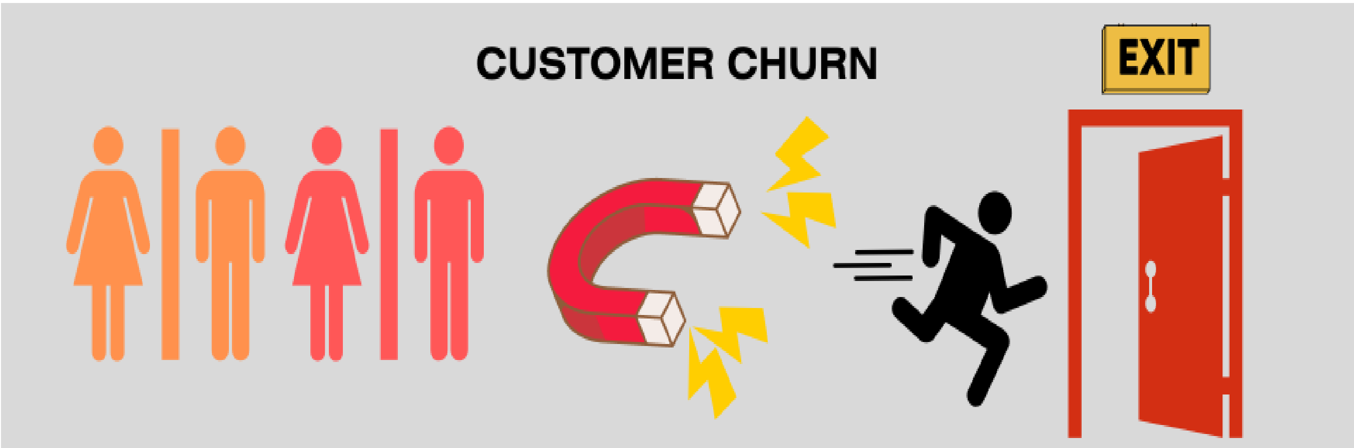
E-commerce Customer Churn Analysis using PostgreSQL

Identifying the reasons why customers are leaving an online e-commerce company.



Customer churn occurs when customers choose to stop engaging with a company, posing a significant challenge for businesses as it affects retention and overall success. E-commerce companies, in particular, struggle to maintain customer loyalty in a fast-paced environment. To combat this issue, it's vital to identify at-risk customers and implement targeted retention strategies. This report analyzes a dataset from an online retail company to uncover insights related to customer churn, guiding decision-making and enabling proactive steps to reduce attrition and build long-term loyalty. By understanding churn factors, companies can enhance customer engagement and offer appealing promotions to mitigate attrition.

Project Approach

To effectively cleanse and prepare the dataset for analysis, I utilized PostgreSQL for data manipulation and Microsoft Power BI for visualizations, which will be presented at the end of this report. The analysis is structured in several phases: data cleaning, data exploration, insights generation, and recommendations, following a systematic methodology to ensure clarity and coherence in findings.

Through this process, the aim is to derive actionable insights that can help improve customer retention strategies and enhance overall satisfaction.

Database and Table Creation

Established the "ecommerce" database and created the required tables for data storage. Subsequently, imported the dataset from a CSV file into the "ecommerce_data" table and verified the import.

Data Output Messages Notifications

	customerid integer	churn integer	tenure integer	preferredlogindevice character varying (50)	citytier integer	warehousetohome integer	preferredpaymentmode character varying (50)	gender character varying (1
1	50014	1	0	Phone	1	15	Credit Card	Male
2	50767	0	10	Phone	1	8	Debit Card	Female
3	50684	0	2	Phone	1	16	Debit Card	Female
4	50709	0	0	Phone	1	16	Debit Card	Male
5	50718	1	0	Phone	1	16	Debit Card	Female
6	50724	0	8	Computer	1	16	UPI	Male
7	50729	0	0	Phone	1	16	Debit Card	Male
8	52982	0	10	Phone	1	10	Cash on Delivery	Male
9	54814	0	24	Computer	1	9	Debit Card	Male
10	50053	1	0	Phone	1	8	Debit Card	Male
11	50054	0	19	Phone	3	6	E wallet	Female
12	51333	0	10	Phone	1	8	Credit Card	Female

Total rows: 1000 of 5630 Query complete 00:00:00.669 Ln 3, Col 1

Data Cleaning

Before starting the analysis, it's crucial to ensure that the dataset is clean and trustworthy. The data cleaning process includes addressing missing values, fixing inconsistencies, and properly formatting the data for analysis. For this project, the dataset was thoroughly cleaned to ensure both accuracy and reliability in the analysis.

1. Finding the total number of customers

```
--/Finding the total number of customers/--
SELECT DISTINCT COUNT(CustomerID) as TotalNumberOfCustomers
FROM ecommerce_data
```

	totalnumberofcustomers bigint
1	5630

In this dataset, there are 5,630 customers.

2. Checking for duplicate rows

```
--/Checking for duplicate rows/--
SELECT CustomerID, COUNT (*) as Count
FROM ecommerce_data
GROUP BY CustomerID
Having COUNT (CustomerID) > 1
```

customerid integer	count bigint

The query returns an empty table, showing there are no duplicate rows.

