**Finding Business Location in Jakarta**

Vito Kosasih

# Introduction

## Business Problem

There is a cafe brand that is well known among young people in Australia who want to open their first branch in Indonesia, to be precise in Jakarta. It takes careful consideration to determine which location is the best choice for them. We will use our data science powers to generate a few most promising neighborhoods based on this criterion. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

To launch this project, the brand needs more knowledge about demographics in Jakarta, such as:

* How many cafes are already established in the area?
* What are the most popular venues?
* What is the total population of young people (theirs target market) in the area?

## Target Audience

**Entrepreneurs** who are passionate about opening a coffee shop in a metropolitan city would be very interested in this project. The project is also for **business owners** and **stakeholders** who want to expand their businesses and wonder how data science could be applied to the questions at hand.

# Data Description

## Data Requirements and Collection

Following data sources will be needed to extract/generate the required information:

* 1st data: contains the latitude, longitude for each neighborhood in Jakarta.
* 2nd data: https://data.jakarta.go.id/dataset/jumlah-penduduk-berdasarkan-tenaga-kerja-jenis-kelamin-dan-umur contains the total population based on age categories in each neighborhood in Jakarta.
* 3rd data: https://developer.foursquare.com/ contains the most popular or common venues of a given neighborhood in Jakarta.

# Methodology

## Analytic Approach

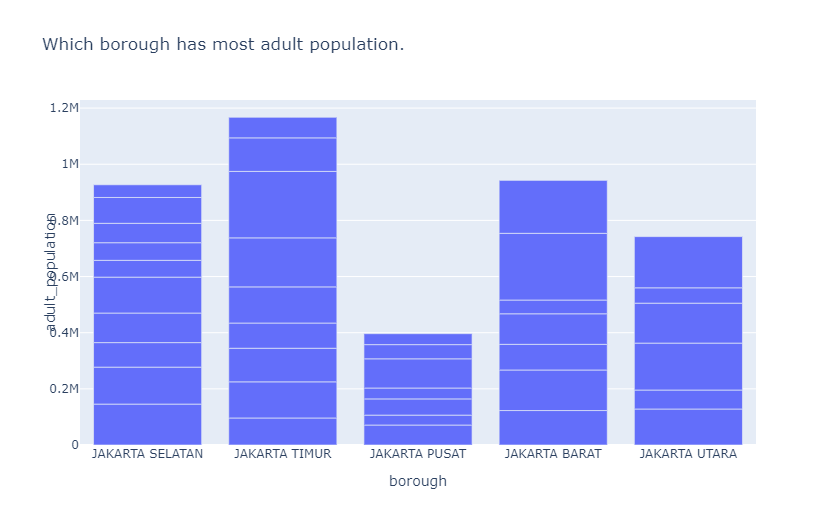
We will display any statistics needed to answer questions about demographics.

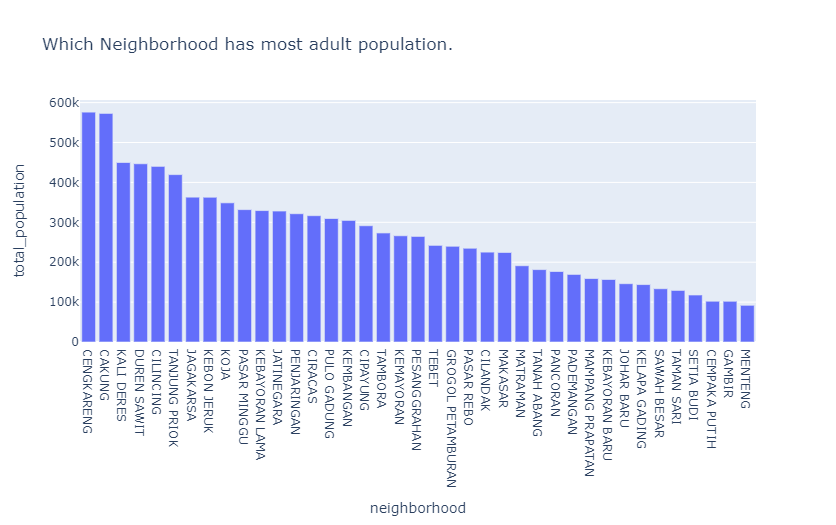
Next, we deal with the problem using the grouping technique, which is k-Means. This approach enables the public to know how similar neighbourhoods are about their demographics. We can then examine each group and identify the distinct place categories that distinguish each group.

