

Explore along Hong Kong Tramways

1. Introduction

1.1 Background

Since 1904, trams have been running from East to West on Hong Kong Island. Today, the tram becomes a nostalgic mode of public transport to explore the city. Connecting the heritage past and new vibes along its tracks, it's an experience for both locals and travellers to hop on and off the trams and discover hidden gems and new attractions along its path.

In this project, I want to look into the neighbourhoods of the tram stops to identify and classify the venues in proximity. For examples, for some tram stops are close to a lot of cafe and restaurants while others might be in the shopping hubs, or in vicinity to arts venues, or close to parks.

1.2 Application of the analysis

By identifying and classification of the establishments close to the tram stops, one would be able to design his/her own itineraries of different themes (or blended) along the tramways. For instance:-

- Foodie Tour
- Shopping Tour
- Arts Tour
- Outdoor Recreational Tour

The analysis would be useful for tour companies, tour guides, guide book publishers/website operators or any activity organizers who are interested in planning a themed travel along the tramways for profit making, educational, cultural or leisure purposes.

2. Data

2.1 Data Required

The main data required for this project include:-

- (a) location of the tram stops and;
- (b) venues close to the respective tram stops.

2.2 Data Acquisition

(a) Location of the tram stops

As there is no ready made data-frame corresponding the tram stops to their latitudes and longitudes, I have to manually prepare the location data for this project.

Names and neighbourhoods of each tram stops can be found in the official website of the Hong Kong Tramways.

The coordinates of each tram stops is then obtained from Google Map.

(b) Venues close to the respective tram stops

Besides, I use the Foursquare API to explore the venues close to each tram stop.

As Foursquare also grouped the high-level venue data, i.e. all venues, into some sub-categories, I am able to obtain venue data with respect to different themes. For examples:

- Arts & Entertainment (4d4b7104d754a06370d81259)
- Food (4d4b7105d754a06374d81259)
- Outdoors & Recreation (4d4b7105d754a06377d81259)
- Shop & Services (4d4b7105d754a06378d81259)

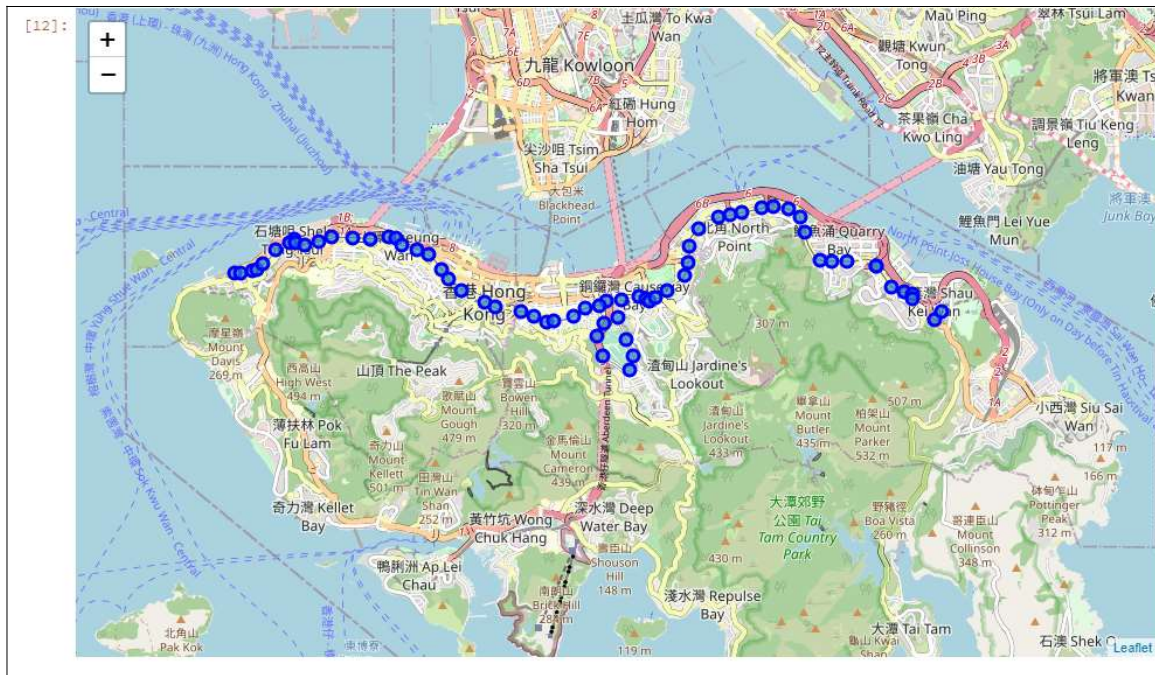
2.3 Data Preparation

Having obtained the location and venues data for each tram stops, I put the data together into data frames. 4 data-frames are prepared for analysis of the four mentioned categories of data.

3. Visualization and Data Exploration

3.1 Folium Library and Leaflet Map

Having scrapped the data of tram stops and their coordinates from the web site Hong Kong Tramways and Google Map, I use the Folium library to create a leaflet map to visualize the distribution of tram stops.



Note: Tram stops are shown as blue spots on the above map

3.2 Foursquare API Venue Data Exploration

Having identify the tram stops, I used the Foursquare API and its sub-categories ID to retrieve the top 100 venues within a radius of 500 m from each tram stops.

Firstly, I worked on the Art & Entertainment venues.

The codes to retrieve Arts and Entertainment venues data are as follows:-

```
url      = 'https://api.foursquare.com/v2/venues/explore?
client_id={}&client_secret={}&ll={},
{}&v={}&radius={}&limit={}&categoryId={}'.format(CLIENT_ID,
neighborhood_latitude,
neighborhood_longitude, VERSION, 500, 100, CATEGORY_ID)

results = requests.get(url).json()
results
```

A sample of the result obtained is as follows:-

```
{'meta': {'code': 200, 'requestId': '5f38b17ccf2d854abd37a3f7'},
 'response': {'headerLocation': 'Kennedy Town',
 'headerFullLocation': 'Kennedy Town, Hong Kong',
 'headerLocationGranularity': 'neighborhood',
 'query': 'arts entertainment',
 'totalResults': 4,
 ...}
```

The above request is repeated for each of the tram stops and put into a data-frame.

