



# Customer Churn Analysis



2023  
Report

Name: Saima naz hussain



E-mail: saimanaz6777@gmail.com



## *Contents*

**01**

***Data Cleaning and processing***

**02**

***Exploratory Data Analysis (EDA)***

**03**

***Visualization***

**04**

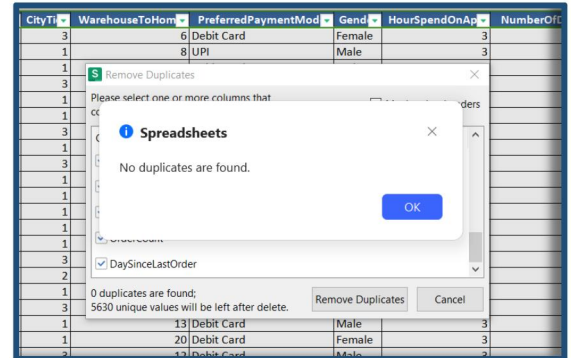
***Recommendations***

## Data Cleaning and processing

A

### Checking for duplicates:

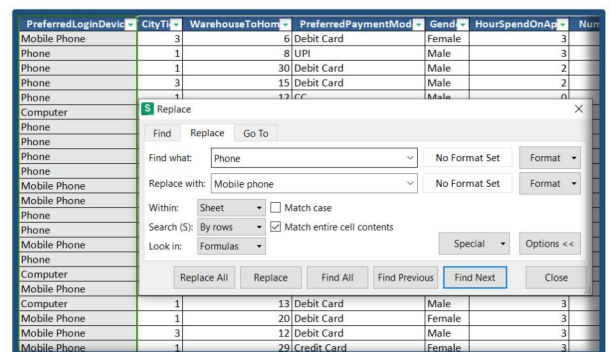
Upon inspection, it has been verified that there are no duplicate entries within the data table. Furthermore, a thorough examination of the unique values comprising the primary key has also confirmed the absence of any duplicates.



B

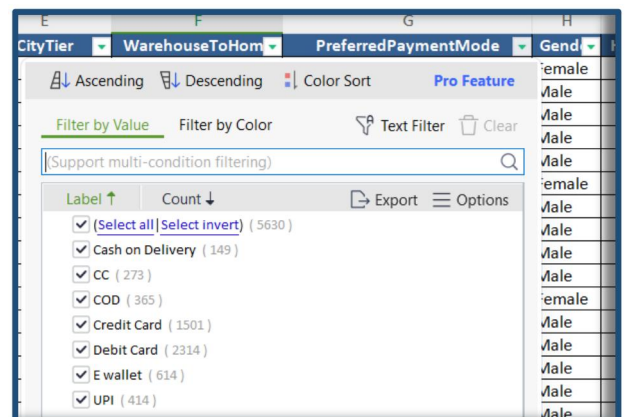
### Discrepancy in data set:

We discovered a **discrepancy in 1,231 values** within the **"Preferred Login Device"** column. In these cases, the entry "Phone" has been used interchangeably with "Mobile Phone." To ensure consistency, we will replace all instances of "Phone" with "Mobile Phone."

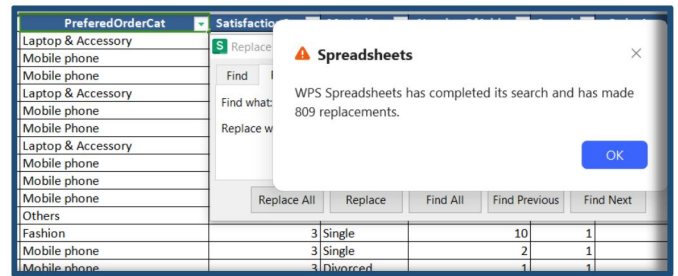


We've identified an inconsistency in the **"Preferred Payment Mode"** column. Specifically, the payment method "Cash on Delivery" has been recorded as both "COD" and "Cash on Delivery," creating a discrepancy. To address this, we will unify these entries by converting the **365 occurrences of "COD" to "Cash on Delivery."**

Additionally, "Credit Card" has been represented as "CC" in 273 instances. To maintain consistency, we will convert these **273 occurrences of "CC" to "Credit Card."**



We've come across an inconsistency in the "Preferred Order Cart" column, where both "Mobile Phone" and "Mobile" have been used. To ensure uniformity, we will convert the **809** occurrences of "Mobile" to "Mobile Phone."



