Analysis of cafes in Manhattan

A business analysis for setting-up new cafes

New York - 2 January 2019

Dona Cherian

Introduction

Manhattan is the economic and administrative capital of New York and one of the original counties among the 13 colonies that originally formed United States of America. Often referred as the cultural, financial, media and entertainment capital of the world Manhattan one of the hottest places for blooming business. Being one of the most densely populated cities in the world and containing one of the biggest moving population Manhattan has an ever increasing demand for Cafes and restaurants. In this short report I aim to present an analysis for the possibility of starting a new cafe in the city of Manhattan and elaborate the most favourable places for the business.

Overview of the city of Manhattan

Often referred as 'the City' Manhattan has one of the largest population in the world. In this analysis we are going to explore the 40 neighbourhoods of Downtown and Midtown Manhattan limited by the spatial co-ordinates 40.7900869, -73.9598295. A list of neighbourhoods under consideration are given below.

List of Neighbourhoods in Manhattan
Marble Hill
Chinatown
Washington Heights
Inwood
Hamilton Heights
Manhattanville
Central Harlem
East Harlem
Upper East Side
Yorkville
Lenox Hill
Roosevelt Island
Upper West Side
Lincoln Square
Clinton
Midtown
Murray Hill
Chelsea

Greenwich Village
East Village
Lower East Side
Tribeca
Little Italy
Soho
West Village
Manhattan Valley
Morningside Heights
Gramercy
Battery Park City
Financial District
Carnegie Hill
Noho
Civic Center
Midtown South
Sutton Place
Turtle Bay
Tudor City
Stuyvesant Town
Flatiron
Hudson Yards

The City has an average population density of 66,940 people per square mile. The City is popular for its rich culture of restaurants and according to a study by Crain's New York in 2017 the City has 309.6 restaurants per 100,000 people. However this number comes well below many other cities. This shows a growing demand for good cafes and restaurants in and around the city of New York.

Background

The city is known for its rich cultural as well as culinary diversity and is house for a wide variety of restaurants and cafes. It is hence very important to consider a couple of factors before starting a new business in the culinary sector. Some of the important factors to consider are

1. number of existing cafes in various neighbourhoods

2. popularity of various cafes in different neighbourhoods
3. locations of interest in various neighbourhoods to estimate the floating population and
4. population density of neighbourhoods.

Problem Description

The most important point to consider is to identify a locality with high population density with lower number of popular cafes. It is important to consider that a place with high density of popular cafes might not be the best choice. Similarly a locality with low number of cafes but low population density is not a good choice as well. Hence it is important to identify the distribution of cafes in different localities with respect to its population. First and foremost we will have to identify a method to segregate and cluster various neighbourhood in the City based on location co-ordinates. Once the city is divided based on neighbourhoods we will have to retrieve the information of various cafes and their popularity across the city. In a later step the population density of various neighbourhoods are analysed together with the cafes in the corresponding locality to identify a distribution map of cafes. A combined analysis of Places of interest and population density is required to identify the right locality for starting the new business.

Available data from internet

We will have to use the json data of the city of New York to retrieve the location data for Manhattan. Fortunately this is available here and we can readily use it for our purpose. Secondly we will need information of various cafes and popular places of interests in the city and this can be retrieved using FourSquare developer API. The data will be retrieved using the personal login in Developer API and sorted and formatted. Lastly the population density of various Boroughs in the city of New York is available in the internet as well. We will have to scarp the data from different web pages and combine it with the data available from FourSquare for our requirement. In the final form the data will be a list of neighbourhoods sorted with respect to the number of cafes, popular places and its population density.

Example data

Location data

	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

Population density data

Borough	Area, km2	Pop. Census 2010	Pop/ km2	Neighbourhoods
anhattan CB 1	4.45	34,420	7,735	Battery Park City, Financial District, TriBeCa
Manhatta n CB 2	4.01	93,119	23,222	Chinatown, Greenwich Village, Little Italy, Lower East Side, NoHo, SoHo, West Village
Manhatta n CB 3	4.56	1,64,407	36,054	Alphabet City, Chinatown, East Village, Lower East Side, Two Bridges
Manhatta n CB 4	5.41	87,479	16,170	Chelsea, Clinton
Manhatta n CB 5	4.25	44,028	10,360	Midtown
Manhatta n CB 6	3.55	1,36,152	38,353	Gramercy Park, Kips Bay, Murray Hill, Peter Cooper Village, Stuyvesant Town, Sutton Place, Tudor City, Turtle Bay, Waterside Plaza
Manhatta n CB 7	5.46	2,07,699	38,040	Lincoln Square, Manhattan Valley, Upper West Side
Manhatta n CB 8	5.13	2,17,063	42,312	Lenox Hill, Roosevelt Island, Upper East Side, Yorkville
Manhatta n CB 9	3.91	1,11,724	28,574	Hamilton Heights, Manhattanville, Morningside Heights
Manhatta n CB 10	3.63	1,07,109	29,507	Harlem, Polo Grounds
Manhatta n CB 11	5.75	1,17,743	20,477	East Harlem, Randall's Island, Spanish Harlem, Wards Island
Manhatta n CB 12	7.64	2,08,414	27,279	Inwood, Washington Heights

This formatted and combined data will be used for the analysis of our problem using k-mean clustering.

Closing statement

A detailed analysis of the above problem with available data will be presented in the repository in one weeks time.