Data-Driven Optimization for B2B HVAC&R Operations

Final report for the BDM capstone Project

Submitted by

Name: Yash Mishra

Roll number: 21f1006461

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive Summary

Blue Star, India's foremost Heating, Ventilation, Air Conditioning, and Refrigeration Company, has its Uttar Pradesh branch office, managed by Area Manager Mr. Pravesh Upadhyay, primarily engages in B2B dealings. They achieved an impressive revenue exceeding ₹12 crores in the HVAC&R sector during the 2022-23 fiscal year. Despite this success, operational challenges due to the absence of a robust data-driven approach hinder decision-making and efficiency.

Recognizing the need for transformative strategies, Mr. Upadhyay aims to acquaint himself and his team with the requirements, processes, and strategies involved in data-driven decision-making. Leveraging SAP Point of Sale (PoS) system, which was used for data collection, the project focuses on enhancing data literacy, showcasing strategies involved in data analysis, evaluating retailer performance and studying demand and supply dynamics. This initiative reflects Blue Star's commitment to technological advancements, fortifying its position in the dynamic HVAC&R industry by fostering adaptability, efficiency, and strategic growth.

Mr. Upadhyay actively explores data intricacies, emphasizing simplified techniques to enhance team efficiency. Meticulously collected sales data shared through SAP software initiates discussions on leveraging data effectively. The project's focus includes evaluating retailer performance and gaining insights into demand and supply dynamics in the B2B HVAC&R sector, ultimately enhancing Blue Star's strategic positioning.

This initiative, spearheaded by Mr. Upadhyay, envisions not only improving data-driven decision-making but also fostering a culture of adaptability and efficiency within Blue Star. The company's background showcases its leadership in the HVAC&R industry, with a robust integrated business model, extensive market presence, and dedication to maintaining industry standards. This project serves to identify bottlenecks and offers solutions to enhance operational efficiency and strategic excellence for sustained growth in the competitive market.

2. Detailed Explanation of Analysis Process/Method

We started with python using "Spyder" to study the various properties of the data provided by Mr. Upadhyay excluding the last row as that was the total row.

```
#%%
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_excel(r"C:\Users\dell\OneDrive\Desktop\bdm ac\Billing Apr22-- 29 mar23 33 div.XLSX")
d=data.copy(deep=True)
#%% dropping the total col
d = d.drop(d.index[-1])
```

Figure 1. Data Loading and Processing

A total of 62 columns were present in the dataset. After discussions we agreed to narrow down our study to "seven plus one" columns. Seven were directly from the dataset and one additional column of "Missing Values".

```
In [3]: len(d.columns)
Out[3]: 62
```

*Figure **Error! No text of specified style in document.**.2. Degree of the Original Dataset*

2.1Total Sales Value

A series of descriptive statistical analyses on the 'Total sales value' column were conducted.

- The .describe() method was initially used to obtain key descriptive statistics for the 'Total sales value' column. This includes the count, mean, standard deviation, minimum, 25th percentile (Q1), median (50th percentile or Q2), 75th percentile (Q3), and maximum values.

```
In [7]: #%%#%% descriptive stats total sales value
   ...: print('DESCRIBE\n',d['Total sales value'].describe())
   ...: print('\nSUM\n',d['Total sales value'].sum())
   ...: print('\nMEAN\n',d['Total sales value'].mean())
   ...: print('\nOrders per Day\n',d['Total sales value'].count()/365)
   ...: print('\nLOWEST SALE\n',d['Total sales value'].abs().min())
   ...: print('\nHIGHEST NEGATIVE\n',d['Total sales value'].min())
   ...: print('\nTOTAL NEGATIVE SALES\n',d['Total sales value'][d['Total
sales value'] < 0].sum())
```

Figure 3 Code for Descriptive Statistics of Total Sales Value

- The .sum() method was employed to calculate the total sum of the 'Total sales value' column, providing an overall monetary value for the dataset.
- The .mean() method calculated the average (mean) value of the 'Total sales value,' providing insight into the typical value per data point.
- The number of orders per day was estimated by dividing the total count of data points by the number of days. This was done to provide an average daily order rate.
- To address negative values, the .abs() method was applied to obtain the absolute values, allowing for a focus on the magnitude rather than the direction of the sales values.
- The minimum absolute value was identified using .abs().min(), revealing the smallest positive value in magnitude.
- The sum of negative values in the 'Total sales value' column was calculated to understand the cumulative impact of negative sales.

2.2Categorical Columns

Then we conducted a detailed analysis of other five categorical columns in the modified dataset using pandas. The choice of these specific columns, namely 'Material', 'Ship to party name', 'Retailer', 'Sales Unit', and 'Segment,' were made for a deliberate focus on key dimensions of the dataset that were likely relevant for business analysis in the context of sales and business operations.

1. The 'Material' column often represents the unique identifier for a specific model. Analyzing sales performance or other metrics based on materials provided insights into the popularity, profitability, and contribution of different products to overall revenue.

2. 'Ship to Party Name' column refers to the name of the entity or customer to whom the products were shipped. Analyzing sales by customer revealed patterns in customer preferences, identifying high-value clients, guiding customer relationship management strategies.

3. The 'Retailer' column focused on understanding sales dynamics across different retailers. Analyzing sales by retailer helped in assessing the performance of various sales channels, identifying top-performing retailers, and tailoring marketing strategies for different retail partners.

4. The 'Sales Unit' column represented sales entity responsible for leads. Analyzing sales by unit provided insights into the efficiency of different sales entities, helping optimize resource allocation and improve overall sales performance.

5. The 'Segment' column is associated with the segmentation of customers based on criteria such as, market segment, product category. Analyzing sales by segment enables a deeper understanding of the performance of different product categories and customer groups, supporting strategic decision-making.

