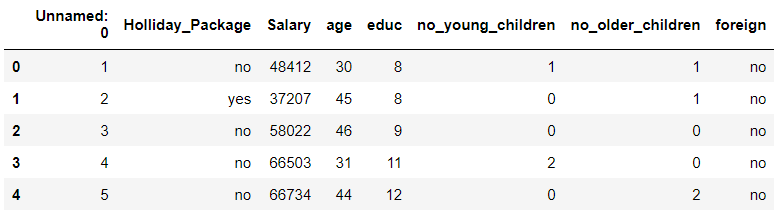
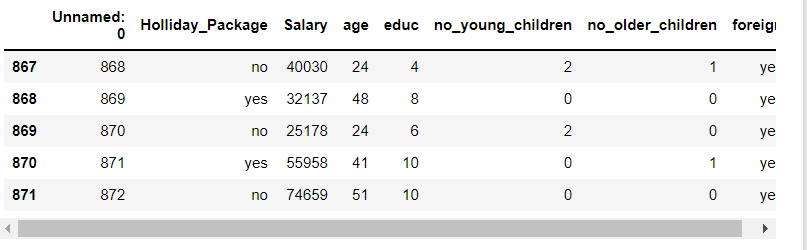
**1.1 Data Ingestion: Read the dataset. Do the descriptive statistics and do null value condition check, write an inference on it. Perform Univariate and Bivariate Analysis. Do exploratory data analysis.**

**HEAD**

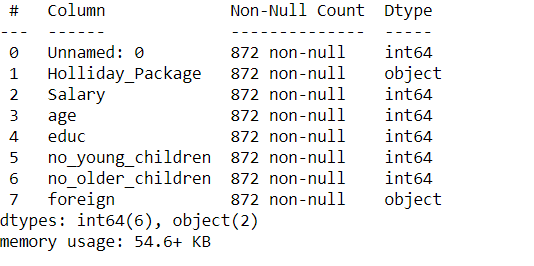
 **TAIL**



**SHAPE**



**INFO**

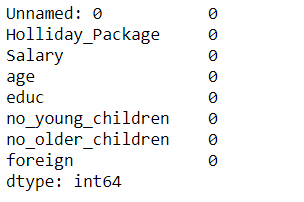


We have been given age, Salary, education, no\_young\_children, no\_older\_children, whether foreigner or not.

