

The background is a high-angle, wide-area photograph of the Shanghai skyline, showing a dense cluster of skyscrapers and buildings. A large, thick yellow letter 'A' is superimposed over the center of the image. Inside the upper part of the 'A', there are two smaller, thin yellow squares. Several thin yellow lines radiate from the top and sides of the 'A' towards the corners of the frame. At the bottom left and bottom right corners, there are small yellow triangles pointing inwards towards the center.

Metro Stations of Shanghai

Introduction

- Shanghai, the economic center of China, millions of people come in and out in one year. For new-comers, it is essential to look for a suitable home and their basic public transportation is metro train and it is also a crucial factor when looking for a new home. So this project aims to accurately point out every metro station and their surroundings.
- The target customer: people who search for a new home or hotel, and want to get well known about one area.

Data

- The list of all the metro stations we used were retrieved from the Wikipedia page:

	Number	Station_Name	Opened	Location	Platform_Level	Platform_Type	Transfers
0	1	莘庄	1996年12月28日	Minhang	At-grade	Side platform	5
1	1	外环路	1996年12月28日	Minhang	At-grade	Side platform	NaN
2	1	莲花路	1996年12月28日	Minhang	At-grade	Side platform	NaN
3	1	锦江乐园	1996年12月28日	Xuhui	At-grade	Side platform	NaN
4	1	上海南站	2004年10月30日	Xuhui	Underground	Island platform	3

Data

- To coordinate the latitude and longitude to each row, by looping through the whole list and creating custom Baidu API queries for each row from their cell values:

	Number	Station_Name	Opened	Location	Platform_Level	Platform_Type	Transfers	Longitude	Latitude
0	1	莘庄	1996年12月28日	Minhang	At-grade	Side platform	5	121.392186	31.116872
1	1	外环路	1996年12月28日	Minhang	At-grade	Side platform	NaN	121.399614	31.126649
2	1	莲花路	1996年12月28日	Minhang	At-grade	Side platform	NaN	121.409334	31.136734
3	1	锦江乐园	1996年12月28日	Xuhui	At-grade	Side platform	NaN	121.415479	31.145542
4	1	上海南站	2004年10月30日	Xuhui	Underground	Island platform	3	121.435865	31.159439

