Report - Product Survey Data Analysis

MA4829 - Machine Intelligence

Group Members:
Colas Taylor
Ho Nguyen Ky Trung - U2120180B
Benjamin Teh Jhen Hing - U2023065K
Mallik Aryan - U2023565F

1. Table of Contents

2. Abstract

This report presents a comprehensive survey data analysis aimed at understanding the preferences and behaviours of potential car buyers in Singapore focusing on the burgeoning trend of vehicle customization. We begin by exploring key demographics, including age, gender, and car ownership status, to provide a foundational understanding of the market composition.
Through a variety of data mining techniques including classification, clustering, PCA, and association rule mining, key insights were interpreted from the dataset, illuminating significant patterns and relationships within the data. The analysis uncovered strong demand for personalised vehicles among Singaporean consumers, with this trend being amongst the younger generation in particular. Findings include detailed insights into the specific interior and exterior components that resonate most with respondents for customization. Moreover, the analysis explores multiple factors influencing purchase decisions, and willingness to spend on customization options. These insights offer actionable recommendations for car manufacturers, design firms, and other stakeholders aiming to cater to the growing demand for personalized vehicles in Singapore. Additionally, the report assesses experience with 3D design software and collects unique design ideas reflecting Singaporean identity. Through these discoveries, businesses can develop targeted product design and customization strategies to enhance customer satisfaction, increase vehicle sales, and hence gain a competitive edge in the market.

3. Introduction

3.1 Background

Global car trends lean personalised. Singapore, a tech-savvy and design-driven market, embraces this shift. Our survey of potential car buyers delves into demographics, purchase drivers, and specific customization preferences. Uncovering their desires, we equip car manufacturers and stakeholders with insights to tailor offerings and win over Singapore's personalised car enthusiasts.

3.2 Objective

Singapore car buyer survey unlocks personalised vehicle demands. Analyse demographics, preferences, budgets, and design aspirations to:

1. Craft solutions: Develop data-driven product design and customization options for manufacturers.
2. Boost sales: Recommend features tailored to specific customer segments, maximising appeal and driving sales.
3. Understand buyers: Gain deep insights into the Singaporean car buyer demographic to inform marketing and product offerings.

Bridge the gap between consumer desires and industry offerings, creating personalised car experiences that resonate with Singaporean drivers and fuel market success.

3.3 Scope

Singapore car buyer survey fuels product & sales strategies. Analyse preferences, motivators, & desired customization (budgets, components) to unlock actionable recommendations. Leverage spending tiers for lucrative interior & exterior customization opportunities, catering to Singapore's personalised car dreams.

4. Methodology

4.1 About the Dataset

50-participant survey captures Singaporean car buyer preferences for customization: demographics (age, gender, car ownership), personalities (customisation and personalisation likelihoods), purchase factors, preferred components (interior and exterior), willingness to spend (personalised and custom components), self-design interest, 3D design experience, and uniquely Singaporean design ideas.

4.2 Data Selection

Our analysis focused on key columns for both product and customer understanding.

For products, we examined:

- components desired for customization,
- individual willingness to spend on personalization,
- individual willingness to spend on customization,
- individual design experience,
- unique design ideas input.

On the customer side, we analysed demographics like:

- gender,
- age,
- marital status,
- purchase decision factors,
- current car ownership status.

These elements were chosen for their inherent relevance to the core aspects of product design and customization strategies.

4. 3 Data Cleaning

Several measures were undertaken to enhance the dataset's clarity and completeness.

- Column names were renamed shorter.
- Null values were dropped in all columns if small percentages (<3%).
- Values with small percentages (<3%), were either merged into broader categories ("100-500" merged into "under 500") or dropped ("gender":"prefer not to say").
- Conversion of repeated data into lists was implemented, particularly for interior and exterior components, as well as purchase factors.
- Data cleaning also involved addressing specific features like car ownership and design experience.
- Null values within the "Design Ideas" column were filled with 'none'.

Meticulous restructuring streamlined the data for clarity, user-friendliness, and analysis. Preserving coherence and completeness, it facilitates exploration of unique values and reduces complexity for deeper insights.

4.4 Data Transformation

4.4.1 Data Encoding

4.4.1.1 Label Encoding

Addressing categorical variables with repeated string values, Label Encoding was applied, ensuring a standardised and clear representation of the data. For gender and marital status, an ascending order is assigned based on counts, providing a standardised perspective. Similarly, age groups, spending categories, car ownership, and likelihood variables are encoded in ascending order to facilitate a structured interpretation of the data. For the binary variables in design experience and design ideas, boolean encoding is employed to convert these aspects to True (have) or False (do not have) format, simplifying the analysis of responders' design-related inclinations.

4.4.1.2 List Expansion

List expansion is implemented for those multiple-choice questions featuring repeated variables in their answers, including purchase factors and customization components. This method disaggregates the unique variables into separate columns and assigns boolean values to represent the presence of the factor or component in the response. This comprehensive encoding process enhances the accessibility to the factors and components in the response for further analysis, providing valuable insight into responders' preferences.

4.4.1.3 Data Scaling

Minmax scaling rescales features to a specific range (0-1) based on the minimum and maximum values in the dataset, ensuring all features contribute equally to subsequent analysis. This ensures features don't dominate each other in distance-based algorithms by rescaling their range, leading to fairer comparisons and potentially faster convergence during analysis.

4.4.2 Dimension Reduction - Principal Component Analysis

Principal component analysis condenses high-dimensional data into fewer, uncorrelated variables capturing the most significant information. The optimal number of principal components is determined by variance explained by each principal component. This can be visualised on Scree plots, and the number is chosen at an "elbow" where variance drops off significantly, indicating less informative components beyond that point.

