HW 9 - Modeling Practice

This homework is meant to give you a chance to do some structured practice with fitting models using the `tidymodels` framework.

We'll continue with the homework 8 data set and analysis you already did!

Data

We will use a dataset from the UCI Machine Learning Repository. This data set is about bike sharing rentals and is available at the assignment link. You can learn more about the data here. The data is available at https://www4.stat.ncsu.edu/~online/datasets/SeoulBikeData.csv

The data description describes the following variables:

- Date : **day/month/year**
- Rented Bike count - Count of bikes rented at each hour
- Hour - Hour of the day
- Temperature-Temperature in Celsius
- Humidity - %
- Windspeed - m/s
- Visibility - 10m
- Dew point temperature - Celsius
- Solar radiation - MJ/m2
- Rainfall - mm
- Snowfall - cm
- Seasons - Winter, Spring, Summer, Autumn
- Holiday - Holiday/No holiday
- Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

To Do:

Add to your previous document where you fit multiple linear regression models using 10 fold cross-validation. You should add (using the first recipe from homework 8 with no quadratics nor interactions):

- a (tuned) LASSO model (note: I got all the same rmse values for every tuning value)
- a (tuned) Regression Tree model
- a (tuned) Bagged Tree model
- a (tuned) Random Forest model

Each of these models should be fit and tuned on the training set. You should take the best model from each family of models (best LASSO, best Regression tree,. . .) and fit it to the entire training data set and see how it predicts on the test set. Include your best MLR model from the last homework here.

- **Compare all final models on the test set using both rmse and mae (mean absolute error).**

- Extract the final model fits for each type and report a summary of each model
  - For the LASSO and MLR models, report the final coefficient tables
  - For the regression tree model, give a plot of the final fit
  - For the bagged tree and random forest models, produce a variable importance plot
    - For the random forest model, this is a bit complicated. Check this out and see if you can get it to work.

For the overall best model, fit it to the entire data set!