**Problem Statement:**

The following dataset credit.csv contains information about credit card debt for hundreds of customers. The response is balance (average credit card debt for each individual) and there are several quantitative predictors: age, cards (number of credit cards), education (years of education), income (in thousands of dollars), limit (credit limit), and rating (credit rating). You need to create linear regression models based on the techniques you have learnt in the course and point out which models would be best suited for this dataset.

<https://drive.google.com/file/d/1Ga7Q25CTCR7EFYmdlR2JEr6m6ptcS_sk/view?usp=sharing>

Note that this is an open ended assignment, you are expected to carry out data exploration, extensive model selection process, hyperparamer tuning, model evaluation and testing etc. A Machine Learning project is generally not linear but will consist broadly of the following major steps.

Define Problem

Prepare Data

Data understanding (EDA)

Model building

Evaluate and compare models

Improve Model

Report Results

Evaluation will be relative based upon your results and efforts put in.

**What will you have to submit?**

You need to submit a jupyter notebook (.ipynb) file with all the work done via the elearn(Taxila) portal. Naming convention of the file should be <BITS-ID-number>.ipynb.

**Deadline:**

1st November, 2023

Note that this is a hard deadline and no extension will be granted further. Copying efforts will be dealt very seriously, students involved, if caught, will be given zero marks.