

CUSTOMER CHURN ANALYSIS - E-COMMERCE PLATFORM

“UNDERSTANDING CHURN PATTERNS AND IMPROVING RETENTION”

CREATED BY SHIKHA PANDEY



PROJECT OVERVIEW

- Objective: Analyze customer churn patterns, identify key drivers, and suggest retention strategies.
- Tools Used: SQL for Database Setup, Data Cleaning, Exploratory Data Analysis (EDA), and Insights Generation.
- Dataset: ~5,500+ e-commerce customer records with features like Tenure, Login Device, Payment Mode, Satisfaction Score, Churn, etc

DATABASE SETUP

-- Create Database

```
CREATE DATABASE IF NOT EXISTS Ecommerce_db;
```

```
USE Ecommerce_db;
```

-- View raw customer data

```
SELECT * FROM customers;
```

DATA PREPARATION AND CLEANING

-- 2.1 Identify Duplicate Records

```
SELECT CustomerID, COUNT(*) AS duplicate_count
FROM customers_data
GROUP BY CustomerID
HAVING COUNT(*) > 1;
```

-- Alternative approach using ROW_NUMBER

```
) WITH duplicate_cte AS (
    SELECT *,
        ROW_NUMBER() OVER (
            PARTITION BY CustomerID, Churn, Tenure, PreferredLoginDevice, CityTier,
            WarehouseToHome, PreferredPaymentMode, Gender, HourSpendOnApp, PreferredOrderCat,
            SatisfactionScore, MaritalStatus, NumberOfAddress, Complain, OrderAmountHikeFromLastYear,
            CouponUsed, OrderCount, DaySinceLastOrder, CashbackAmount
        ) AS row_num
    FROM customers_data
)
SELECT *
FROM duplicate_cte
WHERE row_num > 1;
```

REMOVING DUPLICATES

```
-- View duplicate records  
SELECT * FROM customers_data2  
WHERE row_num > 1;
```

```
-- Disable safe updates to delete duplicates  
SET SQL_SAFE_UPDATES = 0;
```

```
-- Remove duplicates  
DELETE FROM customers_data2  
WHERE row_num > 1;
```

```
-- View cleaned data  
SELECT * FROM customers_data2;
```

STANDARDIZING COLUMNS

-- Check unique payment modes

```
SELECT DISTINCT PreferredPaymentMode FROM customers_data2;
```

-- Standardize payment modes

```
UPDATE customers_data2
SET PreferredPaymentMode = 'Cash on Delivery'
WHERE PreferredPaymentMode = 'COD';
```

```
UPDATE customers_data2
```

```
SET PreferredPaymentMode = 'Credit Card'
WHERE PreferredPaymentMode = 'CC';
```

-- Standardize order categories

```
UPDATE customers_data2
SET PreferredOrderCat = 'Mobile Phone'
WHERE PreferredOrderCat LIKE 'Mobile%';
```

