Seattle SDOT Collisions Data Analysis

Team 2 – Bryan Wang, Hao Tung, Jess Lu, Neelabja Gayen, Pallavi Khabale

AGENDA

01 INTRODUCTION

02 DATASET

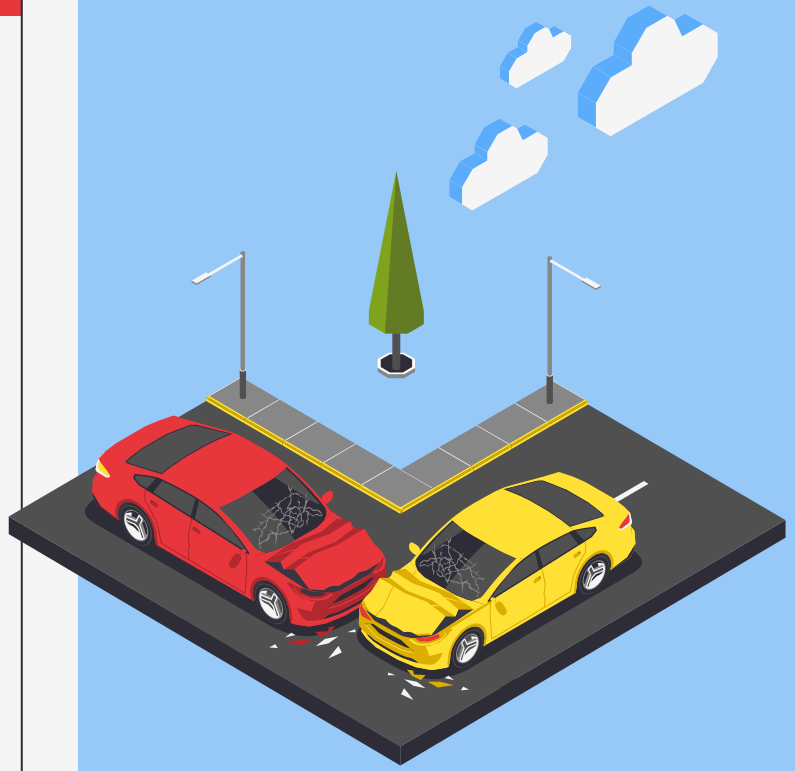
03 DATA CHALLENGES

04 QUESTIONS & INSIGHTS

05 DATA ANALYSIS

06 STRATEGIC VALUE & KEY TAKEAWAYS

INTRODUCTION



INTRODUCTION

38,000

People die every year in crashes in U.S. roadways

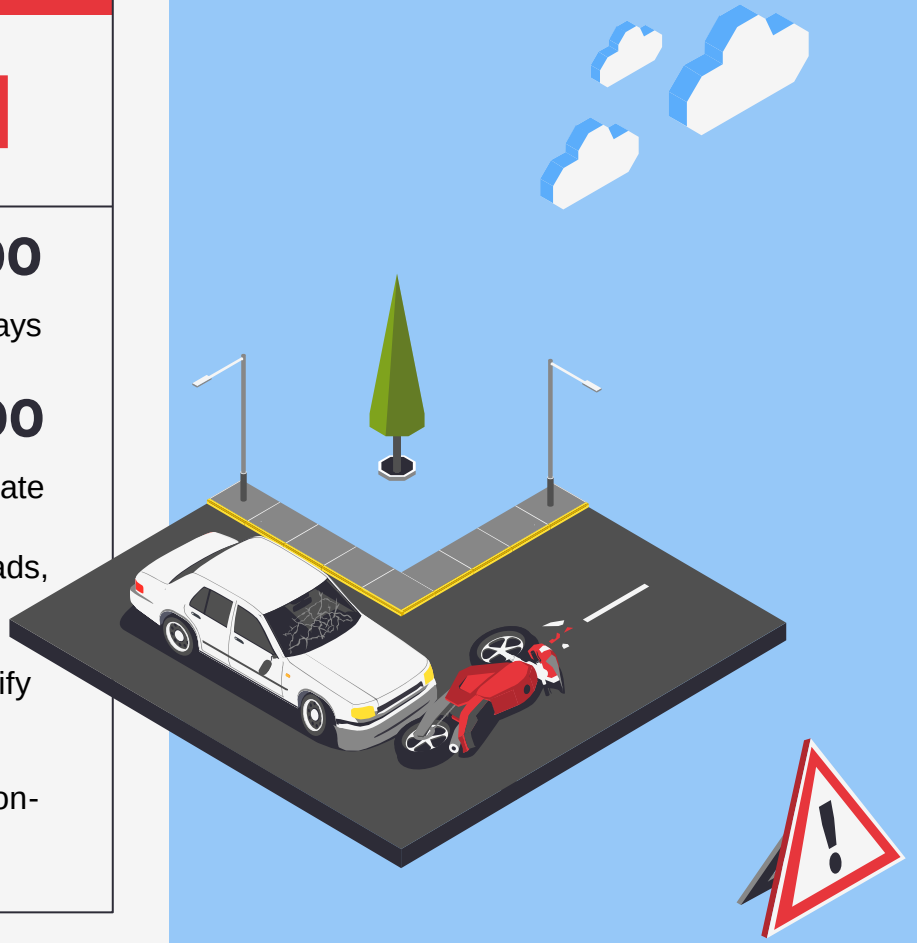
12.4 deaths per 100,000

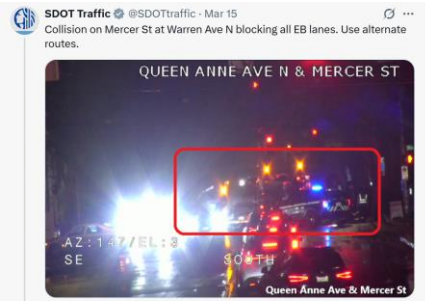
U.S. traffic fatality rate

With increasing number of vehicles on urban and suburban roads, the cases of vehicle accidents are also increasing.

Exploring the key trends, contributing factors can help us identify the potential solutions for improving road safety.

Data can be used to gain insights that can inform better decision-making for city planning and public awareness.





A Rise in Seattle Collisions !

Increasing Collision Trends:

- There are 475 recorded collisions in 2025 with 34 collisions alone from March 1st to March 15th.
- Factors include population growth, increased vehicle usage, and changing road conditions.

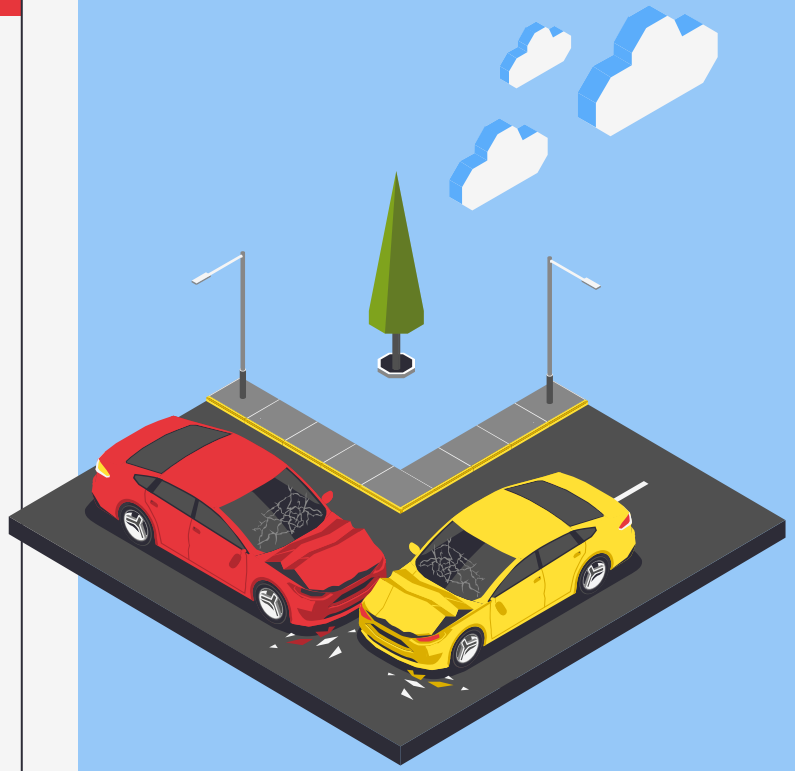
Key Contributing Factors:

- Distracted driving, DUI and speeding remain major causes.
- Weather conditions (rain, fog) impact road safety.
- Infrastructure and road design challenges contribute to accident-prone areas.

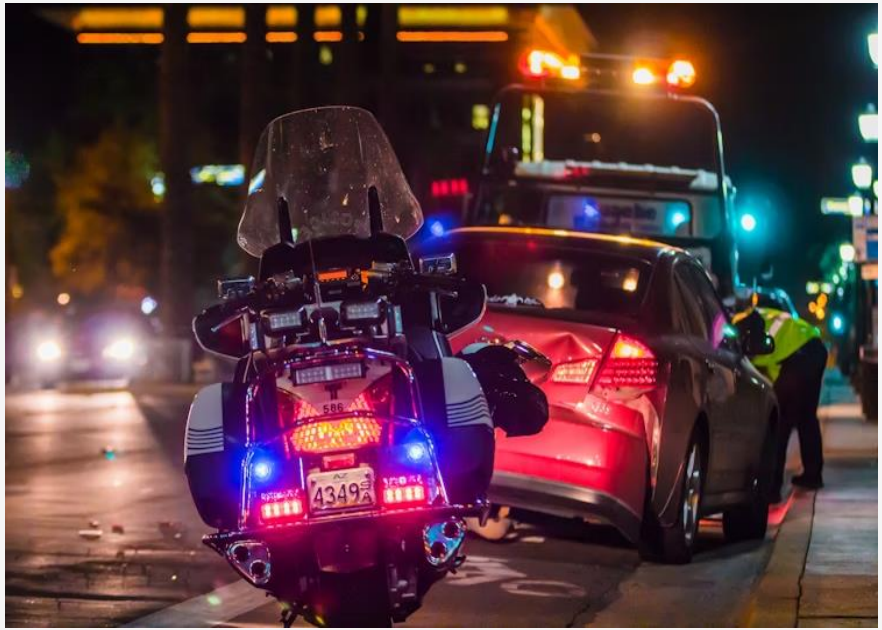
Impact on the City:

- More accidents lead to higher fatalities and injuries.
- Increased congestion and economic costs from traffic delays.
- Growing need for data-driven safety interventions and policy changes.

DATASET



DATASET DESCRIPTION – SDOT Collisions



“ SDOT Collisions All Years ”

SDOT Collisions All Years is provided by SPD and recorded by Traffic Records. This feature includes all types of collisions and will display at the intersection or mid-block of a segment.

Duration: 10/06/2004 - 03/02/2025
Location: Seattle (-122.4202,47.4935,-122.2450,47.7356)

# of COLUMNS: 50
# of RECORDS: 253,884

Source: https://data.seattle.gov/

Dataset Link: https://data.seattle.gov/dataset/SDOT-Collisions-All-Years/qdnv-25h8

DATASET DESCRIPTION – SDOT Collisions

Detailed reports of collisions data maintained by the Seattle Department of Transportation (SDOT) from 2004 - 2025.

There are 50 columns that includes date, time, location, severity, environmental factors, contributing factors, collisions information, and involved parties.

Provides latitude and longitude coordinates for mapping accident hotspots.

Captures primary contributing circumstances such as speeding, DUI, and inattentiveness.

Helps in identifying trends, informing policy decisions, and supporting Vision Zero initiatives.

Location & Time

Severity

Injuries & Fatalities

Under Influence

Speeding

Inattentiveness

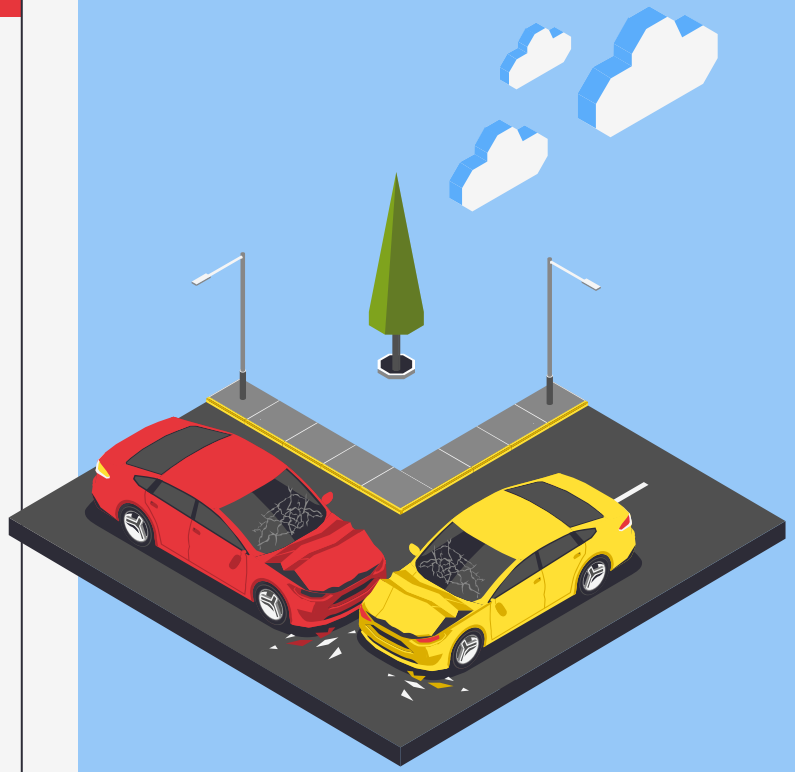
Road Conditions

Weather

Light Conditions

Dataset Link: https://data.seattle.gov/dataset/SDOT-Collisions-All-Years/qdnv-25h8

DATA CLEANING CHALLENGES



DATA CLEANING CHALLENGES

Data Cleaning

- Missing or incomplete entries (e.g., unknown causes or locations)
- Extracting Time for the incident date/time because of missing time values.
- Inconsistent formatting in different columns.

