Battle of the Neighborhoods - House Buying Search

1. Introduction

1.1 Background

As an engineer working 5+ years in the automotive industry since graduating from post secondary school, I am now looking to take the next step in my life which is to purchase a house. At first when I started to look into purchasing a house I was only focused on the house itself, but I soon realized that if I wanted to live there long term I also need to look at the surrounding amenities as I did not want a *commuter* house.

When moving into an area it is important that the house meets your needs while the neighborhood also suits your lifestyle. As a young adult the following list is what I would be looking for in my search for a house:

1. Affordable neighborhood
2. Fitness amenities in the surrounding area
3. Restaurants and coffee shops within walking distance

1.2 Problem Description

Purchasing a house is a very big decision to make. This decision has many obstacles as there are many unknowns which includes the following:

- What city to purchase the house in?
- What are the different neighborhoods in the city?
- Where is each neighborhood within the city?
- How do housing prices vary with location?
- What kind of popular venues are around each neighborhood?
- How many available house listings are there in each location?

With all these unknowns it would take a significant amount of time to answer these questions by combing the internet. Time is definitely a luxury in my current situation with the high number of hours I work every week so an efficient method to summarize findings would be ideal compared to spending hours on end combing the internet with findings being manually written down.

For this project we will be comparing neighborhoods in the cities of Hamilton and Burlington to determine the ideal location for me to purchase a house.

1.3 Target Audience

This project will serve **young adults** to help make an informed decision on where to purchase a house in Hamilton and Burlington by providing an in-depth analysis of neighborhood housing costs, available housing listings, and surrounding venues.

2. Data

Two cities will be analyzed in this project:

- Hamilton
- Burlington

I will be creating separate sets of datasets for each city for my analysis. This will allow me to gain an understanding of the neighborhoods based on their respective cities as I am interested in both the individual neighborhoods in addition to the cities as a whole.

2.1 Neighborhood Datasets

Datasets which includes information of neighborhoods for each city will be created by web scraping the table for L postal codes in Canada in Wikipedia. The coordinates of each neighborhood will be determined by their respective postal code. Several neighborhoods were found to share the same postal code, these neighborhoods were kept grouped together as they will reflect the same location based on the geographical coordinates in our analysis.

Burlington neighborhood dataset:

	PostalCode	City	Neighborhood	Latitude	Longitude
66	L7L	Burlington	Corporate, Elizabeth Gardens , Longmoor , Pinedale , Shoreacres	43.384449	-79.767496
75	L7M	Burlington	Alton Village , Headon Forest , Millcroft , Palmer , Tansley Woods	43.417438	-79.857794
82	L7N	Burlington	Dynes, Roseland	43.351052	-79.783412
91	L7P	Burlington	Brant Hills , Tyandaga , Mountainside	43.372384	-79.879311
100	L7R	Burlington	Downtown , Plains	43.353290	-79.804560
109	L7S	Burlington	Maple	43.317173	-79.810205
118	L7T	Burlington	Aldershot Central , Aldershot South	43.313290	-79.855320

Hamilton neighborhood dataset (Not all data shown due to large size of dataset):

	PostalCode	City	Neighborhood	Latitude	Longitude
6	L9A	Hamilton	Crerar , Bruleville , Hill Park , Inch Park , Centremount , Balfour , Greeningdon , Jerome	43.226064	-79.870944
13	L8B	Hamilton	Waterdown	43.339986	-79.894345
14	L9B	Hamilton	Barnstown , West Chappel , Allison , Ryckmans , Mewburn , Sheldon , Falkirk , Carpenter , Kennedy , Southwest Outskirts	43.202345	-79.898150
21	L9C	Hamilton	Southam , Bonnington , Yeoville , Kernighan , Gourley , Rolston , Buchanan , Mohawk , Westcliffe , Gilbert , Gilkson , Gurnett , Fessenden , Mountview	43.228755	-79.902844
28	L8E	Hamilton	Confederation Park , Nashdale , East Kentley , Riverdale , Lakely , Grayside , North Stoney Creek	43.231113	-79.698031
36	L8G	Hamilton	Greenford , North Gershome , West Stoney Creek	43.216999	-79.741531

2.2 Housing Sales Datasets

Datasets which includes average housing list price and number of active houses for sale by city and neighborhood will be created by web scraping several pages from the website Zoocasa.

