London Bike Share Data Analysis

Caroline Sklaver & Sayra Moore

Description of Dataset

"cnt" - the count of a new bike shares

"is_holiday" - boolean field - 1 holiday / 0 non holiday

"is_weekend" - boolean field - 1 if the day is weekend

"season" - category field meteorological seasons: 0-spring ; 1-summer; 2-fall; 3-winter.

"timestamp" - timestamp field for grouping the data

"t1" - real temperature in C

"t2" - temperature in C "feels like"

"hum" - humidity in percentage

"wind_speed" - wind speed in km/h

"weather_code" - category of the weather

"is_holiday" - boolean field - 1 holiday / 0 non holiday

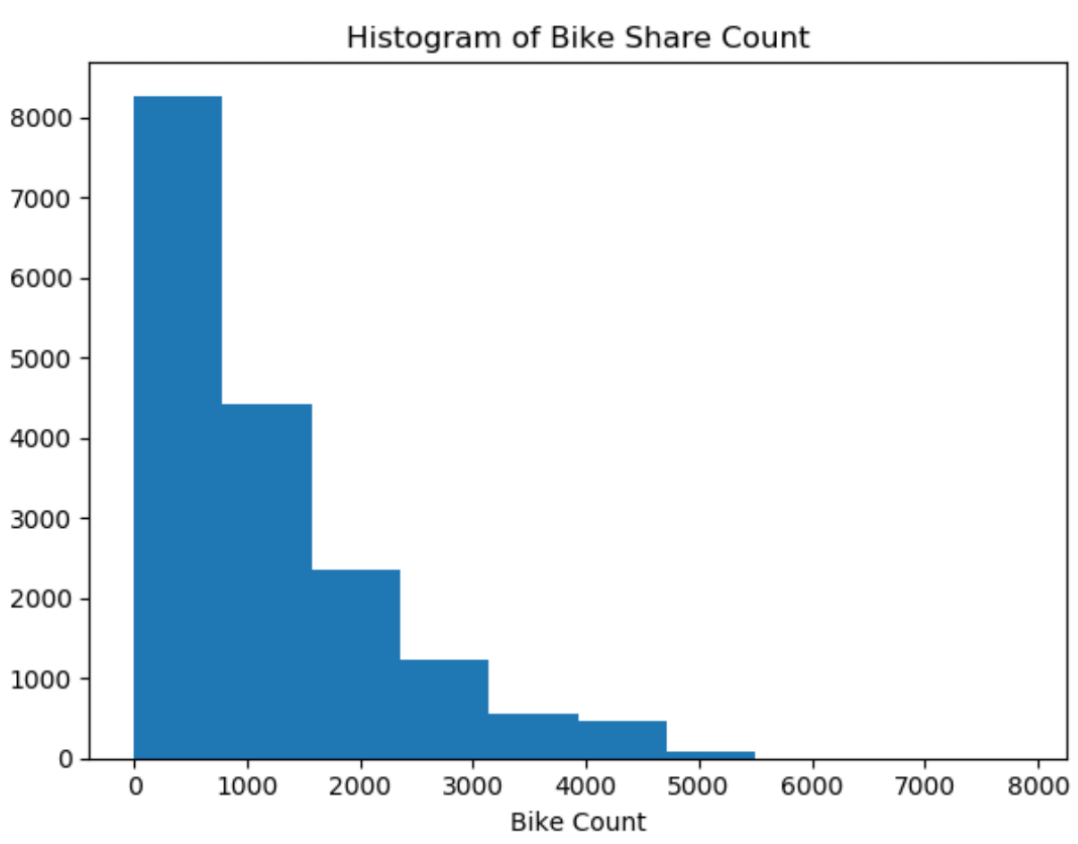
"is_weekend" - boolean field - 1 if the day is weekend

