|  |  |
| --- | --- |
| A blue and orange star with a black background  Description automatically generated | The University of Texas at Arlington  Applied Statistics and Data Science |

**ASDS-6303-002 DATA MINING WITH INFO VISUAL**

* **FINAL PROJECT**

**TITLE: AMERICAN CITIZENS ANNUAL INCOME PREDICTION**

**GROUP 05**

**SRUJANA GUGGILLA [1002165664]**

**MISBAHJABIN SHAIKH [1002189629]**

**TABLE CONTENT**

**Introduction**

**Data Overview**

The dataset is collected from Kaggle holding 25,000 observations and 15 variables. The data type contains Int & Char.

|  |  |
| --- | --- |
| Column | Description |
| age | Age |
| workclass | A general term indicating the employment status of an individual. |
| fnlwgt | Final weight, representing the number of individuals that this row represents (a representative sample). |
| education | Highest level of education achieved by an individual. |
| education.num | Highest level of education achieved by an individual in numerical form. |
| marital.status | Marital status of an individual. Note that Married-civ-spouse refers to a civilian spouse, and Married-AF-spouse refers to a spouse in the Armed Forces. |
| occupation | General type of occupation of an individual. |
| relationship | Relationship of this individual with others, for example, spouse (Husband). Each data point has only one relationship. |
| race | Race |
| sex | Biological sex of an individual. |
| capital.gain | Capital gains of an individual. |
| capital.loss | Capital losses of an individual. |
| hours.per.week | Number of hours the individual reported working per week. |
| native.country | Country of origin. |
| income | Income, less than or equal to $50,000 (<=50K) or more than that (>50K). |