Optimizing Inventory and Service Offerings for a Multi-Service Retail Shop

Mid Term Report for BDM Capstone Project

Submitted by

Name: **Sneha Sarkar**

Roll number: **23f2004880**

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 **"Optimizing Inventory and Service Offerings for a Multi-Service Retail Shop"**. I extend my appreciation to **Cafe.com**, 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 through 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 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: Sneha Sarkar

Date: September 18, 2025

1. Executive Summary and Title

Situated at Suri Bus Stand, Birbhum, West Bengal, "**Cafe.com**" is a multi-service retail outlet that serves the day-to-day requirements of students, professionals, and locals by providing a variety of services in one place. The business was set up in 2018 and has a B2C business model. The outlet deals in document printing and Xerox, photo ID card services like Aadhaar, voter ID, and PVC card prints, and retail of stationery, school accessories, and small items like pens, pencils, rulers, PVC card holder. With their centralized location and handy service mix, it has become a one-stop shop for people looking for quick and cost-effective solutions.

In spite of the popularity of the store, it has some internal and external problems that can restrict its growth and operational effectiveness. Internally, the place keeps sales records on Excel. However, the sheets are not properly formatted or standardized, making them difficult to read and analyze. This lack of structured insights prevents the owner from clearly identifying which services or products generate the most revenue. That leads to stock running low at bad times. Upselling chances slip away. Service just is not as smooth as it could be. Externally, customers act all over the place. Demand goes up and down with stuff like exams or holidays. Nearby shops compete more now. They sell the same things. Plus, suppliers charge more. Space is tight. The business has to change. It needs to grow somehow.

Solving these woes through improved inventory planning, sales analysis, and space management will be critical in improving profitability and customer satisfaction in this dynamic market environment.

2. Proof of Originality

2.1. Video Interaction with the Owner (in Bangla)

Video 2.1: Interaction with the Owner via GMeet [Go to the Video link]

2.2. Letterhead from Business

CAFE.COM

Cafe.com
Suri Bus Stand, Suri
Birbhum, West Bengal - 731101

Date: September 17, 2025

The Head of Department
Indian Institute of Technology Madras
Chennai, Tamil Nadu - 600036

Subject: Data Provision Certificate for Academic Project

Dear Sir/Madam,

This is to certify that **Cafe.com**, a service center offering facilities such as Xerox, printing, photo ID card services, stationery, mobile accessories, and small electronic items, has provided relevant business data to **Sneha Sarkar** for use in her academic project titled:

"Optimizing Inventory and Service Offerings for a Multi-Service Retail Shop"

The dataset includes records of service usage, sales transactions, and customer requests collected during the period **October, 2024** to **April, 2025**. This data has been shared strictly for educational and analytical purposes and remains the intellectual property of **Cafe.com**.

We are pleased to support **Sneha Sarkar** in pursuing this academic endeavor and believe the dataset will help achieve valuable insights. Should you require any additional details or confirmation, you may reach out to us directly.

Thank you.

