Recommend a suitable location for a restaurant in New York state.

KANNA VENKATA RAGHAVENDRA KARTHIK

Introduction

- People love to eat good food and taste various delicacies.
- Different cuisine restaurants have provided this need.
- Restaurants have always played a crucial role in business, social life of the society.
- Many personal and professional events happen at restaurants.
- Location of the restaurant is one of the major contributing factors for a successful restaurant.

Business Problem

Consider the following hypothetical scenario :

‘Resto-grand’ a successful multi-chained restaurant company has its brands around Europe, India, California. They wanted to start their restaurants in New York state and required some suggestions on where to open them.

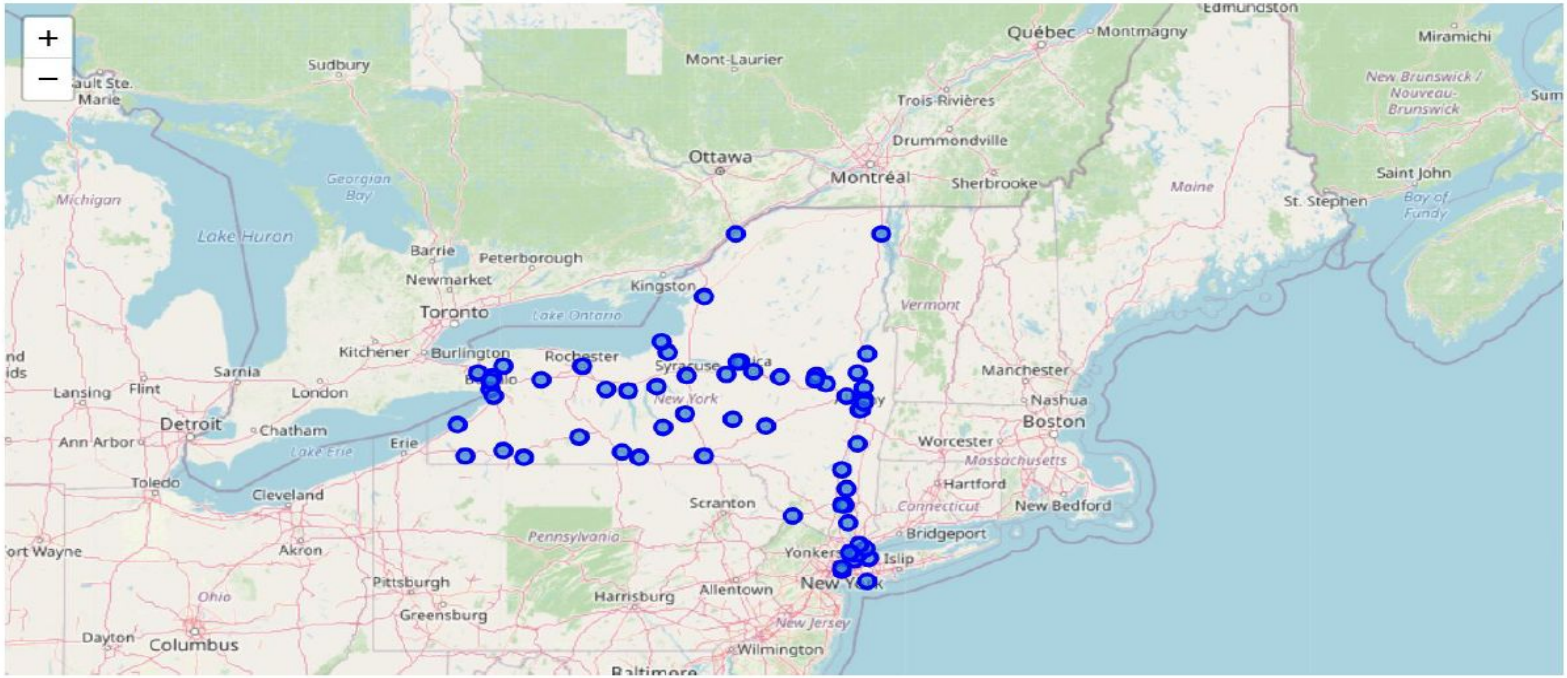
The aim of the project is to find suitable locations for the restaurants in New York state.

Data Acquisition and Cleaning

- List of cities present in New York state can be obtained from wikipedia's page : (https://en.wikipedia.org/wiki/List_of_cities_in_New_York)
- Latitude and Longitude locations of them can be found using python's geopy package.
- We will use foursquare.com's API to find the most common venues present at given geographical coordinates.
- Remove unnecessary columns in the dataset and ensure the dataset has data about city, county, population, latitude and longitude columns.
- The dataset at this stage is of shape (62,5) where each row represent a city the New York state with data of above mentioned 5 columns.