Burlington housing dataset - Coordinates are from neighborhood datasets:

	PostalCode	Neighborhood	Longitude	Latitude	Active Listings	Avg. List Price	Avg. List Price ($)
0	L7L	Corporate,Elizabeth Gardens,Longmoor,Pinedale,Shoreacres	-79.767496	43.384449	99	1188548.0	$1,188,548.00
5	L7M	Alton North,Headon Forest,Millcroft,Palmer,Tansley	-79.857794	43.417438	89	1199763.2	$1,199,763.20
10	L7N	Dynes,Roseland	-79.783412	43.351052	41	1434010.5	$1,434,010.50
12	L7P	Brant Hills,Tyandaga,Mountainside	-79.879311	43.372384	59	862516.0	$862,516.00
15	L7R	Plains	-79.804560	43.353290	14	722528.0	$722,528.00
16	L7S	Maple	-79.810205	43.317173	42	784498.0	$784,498.00
17	L7T	Aldershot Central,Aldershot South	-79.855320	43.313290	78	1064370.0	$1,064,370.00

Hamilton housing dataset - Coordinates are from neighborhood datasets (Not all data shown due to dataset size):

	PostalCode	Neighborhood	Longitude	Latitude	Active Listings	Avg. List Price	Avg. List Price ($)
0	L8B	Waterdown	-79.894345	43.339986	26	8.123960e+05	$812,396.00
1	L8E	Nashdale,Riverdale West,Lakely,Stoney Creek	-79.698031	43.231113	18	9.644300e+05	$964,430.00
5	L8G	Stoney Creek,Gershome,Gershome	-79.741531	43.216999	29	1.047238e+06	$1,047,238.33
8	L8H	McQuesten West,Parkview West,Hamilton Beach C,Normanhurst,Homeside,Crown Point West	-79.791770	43.264400	55	6.884852e+05	$688,485.17
14	L8J	Stoney Creek,Albion Falls	-79.730197	43.178694	10	7.829265e+05	$782,926.50
16	L8K	Gershome,Gershome,Eastmount,Bartonville,Glenview West,Rosedale,Red Hill,Corman,Vincent	-79.802858	43.221302	71	7.368563e+05	$736,856.33
25	L8L	Kernighan,Crown Point West,Stipley,Gibson,Landsdale,North End West,Beasley	-79.844632	43.259358	137	6.199774e+05	$619,977.43

2.3 Foursquare Top Venues Datasets

The Foursquare API will be utilized to gather venue information around the neighborhoods in each city.

Burlington neighborhoods top venues dataset:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aldershot Central,Aldershot South	Train Station	Platform	Road	Coffee Shop	Ice Cream Shop	Hardware Store	Gym / Fitness Center	Gym	Gas Station	Furniture / Home Store
1	Corporate,Elizabeth Gardens,Longmoor,Pinedale,Shoreacres	Construction & Landscaping	Athletics & Sports	Gym / Fitness Center	Skating Rink	Dessert Shop	Hotel	Hardware Store	Gym	Gas Station	Furniture / Home Store
2	Dynes,Roseland	Furniture / Home Store	Sandwich Place	Paper / Office Supplies Store	Discount Store	Dessert Shop	Convenience Store	Coffee Shop	Ice Cream Shop	Clothing Store	Thrift / Vintage Store
3	Maple	Italian Restaurant	Gym	Brewery	Clothing Store	Dessert Shop	Hotel	Hardware Store	Gym / Fitness Center	Gas Station	Furniture / Home Store
4	Plains	Restaurant	Hotel	Gym / Fitness Center	Intersection	Breakfast Spot	Gas Station	Furniture / Home Store	Mediterranean Restaurant	Pub	Hardware Store

