**Introduction**

This assignment is divided into four parts. First is the python section with census dataset, which includes entities such as age, class, qualification, and many more. The dataset contains nearly 30000 entries in total.

We used this dataset to perform a number of operations in Jupyter Notebook (Python). We used the Pandas library to perform operations like data exploration, visualization, and processing.

In a nutshell, this massive data set contains several important pieces of information, the majority of which are from the United States, with a smaller proportion from other countries.

**Task 1: Statistical Exploratory Data Analysis**

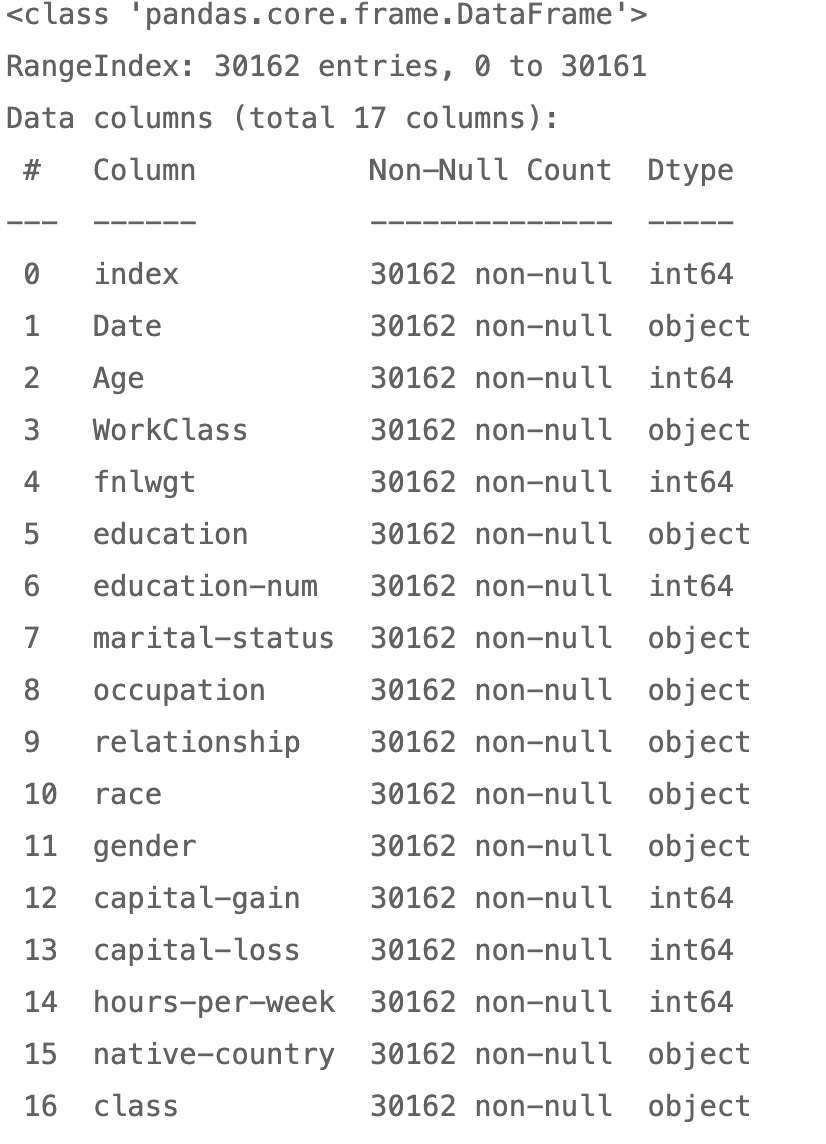
**#For each task below, look for a Pandas function to do the task.**

**#Replace None in each task with your code.**

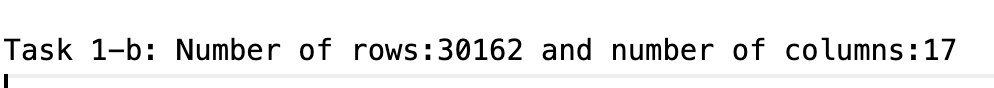
**#Before starting with the tasks in the assignment, we need to remove the rows with missing values**

