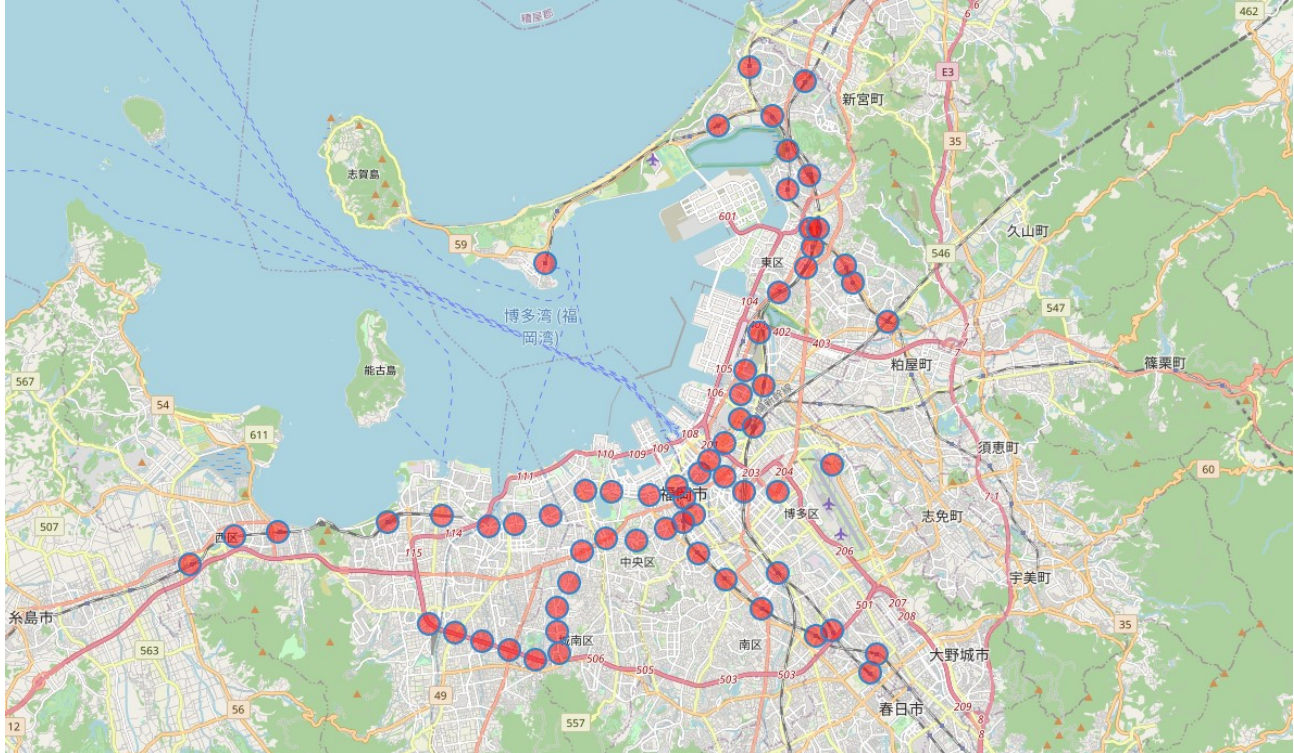IBM Data Science - Capstone Project

The Battle of Neighbourhoods

Find a good place for a new coffee shop in Fukuoka City using Data Science methodologies

Abstract

This project is the capstone assignment for the IBM Data Science Professional Certification on Coursera. [1]

(to be completed)

E. Malamura

July 8, 2021

Table of Content

(to be completed)

1. Introduction

Fukuoka is Japan's fifth largest city and the capital of Fukuoka prefecture situated on the northern side of Kyushu island. According to the preliminary data of the 2020 national census, the city experienced a five-year population growth of 4.9%, the fastest rate among the ordinance-designated cities in Japan. Fukuoka's population is also relatively youthful, with the percentage of 20 to 30 year olds considerably higher than the Japanese average. [2]

Along with the population growth, in recent years there is a trend of growing popularity of cafes and coffee culture. There are many new venues opened in recent years, as in busy districts near big city transport hub stations, as well as more coffee venues are being opened in quieter residential neighbourhoods too.

The goal of this analysis is to identify a good location for opening a new coffee shop in the city, such that: a) the area has a decent foot traffic, b) the area is popular among people, who are most likely to be active customers of a coffee shop. In this analysis we will focus mainly on the young population of Fukuoka City, trying to detect respective locations of interest and gathering, c) the area has few coffee shop venues yet, that would give some competitive advantage to the new venue.

The results of this analysis will be useful for entrepreneurs, who are planning to open a coffee shop venue in the city and thus looking for a promising location.

2. Data Description.

Instead of analysis of districts and neighbourhoods of the city by their names and official borders, we will use train stations as centers of neighbourhoods. That sounds reasonable for Japanese cities, due to high levels of development and popularity of public transport. A subway or a city railroad station is always a hub of foot traffic and a center of a neighbourhood.

In Fukuoka, we will analyse stations of 3 main city railways: Fukuoka City Subway (35 stations), Japan Railroad (18 stations within the city) and Nishitetsu Railroad (16 stations within the city). The list of names of all stations (in Japanese) of 3 operators are found and scraped from the following website of local 'yellow pages'. [3]

Every train station has a page in wikipedia that includes information about their locations, including geographical coordinates. The links to those pages were assembled using the names of stations in Japanese, then the geographical coordinates were scraped for each station from corresponding

wikipedia pages. [4]

The load (foot traffic) of every station is obtained from another source of regional statistical data in CSV format, and then merged with the scraped data by train station names. [5]

At last, but not least, we used the Foursquare API to gather information about venues in the vicinity (500m radius) of each train station [6]. We were interested in 3 large categories of venues, which are:

- Coffee shops;
- Colleges and universities;
- Arts and entertainment venues

The first category will give us information about possible competitors in the vicinity of a station. The other two categories may give us insights about places of gathering of young audience, that is the primary target group of customers, as it is stated in Introduction.

The resulting dataset is summarized in Table 1.

Table 1. Summary of the main dataset of venues.

#	Column	Count Values	Type	Sample
0	Station	1251	object	Tenjin
1	Station Latitude	1251	float64	33.591489
2	Station Longitude	1251	float64	130.399358
3	Daily Passengers	1251	int64	148957
4	Venue	1251	object	Tenjin Style
5	Venue Latitude	1251	float64	33.593734
6	Venue Longitude	1251	float64	130.394685
7	Venue Category	1251	object	Cafe

In the next section we will start with an exploratory analysis intending to understand it's underlying features. We will proceed with analysis of locations from each of the chosen criteria (foot traffic, popularity among the target customers and density of existing coffee shop venues in the area). Then finally we will apply a machine learning technique (clustering) to detect the most promising group of locations for opening a new coffee shop business.

References

1. Coursera IBM Data Science
2. Wiki: Fukuoka City
3. Station names (Mapion)
4. Location data source (Wikipedia)
5. Foot traffic data source (opendata)

6. [Foursquare API](#)