**PROJECT-1**

**Domain:** HBFC **Bank (Personal Loans)**

**-Done by Sidhramappa.H**

1.What percentage of the bank’s customers (according to the data) have availed Personal Loans?

**Sol:** Total number of customers=**5000**

To find total number of customers availed personal loan we used COUNTIF

i.e. **=COUNTIF (K2: K5001,“Yes”)**

Customers using personal loan=**480**

(ii) Calculating the percentage of this 480 customers i.e. **= (480/500)\*100 is equal to 9.6%**

2. Generate a table with min, max, median & average for all numeric variables (age, experience, income, family members, CCAvg, Mortgage). What are your observations?

Sol: First generating the table for these features by taking (**Age, Experience, Income, Family, CCAvg, Mortgage**) in columns and (**Min, Max, Median, Average**) in rows.

Then finding the Min, Max, Median, and Average for each feature

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Column1** | **Age** | **Experience** | **Income** | **Family Members** | **CCAvg** | **Mortgage** |
| Min | 23 | 0 | 8 | 1 | 0 | 0 |
| Max | 67 | 43 | 224 | 4 | 10 | 635 |
| Median | 45 | 20 | 64 | 2 | 1.5 | 0 |
| Average | 45.3384 | 20.1348 | 73.7742 | 2.39723 | 1.937938 | 56.4988 |

From above table we can observer the Minimum, Maximum, Median, and Average for each feature and it’s easy to know the which features as what value and all.

3. Create a new categorical variable for Experience using 4 categories – a. 0 to 10 years b. 11 to 20 years c. 21 to 30 years and d. 30+ years. Plot a bar graph for this new categorical variable

**Sol**: Here with the help of IF statement we have created a new categorial variable for experience

**i.e. =IF(C2<=10,”0-10”, IF(C2<=20,”11-20”, IF(C2<=30,”21-30”,”30+”)))**

Here, the C2 refer to the particular cell of experience column in Excel.

|  |
| --- |
| 30+ |
| 30+ |
| 30+ |
| 30+ |
| 30+ |
| 30+ |
| 30+ |
| 30+ |
| 30+ |
| 30+ |

This are the some of the samples for created as new categorical variable for Experience.

Here to plot this bar chat first in we need to select the newly created Experience categorical column and later plotting the pivot table for the features and later selecting this pivot table values to create the bar chart.

4. Create a scatter plot of the Age and the Experience variable. What do you observe?

Sol: To create a scatter plot, first we need to select the Age and Experience then clicking on the insert button later selecting the scatter chart.

5. What are the top 3 areas (ZIP Codes) where the bank’s customers are located?

Sol: (i) To know the top1area(ZIP Codes), taking the MODE for Income categorical

i.e. =MODE (E2:E5001) which gives the value of top1 area(ZIP Code) i.e. **(94720) is the top 1 area (ZIP Code).**

(ii) To know the top2 area (ZIP Codes) repeating the same with some changes like making another copy of table and filtering the above value (i.e 94720)

i.e., =MODE (E2:E5001) which gives the value of top1 area (ZIP Code) i.e. **(94305) is the top 2 area (ZIP Code).**

(iii) Repeating the above procedure again to know the top3 area (ZIP Codes)

i.e., =MODE (E2:E5001) which gives the value of top1 area (ZIP Code) i.e **(95616) is the top 3 area (ZIP Code).**

6. How many customers have a combination of Fixed Deposits and Credit Cards but not Personal Loan?

Sol: First we need to create a new column for the customers having a combination of fixed deposits and credit cards but not personal loan this can be achieved by if statement as below

**i.e. =IF(AND(M2="Yes",O2="Yes",K1="No"),"yes","no")**

Later to find how many customers having the combination of these three conditions we can use the COUNTIF for the Yes.

**i.e. =COUNTIF(Table24[column1],"yes")**

**147 customers have a combination of fixed deposits and credit cards but not personal loan.**

7. What is the median income of the customers who have availed personal loans and compare it with the median income of those customers who have not availed personal loans? What do you infer?

Sol: Filtering the table for customers who have availed personal loans and calculating the median income.

**i.e. =MEDIAN(Sheet1!D42:D4893)**

Once again filtering the table for customers who have not availed personal loans and calculating the median income.

