Analysis of Commuter Data and Venue Data for Ottawa, ON, Canada

Scott Proulx

June 10, 2020

1 – Introduction

1.1 – Background

As urban populations increase and metropolitan areas of municipalities expand, commuting times between work and home have been increasing globally. Cities across Canada are not immune to these rising statistics. Between 1996 and 2016, the percentage of people working and living within their city's core has decreased drastically. In the Ottawa-Gatineau census metropolitan area (CMA), there has been a drop in city core commuting by over 50% within this timeframe [1]. Indeed the rise of other forms of commuting have come to replace it, including: traditional commuting (outside-to-inside city core), reverse commuting (inside-to-outside city core), commuting between suburbs as well as commuting within a suburb.

Ottawa, Ontario is the capital city of Canada and in 2016 boasted a population of 934,243 as well as a metropolitan population of 1,323,783 [2]. The CMA is divided into 23 distinct wards – a mix of urban and suburban districts, each represented by their own City of Ottawa councillor. Out of the eight largest CMAs in Canada, the Ottawa-Gatineau region exhibited the second-highest median commuting distance in 2016, second only to Toronto [1].

The locations that people choose to live in have a big impact on commuting distances. While real estate costs play a role in where people choose to live, aspects such as nearby venues and amenities are also important to consider. Furthermore, municipal infrastructure such as access to transit plays an integral role in determining the mode of commuting a person may use to get to work each day. The recently opened Confederation Line – a light rail train, with plans for expansion to some of the outer wards – is one example of this type of infrastructure.

1.2 – Problem

For this study, we want to determine if it is possible to cluster groups of communities in the City of Ottawa based on venue availability and commuter data. Additionally, we will explore if these clustered groups have any obvious discernable differences based on venue availability and commuter data.

1.2 – Interest

This study could be of great interest to the City of Ottawa as well as to its transit association OC Transpo. Furthermore, the insights discovered from this analysis could prove useful to other municipalities in Canada and abroad.

2 – Data

2.1 – Description of Data

The data used for this study will be obtained from various sources. First, the geographic data for neighborhoods in Eastern Ontario was scraped from Wikipedia [3], and the neighborhoods belonging to Ottawa were be obtained. The geographical coordinates for each neighborhood were determined using GeoPy. Shape and map data for the wards in Ottawa were obtained from publicly available data on Open Ottawa and will be used to project the ward territories onto a Folium map [4]. The Foursquare Places API was used to obtain venue categories within a defined radius of each neighborhood's geographic center [5]. Finally, census data was obtained from on Open Ottawa, derived from the Statistics Canada 2016 Census for the wards of the City of Ottawa [6]. This data includes counts for the main mode of commuting, and commuting duration for each ward. This data was integrated to perform clustering (K-means and DBSCAN) to determine groups of neighborhoods based on venue availability and commuter data. From this, we were able to draw inferences about the different community groups, with implications for future transit and city planning projects.

2.2 – Cleaning the Data

Neighborhood data scraped from the web required vigorous cleaning. Firstly, only postal codes associated with Ottawa were desired (the website includes postal codes from all of Eastern Ontario). Secondly, there were many discrepancies in the way that neighborhoods were listed, so the formatting had to be cleaned to separate neighborhoods with backslashes as opposed to commas, spaces, and other punctuation. Ward data, particularly shape data to superimpose onto the map, was parsed from a GeoJSON file. To assign each neighborhood to a ward, the shapely package was used to determine whether each coordinate fell within the bounds of the polygon shape of each ward. Census data – including commuting mode and commuting time, were normalized to the total population of each ward to avoid skewing by population size. For venues, frequency in each neighborhood, all venues within a 1-km radius of the neighborhood (up to a limit of 100 venues) were obtained. Categories were consolidated to only include 14 different venue types, which were then normalized to the total number of venues in each neighborhood.

2.3 – Data Structure

Each row in the data set is associated with a single postal code or neighborhood. Each column in the data set is associated with a single feature of the data set. The features of the data can be seen in Table 1. The final data set delivered as input to the machine learning models had 35 rows (neighborhoods) and 25 features.

Table 1 – Features of data set

Feature	Category	Description
Car, truck, van - as a driver	Commuting Mode	Frequency of driving to work as the driver
Car, truck, van - as a passenger	Commuting Mode	Frequency of driving to work as a passenger
Public transit	Commuting Mode	Frequency of taking public transit to work
Walked	Commuting Mode	Frequency of walking to work
Bicycle	Commuting Mode	Frequency of biking to work
Other method	Commuting Mode	Frequency of taking other means to work
Less than 15 minutes	Commuting Time	Frequency of a commute less than 15 minutes
15 to 29 minutes	Commuting Time	Frequency of a commute between 15 and 29 minutes
30 to 44 minutes	Commuting Time	Frequency of a commute between 30 and 44 minutes
45 to 59 minutes	Commuting Time	Frequency of a commute between 45 and 59 minutes
60 minutes and over	Commuting Time	Frequency of a commute 60 minutes or over
Alcohol, Marijuana and Tobacco	Venue Type	Frequency of alcohol, marijuana, or tobacco stores in area
Bars, Pubs and Clubs	Venue Type	Frequency of bars, pubs, or clubs in area
Business and Industry	Venue Type	Frequency of businesses and industry in area
Coffee and Cafes	Venue Type	Frequency of coffee shops and cafes in area
Entertainment	Venue Type	Frequency of entertainment venues in area
Food and Groceries	Venue Type	Frequency of food and grocery stores in area
Health and Fitness	Venue Type	Frequency of health and fitness venues in area
Hospitality	Venue Type	Frequency of hospitality-related venues in area
Medical and Pharmacy	Venue Type	Frequency of clinics, pharmacies, and hospitals in area
Other	Venue Type	Frequency of other venues in area
Outdoor Venues	Venue Type	Frequency of outdoor venues in area
Public Amenities	Venue Type	Frequency of public amenities such as libraries in area
Restaurant and Eateries	Venue Type	Frequency of restaurants and eateries in area
Shopping	Venue Type	Frequency of shopping venues in area

3 – Methodology

3.1 – Implementation of Geographical Data

The shape files from the Ottawa ward data were superimposed as polygons onto a Folium map of Ottawa. In addition, the neighborhoods corresponding to postal codes in Ottawa were also superimposed onto this map. The image can be seen in ***Figure 1***, and many subsequent figures and images were built upon this figure. Evidently, most of the neighborhoods are in the wards surrounding the city center, whereas outer wards are sparsely populated with neighborhoods. In fact, one such ward – Rideau-Gouldbourn (Ward 21) does not have a City of Ottawa postal code located within it, and instead has the standard rural “K0A” National Capital Region assignment. This postal code was excluded from analysis as the geographic location of the postal code is irrelevant to the exact geographic distributions of the communities it is assign to.

