**Capstone Project:**

**The clash of Neighborhoods in Toronto City**

# Introduction

**Problem**

How To Find The Best Neighborhood In Toronto Based On Specific Needs And Requirements For The Robert’s Family ?

**Background**

We are living in Toronto, Canada, since many years and we have a long-time friend in France moving in Toronto with his family. As he is not familiar with our city, he asks our help to **find a neighborhood that match with all his family’s needs**. The Robert’s have four children aged from 3 to 11 years old and three border collie dogs which are very dynamic. The parents work from home so they don’t have any geographic constraints related to their work location, but they really seek for a place that offers the best opportunities in terms of **education**, **sports facilities** and **healthcare institution** for their children. They also look for **green spaces** and **parks** nearby for their dogs and kids. Concerning their lifestyle, this family also have a very healthy diet and want a developed food offer around their home with a various **fresh market**, **farmers market**, **grocery store** and restaurants that serve fresh foods and have good tips.

# Data

The data are retrieving from three different sources. The first is a Wikipedia page to retrieve postal code and Borough of neighborhood located within the city of Toronto in the province of Ontario, Canada. (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>). The second source is a csv file that contain all specific location for neighborhoods in Toronto city (<http://cocl.us/Geospatial_data>). With these two databases, a dataframe containing Postal codes, Borough, Neighborhood, and specific location (i.e latitude and longitude) is created (Figure 1). The dataframe has 10 boroughs and 103 neighborhoods (Figure 2).