3. Checking for null values

Let start by finding the null count of columns containing null values.

```
--/Checking for null values/---
SELECT ColumnName, COUNT(*) AS NullCount
FROM (
    SELECT
        CASE
            WHEN CustomerID IS NULL THEN 'CustomerID'
            WHEN Churn IS NULL THEN 'Churn'
             WHEN tenure IS NULL THEN 'tenure'
            WHEN PreferredLoginDevice IS NULL THEN 'PreferredLoginDevice'
            WHEN CityTier IS NULL THEN 'CityTier'
            WHEN PreferredPaymentMode IS NULL THEN 'PreferredPaymentMode'
            WHEN warehousetohome IS NULL THEN 'warehousetohome'
            WHEN Gender IS NULL THEN 'Gender'
            WHEN NumberOfDeviceRegistered IS NULL THEN 'NumberOfDeviceRegistered'
            WHEN PreferredOrderCat IS NULL THEN 'PreferredOrderCat'
            WHEN hourspendonapp IS NULL THEN 'hourspendonapp'
            WHEN SatisfactionScore IS NULL THEN 'SatisfactionScore'
            WHEN MaritalStatus  IS NULL THEN 'MaritalStatus'
            WHEN NumberOfAddress IS NULL THEN 'NumberOfAddress'
            WHEN Complain IS NULL THEN 'Complain'
            WHEN orderamounthikefromlastyear IS NULL THEN 'orderamounthikefromlastyear'
            WHEN couponused IS NULL THEN 'couponused'
            WHEN ordercount IS NULL THEN 'ordercount'
            WHEN daysincelastorder IS NULL THEN 'daysincelastorder'
            WHEN CashbackAmount IS NULL THEN 'CashbackAmount'
        END AS ColumnName
    FROM ecommerce_data
) AS NullCounts
WHERE ColumnName IS NOT NULL
GROUP BY ColumnName;
```

	columnname text	nullcount bigint
1	tenure	264
2	orderamounthikefromlastyear	265
3	hourspendonapp	255
4	daysincelastorder	307
5	ordercount	258
6	warehousetohome	251
7	couponused	256

The above table shows the number of null values present for each column. The columns with null values include CouponUsed, DaysSinceLastOrder, HourSpendOnApp, OrderAmountHikeFromLastYear, OrderCount, Tenure, and WarehouseToHome.

3.1 Handling null values

We will fill in the nulls using the mean value of the corresponding column in order to deal with these null values.

```
--/Handling Null values/--
UPDATE ecommerce_data
SET tenure = (SELECT AVG(tenure) FROM ecommerce_data)
WHERE tenure IS NULL

UPDATE ecommerce_data
SET orderamounthikefromlastyear = (SELECT AVG(orderamounthikefromlastyear) FROM
ecommerce_data)
WHERE orderamounthikefromlastyear IS NULL

UPDATE ecommerce_data
SET Hourspendonapp = (SELECT AVG(Hourspendonapp) FROM ecommerce_data)
WHERE Hourspendonapp IS NULL

UPDATE ecommerce_data
SET WarehouseToHome = (SELECT AVG(WarehouseToHome) FROM ecommerce_data)
WHERE WarehouseToHome IS NULL

UPDATE ecommerce_data
SET couponused = (SELECT AVG(couponused) FROM ecommerce_data)
WHERE couponused IS NULL

UPDATE ecommerce_data
SET ordercount = (SELECT AVG(ordercount) FROM ecommerce_data)
WHERE ordercount IS NULL

UPDATE ecommerce_data
SET daysincelastorder = (SELECT AVG(daysincelastorder) FROM ecommerce_data)
WHERE daysincelastorder IS NULL
```

Now that the null values have been replaced with their mean, there are no more null values left.

4. Creating a new column from an already existing "churn" column

The churn column in the dataset comprised values of 0 and 1 value, which we discovered while examining it. When a customer's value is 0, it indicates customer did not churn, but when it is 1, customer churned. We'll add a new column named "Churn_Status" that indicates "Stayed" when a customer did not churn and "Churned" when a customer did. This will help because it can be confusing to recall which represents what.

```
ALTER TABLE ecommerce_data
ADD Churn_Status VARCHAR(50)
```

```
UPDATE ecommerce_data
SET Churn_Status =
  CASE
    WHEN Churn = 1 THEN 'Churned'
    WHEN Churn = 0 THEN 'Stayed'
END
```

	customerid integer	churn_status character varying (50)
154	51481	Stayed
155	51482	Churned
156	51538	Churned
157	51541	Stayed
158	51542	Stayed

5. Creating a new column from an already existing “complain” column

We observed that the complain column also had 0 and 1, which is really similar to what we accomplished in the part above. A value of '0' indicates that the consumer did not file any complaints, but a value of '1' indicates that they did. To make things clearer, we're going to add a new column called "Complain_Received" that will indicate "Yes" when a client has complained and "No" when they haven't.

```
--/Creating a new column from an already existing “complain” column/--

ALTER TABLE ecommerce_data
ADD Complain_Recieved VARCHAR(10)

UPDATE ecommerce_data
SET Complain_Recieved =
  CASE
    WHEN Complain = 1 THEN 'Yes'
    WHEN Complain= 0 THEN 'No'
END
```

	customerid integer	complain integer	complain_recieved character varying (10)
1	50001	1	Yes
2	50004	0	No
3	50006	1	Yes
4	50012	1	Yes
5	50013	1	Yes
6	50014	1	Yes
7	50053	0	No
8	50054	0	No

6. Checking values in each column for correctness and accuracy

After going through each column, we noticed some redundant values in some columns and a wrongly entered value. This will be explored and fixed.

