

Coursera Capstone Project

A Cluster Analysis of Hong Kong Venues Data from Foursquare

Final Report for the IBM Applied Data Science Capstone Course via Coursera

Richard Williams

21 January 2021

Table of Contents

Table of Contents	2
Executive Summary	3
Introduction	3
Data	4
Methodology	4
Results	5
Discussion	6
Conclusion	7
References	8
Acknowledgements	8
Appendices	8

Executive Summary

This report shows the results of applying cluster analysis to data about venues in Hong Kong, hereafter abbreviated to "HK".

Venues data was sourced from Foursquare, grouped by HK district and cluster analysis applied to the data using the unsupervised machine learning algorithm, K-Means, in order to group the districts into five clusters using venue type as the variable.

The results include:

- (a) An analysis of the city, which would be important for someone planning to visit or live in HK. The data is especially rich in restaurant data.
- (b) An assessment of the completeness of Foursquare data for a city well known by the report writer.

Introduction

Problem

For someone visiting or intending to live in a large new city, it is important to understand its neighbourhoods and what features they have to offer. The visitor might be keen to know about the restaurants in each area: whether the area she intended to visit contained her favourite type of restaurant, or perhaps contained a variety of cuisine types. This report is intended to help a potential visitor to HK better understand the character of its nineteen districts.

Literature review

The reader will find the Internet to be a good source of information about HK: where to stay, things to do, places to eat. At the time of writing, it seems no cluster analysis has so far been performed on venues data in HK in order to categorise its 19 districts by type of venue, so this is the gap which this report aims to begin to bridge, albeit only partially.

HK has been chosen as the subject for this report as it is a location well known by the report writer who is therefore in the position to validate the results of the unsupervised clustering process.

Data

As mentioned, this report is intended to help a potential visitor to HK better understand the character of its nineteen districts. In order to accomplish this, the following data sources are used:

- **A list of the 19 districts** of HK is sourced from Wikipedia at https://en.wikipedia.org/wiki/Districts_of_Hong_Kong.
- **Latitude and longitude** values of the 19 HK districts are sourced from GeoCoder.
- **Data about venues** in HK sourced from Foursquare. Foursquare is a technology company that has built a massive dataset of crowd-sourced global location data. Users of the Foursquare app upload data about venues they visit. The HK data which has currently been uploaded to Foursquare is weighted towards eateries. The Foursquare data can be accessed using the Foursquare API.

Methodology

A Jupiter notebook is used to import and study the above data, since this provides the platform to run the code of various Python data analysis libraries and to view the results. Screenshots from the notebook's results are added to this report.

After importing a list of the 19 HK districts from Wikipedia, latitude and longitude values for these districts are captured from the GeoCoder API. The combination of these data is shown in [Appendix 1](#).

With the latitude and longitude data for the districts prepared, Foursquare is queried to find what data it contains about venues in these districts. At this stage it is possible to see which are the most common venue types per district.

Using the dataset of venues per district, a K-Means algorithm from scikit-learn is called in order to group the districts into clusters of districts, each cluster having a similar composition of venue types. (K-Means is an unsupervised machine learning algorithm, well-suited to finding the similarity of data like districts based on one or more features.) Since it is necessary to input the required number of clusters in advance of running K-Means then it is requested that K-Means groups the 19 HK districts into 5 clusters.

These 5 clusters are then inspected to determine their characteristics and any similarities and dissimilarities are commented on.

Finally, some comment is offered about the completeness of HK Foursquare data.

Results

This section provides a summary of the results found by following the methodology above,

Foursquare returned a total of 658 venues for HK. Here's a sample of these rows:

District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Central and Western	22.28219	114.14490	Caine Road Garden (平頂花園)	22.283006	114.147930	Garden
Central and Western	22.28219	114.14490	Yoga Room Room	22.284453	114.147928	Yoga Studio

From these 658 venues, it was possible to determine a count of HK restaurant venues by cuisine type and the results are shown in [Appendix 2](#).

Also it was possible to determine the top 10 venue types for 17 of the 19 districts. Two districts didn't have enough reported data. The results are at [Appendix 3](#).

Then K-Means was run in order to classify the above dataset (the top 10 venue types for each of the 17 districts) into 5 clusters. The tabular results are at [Appendix 4](#).

A map of the 5 clusters was drawn using Folium and the results shown at [Appendix 5](#).