These columns appeared to represent critical dimensions in the sales and business operations domain. Analyzing these dimensions provided a granular understanding of product performance, customer relationships, retail partnerships, sales entities, and market segments. The insights gained from this analysis informed strategic decisions, marketing strategies, and resource allocation within the business. The specific choice of columns aligns with common business practices and objectives in sales analysis and optimization.

```
In [9]: #%% Descriptive stats
   ...: c=['Material','Ship to party name', 'Retailer', 'Sales Unit', 'Segment', ]
   ...: for p in c:
   ...:     print(p)
   ...:     print('LENGTH of UNIQUE:',len(d[p].unique()) - (pd.isna(d[p].unique()).sum()))
   ...:     print('\nDESCRIBE :\n',d[p].describe())
   ...:     print('\nVALUE COUNTS :\n',d[p].value_counts().head(10))
   ...:     # print('\nTOP COUNTS :\n',d.groupby(p).apply(lambda x: x['Total sales
value'].count()).sort_values(ascending=False).head(10))
   ...:     print('\nTotal Sales Value :\n',d.groupby(p).apply(lambda x: x['Total sales
value'].sum()).sort_values(ascending=False).head(10))
```

Figure 4. Code to generate Descriptive Statistics for Categorical Columns

Code Explanation:

1. Length of Unique Values: This section calculates the number of unique non-null values for each categorical column, providing insight into the diversity or cardinality of the categories.

2. Descriptive Statistics: Descriptive statistics for categorical columns are a bit different such as count, unique, top, frequent are presented.

   2.1 The 'count' indicates the number of non-null (non-missing) entries column.
   2.2 The 'unique' value represents the number of unique categories or distinct values present.
   2.3 'Top' indicates the most frequently occurring category.
   2.4 'Freq' specifies the frequency of the top category, meaning the number of occurrences of the most common category.

3. Value Counts (Top 10): The value_counts() method is used to display the count of each unique value in the categorical column. This helps identify the most frequent categories and their respective counts, focusing on the top 10.

4. Total Sales Value by Category (Top 10): The data is grouped by the categorical column, and the total sales value is calculated for each group. This section reveals the top 10 categories with

the highest total sales, allowing for an understanding of the sales distribution across different categorical values.

Thus, by employing these analyses, a comprehensive understanding of the unique values, distribution, and total sales contributions associated with each categorical column in the dataset were gained.

2.3 Missing Values Column

Then we focused on analyzing and visualizing null values in each column of the dataset, including the creation of an additional column to represent "Missing Values."

1. Counting NULL values in each column: This loop iterates through each column in the dataset and prints the count of null values for each column. The results are stored in a dictionary where the column name is the key, and the corresponding count of null values is the value.

```python
#%% Null values in each col
nd={}
for i in d.columns:
    print(i,":",d[i].isna().sum())
    nd[i]=d[i].isna().sum()
```

Figure 5. Null Values in each Column

2. Visualizing Null Values: A data frame is created using the dictionary of null values. This data frame is then transposed to have columns as indices. A heat map is generated using seaborn library (sns.heatmap) to visualize the null values in each column. Columns are on the x-axis, and the y-axis represents whether values are missing or not. The resulting heat map provides a clear overview of which columns contain missing values and the extent of the missing data.

```python
#%% visualizing nulls
ndf = pd.DataFrame(list(nd.items()),
columns=['Column Name','Number of NA'])
ndf = ndf.set_index('Column Name')
ndft=ndf.T
plt.figure(figsize=(18,6))
sns.heatmap(ndft)
plt.show()
```

Figure 6. Visualizing all Columns

3. Separating Columns with Non-Zero Null Values: Another dictionary (nod) is created to store columns that have non-zero null values. This step filters out columns with no missing values.

```python
#%% 3. separating cols that have 0 null
nod={}
for i in d.columns:
    if d[i].isna().sum() != 0:
        nod[i]=d[i].isna().sum()
```

Figure 7. Columns without Non-Zero Null Values

4. Displaying Columns with Non-Zero Null Values: Similar to the previous heat map, a new heat map is generated specifically for columns with non-zero null values. The resulting heat map provides a focused view of columns where missing values are present.

```python
nodf = pd.DataFrame(list(nod.items()),
columns=['Column Name','Number of NA'])
nodf = nodf.set_index('Column Name')
nodft=ndf.T
plt.figure(figsize=(10,6))
sns.heatmap(nodf,annot=True,fmt='.0f',
            cmap='magma')
plt.title('Heatmap of missing values',
fontdict={'weight': 'bold','size' : 16},
            pad=12)
plt.show()
```

Figure 8. Visualizing Columns having Null Values

2.4 Visualizations

For data visualizations both python and excel were used to leverage the strengths of each tool for different aspects of the analysis.

Python

- Python is well-suited for handling data manipulation and creating sophisticated visualizations, providing flexibility and customization options.
- With libraries such as pandas and seaborn, it was used for the initial data processing, counting null values, and creating a heat map to visualize the null values in each column.
- Python was chosen for its simplicity and ease of use in handling data manipulation and creating basic visualizations. Its straightforward syntax and rich libraries made it accessible for initial analysis.

Excel

- Excel was used for additional visualizations, particularly for specific subsets of the data or for creating visually appealing reports.
- Mainly Pivot tables were created to draw graphs
- As it is accessible to a broader audience, including stakeholders who prefer a familiar interface for data exploration and reporting it could help the team to adapt efficiently.

The inclusion of different tools reflects a sense of choice, recognizing that different tools can serve different purposes. It acknowledges that Mr. Upadhyay has the option to choose the tool that best aligns with the specific requirements and preferences of the analysis.

