

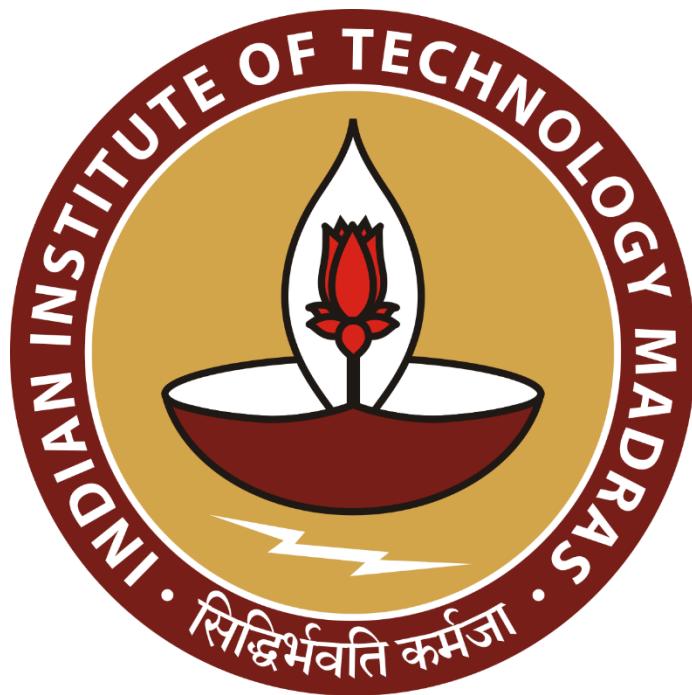
Sales Trends and Customer Behavior in a Medical Shop: An Analytical Study

A Mid-term Submission for the BDM capstone Project

Submitted by

Name: SOUMYADEEP DHALI

Roll number: 22f1001762



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

Contents:

1	Executive Summary and Title	2
2	Proof of originality of the Data	2
3	Metadata and Descriptive statistics	4
4	Detailed explanation of analysis process/methods	6
5	Results and findings	8

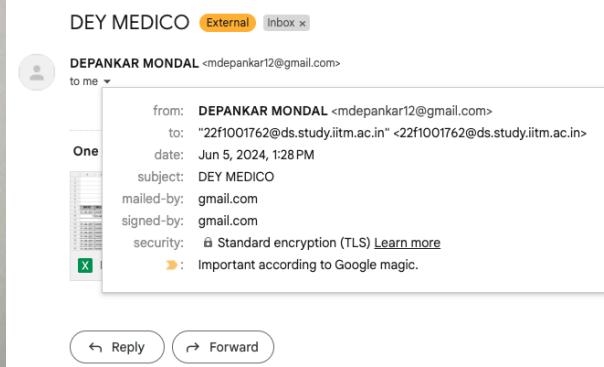
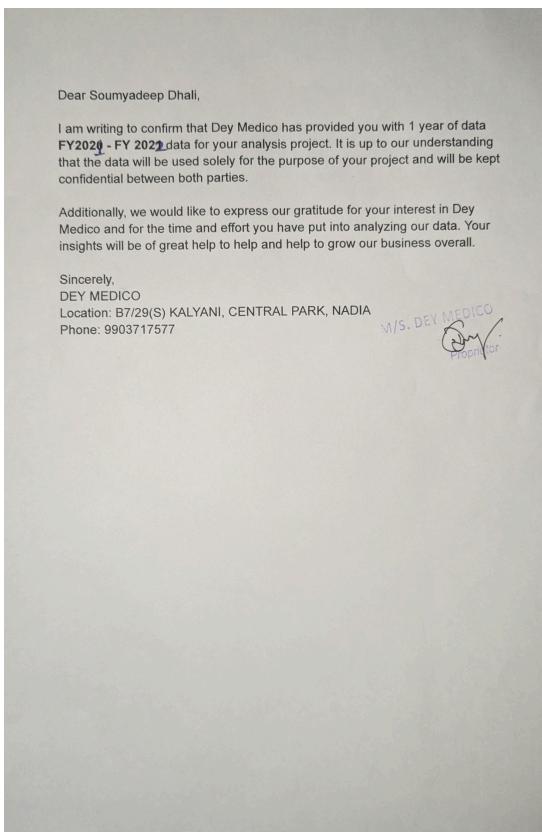
1. Executive Summary

DEY MEDICO is a leading medical shop specializing in the sale of a wide range of medicines, committed to providing high-quality healthcare products and exceptional customer service to the community.

