IBM Data Science- Capstone Project Report
The battle of neighborhoods

Classification of Montreal neighborhoods

Aron Hayoun

Introduction

- Montreal is the second most-populous city in Canada: **daily transportations are a challenge**.

- **Idea:** classifying the different neighborhoods by types of venues could help planning the transportation resources in the different times of the day.

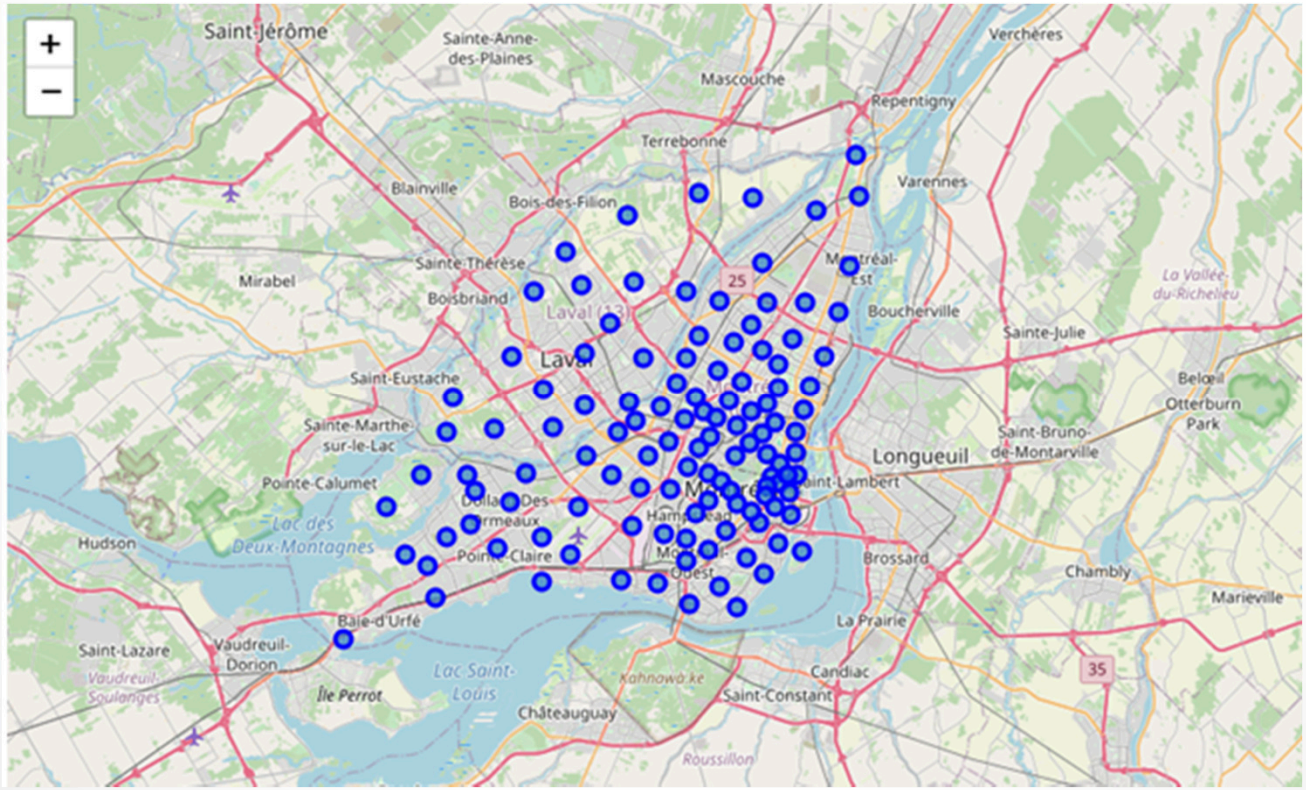
- **Business interest :**

  - A taxi company to plan a better dispatch of the taxi float at each time of the day.