The final step was to examine the data of each cluster manually, to note which districts are in each cluster and which venue types are most common in each district. From this analysis, it is possible to determine which characteristics are common within each cluster and which are dissimilar between clusters. The composition of each cluster is shown in [Appendix 6](#).

A discussion of these results follows next.

Discussion

Based on the results above, the observations drawn in this report divide into two areas:

- An analysis of the city.
- The completeness of Foursquare data.

An Analysis of the City

Looking at the data divided into 5 clusters, it is observed that:

- Out of the seventeen districts which had enough data to use, thirteen all fell into the same cluster, meaning that the top ten reported venues in these districts are more similar to each other than to other clusters.
- Since the Foursquare data is dominated by food-related venues, it would be reasonable to conclude that any one of these thirteen districts would be a similar place to visit or settle from a food-perspective.
- However, since it is known that there are interesting differences between many of these thirteen districts, this indicates that the features which distinguish them were either not captured in the list of top ten venue types per district or that a higher number of clusters should be allowed in order find those distinctions.
- The remaining 4 districts fell into 4 other clusters, making these districts more unique in their top ten values than the others. For example, Kowloon City seems to distinguish itself for hotpot restaurants.

- Regardless of their differences, all districts appear to have a good range of facilities. This means that a visitor could stay virtually anywhere in the city and not find herself stranded without something to eat!
- The top 10 venue types by cluster and district is shown in appendix 5 for the interested reader to browse the data further.

The Completeness of Foursquare Data

Also, regarding the Foursquare data, it is observed that:

- For two of the districts, "Eastern" and "Sai Kung", there was not enough venue data available from Foursquare for them to be used in this exercise.
- In fact with only 658 rows, the size of the dataset of reported venues in Hong Kong is small and perhaps the usage of Foursquare itself should be promoted in Hong Kong in order to capture more data.
- It is natural to assume that Foursquare users who have uploaded data have focussed on the type of venues they are keen to recommend. This is evidenced by the fact that most uploaded venue types are some form of eatery.
- This means however that the data is lacking in more mundane venue types like schools, clinics, hospitals, churches and temples. These are almost completely lacking in the HK dataset.

Conclusion

Regarding a visit to Hong Kong, it can be concluded from the Foursquare dataset that a visitor should have no concern about finding something to eat since there is a good variety of eateries in all districts of HK. The Foursquare data could also be used usefully to navigate to the more popular venues.

However, it is suggested that at this stage that the Foursquare data is not yet as useful as other well known sites on the Internet for researching HK generally. Perhaps the biggest finding of this analysis is that more HK venues data is needed. Perhaps Foursquare is

currently less popular in HK than in other places. It would also be useful to have a wider variety of venue types reported in HK, not just eateries, and it is hoped that this may come naturally over time as the size of the HK dataset increases.

References

- Foursquare data, see <https://foursquare.com/>
- Jupyter notebooks, see <https://jupyter.org/>
- Retrieving latlong values, see <https://geocoder.readthedocs.io/index.html>
- K-Means clustering, see https://en.wikipedia.org/wiki/K-means_clustering
- Drawing maps with Folium, see <https://python-visualization.github.io/folium/modules.html>

Acknowledgements

This report has been compiled as part of the assessment for the Applied Data Science Capstone course provided by IBM and hosted by Coursera. For this reason, the author thanks the IBM course tutors for the idea for this assessment and for the necessary training in order to complete it.

Appendices

See below.

Appendix 1 - Hong Kong's 19 Districts with Location Data added

	District	Population	PopulationGrowth	Density	Latitude	Longitude
0	Central and Western	243266	-2.7	19391	22.282190	114.144860
1	Eastern	555034	-2.8	30861	22.272090	114.271396
2	Southern	274994	-0.6	7080	22.258010	114.153080
3	Wan Chai	180123	-0.1	17137	22.277101	114.173337
4	Sham Shui Po	405869	11.0	43381	22.329351	114.159179
5	Kowloon City	418732	15.6	41802	22.311130	114.183540
6	Kwun Tong	648541	10.4	57530	22.314236	114.226625
7	Wong Tai Sin	425235	0.4	45711	22.336088	114.191726
8	Yau Tsim Mong	342970	22.3	48046	22.309739	114.168521
9	Islands	156801	14.4	886	22.314680	113.932430
10	Kwai Tsing	520572	-0.5	22307	22.314236	114.226625
11	North	315270	12.3	2310	22.513688	114.206956
12	Sai Kung	461864	13.6	3563	22.400012	114.313797
13	Sha Tin	659794	8.6	9602	22.382036	114.202102
14	Tai Po	303926	3.5	2233	22.290868	113.952263
15	Tsuen Wan	318916	10.5	5149	22.370660	114.104790
16	Tuen Mun	489299	-2.5	5894	22.395699	113.973083
17	Yuen Long	607200	15.0	4436	22.445131	114.025732
18	Marine	1201	-60.8	0	22.276640	114.171370

