**I. Data Preparation**

* Include all the packages that are needed for the model.
* Access the dataset from the drive.
* Check the dataset.
* Slice data to prepare for the model.

**II. Building Logistic Regression Model**

* Apply the prepared data into the model.
* Check the logistic regression model score.

**III. Feature Importance**

* Rank the features based on the coefficients.

**IV. Evaluation**

* Use AUC score and ROC curve to evaluate the model.

**V. Conclusion**

- The Logistic Regression model provides the score of 68%, and this is acceptable to predict CVD Risk. The AUC score is 0.64, so we can expect there would be many errors in our predictions. The ROC curve reflects the area that the model will have wrong predictions. From the coefficients, race, mstat, hip, neck20, av\_weight\_kg, tea15, happy25, and hlthlm25 have negative coefficients. When these variables go up, the CVD risk will go down. Therefore, the patient who has high numbers of these variables will tend not to have CVD risk. On the other hand, when the rest of independent variables go up, the CVD risk will go up. Furthermore, the coefficients reflect how much the features impact the CVD risk. The rank indicates the most impactful feature to the least effective feature.