ocean

Navigating Through Doug's Car Scores

Luca Ordronneau
Discord: lucanew#3793 | GitHub: lucaordronneau

Desights

10-28-2023

Introduction

Purpose of Analysis

This report delves into the "DougScore" dataset, aiming to uncover **patterns** and **insights** from Doug DeMuro's comprehensive vehicle reviews.

Dataset's Richness

Spanning numerous vehicle entries, each with detailed scoring across multiple categories, the dataset offers a holistic view of both performance and aesthetic metrics of automobiles.

Engagement

Crafted for car enthusiasts and potential buyers, this analysis **translates complex data patterns into relatable insights**.

Deep Dive into Relationships

We'll highlight intricate relationships between scoring categories, spotlighting **top-performing** cars and offering tangible takeaways.

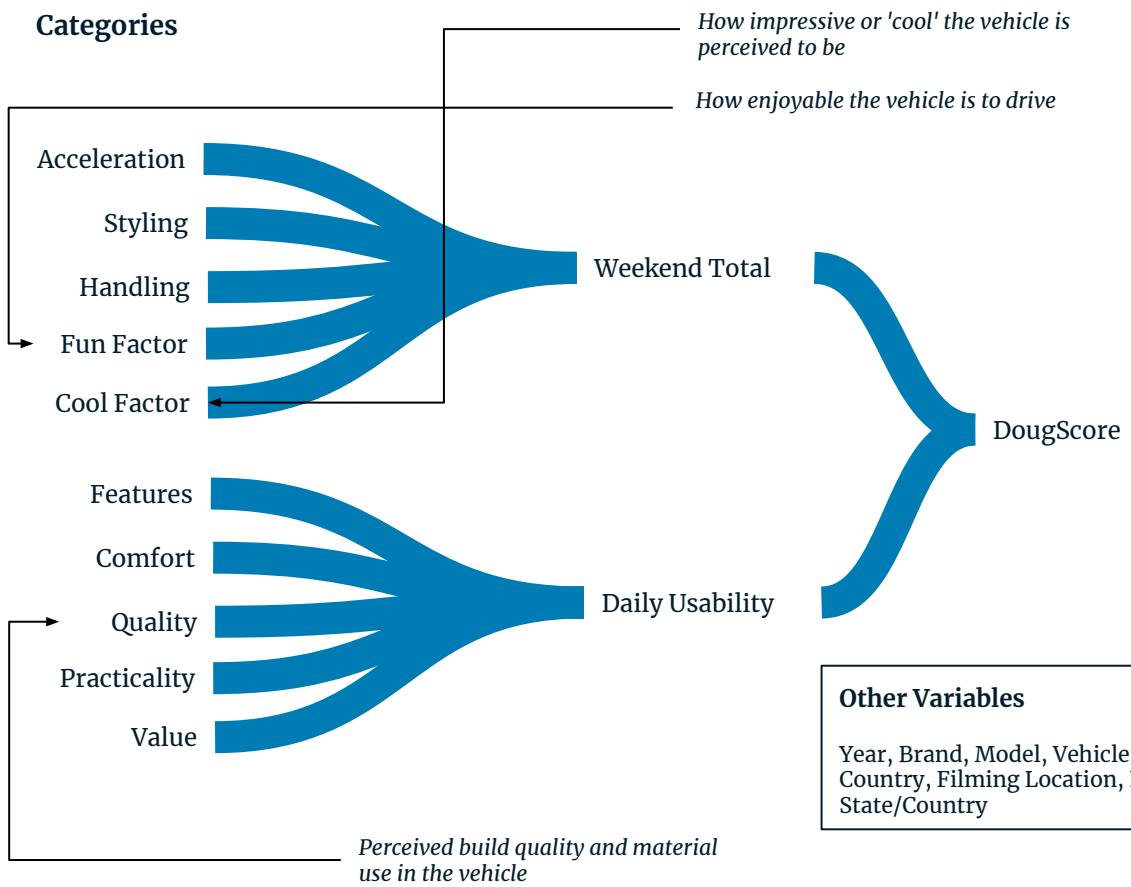
Exploring Dynamics of DougScore

Beyond correlations, the report investigates the **synergies** between individual scores and their collective impact on the final "DougScore".

Implications and Takeaways

The findings present real-world implications, guiding decision-making for prospective buyers and offering fresh perspectives to car aficionados.

Dataset Variables Overview



Other Variables

Year, Brand, Model, Vehicle Country, Filming Location, Filming State/Country

Example

Year: 2020
Brand: Ferrari
Model: Monza SP1
Vehicle Country: Italy

Styling: 9
Acceleration: 10
Handling: 10
Fun Factor: 10
Cool Factor: 10
Weekend Total: 49

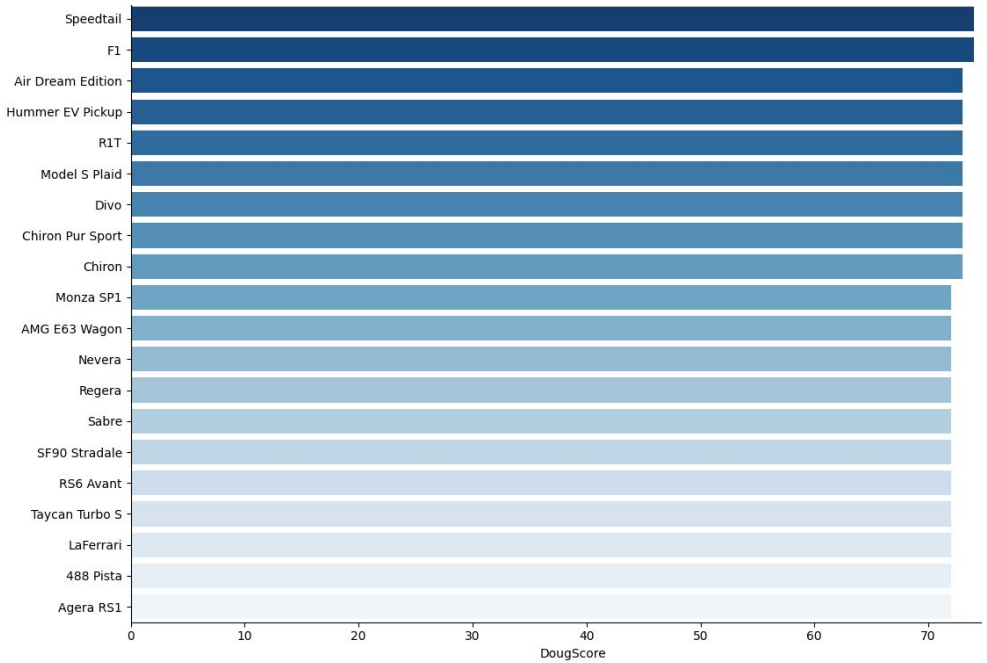
Features: 5
Comfort: 2
Quality: 8
Practicality: 1
Value: 7
Daily Usability: 23

DougScore: 72

Filming Location: 32:13
Filming State/Country: California

General Analysis

Top 20 Cars with Highest DougScore



The most recent years in the dataset show some of the highest average scores, suggesting that vehicles from these years might have been rated more favorably. This could reflect **advancements in vehicle technology, design, and features**.

