

# Amazon Sales Analysis

## Low-Level Design

**Version:** 1.0

**Revision Date:** 22-11-24

Riya Singh

## Document Version Control Table

version	Date	Author	Changes
1.0	22-11-2024	Riya Singh	Descriptions
1.1	24-11-2024	Riya Singh	Diagram

## Contents

<b>1. Introduction.....</b>	<b>4</b>
1.1 Low-Level Documentation (LLD).....	4
<b>2. System Architecture Design.....</b>	<b>6</b>
2.1 Data Collection and Source.....	6
2.2 Data Cleaning and Transformation (ETL Process).....	6
<b>3. Architecture Description.....</b>	<b>8</b>
3.2 Data Loading.....	9
3.3 Data Transformation.....	9
3.4 Data Modeling.....	10
3.5 Dashboard Deployment.....	11
<b>4. Dashboard Report and Insights.....</b>	<b>12</b>
4.1 Overview of the Dashboard.....	12
4.2 Key Visualizations.....	12
4.3 Key Insights Derived from the Dashboard.....	12
4.4 Recommendations Based on Insights.....	14
4.5 Future Enhancements.....	14

# 1. Introduction

## 1.1 Low-Level Documentation (LLD)

Low-Level Documentation (LLD) provides a detailed step-by-step explanation of the processes involved in the project. It includes cleaning the dataset using Python, transforming data for analysis, and building relationships between attributes in Power BI. The LLD ensures that each technical component, such as calculated columns, data flow, and dashboard design, is documented for reproducibility and clarity.

## 1.2 Scope

The scope of this project is to analyze Amazon sales data to identify trends, patterns, and insights that can drive business decisions. It focuses on time-based sales analysis, revenue and profit metrics, and sales performance across various regions and categories. The project also involves designing an interactive Power BI dashboard for visualizing insights in a user-friendly manner. Predictive modeling and real-time data analysis are not covered.

## 1.3 Problem Statement

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise.

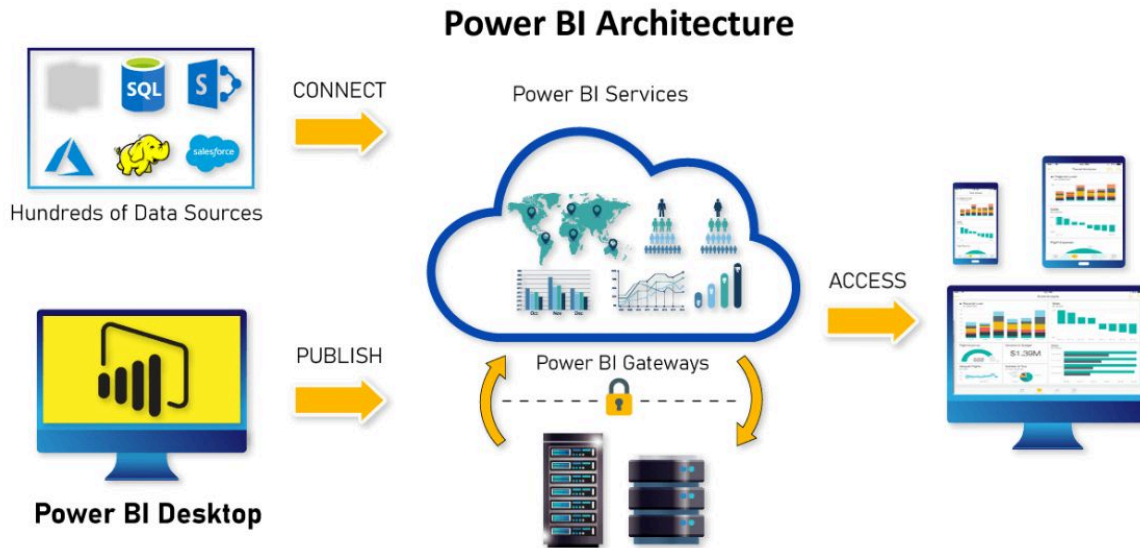
ETL process.

