



PROJECT ON :HDFC BANK DATA SET

Data analysis using Excel



Submitted By:-Ankith H Poojary

SEPTEMBER 24, 2023

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

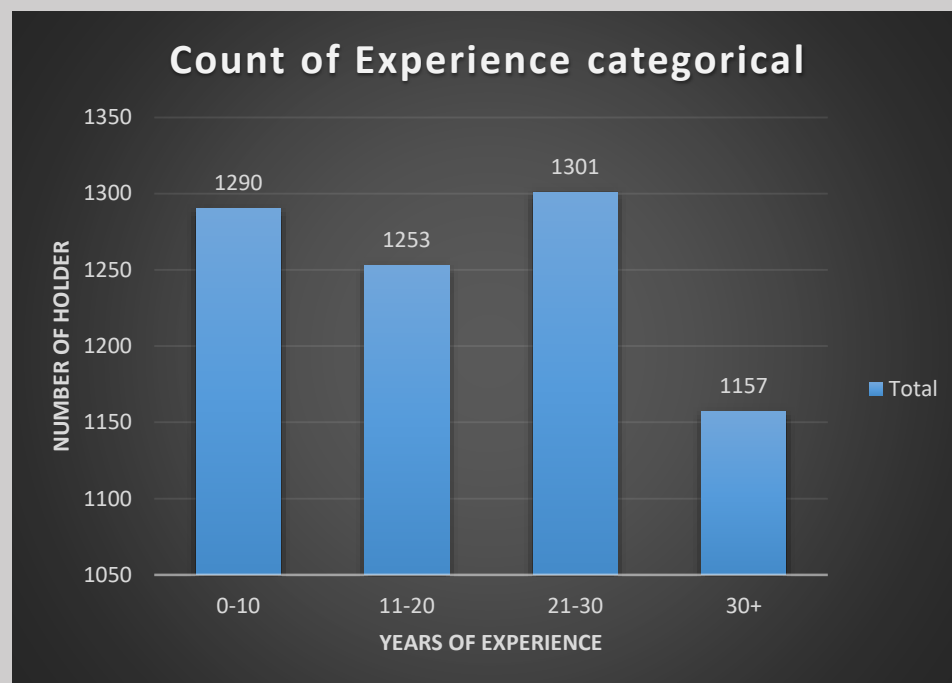
Personal loan	Count of id
No	90.40%
Yes	9.60%
Grand Total	100.00%

Out of total customer of the bank 9.6% customer availed personal loan 90.40% are not availed not

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

Column1	Age	Experience	Income	Family members	CCAVG	Mortgage
minimum	23	0	8	1	0	0
maximum	67	43	224	4	10	635
median	45	20	64	2	1.5	0
average	45	20	74	2	2	56

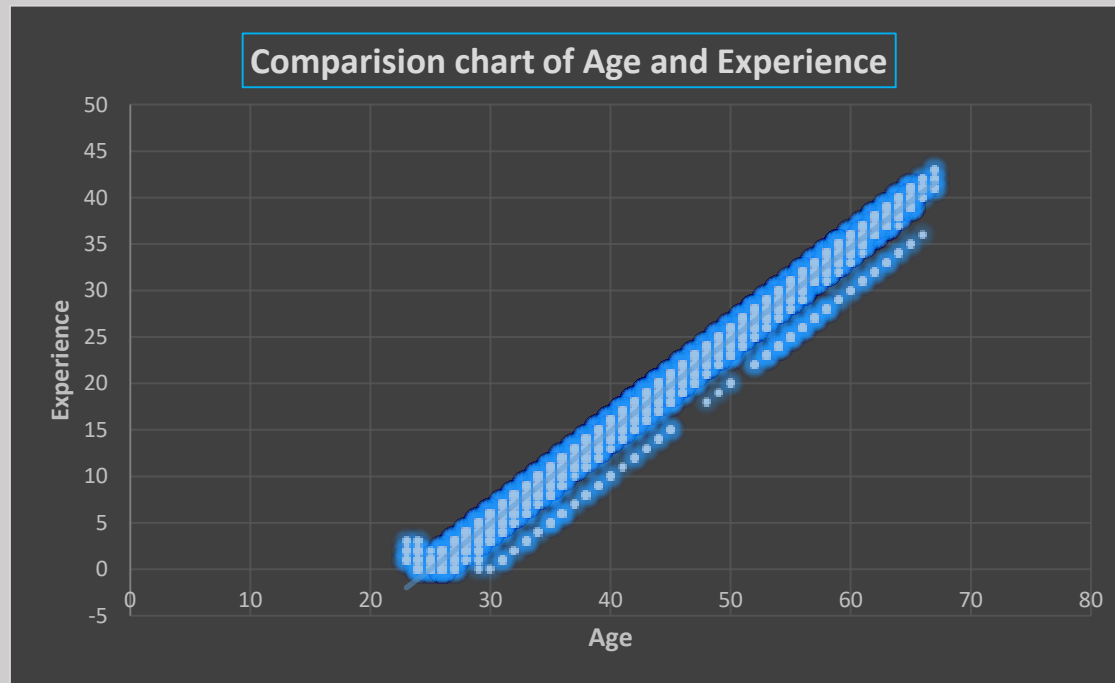
3) Create a new categorical variable for Experience using 4 categories – •0 to 10 years •11 to 20 years •21 to 30 years •30+ years. Plot a bar graph for this new categorical variable



