

1. Introduction/Business Problem

1.1: Background

2020 has been a strange and turbulent year, and for many businesses (especially the small ones) it has been a stressful and worrisome time. Some of the hardest hit were the shops and eateries of downtown cores the world over. Recent and ongoing developments surrounding the COVID-19 Pandemic have seen an exodus of employees and activity out of characteristically bustling business centres, and as workers retreated to the safety of their homes their employers were fast in transitioning towards remote work as a means of keeping their operations up and running. This, however, was and is not a solution for those brick-and-mortar shops in these business centres who relied on the routine traffic that had come to define their metropolitan locales, and with many businesses coming to embrace the ease, efficiency, and cost savings stemming from operating remotely, many of these downtown establishments have been made to face the reality that much of customer base may never return.

Great change, however, can be accompanied by great opportunity, and a young barista with business aspirations of their own thinks they might have an opportunity to seize upon their own small share of said opportunity. This individual, henceforth referred to as 'the client', having years of experience making and serving fine craft coffee beverages and products, knows well that you can take the customer from the coffee, but you can't take the coffee from the customer. The demand is still there, and with the shift occurring so quickly and unexpectedly, that demand is almost certainly untapped. Further, with so many businesses (especially small restaurants, eateries, cafes, etc.) being forced to shut their doors the client has noted that much of the physical capital they require is now available at bargain bin prices through means such as public auctions. For many the office is now the home, but that doesn't mean they aren't still willing to spend \$5 to \$10 on a fancy morning pick-me-up, and the client plans to satisfy the demand by getting ahead of the competition by emulating the tried and true business model of the ice cream truck and adapting it to the provision of fine craft coffee.

1.2: Business Problem

Though their mobile operation is expected to be much cheaper to run than a conventional brick-and-mortar operation, the acquisition and leasing of all the necessary equipment, licenses, and permits (etc) will still be costly, and they intend to finance this operation partly by means of small business loans. As a result, the client is in the process of developing a comprehensive business plan to present to potential lenders/investors, and they have requested your services in the development of a specific piece of that business plan. You have been hired for the purpose of contributing to the market research portion of their business plan. More specifically, you have been tasked with analyzing the various communities around the city of Calgary, Alberta (where they intend to operate) with the goal being the production of intuitive visual aids to help in identifying high Vs low priority neighbourhoods/communities for the client's mobile java truck.

The Business Problem which we seek here to address is outlined explicitly below...

How can the various communities of the City of Calgary be profiled/classified for the purpose of identifying those communities which could be expected to provide the highest likelihood of positive returns for our client?

1.3: Interest

As outlined in the 'Background' section, potential investors and creditors are the intended audience. However, other entrepreneurial minds may wish to replicate the analysis for their respective niche should they too wish to capitalize on the sudden changes foisted upon our economy by COVID-19 and the associated commercial fallout in a similar manner

2. Data Acquisition and Cleaning

2.1: Data Sources and Assumptions

Our data is derived from several sources. First, our geospatial data was obtained from the City of Calgary at their website [here](#). However, at the client's request, the analysis was to be restricted to a very specific subset of Calgary communities. This came after consultation with the client from which the following determinations and assumptions were made.

- ***Business parks and Industrial regions are already the targets of many mobile food service vendors. Though these vendors produce a dissimilar product and there could be room here for future expansion, it is assumed that the highest return is to be found in residential areas.***
- ***The analysis should be restricted to residential areas since these will be the areas to which demand has transitioned away from commercial parks and the downtown core. Uniform data is also difficult/impossible to obtain for non-residential areas (such as industrial sectors) thus preventing uniform analysis.***
- ***Downtown/City Centre is to be disqualified from analysis as this is the region from which workers and the economic/activity they bring with them are transitioning away from and towards their homes in the surrounding residential areas. Further, roadways are narrower, parking is in short supply, restrictions are tighter, and the market is highly saturated downtown and assumed to be the lowest return sector of the city.***
- ***Community sub-districts are assumed to include primarily condo and apartment complexes which makes their occupants more difficult to reach/access. Data is also unavailable for these areas.***

As a result of these assumptions and determinations the geojson file defining the geographic boundaries of those communities under investigation is edited to include only residential communities outside of Calgary's City Centre (downtown). This process is described in the data cleaning section of this document. The modified geojson file and containing the geospatial data actually used for mapping and for forming a template for our 'Community' feature can be found [here](#).

