


Capstone Project - The Battle of Neighborhoods



CHOW KUM SIONG PAUL
PAUL.CHOW.KUM.SIONG@GMAIL.COM

Contents

1. Introduction	2
2. Data required and the source of the data.	2
3. Methodology.....	3
4. Results section	6
5. Discussion, observation and recommendation	6
6. Conclusion.....	7

1. Introduction

As technology advances and the topic of going global takes centre stage in many company expansion plans, the key factors of executing a successful business plan and product marketing campaign still lies within the fundamental concepts of:

- Trading area of a given place of doing business
- Economies of scale in procurement and distribution cost
- Foreseeable volume of demand
- Product positioning and market positioning

The key differentiation between a successful business or otherwise depends largely on how robust, comprehensive and adaptable are the strategies adopted by its management team to address the above factors. The purpose of this project is to examine whether seemingly obvious assumptions based on hearsays, news and qualitative observations about specific market conditions in Singapore still hold true when a data-driven analytical approach is taken instead to presume those assumptions.

This project aims to show aspiring local entrepreneurs, overseas franchiser or even owners of existing family business with long history in Singapore how basic applied data science may possibly help them stand out from the competition by “being at the right place, selling the right thing”, through insights about existing data.

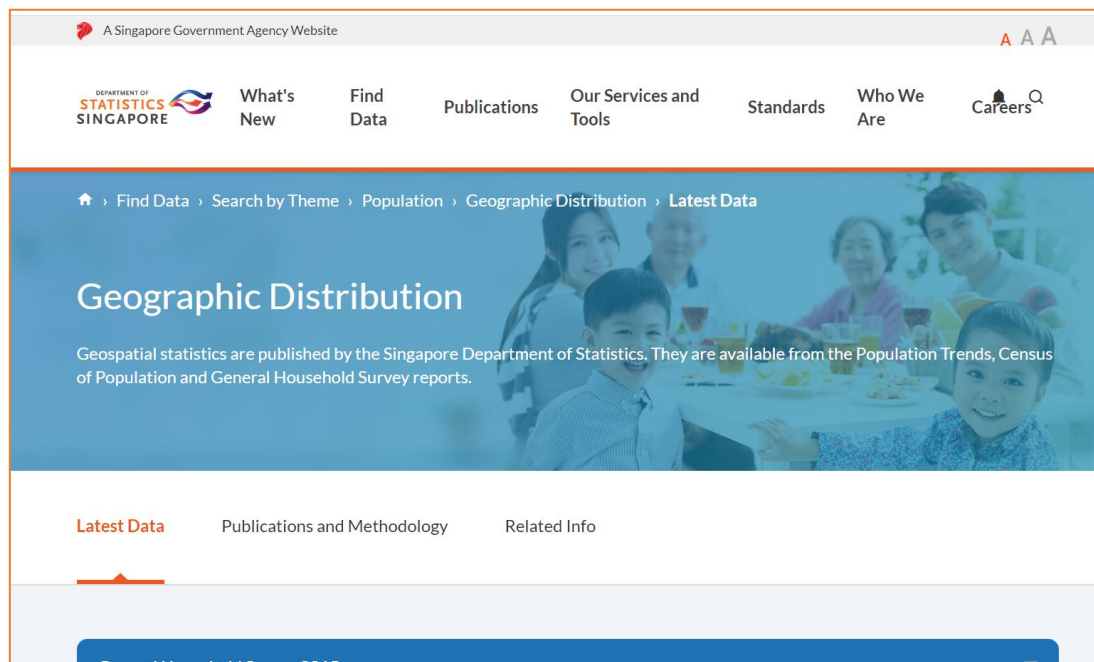
Type of business to be discussed in this project is – supplier of essential ingredients for Asian cuisines.

