Coursera Capstone Project

**A. Introduction**

**A.1. Description & Discussion of the Background**

Toronto is capital of the province of Ontario, southeastern Canada. It is the most populous city in Canada, a multicultural city, and the country’s financial and commercial center. Its location on the northwestern shore of Lake Ontario, which forms part of the border between Canada and the United States, and its access to Atlantic shipping via the St. Lawrence Seaway and to major U.S. industrial centers via the Great Lakes have enabled Toronto to become an important international trading center. This city is the Top in the country for Small Business.

The City of New York, usually called either New York City (NYC) or simply New York (NY), is the most populous city in the United States and thus also in the state of New York. New York is also the most densely populated major city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. The city's fast pace has inspired the term New York minute. Home to the headquarters of the United Nations, New York is an important center for international diplomacy.

An Indian restaurant owner who has multiple restaurant in Toronto is planning to expand restaurant to New York. Seeing the potential of New York, he decides to open a new restaurant in this city. We need to find: Where should the new restaurant be located?

The main purpose of this report is showing a Data Science approach to solve that question.

**A.2. Data Description**

Based on the purpose of the project, New York City neighborhoods we collect the data using FourSquare API which provides the surrounding venues of a given coordinate.

The process of collecting and clean data:

* Find the geographic data of the New York.
* I used **Forsquare API** to get the most common venues of given Borough for the restaurant in New York.
* Count the occurrence of each venue type, then apply one hot encoding to turn each venue type into a column with their occurrence as the value.

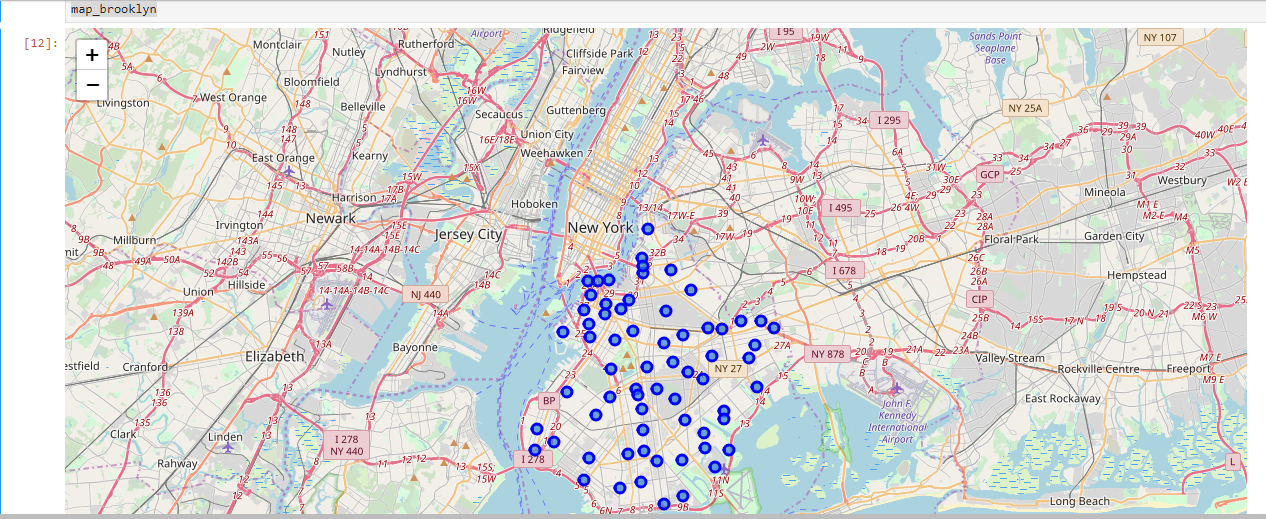
**B. Methodology**

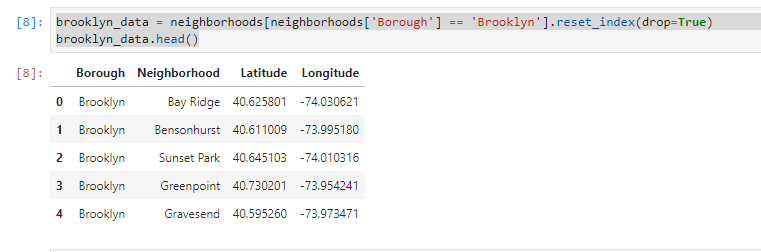
The assumption is that a neighborhoods in New York is a good place to open a new restaurant. Thus, the clustering technique will be used to analyze the dataset. In the end, we will find good neighborhoods in New York which suits the client’s purpose.

