Calgary's passion for Coffee

Introduction

The final part of the IBM Capstone Project in Coursera consists in a small taste of the real work that data scientists perfom daily. The objectives of this project were to define a problem that could be solved using the tools and procedures that were described during the previous courses. This problem must be solved using public data from the internet and must include the usage of Foursquare as a tool to find venues in a particular zone.

The main topic of this assignment is related to the passion of Calgarians for coffee. During the period I have lived in this city (2 years) I have witnessed the love that Calgarians have for coffee and their extreme interest in discovering new places to drink coffee and chat. Almost all meetings in Calgary are held in a café or coffee shop. Calgarians often have a café to chat with friends, one for business, one for meeting new people and one for networking. This passion of Calgarians for coffee transformed a juice/water drinker as me, into an avid coffee fan that has decided to write the introduction for his assignments in a Starbucks!

The main objective of this study is to define which are the most important coffee shops in Calgary depending on the neighbourhood and brand. In the following sections I will explain the problem, data preparation, interpretation and conclusions. All the codes are included in my Github repository.

Data

The data that was used for this assignment came principally from Wikipedia: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary. This table was wrangled using *requests* and *BeautifulSoup*. Because the list of neighbourhoods in Wikipedia included absolutely all the neighbourhoods a sub-set of them was defined taking into account that most of the cafés and coffee shop are located in downtown.

The information from the neighbourhood table was adjusted in order to show only the area of interest. After this, *geopy* was used to calculate the coordinates for each of the neighbourhoods. Once this was done, it was possible to use folium to plot the relative location of each neighbourhood on a Calgary map.

Using Foursquare API, the principal venues for each neighbourhood were found and it was possible to demonstrate that, in these communities, the most popular venue are coffee shops or cafés. Once the list of venues was filtered in order to include only coffee shops and cafés, the location of each of them was mapped according to their location. Finally, the most popular coffee shops were defined and their positions in the neighbourhoods were located in a map.

Methodology

The methodology that was followed can be simplified in three steps: data wrangling, data preparation and analysis. In data wrangling the input data is taken from the source and converted into a dataframe that can be handled by Python. After this step, this data is cleansed and prepared in a format that is suitable for the analysis that will be performed. Once the data is ready, then it can be analyzed through plots and maps.

Results

In the following sections the principal results that were obtained at each stage of the study will be clearly explained as well as the assumptions and simplifications that were taken in each step.

a) Data Wrangling

The origin of the data came from the webpage https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary in Wikipedia. This webpage contains a table just as the one is shown in Figure 1. The important aspects of this table are the name of each neighbourhood, their quadrant, sector and area. Once the information of this table is taken, the irrelevant columns and rows will be removed in order to leave a functional dataframe that will be used to the rest of the study.

Name[9]	Quadrant	Sector[10]	Ward[11]	Type[10]	2012 Population Rank	Population (2012)[9]	Population (2011)[9]	% change	Dwellings (2012)[9]	Area (km²)[10]	Population density
Abbeydale	NE/SE	Northeast	10	Residential	82	5,917	5,700	3.8	2,023	1.7	3,480.6
Acadia	SE	South	9	Residential	27	10,705	10,615	0.8	5,053	3.9	2,744.9
Albert Park/Radisson Heights	SE	East	10	Residential	75	6,234	6,217	0.3	2,709	2.5	2,493.6
Altadore	SW	Centre	11	Residential	39	9,116	8,907	2.3	4,486	2.9	3,143.4
Alyth/Bonnybrook	SE	Centre	9	Industrial	208	16	17	−5.9	14	3.8	4.2
Applewood Park	SE/NE	East	10	Residential	69	6,498	6,404	1.5	2,215	1.6	4,061.3
Arbour Lake	NW	Northwest	2	Residential	26	10,836	10,762	0.7	3,918	4.4	2,462.7
Aspen Woods	SW	West	6	Residential	92	5,271	4,469	17.9	2,281	3.8	1,387.1
Auburn Bay	SE	Southeast	12	Residential	60	7,193	5,769	24.7	2,808	4.5	1,598.4
Aurora Business Park	NE	North	3	Industrial	237	0	0	—	0	2.4	0

