

ONE BHARAT DS ASSIGNMENT SUBMISSION

By : Divjot kaur

- Email : divjotbhullar006@gmail.com
- The entire report was created using pandas, numpy, matplotlib, and seaborn in Python for data analysis and visualization.
- To access the code for this report visit the GitHub link:
<https://github.com/DivKjot/DS-assignments>
- The results are displayed in this PowerPoint presentation.

Bank Statement Analysis

- Transaction Analysis

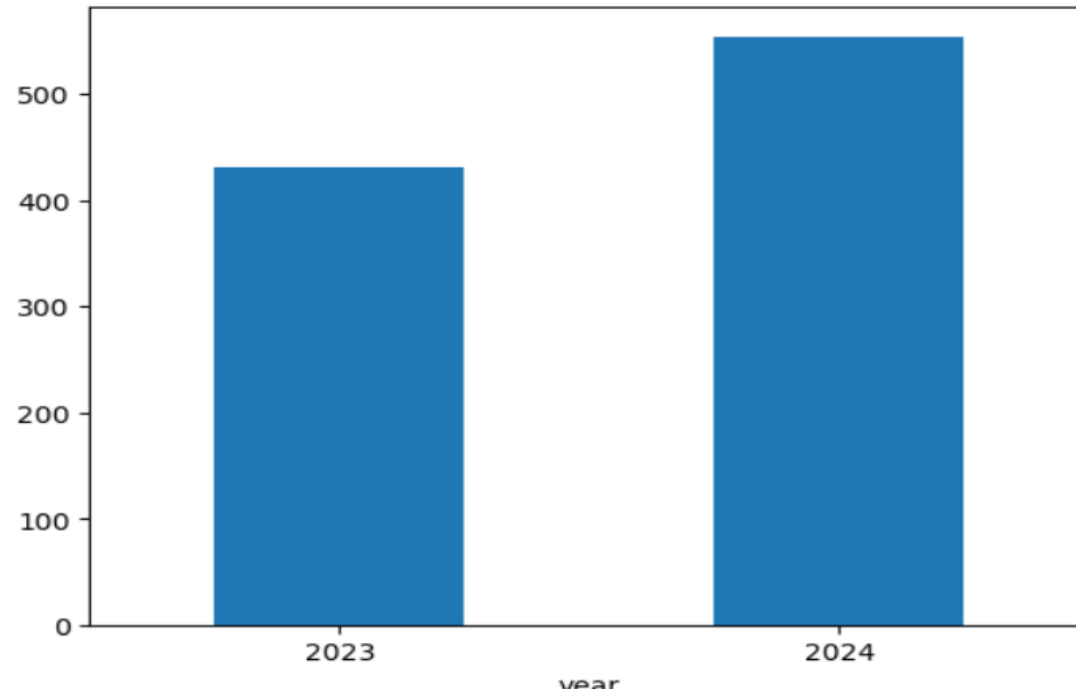
- What is the total number of transactions made over the year?

Year No. of Transactions:

2023 431

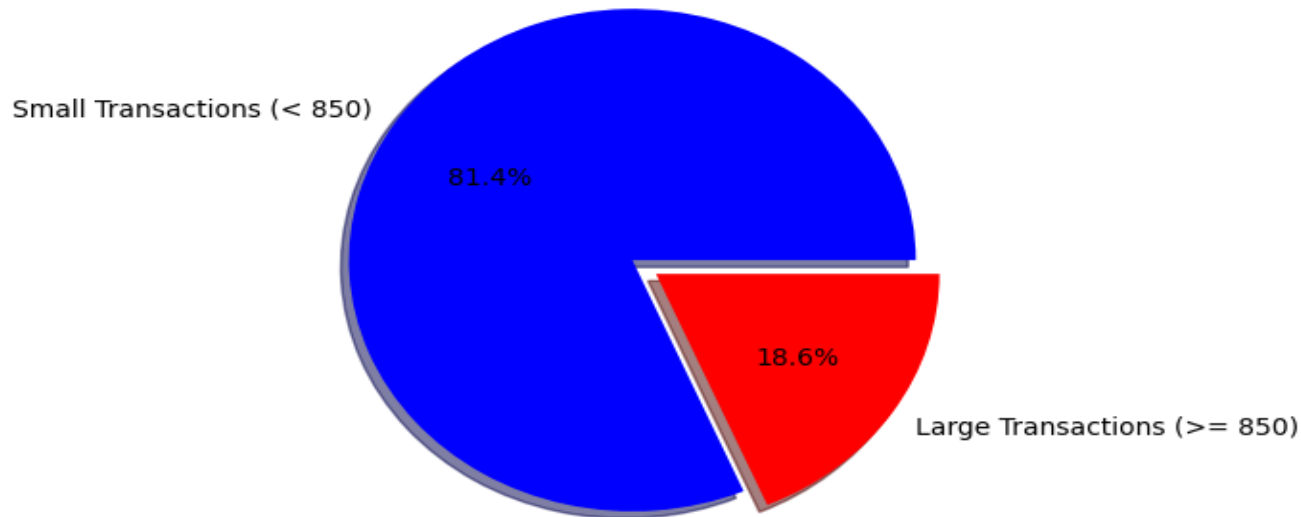
2024 554

- There is a Increase of 123 transactions from the year 2023 to 2024.

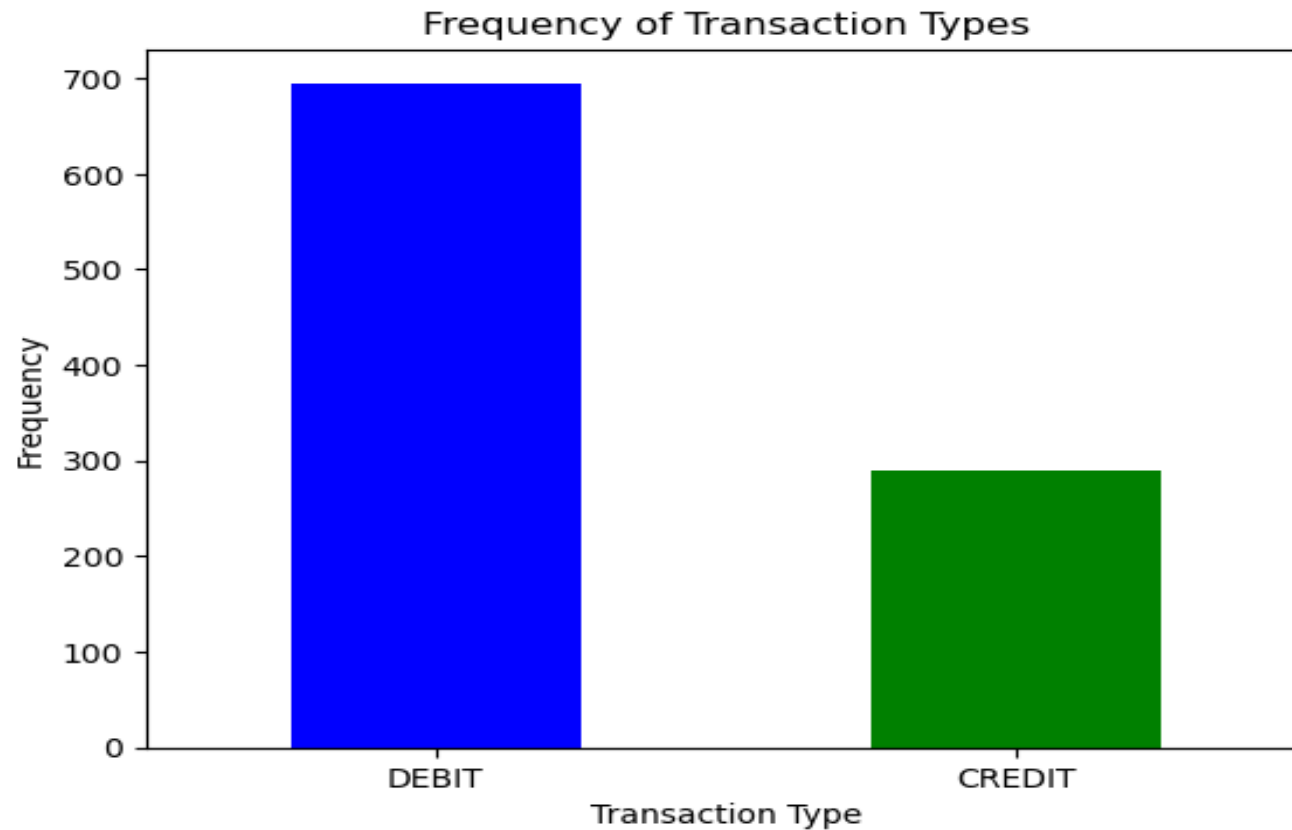


- What is the distribution of transaction amounts (e.g., small vs. large transactions)?(define small and large transactions by yourself)
- let transactions amount < 850 be "Small"
- let transactions amount ≥ 850 be "Large"
- Here number of small transactions : 802
- Number of large transactions : 183