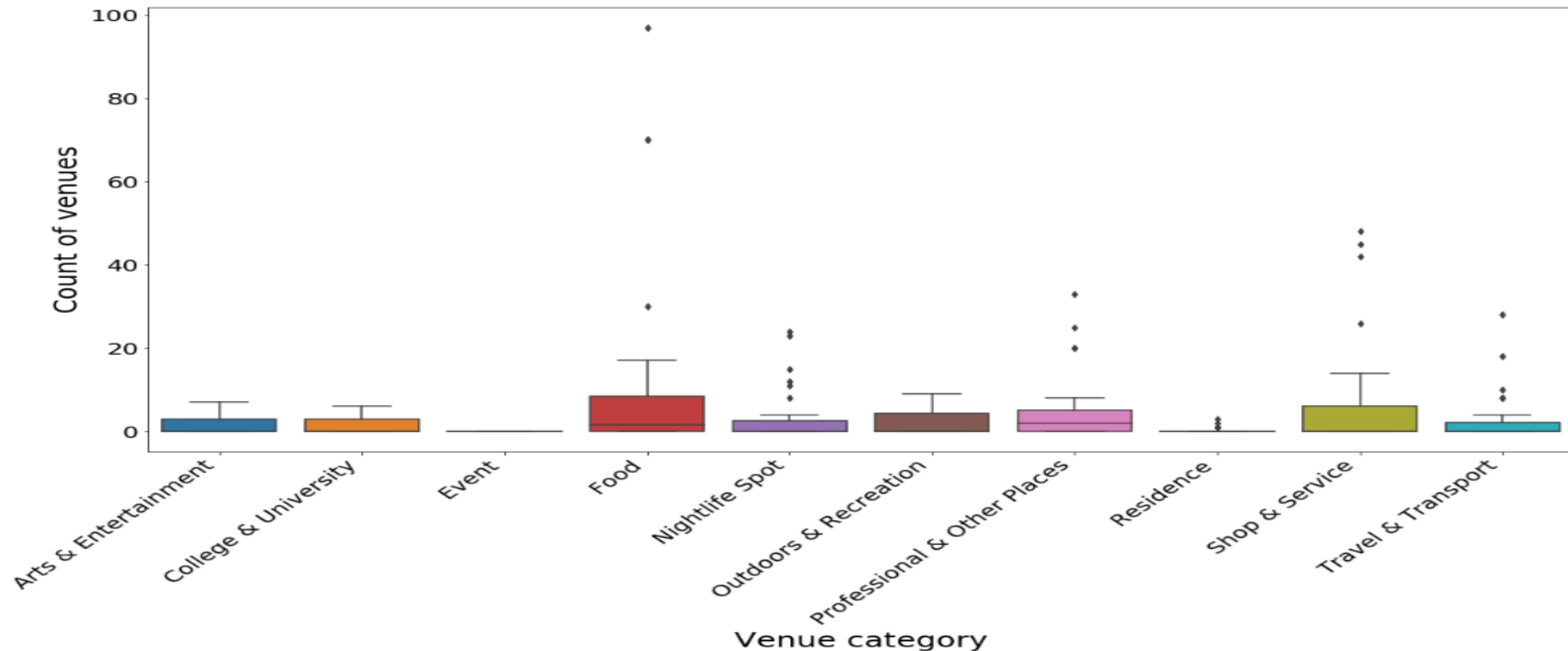
Data

- Through Foursquare Places API, we query each station's location and collect the number of venues, sorted by the top-level categories available from the Foursquare API:

	Number	Station_Name	Opened	Location	Platform_Level	Platform_Type	Transfers	Longitude	Latitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot
0	1	莘庄	1996年 12月28日	Minhang	At-grade	Side platform	5	121.392186	31.116872	0	0	0	1	0
1	1	外环路	1996年 12月28日	Minhang	At-grade	Side platform	NaN	121.399614	31.126649	0	0	0	2	1
2	1	莲花路	1996年 12月28日	Minhang	At-grade	Side platform	NaN	121.409334	31.136734	0	0	0	6	2
3	1	锦江乐园	1996年 12月28日	Xuhui	At-grade	Side platform	NaN	121.415479	31.145542	0	0	0	4	1
4	1	上海南站	2004年 10月30日	Xuhui	Underground	Island platform	3	121.435865	31.159439	0	1	0	3	0

