WGU MSDA D214 Task 3 Capstone

AUTHOR
Tyson Biegler

PUBLISHED
May 19, 2025

Executive Summary: Bike rental forecast analysis using Meta's prophet

Problem statement and hypothesis

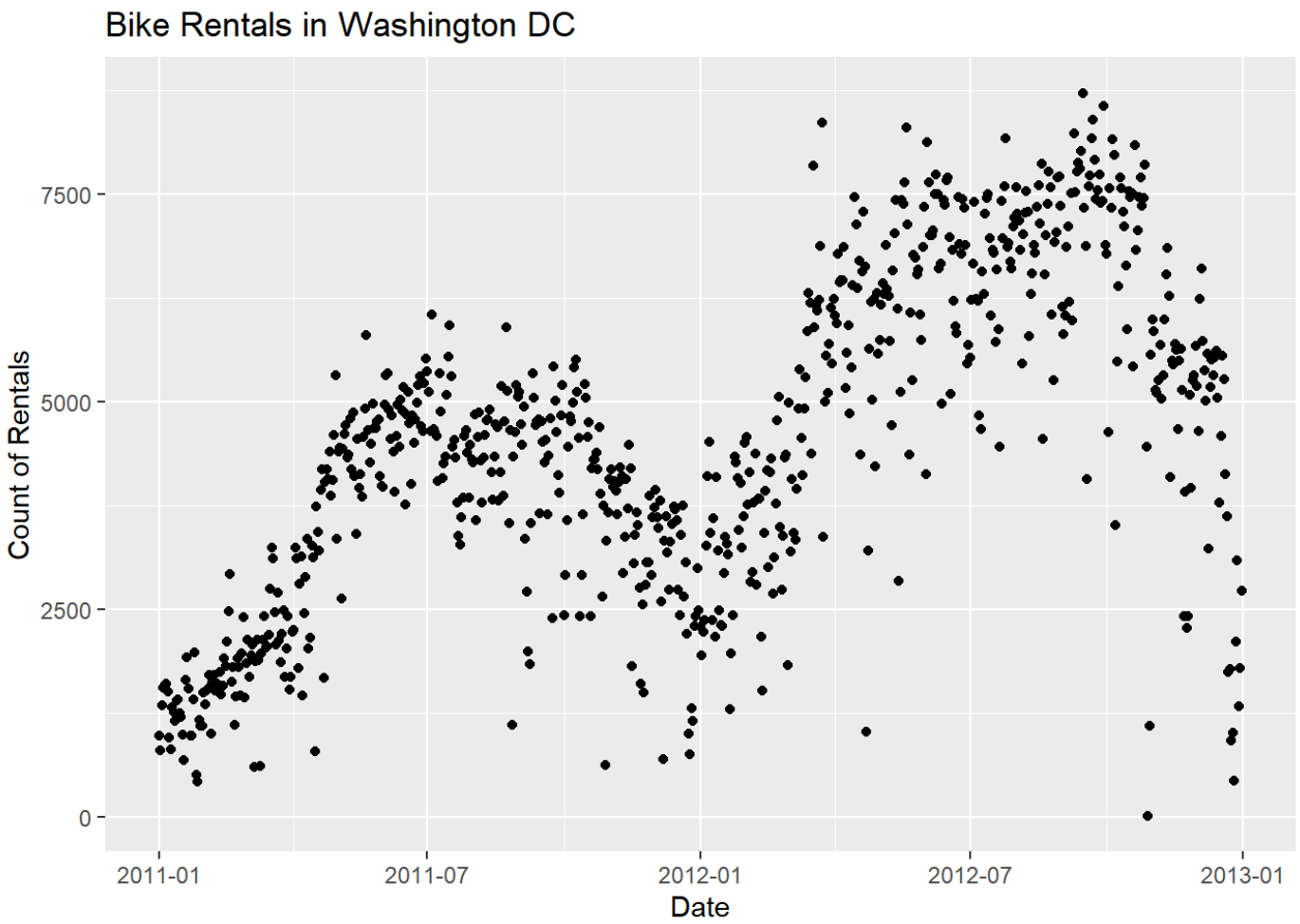
The Goal of this analysis was to assess whether incorporating external variables like temperature, humidity and US holidays improves the forecasting accuracy of daily bike rentals for Capital Bikeshare in Washington DC.

The **null hypothesis** is that these variables do not statistically significantly enhance the forecasting accuracy, while the **alternative hypothesis** is that they do statistically significantly enhance the forecasting accuracy.

Accurately forecasting of bike rental demand is critical given that Washington experiences a high influx of tourism each year (RoadGenius, 2024) as well as the global scale of bike share programs (Fanaee-T, 2013).

