CTA Ridership Analysis

The Impact of COVID on Public Transit Ridership

Team:

Nethra Subramanian

Devanshi Ledwani

Devon Delgado

Karan Mehta

Sigma Consulting Group

Problem Statement

Impact of COVID-19 on Ridership

The COVID-19 pandemic has significantly affected public transportation ridership like the Chicago Transit Authority. As ridership climbs back to pre-pandemic levels, there is a need to maintain efficient operations through capacity planning and resource management.

Objective

Develop robust forecasting models to predict the recovery of ridership levels as the effects of COVID-19 diminish. These forecasts can be utilized to inform CTA personnel on adjustments to operations and capacities to meet future demand.

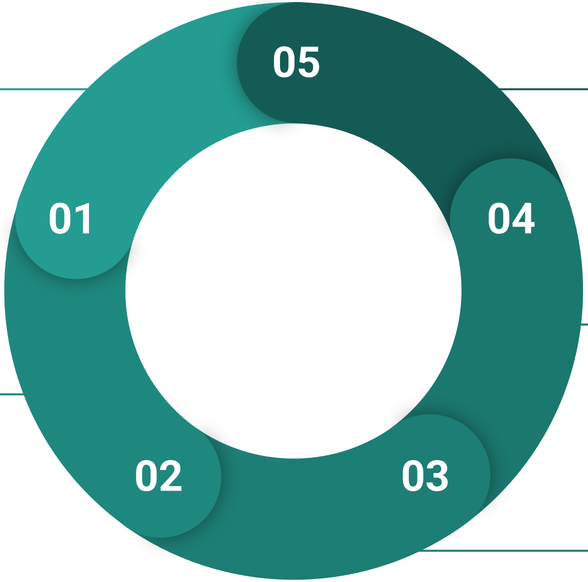
Impact of Accurate Modeling

Accurate forecasts will help in resource allocation, schedule adjustments, and maintaining service quality while adapting to new ridership patterns. This ensures readiness for enhancing the responsiveness of CTA operations.

Chicago Transit Authority Daily Ridership Data from January 2001 to March 2024, sourced from City of Chicago Data Portal

Day Type: Weekday, Saturday, or Sunday/Holiday

Service Date: The day of operation



Bus Boardings: Aggregate bus boardings on a given day

Rail Boardings: Aggregate train boardings on a given day

Total Boardings: Sum of the Bus and Rail Boardings Columns

Assumptions and Hypotheses about the Data and Modeling

01 Dataset Assumptions

- **Stationarity:** Trend stationary pre-COVID, but not post-COVID
- **Seasonality:** Regular repeating patterns (daily, weekly, yearly).
- **Additivity:** Components are additive.
- **No Missing Data:** Appropriately handled missing values.
- **Outliers:** No significant outliers affecting the model.

02 Hypotheses

- **COVID-19 Impact:** Significant changes in ridership patterns due to the pandemic.
- **Post-COVID Recovery:** Gradual recovery and stabilization of ridership.
- **Seasonal Effects:** Influences from holidays, weekends, weather, and events.

03 Modeling Assumptions

- **Linear Relationships:** Level, trend, and seasonality components are linear and additive.
- **Constant Seasonality:** Seasonal patterns remain consistent over time.
- **Model Suitability:** Holt-Winters model effectively captures level, trend, and seasonality.

04 Limitations & Considerations

- **Structural Changes:** Major shifts in ridership patterns not captured by the model.
- **Data Quality:** Accuracy and completeness of the dataset.
- **External Factors:** Economic conditions, fuel prices, and alternative transportation options.

Data Processing and Transformation

Duplicate Data:

62 duplicate rows in the data were found and subsequently dropped.

Out of Order Data:

Ensured correct order of data points by time to produce the best analysis.

