

Capstone Project

IBM Data Science Professional Certificate - Coursera

Introduction

Data Science with its machine learning techniques is taking business by storm. Today internet is flooded with these ML models from e-commerce to human care and list goes on.

In this project we will use one of ML technique to look into Neighbourhoods of Brooklyn, New York and create cluster containing neighbourhoods with similarities.

Data/Libraries Used -

- We have used New York dataset which contains details of all Borough in New York and their corresponding neighbourhoods.
- Next we have used Foursquare API to fetch location specific venues from each neighbourhood.
- We have used Folium library greatly to visualize the neighbourhoods clearly and have a clear understanding. Also Matplotlib is used.
- For data wrangling, analysis and cleaning, we have used Numpy, Pandas.
- For ML we have used skit-learn library.

Methodology

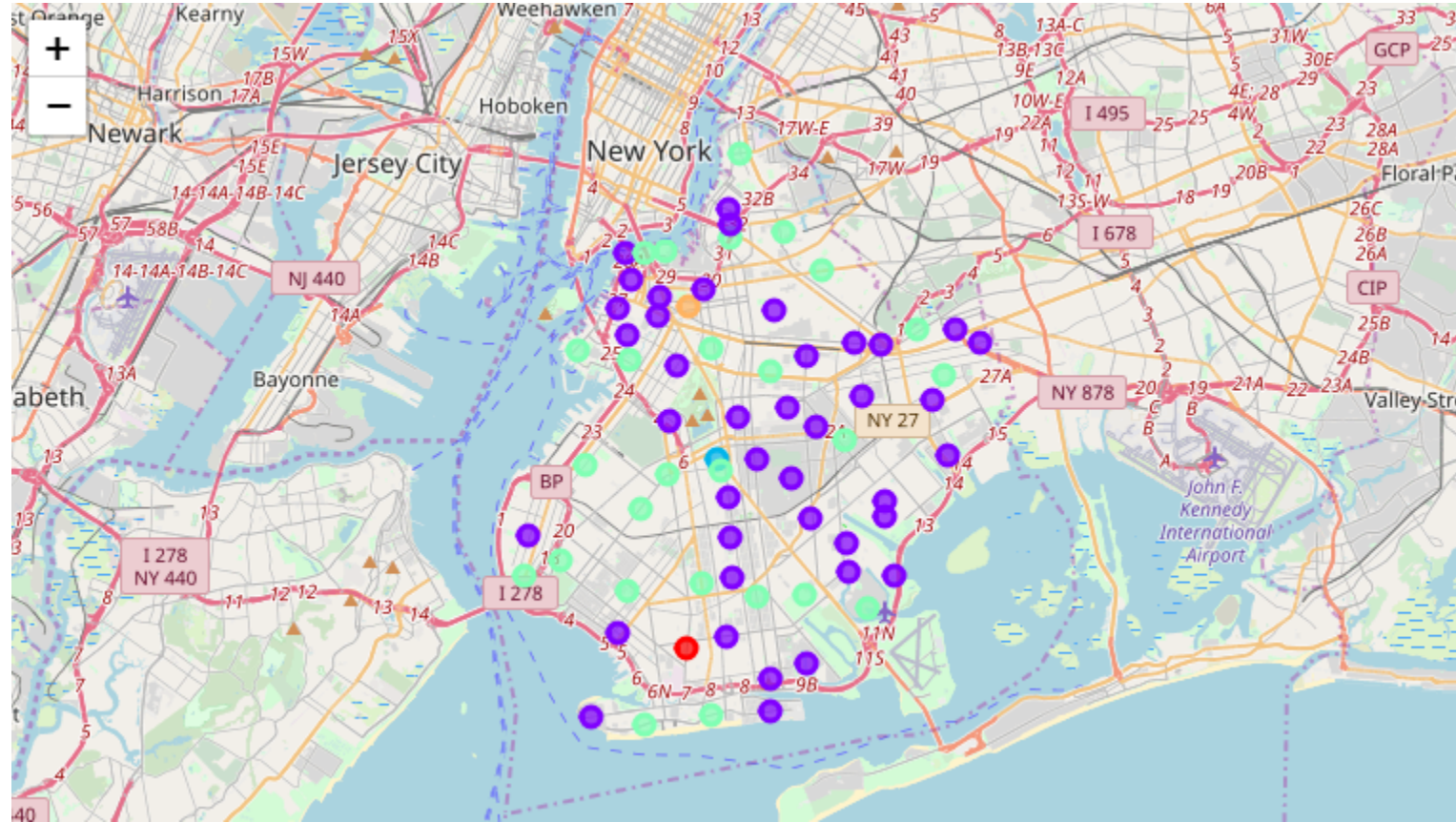
In this project New York data is taken as a base which contains info of all borough and their corresponding neighborhoods. Some exploratory data analysis applied –

- From the dataframe we will subset the dataset to get only the records for BROOKLYN,
- Visualize all neighborhood of BROOKLYN using FOLIUM library.
- Fetching Latitude and Longitude of neighborhoods using Geopy Library
- Normalizing the data fetched from Foursquare API.

Machine learning technique –

- We will use Clustering technique of Unsupervised Learning to segment the neighborhoods and cluster them using K-MEANS clustering to get the clusters with similarities in their venues.

Sample Maps Generated for Brooklyn



Wrapping it up -

- According to results we see that most common venues come out to be restaurants and Coffee shops, which means any new businessman can start restaurants provided that they need to compete with existing established chains but if he provides top class facilities to get to top.

Thanks!!!