|  |
| --- |
| Restaurant recommender system |
|  |

|  |  |
| --- | --- |
| 07/13/2020 | IBM Capstone project |

Restaurant recommender system is a machine learning model, developed to demonstrate as a capstone project to IBM through coursera. It recommends restaurants based on user’s likes and dislikes and his previous interest data. **Table of contents**

|  |  |
| --- | --- |
| **Topic** | **Page number** |
| Introduction section | 2 |
| Data section | 4 |
| Methodology section | 5 |
| Result section | 6 |
| Discussion section | 7 |
| Conclusion section | 7 |

* **Introduction :**

**Problem background:**

New York City, often called simply New York and abbreviated as NYC, is the [most populous city](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population) in the [United States](https://en.wikipedia.org/wiki/United_States). With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2), New York is also the [most densely populated](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population_density) major city in the United States.

Problem description:

Suppose I travel and keep changing places very frequently. This is very hectic and plus i get to experience very different types of environment, of which I do not have much knowledge about. In such situation, food can be an important factor for decided how you rate your trips and plus also recommending it to the people. Food can also attract people around to world to try it out if it were to be the best. In such scenarios, we need to find the right place, at reasonable cost, to serve us the best possible way. So there are few questions that must be addressed, such as:

* How many types of foods are available in the restaurant?
* Which is the nearest to me with good rating?
* How many "similar" restaurants are available nearby me?
* Can we find any veg restaurant?

To address such question, XXYZ company's manager decides to allocate this project to me not just to find out solutions to the questions but also build a system that can help in recommending new places based on their rankings compared to the previously visited by me.

Expectations from this recommender system is to get answer for the questions, and in such a way that it uncovers all the perspective of managing recommendations. It is sighted to show:

* What types of restaurants are present in a area?
* Where are the similar restaurant present based on a preference to particular food?
* How do different restaurants rank with respect to my preferences?

Target audience:

Target audiences for this project does not limit to a person who keeps travelling but everyone. People could simply decide to look for a similar restaurant all the time because they are addicted to a specific category of food. People who rarely use restaurants would prefer to have the most rated restaurants nearby them and all this could be easily handled by our recommender system. So target for this project is basically everyone who is exploring different places or similar places.

Success rate:

With restaurants evolving, new food categories emerge, hybrid food starts to be more popular, we need a system that could help us access vast number of food varieties. It is impossible for a person to ask each and everyone about their visit to a particular place and also not everyone remembers everything. On the other hand, Computers are good at remembering things, and with Machine learning to its peak, it high time technology will by our personal guidance and help us personally based on our likes and dislikes. So people would care about this project as their personal assistance and success rate could certainly increase with time.

* **Data :**

**Data requirements:**

To find a solution to the questions and build a recommender model, we need data and lots of data. Data can answer question which are unimaginable and non answerable by humans because humans do not have the tendency to analyze such large dataset and produce analytics to find a solutions.

Let's consider the base scenario:

Suppose I want to find a restaurant, then logically, I need:

Its geographical coordinates(latitude and longitude) to find out where exactly it is located.

To access location of a restaurant, its Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the restaurants with its labels respectively.

**Data collection :**For this project I used this dataset of New York City

[https://geo.nyu.edu/catalog/nyu\_2451\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572#)

