NEXUS BANK CASE STUDY

CUSTOMER SEGMENTATION AND DEPOSIT DETECTION SYSTEM

AIM OF THE PROJECT

- The problem to be solved with data mining is understand customer's behavior and improve term-deposits.

- The goal is to develop a machine learning system that will help the bank segment their customers and also be able to detected if a customer will make a term-deposit or not.

Benefits of the Deposit Detection System and Customer Segmentation System

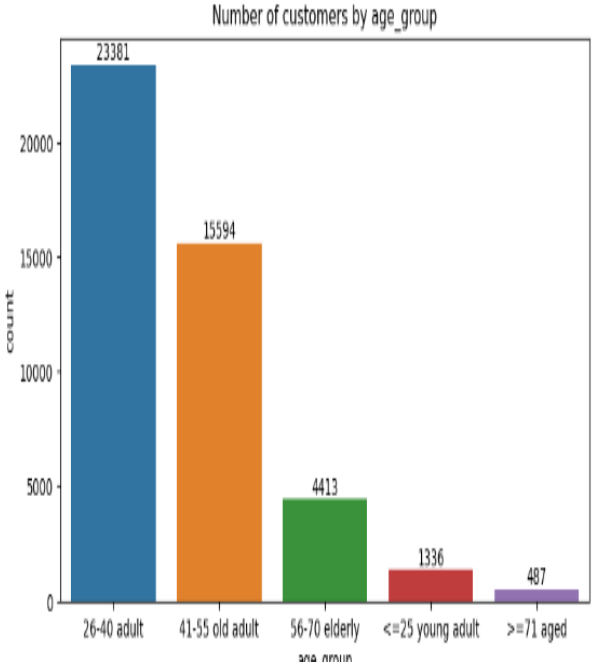
The following are the what Nexux Bank will gain from the customer segmentation system and deposit detection system

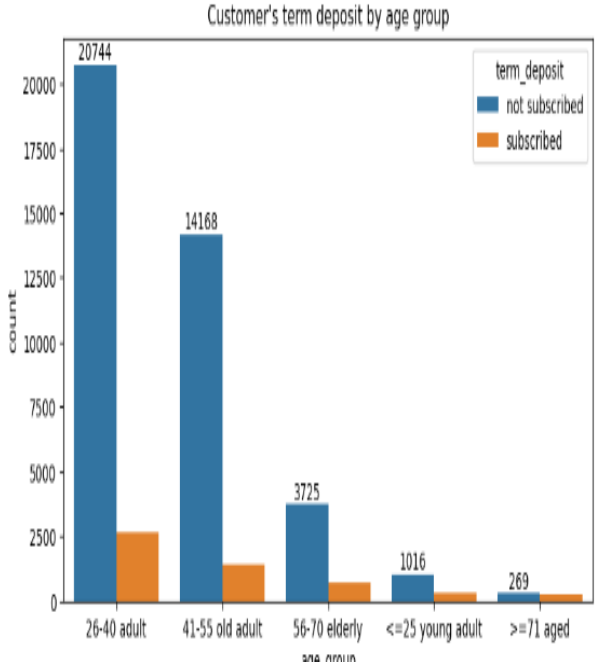
- Improve relationship with Customer
- Ability to create solutions that meets individual's needs
- Optimise operations by minimising defaults and improving customer deposits
- Promote a better and more effective marketing campaign strategy
- Improve bank's efficiency