3. Results and Findings

3.1 Sales Unit and Retailers:

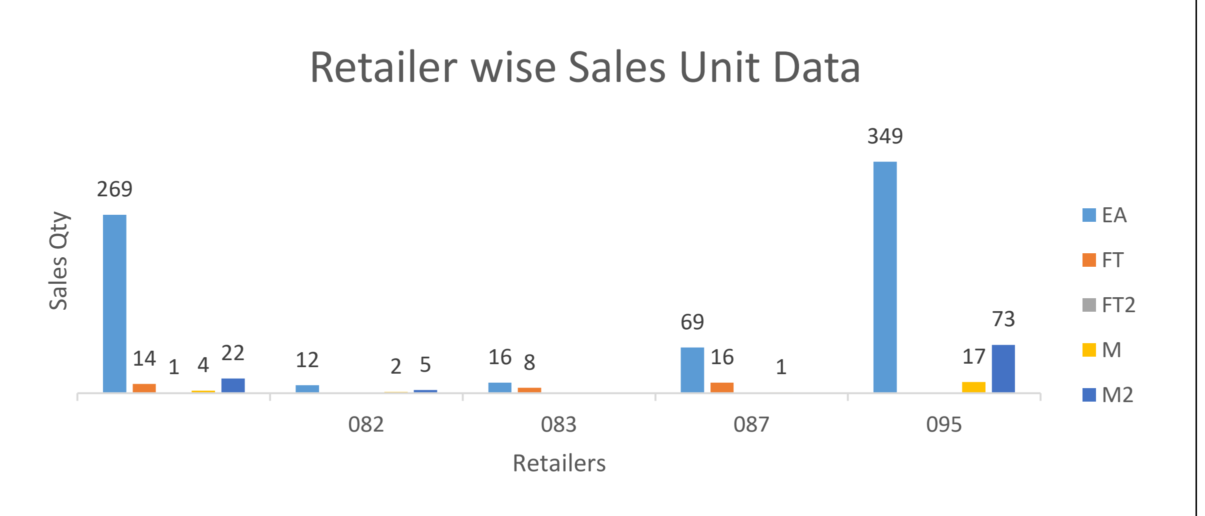


Figure 9. Retailer wise Sales Unit Data

3.1.1 Sales Unit Performance:

The performance analysis of different sales units provides crucial insights into their effectiveness in generating leads and contributing to the overall sales value.

<u>Top Performer: “EA” Sales Unit</u>
The “EA” sales unit emerged as the standout performer, contributing significantly to Blue Star's success. This sales unit generated an impressive 90.33% of the total sales value, showcasing its pivotal role in driving revenue. With a substantial contribution of 715 leads, the “EA” unit not only excelled in terms of value but also demonstrated efficiency in lead generation.

<u>Underperformer: “FT2” Sales Unit</u>
In contrast, the “FT2” sales unit struggled to make any impact, accomplishing only 1 lead. The minimal contribution of “FT2” to the overall sales indicates potential challenges or inefficiencies within this sales unit.

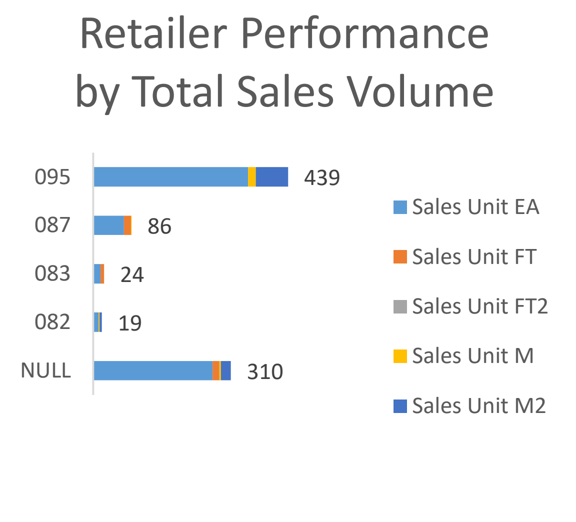


Figure 10. Retailer Performance by Total Sales Volume

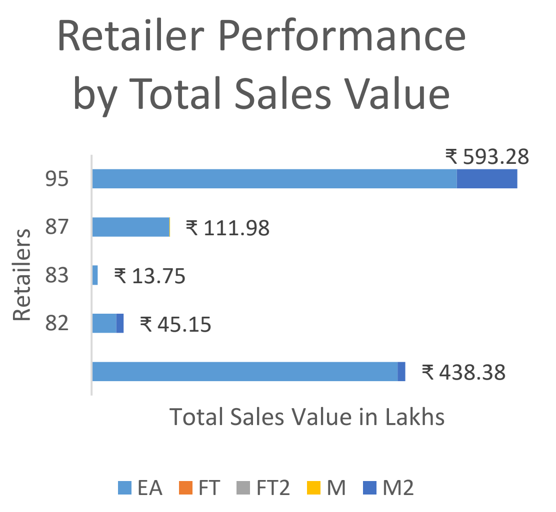


Figure 11. Retailer Performance by Total Sales Value

3.1.2 Retailer Contribution:

Analyzing the performance of different retailers sheds light on their individual contributions to the total sales volume and value.

<u>Top Performer: Retailer “095”</u>
Retailer “095” emerged as the top-performing entity, making a substantial impact on Blue Star's sales outcomes. This retailer accounts for an impressive 50% of the total sales volume, indicating a significant market share. In terms of total sales value, “095” contributes 49.33% to the overall revenue, showcasing its importance in driving financial success.

<u>Underperformer: Retailer “082” and “083”</u>
Retailer “083” and “083” presented contrasting performance metrics, with the lowest total sales volume of 19 orders by “082”. The total sales value generated by “083” is ₹13,74,850.00, marking it as the retailer with the lowest financial contribution.

