ANALYSIS OF FOOD BUSINESS OPPORTUNITY IN JAKARTA USING DATA FROM FOURSQUARE PUBLIC API

by Vicky Yuliandi
vicky.yuliandi.s@gmail.com

I. Introduction

Jakarta is Indonesia's massive capital. This city sits on the northwest coast of the island of Java. Indonesia is the largest economy of ASEAN, and Jakarta is the economic nerve center of the Indonesian archipelago. Hence, it is not surprising if this city attracts interest of people doing their business here.

Some people might be interested to start a food business in Jakarta. Some are experienced, some are new to this kind of business. Despite of how much experience they have, a good businessman should think of fundamental question about how to start their own business. They may have at least two critical questions that we will discuss here. The first question is “what kind of food business that they should do here?”. With many kinds of food venue commonly found in the city, they should know which kind of food venue that they can make to compete with existing similar venues. The second question they need to answer is “where in the city should they start this new food business?”. Obviously, Jakarta is a big city. Ones might get confused in deciding where to start their business in this city.

Data scientists should be challenged in answering this business problem. We can bring an analysis on the related market data to help answering these two questions. The analysis would be useful in guiding them to take right decisions. This report will focus on answering the business questions in the point of view of a data scientist.

II. Data

In order to answer those two critical questions, as a data scientist, we need to have data related to the food business market in Jakarta. One best thing that might come our mind now is to explore

existing food market in all area in Jakarta. Therefore, firstly we need to have a list of area in Jakarta. Then, we will inspect the business market by finding data of existing food venues in those areas. In Indonesia, we have our own administrative division system [1]. There are four levels of government in this country. In this analysis, we will only inspect the market on the third level, that is sub-district (*kecamatan*). We will use list of sub-district in the city of Jakarta from Wikipedia page [2]. The center of the sub-district coordinate will be used to find nearest food venues.

Foursquare has public API that can provide us the data. Foursquare has some account tiers for developers. Each tier has different set of available features [3]. In this project we are using personal tier. In this tier, we can get the data of venues near the specific location with specific category. We can use this feature to find existing food venues in Jakarta. The data should contain longitude, latitude, and the category of the venue. Processing and analysis will be done on this data to help us answer the business problem.

III. Methodology

There are steps that we need to do to answer the business problem. Here they are:

1. Web scraping

   Web scraping is a technique to extract the necessary data from websites. We need to extract the necessary data from the Wikipedia page that contains the list of sub-districts in Jakarta to our analysis.

2. Get longitude and latitude of each sub-district

   The coordinate of each sub-district center is needed when using public Foursquare API in the next step [4].

3. Get list of food venues in each sub-district

   We use the coordinate of sub-district to find the nearest food venues. We limit our exploration by only getting 50 number of venues in radius 2000 meters from the coordinate. The data should contain longitude, latitude, and the category of the venue.

4. Exploratory Data Analysis (EDA)

   Descriptive statistics is useful to describe and summarize the venue data. Creating data visualization of top food venue categories and top venues will help us to answer the first question.

5. Cluster all sub-districts

   5.1. Data Preparation

   Data of available food categories in each sub-district will be considered as a feature in our model. So, we are playing with categorical variable. It is convenient to convert this variable to numerical variable because most machine learning models cannot work with categorical data as a feature.

   5.2. Clustering

   We will use k-Mean, which is a popular clustering algorithm for market segmentation [5], to see the similarity between each sub-district. We divide all sub-districts in Jakarta into 6 clusters.

   5.3. Cluster Visualization

   It is good for us to see the map of the 6 clusters so we can know visually the geographical distribution of all clusters.

   5.4. Cluster Analysis

   We can get the top 10 food venue categories in each cluster to see the dissimilarity between them. This can help us to answer the second question.

See the detailed documentation of this long process at the bottom of this report [6].

IV. Result and Discussion

The result we got from the Exploratory Data Analysis (EDA) are two bar charts. The first one shows the most popular food venue categories in Jakarta (See Figure 1). The most common food venue that we can meet in Jakarta is Indonesian restaurant with 252 total number of venues in the city. The second and the third most popular food venue category are Asian restaurant and Noodle House with total number of venues 142 and 131 respectively.

