Predict Upcoming Marketing Campaign Performance and Who to Target

Predictive modeling plays a crucial role in the realm of marketing campaigns, enabling businesses to make data-driven decisions and optimize their strategies for maximum success. By leveraging historical data and relevant features, predictive models can provide insights into the potential outcomes of upcoming marketing campaigns. In this data science project, we aim to build a predictive model specifically tailored for analyzing the success of a marketing campaign for a bank. By examining various client-related attributes and incorporating information from previous campaigns, our project aims to forecast the likelihood of positive campaign outcomes. This predictive model will empower the bank to make informed decisions, allocate resources effectively, and tailor their marketing efforts to target clients who are more likely to respond positively.

In the realm of marketing, understanding the impact and effectiveness of campaigns is crucial for businesses. By analyzing the provided dataset, we can gain valuable insights into the factors that influence campaign outcomes and identify key trends to optimize marketing efforts. This analysis can help organizations allocate their resources efficiently, target the right audience, and improve the overall return on investment (ROI) of their marketing campaigns.

1. Data

This Kaggle dataset is a synthetic dataset which contains information on bank customers as well as information of outcomes of previous marketing campaigns.

https://www.kaggle.com/datasets/khanimar/marketing-campaign-analysis-data

2. Method

The column target, represents the customer response to the current marketing campaign. It is “yes” or “no”, which we define as “successful” and “failed” respectively.

3. Data Cleaning

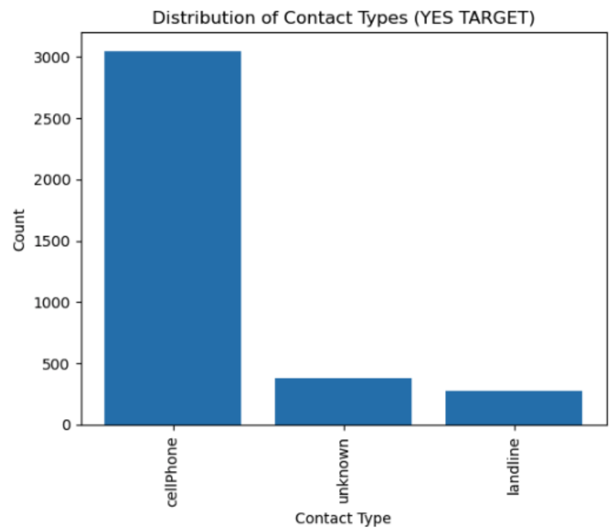
Data Cleaning Notebook: https://github.com/expl0ding/Capstone3/blob/main/Capstone_3_Data_Clean.ipynb

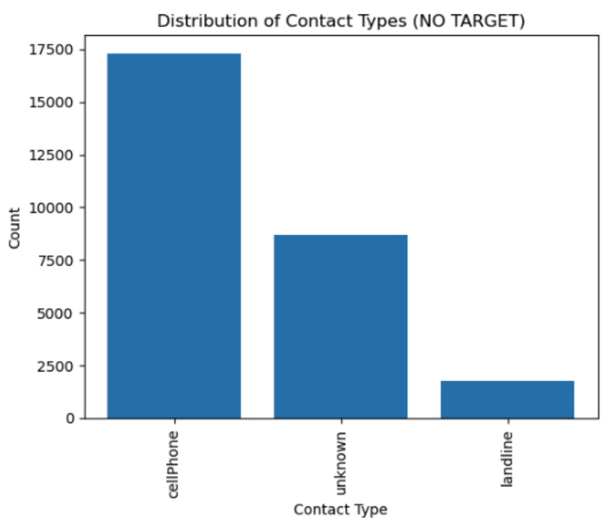
The dataset was fairly clean already. There was only one issue that needed to be addressed. There are newer customers that was not a part of the campaign prior to this one, so there were a lot of null values that needed to be filled in the “day since last campaign” column. Those null values were replaced with -1.