[26]:	Neighborhoods	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kennedy Town Terminus	22.283015	114.126199	Drawgraphy 童畫	22.282480	114.128278	Art Gallery
1	Kennedy Town Terminus	22.283015	114.126199	Channy Stage Arts Workshop	22.282160	114.129360	Dance Studio
2	Kennedy Town Terminus	22.283015	114.126199	DreamLikeBubbles	22.282760	114.129989	Art Gallery
3	Kennedy Town Terminus	22.283015	114.126199	Shc	22.283410	114.130470	Dance Studio
4	Davis Street	22.283102	114.126969	Drawgraphy 童畫	22.282480	114.128278	Art Gallery
5	Davis Street	22.283102	114.126969	Channy Stage Arts Workshop	22.282160	114.129360	Dance Studio
6	Davis Street	22.283102	114.126969	DreamLikeBubbles	22.282760	114.129989	Art Gallery
7	Davis Street	22.283102	114.126969	Shc	22.283410	114.130470	Dance Studio
8	Davis Street	22.283102	114.126969	Galerie Koo	22.285532	114.130828	Art Gallery
9	Davis Street	22.283102	114.126969	Loom Art	22.283388	114.131709	Art Gallery

It is noted that there are different kinds of venue clustered around each tram stops. To tally the frequency of each kind of venue for every tram stops, I proceed with the followings:-

- Create a data-frame with panda one hot encoding for the venue categories.
- Use pandas group by the tram stops (i.e. the “Neighborhoods” column of my data-frame) and obtain the mean of the one-hot encoded venue categories.

```
[30]: # one hot encoding
hk_onehot = pd.get_dummies(hk_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
hk_onehot['Neighborhoods'] = hk_venues['Neighborhoods']

# move neighborhoods column to the first column
fixed_columns = [hk_onehot.columns[-1]] + list(hk_onehot.columns[:-1])
hk_onehot = hk_onehot[fixed_columns]

hk_onehot.head()
```

```
[30]: hk_onehot.shape
```

```
[30]: (591, 26)
```

The Art & Entertainment venue has 591 entries comes in 26 categories.

I grouped the venues with respect to each tram stops, then took the mean of frequency for occurrence of each category.

```
[32]: hk_grouped = hk_onehot.groupby('Neighborhoods').mean().reset_index()
hk_grouped
```

```
[32]:
```

	Neighborhoods	Art Gallery	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio	History Museum	Laser Tag	Mini Golf	Movie Theater	Multiplex	Museum	Music Venue	Opera House	Outdoor Sculpture	Performing Arts Venue
0	Admiralty MTR Station	0.400000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.200000	0.200000	0.000000	0.000000	0.00	0.000000	0.000000
1	Arsenal Street	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.133333	0.066667	0.000000	0.066667	0.00	0.000000	0.133333
2	Bank Street	0.560000	0.000000	0.040000	0.000000	0.040000	0.000000	0.120000	0.040000	0.000000	0.000000	0.000000	0.040000	0.040000	0.040000	0.00	0.000000	0.040000
3	Broadwood Road	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
4	Burrows Street	0.142857	0.000000	0.000000	0.000000	0.000000	0.000000	0.285714	0.285714	0.000000	0.000000	0.000000	0.000000	0.000000	0.142857	0.00	0.000000	0.000000
5	Canal Road West	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.166667	0.000000	0.166667	0.000000	0.000000	0.333333	0.000000	0.166667	0.00	0.000000	0.000000
6	Causeway Bay Terminus	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.100000	0.100000	0.000000	0.500000	0.000000	0.000000	0.00	0.000000	0.000000
7	Chai Wan Road	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
8	Connaught Road West	0.600000	0.000000	0.000000	0.000000	0.000000	0.066667	0.200000	0.000000	0.000000	0.066667	0.000000	0.000000	0.066667	0.000000	0.00	0.000000	0.000000
9	Cotton Tree Drive	0.500000	0.000000	0.000000	0.000000	0.083333	0.000000	0.000000	0.083333	0.000000	0.000000	0.083333	0.083333	0.083333	0.000000	0.00	0.000000	0.000000
10	Davis Street	0.666667	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000

By sorting the data-frame, the most common categories of venues with respect to each trams station are identified as shown below.

```
[35]:
```

	Neighborhoods	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	Admiralty MTR Station	Art Gallery	Multiplex	History Museum	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio
1	Arsenal Street	Art Gallery	Rock Club	Theater	Performing Arts Venue	Movie Theater	Music Venue	Multiplex	Laser Tag	Art Museum	Arts & Entertainment
2	Bank Street	Art Gallery	Dance Studio	Music Venue	Arts & Entertainment	Concert Hall	History Museum	Multiplex	Theater	Museum	Performing Arts Venue
3	Broadwood Road	Arts & Entertainment	Racecourse	History Museum	VR Cafe	Multiplex	Art Museum	Comedy Club	Concert Hall	Country Dance Club	Dance Studio
4	Burrows Street	Dance Studio	History Museum	Art Gallery	Rock Club	Music Venue	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall
5	Canal Road West	Multiplex	VR Cafe	Dance Studio	Laser Tag	Music Venue	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
6	Causeway Bay Terminus	Multiplex	Dance Studio	Art Gallery	Laser Tag	Mini Golf	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
7	Chai Wan Road	History Museum	Movie Theater	VR Cafe	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio
8	Connaught Road West	Art Gallery	Dance Studio	Mini Golf	Country Dance Club	Museum	Rock Club	Racecourse	Art Museum	Arts & Entertainment	Comedy Club
9	Cotton Tree Drive	Art Gallery	Theater	Concert Hall	History Museum	Movie Theater	Multiplex	Museum	Public Art	Performing Arts Venue	Art Museum

3.3 Clustering of tram stops for Art and Entertainment Venues

With venue data arranged in appropriate data-frame, I conducted k-means to constellate the tram stops into 5 clusters.

```
[36]: # set number of clusters
kclusters = 5

hk_grouped_clustering = hk_grouped.drop('Neighborhoods', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(hk_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:15]

[36]: array([4, 4, 1, 0, 4, 2, 2, 4, 1, 1, 1, 1, 1, 2, 3], dtype=int32)
```

