

Traffic accident Segmentation and Clustering in the city of Montreal

Edinson Florez

01/04/2021

Introduction

- ▶ Injuries caused by traffic accidents are among the leading causes of death worldwide and they are the main cause of death among young people, approximately 1.3 million people die each year in the world's roads.
- ▶ Because of these enormous impacts, governments are making big efforts to reduce these numbers. Today, because of these efforts, we have systems that provide valuable information on traffic, with governments providing information such as accident hotspots and other useful statistics

Business Problem

- ▶ The objective of this project is to analyze how venues around traffic accidents have an influence on them. This project is mainly focused on geospatial analysis of the traffic accidents in Montreal City with people injured in 2019.
- ▶ The main audience of this project will be the people in charge of optimizing resources in emergency care (tow trucks, firefighters, traffic agents, ambulances), who can concentrate their units in certain areas of interest in the city with high accidents density.
- ▶ Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: Are venues or points of interest (POI) an important feature influencing car accidents in the city of Montreal?

Data description

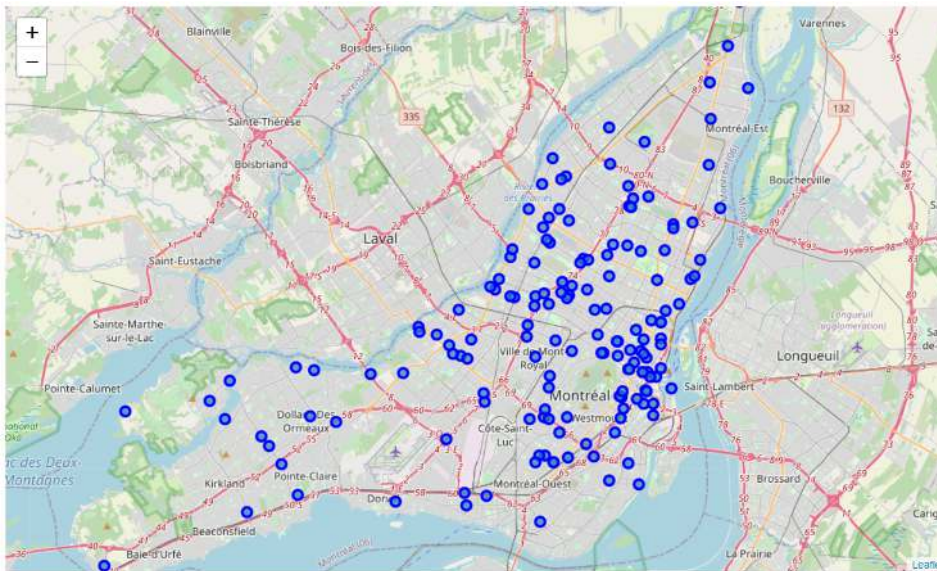
- ▶ To solve the problem, we will need the following data:
- ▶ List of car accidents in the city of Montreal in 2019. This defines the scope of this project, which is limited to the city of Montreal, the biggest city in the province of Quebec and the second-biggest city in Canada.
- ▶ Latitude and longitude coordinates of those accidents. This is required in order to plot the map and to get the venue data.
- ▶ Venue data. I will use this data to perform clustering on the car accidents.

Data Sources

- ▶ The data used in the present study come from Montreal City Open Data Catalog for traffic accidents.
- ▶ The data set has a list of collisions that have occurred in Montreal since 2012. It describes collisions involving at least one motor vehicle traveling on the network and which were the subject of a police report. It includes descriptive, contextual and event location elements, including severity expressed in death, serious injury, minor injury and property damage only.

Methodology

- ▶ To visualize geographic details of accidents in Montreal City, the specialized python folium library was used.



