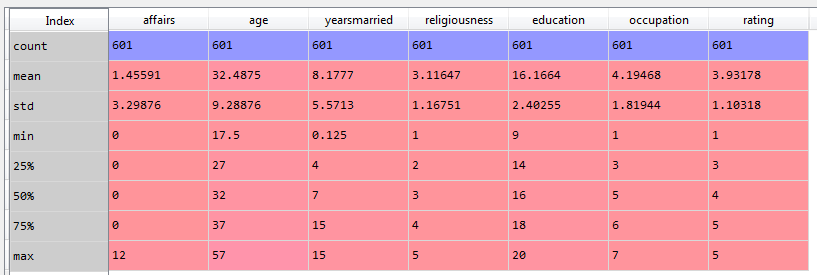
**Affairs Data Logistics Regression**

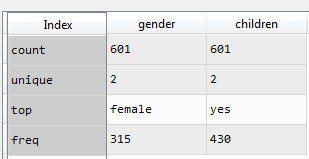
**Descriptive Statistics of the categorical variables is as follows:**



According to the above exploration we can find the following information:

* 75% of people don’t have any affairs.

**Descriptive Stats of the categorical variables and the ones who have subscribed to the term deposit:**



According to the above exploration we can find the following information:

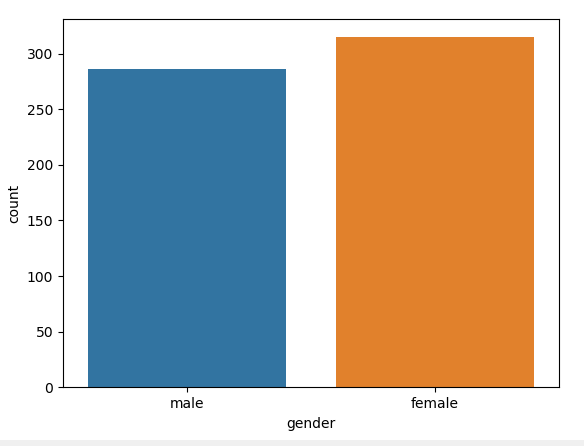
* The number of female is greater than males in the dataset
* Most of the adults have children

The following is the analysis of the categorical variables:

Variable **Gender:**

**Unique elements:**

array(['male', 'female'], dtype=object)

