AllLife Bank is a US bank that has a growing customer base. The majority of these customers are liability customers (depositors) with varying sizes of deposits. The number of customers who are also borrowers (asset customers) is quite small, and the bank is interested in expanding this base rapidly to bring in more loan business and in the process, earn more through the interest on loans. In particular, the management wants to explore ways of converting its liability customers to personal loan customers (while retaining them as depositors).

A campaign that the bank ran last year for liability customers showed a healthy conversion rate of over 9% success. This has encouraged the retail marketing department to devise campaigns with better target marketing to increase the success ratio.

You as a Data scientist at AllLife bank have to build a model that will help the marketing department to identify the potential customers who have a higher probability of purchasing the loan.

**Objective**

1. To predict whether a liability customer will buy a personal loan or not.
2. Which variables are most significant.
3. Which segment of customers should be targeted more.

**Data Dictionary**  
\* ID: Customer ID  
\* Age: Customer’s age in completed years  
\* Experience: #years of professional experience  
\* Income: Annual income of the customer (in thousand dollars)  
\* ZIP Code: Home Address ZIP code.  
\* Family: the Family size of the customer  
\* CCAvg: Average spending on credit cards per month (in thousand dollars)  
\* Education: Education Level. 1: Undergrad; 2: Graduate;3: Advanced/Professional  
\* Mortgage: Value of house mortgage if any. (in thousand dollars)  
\* Personal\_Loan: Did this customer accept the personal loan offered in the last campaign?  
\* Securities\_Account: Does the customer have securities account with the bank?  
\* CD\_Account: Does the customer have a certificate of deposit (CD) account with the bank?  
\* Online: Do customers use internet banking facilities?  
\* CreditCard: Does the customer use a credit card issued by any other Bank (excluding All life Bank)?

**Best Practices for Notebook :**

* The notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
* The notebook should be run from start to finish in a sequential manner before submission.
* It is preferable to remove all warnings and errors before submission.
* The notebook should be submitted as an HTML file (.html) and as a notebook file (.ipynb)

**Submission Guidelines :**

1. There are two parts to the submission:
   1. A well commented Jupyter notebook [format - .ipynb]
   2. File converted to HTML format
2. Any assignment found copied/ plagiarized with other groups will not be graded and awarded zero marks
3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation
4. Submission will not be evaluated if,
   1. it is submitted post-deadline, or,
   2. more than 2 files are submitted

Rubric:

| **Criteria** | **Points** |
| --- | --- |
| **Perform an Exploratory Data Analysis on the data**  - Univariate analysis - Bivariate analysis - Use appropriate visualizations to identify the patterns and insights - Any other exploratory deep dive | 5 |
| **Illustrate the insights based on EDA**  Key meaningful observations on the relationship between variables | 5 |
| **Data Pre-processing**  Prepare the data for analysis - Missing value Treatment, Outlier Detection(treat, if needed), Feature Engineering, Prepare data for modelling and check the split | 5 |
| **Model building - Logistic Regression**  - Build the model and comment on the model statistics - Test assumptions - Filter out key variables that have a strong relationship with the dependent variable | 12 |
| **Model performance evaluation and improvement**  - Comment on which metric is right for model performance evaluation and why? - Comment on model performance - Can model performance be improved? if yes then do it | 6 |
| **Model building - Decision Tree**  - Build the model and comment on the model statistics - Identify the key variables that have a strong relationship with the dependent variable | 5 |
| **Model performance evaluation and improvement**  - Evaluate the model on appropriate metric - Comment on model performance - Can model performance be improved? if yes then do it | 8 |
| **Actionable Insights & Recommendations**  - Compare decision tree and Logistic regression - Conclude with the key takeaways for the marketing team - what would your advice be on how to do this campaign? | 5 |
| **Misclassification analysis**  - Do an analysis of all misclassified samples - Try to find a pattern among those | 5 |
| **Notebook - Overall**  - Conclude with the key takeaways for the business - What would your advice be to grow the business? | 4 |