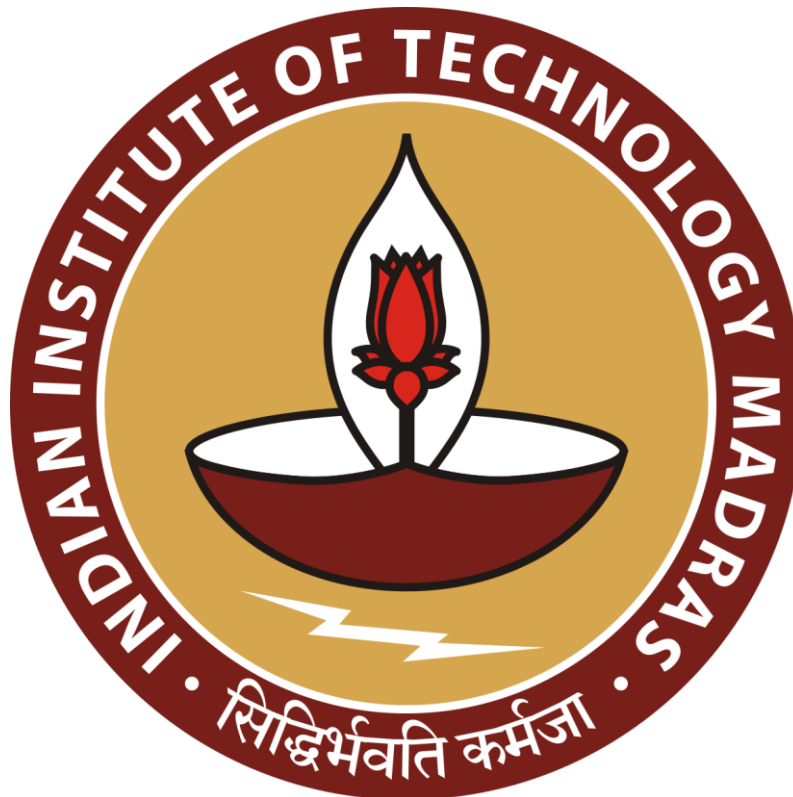**Data-Driven Demand Forecasting and Stock Optimization
for an Army CSD**

A Mid-Term report for BDM capstone project

Submitted by

Name: Akshata Kumar
Roll Number: 22f3001688

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

Contents

Executive Summary:-

This report presents a data-driven approach to demand forecasting and stock optimization for the **Army Canteen Stores Department (CSD)**. The objective is to enhance inventory management by leveraging statistical models, machine learning techniques, and optimization strategies to predict demand fluctuations and ensure optimal stock levels.

The Army Canteen Stores Department (CSD), particularly the **MP Sub Area Unit Run Canteen in Bhopal Cantt**, faces challenges in maintaining optimal inventory levels due to **fluctuating demand, diverse product categories, and variable supplier lead times**. Frequent stockouts of essential items and excessive overstocking of slow-moving goods have resulted in inefficiencies, affecting both cost and service quality. This project addresses these challenges through a data-driven approach to demand forecasting and inventory optimization.

The dataset used includes **3000** records with key metadata fields such as **Item Code, Category, Stock Quantity, Reorder Level, Lead Time, Weekly Demand, EOQ, and Sales Revenue.** It was cleaned and standardized primarily using Python and its powerful libraries, ensuring consistency in categorical labels, date formats, and missing value handling. Descriptive statistics revealed **right-skewed EOQ** and revenue distributions, high variability across categories, and a **greater number of overstock events than stockouts**— indicating an opportunity for smarter procurement strategies.

The analysis employs forecasting models like **ARIMA, SARIMA, Holt-Winters, and Prophet** to project item-level demand, alongside EOQ-based optimization for inventory control. Mid-term findings include the identification of high-impact SKUs, reorder risks, and category-wise stock inefficiencies. Key visualizations such as EOQ **distribution plots, revenue histograms, and reorder alert dashboards** provide actionable insights. These results lay a strong foundation for implementing strategic stocking policies aimed at improving operational efficiency and reducing excess costs within the CSD supply chain.

By integrating advanced analytics and forecasting techniques, this study aims to streamline supply chain operations, minimize stock wastage, and enhance overall efficiency in Army CSD inventory management.