Feature Engineering

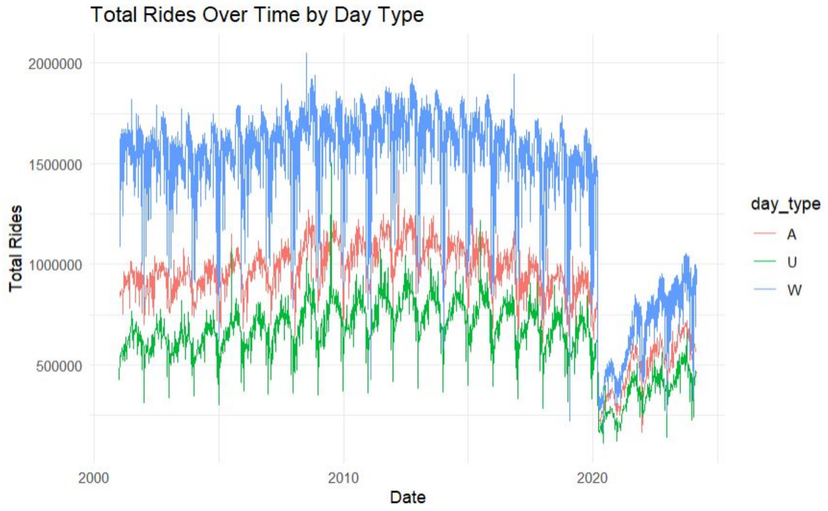
Current Approach:

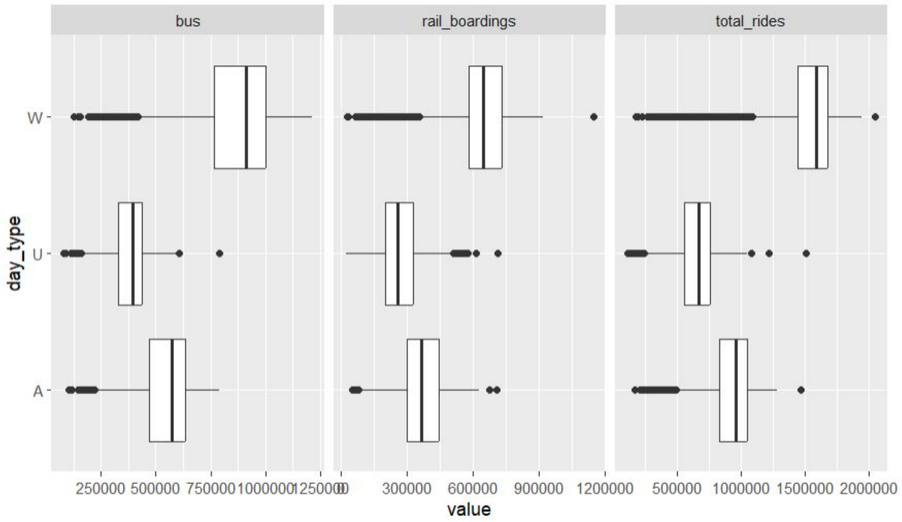
- **No feature engineering** was performed as the data was already suitable for estimating daily aggregates.

Future Work:

- Hourly Aggregates:
    - Consider breaking down the data to **hourly intervals** for more granular modeling, particularly to analyze rush hour and lull periods.
- Enhanced Modeling:
    - By analyzing hourly data, we could improve insights into **daily patterns** and optimize resource allocation during peak times.

