

# RECOMMENDER SYSTEM FOR OPENING A NEW RESTAURANT IN JAKARTA, INDONESIA



**IBM FINAL CAPSTONE PROJECT**

**29/06/2020**

**Dhika Rangga**

## I. INTRODUCTION

Jakarta, officially the Special Capital Region of Jakarta (Indonesian: Daerah Khusus Ibukota Jakarta), is the capital and largest city of Indonesia. Situated on the northwest coast of the world's most populous island of Java, it is the centre of economy, culture and politics of Indonesia with a population of more than ten million as of 2014. Although Jakarta only covers 699.5 square kilometres, the smallest among any Indonesian provinces, its metropolitan area covers 6,392 square kilometres. It is the world's second-most populous urban area after Tokyo, with a population of about 30 million as of 2010. Jakarta's business opportunities, as well as its potential to offer a higher standard of living, have attracted migrants from across the Indonesian archipelago, making it a melting pot of numerous cultures.

Indonesia is the largest economy of ASEAN, and Jakarta is the economic nerve centre of the Indonesian archipelago. Jakarta's nominal GDP was US\$483.8 billion in 2016, which is about 17.5% of Indonesia's. Jakarta ranked at 21 in the list of Cities Of Economic Influence Index in 2020 by CEOWORLD magazine. According to the Japan Center for Economic Research, GRP per capita of Jakarta will rank 28th among the 77 cities in 2030 from 41st in 2015, the largest in Southeast Asia. Savills Resilient Cities Index has predicted Jakarta to be within the top 20 cities in the world by 2028.

The two major population groups in Indonesia that are deemed attractive target consumer bases for the Food Service Industry in Indonesia, especially in Jakarta, are the middle-income group and millennials, due to their large number and unique consumption behaviour. Indonesia has a relatively young population, with 59 percent being aged between 15 and 54 years and a median age of 30 years. This big, yet young slice, gives Indonesia one of the best ever “demographic bonuses” enjoyed by any country today. Jakarta’s increasing urbanization and a rise in the number of office workers has consequently seen growth in both disposable income and household consumption.

## II. PROBLEM BACKGROUND

Food service industry growth has been consistent over the past several years. SWA Magazine stated there were 380 million restaurant visits in Jakarta during 2013 and resultant revenues of IDR 22.23 trillion (approx. USD \$1.5 billion). The same research identified the number of high-end restaurants in Jakarta more than doubling in the period of 2009 to 2014. According to Euromonitor, food service in Indonesia grew at a CAGR of 8.7 percent from 2010 to 2014, reaching USD \$36.8 billion sales in 2014. Indonesia's Trade Ministry statistics state that the Indonesian culinary sector overall grew by 8.16 percent in 2015. More recently, growth by outlet has slowed down although larger food franchises continue to expand steadily. As of 2018, value growth continues to exceed forecasted expectations as does daily sales volumes.

The restaurants business in Jakarta is going through a revolutionary change due to the strong driving force of consumer's market power, vigorous completion and new management practices. As a metropolitan city consisting of diverse culture and nationalities, Jakarta has a vast range of restaurant options from street vendors to upscale dining restaurants and cafes. Food is a major expense in the Indonesian family's monthly budget, requiring on average 60% of total expenditures. Jakarta's market is a vibrant and diverse food service market as well as being fast developing and growing. The region represents a large and exiting potential for food service suppliers with the possibility of creating synergies in terms of distribution, marketing and production. In the context of food & beverage industry, some of the product decision that may be considered would consists from the selection and type of the menu, the taste of the food and beverages to the overall ambiance or feel of the restaurant.

The objective of this project is to analyse and select the best category of food & beverage for anyone who wants to open a new culinary business in the city of Jakarta, Indonesia based on every district in its area but they don't have much knowledge about the area so it's hard to decide what kind of food & beverage category they want to open. Using data science methodology and machine learning techniques like clustering and recommender system, this project aims to provide solutions to answer the culinary business question: In the city of Jakarta, Indonesia, if a someone is looking to open a new food & beverage service in a specific district, what kind of food & beverage category would you recommend that they open it?

### III. TARGET AUDIENCE

This project is particularly useful not only for someone who wants to open or invest in a new food & beverage service but also for everyone who wants to find the most frequent food & beverage service in specific area based on district to know what is the most frequent of its category in the area they are looking for and all this could be easily handled by our recommender system.

### IV. DATA REQUIREMENTS

Data is the main source for starting and completing a recommender model that will answer questions and provide solutions with a mathematical approach in programming and knowledge domain intuition. Data can answer question which are unimaginable and non-answerable by humans because humans do not have the tendency to analyse such large dataset and produce analytics to find a solution. To solve the problem, we will need the following data:

- Geographical coordinates (latitude and longitude) every district in Jakarta area to populate all food & beverage services based on their category and find out what kind of category is the most frequent occurrence.
- Population of each district.
- The list of all food & beverage services in every district using Foursquare API that will be used to perform clustering on the districts.

