**BMW Sales (2010–2024): Data Exploration Report:**

## **Executive summary**

The following analysis covers BMW sales from 2010 to 2024 using the dataset provided and the engineered attributes created in our project notebook. Regarding the analysis conducted, it illustrated that SUVs lead sales and continue to expand, as Hybrid and Electric powertrains gain strength, and Automatic transmissions steadily overtake Manual. Prices are broadly consistent across regions, with differences driven more by product mix than geography. The Covid-19 period appears as a break in the year-over-year (YoY) series during 2020–2021, followed by a strong recovery from 2022 led by SUVs and Hybrid models. The engineered features, such as Age\_Car, Car\_Classification, Period\_of\_Covid19, Price\_per\_KM, and Revenue, turn the raw dataset into a structured story about how product lines, technologies, and consumer preferences evolve over time.

## **Project context**

This work addresses the Sales Forecasting and Optimization project, which aims to use historical sales to build predictive models that support inventory planning, marketing strategies, and sales optimization. The present report focuses on the exploratory data analysis (EDA) stage. Insights gained here will guide the selection of features and modeling approaches in subsequent phases.

**Dataset and feature engineering**

The dataset includes around 50,000 rows, covering the attributes Model, Year, Region, Color, Fuel\_Type, Transmission, Engine\_Size\_L, Mileage\_KM, Price\_USD, Sales\_Volume, and Sales\_Classification. As part of our data cleaning process, we as a team checked for null values and duplicate entries, confirming that no duplicates or nulls were present before feature engineering. During feature engineering, the following additional variables were introducewd in the notebook:

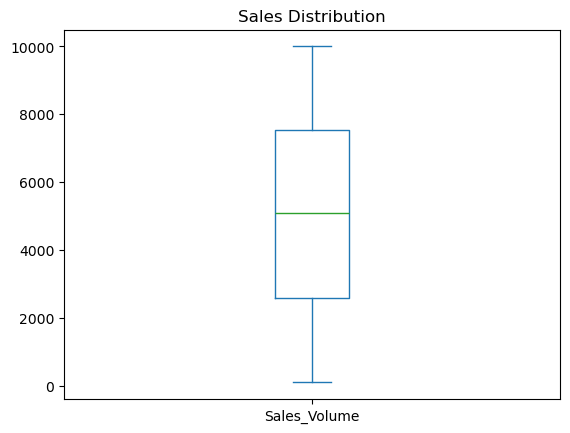
* Age\_Car measures the difference between the current year and the car year.
* Car\_Classification groups models into Series, Sport, SUV, and Other.
* Period\_of\_Covid19 flags the years 2020 and 2021 to isolate pandemic effects.
* Price\_per\_KM relates price to mileage, providing an efficiency perspective.
* Revenue is calculated as sales volume multiplied by price, grounding the analysis financially.  
  These engineered attributes proved critical in clarifying trends that the raw fields alone could not fully explain.

## **Data quality and preparation**

As illustrated in the notebook, we followed various data cleaning steps by removing duplicates and handling missing values. Categorical variables were standardized to avoid inconsistencies (for example, “manual” and “Manual” both mapped to Manual, “i8” and “I8” both mapped into I8). The distributions of numeric attributes fell within realistic ranges: Price\_USD spanned $30,000 to $120,000, Sales\_Volume ran from hundreds to nearly 10,000 units annually, and engine sizes ranged from small efficient builds to high-performance 5.0L models. The dataset is synthetic but balanced, designed to simulate realistic automotive market conditions.

