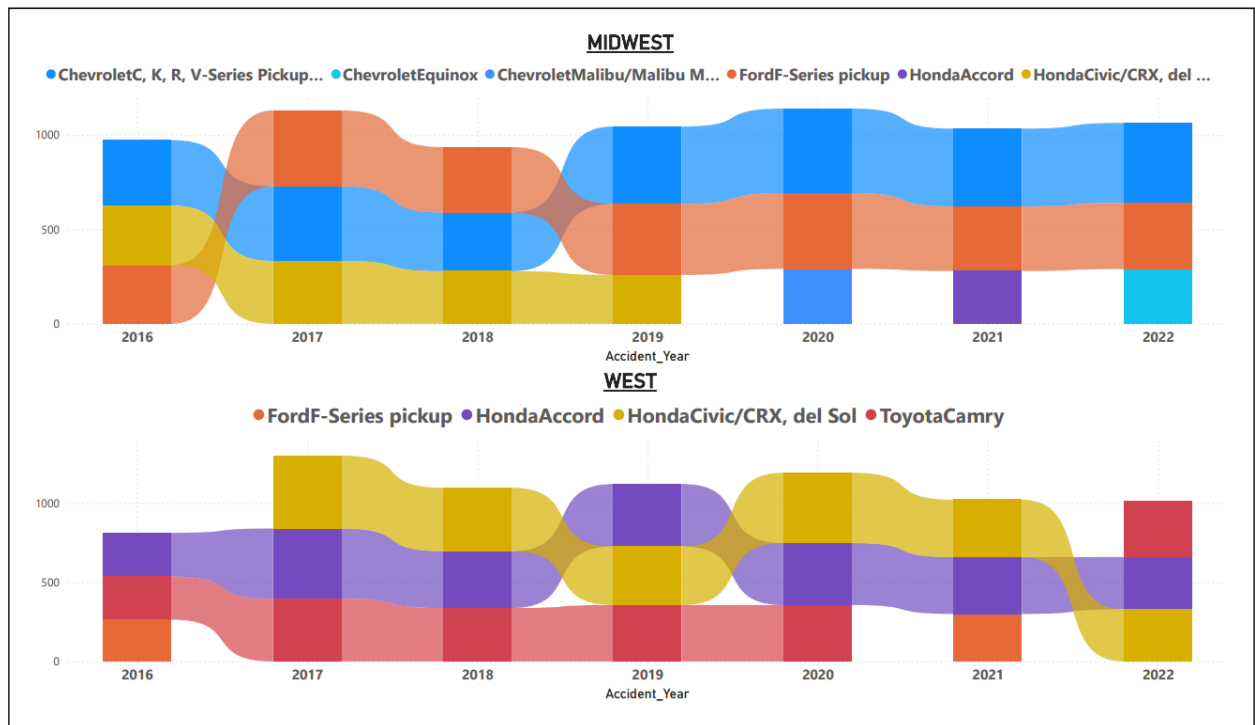


Image Veh-05

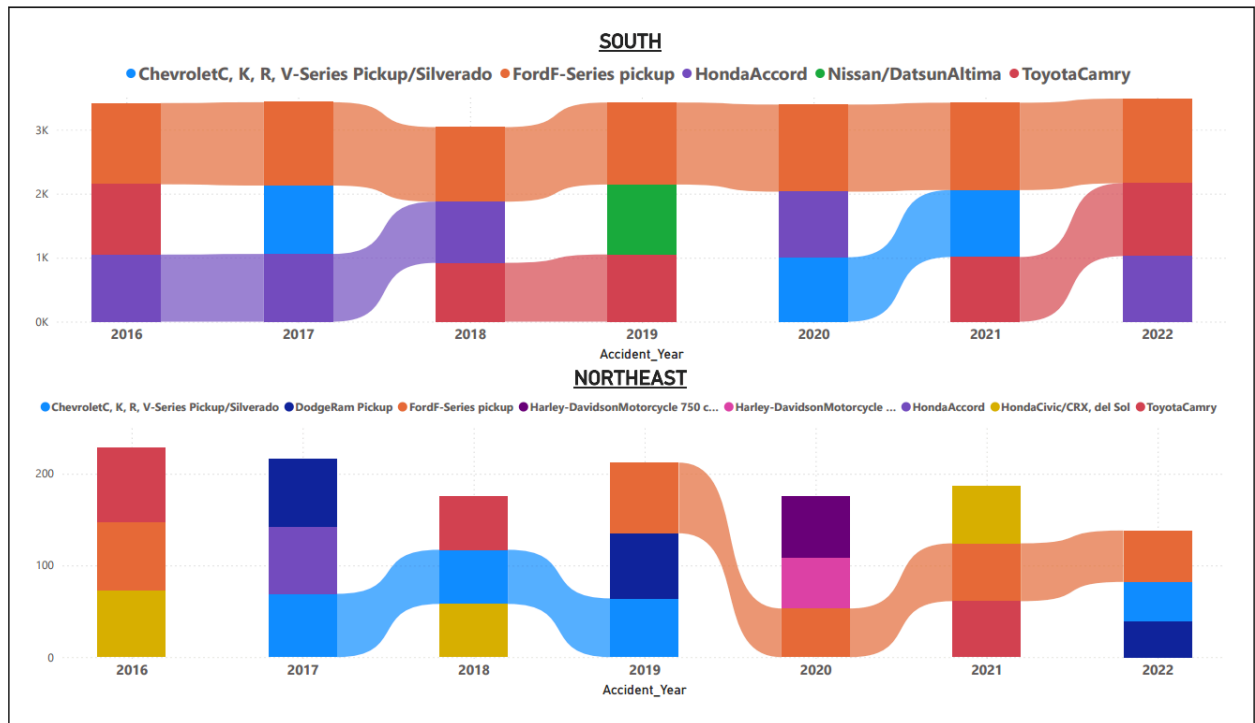


Image Veh-06

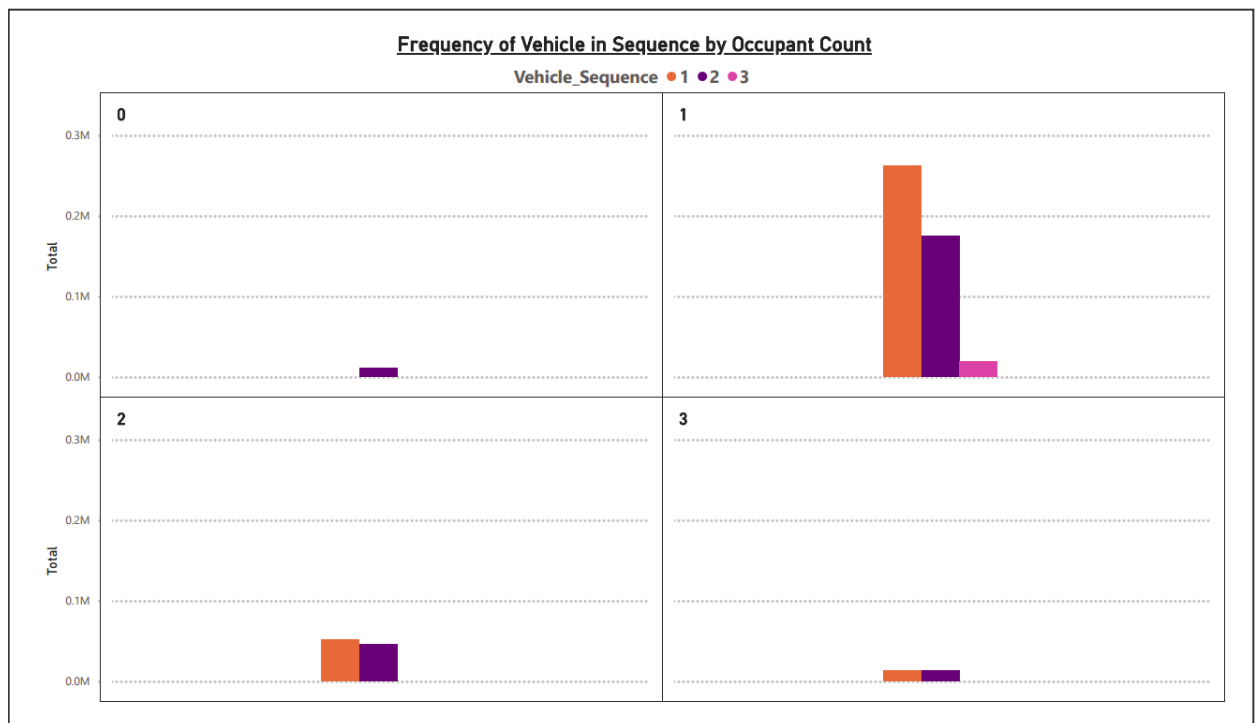


Image Veh-07

Top left- occupant count

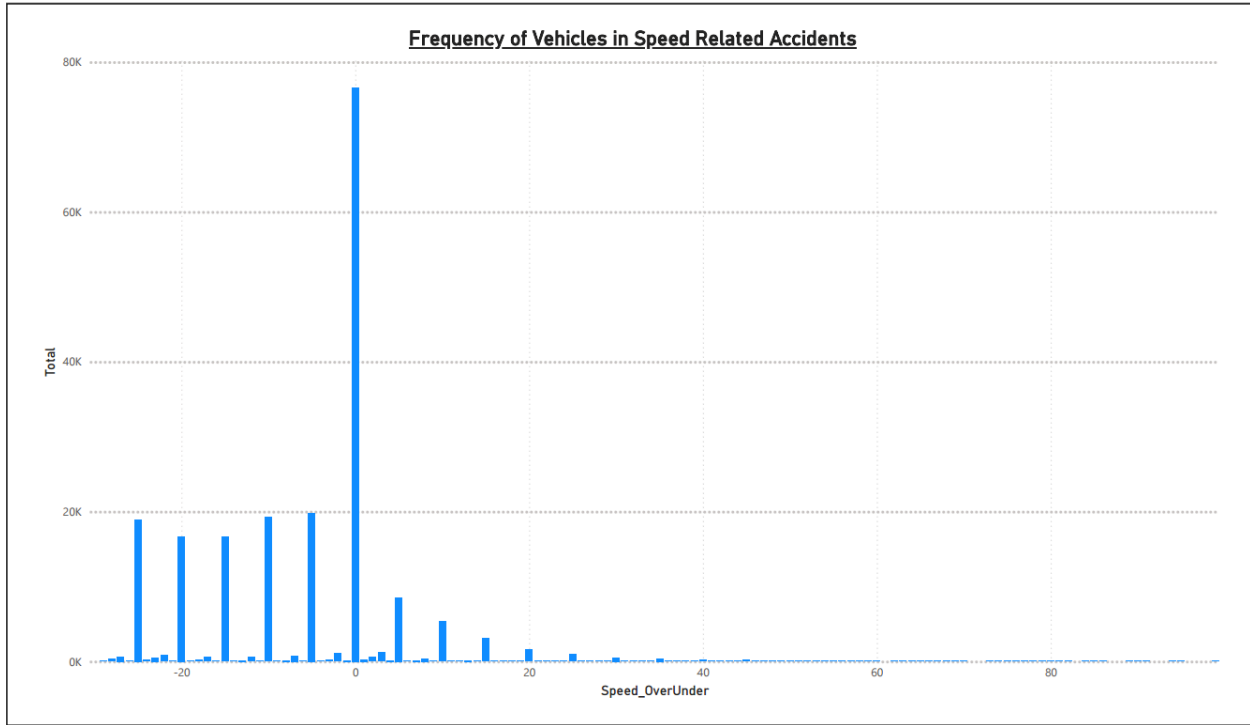


Image Veh-08

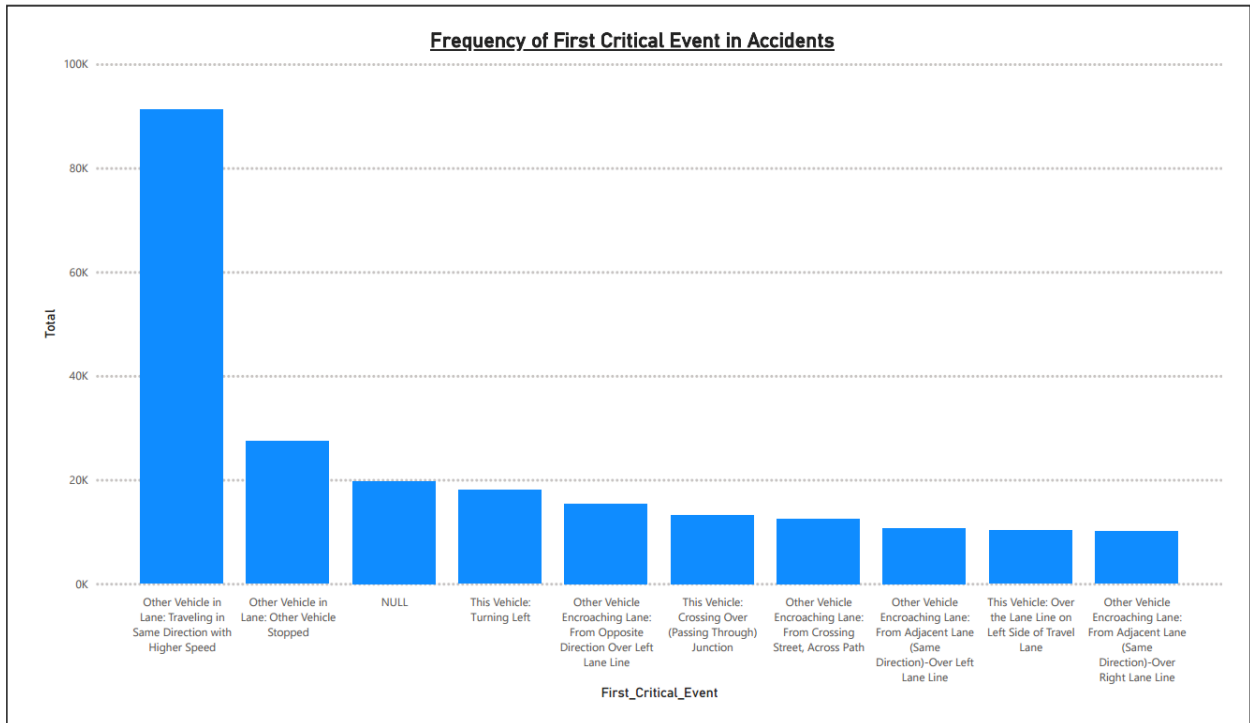


Image Veh-09

SQL Queries for Visualizations:

Image Acc-01, Acc-02, and Acc-03:

```
SELECT
    ddat.date_year AS year_record,
    ddat.month_num AS month_record,
    COUNT(acc.case_num) AS total_accidents
FROM fact_accident AS acc
JOIN dimacc_dates AS ddat
    ON acc.acc_date = ddat.date_id