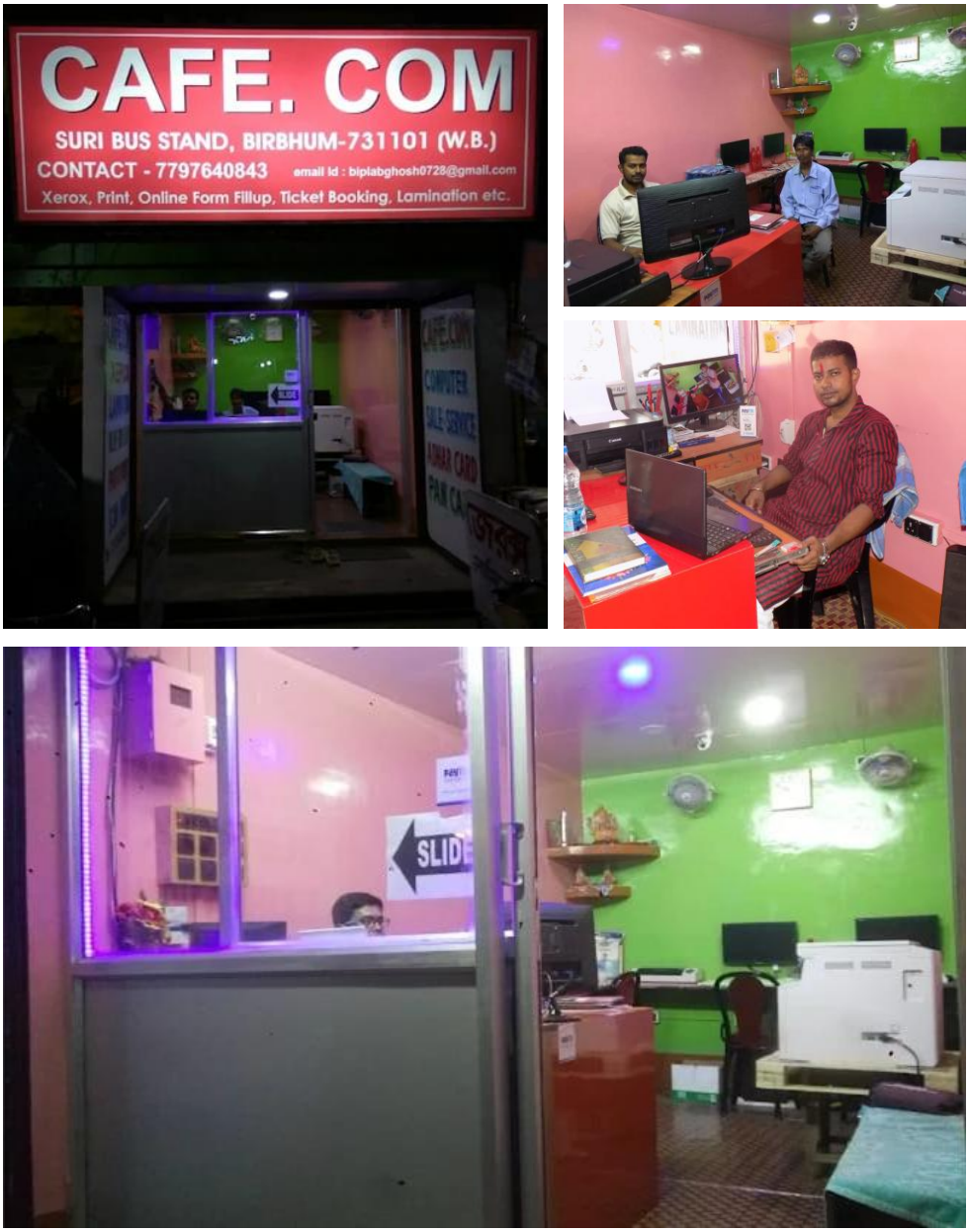
Sincerely,
Biplab Ghosh
Owner
Cafe.com

Biplab

+91 77976 40843
Suri Bus Stand, Suri, Birbhum, West Bengal - 731101
cafe.com.suri@gmail.com

Image 2.1: Proof of Originality from the Shop [Go the Letterhead PDF link]

2.3. Images of the Shop



Images 2.2-2.5: Images of the Shop [Go to the Images folder]

3. Metadata

3.1. Origin of the Dataset

The dataset came from **Cafe.com**'s manually entered datasheet. That is where we pulled it for this project. The business does not keep any structured data around. Still, they track each customer transaction item in a concatenated way. Pretty much everything gets lumped together like that. The cleaning and preprocessing steps for the data show up later in this document under the 'Data Preprocessing' tag.

3.2. Raw Data Format

Column	Data Type	Description	Example	Missing Value (%)
date	String	Date of sell	14 oct 24	0
time	String	Time of sell	6:10 pm	0
customer	String	Customer category	student	14.83%
item	String	Name of SKU	kamlinnotebook+apsara pencel+plastic-cover	0
quantity	String	Count of purchase of SKU	2+4+1	0
price/unit	String	Cost price of each SKU	50+5+5	0
total sell	String	Total Cost price	100+20+5	0
payment	String	Mode of payment	PayTM	8.91%

Table 3.1: Raw data format [Go to the spreadsheet]

3.3. Data Preprocessing

To ensure the dataset was consistent, structured, and ready for analysis, several preprocessing operations were carried out on the raw cyber cafe data. These steps are described below:

- **Data Cleaning**

  Data cleaning started with the raw dataset. It was full of missing values, and invalid entries. It needed careful handling. Some got imputed when it made sense to do so. Other rows with incomplete or unusable info just got dropped. Invalid entries went too, like rows missing a customer type. All that ensured the dataset stayed reliable overall.

- **Standardization of Columns**

  We wanted the data easier to interpret, so column names got put into a consistent format. Data types needed fixing as well. Dates turned into proper datetime objects. Quantities and prices became numeric values. Categorical things like different items got normalized, i.e. stuff like ‘apsara pencil’, ‘youva pencil’, and ‘natraj pencil’ all changed to one standard form ‘Pencil’.

