The Battle of the Neighborhoods-Report

1.Introduction & Business Problem:

Problem Background:

New York City - there's no place quite like it. Even if you think you know New York City well-the world-class museums, amazing food and unforgettable views-there's always something new and exciting to discover. Thrilling shows, fabulous shopping, and first-rate sports draw visitors from around the globe to the world's entertainment capital.

With awesome architecture and parks, New York was rated No. 1 of America's most beautiful cities by Budget Travel. Take in its grandeur from observation decks at icons like the Empire State Building and Top of the Rock, or on a stroll across the Brooklyn Bridge or Brooklyn Heights Promenade.

Problem Description:

If suppose you want to visit New York City for the first time you don't have any idea about that city then the very first problem you face is to pick the best hotel to stay. So for that we will be googling for the hotels checking for the reviews and feedback. Once we find the hotel then we check for the nearby places to visit and lot more. All this process is time consuming one we should search at different sites which is much more difficult.

In the process of automating this task the below mentioned problems also includes:
- ➤How to fetch the hotels data in any given city?
- ➤what features are to be considered at the time of ranking the Hotels?
- ➤How to get the nearby places surrounded to the hotel?

Yes, we are going to work on this problems and will try to find a solution for this...

Target Audience:

To help the user who is going to visit the New York city by providing the details like best rated hotels for stay along with the details of the visiting places surrounded by that hotels. By providing this details we can help they in picking for the best as the reason for the each user will be different the preferences of the user may also differ for each so by providing this details we can help thenm more. We can even do more like including the check-in check-out time and lot more but as its tonly the starting stage I made it only this much.
This would interest anyone who wants to visit New York city.

Success Criteria:

The success criteria of the project will be a good recommendation of borough/Neighborhood choice for the users who wants to visit as it is providing the best 5 places in that location and nearest visiting spots.

2. Data :

One city will be analyzed in this project : *New York City*.

We will be using the below data sets for analyzing New York city

Data 1 : Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.

This dataset exists for free on the web. Link to the dataset is :
 https://geo.nyu.edu/catalog/nyu_2451_34572

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Data 2 : For the below analysis we will get data from Wikipedia as given below :

1. New York Population
2. New York City Demographics
3. Cuisine of New York city

 https://en.wikipedia.org/wiki/New_York_City
 https://en.wikipedia.org/wiki/Economy_of_New_York_City
 https://en.wikipedia.org/wiki/Portal:New_York_City
 https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

Data 3 : New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

Out[46]:

	name	categories	lat	lng
0	Lollipops Gelato	Dessert Shop	40.894123	-73.845892
1	Rite Aid	Pharmacy	40.896649	-73.844846
2	Carvel Ice Cream	Ice Cream Shop	40.890487	-73.848568
3	Dunkin'	Donut Shop	40.890459	-73.849089
4	Shell	Gas Station	40.894187	-73.845862

3. Methodology :

Business Understanding :

Our main goal is to get the best 5 hotels to stay in New York City for the Users.

Analytic Approach :

New York city neighborhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is clustering of Wake field . And second part is clustering of Bronx. This is done because of the following Exploratory data analysis.

Exploratory Data Analysis :

### Data 1- New York city Geographical Coordinates Data.

1. In this we load the data and explore data from **newyork_data.json** file.
2. Transform the data of nested python dictionaries into a pandas dataframe.
3. This data frame contains the geographical coordinates of New York city neighborhoods.
4. This data will used to get Venues data from Foursquare.
5. We used geopy and folium libraries to create a map of New York city with neighborhoods superimposed on top.

