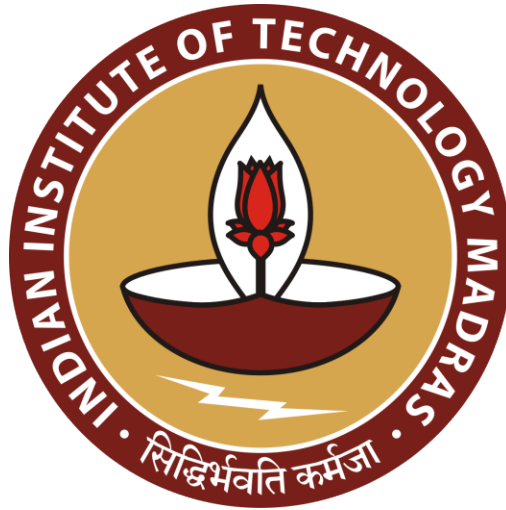


Enhancing Business Dynamics through Data-Driven Solutions at King Pipes



Submitted by:

Suyamoon Pathak

(21f3002588)

Submitted to:

BS Data Science and Applications,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

10th August 2023

1. Executive Summary:

This report presents a strategic initiative to enhance business data management within Samrat Plastic Industries, operating under the brand King Pipes. By addressing key challenges in inventory management, demand prediction, and targeted marketing, this project aims to drive operational efficiency and sustainable growth.

Samrat Plastic Industries, a distinguished player in the plastic manufacturing sector, has built a reputation for high-quality UPVC, CPVC, and SWR pipes and fittings. Despite its accomplishments, the company faces persistent obstacles in inventory control, hindered by stockouts and overstocking. Accurate demand prediction remains elusive, affecting production planning and resource allocation. Furthermore, optimizing targeted marketing efforts necessitates a deeper understanding of customer segments.

To overcome these challenges, this project will implement a rigorous data-driven approach. Time series analysis will empower accurate demand forecasts, ensuring optimal inventory levels. Cluster analysis will identify distinct customer segments, enabling personalized marketing strategies. By employing Python's libraries like pandas, scikit-learn, and Excel's visualization capabilities, we will extract actionable insights from sales, production, and inventory data.

The anticipated outcome is a transformation in Samrat Plastic Industries' operations. Streamlined inventory management will mitigate production disruptions, fostering efficient resource utilization. Informed decision-making, guided by data-derived intelligence, will pave the way for sustained growth and market leadership. Through this project, we aim to empower Samrat Plastic Industries to achieve its vision of becoming a 2000 crore turnover company by 2030, solidifying its position as an industry frontrunner.

2. Proof of Originality of the Data

2.1. Recorded video

Here's the video link for my conversation with Mr. Chiragbhai Joshi, the Sales and Marketing Manager at Samrat Plastic Industries.

Interview with Mr. Chiragbhai Joshi: ([Link](#))

Factory Visit Video Link: ([Link](#))



Figure 1 Suyamoon Pathak (me, on the left) interviewing Mr. Chiragbhai Joshi (on the right)