GROUP BY month_record, year_record
;
```

Image Acc-04, Acc-05, and Acc-06:

```
SELECT
    ddat.acc_date AS Record_Date,
    days.day_name AS Day_of_Week,
    dweath.weather_cond AS Weather,
    acc.region AS Region,
    acc.urbancity AS Zone,
    COUNT(acc.case_num) AS Total_Accidents
FROM fact_accident AS acc
JOIN dimacc_dates AS ddat
    ON acc.acc_date = ddat.date_id
JOIN dimacc_weather AS dweath
    ON acc.weather = dweath.weather_cond
JOIN dimacc_dates AS ddat
    ON acc.acc_date = ddat.date_id
JOIN dimacc_dayofweek AS days
    ON acc.day_week = days.day_id
GROUP BY Record_Date , Day_of_Week , Weather, Region, Zone
;
```

Image Acc-07, Acc-08:

```
WITH month_most_accidents AS(  
SELECT  
    *  
    ,row_number() OVER(PARTITION BY year_record ORDER BY total_accidents  
DESC) AS rownum  
FROM (  
    SELECT  
        ddates.date_year AS year_record,  
        ddates.month_number AS month_record,  
        COUNT(acc.case_num) AS total_accidents  
    FROM fact_accident AS acc  
    JOIN dim_dates AS ddates  
        ON acc.acc_date = ddates.date_id  
    GROUP BY month_record, year_record  
    ) AS a)  
SELECT  
    year_record  
    ,month_record  
    ,total_accidents  
FROM month_most_accidents  
WHERE rownum = 1  
;
```

AND

```
SELECT  
    ddat.acc_date AS Record_Date,  
    days.day_name AS Day_of_Week,  
    dtime.hours AS Time_of_Accident,  
    dtime.timeframe AS Timeframe,  
    dtime.is_peak_hour AS Rush_Hour,  
    COUNT(acc.case_num) AS Total_Accidents  
FROM fact_accident AS acc  
JOIN dimacc_dates AS ddat  
    ON acc.acc_date = ddat.date_id  
JOIN dimacc_dayofweek AS days  
    ON acc.day_week = days.day_id  
JOIN dimacc_time AS dtime
```



```
        ON acc.acc_time = dtime.time_id
GROUP BY Record_Date , Day_of_Week , Time_of_Accident , dtime.timeframe ,
Rush_Hour
```

