GitHub Link
https://gist.github.com/06aa5b21b5e904776df3df33eee8fe59

Finding place for office in Toronto, Canada

Table of Contents

Introduction

The aim of the project is to formulate a methodology for choosing the best places for an office location based on Foursquera data depending on the specifics of the client’s business. Business profile defines the infrastructure requirements that should be located in the vicinity of his office. Internet data on the location of social and technological infrastructure allows to optimize the search for the office that best meets the needs of the company.

This project is being implemented as Capstone project by the IBM Data Science Professional Certificate program, which involves "to leverage the

Foursquare location data to explore or compare neighbourhoods or cities" or to solve other problems with that data.

1. Business Problem

Choosing the right place for the office is an important task when creating and expanding and modifying any business. At the same time, companies are guided both by the rental price or the cost of commercial real estate, and by the infrastructure that will be necessary to perform the functions of this organization. Internet data on rental prices, property values, and the placement of infrastructure can help in choosing the best office location. In particular, data from the Foursquare service will be useful for making such decisions.

We plan to create a methodology for choosing the best place for the office, taking into account the criteria

- the cost of renting commercial real estate,
- quality ratings of the proposed premises,
- requirements for infrastructure facilities, based on the use of Foursquera data and other online resources.

For example, we will use data from Toronto (Canada), in which the company organizing international conferences and forums plans to open its new office. To do business, this company needs:

- low rental prices of commercial real estate;

- the presence of a first-class business or conference centre near the office;
- availability of high-class hotels and restaurants near the office centre;
- proximity of tourist and cultural sites, as well as sports facilities for organizing the rest of conference participants
- with banks and office services.

This methodology can be used

- realtors and brokers engaged in operations with commercial real estate,
- tenants choosing a location for offices,
- potential landlords who form commercial offers for the market.

2. Data Requirements

To analyze the potential locations for the office of the company organizing international meetings and conferences, we need information:

1. About the conference and business centres of Toronto, in which it is possible to rent premises to organize these events.
2. About ratings and rental prices of these areas.
3. About tourist and cultural infrastructure facilities, as well as service facilities located around potential conference venues. We have to assess information whether there are any
4. Office premises for rent,

5. Hotels, restaurants and objects necessary for the organization of rest and leisure of participants of conferences and forums near conference and business centres.

We chose 500 m as the length of the radius within which all the necessary objects should be concentrated.

3. Data Collection

To begin with, we will take a look at Toronto on the Map using the folium library.

We will also fetch the data from three different resources:

- **Regus.ca**: contains information about the office spaces of Toronto, their ratings and rental prices. For our example, we can get information about all rent and purchase options for office space available in Toronto with addresses and coordinates of their location.
- **EventSource.ca**: contains information about the infrastructure objects of Toronto, their ratings and rental prices. For our example, we can get information about the Conference and Business Centres of Toronto, their ratings and rental prices of space on this resource addresses and coordinates of their location.
- **Foursquare**: We will use the Foursquare API to fetch venues in Toronto. Foursquere contains data on hotels and restaurants, cultural and tourist sites, sports facilities in Toronto, as well as the addresses and coordinates of their location. We will search for infrastructure data located 500 meters around the Toronto office spaces.

3.1. Getting data about office space rent in Toronto on Regus.ca

Site **Regus.ca** contains information about the office spaces of Toronto. We implemented this resource parsing to get data about:

- title office space;
- its site link;
- its type;
- its location and post address;
- its min and max rent prices;
- its rating.

A total of **53 records of office space in Toronto** were received.

Then we removed from the DataFrame the lines relative to the office spaces that are located outside of Toronto.

As a result, there are **36 records of office spaces**, each of which was continued to collect information.

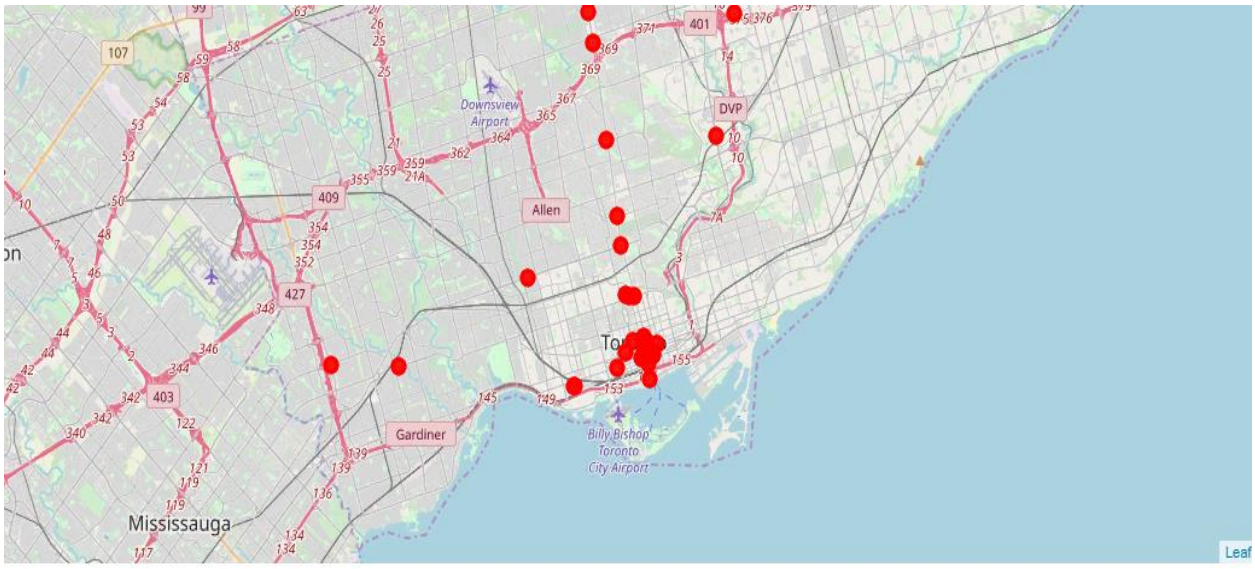


Figure 1. Toronto Office spaces on the map

3.2. Search data about Conference Centres in Toronto on EventSource.ca

Considered in this case, the company organizes international conferences, symposia and seminars. For its business it is extremely important that a high-class conference centres with reasonable rental prices be located near the office. We collected data about Toronto conference centres from a resource www.EventSource.ca using parsing. On this site, data were obtained regarding

- the name and type of the centre,
- its location, including geographic coordinates,
- rental prices
- and ratings, built on the basis of fixing customer ratings using the services offered by the centres.

Initially, data was obtained from **12 Toronto conference centres**.

As for Office Spaces, we removed from the DataFrame the lines relative to the centres that are located outside of Toronto and displayed the resulting objects set on the map with office space objects. As a result, there are **7 conference centres in Toronto**.

Table 1. List conference centres in Toronto

	Name	Link	Type	Postcode	Location	Adress	Latitude	Longitude	Price_range	Rating
0	Twenty Toronto Street	/twenty-toronto-street	Conference Centre	M5C 2B8	Toronto	20 Toronto St	43.650213	-79.376626	Low	0.00
1	Markham Convention Centre	/markham-convention-centre	Convention Centre		Toronto	2901 Markham Rd	43.825604	-79.247333	Low	4.83
2	Centennial	/centennial-	Special Event	M1G 3T8	Toronto	937 Progres	43.784623	-79.22974	Low	4.96

	Name	Link	Type	Postcode	Location	Adress	Latitude	Longitude	Price_range	Rating
	College Event Centre	college-event-centre	Venues			s Ave		6		
3	Vantage Venues	/vantage-venues	Special Event Venues	M5H 1J9	Toronto	150 King St. W, 27th Floor	43.648100	-79.384452	Low	4.96
4	The Bram & Bluma Appel Salon	/the-bram-and-bluma-appel-salon	Special Event Venues	M4W 2G8	Toronto	789 Yonge St 2nd floor	43.672129	-79.386411	high	5.00
5	Beanfield Centre	/beanfield-centre	Conference Centre	M6K 3C3	Toronto	105 Princes' Blvd	43.634178	-79.410696	middle	4.50
6	Oakham House	/oakham-house	Conference Centre	M5B 1E9	Toronto	63 Gould St	43.658005	-79.378062	high	5.00