Proof of Originality of Data:-
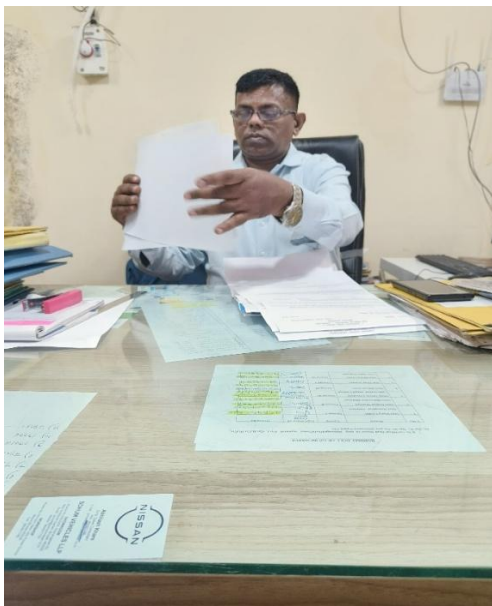
1. **Letter Head:-**
   https://drive.google.com/file/d/1rAe5Rod2Vmot6Q2_poEhQ_L9rYmf8nEb/view?usp=sharing

2. **Video Proof:-**
   https://drive.google.com/drive/folders/1FXV6-GPPv0CMxLucGlGWdDeG6Q_2-Zhn?usp=sharing

3. **Images at the CSD:-**

Metadata:-

The metadata given provides an overview of the structure and attributes of the dataset used for **CSD Stock Optimization & Inventory Management**. It defines key details such as column names, data types, descriptions, and data sources. This ensures clarity, consistency, and ease of analysis.

- Data Source:

Collected from **MP Sub Area Unit Run Canteen, Bhopal Cantt, Madhya Pradesh.** Extracted from **transaction records, stock registers, supplier invoices, and historical procurement data.**

- Data Description:

The dataset used in this project has been collected through raw, real-world inventory records from the MP Sub Area Unit Run Canteen (URC), Bhopal Cantt. It consists of **3000 entries** across **18 columns**, representing item-level stock, sales, procurement, and supplier data required for demand forecasting and inventory optimization.

Column Name	Description	Data Type
Item_Code	Unique alphanumeric ID assigned to each item	String
Item_Name	Name of the item (includes inconsistent casing and typos)	String
Category	Category the item belongs to (e.g., Grocery, Beverages)	String
Stock_Quantity	Current stock level in units	Integer
Reorder_Level	Threshold at which reorder is initiated	Integer
Unit_Price	Retail price per unit (in ₹)	Float
Purchase_Cost	Purchase cost per unit (in ₹)	Float
Last_Purchase_Date	Date of last purchase (format: dd-mm-yyyy)	String / Date
Weekly_Demand	Average weekly demand in units	Integer
Monthly_Sales_Units	Total units sold per month	Integer
Sales_Revenue	Revenue generated per item per month (₹)	Float
Lead_Time_Days	Average days taken to receive stock from supplier	Integer / Null
Ordering_Cost	Cost incurred per procurement order (₹)	Integer
Holding_Cost_Per_Unit	Estimated monthly cost of holding one unit in stock (₹)	Float
EOQ	Computed Economic Order Quantity per item	Float
Stockout_Events	Number of times item went out of stock in last 6 months	Integer
Overstock_Events	Number of times stock exceeded optimal level	Integer
Supplier_Name	Name of the vendor or brand supplying the item	String

