Where To Open A Coffee Shop In Toronto

The Problem

- Client plans on opening a new coffee shop in Toronto.
- Search for a list of neighborhoods which would give it the greatest chance of success.

Analytical Approach

- The factors affecting the potential of a neighborhood were identified.
- Features were selected that serve as indicators for those factors.
- Data was normalized and neighborhoods were clustered using K means clustering algorithm.
- Distance of a cluster centroid from origin gives a measure of sum of effects from positive factors.
- Clusters whose centroids were farthest from origin were selected. Neighborhoods in the selected clusters were short listed to create a list of preferred neighborhoods.

Positive Factors

Table below lists the positive factors which need to be considered while evaluating the prospect of a neighborhood.

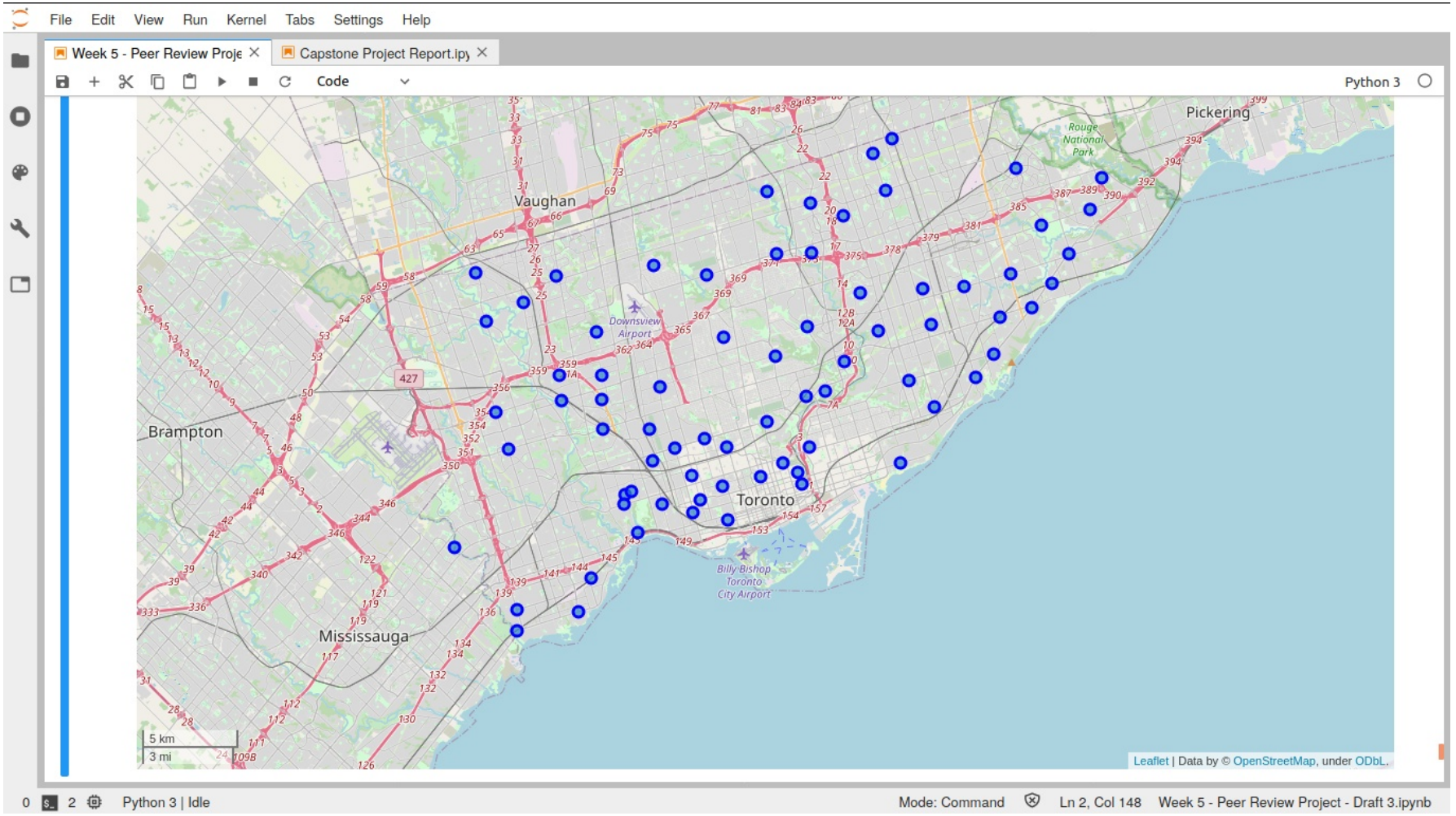
Factor	Rationale
Population	To support a new business
Average Income	Residents‘ disposable income
Walkability	More people exploring venues
Business Atmosphere	Economic strength of a neighborhood
Parks and Playgrounds	Draw crowds enjoying leisure time

Features and Data Sources

The following table lists the features of the model. They serve as indicators of the positive factors listed in the previous slide.