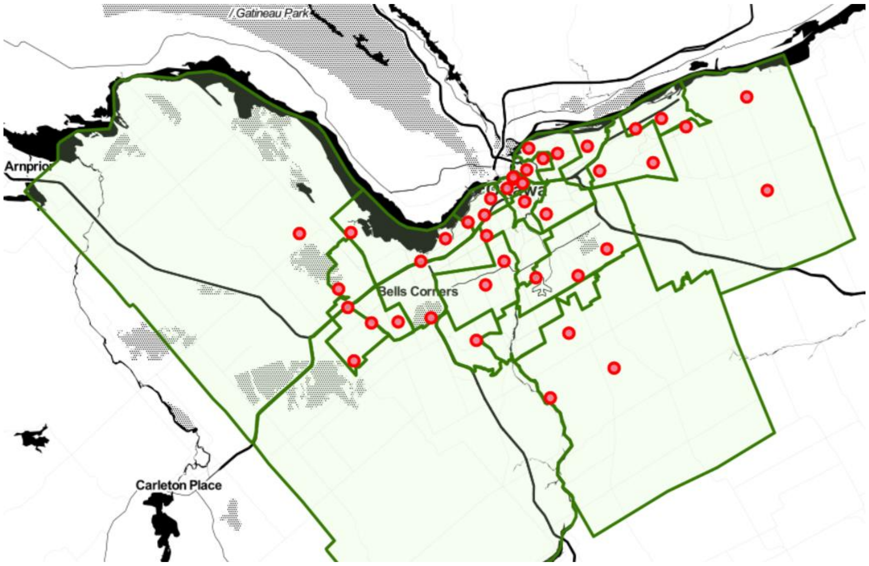

Figure 1 – Superimpositions of neighborhoods and wards on a map of Ottawa

3.2 – Exploratory Analysis of Commuter Data in Ottawa Wards

Exploring the various commuting modes in the wards of Ottawa was of interest to see how people of the many wards and hence their neighborhoods, get to work. The prevalence of driving (Figure 2), taking public transit (Figure 3), walking (Figure 4) and biking (Figure 5) to work were displayed as choropleth maps of the Ottawa wards, while driving as a passenger and taking other modes were not included for simplicity. Some interesting trends can be seen from these images. As was expected, wards on the outer, rural edges of the city tend to be more driver-oriented, while wards on the suburban edges of the city have a mix of drivers and those who take public transit to work. People living in the inner wards closer to downtown were most likely to walk, bike, or take public transit to work, and were unlikely to be drivers. This make a lot of sense, as most people who work in Ottawa for the federal government (the largest employer in Ottawa) work downtown. Another large hub of employment in Ottawa is Kanata – a Western suburban area consisting of two wards (Kanata North and Kanata South) – which is home to many technology companies. It appears that people who live in Kanata however are not biking or walking to work – likely due to the suburban nature of the areas.

