Building Coffee Shop Business Analysis with Data Science

By: Intan Dea Yutami

I. Introduction (Business Problem)

Coffee shop is one place that people can go to, whether for quick refreshment, truly enjoying the coffee, or becoming temporary workspace for individuals.

According to the research by Toffin and MIX MarComm SWA Media Group magazine, the number of coffee shops in Indonesia has reached 2,950 in 2019, which was 3x compared to 2016. This phenomenon was drived by many factors, such as presence of social media than eases promotion, existence of ride hailing platforms that can ease customers to purchase via delivery, increasing purchasing powers, and new culture of hanging out among youngsters in Indonesia.

Businessmen and women who are interested in building new brands of coffee shop, or franchise, should assess every aspect thoroughly as a part of business plan, so the business can profit as much as possible. One aspect that can affect the profit is the location. Building coffee shops in a region full of competitors can be tough, so one can opt to strategize by opting regions with a greater number of populations yet a smaller number of rivals.

Bandung, as now populated by 2.5 million of people, is now the 4th most populated cities in Indonesia, having no exception to this increasing trend of coffee shops. This assignment is for businessmen or women who are interested in building new brand of coffee shops or expanding their brands in Bandung. To summarize, where should someone/a company have a new branch/make new local brand of coffee shop in Bandung?

II. Data

First Data is taken from Open Data Kota Bandung. The data taken from this website is the number of populations in Bandung based on age groups and neighbourhoods (or "Kecamatan" in Indonesian terms). From this data, the age groups will be reorganized into new groups (children, teenagers, young adults, adults, and senior adults). The number of young adults & adults' population per Kecamatan will be included as features.

The other data is taken from FourSquare API, in which venues in each Kecamatan will be collected, and the number of coffee shops per Kecamatan will be included as feature too.

III. Methodology

A. Exploratory Data Analysis in Bandung Population Dataset

First, let's see how the population dataset looks like

	Kecamatan	Kelurahan	Usia 0-4 Thn	Usia 5-9 Thn	Usia 10-14 Thn	Usian 15-19 Thn	Usia 20-24 Thn	Usia 25-29 Thn	Usia 30-34 Thn	Usia 35-39 Thn	Usia 40-44 Thn	Usia 45-49 Thn	Usia 50-54 Thn	Usia 55-59 Thn	Usia 60-64 Thn	Usia 65-69 Thn	Usia 70-74 Thn	Usia 75 Thn
0	Andir	Campaka	1190	1510	1590	1380	1494	1411	1284	1520	1500	1421	1213	913	741	486	249	381
1	Andir	Ciroyom	1383	1795	1793	1490	1547	1525	1434	1638	1601	1317	1221	934	807	576	320	498
2	Andir	Dungus Cariang	1165	1515	1573	1370	1435	1410	1303	1536	1509	1417	1277	966	818	588	359	526
3	Andir	Garuda	672	824	834	830	740	819	724	880	900	786	749	544	486	348	232	341
4	Andir	Kebon Jeruk	584	794	798	766	818	777	751	920	956	812	845	700	637	526	378	562

Figure 1 – Snapshot of Bandung Population Dataset

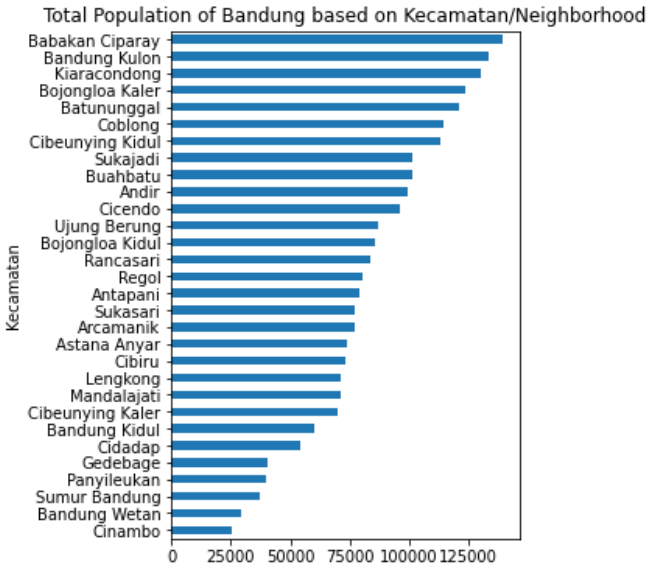
Dataset contains of 151 rows and 18 columns. In this dataset, name of Kecamatan and Kelurahan (borough/smaller regions under it) are also included. The number population is categorized into several age groups in range of 5 years. For the project, the dataset will be recategorized into new age groups. In Bandung, there are 30 Kecamatans as the object of analysis.

```
[211] bdg_pop['Kecamatan'].nunique()
0s

30
```

