Forecasting Bikesharing Usage for DC's Capital Bikshare System

Table of Contents

Introduction

Bikesharing systems are an increasingly popular solution in major urban areas to increase the usage of bicycles as a mode of transport. The riding of bikes helps to improve the lives of both users, as well as non-users, as each bike trip potentially represents a trip that would otherwise have required a car. We used data from the DC Capital Bikeshare from 2011 to 2017.

The purpose of this analysis is to determine variables/factors that help estimate bike usage and develop a model that predicts the usage based on certain predictor variables.

Overview of Project

The idea of the project was to use different data acquisition, cleaning and modeling techniques to determine if we could forecast the bike usage. If we are able to model and forecast the bike usage, then it would allow the company to plan for the best route to increasing our fleet as well as expanding the available stations that are offered.

Some of the questions were: can we use the data to know when our usage is lower to repair bikes? When should we start increasing our fleet to best meet demand? Does weather and seasons have an impact on our business model?

We made an initial hypothesis that we would see a higher usage during the summer months and when the weather was nicer. With our initial exploratory data analysis, we could see that there is a trend and seasonality to the usage. The question then became, can we model this through a time series model. Also could we see what features were key factors in determining bike usage. Does the season or weather have an affect?

Overview of Data

Initial Data Set:

We started by just looking at two years of Capital Bikeshare usage from this dataset: https://archive.ics.uci.edu/dataset/275/bike+sharing+dataset

The dataset contains 2011 and 2012 historical usage data from Washington, DC's public Capital Bikeshare program, one of the first large scale bikeshare programs in the nation. Usage data is broken out by day and by hour. Additional data included a variety of information on weather, season, and whether a day was a holiday:

```
- holiday : whether day is holiday or not (extracted from http://dchr.dc.gov/page/holiday-sched
- weekday : day of the week
- workingday : if day is neither weekend nor holiday is 1, otherwise is 0.
+ weathersit :
 - 1: Clear, Few clouds, Partly cloudy, Partly cloudy
 - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
 - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
 - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp : Normalized temperature in Celsius. The values are divided to 41 (max)
- atemp: Normalized feeling temperature in Celsius. The values are divided to 50 (max)
- hum: Normalized humidity. The values are divided to 100 (max)
- windspeed: Normalized wind speed. The values are divided to 67 (max)
- casual: count of casual users
- registered: count of registered users
- cnt: count of total rental bikes including both casual and registered
```

Data Cleaning Process

The dataset had required minimal cleaning. We had to convert several variables into factor variables (`season`, `holiday`, `weekday`, `workingday`, `weather`). Additionally we noted that the key for our dataset mislabeled the season variable, which was trivial to correct. Fortunately, there was no missing data.

Additional Scraping, Cleaning

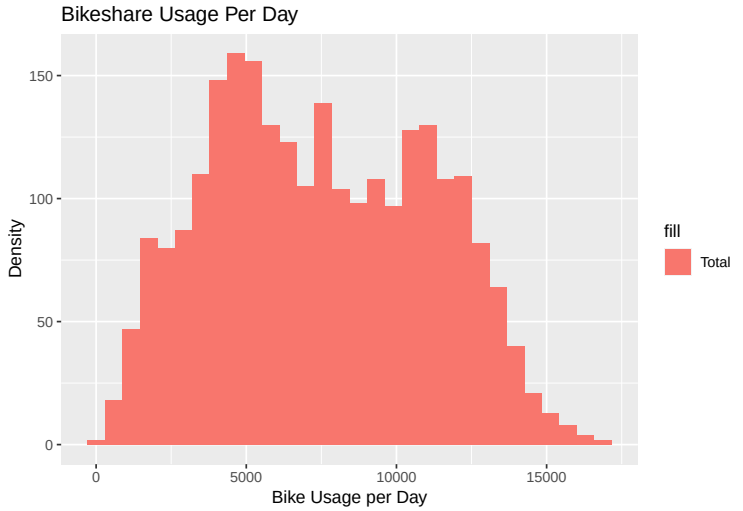
Initial EDA on the dataset we began with indicated we did not have sufficient data to fit any models that could improve upon linear regression. So, we scraped and combined two different datasets. We used python scripts located in the Other Resources directory to do this. Specifically, we ran get_bikeshare_data.py to get bikeshare data, then we ran join_data.py (which imports from noaa.py) to join the data with weather data from NOAA. We used the same process to add the additional weather columns to the 2011 and 2012 as well.

