Finding the Best Location for Chicken Outlet

Bahri Nurhabbibi

1. May, 2021

# 1. Introduction

## 1.1 Background

Quantum Capital, a Singaporean firm specializing in direct investment is planning to open a new franchise of fast food fried chicken outlet in Indonesia. To begin with, they decide to open the franchise outlet in Jakarta - the capital of Indonesia. For the year 2021 alone, they decide to invest & open 5 chicken outlets.

Jakarta is a metropolitan city with supporting areas. The main area is divided into 5 regents, consist of around 40 districts. The analysis will be done among the districts using any available data in the internet.

With high number of populations, Quantum Capital expecting a tight competition from existing multinational outlet such as KFC, McDonald and local brands alike.

It’s important for Quantum Capital to choose the best location for its pilot project. A successful project will act as a good portfolio & attract more investment fund from clients.

## 1.2 Objective

This research’s objective is:

1. provide exploratory data analysis using data available from the internet
2. give recommendation on the best possible location for the chicken outlet
3. provide a base recommendation for future research

## 1.3 Challenge

Jakarta is at very early stage of implementing smart city. Researcher should expect that the amount of open-source data available on the internet should be very limited.

This exploratory research is mainly limited by this.

# 2. Data Acquisition & Cleaning

District name & sub district name, area, population for all district in Jakarta Province is available from various source on the internet.

We use the district name to pull the latitude / longitude data from geocoder.

For venue data, we input the latitude / longitude data to Foursquare API and get return of number of venues around that location. We pull only certain category which we identify as direct or potential competitor. Below are the table of categories regarded as potential competitor.

|  |  |
| --- | --- |
| Category | Sample |
| Fast Food Restaurant | McD, A&W |
| Wings Joint | Wingstop, various steet hawker, steakhouses |
| Fried Chicken joint | Ayam Suharti, KFC, street hawkers |

Combining all data, we get the GDP density, population density & competitor density for each sub-district.

## Data Sources

Below are the links to our data sources:

|  |  |  |
| --- | --- | --- |
| No | Data Name | Source |
| 1 | Population  District name | [National Biro of Statistic](https://jakarta.bps.go.id/indicator/12/417/1/proyeksi-penduduk-provinsi-dki-jakarta-menurut-kecamatan-jenis-kelamin-dan-rasio-jenis-kelamin.html) |
| 2 | Latitude longitude | Geocoder |
| 3 | Area | [Jakarta Government](https://data.jakarta.go.id/dataset/3e22f472-a887-40eb-bc09-7e0f06f8d85f/resource/cabba6c1-78d3-48a7-a2fb-045b343db1a0/download/Luas-Wilayah-dan-kepadatan-Provinsi-DKI-Jakarta-Tahun-2015.csv) |
| 4 | Venue Data | foursquare |
| 5 | Regional Domestic Gross Income | [databoks](https://databoks.katadata.co.id/datapublish/2019/10/14/mencapai-rp-640-triliun-pdrb-jakarta-pusat-terbesar-di-dki) |

The Regional Domestic Gross Income only available at district level & then extrapolated to sub-district level based on the number of people on each sub district.

## 2.2 Data Cleaning

So far, only standard data cleaning is needed, such as cleaning the HTML tags / characters for data gathered via html page and removing empty whitespaces.

Data cleaning also done based on multiple entry in venue categories, example: KFC is registered in both Fried Chicken Joint & Fast-Food Restaurant

There’s no missing or incomplete data.

## 2.3 Final Table

The final table is in the form of:

|  |  |  |
| --- | --- | --- |
| **No** | **Column** | **Comments** |
| 1 | District | District name as Index |
| 2 | People | Number of people in the area |
| 3 | Sub-GDP | GDP of each sub-district |
| 4 | Area(km2) | District area in km square |
| 5 | Venue | Number of relevant venues in a sub-district |
| 6 | People/ km2 | Density of person / km2 in a sub-district |
| 7 | Venue/ km2 | Density of venue / km2 in a sub-district |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **District** | **People** | **Sub-GDP** | **Area (km2)** | **Venue** | **GDP/km2** | **Venue/km2** | **Person/km2** |
| **JAGAKARSA** | 199969 | 104,639,136,785,301 | 25 | 33 | 4,205,753,000,000 | 1.326367 | 8037.339228 |
| **KALI DERES** | 154016 | 80,592,998,370,372 | 22 | 50 | 3,712,252,000,000 | 2.303086 | 7094.242285 |
| **CILANDAK** | 103351 | 54,081,179,712,343 | 18 | 81 | 2,978,039,000,000 | 4.460352 | 5691.134361 |
| **PESANGGRAHAN** | 111287 | 58,233,904,332,300 | 13 | 62 | 4,326,442,000,000 | 4.606241 | 8267.979198 |
| **KEBAYORAN LAMA** | 155328 | 81,279,537,521,252 | 17 | 90 | 4,855,408,000,000 | 5.376344 | 9278.853047 |
| **KEBAYORAN BARU** | 72561 | 37,969,487,291,921 | 13 | 89 | 2,938,815,000,000 | 6.888545 | 5616.176471 |
| **MAMPANG PRAPATAN** | 71859 | 37,602,146,984,057 | 8 | 84 | 4,858,159,000,000 | 10.852713 | 9284.108527 |
| **PANCORAN** | 77249 | 40,422,608,892,016 | 9 | 56 | 4,738,876,000,000 | 6.565064 | 9056.154748 |
| **TEBET** | 106425 | 55,689,732,570,426 | 9 | 55 | 6,167,191,000,000 | 6.090808 | 11785.71429 |
| **SETIABUDI** | 70788 | 37,041,717,540,008 | 9 | 50 | 4,180,781,000,000 | 5.643341 | 7989.616253 |
| **PASAR REBO** | 108507 | 33,987,975,614,159 | 13 | 32 | 2,620,507,000,000 | 2.467232 | 8365.998458 |
| **CIRACAS** | 139111 | 43,574,159,046,525 | 16 | 13 | 2,709,836,000,000 | 0.808458 | 8651.181592 |
| **CIPAYUNG** | 145374 | 45,535,937,468,853 | 28 | 6 | 1,599,998,000,000 | 0.210822 | 5108.011244 |
| **MAKASAR** | 101704 | 31,857,051,359,474 | 22 | 41 | 1,457,322,000,000 | 1.875572 | 4652.516011 |
| **KRAMAT JATI** | 148241 | 46,433,976,545,463 | 13 | 2 | 3,491,276,000,000 | 0.150376 | 11145.93985 |
| **JATINEGARA** | 133114 | 41,695,700,608,285 | 10 | 32 | 4,063,908,000,000 | 3.118908 | 12974.07407 |
| **DUREN SAWIT** | 201592 | 63,145,271,549,390 | 23 | 65 | 2,786,640,000,000 | 2.868491 | 8896.381289 |
| **CAKUNG** | 260735 | 81,670,812,221,865 | 42 | 17 | 1,932,122,000,000 | 0.402176 | 6168.322687 |
| **PULOGADUNG** | 133866 | 41,931,251,841,495 | 16 | 23 | 2,687,901,000,000 | 1.474359 | 8581.153846 |
| **MATRAMAN** | 75661 | 23,699,523,744,486 | 5 | 23 | 4,856,460,000,000 | 4.713115 | 15504.30328 |
| **TANAH ABANG** | 73039 | 100,803,525,441,708 | 9 | 58 | 10,839,090,000,000 | 6.236559 | 7853.655914 |
| **MENTENG** | 34963 | 48,253,585,892,721 | 7 | 27 | 7,389,523,000,000 | 4.134763 | 5354.211332 |
| **SENEN** | 49142 | 67,822,490,002,005 | 4 | 20 | 16,033,690,000,000 | 4.728132 | 11617.49409 |
| **JOHAR BARU** | 59636 | 82,305,604,447,510 | 2 | 14 | 34,728,100,000,000 | 5.907173 | 25162.8692 |
| **CEMPAKA PUTIH** | 42063 | 58,052,529,342,605 | 5 | 80 | 12,351,600,000,000 | 17.021277 | 8949.574468 |
| **KEMAYORAN** | 113581 | 156,756,872,673,431 | 7 | 41 | 21,621,640,000,000 | 5.655172 | 15666.34483 |
| **SAWAH BESAR** | 51639 | 71,268,681,804,028 | 6 | 75 | 11,569,590,000,000 | 12.175325 | 8382.954545 |
| **GAMBIR** | 39738 | 54,843,720,395,988 | 8 | 47 | 7,225,787,000,000 | 6.192358 | 5235.573123 |
| **KEMBANGAN** | 164775 | 56,068,095,562,694 | 24 | 128 | 2,319,739,000,000 | 5.295821 | 6817.33554 |
| **District** | **People** | **Sub-GDP** | **Area (km2)** | **Venue** | **GDP/km2** | **venue/km2** | **person** |
| **KEBON JERUK** | 189778 | 64,575,882,504,609 | 18 | 44 | 3,662,841,000,000 | 2.495746 | 10764.49234 |
| **PALMERAH** | 100456 | 34,182,227,934,128 | 8 | 10 | 4,551,562,000,000 | 1.331558 | 13376.29827 |
| **GROGOL PETAMBURAN** | 123303 | 41,956,391,364,993 | 10 | 26 | 4,199,839,000,000 | 2.602603 | 12342.64264 |
| **TAMBORA** | 115805 | 39,405,042,067,290 | 5 | 19 | 7,297,230,000,000 | 3.518519 | 21445.37037 |
| **CENGKARENG** | 290468 | 98,837,733,769,713 | 27 | 12 | 3,722,702,000,000 | 0.451977 | 10940.41431 |
| **KALIDERES** | 228859 | 77,873,999,589,637 | 30 | 4 | 2,576,050,000,000 | 0.132319 | 7570.592127 |
| **PENJARINGAN** | 175862 | 84,892,126,745,772 | 35 | 88 | 2,392,676,000,000 | 2.480271 | 4956.651635 |
| **PADEMANGAN** | 81321 | 39,255,283,342,012 | 10 | 44 | 3,961,179,000,000 | 4.43996 | 8205.953582 |
| **TANJUNG PRIOK** | 196802 | 95,000,286,177,920 | 25 | 33 | 3,781,859,000,000 | 1.313694 | 7834.474522 |
| **KOJA** | 156325 | 75,461,223,649,980 | 13 | 15 | 5,712,432,000,000 | 1.135503 | 11833.838 |
| **KELAPA GADING** | 83711 | 40,408,984,442,434 | 16 | 90 | 2,506,761,000,000 | 5.583127 | 5192.990074 |
| **CILINCING** | 213540 | 103,080,055,641,879 | 38 | 1 | 2,734,219,000,000 | 0.026525 | 5664.190981 |