Data analysis process

The data used in this analysis is from the UCI Machine Learning Repository, comprising of two years of Capital Bikeshare's 731 daily rental data. The dataset also includes weather data from freemeteo.com. Data preparation was conducted in R using tidyverse, lubridate, and prophet libraries. The dataset was formatted to meet prophet's requirements, ds for date, and y for rental count (Facebook, n.d. - a). Plotting the data reveals a clear upward trend and potential seasonality as it appears that the winter months have consistently fewer rentals than the summer



Two models were created. The first is the baseline model that includes date, and rental count, and the second model is the enhanced model that contains the additional variables of temperature, humidity, and US holidays.

Cross validation was performed with a 540 day training window, a 30 day period, and a 90 day forecast horizon. Model performance was evaluated using the Root Mean Square Error (RMSE), and a paired t-test compared the RMSEs to determine the statistical significance.

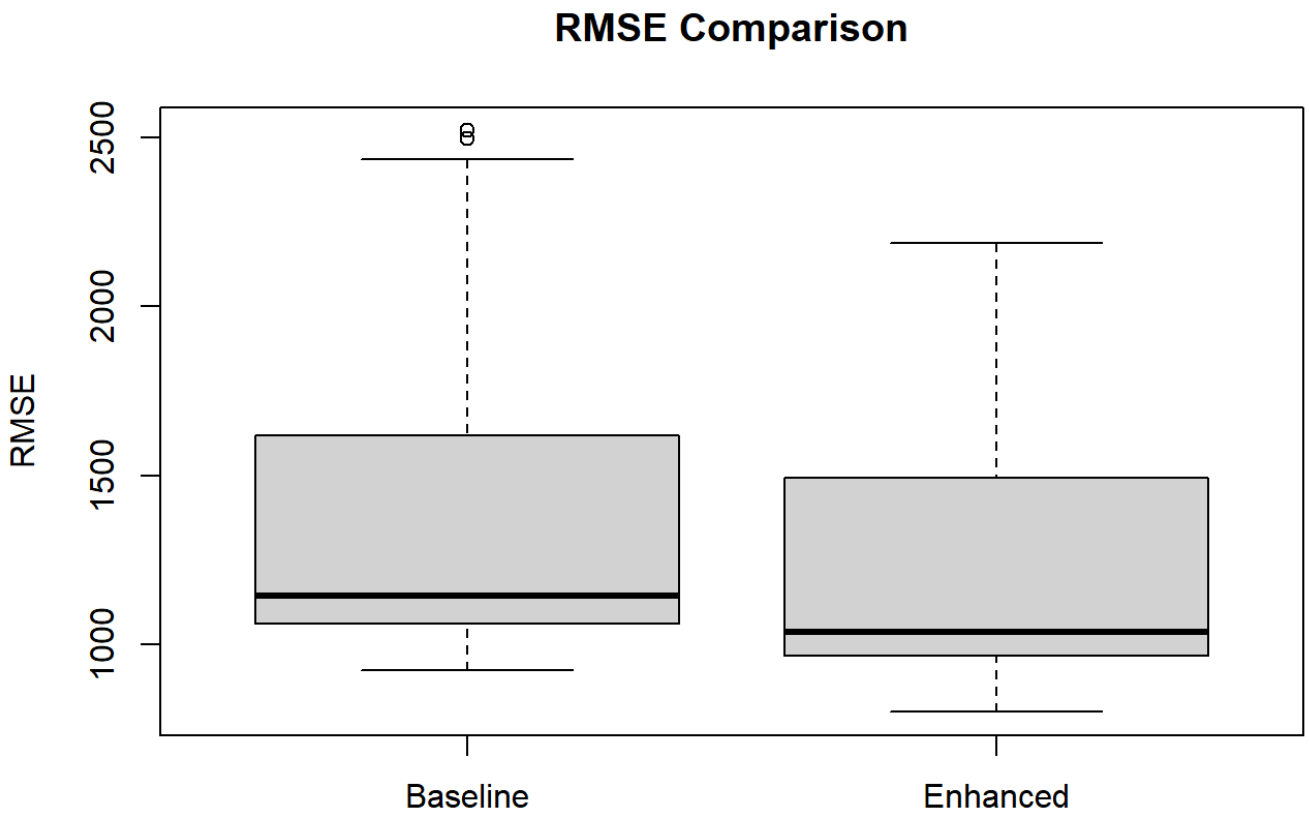
Findings

The enhanced model outperformed the base model with a statistically significant *(p-value < 2.2e-16)* reduction in RMSE

