Key dates:

4/14/23 (8 PM) → Starting chat

4/18/23 (9 pm) → Clean Code chat

4/19/23 → Cleaned Dataset

4/21/23 (7:20 PM) → Checkpoint

4/26/23 (6:45 PM) → Checkpoint

4/28/23 (6:45 PM) → Checkpoint content and presentation

4/30/23 (??) → finish check

5/1/23 → Final review

5/2/23 → Presentation

Python Project

* (4/14/2023)
  + Assigning Roles: Stephani- Python Project
* Cleaned (do research and try out different columns)
* talk about which variables we chose and why
* Analysis
  + Correlation
  + Linear Regression
  + Logistic Regression
  + Random Forest
  + KMean
  + Decision Tree
* PowerPoint Presentation
  + Code
  + Algorithms
  + Visuals
  + Anything else to support your findings

Infrastructure Project

* (4/14/2023)
  + Assigning Roles: Emmanuel - Infrastructure Project
* (4/18/2023)
  + Infrastructure Design [Preliminary Understanding]
    - Assessing the Diabetic dataset structure
    - Understanding overall project scope and goals
    - Research into data architecture/infrastructure from Class and Online
* (4/21/2023)
  + Architecture Schematic [Designed Infrastructure Goals]
    - Collect/Store Data
      * Data Lakes, Cloud Bases, EHR Systems, Datatypes (Structured/Unstructured)
    - Preprocess and Clean Data
      * Techniques for processing, cleaning, and data normalization
    - Model Building
      * Construction of algorithm to analyze and assess data for readmission cues
      * Testing and validations of results
    - Deployment/Implementation
      * Taking the algorith/model and implementing it within a clinical dashboard or EMR system for clinicians to predict readmissions
    - Maintenance
      * Model upkeep, making sure that the model is running correctly and accurately
      * Iterations to improve model with new data
* (4/26/2023)
  + Preliminary Schematic Design
    - Canva + Photoshop + Powerpoint
    - AWS Storage, ML and Deployment
* (4/28/2023)
  + PowerPoint Presentation
    - Upload Infrastructure Schematic
    - Contruct Script
* (4/30/2023)
  + Presentation Introduction
    - Diabetes Statistics
    - Diabetes - Readmission Statistics
    - Dataset Data Explanation [Columns]
    - Goal of Readmissions Project
* (5/1/2023)
  + Final Review and Mock Presentation

Visualization Project

* Tableau Dashboard
  + At least 4 charts
    - Demographics
      * Age
      * Race
    - 30-Day Readmission
  + [Colors chosen must not be Boicey’s Colorblind]
* 4/14/23
  + Assigning Roles: Wenjin Visualization
* 4/18/23
  + Decision on key variables/columns to be used for tableau visualization based on cleaned dataset
* 4/19/23
  + start creating the based charts needed for the visualization
* 4/21/23
  + create dashboard V1.0 for group to have a basic idea of how the dashboard will turn out and decided on a color scheme to be used as a group
* 4/26/23
  + Add additional charts based on the events of the python project and further discuss the color scheme to be used based on V1.5 of dashboard
* 4/28/23
  + Show group the final dashboard product making sure everyone in group is okay with it, further discuss color scheme, charts placement, and other miscellaneous details of dashboard if needed
* 4/30/23
  + Begin adding slides to the presentation, and drafting script
* 5/1/23
  + Practicing presenting the presentation as a group

Python Analysis

* Work with Group members to clean data
* Attempt 1: ML models without getting rid of columns
* Attempt 2: Using RSME to choose columns → ML models
* Attempt 3: Using Correlation Matrix and Anova to choose columns
* Attempt 4: RSME and Anova for dummy variables