What makes the top cars have this high DougScore ? Let's delve deeper into categories reliationships

No. of vehicles
585

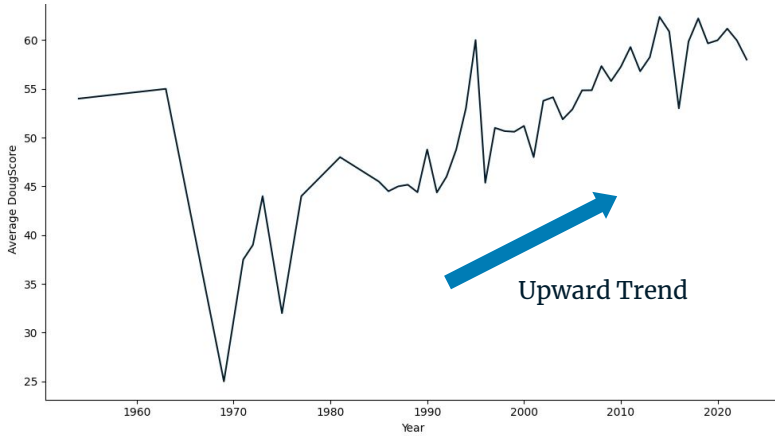
Predominant vehicle origin
Germany

Max DougScore
74

Min DougScore
25

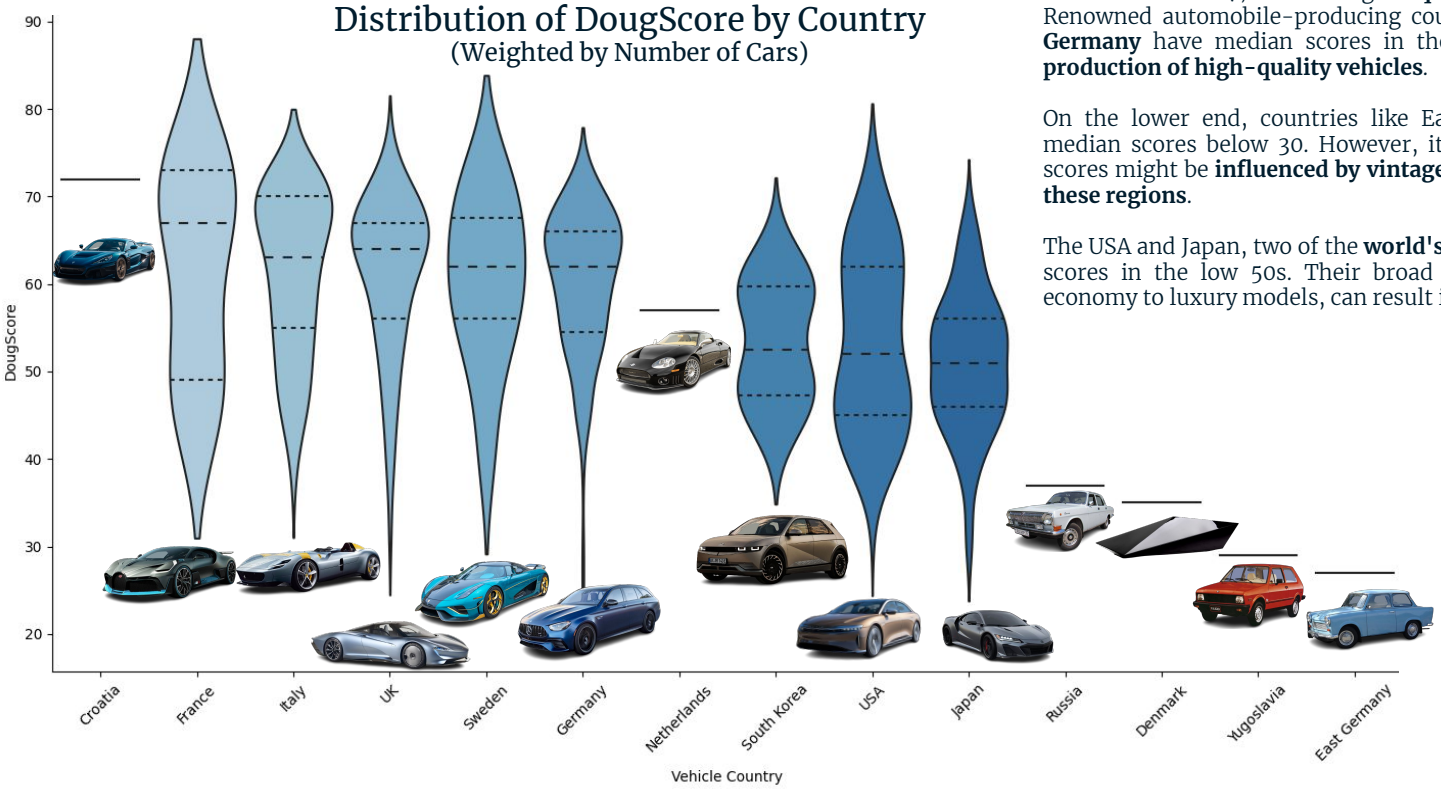
Average DougScore
56.92

Average DougScore Over Time



General Analysis

Country analysis



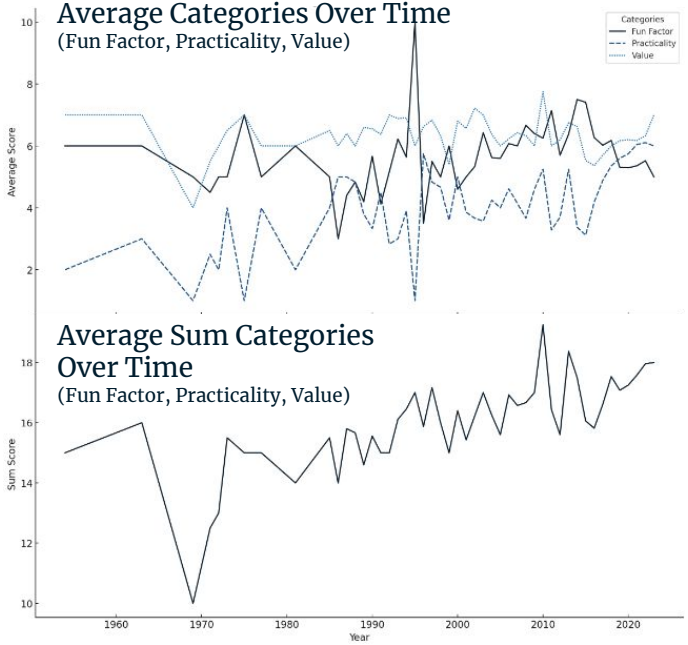
Croatia emerges at the top with a median DougScore of **72**, though it's crucial to note that this is based **on a single car review**, making it less representative of a broader trend. **France** follows closely with a notable median score of 67, showcasing the **quality and appeal** of French vehicles. Renowned automobile-producing countries such as the **UK**, **Italy**, and **Germany** have median scores in the 60s, reflecting their **consistent production of high-quality vehicles**.

On the lower end, countries like East Germany and Yugoslavia have median scores below 30. However, it's essential to consider that these scores might be **influenced by vintage or unique vehicles reviewed from these regions**.

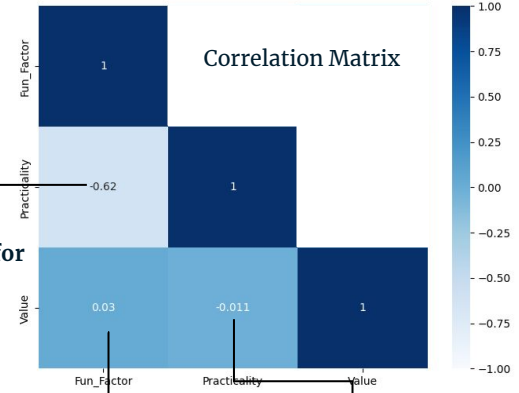
The USA and Japan, two of the **world's major car producers**, have median scores in the low 50s. Their broad range of vehicles reviewed, from economy to luxury models, can result in a diverse score distribution.

Relationship Analysis

Fun factor, practicality, value



As **Fun Factor has declined** and **Practicality increased** since 2010, with **Value remaining stable**, it suggests that the automobile industry is recalibrating its design priorities. Even as fun and excitement might be slightly reduced in newer models, their **practicality** and **consistent** perceived market worth highlight a careful balance in meeting evolving consumer expectations.



Suggests that cars rated **higher for fun tend to be rated lower for practicality**, and vice versa.

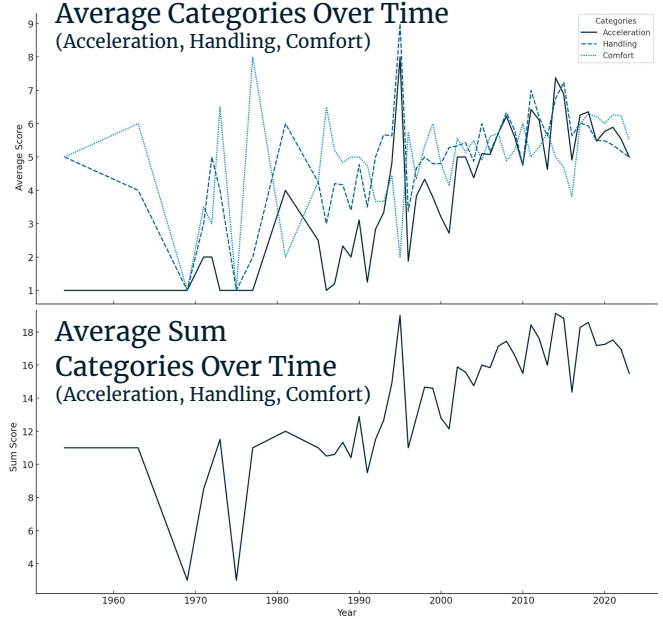
This means that fun factor and value **don't have a strong linear relationship**.