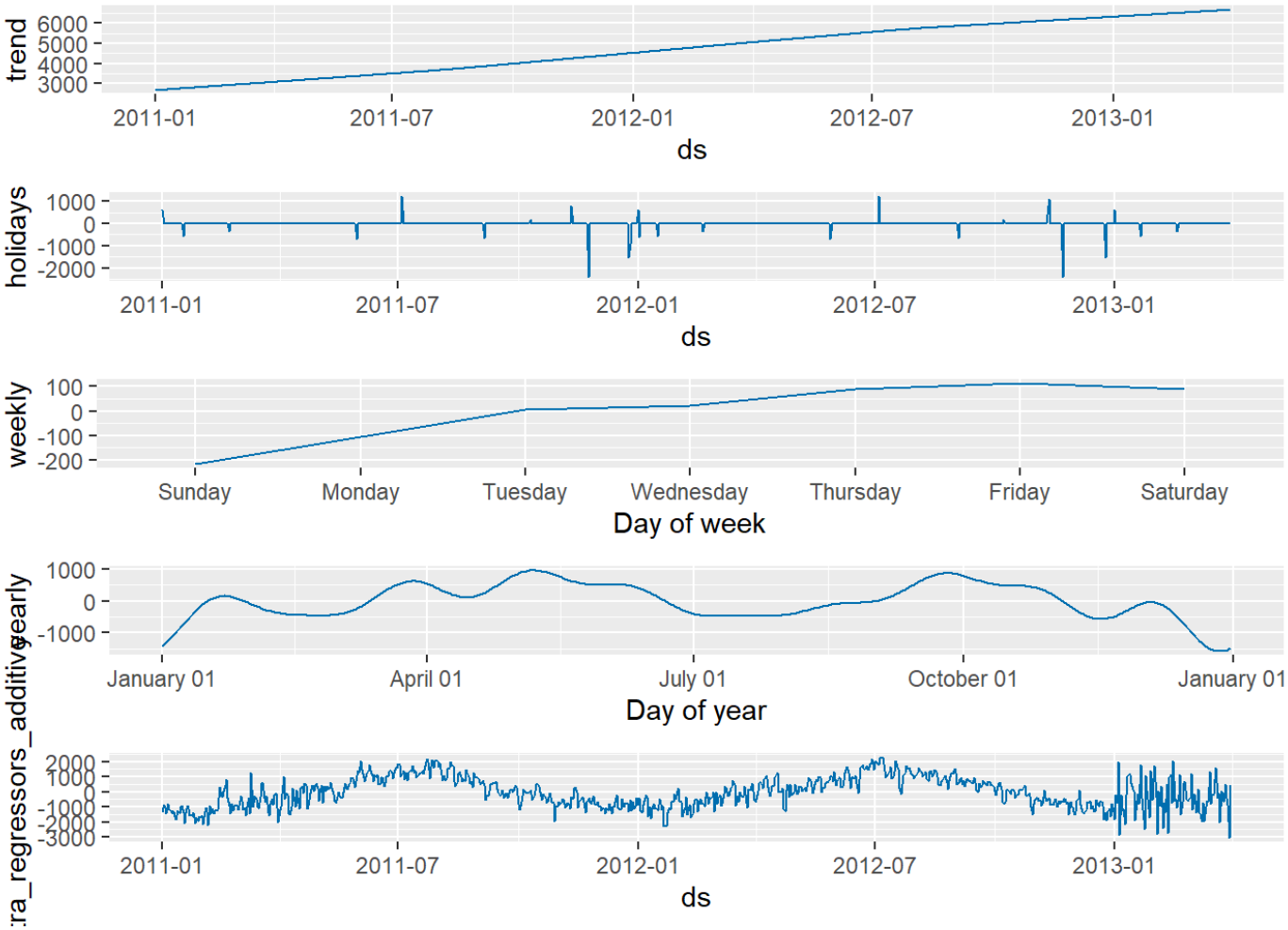
```
    Paired t-test

data:  perf_base$rmse and perf_enhanced$rmse
t = 16.101, df = 81, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
```

```
 115.3527 147.8825
sample estimates:
mean difference
       131.6176
```

