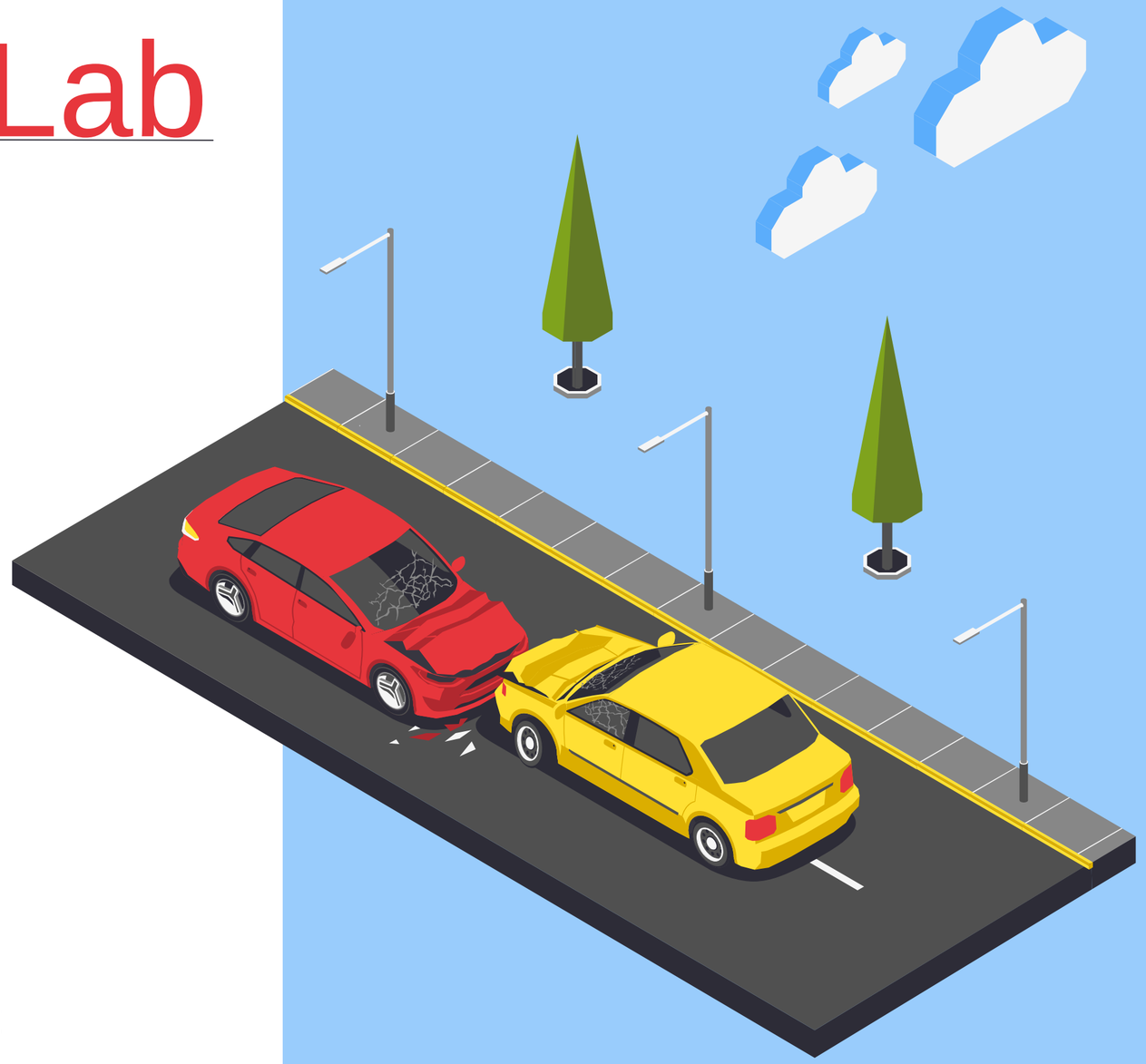
The Foundation Lab



Building Foundations with Data & Insight

TABLE OF CONTENTS






POC REVISITED

THE VALUE OF THE FOUNDATION LAB PROJECT

Current problems

Montreal sees about 21 traffic deaths per year, yet current analyses are mostly reactive and focus on explaining severity after crashes occur.

Existing research identifies broad factors (e.g., weather, road conditions) but does not pinpoint where or when high-risk collisions are likely to happen.

Our solutions

We built a predictive model that estimates collision likelihood across Montreal rather than studying only crash severity.

We localized the analysis to conditions of the space and time that elevate risk.

Impacts

The City of Montréal can target specific infrastructure improvements, e.g. upgrading high-risk corridors or unsafe intersections.

Reducing the number of severe to mortal collisions by 25%.

REVISITING HYPOTHESES.....

H.1 : The probability of severe crashes is higher in specific neighborhoods

Our EDA Findings

- **42% severe/fatal crashes were truly recorded in top 10 areas (e.g., Ville-Marie).**
- **Neighborhood is the top feature for enhanced model accuracy/recall.**

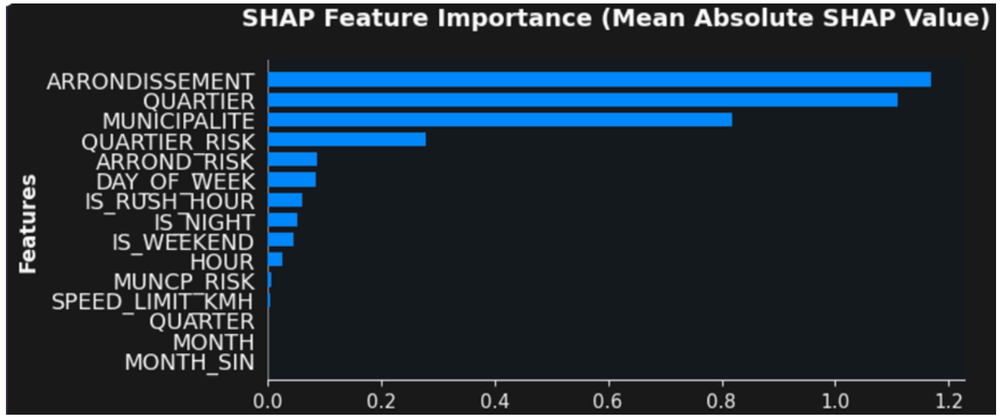
Keep Hypothesis 1

REVISITING HYPOTHESES.....