**DESCRIBE**



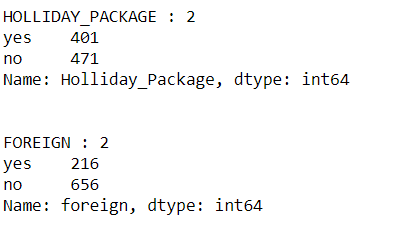
**NULL VALUES**



**DUPLICATES**



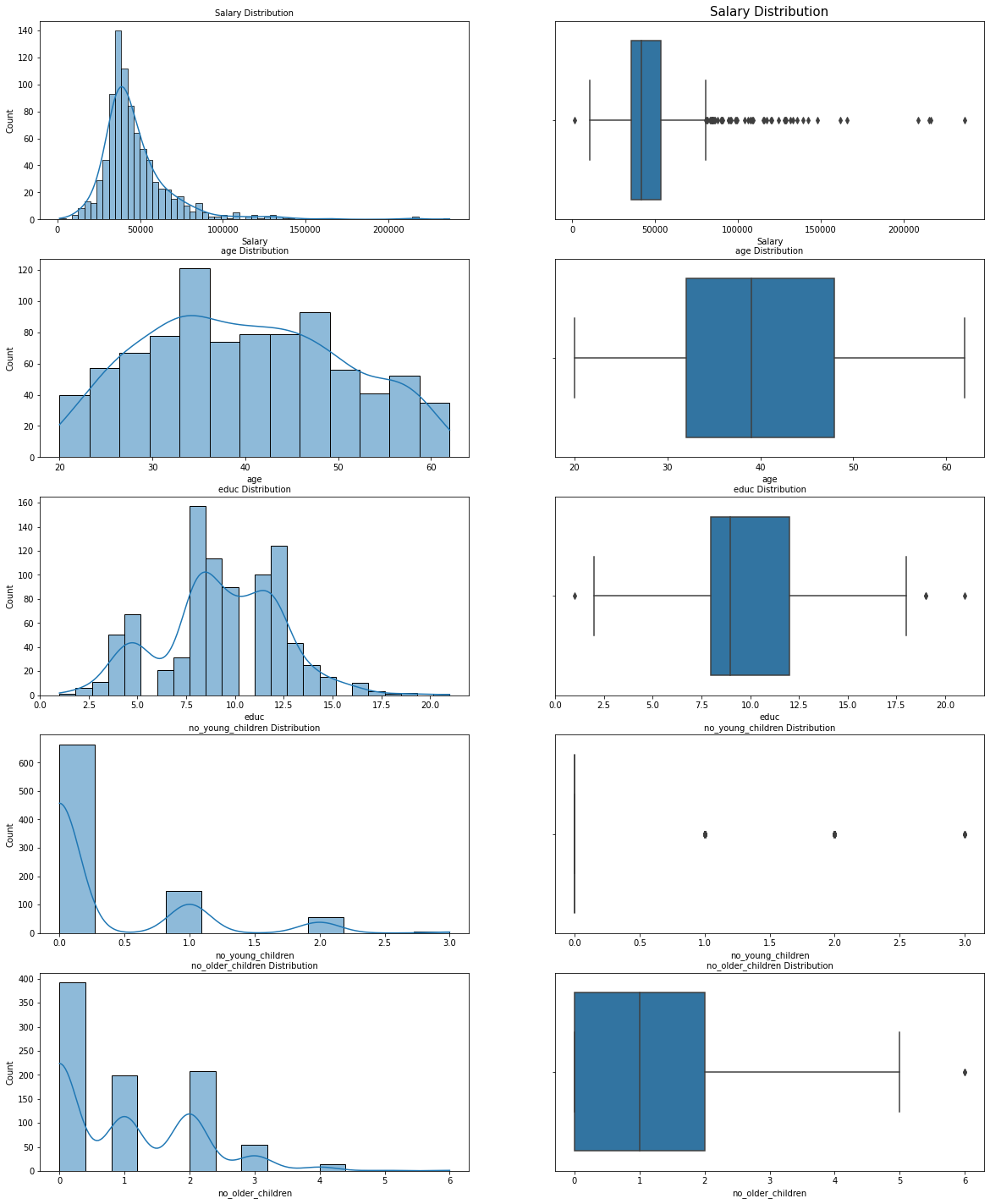
**CATEGORICAL VARIABLES**



**UNIVARIATE AND BIVARIATTE ANYSIS**

We have been given age, Salary, education, no\_young\_children, no\_older\_children, whether foreigner or not.

* Expected relationship in among independent variables:
  + Between age and Salary
  + Between education and Salary
  + Between Holliday\_package and Salary
  + Between age and education
* We’ll do univariate analysis followed by bivariate analysis.

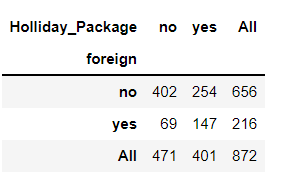
****

**Salary is heavily right skewed.**

**Both children’s distributions are right skewed but older children distribution is less skewed as compared to young children distribution.**

**There are a lot of outliers in the salary which is expected.**

**Age is normally distributed but not perfect.**

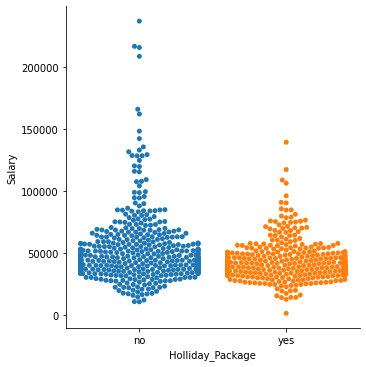


**This shows us that,**

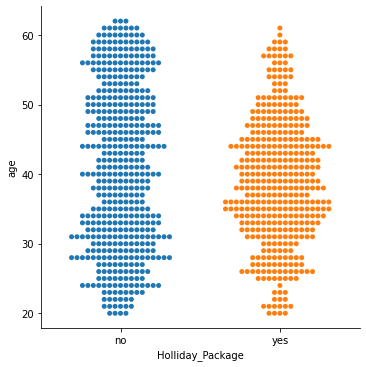
**Out of 216 foreigners, 147 opted for holiday package. (68.05 %)**

