“Distributor Engagement Analysis”

A Mid-Term report for the BDM capstone Project

Submitted by

Name: SADEED AZHAR

Roll number: 21F2001197

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

Contents

1. Executive Summary

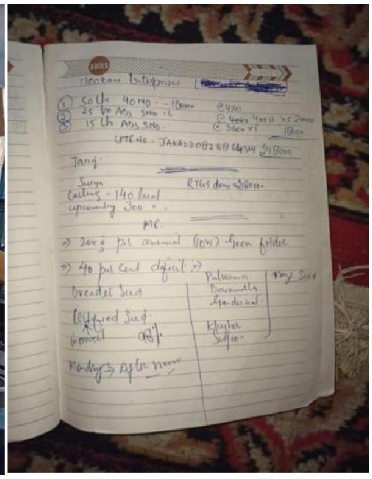
COMPANYNAME, located in New Delhi, specializes in the manufacturing and distribution of electrical appliances, which include LED televisions, fans, geysers, and induction cooktops. The company has successfully expanded its distribution network across multiple states and is now focusing on increasing its presence in Kashmir. Operating primarily under a B2B model, COMPANYNAME sells directly to wholesalers and distributors, while also offering direct sales and services to government agencies and healthcare institutions.

To tackle the challenges of expanding into various districts of Kashmir, I intend to analyze the company's historical sales data to identify promising regions and products by analyzing market trends and demand patterns using tools such as Microsoft Excel and Python (Pandas, Seaborn, and Matplotlib), I will derive insights into market dynamics, distributor performance, and sales outcomes.The metadata section will outline the overall structure and characteristics of the dataset. Descriptive statistics will provide an overview of data distribution, including metrics such as mean, median, standard deviation, skewness, and kurtosis. The analysis will incorporate techniques like data preprocessing, pivoting, and visualization for clear interpretation of data.

The findings are aimed to reveal peak sales patterns, product-specific trends, regional performance, and challenges related to dealer engagement. By leveraging these insights, COMPANYNAME aims to improve its operations, enhance distributor performance, and strengthen its market presence in Kashmir.

2. Proof of originality:

Stock in office



Data

Video interaction: link

Transcription document: link

Letter: link

3. Meta Data

There were multiple xlsx files which were combined into a single excel file,"**Masterfile.xlsx**"
Dataset link: link

The Data sheet consists of the following columns:

Column Name	Description	Units / Format	Values
Date	Transaction date	Date (dd-mm-yyyy)	Jan - June 2023.
Distributor/ Dealer Name	Name of the business.	Text	63 unique names
Area	Geographical region of the dealer grouped into districts.	Text	• Srinagar • Islamabad • Bijbehara • Ganderbal • Baramulla • Budgam
Contact Purpose	Purpose of contact.	Text	• Sale • Distributorship
Contact Mode	Mode of communication.	Text	• Phone • In-person
Remarks	Response received.	Text	• Delivered • Not Interested • Interested • Credit • Bank Guarentee
Product	Product name/model	Text	11 unique products
Price	Net landing price of the product after GST.	Numeric (INR)	422.83 - 22,063.25
Qty	Quantity ordered	Numeric	0 - 5
Amount	Billed amount	Numeric (INR)	0 - 88253
Month	Month extracted from Date	Numeric	1 - 6
Weekday	Day of the week	Numeric	1 - 7 (Monday - Sunday)

4. Descriptive Statistics:

Python library **pandas** was used for performing descriptive analysis.

- The dataset has the shape **(385,12)** and all the columns are non-null.

A) MEAN , MEDIAN & STD. DEV:

- **Numerical Columns:**

Metric	Price	Quantity	Amount
Mean	2,631.92	1.18	6,137.32
Median	0.00	0.00	0.00
Std Dev	5,409.84	1.35	13,246.34
Min	0.00	0.00	0.00
25th %ile	0.00	0.00	0.00
75th %ile	2,669.27	2.00	5,338.54
Max	22,063.25	5.00	88,253.00

As all the medians are 0.00 ,they indicate failed sales interactions, Means for Price ,Qty & Amount are significantly higher than their medians indicating:

- High product price variance across transactions.
- A few high-priced products are skewing the average.
- Large number of inquiries or incomplete sales.
- Small number of units sold per order (1.18), a few larger orders (up to 5 units) exist
- Standard deviation (13,246.34) is very high, reinforcing the fact that most sales are small, while some transactions involve large amounts.