Data Processing

- Renamed the column names.
- Dropped irrelevant columns.
- Replaced missing values.
- Converted X and Y values into longitudes and latitudes.
- Replacing long descriptions with compact ones for better readability.
- Remapping few columns for consistent format.
- Removing biases caused due to outliers.

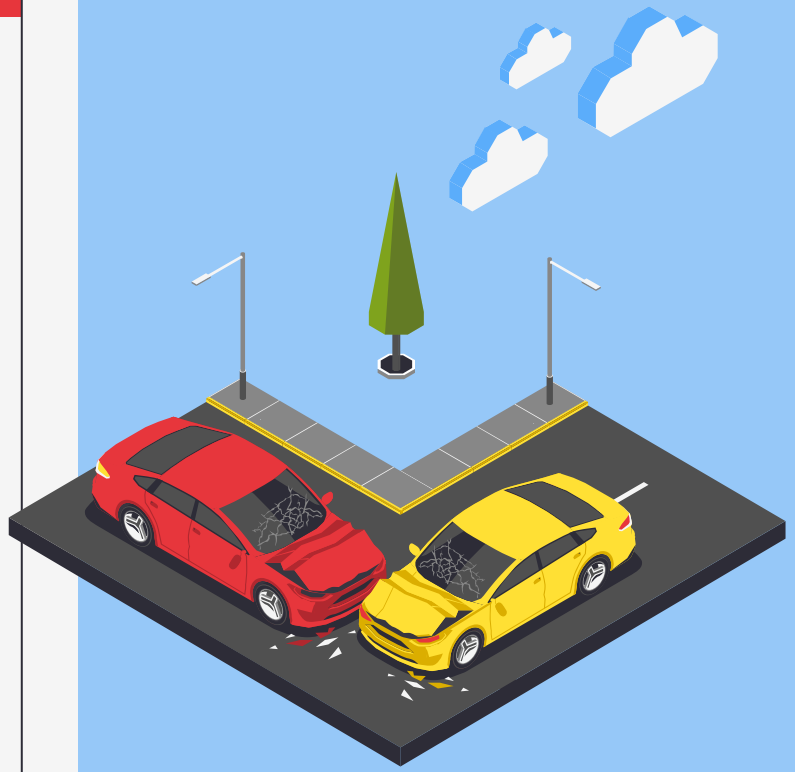
Solution

- Using Python pandas to convert incident time to datetime. If no time found (:) replace by NAN.
- Created time buckets like Late Night, Noon, etc. to group records together for easier analysis.
- Created a remapping scale on a number scale or Y/N scale as required.
- Figured out the logic to convert the geospatial values into longitudes and latitudes.

Challenges

- Not substituting the missing time values as 00:00
- Multiple time buckets available.
- Inconsistent values for categorical columns like severity code and under influence.
- Convert X & Y values into longitudes and latitudes

QUESTIONS & INSIGHTS



QUESTIONS

Time Period Analysis

Are there any Patterns and Trends in Collision Data across all Years?

Speeding/ DUI/ Inattentiveness

Did these three factors contribute to the severity of the accidents?

Collision Analysis

What are the Key Collision Patterns and High-Risk Locations Identified?

Other Conditions Impact

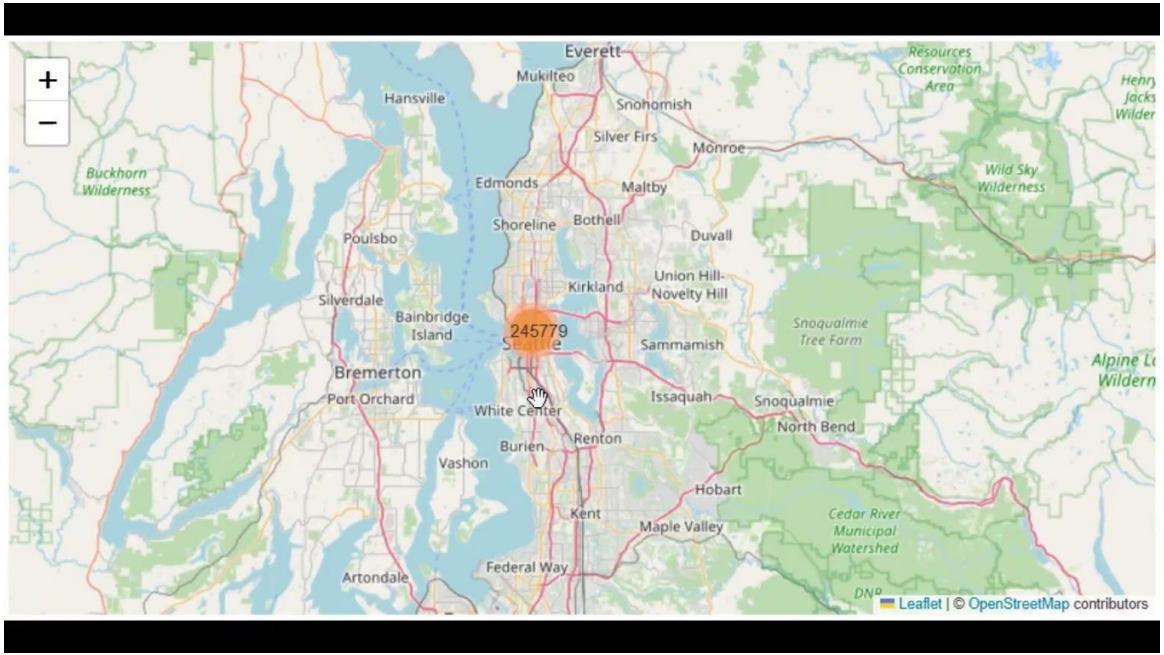
Do Weather, Road & Light Conditions affect collisions?

01

Time Period Analysis

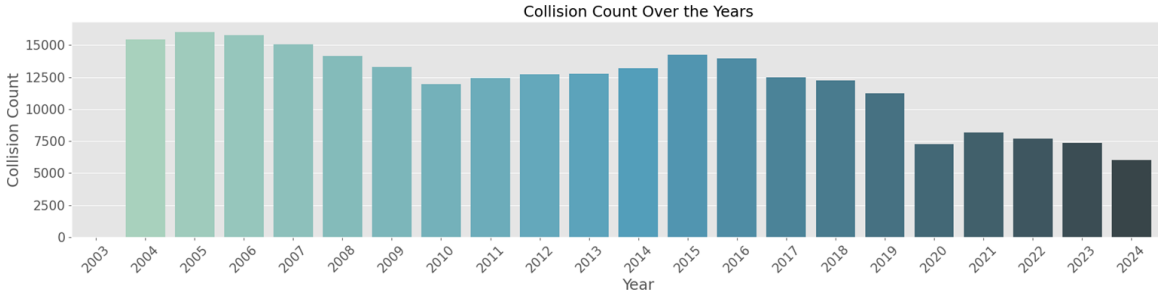
Uncovering Time-Based Patterns and Trends in Collision Data

A look at all the collisions over the years

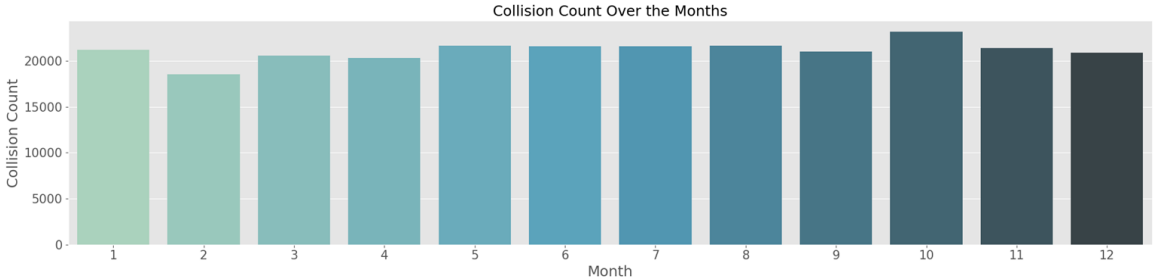


2004 - 2025

Do you know how many car accidents happen in Seattle in a year on average?



The continuous improvement and rising usage of public transportation have likely reduced the number of vehicles on the road, decreasing collision risks.

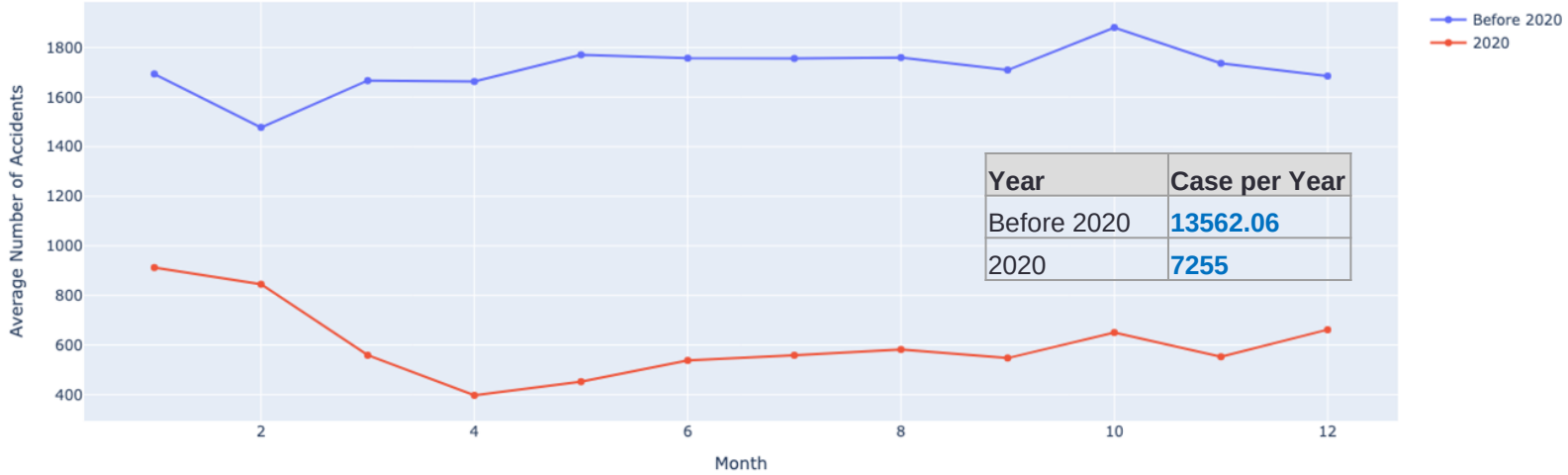


Collision rates are consistently higher from **October to December**, likely influenced by **adverse weather** and **holiday travel surges**.

Yearly Trend:
Collision counts averaged around 12,000 cases from 2004 to 2024, but slightly declined in current years.
The major dip in 2020 is due to Covid Impact.

Monthly Distribution:
Collision counts are relatively higher from October to December compared to other months.
The major holiday season falls under this bucket which is a contributing factor.

A Significant Decrease During 2020!



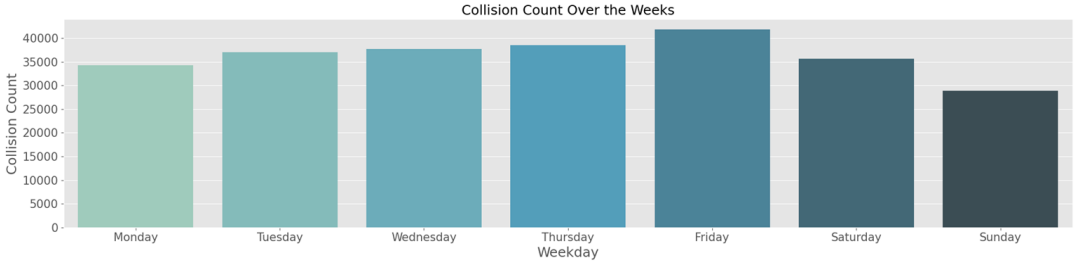
Year	Case per Year
Before 2020	13562.06
2020	7255

The sharp decline in 2020's accident count underscores the impact of reduced mobility. This suggests that traffic volume control can significantly influence accident rates.

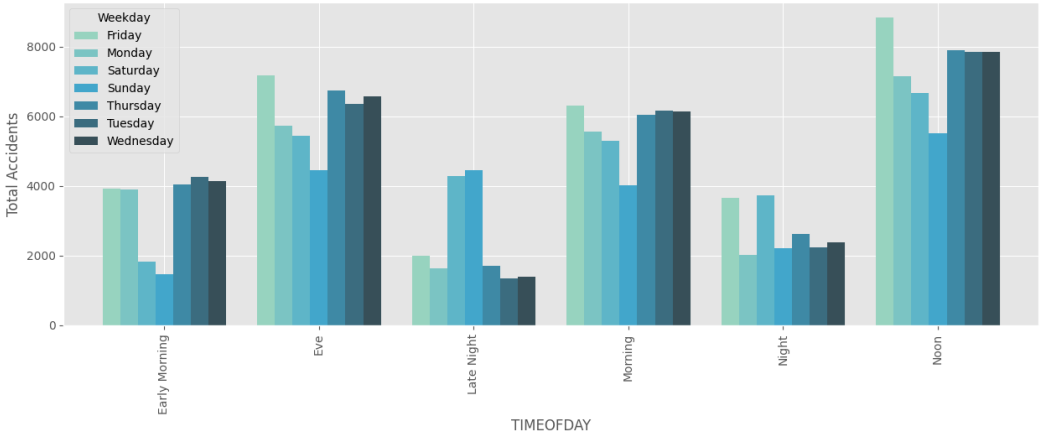
```
SELECT
'Before 2020' AS Year, ROUND(AVG(year_count ), 2) AS "Case per Year"
FROM (
SELECT Year, COUNT(*) AS year_count
FROM sdotcol WHERE Year < '2020' and Year <> '2003'
GROUP BY Year ) AS yearly_counts
Union all
select '2020' as Year, Count(*) as "Case per Year"
from sdotcol Where Year = '2020'
```

Before 2020, the average number of road accidents per year was **13,562.06**, while in 2020, it dropped to **7,255**, reflecting the impact of COVID-19 restrictions on road activity.

Does the day of the week result in more accidents?