- **Time formatting**

  Values were irregular, inconsistent and were in a 12-hour AM/PM structure. So, for better analysis, they were divided into three bins, ‘Morning’, ‘Afternoon’, & ‘Evening’ in order to study the maximum and minimum customer pressure in different shifts.

- **Categorization of Items**

  Instead of analyzing every single item on its own, we grouped them into broader service categories. Things fell under Academic Supplies, Document Services, Basic Stationery, Office Supplies, Digital Services, Photo Services, and Printing Services. This reduced noise in the dataset, allowing clearer insights into customer preferences at the service.

- **Payment Method Consolidation**

  To fix inconsistencies. Variants like upi, PayTM, GooglePay merged under one category, UPI. The number of unique payment methods narrowed to a few meaningful ones; Cash, UPI, Card. This cleaned up payment trend analysis made it more insightful.

- **Final Structuring**

  After all transformations, the dataset is organised into a logical structure. Column order arranged as Date, Time, Item, Category, Quantity, Unit Price, Total Price, Payment Method. This preprocessed dataset turned out consistent, accurate. Ready for detailed analysis like forecasting, segmentation, market basket modelling.

Column	Data Type	Description	Example	Missing Value (%)
Invoice Number	String	Bill ID	INV 00000005	0
Date	DateTime	Date of transaction	2024-10-13	0
Customer Type	String	Customer category	Student	0
Category	String	Category of SKU	Academic Supplies	0
Item	String	Name of SKU	Notebook	0
Quantity	Integer	Count of purchase of SKU	2	0
Unit Price	Float	Cost price of each SKU	40	0
Line Total	Float	Total Cost price	80	0
Payment Method	String	Mode of payment	Cash	0
Purchase Shift	String	Shift of purchase	Evening	0

Table 3.2: Preprocessed data format [Go to the spreadsheet]

4. Descriptive Statistics

4.1. Overview of the Dataset

Between October 2024 and April 2025, a total of 34 SKUs were sold, generating 1,740 transactions overall. The total revenue during this period amounted to ₹4,42,780.00, with an average transaction value of ₹76.53.

4.2. Central Tendency of Numerical Columns

	Quantity	Unit Price	Total Price
Count	5766.000000	5766.000000	5766.000000
Mean	21.065557	14.806191	76.555498
Min	1.000000	1.500000	1.500000
25%	2.000000	2.000000	30.000000
50%	5.000000	10.000000	60.000000
75%	32.750000	25.000000	110.000000
Max	100.000000	129.000000	440.000000
Std	28.660876	17.648897	64.113783

Table 4.1: Descriptive Statistics of the Numerical Columns

4.3. Sales by Month

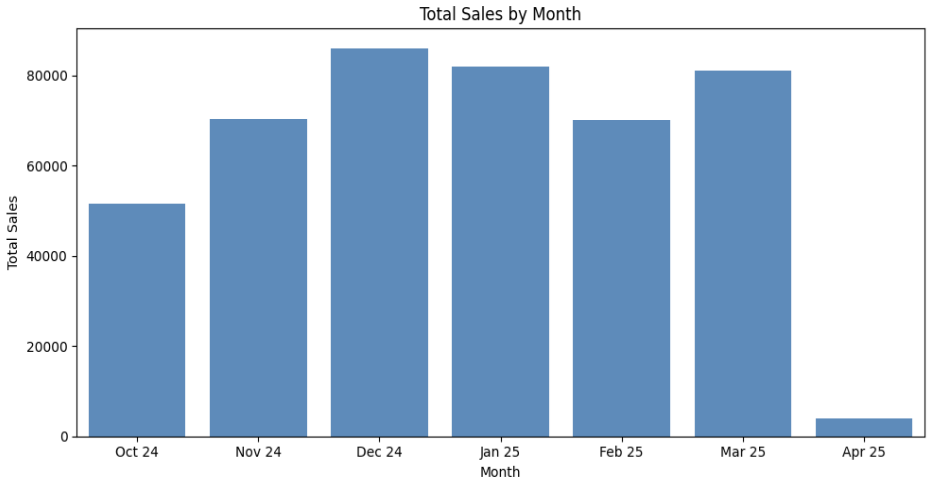


Figure 4.1: Sales by Month Distribution

This chart shows the sum of sales for each month, helping to identify peak revenue periods and seasonal trends. Spikes in sales may indicate successful promotions, holidays, or increased demand.

