A Study on Predictors of Road Traffic Accidents in Metro Manila

CSCI 113i Final Project

Nico Palo, Mireya Reyes, Ja Valdez

Overview of past status report

- Exploring the Kaggle dataset revealed **MCAR and improperly formatted** data entries.
- Supervised learning was chosen as an appropriate machine learning task for the project.
- An outcome variable describing the **overall impact of an accident** was engineered by aggregating the number of lanes blocked, the nature of the accident, and the type and number of vehicles blocked.
- Among Decision Trees, Random Forest Regression, and Logistic Regression, **Decision Trees** was determined to be the most accurate algorithm.

Overview of past status report

Results of Exploratory Data Analysis

Significant Findings

- **Vehicle stalling** and **vehicle collision** are the two common themes in MMDA's reports.
- Most reported accidents occur in **major highways and roads** (i.e. EDSA)
- Road accidents mostly occur in the **morning** and **afternoon**
- Majority of accidents block only **one lane**.

Palo, Reyes, Valdez | CSCI 113i J

Feature Engineering

ImpactLevel_Type	ImpactLevel_Lanes	ImpactLevel_Involved	Final_ImpactLevel
2	1	2	2
2	1	2	2
1	1	2	1
2	1	2	2
2	1	2	2
...	...	...	...
2	1	2	2

IMPACTLEVEL_TYPE

1 - "Stalled", "Accident", "
2 - "Multiple Collision", "Hit and Run"

IMPACTLEVEL_LANES

1 - If 1 or 2 lanes are blocked
2 - If 3 or 4 lanes are blocked

IMPACTLEVEL_INVOLVED

1 - "Motorcycle", "Car", "SUV", "Taxi"
2 - "Truck", "Bus", "Pax", "Ambulance"

Palo, Reyes, Valdez | CSCI 113i J

Results of Data Pipeline

Pre-modeling Dataset Form

Applying all the proposed transformations:

	Latitude	Longitude	Daypart_Afternoon	Daypart_Early Morning	Daypart_Evening	Daypart_Late Evening	Daypart_Morning
0	14.586343	121.061481	0	0	0	0	1
1	14.586343	121.061481	0	0	0	0	1
2	14.589432	121.057243	0	0	0	0	1
3	14.559818	121.040737	0	0	0	0	1
4	14.559818	121.040737	0	0	0	0	1
...	...	...	...	...	...	...	...
22591	14.601442	121.079351	0	0	1	0	0
22592	14.625695	121.048294	0	0	1	0	0
22593	14.625695	121.048294	0	0	1	0	0
22594	14.609768	121.055583	0	0	1	0	0
22595	14.582681	121.076837	0	0	1	0	0

Palo, Reyes, Valdez | CSCI 113i J

Accuracy

Model	Accuracy
Decision Tree	**81.74%**
Random Forest	81.57%
Logistic Regression	81.33%

Palo, Reyes, Valdez | CSCI 113i J

Points of Improvement

Refine pre-processed data

- Improve data quality and extract fundamental elements (*'Location' and 'Involved'*)

Enhance feature engineering

- Strategically utilize other features in dataset (*major highways, vehicles/persons involved*)

Link data to OpenStreetMap

- Initiate deeper analysis of road accident factors through linkage to OSMNX geographical features

Deeper evaluation of Machine Learning Models

- Further training and testing on ML models to determine best fitting model for data