The analysis of the medical shop's sales data aimed to identify trends, optimize inventory, and improve sales strategies. The data collected over a year included daily sales amounts, which were categorized by month for analysis. The findings revealed that July had the highest sales, likely due to a spike in COVID-19 cases, while August had the lowest sales attributed to fewer days the shop was open. Monthly sales showed variability, with total sales peaking in the summer season and dipping in winter. The analysis highlighted significant growth rates between certain months, providing insights into sales trends and periods of growth or decline. Seasonal trends indicated that spring was the most profitable season, followed by summer and winter. The findings underscore the importance of understanding customer purchasing behavior and sales patterns to inform decisions on inventory management and promotional strategies. The structured approach of data collection, cleaning, exploration, modeling, and interpretation facilitated valuable insights, ultimately aiding in data-driven decision-making to enhance business performance.

2. Proof of Originality





[Video link\(proof of originality\)](#)

3. Metadata and Descriptive Statistics

METADATA:

The owner shared a compiled sales report excel sheet **FY 2021-22** containing details of the date the items were sold, the bill number, party name, the taxable amount, the tax on the item, the total amount of the product.

Key	Description
DATE	The date at which the item was sold
BILL NO.	The unique bill number of the items sold
PARTY NAME	The name to whom the products were sold
BILL AMT.	The total bill amount the bill
TAXABLE	The taxable amount to which the tax will be applied
TAX	The total tax amount

DESCRIPTIVE STATISTICS:

After data preprocessing and cleaning, the following is a concise overview of the dataset using descriptive statistics. I have attempted to condense the information by highlighting the descriptive statistics below of the sales data **FY 2020-21**.

DEY MEDICO								
B-7/29(S) KALYANI, CENTRAL PARK NADIA								
Phone : 9903717577								
GSTIN : 19ADJPD7353D1ZM								
SALES BOOK								
01-04-2021 - 31-03-2022								
DATE	BILL NO.	PARTY NAME	BILL AMT.	TAXABLE	TAX	TAX FREE	EXEMPTED	R.OFF
01-04-2021	A000001	RO SANCHITA DAS	8297.00	7067.13	1230.22	0.00	0.00	-0.35
1	03-04-2021		8297.00	7067.13	1230.22	0.00	0.00	-0.35
01-04-2021	A000002	PW SANGITA DAS	4590.00	4097.85	491.76	0.00	0.00	0.39
01-04-2021	A000003	IP PRATAP DAS	5262.00	4616.14	645.86	0.00	0.00	0.00
01-04-2021	A000004	PR SEKHAR DAS	3083.00	2752.45	330.30	0.00	0.00	0.25
01-04-2021	A000005	KQ SUNTI RANI DAS	10024.00	8950.17	1074.04	0.00	0.00	-0.21
01-04-2021	A000018	FX SANJIB DAS	3993.00	3684.59	308.82	0.00	0.00	-0.41
01-04-2021	A000019	ER SAMIR MONDAL	9136.00	8148.42	987.58	0.00	0.00	0.00
01-04-2021	A000020	NG PRADIP DAS	4937.00	4327.38	609.56	0.00	0.00	0.06
7	01-04-2021		41025.00	36577.00	4447.92	0.00	0.00	0.08

Fig 1 : Sales data FY 2020-21

Mean	The average value.
Median	The middle value when arranged in ascending order.
Mode	The largest value.
Standard Deviation	The measure of spread of values from mean in a single sample.
Sample Variance	The average of the squared differences from the mean.
Kurtosis	The measure of how much the data is dispersed between a distribution's center and tails.
Skewness	The degree of asymmetry observed in a probability distribution.
Range	The difference between highest and lowest observation in a data.
Minimum	The minimum value in the dataset.
Sum	The total sale for the year.
Count	The total count of entries.