Suggests that these two factors are largely **independent** of each other.

	Fun Factor	Practicality	Value	Fun Factor, Practicality, Value
Top 1	McLaren Speedtail	Tesla Model X	McLaren F1	Rivian R1T
Top 2	McLaren F1	Kia Carnival SX Prestige	Ford Raptor	GMC Hummer EV Pickup
Top 3	Bugatti Divo	Pacifica Pinnacle	Porsche 911 Carrera 4S	Dodge Ram TRX
Top 4	Bugatti Chiron Pur Sport	Sienna Platinum	Chevrolet Corvette Z06 C5	Ford Raptor
Top 5	Bugatti Chiron	Aztek	Fiat 500 Abarth	Jeep Gladiator Mojave

Relationship Analysis

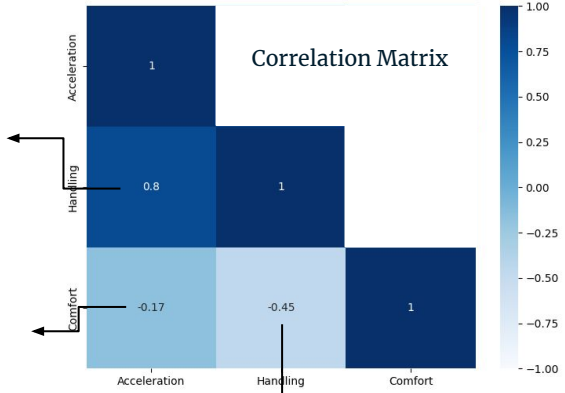
Acceleration, handling, comfort



While modern vehicles are **enhancing their performance** in terms of acceleration and handling, it appears there are inherent design trade-offs or challenges that sometimes **lead to a compromise on the comfort front**. This underscores the complexities manufacturers face when balancing performance dynamics with comfort in vehicle design.

Strong positive correlation. Suggests that cars that **score high in acceleration tend to also score high in handling**, and vice versa.

Indicates that cars with **higher acceleration might slightly compromise on comfort**, but the relationship is not very strong.

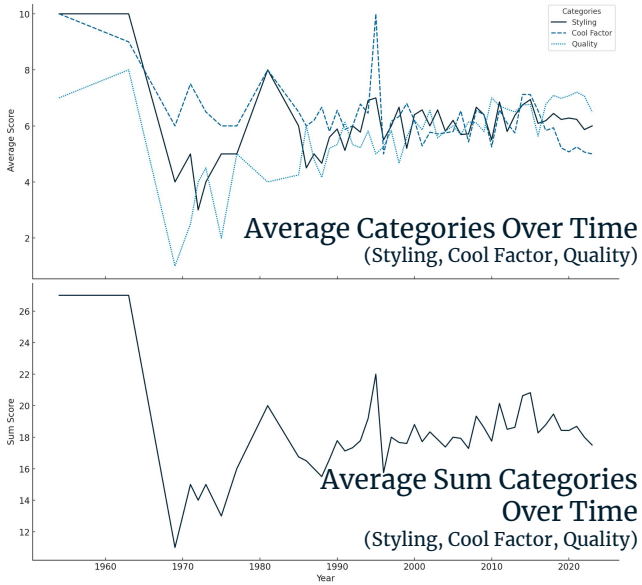


Suggests that vehicles that **score high in handling often have a compromise in terms of comfort.**

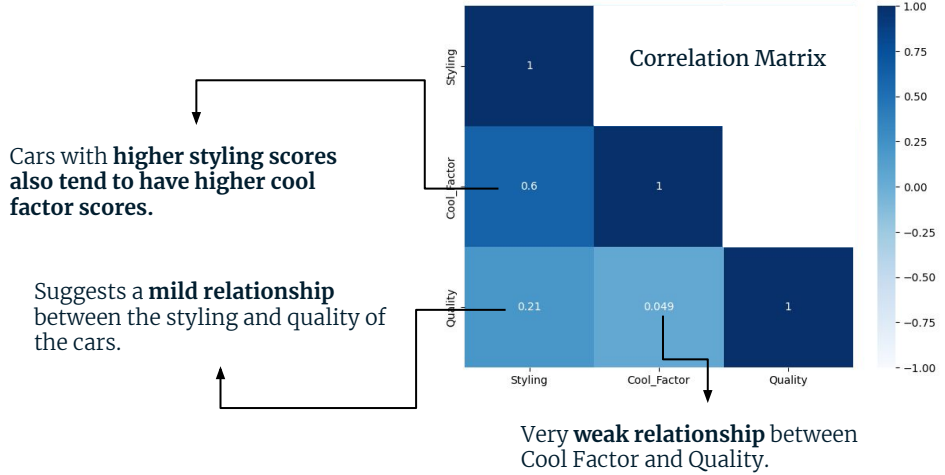
	Acceleration	Handling	Comfort	Acceleration, Handling, Comfort
Top 1	McLaren Speedtail	McLaren Speedtail	Rolls-Royce Cullinan Black Badge	McLaren Speedtail
Top 2	Lucid Air Dream Edition	McLaren F1	Rolls-Royce Cullinan	Lucid Air Dream Edition
Top 3	Tesla Model S Plaid	Ferrari Monza SP1	Rolls-Royce Phantom	Bugatti Divo
Top 4	Bugatti Divo	Rimac Nevera	Rolls-Royce Dawn	Bugatti Chiron Pur Sport
Top 5	Bugatti Chiron Pur Sport	Koenigsegg Regera	Rolls-Royce Wraith	Bugatti Chiron

Relationship Analysis

Styling, cool factor, quality



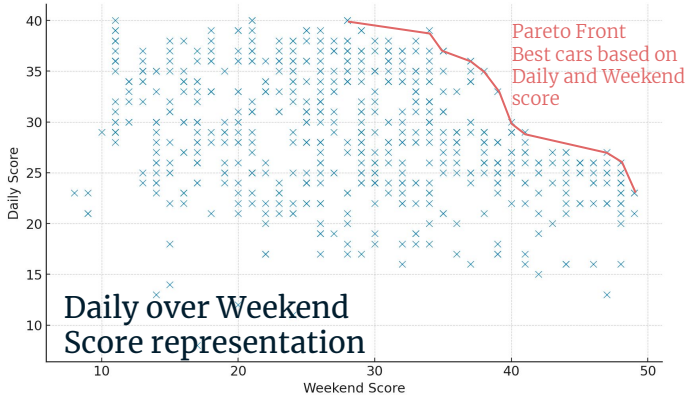
Since the 1990s, vehicle scores for Styling, Cool Factor, and Quality have remained relatively **stable**, with a modest upward trend. The data highlights a consistent emphasis by manufacturers on aesthetics, innovation, and build quality. A positive relationship exists between styling and cool factor, while the link between styling and quality is milder. Overall, modern cars, with their **blend of design appeal and reliability, align well with reviewer preferences and evolving market expectations.**

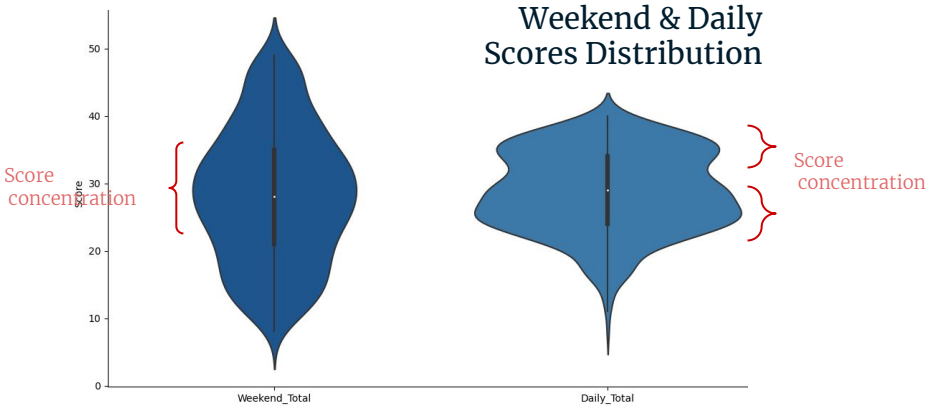


	Styling	Cool factor	Quality	Styling, Cool factor, Quality
Top 1	Aston Martin Lagonda Taraf	McLaren Speedtail	Lexus LC500 Convertible	Porsche Carrera GT
Top 2	Aston Martin Vanquish	McLaren F1	Lexus LC 500	McLaren F1
Top 3	BMW Z8	Bugatti Divo	Lexus LS 500h	Ferrari Monza SP1
Top 4	Ferrari F40	Bugatti Chiron Pur Sport	McLaren F1	Ferrari 250 GT Lusso
Top 5	Vector W8	Bugatti Chiron	Porsche Carrera GT	Mercedes-Benz 300SL Gullwing