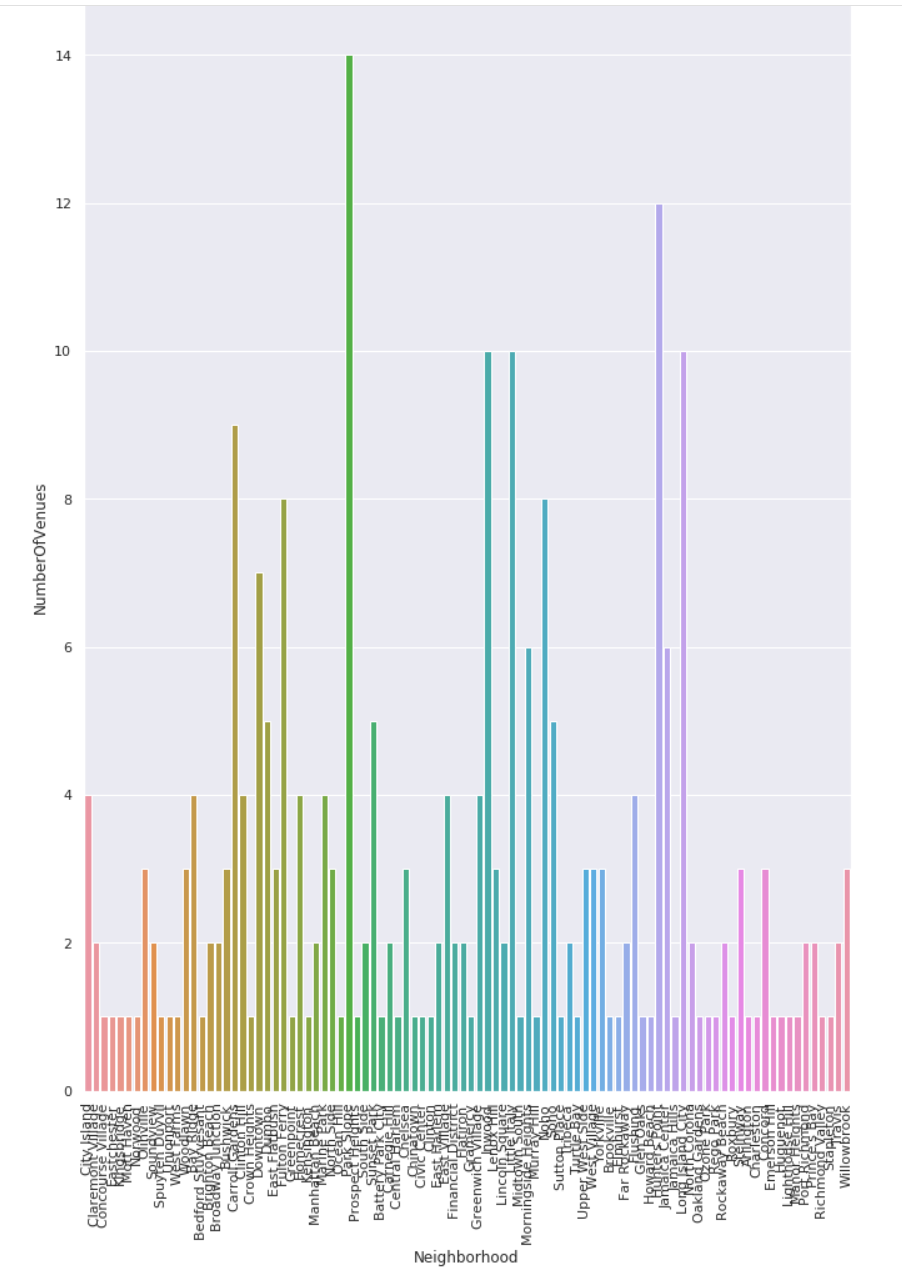
* **Methodology :**

**Exploratory analysis:**

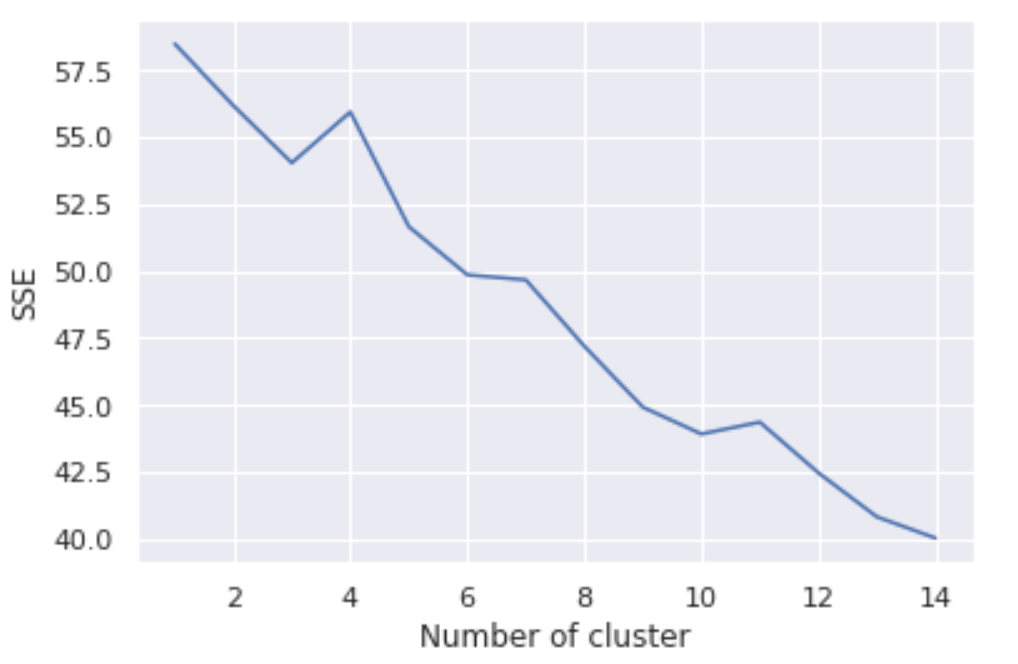
Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched.

Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data.

While exploring the dataset, I found out that Brooklyn, Park Slope has most number of venues.



Also while producing graph for number of cluster, I produced a graph to explore all the values for n\_clusters and then finding the best by exploring the elbow graph.



