Recommend Hotels in Manhattan, NY for Visitors

Author: Meixuan Xu

Date: July 6, 2020

1. Introduction

1.1 Background

New York City is one of the most popular tourist cities. The tourist population in New York has continuously increased in past nine years and welcomed approximately 65.2 million visitors in 2018, including overall 37.9 million visitors who stayed overnight in 2018. When people plan a trip, the first thing they need to do is to book fly and hotel. Many visitors want to book a hotel surrounded by preferred categories of restaurants within walking distance, so they can have nice dining out, and taste different styles and delicious food during their trip.

However, there is no booking website helps visitors to find hotel based on based on the restaurant flavors. For example, Booking.com is a famous website for booking accommodations. This website put together several menus for their customers to search different accommodations separately. If you searched hotels, you got a result of hotel list. If you want to know the surrounding accommodations, such as restaurants and public transport stations, you need to do a lot of extra searches on other search engines like maps.google.com.

1.2 Problem description

This project will leverage foursquare location data to build a recommendation system to help tourist to find hotel in NYC and explore the surroundings of the hotel based on their preferred restaurant flavors.

Suppose a family want to take their vacation in Manhattan, NYC. They want an affordable, comfort hotel which has a lot of preferred style restaurants in a walking distance, so they can easily have nice diners in nearby restaurants during their stay.

From a data scientist's view, this problem is described as a family want a hotel (a GPS point) in a neighborhood that meets below requirements:

1) In a hotel with at least several preferred categories of restaurants.

2) These restaurants are in walking distance.

2. Data Sources and Data Preprocessing

2.1 Data sources

To complete this project, we need two geography data sources, NYC neighborhood geography data and NYC venue geography data.

NYC neighborhood geography data can be downloaded from geo.nyu.edu And 5 boroughs and 306 neighborhoods geography data of New York City have been received from the downloaded file.

NYC venue geography data can be retrieved from a location technology platform - Foursquare.com which provides powerful venue and location

information. We can leverage the free Foursquare API to get all of restaurants and hotel information in every neighborhood in NYC.

2.2 Data Preprocessing

When raw data was downloaded from a website, usually they were just exact what you want. We need do data wrangling to extract the fields that we needed, or maybe we need to create new fields for analysis.

Step 1, since the borough Manhattan is most popular borough in NYC, in this project, I practice a scenario that the visitor wants to find a hotel in borough Manhattan. the NYC neighborhoods geography data has 16 properties. I extracted 4 properties in each record, and then select the Borough Manhattan geography data for next step analysis.

Step 2, Based on the Manhattan neighborhoods geographic data, I used Foursquare API to get all venues for every neighborhood in Manhattan, and then extracted restaurant venues and hotel venues.

Step 3, This step is data standardization. For example, the value of the venue category in the data frame is not standardized. so, I copied the current category and add as a new regional style column and reorganize the venue categories before further analysis. For example, some restaurants resisted their category as "Dim Sum Restaurant", actually this belongs to Chinese style restaurant, so "Dim Sum Restaurant" is replaced with "Chinese restaurant" in the regional style column.

Step 4, In this project, supposed the visitor’s favorite style restaurant are Chinese restaurant, Seafood restaurant, and Japanese restaurant. Thus, we select there three categories of restaurants from the standardized restaurant venues as preferred restaurants.

After data preprocessing, we got preferred restaurant list (see Figure_2_1) and hotel list (see Figure_2_2) in borough Manhattan.

