Case Study for VP Fire System Private Limited

A Midterm report for the BDM capstone Project

Submitted by

Name: Pragya Tripathi

Roll number: 23f1001699

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1 Executive Summary

VP Fire System Private Limited, established in 2018 and headquartered in Delhi, is a leading B2B provider of fire and security solutions. Despite its innovative offerings and growing presence, VP Fire System faces three critical challenges: accurately forecasting seasonal demand, understanding its customer base for targeted marketing, and improving visibility in key sectors like hotels and hospitals.

This report is based on primary data collected directly from VP Fire System Private Limited, ensuring its authenticity and originality. The data includes sales and purchase records spanning October to December 2024, provided by the business owner. Supporting evidence, such as signed letterhead, images of the organization, and a video interaction with the owner, further validates the transparency and credibility of this research.

After cleaning and preprocessing of the datasets, the preliminary analysis reveals high variability, with extreme outliers influencing overall revenue. Large, infrequent transactions significantly impact forecasting accuracy, making the company dependent on bulk orders. Demand patterns are volatile, with seasonal trends confirmed through STL decomposition. SARIMA modeling has been implemented for improved demand forecasting, addressing seasonality and irregular sales spikes. Customer segmentation analysis highlights a heavy reliance on a few high-value clients, with Honeywell Automation India Ltd. being the largest contributor.

These insights emphasize the need for strategic demand planning and targeted marketing to stabilize sales and become more profitable. VP Fire System can improve demand forecasting by integrating external factors like seasonal trends and festivities into its SARIMA model. Expanding into sectors like hotels and hospitals through targeted marketing will help diversify the customer base and enhance market visibility, reducing reliance on a few high-value clients.

2 Proof of Originality

This section validates the primary data collection, ensuring research authenticity. Supporting materials are available via the following Google Drive link.

The provided evidence includes:

- Datasets (Both Original and Cleaned)
- Official signed and stamped Letterhead
- 6 Images of the organisation and owner.
- Short Video of interaction with the owner.

Google drive Link:
https://drive.google.com/drive/folders/19MVGp_lXh28q-bwWgTuF2pY6ZqflWWR5?usp=sharing

3 Metadata and Descriptive Statistics

3.1 Metadata

The sales and purchase data were originally recorded and maintained by VP Fire System in Excel sheets, likely exported from their internal accounting system. The business owner provided these records as a zip file containing two datasets:

- Sales Data – Three separate Excel sheets for October, November, and December 2024.
- Purchase Data – A single consolidated sheet covering the same period.

To prepare the dataset for analysis, the relevant details were extracted, cleaned, and compiled into two structured datasets: “Sales_Data” and “Purchase_Data”, each spanning October to December 2024**.**

The extracted dataset includes key sales (and purchase) metrics essential for understanding revenue trends and buyer behaviour.

Dataset: Sales_Data.xlsx

	DATE	PARTICULARS	BUYER	QUANTITY	UNIT	RATE	VALUE	TYPE	MONTH
0	2024-10-01	Dismantling and ...	Honeywell Automa...	9	JOB	343.0	3087.0	Labor & Miscella...	October
1	2024-10-01	Dismantling and ...	Honeywell Automa...	14	JOB	1176.0	16464.0	Labor & Miscella...	October
2	2024-10-01	Dismantling and ...	Honeywell Automa...	14	JOB	882.0	12348.0	Labor & Miscella...	October
3	2024-10-01	Dismantling and ...	Honeywell Automa...	2	JOB	2744.0	5488.0	Labor & Miscella...	October
4	2024-10-01	Dismantling and ...	Honeywell Automa...	14	JOB	196.0	2744.0	Labor & Miscella...	October

Figure 1: Structure of Sales Data

Overview:

The dataset “Sales_Data.xlsx” contains transactional data for VP Fire System spanning from October 2024 to December 2024, capturing key details about sales activities. It includes 249 entries and provides insights into products/services sold, clients, and revenue generation.

Data Composition:

- Columns: The dataset comprises 6 columns:
    - **DATE:** Represents the transaction date in DD-MM-YYYY format. This helps analyse seasonal trends and identify peak sales periods, aiding in demand forecasting.

