Sales Forecasting and Revenue Analysis For E-Commerce Business

A Proposal report for the BDM capstone Project

Submitted by

Name: Swaraj Choudhary
Email:23f1001245@ds.study.iitm.ac.in
Roll number: 23f1001245

IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

Contents

Declaration Statement

I am undertaking a project titled "Sales Forecasting and Revenue Analysis For E-Commerce business" The dataset used for this project has been sourced from Kaggle, a well-known platform for data science and machine learning datasets. The original dataset is available at: https://www.kaggle.com/datasets/carrie1/ecommerce-data .
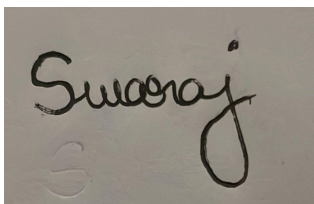
I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered through Secondary source (Kaggle) and carefully analyzed to assure its reliability.

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 information 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 agree that all the recommendations are business-specific and limited to this project exclusively, and cannot be utilized for any other purpose with an IIT Madras tag. I understand that IIT Madras does not endorse this.

Signature of Candidate:

Name: Swaraj Choudhary

Date: 10/07/2025

1   Executive Summary

The business for this project is based on a Secondary Dataset sourced from Kaggle (https://www.kaggle.com/datasets/carrie1/ecommerce-data) which contains transactional records from a UK based online retail company operating in the B2C (business-to-consumer) domain. The business specializes in selling a wide range of gift items and household goods through its e-commerce platform serving customers across multiple countries by selling a wide range of gift items and household goods through its e-commerce platform.

As mid-sized online retailers, the business likely faces challenges such as demand forecasting, product performance evaluation, and inventory optimization. A lack of predictive insights leading to inefficient stocking, and pricing, affecting both profitability and customer satisfaction. The dataset spans December 2010 to December 2011 and includes fields like invoice number, item description, quantity, price, customer location, and purchase date enabling detailed trend analysis.

In order to address these challenges, I will be utilizing tools like Python, Excel, and libraries like Matplotlib and Seaborn for data analysis and visualization. A regression model will be used to forecast sales quantities, while a classification model will identify high-performing products with the expected outcome to generate data driven insights that support improved inventory planning, pricing strategies, and profitability for the e-commerce business.

2   Organization Background

The organization analyzed in this project is a mid-sized e-commerce retailer based in the United Kingdom, operating in the gifts and home decoration industry. The company specializes in selling a wide range of household products, decorative items, party supplies, and novelty gifts through its online platform. It caters primarily to individual customers across the UK and parts of Europe, functioning entirely in the business-to-consumer (B2C) space without physical retail stores.

The business recorded a total revenue of approximately £10.6 million from December 2010 to December 2011, averaging around £35,000 in daily revenue. The highest revenue months include December 2010 (£823,746) and March 2011 (£717,639), indicating strong seasonal demand.

Despite the considerable revenue generation, the business lacks clarity on which products are consistently driving revenue, and faces challenges with inventory planning and demand forecasting. These issues contribute to lost sales opportunities, overstocking of slow-moving items, and understocking of high-demand products. Hence, to address this I will be using the transaction data to build forecasting and classification models that will improve inventory decisions, pricing strategies, and overall profitability.

3   Problem Statements

- Problem statement-1 : The business is experiencing frequent stockouts and excess inventory due to unpredictable customer demand, the company often runs out of popular products while overstocking rest of the items, leading to customer dissatisfaction and increased storage costs. Hence, there is a need for inventory planning and reduction in operational costs.

- Problem statement-2: The business is unable to predict how the product level features impact revenue. Without a data driven model to estimate revenues the business struggles to find the strategic pricing of products and planning to increase profit. Hence, a regression model is needed for revenue and profit forecasting and analysis.

4   Background of the Problem

- Since the business operates in the highly competitive online market, it becomes increasingly difficult to predict customer demand accurately. Seasonal changes in buyer behavior, competitor discounts, etc, make it challenging to maintain the right stock levels. As a result, the business often runs out of popular products or ends up overstocking items that don't sell, leading to higher storage costs and missed sales. This affects customer satisfaction and prevents the business from achieving consistent daily revenue.

- Also, based on the dataset the business lacks a clear system to provide insights into which factors truly influence revenue without which the pricing decisions, product promotions, and restocking efforts are all made without any solid strategy which is limiting the business's ability to grow profitably and make the most of its high-selling items.

- Hence, these challenges make it hard for the business to operate efficiently. To stay competitive and profitable, the company must move from guesswork to a more data-driven approach for managing inventory and forecasting revenue.

5   Problem Solving Approach

5.1   The dataset used in this project is a secondary dataset from Kaggle, titled "E-Commerce Data"(https://www.kaggle.com/datasets/carrie1/ecommerce-data) which contains over 500,000 transactional records from a UK-based online retail company operating in the B2C domain. The data spans from December 2010 to December 2011 and includes key fields such as: InvoiceNo, InvoiceDate, Description, Quantity, UnitPrice, CustomerID, etc making it ideal for our analysis.

