A study on in-depth analysis on banks products and services and exploring the attributes affecting the banks overall performance.

Machine Learning.

Name: M.R.Usmaan

Table of Contents

Chapter 01: Introduction

1.1 Background

The banking sector is one of the important aspects of the economy and it offers countless services to its consumers. Main aspect of the banking sector is their consumers so, bank need to have proper understanding about their consumers preferences to achieve their goal and build the trust upon people. So, focusing this our research is going to study on in-depth analysis on banks services and going to explore the attributes affecting the bank overall performance. In this study we going to explore the features that contribute to customer fixed deposit subscription plan and telemarketing call duration. We are going to group the customers according to their risk profile. So, at the last of this research we going to give recommendation to banks by analyzing our research results so it will be very useful to bank organizations to retain their customers.

Bank organizations must understand the customer behavior and they need to know which factors effect to customer subscription plan and telemarketing call duration. In depth, it is important to analyze which customer group has a risk profile. So, focusing on these the proposed research seeks to explore and analyze in-depth analysis on banks products and services and exploring the attributes affecting the banks overall performance.

1.2 Research Problem

At present, customers are stopping and quitting their relationships with bank, and they use alternative ways for their financial needs. For an example they invest their money in trading and real estate, and they invest in cryptocurrency. So, this activity of consumers hast the biggest negative impact in bank organizations. And bank resources are not planning properly to take effective output from those resources, finally most of the banks which don’t have proper insights about their customers facing many financial issues so these banks should consider to identify the risk profile of the customers. The goal of this research study is to explore and analyze in-depth analysis on banks products and services and exploring the attributes affecting the banks overall performance.

1.3 Research Questions

I. What are the features that affect the decisions of customers to make deposits in bank and which model is suitable to predict it?
II. What are the factors that affect telemarketing phone call duration that makes to customer and which model is very suitable to predict call duration?
III. Which group of customers have risk profile and how we can reduce the risk?

1.4 Objectives

I. Explore and analyze the factors that affect decisions of customers to make deposits in bank and build a model to predict customer decision regarding deposit.
II. Explore the factors that affect telemarketing phone call duration that makes to customer.
III. Identifying the risk profile of groups of customers who have default credit based on age and bank balance of them.

1.5 Expected Limitations

This research study is strongly dependent and based on secondary dataset. So, according to that there is no proper idea about how the data have collected, which bank's data is this, what type of the currency this data have used, gender of the customer, which years data is this and there are some unknown classes in this category, so these have big negative impact to research study.

CHAPTER 2: Literature Review

2.1 Theoretical explanation about the key words in the topic

Key word	Explanation	Reference
Convolutional neural network	A type of artificial neural network used primarily for image recognition and processing	(arm limited.grossary/webpage)
Artificial neural network	An attempt to simulate the network of neurons that make up a human brain so that the computer will be able to earn things and make decisions in a human like manner.	Wikipedia
Data mining models	A set of data statistics and patterns that can be applied to new data to generate predictions and make inferences about relationships	Microsoft.com

2.2 Findings by other researchers

This article introduces a deep learning approach, specifically a deep convolutional neural network, to predict the success of bank telemarketing. The model built on a dataset of 45,211 phone calls, achieves a notable accuracy of 76.70%, outperforming conventional classifiers. The focus is on leveraging relationships among attributes and hierarchical features, with suggestion for future exploration in financial instruments and comparison with other recommendation algorithms. **(predicting the success of bank telemarketing using deep CNN)**

Exploring borrower preferences in loan repayment plans, this article identifies a positive correlation between financial literacy, capability , and the preference for increasing installments. Borrowers are classified into low, medium and high levels for both traits, with recommendations for specific repayment plans based on borrower profiles, considering interest rates and risk factors**.(An analysis of loan repayment plans according to the bank customer profile)**

Highlighting the significance of telemarketing in banking , this research employs AI prediction models, including logistic regression ,decision tree and support vector machine. Results indicate high prediction accuracy , with logistic regression outperforming, emphasizing the effectiveness of telemarketing and the potential of AI – assisted models for market segmentation and promotion in the banking sector**.(Applying AI techniques to predict the success of bank telemarketing)**

Using data mining techniques, particularly classification and regression tree analysis, this study categorizes customers applying for loans into five risk profiles. The goal is to help banks strategically manage customers, enhance competition, and achieve annual targets , emphasizing the significance of understanding customers profiles for informed decision – making**. (data mining applications).**

This article explores the impact of mobile banking on the industry , identifying 17 key dimensions of service quality base don 803 customer reviews from top US banks. Practical advice for mobile banks includes focusing on attributes like accurate information , user friendly experiences, overall convenience, effective mobile applications , and proactive service measures**.(examining the key dimension of mobile banking service quality)**

Focusing on mobile banking adoption in Oman, this study combines structural equation modelling and neural network analysis. Key findings emphasize the importance of satisfaction, trust and service quality in retaining and attracting customers, offering valuable implications for understanding and enhancing mobile banking adoption**. (examining the role of trust and quality dimensions in the actual usage of mobile banking services)**

This paper investigates the success factors of bank telemarketing prediction using machine learning on a Portuguese banking dataset. Key findings reveal significant variables influencing success, with a focus on decision tree methods. The developed model aims to assist bank/finance companies in managing customer records and making faster decisions for telemarketing success**.(how does machine learning predict the success of bank telemarketing)**

Focusing on the use of Artificial Neural Networks (ANN) in the banking sector, this articles aims to build strong customer relationship through CRM strategies. The ANN model, developed through a six-step procedure, improves customer service efficiency, retention, satisfaction, and overall bank growth, with recommendations for further research on internet banking and ATM transactions. **(Neural Network and classification approach in identifying customer behavior in the banking sector)**

Investigating telemarketing practices for long term bank deposits, this study utilizes machine learning algorithms and identifies logistic regression as the most effective in predicting potential customers. The results offer insights for enhancing telemarketing campaigns and provide guidance for future researchers on feature selection and balancing techniques. **(Predicting the success of bank telemarketing for selling long term deposits)**

Emphasizing the significance of telemarketing for banks, this article proposes a machine learning model to predict the success rate using Portuguese bank dataset. Tree classification algorithms, particularly J48 and random forest, demonstrate better performance, showcasing the potential of the model in enhancing both the financial and functional aspects of bank. **(Prediction of bank customer telemarketing success using data mining models)**

2.4 Table for variables their definitions and resources

Variables	Definition
Age	Age of the customer
Job	Type of the job doing by customers
Marital	Marital status of the customer
Education	Education level of the customer
Default	Does customer has credit in default or not ?
Balance	Average yearly balance in account
Housing	Does customer has housing loan?
Loan	Does customer has personal loan?
Contact	Communication type of the customer
Day	Customer’s last contact day of the month
Month	Customer’s last contact month of the year

Duration	Last contact duration with customer, in seconds
Campaign	Number of contacts performed during this campaign and for this client.
Pdays	Number of days that passed by after the client was last contacted from a previous campaign.
Previous	Number of contacts performed before this campaign and for this client
poutcome	Outcome of the previous marketing campaign
Y	Has the client subscribed for the term deposit.

