

Multi-Dimensional Sales Performance Analytics Using Excel

Project Objective: To design a **star-schema-based retail sales data model** and enable multi-dimensional sales, customer, product, and store performance analysis using Excel-based analytical techniques through data cleaning, aggregation, pivot tables, and business rule–driven metrics such as customer loyalty.

Dataset Description: The dataset is structured in a star schema consisting of a Sales Fact table and Customer, Product, and Store Dimension tables.

The **Sales Fact table** captures transaction details such as order date, customer ID, product ID, store ID, quantity, unit price, discount, payment type, and total amount.

The **Customer Dimension table** contains customer demographics including age, gender, location, and loyalty level, linked to the fact table using Customer ID.

Note: Stock is a time- and store-dependent measure and is more appropriately modeled in a separate Inventory Fact table rather than in the Product Dimension or Sales Fact table to maintain correct grain.

The **Product Dimension table** includes product attributes such as category, sub-category, brand, and cost, supporting product-level analysis.

The **Store Dimension table** provides store information including region, city, and store type for location-based analysis.

This structure enables effective multi-dimensional sales performance analysis in Excel.

Sales_Fact Table Column Descriptions:

Column Name	Description
Sales_ID	Unique identifier for each sales transaction
Order_Date	Date on which the sales transaction occurred.
Customer_ID	Foreign key linking to the Customer Dimension Table , enabling customer-level analysis
Product_ID	Foreign key linking to the Product Dimension

	Table , used for product and category performance analysis.
Store_ID	Foreign key linking to the Store Dimension Table , supporting store-wise and regional performance evaluation.
Quantity	Number of units sold in a transaction.
Unit_Price	Price per unit of the product at the time of sale.
Discount	Discount applied to the transaction.
Payment_Type	Mode of payment used (e.g., Cash, Card, UPI, Online).
Total_Amount	Final transaction value after applying quantity, unit price, and discount.

Customer_Dim Table Column Descriptions:

Column Name	Description
Customer_ID	Unique identifier for each customer. Acts as the primary key and is used to link with the Sales Fact table.
Name	Customer name used for customer-level analysis .
Age	Age of the customer at the time of collecting the data.
Gender	Gender of the customer.
City	City where the customer is located.
	State where the city of the customer belongs to.
Country	Country of residence.
Loyalty_Level	Classification of customers based on engagement or spending behavior (e.g., Bronze, Silver, Gold)

Product_Dim Table Column Descriptions:

Column Name	Description
Product_ID	Unique identifier for each product. Serves as the primary key and is used to establish a relationship with the Sales Fact table.
Product_Name	Product name used for Product-wise sales analysis .
Category	Classification of the Product.
Sub Category	Detailed classification of the product with in the category.
Brand	Brand or manufacturer of the product.
Cost	Standard or base cost per unit of the product.
Stock	Represents the current inventory level of the product at the time of data collection.

Store_Dim Table Column Descriptions:

Column Name	Description
Store_ID	A unique number used to identify each store. Serves as the primary key and is used to establish a relationship with the Sales Fact table.
Store_Name	The name of the store
Region	The area or region where the store is located
City	The city where the store is present
Store_Type	The type of store (example: Retail, Outlet, Online).

Data cleaning & transformation:

Sales_Fact table:

- The Sales Fact table was converted into an Excel table to support structured validation and analysis. Data types were validated to ensure consistency across all columns, including text, numeric, date, and currency fields. No duplicate Sales_ID values were found, confirming the uniqueness of transactions.
- Some Customer_ID values (CUST1063, CUST1267, CUST1290, CUST1374, CUST1386, CUST1395, CUST1429, CUST1487) were present in the Customer Dimension but had no corresponding sales records in the fact table. This represents inactive or non-purchasing customers during the analysis period and was treated as a valid business scenario. No blank rows or header issues were identified.
- Missing values in Quantity, Unit Price, and Total Amount were identified using below Excel formulas. All text-based columns were standardized using TRIM and CLEAN functions to remove extra spaces and non-printable characters. Proper alignment was applied to improve readability.

```
=IF(ISBLANK([@[Unit_Price]]),[@[Total_Amount]]/([@[Quantity]*(1-[@Discount]))],[@[Unit_Price]])
```

```
=IF(ISBLANK([@[Quantity]]),[@[Total_Amount]]/([@[Unit_Price]*(1-[@Discount]))],[@[Quantity]])
```

```
=IF(ISBLANK([@[Total_Amount]]),([@[Quantity]*[@[Unit_Price]]*(1-[@Discount])),[@[Total_Amount]])
```

Store_Dim table:

The Store Dimension table was validated for data types, duplicates, and missing values. No duplicate Store_ID values or blank rows were found. All text columns were standardized to ensure consistent formatting, and column alignment was applied. No major data quality issues were identified in this table.