Fridays show the highest collision rates, indicating that end-of-week traffic and behavior may be contributing factors.



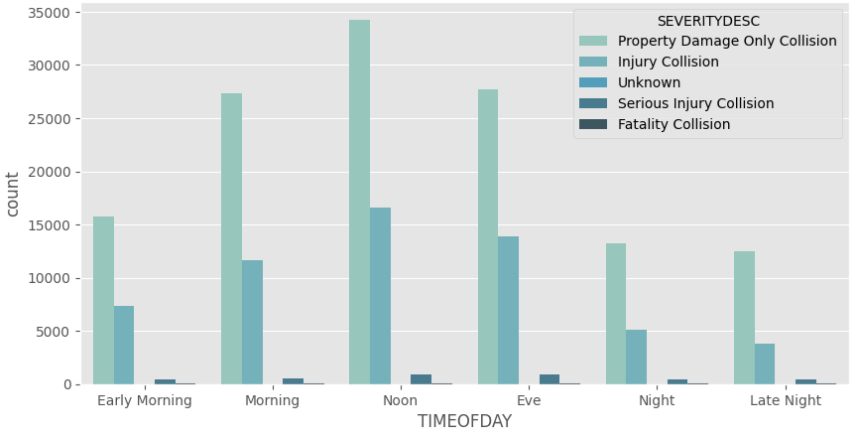
Noon and evening hours are the most accident-prone periods, possibly due to **increased traffic, fatigue, and distractions**.

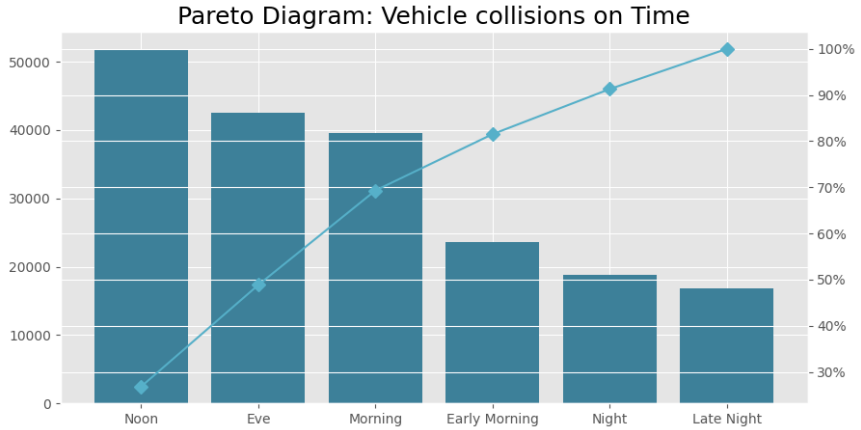
Weekly Distribution:
Collision counts are highest on **Fridays** and lowest on Sundays.

Time of Day Distribution:
Noon periods exhibit the highest collision counts, while night times have the lowest.

Time of Day	From to	%
Late Night	00:00 to 04:00	9%
Early Morning	04:00 to 08:00	12%
Morning	08:00 to 12:00	20%
Noon	**12:00 to 16:00**	**27%**
Eve	16:00 to 20:00	22%
Night	20:00 to 00:00	10%

Is Time of the Day a factor for causing accidents?





Severity by Time of Day:

The highest collision counts occur during noon, evening, and morning hours, with property damage-only collisions being the most common. During late-night, severe collisions are rare. 3 of 4 collisions happening during late-night are just doing property damage.

Pareto Analysis:

A significant portion of collisions happen during noon and evening, accounting for the majority of incidents.

Year	Month	Weekday	TimeOfDay	TotalCase
2015	5	Friday	Noon	73
2005	4	Friday	Noon	72
2012	11	Friday	Eve	70

46.5%

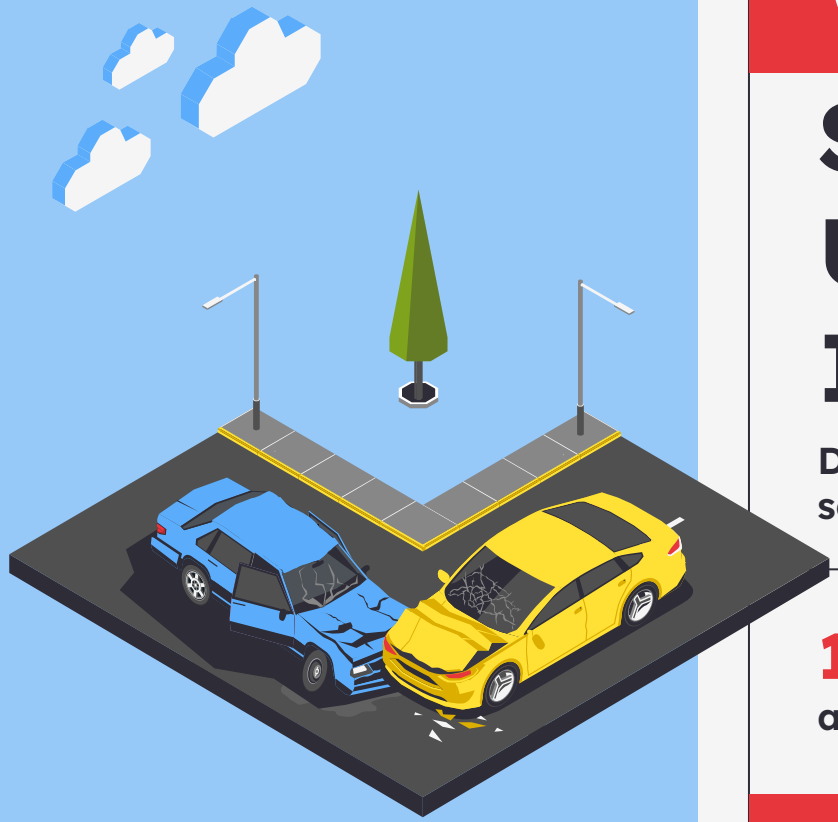
Less collisions in 2020 due to Covid-19

99.75

More cases in October on average

48.83%

Collisions happened at noon and evening



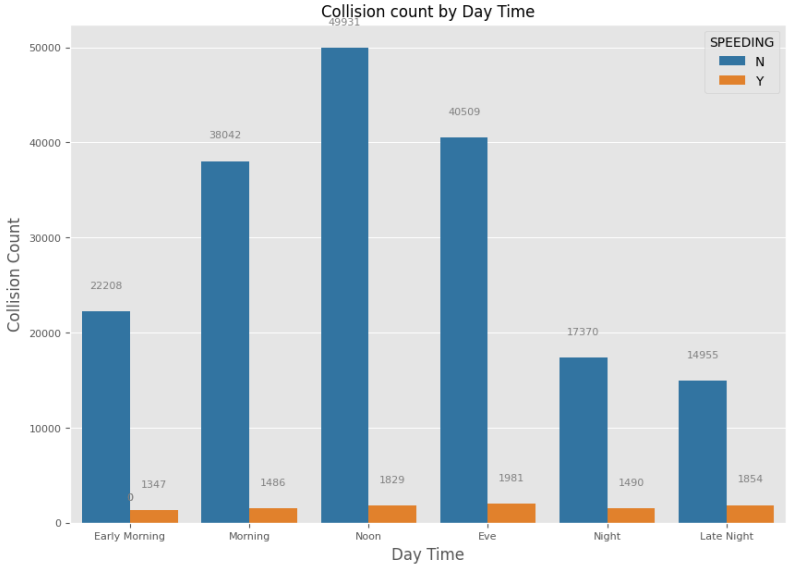
02

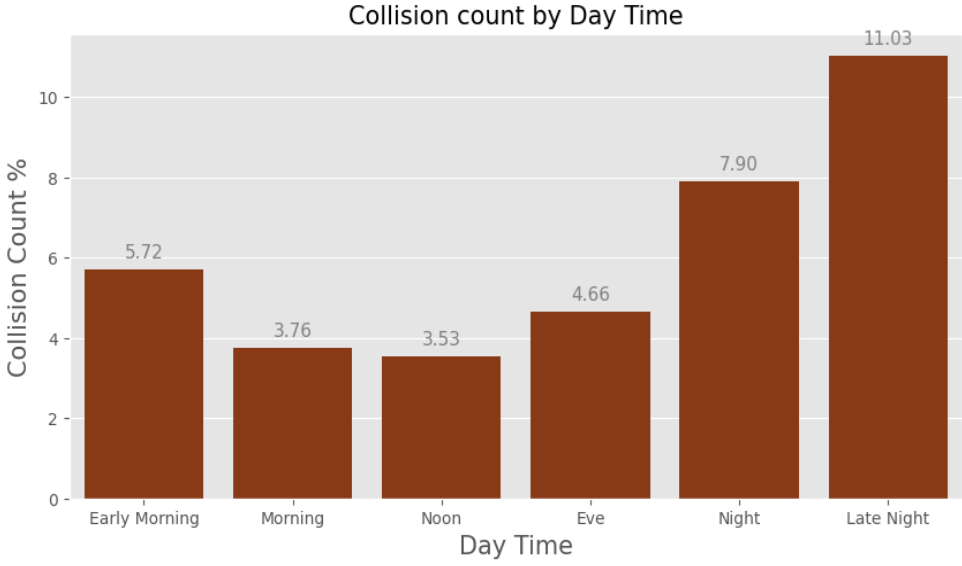
Speeding, Under influence, Inattentiveness

Did these three factors contribute to the severity of the accidents?

19.18% of vehicle collisions are associated with at least one of these factors.

50 When Do Speeding-Related Collisions Occur Most Frequently?

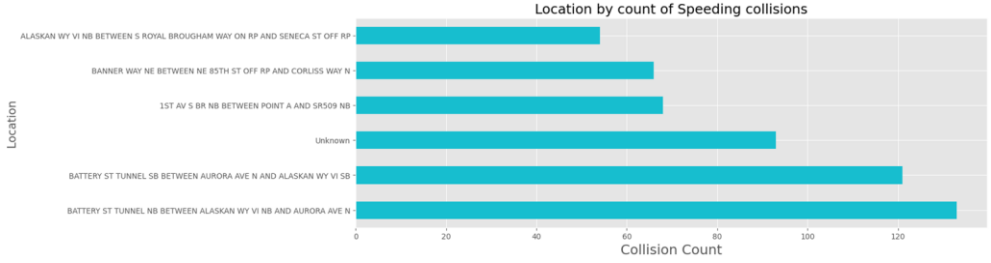




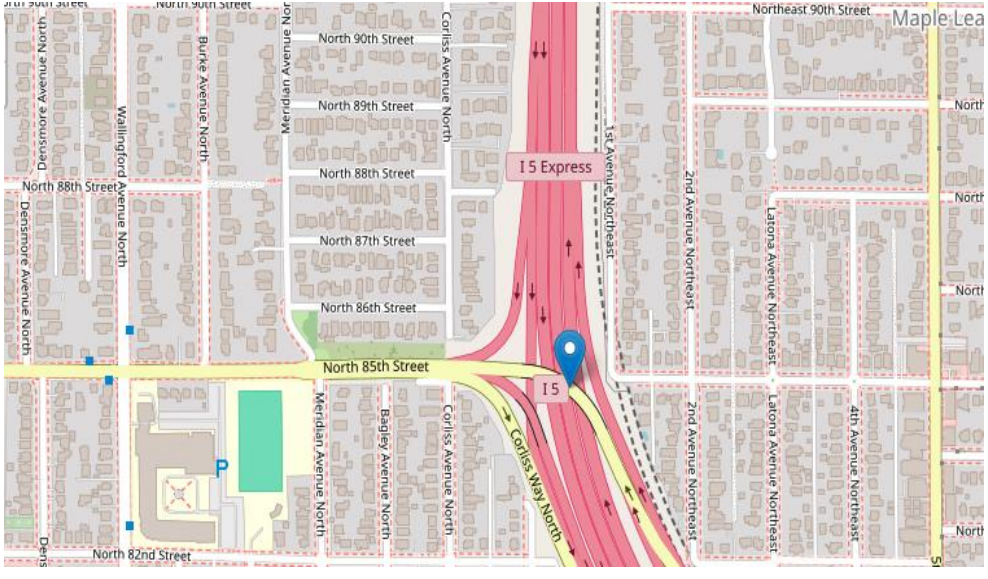
Collision by time:
Collisions caused by speeding remain relatively consistent throughout the day, whereas non-speeding accidents show significant variation depending on the time of day.

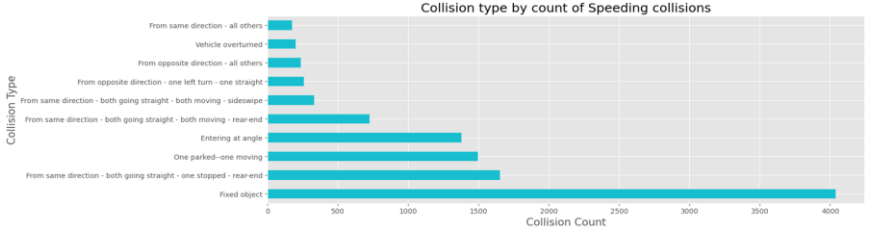
Collisions involving speeding in a day:
During late-night hours, there is a significant increase in the percentage of speeding-related collisions. Accidents involving speeding are nearly **three times more likely** at night compared to noon..

50 Where and How Do Speeding Collisions Occur?



Speeding collision by location:
The Battery Street Tunnel appears among the top two locations for speeding-related incidents. This suggests that the road's design may contribute to higher instances of speeding.

