Competition or Complementation.

Eduardo Cruz

July 17, 2019

1. Introduction.

1.1 Background

The most common form of entrepreneurship is a small business which are very important for the economies of several developed and devoloping countries. For example, in the United States, small businesses represent the 99.9% of all firms, and nearly a half of private sector employees. Also they contributed to the 65.9% of job creation between 2000 and 2017 [1]. Small businesses are a great opportunity to achieve finacial success for a lot of people, which include women and minorities. In fact, in the last two decades there was an increase of businesses owned by women and minorities [2].
There are several topics which we must to consider before starting a new business, and a key is to develop a strategic plan.

Some ideas come to us about a new business that belongs to either, goods sector or service sector. Perhaps our idea is not so revolutionary for an industry and we want to offer a simple good or service for the customers. May our idea is something that we can find hundreds of different businesses which offer the same good or service that we want to sell. Although the majority of these businesses are small businesses, some of them are the biggest and the most popular of a certain zone, and they can sink the success of a new business nearby.

1.2 Problem

In the face of the challenge of starting a new small business there are several strong competitors that can truly obstruct the sucess of this new business. So, the problem here is to decide to compete with them or look for a complementation. In other words, this new business can try to struggle with a bigger

business with several decades of experience, or can take advantage of the popularity of some businesses.

This work aims to obtain the most common options for new businesses based in patterns of the most popupar venues in several cities.

1.3 Interest

The results of this work could be very helpful for anyone who is looking to start a new business and if he or she is not totally sure what kind of business may be the best option in a certain zone.

2. Data acquisition and cleaning

2.1 Data sources

Two cities will be compared: Toronto, Canada and New York, USA in order to obtain data about venues per boroughs and neighborhoods. All of this data will be obtained from FourSquare database.
The data about boroughs, neighborhoods names and coordinates for New York are avaliable at hppts://cocl.us/new_york_dataset and for Toronto postal codes are avaliable at https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and coordinates are avaliable at https://cocl.us/Geospatial_data. Information about name of venues, categories and localization will be categorized, compared and recorded in graphs and tables.

2.2 Data cleaning

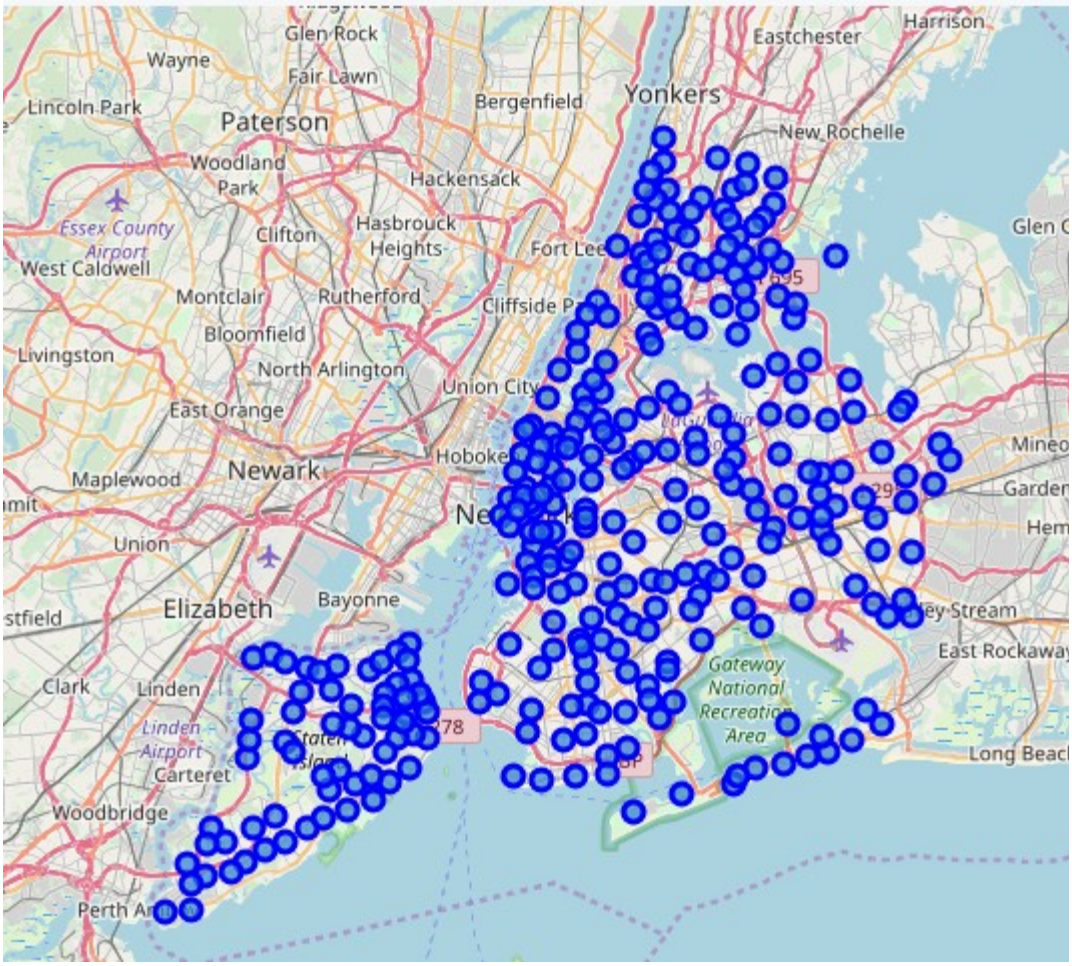
The obtained data will be compiled in different dataframes which will contain the name, category and localization of the venues from all boroughs and neighborhood of Toronto and New York. If there are missing data of some venue, it will be dropped.