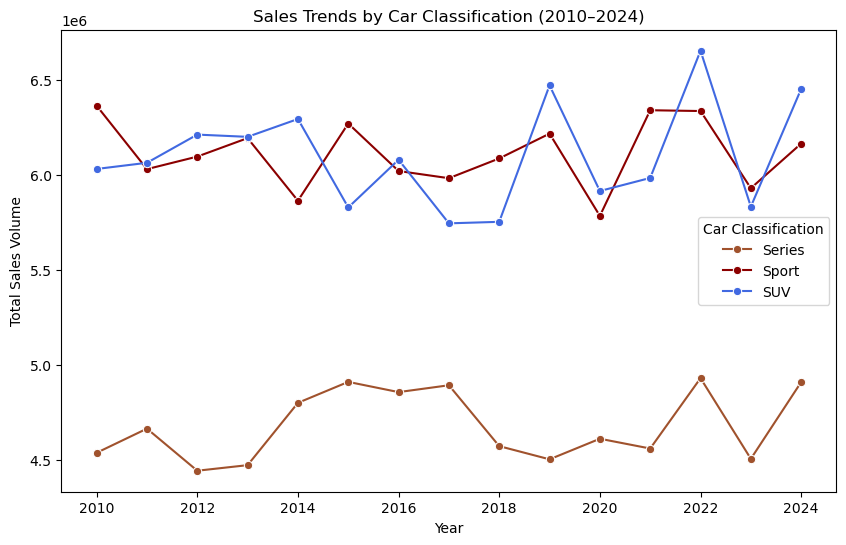
## **Sales Volume Distribution**

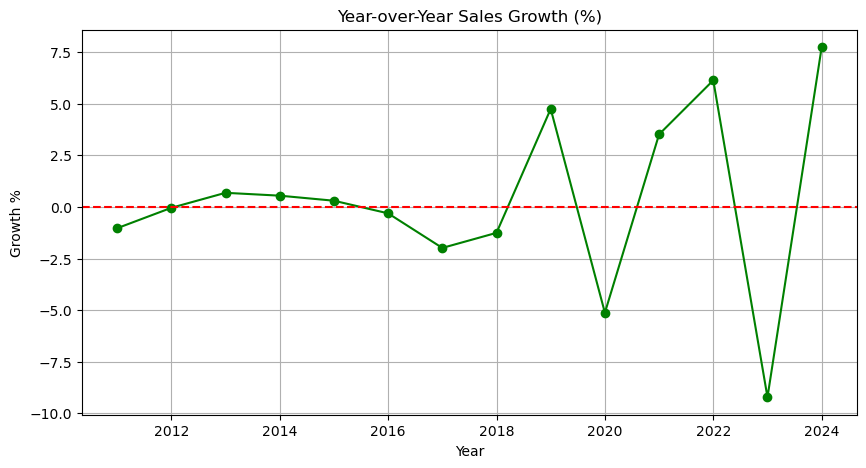
The boxplot of Sales\_Volume highlights how sales are spread across the dataset. Most values fall between about 2,000 and 8,000 units, with a median close to 5,000. A few very low points appear near zero, which likely represent edge cases such as demo or rare sales records. On the upper side, sales reach around 10,000 units, showing that while there is variation across models and regions, most results cluster in a consistent mid-range.



## **Time and YoY behavior**

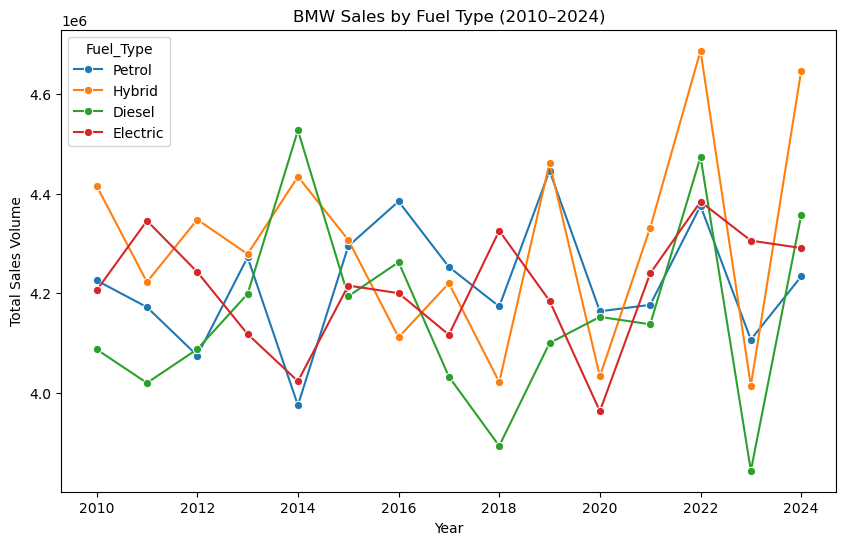
The analysis of Car\_Classification reveals the long-run shift in demand, it highlighted that SUVs dominate and increase their share steadily. While Series models maintain consistent volumes, Sport models remain niche but resilient. Moving on, the YoY growth series exposes the Covid-19 shock in 2020–2021, when growth turned negative across various categories. From 2022 onward, SUVs led the recovery with sharply positive YoY, confirming their role as a major growth factor of BMW’s portfolio.



