**Assessment 1**

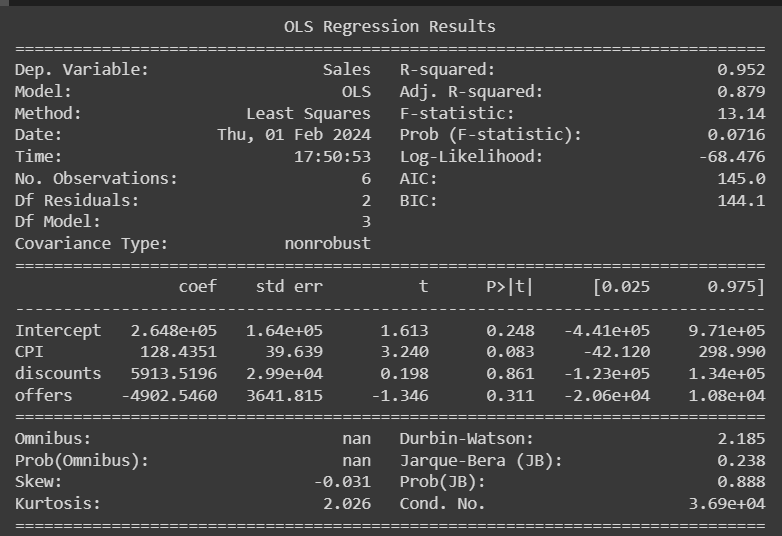
**Model used :** Multiple linear regression

In this case we have used multiple linear regression model to predict the Sales of an organization based on the variables CPI, discounts and offers.

**Dependent variable :** Sales

**Independent variable :** CPI, Discounts and Offers

The model developed has a good accuracy with a R-square value of **0.95**



**Results:**

1.Predicted sales value for 5000 CPI , 3 percentage discounts, 20 rewards offers = **826645**

1. Predicted sales value for 4000 CPI , 8 percentage discounts, 19 rewards offers = **732680**

**Assessment 2**

**Model used :** Logistics regression

Logistics regression Machine learning model is used to predict whether a loan can be offered to a customer by assessing their Cards, debit card, insurance, age and cibil score data.

We have used confusion matrix and ROC curve to assess the accuracy of the model.

Below are the details of the model developed,

**Accuracy :** 70.8%

**Error rate :** 29.1%

**Total number of customer (N) :** 1096

**True Positive (TP) :** 381

**False Positive (FP) :** 197

**False Negative (FN) :** 123

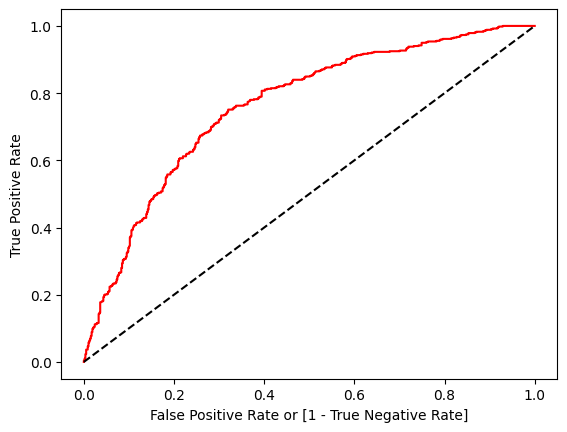
**True Negative (TN) :** 395

**Specificity :** 60.7 %

**Sensitivity :** 75.5%

**False Negative Rate (FNR) :** 24.4%

**False Positive Rate (FPR) :** 33.2%



**Conclusion:**

With the ROC curve we can conclude this is a perfect model and as it has a good accuracy of 70.8% we can use this model to predict a loan can be offered to a customer.

**Assessment 3**

Here we have used multiple customer details and personal information to categorize the customer whose salary is lesser than or equal to 50k or greater than 50K

We have achieved this by using multiple models, Below are the list of ML models built using the data and their results.

**Model name : K Nearest Neighbor**

* We have build the model by first cleaning the data given, we have removed the columns which are not required or does not contribute to the model and then we have filled the values which are not available and have cleaned the data which can be used to build the model.
* Then we have encoded the categorical values to numeric values which can be used to build the model by using Label encoder.
* We have checked the correlation between the columns by heatmap to see which columns contribute more and which does not contribute much.
* Then we have splitted the data for testing and training with 20% and 80% respectively
* Then we have fitted the model using the train data.

KNN model has been built with a **accuracy score** of **0.79**

**Model name : Support Vector Machine**

* We have build the model by first cleaning the data given, we have removed the columns which are not required or does not contribute to the model and then we have filled the values which are not available and have cleaned the data which can be used to build the model.
* Then we have encoded the categorical values to numeric values which can be used to build the model by using Label encoder.
* Then we have splitted the data for testing and training with 20% and 80% respectively
* Then we have fitted the model using the train data.
* SVM separates the customers into two categories whose salary is lesser than or equal to 50K and greater than 50K with a separator

SVM model has been built with a **accuracy score** of **0.81**

**Model name : Decision Tree**

* We have build the model by first cleaning the data given, we have removed the columns which are not required or does not contribute to the model and then we have filled the values which are not available and have cleaned the data which can be used to build the model.
* Then we have encoded the categorical values to numeric values which can be used to build the model by using Label encoder.
* Then we have splitted the data for testing and training with 20% and 80% respectively
* Then we have fitted the model using the train data.
* Decision tree helps us to predict or make a decision about the salary separation

Decision Tree model has been built with a **accuracy score** of **0.82**

**Model name : Random Forest**

* We have build the model by first cleaning the data given, we have removed the columns which are not required or does not contribute to the model and then we have filled the values which are not available and have cleaned the data which can be used to build the model.
* Then we have encoded the categorical values to numeric values which can be used to build the model by using Label encoder.
* Then we have splitted the data for testing and training with 20% and 80% respectively
* Then we have fitted the model using the train data.
* Random forest helps us to predict or make a decision about the salary separation

Random forest model has been built with a **accuracy score** of **0.81**

**Model name : K means clustering**

* We have build the model by first cleaning the data given, we have removed the columns which are not required or does not contribute to the model and then we have filled the values which are not available and have cleaned the data which can be used to build the model.
* Then we have encoded the categorical values to numeric values which can be used to build the model by using Label encoder.
* Then we have splitted the data for testing and training with 20% and 80% respectively
* Then we have fitted the model using the train data.

K means clustering model helps us to cluster the data given into clusters, I this case we have clustered the customer data into two categories as needed.

**Note:** Model built ipynb file will be uploaded to Github for the model details and data with all the details present.