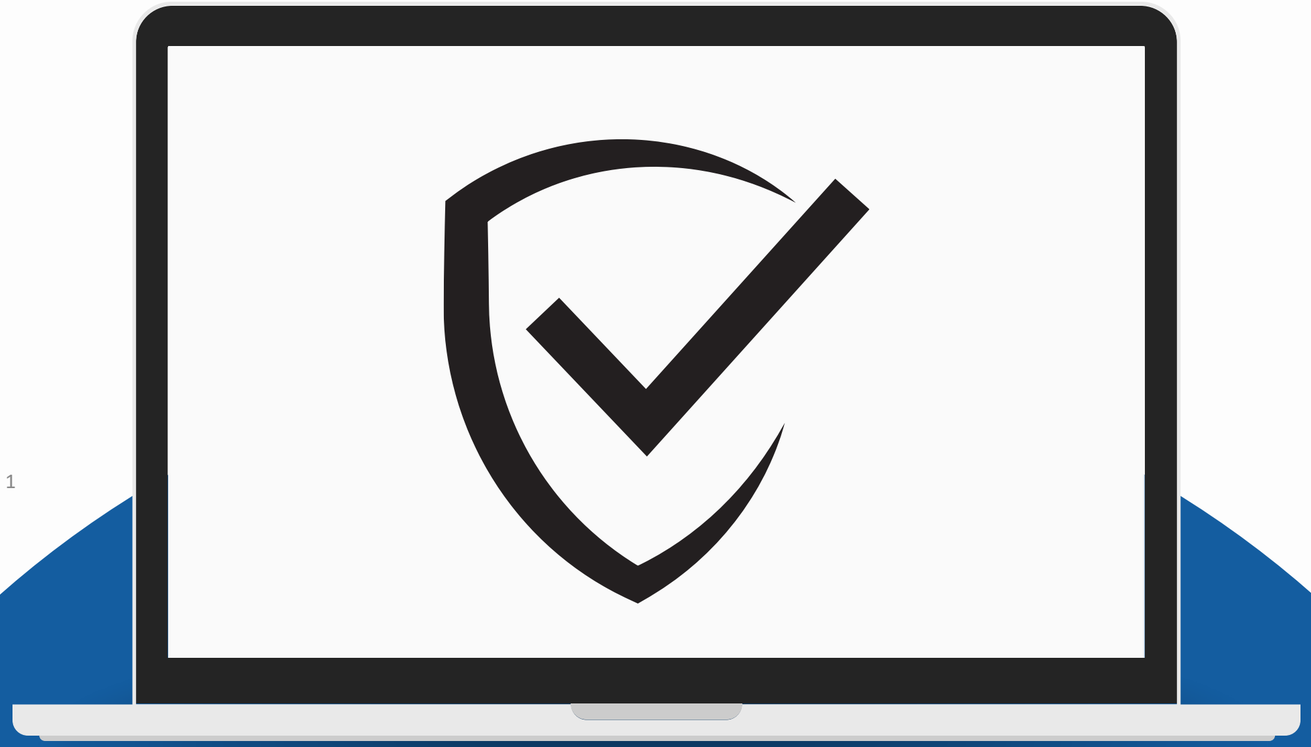
Montreal Traffic Collision Analysis

Shuxi Chen
JaeYoon Lee
Yifei Liu
Hannah Wang



Overview

Problem Statement

Problem Statement

- **Project Focus:** Traffic accident analysis in Montreal, aimed at enhancing road safety and minimizing accidents
- **Objective:**
  - Identify high-risk zones within the city
  - Predict the severity of traffic accidents
  - Provide actionable insights for preventative measures

Risk Mapping

- Focus on analyzing historical traffic data to identify accident trends, high-risk zones, and severity patterns
- Visualization of accident

Predictive Modeling for Severity

- Predict accident severity to proactively prevent accidents
- Reduce fatalities, injuries, and material damage

Data-Driven Policy Making

- Empower Montreal traffic authorities with insights to inform policies on road safety and urban planning
- Recommendations for road maintenance prioritization or public safety campaigns

Key Stakeholders

01

Montreal City Traffic Authorities (le Service des infrastructures du réseau routier)

- Interest: High — Focused on improving traffic management and road safety.
- Power: High — Authority to implement traffic and infrastructure changes.

02

Montreal Municipal Government (City of Montreal)

- Interest: High — Committed to enhancing public safety and urban planning.
- Power: High — Controls budget, policy, and infrastructure approvals.

03

Montreal Police Department (SPVM)

- Interest: High — Concerned with enforcing laws and responding to accidents.
- Power: Medium — Influence over operations but limited in infrastructure decisions.

Key Stakeholders

04

Montreal Public Safety Departments (Fire and Ambulance Services)

- Interest: High — Directly involved in accident response and resource allocation.
- Power: Medium — Can influence deployment but not overall traffic planning.

05

Insurance Company of Canada

- Interest: Medium — Uses data for risk assessment and premium adjustments.
- Power: Low — Minimal influence over public infrastructure.

06

Montreal General Public (Drivers, Pedestrians, Cyclists)

- Interest: High — Direct beneficiaries of improved road safety.
- Power: Low — Limited decision-making power but can influence public opinion.

Exploratory Data Analysis

Data Description

The dataset focuses on traffic accidents in Montreal, containing variables that track accident location, severity, and other contextual factors like weather and road conditions. The dataset comes from official sources **SAAQ (Société de l'assurance automobile du Québec)**, and it includes geographic coordinates, quality, and precision scores for accident locations on Montreal's road network. The methodology and unit of analysis are specificed in the data dictionary attached in the appendix.

shape	(218272, 68)
Categorical variables	33 (object, int)
Numerical variables	35 (int, float)

Dataset Overview

Accident Basic Info
Seq_Num
Year
Acc_Date
Acc_Day
Acc_Time

Accident Details
Acc_Type
Part_Sit
Severity

Geolocation Quality
Loc_Quality
Loc_Accuracy
Loc_Imprecise

Road conditions & Infrastructure
Acc_Type
Surface_Cond
Light_Cond
Environ_Type
Road_Cat
Road_Cond
Road_Aspect
Loc_Code
Pos_Code
Road_Config
Work_Zone
Weather_Cond
Speed_Limit

shape	(218272, 68)
Categorical variables	33 (object, int)
Numerical variables	35 (int, float)

Casualty Counts
Num_Death
Num_Serious_Inj
Num_Minor_Inj
Total_Victims

Vehicle Counts
Num_Veh_Invld
Num_Light_Veh
Num_Heavy_Truck
Num_Equip
Num_Bus
Num_Bike
Num_Moped
Num_Moto
Num_Taxi
Num_Emerg
Num_Snowmobile
Num_OffRoad
Num_Other_Veh
Num_Unspec_Veh

Location Details
Municipality_Code
Street_Num
Street_Num_Suffix
Km_Mark
Distance_Meters
Street
Route_Num
Road_Code
Landmark_Code
Landmark_Type
Near_To
Admin_Region
County
LOC_X
LOC_Y
Lontitude
Latitude
Loc_Detached

Pedestrian
Num_Ped_Death
Num_Ped_Inj
Num_Ped_Vic

Motorcycle
Num_Moto_Death
Num_Moto_Inj
Num_Moto_Vic

Bicycle
Num_Bike_Death
Num_Bike_Inj
Num_Bike_Vic

Dataset Overview

Accident Basic Info

- Seq_Num
- Year
- Acc_Date
- Acc_Day
- Acc_Time

Accident Details

- Acc_Type
- Part_Sit
- Severity

Geolocation Quality

- Loc_Quality
- Loc_Accuracy
- Loc_Imprecise

Road conditions & Infrastructure

- Acc_Type
- Surface_Cond
- Light_Cond
- Environ_Type
- Road_Cat
- Road_Cond
- Road_Aspect
- Loc_Code
- Pos_Code
- Road_Config
- Work_Zone
- Weather_Cond
- Speed_Limit

shape	(218272, 68)
Categorical variables	33 (object, int)
Numerical variables	35 (int, float)

Casualty Counts

- Num_Death
- Num_Serious_Inj
- Num_Minor_Inj
- Total_Victims

Vehicle Counts

- Num_Veh_Invld
- Num_Light_Veh
- Num_Heavy_Truck
- Num_Equip
- Num_Bus
- Num_Bike
- Num_Moped
- Num_Moto
- Num_Taxi
- Num_Emerg
- Num_Snowmobile
- Num_OffRoad
- Num_Other_Veh
- Num_Unspec_Veh

Location Details

- Municipality_Code
- Street_Num
- Street_Num_Suffix
- Km_Mark
- Distance_Meters
- Street
- Route_Num
- Road_Code
- Landmark_Code
- Landmark_Type
- Near_To
- Admin_Region
- County
- LOC_X
- LOC_Y
- Lontitude
- Latitude
- Loc_Detached

Pedestrian

- Num_Ped_Death
- Num_Ped_Inj
- Num_Ped_Vic

Motorcycle

- Num_Moto_Death
- Num_Moto_Inj
- Num_Moto_Vic

Bicycle

- Num_Bike_Death
- Num_Bike_Inj
- Num_Bike_Vic

Dataset Overview

Accident Basic Info

- Seq_Num
- Acc_Date
- Acc_Time

Accident Details

- Acc_Type
- Part_Sit
- **Severity**

Geolocation Quality

- Loc_Quality
- Loc_Accuracy
- Loc_Imprecise

Road conditions & Infrastructure

- Acc_Type
- Surface_Cond
- Light_Cond
- Environ_Type
- Road_Cat
- Road_Cond
- Road_Aspect
- Loc_Code
- Pos_Code
- Road_Config
- Work_Zone
- Weather_Cond
- Speed_Limit

shape	(218128, 59)
Categorical variables	27 (object, int)
Numerical variables	32 (int, float)

Casualty Counts

- Num_Death
- Num_Serious_Inj
- Num_Minor_Inj
- Total_Victims

Vehicle Counts

- Num_Veh_Invld
- Num_Light_Veh
- Num_Heavy_Truck
- Num_Equip
- Num_Bus
- Num_Bike
- Num_Moped
- Num_Moto
- Num_Taxi
- Num_Emerg
- Num_Snowmobile
- Num_OffRoad
- Num_Other_Veh
- Num_Unspec_Veh

Location Details

- Street_Num
- Street_Num_Suffix
- Km_Mark
- Distance_Meters
- Street
- Route_Num
- Road_Code
- Landmark_Code
- Near_To
- Lontitude
- Latitude

Pedestrian

- Num_Ped_Death
- Num_Ped_Inj
- Num_Ped_Vic

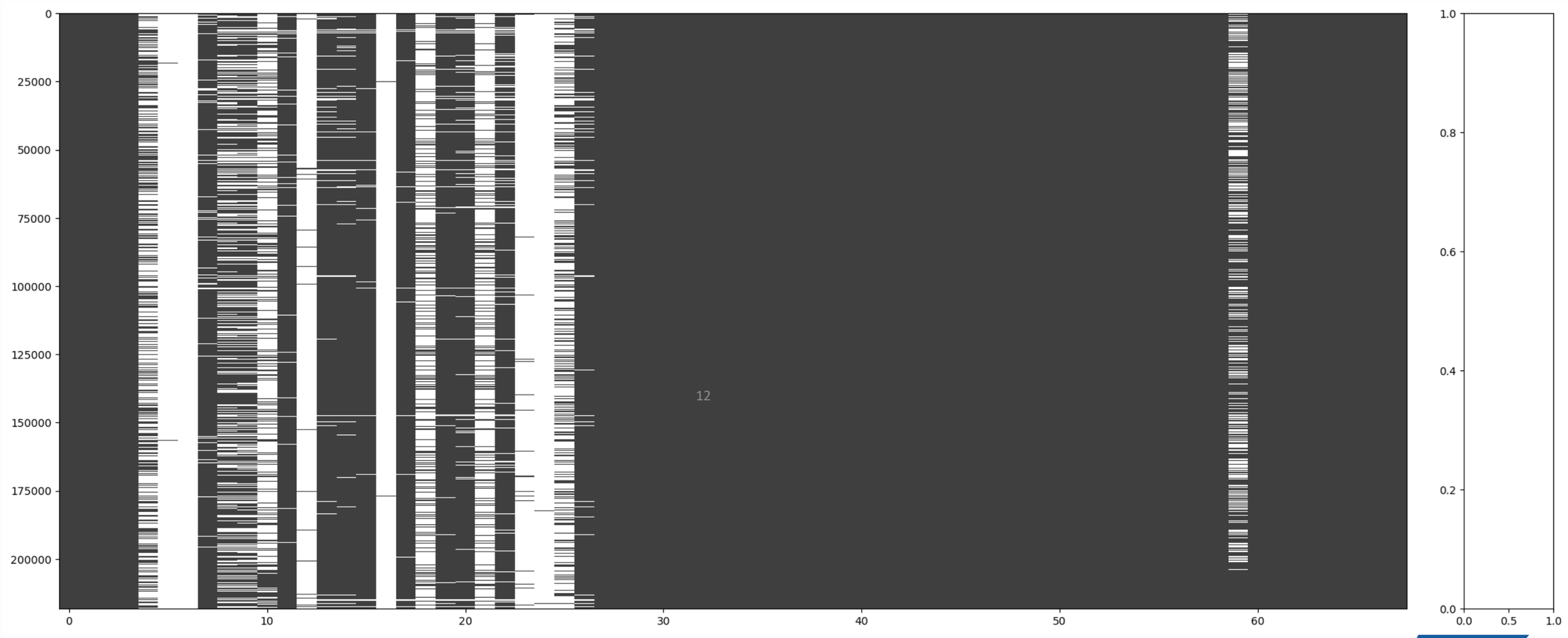
Motorcycle

- Num_Moto_Death
- Num_Moto_Inj
- Num_Moto_Vic

Bicycle

- Num_Bike_Death
- Num_Bike_Inj
- Num_Bike_Vic

Dataset Overview



Dataset Overview

Accident Basic Info

- Seq_Num
- Acc_Date
- Acc_Time

Accident Details

- Acc_Type
- **Severity**

Geolocation Quality

- Loc_Quality
- Loc_Accuracy
- Loc_Imprecise

Road conditions & Infrastructure

- Acc_Type
- Surface_Cond
- Light_Cond
- Environ_Type
- Road_Cat
- Road_Aspect
- Loc_Code
- Road_Config
- Weather_Cond
- Speed_Limit

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Casualty Counts

- Num_Death
- Num_Serious_Inj
- Num_Minor_Inj
- Total_Victims

Vehicle Counts

- Num_Veh_Invld
- Num_Light_Veh
- Num_Heavy_Truck
- Num_Equip
- Num_Bus
- Num_Bike
- Num_Moped
- Num_Moto
- Num_Taxi
- Num_Emerg
- Num_Snowmobile
- Num_OffRoad
- Num_Other_Veh
- Num_Unspec_Veh

Location Details

- Street
- Near_To
- Lontitude
- Latitude

Pedestrian

- Num_Ped_Death
- Num_Ped_Inj
- Num_Ped_Vic

Motorcycle

- Num_Moto_Death
- Num_Moto_Inj
- Num_Moto_Vic

Bicycle

- Num_Bike_Death
- Num_Bike_Inj
- Num_Bike_Vic

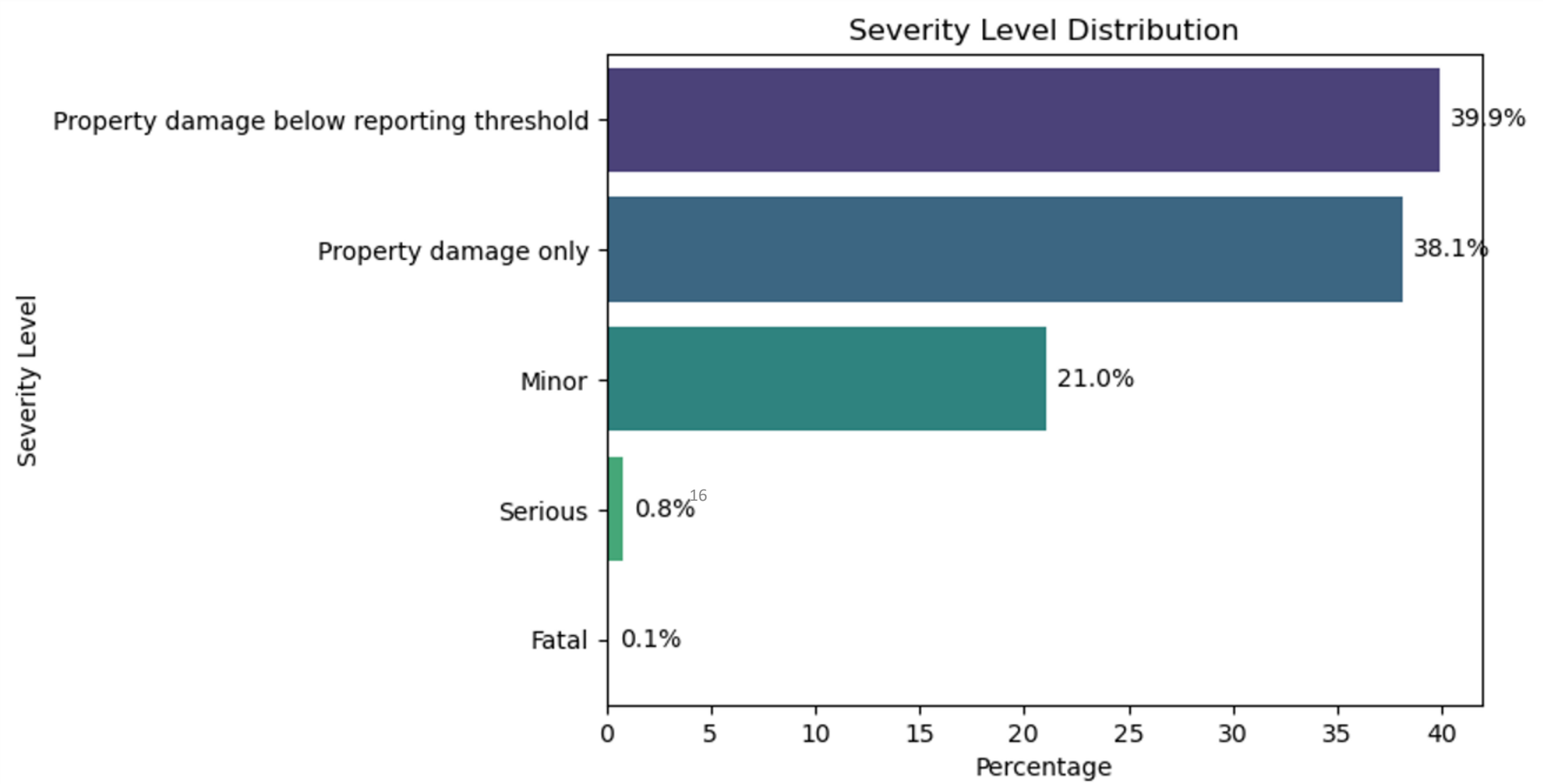
Missing value handling

	feature	NA_count	missing %
0	Street	12264	5.62
1	Near_To	71031	32.56
2	Acc_Type	10054	4.61
3	Surface_Cond	12746	5.84
4	Light_Cond	12909	5.92
5	Environ_Type	7044	3.23
6	Road_Cat	6344	2.91
7	Road_Aspect	9903	4.54
8	Loc_Code	17743	8.13
9	Pos_Code	168943	77.45
10	Road_Config	21947	10.06
11	Work_Zone	213230	97.75
12	Weather_Cond	13589	6.23
13	Speed_Limit	80812	37.05
14	Lontitude	7	0.00
15	Latitude	7	0.00

Rolling window imputation

Seq_Num	Acc_Day	Acc_Date	Acc_Time
Sequential number identifying the accident. Composed of the year of the accident and a sequential number. (YYYY _999, where year and sequential number are separated by a space, an underscore and a space)	LU = Monday MA = Tuesday ME = Wednesday JE = Thursday VE = Friday SA = Saturday DI = Sunday		
SPVM _ 2012 _ 10020	ME	2012/05/09	20:00:00-20:59:00
SPVM _ 2012 _ 10021	ME	2012/05/09	Non précisé
SPVM _ 2012 _ 10022	VE	2012/05/11	17:00:00-17:59:00
SPVM _ 2012 _ 10023	VE	2012/05/11	17:00:00-17:59:00
SPVM _ 2012 _ 10024	VE	2012/05/11	19:00:00-19:59:00
SPVM _ 2012 _ 10025	SA	2012/05/12	20:00:00-20:59:00
SPVM _ 2012 _ 10026	LU	2012/05/14	21:00:00-21:59:00

Target variable - Imbalanced Distribution



Accident Basic Info

- Seq_Num
- Acc_Date
- Acc_Time

Accident Details

- Acc_Type
- Severity

Road conditions & Infrastructure

- Acc_Type
- Surface_Cond
- Light_Cond
- Environ_Type
- Road_Cat
- Road_Aspect
- Loc_Code
- Road_Config
- Weather_Cond
- Speed_Limit