2.3 Feature selection

For each neighborhood of every borough will be recolected the most popular venues, ranking from “top 100” to “top 20” and data of their name, category and localization will be grouped in dataframes. The categories of the most popular venues will be compared to each neighborhood in order to know if there is a common category or the popular venues are distributed in several categories.
For every “top”, will be consulted near venues, and also their category will be recorded in dataframes. Their categories will be displayed in histograms for visualizing the most common venue categories near to a very popular venue. Given the most common categories of the near venues, will be analized if some of these venues also belong to some “top”, in other words, if they are also popular compared to the most popular venue.
Then, will be analized what case happens more frequently: given a popular venue and their nearest venues, their categories are the same or they diverge? This analysis will indicate the cases in which there is competition and cases of complementation.

3. Explanatory Data Analysis.

Two cities were compared: New York and Toronto. First the map of New York was displayed in order to visualize the localization of its neighborhoods (Map 1). Each blue point indicates a neighborhood.
In every neighborhood the 100 most ,popular venues were grouped in a dataframe, then they were sorted by category. A new dataframe which contained the category of the first 20 most popular venues of every neighborhood were generated.



Map 1. **Map of New York.** In blue points are indicated every neighborhood.

Finally in a new dataframe were counted by neighborhoods the categories of the first most popular venue, the second most popular and the third most popular. These results were also plotted.

For example, in Table 1 we can see that “Pizza Place” is category of the first common place in 36 neighborhoods, “Deli / Bodega” is the category of the first common place in 28 neighborhoods, “Italian Restaurant” is the category of the first common place in 23 neighborhoods, and so on. The results were plotted in Graph 1.

	1st Most Common Venue
Pizza Place	36
Deli / Bodega	28
Italian Restaurant	23
Bar	12
Bus Stop	10
Coffee Shop	10
Chinese Restaurant	9
Caribbean Restaurant	9
Park	8
Pharmacy	8
Others	149