Appendix 2 - Count of Reported Venues in HK: Restaurants by Cuisine type

Venue Category	Count
Chinese Restaurant	32
Cantonese Restaurant	24
Japanese Restaurant	23
Hong Kong Restaurant	22
Fast Food Restaurant	21
Italian Restaurant	16
Thai Restaurant	16
Vietnamese Restaurant	9
Hotpot Restaurant	9
Sushi Restaurant	8
Seafood Restaurant	6
Korean Restaurant	6
Dumpling Restaurant	6
French Restaurant	5
Vegetarian / Vegan Restaurant	5
Dim Sum Restaurant	5
Spanish Restaurant	4
Restaurant	4
Indian Restaurant	4
Beijing Restaurant	4
Middle Eastern Restaurant	4
Malay Restaurant	3
Mexican Restaurant	3
Ramen Restaurant	3
Donburi Restaurant	3
American Restaurant	3
Indonesian Restaurant	3
Asian Restaurant	3
Lebanese Restaurant	2
Greek Restaurant	2
Tapas Restaurant	2

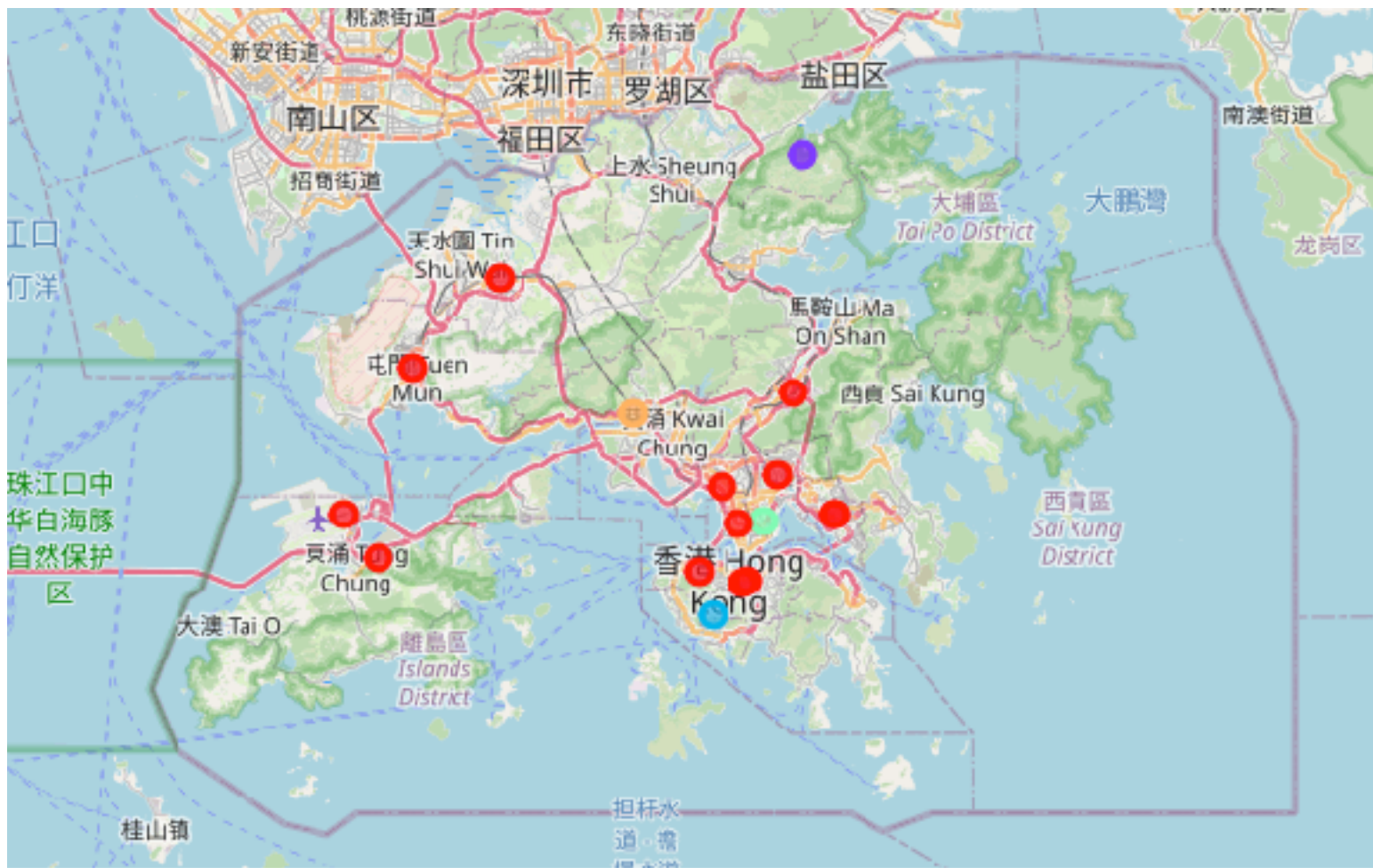
Appendix 3 - The Top 10 Reported Venue Types in HK per District

