**Capstone Project -IndianDhaba**

1. **Introduction**

1.1 **Business Understanding**: Business Clients are interested in figuring out the best locations to open the New IndiaDhaba which includes dishes made from India Flavors and Major Ingredients. Main focus to explore/elevate the taste and name of Indian dishes to across the city of New York.

**Interested**: Authentic Indians Dhabas owners who wants to explore the Indian Cuisines/Dishes.

* It includes to find out the best location where Indians Families residents in large amount and transportations available easily for that place from every corner of city.
* Location have less or no number of Indian restaurant at that location.
* Best to find place which visited frequently by many tourists and local citizens.
* Middle of city will be best one choice.
* Market whether all ingredients easily available.
* In Budget

In this business requirement we have to search the most prominent choice/location for restaurants with more of he requirements includes. Will need to fix all the things in given budget.

First things we have to do is finding the most visited places in New York city.

QUESTIONS:

* Transportations cost?
* What kind of citizens residents at that location?
* Is good idea to open the IndianDahaba at that location?
* What is the backup?
* What kind of foods, Indian dishes most liked in New York?
* Is there any hotels/bars/restaurants which includes the similar kind of dishes?

**2) Data Requirement and Understanding**

In the initial data collection , we have to identify and gather the available data resources. These can be in the form of structured, unstructured, and even semi-structured data relevant to the problem domain.

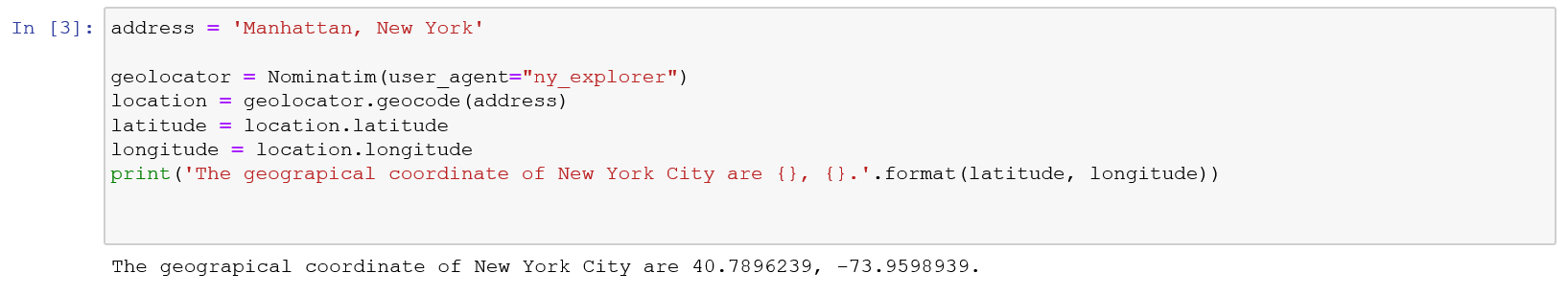
* Web Scraping for find out the most visited and famous places for foods in New York City.
* Link: <https://www.thrillist.com/eat/new-york/most-famous-nyc-restaurants>

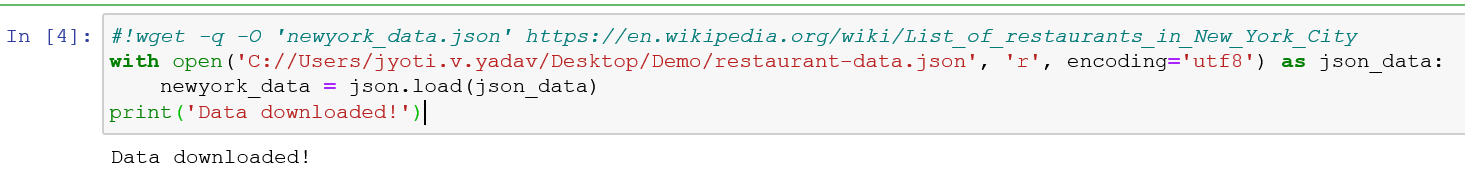
**2.1 Data Collection and Understanding:**

- It's important to understand its content and assess its quality,

- Discover only interesting preliminary insights.

- Determine whether additional data is necessary to fill any gaps in the data.

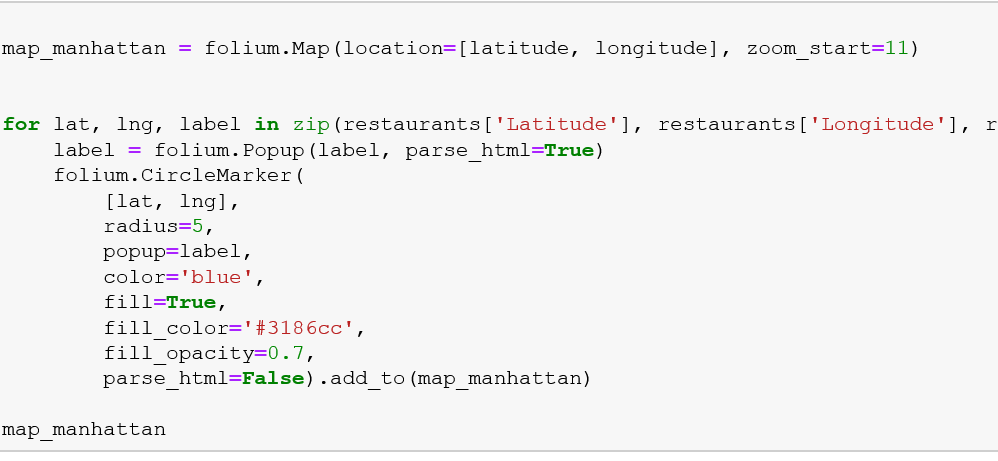




Transform the data into a pandas data frame The next task is essentially transforming this data of nested Python dictionaries into a pandas data frame. So let's start by creating an empty data frame