4.4. Understanding Categories

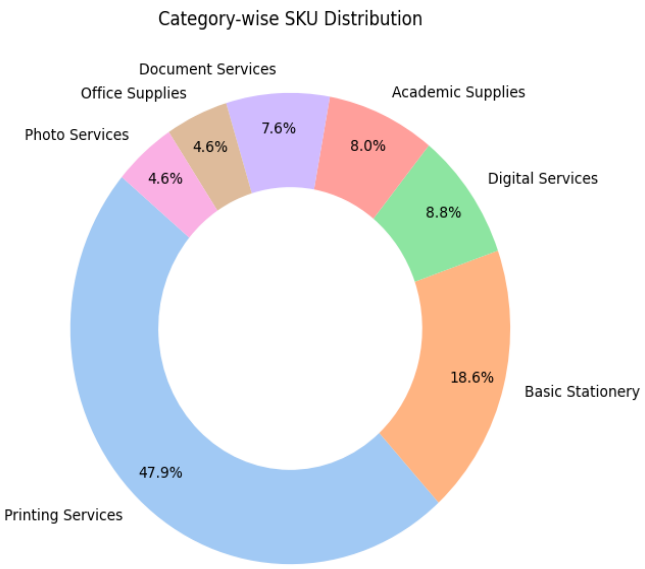


Figure 4.2: Category-wise SKU Distribution

The way items get spread out across different categories really shows what kind of services this café focuses on. Printing and Xerox stuff, along with stationery products, stand out as the big players here. It makes sense that people see the cyber cafe mostly as a spot for quick fixes on school or office work. Then we have those other areas like basic accessories and digital things, photo IDs or browsing sessions. They add some variety, but not a ton. Overall, the place has tried to expand a bit. Still, the main action stays locked in on printing and stationery needs.

4.5. SKU Distribution

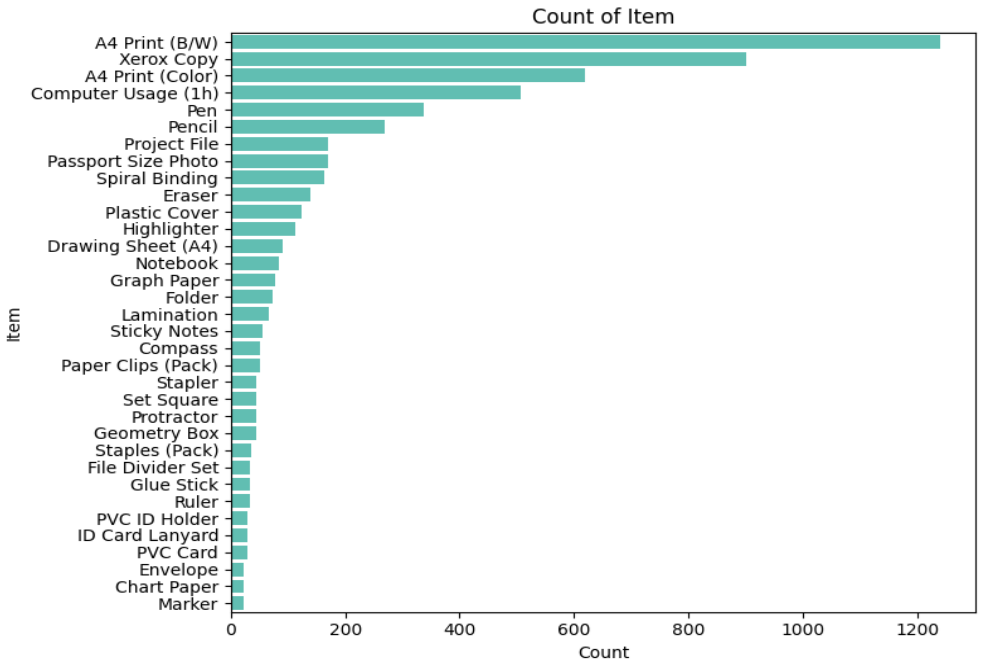


Figure 4.3: SKU Count

Looking at the chart for how many items get bought, we can spot the top sellers right away. Things like photocopies, printouts, and pens, which come under the categories of '**Printing Services**' and '**Basic Stationary**', show up all the time. On the flip side, items like Lamination, Geometry Box, PVC card, which come under the categories of '**Document**

Services' and '**Photo Services**', do not sell as often but bring in more cash per sale. This setup points to that usual trade-off between lots of low-price sales and fewer high-price ones. The cafe pulls in reliable money from the everyday stuff. But those bigger purchases now and then, they can really lift the profits. Maybe pushing some cross-sells, like deals on accessories when someone prints a bunch, could even work.