District	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
Central and Western	Yoga Studio	Café	Thai Restaurant	Italian Restaurant	Mexican Restaurant	Coffee Shop	Tapas Restaurant	Noodle House	Tea Room	Japanese Restaurant
Islands	Airport Lounge	Coffee Shop	Bakery	Clothing Store	Cocktail Bar	Electronics Store	Candy Store	Pizza Place	Chocolate Shop	Airport
Kowloon City	Hotpot Restaurant	Noodle House	Bus Stop	Dessert Shop	Fast Food Restaurant	Theater	Airport	Park	Organic Grocery	Paper / Office Supplies Store
Kwai Tsing	Chinese Restaurant	Japanese Restaurant	Cha Chaan Teng	Fast Food Restaurant	Restaurant	Sushi Restaurant	Coffee Shop	Café	Department Store	Hong Kong Restaurant
Kwai Tung	Chinese Restaurant	Japanese Restaurant	Cha Chaan Teng	Fast Food Restaurant	Restaurant	Sushi Restaurant	Coffee Shop	Café	Department Store	Hong Kong Restaurant
Marine	Coffee Shop	Luncheon Restaurant	Italian Restaurant	Japanese Restaurant	Café	Hong Kong Restaurant	Hotel	Chinese Restaurant	Thai Restaurant	Bakery
North	Waterfall	Airport	Performing Arts Venue	Molecular Gastronomy Restaurant	Multiplex	Noodle House	Organic Grocery	Paper / Office Supplies Store	Park	Perfume Shop
Sha Tin	Fast Food Restaurant	Chinese Restaurant	Karaoke Bar	Hong Kong Restaurant	Noodle House	Sushi Restaurant	Café	Park	Train Station	Bus Station
Sham Shui Po	Woods House	Hong Kong Restaurant	Convenience Shop	Italian Restaurant	Snack Place	Shopping Mall	Ramen Restaurant	Flea Market	Cha Chaan Teng	Cafe
Southern	Waterfall	Grocery Store	Trail	Airport	Paper / Office Supplies Store	Mobile Phone Shop	Molecular Gastronomy Restaurant	Multiplex	Noodle House	Organic Grocery
Tai Po	Bus Stop	Convenience Store	Shopping Mall	Furniture / Home Store	Butcher	Cantonese Restaurant	Park	Fast Food Restaurant	Bus Station	Supermarket
Tuen Mun	Bus Station	Fast Food Restaurant	Bakery	Bus Stop	Dim Sum Restaurant	Hong Kong Restaurant	Supermarket	Noodle House	Organic Grocery	Paper / Office Supplies Store
Tuen Mun	Shopping Mall	Coffee Shop	Fast Food Restaurant	Multiplex	Hong Kong Restaurant	Dessert Shop	Clothing Store	Cantonese Restaurant	Supermarket	Shanghai Restaurant
Wan Chai	Coffee Shop	Cantonese Restaurant	Japanese Restaurant	Hotel	Hong Kong Restaurant	Café	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Bakery
Wong Tai Sin	Chinese Restaurant	Noodle House	Supermarket	Metro Station	Coffee Shop	Shopping Mall	Cha Chaan Teng	Cantonese Restaurant	Park	Bus Station
Yau Tsim Mong	Hotel	Cantonese Restaurant	Dessert Shop	Chinese Restaurant	Hong Kong Restaurant	BBQ Joint	Dim Sum Restaurant	Café	Noodle House	Indian Restaurant
Yuen Long	Woods House	Fast Food Restaurant	Chinese Restaurant	Japanese Restaurant	Shopping Mall	Market	Coffee Shop	Thai Restaurant	Dumpling Restaurant	Dessert Shop