1.1 Image samples of raw data :-

	Item_Code	Item_Name	Category	Stock_Quantity	Reorder_Level	Unit_Price	Purchase_Cost	Last_Purchase_	Weekly_Demand	Monthly_Sales_	Sales_Revenue	Lead_Time_Day	Ordering_Cost	Holding_Cost_P	EOQ	Stockout_Events	Overstock_Even	Supplier_Name
2	CSD0001	Eraser	Stationery	220	45	103.42	69.35	30-04-2025	10	50	5171.0	15	150	5.17	53.86	0	1	Amul
3	CSD0002	Phenyl	household	104	80	442.23	272.13	27-01-2025	35	105	46434.15	15	100	22.11	30.82	0	0	Patanjali
4	CSD0003	Marker	Stationery	86	45	490.34	343.37	29-01-2025	29	87	42659.58	7	100	24.52	26.64	0	0	ITC
5	CSD0004	Atta	GROCERY	83	36	344.22	211.29	27-02-2025	1	3	1032.66	5	100	17.21	5.9	0	1	ITC
6	CSD0005	SanitaryPads	Personal Care	90	25	294.26	188.35	19-03-2025	16	80	23540.8	3	150	14.71	40.39	0	0	Local Vendor
7	CSD0006	Detergent	Household	151	57	427.67	346.66	15-03-2025	6	24	10264.08	8	150	21.38	18.35	0	0	ITC
8	CSD0007	SPICES	GROCERY	111	26	468.53	323.07	18-01-2025	26	78	36545.34	12	100	23.43	25.8	0	0	CocaCola
9	CSD0008	Atta	GROCERY	0	95	259.82	214.51	21-02-2025	19	95	24682.9	7	100	12.99	38.24	1	0	Amul
10	CSD0009	Sanitary Pads	Personal Care	153	62	351.85	228.53	24-02-2025	34	102	35888.7	13	100	17.59	34.06	0	0	Nestle
11	CSD0010	Face Wash	Personal Care	109	77	256.5	208.97	29-04-2025	7	21	5386.5	3	150	12.83	22.16	0	0	ITC
12	CSD0011	Sugar	Grocery	197	25	265.32	202.69	22-12-2024	9	45	11939.4	3	100	13.27	26.04	0	0	Amul
13	CSD0012	Phenyl	Household	80	22	433.19	280.51	21-11-2024	17	85	36821.15	5	150	21.66	34.31	0	1	HUL
14	CSD0013	Tea	Beverages	149	28	182.75	137.38	19-04-2025	31	155	28326.25	7	200	9.14	82.36	0	1	Local Vendor
15	CSD0014	Sanitary Pads	Personal Care	118	43	445.83	288.87	23-01-2025	9	27	12037.41	12	200	22.29	22.01	0	1	Nestle
16	CSD0015	Eraser	Stationery	93	47	150.55	95.37	05-01-2025	3	12	1806.6	13	100	7.53	17.85	0	0	CocaCola
17	CSD0016	Sugar	Grocery	131	27	462.05	301.98	24-02-2025	15	60	27723.0	6	200	23.1	32.23	0	1	ITC
18	CSD0017	Pencil	Stationery	96	81	432.56	262.39	14-01-2025	10	50	21628.0	3	100	21.63	21.5	0	0	Godrej
19	CSD0018	Toothpaste	Personal Care	166	84	475.92	349.0	29-04-2025	13	52	24747.84	4	150	23.8	25.6	0	0	HUL
20	CSD0019	Broom	Household	257	80	151.65	93.45	04-02-2025	1	5	758.25	9	100	7.58	11.49	0	1	ITC
21	CSD0020	Notebook	Stationery	69	35	82.87	55.42	30-12-2024	8	32	2651.84	15	200	4.14	55.6	0	0	Patanjali
22	CSD0021	Shampoo	Personal Care	69	36	442.5	271.62	14-01-2025	15	75	33187.5	15	100	22.12	26.04	0	0	ITC
23	CSD0022	Floor Cleaner	household	159	84	311.04	201.44	22-01-2025	24	96	29859.84	8	100	15.55	35.14	0	0	Nestle
24	CSD0023	pencil	STATIONERY	81	63	76.94	65.23	08-04-2025	25	125	9617.5	9	200	3.85	113.96	0	0	HUL
25	CSD0024	Shampoo	personal care	123	97	145.03	92.51	28-01-2025	11	44	6381.32	7	100	7.25	34.84	0	0	Patanjali
26	CSD0025	Pen	Stationery	141	50	301.09	204.38	26-11-2024	16	48	14452.32	14	200	15.05	35.72	0	0	Godrej
27	CSD0026	SanitaryPads	Personal Care	224	29	30.63	25.96	23-04-2025	1	3	91.89	10	200	1.53	28.01	0	0	Amul
28	CSD0027	Pen	Stationery	237	35	278.37	169.74	10-05-2025	14	56	15588.72	10	200	13.92	40.11	0	1	HUL
29	CSD0028	Pencil	Stationery	29	91	157.55	127.64	27-01-2025	21	84	13234.2	11	150	7.88	56.55	2	0	Patanjali
30	CSD0029	Spices	Grocery	3	40	375.3	306.76	18-03-2025	18	90	33777.0	5	200	18.77	43.79	2	0	Amul
31	CSD0030	Atta	Grocery	146	31	332.28	231.98	23-11-2024	11	44	14620.32	7	150	16.61	28.19	0	1	Local Vendor

1.2 Google drive link for raw data :-

https://drive.google.com/file/d/1Wr4tb9MFgg9Th7apyWmsNJ5s-eSdNPyj/view?usp=sharingi

2. **Cleaned Data :-** To transform a raw, semi-structured inventory dataset from the MP Sub Area CSD into a clean and consistent format suitable for forecasting, optimization, and dashboarding tasks, I followed these steps:

Data Issues:
- Mixed casing in categorical fields
- Inconsistent date formats
- Missing values in Lead_Time_Days
- Typographical anomalies in item/category names

Cleaning & Preprocessing Steps

1. **File Upload & Reading**
- Loaded the raw CSV into a Pandas DataFrame using pd.read_csv().
- Previewed using .head() and .info() to inspect structure.

2. **Standardized Categorical Text Columns**

Issue: Categories and supplier names had inconsistent casing, spacing, and typos.

Fix Applied:
- Unified category labels (e.g., "GROCERY", "grocery", "Grocery" → "Grocery").
- Stripped extra spaces and standardized all to title case.

3. Handled Inconsistent Date Formats

Issue: Last_Purchase_Date column contained both DD-MM-YYYY and YYYY/MM/DD.

Fix Applied:
- Converted all entries into datetime64 type.
- Invalid entries became NaT (automatically handled next).

4. Imputed Missing Values in Lead Time

Issue: 57 entries had missing or blank Lead_Time_Days.
Fix Applied:
- Converted all entries to numeric.
- Imputed missing values using the median for robust estimation.

5. Fixed Missing Dates (if any)

Issue: Some Last_Purchase_Date values were NaT.

Fix Applied:
- Filled missing purchase dates with the dataset's **median date**.

6. Removed Duplicates

Ensured no duplicate items existed based on unique Item_Code.