Do ETL : Extract-Transform-Load some Amazon dataset and find for me

Sales-trend -> month wise , year wise , yearly\_month wise .

## 2. System Architecture Design

The architecture design of the Amazon Sales Analysis project is aimed at outlining the flow of data from its source to the final dashboard. Below is a detailed breakdown of the architecture:



### 2.1 Data Collection and Source

**Source:** The data for this project is sourced from a dataset containing 101 rows and 18 columns, including sales attributes like *region*, *country*, *item type*, *sales channel*, and sales figures (units sold, revenue, profit).

**Format:** The dataset is in a CSV format, which is loaded into Google Colab for data cleaning and transformation.

### 2.2 Data Cleaning and Transformation (ETL Process)

**Extract:** The raw data is loaded from the dataset into Python (Google Colab).

**Transform:** Using Python, the data is cleaned and prepared by:

- Removing or correcting any inconsistencies.
- Creating new columns like *year*, *month*, and *total sales* to assist with time-series analysis.
- Ensuring all columns have the correct data types (e.g., date, numeric).

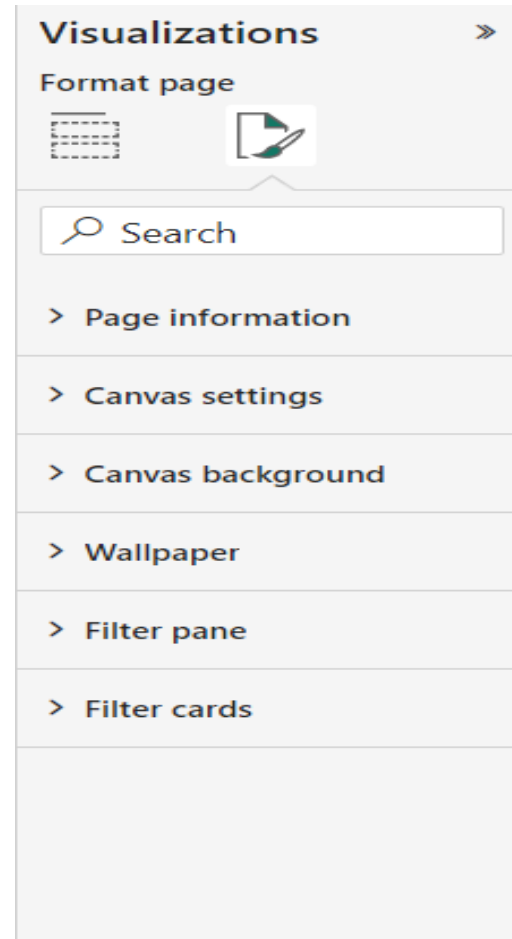
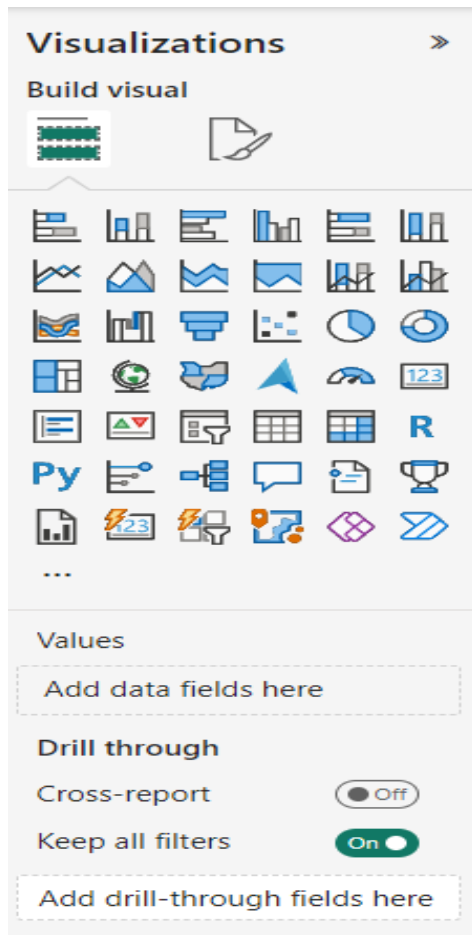
**Load:** The cleaned data is imported into Power BI for analysis and visualization.