- **Categorical Columns:**

Column Name	Unique Values	Mode	Frequency
Distributor/Dealer	63	Prince Electronics	21
Month	6	January (1)	81
Weekday	7	Thursday (4)	72
Area	6	Srinagar	236
Contact Purpose	2	Sale	306
Contact Mode	2	Phone	330
Remarks	7	Not Interested	194
Product	12	Nill	199

- **Distribuor/Dealer:** There are 63 unique distributors, with "Prince Electronics" appearing most frequently (21 times).
- **Month & Weekday:** Sales are spread across 6 months and all 7 weekdays, with January (81 records) and Thursday (72 records) being the most common.
- **Area:** Sales are highly concentrated in Srinagar (236 records).
- **Contact Purpose:** "Sale" is the primary contact purpose (306 times).
- **Contact Mode:** Most contacts were via Phone (330 times), with fewer in-person interactions.
- **Remarks:** "Not Interested" appears 194 times, suggesting a significant number of rejected sales.
- **Product:** There are 12 unique products, but many transactions have "Nill" (199 times), indicating inquiries or unsuccessful sales.

B) SKEWNESS & KURTOSIS:

Since the dataset consisted of mostly zero value transactions, they were temporarily removed for better visualization of the distribution and the following was observed**:**

Columns	Before Removal	After Removal	Change observed
Price	Skew: 2.779 Kurt: 7.120	Skew: 1.730 Kurt: 1.768	Less skewed, fewer extreme outliers
Qty	Skew: 0.63 Kurt: -0.83	Skew: 0.703 Kurt: 0.769	Slightly more spread out
Amount	Skew: 3.281 Kurt: 11.732	Skew: 2.187 Kurt: 4.581	Still right-skewed but much better

- The dataset contains a mix of normally distributed and highly skewed columns.
- Price and Amount show strong right-skewness, meaning most values are low, but there are extreme high values (outliers).
- Qty is slightly right-skewed, suggesting most transactions involve small quantities.
- Price and Amount have high kurtosis, indicating that they have a few very high values.

5. Detailed explanation of analysis:

Data was primarily collected through multiple .xlsx files, with some key information recorded in a register across three meetings. After collection, data cleaning was performed in MS Excel for each dataset.

The initial data contained various inconsistencies, such as differing data types, spelling errors, and misaligned columns. Following a comprehensive cleaning process, the datasets were merged using the pandas library in Python, resulting in a file named "MASTERFILE.xlsx." The merging was validated in MS Excel to ensure all datasets had the same columns in the correct order. Visual Studio Code was used as the integrated development environment (IDE) for this process, and some results were documented in MS Word. Descriptive statistics were generated mainly using the pandas library.To improve the visualization of distributions, numerous '0' or 'Nill' values in the 'Qty', 'Price', and 'Amount' columns were temporarily excluded, as the overall distribution was notably right-skewed. The matplotlib and seaborn libraries were employed for visualizing distributions, and pivot tables were created in MS Excel to summarize various metrics during the analysis.

- **Data Analysis Process:**

The data analysis process was carefully structured to ensure accuracy, consistency, and a logical approach to extracting useful insights whilst minimizing errors, maintain data integrity, and present findings in a way that is easy to understand. Below is a breakdown of the process :

1) Collecting and Cleaning Data in Excel:

The first step in the analysis process is collecting all relevant datasets and ensuring they are properly formatted. It is crucial to check that all datasets have the same column names and structure, as mismatched or missing values can create inconsistencies that may affect the analysis. Since the dataset is relatively small, Microsoft Excel is the preferred tool for this step. It allows for easy manual inspection, quick corrections, and straightforward data cleaning. Excel's features make it convenient to spot and fix missing or incorrect values before merging the data programmatically. Additionally two more columns, **Month & Weekday** were calculated from the **Date** column using the following formulas:

- **Month : =MONTH(**Date**)**
- **Weekday: = WEEKDAY(**Date**)**

2) Merging Datasets Using Pandas:

Once the data is cleaned and structured, it needs to be combined efficiently. Using the Pandas library in Python allows for smooth merging of multiple datasets while maintaining

consistency. Pandas offers built-in functions that help handle missing values, and resolve inconsistencies.

3) Descriptive Statistics & Initial Visualizations (Pandas, Matplotlib, Seaborn):

To gain an initial understanding of the dataset, descriptive statistical methods are applied. This includes calculating key metrics such as the mean, median, standard deviation, skewness, Kurtosis and overall distributions. These statistics provide a summary of the data and highlight important characteristics.Following were used to do so:

- **df.info() , df.describe() , df.describe(include = 'object')**

In addition to numerical analysis, data visualization plays a crucial role in identifying trends, patterns, and potential outliers. Using visualization libraries like Matplotlib and Seaborn, graphical representations such as histograms, scatter plots, and box plots make it easier to detect patterns that may not be obvious when looking at raw numbers.