Hamilton neighborhoods top venues dataset:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Albion Falls,Trenholme,Quinndale,Templemead,Broughton West,Eleanor,Randall,Rushdale,Butler	Sandwich Place	Fast Food Restaurant	Discount Store	Pizza Place	Wings Joint	Construction & Landscaping	Gym	Grocery Store	Gas Station	Furniture / Home Store
1	Ancaster	Department Store	Fast Food Restaurant	Sporting Goods Shop	Shopping Mall	Sandwich Place	Furniture / Home Store	Pharmacy	Vietnamese Restaurant	Gym	Grocery Store
2	Barnstown,Allison,Ryckmans,Mewburn,Sheldon,Falkirk West,Carpenter,Kennedy West	Coffee Shop	Fast Food Restaurant	Restaurant	Wings Joint	Portuguese Restaurant	Grocery Store	Gym	Juice Bar	Kids Store	Pharmacy
3	Central,Strathcona	Beach	Park	Wings Joint	Convenience Store	Gym	Grocery Store	Gas Station	Furniture / Home Store	Food Truck	Flower Shop
4	Crerar,Bruleville,Hill Park,Inch Park,Centremount,Greeningdon,Jerome	Coffee Shop	Motorcycle Shop	Office	Pizza Place	Pharmacy	Wings Joint	Convenience Store	Grocery Store	Gas Station	Furniture / Home Store

3. Methodology

3.1 Analytic Approach

In this project I will be taking two analytic approaches:

1. **Exploratory Data Analysis (EDA)** - To discover details in the data that will contribute to providing useful insights to the reader and young adults. I used the neighborhood datasets as a foundation for building the housing sales datasets. This allowed me to use EDA to discover differences of average house prices and active listings by neighborhoods.

2. **Prescriptive Analytics** - To provide the reader and young adults information regarding similarities and differences between neighborhoods and their surround venues. K-Means Clustering will be used in this situation. The neighborhoods datasets were used as a foundation to build the Foursquare venues datasets. This analysis will provide the reader information on surrounding venues of each neighborhood.

4. Analysis

4.1 Exploratory Data Analysis

Focusing on neighborhoods with the geographical locations based on their postal codes, we use the data from the neighborhoods datasets to visualize the neighborhoods using the folium module for both Hamilton and Burlington.

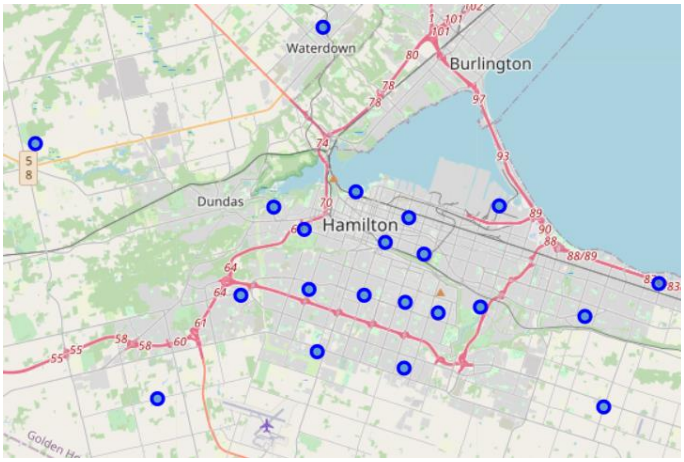

Figure 1: Hamilton Neighborhoods

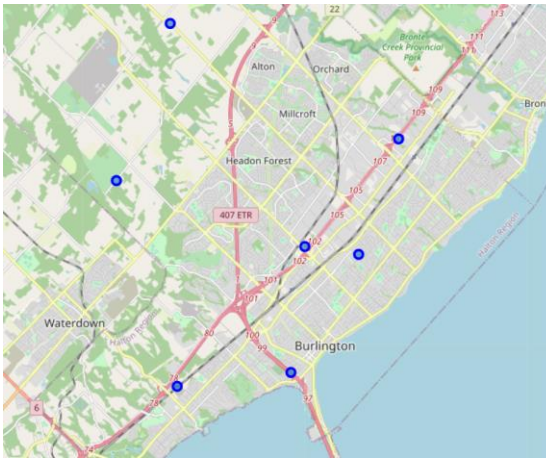

Figure 2: Burlington Neighborhoods

With the neighborhoods mapped out above, we can see that Hamilton has significantly more neighborhoods with 21 in comparison to 7 neighborhoods in Burlington.

Now that the neighborhoods in Hamilton and Burlington are visualized in the figures above, we can now explore the following in our data:

- Housing cost for each neighborhood
- Active number of houses for sale in each neighborhood
- How Hamilton and Burlington compare to one another in regard to active listings and housing costs

Before we explore the data for each individual neighborhood, we will look at the high-level information to understand the difference between Hamilton and Burlington in terms of active number of houses for sale and their respective costs.

In the figures below we have grouped the active number of houses for sale into three groups: low cost, medium cost, and high cost.

