Clustering Neighborhoods for supporting tourists decisions

Thiago F. C. De Oliveira

Business Problem

- What if you can get all the information about the neighborhood you will be staying? How easy is to get to the subway, or how many options would you have to eat if you decide to take a walk?

- Provide a cluster analysis of Toronto's neighborhoods in order to support tourists decision in choosing a place to stay in a vacation travel.

Data

- List with all postal codes, boroughs and neighborhoods of Toronto, avaiable in wikipedia.

- .csv file with latitude and longitude of each neighborhood.

- Venues of each neighborhood acquired with FourSquare API, with its category and price.

- transit.land database to get all subways and bus station.

- Cleaned data contains 11 features.
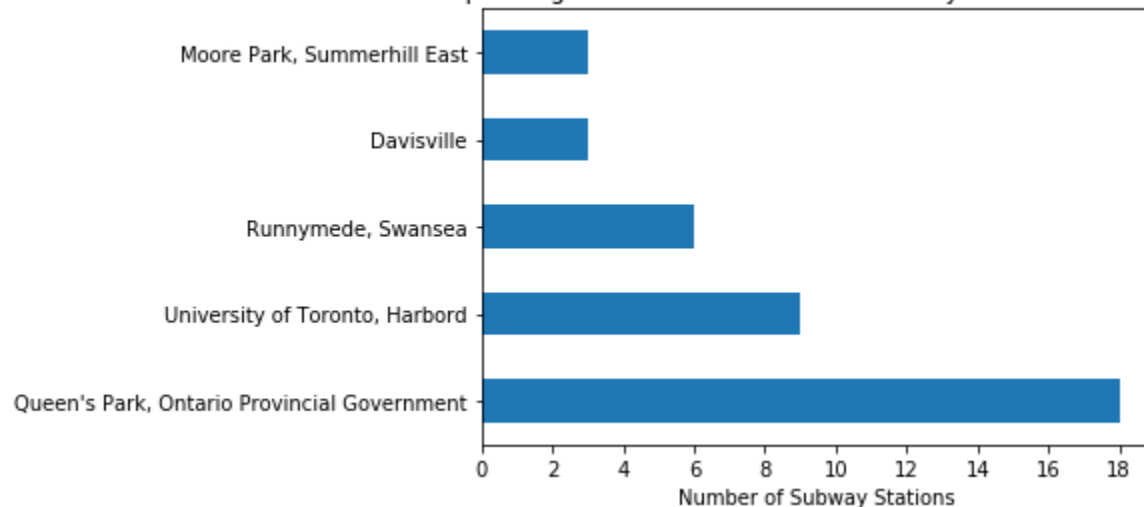
Categories occurence