	Neighborhood	Venue	Latitude	Longitude	Category	Regional_Style
0	Marble Hill	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant	Seafood Restaurant
1	Chinatown	Spicy Village	40.717010	-73.993530	Chinese Restaurant	Chinese Restaurant
2	Chinatown	Wah Fung Number 1 Fast Food 華豐快飯店	40.717278	-73.994177	Chinese Restaurant	Chinese Restaurant
3	Chinatown	Da Yu Hot Pot 大渝火锅	40.716735	-73.995752	Hotpot Restaurant	Chinese Restaurant
4	Chinatown	Xi'an Famous Foods	40.715232	-73.997263	Chinese Restaurant	Chinese Restaurant
...	...	...	...	...	...	...
187	Tudor City	251 Ginza Sushi	40.745567	-73.975706	Sushi Restaurant	Japanese Restaurant
188	Tudor City	Docks Oyster Bar	40.749483	-73.975471	Seafood Restaurant	Seafood Restaurant
189	Flatiron	Sugarfish	40.738951	-73.988955	Japanese Restaurant	Japanese Restaurant
190	Flatiron	HALL	40.740260	-73.992324	Japanese Restaurant	Japanese Restaurant
191	Flatiron	Sushi By Bou	40.740883	-73.993525	Sushi Restaurant	Japanese Restaurant

192 rows × 6 columns

Figure_2_1 Preferred restaurant list in Manhattan

	Neighborhood	Venue	Latitude	Longitude	Category
0	Chinatown	Hotel 50 Bowery NYC	40.715936	-73.996789	Hotel
1	Hamilton Heights	Hotel San Fermin B&B	40.822566	-73.944624	Hotel
2	Upper East Side	The Carlyle	40.774413	-73.963301	Hotel
3	Upper East Side	The Surrey	40.774415	-73.963889	Hotel
4	Upper West Side	The Lucerne Hotel	40.783427	-73.978495	Hotel
...	...	...	...	...	...
58	Tudor City	Millennium Hilton New York One UN Plaza	40.750399	-73.969050	Hotel
59	Flatiron	The New York EDITION	40.741286	-73.987358	Hotel
60	Hudson Yards	Equinox Hotel - Hudson Yards	40.754768	-74.001986	Hotel
61	Hudson Yards	YOTEL New York	40.759171	-73.995268	Hotel
62	Hudson Yards	Cachet Boutique Hotel	40.759773	-73.996460	Hotel

63 rows × 5 columns

Figure_2_2 Preferred restaurant list in Manhattan

3. Data Analysis

3.1 Feature variables decided by visitors

The goal of this project is to recommend hotels that meets the user's two requirements: number of preferred categories of restaurants near the hotels and they are in a walking distance.

A hotel with how many preferred restaurants nearby should be recommended? Different visitors might have various numbers. This number is apparently should be decided by visitor. Here I use a variable "least_num" to let user to decide at least number of the preferred restaurants near a hotel to be recommended.

Also, how many meters away is in a walking distance? Different users might have their own feelings about walking distance. So, this project also allows users to decide their own distance as walking distance. A variable "distance_m" was created for the walking distance in meter.

3.2 Calculation

Since I already have the GPS information of both restaurants and hotels. Here I need to download python module geopy.distance to calculate the distance between hotel and restaurant.

When I calculate the distance between hotel and all restaurants, all of preferred restaurants within the "distance_m" have been saved and counted, and sorted.

3.3 Result

Based on the variable "least_num" and variable "distance_m", two scenarios are demonstrated as below.

Scenario 1, when least_num > 5 and distance_m <= 400 meters, we got 14 hotels recommended. See Figure_3_1