4) Performing Detailed Calculations & Tabulations in Excel:

Certain calculations, such as total revenue per product or sales by region, are best performed in Excel due to its tabular format and ease of use. Excel allows for clear presentation, manual adjustments, and cross-verification of results, making it a practical tool for business-oriented data summaries.Majority of the calculations were performed by creating **pivot tables** on the dataset. Some of the functions include:**SUMIF(),COUNTIF(),AVERAGE(),MAX(),MIN()**

5) Creating Final Visualizations in Excel

To effectively present key insights and trends, visual representations such as charts and graphs are created in Excel. Excel's charting tools offer an intuitive way to display information in a clear and interactive format, making it easier for stakeholders to interpret findings and make informed decisions.

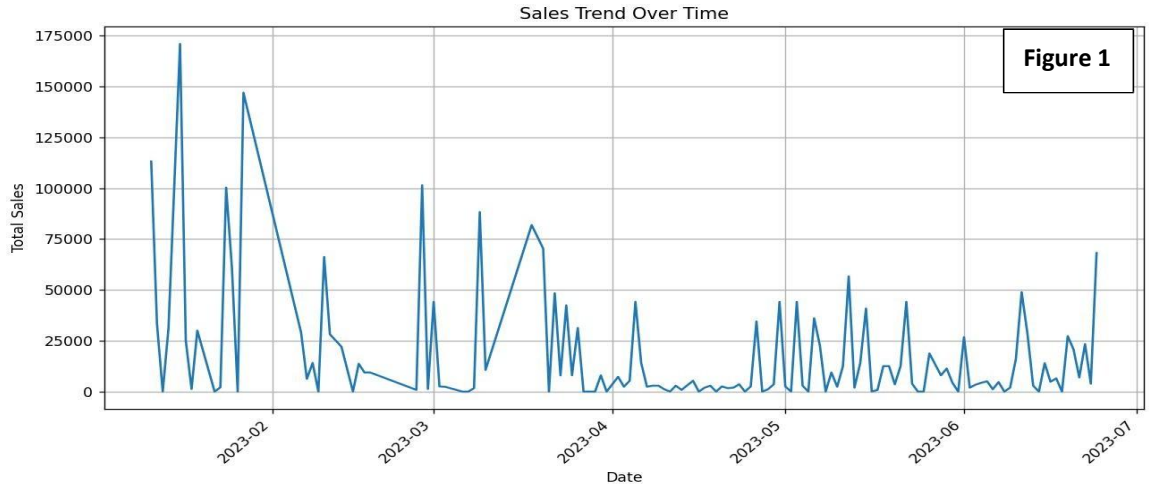
By combining the capabilities of both Python and Excel, this structured approach ensures high-quality data analysis. Excel is used for manual data cleaning, tabulations, and final presentations, while Python provides powerful tools for automation, merging, and in-depth statistical analysis. This balanced workflow enhances accuracy, efficiency, and the overall impact of the analysis.

6. Results & findings

1. Sales Trends:

Monthly Growth Rates: (refer to Figure 1)

February: Decreased by 57.8% , March: Increased by 48.4% ,April: Decreased by 57.1% ,May: Increased by 96.8% , June: Decreased by 15.1% ,



Peak Sales: January recorded the highest sales (**714,937.65)** , followed by a gradual decline.

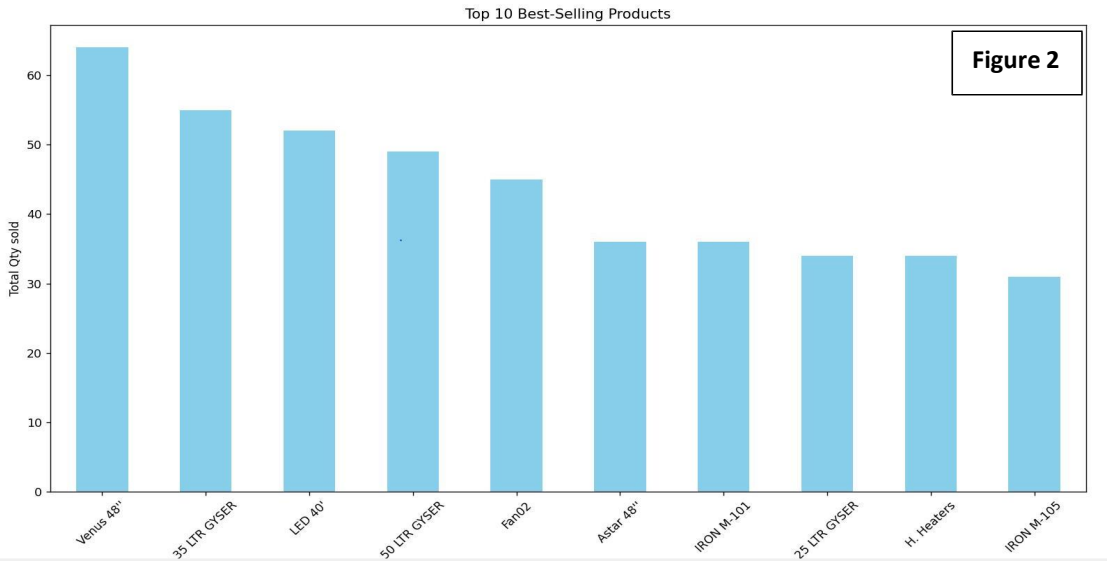
Lowest Sales: April had the least sales **(192,259.82)**