Appendix 4 - The Top 10 Reported Venue Types in HK per District with Cluster Number

District	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
Central and Western	0	Yoga Studio	Café	Thai Restaurant	Italian Restaurant	Mexican Restaurant	Coffee Shop	Tapas Restaurant	Noodle House	Tea Room	Japanese Restaurant
Islands	0	Airport Lounge	Coffee Shop	Bakery	Clothing Store	Cocktail Bar	Electronics Store	Candy Store	Pizza Place	Chocolate Shop	Airport
Kowloon City	3	Hotpot Restaurant	Noodle House	Bus Stop	Dessert Shop	Fast Food Restaurant	Theater	Airport	Pan	Organic Grocery	Paper / Office Supplies Store
Kwai Tsing	0	Chinese Restaurant	Japanese Restaurant	Cha Chuan Teng	Fast Food Restaurant	Restaurant	Sushi Restaurant	Coffee Shop	Café	Department Store	Hong Kong Restaurant
Kwai Tong	0	Chinese Restaurant	Japanese Restaurant	Cha Chuan Teng	Fast Food Restaurant	Restaurant	Sushi Restaurant	Coffee Shop	Café	Department Store	Hong Kong Restaurant
Kowloon	0	Coffee Shop	Cantonese Restaurant	Italian Restaurant	Japanese Restaurant	Café	Hong Kong Restaurant	Hotel	Chinese Restaurant	Thai Restaurant	Bakery
North	1	Waterfall	Airport	Performing Arts Venue	Molecular Gastronomy Restaurant	Multiplex	Noodle House	Organic Grocery	Paper / Office Supplies Store	Park	Bakery Shop
Sha Tin	0	Fast Food Restaurant	Chinese Restaurant	Karaoke Bar	Hong Kong Restaurant	Noodle House	Snack Restaurant	Café	Pan	Train Station	Bus Station
Sham Shui Po	0	Noodle House	Hong Kong Restaurant	Dessert Shop	Italian Restaurant	Snack Place	Shopping Mall	Ramen Restaurant	Flea Market	Cha Chuan Teng	Café
Southern	2	Waterfall	Grocery Store	Tail	Airport	Paper / Office Supplies Store	Mobile Phone Shop	Molecular Gastronomy Restaurant	Multiplex	Noodle House	Organic Grocery
Tai Po	0	Bus Stop	Convenience Store	Shopping Mall	Furniture / Home Store	Butcher	Cantonese Restaurant	Park	Fast Food Restaurant	Bus Station	Supermarket
Tsuen Wan	4	Bus Station	Fast Food Restaurant	Bakery	Bus Stop	Dim Sum Restaurant	Hong Kong Restaurant	Supermarket	Noodle House	Organic Grocery	Paper / Office Supplies Store
Tuen Mun	0	Shopping Mall	Coffee Shop	Fast Food Restaurant	Multiplex	Hong Kong Restaurant	Dessert Shop	Chinese Restaurant	Cantonese Restaurant	Supermarket	Shanghai Restaurant
Wai Loh	0	Lemon Shop	Cantonese Restaurant	Japanese Restaurant	Hotel	Hong Kong Restaurant	Lake	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Bakery
Wong Tai Sin	0	Chinese Restaurant	Noodle House	Supermarket	Metro Station	Coffee Shop	Shopping Mall	Cha Chuan Teng	Cantonese Restaurant	Park	Bus Station
Yai Tsai Mong	0	Hotel	Cantonese Restaurant	Dessert Shop	Chinese Restaurant	Hong Kong Restaurant	BBQ Joint	Dim Sum Restaurant	Café	Noodle House	Indian Restaurant
Yuen Long	0	Noodle House	Fast Food Restaurant	Chinese Restaurant	Japanese Restaurant	Shopping Mall	Market	Coffee Shop	Thai Restaurant	Dumpling Restaurant	Dessert Shop