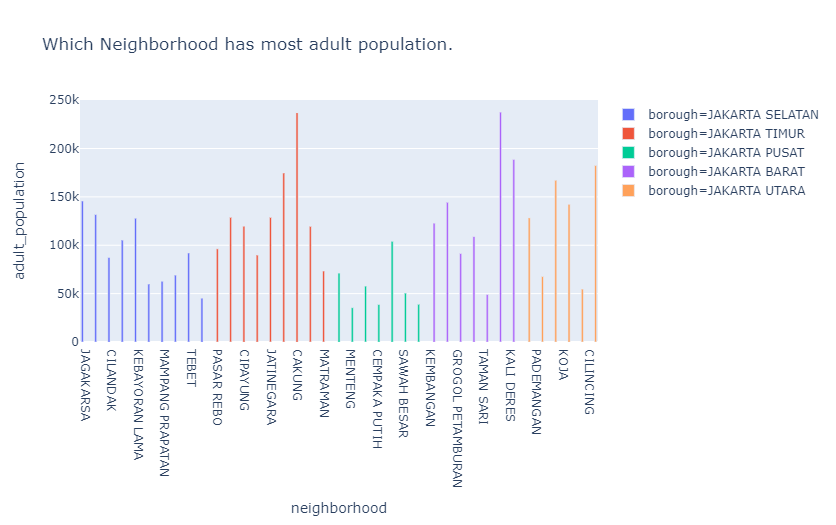
K-Means is one of the popular machine learning algorithms used to group data points based on similar characteristics. The algorithm is fast and efficient for a medium to large sized database and is useful for quickly discovering insights from unlabelled data.

## Exploratory Data Analysis

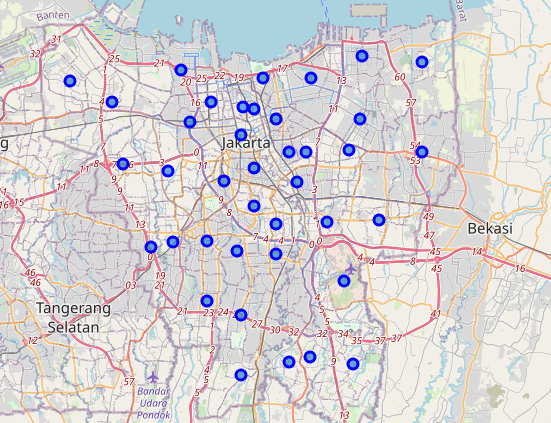
We started this project by analysing the demographics of the city of Jakarta. Starting from seeing the total adult population of each borough to which neighbourhood has the largest total adult population in Jakarta.



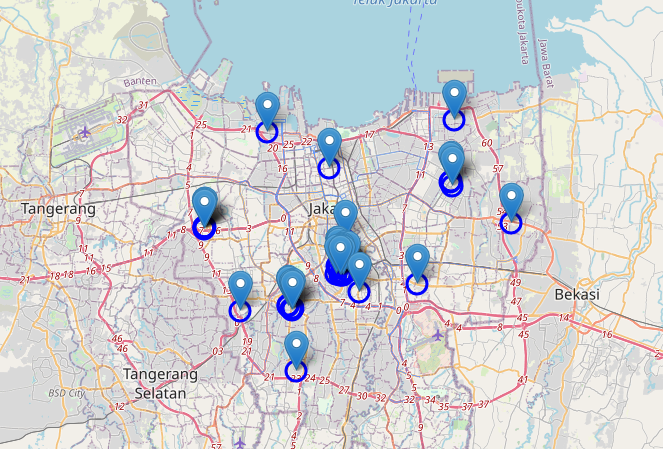




After understanding the population, we continue to look at the location of each neighborhood in Jakarta.

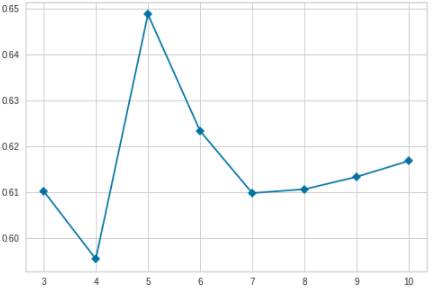


After that we look for in which area there are cafes or coffee shops in Jakarta.