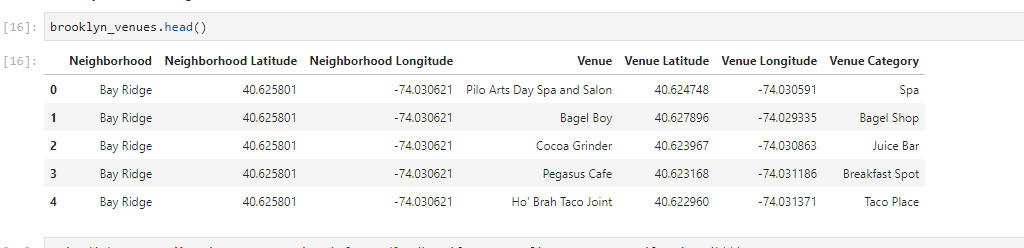
Python data science tools will be used to help analyze the data.

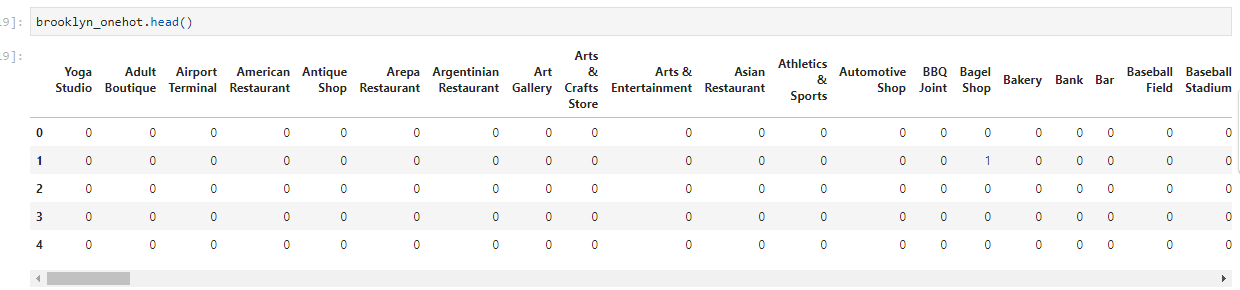
**Map the Neighborhood where specific restaurant is common**