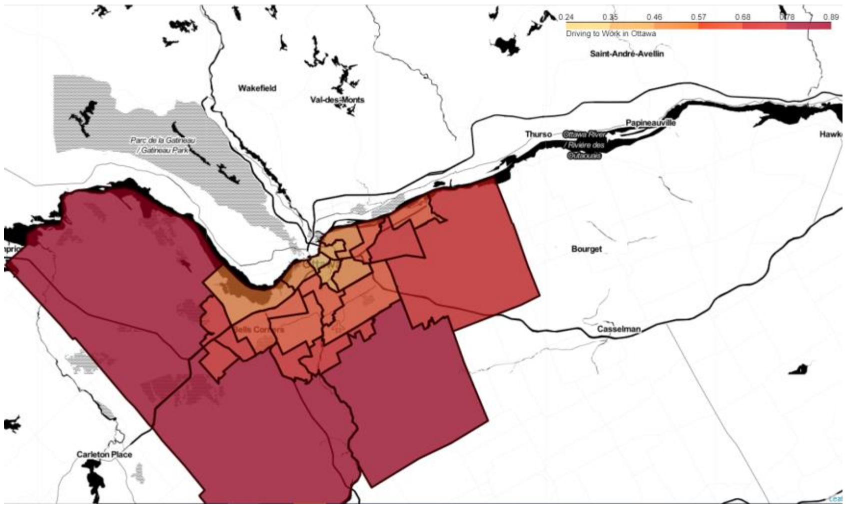

Figure 2 - Prevalence of driving to work in the wards of Ottawa

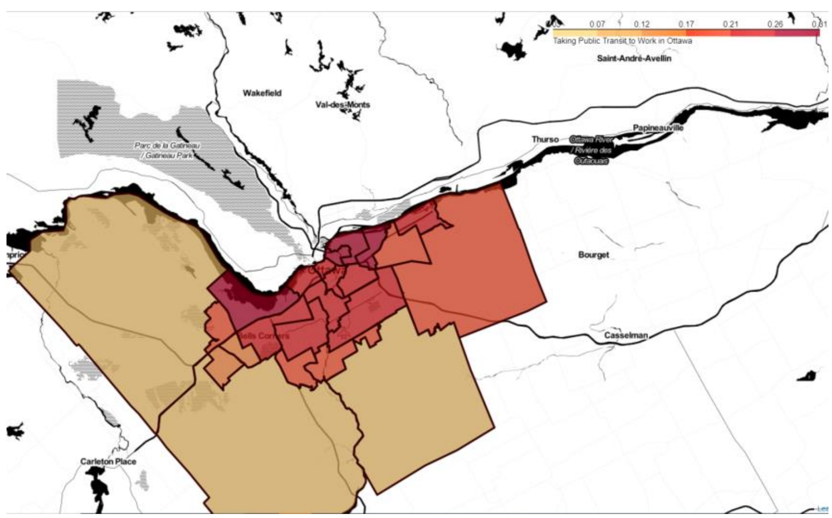

Figure 3 - Prevalence of taking public transit to work in the wards of Ottawa

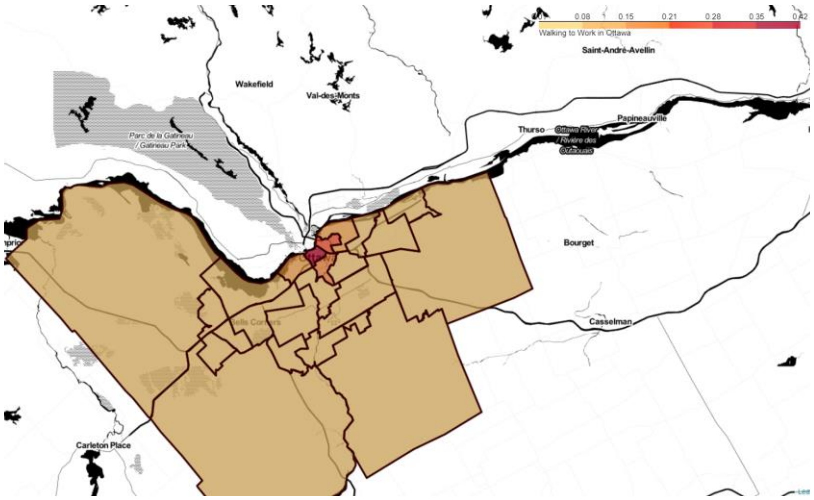

Figure 4 - Prevalence of walking to work in the wards of Ottawa

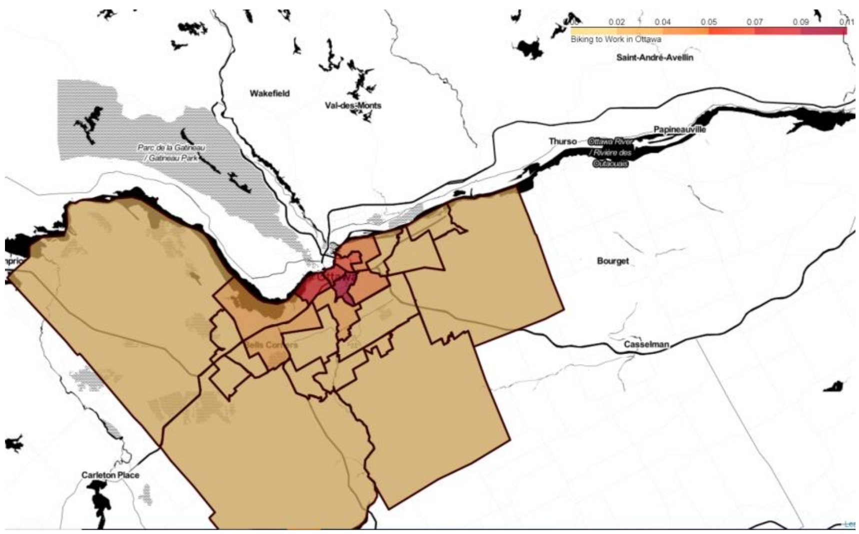

Figure 5 - Prevalence of biking to work in the wards of Ottawa

Another integral feature of a daily commute to work is the time spent commuting each way. For simplicity, only two of the time frames – 60+ minutes (Figure 6) and less than 15 minutes (Figure 7) were plotted on choropleth maps. It would be expected that if everybody was working downtown, then the further people lived from the city center, the longer their commutes would be. Some rural wards such as Cumberland and West Carleton-March exhibit these behaviours, whereas the other rural wards do not. In addition, hour-long commutes are virtually non-existent for those living in central wards of the city. Shorter commute distributions are arguably more interesting. Evidently, downtown wards show the strongest distributions of commutes under 15 minutes, while Kanata North also has a strong frequency of these shorter commutes. Therefore, it appears that wards that are driver-centric tend to have longer commute times, while alternate modes of travel typically lead to shorter commute times.

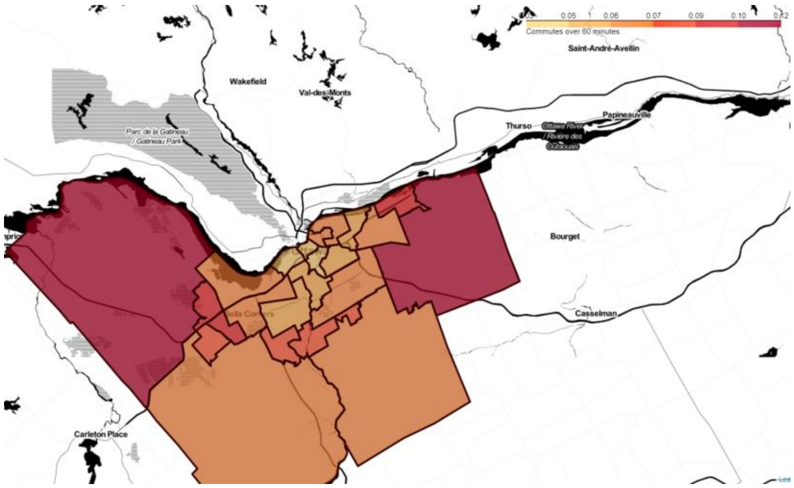

Figure 6 - Prevalence of commutes over 60 minutes in the wards of Ottawa

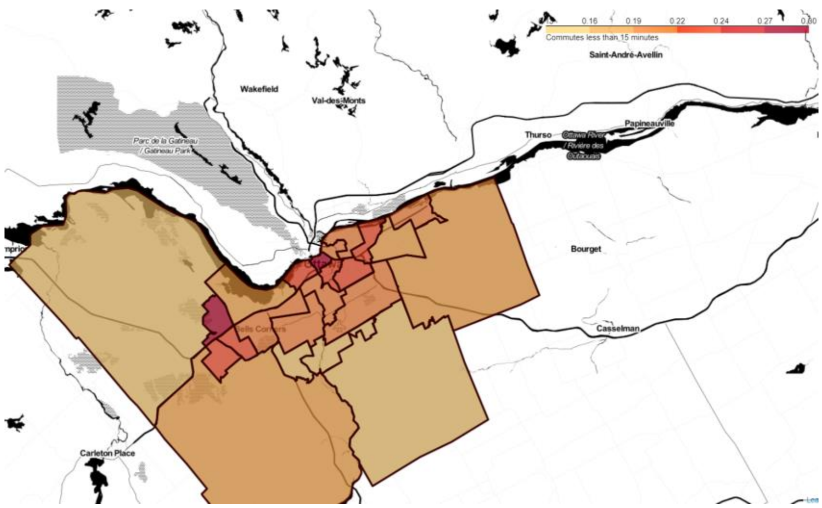

Figure 7 - Prevalence of commutes under 15 minutes in the wards of Ottawa

3.3 – Exploratory Analysis of Venue Data in Ottawa Neighborhoods

Prior to grouping venue categories together, there were a total of 208 unique venue categories. This is a problematic number of categories for plotting, so this list was distilled into 14 unique venue category groups. The distribution of venue categories in each ward can be seen in Figure 8. While there is heterogeneity in the venue distribution between the wards, categories of restaurants and eateries, coffee and cafes, and shopping all have dominant presences in these spaces.