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Casualty Counts

- Num_Death
- Num_Serious_Inj
- Num_Minor_Inj
- Total_Victims

Vehicle Counts

- Num_Veh_Invld
- Num_Light_Veh
- Num_Heavy_Truck
- Num_Equip
- Num_Bus
- Num_Bike
- Num_Moped
- Num_Moto
- Num_Taxi
- Num_Emerg
- Num_Snowmobile
- Num_OffRoad
- Num_Other_Veh
- Num_Unspec_Veh

Location Details

- Street
- Near_To
- Lontitude
- Latitude

Geolocation Quality

- Loc_Quality
- Loc_Accuracy
- Loc_Imprecise

Pedestrian

- Num_Ped_Death
- Num_Ped_Inj
- Num_Ped_Vic

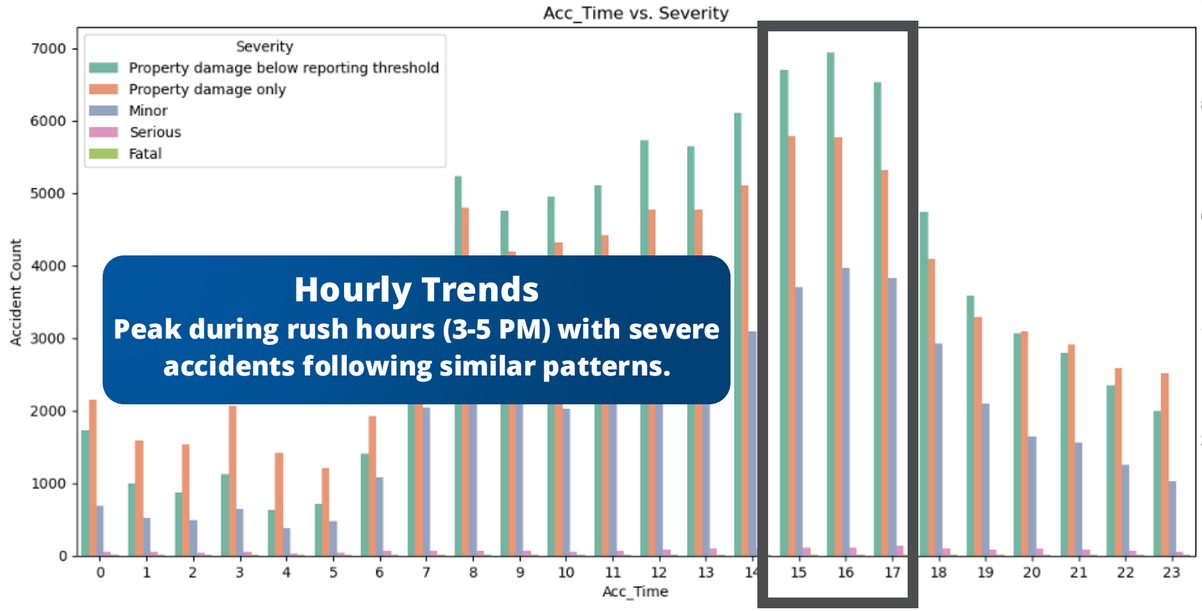
Motorcycle

- Num_Moto_Death
- Num_Moto_Inj
- Num_Moto_Vic

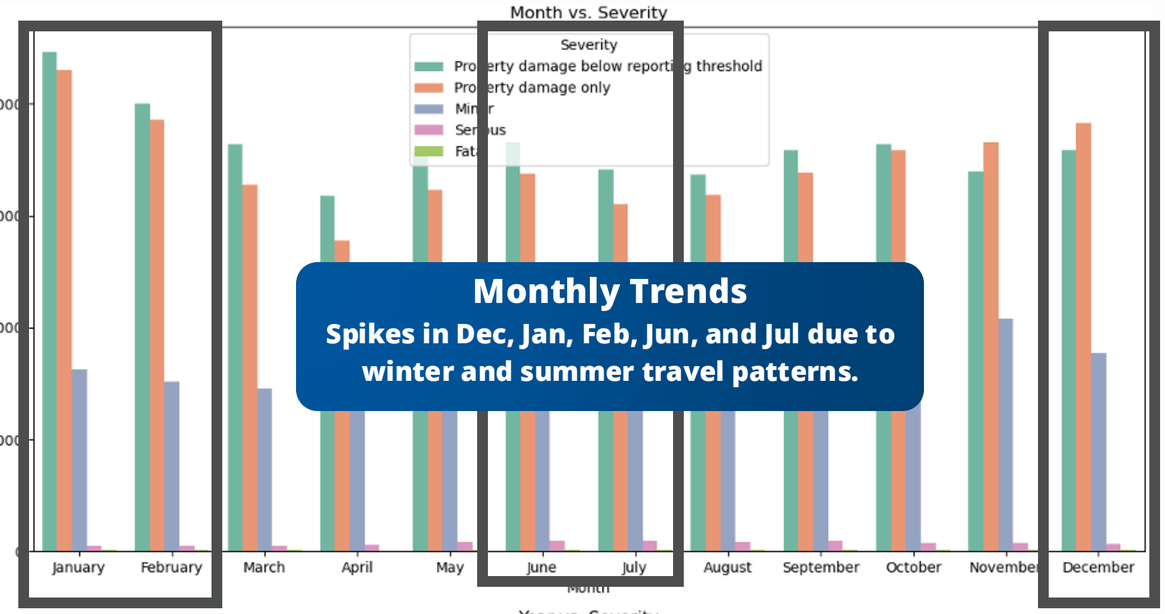
Bicycle

- Num_Bike_Death
- Num_Bike_Inj
- Num_Bike_Vic

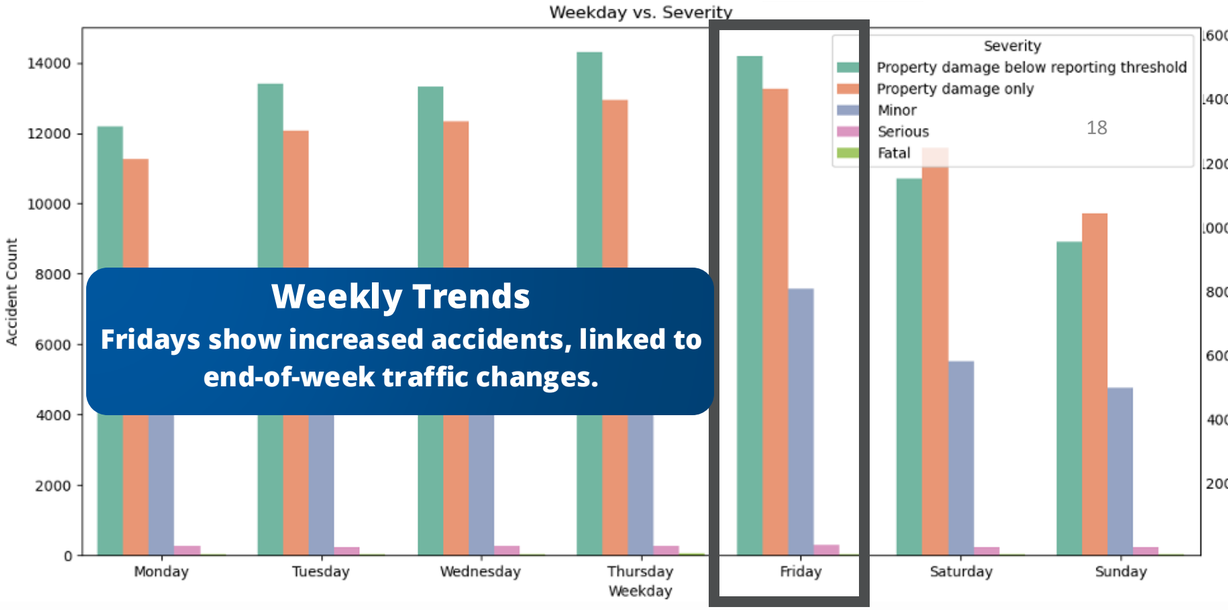
Collision patterns by different timeframes



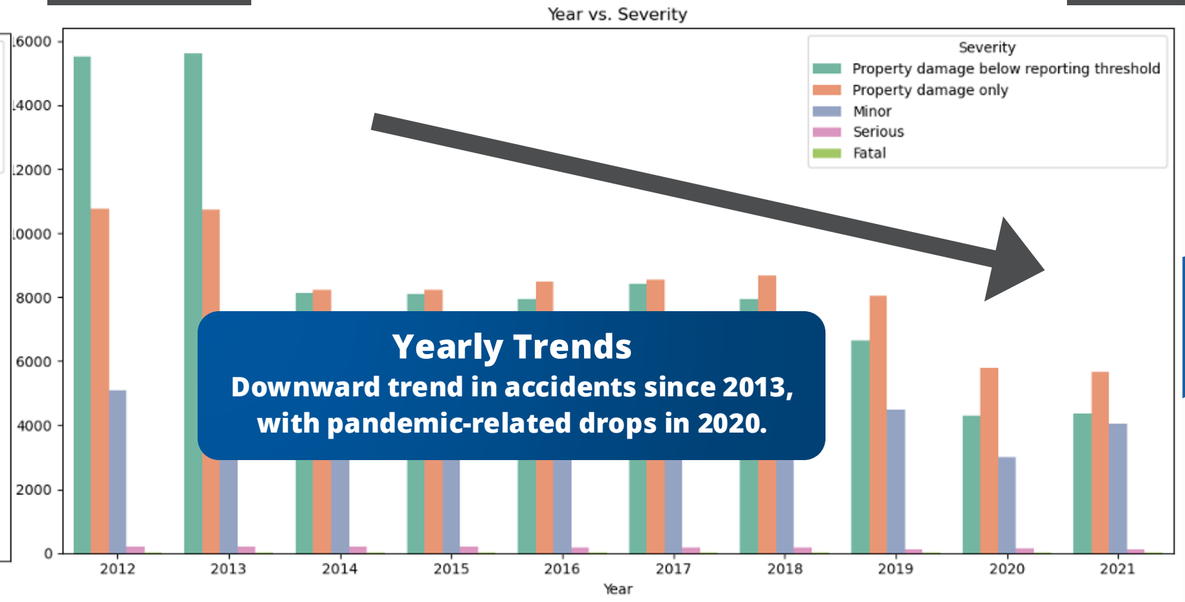
Hourly Trends
Peak during rush hours (3-5 PM) with severe accidents following similar patterns.



Monthly Trends
Spikes in Dec, Jan, Feb, Jun, and Jul due to winter and summer travel patterns.



Weekly Trends
Fridays show increased accidents, linked to end-of-week traffic changes.



Yearly Trends
Downward trend in accidents since 2013, with pandemic-related drops in 2020.

Accident Basic Info
Seq_Num
Acc_Date
Acc_Time

Accident Details
Acc_Type
Severity

Road conditions & Infrastructure
Acc_Type
Surface_Cond
Light_Cond
Environ_Type
Road_Cat
Road_Aspect
Loc_Code
Road_Config
Weather_Cond
Speed_Limit

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Vehicle Counts
Num_Veh_Invld
Num_Light_Veh
Num_Heavy_Truck
Num_Equip
Num_Bus
Num_Bike
Num_Moped
Num_Moto
Num_Taxi
Num_Emerg
Num_Snowmobile
Num_OffRoad
Num_Other_Veh
Num_Unspec_Veh

Location Details
Street
Near_To
Lontitude
Latitude

Geolocation Quality
Loc_Quality
Loc_Accuracy
Loc_Imprecise

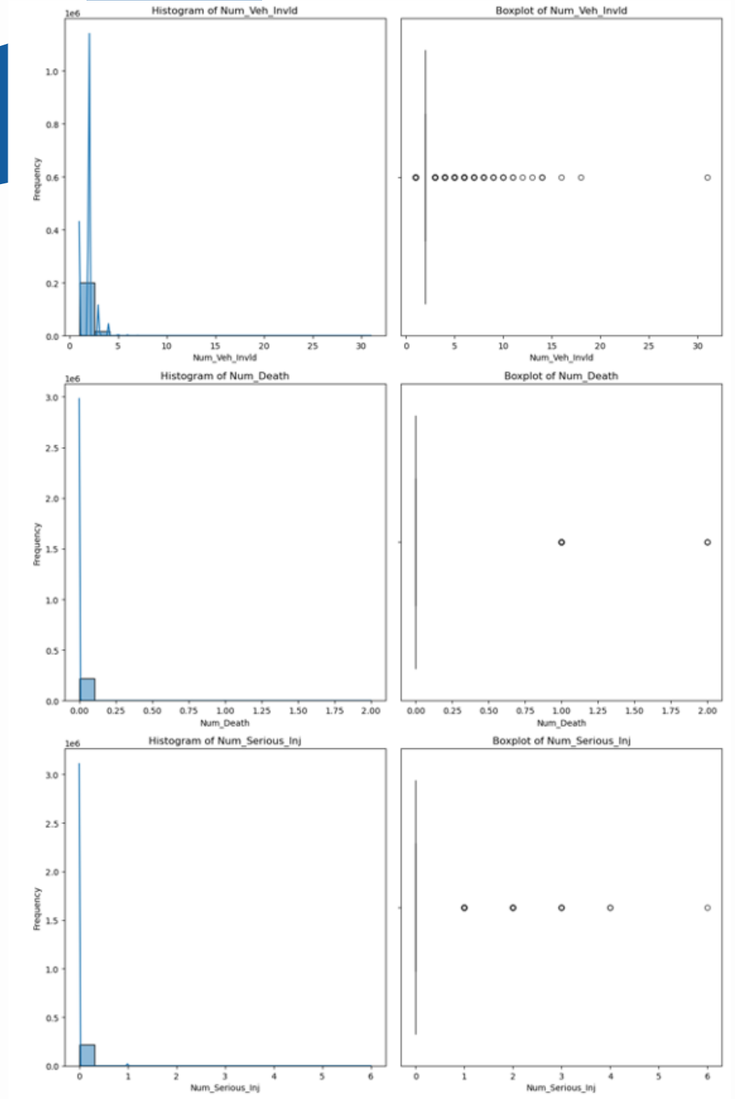
Casualty Counts
Num_Death
Num_Serious_Inj
Num_Minor_Inj
Total_Victims

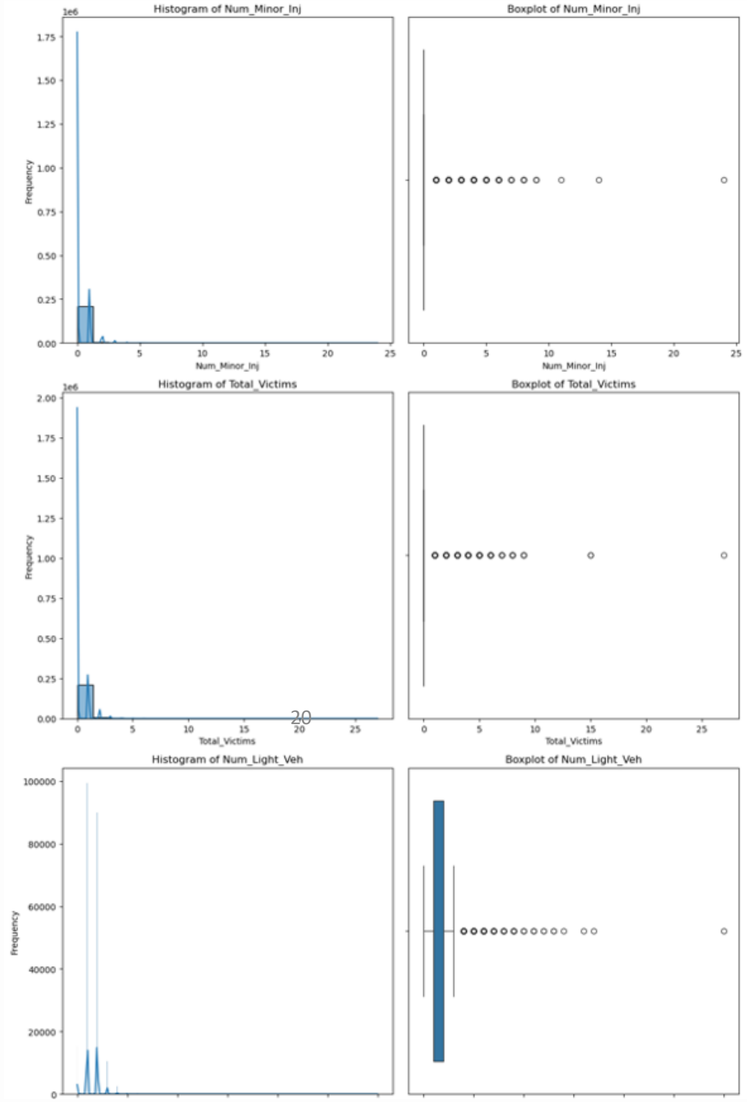
Pedestrian
Num_Ped_Death
Num_Ped_Inj
Num_Ped_Vic

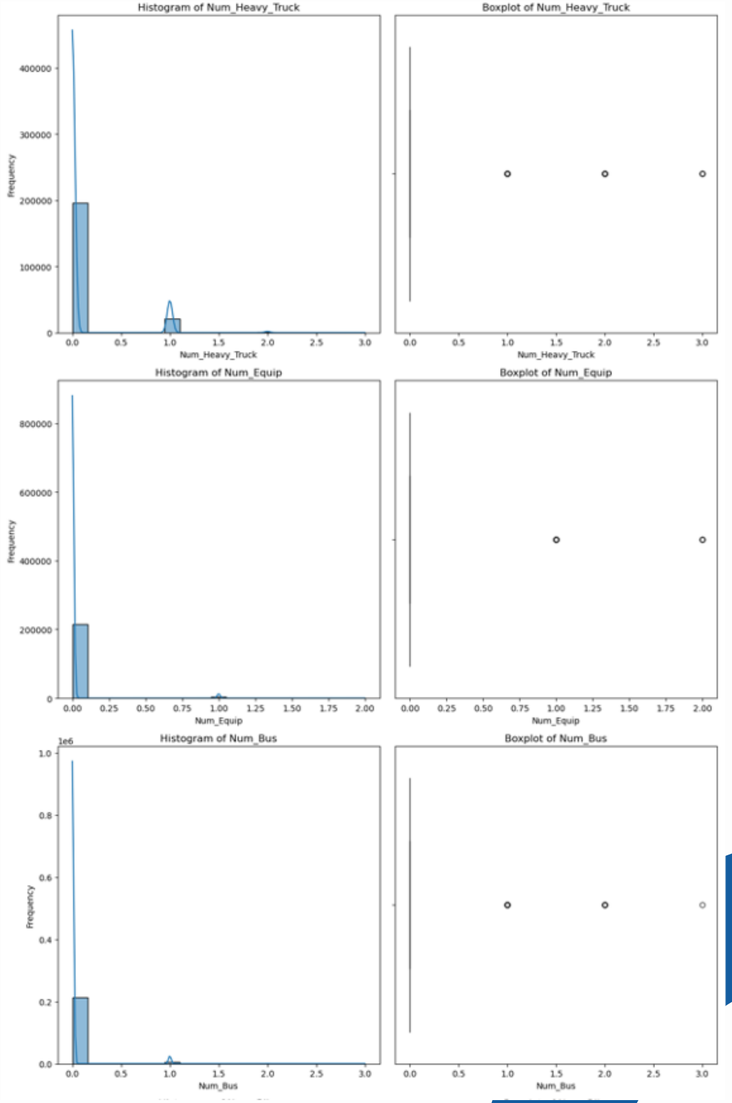
Motorcycle
Num_Moto_Death
Num_Moto_Inj
Num_Moto_Vic

Bicycle
Num_Bike_Death
Num_Bike_Inj
Num_Bike_Vic

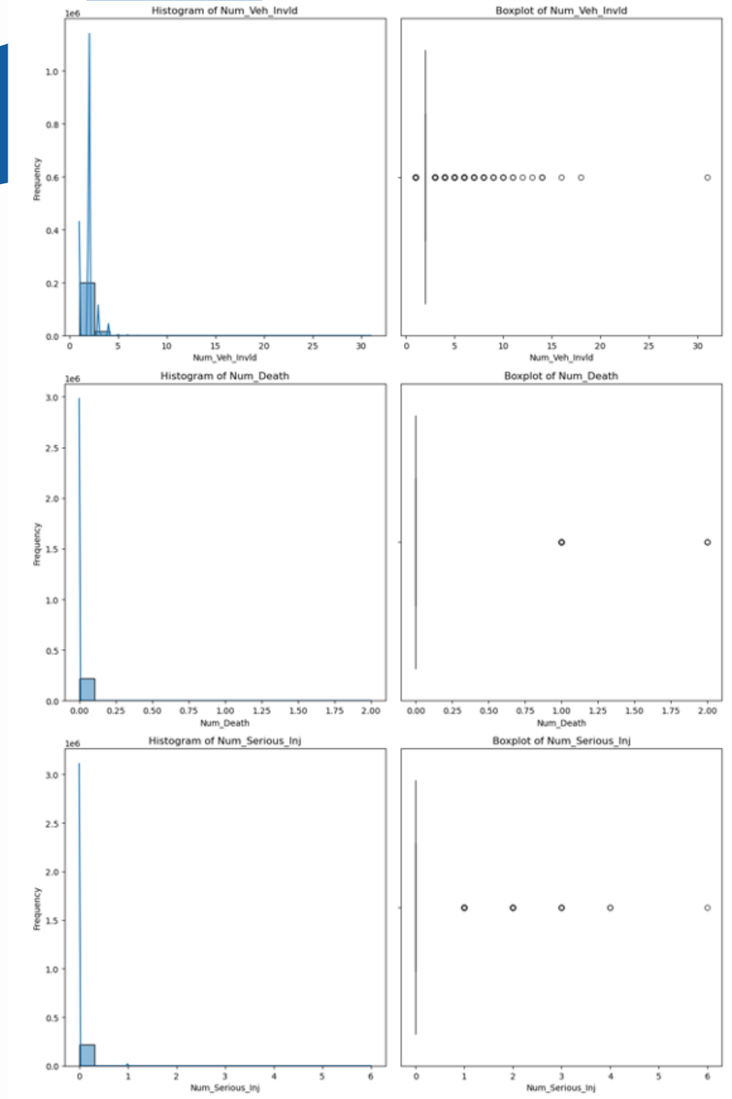
High zero percentage in frequency variables