# 3. Exploratory Data Analysis

### 3.1 Clustering Process

The clustering process is done to 3 data:

1. GDP/ km2
2. People / km2
3. Venue/ km2

Using K-means++ clustering & DBScan as comparison.

K-means clustering ++ is used because this method is considered enough for the purpose of this research, to find groups in the data, robust & easy to implement.

DBScan is presented as alternative method for comparison purpose as it’s approached the clustering process from different angle.

### 3.2 K-Means Clustering

Clustering performance is evaluate using elbow method & silhouette analysis.

Elbow method gives us an idea on what a good k number of clusters would be based on the sum of squared distance (SSE) between data points and their assigned clusters’ centroids. We pick k at the spot where SSE starts to flatten out and forming an elbow.

Chart, line chart

Description automatically generated

Graphic shown that the sum of square distance becoming flattened around k values of 5-6-7.

Second method, which is silhouette analysis can be used to determine the degree of separation between clusters or each sample.

Below are steps to calculate silhouette analysis:

1. Compute the average distance from all data points in the same cluster (ai).
2. Compute the average distance from all data points in the closest cluster (bi).
3. Compute the coefficient:

A picture containing table

Description automatically generated

The coefficient can take values in the interval [-1, 1].

1. If it is 0 –> the sample is very close to the neighboring clusters.
2. It it is 1 –> the sample is far away from the neighboring clusters.
3. It it is -1 –> the sample is assigned to the wrong clusters.

Therefore, we want the coefficients to be as big as possible and close to 1 to have a good cluster.

For the result, we can see that the K in 6 & 7 has similar value.

* The value of n-clusters as 4 and 5 looks to be suboptimal for the given data due to the following reasons:
  + Presence of clusters with below-average**silhouette scores**
  + **Wide fluctuations in the size**of the silhouette plots.
* The value of 6 and 7 for n clusters looks to be the optimal one. The silhouette score for each cluster is above average silhouette scores. Also, the fluctuation in size is similar.
* Cluster number 6 is preferable since it present less complexity.

Based on the model evaluation above, we decide to use k=6.