It should be noted that four wards – West Carleton-March (Ward 5), Gloucester-Southgate (Ward 10), Rideau-Goulbourn (Ward 21) and Gloucester-South Nepean (Ward 22) are not included as no postal codes are geographically located within the boundaries of these wards. Each of these wards does have a postal code geographically situated on the boundaries of bordering wards, so a solution could have been to increase the radius at which venues are found. However, to make a fair comparison, the same radius should be used in each neighborhood and therefore in each ward. A radius of 1 km was chosen, and despite the postal codes situated at the edges of bordering wards, the wards with no neighborhoods within them resulted in having no venues found. Further increasing the radius at which venues are assigned to a neighborhood risks leading to overlapping venues in neighborhoods that are located closer together, such as those downtown.

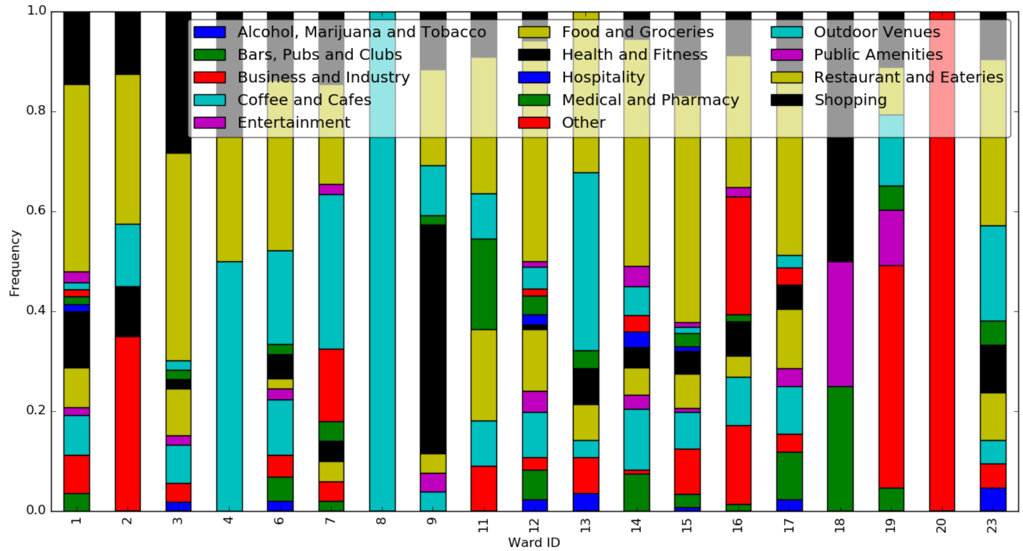

Figure 8 – Mean venue distribution in the wards of Ottawa

In order to visualize the distributions of some of the top categories of venues in Ottawa's wards, the frequencies of restaurants and eateries (Figure 9), shopping venues (Figure 10) as well as coffee shops and cafes (Figure 11) were plotted on choropleth maps. Restaurants and coffee shops would be expected to be located mostly downtown, which for the most part appears to be the case. However, suburban outliers with relatively high venue densities also appear, such as Barrhaven for restaurants, and Stittsville for coffee shops and cafes. It should be noted that coffee shops and cafes include to-go coffee chains such as Tim Hortons, which is an extremely common, immensely popular fast food chain widely distributed in suburban and rural communities alike. It was expected that shopping would be more likely assigned to wards surrounding, but not in downtown wards, which is what appears to have happened. This is likely due to the presence of "big-box" stores in these areas as well as shopping malls. While downtown wards have lots of shopping, this data is only concerned with how many of the venues in the ward are shopping venues compared to all the other venues. As such, because there are plenty of restaurants and coffee shops downtown, the prevalence of shopping venues is much lower than in wards with lower restaurant and coffee shop densities.

