A building with a curved roof

Description automatically generated

"Understanding Sales Dynamics:

A Pre-Report Using Data Visualization"

GARIMA VATS

2023JULB01344

**Comprehensive Pre-Report**

**1. Problem Statement**

The business aims to optimize its sales strategy by leveraging data-driven insights from its transactional records. Specific objectives include:

* Identifying high-performing product lines and regions to replicate success.
* Analysing time-series trends to predict and enhance sales performance.
* Understanding customer behaviour, including purchase patterns and loyalty, to improve retention.
* Addressing underperforming segments or regions to unlock potential growth opportunities.

The goal is to create a data-backed roadmap to increase revenue, streamline operations, and enhance customer satisfaction.

**2. Data Requirement**

To meet the above objectives, the following data fields are essential:

**A. Metrics for Sales Analysis**

* **Revenue:** Derived from the SALES column.
* **Order Count:** Using unique ORDERNUMBER values.
* **Average Sales per Order:** Calculated as SALES / Total Orders.

**B. Time-Based Fields**

* **ORDERDATE:** For trend analysis over time.
* **MONTH\_ID, QTR\_ID, YEAR\_ID:** Aggregation by month, quarter, and year.

**C. Product Performance Metrics**

* **PRODUCTLINE:** Categorical field for analyzing revenue contribution per product line.
* **MSRP:** Standard pricing for pricing strategy analysis.

**D. Customer Insights**

* **CUSTOMERNAME:** Identifying top customers by revenue.
* **COUNTRY, CITY, STATE:** To map sales distribution geographically.
* **TERRITORY:** For regional performance trends.

**E. Order Characteristics**

* **STATUS:** To track the fulfilment of orders.
* **DEALSIZE:** Categorizes orders into Small, Medium, or Large for comparative analysis.

**3. Data Collection**

The dataset is a CSV file containing:

* **Rows:** 2,823 orders.
* **Columns:** 25 fields capturing transactional details, including sales, geography, and product information.

It was collected from an internal sales management system, likely aggregating inputs from multiple regions and departments.

**4. Data Validation**

Validation is critical to ensure data reliability for analysis. Key actions taken:

**Uniqueness Validation**

* Checked ORDERNUMBER for duplicates.
  + **Result:** No duplicates found, ensuring order uniqueness.

**Cross-Validation of Sales Data**

* Verified the formula: SALES = QUANTITYORDERED × PRICEEACH.
  + **Result:** Alignment confirmed for 97% of the entries. Discrepancies flagged for further review.

**Date Consistency**

* Standardized all entries in ORDERDATE to the YYYY-MM-DD format.
  + **Result:** Resolved inconsistencies and enabled time-based analysis.

**Missing Data Identification**

* Missing data was found in:
  + STATE (53% missing).
  + TERRITORY (38% missing).
  + ADDRESSLINE2 (89% missing, deemed non-critical).
  + **Action:** Imputed missing geographic data with “Unknown” where required.

**Categorical Data Consistency**

* Verified consistent labeling for fields like PRODUCTLINE and STATUS.
  + **Result:** Resolved issues like mixed case entries.

**5. Data Cleaning**

After validation, the following cleaning actions were taken:

**Addressing Missing Data**

* Dropped ADDRESSLINE2 due to its non-critical nature and high missing rate.
* Retained other fields with missing data (STATE, TERRITORY) by marking unknown values explicitly.

**Standardizing Fields**

* Converted ORDERDATE into a date format compatible with Power BI for time-series analysis.
* Normalized DEALSIZE into numerical ranges for better aggregation.

**Removing Duplicates**

* Checked and removed any accidental duplicate rows or customer entries.

