# **PROJECT HBFC**

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

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

**9.60%** of the bank customers have availed the Personal Loan.

The evaluation is done using the **Pivot table**.

**STEPS:**

* Select the data and insert the **pivot table.**
* Add the **Personal loan** category to the **rows** and also to the **values** section in the pivot table.
* And set the values to count of personal loan and reflect the values as percentage in the **value field settings**.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Numeric  Variables | min | max | Average | median |
| Age | 23 | 67 | 45.3384 | 45 |
| Experience | 0 | 43 | 20.1348 | 20 |
| Income | 8 | 224 | 73.7742 | 64 |
| Family members | 0 | 4 | 2.3866 | 2 |
| Cc Avg | 0 | 0 | 1.937938 | 1.5 |
| Mortgage | 0 | 635 | 0 | 0 |

* The Avg and Median of Income are different.
* The Avg of Mortgage is 0.

**3. Create a new categorical variable for Experience using 4 categories – 0 to 10 years 11 to 20 years 21 to 30 years and 30+ years. Plot a bar graph for this new categorical variable [Hint – You may make use of if else/nested if statements to accomplish this task. You can refer how Income\_Category has been created in the dataset]**

* The new categorical variable for Experience using the 4 Categories can be obtained using **nested if condition**.
* The condition is set in such a manner that, the Experience below 10 years should fall in the range of 0-10years, Experience above 10 and below 20 in the range of 11-20, Experience above 20 and below 30 in the range of 21-30 and the experience of above 30 years is put in the category 30+.

|  |  |
| --- | --- |
| **exp range** | **Count of exp**  **range** |
| 0-10 | 1171 |
| 11-20 | 1223 |
| 20-30 | 1323 |
| 30+ | 1283 |

* After creating the new column for Experience Variable specifying 4 categories, insert pivot table to analyse the count of Employees within the experience range of 4 categories.
* And plot the Bar graph for the pivot table.

**Bar Graph**

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

According the Scatter plot there is a significant increase in the Experience as the age increases.

* The bank customers of age group 60-70 have the highest experience of 43 years.
* Majority of customers have 5-15 years of experience.
* Customers within the age group 20-30 have less experience.
* Some of them even don’t have any professional experience who are in the age range of 20-30.

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

|  |  |
| --- | --- |
| **Row Labels** | **Count of ZIP Code** |
|  |  |
| 94720 | 169 |
| 94305 | 127 |
| 95616 | 116 |

* Insert **pivot table** and drag the zipcode field into **rows** and **values.**
* And then **sort** the data from largest to smallest.
* The values at the top are the areas where most of the bank customers are located.

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

**147** customers have a combination of Fixed deposits and Credit cards but not Personal loans.

* Create a new column for the customers who have both Fixed Deposits and Credit Cards using **AND** condition.
* And another column for the customers who don’t have Personal Loan but have a combination of Fd and Credit card.
* Then insert Pivot table and drag the new column to rows and values to get the count of customers.

|  |  |
| --- | --- |
| **Row Labels** | **Count of Customers** |
| 0 | 4853 |
| 1 | 147 |
| **Grand Total** | **5000** |

**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?**

* Insert pivot table drag Personal loan field into **Filters** and Income field to **rows**.
* Set the Personal loan filter to YES to calculate the median of the customers who have availed Personal loans.
* Change the filter to NO to calculate the Median of customers who have not availed personal loans.

**MEDIAN: =Median(cells range of income column)**

* 136.5 is the Median income of the customers who have availed Personal loans.
* 108.5 is the Median income of the customers who have not availed Personal loans.

The median income of customers who availed personal loan is greater than the median income of customers who haven’t availed personal loan.

The customers with **high income are more likely to take personal loans**.

**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 [Hint: Please drag Personal Loan to the Columns area while creating the Pivot Table to get the required values]**

**EDUCATION Vs PERSONAL LOAN**

* Select data and insert pivot table.
* Drag the personal loan field into columns and values in the pivot table.
* And Education field into rows.
* In the value field settings select count and show the values as % of row total.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Personal Loan** | **Column Labels** |  |  |
| **Row Labels** | **No** | **Yes** | **Grand Total** |
| Graduate | 87.03% | 12.97% | 100.00% |
| Professional | 86.34% | 13.66% | 100.00% |
| Undergraduate | 95.56% | 4.44% | 100.00% |
| **Grand Total** | **90.40%** | **9.60%** | **100.00%** |

**TD ACCOUNT Vs PERSONAL LOAN**

* Select data and insert pivot table.
* Drag the personal loan field into columns and values in the pivot table.
* And TD account field into rows.
* In the value field settings select count and show the values as % of row total.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Personal Loan** | **Column Labels** |  |  |
| **Row Labels** | **No** | **Yes** | **Grand Total** |
| No | 92.76% | 7.24% | 100.00% |
| Yes | 53.64% | 46.36% | 100.00% |
| **Grand Total** | **90.40%** | **9.60%** | **100.00%** |

**ONLINE Vs PERSONAL LOAN**

* Select data and insert pivot table.
* Drag the personal loan field into columns and values in the pivot table.
* And Online field into rows.
* In the value field settings select count and show the values as % of row total.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Personal Loan** | **Column Labels** |  |  |
| **Row Labels** | **No** | **Yes** | **Grand Total** |
| No | 90.63% | 9.38% | 100.00% |
| Yes | 90.25% | 9.75% | 100.00% |
| **Grand Total** | **90.40%** | **9.60%** | **100.00%** |

**INCOME CATEGORY Vs PERSONAL LOAN**

* Select data and insert pivot table.
* Drag the personal loan field into columns and values in the pivot table.
* And Income Category field into rows.
* In the value field settings select count and show the values as % of row total.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Personal Loan** | **Column Labels** |  |  |
| **Row Labels** | **No** | **Yes** | **Grand Total** |
| 0-50 | 100.00% | 0.00% | 100.00% |
| 100+ | 63.86% | 36.14% | 100.00% |
| 51-100 | 97.76% | 2.24% | 100.00% |
| **Grand Total** | **90.40%** | **9.60%** | **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?**

* The customers with high income are most likely to take Personal loan.
* The categorical variables such as **Income Category, TD Account and Education** appear most important to analyse the customers who take Personal loans.
* Customers who have TD account with the bank opted the personal loan.
* Majority of the customers who are undergraduate didn’t opt for 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 acceptance 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?**

Based on the analysis,

The way to optimise the cost of campaign is to the target the customers who have high **Income, TD Account**.

According to the analysis the customer who have Income category of 100+ have availed the personal loan i.e., if the customer has high income than the probability of taking personal loan is also high.

The customers who also have TD Account with the bank have opted the personal loan.

I suggest the management to approach the customers who have professional background with high experience, whose income is high and also the customers who have Term Deposit Account with the bank.