

# Report of the analysis of the foursquare venues data of Toronto



UNIVERSITY OF  
BIRMINGHAM

Manuel Maas<sup>1</sup>

1

1191 Words

Student ID:000000

**Keywords:** Toronto, data science, foursquare

28<sup>th</sup> December, 2020

---

## Abstract

In this analysis the data provided by the website *foursquare.com* will be used to generate some insights which could be used by estate agents. The analysis will be divided in three parts which will be discussed later.

---

## 1 Introduction

The main idea here is to gather information on the venues of Toronto. One could think of estate agents which could use this data analysis to provide a better service for their costumers.

### 1.1 Part I: Similarity of Neighborhoods in Toronto

In this part a comparison of the neighborhoods of Toronto will be done and it will be determined, based on there numbers of venues, how similar they are. This information could be used to find the neighborhood which is promising for a company to open a new business. If for example a company has a business in one neighborhood in Toronto which does good profit it could use this algorithm to determine the most similar neighborhood and open a new business there. Another example for this application would be a citizen which wants to move in Toronto and wants to find a similar neighborhood. So this algorithm could be used by estate agents.

### 1.2 Part II: Correlation between venues

In this part the correlation of the different kinds of venues will be analysed. For example if a costumer is especially interested in one kind of venue (e.g café) we could look at the correlation and tell him whether it is likely to find another kind of venue (e.g asian restaurant) nearby.

### 1.3 Part III: Dependency of the mean of the frequency of occurrence of venues from the distance to the city core

In this part I want to take a closer look at the dependency of the mean of the frequency of occurrence of venues from the city core. We could find out wether a costumer should live closer to city core or further away. If the costumer for example want's to live near a Asian Restaurant he should move closer to the city core.

## 2 Data

The dataset which will be used is a table with neighborhoods of Toronto and their mean of the frequency of occurrence of each Venue Category. This table is generated by using the data available on *foursquare.com*. From this website all the venues within a 500 meter radius from each neighborhood were extracted. The category of the venues were represented using one-hot encoding and listed in a table with the name of the belonging neighborhood. Doing some transformation to this table a new table was generated with a row for every neighborhood and their mean of the frequency of occurrence of each Venue Category.

## 3 Methodology

In this section the Methodology of each part will be described.

### 3.1 Part I: Similarity of Neighborhoods in Toronto

In order to compare two neighborhoods the euclidean distance is used and function which converts the row to a list will be created. Now with a loop every neighborhood will be compared to each other. As an example the most similar neighborhood to 'Christie' will be determined. In addition the dataframe with the euclidean distances will be visualized.

### 3.2 Part II: Correlation between venues

For this part the ".corr()-function" of pandas, a Python Data Analysis Library, will simply be used. For some venues the correlation will be plotted and some statements about the correlation between some venues will be made.

### 3.3 Part III: Dependency of the mean of the frequency of occurrence of venues from the distance to the city core

In this part the center is defined with its latitude and longitude. The neighborhoods and the center are visualized using folium. For the calculation of the distance to the center a function will be defined. The a new dataframe with all the distances will be generated and the ".corr()-function" will be used to determine the correlation between the distance from the center and the mean of the frequency of occurrence of each venue. In addition the dependency of mean of the frequency of occurrence of the venues 'Pub', 'Asian Restaurant' and 'Café' on the distance to the city core will be plotted.

