Business Report - 2023

SEPTEMBER 13

HBFC BANK
Authored by: EKTA SINGH

Personal loan data analysis – (EDA)

"Finding out potential target customers for personal loans"

HBFC Bank has a growing customer base.
Majority of these customers having deposits and Term Deposit in the bank.
The number of customers having loan accounts is quite small, and the 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.
The management wants to explore ways of converting its customers having deposits to personal loan customers.

At present the bank has small number of asset/loan customers. The bank wants to increase their income by increasing the customer base of "asset customers".
Last year the bank ran a campaign where they successfully converted 9% of the existing "liability customers" to "asset customers".

This has encouraged the bank to have a better targeted marketing campaign to increase the success ratio with minimum budget.

APPROACH USED -

-> Performed preliminary data analysis (EDA) and visualization tools (scatter chart, bar chart, etc) to understand the profile of customers having savings account, who took personal loan in the last marketing campaign VS customers who didn't take it up.
-> Compared the effect of various significant attributes with personal loan using pivot tables.
-> Using EDA identify profile of customers whom bank can target for selling personal loans.

You have been provided with a dataset of 5000 customers. The data include customer demographic information (age, income, etc.), the customer's relationship with the bank (mortgage, securities account, etc.), and the customer response to the last personal loan campaign (Personal Loan).

Data Description:

ID	Customer ID
Age	Customer's age in 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?
TD Account	Does the customer have a Term deposit (Including Fixed and Recurring Deposits) account with the bank?
Online	Does the customer use internet banking facilities?
CreditCard	Does the customer use a credit card issued by the bank?

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

Only 9.60% of the total bank's customer have availed personal loan according to the data of the last year's marketing campaign.

Personal loan	% of Personal Loan
No	90.40%
Yes	9.60%
Grand Total	100.00%

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

	Age (in years)	Experience (in years)	Income (in K/year)	Family members	CCAvg	Mortgage
AVERAGE	45.3384	20.1348	73.7742	2.397230028	1.937938	56.4988
MEDIAN	45	20	64	2	1.5	0
MAX	67	43	224	4	10	635
MIN	23	0	8	1	0	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**

-> We can create categorical variable for Experience column by using nested IF conditions.

-> We can see from the Bar plot below that the most number of people are having 21 to 30 years of experience.

Experience categorical
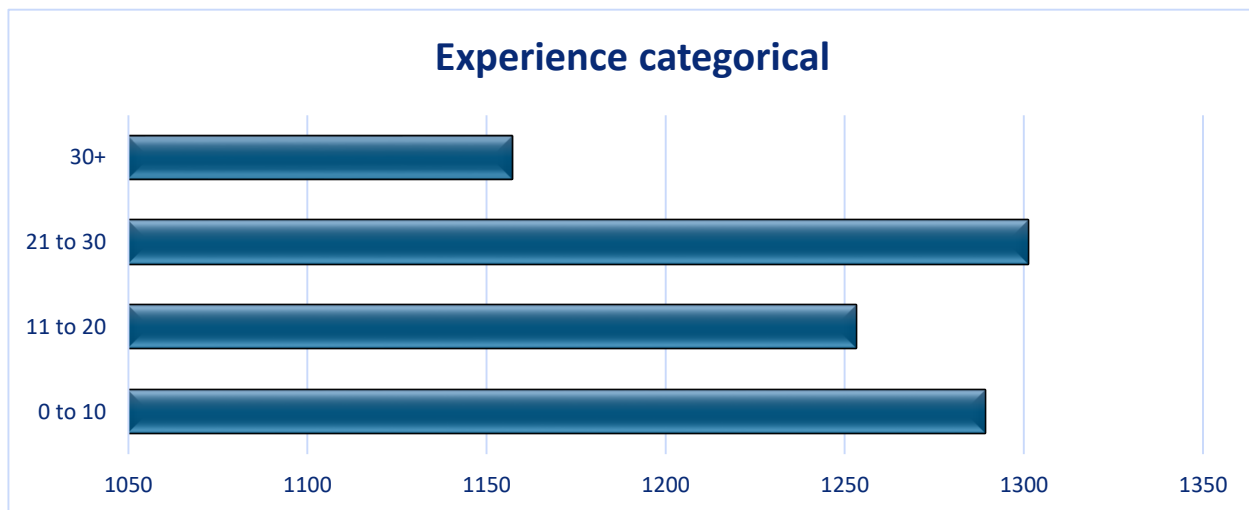
Experience	Count of Experience Categorical
0 to 10	1289
11 to 20	1253
21 to 30	1301
30+	1157
Grand Total	**5000**