Figure 1

In order to take the information from this table, the module *BeautifulSoup* was employed. Figure 2 shows the code that was used to take this data. The table was stored in a variable called *data* that must be complemented with the names of each of the headings.

```
#Import Data from Wikipedia
website_text = requests.get('https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary').text
soup = BeautifulSoup(website_text,'xml')

table = soup.find('table',{'class':'wikitable sortable'})
table_rows = table.find_all('tr')

data = []
for row in table_rows:
    data.append([t.text.strip() for t in row.find_all('td')])
```

Figure 2

b) Data preparation

After adding the headings to the dataframe that was just created, it looked like it is shown in Figure 3. This figure is showing only the first 5 rows. As it was mentioned before, some of the columns were removed because were not important for this project. In addition, some of the rows were also removed. This study was focused only on the cafés in the centre of Calgary, for this reason, all the neighbourhoods that belonged to a sector that was not Centre were removed.

	Name	Quadrant	Sector	Ward	Type	Population 2012 (Rank)	Population 2012	Population 2011	%Change	Dwellings	Area (km2)	Density
0	None	None	None	None	None	None	None	None	None	None	None	None
1	Abbeydale	NE/SE	Northeast	10	Residential	82	5,917	5,700	3.8	2,023	1.7	3,480.6
2	Acadia	SE	South	9	Residential	27	10,705	10,615	0.8	5,053	3.9	2,744.9
3	Albert Park/Radisson Heights	SE	East	10	Residential	75	6,234	6,217	0.3	2,709	2.5	2,493.6
4	Altadore	SW	Centre	11	Residential	39	9,116	8,907	2.3	4,486	2.9	3,143.4

Figure 3

c) Coordinates calculation

The final dataframe looked like Figure 4 were only the first 5 rows are shown. This table is also showing the coordinates for each of the neighbourhoods. These coordinates were calculated using the *geopy* module that finds the latitude and longitude when the name of each neighbourhood is supplied.

	Name	Quadrant	Population 2012	Area (km2)	Latitude	Longitude
1	Altadore	SW	9,116	2.9	51.0151	-114.1007
2	Banff Trail	NW	3,837	1.5	51.0714	-114.1116
3	Bankview	SW	5,221	0.7	51.0338	-114.0995
4	Bel-Aire	SW	424	0.3	37.7625	-97.2660
5	Beltline	SW/SE	19,681	2.9	51.0404	-114.0725
6	Bridgeland/Riverside	NE/SE	5,594	3.1	51.0538	-114.0443
7	Britannia	SW	803	0.5	51.0125	-114.0834
8	Cambrian Heights	NW	2,059	0.9	51.0871	-114.0880
9	Capitol Hill	NW	4,015	1.4	38.8895	-77.0094
10	Chinatown	SW/SE	1,577	0.2	40.7164	-73.9962

Figure 4

d) Neighbourhood map

With the latitude and longitude of each neighbourhood it was possible to plot each of these coordinates in a map using *Folium*. Figure 5 shows this map. Each dot represents a neighbourhood in the Calgary centre.