3.2 Total Sales

Understanding the variability in total sales values is crucial for assessing the risk and uncertainty associated with revenue generation. Here, we delve into the patterns and characteristics of total sales values across different segments and retailers.

```
DESCRIBE
 count       878.00
mean     136962.92
std      197103.65
min     -941028.00
25%       32587.50
50%       85209.58
75%      188154.00
max     1785000.00
Name: Total sales value, dtype: float64
```

Figure 12. Output for Descriptive Statistics of Total Sales Value

```
SUM
 120253441.23000002
MEAN
 136962.91711845092
Orders per Day
 2.4054794520547946
LOWEST SALE
 490.0
HIGHEST NEGATIVE
 -941028.0
TOTAL NEGATIVE SALES
 -5971040.91
```

Figure 13. Other customized Statistics

<u>Significant Variability:</u>

Total sales values exhibit a notable degree of variability, showcasing fluctuations across diverse segments and retailers. The variance in sales values indicates that certain segments or retailers may pose a higher level of risk and unpredictability in terms of revenue generation.

<u>Positive and Negative Values:</u>

The presence of both positive and negative values in the "Total sales value" column suggests diverse transaction outcomes. Positive values represent successful sales transactions, while negative values indicate potential issues such as refunds, canceled orders, or adjustments.

3.3 Demand and Supply Dynamics:

We explored and analyzed the performance of different segments within the HVAC&R industry, with a specific focus on identifying dominant segments based on sales values. The goal is to understand the contribution of each segment to the overall sales volume and value, providing strategic insights for decision-making.