H.2 & H.3 : The probability of severe crashes is higher on cracked roads and/or flat curved roads.

Our EDA Findings

Road defects appeared in fewer than 5% of severe cases, while night-time and high-traffic periods showed far greater predictive power and policy relevance.



New Hypothesis 2: The Probability of Severe Crashes is Higher for vulnerable road users, especially in areas without protected infrastructure.

New Hypothesis 3: The Probability of Severe Crashes is Higher During Night-Time and Rush-Hour Periods on High-Traffic Streets.

DATASET UPDATE

UPDATE ON THE DATA CLEANING TO INCREASE IMPACT

Switched from Postal Codes to Neighborhoods

- Problems Discovered:
  - **noisy patterns, overfitting in model, hard to draw meaningful recommendations** (e.g., "fix postal code H2X 3Y7" is not useful for city planners).
- Why Neighborhoods Are Better:
  - **Larger areas** → more crashes per zone (stronger patterns, less noise)
  - **Actionable for recommendations**: Aligns with city planning (e.g., "target Ville-Marie neighborhood").
  - **Ties to real impact**: Neighborhoods match demographics/traffic flow → easier to deploy interventions and measure lives saved

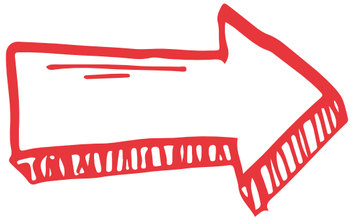
Multi-Class vs Binary – Why Binary Wins

- Why discarded: Too hard to separate similar classes (e.g., Léger vs Dommages material only) → focuses on common minor crashes, ignores rare fatal ones
- **Final choice: Binary model (Severe = Grave + Mortel vs. Everything Else)**
- Why binary wins: Prioritizes catching deaths (recall) over "perfect" labels → real impact to lives/injuries saved per year

DATA TRANSFORMATION PIPELINE TO ADRESS ISSUES FROM THE EDA

INITIAL FEATURES EDA

- 71 raw columns
- French codes & numeric labels
- Missing/unknown values
- Raw timestamps & geo IDs
- Severe crashes extremely rare
- Mixed text & numbers



INITNIAL FEATURES NOW

- 22 as baseline
- Time & neighborhood risk features
- Balanced training set via SMOTENC when encodinc
- Fully numeric, model-ready predictors

FEATURES ENGINEERING

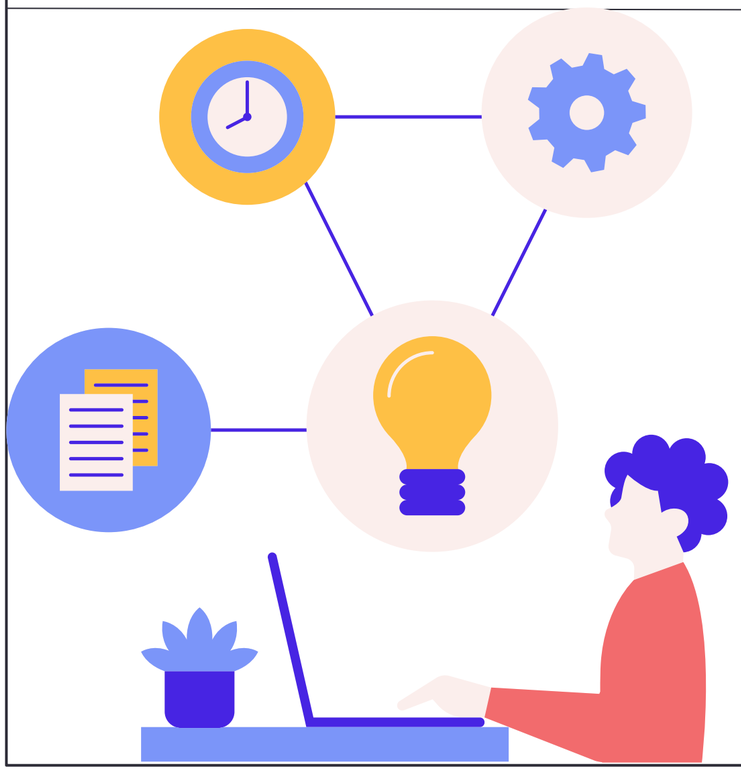
ANOVA F	**Test that checks if a categorical feature creates clearly different severity rates across its groups.**
Gini	**Tree-based score (from Random Forest) that measures how much a feature, numeric or categorical, helps create purer, cleaner splits.**
Gini	**Method that works on both numeric and categorical features by randomly shuffling a column and measuring how much the model's performance drops (bigger drop = more important).**

FEATURE SELECTION RANKING

MASTER FEATURE RANKING — ALL THREE METHODS COMBINED

Rank	Feature	ANOVA F	Gini	Perm	Composite	
1	QUARTIER_RISK	4751.78	0.1940	0.044676	1.67	⭐⭐⭐
2	ARROND_RISK	1812.24	0.3397	0.028562	2.67	⭐⭐⭐
3	QUARTIER	1554.66	0.1280	0.028987	3.67	⭐⭐⭐
4	ARRONDISSEMENT	1397.60	0.1314	0.028400	4.33	⭐⭐
5	MUNICIPALITE	9745.81	0.0515	0.003014	4.67	⭐⭐
6	MUNCP_RISK	689.32	0.0504	0.004175	6.67	⭐⭐
7	HOUR	171.50	0.0495	0.005022	7.67	⭐⭐
8	DAY_OF_WEEK	303.53	0.0227	0.003570	8.00	⭐⭐
9	SPEED_LIMIT_KMH	848.40	0.0180	0.000069	9.00	⭐
10	IS_RUSH_HOUR	250.98	0.0036	0.002883	10.33	⭐
11	IS_WEEKEND	17.89	0.0037	0.002713	11.00	⭐
12	IS_NIGHT	2829.73	0.0076	-0.008306	11.67	⭐
13	MONTH	nan	0.0000	0.000000	nan	⭐
14	QUARTER	nan	0.0000	0.000000	nan	⭐
15	MONTH_SIN	nan	0.0000	0.000000	nan	⭐
16	MONTH_COS	nan	0.0000	0.000000	nan	
17	SEASON	nan	0.0000	0.000000	nan	
18	WEATHER_EN	nan	0.0000	0.000000	nan	
19	SURFACE	nan	0.0000	0.000000	nan	
20	LIGHTING	nan	0.0000	0.000000	nan	

MODEL & VALIDATION



PRIORITIZING RECALL OVER ACCURACY AS OUR MAIN METRICS FOR MODEL EVALUATION.

