**Retail Store Assortment**

**Diagnostic Analytics Use Case**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Version 1.0**

**Use Case Project:** Retail Store Assortment Diagnostic

**Jr. Business Analyst:** Sounak Dutta Chowdhury

**Type of Analytics:** Diagnostic Analytics

**Problem Type:** Anomaly Attribution Analysis

**Data Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No** | **Data Source Details:** | **Granularity:** | **Load Type:** | **Historical Data Range:** |
| 1. | RSA - Product\_Table\_2 | Transaction Level (Based on Customer ID & Product Category) | Batch | 2023-01-01 to 2024-01-01, 1 year of data |

**Data Dictionary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Parameters/Identifiers** | **Definition** | **Data Type** |
| 1. | Transaction ID | Unique identifier for each transaction | int |
| 2. | Date | Date of the transaction | date |
| 3. | Customer ID | Unique identifier for each customer | string |
| 4. | Gender | Gender of the customer | string |
| 5. | Age | Age of the customer | int |
| 6. | Product Category | Category of the product sold | string |
| 7. | Quantity | Number of units sold | int |
| 8. | Price per Unit | Price for each unit of the product | int |
| 9. | Total Amount | Total revenue generated from the transaction | int |

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Metrics** | **Definition** | **Data Type** |
| 1. | Total Quantity Sold to Customers | Sum of quantities purchased by each customer | Integer |
| 2. | Total Revenue per Customer | Total amount spent by each customer | Integer |
| 3. | Average Order Value | Average amount spent per order (Total Amount / Quantity) | Float |
| 4. | Age Group | Grouping customers into age ranges (e.g., 18-25, 26-35, etc.) | Categorical |
| 5. | Price Sensitivity | Categorization of price sensitivity (Low, Moderate, High) | Categorical |
| 6. | Transaction Month | Month of the transaction extracted from the Date column | String |

**Problem Statement**

Retailers face challenges in optimizing store assortment due to an over-reliance on an improper inventory allocation and replenishment. Asymmetric stock allocation to higher-class stores, inconsistent services, and improper inventory management result in inefficiencies, overstocking risks, and reduced customer satisfaction. When certain products or categories experience decreased sales performance, it leads to a loss in overall revenue and profitability. Inconsistent sales trends may be due to an unbalanced product mix, ineffective promotional strategies, supply chain issues, or suboptimal product placement in stores.

Understanding the patterns and root causes of these sales variations along with demographic influences and purchase behavior will help optimize future sales strategies. A more dynamic, data-driven approach is essential to enhance assortment efficiency and minimize working capital exposure.

**Business Objectives**

1. **Identify the Root Causes of Low Sales**: Explore which products, categories, customer demographics, product prices are underperforming in sales.
2. **Quantify the Impact of Sales Performance**: Analyze how product preferences in sales affect total revenue.
3. **Highlight Key Channels**: Identify key distribution channels where these patterns are most prevalent.

**Data Sources**

* RSA - Product\_Table\_2 (CSV): Sales data containing records of different product categories with their price, quantity, demographic details, purchase patterns etc. across various stores.
* **Date Range**: 2023-01-01 to 2024-01-01, 1 year of data.

**Data Quality Observations**

* **Null or Missing Values**: There are no missing sales records or product information.
* **Outliers**: Unusual spikes or drops in sales in the below table like:

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Product Names** | **Monthly** | **Quarterly** |
| 1. | Electronics | December (40),  March (14) | Qtr. 4 (102),  Qtr. 1 (66) |
| 2. | Clothes | May (37),  July (19) | Qtr. 2 (101),  Qtr. 3 (71) |
| 3. | Beauty | October (31),  March (21) | Qtr. 2, Qtr. 4 (82),  Qtr. 3 (71) |

**Methodologies**

**1. Correlation Analysis**

Purpose: To identify relationships between sales depending on the Transaction IDs and various independent variables (e.g., product category, quantity, seasonality, promotions).

How: Calculate correlation coefficients (e.g., Pearson or Spearman) to understand which factors are strongly associated with sales fluctuations.

Tools: Use statistical packages like `pandas`, `NumPy`, or `R` for calculating correlations.