7. Final Data Check

Verified with .isnull().sum() and .info() that:
- No critical nulls remain
- All datatypes are correct
- Consistency across columns is achieved

2.1 Google drive link for cleaned data :-
https://drive.google.com/file/d/1eYIkk0hYYwkSw_uKrtEuLe9SNmDOLfvl/view?usp=sharing

2.2 Image samples for cleaned data :-

Item_Code	Item_Name	Category	Stock_Quan	Reorder_L	Unit_Price	Purchase_	Last_Purcha	Weekly_D	Monthly_	Sales_Rev	Lead_Time	Ordering_	Holding_C	EOQ	Stockout_	Overstock	Supplier_Name
CSD0001	Eraser	Stationery	220	45	103.42	69.35	30-04-2025	10	50	5171	15	150	5.17	53.86	0	1	Amul
CSD0002	Phenyl	Household	104	80	442.23	272.13	27-01-2025	35	105	46434.15	15	100	22.11	30.82	0	0	Patanjali
CSD0003	Marker	Stationery	86	45	490.34	343.37	29-01-2025	29	87	42659.58	7	100	24.52	26.64	0	0	Itc
CSD0004	Atta	Grocery	83	36	344.22	211.29	27-02-2025	1	3	1032.66	5	100	17.21	5.9	0	1	Itc
CSD0005	Sanitarypads	Personal C	90	25	294.26	188.35	19-03-2025	16	80	23540.8	3	150	14.71	40.39	0	0	Local Vendor
CSD0006	Detergent	Household	151	57	427.67	346.66	15-03-2025	6	24	10264.08	8	150	21.38	18.35	0	0	Itc
CSD0007	Spices	Grocery	111	26	468.53	323.07	18-01-2025	26	78	36545.34	12	100	23.43	25.8	0	0	Cocacola
CSD0008	Atta	Grocery	0	95	259.82	214.51	21-02-2025	19	95	24682.9	7	100	12.99	38.24	1	0	Amul
CSD0009	Sanitary Pads	Personal C	153	62	351.85	228.53	24-02-2025	34	102	35888.7	13	100	17.59	34.06	0	0	Nestle
CSD0010	Face Wash	Personal C	109	77	256.5	208.97	29-04-2025	7	21	5386.5	3	150	12.83	22.16	0	0	Itc
CSD0011	Sugar	Grocery	197	25	265.32	202.69	22-12-2024	9	45	11939.4	3	100	13.27	26.04	0	0	Amul
CSD0012	Phenyl	Household	80	22	433.19	280.51	21-11-2024	17	85	36821.15	5	150	21.66	34.31	0	1	Hul
CSD0013	Tea	Beverages	149	28	182.75	137.38	19-04-2025	31	155	28326.25	7	200	9.14	82.36	0	1	Local Vendor
CSD0014	Sanitary Pads	Personal C	118	43	445.83	288.87	23-01-2025	9	27	12037.41	12	200	22.29	22.01	0	1	Nestle
CSD0015	Eraser	Stationery	93	47	150.55	95.37	05-01-2025	3	12	1806.6	13	100	7.53	17.85	0	0	Cocacola
CSD0016	Sugar	Grocery	131	27	462.05	301.98	24-02-2025	15	60	27723	6	200	23.1	32.23	0	1	Itc
CSD0017	Pencil	Stationery	96	81	432.56	262.39	14-01-2025	10	50	21628	3	100	21.63	21.5	0	0	Godrej
CSD0018	Toothpaste	Personal C	166	84	475.92	349	29-04-2025	13	52	24747.84	4	150	23.8	25.6	0	0	Hul
CSD0019	Broom	Household	257	80	151.65	93.45	04-02-2025	1	5	758.25	9	100	7.58	11.49	0	1	Itc