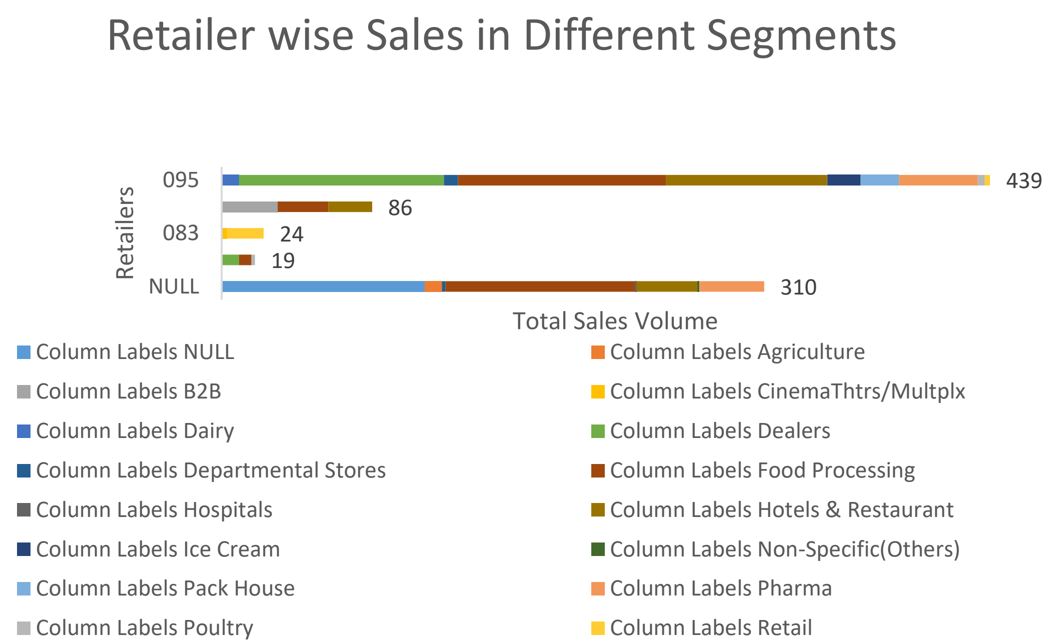


Figure 14. Retailer wise Sales in Different Segments

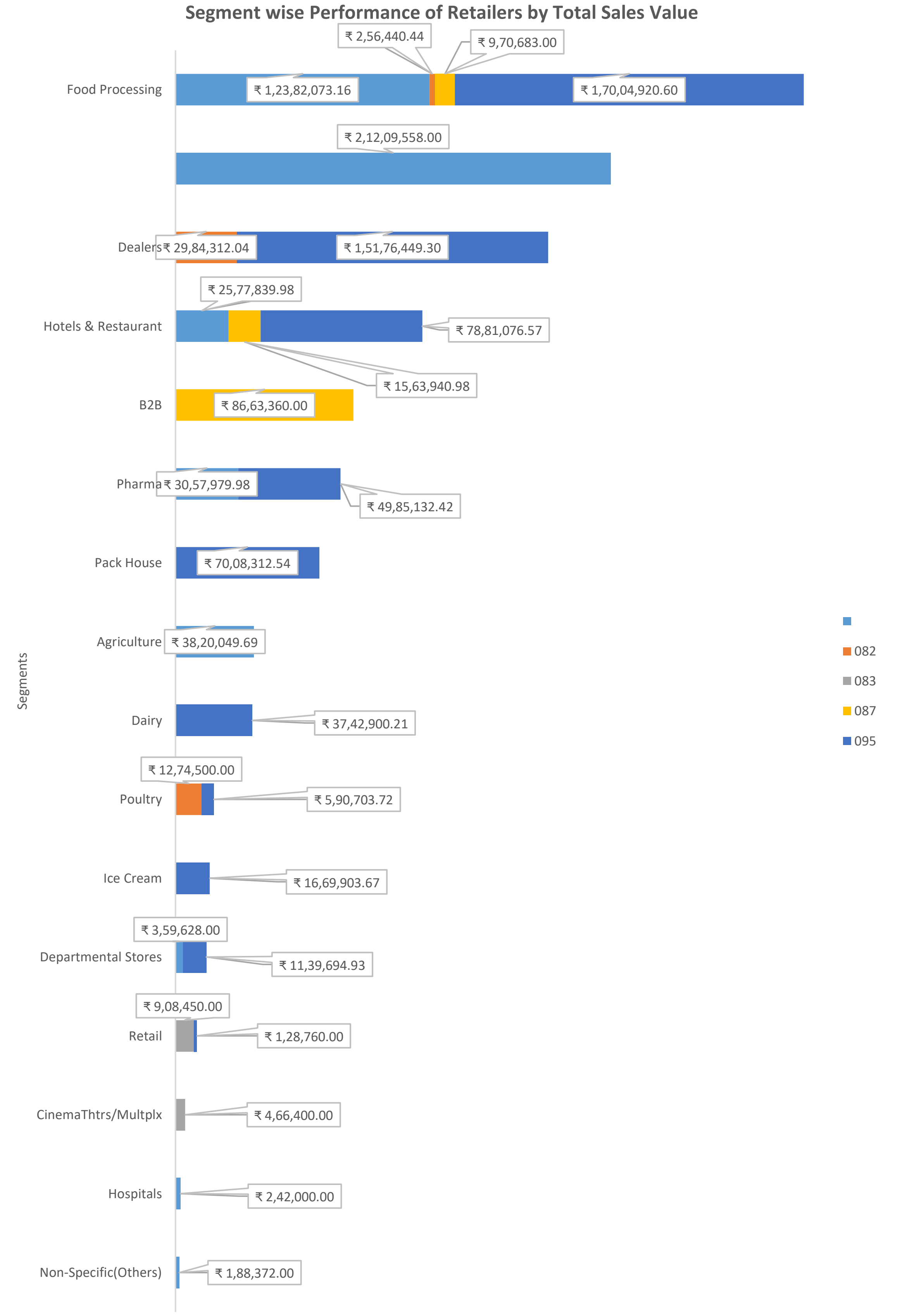


Figure 15. Segment wise Performance of Retailers by Total Sales Value

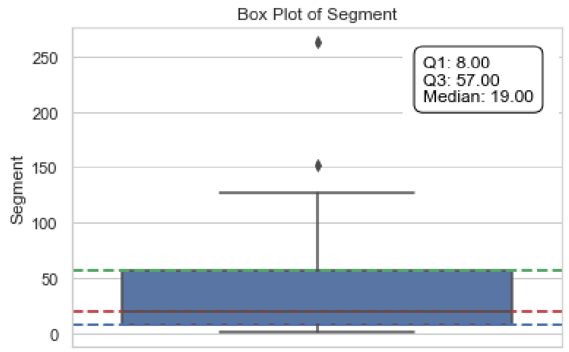


Figure 16. Box Plot of Segments by Total Sales Volume

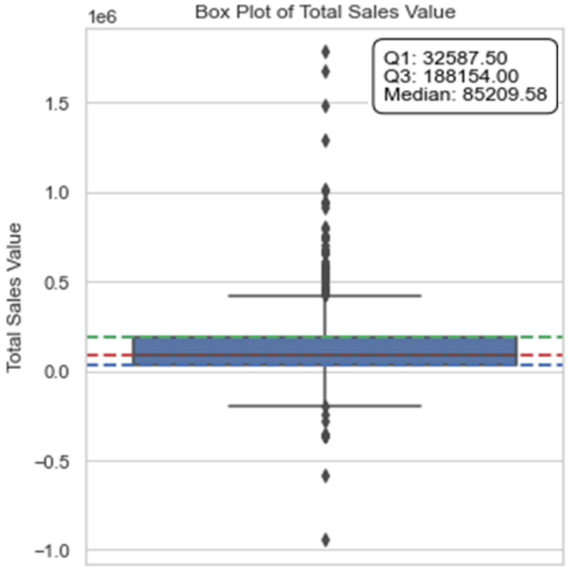


Figure 17. Box Plot of Total Sales Value

3.3.1 Dominant Segments:

Dominance:

Two standout segments, namely "Food Processing" and "Pharma", emerged as dominant contributors to the overall sales values. These segments exhibit high sales values, signifying their significant impact on Blue Star's revenue in the HVAC&R sector. The dominance of these segments suggests a robust demand for HVAC&R products within the food processing and pharmaceutical industries.

Other Significant Contributors:

Beyond "Food Processing" and "Pharma," other segments also make substantial contributions to the overall sales performance. "Hotels & Restaurant," "Dealers," and "B2B" emerge as noteworthy segments that significantly contribute to the overall sales volume and value. The presence of multiple segments with substantial contributions highlights the diverse market presence of Blue Star across various sectors within the HVAC&R industry.

