



AIFL

MODULE
PROJECT

Data Visualisation

TOTAL
SCORE

30

- **DOMAIN:** BFSI
- **CONTEXT:** Analytics as a discipline has totally reinvented itself over the past several years, from primarily a support function to one that creates both value and revenue. BFSI marketing departments should keep up with the same pace of change that we’ve seen in other industries in adopting data driven decision making else it will cost their institutions hundreds of thousands or millions of dollars in hard costs and lost revenue opportunity. Further, the ecosystem in which the BFSI companies operate has changed drastically.
- **DATA DESCRIPTION:** The file Bank.csv contains data on 2999 customers. The data include various information collected by the bank on the customer like their demographics, information (age, income, etc.), the customer’s relationship with the bank, the customer response to the last personal loan campaign (Personal Loan) etc.

<ul style="list-style-type: none">• Age - Customer's age completed in years• Experience - Years of professional experience• Income - Annual income of the customer (\$)• Family - Family size of the customer• CCAvg - Avg. spending on credit cards per month (\$)	<ul style="list-style-type: none">• Education - 1: Undergrad, 2: Graduate, 3: Advanced/Professional• Mortgage: Value of house mortgage if any (\$)• Personal Loan: Did customer have a securities account with the bank?• Credit card: Does the customer use a credit card issued by bank?
--	---

- **CASE STUDY:**

The case is about a bank which has a growing customer base. 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 with minimal budget.

The department wants to analyze and visualize data that will help them identify the potential customers who have higher probability of purchasing the loan. This will increase the success ratio while at the same time reduce the cost of the campaign.

- **STEPS AND TASKS:**
 1. **Univariate Analysis: [10 Marks]**
 - A. Import and Read 'Bank.csv'. [1 Mark]
 - B. Identify and share size of the input data. (Number of rows and number of features) [1 Mark]
 - C. Convert the below listed features into categorical features. [1 Mark]
['Education', 'Personal Loan', 'Securities Account', 'CreditCard']
 - D. What is the mean age of our customer base? [1 Marks]
 - E. Observe if there are missing values in the data and report the number of missing values if there are any. [1 Mark]
 - F. Check the frequency distribution of "Personal loan" and report if the classes are balanced or imbalanced. [2 Mark]
 - G. What is the minimum and maximum income of our customer base given in the dataset? [1 Mark]
 - H. Visualize a boxplot for 'Age' and 'Experience' and share your findings for the same. [2 Mark]
 2. **Bivariate Analysis: [14 Marks]**
 - A. How much percentage of the customers who use our credit card have a personal loan with us? [3 Marks]
 - B. Under education column (1. Undergrad 2. Graduate 3. Advanced/professional) which category of customers listed above are more prone to subscribe to our personal loan? [4 Marks]
 - C. A marketing specialist claims that our customers who have personal Loan tend to have higher average spending on Credit Card (CCAvg) than those who are less experienced. With the help of visual analysis, can you prove or disprove this claim? [3 Marks]
 - D. A data scientist looks at this data and comes up with the following claim:- On average, a customer with more than 100,000 \$ annual income is more likely to subscribe to a personal loan than with a customer with less than \$100,000 income. Do you agree with this claim? Will you direct the sales team to focus on customers with more than 100,000 \$ annual income? [4 Mark]
 3. **Data Transformation: [6 Marks]**
 - A. Report the approx. percentage of negative values of "Experience" in the data. Investigate the column properly. Perform data cleaning on the "Experience" column(i.e. Replace negative values with zeros) [4 Marks]
 - B. Calculate the correlation of all the columns in the above dataset. State in which two features the correlation is maximum. Calculate the correlation between cc_avg & income. [2 Marks]