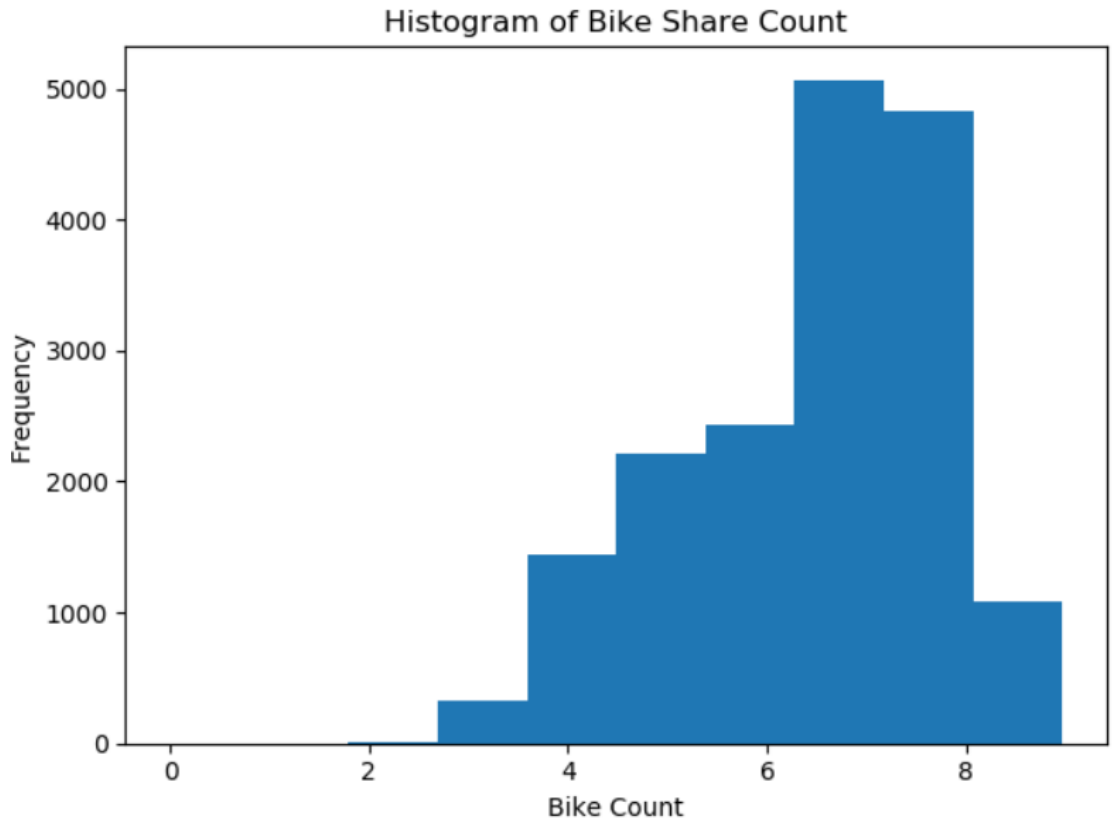
"season" - category field meteorological seasons: 0-spring ; 1-summer; 2-fall; 3-winter.