Table of Contents

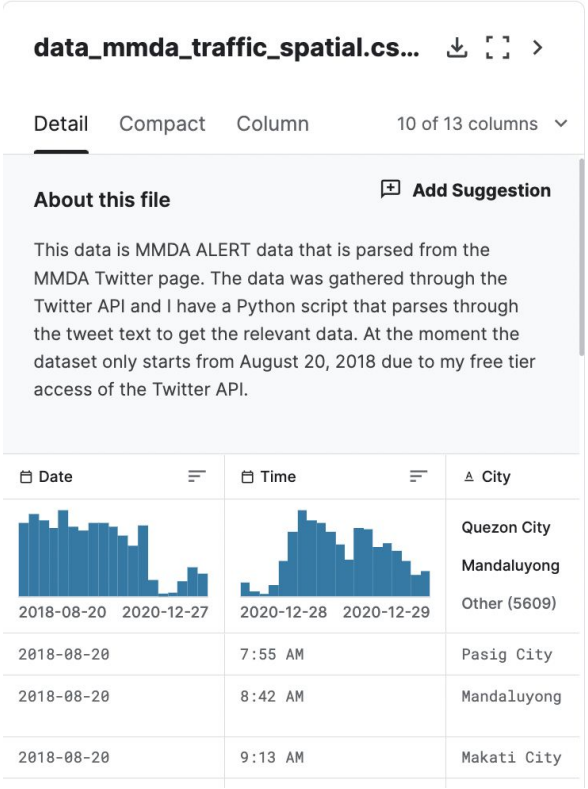
Metro Manila Road Incident Dataset

Sourced from Kaggle

1 Incidents Reported by the Metro Manila Development Authority Twitter Page

2 **<u>Data Size:</u>**

- 13 columns and 17,313 rows



Refining Pre-processed Data

Fixing format in preparation for feature engineering

1 Ñ Values

EDSA MU√ëOZ	ESPA√ëA MACEDA
Para√É¬±aque	OSME√ëA QUIRINO

2 Road Name Format

BONNY SERRANO 20TH AVE.
BONNI SERRANO TUNNEL
B. SERRANO KATIPUNAN

ELLIPTICAL ROAD NORTH AVE.
ELLIPTICAL NORTH AVE

3 Use of Punctuations

MINDANAO AVENUE
C5 LANUZA AVE.
EDSA MAIN AV,

4 Uniformity of Entries

MMDA ALERT: Rallyist at EDSA Camp Aguinaldo Gate 2 NB as of 1:18 PM. Involved Group: "Migrante, Bayan Muna, Anak Pa,Ä¶ https://t.co/4zqjABrBgT	EDSA WHITEPLAINS INVOLVING INVOLVING SUV AND CAR

Integration of OSM Dataset via OSMnx

To enrich the analysis of the study, OSM and its equivalent library (OSMnx) can provide further context on road networks, visualization, and geometries.

Sourced from OpenStreetMap

1 [For Analysis]
Methods used:

- ox.graph_from_place (Road Networks)
- ox.graph_to_gdfs (Convert to GeoDataFrame)
- ox.geometries_from_point, ox.geometries_from_place

2 [Integrated in Modeling]
Data lifted from the OSM:

- Points of Interest
  - Building = True
  - School = True
  - Hospital = True

! **Limitation**

- No explicit tag for Circumferential and Radial (referenced DPWH road data)

Integration of OSM Dataset via OSMnx

Hence, the focus of the OSM integration will be on the (1) **Final Impact Score** and the (2) Road Incidents from the Mode of the City Column—**Quezon City**.

How will the data be used in analysis?

1 **Correlation of Final Impact Score from the POIs**

- Exploring data specifically spatial correlation

2 **Visualization and Geospatial Analysis of Quezon City**

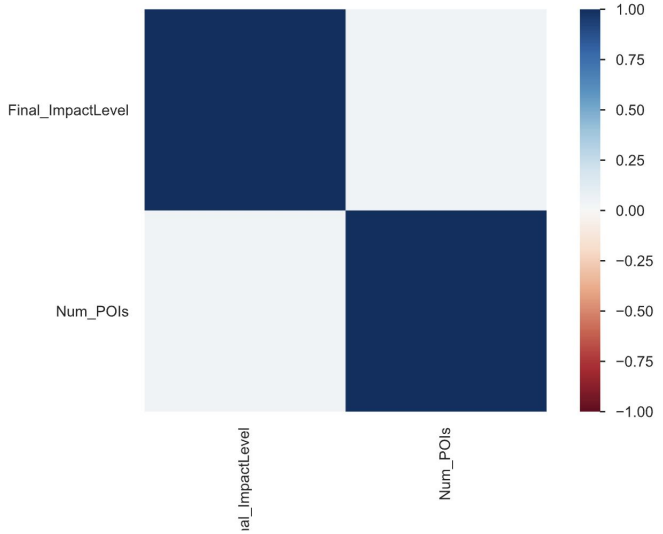
- Exploring geospatial relationships
- Patterns of geographic data (points and polygons)