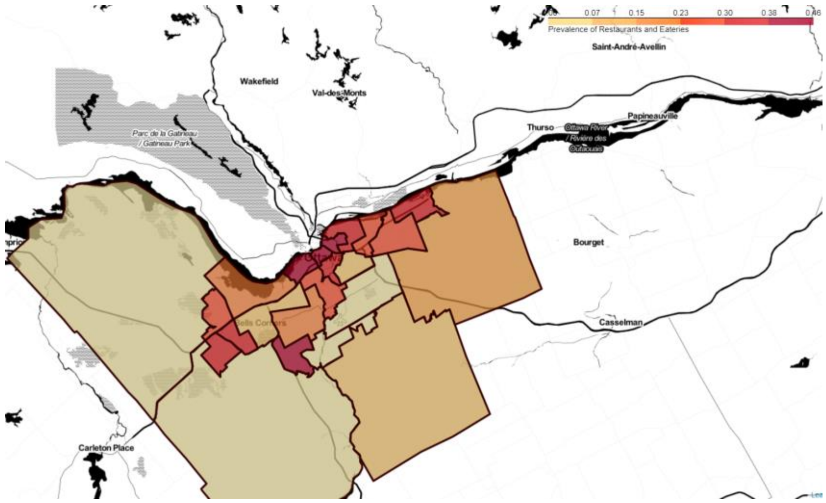

Figure 9 – Distribution of restaurants and eateries in Ottawa

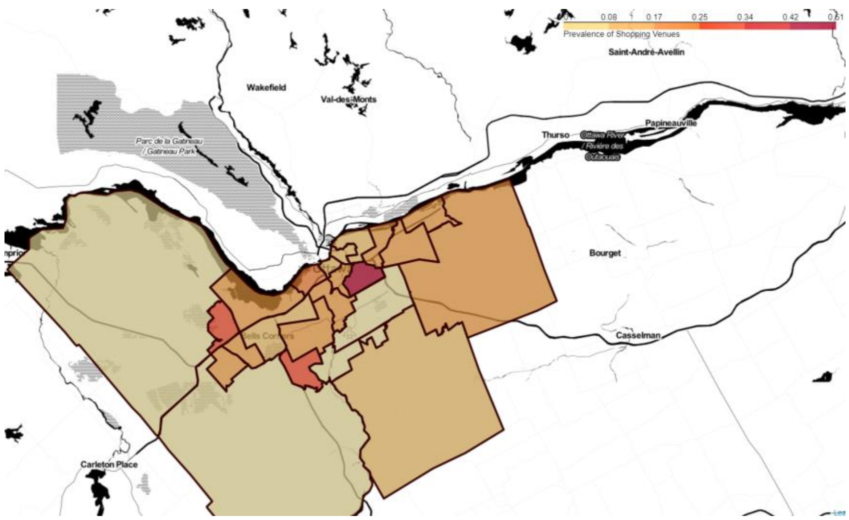

Figure 10 – Distribution of shopping venues in Ottawa

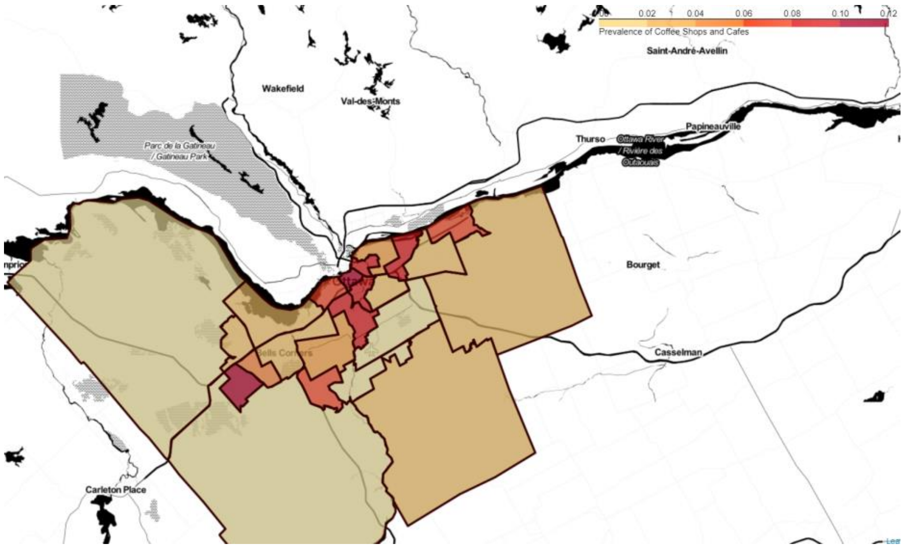

Figure 11 – Distribution of coffee shops and cafes in Ottawa

3.4 – Correlation of all Features

The correlations between all features, converted to numerical formats, are shown in Figure 12. Most of the features do not exhibit strong correlations, but there are many that do – both positive and negative. For instance, there is a strong correlation between people who bike to work and people whose commutes take 15 to 29 minutes. Additionally, walking neighborhoods also appear to have a high density of hospitality venues, such as hotels – this makes sense, as many hotels are located downtown. Some weaker positive correlations include the prevalence of business and industry venues (ex: gas stations, banks, IT services) and commutes of 30 to 44 minutes – which suggests suburban neighborhoods. On the negative correlation side, there are many examples. People who drive to work are unlikely to live near hospitality venues, which again confirms that people who drive do not live downtown. Additionally, densities of outdoor venues, such as parks, are negatively correlated with restaurants and eateries. Restaurant and eatery prevalence also exhibits a negative correlation with commutes of 30 to 44 minutes, suggesting that suburban wards are not restaurant-dense.

Most of these relations are based on commute mode and commute time, which was discussed previously. Indeed, it appears that commute mode and commute time are strongly correlated with each other, both positively and negatively. People who drive to work are strongly correlated to having longer commute times, whereas people who take public transit, walk or especially bike to work have shorter commute times.