Largest	The largest value in the dataset.
Smallest	The smallest value in the dataset.
Confidence Level(95%)	The confidence level indicates the probability with which the estimation of the location of a statistical parameter (e.g., an arithmetic mean) in a sample survey is also true for the population/

For overall sales data **FY 2021-22**

Mean	16673.17505
Median	15661
Mode	19844
Standard Deviation	11046.2112
Sample Variance	122018781.9
Kurtosis	3.87011596
Skewness	1.580373649
Range	75080
Minimum	649
Sum	75729
Count	954
Confidence Level(95%)	701.8416766

4. Detailed explanation of analysis process/methods

Data Analysis Process for Medical Shop Sales Data

a. Define the Objective

- **Identify the problem or question:** I aim to analyze the sales data collected from the medical shop to identify trends, optimize inventory, and improve sales strategies.

- **Set goals:** Specific outcomes include identifying peak sales periods, understanding customer purchasing behavior, and forecasting future sales.

b. Collect Data

- **Data sources:** I gathered sales transaction data from the medical shop's records.
- **Data acquisition:** I ensured the data included relevant fields such as date, sales amount, product details, and customer information.

c. Clean the Data

- **Data quality:** I checked for and addressed missing values, duplicate entries, and errors in the dataset.
- **Data cleaning:** I corrected inaccuracies, filled in missing values, and removed duplicates to ensure the data was accurate and reliable. There were some entries which were of no use to me so I cleaned them out.

d. Explore and Understand the Data

- **Descriptive statistics:** I calculated basic statistics such as, mean, median, mode, maximum, minimum, kurtosis etc.
- **Data visualization:** I created charts and graphs to visualize data distributions and relationships, such as sales trends over time.
- **Identify patterns:** I looked for trends, correlations, and outliers in the data to gain a better understanding of sales behavior.

e. Transform the Data

- **Data normalization:** I scaled the data if necessary to ensure consistency across different metrics.
- **Feature engineering:** I created new features, such as sales per customer and average sales per product, to enhance the analysis.
- **Data aggregation:** I summarized data by grouping it by month, day of the week, or product category to facilitate deeper analysis.

f. Model the Data

- **Select models:** I chose appropriate models for time series analysis to forecast future sales and clustering algorithms to segment customers.
- **Train models:** I used a subset of the data to train the models.
- **Evaluate models:** I assessed model performance using metrics such as accuracy and mean absolute error to ensure reliability.

g. Interpret Results

- **Analyze findings:** I interpreted the results from models and visualizations to draw meaningful conclusions.
- **Draw conclusions:** I made inferences based on the analysis, such as identifying peak sales periods and key products driving sales.
- **Validate results:** I checked the reliability and validity of the conclusions by cross-referencing with known patterns and business context.

h. Communicate Insights

- **Report findings:** I prepared a report summarizing key insights and recommendations, supported by visualizations.
- **Data visualization:** I used charts and graphs to illustrate findings clearly and effectively.
- **Stakeholder communication:** I presented the results to stakeholders, ensuring they understood the implications and potential actions.

i. Make Data-Driven Decisions

- **Implement solutions:** I used the analysis results to inform decision-making, such as optimizing inventory levels and planning promotions.
- **Monitor outcomes:** I tracked the impact of decisions and actions based on the analysis to ensure they achieved the desired results.

j. Review and Iterate

- **Evaluate process:** I assessed the overall data analysis process to identify areas for improvement.
- **Iterate:** I refined and repeated the analysis as necessary, incorporating new data or methods to continually improve insights and decision-making.

By following this structured approach, I systematically analyzed the sales data from the medical shop, extracted valuable insights, and made informed decisions to enhance business performance.

5. Results and Findings

Monthly Sales Analysis for a Medical Shop:

Data Collection and Preparation

- **Data Source:** Sales data for a medical shop collected over one year.
- **Data Structure:** Each entry includes a date and the total sales amount for that day.
- **Preprocessing:** Extracted the month from each date to categorize sales data by month.

Methodology

- Data Extraction:** Extracted the month from each sales date to facilitate monthly analysis.
- Sales Aggregation:** Summarized total sales for each month to observe monthly trends.