**#write the code for it here and save the new data frame with the same name as df\_census.**

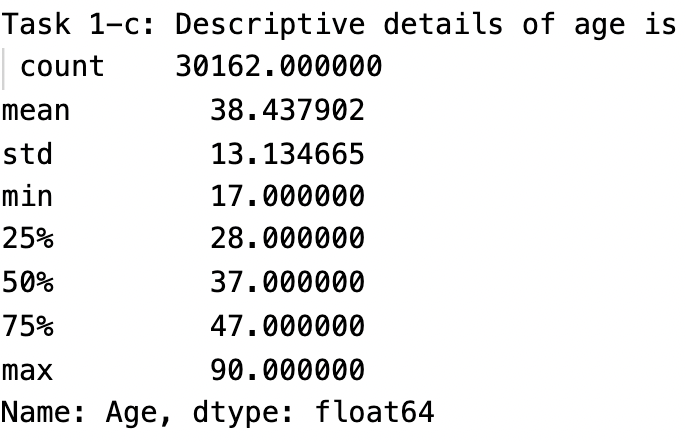
**#Task 1-a: Print the details of the df\_census data frame (information such as number of rows,columns, name of columns, etc)**

****

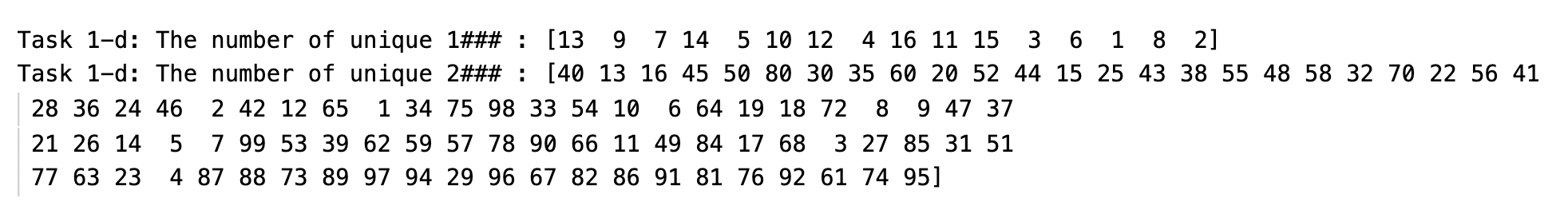
**#Task 1-b: Find the number of rows and columns in the df\_census data frame.**

****

**#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. Pandas has a convenient and powerful syntax for aggregation, filtering, and ranking.