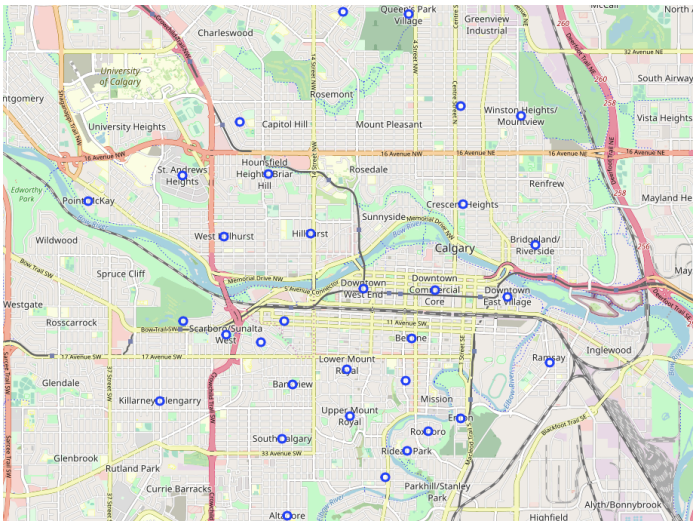

Figure 5

e) Coffee shops and cafés by each neighbourhood

In order to determine which are the coffee shops and cafés in the Calgary centre's neighbourhoods it is necessary first to know which are the different venues that are located in these communities. In order to do this it was necessary to use Foursquare API. After obtaining the url using the credentials, radius and the coordinates for each neighbourhood it was possible to obtain a JSON file with all the venue's information. This JSON file was converted into a dataframe that only showed the coffee shops and cafés as is presented in Figure 6.

	Community	Community Latitude	Community Longitude	Venue	Venue Latitude	Venue Longitude	Venue_Category
0	Altadore	51.0151	-114.1007	Monogram Coffee	51.010967	-114.100054	Coffee Shop
1	Banff Trail	51.0714	-114.1116	Weeds Cappucino Bar	51.070377	-114.104793	Coffee Shop
2	Bankview	51.0338	-114.0995	Starbucks	51.031653	-114.095192	Coffee Shop
3	Beltline	51.0404	-114.0725	Starbucks	51.042703	-114.071435	Coffee Shop
4	Beltline	51.0404	-114.0725	Philosafy	51.037911	-114.076205	Café

Figure 6

Discussion

Figure 7 shows the amounts of venues by each type in the neighbourhoods that are located in Calgary centre. It can be seen how the amount of cafés and coffee shops exceeds any other type of venue. This is showing how important are coffee shops for Calgarians, particularly those that live or work in downtown.

