



Bank Term Deposit Marketing Strategy

Tannaz Mostafid

Outline



Business Understanding

Data Understanding & Exploration

Data Preparation

Modeling

Evaluation

Business Understanding



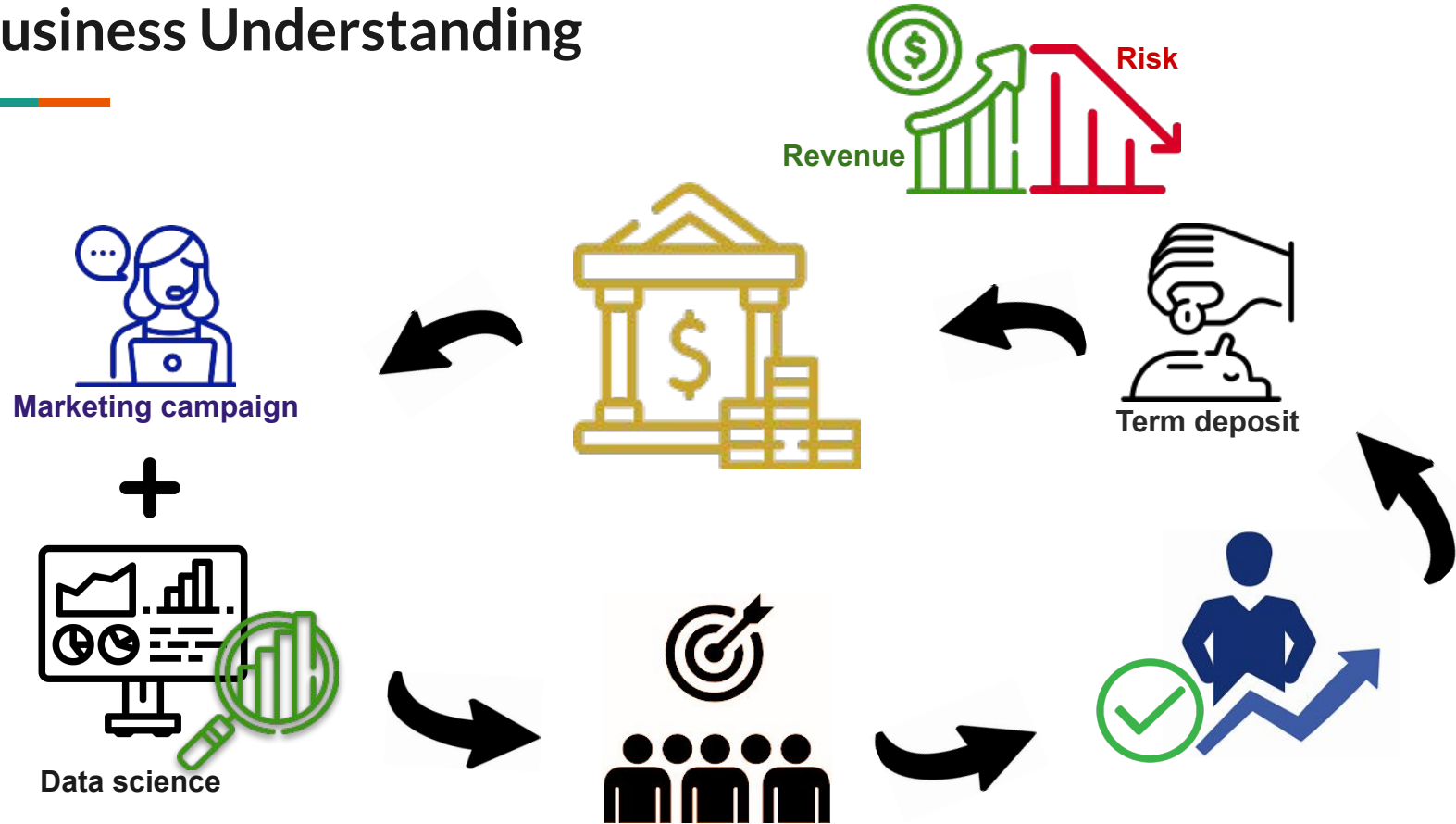
- **What is the Term Deposit?**

- A term deposit is a savings option where you lock away your money for a fixed term and earn a guaranteed interest rate, but can't access the money until the term is up.

