# Forecasting Bikeshare usage with Gradient Boosting regression

by carlie de boer (she/her)

Oct. 2021

---

***Table of Contents***

# Intro

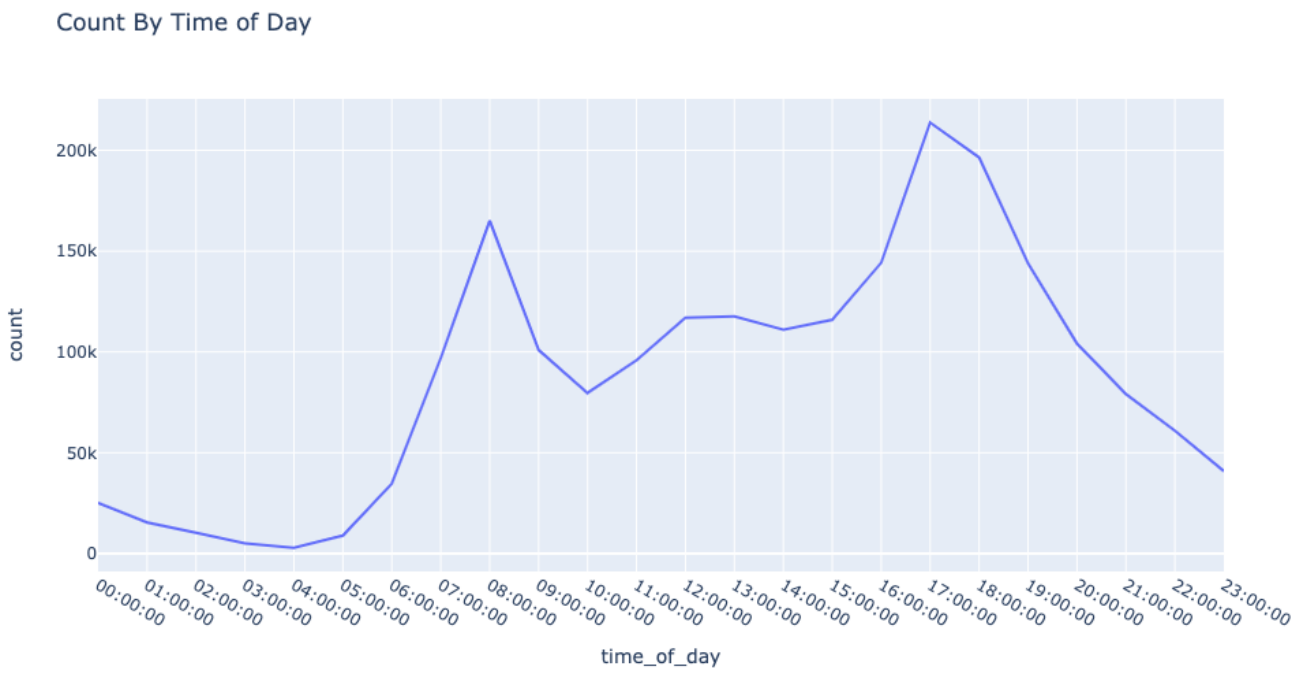
As a resident of a major city and a proponent of sustainability, I'm interested in analyzing this Bike Share dataset from Capital Bikeshare program in Washington, D.C. Bike sharing systems function as a sensor network, which can be used for studying mobility in a city and forecasting demand.

In this project, I'll combine historical usage patterns with weather data in order to forecast bike rental demand using GBM.

