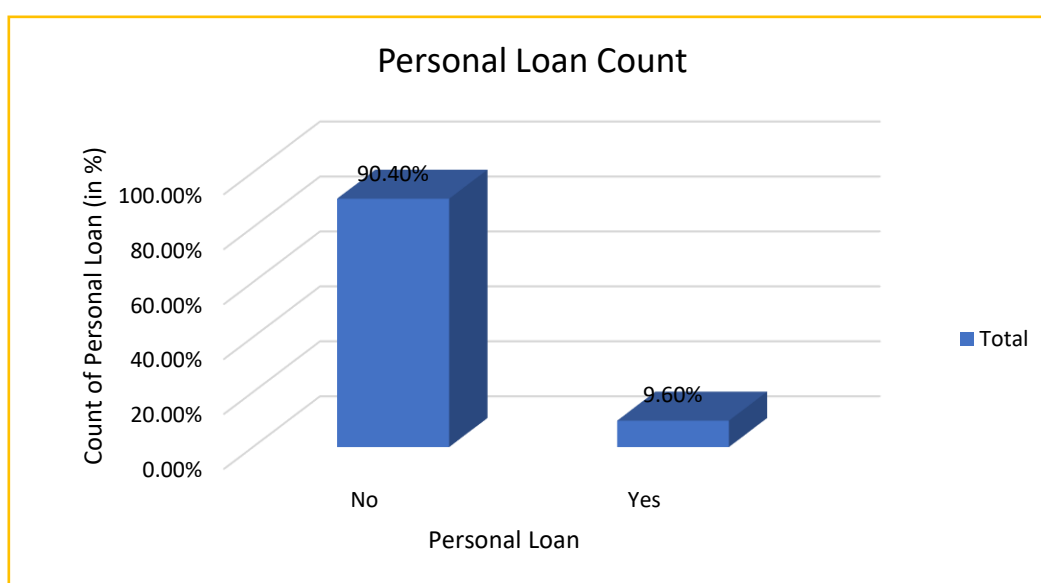# HBFC Bank Loan Data Analysis

Personal loan data analysis – Exploratory data analysis

QUESTION 1

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

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

Count of customers
4520
480
5000

Personal Loan Count

According to the data 9.6% (480 customers) of the bank's customers have availed Personal loans. 90.4% (4520 customers) of the bank's customers have not availed Personal Loans.
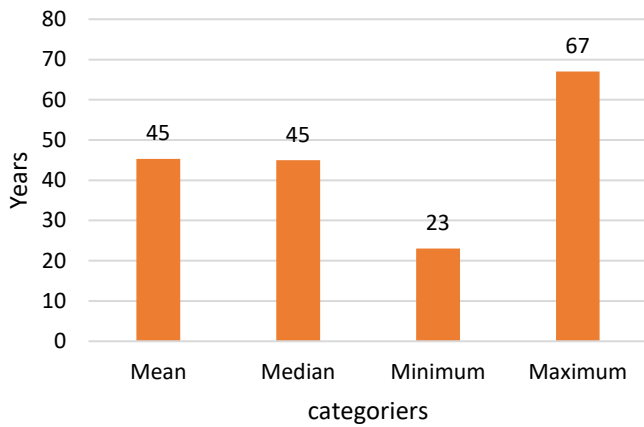
QUESTION 2

Generate a table with min, max, median & average for all numeric variables (age, experience, income, family members, CCAvg, Mortgage).
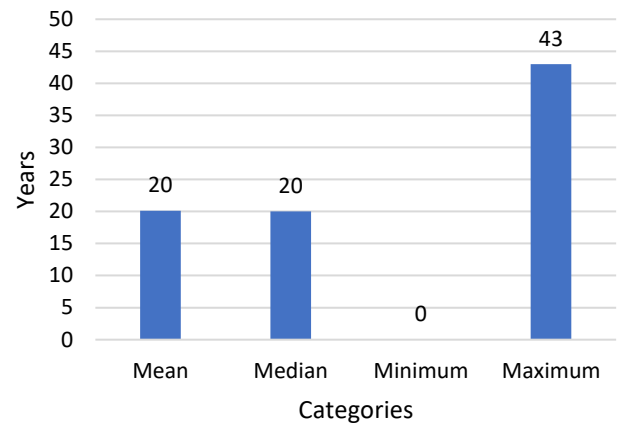
	Age (in years)	Experience (in years)	Income (in K/year)	Family members	CCAvg (in $000/month)	Mortgage(in $000)
Mean	45	20	73.8	2	1.9	56.5
Median	45	20	64	2	1.5	0
Minimum	23	0	8	1	0	0
Maximum	67	43	224	4	10	635