Accuracy

How often the model is correct overall?

In our dataset, severe crashes are extremely rare, so can overestimate by predicting “non-severe” for nearly every case.

Vs

Recall

How many actual severe collisions the model successfully catches?

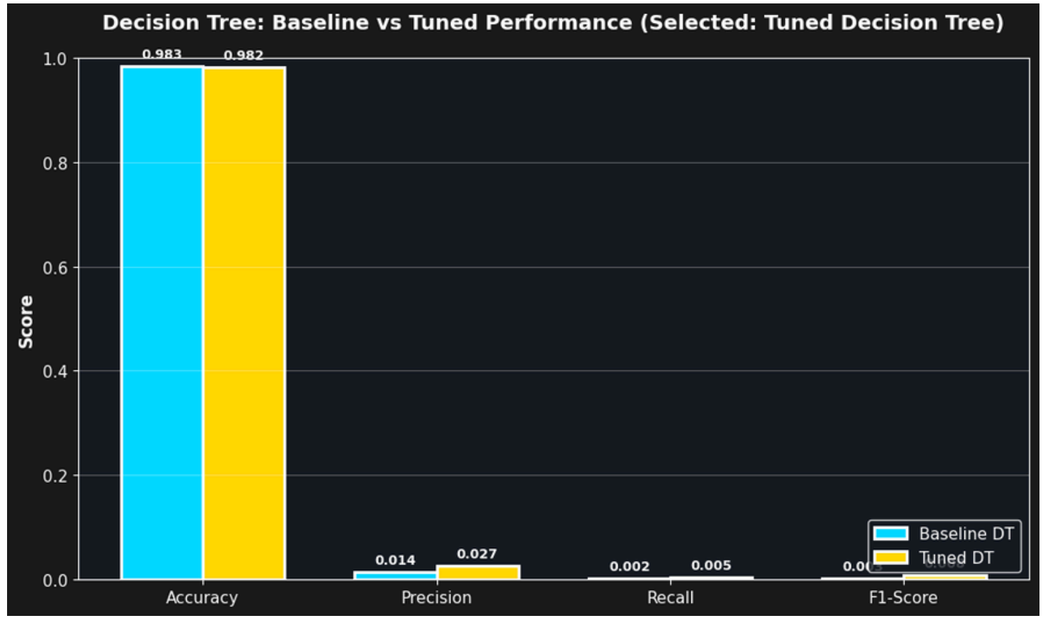
Tells us how well we detect the critical minority class

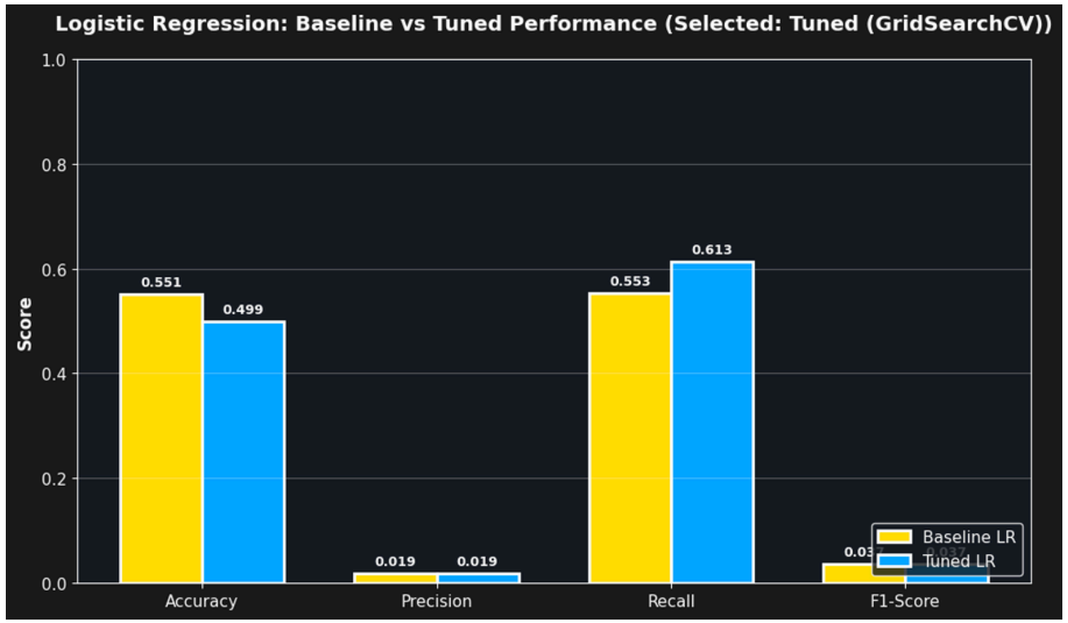
- The dataset is imbalanced, so accuracy looks high even when the model misses severe collisions.
- Missing a severe or fatal collision is far more costly than a false alarm.
- Prioritizing recall ensures we identify high-risk events for prevention and resource planning.