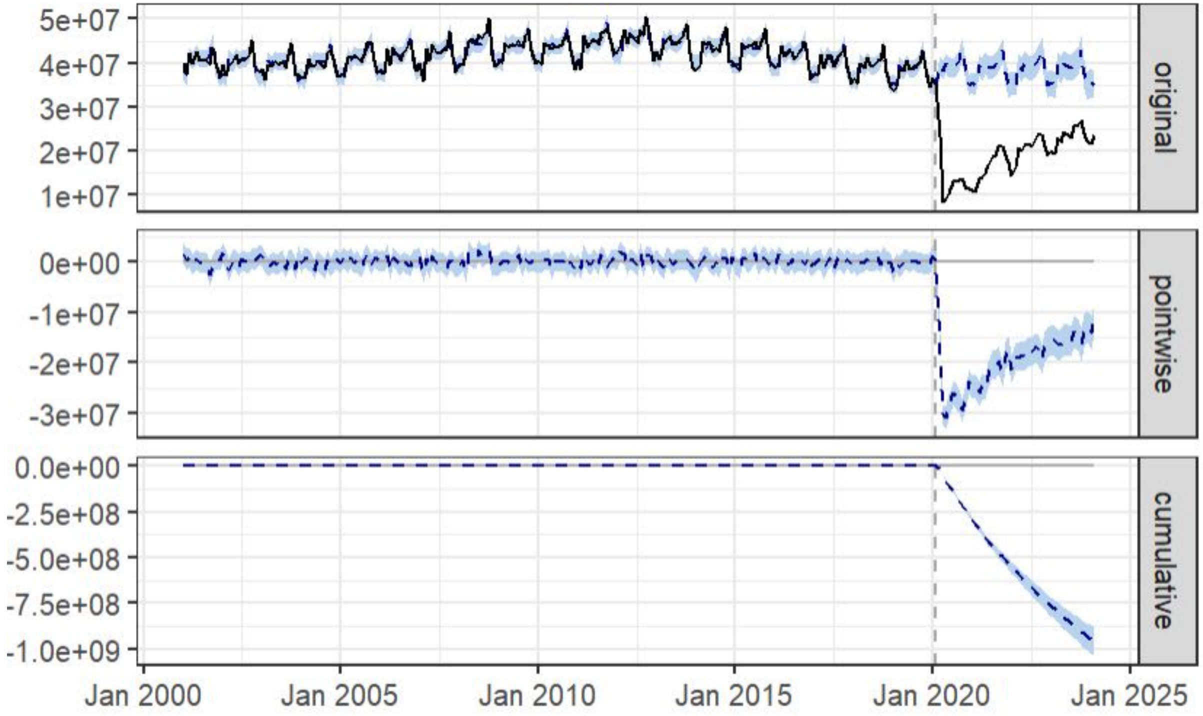
Weekdays Have The Highest Ridership





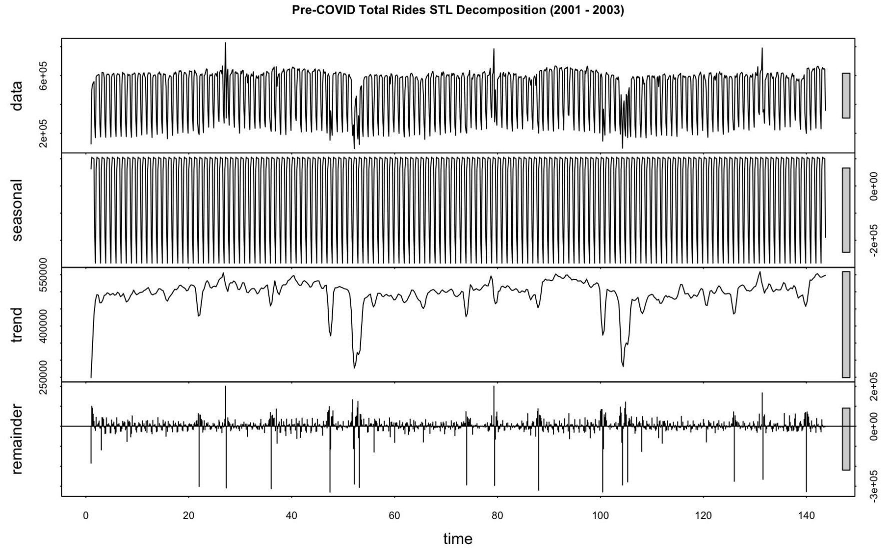
Weekdays have the highest ridership due to daily commuters traveling for work and school, followed by Saturdays which have increased activity for leisure and errands, and then Sundays or holidays which typically have the lowest ridership as many people stay home or engage in less travel.

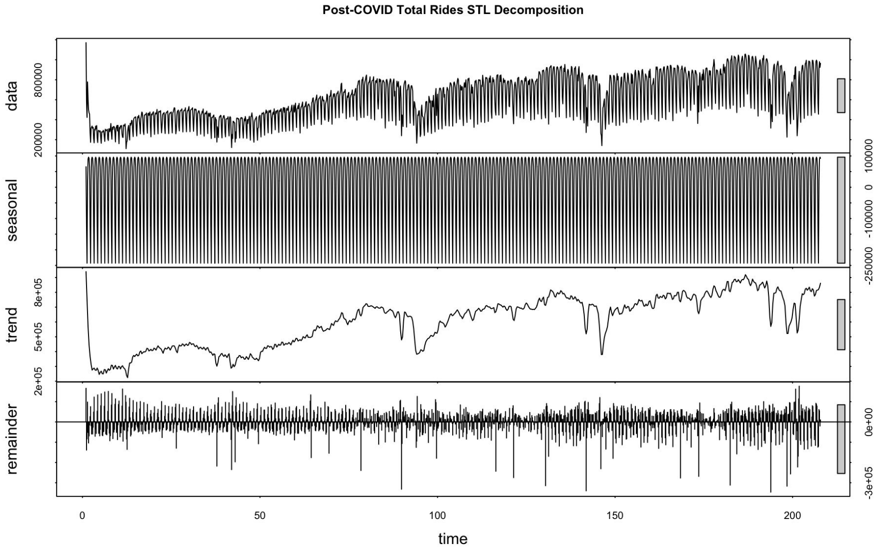
Causal Impact Model



- Monthly Causal Impact model for total ridership shows a -52% decrease in number of riders and also indicates that the causal impact is statistically significant.
- Expected Average response of 38.46M as compared to actual average response of 18.53M.
- CTA started observing a reduction in ridership from May 2013 after LYFT was introduced in Chicago.

