Technical Report Using Excel: Retail Dataset Analysis

INTRODUCTION

This is a report reviewing a retail dataset. I will analyze this dataset by giving my initial insights and detailed observations on the data.

The dataset is a sales data which includes various attributes. It comprises of 2824 entries and 25 columns, featuring numerical and categorical variables. The primary purpose of this review is to identify key patterns, trends, and anomalies within the data, as well as to outline potential areas for deeper analysis.

Initial Insights and Exploratory data analysis

a) ***Key variables and data types***

Numerical Variables: Order number, Quantity ordered, Price each, Order line number, sales, Quarter id, Month id, Year id, MSRP.

Categorical Variables: Order date, Status, Product line, Product code, Customer name, Phone, Addressline1, Addressline2, City, State, Postal code, Country, Territory, Contact last name, Contact first name, Deal size.

b) ***Exploratory data analysis***

To further explore this Retail dataset, these details were noticed in this data collection;

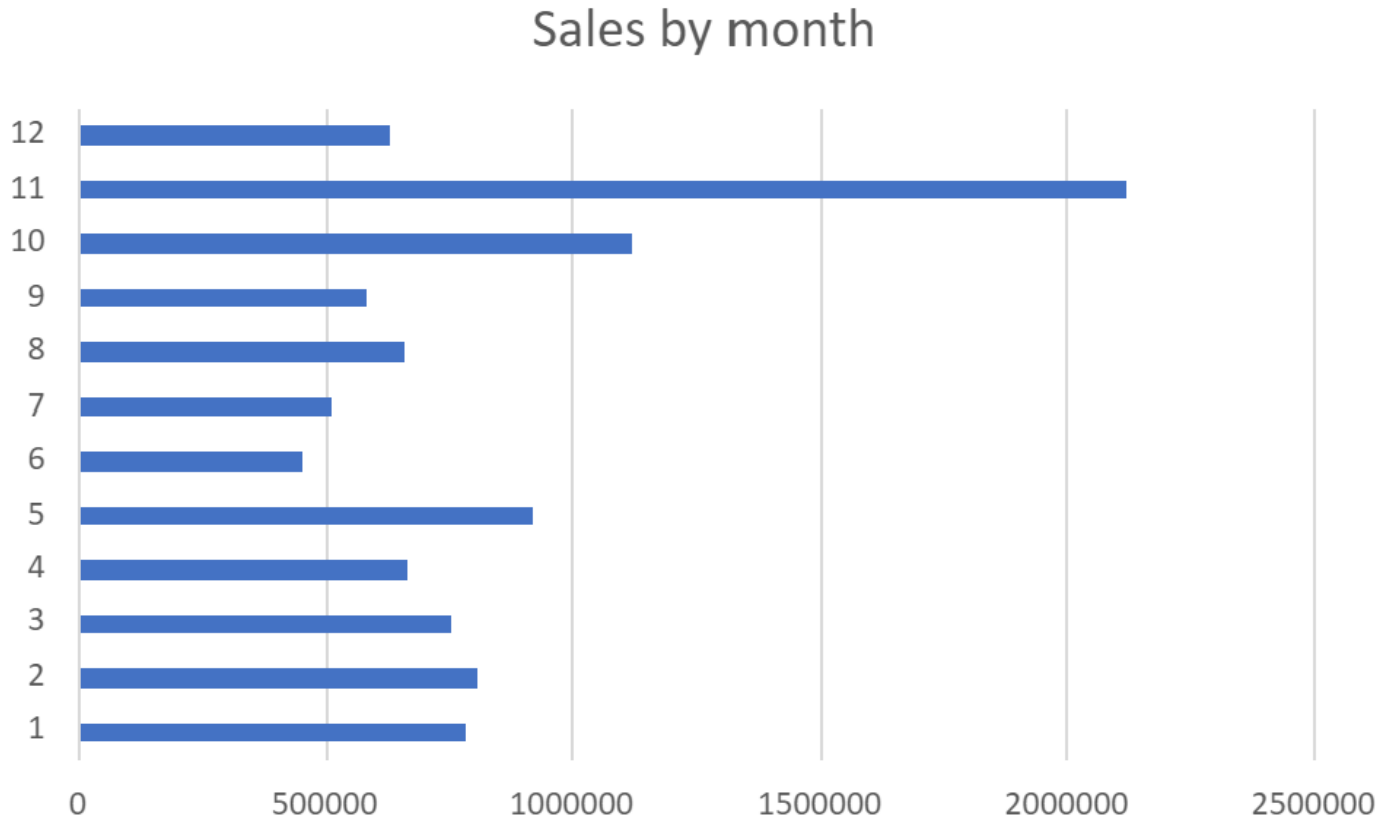
1. These columns have missing values Addressline2, State, Postal code, and Territory.
2. The Order date is in a string format and includes both date and time.
3. Columns like City, State, Country, and Territory give geographical information.
4. Some important variables related to sales are Quantity ordered, Price each, Deal size, sales, order status etc.
5. The status order are represented with Shipped, Cancelled, and Disputed, In Process, On Hold, and Resolved.
6. The products are automobiles; Classic Cars, Motorcycles, Planes, Ships, Trains, Trucks and Buses, Vintage Cars