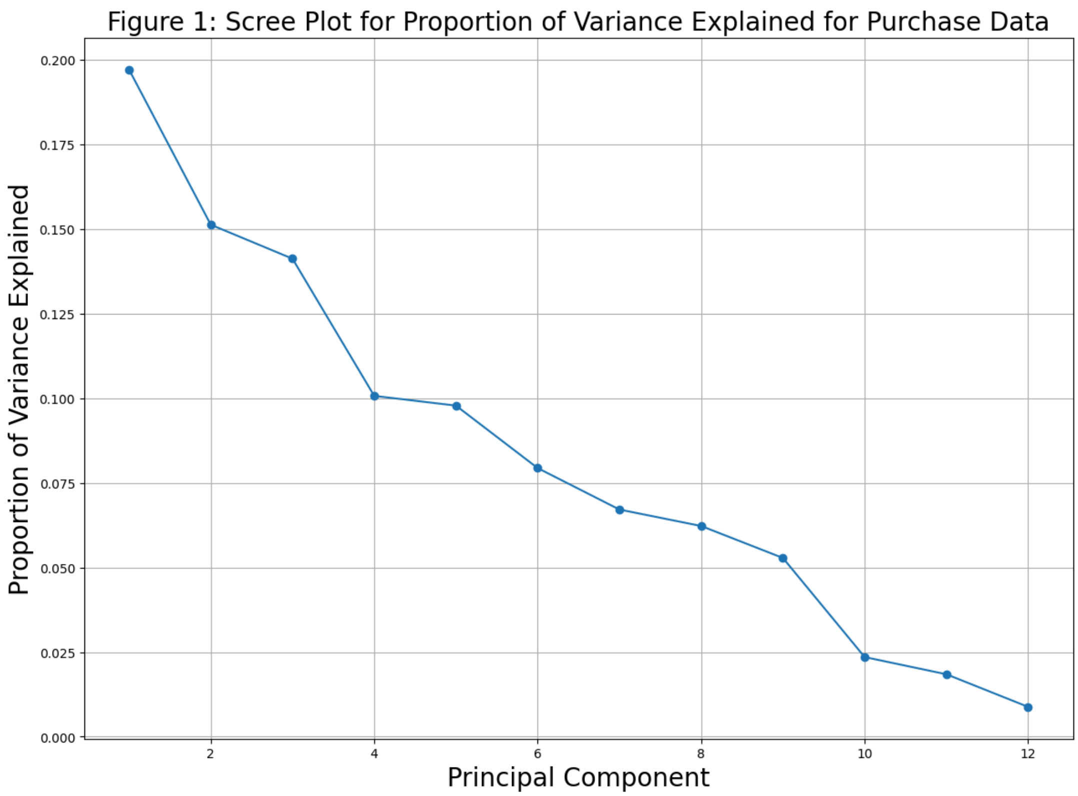

Figure N. Scree Plot for Proportion of Variance for Customer Analysis Data. Elbow at PC=10

For the data set that contains the data for analysis of the customer demographic, the “elbow” is seen when the number of principal components is 10.

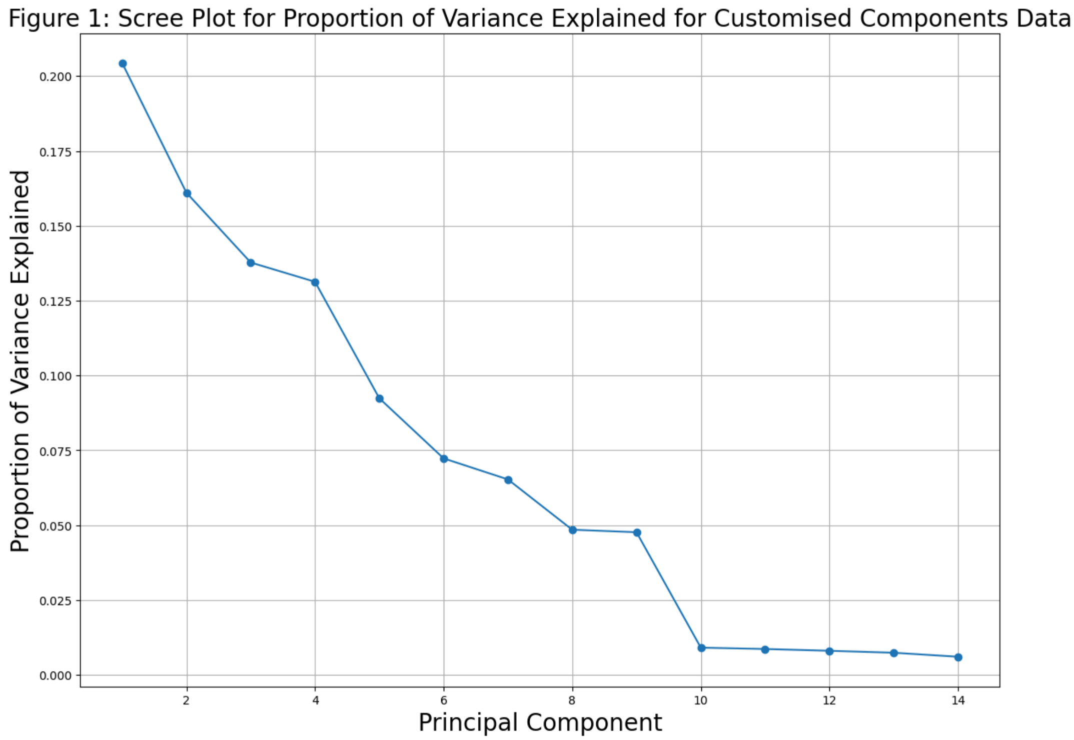

Figure N. Scree Plot for Proportion of Variance for Product Analysis Data. Elbow at PC=10

For the data set that contains the data for analysis of the customisable components, the “elbow” is seen when the number of principal components is 10.

