

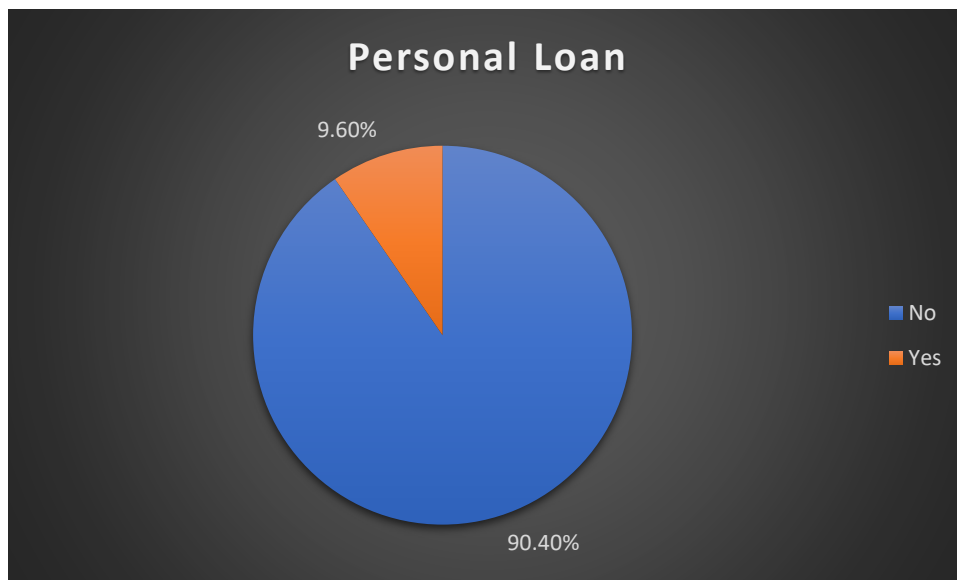
## Week - 1

### BOE – HBFC BANK

#### Assignment – 1

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

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



PIE CHART FOR CUSTOMERS WHO AVAILED PERSONAL LOAN

#### Inference:

- According to the data, there are totally **480 Customers** who have **availed** personal loans out of 5000, with a percentage of **9.60%**.
- **4520 customers** who have **not availed** personal loans i.e., with a percentage of **90.40%**
- The **Orange** colour indicates percentage of peoples who availed personal loans.
- The **Blue** colour indicates percentage of people who have not availed personal loans.

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

Note: Mean is the Average.

- Table for Age & Experience

<i>Age (in years)</i>	<i>Values</i>	<i>Experience (in years)</i>	<i>Values</i>
Minimum	23	Minimum	0
Maximum	67	Maximum	43
Median	45	Median	20
Mean	45.3384	Mean	20.1348

- Table for Income & Family Members

<i>Income (in K/year)</i>	<i>Values</i>	<i>Family members</i>	<i>Values</i>
Minimum	8	Minimum	1
Maximum	224	Maximum	4
Median	64	Median	2
Mean	73.7742	Mean	2.3958

- Table for CCAvg & Mortgage

<i>CCAvg</i>	<i>Values</i>	<i>Mortgage</i>	<i>Values</i>
Minimum	0	Minimum	0
Maximum	10	Maximum	635
Median	1.5	Median	0
Mean	1.937938	Mean	56.4988

