Manhattan Eating Preferences

Adrian Tarniceriu

14.08.2019

1. Introduction

The most populous city in the United States, New York City is home to more than 8 million inhabitants (without including the metropolitan area). On top of this, it attracted 65.2 million tourists in 2018, which poses great challenges in terms of infrastructure, supplies, and alimentation. Being the most densely populated borough of New York, these issues are even more important in Manhattan.

This project focuses on the alimentation aspect of the Manhattan region. More specifically, it studies the eating place distribution based on neighborhoods, where the eating places are grouped in three categories: cafes, restaurants, and fast-food places.

This analysis could help visitors find a location more suitable for their preferences, but also provide entrepreneurs with an insight about possible business opportunities, such as opening a new restaurant (or shutting down an existing one).

2. Data

2.1 Data Source

The data used herein covers the Manhattan borough of New York City. This is sub-divided in 40 distinct neighborhoods. The location data, containing the list of neighborhoods and their coordinates (latitude, longitude), is obtained from *https://cocl.us/new_york_dataset*.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Fig 1. Example of location data

For each neighborhood, the venue data is obtained from Foursquare. In total, there were 1182 venues retrieved for the 40 neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue ID	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	4bf58dd8d48988d1ca941735	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	4bf58dd8d48988d102941735	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	4bf58dd8d48988d147941735	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	4bf58dd8d48988d1e0931735	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	4bf58dd8d48988d148941735	Donut Shop
5	Marble Hill	40.876551	-73.91066	Blink Fitness Riverdale	40.877147	-73.905837	4bf58dd8d48988d176941735	Gym
6	Marble Hill	40.876551	-73.91066	TCR The Club of Riverdale	40.878628	-73.914568	4e39a891bd410d7aed40cbc2	Tennis Stadium
7	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	4bf58dd8d48988d1ce941735	Seafood Restaurant
8	Marble Hill	40.876551	-73.91066	T.J. Maxx	40.877232	-73.905042	4bf58dd8d48988d1f6941735	Department Store
9	Marble Hill	40.876551	-73.91066	Starbucks	40.873755	-73.908613	4bf58dd8d48988d1e0931735	Coffee Shop

Fig 2. Example of venue data

2.2 Data Preprocessing

Because the venues can be of different types (eating places, sport, museums, etc.), and we are only interested in the eating places, we will filter out unrelated data. After this filtering, we will have 455 venues left.

Before continuing with the analysis, we do one more pre-processing step: here, we are only interested in cafes, restaurants, and fast-food places. Thus, each food-related Venue category will be assigned to one of these three categories as shown in the table below

Vanue Category contains:	Assigned to:
'restaurant', 'bodega', or 'diner'	Restaurant
'cafe' or 'coffee'	Cafe
'joint', 'bagel', 'pizza', 'breakfast', 'burger', 'burrito', 'creperie', 'fast food', 'pastry', 'sandwich', 'snack', or 'taco'	FastFood

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue ID	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	4bf58dd8d48988d1ca941735	FastFood
1	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	4bf58dd8d48988d147941735	Restaurant
2	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	4bf58dd8d48988d1e0931735	Cafe
3	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	4bf58dd8d48988d1ce941735	Restaurant
4	Marble Hill	40.876551	-73.91066	Starbucks	40.873755	-73.908613	4bf58dd8d48988d1e0931735	Cafe
5	Marble Hill	40.876551	-73.91066	Subway Sandwiches	40.874667	-73.909586	4bf58dd8d48988d1c5941735	FastFood
6	Marble Hill	40.876551	-73.91066	Boston Market	40.877430	-73.905412	4bf58dd8d48988d14e941735	Restaurant
7	Marble Hill	40.876551	-73.91066	SUBWAY	40.878493	-73.905385	4bf58dd8d48988d1c5941735	FastFood
8	Marble Hill	40.876551	-73.91066	Subway	40.877720	-73.905380	4bf58dd8d48988d1c5941735	FastFood
9	Marble Hill	40.876551	-73.91066	Terrace View Delicatessen	40.876476	-73.912746	4bf58dd8d48988d146941735	Restaurant

Fig 3. Example of clean pre-processed data

At this point, all needed data is grouped in one data-frame. Further processing will be described in the following sections (part 2 or week 5 of the final assignment).