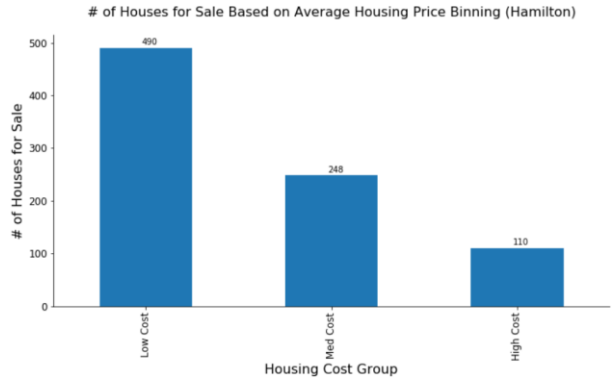

Figure 3: Hamilton Active Housing Listings Grouped by Cost

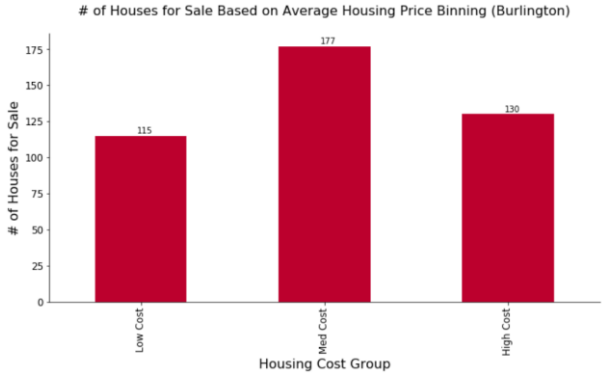

Figure 4: Burlington Active Housing Listings Grouped by Cost

A significance difference between Hamilton (Blue) and Burlington (Red) can be observed in which Hamilton has the highest number of active listings in the “low cost” group whereas Burlington has the highest number of listings in the “medium cost” group.
In total Hamilton has significantly more active listings, but that is expected as it has many more neighborhoods than Burlington.

Next, we want to understand whether there is a difference in the price range for each housing cost group between Hamilton and Burlington as they may not be the same. We found the minimum and maximum value for each cost group for both cities and plotted them together in the graph below.

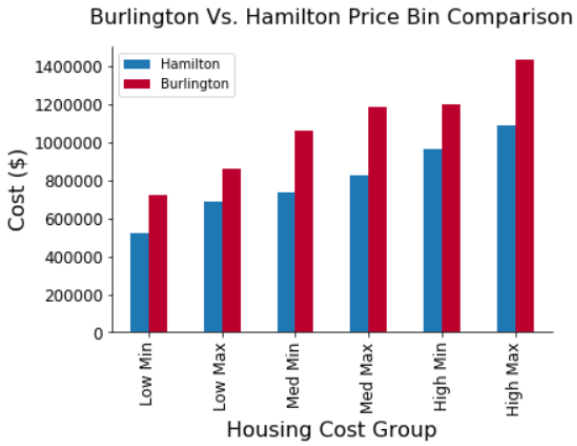

Figure 5: Hamilton vs. Burlington Grouped Housing Cost

As you can see in the figure above, the overall cost of active houses for sale is consistently higher in Burlington across all cost groups. This is a key point to keep in mind when looking into where to purchase a house if cost is a significant factor in the decision.

Now that we understand the differences between the Hamilton and Burlington, we will investigate the average list price and active listings by neighborhood in detail. Below we plotted the average list price and active listings of houses for sale by neighborhood for both Hamilton and Burlington.

The figures below allow us to see how each neighborhood differs by average list price and active listings of houses for sale. For Hamilton we can see that there is one location in the low average list price group which has a significant contribution to the number of active listings in the low-cost group indicated in figure 3. For Burlington the graph reflects figure 4 in which the medium-cost group has the greatest number of active listings.

