CSP-571 : Project Presentation

Exploring Performance and Efficiency in the CTA System: A Data Driven Analysis of Routes and Stations

Presented by:

Ranjitha Aswath (A20526404)

Sanketkumar Patel (A20523237)

Krishna Chatrathi (A20520106)

Presented to:

Prof. Jawahar Panchal

Executive Summary

- This project aimed to evaluate the effectiveness of the Chicago Transit Authority (CTA) system through a comprehensive analysis of various operational aspects. The primary objectives were to identify the busiest routes and stations, monitor daily changes in L-line ridership, detect service delays and disruptions, evaluate neighborhood performance, and compare travel wait times with scheduled service times. By combining historical trends and advanced analyses, this research provides valuable insights that contribute to the overall efficiency and effectiveness of the CTA system, ultimately enhancing public transportation experiences for Chicago residents and visitors

Key Findings

- High Ridership in key Locations
- Impact of Covid19
- Seasonal Variation

Recommendations

- Post Covid Recovery strategies
- Infrastructure and service optimization
- Real-Time Data Integration
- Bus Route Analysis

Methodology

- Proposed methodology focused on evaluating CTA system effectiveness.
- Data gathering from CTA website and Chicago Data Portal.
- Real-time data access through Transit App API (future implementation).
- Data preprocessing for consistency, handling missing values, and outlier detection.
- Geospatial analysis of L-stations with emphasis on a specific rail line (e.g., red line).
- Cleaning process included filtering, merging, and adjusting data types.
- Standardized approach ensures reliable and insightful analysis.

Changes to Plan

- Due to unavailability and no reliable data readily we have opted out of analysis of bus routes and concentrated on “L” lines

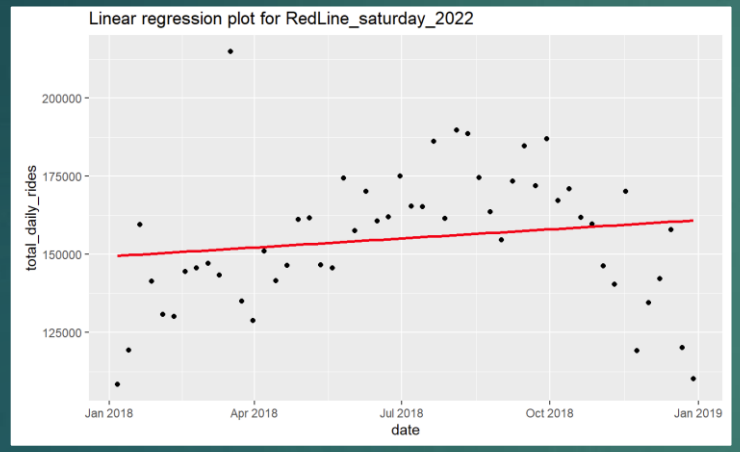
Data Gathering

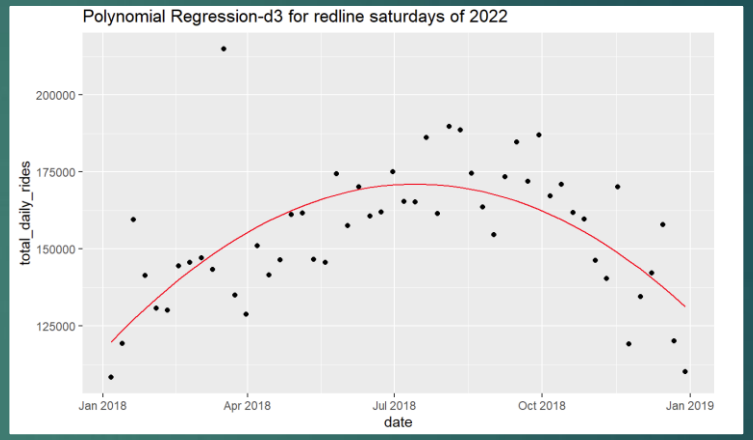
- We obtained data from CTA website
- We used "CTA_-System_Information-List_ofL_Stops.csv" to get station IDs of desired routes
- We used "CTA_-_Ridership_-__L__Station_Entries_-_Daily_Totals.csv" to find out daily totals of RED line rider ship
- We use "CTARailLines.shp"to obtain coordinates of CTA lines and CTA stops
- We used "CTA_-Ridership-_LStation_Entries-Monthly_Day-Type_Averages__Totals.csv"

