

## Introduction/Business Problem

When a municipality wants to know and better serve its citizens, there are many ways to do so. One of them is a survey, which involves going door to door to meet people and asking them a few prepared questions to better understand their feelings about safety, services offered by the municipality or surrounding businesses. Following this type of survey, it is necessary to define which categories of people feel the most left out in order to better target the services to be provided to the greatest number of people, and to adopt a policy accordingly.

As such, the City of Montreal conducted this type of survey in 2016. In this survey, more than 7,000 people were questioned and we are especially interested in their satisfaction with the businesses and services offered in their borough.

The final objective will be to have a recommendation of new services and businesses to develop for the groups who feel the most left out by borough.

## Data

To proceed with this analysis, we will need to retrieve the data from

- the survey conducted by the City of Montreal (<https://donnees.montreal.ca/ville-de-montreal/sondage-satisfaction-citoyens>). These data will allow us to create several people clusters based on their profile (age category, number of years of study, mother tongue, etc.), their neighbourhood of residence and their level of satisfaction with the services and businesses in their postalcode
- the API foursquare. We retrieved the list of services, stores and leisure activities around the center of the postalcode.

The survey data had to be cleaned up a bit in order to eliminate respondents whose postal code is not assigned to the City of Montreal. It was also chosen to remove respondents who did not want to answer a question related to their profile or to the question concerning their level of satisfaction with the services offered and stores in their neighbourhood. After all this cleaning, we reduced the sample from 7614 to 6871 respondents.

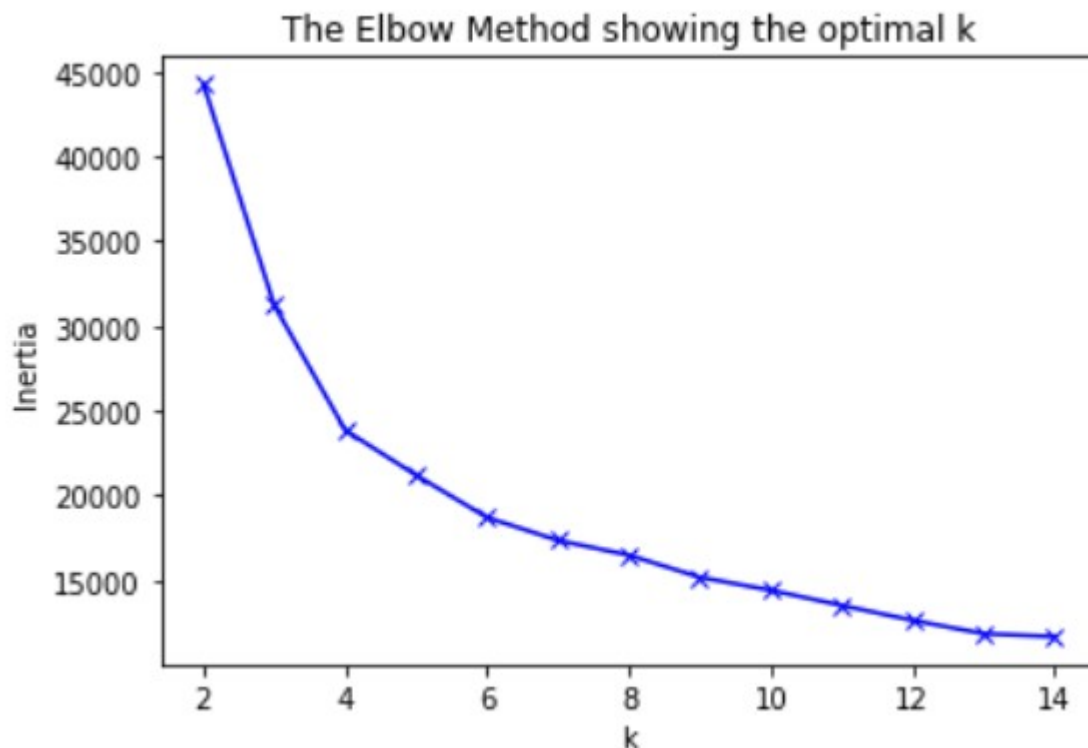
	PostalCode	age_group	household_type	owner_tenant	education_level	occupation	sex	satisfaction_result
0	H1K	3	2	1	4	1	1	3
1	H2M	6	4	1	2	3	2	1
2	H2V	2	5	2	2	1	1	2
3	H1K	4	2	2	4	1	1	1
5	H1E	6	2	1	2	2	1	1

# Methodology

The methodology consists of three main steps:

1. Working with the survey information and grouping respondents by affinity
2. Retrieve venues from the Foursquare API for each postal code
3. Put together the two previous steps, and draw out the list of areas to be developed to improve the level of satisfaction of the group that feels the least satisfied

The first step is to group respondents into different clusters based on characteristics: age category, level of education, household type, occupation, sex, owner or tenant). For this, the k-means algorithm was chosen because it is adequate for an unsupervised clustering approach. The elbow method was followed to determine that the optimal number of clusters is 4.