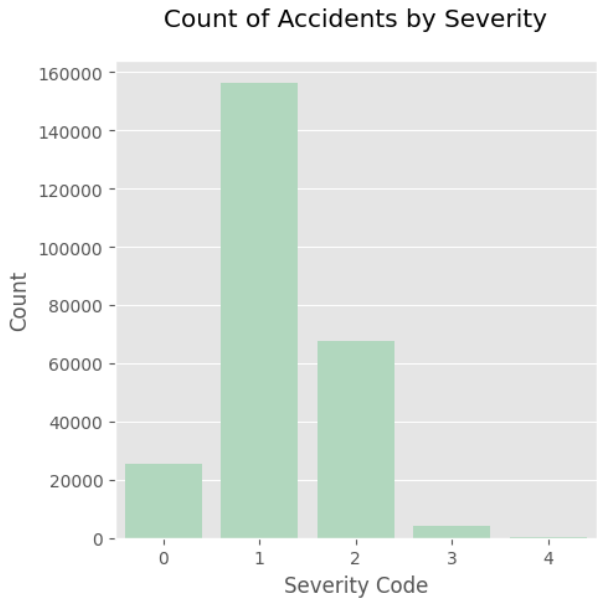




Speeding collision type:
Collisions involving fixed objects at high speeds account for a significant proportion of all collision types. This is likely because speeding can cause tunnel vision, reducing a driver's ability to quickly detect and react to fixed objects on the road.

How Does Speeding Impact Accident Severity?

Severity Count
Accidents are mostly focusing on property damage only collision.



Speeding	Fatality Ratio
N	0.152 %
Y	**1.009 %**

- Based on the risk ratio, the risk of fatality in speeding-related accidents is approximately 6.64 times higher than in non-speeding accidents.

Severity code	Description	Total Case	Speeding(N)	Speeding(Y)
0	Unknown	25349	100%	0%
1	Light (Property damage, light injuries)	156347	95.7%	4.3%
2		67622	93.9%	6.1%
3	Severe (Serious injuries, Fatilities)	4082	87.9%	12.1%
4		484	76%	**24%**

- Speeding is a significant risk factor for high-severity accidents.
- In the most severe accidents (SEVERITYCODE = 4), **24%** involved speeding. In contrast, minor accidents were almost entirely unrelated to speeding.
- Although speeding-related accidents are fewer in number, their fatal risk is **several times higher**, making speeding one of the primary contributors to severe and fatal crashes.

11.03%

Accidents associated with speeding happened at night and late night.

6.64 times

higher is the risk of fatality in speeding-related accidents compared to non-speeding accidents, based on the risk ratio.

12.17%

of speeding-related accidents are associated with driving under the influence.

So, let's take a look at the impact of driving under influence!

By combining strict enforcement, improved infrastructure, and data-driven interventions, Seattle can mitigate the risks associated with speeding.

Increase Late-Night Enforcement:

- Deploy **additional patrol units** during late-night hours, especially in high-risk.
- Implement **random breath testing** to mitigate the combined risk of speeding and DUI.

Implement Speed Monitoring Systems:

- Install **automated speed cameras** in known high-risk zones.
- Utilize **dynamic speed display signs** to remind drivers of their current speed in critical areas.

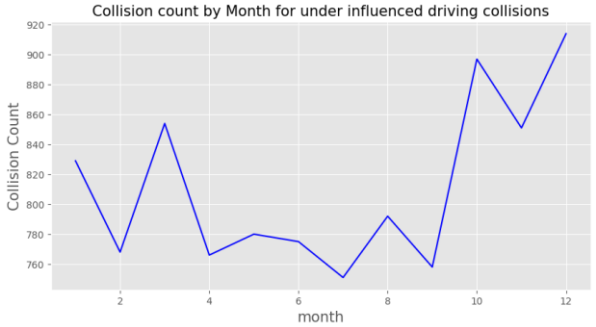
Harsher Penalties for Speeding:

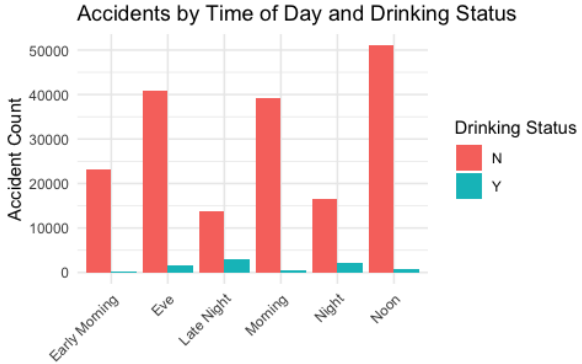
- Enforce stricter fines and license penalties for **repeat speeding offenders**.
- Introduce **graduated penalties** where fines increase based on the severity of speeding.

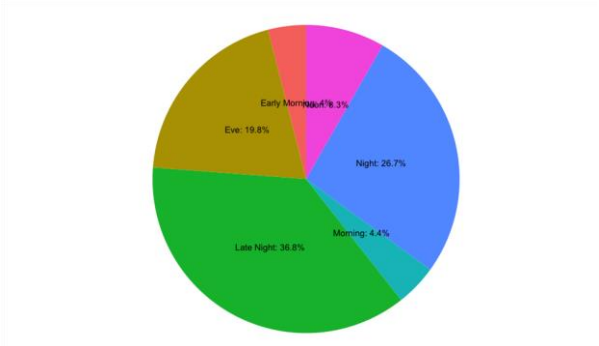
Enhance Road Design in High-Risk Areas:

- Redesign critical zones (e.g., tunnels and sharp curves) with **speed-reducing features** such as rumble strips or sharper curve indicators.

When Are Under Influence Collisions Most Likely to Occur?







The accident counts fluctuate across the months, with **peaks** in **October and December**, and the **lowest** collision counts appear in **July**.

The spike in December could correlate with **holiday seasons**, where drinking and driving behavior might increase.

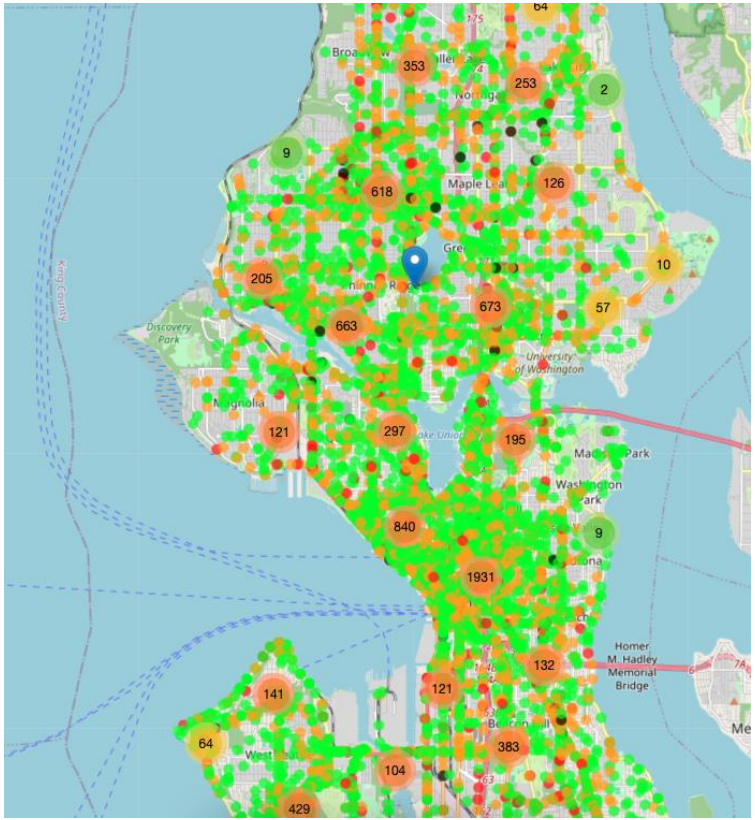
Across all time periods, non-drinking drivers (N) account for the majority of accidents.

However, for **Late Night** and **Night** periods, the proportion of under-influenced accidents significantly **increases** compared to other periods.

The **Late Night** period shows the **highest proportion** of drinking or drug-using related accidents, with more than **36%** of accidents involving alcohol.

Conversely, **Morning** periods exhibit **minimal proportions**, suggesting that drinking and driving are less prevalent.

Where and How Do Under Influence Collisions Occur?



Ranking	Location	Count
1	ALASKAN WY VI SB BETWEEN COLUMBIA ST ON RP AND ALASKAN WY VI SB EFR OFF RP	27
2	BATTERY ST TUNNEL NB BETWEEN ALASKAN WY VI NB AND AURORA AVE N	19
3	1ST AVE BETWEEN BLANCHARD ST AND BELL ST	19
4	NW MARKET ST BETWEEN 3RD AVE NW AND NW 55TH ST	17
5	BATTERY ST TUNNEL SB BETWEEN AURORA AVE N AND ALASKAN WY VI SB	16

Collision Type	N (%)	Y(%)
Other	12.10%	28.70%
Parked Car	24.00%	27.20% ↑
Rear Ended	17.60%	14.80%
Angles	18.90%	10.80%
Pedestrian	4.05%	5.89% ↑
…	…	…

Drivers under the influence are significantly more likely to be involved in collisions with **parked cars or pedestrians** compared to non-drinking drivers.

December

Has the highest number of DUI-related accidents

36.8%

Accidents associated with DUI happened at night and late night.

1.45 times

higher is the risk of hitting a pedestrian in DUI accidents compared to non-influence accidents

What is the impact of inattentiveness?

Recommendations and Improvement Strategies for Reducing Under Influence Collisions

Analysis reveals a **high overlap in locations prone to both speeding and DUI-related collisions**, with areas like the **Battery Street Tunnel** and **Alaskan Way Viaduct** consistently ranking among the top high-risk zones for both behaviors.

Focus on Night and Late-Night Hours:

- Increase DUI Checkpoints during these high-risk periods, particularly near nightlife areas.
- Conduct random sobriety tests to deter drunk driving during holiday seasons.

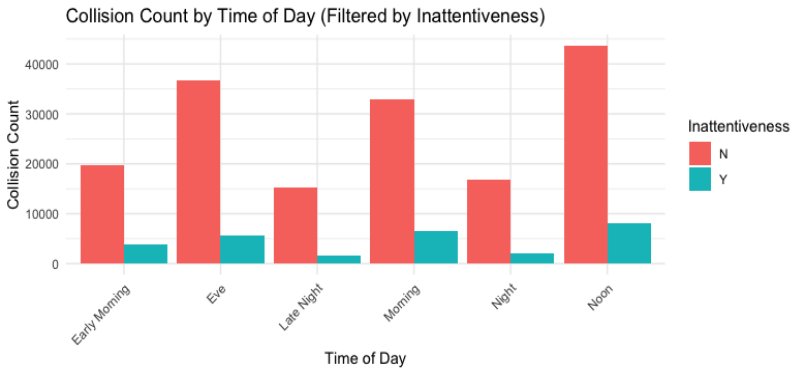
Designing Barrier-Protected Parking Zones for Safer Streets

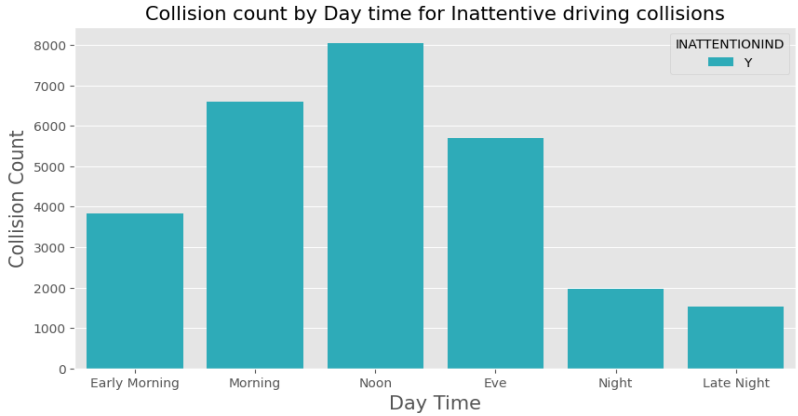
- Creating barrier-protected parking zones to minimize the impact of collisions with parked vehicles.

Enhance Road Design For Safer Pedestrian Pathway:

- Redesigning high-risk pedestrian zones, such as installing raised crosswalks, better street lighting, and speed bumps in nightlife districts.

When Do Inattentiveness-Related Collisions Occur Most Frequently?





- **Morning and Noon Peaks**:

Inattentiveness-related collisions occur most frequently during the **morning** and **noon** periods, suggesting that drivers may be less attentive during commuting hours or midday traffic.

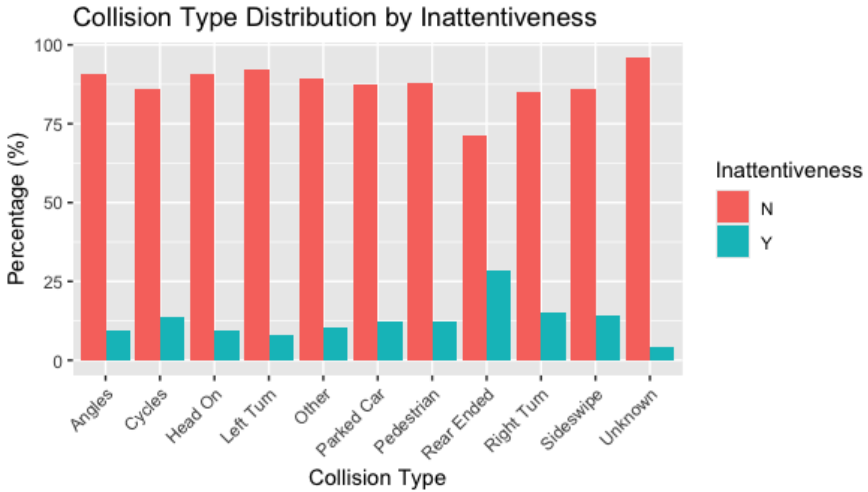
- **Decreased Collisions at Night**:

There is a significant drop in inattentiveness-related collisions during the **night** and **late-night** periods. This could be due to lower traffic volumes or increased driver caution during these hours.

- **Potential Factors**:

The peaks during morning and noon may be influenced by factors such as **rush hour stress, multitasking, or fatigue**, emphasizing the need for targeted safety interventions during these periods.

Inattentiveness Increases Risk of Rear-End Collisions



- **Rear-end collisions account for a significant proportion of inattentiveness-related accidents.**
- **35.64%** of inattentive collisions involve **rear-end** incidents, significantly higher than the **12.30%** for attentive drivers. This suggests that inattentiveness strongly contributes to rear-end accidents, likely due to delayed reaction times.