4.6. Purchase Shift Ratio

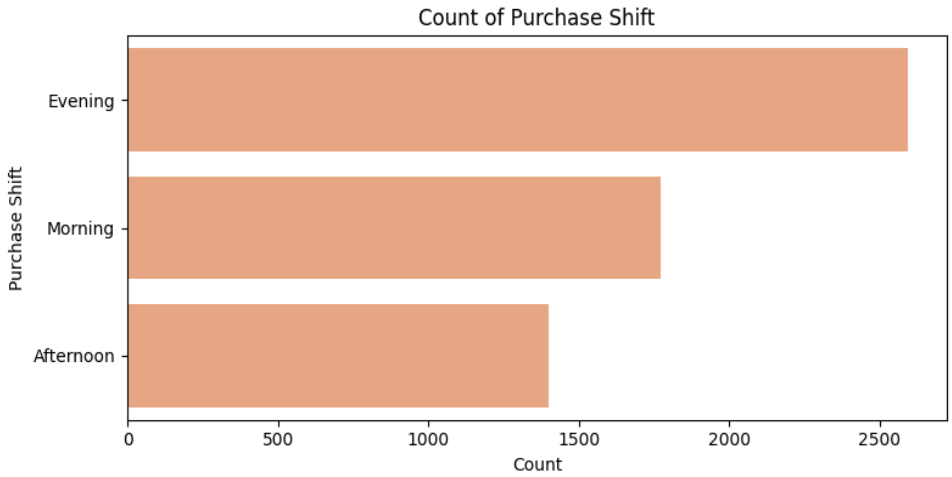


Figure 4.4: Purchase Shift Distribution

Talking about the different convenience slots of the customers, it is observed that the most busy period of the day starts from 5 pm till 10 pm. People are more likely to show up in this slot, followed by the morning period which generally spans from 8 am till 12 pm. It is less likely that there is much sale between 12 pm and 5 pm.

4.7. Customer Reach

Coming to the distribution of different types of customers, the student community has the highest strength, followed by the unknown ones. It is clear from the study that these types of businesses mostly serve the growing youth population, than the regular ones.

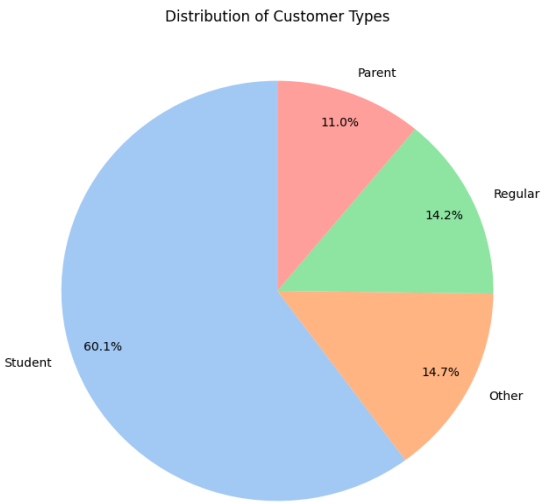


Figure 4.5: Customer Category Distribution

4.8. Studying Payment Method

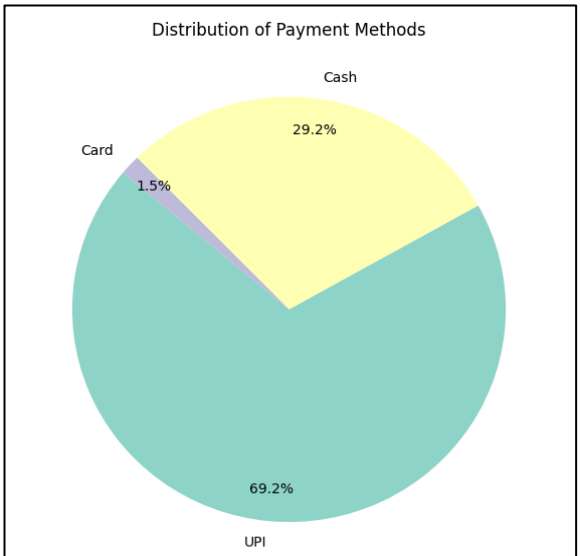


Figure 4.6: Payment Methods Distribution

When it comes to how payments break down, you see what customers like for settling up. Cash handles most of the sales, which fits with those small buys on stationery or copies. That is just how it goes for quick transactions. But UPI and digital options are picking up steam. Especially with younger crowds who want things fast without fumbling for bills. This shift opens doors to nudge more folks toward cashless ways. Small perks like discounts or points for loyalty could do the trick. In the end, it would make bookkeeping easier and cut down on dealing with actual cash.

