

As Ottawa gets 'Ready for Rail', stations should get 'Ready for Growth'

Coursera Capstone Project

By Andrew L Macumber

Introduction

With the eminent opening of the new light rail transit system - The Confederation Line - Canada's capital, Ottawa, Ontario, is set to complete its largest transportation infrastructure project since the building of the Rideau Canal. With an expected increase in commuter traffic at rail stations along the Confederation Line it would be of interest to know how the venues around the stations might change. For example, is there a need for more Food venues at some of the stations. This information is of value to small business owners looking for suitable locations.

We can gain some insight into how venues might change by comparing the relatively new Ottawa stations on the Confederation Line, to stations on a much older and more developed rail line. Line 1 in Toronto, Ontario was opened in 1954 and is one of the busiest lines in North America.

Around each station we can assess the proportions of different types of venue (e.g., Food, Arts & Entertainment, Professional) and then compare each station to see if any differences exist, specifically between the Ottawa stations and the Toronto stations. Differences represent valuable targets for small business owners to establish new business' near Ottawa stations.

Data Description

The Confederation Line is comprised of twelve stations while Line 1 has thirty-eight stations. I chose a subset of fourteen Line 1 stations, representing the original line opened in 1954 from Yonge to Spadina.

The Foursquare API was used to find all venues within a five-minute walking distance (416 m) of each station and collect the venue category (e.g., Food, Arts & Entertainment, College & University) information.

The Foursquare API uses over 900 subcategories for venues organised under ten parent categories. This would make it difficult to do a general comparison between stations as some categories can be quite specific and not found at all stations. A category key was needed that translated any category returned by the Foursquare API to its parent category.

For each station, the percent of each venue category was calculated, and this was used for statistical analyses.

Methodology

After selecting the stations that would be part of the study, I created a list of station names including their city. I used the Geopy library to retrieve the station coordinates from OpenStreetMap Nominatim. This was formatted into a dataframe containing Station Name, City, Latitude and Longitude.

Foursquare has a very detailed venue category labeling scheme containing ten parent categories and 900 subcategories. In order to compare the proportion of different venue types it was necessary to use a standardised venue category convention. I decided to only use the parent category classification. When calling the Foursquare API only one category for the venue is listed, and it is often a subcategory. Thus, I needed a category key that could translate a subcategory back to its parent category. The full category tree was collected using the Foursquare API. This tree included ten parent categories and all subcategories. A dataframe listing every category and its corresponding parent category was used as the category key.

A call to the Foursquare API referencing the dataframe of station coordinates was used to collect all venues within a five-minute walking distance (416 m) of the stations. For each station, I converted the venue category to the parent category using the category key. Next, I grouped the venues by category and counted the number of venues for each parent category. The category counts for each station were merged into a single dataframe containing station information, category counts and total venues. I removed the 'Residence' and 'Event' categories since more than 25% of the stations did not have venues of these types.

To correct for differences in the total number of venues found at each station, I converted the category counts to percentages. Next, I investigated the distributions of each category using box plots. This was used to look for outliers and examine which categories provided the greatest amount of variation.

I performed hierarchical clustering (i.e., complete linkage) on the category counts dataset. Counts were converted to percent and then converted to an array prior to being transformed into a distance matrix. The distance matrix was used for the analysis.

Finally, a decision tree model was fitted to the category count dataset using the Sklearn library. This provided additional insight into how the stations might be divided into groups based on the proportions of their venue types.

Results

Table 01 lists all the train stations including their location (City, Latitude, Longitude) and the number of each type of venue (i.e., parent category) within a five-minute walk (416 m). A total of 4168 venues were returned. There were 1662 venues associated with Ottawa stations ($n = 12$) and 2506 venues associated with Toronto stations ($n = 14$). Toronto stations had a greater number of venues on average (179 ± 41) compared to Ottawa stations (139 ± 41). Toronto stations had on average a greater number of "Food" and "Professional & Other Places" venues (32 ± 12 and 55 ± 27) than Ottawa stations (21 ± 12 and 31 ± 27). In total 393 unique categories were returned, with Toronto stations having 321 and Ottawa stations having 290. Toronto stations had on average a greater number of unique categories (83 ± 16) than Ottawa stations (66 ± 16).

Table 1. Ottawa and Toronto station venue category counts.