**

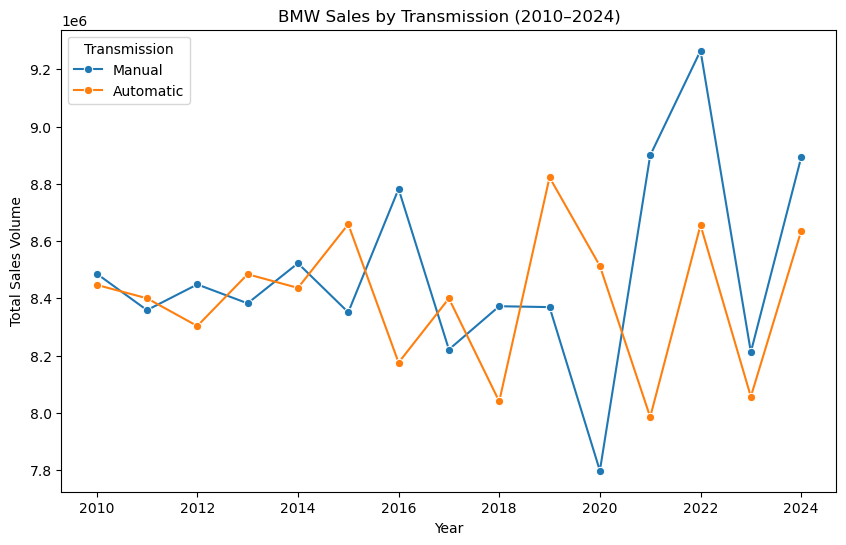
**Fuel mix evolution**

Trends in Fuel\_Type highlight BMW’s gradual shift toward sustainable vehicles. Petrol and Diesel dominated sales in the early years. Hybrid adoption began in the mid-2010s and grew consistently, while Electric vehicles appeared in measurable numbers only after 2020. Breaking down fuel trends by Car\_Classification shows that SUVs were early adopters of Hybrid, Series models followed with gradual adoption, and Sport cars remained mostly Petrol-based. YoY growth lines confirm Hybrid as the most resilient category.