## 4 Results

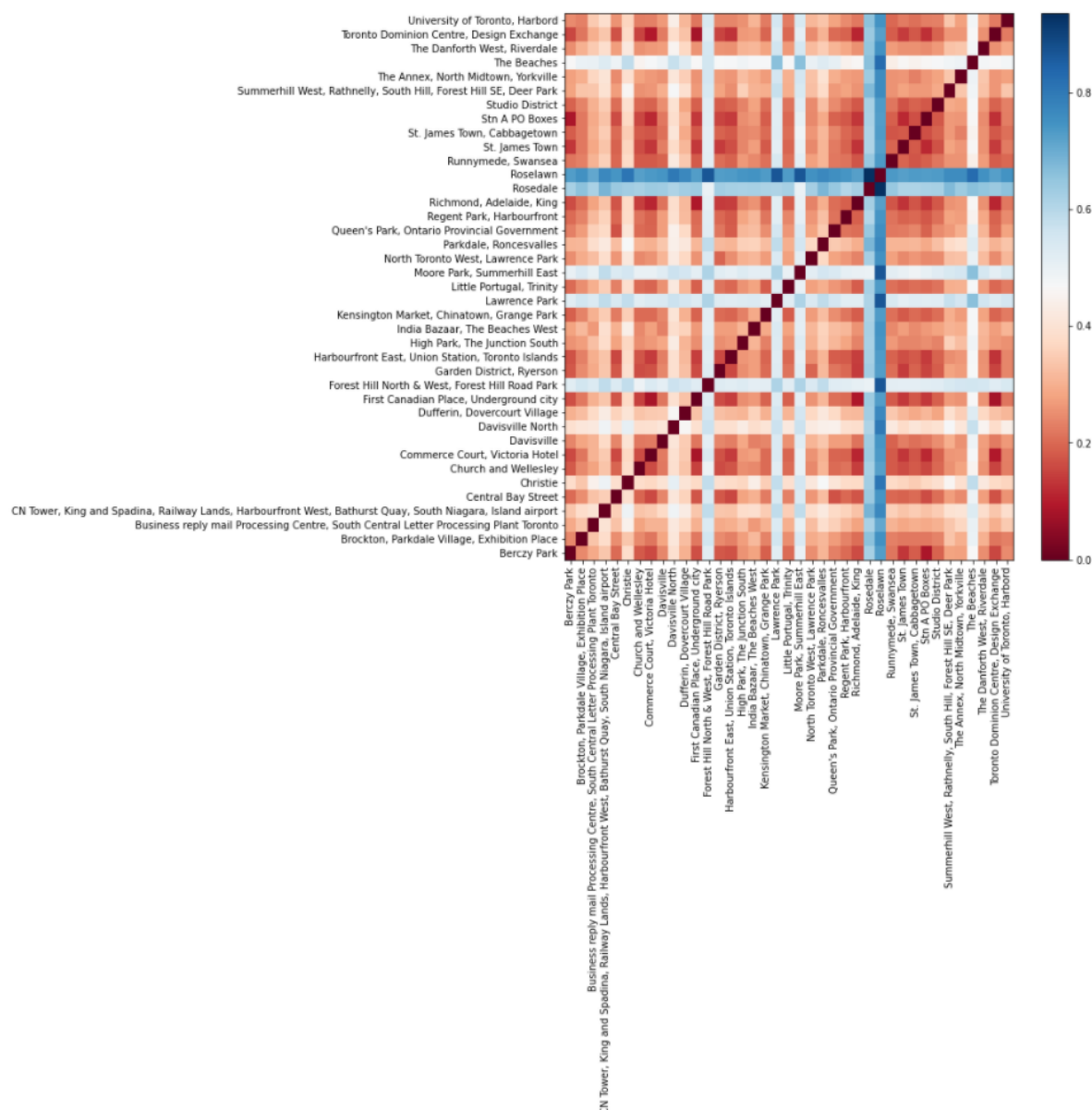
The results of each part will be shown in this section.

## 4.1 Part I: Similarity of Neighborhoods in Toronto

In figure 1 the visualization of the euclidean distance of the different neighborhoods of Toronto based on their mean of the frequency of occurrence of each Venue Category can be seen.

As an example the euclidean distance was used to determine the most similar Neighborhood to 'Christie'. In this example the neighborhood 'Brockton, Parkdale Village, Exhibition Place' was the most similar.