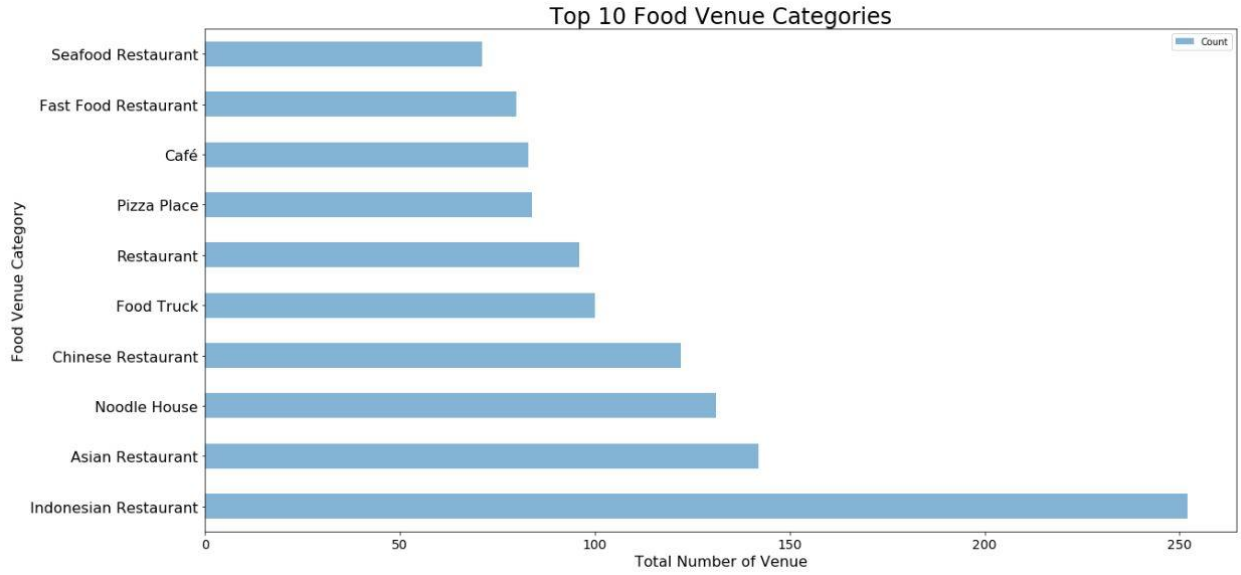

Figure 2: Top 10 food venue categories based on the total number of venues in each category