6.1 Fixing redundancy in "PreferredLoginDevice" Column

Before

```
--/6. Fixing redundancy in "PreferedLoginDevice" Column/--
SELECT DISTINCT PreferredLoginDevice
FROM ecommerce_data
```

	preferredlogindevice character varying (50)
1	Mobile Phone
2	Computer
3	Phone

Although the terms "phone" and "mobile phone" are displayed in the same column, they have the same meaning. Therefore, we will use a phone in place of the mobile phone.

After

```
UPDATE ecommerce_data
SET PreferredLoginDevice = 'Phone'
WHERE PreferredLoginDevice = 'Mobile Phone'
```

	preferredlogindevice character varying (50)
1	Computer
2	Phone

The result above shows the redundancy has been fixed.

6.2 Fixing redundancy in "PreferedOrderCat" Column

Before

```
select Distinct PreferredOrderCat
FROM ecommerce_data
```

	preferredordercat character varying (50)
1	Grocery
2	Mobile
3	Mobile Phone
4	Others
5	Laptop & Accessory
6	Fashion

Notice that mobile phone and phone appear in the column, but they mean the same thing. So we will replace the mobile with mobile phone as the category name.

After

```
UPDATE ecommerce_data
SET PreferredOrderCat = 'Mobile Phone'
WHERE PreferredOrderCat = ' Mobile'
```

	preferredordercat character varying (50)
1	Grocery
2	Mobile Phone
3	Others
4	Laptop & Accessory
5	Mobile Phone
6	Fashion

6.3 Fixing redundancy in "Preferred PaymentMode" Column

Before

```
select Distinct PreferredPaymentMode
FROM ecommerce_data
```

	preferredpaymentmode character varying (50)
1	UPI
2	E wallet
3	CC
4	Debit Card
5	Cash on Delivery
6	COD
7	Credit Card

Credit card and CC are interchangeable, as are cash on delivery and COD. switching from CC to credit card and COD to delivery card.

After

```
UPDATE ecommerce_data
SET PreferredPaymentMode = 'Cash on Delivery'
WHERE PreferredPaymentMode = 'COD'

UPDATE ecommerce_data
SET PreferredPaymentMode = 'Credit Card'
WHERE PreferredPaymentMode = 'CC'
```

	preferredpaymentmode character varying (50)
1	UPI
2	E wallet
3	Debit Card
4	Cash on Delivery
5	Credit Card

6.4 Fixing wrongly entered values in "WarehouseToHome" column

Before

```
SELECT DISTINCT WarehouseToHome
FROM ecommerce_data
ORDER BY WarehouseToHome DESC
```

	warehousetohome integer
1	127
2	126
3	36
4	35
5	34
6	33
7	32

Take note of values 126 and 127, which are unquestionably anomalous and were probably input incorrectly. In order to correct this, we will alter the values to 26 and 27 so that they are within the range of the values in this column.

After

```
UPDATE ecommerce_data
```

```
SET warehousetohome = '27'
WHERE warehousetohome = '127'

UPDATE ecommerce_data
SET warehousetohome = '26'
WHERE warehousetohome = '126'
```

	warehousetohome integer
1	36
2	35
3	34
4	33
5	32
6	31
7	30

Our data has been cleaned and is now ready to be explored for insight generation.

Data Exploration

Answering business questions.

1. What is the overall customer churn rate?

```
--/What is the overall customer churn rate?/--
SELECT
 COUNT(*) AS TotalNumberofCustomers,
 COUNT(CASE WHEN Churn_Status = 'Churned' THEN 1 END) AS TotalNumberofChurnedCustomers,
 CAST(COUNT(CASE WHEN Churn_Status = 'Churned' THEN 1 END) * 100.0 / COUNT(*) AS
DECIMAL(10,2)) AS ChurnRate
FROM ecommerce_data;
```

	totalnumberofcustomers bigint	totalnumberofchurnedcustomers bigint	churnrate numeric (10,2)
1	5630	948	16.84

16.84% of customers in the dataset have ceased to be associated with the company, indicating a large chunk of the customer base has left.