**

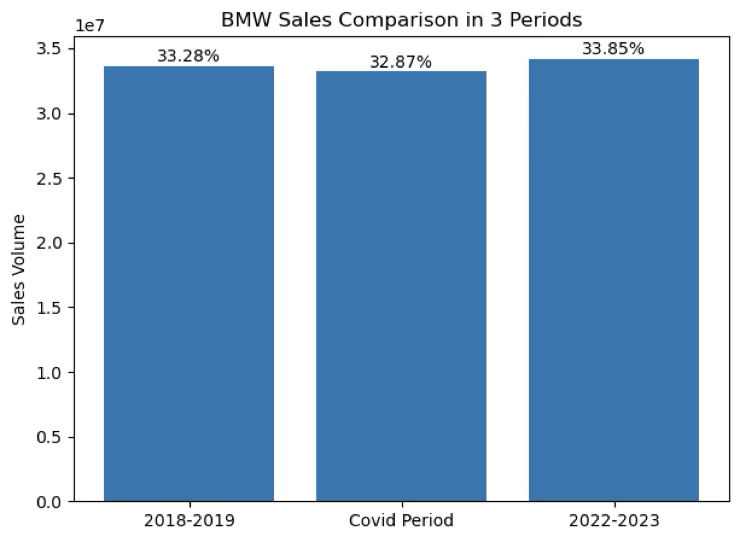
**Transmission trends**

Analysis of the Transmission attribute indicates a gradual but clear shift toward Automatic cars. From near parity in the early years, Automatic steadily overtook Manual, especially within SUVs. Sports cars still maintain a strong Manual presence, reflecting performance preferences. By integrating Age\_Car, it becomes clear that newer models are overwhelmingly Automatic, while older cars tilt Manual, showing lifecycle effects in consumer choice.



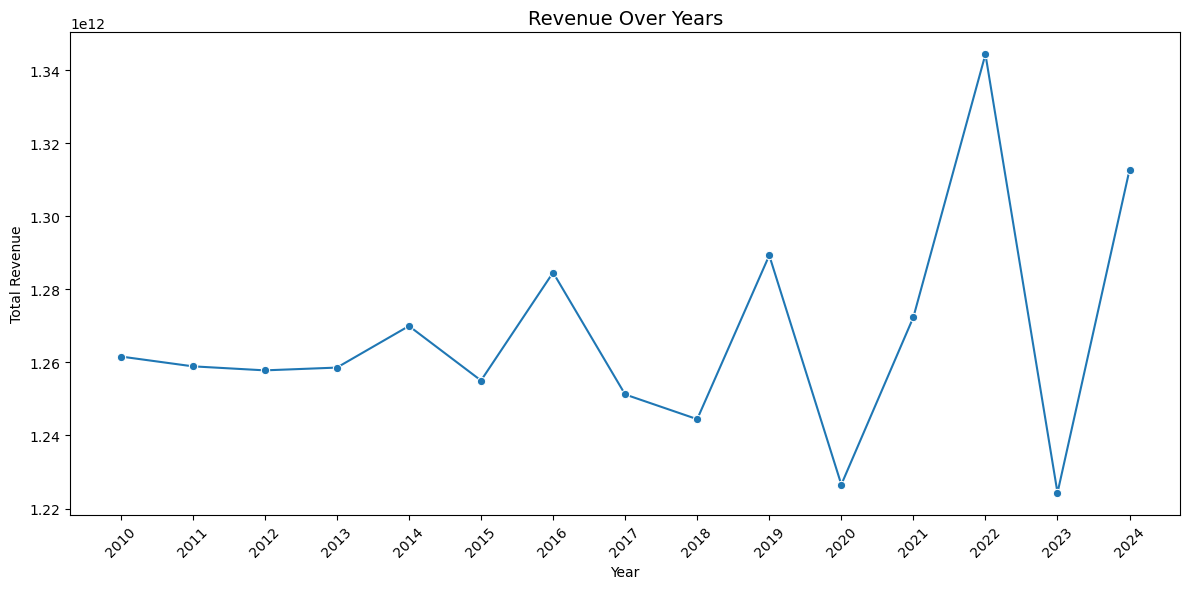
## **Period of Covid19**

During the Covid-19 period, BMW sales experienced a slight decline due to global supply chain disruptions and reduced consumer demand. However, the overall impact was limited, and sales quickly recovered in the following years.