5. Results

5.1 Data Visualisation

5.1.1 Count plots

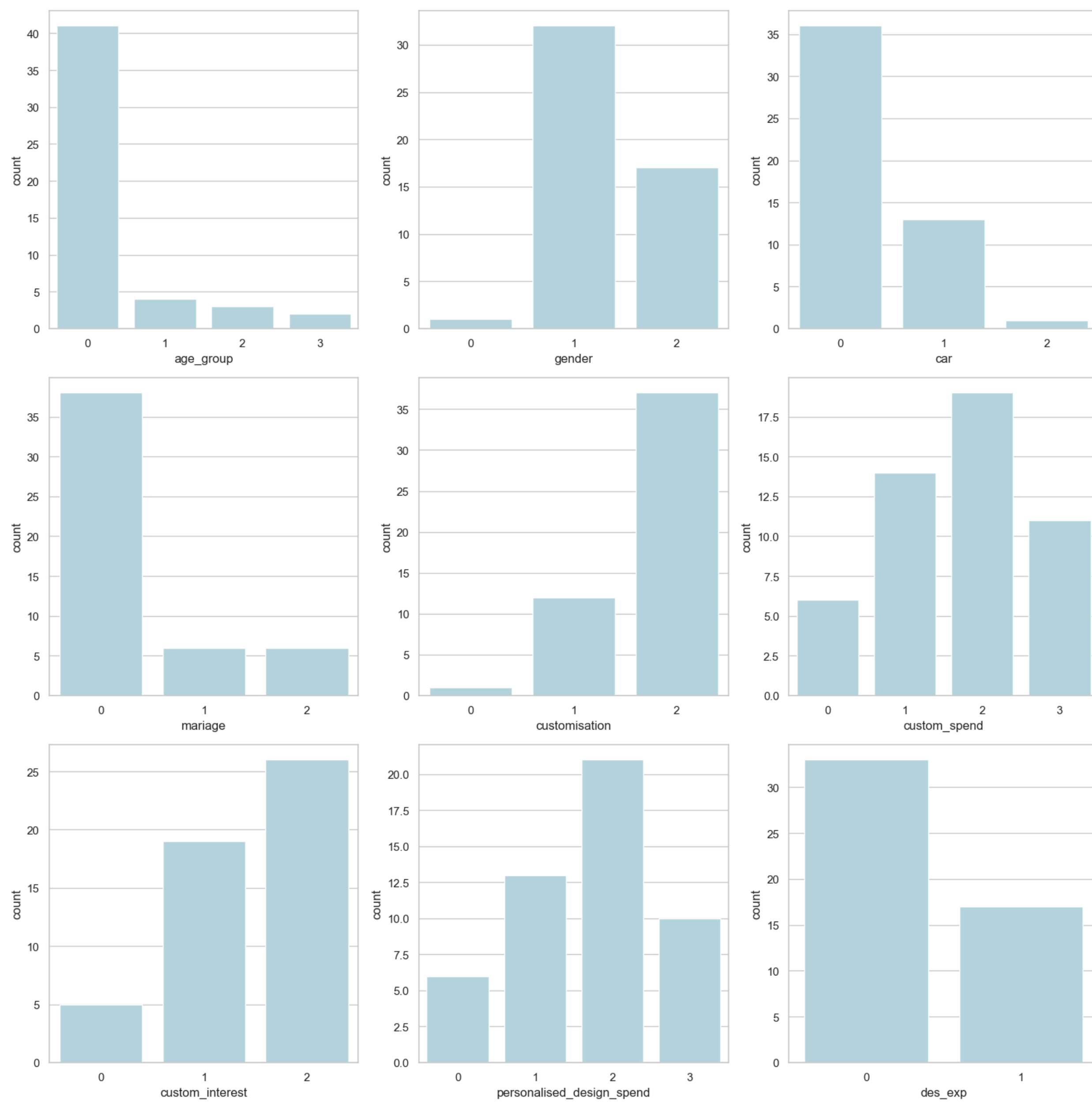

Figure N. Count plots for age group, gender, car ownership, marital status, customisation likelihood, customisation spending, customisation interest, personalised customisation interest, and design experience.

This survey paints a clear picture of potential car buyers in Singapore: young (mostly 20-30s), male, aspiring first-time car owners, single, highly interested in customization, but lacking the design experience or confidence to tackle it themselves. They represent a market eager for personalised options and likely receptive to expert guidance in crafting their dream car. The willingness of survey participants to spend on either a custom component or personalised components roughly resemble a normal distribution.

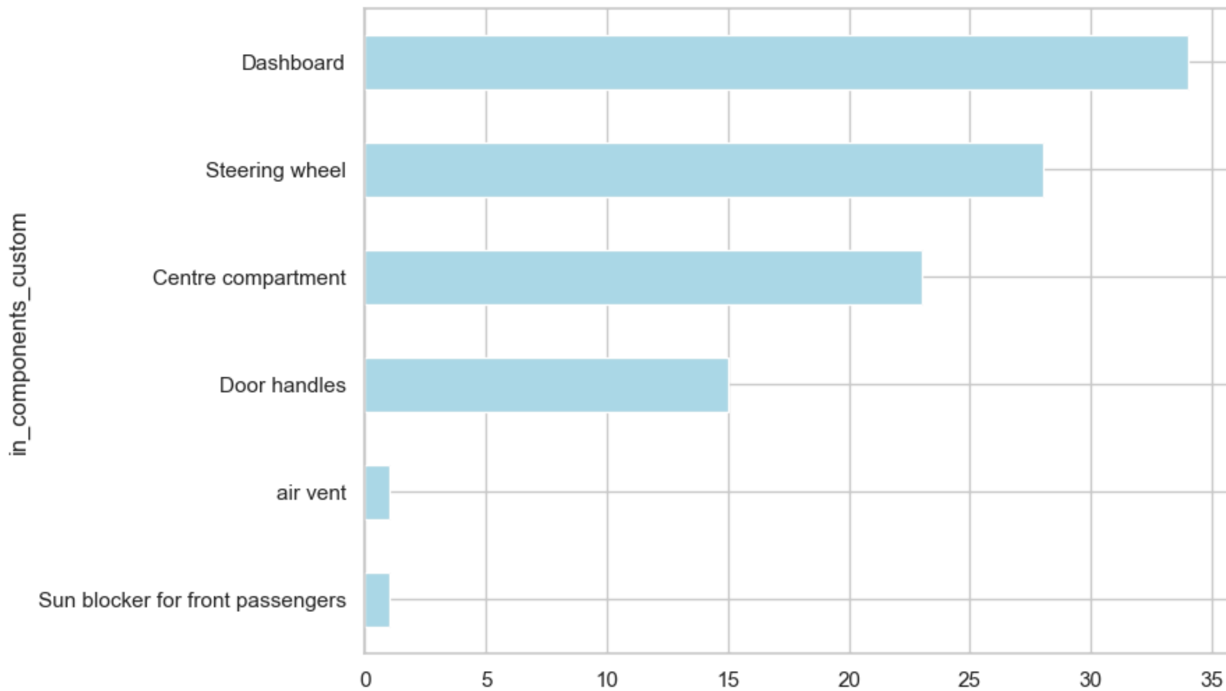

Figure N. Count plots for internal customisable components interest.

The most popular custom interior components are the dashboard, steering wheel, centre compartment, and door handles. There is a large drop in the popularity between the door handles and the less popular components (air vents, sun blockers).

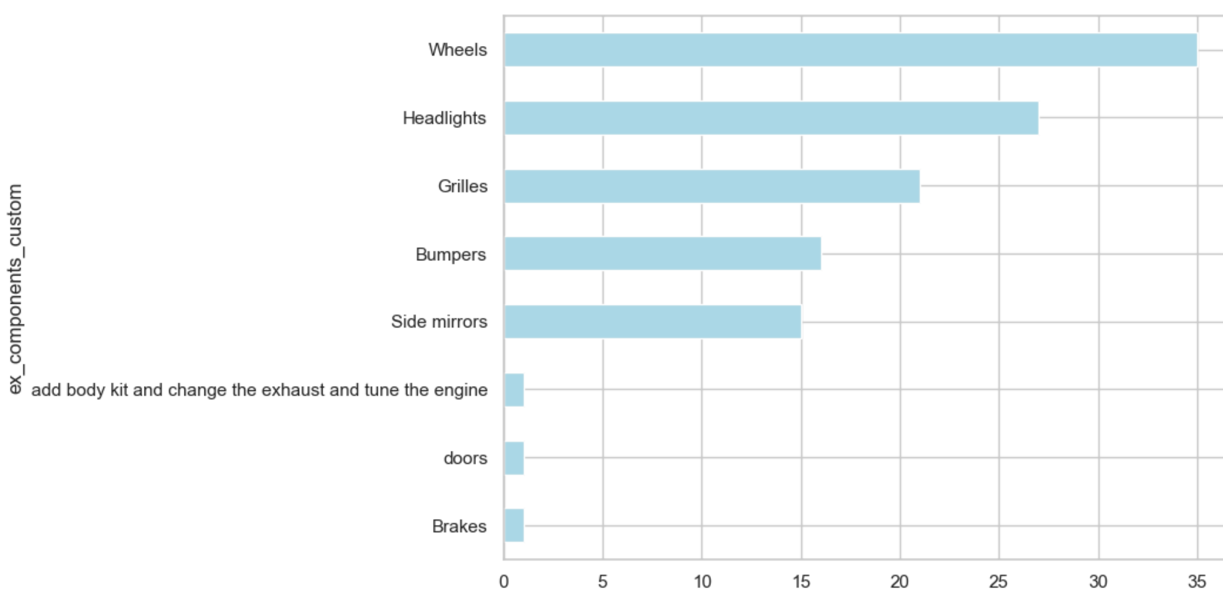

Figure N. Count plots for external customisable components interest.