**#Task 2-a: Find out the sum of Captial Gain for people with education level as Bachelors** and HS-Grad.

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

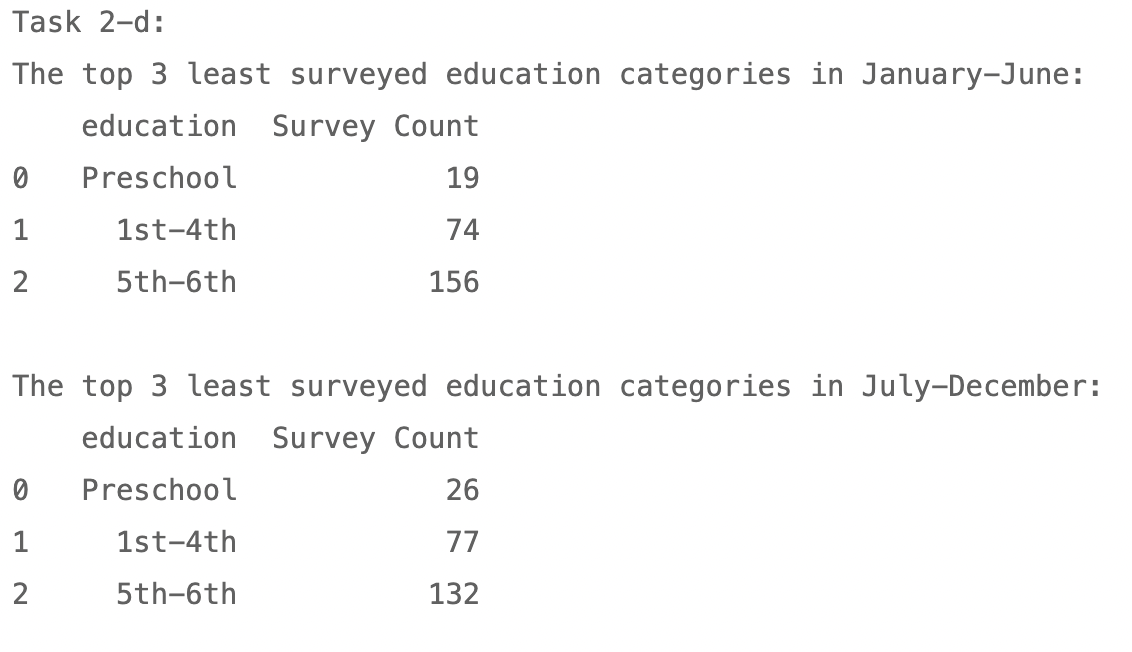


**#Task 2-c: Let us now use multiple filtering criteria**

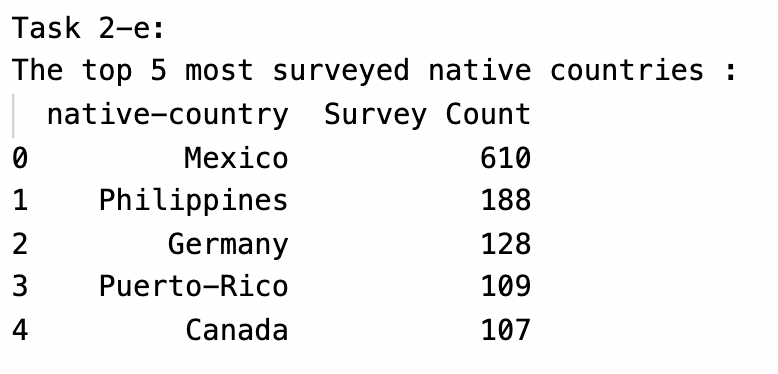
**# Find out the total number of surveys in september and november with workclass as private and age less than 50.**



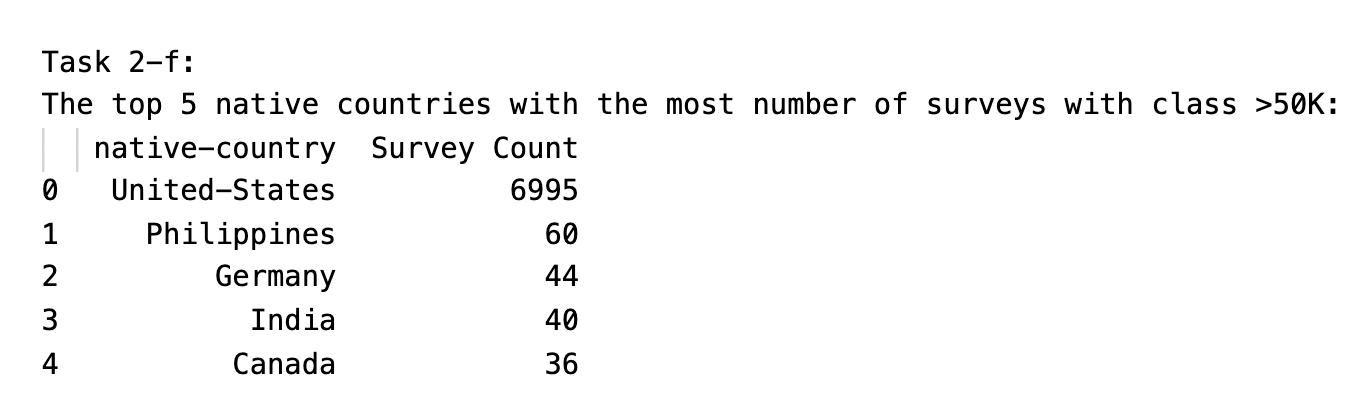
**#Task 2-d: Find out 3 least surveyed education categories, print their names and corresponding number of surveys for periods January-June and July-December.**