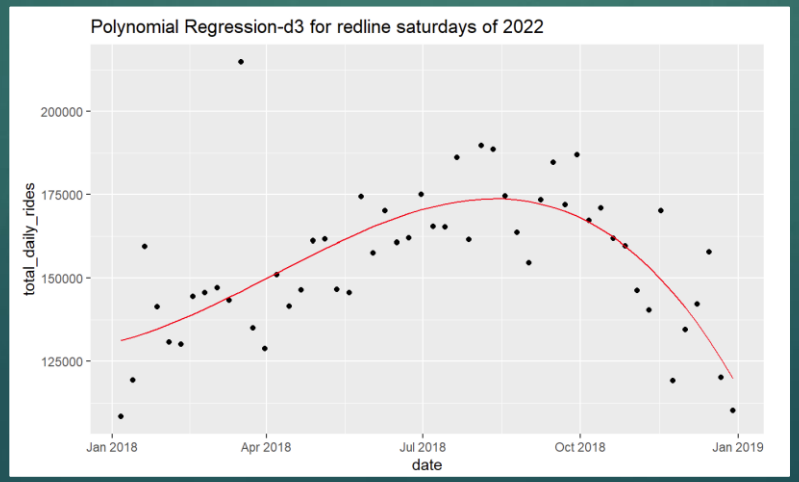
Weekday average Ridership

- Exploratoray Data Analysis
  - Load the "CTA_-_Ridership_-__L__Station_Entries_-_Daily_Totals.csv" file.
  - Changed the format of the date to date format
  - Selected the rows with required year that are required.
  - Merged table with Staion IDs wit respect to stations
  - Removing stations which are not of Red line using station Ids from "CTA_-System_Information-List_ofL_Stops.csv"
  - Ordered colums according to dates and calculated the total ridership of Red line for particular date
  - Seperated dates and corresponding Redline columns , removed duplicates

- Regression Analysis
  - Plotted the Graphs and since only one predictor and dependant variable.
  - Applied Simple linear, polynomial degree 2 and 3 regression analysis and found that Polynomial Degree 3 method is sufficiently suited.







Model Selection

- We have selected Polynomial Regression D3 because it has more accuracry and less RMSE

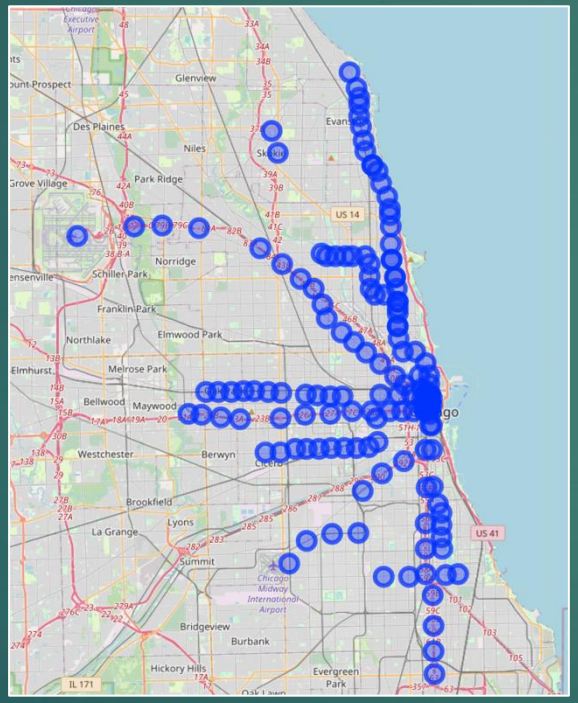
Model	Accuracy (Rsquared)	RSME
Simple Linear Regression	0.1031	18809
Polynomial D-2	0.729	10333
Polynomial D-3	0.745	10021

Geospatial Analysis

- Data gathering
  - We loaded “CTA_-System_Information-List_ofL_Stops.csv”
  - Selected wanted columns for all “L” lines
  - Separated x and y coordinates from location column into separate columns and changed their data type to numeric.
  - Removed Location column to avoid redundancy
  - Loaded “CTA_-Ridership-_LStation_Entries-Monthly_Day-Type_Averages__Totals.csv” to get ridership data.
  - Modifed data type to desired data type for required columns
  - Filtered out specific date
  - Merged both modified loaded files using commom “MAP_ID = station_Id”
  - Removed columns to avoid redundancy

