**# Task 1: Statistical Exploratory Data Analysis**

**Let us start with getting know the dataset. Your first task will be to get some basic information by using Pandas features.**

**Task 1-a: Print the details of the df\_census data frame (information such as number of rows,columns, name of columns, etc)**

A screenshot of a computer

Description automatically generated with medium confidence

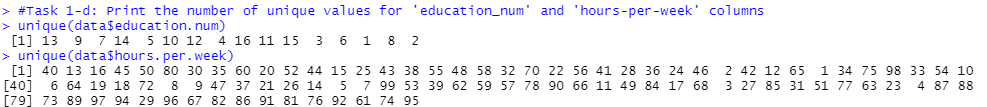
**#Task 1-b: Find the number of rows and columns in the df\_census data frame.**



**#Task 1-c: Print the descriptive details (min, max, quartiles etc) for 'Age' column of the df\_census**



**#Task 1-d: Print the number of unique values for 'education\_num' and 'hours-per-week' columns**



**# Task 2: Aggregation & Filtering & Rank**

**In this task, we will perform some very high level aggregation and filtering operations. Then, we will apply ranking on the results for some tasks.**

**Text

Description automatically generated**

**#Task 2-b: Find out the total number of people surveyed in months may, october and december.**

**Text

Description automatically generated**

Note: The True values are surveyed count for the month

**#Task 2-c: Let us now use multiple filtering criteria**

**# Find out the total number of surveys in september and november with workclass as private and age less than 50.**

A picture containing graphical user interface

Description automatically generated

Note: The True values are surveyed count for the month

**#Task 2-d: Find out 3 least surveyed education categories, print their names and corresponding number of surveys for periods January-June and July-December.**

**Graphical user interface, text, application

Description automatically generated**

**#Task 2-e: Find out top 5 native-countries besides United-States, print their names and number of surveys belonging to each.**

**Text

Description automatically generated**

**#Task 2-f: Find out Top-5 native-countries with the most number of samples belonging to class >50K**

**Text

Description automatically generated with medium confidence**

**# Task 3: Visualization**

**In this task, you will perform a number of visualization tasks to get some intuition about the data. Visualization is a key component of exploration. You can choose to use either Matplotlib or Seaborn for plotting. The default figures generated from Matplotlib might look a bit ugly. So you might want to try Seaborn to get better figures. Seaborn has a variety of styles. Feel free to experiment with them and choose the one you like. We have earmarked 10 points for the aesthetics of your visualizations.**

**#Task 3-a: Draw a histogram for total number of surveys taken each month. Dislpay months with their corresponding numbers(Eg: January is 1)**

**Chart, histogram

Description automatically generated**

**#Task 3-b: Draw a vertical bar chart for total number of surveys taken for each gender for each month. Display months with their corresponding names.**

**# Remember to make the bar chart into a vertical bar chart**

**Graphical user interface

Description automatically generated with low confidence**

**#Task 3-c: Draw a horizontal bar chart for number of surveys taken with respect to age feature keeping the age interval as 15.**

**# Remember to make the bar chart into a horizontal bar chart**

**Chart

Description automatically generated**

**#Task 3-d: Draw a "vertical" bar chart that lists the top-5 native-countries based on the number of samples with class >50K.**

**# Remember to make the bar chart into a vertical bar chart**

**Chart

Description automatically generated**

**#Task 3-e: Now repeat Task 3-d based on education (again top-5)**

**Chart, bar chart

Description automatically generated**

**#Task 3-f: Draw a scatter plot for age vs hours per week.**

**A picture containing scatter chart

Description automatically generated**

**#Task 3-g: Draw a line chart showing average capital gain for each education category.**

**Chart, line chart, histogram

Description automatically generated**

**#Task 3-h: Draw a 'horizontal' bar chart for the top-5 most common occupation.**

**A picture containing table

Description automatically generated**

**#Task 3-i: Draw a 'horizontal' bar chart for the top-5 most common workclass.**

**Chart

Description automatically generated**

**Task 4 :**

The dataset contains a lot of interesting points. Most fascinating thing we found in this data set is that the United States tops in all the categories of comparison.

Working class of a country decides the overall growth of the nation. As we can see through our analysis of this data set ‘United States ‘ has maximum working class people with Bachelorette degree.So United nation will have maximum productivity.

**Graphical user interface

Description automatically generated**

Now we were curious to know which country had second best productivity after United states. We found out Philippines is second in list.

Chart

Description automatically generated