After running the above code, the trams stops are classified into 5 clusters. A data-frame showing the tram stops, their cluster labels and the top 10 venues of each stop is created as shown below.

```
[37]: # add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

[38]: hk_merged = df

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
hk_merged = hk_merged.join(neighborhoods_venues_sorted.set_index('Neighborhoods'), on='Station')
hk_merged.head() # check the last columns!
```

	Station No.	Station	Neighborhood	lat	long	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
0	KTT	Kennedy Town Terminus	Kennedy Town	22.283015	114.126199	1	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Cinema	Bar
1	104W	Davis Street	Kennedy Town	22.283102	114.126969	1	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Cinema	Bar
2	102W	Smithfield	Kennedy Town	22.283427	114.128774	1	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Cinema	Bar
3	100W	Sands Street	Kennedy Town	22.283506	114.129394	1	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Cinema	Bar
4	98W	Kennedy Town Praya	Kennedy Town	22.284337	114.130346	1	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Cinema	Bar

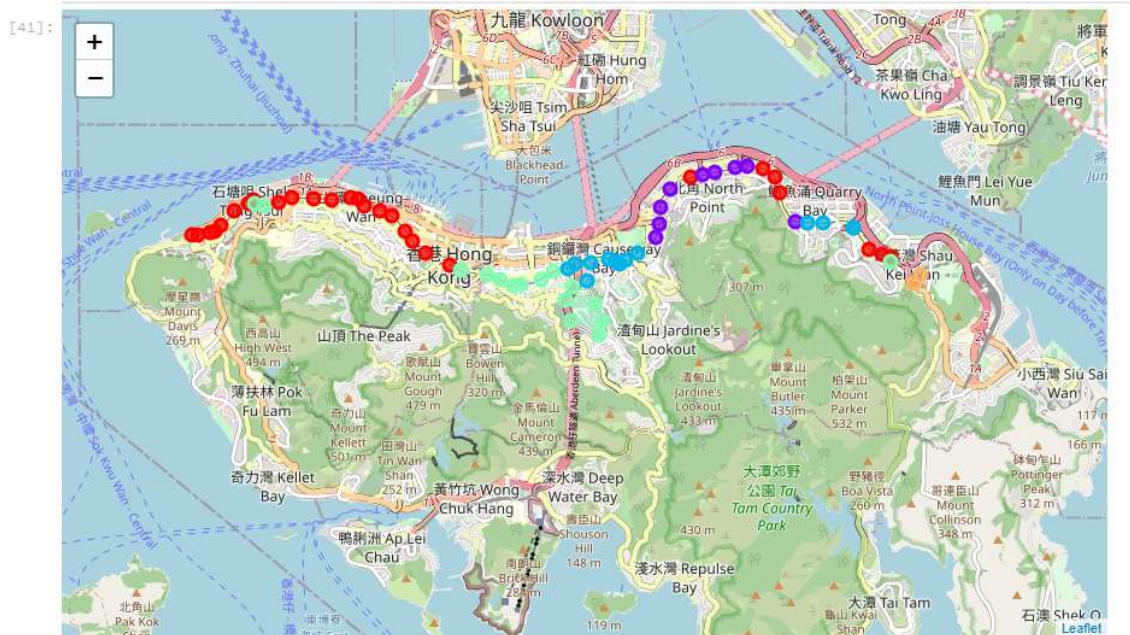
A leaflet map is also generated to visualize the clusters of tram stops.

```
[41]: # create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(hk_merged['lat'], hk_merged['long'], hk_merged['Station'], hk_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
#map_clusters.save('Clusters.html')
```



Notes: Tram stops are shown as coloured spots. The colours of different clusters are:-

- (i) Cluster 0 - Red
- (ii) Cluster 1 - Purple
- (iii) Cluster 2 - Blue
- (iv) Cluster 3 - Green
- (v) Cluster 4 - Orange

3.4 Analysis of the Clusters

Having grouped the tram stops into 5 clusters, I examine each cluster to find out its distinguishing features.

The stations of each clusters and their most common venues are grouped and put into data-frames as shown below:-

An extract of data-frame for Cluster 0

```
[42]: hk_merged.loc[hk_merged['Cluster Labels'] == 0, hk_merged.columns[[1] + list(range(5, hk_merged.shape[1]))]]
```

```
[42]:
```

	Station	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	Kennedy Town Terminus	0	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum	Laser Tag
1	Davis Street	0	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum	Laser Tag
2	Smithfield	0	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum	Laser Tag
3	Sands Street	0	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum	Laser Tag
4	Kennedy Town Praya	0	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum	Laser Tag
5	Queen's Road West	0	Art Gallery	Movie Theater	Dance Studio	Music Venue	Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
6	Whitty Street Depot	0	Art Gallery	Music Venue	Movie Theater	Rock Club	Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
7	Hill Road	0	Art Gallery	Music Venue	Movie Theater	Rock Club	Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
10	Water Street	0	Art Gallery	Music Venue	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio	History Museum
11	Western Street	0	Art Gallery	Street Art	Music Venue	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio
12	Eastern Street	0	Art Gallery	Dance Studio	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum	Laser Tag

An extract of data-frame for Cluster 1

```
[43]: hk_merged.loc[hk_merged['Cluster Labels'] == 1, hk_merged.columns[[1] + list(range(5, hk_merged.shape[1]))]]
```

```
[43]:
```

	Station	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
36	Lau Sin Street	1	Art Gallery	Dance Studio	Music Venue	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum
37	Lau Li Street	1	Dance Studio	Music Venue	Art Gallery	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum
38	Jupiter Street	1	Dance Studio	Music Venue	Art Gallery	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum
39	Fortress Hill	1	Dance Studio	Music Venue	Art Gallery	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	History Museum
41	North Point Terminus	1	Dance Studio	Museum	Opera House	Music Venue	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
42	Shu Kuk Street	1	Dance Studio	Museum	Opera House	Music Venue	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
43	Tin Chiu Street	1	Museum	Dance Studio	Opera House	Music Venue	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
44	Healthy Street West	1	Museum	Music Venue	Dance Studio	Art Gallery	Rugby Stadium	Mini Golf	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall
48	Mount Parker Road	1	Dance Studio	Museum	Public Art	Art Gallery	Rugby Stadium	Mini Golf	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall

