

Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Restaurant in HongKong, China

By: Rachel Gan

October 28

1. Introduction

1.1 Background

For most people who want start a business, opening a restaurant is a good choice, because the catering industry is related to everyone and has a huge consumer base. Additionally, the cost and risk of opening a restaurant is very low.

HongKong has a population of 7.5 million and a population density of 6,300 people per square kilometer, HongKong is also a multi-ethnic society, which means that the catering industry has large demands and diverse needs.

Before opening a restaurant, we need to take a lot of factors into consideration to help us make a better decision. Particularly, the location of the shopping mall is one of the most decisions that will determine whether the restaurant will be a success or a failure.

1.2 Business Problem

The purpose of this top project is to analyze and select the best location in Hong Kong to open a restaurant. The project uses data science methods and machine learning techniques such as clustering to provide solutions to the business problem: If an entrepreneur wants to open a restaurant, where would you recommend the location is?

1.3 Interest

Entrepreneurs or companies that open restaurants may be interested in the location of restaurants, and other real estate agents may also be interested in.

2. Data

2.1 Required data

- * List of neighborhoods in HongKong.
- * Latitude and longitude coordinates of those neighborhoods.
- * Venue data

2.2 Data sources

This wikipedia page (https://en.wikipedia.org/wiki/Category:Places_in_Hong_Kong) contains the list of neighborhoods in HongKong, with a total 67 neighborhoods. We will use Python requests and BeautifulSoup package to extract the neighborhoods data from the wiki page. Then we will

get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude of each neighborhood. After that, we are going to use Foursquare API to get venue data of all neighborhoods, which will provide many categories of the venue data, we will focus on the restaurant categories in order to help us solve the further problem.

Additional, we will also use many other data science skills to help us solve problems, such as we will use machine learning techniques like k-means clustering, data visualization packages like Folium, etc...