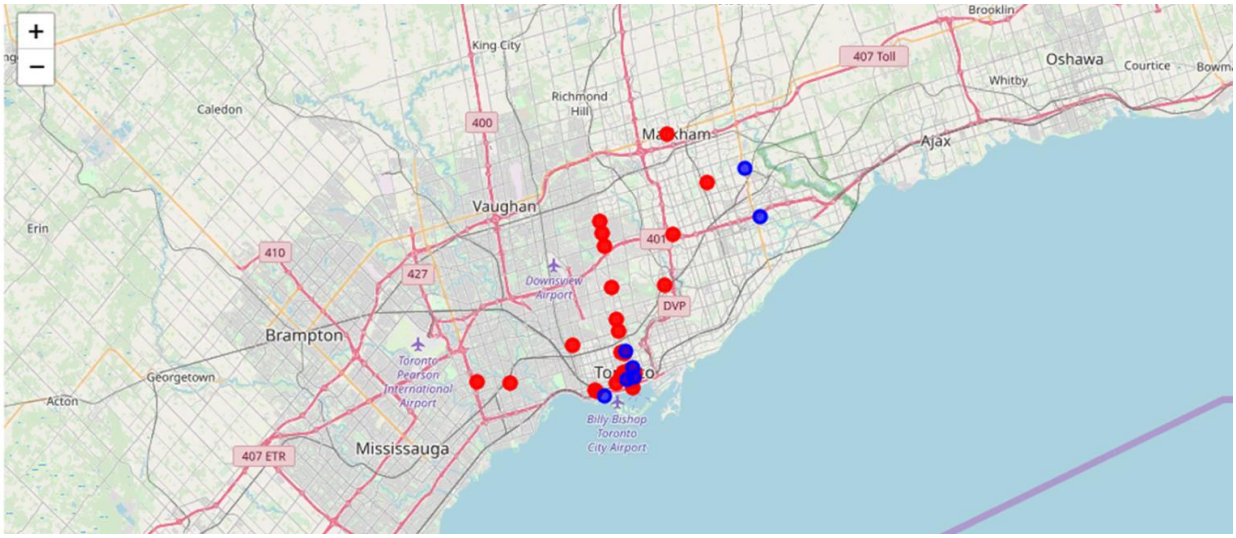


Figure 2. Toronto Conference Centres on the map

Distance to the nearest conference centre and rating of the nearest conference centre for office spaces

One of the most important criteria for choosing an office is the proximity of the conference centre. Therefore, we calculated the distance to the nearest conference centre for each office space from our list, using special function.

After searching nearest CC and adding its features in DataFrame

The data on the nearest conference centre, as well as the quality characteristics and prices of the services provided, were added to the DataFrame containing the list of office premises.

3.3. Getting data about infrastructure around Office spaces in Toronto on Foursquare

The second component of the office space quality criterion is a characteristic of the surrounding infrastructure. Near the office and conference centre where events will be held, there should be

- hotels,
- cafe and restaurants,
- sports facilities (including preferably a pool and spa),
- a bank,
- tourist facilities and cultural facilities. Data on the availability and location of such objects in Toronto were collected using the service Foursquare.

The top 100 venues that are around OS a radius of 500 meters

Using the Foursquare service, we chose infrastructure objects located within a radius of 500 meters around each office space from our list.

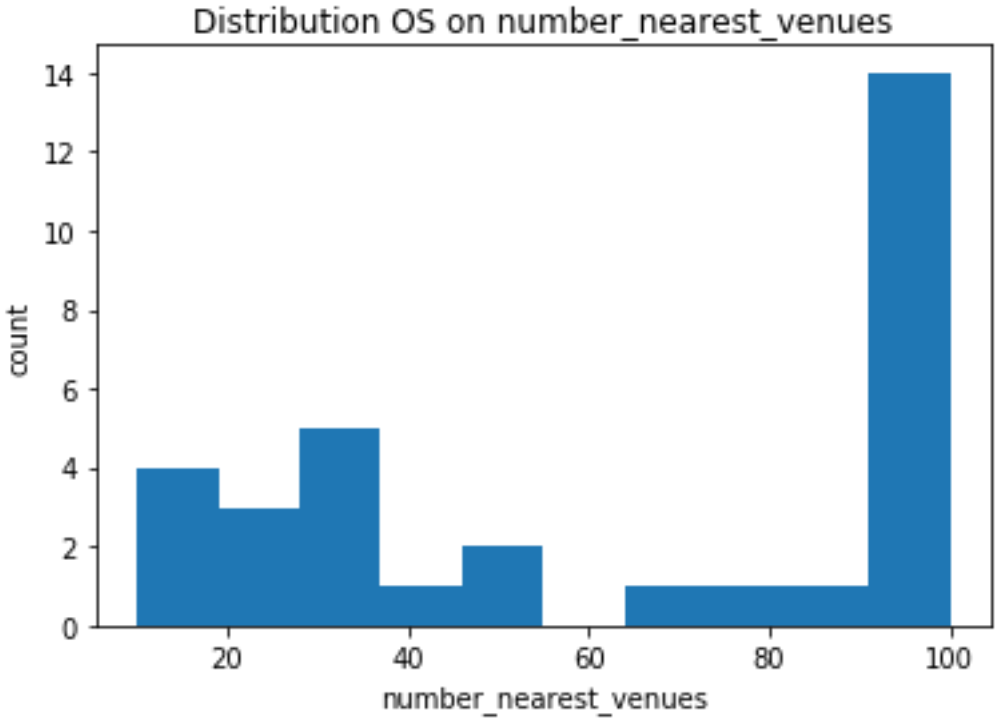


Figure 3. Distribution office spaces on nearest venues number.

4. Data Preparation and Description

4.1.Visualization of data on infrastructure objects around office spaces

Definition of classes of found objects of infrastructure

The service Foursquare offers an extremely broad classification of infrastructure facilities. For the purpose of analysis, we have aggregated the categories of objects indicated on the resource into wider classes.

Next, objects were deleted whose presence is not critical for the choice of office:

```
['Art Studies','Avto and cycle services','Books, arts and supplies shops',
'Business and Home services','Buildings and constructions',           'Cl
othes and shoes shops','Department store','Diners','Food shops - specializ
ed','Food shops - univercal',           'Game shops','Goods for home and
office','Health & Beauty Service','Kids and babies shops','Market',
     'Medicin and Pharmacy','Music and electronics shops','Office and busi
ness centres','Other shopes','Sport shops','Schools and univercity','Other
```

```
goods','Universal Stores',                    'Transport infrustructure','Zoo and
veterinarian']
```

The result was the following distribution of office spaces on venues numbers.

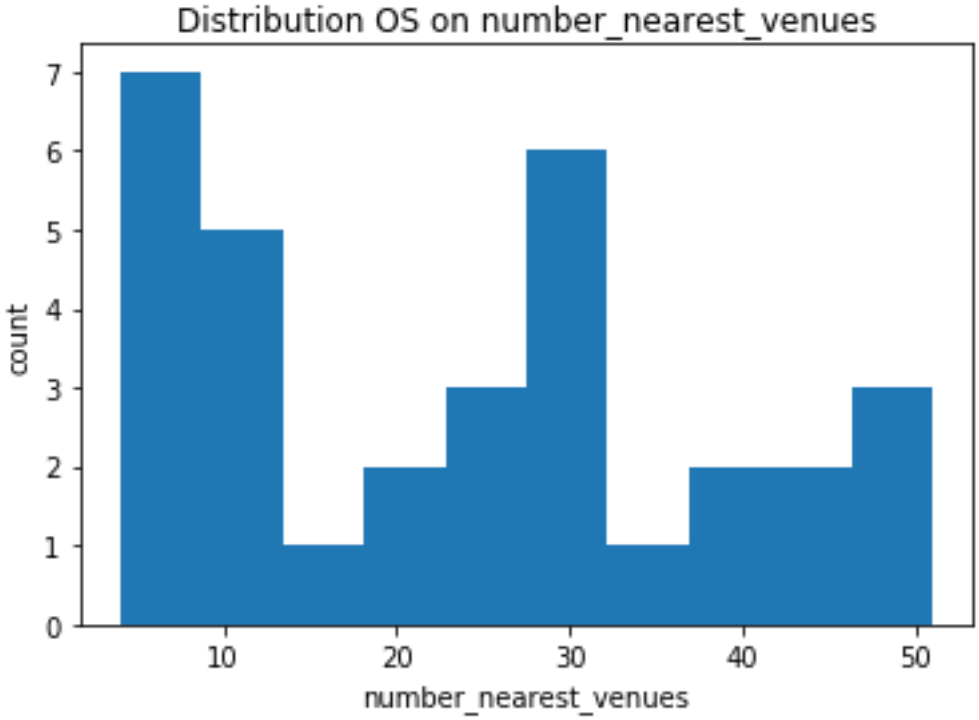
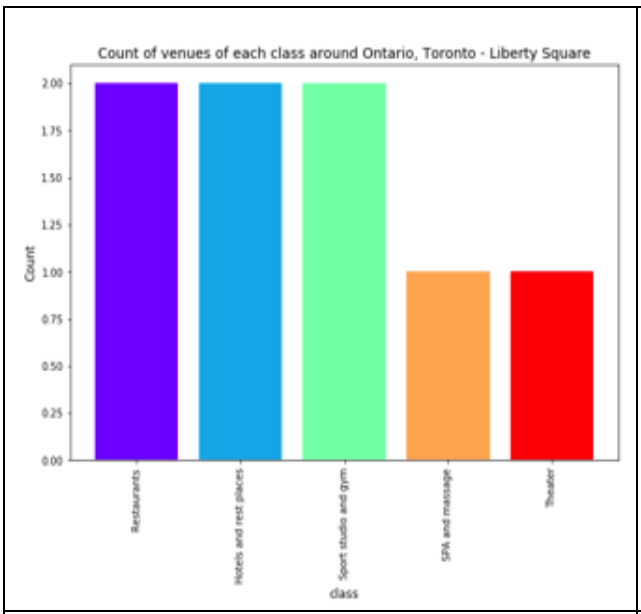