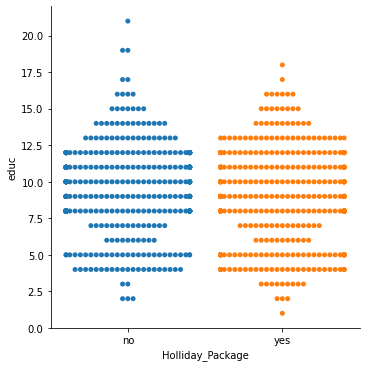
**Out of 656 non foreigners, 254 opted for holiday package. (38.71 %)**

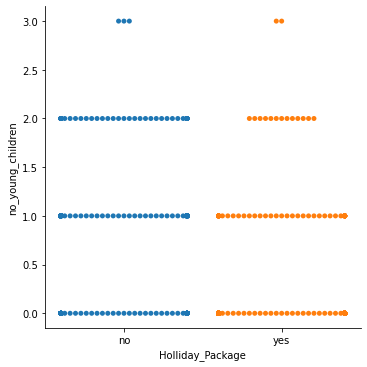
****

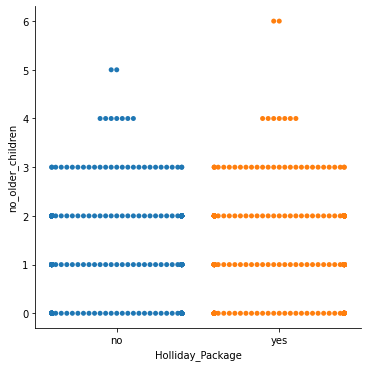
****

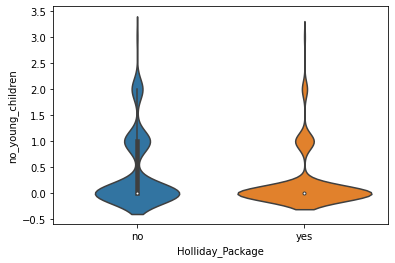
**Person with higher salary don’t have holiday package.**