2. What is the typical tenure for churned customers?

Using the CASE statement, we will first build a new column called "Tenure_Range" that will group the customer tenure into four categories: six months, one year, two years, and more than two years.

```
--/Update typical tenure_range for churned customers?/--
ALTER TABLE ecommerce_data
ADD Tenure_Range VARCHAR(50)

UPDATE ecommerce_data
SET Tenure_Range =
CASE
    WHEN tenure <= 6 THEN '6 Months'
    WHEN tenure > 6 AND tenure <= 12 THEN '1 Year'
    WHEN tenure > 12 AND tenure <= 24 THEN '2 Years'
    WHEN tenure > 24 THEN 'more than 2 years'
END
```

Finding typical tenure for churned customers

```
--/Finding typical tenure for churned customers/--
SELECT Tenure_Range,
    COUNT(*) AS TotalCustomers,
    SUM(churn) AS ChurnedCustomers,
    CAST(SUM (churn) * 1.0 / COUNT(*) * 100 AS DECIMAL(10,2)) AS ChurnRate
FROM ecommerce_data
GROUP BY Tenure_Range
ORDER BY ChurnRate DESC
```

	tenure_range character varying (50)	totalcustomers bigint	churnedcustomers bigint	churnrate numeric (10,2)
1	6 Months	2150	697	32.42
2	1 Year	1584	156	9.85
3	2 Years	1467	95	6.48
4	more than 2 years	429	0	0.00

This indicates that customers with longer tenure with the company—more than two years in this case—have demonstrated a reduced likelihood for turnover as compared to customers in shorter tenure categories.

3. How does the churn rate vary based on the preferred login device?

```
--/How does the churn rate vary based on the preferred login device?/--

SELECT preferredlogindevice,
COUNT(*) AS ChurnedCustomers,
CAST(COUNT(*) * 1.0 / 948 * 100 AS DECIMAL(10,2)) AS ChurnRate
FROM ecommerce_data
WHERE churn = 1
GROUP BY preferredlogindevice
```

```
ORDER BY ChurnRate DESC
```

	preferredlogindevice character varying (50)	churnedcustomers bigint	churnrate numeric (10,2)
1	Phone	624	65.82
2	Computer	324	34.18

The turnover rate for customers who prefer to log in using a phone is somewhat greater than that of those who prefer to log in using their computer. This could mean that users of the platform through a phone have different usage habits, tastes, or experiences, all of which increase the risk of churn.

4. What is the distribution of customers across different city tiers?

```
--/What is the distribution of customers across different city tiers?/---

SELECT  citytier,
COUNT(*) AS ChurnedCustomers,
CAST(COUNT(*) * 1.0 / 948 * 100 AS DECIMAL(10,2)) AS ChurnRate
FROM ecommerce_data
WHERE churn = 1
GROUP BY citytier
ORDER BY ChurnRate DESC
```

	citytier integer	churnedcustomers bigint	churnrate numeric (10,2)
1	1	532	56.12
2	3	368	38.82
3	2	48	5.06

A major metropolitan area classified as:

- City Tier 1 usually has the best infrastructure and economic development. These cities often contain the largest populations as well as important business and commercial hubs.
- City Tier 2 cities are regarded as lesser or secondary urban centres, compared to Tier 1 cities.
- City Tier 3 typically refers to smaller cities or villages with less established infrastructure and a smaller population than Tier 1 and 2 cities.

The outcome implies that customer attrition rates are influenced by the city tier. When compared to Tier 2 and Tier 3 cities, Tier 1 cities have a comparatively Higher turnover rate. Numerous variables, including competition, consumer preferences, and the availability of alternatives in different city levels, could be to blame for this.

5. Is there any correlation between the warehouse-to-home distance and customer churn?

In order to answer this question, we will create a new column called "WarehouseToHome_Range" that groups the distance into very close, close, moderate, and far using the CASE statement.

```
--/Correlation between warehouse to home and churn rate./--

ALTER TABLE ecommerce_data
ADD warehousetohome_range VARCHAR(50)

UPDATE ecommerce_data
SET warehousetohome_range =
CASE
    WHEN warehousetohome <= 10 THEN 'Very close distance'
    WHEN warehousetohome > 10 AND warehousetohome <= 20 THEN 'Close distance'
    WHEN warehousetohome > 20 AND warehousetohome <= 30 THEN 'Moderate distance'
    WHEN warehousetohome > 30 THEN 'Far distance'
END
```

Finding a correlation between warehouse to home and churn rate.

```
SELECT warehousetohome_range,
  COUNT(*) AS TotalCustomer,
  SUM(Churn) AS CustomerChurn,
  CAST(SUM(Churn) * 1.0 /COUNT(*) * 100 AS DECIMAL(10,2)) AS Churnrate
FROM ecommerce_data
GROUP BY warehousetohome_range
ORDER BY Churnrate DESC
```

	warehousetohome_range character varying (50)	totalcustomer bigint	customerchurn bigint	churnrate numeric (10,2)
1	Far distance	469	98	20.90
2	Moderate distance	874	176	20.14
3	Close distance	2318	408	17.60
4	Very close distance	1969	266	13.51

Customer turnover rates appear to be somewhat impacted by the distance between the warehouse and the customer's residence. Clients who live closer to the warehouse typically have lower churn rates, but those who live farther away are more likely to experience turnover. This implies that elements like convenience, shipping costs, and delivery timeframes might have an impact on keeping customers.