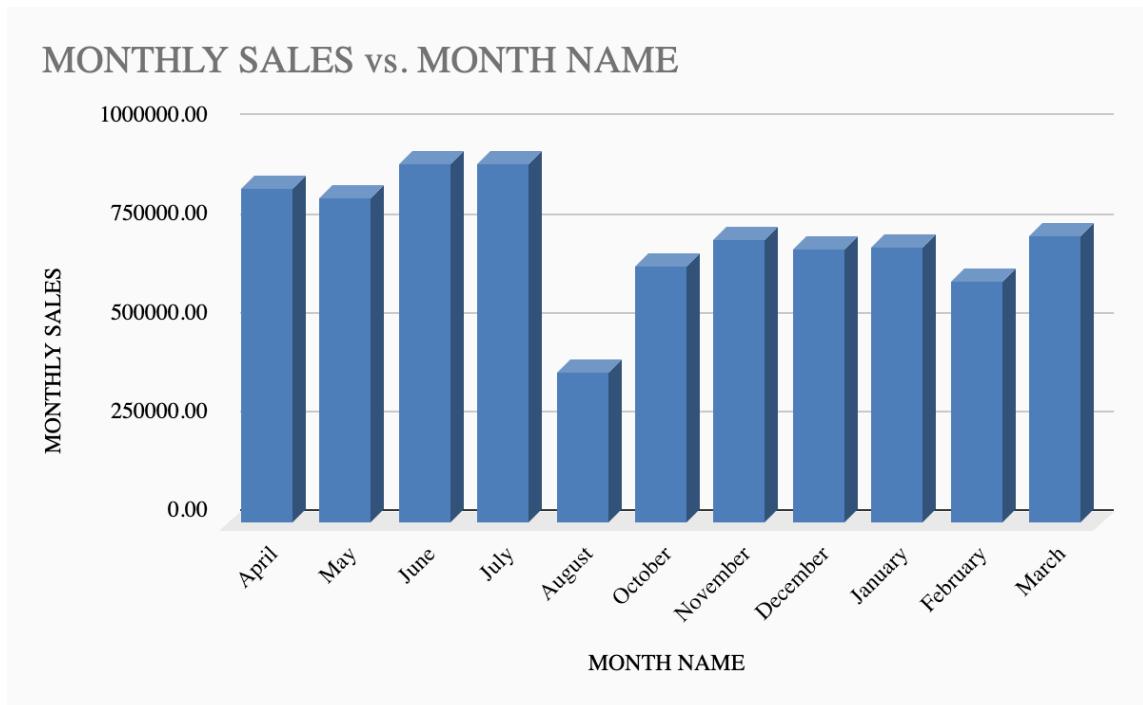


Fig 2: Monthly sales vs. Monthly name

Findings:

Monthly Sales Distribution:

- **April:** Observed sales total of Rs. 847074
- **May:** Observed sales total of Rs. 822663
- **June:** Observed sales total of Rs. 906525
- **July:** Observed sales total of Rs. 906731
- **August:** Observed sales total of Rs. 379203
- **October:** Observed sales total of Rs. 651106
- **November:** Observed sales total of Rs. 714516
- **December:** Observed sales total of Rs. 694398
- **January:** Observed sales total of Rs. 696521
- **February:** Observed sales total of Rs. 611877
- **March:** Observed sales total of Rs. 728248

Peak Sales Months:

- **Highest Sales:** The month of **July** showed the highest sales with a total of Rs.906731. This could be due to the sudden spike of Coronavirus cases in our neighborhood.
- **Lowest Sales:** The month of August showed the lowest sales with a total of Rs.379203 simply because of the fewer days the shop being open.

Growth Rate Analysis:

- **Monthly Growth Rate:** Calculated the growth rate from month to month to understand the sales trends.
 - For example, the growth rate from January to February was calculated as:

$$\text{Growth Rate} = \left(\frac{\text{Sales in February} - \text{Sales in January}}{\text{Sales in January}} \right) \times 100$$

- **Observation:** Identified periods of significant sales growth or decline.