An extract of data-frame for Cluster 2

```
[44]: hk_merged.loc[hk_merged['Cluster Labels'] == 2, hk_merged.columns[[1] + list(range(5, hk_merged.shape[1]))]]
```

	Station	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
30	Canal Road West	2	Multiplex	VR Cafe	Dance Studio	Laser Tag	Music Venue	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
31	Paterson Street	2	Multiplex	Dance Studio	VR Cafe	Laser Tag	Mini Golf	Art Gallery	Racecourse	Science Museum	Art Museum	Arts & Entertainment
32	Pennington Street	2	Multiplex	Dance Studio	VR Cafe	Laser Tag	Mini Golf	Art Gallery	Racecourse	Science Museum	Art Museum	Arts & Entertainment
33	Causeway Bay Terminus	2	Multiplex	Dance Studio	Art Gallery	Laser Tag	Mini Golf	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
34	Shelter Street	2	Multiplex	Art Gallery	Dance Studio	Laser Tag	Mini Golf	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
35	Victoria Park	2	Multiplex	Art Gallery	Dance Studio	Laser Tag	Mini Golf	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
49	Shipyards Lane	2	Museum	Multiplex	Art Gallery	Science Museum	Rugby Stadium	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
50	Kornhill	2	Multiplex	Museum	Dance Studio	History Museum	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Laser Tag
51	Tai Koo Shing Road	2	Multiplex	Museum	Rock Club	Dance Studio	History Museum	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
52	Tai Hong Street	2	Multiplex	Museum	Rock Club	Dance Studio	History Museum	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
59	Foo Ming Street	2	Multiplex	VR Cafe	Dance Studio	Laser Tag	Mini Golf	Music Venue	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall
60	Leighton Road	2	Multiplex	VR Cafe	Racecourse	Dance Studio	History Museum	Laser Tag	Music Venue	Movie Theater	Art Museum	Arts & Entertainment
67	Tin Lok Lane	2	VR Cafe	Multiplex	Dance Studio	History Museum	Music Venue	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club

An extract of data-frame for Cluster 3

```
[45]: hk_merged.loc[hk_merged['Cluster Labels'] == 3, hk_merged.columns[[1] + list(range(5, hk_merged.shape[1]))]]
```

	Station	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
8	Shek Tong Tsui Terminus	3	Art Gallery	Outdoor Sculpture	Music Venue	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio
9	Whitty Street	3	Art Gallery	Music Venue	Street Art	Outdoor Sculpture	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club
23	Admiralty MTR Station	3	Art Gallery	Multiplex	History Museum	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio
24	Arsenal Street	3	Art Gallery	Rock Club	Theater	Performing Arts Venue	Movie Theater	Music Venue	Multiplex	Laser Tag	Art Museum	Arts & Entertainment
25	Gresson Street	3	Art Gallery	Rock Club	Theater	Performing Arts Venue	Dance Studio	Movie Theater	Music Venue	Multiplex	Laser Tag	Art Museum
26	Swatow Street	3	Rock Club	Theater	Performing Arts Venue	Dance Studio	Art Gallery	History Museum	Music Venue	Movie Theater	Mini Golf	Art Museum
27	O'Brien Road	3	Rock Club	Theater	Art Gallery	Dance Studio	Music Venue	Performing Arts Venue	History Museum	Movie Theater	Mini Golf	Art Museum
28	Burrows Street	3	Dance Studio	History Museum	Art Gallery	Rock Club	Music Venue	Movie Theater	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall
29	Tonnochy Road	3	Music Venue	Art Gallery	Multiplex	History Museum	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio
56	Sun Shing Street	3	Art Gallery	Rock Club	History Museum	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio

An extract of data-frame for Cluster 4

```
[46]: hk_merged.loc[hk_merged['Cluster Labels'] == 4, hk_merged.columns[[1] + list(range(5, hk_merged.shape[1]))]]
```

```
[46]:
```

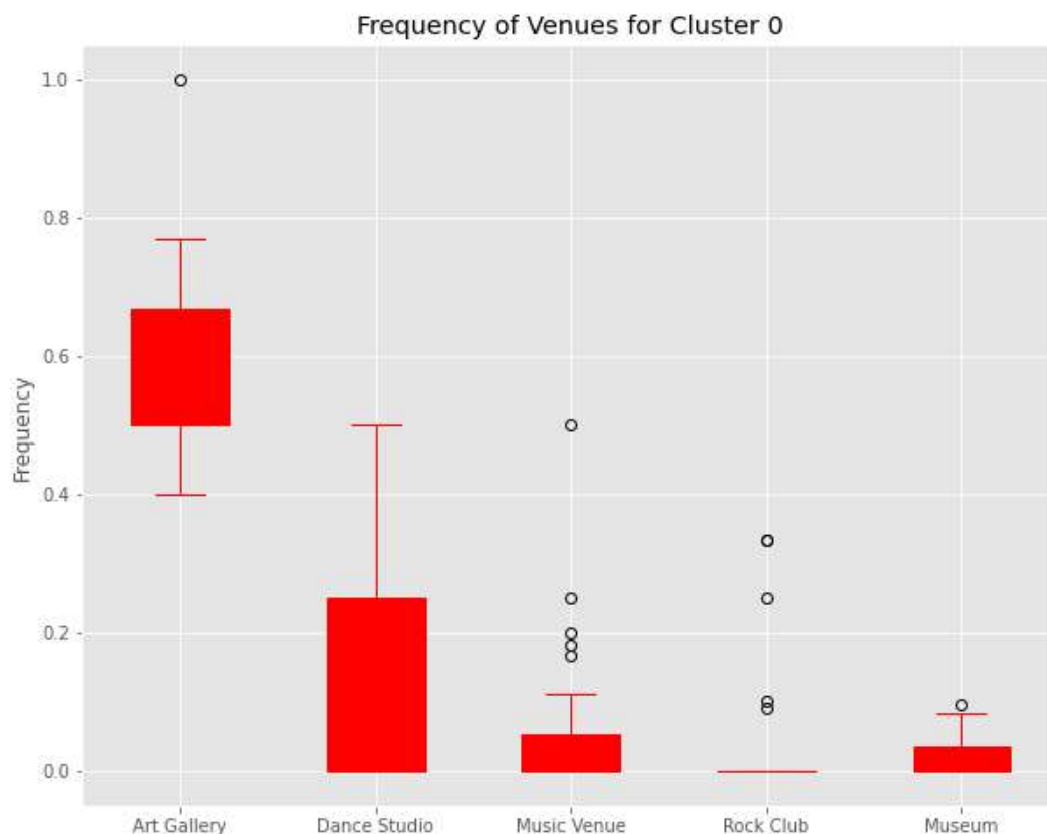
	Station	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
57	Chai Wan Road	4	History Museum	VR Cafe	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio	Laser Tag
58	Shau Kei Wan Terminus	4	History Museum	VR Cafe	Multiplex	Art Museum	Arts & Entertainment	Comedy Club	Concert Hall	Country Dance Club	Dance Studio	Laser Tag

