Introduction

1.1 Business Problem/Objective

The objective of this project is to find an optimal location for our client to open a new **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 south 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.

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.

Methodology

3.1 Obtain coordinates for all neighborhoods

The coordinates for each neighborhood were obtained using the following code and the 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)
```

Once the data was in a pandas dataframe the unneeded columns were dropped and column names were changed. The first five rows of the resulting datasets are as follows:

	Neighborhood	Longitude	Latitude
0	YORKVILLE	-114.076648	50.870403
1	WOLF WILLOW	-114.008637	50.870724
2	WEST SPRINGS	-114.206168	51.059732
3	WOODLANDS	-114.106339	50.942876
4	WINDSOR PARK	-114.083550	51.005040

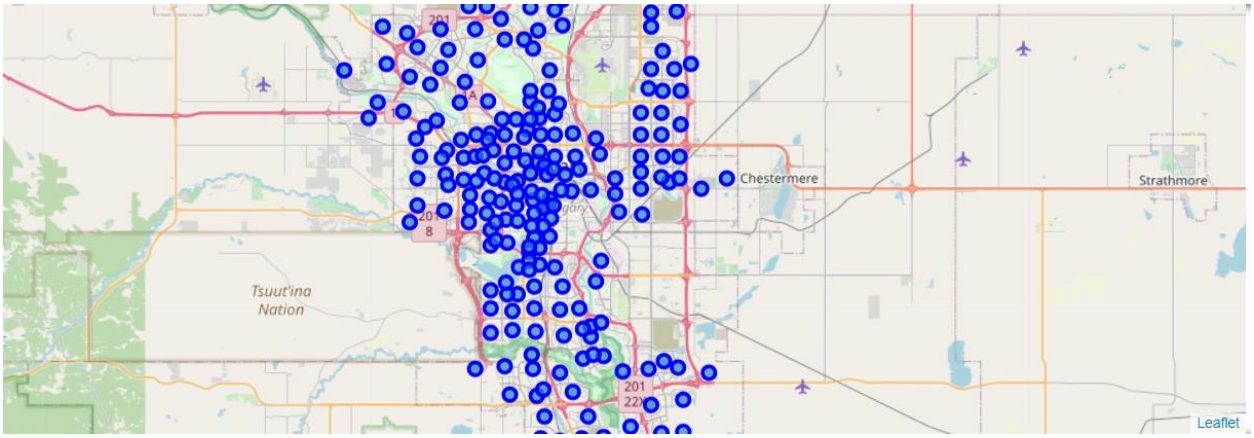
The geographic coordinates of Calgary were obtained using geolocator.

```
address = 'Calgary, AB'

geolocator = Nominatim(user_agent="yyc_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinates of Calgary are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinates of Calgary are 51.0534234, -114.0625892.
```

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



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)
```

Retrieve the geographical coordinates of Aspen Woods

```
neighborhood_latitude = neighborhoods.loc[254, 'Latitude'] # neighborhood latitude value
neighborhood_longitude = neighborhoods.loc[254, 'Longitude'] # neighborhood longitude value

neighborhood_name = neighborhoods.loc[254, 'Neighborhood'] # neighborhood name

print('Latitude and longitude values of {} are {}, {}.'.format(neighborhood_name,
                                                               neighborhood_latitude,
                                                               neighborhood_longitude))
```

```
Latitude and longitude values of ASPEN WOODS are 51.04512921632106, -114.20790071386324.
```

Prepare the get request to retrieve the top 200 venues in Aspen Woods within a radius of 1500 meters

```
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)
url # display URL
```

A snippet of the resulting json file is shown below.

```
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}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4b0586e9f964a520937422e3',
       'name': 'A Ladybug and Cafe',
       'location': {'address': '2132 - 10 Aspen Stone Blvd. SW',
```


The cleaned json data was structured into a pandas dataframe.

	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

3.3 Use the Foursquare API to obtain information for all Neighborhoods

A function was created to repeat the above process for all the neighborhoods in Calgary and create a dataframe called calgary_venues.