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
2983	CSD2982	Notebook	Stationery	121	29	170.73	127.05	17-02-2025	20	80	13658.4	5	200	8.54	61.21	0	0	Cocacola	
2984	CSD2983	Tea	Beverages	83	38	438.54	346.33	27-12-2024	34	102	44731.08	10	100	21.93	30.5	0	1	Hul	
2985	CSD2984	Mop	Household	111	42	276.93	202.13	02-12-2024	27	108	29908.44	7	200	13.85	55.85	0	0	Itc	
2986	CSD2985	Sanitary Pads	Personal C	79	51	152.84	98.69	29-04-2025	1	4	611.36	14	100	7.64	10.23	0	0	Hul	
2987	CSD2986	Eraser	Stationery	79	31	281.06	198	11-12-2024	11	33	9274.98	12	200	14.05	30.65	0	1	Patanjali	
2988	CSD2987	Coffee	Beverages	204	74	152.89	123.86	01-04-2025	34	136	20793.04	8	200	7.64	84.38	0	1	Patanjali	
2989	CSD2988	Juice	Beverages	167	44	49.15	41.72	06-03-2025	50	150	7372.5	12	200	2.46	156.17	0	1	Godrej	
2990	CSD2989	Notebook	Stationery	2	77	447.41	363.1	08-05-2025	16	48	21475.68	10	150	22.37	25.37	1	0	Itc	
2991	CSD2990	Soft Drink	Beverages	149	40	436.61	311.47	11-11-2024	13	52	22703.72	12	200	21.83	30.87	0	0	Patanjali	
2992	CSD2991	Notebook	Stationery	184	65	375.57	250.5	19-01-2025	25	75	28167.75	14	100	18.78	28.26	0	0	Hul	
2993	CSD2992	Shampoo	Personal C	138	84	180.42	130.11	21-01-2025	22	66	11907.72	12	200	9.02	54.1	0	0	Local Vendor	
2994	CSD2993	Toothpaste	Personal C	117	66	404.29	284.73	27-11-2024	6	30	12128.7	15	200	20.21	24.37	0	0	Amul	
2995	CSD2994	Rice	Grocery	90	42	81.37	53.89	01-01-2025	10	30	2441.1	8	200	4.07	54.3	0	0	Hul	
2996	CSD2995	Mop	Household	222	36	133.65	107.42	03-12-2024	1	4	534.6	3	100	6.68	10.94	0	1	Godrej	
2997	CSD2996	Notebook	Household	0	60	23.42	14.9	13-03-2025	7	21	491.82	9	200	1.17	84.73	2	0	Patanjali	
2998	CSD2997	Energy Drink	Beverages	11	84	94.96	68.51	12-01-2025	1	4	379.84	11	150	4.75	15.89	1	0	Patanjali	
2999	CSD2998	Tea	Beverages	197	76	414.73	290.7	23-02-2025	1	5	2073.65	3	150	20.74	8.5	0	0	Patanjali	
3000	CSD2999	Juice	Beverages	132	44	273.92	222.9	28-02-2025	11	33	9039.36	4	150	13.7	26.88	0	1	Local Vendor	
3001	CSD3000	Eraser	Stationery	83	73	444.87	374.44	18-12-2024	10	30	13346.1	9	200	22.24	23.23	0	0	Local Vendor	

Descriptive Statistics :-

1. **EOQ Distribution Analysis -** The EOQ distribution is right-skewed, with most items clustered between 20 - 40 units, indicating optimal small-batch ordering for the majority of inventory. A long tail of higher EOQs suggests a few high-demand or high-cost items requiring bulk procurement strategies.
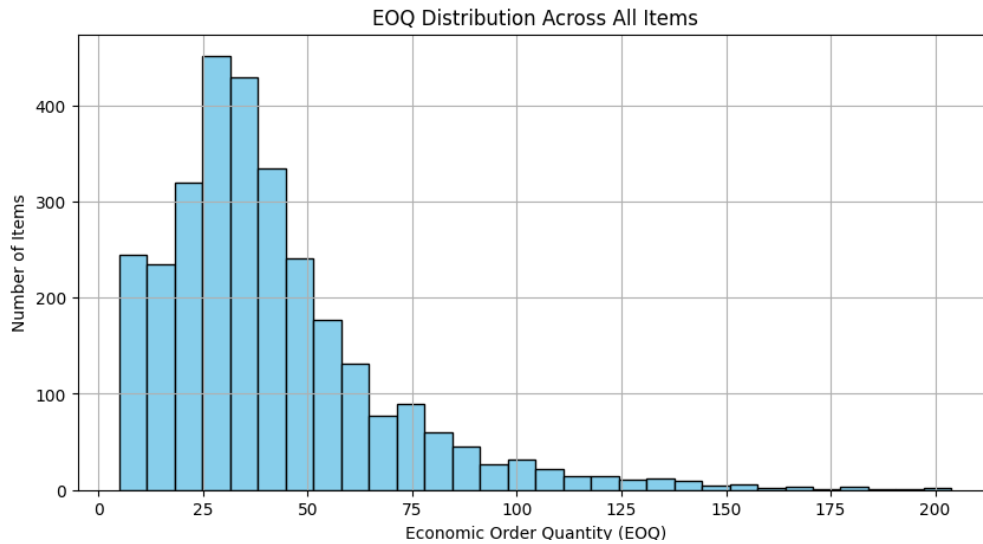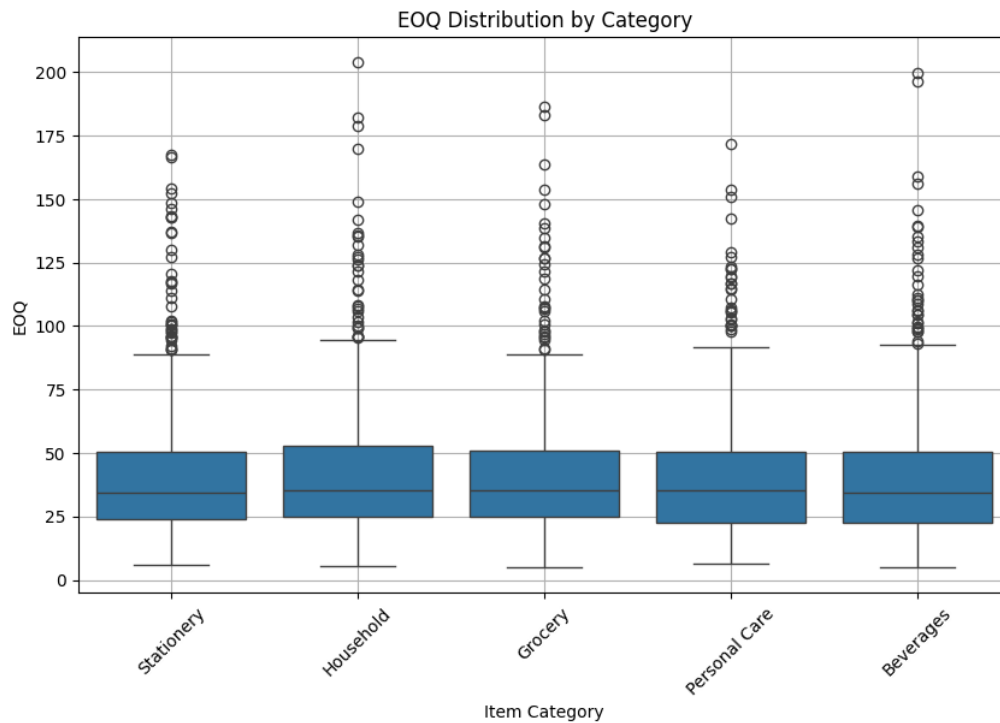