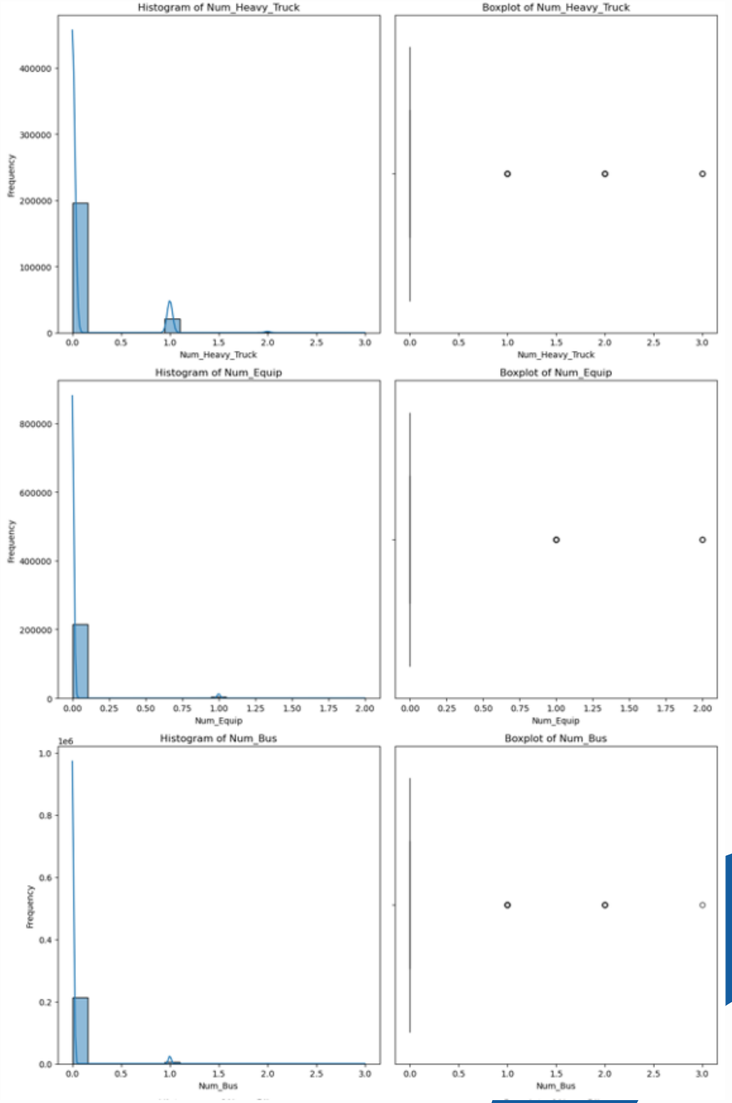




High zero percentage in frequency variables



	Column	Zero_Percentage
0	Num_Veh_Invld	0.000000
1	Num_Death	99.879429
2	Num_Serious_Inj	99.172504
3	Num_Minor_Inj	78.815191
4	Total_Victims	78.020245
5	Num_Light_Veh	7.038986
6	Num_Heavy_Truck	90.037501
7	Num_Equip	98.559103
8	Num_Bus	97.447370
9	Num_Bike	96.020227
10	Num_Moped	99.393934
11	Num_Moto	98.648958
12	Num_Taxi	96.796376
13	Num_Emerg	96.845889
14	Num_Snowmobile	99.997249
15	Num_OffRoad	99.991290
16	Num_Other_Veh	99.452615
17	Num_Unspec_Veh	81.250458
18	Num_Ped_Death	99.928482
19	Num_Ped_Inj	95.154221
20	Num_Ped_Vic	95.086371
21	Num_Moto_Death	99.992206
22	Num_Moto_Inj	99.243563
23	Num_Moto_Vic	99.235770
24	Num_Bike_Death	99.985788
25	Num_Bike_Inj	97.012763
26	Num_Bike_Vic	96.998551



Dummify it!

Accident Basic Info
Seq_Num
Acc_Date
Acc_Time

Accident Details
Acc_Type
Severity

Road conditions & Infrastructure
Acc_Type
Surface_Cond
Light_Cond
Environ_Type
Road_Cat
Road_Aspect
Loc_Code
Road_Config
Weather_Cond
Speed_Limit

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Vehicle Counts
Num_Veh_Invld
Num_Light_Veh
Num_Heavy_Truck
Num_Equip
Num_Bus
Num_Bike
Num_Moped
Num_Moto
Num_Taxi
Num_Emerg
Num_Snowmobile
Num_OffRoad
Num_Other_Veh
Num_Unspec_Veh

Location Details
Street
Near_To
Lontitude
Latitude

Geolocation Quality
Loc_Quality
Loc_Accuracy
Loc_Imprecise

Casualty Counts
Num_Death
Num_Serious_Inj
Num_Minor_Inj
Total_Victims

Pedestrian
Num_Ped_Death
Num_Ped_Inj
Num_Ped_Vic

Motorcycle
Num_Moto_Death
Num_Moto_Inj
Num_Moto_Vic

Bicycle
Num_Bike_Death
Num_Bike_Inj
Num_Bike_Vic

Dummify it!

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Accident Basic Info
- Seq_Num
- Acc_Date
- Acc_Time

Accident Details
- Acc_Type
- Severity

Road conditions & Infrastructure
- Acc_Type
- Surface_Cond
- Light_Cond
- Environ_Type
- Road_Cat
- Road_Aspect
- Loc_Code
- Road_Config
- Weather_Cond
- Speed_Limit

Casualty Counts
- Num_Death
- Num_Serious_Inj
- Num_Minor_Inj
- Total_Victims

Vehicle Counts
- Num_Veh_Invld
- Num_Light_Veh
- Num_Heavy_Truck
- Num_Equip
- Num_Bus
- Num_Bike
- Num_Moped
- Num_Moto
- Num_Taxi
- Num_Emerg
- Num_Snowmobile
- Num_OffRoad
- Num_Other_Veh
- Num_Unspec_Veh

Location Details
- Street
- Near_To
- Lontitude
- Latitude

Geolocation Quality
- Loc_Quality
- Loc_Accuracy
- Loc_Imprecise

Pedestrian
- Num_Ped_Death
- Num_Ped_Inj
- Num_Ped_Vic

Motorcycle
- Num_Moto_Death
- Num_Moto_Inj
- Num_Moto_Vic

Bicycle
- Num_Bike_Death
- Num_Bike_Inj
- Num_Bike_Vic

Dummify it!

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Accident Basic Info
- Seq_Num
- Acc_Date
- Acc_Time

Accident Details
- Acc_Type
- Severity

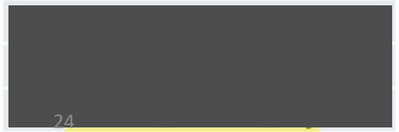
Road conditions & Infrastructure
- Acc_Type
- Surface_Cond
- Light_Cond
- Environ_Type
- Road_Cat
- Road_Aspect
- Loc_Code
- Road_Config
- Weather_Cond
- Speed_Limit

Vehicle Counts
- Num_Veh_Invld
- Num_Light_Veh
- Num_Heavy_Truck
- Num_Equip
- Num_Bus
- Num_Bike
- Num_Moped
- Num_Moto
- Num_Taxi
- Num_Emerg
- Num_Snowmobile
- Num_OffRoad
- Num_Other_Veh
- Num_Unspec_Veh

Location Details
- Street
- Near_To
- Lontitude
- Latitude

Geolocation Quality
- Loc_Quality
- Loc_Accuracy
- Loc_Imprecise

Casualty Counts



- Total_Victims

Pedestrian
- Num_Ped_Death
- Num_Ped_Inj
- Num_Ped_Vic

Motorcycle
- Num_Moto_Death
- Num_Moto_Inj
- Num_Moto_Vic

Bicycle
- Num_Bike_Death
- Num_Bike_Inj
- Num_Bike_Vic

Dummify it!

Accident Basic Info
Seq_Num
Acc_Date
Acc_Time

Accident Details
Acc_Type
Severity

Road conditions & Infrastructure
Acc_Type
Surface_Cond
Light_Cond
Environ_Type
Road_Cat
Road_Aspect
Loc_Code
Road_Config
Weather_Cond
Speed_Limit

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Vehicle Counts
Num_Veh_Invld
Num_Light_Veh
Num_Heavy_Truck
Num_Equip
Num_Bus
Num_Bike
Num_Moped
Num_Moto
Num_Taxi
Num_Emerg
Num_Snowmobile
Num_OffRoad
Num_Other_Veh
Num_Unspec_Veh

Location Details
Street
Near_To
Lontitude
Latitude

Geolocation Quality
Loc_Quality
Loc_Accuracy
Loc_Imprecise

Casualty Counts
Total_Victims

Pedestrian
Num_Ped_Death
Num_Ped_Inj
Num_Ped_Vic

Motorcycle
Num_Moto_Death
Num_Moto_Inj
Num_Moto_Vic

Bicycle
Num_Bike_Death
Num_Bike_Inj
Num_Bike_Vic

Dummify it!

Accident Basic Info

Seq_Num
Acc_Date
Acc_Time

Accident Details

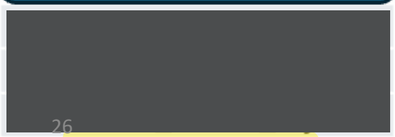
Acc_Type
Severity

Road conditions & Infrastructure

Acc_Type
Surface_Cond
Light_Cond
Environ_Type
Road_Cat
Road_Aspect
Loc_Code
Road_Config
Weather_Cond
Speed_Limit

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Casualty Counts



Total_Victims

Vehicle Counts

Num_Veh_Invld
Num_Light_Veh
Num_Heavy_Truck
Num_Equip
Num_Bus
Num_Bike
Num_Moped
Num_Moto
Num_Taxi
Num_Emerg
Num_Snowmobile
Num_OffRoad
Num_Other_Veh
Num_Unspec_Veh

Location Details

Street
Near_To
Lontitude
Latitude

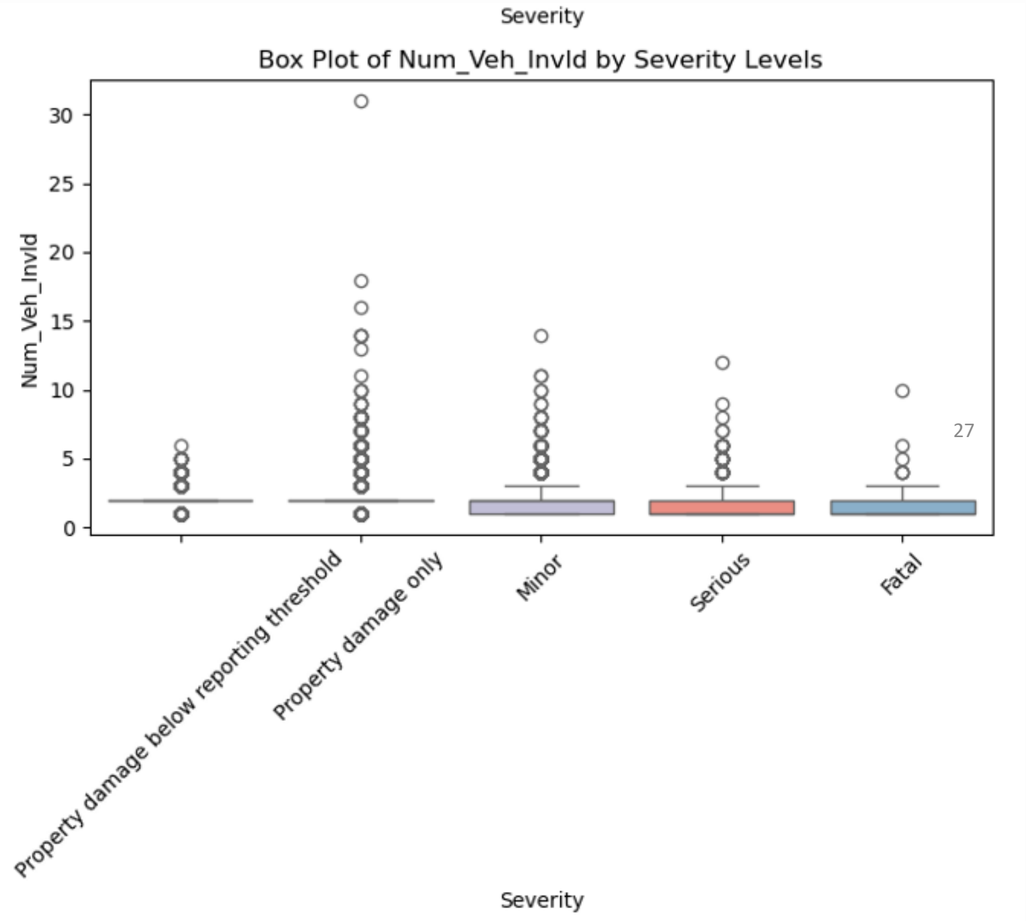
Geolocation Quality

Loc_Quality
Loc_Accuracy
Loc_Imprecise

Total Victims and Vehicles Involved

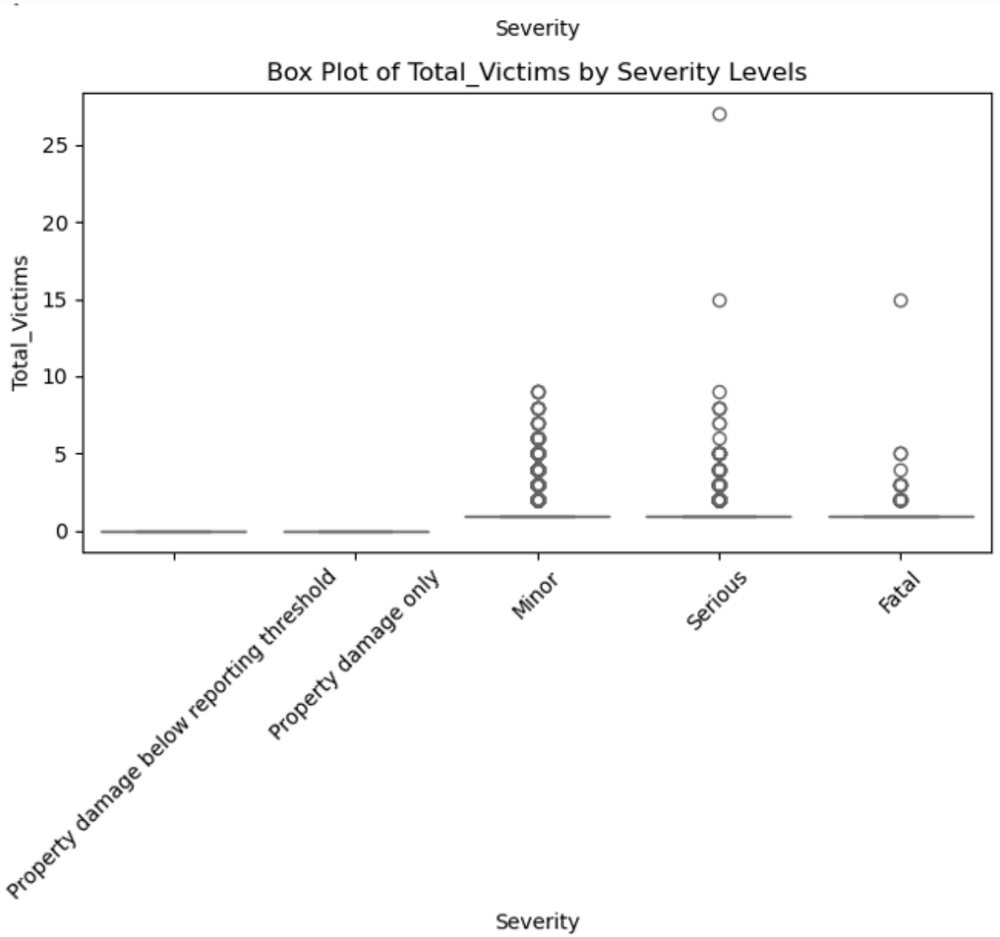
Total Victims

Higher victim counts correlate with increased severity.

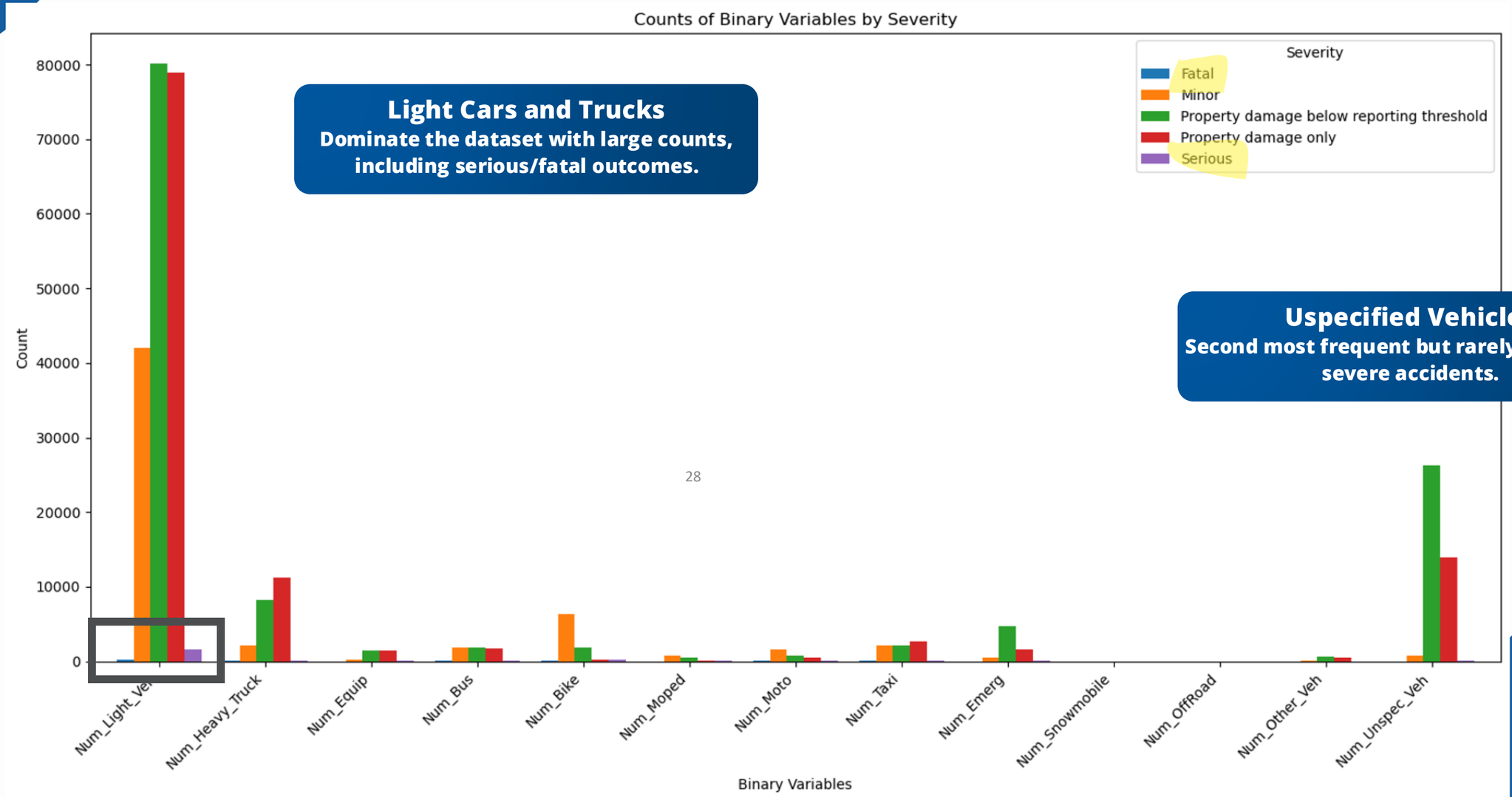


Total Vehicles Involved

While most accidents involve 1-5 vehicles, higher severity levels show more cases with multiple vehicles.