5. Explanation of Analysis Procedure

The preprocessed dataset provides a clean and structured foundation for analysis. It removes inconsistencies and errors, ensuring accuracy in insights derived from it. Using this refined data, several analytical methods—such as **Market Basket Analysis**, **ARIMA Forecasting**, and **K-Means Segmentation**—are applied to explore customer behavior, service demand, and revenue trends. Together, these techniques reveal patterns that guide strategic decision-making and operational improvements for the business.

5.1. 80/20 Revenue Impact Assessment (Pareto Analysis)

On preprocessed revenue data, it finds top services driving most income. Helps prioritize those high-impact ones for marketing, resource allocation.

5.2. Market Basket Analysis (Cross-Selling Insights)

It uses the preprocessed data to spot services bought together often. Association rule mining does the trick there. Metrics such as **Support**, **Confidence**, **Lift** point out chances for cross-selling. That guides bundled promotions, service placement ideas.

5.3. Cyclic Demand Decomposition (Seasonality Analysis)

Preprocessed time-series data breaks demand into **Trend**, **Seasonal**, **Residual** parts. Reveals recurring cycles, improves forecasting accuracy.

5.4. K-Means Customer Segmentation (Behavioral Segmentation)

Clustering groups customers from the preprocessed set by usage frequency, spending habits. It uncovers different segments for targeted marketing and retention efforts.

5.5. Service Utilization Dynamics (Trend Mining)

Analysis on the preprocessed data reveals shifts in service usage over time. It spots peak demand periods, changing customer preferences. That aids operational planning.

5.6. ARIMA Time Series Forecasting (Demand Forecasting)

The model takes preprocessed time-based data to predict future service needs. It picks up trends, seasonality. All that supports better inventory, resource planning.

6. Results and Findings

6.1. Revenue Concentration & Pareto Analysis

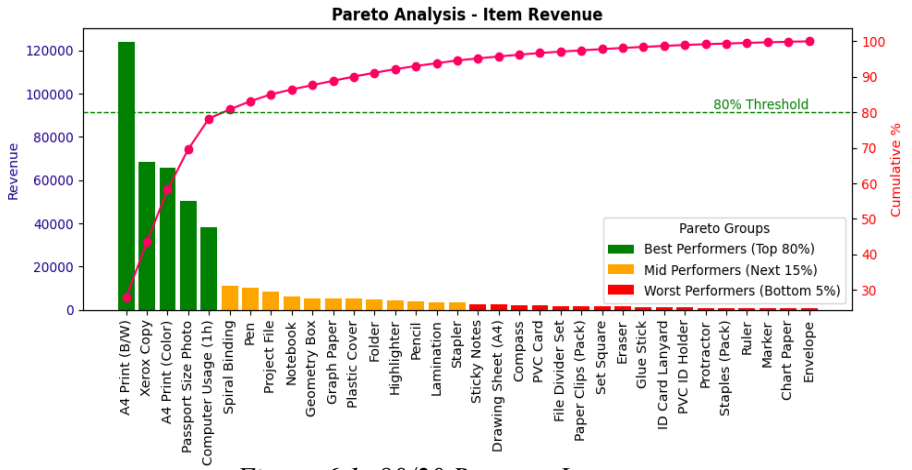


Figure 6.1: 80/20 Revenue Impact

- The Pareto Analysis, reveals items classified into 3 categories based on their influence on the sales. The elbow curve marked by the **pink like**, shows a sharp elbow at the 80% mark showing that the cumulative sales total of the other 20% aren't that significant.

- **Best Performers**: 5 SKUs, namely, A4 Print (B/W & Colors), Xerox Copy, Passport Size photo clipping, and Computer Usage, are responsible for 80% of the total revenue generated. These 5 fall in the services category, signifying that Services offered by the shop are much more important than the products it sells.

- **Mid Performers**: 12 SKUs, some of them are, Spiral Binding, Pen, Files, etc. These SKUs contribute towards the next 15% of the total sales. These are almost all products, reinforcing the idea that products are not much of a sales generator.

- **Worst Performers**: All the rest SKUs, being all products and further solidifying the notion of products being low performers.

6.2. Market Basket Analysis

- The network shows frequently co-purchased items from Market Basket Analysis.
- A4 Print (B/W) appears as the central node, **indicating it is the most frequently purchased item** and often purchased with several other products such as Folder, Stapler, Protractor, and Graph Paper. This can indicate that **people tend to be impulse-buy while getting prints**.

