Opening an Ice Cream Business in Seattle – Analysis

Shaheer Khan

August 27, 2020

Table of Contents

1. Introduction

1.1. Background

Seattle is not only the largest city in the state of Washington, but it is also the largest city in the Pacific Northwestern region of North America. Many of us know Seattle for its rainy weather and tech giants, being as it's the birthplace of companies such as Microsoft and Amazon, to name a few. Thanks to its tech scene, Seattle is now one of the fastest growing major cities in the United States. Any savvy entrepreneur would consider business possibilities in this thriving city, and for me, that means opening my dream ice cream business. Catering to Seattle's outdoorsy population and for my love of ice cream, I was tempted to explore the possibility of opening an ice cream business in this city. [1]

1.2. Business Problem

The audience for this analysis is any business owner trying to figure out where to open an ice cream shop or ice cream truck, given they want to be situated in a populous area and in an area where the general population is young and outdoorsy, in the city of Seattle, Washington, USA.

2. Data

Based on the business problem discussed earlier, factors that will be affecting the recommendations are:

- A populous area
- Young and Outdoorsy Inhabitants

Having said that, it is essential to keep these 2 factors in mind when recommending the best possible neighbourhood for an ice cream business to be opened.

2.1. Neighbourhood Data

Neighbourhood Data for Seattle, Washington was imported from US CENSUS. This data is obtained from seattle.gov (US census) website [2]. This data set includes data on population, housing, and ethnicity information for each neighbourhood. Total number of neighbourhoods provided were **53**. The data was presented in a CSV format, since the dataset was not too big, I chose to remove unwanted headers, footers and columns within the CSV file using Excel. Later high-level data cleaning was carried out in Python.

2.2. Location Data

The location data is provided by Foursquare API. The Foursquare API was used to gather the data regarding most common venues for each respective neighbourhood using the explore function of the Foursquare API. This is discussed in more detail later in the Methodology Section.

2.3. Coordinates Data

The coordinates data for each respective neighbourhood was obtained using geopy and geocoder packages. The code for which can be found on my GitHub [3].

3. Methodology

To begin, the Excel file provided in the source found in **Section 2.1** was downloaded. Then, the excel file was converted into a CSV file. The CSV file was then cleaned in Microsoft Excel as discussed earlier and then it was imported to Python notebook using Pandas. The imported data was stored in a DataFrame called 'census_df'

The CSV file had 14 columns in total named:
1) Neighbourhood
2) Total Population

3) White

4) Black or African American

5) American Indian & Alaska Native

6) Asian

7) Native Hawaiian & Pacific Islander

8) Persons of Color*

9) Population 18 and Over

10) Population Under 18

11) Persons of Color Under 18

12) Housing Units

13) Occupied Housing Units

14) Vacant Housing Units

Out of which columns labelled the following were dropped:

- "Persons of Color*"
- "Persons of Color Under 18"

Then, the first entry of the DataFrame was dropped, later the data type of all the columns except for 'Neighbourhood' column was converted to int64 type. The DataFrame was evaluated for any missing values.

A preview of the DataFrame is shown below:

	Neighbourhood	Total Population	White	Black or African American	American Indian & Alaska Native	Asian	Native Hawaiian & Pacific Islander	Persons of Color*	Population 18 and Over	Population Under 18	Persons of Color Under 18	Housing Units	Occupied Housing Units	Vacant Housing Units
1	Arbor Heights	6030	5082	98	43	326	33	1140	4869	1161	342	2752	2612	140
2	Fauntleroy/Seaview	13723	11777	349	89	625	29	2303	11352	2371	649	6920	6513	407
3	West Seattle Junction/Genesee Hill	17713	15299	342	122	855	46	2910	14511	3202	768	8889	8372	517
4	Alki/Admiral	10542	9240	188	73	554	6	1585	9034	1508	374	6068	5528	540
5	North Delridge	4787	3529	318	52	402	30	1493	4104	683	334	2618	2403	215

Figure 1: DataFrame "census_df" head view

3.1. Adding Latitude and Longitude Coordinates to the Neighbourhoods

Using the Geocoder and Geopy packages, the coordinates were added to the respective neighbourhoods. The coordinates data was added to the existing DataFrame called "census_df."

3.2. Evaluating Missing Values after Coordinates Import

The DataFrame was evaluated for any missing coordinate values, and it was found that one neighbourhood was missing coordinates data. That neighbourhood is called "Pioneer Square/International District", therefore using the source below the coordinates will be changed

Source:

https://geohack.toolforge.org/geohack.php?pagename=Pioneer_Square,_Seattle&params=47_36_06_N_122_19_55_W_dim:500_region:US-WA

3.3. Evaluating the Accuracy of the Coordinates Data Geographically

As shown below in Figure 2 there were, misplaced data points. One is called "North Capitol Hill" and the other is "Belltown" as shown in the Figure 2 and Figure 3 below.
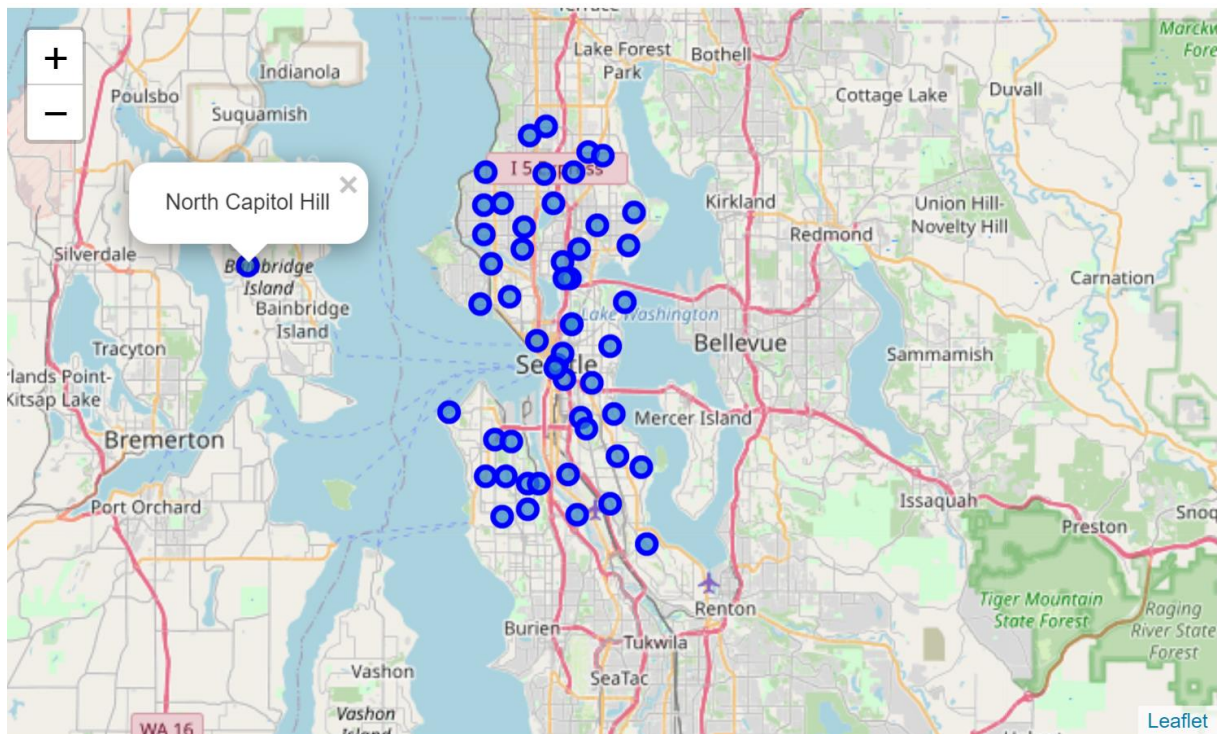
Figure 2: Inaccurate Data Point "North Capitol Hill"
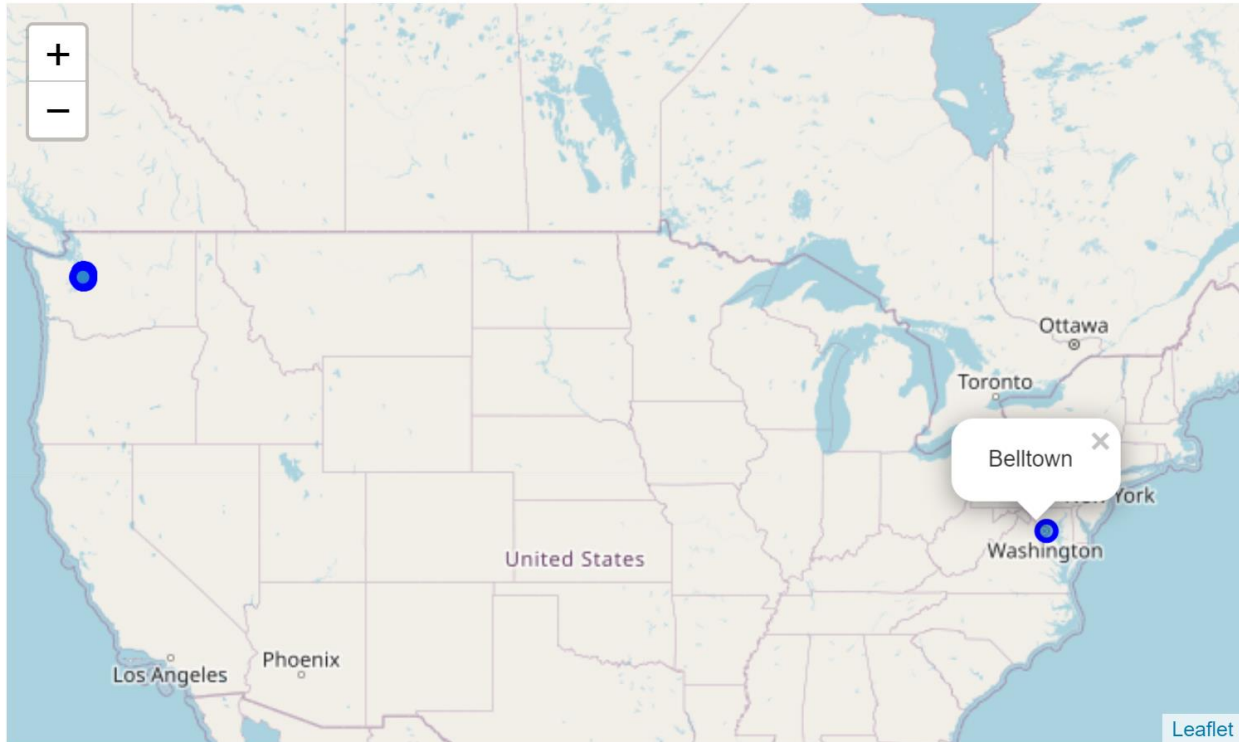
Figure 3: Inaccurate Data Point "Belltown"

