





# Problem



*The goal is to predict whether a customer or a client of the Bank would subscribe to a product known as Bank Term Deposit. To tell whether it would be a YES or a NO*

---

# Stakeholders



*The Management Team*



*The Marketing Team*

# Data Facts

- *The Data we used is related to direct marketing campaigns based on phone calls.*
- *The Data had 4521 entries and included 17 features with client information such as; Age, Type of Job, Education, Marital Status, and Housing Loan/ Personal Loan.*
- *The Data also had information relating to the last contact such as*
  - *How the contact was made (Cellular or Telephone)*
  - *When the contact was mad ( Day, Month, and Duration of the call)*
- *The data also informs of the number of contacts made and, how long it took before the previous contact, and the outcome of the previous campaign.*

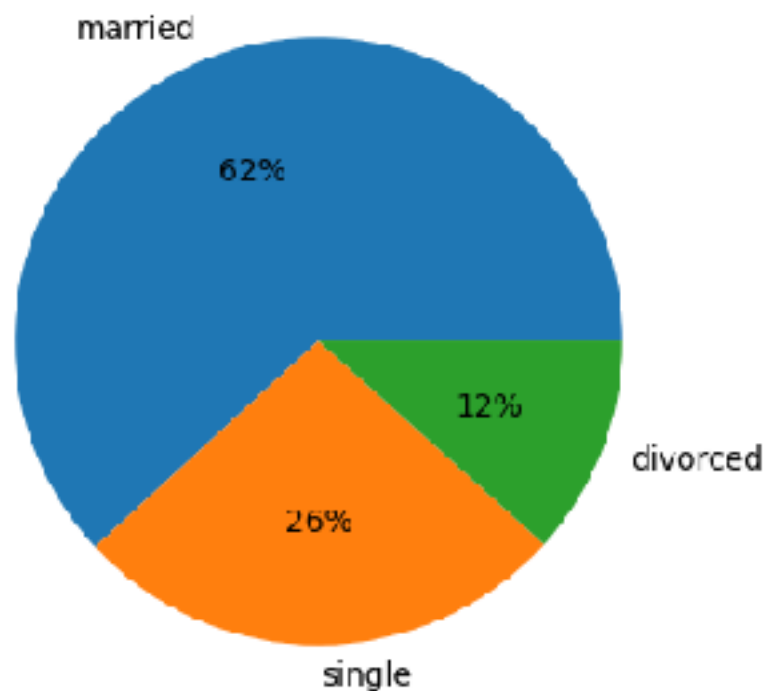




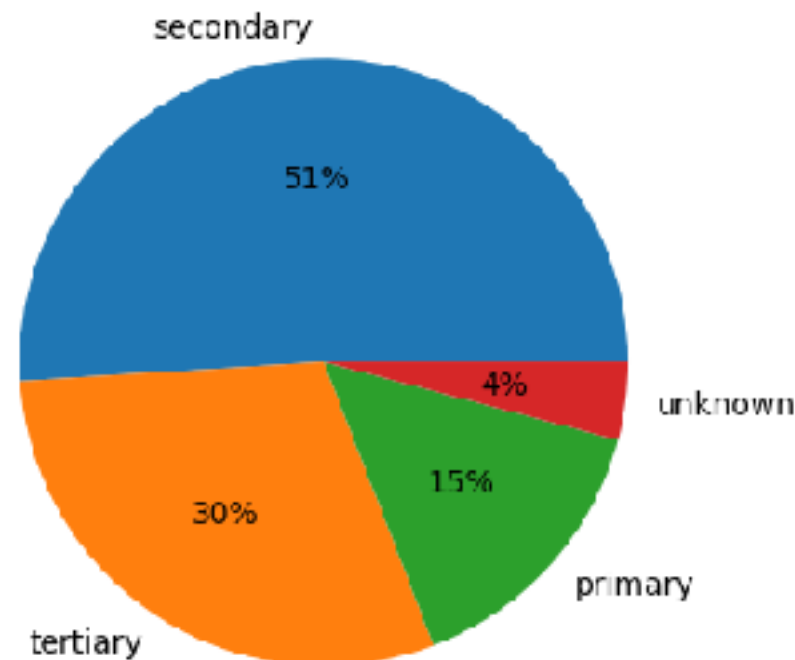