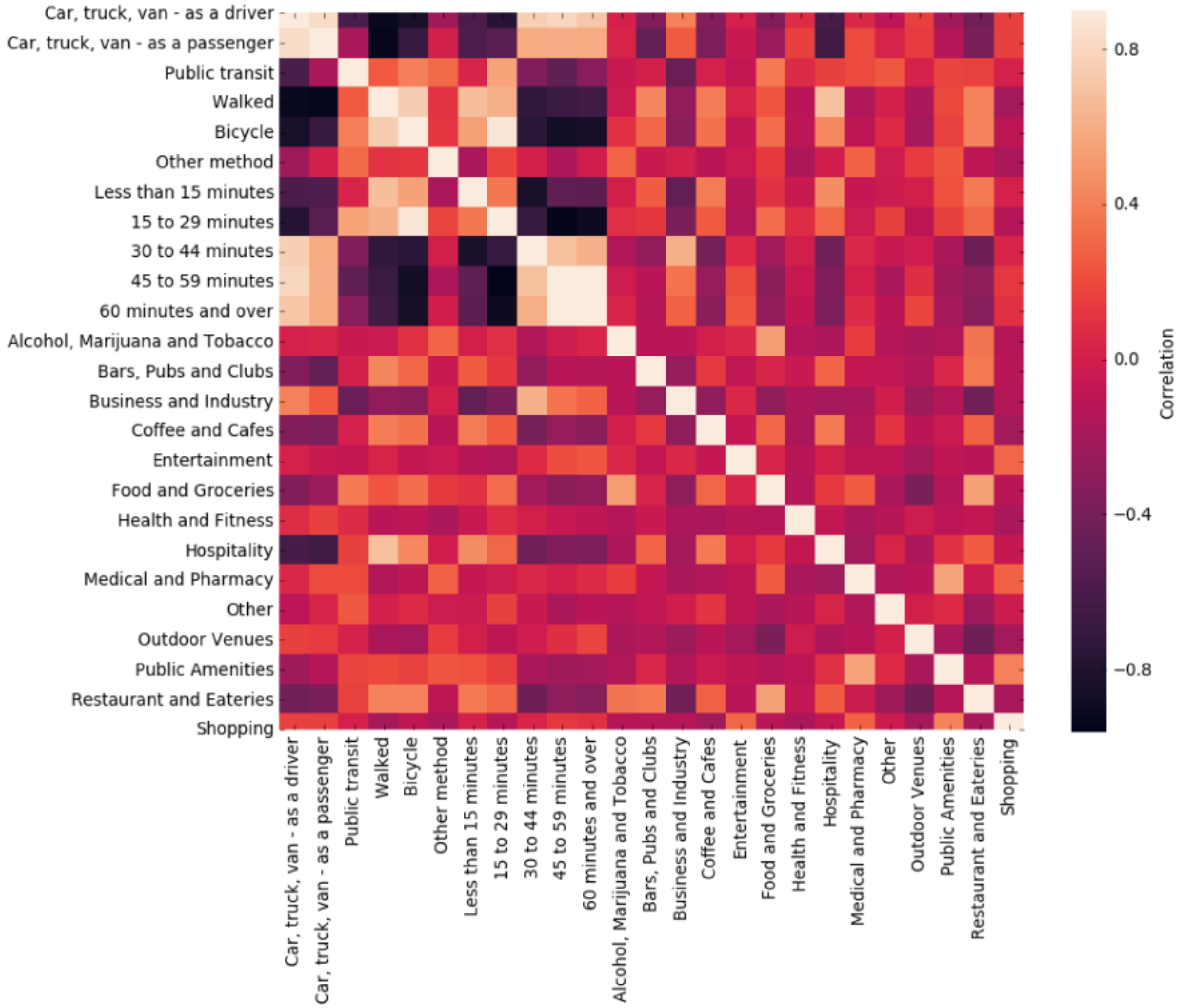

Figure 12 – Correlation plot for all features in Ottawa venue and commuting data

3.5 – Clustering Neighborhoods

The data for this problem lends itself to an unsupervised learning and clustering problem. Therefore, two commonly used methods for this type of problem – K-Means Clustering and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) were used. K-Means is the most used method here and is generally quite fast; however, the number of clusters (k) must be indicated in advance, and the clusters are all hyperspheres. The issue of picking the number of clusters may be addressed by finding the elbow point from a plot of the distortion of the data against different values of k. This hyperparameter optimization can be seen in Figure 13, although no obvious elbow point exists. Therefore, a value of k = 5 was chosen as the best guess. DBSCAN does not require the number of clusters to be assigned prior to calculation and can cluster amorphous shapes other than hyperspheres. In addition, not all points require assignment to a cluster, and may be determined outliers if they do not fit within a designated cluster (K-Means may simply assign an outlier its own cluster or assign it to a pre-existing one). A similar procedure was performed for DBSCAN with values of Epsilon plotted against each neighborhood (Figure 14). The elbow point in this case was slightly easier to determine and was set to a value of Epsilon = 5.2, and the minimum number of samples required for a cluster was set to two.

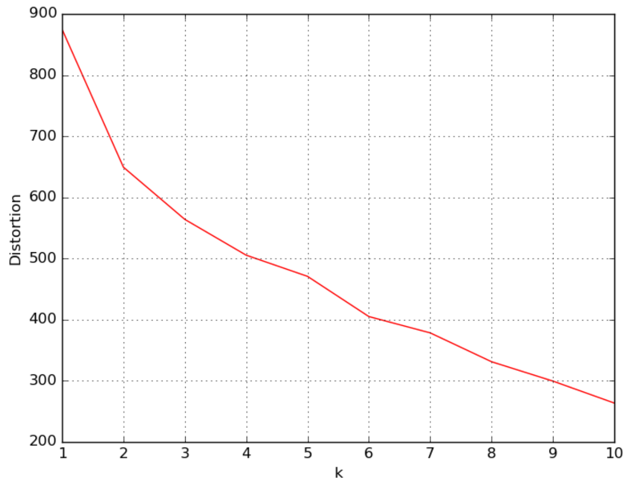

Figure 13 - Determination of number of clusters for K-Means

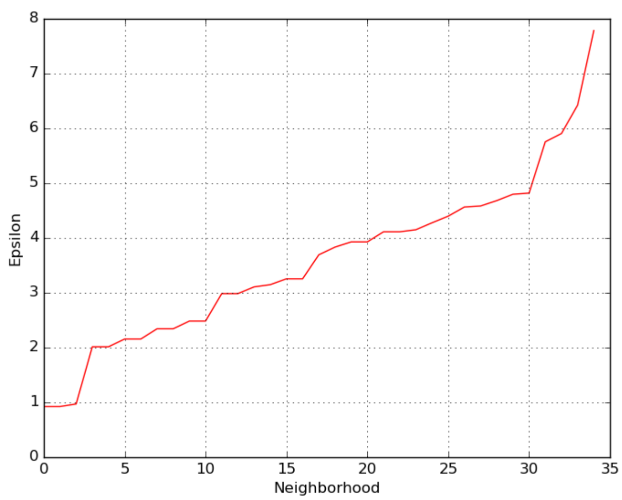

Figure 14 - Determination of Epsilon for DBSCAN

4 – Results

The clustering results for neighborhoods in Ottawa have been plotted on the boundaries of the wards for K-Means (Figure 15) and DBSCAN (Figure 16), and label assignment by both methods has been tabulated in Table 2. It is evident that clustering assignments are different for both methods. For instance, K-Means assigned neighborhoods to five different clusters, whereas DBSCAN was only able to find a single cluster for all neighborhoods (not including outliers). There also does not appear to be any relation to the outlier neighborhoods from DBSCAN and their associated K-Means labels. For instance, the outlier neighborhoods determined by DBSCAN consist of two found downtown, one in a suburban neighborhood and one in a rural ward. It is also difficult to determine the metrics in clustering for the K-Means assignments by visual inspection. However, there are many features that the model is composed of (25 features – 14 of them for venues and 11 for commuter data), and so patterns are difficult to assign by intuition. Unfortunately, six neighborhoods could not be assigned a label for either method due to insufficient venues located in proximity to the coordinates of their associated postal codes.