Feature	Data Source
Population, Average Income	Wikipedia
Average Income	Wikipedia
Walkability	Toronto Open Data
Businesses, Debt Risk Score	Toronto Open Data
Venues (including coffee shops, parks and playgrounds)	Foursquare API

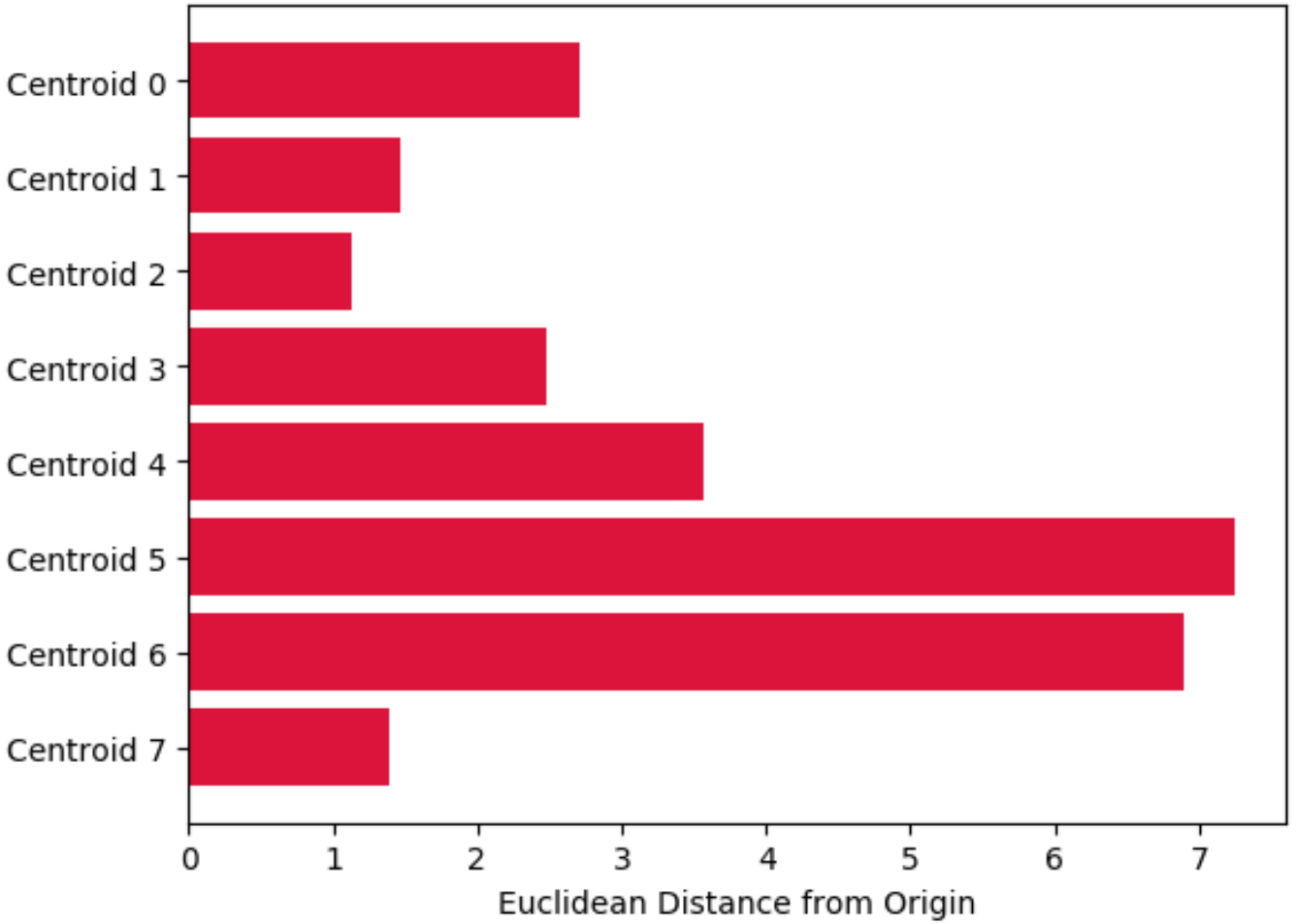
Toronto Neighborhoods



Neighborhoods - Summary

	index	Businesses	Debt Risk Score	Walk Score	Population	Average Income	Latitude	Longitude	Number of Cafes	Number of Parks/Playgrounds
count	74.000000	74.000000	74.000000	74.000000	74.000000	74.000000	74.000000	74.000000	74.000000	74.000000
mean	39.418919	516.472973	737.216216	70.418919	17124.486486	38885.932432	43.712159	-79.387853	0.527027	0.486486
std	21.644707	629.280481	28.177651	11.880693	10406.877090	25578.174258	0.055467	0.110338	1.009855	0.706976
min	1.000000	47.000000	679.000000	42.000000	2790.000000	19521.000000	43.592005	-79.585434	0.000000	0.000000
25%	21.250000	184.250000	718.250000	61.000000	10404.000000	26682.750000	43.670903	-79.473990	0.000000	0.000000
50%	39.500000	356.500000	738.000000	69.500000	14654.500000	30836.500000	43.710416	-79.385243	0.000000	0.000000
75%	57.750000	578.250000	758.000000	79.000000	21350.750000	44076.500000	43.754799	-79.307272	1.000000	1.000000
max	76.000000	4324.000000	791.000000	99.000000	52220.000000	213941.000000	43.823174	-79.165837	4.000000	4.000000