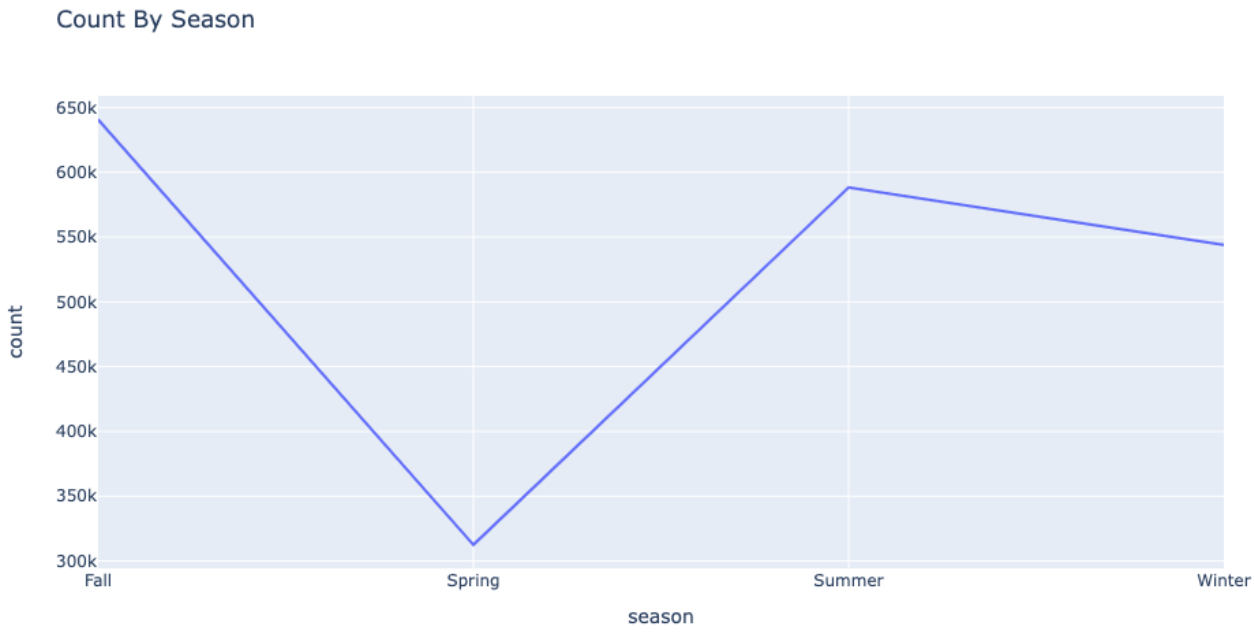
*"How do time based characteristics like time of day and previously occurring trends impact the ability to forecast total rentals? What about the weather?"*

# EDA

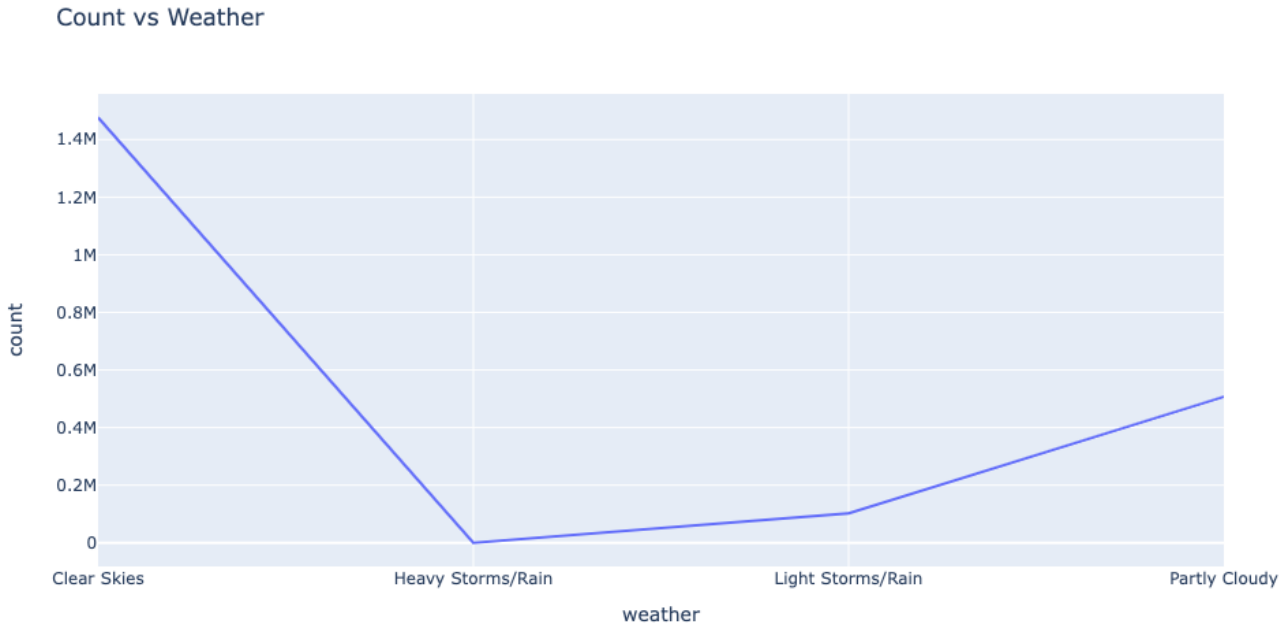
**Line Plots**
Rentals (Count) are most popular around 8 AM and 5 PM, perhaps for commuting to and from work.