EXPLORATORY DATA ANALYSIS

UNIVARIATE ANALYSIS

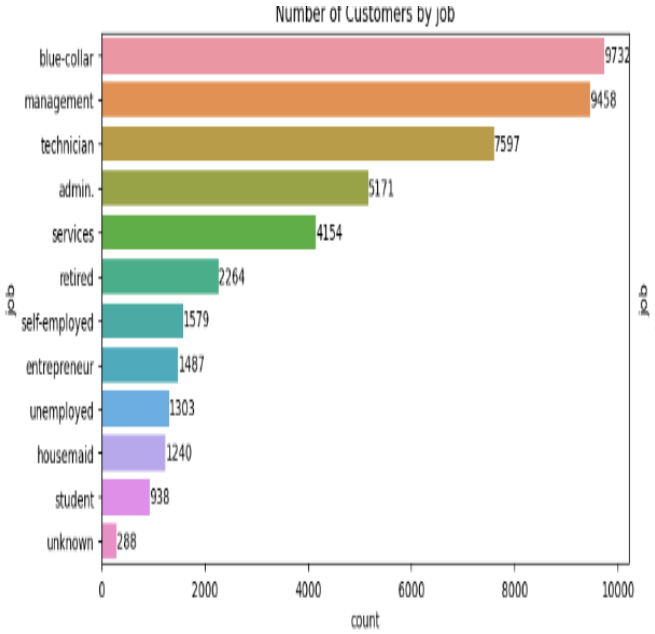
Showing the relationship between customer's demographics and their likelihood of subscription to term-deposit.

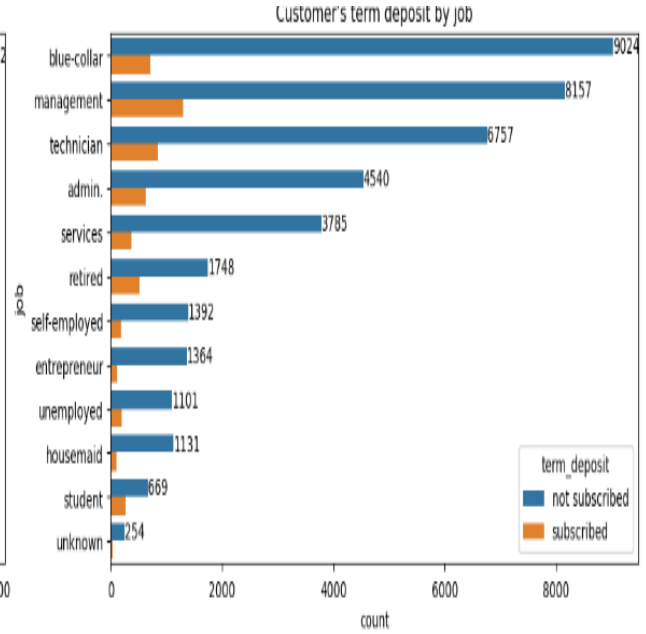




Above: Majority of the customers are with the adult age(26-40year-old),and about 10% of them, 23% of the young adults and 50% of the aged subscribed to term-deposit.

Below: Most of the customers have blue collar jobs. But those with management job, subscribed most followed by blue collar and technician. The least are enterpreneurs and housemaids, strangely the unemployed subscribed more than them

