CIND 820

GitHub Repository Contents

**Capstone Project.Rproj:** project files in R

**InitialAnalysisKellyRevering.Rmd/InitialAnalysisKellyRevering.html:** the R markdown document in both original markdown file and knitted file. This file refers to all the data, both genders. It has the preprocessing and EDM of the dataset. The clustering, kmeans classification, decision tree classification, Naives Bayes classification, and linear regression coding and initial result metrics have been calculated.

**InitialAnalysisKellyRevering\_female.Rmd/InitialAnalysisKellyRevering\_female.html:** the R markdown document in both original markdown file and knitted file. This file refers to only data where gender is female. It has the preprocessing and EDM of the dataset. The clustering, kmeans classification, decision tree classification, Naives Bayes classification, and linear regression coding and initial result metrics have been calculated.

**InitialAnalysisKellyRevering\_male.Rmd/InitialAnalysisKellyRevering\_male.html:** the R markdown document in both original markdown file and knitted file. This file refers to only ddata where the gender is male. It has the preprocessing and EDM of the dataset. The clustering, kmeans classification, decision tree classification, Naives Bayes classification, and linear regression coding and initial result metrics have been calculated.

**README.md:** readme file detailing what preprocessing was completed on what datasets and the EDM of the datasets. As well as what classification/regression analysis was done and packages involved.

**Variable dictionary.docx:** a listing of all variables used in analysis and brief description.

**corr\_all\_copy:** a copy of the correlation coding for whole dataset analysis.

**corr\_female:** a copy of the correlation coding for female dataset analysis.

**corr\_male\_copy**: a copy of the correlation coding for male dataset analysis.

**dt\_analysis:** decision tree code for all analysis.

**nb\_analysis:** Naives Bayes analysis for all datasets.

**age\_levels.R:** R code for changing the age variable to a factor with levels

**change\_to\_factors.R:** R code to change variables to factors

**corr\_all.R:** R code to look at correlation between variables for the whole dataset

**corr\_all.csv:** resulting table of correlation values for the whole dataset

**corr\_all\_plot.png**: plot of resulting table

**corr\_male.csv:** resulting table of correlation values for male dataset

**decision tree.R:** decision tree r code for whole dataset

**gender\_files.R:** R code looking at the gender split in the dataset

**hist\_age.png:** plot of the histogram of the variable age for the whole dataset

**kmeans.R:** R code for kmeans for the whole dataset

**kmeans\_all.R:** redo of kmeans code for the whole dataset

**missing\_values.R:** rRcode looking at missing values in the whole dataset

**summary\_age\_gender\_class.R:** R code looking at age in relation to gender for the whole dataset

The above R script files were generated before the R markdown file. I had extreme difficult trying to commit changes to my repository from R and decided to just do all the coding in the R markdown file and upload the finished code to the repository. Some of the resulting tables and plots differ from the ones generated from the R markdown file.