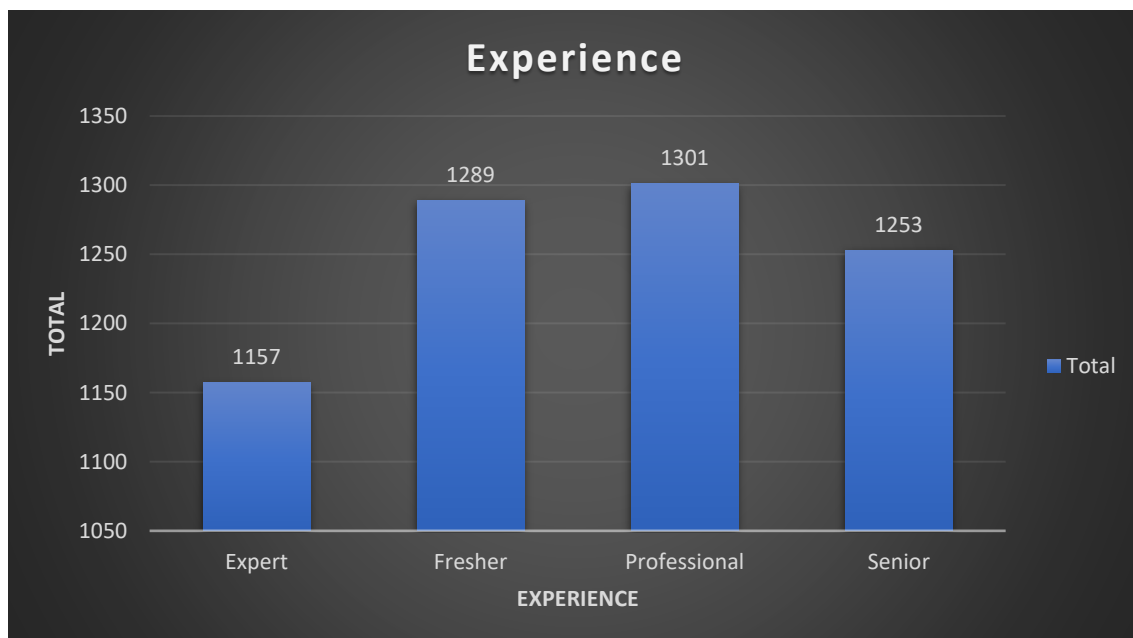
- By taking **Descriptive statistics** for the data given by the bank, we can figure out the Minimum, Maximum, Median, mean (Average) for Age, Experience, Income, Family Members, CCAvg and Mortgage and generated as a table.

### 3) Create a new categorical variable for Experience using 4 categories

- 0 to 10 years - Fresher
- 11 to 20 years - Senior
- 21 to 30 years - Professional
- 30+ years - Expert

Plot a bar graph for this new categorical variable.

Experience	Count of Experience (in years)
Expert	1157
Fresher	1289
Professional	1301
Senior	1253
Grand Total	5000



Create a new categorical variable for Experience that categorizes experience into 4 types as

1. Fresher
2. Senior
3. Professional
4. Expert

A pivot table is created using the categorical data of Experience and total count of them, by this we can create a Bar chart that holds their data as:

- Experience in X – Axis.
- Total Count in Y – Axis.

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



### Using Scatter Plot

By considering 2 Columns in the given data like Experience (in-Years) and Age (in- Years).

- Age (In – Years) as X- Axis.
- Experience (In – Years) as Y – axis.
- **Inference** – the above scatter plots states that when **Age** increases, **Experience** also increases.

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

ZIP Cod	values
94720	169
94305	127
95616	116
90095	71
93106	57
93943	54
92037	54
91320	53
91711	52
94025	52

Top 3	ZIP Code	Values
1	94720	169
2	94305	127
3	95616	116

- By taking Zip Code column to find the top 3 areas:  
Inference - In our analysis, we have established the three leading areas:
  1. 94720 – with 169 customers
  2. 94305 – with 127 customers
  3. 95616 – with 116 customers

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

- Totally **147 customers** have a combination of fixed deposits and Credit cards but not Personal loan

By filtering the bank table with the respective columns like customers with a combination of “Yes” in both **fixed deposit** and **credit card**, but **personal loan** as “No.”

We can find the total customers with this combination.

#### 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 for Customers who availed personal loan is **142.5**
- Median for Customers who have not availed personal loan is **59**.