Integration of POI and Road Incidents

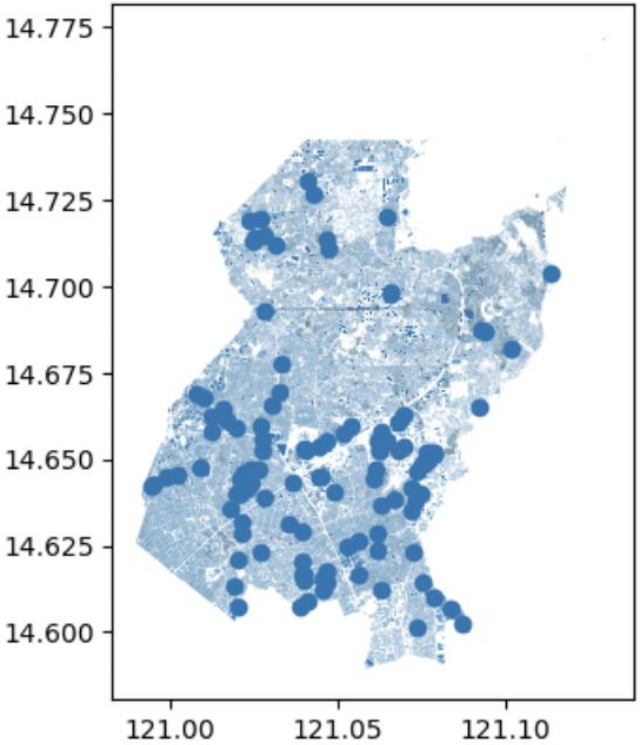
Correlation of Final Impact Score with POIs

- *Very weak positive correlation: closer to 0*

	Num_POIs	Final_ImpactLevel
Num_POIs	1.000000	0.006826
Final_ImpactLevel	0.006826	1.000000



Visualization of Quezon City Geometry



Quezon City

- *Top 1 City for Recorded Road Incidents*
- *Clustering of Geometries between these coordinates (14.675,121.00) to (14.600 to 121.10)*

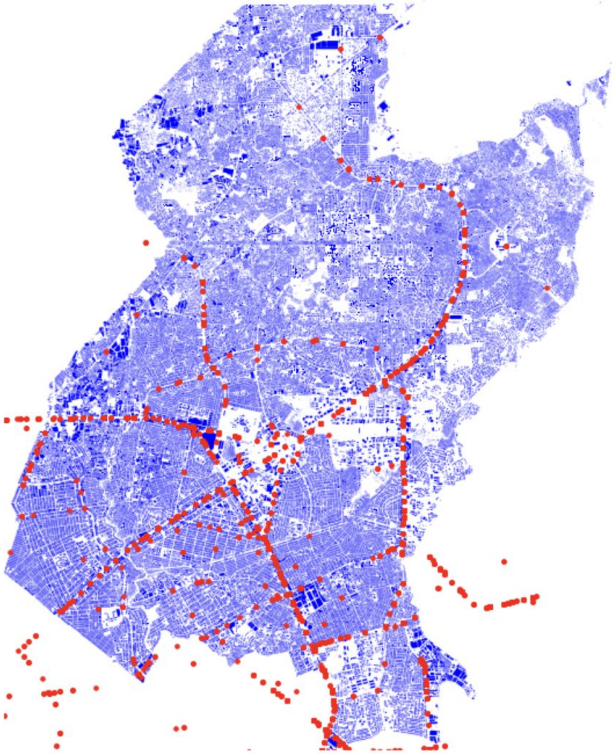
Building

- Points of Interest
  - tags = {"building":True}

geometries_from_place()

