**Relocating from Mumbai to New York City**

**Capstone Project**

**IBM Data Science Professional Certificate**

**Problem Statement**

An Indian student is moving to New York from Mumbai for further studies. As he will use public transportation system for daily commute and will hardly have any time for cooking. Ideally, he wants to live in a neighborhood with both Indian grocery store and Indian restaurant, if not, at least one of them.

**Broader Audience**

People relocating from India to New York wanting to live in neighborhood with easy access to Indian food.

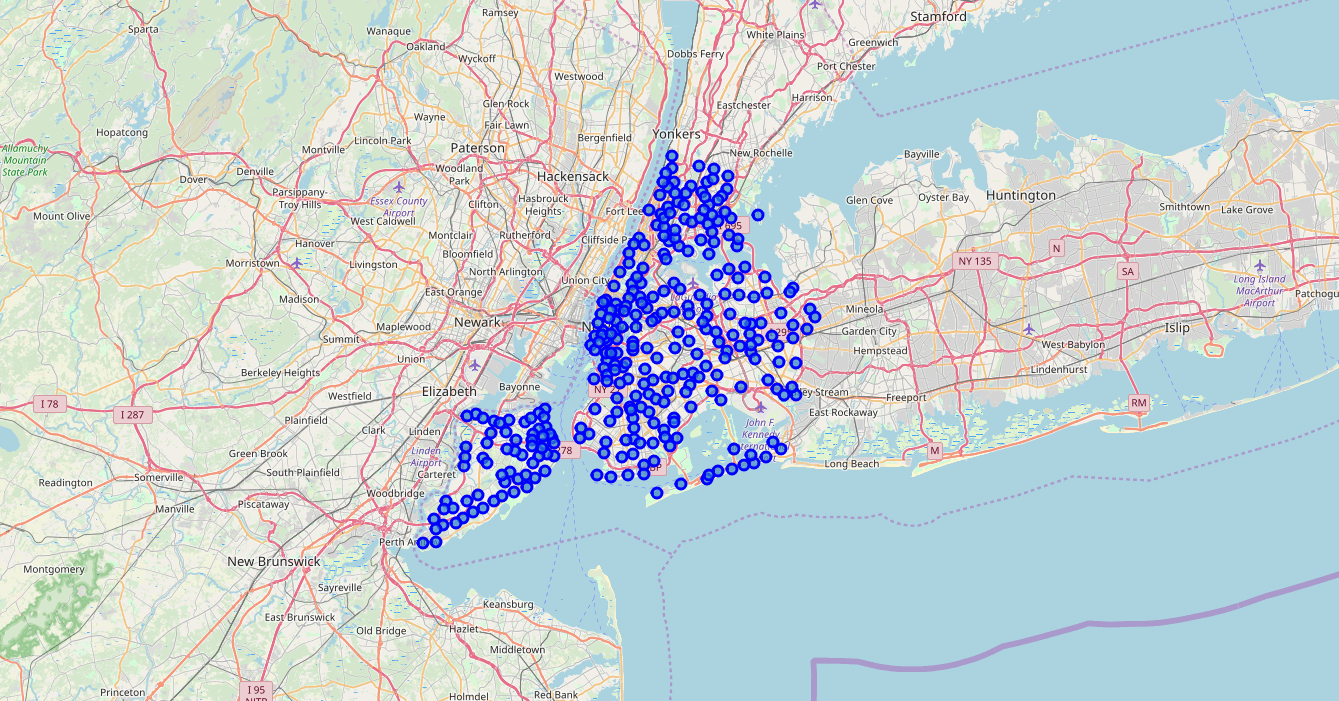
**Data and libraries**

Data Source: New York is pulled from shared drive provided in IBM Coursera program lab.

Other libraries will be used to carry out analysis and derive desirable insights are - Pandas - Library for Data Analysis, NumPy – Library to handle data in a vectorized manner, JSON – Library to handle JSON files, Geopy – To retrieve Location Data, Requests – Library to handle http requests, Matplotlib – Python Plotting Module, Sklearn – Python machine learning Library, Folium – Map rendering Library.