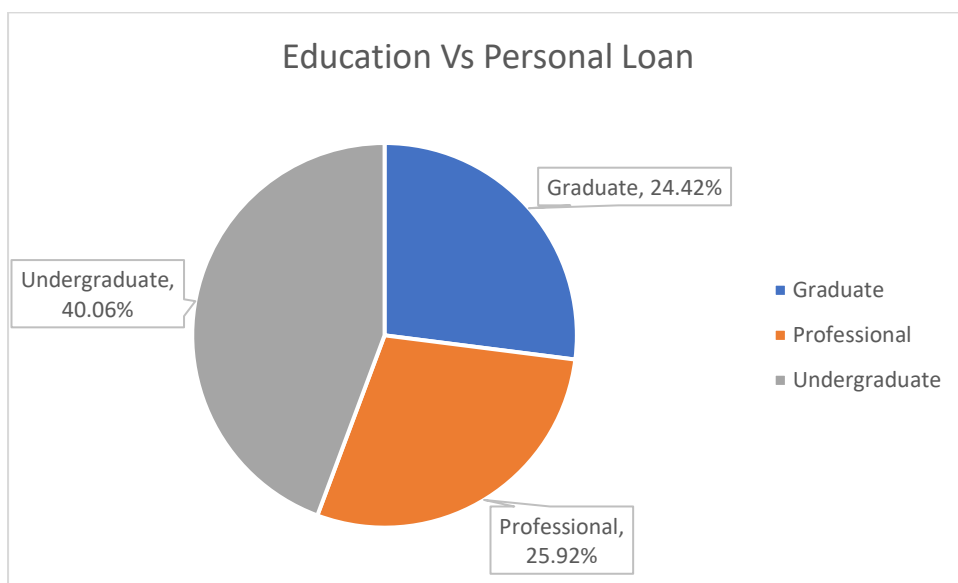
By comparing both the median income of the customers who availed personal loan and not,

**Inference** - Our analysis reveals that the median income of the customers who have **availed personal loans** is **higher** with a value of **142.5**. and comparatively who have **not availed personal loan** is **lower** with a value of **59**.

8) Create 4 separate Pivot Tables. Summarize your data by percentage values.

a) Education vs Personal Loan

Count of Personal Loan	Personal loan		
Education	No	Yes	Grand Total
Graduate	24.42%	3.64%	28.06%
Professional	25.92%	4.10%	30.02%
Undergraduate	40.06%	1.86%	41.92%
Grand Total	90.40%	9.60%	100.00%



The above pivot table and pie chart states that the customers who have not availed personal loans by their educational qualifications.

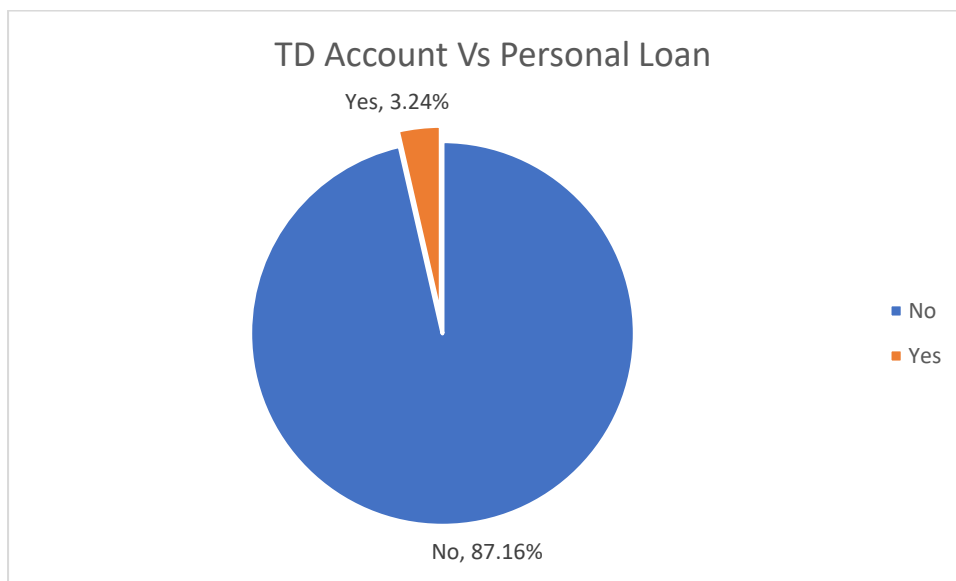
In which only 9.60% of customers have availed personal loans.