Collision Type Distribution by Attentiveness

	Proportion of Accidents (%)	
Collision Type	**Attentive (%)**	**Inattentive (%)**
Angles	16.59	11.84
Cycles	2.72	3.08
Head On	1.08	0.76
Left Turn	6.57	3.91
Other	11.30	9.40
Parked Car	20.83	20.95
Pedestrian	3.58	3.47
Rear Ended	12.30	35.64
Right Turn	1.28	1.59
Sideswipe	8.18	9.35
Unknown	15.57	0.00

52.86%

Accidents associated with inattentiveness happened at morning and noon.

35.64%

of inattentive collisions involve rear-end incidents

16.67%

Friday accounts for the highest proportion, while Sunday has the lowest at 10.39%.

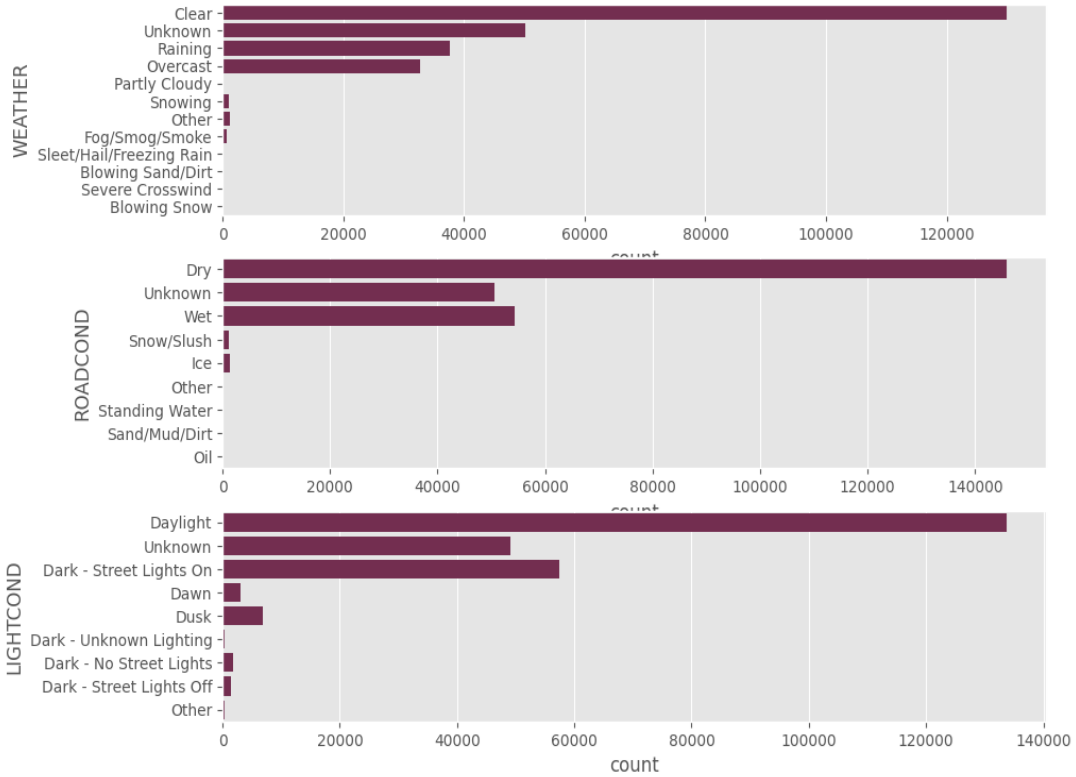
Next, let's explore the insights from statics model.

03

Impact of Other Conditions

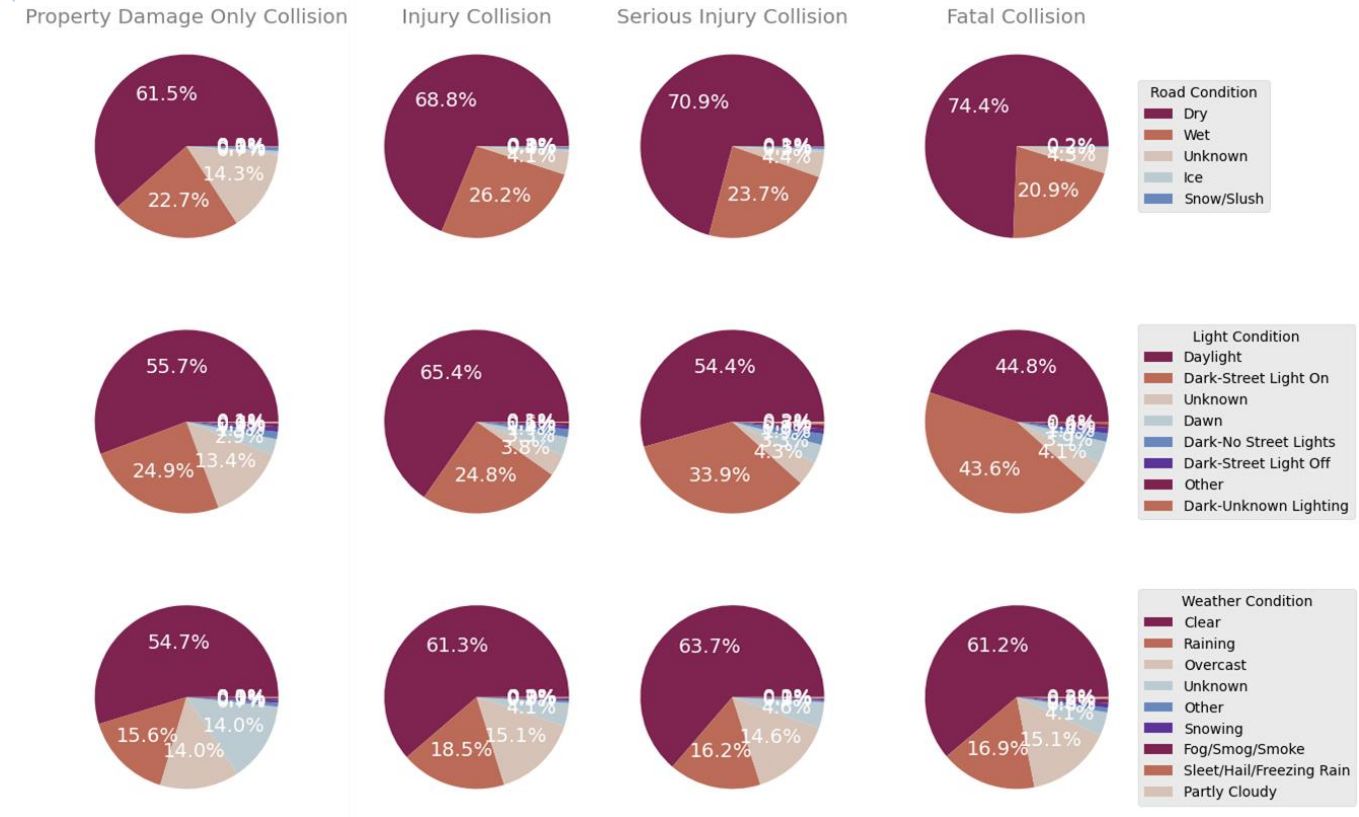
Do Weather, Road & Light Conditions affect collisions?

Impact of Weather, Road and Light Conditions

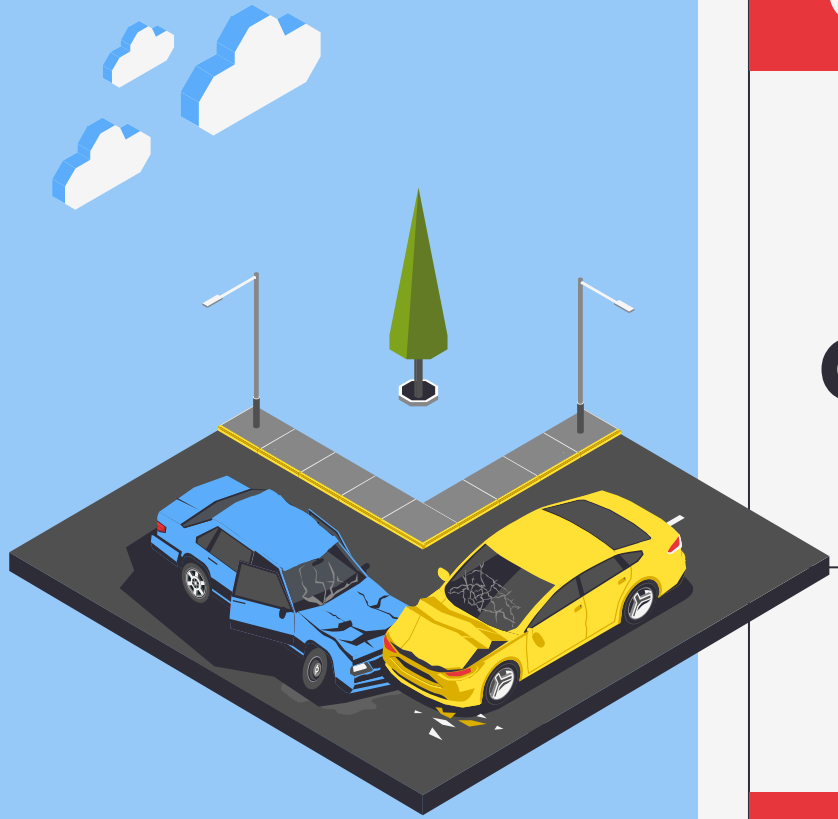


- Dry road surfaces have significantly higher collision counts compared to wet or icy conditions.
- More accidents occur during daylight hours compared to nighttime or low-light conditions.
- While clear weather dominates, there are still a considerable number of accidents under conditions like rain, fog, and snow.
- People could be more cautious while driving in adverse conditions.
- Other factors like traffic density or driver behavior may play a more significant role in collisions.

Damage due to different conditions



- Injury collisions have a relatively higher percentage of wet/icy road conditions compared to property damage-only collisions.
- Fatal collisions have a notable proportion occurring in low-light conditions (e.g., "Dark-Street Light On" and "Dark-No Street Lights"), indicating reduced visibility as a risk factor.
- One is most likely to engage in injury or fatal collision in case of raining or overcast weather conditions.



04

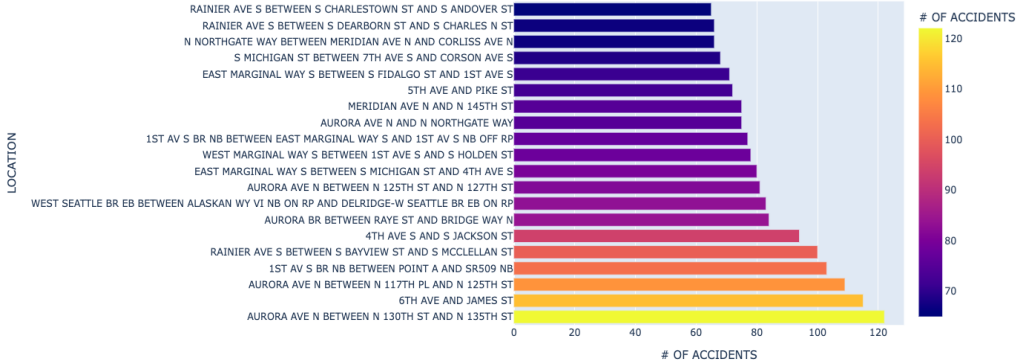
Collision Analysis

What are the Key Collision Patterns and High-Risk Locations Identified?

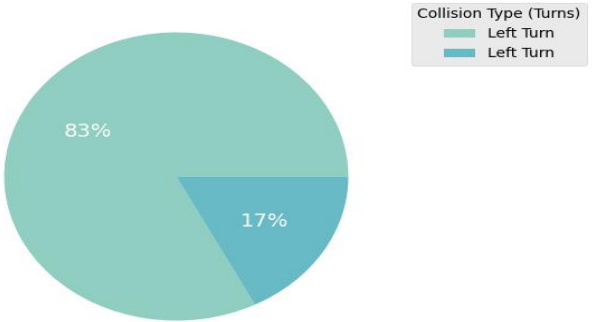
Key Collision Patterns and High-Risk Locations Identified

Highest proportion of all types, representing **45.22%** of the total incidents.

Collision Type	Common Scenarios	(%)
VEHICLE <> VEHICLE, FRONT END AT ANGLE	**Occurs at intersections due to failure to yield or red-light running**	**45.22**
VEHICLE <> VEHICLE, REAR END	Happens due to tailgating, distracted driving, or sudden braking	29.67
VEHICLE <> VEHICLE, LEFT SIDE SIDESWIPE	Occurs during improper lane changes or failing to check blind spots	5.28
MOTOR VEHICLE RAN OFF ROAD - HIT FIXED OBJECT	Results from speeding, poor weather, or loss of control	5.04
MOTOR VEHICLE STRUCK PEDESTRIAN	Happens at crosswalks or intersections due to failure to yield	3.95



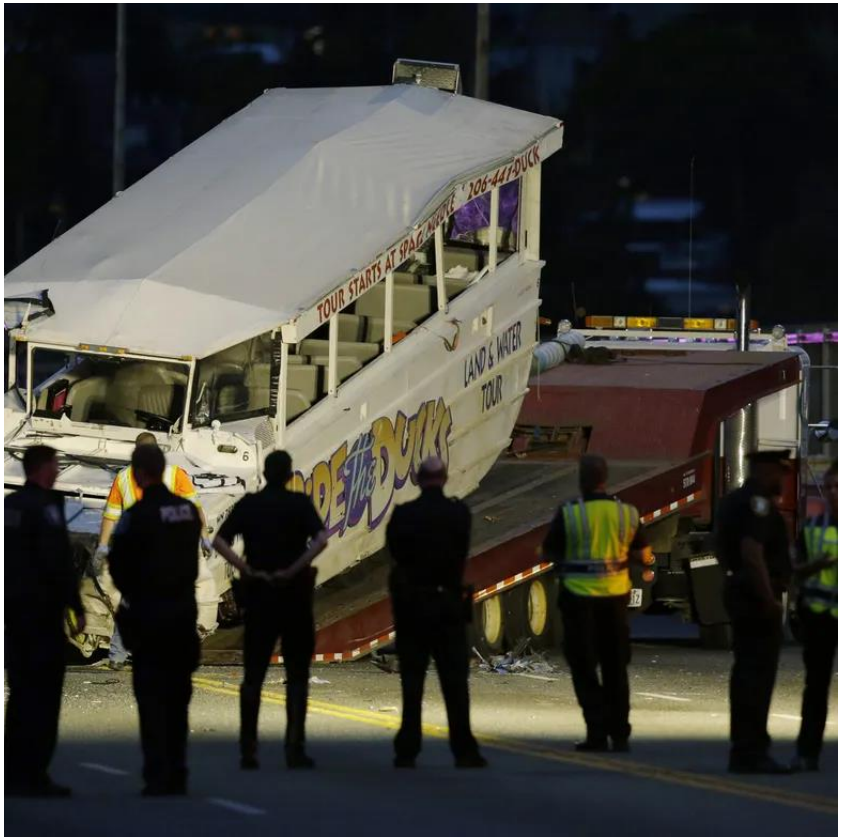
Left Turns Are **4.88 Times More** Common Than Right Turns



The highest number of accidents occurred along **Aurora Ave N between N 130th - N 135th St 98133**, followed by **6th Ave and James St 98104**.

