

Situation:

The bank (HBFC Bank) has hired a consultant to help them identify potential customers who have a higher probability of purchasing a personal loan. The bank has a growing customer base, but the number of customers who are also borrowers is quite small. The bank wants to explore ways of converting its customers with deposits to personal loan customers while retaining them as depositors.

Task:

explore the dataset of 5000 customers provided, which includes customer demographic information, the customer's relationship with the bank, and the customer response to the last personal loan campaign. We need to analyze the data to understand the variables and the impact they have had on Personal Loans so that the bank can leverage the insights to reach out to the right customers who have a higher probability of purchasing the loan. we need to provide recommendations to the bank based on our analysis.

Action:

Data: ID, Age (in years), Experience (in years), Income (in K/year), Income Categorical, ZIP Code, Family members, CCAvg, Education, Mortgage, Personal Loan, Securities Account, TD Account, Online, CreditCard

Descriptive Analysis:

1. What percentage of the bank's customers have availed Personal Loans?

To determine the percentage of the bank's customers who have availed personal loans, we need to use a pivot table. In the pivot table, we will set the column as "personal loan" and the value as the "count of ID" to obtain the percentage of customers who have availed personal loans.

	A	B	C	D
1				
2				
3	Column Labels ▾			
4	No	Yes	Grand Total	
5	Count of ID	90.40%	9.60%	100.00%
6				
7	Availed Personal Loan: 9.60%			

the last campaign, we observed that only 9.60% of customers had purchased a personal loan, leaving 90.40% of customers who have not yet bought the personal loan. To increase the personal loan business, we need to analyze the characteristics of the 9.60% who purchased personal loans and explore ways to identify potential customers who exhibit similar characteristics.

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

Here we have created a table and utilized the functions =MIN(), =MAX(), =MEDIAN(), and =AVERAGE() to obtain the values of age, experience, income, family members, CC avg, and mortgage.

	A	B	C	D	E
1					
2	Column1	Min	Max	Median	Average
3	Age	23	67	45	45.3384
4	Experience	0	43	20	20.1348
5	Income	8	224	64	73.7742
6	Family Members	1	4	2	2.36612
7	CCAvg	0	10	1.5	1.937938
8	Mortgage	0	635	0	56.4988

- The customers in the dataset have an age range of 23 to 67 years with a median age of 45 years.
- The work experience of the customers ranges from 0 to 43 years with a median of 20 years.
- The income of the customers ranges from 8 to 224 (in thousands of dollars) with a median income of 64 thousand dollars.
- Most customers have 2 family members with a range of 1 to 4, and a median of 2 family members.
- The average credit card spending of the customers is 1.93 thousand dollars per month with a range of 0 to 10 thousand dollars, and a median of 1.5 thousand dollars.
- The customers have a mortgage range of 0 to 635 thousand dollars, with most having no mortgage (median of 0 thousand dollars).

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.

Here, we first we open a new sheet and used an IF function to create categories of experience and form it into a table

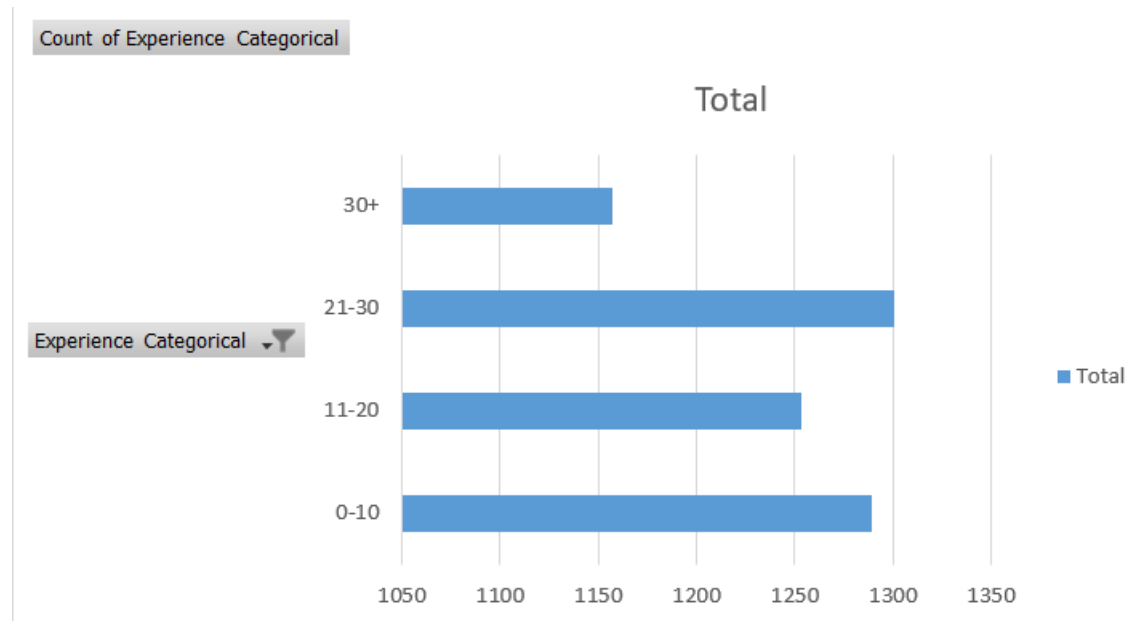
```
=IF(Bank_Personal_Loan_Modelling!C2<=10,"0-10",IF(AND(Bank_Personal_Loan_Modelling!C2>10,Bank_Personal_Loan_Modelling!C2<=20),"11-20",IF(AND(Bank_Personal_Loan_Modelling!C2>20,Bank_Personal_Loan_Modelling!C2<=30),"21-30","30+"))).
```