NOTE: All of the variable sources got from Kaggle

(https://www.kaggle.com/datasets/muhammedsal98/bank-marketing)

2.5 Chapter conclusion

Important terminology like artificial neural network and convolutional neural networks are explained in detail in this chapter . It presents research results on the application of AI and Deep learning to banking, particularly in the area of telemarketing success prediction. Research investigates methods such as decision trees, logistic regression, and deep convolutional neural networks. The impact of mobile banking and borrower preferences in loan repayment arrangements are among the other subjects covered in this chapter. A useful table summarizing the important variables comes at the end. All things considered, the literature study establishes a strong basis by disclosing the various methods that scholars have employed to enhance banking procedures.

Chapter 03: Methodology

3.1 Population, sampling, and sampling techniques:

The population of this research includes individuals who are potential bank customers. This could extent to a broader population if the research aims to have generalizable findings. The sample would like to be a subset of the population, representing individuals who have interacted with banks or have the potential to do so from a specific region.

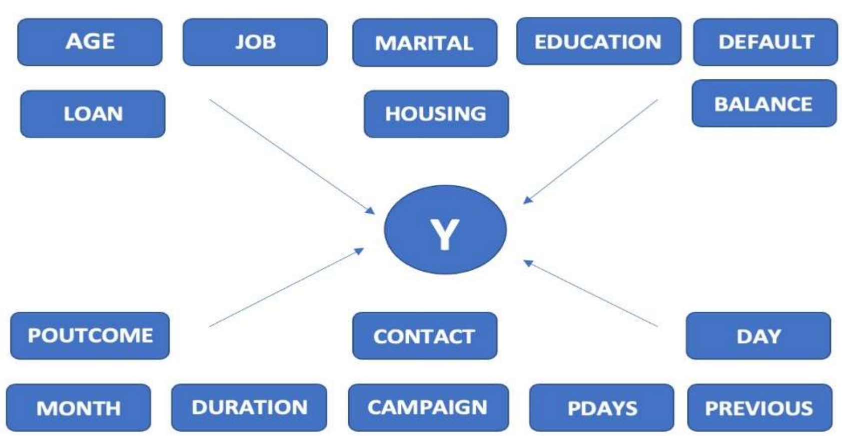
Depending on the research design, random sampling ensures equal chances for individuals in the entire population to be included, while stratified sampling allows for accurate representation pf subgroups based on characteristics like age or income. These sampling techniques collectively contribute to the comprehensive understanding of how customers made decisions regarding bank deposits and loans.

3.2 Type of data to be collected and data sources:

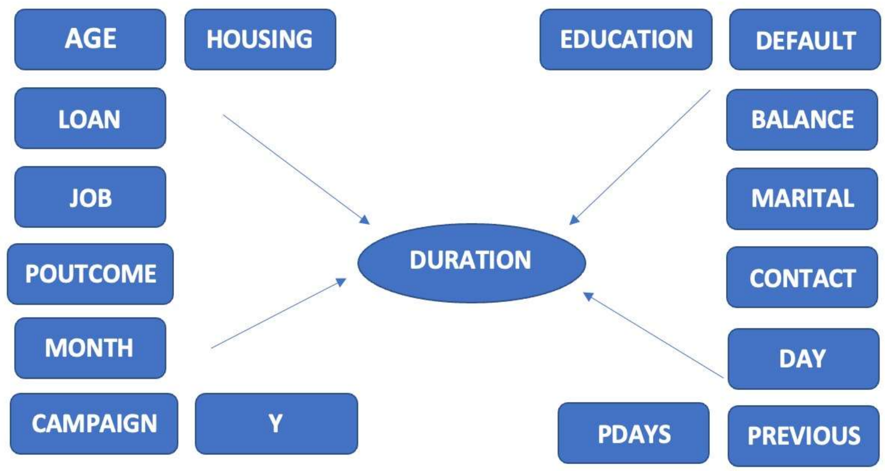
The data to be collected includes both quantitative and qualitative variables. Numeric data includes features like age, balance, day, duration, campaign, pdays and previous. Categorical data includes job, marital status, education, default, housing, loan, contact, month poutcome and y. Data is retrieved from banks’ records, customer databases and possibly though surveys or interviews to gather qualitative insights to make decision making process. Here the secondary dataset was acquired from Kaggle.

3.3 Conceptual framework:

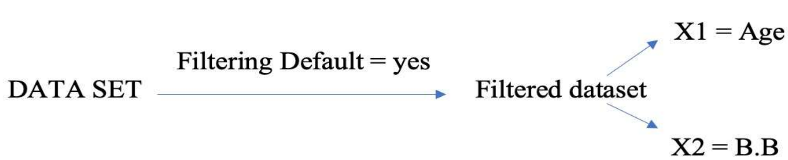
The conceptual framework involves understanding the relationships between the variables. In this case , its about understanding how factors like job, marital status, education etc.. influence deposit decisions , telemarketing call durations and risk of credit defaults.



OBJETIVE 01



OBJECTIVE 02



OBJECTIVE 03

3.4 Operationalization Table:

Variable	Indicators	Measures
AGE	Age of the customer	Numerical
JOB	Type of job	Categorical: - Admin - Unknown - Unemployed - Management - Housemaid - Entrepreneur - Student - Blue – collar - Self-employed - Retired - Technician - Services
MARITAL	Marital status	Categorical: - Married - Divorced - Single
EDUCATION	Education level	Categorical (Binary)
DEFAULT	Has credit in default?	Categorical (Binary)
BALANCE	Average yearly balance in euros	numerical
HOUSING	Housing loan	Categorical (Binary)

LOAN	Personal loan	Categorical (Binary)
CONTACT	Contact communication type	Categorical: - Unknown - Telephone - Cellular
DAY	Last contact day of the month	numerical
MONTH	Last contact month of year	Categorical (Jan – Dec)
DURATION	Last contact duration in seconds	numerical
COMPAIGN	Number of contacts performed during this campaign	numerical
PDAYS	Number of days that passed by after the client was last contacted from a previous campaign	numerical
PREVIOUS	Number of contacts performed before this campaign and for this client	numerical
POUTCOME	Outcome of the previous marketing campaign	Categorical: - Unknown - Other - Failure - Success
Y	Has the client subscribed the term deposit	Categorical (Binary)

3.5 Methods of data Analysis:

A. Classification:

To explore the complexities that impact consumer's choices to deposit money in bank and then build a model that predicts these choices, a classification strategy will be used, which will allow important variables that have a big influence on consumer's deposit decisions to be identified.

