Business Analysis of Car Sales Dataset

Efthymios Mavridis

Mediterranean College – University Of Derby

[efth.mavridis@mc-class.gr](mailto:efth.mavridis@mc-class.gr)​

# Table of Contents

[1. Introduction 4](#_Toc165217055)

[1.1. Aims and objectives. 4](#_Toc165217056)

[1.2. Dataset Information 4](#_Toc165217057)

[1.3. Description of the Car\_Sales Dataset 5](#_Toc165217058)

[2. Analytics Design 5](#_Toc165217059)

[3. Data Analytics 7](#_Toc165217060)

[3.1. Data Preprocessing 7](#_Toc165217061)

[3.1.1. Data Cleaning 7](#_Toc165217062)

[3.1.2. Data Transformation 7](#_Toc165217063)

[3.2. Data Mining 8](#_Toc165217064)

[3.2.1. Distribution of Car Sales Dataset 8](#_Toc165217065)

[3.2.2. Distribution of Car Prices and Annual Income by Category 9](#_Toc165217066)

[3.2.3. Distribution of Car Prices by Gender 13](#_Toc165217067)

[3.2.4. Destribution of Car Prices by Transmission 14](#_Toc165217068)

[3.2.5. Evaluating the Interaction Effects of Gender and Transmission Type on Car Prices 16](#_Toc165217069)

[3.2.6. Comparative Heatmap of Average Car Prices Across Different Companies 18](#_Toc165217070)

[4. Critical Analysis of SAS and Tableau 19](#_Toc165217072)

[4.1. Introduction 19](#_Toc165217073)

[4.2. SAS Overview 19](#_Toc165217074)

[4.3. Tableau Overview 19](#_Toc165217075)

[4.4. Comparison of SAS and Tableau 20](#_Toc165217076)

[4.5. Summary 23](#_Toc165217077)

[5. Conclusion 23](#_Toc165217078)

[5.1 Analytics Insights 23](#_Toc165217079)

[5.2. Limitations and future works 24](#_Toc165217081)

[6. References 25](#_Toc165217082)

[7. Appendix 1 - SAS Code 26](#_Toc165217083)

[7.1. Data Preprocessing 26](#_Toc165217084)

[7.2. Data Cleaning 27](#_Toc165217085)

[7.3. Data Transformation 27](#_Toc165217086)

[8. Appendix 2 – Tableau Configuration 29](#_Toc165217087)

1. Introduction
   1. Aims and objectives.

The aim of the analysis based on the provided dataset is to gain insights into various aspects of the car market. By conducting this analysis, the aim is to provide valuable insights for stakeholders in the automotive industry, including manufacturers, dealerships, and marketers, to better understand market dynamics and consumer behavior. These insights can inform strategic decision-making processes, such as product development, marketing strategies, and sales planning, ultimately leading to improved competitiveness and market positioning.

The objectives of the analysis on the car sales dataset are to:

• Evaluate the dispersion of car buyers' annual income and identify the income levels that influence the choice of car model and Price.

• Analyze the relationship between the gender of car buyers and their preferences in terms of colors, brands, and car models.

• Examine the evolution of car prices over time and determine the factors that affect pricing.

• Analyze the region with the highest demand for cars and identify the most popular brands and models in that region.

• Evaluate the performance of car dealerships based on their region and identify potential factors affecting their sales

The analysis of this dataset will attempt to answer the following questions:

• Customer Segmentation: How can customers be segmented based on their demographic characteristics such as gender, annual income, and location?

• Price Analysis: How do car prices vary across different brands, models, and regions? Are there any patterns or trends in pricing?

• Regional Analysis: What are the regional preferences for car brands and models? Are there any differences in purchasing behavior across different regions?

* 1. Dataset Information

The data is rich in information on car sales, covering transaction specifics, details of the characteristics of the customer acquiring the car, and specifics of the car itself. Hence, it is through this data that meaningful analysis is portrayed in the market trends and customer perspective in the car industry. Thus, an example of this remains a comparative analysis of the patterns and regional disparities with relation to manufacturer performance, popular models, and valuable demographic insights done for stakeholders. Other aspects, like seasonal variations and competitor analysis, gave room for the anticipation of fluctuations in the sale, hence informed strategies on decision-making throughout the whole year. In other words, it helps forecast based on historical data not only future trends of the market but also guides in marketing strategies, advertising campaigns, and investment decisions. Moreover, predictive analysis helps to optimize supply chain management and inventory, reduce operations, and associated costs. All these methods of analysis are integrated to lay the complete framework of maneuvering the complexities of the automotive market, assuring adaptability and sustainable growth.

* 1. Description of the Car Sales Dataset

The dataset used for this analysis is Car Sales and was obtained via the Kaggle website. This dataset contains about 23907 observations, it contains 16 variables such as Car id, Date, Customer Name, Gender, Annual Income, Dealer Name, Company, Model, Engine, Transmission, Color, Price, Dealer\_No, Body Style, Phone, Dealer Region. This structured data can be used to analyze car sales trends, customer preferences, and dealer performance across different regions and time periods.