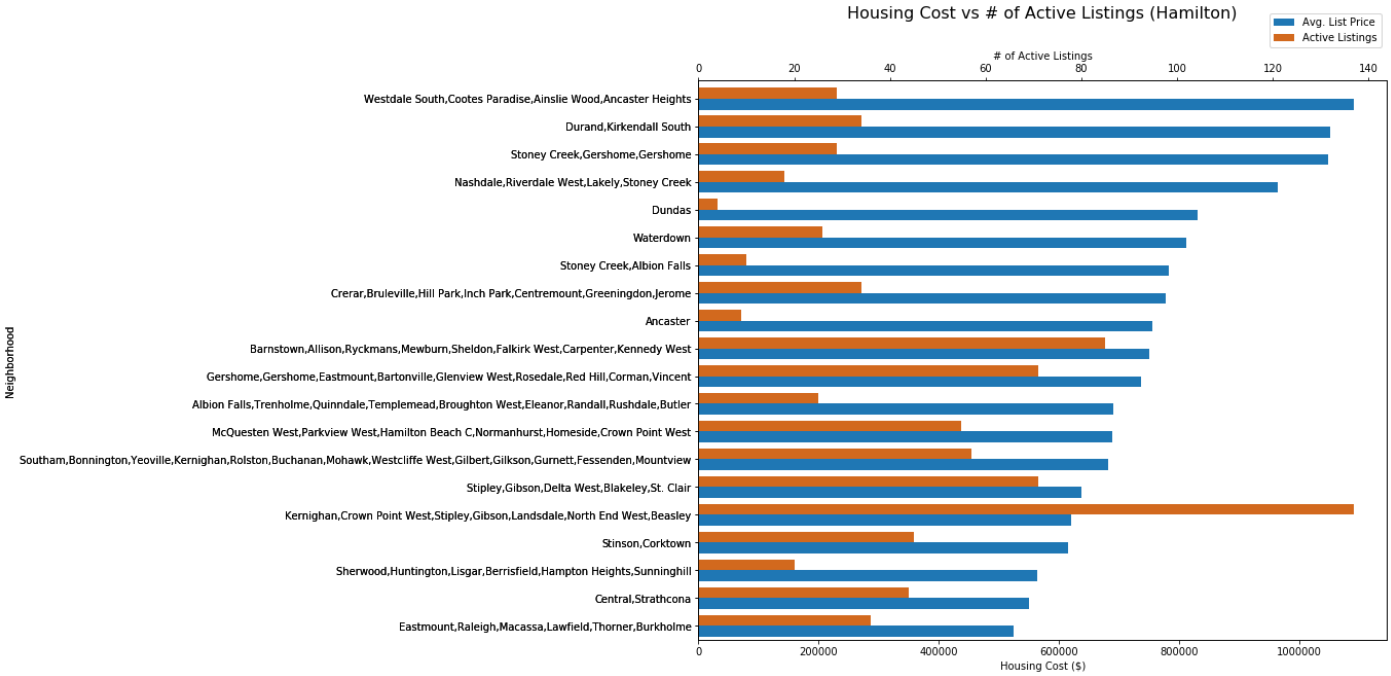

Figure 6: Hamilton Housing Cost and Active Listings by Neighborhoods

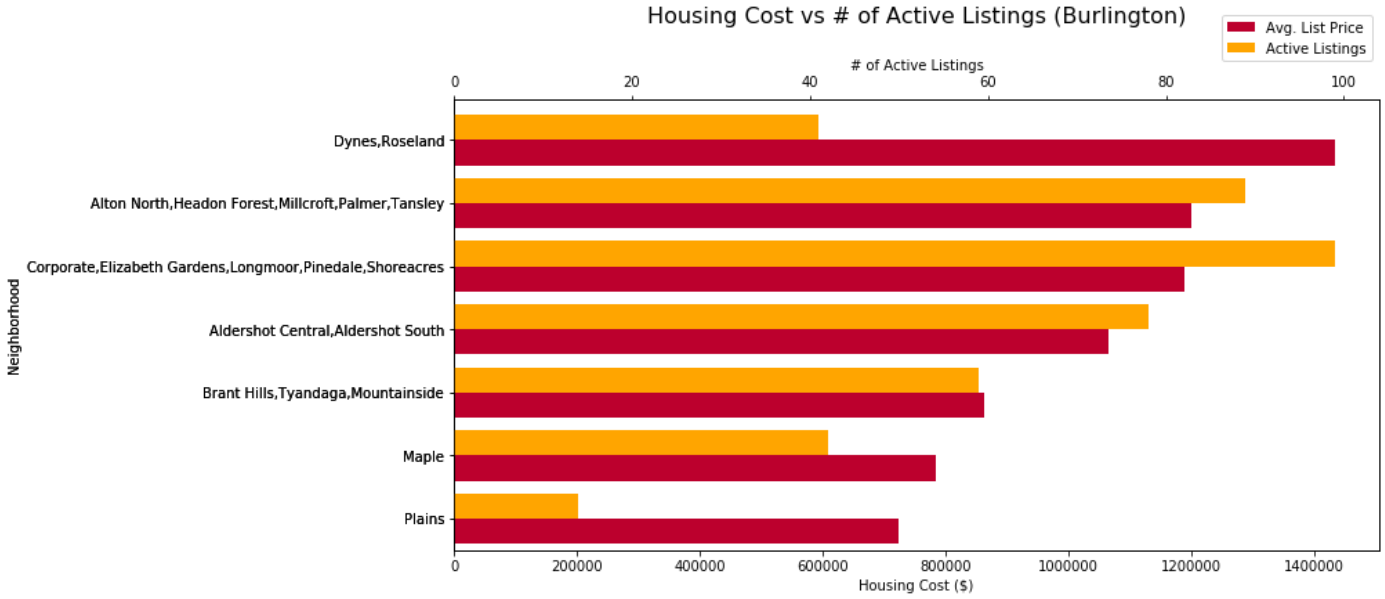

Figure 7:Burlington Housing Cost and Active Listings by Neighborhoods

4.2 Clustering

To gain an understanding for the similarities and differences between neighborhoods and their surrounding venues within a 500m radius we performed K-Means Clustering for both Hamilton and Burlington. In this case we will be focusing on the top 10 common venues surrounding each neighborhood. The clustered neighborhoods are then visualized using folium.

NOTE: The foursquare data collected at the time of this analysis (May 2020) is limited due to the closure of most venues due to the COVID-19 shutdown

We will start with Hamilton in which we fixed K to be 3 for our K-Means clustering.