New York neighborhood visualization
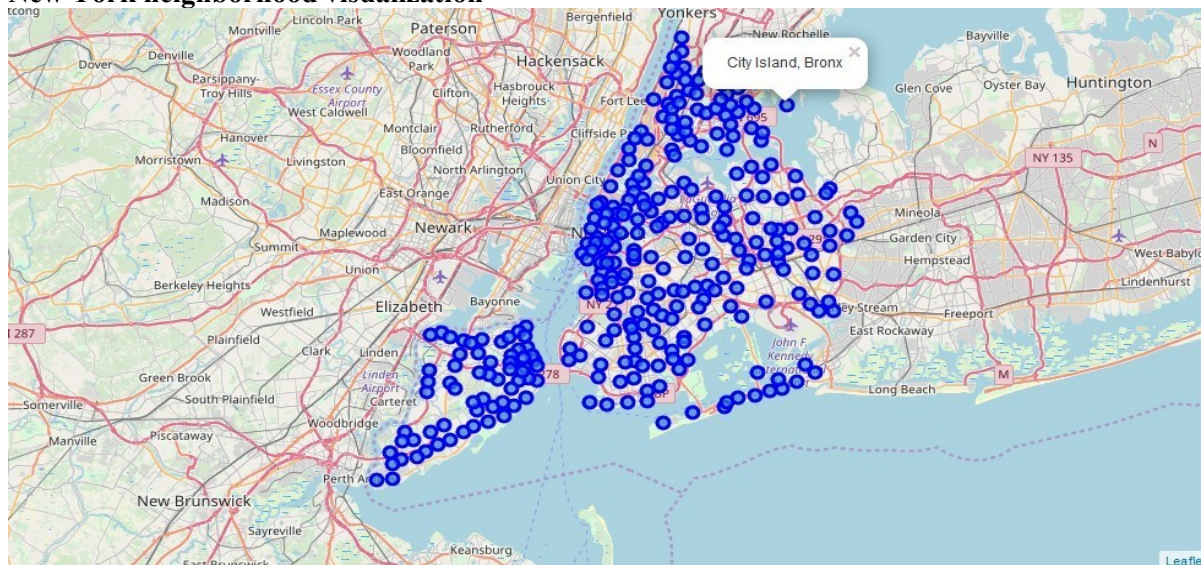
Data 2-

Considering Bronx as a required country in New York City and finding the geographical area and then finding the surroundings plotting on the map as shown
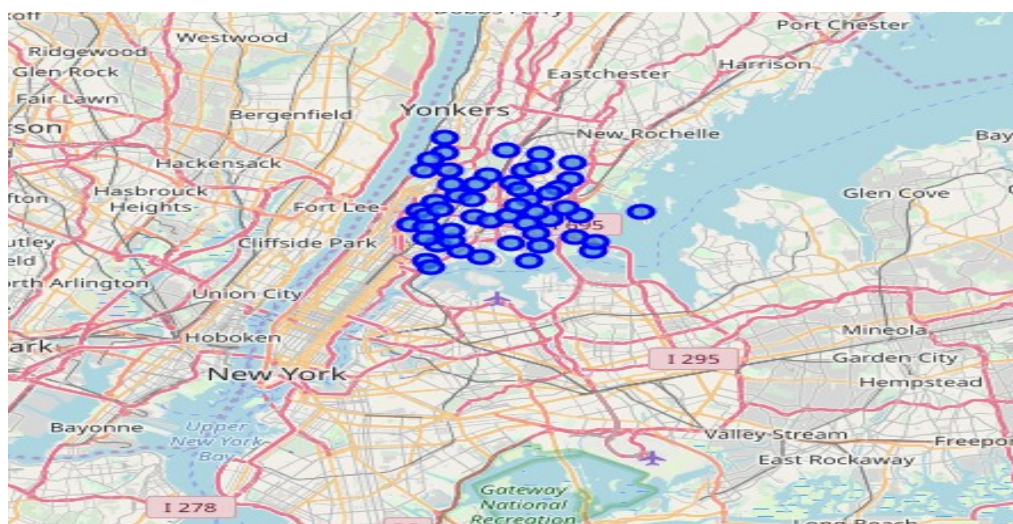
Data 3 :

Now Considering the Wake field as the required city and then finding the latitude and longitude of that area and then fetching popular stores available in that area.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
4	Wakefield	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station

Data 4:

Getting all the venues surrounded by Bronx so that we can find the best venues

Out[50]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
4	Wakefield	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station

RESULTS :

Finally getting the top picks along with the venus filtered on the priority bases and are tabulated for easy understanding

