Introduction:

The goal of analyzing the bike sharing dataset is to understand the behavior customers and find the significant factors which drive the decision of the customers to ride the bike or not. Application of this analysis will be to predict the number of rides on any given day based on the significant predictors.

The predictors in this dataset include following categorical predictors, Year, Season, Working/Non-Working Day, Weather Situation, and following numerical predictors Temperature, Humidity, and Windspeed.

We will fit various regression models such as Linear, K-fold, Ridge and Lasso, PCR, and Random Forest, on both Casual and Registered riders using the above-mentioned predictors to successfully predict the number of riders based on best MSE estimate and adjusted R2. The final model or the set of models will be chosen, based on those estimates.