1. ABSTRACT

The rapid expansion of e-commerce has generated massive amounts of customer and transaction data. To remain competitive, companies must understand customer behavior, predict future trends, and make data-driven decisions. This project presents an **Enterprise E-Commerce Analytics Platform** that performs customer segmentation, churn risk prediction, Customer Lifetime Value (CLV) estimation, and segment-wise revenue forecasting. Raw data is cleaned and transformed through a structured feature engineering pipeline. Machine learning models are used to calculate churn probability and forecast future monthly revenue. An interactive Streamlit dashboard integrates all insights into five modules: Executive Overview, Customer Segments, Churn Analysis, Action Recommendations, and Revenue Forecasting. This platform replicates real-world analytics systems used by e-commerce companies and demonstrates a complete end-to-end data science workflow.

2. INTRODUCTION

E-commerce businesses deal with large volumes of customers who behave differently in terms of purchasing frequency, spending, loyalty, and churn tendencies. Without proper analytics, businesses lose potential revenue due to:

- Lack of customer understanding

- Poor targeting

- Inefficient retention strategies

- Inaccurate forecasting

This project aims to solve these issues by building a **complete analytical ecosystem** that helps businesses monitor customer behavior, estimate future value, and prevent churn.

The developed platform is similar to what real companies such as Amazon, Flipkart, Nykaa, and Meesho use internally to drive business decisions.

3. PROBLEM STATEMENT

E-commerce businesses deal with large volumes of customers who behave differently in terms of purchasing frequency, spending, loyalty, and churn tendencies. Without proper analytics, businesses lose potential revenue due to:

- Lack of customer understanding

- Poor targeting

- Inefficient retention strategies

- Inaccurate forecasting

This project aims to solve these issues by building a **complete analytical ecosystem** that helps businesses monitor customer behavior, estimate future value, and prevent churn.

The developed platform is similar to what real companies such as Amazon, Flipkart, Nykaa, and Meesho use internally to drive business decisions.

4. OBJECTIVES

E-commerce companies face the following challenges:

1. Difficulty in identifying high-value customers

2. Inability to detect customers who are likely to stop buying (churn)

3. Lack of clear segmentation for targeted marketing

4. No accurate method to forecast future revenue trends

5. Poor understanding of customer lifetime value

The project addresses these issues through a structured analytical and predictive system.

5. METHODOLOGY

5.1 Data Collection

Raw datasets (orders, customers, spending details) were stored in:

core/data/raw/

5.2 Data Cleaning

Tasks performed:

- Removing duplicates

- Handling missing values

- Converting date formats

- Fixing invalid entries

- Converting data types

Output:

orders_cleaned.csv

5.3 Feature Engineering

Created features:

- **Recency:** Days since last purchase

- **Frequency:** Number of orders

- **Monetary Value (M):** Total amount spent

- **Average Order Value**

- **Monthly Spending Patterns**

- **CLV Proxy (RFM-based)**

Outputs:

customer_monthly_value.csv

customer_clv_features.csv

dashboard_data.csv

5.4 Customer Segmentation

Used CLV + RFM metrics to classify customers into:

- High Value

- Medium Value

- Low Value

This helps in targeted marketing and prioritization.

5.5 Churn Probability Calculation

A churn score was generated using:

- Recency

- Frequency

- Monetary metrics

- CLV

- Activity patterns

Customers were labeled:

- High Risk

- Medium Risk

- Low Risk

This helps identify customers who may stop buying soon.

5.6 Revenue Forecasting

Steps:

1. Aggregate monthly revenue for each segment

2. Apply **Linear Regression** for future forecasting

3. Predict revenue for the next 3 months

Save results as:

segment_revenue_forecast.csv

5.7 Dashboard Development (Streamlit)

Five tabs were created:

1. **Executive Overview**

2. **Customer Segments**

3. **Churn Risk Analysis**

4. **Action Recommendations**

5. **Revenue Forecast**

Each tab displays visualizations and data insights.

5.8 Deployment

- Version control using **GitHub**

- Live deployment using **Streamlit Cloud**

Result:
An accessible, fully functional analytics platform available online.

6. SYSTEM ARCHITECTURE

- Data ingestion layer

- Data cleaning & preprocessing layer

- Feature engineering layer

- Modeling layer (churn + forecasting)

- Visualization & UI layer (Streamlit)

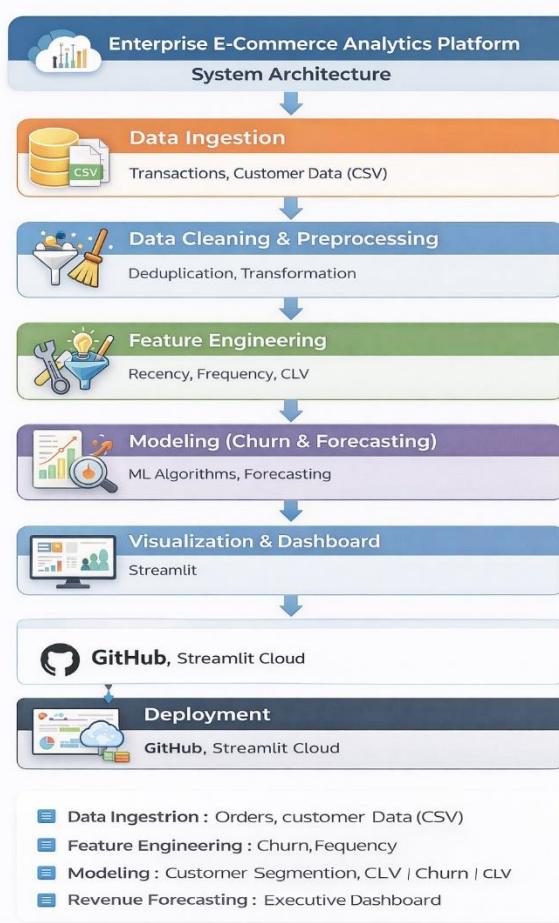- Deployment layer (GitHub + Cloud)

Figure 1: System Architecture

7. RESULTS

1. **Customer Segmentation**
   - Clear separation into High, Medium, Low value groups
   - Helps identify profitable customers

2. **Churn Analysis**
   - Customers with highest churn probability identified
   - Helps retention teams save customers early

3. **Revenue Forecast**
   - Predicts upcoming monthly revenue
   - Useful for budgeting and planning

4. **Executive Dashboard**
   - Real-time insights in a simple interface

- Suitable for CEOs, marketing, and strategy teams

8. ACTIONABLE INSIGHTS

- High-value + high-risk customers need **immediate retention.**

- Low-value customers require **automated campaigns.**

- Segments with rising revenue should be **promoted more.**

- Segments with falling revenue need **intervention.**

- Churn probability can guide **email and discount strategies.**

9. CONCLUSION

The Enterprise E-Commerce Analytics Platform successfully demonstrates a complete data science solution, from raw data processing to machine learning and interactive visualization. The system closely resembles real-world analytics platforms used in e-commerce companies and provides valuable insights for improving customer retention and forecasting revenue. The project highlights practical knowledge in data engineering, feature engineering, modeling, dashboarding, and cloud deployment.

10.FUTURE SCOPE

- Add advanced ML models (XGBoost, LightGBM)

- Implement personalized customer recommendations

- Build an automated email retention system

- Add product-wise forecasting

- Create a real-time data pipeline

- Integrate a database like MySQL or MongoDB

- Convert the app into a SaaS platform

- Add RFM-based automated segmentation