There is ~65% more chance to encounter a collision while taking a Left turn than taking a Right turn. This stat indicates that people are more complacent while taking a left turn than taking a right turn.

2015 Aurora Bridge Collision: Seattle's Deadliest Traffic Accident



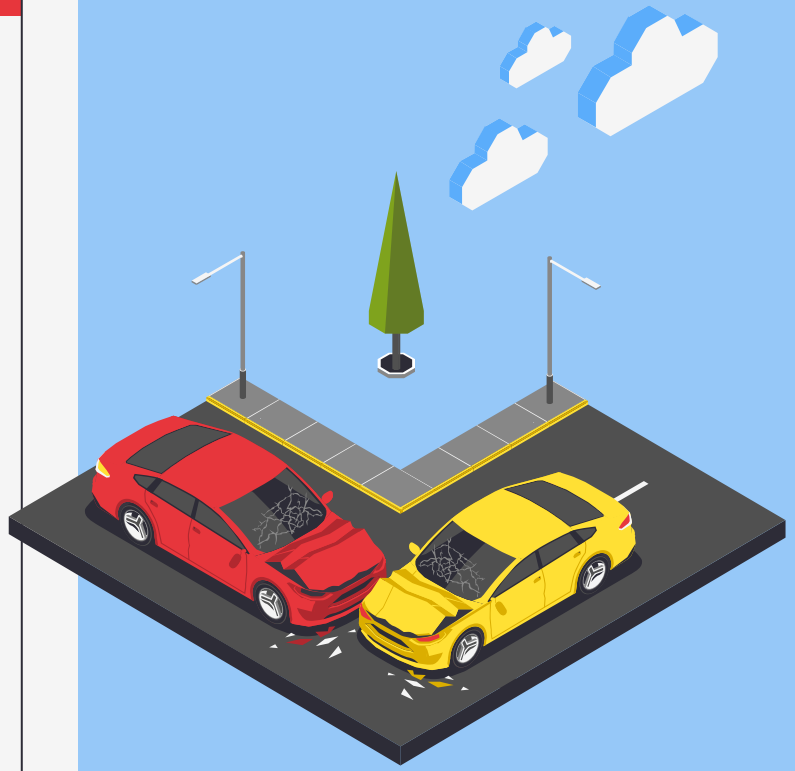
On September 24, 2015, a tragic accident occurred on Seattle's Aurora Bridge involving a Ride the Ducks amphibious vehicle and a charter bus. The collision resulted in five fatalities and left fifty individuals injured.

- **Date:** September 24, 2015
- **Location:** Aurora Bridge, Seattle, WA
- **Vehicles Involved:** Ride the Ducks amphibious vehicle and a charter bus.
- **Casualties:** Five deaths and fifty injuries

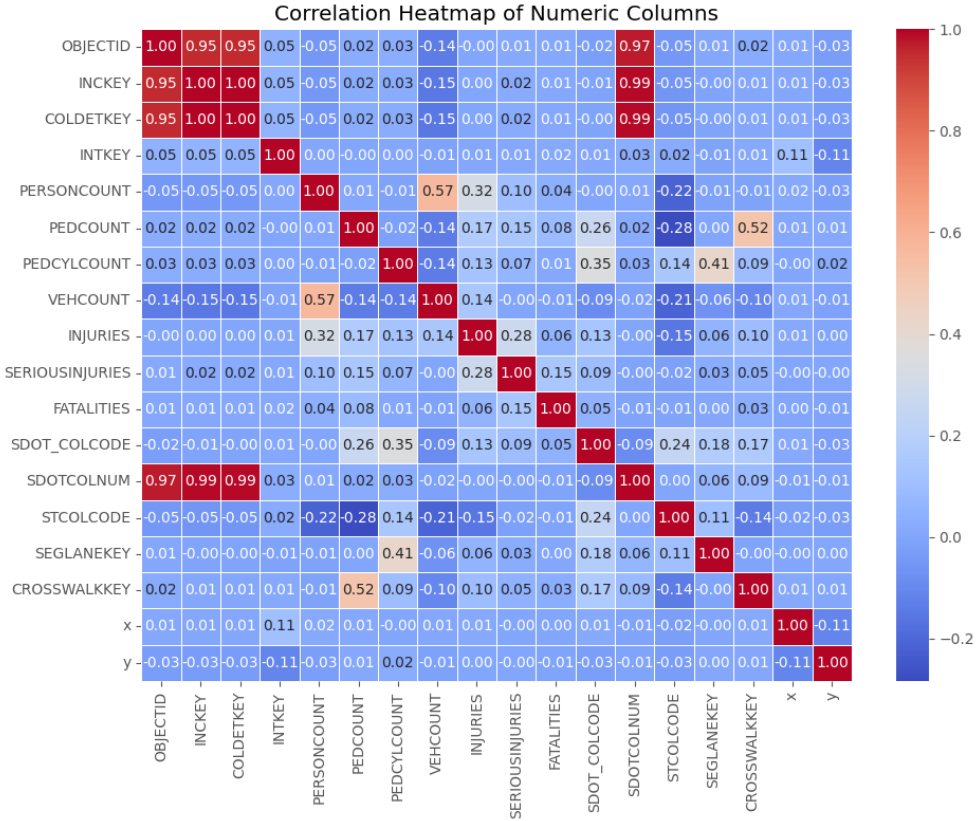
How Did We Find It?

We aimed to identify the most severe collision in terms of casualties. Through an SQL query, we pinpointed this specific record, which led us to uncover the tragic news article detailing this devastating incident.

DATA ANALYSIS



How are numeric columns correlated?



- The dataset contains several identifiers (OBJECTID, INCKEY, COLDETKEY, and SDOTCOLNUM) that are highly correlated, likely representing event or incident IDs.
- Variables like PERSONCOUNT, VEHCOUNT, and INJURIES show meaningful relationships, suggesting these factors could be important for analyzing incidents involving multiple people or vehicles.
- Spatial coordinates (X, Y) and categorical codes (STCOCODE) appear to have little correlation with other numeric variables, implying they may serve a different purpose in analysis (e.g., location mapping).
- A moderate positive correlation between person count and vehicle count suggests that as the number of people involved increases, the number of vehicles involved also tends to increase.
- A weaker but positive correlation between injuries and person count indicates that injuries increase with the number of people involved.

Impact of Speeding, DUI, and Inattentiveness on Severity

```
> # Speeding vs Severity
> chisq_speeding <- chisq.test(table(filtered_data$SPEEDING, filtered_data$SEVERITYCODE))
> print(chisq_speeding)

        Pearson's Chi-squared test

data:  table(filtered_data$SPEEDING, filtered_data$SEVERITYCODE)
X-squared = 775.57, df = 3, p-value < 2.2e-16

>
> # Under Influence (DUI) vs Severity
> chisq_underinfl <- chisq.test(table(filtered_data$UNDERINFL, filtered_data$SEVERITYCODE))
> print(chisq_underinfl)

        Pearson's Chi-squared test

data:  table(filtered_data$UNDERINFL, filtered_data$SEVERITYCODE)
X-squared = 634.28, df = 3, p-value < 2.2e-16

>
> # Inattentiveness vs Severity
> chisq_inattention <- chisq.test(table(filtered_data$INATTENTIONIND, filtered_data$SEVERITYCODE))
> print(chisq_inattention)

        Pearson's Chi-squared test

data:  table(filtered_data$INATTENTIONIND, filtered_data$SEVERITYCODE)
X-squared = 235.5, df = 3, p-value < 2.2e-16
```

All three behaviors are strongly associated with accident severity, with p-values < 2.2e-16, indicating highly significant results.

Impact Strength (Based on Chi-Square Statistic):

- **Speeding (775.57)** has the **strongest influence** on accident severity.
- **DUI (634.28)** is also a major contributing factor.
- **Inattentiveness (235.5)**, while significant, has a relatively smaller impact compared to Speeding and DUI.

Logistic Regression Model Results

This is a **logistic regression model** analyzing the factors contributing to **severe accidents (`SevereAccident`)**, with the following variables:

- `ROADCOND` (Road Conditions)
- `LIGHTCOND` (Lighting Conditions)
- `SPEEDING` (Speeding Behavior)
- `UNDERINFL` (Under Influence - alcohol or drugs)
- `INATTENTIONIND` (Driver Inattention)

Model Type:

- **Logistic Regression** using a **binomial family**.
- Suitable for binary classification where the outcome variable SevereAccident is binary (e.g., 1 for severe accidents and 0 for non-severe).

```
Call:
glm(formula = SevereAccident ~ LIGHTCOND + ROADCOND + SPEEDING +
    UNDERINFL + INATTENTIONIND, family = "binomial", data = filtered_data)

Coefficients:
                                Estimate Std. Error  z value Pr(>|z|)
(Intercept)                    -3.91239    0.02290 -170.817  < 2e-16 ***
LIGHTCONDDark - Street Lights On  0.29430    0.03514    8.375  < 2e-16 ***
LIGHTCONDOther                  0.77083    0.32517    2.371  0.01776 *
ROADCONDIce                    -0.61626    0.22267   -2.768  0.00565 **
ROADCONDSnow/Slush             -0.86993    0.28113   -3.094  0.00197 **
ROADCONDWet                    -0.29143    0.03847   -7.575 3.59e-14 ***
SPEEDINGY                       0.93499    0.04872   19.191  < 2e-16 ***
UNDERINFLY                      0.70670    0.05394   13.101  < 2e-16 ***
INATTENTIONINDY                -0.55965    0.05718   -9.787  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 38800  on 186776  degrees of freedom
Residual deviance: 37939  on 186768  degrees of freedom
AIC: 37957
```

Key Insights from Logistic Regression Model

1. `SPEEDING` and `Driving Under Influence`

These two behaviors are identified as the **most significant risk factors** for severe accidents.

- Both variables have **strong positive estimates** with highly significant p-values, indicating that:
  - Drivers who **speed** or drive **under the influence** are **substantially more likely** to be involved in severe collisions.

2. `INATTENTION`

- **Inattention Shows an Opposite Impact on Severity Compared to Speeding and DUI**
- This suggests that inattentive driving may lead to **less severe collisions**, potentially due to slower speeds or less aggressive behavior.
- However, this does **not mean inattentiveness is safe**. Instead, it emphasizes that:
  - Severe accidents are more often linked to **aggressive behaviors** (like speeding or DUI).
  - Minor distractions (like using a phone or adjusting controls) may lead to **lower-impact accidents**, but they still pose significant risks.

Key Insights from Logistic Regression Model

3. Lighting Conditions

- **Driving in darkness without streetlights** significantly increases the risk of severe accidents.
- **Driving in darkness with streetlights** shows a **slightly elevated risk**, but it's less severe compared to complete darkness.

4. Road Conditions

- **Wet conditions** show a slight **decrease in severe accident risk**, possibly due to reduced traction and longer stopping distances.
- Interestingly, **icy or snowy conditions** show a **negative association** with severe accidents:
    - **Ice** and **Snow/Slush** are linked to **lower accident severity**.
    - This could be because drivers are **more cautious** under dangerous road conditions, reducing speed and increasing focus.

Logistic Regression Results - Odd Ratio

```
> # Calculate Odds Ratios
> exp(coef(model_severity))
                    (Intercept) LIGHTCONDDark - Street Lights On
                      0.0199926                         1.3421878
                  LIGHTCONDOther                      ROADCONDIce
                      2.1615692                         0.5399608
              ROADCONDSnow/Slush                      ROADCONDWet
                      0.4189811                         0.7471945
                     SPEEDINGY                       UNDERINFLY
                      2.5471946                         2.0272832
                INATTENTIONINDY
                      0.5714089
```

The logistic regression model estimates the impact of various factors on accident severity, with results presented in odds ratios (ORs).

Odds ratios indicate how different factors increase or decrease the likelihood of a severe accident.

Understanding Odds Ratios (ORs)

- **OR > 1** → **Increases** the likelihood of a severe accident.
- **OR < 1** → **Decreases** the likelihood of a severe accident.
- **OR = 1** → No significant impact on severity.

Speeding (OR=2.55)

Speeding increases the likelihood of severe accidents by a factor of 2.55, indicating a significantly elevated risk compared to non-speeding incidents.
→ Speeding reduces reaction time and increases impact force, leading to a higher probability of severe or fatal crashes.

Under Influence (DUI) (OR = 2.03)

DUI raises the probability of severe accidents by 2.03, underscoring its association with high-impact collisions.
→ Impairment from alcohol or drugs affects reaction time, motor control, and decision-making, significantly increasing severity.

Other Lighting Conditions (OR = 2.16)

Low-visibility lighting conditions (comparing with daytime) are associated with a 2.16-fold increase in severe accident risk.
→ Poor lighting impairs depth perception and hazard detection, leading to delayed reactions and an increased likelihood of severe collisions.

