2. Data Acquisition, Cleaning and Exploration

New York City (NYC), often called New York (NY), is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2), Situated on one of the world's largest natural harbors, New York City is composed of a total of 5 boroughs and 306 neighborhoods.
in order to analyze New York City to find the best location for Turkish restaurant, the data acquired for this project is a combination of data scraped from a Wikipedia and HTML pages. After accruing and cleaning the data we have four new datasets. We will explore these datasets in order to select the best brought in New York City and then select the best neighborhood in the selected borough to start a new Turkish restaurant.

First dataset

The first dataset is scraped from a Wikipedia page https://en.wikipedia.org/wiki/New_York_City.
The dataset contains the list of New York City boroughs Jurisdiction, population, Gross Domestic Product (GDP), land area and population density.

	Borough	County	Population	Gross Domestic Product billions/US$	Gross Domestic Product per capita(US$)	Land area square miles	Land area square km	persons per sq. mile	persons per square km
0	The Bronx	Bronx	1418207	42.695	30100	42.10	109.04	33867	13006
1	Brooklyn	Kings	2559903	91.559	35800	70.82	183.42	36147	13957
2	Manhattan	New York	1628706	600.244	368500	22.83	59.13	71341	27544
3	Queens	Queens	2253858	93.310	41400	108.53	281.09	20767	8018
4	Staten Island	Richmond	476143	14.514	30500	58.37	151.18	8157	3150

Second dataset

In order to explore competitors in the selected borough in New York City we use the second dataset which scraped from a wikipedia pages:
https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City
https://en.wikipedia.org/wiki/Cuisine_of_New_York_City.

There are two lists in the second dataset one of Michelin starred restaurants 2020 and cuisine of New York City borough of Manhattan

	Name	Borough
0	15 East	Manhattan
4	Agern	Manhattan
5	Ai Fiori	Manhattan
7	Aldea	Manhattan
13	L'Appart	Manhattan
14	Aquavit	Manhattan
15	Aska	Brooklyn
16	L'Atelier de Joël Robuchon	Manhattan
17	Atera	Manhattan
18	Atomix	Manhattan
19	Aureole	Manhattan

Manhattan Michelin starred restaurants 2020

	Borough	Neighborhood	Cuisine
0	Manhattan	Chinatown	Chinese,Vietnamese
1	Manhattan	East Harlem	Puerto Rican, Mexican, Dominican, Chinese-Cuba...
2	Manhattan	East Village	Japanese, Korean, Indian, Ukrainian
3	Manhattan	Greenwich Village	Italian, Middle Eastern
4	Manhattan	Harlem	Italian, African-American, Latin American, We...
5	Manhattan	Koreatown	Korean
6	Manhattan	Nolita	Australian
7	Manhattan	Little Italy	Italian
8	Manhattan	Lower East Side	Puerto Rican, Jewish, Italian, Latin American
9	Manhattan	Murray Hill	Indian, Pakistani and Bangladeshi
10	Manhattan	Upper West Side, Manhattan	Jewish, Chinese-Latino
11	Manhattan	Washington Heights	Dominican, Puerto Rican, Mexican, Jewish
12	Manhattan	Upper East Side	German, Czech, Hungarian

cuisine of Manhattan

Third dataset

In order to segment New York City 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 latitude and longitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572.

New York City 5 boroughs and their neighborhoods geographical coordinates and category data in the third dataset will be utilized as input for the Foursquare API, that will be leveraged to provision neighborhood venues information for selected borough (Manhattan). We will use the Foursquare API to explore neighborhoods in Manhattan. The below is Manhattan dataset obtained using Foursquare API data.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Manhattan neighborhoods and venues geographical coordinates and category

Fourth dataset

The fourth dataset explore Manhattan neighborhoods along with the top 10 most common venues

	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	Battery Park City	Park	Memorial Site	Food Court	Coffee Shop	Building	Gym	Shopping Mall	Burrito Place	Monument / Landmark	Movie Theater
1	Carnegie Hill	Gym / Fitness Center	Café	Bookstore	Pizza Place	Coffee Shop	Italian Restaurant	Gym	Gourmet Shop	Shipping Store	French Restaurant
2	Central Harlem	Chinese Restaurant	African Restaurant	American Restaurant	French Restaurant	Cafeteria	Bookstore	Boutique	Fried Chicken Joint	Spa	Food Truck
3	Chelsea	Ice Cream Shop	Theater	Cupcake Shop	American Restaurant	Speakeasy	Coffee Shop	New American Restaurant	Bar	Café	Nightclub
4	Chinatown	Chinese Restaurant	Hotpot Restaurant	Spa	Sandwich Place	Yoga Studio	Pizza Place	Boutique	Bubble Tea Shop	Spanish Restaurant	Salon / Barbershop

Manhattan neighborhoods top 10 most common venues