- *retrieved the identified buildings in the city*

Visualization of Quezon City Geometry



Thematic Map overlaying road incident data in Quezon City

- *Figure on left: Quezon City—Top 1 City for Recorded Road Incidents*

Buildings

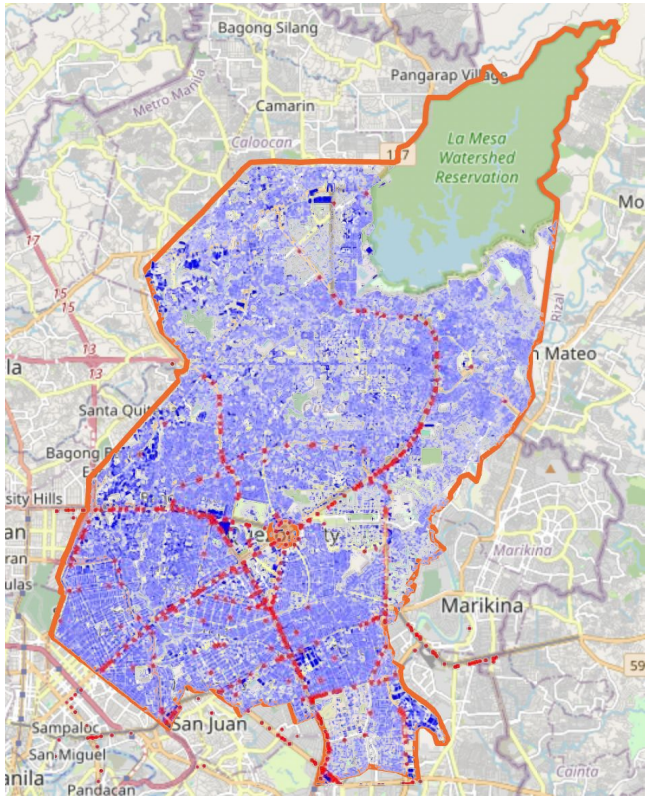
- *represented by the blue markers*

Road Incidents

- *Represented by the red markers*

ox.plot_footprints() and road_incidents_gdf.plot()

Visualization of Quezon City Geometry



Imposed on the Quezon City Map

- *Figure on left: Quezon City—Top 1 City for Recorded Road Incidents*

Quezon City Border

- *Represented by the orange polygon*

Large volume of road incidents along EDSA

- *“Hotspot” for Road Incidents*

Visualization of Quezon City Geometry

One-hot Encoding on Circumferential and Radial Roads involved in Quezon City road incidents

City	Location	Latitude	Longitude	C1 Recto	C2 President	C3 Araneta	C4 EDSA	C5 Katipunar	C6 Southeast	R1: Roxas	R2 Taft	R3 Osmeña	R4 Shaw	R5 Ortigas	R6 Magsaysa	R7 Quezon /	R8 A. Bonifa	R9 Rizal	R10 Del Pan/M
Quezon	EDSA ORTIG.	14.59[illegible]	[illegible]	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Quezon	EDSA ORTIG	14.59[illegible]	[illegible]25	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Quezon	EDSA FARME	14.619717	121.050978	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Quezon	EDSA FARME	14.619717	121.050978	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Quezon	C5 ATENEO I	14.638481	121.07454	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	C5 ATENEO I	14.638481	121.07454	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	C5 ATENEO I	14.638481	121.07454	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	C5 ATENEO I	14.638481	121.07454	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	EDSA BONI	14.638481	121.07454	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Quezon	EDSA ERMIN	14.628035	121.047229	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	EDSA ERMIN	14.628035	121.047229	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	COMMONW	14.659263	121.061261	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	C5 EASTWO	14.60728	121.078528	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	C5 EASTWO	14.60728	121.078528	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	QUEZON AV	14.639754	121.030335	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	QUEZON AV	14.639754	121.030335	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	QUEZON AV	14.639754	121.030335	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	QUEZON AV	14.639754	121.030335	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	EDSA NEPA (	14.627928	121.047259	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Quezon	EDSA NEPA (	14.627928	121.047259	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Quezon	EDSA P. TUA	14.616404	121.052548	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	C5 GREENMI	14.601442	121.079351	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	C5 GREENMI	14.601442	121.079351	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	COMMONW	14.66533	121.070469	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	COMMONW	14.66533	121.070469	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	LUZON F/O	14.665241	121.070339	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Quezon	LUZON F/O	14.665241	121.070339	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Quezon	EDSA P. TUA	14.616404	121.052548	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Quezon	QUEZON AV	14.646323	121.040861	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	QUEZON AV	14.646323	121.040861	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Quezon	B. SERRANO	14.61474	121.070685	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Visualization of Quezon City Geometry

One-hot Encoding on Circumferential and Radial Roads involved in Quezon City road incidents

C1 Recto	52
C2 President Quirino	487
C3 Araneta	280
C4 EDSA	4940
C5 Katipunan /C.P. Garcia	3254
C6 Southeast Metro Manila	249
R1: Roxas	74
R2 Taft	83
R3 Osmeña (formerly South Super)	916
R4 Shaw	72
R5 Ortigas	353
R6 Magsaysay /Aurora	957
R7 Quezon /Commonwealth	2399
R8 A. Bonifacio	263
R9 Rizal	58
R10 Del Pan/Marcos /MacArthur	440

Sum up each column: Top 6 C/R roads

Road Name	Road Incident Reports
C4 EDSA	4940
C5 Katipunan /C.P. Garcia	3254
R7 Quezon /Commonwealth	2399
R6 Magsaysay /Aurora	957
R3 Osmeña (formerly South Super)	916
C2 President Quirino	487

