Joseph Foster V00985606
Corey Gibbons V00991339
Tomas Guevara Mix V00980970

Introduction

The goal of our analysis is to forecast the number of cyclists daily on Victoria's galloping goose trail. Creating a regression model and plotting our results helped unveil reasons for peak days and cycling patterns among pedestrians. The following information outlines our process and findings through our model.

Data

The dataset given contains the number of cyclists on the Galloping Goose trail for each day between January 1st, 2019 to March 14th, 2024. The dataset also tracks the month of that day, as well as if it was a weekend or weekday when the data was recorded. Each month, weekend and weekday are imputed as binary variables (represented as 1 if it is true, and false otherwise). Days between March 13th and April 1st have the number of cyclists listed as NA as the data has not been recorded thus far and will be predicted.

Methods

To estimate the number of cyclists on the Galloping Goose, our model is:

$$Cyclists = B_0 + B_1 Weekend + B_2 cyclists\_lag + *BMonths$$

Where B_0 is the intercept, B_1 represents the change in cyclists depending on if it is Saturday or Sunday and B_2 represents the effect of the number of cyclists on the previous day. *$BMonths$, represents each month as it's own regressor with it's own respective coefficient, e.g. $B_3 January + B_4 February + \ldots + B_{13} November.$ Since the months are binary variables, one month (December in our case) is omitted as it will be represented in the intercept.

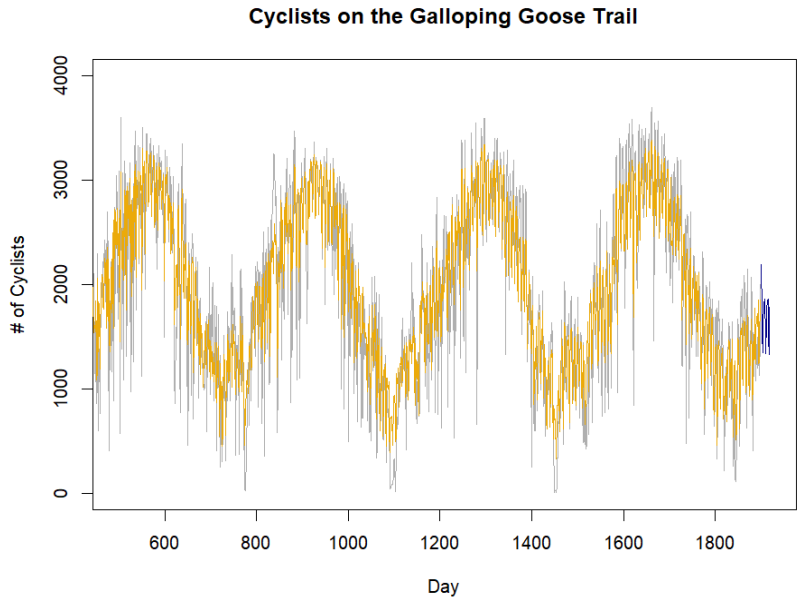
The weekend variable is necessary due to change in individuals schedules, whether it is more time to bike on weekends or not needing to bike to work. The lagged variable is important to account for unforeseen circumstances that may affect multiple days such as a lockdown. Each month variable can account for weather and other seasonal factors.

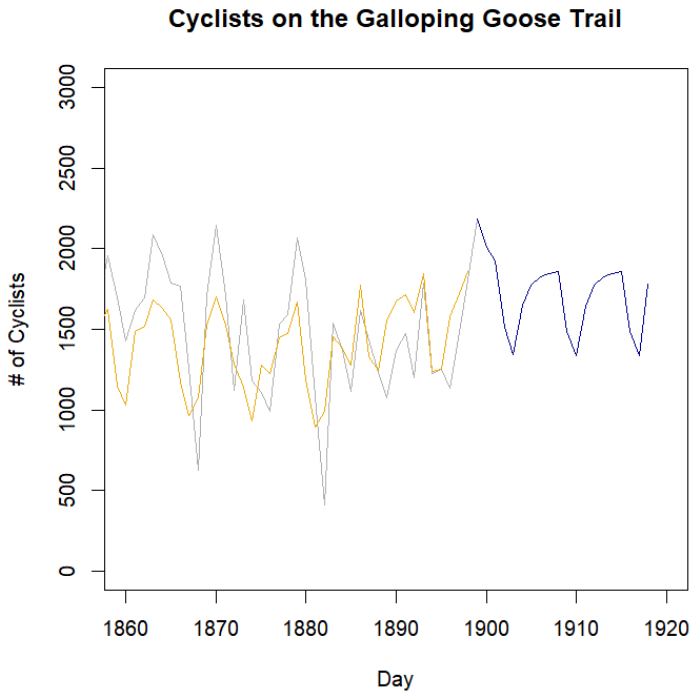
Results

Table 1: Estimated Results	Dependent variable = Number of cyclists
Intercept	701.38180 (45.25524)
Weekend	-374.23771 (25.16724)
cyclists_lag	0.41266 (0.01956)
*Months	
Number of Observations	1899
R-Squared	0.6991

Notes: Standard Error shown in parenthesis.
*Data for each month is not shown due to redundancy, discussed in conclusion.

Figure 1: Observed (grey) and predicted (orange, blue) cyclists on the trail





Conclusion

Winter months have on average less cyclists as the coefficients are smaller compared to summer months as seen in the graph. This is very likely due to the fact that temperature plays a big role in determining if people want to go biking. Each individual month regressor proved significant at the 0.05 level except for January. Although, we opted to keep January in the model since this is due to it being the month after December which is considered in the intercept. We also observed that there were less cyclists on the trail when it was a weekend with a coefficient of about -374. Our assumption is that the majority of cyclists use the Galloping Goose as a form of commuting to work. We observe that the prediction does not account for extreme values as there seems to be a general underprediction in peaks and overprediction in troughs relative to the observed (Figure 1). The model could be improved by considering temperature or precipitation, although these regressors most likely would display high levels of collinearity with each month regressor and the lagged regressor.

References

Galloping Goose Trail. Data.eco. (2014, November 12). https://data.eco-counter.com/public2/?id=100117730