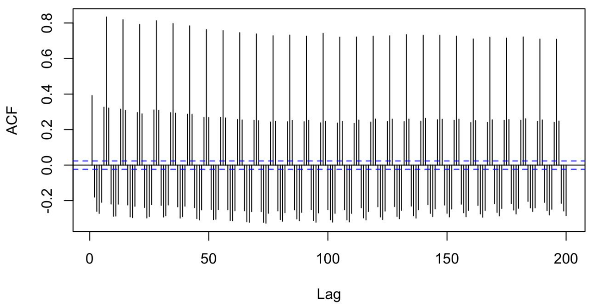
Total Ridership is Trend Stationary in the Pre-COVID era with upward trend in the Post-COVID era



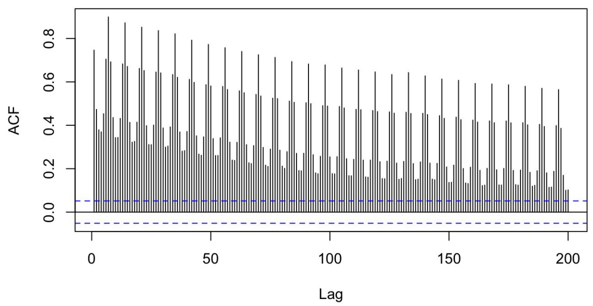


Autocorrelation and Partial Autocorrelation Functions show a Strong Autoregressive Process

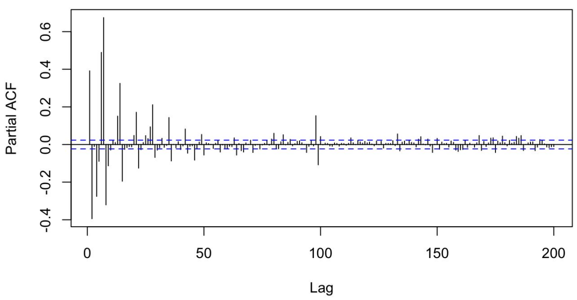
Autocorrelation of Pre-COVID Total Ridership



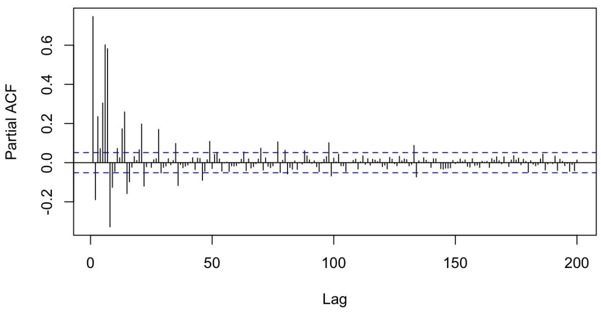
Autocorrelation of Post-COVID Total Ridership