6. Which is the most preferred payment mode among churned customers?

```
--/Which is the most preferred payment mode among churned customers?/--

SELECT preferredpaymentmode,
  COUNT(*) AS TotalCustomer,
  SUM(churn) AS CustomerChurn,
  CAST(SUM(Churn) * 1.0 /COUNT(*) * 100 AS DECIMAL(10,2))AS Churnrate
FROM ecommerce_data
```

```
GROUP BY preferredpaymentmode
ORDER BY Churnrate DESC
```

	preferredpaymentmode character varying (50)	totalcustomer bigint	customerchurn bigint	churnrate numeric (10,2)
1	Cash on Delivery	514	128	24.90
2	E wallet	614	140	22.80
3	UPI	414	72	17.39
4	Debit Card	2314	356	15.38
5	Credit Card	1774	252	14.21

Customer turnover rates appear to be somewhat impacted by the preferred payment method. Payment methods with greater churn rates, such "Cash on Delivery" and "E-wallet," suggest that users of these methods are more likely to stop using them. However, the churn rates of payment methods like "Credit Card" and "Debit Card" are comparatively lower.

7. Is there any difference in churn rate between male and female customers & order count?

```
--/Is there any difference in churn rate between male and female customers?/--

SELECT gender,
  COUNT(*) AS total_customers,
  SUM(churn) AS total_churn,
  SUM(ordercount) as sum_order,
  CAST(SUM(churn) * 1.0 /COUNT(*) * 100.0 AS DECIMAL(10,2)) AS Churnrate
FROM Ecommerce_data
GROUP BY gender
ORDER BY Churnrate DESC
```

	gender character varying (10)	total_customers bigint	total_churn bigint	sum_order bigint	churnrate numeric (10,2)
1	Male	3384	600	9936	17.73
2	Female	2246	348	6997	15.49

```
--/churn rate for churn customer only/--
SELECT  gender,
COUNT(*) AS ChurnedCustomers,
CAST(COUNT(*) * 1.0 / 948 * 100 AS DECIMAL(10,2)) AS ChurnRate
FROM ecommerce_data
WHERE churn = 1
GROUP BY gender
ORDER BY ChurnRate DESC
```

	gender character varying (10)	churnedcustomers bigint	churnrate numeric (10,2)
1	Male	600	63.29
2	Female	348	36.71

The table shows that male customers have a higher churn rate compared to female customers. Males placed 9,936 orders, while females placed 6,997. Despite being more active, males experience slightly more churn than females.

8. How does the average time spent on the app differ for churned and non-churned customers?

```
--/Average time spent on the app differ for churned and non-churned customers?/--

SELECT Churn_status,ROUND(AVG(hourspendonapp),2) AS avg_hourspendonapp
FROM ecommerce_data
GROUP BY churn_status
```

	churn_status character varying (50)	avg_hourspendonapp numeric
1	Stayed	2.93
2	Churned	2.96

Customers who have stayed and those who have churned had similar average hours spent on the app, with a slight variation in points; this suggests that average app usage time is not a significant factor in differentiating between the two groups of users.

9. Does the number of registered devices impact the likelihood of churn?

```
--/Number of registered devices impact the likelihood of churn?/--

SELECT numberofdeviceregistered,
   COUNT(*) AS TotalCustomer,
   SUM(Churn) AS CustomerChurn,
   CAST(SUM(Churn) * 1.0 /COUNT(*) * 100 AS DECIMAL(10,2)) AS Churnrate
FROM ecommerce_data
GROUP BY numberofdeviceregistered
ORDER BY Churnrate DESC
```

	numberofdeviceregistered integer	totalcustomer bigint	customerchurn bigint	churnrate numeric (10,2)
1	6	162	56	34.57
2	5	881	198	22.47
3	4	2377	392	16.49
4	3	1699	254	14.95
5	2	276	26	9.42
6	1	235	22	9.36

There appears to be a relationship between the quantity of devices that users register and the probability of churn. Higher turnover rates are seen in customers with more registered devices—for example, six or five. Customers with fewer registered devices, like two or one, however, exhibit comparatively lower churn rates.

10. Which order category is most preferred among churned customers?

```sql
--/Which order category is most preferred among churned customers?/--

SELECT preferredordercat,
   COUNT(*) AS TotalCustomer,
   SUM(Churn) AS CustomerChurn,
   CAST(SUM(Churn) * 1.0 /COUNT(*) * 100 AS DECIMAL(10,2)) AS Churnrate
FROM ecommerce_data
GROUP BY preferredordercat
ORDER BY Churnrate DESC
```

	preferredordercat character varying (50)	totalcustomer bigint	customerchurn bigint	churnrate numeric (10,2)
1	Mobile Phone	2080	570	27.40
2	Fashion	826	128	15.50
3	Laptop & Accessory	2050	210	10.24
4	Others	264	20	7.58
5	Grocery	410	20	4.88

```sql
--/churn rate for churn custumer only/--
SELECT preferredordercat,
COUNT(*) AS ChurnedCustomers,
CAST(COUNT(*) * 1.0 / 948 * 100 AS DECIMAL(10,2)) AS ChurnRate
FROM ecommerce_data
WHERE churn = 1
GROUP BY preferredordercat
ORDER BY ChurnRate DESC
```

	preferredordercat character varying (50)	churnedcustomers bigint	churnrate numeric (10,2)
1	Mobile Phone	570	60.13
2	Laptop & Accessory	210	22.15
3	Fashion	128	13.50
4	Grocery	20	2.11
5	Others	20	2.11