MONTHLY GROWTH RATE vs. MONTHS

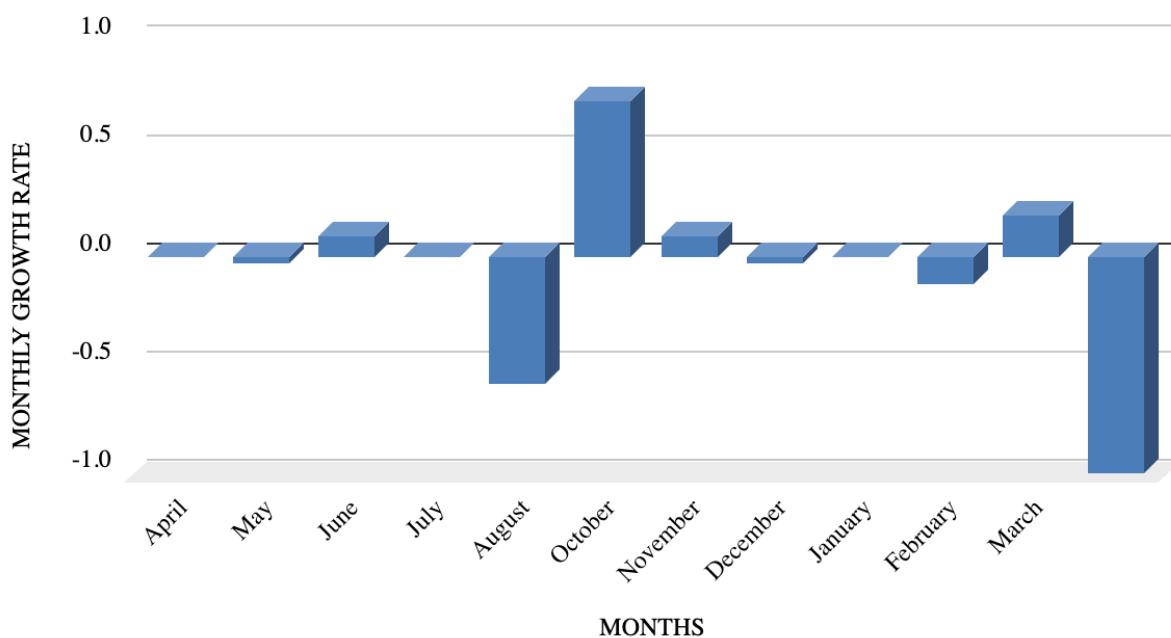


Fig 3: Monthly sales vs. Monthly name

The above chart shows the monthly growth rate during the year. The (+) y-axis shows a positive growth rate and the (-) y-axis shows a negative growth rate.

Seasonal Trends:

- **Winter (Nov, Dec, Jan, Feb):** Total sales were Rs. 2397985.
- **Spring (Mar, Apr, May):** Total sales were Rs. 2843565.
- **Summer (Jun, Jul, Aug, Oct):** Total sales were Rs. 2717312.

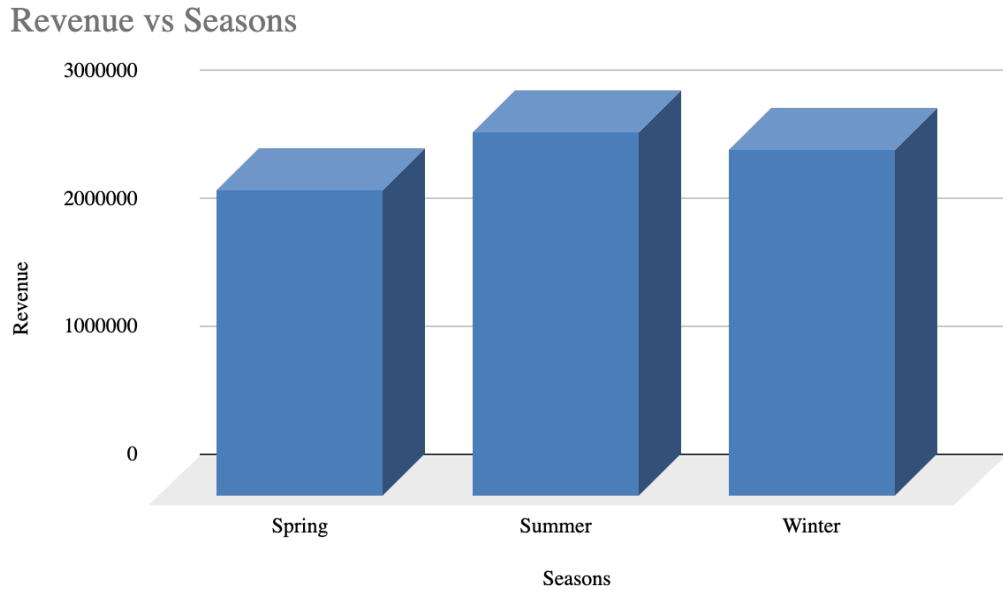


Fig 4: Revenue vs. Seasons

It is clearly evident that their most profitable season is the summer, there is a slight decrease in revenue the winter season.