# Data Insights



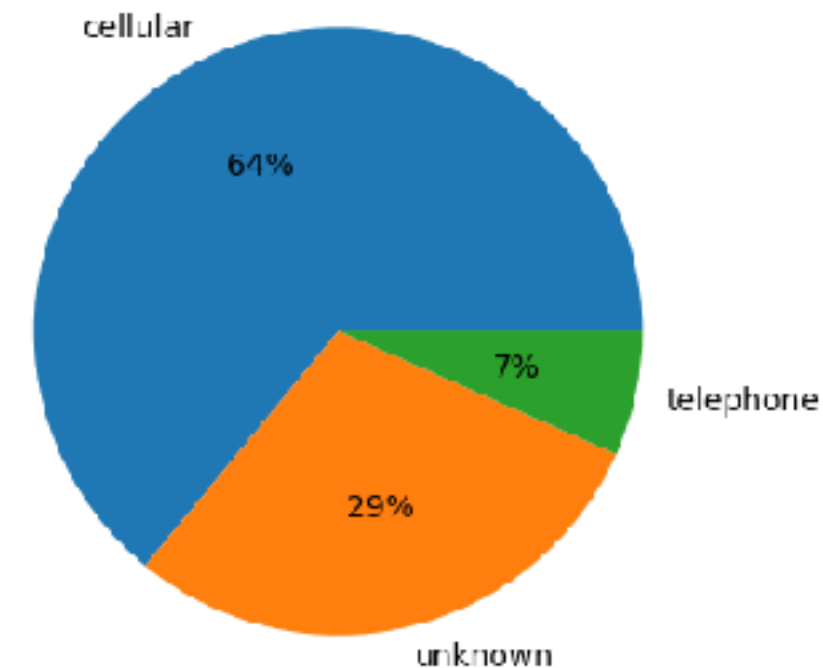
Marital



Education

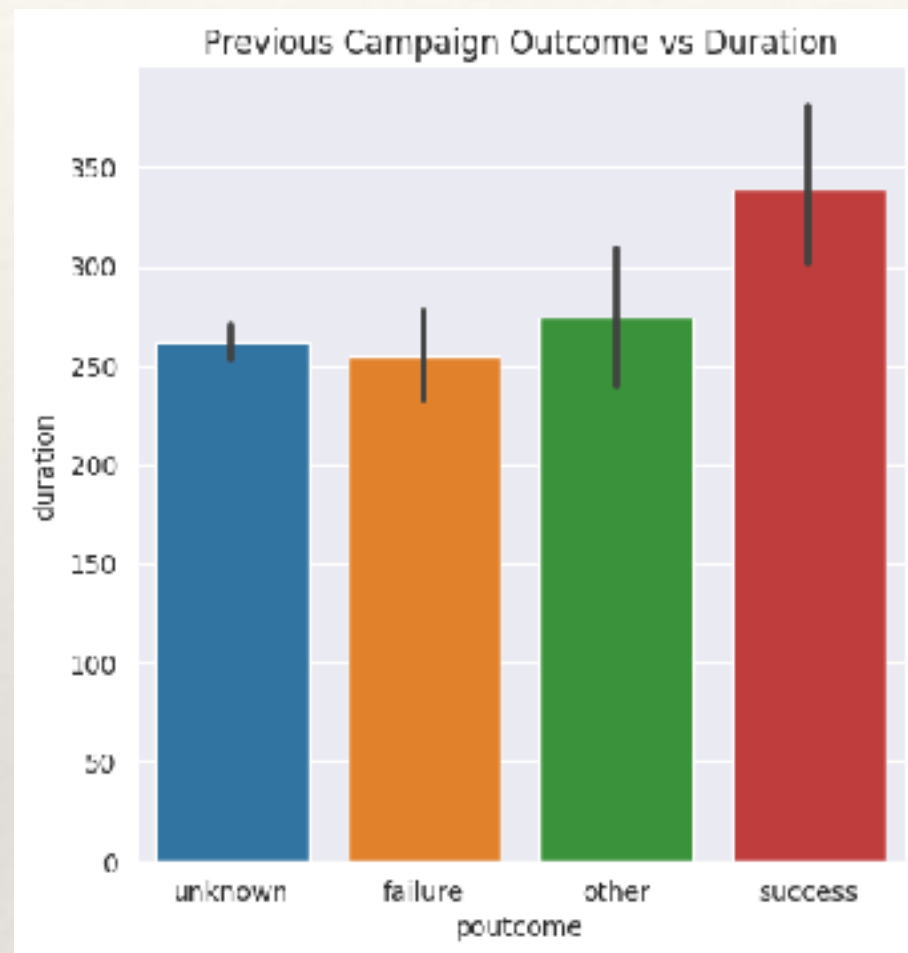


Contact

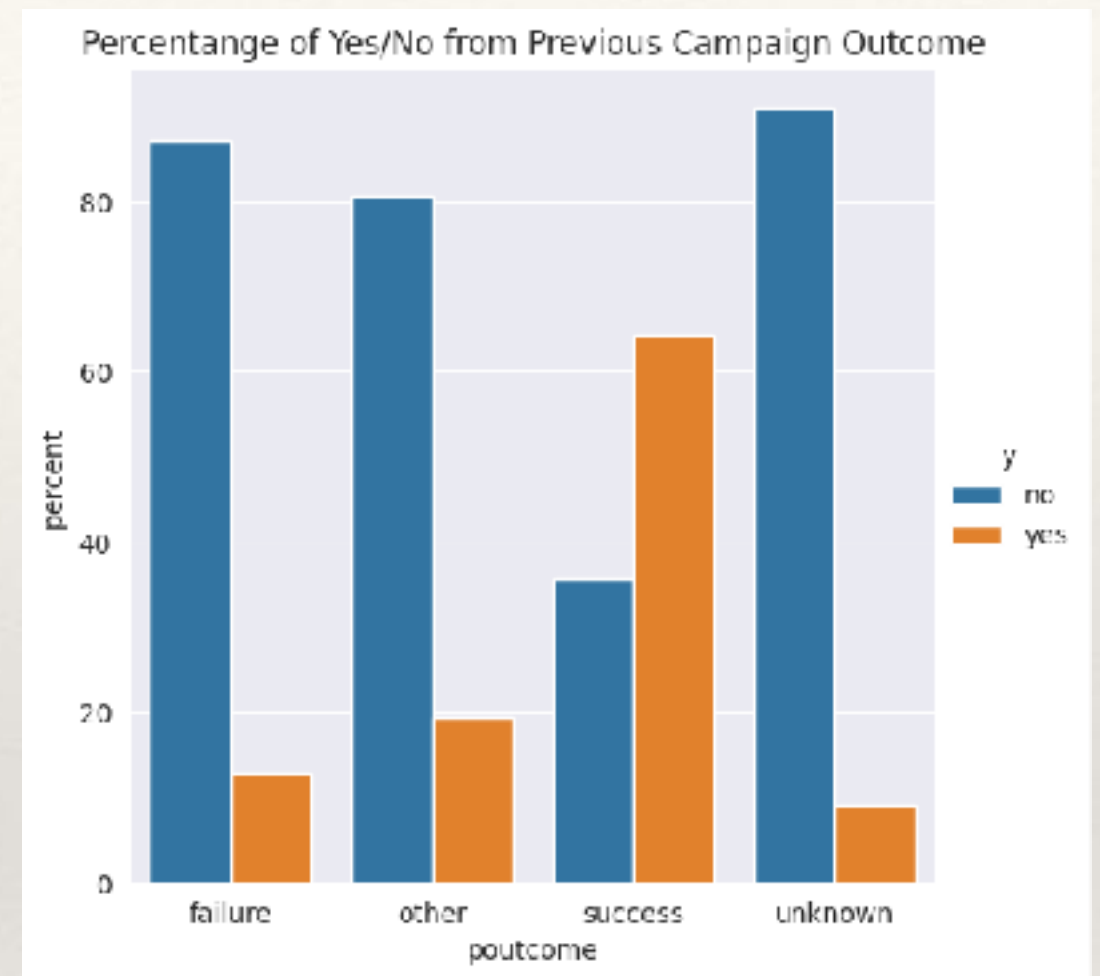


- 62% of the customers are married, compared to 12% Divorced and 26% Single
- 51% of the customers had a Secondary Education Level, compared to 30% Tertiary Level and 15% Primary Level
- 64% of the Customers used Cellular Contact compared to 7% Telephone and 29% Unknown.