4. EDA

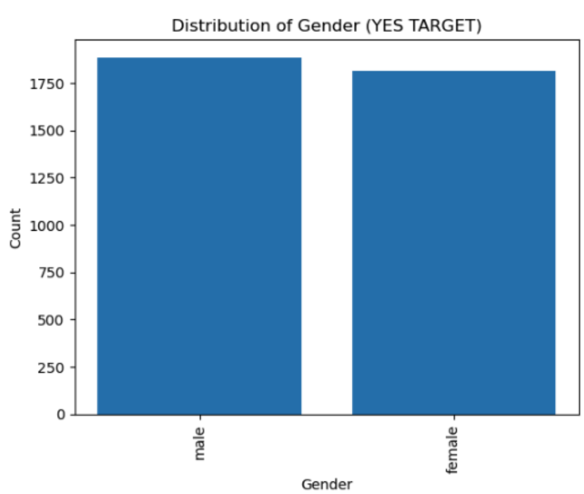
EDA Notebook: https://github.com/expl0ding/Capstone3/blob/main/Capstone3_EDA.ipynb

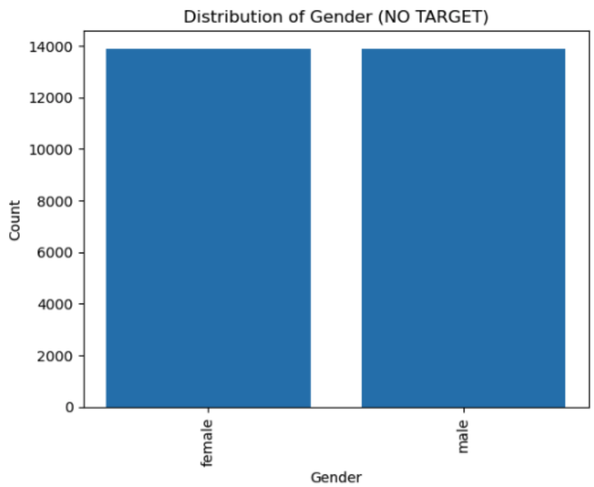
Contact type does not seem to make much of a difference between the converted versus the not converted.



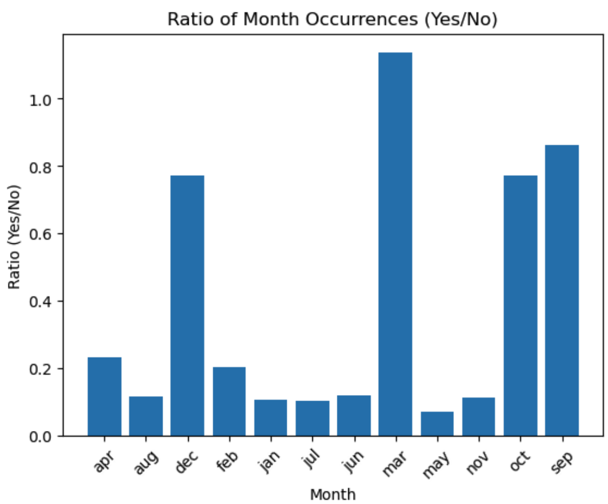


There is also not much of a difference between targeting the different genders.

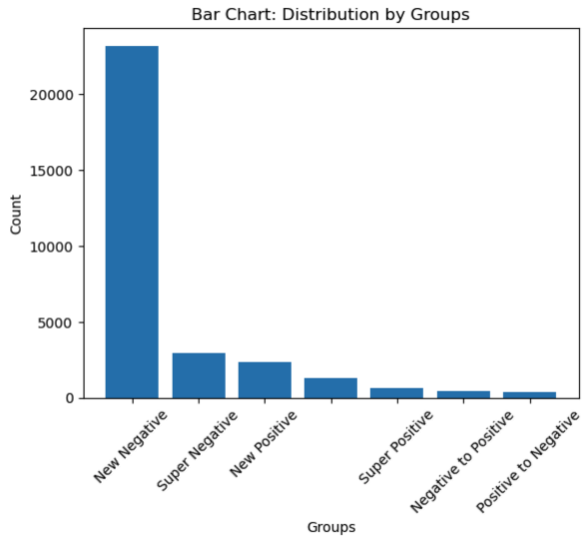




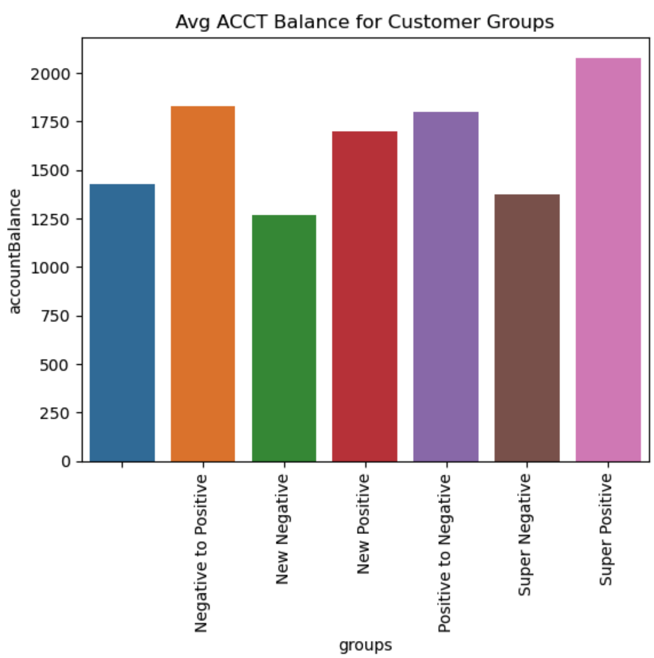
Many other features also did not seem to impact the outcome of the campaign. However, it does seem that the months of March, December, Sept, and October are the highest likelihood for the campaign to turn out successful.