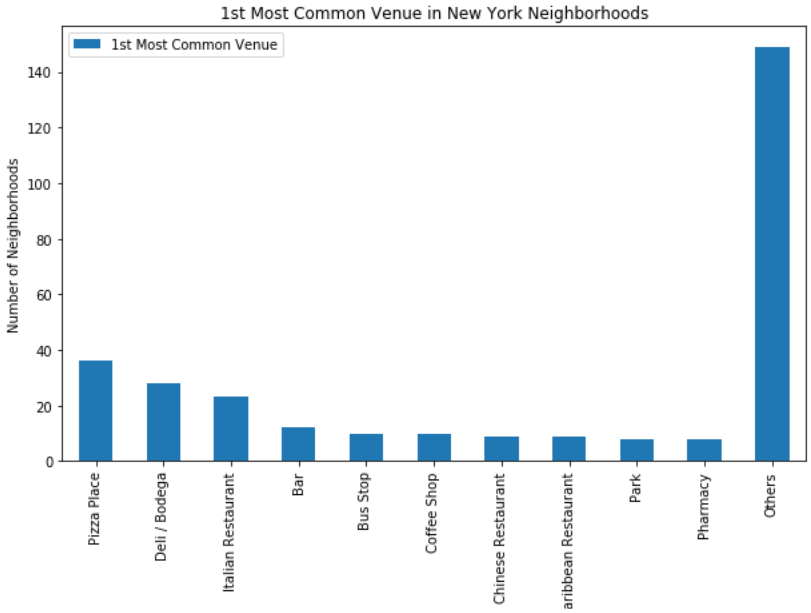


Table 1 and Graph 1. **First Most Common Venue per Neighborhoods.** In the table are indicated the counts of the categories which they are the first most common venue in the neighborhoods of New York.

The same analysis was conducted to the second most popular venue, where also “Pizza Place” and “Deli / Bodega” were very common (Table 2 and Graph 2). “Chinese Restaurant” and “Italian Restaurant” were polular options too.

	2nd Most Common Venue
Pizza Place	20
Deli / Bodega	18
Italian Restaurant	17
Pharmacy	11
Coffee Shop	11
Chinese Restaurant	10
Park	10
Grocery Store	9
Bank	8
Bus Station	8
Others	180

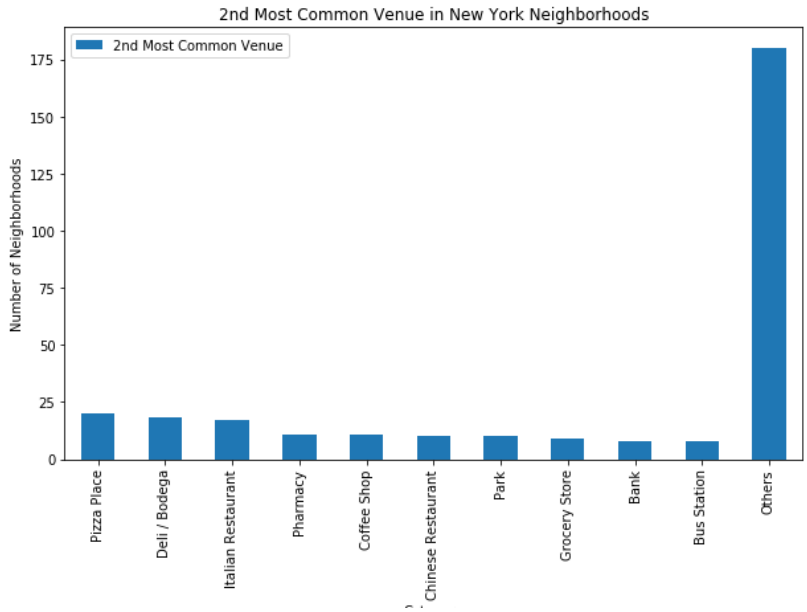


Table 2 and Graph 2. **Second Most Common Venue per Neighborhoods.** In the table are indicated the counts of the categories which they are the second most common venue in the neighborhoods of New York.

The third most common venues in the neighborhoods of New York were also analyzed. “Pizza Place” and “Deli / Bodega” were popular options as in the first and second most common venues tables (Table 2 and Graph 2). But in this third position, several kinds of restaurant are popular places. Categories, such as “Coffee Shop”, “Chinese Restaurant”, “Mexican Restaurant” and “Fast Food Restaurant” are very popular places for this third position.

In these three position there is a “Others” option. Is important to report that in this option were, at least, 60 different categories. Although this option counts for the half of all neighborhoods, the categories are very diverse and could not be a great option to consider them as a popular venue.