## 2.3 Data Modeling and Relationships

- **Power BI Setup:**
  - A **DateTable** is created in Power BI to facilitate time-based analysis.
  - Relationships between tables (e.g., main dataset and DateTable) are established for accurate aggregations and visualizations.

## 2.4 Data Analysis and Visualizations

- **Metrics and KPIs:** Key metrics like total revenue, total cost, and total profit are calculated and displayed.
- **Visuals:**
  - Line charts to show monthly and yearly sales trends.
  - Pie charts and bar graphs for sales performance by region, country, and sales channel.
  - Filter panels for interactive analysis.



## 2.5 Dashboard Deployment

- The final Power BI dashboard is deployed and shared for user interaction.
  - **User Interface:** The dashboard features a Pink-themed layout, with visually appealing graphs and filters for users to explore different aspects of the data.
  - **Accessibility:** The dashboard can be accessed via Power BI Service or embedded into web platforms for wider use.

## 3. Architecture Description

### 3.1 Data Description

The Amazon Sales Analysis project uses a dataset consisting of two tables: **Sales Data** and **Customer Data** for the years 2017, 2018, and 2019. These tables provide detailed sales and customer information required for analysis. Below are the features of both tables:

#### Amazon Sales Data Features:

1. **Order ID:** Unique identifier for each order.
2. **Order Date:** The date on which the order was placed.
3. **Ship Date:** The date on which the order was shipped.
4. **Region:** Geographic region of the order, e.g., Central, South.
5. **Country:** Country where the order was placed.
6. **Sales Channel:** Mode of sales, e.g., Online, Offline.
7. **Order Priority:** Priority assigned to the order, e.g., High, Low.
8. **Item Type:** Category of the item sold.
9. **Units Sold:** Quantity of the items sold in the order.
10. **Unit Price:** Price per unit of the item.
11. **Unit Cost:** Cost per unit of the item.
12. **Total Revenue:** Calculated as the product of *Units Sold* and *Unit Price*.
13. **Total Cost:** Calculated as the product of *Units Sold* and *Unit Cost*.
14. **Total Profit:** Calculated as the difference between *Total Revenue* and *Total Cost*.
15. **Year:** Extracted year from the order date.
16. **Month:** Extracted month from the order date.
17. **Year\_Month:** Concatenated format of year and month for time-series analysis.

#### Customer Data Features:

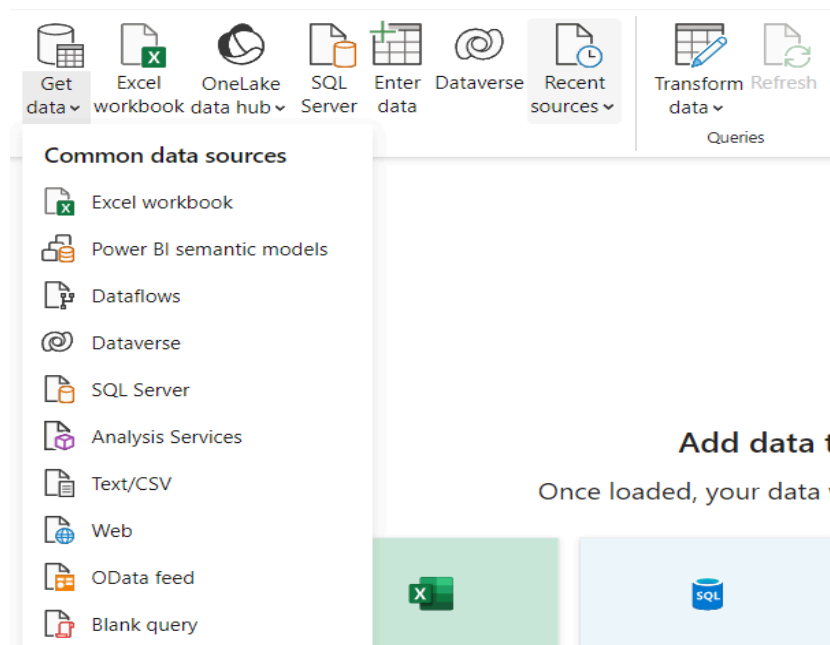
1. **Customer ID:** Unique identifier for each customer.
2. **Region:** Geographic region where the customer resides, e.g., Central, Southern.
3. **Division:** Classification of the customer as Domestic or International.
4. **City:** Name of the customer's city.