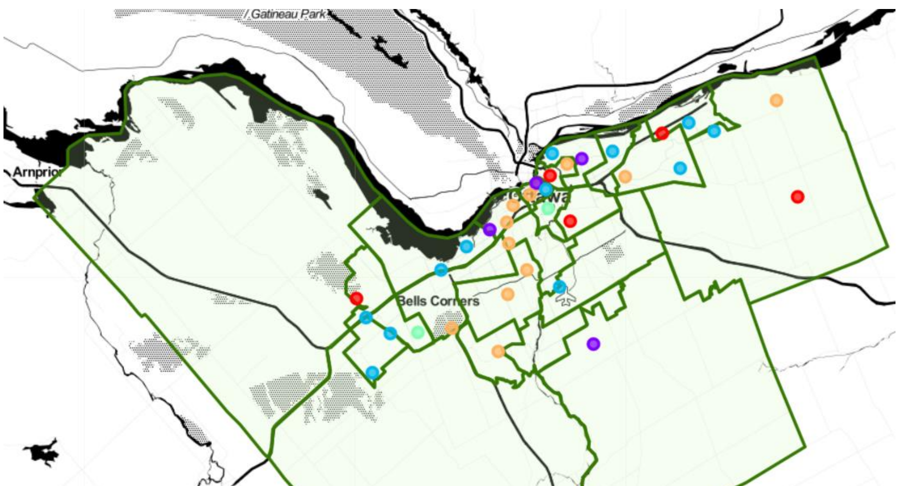

Figure 15 - K-Means clustering results superimposed on map of Ottawa

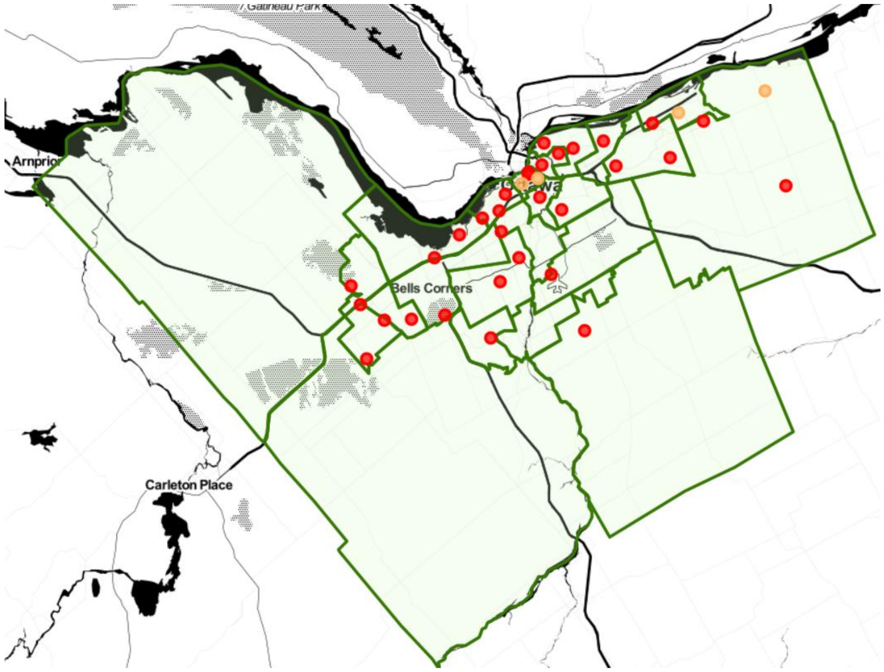

Figure 16 - DBSCAN clustering results superimposed on map of Ottawa

Table 2 - Clustering assignment for all neighborhoods by K-Means and DBSCAN

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Commute		Label	
				Mode	Time (min)	K-Means	DBSCAN
Navan	Business and Industry	Shopping	Restaurant and Eateries	Driving	15 - 29	0	0
Orleans	Restaurant and Eateries	Health and Fitness	Business and Industry	Driving	15 - 29	0	0
Alta Vista / Billings Bridge	Shopping	Public Amenities	Medical and Pharmacy	Driving	15 - 29	0	0
Lower Town / Byward Market / Sandy Hill / University of Ottawa	Restaurant and Eateries	Bars, Pubs and Clubs	Coffee and Cafes	Driving	15 - 29	0	0
Marchwood	Outdoor Venues	Shopping	Restaurant and Eateries	Driving	< 15	0	0
Highland Park / McKellar Park Westboro Glabar Park Carlingwood	Restaurant and Eateries	Shopping	Health and Fitness	Driving	15 - 29	1	0
Overbrook / Forbes / Manor Park / Viscount Alexander Park / Finter Quarries	Restaurant and Eateries	Food and Groceries	Business and Industry	Driving	15 - 29	1	0
Downtown	Restaurant and Eateries	Coffee and Cafes	Shopping	Walking	15 - 29	1	0
South Gloucester	Business and Industry	Shopping	Restaurant and Eateries	Driving	30 - 44	1	0
Queenswood	Outdoor Venues	Restaurant and Eateries	Medical and Pharmacy	Driving	15 - 29	2	-1 (Outlier)
Centretown	Restaurant and Eateries	Outdoor Venues	Medical and Pharmacy	Driving	15 - 29	2	-1 (Outlier)
Fallingbrook	Restaurant and Eateries	Shopping	Food and Groceries	Driving	15 - 29	2	0
Britannia Whitehaven / Bayshore / Pinecrest	Outdoor Venues	Shopping	Other	Driving	15 - 29	2	0
Bells Corners / Arlington Wood / Redwood / Qualicum / Crystal Beach	Restaurant and Eateries	Medical and Pharmacy	Food and Groceries	Driving	15 - 29	2	0
Beacon Hill / Cyrville / Carson Grove	Restaurant and Eateries	Shopping	Health and Fitness	Driving	15 - 29	2	0
Katimavik-Hazeldean / Glen Cairn	Outdoor Venues	Restaurant and Eateries	Health and Fitness	Driving	15 - 29	2	0
Rockcliffe Park / New Edinburgh	Restaurant and Eateries	Coffee and Cafes	Bars, Pubs and Clubs	Walking	15 - 29	2	0
Stittsville	Outdoor Venues	Coffee and Cafes	Restaurant and Eateries	Driving	15 - 29	2	0
Heron Gate / Heron Park / Riverside Park / Hunt Club / Riverside South / YOW	Other	Business and Industry	Restaurant and Eateries	Driving	15 - 29	2	0
Terry Fox / Palladium	Restaurant and Eateries	Shopping	Outdoor Venues	Driving	15 - 29	2	0
Chapel Hill South / Blackburn	Business and Industry	Shopping	Outdoor Venues	Driving	15 - 29	2	0
Bridlewood	Restaurant and Eateries	Outdoor Venues	Shopping	Driving	15 - 29	3	0
The Glebe / Old Ottawa South / Old Ottawa East / Carleton University / Dow's Lake area	Restaurant and Eateries	Shopping	Food and Groceries	Driving	15 - 29	3	0
Cumberland	Restaurant and Eateries	Coffee and Cafes	Outdoor Venues	Walking	15 - 29	4	-1 (Outlier)
Dalhousie Ward	Restaurant and Eateries	Health and Fitness	Business and Industry	Driving	15 - 29	4	-1 (Outlier)
Government of Canada: Ottawa and Gatineau offices (partly in QC)	Restaurant and Eateries	Shopping	Health and Fitness	Driving	15 - 29	4	0
Blackburn Hamlet / Pine View / Sheffield Glen	Shopping	Entertainment	Business and Industry	Driving	15 - 29	4	0
Queensway / Copeland Park / Central Park / Bel Air Carleton Heights	Restaurant and Eateries	Shopping	Health and Fitness	Driving	15 - 29	4	0
Eastern Nepean: Fisher Height / Parkwood Hills / Borden Farm Pine Glen	Health and Fitness	Outdoor Venues	Shopping	Driving	15 - 29	4	0
Centrepointe / Meadowlands / City View / Craig Henry / Tangelwood / Grenfell Glen / Davidson Heights	Restaurant and Eateries	Shopping	Food and Groceries	Driving	15 - 29	4	0
Barrhaven	Restaurant and Eateries	Food and Groceries	Medical and Pharmacy	Driving	15 - 29	4	0
Vanier / McKay Lake area	Restaurant and Eateries	Bars, Pubs and Clubs	Coffee and Cafes	Walking	15 - 29	4	0
Fallowfield Village / Cedarhill Estates / Orchard Estates	Outdoor Venues	Shopping	Restaurant and Eateries	Driving	15 - 29	4	0
Civic Hospital / Island Park / Hintonburg / Mechanicsville / Champlain Park	Restaurant and Eateries	Food and Groceries	Coffee and Cafes	Driving	15 - 29	4	0
Westboro / Carlington	Restaurant and Eateries	Shopping	Business and Industry	Driving	15 - 29	4	0
Riverview / Hawthorne / Canterbury / Hunt Club Park	n/a	n/a	n/a	Driving	15 - 29	n/a	n/a
Beaverbrook / South March	n/a	n/a	n/a	Driving	15 - 29	n/a	n/a
Manotick	n/a	n/a	n/a	Driving	30 - 44	n/a	n/a
Greely	n/a	n/a	n/a	Driving	30 - 44	n/a	n/a
Blossom Park / Greenboro / Leitrim / Findlay Creek	n/a	n/a	n/a	Driving	15 - 29	n/a	n/a
North March	n/a	n/a	n/a	Driving	15 - 29	n/a	n/a

