After investigated the data, checking for data unbalancing, visualizing the features and understanding the relationship between different features, I have come up with predictive model that can predict the default of customers based on different parameters.

First I would love to discuss about some insights made from the data:

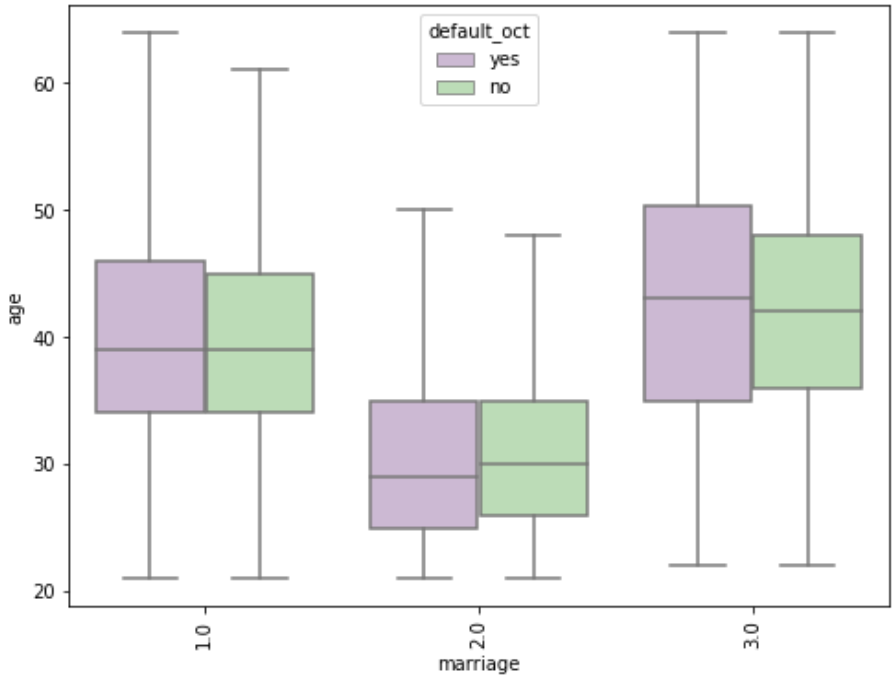
1. False filing of data
   It can be observed that some parameters like sex, marriage, pay were wrongly labeled as they contained values which were not documented

2. Missing values:
   Few customers did not have either pay_amt5, bill_amt5 nor pay_5 addressed at all.

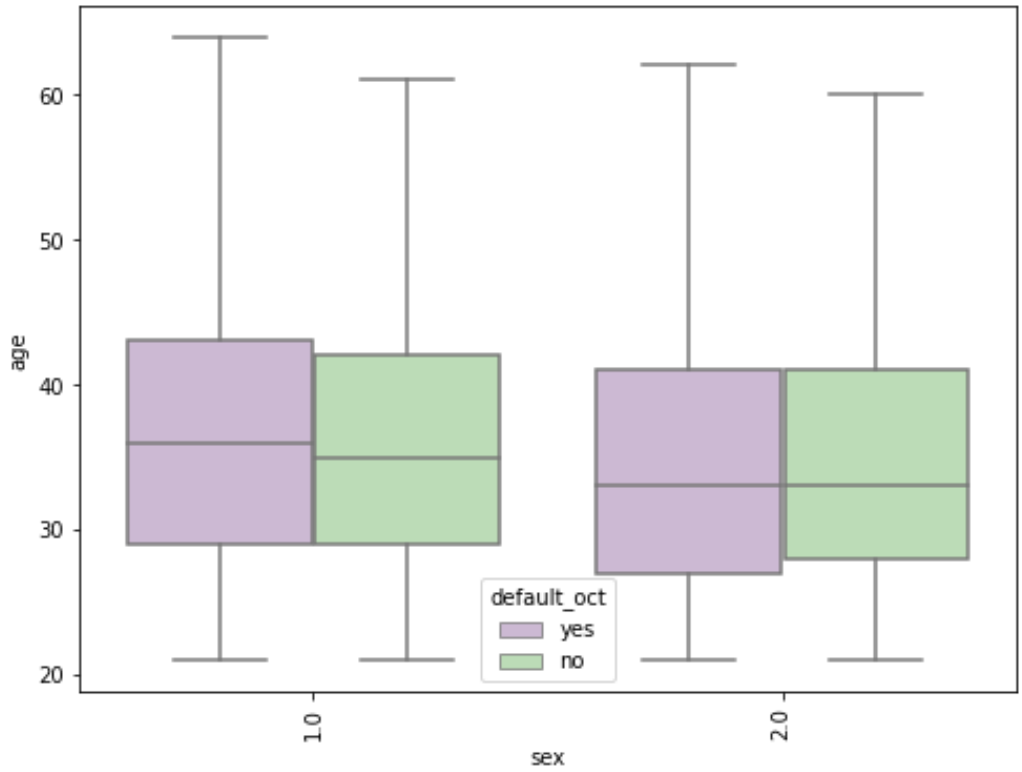
3. Married and “other” labeled people are most likely to default.