  - The city transportation system to help for an expansion of the metro network

Data

- Neighborhoods data: we need the geographic coordinates of each neighborhood:



Data

- Venues data: we use the Foursquare API to find the number of venues in each category:
- Foursquare categories are the following:

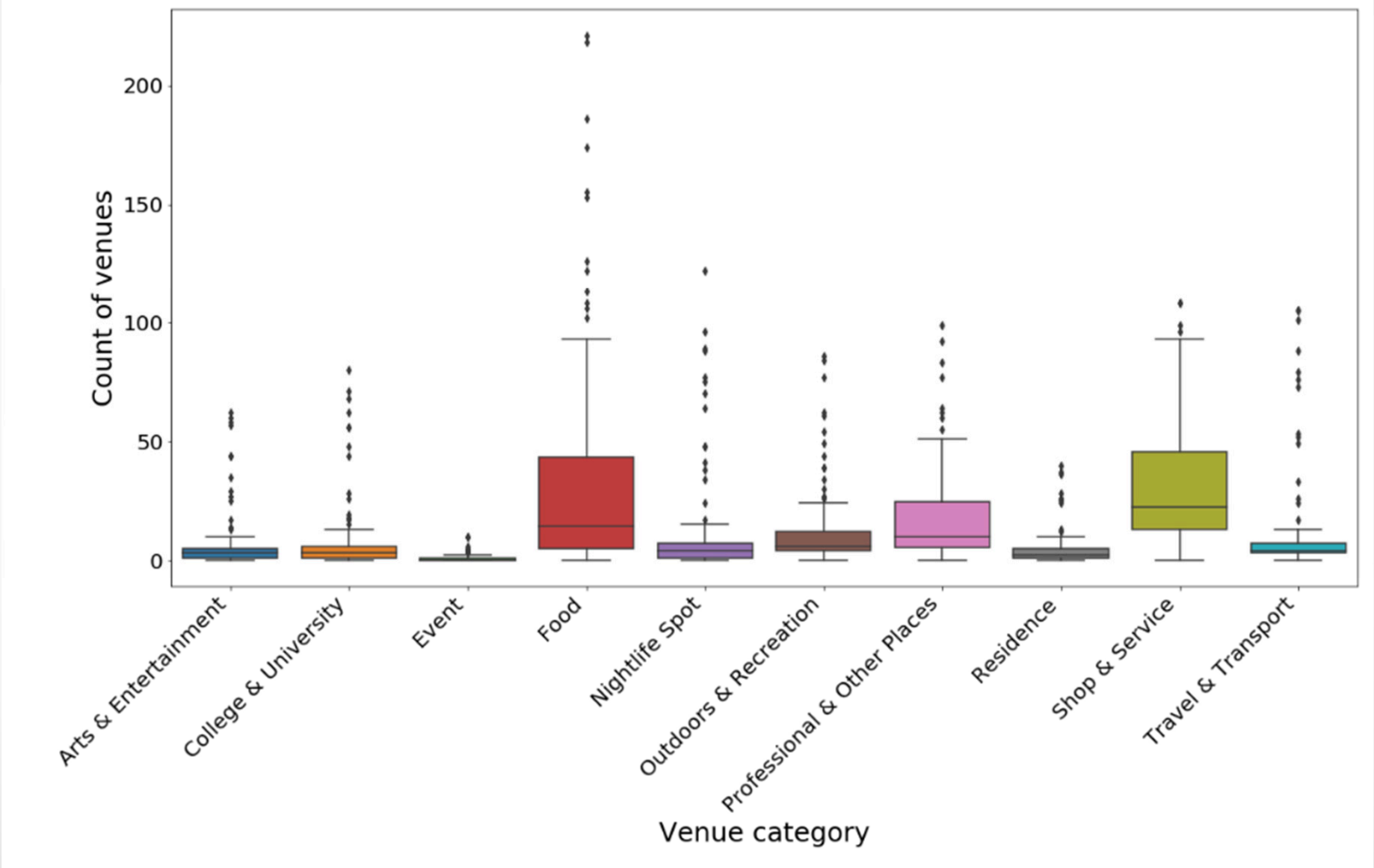
```
Arts & Entertainment
College & University
Event
Food
Nightlife Spot
Outdoors & Recreation
Professional & Other Places
Residence
Shop & Service
Travel & Transport
```