Station	City	Latitude	Longitude	Arts & Entertainment	College & University	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Shop & Service	Travel & Transport	Total Venues	Unique Categories
Bayview	Ottawa	45.40923	-75.72232	10	1	25	11	10	42	51	11	184	99
Blair	Ottawa	45.431026	-75.60842	1	1	26	1	4	20	35	10	105	63
Cyrville	Ottawa	45.422745	-75.62637	2	0	8	2	2	30	36	11	104	60
Hurdman	Ottawa	45.412335	-75.66442	2	11	15	3	14	16	15	26	125	62
Lees	Ottawa	45.416285	-75.67053	3	13	15	3	11	14	10	16	106	61
Lyon	Ottawa	45.418518	-75.70502	2	4	25	4	7	82	21	28	191	73
Parliament	Ottawa	45.421389	-75.69889	3	0	28	5	6	68	34	5	158	69
Pimisi	Ottawa	45.413592	-75.71373	24	2	34	2	21	30	28	16	188	89
Rideau	Ottawa	45.327995	-75.70152	0	1	13	2	7	41	15	5	90	50
St-Laurent	Ottawa	45.420605	-75.63819	3	1	35	0	1	7	133	1	195	71
Tremblay	Ottawa	45.416933	-75.65335	2	1	16	0	4	13	20	24	85	47
uOttawa	Ottawa	45.420677	-75.68247	1	83	11	4	6	8	5	4	131	50
Bloor-Yonge	Toronto	43.670812	-79.38685	2	5	25	2	5	92	49	8	200	81
College	Toronto	43.660662	-79.3828	6	3	36	4	6	42	44	10	175	94
Dundas	Toronto	43.658537	-79.38102	15	6	53	5	10	41	47	5	199	99
King	Toronto	43.648523	-79.37777	4	2	31	7	7	84	37	14	199	75
Museum	Toronto	43.667097	-79.39361	37	36	14	4	12	34	14	5	164	82
Osgoode	Toronto	43.650802	-79.38654	11	1	31	8	15	61	15	25	188	92
Queen	Toronto	43.652908	-79.37946	9	0	36	6	3	31	96	7	198	94
Queen's Park	Toronto	43.65998	-79.39037	3	19	35	1	3	80	23	6	181	85
Spadina	Toronto	43.667235	-79.40369	4	12	25	5	6	43	24	61	195	83
St-Andrew	Toronto	43.647732	-79.38487	4	3	39	5	4	96	29	12	200	78
St-George	Toronto	43.668641	-79.39802	11	23	12	4	4	37	13	7	123	72
St-Patrick	Toronto	43.65482	-79.38827	3	5	46	2	4	82	28	8	197	82
Union	Toronto	43.64469	-79.37997	4	0	23	4	1	17	12	29	97	49
Wellesley	Toronto	43.665403	-79.3836	6	2	41	10	5	26	61	11	190	97

For Figure 1 venue counts for each station were converted to percent and grouped by category. Four categories did not exceed 10% (apart from outliers) at any station: “Outdoor & Recreation”; “Nightlife Spot”; “College & University”; and “Arts & Entertainment”. Both “Professional & Other Places” and “Shop & Service” venues had the greatest range of values, followed by “Travel & Transport” and “Food”. Several outliers were identified (black circles), but the majority fall within the range of other category distributions, apart from two extreme outliers in “Shops & Service” and “College & University”. Those two outlying stations were the only ones to exceed 55% in a single venue category.