January was the best month , possibly due to New Year sales promotions or bulk orders. Fluctuations (March up, April down, May up again) indicate inconsistency in demand and sales efforts.,June saw a decline (-15.1%), possibly due to seasonal trends or stock shortages.

2. Product Trends : (refer to Figure 2)

Top-Selling Products :

- **Venus 48":** 64 units **(14.15%)**
- **35 LTR Geyser:** 55 units **(12.16%)**
- **LED 40'':** 52 units **(11.5%)**



Low-Selling Products:

- **IRON M-105: 31 units (6.8%)**

- **Geyser 25 LTR: 16 units (3.53%)**

The product trends also vary based on seasonality as sales declined towards the end of winter season.

2. Area-Wise Sales Performance: (refer to Figure 3)

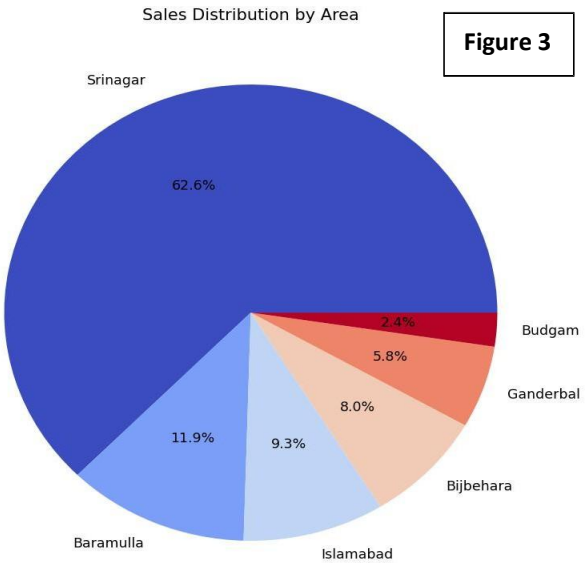


Figure 3

- **Highest Sales Region: Srinagar** (1,588,652 - **62.6%**)
- **Lowest Sales Region: Budgam** (30,410 - 2**.4%**)

Srinagar has dominated the sales as majority of the current business dealers are located in this region but areas such as Budgam and Ganderbal haven't been fully explored yet.

3. **Sales Issues & Challenges:** (refer to Figure 4)

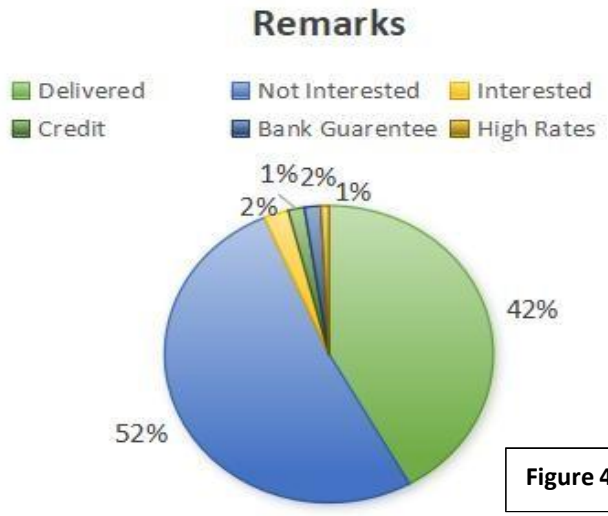


Figure 4

- **"Not Interested" (52%)** is the major reason for lost sales.
- **Credit & Bank Guarantee Issues (12 cases total)** - Financing barriers for buyers.
- **High Prices (3 cases)** : Pricing doesn't seem to be a significant problem for sales.