**STEPS FOLLOWED IN THE PROJECT**

**Data Understanding**

The first step in the project data understanding.

There are total 7 columns in a dataset. Out of which there are 4 categorical columns and 2 continuous variables i.e. Var1, Var2.

Customer\_ID have unique values will not give any relevant information.

**Data Preprocessing**

Next step is Data preprocessing.

Before Data preprocessing we need to analyse the data.

Tested for Null values.There are no Null values in the dataset. Hance, NO missing value treatment necessary.

Verified for outliers, there are outliers in two columns i.e. Var1, Var2.

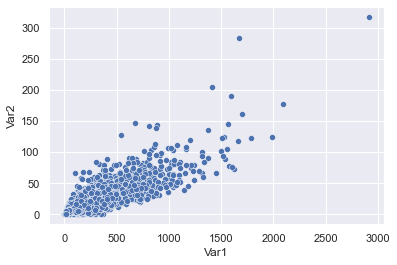
Outlier treatment has been done.

**Exploratory Data Analysis**

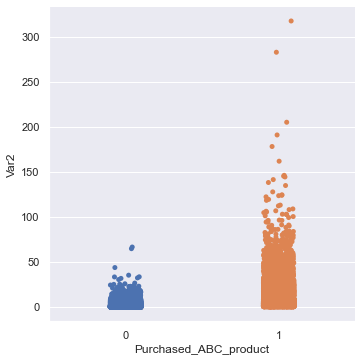
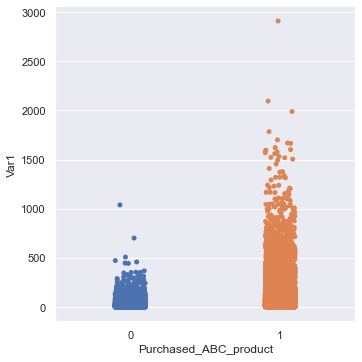
Next is Exploratory Data Analysis. Here we got some inferences.

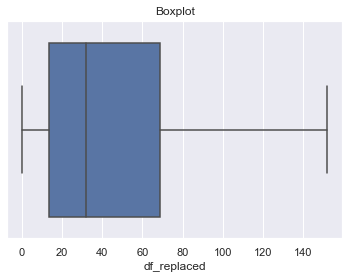
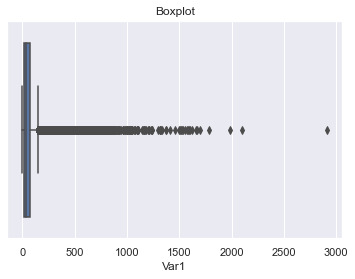
**Inferences**

There is positive correlation between Var1 and Var2.

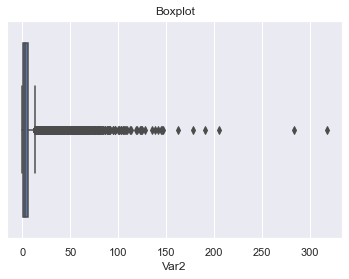
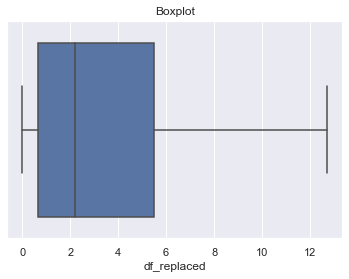


The customers purchased the product are more in number compared to who didn’t purchased the product.





Outliers in Var1 After Outlier treatment

Outliers in Var2 After outlier treatment

There are 4 categorical variables, created dummy variables for these categorical columns.

Output is the Categorical column ‘0’ and ‘1’.

This comes under supervised learning classification model.

**Train\_Test\_Split**

Splitting the data into train and test data.

80% Train data

20% Test data

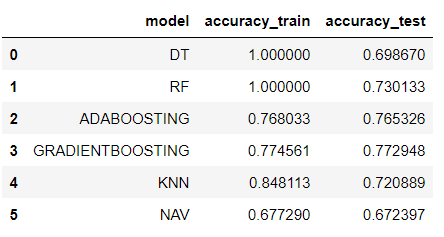
Trained the train data using different classification models.

**Modelling**

Different classification models applied

1. Logistic Regression
2. Navie bayes
3. K Nearest Neighbour
4. Ada boosting
5. Decision tree
6. Random forest

**Accuracy**



Out of all the Models, Gradient boosting gives the best accuracy.