Visualization of Quezon City Geometry

Top 6 C/R Roads in the Quezon City Subset

Road Name	Road Incident Reports
C4 EDSA	4940
C5 Katipunan /C.P. Garcia	3254
R7 Quezon /Commonwealth	2399
R6 Magsaysay/Aurora	957
R3 Osmeña (formerly South Super)	916
C2 President Quirino	487

Mode for locations associated with EDSA

	Location	Count
633	EDSA MAIN AVE	412
751	EDSA SM NORTH	264
733	EDSA SANTOLAN FLYOVER	245
190	C5 GREENMEADOWS	227
729	EDSA SANTOLAN	212

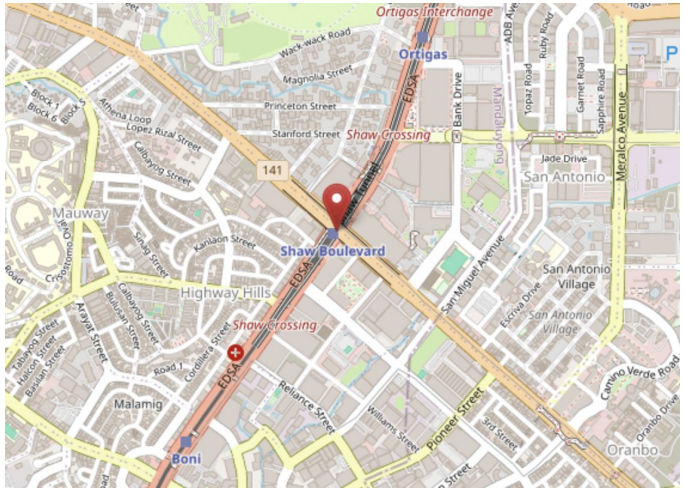
Location	Count	Associated Surfaces
EDSA MAIN AVE	411	Asphalt
EDSA SM NORTH	264	Asphalt
EDSA SANTOLAN FLYOVER	244	Asphalt
C5 GREENMEADOWS	227	Asphalt
EDSA SANTOLAN	212	Asphalt

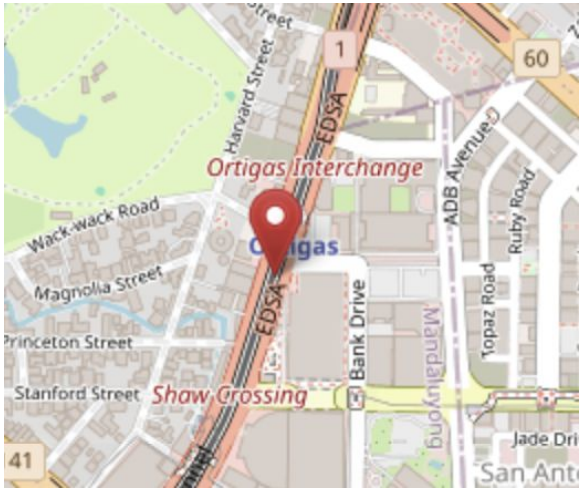
Beyond QC: *Road Incident "Hotspots"*

Areas with Concentrated Incident Reports: EDSA-SHAW, EDSA-ORTIGAS

Modes of the Longitude-Latitude Subset

Longitude	Latitude	Count
121.053565	14.581153	816
121.056314	14.586681	718





Cleaning

To enhance and better inform decision-making for road safety measures—particularly aimed at mitigating the primary conditions that are present in frequent road accidents in Metro Manila—ultimately reducing severe accident frequency

Scaling

To develop a machine learning model that is able to predict accident incidence levels and identify recurring road conditions that frequently lead to accidents

Feature Engineering

Location

Given the scope of the study, two main **road categories** were considered to aggregate the data under “Location” column.

A

Circumferential

traverse the whole or a big portion of the metropolis

B

Radial

area covered is smaller; connects circumferential

Name	Circumferential	Radial
Recto Avenue	TRUE	FALSE
President Quirino Avenue	TRUE	FALSE
Araneta Avenue	TRUE	FALSE
EDSA	TRUE	FALSE
Katipunan Avenue/C.P. Garcia	TRUE	FALSE
Southeast Metro Manila Expressway	TRUE	FALSE
Roxas Boulevard	FALSE	TRUE
Taft Avenue	FALSE	TRUE
Osmeña Highway (formerly South Super Highway)	FALSE	TRUE
Shaw Boulevard	FALSE	TRUE
Ortigas Avenue	FALSE	TRUE
Magsaysay Boulevard/Aurora Boulevard	FALSE	TRUE
Quezon Avenue/Commonwealth Avenue	FALSE	TRUE
A. Bonifacio Avenue	FALSE	TRUE
Rizal Avenue	FALSE	TRUE
Del Pan/Marcos Highway/MacArthur Highway	FALSE	TRUE

Feature Engineering

Location

Given the scope of the study, two main **road categories** were considered to aggregate the data under “Location” column.

