<u>Executive Summary of Analysis</u>

Introduction

The purpose of this study was to maximize the revenue from direct marketing campaigns given the socio-demographic and customer account data. The in-depth analysis was performed where various machine learning classification models were trained to predict propensity of a client towards buying marketing offers like Consumer Loan, Mutual Funds and Credit Card. The optimization strategy was applied to maximize the revenue generated from various Offers. The Classification and Regression Models were created to provide answers to the proposed questions in the study. The panda profiling was used to understand the data insights and the analysis was done on the Jupyter Notebook.

Data Analysis

The generated 'panda profiling' report gave information about the missing data by plotting graphs, correlations between different features and hence important features were explored for further analysis. The missing data was represented using various graphs like Matrix, Heatmap and Dendrogram. In socio-demographic data, the range of age was found to be varied from 0 to 100 (analysed using histogram). The missing values were later corrected by filling with zero, mean and median after analysis. Feature Scaling was performed on some features to follow normal distribution which improved the accuracy in Sale Prediction for Offers.

After thorough analysis of these features, it was found that Propensity can be estimated using ***Classification Models*** and could provide solution to the questions discussed previously (related to higher propensity of clients). Also, the revenues for the marketing offers could be predicted by building a ***Regression Model*** or using ***Mean Values*** from Training data. In order to build a Machine Learning Classification Model, the important features were captured using two ***Feature Selection*** techniques known as ***Recursive Feature Elimination*** (RFE) and ***Correlation*** statistics. The RFE method uses accuracy of model to determine which features contributed the most in predicting the target. The correlations between the target variable and other features were determined which also helped in selection of important features for modelling.

<u>*Machine Learning Classifiers*</u>

Logistic Regression, Random Forest, ***Stacking*** (Logistic Regression + XGBoost + Gradient Boost), and ***Deep Learning***. ***The Deep Leaning model gave the satisfactory results on validation using metrics.***

These algorithms were employed to determine the Propensity for every product offer based on the ***Evaluation Metrics*** (Recall, Precision, AUC, Model Accuracy and Loss), Deep Learning model was best suited which gave an *Accuracy of* ***75%*** with *Loss as* ***0.55***. The ***Regularization*** was performed so that overfitting of training data can be avoided using the *Early Stopping* method. The Validation was performed on 15% of the training set data and required graphs were plotted on various epochs.

<u>*Strategy to Maximize the Revenue for Marketing Offers*</u>

It was assumed that ***'Sale Propensity is proportional to the Revenue',*** higher propensity will leads to higher revenues***.*** The regression model was initially created using Linear Regression to predict the revenue for different product offers (CL, MF and CC). The Evaluation of the regression model was performed and it was found that results from 'R adjusted score' and other metrics were not satisfactory. The relationship was plotted between the actual value and the predicted values (generated from the model) but majority of the points were lying far from the estimated line. *Hence it was assumed to fill the predicted revenue with the product offer's individual mean (as we know from Statistics that Expected Value is the best Predictor)*.

To maximize the revenue, the expected revenues for each marketing offer were estimated using:

Expected Revenue = Propensity to buy any offer * Predicted Revenue

Propensity for each offer was calculated using the classification model and Predicted Revenue is the individual mean. Hence, Expected Revenue for all offers can be easily calculated as discussed above.

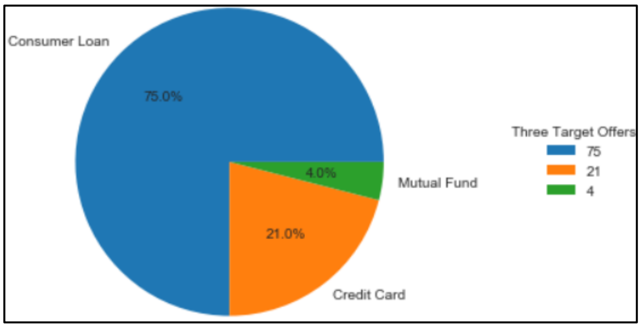
Results and Conclusion

The models used in classification gave good *Accuracy* on the validation set and the 'Adam' optimization algorithm was used to update the weights iteratively. The Sale Propensity was predicted for every offer in the target data (646 records) and the highest propensity list was filtered.

It was assumed that ***'Sale Propensity is proportional to the Revenue'***. Hence the list of 15% of the clients (100 people) having the highest expected revenue was obtained with their respective offers.

Distribution of Offers among Clients:

Marketing Offer	Number of Clients in Each Offer (100 Clients)
Consumer Loan	75
Credit Card	21
Mutual Fund	4



***The total expected revenue for 100 clients was estimated to be 835.50*.**

Further Insights from the Results obtained from the Target Data

- The Consumer Loan clients having good profit were young, working and loyal customers. They had more number of withdrawals as compared to deposits. They had good credits in current accounts as compared to other offer clients (mutual fund and credit card).

- The Mutual Fund clients having maximum profit had more number of withdrawal transactions as compared to deposits. But a good amount is deposited monthly as compared to withdrawal amount. They preferred cashless transactions to withdrawal using cards. The retired customers had less balance in current and saving accounts.

- The Credit Card clients having good profit had more number of withdrawal transactions as compared to deposit. The young credit card customers have more balance in their current account as compared to the retired customers. The retired customers had a good balance in saving accounts as compared to young customers.

- The Expected Revenue and Sale Propensity were related when observed using various scatter plots.