5. **Country:** Name of the customer's country.
6. **Zip Code:** Postal code of the customer's location.

### 3.2 Data Loading

The process of loading the data into Power BI involves the following steps:

1. **Open Power BI Desktop Application**
  - Launch the Power BI Desktop tool on your system.
2. **Connect to Data Source**
  - Click on **Get Data** → Choose the data source type (e.g., Excel or CSV).
  - Browse and select the dataset files (Sales and Customer data).
3. **Load Data into Power BI**
  - After connecting, load both the **Sales Data** and **Customer Data** tables into Power BI.
  - Preview the data to verify correct loading and ensure both tables are available in the data model.
4. **Establish Relationships**
  - Link the **Customer ID (Custkey)** from the Customer table to the **Customer ID (Custkey)** in the Sales table using a one-to-many relationship.
  - Optionally, create additional calculated columns if required for specific analyses (e.g., concatenating year and month).
5. **Create Date Table for Time-Based Analysis**
  - Generate a DateTable in Power BI to support time-based calculations.
  - Relate the **Order Date** column from the Sales table with the DateTable's date field.





### 3.3 Data Transformation

#### Overview

The transformation process converts the original datasets into a format suitable for analysis by modifying and enhancing attributes as needed. The transformations are performed using Power BI's Power Query Editor, which supports the ETL process.

#### Steps in the Transformation Process:

1. **Attribute Format Conversion:**
  - Ensure numerical fields (e.g., Sales Amount, Unit Price) are set as numeric types.
  - Convert date fields (e.g., Order Date, Ship Date) into *DateTime* format for accurate analysis.
2. **Feature Engineering:**
  - Add calculated fields for metrics such as *Total Revenue*, *Total Cost*, and *Total Profit*.
  - Create concatenated fields like *Year\_Month* for time-series analysis.
3. **Dim\_Date Table Creation:**
  - A new Date Dimension Table (Dim\_Date) is created to enhance time-based analysis.
  - The table includes the following features:
    - **Date:** Unique date values.
    - **Day of Week:** Name or numeric value of the day.
    - **Month Name:** Full name of the month (e.g., January).
    - **Month Number:** Numeric representation of the month (e.g., 1 for January).
    - **Year:** Extracted year.
    - **Year Quarter:** The quarter of the year (e.g., Q1, Q2).

### 3.4 Data Modeling

#### Purpose:

Data modeling establishes relationships between tables to ensure efficient querying and analysis in Power BI. It involves organizing data into logical structures and creating relationships to enable seamless reporting.

#### Steps in Data Modeling:

1. **Define Relationships:**
  - A one-to-many relationship is created between the Customer Table and the Sales Table using **Custkey**.
  - A one-to-many relationship is created between the Dim\_Date table and the Sales Table using **Order Date**.

