Location for a new Italian Restaurant

CONTENTS

1 Introduction

1.1 Business Problem/Objective

The objective of this project is to find an optimal location for our client to open a ne w **Italian restaurant** in **Calgary**, Canada.

The criteria that we will use to choose the optimal location are:

- Proximity to other restaurants - look for areas of with low restaurant density
- No Italian restaurants nearby
- Preference will be given to new and upcoming neighborhoods in the west and so uth of the city

We will use our knowledge of data science to identify a few potential neighborhoods based on the above criteria. The advantages of each neighborhood will be identified to assist our client in making the final choice.

2 Data

2.1 Data Sources

The following data sources will be utilized to find the optimal location for a new Italian restaurant:

- The neighborhoods in Calgary and their geographic coordinates as well as demographic information will be obtained from the City of Calgary.
- Neighborhood demographics information including the number of restaurants, the types of restaurants and their locations in each neighborhood will be obtained using the FourSquare API.

3 Methodology

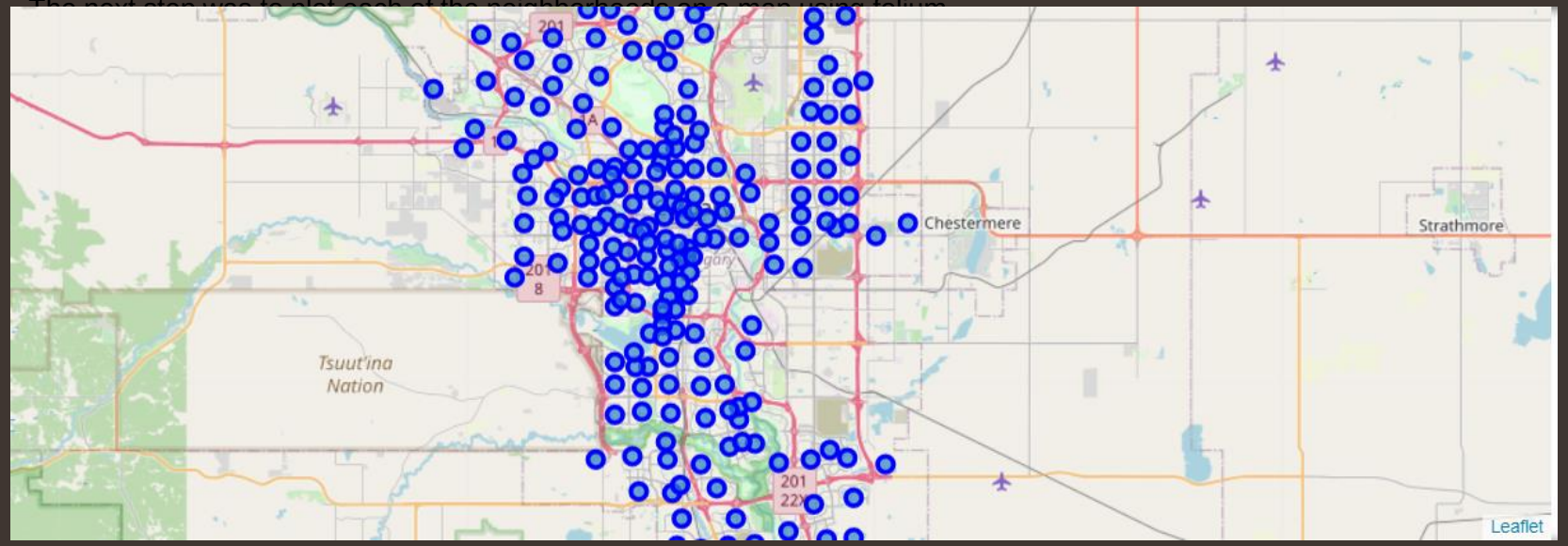
3.1 Obtain coordinates for all neighborhoods

The coordinates for each neighborhood were obtained using the following code and t he json file was converted to a pandas dataframe.

```
# Unauthenticated client only works with public data sets. Note 'None'
# in place of application token, and no username or password:
client = Socrata("data.calgary.ca", None)

# Example authenticated client (needed for non-public datasets):
# client = Socrata(data.calgary.ca,
#                  MyAppToken,
#                  userame="user@example.com",
#                  password="AFakePassword")

# First 2000 results, returned as JSON from API / converted to Python list of
# dictionaries by sodapy.
results = client.get("j9ps-fyst", limit=2000)

# Convert to pandas DataFrame
df = pd.DataFrame.from_records(results)
```

3 Methodology

The next step was to plot the neighborhoods on a map using folium.