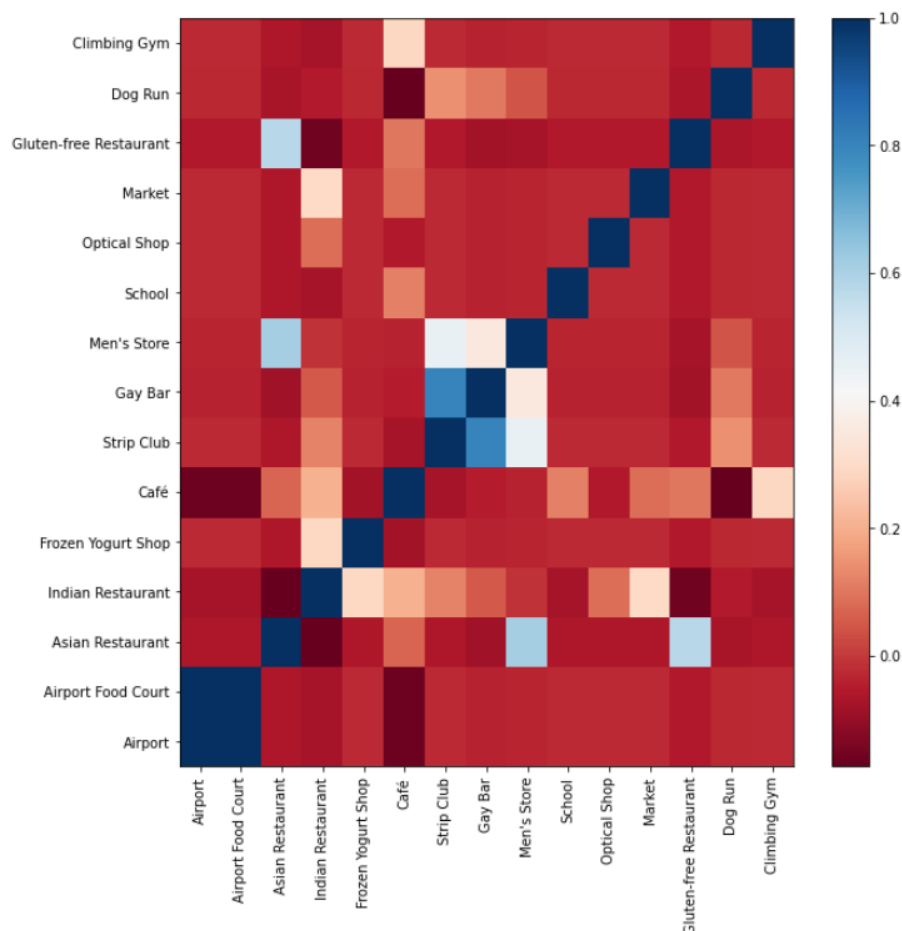
Interesting is that, as it can be seen from figure 1, the Neighborhoods 'Roselawn' and 'Rosedale' are quite different from the other Neighborhoods in this calculation. They are only very similar to themselves.



**Fig. 1.** Euclidean distance of the different neighborhoods of Toronto based on their mean of the frequency of occurrence of each Venue Category.

## 4.2 Part II: Correlation between venues

The visualization of the correlation between some of the venues can be seen in figure 2. The venues which were chosen were "Airport", "Airport Food Court", "Asian Restaurant", "Indian Restaurant", "Frozen Yogurt Shop", "Café", "Strip Club", "Gay Bar", "Men's Store", "School", "Optical Shop", "Market", "Gluten-free Restaurant", "Dog Run" and "Climbing Gym". The visualization was only made on some venues because there were 235 venue categories and this would have been too big to visualize.