Distribution of Transaction Amounts

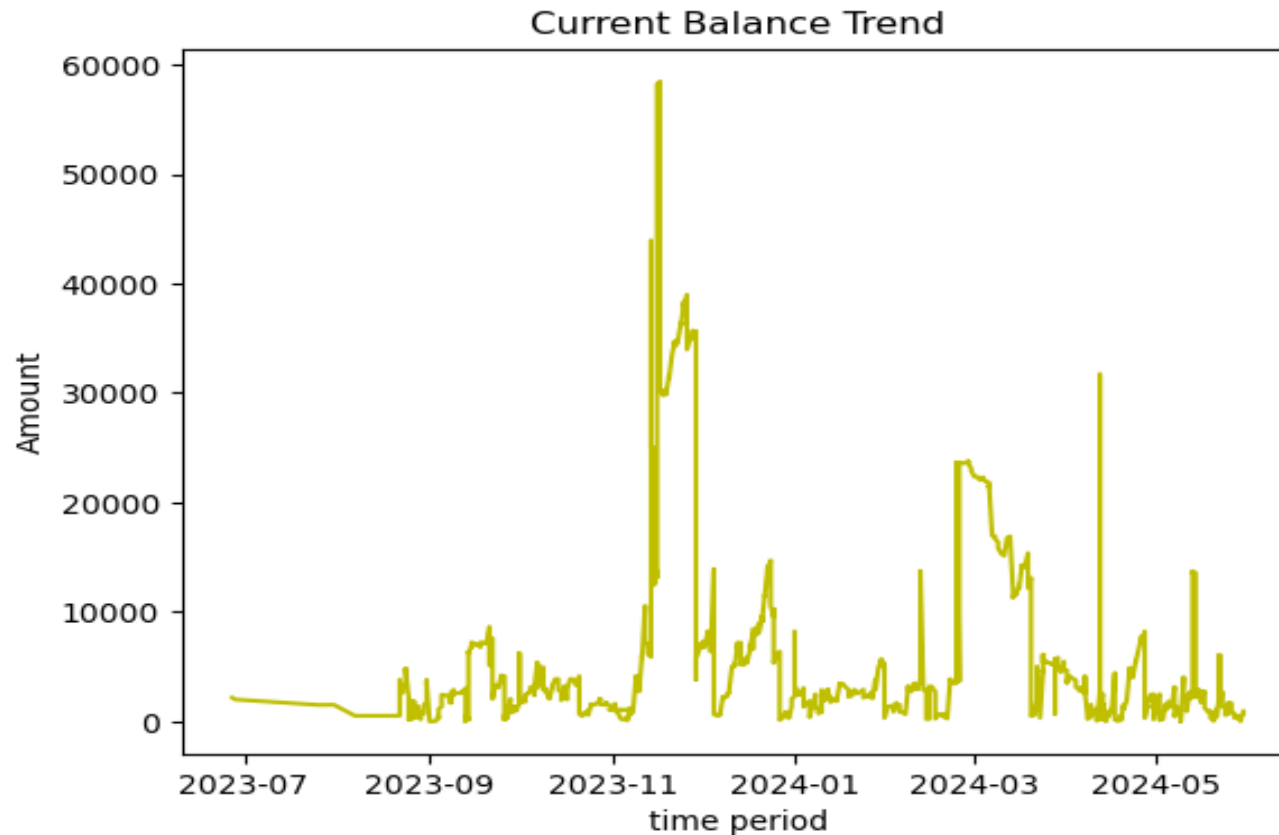


- Analyze the frequency of different transaction types (debit vs. credit).
- DEBIT 695
- CREDIT 290



Balance Analysis

- What is the trend of the account balance over time?



The current balance trend shows significant fluctuations, characterized by a sharp increase in November 2023 followed by a pronounced decrease in December 2023. Despite of sharp spikes at some point of time the average balance remains around Rs. 5901

Spending Patterns

- What are the main categories of expenses (e.g., fuel, Ecommerce, food, shopping, ATM withdrawals, UPI transactions)?

type	mode	
CREDIT	OTHERS	188
	UPI	101
	CASH	1
DEBIT	UPI	688
	CARD	4
	ATM	3

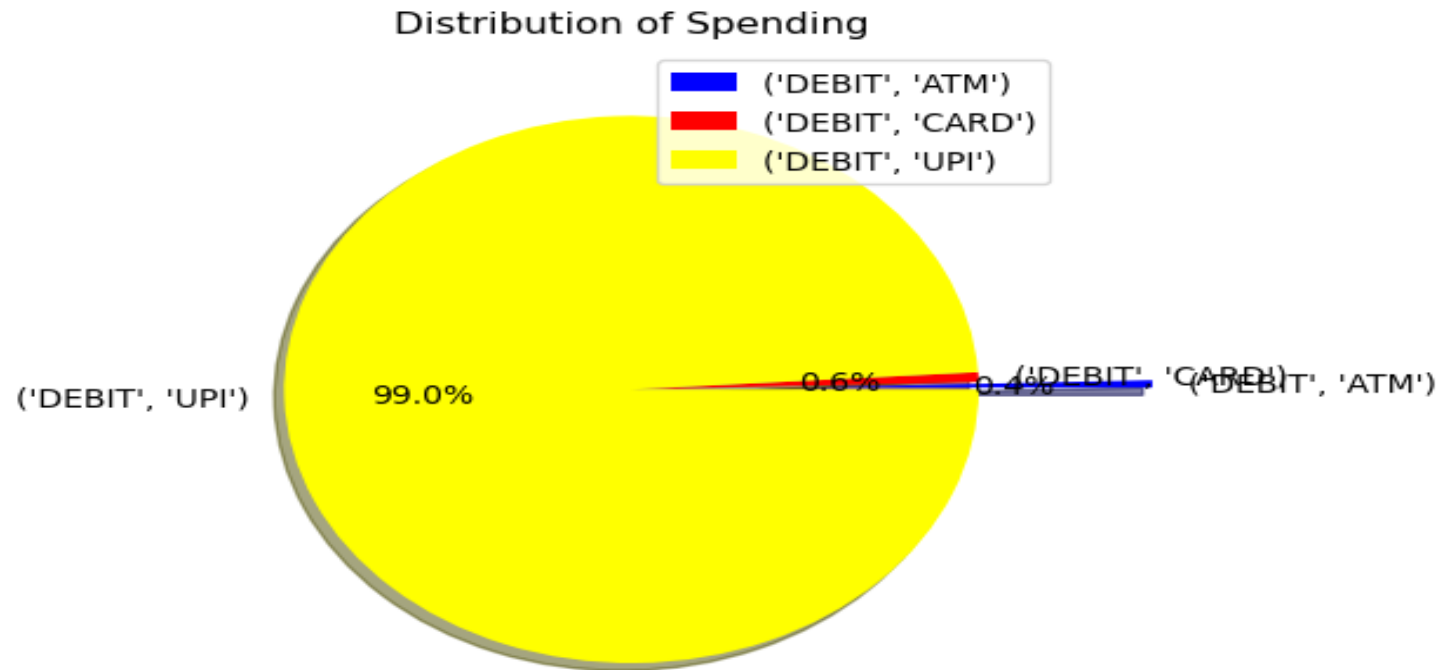
Spending:type mode : DEBIT

ATM 3

CARD 4

UPI 688

- Analyze the frequency and amount of spending in each category.
- UPI is the most commonly used method for spending and CASH is the least used method for spending by customers. Indicating a transformation towards a **cashless economy**.



Income Analysis

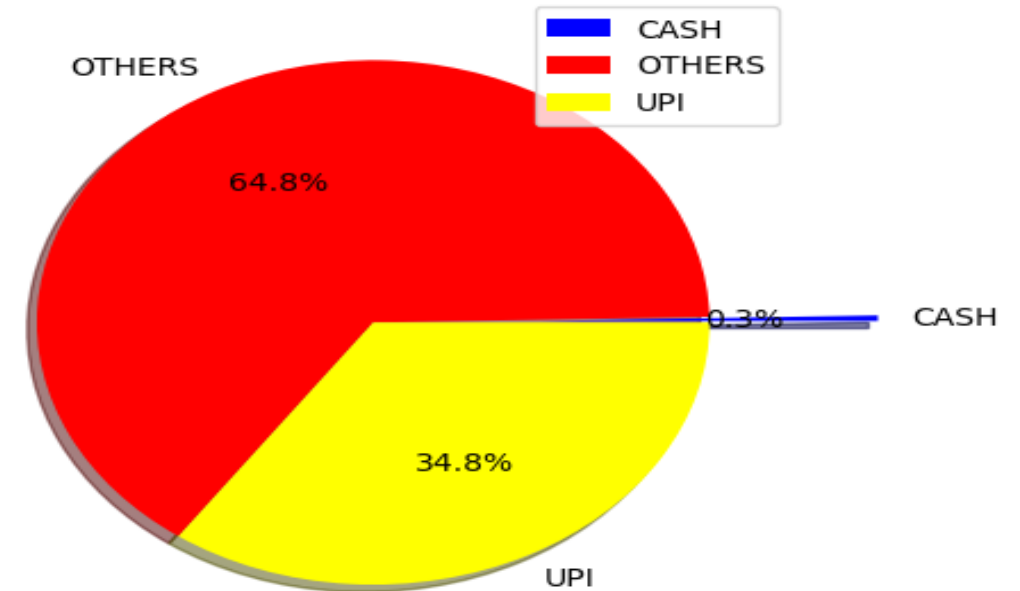
- What are the main sources of income (e.g., salary, UPI credits)?