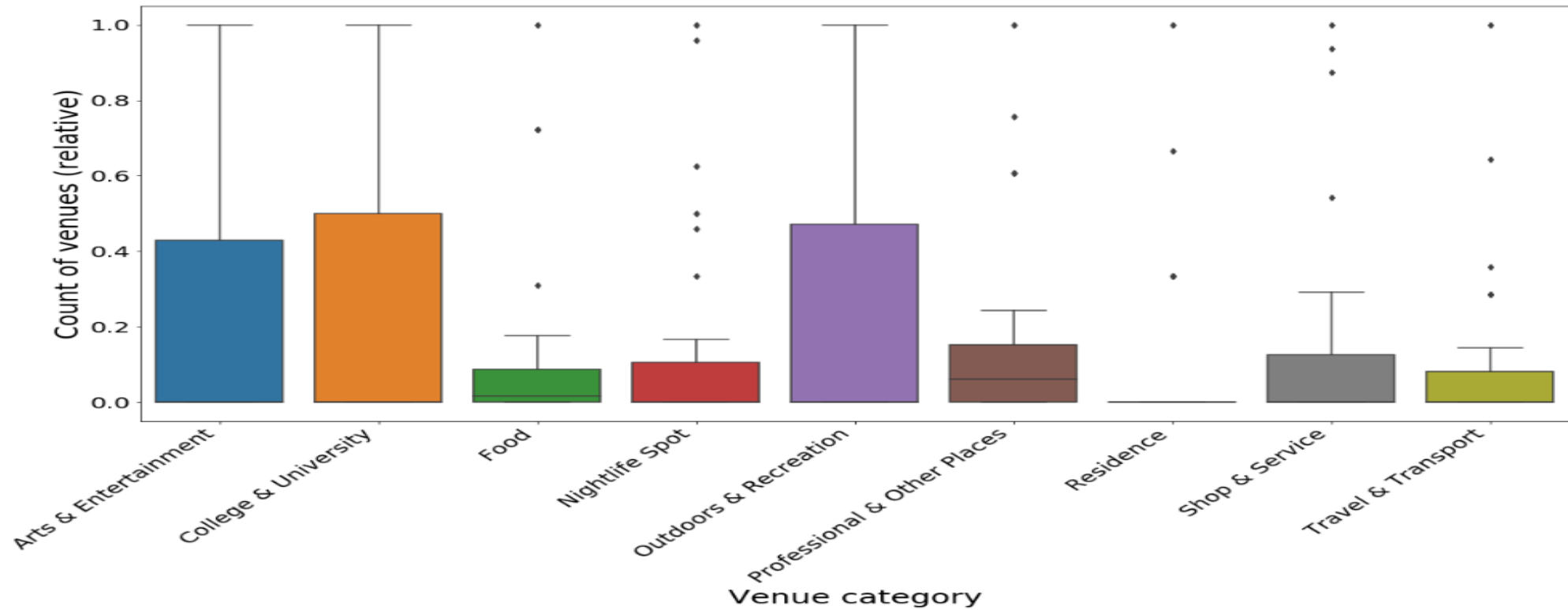
Analysis

- Ten categories of venue: Arts & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoors & Recreation, Professional & Other Places, Residence, Shop & Service, Travel & Transport.
- Using box plot to vividly display the distribution of different venue category:



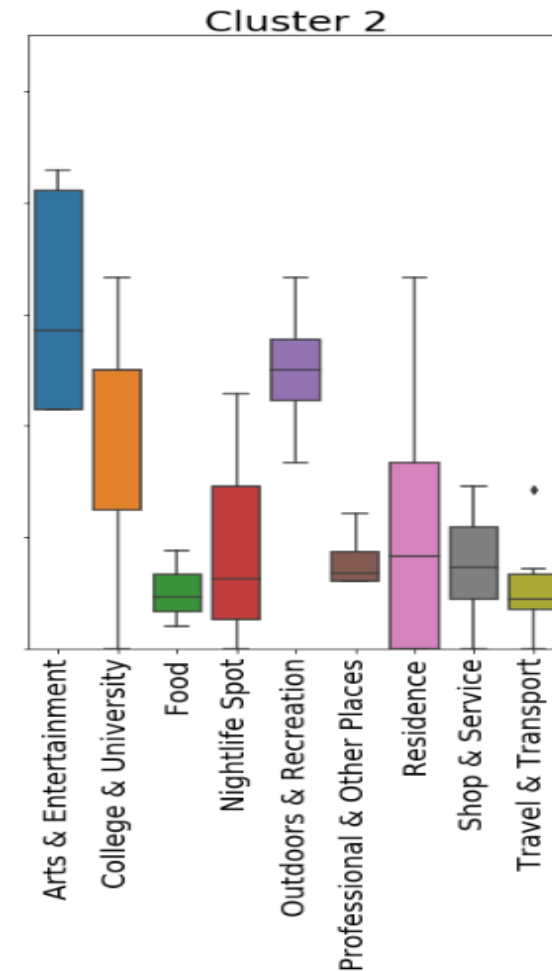
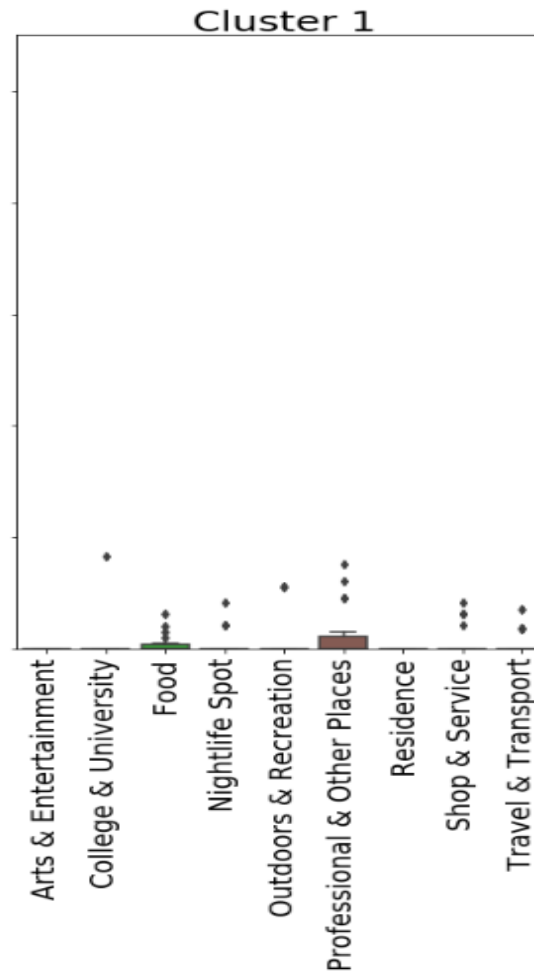
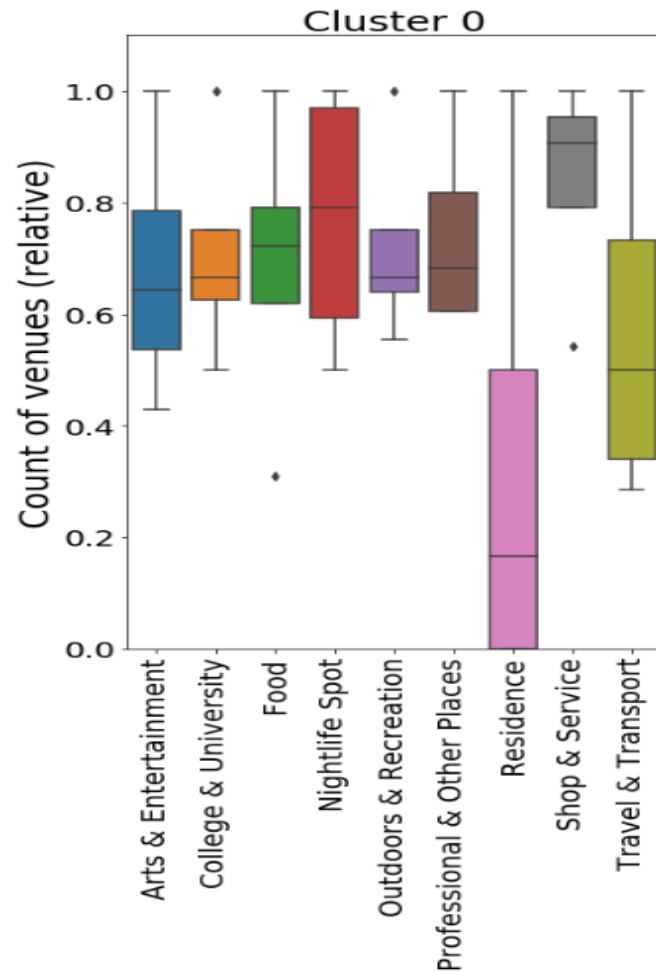
Analysis