Community level profile data is obtained from the City of Calgary's website and can be viewed and downloaded [here](#). Median Household Income and Population are derived from each community profile. After consultation with the client the following assumptions based on their personal industry expertise were made.

- ***Coffee is universally enjoyed within all communities and regions of the city, and that the primary determinants for success when it comes to Craft Coffee providers are income (many coffee beverages are quite expensive), and population relative to existing services (market saturation).***

Therefore, only income and population figures are derived from the data available in the community profiles. Further, since this data is not made available by means of accessing individual PDF files for each community, the inclusion of other features from within the profiles would have increased the time and cost required to produce the analysis, so data from these profiles was limited to these 2 features. A CSV file containing this data can be found [here](#).

Data for venues is obtained via Foursquare's places API. A search radius of 5,000m (5km) is used to make sure each potential competitor in each neighbourhood is identified. These data are used in conjunction with population data in order to determine a measure for market saturation. Consultation with the client resulted in the following assumptions being applied to the acquisition of venue data from the places API.

- ***A potential competitor is understood as stand-alone establishments which are not part of a food court.***
- ***As well, a potential competitor is understood as those establishments for which coffee and coffee related services/products are the primary product/service provided.***

The client's experience leads them to believe that they won't be directly competing with every single food service establishment which sells coffee or some coffee products as part of their menu. As a result, the API call is filtered on the following 2 Foursquare Places categories.

- **Café:** '4bf58dd8d48988d16d941735'
- **Coffee Shop:** '4bf58dd8d48988d1e0931735'

2.2: Data Acquisition and Cleaning

Since the data was obtained from multiple sources with each being sourced for their own specific features there was an abundance of pre-processing and cleaning that needed to be completed prior to carrying out the analysis.

Figure 1: Geospatial Data Pre-Cleaning

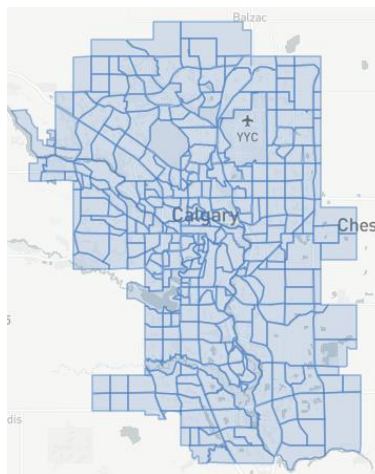
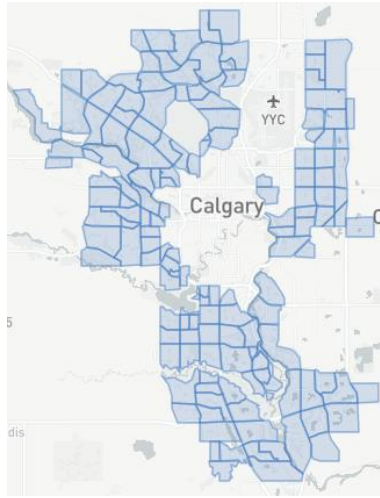


Figure 2: Geospatial Data Post-Cleaning



First the geospatial data contained in our geojson file needed to be cleaned of any entries pertaining to communities being excluded from the analysis. This meant sifting through the file and removing any keys and the corresponding data for which the 'Class' feature value was not listed as 'Residential', and as well it meant removing any keys and corresponding data for which the 'Sector' feature value was 'Centre' which denotes the City's downtown core. As stated above in the Business Problem section, the client's goal is to tap into demand that has since left the downtown core and since relocated to residential areas and suburbs where those working from home no longer have the sort of easy access that they one did to coffee shops and cafes. This actually resulted in the removal of exactly one half of the 'Community Boundaries' from the geojson file from 306 down to 153. Figures 1 Figure 2 above allow the reader to visualize the changes made. These changes were accomplished using the geojson file editor which can be found [here](#). The labels contained within the geojson file acted as the template for the entire dataset since any mapping that included community boundaries would require that the relevant labels be perfectly matched. The strings representing our community labels needed to be perfectly harmonized with those in the geojson file, and as well all of the labels contained in the file had to be contained in the column of the dataset being keyed on. As a result, data drawn from other sources is modified to match the file in later steps.

