Capstone Project - The Battle of Neighborhoods (Part-1)

Data Selection

United States of America has a racially and ethnically diverse population. At the federal level, race and ethnicity have been categorized separately.In 2006, of the 1,266,264 legal immigrants to the United States, 58,072 were from India. Between 2000 and 2006, 421,006 Indian immigrants were admitted to the U.S., up from 352,278 during the 1990–1999 period. According to the 2000 U.S. census, the overall growth rate for Indians from 1990 to 2000 was 105.87 percent.

The average growth rate for the U.S. was 7.6 percent. Indians comprise 16.4 percent of the Asian-American community. In 2000, the Indian-born population in the U.S. was 1.007 million. According to the U.S. Census Bureau, between 1990 and 2000, the Indian population in the U.S. grew 130% – 10 times the national average of 13%. Indian Americans are the third largest Asian American ethnic group, following Chinese Americans and Filipino Americans.

For our restaurant problem, we will focus on the Boroughs of United States and work on getting the data from all the Boroughs.To solve our problem of finding a best location to start an Indian restaurant in US, we need to datasets based on various parameters such as :

1. Population of target audience in all the boroughs of US based on their Asian ethnicity, Age, Gender, Marital Status, Employment Status, Income
2. We also need the data about the required Business floorspace and Rateable Value statistics of each borough.
3. Considering the competitors factor, we also need the data of extsing Licensed Restaurants in each borough.
4. And lastly we will also consider the borough level tourist and domestic annual spend estimates.

All the above required information is available at Wikipedia, which is a free and open data-sharing portal where anyone can access information.

The link for the data : https://en.wikipedia.org/wiki/Indian_Americans

Along with the above datasets we will also use the Foursquare location data to solve our problem.