Coursera capstone project

What would be the best neighborhood to open a new fashion boutique in Montreal ?

Business and data science problem

- This study is aimed at provinding insights to wanna be fashion shop owners in Montreal. We would like to know the best neighborhood to open such a shop depending on neighborhood profiles and mean revenues
- The data science problem is to try to cluster the neighborhoods we have with such data to identify favorable ones

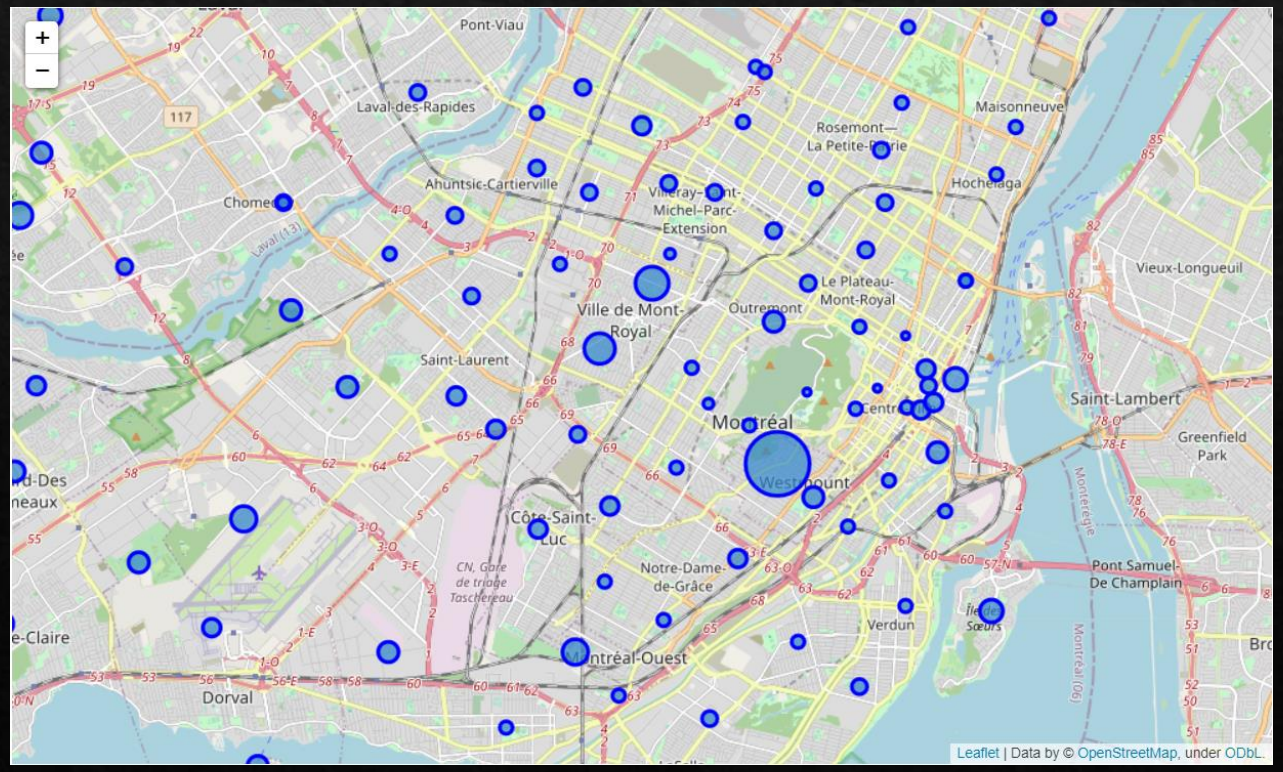
15/06/2020

Data collection and cleaning

- List of Montreal neighborbhood and socio economic data from Statistics Canada, printed in a new article http://www1.journaldemontreal.com/2017/09/salaires/
- Google geographical data (coordinates of each neighborhood), Geocoding API
- Foursquare venue data, Foursquare API

Using average revenue data per neighborhood to get insight

- Using the exhaustive list of neighborhoods and socio economic data we have from statistics Canada (as of 2015), we can quickly get a first insight as to where the richest neighborhoods are in Montreal :