**Methodology**

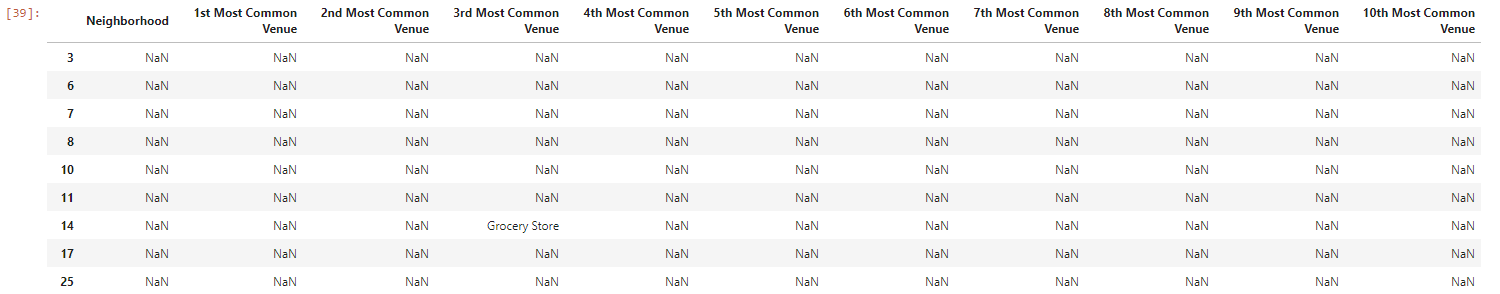
To start the analysis the data provide by the instructor in one of the lab sessions was used. The data for New York City comprised of 5 boroughs and 306 neighborhoods. Using geopy the map of New York city was created with these 306 neighborhoods.

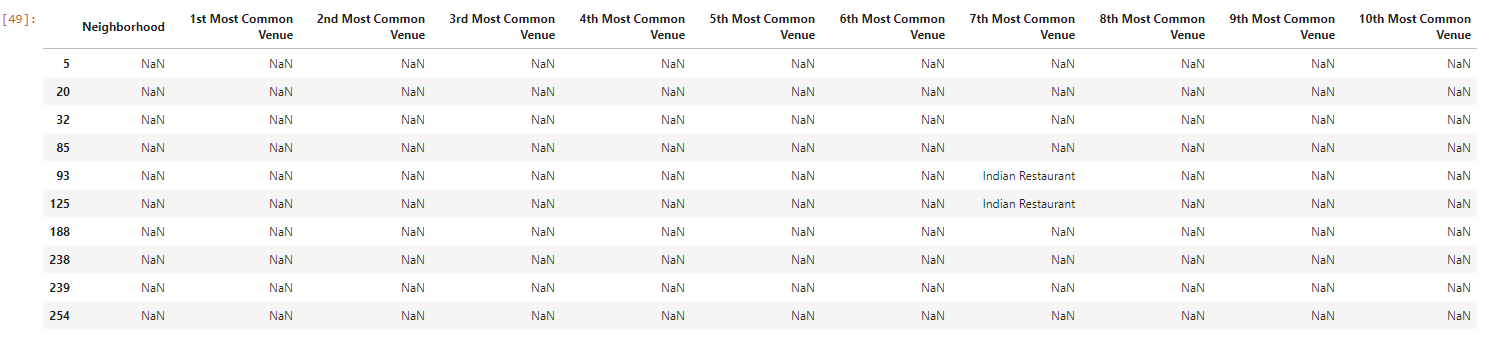


Next, using Foursquare API identified the number of venues as 30362 that were categorized in 446 categories.