Fig 1.1

Fig 1.2

EOQ values across all categories show similar medians (~30–40 units), but each has a wide range and many high-value outliers. This indicates that while most items follow consistent ordering norms, each category contains some bulk-demand items requiring tailored inventory strategies.

2. **Stock Quantity Analysis -** The scatter plot shows a wide spread of stock quantities across all reorder levels, with no strong linear correlation. However, most items maintain stock well above their reorder threshold, indicating generally safe inventory buffers across categories, though some low-stock outliers suggest possible restocking delays.
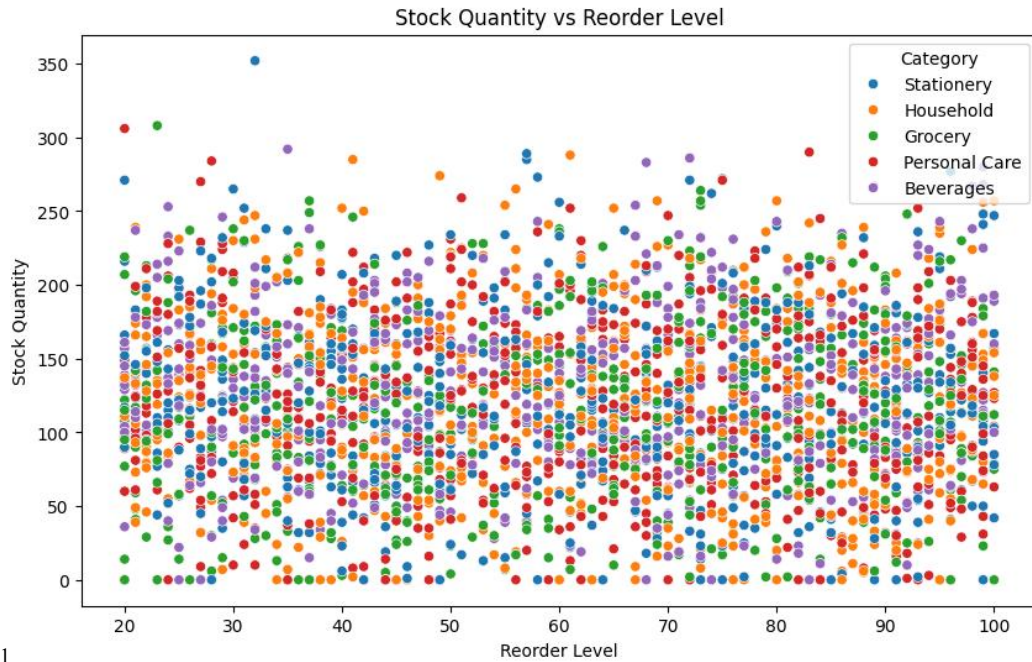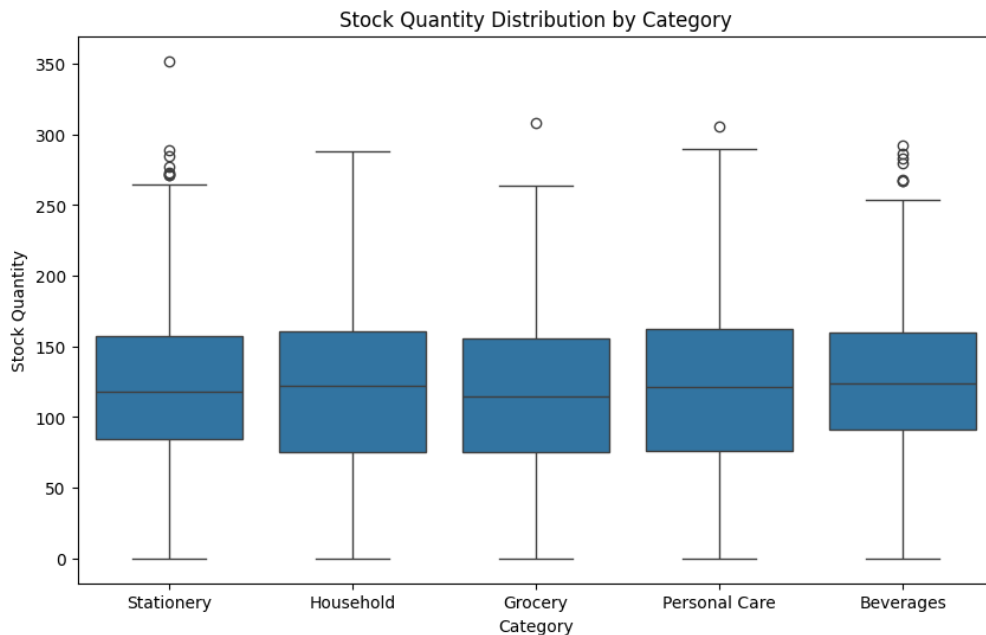