The most popular custom exterior components are the wheels, headlights, grilles, bumpers, and side mirrors. There is a large drop in the popularity between the door handles and the less popular components (customised body kit, exhaust, engine tuning, doors, brakes).

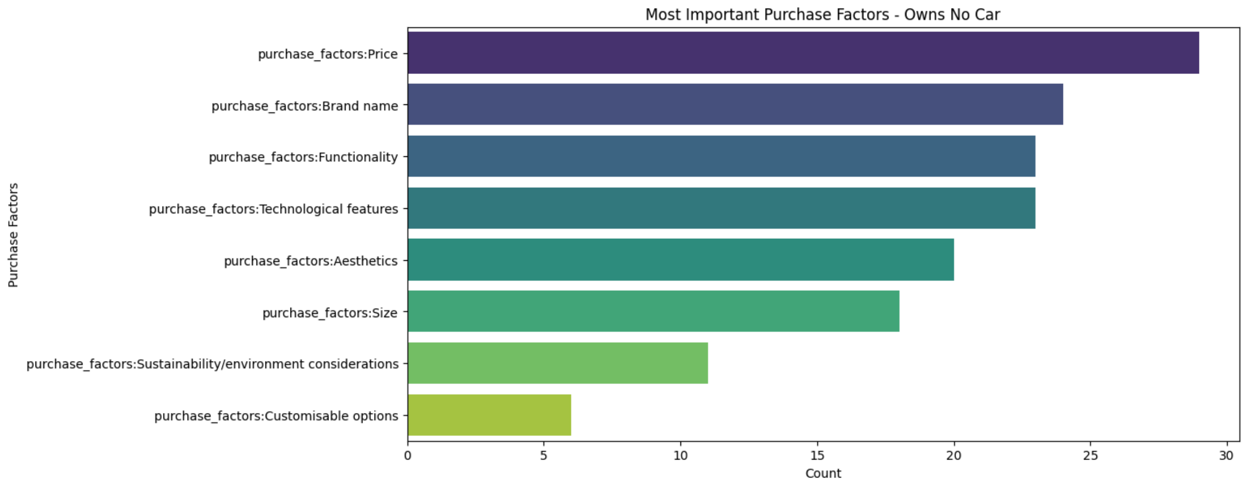

Figure N. Count plots for purchasing factors.

The top 4 common purchasing factors of a car are price, brand name, functionality, and technological features.

5.1.2 Data Correlation

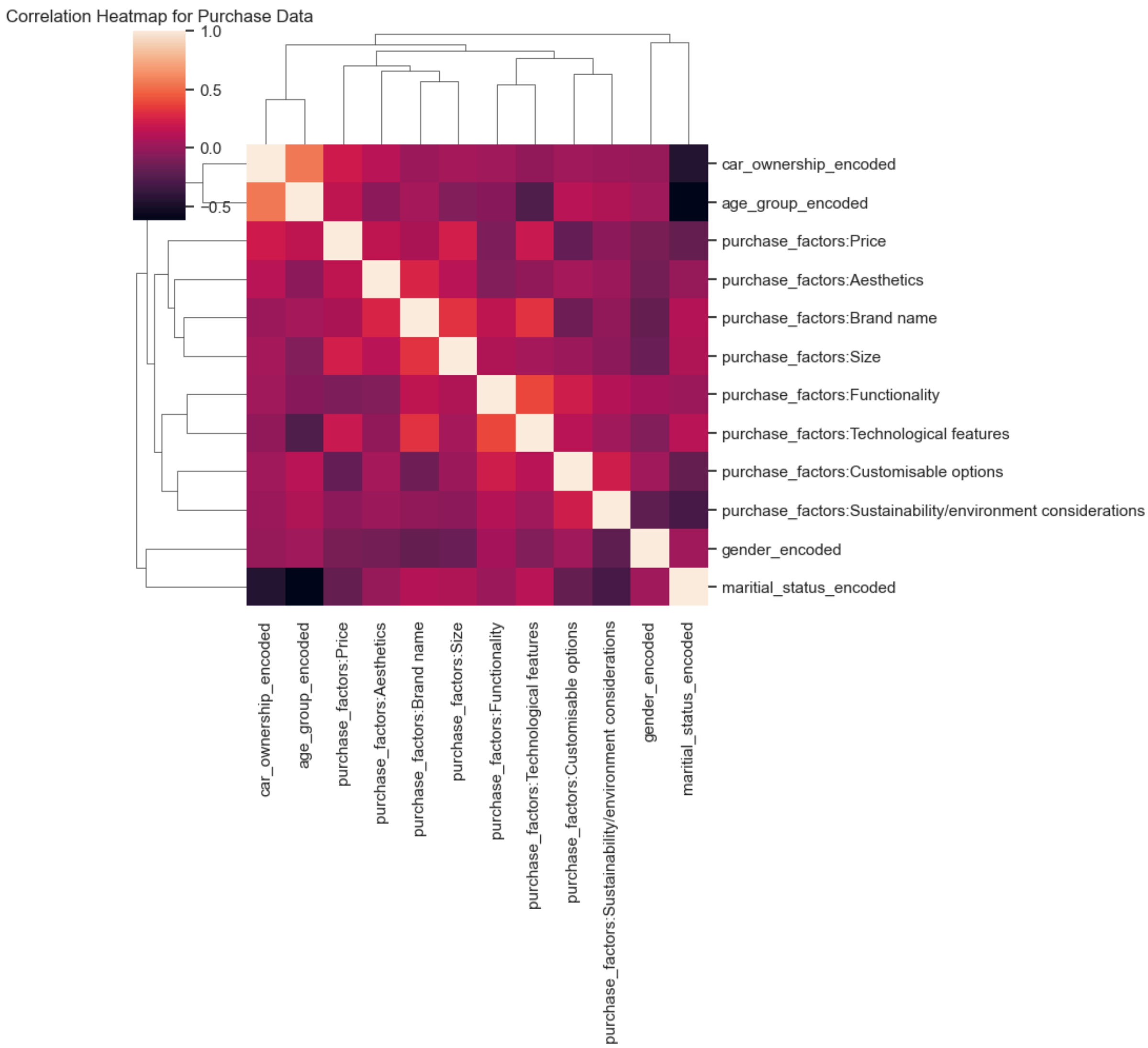

Figure N. Heatmap of correlations of purchasing factors.

This correlation map highlights the priorities of potential car buyers. Car ownership appears tightly linked to both age and marital status. Younger individuals and singles are less likely to own cars. Age also reveals surprising connections: it's negatively correlated with technological features and positively with brand name and size, while older ones value established brands and spaciousness. Technology itself shows strong positive correlations with both brand name and functionality. Sustainability holds negative correlations with age and marital status.

	pearson_correlation to customise_spend
customise_spend_encoded	1.000000
personalise_spend_encoded	0.722008
exterior_components:Wheels	0.337218
gender_encoded	0.332648
interior_components:Steering wheel	0.272192
purchase_factors:Price	-0.255056
exterior_components:Brakes	-0.290175

Figure N. Correlation of customised components to the amount a customer is willing to spend.