Figure 2 – Number of Kecamatan in Bandung

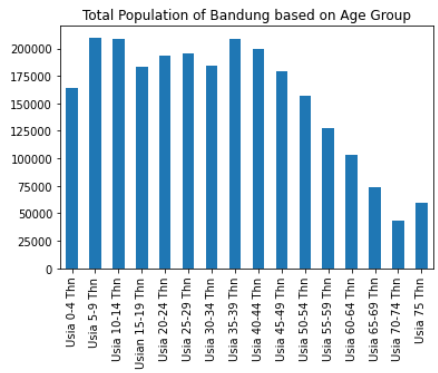
For all the 30 Kecamatans in Bandung, let's see total population for each Kecamatan.



count	30.000000
mean	83012.866667
std	31212.647308
min	25266.000000
25%	70288.500000
50%	79666.000000
75%	101661.500000
max	139374.000000

Figure 3 – Left: Total Population per Kecamatan, Right – Statistical Description of Total Population per Kecamatan

The most populated Kecamatan in Bandung is Babakan Ciparay, with around 139,000 persons, while the least populated is Cinambo (around 25,000 persons). Average total population is around 83,000 persons.



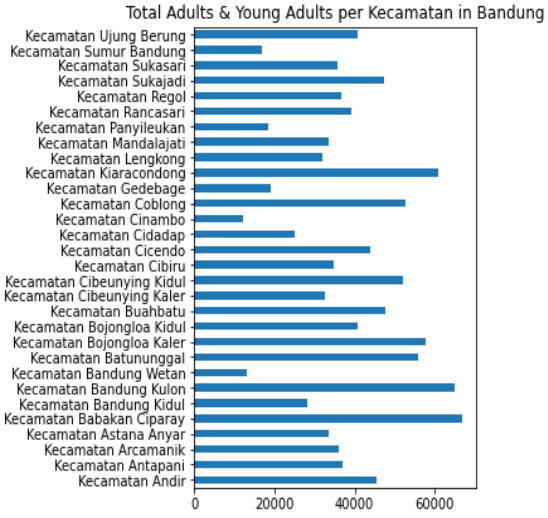


Figure 4 – Left: Total Population Based on Old Age Group, Right: Total Young Adults & Adults per Kecamatan

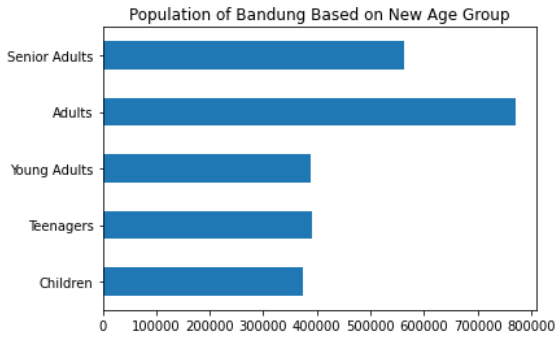


Figure 5 – Population Based on New Age Group

The dataset is recategorized into new age groups with specification:

- Children: 0-9 years old
- Teenagers: 10-19 years old
- Young Adults: 20-29 years old
- Adults: 30-49 years old
- Senior adults: > 49 years old

According to the charts above, Bandung is dominated with Adults age group, followed by Senior Adults. Bandung has big potential for coffee shops expansion, since the number of populations of Adult & Young Adult is quite high. Kecamatan with highest number of adults and young adults are Babakan Ciparay, Kiaracondong, and Bandung Kulon.

