**E-commerce Customer Churn Analysis using SQL**

**Understanding why customers are leaving an online e-commerce company**

Customer churn refers to the situation when customers decide to stop doing business with a company. This is a significant concern for businesses, as it hampers customer retention and overall success. In the fast-paced world of online retail, e-commerce companies face the challenge of retaining customers. To address this, identifying at-risk customers and implementing targeted retention strategies are crucial. This article presents a detailed analysis of a dataset from an online retail company, revealing valuable insights on customer churn. These insights offer essential guidance for decision-making, enabling proactive measures to reduce attrition and foster long-term loyalty. By understanding the factors driving churn, companies can mitigate attrition by providing appealing promotions and ensuring ongoing customer engagement.

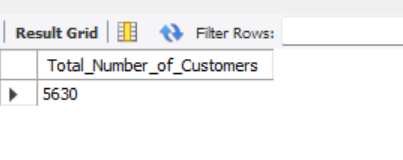
**Project Approach**

I got this dataset from [Kaggle](https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction?sort=most-comments), and it contains information such as customers' personal details, satisfaction scores, preferred payment mode, days since the last order, and cashback amount. I used SQL (Azure Data Studio) to clean and analyze this dataset, and performed visualizations using Microsoft Power BI. This analysis is divided into several stages: data cleaning, data exploration, an insight section, and recommendations.

Data Cleaning

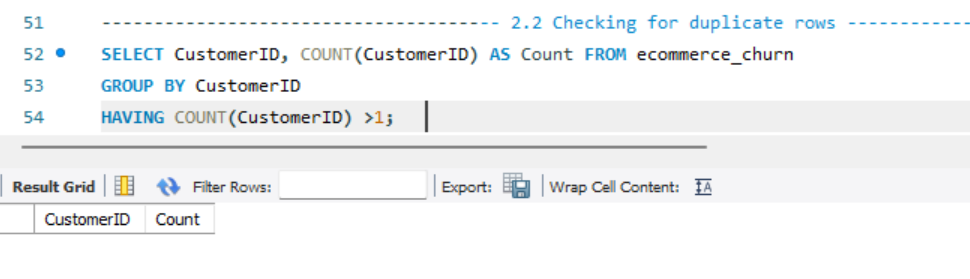
Before embarking on any analysis, it is essential to ensure the dataset is clean and reliable. The data cleaning process involves handling missing values, correcting inconsistencies, and formatting the data for analysis. In this project, we carefully cleaned the dataset to ensure the accuracy and integrity of our findings.

**1. Finding the total number of customers**



There are 5,630 customers found in this dataset.