The survey reveals intriguing spending habits when it comes to car customization. People willing to splurge on personalised components also tend to show strong interest in customising wheels and steering wheels and skews male. On the other hand, prioritising affordability during car purchase correlates to a lower willingness to spend on custom components. Additionally, those interested in customising brakes seem less inclined to invest in other cosmetic modifications.

5.2 Data Clustering

In another analysis, the K-means clustering method was employed. This method iteratively minimises the distances between data points and centroids to find the most optimum solution for all data points.

Clustering is a useful technique utilised to categorise similar data points, shedding light on the underlying structures and relationships within a dataset. By grouping data points based on their similarities, clustering facilitates the identification of patterns and trends. These insights are valuable for conducting data-driven tasks, such as predictive modeling and anomaly detection.

K-means clustering will be performed on the unscaled encoded dataset. To determine the optimal number of clusters, K, the "elbow method." [https://iopscience.iop.org/article/10.1088/1742-6596/1361/1/012015/meta] was used. Once the optimal K is chosen, the algorithm will assign each data point to the closest cluster based on their encoded features, effectively segmenting the market into distinct groups with shared characteristics and potential customization preferences. The scatter plot of the clusters are then plotted from the reduced data set generated from the PCA.

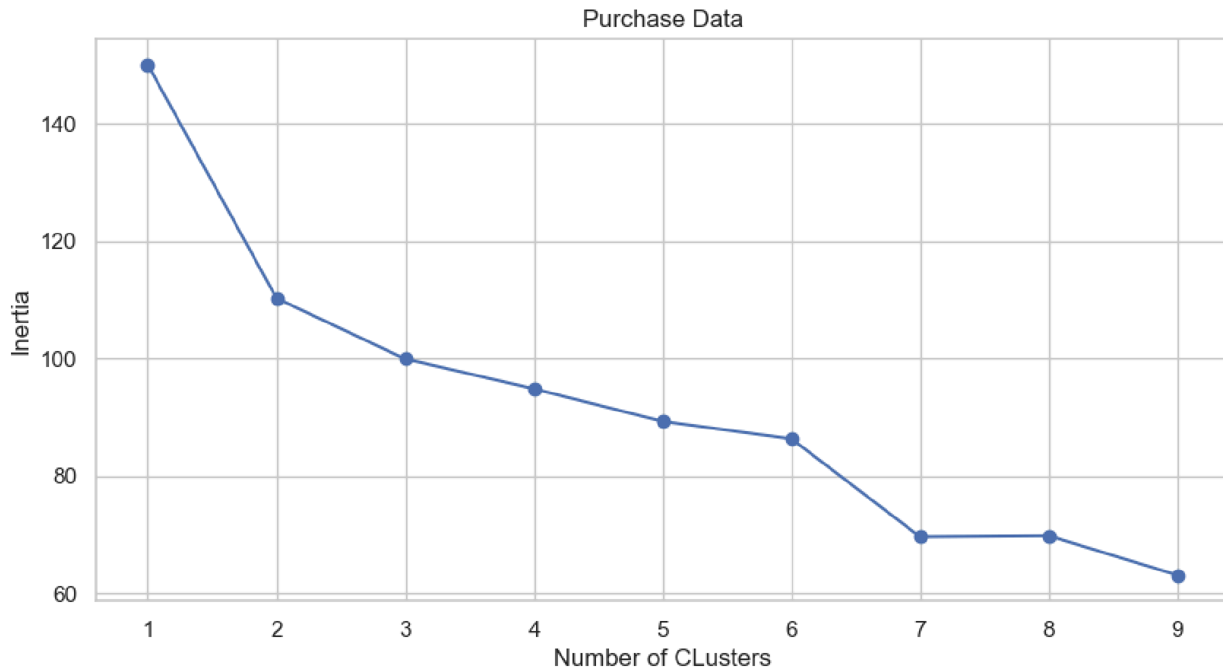

Figure N. Inertia - number of clusters plot for Customer Analysis Data.

This plot shows the decrease in inertia to start diminishing from K = 2. Hence, that will be chosen as the optimum value.

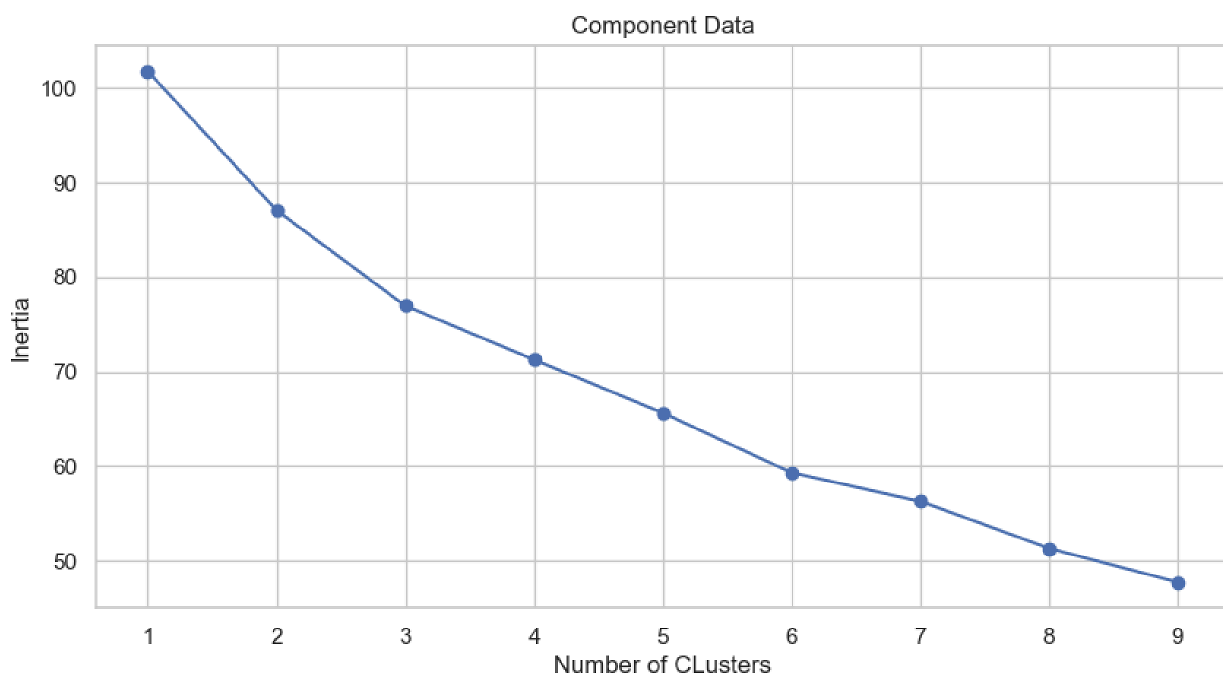

Figure N. Inertia - number of clusters plot for Product Analysis Data.

