

A description of the problem and a discussion of the background. (15 marks)

1. Introduction

1.1 Background

Manhattan, New York City. It is the most densely populated of the five boroughs in New York City. Manhattan is also described as the cultural, financial, media and entertainment capital of the world. It is of no surprise that businesses would want to set up shop in Manhattan.

Derek, a good friend of mine is deciding to open up an office in Manhattan. However, he is clueless of where he should open it. After a brief discussion with his colleagues, he decided that the most important factor that should be taken into account is the number of food places around the area. Since his colleagues and himself love to eat and would like to have a wide variety of restaurants and joints to choose from during lunch time, he would want to have an office with the most food places around the neighborhood in Manhattan. As Derek is aware of my newly acquired data analytics skills, he has approached me to come up with a recommendation of which neighborhood that he should open up his office in.

1.2 Interest

Any individual that would want to set up an office or would like to know which area in Manhattan has the most food places around would find this analysis useful.

A description of the data and how it will be used to solve the problem. (15 marks)

2. Data Acquisition

2.1 Data Sources

We would need various sets of data to ensure accurate analysis. First, we would need a dataset of New York neighborhoods which is already available to us from a previous exercise that can be obtained from this link:
https://cocl.us/new_york_dataset.

In the dataset, there are features that we would need like the boroughs and neighborhoods of New York so that we can find which neighborhoods reside in Manhattan. Next, we would need to make some foursquare api calls to obtain the data of the neighborhood venues in Manhattan. We would need to extract the venue category, longitude and latitudes of the venues.

A combination of the New York dataset from the link and foursquare API would allow us to obtain the longitude and latitudes of the different venues in the neighborhoods in Manhattan. Thus, it allows us to have enough information to identify clusters of food places further in our analysis.

2.2 Data Cleaning

After acquiring the data from the link, what we have is a set of features of the different locations around New York City. However, a lot of these data are redundant and the first thing I did was to extract only what I needed from the dataset like the borough, neighborhoods, latitude and longitude and put them into a data frame. The next step was to filter the data and extract the neighborhoods, latitude and longitude residing in the borough Manhattan. Now, we have a cleaned data frame of Manhattan neighborhoods and their longitude and latitudes (Table 1).

	Borough	Neighborhood	Latitude	Longitude
6	Manhattan	Marble Hill	40.876551	-73.910660
100	Manhattan	Chinatown	40.715618	-73.994279
101	Manhattan	Washington Heights	40.851903	-73.936900
102	Manhattan	Inwood	40.867684	-73.921210
103	Manhattan	Hamilton Heights	40.823604	-73.949688

Table 1: Cleaned data frame of Manhattan Neighborhoods with their Latitude and Longitudes.

Next, we crawled the internet with foursquare API for the venues in Manhattan. The dataset consists of many pieces of information that we do not need. Thus, we extracted only the information we need once again and merged it with the data frame that we made previously (Table 2).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance	Venue Latitude	Venue Longitude
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	This spot is popular	Yoga Studio	376	40.876844	-73.906204
1	Marble Hill	40.876551	-73.91066	Arturo's	This spot is popular	Pizza Place	240	40.874412	-73.910271
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	This spot is popular	Diner	452	40.880404	-73.908937
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	This spot is popular	Pizza Place	516	40.879435	-73.905859
4	Marble Hill	40.876551	-73.91066	Starbucks	This spot is popular	Coffee Shop	441	40.877531	-73.905582

Table 2: Merged dataset of Manhattan neighborhoods and different venues.