Geographical location of 62 states

Using folium, a map rendering software we shall just see the location of cities on the world map. The blue points represent our 62 cities in New York state.



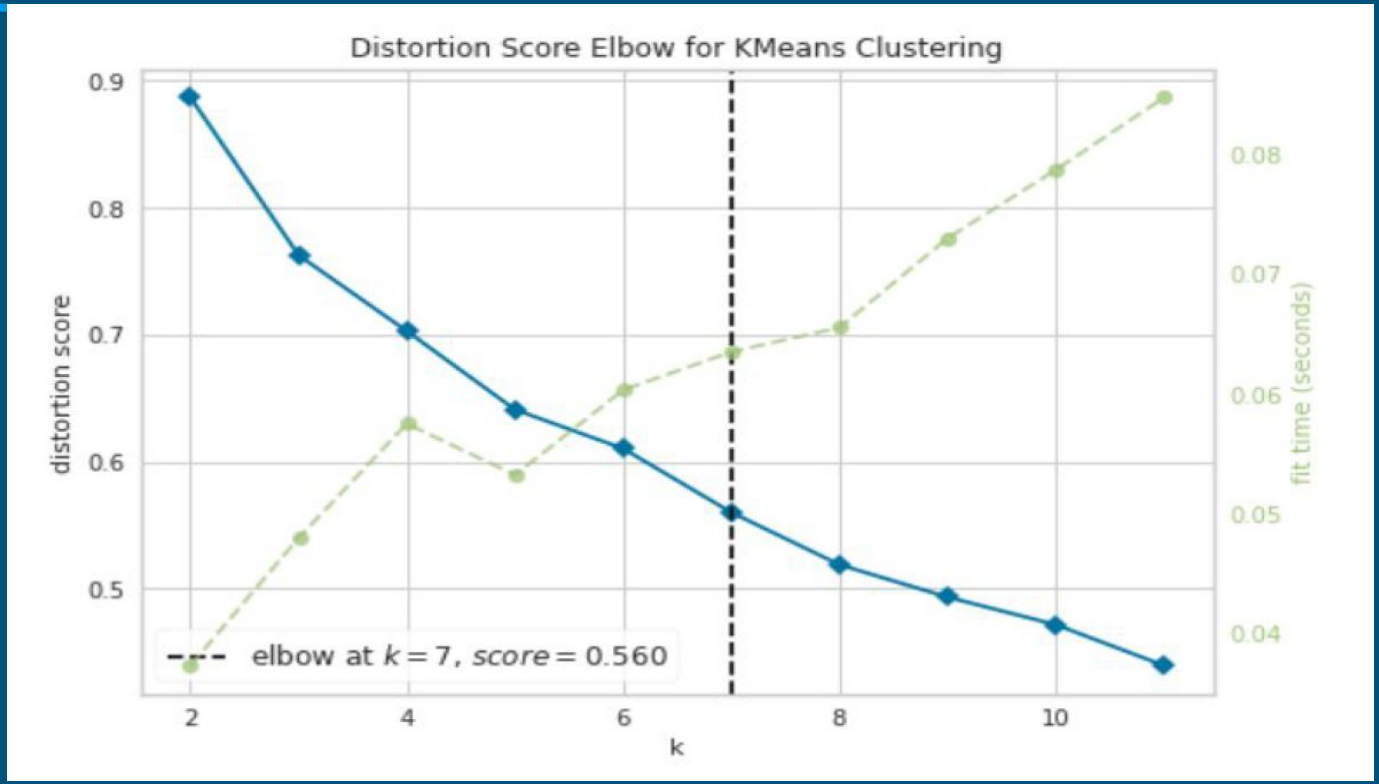
Methodology

- Using foursquare's API obtaining 100 venues for Amsterdam city within 30 km radius as part of data exploration to see how the data is received.
- Later generalize this for all the cities of New York state.
- Create a dataframe having venue's name, category, latitude and longitude of each city.
- Received 293 unique venue categories across 62 cities.
- Perform one-hot encoding for venue category and create a dataframe with city and the frequency of venue category present in it.
- Sort it so that top 10 most common venues of each city are present in a row.

Clustering

- Use the unsupervised KMeans Clustering algorithm to model the data.
- This arranges the similar cities into a cluster and dissimilar cities in different clusters.
- This is done by importing from sklearn library.
- For this algorithm we need to pass a parameter called number of clusters.
- Optimal value for number of clusters (k) to be found.
- We use elbow method to find this.
- Make use of KElbowVisualizer in yellow brick library.