# Data Insights



- We can tell that from the previous campaign, the longer the duration of the phone call, the more likely it was for the client to subscribe to the bank products.*



- And those who said yes to the previous outcome were 60% more likely to say yes to the bank term deposit campaign.*

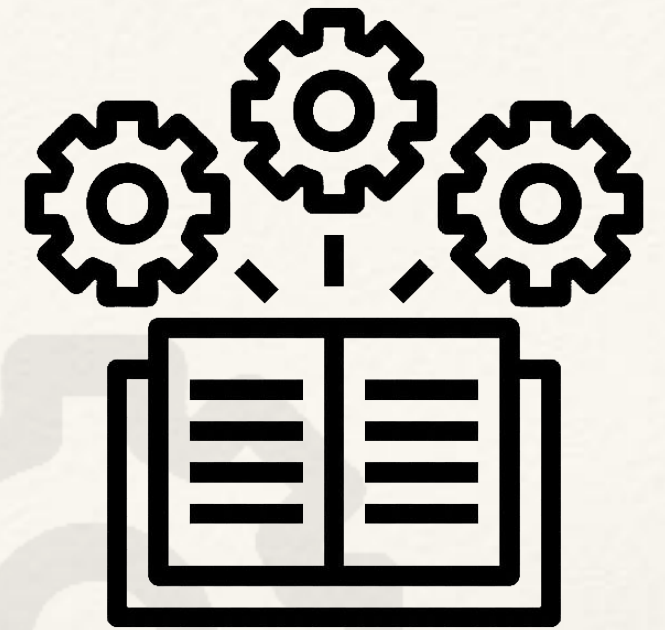
# Machine Learning Models

- *Logistic Regression Model*
- *Random Forrest Classifier*
- *XGB Classifier*
- *KNN Classifier*

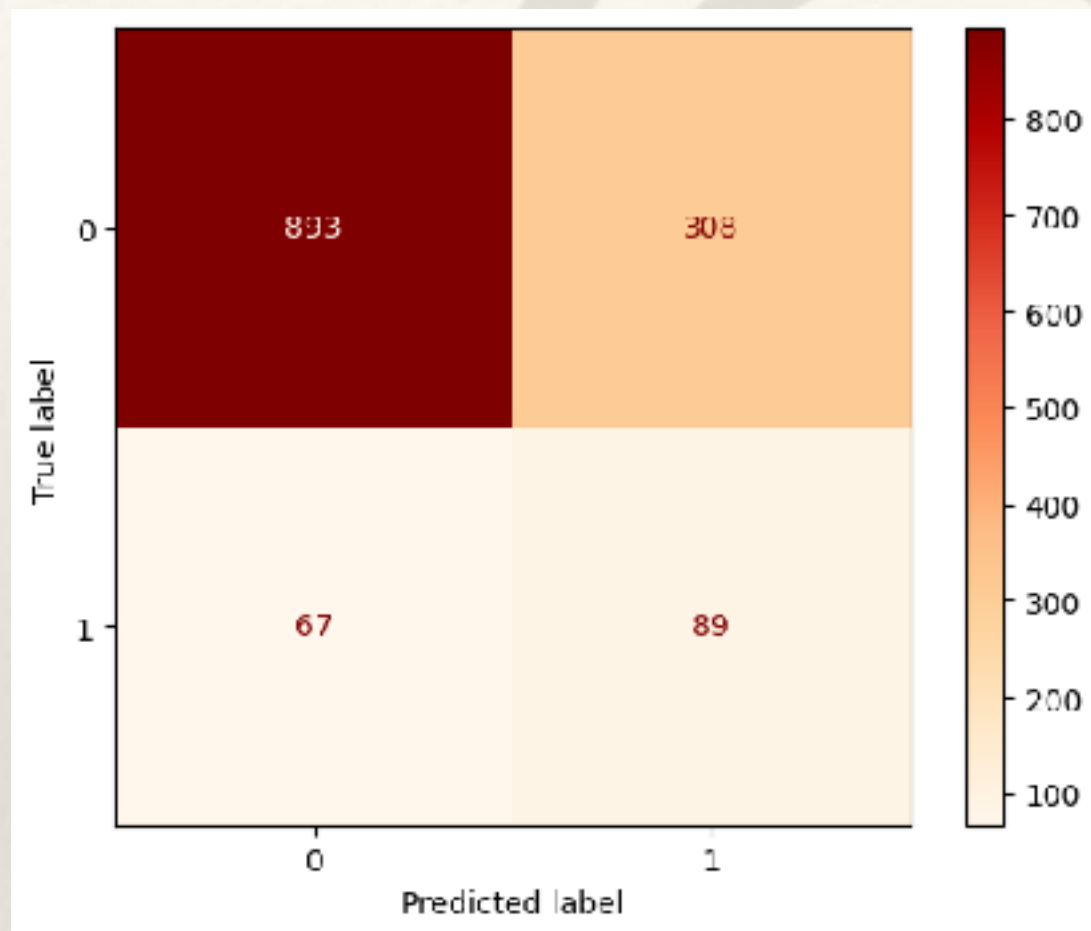
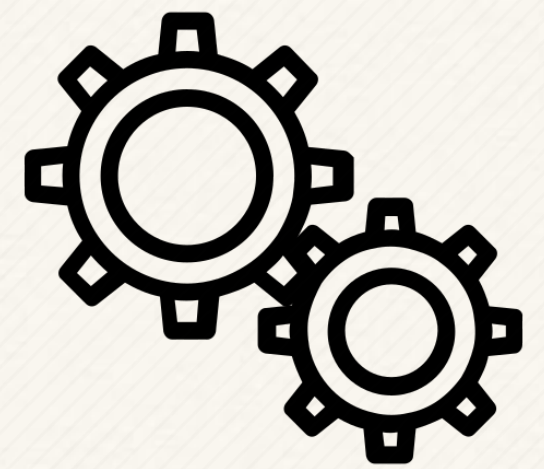
*The models seemed to perform poorly on the data, probably because the target data was quite imbalanced.*