**Fig. 2.** The correlation between some of the venues.

From the correlation between the different venues some statements were made:

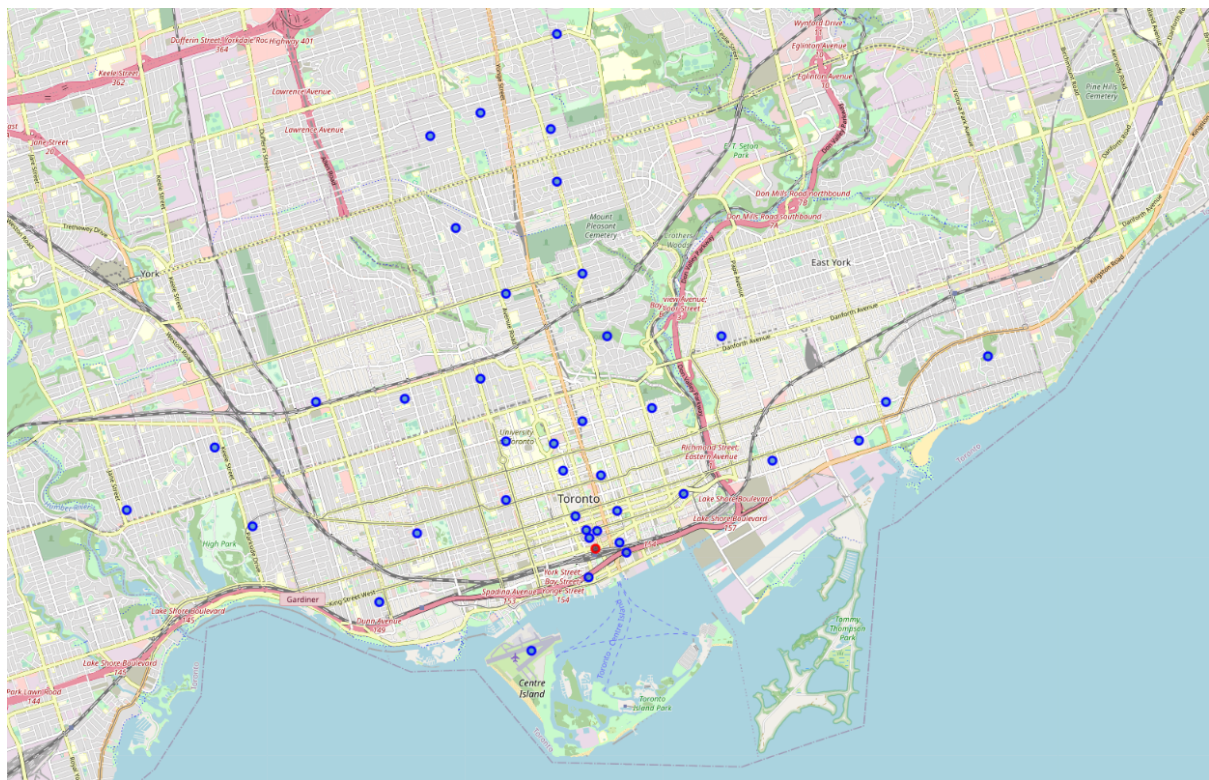
- "Airport" and "Airport Food Court" correlate very strongly because every Airport seems to have a Food Court.
- "School" and "Strip Club" correlate negatively. This makes sense because schools are usually not built near strip clubs.
- "Gluten-free Restaurant" and "Asian Restaurant" are highly correlated. So if you want to eat gluten-free you should look for asian restaurants.

- "Climbing Gym" and "Café" are also highly correlated. So if you want to live near both venues it should not be a problem.
- "Strip Club", "Gay Bar" and "Men's Store" are highly correlated. Also "Men's Store" correlates with "Asian Restaurant" and "Gluten-free Restaurant".