Autocorrelation of Pre-COVID Total Ridership



Autocorrelation of Post-COVID Total Ridership

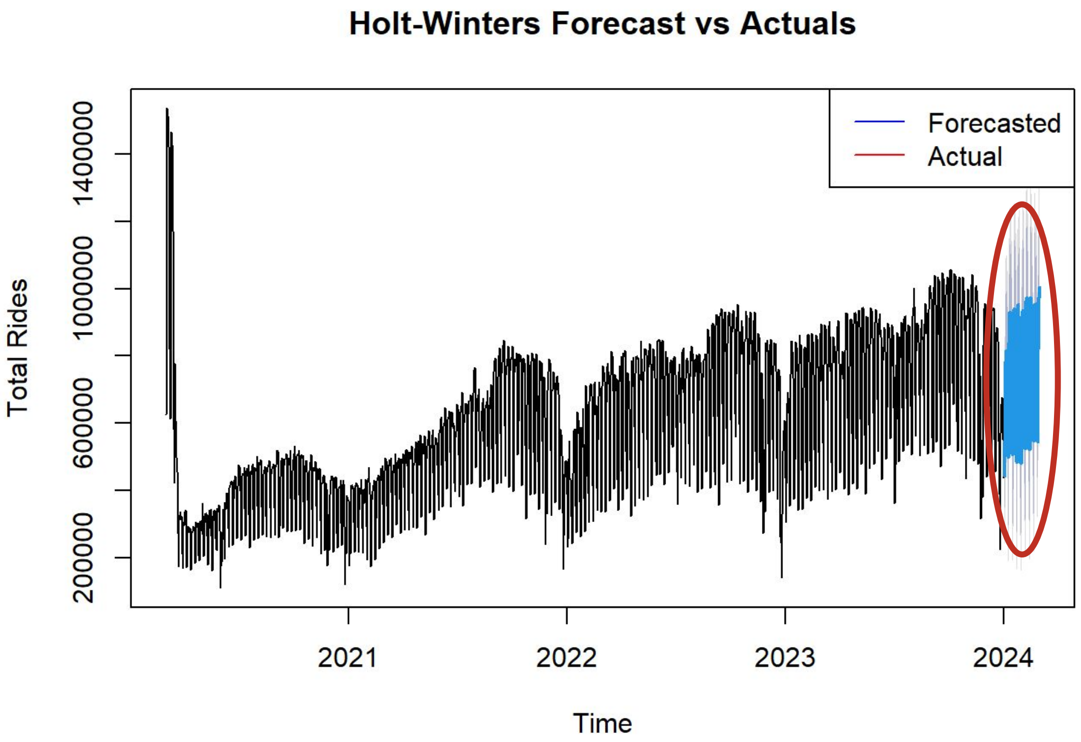


Holt Winters (Post_COVID): Forecasting Public Transportation Ridership

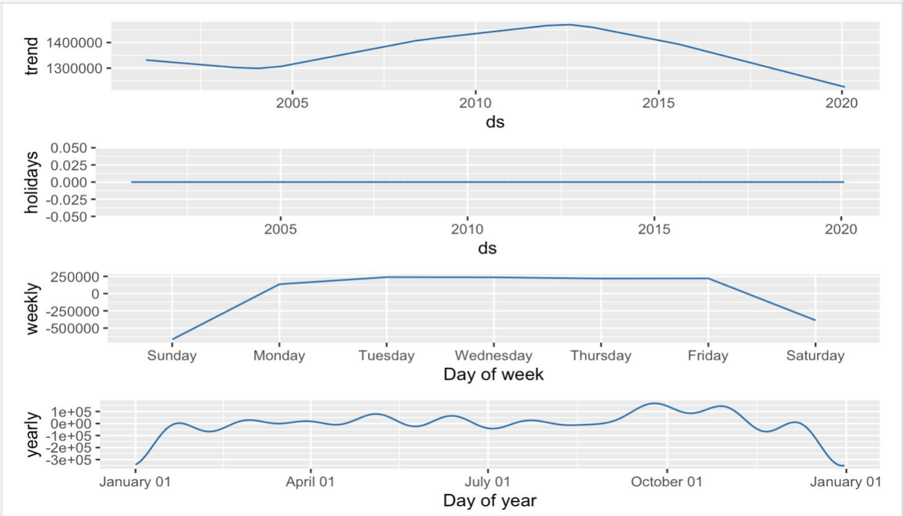
Plot Description:
- **X-Axis (Time):** 2020 to 2024
- **Y-Axis (Total Rides)**
- **Black Line:** Actual rides (2020-2023)
- **Blue Line & Shaded Area**: Forecasted rides for 2024
- **Red Line:** Actual rides for 2024

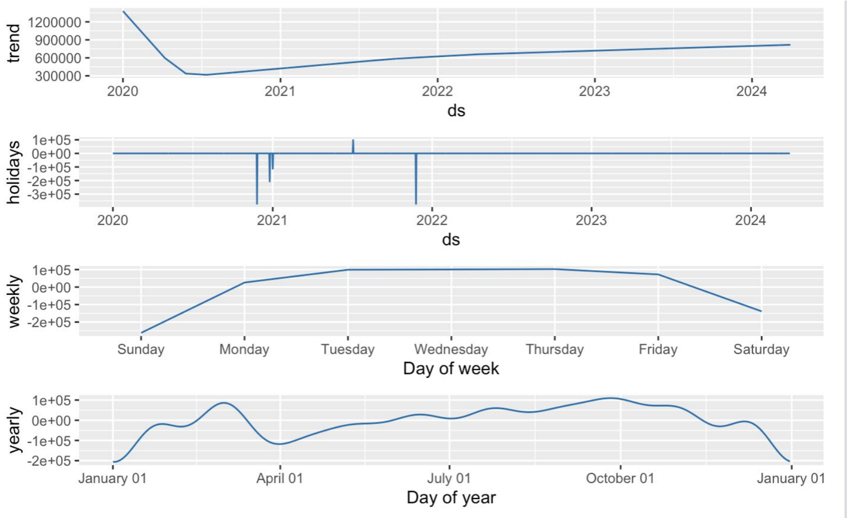
Observations:
- Forecast captures **post-COVID trends** and seasonality.
- Actual 2024 rides align well within forecasted range

