E-Commerce Customer Churn Analysis

21/09/2023

**Presented by:**

Mubeen Arshad

**Data Cleaning:**

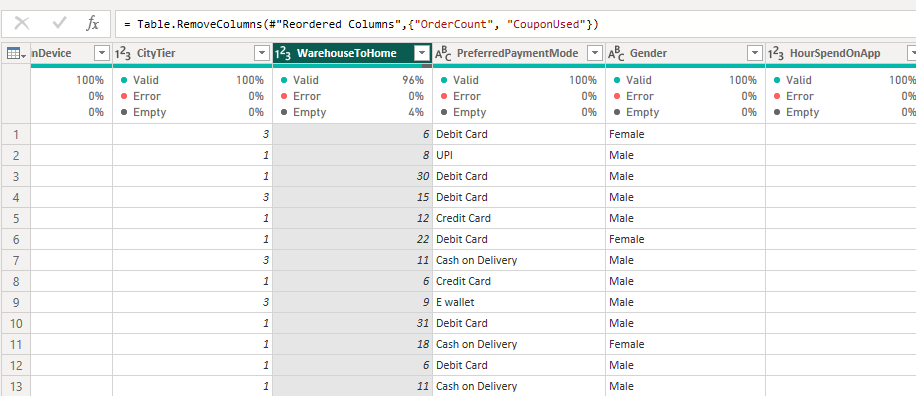
We have performed the data cleaning in MS Excel. There were some columns with missing values and some of the columns contained inconsistent data. Here is the process of dealing with missing values and inconsistent data.

**Missing values:**

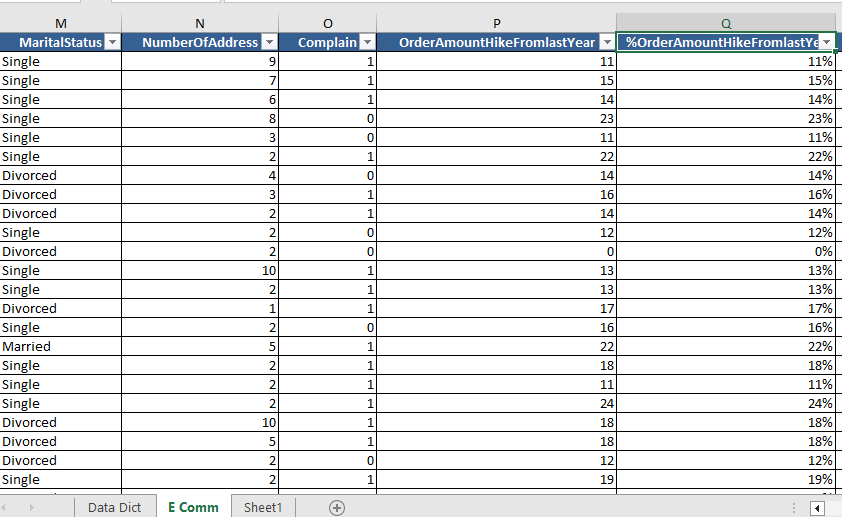
Columns with missing values were WerehouseToHome, OrderAmountHikeFromLastYear, OrderCount, and CouponUsed.

The WerhouseToHome columns consisted of 251 missing values. We have left the WerehouseToHome column as it is because of the data integrity. If the data was genuinely missing because some customers may not have provided this information, it's better to retain the missing values as is rather than make assumptions or fill in the data incorrectly. Also, imputing missing distance data with a mean, median, or any other value could introduce bias into the analysis.

It shows that there is 4% empty data.



The 'OrderAmountHikeFromLastYear' column had 265 missing values. These missing values in the 'OrderAmountHikeFromLastYear' column have been filled in with '0'. By filling missing values with '0', we essentially assume that there was no percentage increase in the order amount from the previous year for the cases with missing data. If we were to remove rows with missing values, we would lose valuable information and potentially reduce the representativeness of the dataset.