- **What are the benefits of term deposits for both banks and customers?**

- Bank: Stable predictable funding source, reduce their liquidity risk.
- Customer : Secure and low-risk investment option, Guaranteed rate of return
- Overall: provide a win-win situation

Business Understanding



Business Understanding



- **Goal**

- Optimizing factors for successful bank deposit subscriptions to efficiently target high-value customers.

- **Objective**

- The objective is to develop a ML model that can effectively classify whether a client subscribes to longer-term deposits by analyzing input features and identifying unknown patterns.

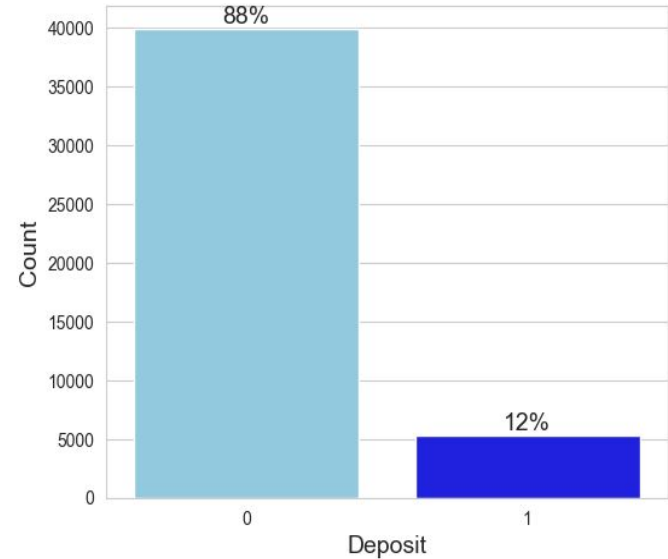
Data Understanding

- Which customer characteristic is important for effective marketing strategies?
- What leads customers to subscribe to term deposit?



Deposit Ratio

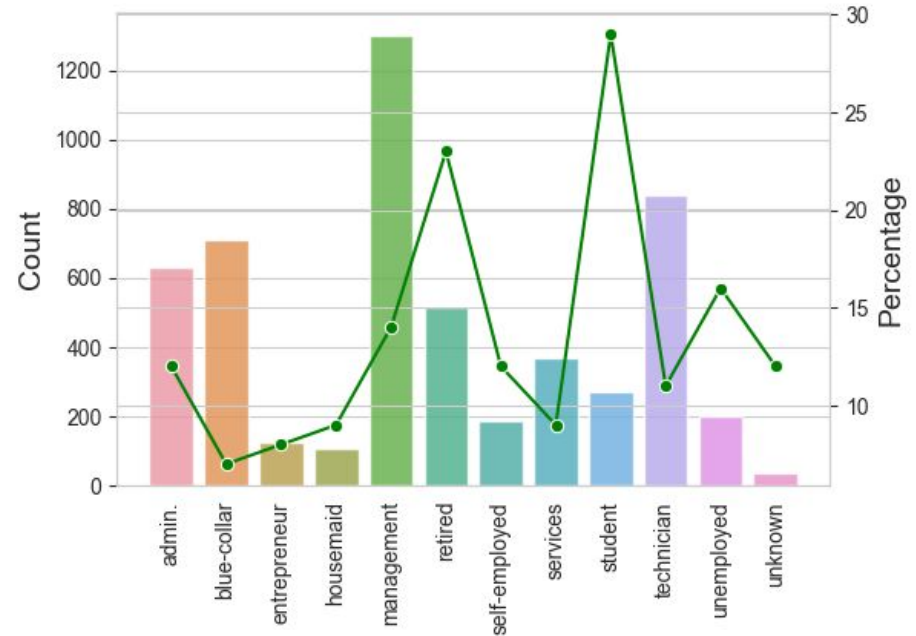
- The value to be predicted is the deposit
- Highly Imbalanced Data
- Around 12% of customers subscribe to a term deposit



Comparing the number of customers who have subscribed (1) and those who have not subscribed (0) a deposit.

Which profession are more likely to get term deposit?