Following this optimal choice of k-means, each respondent was assigned to a cluster.

Initially, we had about 6800 respondents from Montreal spread over about 100 different geographic areas with a distinct postal code.

The distribution of the 6800 respondents was not equitable by postal code and group. As a result, a satisfaction survey value could have a lot of value if the respondent was the only one in his cluster for a postal code.

In order to avoid these pitfalls of non-representative sampling, geographic areas that were under-represented had to be removed. We used the median number of respondents per postal code and the median number of respondents from the same cluster per postal code to define the thresholds. Next, we removed all respondents who were in a geographic area that did not meet these threshold values.

cluster_labels		nb	
count	103.000000	count	367.000000
mean	66.708738	mean	18.722071
std	70.196731	std	20.255718
min	1.000000	min	1.000000
25%	18.000000	25%	5.000000
50%	53.000000	50%	12.000000
75%	93.500000	75%	26.000000
max	424.000000	max	136.000000

Initially, we planned to provide recommendations for new services and businesses to develop for the 100 geographical areas, but after filtering out the too small sample of respondents, there are only 25 geographical areas and 3900 respondents left.

The second step was to retrieve the set of arrivals from each geographical area within a radius of 3km. Of all these visits, we only kept the 10 most common types of visits within this geographical perimeter.

	PostalCode	MostCommonVenue1	MostCommonVenue2	MostCommonVenue3	MostCommonVenue4	MostCommonVenue5	MostCommonVenue6
0	H1A	Restaurant	Italian Restaurant	Grocery Store	Gas Station	Convenience Store	
1	H1B	Restaurant	Gas Station	Coffee Shop	Pharmacy	Supermarket	
2	H1C	Pharmacy	Grocery Store	Convenience Store	Liquor Store	Pizza Place	Fast Food Restaurant
3	H1E	Pharmacy	Coffee Shop	Grocery Store	Italian Restaurant	Liquor Store	
4	H1G	Pharmacy	Fast Food Restaurant	Supermarket	Gas Station	Coffee Shop	
...	...	...	...	...	...	...	...
116	H9P	Coffee Shop	Restaurant	Fast Food Restaurant	Sushi Restaurant	Breakfast Spot	
117	H9R	Coffee Shop	Italian Restaurant	Liquor Store	Fast Food Restaurant	Pharmacy	
118	H9S	Bank	Coffee Shop	Liquor Store	Ice Cream Shop	Shopping Mall	
119	H9W	Coffee Shop	Italian Restaurant	Gas Station	Fast Food Restaurant	Pharmacy	
120	H9X	Bar	Sandwich Place	Convenience Store	Café	Burger Joint	

121 rows × 11 columns

The third and final step is to pool the information gleaned from the survey, the defined respondent groups, and the venues that emerged from the Foursquare API. On the one hand, we isolated the group with the lowest average level of satisfaction for each postal code.

PostalCode	cluster_labels	min_avg_satisf
H1A	2	1.750000
H1E	0	1.555556
H1G	2	1.428571
H1H	2	1.600000
H1K	2	1.527273
H1R	0	1.500000
H1S	0	1.440000
H2E	2	1.416667
H2K	0	1.733333
H2L	0	1.883333
H2V	3	1.562500
H3E	0	1.500000

On the other hand, we determined the 5 geographical areas in each group that showed the highest level of satisfaction. Here an example of the postal code where the cluster 3 is the most satisfied.

	PostalCode	cluster_labels	max_avg_satisf
<b>67</b>	H4E	3	2.362069
<b>19</b>	H1K	3	1.970588
<b>3</b>	H1A	3	1.968750
<b>47</b>	H3E	3	1.846154
<b>103</b>	H9C	3	1.715789

For each cluster, from these 5 geographical areas, we have drawn up a list of the most common types of arrivals.

	cluster_labels	suggested_venues
0	0	[Ice Cream Shop, Clothing Store, Bagel Shop, S...
1	1	[Ice Cream Shop, Supermarket, Coffee Shop, Del...
2	2	[Ice Cream Shop, Supermarket, Coffee Shop, Ita...
3	3	[Korean Restaurant, Coffee Shop, Harbor / Mari...

