**Employment Case Study:**

Mr. X is a final-year student at the School of Management, Uddhampur. He is working on an assignment to predict the salary of an employee based on several factors. He was able to find a dataset from the census database.

Your task as a data analyst is to help him

* Build the best solution that would predict “Salary”
* Identify the most important factors that affect “Salary”
* Also return the predictions of “Salary” in the “hold\_out” dataset.

Note: Kindly mention them in the PPT. The PPT should detail your complete solution approach. You are free to make Assumptions as required,

**Data Set Description:**

**Age:** Age of the person

**Education:** Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.

**Education Yrs:** Number of years the person had undergone formal study.

**Marital Status:** Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.

**Employer:** Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.

**Occupation:** Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.

**Relationship Status:** Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.

**Sex:** Female, Male.

**Capital gain:** Amount gained by the person during that time.

**Capital loss:** Amount lost by the person during that time.

**Hours per Week:** continuous.

**Native Country:** Native Country of the person

**Salary: Salary per year**