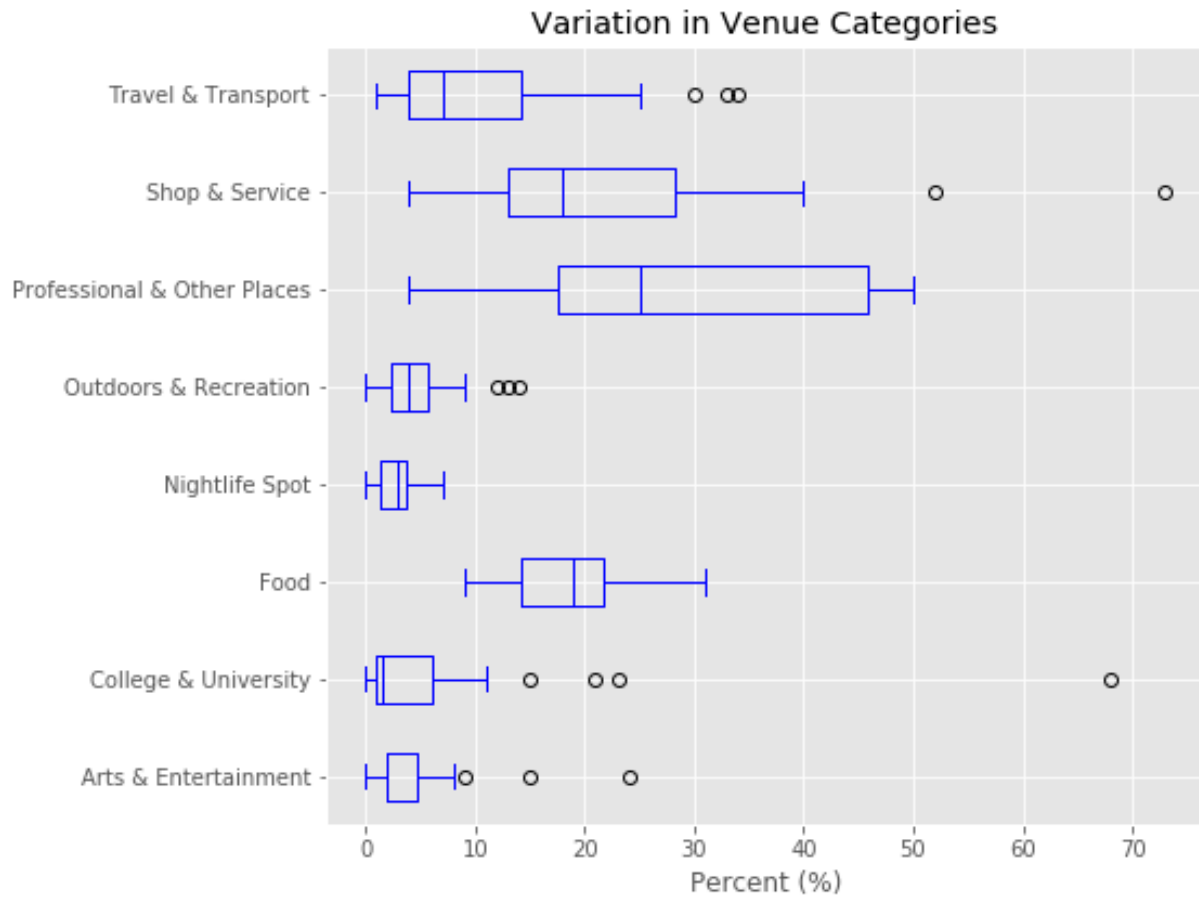


Figure 1. Box plots of venue category counts for all stations. Outliers marked by black circles.

Figure 02 shows the dendrogram of our complete linkage cluster analysis of the station venue category counts. The dashed grey line marks the cut-off resulting in seven clusters, four of which contain only a single member. Two of the single member clusters represent the two extreme outliers identified by the box plots.

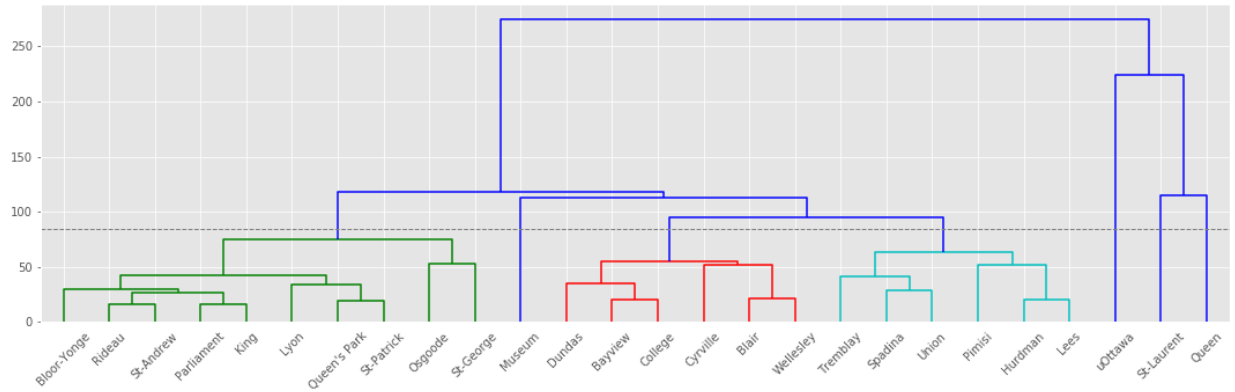


Figure 2. Cluster analysis dendrogram of station venue category counts. Dashed line represents cutoff.