	3rd Most Common Venue
Deli / Bodega	18
Pizza Place	18
Coffee Shop	12
Italian Restaurant	11
Bank	10
Donut Shop	9
Mexican Restaurant	8
Fast Food Restaurant	7
Sandwich Place	7
Chinese Restaurant	7
Others	195

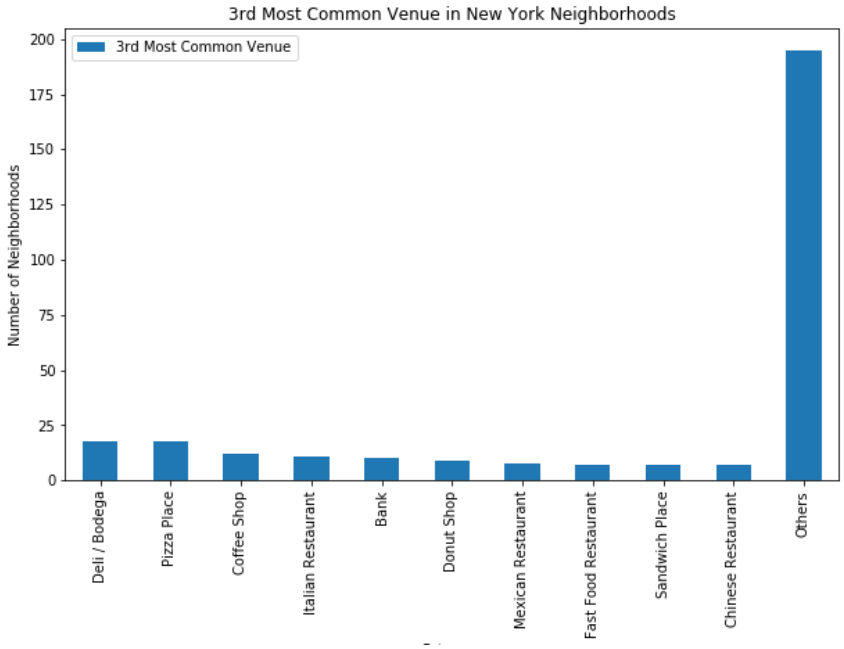


Table 3 and Graph 3. **Third Most Common Venue per Neighborhoods.** In the table are indicated the counts of the categories which they are the third most common venue in the neighborhoods of New York.

The former analysis was conducted in a big and diverse city such as New York. A wider analysis of these venues could be performed in another diverse city in order to complement it. The same process were performed in Toronto, Canada. In the Map 2 is displayed the Map of Toronto and its neighborhoods are indicated in blue points.

Map 2. **Map of Toronto.** In blue points are indicated every neighborhood.

The same analysis performed in New York neighborhoods was perfomed in Toronto neighborhoods. The dataframe griuped and sorted by categories were counted by neighborhood and its counts per category are reported in Table 4 and plotted in Graph 4 as the first most common venue per neighborhood.

The results were similar to the New York fisrt common venue, but here “Coffee Shop” is the most popular venue. Also, “Pizza Place” is a polular venue.

	1st Most Common Venue
Coffee Shop	33
Pizza Place	17
Café	15
Baseball Field	11
Grocery Store	10
Airport Service	7
Playground	6
Pet Store	6
Hardware Store	5
Bank	5
Others	88

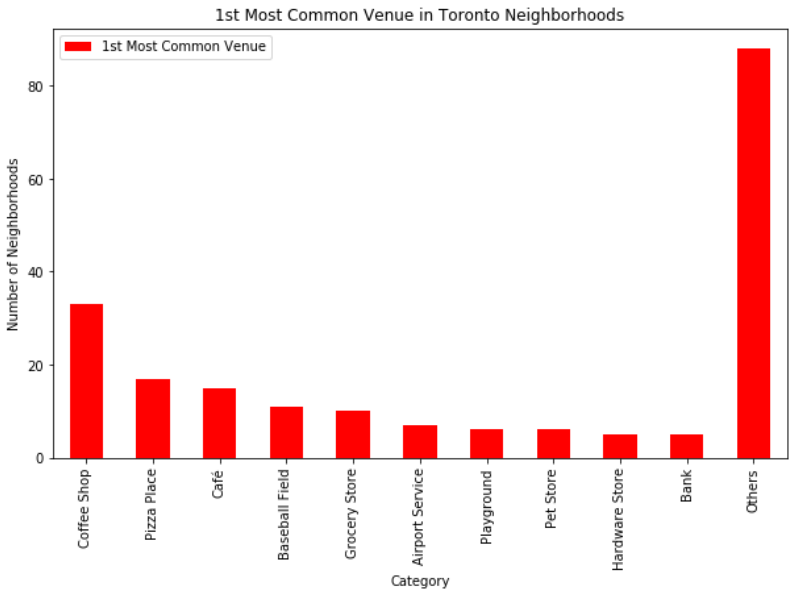


Table 4 and Graph 4. **First Most Common Venue per Neighborhoods.** In the table are indicated the counts of the categories which they are the first most common venue in the neighborhoods of Toronto.

