Project Title: Sales Data Analysis (Jan 2003 – May 2005)  
**Tools Used:**

* Python (Pandas, Matplotlib),
* Jupyter Notebook

**Objective**

This sales analysis project aimed to evaluate customer transactions from January 2003 to May 2005 using Python-based data wrangling and visualization tools. The dataset consisted of 2,823 records and 24 columns, including key variables such as sales amount, order date, product line, and customer information.

**Methodology**

* **Data Import and Cleaning:**
  + Imported dataset into Jupyter Notebook.
  + Removed unnecessary columns like ADDRESSLINE2 with excessive null values.
  + Filled or flagged missing values in STATE, TERRITORY, and POSTALCODE.
  + Converted object-type date fields to datetime format.
  + Standardized data types.
  + Removed outliers from the SALES column (values > 12,000).
  + Created a new REVENUE column using: PRICEEACH × QUANTITYORDERED.

**Key Findings**

* **Total Revenue:**  
  The business generated **$8,273,286.79** in total revenue.
* **Peak Sales Period:**
  + **November 2004:** $1,076,511.51
  + **November 2003:** $1,029,837.66  
    These months saw the highest revenue, indicating strong year-end sales.
* **Top Product Line:**
  + **Classic Cars** were the most frequently sold products, followed by **Vintage Cars** and **Motorcycles**.
* **Customer Insights:**
  + 92 unique customers were identified.
  + **Euro Shopping Channel** was the top customer with **$761,195.05** in total revenue.
* **Geographic Insights:**
  + The **EMEA (Europe, Middle East, and Africa)** region contributed the highest sales.
  + Followed by entries marked as “missing,” then **APAC** and **Japan**.
* **Order Status:**
  + **92.77%** of orders were successfully shipped.
  + Very few were cancelled, disputed, or on hold.
* **Price vs. Quantity Relationship:**
  + No strong correlation found between unit price and quantity ordered.
  + Suggests pricing had limited influence on order volume.

**Limitations**

* **Missing Data:**
  + Geographic fields such as STATE and TERRITORY had many missing entries.
  + Filled with default values, which may limit accuracy in regional insights.
* **Customer Records:**
  + Duplicate entries in customer names (>2,700 times) suggest the need for better customer ID systems.

**Conclusion**

This comprehensive sales analysis reveals strong year-end sales trends, most popular product lines, key customers, and dominant sales regions. With 92% on-time order fulfillment, the business shows supply chain efficiency.  
Future improvements should focus on enhancing data quality and utilizing insights for:

* Demand forecasting
* Customer segmentation
* Region-specific marketing and inventory planning