Table 2 shows the total number of stations in each cluster, the number of Ottawa and Toronto stations and the mean values for each venue category. More than 50% of Toronto stations are found in Cluster 0 ($n = 7$), while only three Ottawa stations are found in Cluster 0. Based on the averages I have highlighted the characteristic venue categories for each cluster. For example, Cluster 0 is defined by higher percentages of “Professional & Other Places”, while Cluster 5 is defined by very high “College & University”.

Table 2. Venue category averages for each cluster. Characteristic categories are highlighted for each cluster.

Cluster	Total Stations	Ottawa Stations	Toronto Stations	Arts & Entertainment	College & University	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Shop & Service	Travel & Transport
0	10	3	7	2.7	4.5	17.5	2.6	4.1	44.7	16.5	7.2
1	6	4	2	4.8	5.8	19.2	2.5	7.8	18.5	16	25.2
2	6	3	3	4.2	1.3	22	3.5	4	24.2	33.2	7.7
3	1	0	1	5	0	19	3	2	16	52	4
4	1	1	0	2	1	19	0	1	4	73	1
5	1	1	0	1	68	9	3	5	7	4	3
6	1	0	1	24	23	9	3	7	22	9	3

Figure 3 shows the decision tree output. Leaf Node 0 is made up of nine (out of ten) members of Cluster 0 that have “Professional & Other Places” greater than 34.5%. Leaf Node 1 is made up of five (out of six) members of Cluster 1 having “Travel & Transport” greater than 15.5%. Leaf Node 2 is comprised of all six members of Cluster 2 and one member from each of Cluster 3 and 4, having “Shop & Service” greater than 21.6%. Leaf Node 3 is comprised of one member from Cluster 1 having less than 11% “College & University”. Leaf Node 4 is comprised of one member from each of Cluster 0, 5 and 6 having “College & University” greater than 11%.