Methodology

We request to the API the number of venues for each category:

Code	Neighborhood	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
H1A	Pointe-aux-Trembles	45.674145	-73.500435	1	1	0	4	0	5	2	0	4	5
H2A	Saint-MichelEst	45.561809	-73.601338	3	2	1	19	8	6	13	2	28	5
H3A	Centre-ville de MontréalNord (Université McGill)	45.504000	-73.575000	57	80	4	186	88	77	83	36	108	88
H4A	Notre-Dame-de-GrâceNord-est	45.472000	-73.615000	17	1	0	84	6	9	25	2	51	6
H9A	Dollard-Des-OrmeauxNord-ouest	45.495801	-73.832858	0	2	0	5	2	4	4	0	2	2

Methodology: Data exploration



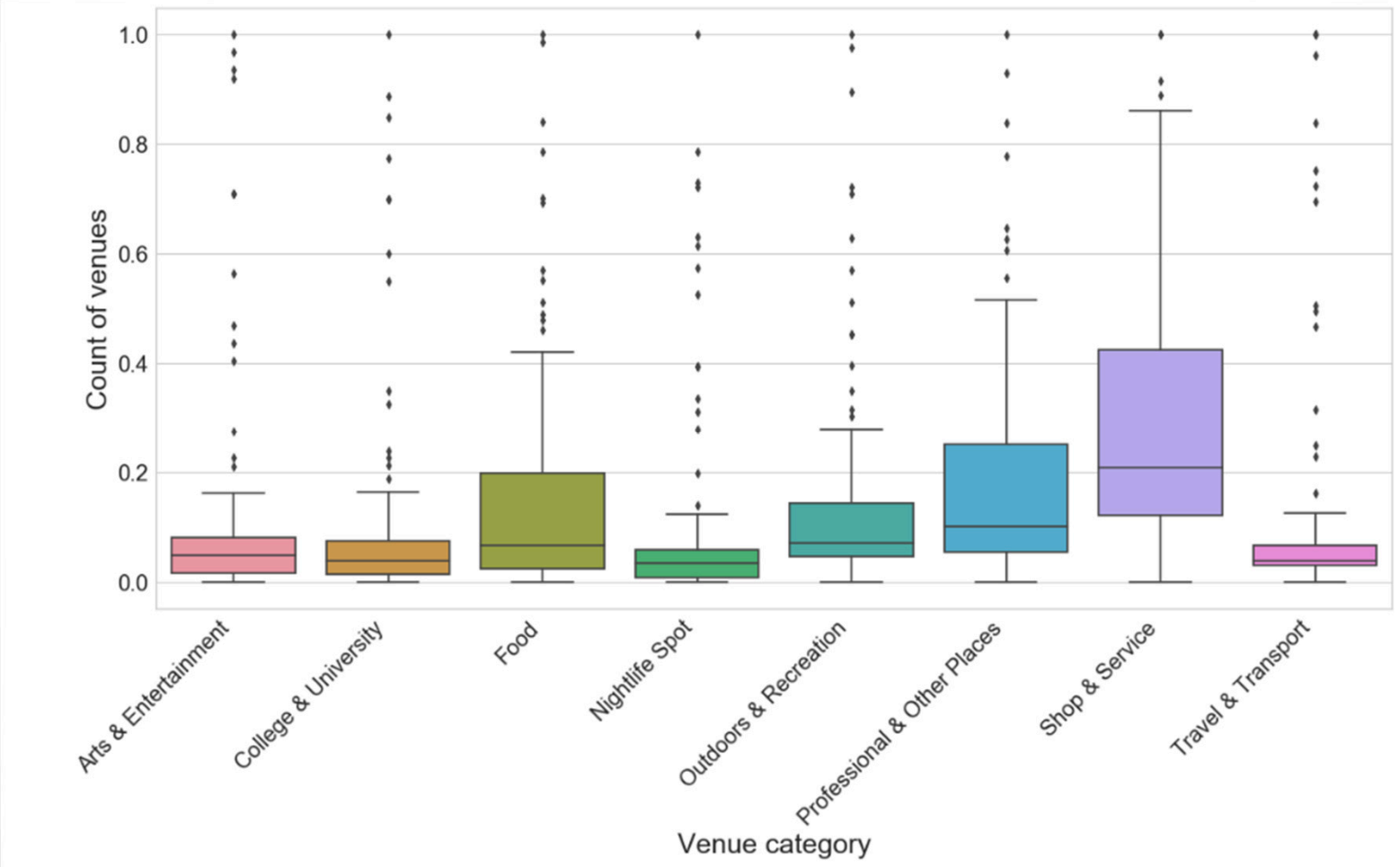
Methodology: Data preparation

event & residence categories has very few numbers in the different areas, we can drop these categories as it won't make any big differences (Also most of residences are not in Foursquare data so it is not relevant to use this category)

We normalize the data with MinMaxScaler

	Arts & Entertainment	College & University	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Shop & Service	Travel & Transport
0	0.016129	0.0125	0.018100	0.000000	0.058140	0.020202	0.037037	0.047619
1	0.048387	0.0250	0.085973	0.065574	0.069767	0.131313	0.259259	0.047619
2	0.919355	1.0000	0.841629	0.721311	0.895349	0.838384	1.000000	0.838095
3	0.274194	0.0125	0.380090	0.049180	0.104651	0.252525	0.472222	0.057143
4	0.000000	0.0250	0.022624	0.016393	0.046512	0.040404	0.018519	0.019048