Dark with Street Lights (OR=1.34)

Driving in darkness with streetlights present (comparing with daytime) is associated with a 1.34 times higher probability of severe accidents compared to well-lit conditions.
→ Despite the presence of streetlights, reduced overall visibility at night still contributes to higher crash severity due to delayed response times.

Inattention (OR = 0.57)

Inattentive driving is associated with a 43% lower probability of severe accidents compared to other contributing factors.
→ Inattention is often linked to **low-speed collisions**, which tend to result in less severe outcomes than those caused by speeding or DUI.

Wet Road Conditions (OR = 0.75)

Wet road conditions (comparing with dry road) are correlated with a 25% reduction in the likelihood of severe accidents.
→ Drivers tend to reduce speed on wet roads, leading to lower impact forces and a reduced probability of severe crashes.

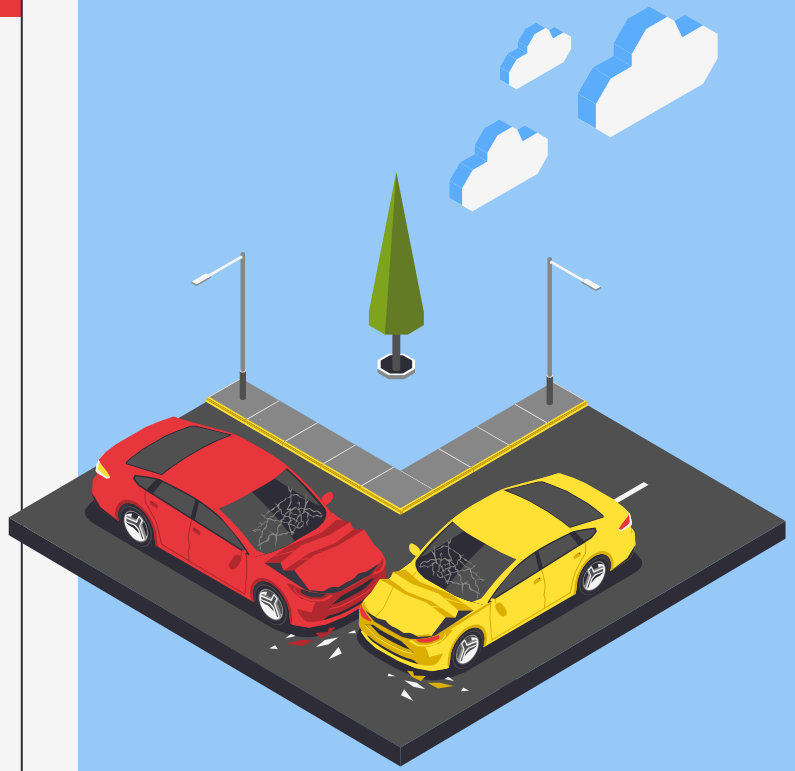
Ice (OR = 0.54)

Icy road conditions (comparing with dry road) decrease the probability of severe accidents by approximately 46%, likely due to altered driving behavior.
→ Drivers adopt more cautious behaviors, such as reduced speed and increased following distances, which mitigate high-impact crashes.

Snow/Slush (OR = 0.42)

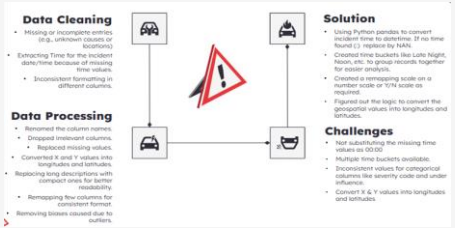
Snowy or slushy road conditions (comparing with dry road) reduce the likelihood of severe accidents by 58%, reflecting the behavioral adjustments.
→ Similar to icy roads, drivers operate vehicles at lower speeds in snowy conditions, thereby reducing the severity of crashes.

STRATEGIC VALUE AND KEY INSIGHTS

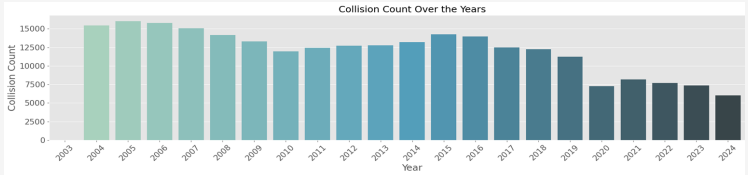


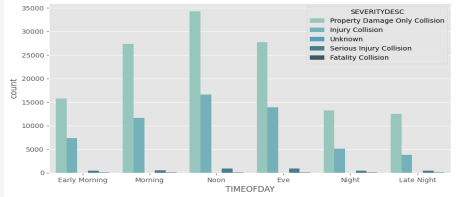
KEY TAKEAWAYS - Seattle Vehicle Collisions Analysis

Summary:
Vehicle collisions in Seattle are influenced by factors such as DUI, Speeding and inattentiveness.



Time Period Analysis

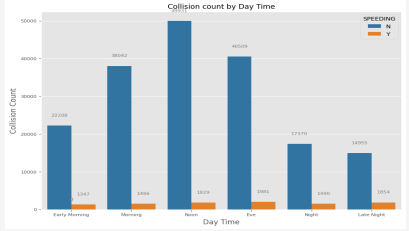


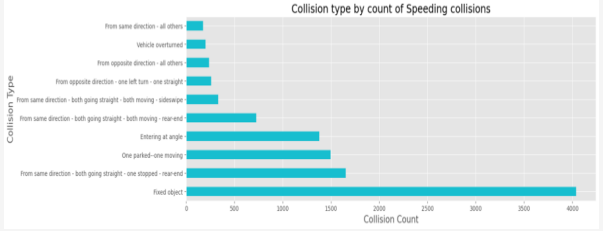


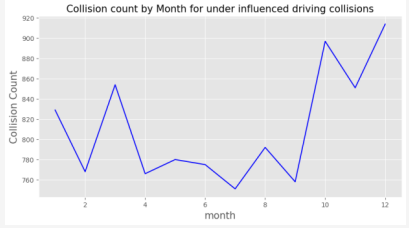
Time of Day	From to	%
Late Night	00:00 to 04:00	9%
Early Morning	04:00 to 08:00	12%
Morning	08:00 to 12:00	20%
Noon	**12:00 to 16:00**	**27%**
Eve	16:00 to 20:00	22%
Night	20:00 to 00:00	10%

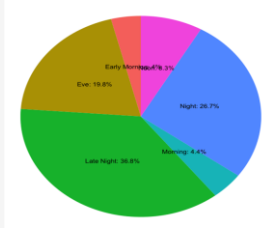
The number of traffic accidents has been gradually decreasing each year, with a notable drop in 2020.The Noon & evening hours show the highest collision counts. Late-night hours have fewer collisions but a higher proportion of severe incidents.

Major factors: Speeding, Under Influence, Inattentiveness



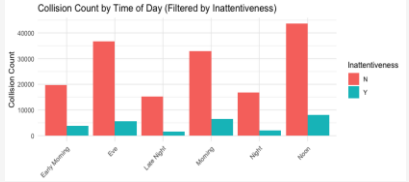


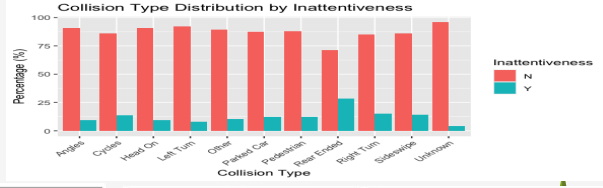




Ranking	Location	Count
1	ALASKAN WY VI SB BETWEEN COLUMBIA ST ON RP AND ALASKAN WY VI SB EFR OFF RP	27
2	BATTERY ST TUNNEL NB BETWEEN ALASKAN WY VI NB AND AURORA AVE N	19
3	1ST AVE BETWEEN BLANCHARD ST AND BELL ST	19
4	NW MARKET ST BETWEEN 3RD AVE NW AND NW 55TH ST	17
5	BATTERY ST TUNNEL SB BETWEEN AURORA AVE N AND ALASKAN WY VI SB	16

Collision Type	N (%)	Y(%)
Other	12.10%	28.70%
Parked Car	24.00%	27.20%
Rear Ended	17.60%	14.80%
Angles	18.90%	10.80%
Pedestrian	4.05%	5.89%
...	...	...





Increase Late-Night Enforcement:
- Deploy **additional patrol units** during late-night hours, especially in high-risk.
- Implement **random breath testing** to mitigate the combined risk of speeding and DUI.

Implement Speed Monitoring Systems:
- Install **automated speed cameras** in known high-risk zones.
- Utilize **dynamic speed display signs** to remind drivers of their current speed in critical areas.

Analysis reveals a **high overlap in locations prone to both speeding and DUI-related collisions**, with areas like the **Battery Street Tunnel** and **Alaskan Way Viaduct** consistently ranking among the top high-risk zones for both behaviors.

Focus on Night and Late-Night Hours:
- Increase DUI Checkpoints during these high-risk periods, particularly near nightlife areas.
- Conduct random sobriety tests to deter drunk driving during holiday seasons.

Harsher Penalties for Speeding:
- Enforce stricter fines and license penalties for **repeat speeding offenders**.
- Introduce **graduated penalties** where fines increase based on the severity of speeding.

Enhance Road Design in High-Risk Areas:
- Redesign critical zones (e.g., tunnels and sharp curves) with **speed-reducing features** such as rumble strips or sharper curve indicators.

Designing Barrier-Protected Parking Zones for Safer Streets
- Creating barrier-protected parking zones to minimize the impact of collisions with parked vehicles.

Enhance Road Design For Safer Pedestrian Pathway:
- Redesigning high-risk pedestrian zones, such as installing raised crosswalks, better street lighting, and speed bumps in nightlife districts.

Key Findings and Strategies for Reducing Seattle Collisions

	Key Findings	Key business improvement directions
Overall	• There are **99.75** more collisions in October than the average for other months. • **48.83%** of collisions occur at noon and in the evening. • **Fridays** typically have more collisions than other days. • **20.85%** of collisions involve parked cars.	• **Reduce October Collisions**: Implement targeted awareness campaigns and increased law enforcement presence in October to address factors contributing to the spike in collisions. • **Improve Noon and Evening Traffic Safety**: Enhance traffic signal timing, increase lighting, and deploy enforcement measures during peak collision hours to reduce accidents. • **Prevent Parked Car Collisions**: Designate clearer parking zones, enforce stricter parking regulations, and encourage the use of backup cameras and parking sensors.
Severity	• **40.08%** of fatal accidents are caused by hitting pedestrians. • **44.83%** of fatal accidents occur at "Dart – street lights on." • **Friday** has the highest number of fatal accidents compared to other days. • In recent years, speeding violations have steadily declined, dropping from 901 cases (8%) in 2008 to **261 cases (2%) in 2024.**	• **Prevent Pedestrian Fatalities**: Implement better pedestrian crossings, add more visible signage, and enforce lower speed limits in high-risk areas. • **Improve Street Lighting Safety**: Upgrade street lighting, install motion-sensor lights in dark areas, and enhance visibility at intersections to reduce accidents. • **Reduce Friday Fatalities**: Increase law enforcement patrols, conduct DUI checkpoints, and promote responsible driving awareness campaigns on Fridays to prevent accidents.
Speeding	• **11.03%** of accidents related to speeding occur at night or late at night. • The risk of fatality in speeding-related accidents is **more than twice** as high as in non-speeding accidents, based on the risk ratio.	• **Enhance Nighttime Speed Enforcement**: Increase police patrols, install speed cameras, and implement stricter penalties for nighttime speeding violations. • **Improve Road Safety at Night**: Enhance street lighting, install reflective road markings, and use warning signs in high-risk speeding areas. • **Educate Drivers on Speeding Risks**: Launch awareness campaigns emphasizing the higher fatality risk in speeding-related accidents, particularly at night.
DUI	• **December** has the highest number of DUI-related accidents. • **Saturday** has the highest number of DUI-related accidents. • **36.8%** of DUI-related accidents occur at night or late at night. • The risk of hitting a pedestrian in DUI-related accidents is **1.45 times higher** than in non-DUI accidents.	• **Increase DUI Enforcement on High-Risk Days**: Implement more DUI checkpoints and police patrols on Saturdays, Sundays, and during peak months like December and October. • **Promote Safe Transportation Alternatives**: Encourage ride-sharing, public transit, and designated driver programs to reduce DUI incidents on weekends and at night. • **Strengthen Public Awareness Campaigns**: Launch educational initiatives highlighting the heightened risk of pedestrian accidents and fatalities in DUI-related crashes.
Inattention	• **52.86%** of accidents associated with inattentiveness happened at morning and noon. • **35.64%** of inattentive collisions involve rear-end incidents • **Friday** accounts for the highest proportion of inattentive collisions at **16.67%**, while Sunday has the lowest at 10.39%.	• **Enhance Awareness During Peak Hours**: Implement public awareness campaigns to remind drivers to stay focused, particularly in the morning and at noon. • **Reduce Rear-End Collisions**: Promote the use of adaptive cruise control and collision warning systems, while also enforcing safe following distances. • **Increase Friday Traffic Safety Measures**: Deploy additional traffic enforcement and safety reminders on Fridays to mitigate the high rate of inattentive collisions.

Transforming Seattle's Road Safety: A Comprehensive Strategy for Smarter, Safer Streets

Project Objective

Aiming 30% less collisions in Seattle by 2035

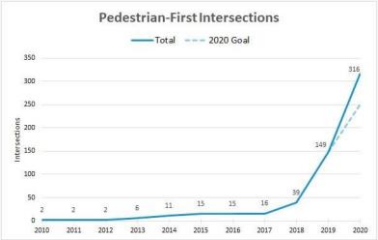
- Smart Traffic Management Systems
- Integrated Public Safety Campaigns
- Urban Infrastructure Optimization
- Advanced Driver Assistance Technologies

Four major processes

1. Smart Mobility Integration

Concept: Utilize technology to optimize traffic flow, enhance transportation efficiency, and reduce congestion-related accidents.
Implementation: Deploy AI-driven traffic management systems, adaptive traffic signals, and real-time monitoring to dynamically adjust traffic patterns in high-risk areas of Seattle.