OBSERVATIONS

Analysis: With the use of the pivot table, these observations were made and can be visualized using the charts below;

1. **Monthly sales performance**

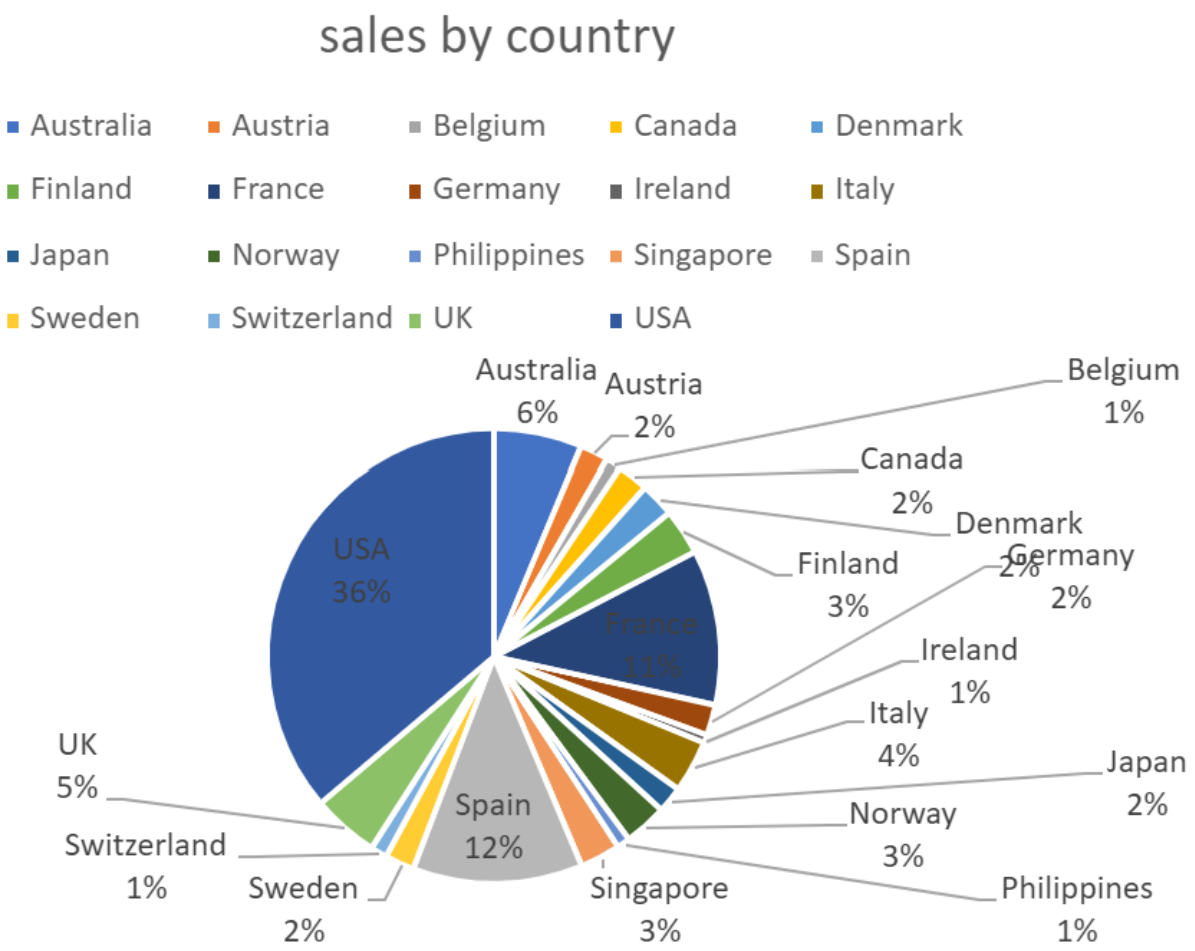
In this analysis results, this shows the general performance of the products monthly with the month of November recording the highest sales which was around $2.1 million, while in June the lowest sales was made.