The average, Median , Min and Max Value for Age(in years) , experience (in Years), income (in k/year), Family Members , CCAvg (in $000/month) and Mortgage (in $000) are plotted in the above table. The bar graph representations of the same are as follows.
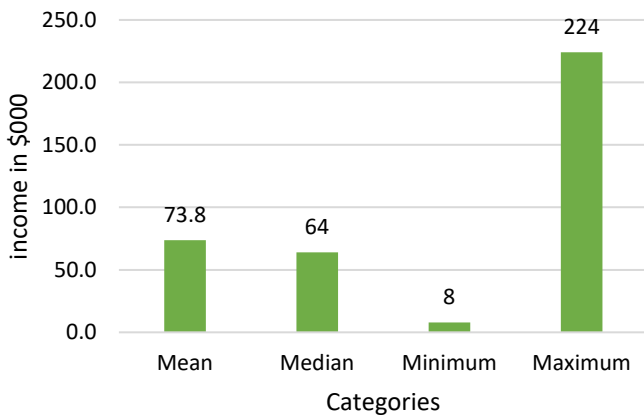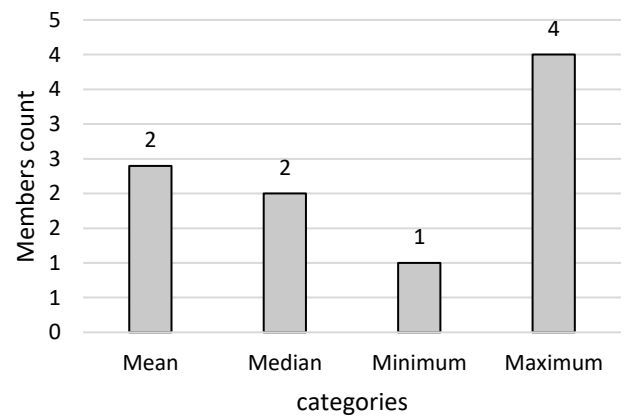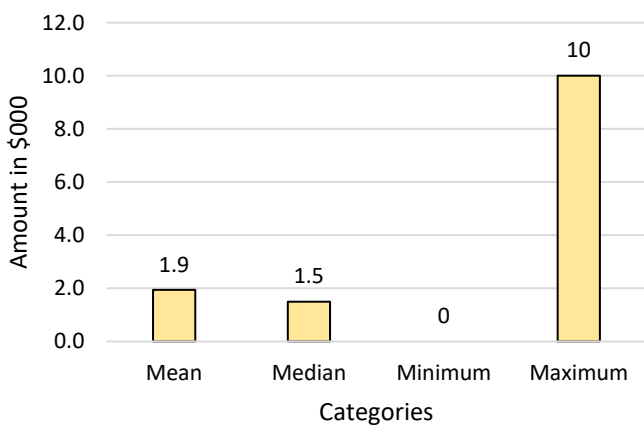
Age (in years)

Category	Years
Mean	45
Median	45
Minimum	23
Maximum	67

Experience (in years)

Category	Years
Mean	20
Median	20
Minimum	0
Maximum	43

Income (in K/year)

Category	income in $000
Mean	73.8
Median	64
Minimum	8
Maximum	224

Family members

Category	Members count
Mean	2
Median	2
Minimum	1
Maximum	4

CCAvg (in $000/month)

Category	Amount in $000
Mean	1.9
Median	1.5
Minimum	0
Maximum	10

Mortgage(in $000)

| Category	Amount in $000
Mean	56.5
Median	0
Minimum	0
Maximum	635

QUESTION 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

IF and AND functions can be used to create a new Column as Experience Categorical.

