Assignment3 - Pandas

Data source - UCI adult dataset, also called as Census Income

Link - UCI Machine Learning Repository: Adult Data Set

Question - Predict whether income exceeds $50K/yr based on census data.

Answer the following

1. Shape of dataset
2. Print column names
3. **How many men and women (**sex**feature) are represented in this dataset?**
4. **What is the average age (**age**feature) of women?**
5. **What is the percentage of German citizens (**native-country**feature)?**
6. **What are the mean and standard deviation of age for those who earn more than 50K per year (**salary**feature) and those who earn less than 50K per year?**
7. **Is it true that people who earn more than 50K have at least high school education? (**education – Bachelors, Prof-school, Assoc-acdm, Assoc-voc, Masters**or**Doctorate**feature)**
8. **Display age statistics for each race (**race**feature) and each gender (**sex**feature). Use**groupby()**and**describe()**. Find the maximum age of men of**Amer-Indian-Eskimo**race.**
9. **Among whom is the proportion of those who earn a lot (>50K) greater: married or single men (**marital-status**feature)? Consider as married those who have a**marital-status**starting with**Married**(Married-civ-spouse, Married-spouse-absent or Married-AF-spouse), the rest are considered bachelors.**
10. **What is the maximum number of hours a person works per week (**hours-per-week**feature)? How many people work such a number of hours, and what is the percentage of those who earn a lot (>50K) among them?**
11. **Count the average time of work (**hours-per-week**) for those who earn a little and a lot (**salary**) for each country (**native-country**). What will these be for Japan?**