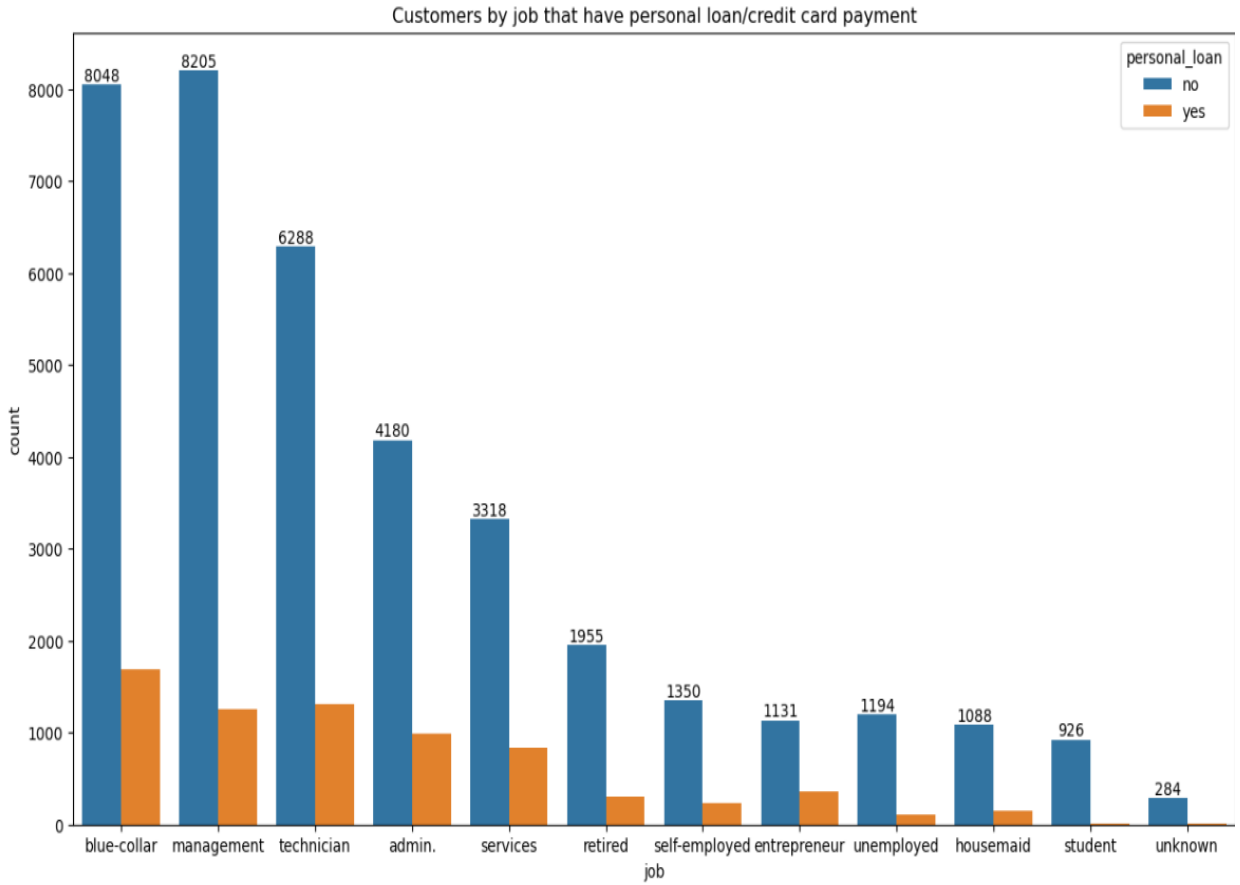


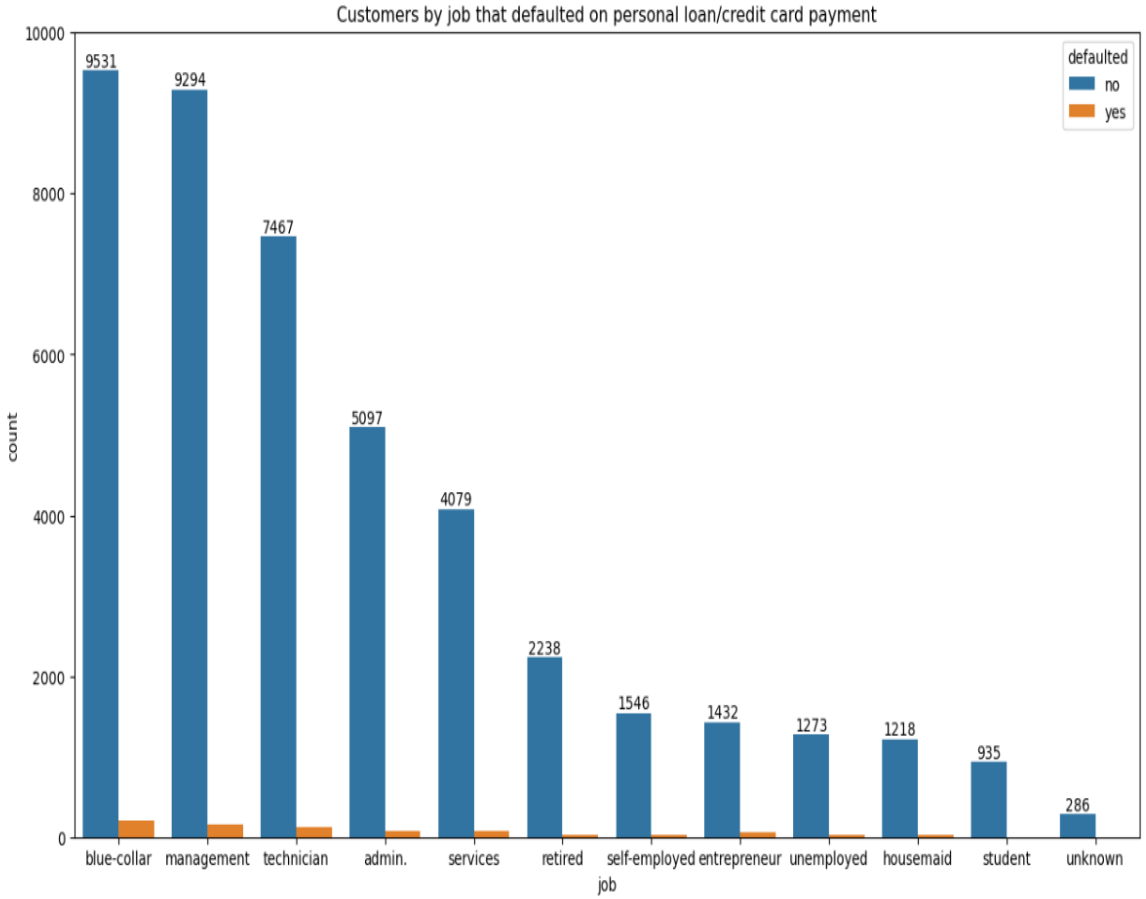


UNIVARIATE ANALYSIS

Showing the relationship between customer's demographics and attitude towards defaulting

About 201 out of 1,684 blue collar job holders with personal loan have defaulted (approx.12%), 13% and 10% defaulted in the management and technician categor respectively.
But contrary to the above the percentage of defaulter were seen to be higher among those that are unemployed and students, having about 28% and 25% of them defaulting. Although the only 12 students were given personal loan.