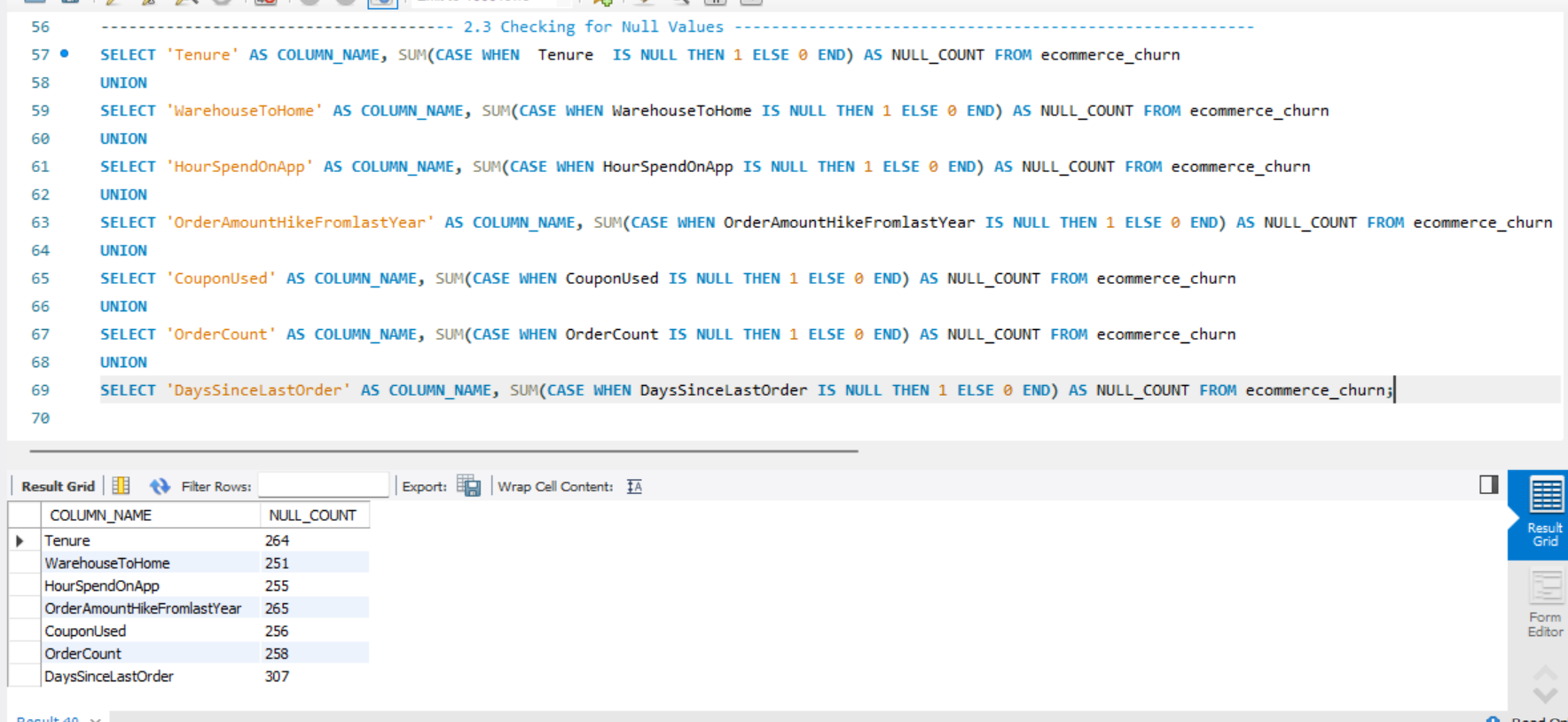
**2. Checking for duplicate rows**



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

**3. Checking for null values**

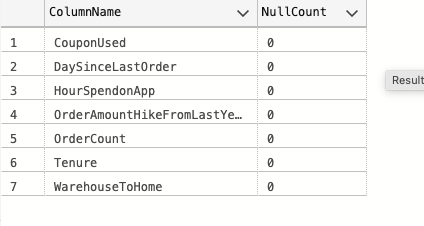
We start out by finding the null count of columns with null values present.



CouponUsed, DaysSinceLastOrder, HourSpendOnApp, OrderAmountHikeFromLastYear, OrderCount, Tenure, and WarehouseToHome all have null values present, and the number of null values present for each column can be seen in the above table.

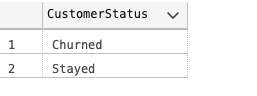
3.1 Handling null values

In order to handle these null values, we will fill the nulls with the mean value of their respective columns. Now that the null values have been replaced with their mean, there are no more null values left.



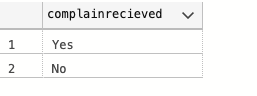
**4. Creating a new column from an already existing “churn” column**

While going through the dataset, we noticed that the churn column contained 0 and 1 values. 0 means that a customer did not churn, while 1 means that a customer churned. It is difficult to remember which is for what, so we will create a new column called ‘CustomerStatus’ that shows “Stayed” when a customer did not churn and “Churned” when a customer churned. The new column “CustomerStatus” has two distinct values called "churned" and "stayed," as earlier described.



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

Very similar to what we did in the section above, we noticed that the complain column also contained 0 and 1. ‘0’ means that the customer did not record any complaints, while ‘1’ means the customer recorded a complaint. For clarity purposes, we will create a new column called ‘ComplainRecieved’ that shows “No” when a customer did not complain and “Yes” when a customer complained.