2.2 Letter of Originality from the Organization

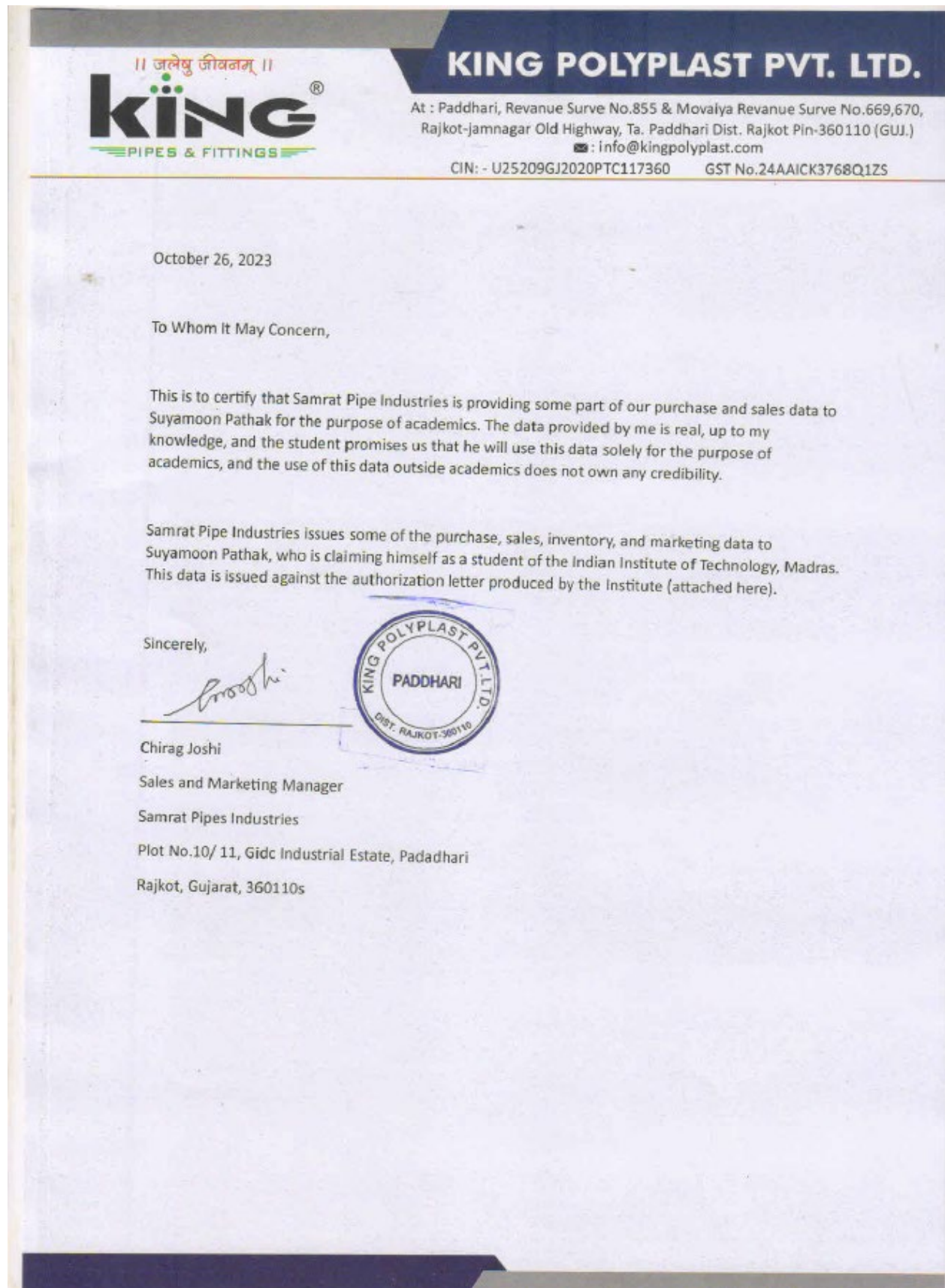


Figure 2 Letter of Originality

2.2 Images Related to the Organization



Figure 4 Samrat Plastic Industries Entrance Gate



Figure 3 Samrat Plastic Industries Inventory


3. Metadata

In the pursuit of understanding and analyzing the operational dynamics of Samrat Plastic Industries, the collected data serves as a critical foundation. The metadata accompanying this data provides essential context and attributes that shed light on the origin, scope, and relevance of the information under investigation. This section outlines the key aspects of the metadata associated with the data obtained from the organization, enabling a comprehensive view of the data's significance.

3.1 Data Source and Origin

The data used for this analysis originates directly from Samrat Plastic Industries (kingpipe.in), a well-established player in the pipe manufacturing industry. This organization, headquartered in Rajkot, Gujarat ([Google Maps link](#)), specializes in the production and distribution of various pipes and fittings, catering to a diverse range of clients and industries. The data provided is divided into two different datasets: one that shows production metrics for their major products (UPVC, CPVC, and SWR pipes), and another that shows the costs and results of their marketing campaigns. These datasets span from June 1, 2023, to September 30, 2023.

3.2 Data Collection Process

The data collection process involved face-to-face interviews with the organization's managers, Mr. Jaydeep Patel and Mr. Chiragbhai Joshi. These discussions encompassed topics such as inventory management, customer segmentation, marketing strategies, production processes, and challenges faced by the industry. Access to the information relevant to the previous four months was then made possible by a visit to their cyber office which can be found in the link [here](#). 

3.3 Data Attributes and Variables

The important features in the production dataset are date, product identity, quantity, supplier information, price, lead time, batch serial numbers, manufacturing details, quality control tests, and operational information on workers and equipment.