Bubble plot of average household income per neighborhood

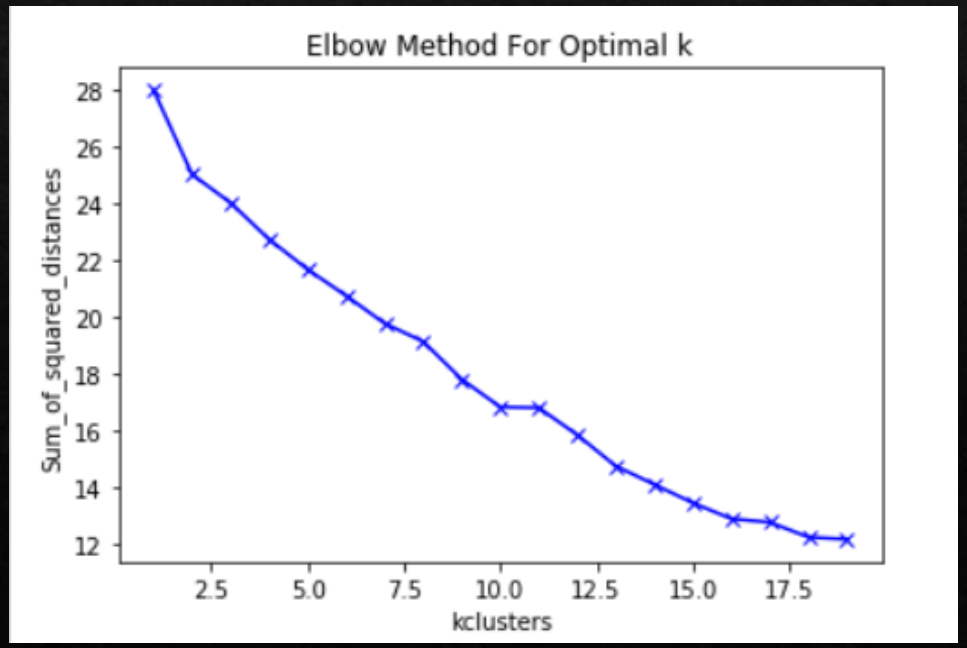
We gather that the richest areas are:

- Ville Mont royal, a rather residential neighborhood
- Westmount, also residential
- Neighborhoods around the gay village and downtown Montreal, which are incidently the most active areas regarding shopping and going out

15/06/2020

1st attempt at clustering neighborhoods based on their most common venues

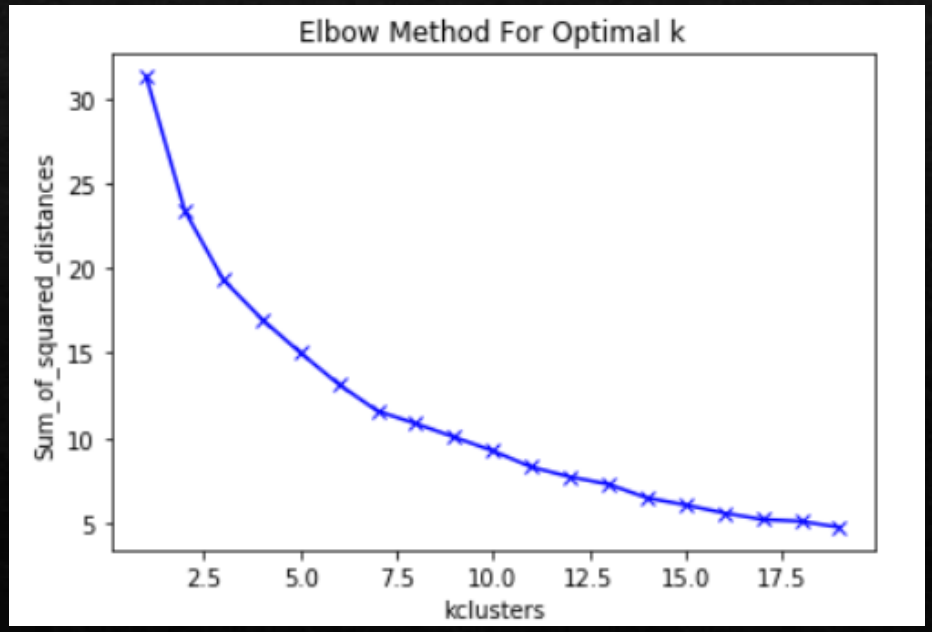
- Using the elbow method to obtain the optimal number of clusters to ask k means to understand, we try to cluster the top 10 venues data per neighborhood using the foursquare categories



- We do not get a clear elbow so we will try to adapt the data

2nd attempt at clustering neighborhoods venue data

- We used foursquare rather detailed venue categorization for the first attempt, we think maybe it made for too specific neighborhood profiles. We try to use a macro-categorization, going from 264 venue categories to 19.



- We get a better looking curve but still no clear elbow

Conclusion and future directions

- We were able to identify neighborhoods with the highest houselhod income in the city
- We were not able to get an optimal number of cluster using the elbow method to cluster venues data with k means
  - We could try alternative methods like analyzing the silhouette data from clustering attemps
  - We could try to reprocess the data to make it easier to cluster
  - We could try another clustering algorithm
- Once clear profiles of neighborhoods are observed, and neighborhood candidates selected for the shop opening, a precise business analysis will need to be conducted to ensure viability of the project