After any communities labeled as 'Centre' or anything other than 'Residential' for their 'Sector' and 'Class' features the file was loaded into a Jupyter Notebook and processed further. The raw CSV geospatial data file which describes the initial dataframe can be found [here](#). The table is too large to present the entirety of its features here so a list is provided instead. Initially the following features were included in the dataset. The only information directly meaningful to the analysis is contained in the feature describing the community name (see list below), however, some of the features are later used for the purpose of filtering during the cleaning and processing of venue data gathered from Foursquare in later steps.

- **type**: This describes the entries type within the geojson file. As usual, each of our boundaries are included under the label 'Feature', thus this feature is removed as it provides nothing of value to the analysis.

- **properties.comm_structure**: Describes the era or building status of the community (whether or not it has finished building out to its boundaries). This feature is also removed as there was little of value to be gleaned from it.
- **properties.name**: This is the community/neighbourhood name. This feature is kept and used as a template for all subsets for the purposes of mapping. This feature was renamed 'Community'.
- **properties.sector**: Describes the geographical subsector sector of the city broken down into N,S, E, W, NW, NE, SW, and SE. This feature is not considered in the analysis, but is used in a subset of the geospatial data for the purposes of filtering venue data from Foursquare later.
- **properties.class.code**: A code corresponding to the class of the case. Since at this point each entry is residential, each code is 1. This feature is removed as it is of no value for the analysis or as a filter for later cleaning/processing.
- **properties.srq**: A variation of the comm_structure feature which references only the development state and excludes the era or years of development. This feature is also removed.
- **properties.class**: A variation of the class.code feature. All are residential and the feature is removed but is also used in a subset of the full geospatial file later for the purposes of filtering venue data from Foursquare.
- **properties.comm_code**: An acronym for the properties.name feature. We did not want to work with acronyms and wanted all visual aids to be intuitive, and so this feature is removed and the community names are favoured.
- **geometry.type**: Describes the shape type of the geographical data. This is not relevant to the analysis and all mapping is handled by Folium by passing the geojson file, so this feature is removed.
- **geometry.coordinates**: Describes the coordinate set defining the boundaries of each community. This is all contained in the geojson file and is handled by Folium during the mapping process, so this feature is removed.

A small subset of the initial file is now summarized in a dataframe (CGY_COMM) which forms the basis for the overall dataset which is to be appended and completed as data is acquired from other sources, cleaned/processed, and appended to our dataframe. From this completed dataframe, several subsets are also derived for different sections of the analysis. A sample of this dataframe is shown below in Figure 3.

Figure 3: Initial CGY_COMM Dataframe

	Community	Sector	Class
0	ABBEYDALE	NORTHEAST	Residential
1	ACADIA	SOUTH	Residential
2	ALBERT PARK/RADISSON HEIGHTS	EAST	Residential
3	APPLEWOOD PARK	EAST	Residential
4	ARBOUR LAKE	NORTHWEST	Residential

Population and Income data were drawn, as mentioned, from the City of Calgary's Community profiles. A hyperlink is provided above to this list of community profiles at the City of Calgary's website. It is reported here that their data was sourced from the Statistics Canada's 2016 Census data, however, I could not locate this data at Statistics Canada's Community Data Project website, so I instead opted to extract each community's income and population figures directly from their respective profiles. Profiles

were missing for a total of 13 of the 153 communities considered in the analysis. However, these communities remained in the dataset since they could still be used in the heatmapping portion and for one of the choropleths maps in this analysis. Both population and income data are of type float. A sample of the compiled CSV file is included below in Figure 3. The features in the data are described in the list below

- **Population:** A measure of the population for each community in the dataset. This feature is used as a measure of community potential on its own during the choropleth mapping section of this analysis, but is also used to determine a measure for market saturation to be used in both the choropleth and mapping sections as well.
- **Median Inc:** A mean measure of the median household income for each community. Median household income was the most current measure I could find compiled at the community level for the city of Calgary.

