Capstone Data Science Project

**Project Name:** The Battle of Neighbourhoods

**Author:** Pankaj Kumar

**Date:** 14th March 2020

**Location:** Sydney, Australia

**Introduction:**

Hello! My name is Pankaj, and like my other 1.33 Billion Indians, I too am fond of our Indian traditional cuisine, and am proud of it!

The one thing we all Indians love unanimously is our Indian food! We are closely connected to the aroma & taste of the variety of spices, which bring the unique taste to our food.

Back in 2009, I planned to travel to New York along with group of Indian friends. Well, we made all necessary arrangements but the most important topic with a big question mark was - where do we find Indian restaurants?

New York is known for its long history of immigration & ethnically diverse culture & food. But none of us had luxury of carrying smartphones, which could give us feedback on Indian restaurants within New York back then.

We managed to do a bit of research about good Indian food in NY by Internet searches, asking hotel staff, little bit of strolling around and found couple of Indian takeaway/Dine-in restaurants around but the lack of taste was pointing to only result - our research wasn't good.

Today, as a part of my Capstone Data Science project, I'll like to utilize my Data Science skills to solve a few basic problems of finding out

* **Which areas in NY have Indian restaurants?**
* **Which is the best location to look for Indian restaurants in NY?**
* **Where should I stay to have multiple options of Indian restaurants around?**

**Data Analysis:**

For this project we would be looking at below data:

(1) New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude.

\* Data source: https://cocl.us/new\_york\_dataset

\* Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of New York City.

(2) Indian restaurants in each neighbourhood of New York City.

\* Data source: Fousquare API

\* Description: By using this API, we will get all the venues in each neighbourhood. We can filter these venues to get only Indian restaurants.

(3) GeoSpace data

\* Data source: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

\* Description : By using this geo space data we will get the New york Borough boundaries that will help us visualize choropleth map.

**Problem solving approach:**

\* Collect the New York City data from https://cocl.us/new\_york\_dataset

\* Using FourSquare API we will find all venues for each neighbourhood.

\* Filter out all venues that are Indian Restaurants.

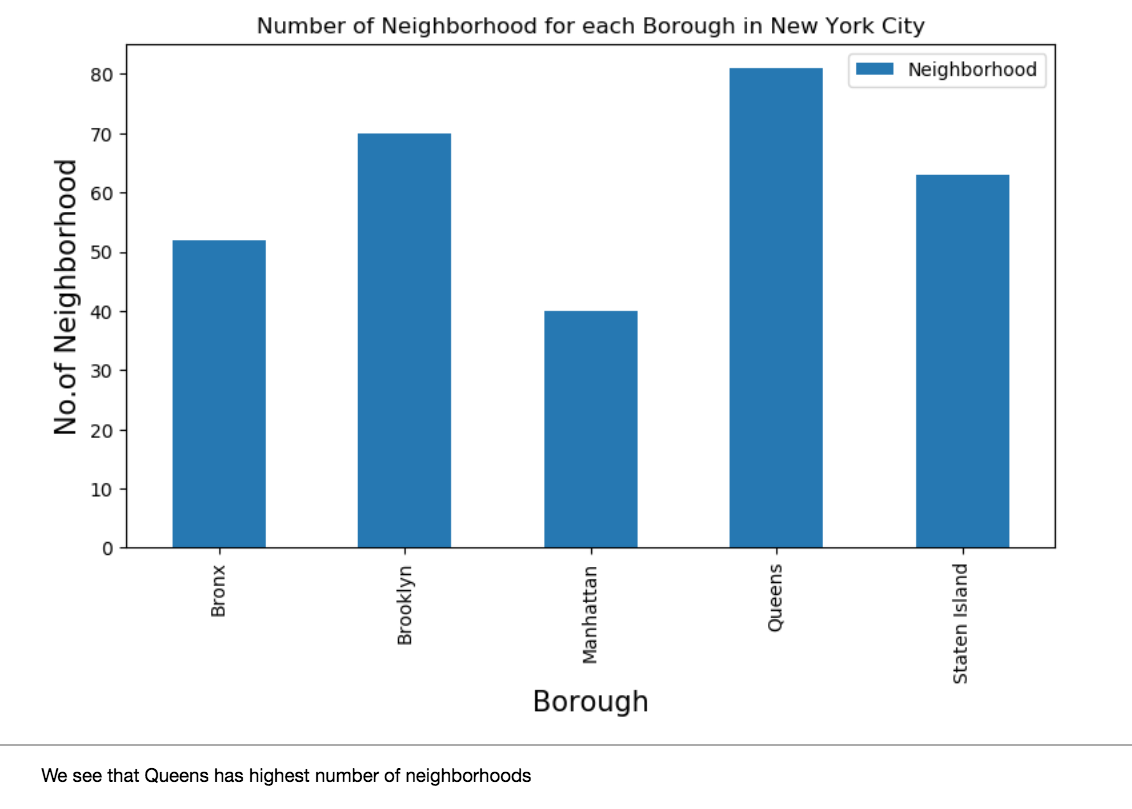
\* Find rating, tips and like count for each Indian Restaurants using FourSquare API.