B. Regression:

To systematically investigate the variables that affect how long tele marketing calls to consumers last , multiple linear regression will be used, giving a more detailed understanding of the variables influencing how long these interactions last.

C. Clustering:

Identifying the risk profiles of the customer groups with default credit is the third purpose as well, with an emphasis on age and bank balance. Clustering technique will be used to meet this goal, enabling the identification of discrete risk categories based on predetermined financial and demographic factors.

Chapter 04: Data Analysis and Statistics Test.

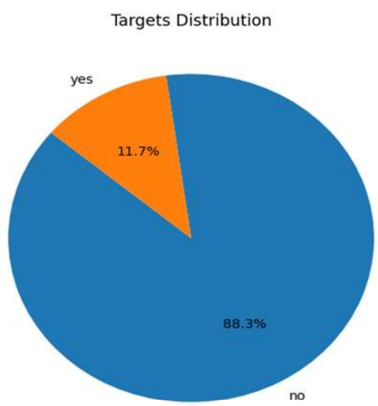
4.1 Data Pre-Processing

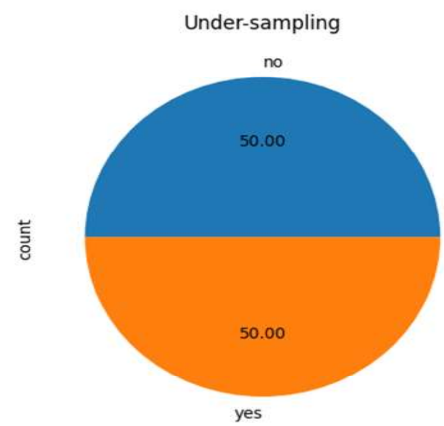
In our dataset there are seventeen variables. In that seven variables are numerical and others are categorical. There is no any null values in our dataset.

Numerical Variables = age, balance, day, duration, campaign, pdays and previous.

Categorical Variables = job, marital, education, default, housing, loan, contact, month, poutcome and y.

First for most in preprocessing We checked the data imbalance. So, our target data is not balanced. We have 39922 entries for no and 5289 entries for yes. So, we did under sampling technique to balance my data.

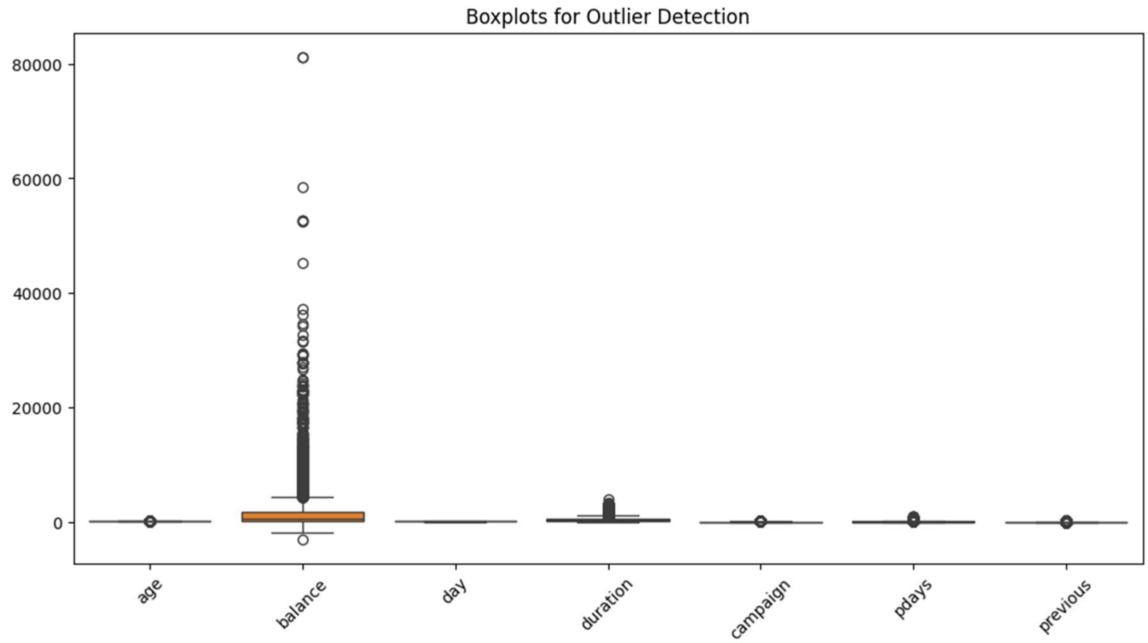




Then we convert object variables to categorical and we get the descriptive statistics of numerical variables.

	age	balance	day	duration	campaign	pdays	previous
count	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000
mean	40.936210	1362.272058	15.806419	258.163080	2.763841	40.197828	0.580323
std	10.618762	3044.765829	8.322476	257.527812	3.098021	100.128746	2.303441
min	18.000000	-8019.000000	1.000000	0.000000	1.000000	-1.000000	0.000000
25%	33.000000	72.000000	8.000000	103.000000	1.000000	-1.000000	0.000000
50%	39.000000	448.000000	16.000000	180.000000	2.000000	-1.000000	0.000000
75%	48.000000	1428.000000	21.000000	319.000000	3.000000	-1.000000	0.000000
max	95.000000	102127.000000	31.000000	4918.000000	63.000000	871.000000	275.000000

Box Plot to numerical variables,



I check for outliers with the help of numerical variables using box and whisker plot visualization. Then we removed the outliers using IQR method. Then I convert my categorical variables to dummy variables. we did one hot encoding because there is no any ordinal categorical variables so I don't need to do ordinal encoding. Then to achieve my second objective I filter and get only the customers who did not deposit and to third objective I choose only customers who have default credit. Then we did feature selection to implement our classification models and regression model. We choose decision tree classifier method to select features for classification models and for the second objective we check the P value of the features to check the significant relationship.

4.2 Statistics Test

We are going to do five machine learning algorithms to solve our our problem and achieve our objectives. In those five algorithms three are classification models, one is regression model and one is clustering model. To do this classification tests I split the dataset 80% to training and 20% to testing.

Classification Models = Logistic Regression, SVM classifier, Random Forest Classifier

Regression Model = Multiple Linear regression, XGBoost Regressor, Random Forest Regressor.

Clustering Model = K- Means Clustering

4.2.1 Explore and analyze the factors that affect decisions of customers to make deposits in bank and build a model to predict customer decision regarding deposit.