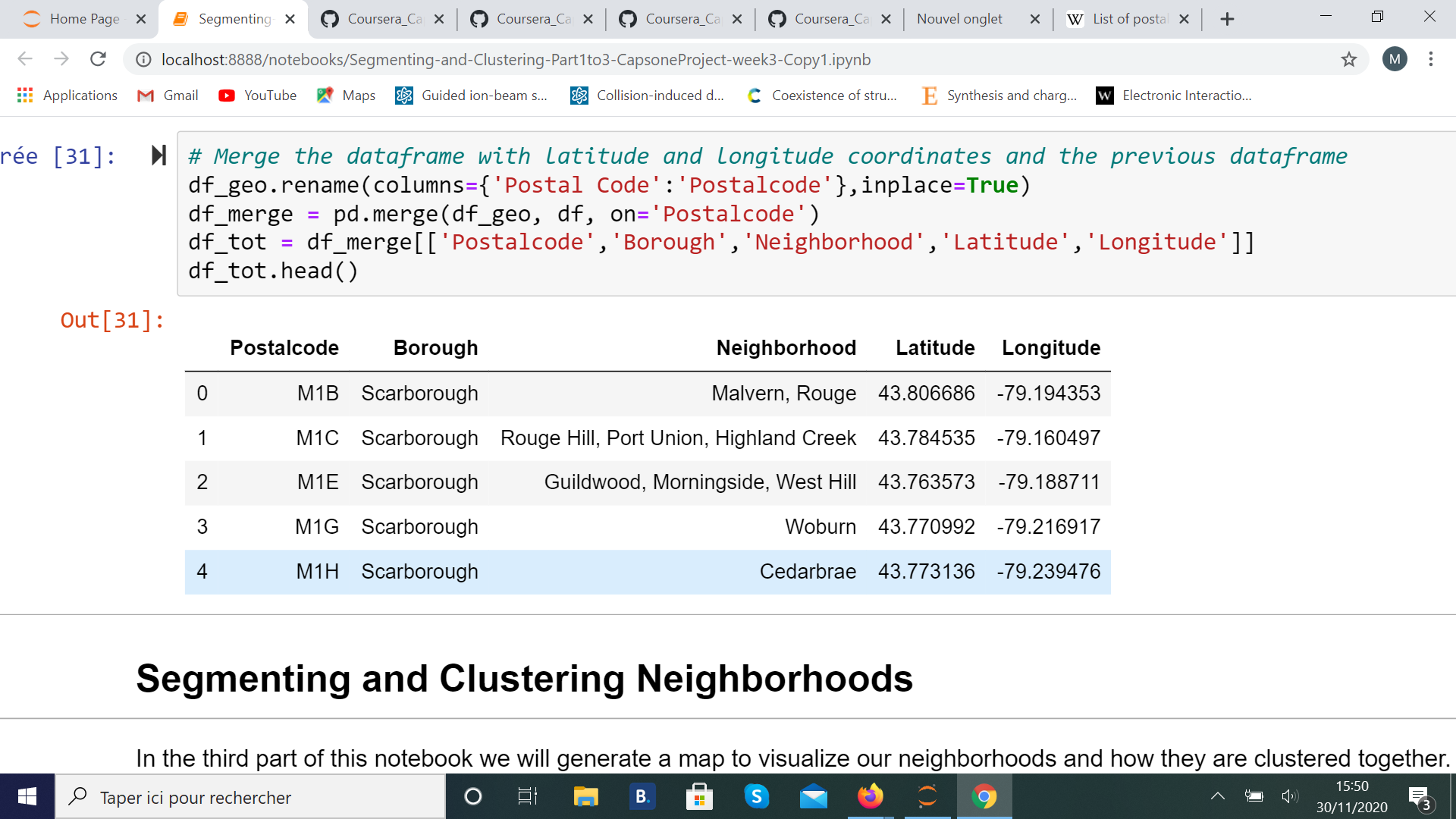


Figure 1 : Dataframe that contains keys information for neighborhood in Toronto city, Canada.