1

Classify

Group based on presence of C/R keywords

2

One-Hot Encoding

Binary: True-or-False if the road is C/R

City	Location	Circumferential	Radial
Pasig	ORTIGAS EMERALD	0	1
Pasig	ORTIGAS EMERALD	0	1
Mandaluyong	EDSA GUADIX	1	0
Makati	EDSA ROCKWELL	1	0
Makati	EDSA ROCKWELL	1	0
Mandaluyong	EDSA GUADIX	1	0
San Juan	ORTIGAS CLUB FILIPINO	0	1
Makati	C5 KALAYAAN	1	0
Quezon	EDSA ORTIGAS ROBINSONS	1	1
Quezon	EDSA ORTIGAS ROBINSONS	1	1
Mandaluyong	EDSA LIGHT MALL	1	0
Mandaluyong	EDSA LIGHT MALL	1	0
Quezon	EDSA FARMERS	1	0
Quezon	EDSA FARMERS	1	0
Pasig	C5 LANUZA	1	0
Quezon	C5 ATENEO KATIPUNAN	1	0
Quezon	C5 ATENEO KATIPUNAN	1	0
Quezon	C5 ATENEO KATIPUNAN	1	0
Quezon	C5 ATENEO KATIPUNAN	1	0
Quezon	EDSA BONI	1	1
Marikina	MARCOS HIGHWAY LRT SANTOLAN	0	1
Marikina	MARCOS HIGHWAY LRT SANTOLAN	0	1
Pasay	EDSA HERITAGE	1	0
Pasay	EDSA HERITAGE	1	0
Pasig	C5 ORTIGAS FLYOVER	1	1
Pasig	C5 ORTIGAS FLYOVER	1	1
Quezon	EDSA ERMIN GARCIA	1	0
Quezon	EDSA ERMIN GARCIA	1	0
Quezon	COMMONWEALTH DILIMAN	0	1
Quezon	C5 EASTWOOD	1	0
Quezon	C5 EASTWOOD	1	0
Pasig	MARCOS HIGHWAY LIGAYA	0	1
Pasig	MARCOS HIGHWAY LIGAYA	0	1

Feature Engineering

‘Involved’ Column

k	Involved	
1	TAXI AND MC	I
1	L300	I
1	SUV AND L30	I
1	L300	I
1	2 CARS	I
1	BUS	I
1	CAR AND BU	I
2	3 CARS	I
2	3 CARS	I
1	BUS AND CAI	I
1	TRUCK	I
1	CLOSED VAN	I
1	BUS	I
1	PUJ AND AUV	I
1	TAXI AND TRU	I
1	CLOSED VAN	I
1	CLOSED VAN	I
1	BUS	I
1	BUS	I

1 Classified involved vehicles/people according to type of transportation

Motorcycle	Public	Commercial/Cargo
MOTORCYCLE	PUJ	TRUCK
	BUS	TANKER
	AUV	SWEEPER
	FX	DUMP TRUCK
	JEEP	FORWARD
	TAXI	CRANE
	SERVICE	MIXER
	VAN	TOW TRUCK
	UV EXPRESS	RNS CONTRACTOR
	AUX	

Emergency/Gov't	Non-Motorist *	Private Car
FIRETRUCK	PEDESTRIAN	EVERYTHING ELSE
PERSONNEL	BICYCLES	
MILITARY	BIKE	
CONSTABLE	PEDICAB	
	KALESA	
	RALLYIST	

2 Performed one-hot-encoding

Involved_Motorcycle	Involved_PublicTranspo	Involved_CargoCommercial	Involved_EmergencyGov't	Involved_Non-Motorist	Involved_PrivateCar
0	1	0	0	1	0
0	0	0	0	0	1
0	0	0	0	0	1
0	0	0	0	0	1
0	0	0	0	0	1
0	0	0	0	0	1
0	0	0	0	1	0
0	1	0	0	0	0
0	0	0	0	0	1
0	1	0	0	0	0
0	0	0	0	0	1
0	0	0	0	0	1
0	1	0	0	1	0
0	0	0	0	0	1
0	0	1	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	0	0	1
0	0	0	0	1	0
0	1	0	0	0	0
0	1	0	0	0	0

