# Naïve Bayes assignment problem

Problem Statement: 1) Prepare a classification model using Naive Bayes

for salary data

Data Description:

age -- age of a person

workclass -- A work class is a grouping of work

education -- Education of an individuals

maritalstatus -- Marital status of an individulas

occupation -- occupation of an individuals

relationship --

race -- Race of an Individual

sex -- Gender of an Individual

capitalgain -- profit received from the sale of an investment

capitalloss -- A decrease in the value of a capital asset

hoursperweek -- number of hours work per week

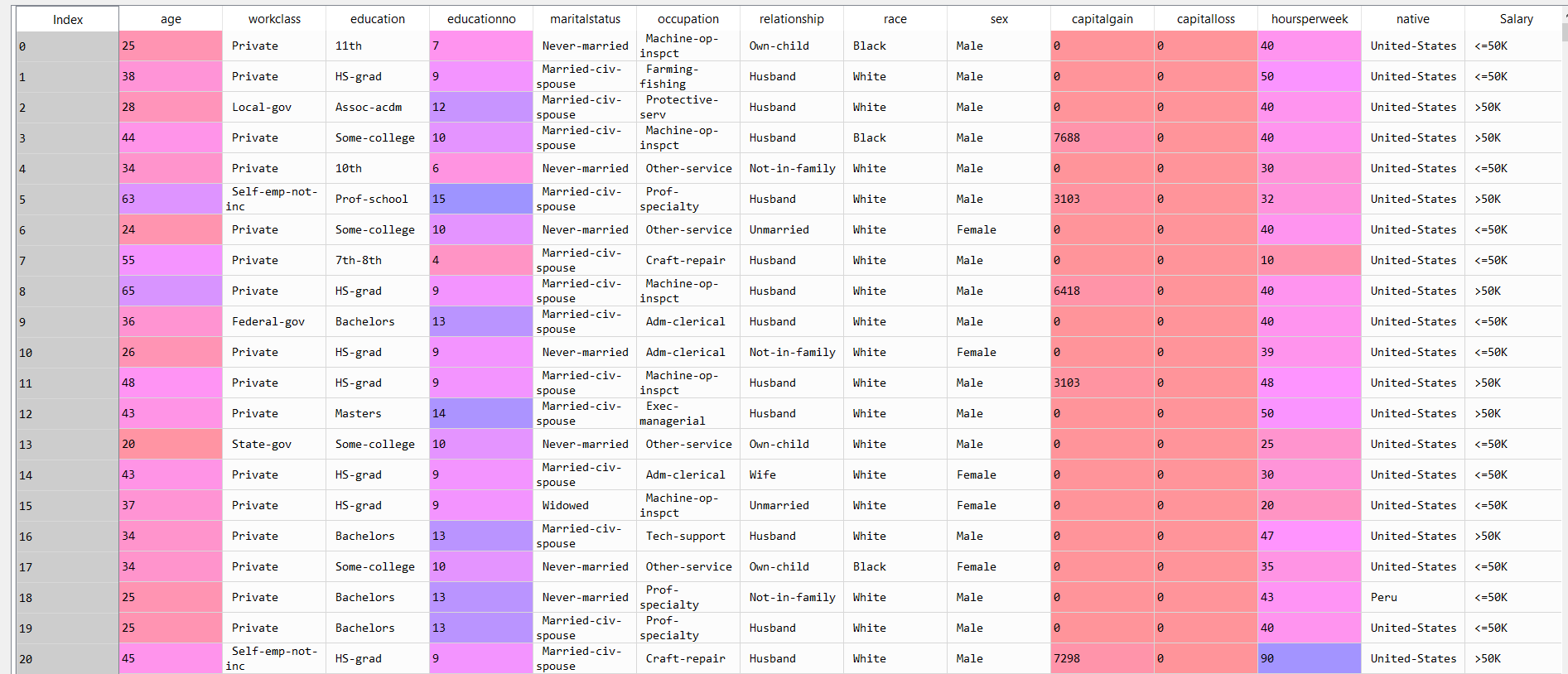
native -- Native of an individual

Salary -- salary of an individual

We import the train data as follows:

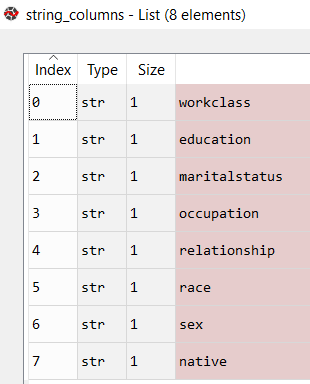


We proceed to import the test data as follows:



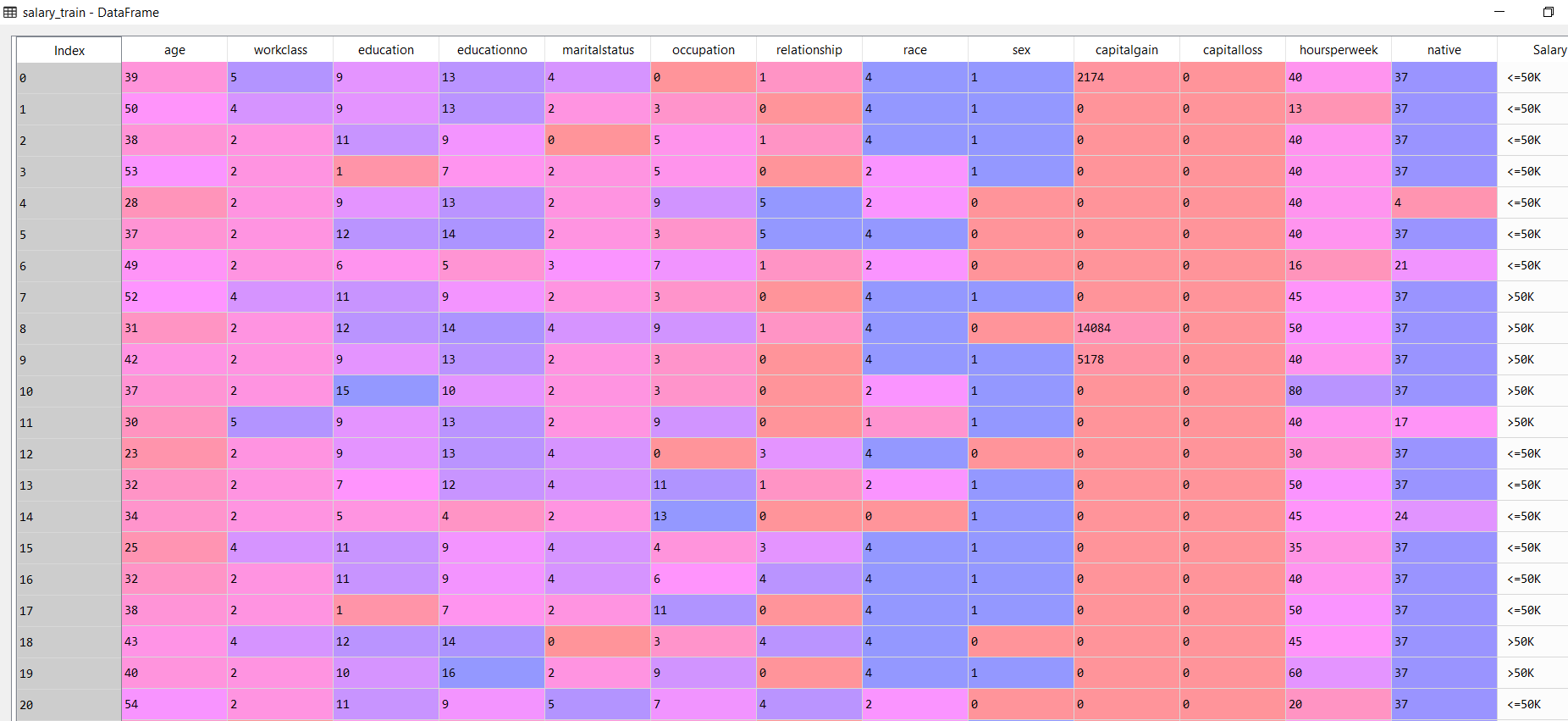
From the above testing and training data we have few features that are categorical in nature and these variables are to be converted to numerical values

We have a list of variables that are identified as being categorical as follows:

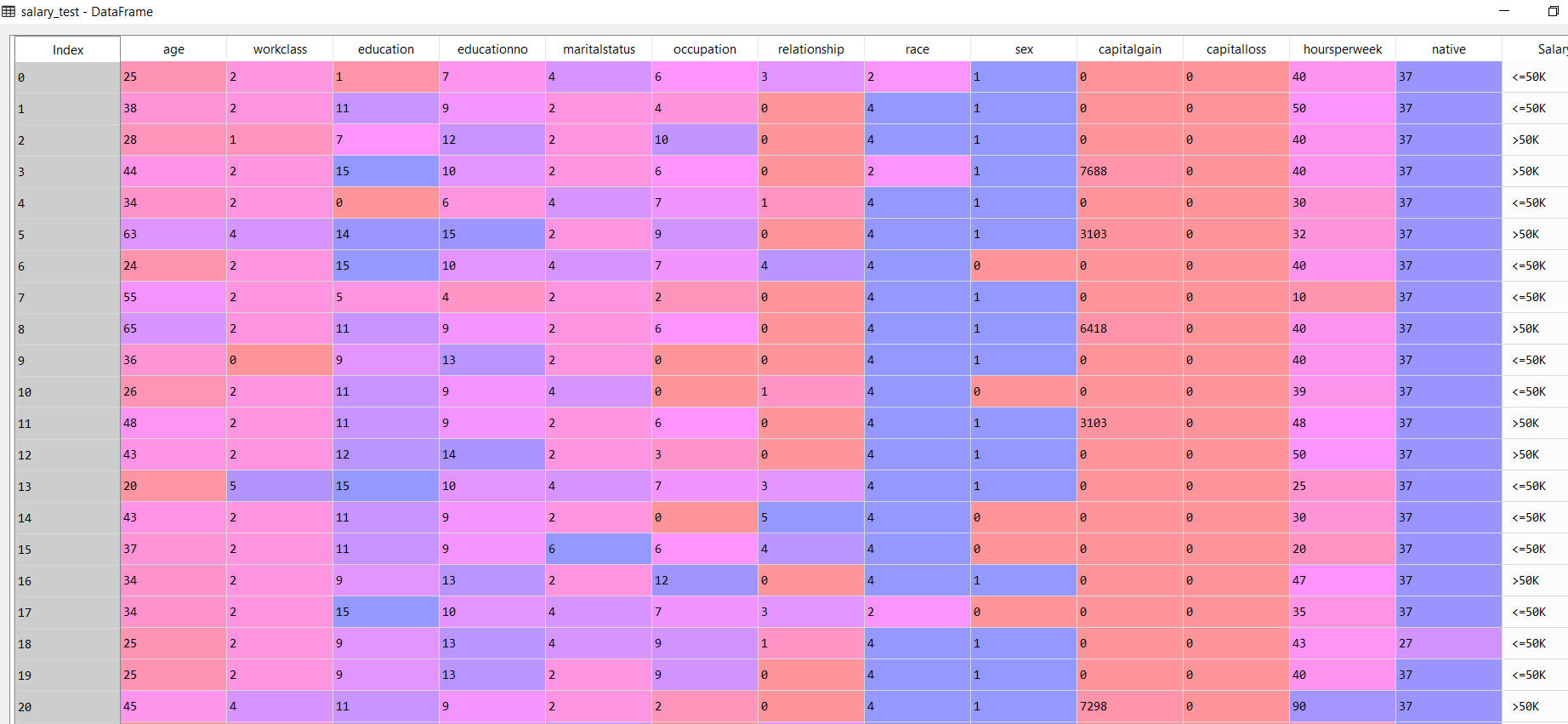


These variables are converted as follows:

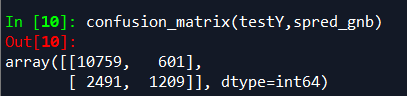
Train data:



Test data:

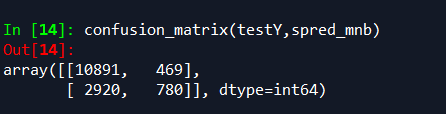


The data is run through the Naïve Bayes Gaussian algorithm and the following are the results:



The accuracy of the above model is about 79%

The data is run through the Naïve Bayes Multinomial algorithm and the following are the results:



The accuracy of the above model is about 76%