**Executive Summary**

**Objective**

This report analyses the customer data to find trends and make predictions on various marketing offers like Customer loan, Credit cards and Mutual funds by designing probabilistic models. These models will be used by the bank to determine the propensity of a customer to buy a marketing scheme and thus target the customers with the highest propensity to maximize the revenue. The various questions answered in this analysis are: -

* Which clients have higher propensity to buy consumer loan?
* Which clients have higher propensity to buy credit card?
* Which clients have higher propensity to buy mutual fund?
* Which clients should be targeted with which offer?
* What would be the expected revenue based on the analysis?

**Technology Used:**

Jupiter Notebook was used as a tool to develop and visualize the models leveraging python libraries like sklearn, pandas, numpy, matlplotlib etc.

**Analysis**

The target variables were first studied to understand the problem we were dealing with, since the target sale variable was discrete, it was understood that it is a classification problem.

Once it was clear, what kind of algorithms we would be needing, the next step involved determining the features (columns) that would be needed for the modelling. To determine the features, a correlation matrix was computed and sorted by their correlation score with the target variable (for each case) to get the columns which are most suitable for determining the target.

In total, three models were designed (one each for Loan, Credit card and Mutual Fund) using Logistic Regression, Random Forest and Gaussian Naïve Bayes. The probabilities computed by these models were then used to decide which clients to target with which offer. Some Visualizations were also done for showing the basic trends which were observed from the datasets.

**Results**

The data was split into training (85%) and testing (15%) analytical datasets using K-Fold Cross validation, the models were trained using the training dataset and predictions were made on the test datasets.

Accuracy scores for various models:

1. Logistic Regression: Used for Predicting Target Sale of Loan with accuracy: 62.93 %
2. Random Forest: Used for Predicting Target Sale of Credit Cards with accuracy: 79%
3. Gaussian Naïve Bayes: Used for Predicting Target Sale of Mutual Funds with accuracy: 71.32%

Consequently, the test data was then used to determine the propensity of the customers to buy these schemes, the score was then mapped with the clients to determine the best case and the expected direct offer market revenue.

Best case: Revenue from all Customers having Sales Target as 1 and having a positive propensity score.

Revenue From sale of Credit Card = 265.634285708

Revenue From sale of Consumer Loan = 497.7578571720001

Revenue From sale of Mutual Fund = 265.49410713699996

Total Best Case Revenue = 981.83

Expected Case: Not all customers are expected to accept the offer, so a threshold propensity of 0.6 was set and the list was filtered to include only those customers who have propensity greater than the limit.

Expected Revenue From sale of Credit Card = 63.93

Expected Revenue From sale of Consumer Loan = 221.583928589

Expected Revenue From sale of Mutual Fund = 32.973392861

Total Expected Revenue = 318.48732144999997

**Limitations**

1. The models could have been trained better with more data and better feature engineering.
2. A very low correlation was observed between the target and the features which impacted the fitness of the model.

**Conclusions**

The results derived from these models can be used as a basis for developing an intelligent system which could predict the needs of customers based on their usage statistics. This will benefit both the customers and the vendors. With more data and better feature engineering, the accuracy of the models can be improved dramatically, and can be deployed in real world scenarios.