

Predicting the best location for opening a Restaurant in New York City

IBM Applied Data Science Project

1. Introduction

1.1 Background

Opening a new business in any field is a big challenge with heavy competition everywhere. With various trending customer preferences, one must explicitly adapt to the needs of these customers. This is where data science comes into picture where an informed decision on opening any business is advantageous. The need to leverage the data covering current customer preferences of restaurants and trending food categories aspects can be very crucial for opening a restaurant. One such way can be to analyse various locations that include several restaurants and provide details as to which locations can work for a particular food restaurant category (Italian, Japanese, Indian, etc.).

1.2 Problem

In this project, an attempt to find the best location for opening a restaurant business is carried out. The expected outcomes are:

1. Determine the best restaurant category (Chinese, Italian, Mexican, etc.) for a new business
2. Determine the best neighbourhood to open the restaurant

1.3 Interest

This project focuses on location analysis of opening a restaurant, hence, any stakeholders or entrepreneurs interested in doing a new business such as food restaurants which is a very demanding market. An overview of the venue feature analysis can help the stakeholders to open their restaurants in a particular location of high demand.

2. Data Acquisition and Cleaning

The city under consideration is NYC (New York City), which is divided into 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood. Luckily, this dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

The relevant data of boroughs and neighbourhoods is extracted from the features key. The data thus obtained was then transformed to a Pandas dataset.

The Manhattan borough is selected for its most popularity with 3071 business venues. Amongst these businesses, we considered Restaurant business with 816 business venues. Next, all the existing restaurant businesses and their different categorical data is obtained from **Foursquare Location API Dataset**. The data consists of the location, restaurant names, neighbourhood names, and restaurant categories. There are also over 76 unique restaurants to analyse and setup

the most favourable restaurant category in NYC according to the neighbourhood location trends and advantages.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue_Category	Venue ID
0	Marble Hill	40.876551	-73.910660	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant	4b9c9c6af964a520b27236e3
1	Marble Hill	40.876551	-73.910660	Boston Market	40.877430	-73.905412	American Restaurant	585c205665e7c70a2f1055ea
2	Chinatown	40.715618	-73.994279	Kiki's	40.714476	-73.992036	Greek Restaurant	5521c2ff498ebe2368634187
3	Chinatown	40.715618	-73.994279	The Fat Radish	40.715323	-73.991950	English Restaurant	4c9d482e46978cfa8247967f
4	Chinatown	40.715618	-73.994279	Da Yu Hot Pot 大渝火锅	40.716735	-73.995752	Hotpot Restaurant	5d992946dbf3ca0008d05211

Foursquare Specific Venue Dataset is also used that contains more exploratory data containing the restaurant ratings, number of reviews, total likes, and price category. The data is further cleaned to the features of the top 5 restaurants that are most common in NYC.

Feature Analysis: The features that were considered for the specific neighbourhood location analysis were: Average Ratings, Total Reviews, Total Likes, and the Price Category. These features are a good start for indicating which restaurant categories are most popular and unpopular amongst the customers. This data was gathered using the premium call from the Foursquare Venue API dataset for each venue.

	Neighborhood	Venue	Venue_Category	Venue ID	Price_Category	Ratings	Total_Reviews	Likes
0	Marble Hill	Land & Sea Restaurant	Seafood Restaurant	4b9c9c6af964a520b27236e3	Moderate	7.5	70	42
1	Marble Hill	Boston Market	American Restaurant	585c205665e7c70a2f1055ea	Moderate	7.1	5	4
2	Chinatown	Spicy Village	Chinese Restaurant	4db3374590a0843f295fb69b	Cheap	8.2	690	500
3	Chinatown	Wah Fung Number 1 Fast Food 華豐快餐店	Chinese Restaurant	4a96bf8ff964a520ce2620e3	Cheap	8.2	281	192
4	Chinatown	Xi'an Famous Foods	Chinese Restaurant	5894c9a15e56b417cf79e553	Cheap	8.5	138	104