5.2   To address the issue of stock imbalance caused by unpredictable demand, a time series a time series approach will be applied in which he data will be aggregated by product on a weekly basis to create time series for quantity sold. Forecasting methods such as Moving Averages, ARIMA will be used to detect seasonality and predict future demand which will help the business plan inventory more accurately, reduce overstocking, prevent stockouts, and improve customer satisfaction.

5.3   To understand what factors influence product-level revenue, a multiple linear regression model will be developed through which revenue will be calculated as the product of quantity sold and unit price then variables such as product type, unit price, quantity, and time of sale will serve as predictors in the model hence giving us insights on what drives revenue the most, helping the business with pricing, promotion, and product prioritization strategies aimed at increasing overall profit.

5.4   The primary tools used for this project will include Python in which libraries such as Pandas, Seaborn, Matplotlib, Statsmodels, and Scikit-learn will be used to support statistical modeling and visual analysis and Microsoft Excel will be used during the early stages for basic cleaning, summary statistics, and quick business-friendly exploration of trends. This combination of tools will help ensure both technical depth and practical usability of the insights generated.

6 Expected Timeline

6.1 Work Breakdown Structure
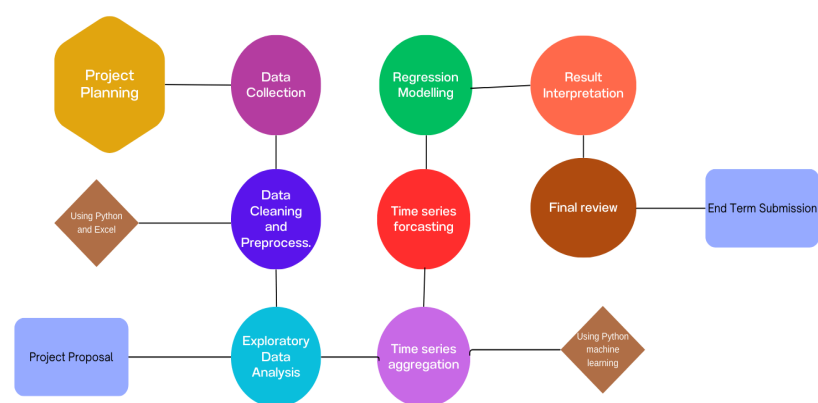
6.2 Gantt Chart
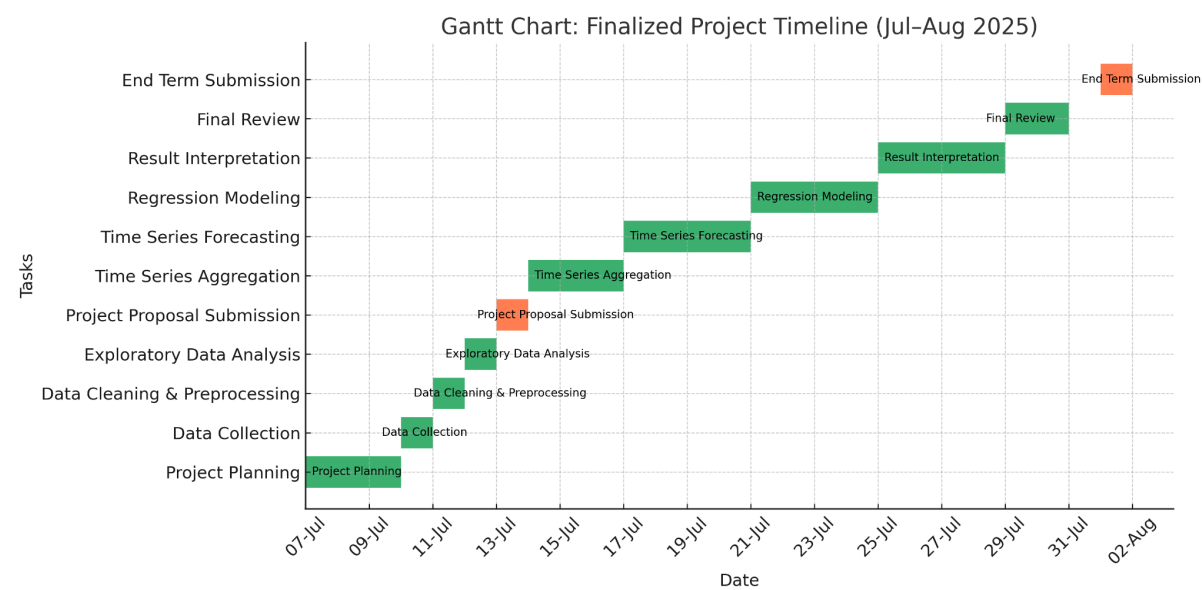
7   Expected Outcome

- This project will deliver a data-driven insight to support better decision-making in inventory planning and revenue optimization. The first key outcome is a time series forecasting model that will predict weekly product demand, helping reduce stockouts and overstocking, thus improving inventory efficiency.

- The second outcome will be a regression model to identify the primary drivers of product-level revenue, by analyzing the price, quantity, category, and time of sale, of the products the model will help the business make informed pricing and stocking decisions.

- Alongside the models, visual charts and summary reports will be generated using Python and Excel to highlight patterns and trends in customer behavior and product performance.

- All findings will be based entirely on the analysis of the Kaggle provided transactional dataset, ensuring that the final recommendations are reliable, measurable, and actionable for the business.