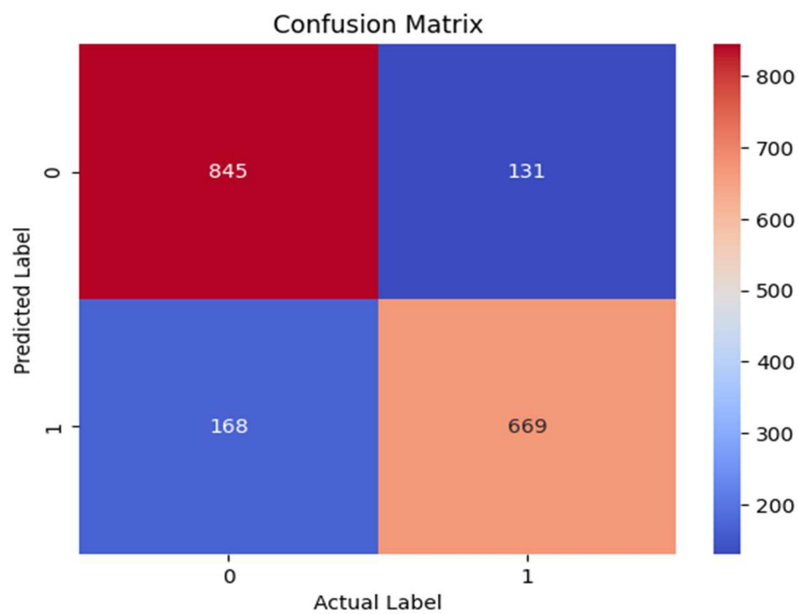
Logistic Regression

We did hyper parameter tuning to get suitable regularization parameter (C) and penalty..

```
Best Hyperparameters: {'C': 100, 'penalty': 'l2'}
```

We evaluate the model using classification report and confusion matrix.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.87      0.85       976
           1       0.84      0.80      0.82       837

    accuracy                           0.84      1813
   macro avg       0.84      0.83      0.83      1813
weighted avg       0.84      0.84      0.83      1813
```




```
In [62]: logistic_model.score(X_train_scaled,y_train)
Out[62]: 0.8300455235204856

In [63]: logistic_model.score(X_test_scaled,y_test)
Out[63]: 0.8350799779371207
```

Approximately test and train accuracy is equal so model did not overfit and accuracy is 83% so this model is good.

Random Forest Classifier

We did hyper parameter tuning to parameters n_estimators, max_depth, min_samples_split, min_samples_leaf.

```
Best Hyperparameters: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_s
plit': 10, 'n_estimators': 15}
Best Accuracy: 0.8507391066371577
Best Model Accuracy on Test Data: 0.8477661334804192
```

Model Evaluation,

```
In [68]: random_forest_classifier.score(X_train_scaled,y_train)
Out[68]: 0.9733756380190371

In [69]: random_forest_classifier.score(X_test_scaled,y_test)
Out[69]: 0.8450082735797021
```

The test accuracy and train accuracy have big different. Train set have high accuracy si model have overfit, We try to prevent overfit using hyperparameter tuning but fail to prevent it.

Support Vector Machine Classifier

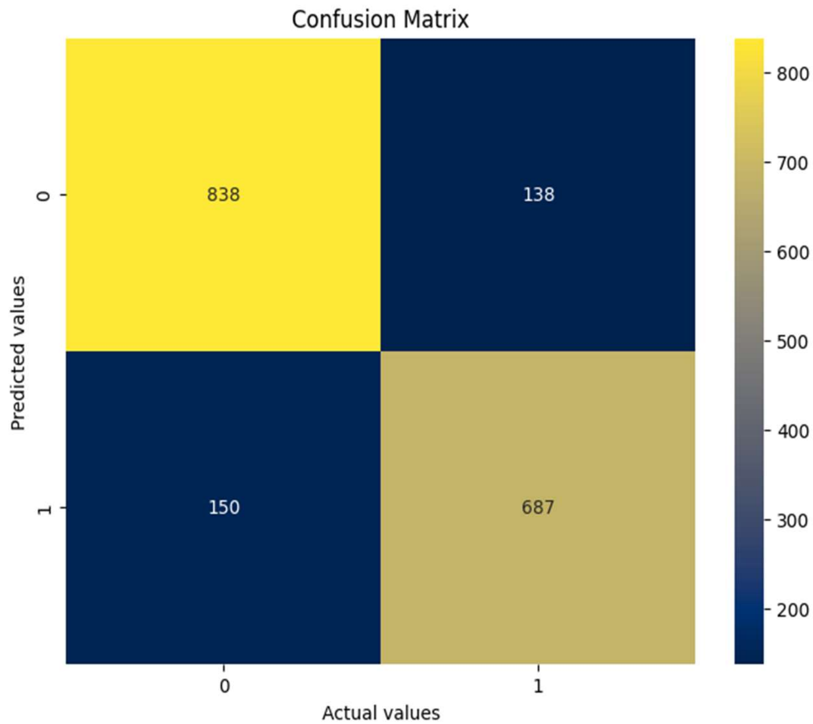
We skipped hyperparameter tuning to this model because of computational cost.

Evaluation,

```
In [75]: classifier.score(X_test_scaled, y_test)
Out[75]: 0.8411472697186982

In [76]: classifier.score(X_train_scaled, y_train)
Out[76]: 0.831976824389571
```

Train set and test set accuracy is approximately equal so this model prevent from overfit.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.86      0.85       976
           1       0.83      0.82      0.83       837

    accuracy                           0.84      1813
   macro avg       0.84      0.84      0.84      1813
weighted avg       0.84      0.84      0.84      1813
```



So, this models accuracy is 84% and this is equal to logistic regression model but we going to decide the model using ROC curves.

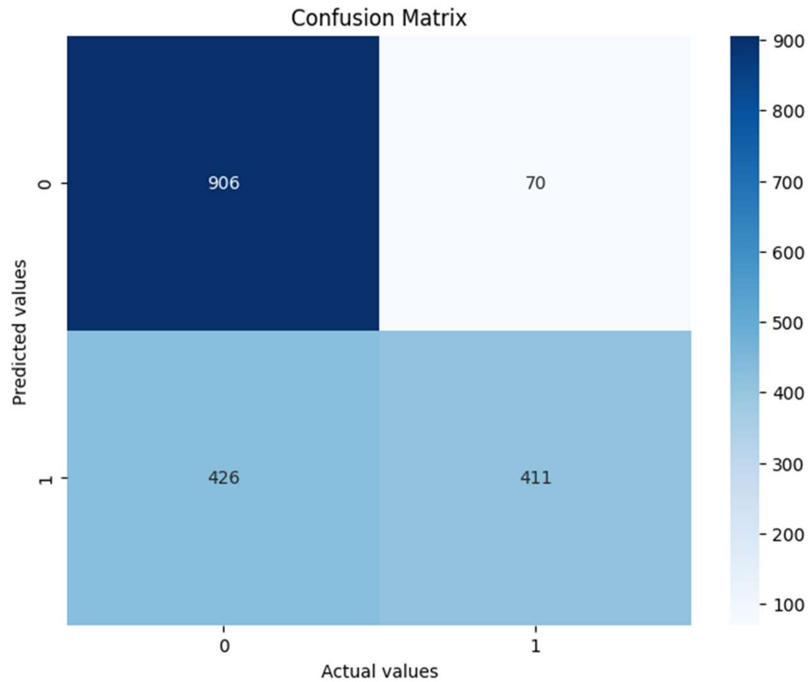
Naïve Bayes

Accuracy of training and testing set,

```
In [84]: testscore=nb_classifier.score(X_train, y_train)
         print("Test Set Accurac:", testscore)

         Test Set Accurac: 0.7194095737343081

In [85]: testscore=nb_classifier.score(X_test, y_test)
         print("Test Set Accurac:", testscore)

         Test Set Accurac: 0.7264202978488693
```

