Opening a new Dim Sum Restaurant in Manhattan

IBM Data Science Certificate Capstone Project

Prepared by:

Henry Li

09/2020

Introduction:

2020 proved to be a tough year for the restaurant industry. Many of the Chinese restaurants did not survive the four-month shutdown brought by Covid-19 and closed their businesses permanently. This fact posed many challenges for New Yorkers to dine out after Covid-19 but also brought about a lot of opportunities for new players to join the restaurant industry.

Dim Sum is one of the most popular choices if you feel like having some Chinese food. It originated in Southern China hundreds of years ago and served small dishes from mid-morning to mid-afternoon. Tim Ho Wan in Manhattan serves Dim Sum and is very popular, but they are faced with short of cash problem as well. Therefore, opening some new Dim Sum places is actually a pretty good investment/entrepreneurial choice.

One of my friends is looking at a business opportunity of opening a new Dim Sum restaurant in Manhattan, New York City. The goal of this project is to leverage data science and machine learning techniques to help him determine the best possible neighborhood to open the restaurant.

Business Problem:

This project is designed to help investors/entrepreneurs to navigate through the business opportunity of opening a new Dim Sum restaurant in Manhattan, New York City. We will explore the location data by data visualization as well as machine learning techniques to determine which neighborhood in Manhattan will be the best location for opening new Dim Sum parlors.

Data:

In order to solve this problem, the following data would be needed:

1. Neighborhood location data (including Latitudes and Longitudes) for Manhattan Neighborhoods, obtained from Wikipedia.
2. Neighborhood demographics data from NYU Furman Center https://furmancenter.org/neighborhoods.
3. Venue data provided by Foursquare API. In order to determine which neighborhood is the most suitable one, I need to look at existing restaurants, local demographics, traffic information, residential vs business usage, etc.