

# Data Science Capstone Project

Finding the best location for a bar in the city of Toronto

## Business Understanding

When opening a bar, of course, there is always the question of the ideal location. There are two problems in particular: First, the location must be chosen so that there are as many potential customers as possible. On the other hand, there should not be too many competing bars in the immediate vicinity.

For an entrepreneur, it is therefore of the highest interest to choose the location for the new opening in such a way that as many customers as possible, and thus as high a return as possible, can be brought in.

In this project, this question will be answered using data on the example of the city of Toronto. For this purpose, each neighborhood will be evaluated individually and a recommendation for suitable neighborhoods will be given.

Since bars are often visited after a visit to a restaurant and a high number of restaurants usually corresponds to a high number of walk-in customers, both the number of restaurants and the number of bars are included in the evaluation.

## Data Understanding

Since we want to evaluate the question of the ideal place to open a bar for each neighborhood in Toronto individually, it is important to get an overview of all neighborhoods. On Wikipedia there is an overview of all postal codes with the corresponding neighborhoods. We use this page to generate a complete list of all neighborhoods in Toronto.

As a second step, we use the API of Foursquare to get a list of restaurants and bars for each neighborhood. We would like to select neighborhoods with a high number of restaurants, as these also correlate with high foot traffic. On the other hand, there should be comparatively few bars in the neighborhood.

The number of restaurants and bars are then used to get a closer selection for suitable neighborhoods.

The dataset then contains a list of all neighborhoods with their coordinates and the number of restaurants and bars.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Restaurent Count	Bar Count
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	24	12
1	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	3	1
2	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494	31	9
3	M3B	North York	Don Mills	43.745906	-79.352188	5	1
4	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	5	2

## Methodology

For analysis, the data are first clustered into 3 groups in terms of the number of restaurants in each neighborhood. Thus, we obtain neighborhoods with a 'high', a 'medium' and a 'low' number of restaurants.

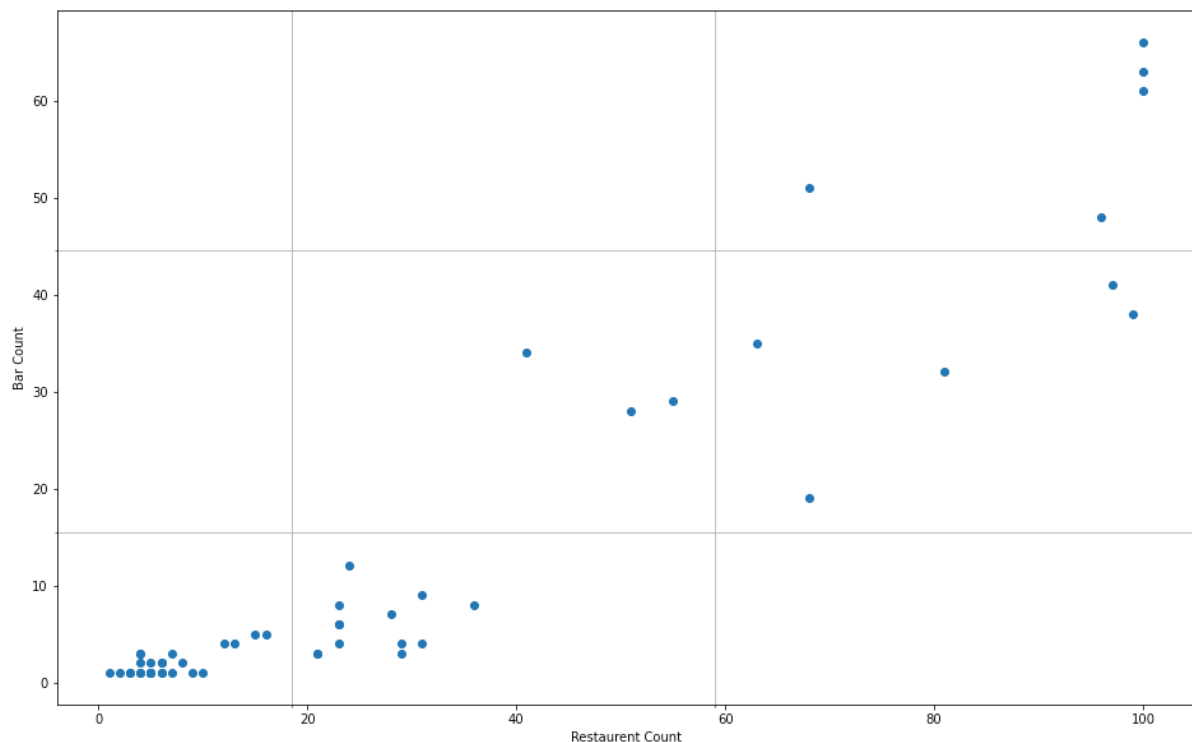
As an algorithm, k-means clustering is used, since we already have a precise idea of the number of groups.