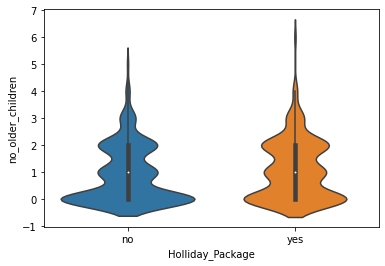
****

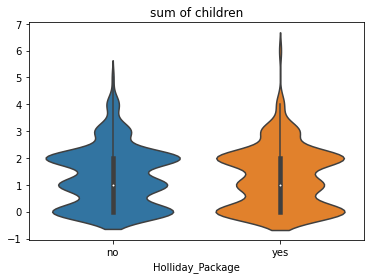
****

****

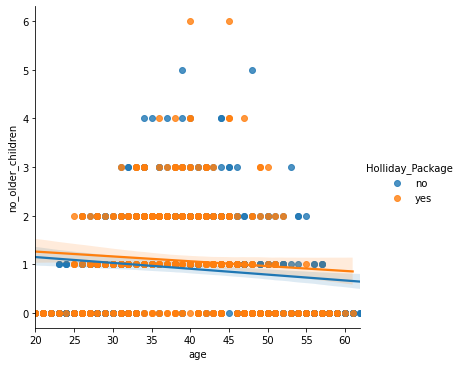
****

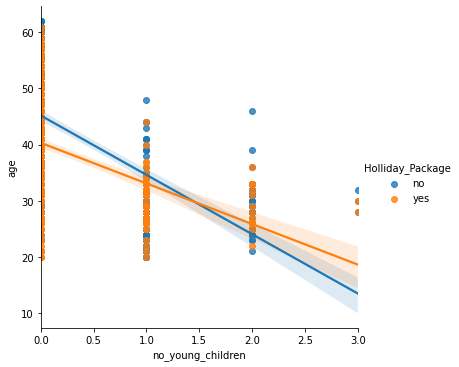
****

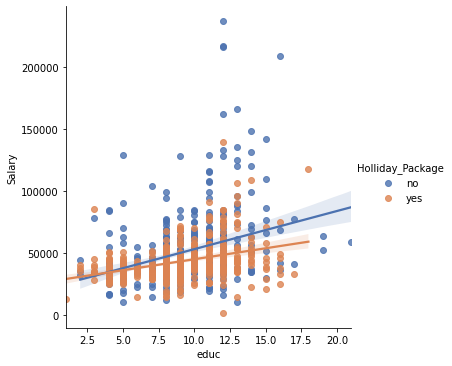
****

****

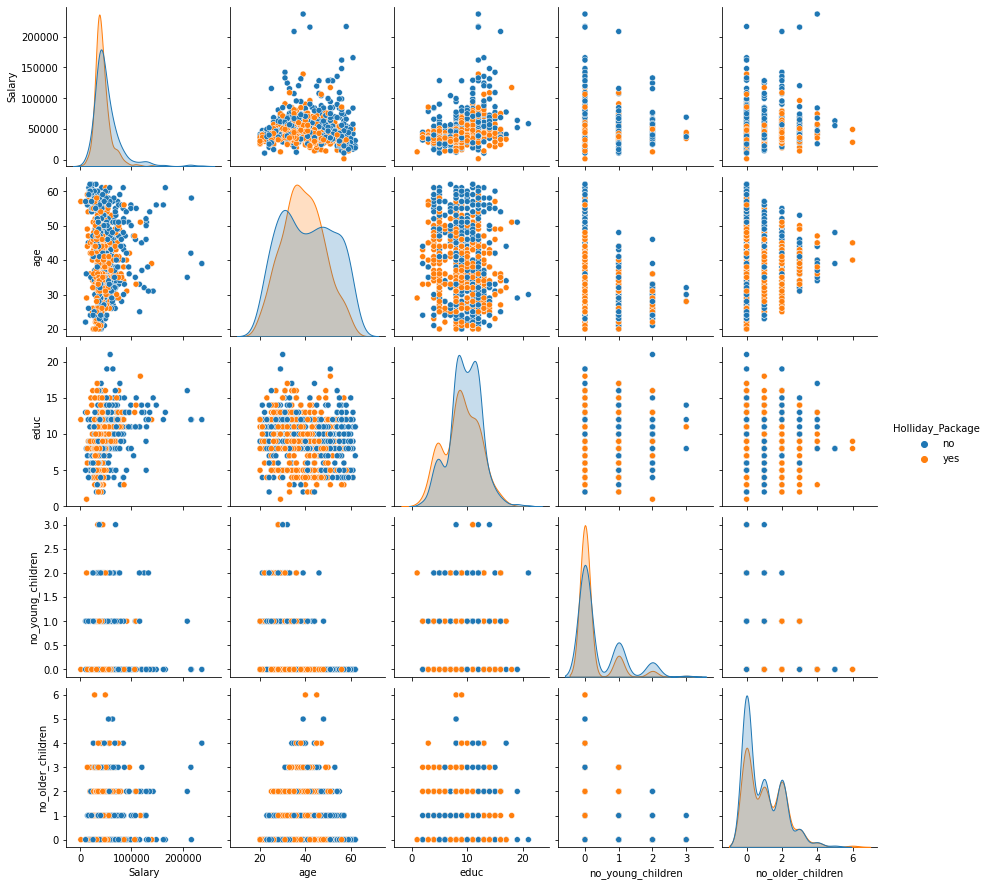
****

****

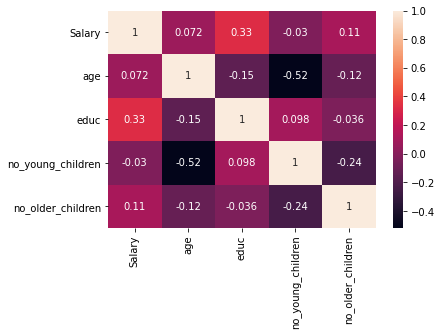
****

****

**Above a certain salary people don’t take holiday packages.**

****

**As no of young children increase. The holiday package decrease.**

****

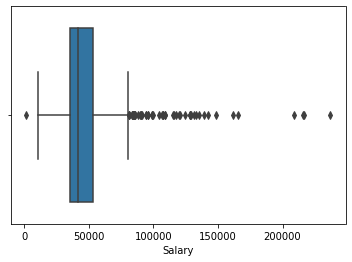
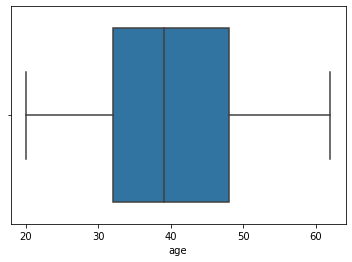
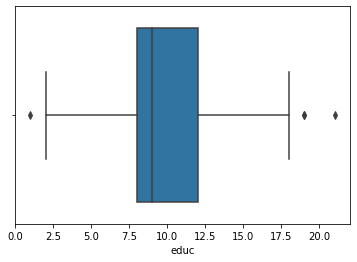
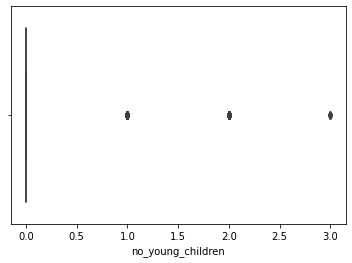
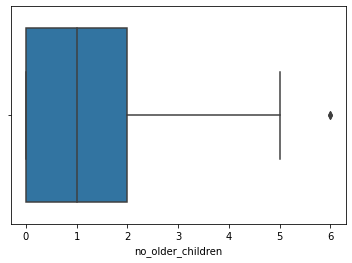