Experience Categorical	Count of Experience (in years)	Percentage of Customers (out of 5000)
0 to 10 Years	1289	25.78%
11 to 20 Years	1253	25.06%
21 to 30 Years	1301	26.02%
30+ Years	1157	23.14%
		100.00%
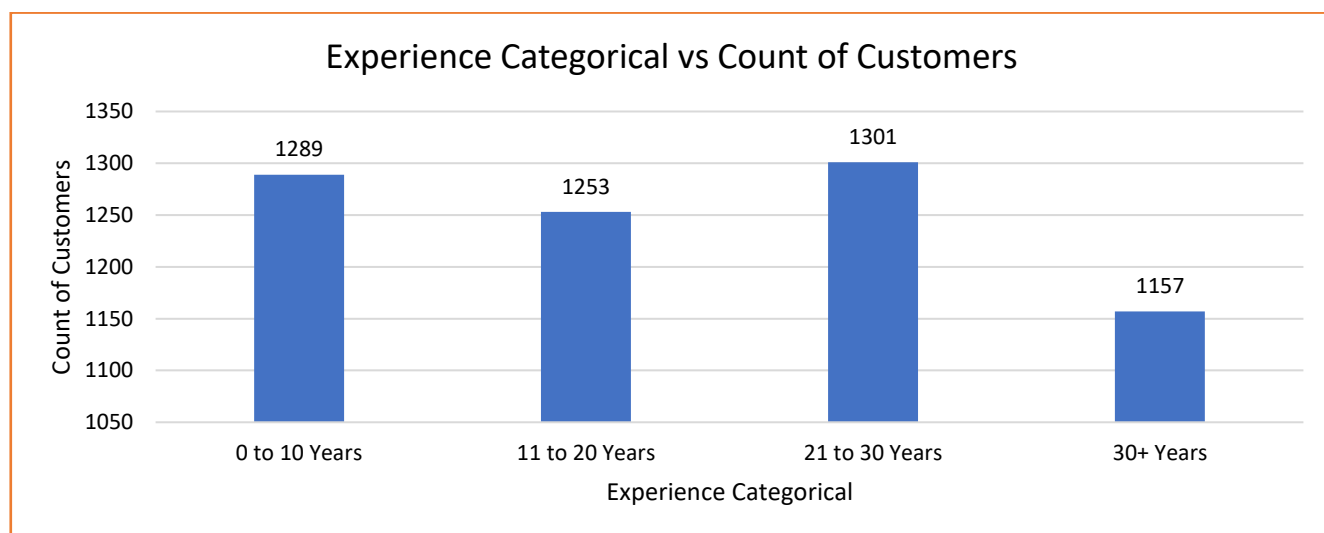
We can infer that 1289 customers fall under the category 0 to 10 years that is 25.78% of the customers,1253 customers fall under the category 11 to 20 years that is 25.06% of the customers,1301 customers fall under the category 21 to 30 years that is 26.02% of the customers and 1157 customers falls under the category 30+ years that is 23.14% of the customers.

QUESTION 4

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

Age vs Experience (in years)
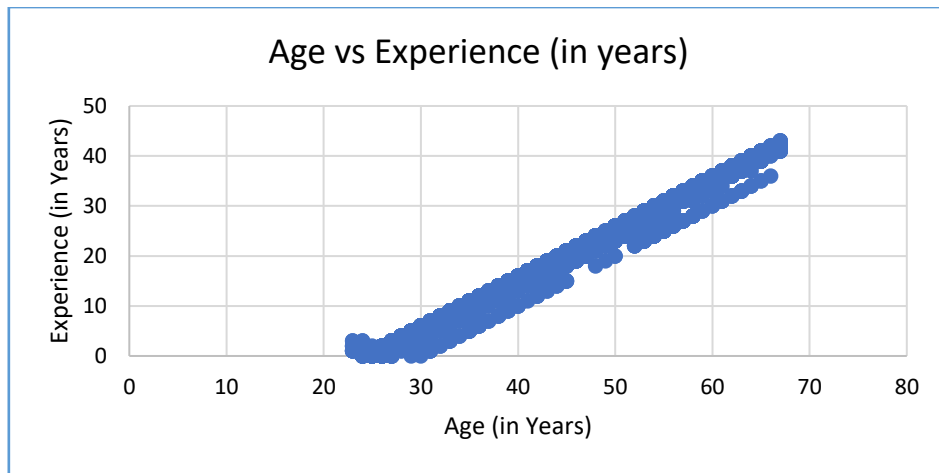
The Scatter plot of the Age and Experience Variable shows a strong positive correlation of value 0.994. This shows that with increase in age, Experience also increases. Similarly, when age is less Years of Experience is also less. In other words, we can say Experience is directly proportional to age.

QUESTION 5

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

Zip Code	Count of Zip Code	Percentage %
94305	116	2.32%
95616	127	2.54%
94720	169	3.38%

The top 3 areas (ZIP Codes) where the bank's customers are located in ascending order are 95616 with a customer count of 116 that is 2.32%, 94305 with a customer count of 127 that is 2.54% and 94720 with a customer count of 169 that is 3.38%.