3 Methodology

3.2 Use the Foursquare API to obtain information for Aspen Woods neighborhood

Define the Foursquare credentials and version

```
CLIENT_ID = '' # your Foursquare ID
CLIENT_SECRET = '' # your Foursquare Secret
VERSION = '20210331' # Foursquare API version
LIMIT = 200 # A default Foursquare API limit value
radius = 1500
print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

Prepare a get request for the top 200 venues withing a radius of 1500m

```
LIMIT = 200 # limit of number of venues returned by Foursquare API
radius = 1500 # define radius

# create URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
```

3 Methodology

Obtain the json file, cleanse and structure into a dataframe

```
results = requests.get(url).json()
results
```


```
{'meta': {'code': 200, 'requestId': '6080b22c2ce7be68aa95fa02'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Aspen Woods',
  'headerFullLocation': 'Aspen Woods, Calgary',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 20,
  'suggestedBounds': {'ne': {'lat': 51.05862922982107,
    'lng': -114.18646817946073},
   'sw': {'lat': 51.031629202821044, 'lng': -114.22933324826575}},
```


	name	categories	lat	lng
0	A Ladybug and Cafe	Coffee Shop	51.041340	-114.212507
1	Denim and Smith Barbershops-Aspen	Salon / Barbershop	51.039860	-114.208736
2	Blush Lane Organic Market	Grocery Store	51.041304	-114.213076
3	Diner Deluxe Aspen	Restaurant	51.039636	-114.209193
4	Original Joe's Restaurant & Bar	Restaurant	51.039232	-114.208116

3 Methodology

3.3 Use the Foursquare API to obtain information for all Neighborhoods

Create a function to repeat the Aspen Woods process for all neighborhoods and create a dataframe

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	YORKVILLE	50.870403	-114.076648	Kildares Ale House	50.877934	-114.071518	Bar