	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

Since the dataframe contained a number of venues that were not restaurants (i.e. bars, grocery stores, etc.) the Venue Category was used to filter out those that did not have a venue category containing the words ‘Rest”, ‘Diner’ and ‘Steakhouse’ and create a new restaurants dataframe.

The restaurants dataframe was grouped by neighborhood and a total count of restaurants in each neighborhood was obtained.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
ABBEYDALE	2	2	2	2	2	2
ACADIA	25	25	25	25	25	25
ALBERT PARK/RADISSON HEIGHTS	13	13	13	13	13	13

3.4 Analyze each Neighborhood

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.

```
calgary_grouped = calgary_onehot.groupby('Neighborhood').mean().reset_index()
calgary_grouped
```

	Neighborhood	American Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Dim Sum Restaurant	Diner	Eastern European Restaurant	Empanada Restaurant	I
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	

We then printed the top 5 restaurant categories in each neighborhood.

```
num_top_venues = 5

for hood in calgary_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = calgary_grouped[calgary_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----ABBEYDALE----
                       venue  freq
0       Fast Food Restaurant   1.0
1    Scandinavian Restaurant   0.0
2   Mediterranean Restaurant   0.0
3         Mexican Restaurant   0.0
4  Middle Eastern Restaurant   0.0
```

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

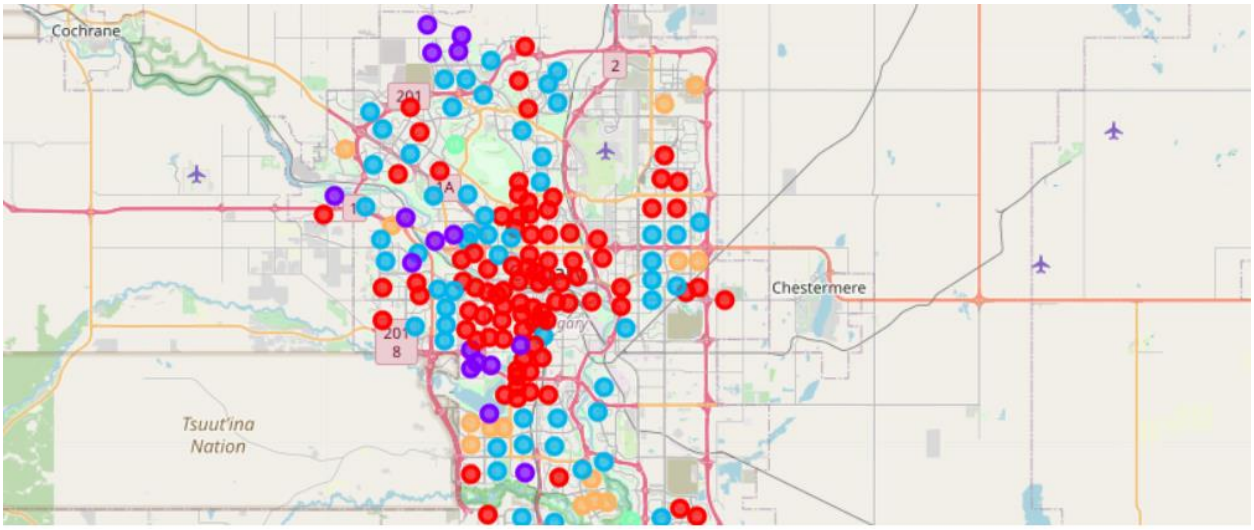
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ABBEYDALE	Fast Food Restaurant	Vietnamese Restaurant	Japanese Restaurant	Indonesian Restaurant	Indian Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant
1	ACADIA	Fast Food Restaurant	Vietnamese Restaurant	Mexican Restaurant	American Restaurant	Italian Restaurant	Restaurant	Chinese Restaurant	Greek Restaurant	Japanese Restaurant	Swiss Restaurant
2	ALBERT PARK/RADISSON HEIGHTS	Indian Restaurant	Asian Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Seafood Restaurant	Falafel Restaurant	Italian Restaurant	Restaurant	Diner	Dim Sum Restaurant

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 similarities. The Cluster Label was then added to the dataframe.