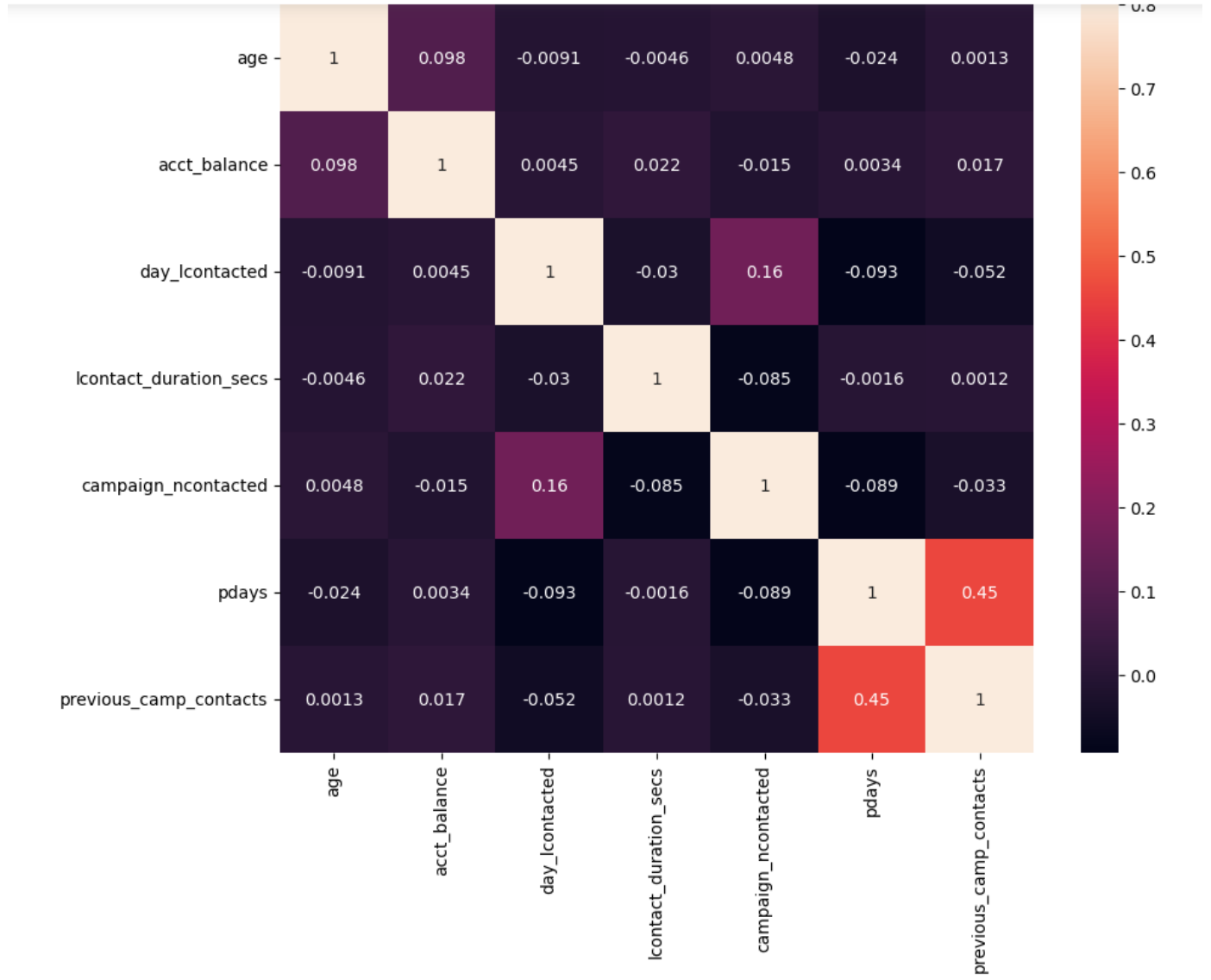


CORRELATION/MULTIVARIATE ANALYSIS

Insights:
From the heatmap, we can see that there is a positive relationship between

-Days that passed after the customer was last contacted from previous campaign and number of calls made before this campaign

-Day of the month when customer was last contacted and number of calls made in this campaign