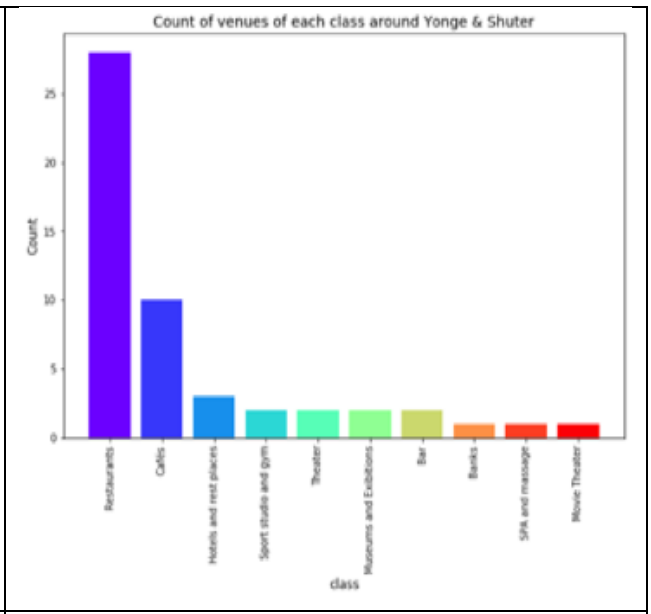
Figure 4. Final distribution office spaces on nearest venues number

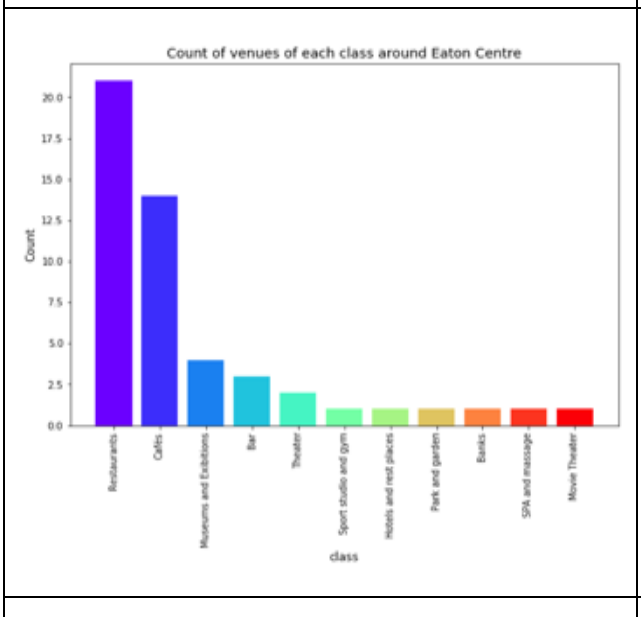
Get distribution plots venues around office spaces

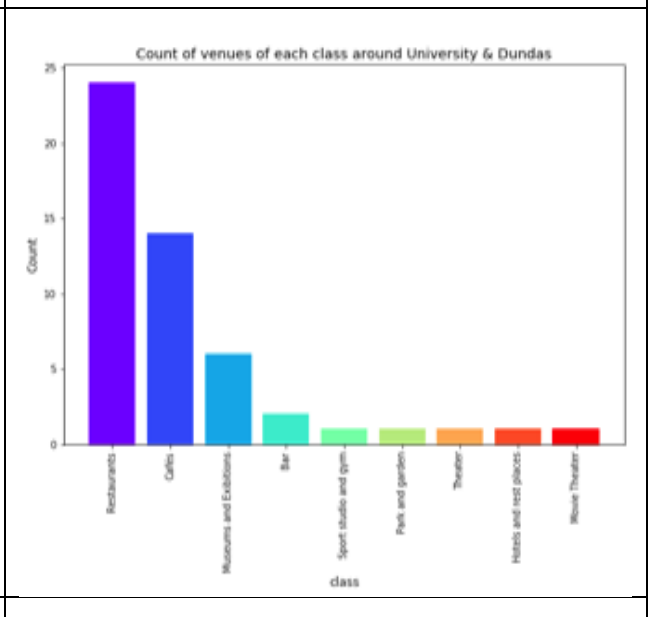
For a better understanding of the characteristics of the infrastructure, for each office space, from our list, we constructed the distribution of infrastructure objects to the allocated classes.

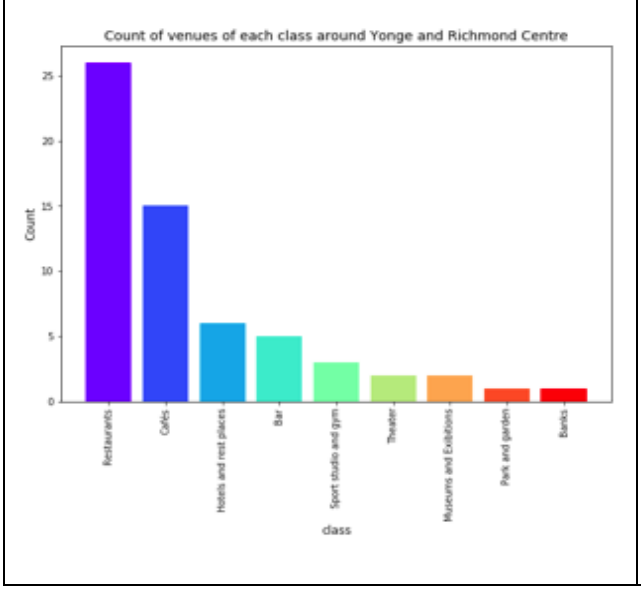
Table 2. Venues around Office Spaces in Toronto

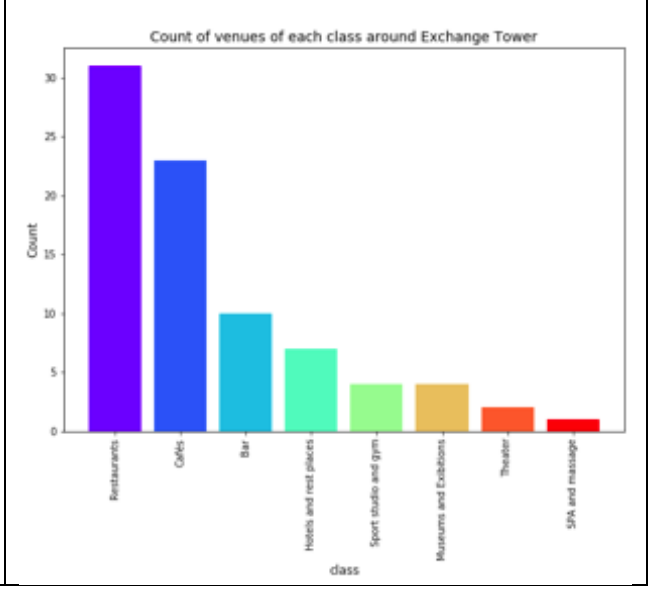

Count of venues of each class around Ontario, Toronto - Liberty Square
Count
class
Restaurants
Hotels and rest places
Sport studio and gym
SPA and massage
Theater