**Table 1 - Description of Variables in the Car Sales dataset**

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| Car\_id | Unique identifier for each car sold. |
| Date | The date on which the transaction occurred.The format of this column is dd/mm/yyyy. |
| Customer\_Name | Name of the customer who purchased the car. |
| Gender | Gender of the customer (noted as 'Male' which may need verification for accuracy). |
| Annual Income | Annual income of the customer in USD. |
| Dealer Name | Name of the dealership where the car was sold. |
| Company | Manufacturer of the car. |
| Model | Model of the car sold. |
| Engine | Type of engine in the car, such as 'Double Overhead Camshaft' or 'Overhead Camshaft'. |
| Transmission | Type of car transmission (e.g., Automatic or Manual). |
| Color | Color of the car. |
| Price | Sale price of the car in USD. |
| Dealer\_No | Dealer identification number (with a slight column label issue of trailing space). |
| Body Style | Body style of the car, like SUV, Hatchback, or Passenger. |
| Phone | Contact phone number of the dealer. |
| Dealer Region | Region where the dealer is located. |

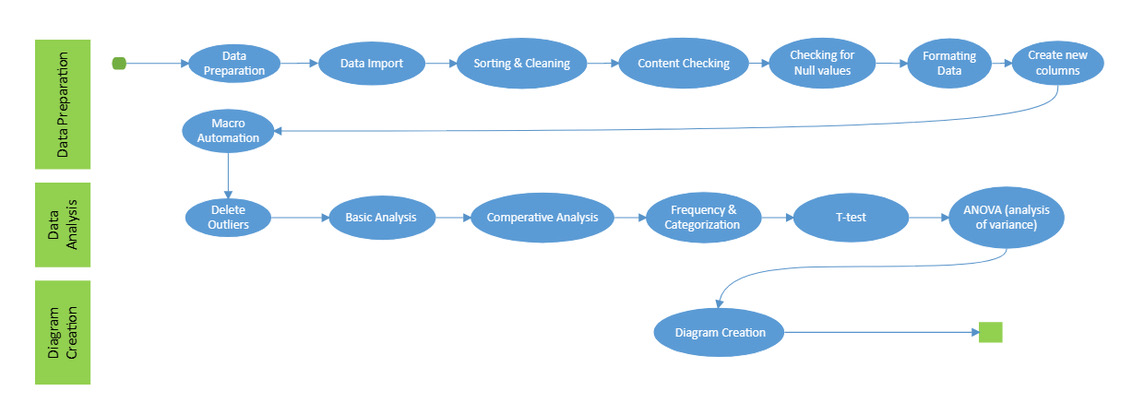
1. Analytics Design

The analytics design that we follow in this process are the steps below.

* Data Cleaning
* Data Integration
* Data Selection
* Data Transformation
* Data Mining
* Pattern evaluation
* Results presentation

The analytics design is represented via Unified Modeling Language (UML) diagram. UML is a standardized modelling language which is used for visualizing, specifying, constructing, and documenting the artifacts of software systems, such as business modelling and other non-software systems. UML is a set of graphical notation techniques that create visual models of software-intensive systems. These visual models can represent various aspects of the system, including behavior, interactions, structure and architecture. Overall UML provides a standardized and widely accepted means of communication among stakeholders, enabling clearer understanding, better collaboration, and more effective development and management of software systems and other complex systems.

The analytics design is represented in a UML Activity Diagram in figure 1.



**Figure 1 – UML Activity Diagram for Car Sales Analytics Design**

1. Data Analytics

Through SAS University Edition and Tableau Public, we dissected their fairly complex data patterns critically. In the process, we used SAS to assist in-depth investigations, pre-processing, predictive modeling, and statistical analysis. On its part, Tableau translated its findings into interactive dashboards, enhancing the capability of SAS. This synergy approach allowed more comprehensive insights from the car sales dataset, exemplifying the effectiveness of using specialized tools for such analyses in a university setting.

* 1. Data Preprocessing

The data preprocessing required for the study needs to be very accurate and reliable. It involved the cleaning and preparation of the data with due care for inconsistencies, missing values, and outliers. Major decision support came from exploratory data analysis (EDA), which gave light into the variable's distribution and characteristics. All throughout the process, we improved quality toward the dataset to ensure a solid foundation for robust analysis.

* + 1. Data Cleaning

Data cleaning is paramount to obtain accurate findings. This involves a systematic approach that tackles inconsistencies, missing values, and outliers. We sought to keep the data integrity by industry standards of data imputation and outlier detection, which could allow a minimization of bias in robust analysis. We first removed the duplicate values and sorted the data using Proc Sort. After that, we checked for null values so that our data would become clearer and more transparent.

* + 1. Data Transformation

This includes turning the data into something that can provide sense in the analysis with the use of industry-standard techniques for the uncovering of trends and patterns. It optimizes the data for use in statistical analysis, predictive modeling, and visualization. Therefore, careful transformation can prove to give valuable insights that are quite implementable for making decisions within the automotive industry. To transform our data, we introduced additional columns so we could search for patterns and relationships. We converted Price and Annual Income from numeric values to currency format.

* 1. Data Mining

This was an important step that key insights were drawn from key variables like Price and Annual Income. Sophisticated data mining techniques helped to unveil the hidden relationships and trends. Clustering and basic statistics helped to identify and give meaning to the patterns found, unveiling the intentions or tendencies present in the customer or market data. Deleting the outliers first, we do some basic statistics, such as average Price and average Annual Income, to check if the clustering results we got by using Price and Annual Income are reasonable or not. We used the average Price also to understand whether there is an interaction effect of price with Gender, Transmission, and Company.

.

* + 1. Distribution of Car Sales Dataset

An initial view from the results of the distribution of data for both Price and Annual Income variables.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 2 - Summary statistics for Price and Annual\_income

It can be derived from the provided statistics that there is a remarkable discrepancy both in the prices of the cars and the annual incomes of the individuals. The mean values only give a hint of the situation, but the large dispersion of minimum and maximum values shows great differences.

The results of such an analysis will help highly in deriving the profile of customers and their decisions on pricing to a large extent, thus helping derive a better targeting of marketing and sales strategies. Such data is very helpful, especially with regard to customers in different income brackets, as it would enable one to understand their purchasing power and how they prefer to spend it.

* + 1. Distribution of Car Prices and Annual Income by Category

At this part of code, we created a macro directive for automating graph generation. Depending on the value assigned to the macro variable &plot\_type, a specific type of graph is executed. For instance, if the &plot\_type is set to 'BOXPLOT', a box plot for car prices by price category is generated.

From the histogram and the boxplot, it is evident that there are outliers, which could skew the results. For this reason, we will remove them in the subsequent steps.