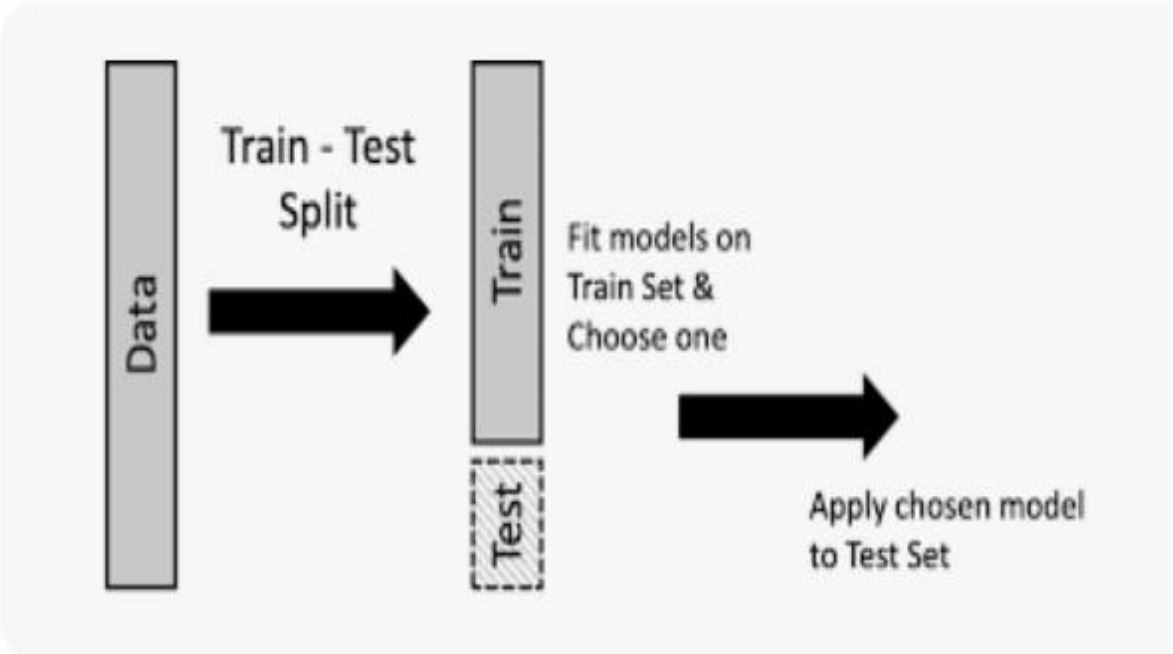


DATA PREPROCESSING

- Feature Engineering- Split Data into independent and dependent variable
- Data Normalisation- Using the MinMaxScaler
- Encoding – LabelEncoding to change categorical data to numerical data