BUILDING ROBUSTNESS THROUGH A BROAD SPECTRUM OF MODELING APPROACHES

Model / Criteria	Logistic Regression	Decision Tree	XGBoost	Random Forest
Rationale	Linear, interpretable baseline	Non-linear rule learner	Variance-reduced ensemble	High-performance booster
Recall	**0.5480**	**0.0228**	**0.0016**	**0.0081**
Pros	Simple, interpretable, fast	Clear rules, flexible	Robust, stable, accurate	Strong accuracy, regularized
Cons	Misses non-linearities	Overfits easily	Less interpretable	Complex, tuning-heavy

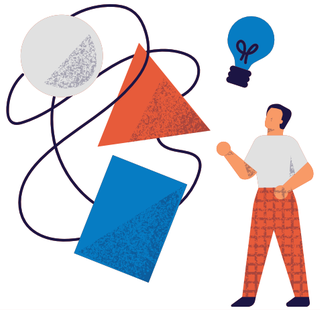
FINAL MODEL SELECTION BASED ON UNTUNED/TUNED METRICS COMPARISONS





TUNED LOGISTIC REGRESSION IS OUR BEST MODEL BECAUSE OF ITS RECALL VALUE

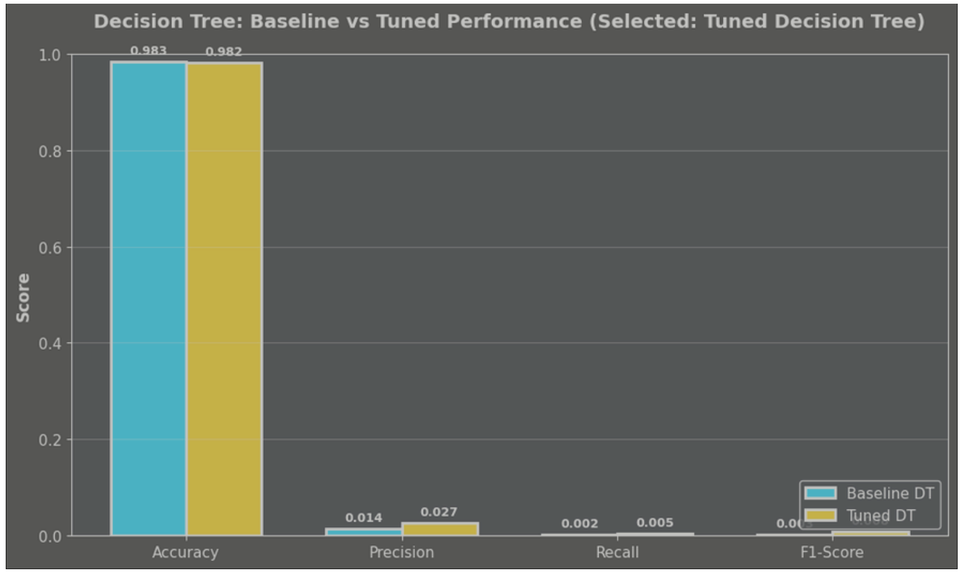
Criteria \ Model	Logistic Regression	Decision Tree
Recall	.613	.005
Precision	0.019	0.0228
ROC-AUC	.61	.52
F1-Score	.0368	.0083