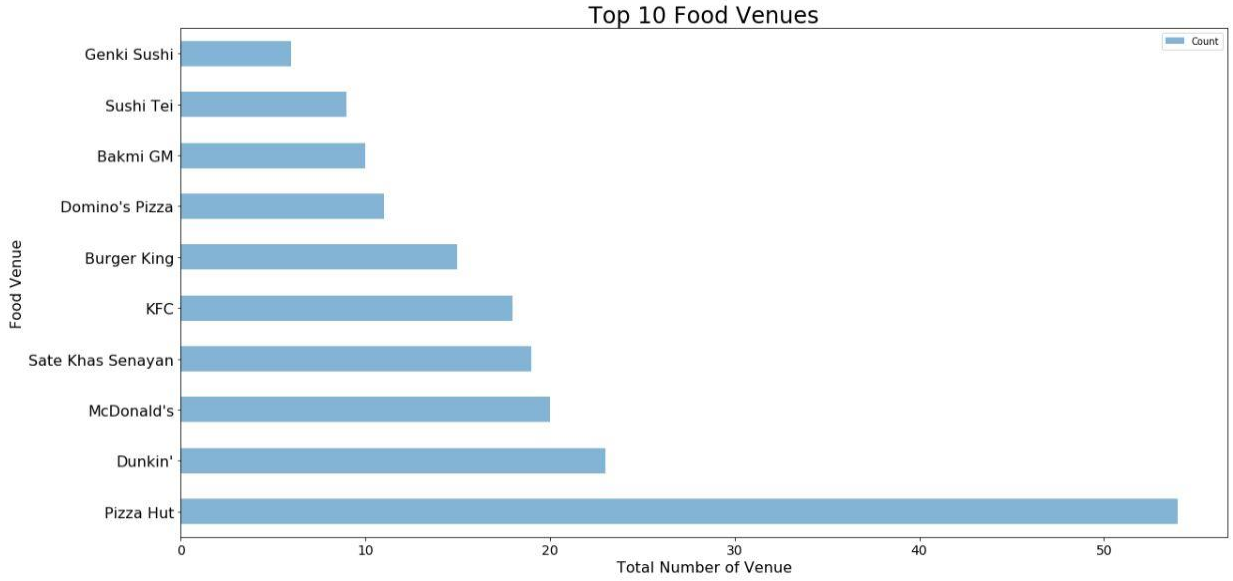

Figure 1: Top 10 food venues based on total number of same venues

Top common venues that we can meet in Jakarta are showed in Figure 2. Pizza Hut has the highest number of venues in Jakarta that is 54 venues. Dunkin' is the second most common venue with 23 number of venues, followed by McDonald's on the third position with 20 venues. From the result we can see that the most common venue category is not dominated by one venue because Pizza Hut is not a Indonesian Food Restaurant, but instead an Italian food restaurant. This means the competition between Indonesian food restaurants in Jakarta is still possible because it is not

dominated by one venue brand. So, it seems like building new Indonesian food restaurant look promising.

One might also think that we should choose the less popular food venue type to avoid large number of same competitors. There is nothing wrong with it. But sometimes being a little bit different is better than being totally different. A little innovation in popular trend can be a good idea to start a business. You can follow any idea that comes to your mind after seeing the data, but this time we need you to leave it for a moment to see the flow on how we answer the business problem.

So, let's say that we all agree to build a new Indonesian food restaurant in Jakarta. We also want to avoid places with many same competitors this time. The best way to find the right place in Jakarta to start our new business is by finding sub-district cluster with least number of Indonesian restaurants. From the cluster analysis (See Table 1), we have top 10 food venues in each sub-district cluster. Based on the data, sub-districts in cluster 2 have Indonesian restaurant the lowest rank of popularity. So, in these sub-districts there are fewer number of Indonesian restaurants compared to sub-districts in other clusters. In Figure 3, we can see visually the location of sub-districts that are in cluster 2.

Table 1: Cluster Analysis

Cluster	0	1	2	3	4	5
1st Most Common Food Venue	Indonesian Restaurant	Indonesian Restaurant	Noodle House	Indonesian Restaurant	Asian Restaurant	Indonesian Restaurant
2nd Most Common Food Venue	Asian Restaurant	Restaurant	Chinese Restaurant	Asian Restaurant	Diner	Pizza Place
3rd Most Common Food Venue	Food Truck	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Indonesian Restaurant	Restaurant
4th Most Common Food Venue	Fast Food Restaurant	Japanese Restaurant	Indonesian Restaurant	Caffé	Seafood Restaurant	Food Truck
5th Most Common Food Venue	Noodle House	Caffé	Seafood Restaurant	Chinese Restaurant	Wings Joint	Asian Restaurant

6th Most Common Food Venue	Restaurant	Fast Food Restaurant	Food Truck	Food Truck	Food	Seafood Restaurant
7th Most Common Food Venue	Caffé	Asian Restaurant	Caffé	Noodle House	Donut Shop	Noodle House
8th Most Common Food Venue	Food Court	Steakhouse	Pizza Place	Restaurant	Dumpling Restaurant	Fast Food Restaurant
9th Most Common Food Venue	Pizza Place	Bakery	Restaurant	Sushi Restaurant	Eastern European Restaurant	Caffé
10th Most Common Food Venue	Donut Shop	Noodle House	Japanese Restaurant	Korean Restaurant	Fast Food Restaurant	Indonesian Meatball Place

