Introduction/Business Problem

Starting a new business in a city is always a challenging one. That too if we need to start a business in a city like New York, it will be more difficult. Machine learning models come as a big help in these scenarios where a customer needs a proper insight into the data of existing business in his location of preference.

This project is intended for a customer who is planning to start a new restaurant in the boroughs of New York city. It will analyze the existing restaurants in the the New York city boroughs of Bronx, Manhattan, Brooklyn and Queens. This will give an insight into the restaurants types like Italian, Chinese etc. which are already there in these boroughs in abundance and those which are very less. Based on this analysis the customer can conclude on the type of Restaurant he can start in these areas.

Data

Our dataset contains information on the Boroughs in the New York city. This is taken from the website https://geo.nyu.edu/catalog/nyu_2451_34572 . The json file available from this website will give information on the Boroughs and their latitude and longitude. This website is maintained by New York University Library which has Spatial data repository of New York city.

▼ root: {} 5 keys
    type: "FeatureCollection"
    totalFeatures: 306
  ▼ features: [] 306 items
    ▼ 0: {} 5 keys
        type: "Feature"
        id: "nyu_2451_34572.1"
      ▼ geometry: {} 2 keys
          type: "Point"
        ▼ coordinates: [] 2 items
            0: -73.84720052054902
            1: 40.89470517661
        geometry_name: "geom"
      ▼ properties: {} 8 keys
          name: "Wakefield"
          stacked: 1
          annoline1: "Wakefield"
          annoline2: null
          annoline3: null
          annoangle: 0
          borough: "Bronx"
        ▶ bbox: [] 4 items
    ▶ 1: {} 5 keys

 We have also used the Foursquare API to explore the boroughs of the New York city. The foursquare API will give us information on the Restaurants and their type like Chinese, Italian etc. and their location in each Borough.

▼ root: {} 2 keys
  ► meta: {} 2 keys
  ▼ response: {} 1 key
    ▼ venues: [] 50 items
      ▼ 0: {} 6 keys
          id: "3fd66200f964a5209ae31ee3"
          name: "Arturo's Restaurant"
        ▼ location: {} 12 keys
            address: "106 W Houston St"
            crossStreet: "at Thompson St."
            lat: 40.727407
            lng: -74.000378
          ► labeledLatLngs: [] 1 item
            distance: 65
            postalCode: "10012"
            cc: "US"
            city: "New York"
            state: "NY"
            country: "United States"
          ► formattedAddress: [] 3 items
        ▼ categories: [] 1 item
          ▼ 0: {} 6 keys
              id: "4bf58dd8d48988d110941735"
              name: "Italian Restaurant"
              pluralName: "Italian Restaurants"
              shortName: "Italian"
            ► icon: {} 2 keys
              primary: true
          referralId: "v-1545202601"
          hasPerk: false
      ► 1: {} 7 keys

From the Boroughs available in the data, four boroughs namely Bronx, Manhattan, Brooklyn and Queens are chosen for analysis.

Methodology

PreProcessing:

Scaling and distribution:

The data from the json file is totally unstructured one. It was normalized using python dataframe and cleaned. The data was grouped based on Borough. The restaurant type was averaged to find the frequency of each type in a borough.

Modeling:

I have used the k-means clustering to cluster and analyze the restaurants in the boroughs
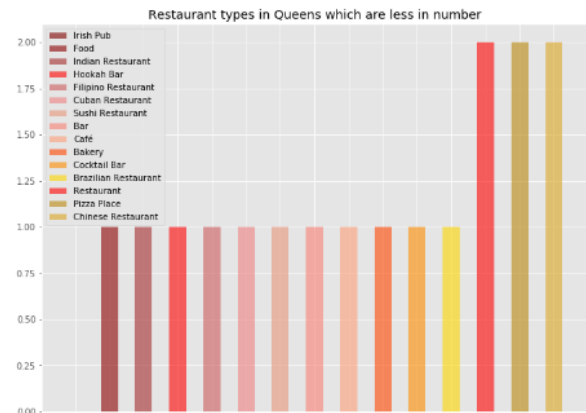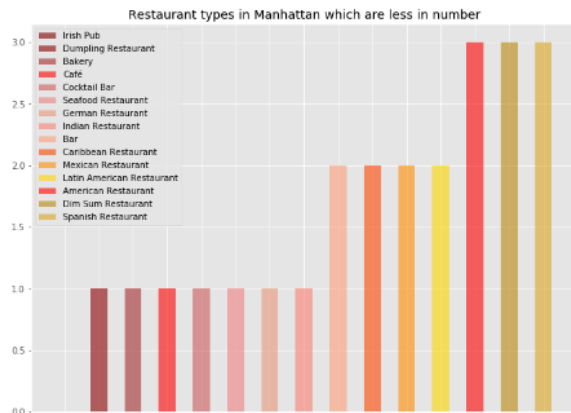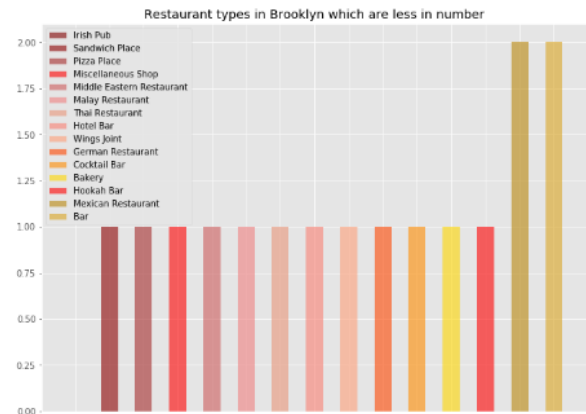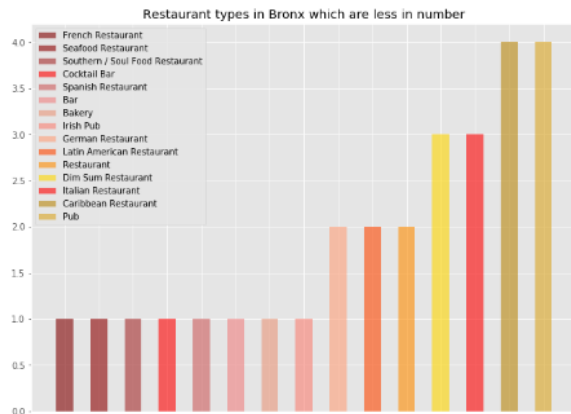
Results