Table 1: Community Profile Data Sample

Community	Population	Median Inc
ABBEYDALE	6150	81232
ACADIA	10435	72552
ALBERT PARK/RADISSON HEIGHTS	6640	64429
APPLEWOOD PARK	6850	84965
ARBOUR LAKE	10760	109790
ASPEN WOODS	9060	199759
AUBURN BAY	14850	136961
BAYVIEW	740	260339
BEDDINGTON HEIGHTS	11840	88241

After manually compiling all of the Community Profiles data for income and population a CSV file is saved to be read into the Notebook where the data processing and analysis is executed.

Instead of extracting community coordinates from the geojson file, it was simpler to use the Nominatim API and geocoding to assign coordinates to each community. Once this was accomplished the coordinates were appended to the CGY_COMM dataframe as can be seen in Figure 4 below.

Figure 4: CGY_COMM Dataframe with Coordinates Appended

	Community	Sector	Class	Latitude	Longitude
0	ABBEYDALE	NORTHEAST	Residential	51.058838	-113.929413
1	ACADIA	SOUTH	Residential	50.968655	-114.055587
2	ALBERT PARK/RADISSON HEIGHTS	EAST	Residential	51.044845	-113.990195
3	APPLEWOOD PARK	EAST	Residential	51.044858	-113.928931
4	ARBOUR LAKE	NORTHWEST	Residential	51.136786	-114.202355

Community population and income data is now also appended to the CGY_COMM dataframe to produce the dataframe below in Figure 5.

Figure 5: CGY_COMM with Community Profile Data Appended

	Community	Sector	Class	Latitude	Longitude	Population	Median Inc
0	ABBEYDALE	NORTHEAST	Residential	51.058836	-113.929413	6150.0	81232.0
1	ACADIA	SOUTH	Residential	50.968655	-114.055587	10435.0	72552.0
2	ALBERT PARK/RADISSON HEIGHTS	EAST	Residential	51.044845	-113.990195	6640.0	64429.0
3	APPLEWOOD PARK	EAST	Residential	51.044658	-113.928931	6850.0	84965.0
4	ARBOUR LAKE	NORTHWEST	Residential	51.136786	-114.202355	10760.0	109790.0

The next step was to acquire and process venue data from the Foursquare Places API. The API call, as mentioned previously, applied a search radius of 5 km and filtered on the categories 'café' and 'coffee shop'. Each community's coordinates are passed as a parameter in the API call and data for all coffee shops and cafes identified in the call were returned. A subset of this data was extracted and transformed into the dataframe CGY_JAVA. The top of this dataframe is displayed in the Figure 6 below.

Figure 6: CGY_JAVA Dataframe from Foursquare API Call

	Community	Community Latitude	Community Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ABBEYDALE	51.058836	-113.929413	Tim Hortons	51.074283	-113.957754	Coffee Shop
1	ABBEYDALE	51.058836	-113.929413	Tim Hortons	51.038402	-113.931031	Coffee Shop
2	ABBEYDALE	51.058836	-113.929413	Tim Hortons	51.058854	-113.982534	Coffee Shop
3	ABBEYDALE	51.058836	-113.929413	Starbucks	51.038989	-113.908861	Coffee Shop
4	ABBEYDALE	51.058836	-113.929413	Starbucks	51.064950	-113.956489	Coffee Shop

The features in the above CGY_JAVA dataframe are described in the list below.