This would be the list of recommended visits to improve the satisfaction of the least satisfied group, from which we would deduct the 10 most frequent visits for the geographic area in question.

## Results

As a result, we are able to recommend a list of new services and businesses to develop in each geographical area (i.e. PostalCode), in order to improve the satisfaction of the least satisfied cluster.

PostalCode	Recommended venues to develop
H1A	['Ice Cream Shop', 'Supermarket', 'Coffee Shop', 'Sushi Restaurant', 'Dessert Shop', 'Fast Food Restaurant', 'Automotive Shop', 'Gym', 'Brewery', 'Indian Restaurant', 'Bar', 'French Restaurant', 'Café', 'Breakfast Spot', 'Sandwich Place', 'Tea Room', 'Bakery', 'Portuguese Restaurant', 'Concert Hall', 'Pizza Place', 'Greek Restaurant', 'Hotel', 'Burger Joint', 'Bank']
H1G	['Ice Cream Shop', 'Italian Restaurant', 'Sushi Restaurant', 'Dessert Shop', 'Automotive Shop', 'Gym', 'Brewery', 'Indian Restaurant', 'Bar', 'French Restaurant', 'Café', 'Tea Room', 'Bakery', 'Portuguese Restaurant', 'Concert Hall', 'Pizza Place', 'Park', 'Greek Restaurant', 'Hotel', 'Burger Joint', 'Bank']
H1H	['Italian Restaurant', 'Sushi Restaurant', 'Dessert Shop', 'Automotive Shop', 'Gym', 'Brewery', 'Indian Restaurant', 'Bar', 'French Restaurant', 'Café', 'Tea Room', 'Bakery', 'Portuguese Restaurant', 'Concert Hall', 'Pizza Place', 'Greek Restaurant', 'Restaurant', 'Hotel', 'Burger Joint', 'Bank', 'Grocery Store']
H1K	['Ice Cream Shop', 'Supermarket', 'Sushi Restaurant', 'Dessert Shop', 'Automotive Shop', 'Gym', 'Brewery', 'Indian Restaurant', 'Bar', 'French Restaurant', 'Café', 'Tea Room', 'Portuguese Restaurant', 'Concert Hall', 'Pizza Place', 'Park', 'Greek Restaurant', 'Gas Station', 'Hotel', 'Burger Joint', 'Bank']
H2E	['Ice Cream Shop', 'Supermarket', 'Sushi Restaurant', 'Dessert Shop', 'Fast Food Restaurant', 'Automotive Shop', 'Gym', 'Brewery', 'Indian Restaurant', 'Bar', 'French Restaurant', 'Sandwich Place', 'Tea Room', 'Portuguese Restaurant', 'Concert Hall', 'Pizza Place', 'Park', 'Greek Restaurant', 'Restaurant', 'Gas Station', 'Hotel', 'Burger Joint', 'Bank', 'Grocery Store']
H3H	['Coffee Shop', 'Italian Restaurant', 'Dessert Shop', 'Fast Food Restaurant', 'Automotive Shop', 'Gym', 'Brewery', 'Indian Restaurant', 'Bar', 'French Restaurant', 'Breakfast Spot', 'Sandwich Place', 'Tea Room', 'Portuguese Restaurant', 'Concert Hall', 'Greek Restaurant', 'Restaurant', 'Gas Station', 'Pharmacy', 'Hotel', 'Burger Joint', 'Bank']
H4E	['Ice Cream Shop', 'Supermarket', 'Coffee Shop', 'Italian Restaurant', 'Sushi Restaurant', 'Dessert Shop', 'Fast Food Restaurant', 'Automotive Shop', 'Indian Restaurant', 'Bar', 'Sandwich Place', 'Tea Room', 'Portuguese Restaurant', 'Concert Hall', 'Greek Restaurant', 'Gas Station', 'Pharmacy', 'Hotel', 'Burger Joint', 'Bank', 'Grocery Store']
H4G	['Ice Cream Shop', 'Supermarket', 'Coffee Shop', 'Italian Restaurant', 'Sushi Restaurant', 'Dessert Shop', 'Fast Food Restaurant', 'Automotive Shop', 'Gym', 'Indian Restaurant', 'Breakfast Spot', 'Sandwich Place', 'Tea Room', 'Bakery', 'Portuguese Restaurant', 'Concert Hall', 'Greek Restaurant', 'Gas Station', 'Pharmacy', 'Hotel', 'Burger Joint', 'Bank']