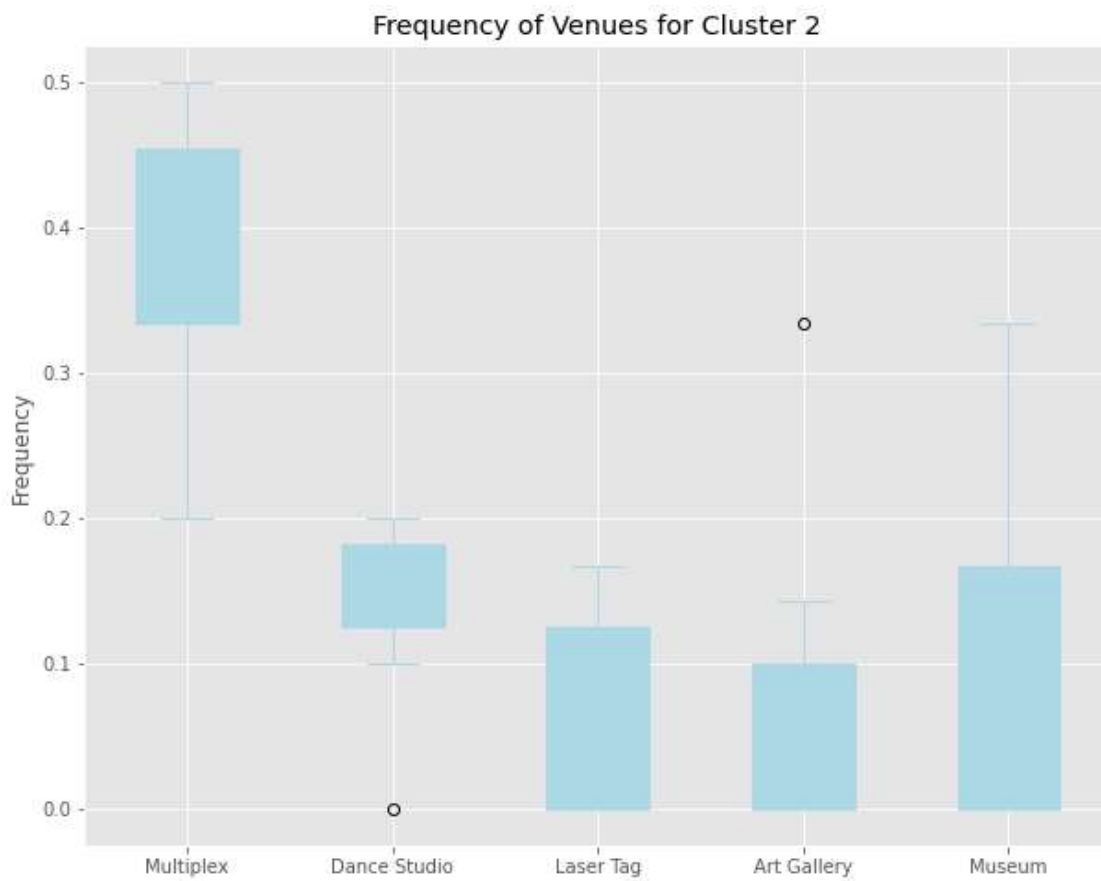
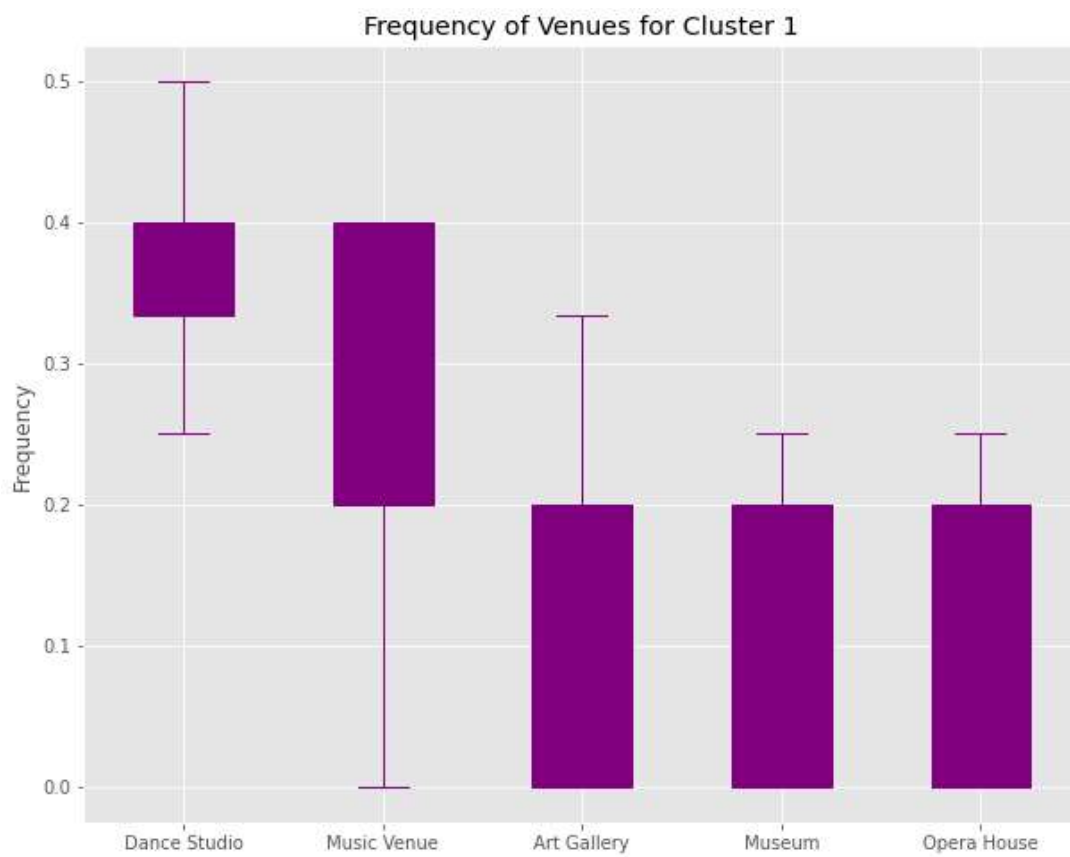
By examining the data-frames, it might gives a preliminary idea of the distinguishing venues for the clusters.