QUESTION 6

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

Personal Loan	No	
Count of TD Account	**Credit Card**	
TD Account	**Yes**	**Grand Total**
Yes	147	147
Grand Total	**147**	**147**

| Count of customers who have a combination of Fixed Deposits and Credit Cards but not Personal Loan	147

Customer Percentage:	2.94%

From the data we can infer that 147 customers (i.e., 2.94%) have a combination of Fixed deposits and Credit Cards but does not have Personal Loan.

QUESTION 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?

Column	Median (in K/year)
Median Income of Customers availed Personal Loans	142.5
Median Income of Customers haven't availed Personal Loans	59

Income of Customers availed Personal Loans

Income of customers who haven't availed Personal Loans
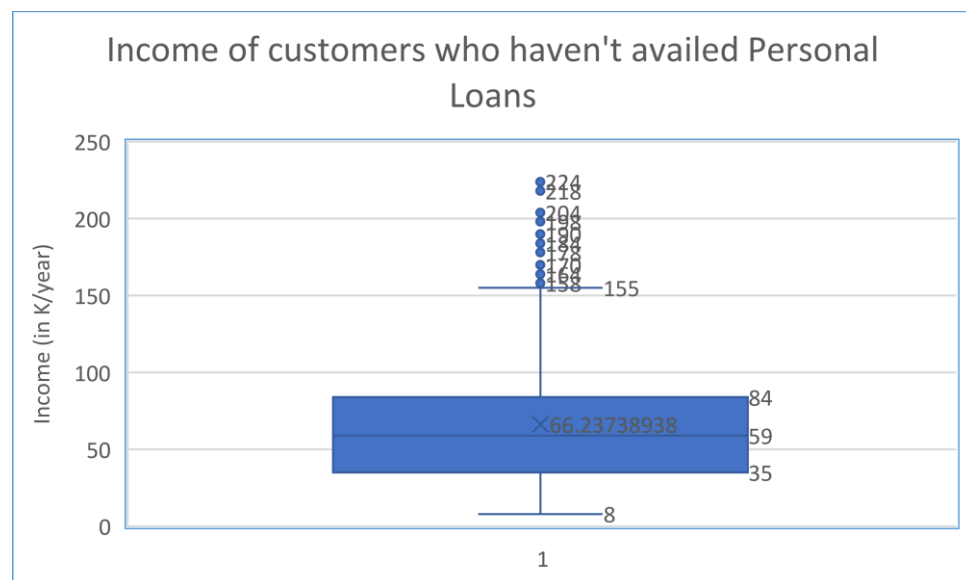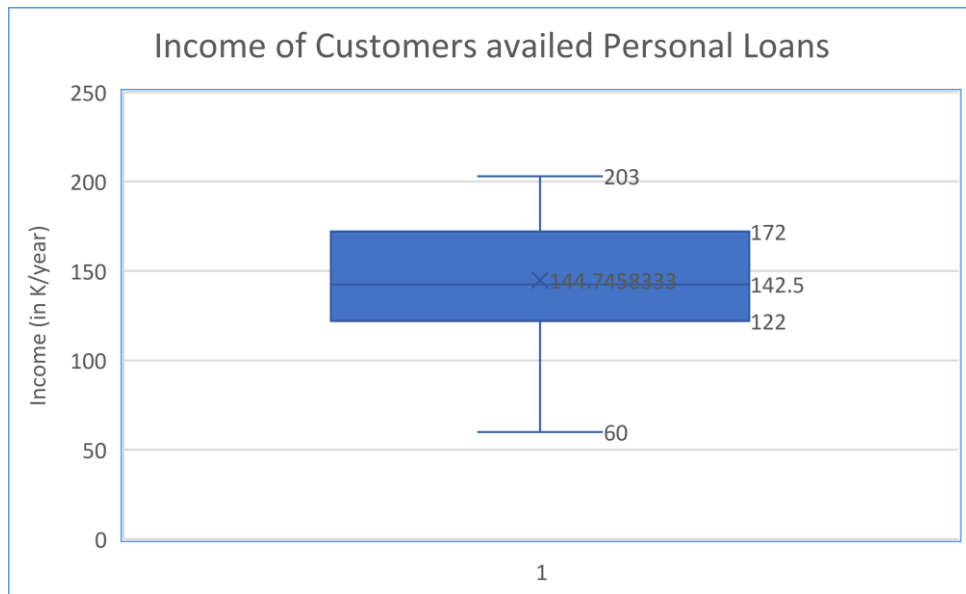
     The median income of people who have availed personal loans is $142500 per year which is very higher than the income of the people who have not availed personal loans, that is $59000. We can infer that people with high income per annum has availed personal loans than people with low income per annum.