The correct coordinates were collected from the internet as shown in the Python Notebook [3]. A preview of the final "census_df" DataFrame alongside Seattle Neighbourhood map is shown below in Figure 4 and Figure 5:

	Neighbourhood	Total Population	White	Black or African American	American Indian & Alaska Native	Asian	Native Hawaiian & Pacific Islander	Population 18 and Over	Population Under 18	Housing Units	Occupied Housing Units	Vacant Housing Units	Latitude	Longitude
1	Arbor Heights	6030	5082	98	43	326	33	4869	1161	2752	2612	140	47.546653	-122.383186
2	Fauntleroy/Seaview	13723	11777	349	89	625	29	11352	2371	6920	6513	407	47.565040	-122.376154
3	West Seattle Junction/Genesee Hill	17713	15299	342	122	855	46	14511	3202	8889	8372	517	47.578940	-122.410670
4	Alki/Admiral	10542	9240	188	73	554	6	9034	1508	6068	5528	540	47.564730	-122.363860
5	North Delridge	4787	3529	318	52	402	30	4104	683	2618	2403	215	47.547040	-122.368940

Figure 4: Final cleaned version of "census_df" Dataframe
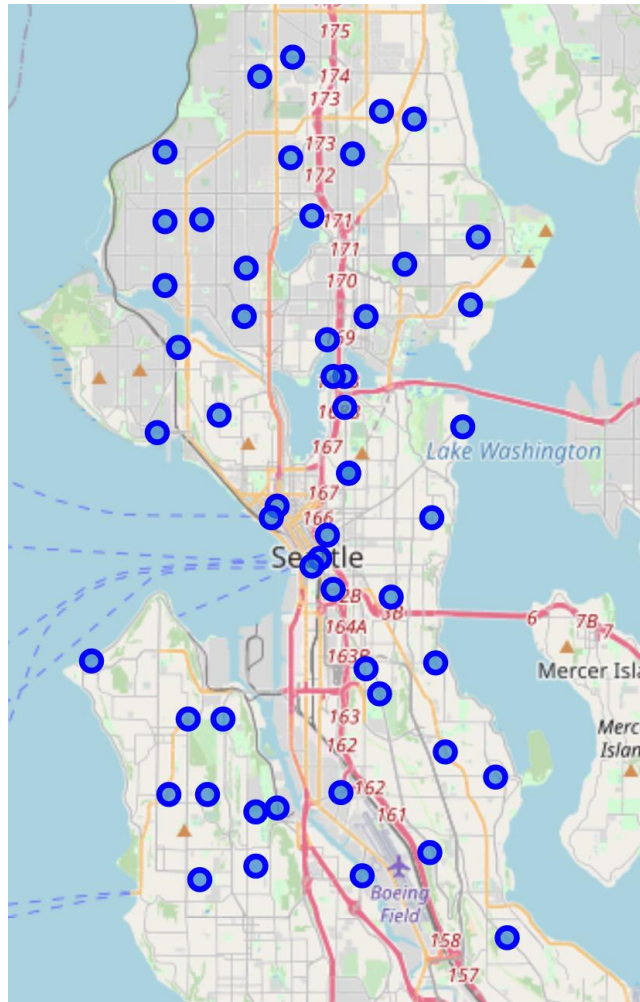
Figure 5: Neighbourhoods in Seattle

3.4. Creating a DataFrame for each Feature

3.4.1. Population by Age DataFrame

A preview of Population by Age DataFrame can be seen below in Figure 6:

	Neighbourhood	Population 18 and Over	Population Under 18	Latitude	Longitude
1	Arbor Heights	4869	1161	47.546653	-122.383186
2	Fauntleroy/Seaview	11352	2371	47.565040	-122.376154
3	West Seattle Junction/Genesee Hill	14511	3202	47.578940	-122.410670
4	Alki/Admiral	9034	1508	47.564730	-122.363860
5	North Delridge	4104	683	47.547040	-122.368940

Figure 6: Population by Age DataFrame

3.4.2. Race/Ethnicity DataFrame

A preview of Race/Ethnicity DataFrame can be seen below in Figure 7:

	Neighbourhood	White	Black or African American	American Indian & Alaska Native	Asian	Native Hawaiian & Pacific Islander	Latitude	Longitude
1	Arbor Heights	5082	98	43	326	33	47.546653	-122.383186
2	Fauntleroy/Seaview	11777	349	89	625	29	47.565040	-122.376154
3	West Seattle Junction/Genesee Hill	15299	342	122	855	46	47.578940	-122.410670
4	Alki/Admiral	9240	188	73	554	6	47.564730	-122.363860
5	North Delridge	3529	318	52	402	30	47.547040	-122.368940

Figure 7: Race/Ethnicity DataFrame

3.4.3. Housing DataFrame

A preview of the Housing DataFrame can be seen below in Figure 8:

	Neighbourhood	Housing Units	Occupied Housing Units	Vacant Housing Units	Latitude	Longitude
1	Arbor Heights	2752	2612	140	47.546653	-122.383186
2	Fauntleroy/Seaview	6920	6513	407	47.565040	-122.376154
3	West Seattle Junction/Genesee Hill	8889	8372	517	47.578940	-122.410670
4	Alki/Admiral	6068	5528	540	47.564730	-122.363860
5	North Delridge	2618	2403	215	47.547040	-122.368940

Figure 8: Housing DataFrame

3.5. Exploratory Data Analysis:

3.5.1. Population per Neighbourhood Analysis

3.5.1.1. Ten Neighbourhoods with most Population Under 18 in Seattle

The Population DataFrame created earlier is used to derive 10 neighbourhoods with the most population with the age under 18 as shown graphically in Figure 10. Later this data is plotted on a map to find out which areas in the Seattle region possess these neighbourhoods. A preview of which is provided in Figure 9. The data seems scattered, being almost

equally divided between southern and northern Seattle, so no conclusive comment could be made.

	Neighbourhood	Population Under 18	Latitude	Longitude
30	Ravenna/Bryant	4227	47.68033	-122.27290
13	South Beacon Hill/NewHolly	3892	47.57686	-122.31271
38	Greenwood/Phinney Ridge	3870	47.68508	-122.33232
17	Columbia City	3692	47.55115	-122.26680
50	Queen Anne	3623	47.61576	-122.34464
16	Rainier Beach	3389	47.55687	-122.28452
3	West Seattle Junction/Genesee Hill	3202	47.57894	-122.41067
31	Wedgwood/View Ridge	3179	47.66398	-122.27582
48	Magnolia	2655	47.65399	-122.37969
32	Laurelhurst/Sand Point	2645	47.70013	-122.31765

Figure 9: Ten Neighbourhoods with most Population Under 18 in Seattle
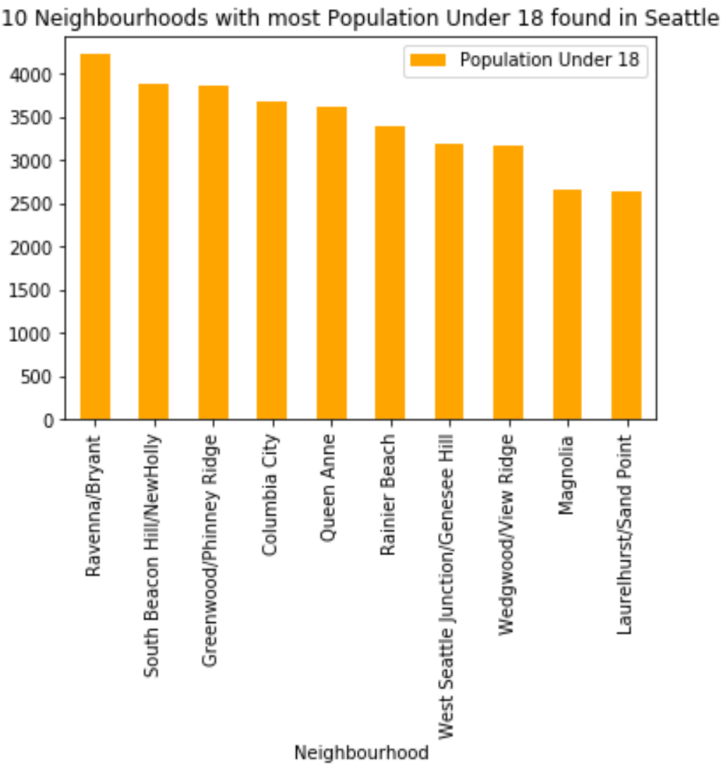
Figure 10: Ten Neighbourhoods with most Population Under 18 in Seattle in a Bar Graph

3.5.1.2. Ten Neighbourhoods with most Population 18 and Over in Seattle

The Population DataFrame created earlier is used to derive 10 neighbourhoods with the most population with the age 18 and over. Later this data is plotted on a map to find out which areas in the Seattle region possess these neighbourhoods. A preview of which is provided below in Figure 11. Majority of the population which is 18 and over lives in the northern part of Seattle.