H8N	['Ice Cream Shop', 'Italian Restaurant', 'Sushi Restaurant', 'Dessert Shop', 'Automotive Shop', 'Gym', 'Brewery', 'Indian Restaurant', 'Bar', 'French Restaurant', 'Sandwich Place', 'Tea Room', 'Bakery', 'Portuguese Restaurant', 'Concert Hall', 'Park', 'Greek Restaurant', 'Restaurant', 'Gas Station', 'Hotel', 'Burger Joint']
H1E	['Ice Cream Shop', 'Clothing Store', 'Bagel Shop', 'Supermarket', 'Cheese Shop', 'Vietnamese Restaurant', 'Gym', 'Brewery', 'Bar', 'French Restaurant', 'Café', 'Breakfast Spot', 'Sandwich Place', 'Middle Eastern Restaurant', 'Bakery', 'Asian Restaurant', 'Park', 'Bank']
H1R	['Clothing Store', 'Bagel Shop', 'Cheese Shop', 'Italian Restaurant', 'Vietnamese Restaurant', 'Gym', 'Brewery', 'Bar', 'French Restaurant', 'Café', 'Breakfast Spot', 'Middle Eastern Restaurant', 'Bakery', 'Asian Restaurant', 'Pizza Place', 'Liquor Store', 'Bank']
H1S	['Clothing Store', 'Bagel Shop', 'Cheese Shop', 'Italian Restaurant', 'Vietnamese Restaurant', 'Gym', 'Brewery', 'Bar', 'French Restaurant', 'Café', 'Middle Eastern Restaurant', 'Bakery', 'Asian Restaurant', 'Pizza Place', 'Park', 'Gas Station', 'Bank']
H2K	['Ice Cream Shop', 'Clothing Store', 'Bagel Shop', 'Supermarket', 'Cheese Shop', 'Italian Restaurant', 'Vietnamese Restaurant', 'Fast Food Restaurant', 'Gym', 'Brewery', 'Bar', 'French Restaurant', 'Breakfast Spot', 'Sandwich Place', 'Middle Eastern Restaurant', 'Asian Restaurant', 'Liquor Store', 'Gas Station', 'Pharmacy', 'Bank', 'Grocery Store']
H2L	['Ice Cream Shop', 'Clothing Store', 'Bagel Shop', 'Supermarket', 'Coffee Shop', 'Cheese Shop', 'Italian Restaurant', 'Vietnamese Restaurant', 'Fast Food Restaurant', 'Gym', 'Brewery', 'Bar', 'French Restaurant', 'Breakfast Spot', 'Sandwich Place', 'Middle Eastern Restaurant', 'Asian Restaurant', 'Liquor Store', 'Gas Station', 'Pharmacy', 'Bank', 'Grocery Store']
H3E	['Ice Cream Shop', 'Clothing Store', 'Bagel Shop', 'Supermarket', 'Coffee Shop', 'Cheese Shop', 'Italian Restaurant', 'Vietnamese Restaurant', 'Fast Food Restaurant', 'Gym', 'Sandwich Place', 'Middle Eastern Restaurant', 'Asian Restaurant', 'Liquor Store', 'Gas Station', 'Pharmacy', 'Bank']
H4C	['Clothing Store', 'Bagel Shop', 'Supermarket', 'Coffee Shop', 'Cheese Shop', 'Italian Restaurant', 'Vietnamese Restaurant', 'Fast Food Restaurant', 'Brewery', 'Bar', 'French Restaurant', 'Breakfast Spot', 'Sandwich Place', 'Middle Eastern Restaurant', 'Asian Restaurant', 'Gas Station', 'Pharmacy', 'Bank', 'Grocery Store']
H4L	['Ice Cream Shop', 'Clothing Store', 'Bagel Shop', 'Supermarket', 'Cheese Shop', 'Vietnamese Restaurant', 'Brewery', 'Bar', 'French Restaurant', 'Middle Eastern Restaurant', 'Bakery', 'Asian Restaurant', 'Pizza Place', 'Park', 'Restaurant', 'Gas Station', 'Bank']
H4R	['Ice Cream Shop', 'Clothing Store', 'Bagel Shop', 'Supermarket', 'Cheese Shop', 'Italian Restaurant', 'Vietnamese Restaurant', 'Fast Food Restaurant', 'Brewery', 'Bar', 'French Restaurant', 'Café', 'Middle Eastern Restaurant', 'Bakery', 'Asian Restaurant', 'Pizza Place', 'Liquor Store']
H8Y	['Ice Cream Shop', 'Clothing Store', 'Bagel Shop', 'Supermarket', 'Coffee Shop', 'Cheese Shop', 'Vietnamese Restaurant', 'Brewery', 'Bar', 'French Restaurant', 'Café', 'Breakfast Spot', 'Sandwich Place', 'Middle Eastern Restaurant', 'Bakery', 'Asian Restaurant', 'Pizza Place', 'Park', 'Liquor Store', 'Gas Station', 'Bank']
H9C	['Ice Cream Shop', 'Clothing Store', 'Bagel Shop', 'Supermarket', 'Coffee Shop', 'Cheese Shop', 'Italian Restaurant', 'Vietnamese Restaurant', 'Fast Food Restaurant', 'Gym', 'Brewery', 'Bar', 'Café', 'Middle Eastern Restaurant', 'Bakery', 'Asian Restaurant', 'Pizza Place', 'Liquor Store', 'Restaurant', 'Gas Station', 'Bank']
H2V	['Korean Restaurant', 'Coffee Shop', 'Harbor / Marina', 'Italian Restaurant', 'Auto Dealership', 'Golf Course', 'Fast Food Restaurant', 'Gym', 'Brewery', 'Thai Restaurant', 'French Restaurant', 'Sandwich Place', 'Hockey Arena', 'Liquor Store', 'Convenience Store', 'Athletics & Sports', 'Gas Station', 'Pharmacy', 'Shopping Mall', 'Grocery Store']
H4A	['Korean Restaurant', 'Coffee Shop', 'Harbor / Marina', 'Italian Restaurant', 'Auto Dealership', 'Golf Course', 'Fast Food Restaurant', 'Brewery', 'Thai Restaurant', 'French Restaurant', 'Breakfast Spot', 'Sandwich Place', 'Hockey Arena', 'Convenience Store', 'Athletics & Sports', 'Gas Station', 'Pharmacy', 'Shopping Mall', 'Grocery Store']
H8S	['Korean Restaurant', 'Coffee Shop', 'Harbor / Marina', 'Auto Dealership', 'Golf Course', 'Fast Food Restaurant', 'Gym', 'Bar', 'Thai Restaurant', 'French Restaurant', 'Café', 'Breakfast Spot', 'Sandwich Place', 'Bakery', 'Pizza Place', 'Hockey Arena', 'Liquor Store', 'Restaurant', 'Convenience Store', 'Athletics & Sports', 'Shopping Mall', 'Grocery Store']