In the second most common venue in Toronto also are “Pizza Place” and “Coffee Shop” (Table 5 and Graph 5). In this group are new categories that are different to New York venues, such as “Women’s Store”. In this section are two open places: “Park” and “Pool” as most common venues.

	2nd Most Common Venue
Coffee Shop	23
Park	20
Café	15
Pizza Place	13
Women's Store	10
Pool	9
Bakery	8
Sandwich Place	7
Airport Terminal	7
Fast Food Restaurant	6
Others	85

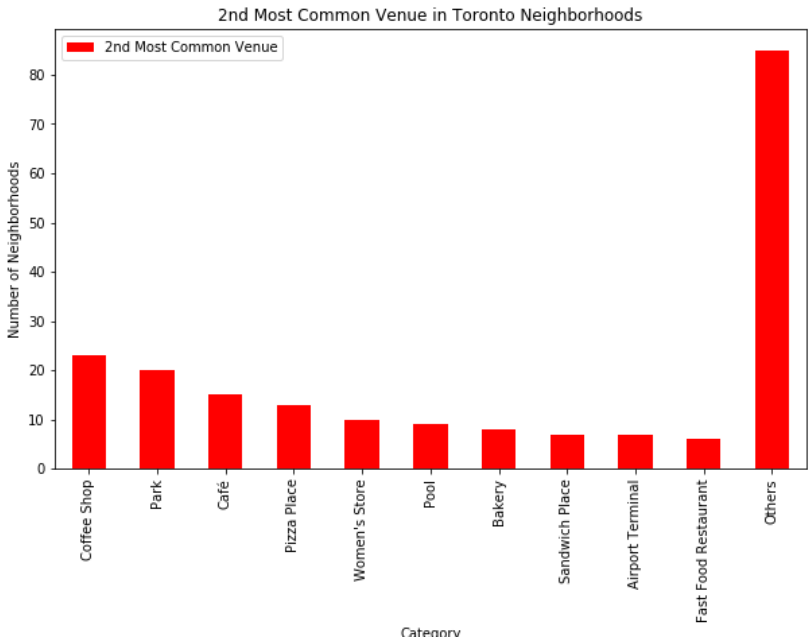


Table 5 and Graph 5. **Second Most Common Venue per Neighborhoods.** In the table are indicated the counts of the categories which they are the second most common venue in the neighborhoods of Toronto.

In the third position are several stores, such as “Women’s Store”, “Convenience Store” and “Discount Store” (Table 6 and Graph 6). As in the case of New York, there is a option “Others” which counts for more or less than the half of total neighborhoods of Toronto. In this option are a lot of categories, for which there are only one nighborhood per category, and would be complicated analyze everyone of them. Compared to New York there are options that are the same, such as “Pizza Place”, “Coffee Store” and “Restaurants”, but in the case of Toronto, there are several important options that are different from New York, such as several kinds of “Stores”.

	3rd Most Common Venue
Women's Store	21
Coffee Shop	20
Convenience Store	10
Park	9
Drugstore	8
Discount Store	8
Fast Food Restaurant	7
Hotel	7
Airport Lounge	7
Supermarket	6
Others	100

