



Portfolio Project – Customer Churn & Retention Analysis

Case Study: Zepto (Quick Commerce)

Prepared by:

Nischal Sachdeva

(Business Analyst Portfolio Project)

Disclaimer:

This project is an independent, self-initiated case study created for portfolio and learning purposes. It is not affiliated with Zepto and does not use any confidential company data. All datasets used are either publicly available or simulated.



Problem Statement and Objectives:-

Problem Statement

Zepto, one of India's fastest-growing quick-commerce companies, is currently facing a high customer churn rate of **25%**, which has been driving up Customer Acquisition Costs (CAC), lowering Customer Lifetime Value (CLV), and negatively impacting profitability. To sustain growth in a highly competitive market, Zepto aims to reduce churn to **8%** by identifying the key drivers of customer attrition, quantifying their impact, and implementing data-driven retention strategies.

Project Objectives

This project aims to:

1. Define & measure churn for Zepto customers using order engagement data.
2. Analyze churn cohorts to uncover patterns in customer behavior and retention.
3. Identify key churn drivers (delivery delays, product unavailability, discounts reliance etc.)
4. Segment customers by risk level and build actionable profiles.
5. Recommend retention strategies with measurable KPIs and ROI impact.
6. Design a dashboard for ongoing churn & retention monitoring.

Project Scope

In-Scope:

- Customer transactions (orders, refunds, cancellations).
- Delivery metrics (SLA breaches, delays).
- Marketing & engagement (offers, push notifications, campaigns).
- Customer support tickets.

Out-of-Scope:

- Corporate/B2B accounts.
- Fraudulent or duplicate customers.
- Non-digital marketing channels.

Success Criteria

- Clear churn definition and agreed KPIs.
- Retention dashboard (Power BI/Excel).
- Identification of top 3–5 churn drivers.
- 3+ actionable recommendations to improve retention by ≥ 8 percentage points.

Methodology & Approach

The Churn Analysis & Retention Project for Zepto followed a structured Business Analysis lifecycle. Each stage was approached using BA best practices, supported by analytical tools (SQL, Python, Excel, Power BI). The process ensured that business needs were systematically translated into actionable insights and validated solutions.

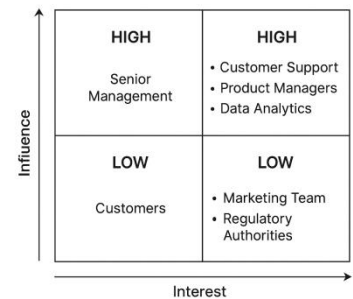
Steps	Purpose	Application in Zepto Churn Project	Key Deliverables
1. Business Need Identification	Define the problem	High churn rate observed; need to improve retention & customer lifetime value.	Problem Statement
2. Stakeholder Analysis	Identify & classify stakeholders	Operations, Marketing, Product, Customer Support, Leadership mapped via RACI.	Stakeholder Register
3. Requirements Elicitation	Gather business & functional needs	Workshops & hypothetical discussions → churn reduction KPIs identified.	Requirements Catalog
4. Business Requirements Document (BRD)	Document high-level business needs	Captured business goals, success metrics (Churn %, Retention Rate, CLV).	BRD
5. Functional Requirements Document (FRD)	Translate into system requirements	Defined SQL queries, dashboards, Python model outputs.	FRD
6. User Stories and Epics	Break requirements into manageable item	INVEST-based user stories for dashboards, churn prediction features.	User Story Backlog
7. Modeling and Visualization	Understand system & process flows	Use Case Diagrams (customer–system), BPMN (Order→Delivery→Feedback).	UML Diagrams, BPMN Models
8. Gap Analysis	Compare current vs. desired state	Current churn 25% vs. target ≤15%; identified gaps in process & engagement.	Gap Analysis Report
9. Root Cause Analysis (Ishikawa)	Identify churn drivers	Found delivery delays, pricing, poor personalization as root issues.	RCA Diagram
10. Business Process Improvement	Recommend improvements	Loyalty program, targeted offers, UX enhancements.	BPI Recommendations
11. Risk Analysis	Anticipate risks	Highlighted data inaccuracies, customer behavior shifts.	Risk Register
12. Requirement Traceability Matrix (RTM)	Ensure requirements coverage	Mapped needs → functional requirements → test cases.	RTM
13. User Acceptance Testing (UAT)	Validate solution fit	UAT scenarios for dashboards and churn model.	UAT Scripts
14. Tools and Execution	Apply technical tools	SQL (data extraction), Excel (metrics), Python (prediction), Power BI (dashboards).	Working Dashboards & Model
15. Final Deliverables	Present insights & solutions	Consolidated BA documents, dashboards, KPIs, recommendations.	Final Project Report

Stakeholder Management

Purpose:

To identify, analyze, and engage stakeholders to ensure project success, address expectations, and manage communications efficiently.