Classification Report and Confusion Matrix,

	precision	recall	f1-score	support
0	0.68	0.93	0.79	976
1	0.85	0.49	0.62	837
accuracy			0.73	1813
macro avg	0.77	0.71	0.70	1813
weighted avg	0.76	0.73	0.71	1813



This models accuracy is 73% and this is lower than logistic regression and SVM.

XGBoost Classifier

Hyper Parameter Tuning,

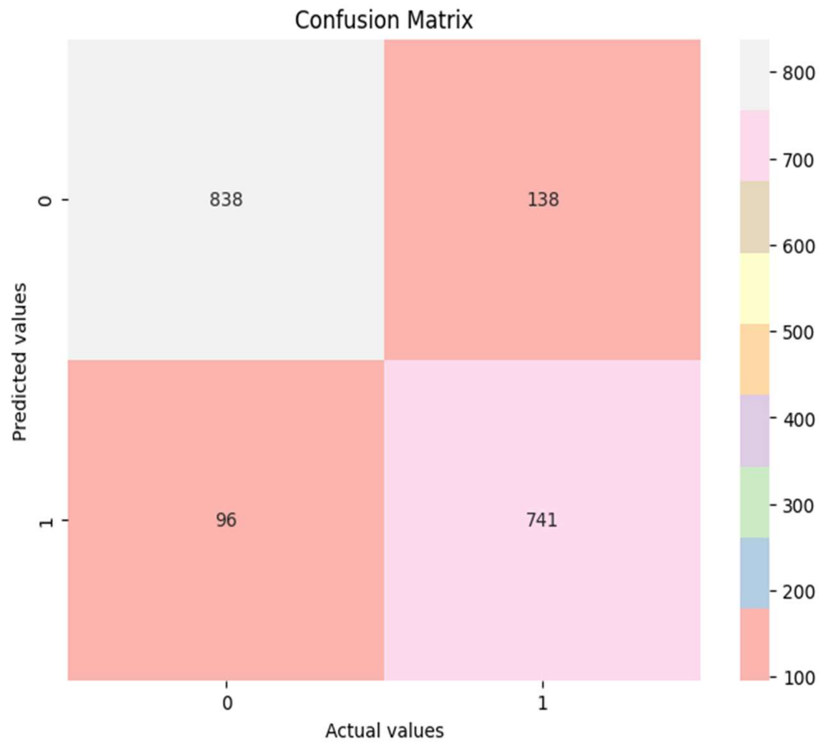
```
Best Hyperparameters: {'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 20
0, 'subsample': 0.9}
```

Model Evaluation,

```
In [92]: accuracy = xg_classifier.score(X_train_scaled,y_train)
         print(accuracy)

         0.9001241550558697

In [93]: accuracy = xg_classifier.score(X_test_scaled,y_test)
         print(accuracy)

         0.8709321566464424
```

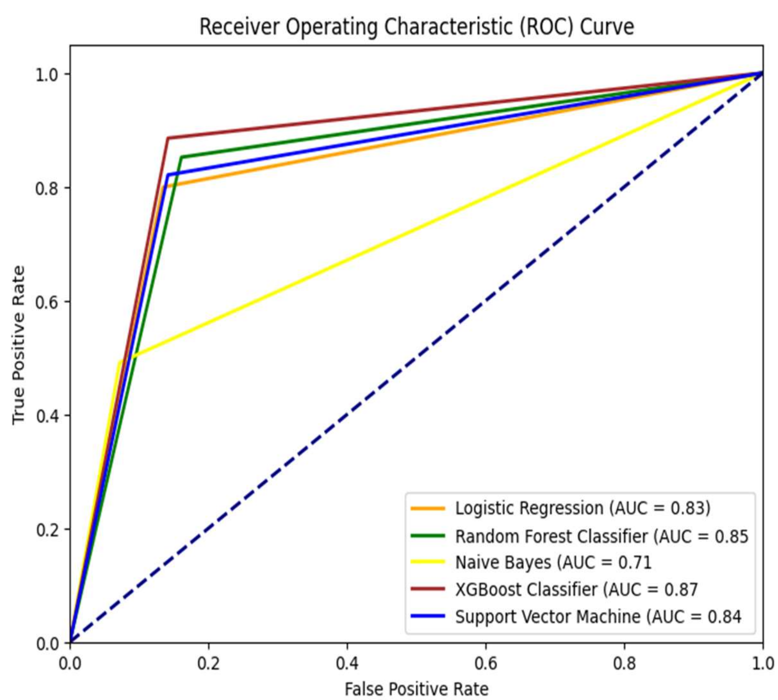
Classification Report & Confusion Matrix

	precision	recall	f1-score	support
0	0.90	0.86	0.88	976
1	0.84	0.89	0.86	837
accuracy			0.87	1813
macro avg	0.87	0.87	0.87	1813
weighted avg	0.87	0.87	0.87	1813



In this case training set accuracy is greater than testing set accuracy so this model also overfitted like random forest classifier.

ROC Curve



According to ROC curve accuracy is high for XGBoost classifier and Random Forest classifier but we wont choose that models because we already discuss above that models are overfitting so we avoid this model. So we choose SVM as our best classification model.

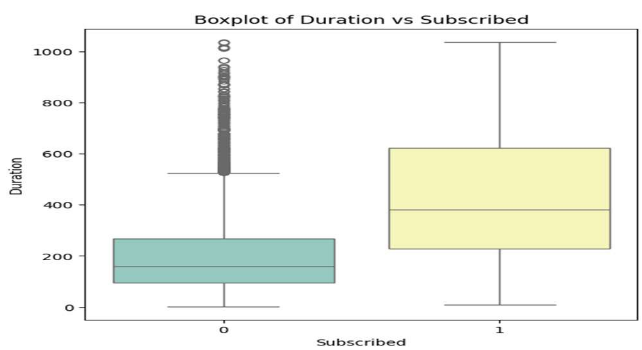
4.2.2 Explore the factors that affect telemarketing phone call duration that makes to customer.