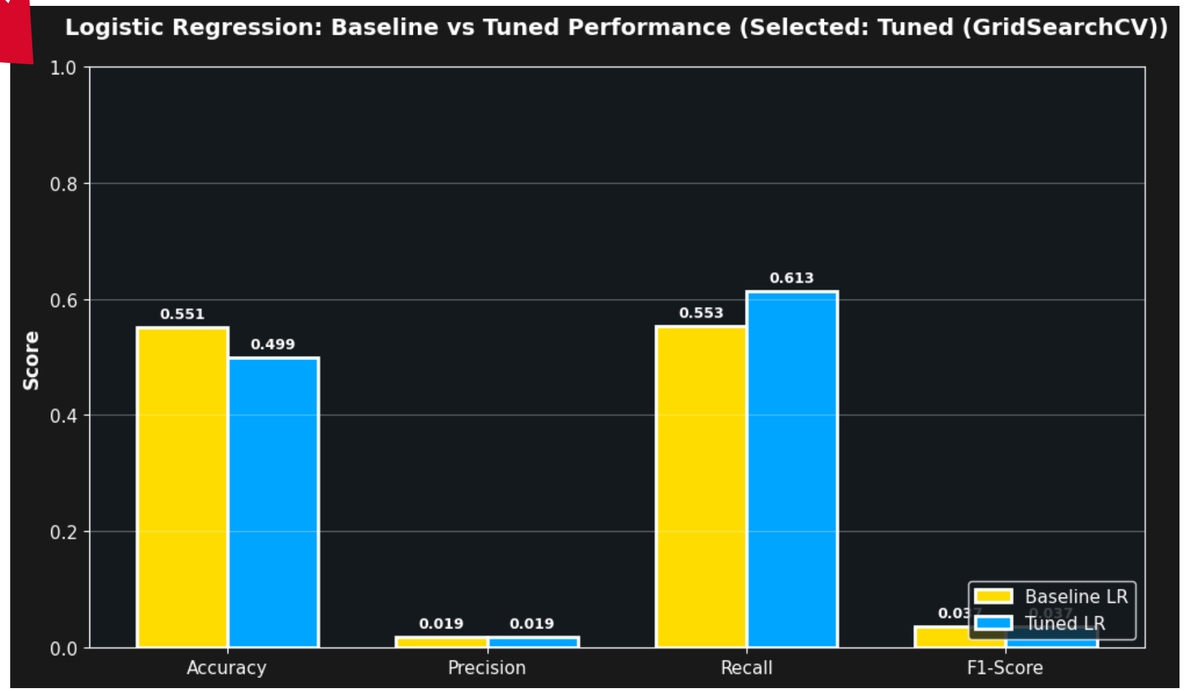


Cross-Validation Summary

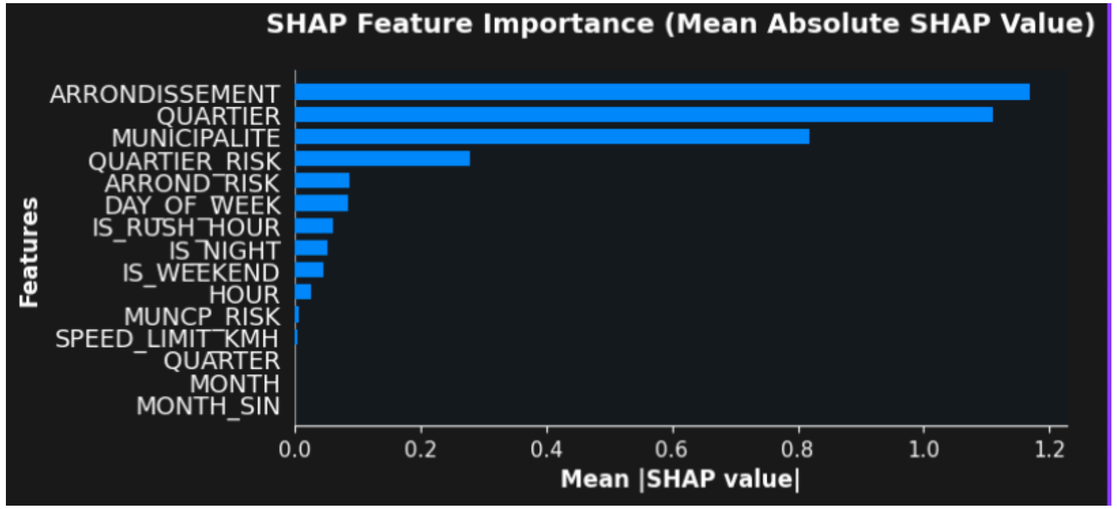
- **Accuracy: 0.5970 ± 0.0023**
- **Precision: 0.5903 ± 0.0024**
- **Recall: 0.6346 ± 0.0028**
- **F1-Score: 0.6116 ± 0.0021**

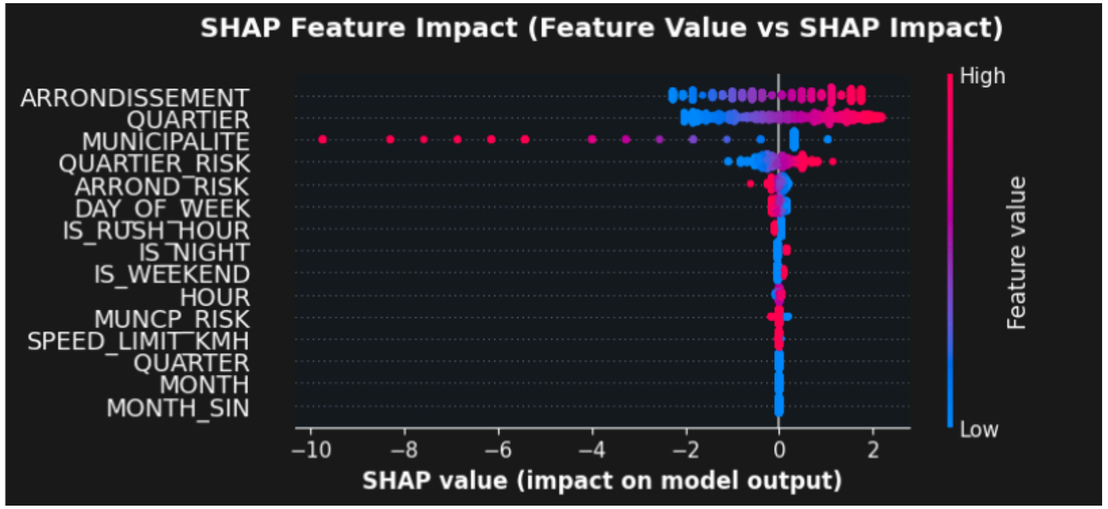
FINAL MODEL SELECTION BASED ON UNTUNED/TUNED METRICS COMPARISONS

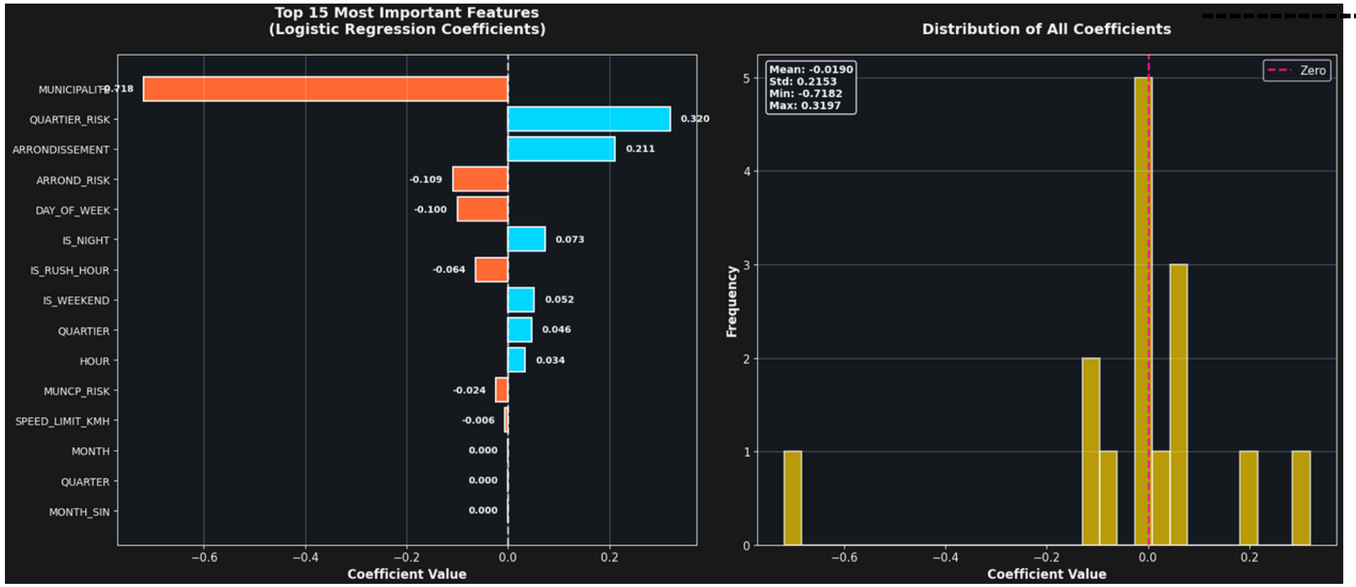




LOCATION-BASED FACTORS AND TIME-BASED FACTORS ROAD DOMINATE ROAD COLLISION SEVERITY PREDICTIONS FOR FATAL OR SEVERE COLLIONS.