Stakeholder Analysis



Stakeholder Analysis Table:-

Stakeholder	Role in Project	Influence	Interest	Communication Frequency	Notes / Actions Required
Customer Support Team	Provide insights on customer complaints, churn reasons	High	High	Weekly	Share churn data & patterns; involve in solution planning
Marketing Team	Design retention campaigns and offers	Medium	High	Bi-weekly	Align marketing campaigns with churn reduction strategies
Product Managers	Approve changes in product or services	High	Medium	Weekly	Review proposed solutions; ensure feasibility
Data Analytics Team	Analyze churn trends, create predictive models	High	High	Daily	Provide analytics support and dashboards
Customers	End-users of products/services	Low	High	Surveys / Feedback	Gather feedback on service experience, loyalty incentives
Senior Management	Approve budget, oversee project success	High	Medium	Monthly	Ensure alignment with business objectives and ROI
Regulatory Authorities	Ensure compliance with data privacy laws	Medium	Low	As needed	Confirm adherence to legal requirements

RACI MATRIX

A RACI matrix identifies who is

- **Responsible (R)**
- **Accountable (A)**
- **Consulted (C), and**
- **Informed (I)**

for each task.

Here's a professional RACI matrix based on stakeholders and their roles:

Task / Activity	Customer Support Team	Marketing Team	Product Managers	Data Analytics Team	Customers	Senior Management	Regulatory Authorities
Provide insights on customer complaints	R	I	C	C	I	A	I
Design retention campaigns and offers	C	R	C	I	I	A	I
Approve changes in product or services	I	I	A/R	C	I	C	I
Analyze churn trends and create predictive models	C	I	C	R	I	A	I
Gather feedback on products/services	I	I	I	C	R	C	I
Approve budget and oversee project success	I	I	C	C	I	A/R	I
Ensure compliance with data privacy laws	I	I	I	C	I	C	A/R

Key Takeaways:-

Engage key stakeholders and document needs to align actions, drive retention, and prevent scope creep.

Requirements Elicitation & User Stories

1. Requirements Elicitation

To ensure that the churn prediction and retention project addresses real business needs, requirements were gathered using multiple elicitation techniques:

- **Stakeholder Interviews** – Conducted with senior management, marketing, customer support, and data teams to understand pain points.
- **Workshops** – Brainstorming sessions with cross-functional stakeholders to refine KPIs and define scope.
- **Document Analysis** – Reviewing historical churn data, customer feedback, and CRM system documentation.
- **Surveys & Questionnaires** – Collecting insights directly from customers regarding dissatisfaction factors.

2. Business Requirements Document (BRD)

The BRD defines the **why** and **what** of the project at a high level:

- **Objective:** Reduce churn rate from 25% to ~8% through data-driven interventions.
- **Business Needs:** Improve retention, increase customer lifetime value, and optimize loyalty programs.
- **KPIs:**
 - Churn % reduction (from 25% to ≤8%)
 - Increase in Net Promoter Score (NPS)
 - Growth in Average Customer Lifetime Value (CLV)
 - Reduction in SLA breaches & support tickets