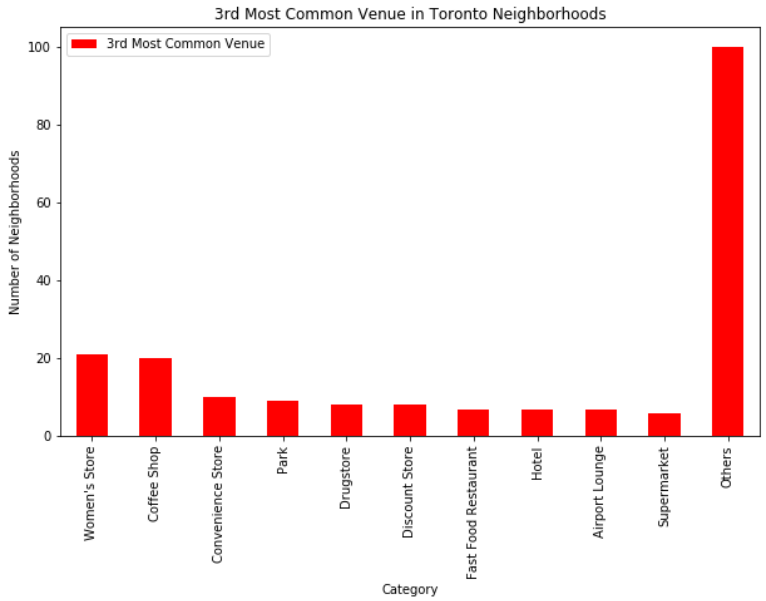


Table 6 and Graph 6. **Third Most Common Venue per Neighborhoods.** In the table are indicated the counts of the categories which they are the third most common venue in the neighborhoods of Toronto.

Now that we know the categories that the most popular venues belong to, it is time to analyze if there are some kind of competence in the neighborhoods. In other words, we want to know if two or more popular venues compete with each other in a certain neighborhood. Let me clear this with a hypothetical example: in any of the two cities, in a given neighborhood, we know the 20 most popular venues; in this list may be two venues which belong to the same category, such as “Pizza Place”. These two venues are “Mario’s” and “Luigi’s”, respectivelly. Perhaps “Mario’s” is the first most popular venue in its neighborhood, and “Luigi’s” is the third most popular venue. This indicate a strong competition between them. So, if someone is wondering if would be a good idea to start a new “Pizza Place” in that neighborhood, would be a complicated scenario because of the future competence with “Mario’s” and “Luigi’s”.

The data recolected from FourSquare were sorted and grouped by neighborhoods. Were grouped in a new data frame the first 20 most common venues and their category for every neighborhood. Then, was calculated the total number of different categories in these 20 venues for every neighborhood.

Surprinsigly, in all of the neighborhood of New York and Toronto, all categories for the first 20 most common venues were different (Table 7 and Table 8). This result indicates that there are not two or

more most popular venues in the first 20 places. In our hypothetical example this means that only "Mario's" controls all the "Pizza Place" market, and "Luigi's" have no place as popular venue.

	Unique Categories	Neighborhoods
0	20	Allerton
1	20	Annadale
2	20	Arden Heights
3	20	Arlington
4	20	Arrochar
5	20	Arverne
6	20	Astoria
7	20	Astoria Heights
8	20	Auburndale
9	20	Bath Beach

	Unique Categories	Neighborhoods
0	20	Adelaide
1	20	Agincourt
2	20	Agincourt North
3	20	Albion Gardens
4	20	Alderwood
5	20	Bathurst Manor
6	20	Bathurst Quay
7	20	Bayview Village
8	20	Beaumond Heights
9	20	Bedford Park

Table 7 and Table 8. **Different Most Popular Categories per Neighborhood.** In the table are indicated the counts of the categories which they are the first 20 most popular venues per neighborhood from New York (left) and Toronto (Right). Only the first 9 neighborhoods are displayed.

4. Conclusions.

The first three places of most common venues where similar to each other.

In these venues, the most common categories were Pizza Place, Coffee Shop and several kinds of restaurants.

The 20th most common venues belong to unique categories.

This indicates a complementation instead of competition.

5. Future Directions.

With the results reported here, someone who desires to start a new small business in a given neighborhood, could look for, at least, the 10 most popular venues in order to discard some ideas. This results also indicates a pattern of ten most common venues. It is clear that start a new “Pizza Place” in both cities would be very difficult due to the competence. But if in the neighborhood there is no “Chinese Restaurant” as popular venue, would be a great idea to start one, studying the service and food quality of other chinese restaurant for reveal why they are not a popular venue and apply the correct keys to our business.

6. References.

[1] U.S. S.B.A (2018) “Frequently Asked Questions about Small Business”, August 2018, U.S. S.B.A. Office of Advocacy. Avaliable at https://www.sba.gov/sites/default/files/advocacy/Frequently-Asked-Questions-Small-Business-2018.pdf

[2] Skripak, S. J. (2016). “Fundamentals of Business”, Virginia Tech. Avaliable at https://vtechworks.lib.vt.edu/handle/10919/70961