2021

Traffic accident Segmentation and Clustering in the city of Montreal

Florez, Edinson

01/04/2021

Contents

1. Introduction

1.1 Background

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 [1]. Specifically in Quebec Province in Canada, 333 people were killed and 1,334 people were seriously injured in rad accidents in 2019 [2]. In addition to the tragic impact on human lives, road traffic injuries have a significant effect on the world's economy. One study estimates that road traffic injuries will cost the world economy US$1.8 trillion in the period 2015-30 [3].

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.

1.2 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?

2. Data

2.1 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.

2.2 Data Sources

The data used in the present study come from Montreal City Open Data Catalog for traffic accidents. [4]

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.

2.3 Data Cleaning

Original dataset is a 190553 rows x 68 columns size. To simplify the study only accidents for 2019 were considered. Additionally, only accidents were seriously injured or killed people were kept.

The final result is a dataframe containing 160 rows x 6 columns as shown below:

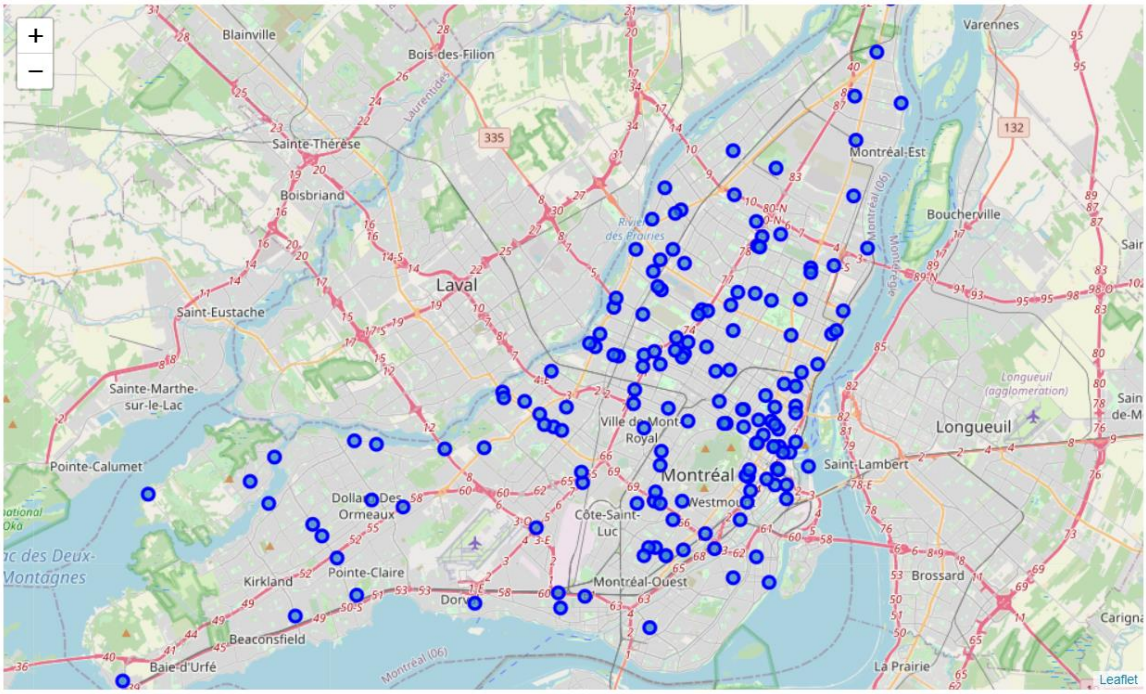
	index	NO_SEQ_COLL	NB_MORTS	NB_BLESSES_GRAVES	latitude	longitude
0	129	SPVM _ 2019 _ 130	0	1	45.48505	-73.86485
1	165	SPVM _ 2019 _ 166	1	0	45.50952	-73.80060
2	210	SPVM _ 2019 _ 211	0	3	45.48921	-73.93608
3	246	SPVM _ 2019 _ 247	0	1	45.50787	-73.76006
4	287	SPVM _ 2019 _ 288	0	1	45.49430	-73.87576
...	...	...	...	...	...	...
155	18718	SPVM _ 2019 _ 18719	0	1	45.61275	-73.51807
156	18735	SPVM _ 2019 _ 18736	1	0	45.67214	-73.50408
157	18936	SPVM _ 2019 _ 18937	0	1	45.65120	-73.48977
158	19122	SPVM _ 2019 _ 19123	0	1	45.43823	-73.84873
159	19261	SPVM _ 2019 _ 19262	0	1	45.41145	-73.95116

160 rows × 6 columns

3. Methodology

The database and support documents used in the present study were hosted in GitHub.

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



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.

For example, for the first accident in 2019, the most common venues classed by categories are:

	name	categories	lat	lng
0	La Roulotte	Diner	45.482877	-73.866884
1	Harvey's	Restaurant	45.483984	-73.860282
2	Club Piscine	Furniture / Home Store	45.482992	-73.860296
3	Second Cup	Coffee Shop	45.484049	-73.860264
4	Bulk Barn	Food & Drink Shop	45.484398	-73.858569

The same process was used to analyze all the accidents in Montreal City in 2019 with the following results:

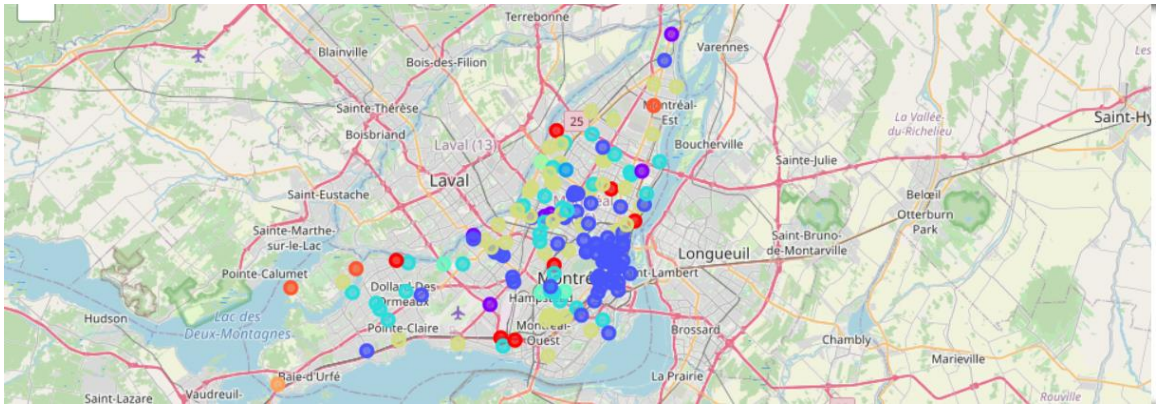
	Collision	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	SPVM _ 2019 _ 10051	Bakery	Grocery Store	Restaurant	Spanish Restaurant	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

Unsupervised learning K-means algorithm to cluster traffic accidents was used. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

4. Results

Using the K-means algorithm and setting the number of clusters to 10. All accidents get segregated according to the surrounding venues.

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



After examining all the ten resultant clusters, the features that make each accident cluster different can be grouped and discriminated. It has been concluded that each accident possesses a certain type of venues and spots that make them different from the others in the way they are configured and the venues they have.

By looking at each cluster's characteristics, one can easily know what are the city areas and the surrounding venues influencing car accidents in Montreal.

The following list contains the features that characterize each cluster, and also a name which they can be classified:

Cluster 1: The park cluster. It is not a very large group of accidents are located in a low population density zone with parks and small restaurants around, widely dispersed within the city, is not important enough to focus emergency resources around.

Cluster 2: The cafe cluster. This is another not very large group of accidents. Here the most common places to find are coffees and restaurants. Not important enough to focus emergency resources around.

Cluster 3: The Downtown cluster. The largest cluster. Very commercial venues in this cluster like hotels, gas stations, stores, restaurants located in downtown city. Accidents here seem to be caused more by the high population density than surrounding venues. This group of accidents already has a significant number of accident care resources such as fire stations, police and medical centers.

Cluster 4: One cluster of one. Not important enough to focus emergency resources around.

Cluster 5: The commercial cluster. It is a large group of accidents. Unlike cluster 3, it contains venues outside the city downtown. The most common venues to find in this cluster are restaurants, stores, pharmacies.

Cluster 6: The construction zone cluster. It's a group of only 3 accidents in a construction zone. Not important enough to focus emergency resources around.

Cluster 7: Another cluster of one. Not important enough to focus emergency resources around.

Cluster 8: The restaurant cluster. It is a large group of accidents in residential zones residential areas on the outskirts of the city. The most common venues in this cluster are restaurants, shopping malls, bakeries, stores. It's an important group, generally in neighborhoods with accident care services like police and fire stations. This is the group to follow by the responsible for accidents care in the city.

Cluster 9: Another cluster of one. Accident happened in industrial zone. Not important enough to focus emergency resources around.

Cluster 10: The developing area cluster. It is a small group of accidents in not very high population density and in developing the city area. Not important enough to focus emergency resources around.

5. Discussion

160 car accidents with seriously injured or killed people in the city of Montreal (Canada) were analyzed and clustered according to venues surrounding each accident.

The main audience of this project was the people in charge of city resources in emergency care and the objective was to determine if venues surrounding car accidents have an impact on it and help to optimize the resources according to venues zones in the city.

According to the results, three main groups of accidents were identified. The first group is geographically concentrated in the downtown city area, venues are very varied including hotels, gas stations, stores and restaurants. Accidents here seem to be caused more by the high population density than surrounding venues. This group of accidents already has a significant number of accident care resources such as fire stations, police and medical centers.

The second group is the commercial cluster. It is a large group of accidents. Unlike cluster 3, it contains venues outside city downtown and scattered in the city. Venues around this group of accidents are very varied like restaurants, stores, pharmacies, gas stations. This group seems more impacted by its proximity to highways, boulevards or main streets than venues around accidents. Special groups, , such as tow trucks, firefighters, traffic agents and ambulances, already serve this kind of high traffic ways in emergency care.

The third main group is composed of accidents in residential zones on the outskirts of the city. Venues are very varied like small restaurants, banks, parks and gyms. Generally, these neighborhoods have already accident care services like police and fire stations.

6. 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.

7. References

[1] Global status report on road safety https://apps.who.int/iris/bitstream/handle/10665/44122/9789241563840_eng.pdf;jsessionid=416EA929042797C4C3A47DA5798D6E28?sequence=1

[2] https://saaq.gouv.qc.ca/en/saaq/documents/road-safety-record/

[3] Chen, S., Kuhn, M., Prettner, K., & Bloom, D. E. (2019). The global macroeconomic burden of road injuries: estimates and projections for 166 countries. The Lancet Planetary Health, 3(9), e390-e398.

[4] https://donnees.montreal.ca/