- **PARTICULARS:** Describes the product or service sold (e.g., Smoke Detectors, Installation, AMC). Understanding product-wise sales allows the company to identify best-selling items and adjust inventory or marketing strategies accordingly.
- **TYPE:** Categorizes each transaction as Product, Installation, Maintenance, Labor & Miscellaneous Services, or Infrastructure, aiding in trend analysis.
- **BUYER:** Identifies the client or organization purchasing the product/service. This enables customer segmentation and helps tailor marketing efforts to different buyer groups for better engagement and retention.
- **QUANTITY:** Indicates the number of units or jobs completed. Tracking quantity sold helps in analysing demand fluctuations and optimizing stock levels to prevent shortages.
- **UNIT:** Specifies the measurement standard for each transaction, which varies based on the nature of the product or service. It may represent individual units for products like smoke detectors, meters for cabling or piping installations, jobs for completed services such as system installation or maintenance, or days for labor-based work.
- **RATE:** Specifies the price per unit or job. Monitoring rate trends provides insights into pricing strategies, ensuring competitive pricing while maintaining profitability.
- **VALUE:** The total monetary value of each transaction, calculated as QUANTITY × RATE. This helps in identifying high-value clients and high-revenue products, allowing for better sales strategies and business expansion efforts.

<u>Data Integrity & Anomalies:</u>

- The dataset is complete with no missing values across any columns.
- Wide-ranging values in the VALUE column (₹904.00 to ₹250,000.00) suggest potential outliers or diverse transaction sizes.

Dataset: Purchase_Data.xlsx

	Date	Particulars	Supplier	Quantity	UNITS	Rate	Value
0	2024-10-07	Fire Alarm Panel	JP Fire Safety	1	Unit	4700.0	4700.0
1	2024-10-07	Manual Call Point	JP Fire Safety	2	Unit	325.0	650.0
2	2024-10-07	Hooter	JP Fire Safety	2	Unit	700.0	1400.0
3	2024-10-07	Response Indicator	JP Fire Safety	15	Unit	50.0	750.0
4	2024-10-07	Smoke Detector	JP Fire Safety	30	Unit	570.0	17100.0

Figure 2: Structure of Purchase Data

Overview:

The dataset “Purchase_Data.xlsx” contains transactional data for VP Fire System spanning from October 2024 to December 2024, documenting key procurement details. It includes 90 entries and provides insights into items/services purchased, suppliers, quantities, and costs.

Data Composition:

The dataset structure closely mirrors that of "Sales_Data.xlsx", with the same column names, except:

- It excludes the "TYPE" column, meaning particulars are not categorized.
- Instead of “BUYER”, this dataset includes “SUPPLIER”, identifying vendors providing the goods/services.

Data Characteristic:

- The presence of 3 missing values in the RATE column may require imputation or closer review.
- The VALUE column displays a broad range, suggesting variations in purchase scale, from small to high value equipment and services.

3.2 Descriptive Statistics

Numerical Summary

Value Analysis

	count	mean	std	min	25%	50%	75%	max	IQR	Skewness	Kurtosis
VALUE	249.00	40787.36	84175.06	180.00	6900.00	16464.00	41400.00	850000.00	34500.00	6.28	49.85

Figure 3: Sales Value Distribution

The sales value distribution reveals substantial variability, with a mean of ₹40,787.36 significantly influenced by high-value transactions and a median of ₹16,464.00 indicating that most sales are smaller. The wide range of transaction sizes, from ₹180 to ₹8,50,000, and a high standard deviation of ₹84,175.06 confirm the spread in sales values. An interquartile range (₹6,900–₹41,400) shows that the middle 50% of transactions are moderate-sized, while skewness (6.28) and kurtosis (49.85) highlight the presence of extreme outliers and a positively skewed distribution driven by rare but exceptionally large transactions.

This variability complicates demand forecasting and revenue prediction due to the disproportionate impact of high-value transactions on overall sales. The company may be overly reliant on premium or bulk orders, making it vulnerable to fluctuations in these sales.

