

You can find adult dataset of UCI [here](#) (adult.data)

The description of the data is available [here](#).

## **Assignment 9**

**(Due 06<sup>th</sup> Nov, 2022)**

1. Write Python/Java code to cluster the data on at least 3 attributes (for example, age, workclass, education).
2. Execute the exercise using Map-Reduce.
3. Execute the above exercise as a map reduce job using Apache Spark.
4. Using spark transformations/actions, find
  - i) The country from which highest number of adults are there except USA
  - ii) No of people who are at least Masters and works in “Tech-support”
  - iii) No of Unmarried male who works in “Local\_govt”