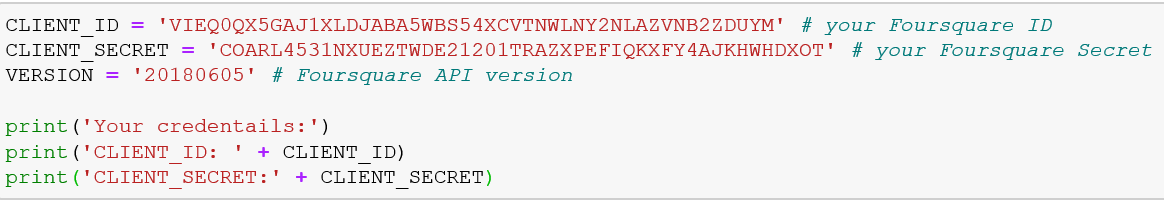
Let's get the geographical coordinates of Manhattan, NY.



**2.2 Foursquare API**

Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and

segment them. Define Foursquare Credentials and Version



3) **Methodology**

In this project we will direct our efforts on detecting areas of New York that have low restaurant density, particularly those with low number of Indian restaurants.

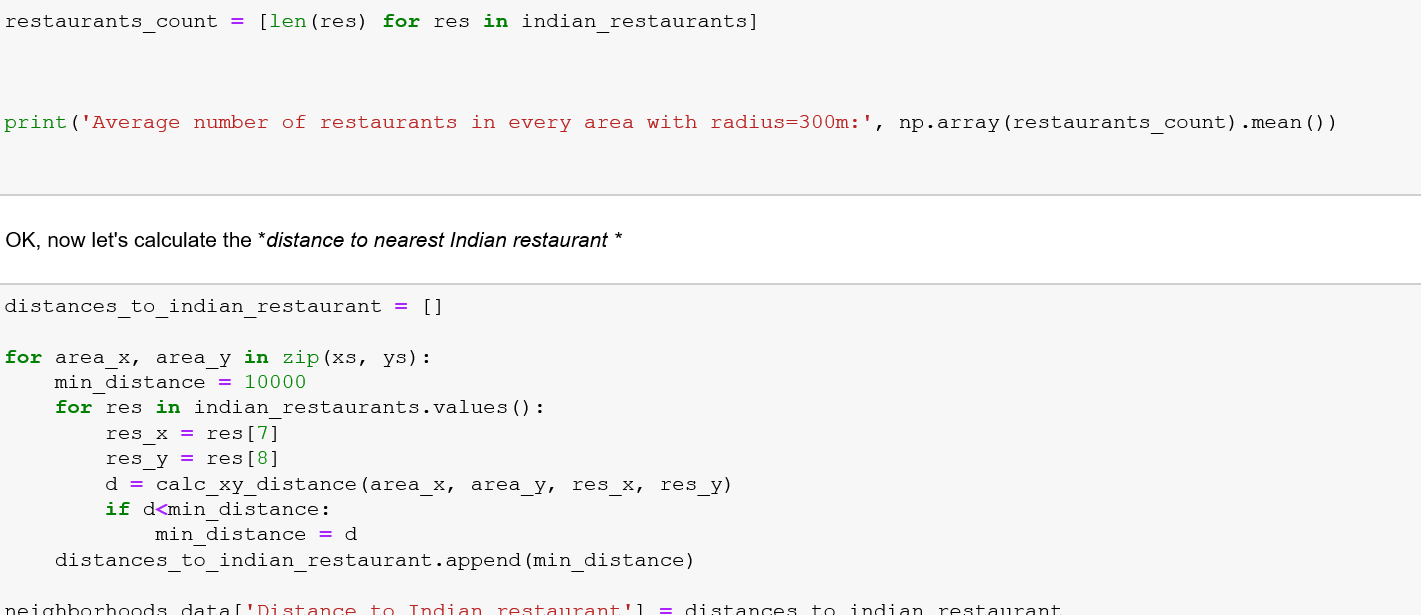
In first step we have collected the required **data: location and type (category) of every restaurant . We have also \*\*identified Indian restaurants** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of New York to identify a few promising areas close to center with low number of restaurants in general (and no Indian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **less number of restaurants \*, and we want locations \*without Indian restaurants in that location**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

4) **Analysis**

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the **number of restaurants in every area candidate**:



5) **Results and Discussion**

Our analysis shows that although there is a great number of restaurants in New York, there are pockets of low restaurant density fairly close to city center. So we focused our attention to areas with low density of Indian restaurants.

After directing our attention to this more narrow area of interest those locations were then filtered so that those with less restaurants .

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Indian restaurants particularly. This does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close in New York - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area.

6) **Conclusion** Purpose of this project was to identify New York areas close to center with low number of restaurants (particularly Indian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Indian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water)etc.