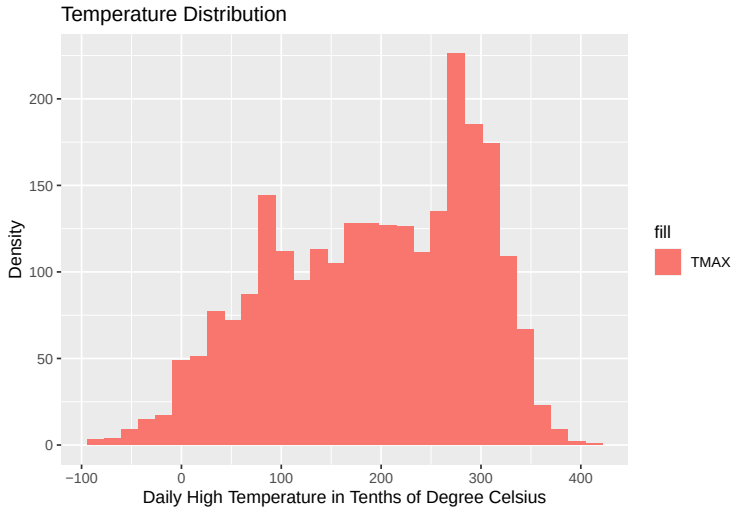
- Capital Bikeshare usage data from 01-01-2013 through 12-31-2017, from Capital Bikeshare (https://ride.capitalbikeshare.com/system-data)
- Weather data for DC for the same time period, from NOAA (https://www.ncdc.noaa.gov/cdo-web/webservices/v2)

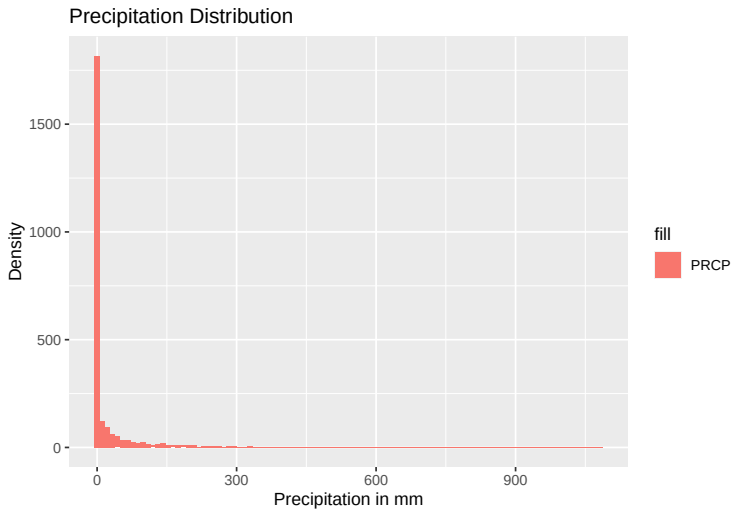
The Bikeshare data only counted trips taken, and did not distinguish between registered or casual users.

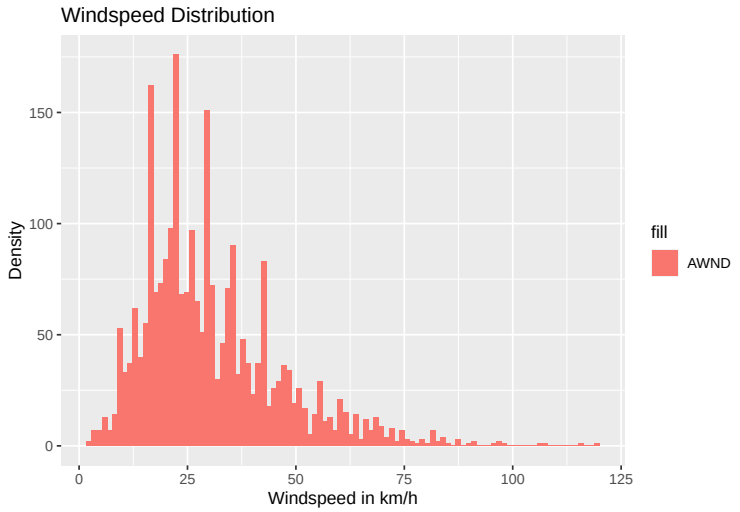
The Weather data included the day's high temperature (in tenths of degree Celsius), precipitation (tenths of mm), and average daily windspeed (km/h).