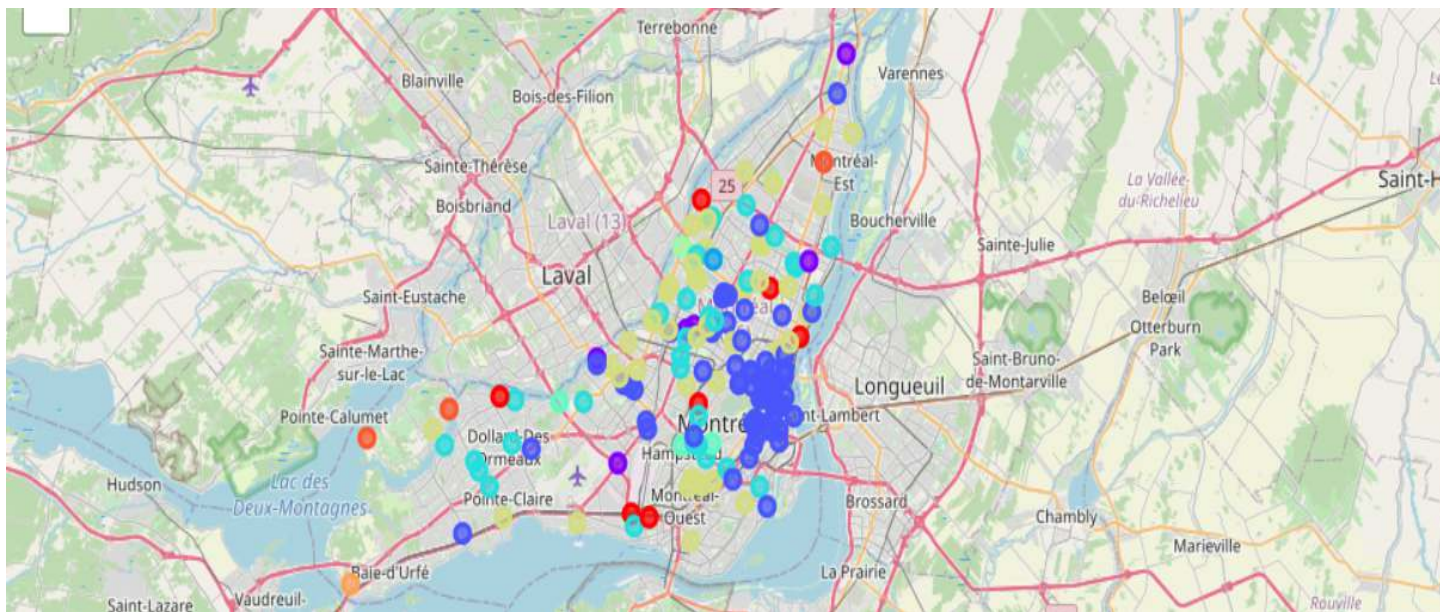
Methodology

- Foursquare API was used to explore the venues around traffic accidents and segment them. I designed the limit as 100 venues and the radius 500 meters for each accident from their given latitude and longitude information.

Collision	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 SPVM_2019_10051	Bakery	Grocery Store	Restaurant	Spanish Restaurant	Mediterranean Restaurant	Sandwich Place	Gas Station	Dog Run	Metro Station	Thai Restaurant
1 SPVM_2019_10193	Coffee Shop	Pharmacy	Fast Food Restaurant	Deli / Bodega	Chinese Restaurant	Italian Restaurant	Bakery	Bank	Gas Station	Discount Store
2 SPVM_2019_10227	Vietnamese Restaurant	Coffee Shop	Chinese Restaurant	Pharmacy	Bank	Discount Store	Thai Restaurant	Breakfast Spot	Multiplex	Food & Drink Shop
3 SPVM_2019_10366	Coffee Shop	Clothing Store	Men's Store	Department Store	Fast Food Restaurant	Sporting Goods Shop	Shopping Mall	Shoe Store	Liquor Store	Electronics Store
4 SPVM_2019_10472	Italian Restaurant	Gym	Vietnamese Restaurant	Rental Car Location	Asian Restaurant	Restaurant	Salon / Barbershop	Grocery Store	Park	Gas Station

Results

- ▶ The following Montreal map shows the clustered accidents after using the K-means algorithm:



Results

After examining all the ten resultant clusters, the features that make each accident cluster different can be grouped and discriminated.

- ▶ Cluster 1: The park cluster.
- ▶ Cluster 2: The cafe cluster.
- ▶ Cluster 3: The Downtown cluster. The largest group of accidents.
- ▶ Cluster 4: One cluster of one.
- ▶ Cluster 5: The commercial cluster. It is a large group of accidents, important.
- ▶ Cluster 6: The construction zone cluster.
- ▶ Cluster 7: Another cluster of one. Not important enough to focus emergency resources around.
- ▶ Cluster 8: The restaurant cluster. It is another large group of accidents in residential zones on the outskirts of the city.
- ▶ Cluster 9: Another cluster of one.
- ▶ Cluster 10: The developing area cluster.

Conclusion

- ▶ Venues around car accidents in the Montreal City seems not to be an important feature influencing car accidents.
- ▶ However, the data analysis allowed to determine three main groups of accidents. The first one caused by the high density of population and vehicles in the city downtown. The second group dispersed in the city and caused by proximity to highways and the third group, in residential areas on the outskirts of the city with very varied venues.