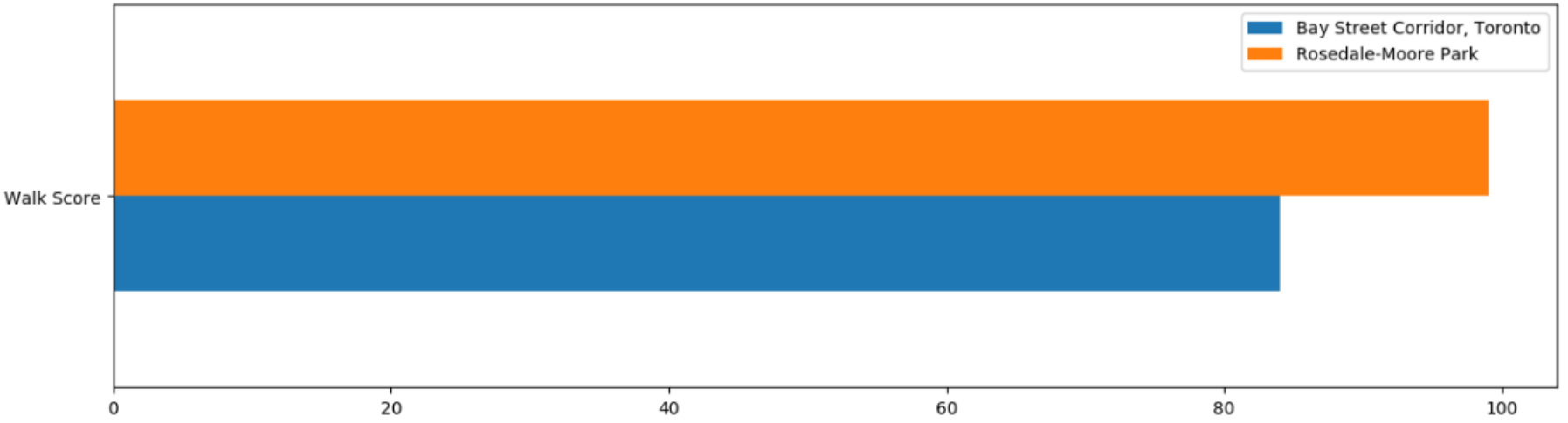
Distance to Centroids from Origin



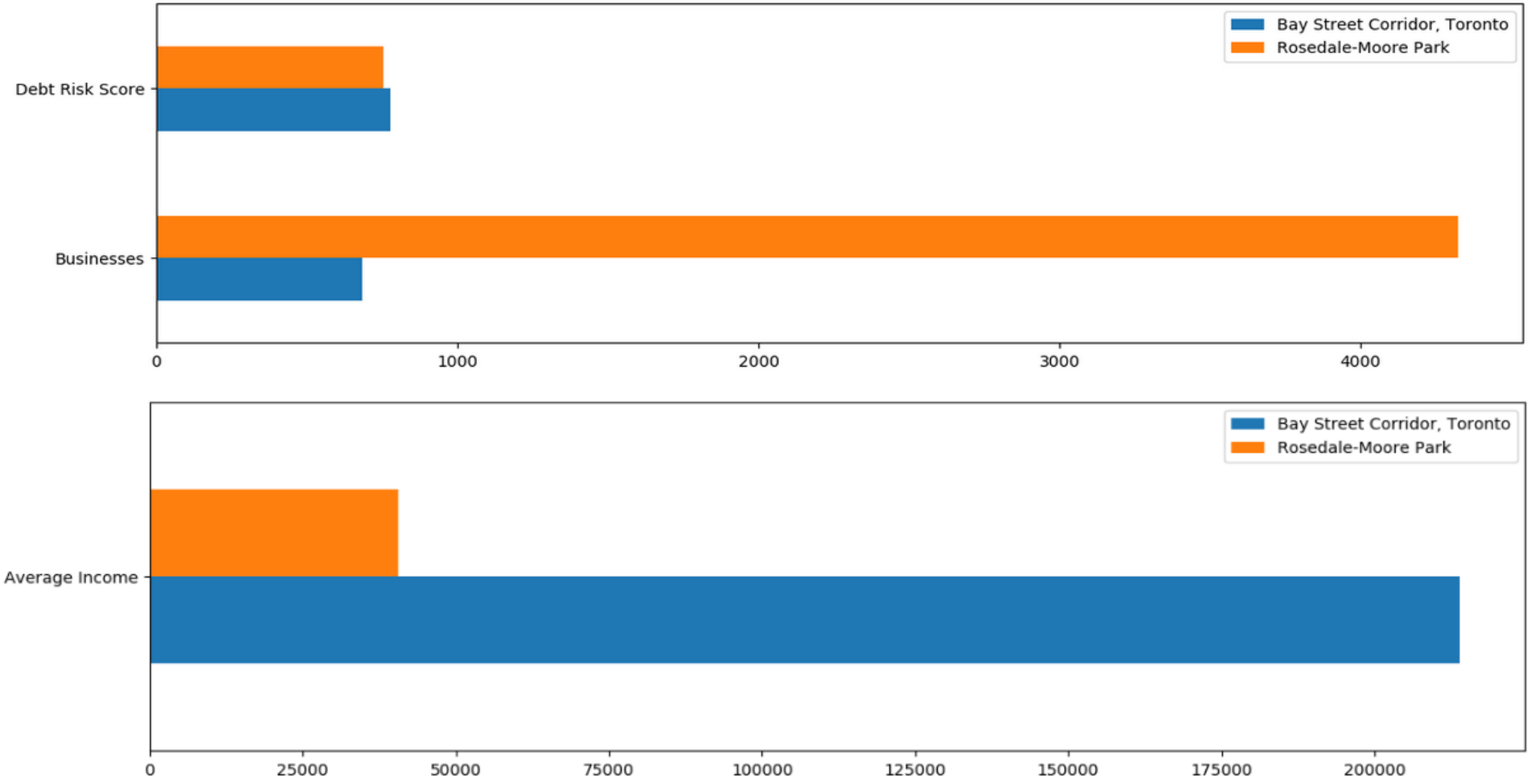
Preferred Neighborhoods Contd.

The two neighborhoods selected finally are:
1. Bay Street Corridor
2. Rosedale-Moore Park

The charts below illustrate how the two neighborhoods compare in a head to head match-up on various metrics.



Preferred Neighborhoods – Economic Factors



Observations and Conclusions

- 74 Toronto neighborhoods were clustered based on 8 features all of which positively affect the potential of place.
- Two clusters stood out with their centroids being farthest from the origin.
- The two clusters had one neighborhood each:Bay Street Corridor and Rosedale-Moore Park. The two neighborhoods made it to the final short list.
- Bay street corridor is a busy thoroughfare and Rosedale-Moore Park is an affluent neighborhood with lots of greenspace.
- Inspite of their differences, both neighborhoods have more than enough financial strength to easily integrate a new business.
- The two neighborhoods are also very walkable with the ability to draw new people.