H3K	['Ice Cream Shop', 'Supermarket', 'Coffee Shop', 'Deli / Bodega', 'Italian Restaurant', 'Sushi Restaurant', 'Pool', 'Fast Food Restaurant', 'German Restaurant', 'Automotive Shop', 'Gym', 'Beer Bar', 'Thai Restaurant', 'Breakfast Spot', 'Sandwich Place', 'Tea Room', 'Vegetarian / Vegan Restaurant', 'Plaza', 'Liquor Store', 'Gas Station', 'Pharmacy', 'Burger Joint', 'Bank', 'Record Shop']
H4N	['Ice Cream Shop', 'Supermarket', 'Deli / Bodega', 'Sushi Restaurant', 'Pool', 'German Restaurant', 'Automotive Shop', 'Brewery', 'Beer Bar', 'Thai Restaurant', 'Breakfast Spot', 'Sandwich Place', 'Tea Room', 'Bakery', 'Vegetarian / Vegan Restaurant', 'Plaza', 'Pizza Place', 'Park', 'Restaurant', 'Gas Station', 'Burger Joint', 'Record Shop']
H8R	['Coffee Shop', 'Deli / Bodega', 'Sushi Restaurant', 'Pool', 'Fast Food Restaurant', 'German Restaurant', 'Gym', 'Brewery', 'Beer Bar', 'Thai Restaurant', 'Breakfast Spot', 'Tea Room', 'Bakery', 'Vegetarian / Vegan Restaurant', 'Plaza', 'Pizza Place', 'Liquor Store', 'Restaurant', 'Gas Station', 'Burger Joint', 'Record Shop']

## Discussion

Throughout this analysis, we realized that the way to obtain quality recommendations was not to consider geographic areas that did not have a sufficiently representative sample. The counterpart to this is that we can only make recommendations for 20-25% of the geographic areas of the city of Montreal. But these are of a much better quality.

By taking into account the most satisfied groups as a benchmark, we offer the public authorities the possibility of setting up financial support measures for certain types of businesses and services that will increase the overall satisfaction of the population.

## Conclusion

In conclusion, it was necessary to make a choice between the quality of the recommendations provided, and the quantity of recommendations to be provided for all zones. This shows that for this type of analysis, a larger number of respondents was needed to provide quality and geographically focused recommendations. Perhaps we could continue the survey and obtain additional respondents and use the rate of Foursquare venues.