5 – Discussion

Many points can be made about the clustering results. First, the lack of a clear elbow point for the optimization of the number of clusters for K-Means clustering led to a rather arbitrary assignment of k = 5. Therefore, the cluster assignment is thrown into question. The only way to truly decrease distortion is to continue increasing the value of k; however, this can lead to some uninteresting results, and possibly no clustering at all. It is likely to be an issue with sample size, as a larger number of samples could allow a higher number of clusters with larger numbers of samples contained within them. This was not as big of an issue for DBSCAN, as hyperparameter optimization was easier by visual inspection.

To rectify this issue for K-Means, we would indeed need more data points. However, Ottawa can be considered a medium-sized city compared to those such as Toronto, Montreal, or New York City, and it therefore does not have the same density or the same number of postal code regions. However, one could immensely increase the number of samples within a given postal code region by using the last three digits of the postal code. We only addressed overall postal code regions, which encompass the first three digits of a postal code, whereas a full postal code for a Canadian address has the format of A1B 2C3. By harnessing those last three digits, we could have greatly increased our density and number of samples.

Another issue that affected our number of samples is venue proximity in regions sparsely populated with venues. This was a big enough issue so as to exclude six neighborhoods which each did not have a single venue within the designated radius from the postal code. At first, a radius of 500 m was used, and even more neighborhoods were excluded. Upon increasing the radius to 1,000 m, only six neighborhoods were excluded. Evidently, there is a distance that we could increase the radius to so that the FourSquare Places API could find at least one venue within the set radius. While this makes sense for rural regions where driving further distances to visit restaurants or shopping centers is common, it wouldn't make much sense for venues located closer to the city center, and would lead to a lot of overlap between neighborhoods. A possible solution to this issue would be to use a sort of dynamic venue radius that increases the further the neighborhood is located away from Ottawa's city center.

Looking at the clustering results, there is a disparity between DBSCAN and K-Means. While K-Means assigned neighborhoods to each of the pre-determined clusters, DBSCAN did not see any differences between neighborhoods (aside from outliers). If perhaps the outliers belonged to their own unique cluster assigned by K-Means, or were individually assigned to single-neighborhood clusters by K-Means, then one could argue that DBSCAN is superior to K-Means, as the latter is simply trying to fill in the number of pre-defined clusters provided by the user. However, this does not appear to be the case and so other reasons must be responsible. The results must be based on the differences between the algorithms. It is possible that no intra-structural clusters can be found within the structures of clusters.

It is difficult to see patterns in the data by visual inspection. For instance, one could assume that neighborhoods found in suburban wards could occupy one cluster, and those in rural wards another cluster, and then many of the urban wards could form several clusters. However, this does not appear to be the case for K-Means. With 25 features, it is simply too difficult to conclusively state what makes neighborhoods in a given cluster similar. However, if a person who spent their entire lives in Bridlewood wanted to move to a different neighborhood, they may find comfort in moving to the Glebe or perhaps to Dow's Lake. This study therefore shows that unsupervised learning can uncover patterns that are extremely difficult to interpret by traditional means.

6 – Conclusions and Future Work

6.1 – Conclusions

Venue and commute mode / time data for the City of Ottawa was compiled into an unlabeled data set with 25 features. While restaurants, shopping venues and coffee shops proved to be some of the most popular venues in the city, the distribution of 14 unique venue categories in each ward and neighborhood was relatively heterogeneous. Commuter data generally followed expected trends, with wards further from the city center showing longer commute times by driving, whereas wards closer to the city center were shorter and by means such as walking, public transportation or biking. K-Means assigned neighborhoods two five different clusters, while DBSCAN was only able to find a single cluster with four outlier neighborhoods.

6.2 – Future Work

Increasing the number of data points in this data set is important to increase model accuracy. Therefore, two different methods discussed can be used to accomplish this goal. First, use of full postal codes as opposed to using just the first three letters would drastically increase "neighborhood" density. In addition, implementing a venue proximity radius that is proportional to a neighborhood's distance from the city's center would help to alleviate low-density issues in rural wards.

7 - References

[1] K. Savage, "Results from the 2016 Census: Long commutes to work by car," *Insights Can. Soc.*, no. 75, pp. 1–12, 2019.

[2] CBC News, "Ottawa's population blooms to 1M," *CBC News*, Ottawa, ON, 11-Jun-2019.

[3] Wikipedia, "List of postal codes of Canada: K," *Wikipedia*, 2020. [Online]. Available: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_K. [Accessed: 01-Jun-2020].

[4] City of Ottawa, "Wards," *Open Ottawa*, 2019. [Online]. Available: https://open.ottawa.ca/datasets/wards?geometry=-78.414%2C44.911%2C-73.188%2C45.587. [Accessed: 01-Jun-2020].

[5] Foursquare.com, "Foursquare Places API." .

[6] City of Ottawa, "2016 Census Ward Data," *Open Ottawa*, 2020. [Online]. Available: https://open.ottawa.ca/datasets/2016-census-ward-data-1. [Accessed: 01-Jun-2020].