- Income:

mode

CASH	1
OTHERS	188
UPI	101

- Income received from the others methods is the highest which might include Salary ,receipts etc. UPI is the 2nd most common source of income .
- CASH is again the least used method for receiving Income by customers.
- Only 0.3% of the total payments received are cash Payments



Alert Generation

- Generate alerts for low balance or high expenditure periods.
- LOW balance Alert is created for transactions with amount < Rs. 10

Low Balance Alerts:					
	amount	type	mode	currentBalance	valueDate
5	1.0	DEBIT	UPI	524.80	2023-08-22
21	1.0	CREDIT	OTHERS	1155.80	2023-08-25
22	1.0	DEBIT	UPI	1154.80	2023-08-25
23	1.0	DEBIT	UPI	1153.80	2023-08-25
25	1.0	CREDIT	OTHERS	1175.80	2023-08-25
29	1.0	DEBIT	UPI	195.80	2023-08-26
35	1.0	DEBIT	UPI	1814.80	2023-08-27
64	1.0	DEBIT	UPI	2898.80	2023-09-09
95	1.0	CREDIT	UPI	5151.80	2023-09-21
114	1.0	DEBIT	UPI	4153.80	2023-09-26
137	1.0	DEBIT	UPI	6237.80	2023-10-01
175	5.0	DEBIT	UPI	2958.80	2023-10-10
226	2.0	CREDIT	OTHERS	1547.80	2023-10-30
273	1.0	CREDIT	UPI	5892.80	2023-11-14
274	1.0	CREDIT	UPI	5893.80	2023-11-14
276	1.0	DEBIT	UPI	43891.80	2023-11-14
284	1.0	DEBIT	UPI	22970.80	2023-11-15
300	1.0	DEBIT	UPI	30369.80	2023-11-19
321	1.0	DEBIT	UPI	38972.80	2023-11-26
335	1.0	DEBIT	UPI	3824.80	2023-11-29
336	1.0	DEBIT	UPI	3823.80	2023-11-29
367	5.0	DEBIT	UPI	2629.80	2023-12-10
413	1.0	CREDIT	UPI	10256.41	2023-12-25
457	1.0	CREDIT	UPI	878.61	2024-01-10
503	7.0	DEBIT	UPI	5623.81	2024-01-30
542	1.0	DEBIT	UPI	2923.81	2024-02-14
546	4.0	CREDIT	UPI	223.81	2024-02-15
653	5.0	DEBIT	UPI	589.11	2024-03-21
688	1.0	DEBIT	UPI	5473.11	2024-03-31
784	1.0	DEBIT	UPI	5580.21	2024-04-24
785	1.0	CREDIT	UPI	5581.21	2024-04-24
793	1.0	DEBIT	UPI	8230.21	2024-04-27
801	1.0	CREDIT	UPI	1309.21	2024-04-28

- A Low Alert is generated for transactions with amount less than Rs. 10.
- This table displays the type (Credit or Debit) of Low transactions along with mode and current balance.
- Since the number of low transaction are quite large the bank can decide to increase the limit for such transactions as they increases the non-beneficial traffic on the site.

- High balance Alert is created for transactions with amount > Rs. 15000

High Balance Alerts:

	amount	type	mode	currentBalance	valueDate
275	37999.0	CREDIT	UPI	43892.80	2023-11-14
277	16500.0	DEBIT	UPI	27391.80	2023-11-14
291	45000.0	CREDIT	CASH	58180.80	2023-11-16
295	21000.0	DEBIT	UPI	35400.80	2023-11-17
332	19000.0	DEBIT	UPI	16625.80	2023-11-29
570	20000.0	CREDIT	UPI	23483.91	2024-02-24
572	20000.0	DEBIT	UPI	3683.91	2024-02-25
576	20000.0	CREDIT	UPI	23664.91	2024-02-25
743	30000.0	CREDIT	UPI	31701.21	2024-04-12
744	30000.0	DEBIT	UPI	1701.21	2024-04-12

- A High Alert is generated for transactions with amount more than Rs. 15000.
- This table displays the type (Credit or Debit) of High transactions along with mode and current balance.
- Bank can decide to offer special discounts to such customers making high amount of transactions.

- Identify any unusual or suspicious transactions.
- Suspicion Alert for all transactions with amount exceeding Rs.25000

Suspicious Transactions:

	amount	type	mode	currentBalance	valueDate
275	37999.0	CREDIT	UPI	43892.80	2023-11-14
291	45000.0	CREDIT	CASH	58180.80	2023-11-16
743	30000.0	CREDIT	UPI	31701.21	2024-04-12
744	30000.0	DEBIT	UPI	1701.21	2024-04-12

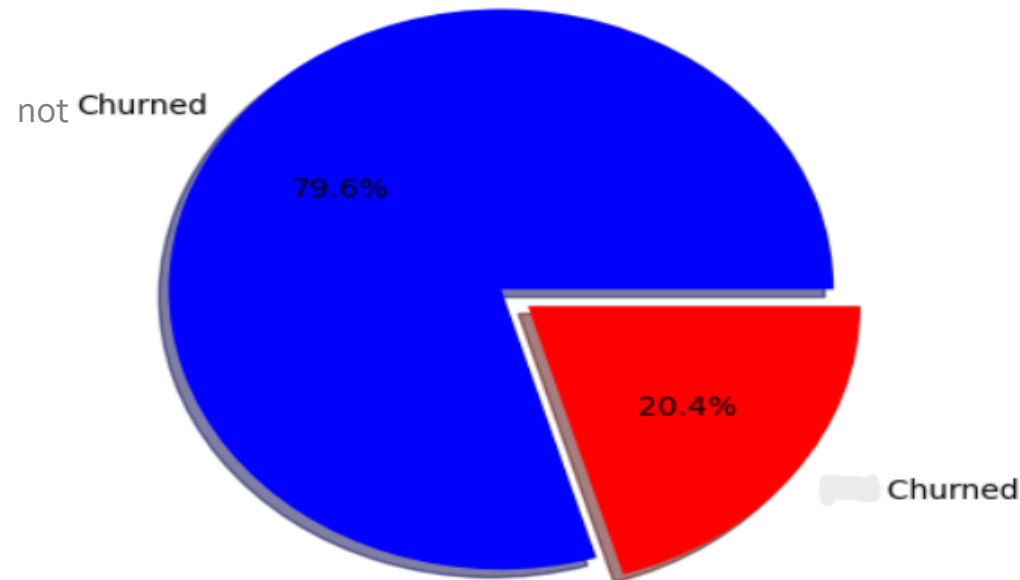
- A suspicion Alert is generated for transactions with amount exceeding not only the high threshold of Rs. 15000 but also an amount of Rs. 25000.
- This displays the type of suspicion transactions along with mode and current balance.
- Bank may decide to monitor the activities of such customers and further investigation.