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



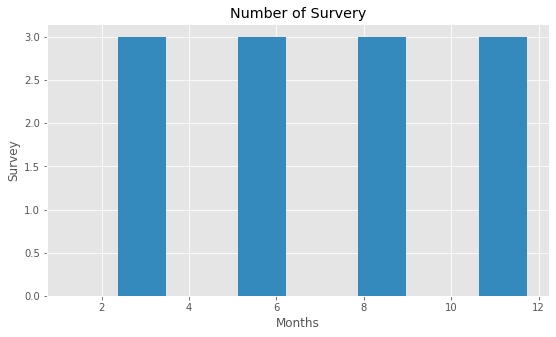
**#Task 2-f: Find out Top-5 native-countries with the most number of samples belonging to class >50K**



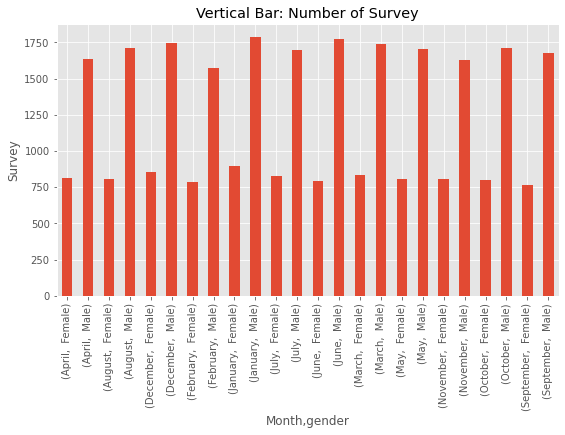
**Task 3: Visualization**

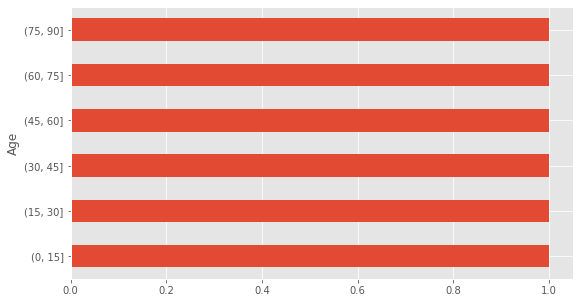
In this task, we will perform a number of visualization tasks to get some intuition about the data. Visualization is a key component of exploration. We choose to use either Seaborn for plotting graphs

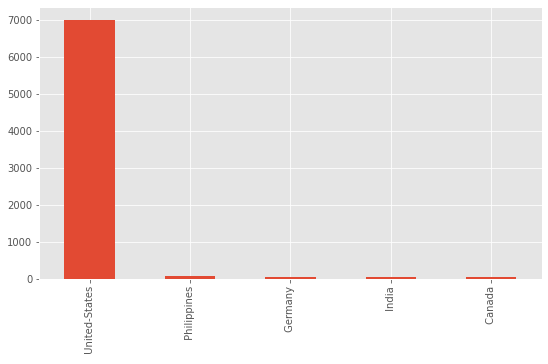
**#Task 3-a: Draw a histogram for total number of surveys taken each month. Dislpay months with their corresponding numbers(Eg: January is 1)**