After creating, we used a pivot table in the same existing sheet where we set the rows to the experience category, which we had earlier created, and the values to the count of experience from actual dataset.

Row Labels	Count of Experience Categorical
0-10	1289
11-20	1253
21-30	1301
30+	1157
Grand Total	5000

Plot a bar graph for this new categorical variable?

We use above pivot table to create the below bar graph

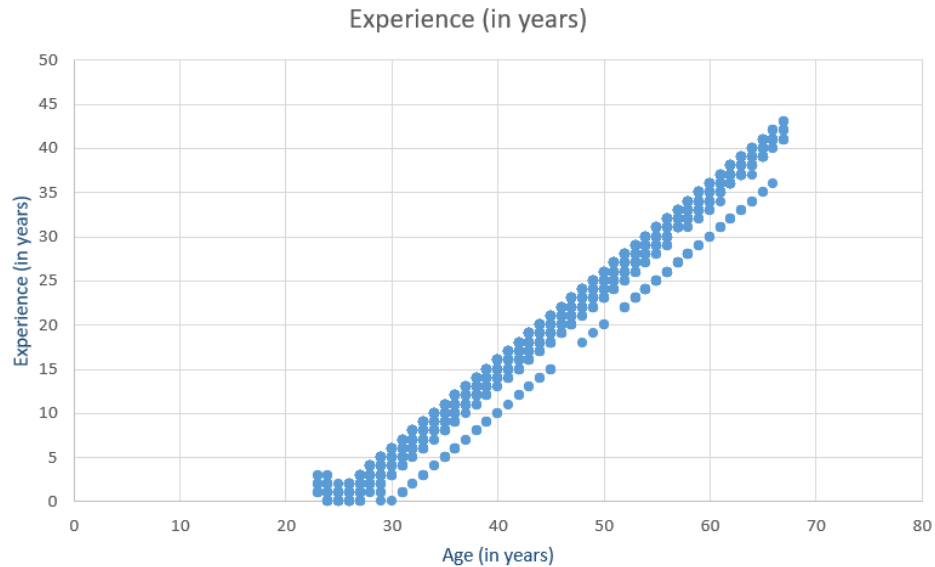


So based on this table and bar graph, it appears that the majority of customers fall into the 0-30 years of experience range. The highest count is in the 21-30 range with 1301 customers, followed by the 0-10 range with 1289 customers and the 11-20 range with 1253 customers.

The lowest count is in the 30+ range with 1157 customers.

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

From below scatter plot we can observe the scatter plot of age and experience, it becomes evident that with an increase in age, there is also a corresponding increase in experience, indicating a positive correlation between these two variables.



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

	A	B
1	ZIP Code	Count of ZIP Code
2	94720	169
3	94305	127
4	95616	116
5	Grand Total	412

To create a table similar to the one above, we need to follow the process below:

- First, we need to create a pivot table on a new sheet using the actual bank dataset.
- Second, we need to set ZIP code in the row and count of ZIP code in the values.
- This will generate the data, but we need to filter it out to show only the highest values.
- To do this, we can use the value filter option and select "Top 10," inputting the number 3.
- Now that we have filtered the data, we need to sort it in descending order by count of ZIP code.
- To do this, we can use the "More sort options" and select "Descending (Z to A) by: Count of ZIP Code.

The data shows that the top three ZIP codes with the highest count of customers are 94720 with 169 customers, 94305 with 127 customers, and 95616 with 116 customers. The grand total count of customers across all ZIP codes is 412.

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

Out of 5000 customers, only 147 have a combination of Fixed Deposits and Credit Cards but do not have Personal Loan. Here, we are also using a Pivot table. We will filter by credit card, TD account, and personal loan. In the values section, we will use the count of ID.

	A	B	C
1	CreditCard	Yes	
2	TD Account	Yes	
3	Personal Loan	No	
4			
5	Count of ID		
6		147	

Based on this observation, the bank could target these 147 customers with personalized loan offers based on their financial history and creditworthiness. Since they already have a relationship with the bank, it might be easier to convince them to take a 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?

