R code Instructions

**R Version:** 4.0.2

**Datasets:**  train.csv and test.csv

**Execution Time**: 3 hours

1. Make sure all the **included packages/libraries** in R file are installed.
2. The **datasets will be uploaded** as part of the Final Report submission. Change the code(line 75-76) to point to the correct path where you store the datasets.
3. Make sure that **plots window in the R studio is big enough** to accommodate the visualizations.
4. Ignore the warning messages.
5. You can run the code up to the Modeling section (line 1363) and then choose to run a specific model. Or you can execute the entire code and it will take around 3 hours to complete.
6. If you have any questions about the R code please send email to our team.
7. Xgboost model was run with kfold /5 repeats. But the code was commented out to reduce the execution time.

The R code is organized into following main sections:

* Import and Prepare Data for Visualizations
* Additional Data Preprocessing
* Data Exploration and Visualization
* Data partition into Training/Test sets
* Model selection & Tuning