RetailSense: A Data-Driven Approach to Retail Optimization

A Proposal report for the BDM capstone Project

Submitted by

Name: Kandagatla Uday Kiran

Roll number: 22f3001962

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

Declaration Statement

I am working on a Project titled "**RetailSense: A Data-Driven Approach to Retail Optimization**". I extend my appreciation to **UC Irvine Machine Learning Repository**, for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered from primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the principles of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I understand that all recommendations made in this project report are within the context of the academic project taken up towards course fulfillment in the BS Degree Program offered by IIT Madras. The institution does not endorse any of the claims or comments.

Signature of Candidate:

Name: Kandagatla Uday Kiran

Date: 08/05/2025

1 Executive Summary and Title

RetailSense: A Data-Driven Approach to Retail Optimization

The organization is a UK-based online retail organization that sells gift and homeware products on its e-commerce platform to B2C and B2B customers in multiple countries.

However, the company is experiencing operational challenges including inconsistent customer retention, a high incidence of returns, and no actionable segmentation or bundling approaches. An initial review of the data highlighted a very high percentage of transactions that were missing Customer IDs, as well as a very high number of returns, indicating a data quality issue and a significant risk of leakage from revenue.

This project uses a methodology to take the data and apply an analytical framework to the issues. For example, RFM analysis will be used to group customers by purchasing behavior, cohort analysis will be used to assess retention, market basket analysis was used to determine products which were frequently purchased together, and return analysis was used to identify products with high return counts, customer segments with high return counts, and trends based on monthly sales data for demand planning.

The results allow for actionable strategies to improve retention, agree on quantities and types of inventory based on customer preferences, reduce losses caused by returns, and increase revenues seized from more effective marketing and/or reduced bundling—demonstrating the potential usefulness of data driven decision making in a retail context.

2 Organization Background

The company is an online retail business based in the UK, selling distinctive gift items, decorative products, and homeware items. The company is primarily an e-commerce business that serves individual and wholesale (B2B) customers including international customers, with product sales all over Europe.

As a niche player in the online retail industry, this organization offers unique products and seasonal items that are aesthetically-based instead of offering widely available or conventional, traditional, and common items. The business model relies heavily on buying in bulk and bulk sales with expiry campaigns that must occur within designated timeframes. When engaging in bulk selling, it is critical that this organization accurately anticipates demand and manages inventory accordingly to maximize return on sales.

The organization is growing but has growing operational issues including fluctuating sales patterns, poor customer segmentation and customer returns that are excessive, which are affecting profitability and long-term customer retention. In a competitive e-commerce retail environment, the organization is seeking avenues to utilize data-driven methods for decision making, optimize resources while maximising return on resources and achieve sustained growth in the global e-commerce market.

3 Problem Statement

The purpose of this project is to use the power of data analytics to address key operational and marketing issues, including how to segment customers by behavior and value; how to create product bundles with likely product combinations; and how to assess product returns to offset losses--to name a few. Addressing these problems will help the firm be more strategic in targeting customers, drive improvements in sales performance, and increase operational efficiencies.

3.1 Problem statement 1:

The business lacks a data-driven method to classify customers based on value, frequency, and engagement, leading to generic, untargeted marketing efforts.

3.2 Problem statement 2:

There is no current system to identify frequently bought together products, missing cross-selling and bundling opportunities.

3.3 Problem statement 3:

Frequent product returns from certain items and customers result in lost revenue and increased logistical costs.

3.4 Problem statement 4:

The business lacks visibility into customer retention over time and cannot identify when and why customers churn.

4 Background of the Problem

Online retail is a fast-paced, data-driven environment where customer retention, inventory management, and profitability are of utmost importance. The organization has accumulated much transaction data, but it still suffers from persistent issues.

First, it is clear there are not enough incentivization, or retention tactics, since most customers only make one purchase from the store. The business has no structured segmentation so it cannot tell if a customer is a loyal or high-value customer or a one-time buyer. Thus, they are not able to market to people sub-segmented in a way that helps better the marketing efforts.

Second, customers repeatedly purchased multiple items in a transaction, yet analysis has not been conducted to determine product pair recommendations. This lack of analysis does not allow for the creation of a profitable bundle or recommend a bundle to the consumer.

Returns are another burdensome problem. Returns at a high rate, especially on certain products, creates inventory that is lost until resold. Less than positive return experiences from the customer can come from quality problems, expectations not aligned with the product's reality or as well customer instruction. Returns and solicitations can result in customer dissatisfaction and even fraud matters as well.

Fluctuating trends of sales related to seasonality and demand variability creates likewise problems with demand planning and inventory control. The business experiences stockouts at peak demand and excess at off-peak. These both compromise revenue for the retailer.

Addressing these issues requires a systematic analytical approach using the organization’s own data to drive decisions. By identifying patterns in purchasing, returns, and customer behavior, the company can adopt smarter strategies for long-term business growth.

5 Problem Solving Approach

I used a systematic data science approach to tackle the mentioned business problems using data scrubbing, segmentation, association mining and trends analysis on the UCI Online Retail data set.