1	YORKVILLE	50.870403	-114.076648	Sobeys - Silverado	50.879006	-114.072621	Grocery Store
2	YORKVILLE	50.870403	-114.076648	Dairy Queen / Orange Julius	50.878756	-114.072321	Ice Cream Shop
3	YORKVILLE	50.870403	-114.076648	Starbucks Silverado	50.877897	-114.072121	Coffee Shop
4	YORKVILLE	50.870403	-114.076648	Beauty Boutique by Shoppers Drug Mart	50.878755	-114.072320	Cosmetics Shop

3 Methodology

3.4 Analyze Each Neighborhood

Use One Hot One hot encoding was used to convert categorical data to numerical values so the data could be processed.

	Neighborhood	American Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Dim Sum Restaurant	Diner	Eastern European Restaurant	Empanada Restaurant	Res
6	YORKVILLE	0	0	1	0	0	0	0	0	0	0	
14	WEST SPRINGS	0	0	0	0	0	0	0	0	0	0	
18	WEST SPRINGS	0	0	0	0	0	0	0	0	0	0	
19	WEST SPRINGS	0	0	0	0	0	0	0	0	0	0	

Next, the data was grouped by neighborhood and the mean of the frequency of occurrence of each restaurant category was calculated.

	Neighborhood	American Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Dim Sum Restaurant	Diner	Eastern European Restaurant	Empanada Restaurant	
0	ABBEYDALE	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1	ACADIA	0.142857	0.000000	0.000000	0.000000	0.000000	0.047619	0.000000	0.000000	0.000000	0.000000	
2	ALBERT PARK/RADISSON HEIGHTS	0.000000	0.000000	0.181818	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
3	ALTADORE	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

3 Methodology

Next, we created a dataframe, neighborhoods_venue_sorted, and displayed the top 10 restaurants categories in each neighborhood.

	Neighborhood	American Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Dim Sum Restaurant	Diner	Eastern European Restaurant	Empanada Restaurant	Res
6	YORKVILLE	0	0	1	0	0	0	0	0	0	0	