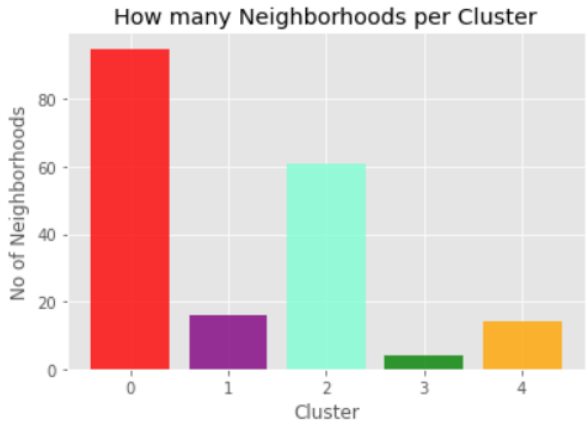
	Neighborhood	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7 C
0	YORKVILLE	-114.076648	50.870403	3	Asian Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Indonesian Restaurant	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

The resulting clusters were plotted on a map using folium.



3.6 Examine the Resulting Clusters

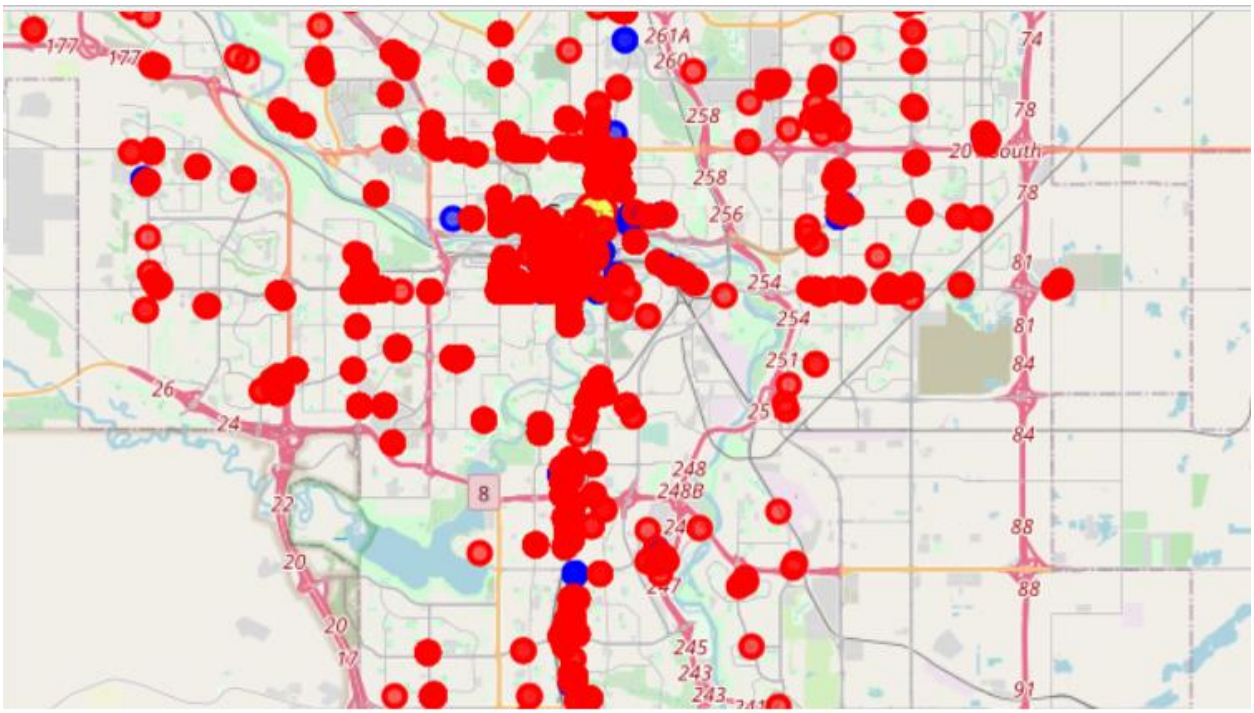
Cluster 0 contained the highest number of neighborhoods followed by cluster 2.