Acc-12, Acc-13, and Acc-14

```
DROP VIEW accident_traits;
CREATE VIEW accident_traits AS
SELECT
    dates.acc_date AS Accident_Date,
    regi.region_name AS Accident_Region,
    weat.weather_cond AS Weather,
    urty.uc_name AS Urban_or_City,
    acc.vehicles AS Vehicles_Involved,
    collision.mancoll_status AS Collision,
    road.road_status AS Roadway,
    junct.junct_status AS Junction,
    inter.intersect_status AS Intersection,
    hwy.hwy_status AS Highway,
    wzzone.workzone_status AS Work_Zone,
    acc.injuries AS Injuries,
    injsev.injury_status AS Most_Severe_Injury,
    COUNT(case_num)
FROM fact_accident AS acc
LEFT JOIN dimacc_dates AS dates
    ON acc.acc_date = dates.date_id
LEFT JOIN dimacc_weather AS weat
    ON acc.weather = weat.weather_id
LEFT JOIN dimacc_region AS regi
    ON acc.region = regi.region_id
LEFT JOIN dimacc_urbancity AS urty
    ON acc.urbancity = urty.uc_id
LEFT JOIN dimacc_manor_collision AS collision
    ON acc.collision = collision.mancoll_id
LEFT JOIN dimacc_roadway_relation AS road
    ON acc.roadway = road.road_id
LEFT JOIN dimacc_relatedjunction AS junct
    ON acc.related_junct = junct.junct_id
LEFT JOIN dimacc_intersection AS inter
    ON acc.intersection = inter.intersect_id
```

```

LEFT JOIN dimacc_interstate_hwy AS hwy
    ON acc.highway = hwy.hwy_id
LEFT JOIN dimacc_workzone AS wzone
    ON acc.workzone = wzone.workzone_status
LEFT JOIN dimacc_injury_severity AS injsev
    ON acc.max_injury = injsev.injury_id
GROUP BY
    Accident_Date,
    Weather,
    Accident_Region,
    Urban_or_City,
    Vehicles_Involved,
    Collision,
    Roadway,
    Junction,
    Intersection,
    Highway,
    Work_Zone,
    Injuries,
    Most_Severe_Injury
;

```

Veh-01:

```

WITH vyear_by_year AS (
SELECT
    *,
    ROW_NUMBER() OVER(PARTITION BY Accident_Year ORDER BY Total DESC) AS
rownum
FROM (
    SELECT
        dts.date_year AS Accident_Year,
        fv.veh_year AS Vehicle_Year,
        COUNT(fv.veh_year) AS Total
    FROM fact_accident AS fa
    JOIN dimacc_dates AS dts
        ON fa.acc_date = dts.date_id
    JOIN fact_person AS fp

```

```

        ON fa.case_num = fp.case_num
    JOIN fact_vehicle AS fv
        ON fv.veh_id = fp.veh_id
    GROUP BY Accident_Year, Vehicle_Year
    ) AS a)
SELECT *
FROM vyear_by_year
WHERE rownum BETWEEN 1 AND 10
;

```

Image Veh-02:

```

WITH vmake_by_year AS (
SELECT
    *,
    ROW_NUMBER() OVER(PARTITION BY Accident_Year ORDER BY Total DESC) AS
rownum
FROM (
    SELECT
        COUNT(fv.veh_id) AS Total,
        dts.date_year AS Accident_Year,
        dmod.make_id AS Make_ID,
        dmake.make_name AS Vehicle_Make
    FROM fact_person AS fp
    JOIN fact_vehicle AS fv
        ON fv.veh_id = fp.veh_id
    JOIN fact_accident AS fa
        ON fa.case_num = fp.case_num
    JOIN dimacc_dates AS dts
        ON fa.acc_date = dts.date_id
    JOIN dimveh_model AS dmod
        ON dmod.model_id = fv.make_model
    JOIN dimveh_make as dmake
        ON dmod.make_id = dmake.make_id
    GROUP BY Accident_Year, Make_ID, Vehicle_Make
    ) AS a)
SELECT Accident_Year, Make_ID, Vehicle_Make, Total, rownum
FROM vmake_by_year
WHERE rownum BETWEEN 1 AND 10