Fig 2.1

Fig 2.2

Stock quantities across categories are fairly consistent, with median levels around 120–130 units. However, all categories exhibit high variability and multiple outliers, indicating the presence of both low-stock and overstocked items that may need closer inventory control.
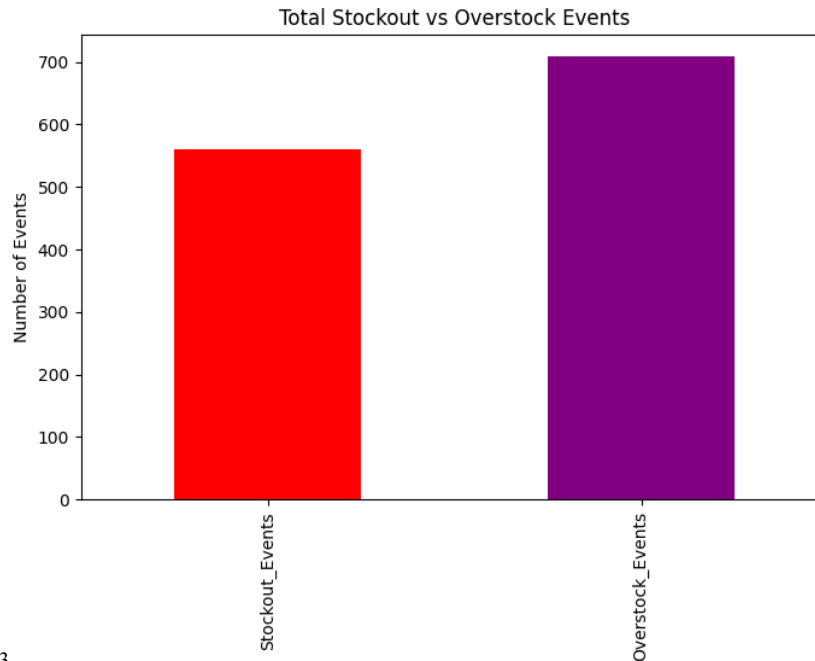
Fig 2.3

The chart shows that overstock events (≈700) outnumber stockout events (≈560), suggesting that the current inventory policy leans towards excess safety stock. While this minimizes unavailability, it may be inflating holding costs and tying up working capital.

3. **Sales Revenue Analysis** - The sales revenue distribution is highly right-skewed, indicating that most inventory items generate low to moderate monthly revenue (under ₹10,000), while a small number of high-performing items contribute disproportionately to total sales. This suggests a classic **Pareto pattern**, where a few key items may drive a majority of revenue and should be prioritized in stock and reorder strategies.
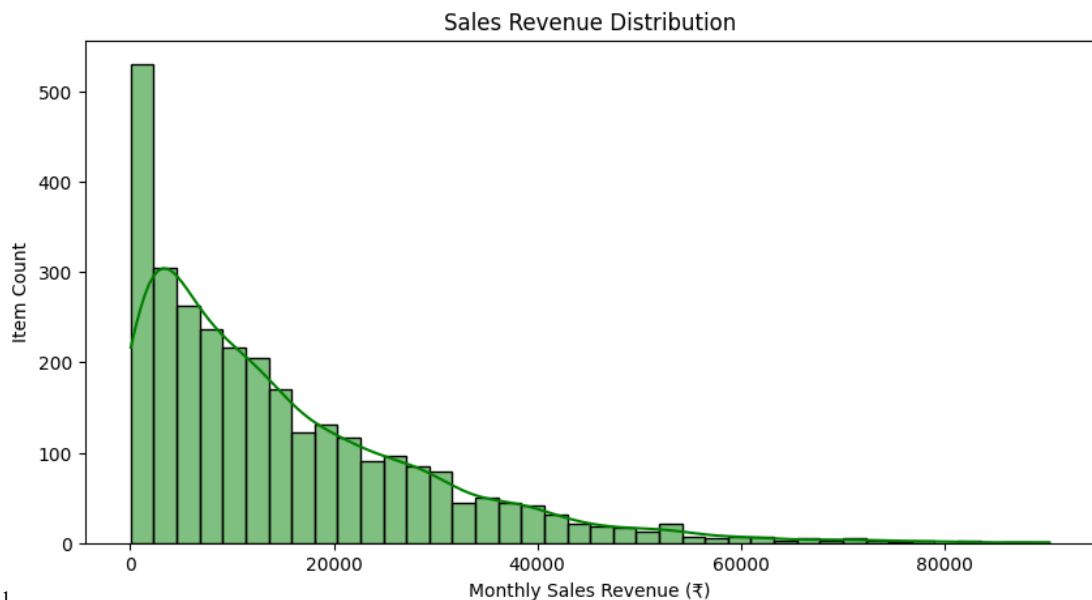