	Neighbourhood	Population 18 and Over	Latitude	Longitude
50	Queen Anne	31835	47.61576	-122.34464
38	Greenwood/Phinney Ridge	20078	47.68508	-122.33232
30	Ravenna/Bryant	19960	47.68033	-122.27290
29	University District	18445	47.67393	-122.29914
24	Capitol Hill	18397	47.62396	-122.31882
22	Central Area/Squire Park	14623	47.60878	-122.32643
3	West Seattle Junction/Genesee Hill	14511	47.57894	-122.41067
23	First Hill	14357	47.62396	-122.31882
46	Wallingford	14193	47.64708	-122.32477
45	Fremont	14016	47.65555	-122.32650
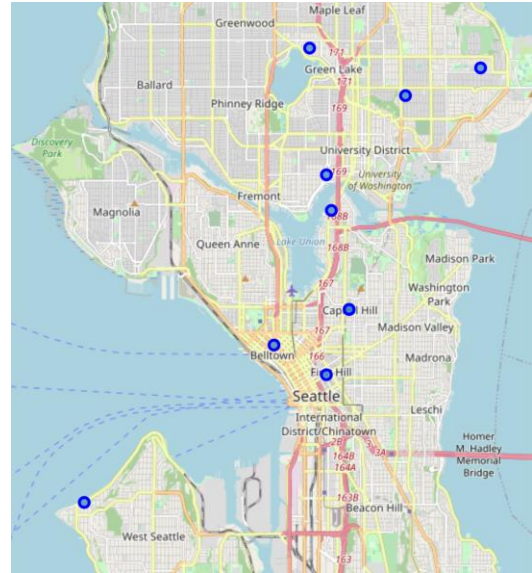
Figure 11: Ten Neighbourhoods with most Population 18 and Over in Seattle

3.5.2. Race/Ethnicity per Neighbourhood Analysis

3.5.2.1. Ten Most White Neighbourhoods in Seattle

The Race/Ethnicity DataFrame created earlier is used to derive 10 most White neighbourhoods found in Seattle. Later this data is plotted on a map to find out which areas in the Seattle region possess these neighbourhoods. A preview of which is provided in Figure 12. Majority of the white population lives in the northern part of Seattle.

	Neighbourhood	White	Latitude	Longitude
50	Queen Anne	30014	47.61576	-122.34464
38	Greenwood/Phinney Ridge	20051	47.68508	-122.33232
30	Ravenna/Bryant	19394	47.68033	-122.27290
3	West Seattle Junction/Genesee Hill	15299	47.57894	-122.41067
24	Capitol Hill	14636	47.62396	-122.31882
45	Fremont	13293	47.65555	-122.32650
29	University District	12527	47.67393	-122.29914
43	Sunset Hill/Loyal Heights	12466	47.66867	-122.38453
46	Wallingford	12239	47.64708	-122.32477
31	Wedgwood/View Ridge	12112	47.66398	-122.27582
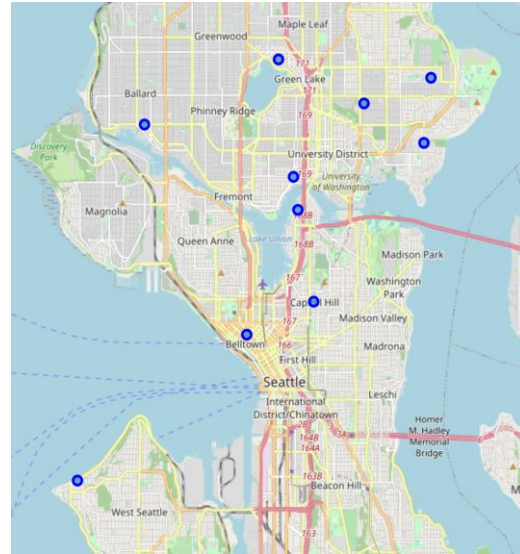
Figure 12: Ten Most White Neighbourhoods in Seattle

3.5.2.2.   Ten Most Black Neighbourhoods in Seattle

The Race/Ethnicity DataFrame created earlier is used to derive 10 most Black neighbourhoods found in Seattle. Later this data is plotted on a map to find out which areas in Seattle possess these neighbourhoods. A preview of which is shown in Figure 13. The only outlier in the map is shown in the north most part of the map which encompasses the neighbourhood called "Cedar Park." So, it is safe to say that majority of the black population lives in the southern part of Seattle.

	Neighbourhood	Black or African American	Latitude	Longitude
16	Rainier Beach	4630	47.55687	-122.28452
17	Columbia City	4521	47.55115	-122.26680
13	South Beacon Hill/NewHolly	4073	47.57686	-122.31271
22	Central Area/Squire Park	3668	47.60878	-122.32643
6	High Point	2319	47.54286	-122.35186
23	First Hill	2245	47.62396	-122.31882
20	Madrona/Leschi	2069	47.59427	-122.30401
15	North Beacon Hill/Jefferson Park	1735	47.51235	-122.26277
18	Seward Park	1510	47.57827	-122.28782
35	Cedar Park/Meadowbrook	1151	47.71868	-122.35030
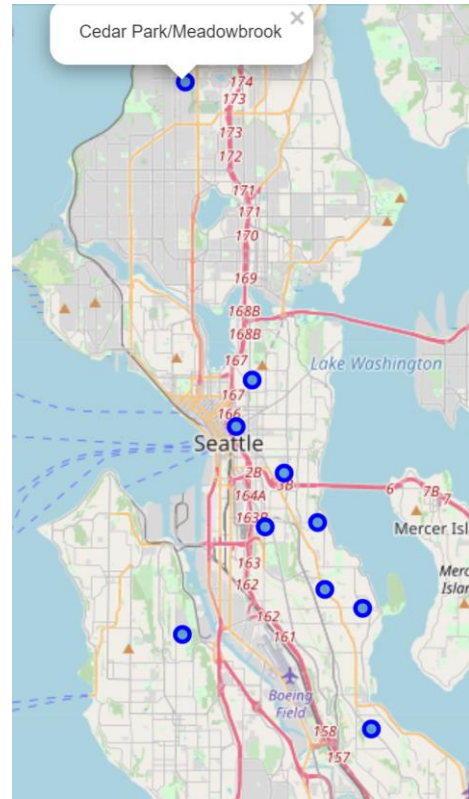
Figure 13: Ten Most Black Neighbourhoods in Seattle

3.5.2.3. Ten Most American Indian & Alaska Native Neighbourhoods

The Race/Ethnicity DataFrame created earlier is used to derive 10 most American Indian & Alaska Native neighbourhoods found in Seattle. Later this data is plotted on a map to find out which areas in Seattle possess these neighbourhoods. A preview of which is shown in Figure 14. Most of the American Indian and Alaska Native population lives in the northern part of Seattle.

	Neighbourhood	American Indian & Alaska Native	Latitude	Longitude
23	First Hill	252	47.62396	-122.318820
50	Queen Anne	209	47.61576	-122.344640
8	Roxhill/Westwood	205	47.52987	-122.351690
38	Greenwood/Phinney Ridge	191	47.68508	-122.332320
24	Capitol Hill	169	47.62396	-122.318820
47	Cascade/Eastlake	161	47.63348	-122.387026
33	Northgate/Maple Leaf	147	47.71030	-122.307200
34	Olympic Hills/Victory Heights	141	47.70841	-122.295860
35	Cedar Park/Meadowbrook	135	47.71868	-122.350300
15	North Beacon Hill/Jefferson Park	130	47.51235	-122.262770
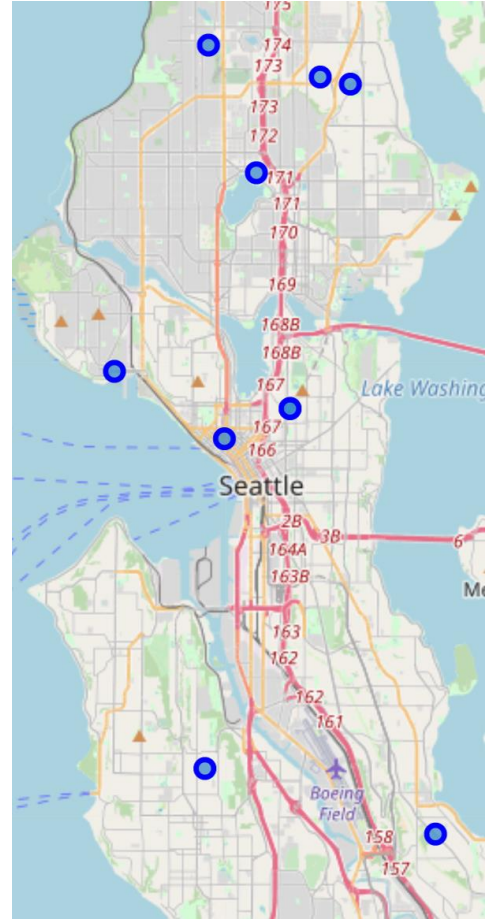
Figure 14: Ten Most American Indian & Alaska Native Neighbourhoods

3.5.2.4. Ten Most Asian Neighbourhoods in Seattle

The Race/Ethnicity DataFrame created earlier is used to derive 10 most Asian neighbourhoods found in Seattle. Later this data is plotted on a map to find out which areas in the Seattle possess these neighbourhoods. A preview of which is shown in Figure 15. The data is very scattered although looking at the DataFrame, there is a clear distinction in the top 6 and the remaining neighbourhoods, hence the top 6 neighbourhoods in this DataFrame were plotted for which the results can be seen in Figure 16. Most of the most Asian population lives in the southern part of Seattle.