The OrderCount column consisted of 258 missing values. The column OrderCount missing values are filled with using the following formula,

**=IF(ISBLANK(S2),IF(R2=0,R2+1,R2),S2)**

It shows that if the "OrderCount" value was missing or empty, we replaced it with the same value as the "CouponUsed" column. Because we were assuming that the number of orders is equal to or greater than the number of coupons used. If "CouponUsed" is zero and "OrderCount" is missing, we assumed that at least one order was placed, so set the "OrderCount" to 1. This assumption ensures that there's at least one order even when the exact count is missing.

A screenshot of a computer

Description automatically generated

We have performed this action because the column labeled 'DaySinceLastOrder' contains no empty or missing values. Therefore, if there is a value present in the 'DaySinceLastOrder' column, it indicates that at least one order was placed. We assume that a value of '0' in the 'DaySinceLastOrder' column signifies that the order was placed on the same day as the provided data.

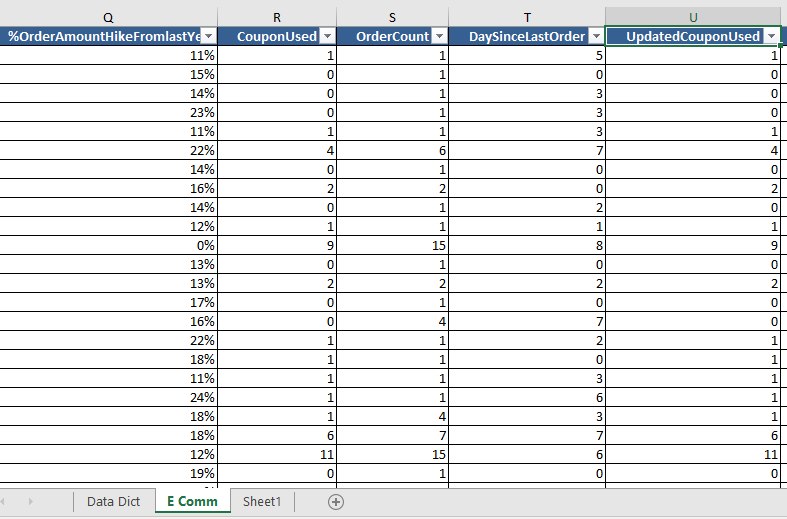
A screenshot of a spreadsheet

Description automatically generated

The CouponUsed column had 256 missing values. The column CouponUsed missing values are replaced or filled with the following formula,

**=IF(ISBLANK(R2),S2-1,R2)**

When 'CouponUsed' is missing, we assume that the customer used one fewer coupon than the number of orders placed. This is a reasonable assumption since customers typically use coupons when they place orders. Therefore, if they placed two orders, we assume they used one coupon. This approach is employed because we cannot replace the missing values with 0, as 0 is already present in the column. We might consider using 0 only if the column does not already contain 0.



This is not the exact best practice, but this is the closest best possible solution to deal with the missing values and it will reduce the minimum possible damage.

**Data Inconsistency:**

The data was inconsistent, and we changed it in the columns of PreferredLoginDevice, PreferredPaymentMode, and in the PreferredOrderCategory column.

• In the 'PreferredLoginDevice' category, there are two overlapping categories: mobile phone and phone.

• In the 'PreferredOrderCategory' category, there are also two overlapping categories: mobile and mobile phone.

• Additionally, in the 'PreferredPaymentMode' category, there are overlapping categories such as CC and Credit Card, as well as COD and Cash on Delivery.

**Analysis and Insights:**

**Customer Churn Overview:**

A number on a white background

Description automatically generated

The churn rate is found by creating the measure and applying the following DAX,

**Churn\_Rate = DIVIDE(COUNTROWS(FILTER('E Comm (3)', 'E Comm (3)'[Churn] = 1)), COUNTROWS('E Comm (3)'))**

**Insights:**

The churn rate of 16.84% shows that a specific portion of customers in the dataset have ended their association with the company.