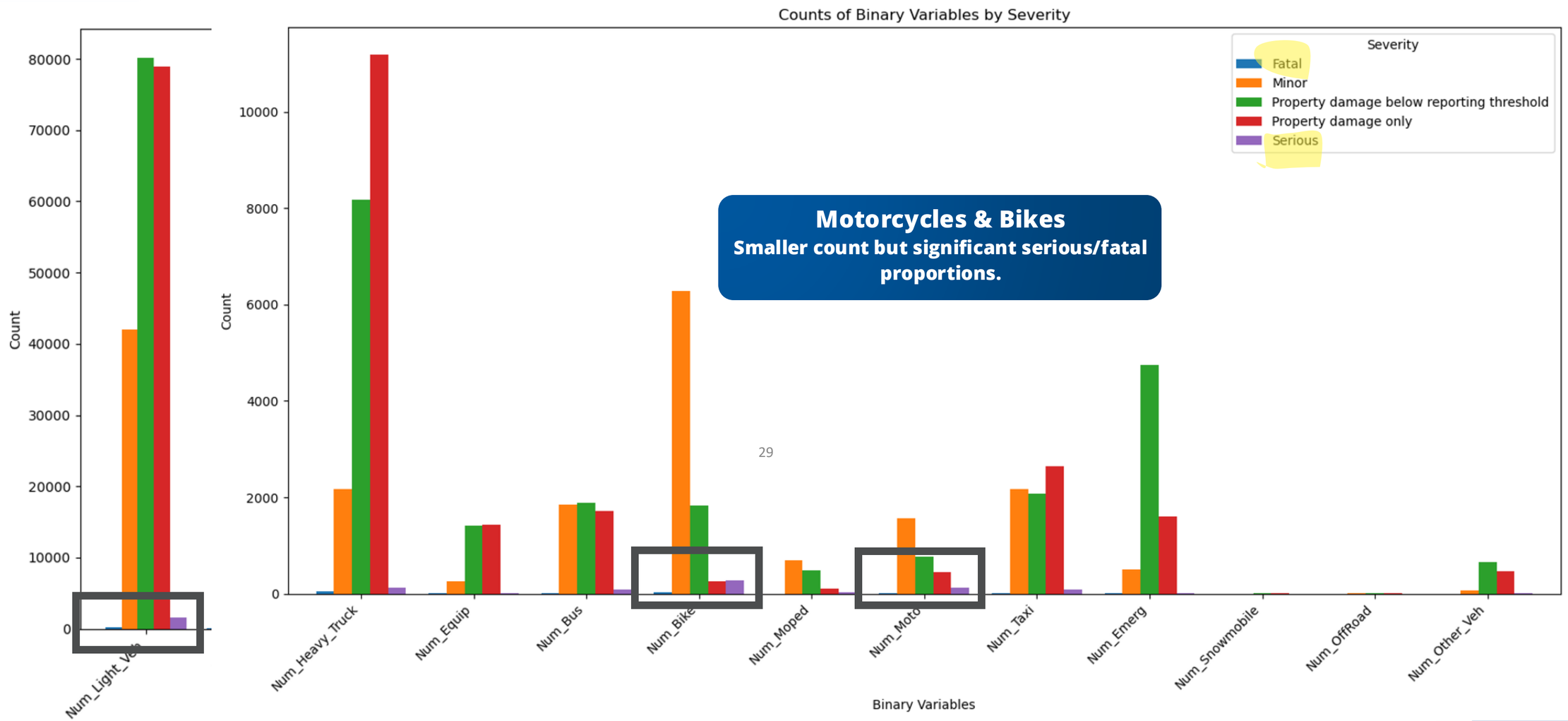
Vehicles Involvement by Severity



Light Cars and Trucks
Dominate the dataset with large counts, including serious/fatal outcomes.

Uspecified Vehicles
Second most frequent but rarely involved in severe accidents.

Vehicles Involvement by Severity



Accident Basic Info

- Seq_Num
- Acc_Date
- Acc_Time

Accident Details

- Acc_Type
- Severity

Road conditions & Infrastructure

- Acc_Type
- Surface_Cond
- Light_Cond
- Environ_Type
- Road_Cat
- Road_Aspect
- Loc_Code
- Road_Config
- Weather_Cond
- Speed_Limit

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Casualty Counts

- Total_Victims

Vehicle Counts

- Num_Veh_Invld
- Num_Light_Veh
- Num_Heavy_Truck
- Num_Equip
- Num_Bus
- Num_Bike
- Num_Moped
- Num_Moto
- Num_Taxi
- Num_Emerg
- Num_Snowmobile
- Num_OffRoad
- Num_Other_Veh
- Num_Unspec_Veh

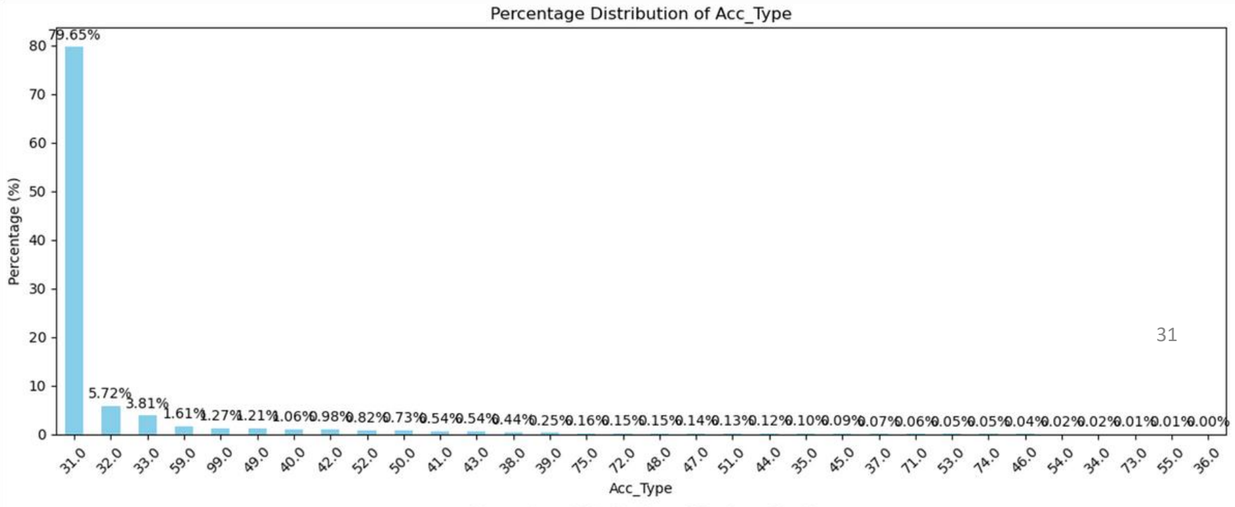
Location Details

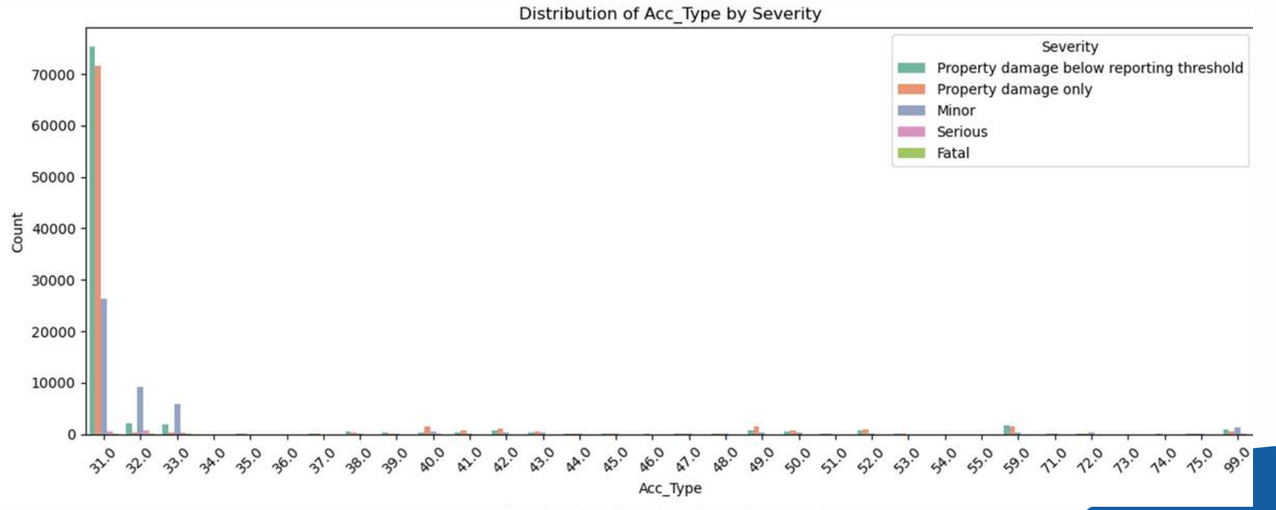
- Street
- Near_To
- Lontitude
- Latitude

Geolocation Quality

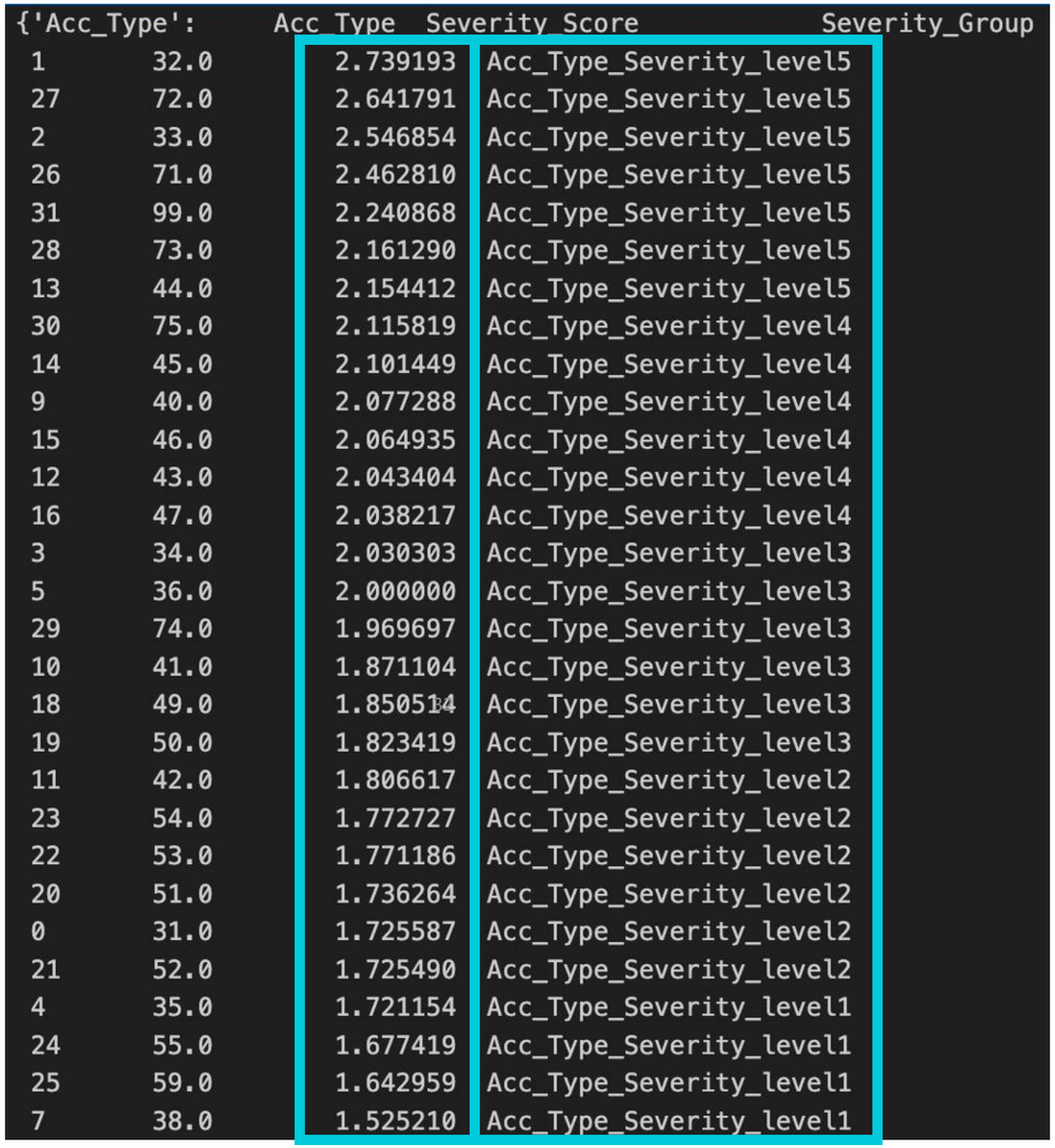
- Loc_Quality
- Loc_Accuracy
- Loc_Imprecise

High Dimension Issue





Target Encoding



{'Acc_Type':

	Acc_Type	Severity_Score	Severity_Group
1	32.0	2.739193	Acc_Type_Severity_level5
27	72.0	2.641791	Acc_Type_Severity_level5
2	33.0	2.546854	Acc_Type_Severity_level5
26	71.0	2.462810	Acc_Type_Severity_level5
31	99.0	2.240868	Acc_Type_Severity_level5
28	73.0	2.161290	Acc_Type_Severity_level5
13	44.0	2.154412	Acc_Type_Severity_level5
30	75.0	2.115819	Acc_Type_Severity_level4
14	45.0	2.101449	Acc_Type_Severity_level4
9	40.0	2.077288	Acc_Type_Severity_level4
15	46.0	2.064935	Acc_Type_Severity_level4
12	43.0	2.043404	Acc_Type_Severity_level4
16	47.0	2.038217	Acc_Type_Severity_level4
3	34.0	2.030303	Acc_Type_Severity_level3
5	36.0	2.000000	Acc_Type_Severity_level3
29	74.0	1.969697	Acc_Type_Severity_level3
10	41.0	1.871104	Acc_Type_Severity_level3
18	49.0	1.850514	Acc_Type_Severity_level3
19	50.0	1.823419	Acc_Type_Severity_level3
11	42.0	1.806617	Acc_Type_Severity_level2
23	54.0	1.772727	Acc_Type_Severity_level2
22	53.0	1.771186	Acc_Type_Severity_level2
20	51.0	1.736264	Acc_Type_Severity_level2
0	31.0	1.725587	Acc_Type_Severity_level2
21	52.0	1.725490	Acc_Type_Severity_level2
4	35.0	1.721154	Acc_Type_Severity_level1
24	55.0	1.677419	Acc_Type_Severity_level1
25	59.0	1.642959	Acc_Type_Severity_level1
7	38.0	1.525210	Acc_Type_Severity_level1

Target Encoding

Accident Type (Acc_Type)

Severity Level	Severity Score Range	Characteristics
Level 5 (Highest)	2.240 - 2.739	Pedestrian collision (32), No collision: rollover (72), Cyclist collision (33), Without collision: rollover (71), Other (99), Without collision: submersion (73), Fixed obje
Level 4	2.038 - 2.115	Without collision: leaves roadway (75), Fixed object: impact attenuator (45), Fixed object: guardrail end (46), Fixed object: floor lamp (40), Fixed object: tree (43), F
Level 3	1.823 - 2.030	Collision with train (34), Collision with moose/bear (36), Without collision: fire/explosion (74), Fixed object: support light (41), Fixed object: building/wall (49), Fixec
Level 2	1.725 - 1.806	Fixed object: utility pole (42), Fixed object: rock face (54), Fixed object: ditch (53), Fixed object: hydrant (51), Collision with road vehicle (31), Fixed object: fence/ba
Level 1 (Lowest)	1.397 - 1.721	Collision with deer (35), Fixed object: culvert (55), Fixed object:

Surface Condition (Surface_Cond)

Severity Level	Severity Score Range	Characteristics
Level 5 (Highest)	1.989 - 2.472	Water accumulation (13), Oily surfaces (20)
Level 4	1.905 - 1.920	Wet (12), Sand/gravel on road (14)
Level 3	1.834 - 1.843	Dried (11), Icy (18)
Level 2	1.684 - 1.834	Slush/melting snow (15), Muddy (19)
Level 1 (Lowest)	1.496 - 1.670	Snow-covered (16), Compacted snow (17), Other (99)

Road Configuration (Road_Config)

Severity Level	Severity Score Range	Characteristics
Level 5 (Highest)	2.025	Separated by crossable fittings (4)
Level 4	2.007	Separated by impassable structure (5)
Level 3	1.991	Two directions, multiple lanes per direction (3)
Level 2	1.818	Two directions, one lane per direction (2)
Level 1 (Lowest)	1.488 - 1.703	One way (1), Other (9)

Environment Type (Environ_Type)

Severity Level	Severity Score Range	Characteristics
Level 5 (Highest)	1.998 - 2.069	Rural (5), Forestry (6)
Level 4	1.963	Industrial/manufacturing (4)
Level 3	1.839 - 1.916	School (1), Other (7)
Level 2	1.828	Residential (2)
Level 1 (Lowest)	1.665 - 1.820	Business/commercial (3), Other (9)

(Appendix.

Severity Patterns Across Key Factors

Accident Type

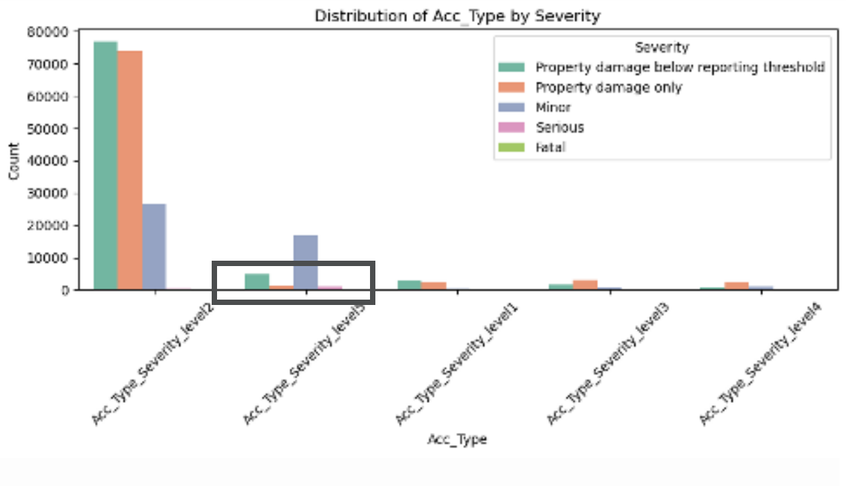
Includes pedestrian, cyclist, rollover, and fixed-object collisions (e.g., guardrails)

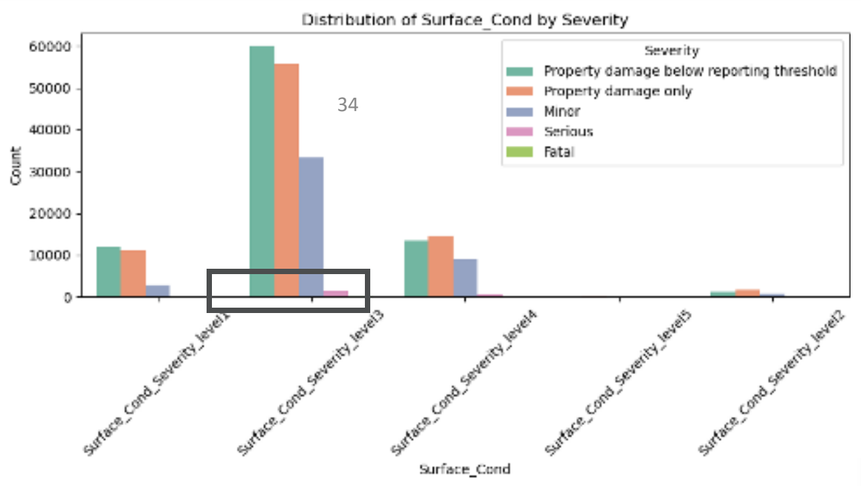
Surface Conditions

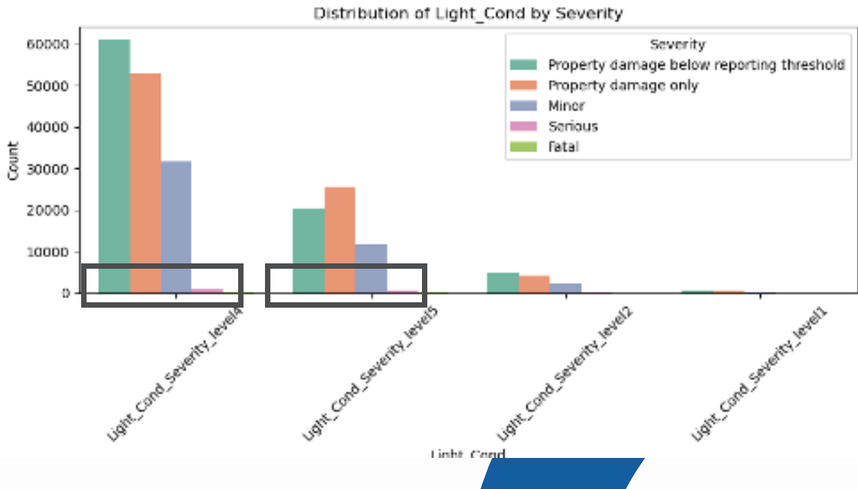
Includes icy and sand/gravel-covered roads, leading to the highest severity outcomes compared to dried or wet conditions.