## 2. Star Schema Design:

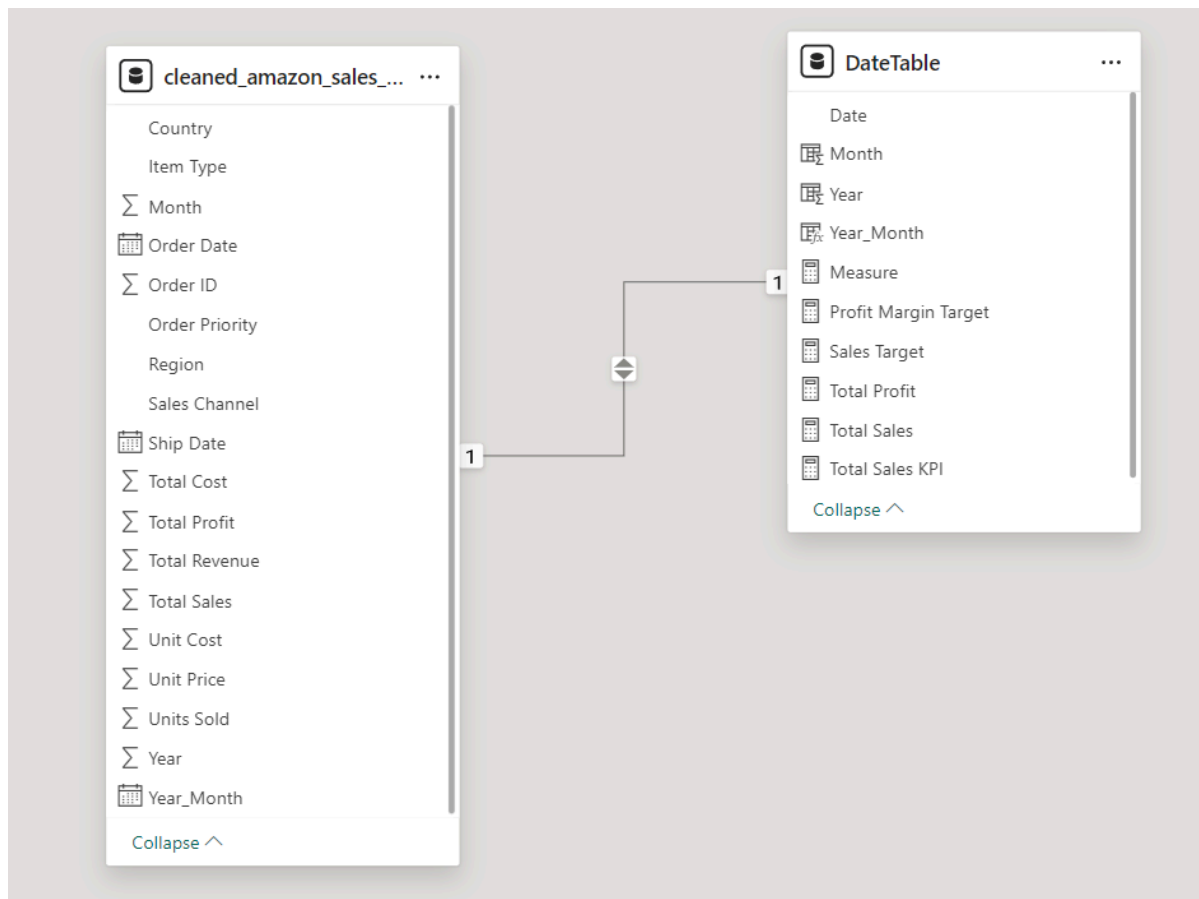
- The model follows a star schema with the Sales Table as the fact table and the Customer Table and Dim\_Date as dimension tables.

## 3. Calculated Measures:

- Create measures for KPIs, such as:
  - **Total Sales Amount** =  $\text{SUM}(\text{Sales}[\text{Total Revenue}])$
  - **Total Profit Margin** =  $\text{SUM}(\text{Sales}[\text{Total Profit}]) / \text{SUM}(\text{Sales}[\text{Total Revenue}])$

## 4. Time Intelligence Measures:

- Add measures for time-based calculations like:
  - **Yearly Sales**: Calculate sales aggregated by year.
  - **Monthly Profit**: Calculate profit for each month.



### 3.5 Dashboard Deployment

**Overview:**

The final step is the deployment of the Power BI dashboard to allow end-users to interact with the insights.

**Steps for Deployment:**

1. **Publish to Power BI Service:**
  - Once the dashboard is ready, it is published to Power BI Service for accessibility.
2. **User Accessibility:**
  - Users can access the dashboard through shared links or embedded visualizations.
3. **Interactivity Features:**
  - Filters, slicers, and drill-through capabilities are included to enable users to explore specific regions, periods, or products.