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allerton	Pizza Place	Deli / Bodega	Chinese Restaurant	Cosmetics Shop	Supermarket	Dessert Shop	Bus Station	Spa	Spanish Restaurant	Check Cashing Service
1	Baychester	Donut Shop	Mattress Store	Supermarket	Discount Store	Pizza Place	Convenience Store	Electronics Store	Fast Food Restaurant	Sandwich Place	Bus Station
2	Bedford Park	Chinese Restaurant	Diner	Supermarket	Food Truck	Bus Station	Mexican Restaurant	Sandwich Place	Pizza Place	Deli / Bodega	Pub
3	Belmont	Italian Restaurant	Deli / Bodega	Pizza Place	Bakery	Dessert Shop	Donut Shop	Bank	Coffee Shop	Mexican Restaurant	Eastern European Restaurant
4	Bronxdale	Breakfast Spot	Doctor's Office	Mexican Restaurant	Spanish Restaurant	Supermarket	Garden	Eastern European Restaurant	Chinese Restaurant	Gym	Bank

Graph plotting the hotel along with the surrounded venues:
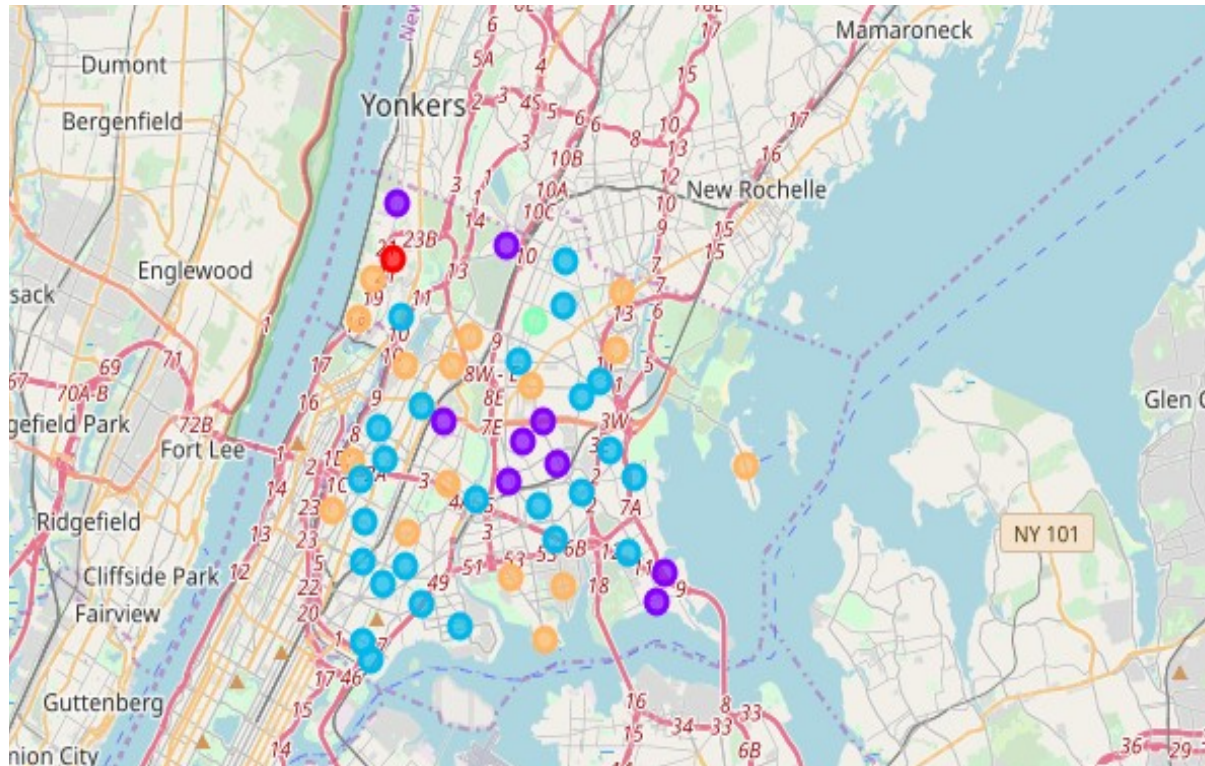
DISCUSSION:

We can further increase this by adding the check-in check-out time, stay duration, price range, external facilities provided like transport if needed and lot more.

CONCLUSION:

Finally We will Generate Map of all Nearby Places along with the 5 Hotels and make 5 clusters of Hotels + Nearby Places using K-Means Algorithm where the user can pick on his choice with the provided data.