14	WEST SPRINGS	0	0	0	0	0	0	0	0	0	0	
18	WEST SPRINGS	0	0	0	0	0	0	0	0	0	0	
19	WEST SPRINGS	0	0	0	0	0	0	0	0	0	0	

3 Methodology

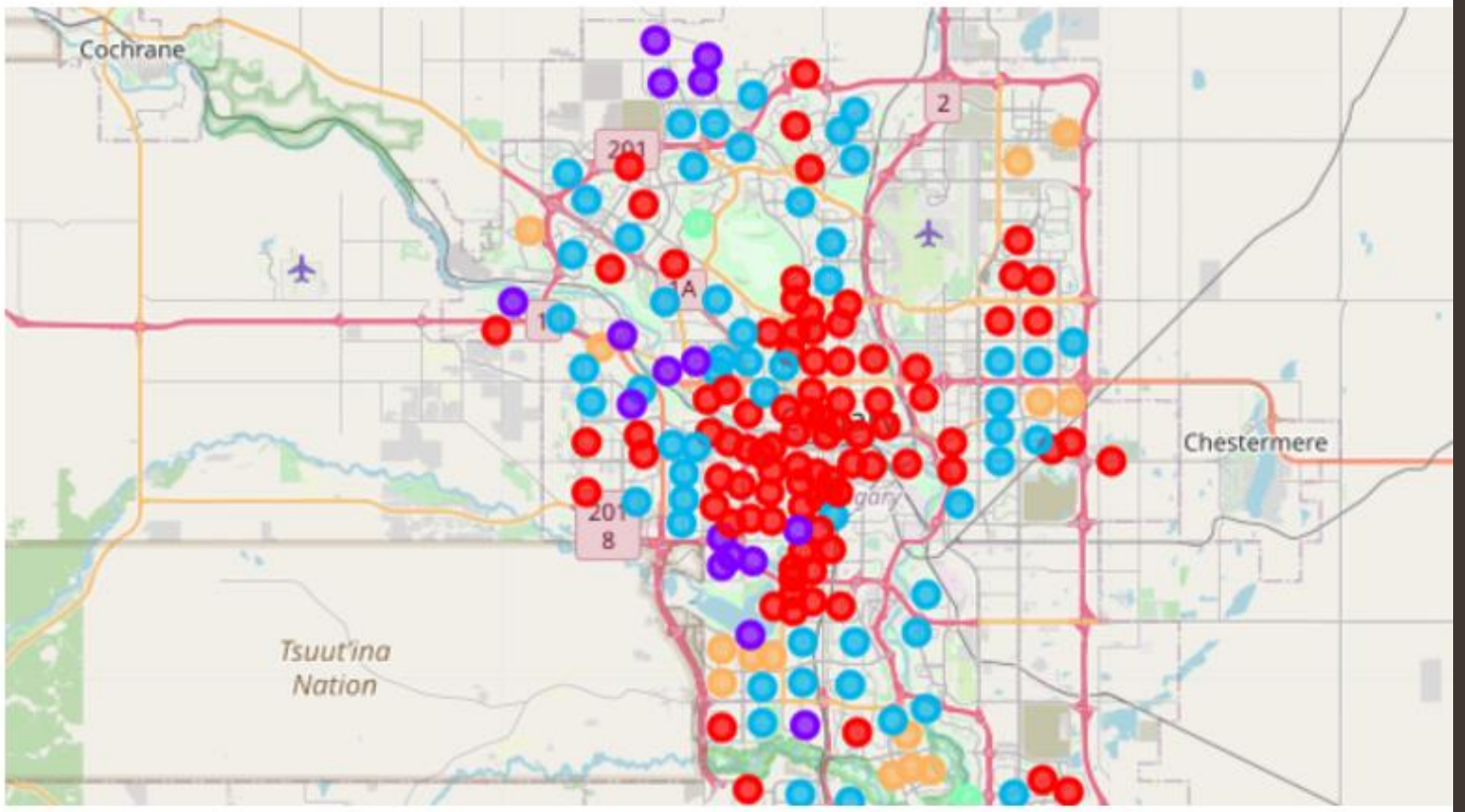
3.5 Cluster Neighborhoods

K-means clustering, a simple and popular unsupervised machine learning algorithm, was used to cluster the neighborhoods into 5 clusters. Each cluster represents a collection of data points that have been aggregated together because of certain

	Neighborhood	Longitude	Latitude	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	7 C
0	YORKVILLE	-114.076648	50.870403	3	Asian Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Indonesian Restaurant	Indian Restaurant	Hong Kong Restaurant	Re
2	WEST SPRINGS	-114.206168	51.059732	1	Restaurant	Fast Food Restaurant	Italian Restaurant	Tapas Restaurant	Sushi Restaurant	Middle Eastern Restaurant	Re
3	WOODLANDS	-114.106339	50.942876	1	Sushi Restaurant	Seafood Restaurant	Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Falafel Restaurant	Re

3 Methodology

The resulting clusters were plotted on a map using folium.

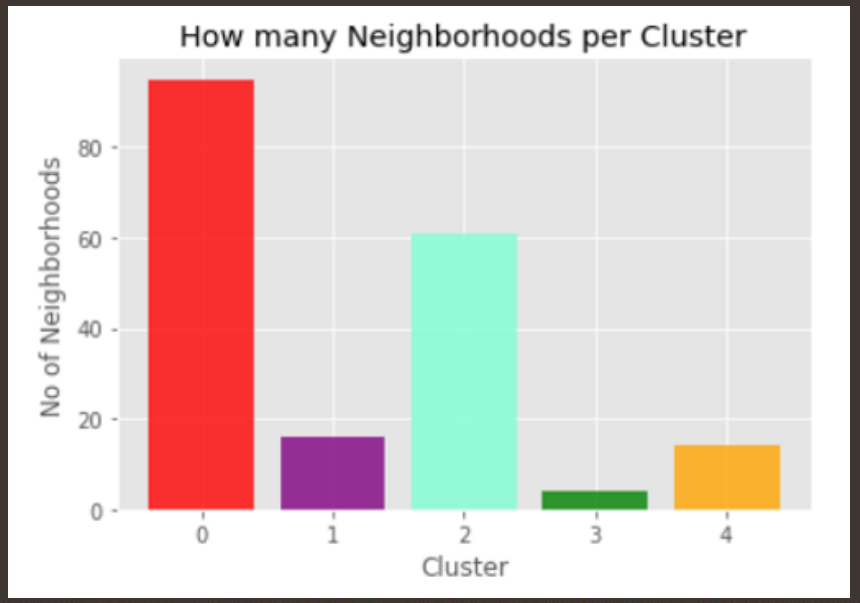


3 Methodology

3.6 Examine the Resulting Clusters

Cluster 0 contained the highest number of neighborhoods followed by cluster 2.

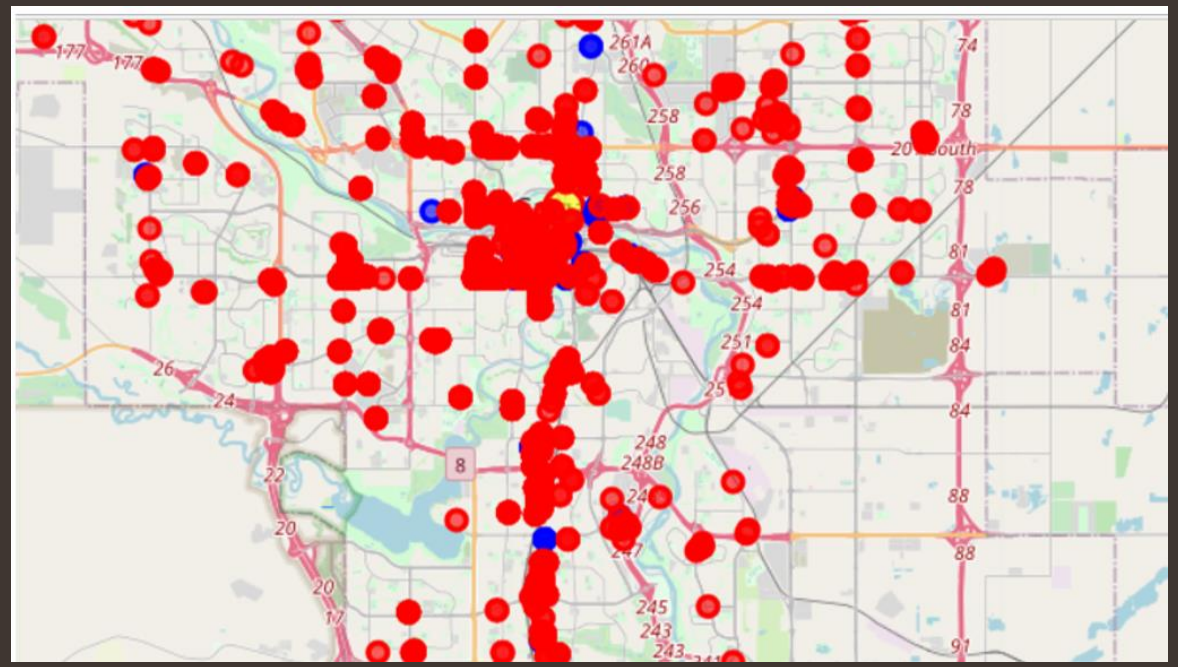
Cluster 0 contains the majority of our candidate neighborhoods and this cluster appe ars to be comprised of mostly oriental restaurants. Cluster 2 contains the remainder o f the candidate neighborhoods and appears to contain mostly Fast Food restaurants.



3 Methodology

3.7 Examine Density of Italian vs Non-Italian Restaurants

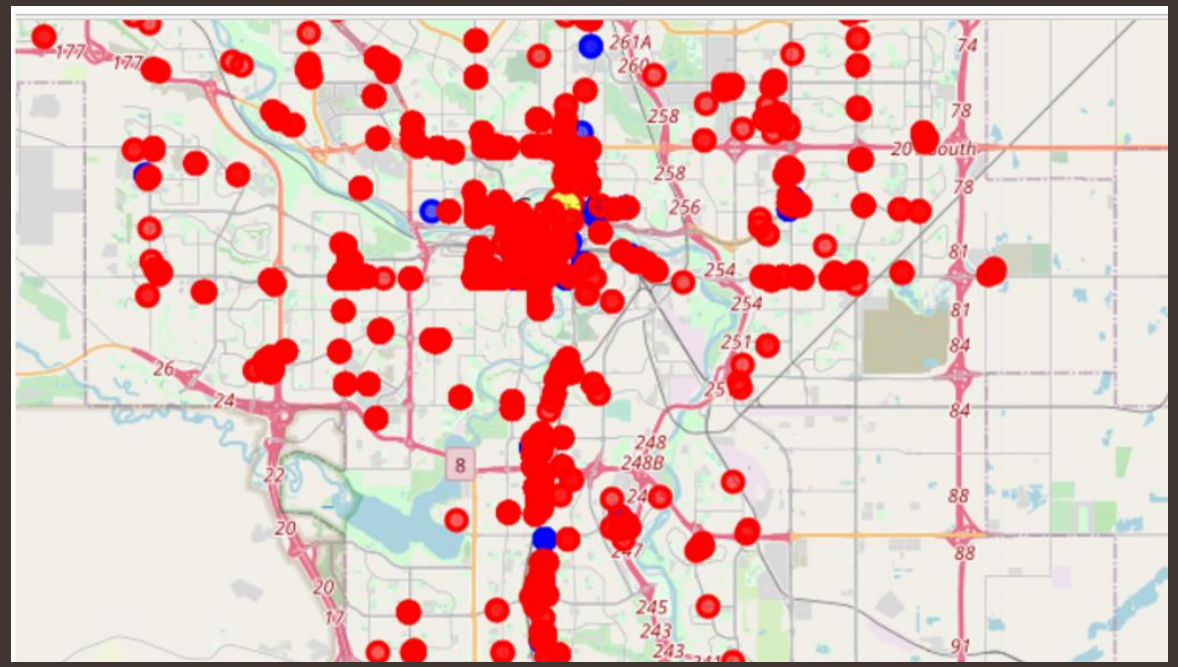
Folium was used plot the Italian (blue) and non-Italian (red) restaurants on a map.



3 Methodology

3.8 Key Demographics of Candidate Neighborhoods