Rate Analysis

	count	mean	std	min	25%	50%	75%	max	IQR	Skewness	Kurtosis
RATE	247.00	16809.25	34589.21	9.00	1600.00	3465.00	16000.00	250000.00	14400.00	3.99	18.68

Quantity Analysis by Unit

	UNIT	Days	JOB	MTR	Unit
QUANTITY	count	1.00	120.00	10.00	118.00
	mean	90.00	8.65	291.10	30.40
	min	90.00	1.00	21.00	1.00
	25%	90.00	1.00	131.25	2.00
	50%	90.00	1.00	200.00	4.00
	75%	90.00	3.00	300.00	12.50
	max	90.00	385.00	800.00	500.00

Figure 4: Quantity and Rate Summary Across Unit Types

The chart reveals significant variations in both quantity and rate across different unit types. JOB-based offerings exhibit high variability in quantity, with a median of 1 but a maximum of 385, suggesting that while most jobs are small, occasional large projects inflate the mean (8.65). Similarly, JOB rates show wide dispersion, ranging from ₹1 to ₹250,000, with a high mean of ₹26,588.69, reflecting diverse pricing patterns. MTR-based products demonstrate bulk purchasing behavior with the highest average quantity (291.10) and a maximum of 800, while rates remain relatively stable. Unit-based items show positively skewed distributions for both quantity and rate, with most orders being small (median quantity: 4), but occasional large orders (maximum quantity: 500) and high rates (maximum: ₹88,500) significantly influencing the averages.

The distinct purchasing patterns suggest diverse customer needs requiring tailored strategies. Additionally, the MTR-based products are in high quantities among all unit types, indicating bulk purchasing behavior. This suggests these products are frequently required in large-scale applications presenting an opportunity to target new segments like hotels and hospitals effectively.

<u>**Categorical Analysis**</u>

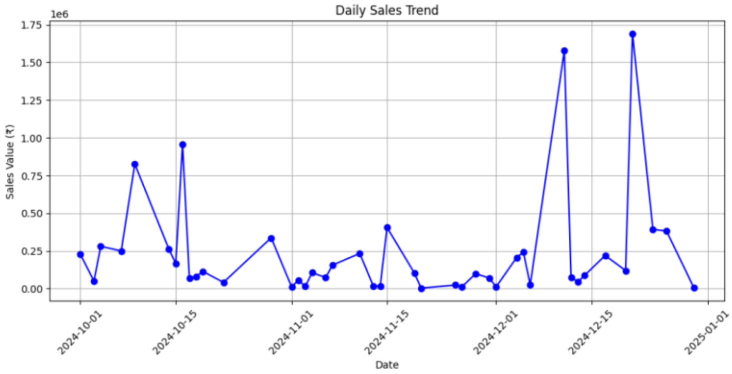

Figure 5: Daily Sales Trend Chart

The daily sales trend shows significant fluctuations in sales values over time, with sporadic spikes in October, November, and December. These peaks, particularly in mid-October and mid-December, align with the high standard deviation observed in the numerical statistics, reflecting the extreme variability in sales during these periods. The mean sales value is skewed upward due to these high-demand spikes, while the median remains relatively low, indicating that most days have minimal sales activity. The irregularity in sales patterns is further emphasized by the sporadic nature of these spikes, making demand forecasting challenging without incorporating additional contextual factors like holidays or promotions.

4 Detailed Explanation of Analysis Process/Method

4.1 Data Cleaning and Preprocessing

The original dataset required thorough cleaning and preprocessing to ensure accurate analysis. Key preprocessing steps included:

- **Convert DATE to a proper datetime format:** The "DATE" column which is in string format is converted into datetime format to manipulate time-series data.
- **Fill missing dates with zero sales:** Since missing dates indicate zero transactions rather than missing data, filling gaps maintained the continuity of the time series, preventing distortions in trend analysis.
- **Aggregate sales per day**: As there are multiple transactions per day, we sum up the "VALUE" column for each day to get the total daily sales as ARIMA/SARIMA works best with a single time series value per timestamp.
- **Standardize Customer Names:** Removed inconsistencies in buyer names (e.g., "H&M Hennes & Mauritz Retail Pvt. Ltd." vs. "H&M HENNES AND MAURITZ INDIA PRIVATE LIMITED") for accurate customer segmentation.
- **Handling Different Units for Quantity and Rate:** The dataset contained various unit types for both QUANTITY and RATE, such as JOB, Unit, MTR, Days. For more structured analysis,

ensuring that calculations like averages and trends were meaningful across different product types, this inconsistency was dealt by introducing a new column called UNIT categorizing each transaction based on its respective unit type.

- **Categorizing Products and Services**: The company deals with both products (such as smoke detectors, control modules, and cables) and services (such as installation, maintenance, and dismantling). To enhance analysis, a new column "TYPE" was introduced, classifying each transaction into categories like Product, Installation, Maintenance, Labor & Miscellaneous Services, and Infrastructure. This helps in filtering and analysing sales trends more effectively.