Important variables included in the marketing dataset include the weekly spending on different advertising platforms (Facebook, Instagram, Google Ads, IndiaMart ads), event expenses (Meet and Greet, Industry Participation), engagement metrics, click-through rates, attendance at events, builder merchandise distribution, leads generated, and sales conversion rates.

3.3.1. Units of measurements

The units of measurement in the datasets offered are essential for comprehending and evaluating the quantitative features of the data. The standard units listed below have been used for a variety of attributes:

- Prices: All prices are expressed in Indian Rupees (INR).
- Weights: Samrat Plastic Industries consistently records weights in grams.
- Dimensions: Pipe dimensions are given in the format length (meter) * diameter (inch), following a standard procedure.
- Thickness: Millimeters (mm) are used to represent thickness measurements.

3.3.2. Encodings

The exact names of the suppliers have been encoded and replaced with categorical variables (e.g., Supplier_1, Supplier_2, etc.) in order to protect supplier confidentiality and adhere to privacy rules. To protect sensitive information, a whole column of supplier phone numbers has also been removed from the dataset.

3.4 Data Limitations

Among the data's limitations are the lack of some precise information, such as consumer demographics, and the marketing data's finer segmentation. Furthermore, the retroactive nature of the data may mean that it lacks real-time insights.

3.5 Anticipated Data Refinement

In order to get a more nuanced and complete analysis, future refinement tactics may involve augmenting the datasets with other factors such as consumer demographics, regional sales breakdowns, increased marketing segmentation, and more comprehensive production metrics.

4. Descriptive Statistics

4.1. The Production Data

4.1.1. Quantitative Attributes

Quantity Statistics by Product Type are given below:

	Mean	Median	Mode	Range	Variance	StdDev
Product_ID						
CPVC_001	1388.043478	1400.0	500	2950	7.500205e+05	866.037258
CPVC_002	1550.000000	1500.0	1500	2950	7.046667e+05	839.444261
CPVC_003	1546.428571	1450.0	100	2800	8.582555e+05	926.420798
SWR_001	4144.230769	4375.0	1100	5900	3.507123e+06	1872.731408
SWR_002	4252.777778	4225.0	1150	5650	2.505278e+06	1582.806930
SWR_003	3633.333333	3550.0	1050	5900	2.747917e+06	1657.684128
UPVC_001	4677.685185	4660.0	4835	595	2.672520e+04	163.478442
UPVC_002	4679.512195	4685.0	4395	585	3.469226e+04	186.258573
UPVC_003	1441.666667	1500.0	1500	1700	2.517424e+05	501.739399
UPVC_004	657.692308	700.0	500	500	2.826923e+04	168.134561

Similarly, Price per unit statistics by Product Type are given below:

	Mean	Median	Mode	Range	Variance	StdDev
Product_ID						
CPVC_001	77.0	77.0	77	0	0.0	0.0
CPVC_002	127.0	127.0	127	0	0.0	0.0
CPVC_003	83.0	83.0	83	0	0.0	0.0
SWR_001	49.0	49.0	49	0	0.0	0.0
SWR_002	51.0	51.0	51	0	0.0	0.0
SWR_003	44.0	44.0	44	0	0.0	0.0
UPVC_001	152.0	152.0	152	0	0.0	0.0
UPVC_002	167.0	167.0	167	0	0.0	0.0
UPVC_003	129.0	129.0	129	0	0.0	0.0
UPVC_004	179.0	179.0	179	0	0.0	0.0

4.1.2. Categorical Attributes

The Supplier frequency distribution according to the dataset is:

```
Supplier_90    21
Supplier_39    21
Supplier_44    19
Supplier_17    14
Supplier_28    13
..
Supplier_71     1
Supplier_72     1
Supplier_73     1
Supplier_74     1
Supplier_57     1
Name: Supplier, Length: 112, dtype: int64
```

Similarly, the geographical distribution of suppliers is given below:

```
Hisar          21
Hubli          21
Patiala        19
Nagpur         14
Indore         13
..
Silvassa       1
Panipat        1
Puducherry     1
Itarsi         1
Amaravati      1
Name: Supplier_Location, Length: 112, dtype: int64
```

The Material Composition frequency is given below:

```
CPVC    121
SWR     121
UPVC    120
Name: Material_Composition, dtype: int64
```