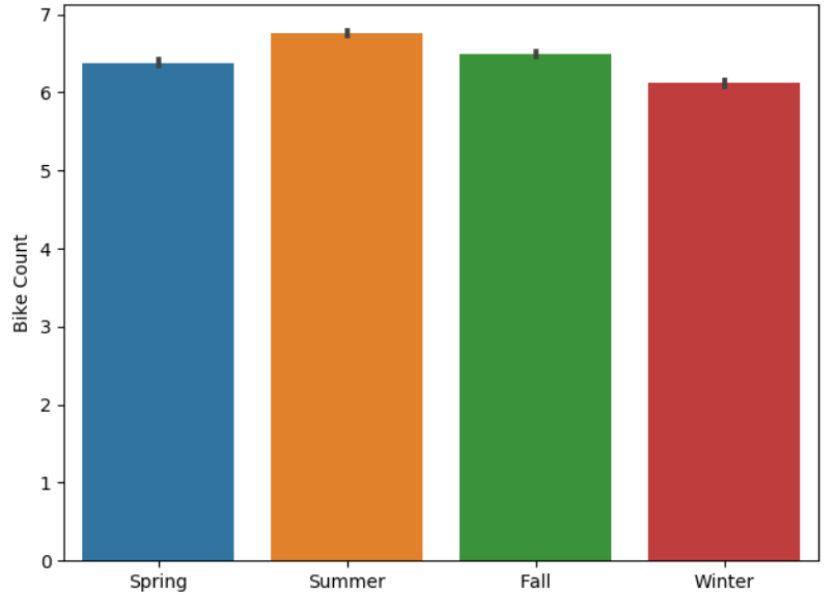
Descriptive Statistics

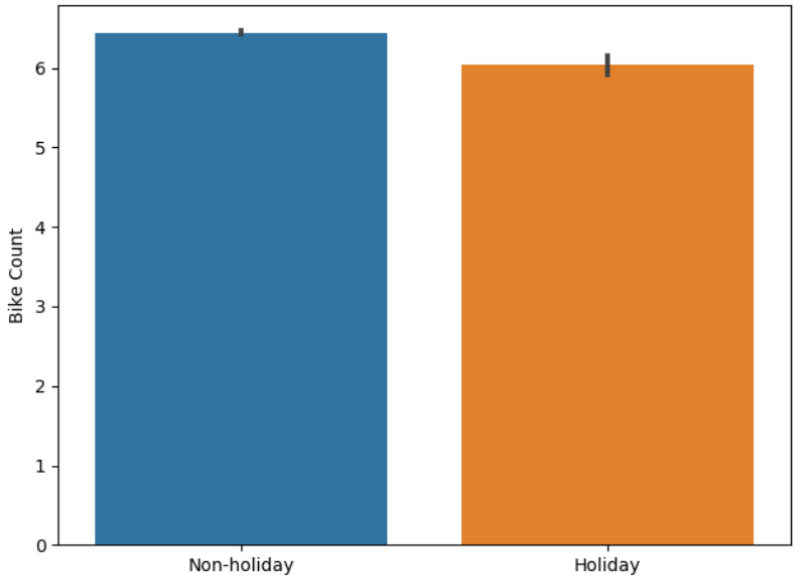
Bike Share Count Distribution

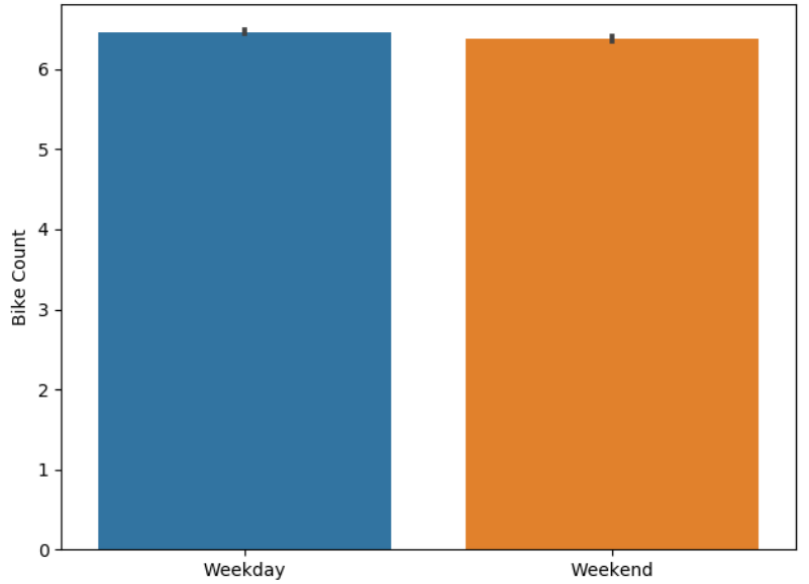
RIght-skewed



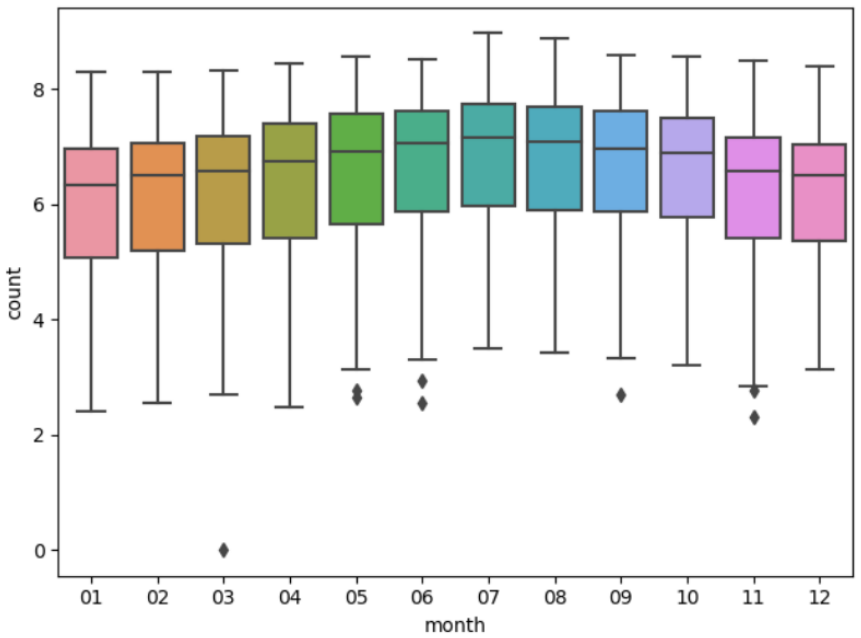


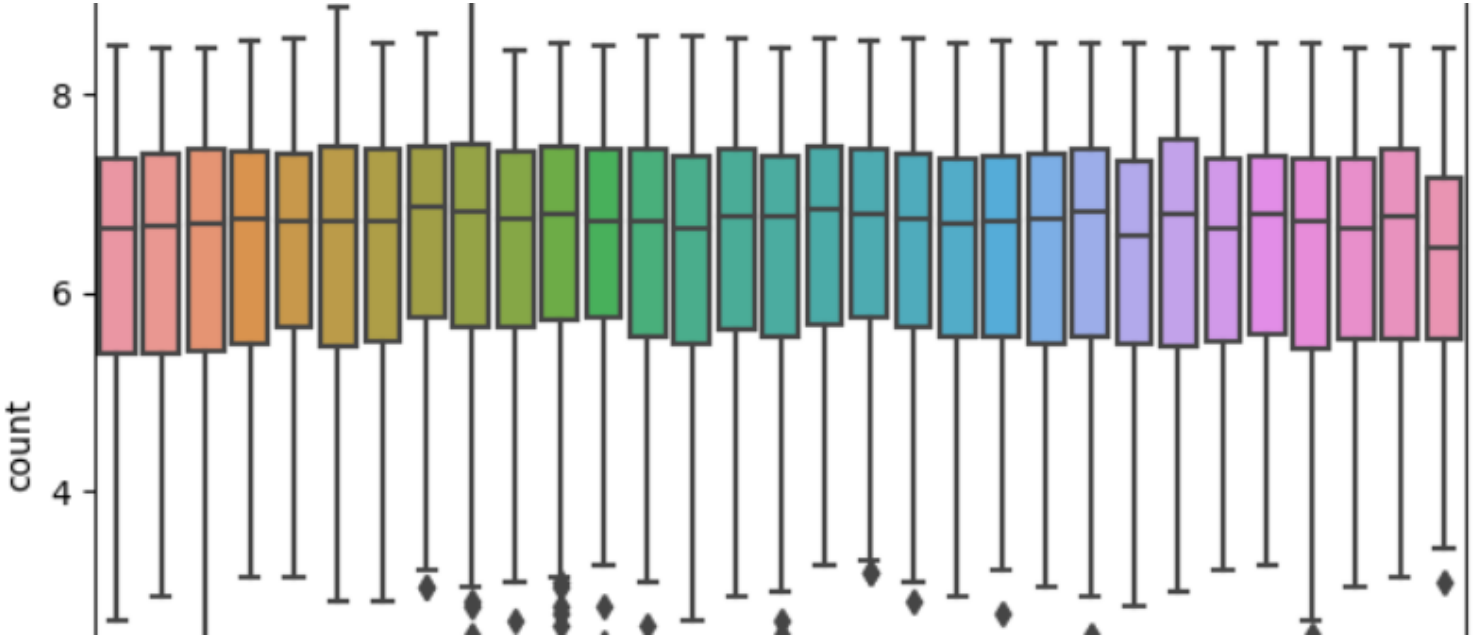


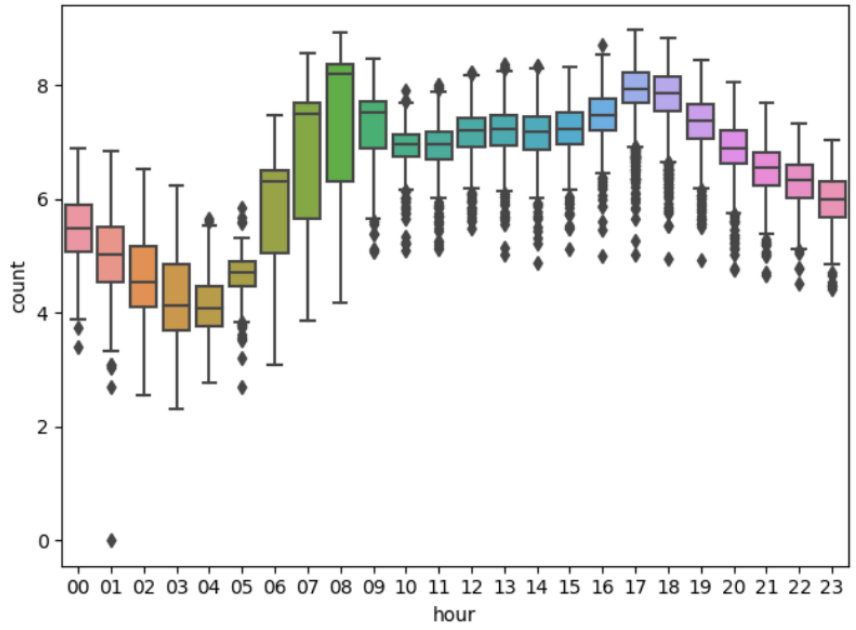




Season, Holiday, Weekend



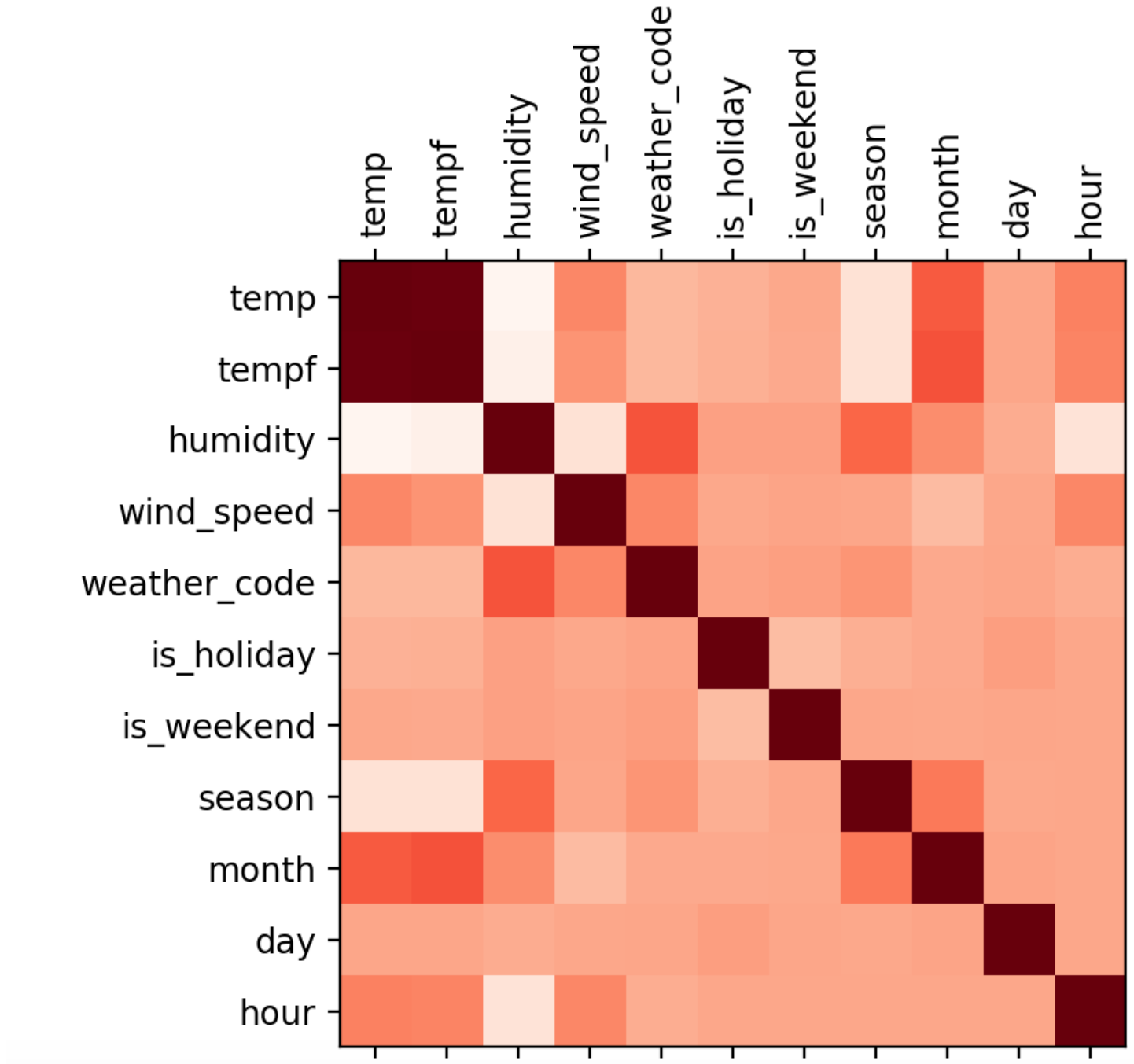




Month, Day, & Hour

Correlation Plot

- Temperature & Temperature "feels"
- Humidity & Weather code, Season
- Month & Temperature



Description of Model

Simple Linear Regression

Multiple Linear regression

Simple Linear Regression

