Business Data Analysis of E-commerce Sales

A Proposal report for the BDM capstone Project

Submitted by

Name: HARMEET KAUR

Roll number: 23F2001610

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 "**Business Data Analysis of E-commerce Sales**". I extend my appreciation to Kaggle, 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 secondary 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: (**Digital Signature**)

Name: HARMEET KAUR

Date: 11/11/2025

1 Executive Summary and Title

This project focuses on analyzing the yearly sales performance of a mid-scale Indian e-commerce company operating under a **Business-to-Consumer (B2C)** model. The company serves customers across multiple regions in India and offers a diverse product portfolio including Electronics, Clothing, Home and Kitchen items, Books, Sports equipment, Beauty products, and Toys. The study uses secondary data containing monthly transactional records with information on units sold, pricing, discount rates, revenue, and profit across different categories and regions.

The **Organization Background** outlines the company's multi-category operations and the challenges associated with managing sales and profitability in a competitive and seasonally influenced e-commerce environment. The **Problem Statement** identifies inconsistent profitability across product categories and significant fluctuations in monthly sales caused by seasonal demand variations, leading to pricing inefficiencies, inventory imbalance, and revenue instability. The **Background of the Problem** explains that aggressive discounting, decentralized pricing decisions, and limited use of analytical insights contribute to these issues.

To address these challenges, the **Problem-Solving Approach** adopts structured data cleaning, descriptive analysis, time-series evaluation, and correlation analysis. Key analyses include category-wise profit comparison, assessment of discount–profit relationships, and month-wise revenue trend evaluation. The **Expected Timeline** presents a phased project plan supported by a Work Breakdown Structure and an updated Gantt chart. The **Expected Outcomes** emphasize data-driven insights for improving business performance.

2 Organization Background

The organization under study is a fictional mid-scale Indian e-commerce company operating under a **Business-to-Consumer (B2C)** model. The company functions entirely through an online platform, offering a wide range of consumer products such as Electronics, Clothing, Home and Kitchen items, Books, Sports equipment, Beauty products, and Toys. It serves customers across multiple regions in India, catering to diverse demographic and geographic segments.

Customer purchasing behavior varies significantly due to factors such as regional preferences, income levels, festive seasons, and promotional campaigns. As a result, demand patterns are highly dynamic, with sharp sales peaks during festivals and discount events, and reduced demand during off-season periods. These fluctuations create challenges in maintaining consistent sales performance and effective inventory management.

The dataset used for this project is secondary data sourced from **Kaggle**, designed for analytical and educational purposes. It includes structured transactional records such as order dates, product categories, quantities sold, prices, discounts, revenue, profit, payment methods, and regional details. This dataset closely reflects real-world e-commerce operations and enables meaningful business data analysis.

The organization relies heavily on discount-based promotions to drive sales; however, the lack of systematic data analysis has resulted in inconsistent profit margins and inefficient inventory planning. This project aims to address these issues through structured data analysis to improve profitability, forecasting accuracy, and operational efficiency.

3 Problem Statement

3.1 Problem statement 1:

Despite stable overall sales volume, profit margins vary significantly across product categories. The business suspects that certain categories generate high revenue but yield lower profits due to higher discount rates and lower unit prices.

3.2 Problem statement 2:

The company's monthly sales show considerable fluctuations throughout the year. Peak months generate surges in demand, while certain periods exhibit sharp declines. These unpredictable shifts cause issues such as overstocking, stock-outs and revenue instability.

4 Background of the Problem

The company operates across multiple product categories such as Electronics, Clothing, and Home & Kitchen. While overall sales remain stable, sharp variations in profit margins across regions and product lines have been observed. Additionally, seasonal demand fluctuations cause unpredictable sales patterns, leading to challenges in inventory management and forecasting accuracy.

The primary causes include inconsistent pricing strategies, aggressive discounting practices, and inadequate use of sales and profit data for decision-making. Seasonal sales surges and off-season declines also disrupt the company's ability to maintain balanced inventory levels and steady cash flow.

Internally, the company struggles with inefficient data analysis, lack of integrated reporting systems, and limited use of predictive analytics. Regional managers often make independent pricing and discounting decisions without centralized oversight. Moreover, forecasting models fail to capture real-time market dynamics, leading to inaccurate sales predictions.

Externally, intense market competition, fluctuating consumer preferences, and economic factors such as inflation and changing purchasing power affect profitability. Seasonal demand shifts during festivals and promotional events further complicate inventory planning and profit stability across regions.

5 Problem Solving Approach

To address the company's challenges related to inconsistent profit margins and seasonal demand fluctuations, a structured data-driven problem-solving approach is adopted. This approach integrates systematic data collection, robust analytical methods, and appropriate software tools to extract meaningful business insights.

5.1 Data Collection and Data Cleaning

The project uses **secondary data sourced from Kaggle**, representing transactional records of an Indian e-commerce company. The dataset includes variables such as order dates, product categories, regions, units sold, selling prices, discount percentages, revenue, and profit values.