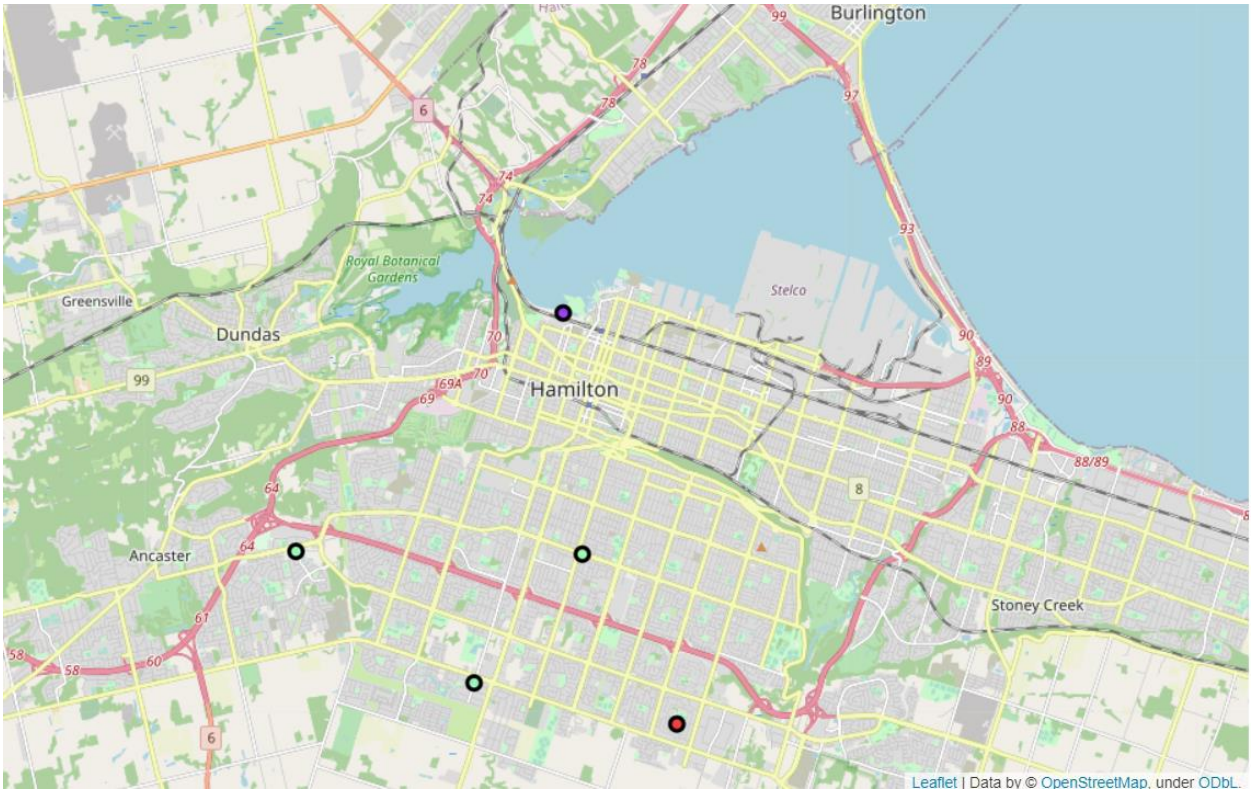

Figure 8: Hamilton Neighborhoods Clustered by Common Venues

As you can see in the figure above the foursquare data is quite limited as we only have data to show 5 neighborhoods in comparison to the existing 21 neighborhoods in Hamilton. With the data available for Hamilton we examine the clusters.

Hamilton Cluster 0 (Red): Mostly food take out venues.

Hamilton Cluster 1 (Purple): Only cluster containing a beach and park in its top 10 venues in Hamilton.

Hamilton Cluster 2(Green): Mostly coffee shops and fast food restaurants.

Next, we clustered neighborhoods in Burlington. We fixed the K to be 4 in our K-Means clustering in this situation.