Relationship Analysis

Weekend score & Daily score





Correlation
-0.31

Daily Total 50.5 %

Weekend Total 49.5 %

Min Weekend Score
8

Max Weekend Score
49

Min Daily Score
8

Max Daily Score
40

	Weekend Score	Daily Score
Top 1	Ferrari Monza SP1	Rolls-Royce Cullinan Black Badge
Top 2	Ferrari LaFerrari	Rolls-Royce Cullinan
Top 3	McLaren Elva	Jeep Grand Wagoneer
Top 4	McLaren Speedtail	BMW X7
Top 5	Bugatti Chiron Pur Sport	Kia Carnival SX Prestige

The scatter plot with its **Pareto front** highlights models that excel in both everyday functionality and weekend fun, ideal for buyers seeking a car that satisfies all aspects of their lifestyle. The violin plot further delves into how different vehicles stack up in terms of daily and weekend scores, revealing the typical ranges and concentration of these ratings. This comparison is crucial for understanding the common trade-offs made by vehicles between practicality and enjoyment.

Relationship Analysis

Weekend score, daily score, brand

Unique Brand

84

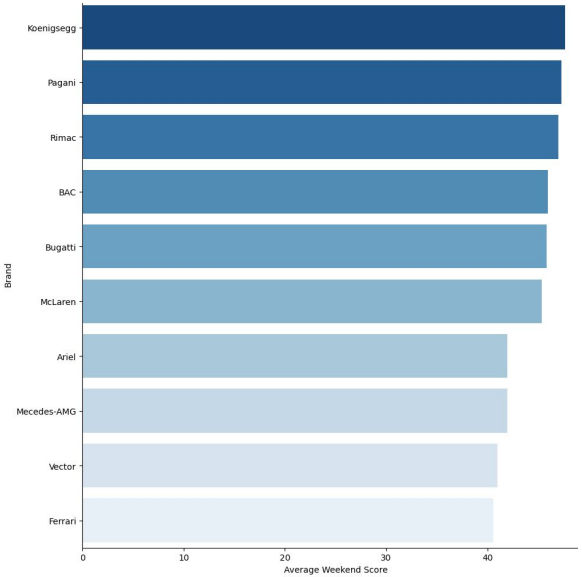
Correlation by Brand

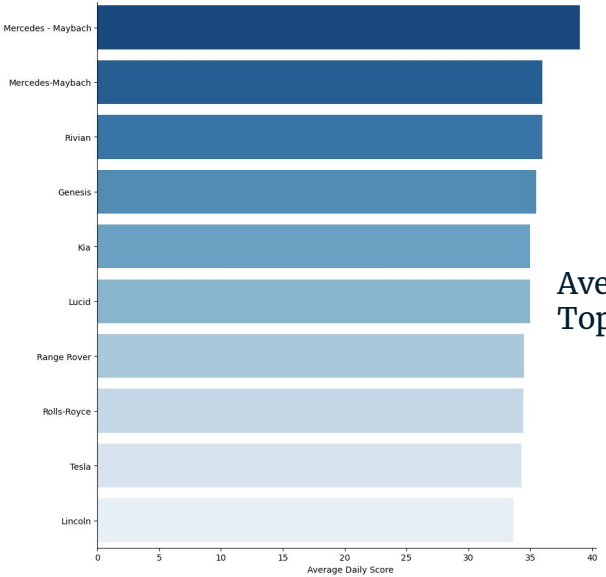
−0.12

This correlation is based on the average scores for each brand. It's a **weaker negative correlation** compared to the individual vehicle correlation *(-0.31, slide 9)*. This suggests that, on average, brands with a higher Weekend Score don't necessarily have much lower Daily Scores and vice versa.

The reduced strength in correlation may indicate that while individual models within a brand might lean heavily towards weekend appeal or daily usability, the brand as a whole might offer a more balanced portfolio of vehicles.

Average Weekend Score Top Cars Model

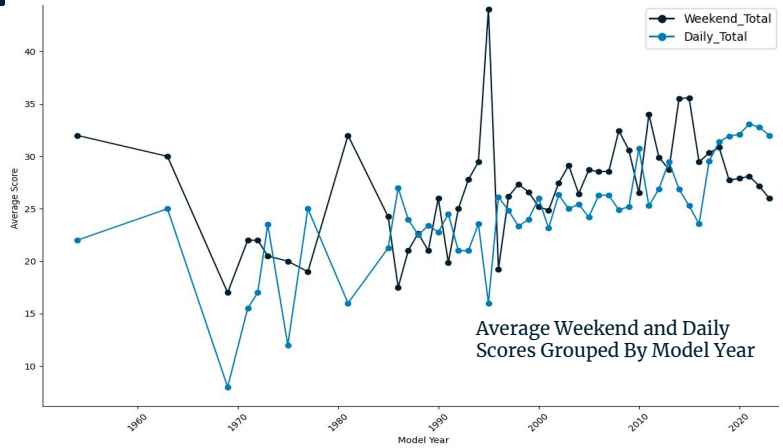


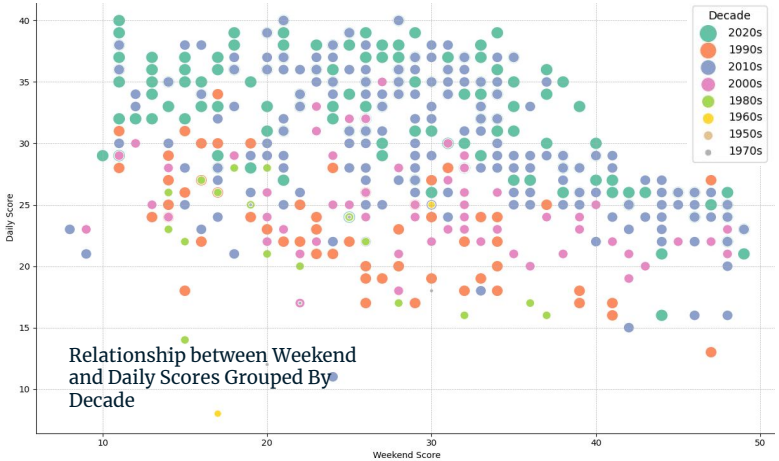


Average Daily Score Top Cars Model

Relationship Analysis

Weekend score, daily score, model year





Correlation by Model Year	Eta Squared (η^2) Correlation (Weekend score & Year)	Eta Squared (η^2) Correlation (Daily score & Year)
0.195	**0.0996**	**0.4658**

The scatter plot gives us a unique view of car scores over the decades. Think of **each dot as a car, and the bigger the dot, the newer the car's model year.**

What stands out from the plot:

- **Newer Cars Shine:** Cars from recent years, especially the 2020s, generally have better scores for both everyday use and weekend fun.
- **Older Cars Have Trade-offs:** For cars from the 1990s and 2000s, if they were good for weekends, they weren't as practical for daily use, and vice versa.