Pre-COVID and Post-COVID Trends
- Initial Sharp Drop:
  - Early 2020 saw a **significant decline in total rides** due to the COVID-19 pandemic.
  - Lockdowns and social distancing measures impacted public transportation usage.
- Gradual Recovery:
  - Post initial drop, a **gradual recovery** with fluctuations in total rides.
  - Reflects changing public behavior and transportation policies over time.



Prophet Model





Pre Covid:

Trend: Steady increase until 2015, followed by a decline.

Weekly Seasonality: Higher ridership on weekdays, lower on weekends.

Yearly Seasonality: Peaks around mid-year, lowest at the beginning of the year.

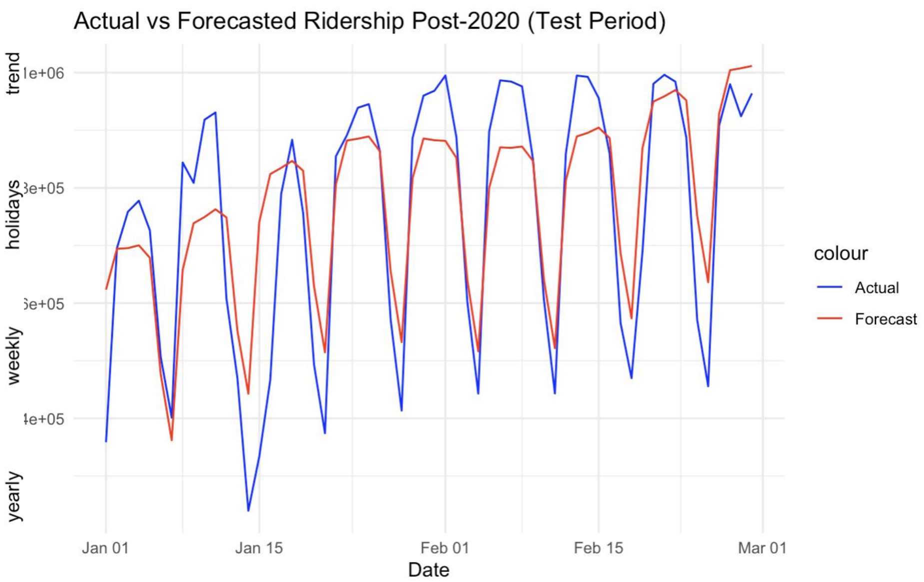
Post Covid:

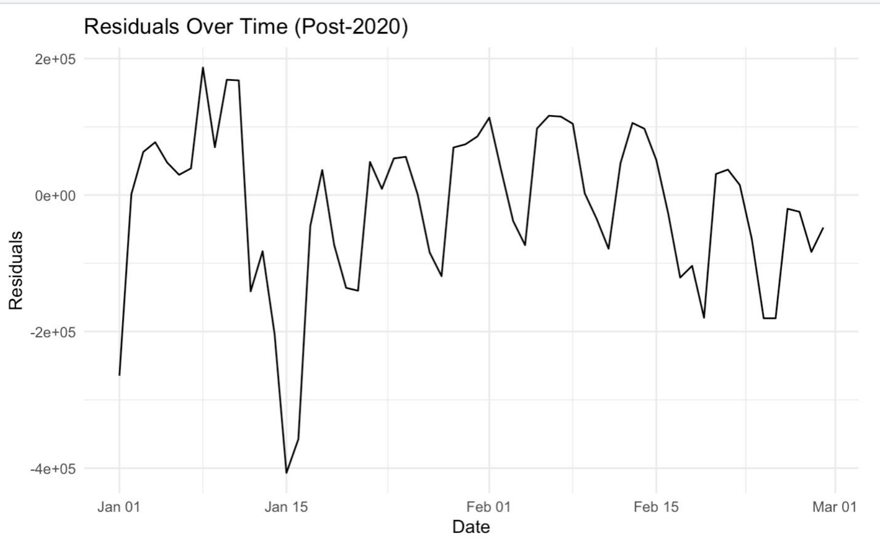
Trend: Sharp decline in 2020, gradual recovery afterward.