Toronto Venues Classification
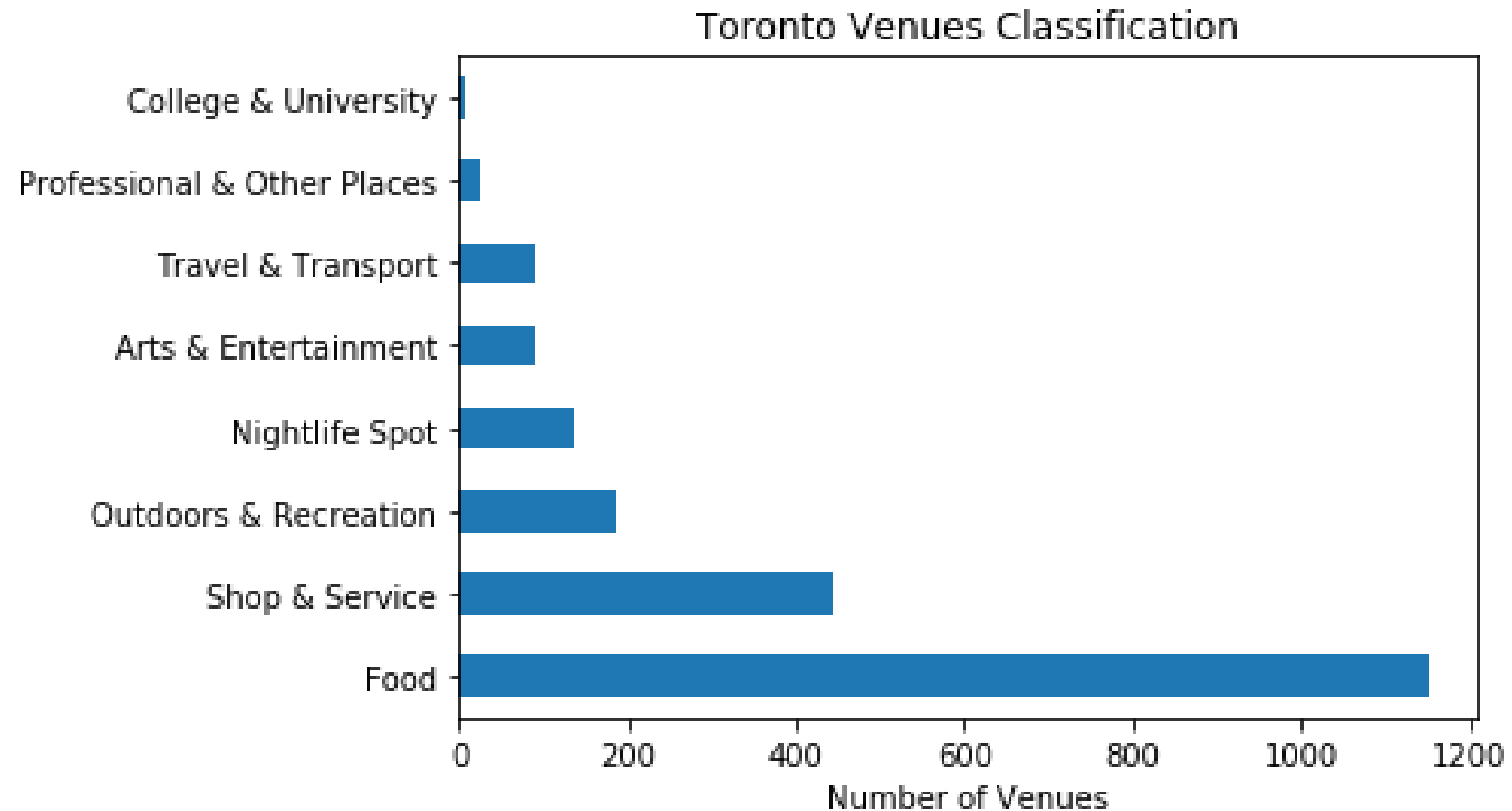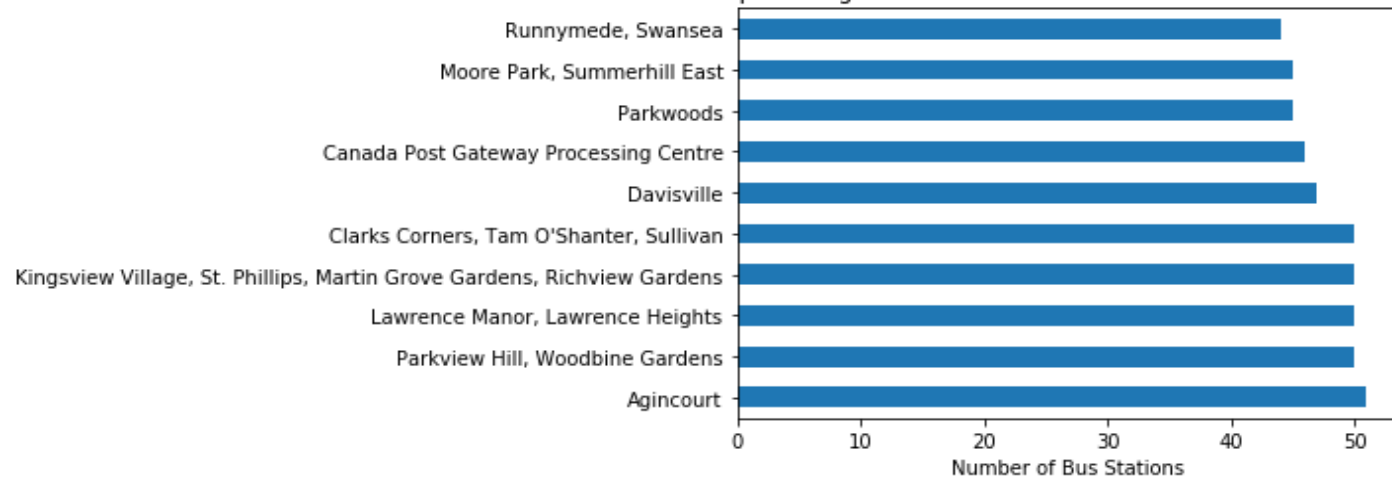