Churn Modelling Data Analysis

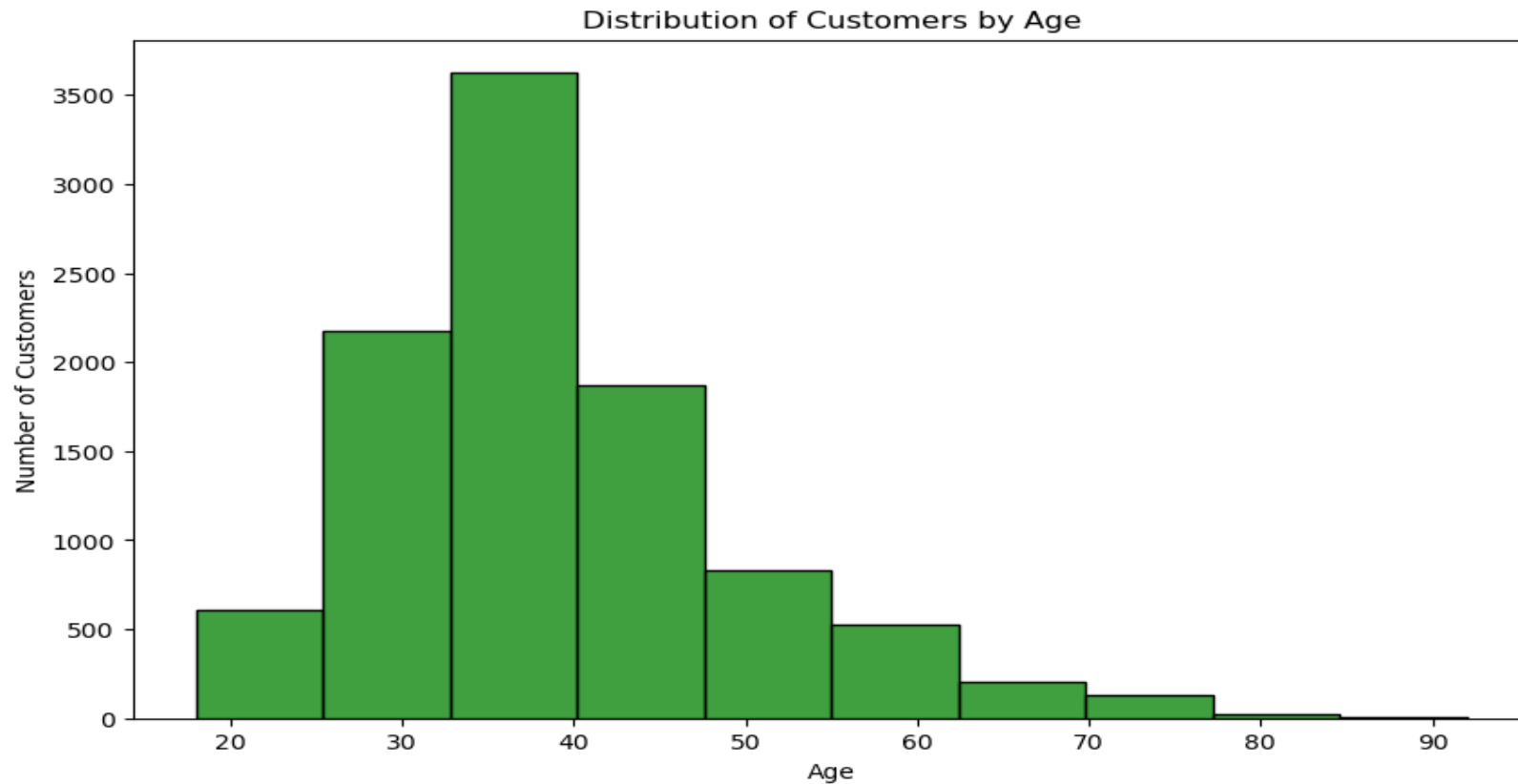
- Distribution of customers(churned / Not churned)
- Churned (status 1): 2037
- Not - Churned (status 0): 7963
- Number of customers Churned is < Number of customers not churned

but number Churned customers are in significant numbers which indicates that there is a need for the company to change its measures to reduce the rate of churned customers.

Distribution of Churned Customers



- What is the distribution of customers across different age groups?
- Histogram is used to show the distribution of customers across different age groups majority of the customers lie in age group of 25-45 and very few are above the age of 65 (senior citizens)



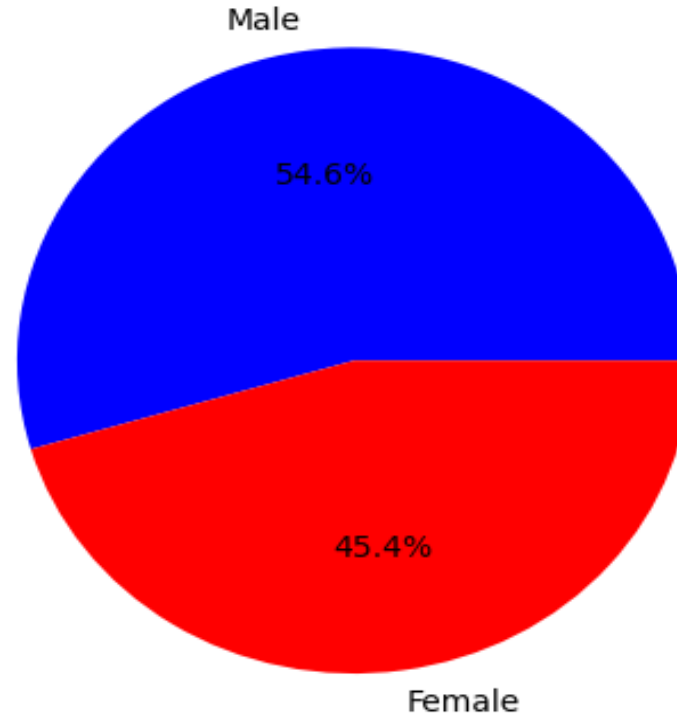
- Analyze the gender distribution of customers.

No. of Male : 5457

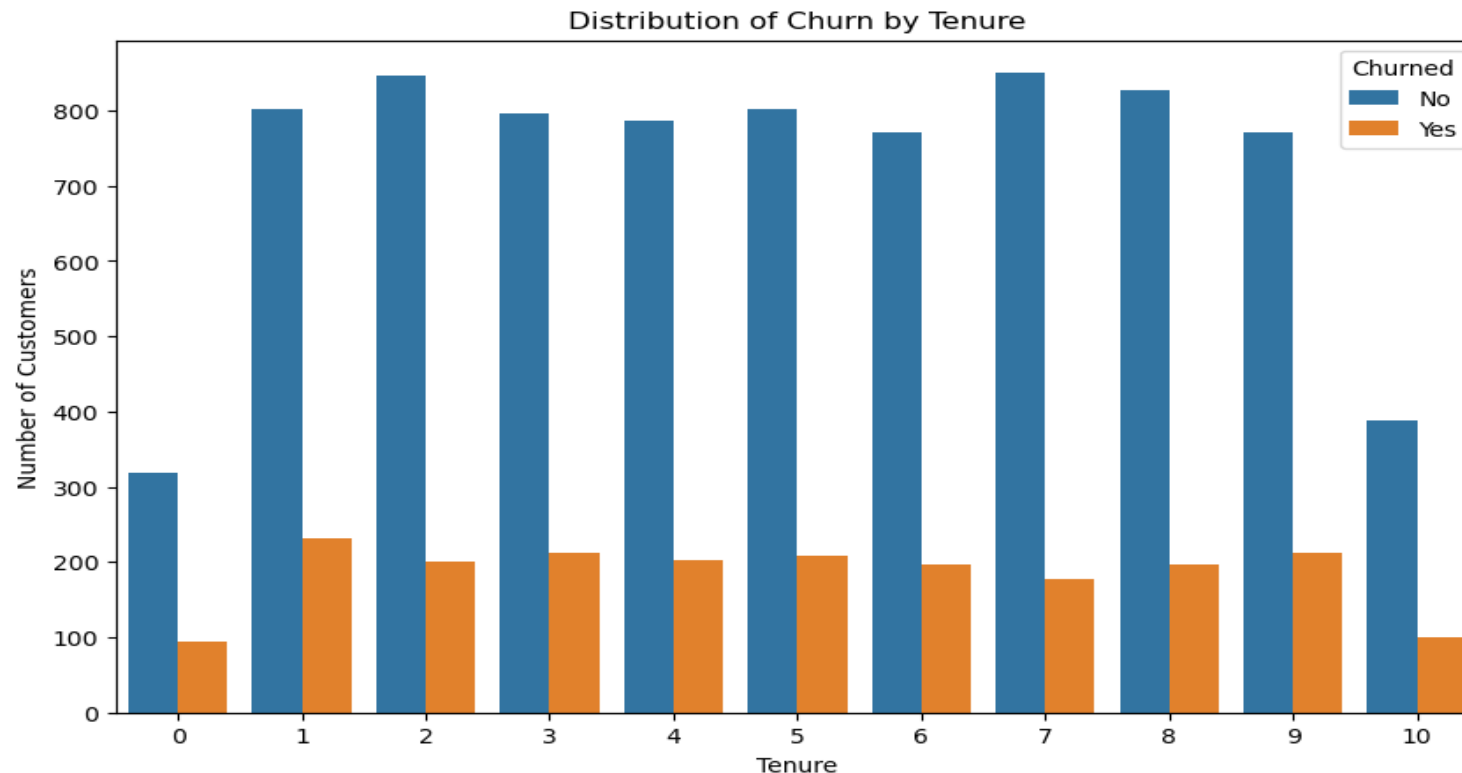
No. of Female : 4543

- Distribution of male and female customers is quite well distributed there is no gender based bias.