Here first we have created two tables by filtering the personal loan by “Yes” and “No”

	A	B	C	D	E	F
1	Income (in K/year)	Personal Loan		Income (in K/year)	Personal Loan	
2	170	Yes		112	No	
3	134	Yes		32	No	
4	133	Yes		20	No	
5	185	Yes		105	No	
6	105	Yes		114	No	
7	195	Yes		21	No	
8	95	Yes		78	No	
9	184	Yes		79	No	
10	154	Yes		75	No	
11	160	Yes		41	No	
12	89	Yes		18	No	
13	94	Yes		51	No	
14	170	Yes		15	No	
15	143	Yes		42	No	
16	172	Yes		18	No	
17	182	Yes		35	No	
18	122	Yes		18	No	
19	138	Yes		53	No	
20	168	Yes		55	No	
21	179	Yes		144	No	
22	143	Yes		11	No	
23	181	Yes		105	No	
24	101	Yes		70	No	
25	174	Yes		39	No	

Next we have calculate the median by the following function:

=MEDIAN(A2:A481) For First Table

=MEDIAN(D2:D4521) For Second Table

HBFC BANK BUSINESS REPORT

Below are the Values we get from these functions

The Median income of availed perrsonal loans				
142.5				
The Median income of not availed perrsonal loans				
59				

From this observation we can know that there is significant difference between the median income of customers who have availed personal loans and those who have not. Customers who have availed personal loans have a median income of 142.5, which is more than twice the median income of those who have not availed personal loans (59).

This suggests that income is an important factor in determining whether a customer is likely to avail a personal loan. Customers with higher income levels are more likely to avail personal loans compared to customers with lower income levels.

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

Education vs Personal Loan

A				
Count of ID	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%	

The highest loan takers are professionals, with 13.66% of them availing personal loans, followed by graduates with 12.97%. On the other hand, undergraduate customers have the lowest personal loan availing rate, with only 4.44% out of 100%.

Overall, the data suggests that customers with higher education levels (Graduate and Professional) are slightly more likely to avail personal loans compared to Undergraduate customers.

TD Account Vs Personal Loan

B				
Count of ID	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%	

The table indicates TD account holders' response to the last personal loan campaign. Among customers with no TD accounts, 92.76% did not take personal loans, while 7.24% did. Among TD account holders, 53.64% did not take personal loans, while 46.36% did. This suggests that TD account holders are more likely to take personal loans, making them a more effective target for personal loan campaigns.

Online vs Personal Loan

C			
Count of ID	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%

The table shows the percentage of customers who used net banking and their response to personal loan campaigns. Overall, there is little difference between the groups, with 9.38% of non-net banking users availing loans compared to 9.75% of net banking users.

The difference between the two groups is minimal, suggesting that using net banking services may not have a significant impact on the likelihood of availing personal loans.

Income Category vs Personal Loan

D			
Count of ID	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%

Out of the total 5000 customers, all customers with an income between 0-50 did not avail personal loans. For customers with an income of 51-100, only 2.24% availed personal loans while 97.76% did not.

In contrast, for customers with an income of 100 or more, 36.14% availed personal loans while 63.86% did not. This suggests that higher-income customers are more likely to avail personal loans than those with lower incomes.

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?

- Customers with higher education levels (Graduate and Professional) are slightly more likely to avail personal loans compared to Undergraduate customers.
- TD account holders are more likely to take personal loans, making them a more effective target for personal loan campaigns.

- Using net banking services may not have a significant impact on the likelihood of availing personal loans.
- Higher-income customers are more likely to avail personal loans than those with lower incomes.

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?

A. Target customers with higher incomes: Customers with higher incomes are more likely to avail personal loans than those with lower incomes. Therefore, the bank should focus on identifying customers with higher incomes and targeting them with personalized loan offers based on their financial history and creditworthiness.

B. Focus on TD account holders: The data shows that TD account holders are more likely to take personal loans. Therefore, the bank should target its TD account holders with personalized loan offers.

C. Target customers with higher education levels: Customers with higher education levels, particularly graduates and professionals, are slightly more likely to avail personal loans compared to undergraduate customers. Therefore, the bank should target its customers with higher education levels with personalized loan offers.

D. Target customers who have fixed deposits and credit cards: The data shows that out of 5000 customers, only 147 have a combination of fixed deposits and credit cards but do not have personal loans. These customers already have a relationship with the bank and might be easier to convince to take a personal loan. Therefore, the bank should target these 147 customers with personalized loan offers based on their financial history and creditworthiness.

E. Improve online loan application process: The data shows that customers who use net banking are slightly more likely to avail personal loans than those who do not. Therefore, the bank should consider improving its online loan application process to make it more user-friendly and accessible to customers.

F. Increase marketing efforts in areas with high customer density: The data shows that certain ZIP codes have a higher concentration of bank customers. Therefore, the bank should focus its marketing efforts on these areas to increase personal loan purchases.