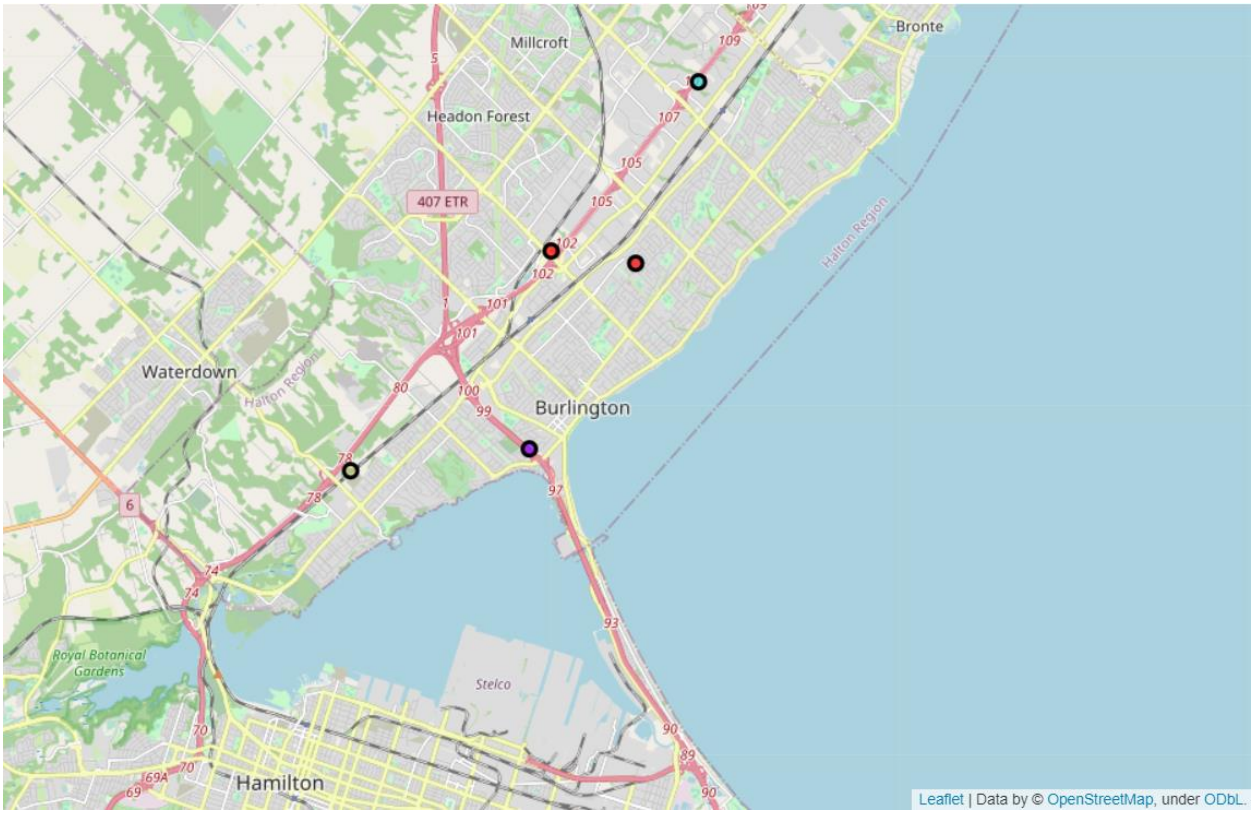

Figure 9: Burlington Neighborhoods Clustered by Common Venues

For Burlington, the foursquare data was not as impacted by the COVID shutdown. We were able to obtain data from 5 out of 7 neighborhoods in which the 2 missing neighborhoods do not have any foursquare data as they are rural. With the available data we examine the Burlington clusters.

Burlington Cluster 0 (Red): Heavy mix of restaurants and stores.

Burlington Cluster 1 (Purple): Mix of fitness, restaurants, stores, and is only area with a brewery.

Burlington Cluster 2 (Teal): Mostly fitness venues.

Burlington Cluster 3 (Beige): Commuter area as the following are within the top 4 venues - train station, platform, and road

5. Result

From the analysis we were able to gain information which would provide us with insight in the differences between Hamilton and Burlington when purchasing a home. Below is a summary of results:

- Hamilton has significantly more neighborhoods in comparison to Burlington (Based on postal codes)
    - Hamilton with 21 neighborhoods and Burlington with 7 neighborhoods
- Overall Burlington is much more expensive to purchase a house when comparing the cost groups between the two cities (Low cost, medium cost, high cost)
- Hamilton has the most active listings in its “low cost” category
- Burlington has the most active listings in its “medium cost” category
- Hamilton neighborhoods are mostly food oriented when looking at nearby top venues
- Burlington neighborhood venues are more diverse than Hamilton in which they also contain fitness venues in the top 10 lists
    - Was able to identify an area in Burlington as a commuter area as the top venues were transit related

6. Discussion

With the analysis completed above we can see that the neighborhood of Maple in Burlington would be an ideal choice for me look to purchase a house based on my criteria which were: affordable neighborhood, fitness amenities, restaurants, and coffee shops within walking distance. The number of active listings in Maple is on the lower end of the spectrum when looking at active listings by neighborhoods in Burlington so the options would be less in comparison to the other Burlington neighborhoods.

If I were to prioritize affordability over my other criteria, I would search for a house to purchase in Hamilton as it has the lowest costs. Once I take into considerations the criteria other than affordability, Hamilton is no longer my preferred city as the top venues surrounding the neighborhoods are mostly food related whereas I am also looking for fitness venues.