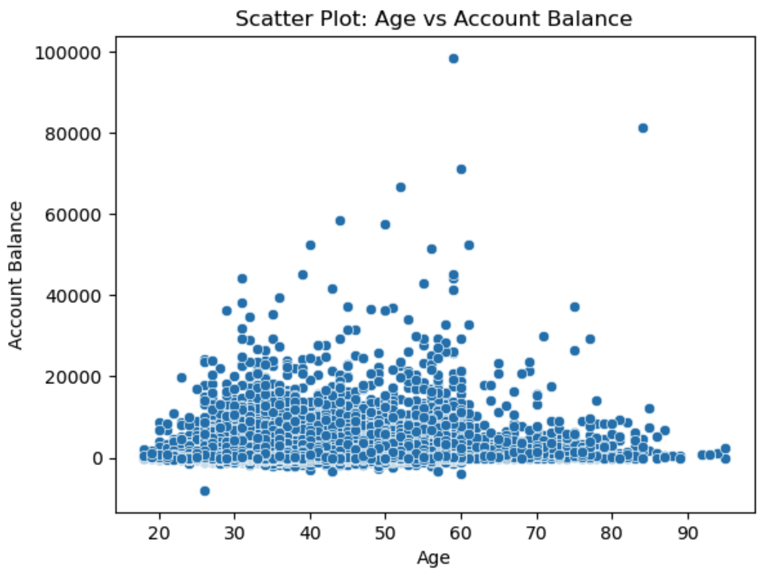
It does look like a majority of the customers are new who did not convert in this campaign.



And those with higher account balances tend to convert.



It does seem like there are a few outliers of those with extremely high account balances, we'll get rid of those and look at the ones that have below $60,000.



4. Machine Learning

Modeling Notebook: https://github.com/expl0ding/Capstone3/blob/main/Capstone3_Modeling.ipynb
Model Metrics File: https://github.com/expl0ding/Capstone3/blob/main/Model_Metrics_Capstone3.txt

After experimenting with several models, such as Decision Tree, Logistic Regression, Random Forest, Gradient Boosting, the Random Forest model had the highest accuracy of 91%, as well as the best ROC curve for True Positive Rate, making it the most suitable model for our problem.

Based on this model, the most important features for accurate classification are the duration of the campaign, account balance, and age.

5. Final Predictions + Improvements

In conclusion, the random forest model developed to predict the outcome of a marketing campaign based on customer and campaign attributes has proven to be most effective as it incorporates important features to make accurate predictions. By utilizing this model, marketers can optimize their efforts by focusing on specific customer segments and tailoring campaigns accordingly.

One key advantage of using this model is the ability to allocate resources more efficiently. By identifying the customers who are most likely to respond positively to a campaign, marketers can prioritize their efforts and allocate resources where they are most likely to yield favorable results. This targeted approach enables better utilization of time, money, and manpower, leading to a higher return on investment. Moreover, the model provides insights into the importance of different customer attributes in determining campaign outcomes. Campaign duration, customer account balance, and customer age have emerged as top features, indicating their significant influence on predicting marketing success. With this knowledge, marketers can focus on these influential factors during campaign planning and execution, ensuring that their efforts are aligned with the characteristics of the target audience.

By leveraging the power of the random forest model, marketing teams can move away from a one-size-fits-all approach and adopt a more personalized and customer-centric strategy. This allows for better engagement, improved customer satisfaction, and ultimately, higher conversion rates. The model empowers marketers to make data-driven decisions, refine their targeting strategies, and create more impactful campaigns that resonate with individual customers.

In conclusion, the random forest model for predicting marketing campaign outcomes provides marketers with a powerful tool to optimize their efforts. This personalized approach leads to better engagement, higher conversion rates, and ultimately, improved marketing success.