The highest rate of customer churn is shown among customers who order primarily items in the "Mobile Phone" category; this suggests that specific retention measures may be necessary for this particular group. The "Grocery" & “Others” category, on the other hand, has the lowest turnover rate, which may indicate that customers in this area are more loyal and retain customers.

11. Is there any relationship between customer satisfaction scores and churn?

```sql
--/Is there any relationship between customer satisfaction scores and churn?/--

SELECT satisfactionscore,
   COUNT(*) AS TotalCustomer,
   SUM(Churn) AS CustomerChurn,
   CAST(SUM(Churn) * 1.0 /COUNT(*) * 100 AS DECIMAL(10,2)) AS Churnrate
FROM ecommerce_data
GROUP BY satisfactionscore
ORDER BY Churnrate DESC
```

	satisfactionscore integer	totalcustomer bigint	customerchurn bigint	churnrate numeric (10,2)
1	5	1108	264	23.83
2	3	1698	292	17.20
3	4	1074	184	17.13
4	2	586	74	12.63
5	1	1164	134	11.51

In comparison to other satisfaction score categories, the finding shows that consumers with greater satisfaction scores—especially those who gave themselves a 5-star rating—have a comparatively higher churn rate. As a result, proactive customer retention techniques are crucial for keeping customers happy at all satisfaction levels. This implies that even extremely satisfied customers may still leave.

12. Does the marital status of customers influence churn behavior?

```sql
--/Does the marital status of customers influence churn behavior?/--

SELECT maritalstatus,
   COUNT(*) AS TotalCustomer,
   SUM(Churn) AS CustomerChurn,
   CAST(SUM(Churn) * 1.0 /COUNT(*) * 100 AS DECIMAL(10,2)) AS Churnrate
FROM ecommerce_data
GROUP BY maritalstatus
ORDER BY Churnrate DESC
```

	maritalstatus character varying (25)	totalcustomer bigint	customerchurn bigint	churnrate numeric (10,2)
1	Single	1796	480	26.73
2	Divorced	848	124	14.62
3	Married	2986	344	11.52

In comparison to customers in other marital categories, single customers have the highest rate of churn. This suggests that customers who are single might be more prone to break off their business relationship with the company. Married customers, on the other hand, have the lowest customer turnover rate, followed by divorced customers.

13. How many addresses do churned customers have on average?

```sql
--/How many addresses do churned customers have on average?/--

SELECT ROUND(AVG(numberofaddress),2) as avg_churned_num_address
FROM ecommerce_data
WHERE churn_status = 'Churned'
```

	avg_churned_num_address numeric
1	4.47

On average, customers who churned had four addresses associated with their accounts.

14. Do customer complaints influence churned behavior?

```sql
--/Do customer complaints influence churned behavior?/--

SELECT complain_recieved,
   COUNT(*) AS TotalCustomer,
   SUM(Churn) AS CustomerChurn,
   CAST(SUM(Churn) * 1.0 /COUNT(*) * 100 AS DECIMAL(10,2)) AS Churnrate
FROM ecommerce_data
GROUP BY complain_recieved
ORDER BY Churnrate DESC
```

	complain_recieved character varying (10)	totalcustomer bigint	customerchurn bigint	churnrate numeric (10,2)
1	Yes	1604	508	31.67
2	No	4026	440	10.93

It's important to address and resolve customer problems, as seen by the higher percentage of customers who stopped using the company's services filing complaints. This is essential for reducing the amount of departing clients and fostering enduring loyalty.

15. How does the use of coupons differ between churned and non-churned customers?

```sql
--/Used of coupons differ between churned and non-churned customers, How?/--

select churn_status, sum(couponused) as sum_couponused
FROM Ecommerce_data
GROUP BY churn_status
```

	churn_status character varying (50)	sum_couponused bigint
1	Stayed	8292
2	Churned	1630

The greater use of coupons by returning customers suggests a higher degree of involvement and commitment to the business.

16. What is the average number of days since the last order for churned customers?

```
--/Average number of days since the last order for churned customers?/--

SELECT Round(AVG(daysincelastorder)) as avg_day_last_order
FROM Ecommerce_data
WHERE churn_status= 'Churned'
```

	avg_day_last_order numeric
1	3

The average short time elapsed between a churned client's last order and their current status suggests that the customer quit doing business with the company not too long ago.