This box plot allows for the comparison of price distributions across different categories, highlighting any variations or outliers within each category.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, διάγραμμα, οθόνη

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 3 - Boxplot of Car Prices by Price Category

Εικόνα που περιέχει κείμενο, διάγραμμα, στιγμιότυπο οθόνης, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 4 - Histogram of Car Prices Without Outliers

This histogram provides insights into the distribution pattern and central tendency of car prices.

The following table shows the result in searching about the differences in Annual Income and Price among the different clustering groups.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά, αριθμός

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 5 - Statistics for Price and Annual Income

The Annual Income variable has an R-Squared (R²) of 0.396079, indicating how much of the variation in Annual Income is explained by the clustering model relative to the groups formed.



Figure 6 - Statistical measure of the clustering model.

This criterion refers to a cumulative statistical measure of the clustering and is 410545 for the model.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά, αριθμός

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 7 - Cluster Summary for Price and Annual Income

The means and standard deviations for the variables `Price` and `Annual Income` seem to vary significantly across groups.

Based on the above, we can draw the following conclusions:

There are significant differences in annual income across clustering groups, as evidenced by the markedly different mean values and standard deviations of `Annual Income` for each group.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, λογισμικό

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 8 - Scatter Plot of Clustering Results on Price and Annual Income

In this scatter plot, the car prices will be against the annual income.

Visual representation reveals some clusters through the Price-to-Annual-Income ratio that lets us target different groups of customers. Clusters show diversified buying behaviors: this is significant for marketing that is segmented and targets strategies towards meeting the diverse customer preference. The distribution of cars according to price categories is as follows:

.Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά, αριθμός

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 9 - Frequency of Price Categories

This tables shows the results of the distribution and the frequency across different price categories.

This distribution of cars reveals significant trends regarding customer preferences based on price categories and can be used for strategic targeting in the market.

The distributions are clearly visible in the following Bar plot.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, οθόνη, ορθογώνιο παραλληλόγραμμο

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 10 - Distribution of cars across Price Categories

* + 1. Distribution of Car Prices by Gender

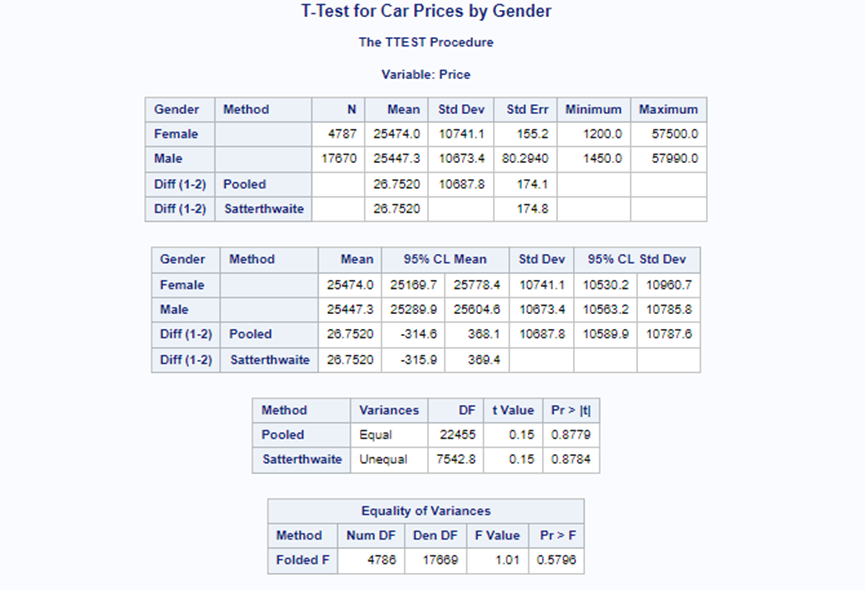


Figure 11 - Gender Based analysis of Car Prices

The T-Test results indicate that there are mean price differences between female and male customers, The t-test indicates no statistically significant difference in car prices between Genders (p-value > 0.05). This is demonstrated by both the pooled and Satterthwaite methods. The equality of variances is not rejected, suggesting that the variance in car prices for females and males is similar.



Figure 12 - Histogram and Kernel density Distribution of Car Prices

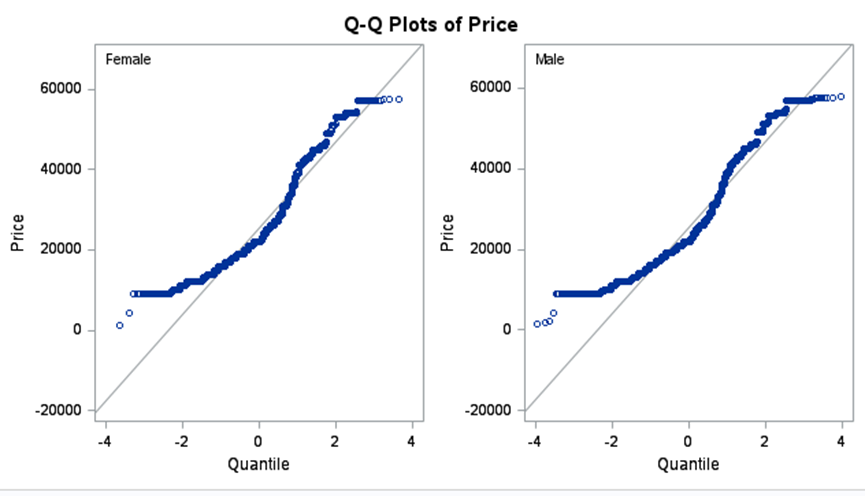
Both histograms and kernel density plots reveal the right-skewed distribution of car prices against male and female segments. The right-skewed distribution informs that there are more instances of lower-priced car purchases and less high-priced car purchases in this segment. The markets' prices, age, and Engine Power are highly skewed variables, with most of the distribution being skewed to the right.