2. Data required and the source of the data.

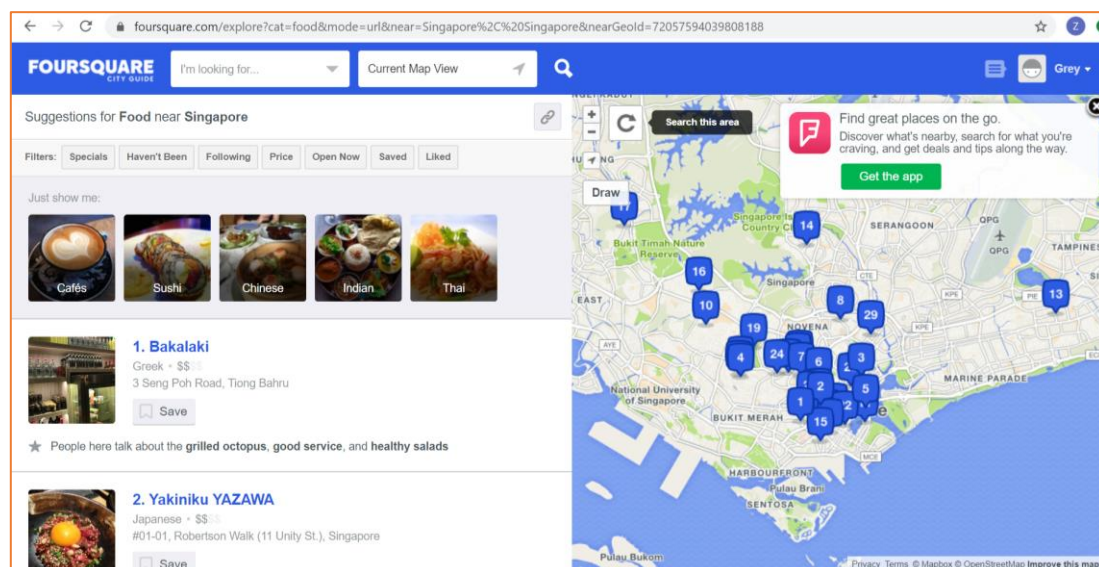
The key data required for the purpose of this project are as follows:

- Locations of selected neighbourhoods in Singapore using <https://www.gps-coordinates.net/> to assess the possible operational issues that comes with travelling time and distance.

- b. No of residents within each neighbourhood to determine the foreseeable demand using data from SingStats.



- c. Foursquare data to identify common venues by consumers.



- d. K-means clustering to segment data from Foursquare, based on similar characteristics.

With the outcomes from the above data processing, the information produced will help businesses make an informed decision and strategize on where the business should be conducted to achieve the highest potential revenue at the lowest possible cost.

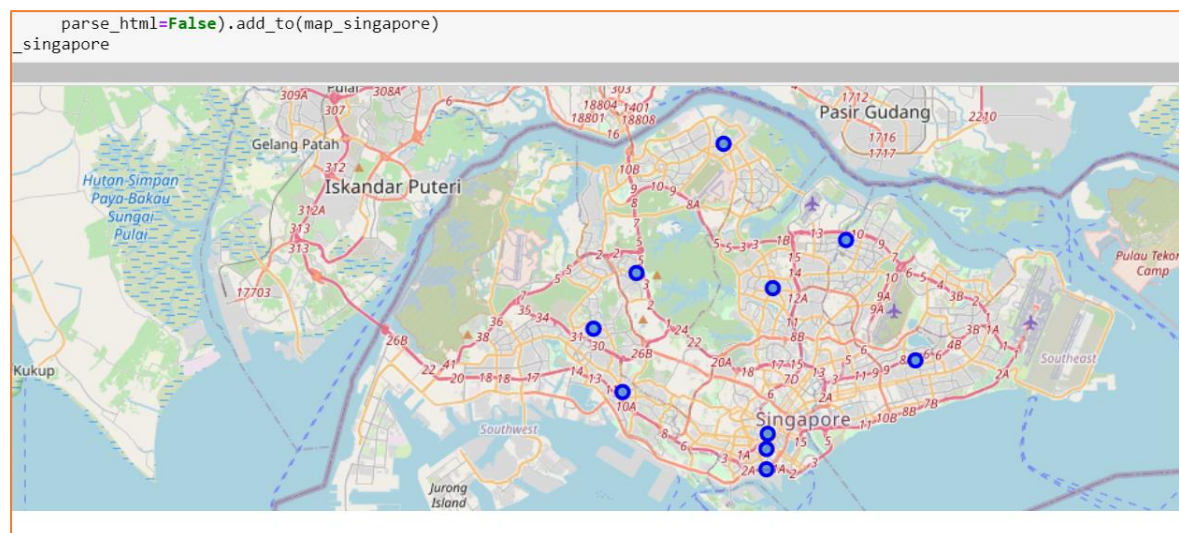
3. Methodology

- Import the necessary libraries in Pandas
- Download and explore the data set of Singapore residents

10J:

	Borough	Neighborhood	Latitude	Longitude	Average no of residents
0	Ang Mo Kio	Ang Mo Kio Town Centre	1.371285	103.846994	1899505
1	Bedok	Bedok North	1.331920	103.924744	1942315
2	Clementi	Clementi Central	1.315100	103.765231	1906745
3	Downtown Core	Anson	1.272658	103.843801	0
4	Sembawang	Sembawang Central	1.449093	103.820056	1922410
5	Bukit Panjang	Bangkit	1.379323	103.772566	1910695
6	Bukit Batok	Bukit Batok Central	1.349091	103.749619	1911280
7	Sengkang	Anchorvale	1.397140	103.886659	1923095
8	Outram	Chinatown	1.283737	103.843798	1906700
9	Museum	Fort Canning	1.292383	103.844242	1900670

- Use geopy library to get the latitude and longitude values of Singapore.
- Create a map of Singapore with 5 neighborhoods superimposed on top



- Define Foursquare Credentials and Version
- Explore Neighborhoods in Singapore

```
In [157]: singapore_venues.groupby('Neighborhood').count()

Out[157]:
```

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Anchorvale	11	11	11	11	11	11
Ang Mo Kio Town Centre	47	47	47	47	47	47
Anson	79	79	79	79	79	79
Bangkit	9	9	9	9	9	9
Bedok North	14	14	14	14	14	14
Bukit Batok Central	23	23	23	23	23	23
Chinatown	100	100	100	100	100	100
Clementi Central	67	67	67	67	67	67
Fort Canning	97	97	97	97	97	97
Sembawang Central	15	15	15	15	15	15

```
In [158]: print('There are {} uniques categories.'.format(len(singapore_venues['Venue Category'].unique())))

There are 145 uniques categories.
```

g. Analyze each Neighborhood

```
In [162]: singapore_grouped = singapore_onehot.groupby('Neighborhood').mean().reset_index()
singapore_grouped

Out[162]:
```

	Neighborhood	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Australian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bank	...	Trail	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant	Waterfront	Whisky Bar
0	Anchorvale	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	...	0.090909	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Ang Mo Kio Town Centre	0.000000	0.000000	0.000000	0.021277	0.00	0.000000	0.000000	0.021277	0.021277	...	0.000000	0.021277	0.000000	0.000000	0.000000	0.000000	0.000000
2	Anson	0.012658	0.000000	0.000000	0.012658	0.00	0.012658	0.012658	0.025316	0.000000	...	0.000000	0.012658	0.000000	0.000000	0.012658	0.000000	0.000000
3	Bangkit	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Bedok North	0.000000	0.000000	0.000000	0.142857	0.00	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Bukit Batok Central	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Chinatown	0.000000	0.010000	0.000000	0.000000	0.01	0.020000	0.000000	0.000000	0.000000	...	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Clementi Central	0.014925	0.000000	0.014925	0.044776	0.00	0.014925	0.000000	0.014925	0.000000	...	0.000000	0.000000	0.014925	0.014925	0.000000	0.000000	0.000000
8	Fort Canning	0.010309	0.010309	0.000000	0.010309	0.00	0.000000	0.000000	0.030928	0.000000	...	0.000000	0.020619	0.000000	0.000000	0.000000	0.010309	0.010309
9	Sembawang Central	0.000000	0.000000	0.000000	0.066667	0.00	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