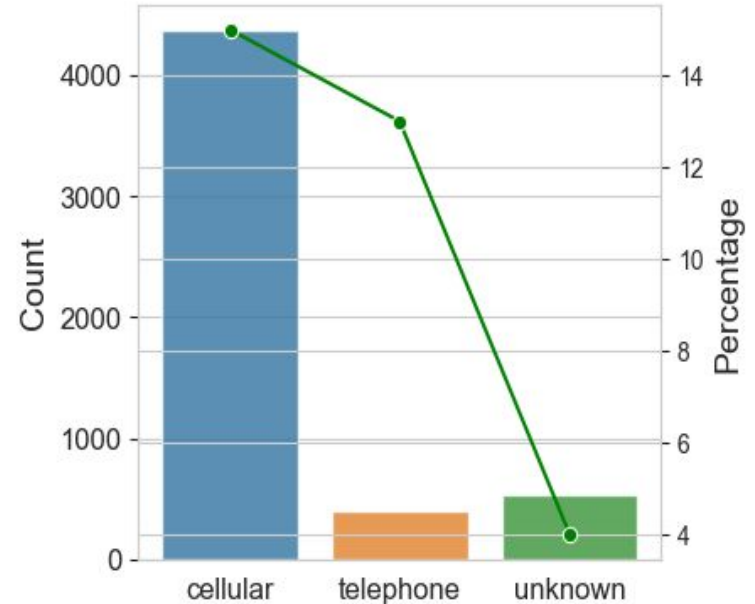
- Students and retirees may be more inclined to subscribe to a term deposit due to their stable income and risk-averse nature.
- Blue-collar workers less likely to invest in term deposits due to their lower disposable income or preference for short-term financial goals.



Customer professions and subscription: Count is total subscriptions per profession and percentage is subscriptions relative to profession group size.

How means of communication can play a role ?

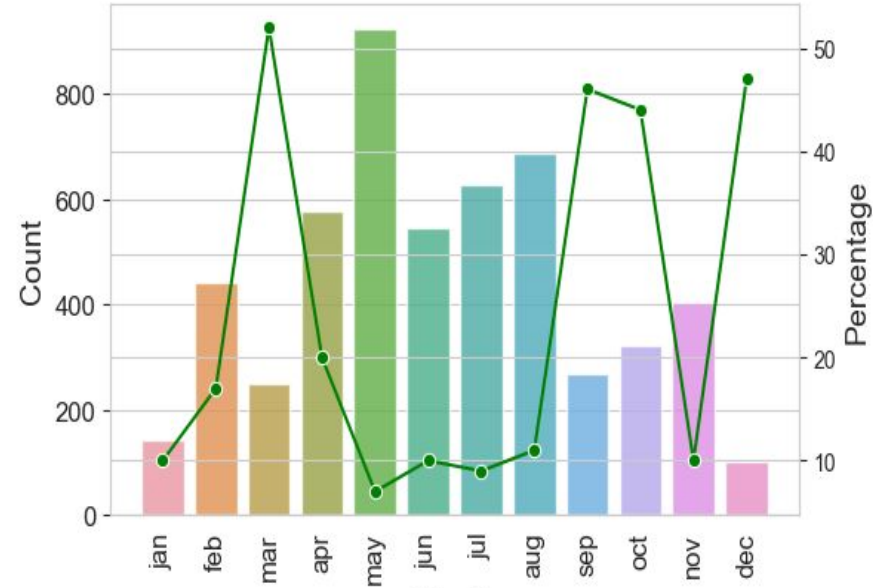
- The most common mode of contact is through cellular.
- The success rate for cellular is slightly better for phone calls.
- May be due to people being less available to answer calls.



Means of communication and subscription: Count is total subscriptions per contact method and percentage is subscriptions relative to group size.

Is specific month(s) leads to more subscription?