B. Exploratory Data Analysis in Bandung Venues

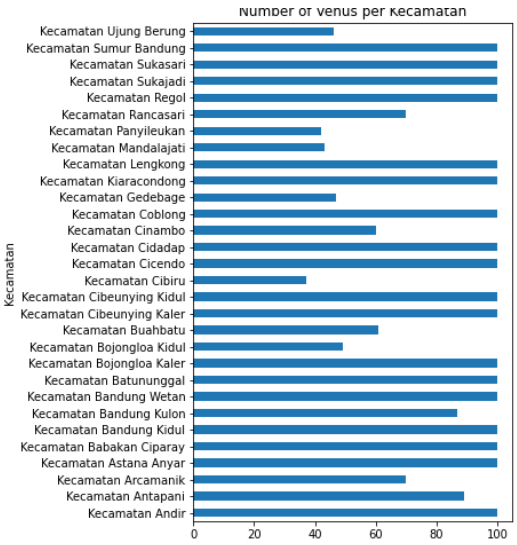
In this part, the venues per Kecamatan will be retrieved using Foursquare API. First, we will retrieve the coordinate of each Kecamatan through geocoder libary. The coordinate of Bandung will also be retrieved.

```
(2501, 7)
```

	Kecamatan	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Kecamatan Andir	-6.9139	107.57435	Noah's Barn Coffeenery	-6.914821	107.577483	Coffee Shop
1	Kecamatan Andir	-6.9139	107.57435	West Point Hotel	-6.912569	107.577947	Hotel
2	Kecamatan Andir	-6.9139	107.57435	Bakmie Royal	-6.920586	107.579469	Noodle House
3	Kecamatan Andir	-6.9139	107.57435	Starbucks	-6.903986	107.579471	Coffee Shop
4	Kecamatan Andir	-6.9139	107.57435	Dapoer Pandan Wangi	-6.928557	107.581916	Sundanese Restaurant
5	Kecamatan Andir	-6.9139	107.57435	d`Groove Sport & Wellness Center	-6.920883	107.574669	Gym / Fitness Center
6	Kecamatan Andir	-6.9139	107.57435	23 Paskal	-6.915411	107.594286	Shopping Mall
7	Kecamatan Andir	-6.9139	107.57435	Domino's Pizza	-6.917921	107.581150	Pizza Place
8	Kecamatan Andir	-6.9139	107.57435	Yogya Sumbersari Junction (Suju)	-6.931714	107.575693	Department Store
9	Kecamatan Andir	-6.9139	107.57435	UNIQLO (ユニクロ) (UNIQLO)	-6.915225	107.594777	Clothing Store

Figure 6 – Snapshot of Venue Dataset

According to the picture above, there are 2501 venues collected around Bandung. Radius is set as 3000 and limit of total venues collected per request is 100.



Neighborhood	
Kecamatan Andir	Hotel
Kecamatan Antapani	Convenience Store
Kecamatan Arcamanik	Convenience Store
Kecamatan Astana Anyar	Hotel
Kecamatan Babakan Ciparay	Indonesian Restaurant
Kecamatan Bandung Kidul	Indonesian Restaurant
Kecamatan Bandung Kulon	Coffee Shop
Kecamatan Bandung Wetan	Hotel
Kecamatan Batununggal	Hotel
Kecamatan Bojongloa Kaler	Hotel
Kecamatan Bojongloa Kidul	Noodle House
Kecamatan Buahbatu	[Asian Restaurant, Bakery, Café, Indonesian Re...
Kecamatan Cibeunying Kaler	Hotel
Kecamatan Cibeunying Kidul	[Bakery, Coffee Shop]
Kecamatan Cibiru	Café
Kecamatan Cicendo	Hotel
Kecamatan Cidadap	[Coffee Shop, Hotel]
Kecamatan Cinambo	Convenience Store
Kecamatan Coblong	Hotel
Kecamatan Gedebage	Indonesian Restaurant
Kecamatan Kiaracondong	Indonesian Restaurant
Kecamatan Lengkong	Indonesian Restaurant
Kecamatan Mandalajati	Convenience Store
Kecamatan Panyileukan	Café
Kecamatan Rancasari	Indonesian Restaurant
Kecamatan Regol	Indonesian Restaurant
Kecamatan Sukajadi	Hotel
Kecamatan Sukasari	Coffee Shop
Kecamatan Sumur Bandung	Hotel
Kecamatan Ujung Berung	Café

Figure 7 – Left: Number of Venues per Kecamatan, Right: Most Common Venues per Kecamatan

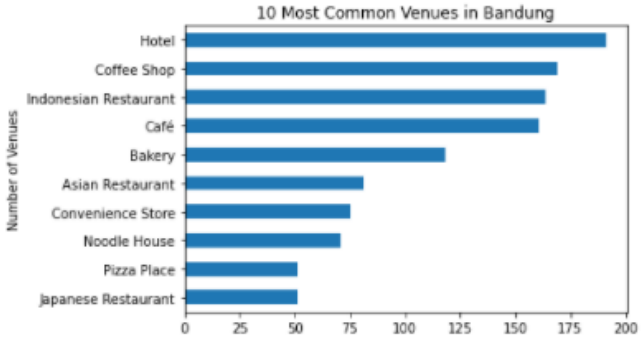


Figure 8 – 10 Most Common Venues in Bandung

According to the chart above, it seems that there are some more crowded Kecamatans, where, in terms of business/number of venues, they have 100 as maximum values. The others are less crowded. There are 167 unique categories of venues in total.