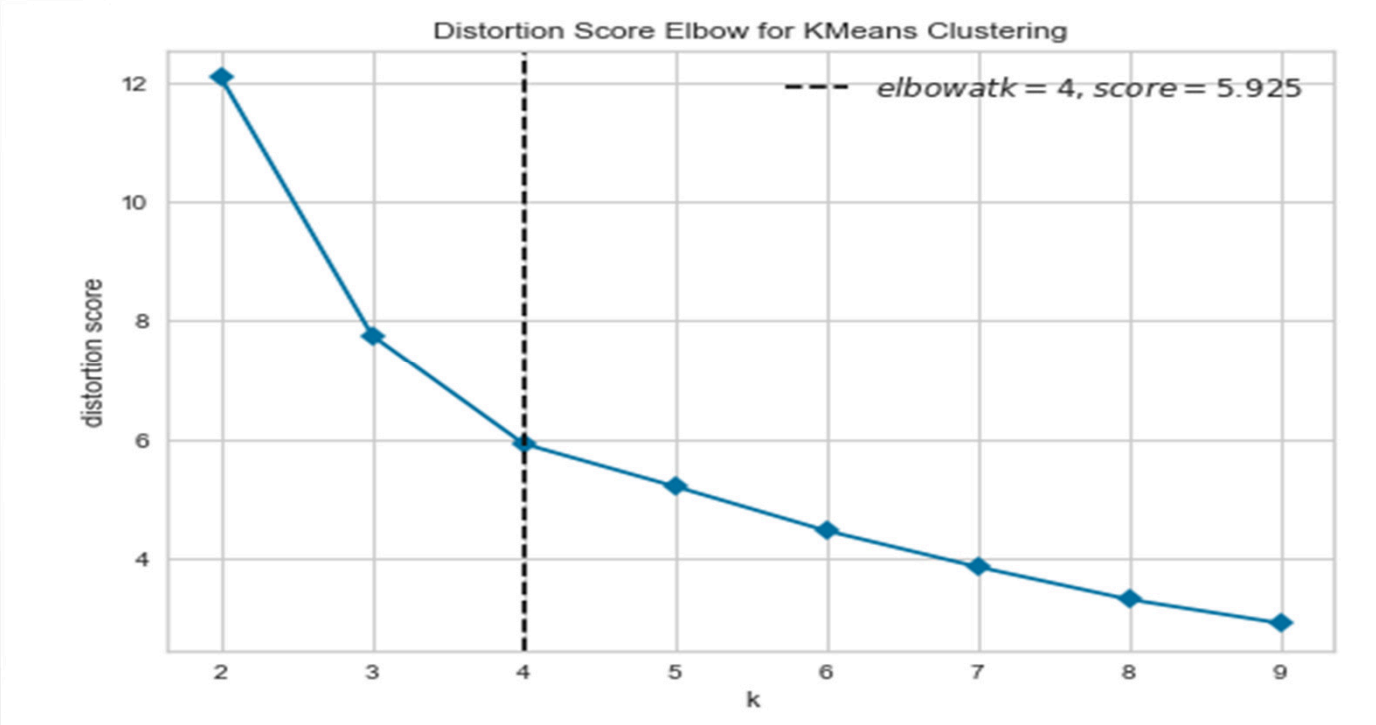
Methodology: Data preparation



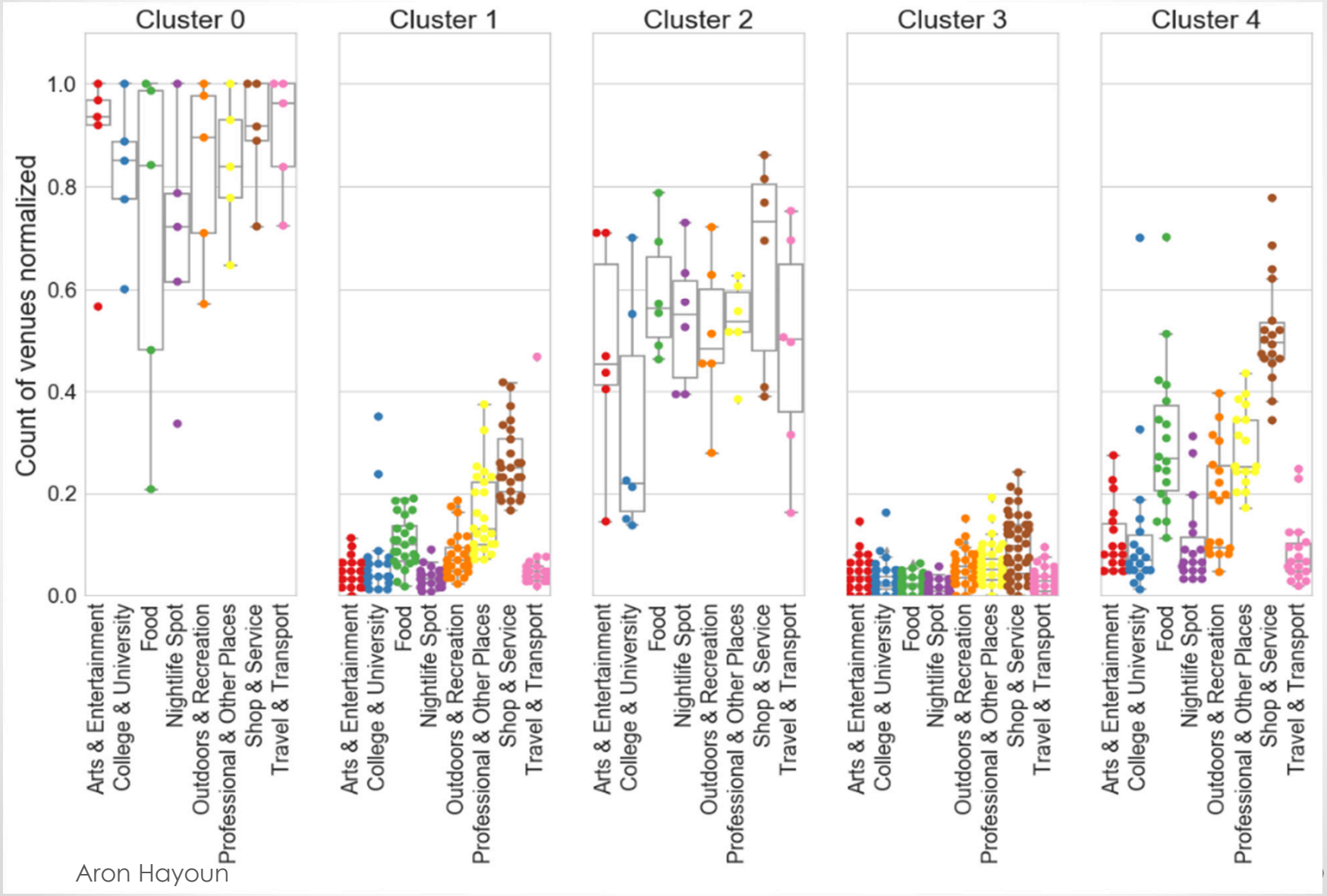
