## Front End Engineering-II /Artificial

## Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

Dataset Adult

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

MR. RAJIV BHARDWAJ TAMANNA

2210990885 (G-13)

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Description About the case study**

# **Display Top 10 rows**

# **Display Last 10 rows**

# **Check shape of our dataset(number of rows and number of columns)**

# **Total number rows,total number of columns,datatypes of each column and memory requirement**

# **Fetch Random Samples from DataSet(50%)**

**Check NULL values in dataset**

**Perform Data Cleaning(Replace '?' with NaN) and plot in graph with seaborm library**

**Drops all rows having missing values**

**Check for duplicate data and drop them**

**Get overall Statistics about the dataframe**

**Bivariate Analysis on graph**

**Replace Salary values ['<=50k','>50k'] with 0 and 1**

**Which Workclass getting the Highest Salary?**

**Covert workclass Columns Datatype to category Data**

**Library**

**library used in this case study is pandas**

**Method**

**read\_csv():**

**Description: Reads a CSV file and converts it into a data frame.**

**tail():**

**Description: Displays the last few rows of the data frame.**

**head():**

**Description: Displays the first few rows of the data frame.**

**shape():**

**Description: Returns the shape (number of rows, number of columns) of the data frame.**

**info():**

**Description: Provides basic information about the data frame, such as column types and missing values.**

**sample():**

**Description: Randomly sample data from daraframes or series.**

**isnull():**

**Description: Returns True/False for each value in the data frame, indicating whether the value is missing (NaN) or not.**

**sum():**

**Description: Calculates the sum of values in each column of the data frame.**

**drop():**

**Description: Removes specific rows or columns from the data frame.**

**dropna():**

**Description: To remove missing values.**

**duplicated():**

**Description: To identify duplicated rows.**

**Describe():**

**Description: To generate descriptive statistics of the numerical columns in a dataframe.**

**Countplot():**

**Description: Plot the count of observations in each category.**

**value\_counts():**

**Description: Counts the unique values in a specific column of the data frame.**

**nunique():**

**Description: Returns the count of unique values in a specific column of the data frame.**

**apply():**

**Description: Apply a function along an axis of a dataframe or series.**

**Groupby():**

**Description: Split the dataframe into groups based on some criteria.**

**contains():**

**Description: Checks if a specified substring or value is present in a column of the data frame.**

**Astype():**

**Description: Used to convert the datatype of a series or dataframe to a specific datatype.**