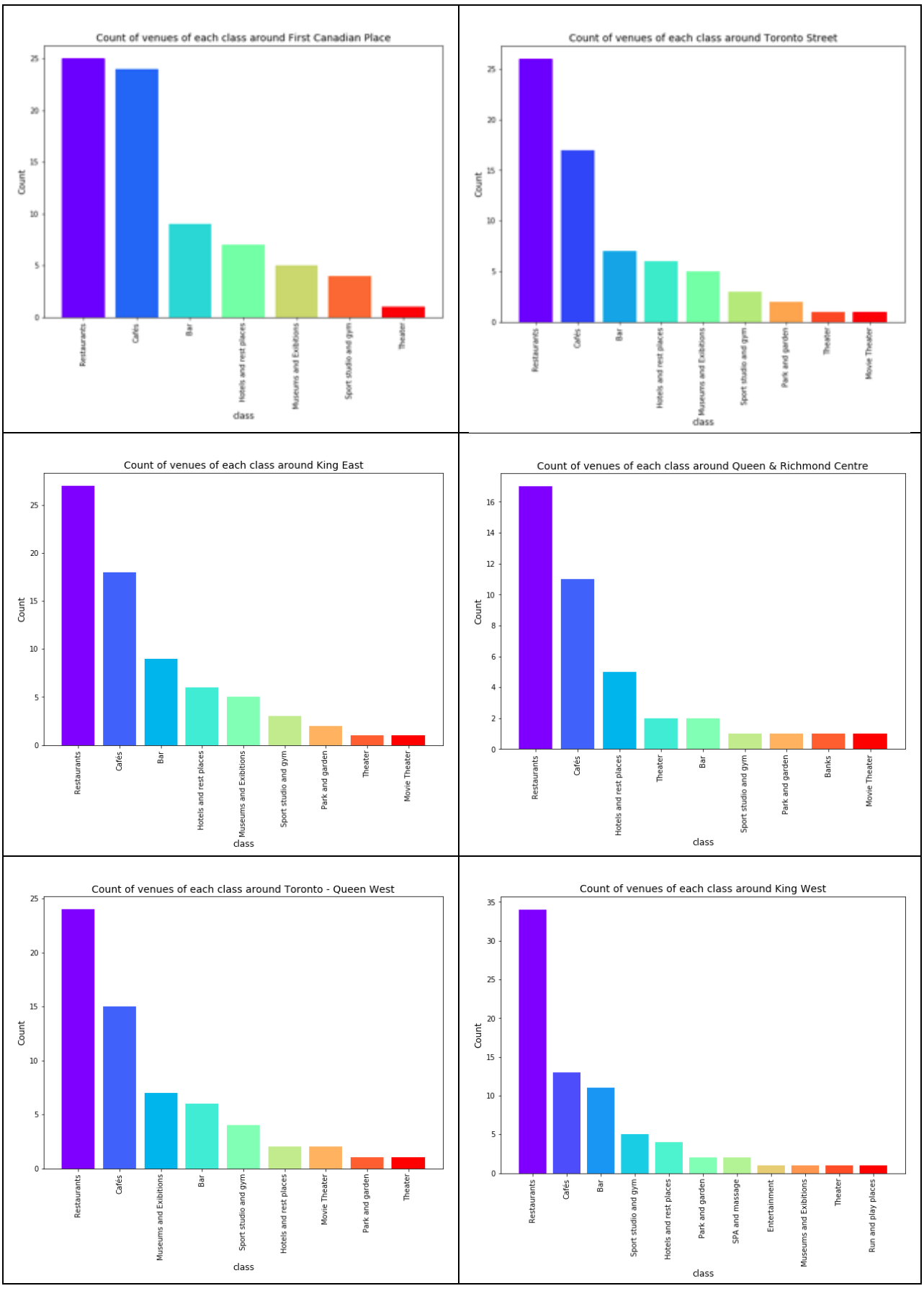

Count of venues of each class around Yonge & Shuter
Count
class
Restaurants
Cafés
Hotels and rest places
Sport studio and gym
Theater
Museums and Exibitions
Bar
Banks
SPA and massage
Movie Theater

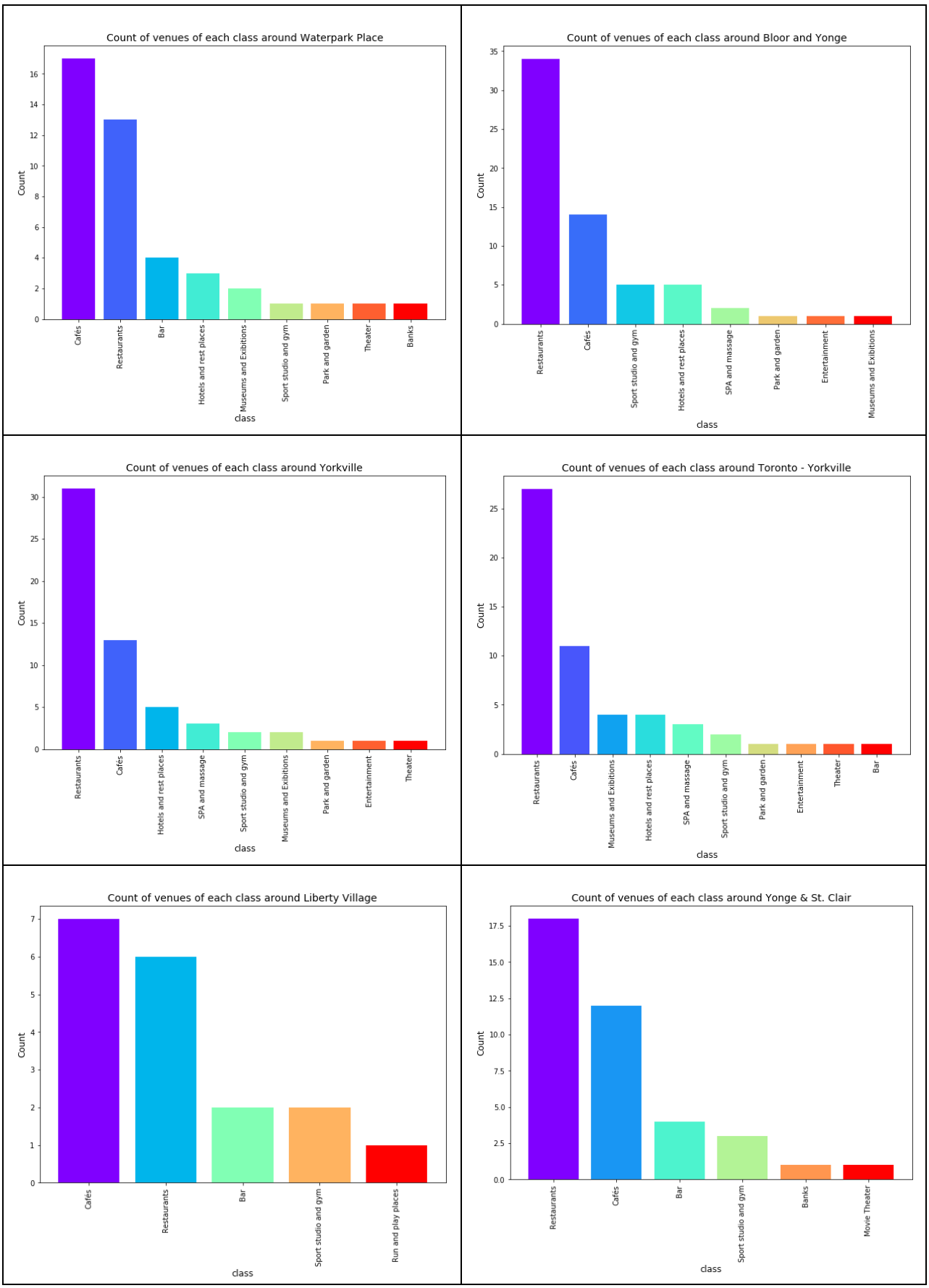

Count of venues of each class around Eaton Centre
Count
class
Restaurants
Cafés
Museums and Exibitions
Bar
Theater
Sport studio and gym
Hotels and rest places
Park and garden
Banks
SPA and massage
Movie Theater