     From the box plots we can say that that minimum income among the customers who have availed Personal Loan, that is $60000 is greater than the median income of Customers who have not availed personal loans, that is $59000.
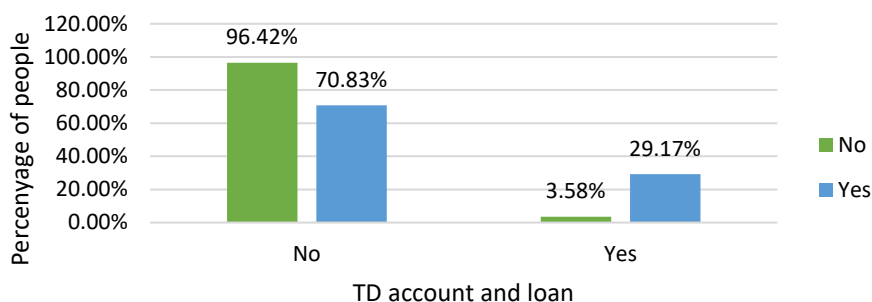
QUESTION 8

Create 4 separate Pivot Tables. Summarize your data by percentage values.
• Education vs Personal Loan
• TD Account Vs Personal Loan
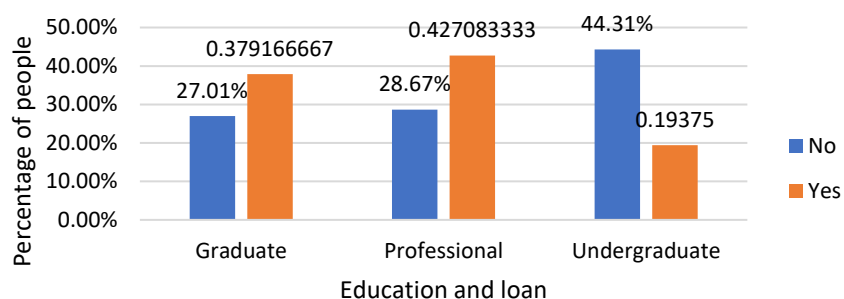• Online vs Personal Loan
• Income_Category vs Personal Loan

Count of Personal Loan	Personal loan		
TD account	No	Yes	Grand Total
No	96.42%	70.83%	93.96%
Yes	3.58%	29.17%	6.04%
Grand Total	100.00%	100.00%	100.00%

Count of Personal Loan	Personal Loan		
Education	No	Yes	Grand Total
Graduate	27.01%	37.917%	28.06%
Professional	28.67%	42.71%	30.02%
Undergraduate	44.31%	19.38%	41.92%
Grand Total	100.00%	100.00%	100.00%

Count of Personal Loan	Personal Loan		
Online	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%

Online vs Personal Loan

Count of Personal Loan	Personal Loan		
Income Category	No	Yes	Grand Total
0-50	42.35%	0.00%	38.28%
100+	17.12%	91.25%	24.24%
51-100	40.53%	8.75%	37.48%
Grand Total	100.00%	100.00%	100.00%
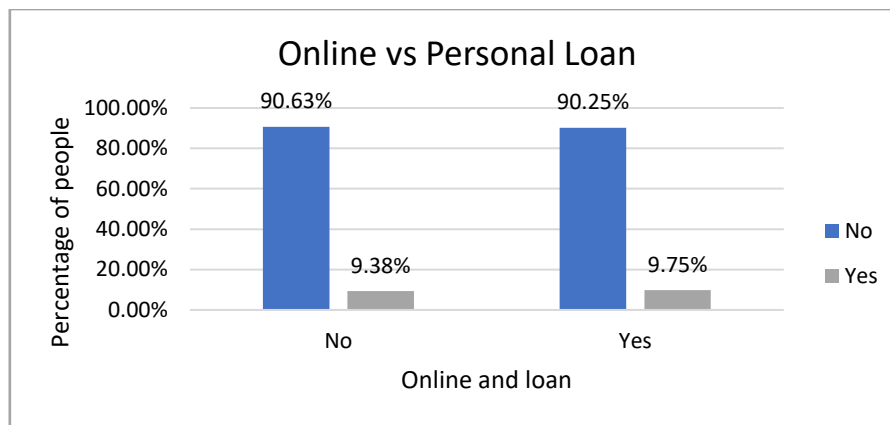
Income category vs Loan

QUESTION 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?