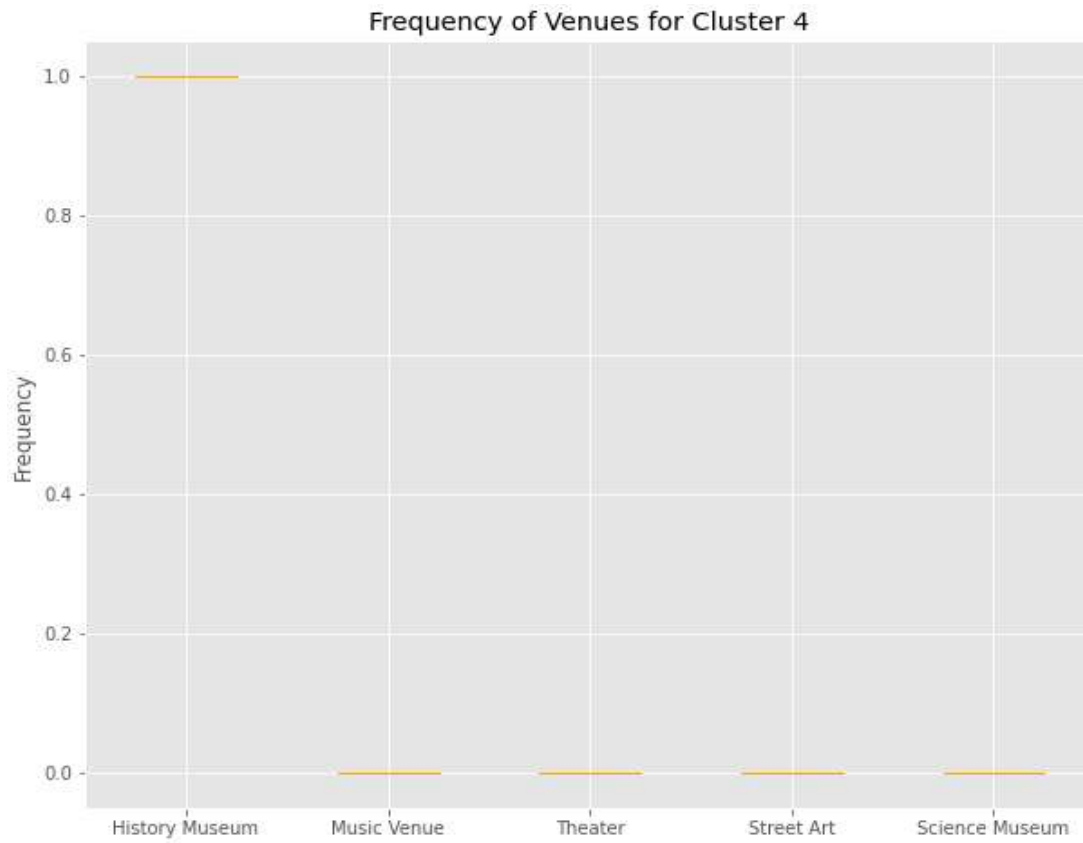
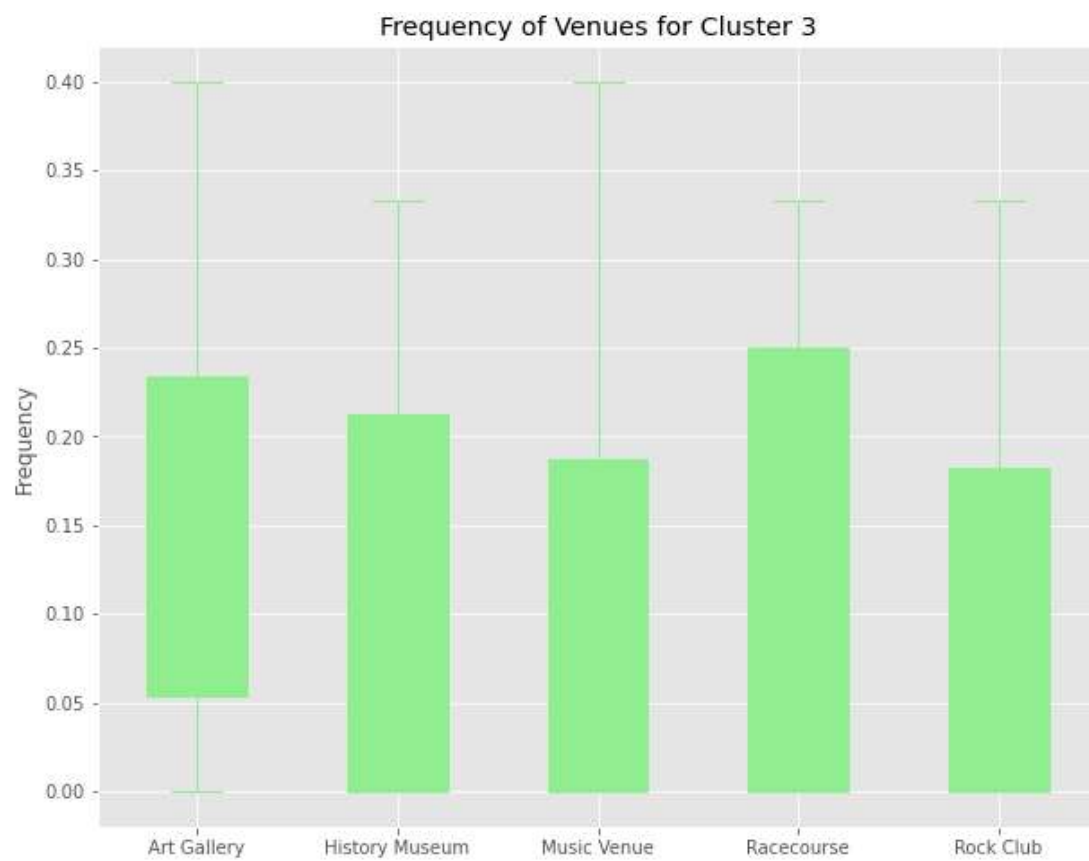
However, some clusters comprise tens of entries and has a varieties of common venues, such as Cluster 3 above, it is hard to tell the character of the cluster by plainly reading the rows of data in the data-frame.