Figure 13 - Quantile plots for both genders according to the distribution of Price

The Q-Q plots for both Genders show that the distribution of Prices deviates from normality, especially at the higher quantiles where we can see a clear deviation from the line, indicating a heavier tail than a normal distribution.

* + 1. Distribution of Car Prices by Transmission

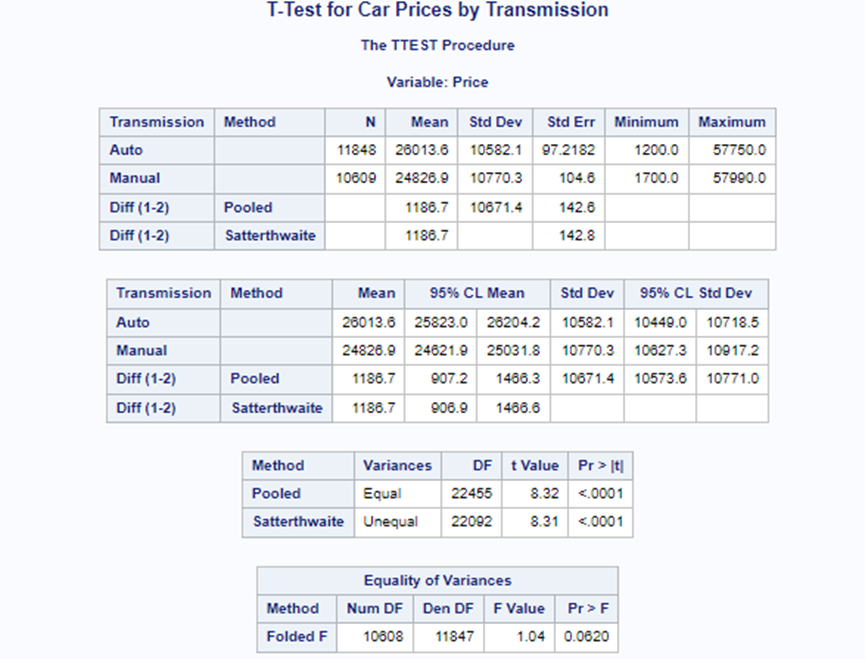


Figure 14 - Transmission Type analysis on Car Prices

The T-Test results illustrate a clear difference in mean car prices between automatic and manual Transmission types. Cars with automatic Transmission have a higher average price compared to those with manual Transmission. This difference is statistically significant with a p-value < 0.0001, suggesting strong evidence against the null hypothesis of equal means.

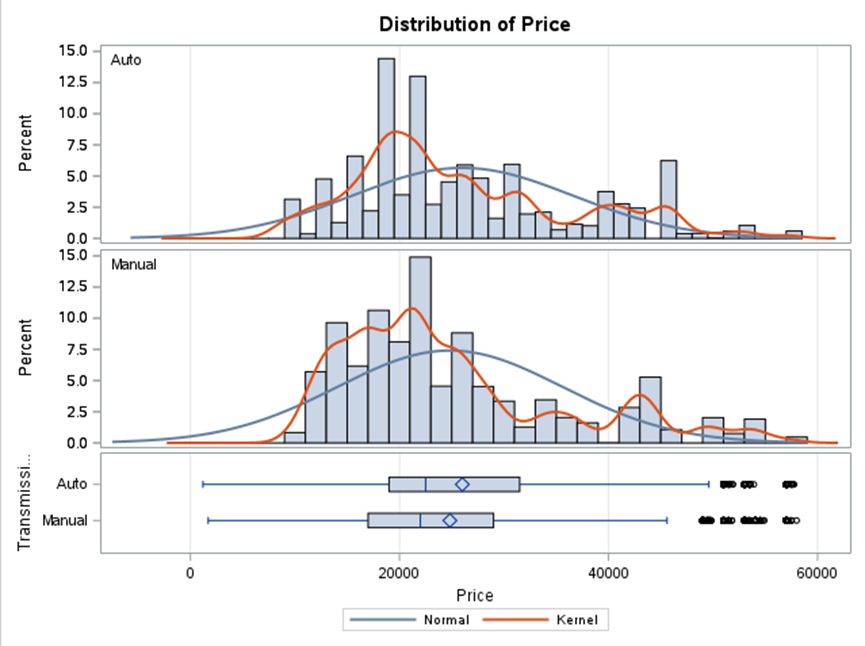


Figure 15 - Histograms and Kernel density plots of both Transmission Types and Distribution of Price

From the histograms and overlaying the kernel density plots on top, it can be seen that both types of Transmission have a price distribution skewed towards the right. The presence of outliers in the data is clearly visible from the box plots, more so for the cars highly priced with an automatic Transmission. The skewness makes it show that though most cars are of low price, there are many cars that belong to the high price category, mostly in the automatic.

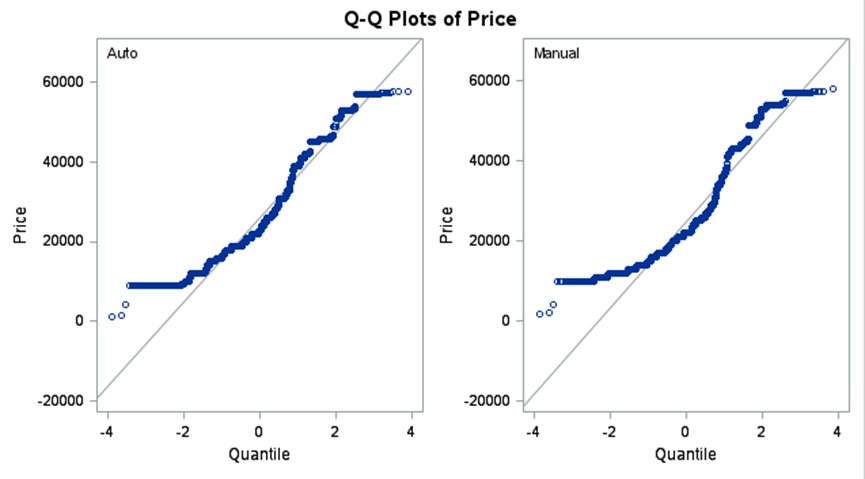


Figure 16 - Quantile Plots for Price Based on Transmission type

The Q-Q plots for Price based on Transmission type show a deviation from normality in the higher Price range, particularly for cars with automatic Transmission. The data points veer off from the straight line, indicating that high-priced car sales are less frequent and thus, do not follow a normal distribution.