Data cleaning is a critical initial step to ensure accuracy and reliability. The dataset is examined for duplicate entries, missing values, and inconsistencies in date formats, category names, and regional labels. Data preprocessing involves removing duplicates, correcting data entry errors, and standardizing categorical and temporal fields. Missing values are handled using appropriate techniques such as averaging, filtering, or logical imputation. This step ensures the dataset is consistent and suitable for further analysis.

5.2 Analysis Methods

Descriptive analysis is conducted to summarize key metrics such as total sales, revenue, and profit across product categories, regions, and time periods. **Time series analysis** is applied to evaluate month-wise sales trends and identify seasonal demand patterns, peak periods, and off-season declines.

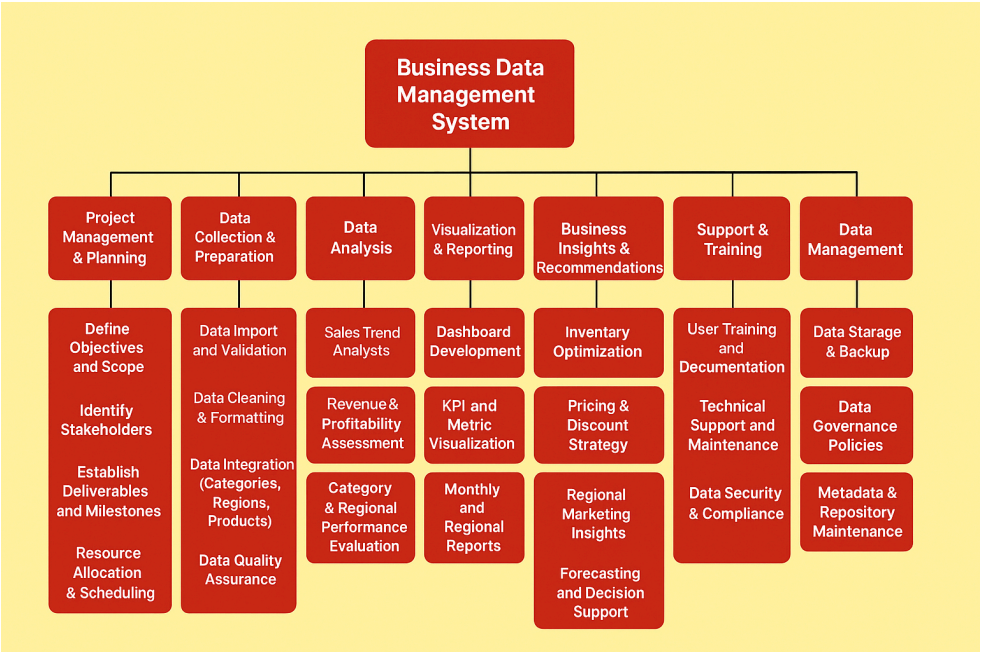
Additionally, **correlation analysis** is performed to examine the relationship between discount rates, sales volume, and profitability. This helps assess the impact of discounting strategies on profit margins. Trend and regression analysis are also used to understand how pricing and discount variables influence revenue and profit performance.

5.3 Tools Used

The analysis is carried out using **Microsoft Excel** and **Python**. Excel is used for data cleaning, PivotTables, visualization, forecasting, and dashboard creation, while Python supports data manipulation and trend analysis. These tools together enable effective data-driven decision-making and business insight generation.

6 Expected Timeline

6.1 Work Breakdown Structure:



6.2 Gantt chart

E-Commerce Project Gantt Chart													
Project Start Date	21-11-2025												
Display Week	1												
Project Lead:													
			Week 1(21-27 Nov 2025)									Week 2(28Nov-4Dec 2025)	
TASK	**START**	**END**	**21-Nov**	**22-Nov**	**23-Nov**	**24-Nov**	**25-Nov**	**26-Nov**	**27-Nov**	**28-Nov**	**29-Nov**	**30-Nov**	**01-Dec**
E-Commerce Project													
Collecting Data	21-11-2025	23-11-2025											
Data Cleaning	24-11-2025	25-11-2025											
Finding Insights	26-11-2025	29-11-2025											
Preparing Report	29-11-2025	07-12-2025											
Final Presentation													
Project Approval													

Expected timeline for completion of project.

7 Expected Outcome

The expected outcome of this Business Data Management project is to develop a comprehensive understanding of sales performance across product categories, regions, and time periods. Through the analysis of key metrics such as units sold, revenue, profit margins, and discount rates, the project aims to identify both high-performing and underperforming product categories. This understanding is expected to support improved inventory planning, optimized pricing decisions, and more efficient allocation of business resources.

The analysis will also reveal seasonal trends and month-wise sales patterns, enabling proactive decision-making during peak-demand and low-demand periods. By examining regional sales performance, the project will provide insights into geographic variations, helping the organization tailor marketing, distribution, and operational strategies more effectively. Furthermore, evaluating the relationship between discounts and profitability will allow the business to refine its promotional strategies in a way that maximizes revenue without adversely affecting profit margins.

Overall, the project is expected to enhance sales forecasting accuracy, reduce operational inefficiencies such as overstocking and stock-outs, and strengthen the organization's ability to make data-driven decisions. The insights generated will empower stakeholders to improve business performance, support strategic planning, and drive sustainable growth in a competitive e-commerce environment.