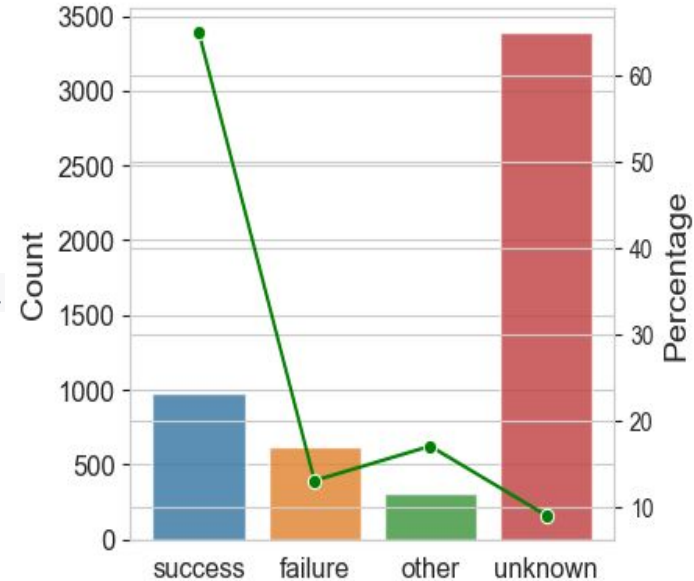
- March and December have the highest deposit subscription ratios, while May has the lowest.
- Use this information to strategically plan marketing
- May investigate the reasons for the lower subscription rate in May.



Month and subscription: Count is total subscriptions per month and percentage is subscriptions relative to the number of contacts.

Successful Previous Campaign lead to more Subscription

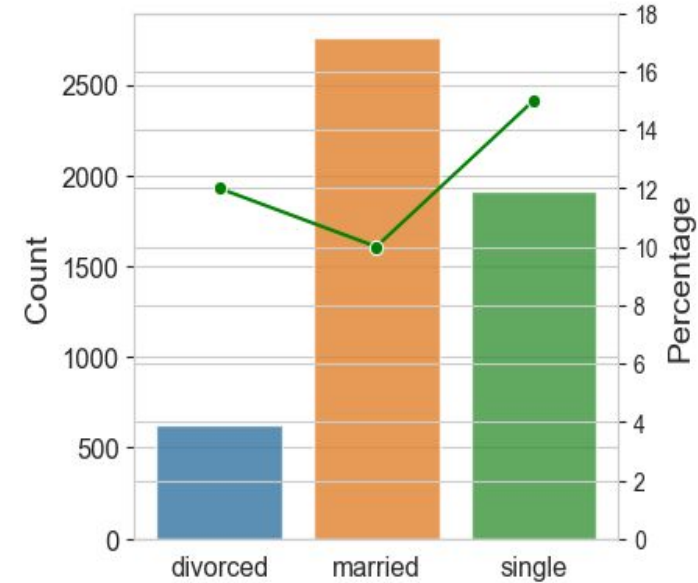
- Previous campaign success positively affects clients' acceptance
- Marketing efforts can influence customers' decisions to subscribe.
- Analyze successful campaigns to replicate factors in future efforts.
- Investigate unsuccessful campaigns for improvement.



Previous campaign result and subscription: Count is total subscriptions per group and percentage is subscriptions relative to group size.

Single Customers are more likely to subscribe

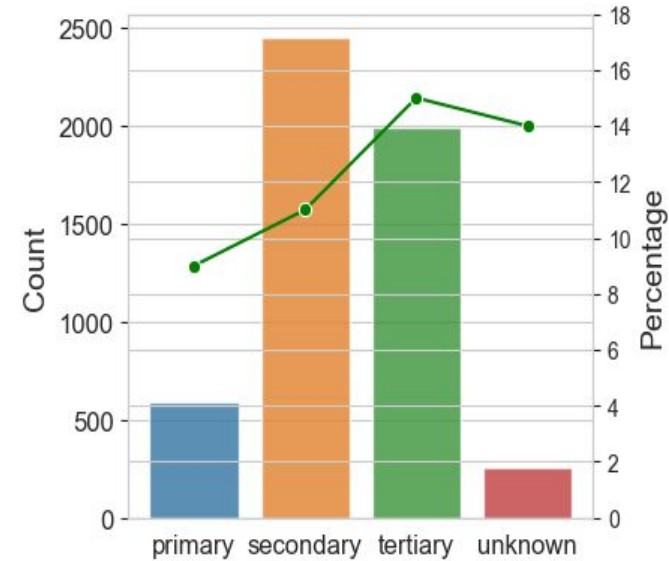
- Marital status impact a customer's likelihood of subscription
- Single customers are 4% more likely to subscribe
- Tailoring marketing efforts to specific demographics, such as singles, may encourage more subscriptions.