1. Data Cleaning & Feature Engineering

I clean data to remove cancellations orders (negative quantity), update null values in the customer and product fields and remove duplicates. Features were created include, per transaction revenue, invoice month, and customer tenure.

2. RFM Analysis

I utilized a Recency, Frequency and Monetary (RFM) approach to segment customers by their recency, frequency and monetary metrics to determine loyal customers, frequent buyers, and at-risk segments for our campaigns.

3. Cohort analysis

I grouped customers into cohorts based on their first month purchased. I then tracked how many of the customers from each cohort returned in subsequent measures, which was a visual representation of retention and drop out trends.

4. Market Basket Analysis

Using the Apriori algorithm, I analyzed the transaction records to identify which products are frequently purchased in combination. The association rules obtained from the data showed product pairs with high-confidence selected which can aid in product and bundle offer design and future personalized product recommendations or promotions.

5. Return Analysis

By filtering the data for negative quantities I identified the products and customers associated with the most returns. Identifying the products returned frequently corralled quality issues and could help highlight potential changes in policies that encourage returns, creating a loss of revenue.

6. Trend Visualization

I aggregated monthly revenue and plotted sales trends to monitor for growth and seasonal trends. This aids in predicting demand and planning stock levels accordingly.

7. CLV and Pareto Analysis

I approximated customer lifetime value (CLV) through analysis of frequency and monetary value. The Pareto analysis indicated the top 20% of customers and products represented 80% of the revenue which is important for focusing business strategies.

By systematically following the approach, I have successfully addressed each core potential business problem—retention, segmentation, bundling, returns, and 'future' trend reporting; and provided points of action. The methodologies adopted best-practice and/or industry benchmarks provide no need for additional interpretation. Furthermore the methods used can be relatively easily applied/ scaled to other datasets or applied in living operation.

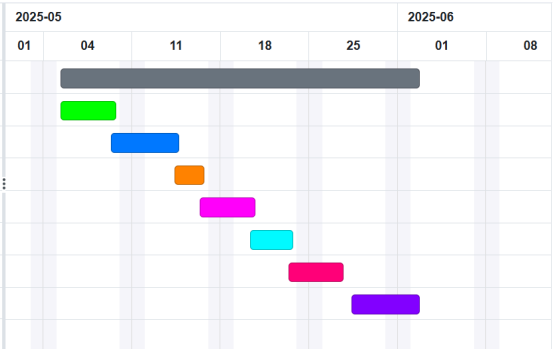
6 Expected Timeline

6.1 Work Breakdown Structure:

Task name	Start	End	Duration
Proposal	05/05/2025	09/05/2025	5 Days
Data Cleaning & Feature Creation	09/05/2025	14/05/2025	4 Days
EDA	14/05/2025	16/05/2025	3 Days
RFM & Cohort Analysis	16/05/2025	20/05/2025	3 Days
Market Basket + Return Analysis	20/05/2025	23/05/2025	4 Days
Forecasting & Visualization	23/05/2025	27/05/2025	3 Days
Final Report	28/05/2025	02/06/2025	4 Days

6.2 Gantt chart

Task Name	Start	End	Duration
BDM PROJECT	2025-05-05	2025-06-02	21 days
Proposal	2025-05-05	2025-05-09	5 days
Data Cleaning & Feature Creation	2025-05-09	2025-05-14	4 days
EDA	2025-05-14	2025-05-16	3 days
RFM & Cohort Analysis	2025-05-16	2025-05-20	3 days
Market Basket + Return Analysis	2025-05-20	2025-05-23	4 days
Forecasting & Visualization	2025-05-23	2025-05-27	3 days
Final Report	2025-05-28	2025-06-02	4 days



7 Expected Outcome

The analysis and forecasting techniques applied to the **Online Retail dataset** will provide valuable insights to improve business decision-making.

7.1 Customer Segmentation:

RFM (Recency, Frequency, Monetary) analysis will segment customers to determine who is loyal, who spends high dollars, and who should be targeted for marketing as "at risk" users.

7.2 Retention Insights:

Cohort analysis will determine how customer retention changes over time and when the majority of customers tend to churn so that engagement can be designed around those timeframes to either improve retention or decrease churn to create long-term loyal retention.

7.3 Product Bundling Opportunities:

Market Basket Analysis will be used to identify product combinations that frequently get purchased to create bundles of products and potentially increase average order value.

7.4 Return Pattern Identification:

Return Analysis will identify the product and the customer that tends to return the most amount of items so that the company can either review product quality, refine return policies, or mitigate loss (via damage and restock fees).

7.5 Sales Trend Visualization:

Monthly sales analysis to display seasonal sales patterns and business trends for the growth of the company will define planning for inventory or forecasts for demand.

7.6 Strategic Business Recommendations:

The final deliverable will summarize the analysis and provide necessary recommendations for driving customer retention, effective marketing, identified losses associated with return to mitigate in the future, and overall improved profitability.