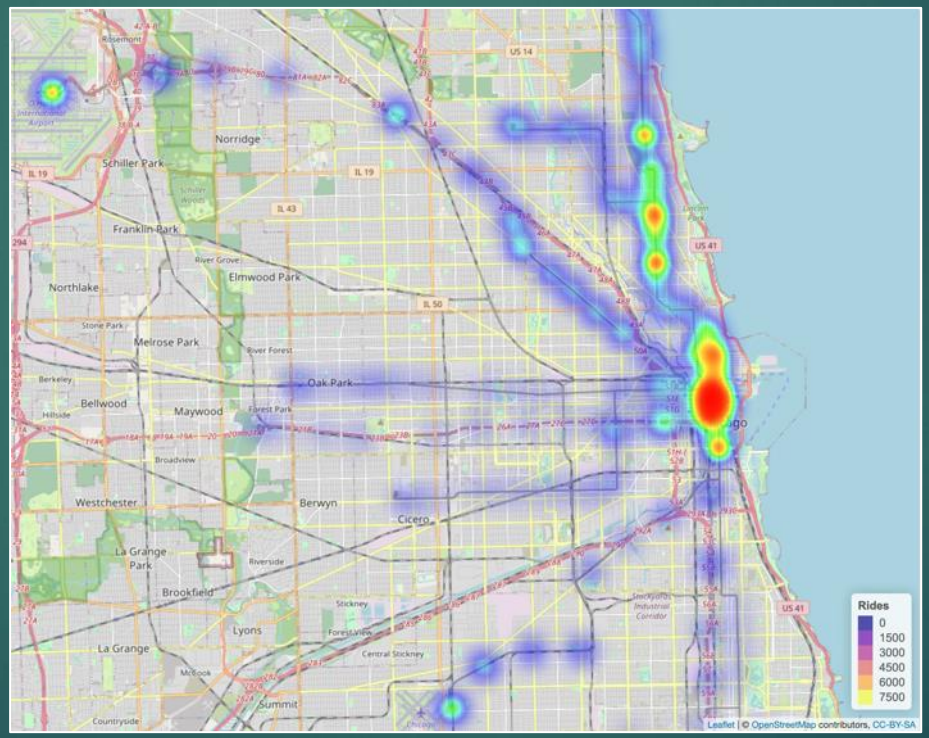
Visualisation

- We mapped all the “L” station on to the map using function **leaflet** package

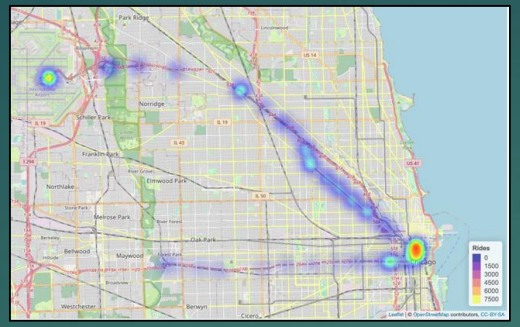


Visualisation

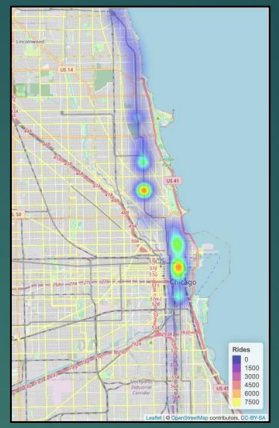
- We mapped busyness all the “L” station on to the map using function **leaflet** package, to get heatmapping, and also with little change to inputs got heat maps for individual lines.