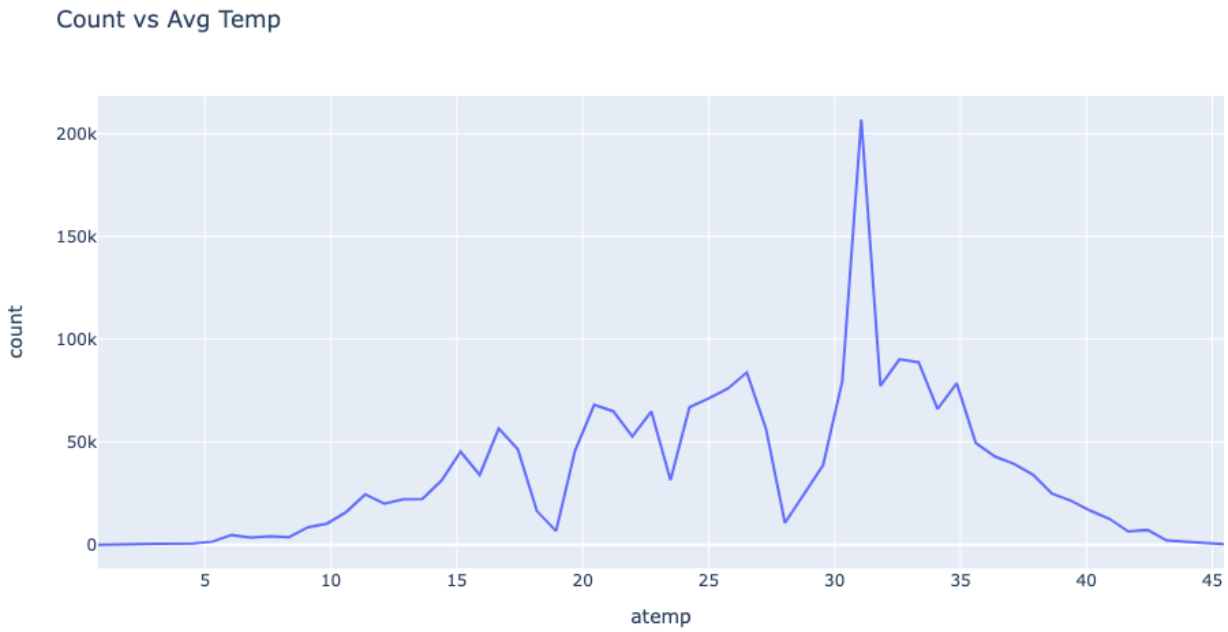
Rentals (Count) are lowest in Spring, highest in Fall.