* + 1. Evaluating the Interaction Effects of Gender and Transmission Type on Car Prices

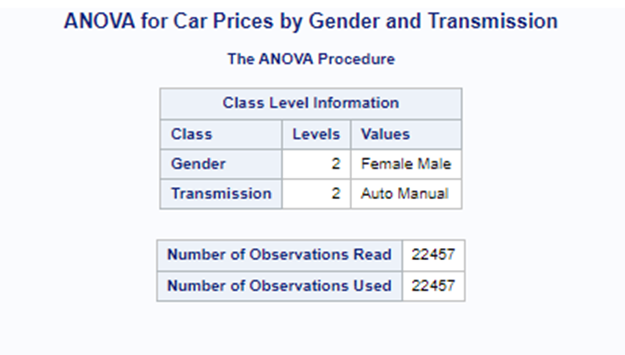


Figure 17A - Analysis of Variance for Car Prices by Gender and Transmission

The ANOVA (Analysis of Variance) output shows how car prices are influenced by Gender and Transmission type, including their interaction.

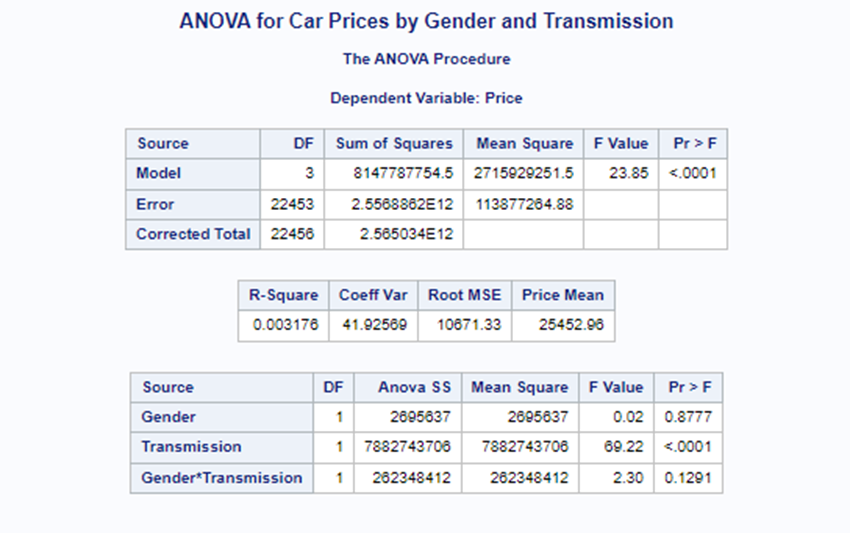


Figure 18B - Analysis of Variance for Car Prices by Gender and Transmission

The results indicate that the model is statistically significant (p < .0001), suggesting that at least one of the factors has a significant effect on car prices. However, when looking at the individual factors the effect of Transmission on Price does not differ significantly between Genders.

The R-square value is relatively low (0.003176), indicating that only a small fraction of the variance in car prices is explained by these factors.

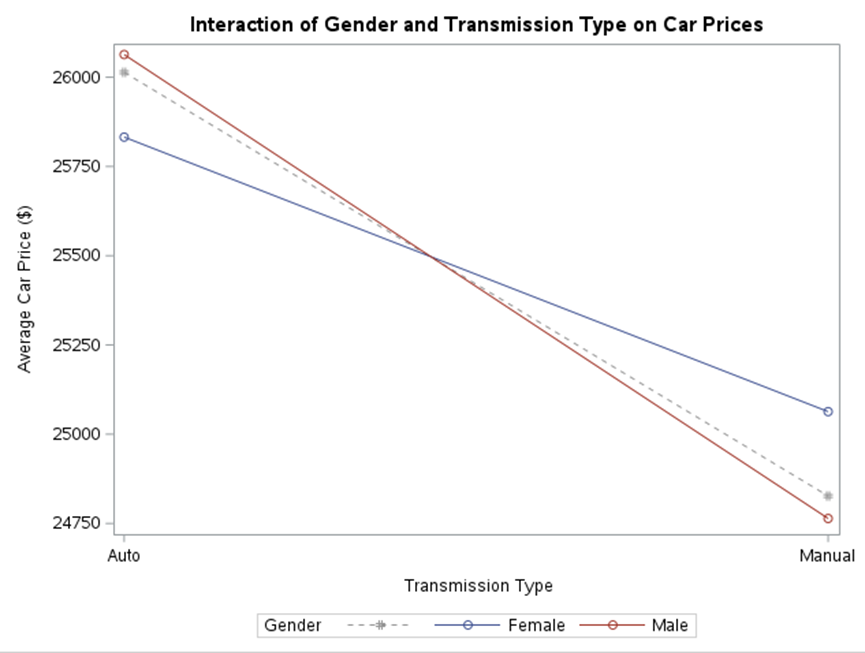


Figure 18 - Interaction of Gender and Transmission Type on Car Prices

The interaction plot is drawn between the Transmission type and Gender with respect to the average car price. Interaction in the slope and the intercept terms appears in the plot for each gender across the two types of transmission.

The scatter plot indicates that automatic Transmission cars are more expensive than manual Transmission cars for males and females. In spite of the non-significance of the interaction term in the ANOVA, the plot was almost parallel, showing that the lines are almost parallel, giving a clearer reinforcement to the former conclusion in the sense that the difference in Price is consistent between gender for each type of Transmission.

