Lab 1 interpretation

Preprocessing and Data Loading:  
  
The read\_csv function from the tidyverse package is used to load the data from a CSV file called "studentInfo.csv".  
Mutation results in the creation of two new variables: disability\_status, a factor indicating whether a student has a disability, and pass\_status, which shows whether a student passed or not.  
Once these alterations occur, the data is analyzed.

Engineering Features:  
  
After being transformed into a factor with predetermined levels, the variable imd\_band is transformed again into the numeric variable imd\_band\_numeric. Most likely, this is for easier to understand model or interpretation.

Data division:  
  
Using the initial\_split function from the sample package, the data is divided into training and testing sets. 20% of the data is set aside for testing, and the remaining 80% is kept for training.  
Model Details:  
  
Using logistic\_reg() from the parsnip package, a logistic regression model is specified with the engine set to "glam" and the mode set to "classification".

Creation of Workflows:  
  
Workflow() from the workflows package is used to create a workflow. The workflow is expanded to include the model and recipe.  
Model Fitting:  
  
Fit() is used to fit the model to the training set of data.  
Assessment of the Model:  
  
Last\_fit() is used to assess the final model based on the testing data.  
The function collect\_predictions() is used to gather predictions.  
By comparing the expected class with the actual pass status and tallying the accurate predictions, the model's accuracy is determined.

Conclusion:  
  
The final draft of the document is ready, maybe for reporting or presenting.