## Clustering the Neighborhoods

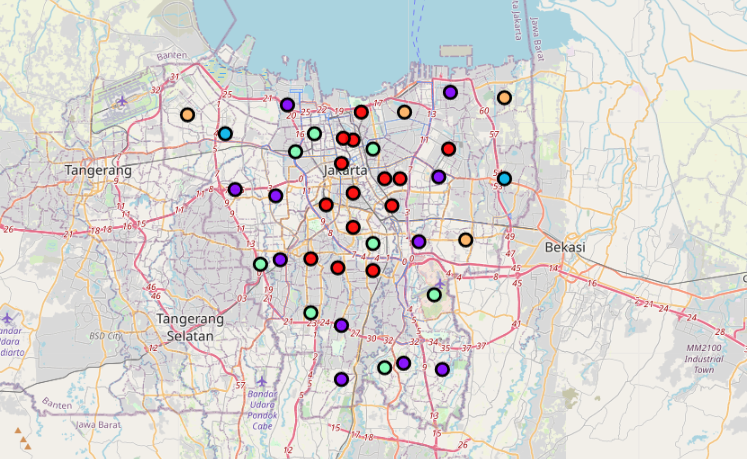
We will run the k-Means algorithm to build a clustering model with different number of clusters (k). The features will be the average occurrence of each place category. With the Silhouette Score Elbow, we can measure and plot clustering performances.



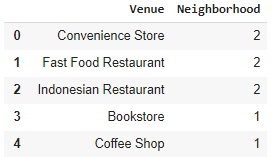
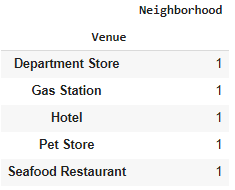
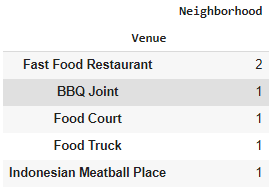
We can inspect that the best k value for this task is 5. Hence, we will have **5 cluster neighborhoods** at the end.

# Results

Finally, Let’s visualize the resulting clusters!



As a result, we can examine venues listed inside each cluster and define the discriminating venue categories that distinguish them.

The list of top 3 venues in cluster 1-5 respectively

# Discussion

As a result, we can group the neighborhoods in DKI Jakarta province into 5 clusters.

* Cluster 1: This cluster has quite several culinary tourism destinations. In addition, there are not many coffee shops or cafes that have opened their business in this cluster. However, the number of adult populations in this cluster tends to be low.
* Cluster 2: Like the first cluster. Culinary tourism in this cluster has developed well and there are not many competitors in the coffee shop or cafe sector. Unlike the first cluster, there are more adult population in this cluster.
* Cluster 3: The neighborhood in this cluster has the most adult population among other clusters. However, the location of the two neighborhoods is quite far from the city center so that the accessibility for this cluster is not good. Not only that, but there are also only a few culinary destinations available in this cluster, making this cluster not the first choice when someone wants to have a culinary tour.
* Cluster 4: This cluster has a character like cluster 1, there are many culinary tourism destinations. However, there are not many adult populations in this cluster. And unlike cluster 1 which tends to be in the middle of the city, cluster 4 is located slightly on the outskirts of the city.
* Cluster 5: Like cluster 3. However, the potential in this cluster looks a little better because there are already several culinary tourism destinations available.

The main objective of this project is to find the best location to open a new cafe in Jakarta. There are many parameters that can be used to assess each location. but here we look at the demographics and businesses situation.

Demographics: The most important thing in a culinary business is the customer. No matter how delicious our menu is, it's meaningless if we don't have customers. Therefore, in this project we are looking for a location where our target market is located.

Businesses situation:

* The business situation of a region can play a positive role.
* Cluster 3 as an example, although there are many target markets there, the cluster does not attract foodies to come to the cluster. This makes it quite difficult for a businessperson to be able to continue to grow his business. Because of that, we consider this cluster is not recommended.
* Cluster 2 to be precise at Jagaraksa, is our top choice for our business location. This is based on the consideration of the attractiveness of approaching the neighborhood because culinary tourism there has been well developed, besides that Jagaraksa also has the 7th largest adult population in Jakarta with 363,000 people.

# Conclusion

Finding the best location to open a business is not an easy matter, there are lots of uncertainties. It takes deep and mature consideration to be able to find the best location. Fortunately, with the available data, we can get valuable insights, which may not be obtained just by surveying the location. This can greatly help everyone to make more informed decisions.

Using this project as an example, I hope it will inspire everyone to deal with a similar case in the future. I would be incredibly happy if any of you wanted to have a discussion!