Top 10 Most Important Features (by absolute coefficient value):

Feature	Coefficient	Abs_Coefficient
MUNICIPALITE	-0.718211	0.718211
QUARTIER_RISK	0.319687	0.319687
ARRONDISSEMENT	0.211284	0.211284
ARROND_RISK	-0.108891	0.108891
DAY_OF_WEEK	-0.100144	0.100144
IS_NIGHT	0.072602	0.072602
IS_RUSH_HOUR	-0.064035	0.064035
IS_WEEKEND	0.051667	0.051667
QUARTIER	0.046245	0.046245
HOUR	0.033982	0.033982

Our function

$$\text{logit}(p)=\beta 0$$

$$-0.718\text{Municipality}+0.320\text{QuartierRisk}+0.211\text{Arrondissement}-0.109\text{ArrondRisk}-0.100\text{DayOfWeek}+0.073$$

$$\text{Night}-0.064\text{RushHour}+0.052\text{Weekend}+0.046\text{Quartier}+0.034\text{Hour}-0.024\text{MunCpRisk}-0.006\text{SpeedLimit}$$

RESULTS & RECOMMENDATIONS

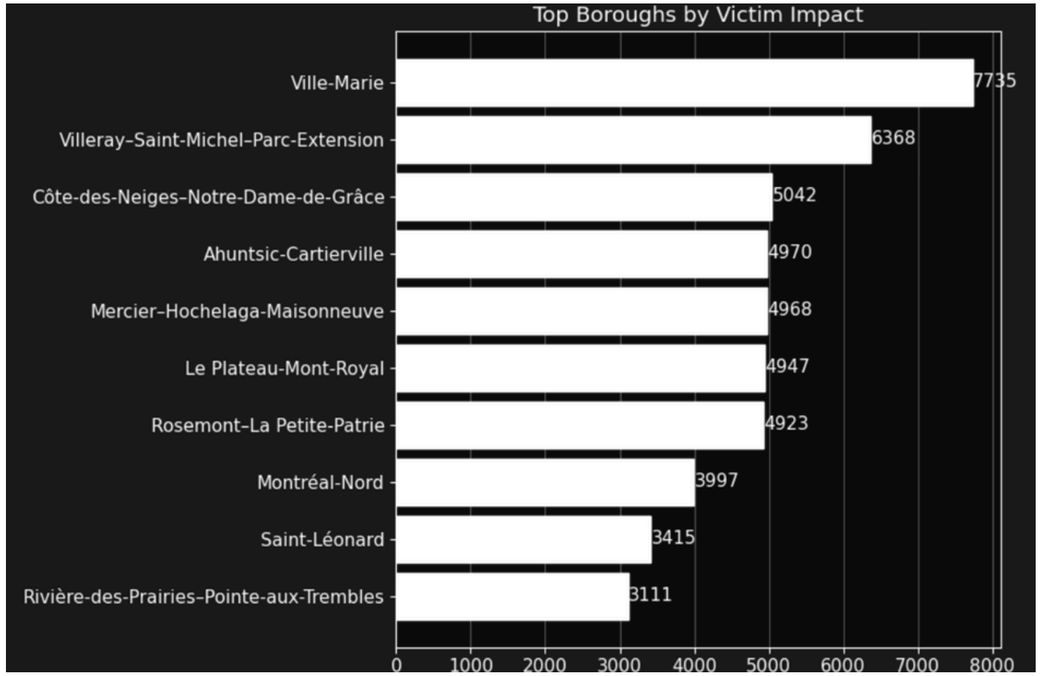
Recommendations for Collision Prevention in Montreal (2025 - 2030)

Key Impact Projections based on our model:

- Currently 21 deaths/year → Potential 25% reduction = **5–10 lives saved per year**
- **$1.7 M – $3.4 M** SAAQ direct payout prevented per year

RECOMMENDATION 1

PRIORITIZE INTERVENTIONS IN CRITICAL NEIGHBORHOODS WHERE SEVERE/FATAL COLLISIONS HAS OCCURENCE HIGHEST.



WHY ?

42% of all severe/fatal collisions occur in just 10 neighborhoods

EXPECTED RESULT

25–40% reduction in city-wide fatalities

Source: Vision Zero cities

RECOMMENDATION 2

CONCENTRATE INTERVENTIONS DURING LETHAL/SEVERE TIME RISK

HOURS :

4pm - 8pm

Highest severe/fatal ratio

10pm - 5am

38% of incidents

WEEKDAY :

MONDAY

High severe/fatal ratio

THURSDAY

High severe/fatal ratio

FRIDAY

High severe/fatal ratio

EXPECTED RESULT :

15–25% fatality reduction

Source: WHO 2018 + SAAQ 2025 Unsafe Driving Behaviors Campaign

RECOMMENDATION 3

PRIORITIZE SPEED ENFORCEMENT ON STREETS WITH SEVERE/LETHAL COLLISIONS RISK

WHAT?

70% of all death from car collision occurs on streets that have **speed limit of 40 and 50 km/hours**

WHY IT MATTERS?

4-5% fewer deaths per 1 km/h slower

Source Speed reduction impact

HOW?

1. MUNICIPALITY ACTIONNABLE STEPS

- **Targeted visibility & education campaigns** during (increasing police presence + ads on the road)
- Pop-up **road-safety checkpoints**

2. EXAMPLE OF STREETS

- Sherbrooke Street
- Jean-Talon Street
- Lacordaire Street
- Wellington Street
- St-Michel Street

EXPECTED RESULT

20-40% in fatality drop

Source : Cochrane 2017 systematic review

RECOMMENDATION 4

PRIORITIZE REAL-TIME ALERTS: INTEGRATE RISK FACTORS (TIME, DAY OF WEEK, SPEED LIMIT, LOCATION) INTO WAZE.

WHY?

- Drivers slow down when warned → proven behaviour change
- **Proven precedent:** Israel Waze + police pilot (2018–2021) → 17% drop in severe injuries

HOW TO DO IT?

1. City of Montréal / MTQ shares weekly anonymized SAAQ crash data
2. Our model re-scores every road segment **in real-time** (cloud pipeline, <2 sec latency)
3. **Severity score > 0.65** → auto-push alert to Waze Partner Program & 511 Québec
4. **Alert radius: 500m** upstream, active only when risky conditions match (Time, weekday, location, etc.)