Chart

Description automatically generated

Chart

Description automatically generated with medium confidence

Chart

Description automatically generated

Chart

Description automatically generated

### 3.3 DBScan Clustering

DBscan are affected by 2 parameters:

1. Minimum samples (“MinPts”): the fewest number of points required to form a cluster
2. ε (epsilon or “eps”): the maximum distance two points can be from one another while still belonging to the same cluster

minimum point, if it is possible, should follow below rule :

If your data has more than 2 dimensions, choose MinPts = 2\*dim, where dim= the dimensions of your data set (Sander et al., 1998).

Based on above, the ε recommended is 6.

While the k value is determined using calculation of the average distance between each point and its k nearest neighbours. We’ll find the optimal value for ε at the point of maximum curvature (i.e. where the graph has the greatest slope).

N=6

Chart, line chart

Description automatically generated

Maximum curvature happened at value around 1.5

However, inputting ε=1.5 and minimum samples=6 do not return satisfying cluster. The data is clustered into 1 cluster & 3 outliers.

array([ 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, -1, -1, -1, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0])

Further investigation, iterating variation of ε from 0.5-1.6 & minimum samples from 2-6 do not provide satisfying result with too many numbers in outliers or too small number cluster. The calculation is presented in our Jupyter Notebook notes.

We decide that the DBScan is not viable approach.

# 4. Findings & Discussion

The classification result of district clustering using k=6 is presented on the table on the next page.

The table is visualized as below:

Chart, scatter chart

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sub District** | **GDP/km2** | **Venue/km2** | **Person/km2** | **clus\_km** |
| KOJA | 5,712,432,000,000 | 1.135503 | 11,833.84 | 0 |
| KEBON JERUK | 3,662,841,000,000 | 2.495746 | 10,764.49 | 0 |
| MATRAMAN | 4,856,460,000,000 | 4.713115 | 15,504.30 | 0 |
| PALMERAH | 4,551,562,000,000 | 1.331558 | 13,376.30 | 0 |
| GROGOL PETAMBURAN | 4,199,839,000,000 | 2.602603 | 12,342.64 | 0 |
| TAMBORA | 7,297,230,000,000 | 3.518519 | 21,445.37 | 0 |
| JATINEGARA | 4,063,908,000,000 | 3.118908 | 12,974.07 | 0 |
| KRAMAT JATI | 3,491,276,000,000 | 0.150376 | 11,145.94 | 0 |
| CENGKARENG | 3,722,702,000,000 | 0.451977 | 10,940.41 | 0 |
| TANAH ABANG | 10,839,090,000,000 | 6.236559 | 7,853.66 | 1 |
| MENTENG | 7,389,523,000,000 | 4.134763 | 5,354.21 | 1 |
| KELAPA GADING | 2,506,761,000,000 | 5.583127 | 5,192.99 | 1 |
| GAMBIR | 7,225,787,000,000 | 6.192358 | 5,235.57 | 1 |
| PADEMANGAN | 3,961,179,000,000 | 4.43996 | 8,205.95 | 1 |
| SETIABUDI | 4,180,781,000,000 | 5.643341 | 7,989.62 | 1 |
| TEBET | 6,167,191,000,000 | 6.090808 | 11,785.71 | 1 |
| PANCORAN | 4,738,876,000,000 | 6.565064 | 9,056.15 | 1 |
| KEBAYORAN BARU | 2,938,815,000,000 | 6.888545 | 5,616.18 | 1 |
| KEBAYORAN LAMA | 4,855,408,000,000 | 5.376344 | 9,278.85 | 1 |
| PESANGGRAHAN | 4,326,442,000,000 | 4.606241 | 8,267.98 | 1 |
| CILANDAK | 2,978,039,000,000 | 4.460352 | 5,691.13 | 1 |
| KEMBANGAN | 2,319,739,000,000 | 5.295821 | 6,817.34 | 1 |
| PENJARINGAN | 2,392,676,000,000 | 2.480271 | 4,956.65 | 2 |
| TANJUNG PRIOK | 3,781,859,000,000 | 1.313694 | 7,834.47 | 2 |
| KALIDERES | 2,576,050,000,000 | 0.132319 | 7,570.59 | 2 |
| JAGAKARSA | 4,205,753,000,000 | 1.326367 | 8,037.34 | 2 |
| PULOGADUNG | 2,687,901,000,000 | 1.474359 | 8,581.15 | 2 |
| CAKUNG | 1,932,122,000,000 | 0.402176 | 6,168.32 | 2 |
| DUREN SAWIT | 2,786,640,000,000 | 2.868491 | 8,896.38 | 2 |
| MAKASAR | 1,457,322,000,000 | 1.875572 | 4,652.52 | 2 |
| CIPAYUNG | 1,599,998,000,000 | 0.210822 | 5,108.01 | 2 |
| CIRACAS | 2,709,836,000,000 | 0.808458 | 8,651.18 | 2 |
| PASAR REBO | 2,620,507,000,000 | 2.467232 | 8,366.00 | 2 |
| KALI DERES | 3,712,252,000,000 | 2.303086 | 7,094.24 | 2 |
| CILINCING | 2,734,219,000,000 | 0.026525 | 5,664.19 | 2 |
| JOHAR BARU | 34,728,100,000,000 | 5.907173 | 25,162.87 | 3 |
| SAWAH BESAR | 11,569,590,000,000 | 12.175325 | 8,382.95 | 4 |
| CEMPAKA PUTIH | 12,351,600,000,000 | 17.021277 | 8,949.57 | 4 |
| MAMPANG PRAPATAN | 4,858,159,000,000 | 10.852713 | 9,284.11 | 4 |
| SENEN | 16,033,690,000,000 | 4.728132 | 11,617.49 | 5 |
| KEMAYORAN | 21,621,640,000,000 | 5.655172 | 15,666.34 | 5 |