Weekly Seasonality: Similar pattern to pre-COVID, but lower weekday peaks.

Yearly Seasonality: Lower overall ridership with disruptions.

Model Performance and Forecast



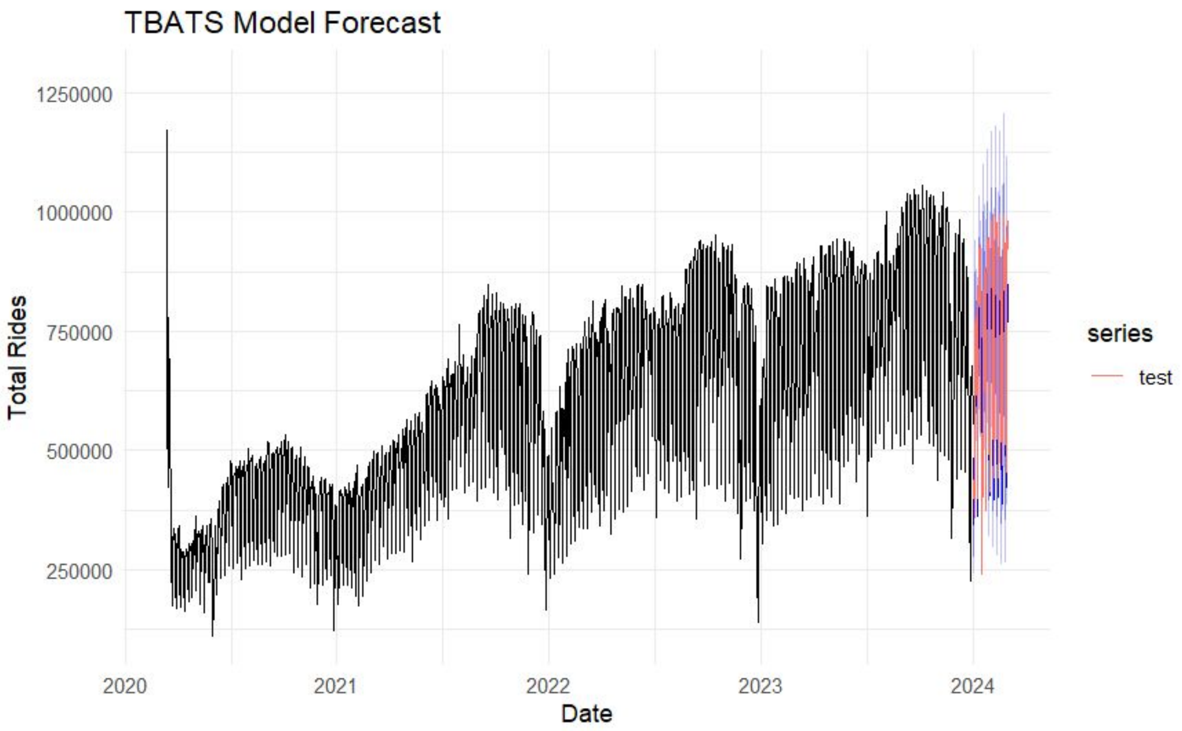


- Actual vs. Forecasted Ridership: The blue line represents actual ridership, while the red line represents forecasted values. The model captures the overall trend and seasonality well, though some peaks and troughs are not perfectly aligned.

The graph shows the difference between actual and forecasted ridership.
While the model performs reasonably well, there are periods with larger residuals, indicating some difficulty in capturing the full variability in ridership during the pandemic.

TBATS Model

- TBATS can handle multiple seasonality modelling.
- The fit takes into account weekly, monthly, and yearly patterns, with adjustments for seasonal amplitude and direction.
- Train Data: 2020/03 - 2023/12
- Test Data: 2024/01 - 2024/02




```
TBATS(0.606, {0,0}, -, {<7,3>, <30,7>, <365,7>})
```

ARIMAX with Fourier Multi-Seasonality

Model Components

Multiple Seasonality

- Models multiple seasonality at the weekly and annual levels
- Uses Fourier series as external regressors

Intervention Analysis

- Uses an intervention as an external regressor
- Accounts for the decline in ridership starting on March 13, 2020