In summary, yes, type of Transmission significantly affects car pricing, while Gender does not have a significant effect on car pricing, and neither is the interaction of Gender and Transmission type significant to the pricing. This analysis will help car sellers and manufacturers understand how these may affect the pricing strategies for the car.

* + 1. Comparative Heatmap of Average Car Prices Across Different Companies

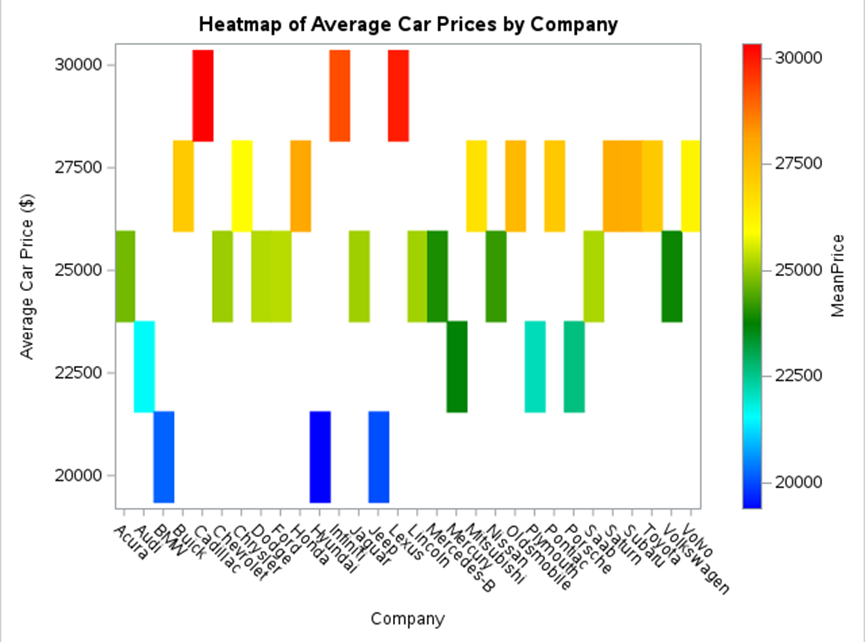


Figure 19 - Heatmap of Average Car Prices by Company

This heatmap provides a visual comparison of the average car prices offered by various automotive companies. Each column represents a company, and the color intensity reflects the company's average car price, with warmer colors (red) indicating higher prices and cooler colors (blue) indicating lower prices.

Analysis of the heatmap reveals the following insights:

• Some companies have significantly higher average car prices, as indicated by the red and orange blocks. These could be premium brands that typically offer higher-end vehicles.

• Other companies, shown in blue and green, have lower average prices, suggesting that these may focus on budget or economy car segments.

• The variation in colors across the heatmap indicates a diverse range of pricing strategies and customer segments targeted by these companies.

Overall, the heatmap effectively highlights the disparities in pricing strategies among different car manufacturers and can serve as a tool for analyzing market positioning and brand perception based on pricing.

1. Critical Analysis of SAS and Tableau
   1. Introduction

SAS and Tableau were used for data analysis and visualization respectively. SAS provided robust insights through data manipulation, statistical analysis, and predictive modeling, while Tableau facilitated clear and compelling communication of findings via interactive dashboards. Leveraging both platforms enabled comprehensive analysis and impactful presentation of results.

* 1. SAS Overview

Originally, SAS was conceived to use in agricultural research by Anthony James Barr in 1966. Since then, it has evolved as the most robust suite of statistical analysis and data management tools. The suite handles data issues such as data manipulation, data analysis, predictive modeling, machine learning, and other complexities. The new shift of huge datasets, SAS becomes more and more useful in a wide array of fields such as healthcare, finance, government, and academia. It has thus grown to be an indispensable part of research that brings ease in terms of data analysis, experimentation, and predictive modeling. With this, organizations in all industries are empowered to make data-driven decisions more productively by confirming its standing as one of the cornerstones in statistics and data science.

* 1. Tableau Overview

Tableau was founded in 2003 by Chris Stolte, Christian Chabot, and Pat Hanrahan as a research project at Stanford University with an approach based on the idea of making data exploration and understanding easy. Tableau is able to come up with dynamic visualizations of different data sources, enabling any person to discover insights, tell their data stories convincingly, and take action into making informed decisions—all of this through an intuitive interface. This level of acceptance of Tableau in academia and industry is very high. It has proven to democratize the analysis and visualization of data, enabling users to draw actionable insights across sectors. Its impact transcends industries to driving innovation and organizational success through data-driven decision-making and performance monitoring.

* 1. Comparison of SAS and Tableau

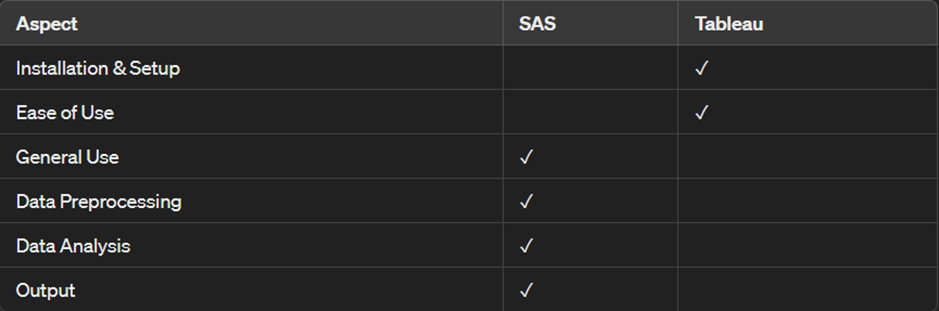
SAS and Tableau differ in installation complexity, with SAS requiring more expertise for setup, while Tableau offers a simpler process. SAS has a steeper learning curve, especially for non-technical users, whereas Tableau's intuitive interface makes it more accessible. While SAS is a comprehensive analytics platform, Tableau excels in data visualization. SAS provides robust data preprocessing tools, while Tableau offers basic capabilities. SAS boasts advanced statistical analysis features, while Tableau focuses on visual analysis. Both tools offer customizable output formats, with SAS catering to complex analytical needs and Tableau emphasizing interactive visualizations for broader audiences. Ultimately, the choice depends on specific user requirements and objectives.