Folium was used plot the Italian (blue) and non-Italian (red) restaurants on a map.



4 Results and Discussion

In Calgary, Italian Restaurant makeup just over 6% of the total number of restaurants in the city. The highest density of Italian restaurants is in the neighborhoods surround ing the downtown core. As you go farther into the bedroom communities the density drops significantly.

The location preference for a new Italian restaurant was upcoming neighborhoods in the south or west sides of the city. These neighborhoods are still in a developing state so the demographics will change over time. These neighborhoods also have very low restaurants density in comparison to other the neighborhoods so competition will low. However, finding an available location for the restaurant may be a more daunting task as very little land in the neighborhoods is zoned for commercial use.

The candidate neighborhoods also have very low restaurants density in comparison to other the neighborhoods so competition should be low. The only neighborhood that had an Italian restaurant was West Springs. However, finding an available location for the restaurant may be a more daunting task as very little land in the neighborhoods is zoned for commercial use.

5 Conclusion

The purpose of this project was to identify potential neighborhoods in Calgary for a n ew Italian restaurant in order to assist our client to make their decision. Foursquare w as the primary source for the restaurant data while neighborhood demographics and geographical information was sourced from the City of Calgary.

The final decision on an optimal location for a new Italian restaurant will be made by the client based on a number of factors such as, future growth rates, availability of sp ace, accessibility, parking, etc.