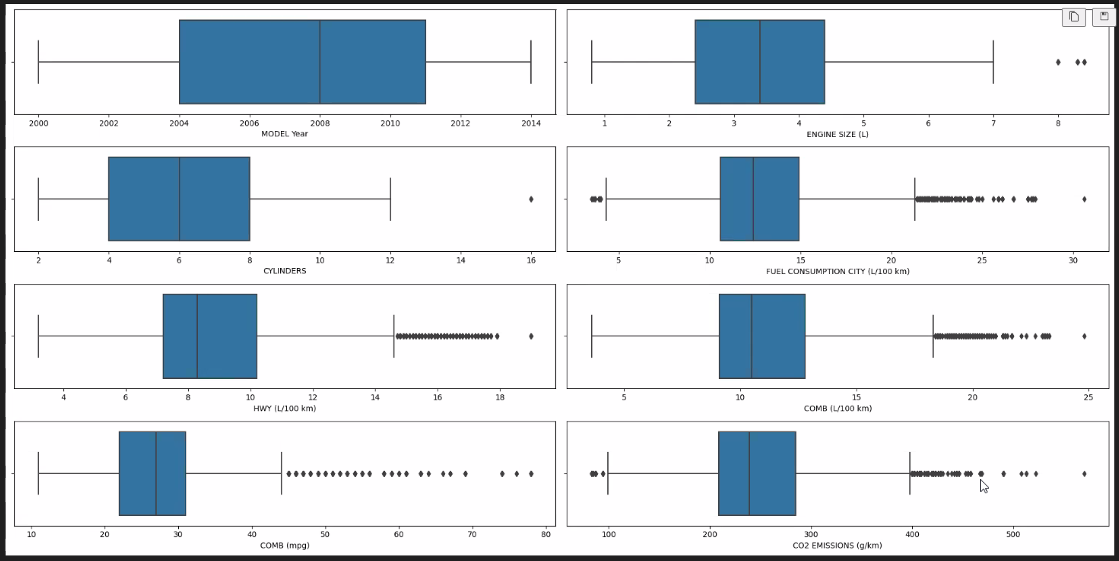
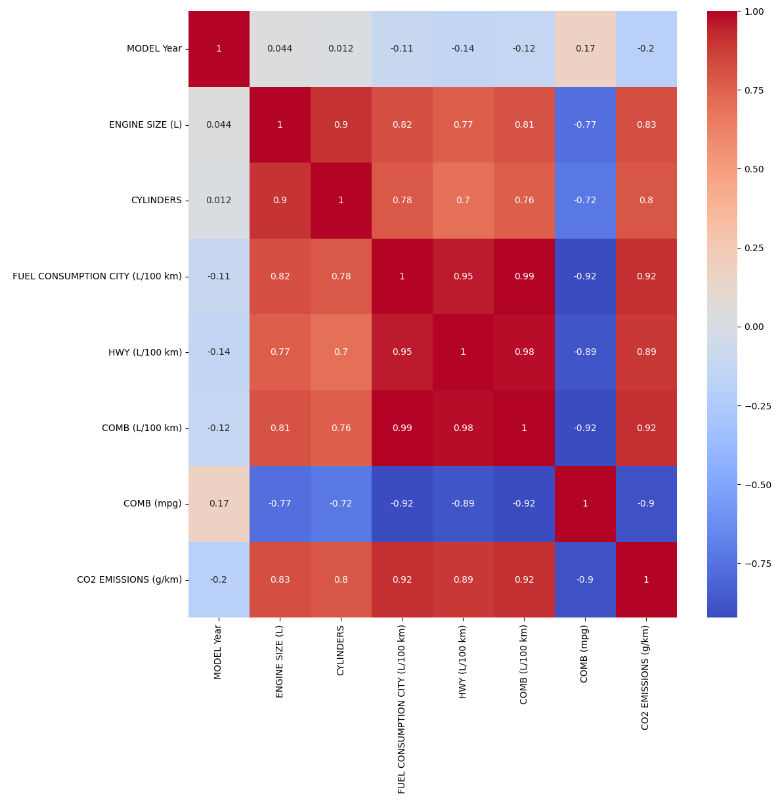
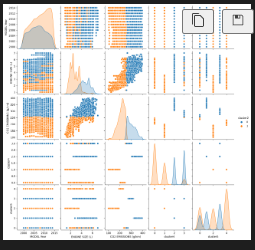
For our objective is about **sustainable development for vehicle**. In the context of vehicle data, for clustering analysis, it would be advisable to include variables that directly relate to vehicle performance and environmental impact. Therefore, ENGINE SIZE (L) and CYLINDERS are essential as they give a sense of the vehicle’s power and potential fuel consumption. The FUEL CONSUMPTION CITY (L/100 km), HWY (L/100 km), and COMB (L/100 km) are critical for understanding the vehicle's efficiency in different driving conditions. CO2 EMISSIONS (g/km) is a direct indicator of the vehicle's environmental impact and is therefore crucial for clustering. Conversely, MODEL Year should be excluded as it does not inherently impact the vehicle's performance or emissions. It's a categorical variable that could introduce bias, as newer models might be more efficient due to advancements in technology rather than their inherent characteristics. COMB (mpg) could be excluded to prevent multicollinearity since it is inversely related to COMB (L/100 km) and does not provide additional information.