```

```
;
```

Image Veh-03:

```
WITH vmodel_by_year AS(
SELECT
    *,
    ROW_NUMBER() OVER(PARTITION BY Accident_Year ORDER BY Total DESC) AS
rownum
FROM (
    SELECT
        COUNT(fv.veh_id) AS Total,
        dts.date_year AS Accident_Year,
        dmod.model_id AS Model_ID,
        dmake.make_name AS Vehicle_Make,
        dmod.model_name AS Vehicle_Model
    FROM fact_person AS fp
    JOIN fact_vehicle AS fv
        ON fv.veh_id = fp.veh_id
    JOIN fact_accident AS fa
        ON fa.case_num = fp.case_num
    JOIN dimacc_dates AS dts
        ON fa.acc_date = dts.date_id
    JOIN dimveh_model AS dmod
        ON dmod.model_id = fv.make_model
    JOIN dimveh_make as dmake
        ON dmod.make_id = dmake.make_id
    GROUP BY Accident_Year, Model_ID, Model_Year, Vehicle_Make
    ) AS a)
SELECT Accident_Year, Model_ID, Vehicle_Make, Vehicle_Model, Total, rownum
FROM vmodel_by_year
WHERE rownum BETWEEN 1 AND 10
;
```

Image Veh-04:

```
WITH yearmakmod_by_year AS(
SELECT
    *,
```

```

        ROW_NUMBER() OVER(PARTITION BY Accident_Year ORDER BY Total DESC) AS
rownum
FROM (
    SELECT
        COUNT(fv.veh_id) AS Total,
        dts.date_year AS Accident_Year,
        dmod.model_id AS Model_ID,
        fv.veh_year AS Model_Year,
        dmake.make_name AS Vehicle_Make,
        dmod.model_name AS Vehicle_Model
    FROM fact_person AS fp
    JOIN fact_vehicle AS fv
        ON fv.veh_id = fp.veh_id
    JOIN fact_accident AS fa
        ON fa.case_num = fp.case_num
    JOIN dimacc_dates AS dts
        ON fa.acc_date = dts.date_id
    JOIN dimveh_model AS dmod
        ON dmod.model_id = fv.make_model
    JOIN dimveh_make as dmake
        ON dmod.make_id = dmake.make_id
    WHERE Model_ID NOT LIKE '99%'
    GROUP BY Accident_Year, Model_ID, Model_Year, Vehicle_Make
    ) AS a)
SELECT Accident_Year, Model_ID, Model_Year, Vehicle_Make, Vehicle_Model,
Total, rownum
FROM vmake_by_year
WHERE rownum BETWEEN 1 AND 10
;

```

Image veh-05 and veh-06:

```

WITH vmodel_by_year AS(
SELECT
    *,
    ROW_NUMBER() OVER(PARTITION BY Accident_Year ORDER BY Total DESC) AS
rownum
FROM (
    SELECT

```

```

COUNT(fv.veh_id) AS Total,
dts.date_year AS Accident_Year,
fv.veh_year AS Model_Year,
dmake.make_name AS Vehicle_Make,
dmod.model_name AS Vehicle_Model,
dreg.region_name AS US_Region
FROM fact_person AS fp
JOIN fact_vehicle AS fv
    ON fv.veh_id = fp.veh_id
JOIN fact_accident AS fa
    ON fa.case_num = fp.case_num
JOIN dimacc_dates AS dts
    ON fa.acc_date = dts.date_id
JOIN dimveh_model AS dmod
    ON dmod.model_id = fv.make_model
JOIN dimveh_make AS dmake
    ON dmod.make_id = dmake.make_id
JOIN dimacc_region AS dreg
    ON fa.region = dreg.region_name
WHERE Model_ID NOT LIKE '99%'
GROUP BY Accident_Year, Model_Year, Model_Year, Vehicle_Make,
US_Region
) AS a)
SELECT Accident_Year, Model_ID, Vehicle_Make, Vehicle_Model, US_Region,
Total, rownum
FROM vmodel_by_year
WHERE rownum BETWEEN 1 AND 5
;

```

Image veh-05 and veh-06:

```

SELECT
COUNT(fv.veh_id) AS Total,
ROUND((COUNT(fv.veh_id) / (SELECT COUNT(veh_id) FROM fact_vehicle)) *
100, 2) AS Percentage,
veh_id % 10 AS Vehicle_Sequence,
dm.moving_name AS Moving_Status,
fv.occupants AS Occupants

```

```
FROM fact_vehicle AS fv
LEFT JOIN dimveh_moving AS dm
    ON fv.moving = dm.moving_id
GROUP BY Moving_Status, Vehicle_Sequence, occupants
ORDER BY Total DESC
LIMIT 10
;
```

Image