Distribution of Prices
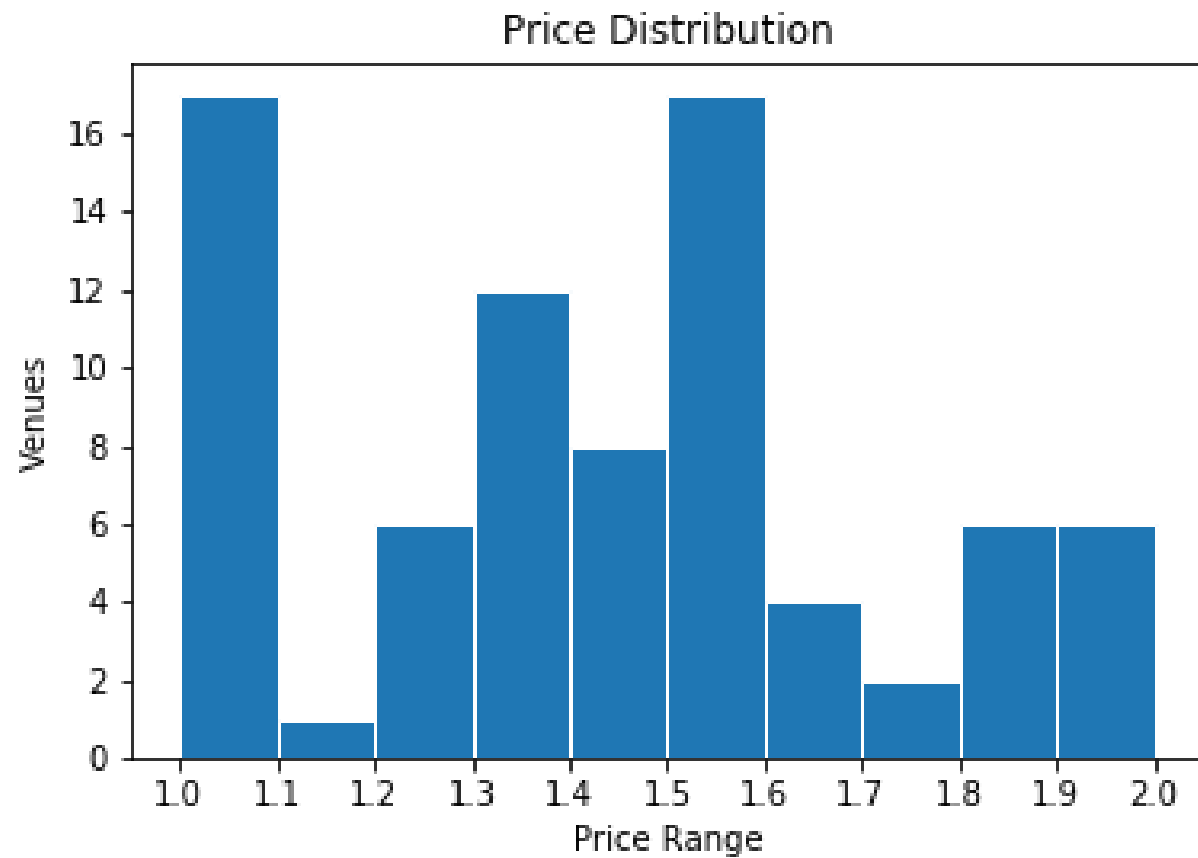
Mobility