Figure 6.2: Cross-Selling Insights

- Surprisingly, this is not true for A4 Print (Colour), which indicates that people **tend to buy B/W prints above Coloured prints**.
- The diagram highlights **cross-selling opportunities** and suggests bundling related services to boost sales.

6.3. Cyclic Demand Decomposition

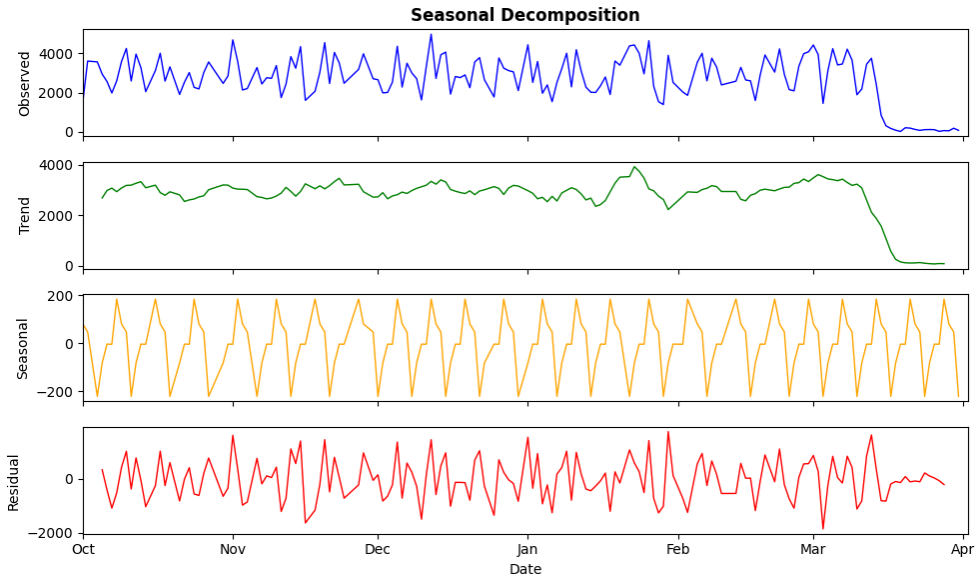


Figure 6.3: Seasonality Analysis

- **Observed:** Shows actual demand with periodic fluctuations and a sharp drop around March–April.
- **Trend:** Remains steady from October to February, then declines sharply toward April.
- **Seasonal:** Displays consistent cyclical patterns, indicating strong recurring demand.
- **Residual:** Captures random variations, relatively stable but reduces near the end.
- **Overall:** Demand shows regular seasonality with a stable pattern that drops significantly in the final months.

6.4. Customer Segmentation with K-Means Clustering

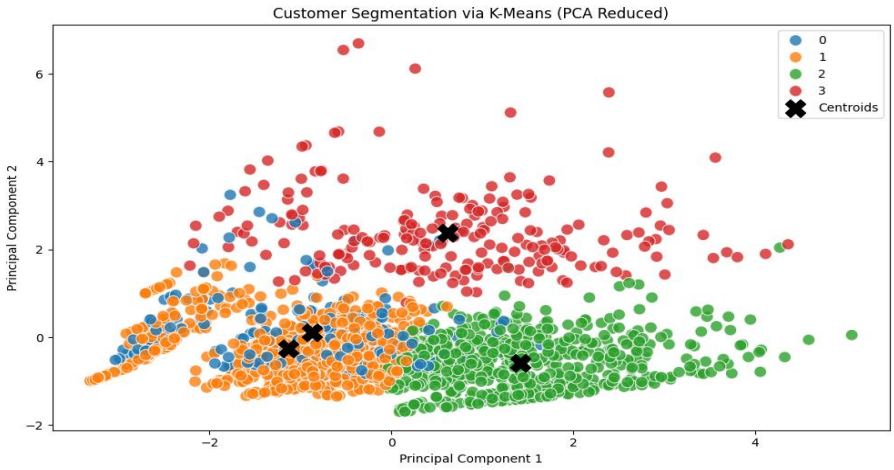


Figure 6.4: K-Means Customer Segmentation

- K-means clustering was applied on the preprocessed data, and for the sake of visualisation, PCA was applied on it.
- It is visible that customers can be **classified into four major categories**. These categories represent **the amount of purchase** they have done, **the payment** they have used, **the categories of the items** they have purchased.