### V. DATA SOURCES

In this final project, each of the data needed in the previous chapter can be found from 4 different sources:

#### 1. List of districts in Jakarta

- [https://id.wikipedia.org/wiki/Daftar\\_kecamatan\\_dan\\_kelurahan\\_di\\_Daerah\\_Khusus\\_Ibukota\\_Jakarta](https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Daerah_Khusus_Ibukota_Jakarta)
- This page contains list of neighbourhoods (districts) from each borough in Jakarta, with total of 42 neighbourhoods from 5 boroughs.
- To extract all information from this site, we will use web scraping technique with the help of Python requests and bs4 packages.

## 2. Geographical coordinates for every district in Jakarta

- <https://github.com/ArrayAccess/Indonesia-Postal-And-Area/>
- This GitHub account provides us information details about country code, postal code, and geo location (latitude & longitude) for every province, borough and neighbourhood in Indonesia.
- All data are provided in 3 datatype formats (.csv, .json, and .sql) but we will only use csv and json data formats.
- We will find all the borough's code from json file based on the data that we gather from point 1, after that we use those codes to find all neighbourhoods in csv file including latitude & longitude information from each neighbourhood.

## 3. Population

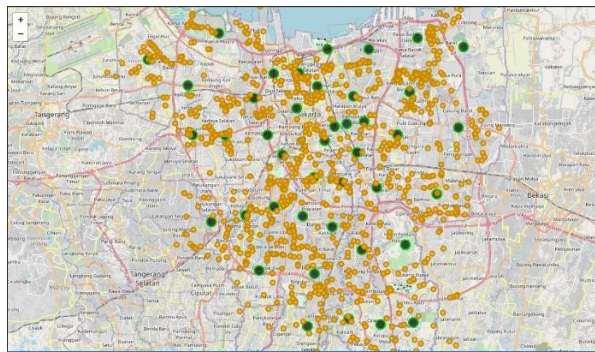
- <https://data.jakarta.go.id/sr/dataset/jumlahpendudukmenurutkecamatanjeniskelamindkijakarta/resource/ec830fce-7b59-4c65-b74e-b226630f06b2>
- This site is one of the official websites of the Indonesian government specifically in the Jakarta area which provides population data for each district in the form of CSV files but the latest data is provided only in 2014 while this project begins in 2020 so that there should be many changes, especially in the size of the number residents of each district within a period of 6 years.
- With this data, we will gather all population of each district and join it with the table from point 2.

## 4. The list of all food & beverage services

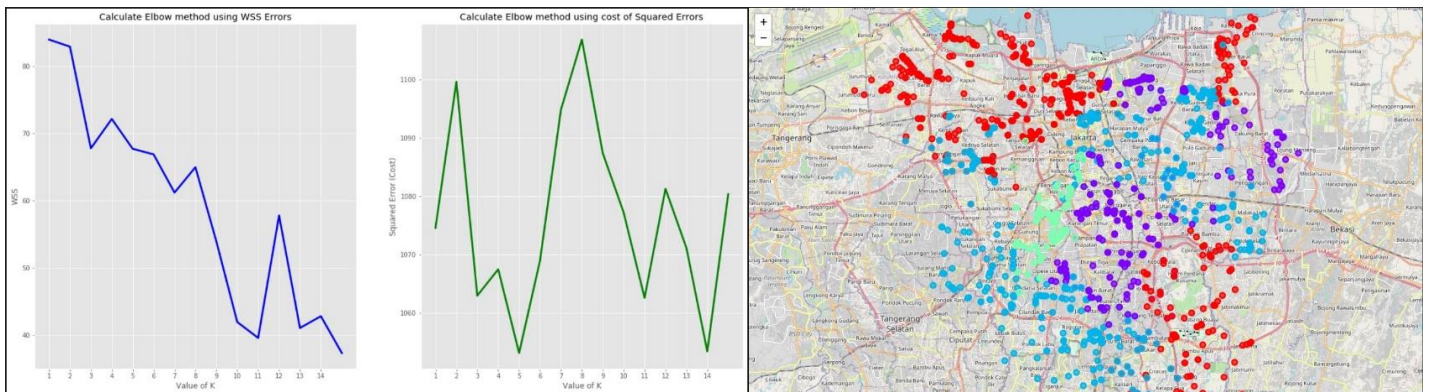
- [Foursquare API](#)
- Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers.
- Foursquare API will provide many categories of the food & beverage service data, use of Foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster in order to help us to solve the business problem put forward.
- Foursquare API leverages the power of finding nearest venues in a radius and also corresponding coordinates, venue location and names.

## VI. METHODOLOGY

- Collect all information about boroughs and neighborhoods in Jakarta by doing web scraping ([https://id.wikipedia.org/wiki/Daftar\\_kecamatan\\_dan\\_kelurahan\\_di\\_Daerah\\_Khusus\\_Ibukota\\_Jakarta](https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Daerah_Khusus_Ibukota_Jakarta)) where all list names are available on the website.
- Based on data about all boroughs and neighborhoods in Jakarta that have been obtained, then look for information about geolocation (longitude and latitude) of all the neighborhoods that have been obtained and then merged with the dataset that has been created.
- Combine with population data in each neighborhood.
- After all the required information about each neighborhood has been obtained, the next step is to execute the foursquare API to pull all the information about the restaurants contained in each neighborhood and then visualize the data using the Folium library.