3. Functional Requirements Document (FRD)

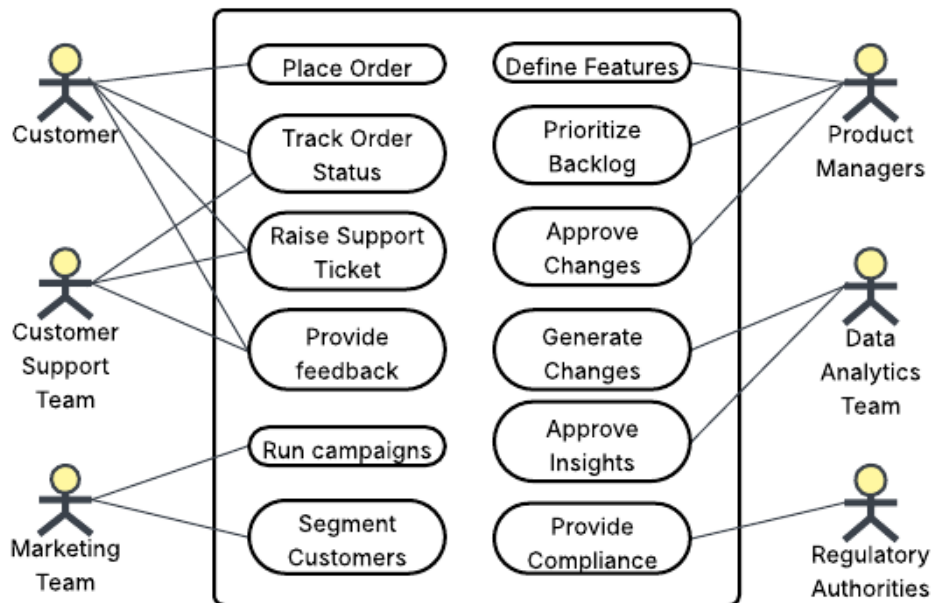
The FRD specifies the **how** of the solution:

- **Data Management:** Ingest and clean customer data (10,000+ records, multiple attributes).
- **Churn Prediction Model:** Build ML model (Python) to classify customers as “at-risk” or “loyal.”
- **Dashboard & Reporting:** Power BI dashboards for real-time monitoring of churn drivers and KPIs.
- **Alerts & Campaign Triggers:** Automated retention campaigns triggered for at-risk customers.
- **User Access:** Role-based access for executives, analysts, and customer service managers.

4. User Stories (Agile Framework)

Stakeholder	User Story	INVEST Validation
Support Lead	<i>As a support lead, I want real-time churn alerts so that I can engage customers proactively.</i>	✔Independent, ✔Valuable, ✔Testable
Marketing Manager	<i>As a marketer, I want churn segmentation by region & age so that I can design tailored campaigns.</i>	✔Negotiable, ✔Small, ✔Estimable
Product Manager	<i>As a product manager, I want insights into churn drivers so that I can prioritize product fixes.</i>	✔Independent, ✔Valuable, ✔Testable

Use Case Diagram

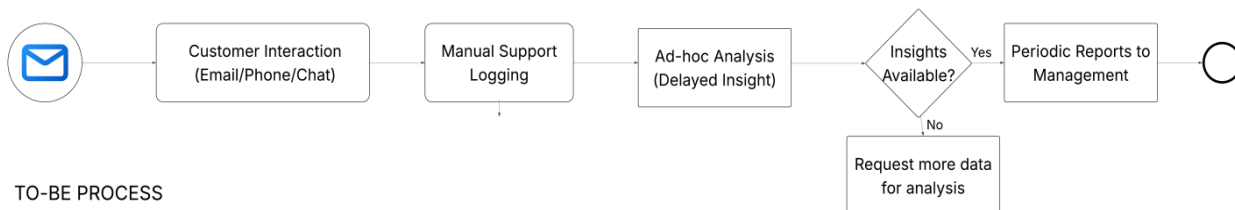


Use Case Diagram for Churn Analysis System

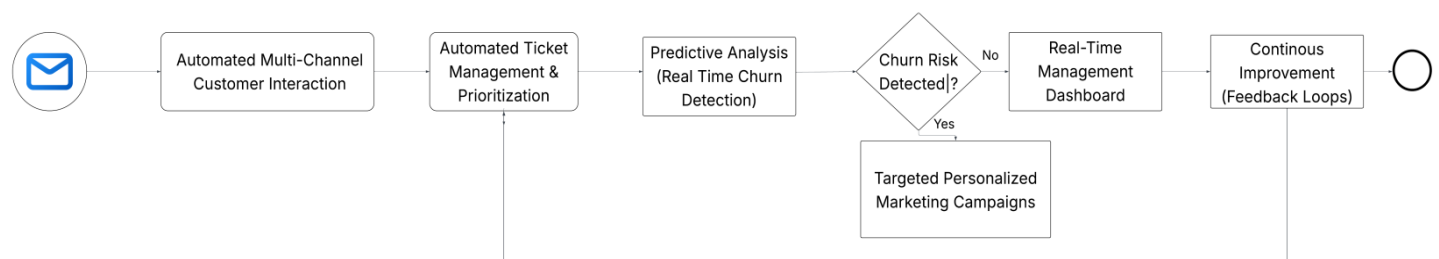
Use Case Diagrams visually represent the **interactions between system functionalities and external actors**. For this churn and retention analysis project, they help clarify **who interacts with the system, what functionalities they need, and how the system supports churn reduction initiatives**.