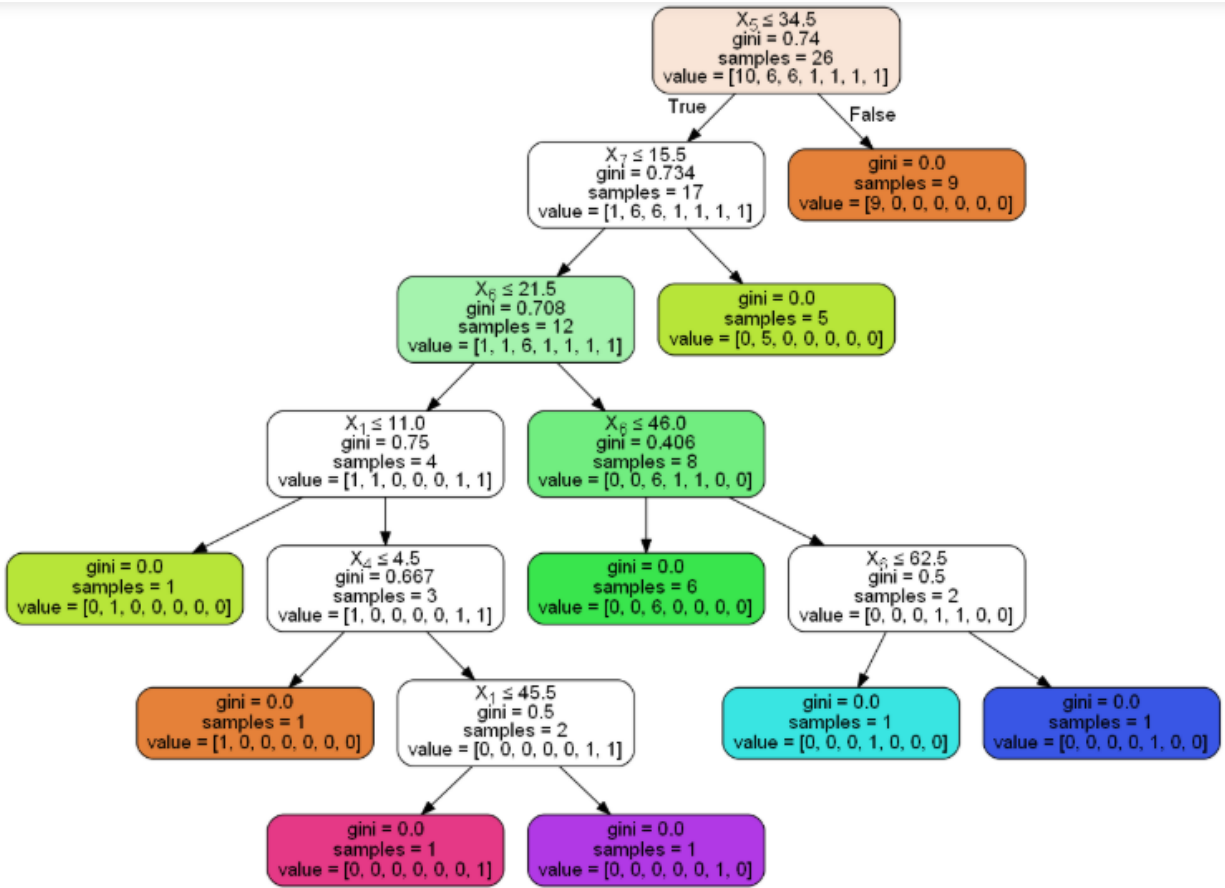


Figure 3. Decision tree model fitted with venue category counts.

Discussion

For the most part the hierarchical cluster analysis and the decision tree model arrived at similar groupings. The decision tree model separated St-George Station from Cluster 0 and Pimisi Station from Cluster 1. Unlike the cluster analysis, the decision tree model grouped St-George with uOttawa Station and Museum Station based on their high percentages of “College & University” and formed a group containing all members of Cluster 2, St-Laurent Station and Queen Station based on their higher percentages of “Shop & Service”.

A significant portion of Toronto stations, twelve out of fourteen, fall within Cluster 0, 1 and 2 highlighting that stations with these proportions of venue categories are characteristic of well-established metro stations. An even mix of Ottawa and Toronto stations is seen in Cluster 1 and 2, both clusters are characterised by an even spread of “Food”, “Professional & Other Places” and “Shop & Service” venues, with Cluster 2 having elevated percentages of “Travel & Transport” venues.

Leaf Node 0 / Cluster 0 contained Stations with high percentages “Professional & Other Place” venues. Toronto stations formed the majority of Leaf Node/Cluster 0, with only three Ottawa stations present. This is also the case for Ottawa stations in general as Toronto stations had on average a greater number of “Professional & Other Places” venues (55 ± 27) than Ottawa stations (31 ± 27). To be more in line with

Toronto stations, Ottawa stations need a greater amount of “Professional & Other Place” venues located in proximity to them.

Two extreme outliers that were comprised of a single dominant venue category were seen in the box plots and the cluster analysis. Seventy-three percent of St-Laurent Station venues are “Shops & Services” while 64% of uOttawa Station venues are “College & University”. These outliers are not surprising as St-Laurent is one of Ottawa’s larger shopping malls and the University of Ottawa is Ottawa’s largest University. Yet this lack of diversity in venue category types at these stations may highlight an opportunity for an increase in other venue category types (e.g., Food, Professional & Other Services) so that the proportion of venue types approaches those of Cluster 0, 1 and 2, in which the majority of Toronto Stations cluster into.

The hierarchical cluster analysis also identified another station that formed its own cluster, Museum Station in Cluster 6. It is characterised by relatively high percent (24%) of “Arts & Entertainment” venues. The closest Ottawa station is Pimisi Station with 15% of its venues being “Arts & Entertainment”. The lack of an Ottawa Station fitting this profile highlights the potential need for such a stop, perhaps during later phases of the light rail expansion a station could be placed closer to “Arts & Entertainment” venues.

Conclusion

Toronto Stations, representing an older and more established metro system, have had time to develop and showcase the typical configuration of venue types at metro stations. Indeed, most Toronto Stations were characterised by even levels of “Food”, “Professional & Other Services” and “Shops & Service” venues, with one cluster having elevated percentages of “Professional & Other Services”. By comparing Ottawa stations to the Toronto stations, we can investigate whether there is room for development in the venue types. “Professional & Other Places” venues were underrepresented at Ottawa stations and represent a target for new venue growth as the Confederation Line matures. In a few cases, Ottawa stations showed a lack of diversity of venue types, being dominated by a single type. This represents an opportunity to increase other types of venues at these stations. Representing a valuable target for entrepreneurs looking for suitable business locations.