BUILDING DEPOSIT DETECTION MODEL

TRAIN-TEST SPLIT IS 80% : 20%

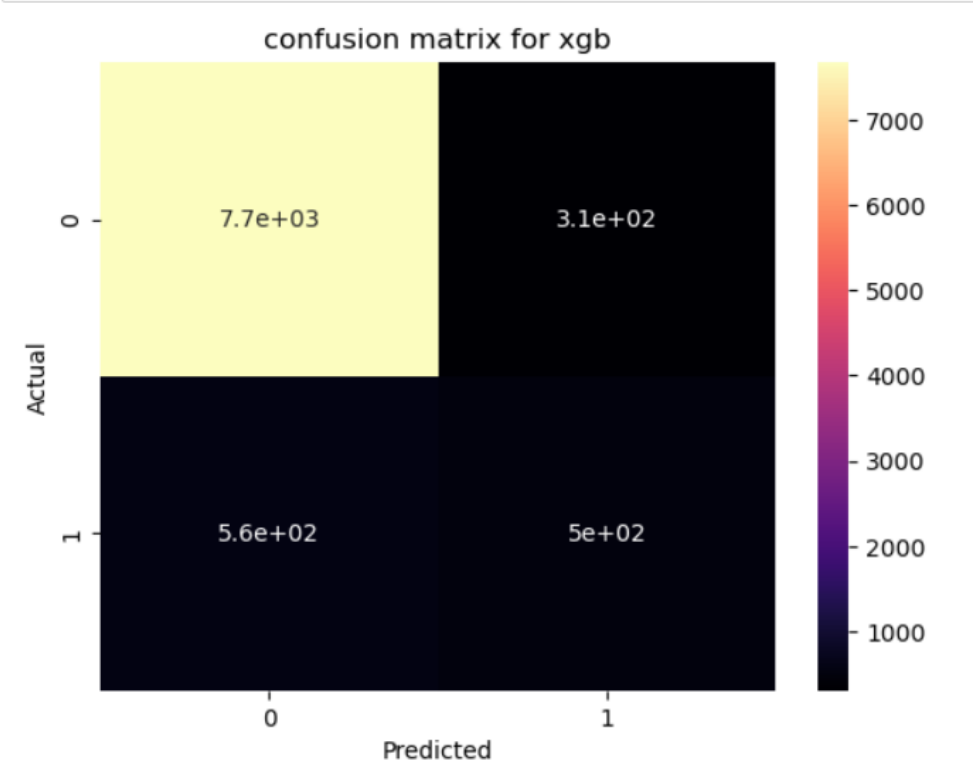


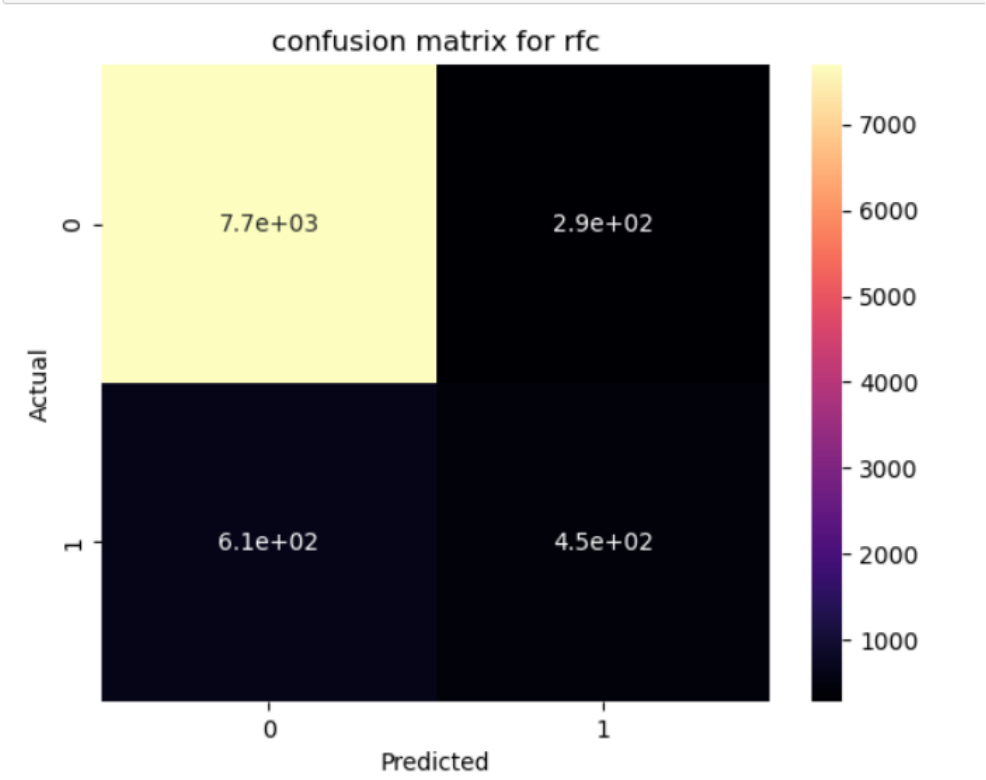
MODEL EVALUATION

Evaluation Metrics	SGD Classifier	K-Neighbors Classifier	Random Forest Classifier	Logistic Regression Classifier	XGB Classifier	Naive Bayes	SVC	Decision Tree Classifier
accuracy	88.67%	87.9%	90.05%	88.48%	90.38%	83.77%	88.25%	86.64%
precision	57.92%	44.98%	61.0%	57.05%	61.9%	35.73%	0.0%	43.52%
recall	13.08%	13.08%	42.52%	8.0%	47.22%	47.7%	0.0%	45.81%
roc	55.91%	55.47%	69.45%	53.6%	71.68%	68.13%	50.0%	68.95%
f1_score	21.34%	20.26%	50.11%	14.03%	53.58%	40.85%	nan%	44.64%

CONFUSION MATRIX

Below: The lower the FN that is false negative (type 2 error) the better the model . Therefore the best model here is the XGB Classifier, the false negative is 560 and its f1_score is 53%.