* **Result :**

The result of the recommender system is that it produces a list of top restaurants and the most common venue item that the user can enjoy. During the runtime of the model, a simulation was done by taking ‘West Village’ as the neighborhood and then processed through our model so that it could recommend neighborhoods with similar characters as that of ‘West Village’.

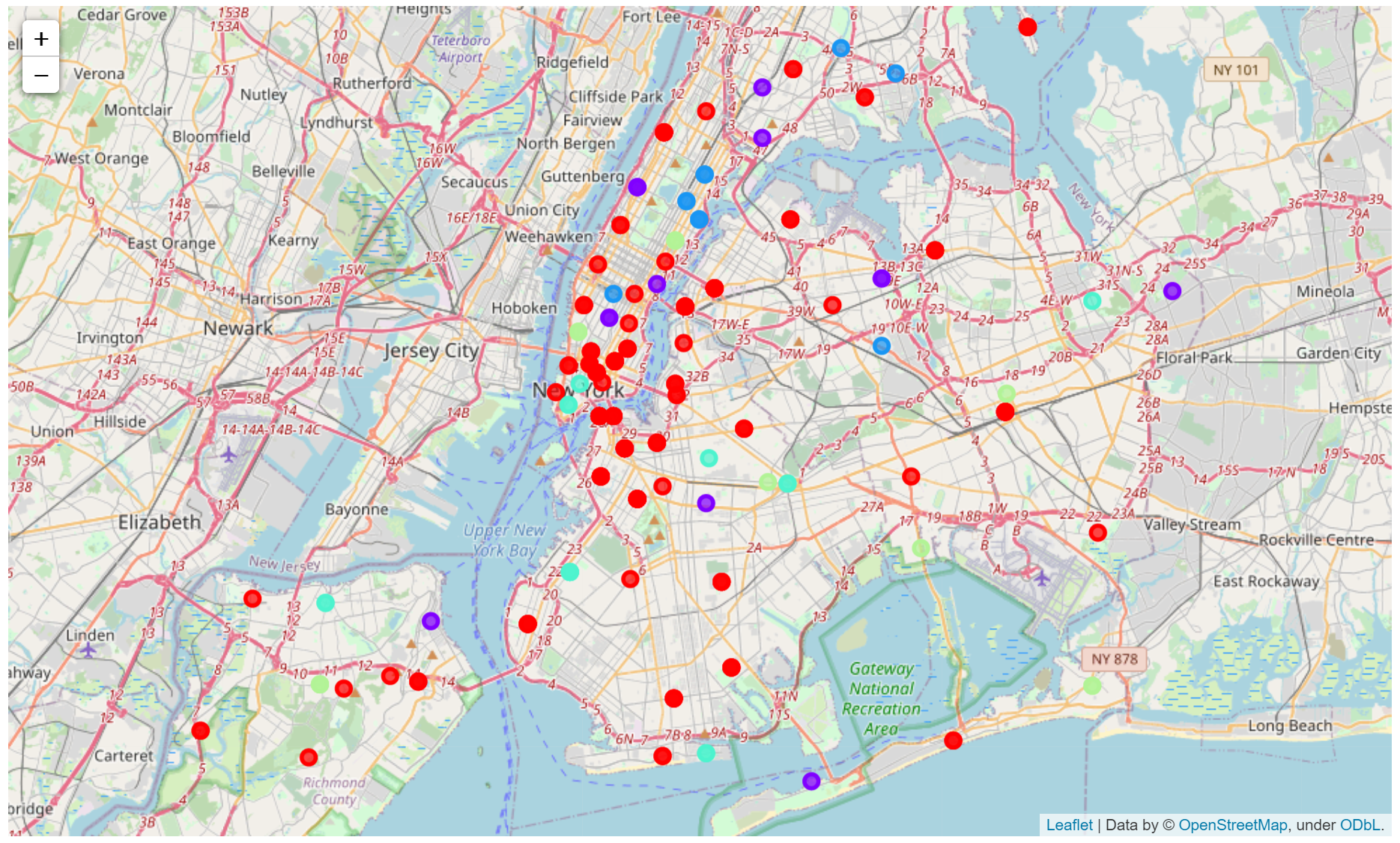
The following image shows the result:



* **Discussion :**

Also during clustering, similar neighborhoods must be dumped into the right cluster.

The following graph shows the clusters:



Another observation that we can make is that choosing number of clustering could produce very diverse results. Some may be over fitted or some may be under fitted. Hence analysis of number of clusters must be done. Ref elbow\_graph in the Methodology section.

* **Conclusion :**

The recommender system is a system that considers factors and makes use of Foursquare API to determine nearby venues. It is a powerful data driven model whose efficiency may decrease with more data but accuracy will increase. It will help users to finish their hunger by providing the best recommendation to fulfil all their needs.