## Changing data type:

Column name	Data type before	Data type after
Customer ID	General	General
Churn	General	Number
Tenure	General	Number
Preferred Login Device	General	Text
CityTier	General	Number
Warehouse To Home	General	Number
Preferred Payment Mode	General	Text
Gender	General	Text
Hour Spend On App	General	Number
Number Of Device Registered	General	Number
Preferred Order Cat	General	Text
Satisfaction Score	General	Number
Marital Status	General	Text
Number Of Address	General	Number
Complain	General	Number
Order Amount Hike Fromlast Year	General	Error ( percentage )
Coupon Used	General	Number
Order Count	General	Number
Day Since Last Order	General	Number

**Error explanation:** The "Order Amount Hike from Last Year" column's data type needs to be changed to a percentage, as specified in the data dictionary. However, there's an issue where the column is currently displayed in a general format. Directly converting it into a percentage format would incorrectly multiply the values by 100.

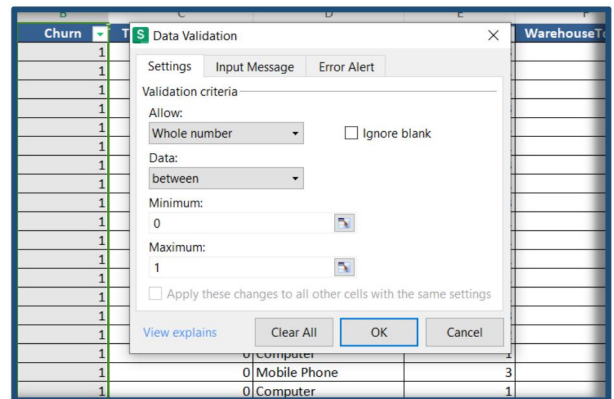
To address this, I've decided to take a two-step approach. First, I will divide the entire column

by 100, effectively converting the values into their decimal form. Then, I will use an automatic correction process to adjust the data type to the desired percentage format. This method ensures that the values are properly scaled to percentages without distorting their meaning.

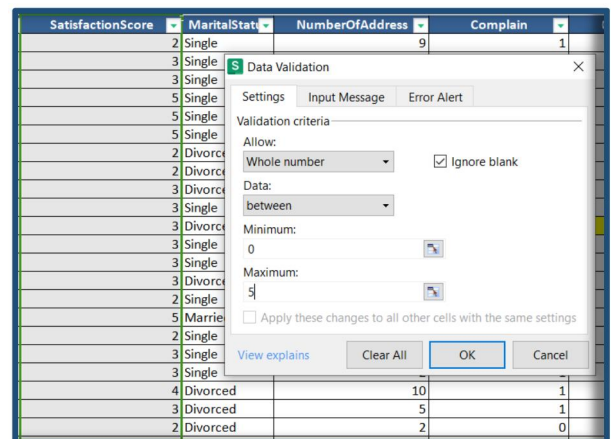
## D

### Data validation:

In the **"churn"** column, I have limited the data to contain only the values "1" and "0" as per the data dictionary specifications. In this context, "1" represents "yes," indicating that churn has occurred, while "0" represents "no," indicating that there has been no churn. This restriction ensures that the data conforms precisely to the predefined categories, facilitating accurate analysis and interpretation in line with the established data dictionary guidelines.



In the **"satisfaction"** column, I've restricted the data to a **range of values from 0 to 5**, in accordance with the data dictionary. This means that the ratings for satisfaction can only fall within this specific range. To clarify, a rating of "5" indicates the highest level of satisfaction, while a rating of "0" signifies the lowest level of satisfaction. By adhering to these defined criteria, we ensure that the data accurately reflects the satisfaction levels as intended in the data dictionary, making it easier to analyze and understand.



## E

### Blank values

I've identified blanks in four columns:

1. "Warehouse to Home": There are 251 missing values.
2. "Order Amount Hike from Last Year": There are 265 missing values.
3. "Coupon Used": There are 265 missing values.
4. "Order Count": There are 258 missing values.