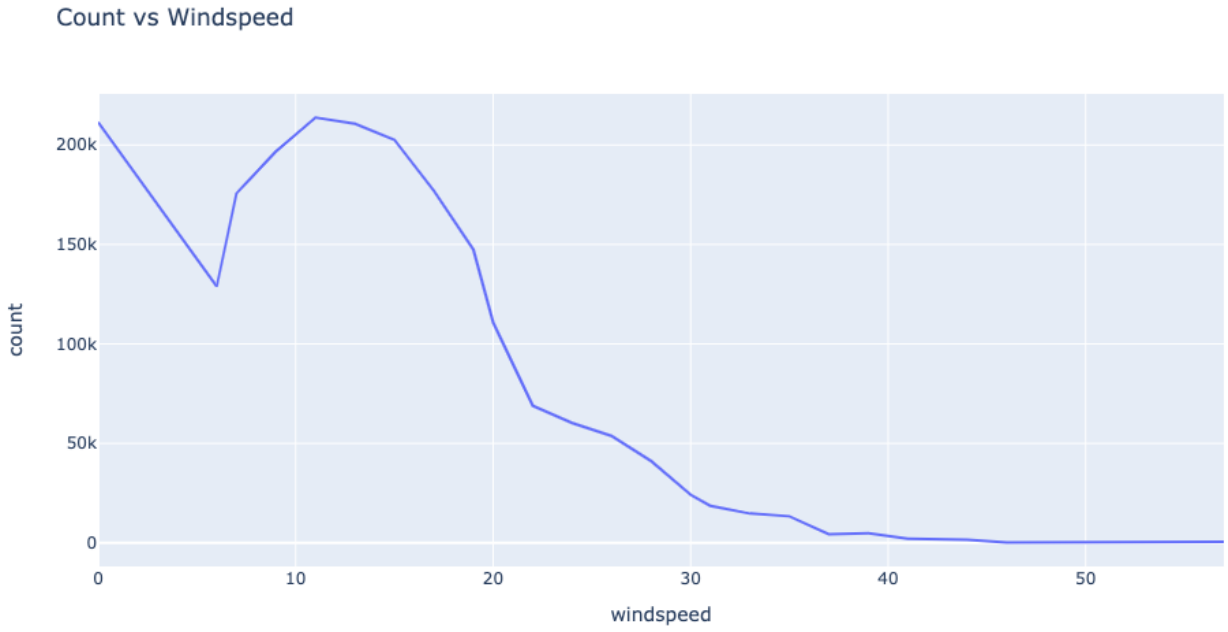

Rentals (Count) are higher when it's not raining, for obvious reasons!



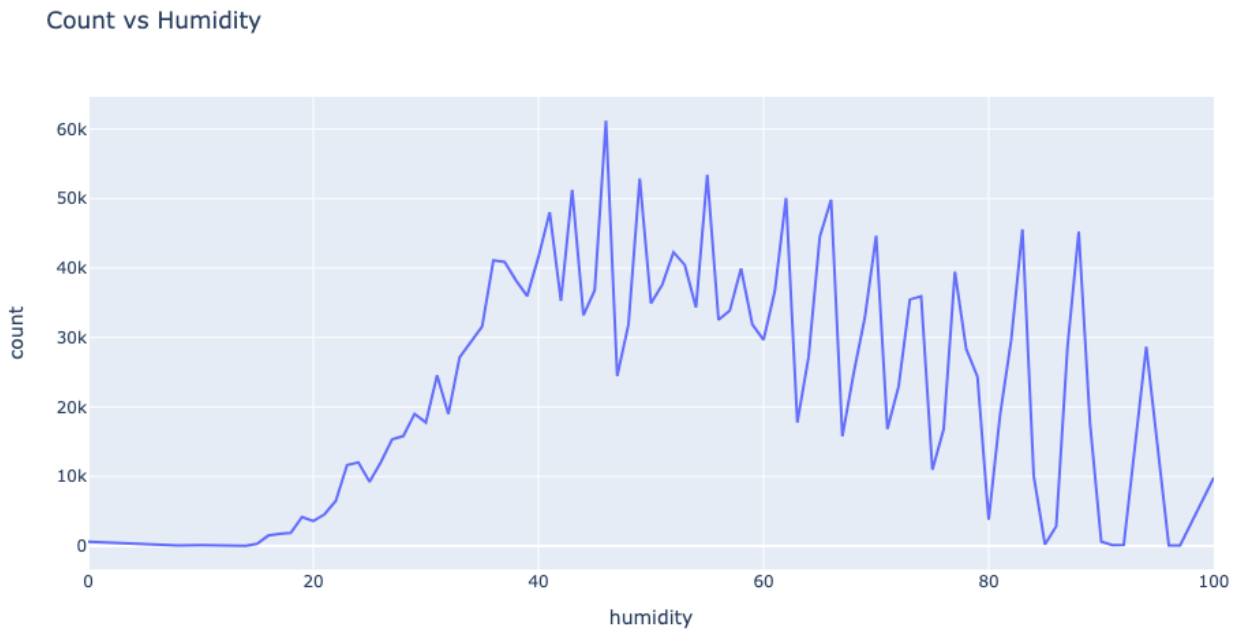
Rentals (Count) are highest at around 30 degrees Avg Temp.




Rentals (Count) go down drastically after 20+ for Windspeed.



There doesn't seem to be a correlation between Count vs Humidity.




**Challenges**
The dataset only includes days 1-18 of the month and data for two years.

There's also a lot of TimeSeries/Time-centric data, so it will be the centerpiece.