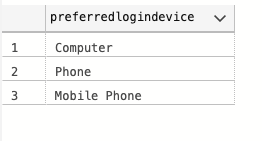
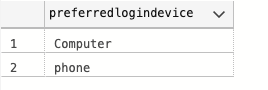


**6. Checking values in each column for correctness and accuracy**

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

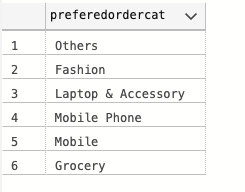
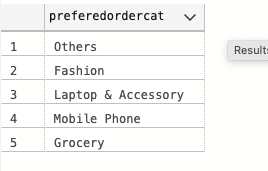
**6.1 Fixing redundancy in “PreferedLoginDevice” Column**

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

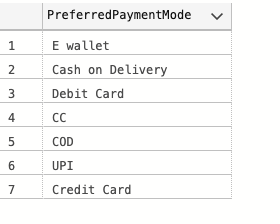
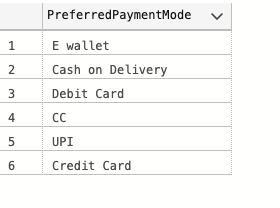
**6.2 Fixing redundancy in “PreferedOrderCat” Column**

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

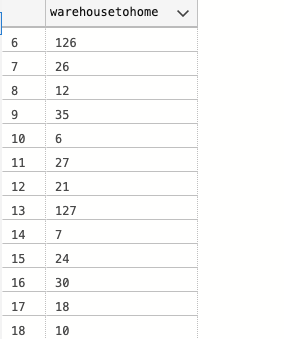
**6.3 Fixing redundancy in “PreferredPaymentMode” Column**

Notice that Cash on Delivery and COD both appear in the column, but they mean the same thing.

**6.4 Fixing wrongly entered values in “WarehouseToHome” column**

The table shows a sample of the entire result. However, notice the 126 and 127 values; they are definitely outliers and most likely wrongly entered. To fix this, we will correct the values to 26 and 27 to fall within the range of the values found in this column.

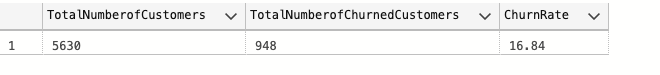


*Our data has been cleaned and is now ready to be explored for insight generation.*

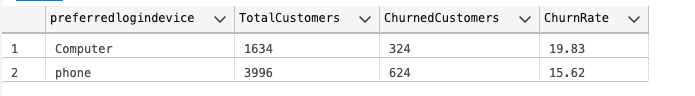
**Data Exploration**

Answering business questions.

**1. What is the overall customer churn rate?**

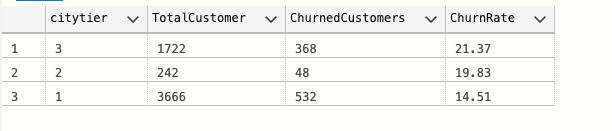
The churn rate of 16.84% indicates that a significant portion of customers in the dataset have ended their association with the company.

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



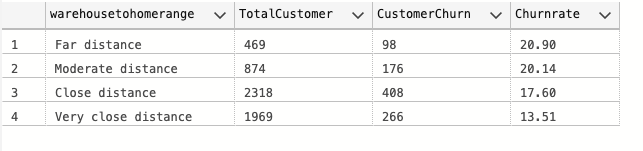
The preferred login device appears to have some influence on customer churn rates. Customers who prefer logging in using a computer have a slightly higher churn rate compared to customers who prefer logging in using their phones. This may indicate that customers who access the platform via a computer might have different usage patterns, preferences, or experiences that contribute to a higher likelihood of churn. Understanding these preferences can help businesses optimize their platform and user experience for different login devices, ensuring a seamless and engaging experience for customers.