To address this, we'll tackle each column one by one and decide how to handle these missing values for each specific situation.

## Order Count:

As specified in our data dictionary, "Coupon Used" represents the total number of coupons used in the last month, while "Order Count" represents the total number of orders placed in the last month. Since coupons are used when orders are placed, it's logical to expect that "Order Count" should always be greater than or equal to "Coupon Used" for each record.

## Order Count $\geq$ Coupon Used

To address the missing values in the "Order Count" column, I've decided on the following approach:

If there's a blank in the "Order Count" column, I will replace it with a value equal to the corresponding "Coupon Used" value plus 1. The reason for adding 1 is that we have data indicating the number of days since the last order for each customer, and if this value is less than 30, it suggests that a customer has placed an order in the last month.

However, I will only add 1 if "Coupon Used" is not zero. This condition ensures that we don't create artificial orders when there have been no coupon usages, as it indicates that no orders were placed.

This adjustment ensures that, at the very least, the "Order Count" will match the minimum correct value, aligning with the relationship between **'orders count'**, **'coupon usage'** and **'Day Since Last Order'** in the last month.

`=IF(ISBLANK(R2),IF(Q2=0,Q2+1,Q2),R2)`

fx   =IF(ISBLANK(R2),IF(Q2=0,Q2+1,Q2),R2)			
Q	R	S	U
CouponUsed	OrderCount	DaySinceLastOrder	OrderCount2
0		0	1
0		0	1
0		9	1
1		2	1
0		10	1
0		2	1
2		8	2
1		4	1



## Coupon Used:

As per our data dictionary, "Coupon Used" represents the total number of coupons used in the last month, and "Order Count" represents the total number of orders placed in the last month. This confirms that both data columns pertain to the previous month.

## Coupon Used $\leq$ Order Count

Given that coupons are only used when orders are placed, we can safely assume that the "coupon Used" should always be less than or equal to the "Order Count". To address the missing values in the "Coupon Used" column, I've decided to apply the following approach:

If there's a blank in the "Coupon Used" column, I will replace it with the corresponding value from the "Order Count" column -1. This correction ensures that the "Coupon Used" column reflects the maximum correct value possible, as it is based on the actual order counts from the same time frame.

`=IF(ISBLANK(Q2),R2-1,Q2)`

=IF(ISBLANK(Q2),R2-1,Q2)			
	Q	R	V
	CouponUsed	OrderCount	Coupon Used
4%		9	8
1%		2	1
3%		11	10
8%		15	14
2%		7	6
4%		4	3
4%		1	0
6%		1	0
1%		1	0
4%		11	10

## Order amount hike from last year:

The column labeled 'order amount hike from last year' in the data dictionary is defined as '**Percentage increase in orders compared to the previous year.**' This column exclusively displays percentage increases. However, it is essential to acknowledge that not every customer will experience a growth in their order amount from the previous year. Consequently, blank values in this column are indicative of no growth or a zero percent increase, and **I have chosen to represent such cases with the value zero.**

## Warehouse to home:

The data dictionary defines the 'warehouse to home' column as the 'Distance between the customer's home and the warehouse.' In the context of e-commerce data, it's important to recognize that customers can reside at varying distances from the warehouse due to the geographical diversity of e-commerce customers. Therefore, it is not appropriate to calculate a mean or median distance for this column, as the distances can greatly vary.

To handle this, **I have chosen not to impute any values for the blank entries in this column.** Instead, we will retain the data as is, reflecting the actual distances between each customer's home and the respective warehouse locations in the dataset.



## Exploratory Data Analysis (EDA):

A

### Customer Churn Overview

The definition of churn rate is **the percentage of users who have stopped using an app.**

To calculate your churn rate, divide churned customers over a period of time by the number of customers you had at the start of that time frame.

#### Churn rate formula

##### Lost customers

(end of period)

##### Total customers

(start of period)

$$\frac{\text{Lost customers}}{\text{Total customers}} \times 100 = \text{Churn rate}$$

=COUNTIF('E Comm'!B2:B5631,"1")/COUNTIF('E Comm'!C2:C5631,">0")		
B	C	D
Churn customer during the month	948	
No of customer start of the period	4858	
<b>churn rate</b>	<b>20%</b>	

**Explanation of formula:** In Column B2:B5631, we track customers who have churned, and '1' represent a churned customer. To calculate the number of customers who churned during a specific period, we count the '1' values in this column.