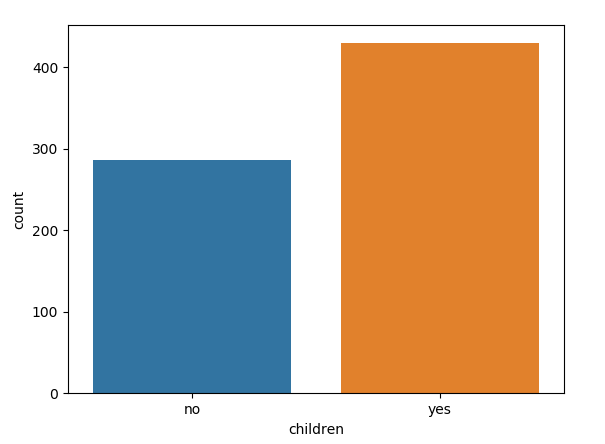


There are a greater number of females than males

**Children:**

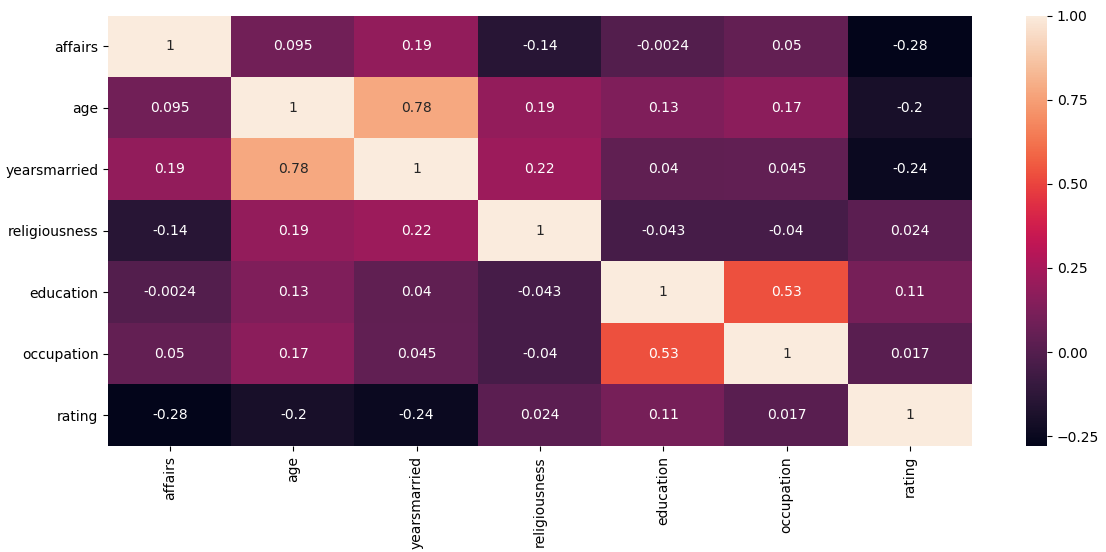
**Unique elements:**

array(['no', 'yes'], dtype=object)



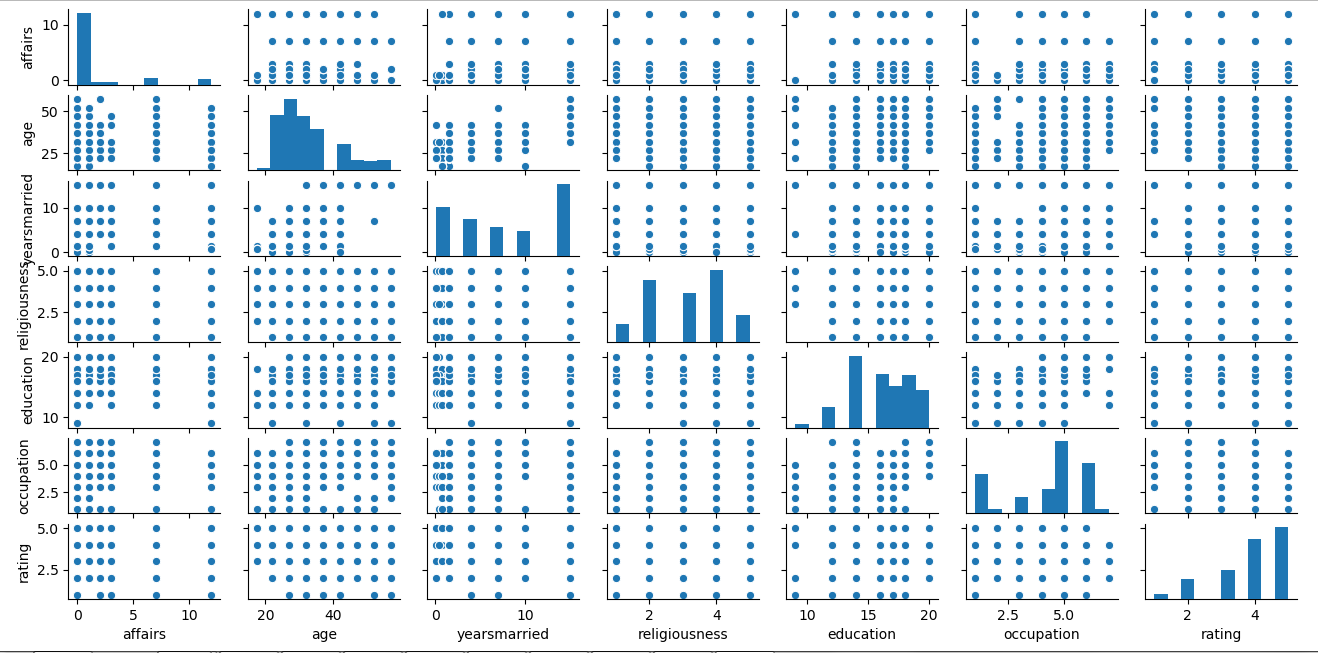
Most adults have children

The following is the correlation of the continuous variables:



According to the above heat map, the correlation between the variables is negligible

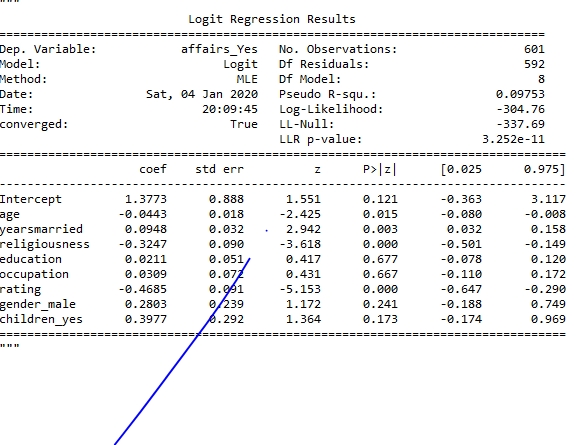
The following is the pair plot of the continuous variables:



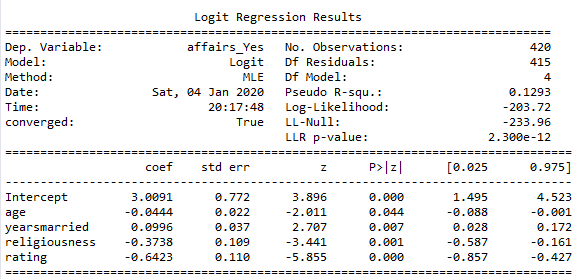
From the above pair plot, the following are the findings:

1. The variables affairs is strongly skewed
2. The scatterplots between the variables indicate that there is low or no correlation relationship between the variables

We split the data in train and test, the following is the result of the logistic regression

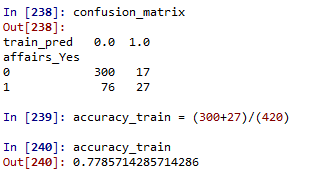


The above results consist of variables that are insignificant to the outcome. So, we remove these variables.



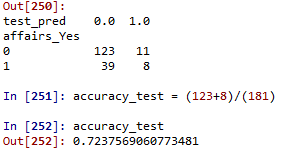
We split the data into test and train

**The confusion matrix of the train data is as follows:**



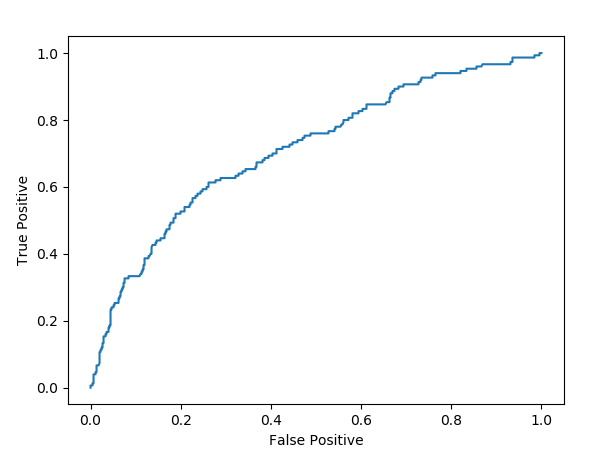
Train data provides an accuracy of 77.85%

**The confusion matrix of the test data is as follows:**



Test data provides an accuracy of 72.37%

The following is the ROC curve



The area under the curve is about 0.71