Marital status and subscription: Count is total subscriptions per group and percentage is subscriptions relative to group size.

How different level of education can impact?

- Education level can impact a customer's likelihood of subscription
- Tertiary-educated customers are twice as likely to subscribe compared to those with lower education levels.
- Marketing efforts could be targeted towards higher-educated customers to encourage more subscriptions.



Education and subscription: Count is total subscriptions per group and percentage is subscriptions relative to group size.

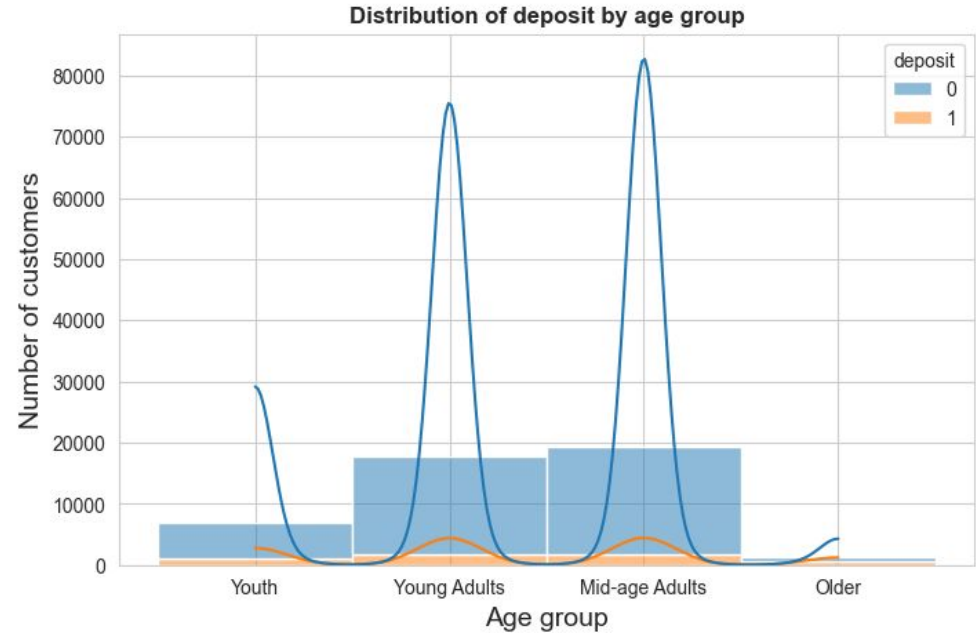
Look at some important Numerical Features



Age, Balance and Duration

Age: target the youngest and the oldest

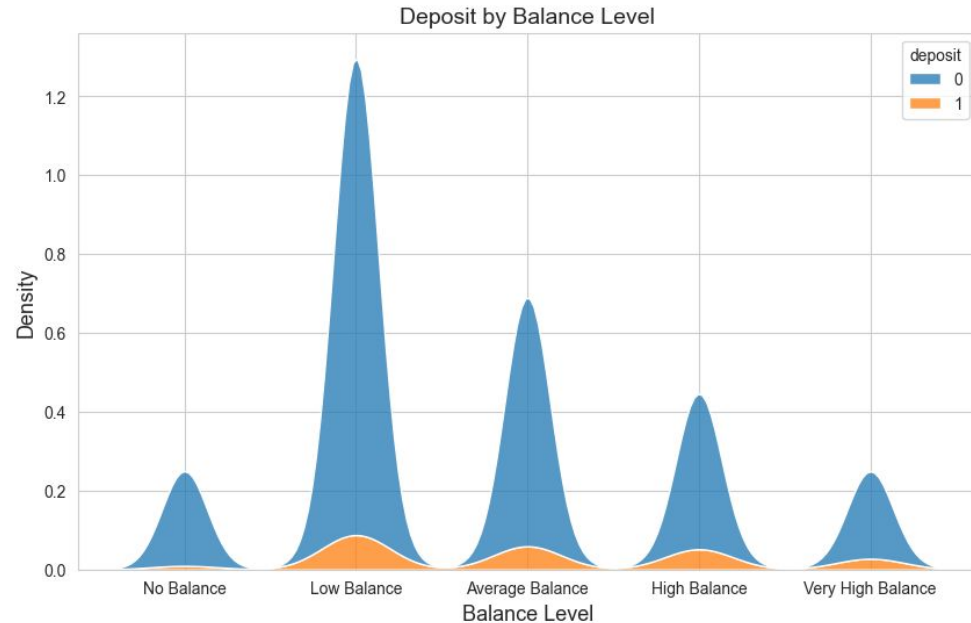
- Most Subscribers are Youth or Older.
- Bank contacted, young adults, middle-aged.
- Age is a significant factor
- Older customers are more ready to deposit



