**Project Folder Description**

1. Data Dictionary
   1. Data Dictionary of Selected and Regrouped Variables
   2. Original Data Manual
2. Datasets
   1. Dataset combined
   2. Dataset Cleaned (After regrouping)
   3. Datasets Raw
      1. Vehicle Dataset
      2. Person Dataset
      3. Accident Dataset
      4. xxxx
3. Codes
   1. Data Cleaning
      1. Data cleaning R Script
   2. Data Processing
      1. Data processing Python Script
   3. Machine learning model
      1. MARS R Script
      2. Logistic Regression R Script
      3. Neural Network Python Script
      4. Random Forest R Script
      5. Random Forest on top 6 important features R script
4. Tableau Dashboard: BC2407\_final.twbx
5. Project Report
6. Power point Slides

All code is downloadable on Github: <https://github.com/weikangg/BC2407-Analytics-II-Project>

Raw Datasets:

Dataset Download: <https://static.nhtsa.gov/nhtsa/downloads/FARS/2020/National/FARS2020NationalCSV.zip>

(Updated 19/01/2023)

Data User Guide (Full Dataset):

<https://static.nhtsa.gov/nhtsa/downloads/FARS/Links%20for%20FARS%20Manuals.pdf>