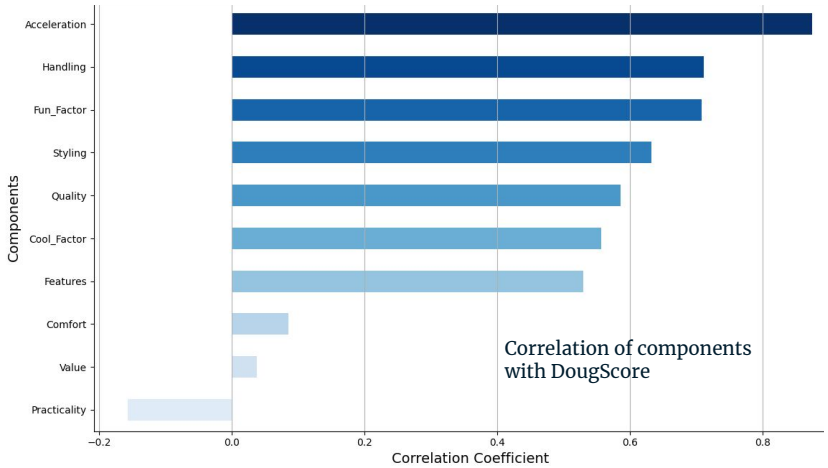
Now, let's dive a bit into the numbers:

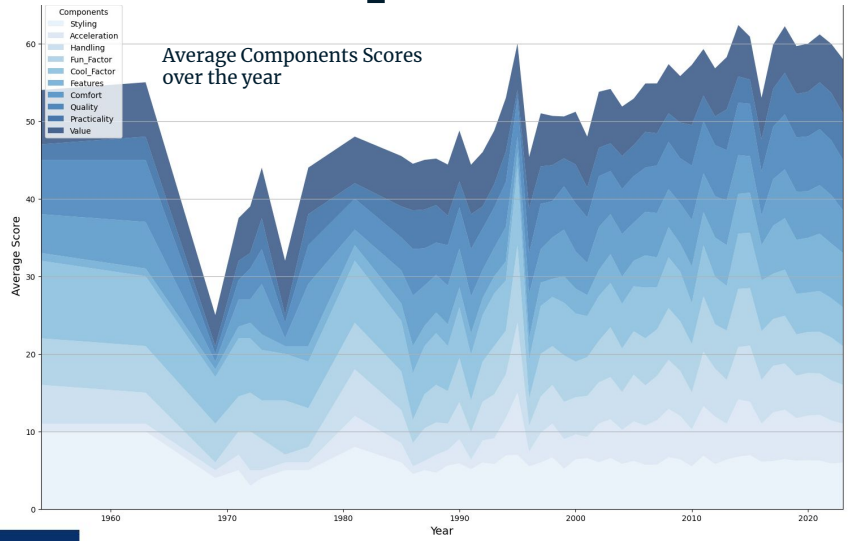
- When we **group the cars by their model year**, the **average correlation** between Weekend Score and Daily Score is **0.195**. This means that, on average, there's a positive but relatively weak relationship between how cars score for weekends and daily use within each year group.
- For the weekend scores, **only about 10% (9.96% to be exact) is influenced by the car's model year**. This means that the Weekend Score of a car, which reflects its appeal for leisure and fun, isn't necessarily tied to how new it is.
- On the other hand, for daily use scores, **a whopping 46.58% is influenced by the car's model year**. So, newer cars tend to be more practical for everyday tasks.

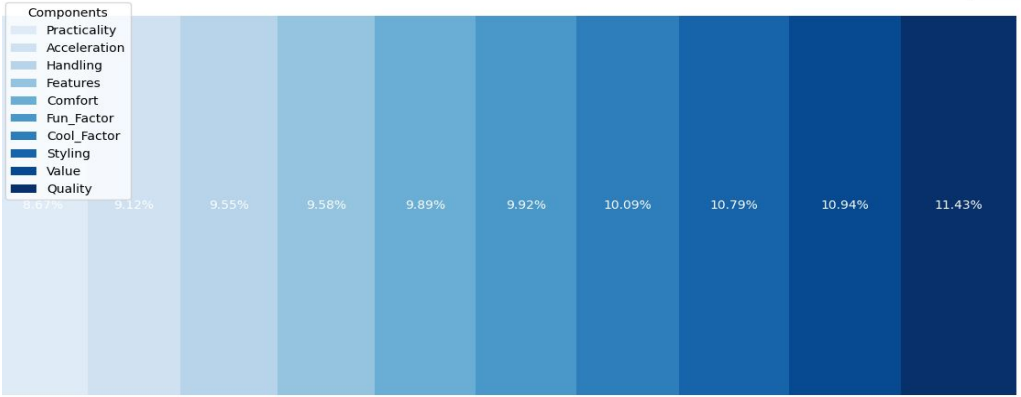
What does this mean for businesses and consumers?

As cars have evolved, it seems manufacturers have focused more on making cars practical, comfortable, and valuable for everyday use. They're responding to what most people want: cars that are reliable for daily tasks but can still offer some fun during weekends.

Interaction effects & non-linear relationships



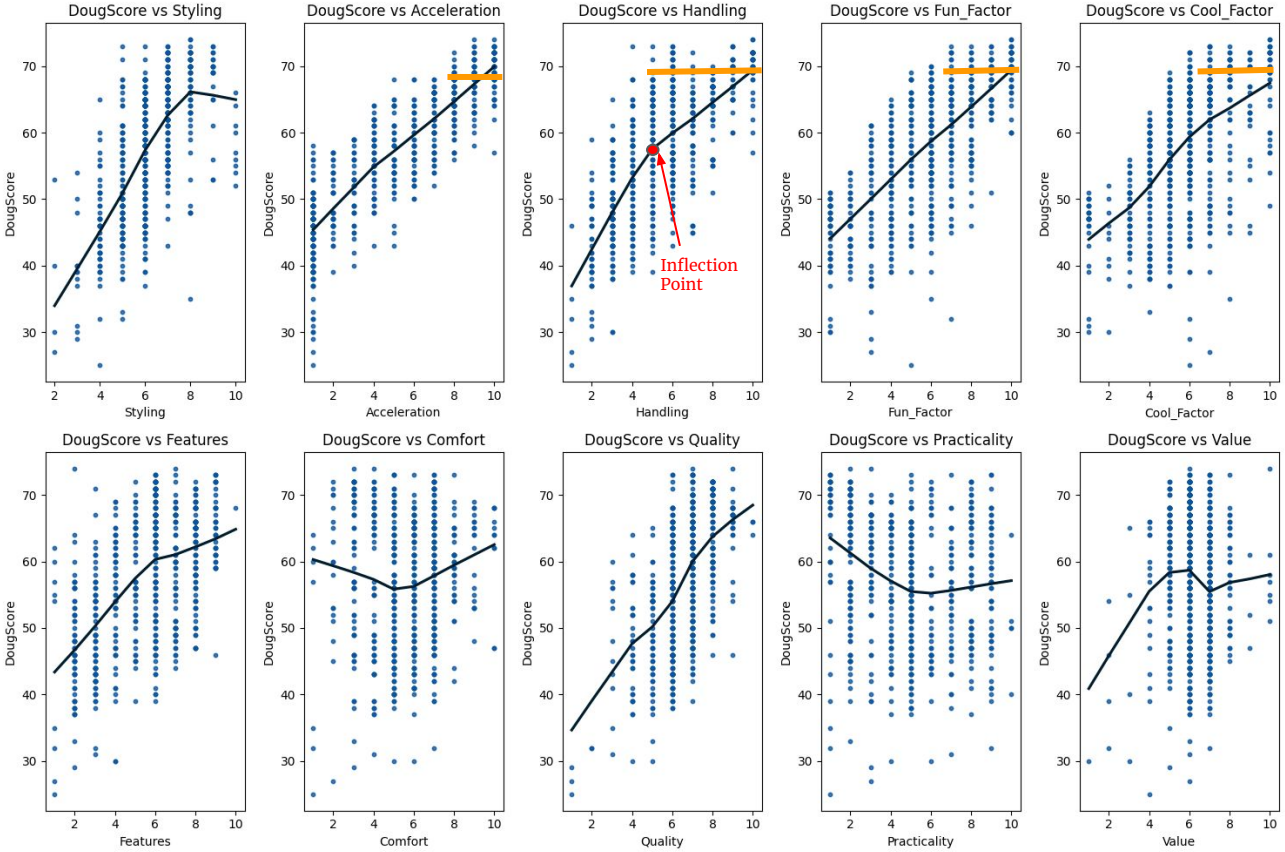




Components average % representation on the DougScore

In the DougScore dataset, **Quality (11.43%)**, **Value (10.94%)**, and **Styling (10.79%)** emerge as paramount components. The data reveals a strong interplay between **Handling, Acceleration, and Fun Factor, with high scores** in one often predicting high scores in the others. However, a trade-off is observed between Practicality and Value. Notably, there's an **upward score trend since the 1980s**, signaling consistent improvements in vehicles. This uptrend, combined with the prominent representation and correlation patterns, indicates **that while intrinsic quality, value, and aesthetics dominate evaluations, a vehicle's performance and practicality significantly influence its overall appeal.**