	Hotel	Count	Restaurants	Latitude	Longitude
0	Hotel 50 Bowery NYC	24.0	,Spicy Village,Wah Fung Number 1 Fast Food 華豐快...	40.715936	-73.996789
1	The Renwick Hotel, Curio Collection by Hilton	15.0	,Café Zaiya,Kajitsu,Sushi Ryusei,Tempura Matsu...	40.750184	-73.977604
2	citizenM Bowery	14.0	,Spicy Village,Wah Fung Number 1 Fast Food 華豐快...	40.720599	-73.993574
3	Mercer Hotel	11.0	,Tomoe Sushi,Blue Ribbon Sushi,Lure Fishbar,Bo...	40.724828	-73.998553
4	Shelburne Hotel & Suites by Affinia	11.0	,Kajitsu,Sushi Ryusei,Tempura Matsui,Momosan R...	40.748419	-73.977940
5	AKA United Nations	10.0	,KaoruMC,Crave Fishbar,Yama 49,Aburiya Kinnosu...	40.752640	-73.971340
6	The William	10.0	,Koi New York,Café Zaiya,Sushi Ginza Onodera,K...	40.750673	-73.980077
7	The Langham, New York, Fifth Avenue	9.0	,Koi New York,Café Zaiya,Sushi Ginza Onodera,Z...	40.750144	-73.983532
8	The Bernic Hotel, Tapestry Collection by Hilton	9.0	,KaoruMC,Crave Fishbar,Yama 49,Aburiya Kinnosu...	40.754432	-73.972685
9	Andaz 5th Avenue - a concept by Hyatt	8.0	,Xi'an Famous Foods,Koi New York,Café Zaiya,Su...	40.752829	-73.980942
10	Soho Grand Hotel	8.0	,Blue Ribbon Sushi,Boqueria,SUGARFISH by sushi...	40.723911	-74.005224
11	Millennium Hilton New York One UN Plaza	8.0	,KaoruMC,Yama 49,Aburiya Kinnosuke,Pescatore S...	40.750399	-73.969050
12	Sofitel New York	6.0	,Málà Project,Xi'an Famous Foods,Koi New York,...	40.755787	-73.981762
13	The Algonquin Hotel, Autograph Collection	6.0	,Málà Project,Xi'an Famous Foods,Koi New York,...	40.755993	-73.982317

Figure_3_1

Scenario 2, when least_num > 3 and distance_m <= 400 meters, we got 21 hotels to recommend. Figure_3_2 shows the first 10 records

	Hotel	Count	Restaurants	Latitude	Longitude
0	Hotel 50 Bowery NYC	24.0	,Spicy Village,Wah Fung Number 1 Fast Food 華豐快...	40.715936	-73.996789
1	The Renwick Hotel, Curio Collection by Hilton	15.0	,Café Zaiya,Kajitsu,Sushi Ryusei,Tempura Matsu...	40.750184	-73.977604
2	citizenM Bowery	14.0	,Spicy Village,Wah Fung Number 1 Fast Food 華豐快...	40.720599	-73.993574
3	Mercer Hotel	11.0	,Tomoe Sushi,Blue Ribbon Sushi,Lure Fishbar,Bo...	40.724828	-73.998553
4	Shelburne Hotel & Suites by Affinia	11.0	,Kajitsu,Sushi Ryusei,Tempura Matsui,Momosan R...	40.748419	-73.977940
5	AKA United Nations	10.0	,KaoruMC,Crave Fishbar,Yama 49,Aburiya Kinnosu...	40.752640	-73.971340
6	The William	10.0	,Koi New York,Café Zaiya,Sushi Ginza Onodera,K...	40.750673	-73.980077
7	The Langham, New York, Fifth Avenue	9.0	,Koi New York,Café Zaiya,Sushi Ginza Onodera,Z...	40.750144	-73.983532
8	The Bernic Hotel, Tapestry Collection by Hilton	9.0	,KaoruMC,Crave Fishbar,Yama 49,Aburiya Kinnosu...	40.754432	-73.972685
9	Andaz 5th Avenue - a concept by Hyatt	8.0	,Xi'an Famous Foods,Koi New York,Café Zaiya,Su...	40.752829	-73.980942

Figure_3_2

4. Results Visualization

Now we've found all of the recommended hotel list, let's cast the results on the map. The recommended hotels are green with hotel icon, and the not recommended hotels are gray with hotel icon. Preferred restaurants are orange with bold circle marker. Data visualization helps user better understand data than just data. Figure_4_1 shows the whole picture of the recommended hotels and preferred restaurants. Figure_4_2 shows the enlarged part.