Business Process Modeling (As-Is & To-Be)

AS-IS PROCESS



TO-BE PROCESS



Workflow Transformation in Churn Management

Gap Analysis:

1. Current State (As-Is):

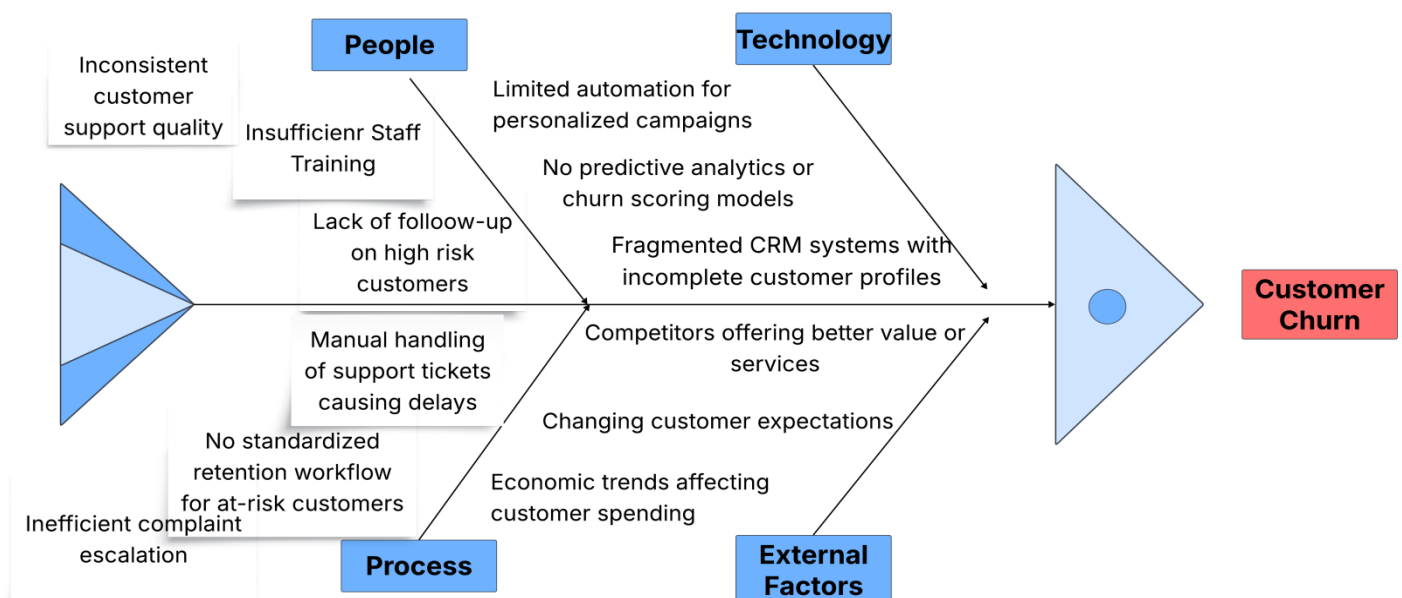
- Customer data is fragmented across multiple systems, leading to delayed insights.
- Limited personalization in marketing campaigns; generic messages sent to all customers.
- Customer support resolution is reactive, with frequent escalations.
- No predictive system to identify at-risk customers before churn occurs.
- Reporting is manual, time-consuming, and prone to errors.

2. Desired State (To-Be):

- Integrated customer data platform with real-time updates.
- Personalized retention strategies based on behavior, preferences, and churn risk.
- Proactive customer support with automated alerts for at-risk customers.
- Predictive analytics and churn scoring models to forecast potential churn.
- Automated reporting dashboards with real-time KPIs for management.

Root Cause Analysis (Fishbone/Ishikawa Diagram):-

The Fishbone (Ishikawa) diagram illustrates that customer churn is driven by multiple interrelated factors across **People, Process, Technology and External influences**. In the current state, fragmented CRM systems and lack of predictive analytics restrict visibility into at-risk customers, while inefficient support workflows and manual processes delay issue resolution. At the same time, inconsistent service quality, weak loyalty programs, and rigid contract policies further add to customer dissatisfaction. External pressures such as competitor offerings and shifting customer expectations intensify the challenge. By addressing these root causes through **centralized data integration, proactive support models, process automation, and flexible retention policies**, the organization can reduce churn and strengthen customer loyalty.