In our dataset we filter and get only the customers who did not subscribed the term deposit and then we start to do data analysis in the telemarketing call duration of customers that who did not subscribe the term deposit. So, we can submit and present this analysis to marketing department and other departments to focus on that group of customers and marketing them to subscribe the term deposit.

Data Visualization

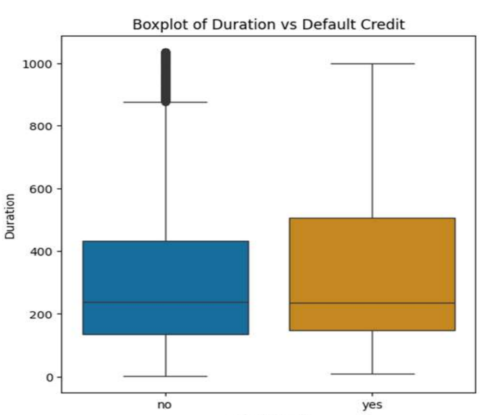
In this section we will get the insights from box plot that which and how categorical factors affect the tele marketing call duration.

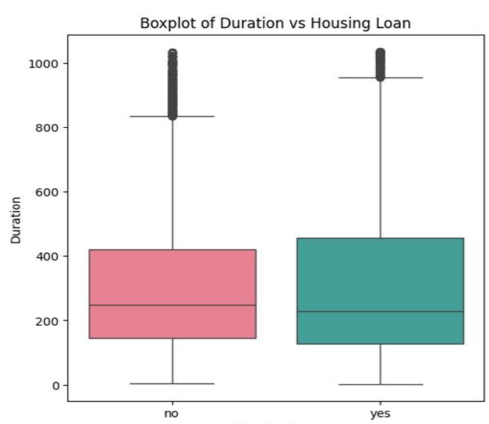
Subscribed vs Duration

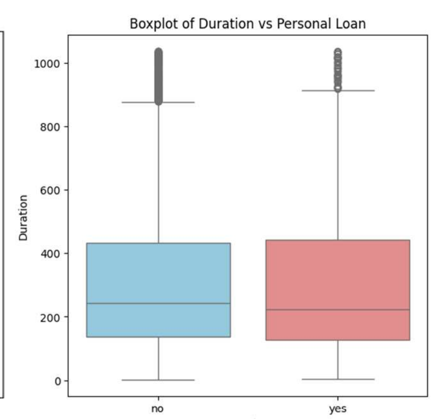


So, according to this plot we can assume that customers who have subscribed to the term deposit have large call duration than customers who have not subscribed.

Default Credit vs Duration

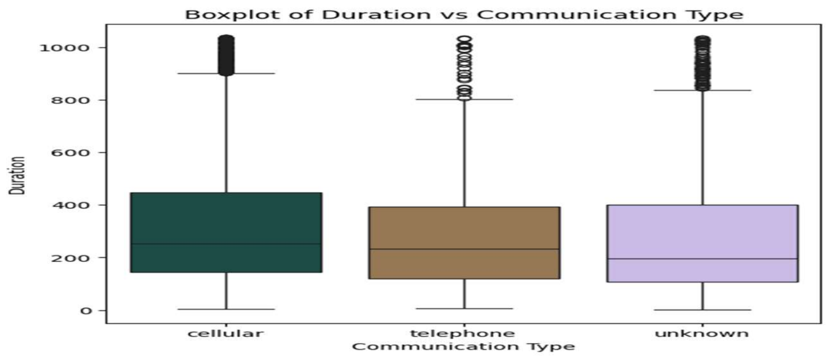






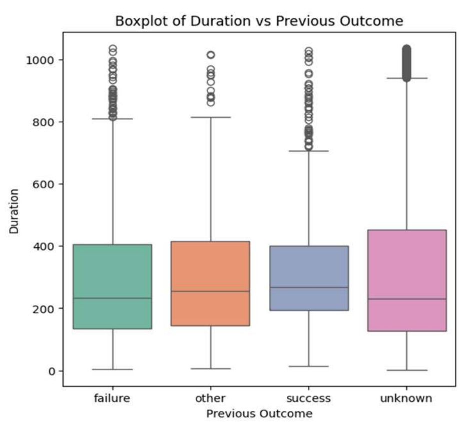
When we look into these three box plots we can decide that every median duration are equally affect the each classes but there spreads are different. So, we cannot get proper insights from these box plots.

Communication Type vs Duration



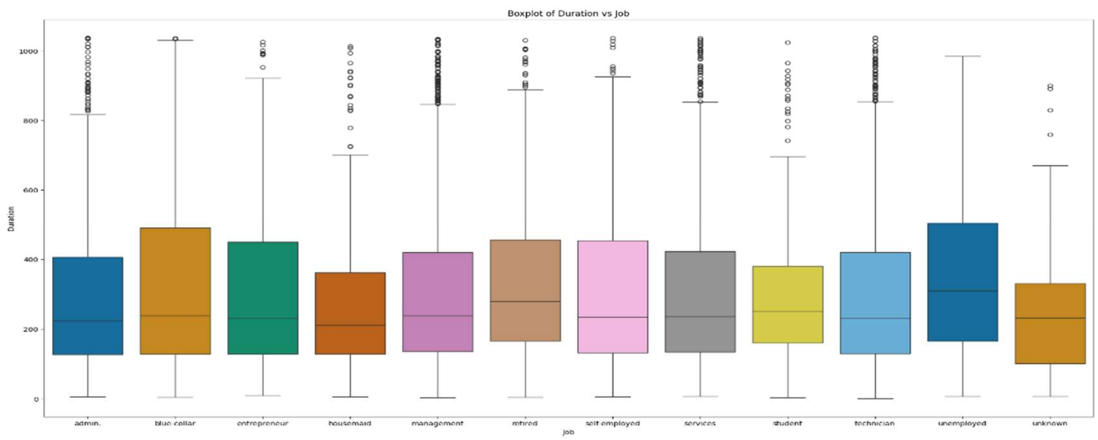
In this case cellular communication type customers have high median duration in and lowest median duration is unknown type. But we can focus on cellular but we cannot focus on unknown type because we don’t have proper description about cellular.