4. The higher is the education, the lower the probability of customer defaulting.

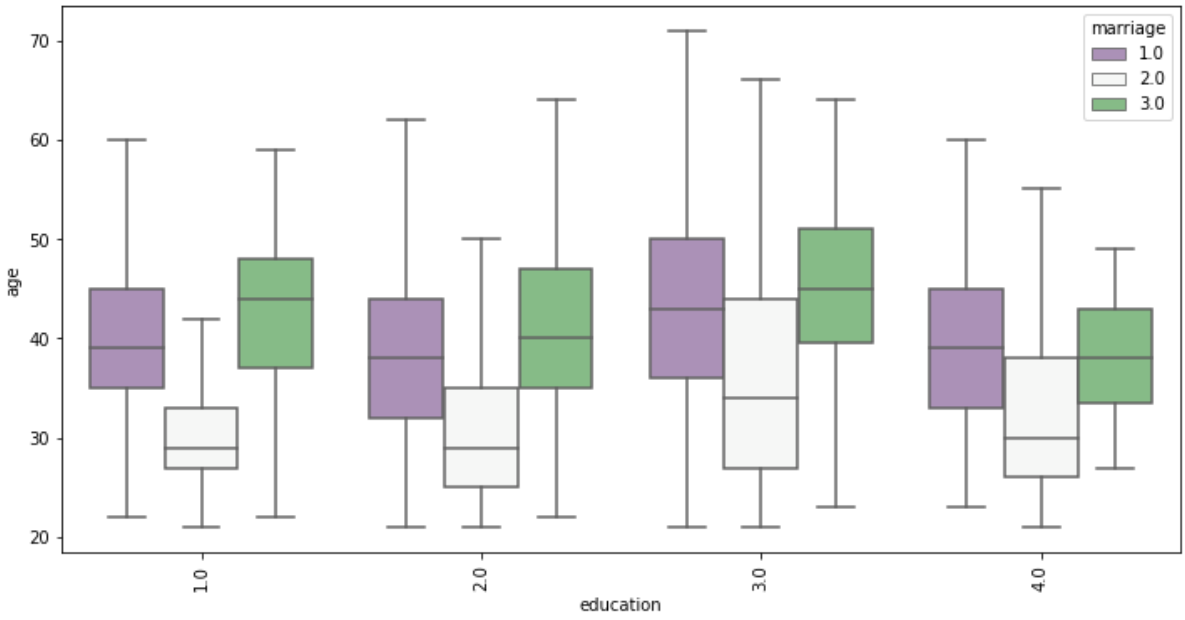
5. Customers who defaulted with “others” as marraige status were of high age.



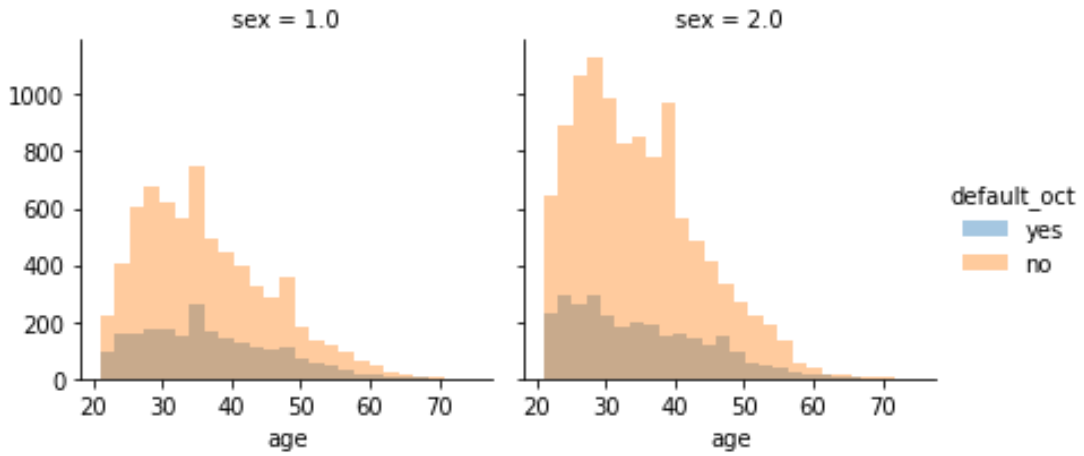
6. On an average, males were of higher age than females for both default and non-default categories.



7. Boxplots with age distribution grouped by education and marriage



8. From below observation we can conclude that age combined with other categories like marriage and sex gives more descriptive features. For example- Males do not default more between 25-40 age group, whereas women do not default more between 20-35 age group.



From this data then we can predict if a customer is going to default in Oct or not. After trying out multiple predictive models, I chose the one which provides the highest prediction accuracy.

	score
ADAboost	0.817157
RandomForest	0.813120
Decision Trees	0.817990
Bagging_DT	0.804783

In this way, the model will be able to predict beforehand which customer might default in the next months, and hence your financial services can concentrate only on those customers by providing them with lower-value card that supports balance management. At the same time the customers which are predicted not to default can have the high-LTV traditional card.

There can be again to approaches followed. One being by using a predictive model which will predict a customer defaulting with higher accuracy but might also contain customers who might not default (higher false positive). By this the financial service might be compromising with the penalizing few customers who would have not defaulted. The other approach being by choosing a predictive model which will predict a customer defaulting with a little lower accuracy but will not falsely predict the customers who won't default (lower false positives). In this way the not all defaulters will be predicted correctly, but at the same time non-defaulters will not be penalized.

Lastly, I would love to offer few other suggestions from the visualizations stated above. For example, customer categories with high defaults (married and lower educated customers) can be focused upon by offering a group focused education program, where they are taught about the consequences of defaulting, which will result in significant reduction of the number of defaulters.