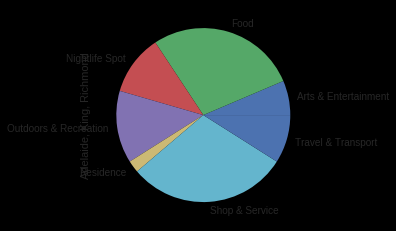
**Introduction/Business Problem** .

A company want to relocate from Toronto to New York. It wants to find a similar borough by venue composition. In other words, we don't use the venue number directly, in stead we use the percentage contribution of categories , say Arts & Entertainment, Food, Nightlife Spot, Outdoors & Recreation, Residence, Shop & Service and Travel & Transport.



**Data**

Toronto Data: Use the Notebook to build the code to scrape the following Wikipedia page, https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M.

We get the Toronto’s information from the previous exercise.

New York Data: The link to the dataset of New York: https://geo.nyu.edu/catalog/nyu\_2451\_34572

Foursquare Data: We get the venues and their categories data from Foursquare.com .

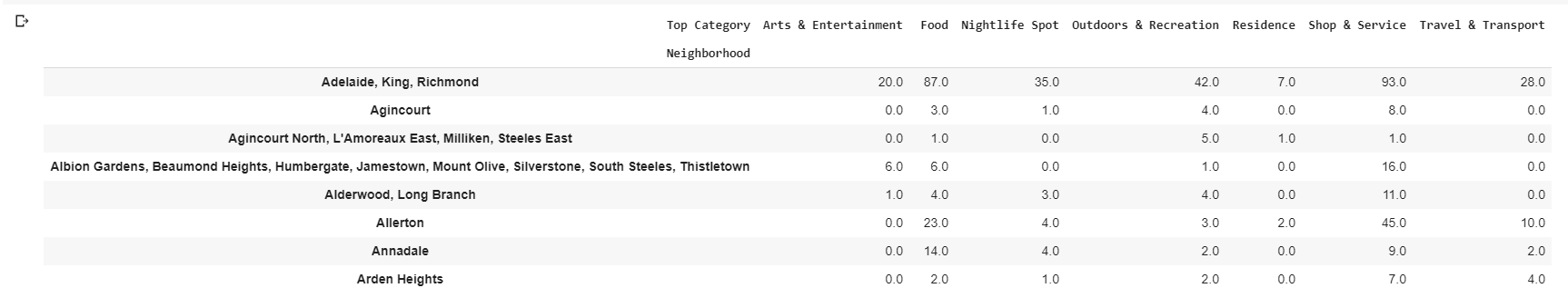
The primary challenge is to fetch data from Foursquare. For every call, there is a limit of 50 venues to return. So in order to fetch more data, If we make venue search calls by category. If we fetch data for all categories, the dataset could have millions of rows. But due to the calle limit, It definitely fail for every time. So we categorize the data into only 7 top categories. As a balance, we fetch data by 7 primary categories. And in order to limit the results per call, we choose a small radius.

In order to break the call limit, I applied 5 Foursquare accounts to download data.

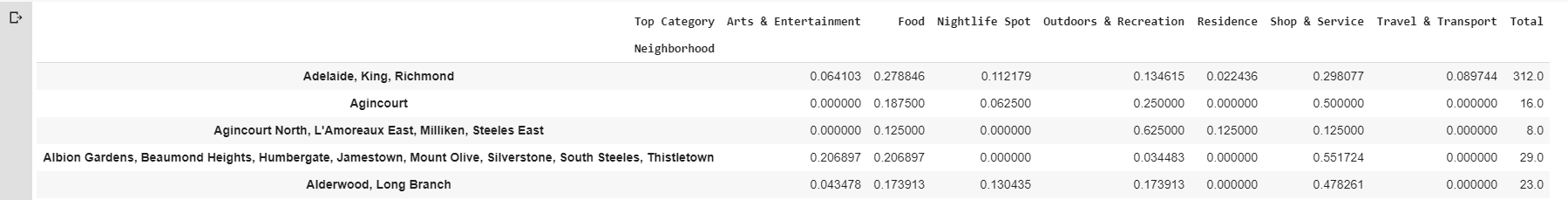
**The Method**

We collect venue data from Foursquare.