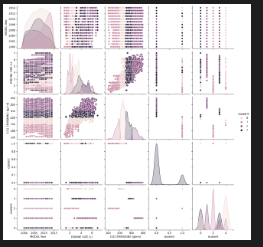
The MODEL YEAR data is more evenly distributed with no significant outliers, indicating that the dataset contains a range of different model years. The plots for ENGINE SIZE (L) and CYLINDERS show more outliers, which may indicate that there are vehicles with unusual engine sizes or cylinder counts. In the graphs for FUEL CONSUMPTION CITY (L/100 km), HWY (L/100 km), and COMB (L/100 km), we can see that some of the data points are far away from the main clusters, which may indicate that some of the vehicles have unusually high or low fuel efficiency. The boxplot for CO2 EMISSIONS (g/km) similarly shows some unusually high emitting vehicles that emit far more than most other vehicles.

For the cluster analysis, we chose CO2 EMISSIONS and ENGINE SIZE as variables. This is due to the high positive correlation between these two variables, which suggests that they are effective in reflecting the environmental impact and performance characteristics of vehicles. By including these variables in the cluster analysis, we can hopefully identify different types of vehicles and provide data support for the development of targeted environmental and marketing strategies. Targeted Marketing and Product Development Segmentation: Focus on Fuel Efficiency: Segments characterized by smaller displacement and lower CO2 emissions can identify consumer segments that prioritize fuel economy and environmental friendliness. These insights can guide the development and marketing of eco-friendly models, such as hybrid or all-electric models, targeting environmentally conscious consumers. Performance-oriented segments: Conversely, groups characterized by larger displacement and higher CO2 emissions can target market segments that value performance and may have a greater preference for larger, more powerful vehicles. We don't evaluate vehicle brands because every manufacturer has a different position on the production of the car and the model of the car. This information can help develop vehicles that match these preferences, while implementing technologies to improve their environmental impact.We decided to retain vehicle information for analysis for only three fuel type groups, E, X, and Z, because vehicle data for fuel types D and N are sparse and may not be sufficient to support reliable statistical analysis or market segmentation.

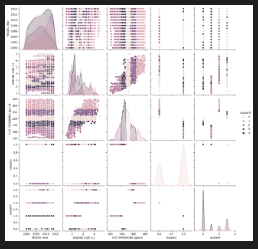
cluster，compare with different custering method, elbow and silhoutee score, for collaborative clustering. K mean is better. The silhoutee score. Make 3 clustering solution and choice the better one. Solution: 1 with 2 cluster, 2 for cluster 4, 3 for clsuter 5.



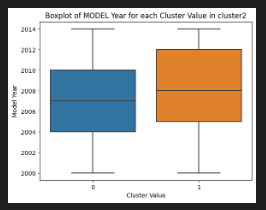
Cluster with 2

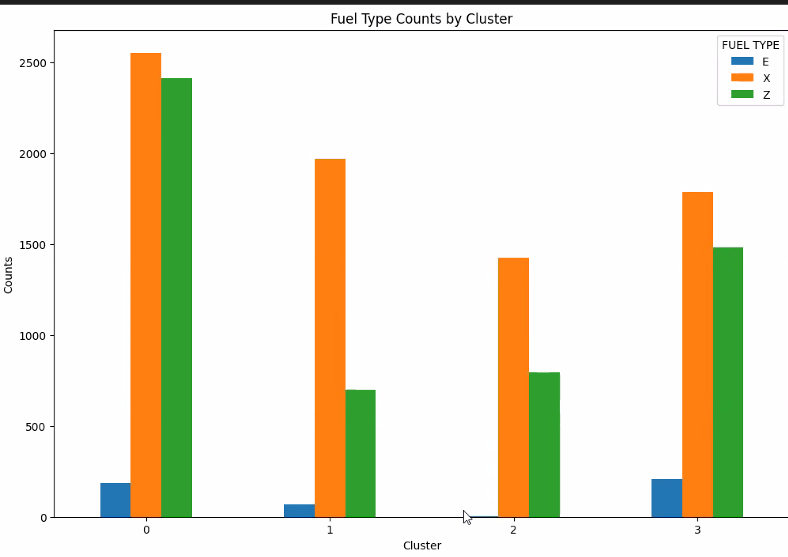


Cluster with 4



Cluster with 5





**Comparative Analysis**

In our clustering solution, we categorize brands into four distinct Clusters, with each Cluster representing a unique market segment characterized by specific attributes such as model year, engine size, CO2 emissions, vehicle class, and fuel type.