Interaction effects & non-linear relationships

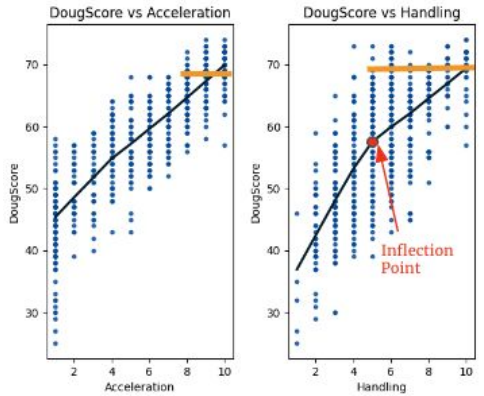


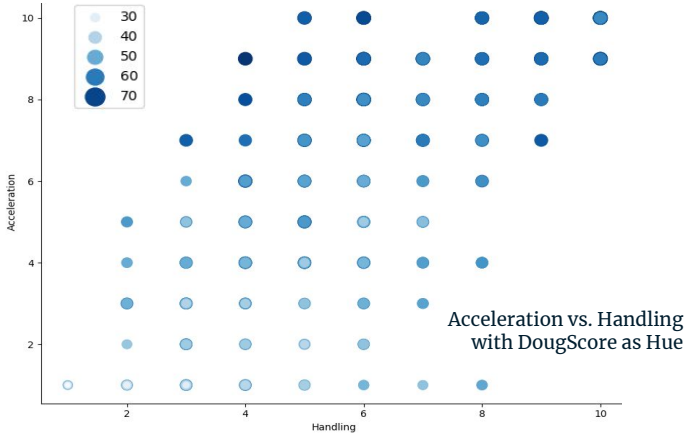
These scatter plots **reveal the impact of various features on DougScore**. Noticeable trends, such as the clear positive correlation between Acceleration and DougScore, suggest strong influences. In contrast, features like Comfort show varied impacts, with **plateaus and inflection points indicating different thresholds for influencing overall scores**.

Let's delve deeper into this analysis for more nuanced understanding.

Interaction effects & non-linear relationships

Acceleration & Handling





Acceleration vs. Handling with DougScore as Hue

Acceleration's Impact
For every 1-point increase in the Acceleration score, the overall DougScore increases by approximately **3.42 points**.
Relevance: Acceleration plays a significant role in determining a car's appeal. A faster car tends to be more appealing, especially for performance enthusiasts.

Handling's Impact
A 1-point boost in the Handling score corresponds to an increase of about **1.08 points** in the DougScore.
Relevance: Good handling ensures a smoother and safer driving experience. For consumers, this translates to better control and overall driving comfort.

Acceleration-Handling Interaction Impact
The combined effect of Acceleration and Handling slightly diminishes their total impact on the DougScore by about **0.15 points**.
Relevance: Cars that excel in both acceleration and handling **might face diminishing returns in terms of overall appeal**. For businesses, this suggests a balanced focus on both attributes might yield better results.

Methodology
Linear regression was employed to derive relationships, we introduced an **interaction term** in the regression model to capture combined effects.

The relationships and interaction effects are **statistically significant** with p-values close to 0 (< 0.05), ensuring the findings are not due to random chance.

For manufacturers: *Focusing on improving both acceleration and handling is key, but it's essential to strike a balance to maximize appeal.*
For consumers: *When evaluating cars, consider both acceleration and handling as they significantly influence the overall appeal and driving experience.*

Interaction effects & non-linear relationships

Acceleration & Handling

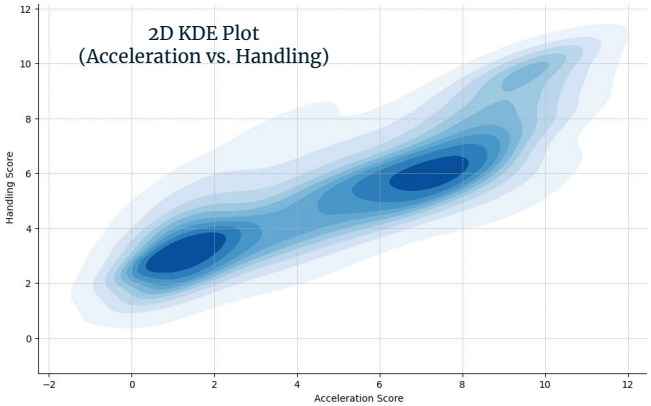
McLaren P1 GTR

Acceleration: 10
Handling: 10
DougScore: 64

Dodge Challenger SRT Demon

Acceleration: 10
Handling: 5
DougScore: 65

***Note**: With a perfect acceleration score but a moderate handling score, this car still receives a high DougScore. This showcases the significant influence acceleration has on the overall score.*



In our analysis of Doug DeMuro's car review scores, we've observed **distinct plateaus** in the impact of Acceleration and Handling on the overall DougScore. Specifically, after vehicles achieve an Acceleration score of around 8-9 and a Handling score starting from 6, **further improvements in these metrics don't seem to significantly elevate the overall DougScore**.

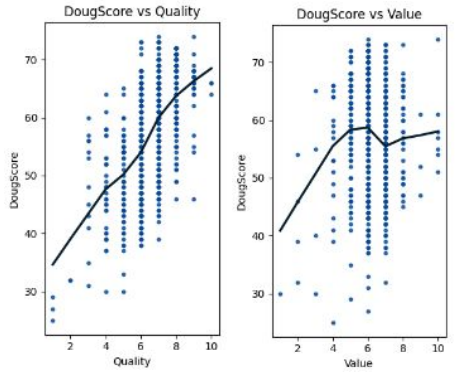
One hypothesis behind this pattern is the inherent limitation of the DougScore scale. It's possible that once a vehicle reaches a certain threshold *(Acceleration: 8, Handling: 6)* of performance, additional enhancements might not markedly increase its perceived quality or appeal. From a consumer's perspective, while high performance is undoubtedly desirable, **there may be a limit beyond which extra gains in acceleration or handling aren't as discernible or valuable for daily driving**. Furthermore, as vehicles achieve top-tier performance scores, other attributes like comfort, features, or overall quality might become more influential in determining the overall DougScore, leading to **observed plateaus in performance metrics.**

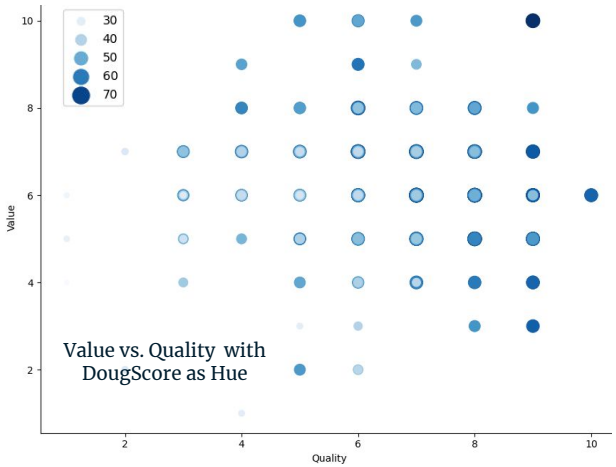
The **2D KDE plot** provides further insights. It reveals **two primary density regions** centered around *(Acceleration: 1 & Handling: 3)* and *(Acceleration: 8 & Handling: 6)*. These regions indicate common combinations of Acceleration and Handling scores among vehicles. The vehicles in the latter, denser region likely represent a segment of performance-oriented cars that strike a balance between speed and agility. Additionally, there's a hint of a smaller cluster around the top scores *(Acceleration: 10, Handling: 10)*, which represent elite vehicles that excel in both metrics.

It provides a clear visualization of market concentrations, enabling informed business decisions aligned with prevalent trends and consumer preferences.

Interaction effects & non-linear relationships

Value & Quality





Value vs. Quality with DougScore as Hue

Quality's Impact
Quality has a strong positive impact on the DougScore, with a significant p-value less than 0.05.

Values' Impact
Value has a positive impact on the DougScore. However, even if value increases, the trend is not entirely linear, and both linear and quadratic fits are not relevant. From a statistical perspective, it's not possible to conclusively state that as value goes up, the DougScore consistently goes up.

