Toy Company Sales Analysis

**Project Objective**

The objective of this project is to create an **interactive dashboard** with cleaned sales data using **R/Python**, **DAX**, and **Power BI** to fulfill business needs. The dashboard helps in identifying sales trends, product performance, and country-wise contributions.

**Dataset**

* The dataset was collected from **Kaggle** and is also provided in the Dataset folder.
* It contains transaction-level sales data, including order details, product line, country, status, and revenue information.

**Data Cleaning**

Data cleaning was performed using both **R** and **Python** scripts (available in the Code folder). The process ensured data quality and integrity by following these steps:

1. **Duplicate Handling**
   * Composite key (ORDERNUMBER + ORDERLINENUMBER) created.
   * Duplicate rows removed to ensure unique transactions.
2. **Validation of Numeric Fields**
   * Checked for invalid values in SALES, PRICEEACH, and QUANTITYORDERED.
   * Negative or zero values were flagged for review.
3. **Sales Integrity Check**
   * Confirmed that SALES = QUANTITYORDERED × PRICEEACH.
   * Discrepancies corrected by recalculating SALES.
4. **Date Validation**
   * Standardized ORDERDATE into proper date format.
   * Extracted YEAR, MONTH, and QUARTER from ORDERDATE.
   * Validated against dataset fields (YEAR\_ID, MONTH\_ID, QTR\_ID).
5. **Missing Values Assessment**
   * Summarized missing data for further inspection.
6. **Standardization of Categorical Fields**
   * Standardized STATUS values into a consistent title case format.
7. **Column Selection**
   * Final dataset retained only essential fields:  
     SALES, STATUS, QUANTITYORDERED, QTR\_ID, MONTH\_ID, YEAR\_ID, PRODUCTLINE, COUNTRY, DEALSIZE.
8. **Output**
   * Final cleaned dataset saved as sales\_data\_cleaned.csv.

**Outcome**: A reliable and consistent dataset free from duplicates, inconsistencies, and formatting errors — ready for analysis and visualization.

**Data Analysis**

* The cleaned dataset was imported into **Power BI** for analysis.
* **Matrix visualizations** were used to summarize data, providing a complete overview of sales performance across years, quarters, products, and countries.
* **DAX Measures** were created to compute:
  + **Average Order Value (AOV)** → Helps evaluate transaction efficiency.
  + **Top Country in Sales** → Identifies the region with highest revenue contribution.

**Data Visualization**

The Power BI dashboard was designed to provide both a **holistic view** and **deep insights** into sales data:

1. **Time-Series Analysis**
   * Line and column charts visualize sales trends by **year and quarter**.
   * Highlights seasonality and growth patterns.
2. **Product Performance**
   * Bar/column charts show sales by **product line**.
   * Helps identify top-performing product categories.
3. **Geographical Insights**
   * Map visualization displays sales by **country**.
   * Highlights regions with highest revenue generation.
4. **Sales Status Tracking**
   * Pie/donut chart of STATUS field to analyze order fulfillment (Shipped, Cancelled, On Hold, etc.).
5. **KPI Cards**
   * Key performance indicators (KPIs) like **Total Sales**, **Average Order Value**, and **Top Country** displayed for quick reference.

Images of this dashboard is presented in Dashboard\_page-1&2 .PNG file.

**Business Objective**

The core business objectives achieved through this project are:

* **Summarize sales by year and quarter** to track performance over time.
* **Identify top-selling product lines** to optimize product focus and marketing strategies.
* **Determine the country with the highest sales** to prioritize regional efforts.
* **Improve decision-making** by providing interactive and visual insights into company sales performance.

### 1. ****Summarize sales by year and quarter****

* **2003**: Q1 = 372k, Q2 = 473k, Q3 = 527k, Q4 = **1.53M**
* **2004**: Q1 = 693k, Q2 = 639k, Q3 = 921k, Q4 = **1.66M**
* **2005**: Q1 = **889k**, Q2 = 590k

Sales peaked in **Q4 of each year** (holiday season effect).  
The highest sales overall occurred in **Q4 2004 (1.66M)**.

### 2. ****Top-selling product lines****

* **Classic Cars** → **2.97M** (highest revenue generator)
* **Vintage Cars** → 1.64M
* **Motorcycles** → 971k
* **Trucks and Buses** → 947k
* **Planes** → 878k

**Classic Cars** are the clear market leader, followed by **Vintage Cars**.

### 3. ****Country with the highest sales****

* **USA** → **2.99M** (top country)
* **Spain** → 1.02M
* **France** → 919k
* **Australia** → 522k
* **UK** → 413k

**USA is the biggest contributor** to sales by a large margin.

# Summary

* **Quarterly Trend**: Sales spike every Q4, showing strong seasonality.
* **Top Product Line**: Classic Cars dominate total sales, making them the core revenue driver.
* **Top Country**: USA leads all other markets, followed by Spain and France.

These insights can guide the toy company to:

* Focus marketing and stock strategies heavily on **Classic Cars**.
* Prepare for **Q4 seasonal demand** with promotions and inventory planning.
* Prioritize **USA** for expansion and retention, while developing Spain and France as secondary strongholds.