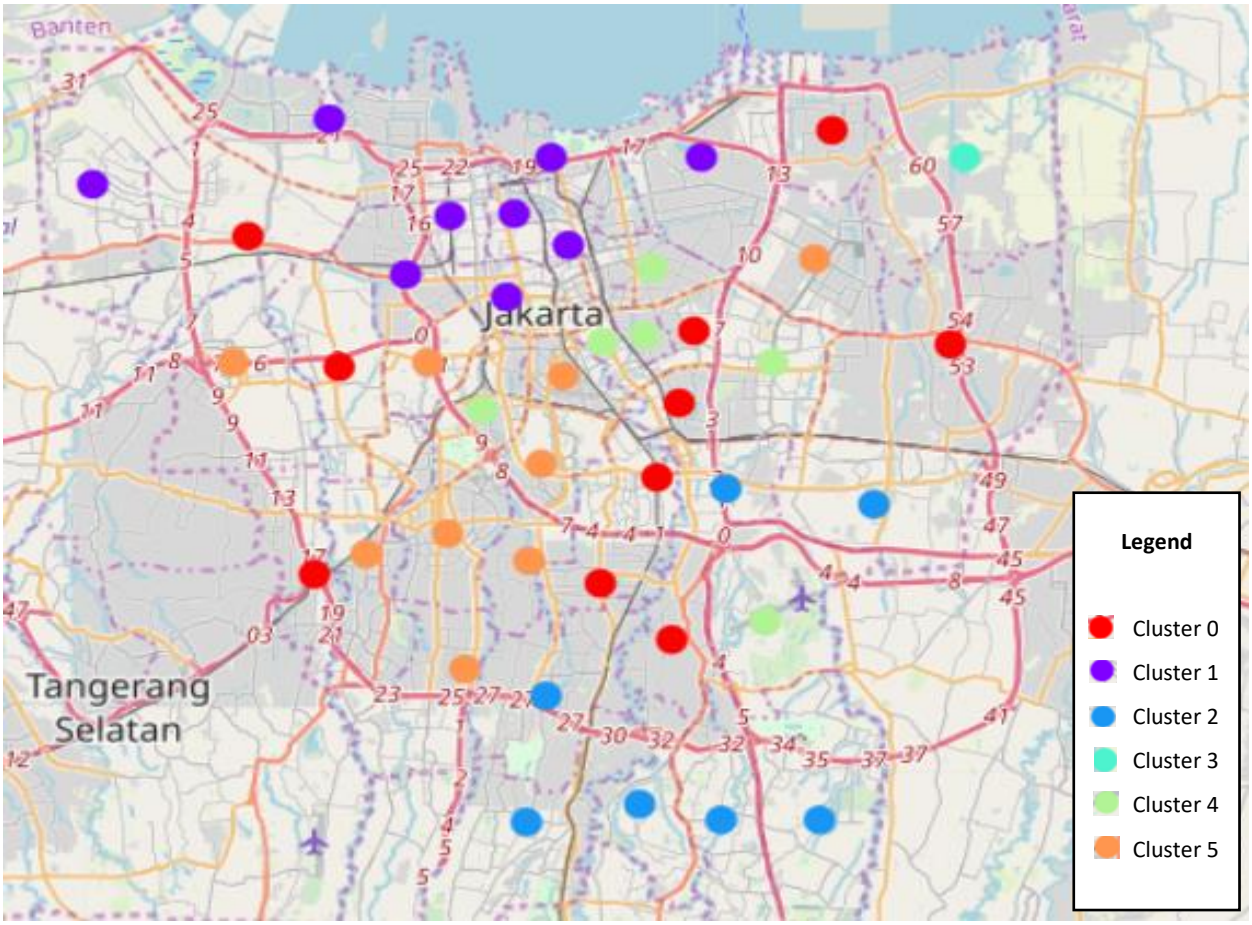

Figure 3: Map of sub-district cluster

We can end with other kind of plan, if we want to do other type of food business other than Indonesian food restaurant. Suppose that you want to do food business with less popularity, we

can do the same process we did previously in choosing the right place to do that business. We just need to find the sub-districts cluster with least number of competitors with same category.

V. Conclusion

The most popular food venue category in Jakarta is Indonesian restaurant. Pizza Hut as an Italian food restaurant is dominating in the market based on the number of venues. But, the most popular food venue category is not dominated by one venue brand. Therefore, starting new Indonesian restaurant in Jakarta still looks promising because the competition there is still possible. Based on the cluster analysis, sub-districts in the cluster 2 are the best place to build our new Indonesian restaurant.

VI. Further Direction

Use premium tier developer account to get more detailed data of the venue, such as menu, rating, likes, visits count, etc. This will improve the analysis and help us more in answering what kind of food that is recommended to offer to the market in our food venue.

VII. Reference

[1] https://en.wikipedia.org/wiki/Subdivisions_of_Indonesia

[2] https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Daerah_Khusus_Ibukota_Jakarta

[3] https://developer.foursquare.com/docs

[4] https://developer.foursquare.com/docs/api/venues/explore

[5] https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

[6] https://github.com/vickyysiahaan/Food_Business_Opportunity/