	Neighbourhood	Asian	Latitude	Longitude
13	South Beacon Hill/NewHolly	7296	47.57686	-122.31271
14	Beacon Hill	5087	47.57074	-122.30785
15	North Beacon Hill/Jefferson Park	5066	47.51235	-122.26277
17	Columbia City	4948	47.55115	-122.26680
29	University District	4446	47.67393	-122.29914
16	Rainier Beach	4325	47.55687	-122.28452
30	Ravenna/Bryant	2586	47.68033	-122.27290
50	Queen Anne	2409	47.61576	-122.34464
34	Olympic Hills/Victory Heights	2346	47.70841	-122.29586
46	Wallingford	2286	47.64708	-122.32477
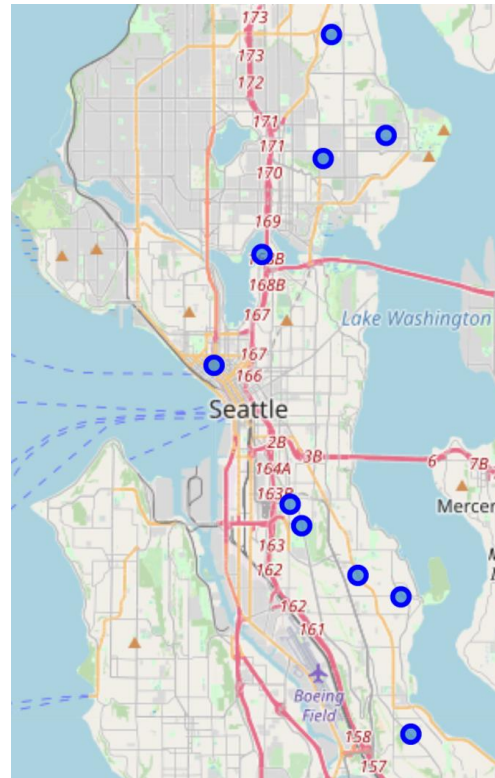
Figure 15: Ten most Asian neighbourhoods found in Seattle

	Neighbourhood	Asian	Latitude	Longitude
13	South Beacon Hill/NewHolly	7296	47.57686	-122.31271
14	Beacon Hill	5087	47.57074	-122.30785
15	North Beacon Hill/Jefferson Park	5066	47.51235	-122.26277
17	Columbia City	4948	47.55115	-122.26680
29	University District	4446	47.67393	-122.29914
16	Rainier Beach	4325	47.55687	-122.28452
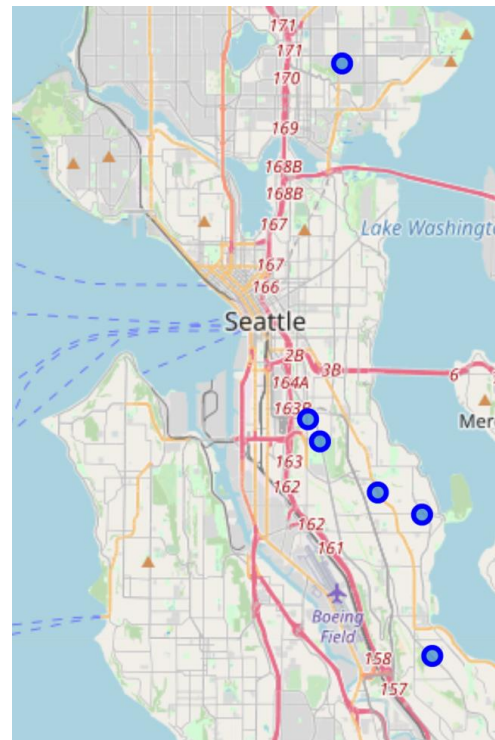
Figure 16: Six most Asian neighbourhoods found in Seattle

3.5.2.5. Ten Most Native Hawaiian & Pacific Islander Neighbourhood

The Race/Ethnicity DataFrame created earlier is used to derive 10 most Native Hawaiian & Pacific Islander neighbourhoods found in Seattle. Later this data is plotted on a map to find out which areas in Seattle possess these neighbourhoods. A preview of which is provided in Figure 17 below. Most of the Native Hawaiian & Pacific Islander population lives in the southern part of Seattle.

	Neighbourhood	Native Hawaiian & Pacific Islander	Latitude	Longitude
16	Rainier Beach	172	47.55687	-122.28452
17	Columbia City	152	47.55115	-122.26680
6	High Point	119	47.54286	-122.35186
13	South Beacon Hill/NewHolly	102	47.57686	-122.31271
8	Roxhill/Westwood	98	47.52987	-122.35169
9	Highland Park	93	47.52722	-122.31445
33	Northgate/Maple Leaf	83	47.71030	-122.30720
23	First Hill	83	47.62396	-122.31882
7	Riverview	74	47.52648	-122.37178
14	Beacon Hill	69	47.57074	-122.30785
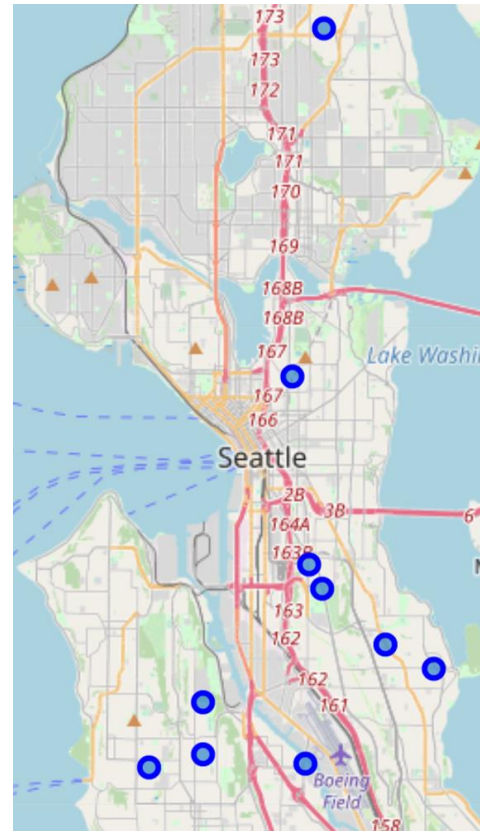
Figure 17: Ten Most Native Hawaiian & Pacific Islander Neighbourhood

3.5.3. Housing per Neighbourhood Analysis

3.5.3.1. Ten Neighbourhoods with most Housing Units in Seattle

The Housing DataFrame created earlier is used to derive 10 neighbourhoods with most Housing Units found in Seattle. Later this data is plotted on a map to find out which areas in the Seattle possess these

neighbourhoods. A preview of which is provided in Figure 18 below. The majority of such neighbourhoods are found in the northern part of Seattle.

	Neighbourhood	Housing Units	Latitude	Longitude
50	Queen Anne	21160	47.61576	-122.344640
24	Capitol Hill	14333	47.62396	-122.318820
38	Greenwood/Phinney Ridge	11664	47.68508	-122.332320
47	Cascade/Eastlake	10834	47.63348	-122.387026
30	Ravenna/Bryant	10777	47.68033	-122.272900
23	First Hill	9159	47.62396	-122.318820
3	West Seattle Junction/Genesee Hill	8889	47.57894	-122.410670
22	Central Area/Squire Park	8879	47.60878	-122.326430
45	Fremont	8819	47.65555	-122.326500
46	Wallingford	7930	47.64708	-122.324770
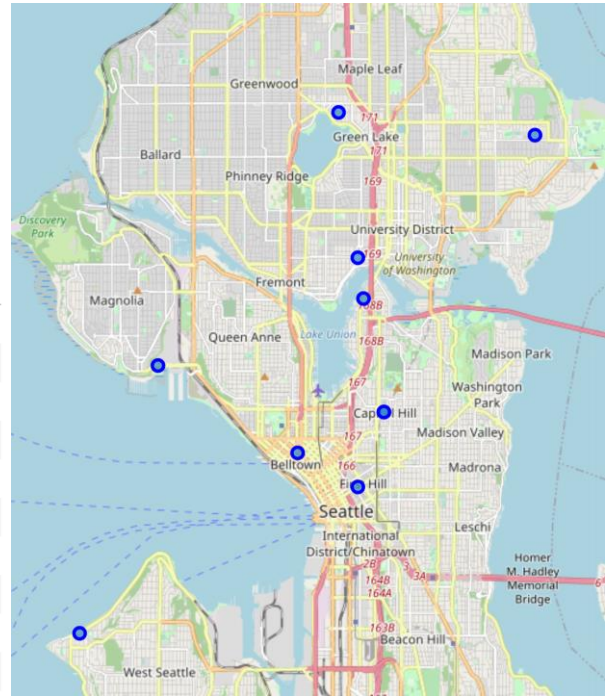
Figure 18: Ten Neighbourhoods with most Housing Units in Seattle

3.5.3.2.   Ten Neighbourhoods with most Occupied Housing Units in Seattle

The Housing DataFrame created earlier is used to derive 10 neighbourhoods with most Occupied Housing Units found in Seattle. Later this data is plotted on a map to find out which areas in Seattle possess these neighbourhoods. A preview of which is shown in Figure 19. The majority of such neighbourhoods are found in the northern part of Seattle.