Value-Quality Interaction Impact
An interaction term between Value and Quality was included in the regression model. The coefficient for the interaction term is **-0.7178**. This indicates that the effect of Quality on the DougScore decreases by about 0.718 points for each one-unit increase in Value (and vice versa). The negative coefficient suggests that there's a **diminishing return when both Quality and Value are high**.

Our findings show that while value and quality both play a role, the relationship isn't straightforward, and other factors may also come into play.

Interaction effects & non-linear relationships

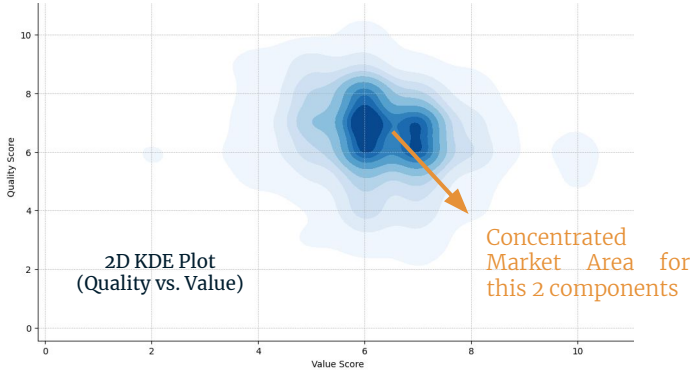
Value & Quality

Rivian R1T

Quality: 6 (Average)
Value: 8 (High)
DougScore: 73

McLaren F1

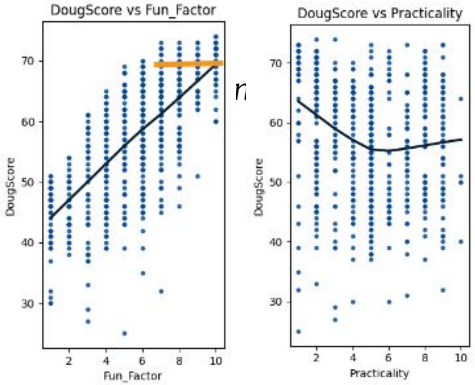
Quality: 9 (High)
Value: 10 (High)
DougScore: 74

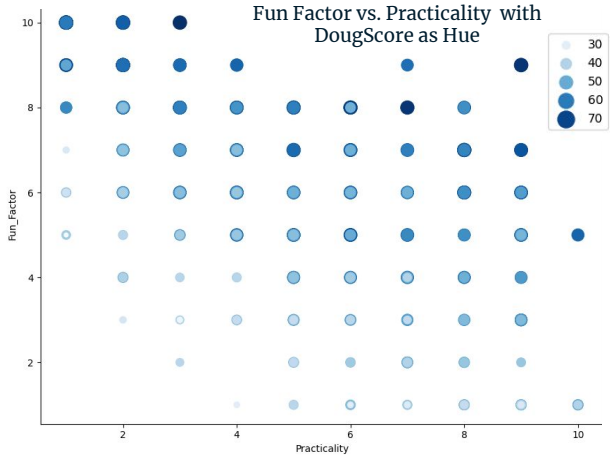


The importance of **Value increases as the score rises, but only up to a point**. After that, even significant improvements in value **don't drastically increase** the overall score.
Quality matters consistently in car ratings. While its **impact also diminishes after a certain level**, it remains influential across the score range.
While it's great to have a car with high Value and Quality, **improvements in these areas alone don't always lead to a substantially higher rating**. It's essential to balance these with other factors like performance, style, and fun.

This is exemplified in our case studies: the Rivian R1T, with an average Quality score and high Value, achieves a total score of 73, underscoring the limited influence of high Value scores. In contrast, the McLaren F1's high scores in both categories only slightly elevate its total score to 74, **indicating the nuanced but robust impact of these factors.**

Interaction effects & non-linear relationships





Practicality's Impact
Impact: For every 1-point increase in the Practicality score, the overall DougScore initially increases by about 3.86 points. However, this impact diminishes, as indicated by the **squared term** of Practicality (-0.18).
Relevance: Practicality, encompassing aspects like storage space, comfort, and daily usability, is important for everyday drivers and family car buyers. However, the diminishing returns suggest that beyond a certain point, additional practicality doesn't significantly enhance a car's overall appeal.

Fun Factor's Impact
Impact: A 1-point increase in the Fun Factor score corresponds to an **increase of about 2.71 points** in the DougScore, with an additional 0.13 points for each square of the Fun Factor score.
Relevance: The Fun Factor, reflecting a vehicle's excitement and engagement levels, plays a crucial role in its overall score. Unlike Practicality, the impact of Fun Factor on DougScore increases, albeit at a **slower pace at higher scores**, highlighting its continued importance in performance and luxury vehicles.

Practicality-Fun Factor Interaction Impact
Impact: The interaction between Practicality and Fun Factor isn't explicitly calculated in our model, but the significant negative correlation (-0.623) between them implies that high scores in one may often be at the expense of the other.
Relevance: This suggests that cars are often designed with a focus on either Practicality or Fun Factor, but rarely both.

Methodology
Approach: We used **linear and quadratic regression** models to understand the relationships, accounting for non-linear effects with squared terms.
Statistical Significance: The relationships and non-linear effects are statistically significant ($p < 0.05$), indicating that the findings are robust and not due to random chance.

***For Manufacturers**: Understanding the diminishing returns of Practicality and the continually important, albeit slower-growing impact of Fun Factor is crucial in designing vehicles that cater to specific market segments.*
***For Consumers**: When considering a vehicle, it's important to weigh Practicality against Fun Factor based on personal needs and preferences. High Practicality might not always equate to a high overall appeal, and the thrill of a fun driving experience continues to be a significant factor in overall satisfaction.*

Interaction effects & non-linear relationships

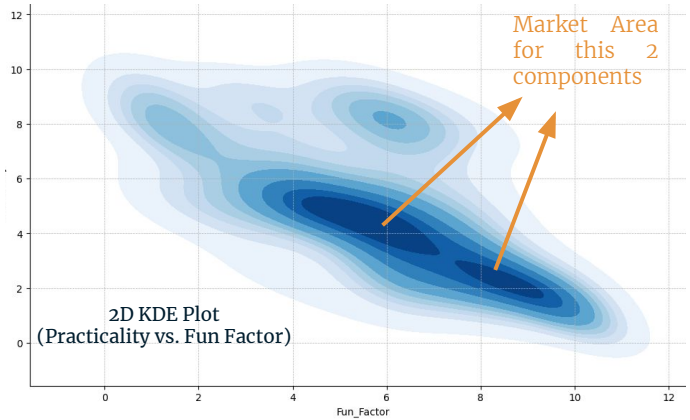
Practicality & Fun Factor

2020 McLaren Speedtail

Fun Factor: 10
Practicality: 2
DougScore: 74

2022 Kia Carnival SX Prestige

Fun Factor: 1
Practicality: 10
DougScore: 51



The McLaren Speedtail, with its perfect Fun Factor score, is a prime example of a high-performance, luxury sports car. It's designed for exhilarating speed and handling, evident in its max scores in Acceleration and Handling. However, its Practicality score is low, reflecting limited daily usability and comfort.

In stark contrast, the Kia Carnival SX Prestige scores exceptionally high in Practicality, making it an excellent choice for families or those needing spaciousness and utility. Its Fun Factor, however, is minimal, indicating a lack of the thrilling driving experience that the McLaren offers.

Performance vs. Utility: The McLaren Speedtail and Kia Carnival represent two ends of the automotive spectrum - performance and utility. The McLaren is all about speed, style, and driving thrill, while the Kia focuses on comfort, space, and practical daily use.
DougScore Differences: Despite their stark differences in Fun Factor and Practicality, both cars score relatively well overall, highlighting Doug DeMuro's balanced approach to scoring that accommodates different vehicle types and consumer needs.

Clustering & Consumer-Business Impact

Proposing similar cars to a client who likes a specific model can be highly beneficial from both a business and consumer perspective.

For Consumers:
Broader Options: Consumers may have a preferred model in mind, but they might not be aware of **similar alternatives** that offer comparable features, performance, or value. Providing options ensures they make the most informed decision.
Better Fit for Needs: While a consumer might like a particular car, there may be alternatives that better fit their specific needs, budget, or preferences.
Price and Availability: A preferred model might not be readily available or could be priced outside a consumer's budget. Similar models might offer more practical choices in terms of availability or cost.
New Discoveries: Consumers may discover models they hadn't considered or were unaware of, which could be more appealing once compared side by side with their initial choice.