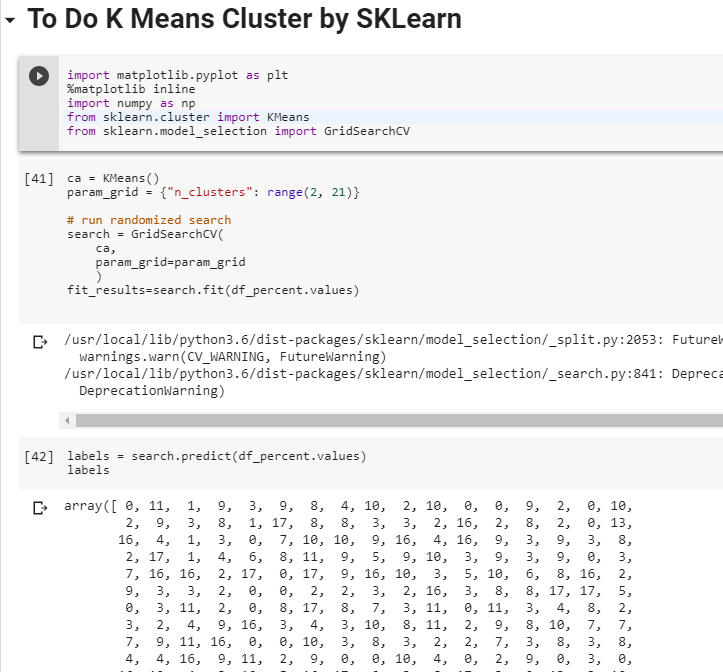
We statistics the venue data and form the pivot table as the following:



We use category count per borough/ Total venues per borough to make the data percent based.



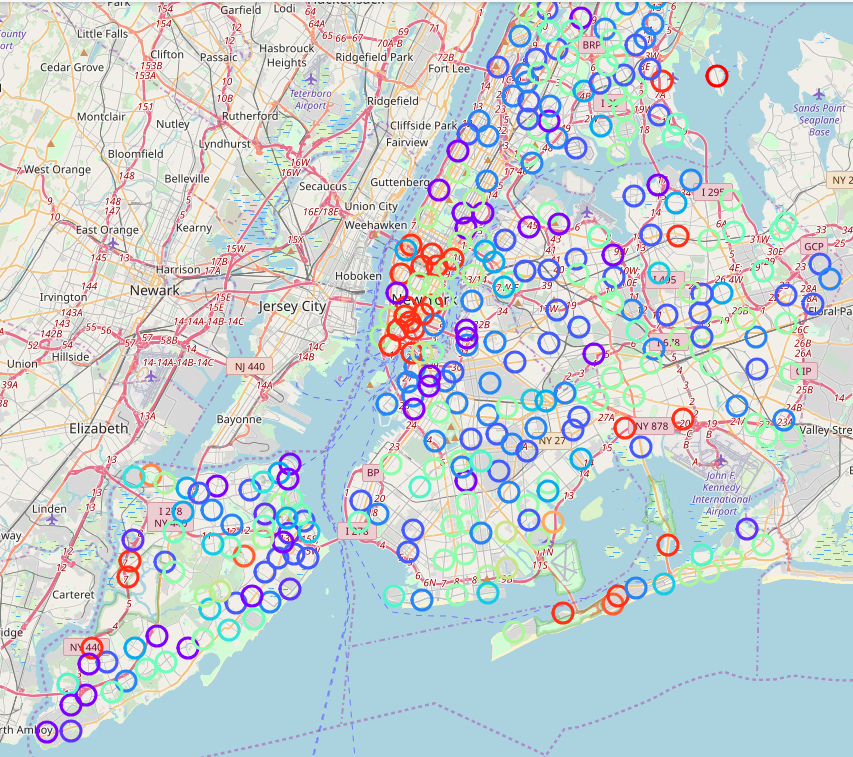
We use the K means cluster of SKLearn to cluster the boroughs. In order to find the best K parameter. We will use GridSearch.



**The Results**

So we can cluster the boroughs of Toronto and New York together.



Visually, the cluster of New York is as the following:

For the company knows its location cluster, it can easy find the area that have the same cluster in New York.