DC Capital Bikeshare Demand Prediction

Smart planning today,
better community tomorrow

Problem statement

Vision: *Enhance* the utilization of DC capital bikeshare by predicting the usage of bikes ahead of time. This will benefit the community through:

- Help tourists/residents plan for transportation ahead of time
- Inform DC capital bikeshare on bike occupancy for better resource allocation

Demo

Predict capital bikeshare

- **<=25%: worry free on missing your bike!**
- **25% ~ 60%: your bike should be available but don’t wait till last minute.**
- **>61%: leave earlier to get your bike before it’s gone!**

Data description

UCI DC bikeshare data

- 2011 – 2012 hourly data (17,379 rows * 18 cols)
  - Train data: first 19 days of each month
  - Test data: the 20th day and forward of each month
- Predictors: 16 predictors on information about the day/hour
- Response variable: hourly count of total bike rentals

Model outputs:

- percentage of the capital bike occupancy

Final model and score

RMSLE Scorer

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i+1)-\log(a_i+1))^2}$$

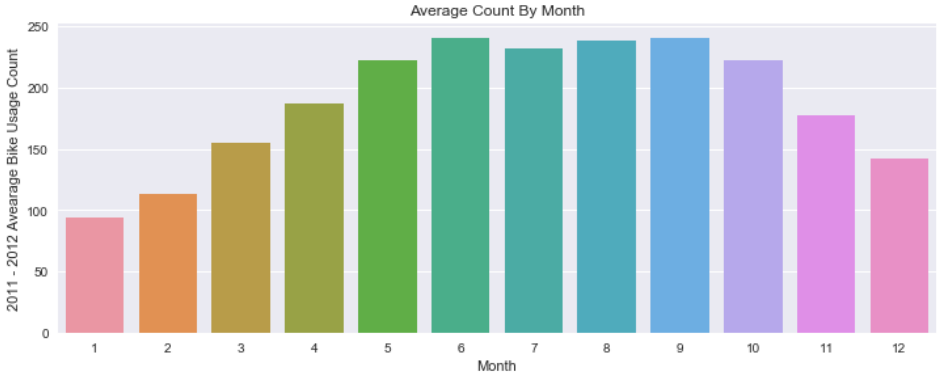
Where:

- n is the number of hours in the test set
- p_i is your predicted count
- a_i is the actual count
- $\log(x)$ is the natural logarithm

RMSLE Scores

	Train	Test
Linear Regression	1.0	1.018
Random Forest	0.138	0.493

EDA discovery on bike usage by month & season

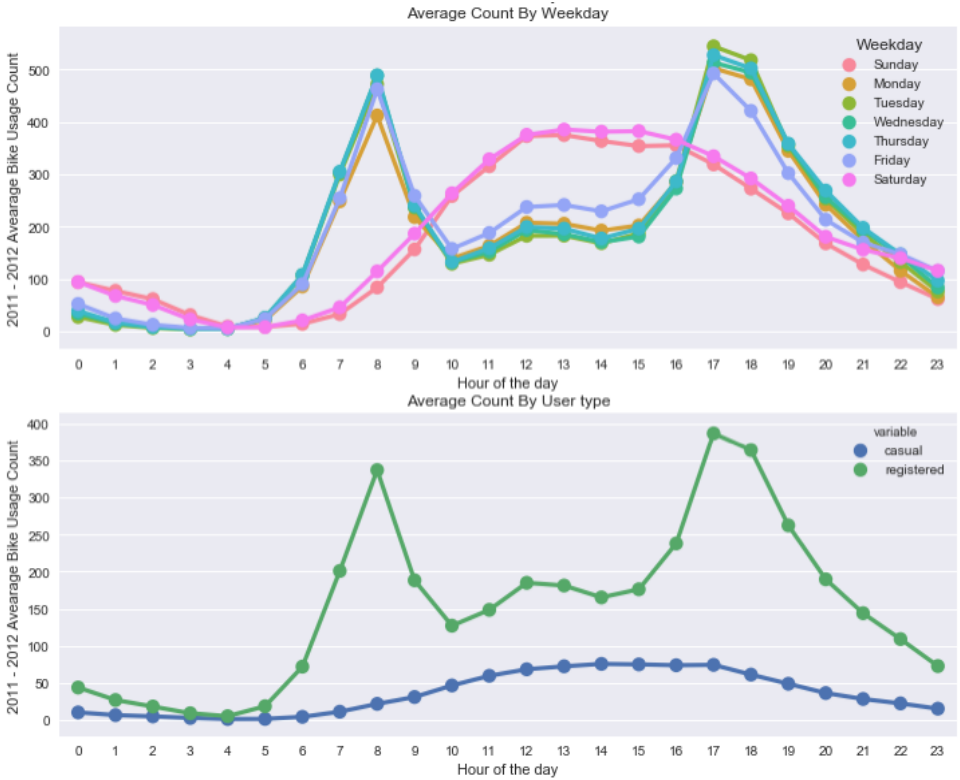


On average, bike usage is highest during June to September.

More people use capital bikes in fall and summer than winter and spring.

The average hourly usage is consistent in each season.

EDA discovery on weekday commuting patterns



Higher usage of bikes during morning and evening commute times on weekdays whereas on weekends, demands for bikes start to grow at 10am till 4pm.

Registered bikeshare users use the bikes more than casual users and their usage patterns align with commuting hours expectation.

Thank you!

Jamie Chen
jamie.chen@u.northwestern.edu

About Jamie:

- Curious data scientist who wants to bring more data-driven insights to make an impact!
- Lived at many places:
  - Shenyang, China
  - Blue Springs, MO
  - Williamsburg, VA
  - Palo Alto, CA
  - Washington, D.C.
  - Evanston, IL
- Hobbies: running, swimming, hiking, music, cooking

Appendix - Webapp looks

Welcome to DC bikeshare usage prediction

A prediction model was built using historical DC bikeshare data to predict the usage of bikeshare in a given hour to help you plan for your transporation methods accordingly and inform the district on bikeshare demands for better resource relocation.

Please enter the following information below:

Select a month:
April

Please enter hour of the day:
8 am

Please indicate whether it is a weekday:
Wednesday

Please select the whether condition:
Clear, Few clouds, Partly cloudy, Partly cloudy

Please enter the temperature (°C):
27

Please enter the humidity level (%):
7

Please enter the windspeed level (km/h):
5

Predict demand

Home Realtime More About Capital Bikeshare

Prediction

Here's the predicted bikeshare occupancy based on your input:"

39.56%

Appendix - Predictors

2.0.3 Variables:

- instant: record index
- dteday : date
- season : season (1:spring, 2:summer, 3:fall, 4:winter)
- yr : year (0: 2011, 1:2012)
- mnth : month (1 to 12)
- hr : hour (0 to 23)
- holiday : weather day is holiday or not (extracted from http://dchr.dc.gov/page/holiday-schedule)
- 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