**2. Time Series Analysis**

Purpose: To explore seasonal patterns and trends in retail sales over time.

How: Decompose time series data into trend, seasonality using methods like moving averages or seasonal decomposition (STL).

Tools: Python’s `statsmodels` or `R` packages for time series decomposition.

**3. Anomaly Detection**

Purpose: To detect unusual patterns in sales data, such as sharp drops or spikes for finding outliers.

How: Use statistical methods like Z-score to identify anomalies.

Tools: Python’s `scikit-learn` or `PyOD` for anomaly detection.

**4. Regression Analysis:**

Understanding Key Sales Drivers, Identifying Relationships Between Sales and Promotions, Measuring Store-Level or Product Category Differences etc.

Purpose: Regression helps to determine which factors (e.g., product category, promotion, seasonality) significantly impact sales (Transaction IDs). It helps assess the effectiveness of promotions on sales. Also, it can explain how sales vary across different seasons or product categories.

How: By running a multiple linear regression model, we can understand how much sales increase or decrease when promotions are applied or when products are sold in different seasons in a year. Also, a binary variable (`1` for promotion applied, `0` for no promotion) to see how sales differ between promotional and non-promotional periods. This helps businesses diagnose performance issues at specific stores or with certain product lines, enabling focused improvement strategies.

Tools: scikit-learn for regression models.

**Observations**

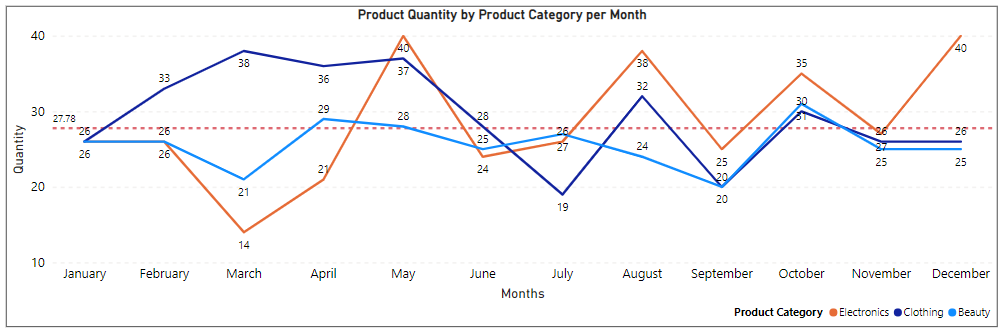
**Product Categories:**

1. Electronics
2. Clothes
3. Beauty

**Sales Variation Factors:**

1. Promotional Activities (Marketing Campaign, Social Media Ads etc.)
2. Sales and Discounts (Prime Day, Black Friday, Diwali, Christmas etc.)

**1. Month Level Sales Trend**



* Observation of sales variation on a Monthly basis.
* Average sales per month are nearly 28 units/ category.

1. Electronics: (Fan, AC, TV, Phones, Laptops etc.)

* Significant decreases in sales during non-promotional periods (Feb – Mar).
* Significant increase in sales during summer season (due to heat, back to school etc.)
* Certain increase due to sales (Prime Day in July, Diwali in Oct, Christmas in Dec) and promotional activities for new product launches. (New models, new colors, new features etc.)

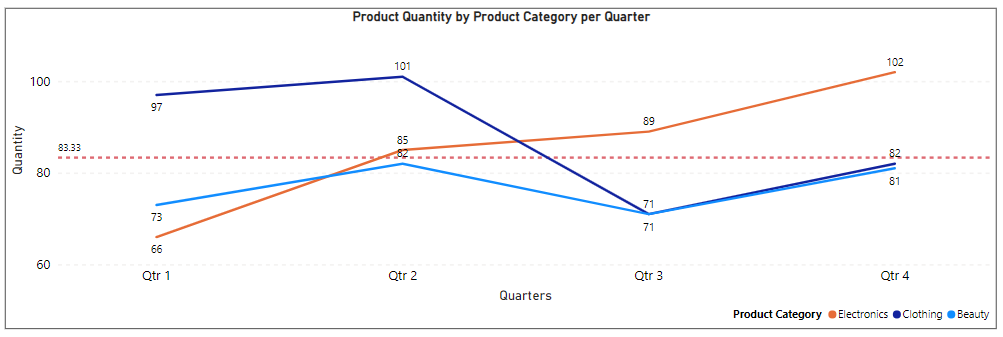
1. Clothes: (T-Shirts, Jeans, Sweater, Tops, Skirts, Coats, Wedding Dress etc.)