*The **Logistic regression Model** was chosen as the final regression model because of its speed and with further Hyper Parameter tuning, the f1-score was improved to 0.32.*

*The f1 score is an average of the mistakes it can make in predicting those who will say yes or no. Therefore the higher f1 score makes it the best model to solve the problem because it will make slightly fewer mistakes than the others.*



# Machine Learning Model - Performance



	precision	recall	f1-score	support
0	0.93	0.74	0.83	1201
1	0.22	0.57	0.32	156
accuracy			0.72	1357
macro avg	0.58	0.66	0.57	1357
weighted avg	0.85	0.72	0.77	1357

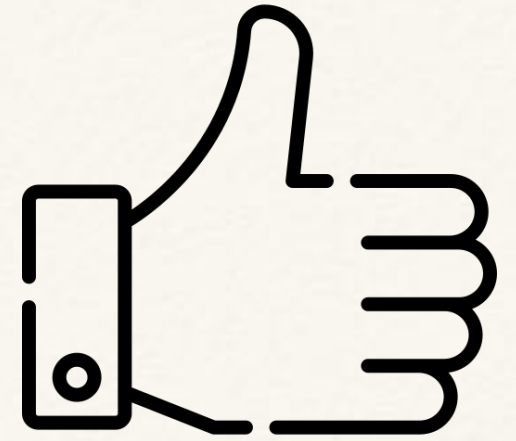
*We can make the following observations from our Performance metrics above*

- For 156 customers of who said yes to the Term Deposit, the model incorrectly predicts 67 of them saying No. This is quite a large number.*
- For 1201 customers who said No to Term Deposit, the model incorrectly predicts 308 of them saying Yes to Term Deposit.*

*The data, sampling and modeling will be tuned to get better performance*



# Recommendations



- *The marketing team needs to target those aged between 40 and 50 as they seem to be employed in blue-collar or management jobs.*
- *Those who have shown positive feedback to previous campaigns need to be targeted more in campaigns as they have a higher chance of saying yes to new bank products.*
- *Longer call durations, normally correlate with a yes. The assumption is the client is at ease and more trusting of the person on the marketing of the phone.*
- *More training needs to be done to engage clients in a personal and friendly way to increase the potential of saying yes to bank product marketing campaigns*





*- Timothy Gambu*