**README:**

*Functions\_Dataholics.R*

Contains mainly own written functions. Also contains some of the functions written by Prof. Nick F Ryman-Tubb, presented to us in the computer labs.

*main.R*

Executes the program. Has each needed filename as a global variable. Reads in the functions file, installs the needed packages and executes the different parts of the program.

**You will be asked to enter your JAVA\_HOME path**. On Windows there was a slight issue regarding the file path. Changing the path to “C:/Program Files/Java/jre1.8.0\_161” fixes the issue should it occur.

*Cleaning.R*

This file reads in the original data files, drops not needed columns and drops entries with NULL values or certain, pre discussed values.

Delivers a bar graph for each parameter and compares the distribution of original and cleaned data set’s Accident\_Severity.

All steps are described inside the report.

*Combining\_Aggregating.R*

Changes Junction\_Control value based on the value of Junction\_Detail.

Combines some values of certain columns.

Will also deliver a comparison of the Accident\_Severity between the original and the cleaned data sets.

All steps are described inside the report.

*Prediction\_Parameter.R*

You will be asked if you want to execute your own clustering or if you just want to execute the hard-coded clustering we prepared.

By typing “y” and pressing enter, you will execute your own kmeans cluster calculation for 2 to 15 clusters, 3 times, to plot the variances for the different cluster numbers.

After that a kmeans cluster will be performed on the local minimum of 8 clusters with 2500 starting points. This will produce a similar cluster to the one we decided for and hard-coded.

That will be executed afterwards.

*TimeAnalysis\_Hyp2.R*

Executes the analysis on the time data (Hypothesis 2 of the report)

*AgeAnalysis\_Hyp3.R*

To run k-means clustering for different values of k which increases run time significantly (total runtime ~10 mins) first ensure the global variable “Multi\_k\_Plot” is set to TRUE. Then type “y” when prompted. To disable this set “Multi\_k\_plot” to FALSE please press ENTER when prompted.

*EnvironmentAnalysis\_Hyp1\_Regression.R*

Executes a multivariate linear regression, a mars regression and a mars regression with cross validation of both the Accident\_Severity and the Custom\_Severity, created in Prediction\_Parameter.R