process:-
step1:-created new column as Experience category used nested if to fill the experience in different category
step2:-created pivot table dragged experience category to Row and customer ID to values then choose the option as count
step 3:-selected the result of pivot table then hit inset tab there selected Bar chart
Step4:-customize the chart by adding title , axis title ,Label,design etc.

Insights;-Most of our customers are with the experience of 21-30 followed by 1-10 years of experience

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



Process:-selected age and Experience column and inserted scatter slot

Observation:-Age and Experience have a direct relationship as age increase Experience also increases hence positive sloped curve

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

Zip Code	Count of id
94720	169
94305	127
95616	116

Process:-created pivot table by dragging zip code to row and Id to values and chooses as count

Observation:-Top 1 city-94720

Top2 city-94305

Top3 city-95616

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

TD Account	Yes
Personal Loan	No
Credit Card	Yes

Count of id

147

Process:-in pivot table dragged TD, Personal loan, Credit card to filter and customer ID to values to count

Observation:-147 customer have the combination of TD and credit card but not 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?

	Availed Persona loan	Not Availed personal Loan
Median	142.5	59

In pivot table filtered personal loan to yes and no then dragged income to values and id to row

From The Median it shows that 50% of the customer who Aailed loan are below the Salary of \$ 142,500 And 50% is Above \$ 142,500.
 On the Other hand Median of the persons salary who not Aailed loan it shows that 50% of customer Are below \$ 59,000 And rest 50% is above \$ 59,000

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

A.> Education Vs Personal Loan

Level of Education	Personal loan		Grand Total
	No	Yes	
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%

B>

TD Vs Personal Loan

Term Deposits	Personal loan		Grand Total
	No	Yes	
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

T Online	Personal loan		Grand Total
	No	Yes	
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

T Online	Personal loan		Grand Total
	No	Yes	
0-50	100.00%	0.00%	100.00%
51-100	97.76%	2.24%	100.00%
100+	63.86%	36.14%	100.00%
Grand Total	90.40%	9.60%	100.00%

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 analyse which customers are most likely to take personal loans and why?

In the above tables we compared the conversion rate of customer to loan under different categories like Education, TD, Online User and income category.

- **In the first Table it shows that the conversion rate of Graduates and professionals are more than undergraduates**
- **In the last table comparison with Income category it's clear that the conversion rate income category 0-50 is 0 and the 51-100 is 2.24% which is minimum, most of the customer who took loan are above 100+.**
- **Category of income is the most important variable because in all other categories the conversion rate is normal but income category impacts most as 0-50 are not taken loan and customer above 100+ said yes to loan so we can consider this factor to reach the target customer for the 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 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?

Suggestions:-

- 1. Target to customer who have the income of above \$ 51000, as below \$ 51000 are not accepted the loan**
- 2. Target between the age of 26 to 60**
- 3. we can exclude the customer who have already taken loan previous year**
- 4. Look for the customer who have TD and not taken Loan. we can give loan against TD**