SALES DATA ANALYSIS

POWER BI PROJECT

SANJANA DALVI

**Index**

1. **Introduction**

* Introduction to the Dataset
* Objective of the Analysis and Dashboard Creation

1. **Data Sources**
   * Source Description
   * Data Extraction Process
   * Data Quality and Preparation
2. **Data Model**
   * Schema Design
   * Relationships
   * Data Transformation
3. **Visualizations**
   * Dashboard Design
   * Key Metrics and KPIs
   * Interactive Elements
4. **Conclusion**
   * Summary of Results

**Introduction to the Dataset**

This document presents an analysis of a detailed sales dataset designed to provide insights into various aspects of sales performance and customer interactions. The dataset is divided across three pages, each focusing on different dimensions and attributes relevant to the sales analysis.

The first page of the dataset contains core sales data, including a range of variables essential for understanding sales transactions. It features columns such as Row ID, Order ID, Order Date, and Ship Date, which track the timeline and specifics of each order. Shipping details are captured under Ship Mode and Shipping Cost, while customer information is detailed through Customer ID, Customer Name, Segment, and location fields including City, State, Country, and Postal Code. The page also includes Market, Region, Product ID, Category, Sub-Category, and Product Name, providing a comprehensive view of the products and their classifications. Key financial metrics such as Sales, Quantity, Discount, Profit, and Order Priority are also included, offering a full picture of the sales performance and customer interactions.

The second page focuses on returned orders, capturing critical information regarding order returns. It includes a Returned column that indicates whether an order was returned, alongside Order ID and Market. This page is crucial for analyzing the return rates and understanding their impact on overall sales performance and customer satisfaction.

The third page provides additional details related to individuals and geographical regions. It includes columns for Person, which contains information related to individuals involved with the orders, and Region, which provides further geographical context within the market. This page helps in understanding the human and regional factors influencing sales and customer interactions.

Overall, this dataset is structured to offer a comprehensive view of sales data, customer details, and order returns. The analysis will leverage these insights to identify trends, assess performance, and uncover opportunities for improving sales and operational efficiency.

**Objectives of the Analysis and Dashboard Creation**

The primary objectives of the analysis and dashboard creation for this sales dataset are to provide a clear, actionable understanding of sales performance, customer behavior, and operational efficiency. Here’s a detailed look at the goals:

1. Enhance Sales Performance Insights
2. Understand Customer Behaviour
3. Optimize Operational Efficiency
4. Identify Market and Regional Trends
5. Visualize Key Metrics and KPIs
6. Improve Decision-Making
7. Monitor and Track Performance

**Data Sources and Preparation**

**Source Description**

The sales dataset is sourced from a comprehensive collection of transactional data across various business units. It includes detailed records of sales transactions, customer interactions, and product information. The primary source of this data is the company’s sales management system, which aggregates information from online and offline sales channels. Additional data is obtained from customer relationship management (CRM) systems, shipping and logistics platforms, and market research databases. The dataset comprises multiple pages, each focusing on different aspects: core sales transactions, return information, and individual and regional details.

**Data Extraction Process**

The data extraction process involves several key steps to ensure that the dataset is accurately and efficiently gathered from its sources:

1. **Data Collection**: Data is retrieved from various sources such as sales databases, CRM systems, and logistics platforms. This involves exporting data files (e.g., CSV, Excel) or connecting to databases via queries.
2. **Data Consolidation**: Extracted data is consolidated into a single repository to facilitate analysis. This may involve merging files from different sources and ensuring that all relevant fields are included.
3. **Data Validation**: Initial checks are performed to ensure that the extracted data is complete and accurate. This includes verifying that all necessary fields are present and that the data adheres to predefined formats and standards.
4. **Data Integration**: Data from different pages or sources is integrated into a unified format suitable for analysis. This includes aligning data structures and ensuring consistency across the dataset.

**Data Quality and Preparation**

Ensuring high data quality and preparing it for analysis involves several critical steps:

1. **Data Cleaning**: Identifying and rectifying errors or inconsistencies in the dataset. This includes handling missing values, correcting incorrect entries, and removing duplicates.
2. **Data Transformation**: Converting raw data into a format suitable for analysis. This involves normalizing values, creating derived fields (e.g., calculating total sales), and categorizing data (e.g., segmenting customers).
3. **Data Validation**: Performing further checks to ensure the accuracy and reliability of the data. This involves comparing data against source records, conducting consistency checks, and validating calculations.

