# 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 |

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

## Methodology

In an attempt to search for an appropriate location for starting a cafe in Manhattan, the data described in the above section is throughly analysed. At first the location data is downloaded and JSON files are converted into pandas dataframe. Using Folium maps a map of the City is constructed based on different neighbourhoods.

In the second part of the analysis the list of existing cafes in the region under consideration is retrieved from FourSquare and the relevant data is converted in to pandas data from.The tailored data was used for further analysis. The table created with the cafe names classified according to postal codes and location co-ordinates.

In the next step the number of cafes in each location are classified based on the postal codes to estimate the density of cafes in each location. After this one hot encoding is done to convert the categorical variables, in this case the names of cafes. The mean of frequency of occurrence of cafes are grouped by postal codes. the data is not ready for analysis. The preprocessed data is now ready to be used for K-mean clustering.

K-mean clustering is an efficient method to for pattern recognition, data segregation and comparison. In our problem the aim is to identify the pattern of cafes retribution among various neighbourhoods in Manhattan. Hence k-mean can be used as the tool here.

## Results

k-mean analysis is performed assuming 5 clusters and the results shows that some of the locations have a higher density of cafes. How ever considering the population density of the locality too, these places can be identified as one of the favourable spots for cafes.

## Observations and recommendations

Based on the analysis of the cafe distribution density and the population pertaining to different neighbourhood regions  under postcode 10023, 10024, 10025 are most favourable. This zip codes comprises the neighbourhood Upper West Side and our analysis points that this is one of the favourable locations for business.

## Conclusion

In this project we aimed to analyze a favourable spot for starting a new cafe in Manhattan and considered various data available online. using k-mean clustering cafes are clustered and trained to reach the result. According to our analysis Upper West Side neighbourhood is best suitable.