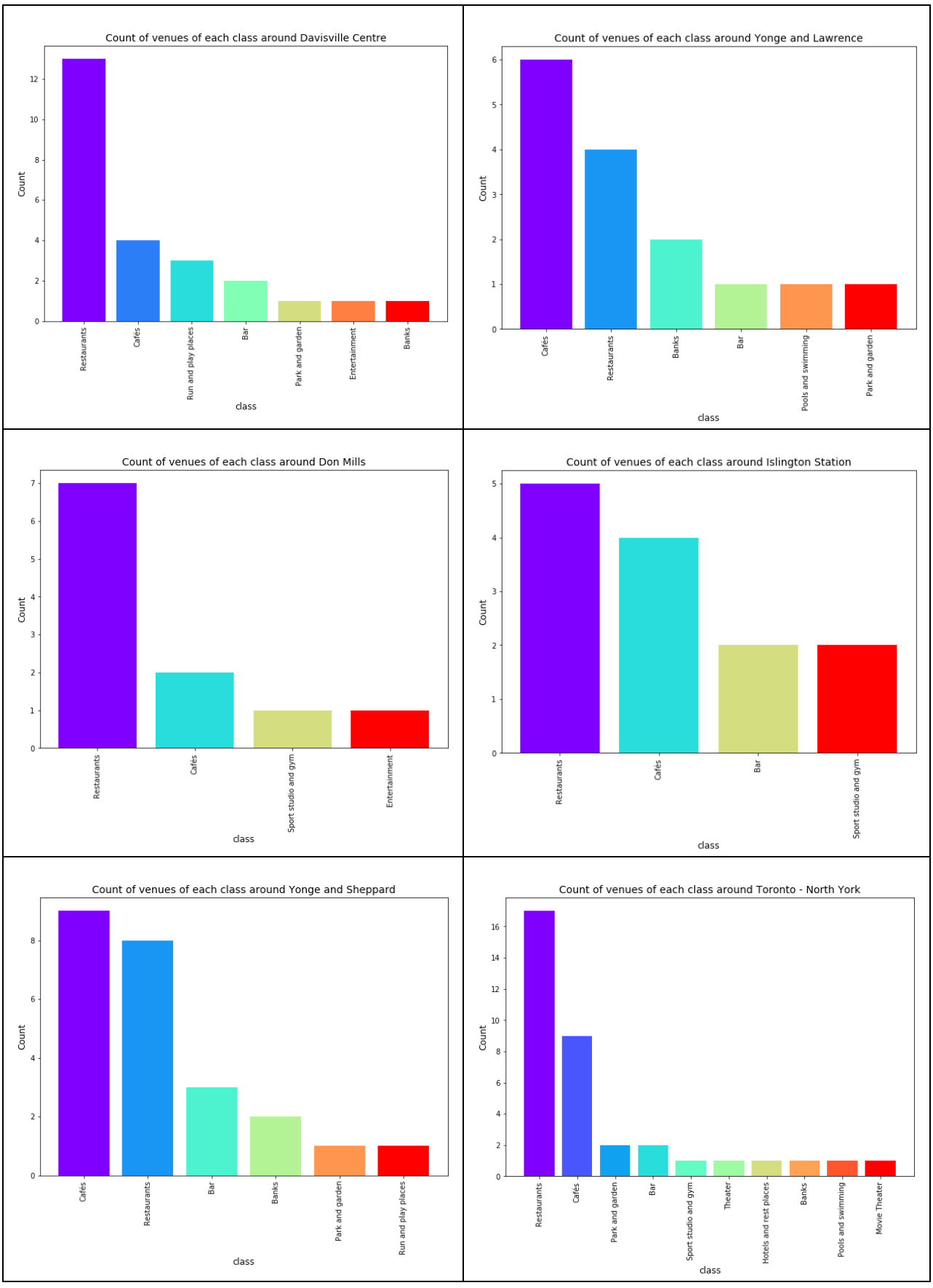

Count of venues of each class around University & Dundas
Count
class
Restaurants
Cafés
Museums and Exibitions
Bar
Sport studio and gym
Park and garden
Theater
Hotels and rest places
Movie Theater

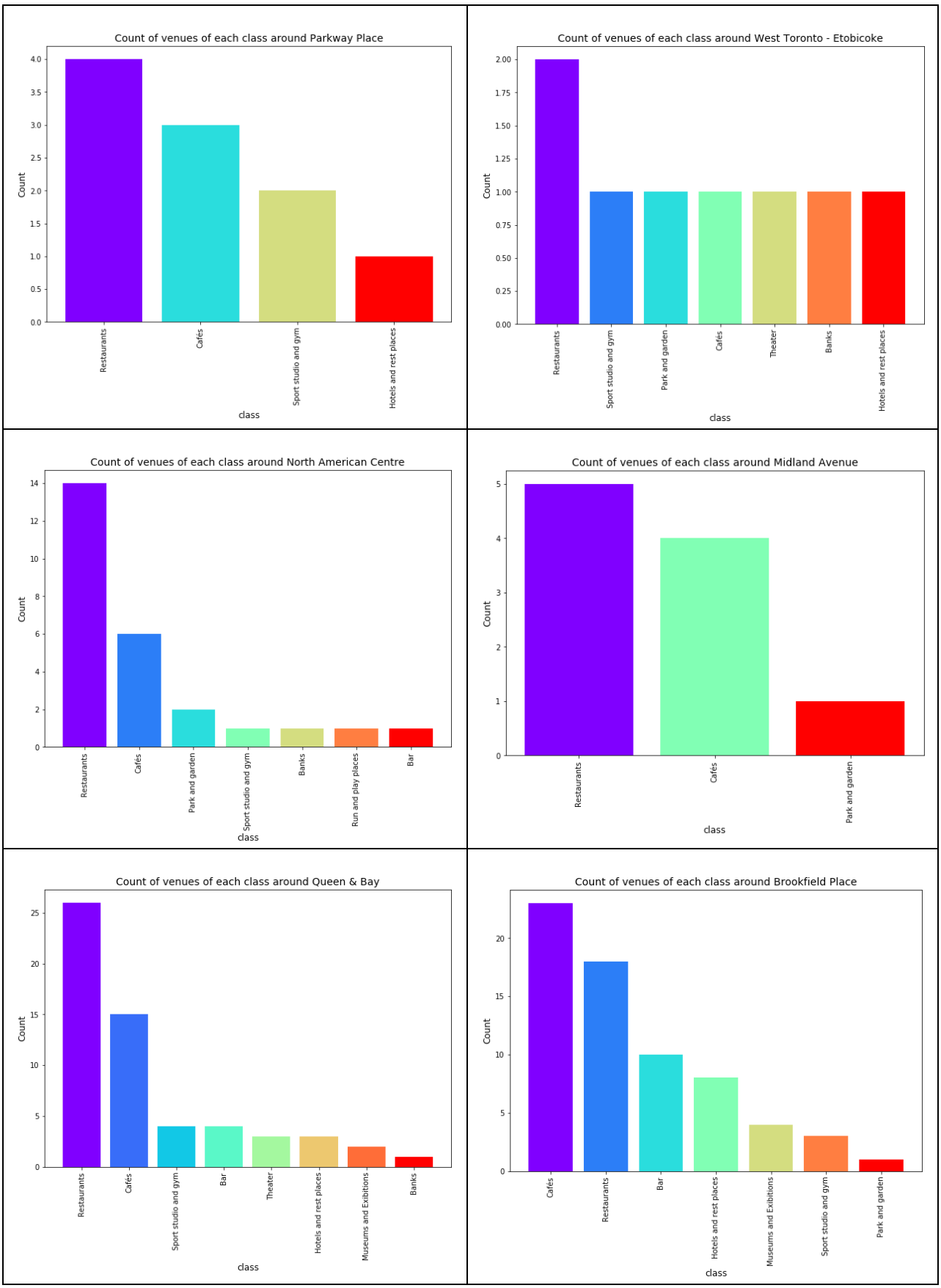

Count of venues of each class around Yonge and Richmond Centre
Count
class
Restaurants
Cafés
Hotels and rest places
Bar
Sport studio and gym
Theater
Museums and Exibitions
Park and garden
Banks


Count of venues of each class around Exchange Tower
Count
class
Restaurants
Cafés
Bar
Hotels and rest places
Sport studio and gym
Museums and Exibitions
Theater
SPA and massage


Count of venues of each class around First Canadian Place
Count
class
Restaurants
Cafés
Bar
Hotels and rest places
Museums and Exibitions
Sport studio and gym
Theater
Count of venues of each class around Toronto Street
Count
class
Restaurants
Cafés
Bar
Hotels and rest places
Museums and Exibitions
Sport studio and gym
Park and garden
Theater
Movie Theater
Count of venues of each class around King East
Count
class
Restaurants
Cafés
Bar
Hotels and rest places
Museums and Exibitions
Sport studio and gym
Park and garden
Theater
Movie Theater
Count of venues of each class around Queen & Richmond Centre
Count
class
Restaurants
Cafés
Hotels and rest places
Theater
Bar
Sport studio and gym
Park and garden
Banks
Movie Theater
Count of venues of each class around Toronto - Queen West
Count
class
Restaurants
Cafés
Museums and Exibitions
Bar
Sport studio and gym
Hotels and rest places
Movie Theater
Park and garden
Theater
Count of venues of each class around King West
Count
class
Restaurants
Cafés
Bar
Sport studio and gym
Hotels and rest places
Park and garden
SPA and massage
Entertainment
Museums and Exibitions
Theater
Run and play places


Count of venues of each class around Waterpark Place
Count
class
Cafés
Restaurants
Bar
Hotels and rest places
Museums and Exibitions
Sport studio and gym
Park and garden
Theater
Banks
Count of venues of each class around Bloor and Yonge
Count
class
Restaurants
Cafés
Sport studio and gym
Hotels and rest places
SPA and massage
Park and garden
Entertainment
Museums and Exibitions
Count of venues of each class around Yorkville
Count
class
Restaurants
Cafés
Hotels and rest places
SPA and massage
Sport studio and gym
Museums and Exibitions
Park and garden
Entertainment
Theater
Count of venues of each class around Toronto - Yorkville
Count
class
Restaurants
Cafés
Museums and Exibitions
Hotels and rest places
SPA and massage
Sport studio and gym
Park and garden
Entertainment
Theater
Bar
Count of venues of each class around Liberty Village
Count
class
Cafés
Restaurants
Bar
Sport studio and gym
Run and play places
Count of venues of each class around Yonge & St. Clair
Count
class
Restaurants
Cafés
Bar
Sport studio and gym
Banks
Movie Theater


