**Introduction**

Chicago is an international city that takes important places in both economy and politics in Asian area. For personal interests, as I’ve spent four there, I wanna digging into the food culture of the neighborhood there. As there is no official division of neighborhoods and the public transportation, CTA services, especially the metra, connects people in the city to their daily activities spot, we will simply use subway stations to represent the neighborhood. In this research, we will look into the restaurant categories 1.5 miles along the subway and use different food types as attributes to build up a unsupervised machine learning classifier model to classify the stations into 3 clusters and study how neighborhood differs from each other and make recommendations on what types of restaurant are specifically popular for clusters.

**Data**

The original list of Chicago CTA Stations is downloaded from CTA website [1]. The venues around stations are retrieved through FourSquare Places API [2]. The data is then merged, cleaned and preprocessed for explicability, understandability and meeting business requirements purpose.

**Methodology**

The method used here are descriptive statistics and the algorithm used for the unsupervised classifier is KMeans.

1. The CTA Station Data

After the combination of list of CTA Stations and use station names to look up the geographic data, we get a table listing the CTA Station name, geographic coordinates and hierarchical division of the station.

An interactive visualization of the data (see in github) with a Folium map is like:

![A close up of a map

Description automatically generated]()

1. A visualization of Food Venues’ location is:
2. The variance ’Venue Category‘ is then processed with one hot encoding to a dummy matrix with each unique value of category into a single attribute. The sum occurrence of each category within an CTA Station neighborhood is calculated.

The top 8 most frequent venues are then showed, for instance:

A screenshot of a cell phone

Description automatically generated

1. KMeans

K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. In this case, we want to cluster CTA Stations together and find inner similarities of restaurant distributions.

In order to look for the bust K clusters, we used the yellowbrick KElbowVisualizer package:

A close up of a map

Description automatically generated

The KMeans is then performed and with kcluster of 3.

**Results**

By clustering CTA Stations into 3, we find the distribution of the 3 clusters like below with the blue spot representing cluster 0, red dot cluster 1 and green dot cluster 2.

A view of top5 venue categories under a cluster is:

![A close up of a map

Description automatically generated]()

A picture containing screenshot

Description automatically generated

**Discussion**

From the bar chart above, we can see that the taste of each cluster varies.

**cluster0: Very much coffee and park, very much fast food, much American food**

**cluster1: Very American Food, some fast food, dumpling and bakery**

**cluster2: Very much coffee, very much American Food, some snack, vegan and sushi**

**Conclusion**

This study looks into how food business differs around Chicago CTA Station neighborhoods. The results provide 3 categories of food selection, or rather lifestyle, among CTA Stations all around Chicago. Hopefully, this short study can provide some insights into where to find food and business opportunities around Chicago and entertains you a little bit by providing some cute visualizations.

**References**

[1] <https://data.cityofchicago.org/Transportation/CTA-System-Information-List-of-L-Stops/8pix-ypme>

[2] <https://foursquare.com/developers/apps>