2. Vision Zero Safety Initiatives



Concept: Adopt a "Zero Fatalities" approach by designing safer streets and prioritizing pedestrian and cyclist protection.
Implementation: Lower speed limits in urban areas, enhance pedestrian crossings, and implement more traffic-calming measures .

3. Sustainable & Multi-Modal Transportation

Concept: Encourage the use of public transit and alternative transportation modes to reduce vehicle dependency and enhance road safety.

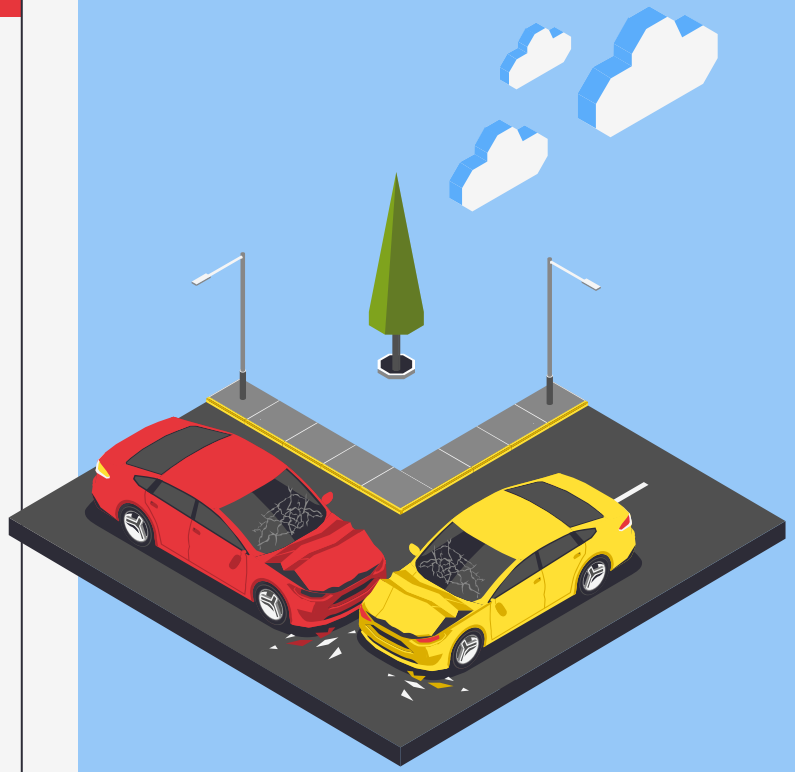
4. Proactive Law Enforcement & Public Awareness

Concept: Strengthen traffic law enforcement and public education to reduce reckless driving behaviors and promote safer roads.
Implementation: Increase DUI and speeding enforcement, install more automated traffic monitoring systems, and launch safety awareness campaigns in Seattle communities to encourage responsible driving.

Roadmap

- Data-Driven Risk Assessment
- Policy & Regulation Enhancement
- Technology-Enabled Traffic Monitoring
- Infrastructure Upgrades
- Public Awareness & Education
- Collaboration with Stakeholders
- Emergency Response Optimization
- Continuous Performance Evaluation

OTHER FACTORS AND DATA



OTHER FACTORS AND DATA

Construction Zones

Analyzing how temporary road work affects collision patterns.

Insurance Claims & Costs

Understanding the financial impact of collisions.

Type, Age & Condition of Vehicles

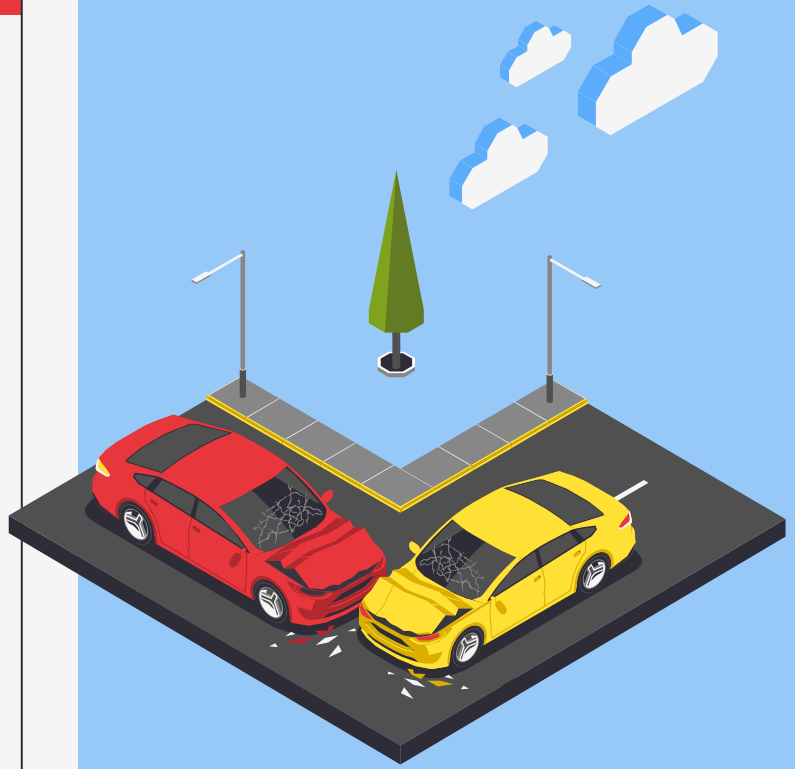
Older or poorly maintained vehicles may contribute to higher accident rates. Assessing if autonomous/ electric vehicles have different collision trends.

Driver & Behavioral Data

Identifying patterns between previous violations (e.g., speeding tickets) and collisions.

Looking at jaywalking incidents or areas with high pedestrian movement.

APPENDIX



APPENDIX

Data Cleaning & Analysis Code - Python

```
8. Convert x and y values into longitudes and latitudes

# Initialize the transformer (State Plane EPSG:2926 to WGS84 EPSG:4326)
transformer = Transformer.from_crs("EPSG:2926", "EPSG:4326", always_xy=True)

# Ensure 'X' and 'Y' columns exist before conversion
if 'Longitude' in df.columns and 'Latitude' in df.columns:
    # Perform vectorized transformation
    df['Longitude'], df['Latitude'] = transformer.transform(df['Longitude'].to_numpy(), df['Latitude'].to_numpy())

    print("Transformation complete. X and Y have been replaced with Longitude and Latitude.")
    print(df.head())

else:
    print("Error: Columns 'X' and 'Y' not found in DataFrame.")

Transformation complete. X and Y have been replaced with Longitude and Latitude.
   OBJECTID SE_ANNO_CAD_DATA  INCKEY  COLDETKEY REPORTNO      STATUS  \
0  30449598    System.Byte[]  363281     364761  3919347     Matched
1  30449599    System.Byte[]    1147       1147  3548946   Unmatched
2  30449600    System.Byte[]  371998     373478  3923038     Matched
3  30449601    System.Byte[]   61600      61600  2617309     Matched
4  30449602    System.Byte[]   68500      68500  2792365     Matched
```

Questions & Insights - SQL

```
SELECT
    CAST(SUM(CASE WHEN sdotcol.INATTENTIONIND  = 'Y' THEN 1 ELSE 0 END) AS FLOAT) /
    NULLIF(SUM(CASE WHEN INATTENTIONIND <> 'Unknown' THEN 1 ELSE 0 END), 0) AS ratio
FROM sdotcol;

select  CAST(SUM(CASE WHEN underinfl = 'Y'  THEN 1
                      ELSE 0 END) AS FLOAT) from sdotcol

SELECT
    CAST(SUM(CASE WHEN sdotcol.INATTENTIONIND  = 'Y' THEN 1 ELSE 0 END) AS FLOAT) /
    NULLIF(SUM(CASE WHEN INATTENTIONIND <> 'Unknown' THEN 1 ELSE 0 END), 0) AS ratio
FROM sdotcol;

SELECT
    CAST(SUM(CASE WHEN sdotcol.TIMEOFDAY  = 'Noon' THEN 1 ELSE 0 END) AS FLOAT) /
    NULLIF(SUM(CASE WHEN TIMEOFDAY <> 'Unknown' THEN 1 ELSE 0 END), 0) AS ratio
FROM sdotcol
where INATTENTIONIND = 'Y';
```

Logistics Regression Code - R

```
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -3.89246    0.16997 -22.901  < 2e-16 ***
ROADCONDIce                     -0.63943    0.22259  -2.873  0.00407 **
ROADCONDSnow/Slush              -0.87998    0.28110  -3.131  0.00175 **
ROADCONDWet                     -0.29786    0.03838  -7.761 8.41e-15 ***
LIGHTCONDDark - Street Lights On  0.27640    0.17098   1.617  0.10597
LIGHTCONDDaylight               -0.01929    0.17059  -0.113  0.90997
SPEEDINGY                        0.93760    0.04847  19.346  < 2e-16 ***
UNDERINFLY                       0.70433    0.05363  13.133  < 2e-16 ***
INATTENTIONINDY                 -0.55498    0.05694  -9.746  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 39062  on 188223  degrees of freedom
Residual deviance: 38196  on 188215  degrees of freedom
AIC: 38214
```

Python Notebook:
https://colab.research.google.com/drive/1U2C5Ou29LnW_HVrq2qjApSS0E0vod5xD?usp=sharing

APPENDIX - Logistics Regression Code-R

```
> # Filter and create binary variable for severe accidents
> filtered_data <- data %>%
+     filter(SEVERITYCODE != 0,
+            !LIGHTCOND %in% c("Unknown",  "Dark - Unknown Lighting" , "Da
rk - No Street Lights" , "Dawn" , "Dusk" , "Dark - Street Lights Off"),
+            !ROADCOND %in% c("Unknown", "Oil", "Sand/Mud/Dirt", "Other",
"Standing Water"),
+            UNDERINFL %in% c("Y", "N")) %>%
+     mutate(
+         SevereAccident = ifelse(SEVERITYCODE %in% c(3, 4), 1, 0),
+
+         # Convert LIGHTCOND to factor and set 'Daylight' as reference
+         LIGHTCOND = factor(LIGHTCOND),
+         LIGHTCOND = relevel(LIGHTCOND, ref = "Daylight")
+     )
>
> # Logistic regression model
> model_severity <- glm(SevereAccident ~ LIGHTCOND + ROADCOND + SPEEDING +
UNDERINFL + INATTENTIONIND,
+                       data = filtered_data, family = "binomial")
>
> # View model summary
> summary(model_severity)

Call:
glm(formula = SevereAccident ~ LIGHTCOND + ROADCOND + SPEEDING +
    UNDERINFL + INATTENTIONIND, family = "binomial", data = filtered_data)

Coefficients:
                              Estimate Std. Error  z value Pr(>|z|)
(Intercept)                  -3.91239    0.02290 -170.817  < 2e-16 ***
LIGHTCONDDark - Street Lights On  0.29430    0.03514    8.375  < 2e-16 ***
LIGHTCONDOther                0.77083    0.32517    2.371  0.01776 *
ROADCONDIce                  -0.61626    0.22267   -2.768  0.00565 **
ROADCONDSnow/Slush           -0.86993    0.28113   -3.094  0.00197 **
ROADCONDWet                  -0.29143    0.03847   -7.575 3.59e-14 ***
SPEEDINGY                     0.93499    0.04872   19.191  < 2e-16 ***
UNDERINFLY                    0.70670    0.05394   13.101  < 2e-16 ***
INATTENTIONINDY              -0.55965    0.05718   -9.787  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 38800  on 186776  degrees of freedom
Residual deviance: 37939  on 186768  degrees of freedom
AIC: 37957
```

```
# Filter and create binary variable for severe accidents
filtered_data <- data %>%
   filter(SEVERITYCODE != 0,
          !LIGHTCOND %in% c("Unknown",  "Dark - Unknown Lighting" , "Dark - No Street Lights" , "Dawn" ,
"Dusk" , "Dark - Street Lights Off"),
          !ROADCOND %in% c("Unknown", "Oil", "Sand/Mud/Dirt", "Other", "Standing Water"),
          UNDERINFL %in% c("Y", "N")) %>%
   mutate(
      SevereAccident = ifelse(SEVERITYCODE %in% c(3, 4), 1, 0),

      # Convert LIGHTCOND to factor and set 'Daylight' as reference
      LIGHTCOND = factor(LIGHTCOND),
      LIGHTCOND = relevel(LIGHTCOND, ref = "Daylight")
   )

# Logistic regression model
model_severity <- glm(SevereAccident ~ LIGHTCOND + ROADCOND + SPEEDING + UNDERINFL +
INATTENTIONIND,
                      data = filtered_data, family = "binomial")

# View model summary
summary(model_severity)
```

```
> # Calculate Odds Ratios
> exp(coef(model_severity))
                  (Intercept) LIGHTCONDDark - Street Lights On
                    0.0199926                       1.3421878
                LIGHTCONDOther                      ROADCONDIce
                    2.1615692                       0.5399608
          ROADCONDSnow/Slush                      ROADCONDWet
                    0.4189811                       0.7471945
                    SPEEDINGY                      UNDERINFLY
                    2.5471946                       2.0272832
                INATTENTIONINDY
                    0.5714089
```

RESOURCES

We referenced these websites below to support our analysis and summary.

Dataset

- https://data.seattle.gov/dataset/SDOT-Collisions-All-Years/qdnv-25h8

The Seattle Times

- https://www.seattletimes.com/seattle-news/ride-the-ducks-vehicle-collides-with-bus-on-aurora-bridge/

Slide Template

- https://slidesgo.com/theme/traffic-collision-prevention#search-traffic+collision&position-1&results-1&rs=search

Kaggle

- https://www.kaggle.com/code/anmoltripathi/vehicle-collision-data-analysis-sdot#Collisions-due-to-inattention
- https://www.kaggle.com/code/hhp07022000/seattle-collison-data-severity-cross-val-100
- https://www.kaggle.com/code/mohamedjafirashraf/predict-the-severity-of-the-collision

Medium

- https://medium.com/intellyticssolutions/data-analysis-beyond-plotting-of-colourful-visualisations-35886facad21

THANKS!