4.1.3. Temporal Analysis

Frequency Distribution of Units Manufactured on Different Dates:

```

19-06-2023    27
13-07-2023    26
02-06-2023    24
02-07-2023    24
17-07-2023    22
13-06-2023    21
24-06-2023    18
21-07-2023    18
21-06-2023    18
11-07-2023    15
07-07-2023    15
28-06-2023    15
24-07-2023    15
07-06-2023    15
29-05-2023    15
27-05-2023    15
15-06-2023    14
18-05-2023    12
11-06-2023    12
25-05-2023    12
15-05-2023     9
Name: Manufacturing_Date, dtype: int64

```

Lead Time Statistics for Different Suppliers:

```

              mean  median  <lambda_0>
Supplier
Supplier_1      4.916667    5.0         5
Supplier_10     5.250000    5.0         5
Supplier_100    4.000000    4.0         4
Supplier_101    6.000000    6.0         6
Supplier_102    5.500000    5.5         5
...             ...      ...         ...
Supplier_95     6.000000    6.0         6
Supplier_96     5.000000    5.0         5
Supplier_97     4.000000    4.0         4
Supplier_98     5.500000    5.5         5
Supplier_99     6.000000    6.0         6

[112 rows x 3 columns]

```

4.1.4. Other Statistics

Summary Statistics for Reorder Levels:

```

count      362.000000
mean       2156.353591
std        978.735669
min         800.000000
25%        1500.000000
50%        2000.000000
75%        3000.000000
max         4000.000000
Name: Reorder_Level, dtype: float64

```

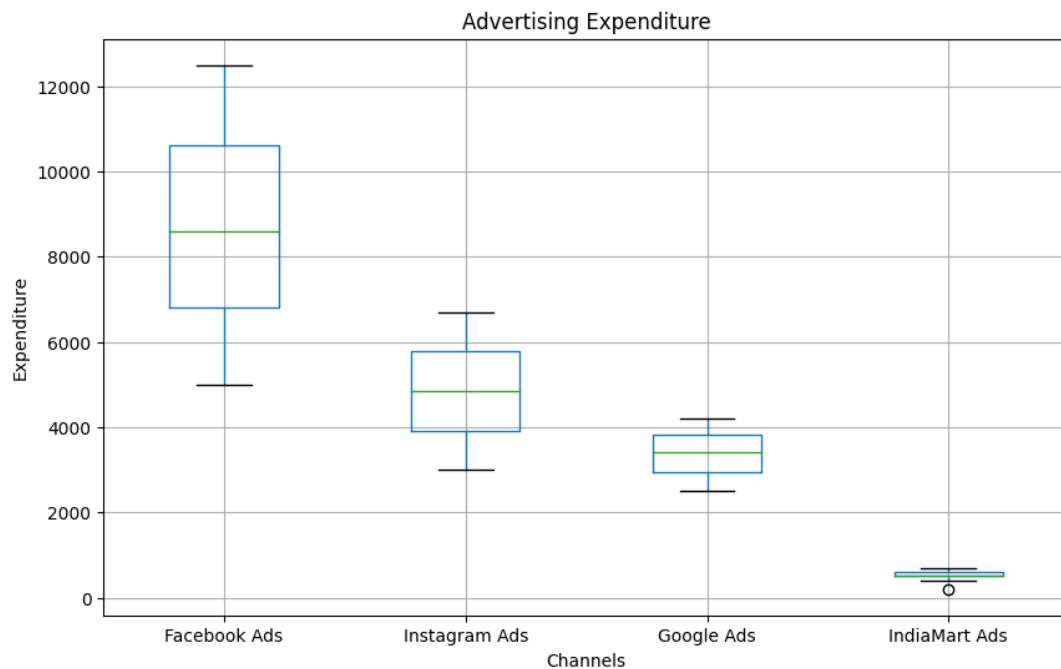
Physical Attributes Across Different Product Types:

Material_Composition	Weight		Area_square_meters	Area_square_meters		Thickness_mm	Thickness_mm		
	mean	median <lambda_0>		mean	median <lambda_0>		mean	median <lambda_0>	
CPVC	2200.000000	2200.0	0	0.053109	0.06350	0.025400	4.5	4.5	0.0
SWR	662.727273	670.0	220	0.053450	0.06350	0.028575	2.0	2.0	0.0
UPVC	3055.416667	2500.0	2200	0.058314	0.05715	0.088900	3.5	3.5	0.0