For others that are looking to purchase a house, the exploratory data analysis would be helpful when it comes to comparing Hamilton and Burlington on a high level in addition to comparing neighborhoods within the cities when looking for more details. When it comes to understanding top common surrounding venues, the clusters would help to identify alternative neighborhoods to investigate purchasing a house.

An additional item that would be beneficial in terms of purchasing a house would be the use of crime data. With crime data involved the analysis would be much improved as many people also take into consideration crime rates when selecting a neighborhood move into, especially if they are looking to start a family in the area. Unfortunately, any crime data referring to specific neighborhoods are no longer publicly available.

This analysis should be used to provide an initial understanding of Hamilton and Burlington neighborhoods when purchasing a house as it only reflects the current condition of the housing market and top common venues surrounding the analyzed neighborhoods. The current condition is a very abnormal time in our life as the economy has been temporarily shutdown due to a global pandemic occurring so when life is back to normal, these items should be re-visited.

7. Conclusion

In this study we used data science to explore and analyse various aspects of Hamilton and Burlington neighborhoods when it comes to finding an ideal location to purchase a house. We collected existing datasets by scraping several websites and had combined them with data collected from the Foursquare API. We completed Exploratory Data Analysis and Clustering on these datasets to provide valuable information which would assist in finding a solution to our problem. Based on these results we were able to find a strong contender in terms of an ideal location to purchase a house under the criteria I had set for myself

We recognize this study is based on limited data which has also been impacted during an abnormal global event. This study still provides a good starting point in terms of finding an ideal location to purchase a house when looking into Hamilton and Burlington. For those that are not searching for a neighborhood to purchase a house based on my criteria I had set, they can still use the information found in this study to assist in making a decision of an area based on their own unique criteria. This study can be repeated for other cities around the globe.