In the second part of our calculation, we want to find the number of customers who were with us at the beginning of the period. To do this, we refer to the 'tenure' column (Column C). We exclude customers with a 'tenure' of zero because a zero tenure means they joined during that specific period. We only want customers who were already with us before this period.

Finally, we divide the number of churned customers by the number of customers at the start of the period to determine the churn rate.

## B

### Churn by Tenure

Do customers with longer tenure tend to churn less? Calculate the average tenure for both churned and retained customers and visualize it.

fx =AVERAGEIF('E Comm'!B2:B5631,"1",'E Comm'!C2:C5631)	
B	C
average tenure of churn customer	3.0907173
average tenure of retained customer	11.05275523

**Explanation of formula:** I'm using the 'AVERAGEIF' function to calculate the average tenure, in months, of customers who have churned. In Column B, which is the 'Churn' column where we have a '1' to represent churned customers, this serves as our condition for the average. So, we're calculating the average based on a specific condition.

We're fetching the tenure information from Column C, which represents the tenure in months. The 'AVERAGEIF' function takes into account only those customers who meet the churn condition (Column B having a '1') and calculates the average tenure using the values from Column C.

For retained customer I only change condition from 1 to 0. zero is retained customer.

## C

### Churn and Customer Satisfaction

Does customer satisfaction score correlate with churn? Calculate the average satisfaction score for both churned and retained customers

fx =AVERAGEIF('E Comm'!B2:B5631,"1",'E Comm'!L2:L5631)	
B	C
average satisfaction score of churn customer	3.39
average satisfaction score of retained customer	3.00

**Explanation of formula:** I'm using the 'AVERAGEIF' function to calculate the average customer satisfaction score, of customers who have churned. In Column B, which is

the 'Churn' column where we have a '1' to represent churned customers, this serves as our condition for the average. So, we're calculating the average based on a specific condition.

We're fetching the satisfaction score information from Column L, which represents the satisfaction score. The 'AVERAGEIF' function takes into account only those customers who meet the churn condition (Column B having a '1') and calculates the average satisfaction score using the values from Column L.

For retained customer I only change condition from 1 to 0. zero is retained customer.



## Churn and Order Count

Are customers who place more orders less likely to churn? Calculate the average order count for both churned and retained customers.

fx =AVERAGEIF('E Comm'!B2:B5631,"1",'E Comm'!U2:U5631)	
B	C
Churn Customer	2.84
Retained Customer	3.02

**Explanation of formula:** I'm using the 'AVERAGEIF' function to calculate the average Order count, of customers who have churned. In Column B, which is the 'Churn' column where we have a '1' to represent churned customers, this serves as our condition for the average. So, we're calculating the average based on a specific condition.

We're fetching the order count information from Column U, which represents the order count. The 'AVERAGEIF' function takes into account only those customers who meet the churn condition (Column B having a '1') and calculates the average order count using the values from Column L.

For retained customer I only change condition from 1 to 0. zero is retained customer.



## Churn by Preferred Payment Mode

Are there any differences in churn rates based on the preferred payment mode?

fx =COUNTIFS('E Comm'!G2:G5631,B19,'E Comm'!B2:B5631,"1")/COUNTIFS('E Comm'!G2:G5631,B19)			
	B	C	D
	Credit Card	0.14	
	Debit Card	0.15	
e	Cash on Delivery	0.25	
	E wallet	0.23	
	UPI	0.17	

This calculation show the churn rate of each payment mode

In this formula, we are using the COUNTIFS function to calculate a specific ratio.

In the 'E Comm' worksheet, Column G represents the 'Preferred Payment Mode,' and Column B represents the 'Churn' column.

We are interested in calculating the churn rate for a particular payment mode, which is indicated by the value in cell B19 (Credit Card).