Gender Distribution of Customers



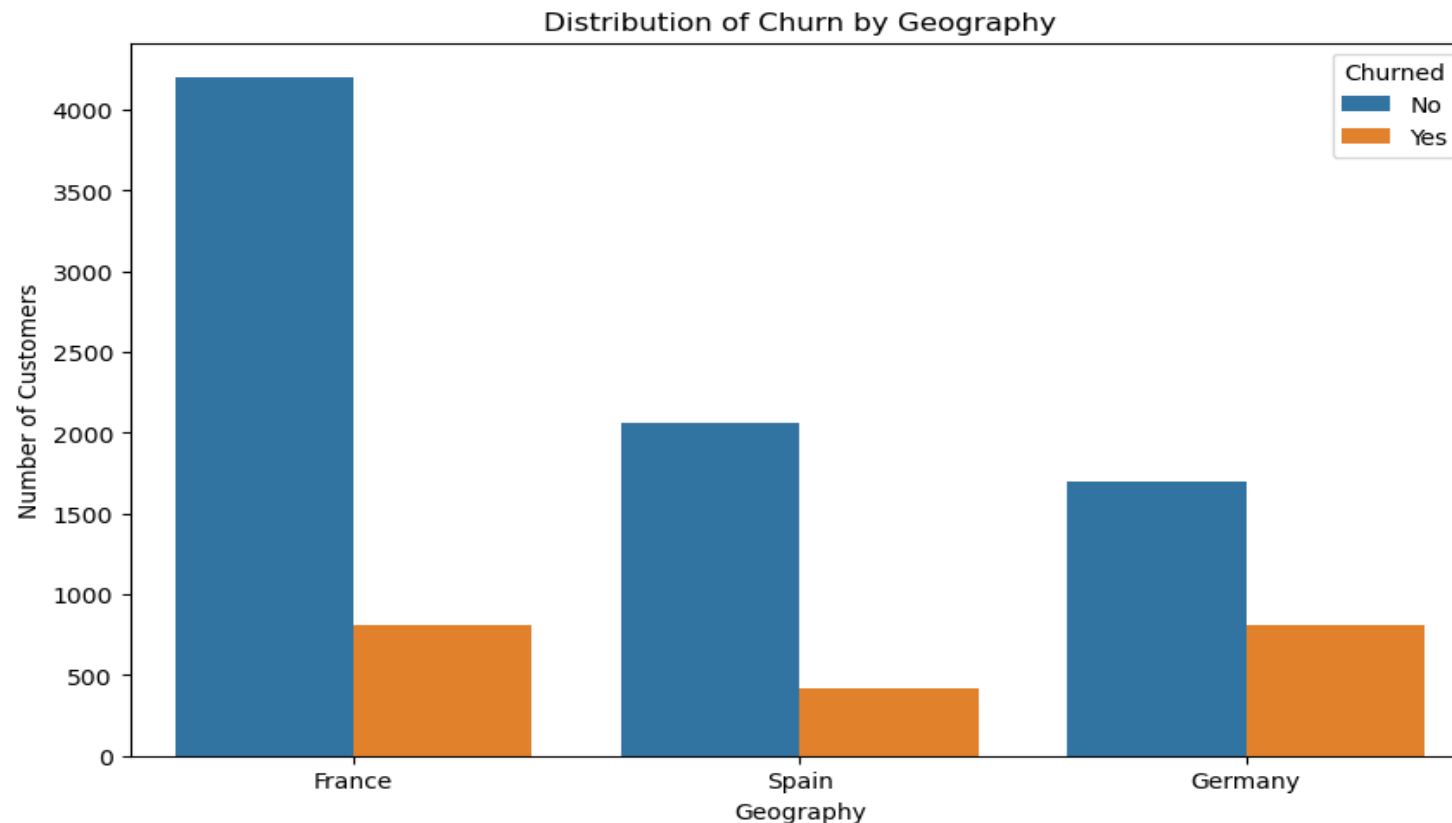
- What percentage of customers have churned?
- Percentage of customers who have churned : 20.3697
- What are the main reasons for customer churn?
- In order to identify the reason for churned customer let us examine various measures such as Tenure ,Geography and Gender



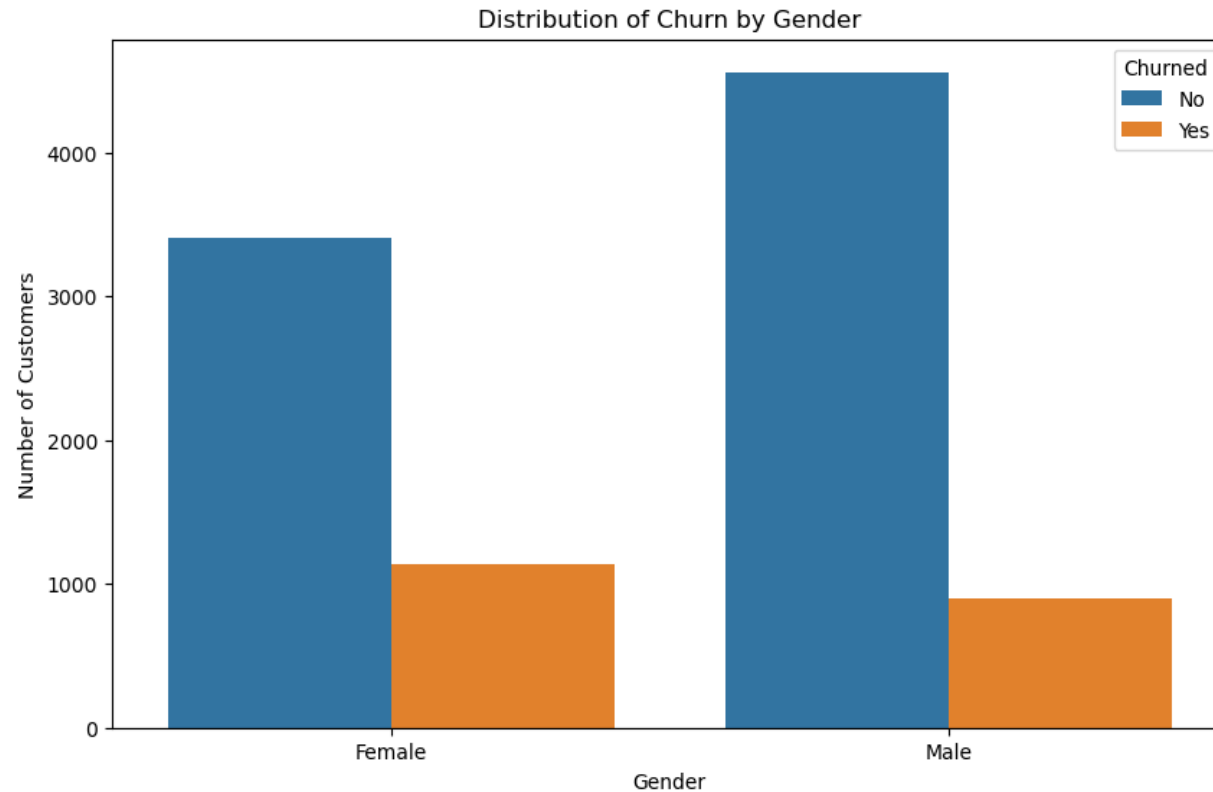
- Customers with higher Tenure rate do not churn often indicating customer loyalty.

- Company should provide discounts on long term schemes in order to retain customers for longer duration and in turn reduce the churning rate.

- Distribution of Churn by Geography:
- France has the least churn rate indicating that company has well researched the market in that location and customers are happy.
- Whereas Germany has the Highest churning rate indicating that the Company should adopt methods to understand the customer needs and adopt methods to reduce the churn rate as the Churned customers increases the cost for the company.



- Gender based churned rate is not significant
- Churning rate is higher in females as compared to males suggesting that company could include products/services to cater to the female customers as well.



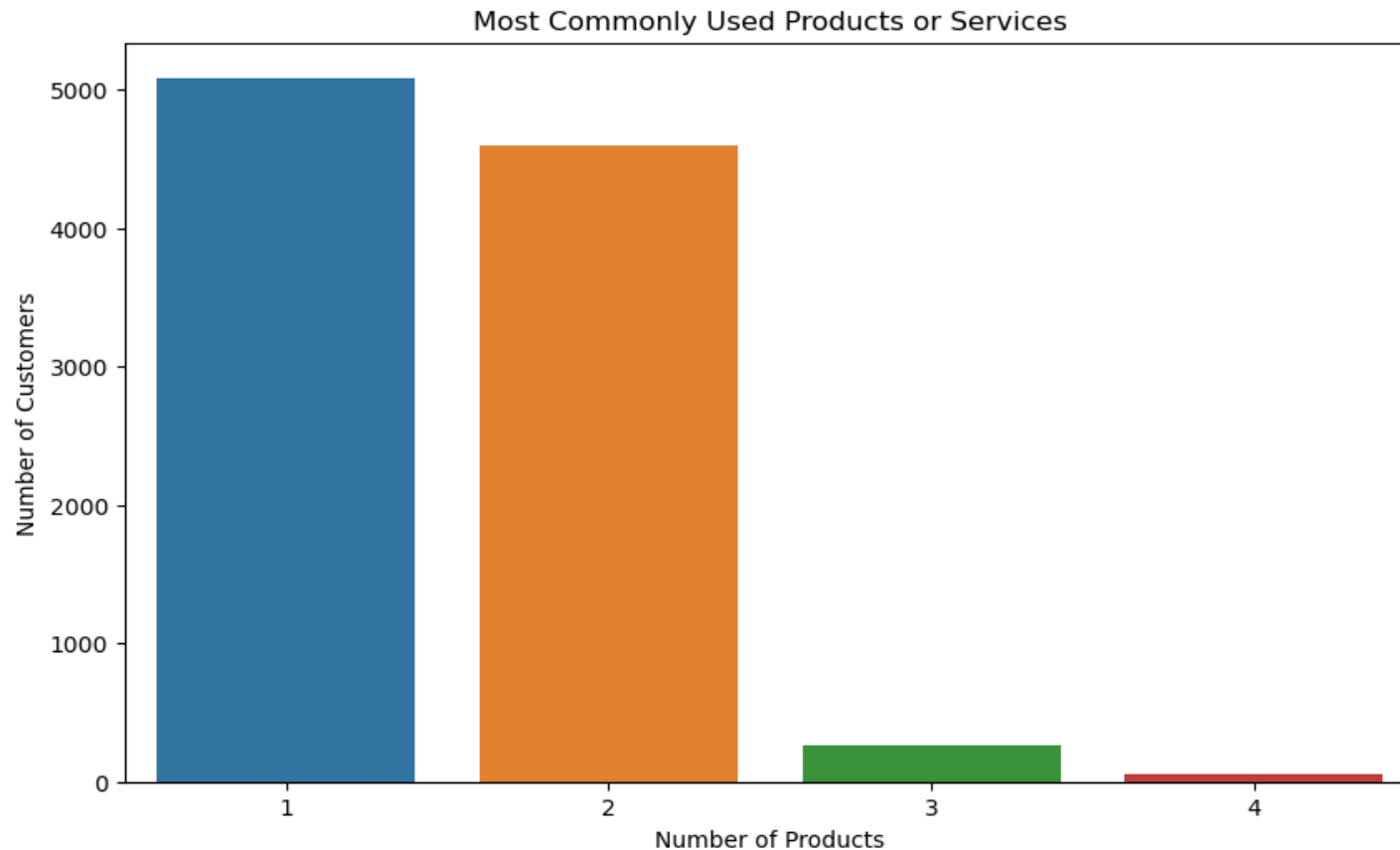
- Trends among customers who have churned.