Count of venues of each class around Davisville Centre
Count
class
Restaurants
Cafés
Run and play places
Bar
Park and garden
Entertainment
Banks
Count of venues of each class around Yonge and Lawrence
Count
class
Cafés
Restaurants
Banks
Bar
Pools and swimming
Park and garden
Count of venues of each class around Don Mills
Count
class
Restaurants
Cafés
Sport studio and gym
Entertainment
Count of venues of each class around Islington Station
Count
class
Restaurants
Cafés
Bar
Sport studio and gym
Count of venues of each class around Yonge and Sheppard
Count
class
Cafés
Restaurants
Bar
Banks
Park and garden
Run and play places
Count of venues of each class around Toronto - North York
Count
class
Restaurants
Cafés
Park and garden
Bar
Sport studio and gym
Theater
Hotels and rest places
Banks
Pools and swimming
Movie Theater


Count of venues of each class around Parkway Place
Count
class
Restaurants
Cafés
Sport studio and gym
Hotels and rest places
Count of venues of each class around West Toronto - Etobicoke
Count
class
Restaurants
Sport studio and gym
Park and garden
Cafés
Theater
Banks
Hotels and rest places
Count of venues of each class around North American Centre
Count
class
Restaurants
Cafés
Park and garden
Sport studio and gym
Banks
Run and play places
Bar
Count of venues of each class around Midland Avenue
Count
class
Restaurants
Cafés
Park and garden
Count of venues of each class around Queen & Bay
Count
class
Restaurants
Cafés
Sport studio and gym
Bar
Theater
Hotels and rest places
Museums and Exibitions
Banks
Count of venues of each class around Brookfield Place
Count
class
Cafés
Restaurants
Bar
Hotels and rest places
Museums and Exibitions
Sport studio and gym
Park and garden

According to the presented schedules, the leaders in the completeness of infrastructure have become

- Yonge & Shuter - 100 venues around,
- Eaton Centre - 100 venues around,
- Yonge and Richmond Centre - 100 venues around,
- Queen & Richmond Centre - 71 venues around,
- Waterpark Place - 77 venues around,
- Toronto - North York - 52 venues around,
- West Toronto - Etobicoke - 9 venues around,
- Queen & Bay are - 100 venues around.

Although around the West Toronto - Etobicoke found only 9 infrastructure facilities, their list is full: there is a hotel, restaurants, bank, theatre and sports infrastructure.

4.2. Visualization project objects on Toronto map

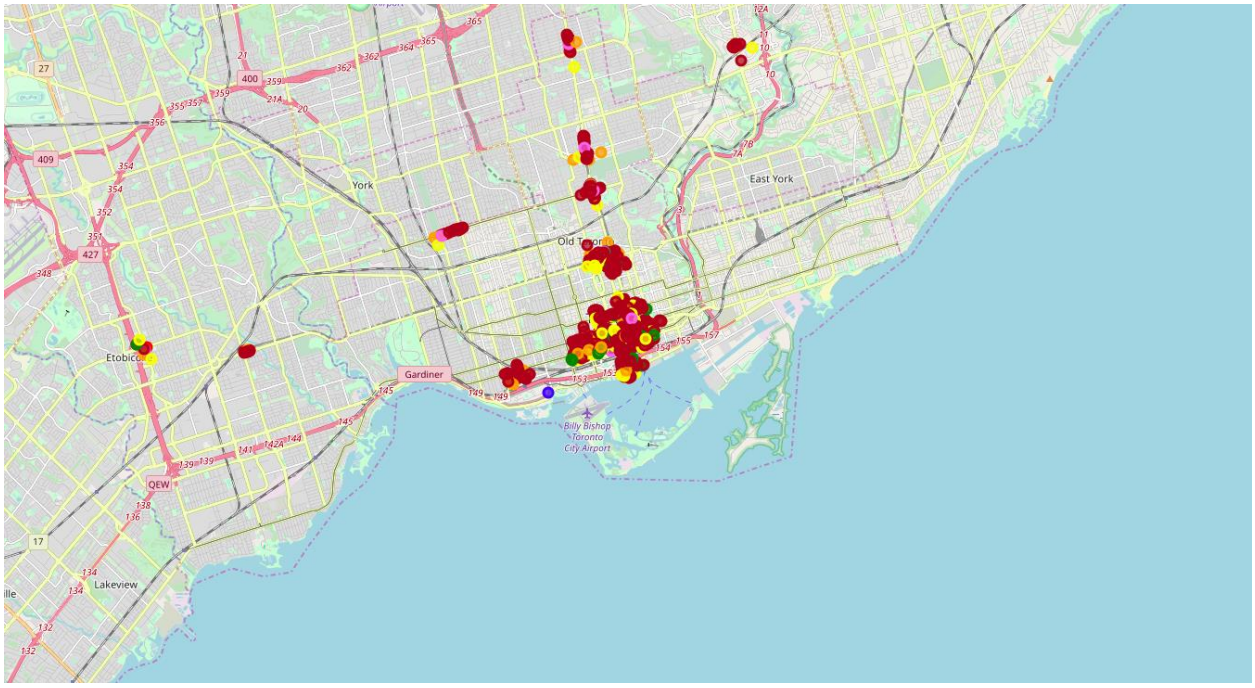


Figure 5. Location of office centres, conference centres and infrastructure facilities on a Toronto map

5. Methodology: Modelling and Evaluation

Our project involves the creation of a methodology for ranking potential office spaces by their attractiveness to the customer, as well as depending on the price terms of the lease. For ranking purposes it is necessary to develop an indicators of the attractiveness of the rental object. We decided to use 2 type of indicators:

- The nearest conference centre Availability and Quality indicator
- The Infrastructure Quality Indicator.

Further, these indicators will be constructed and our office objects will be clustered taking into account their locations, the quality indicators and price characteristics.

5.1 Creating Conference Centres Availability and Quality indicator (Integral rating CC)

Integral rating CC methodology

The method of forming the integral rating of the conference centres is based on the rating of this conference centres, calculated on the basis of customer surveys on resource *EventSource.ca*. Further, this rating is adjusted in accordance with the rental price level of the conference centres premises (coefficient k1) and the proximity of this object to the office space (coefficient k2). The distance to the conference centre more than 10 km is considered unacceptable and leads to a reset of the conference centre rating.

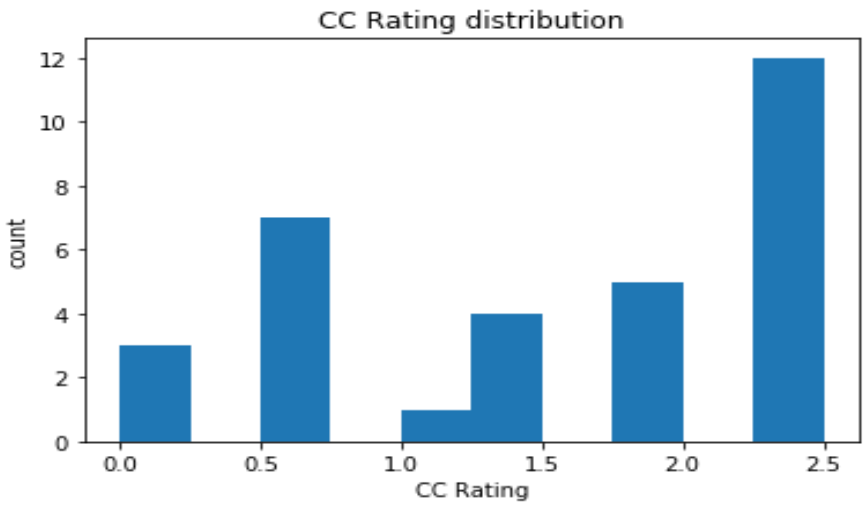


Figure 6. Distribution of office centres on Integral rating CC

5.2 Creating Infrastructure Quality indicator

Infrastructure QI methodology

The method of forming the integral rating of the infrastructure is based on the average number of infrastructure facilities located within a radius of 500 m around the office space. Further, this rating is adjusted in accordance with the presence in the neighbourhood of hotels, restaurants, swimming pools or SPA, bank (coefficients k3, k4, k5, k6).