### 4.3 Part III: Dependency of the mean of the frequency of occurrence of venues from the distance to the city core

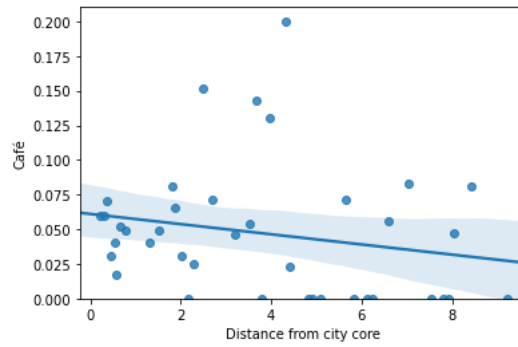
In figure 3 the map of Toronto was visualized using the Python library 'Folium'. In addition in this map the different neighborhood are shown as well as the city core. The latitude of 43.645441936641646 and the longitude -79.38023871946247 were used as the city core.

As an example the dependency of mean of the frequency of occurrence of the venues

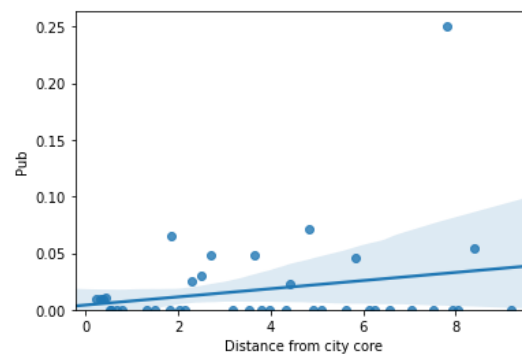


**Fig. 3.** Map of Toronto with the neighborhoods and the city core.

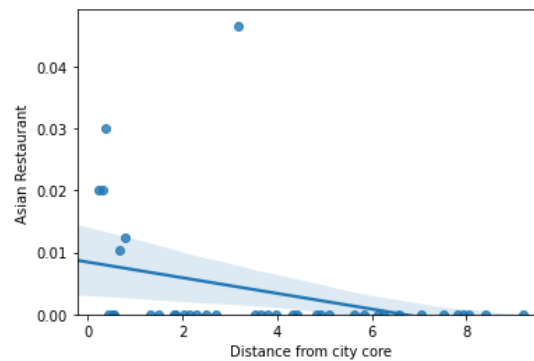
'Pub', 'Asian Restaurant' and 'Café' are shown in the figures 4-6. In this figures it can be seen that 'Asian Restaurants' and 'café' have a negative correlation on the distance from city core. This means Asian Restaurants are most likely near the city core. The venue category 'pubs' seems to have a positive correlation on the distance from city core. This means pubs are most likely not near the city core.



**Fig. 4.** Relationship between the mean of the frequency of occurrence of the venue category 'Café' and the distance from the city core.



**Fig. 5.** Relationship between the mean of the frequency of occurrence of the venue category 'Pub' and the distance from the city core.



**Fig. 6.** Relationship between the mean of the frequency of occurrence of the venue category 'Asian Restaurant' and the distance from the city core.

## 5 Discussion

### 5.1 Part I: Similarity of Neighborhoods in Toronto

The Similarity of Neighborhoods in Toronto was calculated using the venue data and the euclidean distance. This dataset can be used to suggest a citizen who wants to move within Toronto. If a costumer of an estate agent for example wants to live in a similar neighborhood

to 'Christie' the estate agent could suggest 'Brockton, Parkdale Village, Exhibition Place' based on this data analysis.

## 5.2 Part II: Correlation between venues

The correlation between venues could be generated and plotted. If a costumer of an estate agent for example wants to live near a "Climbing Gym" and "Café" the estate agent could tell him that this is not a problem at all.

## 5.3 Part III: Dependency of the mean of the frequency of occurrence of venues from the distance to the city core

The dependency of the mean of the frequency of occurrence of venues from the distance to the city core was successfully created. If a costumer of an estate agent for example wants to live near a "Café" the estate agent could suggest to him that it could be a smart option to move to the city core because it is more likely to find a "Café" there.

# 6 Conclusion

Some useful information using the data provided by *foursquare.com* could be generated. The results can become handy for real estate agents who could use this data to provide better service for their costumers.