Fig 3.1

Detailed Explanation of Analysis method / process used :-

1. EOQ Analysis: Optimizing Inventory Reordering
Tools: Python (Pandas, Matplotlib, Seaborn)
Process:
- Loaded the cleaned inventory dataset into a Pandas DataFrame.
- Calculated the Economic Order Quantity (EOQ) for each item using the EOQ formula:

$$EOQ = \sqrt{\frac{2DS}{H}}$$

  where D = monthly demand, S = ordering cost, and H = holding cost per unit.
- Used Matplotlib to plot a histogram to visualize the EOQ distribution and identify central tendencies.
- Further broke down EOQ across item categories using a box plot to compare ordering strategies across departments.

2. Stock Quantity vs Reorder Level: Evaluating Inventory Risk Zones
Tools: Pandas, Seaborn
Process:
- Plotted a scatterplot of Stock_Quantity vs Reorder_Level to visually assess how close items are to their reorder points.
- Used hue='Category' to distinguish between item types.
- This helped identify items either understocked (near or below reorder level) or excessively overstocked relative to typical reorder thresholds.

3. Stockout and Overstock Events: Operational Risk Identification
Tools: Pandas, Matplotlib
Process:
- Aggregated the total number of Stockout_Events and Overstock_Events across all items using .sum() for each column.
- Plotted a bar chart to compare the two event types side by side.
- This allowed quantification of how frequently the inventory strategy failed to align supply with demand, either due to under-ordering or over-buffering.

4. Sales Revenue Distribution: Identifying High-Impact Items
Tools: Pandas, Seaborn
Process:
- Plotted a histogram of Sales_Revenue using Seaborn's histplot with a KDE overlay.
- Examined the distribution to determine whether sales were evenly spread across items or concentrated among a few.
- Identified revenue-generating outliers as potential candidates for strategic stocking or promotions.

5. Stock Distribution by Category: Identifying Inventory Holding Patterns
Tools: Seaborn (Boxplot)
Process:
- Used a boxplot to visualize the spread and central tendency of stock levels (Stock_Quantity) across categories like Grocery, Beverages, etc.
- This revealed which categories consistently held more stock and which were leaner or more at risk of stockouts.

Detailed Insights Concluded from the Descriptive Analysis :-

1. **EOQ Distribution**
- Most items have EOQ values clustered between 20–40 units, suggesting optimal small-batch ordering.
- However, a long tail of high EOQ values indicates some bulk-demand items that need separate ordering strategies.

2. **Stock vs Reorder Level**
- While most items are stocked above their reorder level, the wide spread reveals inconsistencies in safety stock management.
- A few items dangerously hover near or below reorder thresholds.

3. **Stockout vs Overstock Events**
- Overstock events (~710) are more frequent than stockouts (~560), highlighting a tendency to overstock.
- While this reduces service disruption, it also increases holding cost and shelf-space inefficiency.

4. **Sales Revenue Distribution**
- Highly skewed: a small number of items generate the majority of monthly revenue.
- Indicates the need to focus forecasting and stocking efforts on top-performing items.

5. **Stock Distribution by Category**
- All categories showed wide stock variability, with medians clustered around 120–130 units.
- Outliers in every category indicate both potential overstock and understock risks that could be optimized further.

Further Investigation Plan:-

- **Time-Series Forecasting:** Apply and compare advanced time-series models (ARIMA, Prophet, SARIMA, Holt-Winters) to forecast item-level demand trends over the next 2 months.

- **ABC Categorization:** Use EOQ and sales value to segment items into A, B, and C categories for priority stocking.

- **Category-Level Optimization:** Tailor reorder policies by category based on observed EOQ and demand volatility.

- **Supplier Performance Analysis:** Explore lead time variability and its correlation with stockout events.

- **Dashboards for Monitoring:** Develop visual dashboards using Plotly/Matplotlib to auto-flag items nearing reorder level or incurring repeated stockouts.