This plot shows the decrease in inertia to be roughly constant. The point = 3 is is selected to

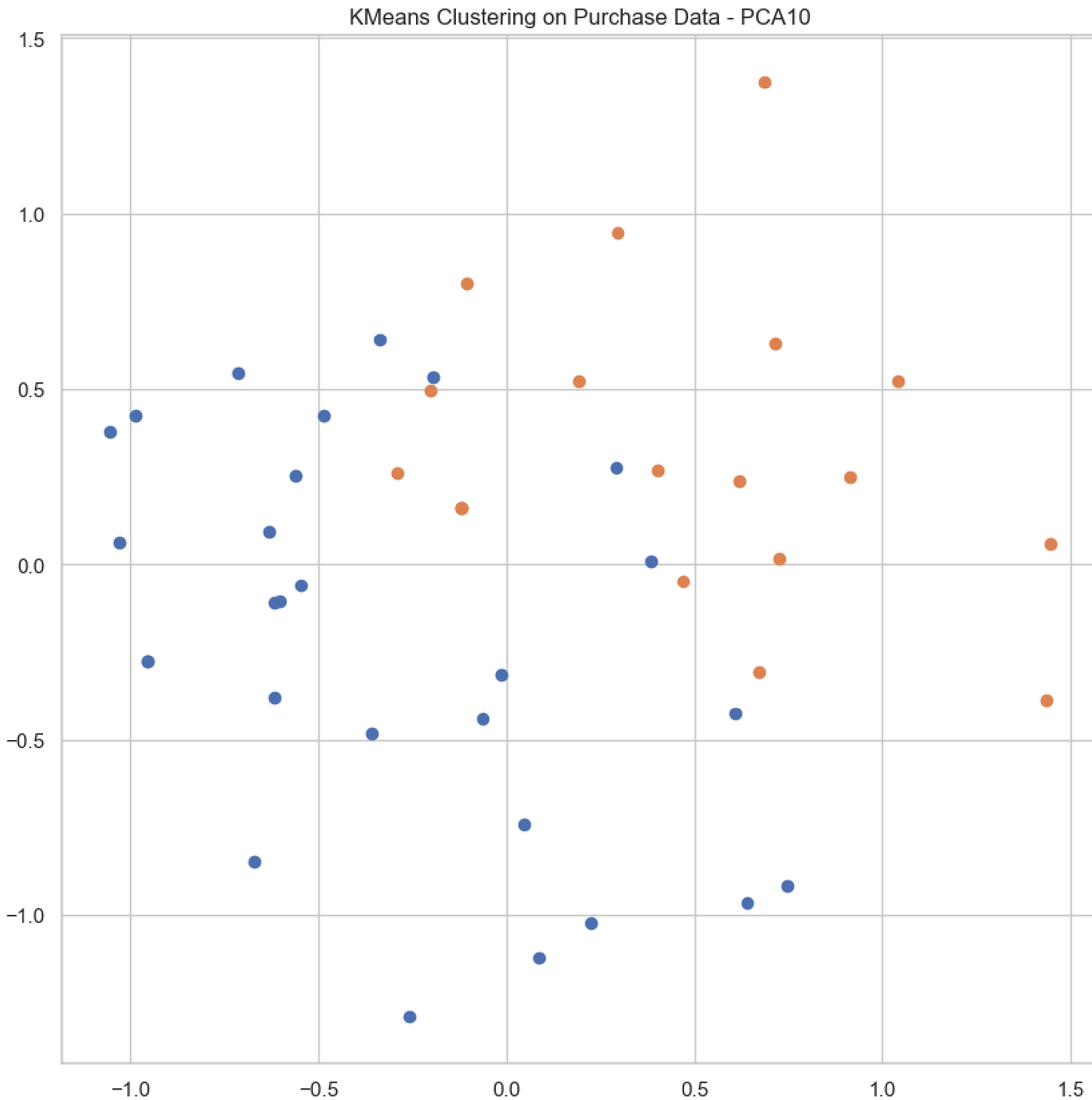

Figure N. KMeans Clustering on Customer Analysis Data with PCA10

As a result of the elbow method, the clusters are easily identifiable.

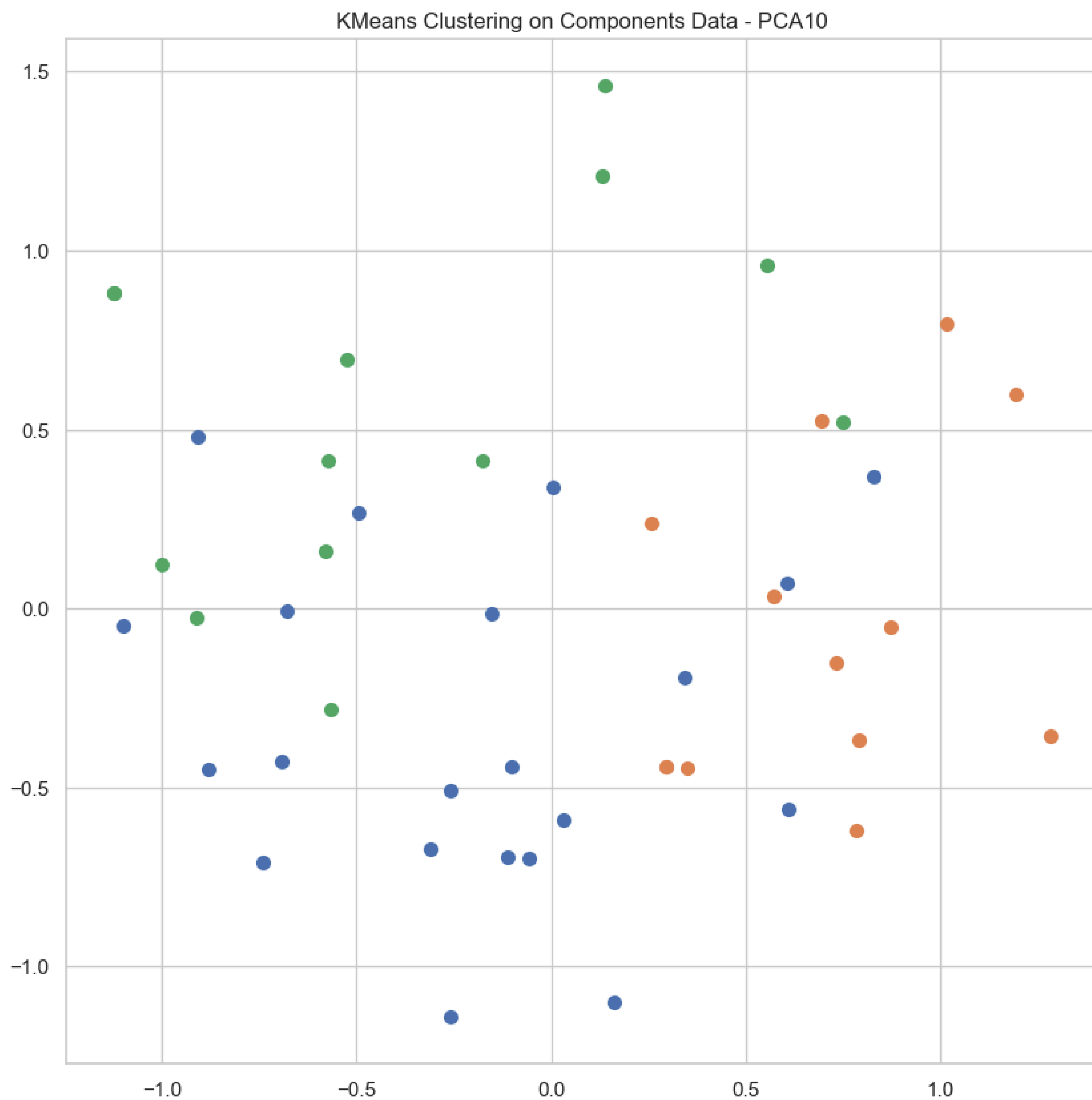

Figure N. KMeans Clustering on Product Analysis Data with PCA10

The clusters although have areas distinct to it. However, there are large overlaps between the clusters.