It also can be concluded that Indonesian Restaurant has been the most common venue categories in most of Kecamatans in Bandung, followed by Hotels. Meanwhile, in terms of number of venues per category, Hotel is the most common venues, followed by Coffee Shop and Indonesian Restaurant.

C. Machine Learning Algorithms Used

In this project, we will discover Kecamatans in Bandung that are suitable for people who want to expand their business in coffee shops, by using K-Means clustering with the features of number of existing coffee shops and total populations of adults and young adults per Kecamatan.

The reason of selecting these features are that we want to minimize the possibilites of competitions in the same area and maximize the purchases by selecting areas having more people

with higher possibilities in purchasing powers and eagerness to spend money or time in coffee shops, hence the young adults and adults.

Features will be clustered into 3 groups by using K-Means. The reason of applying this algorithm is features are quite simple, so the process will be quicker yet easier to interpret.

IV. Analysis and Results

A. Pre-Processing

First, for every venue that have been retrieved, we will create one-hot encoding based on the venue type. Then, for each Kecamatan, we will count the average of each venue type, so that we can get the representation of total venue type per Kecamatan as features. For further analysis, we also create 10 most common venue types per Kecamatan.

	Kecamatan	African Restaurant	Airport	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store
0	Kecamatan Andir	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
1	Kecamatan Antapani	0.0	0.0	0.0	0.0	0.022472	0.0	0.0	0.0
2	Kecamatan Arcamanik	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
3	Kecamatan Astana Anyar	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
4	Kecamatan Babakan Ciparay	0.0	0.0	0.0	0.0	0.010000	0.0	0.0	0.0

	Kecamatan	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Kecamatan Andir	Hotel	Indonesian Restaurant	Coffee Shop	Noodle House	Bakery
1	Kecamatan Antapani	Convenience Store	Café	Noodle House	Indonesian Restaurant	Grocery Store
2	Kecamatan Arcamanik	Convenience Store	Grocery Store	Noodle House	Indonesian Restaurant	Café
3	Kecamatan Astana Anyar	Hotel	Coffee Shop	Indonesian Restaurant	Noodle House	Bakery
4	Kecamatan Babakan Ciparay	Indonesian Restaurant	Convenience Store	Café	Pizza Place	Coffee Shop
5	Kecamatan Bandung Kidul	Indonesian Restaurant	Café	Asian Restaurant	Convenience Store	Pizza Place

Figure 9 – Snapshots of Derived Features

The features that will be included for clustering is Averaged number of coffee shop and total young adults & Adults population per Kecamatan. The reason why these age groups are included is that adults have the purchasing power and some of young adults may have too, but they have the eagerness to spend time and money in the coffee shop. The population features will be scaled before clustering.

```
bdg_coffee.head()
```

	Kecamatan	Coffee Shop	Young Adults	Adults
0	Kecamatan Andir	0.070000	0.589805	0.618927
1	Kecamatan Antapani	0.011236	0.454715	0.455771
2	Kecamatan Arcamanik	0.014286	0.432564	0.438841
3	Kecamatan Astana Anyar	0.090000	0.361135	0.402967
4	Kecamatan Babakan Ciparay	0.040000	1.000000	1.000000

