**Executive Summary**

The Chicago Transit Board has approved nine renovation projects with the goal of modernizing and improving the city’s buses and trains, stations, tracks, and implemented technology so that transit is more accessible to Chicago’s residents (“System improvement projects,” 2023. The high cost of these renovations may have room for improvement when the scheduling is paired with Time Series Modeling (TSM). This analysis seeks to answer whether CTA boarding rates can be reliably forecasted at 90 percent accuracy via Time Series Analysis:

* Hypothesis – CTA boarding rates can be forecasted with 90 percent accuracy.
* Null Hypothesis – CTA boarding rates cannot be forecasted with at least 90 percent accuracy.

Chart

Description automatically generatedThe analysis was done through Python and utilized multiple different Python packages including:pandas, numpy, sklearn, statsmodels, pmdarima, and scipy. The TSM requirements of having a stationary and univariate data as well as the various packages needed to create a working TSM model requires a large amount of memory and processing power that may not be available to the Chicago Transit Board at this time. The analysis first begins by modifying the ridership dataset so that it is univariate and then applying the Augmented Dickey Fuller Test (ADF) to determine stationarity.

Chart

Description automatically generated with low confidenceWith a non-significant p-value of .4815 and critical values whose absolute value is greater than the ADF statistic, the modified dataset was determined to be non-stationary. The dataset was then made to be stationary through differencing and split into train and test datasets for modeling.

Model selection was determined after plotting the seasonal\_decompose() function to visualize the Trend, Seasonality, and Residuals of the dataset:

A picture containing background pattern

Description automatically generated

Graphical user interface, chart, line chart

Description automatically generatedChart, line chart, histogram

Description automatically generatedThen the acf() and psd() functions to visualize the autocorrelation and spectral density of the dataset

Next, the order of the ARIMA model was discovered by using a stepwise approach of finding the lowest AIC value of -2531.143 and the associated PDQ value of (0,0,0). The model was then forecasted with an interval of 1 step and plotted using 95% confidence interval to visualize the predicted values of -0.000022.

Graphical user interface, application, table

Description automatically generated

Chart

Description automatically generatedEven with a low Mean Absolute Error (MAE) value of .000609 and .000807 for both the test and train set respectively, the model lacks statistical soundness based on the plots visualized by the plot\_diagnostics() function.

Based on these findings, the Chicago Transit Board can pursue TSM with a larger dataset in order to have an appropriately trained model to improve performance or abandon predictive modeling altogether and apply another statistical method, like clustering. Expected benefits of the analysis can be influenced by the drastic drop from 179,071,205 to 62,340,303 in ridership between the year 2019 and 2020. In addition, based on the fact that the ridership data collection process is automated – the Chicago Transit Board should try to provide a dataset with a daily frequency rather than a monthly frequency to significantly increase the original dataset size of 264 to around 8,034.

**References**

Athanasopoulos, G. & Hyndman, R. (2023). Forecasting: Principles and Practice. OTexts. Retrieved from <https://otexts.com/fpp2/prediction-intervals.html>

Chicago Data Portal. (2023). Retrieved from https://data.cityofchicago.org/

*Chicago Transit Board approves plan for rehabilitation of the Western Brown Line Station.* (2023). 2023 Chicago Transit Authority. Retrieved from https://www.transitchicago.com/chicago-transit-board-approves-plan-for-rehabilitation-of-the-western-brown-line-station/

*CTA ridership jumps nearly 25% in 2022, but still trails far behind pre-pandemic levels.* (2023). CBS Chicago. Retrieved from https://www.cbsnews.com/chicago/news/cta-ridership-2022-covid-19-pandemic/

*CTA - Ridership - 'L' Station Entries - Monthly Day-Type Averages & Totals.* (2023). Chicago Transit Authority. Retrieved from https://data.cityofchicago.org/Transportation/CTA-Ridership-L-Station-Entries-Monthly-Day-Type-A/t2rn-p8d7

Pandian, S. (2021). *Time series analysis and forecasting: Data driven insights (Updated 2023)*. Analytics Vidhya. Retrieved from https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/#What\_Are\_the\_limitations\_of\_Time\_Series\_Analysis?

Ridership Readme. (2011). Chicago Transit Authority. Retrieved from https://data.cityofchicago.org/api/assets/7971211C-57DF-4650-8EBD-BE295CCD3151?download=true

System improvement projects. (2023). 2023 Chicago Transit Authority. Retrieved from https://www.transitchicago.com/projects/

Zuniga, J. (2023). Arima Models. Amazon AWS. Retrieved from https://rstudio-pubs-static.s3.amazonaws.com/363796\_2c2581467b02403bbdcc77c6b5660ec7.html#:~:text=The%20ARIMA(0%2C%200%2C,variety%20of%20time%20series%20models.