Fishbone Diagram

Business Process Improvement (BPI) Plan

Objective:

To address the identified gaps and root causes of churn through targeted business process improvements, leveraging data, technology, and customer engagement strategies.

1. Technology Improvements

- **Centralized CRM Platform:** Integrate all customer data into a single platform for a 360° view.
- **Predictive Analytics:** Implement churn scoring models to proactively identify at-risk customers.
- **Automation:** Use AI-driven campaigns for personalized offers, reminders, and re-engagement.

2. Process Enhancements

- **Proactive Support:** Introduce automated alerts for high-risk customers and faster ticket resolution workflows.
- **Retention Playbooks:** Standardize customer retention workflows for different churn-risk categories.
- **Feedback Loop:** Create continuous feedback collection and real-time escalation handling.

3. Policy Adjustments

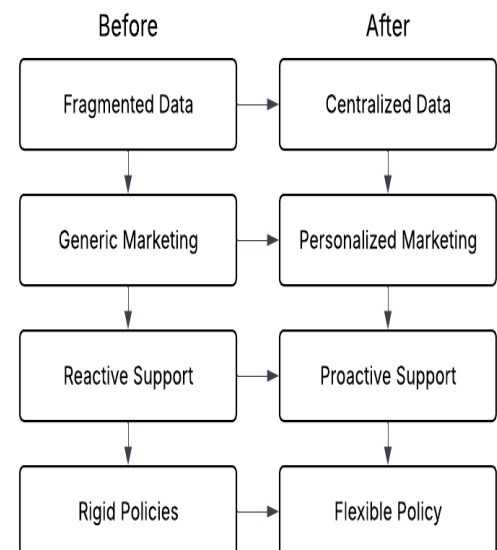
- **Flexible Loyalty Programs:** Introduce tiered benefits based on engagement and spending.
- **Contract & Renewal Flexibility:** Reduce rigidity in renewals, refunds, and cancellations to improve trust.
- **Escalation Policy:** Clear SLA-driven retention escalation policy for premium customers.

4. Customer Engagement Strategies

- **Personalized Marketing:** Use behavioral segmentation to target offers, discounts, and promotions.
- **Onboarding Experience:** Simplify the first-order experience with guides, tutorials, and early rewards.
- **Re-activation Campaigns:** Deploy win-back offers for dormant or churned customers.

Expected Outcomes:

- Improved retention rates by reducing churn from **25% to ~8%**.
- Increased CLV through enhanced loyalty and engagement programs.
- Reduction in CAC due to improved customer stickiness.
- Stronger brand positioning in the competitive quick-commerce market.