Lighting Conditions

Includes nighttime with illumination, leading to more severe accidents compared to daytime or natural light.







Severity Patterns Across Key Factors

Environment Type

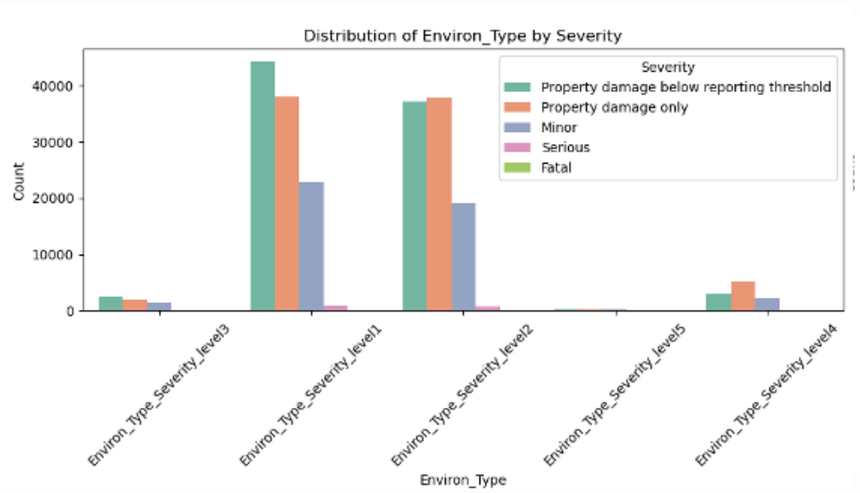
Includes business/commercial areas, residential areas, and other zones.

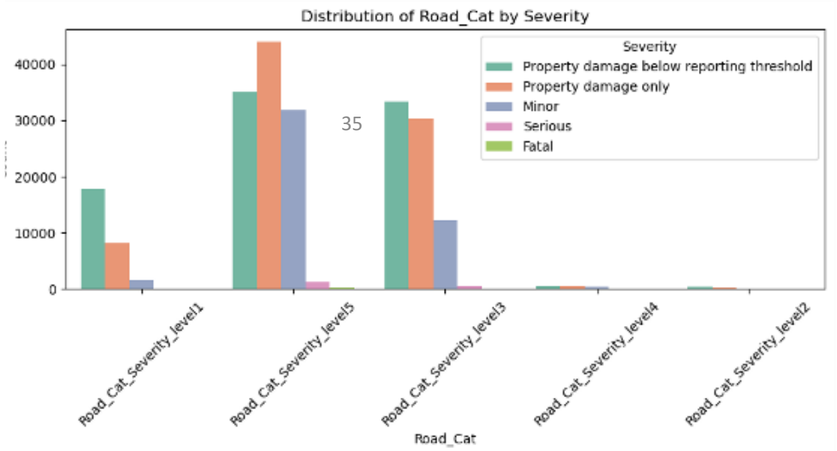
Road Category

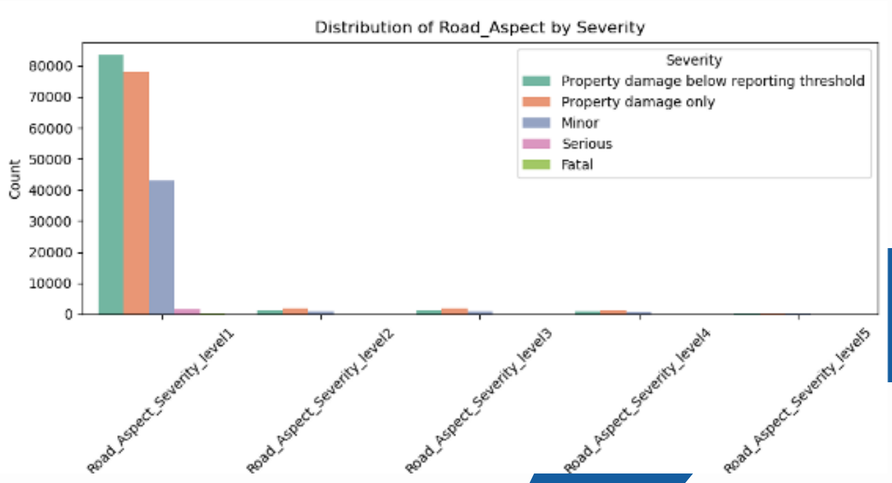
Includes main arteries, numbered roads, and highway ramps.

Road Aspect

Includes straight flat roads and straight down slopes.



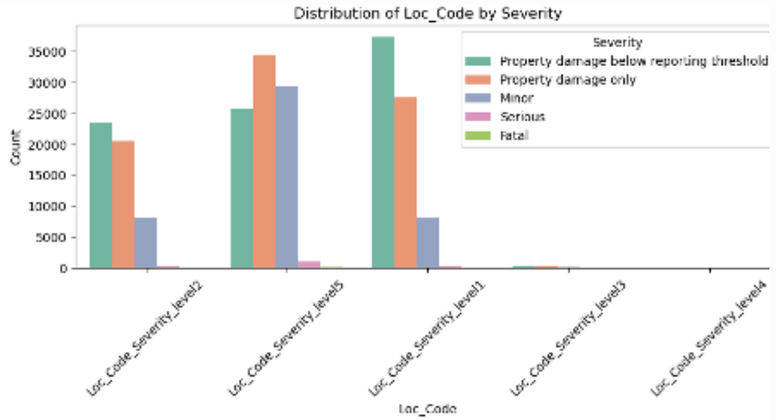




Severity Patterns Across Key Factors

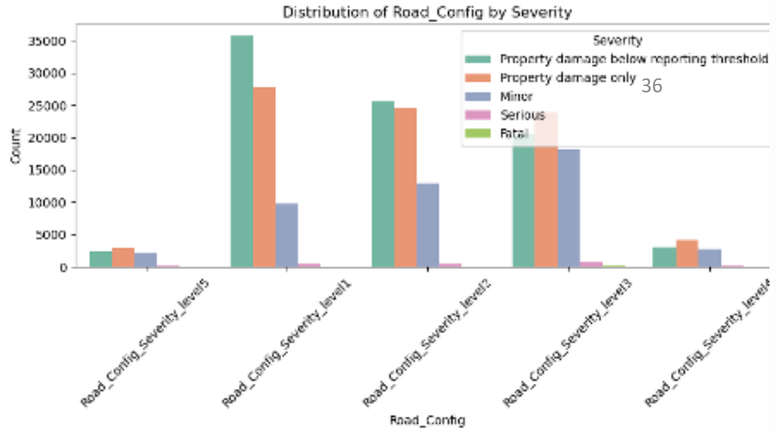
Location Code

Includes intersections, near intersections, shopping centers, and other points.



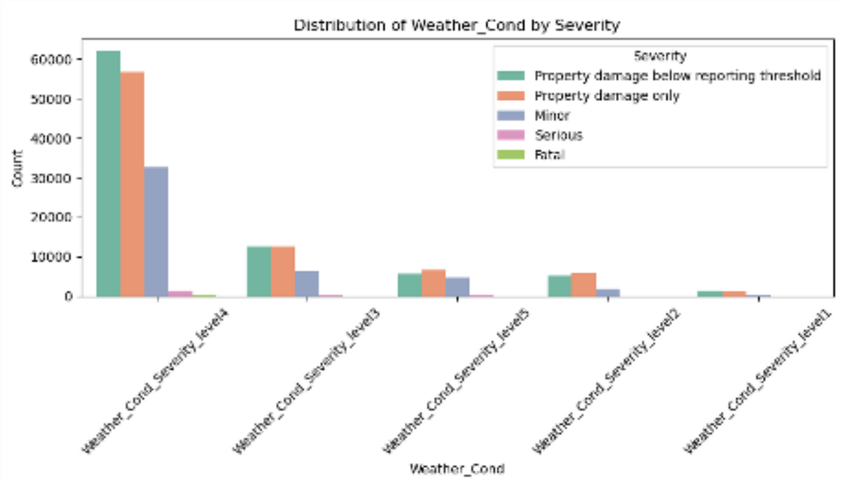
Road Configuration

Includes one-way and multi-lane two-way roads.



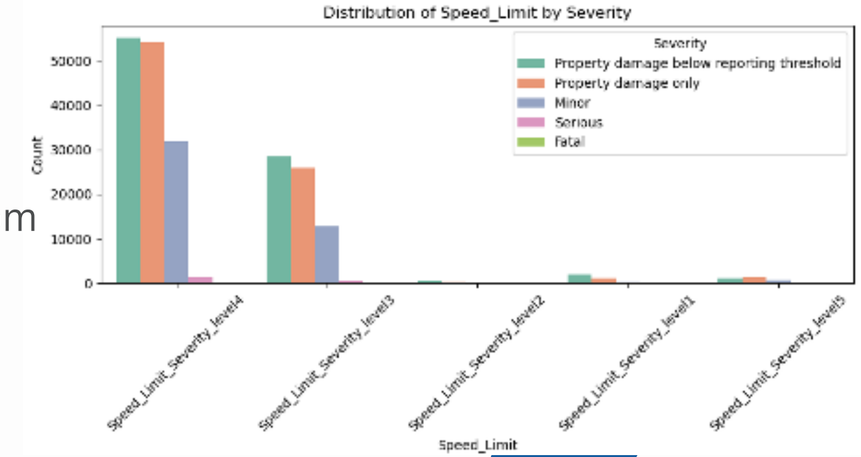
Weather

Includes clear weather and strong winds



Speed Limit

Includes zones from 30 to 90 km/h.



Accident Basic Info

Seq_Num
Acc_Date
Acc_Time

Accident Details

Acc_Type
Severity

Road conditions & Infrastructure

Acc_Type
Surface_Cond
Light_Cond
Environ_Type
Road_Cat
Road_Aspect
Loc_Code
Road_Config
Weather_Cond
Speed_Limit

shape	(218128, 49)
Categorical variables	18 (object, int)
Numerical variables	31 (int, float)

Casualty Counts

Total_Victims

Vehicle Counts

Num_Veh_Invld
Num_Light_Veh
Num_Heavy_Truck
Num_Equip
Num_Bus
Num_Bike
Num_Moped
Num_Moto
Num_Taxi
Num_Emerg
Num_Snowmobile
Num_OffRoad
Num_Other_Veh
Num_Unspec_Veh

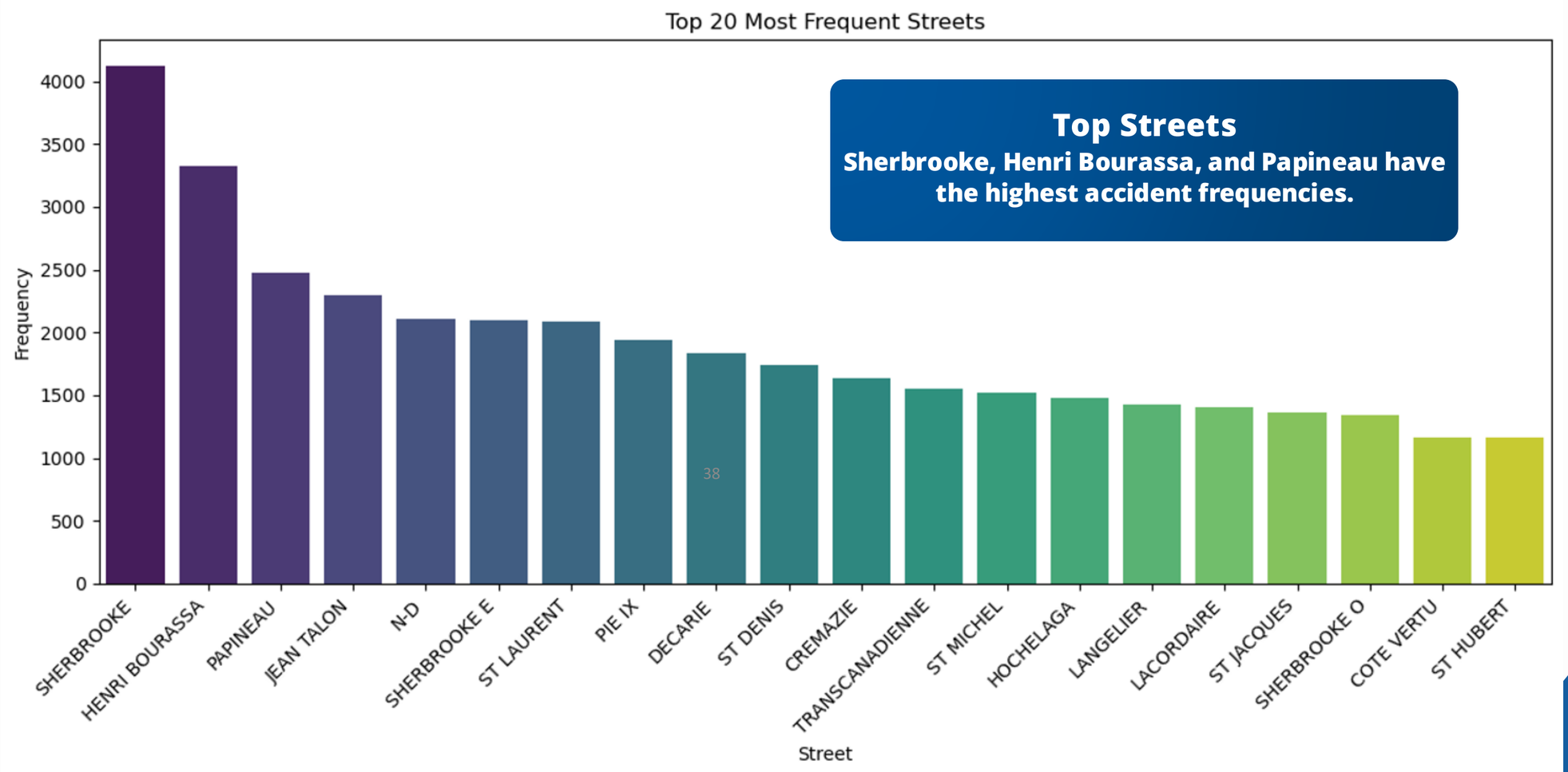
Location Details

Street
Near_To
Lontitude
Latitude

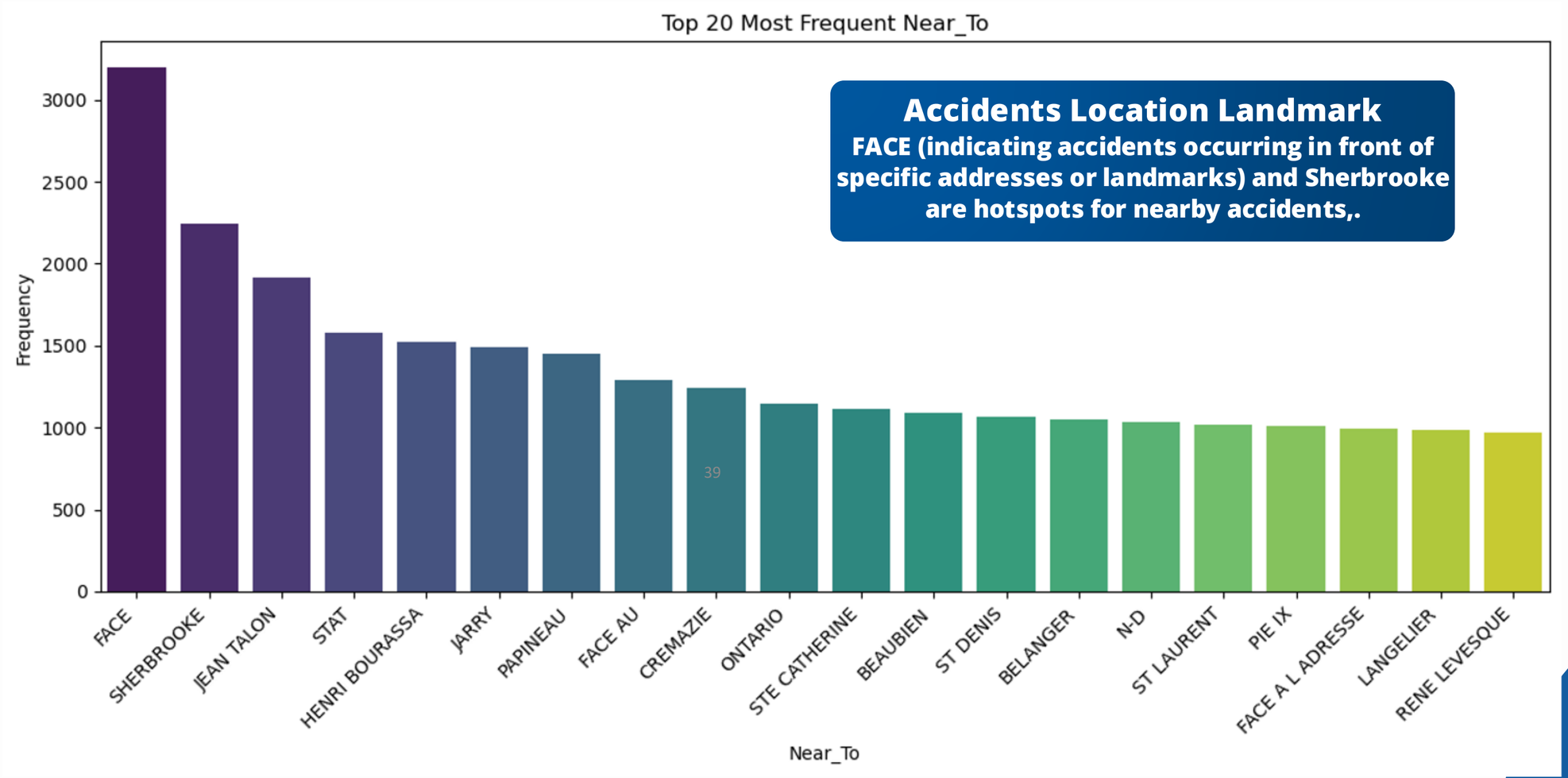
Geolocation Quality

Loc_Quality
Loc_Accuracy
Loc_Imprecise

Severity Patterns Across Key Factors



Severity Patterns Across Key Factors



EDA Summary

Time-Related	
Accident Time	Clear peak between 3:00 p.m. and 5:00 p.m., with minor and property damage incidents often occurring during rush hour; serious accidents follow a similar pattern.
Weekday	Fridays see increased accidents, likely due to end-of-week traffic flow changes.
Month	Peaks in January, February, June, and July due to winter weather and summer travel rates, with a December spike tied to holiday travel.
Year	Overall downward trend in accidents since 2013, with a notable drop in 2020
Vehicles Involvement	
Light vehicles	most commonly involved in accidents
Bikes, Motorcycles	high proportions of severe accidents relative to their counts, potential indicators of serious outcomes
Numerical variables	
Total_Victims	Strongest predictors of accident severity. Higher counts correlate with increased severity, indicating a direct relationship between the number of victims/vehicles and accident impact.
Num_Veh_Invld	
Location Details	
Street	Major roads and locations emerge as accident hotspots, suggesting the need for targeted safety interventions.
Near to Landmark	
Road conditions & Infrastructure	
Accident Types	Pedestrian and cyclist collisions, rollovers, and collisions with fixed objects (e.g., guardrails).
Surface Conditions	Icy or gravel-covered roads are associated with severe outcomes, as well as nighttime or illuminated conditions in business/commercial areas.
Road Categories	Main arteries, numbered roads, and highway ramps show higher severity, likely due to increased speeds and traffic volumes
Road Configuration	One-way streets and two-lane, two-directional roads are notably associated with higher severity scores.
Credibility Score	No clear relationship with accident severity. May serve better as a data quality indicator or weighting factor in analysis to account for data accuracy.

Key Predictors From EDA

Acc_Type, Surface_Cond, Light_Cond, Road_Cat, Road_Config, Num_Heavy_Truck, Num_Moto, Num_Bike, Num_Taxi, Num_Veh_Invld, Total_Victims, Speed_Limit, Weather_Cond

Feature Selection

What we've done?