5.3 Data Classification

Data Classification is the process of organising data by relevant categories. The objective is to organise and structure data in a meaningful way that facilitates analysis, retrieval, storage, and security.

Furthermore, data classification also serves as an important tool in product design and development, allowing businesses to leverage the information that they gain to enhance their future design decisions.

This analysis uses two separate classifiers(DecisionTreeClassifier/RandomForestClassifier) to identify the importance of the car components, listed in terms of consumer customization preferences.

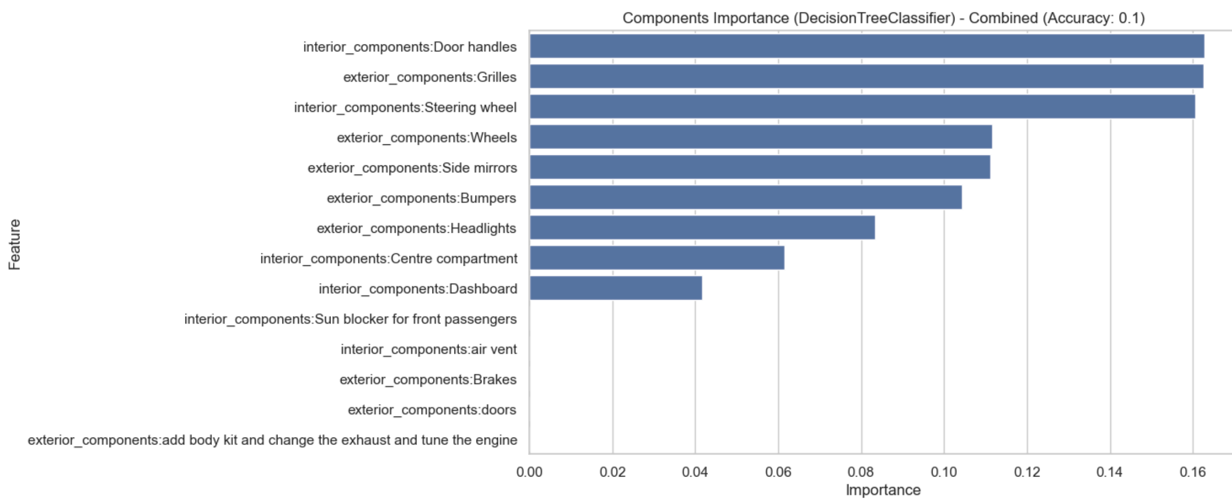

Figure N. Customisable component feature importance to classification of willingness to pay for customisation using Decision Tree Classifier

The Decision Tree Classifer has determined the most important interior and exterior components of a car, and the top 5 will be listed in descending order of importance:

Door handles **(Interior)**, Grilles **(Exterior)**, Steering wheel **(Interior)**, Wheels **(Exterior)**, Side mirrors **(Exterior)**

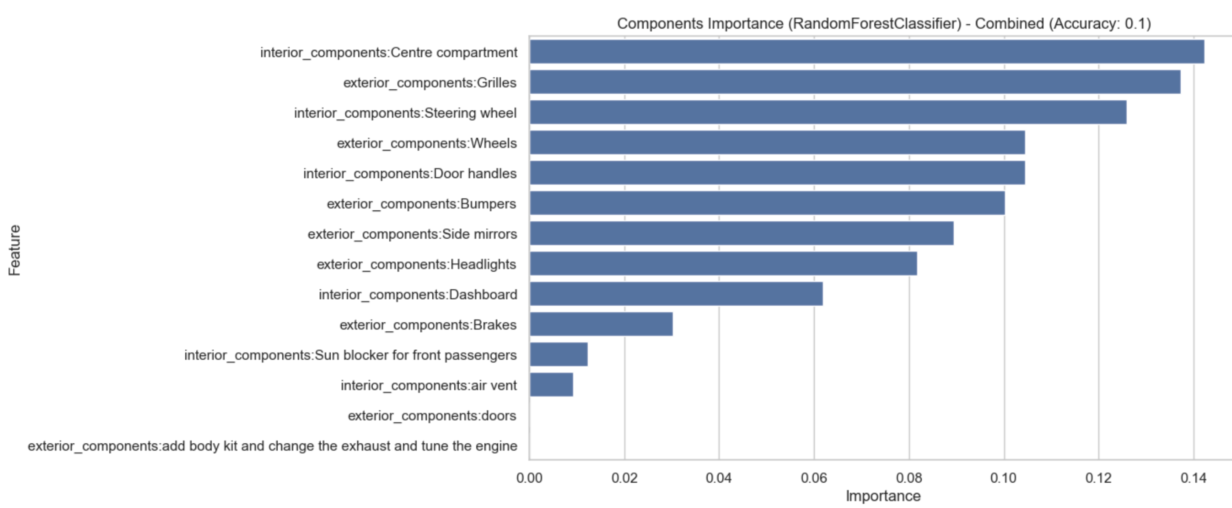

Figure N. Customisable component feature importance to classification willingness to pay for customisation using Random Forest Classifier.

Centre Compartment **(Exterior)**, Grilles **(Exterior)**, Steering wheel **(Interior)**, Wheels **(Exterior)**, Door handles **(Interior)**

5.4 Association Rules Creation

Association Rule Creation uncovers frequent patterns and relationships within data, revealing hidden connections between items or events.

The technique was performed on the Product Analysis Data where the minimum confidence of 80% and minimum support = 10% was set to ensure the rules are generalisable.

5.4.1 Association Rules on Interior Components

The only association rule meeting the criteria suggests that when both dashboards and door handles are chosen for customization, there's a strong (likely high confidence) likelihood that customers will also choose to customise the steering wheel. This points to a potential "interior styling bundle" preference among buyers interested in those specific components.

5.4.2 Association Rules on Exterior Components

While 10 association rules met the criteria, the top 3 outputs ranked are headlights (5), wheels (4), and grilles (3). The specific input combinations are in the appendix.

5.4.3 Association Rules on All Components

While 125 association rules were identified, the most sought-after custom components (ranked outputs) were wheels (45), headlights (30), steering wheel (22), dashboard (18), side mirrors (8), and grilles (2). The specific input combinations can be found in the appendix.

6. Discussion and Analysis

6.1 On Visualised Data

Data visualisation of the survey participants reveals key insights into the target audience for car preferences and customization trends. Among the respondents, a significant majority, falling within the 20-30 age group, suggests a younger audience with a likely inclination towards modern and trendsetting features. Furthermore, the survey indicates a noticeable gender bias, with a higher percentage of male participants. The marital status distribution shows that most respondents identify as single. Individuals without family constraints are keener on personalised vehicles. In terms of car ownership, a noteworthy revelation is that the majority of respondents do not currently own a vehicle. This finding suggests an untapped market of potential first-time car buyers or individuals looking to upgrade their transportation preferences.