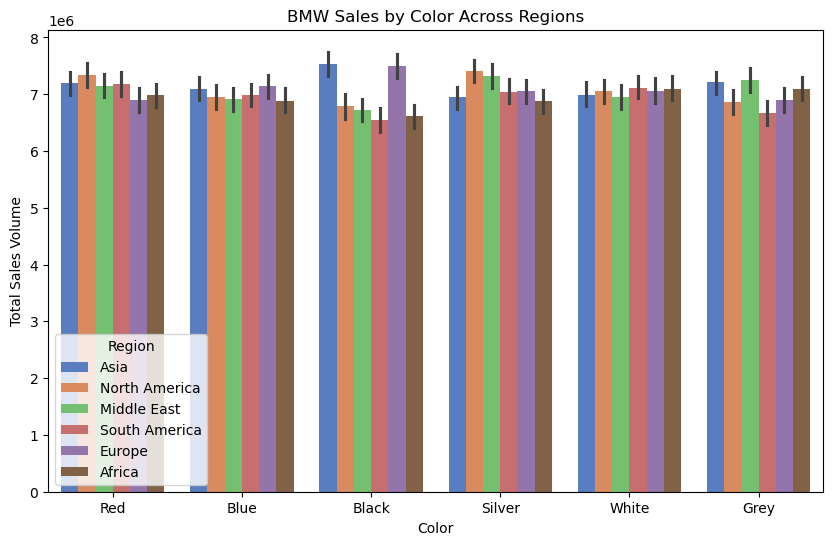
## **Revenue concentration**

The addition of Revenue shows that SUVs not only sell in larger volumes but also contribute the most financially. Even in regions where sales volumes are comparable, revenue differences appear when the mix shifts toward Series or luxury SUVs. This attribute aids the analysis in financial terms and underlines why SUVs are central to both sales forecasting and optimization.

**

## **Regional insights**

Regional analysis shows that Asia and Europe lean toward higher-end models, while South America and Africa skew more toward mid-range. The Middle East sits between the two. Medians across all six regions cluster closely, reinforcing BMW’s global pricing consistency. Color choices play a minor role, though Black is slightly favored in Europe and Asia and Grey in the Middle East. These variations are subtle and more relevant to localized marketing than to forecasting.



**Age\_Car, Mileage\_KM, and Engine\_Size\_L**

Age\_Car provides clearer insights than Mileage\_KM alone. Younger cars are typically associated with higher prices, Automatic transmissions, and Hybrid powertrains, while older cars lean towards Manual and Petrol.

**Seasonality, events, and stability**

At the annual level, classic month-to-month or quarter-to-quarter seasonality is not observable. The defining temporal event is the Covid-19 disruption, where Period\_of\_Covid19 highlights 2020–2021 as a distinct period of depressed sales. From 2022, the rebound aligns with the upward trend of SUVs and Hybrid, showing stability once the temporary shock is accounted for.

## **Limitations and notes**

The dataset is synthetic but designed to mimic real-world distributions. Some expected relationships, such as between mileage and price, appear weaker than in reality, but engineered fields such as Age\_Car, Price\_per\_KM, and Revenue provide structure and explanatory power. The use of annual data ensures robustness but limits the ability to analyze finer seasonal cycles.

## **Key findings**

The analysis shows that SUVs are the dominant and growing classification, especially the X1 class being the highest in terms of revenue in multiple years, Hybrid and Electric powertrains are the most important long-term growth segments, and Automatic transmissions are replacing Manual in mainstream adoption. Pricing is globally consistent across regions. Revenue analysis confirms that SUVs drive both volume and profit contribution. Color and engine size provide minimal explanatory power, while Age\_Car and Period\_of\_Covid19 are crucial for lifecycle and anomaly detection.

## **Conclusion**

The exploratory data analysis of BMW sales from 2010 to 2024 highlights clear structural drivers of demand. SUVs, Hybrid adoption, and the tilt toward Automatic transmissions define the portfolio’s direction. Covid-19 disruptions are identifiable and containable through engineered indicators. Pricing remains globally uniform, while revenue concentration highlights the financial weight of SUVs. The engineered variables as follows: Age\_Car, Car\_Classification, Period\_of\_Covid19, Price\_per\_KM, and Revenue—proved essential in highlighting patterns that raw features alone could not. Together, these insights create a strong foundation for the next stage of the project: building forecasting models that incorporate these structural drivers, handle shocks explicitly, and generate accurate, actionable predictions for BMW’s global sales strategy.