Finding the best place to open a Chinese Restaurant in Manhattan, NY (Report)

Raafat El Jamal

10-Mar-2021

1 Introduction

1.1 Background

Manhattan is a great place for starting a new business as many tourists visit it frequently as it has almost everything a tourist might need such as shops, restaurants, historical places, entertainment, and hotels. It has a strong economy and high population density. There is a high competition between business owners which lets anyone think about starting business there. In this project the investor is wants to start a Chinese Restaurant. It is not easy to find the best place in Manhattan and that is what we will work on.

1.2 Problem

I think the best way is to look at Neighborhoods that do not have Chinese restaurants so that this will be the first Chinese restaurant there which gives an advantage for a good start with less competition, then the highest rated venues there, then select the densest place which gives another advantage as this location is highly demanded. This is how to keep squeezing the search to find the golden location.

1.3 Interest

Any investor would be interested in the results as it saves a lot of research time.

2 Data

2.1 Data Sources

The main dataset is the New York data with all Boroughs and Neighborhoods which can be found here. I extracted Manhattan's data from it then from Foursquare API I explored all Neighborhoods in Manhattan to get Venues' names, categories, IDs, longitudes, and latitudes. After that I extracted Venues' ratings from Foursquare API to build the final table required for analysis.

2.2 Data Cleaning

New York data were extracted from the json file and converted to a dataframe, then I took specific columns (Borough, Neighborhood, Latitude, Longitude). Manhattan data were extracted from the dataframe, then connected to Foursquare API to explore Neighborhoods in Manhattan. Downloaded the json file from Foursquare API then extracted from it (Venue name, category, ID, longitude, latitude) and added them to the Manhattan dataframe as additional columns. I deleted all the Neighborhoods that have Chinese restaurants in it and kept the rest to explore, then based on the Venue ID I extracted Venue Ratings from Foursquare for each venue and added the ratings to the table. This operation was complicated as Foursquare limits API calls to ratings per day, so I divided a copy of the main table into three parts and ran the API call for each part in a separate day, then merged the three ratings results in one file and took the ID and rating columns only and removed the duplicates, then merged the ratings to the main table again and removed venues with empty ratings. Finally, the dataframe was ready for analysis.

2.3 Data Readiness

The final table has Manhattan venues with their coordinates, categories, neighborhoods and ratings, these data are required for my analysis to find the golden location for starting a new Chinese restaurant.