- **Community (Object)**: Refers to the community corresponding to the coordinates passed in the API call. It is important to note that this does not necessarily represent the community to which the venue belongs since the 5km search radius of the call results in overlapping search radii as the loop iterates across each community in our dataset. As a result, this feature is removed from the dataframe.
- **Community Latitude (Float)**: Describes the latitude passed in the API call which identified this case/venue for the given search radius and categories. It is dropped for the same reasons as the 'Community' feature above.
- **Community Longitude (Float)**: Describes the longitude passed in the API call which identified this case/venue for the given search radius and categories. It is dropped for the same reasons as the 'Community' feature above.
- **Venue (Object)**: This is the name of the venue. This feature was not used in analysis, but was held for reference purposes.
- **Venue Latitude (Float)**: Latitude position for the given venue. This feature is kept in the dataframe for use in all sections of the analysis to follow.
- **Venue Longitude (Float)**: Longitude position for the given venue. This feature is kept in the dataframe for use in all sections of the analysis to follow.

- **Venue Category (Object)**: Describes the category to which the venue belongs according to Foursquare's categorization of venues. This feature is not used in the analysis portion but is held as a filter feature for cleaning/processing the data.

A total of 4,682 venue results were returned and translated into the initial CGY_JAVA dataframe. However, as was already explained, overlapping search radii resulting from the specified search radius are known to have resulted in the duplication of many of the returned venues. In addition, though the call filtered on the above specified categories, 138 of the venues returned possessed the Venue Category label 'Tea Room'. The duplicates were filtered from the dataframe using the 'drop_duplicates' method in Pandas by passing a subset parameter consisting of the 'Venue', 'Venue Latitude', and 'Venue Longitude' features to ensure that any distinct venues potentially sharing a space weren't incidentally dropped from the set. Those belonging to the 'Tea Room' category were simply removed by slicing the subset from the dataframe. After filtering out duplicates and unwanted categories a total 386 entries/venues remained, meaning that a total of 4,296 (or 91.8%) of the initial results were either duplicates or are assumed to not be potential competitors to the client.

Figure 7: CGY_JAVA Post Cleaning

	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Tim Hortons	51.074283	-113.957754	Coffee Shop
1	Tim Hortons	51.038402	-113.931031	Coffee Shop
2	Tim Hortons	51.056854	-113.982534	Coffee Shop
3	Starbucks	51.038989	-113.908861	Coffee Shop
4	Starbucks	51.064950	-113.956489	Coffee Shop
...

Further cleaning was also necessary due to the far-reaching nature of the specified search radius. Mapping the venue data revealed 5 venues which appeared to reside outside of Calgary. By running a simple loop passing the coordinates belonging to the indexes of these venues confirmed these suspicions. A map of the results and a list of these communities are provided below.

Figure 8: Mapped Venue Data Post-Filtering



- **Shell, 293105, CrossIron Lane, Balzac, Rocky View County, Alberta, T4A 0V1, Canada**
- **CrossIron Mills, Bass Pro Way, Balzac, Rocky View County, Alberta, T4A 0V1, Canada**
- **CrossIron Mills, Bass Pro Way, Balzac, Rocky View County, Alberta, T4A 0V1, Canada**
- **CrossIron Mills, Bass Pro Way, Balzac, Rocky View County, Alberta, T4A 0V1, Canada**
- **CrossIron Mills, Bass Pro Way, Balzac, Rocky View County, Alberta, T4A 0V1, Canada**

These observations were then removed from the venues set (CGY_JAVA) as they are too far from the city to be considered worthwhile to the client. An updated map is provided below.

Figure 9: Mapped Venue Data Post-Cleaning



Now that the venue data has been obtained accurate community labels need to be assigned to each of the venues. As explained in the description of the data the existing community labels produced by our API call loop pertained to the community to which the coordinates passed in the URL belonged and did not necessarily pertain to the community to which the actual venue coordinates belong. To obtain accurate community labels for each of the venues returned by the Foursquare API call(s) reverse geocoding was applied via a loop very similar to that which provided us with community coordinates in an earlier step. The results of this were then incorporated with the venues dataset. With accurate community label assignments included the venue set now contains all of the necessary data and it is deemed complete. A visual sample of the venue set is provided below in Figure 10. As may have already been assumed, the number of unique community results returned (130) is fewer than that of the number of communities passed in our API call and as well is fewer than that contained in our community profile set (CGY_COMM). This is simply a result of not every community containing a venue that matched our search criteria.