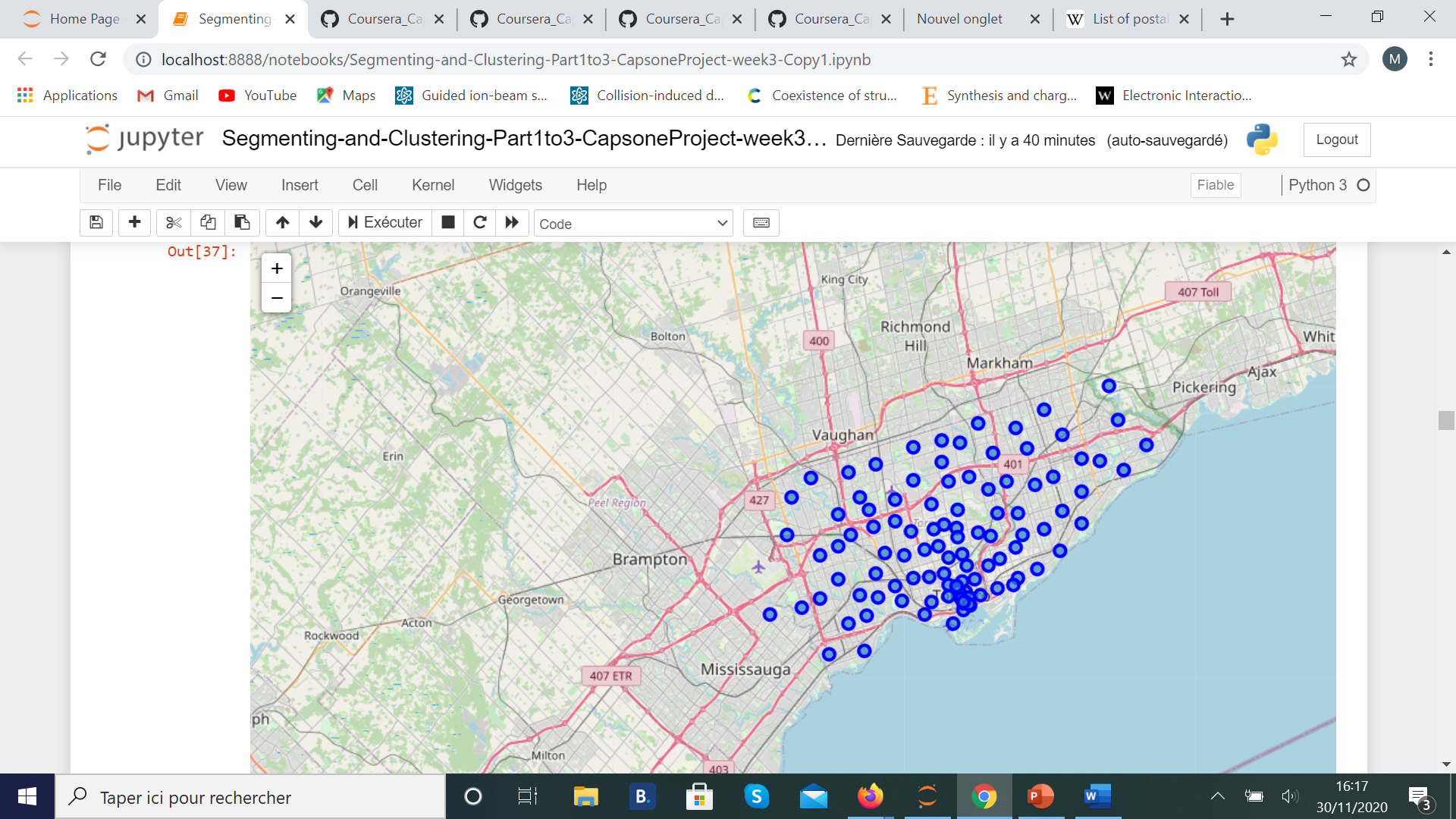


Figure 2 : Map of all neighborhoods retrieve from database for Toronto city, Canada.

The third source is the Foursquare website to retrieve database for the venue in neighborhoods in Toronto. The features available in the dataset are for example: venues, store reviews, longitude and latitude, names of the neighborhood, categories of venues, types of stores. ­­From this data we can create several dataframe, one example is given is the figure above (Figure 3).



Figure 3 : Example of dataframe that can be created with data from the Foursquare website.

When we analyze data in this database, we see that there is 2161 venues listed in Toronto city with 276 uniques categories of venue. ­­

# Methodology

The goal of the project is to find the neighborhood that fulfills the majority of the requirements an individual needs in locating the best neighborhood. From our data, we will extract information about the venues and their specific location (longitude and latitude). The first step is the segmentation of the Toronto city into neighborhoods and provide a detailed list of venues available in all this neighborhood, then we will cluster neighborhoods using the k-means algorithm. Finally, we will compare this cluster and ranked them to determine the best in terms of number of listed requirements.

# Results and discussion

The data collected was cleaned and organized into formats that could be applied in exploratory data analysis methods. A number of bar graphs were produced in order to visualize the data, and help guide the exploratory data analysis in order to draw useful results from the data.

Data Cleaning and Exploratory Data Analysis Completed

The raw Neighborhood data scraped from the Wikipedia page came from a table that also included information on other cities and towns.

This dataset was explored on a map using the Folium library (Figure 2) in order to get a better understanding of the number of neighborhoods in the City of Toronto, and their distribution in the city.

The Foursquare API was called upon to retrieve information on venues local to each of these neighborhoods (Figure 3).

If we look at the 20 first neighborhoods in terms of the number of total venues (Figure 4), we see that there the 6th first neighborhoods count 100 venues (all categories included) and the 20th neighborhood, Runnymede, Swansea, count only 30 different venues. This is an important feature because independently of the list of requirements and therefore the category of venue desired, the Robert’s family as everyone seek for a large number of services available near their future home and a to ensure a dynamic neighborhood lifestyle.