- Apply One-Hot Encoding in all categories to analyze all neighborhoods in the dataset and grouped by boroughs and neighborhoods so that the average values for each category are obtained in each neighborhood.
- After the Exploratory data analysis has been done, then prepare a machine learning model to cluster using the K-Means algorithm and then after all the data has a label then visualize it all.



- The last step is to apply the recommender system algorithm by using the Weighted Rating (WR) formula to get the best recommendation value in each neighborhood so that the results can be used by the user as a reference for opening a new restaurant.



## VII. RESULT

The results produced by the recommender system are sorting the ranking values for each category in each neighborhood to provide recommendations on what categories are suitable if someone wants to build a new restaurant in a certain area in Jakarta that is supported by the determining parameters as a reference basis development of the model making it easier for users to choose without having to do a lot of calculations. Each neighborhood has a food & beverage service category ranking from first to third as shown in the following figure:

	neighborhood	1st recommended category	2nd recommended category	3rd recommended category
0	CAKUNG	Pizza Place	Japanese Restaurant	Seafood Restaurant
1	CEMPAKA PUTIH	Pizza Place	Seafood Restaurant	Steakhouse
2	CENGKARENG	Steakhouse	Pizza Place	Sushi Restaurant
3	CILANDAK	Seafood Restaurant	Steakhouse	Sushi Restaurant
4	CILINCING	Seafood Restaurant	Japanese Restaurant	Steakhouse
5	CIPAYUNG	Seafood Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant
6	CIRACAS	Pizza Place	Seafood Restaurant	Steakhouse
7	DUREN SAWIT	Seafood Restaurant	Steakhouse	Padangnese Restaurant
8	GAMBIR	Pizza Place	Steakhouse	Sushi Restaurant
9	GROGOL PETAMBURAN	Seafood Restaurant	Pizza Place	Steakhouse
10	JAGAKARSA	Seafood Restaurant	Pizza Place	Steakhouse
11	JATINEGARA	Indonesian Meatball Place	Padangnese Restaurant	Pizza Place
12	JOHAR BARU	Pizza Place	Seafood Restaurant	Steakhouse
13	KALIDERES	Seafood Restaurant	Steakhouse	Padangnese Restaurant
14	KEBAYORAN BARU	Japanese Restaurant	Pizza Place	Steakhouse
15	KEBAYORAN LAMA	Japanese Restaurant	Pizza Place	Steakhouse
16	KEBON JERUK	Seafood Restaurant	Indonesian Restaurant	Japanese Restaurant
17	KELAPA GADING	Pizza Place	Steakhouse	Vegetarian / Vegan Restaurant
18	KEMAYORAN	Pizza Place	Seafood Restaurant	Steakhouse
19	KEMBANGAN	Seafood Restaurant	Japanese Restaurant	Steakhouse
20	KOJA	Seafood Restaurant	Pizza Place	Steakhouse
21	KRAMAT JATI	Pizza Place	Steakhouse	Middle Eastern Restaurant
22	MAKASAR	Seafood Restaurant	Steakhouse	Sushi Restaurant
23	MAMPANG PRAPATAN	Japanese Restaurant	Pizza Place	Steakhouse
24	MATRAMAN	Padangnese Restaurant	Pizza Place	Seafood Restaurant
25	MENTENG	Pizza Place	Seafood Restaurant	Steakhouse
26	PADEMANGAN	Pizza Place	Seafood Restaurant	Steakhouse
27	PALMERAH	Steakhouse	Seafood Restaurant	Sushi Restaurant

## VIII. DISCUSSION

In this project, the author has some difficulties that are quite time consuming in its completion and there are even some decisions to change the approach used due to several things, namely:

- The Foursquare API used by the author is `/search` which searches for all venues that have a restaurant category but it turns out that each restaurant category has a different id so it must manually input data for all restaurant categories so that it can be entered into the API parameters,
- when running the foursquare API to pull all the restaurant data in Jakarta based on all ID categories that have been collected previously, it turns out that the results received are still many data that the categories do not match so the author must manually delete the data that is not in accordance with desired category,
- the most recent population data for each neighborhood can be found only for 2014 while the project is taking place in 2020 so it can be said that the population data used for this project is not relevant enough,
- The author could not find any data on population income for each neighborhood, so the author decided not to use the income of the population where in fact the data was useful to be able to provide insight into the price of food that could be used by users when wanting to provide price ranges based on each area.

## IX. CONCLUSION

In this final project, I analyze the distribution of restaurant data in each neighborhood based on each food & beverage category by dividing it into several clusters using the K-Means clustering method where previously I first sought the best k value using the elbow method so that the best value was obtained for k is 4. After all the restaurant data has been divided into several clusters, then in the recommender system algorithm I use the WR (Weighted Rating) approach to get a rating of each category in each neighborhood so that the results obtained recommendations that can be given to users in determining food & beverage categories what is suitable for each neighborhood so that users can easily determine.