$$y = b_0 + b_1 * x_1$$

Multiple Linear Regression

Dependent variable (DV) Independent variables (IVs)

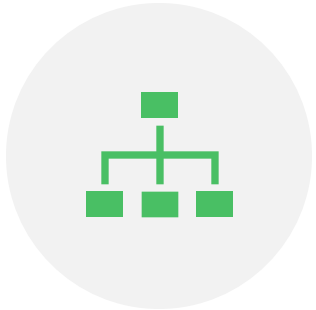
$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \ldots + b_n * x_n$$

Demo: Graphical User Interface

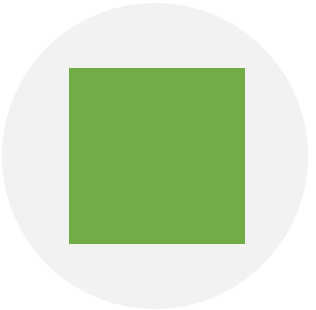
Conclusions

THE 1ST TIME RUNNING THE MODEL WE GOT A 30.7% R-SQUARED

THE 2ND TIME RUNNING THE MODEL WE GOT A 44.8% R-SQUARED



WE COULD NOT RUN CLASSIFICATION MODELS BECAUSE OUR TARGET WAS NOT BINARY



OVERALL, NOT A GREAT MODEL IN PREDICTING COUNT OF BIKE SHARES

References

1. Bogicevic, M. (2018, November 2). Multiple Linear Regression Using Python. Retrieved from https://medium.com/@manjabogicevic/multiple-linear-regression-using-python-b99754591ac0.
2. Columbia Mailman School of Public Health. (2019, June 10). Tutorial: Understanding Linear Regression and Regression Error Metrics. Retrieved from https://www.dataquest.io/blog/understanding-regression-error-metrics/.
3. Evaluating a Linear Regression Model. (n.d.). Retrieved from https://www.ritchieng.com/machine-learning-evaluate-linear-regression-model/.
4. freeCodeCamp.org. (2018, October 8). Machine learning: an introduction to mean squared error and regression lines. Retrieved from https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/.
5. Mavrodiev, H. (2019, October 10). London bike sharing dataset. Retrieved from https://www.kaggle.com/hmavrodiev/london-bike-sharing-dataset.
6. Menon, A. (2018, September 13). Linear Regression Using Least Squares. Retrieved from https://towardsdatascience.com/linear-regression-using-least-squares-a4c3456e8570.
7. Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.
8. Stanford, M. (2017, December 20). Simple Linear Regression in Python. Retrieved from https://medium.com/@mjfstanford/simple-linear-regression-in-python-905b759ef0e6.