Where is the Best place for new restaurant in SEOUL, Korea?

IBM Applied Data Science Capstone Project

Samuel Sung

1. Introduction

   1.1 Background

SEOUL is the biggest city in KOREA. These days, the number of foreign tourists visiting Seoul is increasing every year. There are a growing number of people looking for not only Korean food but also various world foods.

   1.2 Problem

Mr ABC is a very talented Turkish chef. He is looking for a place opted in his new Turkish restaurant. He would like to find place that restaurants in various categories is crowded by customers, visitors. He wants to find place that customers can access by public transportation like subway, or bus, as possible.

2. Data acquisition and cleaning

   2.1 Data sources

Seoul Metropolitan City Office provides various public data such as transportation, demographics, etc. through open data plaza(https://data.seoul.go.kr/) as below

- GEO-JSON Data about 25 sub-areas.

- SubWay Station Data

- In/Out population Data

- Location Data

- Foursquare Venue Data

- Venue Name, ID, Location, Category, Likes

   2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into several files like csv format. Although some data had various problems such as duplication, various encoding issues, and support

only in Korean, the basic data for data analysis was obtained through data correction.

2.3 Feature selection

I found some meaningful derived features using exploratory Data analysis about subway station In/Out customer population info as below

- IN_AM : 06-10 customer count into subway station
- OUT_AM : 06-10 customer count out of subway station
- IN_LUNCH : 11-14 customer count into subway station
- OUT_LUNCH : 11-14 customer count out of subway station
- IN_DINNER : 17-20 customer count into subway station
- OUT_DINNER : 17-20 customer count out of subway station
- IN_NIGHT : 22-01 customer count into subway station
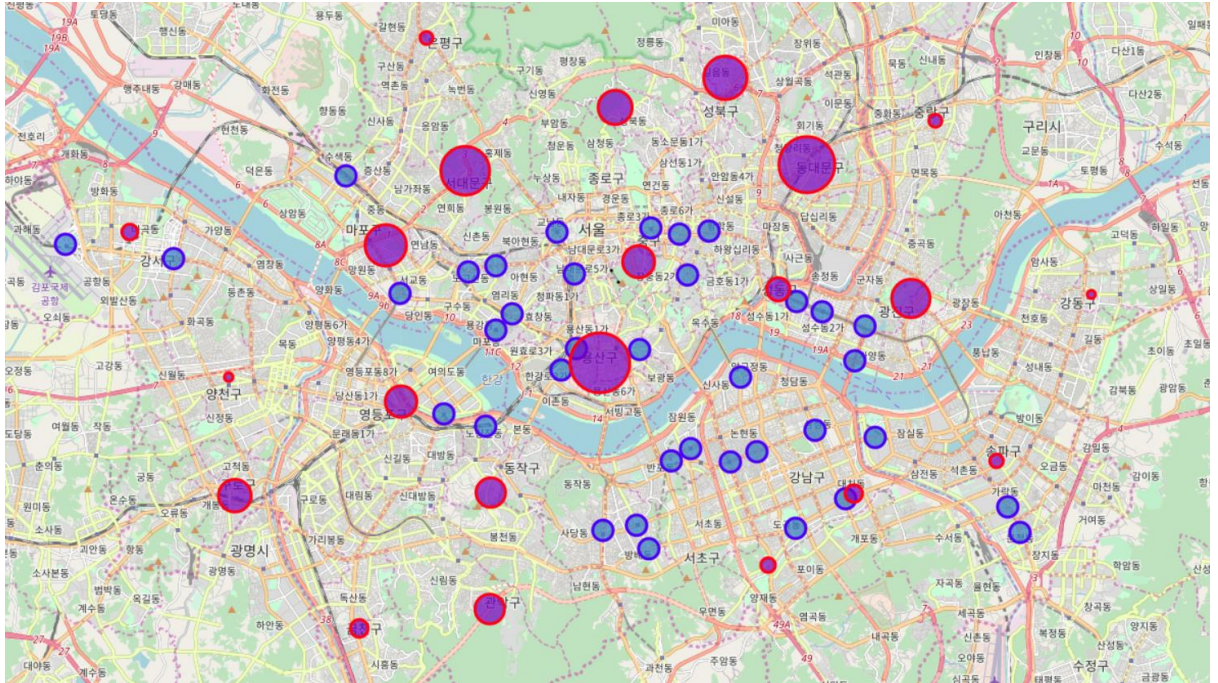- OUT_NIGHT : 22-01 customer count out of subway station

Meaningful Ratio Features

- IN_AM_OUT_DINNER_RATIO : IN_AM / OUT_DINNER *100 (%) - if this ratio is below 100 %, the customer counts in dinner time are more than customer counts in AM(get to work) time using the subway station
- INOUT_R_NIGHT : IN_NIGHT / OUT_NIGHT *100 (%) - if this ratio is over 100 %, the customer counts who go back home after dinner in night time are more than customer counts go out the subway station who lived in nearby station
- OUT_DINNER_IN_NIGHT_RATIO : OUT_DINNER / IN_NIGHT *100 (%) - This ratio provide additional information about how much the customers who go out the station in dinner time go back in Night time.

3. Exploratory Data Analysis

We want to find the best place to open a new restaurant by marking the location of the subway station selected on the map and the foreign residents of each distinction with markers.

- BLUE : selected subway station by crowed population in dinner time
- RED : Foreign Resident Count (The size of the circle is proportional to the number of foreign resident.)

4. Analysis

The Yongsan-gu area was chosen as the best place for the new restaurant through data analysis.

Reason

- 3 selected subway Station
- TOP 1 foregin resident counts

In order to analyze the existing commercial districts, I will use the Foursquare API to look at the current status of restaurants in question.

As result of Foursquare API analysis about One of three selected subway station, Itaewon station, Turkish Restaurant is not yet popluar nearby Itaewon Station Area.

station	Itaewon
1st Most Common Venue	Coffee Shop
2nd Most Common Venue	Korean Restaurant
3rd Most Common Venue	Lounge
4th Most Common Venue	Bar
5th Most Common Venue	Pizza Place
6th Most Common Venue	Café
7th Most Common Venue	Vegetarian / Vegan Restaurant
8th Most Common Venue	BBQ Joint
9th Most Common Venue	Indian Restaurant
10th Most Common Venue	Salad Place

5. Conclusion

The Yongsan-gu area was chosen as the best place for the new restaurant through data analysis.

*Reason

- 3 selected subway Station

- TOP 1 foregin resident counts

As result of using the Foursquare API to look at the current status of restaurants in question, Turkish Restaurant is not yet popluar nearby Itaewon Station Area. But Itaewon Station is the best place for a new restaurant because it is a place where locals and foreigners come to enjoy food from various countries.

6. Future directions

We expect to further develop the solution by further analyzing the real estate rental data or customer assessments of actual competitive restaurants. If possible, the analysis will be expanded to various areas including Hongdaehapjeong, Seongsu-dong, Gangnam, Cheongdam, and Samsung, which have recently emerged as hot places outside Yongsan