assignment-1

K.Janaki Kiruthika

1)What percentage of the bank’s customers (according to the data) have availed Personal Loans vs the ones who have not availed it?

From the pivot Table, extracting the values of Personal loan in column, finding the count of values in percentage, the following is obtained. 9.6% has availed personal loans.

|  |  |
| --- | --- |
| **Row Labels** | **Count of Personal Loan** |
| No | 90.40% |
| Yes | 9.60% |
| (blank) | 0.00% |
| **Grand Total** | **100.00%** |

To make the values more available for analysis, Pie chart is made from the above values.

From the above, we can conclude that 9.6% has availed the personal loan and 90.40% has not availed the personal loan from the bank.

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

By applying the formulas to find MIN,MAX,AVERAGE,MEDIAN and selecting the range of the Age ,Experience, Income, family members, cc avg, mortgage the following table is displayed.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **NO** | **Age** | **Experience** | **Income** | **Family members** | **CCAvg** | **Mortgage** |
|  | (in years) | (in years) | (in K/year) |  |  |  |
| MIN | 23 | 0 | 8 | 1 | 0.0 | 0 |
| MAX | 67 | 43 | 224 | 4 | 10 | 635 |
| AVERAGE | 45.3384 | 20.1348 | 73.7742 | 2.3972 | 1.9379 | 56.4988 |
| MEDIAN | 45 | 20 | 64 | 2 | 1.5 | 0 |

3) Create a new categorical variable for Experience using 4 categories

– • 0 to 10 years

• 11 to 20 years

• 21 to 30 years

• 30+ years.

Plot a bar graph for this new categorical variable.

By using the nested if condition , ie) =IF(A2<=10,"0-10",IF(A2<=20,"11-20",IF(A2<=30,"21-30","30+"))) across the experience column, we can obtain the following values from the Pivot table created between the extracted values from the categorial value.

|  |  |
| --- | --- |
| **Row Labels** | **Sum of Experience (in years)** |
| 0-10 | 6.86% |
| 11-20 | 19.60% |
| 21-30 | 32.98% |
| 30+ | 40.56% |
| **Grand Total** | **100.00%** |

After extraction for the better understanding , we can create a Bar chart

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

By listing the values of the age and experience from the sample table, scatter plot is made.

The trendline shows that there is a linear increase in age and experience together.

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

Step 1: Removing Zip code duplicates we get the filtered 468 values out of 5001 customers.

Step 2: Creating Pivot table, we can get the count of each zip code values.

Step 3: sort the highest to lowest to find the top 3 zip code areas.

|  |  |
| --- | --- |
| **Row Labels** | **Count of ZIP Code without duplicates** |
| 94720 | 169 |
| 94305 | 127 |
| 95616 | 116 |

OTHER METHOD is to use : COUNTIF($A$2:$A$5000,A2)) to obtain the values after filtering.

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

By using filtering technique, we can extract the customer who has availed Fixed Deposits and Credit Cards but no personal loan, 147 customers had no 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?

By filtering, customers with and without personal loans are listed. Median is calculated from formula.

The median income of the customers who have availed personal loans is 142.5k.

The median income of the customers who have not availed personal loans is 59k.

From the above we can infer that the customers with an average salary of 142.5k / year have availed personal loans and customers with salary less than that had not availed personal loans.

8) Create 4 separate Pivot Tables. Summarize your data by percentage values.

• Education vs Personal Loan

• TD Account Vs Personal Loan

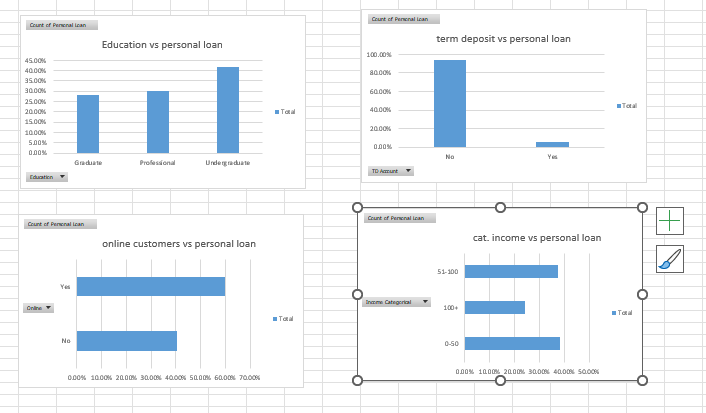
• Online vs Personal Loan

• Income\_Category vs Personal Loan

*a) • Education vs Personal Loan*

|  |  |
| --- | --- |
| **Row Labels** | **Count of Personal Loan** |
| Graduate | 28.06% |
| Professional | 30.02% |
| Undergraduate | 41.92% |
| **Grand Total** | **100.00%** |
| **b)** • TD Account Vs Personal Loan |  |
|  |  |
|  |  |
| **Row Labels** | **Count of Personal Loan** |
| No | 93.96% |
| Yes | 6.04% |
| **Grand Total** | **100.00%** |
|  |  |
| **c)** **Online vs Personal Loan** |  |
| **Row Labels** | **Count of Personal Loan** |
| No | 40.32% |
| Yes | 59.68% |
| **Grand Total** | **100.00%** |
| **d)** **Income\_Category vs Personal Loan** |  |
| **Row Labels** | **Count of Personal Loan** |
| 0-50 | 38.28% |
| 100+ | 24.24% |
| 51-100 | 37.48% |
| **Grand Total** | **100.00%** |

BAR and Column charts are also created to analyze the above values.



9)Analyze 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 analyze which customers are most likely to take personal loans and why?

From the above pivot tables, UG graduates, online customers availed more personal loans than the other customers. If customers had made Term deposits , the customers availing personal loan is less when compared to other categorical values Income category from 0-50k will avail more personal loans than other category . I think analysis online customers and 0-50k , 51-100k income category has to be focused more as they are more likely to personal loans for their needs

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 optimize 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?

The bank should out reach to online customers as their first priority as 60% of the customers have availed personal loan as they are the target customers. Income category with 0 to 50k has to be the next target. This focus will help the bank to optimize the cost of this campaign without significant reduction in number of acceptances of offers.