**No of young children significant negative correlation with age.**

**------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

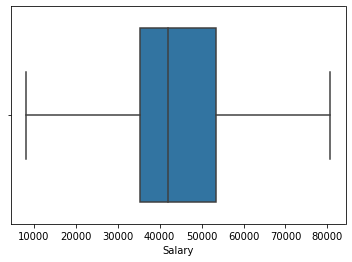
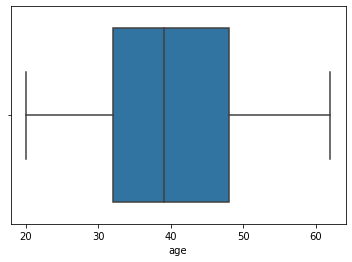
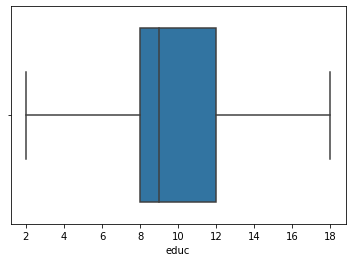
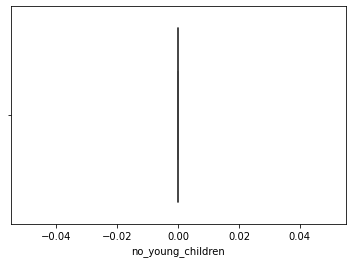
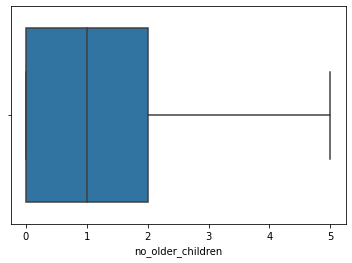
**1.2. Do not scale the data. Encode the data (having string values) for Modelling. Data Split: Split the data into train and test (70:30). Apply Logistic Regression and LDA (linear discriminant analysis).**

**1.3. Performance Metrics: Check the performance of Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and get ROC\_AUC score for each model Final Model: Compare Both the models and write inference which model is best/optimized.**