Customer_Dim table:

- Customer Dimension table was converted into an Excel table and validated for duplicates and missing rows. Data types were verified, and text columns were standardized using TRIM and CLEAN functions. Inconsistent Customer_ID formats were corrected using the SUBSTITUTE function.
- Customer names were reformatted for consistency. The State column was removed as it contained a single constant value and did not add analytical value.

- Loyalty_Level was derived based on cumulative customer spending calculated from the Sales Fact table. Customers without sales transactions were assigned zero total sales before loyalty classification. Loyalty categories were assigned using predefined spending thresholds. Proper alignment was applied to all columns.
- I have created a separate column “Total Sales Amount” in Customer_Dim table and filled the values using below formula for each customer.

=SUMIF(Table4[Customer_ID],[Customer_ID],Table4[Total_Amount])

- Now in loyalty column, filled the values using the below formula. It gives the loyalty value as

“Platinum” if Total Sales Amount >=10000,

“Gold” if Total Sales Amount>=5000,

“Silver” if Total Sales Amount >=2000 else it assigns “Bronze”

=IF([@[Total Sales Amount]]>=10000,"Platinum",IF([@[Total Sales Amount]]>=5000,"Gold",IF([@[Total Sales Amount]]>=2000,"Silver","Bronze")))

- Age groups were created using logical conditions to support age-based analysis.

=IF([@Age]<13,"Child",

IF(AND([@Age]>=13,[@Age]<=19),"Teen",

IF(AND([@Age]>=20,[@Age]<=39),"Adult",

IF(AND([@Age]>=40,[@Age]<=59),"Middle Aged",

IF([@Age]>=60,"Senior Citizen",""))))

Product_Dim table:

- The Product Dimension table was validated for unique Product_ID values and correct data types. Inconsistent Product_ID formats were standardized using the SUBSTITUTE function. Text columns were cleaned using TRIM and CLEAN functions.
- The Cost column was renamed to Base_Cost to avoid confusion with sales values. Missing Base_Cost values were handled using average imputation for analytical continuity. The Stock column was removed, as stock is not analytically appropriate in a multi-store sales analysis. Column alignment was applied throughout the table.
- Below formula fills the empty cell values with the Average value.

=IF(ISBLANK([@Cost]),AVERAGE([Cost]),[@Cost])

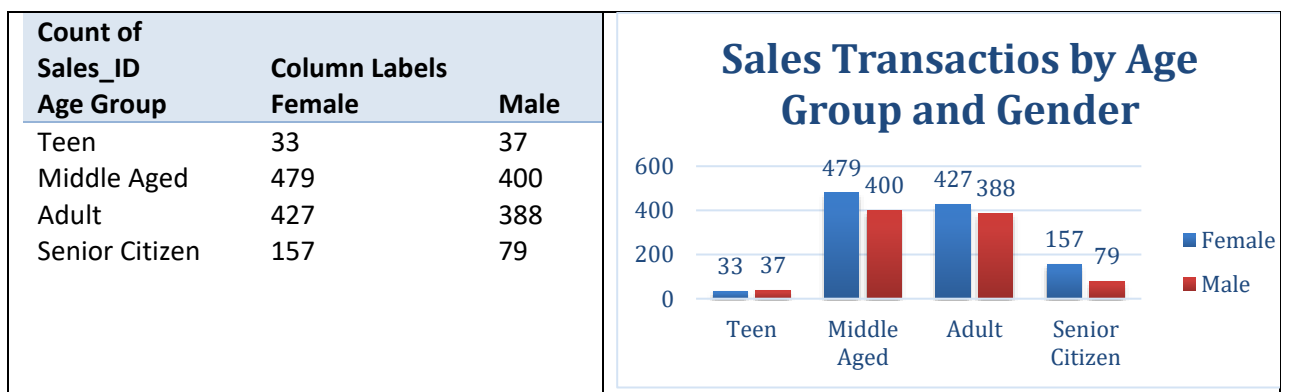
- Stock value in Product Dim table is not analytically appropriate when we have multiple stores and cities. So I have deleted this column from the Product_Dim table.