**Data Model**

**Schema Design**

The schema design represents the structured layout of the dataset, detailing how different data elements are organized and related. For this sales dataset, the schema typically includes the following components:

* **Tables**:
  + **Sales**: Contains core transactional data such as Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Product ID, Sales, Quantity, Discount, Profit, and Shipping Cost.
  + **Customers**: Includes customer-related details like Customer ID, Customer Name, Segment, City, State, Country, and Postal Code.
  + **Products**: Contains information about products including Product ID, Product Name, Category, and Sub-Category.
  + **Returns**: Tracks returned orders with columns such as Order ID and Returned status.
  + **Regions**: Provides geographical context with Region and Market details.
  + **Individuals**: Includes information related to individuals associated with the orders.
* **Attributes**:
  + **Dimensions**: Customer, Product, and Time dimensions to categorize and filter data.
  + **Measures**: Sales, Profit, Quantity, Discount, and Shipping Cost as quantitative metrics to be analyzed.

**Relationships**

Defining relationships between tables ensures that data can be accurately linked and analyzed across different dimensions. Key relationships include:

* **Sales to Customers**: Linked via Customer ID, allowing analysis of sales performance by customer demographics.
* **Sales to Products**: Linked via Product ID, enabling product-level analysis of sales, profitability, and inventory.
* **Sales to Returns**: Linked via Order ID, facilitating the analysis of returned orders and their impact on overall sales and profit.
* **Sales to Regions**: Linked via Market and Region, providing insights into geographical performance and trends.
* **Sales to Individuals**: Linked via relevant identifiers to include personnel data associated with orders.

These relationships enable comprehensive querying and reporting, allowing for multidimensional analysis of sales data.

**Data Transformation**

Data transformation involves modifying and processing raw data to make it suitable for analysis. Key transformation steps for this dataset include:

* **Data Aggregation**: Summarizing data at different levels, such as total sales by region or average profit per product category.
* **Calculated Columns**: Creating new fields derived from existing data, such as Total Sales (Sales \* Quantity) or Profit Margin (Profit / Sales).
* **Data Normalization**: Standardizing data formats and units to ensure consistency, such as converting dates to a standard format or normalizing product names.
* **Data Filtering**: Removing or excluding irrelevant data to focus on specific segments or time periods.
* **Data Integration**: Combining data from different sources into a unified format, ensuring that all relevant information is available for analysis.

**Visualization**

**Dashboard Design**

1. **Layout and Structure**: Design a clear and intuitive layout. Use a grid system to organize content, ensuring that the most important information is prominently displayed.
2. **Color Scheme**: Choose a color palette that is both visually appealing and functional. Ensure that colors are used consistently to represent the same type of data across different charts.
3. **Data Storytelling**: Arrange elements to guide the user through the data story. Start with high-level metrics and allow users to drill down into more detailed data.