	Neighbourhood	Occupied Housing Units	Latitude	Longitude
50	Queen Anne	19359	47.61576	-122.344640
24	Capitol Hill	13177	47.62396	-122.318820
38	Greenwood/Phinney Ridge	11085	47.68508	-122.332320
30	Ravenna/Bryant	10177	47.68033	-122.272900
47	Cascade/Eastlake	8853	47.63348	-122.387026
3	West Seattle Junction/Genesee Hill	8372	47.57894	-122.410670
45	Fremont	8292	47.65555	-122.326500
22	Central Area/Squire Park	8216	47.60878	-122.326430
23	First Hill	7628	47.62396	-122.318820
46	Wallingford	7425	47.64708	-122.324770
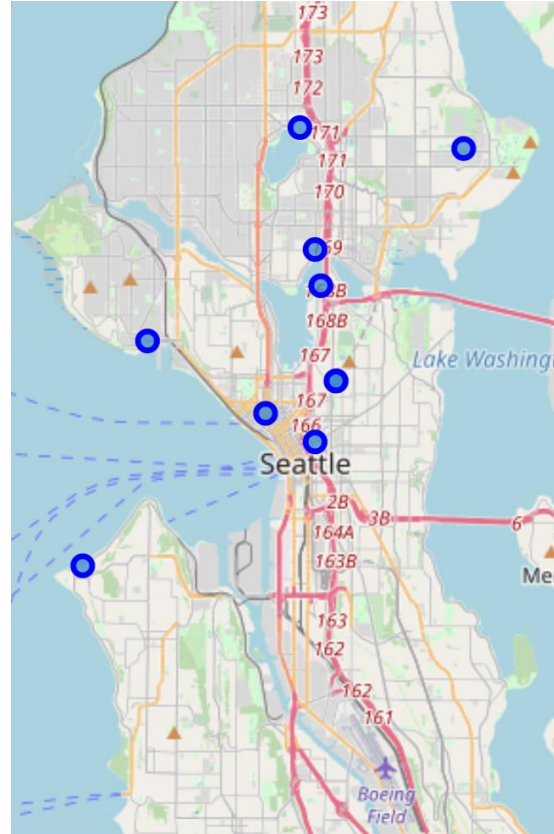
Figure 19: Ten Neighbourhoods with most Occupied Housing Units in Seattle

3.5.3.3. Ten Neighbourhoods with most Vacant Housing Units in Seattle

The Housing DataFrame created earlier is used to derive 10 neighbourhoods with most Vacant Housing Units found in Seattle. Later this data is plotted on a map to find out which areas in Seattle possess these neighbourhoods. A preview of which is shown in Figure 20. The majority of such neighbourhoods are found in the northern part of Seattle.

	Neighbourhood	Vacant Housing Units	Latitude	Longitude
47	Cascade/Eastlake	1981	47.63348	-122.387026
50	Queen Anne	1801	47.61576	-122.344640
23	First Hill	1531	47.62396	-122.318820
24	Capitol Hill	1156	47.62396	-122.318820
51	Belltown	981	47.61322	-122.346500
36	Broadview/Bitter Lake	774	47.69914	-122.339680
33	Northgate/Maple Leaf	733	47.71030	-122.307200
44	Ballard	677	47.66137	-122.356080
22	Central Area/Squire Park	663	47.60878	-122.326430
30	Ravenna/Bryant	600	47.68033	-122.272900
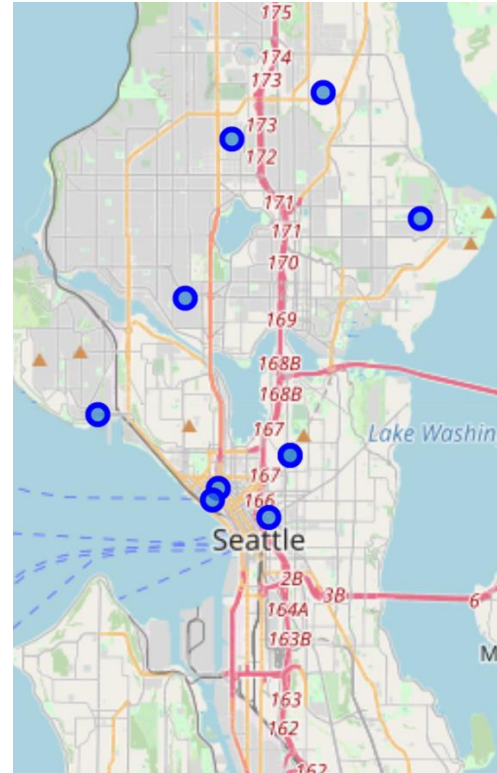
Figure 20: Ten Neighbourhoods with most Vacant Housing Units in Seattle

3.5.4. Populous Neighbourhoods Analysis

3.5.4.1. Ten Most Populous Neighbourhoods

The "census_df" DataFrame created earlier is used to derive 10 most populous neighbourhoods found in Seattle. Later this data is plotted on a map and a bar graph see Figure 22 to find out which areas in Seattle possess these neighbourhoods. A preview of which is shown in Figure 21. The majority of such neighbourhoods are found in the northern part of Seattle.

	Neighbourhood	Total Population	Latitude	Longitude
50	Queen Anne	35458	47.61576	-122.34464
30	Ravenna/Bryant	24187	47.68033	-122.27290
38	Greenwood/Phinney Ridge	23948	47.68508	-122.33232
24	Capitol Hill	19078	47.62396	-122.31882
29	University District	19051	47.67393	-122.29914
3	West Seattle Junction/Genesee Hill	17713	47.57894	-122.41067
22	Central Area/Squire Park	16969	47.60878	-122.32643
17	Columbia City	16883	47.55115	-122.26680
46	Wallingford	16014	47.64708	-122.32477
45	Fremont	15626	47.65555	-122.32650

Figure 21: Ten Most Populous Neighbourhoods
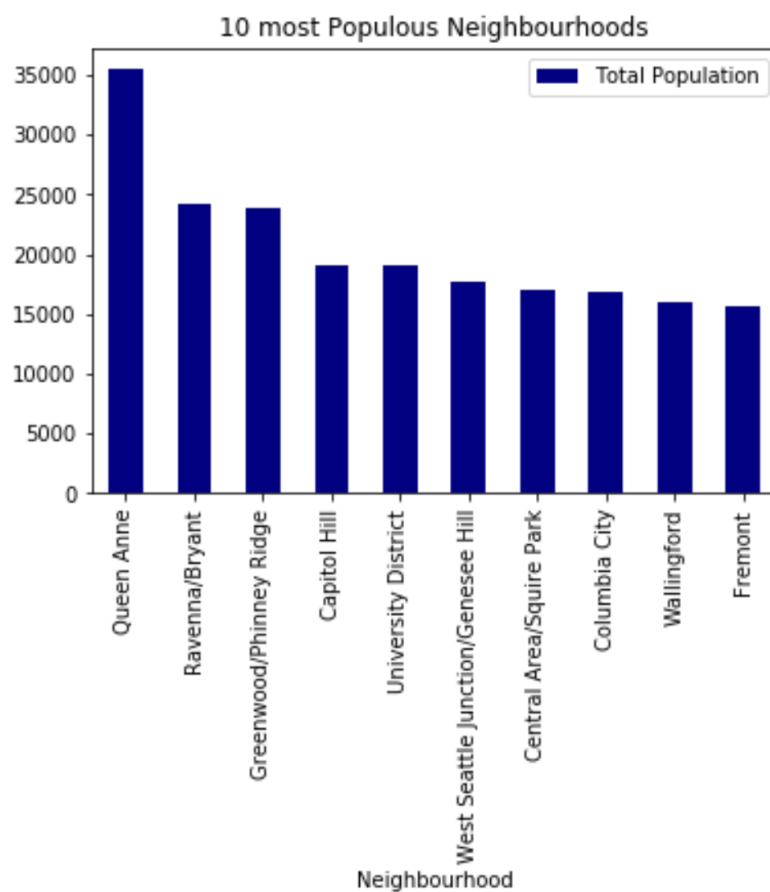
Figure 22: Ten Most Populous Neighbourhoods graphically

3.5.4.2. Ten Least Populous Neighbourhoods

The "census_df" DataFrame created earlier is used to derive 10 least populous neighbourhoods found in Seattle. Later this data is plotted on a map see Figure 23 and a graph see Figure 24 to find out which areas in Seattle possess these neighbourhoods. The majority of such neighbourhoods are found to be scattered, with majority being found in the southernmost part of Seattle, with second most found in downtown area.