17. Order amount hike from last year

```
--/Orderamounthike from last year/--

SELECT SUM(orderamounthikefromlastyear) AS total_order_amount
FROM Ecommerce_data
```

	total_order_amount bigint
1	88513

18. Is there any correlation between cashback amount and churn rate?

Using the CASE statement, we will first create a new column called "CashbackAmount_Range" that will classify the cashback amount into four categories: low (less than 100), moderate (between 100 and 200), high (between 200 and 300), and very high (more than 300).

```
--/Update cashbackamount_range for churned customers?/--

ALTER TABLE Ecommerce_data
ADD cashbackamount_range VARCHAR(50)

UPDATE Ecommerce_data
```

```
SET cashbackamount_range =
 CASE
    WHEN cashbackamount <= 100 THEN 'Low Cashback Amount'
    WHEN cashbackamount > 100 AND cashbackamount <= 200 THEN 'Moderate Cashback Amount'
    WHEN cashbackamount > 200 AND cashbackamount <= 300 THEN 'High Cashback Amount'
    WHEN cashbackamount > 300 THEN 'Very High Cashback Amount'
END
```

Finding the correlation between cashback amount range and churned rate

```
--/Is there any correlation between cashback amount and churn rate?/--
SELECT cashbackamount_range,
 COUNT(*) AS total_customers,
 SUM(churn) AS total_churn,
 CAST(SUM(churn) * 1.0 /COUNT(*) * 100.0 AS DECIMAL(10,2)) AS Churnrate
FROM Ecommerce_data
GROUP BY cashbackamount_range
ORDER BY Churnrate DESC
```

	cashbackamount_range character varying (50)	total_customers bigint	total_churn bigint	churnrate numeric (10,2)
1	Moderate Cashback Amount	4305	814	18.91
2	High Cashback Amount	1157	124	10.72
3	Very High Cashback Amount	156	10	6.41
4	Low Cashback Amount	12	0	0.00

The churn rate was shown to be lower for customers who received extremely high and greater cashback amounts than for those who received moderate cashback amounts. 100% of the customers who received smaller payback amounts were kept as long as possible. This implies that providing larger payback payments can enhance customers loyalty and lower churn rate.

Summary of Insights

- The dataset contains 5,630 customers, with a churn rate of 16.84%.
- Phone users have slightly higher churn rates than computer users, and Tier 1 cities show Higher churn than Tier 2 and Tier 3 cities.
- Proximity to warehouses is linked to lower churn rates, emphasizing the importance of logistics.
- Payment methods matter: "Cash on Delivery" and "E-wallet" users have higher churn compared to "Credit Card" and "Debit Card" users.
- Longer customer tenure is associated with lower churn, while single customers have the highest churn rates.
- More registered devices correlate with higher churn, indicating a need for consistent experiences.
- The "Mobile Phone" category has the highest churn, and surprisingly, highly satisfied customers (rating 5) also have higher churn.

- Churned customers average four associated addresses and often have complaints, suggesting a need for better issue resolution.
- Non-churned customers use coupons more, highlighting the effectiveness of loyalty programs.
- Churned customers often have a short time since their last order, indicating recent disengagement.
- Higher cashback amounts lead to lower churn rates, suggesting that increased incentives enhance loyalty.

Business Recommendations

1. Targeted Retention Campaigns: Focus on retaining new customers early by offering personalized experiences, as short-tenure customers are more likely to churn.

2. Enhance Phone Experience: Given that phone users show higher churn rates, continue improving the phone platform for ease of use.

3. Improve Logistics: Customers who live farther from warehouses have higher churn rates. Enhance delivery speed or offer perks like free shipping to customers in more distant locations.

4. Optimize Payment Options: Consider encouraging users to switch to more secure and seamless payment methods like credit or debit cards, which show lower churn rates compared to "Cash on Delivery" and "E-wallet" users.