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

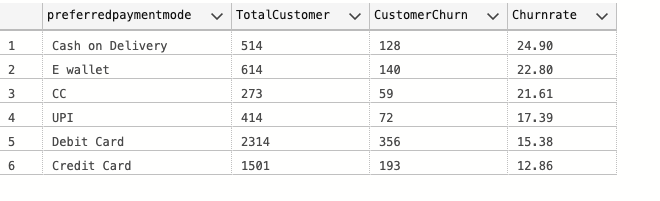


**4. Is there any correlation between the warehouse-to-home distance and customer churn?**

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

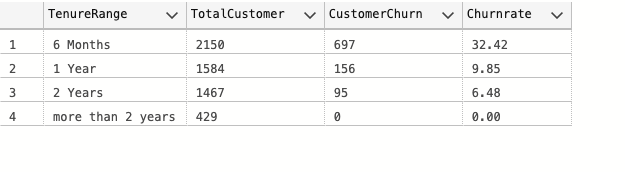


**5. Which is the most preferred payment mode among churned customers?**

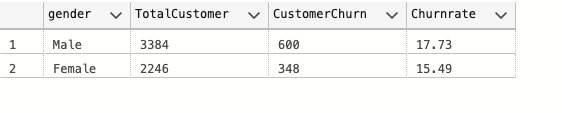


**6. What is the typical tenure for churned customers?**

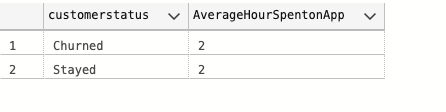
First, we will create a new column called “TenureRange” that groups the customer tenure into 6 months, 1 year, 2 years, and more than 2 years using the CASE statement.



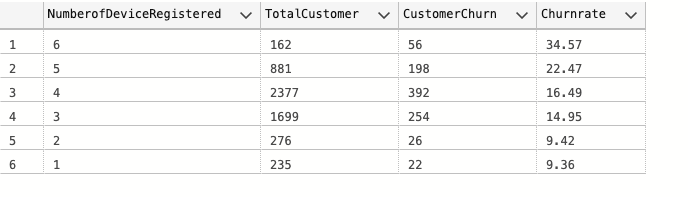
**7. Is there any difference in churn rate between male and female customers?**



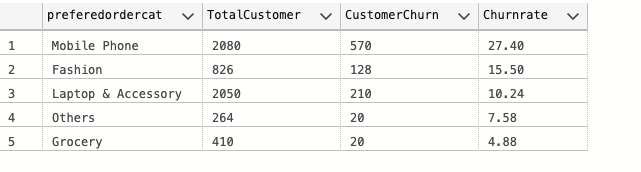
**8. How does the average time spent on the app differ for churned and non-churned customers?**



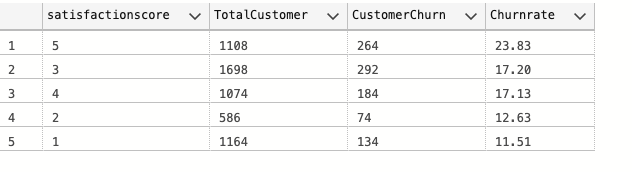
**9. Does the number of registered devices impact the likelihood of churn?**



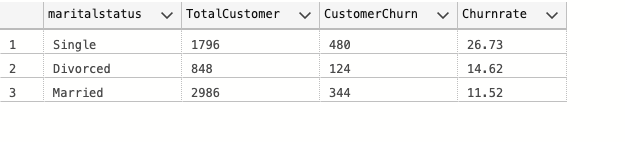
**10. Which order category is most preferred among churned customers?**



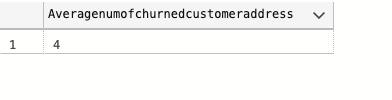
**11. Is there any relationship between customer satisfaction scores and churn?**



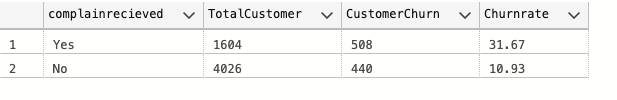
**12. Does the marital status of customers influence churn behaviour?**



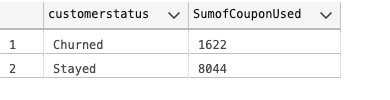
**13. How many addresses do churned customers have on average?**