CHECKING MISSING VALUES

```
-- Check missing/null values
SELECT
    SUM(CASE WHEN CustomerID IS NULL OR CustomerID = '' THEN 1 ELSE 0 END) AS missing_CustomerID,
    SUM(CASE WHEN Churn IS NULL OR Churn = '' THEN 1 ELSE 0 END) AS missing_Churn,
    SUM(CASE WHEN Tenure IS NULL OR Tenure = '' THEN 1 ELSE 0 END) AS missing_Tenure,
    SUM(CASE WHEN PreferredLoginDevice IS NULL OR PreferredLoginDevice = '' THEN 1 ELSE 0 END) AS missing_PREFERREDLOGINDEVICE,
    SUM(CASE WHEN CityTier IS NULL OR CityTier = '' THEN 1 ELSE 0 END) AS missing_CityTier,
    SUM(CASE WHEN WarehouseToHome IS NULL OR WarehouseToHome = '' THEN 1 ELSE 0 END) AS missing_WarehouseToHome,
    SUM(CASE WHEN PreferredPaymentMode IS NULL OR PreferredPaymentMode = '' THEN 1 ELSE 0 END) AS missing_PREFERREDPAYMENTMODE,
    SUM(CASE WHEN Gender IS NULL OR Gender = '' THEN 1 ELSE 0 END) AS missing_Gender,
    SUM(CASE WHEN HourSpendOnApp IS NULL OR HourSpendOnApp = '' THEN 1 ELSE 0 END) AS missing_HourSpendOnApp,
    SUM(CASE WHEN Number0fDeviceRegistered IS NULL OR Number0fDeviceRegistered = '' THEN 1 ELSE 0 END) AS missing_Number0fDeviceRegistered,
    SUM(CASE WHEN PreferedOrderCat IS NULL OR PreferedOrderCat = '' THEN 1 ELSE 0 END) AS missing_PreferedOrderCat,
    SUM(CASE WHEN SatisfactionScore IS NULL OR SatisfactionScore = '' THEN 1 ELSE 0 END) AS missing_SatisfactionScore,
    SUM(CASE WHEN MaritalStatus IS NULL OR MaritalStatus = '' THEN 1 ELSE 0 END) AS missing_MaritalStatus,
    SUM(CASE WHEN Number0fAddress IS NULL OR Number0fAddress = '' THEN 1 ELSE 0 END) AS missing_Number0fAddress,
    SUM(CASE WHEN Complain IS NULL OR Complain = '' THEN 1 ELSE 0 END) AS missing_Complain,
    SUM(CASE WHEN OrderAmountHikeFromlastYear IS NULL OR OrderAmountHikeFromlastYear = '' THEN 1 ELSE 0 END) AS missing_OrderAmountHikeFromlastYear,
    SUM(CASE WHEN CouponUsed IS NULL OR CouponUsed = '' THEN 1 ELSE 0 END) AS missing_CouponUsed,
    SUM(CASE WHEN OrderCount IS NULL OR OrderCount = '' THEN 1 ELSE 0 END) AS missing_OrderCount,
    SUM(CASE WHEN DaySinceLastOrder IS NULL OR DaySinceLastOrder = '' THEN 1 ELSE 0 END) AS missing_DaySinceLastOrder,
    SUM(CASE WHEN CashbackAmount IS NULL OR CashbackAmount = '' THEN 1 ELSE 0 END) AS missing_CashbackAmount,
    SUM(CASE WHEN row_num IS NULL OR row_num = '' THEN 1 ELSE 0 END) AS missing_row_num
FROM customers_data2;
```

UPDATING MISSING VALUES

```
-- Update missing Tenure with average
UPDATE customers_data2
) SET Tenure = (
) SELECT avg_tenure FROM (
    SELECT ROUND(AVG(Tenure), 0) AS avg_tenure
    FROM customers_data2
    WHERE Tenure IS NOT NULL AND Tenure <> ''
) AS t
)
WHERE Tenure IS NULL OR Tenure = '';

-- Update missing HourSpendOnApp with average
UPDATE customers_data2
) SET HourSpendOnApp = (
) SELECT avg_HourSpendOnApp FROM (
    SELECT ROUND(AVG(HourSpendOnApp), 0) AS avg_HourSpendOnApp
    FROM customers_data2
    WHERE HourSpendOnApp IS NOT NULL AND HourSpendOnApp <> ''
) AS t
)
WHERE HourSpendOnApp IS NULL OR HourSpendOnApp = '';
```

EXPLORATORY DATA ANALYSIS

(EDA)

TOTAL CUSTOMERS, CHURNED CUSTOMERS, CHURN RATE

-- Total Customers, Churned Customers, Churn Rate

SELECT

```
COUNT(*) AS Total_Customers,  
SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) AS Churned_Customers,  
ROUND(100.0 * SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) / COUNT(*), 2) AS Churn_Rate_Percent  
FROM customers_data2;
```