01

Statistical tests

- ANOVA Test for Numerical predictors
- Chi-Square test - Categorical predictors

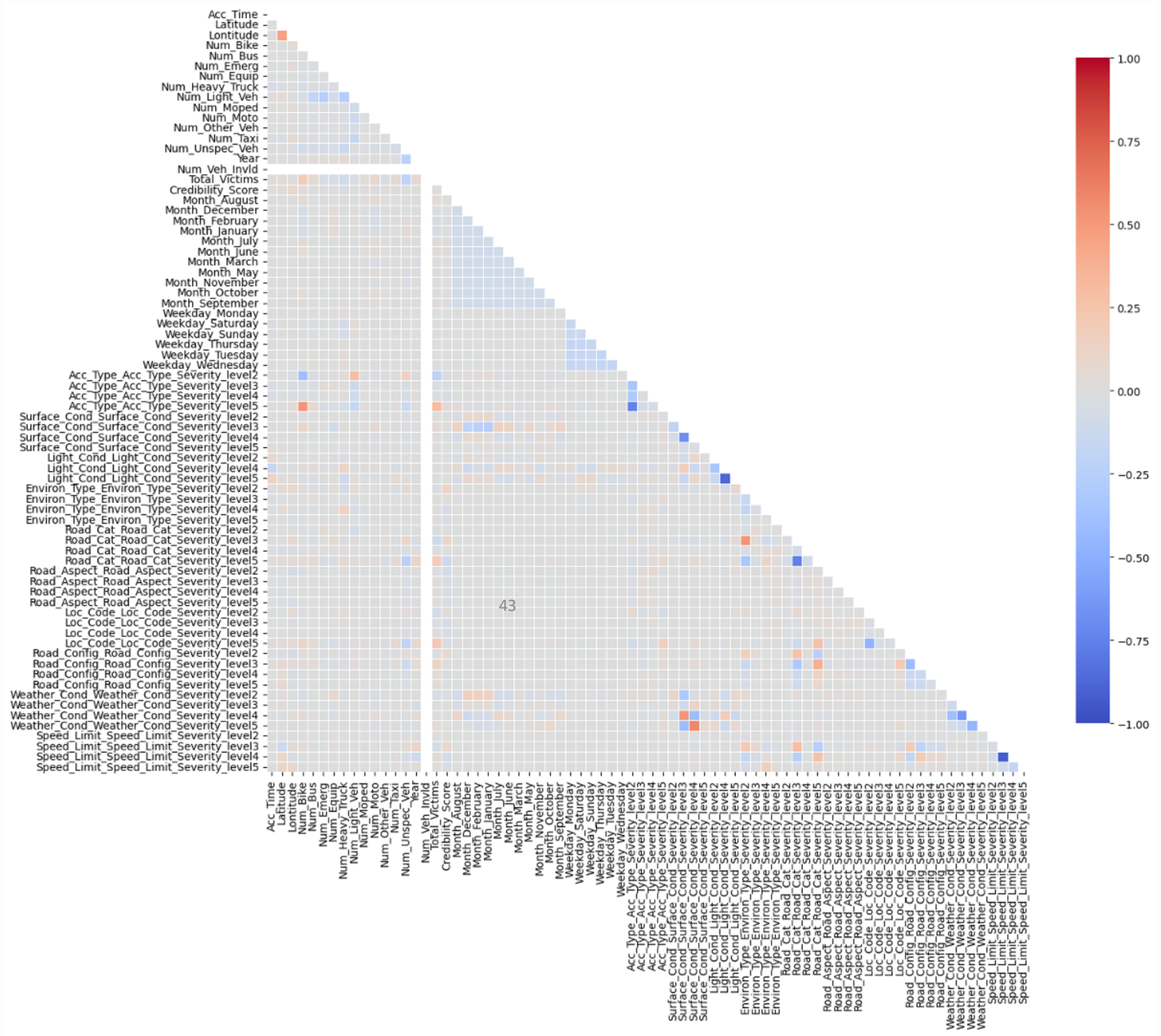
02

Pearson Correlation Matrix

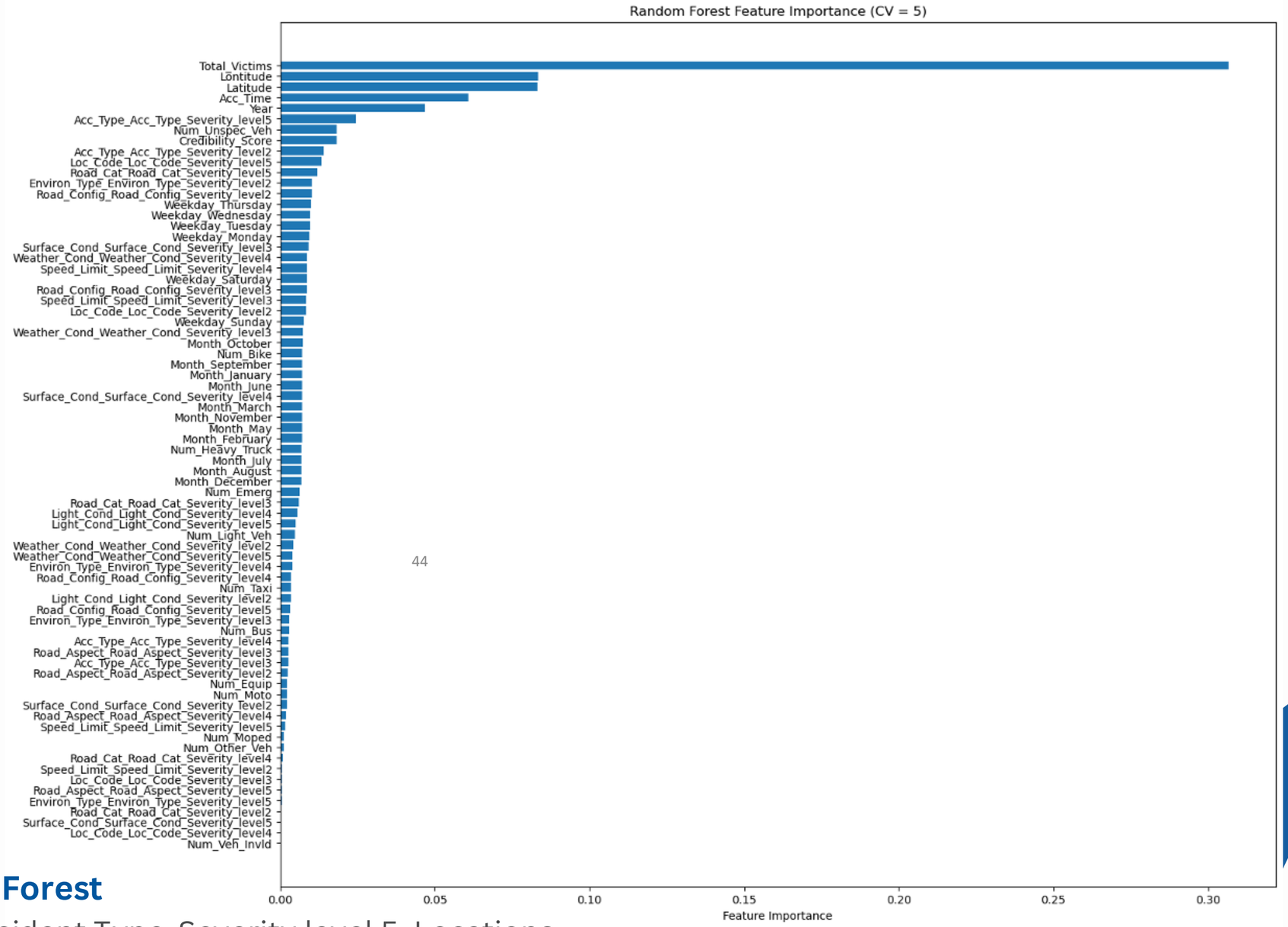
03

Random Forest Feature Importance

Pearson Correlation Matrix



Feature Importance

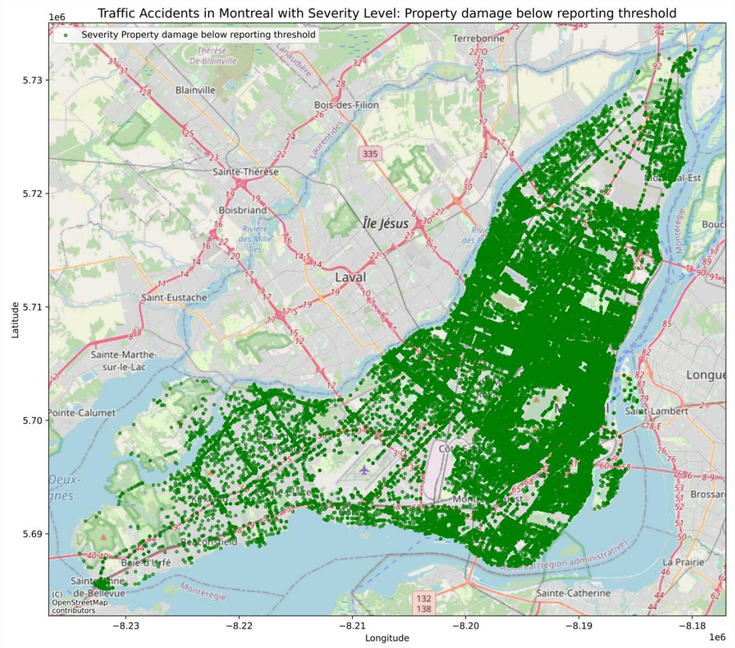


Key Predictors From Random Forest

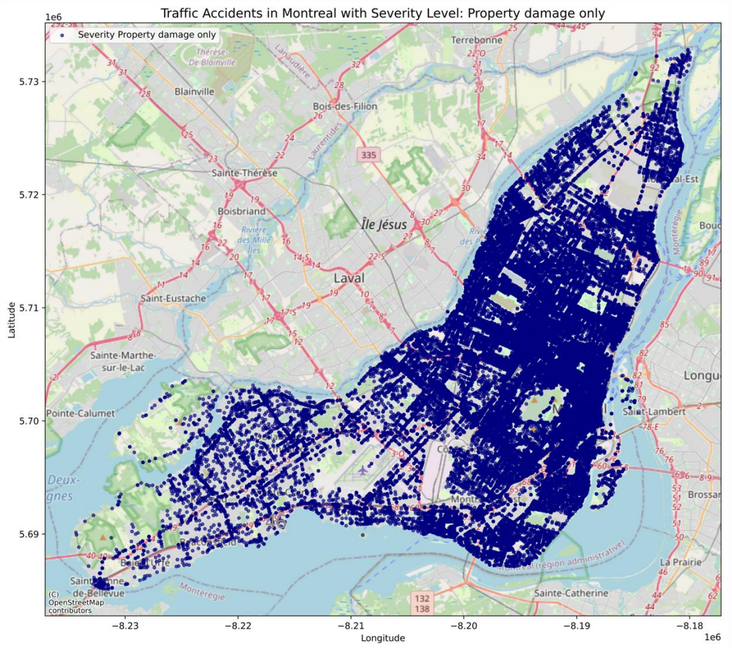
Total_Victims, Accident Time, Accident Type_Severity level 5, Locations,.....

Clustering

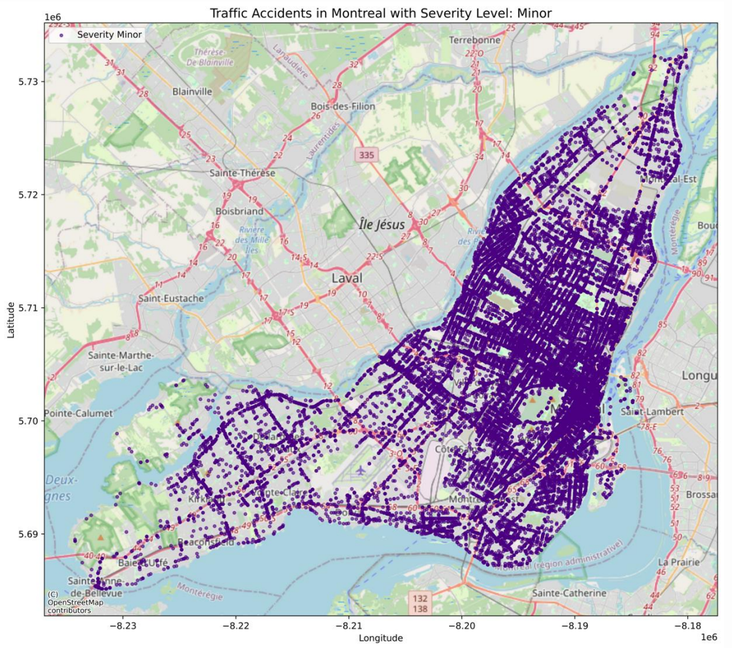
Spatial distribution of accident severity in Montreal



Property Damage Below Reporting Threshold

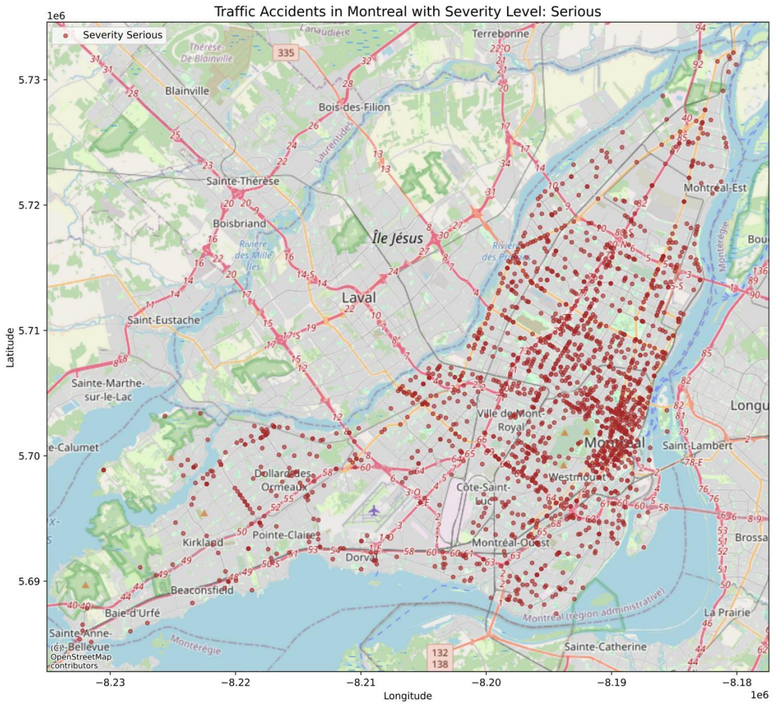


Property Damage Only



Minor Severity

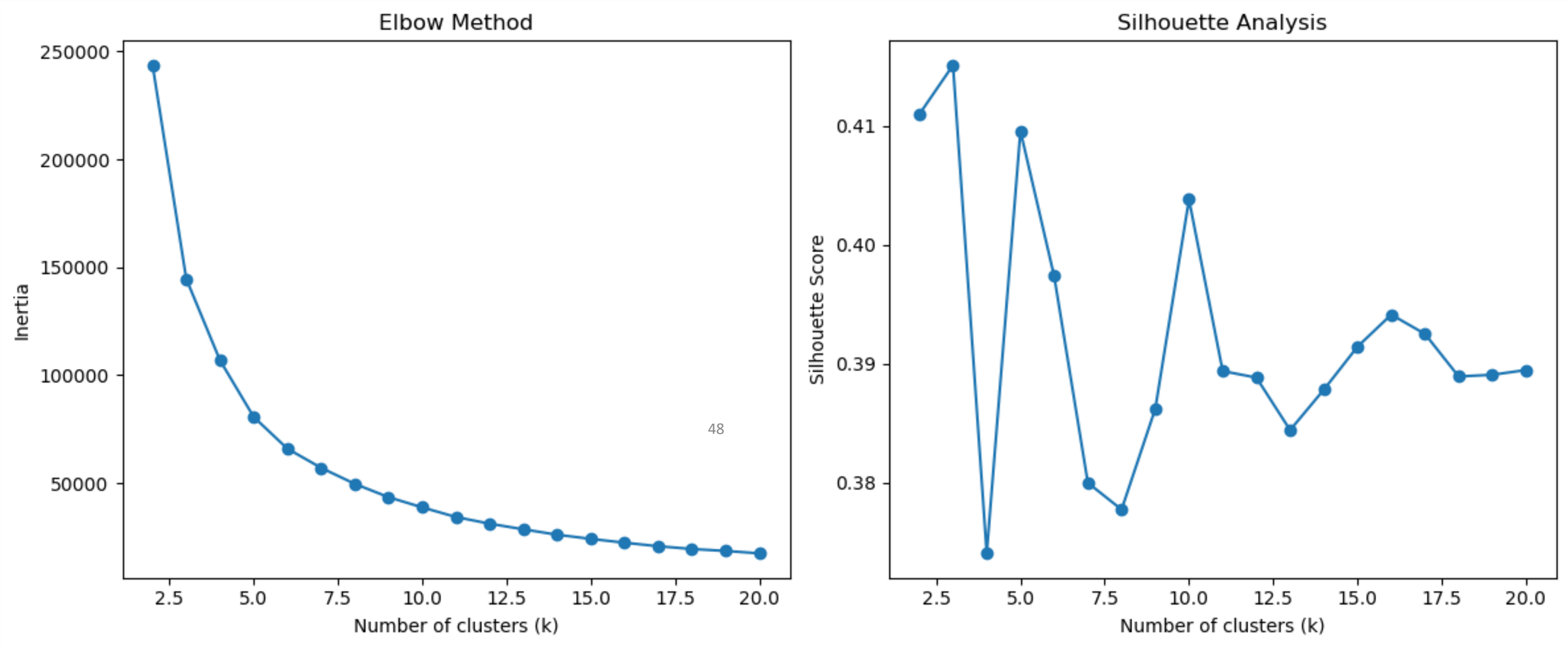
Spatial distribution of accident severity in Montreal