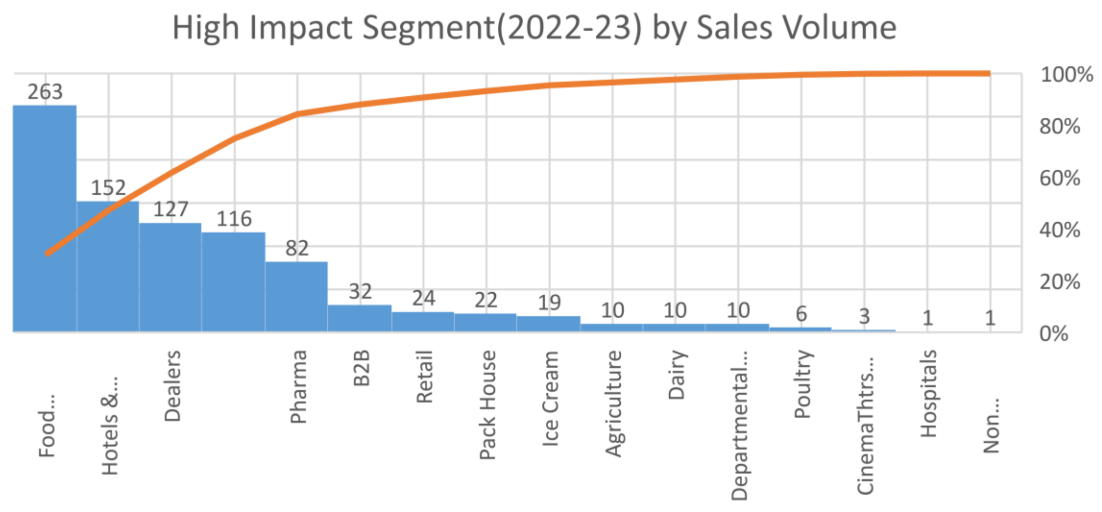


Figure 18. High Impact Segment(2022-23) by Sales Volume

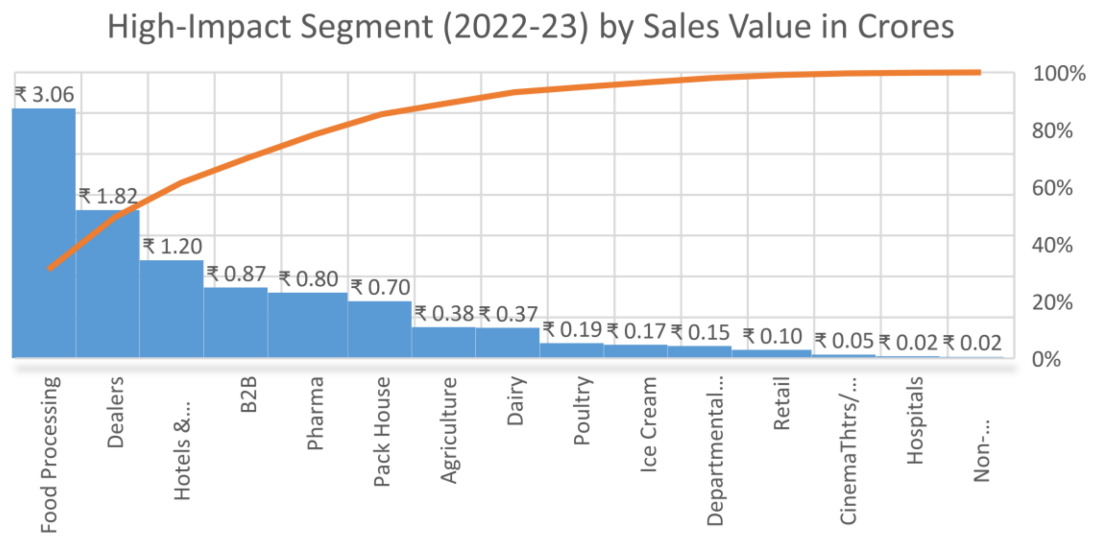


Figure 19. High-Impact Segment (2022-23) by Sales Value in Crores

3.3.2 Detailed Segment Analysis:

Top Segment: "Food Processing"
"Food Processing" emerges as the top-performing segment, representing 29.95% of the total sales volume. The segment recorded sales for 263 products, amounting to a total worth of ₹3,06,14,117.20. In terms of overall sales value, "Food Processing" contributes 25.45% to the total revenue, emphasizing its significance in financial terms.

Lowest Performers: "Hospitals" and "Non-Specific(Others)"
In contrast, some segments, such as "Hospitals" and "Non-Specific(Others)," exhibit lower sales volumes, each contributing only 1 to the total sales volume.

3.4 Seasonal Trends:

We delved into time frame data to identify the seasonal trends observed in Blue Star's HVAC&R sales. By examining quarterly variations and specific segment performances, we gained valuable insights into the temporal dynamics of the business.