Exploratory Data Analysis

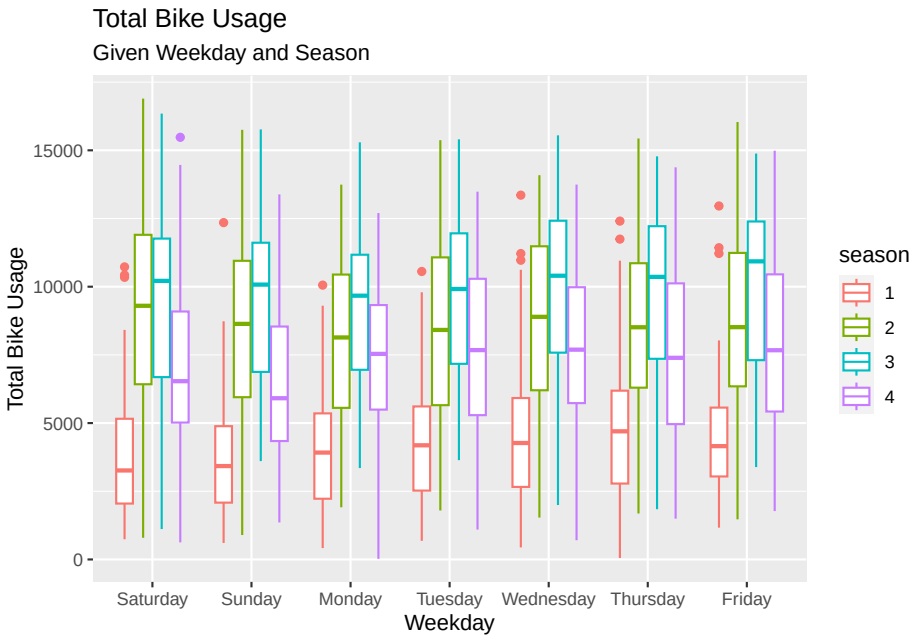






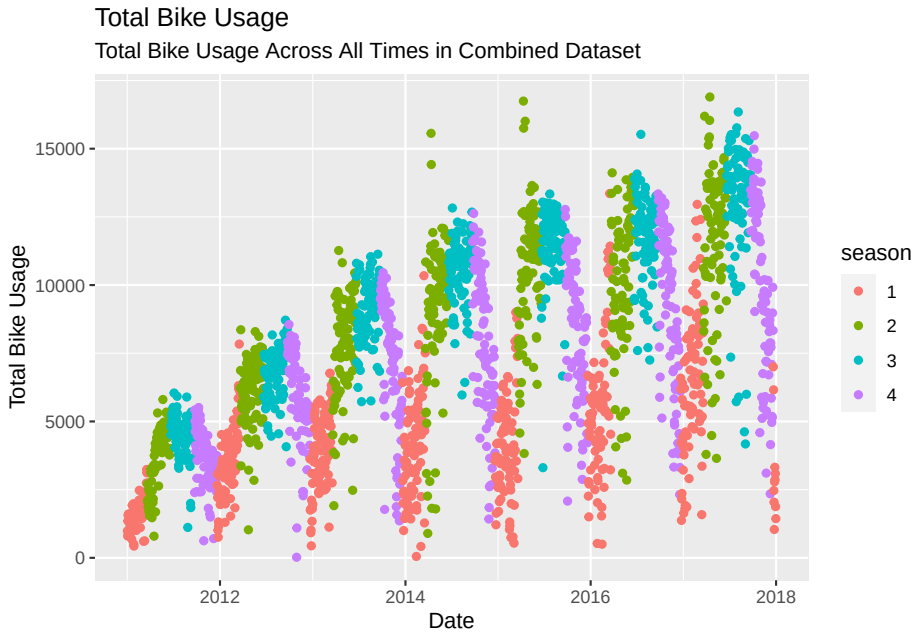


We see that windspeed and precipitation exhibit rightward skew, which makes sense: * Most days have no precipitation * Most days have moderate windspeed, while a few have very high windspeed



We can see in the above chart the effect of season (1 = Winter, 2 = Spring, 3 = Summer, 4= Fall) as well as the effects of the day of the week on total bikeshare usage. Usage is lowest in the winter, highest in the summer, with spring and fall in between.

Fortunately, there is not a huge amount of variation in usage between weekdays and weekends.



The most challenging aspect of our dataset is that usage was not static - it grew from 2011 through 2017. While this was obviously good for Capital Bikeshare, it meant that models would need to take into account not just variation within a year, but overall increasing usage over time.

Key Predictors

Overview of Modeling

Model Types and Comparison Model Selection and Optimization Model Performance Detailed evaluation and interpretation of results

#Conclusion Overall conclusion and key learnings

Works Cited

Bean, R., Pojani, D., & Corcoran, J. (2021). How does weather affect bikeshare use? A comparative analysis of forty cities across climate zones. *Journal of Transport Geography*, 95. https://doi.org/10.1016/j.jtrangeo.2021.103155.

Eren, E., & Uz, V. E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. *Sustainable Cities and Society*, 54. https://doi.org/10.1016/j.scs.2019.101882

Ashgar, H. I., Elhenawy, M., & Rakha, H. A. (2019). Modeling bike counts in a bike-sharing system considering the effect of weather conditions. *Case Studies on Transport Policy*, 7(2), 261-268. https://doi.org/10.1016/j.cstp.2019.02.011