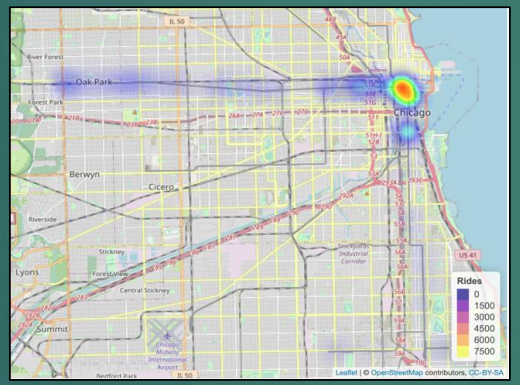
Blue Line



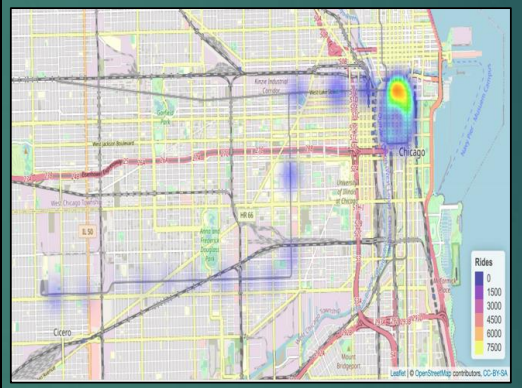
Red Line



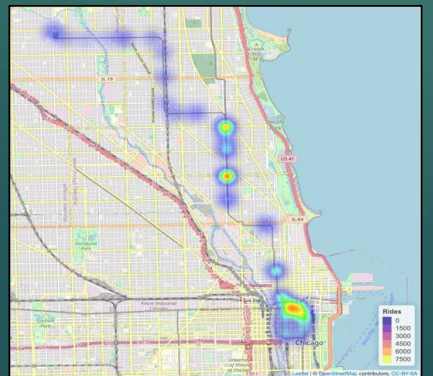
Green Line



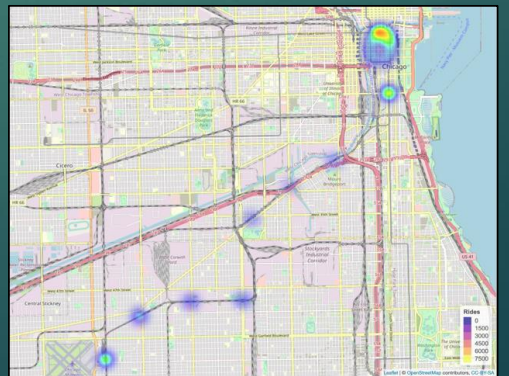
Pink Line



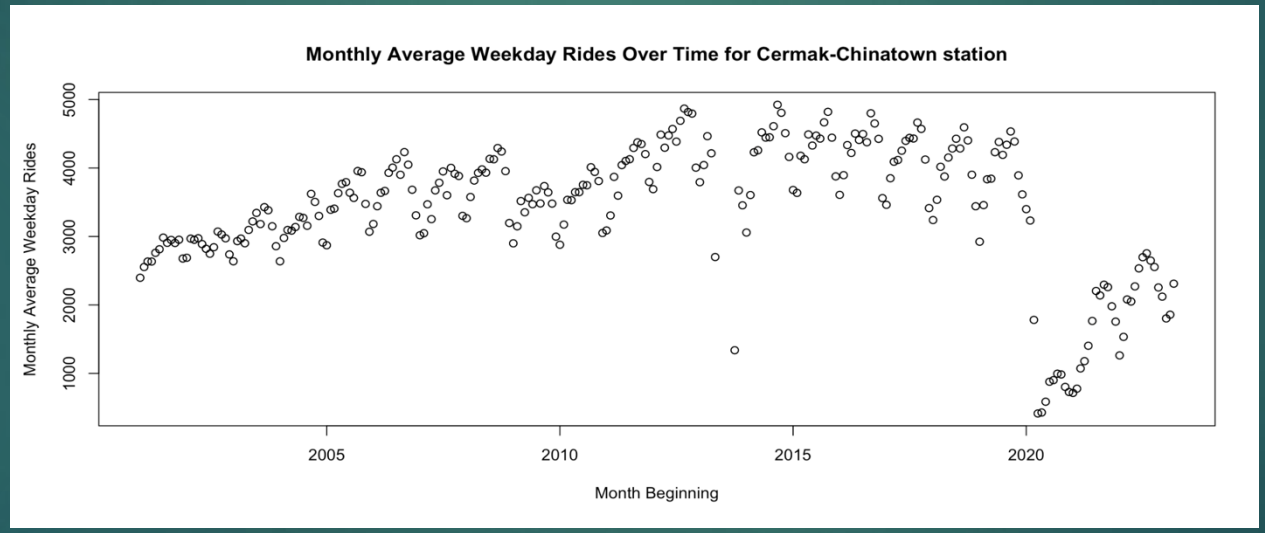
Brown Line



Orange Line

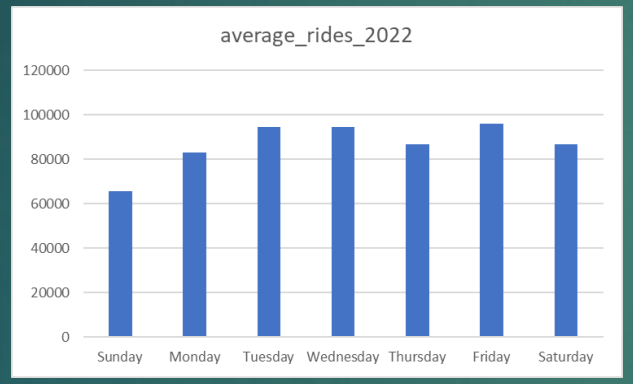