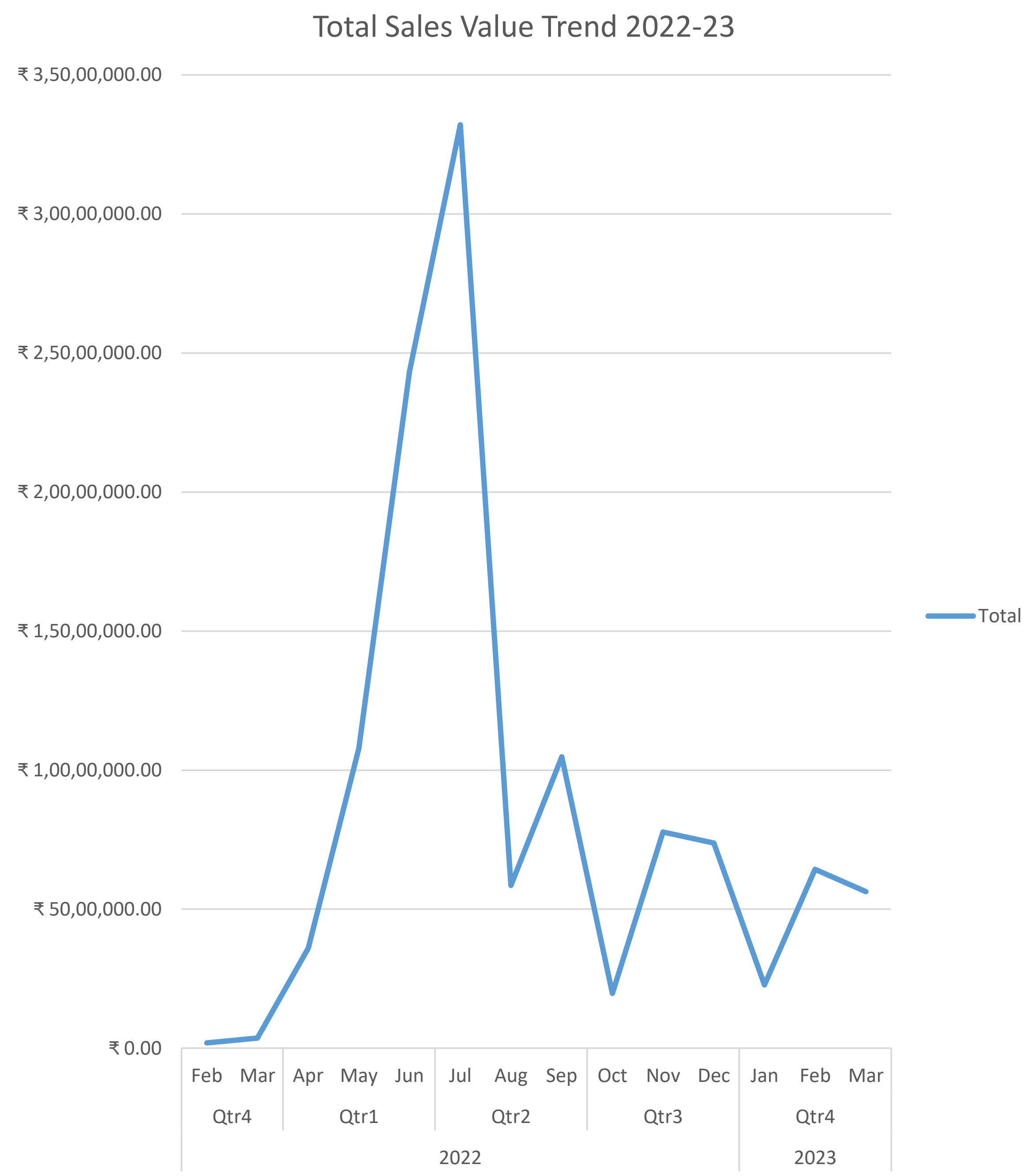


Figure 20. Total Sales Value Trend 2022-23

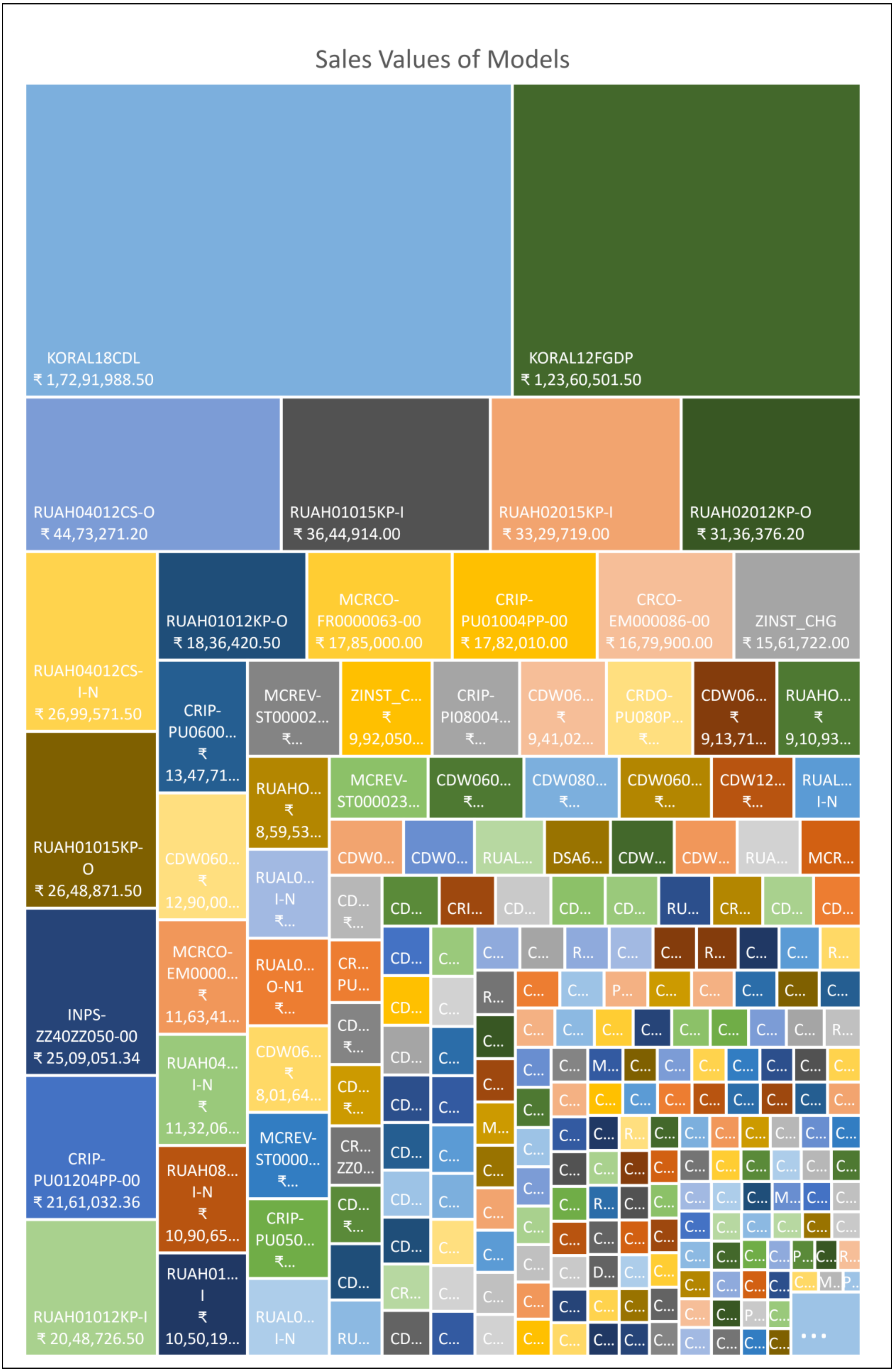
<u>Quarterly Sales Trends:</u>

Trends in total sales values exhibited noticeable fluctuations across different quarters. In Qtr2 of 2022, sales values reached ₹4,95,41,789.40, indicating a substantial increase. On the other hand, Qtr4 of 2023 witnessed a significant rise in sales, reaching ₹1,43,32,317.00, possibly influenced by expanding market demands.

Seasonal Correlation with Sales Quantity:
Seasonal trends in sales quantities correlated with peaks occurring in the summer months. The aggregate total sales value during May 2022 was ₹1,07,87,286.00. However, interestingly, certain segments, including Poultry, Cinema, and Stores, experienced growth after the summer season. For instance, in Jan 2023, Poultry, Cinema, and Departmental Stores contributed ₹38,30,926.65.