So, here's how the formula works:

COUNTIFS('E Comm'!G2:G5631, B19, 'E Comm'!B2:B5631, "1") counts the number of instances where both conditions are met:

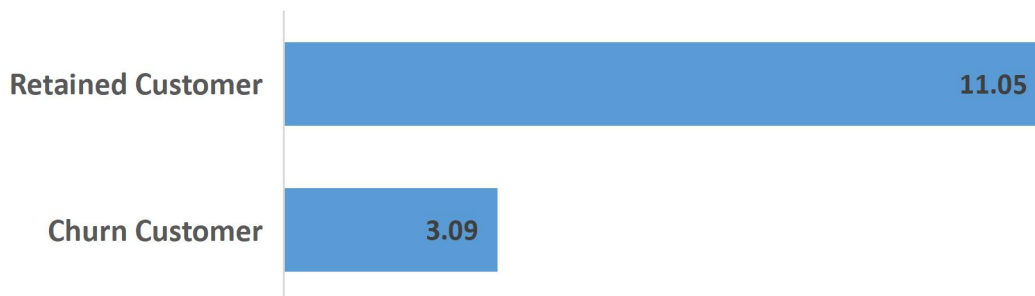
In Column G ('Preferred Payment Mode'), the value matches what is in cell B19.

In Column B ('Churn'), the value is "1," which typically represents churned customers. COUNTIFS('E Comm'!G2:G5631, B19) counts the total number of instances where the condition in Column G matches the value in cell B19, without considering the churn condition.

Finally, we divide the first count by the second count to determine the churn rate for the specific payment mode. This formula helps us understand what percentage of customers, using the payment mode indicated in cell B19, have churned out of the total customers using that payment mode.

## Visualization

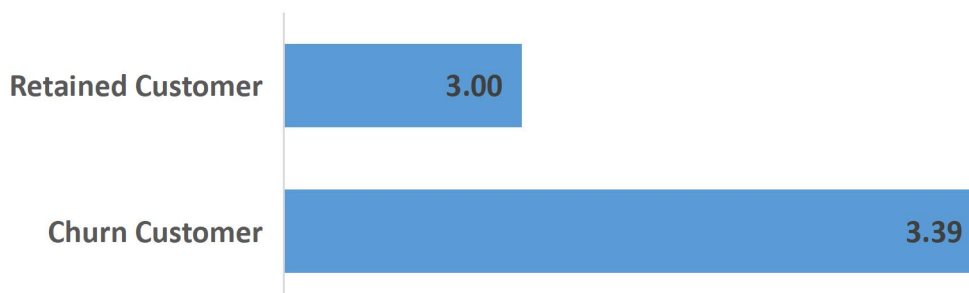
Average Tenure In Months



### Insight:

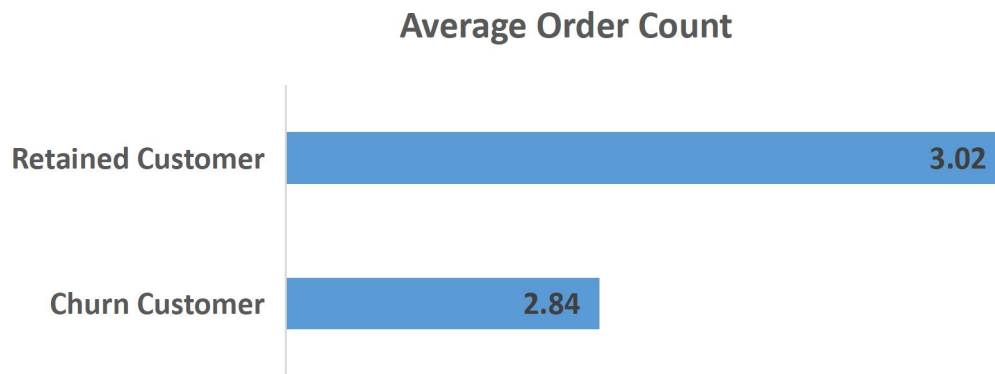
The average tenure for retained customers is 11.05 and average tenure of churn customer is 3.09 which is significantly higher than that of churned customers, it suggests that customers with longer tenure are less likely to churn.