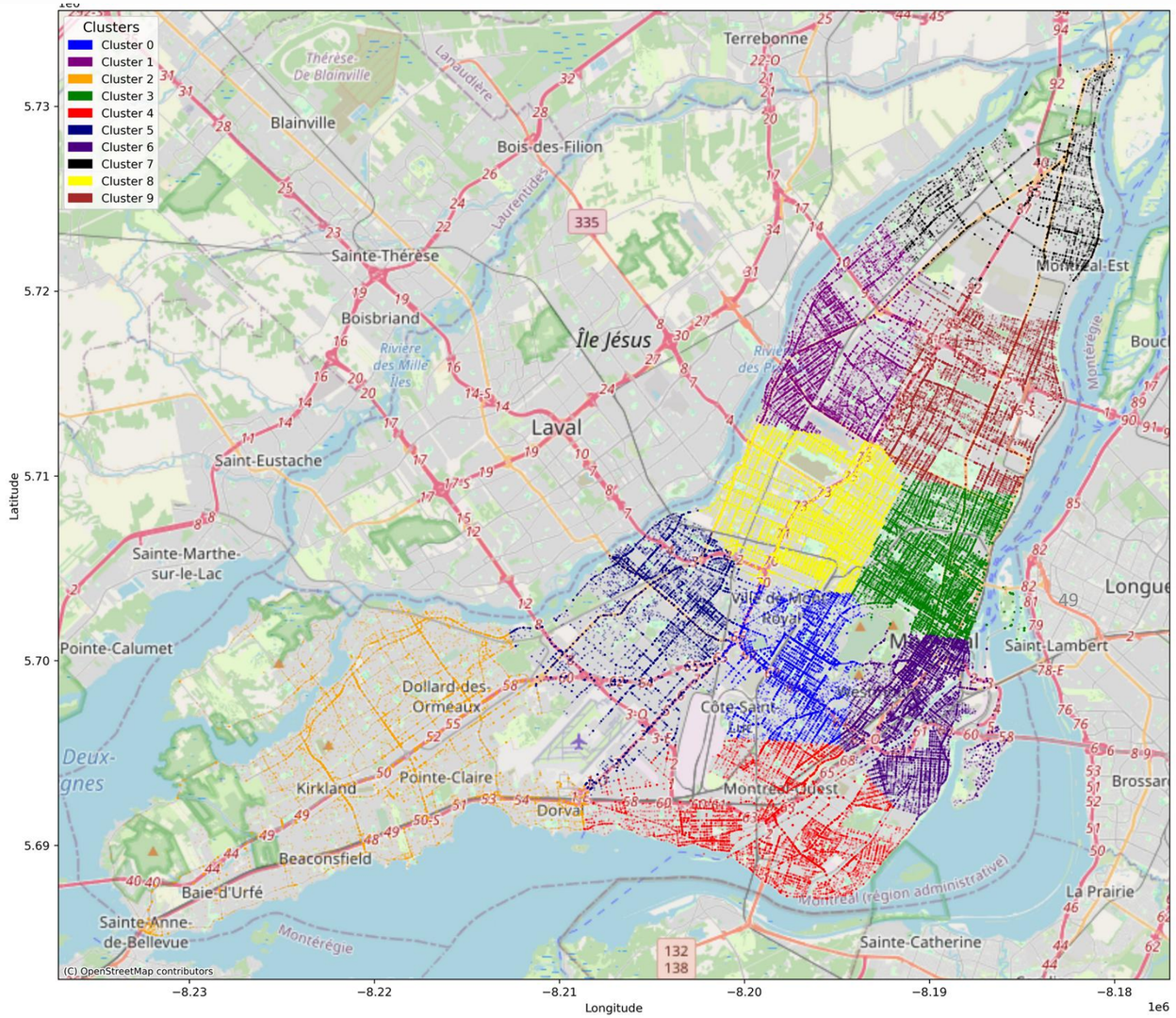
Serious Severity

Fatal Severity

Traffic accidents clusters in Montreal

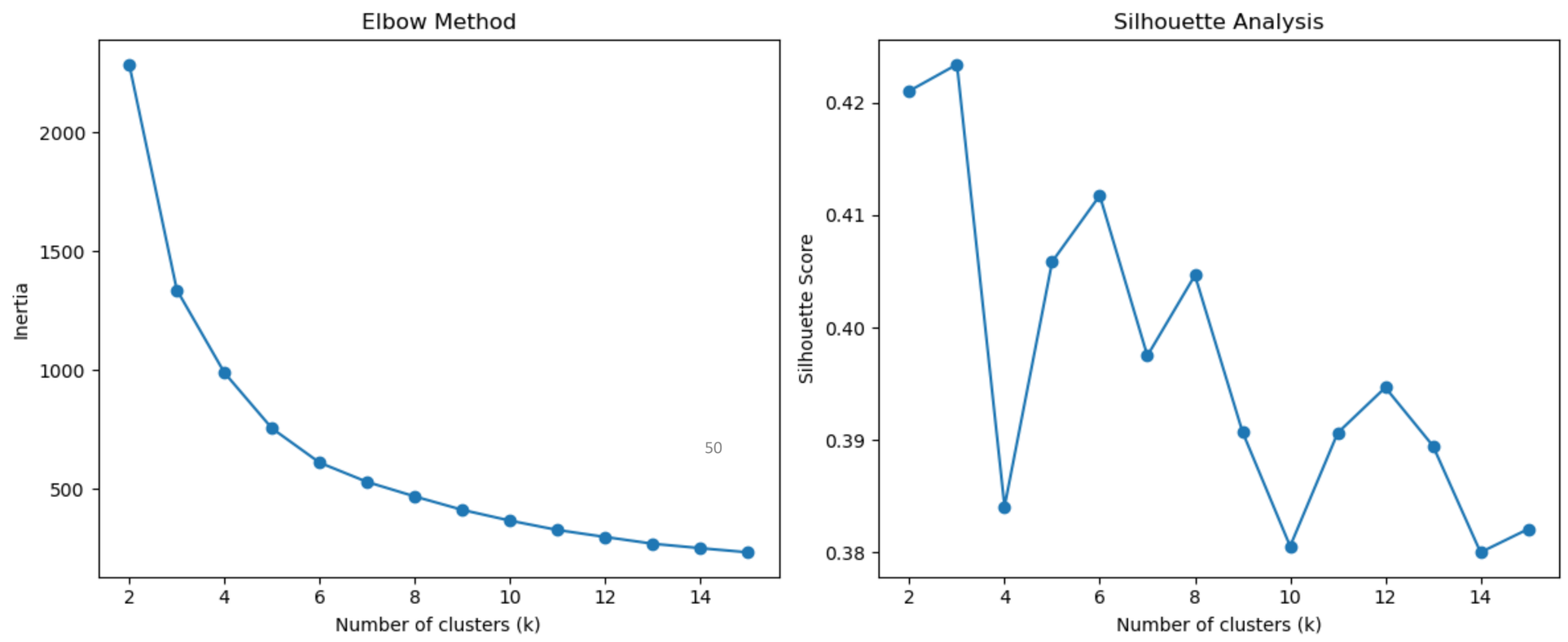


Traffic accidents clusters in Montreal

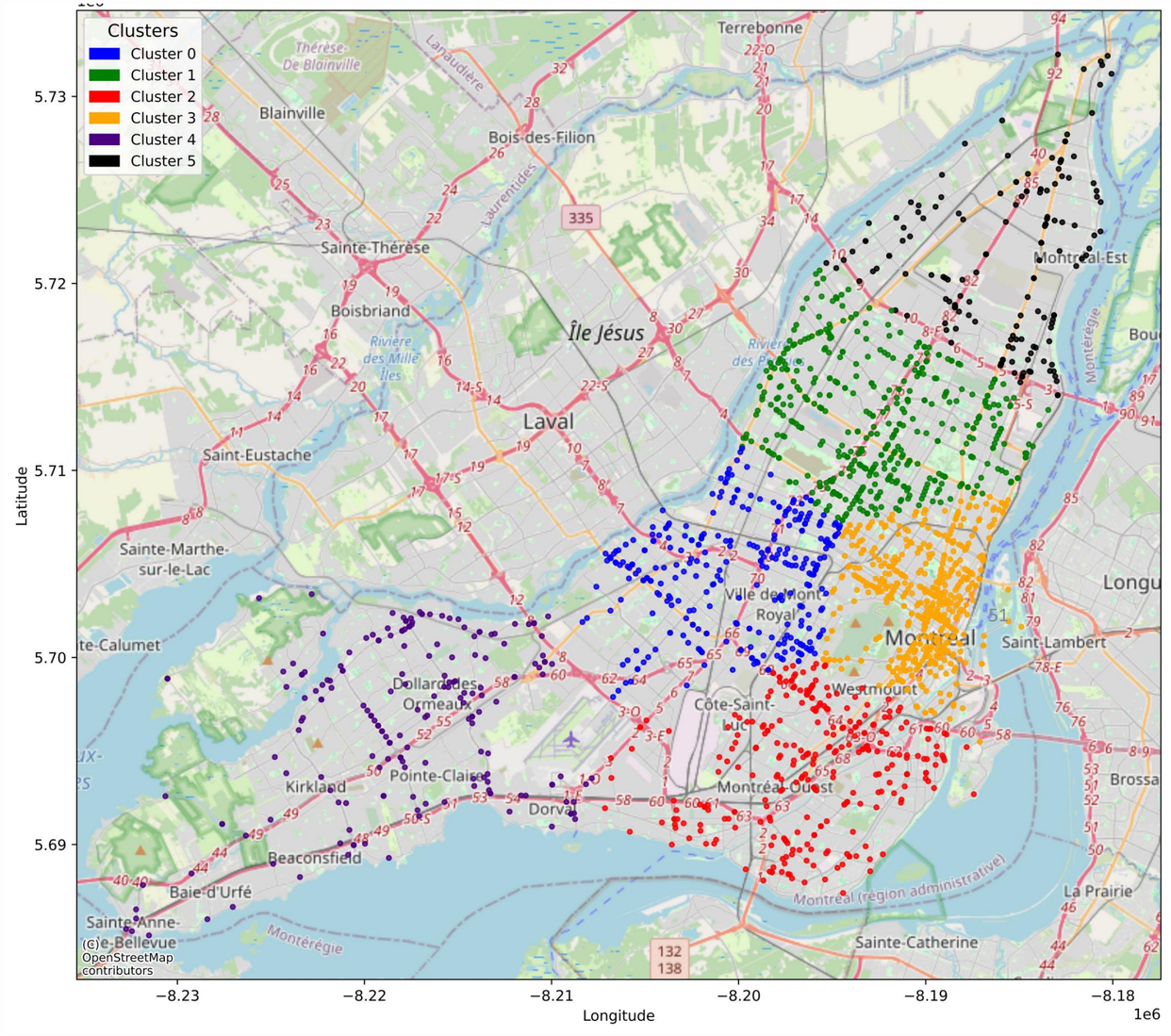


	Cluster	Size	Avg_Victims	Avg_Severity	Most_Common_Road_Config	Most_Common_Environment
	0	17864	0.2496	0.7963	Severity_level1	Severity_level2
	1	20585	0.3316	0.9061	Severity_level2	Severity_level2
	2	17769	0.2510	0.8243	Severity_level2	Severity_level1
	3	37183	0.2707	0.8168	Severity_level1	Severity_level2
	4	17643	0.2517	0.8001	Severity_level2	Severity_level2
	5	15201	0.2500	0.8351	Severity_level2	Severity_level1
	6	26965	0.2328	0.7603	Severity_level1	Severity_level1
	7	7653	0.2950	0.9053	Severity_level3	Severity_level2
	8	34477	0.2783	0.8538	Severity_level1	Severity_level1
	9	22788	0.2966	0.8687	Severity_level3	Severity_level1

Serious & Fatal Accidents Clusters in Montreal



Serious & Fatal Accidents Clusters in Montreal



	Cluster	Size	Avg_Victims	Avg_Severity	Most_Common_Road_Config	Most_Common_Environment
	0	327	1.3303	3.1468	Severity_level3	Severity_level2
	1	474	1.3354	3.1203	Severity_level3	Severity_level1
	2	299	1.4214	3.1405	Severity_level3	Severity_level2
	3	590	1.2695	3.1220	Severity_level3	Severity_level1
	4	200	1.3500	3.1150	Severity_level2	Severity_level1
	5	157	1.4013	3.1338	Severity_level3	Severity_level1

Predictive Model

Classification Model

Collision severity prediction is multi-class classification problem

Logistic Regression

- Uses logistic function to model probability of each class
- Identify linear relationship between features and severity
- Simple and interpretable results
- Fast training and prediction
- Less prone to overfitting

Random Forest

- Ensemble of decision trees using bagging
- Handles complex interactions between collision factors
- Provides feature importance
- Reduces overfitting through ensemble approach
- Handles outliers well

Gradient Boosting

- Sequential ensemble method using gradient descent optimization
- Progressive learning approach
- Capturing subtle patterns
- Better handling of imbalanced severity classes

XGBoost

- Second-order derivatives (Hessian) for optimization
- Improves in accuracy and convergence speed
- Efficiently handles large dataset
- Regularization to prevent overfitting

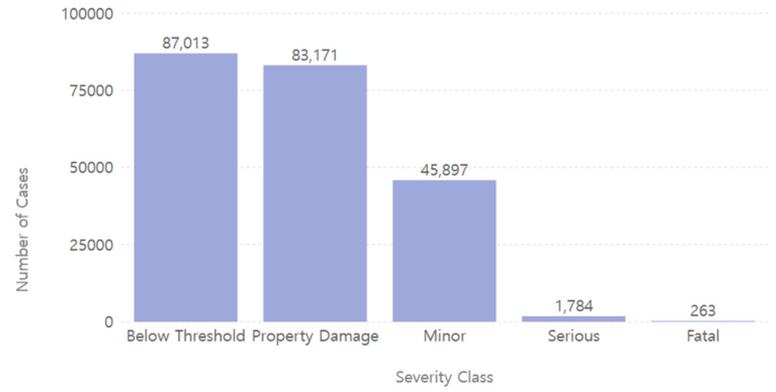
Model performance Comparison

Model	Performance Score
Random Forest	0.7005
Gradient Boosting	0.6892
XG Boosting	0.6679
Logistic Regression	0.6868

Random Forest

Class	Precision	Recall	F1-Score
Damage Below Reporting Threshold	0.65	0.58	0.62
Property Damage Only	0.61	0.67	0.64
Minor	0.96	1.00	0.98
Serious	0.00	0.00	0.00
54			
Fatal	0.00	0.00	0.00
Accuracy			0.70
Macro Avg	0.44	0.45	0.45
Weighted Avg	0.69	0.70	0.70

Severity Distribution



SMOTE

SMOTE Technique Selection

Technique	Total Samples	Avg Borderline Score	Borderline Instances
Regular SMOTE	435,065	0.301441	132,027
Borderline SMOTE	435,065	0.297944	127,925
SMOTE-Tomek	389,785	0.255501	97,369
SMOTE-ENN	**233,591**	**0.043663**	**8,253**

55

- **SMOTE-ENN** demonstrates the lowest borderline instances (8,253) and cleanest decision boundaries (0.043663 avg borderline score)
- Tomek removes overlapping samples between classes, while ENN removes any instances whose class differs from majority of its neighbors. This cleaning helps achieve clearer class boundaries
- Regular SMOTE and Borderline SMOTE actually increased the original borderline instances (122,765) to 132,027 and 127,925 respectively

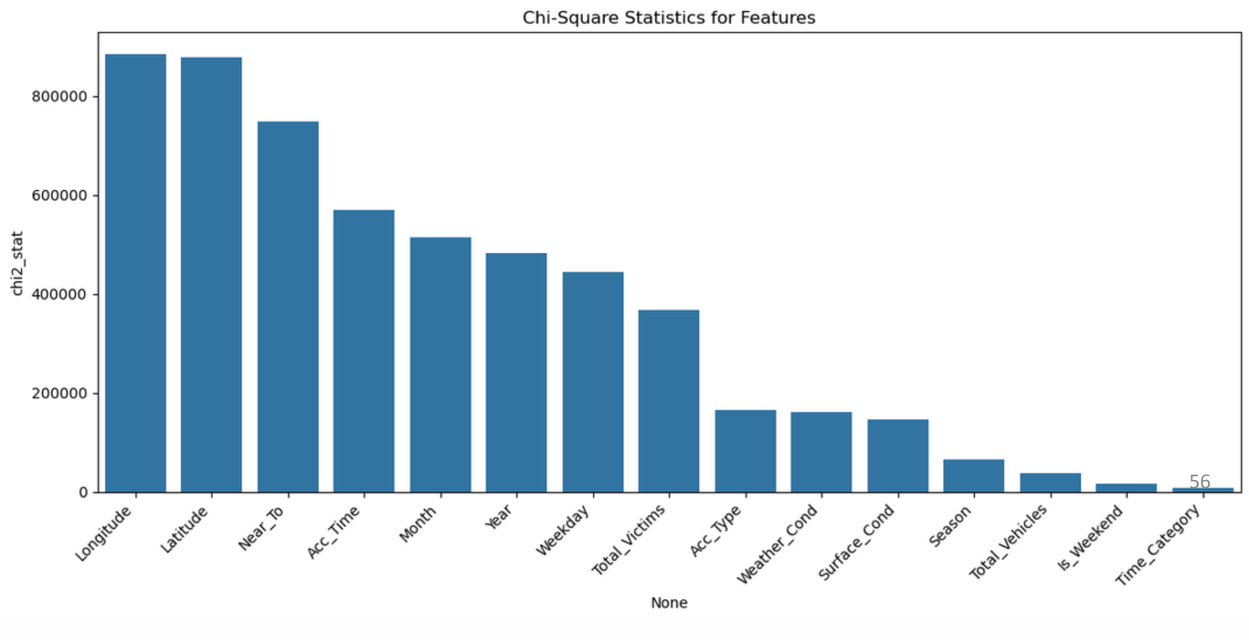
What can we do?

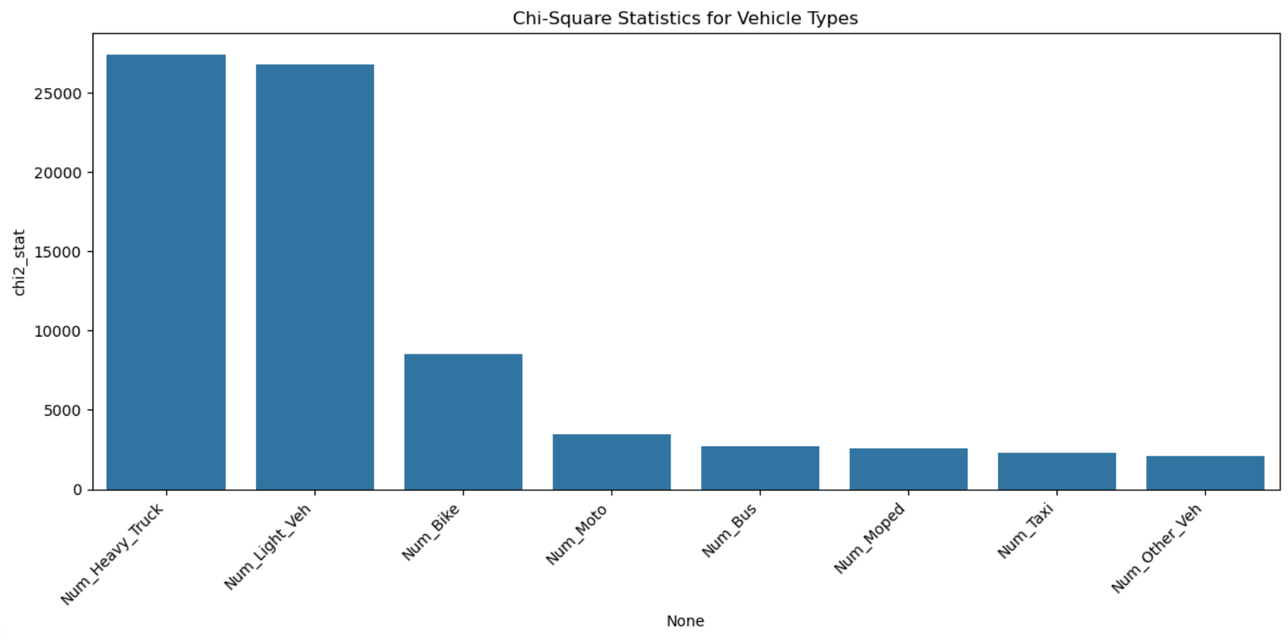
- Severe collisions are naturally rare
- Traditional oversampling can lead to overfitting
- SMOTE reduces the risk of overfitting by generating synthetic samples

Problem:

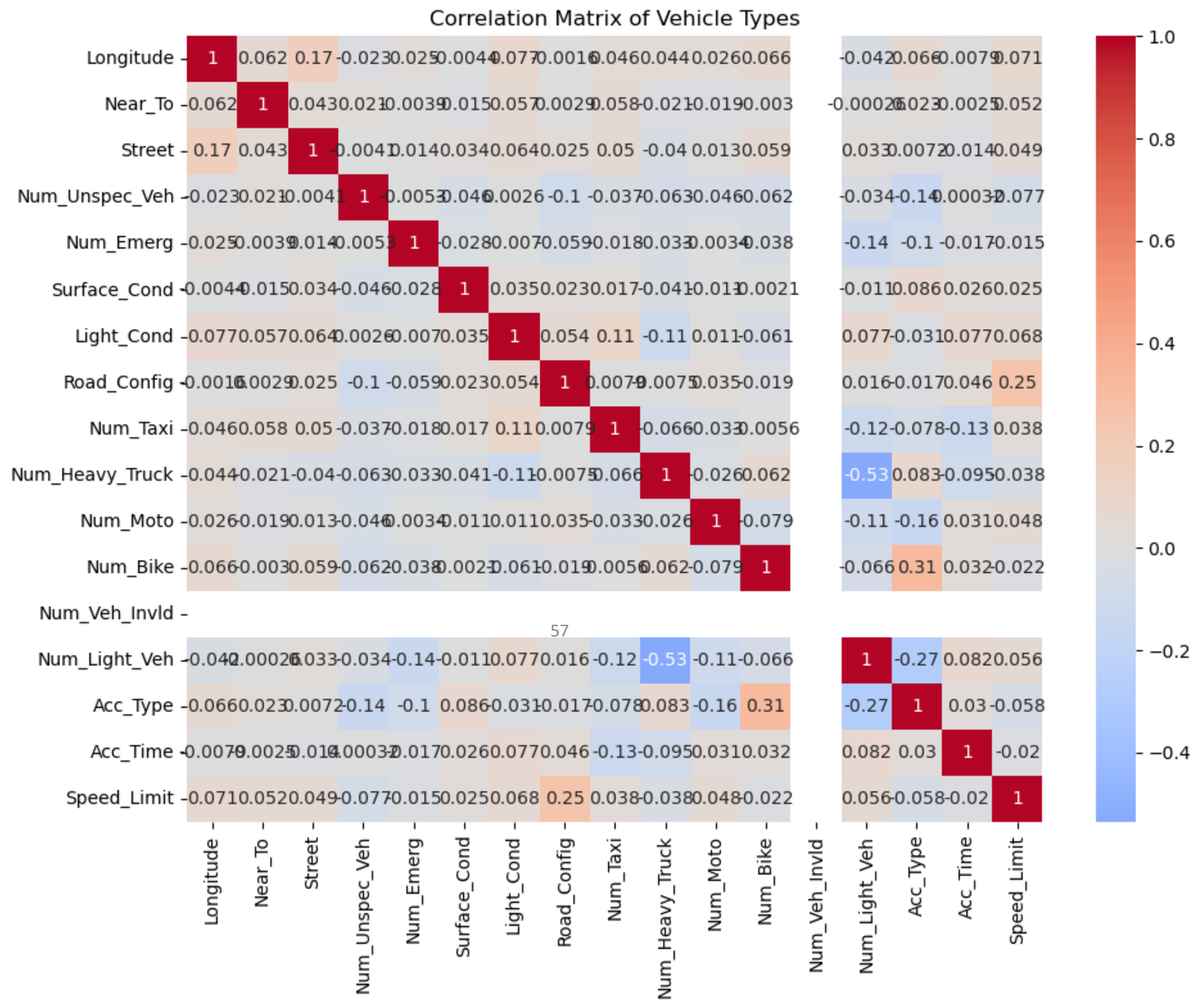
- Large number of borderline instances (122,765) which indicates classification uncertainty
- Applying SMOTE without additional steps might amplify the complexity or introduce noise