Optimal value of K is 7



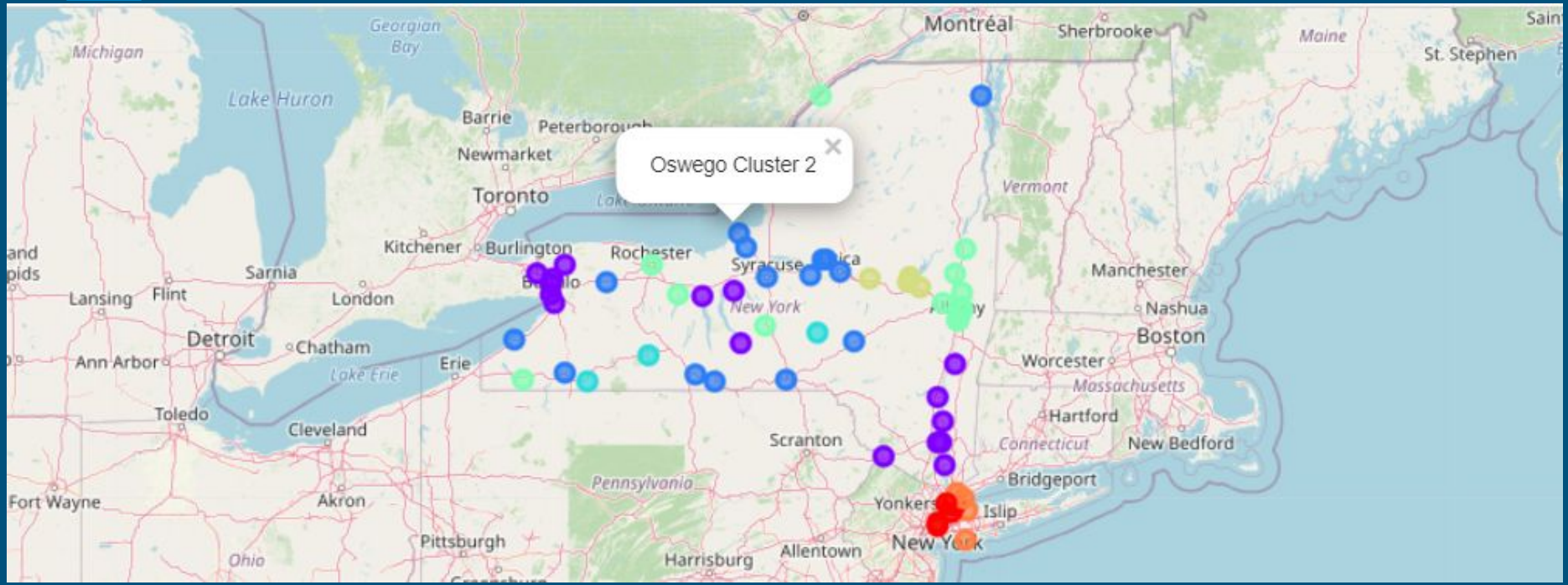
Results

- Cluster 0 contains cities where parks, bakeries, gyms are priority in the top most venues.
- Cluster 1 contains the cities where cafes, bars, restaurants, ice cream shops are priority in the top most venues.
- Cluster 2 contains the cities where many restaurants, pizza places and bars are priority in the top most venues.
- Cluster 3 contains the cities where discount stores, sandwich and pizza places are priority in the top most venues.
- Cluster 4 contains the cities where coffee shops, pizza places , ice cream shops are priority in the top most venues.
- Cluster 5 contains the cities where convenience stores, sandwich places and restaurants are priority in the top most venues.
- Cluster 6 contains the cities where pizza places and bakeries are priority in the top most venues.

Discussions and recommendations

- The cities from Cluster-2 are most suitable location for restaurants.
- We advise the Resto-grand’s business owners to consider these.
- These are the cities that are well represented by the restaurants, other bar and pizza places.
- This shows there is very high demand in the locality for the restaurant and in turn can be successful provided proper ambience and taste of the food.

Blue circles are our restaurant recommendation choice



Conclusion

- We discussed the process of solving our problem of finding suitable locations for opening a restaurant in New York state.
- Python libraries such as pandas, scikit-learn, folium, yellowbrick to name a few.
- Used the foursquare api to find the popular venues and its categories present in the city within a 30km radius.
- KMeans unsupervised machine learning algorithm was used to group these cities into clusters after knowing the top most venue category in each city.
- Cities from Cluster 2 are the best suited for a restaurant.