Performance of the sales by months

2. **Total sales by country**

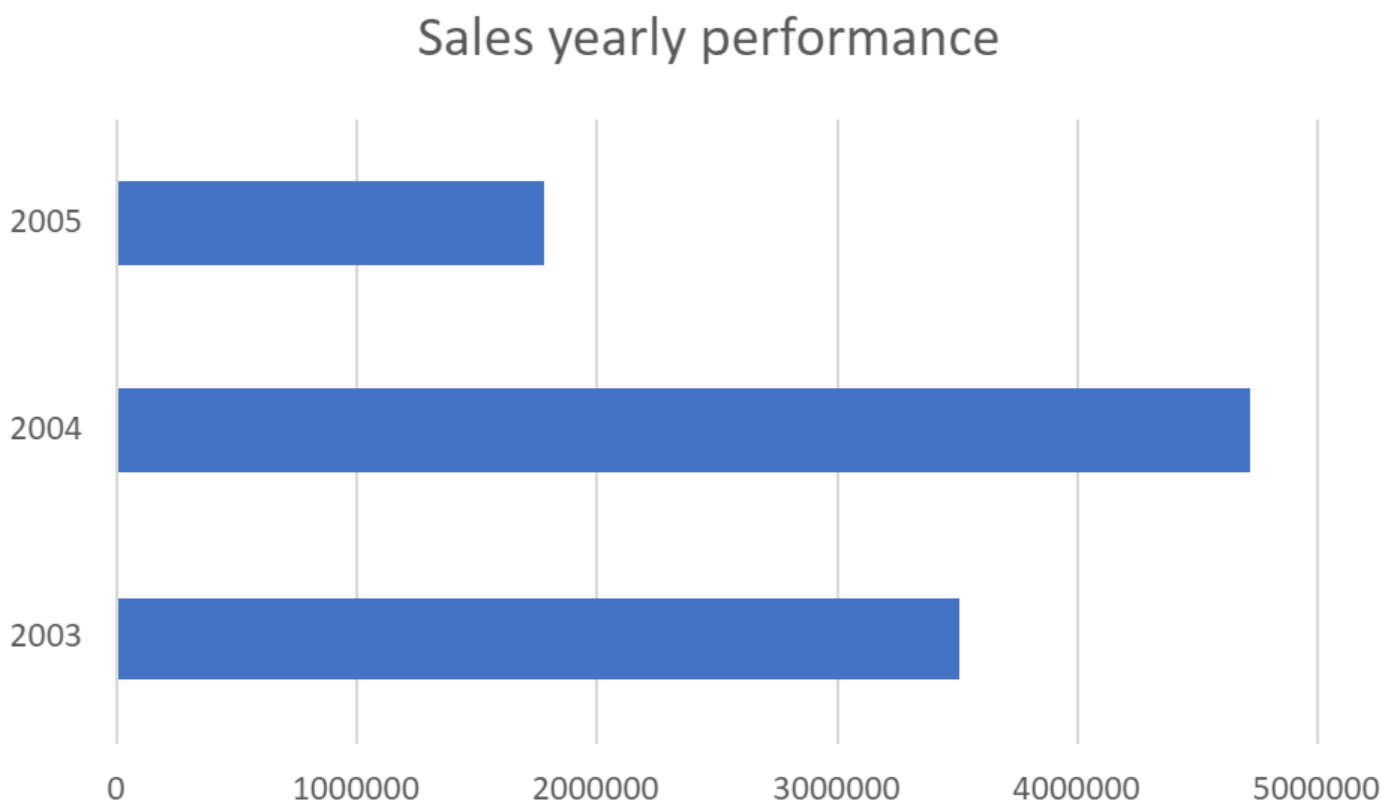
The product order are distributed in various countries while the USA, Spain and France have the most significant orders with the USA having the largest orders. Countries like Sweden, Ireland, Belgium, and Philippines record the least orders.



