

1. Introduction

Scenario:

I am a data scientist residing in Hong Kong. One day, a friend of mine asked me, where is the most suitable place for her to open her own restaurant.

Business Problem:

The challenge is to find a suitable district to open a new restaurant, where the market is not saturated yet.

2. Data Description and Processing

2.1. Data description

The following data is required to answer the issues of the problem:

- List of districts and regions in HK
- Venues for each district (that can be clustered)
- Coordinates for the districts and venues (for mapping purpose)

The data will be used as follows:

- Use Foursquare and geopy data to map top 10 venues for all HK districts and clustered in groups (as per Course LAB)
- Data will be searched in open data sources if available, such as Wikipedia.

2.2. Data cleaning

The district data is scrapped from the Wikipedia using BeautifulSoup. We dropped some unnecessary columns, getting the district names and the corresponding regions. After that, we use geopy to get the coordinates of each district. There are 18 districts in total, but one of them lack information, and we simply dropped it.

| | District | Area | Region | lat | lng |
|----|---------------------|--------|------------------|-----------|------------|
| 0 | Central and Western | 12.44 | Hong Kong Island | 22.285345 | 114.150452 |
| 1 | Eastern | 18.56 | Hong Kong Island | 22.273280 | 114.233118 |
| 2 | Southern | 38.85 | Hong Kong Island | 22.145000 | 114.093200 |
| 3 | Wan Chai | 9.83 | Hong Kong Island | 22.277465 | 114.173026 |
| 4 | Sham Shui Po | 9.35 | Kowloon | 22.330095 | 114.160940 |
| 5 | Kowloon City | 10.02 | Kowloon | 22.331013 | 114.190366 |
| 6 | Kwun Tong | 11.27 | Kowloon | 22.312126 | 114.226501 |
| 7 | Wong Tai Sin | 9.30 | Kowloon | 22.342063 | 114.194388 |
| 8 | Yau Tsim Mong | 6.99 | Kowloon | 22.318775 | 114.161839 |
| 10 | Kwai Tsing | 23.34 | New Territories | 22.350627 | 114.184916 |
| 11 | North | 136.61 | New Territories | 22.382249 | 114.272828 |

Using geopy, we get all the venues in each district. We then apply one-hot coding to the dataset, and calculate the frequency of each venue. Finally the top 10 venues are sorted according to frequency.

| | Neighbourhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---------------------|-----------------------|-----------------------|-----------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 0 | Central and Western | Café | Japanese Restaurant | Coffee Shop | Cocktail Bar | French Restaurant | Indian Restaurant | Chinese Restaurant | Wine Bar | Bar | Italian Restaurant |
| 1 | Eastern | Bus Stop | Zoo | Gastropub | Furniture / Home Store | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Flea Market | Fish & Chips Shop |
| 2 | Kowloon City | Thai Restaurant | Dessert Shop | Café | Chinese Restaurant | Fast Food Restaurant | Cha Chaan Teng | Coffee Shop | Halal Restaurant | Bakery | Noodle House |
| 3 | Kwai Tsing | Scenic Lookout | Sushi Restaurant | Nature Preserve | Trail | Zoo | Fish & Chips Shop | French Restaurant | Food Truck | Food Court | Flea Market |
| 4 | Kwun Tong | Chinese Restaurant | Café | Coffee Shop | Cha Chaan Teng | Hong Kong Restaurant | Japanese Restaurant | Sushi Restaurant | Fast Food Restaurant | Restaurant | Department Store |