Cluster 0 is characterized by smaller engine size and lower CO2 emissions. Vehicles in this cluster are likely to have strong market reputations, proven success, and customer loyalty, attributed to their older model years. Having the extensive data on traditional fuel consumption indicates their dominance in models and makes powered by conventional fuels.

Cluster 1 features the newest model, mostly from around 2010 - the period of rapid technological advancement and the widespread adoption of new technologies. This cluster contains the smallest engine size and lowest CO2 emissions, as the response to sustainable market trends of fuel-efficiency and environmental sustainability.

Cluster 2 is distinguished by its focus on SUVs, trucks, and vans, representing the segment with the oldest model year, the largest engine sizes, and the highest CO2 emissions. Brands within this cluster are likely to target specific marketing segments, emphasizing attributes such as driving safety and suitability for commercial and construction purposes.

Cluster 3 contains a broad spectrum of data across all dimensions. It has a similar model year with Cluster 0 and second largest engine sizes and CO2 emissions. It has diverse vehicle classes, but mainly focuses on SUVs and pick-up trucks, this cluster appears to target both residential and commercial markets, offering customers multiple options.

**Model Year**

The model year of vehicles is a critical indicator of the strategic direction adopted by a brand. It enables clients and stakeholders to predict the technology of a vehicle based solely on its model year. For this dimension, brands in Cluster 0 have been consistently offering products over several years, aligning with the trend of sustainability. This makes them a preferred choice for clients prioritizing reliable collaboration and proven success. Whereas brands in Cluster 1, as the representative of the forefront of technologies, can be the optimal solution for clients seeking market differentiation and advancements in technology.

**Engine Size & CO2 Emissions**

The relationship between engine size and CO2 emissions across the clusters exhibits a strong positive correlation, larger engine sizes tend to result in higher CO2 emissions.

Vehicles in Cluster 2 possess the largest engines among all clusters, the inherent nature of larger vehicles such as SUVs, trucks, and vans. However, it arises challenges in terms of compliance with environmental regulations and achieving social recognition for eco-friendliness.

In contrast, vehicles and brands in Cluster 1, with its smaller engine sizes and lower CO2 emissions, are benefiting from growing customers and policy support. This cluster is particularly appealing to customers with a strong environmental ethic, offering an opportunity to brands to differentiate themselves by supporting increasingly sustainable practice in the market.

**Vehicle Class & Fuel Type**

The analysis of vehicle class and fuel type across the clusters reveals distinct marketing segments, each targeting specific consumer preferences and needs. Cluster 0 and 3 demonstrate extensive market penetration within traditional fuel categories, but have differentiated vehicle classes targeted to customers with various needs. On the other hand, brands within Cluster 1 are likely to extend their product lines to vehicles powered by alternative energy sources, reflecting a strategic decision towards innovation in fuel technology. This approach not only diversifies their portfolio but also aligns with growing consumer demand for environmental sustainability.

**Makes**

The distribution of different makes across each cluster underline the strategic approaches that vehicle manufacturers target diverse market segments by offering a variety of products.

Chevrolet stands out for its extensive presence across all clusters, highlighting its successful mass marketing strategy. Its equal distribution across clusters emphasizes Chevrolet's ability to meet a wide range of consumer preferences and needs, reinforcing its leading position in mass-marketing.

GMC, with major strengths in Cluster 1 and 3, demonstrates a strategy focused on differentiation. This approach highlights GMC’s commitment to providing products with specific use and value to targeted customer segments, which distinguishes itself in a competitive market.

BMW, with a dominant presence in Cluster 2, is the representative of market concentration strategy. The approach indicates BMW focus comprehensively on understanding and serving a specific market segment. The company has maximized its performance within its target market by developing or adapting its products to meet the precise needs.

**Implications**

The Clusters analysis presents a comprehensive landscape of current vehicle market, by having diverse clusters and illuminating the correlation between various vehicle attributes and market trends. For example, as data stated, newer vehicle models tend to have smaller engine sizes and lower CO2 emissions. A key insight from this analysis is the industrial transitions to sustainability. The trend is driven by customer demands for greener vehicles, stringent emissions regulations, and advancements in technology.

This analysis enables the client to align their strategic objective with the broader industrial development in sustainability. By identifying attributes correlated with sustainable practices, such as reduced engine sizes, enhanced fuel efficiency, and the development of vehicles powered by alternative energy sources, the client can not only meet regulatory requirements but also meet the evolving customer preferences.

The company can benefit from focusing on sustainability in multiple ways. It helps the client to gain a competitive edge in the crowded market. Additionally, by prioritizing innovations of hybrid-electric vehicles and explorations of alternative energy sources, the company can create and enhance its brand reputation as innovative and environmentally responsible.