Country sales performance

3. Sales performance per year

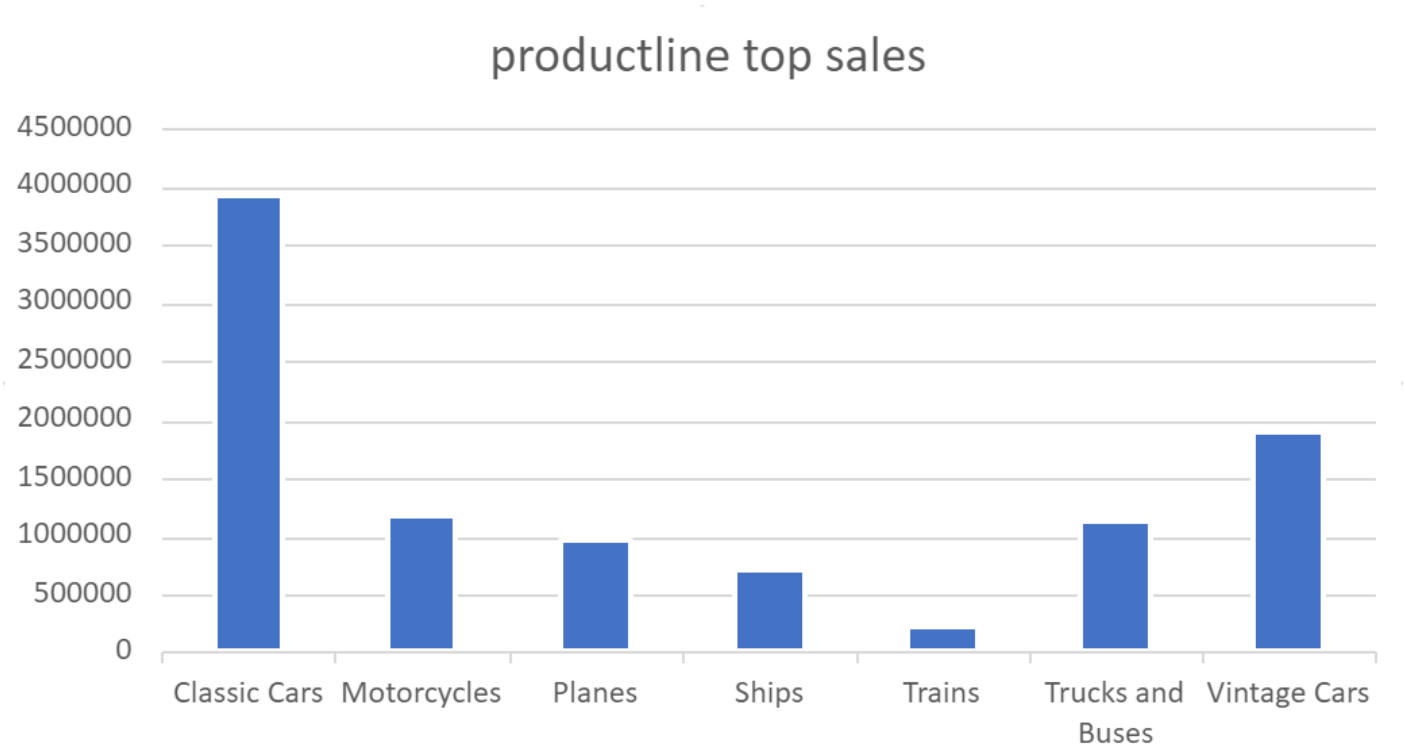
The sales pattern by year shows the growth of the sales from 2003–2004 where the year 2004 made the greatest contribution with a decline in the year 2005.



Sales per year

4. Product line sales performance

There is a significant demand for the classic cars followed by vintage cars, while the trains are in the least demand.



Sales by product

CONCLUSION

Summary

Reviewing this dataset, the initial insight shows the dataset with a range of numerical and category information. The anomalies in this data shows there are missing values in a number of columns which require more analysis. The sales in November indicate the the highest number of sales, the product line shows the classic cars bein the most ordered, The year of 2004 had the best year performance, and The country USA had the greatest sales contribution.

Further analysis

1. Carefully analyze the columns with missing data and update
2. Considering historical data, develop predictive models to forecast sales in the future.
3. Analyze how effectively various product lines and deal sizes are performing.
4. Split up your clientele according to their purchasing patterns and location.