Methodology: Clustering

Number of clusters:

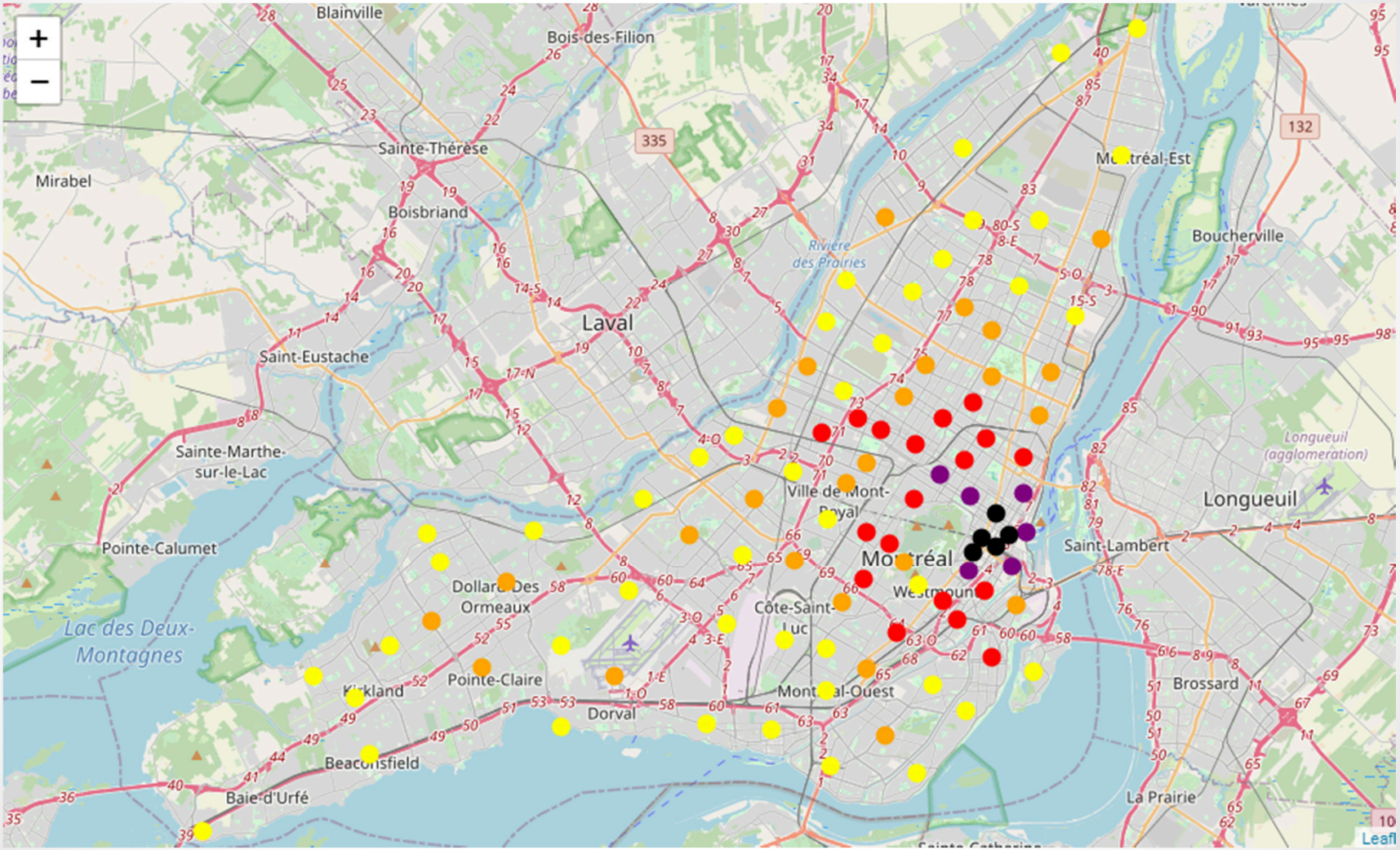
- 5 clusters is empiricaly a good choice
- Elbow method says 4
- We stick with 5