	RowNumber	CustomerId	CreditScore	Age	Tenure \
count	2037.000000	2.037000e+03	2037.000000	2037.000000	2037.000000
mean	4905.917526	1.569005e+07	645.351497	44.837997	4.932744
std	2866.855245	7.269262e+04	100.321503	9.761562	2.936106
min	1.000000	1.556571e+07	350.000000	18.000000	0.000000
25%	2419.000000	1.562736e+07	578.000000	38.000000	2.000000
50%	4871.000000	1.568896e+07	646.000000	45.000000	5.000000
75%	7404.000000	1.575309e+07	716.000000	51.000000	8.000000
max	9999.000000	1.581566e+07	850.000000	84.000000	10.000000

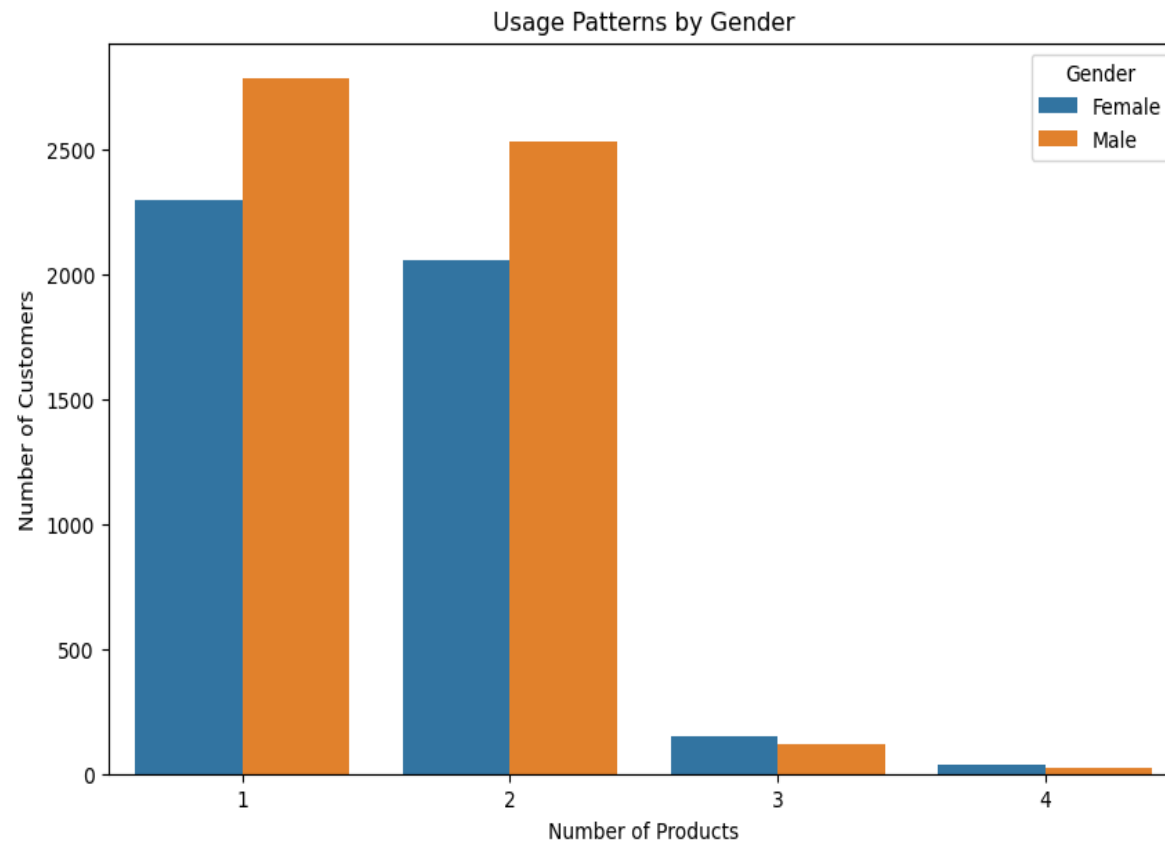
	Balance	NumOfProducts	HasCrCard	IsActiveMember \
count	2037.000000	2037.000000	2037.000000	2037.000000
mean	91108.539337	1.475209	0.699067	0.360825
std	58360.794816	0.801521	0.458776	0.480358
min	0.000000	1.000000	0.000000	0.000000
25%	38340.020000	1.000000	0.000000	0.000000
50%	109349.290000	1.000000	1.000000	0.000000
75%	131433.330000	2.000000	1.000000	1.000000
max	250898.090000	4.000000	1.000000	1.000000

	EstimatedSalary	churned
count	2037.000000	2037.0
mean	101465.677531	1.0
std	57912.418071	0.0
min	11.580000	1.0
25%	51907.720000	1.0
50%	102460.840000	1.0
75%	152422.910000	1.0
max	199808.100000	1.0

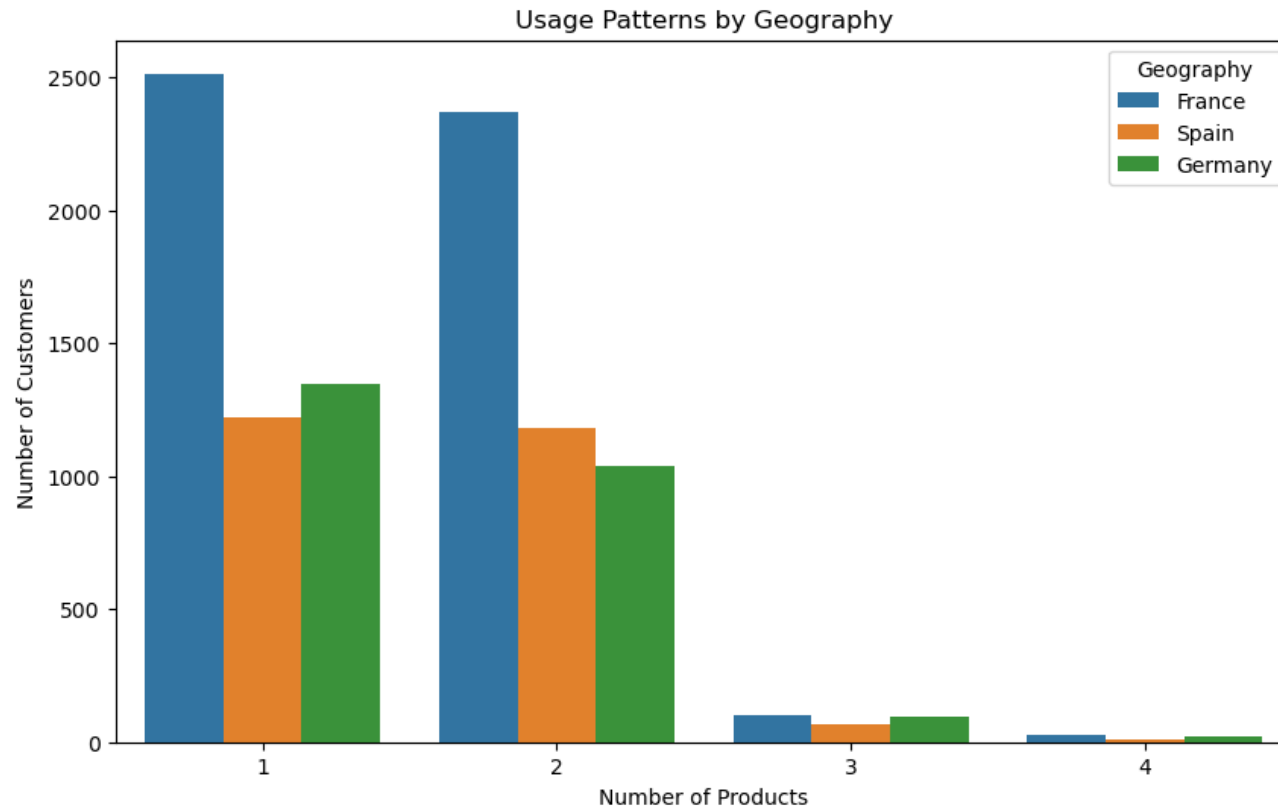
- What are the most commonly used products or services?
- Product id 1 and 2 are the most famous products among customers
company can eliminate the least populous products among the customers.