- The most frequent venue categories: Shop&Service, Professional&OtherPlaces, Travel&Transport, and Food.
- The least frequent venue categories: Event
- Discard event from both the data frame and the list of categories and normalize the data using MinMaxScaler (scale from 0 to 1):



Methodology

- The classifier of project will use K-Means Clustering. For the scope of this class project, we will use 3 clusters in our analysis:



Results

- Summarize each of our classified clusters:
 - Cluster 0 has the highest number of venues nearby, especially for Shop&Service, Nightlife Spot, and Travel&Transport;
 - Cluster 1 on average has the least number of venues near its stations, and appears as the lowest density area;
 - Cluster 2 has the lowest number of Residence venues, and is between the other 2 clusters in nearby venue density.
-
- From map of the Shanghai line 1 Metropolitan Region:
 - Cluster 0 most likely has the highest number of people passing by and creating venues and check-ins, as they are in the densely populated areas of the city (offices and department stores i.e. Xujiahui);
 - Cluster 1 marks stations that are not in areas as developed as in the other two clusters;
 - Cluster 2 seems to mark stations where there are populated by different universities.

Discussion

- First, using the Foursquare database to get the number of venues around each station can make our results a bit biased towards the Food and Travel&Transport categories, as these 2 types of locations are the most commented and checked-into places. The significance of a location or building also is not shown, so some key landmarks or important areas might not be highlighted.
- However, with the main theme being density and having users being able to click and reveal the top 3 categories of each rendered location on the map, we were able to answer the questions and challenges asked in the beginning of the project. Users can use the interactive map of Shanghai's Metro to find out more about the surroundings of each station by clicking on a circled area, and a popup will inform them about the top 3 types of venues around the station. By familiarizing themselves with the color scheme of the map (blue for high-density, white for medium-density, red for sparsely-dense areas), users can view the overall status of the Shanghai Metropolitan Region.

Conclusion

- We have shown how to use the Baidu API, the Foursquare Places API, and the Python Folium library to retrieve the locations and nearby number of venues around each of Shanghai's line1 metro stations, and plot most of them onto an interactive map of the Shanghai Metropolitan Region. The data collected can be useful to others in the future in other areas of research, especially if combined with more data from other sources, such as social media feeds or census data.

An aerial night photograph of a city, showing a grid of streets and illuminated buildings. Overlaid on the image is a large, bright yellow triangle. Inside this triangle, the word "THANKS" is written in a bold, yellow, sans-serif font. Several thin yellow lines and small yellow triangles are scattered around the main triangle, creating a dynamic, geometric design.

THANKS