Steps and Methodology

Step 1: Pre-COVID Baseline Model

- Acquire a baseline ARIMA model for pre-COVID ridership

Step 2: ARIMAX Model with Pulse Intervention

- Train an ARIMAX model with a pulse on March 13, 2020 and acquire ARMA coefficients of the intervention

Step 3: Build a model for the decay of the intervention effect over time

- Build a model of how the intervention effect dies down as time moves on after the onset of COVID, training an Arima model with this effect as an external regressor

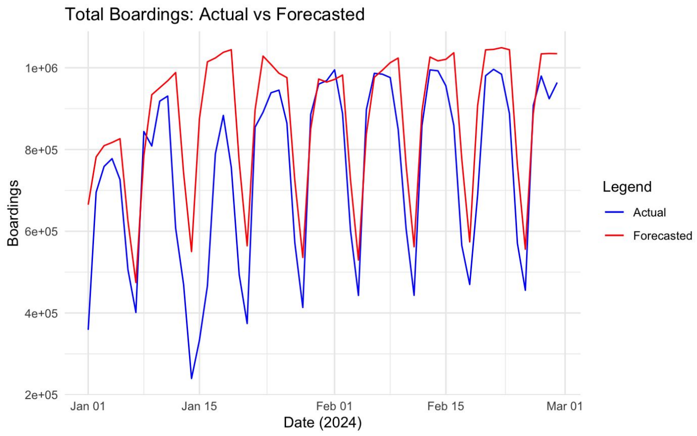
ARIMAX with Fourier Multi-Seasonality

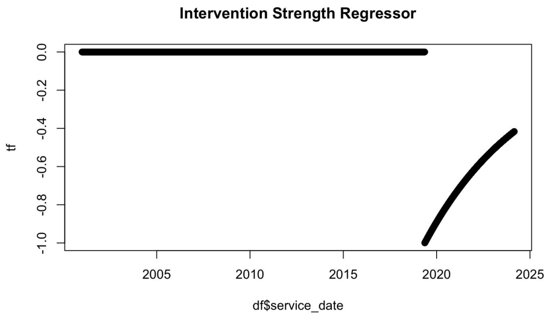
The model captures seasonality well with Fourier transforms, but the forecasts are slightly narrower.

The Intervention Strength Regressor displays how the effects of COVID are dying down at a logarithmic rate, showcasing how its effect on ridership are still felt 4 years later.

More robust forecasts can be captured by including more holidays as external regressors.

RMSE	MAE	MAPE
171,792.1	126,585.2	23.5%





Model metrics forecasting into the first two months of 2024

Model metrics were performed on the total aggregated bus and rail ridership

Model	RMSE	MAE	MAPE
Holt-Winters	196,500	154,658	26.33%
Prophet	123,438.3	103,872.66	19.4%
TBATS	136,913.37	112,943.97	16.74%
ARIMAX	171,792.1	126,585.2	23.5%

TBATS Model performed the best, due to its robust attention to multi-seasonality and its ability to capture complex trends in the data.

Conclusion and Future Work

COVID-19 Impact

- The pandemic significantly affected CTA ridership, with recovery expected to take years. The **TBATS model provided the most accurate forecasts** due to its handling of multiple seasonalities and variance stabilization.

Improving Forecast Accuracy

- To improve accuracy, incorporate additional features like weather, events, and economic indicators. Use periodograms to better identify seasonal patterns in the data.

Appendix (Contributions)

Devon	EDA on Bus, Train, and Total Ridership, concluded that Bus and Train behave similar enough that we can just use Total Ridership to forecast. Built ARIMAX with Fourier Multi-Seasonality Modeling.
Devanshi	Built Prophet models to predict CTA ridership trends, splitting data into pre-2020 and post-2020 periods to account for COVID-19's impact. EDA included data cleaning and visualizations. Models incorporated holidays and generated 30-day forecasts, validated by RMSE and MAPE. This provided actionable insights for CTA planning.
Karan	Build a Causal Impact Model to find the statistical significance of the intervention event (COVID-19 Pandemic), which is statistically significant. EDA on Bus, Rail Boardings and total ridership as well as Causal Impact model on all 3. TBATS model on the total ridership considering all the seasonalities in the data.
Nethra	EDA on bus, rail boardings, and total rides by day_type, revealed key patterns in ridership across different days; developed a Holt-Winters model to forecast total ridership, capturing trends and seasonal variations post-COVID, provided reliable predictions validated against actual data