Age group vs Nr Customers

Balance: Any relation between subscription rate and balance?

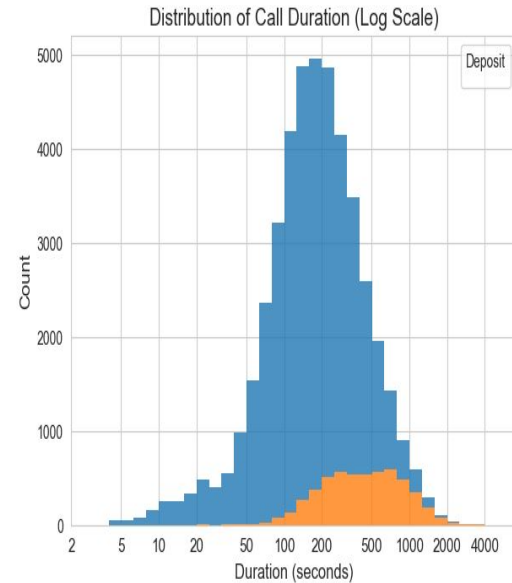
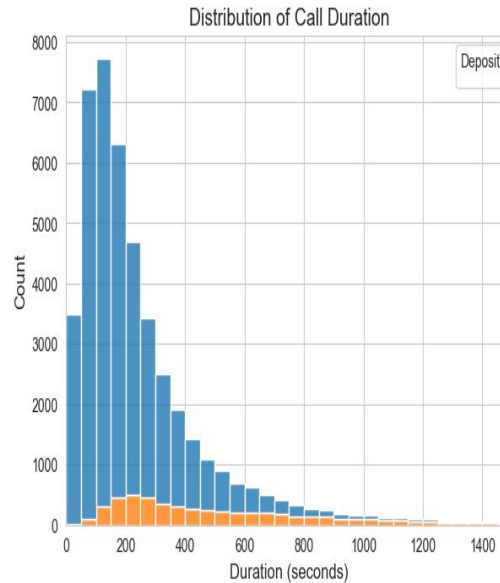
- The kernel density plot, visualize
 - Relation between balance and deposit
- "No Balance" : not making a deposit
- "Low Balance" :the lowest deposit density.
- Highest density belong to higher balances, "High Balance" and "very high Balance" respectively



The plot shows the density distribution of deposit behavior based on balance level.

Longer call durations leads to more Deposit

- Duration feature is important for predicting class labels.
- The right-skewed plot indicates that most calls are relatively short.
- The majority of the call duration lies in the range of 50 to 900 seconds.
- Customers who made a deposit tended to have longer call durations



Log transformation made the distribution more symmetric and easier to compare visually between the two groups