**Churn by Tenure:**

A graph with a bar and a number of bars

Description automatically generated with medium confidence

**Insights:**

This analysis indicates that the customers who churned have less average tenure period which is 3.1 and the customers who are retained with the company have an average of longer tenure around 11.1.

**Churn and Customer Satisfaction:**

A graph of a graph

Description automatically generated with medium confidence

**Insights:**

The average satisfaction score for churned customers is 3.39 and retained customers is 3.00. which indicated that the customers who are more satisfied tend to churn more likely and customers who are not satisfied are retained with the company.

**Churn and Order Count:**

A graph of a bar chart

Description automatically generated

**Insights:**

The analysis shows that the average order count for retained customers is 3.02, which is more than the churned customers which is 2.84.

But this order count and satisfaction score reveals that the churned customers have less order count, but they are more satisfied although churned customers have more order count, but they are not satisfied.

**Churn rates based on Preferred Payment Mode:**

We have performed this analysis by creating measures and applying DAX to each individual payment mode category.

ChurnedCODPercentage =

DIVIDE(

    CALCULATE(

        COUNTROWS('E Comm (3)'),

        'E Comm (3)'[Churn] = 1,

        'E Comm (3)'[PreferredPaymentMode] = "Cash on Delivery"

    ),

    CALCULATE(

        COUNTROWS('E Comm (3)'),

        'E Comm (3)'[PreferredPaymentMode] = "Cash on Delivery"

    )

)

The same DAX is used for other payment mode categories.

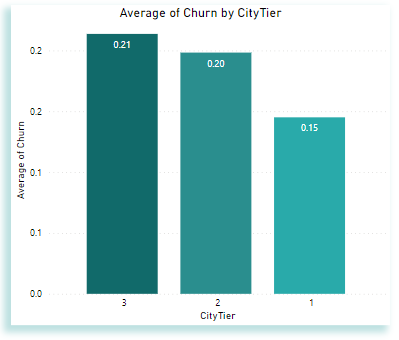
A pie chart with text and numbers

Description automatically generated

**Insights:**

From the data provided, it is clear that “Cash on Deliver” and “E-wallet” are the most preferred payment modes among churned customers, with 24.90% and 22.80% of churned customers respectively. This indicates that these two payment modes might be more commonly associated with churn. While “Debit card” and “Credit card” have lower churn rates.

**Churn by City Tier:**



**Insights:**

The analysis shows that the city tier has an impact on customer churn rates. Tier 1 cities have lower churn rates than Tier 2 and Tier 3 cities, possibly due to competition, customer preferences, or the availability of alternatives across different city tiers.

**Recommendations:**

1. Collect feedback from users who have churned to understand their reasons for leaving. Conduct surveys or use feedback forms within the app to gather insights into areas that need improvement. Also, compare the app's churn rate to industry benchmarks and competitor rates. If the app's churn rate is significantly higher, investigate what competitors are doing differently and consider adopting successful practices.
2. Implement retention programs that reward and engage long-tenured customers. These programs can include loyalty rewards, exclusive offers, and personalized communication to demonstrate appreciation for their long-standing relationship with the company. Focus on improving the onboarding experience for new customers, as the analysis suggests that shorter tenure is associated with higher churn.
3. Implement early intervention strategies for customers showing signs of dissatisfaction, even if they place a high number of orders. Proactively reach out to understand and resolve their concerns to prevent churn. Also, recognize that order count alone does not guarantee customer satisfaction or retention. Focus on improving the overall customer experience, which may lead to increased satisfaction and higher retention rates, even for customers with lower order counts.
4. Consider offering specific incentives or promotions for users of payment modes associated with higher churn rates. These incentives could include discounts, cashback offers, or loyalty programs to encourage continued usage.
5. Tailor retention strategies based on the different city tiers. Understand the preferences and needs of customers in each tier to provide targeted offerings and incentives that resonate with them.