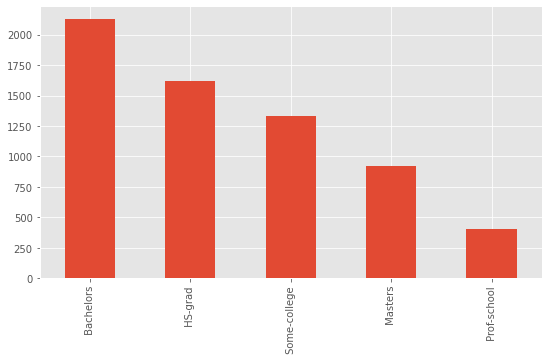
**#Task 3-b: Draw a vertical bar chart for total number of surveys taken for each gender for each month. Display months with their corresponding names.**

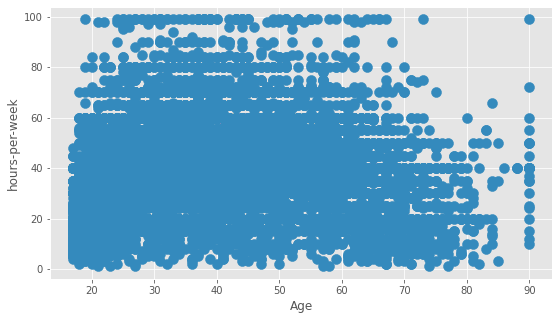


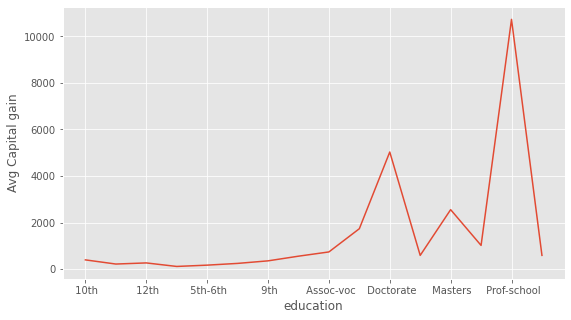
**#Task 3-c: Draw a horizontal bar chart for number of surveys taken with respect to age feature keeping the age interval as 15.**

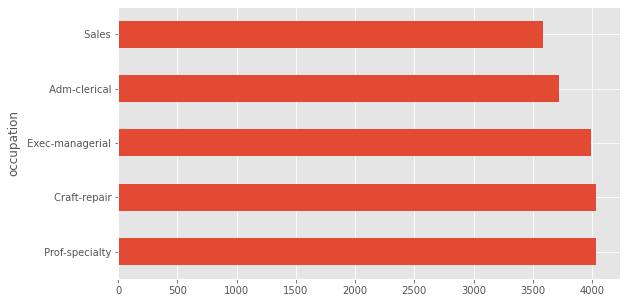
**#Task 3-d: Draw a "vertical" bar chart that lists the top-5 native-countries based on the number of samples with class >50K.**

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

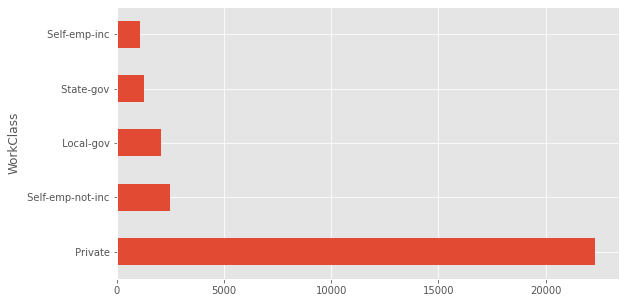


**#Task 3-f: Draw a scatter plot for age vs hours per week.**

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

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

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



**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.

For instance, the comparison of education level vs income, on plotting a graph between people who completed their ‘Doctorate’ vs ‘Income > 50K’, we found that the US tops in this category with the highest number of doctorates and highest number of people with income more than $50,000 with this degree.

This analysis shows that many people in the United-States generate income more than $50,000 if they hold a Doctorate degree.

