**Problem Statement :**

* Bank Loan Prediction

Description:

The data is about a Bank which Has a growing customer Base . Majority of these customers are liablity customers (depositors) with variying size of deposits .The Number of Customers who are also borrowers (assest Customers ) is quite Small , and 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 Management want to explore ways of converting it liablity Customers to personal Loan Customers.These are Some details are : ID,Age,Experience,Income,Zipcode,Family,CCAvg,Education,Mortage,Personal\_Loan,Securities\_Account,Cd\_Account(Any FD),Online(Yes,No),CreditCard.

It is Classification Problem Where we need to find a probability of Potential Customer as well Identify potential Customer who will Buy Loans !

----------------------------------------------------------------------------After Generating a Problem Statement there Has time to Complete Hypothesis Generation :

**Hypothesis Generation :**

Below are Some Factors which I think that identify Potential Customers and Its Probabilities :

Age : It is Main Factor a person who has above 25 who will take loan .

Income : Income will Also help to Identify to Probablity or Likelihood of Buying Personal Loan

CCAvg : I considered only to visualize to see

Education : It considered to see

Mortage : check to gain Insights

Online : - If It Like Any Scheme online related to Loan

CreditCard : Number of credit holder is likely or equal to Loan borrower so its Importance More

There are Some variable which may be affect the Target variable

---------------------------------------------------------------------------

**Understanding the data :**

Columns of data :

Bank\_Cust.columns:

Index(['ID', 'Age', 'Experience', 'Income', 'ZIP Code', 'Family', 'CCAvg',

'Education', 'Mortgage', 'Personal Loan', 'Securities Account',

'CD Account', 'Online', 'CreditCard'],

dtype='object')

**Data Description:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data Description:** | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ID | Customer ID | |  |  |  |  |  |  |
| Age | Customer's age in completed years | | | |  |  |  |  |
| Experience | #years of professional experience | | | |  |  |  |  |
| Income | Annual income of the customer ($000) | | | | |  |  |  |
| ZIPCode | Home Address ZIP code. | | |  |  |  |  |  |
| Family | Family size of the customer | | | |  |  |  |  |
| CCAvg | Avg. spending on credit cards per month ($000) | | | | | |  |  |
| Education | Education Level. 1: Undergrad; 2: Graduate; 3: Advanced/Professional | | | | | | | |
| Mortgage | Value of house mortgage if any. ($000) | | | | |  |  |  |
| Personal Loan | Did this customer accept the personal loan offered in the last campaign? | | | | | | | |
| Securities Account | Does the customer have a securities account with the bank? | | | | | | |  |
| CD Account | Does the customer have a certificate of deposit (CD) account with the bank? | | | | | | | |
| Online | Does the customer use internet banking facilities? | | | | | |  |  |
| Credit Card | Does the customer use a credit card issued by Universal Bank? | | | | | | |  |

# Print the data types of variables :

**ID int64**

**Age int64**

**Experience int64**

**Income int64**

**ZIP Code int64**

**Family int64**

**CCAvg float64 ----**

**Education int64**

**Mortgage int64**

**Personal Loan int64**

**Securities Account int64**

**CD Account int64**

**Online int64**

**CreditCard int64**

We can see there are two format of data types :

Int64 : It represents the integers variables .Int Variables in our Dataset are id,Age,Experience,Income,Zip,Family,Education,Mortage,PersonalLoan,Securities\_Account,Cd\_acc,CreditCard

Float64 : It represents the variable Which have Some Decimal values Involved .They are Also Numerical Value .

Numerical Variables in Our Data Sets are : CCavg

Lets Looking at the Shape of the Data sets:-

Bank\_Cust.shape

**(5000, 14)**

We have 5000 rows and 14 columns in a Data Sets .

----------------------------------------------------------------------z=