**BEFORE REMOVING OUTLIERS**

**    **

**AFTER REMOVING OUTLIERS**

**    **

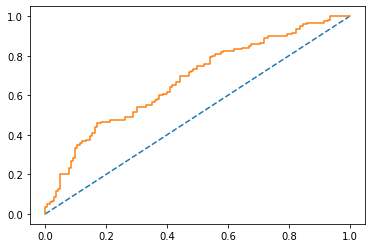
**LOGISTIC REGRESSION**

**Model score =** 0.6327868852459017

**AUC AND ROC**

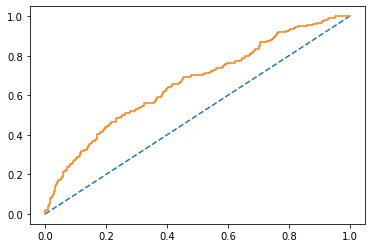
**TEST:**

AUC: 0.675

****

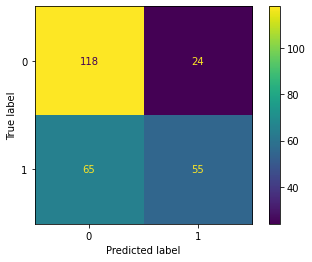
**TRAIN:**

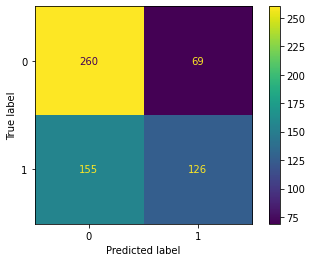
AUC: 0.661

****

**CONFUSION MATRIX**

**TEST:**

**  
TRAIN:**

****

**APPLYING GRID CV:**

**To increase our model’s efficiency**

grid={'penalty':['l1','l2','none'],

'solver':['sag','lbfgs','liblinear'],

'tol':[0.0001,0.00001]}

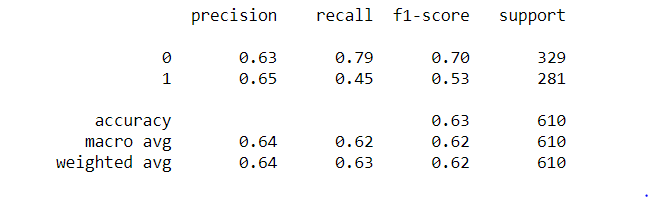
**BEST PARAMETERS AND BEST ESTIMATORS**

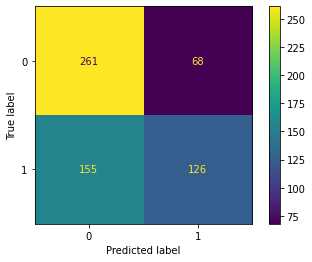
{'penalty': 'l2', 'solver': 'liblinear', 'tol': 1e-05}

LogisticRegression(max\_iter=10000, n\_jobs=2, solver='liblinear', tol=1e-05)

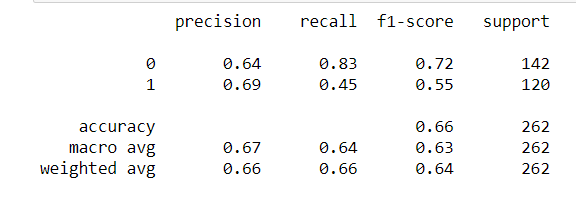
**CONFUSION MATRIX FOR BEST MODEL**

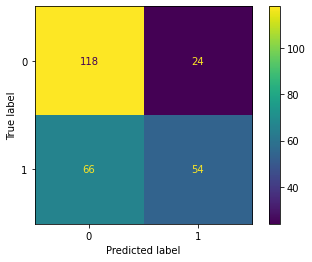
**TRAIN:**



****

**TEST:**



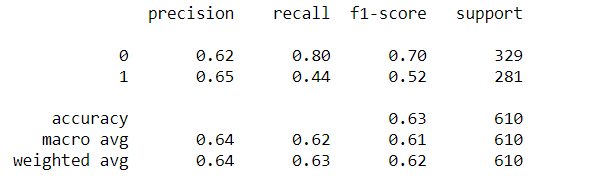
****

**LDA**

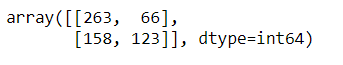
**Model score**

0.6327868852459017

**Classification report train**



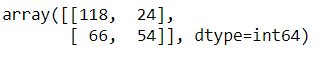
**Confusion matrix train**



**Classification report test**



**Confusion matrix test**



**Changing cut off value for better accuracy and f1 score**

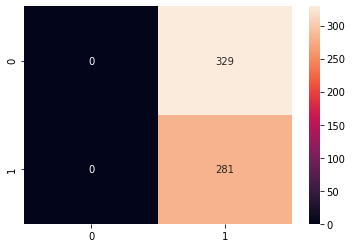
Changing cut off 0.1 to 0.9 with 0.1 increment

