Multinomial Regression (Module -10)

1. You need to predict the type of program a student is in based on other attributes --> mdata.csv

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

	X	id	female	ses	schtyp	prog	read	write	math	science	honors
1	1	45	female	low	public	vocation	34	35	41	29	not enrolled
2	2	108	male	middle	public	general	34	33	41	36	not enrolled
3	3	15	male	high	public	vocation	39	39	44	26	not enrolled
4	4	67	male	low	public	vocation	37	37	42	33	not enrolled
5	5	153	male	middle	public	vocation	39	31	40	39	not enrolled
6	6	51	female	high	public	general	42	36	42	31	not enrolled
7	7	164	male	middle	public	vocation	31	36	46	39	not enrolled
8	8	133	male	middle	public	vocation	50	31	40	34	not enrolled
9	9	2	female	middle	public	vocation	39	41	33	42	not enrolled
10	10	53	male	middle	public	vocation	34	37	46	39	not enrolled
11	11	1	female	low	public	vocation	34	44	40	39	not enrolled
12	12	128	male	high	public	academic	39	33	38	47	not enrolled

Hints:

1. Business Problem

    1.1. Objective
    1.2. Constraints (if any)

2. Data Pre-processing

    2.1 Data cleaning, Feature Engineering, EDA etc.

3. Model Building

    3.1 Partition the dataset
    3.2 Model(s) - Reasons to choose any algorithm
    3.3 Model(s) Improvement steps
    3.4 Model Evaluation
    3.5 Python and R codes

4. Deployment

    4.1 Deploy solutions using R shiny and Python Flask.

5. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

Note:

1. For each assignment the solution should be submitted in the format

2. Research and Perform all possible steps for improving the model(s) accuracy.

    Ex: Transformations, Feature Engineering, Hyper Parameter tuning, Outlier treatment, etc.

3. All the codes (executable programs) are running without errors

4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here.