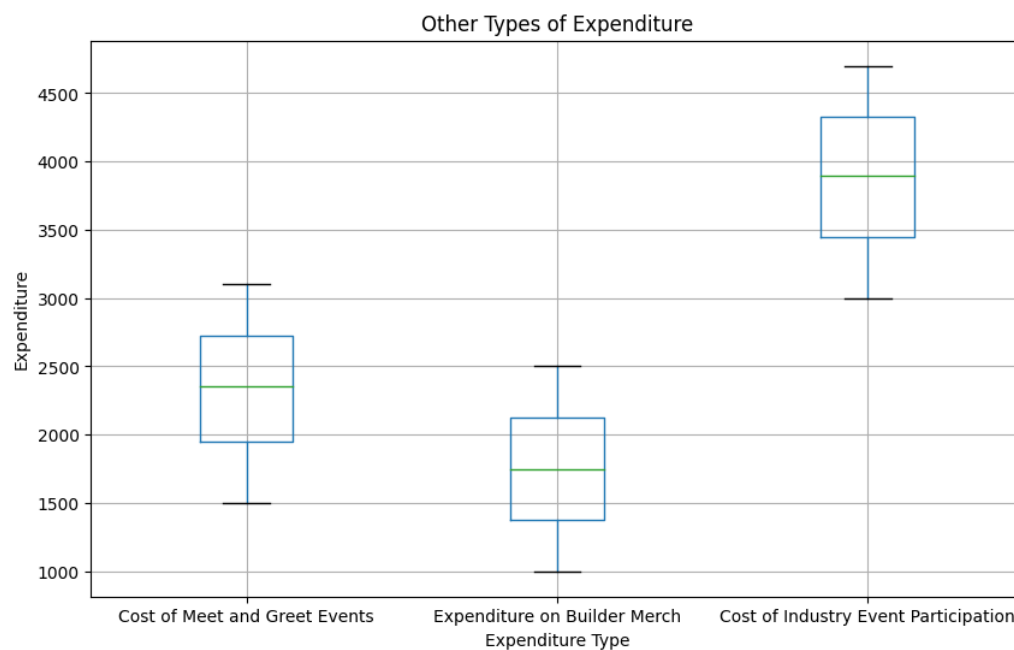
4.2 Marketing Data

4.2.1. Advertising Expenditure:

Advertising on the Internet:

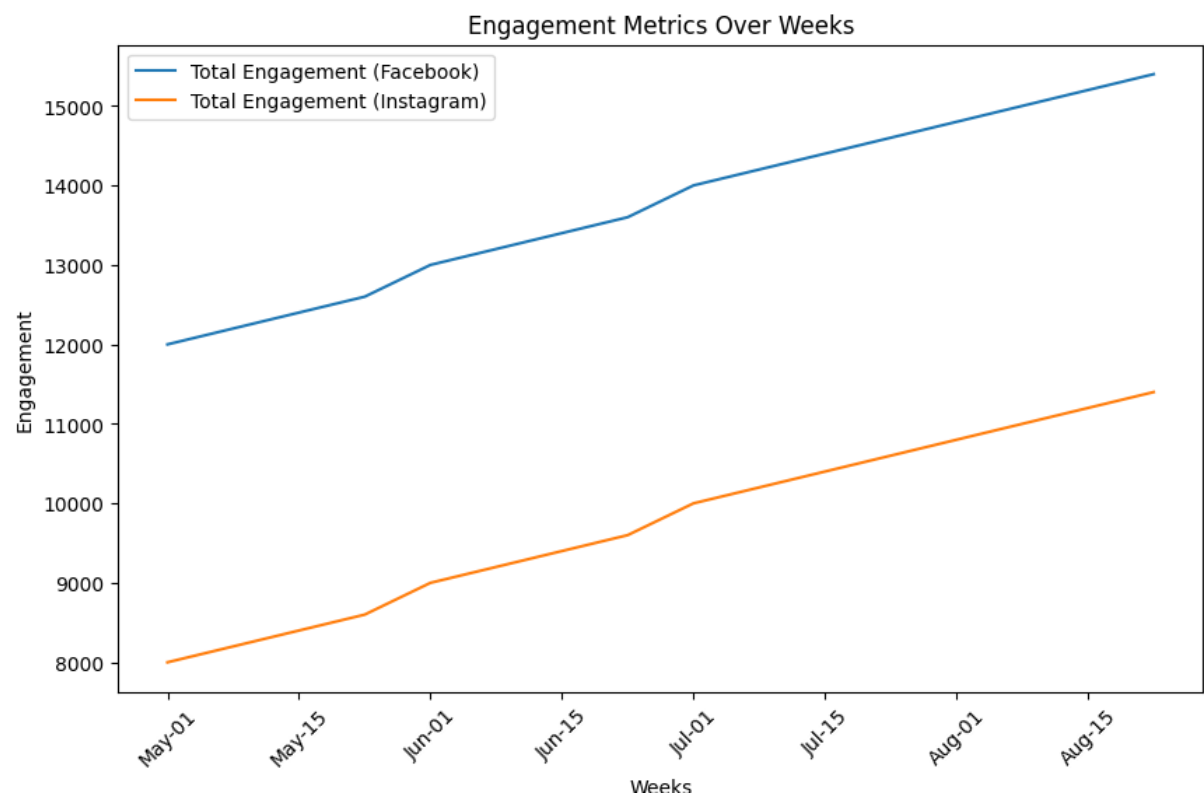


Advertising on Events:

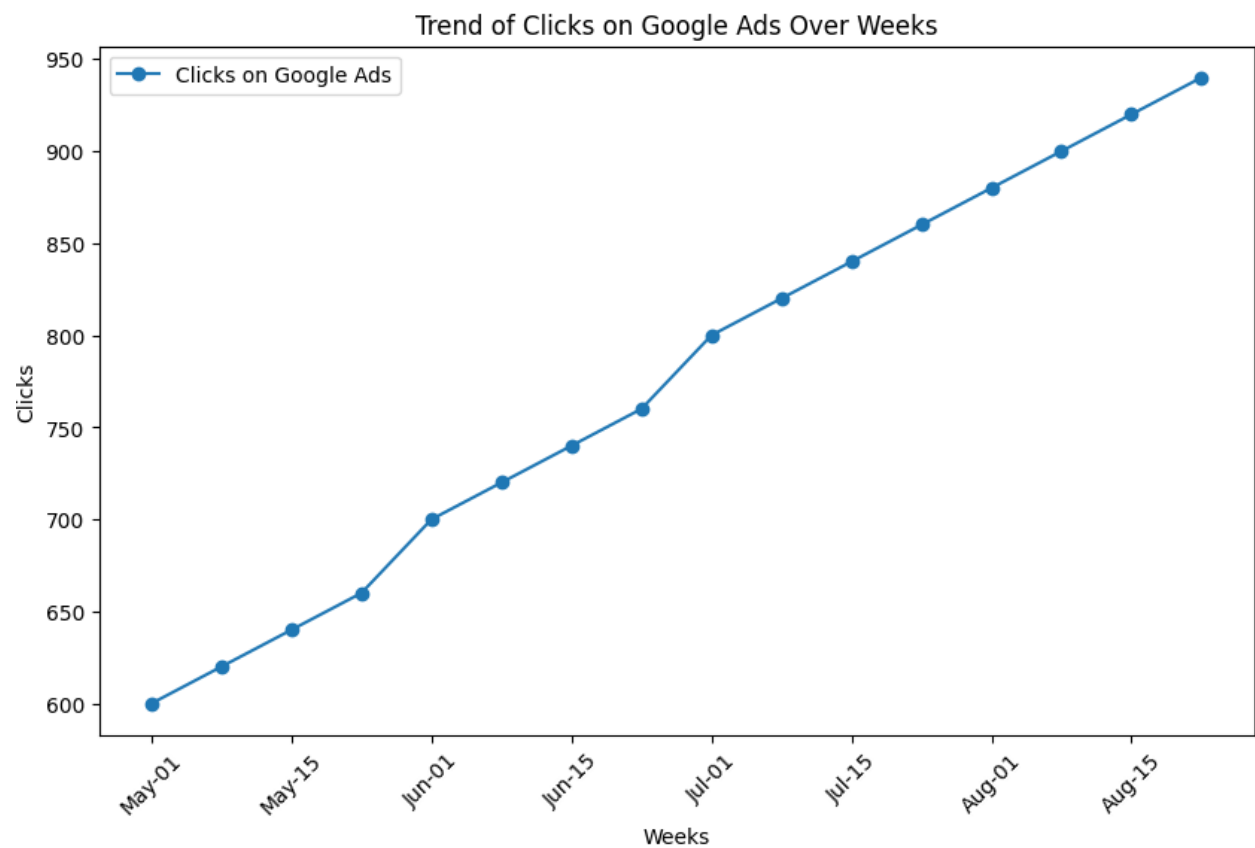


4.2.2. Engagement Metrics:

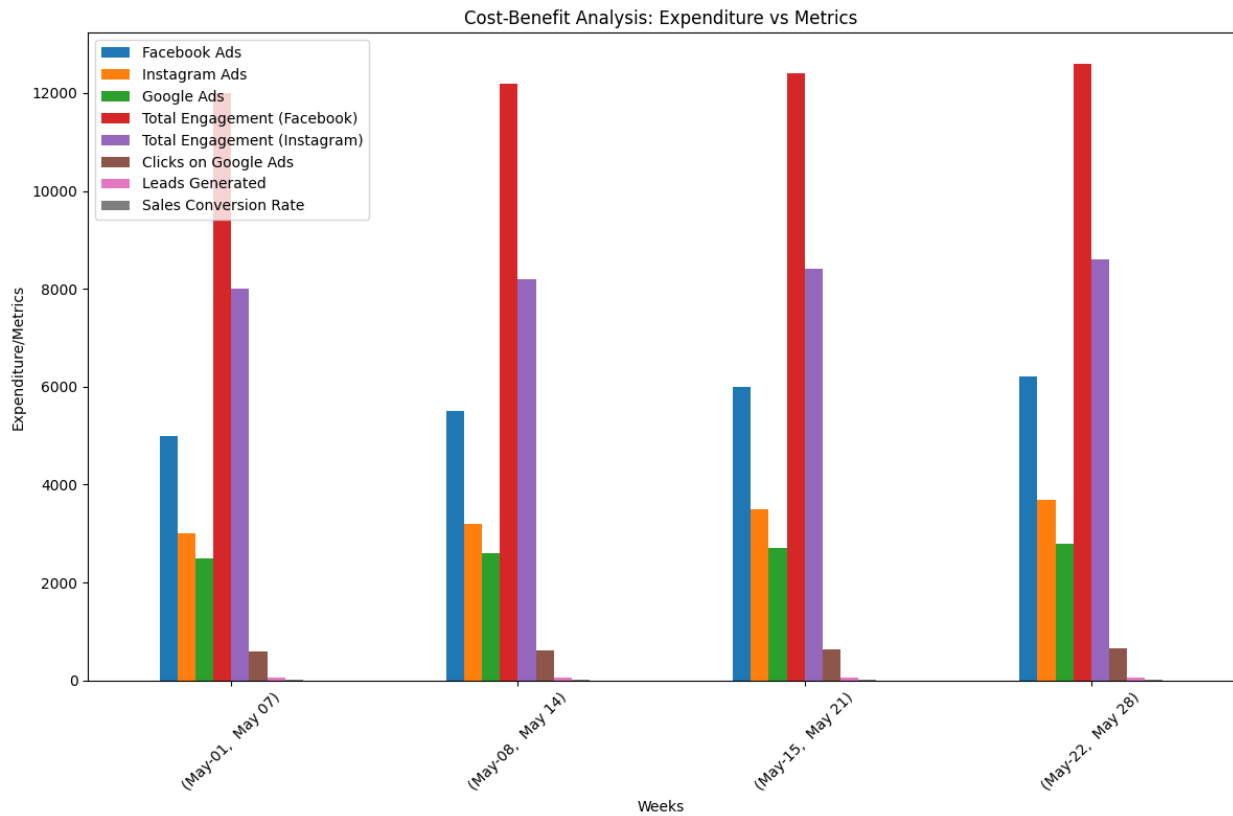
Facebook vs Instagram Engagement:



Google Ad Clicks:



4.2.3. Cost-Benefit Analysis:



5. Detailed Explanation of Analysis Process/Method

In this section, we delve into the methodology and analytical approaches employed to extract insights from the available data. While the current analysis is based on a qualitative understanding of the organization's operations and challenges, we provide a framework for conducting comprehensive quantitative analyses in the final submission, building upon the preliminary insights gained.