Map of Brooklyn, New York**:**





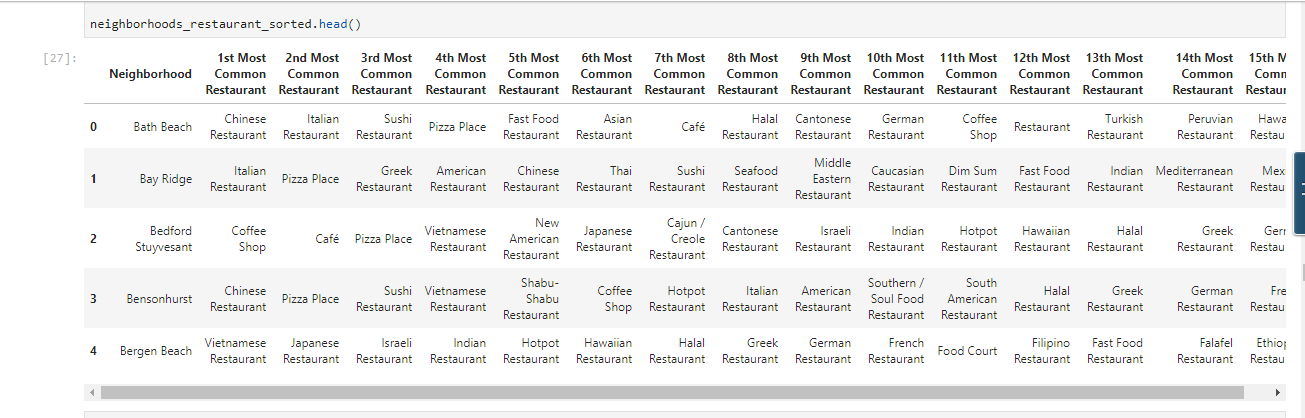


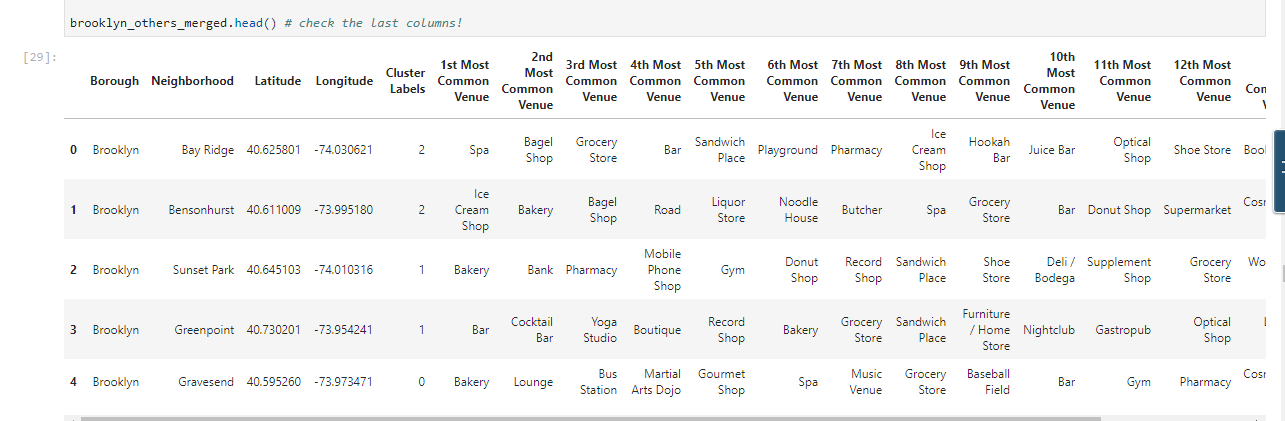


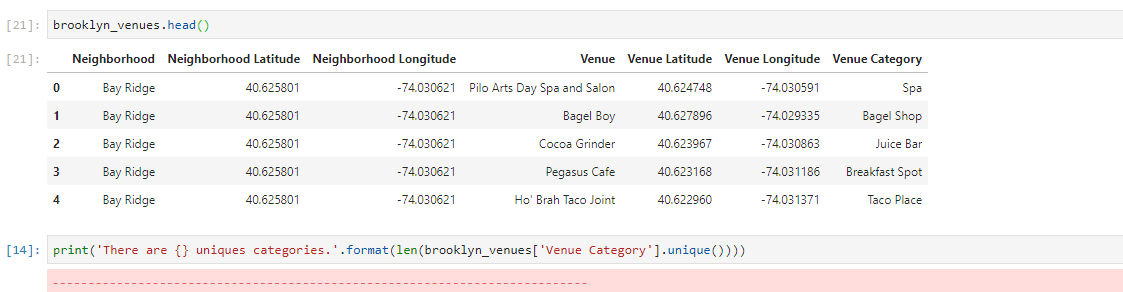


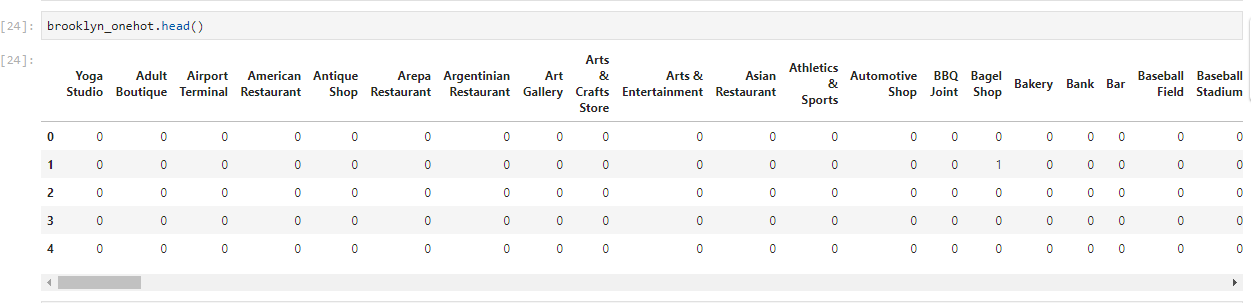


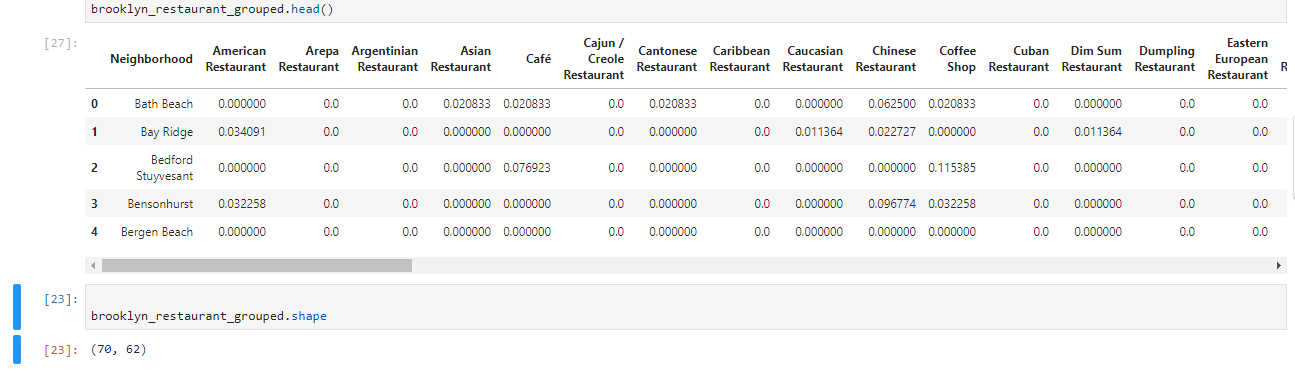