Data preparation & Modeling

Data Preparation for Modeling



- **Remove Unrelated Variables**

- Day,Month:Reduce the dimensionality; emp.var.rate: Avoid Multicollinearity

- **Apply imputation on some features which is labeled as unknown**

- By Randomly distributed or Segmentation

- **Encode categorical variables**

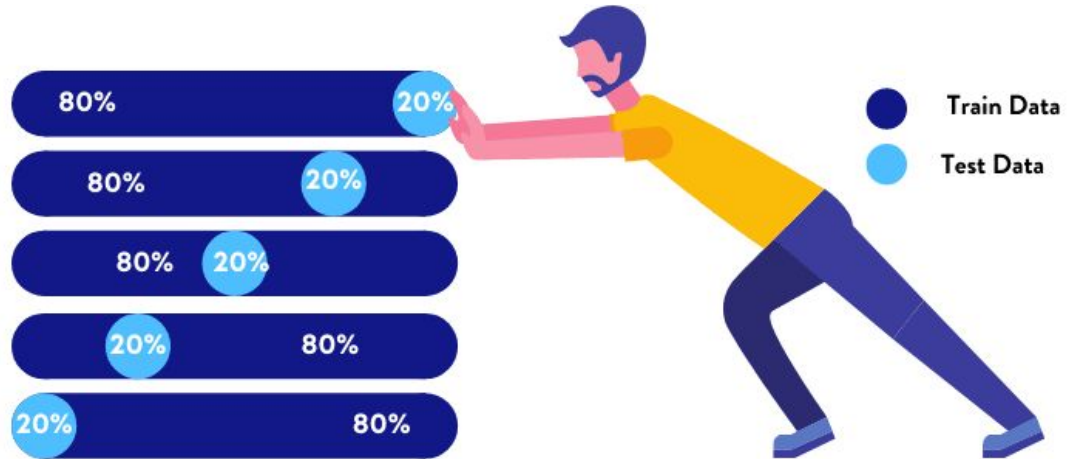
- Dummy encoding: better for Imbalanced data

- **Feature Scaling**

- Robust Scaler : unaffected by outliers

Select best model: Cross Validation

- Use Stratified k-fold
- No winner, perform similarly
- Logreg(0.90), KNN and RF(0.89)



Cross Validation: Retrieved from Google

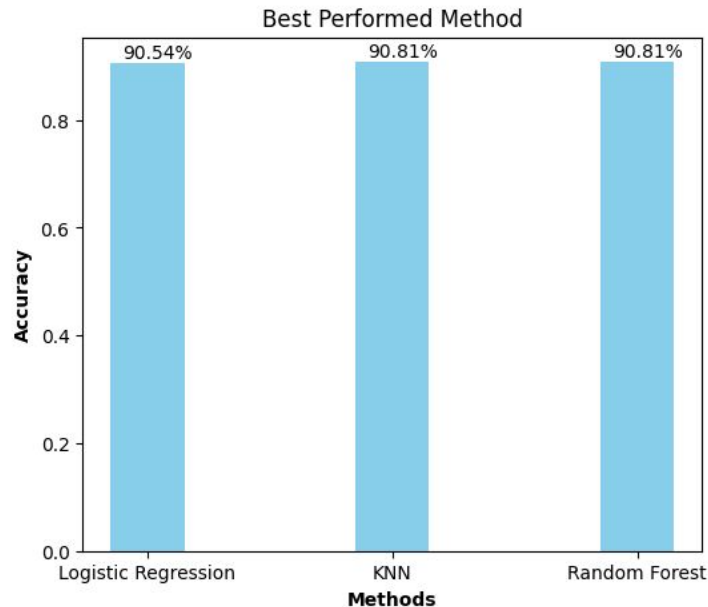
Modeling



- **Optimization**
 - Tune the hyperparameters of each model
- **Experiment on different class weight**
 - Minimize the misclassification on imbalance data
- **Undersampling and Oversampling techniques**
 - Model were tuned on the entire training data and compare o test set
- **Additional features were added**
 - Improve accuracy & ROC

Results: Model Evaluation Based on Accuracy

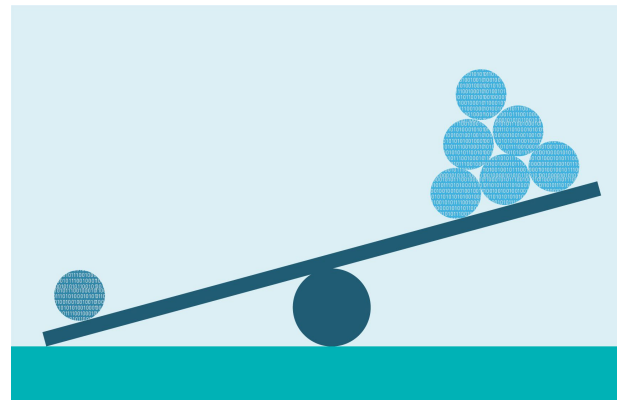
- All the models, have quite good performance
- Adding features improved the performance
- Better picture, other evaluation metrics may be more informative.