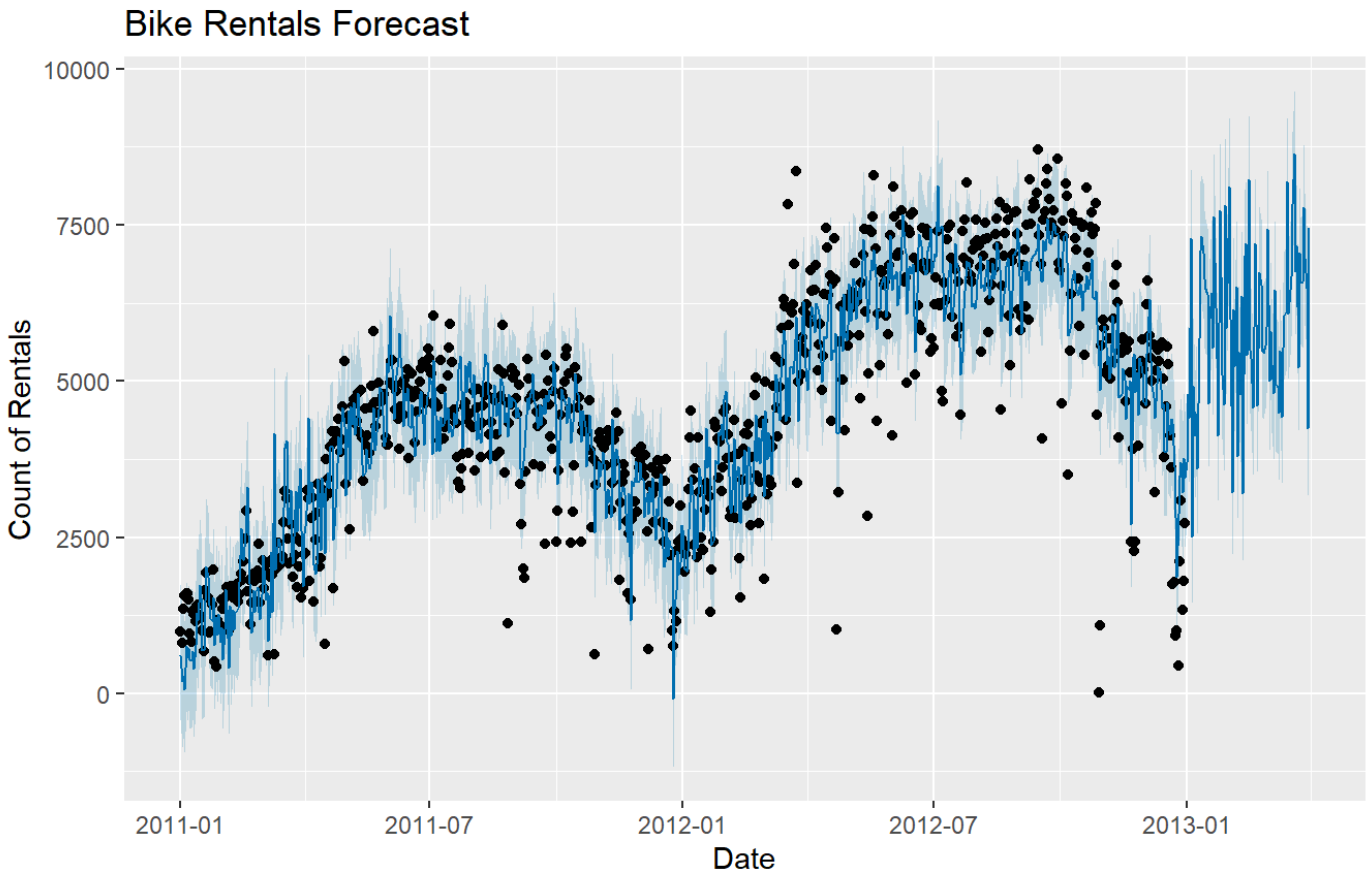
This supports the alternative hypothesis that these regressors do significantly improve the forecasting accuracy.



Component plots reveal an upward trend in rentals, weekly seasonality that appears to peak on Fridays, yearly seasonality with low rentals in January and highs in April, May, and October, and Holiday effects such as the consistent reduction in rentals during December.



The 90 days forecast appears to capture similar patterns and trends that are displayed in the historical data.



Limitations

A key limitation is the use of simulated temperature and humidity values for the 90-day forecast, that was simulated from random sampling of historical winter data through January to March. These values may not reflect actual weather conditions, potentially reducing forecast reliability. Prophet's requirement for known or forecasted regressor values for future dates adds complexity, as future weather data must be estimated (Facebook, n.d. -b). Additionally, the dataset lacks financial data like the costs of rentals limiting insights into pricing impacts on rental demand.

Proposed actions

Capital Bikeshare should integrate weather forecasts and holiday calendars into their inventory management strategies to optimize bike allocation. Real time weather data and holiday schedules can inform daily and seasonal planning, ensuring sufficient bike availability during peak demand periods and holidays.

Expected benefits

Implementing this enhanced model's forecast data into Capital Bikeshare's inventory allocation strategies would ensure that bikes are distributed more efficiently and that demand is being met throughout the year, particularly on holidays and time when weather is predicted to be extreme.

Sources

- Fanaee-T, H. (2013). Bike Sharing Dataset. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Bike+Sharing
- Facebook. (n.d. - a). Quick start. Prophet Documentation. https://facebook.github.io/prophet/docs/quick_start.html
- Facebook. (n.d. - b). Seasonality, Holiday Effects, And Regressors, Prophet Documentation. https://facebook.github.io/prophet/docs/seasonality%2C_holiday_effects%2C_and_regressors.html
- RoadGenius. (2024). Top 20 Most Visited Cities in the U.S. https://roadgenius.com/most-visited-cities-us/