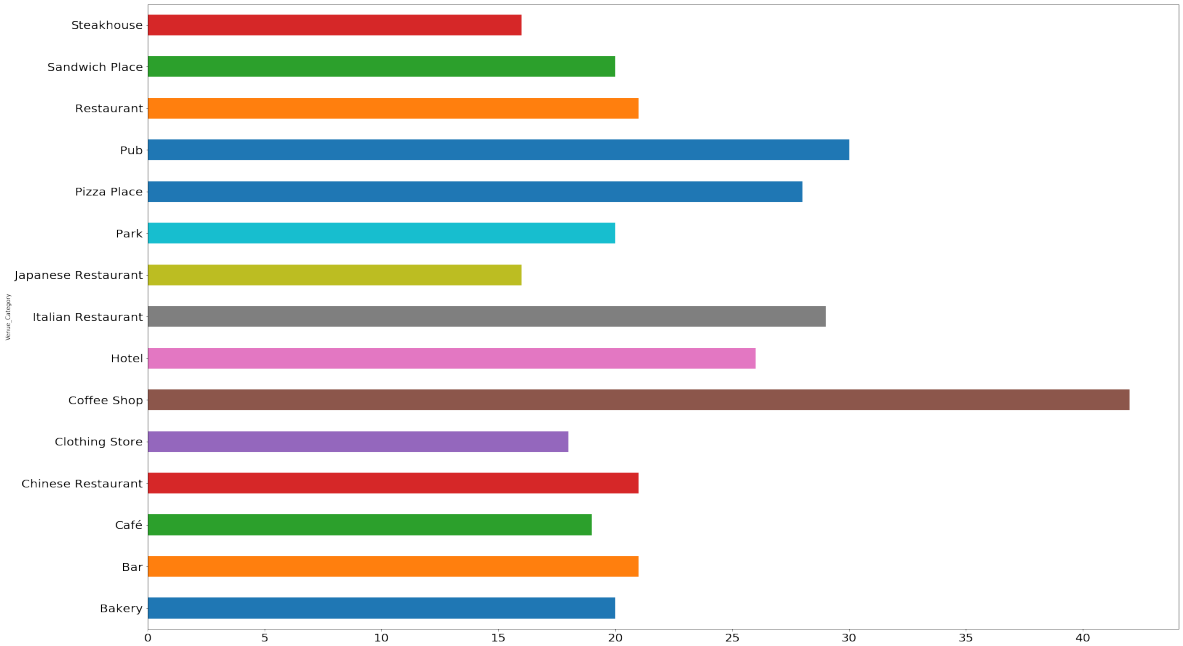

Figure 7

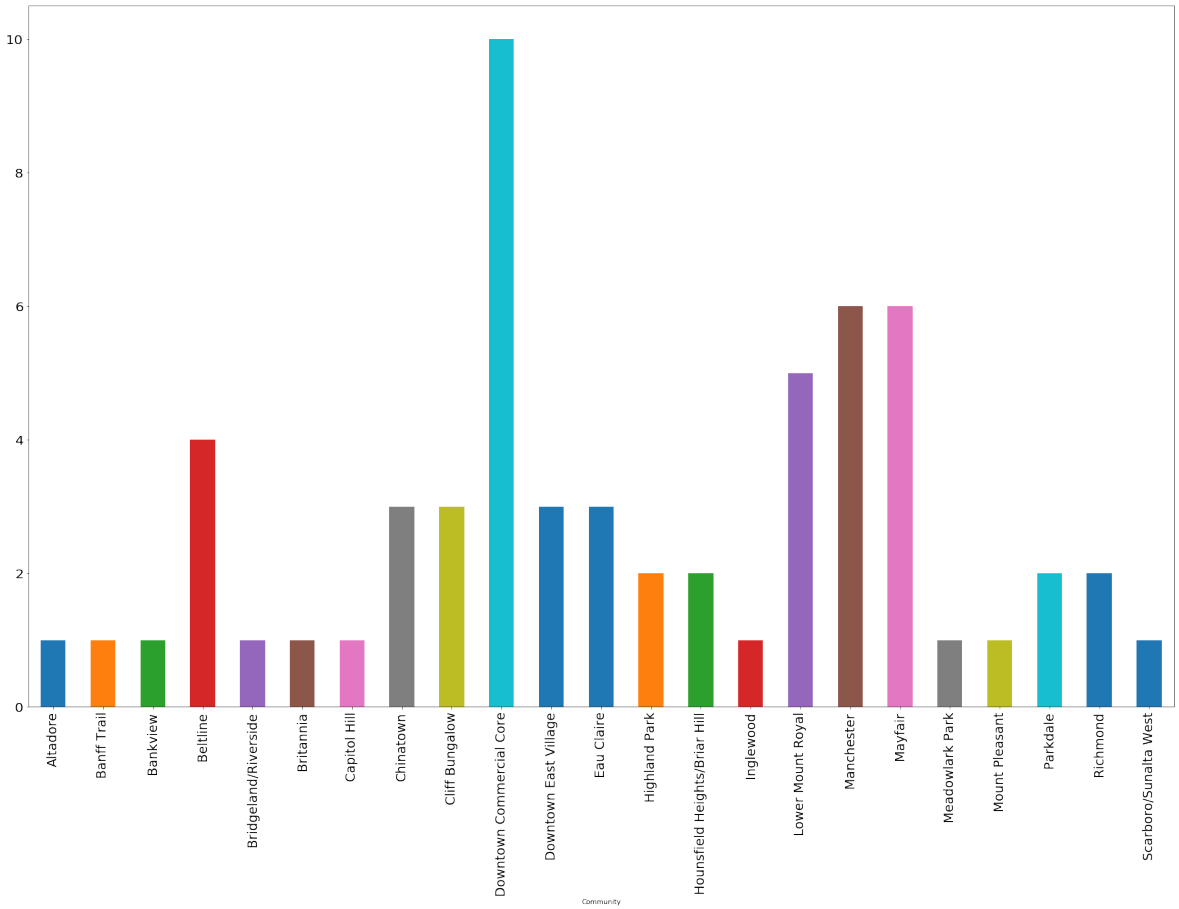

Figure 8

It is important to consider that the Downtown Commercial Core that is where most of the big companies are located is the neighbourhood in Calgary that has the highest number of coffee shops or cafés as it is show in Figure 8.