0.1

Accuracy Score 0.4607

F1 Score 0.6308

Confusion Matrix

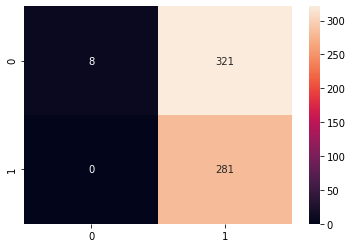


0.2

Accuracy Score 0.4738

F1 Score 0.6365

Confusion Matrix

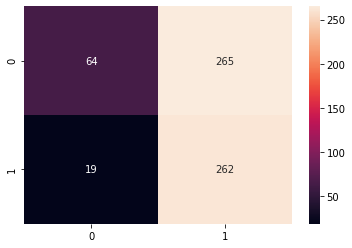


0.3

Accuracy Score 0.5344

F1 Score 0.6485

Confusion Matrix

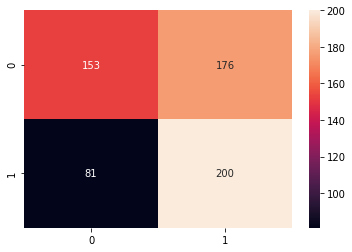


0.4

Accuracy Score 0.5787

F1 Score 0.6088

Confusion Matrix

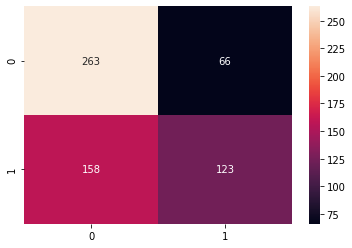


0.5

Accuracy Score 0.6328

F1 Score 0.5234

Confusion Matrix

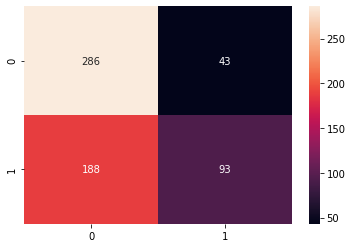


0.6

Accuracy Score 0.6213

F1 Score 0.446

Confusion Matrix

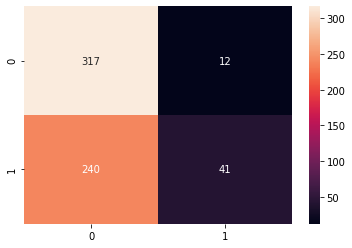


0.7

Accuracy Score 0.5869

F1 Score 0.2455

Confusion Matrix

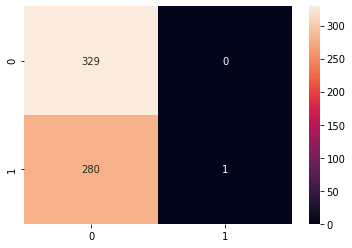


0.8

Accuracy Score 0.541

F1 Score 0.0071

Confusion Matrix

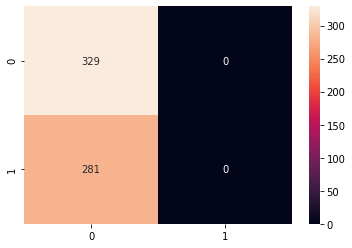


0.9

Accuracy Score 0.5393

F1 Score 0.0

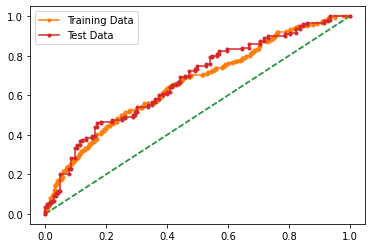
Confusion Matrix



**Auc and roc for test and train**

AUC for the Training Data: 0.661

AUC for the Test Data: 0.675

****

**Output of both the LDA and logistic regression is same.**

**------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

**1.4. Inference: Basis on these predictions, what are the insights and recommendations.**

**We noticed the patterns listed below:**

* **People above 50 and below 25 don’t take holiday package.**
* **Foreigners seem to be taking packages more:**
  + **Out of 216 foreigners, 147 opted for holiday package. (68.05 %)**
  + **Out of 656 non foreigners, 254 opted for holiday package. (38.71 %)**
* **Top tier salary men don’t take holiday package.**
* **No of younger children inversely proportional to holiday package.**

**Recommendations:**

* **Make tour specifically to the interest of young and old people like for younger people we can arrange cheap getaway destinations, where as for old people we can arrange spiritual or calming destinations.**
* **We can encourage people with high number of younger children for some combo package with cheaper alternatives or provide facilities with child care.**

**------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**