Word Cloud visualization:
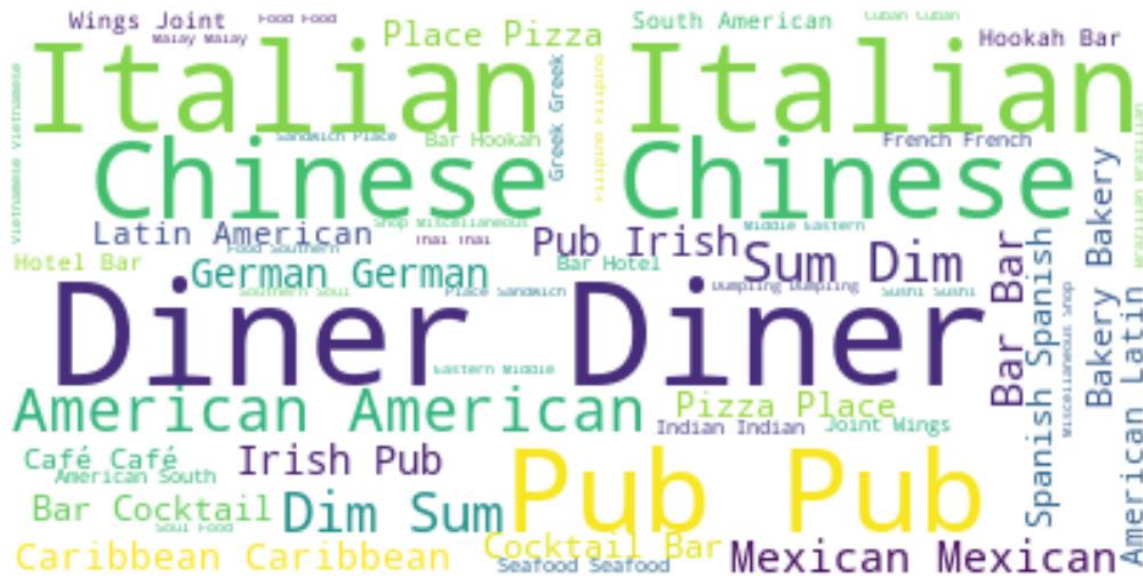
The above visualization will help in identifying the major types of restaurants that are present all over the boroughs.

Bar chart visualization:

This visualization will help the customer to identify the top restaurant type in Bronx, Brooklyn, Manhattan and Queens Boroughs. Based on this he can make a decision on the type of restaurant he should start.

Restaurant types in Bronx which are less in number

Restaurant types in Brooklyn which are less in number

Restaurant types in Manhattan which are less in number

Restaurant types in Queens which are less in number

This report will give an idea on the lesser type of restaurants that are present in these boroughs. This will give a clue whether to start these type of business as they are less in these areas.

Clustering:

The hotels were clustered using the k-means clustering and each cluster was analyzed for similarity and dissimilarity.

Discussion

Based on the bar chart visualization we can see that mostly Italian, Chinese, Diners and Pubs are prominent in these boroughs. Also it can be seen that Indian, Irish Pub, Seafood, Malay and Cuban restaurants are less in these boroughs. Based on the type of restaurant the customer is interested he can decide a borough where that type of restaurants are less but also there are communities that like these type of foods.

Conclusion

This analysis is based on the data available from foursquare on the restaurants in these boroughs. This analysis can be further enhanced by checking at the feed back provided by the user, the number of checkins. We should also collect the data regarding the community , regional details of the people staying in these areas. This will help in making a more precise analysis for making a decision on the type of restaurant to be started in which borough.