Figure 9 shows a map with the location of the coffee shops in Calgary centre. As it was mentioned before, the zone in downtown is the one with the highest number of coffee shops.

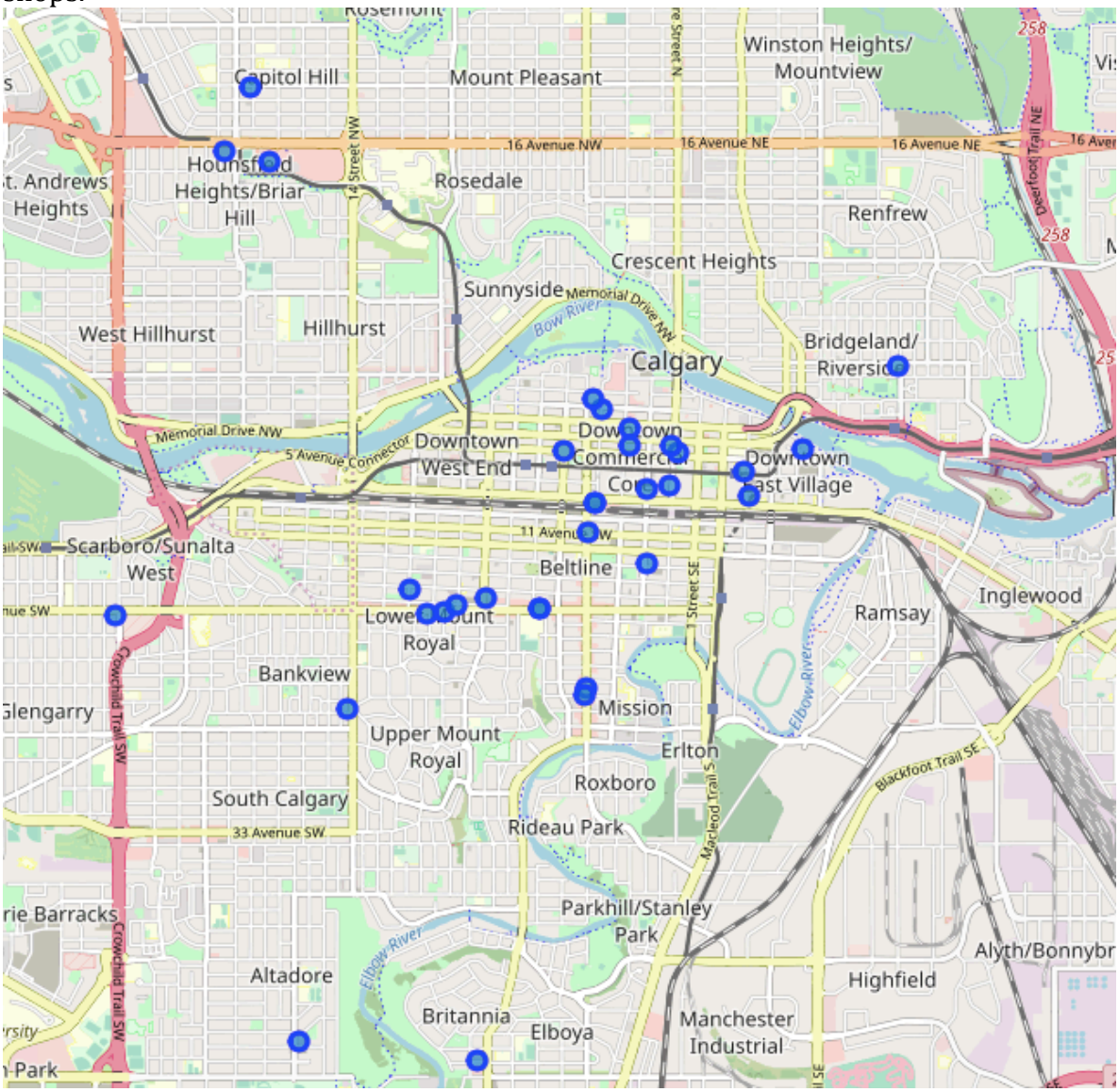

Figure 9

Of the 39 different types of coffee shops in Calgary, there are 9 that have at least two different locations. These coffee shops or