## 4. Dashboard Report and Insights

### 4.1 Overview of the Dashboard

The Power BI dashboard created for the Amazon Sales Analysis project provides a comprehensive view of sales performance across various dimensions, such as time, geography, and product categories. The dashboard is designed to be visually appealing with an interactive blue-themed layout, offering actionable insights through intuitive visualizations and key performance indicators (KPIs).

### 4.2 Key Visualizations

1. **Yearly and Monthly Sales Trends:**
  - **Line Charts:** Display the sales performance over the years (2017, 2018, and 2019) and provide insights into the growth or decline of revenue over time.
  - **Monthly Breakdown:** Highlight peak sales months and seasonal trends, aiding in inventory and sales strategy planning.
2. **Region and Country Sales Analysis:**
  - **Map Visualizations:** Represent sales distribution geographically, showing regions and countries contributing the most to revenue.
  - **Bar Charts:** Compare sales performance across regions and countries.
3. **Profit and Cost Metrics:**
  - **KPI Cards:** Display critical metrics like total sales, total profit, and profit margins in real time.
  - **Pie Charts:** Show the proportion of total cost versus total revenue, helping to identify areas of high profitability or cost inefficiencies.
4. **Product Category Insights:**
  - **Bar and Column Charts:** Highlight the performance of different item types (e.g., electronics, clothing) based on sales and profit.
  - **Filters:** Allow users to drill down into specific product classes or categories to identify top-performing items.
5. **Interactive Features:**
  - **Slicers and Filters:** Enable filtering by year, region, sales channel, and item type for a more tailored analysis.
  - **Drill-Through Pages:** Offer in-depth details for specific metrics or regions when selected.

### 4.3 Key Insights Derived from the Dashboard

### 1. Sales Trends Over Time:

- Sales have shown consistent growth from 2017 to 2019, with peak performance in Q4 of each year, suggesting seasonal shopping spikes.

### 2. Top Performing Regions and Countries:

- The **Central region** has the highest sales volume, while countries like **USA** and **Germany** drive the majority of revenue.

### 3. Product Categories Driving Revenue:

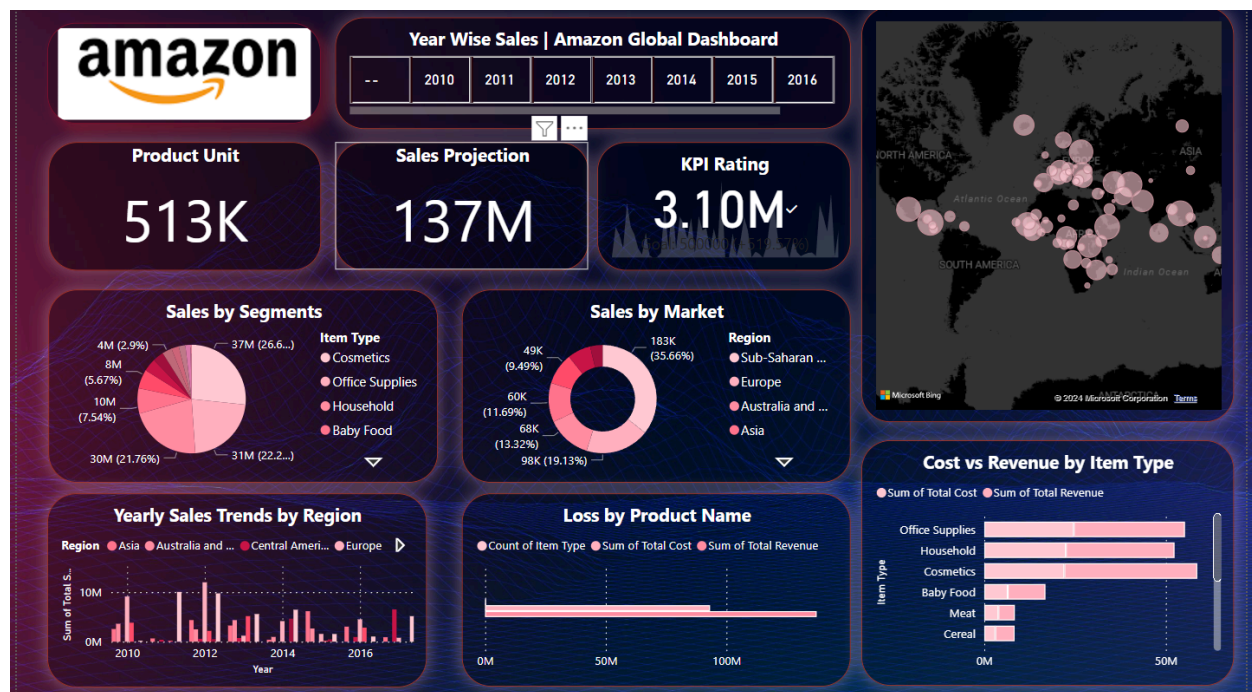
- **Electronics** and **Office Supplies** are the top-selling categories, contributing significantly to overall sales and profits.

