**Problem Statement:**

Census-income data plays the most important role in the democratic system of government, highly affecting the economic sectors. Census-related figures are used to allocate federal funding by the government to different states and localities.

Census data is also used for post census residents estimates and predictions, economic and social science research, and many other such applications. Therefore, the importance of this data and its accurate predictions is very clear to us. The main aim is to increase awareness about how the income factor actually has an impact not only on the individual lives of citizens but also an effect on the nation and its betterment. You will have a look at the data pulled out from the 1994 Census bureau database, and try to find insights into how various features have an effect on the income of an individual.

The data contains approximately 32,000 observations with over 15 variables.

The strategy is to analyze the data and perform a predictive task of classification to predict whether an individual makes over 50K a year or less by using a logistic regression algorithm.

|  |  |
| --- | --- |
| **Column Names** | **Description** |
| Age | Age of the individual |
| Workclass | department of the working individual |
| fnlwgt | Final weight of the individual |
| education | The education degree of the individual |
| education-num | Number of years of education |
| marital-status | Marital status of the individual |
| occupation | Occupation of the individual |
| relationship | Relation value |
| race | Ethnicity of the individual |
| sex | Female, Male |
| capital-gain | capital gain of the individual |
| capital-loss | capital loss of the individual |
| hours-per-week | number of working hours |
| native-country | The native country of the individual |
| Annual-Income | Annual income either >50K or <=50K |