Capital Bike Sharing Data

Lukas Ronkin • Sarah Kaunitz • Sung Hyuk Lee

Introduction to Our Data

- Background
  - Daily count of rental bikes in Capital Bikeshare, Washington DC
  - System designed for shared use
  - Convenient and efficient
  - UCLA introduced system recently
- Data Collection
  - Hadi Fanaee-T, LIAAD in University of Porto, Portugal
  - 2011 - 2012
  - Came directly from Capital Bikeshare
  - Weather and holiday schedules collected separately.

What we are looking for

Guiding Questions:

- How do environmental and seasonal settings impact the rental behavior of bikeshare users?
- Can we predict the amount of bikers for a given day?

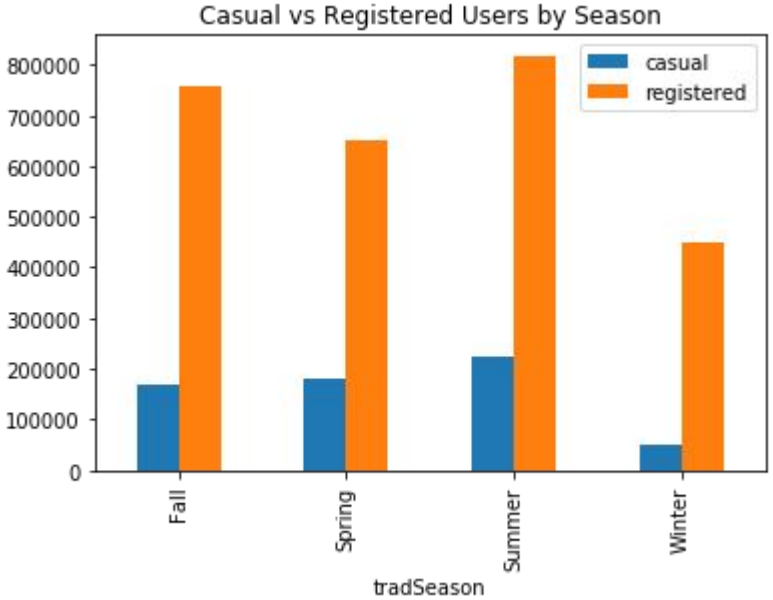
Statistical Python Methods Used:

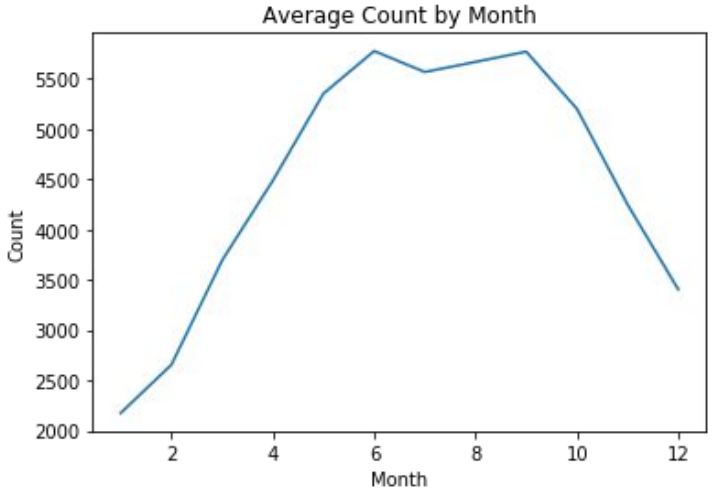
- Created new variable
- Plots (scatter, box, histogram, line, bar)
- Linear regression
- Multiple linear regression

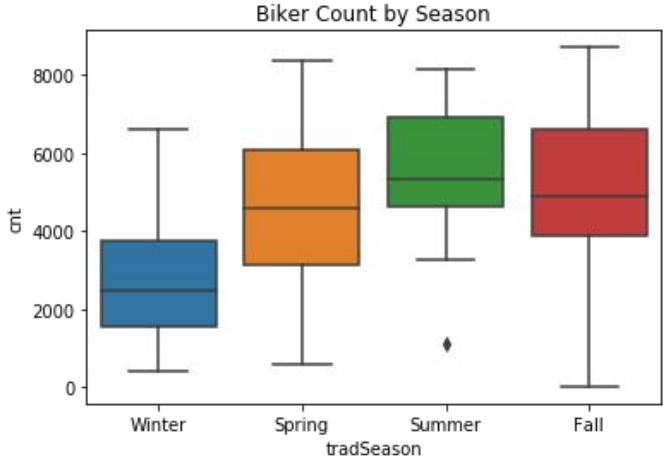
Data:

- Date, seasons, different weather conditions, casual users, registered users, total count of bikers

Exploratory Data Analysis

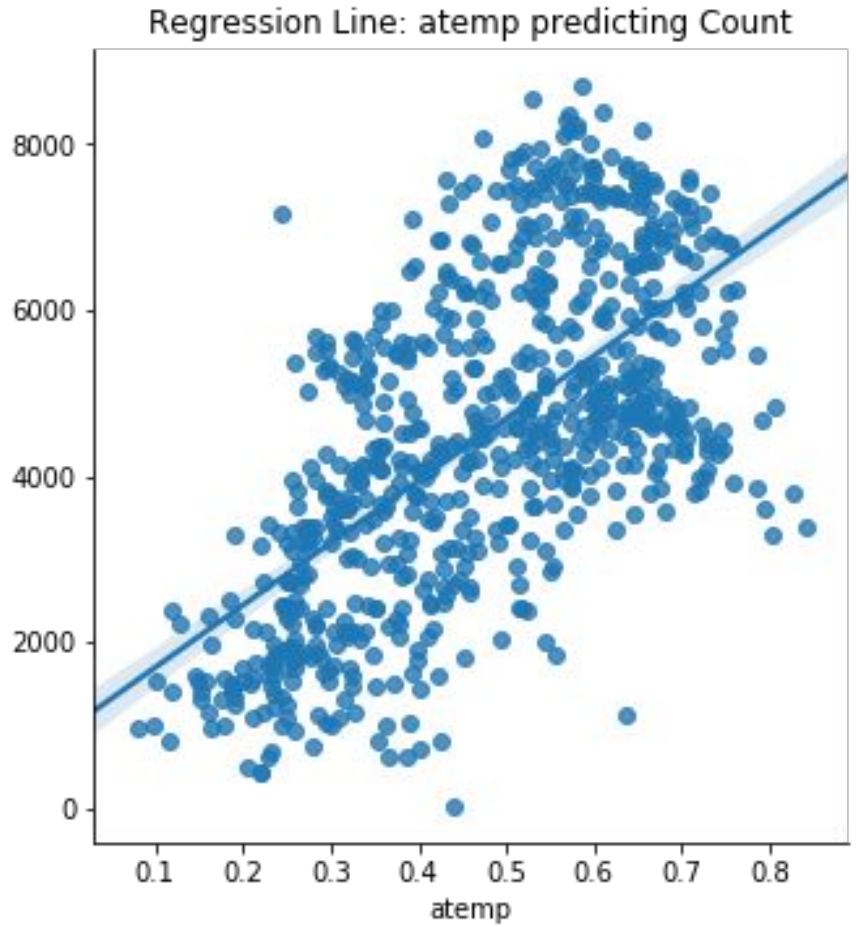






Data Modeling

- Prediction of daily bike rental users based on the environmental and seasonal settings using Linear Regression.
- 3 Models
  - MODEL 1: Expected Count ~ atemp
  - MODEL 2: Expected Count ~ atemp + humidity
  - MODEL 3: Expected Count ~ atemp + humidity + windspeed
- Strong positive linear correlation between temperature and count
- Found that adding humidity and windspeed does increase R^2 and the predictors are statistically significant



Conclusion

- Our analysis shows that “feels like temperature” was the most significant predictor of daily bikeshare users, humidity and windspeed also play a factor.
- Regardless of seasonal settings, bikeshare interest increases with temperature.
    - For example, we might expect a low bikeshare turnout for cold Summer days but a high turnout for warm Winter days.
- We might expect the bikeshare systems in warm places such as UCLA to succeed because of less variability in temperature.