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

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.



Italian restaurants make up just over 6% of the total restaurants in the city. The density of restaurants, in particular Italian restaurants, is highest in the neighborhoods surrounding the downtown core. The farther one goes from the city center the fewer restaurants can be found. The new und upcoming

neighborhoods of the south and west sections of the city exhibit extremely low restaurant density. Only one of the candidate neighborhoods contained an Italian Restaurant.

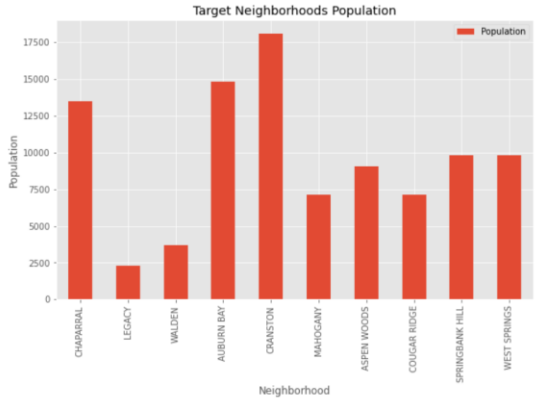
3.8 Key Demographics of Candidate Neighborhoods

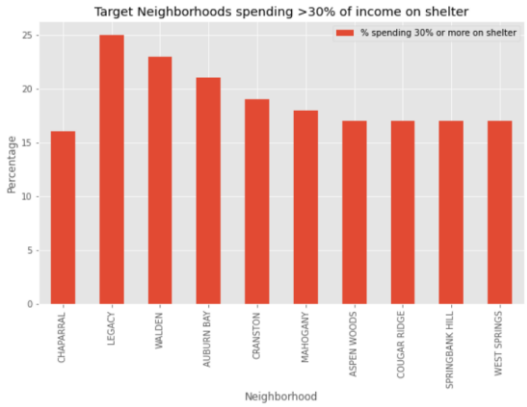
The Preference was to find a location in the new and upcoming neighborhoods of the west and south parts of the city. The information in Table 1 was obtained from the most recent City of Calgary census.

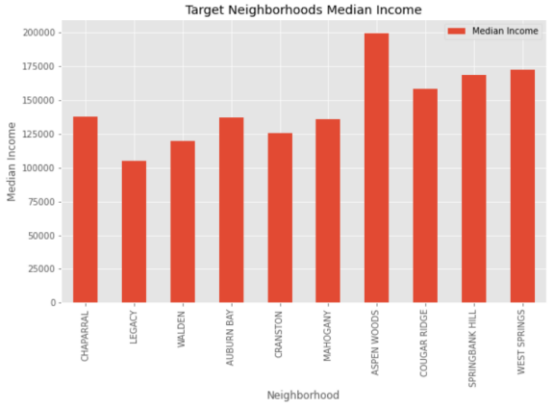
Table 1 – Major Developing Neighborhoods

Neighborhood	Area	Population	Median Income	% spending >30% of income on shelter
Chaparral	South	13,475	137,640	16
Legacy	South	2,315	104,855	25
Walden	South	3,675	119,595	23
Auburn Bay	South	14,850	136,961	21
Cranston	South	18,120	125,439	19
Mahogany	South	7,125	135,818	18
Aspen Woods	West	9,060	199,759	17
Cougar Ridge	West	7,110	158,306	17
Springbank Hill	West	9,835	168,672	17
West Springs	West	9,820	172,358	17

Matplot was then used to visualize the census data in bar charts.







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 surrounding 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.

Conclusion

The purpose of this project was to identify potential neighborhoods in Calgary for a new Italian restaurant in order to assist our client to make their decision. Foursquare was 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 space, accessibility, parking, etc.