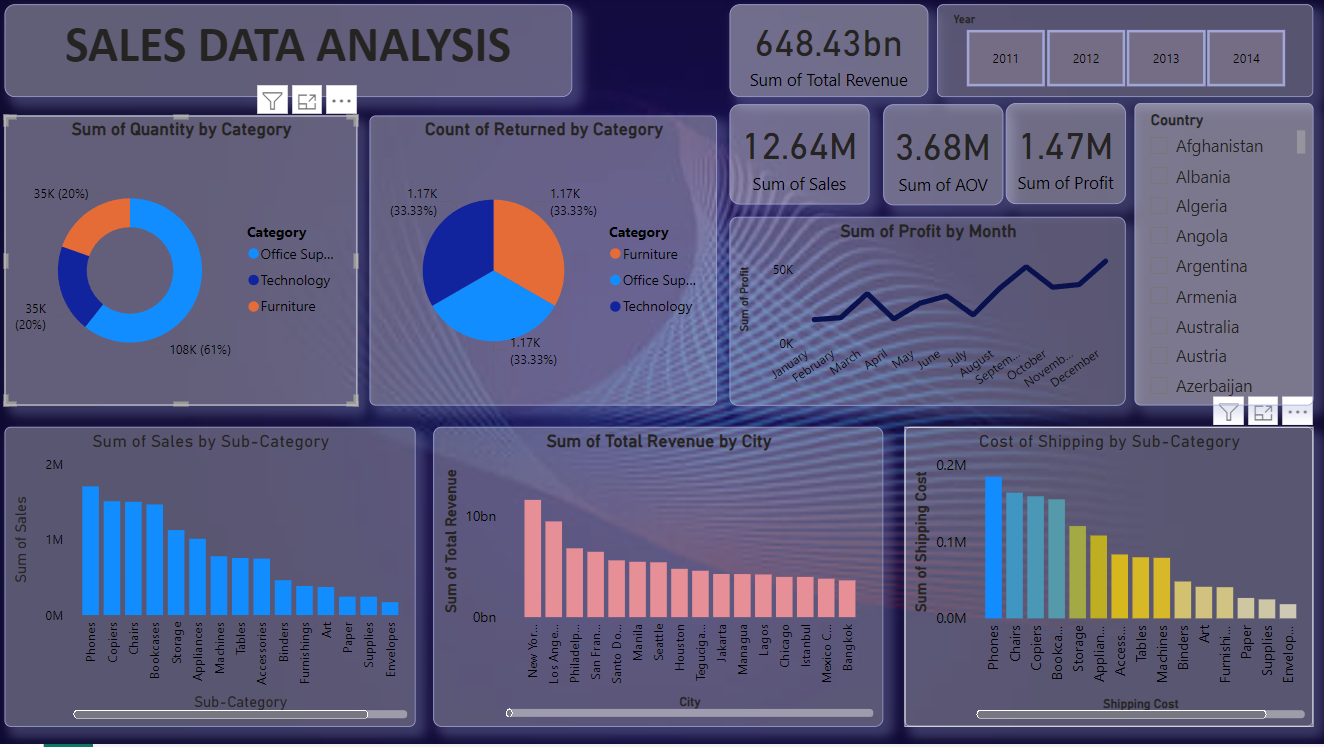
**Key Metrics and KPIs**

1. **Sales Performance**: Track metrics such as Total Sales, Sales Growth Rate, and Sales Targets vs. Actuals.
2. **Customer Metrics**: Include KPIs like Customer Acquisition Cost (CAC), Customer Lifetime Value (CLV), and Retention Rates.
3. **Product Metrics**: Focus on metrics like Product Sales Volume, Return Rates, and Product Profit Margins.

**Interactive Elements**

1. **Filters**: Allow users to filter data by different dimensions such as time period, region, or product category.
2. **Drill-Downs**: Enable users to click on summary metrics to view more detailed breakdowns or related data.
3. **Tooltips**: Provide additional context or explanations for data points when users hover over or click on them.
4. **Dynamic Charts**: Use charts that update in real-time based on user selections or interactions, helping to explore different scenarios or trends.

**Dashboard Overview**

****

* **Title**: "Sales Data Analysis"
* **Objective**: Provide insights into sales performance, profit trends, product returns, and operational costs.
* **Time Frame**: Data filtered for the years 2011 to 2014.

**2. Key Metrics**

* **Total Revenue**: 648.43bn
* **Sum of Sales**: 12.64M
* **Sum of Average Order Value (AOV)**: 3.68M
* **Sum of Profit**: 1.47M

**3. Visualizations & Insights**

* **Sum of Quantity by Category**:
  + **Technology**: 108K (61%)
  + **Office Supplies**: 35K (20%)
  + **Furniture**: 35K (20%)
  + **Insight**: Technology is the dominant category, accounting for more than half of the total quantity sold.
* **Count of Returned Items by Category**:
  + **Furniture, Office Supplies, Technology**: 1.17K each (33.33%)
  + **Insight**: Return rates are evenly distributed across all categories, indicating a consistent quality issue or customer behavior.
* **Sum of Profit by Month**:
  + **Insight**: Monthly profit shows fluctuation, with notable peaks in specific months, indicating possible seasonal effects or promotional periods.
* **Sum of Sales by Sub-Category**:
  + **Top Sub-Categories**: Phones, Copiers, Chairs, Bookcases
  + **Insight**: Phones and Copiers lead in sales, indicating high demand or market dominance in these sub-categories.

**Sum of Total Revenue by City**:

* **Top Cities**: New York, Los Angeles, Chicago
* **Insight**: Revenue is concentrated in major metropolitan areas, highlighting key markets for the business.
* **Cost of Shipping by Sub-Category**:
* **Top Costs**: Phones, Chairs, Copiers
* **Insight**: Shipping costs are highest for Phones and Chairs, likely due to their size, weight, or fragility.

**Conclusion**

In summarizing your analysis, you’ll want to provide a concise overview that encapsulates the key points and actionable takeaways. Here’s how to craft an effective conclusion:

1. **Recap Key Findings**: Briefly restate the most critical insights derived from your data analysis. This reminds your audience of the core conclusions without going into too much detail.
2. **Reinforce Trends and Patterns**: Summarize the main trends and patterns observed. Highlight any significant changes or observations that could impact decision-making.
3. **Summarize Recommendations**: Restate the primary recommendations based on your findings. Ensure they are actionable and aligned with the key issues identified during your analysis.
4. **Emphasize the Impact**: Explain how implementing your recommendations could benefit the organization or project. This could include potential improvements in performance, efficiency, or profitability.
5. **Call to Action**: End with a clear call to action, urging stakeholders to take the next steps or implement the suggested changes. Provide a timeline or outline the necessary actions to facilitate follow-through.

**Thank You**