DSCI Paper Rough Draft

**Goal: What does scooter usage look like in the future?**

(might exclude this since I used a lot of different models and not just ARIMA) ARIMA – What is this?

https://en.wikipedia.org/wiki/Autoregressive\_integrated\_moving\_average

Autoregressive integrated moving average (ARIMA) is a type of time series analysis used for understanding historical data and predicting future values. Autoregressive (AR) means “that the evolving variable of interest is regressed on its prior values” (Wikipedia). Moving average (MA) means “that the regression error is a linear combination of error terms whose values occurred contemporaneously and at various times in the past” (Wikipedia). Integrated (I) means “that the data values have been replaced with the difference between each value and the previous value.

**Methodology**

Using the “Date” column, we created month, year, and week number columns. We used separate data frames for weekly and monthly data. In each of these data frames, we grouped by the time frame of interest to get the sum of scooter rides in each interval. We then created two new columns for the normal and college prices. In both analyses data from 2024 was used as testing data, and previous data was used as the training data in the form of a time series. The accuracy of the models in both analyses was measured using the mean absolute percentage error (MAPE). In both analyses, several models were tested, but only the most accurate models will be included in this paper. The college price decrease initiated in July 2024 caused a huge spike in ridership that was not captured by most models. Thus, in both analyses, the final model is a combination of two models: one that was accurate before the price change and one that was accurate after the price change.

Weekly Rides

Before the price change, a hybrid model composed of an ARIMA, STLM, and THETAM model (weighted equally) most accurately predicted ridership when fit on the testing data, with a MAPE of 29.69. The plot below compares the actual ridership values to the forecasted values; the date of the price change is marked by a dotted black line.

[plot]

An STLF model most accurately captured the spike in ridership when fit on the testing data, with an overall MAPE of 47.12.

[plot]

The final model for predicting weekly ridership uses the hybrid model before the price change and the STLF model after the price change. This model accurately captures the trend before the price change and the spike after the price change and has a MAPE of 29.04 on testing data.

[plot]

Monthly Rides

Before the price change, a hybrid model composed of an ARIMA, STLM, THETAM, NNETAR, ETS, and TBATS model (weighted equally) most accurately predicted ridership when fit on the testing data, with a MAPE of 14.39.

[plot]

A boosting model most accurately captured the spike in ridership when fit on the testing data, with an overall MAPE of 29.39.

[plot]

The final model for predicting weekly ridership uses the hybrid model before the price change and the boosting model after the price change. This model accurately captures the trend before the price change and the spike after the price change and has a MAPE of 8.86 on testing data.

[plot]

**Results**

The predictions for future weekly and monthly ridership will be made assuming that the normal and college prices remain the same.

Weekly

Since the STLF model is used for predicting after the price change, this model was used to predict ridership through the end of 2025. The plot is shown below, beginning at the end of the testing data.

[plot]

In 2025, the week with the lowest ridership (around 122) will be the first week of January 2025 and the week with the highest ridership (around 7300) will be the week of August 25. The January prediction seems slightly low. The August prediction seems accurate, since it is slightly higher than that week in the previous two years.

Monthly

Since the boosting model is used for predicting after the price change, this model was used to predict ridership through the end of 2025. The plot is shown below, beginning at the end of the testing data. \*Note: this will not track a trend, will now produce the same predictions for each January, February, etc.

[plot]

In 2025, SPIN can expect January to be the month with the lowest ridership (around 2350) and September to be the month with the highest ridership (around 22900). The January prediction is slightly lower than all previous January values. The September prediction is about the same as the September 2022 value.