Road Accident Factors Research

Studies mostly relied on DOH's Online National Electronic Injury Surveillance System (**ONEISS**) and Metro Manila Accident Reporting and Analysis System (**MMARAS**)

Key Findings

1. Fatal accidents usually occur from **6:00PM to 5:00AM** (Evening, Late Evening, and Early Morning)
2. Most at-risk road users are **Motorcycles** (62%) and **Pedestrians** (14%)
3. **Motorcycles** were recorded as the vehicle most involved in TVC
4. Pedestrian fatality risk increases in **multi-lane** roads, **high-volume** roads, high speed areas

(Lu, Herbosa, & Lu, 2022) , (Sigua, Latonero, Kamid, & Avendano, 2023) and (Verzosa & Miles, 2016)

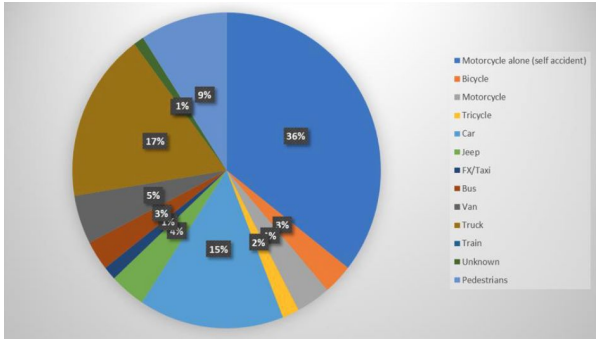


Table 6. Cities/Municipalities and Number of Road Crashes (2019 data)

Cities/Municipality with worst locations

Cities/Municipality	Fatal and Non Fatal Crashes	Rank
Central (Quezon)	6,157	1
Western (Manila)	1,950	2
Eastern (Marikina)	1,450	3

		Cities		
		1	2	3
		Quezon City	Manila City	Marikina City
cations/Roads	1	Commonwealth Ave.	Roxas Blvd	J. P. Rizal St.
	2	EDSA	Radial Road 10	Marcos Highway
	3	C-5 Road	Rizal Ave.	A. Bonifacio Ave.

Feature Engineering

‘Impact’ Column

Evaluates the average impact of a road incident based on the existing features

1 = Low Impact
2 = High Impact

All 4 features were averaged and rounded off to get the “Final_Impact” feature

ImpactLevel_Type	ImpactLevel_Lanes	ImpactLevel_Involved	Impact_Location	Final_ImpactLevel
2	1	2	1	2
2	1	1	1	1
1	1	1	1	1
2	1	1	1	1
2	1	1	1	1
1	1	1	1	1
2	1	2	1	2
2	1	1	1	1
2	1	1	2	2
2	1	1	2	2
2	2	1	1	2
2	2	1	1	2
2	1	2	2	2
2	1	1	2	2
1	1	1	1	1
2	1	1	2	2
2	1	1	2	2

Feature Engineering

‘Impact’ Column

‘Type’

- 1 = “Stalled”, “Accident”
- 2 = “Multiple Collision”, “Hit and Run”

‘Lanes’

- 1 = “1”, “2”
- 2 = “3”, “4”

‘Involved’

- 1 = “Private Car”, “Public Transpo”, “Commercial/Cargo”, “Emergency/Gov’t”
- 2 = “Motorcycle”, “Non-Motorist”

‘Location’

- 1 = Everything not classified as 2
- 2 = “Commonwealth”, “EDSA”, “C5”, “Roxas Blvd”, “Radial Road 10”, “Rizal Ave”, “JP Rizal”, “Marcos Highway”, “A Bonifacio”

‘Final’

- 1 = Low Impact Road Accident
- 2 = High Impact Road Accident

Feature Engineering

Number of POIs

To provide additional context to the roads and the buildings located along these roads, the number of POIs were considered:

**NOTE: Due to the large volume of data, a separate df was created to temporarily store all unique combinations of Longitude and Latitude—to be paired with the sum of POIs near the coordinates*