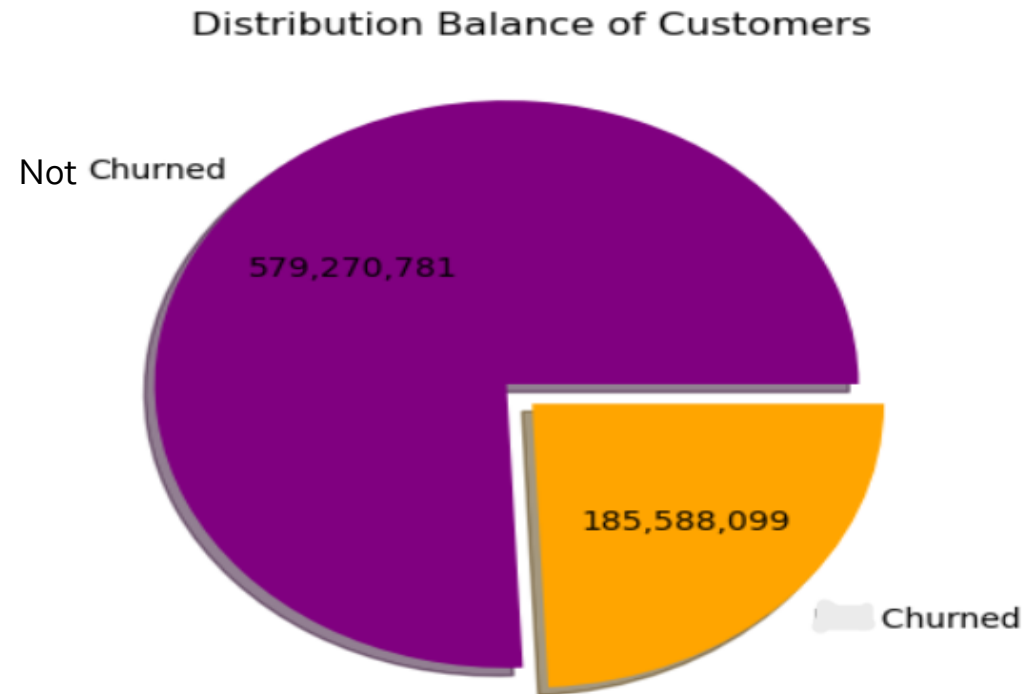
- Analyze the usage patterns of different customer segments.
- Usage patterns of different customer by GENDER:
- There is no significant difference in the preference of different products among male or female both prefer product 1 and 2 over 3 and 4.



- Usage patterns of different customer by *GEOGRAPHY*:
- Product 1 and 2 are most preferred in all the locations .



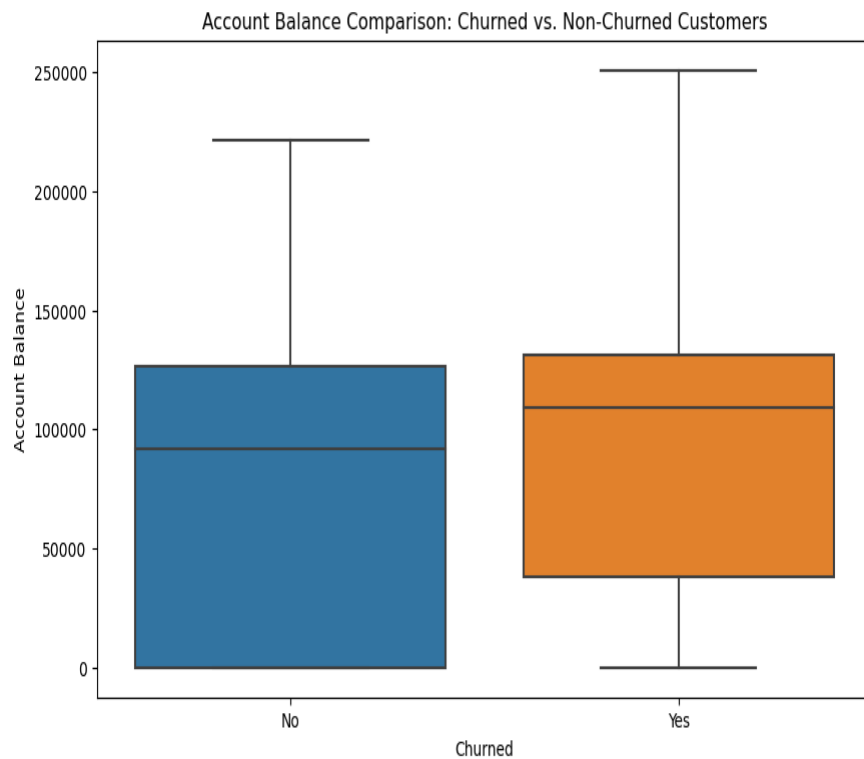
- What is the average account balance of customers?
- Average account balance of customers: Rs76485.89



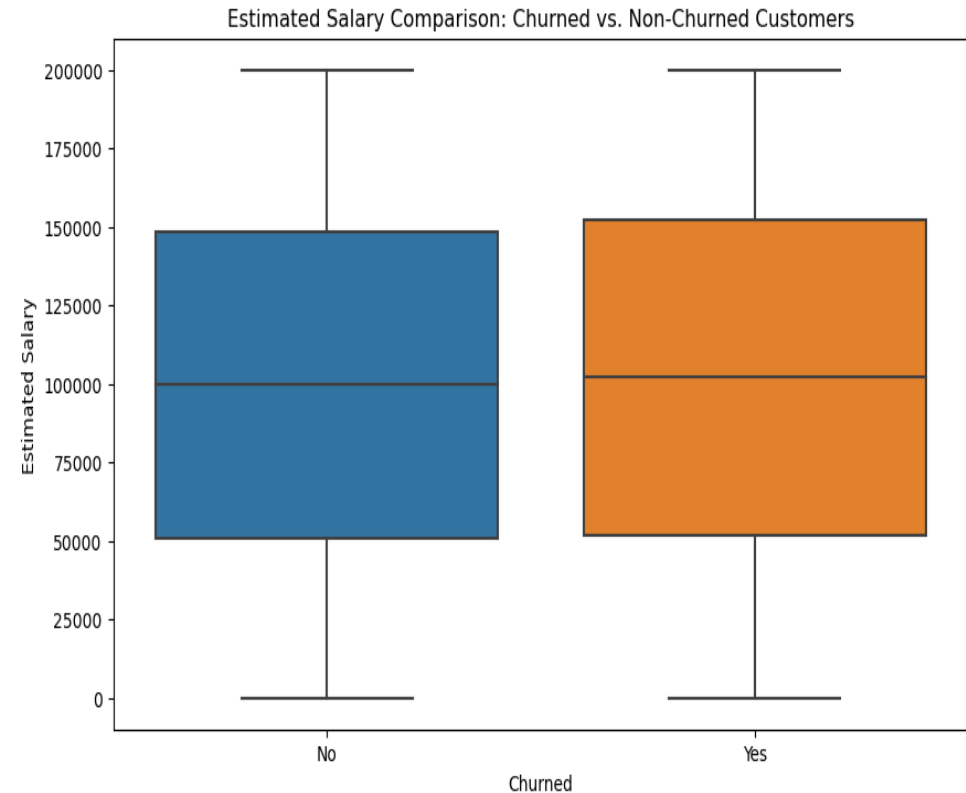
- Compare the financial characteristics of churned vs. non-churned customers.
- Churned (status 1) : 185,588,094.63
- Not - Churned (status 0) : 579,270,798.25

- financial characteristics of churned vs. non-churned customers.

- The churned and non-churned customers have a balance of 1 Lac (slightly below and above respect)
- The min bal for churned and non-churned customers is 50k and 0 resp



- The churned and non-churned customers have an approximate salary of 1 Lac with min being 50K and max around 1.5 Lac
- There is no significant difference between the two groups on basis of their estimated salary.