cafés are: Starbucks, Second Cup, Tim Hortons, Philosafy, Phil & Sebastian Coffee Roasters and Monogram Coffee. Figure 10 shows the distribution of these coffee shops.

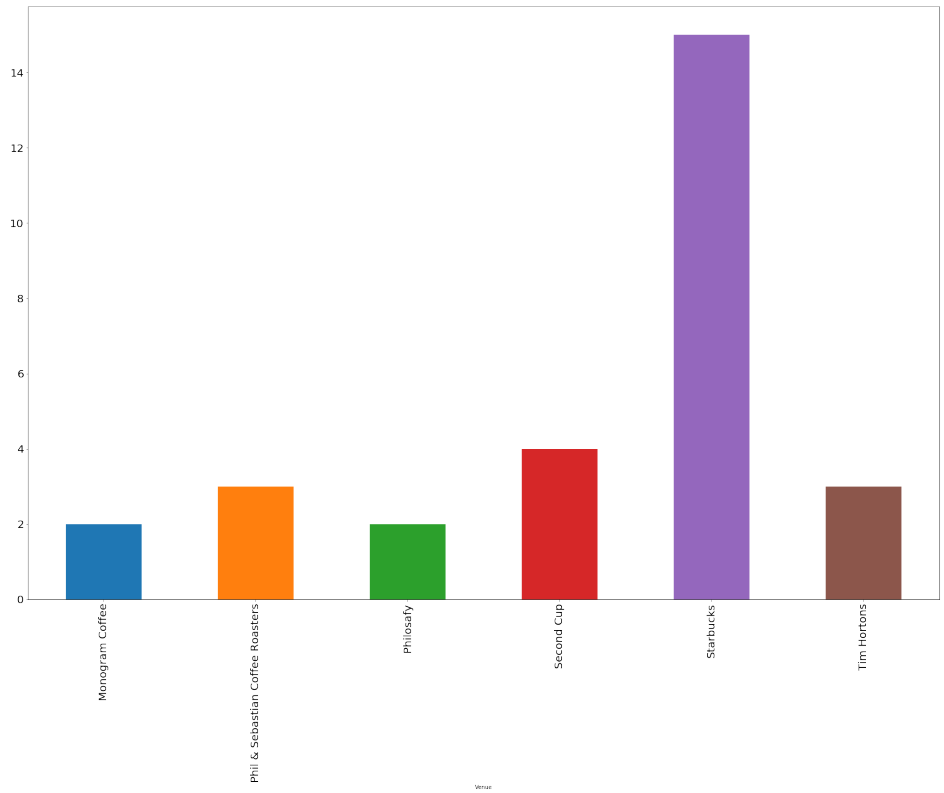

Figure 10

Finally, all these different types of coffee shops or cafés were located in a Calgary map according to their type. This is shown in Figure 11.