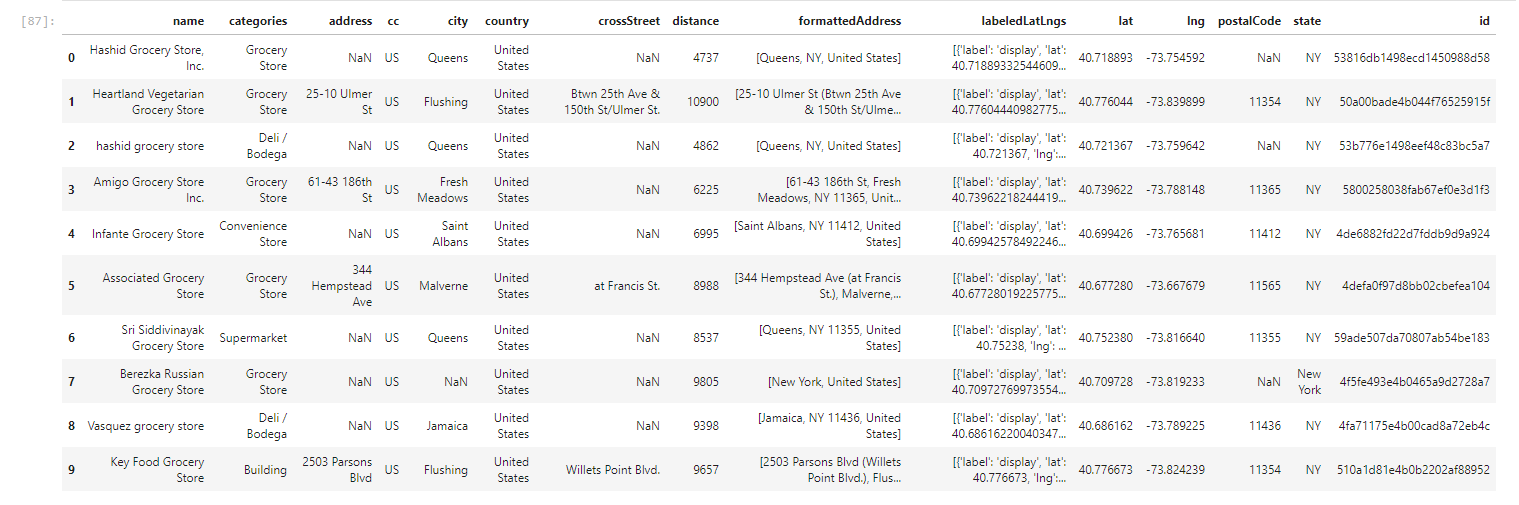


Using the K mean created five clusters of these neighborhoods and Identified neighborhoods with Grocery stores and Indian Restaurants.

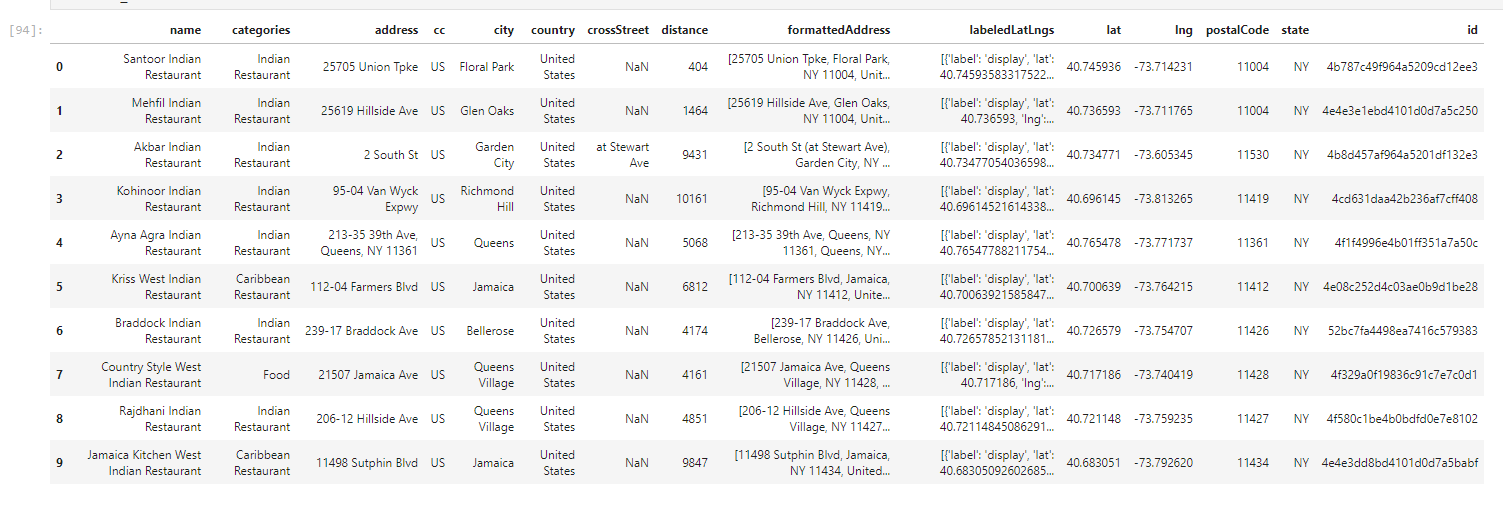




The categories for Grocery store does not provide additional information in terms of what kind of grocery stores they are. Therefore, the data was analyzed to get names of the grocery store in identified neighborhoods. Based on the names of the grocery store, the “Sri Siddhivinakay Grocery Store” was assumed as an Indian Grocery store.



Once the Indian grocery store was identified, the neighborhoods data was analyzed to get names of Indian Restaurants.



Finally, neighborhood Queens in Queens Borough was identified as ideal location with both Indian grocery store and Indian Restaurant.

**Next Steps**

In this analysis the data was limited to top 10 ranked venues, to get detailed list of grocery stores and Indian Restaurants the data analysis can be expanded to entire dataset instead of top 10.