Under the circumstances, I have visualize the five most common venues of each cluster in boxplots

The boxplots for the five clusters of Art & Entertainment venues are shown as below:-







4. Results and Discussions

Based on the above analysis, the following observations are noted.

- Cluster 0 (red) has a concentration of Art Galleries nearby.
- Cluster 1 (purple) is dominated by Dance Studio and Music Venues
- Cluster 2 (blue) has many multiplexes (i.e. cinemas) nearby.
- Cluster 3 (green) is a cluster of diverse venues where you can find different kinds of establishments including historical museum and racecourse.
- Cluster 4 (orange) does not have many Art and Entertainment venues around.

In planning Arts and Entertainments itineraries along the Hong Kong Tramway, the above results would be very useful in designing the routes.

For example, if the theme of the tour is for art appreciation, one might want to go for Art Galleries and design several stops in the red spots throughout a tram journey from east to west (or vice versa).

Alternatively, if one pursues active experiences during the tour, he can stop at the purple spots for dance or music sessions.

For film lovers, the blue spots with clusters of multiplexes are the places to go.

And if one would like to do a variety of activities in a tight schedule, he might go for the green spots to enjoy a bit of everything.

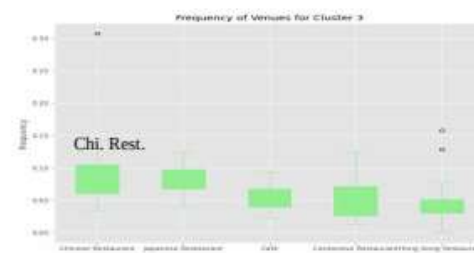
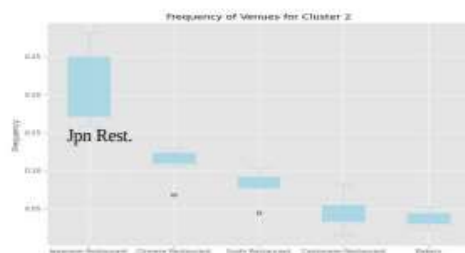
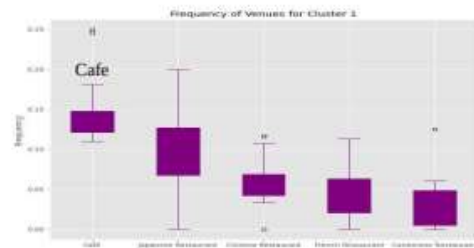
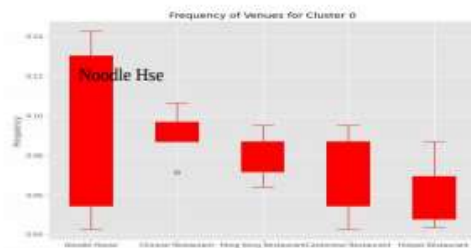
Meanwhile, while there are not many venues Art & Entertainment venues nearby the orange stops, one might stay on board to enjoy the ride during his/her tram travel.

Certainly, one might also design a blended journey with a variety of Art & Entertainment activities during the tram travel. The above clustering would provide a very useful reference in deciding where to do what throughout the trip.