Methodology:Clustering

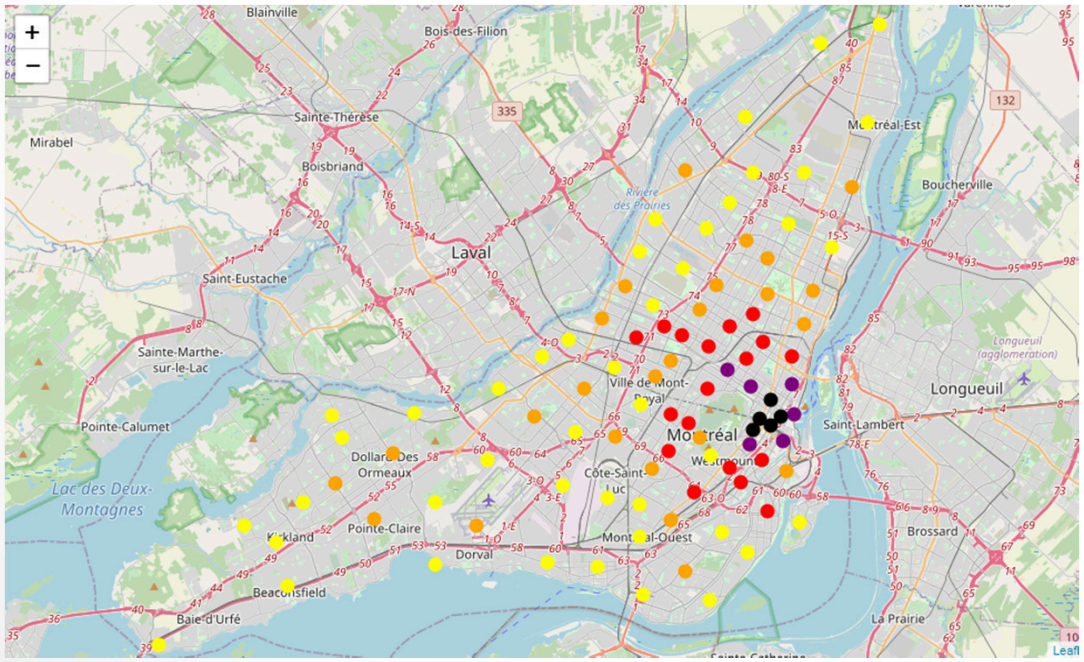


Methodology:Clustering



Results

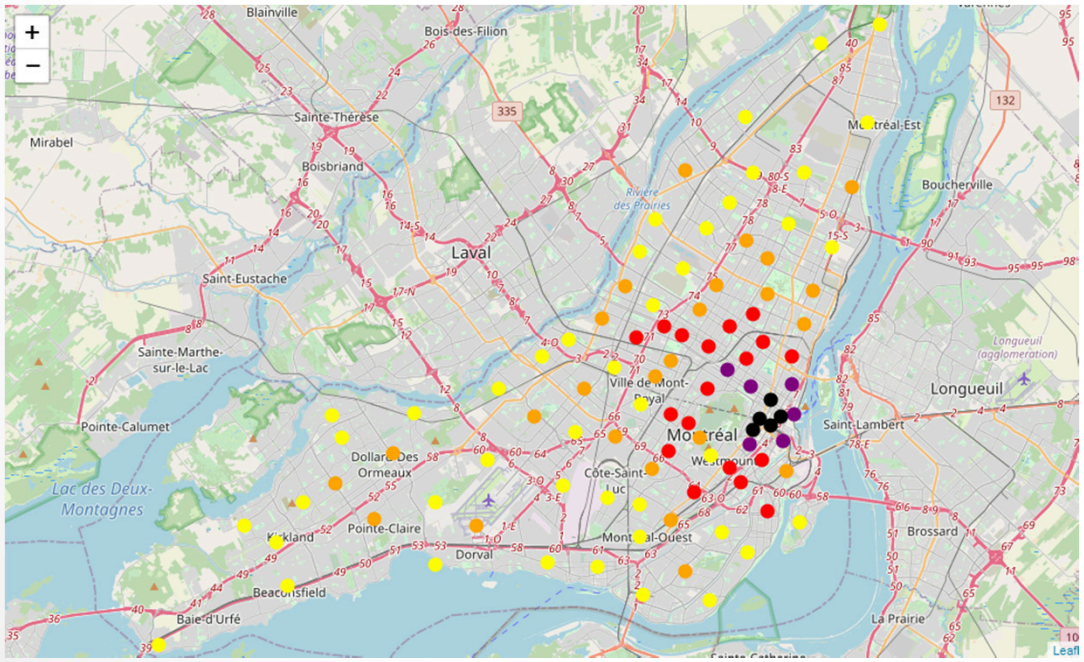
- Cluster 0 (Black) has the highest scores for all venue categories. These neighborhood are the economic and cultural center of Montreal
- Cluster 2 (Purple) has high score for all venue categories
- Cluster 4 (Red) has lower scores with best scores in Shops, Professional places & Food. These areas are mostly commercials
- Cluster 1 (Orange) has low marks with better scores for Food and Shops. This cluster correspond to commercial & residential suburban areas
- Cluster 3 (Yellow) is mostly residential, with very low scores everywhere



Results

After viewing the map:

- Cluster 0 correspond to downtown Montreal
- Cluster 2 is the downtown immediate periphery.
- Cluster 4 corresponds to secondary cultural and economic centres where the first big waves of immigration (Italians, Greeks…) settled. These area are in a process of gentrification.
- Clusters 1 and 3 aren't so clearly geographically distributed but most of the orange points (cluster 1) are located close to downtown



Discussion

- The study relies only on the Foursquare Data. We cannot be certain that the data is complete since some venues could be missing.

- Also the study doesn't consider the size or importance of the venues and one venue such as a train station surely has a really bigger impact on the attractivity of an area than a restaurant.

Conclusion

The clustering method proved to be a good way to understand the economic and cultural dynamic of a city. The more various, complete and precise is the data, the more the clusters will give clear insights and the more we will give credit to the results.