

EXPLORE ALONG HONG KONG TRAMWAYS

Since 1904



Applied Data Science Capstone Project

2020

BACKGROUND

- Since 1904
- Running east to west of Hong Kong Island
- Today, its a nostalgic mode of public transport to explore the city



ANALYSIS AND ITS APPLICATION

- In this project, we'll explore the neighborhoods along the tramways
- Identify and classify the venues near tram stops
- Planning tram tours of different themes, such as:-
 - Arts & Entertainment Tour
 - Foodie Tour
 - Outdoor & Recreation Tour
 - Shopping Tour





DATA ACQUISITION & PREPARATION

- Data of tram stops are scraped from www.hktramways.com
- Coordinates of tram stops are scraped from www.google.com.hk
- Venue data are obtained from Foursquare API which are classified in 4 sub-categories
- Hundreds of venue data are obtained and grouped with respect to the 67 stops
- The most common kind of venues are sorted out



USING FOLIUM LIBRARY CREATE A LEAFLET MAP

to visualize the tram stops along Hong Kong Island

[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	Perf Art
0	Admiralty MTR Station	0.400000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.200000	0.200000	0.000000	0.000000	0.00	0.000000	0	
1	Arsenal Street	0.200000	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	
2	Bank Street	0.560000	0.000000	0.040000	0.000000	0.040000	0.000000	0.120000	0.040000	0.000000	0.000000	0.040000	0.040000	0.040000	0.00	0.000000	0	
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	0.00	0.000000	0	
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	0.142857	0.00	0.000000	0	
5	Canal Road West	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.166667	0.000000	0.166667	0.000000	0.333333	0.000000	0.166667	0.00	0.000000	0	
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	0.00	0.000000	0	
7	Chai Wan Road	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000	0.500000	0.000000	0.000000	0.00	0.000000	0	
8	Connaught Road West	0.600000	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	
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	0.000000	0.00	0.000000	0	
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	0.00	0.000000	0	

CLUSTERING OF ART & ENTERTAINMENT VENUES

- There are 591 venues in 26 categories
- Venues are grouped by their neighborhoods
- Apply one-hot encoding to each categories, taking mean of frequency
- Put the data into a data-frame for clustering

```
[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)
```

CLUSTERING OF ART & ENTERTAINMENT VENUES

- Apply k-means clustering
- Constellate the trams stations into 5 clusters

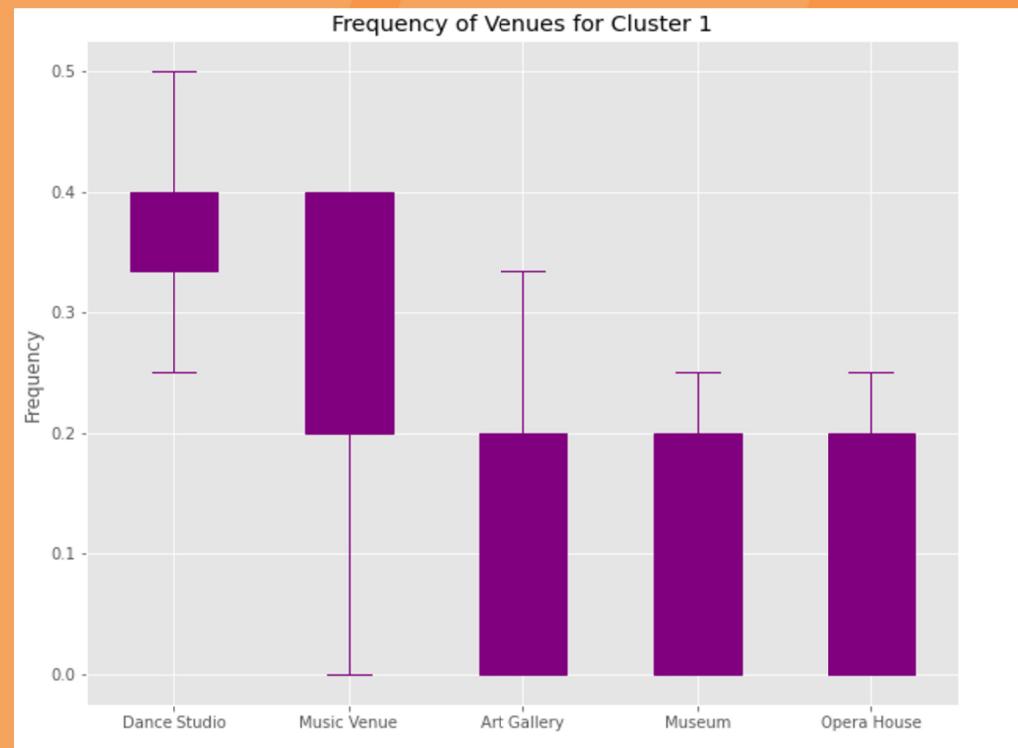
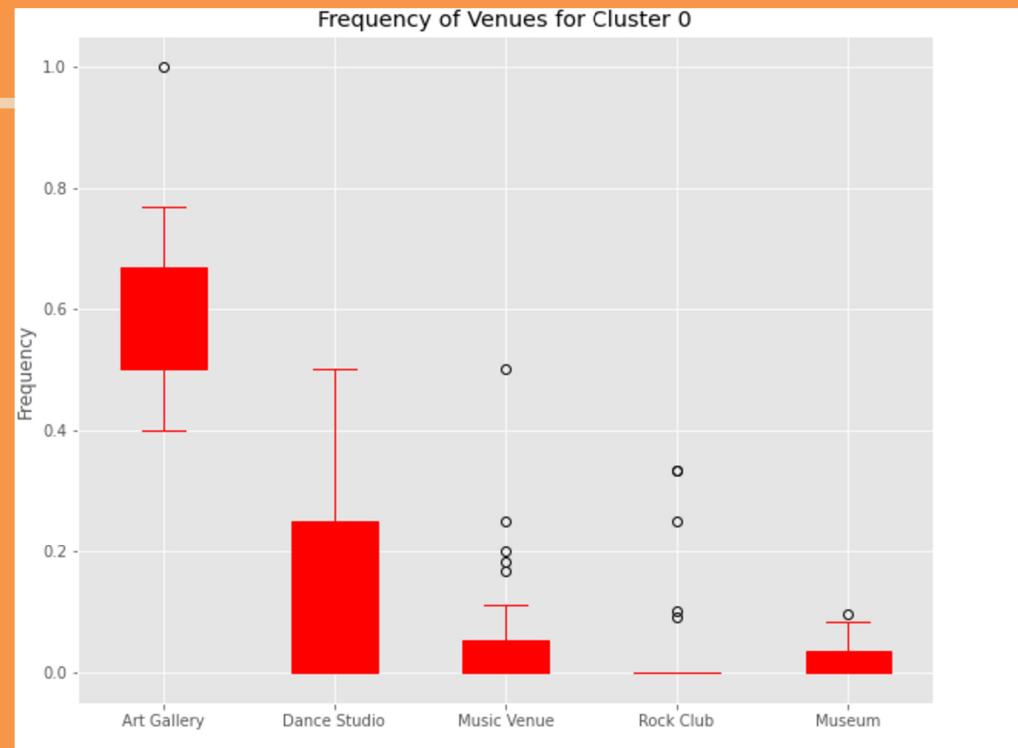