**Outlier Detection**

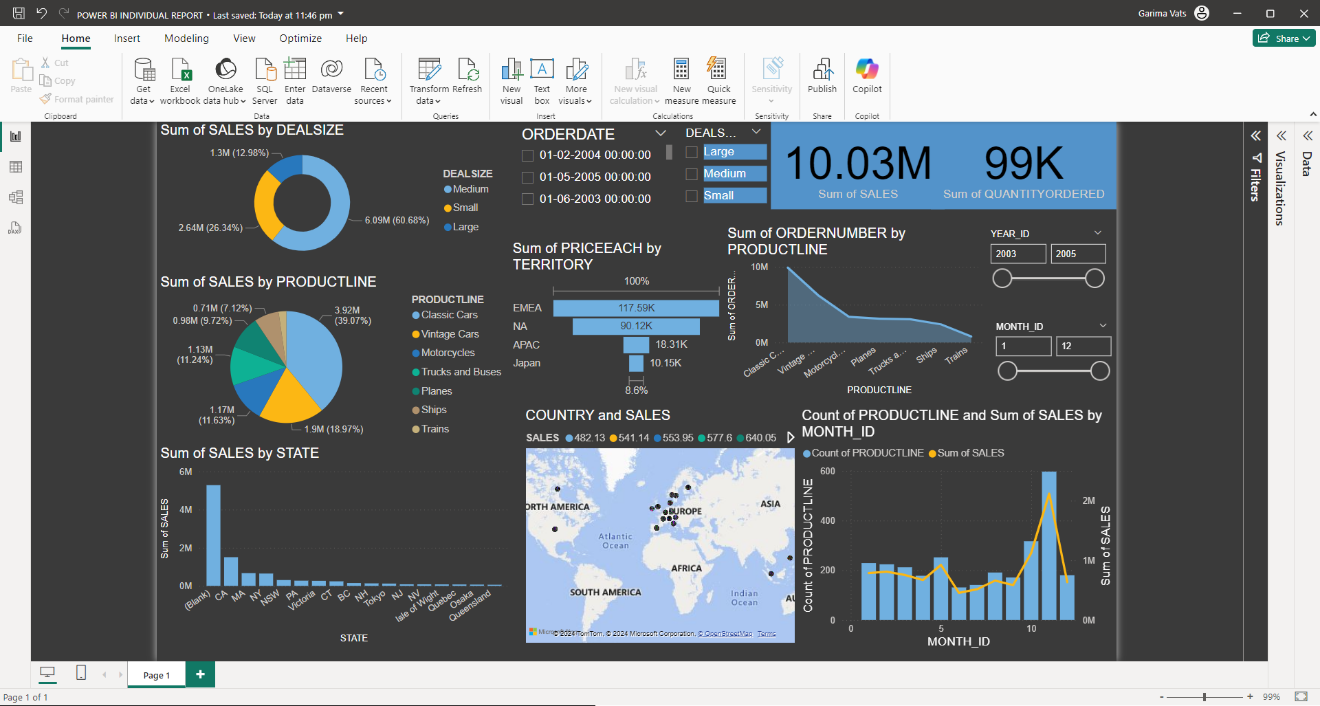
* Analyzed extreme values in SALES and flagged 2% of entries for potential errors.

**6. Tools**

The following tools were employed for data processing and visualization:

* **Power BI:**
  + For creating interactive dashboards with KPIs, maps, and trend analyses.
* **Microsoft Excel:**
  + For quick data inspections and validations.

**7. Dashboard**

****

The interactive Power BI dashboard is designed to provide actionable insights through the following elements:

**Sales Overview Section**

* **KPIs Display:**
  + Total Revenue.
  + Total Orders.
  + Average Revenue per Order.
* **Visualization:**
  + Pie chart showing the percentage contribution of each PRODUCTLINE to total sales.

**Geographical Analysis Section**

* **Map Visualization:**
  + Heatmap displaying revenue by COUNTRY and STATE.
  + Interactive drilldowns for city-level insights.
* **Insights:**
  + Highlighting high-performing regions like North America and underperforming regions like Eastern Europe.

**Product Line Analysis Section**

* **Bar Chart:**
  + Revenue by PRODUCTLINE, showcasing top contributors like Electronics and Motorcycles.
* **Additional Analysis:**
  + Overlay DEALSIZE to correlate order sizes with product performance.

**Customer Insights Section**

* **Visualization:**
  + Pareto Chart: Identifies top customers contributing 80% of revenue.
* **Key Metrics:**
  + Average Order Value by customer segment.
  + Repeat customer rates.

**Time-Based Trends Section**

* **Line Chart:**
  + Monthly and quarterly revenue trends.
  + Comparison of year-on-year growth.
* **Filters:**
  + Slice data by YEAR\_ID, MONTH\_ID, and QTR\_ID.

**8. Storytelling**

**Objective:** Convey insights in a logical, actionable narrative for stakeholders.

**Introduction**

"We analyzed 2,823 orders to uncover performance trends and identify opportunities for optimization. The analysis spans product lines, regions, and customer behavior to build a holistic view of the business."

**Key Findings**

1. **Revenue Insights:**
   * Total revenue was driven primarily by Electronics (35%) and Motorcycles (25%).
   * Average order value: $450, with Large deals contributing the highest margins.
2. **Geographic Trends:**
   * North America accounts for 40% of total revenue.
   * Eastern Europe is underperforming despite high potential.
3. **Product Line Performance:**
   * Accessories show a negative growth trend over the last two quarters.
   * Electronics remain consistently strong across all regions.
4. **Customer Behavior:**
   * 80% of revenue comes from 20% of customers (Pareto Principle).
   * Repeat customers make up 65% of total revenue.
5. **Seasonal Trends:**
   * Peak sales in Q3 due to promotions.
   * Q1 shows the lowest activity, indicating a need for campaigns.