Feature Selection

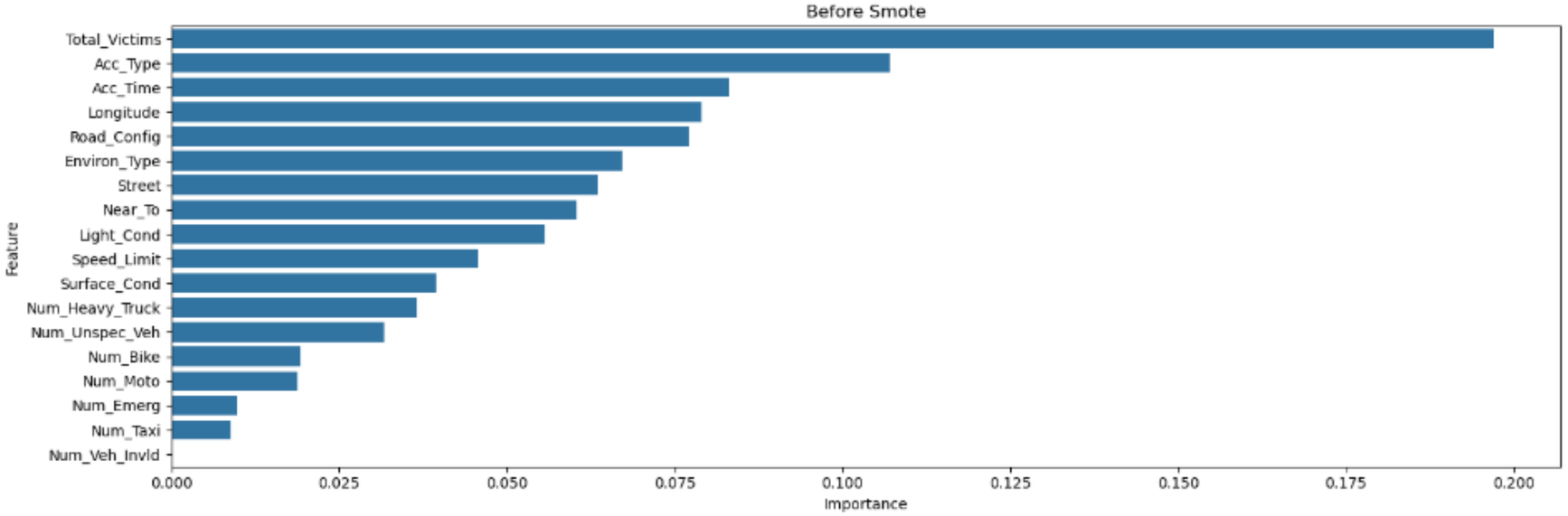


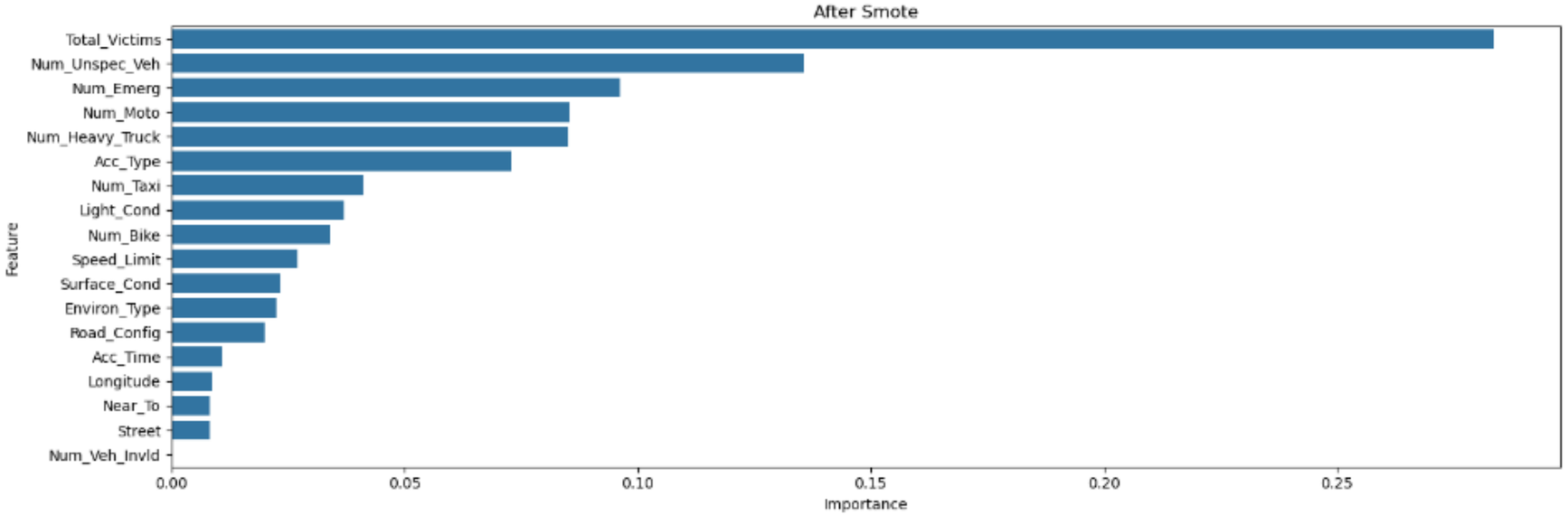


Pearson Correlation Matrix



Feature Importance





Feature Selection

- Total_Victims
- Longitude
- Near_To
- Street
- Num_Unspec_Veh
- Num_Emerg
- Surface_Cond
- Light_Cond
- Environ_Type
- Road_Config
- Num_Taxi
- Num_Heavy_Truck
- Num_Moto
- Num_Bike
- Num_Veh_Invld
- Acc_Type
- Acc_Time
- Speed_Limit

Final Performance

0.95
Accuracy

0.91
Macro Average

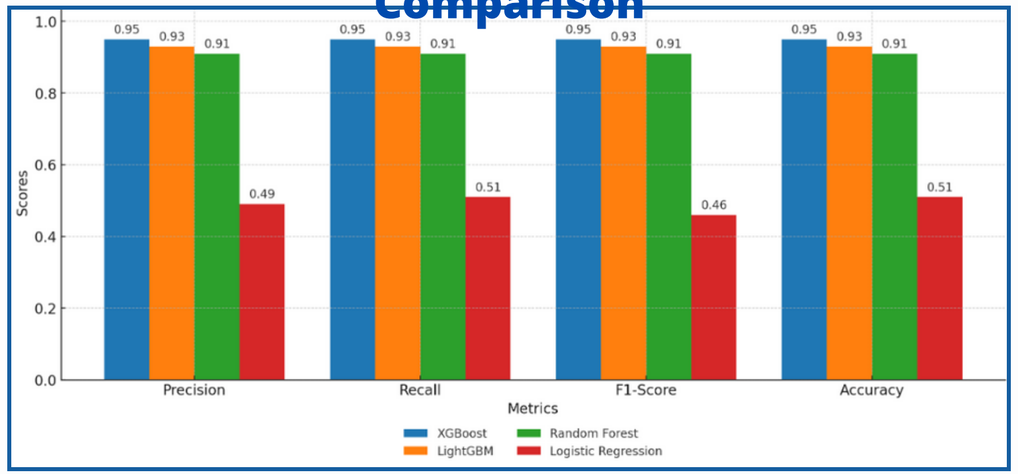
0.95
Weighted F-1 Score

0.95
Weighted Average

XGBoost

Severity Level	Precision	Recall	F-Score
0	0.83	0.81	0.82
1	0.81	0.83	0.82
2	0.98	0.93	0.95
3	0.95	0.98	0.96
4	0.99	1.00	1.00

Model Performance Comparison



Business Value

Business Value

Cost Reduction

- *Emergency Services Optimization*
  - Reduce emergency response time in high-risk clusters
  - Decrease preventable fatalities in severity level 4 clusters due to faster response.

- *Infrastructure Investment Prioritization*
  - Upgrade road infrastructure in high-risk zones with frequent severity levels 3 and 4
  - e.g., downtown, St. Laurent Boulevard, major intersections in the Plateau-Mont-Royal borough

Public Awareness Campaigns

- *Targeted Safety Messaging*
  - Run campaigns focused on high-risk times and areas, such as pedestrian-heavy zones in Plateau-Mont-Royal or bike lanes in Outremont.

- *How to Implement*
  - Distribute digital content on social media and traditional media platforms (radio, billboards)
  - Install electronic road signs displaying real-time safety messages

Business Value

Policy and Governance Value

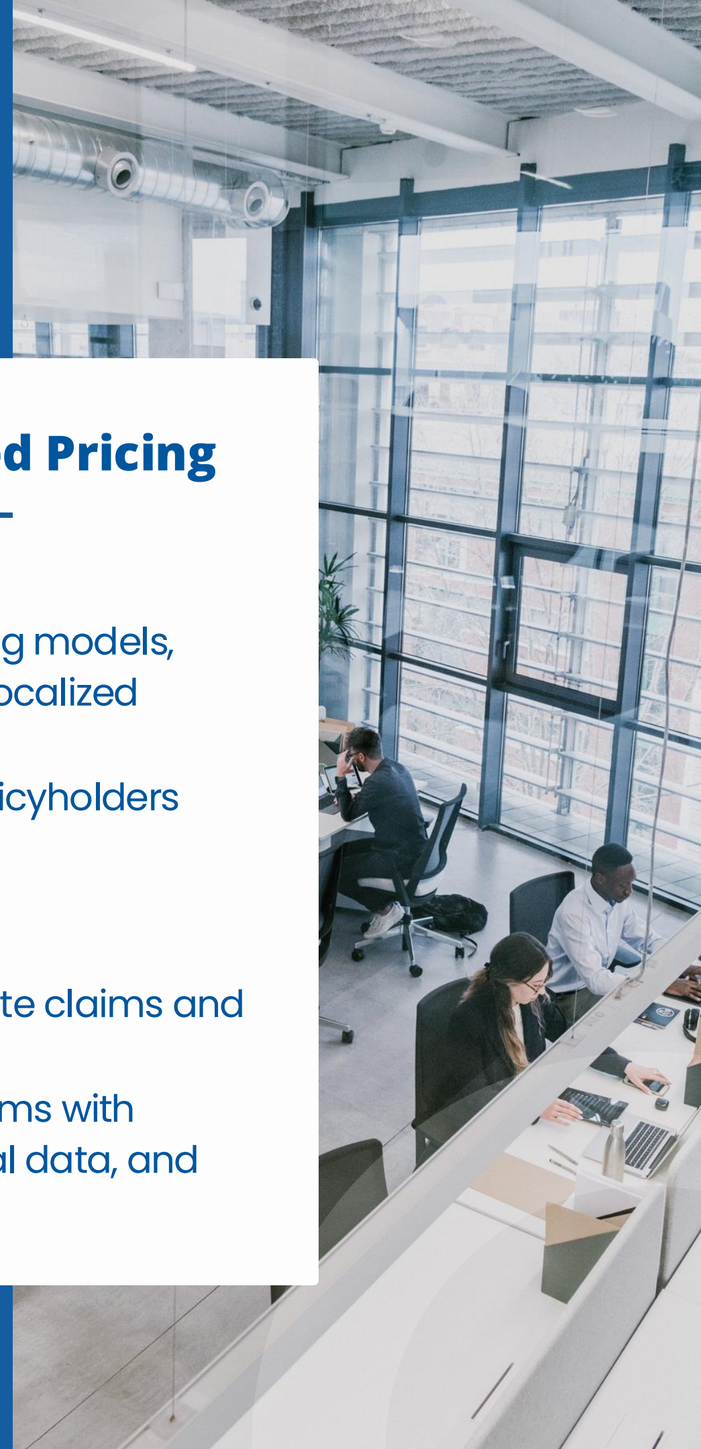
- *Evidence-Based Policy Making*
  - Enables the government to design effective, targeted road safety policies
  - Demonstrates accountability and improves citizen satisfaction with public services

- *Public Trust*
  - Proactive safety measures enhance public trust in government institutions
  - Higher approval ratings for municipal and provincial governments

Insurance and Risk-Based Pricing

- *Dynamic Risk Assessment*
  - Use predictions to refine pricing models, offering premiums based on localized accident risks
  - Differentiate premiums for policyholders based on driving patterns

- *Fraud Detection and Prevention*
  - Leverage predictions to validate claims and identify inconsistencies
  - Cross-reference accident claims with severity predictions, geospatial data, and temporal trends.



le Service des infrastructures du réseau routier (SIRR)

01

Sherbrooke Street

- Install protected turn signals at intersections known for high left-turn collision rates
- Upgrade drainage systems in areas where water accumulation has been reported
- Increase winter maintenance efforts in regions where icy conditions are common

02

Henri-Bourassa Boulevard

- Install pedestrian protection systems where "collision with pedestrian" is frequent

03

Data Integration and Analytics

- *Integrate Traffic Data*
  - Aggregate historical accident data, weather conditions, traffic patterns, and real-time road conditions from existing municipal and provincial databases.
  - Include IoT devices (e.g., traffic cameras, vehicle counters) for live updates.
- *Partners*
  - Establish a central data repository for storing and managing accident prediction data.

Service de Police de la Ville de Montréal (SPVM)

01

Target enforcement

- *Commercial areas (CD_ENVRN_ACCDN = 3) during peak hours*
  - Peak hours coverage: 3:00-7:00 PM
  - Loading zone monitoring
  - Double-parking enforcement
  - Intersection blocking prevention

02

Emergency Response

- *Sherbrooke East Corridor*
  - 2 Advanced Life Support units (7AM-11PM)
  - 1 Rapid Response Vehicle (24/7)
  - Coverage radius: 3-minute response time
- *Henri-Bourassa/Highway Intersections*
  - 2 ALS units (Peak hours)
  - 1 Special Operations unit (Winter conditions)
  - Coverage radius: 4-minute response time

Centre de gestion de la mobilité urbaine (CGMU)

01

Real-time Monitoring Network

- *Central Command Center (Guy Street):*
  - 24/7 Operations Center
  - 250+ CCTV camera feeds
  - Traffic flow visualization
  - Weather condition monitors
- *René-Lévesque Boulevard:*
  - 25 intersection cameras
  - Bus lane monitoring
  - Pedestrian counting systems
  - Bicycle flow detection
- *Pie-IX Boulevard:*
  - BRT corridor monitoring
  - Signal priority system

02

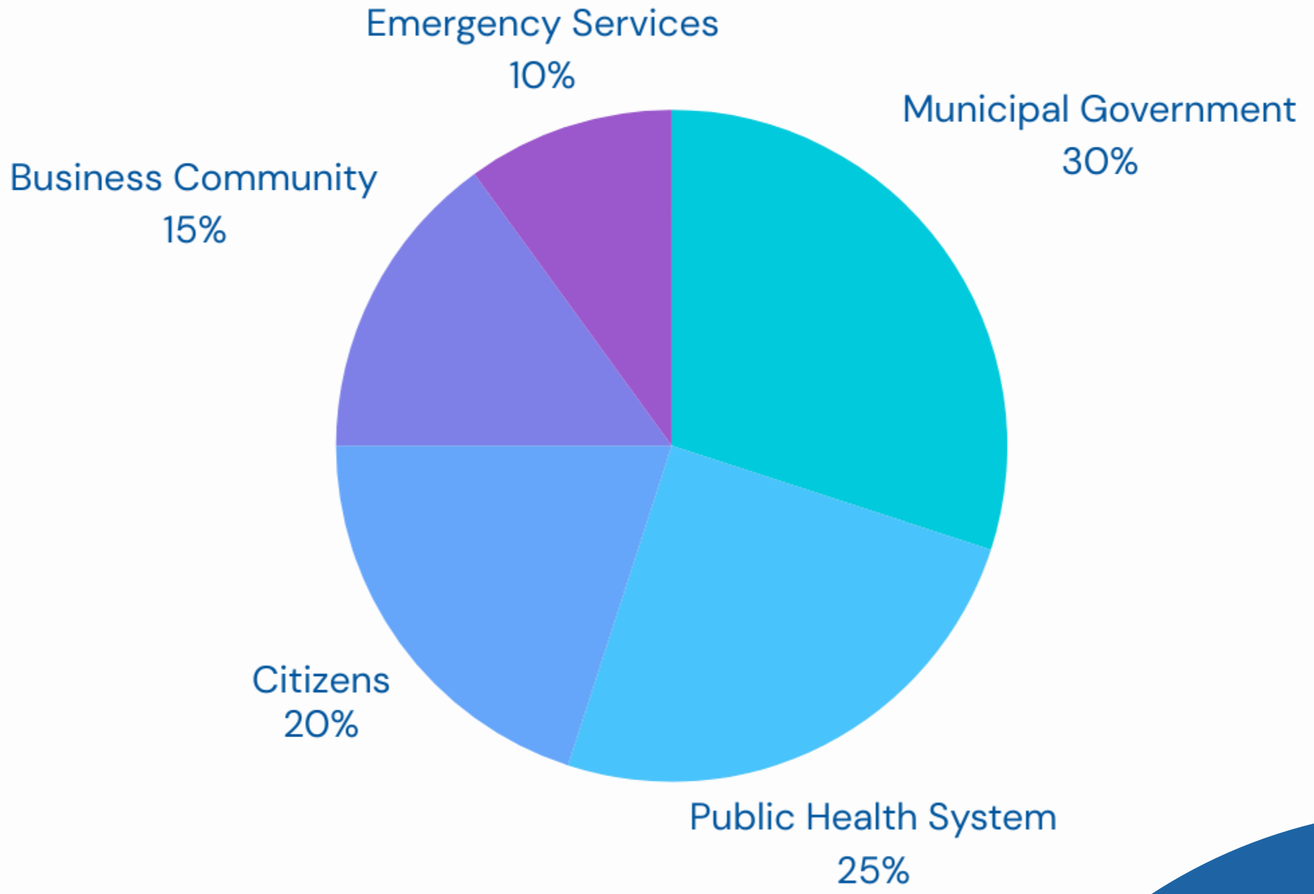
Road Condition Monitoring

- Ice detection systems
- Water accumulation sensors
- Temperature monitoring
- Friction measurement
- Visibility sensors
- Wind speed monitoring

Stakeholder Value Distribution

- **Cross-Reference Sources**
  - SAAQ accident statistics
  - Emergency service records
  - Healthcare system data
  - Municipal budget reports
  - Academic studies
  - Industry reports
  - Comparable city data

- **Distribution Rationale**
  - Direct Impact (40% weight)
  - Implementation Role (30% weight)
  - Long-term Benefits (30% weight)



Future Analysis

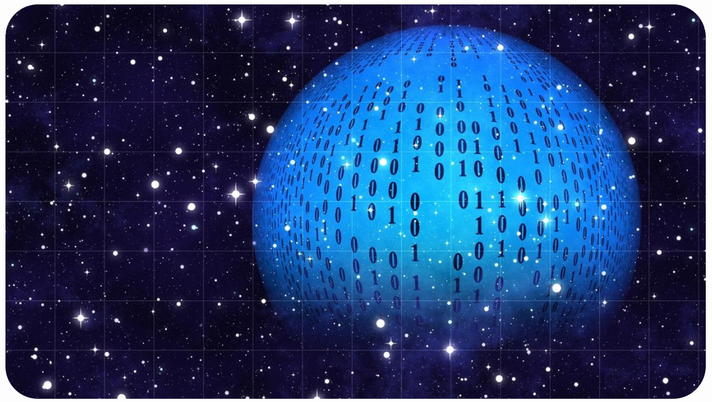
Real-Time Data Integration

Key Objectives

- Utilize real-time data feeds to predict traffic accident risks dynamically
- Enable authorities to implement proactive measures to reduce accident likelihood
- Provide real-time alerts to road users to avoid high-risk areas

Data Sources

- **Traffic Data:** Vehicle density, flow, and speed collected via sensors and GPS
- **Weather Data:** Real-time weather updates, including precipitation, temperature, and visibility conditions
- **Road Conditions:** Real-time updates on surface conditions like ice, snow, or gravel through IoT-enabled road infrastructure
- **Incident Reports:** Data from emergency services, social media, or community reporting apps on accidents or road blockages



Challenges and Solutions

- **Data Integration:**
  - Challenge: Integrating diverse data sources in real-time.
  - Solution: Use robust APIs and data pipelines with standardized formats.
- **Accuracy and Timeliness:**
  - Challenge: Ensuring predictions are both accurate and timely.
  - Solution: Use high-frequency updates and continuous model retraining.

Summary

To improve road safety and optimize resource allocation in Montreal by predicting traffic accidents' severity using advanced machine learning models. The project identifies high-risk zones, timeframes, and contributing factors, providing actionable insights for stakeholders.

Deliverables

- A robust predictive model that forecasts accident severity
- Identification of accident hotspots and temporal patterns
- Actionable insights tailored for stakeholders

Business Value

- Financial Impact
- Social Value
- Environmental Benefits
- Strategic Value

Stakeholder Benefits

- City of Montreal
- Insurance Companies
- Residents
- Businesses

REFERENCES

Hebert, A., Guedon, T., Glatard, T., & Jaumard, B. (2019). High-resolution road vehicle collision prediction for the City of Montreal. 2019 IEEE International Conference on Big Data (Big Data), 1804–1813. https://doi.org/10.1109/bigdata47090.2019.9006009

Muktar, B., & Fono, V. (2024). Toward safer roads: Predicting the severity of traffic accidents in Montreal using machine learning. Electronics, 13(15), 3036. https://doi.org/10.3390/electronics13153036

Transport Canada. Canadian Motor Vehicle Traffic Collision Statistics 2021. Available online: https://tc.canada.ca/en/road- transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2021

Ville de Montréal. Collisions Routières, [Jeu de données]. Dans Données Québec, 2018. Mis à jour le 19 Décembre 2022. 2022.
Available online: https://www.donneesquebec.ca/recherche/dataset/vmtl-collisions-routieres

Service de Police de la Ville de Montréal. (2023). Rapport annuel 2022.

Statistics Canada. (2023). Transportation in Canada 2022. https://www144.statcan.gc.ca/tdih-cdit/index-eng.htm

THANK YOU