Figure 10: Venue Dataset Including Community Labels

	Community	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Albert Park/Radisson Heights	Tomoka Coffee	51.037764	-113.982446	Coffee Shop
1	Altadore	Monogram Coffee	51.010967	-114.100054	Coffee Shop
2	Applewood Park	Tim Hortons	51.038402	-113.931031	Coffee Shop
3	Applewood Park	Starbucks	51.038989	-113.908861	Coffee Shop
4	Applewood Park	Tim Hortons	51.038459	-113.910788	Coffee Shop

With accurate community labels assigned to each venue the venue data could be grouped according to the Community in order to provide a count of the total number of venues present in any given community. This subset is then to be merged with the community profiles set in order to complete it. However, before accomplishing this the community labels produced from the reverse geocoding process needed to be harmonized with those contained in our geojson data to make sure that consistency was maintained across all of our data sources so that any mapping could be completed without issue. As the reader may already have assumed or deduced from the maps above the venue data also contains venues belonging to communities in those regions of the city which were filtered from our initial data (those classed as non-residential or those within the centre sector). To remedy this a filter dataframe (FILTER_DF) was created from the unmodified geojson file and a function was devised to cross reference both the grouped and ungrouped venue data. From the ungrouped venue data (CGY_JAVA) a grouped subset was derived and both were referenced against the filter dataframe to detect any discrepancies. A screenshot of the code for the function is provided for the readers reference below, as is a capture of the grouped subset of the venue data which is to be appended to our community profile set following harmonization. The test argument is a list of community labels to be harmonized and the target argument is a list of community labels being referenced against. The returned dataframe (CROSS_DF) is a dataframe for which a 'False' bool value is returned after passing an community from the test list and checking for an equivalent string in the target list. This dataframe represents those community labels which are either not present or are mislabeled.

Figure 11: Cross Reference Function for Label Harmonization

```
def CrossRef(test,target):  
  
    #cross_dict = dict.fromkeys(['Community','Bool'])  
    cross_dict = {'Community':[],  
                  'Bool':[]}  
  
    for comm in test:  
  
        cross_dict['Community'].append(comm)  
        cross_dict['Bool'].append(comm in target)  
        CROSS_DF = pd.DataFrame(cross_dict)  
        CROSS_DF = CROSS_DF[CROSS_DF['Bool'] == False]  
  
    return (CROSS_DF)
```

Figure 12: Grouped Venue Set (CGY_JAVA_GROUPED) Subset for Cross Referencing

	Community	Venue Count
0	ALBERT PARK/RADISSON HEIGHTS	1
1	ALTADORE	1
2	APPLEWOOD PARK	3
3	ARBOUR LAKE	5
4	ASPEN WOODS	3

For both the grouped and ungrouped venue data an obvious discrepancy can be noted from visual inspection of the community labels contained in the filter data and those against which it is being referenced. The geojson data (which we are harmonizing to) contains capitalized labels whilst the labels return by the reverse geocoding process are a mix of upper- and lower-case characters. This is adjusted for prior to generating lists for each and passing through the cross reference function.

The grouped venue data is passed through the function first and the following dataframe is returned signalling that two labels either aren't present in the filter or data or are mislabeled. A search of string characters in the filter data confirmed that the issue was one of inconsistency in community labels and the labels were adjusted. Afterwards the function was run again to ensure that all labels in our venue set were consistent with those in the geojson file. The dataframe returned by the function is shown below for the readers reference in Figure 13.

Figure 13: Grouped Cross Reference Return DF

	Community	Bool
81	PARKHILL/STANLEY PARK	False
100	SCARBORO/SUNALTA WEST	False

Before appending the grouped data to our community profile set unwanted community classes as well as any communities belonging to the 'Centre' sector of Calgary needed to be filtered from the data. To accomplish this the now adjusted grouped venue subset was merged with the filter set to produce the dataframe described by the snippet below in Figure 14.