1

Conversion

Convert current dataframe to geodataframe

2

Exception-handling

Check for invalid values

	Longitude	Latitude	geometry
0	121.06148	14.58634	POINT (121.061 14.586)
2	121.05724	14.58943	POINT (121.057 14.589)
3	121.04074	14.55982	POINT (121.041 14.560)
6	121.04675	14.60185	POINT (121.047 14.602)
7	121.06294	14.55608	POINT (121.063 14.556)

Check for Invalid Values

```
from shapely.geometry import Point

# Assuming gdf is your GeoDataFrame
invalid_points = gdf[~gdf.geometry.apply(Point).is_valid]

print("Invalid points:")
print(invalid_points[['Longitude', 'Latitude']])
```

```
Invalid points:
Empty DataFrame
Columns: [Longitude, Latitude]
Index: []
```

Feature Engineering

*NOTE: Due to the large volume of data, a separate df was created to temporarily store all unique combinations of Longitude and Latitude—to be paired with the sum of POIs near the coordinates

Number of POIs

To provide additional context to the roads and the buildings located along these roads, the number of POIs were considered:

3

Apply enrich_data()

Count the POIs located within 500m

	Longitude	Latitude	geometry	Num_POIs
0	121.06148	14.58634	POINT (121.061 14.586)	247
2	121.05724	14.58943	POINT (121.057 14.589)	296
3	121.04074	14.55982	POINT (121.041 14.560)	2097
6	121.04675	14.60185	POINT (121.047 14.602)	1027
7	121.06294	14.55608	POINT (121.063 14.556)	5188
...	...	...	...	...
229	121.03394	14.66791	POINT (121.034 14.668)	2734
238	121.08106	14.62775	POINT (121.081 14.628)	1828
244	121.03552	14.64292	POINT (121.036 14.643)	899
251	121.05015	14.62156	POINT (121.050 14.622)	2042
253	121.06801	14.66371	POINT (121.068 14.664)	1842

3a

Create Bounding box

Overlays a polygon according to coordinates and distance

***Limits the scope of the POIs affiliated with the location*

3b

Retrieve POI data

Stores all POIs into a variable, to be counted for the Num_POIs

Models Used

1

Decision Trees

Examines features sequentially and arrives at a predicted impact value

2

Random Forest

Takes the most frequent predicted value made among multiple decision trees

3

Logistic Regression

Given the independent observations, it predicts the value of a categorical variable

Preparing the data for machine learning

- Training and test data were split 80/20
- Sci-kit Learn libraries were used to preprocess the data, fit the model, and test the model
- Performance metrics used are accuracy and confusion matrix
- Independent variables: Direction, Daypart, Num_POIs
- Dependent variable: Final_ImpactLevel
- One-hot encoding was performed on Direction, and Daypart features using pd.get_dummies()
- We assume that null values for Num_POIs are equivalent to zero

Accuracy and Confusion Matrix

1

Decision Trees

71.11% accuracy

[2151 384]
[922 1063]

2

Random Forest

71.24% accuracy

[2113 422]
[878 1107]

3

Logistic Regression

57.50% accuracy

[1713 822]
[1099 886]

The significance of our project to mitigating accident incidence

Insights and Recommendations

Essence of R&D in Infrastructure and Urban Planning

The study emphasizes the need for investing in R&D initiatives to enable proper **documentation of procedures and reports**—describing and assessing the current state of the region's roads.

Importance of considering Driver Behavior as a factor affecting road incident occurrence

The study highlights the need to understand the **behavior of drivers** leading up to road incidents—and how these states could affect the incident's impact level.

Insights and Recommendations

Importance of collaboration among government units

Although the MMDA handles traffic incidents, collaboration is an integral part of working towards our common goal. The Department of Public Works and Highways—responsible for infrastructure in the Philippines—**should work hand-in-hand** with enforcement agencies to ensure that road and traffic data are up-to-date, roads are well maintained and most of all, **road users are safe**.

Creating a centralized database for road incidence data

Consolidating data from various sources and government agencies to ensure data accessibility, accuracy, and consistency

Insights and Recommendations

Need for more commuter-centric urban planning and improved road infrastructure for public transportation

Pedestrians, commuters, and motorcyclists face higher road fatality rates despite these being the main modes of transportation for almost **75%** of the Philippine population. This solidifies the need for improved road conditions for non-private vehicle road users and less car-centric urban planning developments.

Most of all: Importance of Decision Making grounded on Data

Data can be further utilized to advance road safety initiatives and provide solid foundation for **strategic planning and intervention.**

Thank you!