**Inference:**

- In this data we can clearly state that the percentage of professional customers have got approved for the personal loans are higher than the others.
- Undergraduate's customers percentage is comparatively lower than the other two educational qualification.

### 8b) TD Account Vs Personal Loan

Count of Personal Loan	Personal loan		
TD Account	No	Yes	Grand Total
No	87.16%	6.80%	93.96%
Yes	3.24%	2.80%	6.04%
Grand Total	90.40%	9.60%	100.00%

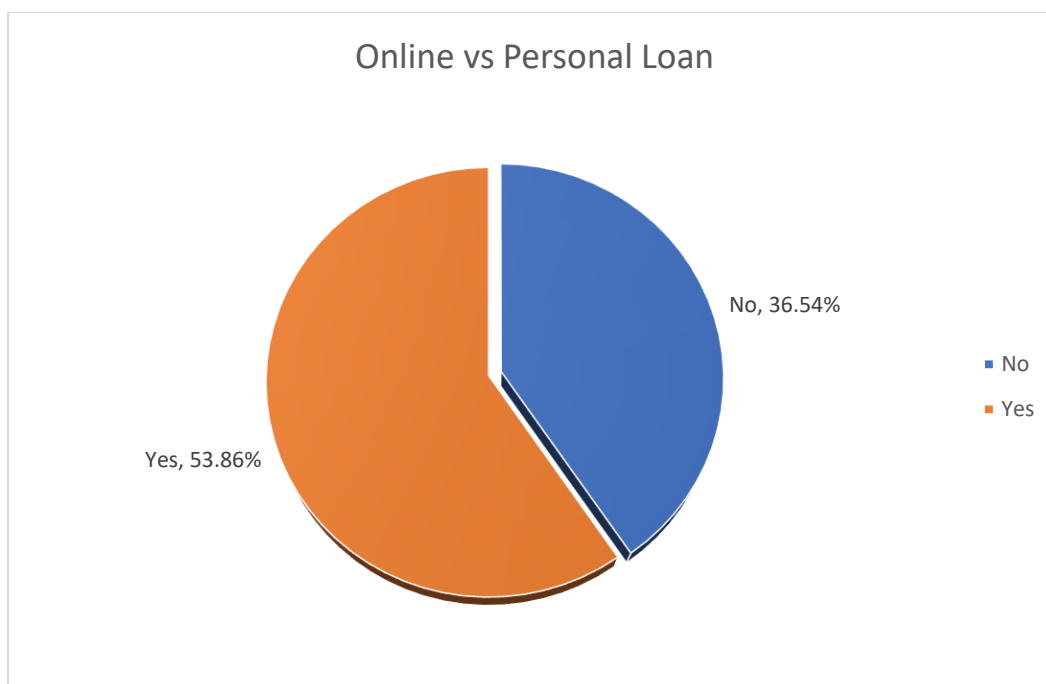


The above pivot table and pie chart shows the customers who have fixed deposit in the bank and who have availed to personal loan respectively.

**Inference** - The above pie table states that the customers who does not have a fixed deposit account in the bank has the highest chance of getting personal loan, customers who does have fixed deposit account has the lowest percentage of availing personal loans.

### 8C) Online vs Personal Loan

Count of Personal Loan	Personal Loan		Grand Total
	No	Yes	
No	36.54%	3.78%	40.32%
Yes	53.86%	5.82%	59.68%
Grand Total	90.40%	9.60%	100.00%