Appendix 5 - Map of Clusters (indicated by different colours)



Appendix 6 - The Top 10 Reported Venue Types per Cluster and District

Cluster 1											
District	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	
Central and Western	Yoga Studio	Café	Thai Restaurant	Italian Restaurant	Mexican Restaurant	Coffee Shop	Tapas Restaurant	Noodle House	Tea Room	Japanese Restaurant	
Wan Chai	Coffee Shop	Cantonese Restaurant	Japanese Restaurant	Hotel	Hong Kong Restaurant	Café	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Bakery	
Sham Shui Po	Noodle House	Hong Kong Restaurant	Dessert Shop	Italian Restaurant	Snack Place	Shopping Mall	Ramen Restaurant	Flea Market	Cha Chaan Teng	Café	
Kwun Tong	Chinese Restaurant	Japanese Restaurant	Cha Chaan Teng	Fast Food Restaurant	Restaurant	Sushi Restaurant	Coffee Shop	Café	Department Store	Hong Kong Restaurant	
Wong Tai Sin	Chinese Restaurant	Noodle House	Supermarket	Metro Station	Coffee Shop	Shopping Mall	Cha Chaan Teng	Cantonese Restaurant	Park	Bus Station	
Yau Tsim Mong	Hotel	Cantonese Restaurant	Dessert Shop	Chinese Restaurant	Hong Kong Restaurant	BBQ Joint	Dim Sum Restaurant	Café	Noodle House	Indian Restaurant	
Islands	Airport/Lounge	Coffee Shop	Bakery	Clothing Store	Cocktail Bar	Electronics Store	Candy Store	Pizza Place	Chocolate Shop	Airport	
Kwai Tsing	Chinese Restaurant	Japanese Restaurant	Cha Chaan Teng	Fast Food Restaurant	Restaurant	Sushi Restaurant	Coffee Shop	Café	Department Store	Hong Kong Restaurant	
Sha Tin	Fast Food Restaurant	Chinese Restaurant	Karaoke Bar	Hong Kong Restaurant	Noodle House	Sushi Restaurant	Café	Park	Train Station	Bus Station	
Tai Po	Bus Stop	Convenience Store	Shopping Mall	Furniture / Home Store	Butcher	Cantonese Restaurant	Park	Fast Food Restaurant	Bus Station	Supermarket	
Tuen Mun	Shopping Mall	Coffee Shop	Fast Food Restaurant	Multiplex	Hong Kong Restaurant	Dessert Shop	Clothing Store	Cantonese Restaurant	Supermarket	Shanghai Restaurant	
Yuen Long	Noodle House	Fast Food Restaurant	Chinese Restaurant	Japanese Restaurant	Shopping Mall	Market	Coffee Shop	Thai Restaurant	Dumpling Restaurant	Dessert Shop	
Marine	Coffee Shop	Cantonese Restaurant	Italian Restaurant	Japanese Restaurant	Café	Hong Kong Restaurant	Hotel	Chinese Restaurant	Thai Restaurant	Bakery	
Cluster 2											
District	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	
North	Waterfall	Airport	Performing Arts Venue	Molecular Gastronomy Restaurant	Multiplex	Noodle House	Organic Grocery	Paper / Office Supplies Store	Park	Perfume Shop	
Cluster 3											
District	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	
Southern	Waterfall	Grocery Store	Trail	Airport	Paper / Office Supplies Store	Mobile Phone Shop	Molecular Gastronomy Restaurant	Multiplex	Noodle House	Organic Grocery	
Cluster 4											
District	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	
Kowloon City	Hotpot Restaurant	Noodle House	Bus Stop	Dessert Shop	Fast Food Restaurant	Theater	Airport	Park	Organic Grocery	Paper / Office Supplies Store	
Cluster 5											
District	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	
Tsuen Wan	Bus Station	Fast Food Restaurant	Bakery	Bus Stop	Dim Sum Restaurant	Hong Kong Restaurant	Supermarket	Noodle House	Organic Grocery	Paper / Office Supplies Store	