EXPECTED RESULT

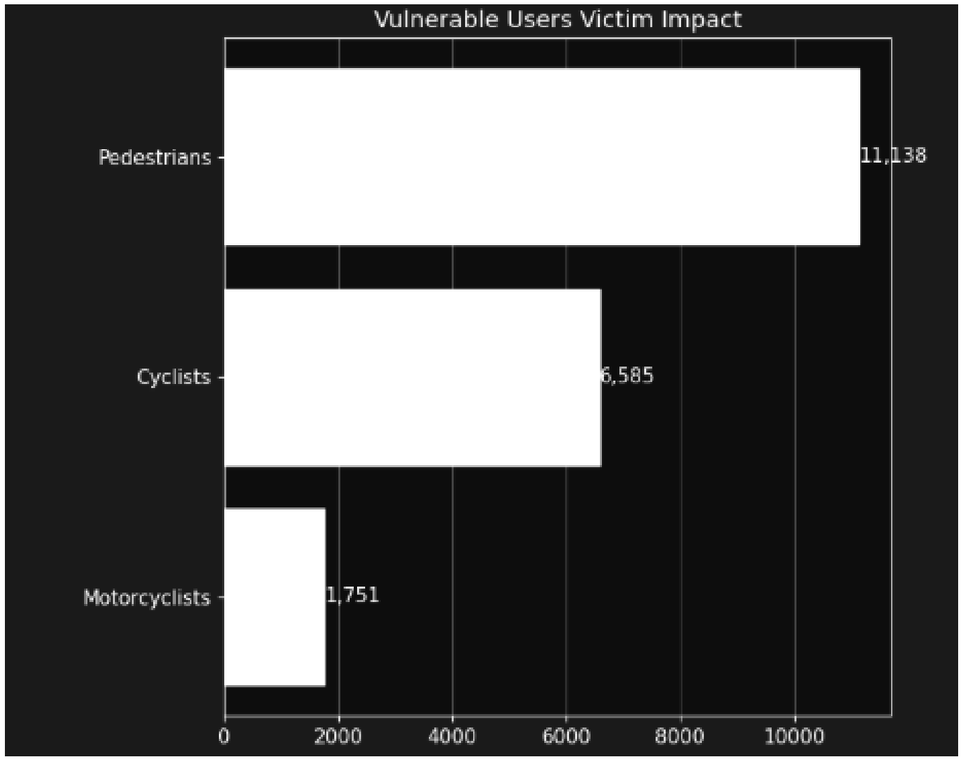
- **12–19%** overall **reduction in severe crashes**
- **≈ 3–5 lives saved per year** in Montréal with only 8% prevention success

Source:

- *Israel Ministry of Transport pilot (2018–2021)*
- *Stockholm Vision Zero dynamic alerts evaluation, Waze for Cities Partner documentation*

RECOMMENDATION 5

IMPLEMENT PROTECTED BIKE LANES AND PEDESTRIAN ZONES IN NEIGHBORHOODS WITH HIGH VULNERABLE USER DEATHS.



WHY ?

70% of fatalities are vulnerable road users

Pedestrians & Cyclists face highest escalation risk

EXPECTED RESULT

10-30% of fatalities/seriously injured cyclists*

With protected infrastructure*

Source: Teschke et al. + Montréal 2025 bike network expansion plan

BIG PICTURE : COMBINING ALL THE RECOMMENDATION TO PREVENT LETHAL and SEVERE COLLISION

⚠ EXTREME SEVERITY RISK EXAMPLE

- René-Lévesque Blvd – Ville-Marie Friday 1:00 AM
- Speed limit 50 km/h Night + Weekend
- **High pedestrian nightlife zone → 6–8× higher chance of fatal/severe crash**

EXPECTED RESULT

Combined alerts + ads + patrols = multiplicative effect (Israel + Stockholm pilots show 22–37% drop of fatal/severe crash when all three are used)

THREE AUTOMATIC ACTIONS TRIGGERED SIMULTANEOUSLY

- **Driver Alert (Waze)**
  - *“Slow down – extreme risk zone ahead – 6× higher fatality risk”*
- **Targeted Safety Ads**
  - Spotify digital billboards on Ste-Catherine & René-Lévesque)
  - 30-second spot launches instantly in a 2 km radius.
- **Police & Peace Officer Dispatch (SPVM + contrôleurs routiers)**
  - Auto-dispatch 2–3 extra patrols to the street
  - Mission:
    - High-visibility presence on sidewalks
    - Guide intoxicated/reckless pedestrians safely to curb or metro •
    - Enforce speed & signalisation for drivers

“It’s 1 AM in Ville-Marie – 70% of tonight’s deaths will be pedestrians or bicycle. Look twice. Save a life.”

LIMITATIONS & NEXT STEPS

LIMITATIONS

- No real-time traffic volume-at-moment data
- No individual victim vulnerability data (helmet, seatbelt, disability)
- No socioeconomic or built-environment variables

MITIGATIONS

Integrate real-time traffic micro- feeds into future models.

Work with SAAQ to access anonymized vulnerability indicators.

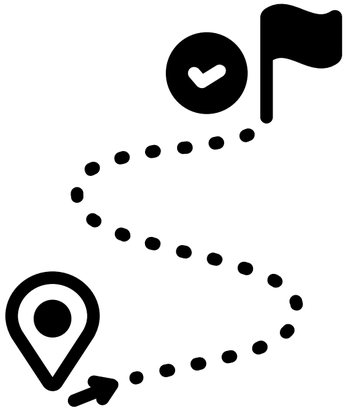
Add census or land-use layers to refine neighborhood-risk modelling.

KPIS

- % of severe-crash drop linked to impaired/offenced in high risk street
- % of Reduction in road users injured on now protected-lane segments.
- Fatality reduction in the 10 prioritized neighborhoods

NEXT STEPS : IMPLEMENTATION TIMELINE

Recommendation 2025-2026 (Pilot)	Primary KPI
Prioritise 2 critical neighbourhoods (Ville-Marie + Plateau)	% reduction in severe and lethal colision associated to that risk
Send 2-3 cops	% reduction in severe and lethal colision associated to that risk
Try Waze Alert	% reduction in severe and lethal colision associated to that risk
During night hours 22:00–05:00)	% reduction in severe and lethal colision associated to that risk
Protected bike lanes & pedestrian zones (if possible)	% reduction in severe and lethal colision associated to that risk



THANKS!