\* Using rating for each restaurant, we will sort that data.

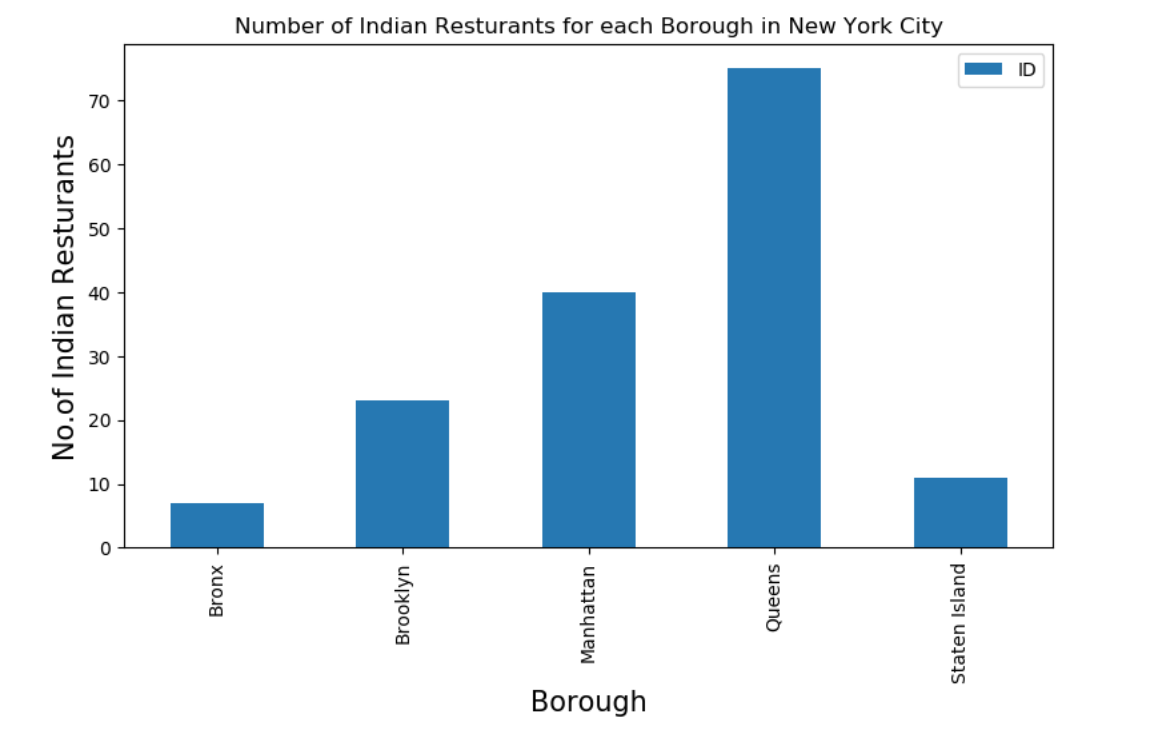
**Jupyter Notebook link:** <https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/4ead17c3-f1b0-4030-a167-338a342e8b75/view?access_token=b7666b9afa623224beef42d754490834cd60c48bfc5721d9bd9642672116da5e>

