Customer Lifetime Value & Churn Analysis

---

1. Project Overview

This project analyzes transactional retail data to identify high-value customers, detect churn risk, and evaluate revenue concentration using RFM (Recency, Frequency, Monetary) modeling.

The objective is to help businesses:

- Maximize customer lifetime value
- Reduce revenue loss due to churn
- Identify revenue concentration risks
- Implement data-driven retention strategies

---

2. Dataset Summary

- Source: Online Retail II Dataset
- Time Period: 2009–2011
- Records: ~1 Million transactions (after filtering)
- Customers Identified: 5,878 unique customers
- Database Used: PostgreSQL
- Visualization Tool: Power BI

---

3. Data Engineering (Python)

3.1 Data Loading

- Combined 2009 and 2010 sheets
- Used pandas for ingestion

3.2 Data Cleaning

- Removed missing Customer IDs
- Removed negative quantities (returns)
- Removed zero-price transactions
- Converted InvoiceDate to datetime
- Created TotalAmount column

3.3 Feature Engineering

Created RFM metrics:

- Recency → Days since last purchase
- Frequency → Number of invoices
- Monetary → Total revenue per customer

3.4 RFM Scoring

Used quintile scoring:

- r_score
- f_score
- m_score
- Combined into rfm_score

3.5 Segmentation Logic

Customers categorized into:

- Champions
- Loyal Customers
- At Risk
- New Customers
- Others

3.6 Business Metrics Created

- Average Order Value
- CLV (simple monetary-based)
- Churn Flag (Recency > 90 days)
- Revenue At Risk

---

4. Database Integration (PostgreSQL)

The RFM dataset was loaded into PostgreSQL using SQLAlchemy.

Table: `customer_rfm`

---

5. Business Analysis using SQL

Q1. Total Customers

5,878 unique customers

Q2. Total Revenue

₹17.74 Million total lifetime revenue

Q3. Revenue Contribution by Segment

- Champions contribute ~47% of total revenue
- Revenue is highly concentrated

Q4. Customer Distribution

- Majority customers fall under “Others”
- Champions are a small but high-impact segment

Q5. Churn Rate by Segment

- At Risk segment → 100% churn
- Others → 71% churn
- Champions → 0%

Q6. Revenue at Risk

₹3.47 Million revenue exposed to churn

Q7. Revenue Concentration (Pareto Analysis)

Top 23% of customers generate 80% of total revenue.

Insight:
Revenue is highly concentrated among a small percentage of customers, creating dependency risk if high-value customers churn.

---

6. Dashboard (Power BI)

Built an executive-level dashboard including:

KPI Cards

- Total Customers
- Total Revenue
- Churn Rate %
- Revenue At Risk

Visualizations

- Revenue Contribution by Segment (Donut)
- Customer Distribution (Bar)
- Revenue at Risk by Segment (Column)
- Churn Rate by Segment
- Top 7 High-Value Customers Table
- Segment Filter

---

7. Key Business Insights

1. Revenue concentration risk is high.
2. 20% of total revenue is exposed to churn.
3. Champions are small in number but drive almost half of revenue.
4. At Risk and Others segments require proactive retention.
5. Business heavily depends on elite customers.

---

8. Business Recommendations

- Implement loyalty programs for Champions.
- Target At Risk customers with personalized retention campaigns.
- Monitor revenue concentration monthly.
- Introduce predictive churn modeling in next phase.
- Reduce dependency risk by expanding mid-value segments.

---

9. Future Improvements

- Build ML churn prediction model
- Advanced CLV prediction using probabilistic models
- Deploy dashboard using Power BI Service
- Automate pipeline with Airflow