BACK-UP SLIDES INDEX

STATISTICAL METHODS :

- **MODEL TRAINING - BASELINE EVALUATION**

RESULTS & INSIGHTS

- **Day of the Week Graph**
- **Financial Projection Assumptions**

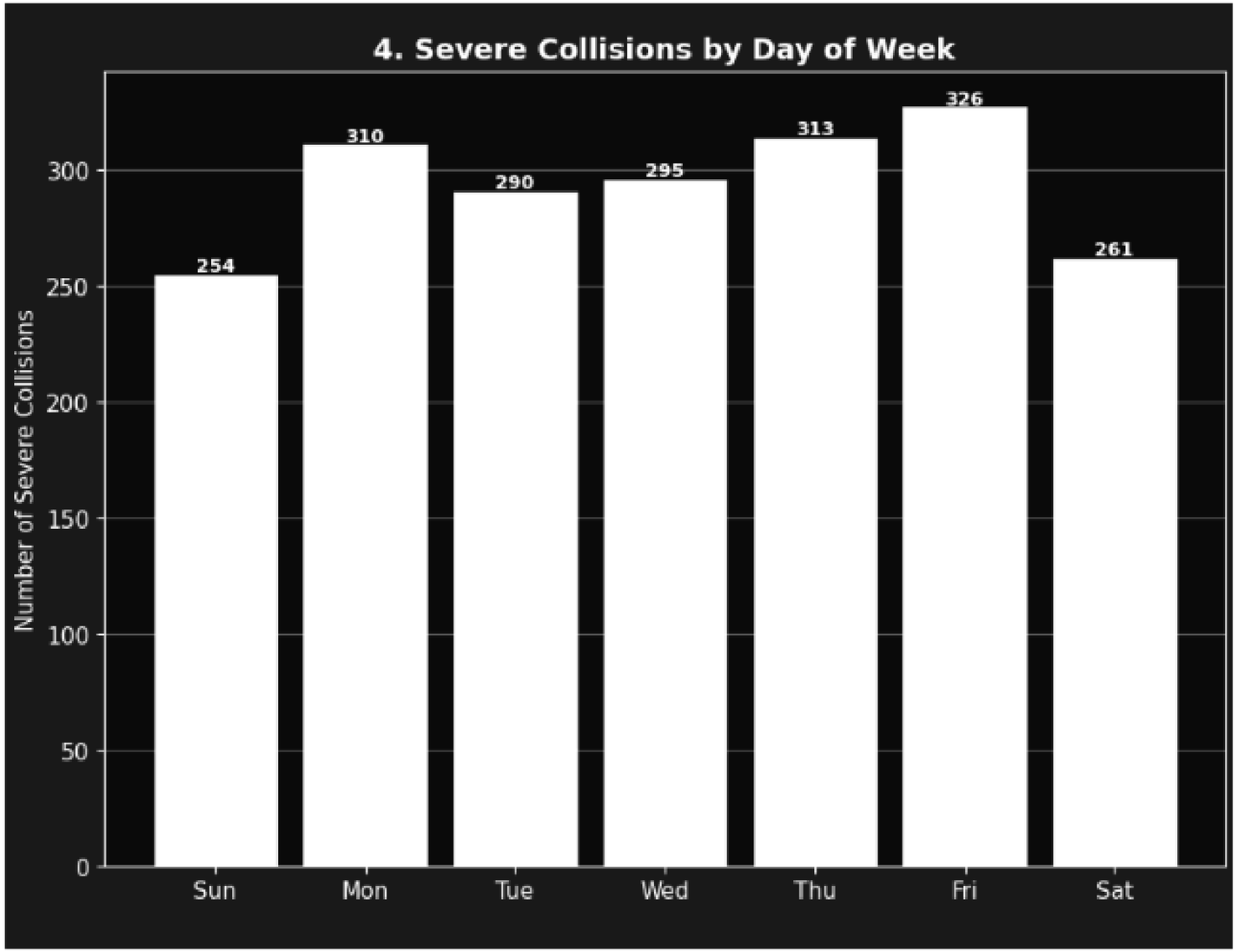
REFERENCES

Société de l'assurance automobile du Québec (SAAQ). (2023). Collisions routières – Ville de Montréal [Data set]. Données Québec. https://www.donneesquebec.ca/recherche/dataset/vmtl-collisions-routieres

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

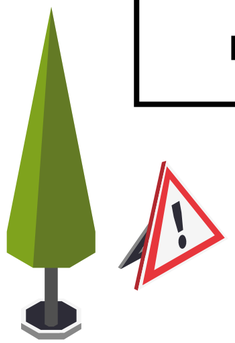
Muktar, B. (2023). Montreal Road Collision Dataset (2012–2021) [Data set]. Kaggle. https://www.kaggle.com/datasets/bmuktar/montreal-road-collision-dataset-2012-2021

DAYS OF THE WEEK HAS AN IMPACT ON SEVERE COLLISIONS



MODEL TRAINING - BASELINE EVALUATION

Decision Criteria	Logistic Regression	Decision Tree	XGBoost	Random Forest
Accuracy	0.5590	0.9843	0.9882	0.9837
Precision	0.0116	0.0318	0.0062	0.0108
Recall	0.5480 ★	0.0228 ★	0.0016	0.0081
F1-score	0.0228	0.0265	0.0026	0.0093
ROC-AUC	0.5743	0.5186	0.5292	0.5297



Financial Impact Projection

Assumption	Detail	Rationale / Source
Baseline fatalities	21 deaths/year in Montreal	SAAQ public statistics
Achievable reduction	25 % (conservative) – 50 % (optimistic) over 6 years with full implementation	Aggregated evidence from recommendations 1–5
Scope	Montréal only (island + agglomerated municipalities in dataset)	Matches dataset coverage
Per-fatality indemnity (2025)	$169,524 – $487,500 (spouse lump-sum) + $8,556 funeral	SAAQ 2025 official tables
Average indemnity used	$337,000 (midpoint)	Reflects typical victim income distribution
Implementation timeline	Recommendations begin 2025, linear scale-up to full effect by 2030	Standard policy rollout assumption
No double-counting	Reductions are additive but capped at 50 % to remain conservative	Avoids over-optimism