3.5 Material/Model Analysis:

In the examination of the Material/Model aspect, a total of 245 products spanning various categories were sold, contributing to Blue Star's overall sales performance.



Sales Values of Models
KORAL18CDL
₹ 1,72,91,988.50
KORAL12FGDP
₹ 1,23,60,501.50
RUAH04012CS-O
₹ 44,73,271.20
RUAH01015KP-I
₹ 36,44,914.00
RUAH02015KP-I
₹ 33,29,719.00
RUAH02012KP-O
₹ 31,36,376.20
RUAH04012CS-I-N
₹ 26,99,571.50
RUAH01012KP-O
₹ 18,36,420.50
MCRCO-FR0000063-00
₹ 17,85,000.00
CRIP-PU01004PP-00
₹ 17,82,010.00
CRCO-EM000086-00
₹ 16,79,900.00
ZINST_CHG
₹ 15,61,722.00
RUAH01015KP-O
₹ 26,48,871.50
INPS-ZZ40ZZ050-00
₹ 25,09,051.34
CRIP-PU01204PP-00
₹ 21,61,032.36
RUAH01012KP-I
₹ 20,48,726.50

Figure 21. Sales Values of Models

3.5.1 Total Sales Volume:

A comprehensive inventory of 245 products, spread across different categories, was successfully marketed and sold. Among the diverse range, specific models stood out in terms of total sales volume. Notably, the top-performing models included "INPS-ZZ40ZZ050-00," "KORAL12FGDP," and "KORAL18CDL." The sales volumes for these models were impressive, with "INPS-ZZ40ZZ050-00" leading the pack by selling 60 units, followed closely by "KORAL12FGDP" with 56 units and "KORAL18CDL" with 52 units.

3.5.2 Total Sales Value:

The success of models was not solely gauged by sales volume; the analysis extended to total sales value, providing a monetary perspective. The models that emerged as the most financially successful were "KORAL18CDL" and "KORAL12FGDP." These models generated substantial revenue, contributing ₹1,72,91,988.50 and ₹1,23,60,501.50, respectively.

Despite a lower sales volume compared to "INPS-ZZ40ZZ050-00," the financial performance of "KORAL18CDL" and "KORAL12FGDP" showcased their higher value propositions, demonstrating their significance in driving revenue for Blue Star.

3.6 Missing Values:

We explored the pervasive issue of missing values within the dataset, primarily attributed to data illiteracy. The challenges posed by missing data were further analyzed, shedding light on the data quality issues faced by Blue Star in its HVAC&R sales records.

Figure 22. Number of missing values

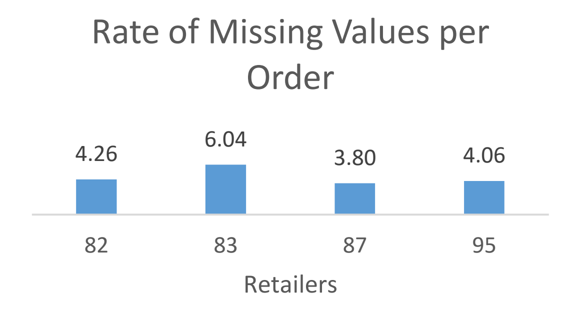


Figure 23. Rate of Missing Values per Order

3.6.1 Data Quality Issues:

<u>Prevalence of Missing Values:</u>

The analysis revealed a substantial problem of missing values, indicating potential gaps in the dataset. This challenge was primarily attributed to data illiteracy, suggesting a need for enhanced training and awareness.

Retailer-Specific Analysis:

Among the sales units, "095" emerged with the highest number of missing values. Notably, retailer "083" exhibited a recurrent pattern of failing to record data, missing valuable opportunities to capture industry-specific insights.

Impact on Industry-Specific Data:

The missing data from retailer "083" and “095” raised concerns, as it hindered the comprehensive study of industry-specific trends. The absence of this data posed challenges to the depth of the analysis, potentially obscuring crucial insights.

3.6.2 Analysis Methodology:

Operational Successes:

The analysis, conducted using both Python and Excel, uncovered operational successes, such as the dominance of the "EA" sales unit and the significant contribution of retailer "095" to the total sales volume and value.

Challenges Identified:

Despite operational successes, challenges were identified, including a high standard deviation in total sales, negative sales figures, and, notably, missing data. These challenges necessitated strategic focus to ensure a comprehensive understanding of the B2B HVAC&R industry landscape.

4. Interpretation of Results and Recommendation

A thorough analysis of Blue Star's B2B HVAC&R operations reveals a set of strategic recommendations. Firstly, replicating successful strategies from the "EA" sales unit across other units is advised, aiming to enhance overall unit performance. Additionally, addressing challenges faced by the "FT2" sales unit and implementing improvement strategies is crucial for optimal performance. Learning from the success of retailer "095" and incorporating its key success factors into the broader retail strategy is recommended, while providing targeted support for the improvement of retailer "083" is essential for overall retail success.

Furthermore, in the realm of total sales values, a detailed investigation into instances of negative sales is recommended, along with the development and implementation of risk mitigation strategies to handle variability. Addressing operational challenges contributing to high standard deviation is crucial for optimizing efficiency. Regarding top segments by sales, tailored strategies for high-performing segments like "Food Processing" and "Pharma," diversification opportunities, and targeted improvements for low performers are recommended.

In terms of seasonal trends, adapting strategies to accommodate seasonal fluctuations, segment-specific seasonal planning, and data-driven decision-making are emphasized. Finally, addressing data illiteracy through training, implementing retailer-specific data recording policies, and maintaining a strategic focus on data quality are crucial for overcoming challenges related to missing values. Overall, these recommendations provide a comprehensive roadmap for enhancing performance, addressing challenges, and fostering sustained growth in the competitive B2B HVAC&R industry.