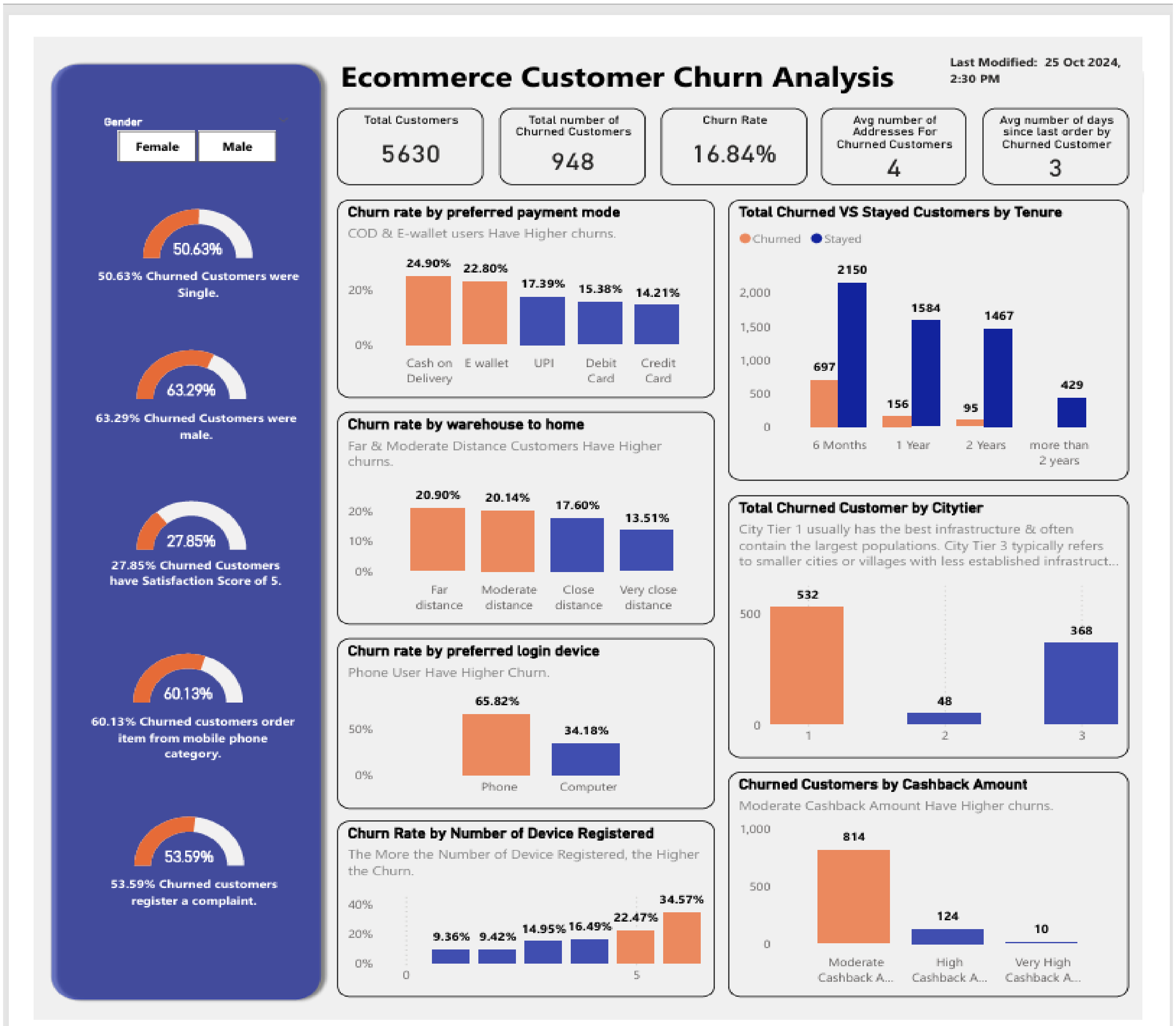
5. Complaint Resolution: Addressing complaints quickly and thoroughly is key to reducing churn. Implement a robust feedback and resolution system to handle customer issues more effectively.

6. Personalized Promotions: Non-churned customers tend to use more coupons, so implementing loyalty programs with personalized offers can help maintain customer engagement.

7. Incentivize Longer Engagement: Offer rewards or discounts to customers who reach certain tenure milestones to encourage long-term loyalty.

8. Focus on High-Churn Product Categories: Products like "Mobile Phones" show high churn rates. Develop specific retention strategies for these categories, such as bundling or exclusive deals.

By focusing on these strategies, the company can reduce churn and enhance customer loyalty in key areas.


Ecommerce Customer Churn Analysis
Last Modified: 25 Oct 2024, 2:30 PM
Gender
Female
Male
Total Customers 5630
Total number of Churned Customers 948
Churn Rate 16.84%
Avg number of Addresses For Churned Customers 4
Avg number of days since last order by Churned Customer 3
50.63%
50.63% Churned Customers were Single.
63.29%
63.29% Churned Customers were male.
27.85%
27.85% Churned Customers have Satisfaction Score of 5.
60.13%
60.13% Churned customers order item from mobile phone category.
53.59%
53.59% Churned customers register a complaint.
Churn rate by preferred payment mode
COD & E-wallet users Have Higher churns.
24.90% 22.80% 17.39% 15.38% 14.21%
20% 0%
Cash on Delivery
E wallet
UPI
Debit Card
Credit Card
Total Churned VS Stayed Customers by Tenure
Churned
Stayed
2150 1584 1467 429 697 156 95
2,000 1,500 1,000 500 0
6 Months
1 Year
2 Years
more than 2 years
Churn rate by warehouse to home
Far & Moderate Distance Customers Have Higher churns.
20.90% 20.14% 17.60% 13.51%
20% 10% 0%
Far distance
Moderate distance
Close distance
Very close distance
Total Churned Customer by Citytier
City Tier 1 usually has the best infrastructure & often contain the largest populations. City Tier 3 typically refers to smaller cities or villages with less established infrastruct...
532 48 368
500 0
1 2 3
Churn rate by preferred login device
Phone User Have Higher Churn.
65.82% 34.18%
50% 0%
Phone
Computer
Churned Customers by Cashback Amount
Moderate Cashback Amount Have Higher churns.
814 124 10
1,000 500 0
Moderate Cashback A...
High Cashback A...
Very High Cashback A...
Churn Rate by Number of Device Registered
The More the Number of Device Registered, the Higher the Churn.
9.36% 9.42% 14.95% 16.49% 22.47% 34.57%
40% 20% 0%
0 5