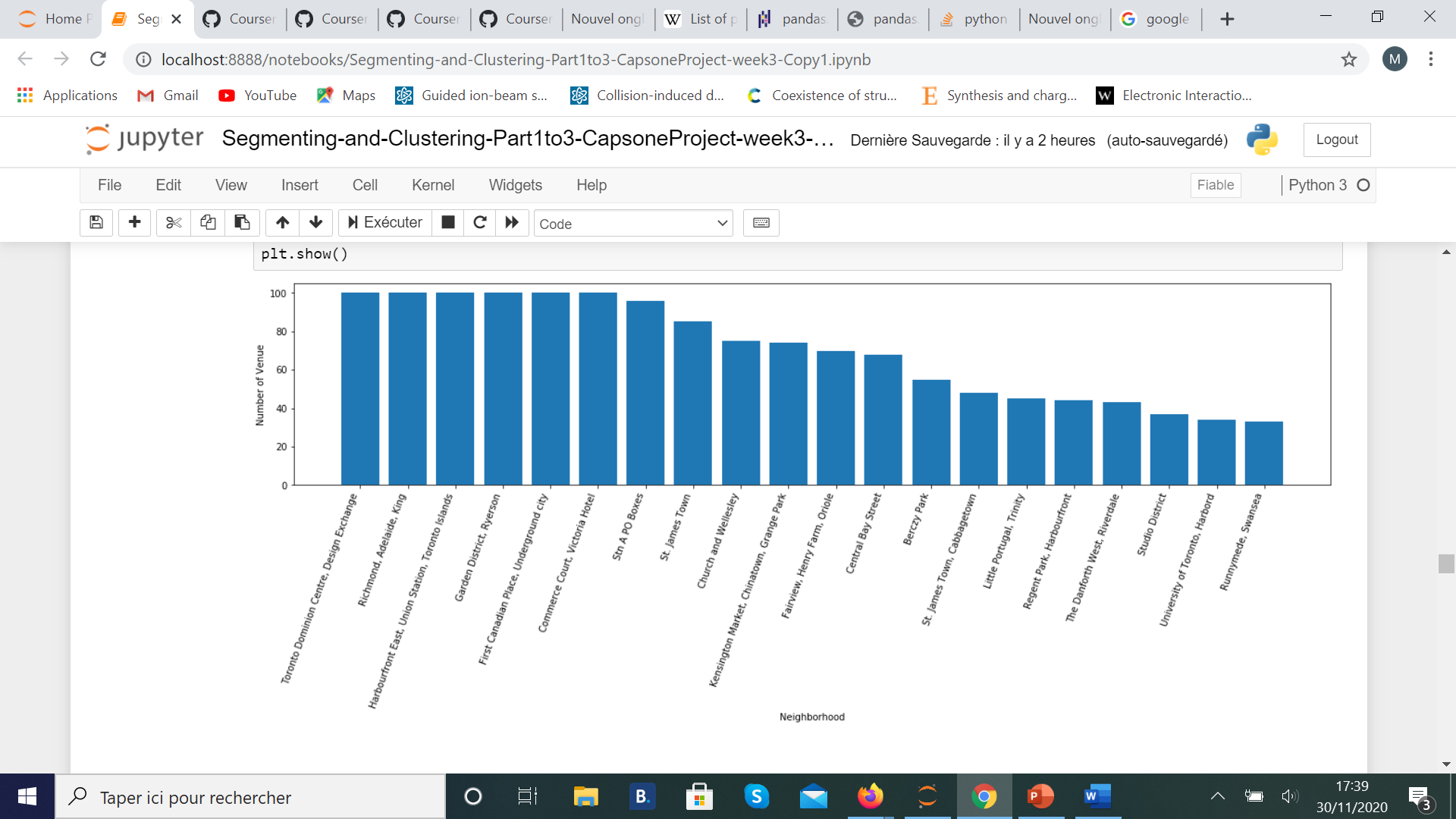