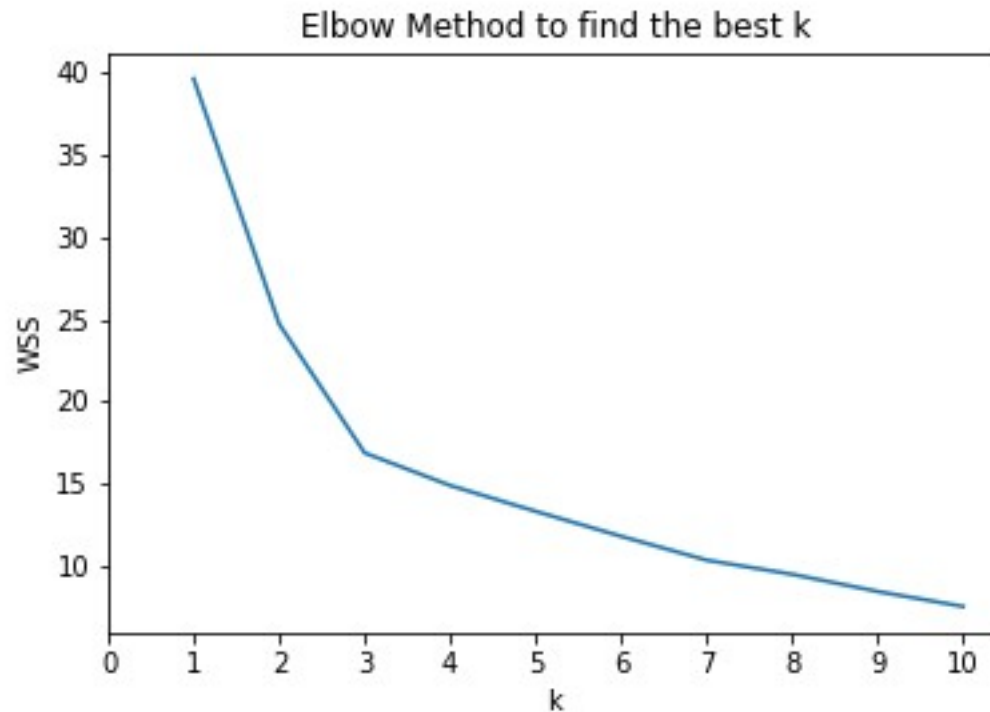
Top 5 Neighborhoods in Number of Subway Stations in Toronto