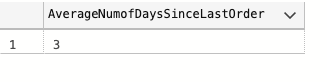
**14. Do customer complaints influence churned behavior?**



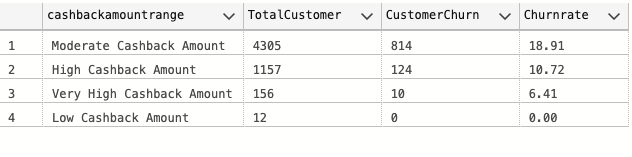
**15. How does the use of coupons differ between churned and non-churned customers?**



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



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



Actionable Insights

* The dataset includes 5,630 customers, providing a substantial sample size for analysis. The overall churn rate is 16.84%, indicating significant customer attrition.
* Customers who prefer logging in with a computer have slightly higher churn rates compared to phone users, suggesting different usage patterns and preferences.
* Tier 1 cities have lower churn rates than Tier 2 and Tier 3 cities, possibly due to competition and customer preferences.
* Proximity to the warehouse affects churn rates, with closer customers showing lower churn, highlighting the importance of optimizing logistics and delivery strategies.
* “Cash on Delivery” and “E-wallet” payment modes have higher churn rates, while “Credit Card” and “Debit Card” have lower churn rates, indicating the influence of payment preferences on churn.
* Longer tenure is associated with lower churn rates, emphasizing the need for building customer loyalty early on.
* Male customers have slightly higher churn rates than female customers, although the difference is minimal.
* App usage time does not significantly differentiate between churned and non-churned customers.
* More registered devices correlate with higher churn rates, suggesting the need for consistent experiences across multiple devices.
* “Mobile Phone” order category has the highest churn rate, while “Grocery” has the lowest, indicating the importance of tailored retention strategies for specific categories.
* Highly satisfied customers (rating 5) have a relatively higher churn rate, highlighting the need for proactive retention strategies at all satisfaction levels.
* Single customers have the highest churn rate, while married customers have the lowest, indicating the influence of marital status on churn.
* Churned customers have an average of four associated addresses, suggesting higher mobility.
* Customer complaints are prevalent among churned customers, emphasizing the importance of addressing concerns to minimize churn.
* Coupon usage is higher among non-churned customers, showcasing the effectiveness of loyalty rewards and personalized offers.
* Churned customers have had a short time since their last order, indicating recent disengagement and the need for improved customer experience and retention initiatives.
* Moderate cashback amounts correspond to higher churn rates, while higher amounts lead to lower churn, suggesting the positive impact of higher cashback on loyalty.

Recommendations

* Enhance the user experience for customers who prefer logging in via a computer. Conduct research to identify and address any issues they might be facing, making improvements to ensure a smoother and more enjoyable experience.
* 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.
* Optimize logistics and delivery to improve customer satisfaction. Focus on reducing delivery times, lowering shipping costs, and finding ways to make the process more convenient, especially for customers living further away.
* Simplify payment processes, particularly for options like “Cash on Delivery” and “E-wallet.” Enhance security measures and offer incentives for customers to use more reliable payment methods such as “Credit Card” and “Debit Card.”
* Improve customer support and complaint resolution. Address customer complaints promptly and effectively, providing satisfactory resolutions. Actively listen to customer feedback, make necessary improvements, and demonstrate a commitment to addressing their concerns.
* Develop targeted retention initiatives for specific order categories, such as the “Mobile Phone” category. Offer exclusive discounts, rewards, or promotions to incentivize continued engagement and loyalty.
* Ensure a consistent experience across different devices. Implement features like cross-device syncing, personalized recommendations, and easy account management to encourage usage and retention across multiple devices.
* Proactively engage and reward satisfied customers across all satisfaction levels. Regularly communicate with them through personalized messages, exclusive offers, and loyalty programs to maintain their loyalty and reduce the risk of churn.
* Consider increasing cashback incentives to retain customers, especially those who are more likely to churn. Conduct A/B testing to determine the most effective cashback levels that encourage higher customer retention rates.

By implementing these recommendations, this company can improve customer retention, reduce churn rates, and build long-term loyalty, leading to sustainable growth and success.