### 4. Profitability Insights:

- High-margin items generate the most profit, while certain regions exhibit lower profitability due to higher costs or discounts.

### 5. Impact of Discounts:

- Discounted items drive significant sales volume but reduce profit margins, highlighting the need for balanced discount strategies.



#### 4.4 Recommendations Based on Insights

1. **Focus on High-Margin Products:**
  - Invest in promoting high-margin items to maximize profitability.
2. **Seasonal Sales Strategies:**
  - Increase marketing efforts and stock levels in Q4 to capitalize on peak sales seasons.
3. **Region-Specific Strategies:**
  - Address low-profit regions by analyzing cost structures and pricing strategies.
4. **Optimize Discounts:**
  - Identify the optimal discount percentage that balances sales volume and profitability.

#### 4.5 Future Enhancements

1. **Incorporate Additional Data Sources:**
  - Add data from recent years (2020 and beyond) to track trends post-2019.
  - Include customer feedback or reviews to analyze satisfaction and its impact on sales.
2. **Advanced Predictive Analytics:**
  - Use machine learning techniques to forecast future sales and customer behavior.
3. **Enhanced Visualizations:**
  - Introduce advanced Power BI visualizations like decomposition trees and smart narratives for deeper insights.
4. **Automated Reports:**
  - Set up scheduled updates to ensure the dashboard reflects real-time data.

## 5. Unit Test Cases

TEST CASE DESCRIPTION	EXPECTED RESULTS
Sales Trend Analysis (Yearly and Monthly)	When a line chart is interacted with, it shows a breakdown of yearly and monthly sales trends.
Region-wise Sales Performance	A map visual or bar chart displays sales across regions, dynamically updating based on filters.
Top Performing Products	Displays a bar chart listing the top products by sales or profit when interacted with.
Relation Between Discounts and Sales Quantity	A scatter plot shows the correlation between discount percentages and the corresponding sales quantity.
Profit Margin Across Regions	A pie or bar chart displays the profit margin percentage for each region.
Monthly and Yearly Profit Trends	A time-series line graph compares monthly and yearly profits over the three years (2017-2019).
Customer Distribution by Country and Division	A bar or column chart dynamically updates to show the distribution of customers by country and division.
Sales by Product Category	A column chart displays sales broken down by product categories (e.g., Electronics, Office Supplies).
Min, Max, and Average Sales Comparison	A bar graph shows the comparison of minimum, maximum, and average sales for various product categories.
Sales Channel Analysis	A pie chart displays the percentage of sales from online vs. offline channels, dynamically updating with filters.
Yearly and Monthly Profit Analysis by Region	A line chart displays profit trends across regions, segmented by year and month.
Relationship Between Sales and Order Priority	A bubble chart visualizes the relationship between order priority (e.g., High, Medium, Low) and sales figures.
Total Profit vs. Total Revenue Across Regions	A stacked bar chart compares total profit and total revenue in each region.