Descriptive Statistical Analysis using Excel Tool Pak:

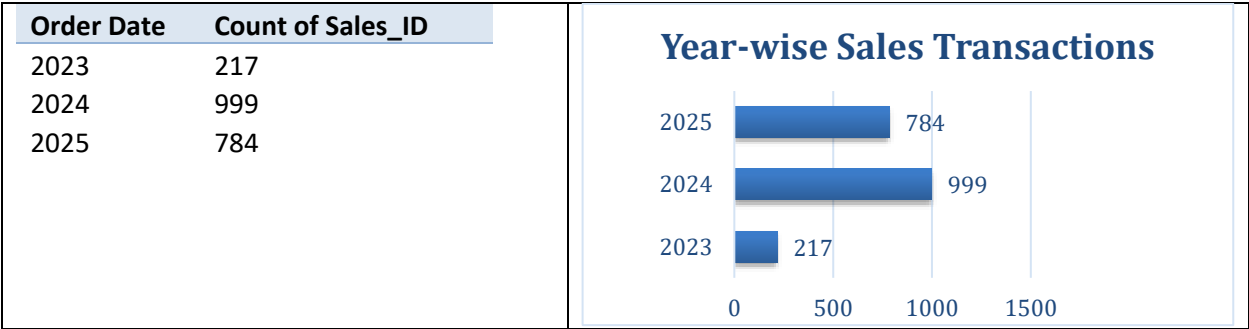
- The given data was analyzed using Excel's Analysis Tool Pak to generate descriptive statistical values for total sales in sales fact table.

Mean	1087.197349
Standard Error	18.3272372
Median	868.7182
Mode	#N/A
Standard Deviation	819.6189642
Sample Variance	671775.2464
Kurtosis	0.045826153
Skewness	0.87904704
Range	3889.7838
Minimum	18.459
Maximum	3908.2428
Sum	2174394.698
Count	2000

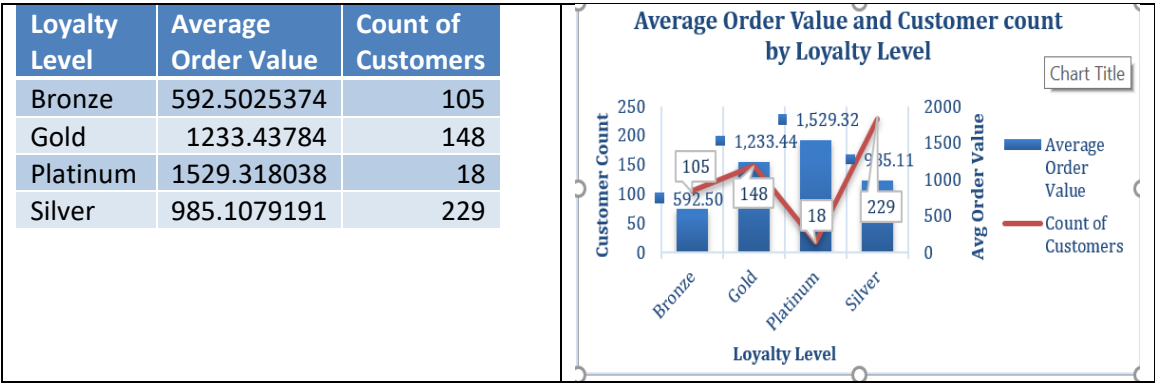
- The dataset contains 2000 records with a mean value of 1087.19 and a median of 868.72. Here median is lower than the mean, it indicates the presence of higher-value transactions.
- The standard deviation of 819.62 indicates high variability in sales values showing that sales performance is not consistent and varies significantly across transactions.
- The distribution is positively skewed with a skewness of 0.88, suggesting the presence of higher-value transactions.
- Kurtosis is close to zero, indicating near-normal distribution.
- The minimum and maximum values range from 18.46 to 3908.24, showing a wide spread in sales performance.
- **Visualization and Insights Arrived with screenshot of visualization:**



- The chart shows sales transactions across different age groups segmented by gender.
- Higher transactions among Middle-aged and Adult groups indicate stronger purchasing activity in these segments.
- Middle-aged and Adult customers are likely to continue driving the majority of sales transactions.
- Business efforts should focus on retaining these two age groups while targeting Teens and Seniors to improve their sales transactions.