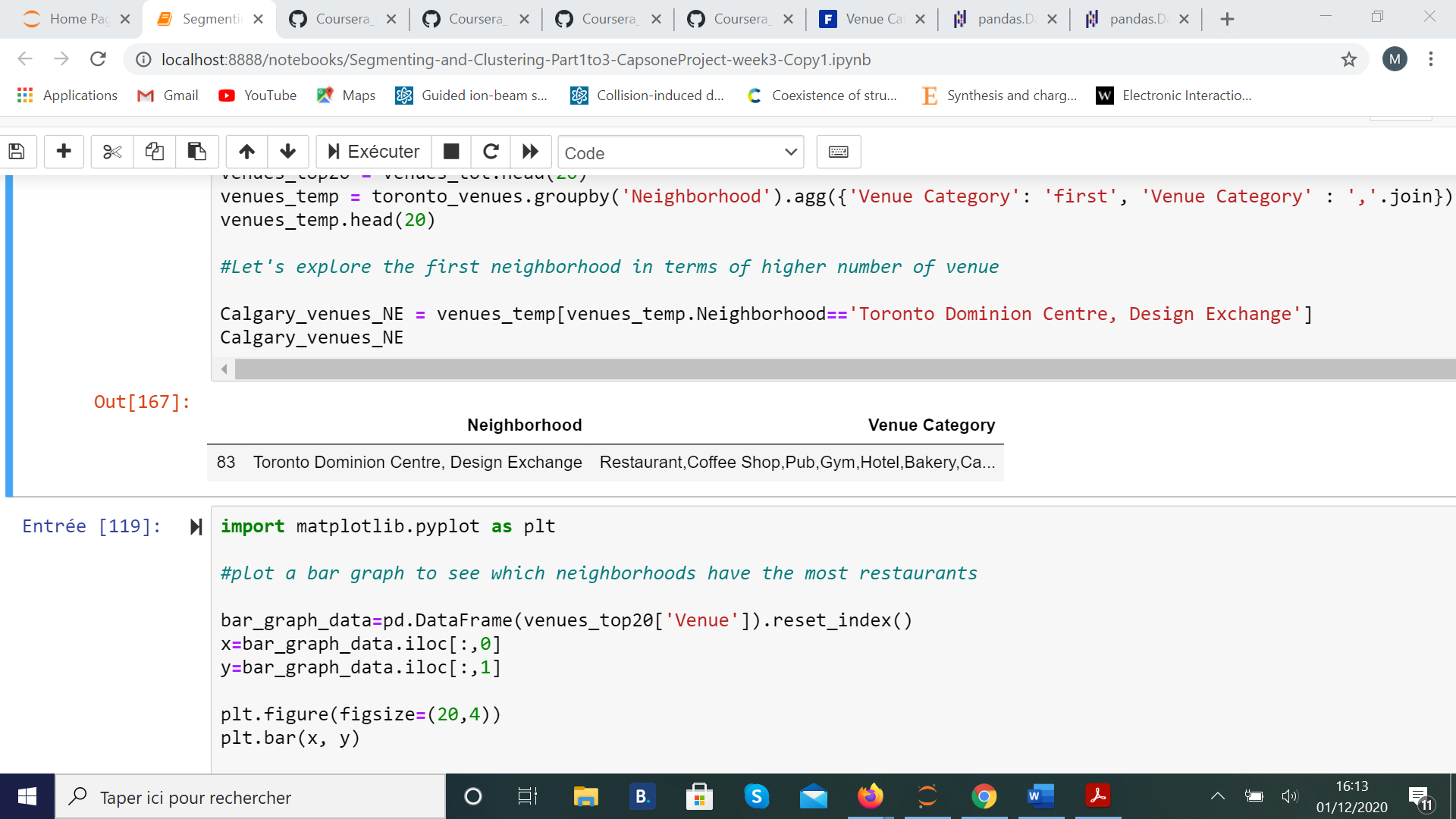