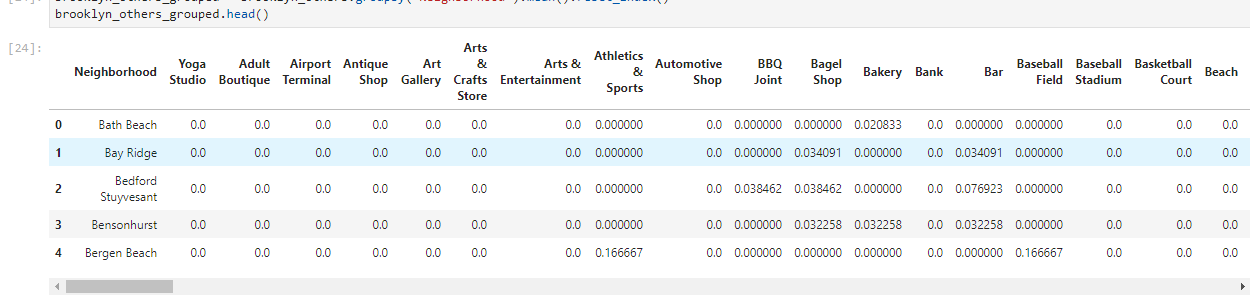






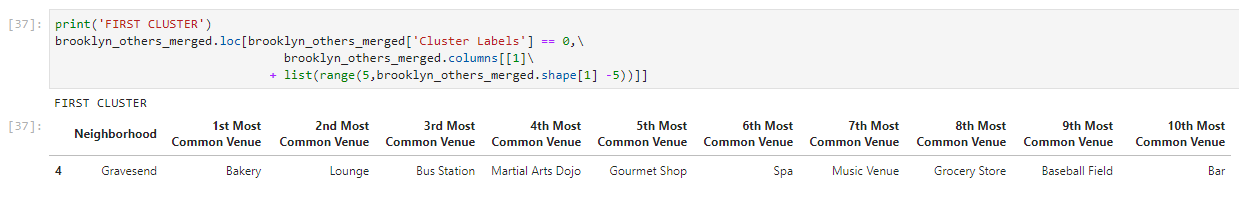


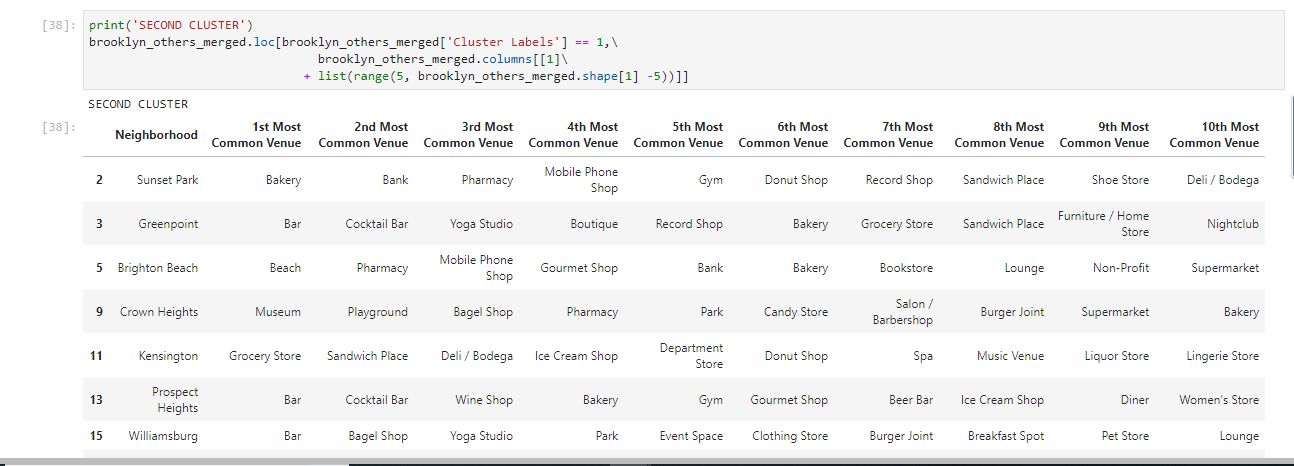


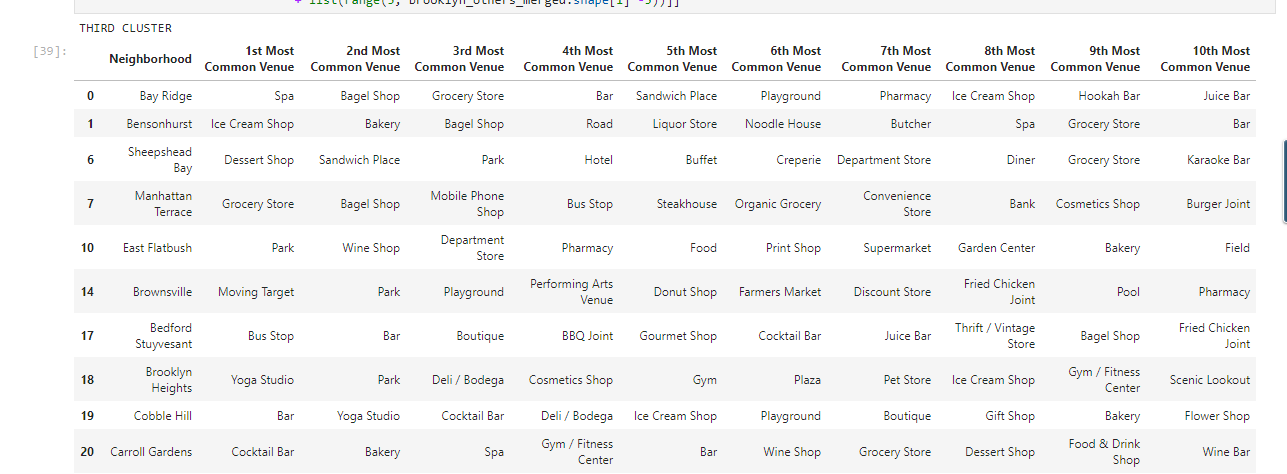


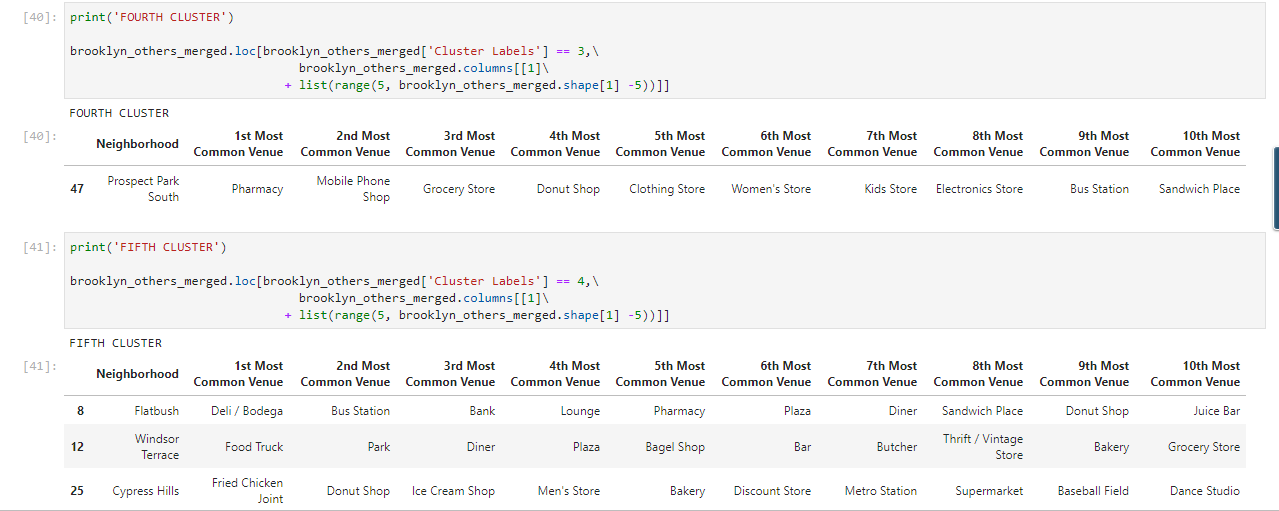


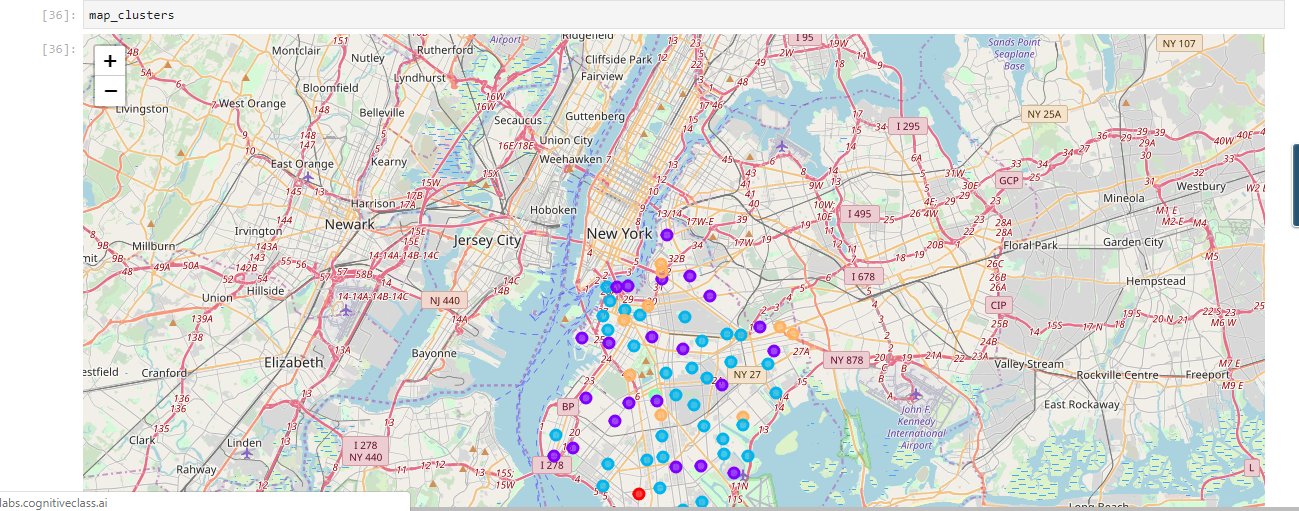
* Venue clustering by neighborhood and analysis of 'best' fit for new location











**C. Results**

* The top venues provide an idea of how diverse the Brooklyn area is in New York City.
* The clustering provides an insight of the similarities in the neighborhoods.

**D. Conclusion**

We can conclude that using Foursquare for location data and combining machine leaning algorithms we can design/develops a system with will be useful for the small Business owners in the city to open a new restaurant.