Total_Customers	Churned_Customers	Churn_Rate_Percent
4293	770	17.94

CHURN BY CITY TIER (DEMOGRAPHICS)

```
-- Churn by City Tier (Demographics)
) WITH City_Churn AS (
    SELECT
        CityTier,
        COUNT(*) AS Total_Customers,
        SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) AS Churned_Customers
    FROM customers_data2
    GROUP BY CityTier
)
SELECT
    CityTier,
    Total_Customers,
    Churned_Customers,
    ROUND(100.0 * Churned_Customers / Total_Customers, 2) AS Churn_Rate_Percent
FROM City_Churn
ORDER BY Churn_Rate_Percent DESC;
```

CityTier	Total_Customers	Churned_Customers	Churn_Rate_Percent
3	1358	304	22.39
2	184	40	21.74
1	2751	426	15.49

CHURN BY PREFERRED PAYMENT MODE

```
-- Churn by Preferred Payment Mode
) WITH Payment_Churn AS (
    SELECT
        PreferredPaymentMode,
        COUNT(*) AS Total_Customers,
        SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) AS Churned_Customers
    FROM customers_data2
    GROUP BY PreferredPaymentMode
)
SELECT
    PreferredPaymentMode,
    Total_Customers,
    Churned_Customers,
    ROUND(100.0 * Churned_Customers / Total_Customers, 2) AS Churn_Rate_Percent
FROM Payment_Churn
ORDER BY Churn_Rate_Percent DESC;
```

PreferredPaymentMo...	Total_Customers	Churned_Customers	Churn_Rate_Percent
Cash on Delivery	405	98	24.20
E wallet	477	112	23.48
UPI	333	68	20.42
Debit Card	1765	294	16.66
Credit Card	1313	198	15.08

CHURN RATE BY SATISFACTION SCORE

-- Churn rate by Satisfaction Score

SELECT

```
SatisfactionScore,  
COUNT(*) AS Total_Customers,  
SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) AS Churned_Customers,  
ROUND(100.0 * SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) / COUNT(*), 2) AS Churn_Rate_Percent
```

FROM customers_data2

GROUP BY SatisfactionScore

ORDER BY SatisfactionScore **desc**;

SatisfactionScore	Total_Customers	Churned_Customers	Churn_Rate_Percent
5	844	214	25.36
4	812	142	17.49
3	1304	244	18.71
2	442	60	13.57
1	891	110	12.35

CHURN RATE BY NUMBER OF COMPLAINTS

-- Churn rate by Number of Complaints

```
SELECT  
    Complain,  
    COUNT(*) AS Total_Customers,  
    SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) AS Churned_Customers,  
    ROUND(100.0 * SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) / COUNT(*), 2) AS Churn_Rate_Percent  
FROM customers_data2  
GROUP BY Complain  
ORDER BY Churn_Rate_Percent DESC;
```

Complain	Total_Customers	Churned_Customers	Churn_Rate_Percent
1	1203	404	33.58
0	3090	366	11.84

CHURN BY APP USAGE

-- Churn by App usage

```
SELECT
    CASE
        WHEN HourSpendOnApp < 2 THEN 'Low Usage (<2 hrs)'
        WHEN HourSpendOnApp BETWEEN 2 AND 4 THEN 'Medium Usage (2-4 hrs)'
        ELSE 'High Usage (>4 hrs)'
    END AS Usage_Group,
    COUNT(*) AS Total_Customers,
    SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) AS Churned_Customers,
    ROUND(100.0 * SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) / COUNT(*), 2) AS Churn_Rate_Percent
FROM customers_data2
GROUP BY Usage_Group
ORDER BY Churn_Rate_Percent DESC;
```

Usage_Group	Total_Customers	Churned_Customers	Churn_Rate_Percent
Medium Usage (2-4 hrs)	4262	770	18.07
Low Usage (<2 hrs)	28	0	0.00
High Usage (>4 hrs)	3	0	0.00