In the same way, clustering is then performed with respect to the number of bars.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Restaurent Count	Bar Count	Restaurent Occurance	Bar Occurance
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	24	12	Medium	Low
1	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	3	1	Low	Low
2	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494	31	9	Medium	Low
3	M3B	North York	Don Mills	43.745906	-79.352188	5	1	Low	Low
4	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	5	2	Low	Low

Ideally, we would like to find neighborhoods where there is a high number of restaurants, but only a low number of bars.

A first scatterplot shows that basically the number of restaurants and the number of bars in a neighborhood are correlated.



## Results

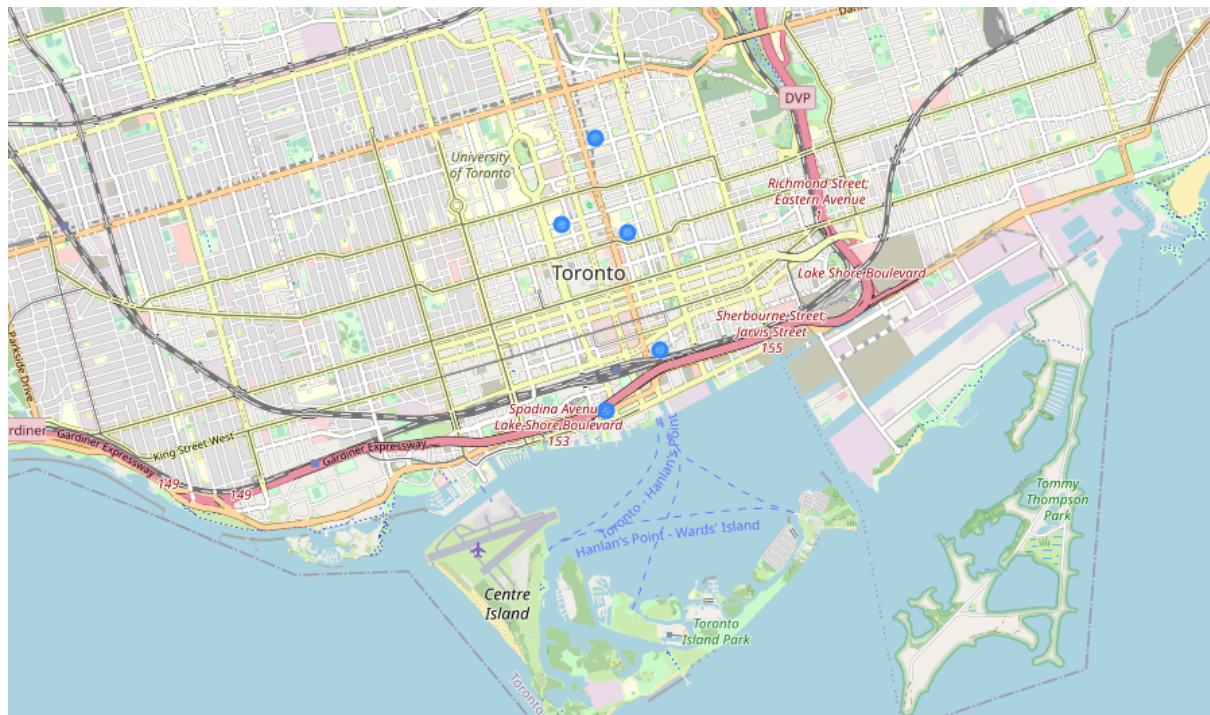
As discussed in the previous section, the scatter plot shows a correlation between the number of restaurants and the number of bars. This is not particularly surprising and was to be expected.

However, we would like to select here the neighborhoods that show a comparatively good relationship between the number of restaurants and bars.

		Neighborhood
Restaururent Occurance	Bar Occurance	
High	High	5
	Medium	5
Low	Low	27
Medium	Low	13
	Medium	3

It turns out that there is no neighborhood with a high number of restaurants and a low number of bars.

However, there are 5 neighborhoods where there is a high number of restaurants and only a moderate number of bars. These neighborhoods are marked on the map below:



## Discussion

With these 5 neighborhoods, there is one neighborhood that stands out. This is the neighborhood "Central Bay Street" with 68 restaurants and only 19 bars.

Based on the data, we would suggest this neighborhood as the appropriate place to open a new bar.

## Conclusion

In this report, we used data from Foursquare to identify an ideal place to open a new bar in Toronto.

To do this, we determined the number of restaurants and bars for each neighborhood and then divided them into categories. There are a total of 5 neighborhoods with a high number of restaurants and only a moderate number of bars. Of these 5 neighborhoods, "Central Bay Street" in particular stands out.

We recommend this neighborhood as the most suitable place for the new opening.