

Coursera Capstone

IBM Applied Data Science Capstone

“The Battle of Neighbourhoods”

**Finding a Neighbourhood for an Indian family to settle
down in Queens, New York**

By: Abhinav Yadav
December 2020

1. Introduction:

The United States of America is the country with one of the largest immigrant population in the world. After the IT boom , this immigration in recent times has been driven by IT workers from Asia coming into the states. A large number of them belong to India which due to its high population and educational system is able to provide one of the cheapest most talented manpower for IT firms. Indians flock by the 1000's to America chasing the American Dream. This along with the America's policy of helping the best brains of the world settle into this country has led to a large number of highly skilled immigrant population. New York has one of the largest Indian population . Out of all the Boroughs - Queens with 6.2 % Percent of Indian Americans in Municipality Population seems to be one that has welcomed the India community with open arms and is a popular location for Indians to settle down. Indians culturally have a strong sense of family and generally prefer to be around theirs . This means they generally come to America with their spouses and Children . While to settle down the future life of their Children influences a decision to a large extent . People prefer living in neighbourhoods that have schools and parks . The presence of Doctor's Offices is also a big influencer.

In this project we will try to find an optimal location for an India family to settle. Specifically, this report will be targeted to people interested in settling down in Queens , New York, USA.

Since there are lots of Neighbourhoods in Queens we will try and find a location that has :

- 1- Schools
- 2- Parks
- 3- Doctor's Office

We will also try and include neighbourhoods with Indian Restaurants (Although this will be an additional clincher not an influencer)

We will use our data science powers to generate a few most promising neighbourhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2. Problem Statement:

The objective of this capstone project will be to find a suitable neighbourhood for an Indian family with kids to settle down .This would be done by leveraging Data Science and Machine Learning (k - means). The main Business question that will be answered in the Capstone Project will be : “Which neighbourhoods in Queens New York are best suitable for an Indian Family with kids to settle down in ?”

3. Stakeholders/Target Audience

The Capstone Project will be particularly useful for people looking to move into new towns, cities or even countries. The project can also be modified to go beyond the current scope as it can be used to scope out other businesses in the area be it gyms , schools etc . The project can be used in an advisory capability by property consultants, realtors etc who can use the project to give their customer an overview of the area and allow them to make a better informed decision. With basic knowledge, customers themselves can use this project to better understand the option available with them to make informed decisions.

4. Foursquare API

The project largely relies on utilising the Foursquare API, mainly the Places API to gather data related to locations. Foursquare is a location technology platform that allows the user to access its upto date database through an API to provide details of location/ venues the user might be interested in. The details include Name, Category, Location (latitude, longitude) , Ratings , reviews , menu etc as per the customer needs. We will be leveraging the API in this project to find out the venues that are present in the Queens Area to look for a suitable neighbourhood for settling down.

5. Data

Data used in this project is the New York dataset and was sourced from: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

The coordinates of places if and when required can be sought by using geopy. Geopy is a Python client for several popular geocoding web services. Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources

Finally Foursquare API will be used for identifying and analysing areas of interests which basically involves using the API to gather the following details –

- a. Number of venues in a particular area based on the radius provided by the user based on neighbourhood details.
- b. Name of the venue
- c. Category of venues (Schools , Parks etc)

Location of the Venue.