XGB CLASSIFIER- F1 SCORE(53%)

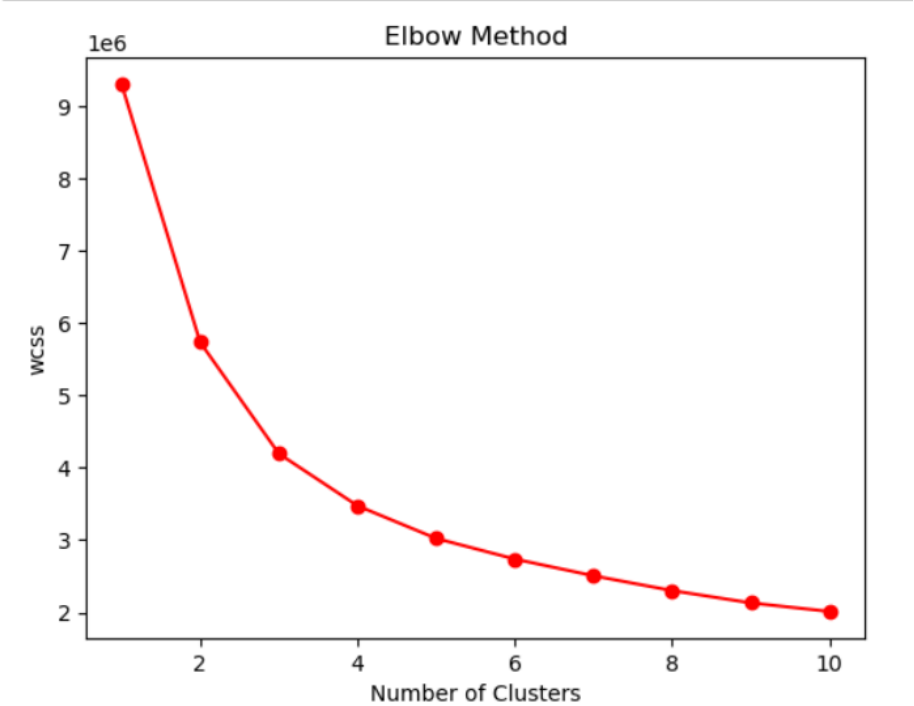
RANDOM FOREST- F1 SCORE(50.11%)

DATA PREPROCESSING

- Data Normalisation of the dataset- Using the MinMaxScaler
- Encoding – LabelEncoding to change categorical data to numerical data

BUILDING CUSTOMER SEGMENTATION MODEL

UNSUPERVISED MACHINE LEARNING USING KMEANS CLUSTERS



MODEL EVALUATION

DIFFERENT VARIABLE WERE B

Number of clusters was defined using the Elbow method, 2 clusters was used because it gave a higher silhouette coefficient of 0.3447, compared to using 3 clusters.

Cluster 1- No personal loan, 43year-old and above, Retired, housemaid, married.

Cluster 2- No personal loan, 18-42year-old, Students, Single, primary and secondary as highest level of education

RECOMMEDATIONS

Base on the insights: -
The bank should have customized products that will serve the clusters of customers e.g student/school fees term-deposit for the students and retirement term deposit for the aged/retired people. With this the bank will be able to meet the needs of targeted customers and also get new customers to save with the bank

In this campaign the customers that were contacted within 1 to 4 times over 1st to 31st of the month subscribed to term-deposit, therefore its advisable to limit the marketing campaign calls to less than 4 to avoid unproductive calls and minimize cost of marketing.

A short term maturity term-deposit should be packaged for those in the low income earners carder

Challenges

Information on the Income status of the customers were not provider, This would have enabled one to understand their spending capacity .

THANK YOU

FROM-

PHARM.(MRS.) MARY KENN-NWOKORO

DATA SCIENTIST

10ALYTICS