Figure 14: Grouped Venue Filter Dataframe

	Community	Venue Count	Sector	Class
0	ALBERT PARK/RADISSON HEIGHTS	1	EAST	Residential
1	ALTADORE	1	CENTRE	Residential
2	APPLEWOOD PARK	3	EAST	Residential
3	ARBOUR LAKE	5	NORTHWEST	Residential
4	ASPEN WOODS	3	WEST	Residential
5	AUBURN BAY	6	SOUTHEAST	Residential
6	AURORA BUSINESS PARK	1	NORTH	Industrial
7	BANFF TRAIL	1	CENTRE	Residential
8	BEDDINGTON HEIGHTS	1	NORTH	Residential
9	BELTLINE	6	CENTRE	Residential

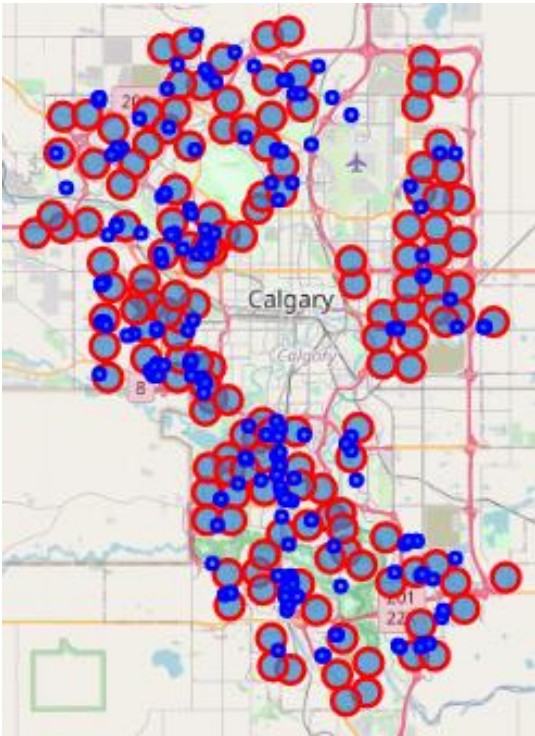
Using the ‘Class’ and ‘Sector’ labels present in the grouped venue data filter dataframe shown above a fully harmonized list of communities and their respective venue counts was produced. After doing so a total of 75 communities remained in the set, implying the same number of unique community labels would exist for the ungrouped venue set which would serve as its own standalone dataset after undergoing the same harmonizing and filtering process. Having explained the process harmonization/filtering process in detail for the grouped venue set (CGY_JAVA_GROUPED) I won’t repeat it again for the ungrouped set since the same process was repeated and the same results obtained (which could be assumed since the grouped set is merely a subset of the ungrouped set).

With the grouped venues set fully harmonized with our community profile and geospatial data the community profile set (CGY_COMM) was completed by merging the two to produce the dataframe represented by the Figure 15 below. This dataframe is referred to as CGY_COMM_FULL.

Figure 15: CGY_COMM_FULL Community Profile Dataset

	Community	Sector	Class	Latitude	Longitude	Population	Median Inc	Venue Count
0	ABBEYDALE	NORTHEAST	Residential	51.058836	-113.929413	6150.0	81232.0	NaN
1	ACADIA	SOUTH	Residential	50.968655	-114.055587	10435.0	72552.0	NaN
2	ALBERT PARK/RADISSON HEIGHTS	EAST	Residential	51.044845	-113.990195	6640.0	64429.0	1.0
3	APPLEWOOD PARK	EAST	Residential	51.044658	-113.928931	6850.0	84965.0	3.0
4	ARBOUR LAKE	NORTHWEST	Residential	51.136786	-114.202355	10760.0	109790.0	5.0

Figure 16: Map of Complete Community and Venue Datasets



A visual summary of the completed base datasets (CGY_COMM_FULL and CGY_JAVA) is provided in Figure 16 above. by mapping the communities as red markers and all of the venues as blue markers on a map of the city of Calgary.