6.5. Service Popularity & Trend Mining

- The graph shows the revenue trends of the top three service categories — **Printing Services**, **Digital Services**, and **Photo Services** — over time.
- **Printing Services** consistently generate the highest revenue, indicating strong and steady customer demand.
- **Photo Services** show moderate and frequent fluctuations, suggesting irregular but recurring usage patterns.
- **Digital Services** maintain low but stable revenue, implying limited yet consistent demand.

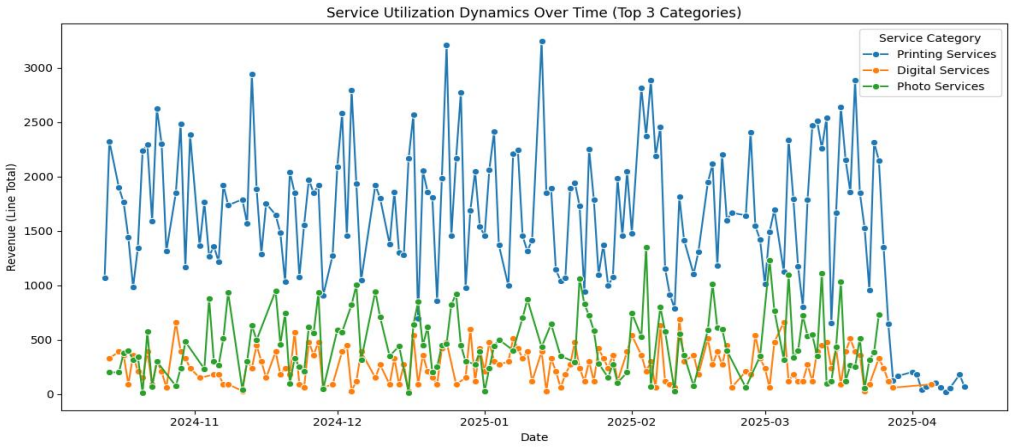


Figure 6.5: Service Utilization Dynamics

- A noticeable decline across all categories is observed toward **March–April 2025**, likely due to reduced activity or external factors impacting service utilization.
- Overall, **Printing Services** dominate revenue generation, followed by **Photo and Digital Services**, highlighting key focus areas for business optimization.

6.6. ARIMA Time Series Forecasting

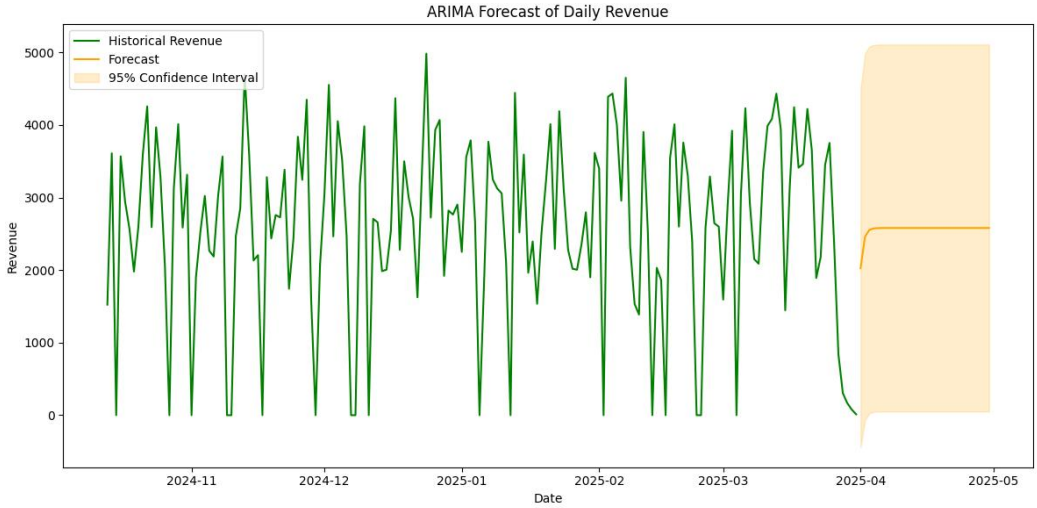


Figure 6.6: Demand Forecasting

- The **past trend** shows high volatility — no consistent increase or decrease over time, just random ups and downs.
- The **ARIMA forecast** predicts that revenue will likely **stabilize around the recent average**, rather than growing or dropping significantly.
- This suggests the store's **overall revenue pattern is steady but irregular**, with daily variations likely due to customer flow or category mix rather than a systematic trend.
- The **wide forecast interval** indicates **uncertainty** — the model can't confidently predict exact values because of the noisy historical pattern.