Average Satisfaction Score



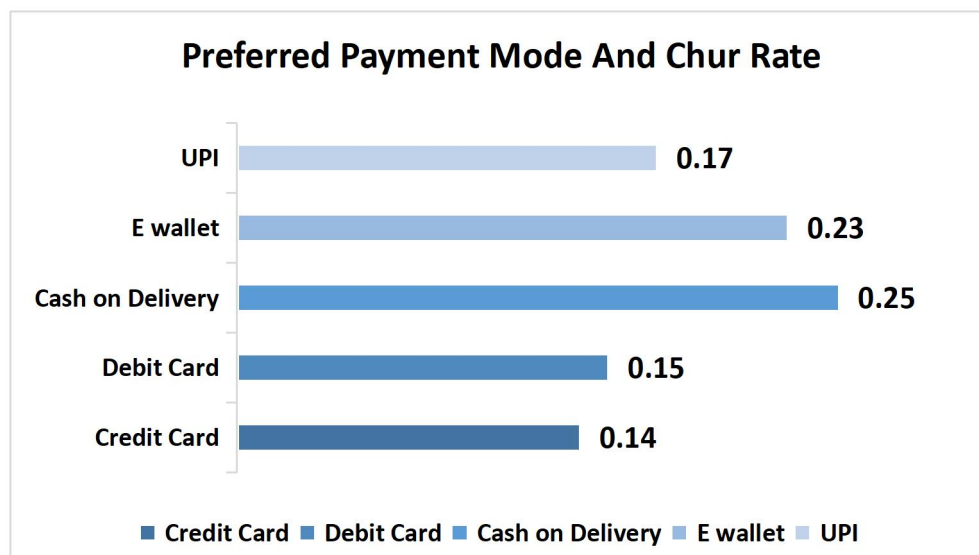
### Insight:

The data shows that churned customers have a higher average satisfaction score (3.39) compared to retained customers (3.00). This unexpected trend suggests a need for further analysis to understand why satisfied customers are churning. Consider gathering feedback, segmenting customers, and exploring external factors affecting churn. Tailor retention strategies accordingly.



### Insight:

Customers who place more orders, on average, are slightly less likely to churn. Consider strategies to encourage repeat purchases and segment your customer base for more effective retention efforts. Keep monitoring customer behavior to detect trends that impact churn.



### Insight:

Payment mode significantly influences churn rates:

- Cash on Delivery (COD) has the highest churn rate (25%).
- Debit and Credit Cards have relatively lower churn rates (15% and 14%, respectively).
- E-Wallet and UPI users fall in between (23% and 17%).

Tailor retention strategies for each payment mode, promote digital payment adoption, and continuously monitor churn rates by mode for adjustments.

## Recommendations

### 1. Tenure and Churn

The average tenure for **retained customers is 11.05** and average tenure of **churn customer is 3.09** which is significantly higher than that of churned customers, it suggests that customers with longer tenure are less likely to churn.

I recommend that the company prioritize customer loyalty by implementing a program that rewards long-term commitment. Consider offering incentives or discounts to customers who reach an 11-month milestone. Additionally, it's vital to gather insights from churned customers to understand the reasons behind their early departure. By identifying the root causes of churn, the company can make informed improvements to enhance the overall customer experience and reduce attrition.

### 2. Customer Satisfaction and Churn

Recommend investigating the reasons behind why satisfied customers choose to churn more often.

### 3. Churn and Order Count

Based on the insight that customers who place more orders are slightly less likely to churn, my recommendations are to encourage repeat purchases with incentives, engage customers, and improve product quality.

### 4. Churn by Preferred Payment Mode

Based on the provided information that **payment mode significantly influences churn rates**, here are some insights and recommendations:

- **Cash on Delivery (COD) Challenge:** The high churn rate of 25% among COD users indicates potential issues with this payment method. Consider improving the COD process, enhancing communication with COD customers, and implementing



measures to reduce fraud or non-deliveries associated with this mode.

- **Credit Card and Debit Card Reliability:** Customers using credit and debit cards have relatively lower churn rates at 14% and 15%. This suggests that these payment methods offer convenience and security. **Encourage more customers to use these options and ensure a smooth payment experience.**
- **E-Wallet and UPI Optimization:** E-Wallet and UPI users fall in between with churn rates of 23% and 17%. Explore ways to enhance the user experience for these payment modes, such as improving app functionality, transaction speed, and customer support.
- **Customer Education:** Educate customers about the benefits and convenience of using credit/debit cards or digital payment methods. This can help shift more customers to lower-churn-rate payment options.