Figure 10 – Snapshot of Final Features

B. Result

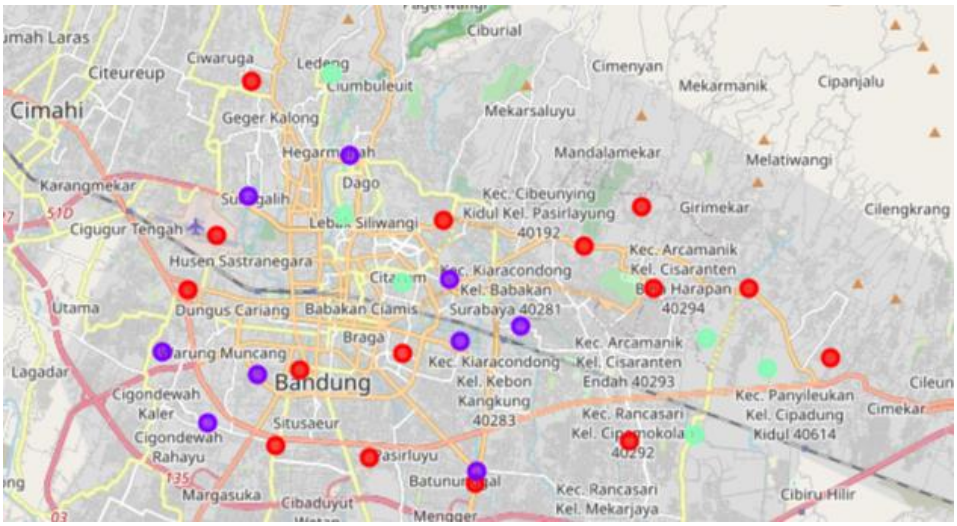


Figure 11 – Clusters of Kecamatan based on K-Means

Cluster 0 is denoted by red, 1 is purple, 2 is light green.

According to the result above, Cluster 0 has the most Kecamatans and Cluster 2 has the least ones. Most of the Kecamatans in Cluster 2 are in the eastern Bandung, while Cluster 1 tends to be in central-western Bandung. Meanwhile Cluster 0 is spread around the city.

	Kecamatan	Young Adults	Adults	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Kecamatan Andir	0.589805	0.618927	Hotel	Indonesian Restaurant	Coffee Shop	Noodle House	Bakery
1	Kecamatan Antapani	0.454715	0.455771	Convenience Store	Café	Noodle House	Indonesian Restaurant	Grocery Store
2	Kecamatan Arcamanik	0.432564	0.438841	Convenience Store	Grocery Store	Noodle House	Indonesian Restaurant	Café
3	Kecamatan Astana Anyar	0.361135	0.402967	Hotel	Coffee Shop	Indonesian Restaurant	Noodle House	Bakery
5	Kecamatan Bandung Kidul	0.289105	0.296504	Indonesian Restaurant	Café	Asian Restaurant	Convenience Store	Pizza Place
10	Kecamatan Bojongloa Kidul	0.498851	0.537308	Noodle House	Indonesian Restaurant	Café	Bakery	Coffee Shop
12	Kecamatan Cibeunying Kaler	0.365401	0.377709	Hotel	Café	Coffee Shop	Bakery	Indonesian Restaurant
14	Kecamatan Cibiru	0.431470	0.403381	Café	Department Store	Supermarket	Indonesian Restaurant	Fried Chicken Joint
15	Kecamatan Cicendo	0.568858	0.589919	Hotel	Coffee Shop	Indonesian Restaurant	Café	Bakery
21	Kecamatan Lengkong	0.356049	0.359786	Indonesian Restaurant	Hotel	Bakery	Coffee Shop	Café
22	Kecamatan Mandalajati	0.392201	0.390917	Convenience Store	Indonesian Restaurant	Breakfast Spot	Seafood Restaurant	Donut Shop
24	Kecamatan Rancasari	0.485233	0.498649	Indonesian Restaurant	Fast Food Restaurant	Convenience Store	Asian Restaurant	Noodle House
25	Kecamatan Regol	0.437486	0.454062	Indonesian Restaurant	Noodle House	Café	Coffee Shop	Bakery
27	Kecamatan Sukasari	0.412711	0.436773	Coffee Shop	Hotel	Asian Restaurant	Café	Clothing Store
29	Kecamatan Ujung Berung	0.521221	0.521508	Café	Department Store	Convenience Store	Indonesian Restaurant	Bakery

Figure 12 - Characteristics of Cluster 0

	Kecamatan	Young Adults	Adults	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Kecamatan Babakan Ciparay	1.000000	1.000000	Indonesian Restaurant	Convenience Store	Café	Pizza Place	Coffee Shop
6	Kecamatan Bandung Kulon	0.965434	0.965064	Coffee Shop	Café	Indonesian Restaurant	Chinese Restaurant	Department Store
8	Kecamatan Batununggal	0.808248	0.796090	Hotel	Indonesian Restaurant	Bakery	Coffee Shop	Café
9	Kecamatan Bojongloa Kaler	0.821757	0.843828	Hotel	Coffee Shop	Chinese Restaurant	Noodle House	Snack Place
11	Kecamatan Buahbatu	0.657351	0.642420	Asian Restaurant	Café	Indonesian Restaurant	Bakery	Hotel
13	Kecamatan Cibeunying Kidul	0.706191	0.738736	Coffee Shop	Bakery	Café	Hotel	Steakhouse
18	Kecamatan Coblong	0.726099	0.752661	Hotel	Café	Asian Restaurant	Coffee Shop	Bakery
20	Kecamatan Kiaracondong	0.887114	0.891138	Indonesian Restaurant	Café	Pizza Place	Hotel	Coffee Shop
26	Kecamatan Sukajadi	0.625738	0.652098	Coffee Shop	Hotel	Café	Japanese Restaurant	Clothing Store