# ML

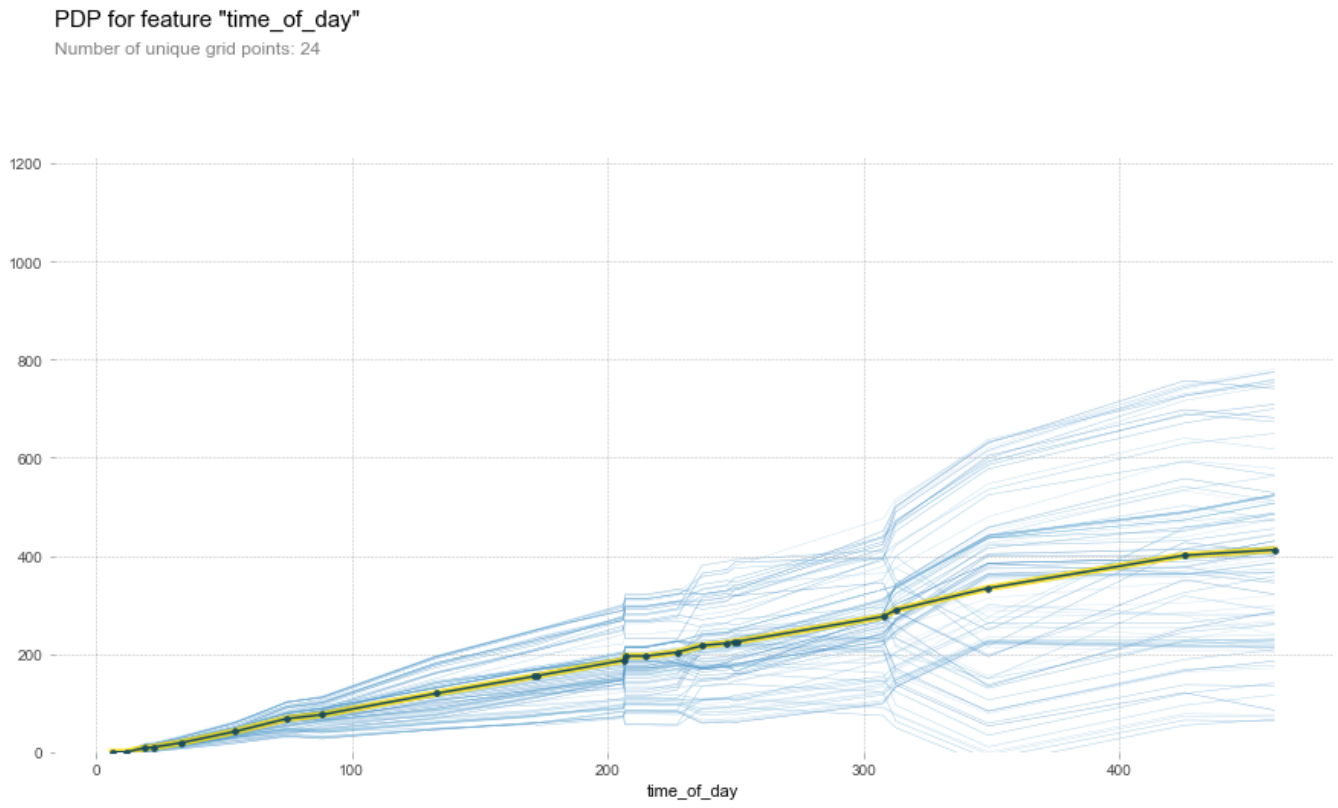
Using Ordinal Encoding, I'll run a Gradient Boosting Regressor model to predict which variables forecast bike rental demand.

I'll also validate by (1) adding more TimeSeries variables that repeat, (2) doing more boosting rounds, and (3) changing model parameters.

Time of Day and The Passing of Time are contributing as well as Working Day, Temperature and Average Temperature.

**PDP Plot**

There's a linear relationship between Time of Day and Count, which means our model is fit accurately.



# Conclusion

Overall, Time of Day and The Passing of Time are contributing to the model the most. It also seems the predominant usage is during weekdays, which is logical because Washington, D.C is a commuter city. I thought weather conditions would matter more. This may be for a variety of reasons:

- Users don’t change their mode of transportation based on weather conditions, especially for commuting
- Washington, D.C has a bike-friendly infrastructure

I hope that analysis of dataset like this can help cities and private institutions plan better for more sustainable models of transportation.

*“"There's no such thing as bad weather, only inappropriate clothing.”*

*- Sir Ranulph Fiennes*