	Neighbourhood	Total Population	Latitude	Longitude
11	Georgetown	1287	47.543625	-122.344164
12	Duwamish/SODO	2354	47.533040	-122.289970
21	Judkins Park	2925	47.603570	-122.329450
10	South Park	3873	47.547510	-122.321490
52	Downtown Commercial Core	4070	47.595820	-122.324740
25	North Capitol Hill	4157	47.639320	-122.320360
7	Riverview	4596	47.526480	-122.371780
5	North Delridge	4787	47.547040	-122.368940
27	Madison Park	4962	47.646900	-122.320280
53	Pioneer Square/International District	5333	47.601667	-122.331944

Figure 23: Ten Least Populous Neighbourhoods on a Map
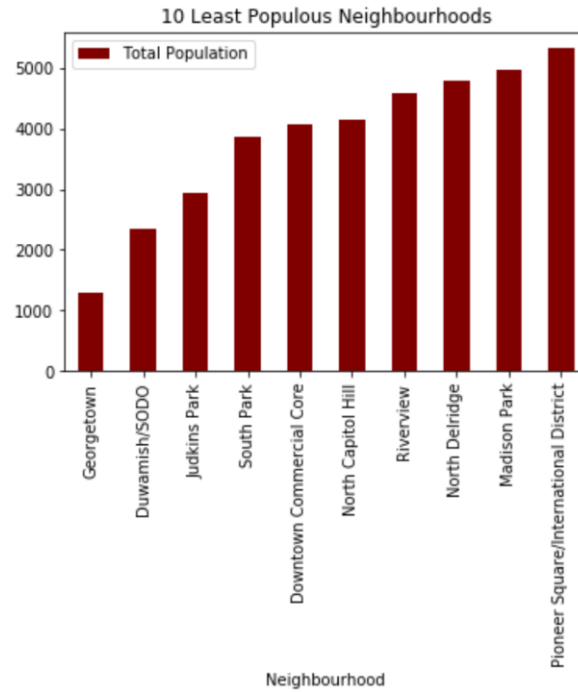
Figure 24: Ten Least Populous Neighbourhoods on a graph

3.6. Venue Analysis

3.6.1. Obtaining Nearby Venues using Foursquare API

A function was devised to derive nearby venues in Seattle's given neighbourhoods. The venues are stored in a new DataFrame called "seattle_venues" as shown in the notebook, a preview of which is shown in Figure 25 below.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Arbor Heights	47.546653	-122.383186	Beveridge Place Pub	47.545380	-122.387385	Pub
1	Arbor Heights	47.546653	-122.383186	The New Bridge	47.546570	-122.387371	Pub
2	Arbor Heights	47.546653	-122.383186	Pet Elements	47.544600	-122.387438	Pet Store
3	Arbor Heights	47.546653	-122.383186	Thriftway	47.544537	-122.386412	Grocery Store
4	Arbor Heights	47.546653	-122.383186	Whiskey West	47.545124	-122.387549	Whisky Bar

Figure 25: Nearby Venues per Neighbourhood

Next, the "seattle_venues" DataFrame was analyzed for the number of venues returned for each neighbourhood. As a result, 10 neighbourhoods with the most venues is shown in Figure 26 and Figure 27:
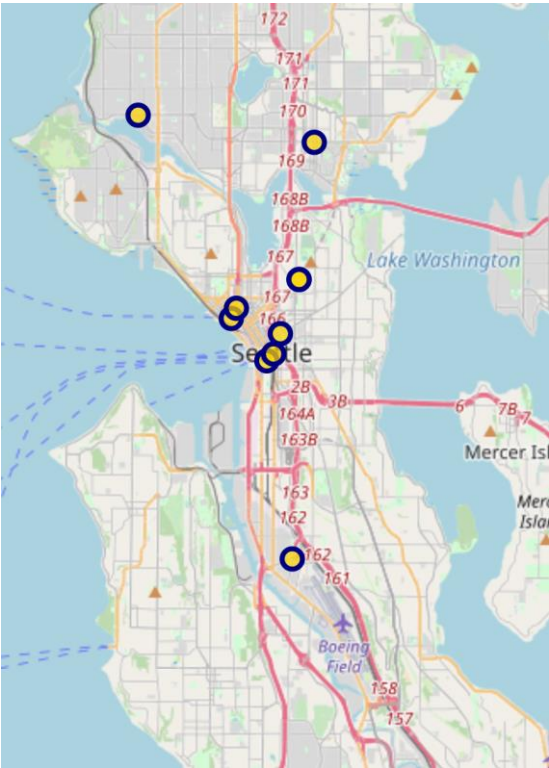
Figure 26: Ten Neighbourhoods with the most Venues on a map

	Neighbourhood	Venue
0	Pioneer Square/International District	100
1	Belltown	100
2	Montlake/Portage Bay	100
3	Judkins Park	92
4	Sunset Hill/Loyal Heights	85
5	Queen Anne	83
6	South Park	57
7	Central Area/Squire Park	53
8	Capitol Hill	50
9	First Hill	50

Figure 27: Ten Neighbourhoods with the most Venues on a graph

3.6.2. Neighbourhoods' Venue Analysis

Firstly, using the venue category of the recently created "seattle_venues" DataFrame, one hot encoding was used to create a DataFrame which encompasses the most common venues for each neighbourhood. Now another DataFrame called "seattle_grouped" was created which depicted the mean of the frequency of occurrence of each venue category with respect to the neighbourhood. A preview of which is shown below in Figure 28.

```
seattle_grouped = seattle_onehot.groupby('Neighbourhood').mean().reset_index()
seattle_grouped.head()
```

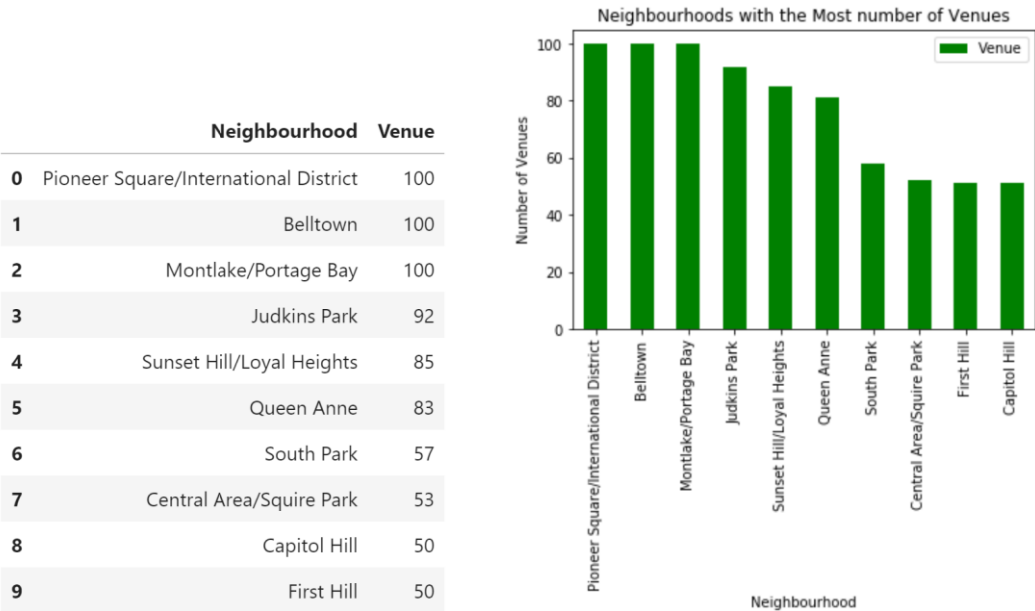|	Neighbourhood	ATM	Advertising Agency	African Restaurant	American Restaurant	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	A Museun
0	Alki/Admiral	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.
1	Arbor Heights	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.
2	Ballard	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.
3	Beacon Hill	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.
4	Belltown	0.0	0.0	0.0	0.010417	0.0	0.0	0.0	0.0	0.

Figure 28: DataFrame "seattle_grouped" one hot encoded

Using the recently created DataFrame, each neighbourhood along with the 10 most common venues was printed, a preview of which is shown below in Figure 29:

```
----Alki/Admiral----                          ----Ballard----
                 venue  freq                                venue  freq
0                 Park  0.23          0                      Park  0.19
1          Bus Station  0.15          1        Convenience Store  0.06
2                  Gym  0.08          2              Coffee Shop  0.06
3          Soccer Field 0.08          3            Grocery Store  0.06
4           Coffee Shop 0.08          4              Music Store  0.06
5           Skate Park  0.08          5                   Winery  0.06
6   Arts & Crafts Store 0.08          6                 Bus Stop  0.06
7        Sandwich Place 0.08          7              Gas Station  0.06
8             Bus Stop  0.08          8                     Food  0.06
9             Bus Line  0.08          9         Volleyball Court  0.06


----Arbor Heights----                         ----Beacon Hill----
                 venue  freq                                venue  freq
0           Pizza Place 0.14          0              Golf Course  0.17
1                  Pub  0.10          1                    Trail  0.08
2        Sandwich Place 0.10          2        Mexican Restaurant 0.08
3   Rental Car Location 0.05          3                Club House 0.08
4   Comfort Food Restaurant 0.05      4               Coffee Shop 0.08
5        Supplement Shop 0.05         5            Pilates Studio 0.08
6    Sporting Goods Shop 0.05         6                Playground 0.08
7           Gas Station  0.05         7                     Park 0.08
8     Chinese Restaurant 0.05         8        Golf Driving Range 0.08
9             Pet Store  0.05         9               Sports Club 0.08
```

Figure 29: A preview of most common Venues per Neighbourhood

To make it presentable, the previous information was put into a DataFrame called "neighbourhoods_venues_sorted." A preview of which is shown below in Figure 30.

	Neighbourhood	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	Alki/Admiral	Park	Bus Station	Arts & Crafts Store	Gym	Sandwich Place	Coffee Shop	Soccer Field	Skate Park	Bus Stop	Bus Line
1	Arbor Heights	Pizza Place	Pub	Sandwich Place	Storage Facility	Pet Store	Supplement Shop	Fried Chicken Joint	Chinese Restaurant	Gas Station	Sporting Goods Shop
2	Ballard	Park	Grocery Store	Food	Music Store	Spa	Bus Stop	Gas Station	General Travel	Pharmacy	Coffee Shop
3	Beacon Hill	Golf Course	Pub	Pilates Studio	Mexican Restaurant	Park	Trail	Club House	Golf Driving Range	Coffee Shop	Sports Club
4	Belltown	Bar	Coffee Shop	Breakfast Spot	Hotel	Seafood Restaurant	Gym	Bakery	Pizza Place	Italian Restaurant	Sushi Restaurant

Figure 30: Most Common Venues per Neighbourhood in a DataFrame

3.6.3. Neighbourhood Clustering

The machine learning model used in the analysis is called "K-means Clustering." To begin modeling, an optimum k value or in other words the number of clusters needed to successfully cluster neighbourhoods is required.