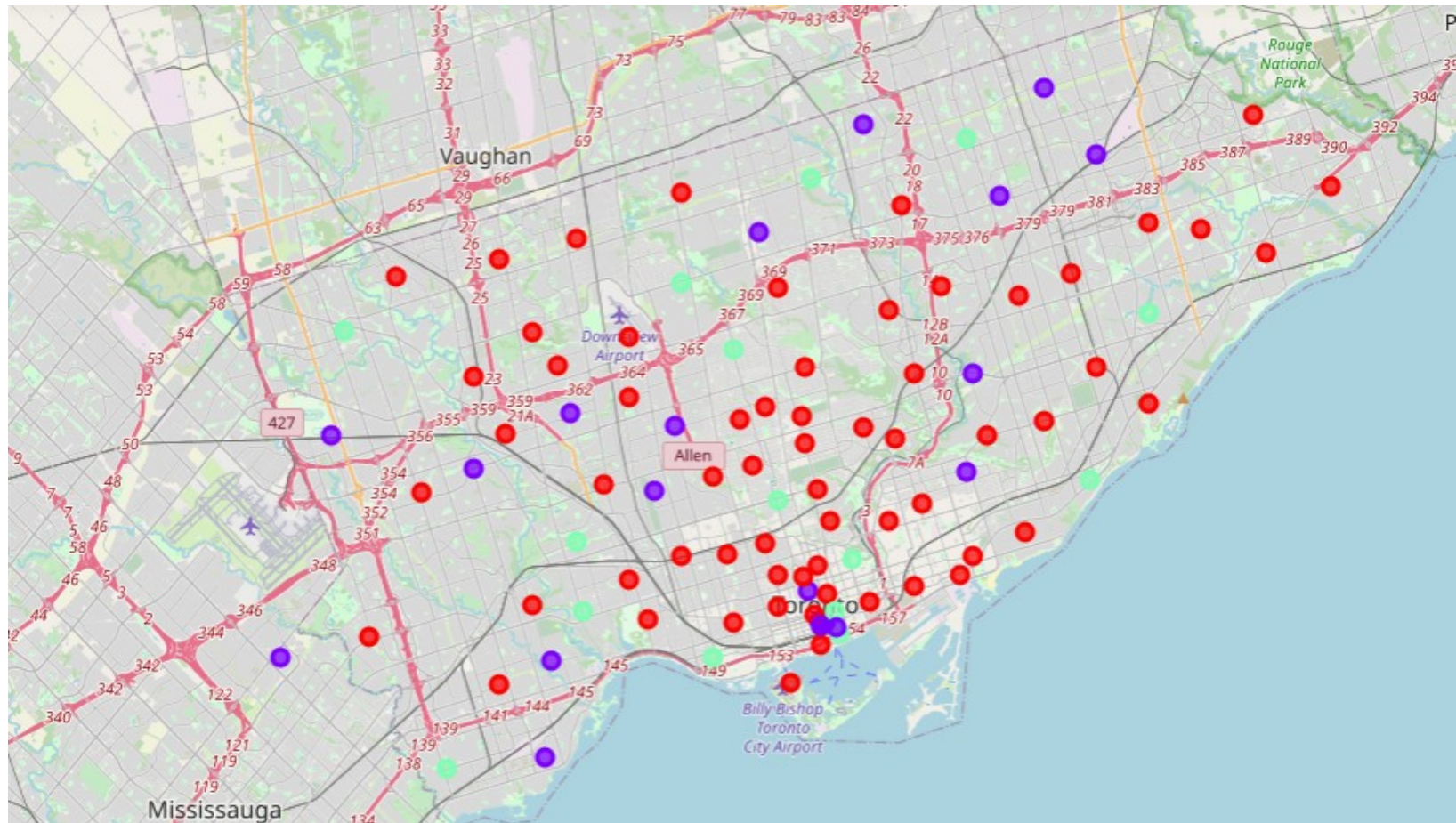Top 10 Neighborhoods in Number of Bus Stations in Toronto

Regression Model: k-Means Clustering

- The best k, by the elbow method, is k=3

Elbow Method to find the best k

Toronto's Clusters

Clusters analysis and occurence

Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	#_Bus	#_Subway	price
Red	Food	Shop & Service	Outdoors & Recreation	0.148459	0.030864	0.560112
Purple	Food	Shop & Service	Travel & Transport	0.225490	0.000000	0.604491
Blue	Food	Shop & Service	['Nightlife Spot' 'Travel & Transport']	0.057516	0.022222	0.582284

Cluster	Ocurrence
Red	63
Purple	20
Blue	15

Conclusions

- This classification provide insight to tourists.

- When choosing a place to stay on Airbnb, or book a hotel, it is possible to know a priori the characteristics of the neighborhood, and take this into account when planning your trip.