Figure 20A - Comparison of SAS and Tableau

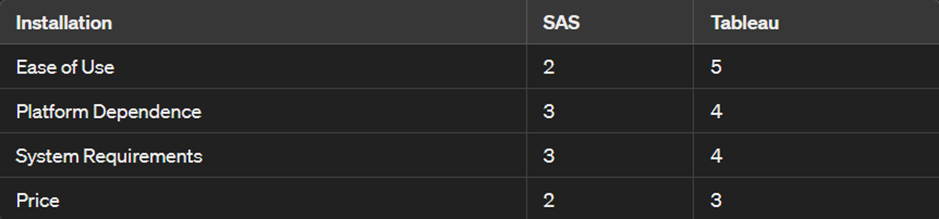


Figure 21B - Comparison between SAS and Tableau

In general, SAS is harder to learn when compared to Tableau, as it has broader functionality and contains programming.

Platform Dependence: Both SAS and Tableau are multi-platformed (Windows, Linux, etc.). However, SAS may tend to have more dependencies and requirements.

System Requirements: Both SAS and Tableau fall within mid-range system requirements, with Tableau hardware requirements possibly falling below those of SAS.

Price: Both SAS and Tableau are commercial software, but SAS is usually more expensive, especially when purchasing enterprise licenses.



Figure 21C - Comparison of SAS and Tableau

Efficiency: SAS is highly efficient for advanced statistical analysis and data processing tasks, while Tableau focuses more on data visualization and exploration.

User Friendliness: Tableau is known for its user-friendly interface and intuitive design, making it accessible to users of all skill levels.

Ease of Use: Tableau has a shorter learning curve compared to SAS, thanks to its intuitive interface and drag-and-drop functionality.

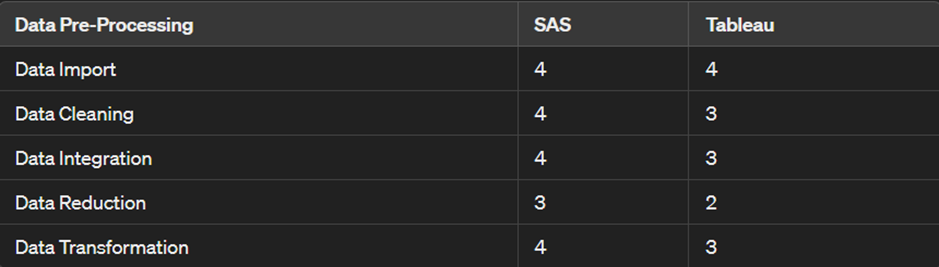


Figure 21 - Comparison in Data Preprocessing of SAS and Tableau

Data Import: Both SAS and Tableau offer robust tools for importing data from various sources, scoring equally in this aspect.

Data Cleaning: SAS and Tableau both provide capabilities for data cleaning, but SAS may offer slightly more advanced features in this regard.

Data Integration: SAS is known for its strong data integration capabilities, allowing users to combine data from multiple sources seamlessly.

Data Reduction: SAS may offer more options for data reduction techniques such as sampling and aggregation compared to Tableau.

Data Transformation: Both SAS and Tableau offer tools for data transformation, but SAS may provide more advanced options for complex transformations.



Figure 22 - Comparison in Data Analysis of SAS and Tableau

Statistical Analysis: SAS provides the best statistical analysis in terms of providing a whole range of procedures on descriptive statistics, inferential statistics, regression analysis, and many more. Tableau offers only rudimentary statistical functions and may lack the same degree of consideration that SAS has to statistical analysis.

Classification Analysis: The classification analysis carried out by SAS exhibits strong capability; the reason is that it includes a wide array of machine learning algorithms and model validation techniques. Tableau has been classified as one of the basic classification analysis functions provided by Tableau and may not be very exhaustive compared to SAS.

Tools for Descriptive Analysis. Descriptive analysis, offered through both SAS and Tableau, allows its customers to summarize and visualize data. Hence, it fetches an equal score in this criterion since they perform equal functions in offering their descriptive analysis tasks.

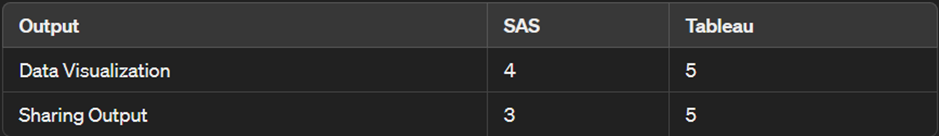


Figure 23 - Data Visualization and Sharing output comparison of SAS and Tableau

Data Visualization: Tableau excels in data visualization, offering a wide range of interactive and visually appealing options for creating charts, graphs, maps, and dashboards. SAS also provides data visualization capabilities, but Tableau may offer more flexibility and interactivity.

Sharing Output: Tableau makes it easy to share visualizations and dashboards with others, allowing for collaboration and real-time updates. While SAS provides options for sharing output, such as reports and graphs, Tableau's sharing capabilities are often more user-friendly and accessible.