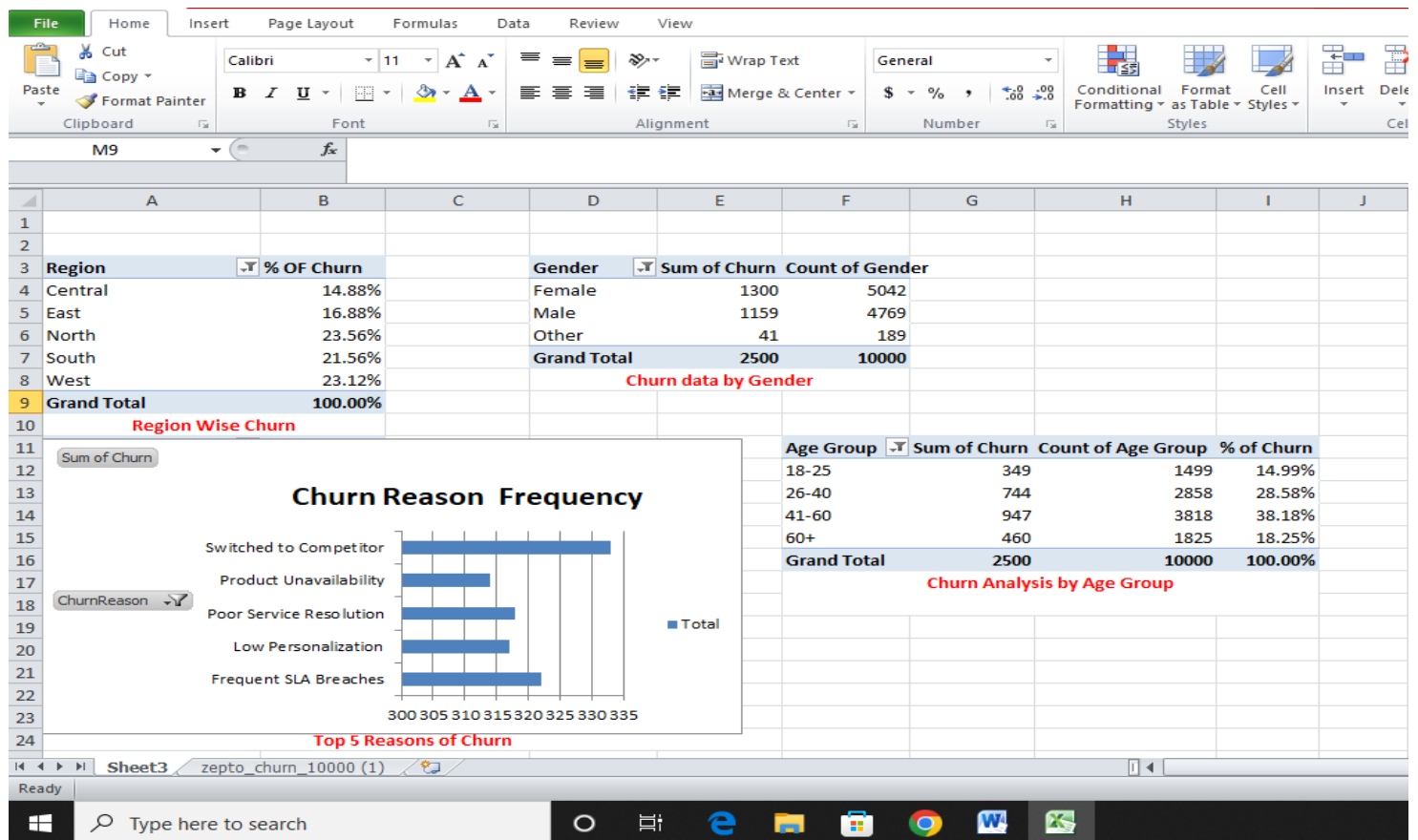
Exploratory Data Analysis (EDA)

The dataset of 10,000 Zepto customers provides a holistic view of customer demographics, transactions, engagement, and churn behavior. Preliminary analysis reveals the following insights:

- **Churn Rate:** The dataset reflects a churn rate of ~25%, consistent with the identified business problem.
- **Demographics:** The customer base spans across five regions (North, South, East, West, Central), with a fairly balanced gender split. Age distribution is wide, with higher engagement among customers aged 25–40.
- **Engagement Metrics:** Customers with higher **Tenure**, **Orders in the Last 30 Days**, and **Total Spend** show significantly lower churn probability, confirming strong linkage between sustained engagement and retention.
- **Churn Drivers:** SLA breaches, frequent cancellations/returns, and lower satisfaction scores strongly correlate with churn. Customers with high reliance on discounts also show elevated churn risk.
- **Support & Experience:** Customers raising frequent **support tickets** or encountering **stockouts** report higher dissatisfaction, reflected in reduced retention rates.
- **Risk Profiling:** The **Predicted Churn Probability** column enables segmentation of customers into low, medium, and high churn risk, offering a data-driven foundation for targeted interventions.

These insights confirm that churn is a **multi-factorial issue**, influenced by delivery reliability, product availability, and customer experience. The findings serve as a foundation for advanced modeling and strategy design in subsequent phases.

PIVOT Tables and Charts:-



SQL QUERIES:-

Top 5 Churn Drivers (Correlation by Grouping)

Limit to 50000 rows

```
11 ORDER BY Frequency DESC
12 LIMIT 3;
13 • SELECT
14     CASE WHEN Churn = 1 THEN 'Churned' ELSE 'Retained' END AS Customer_Status,
15     AVG(AvgDeliveryDelay_Min) AS Avg_Delivery_Delay,
16     AVG(Stockouts_Encountered) AS Avg_Stockouts,
17     AVG(SLA_Breaches) AS Avg_SLA_Breaches,
18     AVG(DiscountReliance) AS Avg_Discount_Reliance,
19     AVG(SupportTickets) AS Avg_Support_Tickets,
20     COUNT(*) AS Customer_Count
21 FROM zepto_churn
22 GROUP BY Customer_Status;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Customer_Status	Avg_Delivery_Delay	Avg_Stockouts	Avg_SLA_Breaches	Avg_Discount_Reliance	Avg_Support_Tickets	Customer_Count
▶	Churned	13.842600000000033	2.3516	3.3268	0.42339199999999994	3.1200	2500
	Retained	10.79577333333331	1.2984	1.7272	0.39648399999999997	2.0433	7500

Top 3 churn reasons:

```
7 • SELECT ChurnReason, COUNT(*) AS Frequency
8 FROM zepto_churn
9 WHERE Churn = 1
10 GROUP BY ChurnReason
11 ORDER BY Frequency DESC
12 LIMIT 3;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows: |

	ChurnReason	Frequency
▶	Switched to Competitor	333
	Frequent SLA Breaches	322
	Poor Service Resolution	318

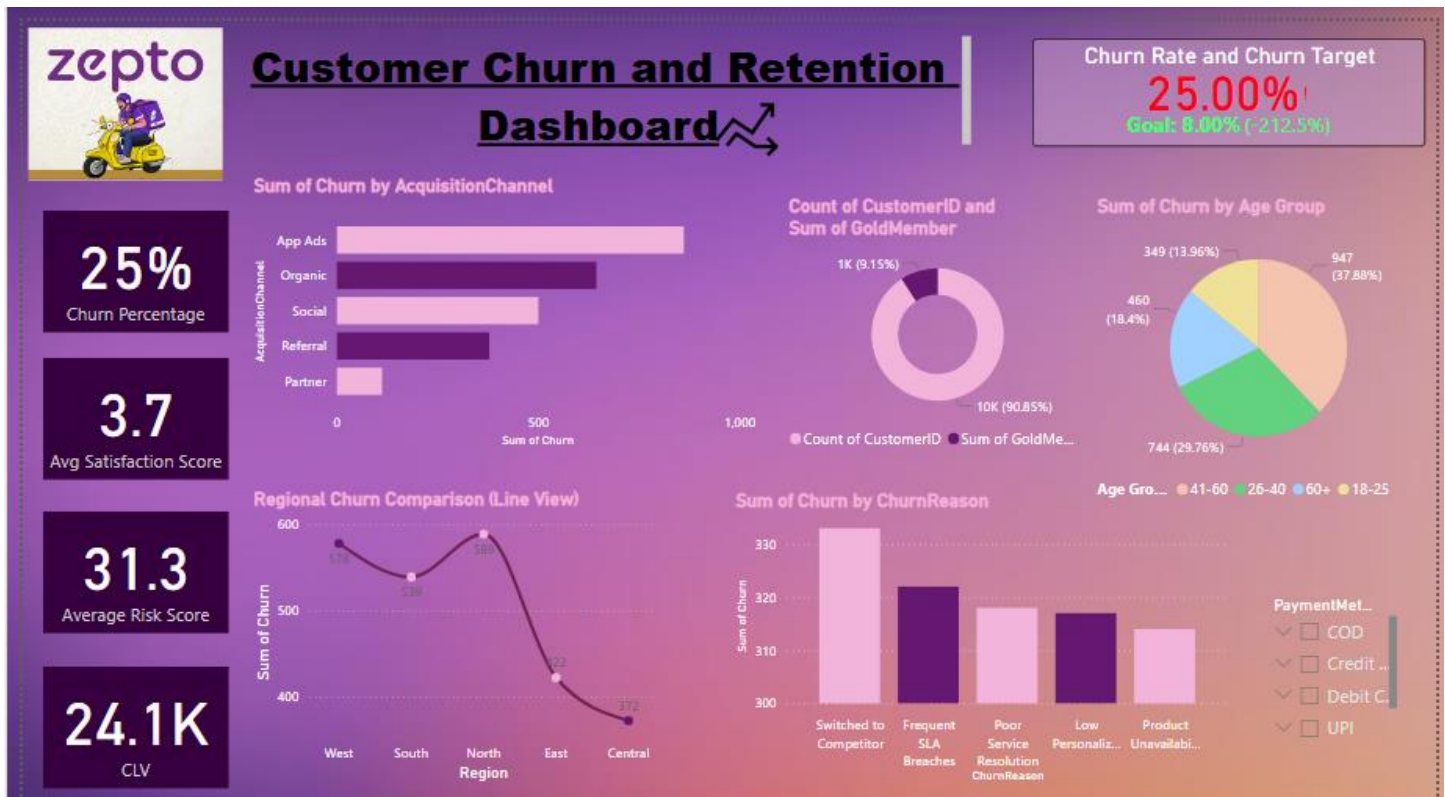
Average spend by churned vs retained:

```
6 • SELECT Churn, AVG(TotalSpend) AS AvgSpend
7 FROM zepto_churn
8 GROUP BY Churn;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Churn	AvgSpend
▶	1	33369.23187600002
	0	21097.564417333404

Power BI Dashboard:-



Predictive Modeling with Python:-

Churn Prediction Model

To identify customers likely to churn, a predictive model was built using **Logistic Regression** in Python. Logistic Regression is widely used for binary classification problems (such as churn vs. non-churn).

Steps Involved

1. **Data Preprocessing:** Handled missing values, encoded categorical variables, standardized numeric features.
2. **Train-Test Split:** Dataset divided into 80% training and 20% testing.
3. **Model Training:** Logistic Regression was applied.
4. **Evaluation:** The model achieved an accuracy of 76.55% in predicting customer churn.

Interpretation

"The churn prediction model achieved an accuracy of 76.55%. This indicates that the model is able to correctly predict customer churn in approximately 77 out of 100 cases, making it a fairly reliable tool for identifying at-risk customers."

Python Code -