Office Supplies Data Analysis

- Sales Analysis
- Sales column is obtained by multiplication of price and units column.

```
In [5]: off_data["Sales"] = off_data["Units"]*off_data["Unit Price"]  
off_data["Sales"].head()
```

```
Out[5]: 0    309.38  
1     57.71  
2    686.95  
3   1619.19  
4   1005.90  
Name: Sales, dtype: float64
```


- What are the total sales for each product category?

ITEM	SALES
Binder	9577.65
Desk	1700.00
Pen	2045.22
Pen Set	4169.87
Pencil	2135.14

- Which product category has the highest sales?
- Item with highest sales is **Binder** with sales amount = **Rs. 9577.65**

- Identify the top 3 best-selling products.

ITEM	SALES
Binder	9577.65
Pen Set	4169.87
Pencil	2135.14

Customer Analysis

- Who are the top 10 customers by sales?

```
Cust_by_sale = off_data.groupby("Rep")["Sales"].sum()  
Cust_by_sale
```

```
Rep  
Alex      2812.19  
Bill      1749.87  
James     1283.61  
Matthew   3109.44  
Morgan    1387.77  
Nick       536.75  
Rachel     438.37  
Richard   2363.04  
Smith     1641.43  
Susan     3102.30  
Thomas    1203.11  
Name: Sales, dtype: float64
```

```
top_10_cust = Cust_by_sale.nlargest(10)  
top_10_cust
```

```
Rep  
Matthew    3109.44  
Susan      3102.30  
Alex       2812.19  
Richard    2363.04  
Bill       1749.87  
Smith      1641.43  
Morgan     1387.77  
James      1283.61  
Thomas     1203.11  
Nick        536.75  
Name: Sales, dtype: float64
```

- What is the total number of unique customers?
- There is a total of 11 unique customers.
- Analyze customer purchase frequency.

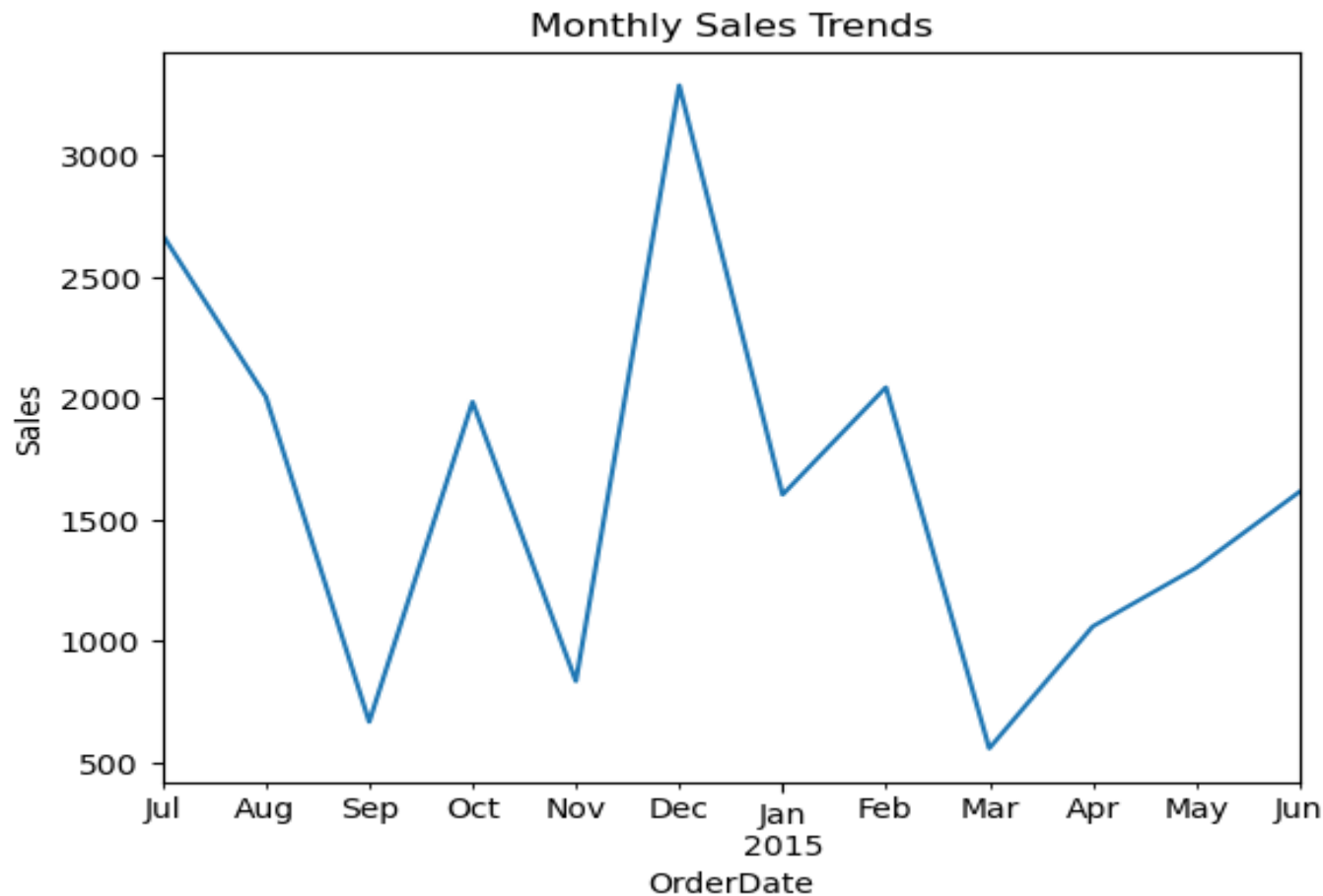
```
cust_purchase_frequency = off_data['Rep'].value_counts()  
cust_purchase_frequency
```

```
Richard    8  
Bill       5  
Alex       5  
Matthew    4  
James      4  
Rachel     4  
Morgan     3  
Susan      3  
Smith      3  
Nick       2  
Thomas     2  
Name: Rep, dtype: int64
```

- Richard is the most frequent customer

Time Series Analysis

- What are the monthly sales trends over the past year?

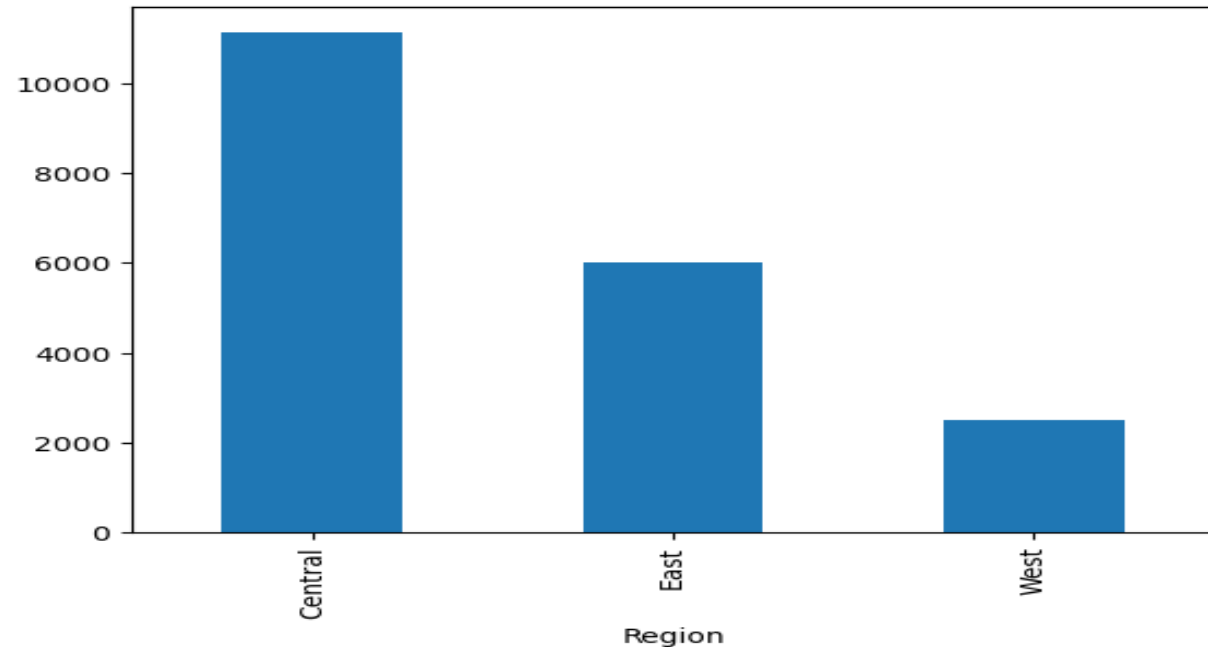


- Fluctuations in the sales amount are quite significant during the business cycle .
- Sep , nov ,March have the least sales
- Dec has the max Sales .
- From Dec onwards sales starts to fall till march and then rose subsequently.
- A cyclic pattern could be observed .

Geographical Analysis

- Which regions generate the most sales?
- What are the sales trends across different regions?

REGION	SALES
Central	11139.07
East	6002.09
West	2486.72



- Central is the region with max sale of amount = Rs. 11139.07

THANKYOU!