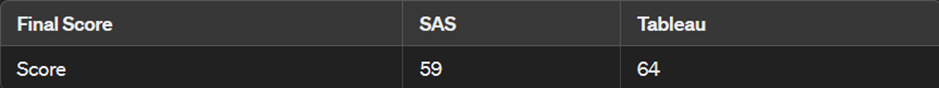


Figure 24 - Final score in the comparison of SAS and Tableau

* 1. Summary

Both SAS and Tableau are strong and competitive platforms for analytical and visual data representation, and each has its own advantages. In terms of tools, SAS is reputed for predictive modeling and has other advanced statistical analysis features. It works best for daunting data processing jobs and full statistical analysis. The learning curve is certainly steeper in comparison and needs much expertise. On the other hand, Tableau has an intuitive interface and is user-friendly in design, dealing exactly with interactive data visualization and its exploration. It gives you a good way of creating visually appealing dashboards and reports that make it easy to share and collaborate. This will basically be a choice between SAS and Tableau, depending on the specifications at hand, use cases, and the organizational preferences. SAS is best used in advanced statistical analysis and data management software, while Tableau excels at data visualization and exploration.

1. Conclusion

5.1 Analytics Insights

Finally, our insights from the analytics provide an all-inclusive understanding of the market dynamics of the automotive industry, majoring on the interplay between the car prices and annual income. From these key variables, enormous tendencies and patterns have been detected, which help in the shaping of consumer behavior and segmentation for the market.

Indeed, our market analysis on prices for cars does clearly point towards such a broad dispersion, from economical to luxury cars, indicating market served has varied purchasing power and preferences. Similarly, the annual income does show broad dispersion, which further gives the evidence of moderate-income group and high net worth individuals.

Using methodologies ranging from clustering to t-tests and ANOVA, with visualizations ranging from histograms, box plots, and heatmaps, many valuable insights have been drawn. These may include the grouping of customers into different segments with variance in income levels and difference in pricing dynamics such as regarding cars for sale and factors like gender and transmission type that affect car pricing.

Further, our results underline the significance of adapted sales and marketing strategies only with targeted customer segments in order that the price premium is translated into actual purchase. Precisely, understanding the nuanced dynamics between price, income, and consumer behavior, automotive companies could optimize their offerings, positioning, and pricing strategy to resonate across diverse customer preferences and market dynamics.

* 2. Limitations and future works

Upon reflecting on our analysis, it is essential to consider the limitations and avenues for future research. Our study has provided insights into the complex relationship between car prices, annual income, and various demographic factors, yet there are constraints to our approach. Our data might not account for critical variables such as regional economic factors, broader consumer preferences, or the evolving dynamics of the market. Expanding our dataset and refining our models will enhance the depth and accuracy of our findings. Given the dynamic nature of the automotive market, continuous monitoring and adaptation are imperative. Future research could involve longitudinal studies to trace pricing dynamics over time and predictive modeling to foresee market shifts. Additionally, exploring aspects like customer satisfaction, brand loyalty, and the impact of emerging technologies could further enrich our understanding of consumer behavior and market trends. In conclusion, while our analysis provides valuable insights, ongoing refinement and exploration are crucial to fully grasp the pricing strategies and market dynamics in the automotive industry.

1. References

Stokes, M.E., Davis, C.S., and Koch, G.G., 2006. Categorical Data Analysis Using the SAS System. Biometrics, 62(1), pp.302-303.

Marasinghe, M.G. and Koehler, K.J., 2021. Statistical Data Analysis using SAS. Journal of the Royal Statistical Society: Series A (Statistics in Society), 183(4), pp.1829-1850.

Mohammad Nazir, N. and Shavarebi, K., 2019. A review of global automotive industry’s competitive strategies. World Journal of Science, Technology and Sustainable Development, 16(4), pp.170-183.

Deloitte, 2024. Global Automotive Consumer Study. Deloitte United States.

Agresti, A., 2007. An Introduction to Categorical Data Analysis, 2nd Edition. Biometrics, 63(4), p.1307.

Stokes, M.E., Davis, C.S., and Koch, G.G., 2000. Categorical Data Analysis Using the SAS System. Biometrics, 56(2), pp. 656-659.

Marasinghe, M.G., and Koch, G.G., 2021. Advanced Techniques in SAS for Categorical Data Analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 183(4), pp.1829-1850.

Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. "Introduction to Linear Regression Analysis." John Wiley & Sons, 2012.

Kutner, Michael H., Christopher J. Nachtsheim, and John Neter. "Applied Linear Statistical Models." McGraw-Hill Education, 2004.

SAS Institute Inc. "SAS/STAT® 15.1 User's Guide." SAS Institute Inc., 2020.

Sheskin, David J. "Handbook of Parametric and Nonparametric Statistical Procedures." CRC Press, 2011.

1. Appendix 1 - SAS Code
   1. Data Preprocessing

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα

Εικόνα που περιέχει κείμενο, γραμματοσειρά, στιγμιότυπο οθόνης, γράμμα

Περιγραφή που δημιουργήθηκε αυτόματα

Εικόνα που περιέχει κείμενο, γραμματοσειρά, στιγμιότυπο οθόνης, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα

* 1. Data Cleaning

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα

* 1. Data Transformation

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

Περιγραφή που δημιουργήθηκε αυτόματα

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

Περιγραφή που δημιουργήθηκε αυτόματα

1. Appendix 2 – Tableau Configuration

Εικόνα που περιέχει στιγμιότυπο οθόνης, διάγραμμα, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 25 - Histogram car prices.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, σχεδίαση

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 26 - Boxplot of Car Prices by Category

.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, διάγραμμα, παράλληλα

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 27 - Scatter Plot of Clustering Results Based in Price and Annual Income.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 28 - - Distribution of Cars Across Price Categories.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, διάγραμμα, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 29 - Interaction of Gender and Transmission Type on Car Prices.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, διάγραμμα, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 30 - T-Test for Car Prices By Gender.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γράφημα, διάγραμμα

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 31 - ANOVA for Car Prices by Gender and Transmission.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 32 - Interaction of Gender and Transmission Type on Car Prices.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης

Περιγραφή που δημιουργήθηκε αυτόματα

Figure 33 - Heatmap of Average Car Prices by Company.