The high interest in customising vehicles, particularly if offered without extra charges, unveils a potential service market for car customization. Moreover, the inclination towards seeking professional help for customization, emphasises the importance of user-friendly and accessible customization interfaces.

The approximately normal distribution of willingness to spend on personalisation design and customisation surcharges suggests potentially different budget ranges and preferences for the customization program.

Manufacturers should focus more on improving the top factors that responders take into consideration while buying a car such as price, brand, function, and technical features. For customised components, wheels, headlights, and grilles are chosen the most among exterior components, while dashboard, steering wheel, and centre compartment are chosen the most among interior components. This highlights specific areas where the company can focus its customization efforts to attract customers.

6.2 On Data Correlation

The correlation map offers valuable insights into the potential car buyers' preference in customisation by showing the relationship between various factors.

Younger individuals and singles are less likely to own cars, possibly reflecting different transportation needs or financial considerations. Negative correlation between age and technological features suggests that younger buyers might not prioritise advanced technology in their purchasing decisions. On the contrary, there are positive correlations between age and brand name and size, indicating that older individuals value established brands and spacious vehicles. Conversely, negative correlations between sustainability with age and marital status hints at varying environmental concerns based on life stages.Technology itself shows strong positive correlations with both brand name and functionality, implying buyers would prefer reliable and well-equipped vehicles.

Those willing to invest in personalised components also tend to customise wheels and steering wheels, and exhibit a strong interest in customization and skew towards male responders. This suggests that visual and tactile enhancements play a crucial role in the customization preferences of certain demographic groups. Additionally, the contradictory interest in customising brakes and investing in other cosmetic modifications reflect a preference for performance upgrades over purely aesthetic ones.

The correlation result also highlights the importance of considering budget constraints and financial priorities in understanding consumer behaviour regarding car customization.

These insights highlight the diverse motivations behind customization choices, suggesting car makers and designers should cater to both visual appeal and performance-driven desires to attract various customer segments.

6.3 On Data Clustering

For the graph on Purchase Data, the clusters can be shown to separate individuals with different priorities in purchase factors. One cluster represents individuals who prioritise factors such as price, functionality, and technological features. This cluster can be shown to predominantly consist of younger individuals who can be inferred to be more technologically inclined and value the practical aspects of a product. Customization options seem to not be as significant for this group when compared to the others.
Another cluster comprises people who tend to prioritise brand name, aesthetics, and customization options. It also suggests a trend towards younger consumers, but not as much as the other cluster. This group seems to be willing to pay more for customization options when compared to the others.

This can be of great help to businesses during the design and development phase of the manufacturing process.

- Understanding differences in preferences among consumers allows car manufacturers to tailor their designs to meet specific market demands.
- Designers can allocate their resources more effectively by understanding the priority of purchase factors amongst their consumers
- Manufacturers can also gain a competitive advantage by aligning product offerings to consumer desires.
- Marketing efforts can be tailored to specific segments if there is knowledge about their factor priorities.

For the graph on Components Data, the clusters have separated individuals based on their customization preferences. The orange cluster appears to be individuals who have preferences

for customization options that are more extensive than average. The green cluster appears to be for individuals who have preferences for moderate levels of customization.
And finally, in the blue cluster are individuals who have preferences for minimal customization or standardised options.

This can be useful to businesses as:
- Clustering will allow them to conduct effective market segmentation, dividing the customer base into meaningful groups based on similarities in preferences.
- Tailoring products to meet the specific preferences that the different customer segments may have will increase customer satisfaction and loyalty.

6.4 On Data Classification

The most important components by consumer customization preferences are determined to be Door handles **(Interior)**, Grilles **(Exterior)**, Steering wheel **(Interior)**, Wheels **(Exterior)**, Side mirrors **(Exterior)**.
- allows manufacturers to identify the areas where consumers seek individualization and personalization in their vehicles.
- By offering customizable options for these components, manufacturers can enhance their overall appeal.
- By focusing on high-priority customizable components, manufacturers can optimise their supply chain and reduce costs associated with an excess inventory.

6.5 On Association Rules Creation

The results of association rules creation offers valuable insights for driving incremental sales through targeted customization options. It reveals that customers seeking customization are more likely to be interested in exterior components like headlights (5 mentions) and wheels (4 mentions), followed by grills (1 mention).

Interior preferences lean towards combinations, with the pairing of dashboards and door handles appearing once alongside steering wheels. The most significant overlap lies in both interiors and exteriors, where wheels (45 mentions) and headlights (30 mentions) reign supreme, followed by steering wheels (22 mentions) and dashboards (18 mentions). These findings suggest that offering bundled packages focused on popular combinations like "Headlights & Wheels" or "Dashboard & Steering Wheel" could entice customization-minded customers.

Additionally, highlighting unique grill options could cater to a niche segment seeking personalization for that specific component. Remember, catering to both individual preferences and identified trends can maximise your chances of securing incremental sales through effective customization strategies.

7. Conclusion

In conclusion, this report provides insights into the personalised car market in Singapore which may be utilised to bridge the gap between consumer desires and industry offerings. These data-driven insights reveal the most desired interior and exterior custom components in the market. This informs manufacturers on which components would play a bigger role in mass customisation and which could play a role in niche personalisation. This equips sales teams on how to market custom components. This would inform all stakeholders about making informed decisions about car sales, car customisation, and car personalisation. Thus, this would lead to greater revenue.

8. Appendix

1. https://github.com/bentjh01/MA4829_Machine_Intelligence.git

9. References

1. Ahlquist, J. (2023). Principal component analysis applied to survey data: Methodological aspects and application [Online]. ResearchGate. Retrieved February 17, 2024, from https://www.researchgate.net/publication/371563950_Principal_Component_Analysis_applied_to_survey_data_Methodological_aspects_and_application

2. Datacamp (n.d.). Principal component analysis in Python [Online]. Retrieved February 17, 2024, from https://www.datacamp.com/tutorial/principal-component-analysis-in-python

3. District Data Labs (2020, September 11). Principal component analysis with Python [Online]. Medium. Retrieved February 17, 2024, from https://medium.com/district-data-labs/principal-component-analysis-with-python-4962cd026465

4. https://www.youtube.com/watch?v=QdBy02ExhGI

5. Singh, J. (2021, November 10). Understanding k-means clustering in machine learning with examples [Online]. Analytics Vidhya. Retrieved February 17, 2024, from https://www.analyticsvidhya.com/blog/2021/11/understanding-k-means-clustering-in-machine-learningwith-examples/

6.https://www.youtube.com/watch?v=5w5iUbTlpMQ&t=1131s&pp=ygUeaG93IHRvIGRvIGNsdXN0ZXJpbmcgaW4gcHl0aG9u.

7. Stack Exchange (2024, February 16). Interpret clustering plotted in the first two principal components [Online]. Stack Exchange. Retrieved February 17, 2024, from

https://stats.stackexchange.com/questions/86479/interpret-clustering-plotted-in-the-first-two-principal-components