Fare is Fair: Creating Chicago transit zones for 'L' travel

Nathaniel Smith

Intent

The City of Chicago currently has a flat rate of $2.50 for entry into the Elevated Rail ("L") transit system. Other metropolitan transit systems such as Transport for London's Tube network have fare zones where, depending on your entry and exit into the system, a fare may be higher or lower. While the "L" infrastructure is currently incapable of tracking when passengers exit the system, is it still possible for Chicago to benefit from zone-based fares?

Using data on station popularity, community socioeconomic hardship, and taxi pick-up and drop-off, fare zones can be created for the "L" network that encourage transit in burdened areas of the network while offsetting this fare deficit by charging riders in more stressed parts of the network a higher fare.

Goals

Would a zone based system for the 'L' benefit Chicago?

To answer this question, three goals are presented:

1. *How are the zones determined?* Use k-means to create varying zones, taking into account station popularity, community hardship, and area taxi demand. Each k-means cluster represents a zone, with all stations residing in a zone.
2. *What would the zones look like?* Web-based visualizations are created to demonstrate a CTA map with the new zones.
3. *Are the metrics used significant?* Web-based visualizations of the metrics are created to demonstrate their relation to each other. Correlation tests on metrics are also performed.

Data Utilized

The following data sets were retrieved to perform transformations:

1. 'L' Station Ridership - Station Entry Totals (2012-2018)
2. CTA 'L' (Rail) Lines
3. Chicago Neighborhood Socioeconomic Indicators (US Census - 2010)
4. Chicago Taxi Trip Data (2018, Limit 3.5M entries grabbed.)
5. 'L' Station Location Data
6. Census Tract Boundaries Location Data

Analysis

1. Perform k-means clustering on station popularity, neighborhood hardship index, and area taxi ridership to assign stations a zone. All data scaled.
2. Visualize 'L' map for varying amount of clusters/zones.
3. Perform calculations on correlation between metrics and visualize the k-means clusters and station metrics.

Results

K-Means Analysis:

Stations with their associated metrics are analyzed using the k-means algorithm with varying number of clusters.

Metrics:

1. Station Popularity
2. Socioeconomic Hardship Index of Community Area
3. Taxi Demand in Census Community Area

Note: All metrics are scaled prior to k-means analysis.

Figure 1: 8 clusters/zones

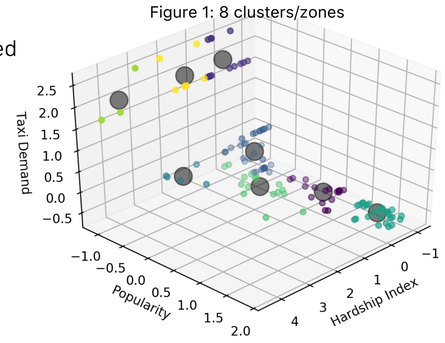


Figure 2: 6 clusters/zones

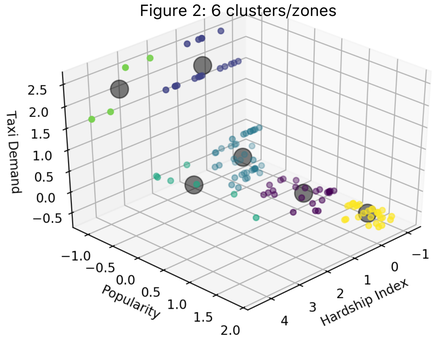
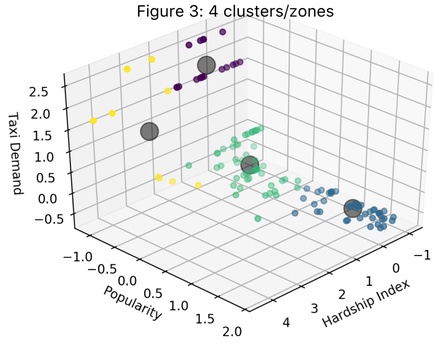


Figure 3: 4 clusters/zones



Correlation Coefficient Data:

Popularity and Hardship Data: -0.5074 | Popularity and Taxi Demand Data: 0.5199 | Hardship and Taxi Demand Data: -0.6134

Visualization & Conclusion

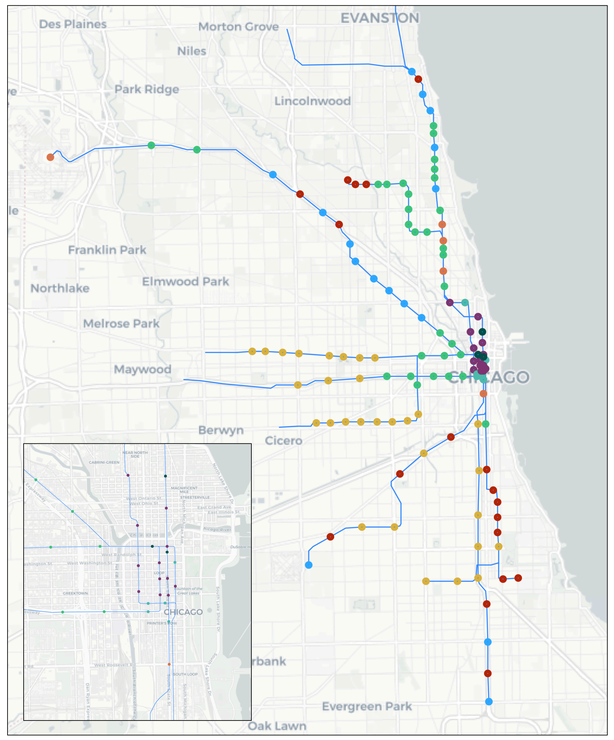


Figure 4: Generated "L" map network from k-means clustering. K-means (k=8) clustering performed on all stations in Chicago proper with three metrics.

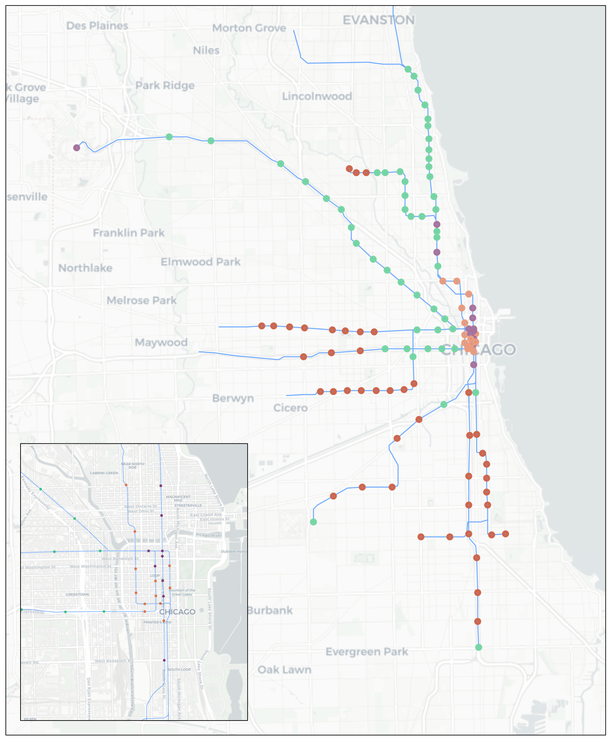


Figure 5: Alternate Generated "L" map network from k-means clustering. K-means (k=4) clustering performed on all stations in Chicago proper with three metrics.

Conclusion:

Utilizing k-means analysis on station metrics does prove to be useful for generating theoretical "fare zones" for the CTA 'L' network. In all visualizations, a strong disparity between stations in the downtown ("Loop") area and the more distant neighborhoods is shown.

Basing fares off of these zones would promote transit in disadvantaged areas while possibly increasing revenue due to a likely higher fare in the Loop.

Stations in the South of Chicago are in their own zone compared to stations in the North due to the severe socioeconomic disparities in the city. As such, Southern stations should see a lower fare than Northern stations.

Overall, the CTA could see benefits from a zone-based approach to fares.