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

The scatter plot for Age and Experience variable can be created as followed-

Age Vs Experience(in years)
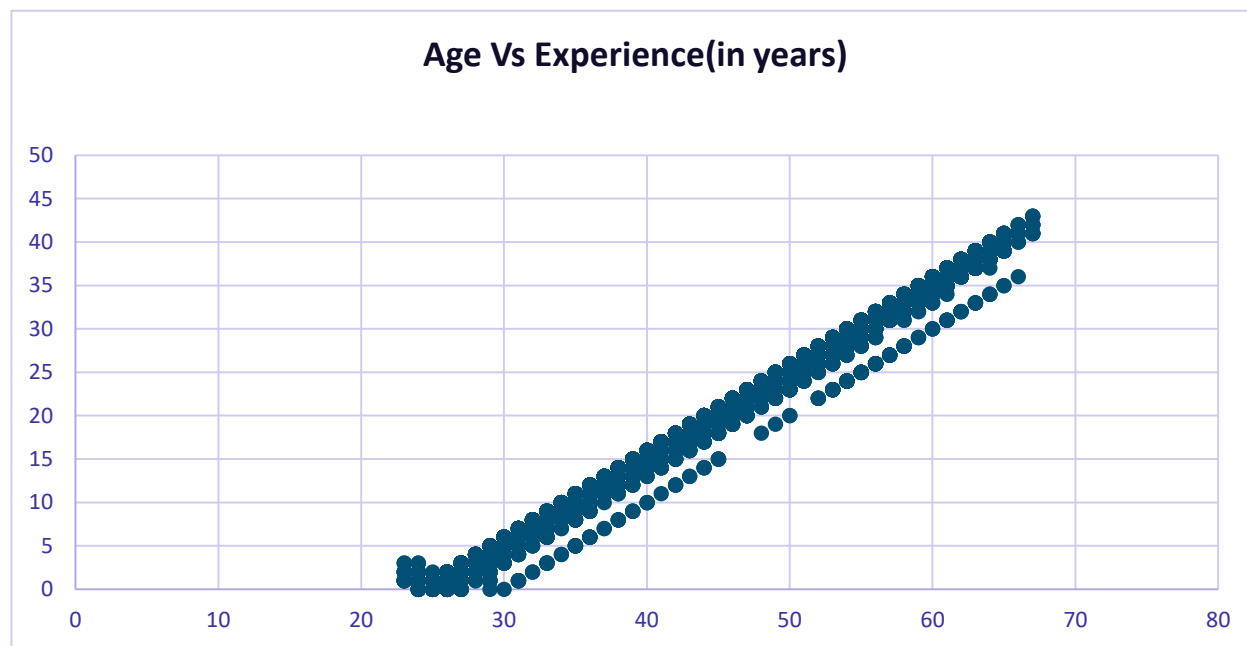
Observations-

-> There is a significant correlation between age and experience, as the age increases experience also increases.

-> There are not any outliers (data points that are far away from the main cluster) present in the dataset.

-> Density of the data points is uniform throughout the plot.

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

The top 3 areas (ZIP Codes) where the bank's customers are located are as followed-

ZIP Code	Count of ZIP Code
94720	169
94305	127
95616	116

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

The total count of the people who have combination of fixed deposit and credit card but not Personal loan= 147

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?

Median income of the people who have availed personal loan = 142.5
Median income of the people who have not availed personal loan = 59

-> Median gives us an idea of the "central" value within our dataset and it does not get affected by the outliers present in the data.
-> It does not convey any information about the spread of the data.
-> Here, we can say that if the median income of the people availed loan is 142.5 then roughly half of the income points lies below and above this median value.

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**

Education Vs Personal Loan	Personal loan	
Education loan	No	Yes
Graduate	87.03%	12.97%
Professional	86.34%	13.66%
Undergraduate	95.56%	4.44%

TD a/c Vs Personal Loan	Personal loan	
TD a/c	No	Yes
No	92.76%	7.24%
Yes	53.64%	46.36%

Online Vs Personal Loan	Personal loan	
Online	No	Yes
No	90.63%	9.38%
Yes	90.25%	9.75%

Income cat. Vs Personal Loan	Personal Loan	
Income cat.	No	Yes
0-50	100.00%	0.00%
100+	63.86%	36.14%
51-100	97.76%	2.24%

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?

If we were to focus on customer behavior on personal loans further TD account and Income level are the categorical variables that should be considered for study.

They offer the most significant variance and insights into the customer behavior concerning personal loans.

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?

-> If someone already has an account in our bank then it is more likely that they will be interested in taking personal loan from our bank. So, focus on these customers.

-> People who took the personal loan last year, it is more likely for them to take it this year also.

-> People who have finished college or a working professional are more likely to take a loan than people who are still in the college.

-> People under the higher income group are more likely to take a loan and pay it on time than people of the lower income group.

-> Instead of spending a lot of money on advertisement, concentrate more on the people who have higher chances of availing a loan from our bank.

-> Everyone will not buy our product (personal loan) so focus on the target customers.

END OF THE REPORT