

README FOR PROJECT_1

R Project File: DSE6630_Project_1.Rproj

- Always open this file first; it will make everything sync together nicely if you do!

Knitted HTML File: Project_1.html

- This is a knitted version of the RMarkdown file that is **human readable** and will make it much easier to learn concepts and do your assignment!
- You should update this regularly as you make changes as a best practice for your teammates.

R Markdown File: Project_1.Rmd

- This is the working file where you will answer questions and do work.

Provided .Rdata files:

FY2024_data_files:

- This directory contains all of the necessary data files for 2024.
 1. `readmissionsClean2024.Rdata` = Cleaned readmissions data
 2. `pneumoniaFull2024.Rdata` = The full, merged, tidied dataset for pneumonia
 3. `pneumoniaAnalyzeNoEncoding2024.Rdata` = Fully prepared pneumonia dataset without any encoding
 4. `pneumoniaAnalyzeFullyEncoded2024.Rdata` = Fully prepared pneumonia dataset with all encoding from Demo 1
 5. `pneumoniaTrain.Rdata` = The initial training partition
 6. `pneumoniaTest.Rdata` = The initial testing partition
 7. `imputedTrain.Rdata` = Training partition after encoding and imputation
 8. `imputedTest.Rdata` = Testing partition after encoding and imputation
 9. `readyTrain.Rdata` = Imputed/Encoded training partition after transformations and removed redundancy
 10. `readyTest.Rdata` = Imputed/Encoded testing partition after transformations and removed redundancy

FY2025_data_files:

- This directory contains all of the necessary data files for 2025.
 1. `readmissionsClean2025.Rdata` = Cleaned readmissions data
 2. `pneumoniaFull2025.Rdata` = The full, merged, tidied dataset for pneumonia
 3. `pneumoniaAnalyzeNoEncoding2025.Rdata` = Fully prepared pneumonia dataset without any encoding
 4. `pneumoniaAnalyzeFullyEncoded2025.Rdata` = Fully prepared pneumonia dataset with all encoding from Demo 1
 5. `pneumoniaTrain.Rdata` = The initial training partition
 6. `pneumoniaTest.Rdata` = The initial testing partition
 7. `imputedTrain.Rdata` = Training partition after encoding and imputation
 8. `imputedTest.Rdata` = Testing partition after encoding and imputation
 9. `readyTrain.Rdata` = Imputed/Encoded training partition after transformations and removed redundancy
 10. `readyTest.Rdata` = Imputed/Encoded testing partition after transformations and removed redundancy

Source Code (.R files):

1. `answers_demo_1.R` = Contains the code for my solutions to Demo_1. Running produces `pneumoniaAnalyzeFullyEncoded2024.Rdata`
2. `doOrdinalEncoding.R` = Contains a function by the same name that performs ordinal encoding of any categorical predictors passed
3. `doFrequencyEncoding.R` = Contains a function by the same name that performs frequency encoding on categorical predictors passed
4. `reRunDemoData.R` = When run, completely processes the hospital readmissions data to do all the tasks from Demo_1 EXCEPT encoding. Produces `pneumoniaAnalyzeNoEncoding2024.Rdata`.

Project__1__cache

- Cache files that can speed up the running of knitting

Project__1__files

- Cache image files to speed up knitting

References file: `references.bib`

- Necessary for knitting correctly. It includes the cited bibliography in knitted report.