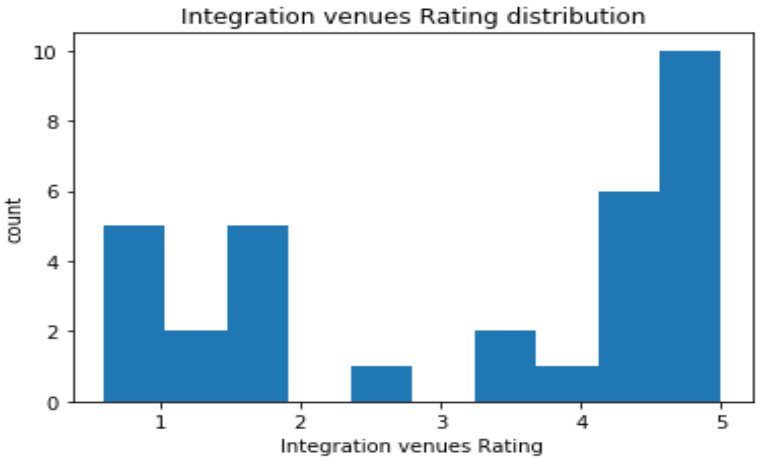


Figure 7. Distribution of office centres on Infrastructure Quality indicator

5.3. Distribution OS prices

The third factor that determines the choice of office space is the rental price. Consider how our objects are distributed by price level, as well as average prices.

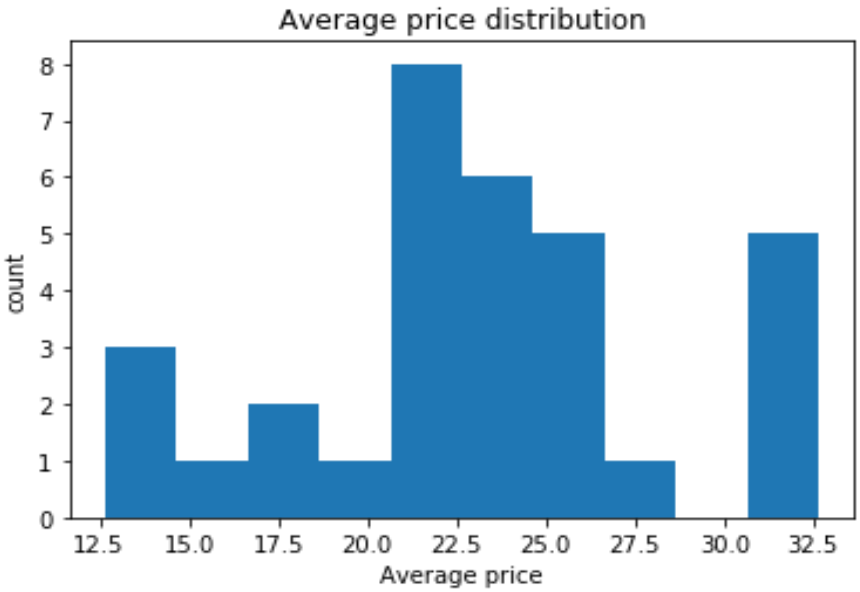


Figure 8. Distribution of office centres on Average rent prices

5.4. Joint distribution of office spaces on Integral rating CC, Infrastructure Quality indicator and Average prices

We visualize the distribution of our office spaces in three ways, which will be used in the process of clustering

- Integral rating CC (R_CC)
- Infrastructure Quality indicator (R_Inf)
- Average prices.

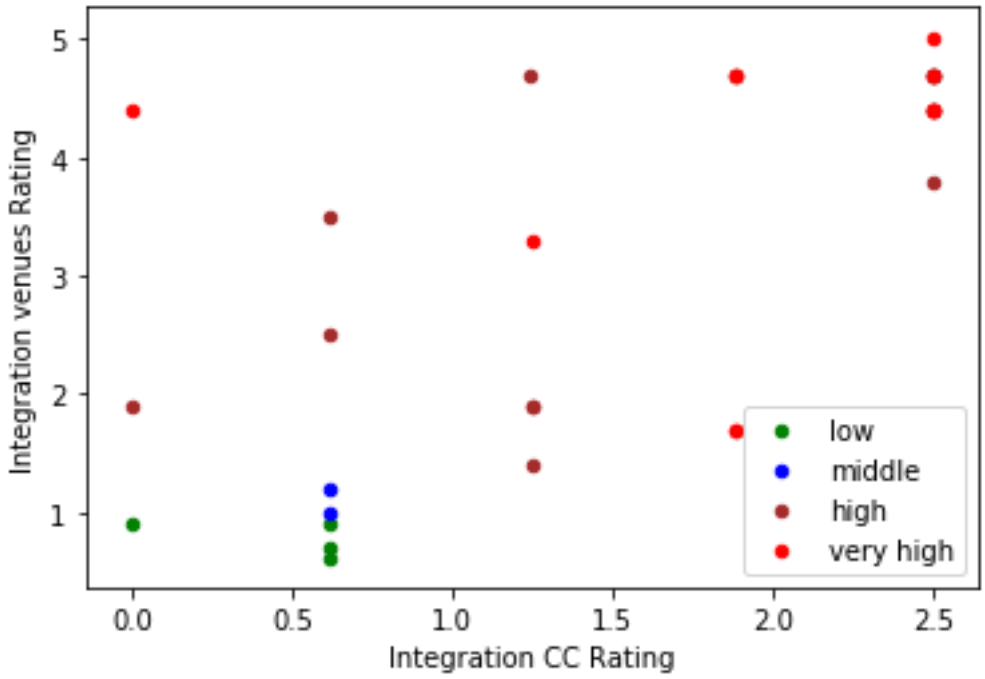


Figure 9. Join distribution of office centres on rated factors

An important result

In the figure above we see three types of objects:

- with low quality ratings and low prices,
- with low ratings and high prices,
- with high ratings and high prices.

This determined the choice of the number of classes for clustering.

5.5. Clustering

Machine learning

We will now cluster all these office spaces based on their average price, Integral rating CC, Infrastructure Quality indicator and locations, using ***KMeans clustering algorithm***.

We'll cluster the object into ***three separate groups***, since the previous data analysis showed exactly the most likely number of clusters.

```
from sklearn.cluster import KMeans NO_OF_CLUSTERS = 3  clustering = df_of.
drop(['Name', 'Link', 'Type', 'Postcode', 'Location','Adress', 'Price_rang
e', 'Rating', 'Nearest CC',\        'Nearest CC distance', 'Nearest CC ran
```

```
ge', 'Nearest CC price rate',\          'Hotels amount', 'Entertainment amou
nt', 'Pools and SPA amount',\          'Sport objects amount', 'Banks amount
', 'Restorants and cafe amount',\          'Total venues amount'], 1) kMeans
 = KMeans(n_clusters = NO_OF_CLUSTERS, random_state = 0).fit(clustering) d
f_of.insert(0, 'cluster_labels', kMeans.labels_)
```

6. Results and Discussion

6.1. Location of clusters on the Toronto map

- the first is green,
- the second is red,
- the third is blue

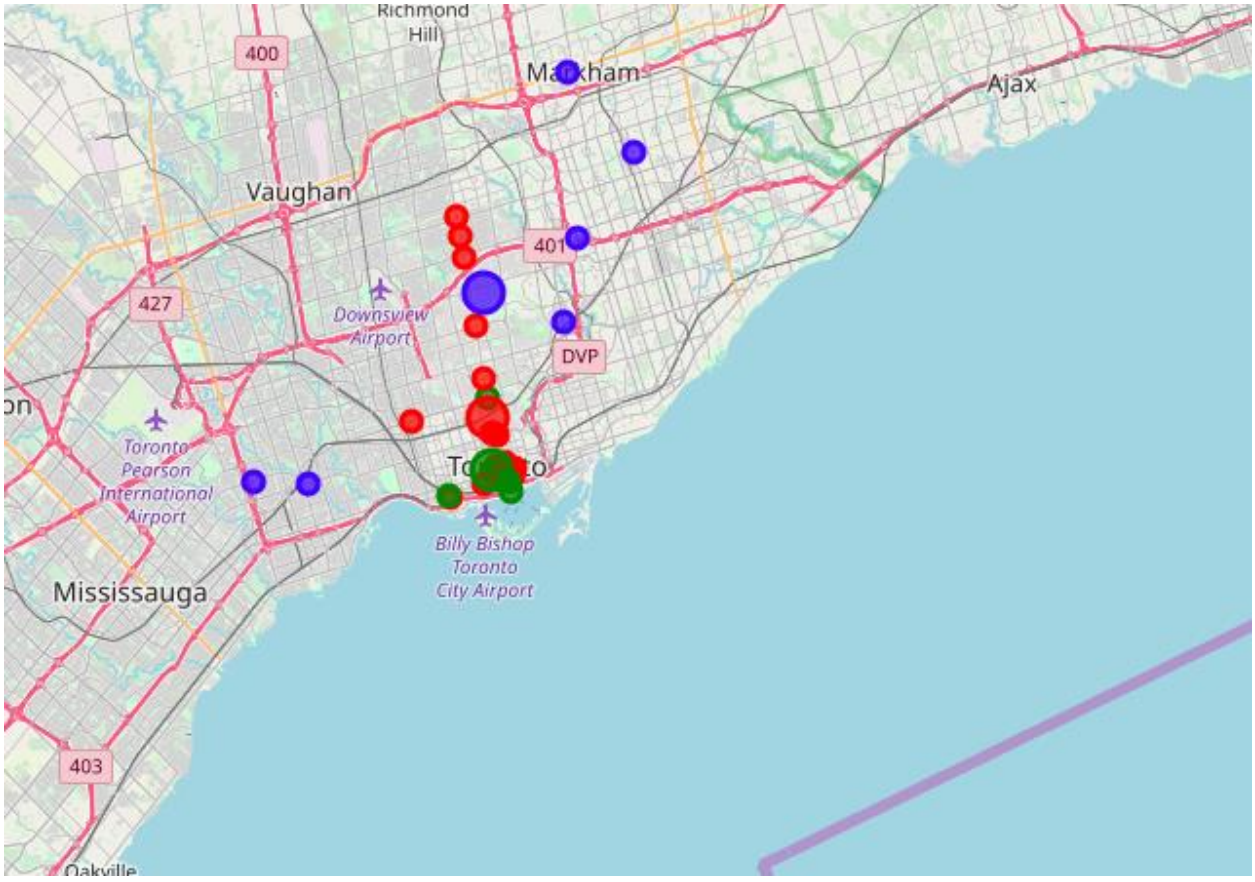


Figure 10. Location of office centres clusters on a Toronto map

We see that geographically the objects of the first cluster are located closer to the coast and the historic center of Toronto, the objects of the second and third cluster are mostly removed inland.