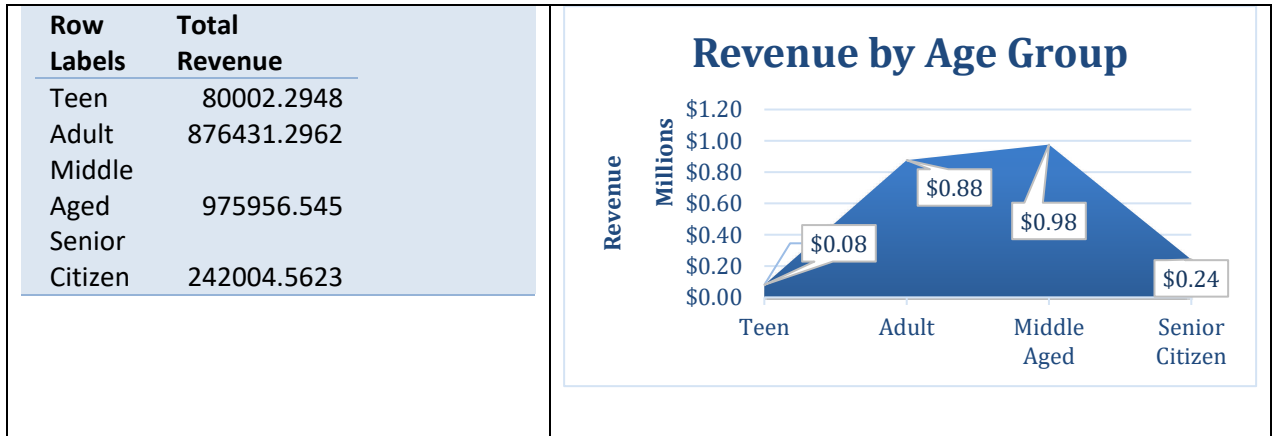


- The chart illustrates the distribution of sales transactions across different order years.
- Higher transactions in recent years indicate increased customer activity and improved sales performance.
- If the current trend continues, sales transactions are expected to remain strong in upcoming years.
- Sustained marketing and operational strategies should be applied to maintain and further enhance transaction growth.

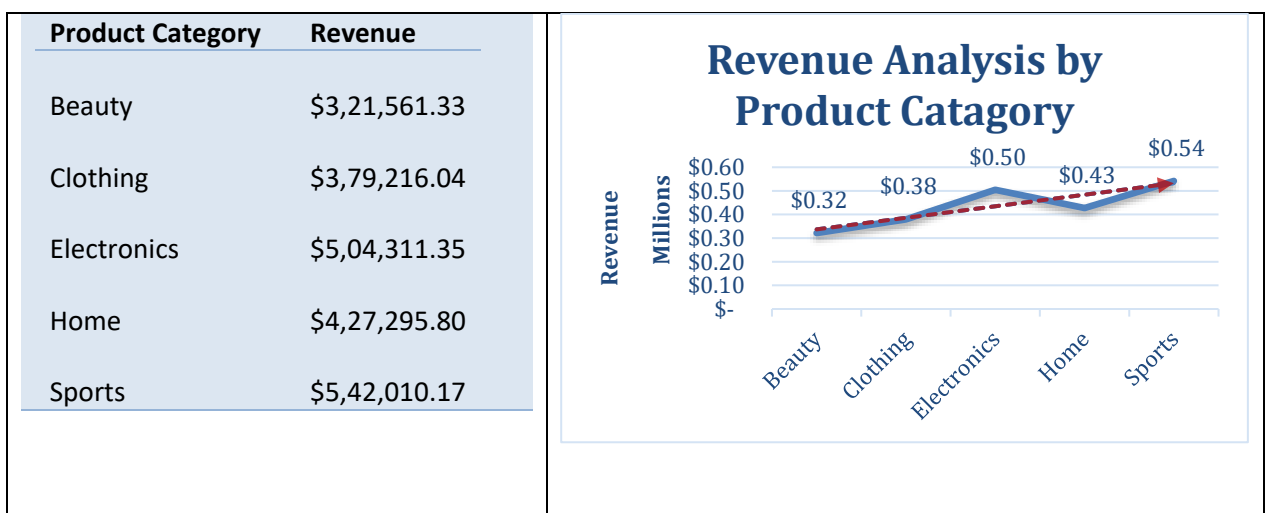


- Shows the average order value and number of customers across different loyalty levels.

- Explains why Platinum customers generate higher order values despite having fewer customers.
- Indicates that higher loyalty levels are likely to contribute more revenue per customer in the future.
- Recommends focusing on retaining and upgrading customers to higher loyalty tiers to maximize revenue.

