For our initial modeling attempts, we first consolidated all available monthly Divvy data from 2020 to 2024. We constructed a time series dataframe based on the year and month, including the total count of ridership, the count segmented by ridership type, and exogenous variables such as membership types. (Due to tokenization errors, the data for May, June, July, and August 2024 could not be joined and were excluded from the dataset.)

The ACF and PACF graphs of all the ridership counts are fairly consistent with sinusoidal trend and significant lag points in the beginning with ACF and sharp decline after lag 1 in PACF. Residuals fit nearly all assumptions as well. We also analyzed the casual and membership riders individually, seeing seasonal trends here also with basic SARIMAX hyperparameters. The latest data have seen some unexpected turbulence and could not be fit by any model well, although SARIMAX was able to predict the peak usage for casual riders amidst the missing data.

When examining the initial total ridership time series, we observed a consistent seasonal trend across all ridership types: ridership increases from winter to spring, peaks in summer, and declines in fall. Additionally, there does not appear to be a clear upward or downward trend in ridership over the years, as the counts remain relatively stable. Given the presence of seasonality, we tested the total ridership counts using SARIMAX and Prophet models. The Prophet model outperformed the ARIMA models, demonstrating strong predictive accuracy and reasonable error metrics.

Next, we analyzed the time series for classical bikes, which showed a strong correlation with member user counts. At this stage, we tested SARIMA, Auto ARIMA, and Prophet models. Auto ARIMA emerged as the most reliable for forecasting, exhibiting the lowest residual autocorrelation and error metrics.

Electric bikes and docked bikes had the fewest data points, with several months containing null values. As a result, when performing time series analysis on these categories, we must be cautious of potential excessive volatility and reduced model predictability.

For potential next steps, we plan to continue fine-tuning and optimizing the hyperparameters for each model to identify the best-performing configurations. Additionally, we will test the forecasting capabilities of these models using the 2025 data that Divvy has released or will release in the coming months. We also intend to incorporate additional exogenous variables, such as other ridership counts, alongside membership types. Finally, we will compare all models for each ridership type to determine the optimal approach for forecasting ridership far into the future.