6.2. First cluster features

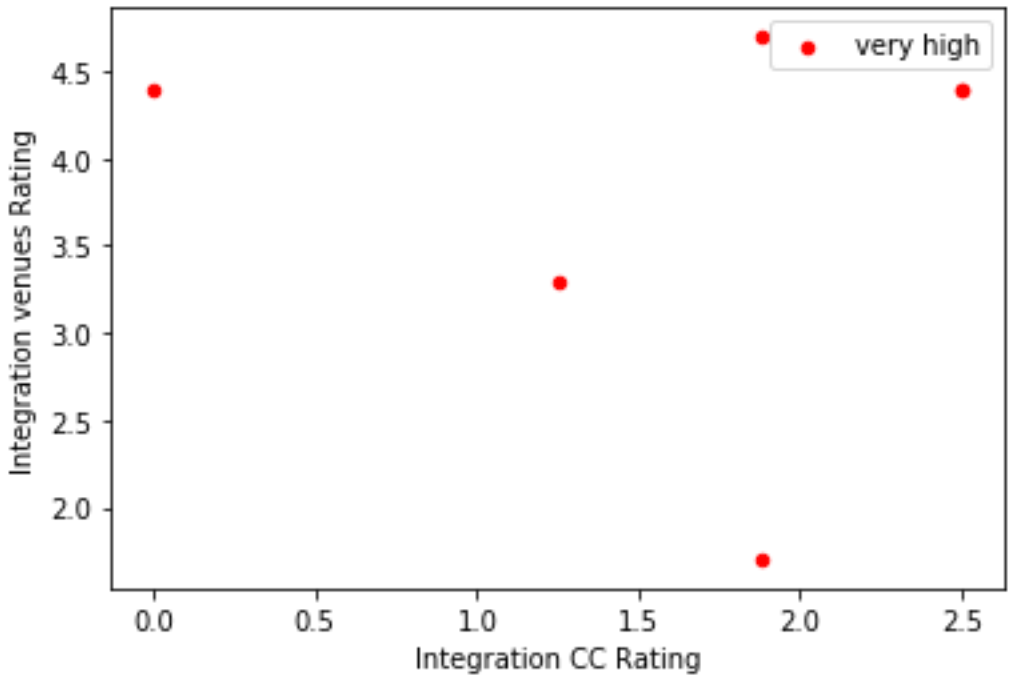


Figure 11. First cluster objects distribution

The first class contains premium properties with a high level of demand for them, which determines the highest rental prices.

6.3. Second cluster features

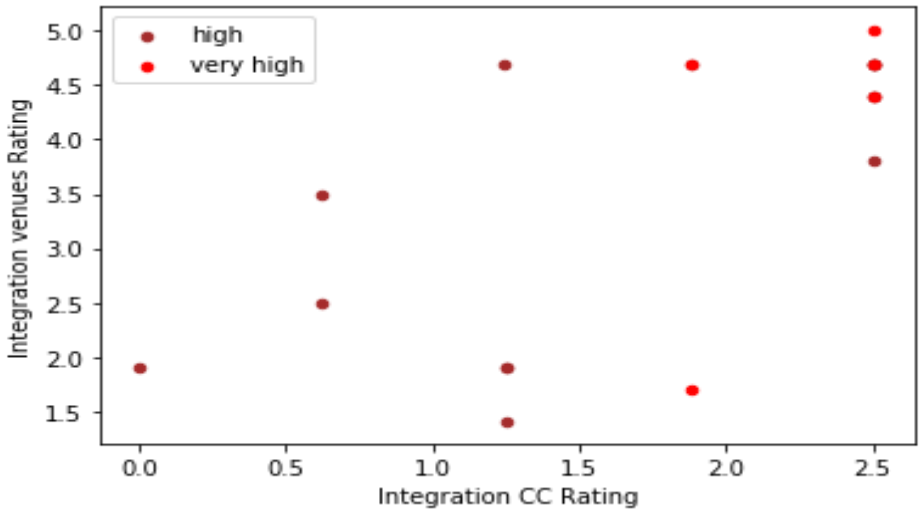


Figure 12. Second cluster objects distribution

The second cluster included office premises from the top two price categories. These objects have high or average values of 'Integration venues Rating' and 'Integration CC Rating'.

6.4. Third cluster features

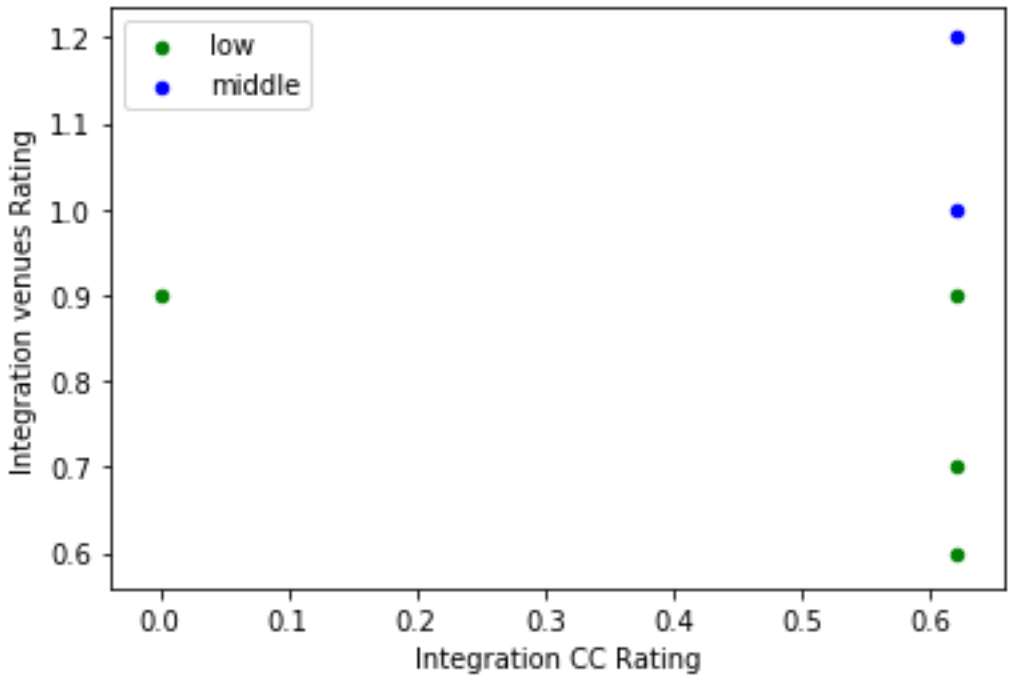


Figure 13. Third cluster objects distribution

The third cluster includes office premises from the two lower price categories. These objects have the lowest values of 'Integration venues Rating' and 'Integration of CC Rating'

7. Conclusion

Machine learning allowed us to identify clusters of office spaces that differ in the level of prices and quality of service.

- Although objects with low prices were included in the ***third cluster***, their quality characteristics least meet the requirements of the client.
- ***The second cluster*** will be most preferable, since it includes objects with first-class characteristics, among which you can choose objects with an acceptable price level.

- ***The first cluster*** - the most expensive, but provides all kinds of services necessary for the client.

For example, **we can recommend Exchange Tower**.

Near the Exchange Tower there is

- a high-class Oakham House conference centre,
- 7 hotels,
- 33 restaurants and cafes,
- a swimming pool. OS has
- an Integration CC Rating of 2.5
- an Integration Venues Rating of 4.7

Thus, we have solved the task of developing a methodology for finding office space that meets the criteria for customers. Using information from the FouthSquere resource has proven extremely useful.