The mean value for each cluster is presented below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cluster** |  | **Venue/km2** | **Person/km2** | **GDP/km2** |
| 0 |  | 2.168701 | 13,369.71 | 4,617,583,000,000 |
| 1 |  | 5.501022 | 7411.181 | 4,955,971,000,000 |
| 2 |  | 1.360721 | 7044.697 | 2,707,472,000,000 |
| 3 |  | 5.907173 | 25,162.87 | 34,728,100,000,000 |
| 4 |  | 13.34977 | 8872.213 | 9,593,117,000,000 |
| 5 |  | 5.191652 | 13,641.92 | 18,827,660,000,000 |

Furthermore, we do a simple heuristic process to understand the cluster & put a scoring.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cluster** | **Venue/km2** | **Person/km2** | **GDP/km2** | **Comments** |
| 0 | LOW | MEDIUM | LOW | Cluster with very high person density. |
| 1 | MEDIUM | LOW | LOW | Cluster with very low venue density. |
| 2 | VERY LOW | LOW | VERY LOW | Cluster with high venue density & medium person density. |
| 3 | MEDIUM | VERY HIGH | VERY HIGH | Cluster with medium venue density & high person density. |
| 4 | VERY HIGH | LOW | MEDIUM | Cluster with very high venue density |
| 5 | MEDIUM | MEDIUM | HIGH | Cluster with medium venue density & medium person density. |

Consideration point:

1. choose very sub-district with **medium/high / very high GDP density** as it represents high purchasing power. Thus cluster 3,4,5 will put into best candidate.
2. Avoid **very high person density** in cluster 3 should be avoided since it might contain big social problem. We have 4 & 5 as our candidate.
3. Very high venue density should be avoided since it represents high competition.

Green color signifies favorable condition while the red color signify the unfavorable condition.

result: we choose cluster: 5. SENEN & KEMAYORAN

High population density with medium or low venue density, as medium / low value density presents less competition, presented as cluster 0 and cluster 3.

Quick search on the internet to Tambora & Johar Baru, shown that these two districts is considered two most dense area in Asia and often ravaged by population problem such as slum, low-income area, waste problem, and often small local riot between its populous residence. We exclude this from the consideration.

# 5. Conclusion

We come to conclusion to recommend below district as a possible location for chicken outlet.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District Name | venue/km2 | person/km2 | GDP/km2 | cluster |
| SENEN | 4.491726 | 11617.49409 | 0.000387 | 5 |
| KEMAYORAN | 5.241379 | 15666.34483 | 0.000335 | 5 |

# 6.Future Direction

The recommendation on this research only considers 4 aspect: GDP,population, number of venue & area. Therefore, do not capture other aspect of consideration

* Purchase parity
* Elite / non elite area which will affect the image of the outlet
* Rent Price
* Legal / Permit
* Property Availability

We recommend qualitative research is done to enhance the result of this paper by doing further exploration on mentioned above aspect.