

1. Introduction

1.1. Description of the Problem:

New York City, more fondly known as the Big Apple is one of the busiest cities in the world and in fact it is a city that never sleeps. Being the most populous as well as the most densely populated city in the US, NYC has been a major point of entry for immigrants, with roughly 36% of the population being foreign born. The challenge is to help these immigrants identify the neighbourhood where they can find the food that they are comfortable with.

1.2 Description of the Background:

Shivaram, a data science enthusiast has moved to New York City from Chennai, India as part of his new job. He wants to use his skills to identify the right neighbourhood for him to get settled where he will have the best indian restaurants that can make him feel at home.

2. Data Section

2.1.1 Dataset 1:

The below link contains information about the boroughs and the different neighbourhoods under each borough of NYC along with its latitude and longitude coordinates.

https://geo.nyu.edu/catalog/nyu_2451_34572

2.1.2 Dataset 2:

The below link contains information about the demographics of NYC and segmentation of ethnic races across different boroughs.

https://en.wikipedia.org/wiki/Demographics_of_New_York_City

2.1.3 Dataset 3:

The Foursquare API will be used to obtain the geographical location data for the NYC Area. These will be used to explore the venues in the neighbourhoods of NYC. The venues will provide the categories needed for the analysis and eventually, these will be used to determine the viable neighbourhood.

2.2 How data will be used to solve the problem:

The first two datasets will be analysed by considering the venues within the neighbourhood of NYC postal code areas. These areas' restaurants would be checked in terms of the types of restaurants within a certain mile radius. Due to Foursquare restrictions, the number of venues will be limited to 100 venues.

3. Methodology

- I analysed the demographics of New York City to better understand how the population spread is across the city, based on which I narrowed down the boroughs which I should concentrate on.
- From those boroughs, to go with my problem statement, since the hero of my problem is an Indian, I analysed how each of those boroughs are diversified in terms of race and ethnicity and picked the one with higher concentration of indian immigrants.
- I used BeautifulSoup package to scrape the wikipedia pages to gather the above mentioned information. This was performed to reduce the load on the code and also to overcome API limitations.
- After that, using the geolocation in Python, I pulled the lat-long coordinates of all the neighbourhoods in the borough which I finalized.
- Post that, using Foursquare API, I was able to pull 100 venues for each of the neighborhoods in my borough. This was due to the limitation in the free API provided by Foursquare.
- Then, I grouped the list of similar restaurants available in each of the neighborhoods to understand which neighborhoods had the most number of options.
- Finally, I created 5 clusters from the list of neighborhoods in my borough and segregated them to understand which segment provides the good number of indian restaurants.

4. Results

- Astoria has the highest number of Indian Restaurants (5), followed by Bayside and Jamaica Hills (3 each).
- Almost all the neighborhoods having Indian restaurants are under Cluster 1, which would be the most ideal segment of Queens if Shivaram wants to enjoy Indian food every day.
- You can refer my notebook here in this [link](#).

5. Conclusion

- I would like to take this opportunity to thank IBM and Coursera for giving me this platform that enabled me to learn the basics of Data Science using Python, which I will now be able to explore further on my own and use it to reach greater heights in my Career.