For Businesses:
Increased Sales Opportunities: By presenting customers with similar options, businesses can **increase the likelihood of a sale**, especially if the initial choice is unavailable or unsuitable.
Customer Satisfaction: Providing customers with more choices that align with their preferences enhances customer satisfaction, leading to better reviews, repeat business, and referrals.
Market Intelligence: Understanding which cars are similar helps businesses in inventory management, marketing strategies, and identifying emerging market trends.
Competitive Edge: Businesses that can effectively match customers with the right product based on sophisticated analysis (like clustering similar vehicles) can differentiate themselves from competitors.

Augmented DougScore with NLP ?

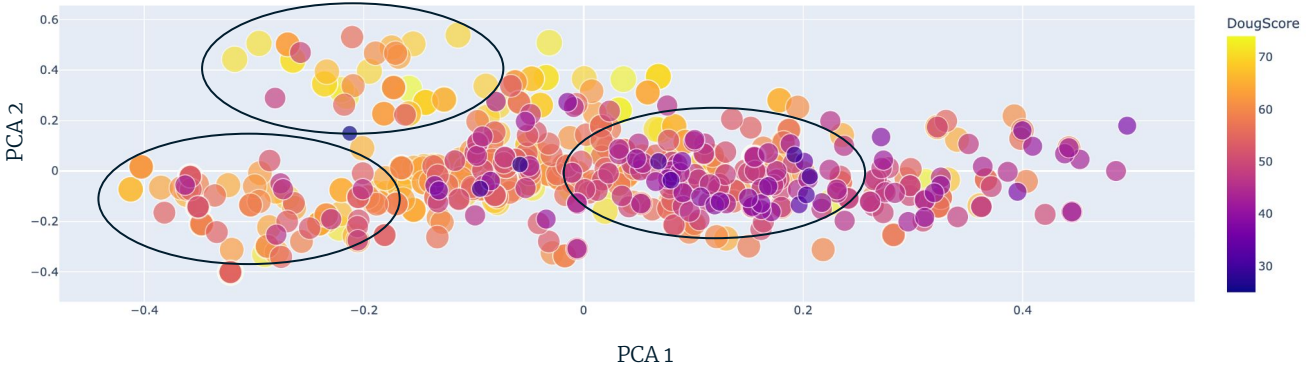
The development of **Large Language Models** (LLMs) and libraries like **LangChain** has revolutionized how we approach text-based applications. In the automotive context, such advancements unlock deep value by extracting rich information about a **car's history, global reputation, and various nuanced details.** We can now envisage a car recommender system finely tuned to an owner's preferences, previous selections, and feedback. This innovation would enable highly personalized and pertinent recommendations. Imagine a tool where, upon a user's query, it sifts through extensive data to present the most relevant cars, **crafting tailored suggestions**. This approach not only refines the user experience but also aids in understanding evolving automotive trends and consumer inclinations, offering a more sophisticated and insightful perspective in the realm of car recommendations.

Methodology:
We aim to extract **Wikipedia descriptions** for each vehicle model, detailing its history, production, and accolades, as exemplified by the Porsche 918 Spyder. Using a **SentenceTransformer** model *(Contextualization model)*, these descriptions will be embedded to identify clusters and patterns, and potentially correlate with the DougScore. Additionally, we'll employ Principal Component Analysis (PCA) to reduce these embeddings to two dimensions, facilitating an insightful visualization of the clustering.

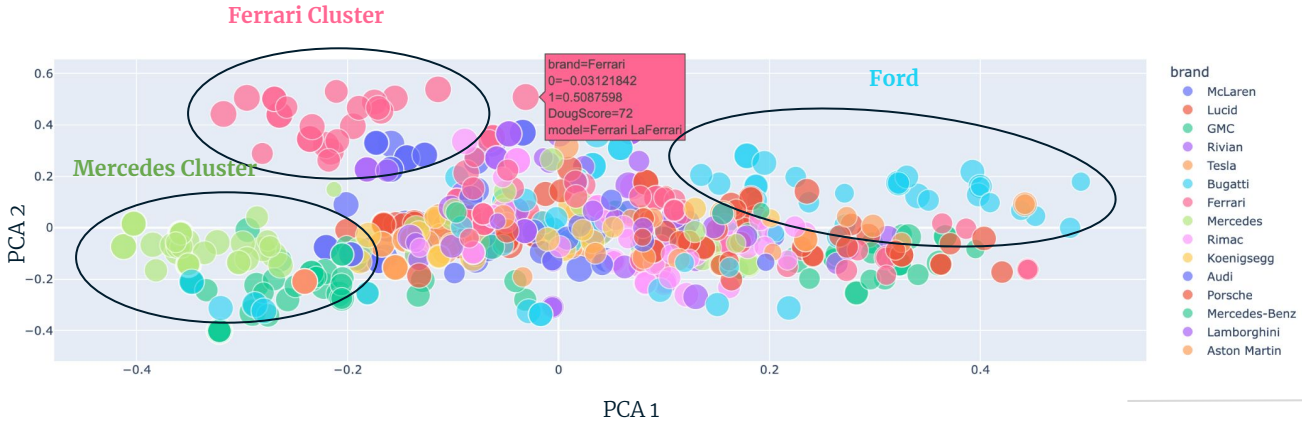
Wikipedia description for Porsche 918 Spyder
*(Extracted with **LangChain** Librairy)*

The Porsche 918 Spyder is a sports car manufactured by German marque Porsche. The 918 Spyder is a plug-in hybrid powered by a mid-mounted naturally aspirated 4.6 L (4,593 cc) V8 engine, developing 447 kW (608 PS, 599 hp) at 8,700 RPM, with two electric motors delivering an additional 210 kW (286 PS, 282 hp) for a combined output of 652 kW (875 hp) and 1,280 N·m (944 lbf·ft) of torque. The 918 Spyder's 6.8 kWh lithium-ion battery pack delivers an all-electric range of 19 km (12 mi) under the US Environmental Protection Agency's five-cycle tests.Production began on 18 September 2013, with deliveries initially scheduled to begin in December 2013, and a starting price of ≈€781,000 (US$845,000 or £711,000). The 918 Spyder was sold out in December 2014 and production ended in June 2015.The 918 Spyder was first shown as a concept at the 80th Geneva Motor Show in March 2010. On 28 July 2010, after 2,000 declarations of interest, the supervisory board of Porsche AG approved series development of the 918 Spyder. The production version was unveiled at the September 2013 Frankfurt Motor Show. Porsche also unveiled the RSR racing variant of the 918 at the 2011 North American International Auto Show, which combines hybrid technology first used in the 997 GT3 R Hybrid, with styling from the 918 Spyder. However, the 918 RSR did not make it to production. The 918 Spyder was the second plug-in hybrid car manufactured by Porsche, after the 2014 Panamera S E-Hybrid.

Augmented DougScore with NLP ?



The clustering observed is **closely aligned with the DougScore**, categorizing into High, Middle, and Low score groups based solely on the vehicle descriptions. This suggests that the **textual descriptions effectively capture essential information** about the car's features, quality, and performance, reflecting in their respective DougScores.



Propose similar cars models based on their characteristics, history, features...

Conclusion

The analysis of the DougScore dataset paints a detailed picture of the automotive landscape, highlighting a significant trend: **cars today are increasingly engineered to marry performance with everyday usability**. This evolution in vehicle design reflects a broader shift in consumer priorities, where the demand for multifunctional, efficient, and performance-oriented cars is becoming more prominent.

In addition to DougScore, leveraging Natural Language Processing (NLP) to examine car descriptions offers a unique glimpse into how **vehicles are perceived and described online**. This approach underscores the potential of NLP not just in extracting meaningful patterns from textual data, but also in **enriching our understanding of market trends and customer preferences**, particularly in complex industries like automotive.

For businesses and consumers, the insights derived from this analysis are invaluable. By understanding these evolving trends and preferences, companies can strategize product development, marketing, and customer engagement more effectively. Consumers can **utilize this data to make informed decisions**, aligning their purchases with vehicles that best meet their evolving needs. Essentially, this kind of report analysis empowers both sides to navigate the automotive market with greater awareness and strategic foresight.

Ferrari Purosangue (2022)
(DougScore to perform in the future)