**Data Interpretation with observations:**

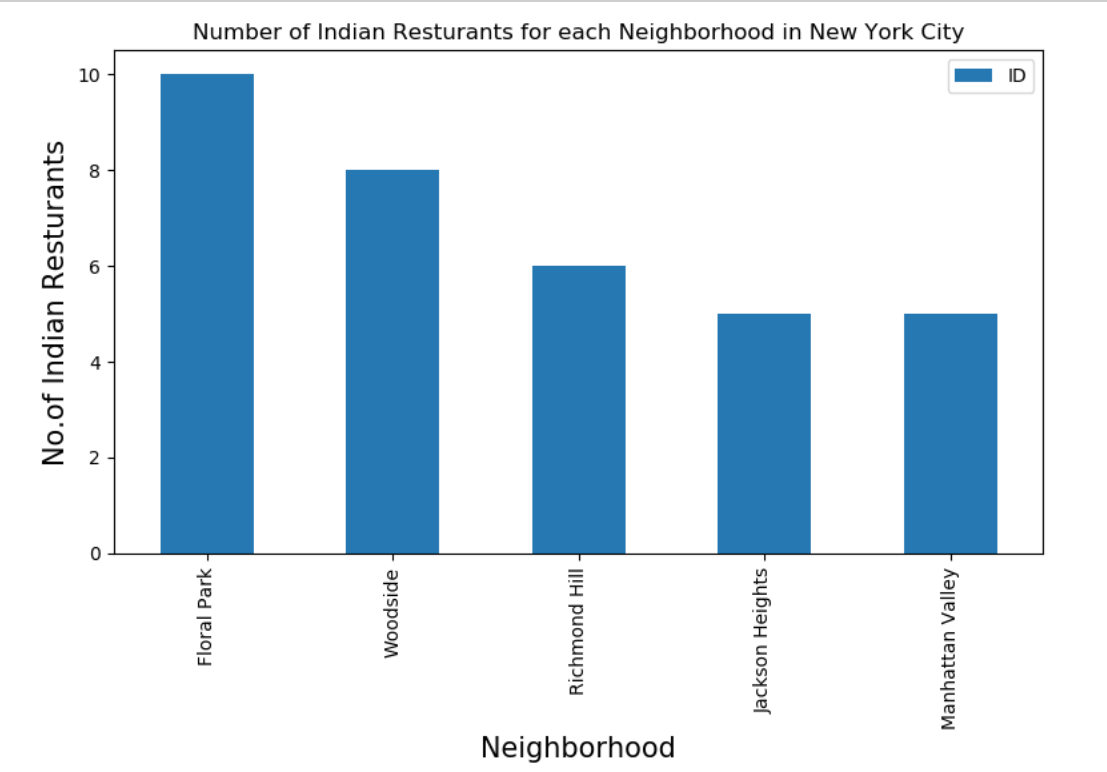
1. There are total of 306 different Neighbourhoods in New York



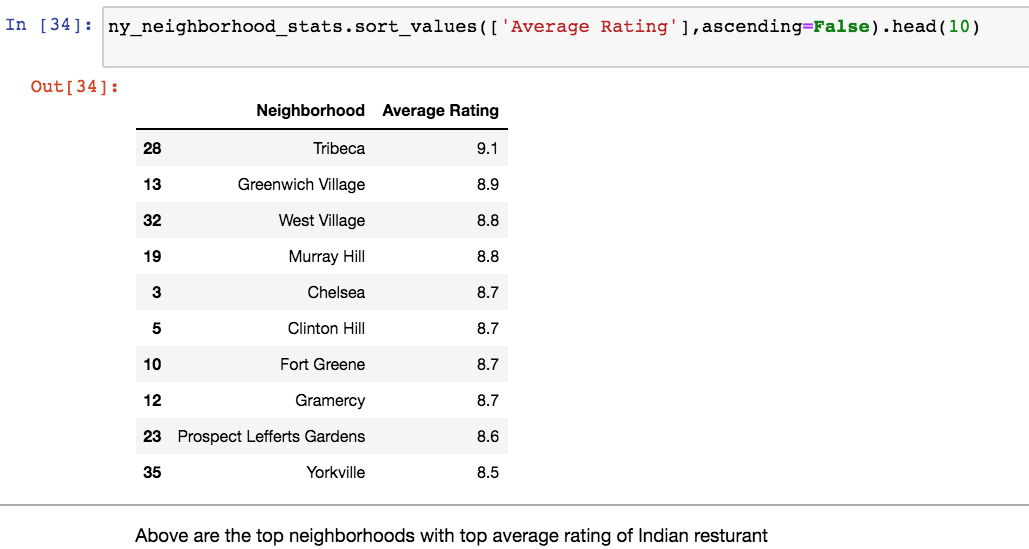
2. We got 156 Indian Restaurants across New York City