Figure 13 - Characteristics of Cluster 1

	Kecamatan	Young Adults	Adults	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	Kecamatan Bandung Wetan	0.000000	0.019357	Hotel	Coffee Shop	Bakery	Café	Indonesian Restaurant
16	Kecamatan Cidadap	0.229107	0.235014	Coffee Shop	Hotel	Café	Asian Restaurant	Indonesian Restaurant
17	Kecamatan Cinambo	0.000766	0.000000	Convenience Store	Café	Indonesian Restaurant	Department Store	Food Truck
19	Kecamatan Gedebage	0.109877	0.135306	Indonesian Restaurant	Pizza Place	Rest Area	Fast Food Restaurant	Department Store
23	Kecamatan Panyileukan	0.164898	0.089285	Café	Convenience Store	Fried Chicken Joint	Indonesian Restaurant	Donut Shop
28	Kecamatan Sumur Bandung	0.087672	0.085948	Hotel	Bakery	Coffee Shop	Café	Steakhouse

Figure 14 - Characteristics of Cluster 2

Looking at the screenshots of each clusters, it seems the clusters are highly dependant on targeted population rate. Cluster 2 has the least number of young adults and adults, meanwhile Cluster 1 has the highest number.

To sum up how dominant coffee shops in each Kecamatan, we can see the table below.

Clusters	Number of Kecamatan with Coffee Shops as		
	1st Most Common Venues	2nd Most Common Venues	3rd Most Common Venues
Cluster 0	2	2	2
Cluster 1	3	1	0
Cluster 2	1	1	1

Table 1 – Number of Kecamatans with Coffee Shops as 3 Most Common Venus

According to the table above, we can find coffee shops most easily in Kecamatans of Cluster 0 since it has been on the top 3 most common venues of most Kecamatan in cluster 0. Cluster 2 does not really have a lot of coffee shops among the clusters.

V. Discussion

It can be concluded that the clusters are clearly built based on the total of populations first, then followed by the number of coffee shops.

For recommendation of where to build new coffee shops, people should choose Kecamatans in Cluster 1, since this cluster has the highest rate of adults & young adults but have moderate numbers of coffee shops.

If to be detailed, I would further recommend the Kecamatans also based on location near the city center/most visited place or from Cluster 1. In Bandung, most hang out/visited places are in city center towards the north.

Here are top 3 Kecamatans in cluster 1 for building new coffee shops:

1. Kiaracondong (targetted population is around 0.9, coffee shop is the 5th most common place)
2. Batununggal (targetted population is around 0.8, coffee shop is the 4th most common place)
3. Coblong (targetted population is around 0.73, coffee shop is the 4th most common place, region is often visited)

V. Conclusion

Coffee shop is one example of top growing business in Indonesia that targets young adults and adults as the consumers. Bandung, the 4th biggest city in Indonesia is also an interesting target for businessmen/women who intends to expand their coffee shop business, since Bandung has supportive demographic of targeted customers.

Using K-Means, neighborhoods or Kecamatans in Bandung are clustered into 3 clusters based on the total of targeted populations and number of coffee shops. The clusters derived from the K-Means are clearly built on the total targeted population first, in which the Cluster 1 has the highest rate of targeted population and also least amount of Kecamatans that have coffee shops as the most common venues.

Kiaracondong, Batununggal, and Coblong are three recommended Kecamatans from Cluster 1, since they have the combination of highest rate of targeted population, amount of coffee shops, and the location near the city center/most visited places.

My suggestions for further development are:

1. Leveraging other dataset that contains income rate per Kecamatan, to add the knowledge on purchasing power
2. Include data about location/Kecamatans with traffic jams rate, since this may affect delivery punctuality
3. Numbers of ratings submitted to Google or ride hailing app, to give an insight on purchase rates in the area, or even the rating themselves can provide knowledge on where to compete easier.