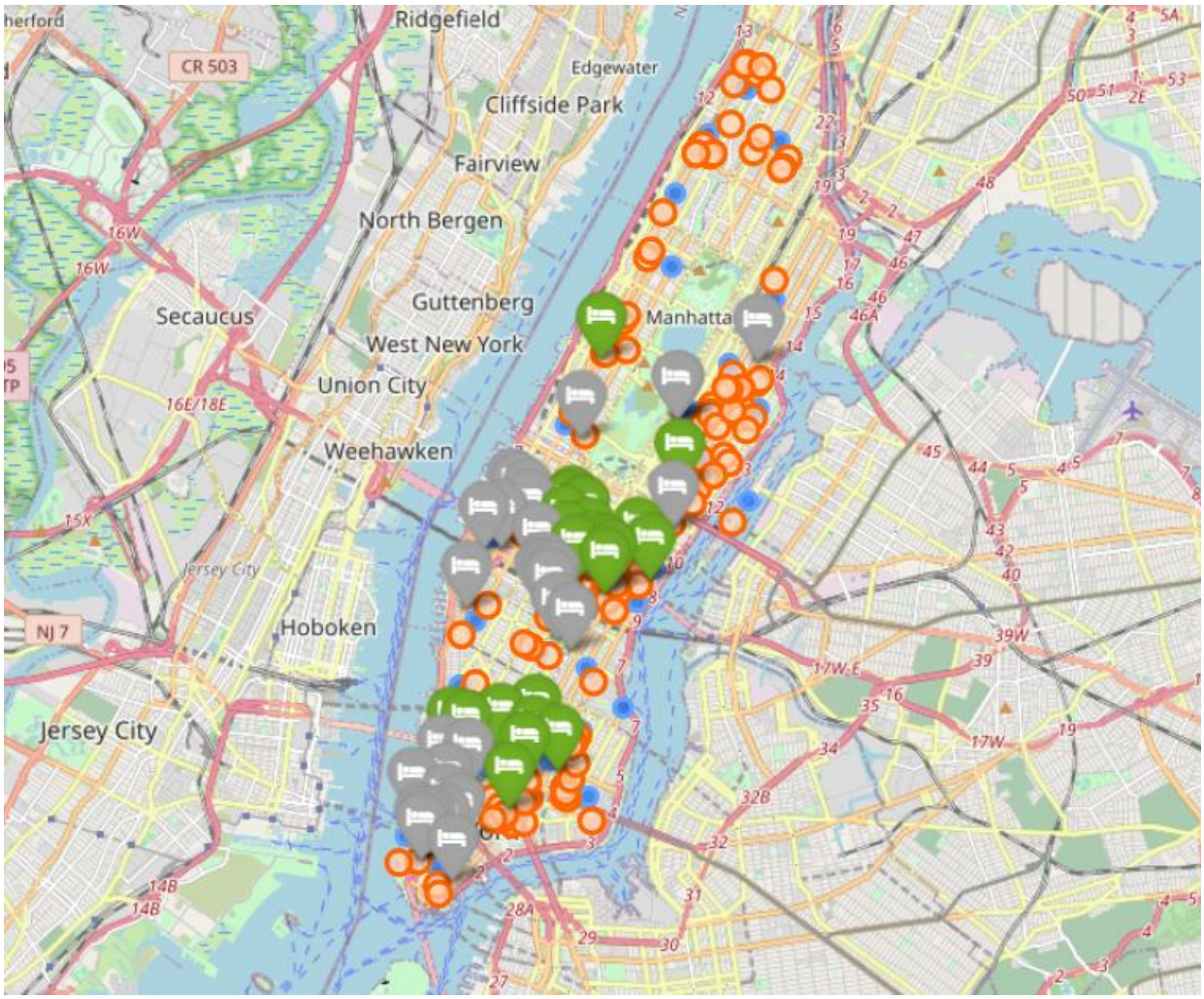

Figure_4_1 shows the whole view of the recommended hotels

Figure_4_2 shows the enlarged part of result

5. Discussion

Although visitor usually only need one hotel during his/her visit, however it has a lot of recommendations in this project. There are 15 recommended hotels in scenario 1 and 24 recommended hotels in scenario 2. This is because besides dining, user might have other considering, such as price, transportation station nearby. So, this project just provides a recommended hotel list for user to pick the one meet use's other requirements.

In this project, I chose free Foursquare API to retrieve restaurant venues and hotel venues information. Comparing with the results from map.google.com, I got less hotels from Foursquare API then from map.google.com. Since this project is only for study use not for business purpose, it's acceptable that we did not get all of the hotel records in Manhattan.