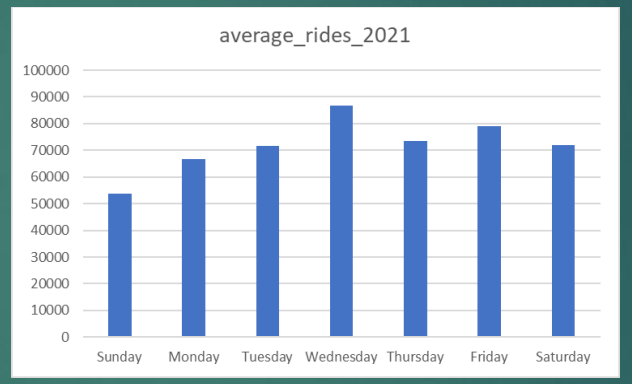
- Using same data frame we plotted change in rider ship over the time for a particular station (Cermak-Chinatown)

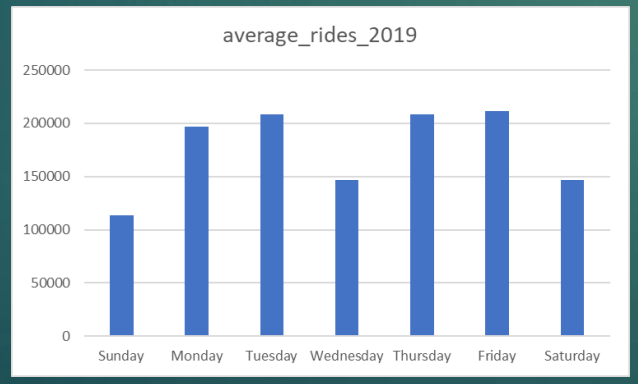


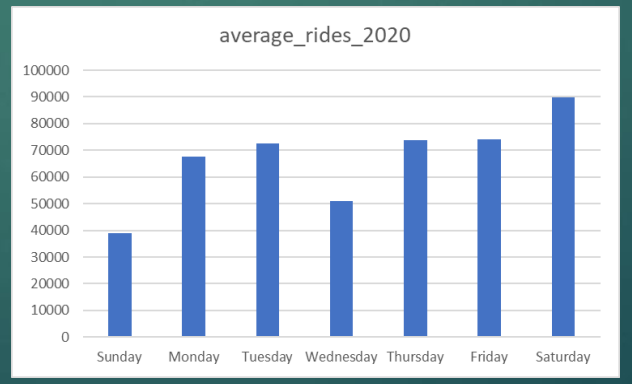
Results

- Impact of COVID-19 on Ridership:

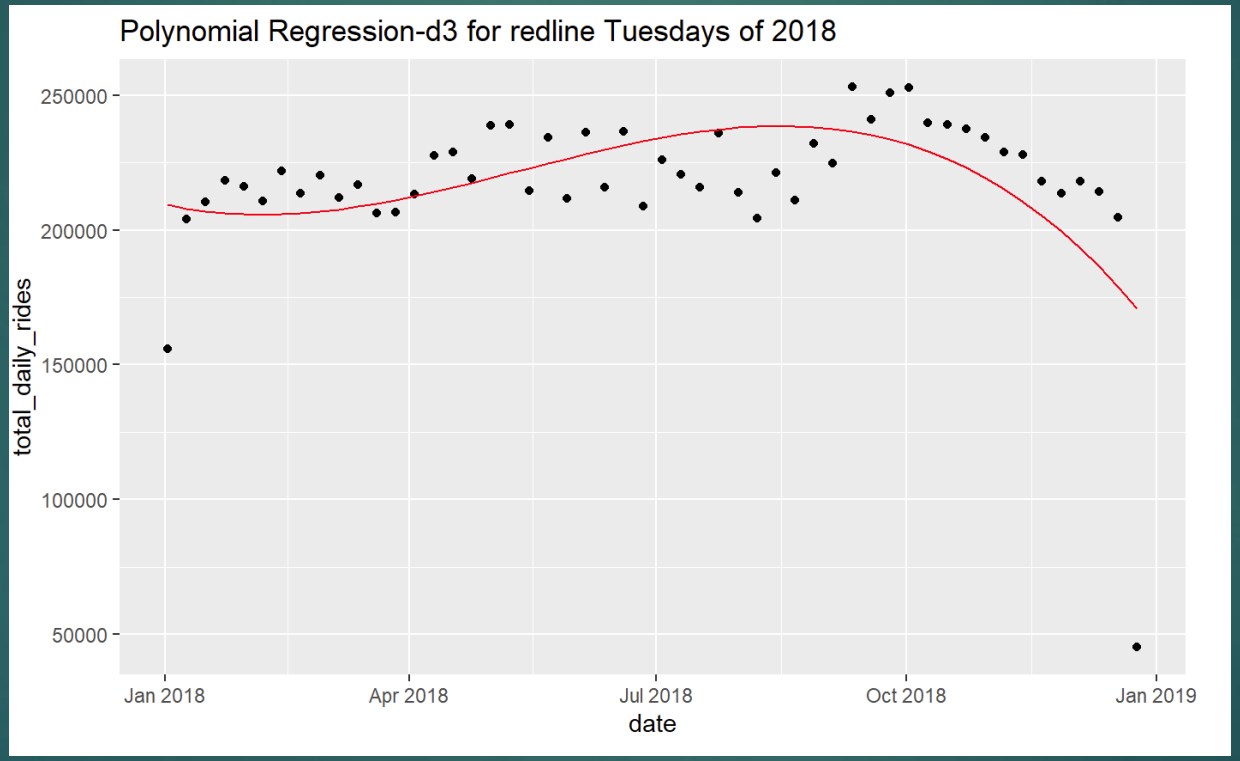




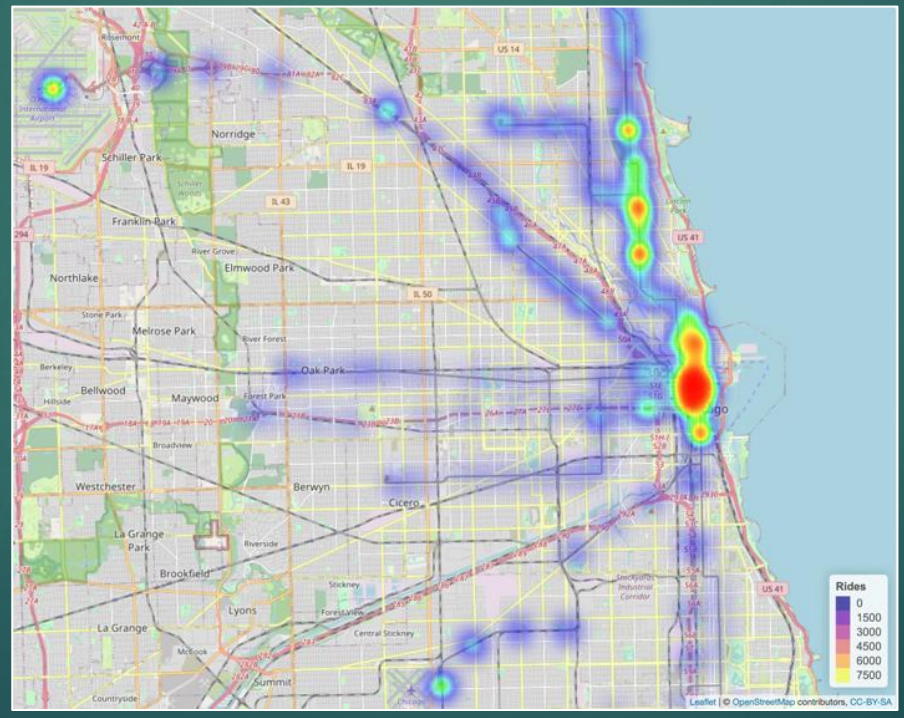




- Seasonal Variation



- High Ridership in Loop and Airport Terminals:



Future Aspirations

- By using API we will get real time data of wait times and delays of all transportation under CTA.
- We will comment on disruption on delays.
- Analyzing data for all routes and station and comment on effective routes with minimum service delays and suggest the optimization.
- We will track weather data for location of station along with API data and analyze how weather effects service delays.