Previous Outcome vs Duration



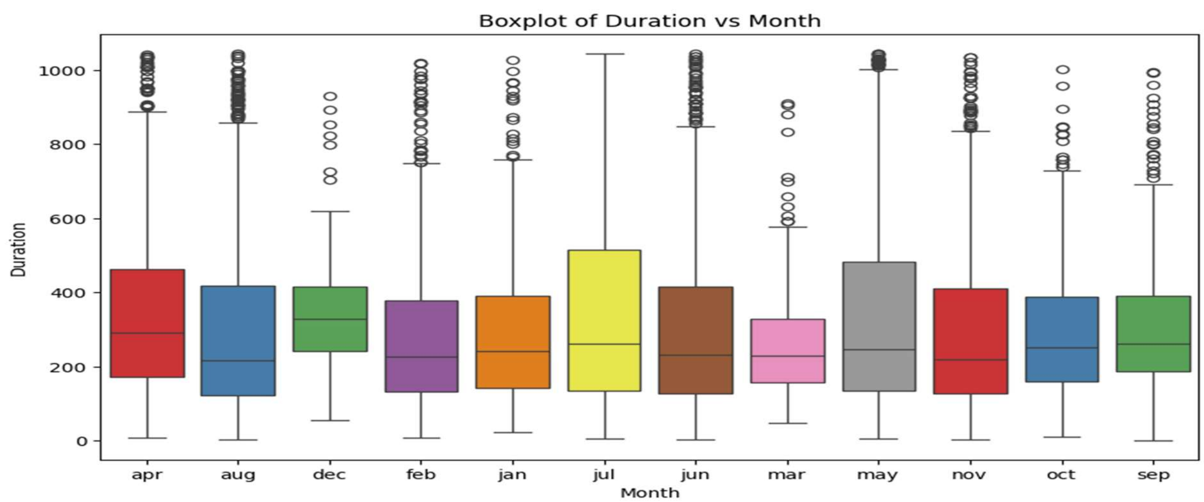
According to this box plot we have highest median duration for success group and lowest to unknown group.

Job vs Duration



According to this box plot median call duration is high for unemployed customers and lowest for housemaids. Other job types median duration is between 200 and 290.

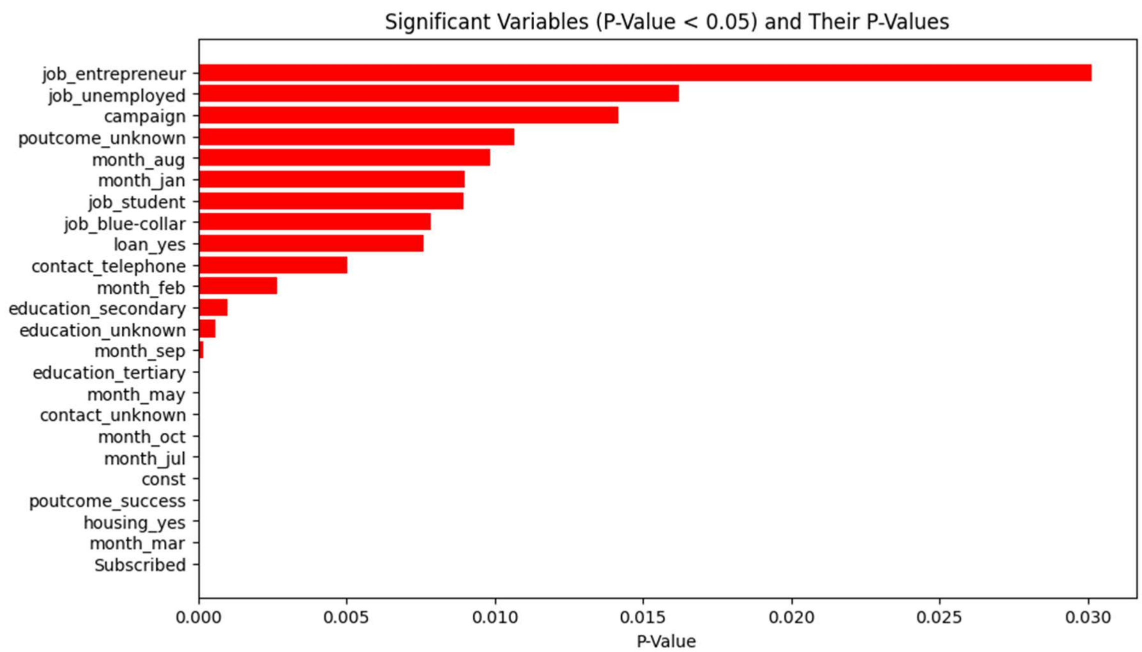
Month vs Duration



According to these box plots we have the highest median duration on December and lowest on August.

Regression Model

We did multiple linear regression model to choose the proper independent variables for our dependent variable (Subscribed) by looking into P values. We choose significant level as a 0.05. So, features that have relationship with dependent variable are below,



Selected Features,

Subscribed,month_mar,housing_yes,poutcome_success,month_jul,month_oct,month_may,job_blue-collar,job_unemployed, job_services,month_sep,contact_unknown,campaign,month_jan,month_aug,loan_yes, job_self-employed,job_student,job_retired, job_technician,poutcome_unknown,job_management,job_housemaid,contact_telephone, job_entrepreneur

Predictive Model,

We build a predictive model to predict tele marketing call duration of customers. So, we took three regression model to build the predictive model. Those are linear regression, random forest classifier and XGBoost regressor.

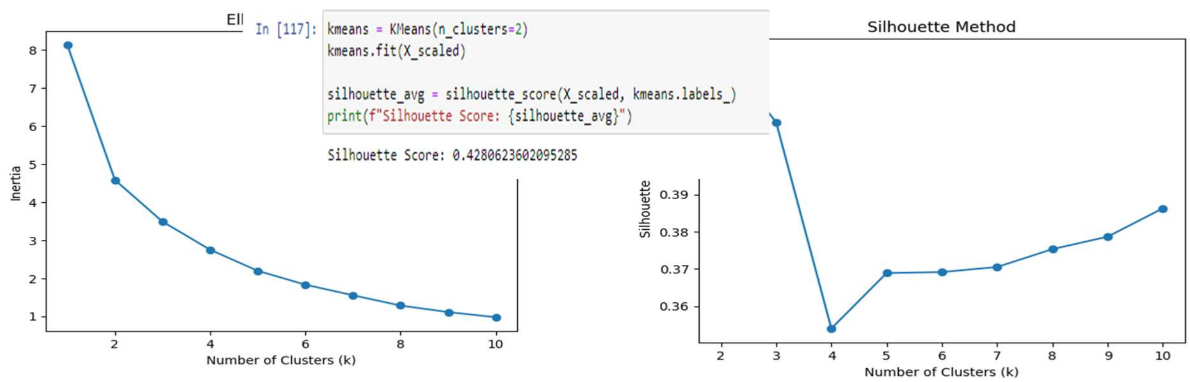
Model	MSE to Test and Predict (y)	neg_MSE to Cross Validation
Linear Regression	41092.53	39698.02
Random Forest Regressor	42259.58	41962.78
XGBoost Regressor	39652.94	39570.91

So, according to this table best model to predict the tele marketing call duration of the customer is XGBoost regressor. By analyzing this scores that we have choosed XGBoost regressor as our regression model.