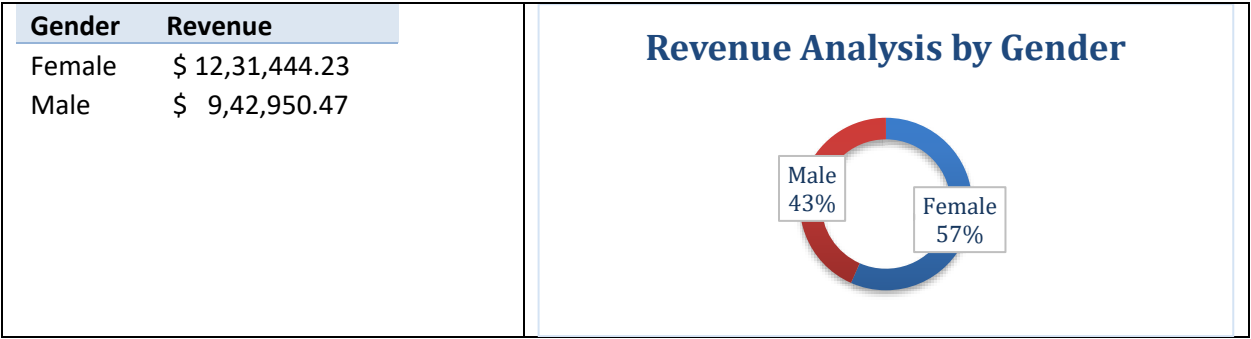


- Summarizes the total revenue contribution of each age group.
- Indicates that higher purchasing power and spending capacity drive greater revenue among adult and middle-aged customers.
- Implies that future revenue growth will be driven mainly by customers in the adult and middle-aged segments.
- Suggests prioritizing personalized offers and premium products for high-revenue age groups while nurturing lower-revenue segments.



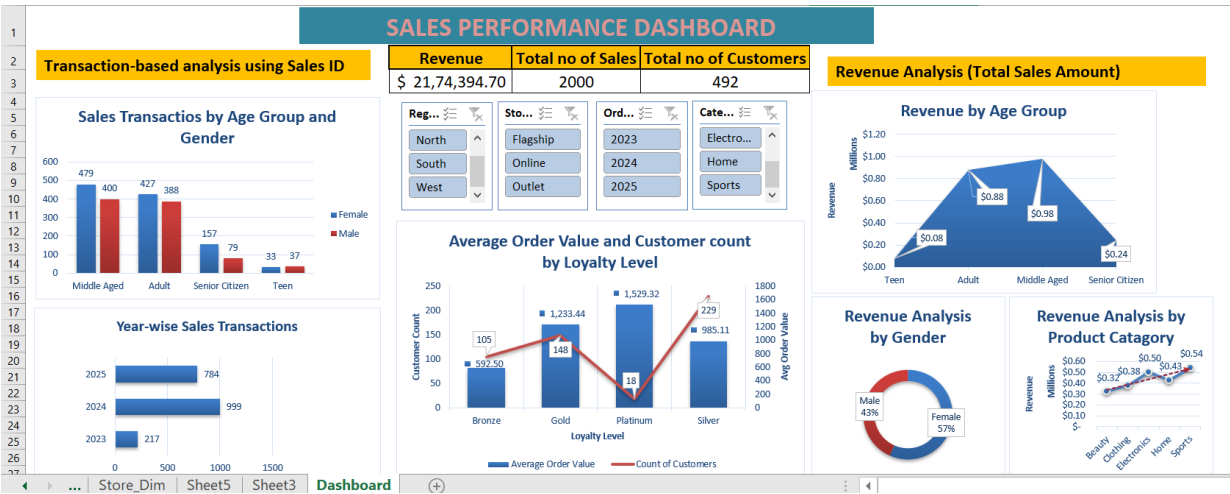
- Presents total revenue generated by each product category.

- Highlights that Electronics and Sports contribute higher revenue compared to Beauty and Clothing.
- Indicates that categories with consistently higher revenue are likely to remain key revenue drivers.
- Recommends strengthening high-performing categories while improving marketing strategies for lower-revenue categories.



- Shows the proportion of total revenue contributed by male and female customers.
- Indicates that female customers contribute a higher share of revenue compared to male customers.
- Suggests that female customers are likely to continue being a major revenue contributor.
- Recommends designing targeted campaigns to retain female customers while increasing engagement among male customers.

Conclusion:



a few strong product categories, with clear opportunities to improve performance through targeted strategies.