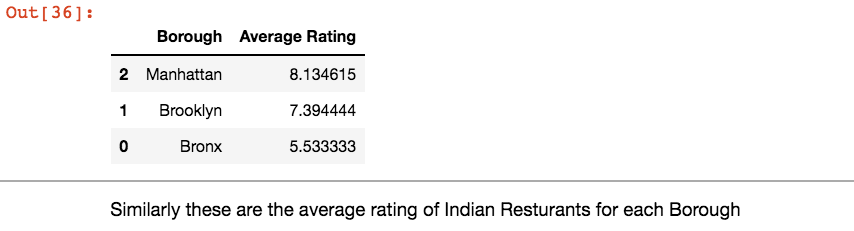


1. Indian Restaurant Count by Neighbourhood

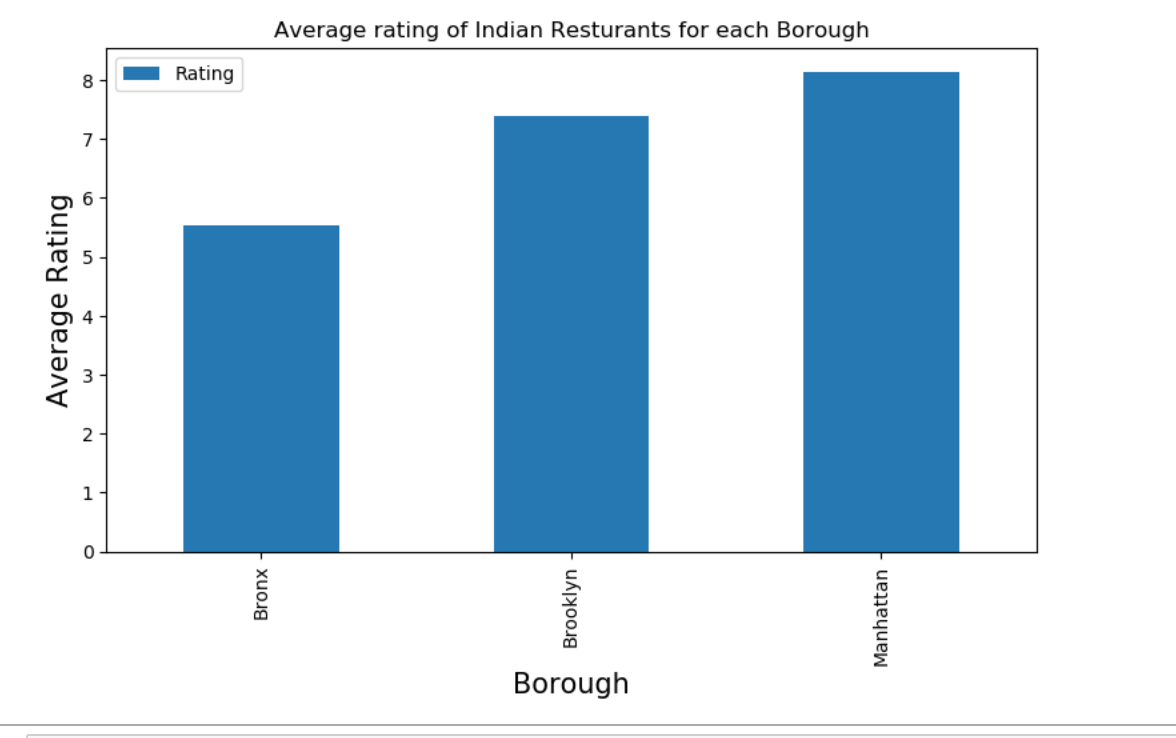


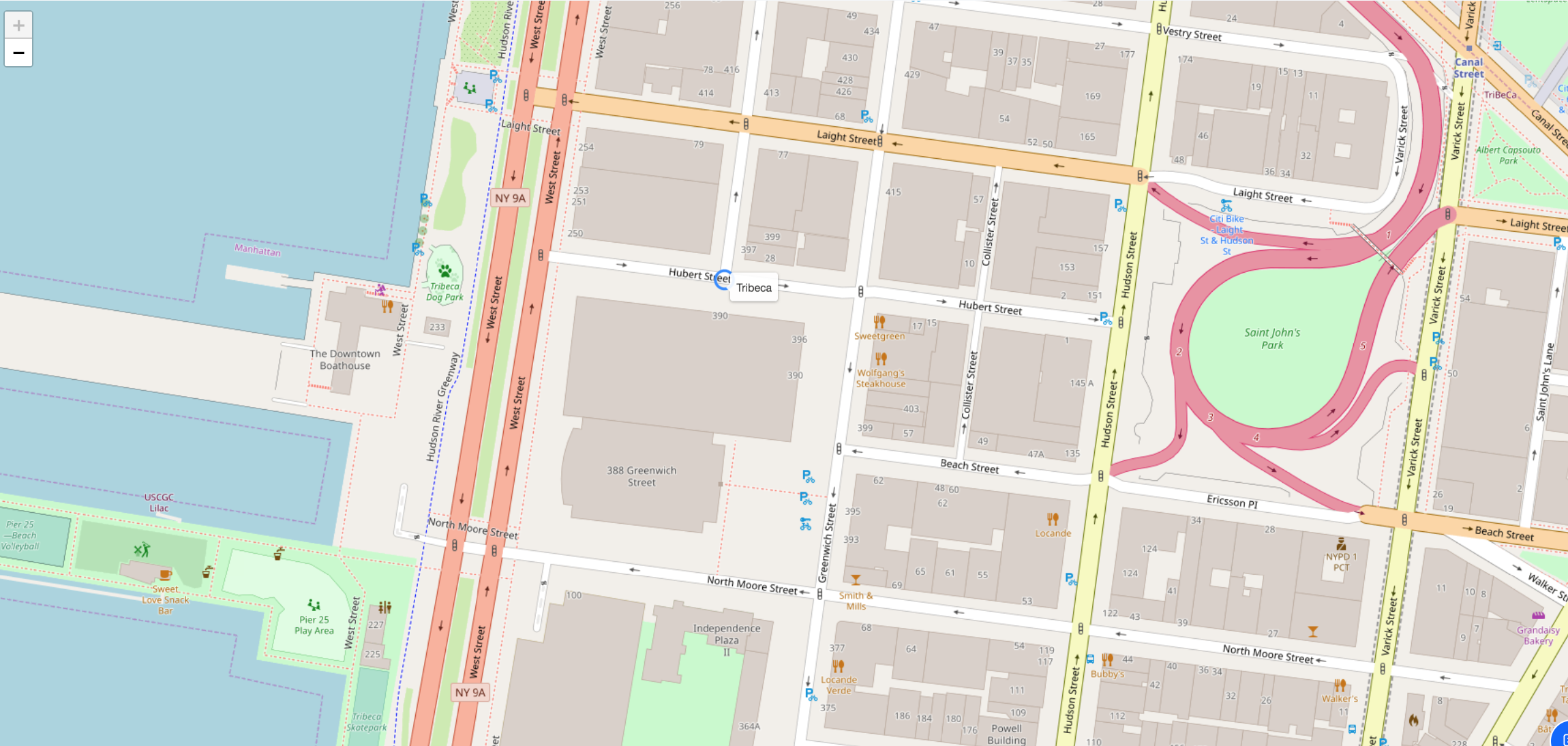
4.Floral Park in Queens has the highest number of Indian Restaurants with a total count of 11.





5. Average rating of Indian Restaurants for each Borough





**Conclusion:**

* There are total of 306 different Neighbourhoods in New York
* We have 156 Indian Restaurants across New York City
* Floral Park in Queens has the highest number of Indian Restaurants with a total count of 11
* Tribeca has the highest average rating = 9.1 for Indian restaurants among all neighbourhoods
* Top 3 Boroughs with top average ranking of Indian restaurants are: Manhattan (8.13), Brooklyn (7.39) and Bronx (5.53)
* I would like to choose Tribeca as my preferred location of best-rated Indian Restaurant in New York.

**Limitations**

* The ranking is purely on basis of rating of restaurants
* The accuracy of data depends purely depends on the data provided by FourSquare