Data Mining Report: Rossman Store Sales

Four files, one directory

Input

Report.pdf

Test Scores.png

R\_EBoosting\_Feature\_Experiments.R

Analytics knit.Rmd

Input –Kaggle Input data for the R code (Rossman Sales)

Report.pdf – The pdf of our final report

Test Scores.png – A picture of our test results on Kaggle as proof of the models’ performance and our ranking

R\_EBoosting\_Feature\_Experiments.R - Contains all the code for generating our experimental results. Also includes the code for generating feature importances and goodness fit of the model. Individual blocks are explained in the document. To train on different feature sets use one of the vectors of features, below the comment “CHOOSE FEATURE SET HERE “

Analytic kit.Rmd – Contains the code for generating figures tables used mostly in the presentation.

In order, the train, test feature value ratios, correlation heatmap, Facebook prophet trends, and mean sales grouped by assortment, dayofweek and promotion.

Note: to run make sure all the packages are installed in R studio if not already. This can be done by uncommenting the install.packages functions the beginning of the files.