Figure 4 : Count of venues within the 20 first neighborhoods in terms of the higher number of venues.

Further examining of venue in each neighborhood, the dataset was further refined to display the frequency of each type of restaurant in each community. This information is critical to understand when answering both *Where to open a new restaurant* and *What type of restaurant to open.* The one-hot encoding methodology was applied in order to sort the dataset.



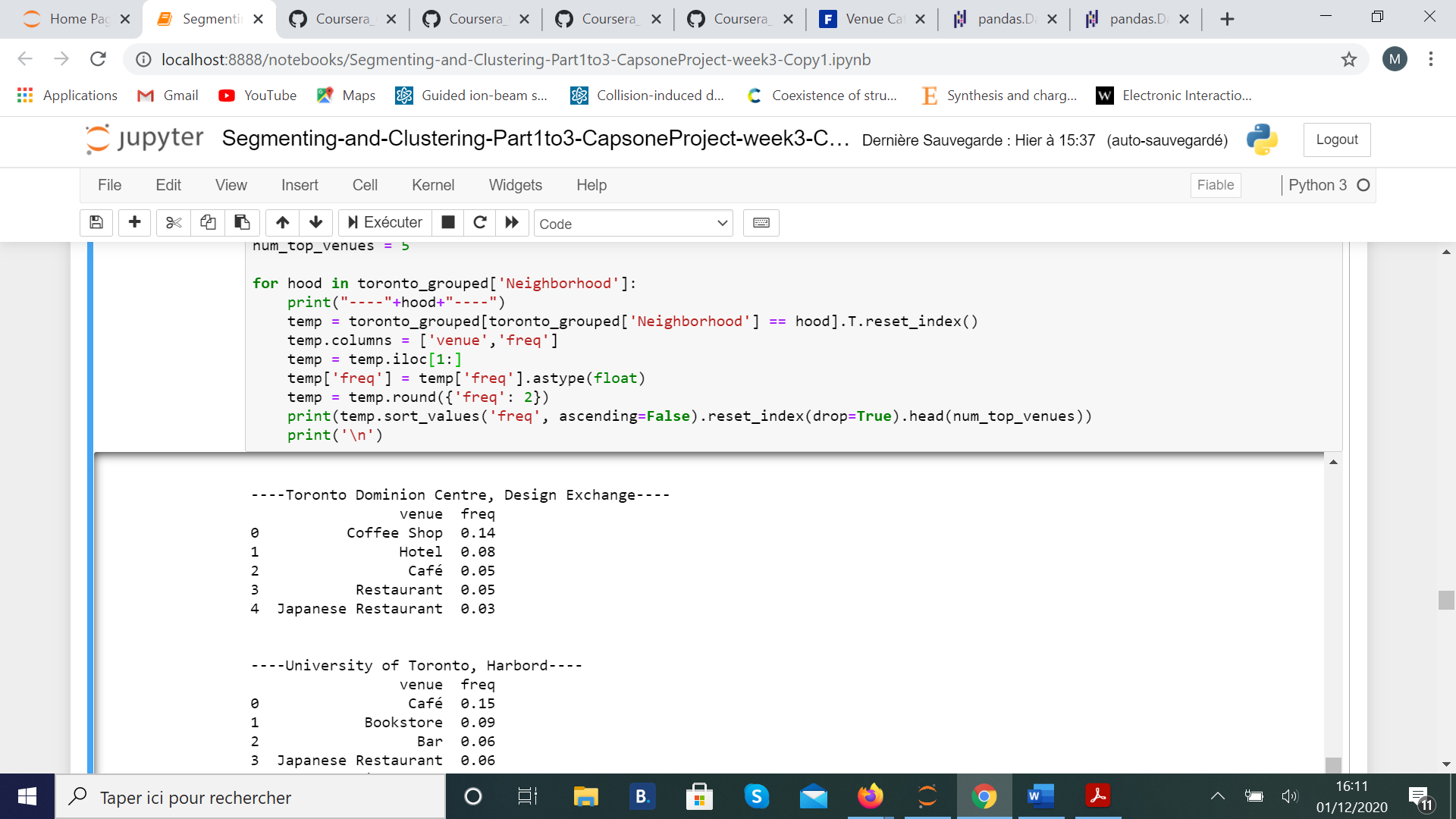


Figure 5 : Count of venues within the 20 first neighborhoods in terms of the higher number of venues.

The five most popular venue types in the Toronto Dominion Centre were displayed. We can see that the most popular venues in these neighborhoods are coffee shop and Hotel. Then we use the k-means algorithm to group neighborhoods into cluster. Each cluster will present similar most common venue. Utilizing the Folium library, these clusters were displayed on a map of Toronto to gain an appreciation for any geographical influences on this dataset.

