**Problem Statement:**

For this dataset, you will have to predict the income of an employee based on their different attributes. The prediction task is to determine whether a person makes over $50K a year or not i.e., there are only 2 categories ($50K, <=$50K per year) on which each employee will be classified.

**Dataset description:**

• age: continuous.

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

• fnlwgt: continuous.

• 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-num: continuous.

• marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.

• 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: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.  
race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.

• sex: Female, Male.

• capital-gain: continuous.

• capital-loss: continuous.

• hours-per-week: continuous.

• native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

**Evaluation criteria:**

• EDA and pre-processing

• Feature importance

• Modelling and results

• Business solution/ Interpretation of results obtained

' relationship', ' marital-status', ' capital-gain', 'age', ' education-num', ' occupation', ' hours-per-week', ' fnlwgt'