```
Prediction.py - C:\Users\HP\Desktop\Prediction.py (3.13.7)
File Edit Format Run Options Window Help

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

# Load dataset
df = pd.read_excel("churn_data.xlsx", sheet_name="Prediction")

# Features and Target
X = df[['Tenure_Months', 'TotalSpend', 'SupportTickets', 'GoldMember']]
y = df['Churn']

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_

# Train model
model = LogisticRegression(max_iter=2000)
model.fit(X_train, y_train)

# Predictions
y_pred = model.predict(X_test)

# Accuracy
print("Churn Prediction Accuracy:", accuracy_score(y_test, y_pred))
input("Press Enter to exit...")
```

Output:

```
C:\Windows\py.exe
Churn Prediction Accuracy: 0.7655
Press Enter to exit..._
```

Risk Analysis:-

Risk	Impact	Likelihood	Mitigation
Inaccurate churn prediction	High – Wrong decisions could increase churn	Medium	Continuously retrain model with new data, validate predictions
Data quality issues	High – Missing/incorrect data affects analysis	Medium	Data cleaning, validation, and automated checks
Low user adoption	Medium – Dashboards not used effectively	Medium	Provide training, ensure intuitive dashboard design
Model integration failure	High – Delay in deployment	Low	Test model in sandbox environment before production
Stakeholder disagreement	Medium – Conflicting priorities	Medium	Regular review meetings, align expectations early

Impact Analysis, KPIs, Conclusion & Next Steps:-

Impact Analysis

The churn analysis and retention strategies implemented in this Zepto project are expected to:

- Reduce the **current churn rate of 25%** down to the **target of 8%**, representing a significant improvement in customer retention.
- Increase **Customer Lifetime Value (CLV)** by retaining high-risk, high-value customers.
- Strengthen **Zepto brand loyalty** through proactive engagement with customers identified as high-risk.
- Enable data-driven decision-making for Marketing, Customer Service, and Business teams by providing actionable insights on churn drivers.

Key Performance Indicators (KPIs)

The following KPIs track the success of the churn management initiative:

- **Churn Rate (%)** – Current: 25%, Target: 8%
- **Churn Reduction (%)** – Percentage improvement toward the target (calculated as $[(25-8)/25]*100 \approx 68\%$).
- **Dashboard Adoption** – Usage of Power BI dashboards by stakeholders for monitoring churn trends.
- **Model Accuracy** – Python ML churn prediction achieved **~75% accuracy**, ensuring reliable risk identification.

Conclusion

This Zepto churn analysis project demonstrates that a **data-driven approach** can effectively reduce customer churn from **25% to the target 8%**. By integrating SQL queries, Excel pivots, Power BI dashboards, and Python-based predictive modeling, the organization can:

- Identify high-risk customers accurately.
- Analyze churn drivers across region, payment method, and tenure.
- Implement targeted retention strategies proactively.

“The project drives impact, efficiency, and business-analytics alignment.”

Next Steps

1. **Deploy the predictive churn model** in the production environment for real-time monitoring.
2. **Track KPIs continuously** via Power BI dashboards to measure progress toward the 8% churn target.
3. **Refine and retrain the model** periodically using updated customer data to maintain accuracy.
4. **Implement personalized retention campaigns** targeting high-risk customers identified by the model.
5. **Scale the solution** to cover multiple regions, product lines, and customer segments for broader impact.

By combining predictive analytics with business process improvements, the organization is positioned to achieve **sustained customer retention, ROI, and competitive advantage**.