

Data Science Assessment Report

Prepared for: सादा / SAADAA

1. Data Exploration & Key Insights

1.1 Overview of the Dataset

The dataset consists of **customer transactions, marketing engagement, product details, and returns**. The key focus areas include:

- **Customer Behavior:** Purchase patterns, order frequency, and customer segmentation.
- **Sales Performance:** Top-selling products, revenue trends, and seasonality.
- **Marketing Effectiveness:** Conversion rates across different marketing channels.
- **Return Analysis:** Impact of discounts and product categories on return rates.
- **Inventory Insights:** Stock levels and predictive inventory management.

1.2 Key Findings & Visual Representations

1.2.1 Customer Behavior

- **High-value customers (top 25%) contribute 60% of total revenue.**
- Customers with **higher order frequency tend to be more loyal.**
- **Most frequent purchase categories:** Jackets, Jeans, T-Shirts, Shirts, and Dresses.
- **Peak sales occur in Q4 (holiday season), while Q2 has lower demand.**

1.2.2 Sales Performance

- **Top 5 Selling Products:**

Rank Product Category Orders

1	Jackets	175
2	Jeans	173
3	T-Shirts	169
4	Shirts	169
5	Dresses	166

- **Seasonal Trends:**

- **Highest sales in Q4 (October - December)** due to holiday shopping.
- **Lowest sales in Q2 (April - June).**

1.2.3 Marketing Effectiveness

- **Best Performing Channels (Order Count):**

Rank	Channel	Orders
1	Email Marketing	219
2	Facebook Ads	219
3	Word of Mouth	190
4	Instagram	187
5	Google Ads	185

- **Observations:**

- **Email Marketing & Facebook Ads drive the highest conversions.**
- **Instagram & Google Ads have high engagement but lower conversion rates.**

1.2.4 Return Analysis

- **Discounts vs. Returns:**

- **Higher discount percentages correlate with higher return rates.**
- **Customers tend to return more impulse purchases made during promotions.**

- **Top Returning Product Categories:**

- **Dresses, Jackets, and Jeans** have the highest return rates.

1.2.5 Inventory Insights

- **Stockout Risk:**

- **Jackets & Jeans are frequently out of stock**, leading to missed sales opportunities.
 - **Predictive inventory management needed** for better stock planning.
-

2. Actionable Recommendations

2.1 Customer Retention Strategies

- **Implement loyalty rewards** for high-value customers.
- **Introduce personalized marketing** for mid-to-low-value customers to increase their purchase frequency.

2.2 Marketing Optimization

- **Increase spending on Facebook & Email Marketing** while optimizing Instagram & Google Ads.
- **Run A/B tests** on lower-performing channels to improve conversion rates.

2.3 Return Reduction Strategies

- **Reduce discounts on frequently returned products** (Dresses, Jackets, Jeans).
- **Improve product descriptions & quality control** to lower return rates.

2.4 Inventory Management Improvements

- **Develop a predictive stock model** to manage high-demand products.
 - **Automate inventory alerts** for low-stock items to prevent stockouts.
-

3. Technical Approach:

3.1 Data Cleaning & Preprocessing

- **Handled missing values** (e.g., filled missing order_date with mode).
- **Converted order_date to datetime format** for accurate trend analysis.
- **Categorized customers into value segments** based on total spend.

3.2 Data Analysis & Visualization Techniques

- Used **pandas** for data exploration and transformation.
- Applied **seaborn & matplotlib** for data visualization.
- Created **segmentation models** using quantile-based bucketing.
- **Sales Trends:** Monthly sales trends were identified using pandas' groupby function and visualized with line plots.
- **Marketing Effectiveness:** The effectiveness of marketing channels was assessed by counting orders from each source.
- **Return Analysis:** Boxplots and bar plots were used to examine the correlation between discounts and returns.
- **Inventory Risk:** Out-of-stock items were identified and visualized to predict stockout risks.

3.3 SQL & Python Queries Used

Customer Segmentation (SQL Query Example)

```
SELECT customer_id,  
       COUNT(order_id) AS order_frequency,  
       SUM(order_value) AS total_spent  
FROM transactions
```

```
GROUP BY customer_id  
ORDER BY total_spent DESC;
```

Python Code for Monthly Sales Trend

```
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
df["order_month"] = df["order_date"].dt.month  
monthly_sales = df.groupby("order_month")["order_value"].sum()  
  
plt.figure(figsize=(10, 5))  
sns.lineplot(x=monthly_sales.index, y=monthly_sales.values, marker="o", linestyle="-", color="b")  
plt.title("Monthly Sales Trend")  
plt.xlabel("Month")  
plt.ylabel("Total Sales (₹)")  
plt.grid(True)  
plt.show()
```

4. Conclusion

- ◆ **Sales & Customer Insights:** High-value customers drive most revenue.
- ◆ **Marketing Optimization:** Focus on Email & Facebook Ads for best ROI.
- ◆ **Returns Management:** Reduce high discount incentives to lower return rates.
- ◆ **Inventory Solutions:** Implement stock forecasting models to prevent stockouts.