10 rows × 146 columns

h. Obtain top 10 most common venues of each neighborhood

```
neighborhoods_venues_sorted

Out[167]:
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Anchorvale	Pool	Beer Garden	Food Court	Hockey Arena	Italian Restaurant	Park	Gym / Fitness Center	Trail	Gym	Bridge
1	Ang Mo Kio Town Centre	Food Court	Coffee Shop	Dessert Shop	Bubble Tea Shop	Snack Place	Japanese Restaurant	Fast Food Restaurant	Supermarket	Seafood Restaurant	Sandwich Place
2	Anson	Japanese Restaurant	Coffee Shop	Hotel	Café	Food Court	Soup Place	Ramen Restaurant	Salad Place	Indian Restaurant	Soba Restaurant
3	Bangkit	Food Court	Miscellaneous Shop	Basketball Court	Bike Trail	Park	Noodle House	Market	Flea Market	French Restaurant	Food
4	Bedok North	Coffee Shop	Asian Restaurant	Food	Plaza	Park	Other Great Outdoors	Steakhouse	Supermarket	Food Court	Fast Food Restaurant
5	Bukit Batok Central	Coffee Shop	Fast Food Restaurant	Food Court	Chinese Restaurant	Shopping Mall	Multiplex	Frozen Yogurt Shop	Café	Bowling Alley	Sandwich Place
6	Chinatown	Chinese Restaurant	Food Court	Hotel	Café	Vegetarian / Vegan Restaurant	Spa	Korean Restaurant	Japanese Restaurant	Italian Restaurant	French Restaurant
7	Clementi Central	Food Court	Coffee Shop	Chinese Restaurant	Asian Restaurant	Dessert Shop	Fried Chicken Joint	Fast Food Restaurant	Japanese Restaurant	Shopping Mall	Noodle House
8	Fort Canning	Japanese Restaurant	Hotel	Bar	Nightclub	Bakery	Spa	Pub	Ramen Restaurant	Restaurant	Seafood Restaurant
9	Sembawang Central	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Bus Station	Playground	Asian Restaurant	Shopping Mall	Flower Shop	Japanese Restaurant	Supermarket

- i. Run k-means to cluster the neighborhood into 5 clusters.

singapore_merged # check the last columns!

Out[173]:

	Borough	Neighborhood	Latitude	Longitude	Average no of residents	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ang Mo Kio	Ang Mo Kio Town Centre	1.371285	103.846994	1899505	1	Food Court	Coffee Shop	Dessert Shop	Bubble Tea Shop	Snack Place	Japanese Restaurant	Fast Food Restaurant	Supermarket	Seafood Restaurant	Sandwich Place
1	Bedok	Bedok North	1.331920	103.924744	1942315	1	Coffee Shop	Asian Restaurant	Food	Plaza	Park	Other Great Outdoors	Steakhouse	Supermarket	Food Court	Fast Food Restaurant
2	Clementi	Clementi Central	1.315100	103.765231	1906745	1	Food Court	Coffee Shop	Chinese Restaurant	Asian Restaurant	Dessert Shop	Fried Chicken Joint	Fast Food Restaurant	Japanese Restaurant	Shopping Mall	Noodle House
3	Downtown Core	Anson	1.272658	103.843801	0	2	Japanese Restaurant	Coffee Shop	Hotel	Café	Food Court	Soup Place	Ramen Restaurant	Salad Place	Indian Restaurant	Soba Restaurant
4	Sembawang	Sembawang Central	1.449093	103.820056	1922410	3	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Bus Station	Playground	Asian Restaurant	Shopping Mall	Flower Shop	Japanese Restaurant	Supermarket
5	Bukit Panjang	Bangkit	1.379323	103.772566	1910695	0	Food Court	Miscellaneous Shop	Basketball Court	Bike Trail	Park	Noodle House	Market	Flea Market	French Restaurant	Food
6	Bukit Batok	Bukit Batok Central	1.349091	103.749619	1911280	1	Coffee Shop	Fast Food Restaurant	Food Court	Chinese Restaurant	Shopping Mall	Multiplex	Frozen Yogurt Shop	Café	Bowling Alley	Sandwich Place
7	Sengkang	Anchorvale	1.397140	103.886659	1923095	4	Pool	Beer Garden	Food Court	Hockey Arena	Italian Restaurant	Park	Gym / Fitness Center	Trail	Gym	Bridge
8	Outram	Chinatown	1.283737	103.843798	1906700	2	Chinese Restaurant	Food Court	Hotel	Café	Vegetarian / Vegan Restaurant	Spa	Korean Restaurant	Japanese Restaurant	Italian Restaurant	French Restaurant
9	Museum	Fort Canning	1.292383	103.844242	1900670	2	Japanese Restaurant	Hotel	Bar	Nightclub	Bakery	Spa	Pub	Ramen Restaurant	Restaurant	Seafood Restaurant

- j. Examine the 5 clusters to identify segmentation and locations that will achieve the highest potential in revenue and economies of scale for a business.

4. Results section

In [187]: df_cluster_0=singapore_merged.loc[singapore_merged['Cluster Labels'] == 0, singapore_merged.columns[[1] + list(range(2, singapore_merged.shape[1]))]]

Out[187]:

	Neighborhood	Latitude	Longitude	Average no of residents	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Bangkit	1.379323	103.772566	1910695	0	Food Court	Miscellaneous Shop	Basketball Court	Bike Trail	Park	Noodle House	Market	Flea Market	French Restaurant	Food

In [188]: df_cluster_1=singapore_merged.loc[singapore_merged['Cluster Labels'] == 1, singapore_merged.columns[[1] + list(range(2, singapore_merged.shape[1]))]]