4.2 SARIMA for Demand Forecasting

The daily sales trend chart reveals significant fluctuations in sales, with multiple sharp peaks and long periods of low or zero sales. This indicates a highly volatile demand pattern rather than a steady trend. Since the seasonality appeared dynamic rather than static, Seasonal-Trend Decomposition (STL) was used, which confirmed the presence of seasonality in the data.

Traditional methods like Simple Moving Average (SMA) or Linear Regression fail to capture irregular sales spikes and seasonality, making them unsuitable for this dataset. ARIMA is effective for trends and autoregressive patterns but does not explicitly model seasonality. Instead, SARIMA (Seasonal ARIMA) is preferred, as it integrates both seasonal and non-seasonal components, improving demand forecasting.

SARIMA is defined as:

$$\mathrm{SARIMA}(p,d,q)\times(P,D,Q,s)$$

where:
p,d,q represent the non-seasonal autoregressive, differencing, and moving average terms.
P,D,Q,s represent the seasonal counterparts with seasonality

For the analysis, SARIMA(1,0,1)(0,0,1,7) was selected which effectively captures regular weekly patterns and general trends but lacks the flexibility to handle irregular events like holiday-driven sales spikes. This limitation stems from SARIMA's assumption of consistent patterns rather than exceptional events. A more appropriate approach would involve SARIMAX with holiday indicators and Prophet’s decomposition framework, which models both regular seasonality and exceptional events, making it more adaptable to fluctuations in sales patterns.

4.3 Customer Segmentation

The Top 10 Customers by Revenue chart highlights a significant disparity in the customer base, with a few high-value buyers contributing disproportionately to total revenue. To further explore these patterns, boxplots of RFM metrics (Recency, Frequency, Monetary) were created, confirming the skewed distribution. The monetary boxplot revealed high-value outliers, while the frequency and recency boxplots showed varying levels of engagement, suggesting the need for a more structured segmentation approach.

RFM scoring and K-Means clustering will be performed to categorize customers based on their purchasing behaviour. This method is preferred over alternatives like decision trees or random forests, which require predefined labels and may overfit transactional data with limited features. Unlike rule-based segmentation, which relies on static thresholds, clustering can capture the dynamic relationships between recency, frequency, and monetary value, providing a more adaptive and data-driven approach to customer segmentation.

5 Results and Findings

5.1 Seasonal Demand

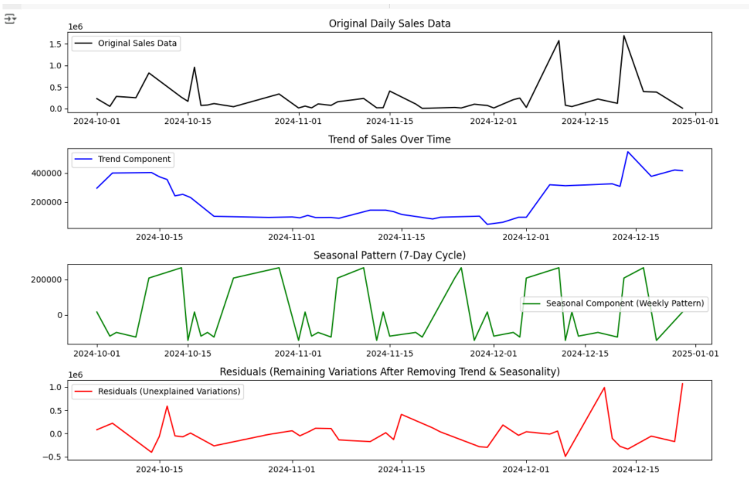

Figure 6: Seasonal Decomposition of Time Series

The Seasonal-Trend Decomposition (STL) reveals key sales patterns affecting demand forecasting. The trend component shows a dip in November, aligning with the company's challenge of manpower and inventory shortages leading to lost sales. The seasonal component highlights a consistent weekly pattern, proving that day-of-week effects remain crucial even during slow periods. The residuals capture sharp spikes in December, likely driven by new year festive shopping and promotions, emphasizing the need for better forecasting to manage demand surges.

