**Problem Statement:**

A University would like to effectively classify their student based on the Program they are enrolled into, perform multinomial regression on the given dataset and provide insights in the documentation.

**prog**: is a categorical variable indicating what type of program a student is in: “General” (1), “Academic” (2), or “Vocational” (3)

**Ses**: is a categorical variable indicating someone’s socioeconomic class: “Low” (1), “Middle” (2), and “High” (3)

**read, write, math, science** are their scores on different tests

**honors**: Whether they have enrolled or not

Solution:

There are five basic steps when you’re implementing Multinomial regression:

1. Import the packages and classes you need.
2. Provide data to work with and eventually do appropriate transformations.
3. Create a regression model and fit it with existing data.
4. Check for collinearity between independent variables and between independent and dependent variables .
5. Check for overfitting issue
6. Apply the model for predictions.

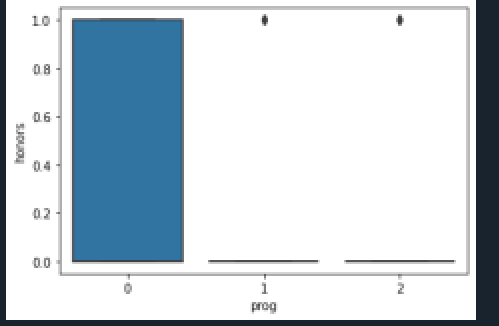
The first step is to import the package numpy and

import numpy as np

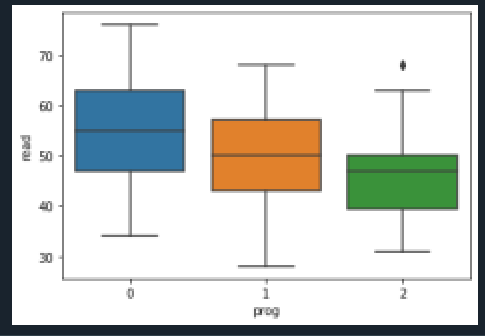
Now, you have all the functionalities you need to implement logistic regression.

relavancy check of dependent variable with independent variables by multiple plots

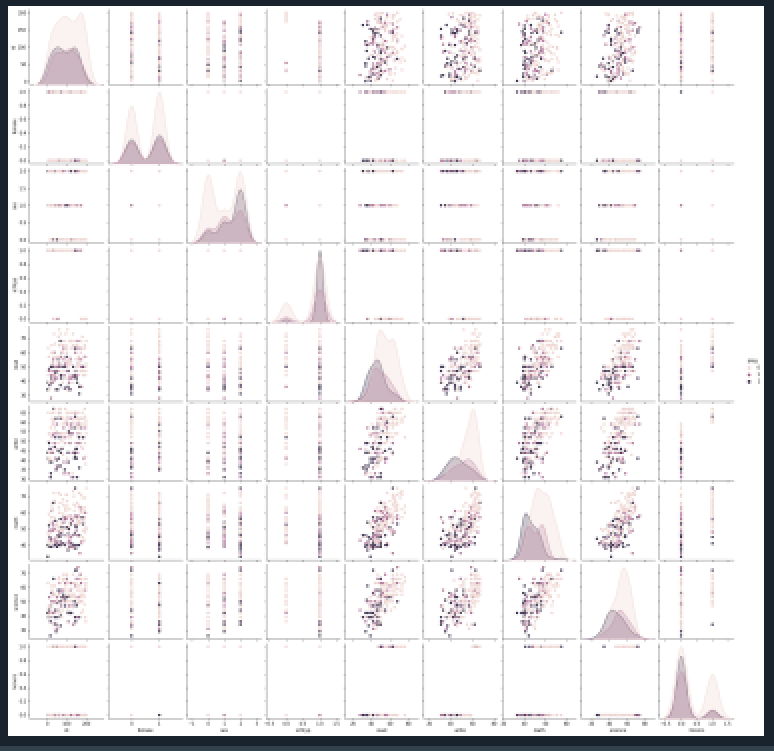
Boxplot of independent variable distribution for each category of choice

sns.boxplot(x = "prog", y = "honors", data = mode)

sns.boxplot(x = "prog", y = "read", data = mode)



Pairplot of loan data set



Building the Model

Train and Test Data Sets

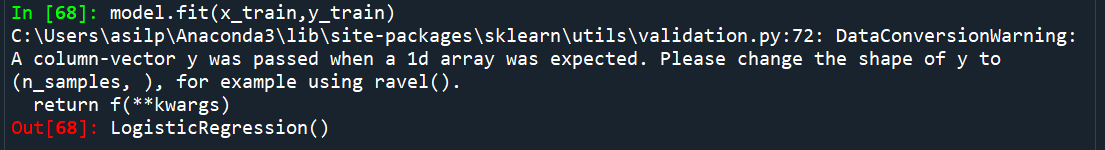
We must now divide our dataset into two distinct parts. A training set and a testing set. We can use scikit-learn for this.

Prior to segmenting the data, we will need to define our input and output variables i.e. X and y. For the purpose of this article, we will stick to numerical data when defining the input values for X.

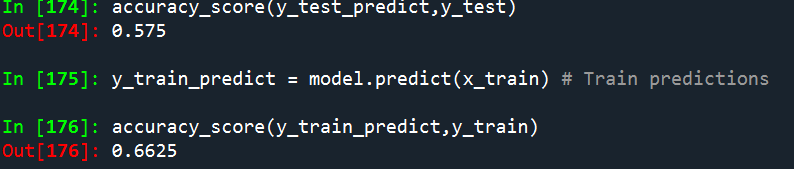


Training and testing the model

We will go ahead and import the Logistic Regression from the sklearn.linear\_model library:



By importing scikit-learn's ‘accuracy\_score’ we can swiftly assess the accuracy of our model and using confusion matrix we can visualize the accuracy of model



# Conclusion

The obtained result shows the power of Machine Learning algorithms. A multinomial Logistic Regression model has proven to be very successful in solving this classification problem.