DEVICE USAGE VS CHURN

-- Device usage vs churn

SELECT

```
PreferredLoginDevice,  
COUNT(*) AS Total_Customers,  
SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) AS Churned_Customers,  
ROUND(100.0 * SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) / COUNT(*), 2) AS Churn_Rate_Percent  
FROM customers_data2  
GROUP BY PreferredLoginDevice  
ORDER BY Churn_Rate_Percent DESC;
```

PreferredLoginDev...	Total_Customers	Churned_Customers	Churn_Rate_Percent
Phone	1056	224	21.21
Computer	1268	262	20.66
Mobile Phone	1969	284	14.42

CHURN BY ORDER CATEGORY PREFERENCE

-- Churn by Order Category Preference

```
SELECT  
    PreferredOrderCat,  
    ROUND(100.0 * SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) / COUNT(*), 2) AS Churn_Rate_Percent  
FROM customers_data2  
GROUP BY PreferredOrderCat  
ORDER BY Churn_Rate_Percent DESC;
```

PreferredLoginDevice	Total_Customers	Churned_Customers	Churn_Rate_Percent
Phone	1056	224	21.21
Computer	1268	262	20.66
Mobile Phone	1969	284	14.42

CASHBACK BUCKET ANALYSIS

```
-- Cashback Bucket Analysis
) WITH CashbackBuckets AS (
    SELECT
        CASE
            WHEN CashbackAmount = 0 THEN 'No Cashback'
            WHEN CashbackAmount BETWEEN 1 AND 100 THEN 'Low Cashback'
            WHEN CashbackAmount BETWEEN 101 AND 500 THEN 'Medium Cashback'
            ELSE 'High Cashback'
        END AS Cashback_Group,
        COUNT(*) AS Total_Customers,
        SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) AS Churned_Customers
    FROM customers_data2
    GROUP BY Cashback_Group
)
SELECT
    Cashback_Group,
    Total_Customers,
    Churned_Customers,
    ROUND(100.0 * Churned_Customers / Total_Customers, 2) AS Churn_Rate_Percent
FROM CashbackBuckets
ORDER BY Churn_Rate_Percent DESC;
```

Cashback_Group	Total_Customers	Churned_Customers	Churn_Rate_Percent
Medium Cashback	4281	770	17.99
No Cashback	4	0	0.00
Low Cashback	8	0	0.00

RETAINED CUSTOMERS - AT RISK DUE TO INACTIVITY

-- Retained customers – At risk due to inactivity?

SELECT

```
CustomerID,  
DaySinceLastOrder,  
SatisfactionScore,  
RANK() OVER(ORDER BY DaySinceLastOrder DESC) AS Risk_Rank  
FROM customers_data2  
WHERE Churn = 0 AND DaySinceLastOrder > 30;
```

CustomerID	DaySinceLastOrd...	SatisfactionScore	Risk_Rank
54306	31	5	1



CUSTOMER CHURN – KEY INSIGHTS

- 📌 Overall churn: 770 of 4,293 customers → 17.94%
- 📌 City Tier: Tier 3 & 2 churn higher than Tier 1
- 📌 Payment Mode: COD & E-Wallet highest; Credit Card lowest
- 📌 Satisfaction: High scorers churn more (25.35%) than low scorers (12.35%)
- 📌 Complaints: With complaints churn 3x higher (33.58% vs 11.84%)
- 📌 App Usage: Medium users churn most (18.07%)
- 📌 Device: Phone (21.21%) > Computer (20.66%) > Mobile App (14.42%)
- 📌 Order Category: Mobile Phones churn highest (26.44%); Grocery lowest
- 📌 Cashback: Medium cashback churn highest (17.99%)

KEY RECOMMENDATIONS

- 🎯 Focus on Tier 2 & 3 cities with localized offers and engagement.
- 🎯 Offer payment-based rewards to shift users from COD & wallets.
- 🎯 Strengthen customer support to reduce churn from complaints.
- 🎯 Launch loyalty & cashback programs for medium-risk users.
- 🎯 Enhance mobile app experience and run category-based offers.
- 🎯 Recheck satisfaction drivers to address hidden churn causes.

THANK YOU!