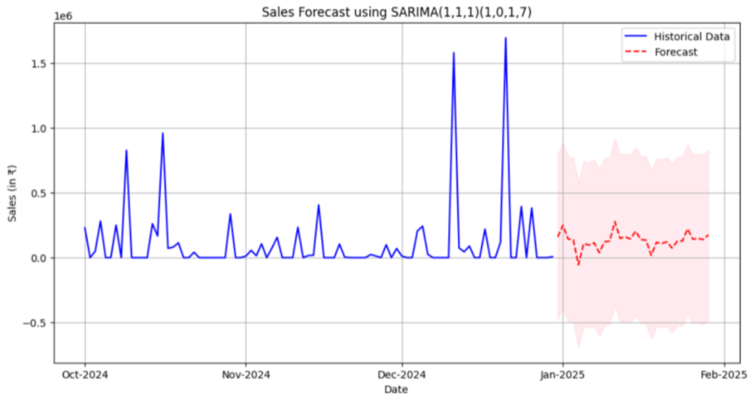

Figure 7: Sales Forecasting using SARIMA

The SARIMA(1,1,1)(1,0,1,7) model effectively captures the sales dynamics by implementing first-order differencing (d=1) to address the non-stationarity confirmed by the ADF test. The weekly seasonality parameter (s=7) accurately models the consistent seven-day sales pattern clearly visible in the seasonal component of the STL decomposition. The reliable forecast is validated by the Ljung-Box test results (p-values>0.05), confirming that residuals exhibit no significant autocorrelation.

5.2 Customer Behavior Insights

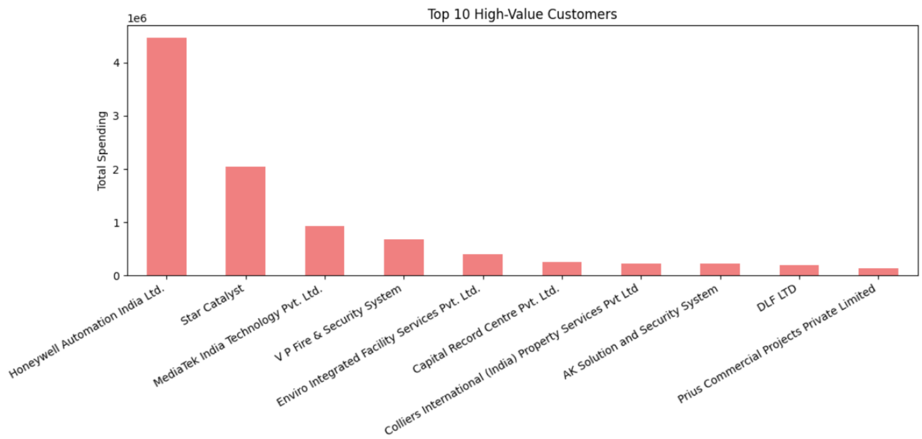

Figure 8: Top 10 High-Value Customers Chart

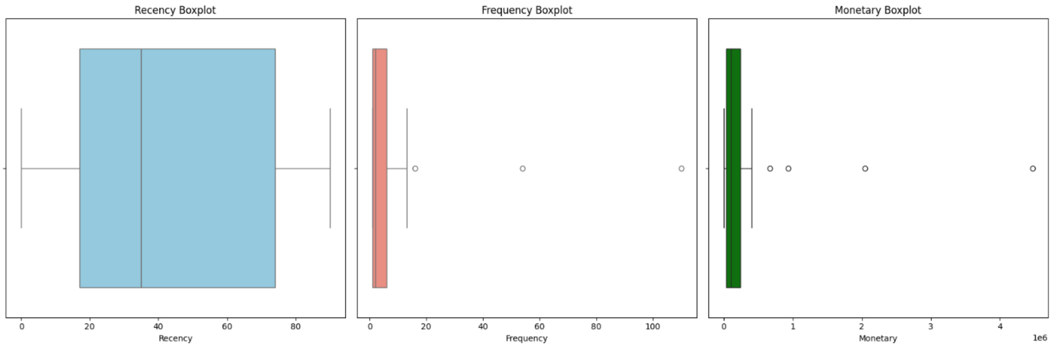

Figure 9: Boxplots of RFM Metrics

The RFM boxplots provide key insights into customer behavior. Recency shows that most purchases occur within 20-60 days, indicating regular engagement. Frequency is highly right-skewed, with a few customers making significantly more purchases than the rest. Monetary exhibits the strongest skewness, where most customers contribute minimal revenue, but a few outliers spend up to 4.5 million.

The monetary boxplot and Top 10 High Value Customers Chart reveal a classic Pareto distribution, where a small percentage of customers (e.g., Honeywell Automation India Ltd. with ~4.4M and Star Catalyst with ~2M) contribute disproportionately to revenue.