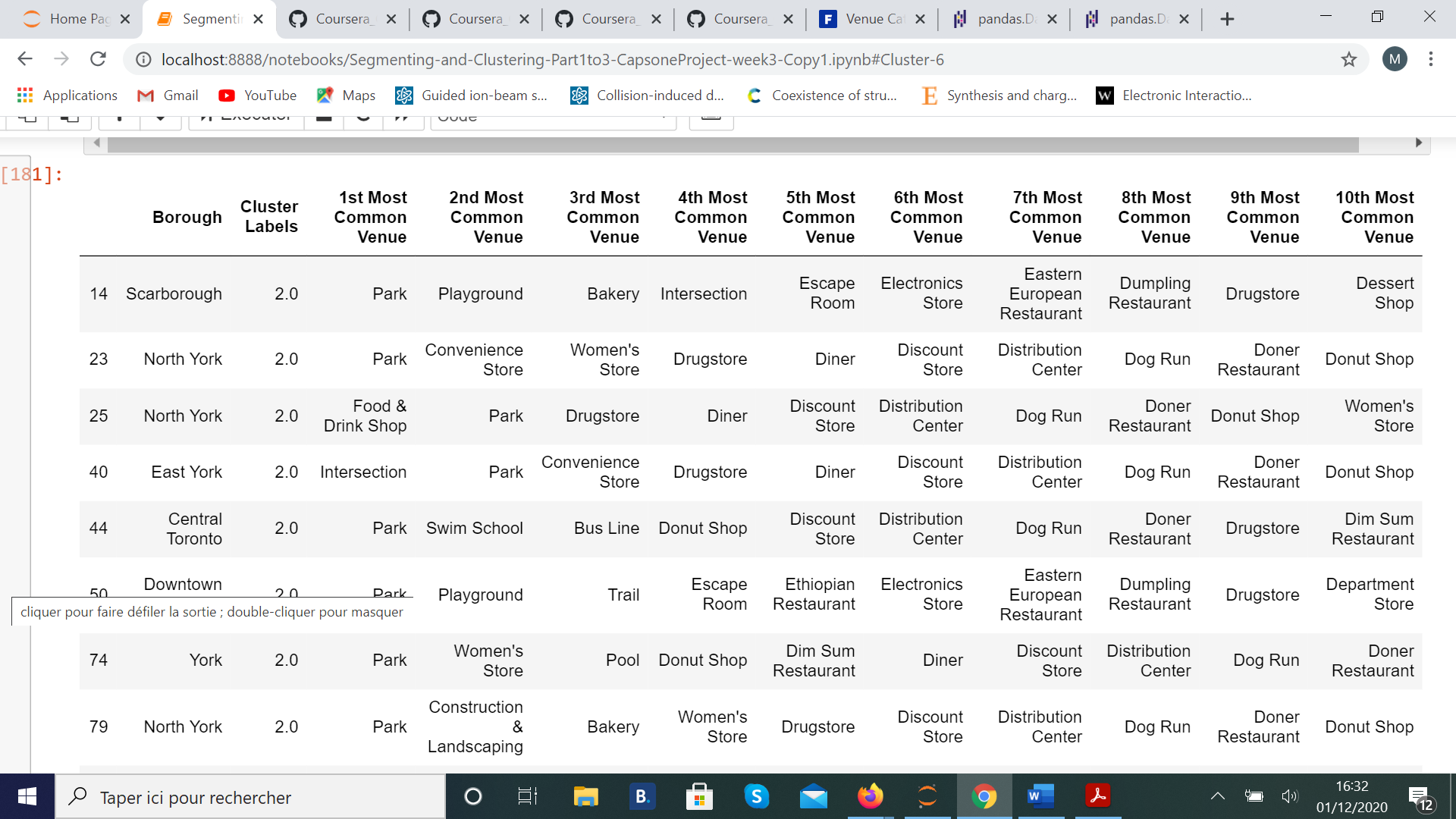


Figure 6 : Count of venues within the 20 first neighborhoods in terms of the higher number of venues.

# Conclusion

The k-means algorithm to group neighborhoods into cluster, each cluster present similar most common venue. When we analyze each of the 7 clusters, we find that only cluster 2 exhibits requirements of Robert’s family. Indeed, we see that we can find “Park” as 1st most common venue and “Playground” or “Swim school” which are among the requirements of the Robert’s family.