Five Tests were carried out, the details of which are shown in the Python Notebook [3]. Results of the tests are displayed below and in Figure 31:

1) Test 1: Elbow Method resulted in k value to be inconclusive.
2) Test 2: Elbow Method with different parameters resulted in k value to be inconclusive.
3) Test 3: Elbow Method with different parameters resulted in k value to be inconclusive.
4) Test 4: Elbow Method with different parameters resulted in k value to be 3.
5) Test 5: Silhouette Method resulted in k value to be 3.
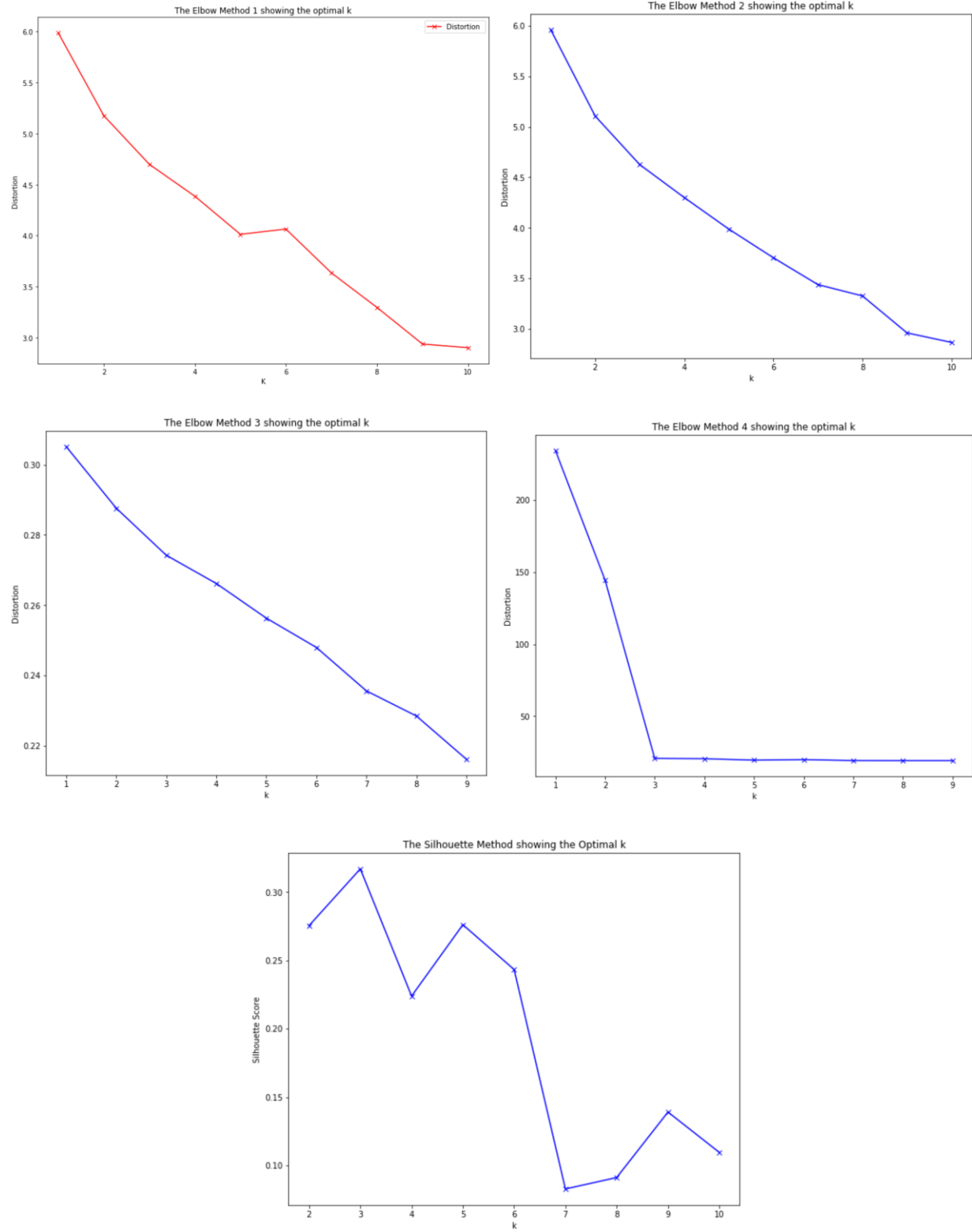
Hence, k was chosen to be 3.

Figure 31: K optimum Graphical Results

4. Results

The resulting DataFrame with cluster labels is called "seattle_merged." This DataFrame encompasses the cluster label as well as the 10 most common venues for each neighbourhood. A preview of the DataFrame is shown below in Figure 32 and the clusters are depicted on the Seattle map as shown in Figure 33:

	Neighbourhood	Latitude	Longitude	Cluster Labels	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
1	Arbor Heights	47.546653	-122.383186	2	Pizza Place	Pub	Sandwich Place	Supplement Shop	Japanese Restaurant	Storage Facility	Fried Chicken Joint	Chinese Restaurant	Gas Station	Shipping Store
2	Fauntleroy/Seaview	47.565040	-122.376154	2	Coffee Shop	BBQ Joint	Golf Course	Gym	Sports Bar	Brewery	Taco Place	Furniture / Home Store	Spa	Salon / Barbershop
3	West Seattle Junction/Genesee Hill	47.578940	-122.410670	2	Coffee Shop	Ice Cream Shop	Brewery	Seafood Restaurant	Art Gallery	Mexican Restaurant	Park	Sandwich Place	Market	Thai Restaurant
4	Alki/Admiral	47.564730	-122.363860	0	Bus Station	Park	Arts & Crafts Store	Sandwich Place	Skate Park	Soccer Field	Construction & Landscaping	Coffee Shop	Gym	Falafel Restaurant
5	North Delridge	47.547040	-122.368940	1	Park	Trail	Playground	Zoo Exhibit	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market	Fish & Chips Shop

Figure 32: Ten most common venues per neighbourhood with Cluster Labels
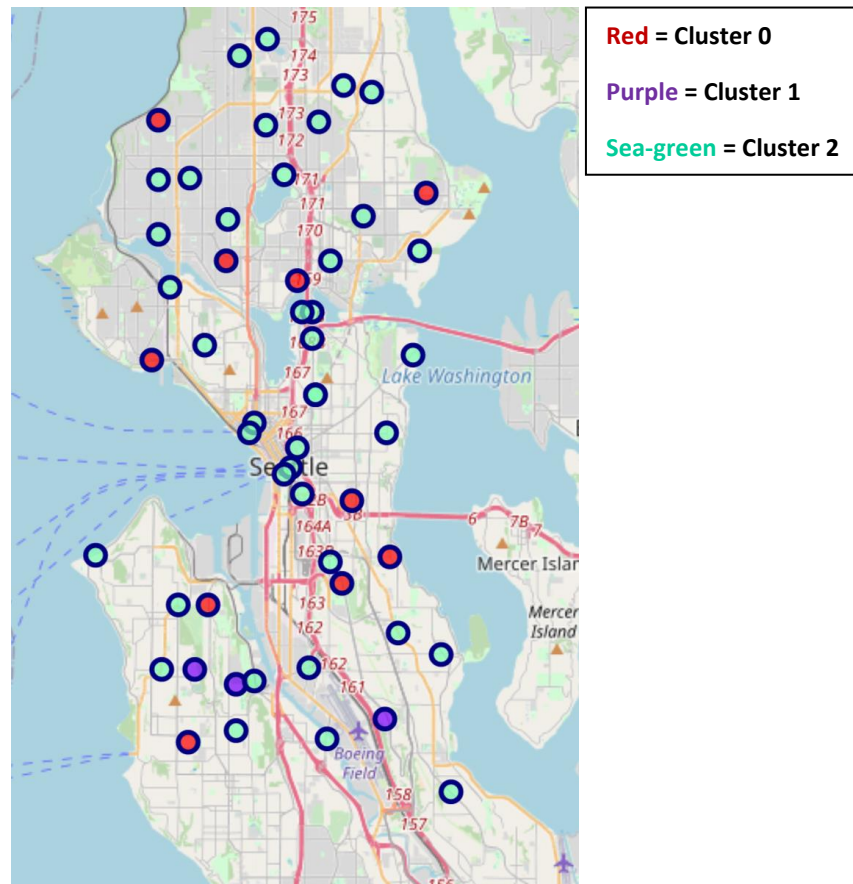
Visualizing the Clusters:

Figure 33: Cluster Labels in Seattle

<u>Examining the Clusters</u>:

Cluster 0 and Cluster 1 comprise mostly of outdoors activities venues, with Parks being the most common, followed by Trails as shown in Figure 34 and Figure 35.

