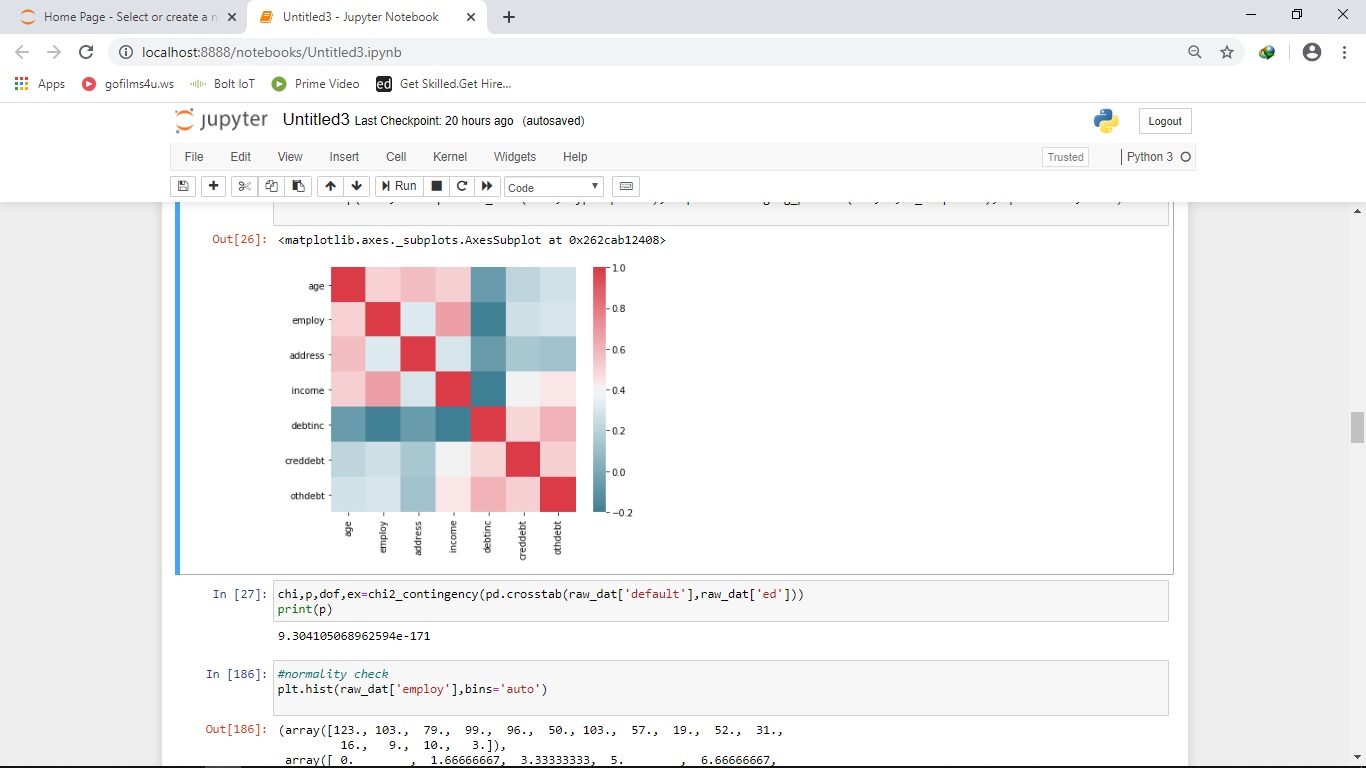
# Bank Loan Default Case

**The correlation heat map for the continous variables:**



From the above correlation heat map we can summarize that,

* EMPLOYMENT STATUS is negatively correlated to Individual’s debt Numerical payment to his or her gross income(debtinc).

Therefore, we can say that the debt payment of the individual is dependent on the his/her employment status.

* Income is negatively correlated to Individual’s debt Numerical payment to his or her gross income(debtinc).

Hence we can say that the debt payment of the individual is dependent on the his/her income.

* EMPLOYMENT STATUS is positively correlated to INCOME.

Which indicates that the income is dependent on the employment status of an individual.

* ANY OTHER DEBTS is positively correlated Individual’s debt Numerical payment to his or her gross income(debtinc).

Which shows that the other debts are also dependent on the payments of debts done by the individual.

**CHI-SQUARE test result for categorical variables:**

The critical value for the categorical variables **EDUCATION, DEFAULT** turned out to be less than 0.05, which indicates that the variables are dependent to each other.

**Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
| **Actual values\predicted values->** | **NO** | **YES** |
| **NO** | True Negative | False Positive |
| **YES** | False Negative | True Positive |

True Negative: The actual value is NO and the predicted value is NO (prediction is true)

True Positive: The actual value is YES and the predicted value is YES (prediction is true)

False Positive: The actual value is NO and the predicted value is YES (prediction is false)

False Negative: The actual value is YES and the predicted value is NO (prediction is false)

**Machine Learning Algorithm:**

We can select KNN algorithm for the development of the model as the **false positive rate is very low** compared to other algorithms, we also choose Logistic Regression as this algorithm also gives approximately same result to that of KNN.

Low **False positive rate** is **selected** for the model because the rate of error is low, where the predicted value is YES (indicating the acceptance of the loan) but the actual result turns out to be NO (the loan will not be accepted).

We can **ignore** the **False Negative rate** because the here predicted value is NO (the loan will not be accepted) but the actual result turns out to be YES (indicating the acceptance of the loan) which is a positive symbol for the profit of the bank.