Business Report

Vinay Kumar Govardhanam

Table of Contents

**1. What percentage of the bank’s customers (according to the data) have availed Personal Loans? 2**

**2. Generate a table with min, max, median & average for all numeric variables (age, experience, income, family members, CreditCardAvg, Mortgage). What are your observations? 3**

**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 tasks. You can refer how Income Category has been created in the dataset] 4**

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

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

**6. How many customers have a combination of Fixed Deposits and Credit Cards but not Personal Loan? 5**

**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? 6**

**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] 7**

**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? 8**

**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? 9**

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

There are only 480 customer’s (10%) have availed Personal Loans

Our Business Objective is to improve the percentage of availed Personal Loan

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

There are 6 numerical features which are

* Age (in years)
* Experience (in years)
* Income (in K/year)
* Family Members
* CCAvg
* Mortgage

Although, we can convert the above Categorical features into normalised(Hot Encoding)

* Education (Graduate=1, Undergraduate=0, Professional=2)
* Personal Loan (Yes=1, No=0)
* Securities Account (Yes=1, No=0)
* TD Account (Yes=1, No=0)
* Online (Yes=1, No=0)
* CreditCard (Yes=1, No=0)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Features | Minimum | Maximum | Median | Average |
| Age (in years) | 23 | 67 | 45 | 45.3384 |
| Experience (in years) | 0 | 43 | 20 | 20.1348 |
| Income (in K/year) | 8 | 224 | 64 | 73.7742 |
| Family members | 1 | 4 | 2 | 2.39723003 |
| CCAvg | 0 | 10 | 1.5 | 1.937938 |
| Mortgage | 0 | 635 | 0 | 56.4988 |

**The findings for the above numerical variables**

# Age (in years)

- There are 45 distinct records (1%)

- the range is 23 – 67

# Experience (in years)

- There are 44 distinct records (1%)

- the range is 0 - 43

# Income (in K/year)

- There are 162 distinct records (3%)

- The range is 8 – 224

- There are some outliers which differs from the mean

- 180, 190, 192, 198, 200, 202, 204, 218, 224 are can be minimized by IQR

# Income Categorical

- There are 3 distinct records (0-50, 51-100, 100+)

- The category for 0 - 50 is 1914

- The category for 51-100 is 1874

- The category for 100+ is 1212

# Family members

- There are 4 distinct records (1, 2, 3, 4)

- There are 18 blank records found

- The mean differ from the median is the reason for the blank records are imputed by median

# CCAvg

- There are 108 distinct records (2%)

# Mortgage

- There are 347 distinct records (7%)

1. **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 tasks. You can refer how Income Category has been created in the dataset]**

* We can infer, there is high range of experienced customer’s in the category “21 – 30”
* The least range of experienced customer’s in the category “30+”
* The mean is 1250 for the four categories and it is differ from the highest of 51 and lowest of 93

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

* We can see a positive relationship in Age and Experience variable
* The Age is highly correlated to Experience variable
* If we increase the age there is highly chance of increase in experience
* We can infer, age is independent and Experience is dependent on age
* We cannot say, it’s a causation sometimes it will be odd or normal (it’s a regression problem)

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

* The topmost 3 areas in terms of zip code where the bank’s customers are located
* The zip code of 94720 having highest bank’s customers of 169
* The zip code of 94305 having second highest bank’s customers of 127
* The zip code of 95616 having third highest bank’s customers of 116
* We can improve in marketing campaign of personal loan offers in the top 3 areas
* We can say, highest customer count in 94720 having personal loan is 11%

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ZipCode** | **No** | | **Yes** | | **Counts** |
| 94720 | 150 | 89% | 11% | 19 | 169 |
| 94305 | 114 | 90% | 10% | 13 | 127 |
| 95616 | 110 | 95% | 5% | 6 | 116 |

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

There are 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?**

* The customers who have availed personal loans having 142.5 median income
* The customer who not have availed personal loans having 59 median income
* The customer having an greater than or equal to the median income of 142.5 can have the higher chance of availing the personal loan
* The customer having an lesser than or equal to the median income of 59 cannot avail the personal loan
* From the figure, we conclude that the customer income having outliers that can differ from having personal loan is “No”



**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** | | |
| **Count of Education Vs Personal Loan** | **Personal Loan** |  |
| **Education** | **No** | **Yes** |
| Graduate | 87% | 13% |
| Professional | 86% | 14% |
| Undergraduate | 96% | 4% |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **TD Account Vs Personal Loan** | | |
| **Count of TD Account Vs Personal Loan** | **Personal Loan** |  |
| **TD Account** | **No** | **Yes** |
| No | 93% | 7% |
| Yes | 54% | 46% |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **Online Vs Personal Loan** | | |
| **Count of Personal Loan** | **Personal Loan** |  |
| **Online** | **No** | **Yes** |
| No | 91% | 9% |
| Yes | 90% | 10% |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **Income Category Vs Personal Loan** | | |
| **Count of Personal Loan** | **Personal Loan** |  |
| **Income Categorical** | **No** | **Yes** |
| 0-50 | 100.00% | 0.00% |
| 100+ | 63.86% | 36.14% |
| 51-100 | 97.76% | 2.24% |

**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?**

**Education Vs Personal Loan**

* + - **We can predict, Professional customers are the highly to reject personal loan**
    - **By the data, Graduate customers are more likely to reject personal loan**
    - **We can say, Undergraduate customers are not likely to have personal loan**
    - **But, there is slight chance of 13-14% of avail personal loans in the field of Graduate or Professional**
    - **Comparatively, we can defer the undergraduate are poor in avail the personal loans & there is slight chance of 4% of avail personal loan. To address this issue, we have to make aware of offers and benefits in marketing campaign**

**TD Account Vs Personal Loan**

* + - **By the data, customers are not having TD account will more likely to say no to personal loans**
    - **But, there is a slight chance of 7% of avail personal loans by the non TD Account**
    - **We can say, customers are having TD account will more likely to say no to personal loans**
    - **But, there is a slight chance of 46% of avail personal loans by the TD Account**

**Online Vs Personal Loan**

* + - **We can say, having Online or Offline will more likely to say no to personal loans**
    - **But, there is a slight chance of 10% of avail personal loans by the online customers**

**Income Category Vs Personal Loan**

* + - **We can say, Income Category of 0 - 50 are more likely to fail in availing personal loan**
    - **We can say, Income Category of 51 - 100 are also fail to avail personal loan**
    - **We can predict, Income Category of 100+ is slight chance of avail personal loan**

In the above analysis, most of the time customers are rejecting the personal loans

But, there is slight possible outcome of avail personal loan

* + - Graduate & Professional are having slight chance of 13-14%
    - TD account customers are probability of 46%
    - Online customers are having slight chance of 10%
    - Customer income of 100+ are having slight chance of 36.14%

We can say, Whether the customer is Graduate or Professional and having TD account and having online and earning income of 100+ is likely to avail personal loan (105%)

**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?**

**From the analysis, we can optimise the cost of this campaign by identifying the correct target base (without significant reduction in number of acceptance of offers)**

**From the dataset, in the family member feature there are 18 missing values which are filled by the median value 2**

**The major challenges/hurdlers are**

* + - **How can we improve the avail personal loan**
    - **How can we know the customer from defaulting**
    - **How can we know whether the customer will opt for another loan when the previous loan is going to be completed**
    - **Whether a customer is eligible for personal loan in spite of criteria check**