	Neighbourhood	Cluster Labels	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
4	Alki/Admiral	0	Bus Station	Park	Arts & Crafts Store	Sandwich Place	Skate Park	Soccer Field	Construction & Landscaping	Coffee Shop	Gym	Falafel Restaurant
7	Riverview	0	Eye Doctor	Pool	Park	Soccer Field	Zoo Exhibit	Food	Flower Shop	Flea Market	Fish Market	Fish & Chips Shop
14	Beacon Hill	0	Park	Golf Course	Pub	Playground	Mexican Restaurant	Golf Driving Range	Skate Park	Coffee Shop	Trail	Dog Run
18	Seward Park	0	Harbor / Marina	Park	Pizza Place	Pet Store	Zoo Exhibit	Eye Doctor	Food	Flower Shop	Flea Market	Fish Market
20	Madrona/Leschi	0	Park	Dance Studio	Yoga Studio	Trail	Rental Service	Bus Station	Tunnel	Skate Park	South American Restaurant	Café
30	Ravenna/Bryant	0	Art Gallery	Bus Stop	Park	Zoo Exhibit	Fabric Shop	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
40	Haller Lake	0	Park	Construction & Landscaping	Garden Center	Bus Stop	Café	Zoo Exhibit	Falafel Restaurant	Food & Drink Shop	Food	Flower Shop
44	Ballard	0	Park	Music Store	Volleyball Court	Gas Station	Coffee Shop	Grocery Store	Bus Stop	Spa	Mexican Restaurant	Convenience Store
45	Fremont	0	Park	Seafood Restaurant	Art Gallery	Café	Boat or Ferry	Lounge	Harbor / Marina	Pizza Place	Bus Stop	Bus Line
47	Cascade/Eastlake	0	Harbor / Marina	Burger Joint	Bus Line	Sporting Goods Shop	Seafood Restaurant	Park	Zoo Exhibit	Fabric Shop	Flower Shop	Flea Market

Figure 34: Cluster 0

	Neighbourhood	Cluster Labels	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
5	North Delridge	1	Park	Trail	Playground	Zoo Exhibit	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market	Fish & Chips Shop
6	High Point	1	Trail	Park	Baseball Field	Zoo Exhibit	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market	Fish & Chips Shop
12	Duwamish/SODO	1	Park	Grocery Store	Zoo Exhibit	Eye Doctor	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market	Fish & Chips Shop

Figure 35: Cluster 1

Cluster 2 comprises of restaurants, eateries, casual hangout venues, with restaurants and coffee shops being the most common. A preview of Cluster 2 is shown below in Figure 36. The full DataFrame can be found in the Python Notebook [3].

	Neighbourhood	Cluster Labels	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
1	Arbor Heights	2	Pizza Place	Pub	Sandwich Place	Supplement Shop	Japanese Restaurant	Storage Facility	Fried Chicken Joint	Chinese Restaurant	Gas Station	Shipping Store
2	Fauntleroy/Seaview	2	Coffee Shop	BBQ Joint	Golf Course	Gym	Sports Bar	Brewery	Taco Place	Furniture / Home Store	Spa	Salon / Barbershop
3	West Seattle Junction/Genesee Hill	2	Coffee Shop	Ice Cream Shop	Brewery	Seafood Restaurant	Art Gallery	Mexican Restaurant	Park	Sandwich Place	Market	Thai Restaurant
8	Roxhill/Westwood	2	Baseball Field	Burger Joint	Convenience Store	Café	Theater	Playground	Zoo Exhibit	Fish & Chips Shop	Fast Food Restaurant	Filipino Restaurant
9	Highland Park	2	Mexican Restaurant	Coffee Shop	Pizza Place	Grocery Store	Harbor / Marina	Bar	Bakery	Café	Brewery	Restaurant
10	South Park	2	Park	Bar	Café	Dive Bar	Brewery	Pizza Place	Dessert Shop	Lounge	Gas Station	Coffee Shop
11	Georgetown	2	Sandwich Place	Wine Bar	Coffee Shop	Fish & Chips Shop	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Zoo Exhibit
13	South Beacon Hill/NewHolly	2	Mexican Restaurant	Café	Coffee Shop	Pub	Peking Duck Restaurant	Mediterranean Restaurant	Park	Fried Chicken Joint	Scenic Lookout	Bakery
15	North Beacon Hill/Jefferson Park	2	Garden	Bus Stop	Baseball Field	Zoo Exhibit	Falafel Restaurant	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
16	Rainier Beach	2	Pizza Place	Coffee Shop	Bar	Ice Cream Shop	African Restaurant	Diner	Pet Store	Gastropub	Bookstore	Sushi Restaurant
17	Columbia City	2	Music Venue	Burger Joint	Zoo Exhibit	Deli / Bodega	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market

Figure 36: Cluster 2

5. Discussion

In order to recommend which neighbourhood(s) are best to open ice cream shops or trucks in, areas should be populous, and they should preferably have young and outdoorsy population.

Cluster 0 and Cluster 1 encompass majority of neighbourhoods, with outdoors activities' venues being the most commonly found venues, such venues include parks, trails, and playgrounds to name a few. Parks were the most common, followed by trails. Next, the condition of such neighbourhoods to have a young population (see Figure 37) and be a populous area (see Figure 38) must be satisfied, in order to have a chance at having a successful ice cream business.

Top 10 Neighbourhoods with most Population Under 18 in Seattle

```
top10_neigh_Pop18Under = pop_18_df[['Neighbourhood','Population Under 18','Latitude','Longitude']].sort_values(['Population Under 18'], ascending=False).head(10)

top10_neigh_Pop18Under
```

	Neighbourhood	Population Under 18	Latitude	Longitude
30	Ravenna/Bryant	4227	47.68033	-122.27290
13	South Beacon Hill/NewHolly	3892	47.57686	-122.31271
38	Greenwood/Phinney Ridge	3870	47.68508	-122.33232
17	Columbia City	3692	47.55115	-122.26680
50	Queen Anne	3623	47.61576	-122.34464
16	Rainier Beach	3389	47.55687	-122.28452
3	West Seattle Junction/Genesee Hill	3202	47.57894	-122.41067
31	Wedgwood/View Ridge	3179	47.66398	-122.27582
48	Magnolia	2655	47.65399	-122.37969
32	Laurelhurst/Sand Point	2645	47.70013	-122.31765

Figure 37: Ten Neighbourhoods with the Most Population under 18 DataFrame

Population Analysis Dataframe:

Top 10 Most Populous Neighbourhoods

```
Most_Pop_neigh = census_df[['Neighbourhood','Total Population','Latitude','Longitude']].sort_values(['Total Population'], ascending=False).head(10)

Most_Pop_neigh
```

	Neighbourhood	Total Population	Latitude	Longitude
50	Queen Anne	35458	47.61576	-122.34464
30	Ravenna/Bryant	24187	47.68033	-122.27290
38	Greenwood/Phinney Ridge	23948	47.68508	-122.33232
24	Capitol Hill	19078	47.62396	-122.31882
29	University District	19051	47.67393	-122.29914
3	West Seattle Junction/Genesee Hill	17713	47.57894	-122.41067
22	Central Area/Squire Park	16969	47.60878	-122.32643
17	Columbia City	16883	47.55115	-122.26680
46	Wallingford	16014	47.64708	-122.32477
45	Fremont	15626	47.65555	-122.32650

Figure 38: Ten Most Populous Neighbourhoods DataFrame

Now to verify if any of the **13 neighbourhoods** found in **Cluster 0** and **Cluster 1** happen to have the characteristics laid out previously. As shown below, only **2 neighbourhoods** were found to fit the criteria out of which only **1 neighbourhood** satisfied **both conditions**. The striked-out neighbourhoods below were found to be unsatisfactory regarding the conditions laid out earlier. Only one neighbourhood called "Ravenna/Bryant" was found in the most populous neighbourhood DataFrame and most population under 18 neighbourhoods DataFrame.

- ~~Alki/Admiral~~
- ~~Riverview~~
- ~~Beacon Hill~~
- ~~Seward Park~~
- ~~Madrona/Leschi~~
- **Ravenna/Bryant (Found in Both)**
- ~~Haller Lake~~
- ~~Ballard~~
- **Fremont (Found in Populous DataFrame)**
- ~~Cascade/Easlake~~
- ~~North Delridge~~
- ~~High Point~~
- ~~Duwamish/SODO~~

Therefore, it is recommended to open an ice cream shop or truck in Ravenna/Bryant neighbourhood, in the city of Seattle, Washington, USA.

6. Conclusion

Although Ravenna/Bryant neighbourhood was chosen as a suitable place to open an ice cream shop or truck, other neighbourhoods found in Cluster 0 and Cluster 1 could also be a viable candidate, given further investigation is carried out. Other business stakeholders could explore the neighbourhoods found in Cluster 0 and Cluster 1 to see the viability of opening a

business which also caters to an outdoorsy population. Potential businesses in this cluster could include, smoothie shop/truck, snack shop/truck etc.

Following DataFrames shown below were not utilized but could be used for future research for other types of businesses:

- Race/Ethnicity DataFrame
- Housing DataFrame
- Populous Neighbourhoods DataFrame

7. References

[1] Wikipedia Contributors, "Seattle," Wikipedia, 27-Aug-2020. [Online]. Available: https://en.wikipedia.org/wiki/Seattle. [Accessed: 27-Aug-2020]

[2] "About Seattle - OPCD | seattle.gov," Seattle.gov, 2019. [Online]. Available: https://www.seattle.gov/opcd/population-and-demographics/about-seattle. [Accessed: 27-Aug-2020]

[3] ShaheerKhan200, "ShaheerKhan200/Ice-Cream-Business-in-Seattle---Analysis," *GitHub*, 2020. [Online]. Available: https://github.com/ShaheerKhan200/Ice-Cream-Business-in-Seattle---Analysis. [Accessed: 11-Sep-2020].