MAP SHOWING THE CLUSTERING OF ART & ENTERTAINMENT VENUES

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4

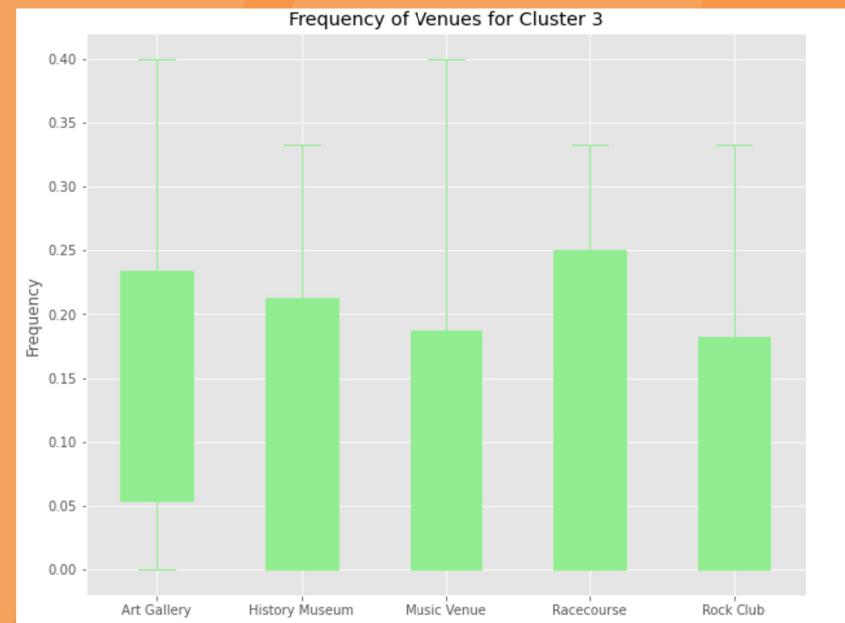
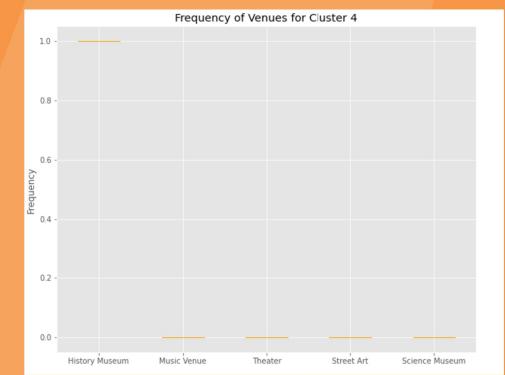
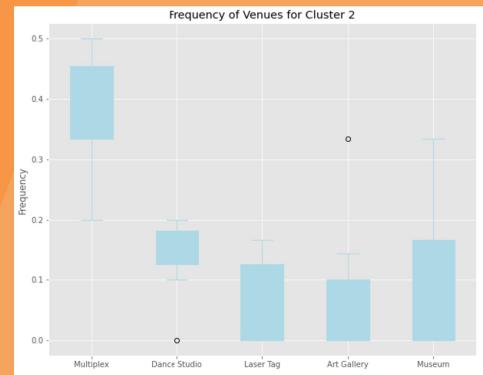
ANALYSIS OF THE CLUSTERS

- Cluster 0 is dominated by Art Galleries
- Cluster 1 has most Dance Studios and Music Venues



ANALYSIS OF THE CLUSTERS

- Cluster 2 is a Multiplex (i.e. cinema) bunch
- Cluster 3 is featured by History Museum & Racecourse
- Cluster 4 has not much to do



CONCLUSION

- Built useful models to cluster venues along the tramways
- Inclusion of other information such venue capacity, features, ratings can give more comprehensive insight in itinerary planning
- Applicable for other categories of venues such as eateries, shops or recreational facilities.