5.1. Analysis of Quality Control:

Justification:

- Problem Alignment: Focuses on the issue of uneven product quality, which is stated in the problem statement.
- Variability identification makes it possible to determine how different items, batches, or suppliers differ in terms of quality.

Why is it Appropriate?

- Root Cause Identification: Allows for the identification of certain material suppliers or variables that contribute to variances in quality, hence enabling targeted changes.
- Long-term Impact: Improving product quality has a favorable effect on the bottom line by maintaining consumer satisfaction and loyalty.

5.2. Assessment of Supplier Performance:

Justification:

- Problem Alignment: By concentrating on aspects relating to suppliers, this technique helps tackle the issue of ineffective inventory management.
- Data Leverage: Makes use of information that is readily available about pricing, lead times, and quantity across several vendors.

Why is it Appropriate?

- Optimization Potential: Reduces stockouts and overstocking by making dependable suppliers easier to find and work with.
- Cost optimization: Finding suppliers with shorter lead times and competitive prices helps reduce inventory costs and boost operational effectiveness.

5.3. Time Series Analysis for Marketing Expenditure

Justification:

- Problem Alignment: The difficulty of precisely forecasting consumer demand is in line with the requirement to recognize patterns and trends in marketing expenditure over time.
- Data Suitability: Weekly data makes it possible to analyze trends throughout distinct periods, which helps to spot long-term or seasonal patterns in ad spending.

Why is it Appropriate?

- Forecasting Capability: Time series analysis makes it possible to project future marketing expenditures by using data from the past, which helps with budgetary decisions.
- Impact Evaluation: It helps allocate resources to high-performing platforms by making it easier to evaluate the success of marketing initiatives over time.

5.4. Analysis of the correlation between engagement metrics and Ad spend:

Justification:

- Problem Alignment: Analyzing the relationship between ad expenditure and engagement metrics as well as recognizing customer segmentation are challenging tasks.
- The purpose of the objective assessment is to learn more about how various ad platforms affect consumer engagement.

Why is it Appropriate?

- Making strategic decisions: By determining the best platforms for interaction and conversion, strategic decision-making aids in the direction of marketing initiatives.
- Quantitative insights: Supports data-driven decision-making by offering quantifiable connections between engagement measures and ad spend.

6. Results and Findings

6.1. Product Quantity:

- The mean amounts of the CPVC products vary; CPVC_001 averages approximately 1388 units, whereas CPVC_002, CPVC_003 averages approximately 1550 units.
- Compared to other SWR types, SWR_001 has a significantly greater mean amount of about

4144 units, making it stand out.

- The mean quantities of UPVC items are moderate, particularly for UPVC_001 and UPVC_002.

6.2. Trends in Prices Per Unit:

- Consistent pricing is evident in the uniform pricing observed for various CPVC and SWR product types.
- The cost of UPVC goods varies somewhat, with UPVC_004 having the highest average cost per unit.

6.3. Geographic Distribution and Supplier:

- A number of suppliers make equal contributions; the most frequent contributors are Supplier_90, Supplier_39, and Supplier_44.
- The suppliers are dispersed throughout different geographical areas, with several of them situated in cities such as Hisar and Hubli.

6.4. Lead times and dates for manufacturing:

- Production appears to be centered around certain dates, with the highest number of units produced on June 19, 2023.
- variable suppliers have slightly variable lead times; the median lead times are often between five and six units.

6.5. Overview of Reorder Levels:

- Reorder levels vary widely, usually between 1500 and 4000 units, with 2000 units being the greatest number.

6.6. Physical Characteristics Among Material Kinds:

- The weight and area square meters of the various material types differ greatly, with UPVC having a significantly higher weight than SWR and CPVC.
- The pipes with the same materials all have the same thickness.

6.7. Expenditure and Advertising Trends:

- The average spending on advertising across platforms shows a consistent trend, particularly on Facebook and Instagram.
- Other categories of spending display somewhat variable standard deviations but constant mean values.

6.8. Event Participation and Engagement Metrics:

- Different data show that engagement on social media sites like Facebook and Instagram stays constant.
- Attendance at different events demonstrates a moderate to high degree of involvement.

6.9. Sales Conversion and Generated Leads:

- The measurements of leads generated and sales conversion rates show a moderate standard deviation and steady mean values.