The Best performed Model based on Accuracy

Which one is more important Accuracy or ROC?

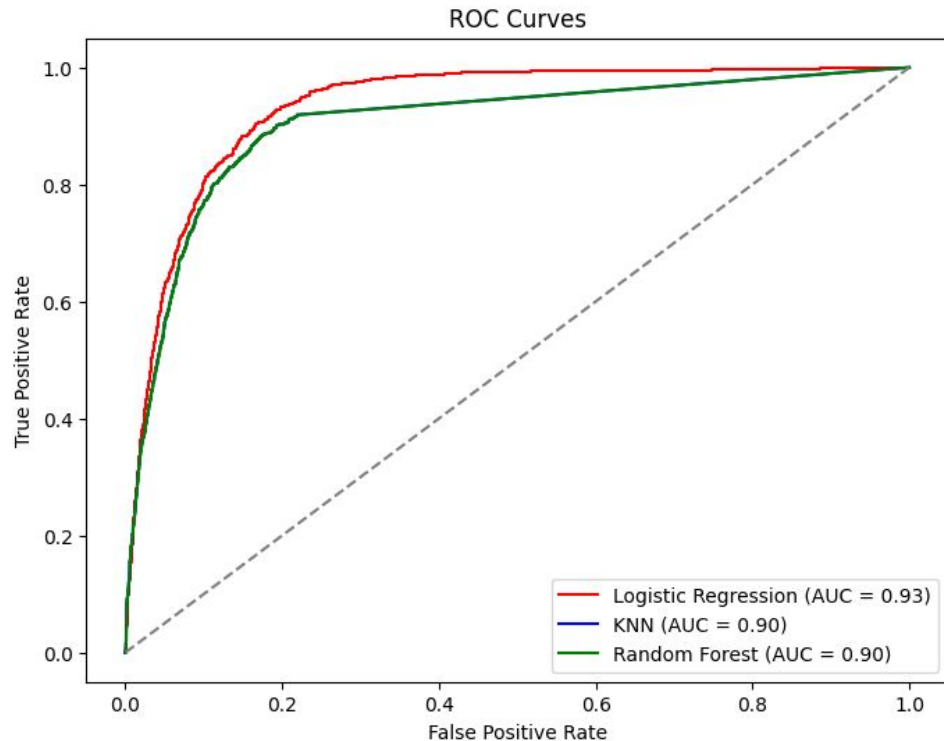
- Accuracy, be misleading because might ignore minority class.
- ROC considers TPR and FPR across different thresholds
- More comprehensive picture,
 - Precision, recall, and F1-score are also useful



Cross Validation: Retrieved from Google

ROC Curve & AUC

- Adding more features improved the accuracy
 - Better representation by separating the classes
 - Reduce Noise
 - Interaction effect
- The ROC curve shows a **logistic regression** perform better
 - Linear relationship
 - Less prone to overfitting
 - Sensitive to feature scaling



Further Technical Exploration



- The Logistic Regression model demonstrated better recall and f1-score for the minority class, indicating its suitability for predicting term deposit subscriptions for the bank.
- More tuning the hyperparameters
- Explore other classification methods such as Naive Bayes to compare their effectiveness in handling imbalanced data.
- Research on `class_weight` parameters could help to minimize misclassification and improve model performance.

Recommendation



- Focus term deposit campaign on **students, retirees, and high-balance customers.**
- Prioritize **contacting** customers **less than 10 times** and use **cellular** contacts.
- Reach out to customers who **agreed to previous campaigns** and increase memory retention with more recent contacts.
- **Avoid making too many calls**, with almost no chance of success after 26 calls.

How Improve the Result?



- Ensuring data completeness
 - Missing value (unknown values)
 - Reduced sample size, bias and inaccurate conclusion
- Collecting additional relevant features
 - Better understanding the relationship between variables
- Document the data
 - in a way that is easy to understand and use

Thank you for your attention!