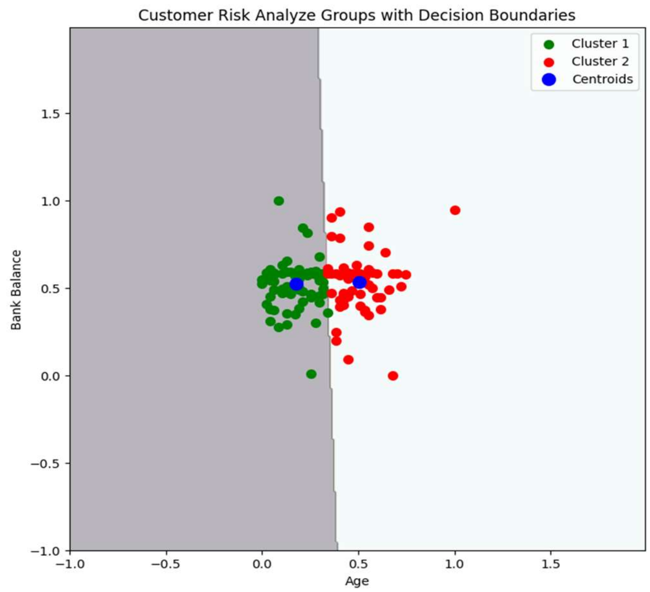
4.2.3 Identifying the risk profile of groups of customers who have default credit based on age and bank balance of them.

We going to group the customers and we going to identify the risk profile customers through our dataset using clustering algorithm. For this research first we filter and get the customers who have default credit because if he / she had a default credit they need to pay money back to bank so, when we look into our dataset the customers who have default credit have housing loan or personal loan or neither these types of loan. So however if customer have a default loan we can consider them as a loan borrower. So we going to analyze the risk by grouping these customers. For this we have used K means clustering algorithm.

To choose K (Clusters) we have used elbow method and Silhouette Method.



We get silhouette score as 0.43 so this prove us that our object is good matched for its own cluster and poorly matched for its neighbor cluster.



According to this plot we can decide that the customers who have age more than 42 (42 = 0.42) we need to analyze those group of customers (cluster 2) and bank need to get the money back from them as soon as possible. More than that bank should focus on group 1 customers, and they need to monitor them.

Chapter 05: Discussion, Recommendations and Conclusion

5.1 Discussion

To achieve our objectives, we need brief fundamental knowledge of machine learning but not only machine learning, we need descriptive statistics, data preprocessing and data visualization. This research study will be very useful to bank organizations to customer retention and to achieve their goals. We built a predictive model using logistic regression to predict whether the customer will subscribe to the term deposit or no. and we identify the features that influence customers subscription. Next, we identify the factors that affect tele marketing call duration of the customers, and we build a predictive model using XGBoost regressor to predict that. So, this will be useful to manage the resources in a bank. Finally, we segment the customers to identify the risk profile. So these researches we use secondary dataset where we get it from Kaggle so we tune the models and did our best to achieve the objectives.

5.2 Recommendation

According to our research we will recommend some ideas to bank organization by the help of our research results.

- According to data visualization we can assume that some categories don't influence the term subscription of the customer directly. Those categories are default loan, personal loan, marital status and education level.
- Customers who have housing loan did not subscribed so we can make and adjust some terms and conditions of the subscription to customers who have get housing loan.
- We need to adjust the terms and conditions of the subscription plan to customers whose jobs are blue collar, technician and services.
- To build a predictive model to predict whether customers accept the subscription plan or not we need to build a model using SVM.
- To build a predictive model to predict telemarketing call duration to customers we need to build a model using XGBoost Regressor.
- While doing tele marketing we need to plan and focus the call duration of customers by analyzing other features of that customer. Such as job and etc.…
- We need to plan other way to contact and market the term deposit plan to customers whose occupation is housemaid and admin
- In some months banks tele marketing call duration is low we need to analyze why is that and how to schedule marketing plan on those months. (August, February, January, and March)
- Banks need to focus to group of customers who are under cluster 2 because they have get loan in bank, their age is also high and they have good bank balance also. So, we need to focus on those customers to get back the loan or settle down it.

5.3 Conclusion

This research goals are to analyze the factors that contribute to term deposit subscription and the telemarketing call duration, build a predictive model to predict telemarketing call duration and whether customer accept or reject the term subscription and to analyze the risk profile customers of the bank by analyzing their bank balance and age.

This dataset is secondary data set so, according to that there are some areas should to investigate more,

- There is a category call 'unknown' so it's a big mistake we can't get proper insights on categorical data which include class unknown.
- This dataset did not mention the way data was collected and which specific area and bank is this.

References

Examining the key dimensions of mobile banking service quality: an exploratory study, jun,M. and palacios,S. (2016),

https://www.emerald.com/insight/content/doi/10.1108/IJBM-01-2015-0015/full/html

Examining the role of trust and quality dimensions in the actual usage of mobile banking services: an empirical investigation, Sujeet kumar sharma,Manisha Sharma, 2019 feb , https://www.sciencedirect.com/science/article/abs/pii/S0268401218307321

Predicting the success of bank telemarketing using deep CNN, kee hoon kim ,chaan,seok lee, ang muk jo, sung bae cho, 2015, https://ieeexplore.ieee.org/abstract/document/7492828

How does machine learning predict the success of bank telemarketing , younkeun choi, jae choi,2022, https://www.researchsquare.com/article/rs-1695659/v1

Prediction of bank customer telemarketing success using data mining models, yujen sun, xinru cheng, haiming li, yuwen lin,joowan sung, anlan wang, 2019, https://journals.adrri.org/index.php/adrrijet/article/view/461

Predicting the success of bank telemarketing for selling long-term deposits an application of machine learning algorithms, premkumar borugadda,Dr. Prabakar nandru,Dr, chendragiri madavaiah,2021, https://journal.stic.ac.th/index.php/sjhs/article/view/296

Applying AI techniques to predict successs of bank telemarketing, kun huang chen, hsuan wen chui, 2020, https://dl.acm.org/doi/abs/10.1145/3417188.3417198#:~:text=We%20use%20python%20to%20build,regression%20analysis%20performed%20the%20best.

Determination of bank customer risk profile: data mining applications ,tanerersoz,filiz ersoz, seyma ozbilge,2016, https://www.researchgate.net/publication/315831291_Determination_of_the_Bank's_Customer_Risk_Profile_Data_Mining_Applications

An analysis of loan repayment plans according to the bank customer profile, SC. Rambaud,MD.Alveriz,2018, https://www.semanticscholar.org/paper/An-Analysis-of-Loan-Repayment-Plans-According-to-Rambaud-%C3%81lvarez/0c3f78d59a4facdf93c3895ed684bcbf6c9731c6

Neural netowkr and classification approach in identifying customer behaviour in the banking sector: a case study of an internat, francisca,Sanjay,ricardo,Luis, 2012, https://onlinelibrary.wiley.com/doi/abs/10.1002/hfm.20398