Count of Personal Loan	Personal Loan		
Education	No	Yes	Grand Total
Graduate	27.01%	37.92%	28.06%
Professional	28.67%	42.71%	30.02%
Undergraduate	44.31%	19.38%	41.92%
Grand Total	100.00%	100.00%	100.00%

On analyzing the pivot table, we can see that the people who have availed personal loans are mostly Graduate (37.92%) and Professional (42.71%) category. Very small percentage of customers who are graduates (19.38%) availed personal loans. Thus, we can target more Graduates and Professional customers to awail personal loan.

Count of Personal Loan	Personal loan		
TD account	No	Yes	Grand Total
No	96.42%	70.83%	93.96%
Yes	3.58%	29.17%	6.04%
Grand Total	100.00%	100.00%	100.00%

The data shows that customers who have TD account have availed Personal Loans (29.17%) than customers who don't have TD account. 96.42% of customers who don't have TD account have not availed Personal account. Thus, we can target customers with TD account.

Count of Personal Loan	Personal Loan		
Online	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 data shows that around 90-91% of both the customers who use and don't use Online transaction facilities have not availed Personal loans. This shows that online banking facility doesn't impact much on the personal loan. This is an anomaly. This data can't be used to predict whether to target the customer or not.

Count of Personal Loan	Personal Loan		
Income Category	No	Yes	Grand Total
0-50	42.35%	0.00%	38.28%
100+	17.12%	91.25%	24.24%
51-100	40.53%	8.75%	37.48%
Grand Total	100.00%	100.00%	100.00%

No customers with income category 0-50 have availed personal loans. A very small percentage of customers that is 8.75% have availed loan in the category 51-100. Most of the customers who have availed Personal loans, that is 91.25% who have availed personal loans falls under the category 100+ . So, we can target customers of category 100+.

The following categorical variables don't show much significant relation to target Customers.

Count of Personal Loan	Column Labels		
Credit card	No	Yes	Grand Total
No	70.64%	70.21%	70.60%

Yes	29.36%	29.79%	29.40%
Grand Total	**100.00%**	**100.00%**	**100.00%**

Count of Personal Loan	Column Labels		
Experience categorical	**No**	**Yes**	**Grand Total**
0 to 10 Years	25.58%	27.71%	25.78%
11 to 20 Years	25.11%	24.58%	25.06%
21 to 30 Years	26.15%	24.79%	26.02%
30+ Years	23.16%	22.92%	23.14%
Grand Total	**100.00%**	**100.00%**	**100.00%**

Count of Personal Loan	Column Labels		
credit card	**No**	**Yes**	**Grand Total**
No	70.64%	70.21%	70.60%
Yes	29.36%	29.79%	29.40%
Grand Total	**100.00%**	**100.00%**	**100.00%**

Count of Personal Loan	Personal Loan		
Securities Account	**No**	**Yes**	**Grand Total**
No	89.78%	87.50%	89.56%
Yes	10.22%	12.50%	10.44%
Grand Total	**100.00%**	**100.00%**	**100.00%**

Categorical Variables such as **Education, TD account, Income Category** appear most important if you want to analyze which customers are most likely to take personal loans because the relationship between these categorical variables and Personal loan variable is significant compared to others and can be used to target potential customers who most likely avail Personal Loan.

QUESTION 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 following **106 customers** can be targeted and will most likely avail Personal loan. These customers have the educational category of **Graduate and Professional** and falls under the **income category 100+** (i.e., $100000 per year). These customers also **have TD Account**, because these categories have significant relationship with personal loan category.

Count of targetable customers	106

Customers with the following Customer ID matches out Target base based on our analysis.

ID

30
39
48
76
132
200
244
248
289
300
439
464
483
567
571
672
723
773
783
933
972
982
1015
1025
1039
1051
1067
1106
1138
1238
1274
1293
1406
1412
1419
1445
1525
1584
1593
1610
1632
1660
1823
1913
1919
1936
1938
2006
2015
2047
2145

2179
2218
2231
2318
2357
2360
2362
2401
2434
2444
2591
2614
2708
2803
2810
2813
2833
2839
2842
2899
2957
2987
3007
3038
3055
3143
3161
3210
3323
3328
3358
3369
3374
3450
3563
3662
3949
3987
4009
4036
4148
4283
4311
4329
4346
4409
4423
4586
4632
4672
4720

4824
4884
4928
4963