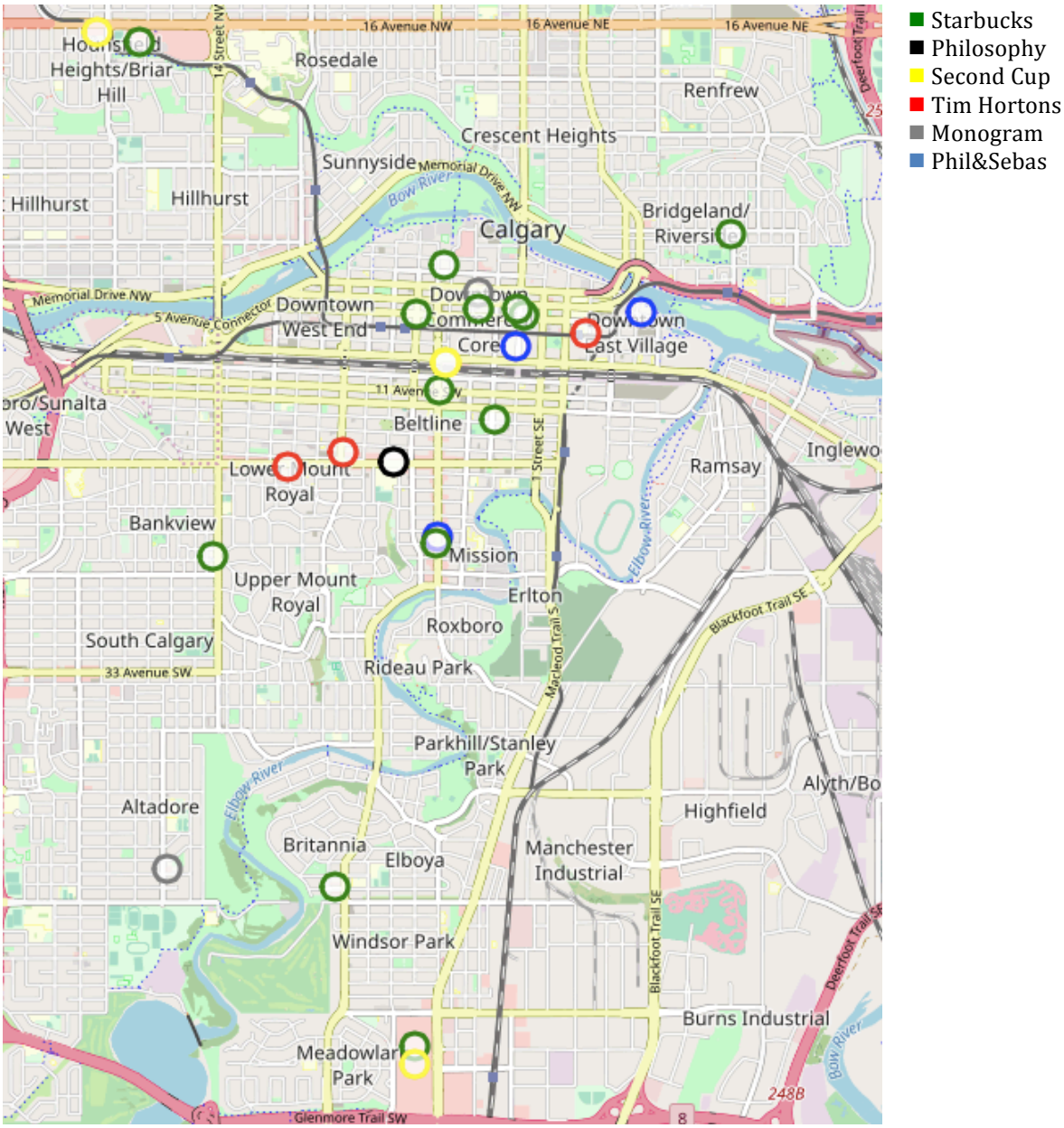

Figure 11

Conclusion

In this short study it was possible to find and locate on a map the different venues in a particular zone of Calgary. The coffee shops and cafés were separated from the rest because they were predominantly more than the rest of venues. It was concluded that the neighbourhood in Calgary with the highest number of coffee shops was the Downtown Commercial Core and that the type of coffee shop that had more locations was Starbucks, followed by Tim Hortons.