**i.e. =MEDIAN(Sheet1!D2:D5001)**

**In the above two cases we got the same median value i.e. 64** is the median income for the customers who have availed personal loans as well for customers who have not availed personal loan

8. Create 4 separate Pivot Tables. Summarise your data by percentages.

• Education vs Personal Loan

• TD Account Vs Personal Loan

• Online vs Personal Loan

• Income\_Category vs Personal Loan

Sol: For creating the pivot table for Education vs Personal loan, first selecting the full table then clicking on insert button then selecting pivot table.

Putting the Education in rows and Personal loan in columns as well as in values, later right clicking on the personal loan in values and selecting value field settings then clicking on show value as in that select the % of the column total.

Which creates the table with personal loan column and education with there attributes and the percentage of no and percentage of yes for education and the grand total also generated.

Repeating the same for other 3

i.e., for • TD Account Vs Personal Loan

• Online vs Personal Loan

• Income\_Category vs Personal Loan

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Count of Personal Loan** | **Eduction** |  |  |
| **Row Labels** | **No** | **Yes** | **Grand Total** |
| Graduate | 27.01% | 37.92% | 28.06% |
| Professional | 28.67% | 42.71% | 30.02% |
| Undergraduate | 44.31% | 19.38% | 41.92% |
| **Grand Total** | **100.00%** | **100.00%** | **100.00%** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| **Count of Personal Loan** | **TD Account** |  |  |
| **Row Labels** | **No** | **Yes** | **Grand Total** |
| No | 96.42% | 70.83% | 93.96% |
| Yes | 3.58% | 29.17% | 6.04% |
| **Grand Total** | **100.00%** | **100.00%** | **100.00%** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| **Count of Personal Loan** | **Online** |  |  |
| **Row Labels** | **No** | **Yes** | **Grand Total** |
| No | 40.42% | 39.38% | 40.32% |
| Yes | 59.58% | 60.63% | 59.68% |
| **Grand Total** | **100.00%** | **100.00%** | **100.00%** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| **Count of Personal Loan** | **Income\_Categorical** | |  |
| **Row Labels** | **No** | **Yes** | **Grand Total** |
| 0-50 | 42.35% | 0.00% | 38.28% |
| 100+ | 17.12% | 91.25% | 24.24% |
| 51-100 | 40.53% | 8.75% | 37.48% |
| **Grand Total** | **100.00%** | **100.00%** | **100.00%** |

9. Analyse the Pivot tables created in the previous question and state any anomaly that you observe. Which categorical variables appear most important for your further study if you want to analyse which customers are most likely to take personal loans and why?

Sol: In 1st pivot table i.e., Education Vs Personal Loan

Graduate, Professional’s “Yes” percentages is more and undergraduate’s “Yes” is less which means customers many customers are Graduate and Professional compared to undergraduate.

In 2nd pivot table i.e., TD Account Vs Personal Loan

The percentage of customers having the TD Account is very less compared to percentage of customers who don’t have the TD Account.

In 3rd pivot table i.e., Online Vs Personal Loan

The customer’s percentage is more for online which means customers are preferring for the online

In 4th pivot table i.e., Income\_categorical Vs Personal Loan

The income between 0-50, the income\_categorical is “NO”

The income between 51-100, the income\_categorical is “No”

The income above 100, the income\_categorical is “Yes”

According to my analyse the Education Vs Personal Loan categorical are more important for future and this categorical customers are most likely to take personal loan because most of them are graduate and professionals they might need the Personal loan for their future and they also aware of everything about the Personal loan.

10. In the last campaign, bank reached out to 5000 customers out of which 480 customers accepted the personal loan offer. The bank incurred a huge cost in running a marketing campaign to reach out to so many customers. This is where you as a strategic business consultant step in. You are tasked to optimise the cost of this campaign by identifying the correct target base (without significant reduction in number of acceptances of offers). The bank can then send Personal Loan offers to these target customers who have a higher chance of accepting the offer. Based on your analysis, what strategy would you suggest to the management of HBFC bank?

Sol: The Bank can offer the personal loan to all the categories based on the requirements of the customers and the bank can offer the more personal loans to the education sector because they are one who requires the personal loans and in education sectors most of the peoples are graduate and the professional, they might aware everything about the personal loans.

The term deposit account hold customers might also need this personal loan for the depositing the money in bank in case where they are not able to pay the money at any situation.

Offering the personal loan for the customers having the more income, this means that by providing the personal loan to such categorical customers with mortgage which will be safer for the bank, and it will also create the responsibility about their personal loan.

At present all the transactions will be in online mode so bank should look for his customers having more transactions in online then to decide to give or not give loan to his customers, if is online transactions are more then bank should give him a personal loan.