* Massive rise from Jan to march and the trend goes to May. (Winter, Spring, Vacations etc.)
* Massive drop from May to July (Summer issues, no festivals etc.)
* For promotional activities and festive season then, uptrend resumes till the end of the year.

1. Beauty Products: (Creams, Moisturizer, Lipstick, Cosmetics etc.)

* Trend is moderate over the year.
* Increase demands from March to April (for normal sales) and September to October (for Festive Season)

**2. Quarter Level Sales Trend**



* Observation of sales variation on a Quarterly basis.
* Average sales per quarter are nearly 84 units/ category.

1. Electronics:

* The quarterly sales analysis shows a sharp rise (from Qtr. 1 to Qtr. 4) during promotional weeks, seasonal sales, festive seasons etc. Also, new product demands with affordable price increase the sales.

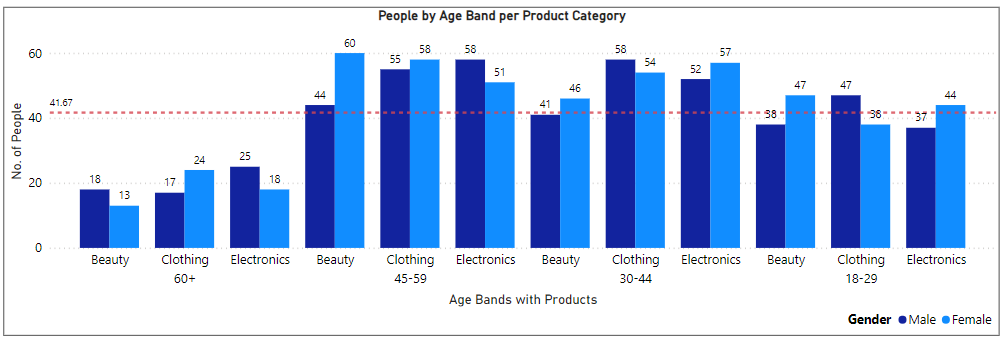
1. Clothes:

* Significant drop from Qtr. 2 to Qtr. 3 (No vacations, Less Holidays etc.)
* Descent rise in sales from Qtr. 3 to Qtr. 4 (Festive season, Celebrations etc.)

1. Beauty Products:

* Moderate ups and downs throughout the entire year.

**3. Age Band Level Product Category Analysis**



* Observation of sales variation on an Age Band basis.
* Average sales per quarter are nearly 84 units/ category.

1. Electronics:

* This category exhibits consistent sales throughout different age bands over the year.

Sharp rise in sales in the age group of 30 to 60 for both the demographics.

* From younger generation to mid (range from 18 to 45) this category is female dominated. The trend reverses for the older generations.

1. Clothes:

* From younger generation to mid (range from 18 to 45) this category is male dominated. The trend reverses for the older generations.

1. Beauty Products:

* This has flat patterns throughout the entire year.
* This is particularly female dominated from teenage to old age groups.

**Conclusion**

Overall conclusions from the above observations are done with quantification of the impact of identified independent variables on the target variable.

* **Impact of Promotions**: Promotions significantly impact on sales (Electronics in Qtr. 3 and Qtr. 4), but certain categories underperform regardless of promotional activity (Beauty and Clothing in Qtr. 3, Qtr. 4).
* **Sales Gaps**: Identified gaps in sales across several product categories, indicating areas for improvement in supply chain or promotion strategies.

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Product Names** | **Monthly** |
| 1. | Electronics | February - March (Massive Drop) |
| 2. | Clothes | May - July (Massive Drop) |
| 3. | Beauty | July - September (Normal Drop) |

**Definition of Done (DoD)**

* Completion of the EDA report with detailed insights into purchase patterns, promotional effectiveness, seasonal sales performance, Customer preferences.
* Generation of actionable insights for further analysis and potential forecasting model development.