Wentao,Cui

December 21, 2020

# Coursera Capstone

## IBM Data Science Capstone Project

## Analyzing subway economy, and opening a new “hot pot restaurant” in Beijing, China

## Introduction

Hotpot also known as soup-food, is one of the most social of dining formats. There are many regional version of hot pot throughout China. For the cities in northern China, like Beijing, lamb is a common choice. The assumption is like this, one of my best friends want to open a new traditional old Beijing style hotpot restaurant in Beijing. Because I was born in Beijing, and be very familiar with Beijing, I am asked to provide suggestions about where are proper places to open this restaurant.

Because Beijing is in poor traffic conditions, public transportation is the first choice for most people. So I want to do analysis which focus on public transportation ,especially for subway, and try to find best places around the subway station.

## Business Problem

The objective of the capstone project is to analyze and select the best locations, which are used to open “hotpot” restaurant in Beijing. Data science methodology and machine learning algorithms will be used and aimed to provide solutions for this.

## Target Audience

The entrepreneur who want to find proper location to open “hotpot” restaurant in Beijing. The output of project could also be meaningful for other business decision, e.g. is it better to open “convenience store“ near the “GuoMao” subway station and so on.

## Data

- Select POIs of subway station in Beijing from Baidu map, and each POI distribute in various regions (district) in Beijing
- Get longitude and latitude of each POI based on coordinate system of Baidu
- Save POI's name, longitude and latitude in .csv file as dataset
- Retrieve and explore venues, which locate in 3 kilometers near each subway station by using Foursquare APIs
- Foursquare APIs provide many categories of venue, for this project, I am particularly interested in the “Chinese restaurant” and “Hot Pot”.

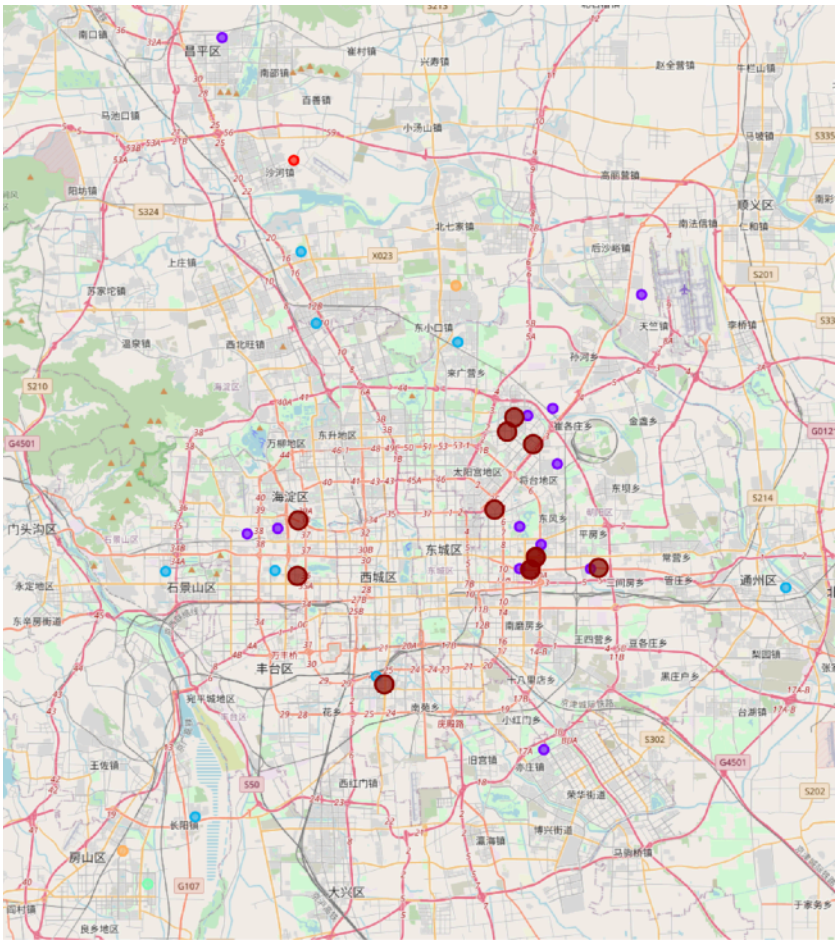
## Methodology

- As above “DATA” section mentioned, select subway POIs and create dataset .csv file manually. The .csv file includes “Subway station name”, “longitude” and “latitude” of each subway station
- Use Foursquare APIs to get top 100 venues that within a radius 3000 meters around each subway station
- Filter all “Hotpot Restaurant” and “Restaurant” type venues, since we are just analyzing “Restaurant” data
- Perform clustering on the “Restaurant” data by using k-means clustering algorithm
- Execute observation and  analysis for each cluster, and try to find best locations

## Results

As below k-means output shows that,

- Because the output of Foursquare APIs does NOT include so much “hotpot” type data, so I added each found “hotpot” type venue into map with big dark red color circle marker.  Most found “hotpot” data locate in cluster 1
- Most found restaurants data locate in cluster 1 and 2.
- Cluster 1 (purple color) locate in east and northeast areas of Beijing, which include few large commercial and business districts, such as “CBD” and “WangJing”
- Cluster 2 (light blue) mainly locate in north and northwest areas of Beijing, which include few large-scale residential areas, such “Tiantongyuan”, “BeiYuan”,”HuiLongGuan”
- Cluster 3,4 mainly locate in “FangShan” and the north of “TianTongYuan” areas, which far from downtown of Beijing
- Cluster 0 could be some kind of outlier, possibly because of dataset, and is ignored in this project



## Limitations and About Future Research

There are others import factors, such as average house price, population density, income of residents that could also affect the location selection. Future's research could make use these factors to obtain more interesting results.

## Conclusion

In this project, we have simply experienced whole process of data science, which including "identify business problem", "collect and prepare data", "select proper model" and "model output analysis".

For this project, best location for opening "hotpot" restaurant should be meet below criterion,

- In order to avoid competition, the best area should include as few "hotpot" restaurants as possible. For this consideration, cluster 2 could be better option.
- Because most "Chinese restaurant" provide "hotpot" services, so the best areas should include as few "Chinese restaurant" as possible. For this consideration, cluster 3 and 4 could be better selections.

## References

Foursquare Developers Documentation. *Foursquare*. Retrieved from https://developer.foursquare.com/docs

Baidu map system.