Out[188]:

	Neighborhood	Latitude	Longitude	Average no of residents	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ang Mo Kio Town Centre	1.371285	103.846994	1899505	1	Food Court	Coffee Shop	Dessert Shop	Bubble Tea Shop	Snack Place	Japanese Restaurant	Fast Food Restaurant	Supermarket	Seafood Restaurant	Sandwich Place
1	Bedok North	1.331920	103.924744	1942315	1	Coffee Shop	Asian Restaurant	Food	Plaza	Park	Other Great Outdoors	Steakhouse	Supermarket	Food Court	Fast Food Restaurant
2	Clementi Central	1.315100	103.765231	1906745	1	Food Court	Coffee Shop	Chinese Restaurant	Asian Restaurant	Dessert Shop	Fried Chicken Joint	Fast Food Restaurant	Japanese Restaurant	Shopping Mall	Noodle House
6	Bukit Batok Central	1.349091	103.749619	1911280	1	Coffee Shop	Fast Food Restaurant	Food Court	Chinese Restaurant	Shopping Mall	Multiplex	Frozen Yogurt Shop	Café	Bowling Alley	Sandwich Place

In [189]: df_cluster_2=singapore_merged.loc[singapore_merged['Cluster Labels'] == 2, singapore_merged.columns[[1] + list(range(2, singapore_merged.shape[1]))]]

Out[189]:

	Neighborhood	Latitude	Longitude	Average no of residents	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Anson	1.272658	103.843801	0	2	Japanese Restaurant	Coffee Shop	Hotel	Café	Food Court	Soup Place	Ramen Restaurant	Salad Place	Indian Restaurant	Soba Restaurant
8	Chinatown	1.283737	103.843798	1906700	2	Chinese Restaurant	Food Court	Hotel	Café	Vegetarian / Vegan Restaurant	Spa	Korean Restaurant	Japanese Restaurant	Italian Restaurant	French Restaurant
9	Fort Canning	1.292383	103.844242	1900670	2	Japanese Restaurant	Hotel	Bar	Nightclub	Bakery	Spa	Pub	Ramen Restaurant	Restaurant	Seafood Restaurant

In [190]: df_cluster_3=singapore_merged.loc[singapore_merged['Cluster Labels'] == 3, singapore_merged.columns[[1] + list(range(2, singapore_merged.shape[1]))]]

Out[190]:

	Neighborhood	Latitude	Longitude	Average no of residents	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Sembawang Central	1.449093	103.820056	1922410	3	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Bus Station	Playground	Asian Restaurant	Shopping Mall	Flower Shop	Japanese Restaurant	Supermarket

In [191]: df_cluster_4=singapore_merged.loc[singapore_merged['Cluster Labels'] == 4, singapore_merged.columns[[1] + list(range(2, singapore_merged.shape[1]))]]

Out[191]:

	Neighborhood	Latitude	Longitude	Average no of residents	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Anchorvale	1.397140	103.886659	1923095	4	Pool	Beer Garden	Food Court	Hockey Arena	Italian Restaurant	Park	Gym / Fitness Center	Trail	Gym	Bridge

Based on the results, it is interesting to compare the tradeoffs and advantages between Cluster 1 & 2 in terms of the distance between each similar neighborhoods and the total average no of residents (ie demand volume).

5. Discussion, observation and recommendation

Though Cluster 1 presents the best potential for revenue justified by the high demand in volume which will be generated by the average no of residents in the neighborhoods, the distance between the

neighborhoods within that same cluster is rather far. This will incur more distribution cost during daily operations. Moreover, the first common venue within that cluster is food court or coffeeshop. What this means is that the low market prices of food at these forms of eateries will in turn expect the suppliers of essential ingredients to sell at the corresponding low prices.

In contrast, Cluster 2 has Japanese and Chinese Restaurants as most common venues based on Foursquare data. Based on the location map plotted, those two neighborhoods and the Central Business District are within a relatively shorter distance as compared to other neighborhoods within their respective clusters. In terms of daily operations, this means that the business in Cluster 2 stands to benefit from the economies of scale in terms of distribution and overhead costs.

6. Conclusion

In conclusion, the outcome of this project recommends that the most suitable location to conduct a business to supply essential ingredients for Japanese and Chinese Cuisines is within downtown core areas near Anson Road, Chinatown and Fort Canning. Further to this project, other important factors like property rental price can be included as well as increasing the no. of neighbourhoods in the clustering process to make the outcome more comprehensive and accurate.