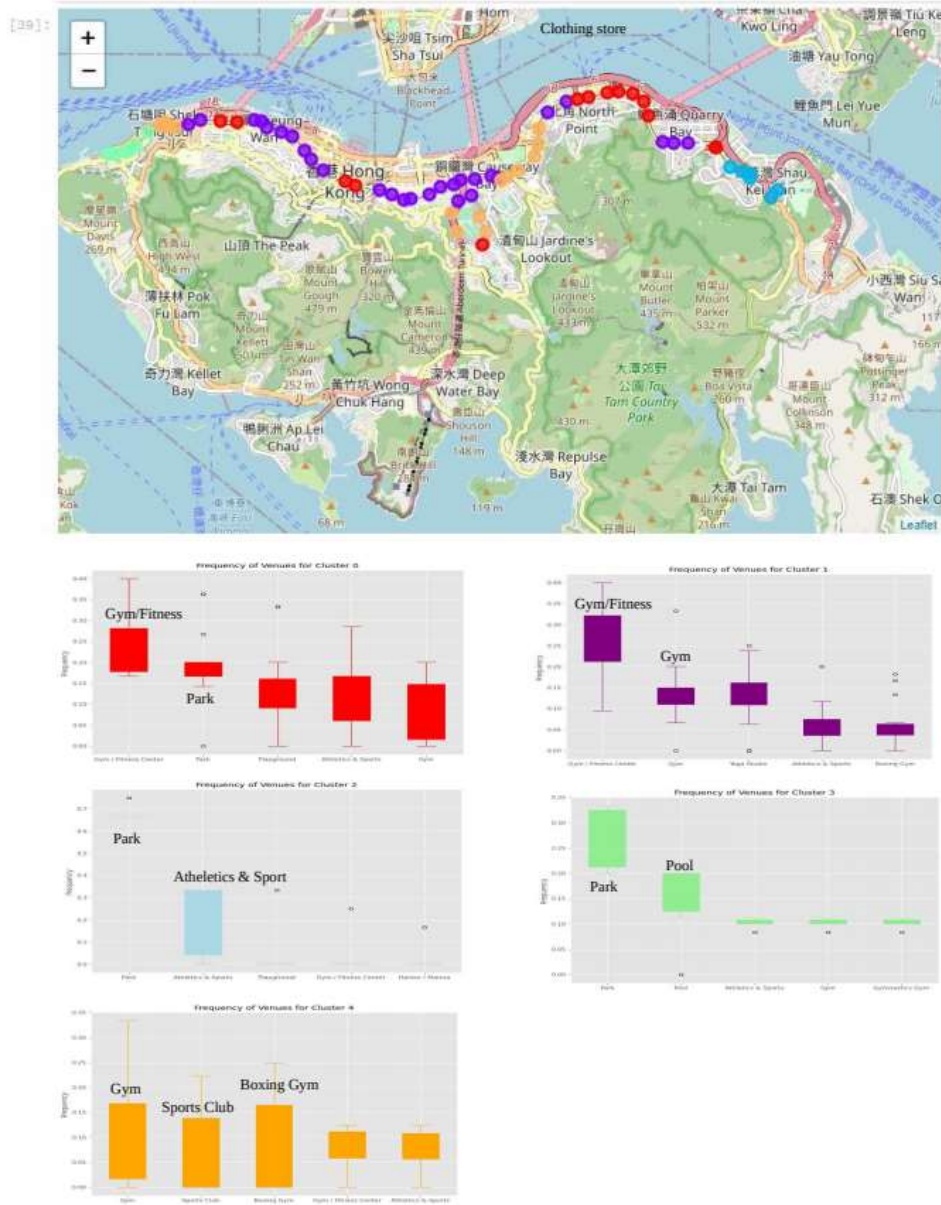
In the course of my capstone project, I have repeated the above process for other categories of data available in the Foursquare API namely Food venue, Outdoors & Recreation venue as well as Shop & Services venues. The results are shown as follows for your reference.

I believe the data analysis for these other categories of data are also very useful for itinerary planning of other theme tours along the Hong Kong Tramways as illustrated above. One might also consider a comprehensive tour comprising different elements along the journey.

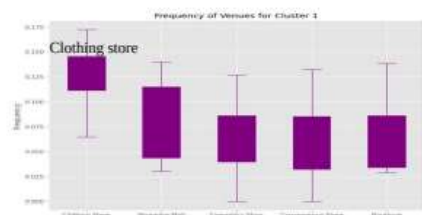
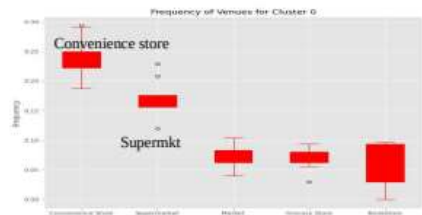
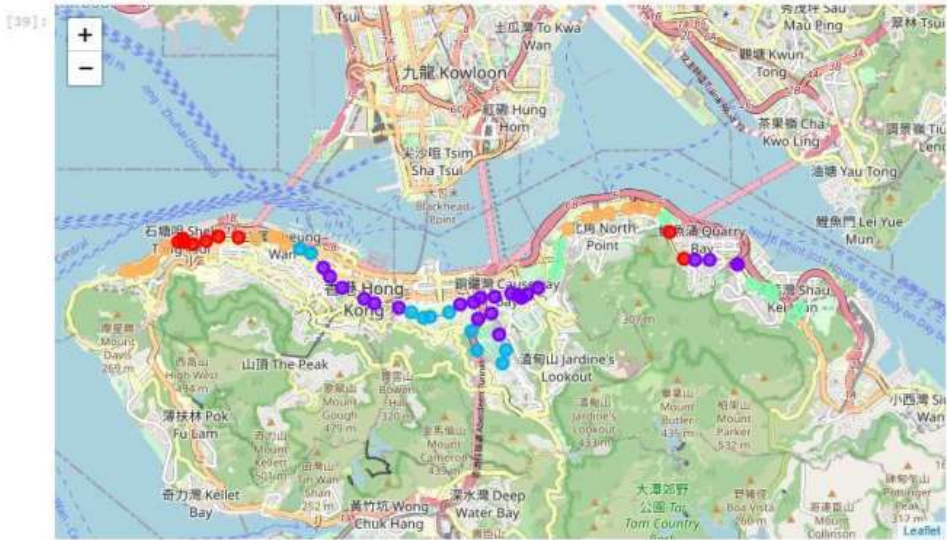
Clustering of Tram Stations for Food Venues



Clustering of Tram Stations for Outdoor & Recreation Venues



Clustering of Tram Stations for Shop & Services Venues



5. Discussions

In the course of my analysis, I found the highest number of venues in the Foursquare data along the Hong Kong Tramways are eating places. At the beginning of my project, I have adopted the high-level venues (i.e. venues comprising all sub-categories) in the API, as a result, the most common venues for all tram stops are eateries, I am therefore, unable to categorize the tram stops by their proximity to different kind of venues.

Under the circumstance, I resorted to the use of sub-category data instead of the high level data. These sub-category data allow me to perform meaningful clustering and analysis as illustrated above.

Besides, as the above analysis are based on the occurrence of most common venues close to each tram stops, the results might provide a good reference for the geographical distribution of the venues. However, characters of the venues such as capacity, popularity or special features have not yet been accounted. In drawing up actual itineraries for travel, it is believed that these other information are also worth to be considered.

6. Conclusion

In this project, I've a good glimpse of how data-science project is carried out using real life data in my neighbourhood. I assembly the primitive data from difference sources including the Hong Kong Tramways website, Google Map and Foursquare API etc., then use python libraries to scrap the data, perform analyses and visualized the results on Folium maps and boxplots.

Uses and limitations of the project are discussed above. It reveals that Foursquare API is a useful sources of information to explore the establishments in an area and their geographical distribution. However, in making actual uses of the information, such as planning travel itinerary as illustrated by my project above, more aspects should be considered apart from frequency or occurrence or geographical distribution, such as venue size, rating or venue features. This might require a combination of data from the Foursquare API as well as other sources to give comprehensive insights of the problem.