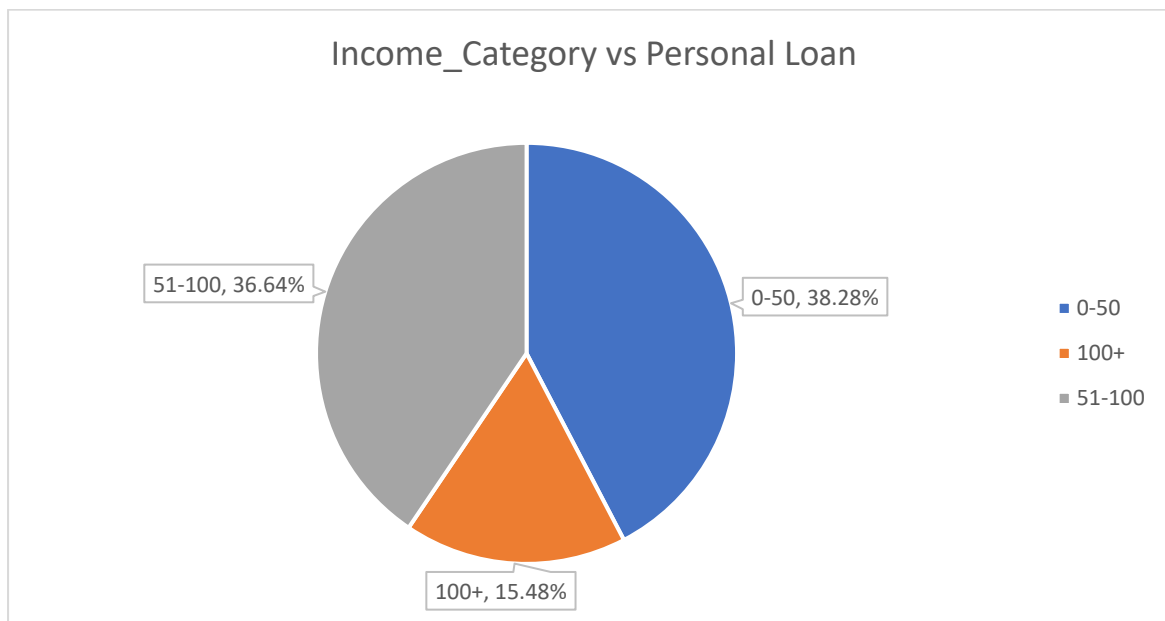
The above pivot table and pie chart states the data between online customers who have availed for personal loans.

**Inference-** in which the rate of total percentage of customers who use online Net banking has the higher rate of getting personal loans when compared to the customers who does not have online Net banking has lower percentage rate of getting personal loan.



#### 8d) Income Category vs Personal Loan

Count of Personal Loan	Personal Loan		Grand Total
	No	Yes	
Income_Category			
0-50	38.28%	0.00%	38.28%
100+	15.48%	8.76%	24.24%
51-100	36.64%	0.84%	37.48%
Grand Total	90.40%	9.60%	100.00%



The above pivot table and pie chart states the data between the customers Income category and Personal loan.

#### Inference:

The table states that the customers in the category of 100+ have the highest acceptance rate for personal loan at 8.67%. this category group represents the individual customer with income rate of 100+. In other side the customers with 0 – 50 has a zero percent rate for personal loan which is the lowest rate in the table.

**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?**

From the analysis of the above pivot tables and pie charts we can find that there is a discrepancy in the **income category vs personal loan** table where customers in the income category with 100+ income rate has the highest acceptance rate for personal loan with 8.76%. Still the customers in this 100+ rate have not availed personal loan. By sorting this category customers we can increase the rate of personal loan by focusing on the rest of this category.

In between there lies the customers with 51-100 as their income category with the lowest percent of acceptance rate in personal loans by the bank, with the rate of just 0.84% which is nearest to 1 %.

By considering this income categorical vs personal loan pivot table which appeared to be one of the most important analyses. By which the income category of the customers lies between the range of 51-100. There is a possibility to increase the personal loan availing customers by changing the range of it accordingly.

**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?**

1. By analysing the previous campaign data, we can observe that customer with 100+ income rating has a highest possibility of accepting the personal loan.
2. Yet few of the same income category customers have not availed personal loan.
3. The HBFC Bank can Target the higher income rate category customers for more personal loans.
4. Based on the insights, the bank may consider targeting its personal loan marketing campaigns more heavily towards customers in the 100+ income category, as they have a significantly higher likelihood of acceptance.
5. The bank could also review and approach to customers in the other 0- 50 and 51 – 100 income categories also to understand why their acceptance rates are so low.