

**INSY 669-075**

Text Analytics

Presented to Professor Changseung Yoo

Group Assignment 1

Prepared by:

|  |  |
| --- | --- |
| Ali Al Samuraee | 260739157 |
| Cristina Esposito | 260744222 |
| Alya Gabsi | 260771940 |
| Siraj Hatoum | 260509637 |
| Yashica NA | 260945954 |

Wednesday, February 2nd, 2022

Contents

[**Overview** 1](#_Toc94702863)

[**1.0 Introduction and Industry Analysis** 1](#_Toc94702864)

[Data Overview 2](#_Toc94702865)

[Data Preprocessing 3](#_Toc94702866)

[**2.0 Top 10 Brands and The Associations Between Them** 3](#_Toc94702867)

[**3.0 Top 5 Attributes and Their Associations with The Top 5 Brands** 6](#_Toc94702868)

[Additional Data Processing 6](#_Toc94702869)

[Results 6](#_Toc94702870)

[**4.0 Managerial Advice** 9](#_Toc94702871)

[**5.0 Aspirational Brand Analysis** 11](#_Toc94702872)

[**Conclusion** 13](#_Toc94702873)

# **Overview**

This report outlines the results and insights from a text-based analysis performed on user comments from an Edmund’s online forum on sedans. The analysis relies on brand and attribute mentions and co-mentions to identify patterns and trends in users’ top of mind brand associations as well as brand and attribute associations.

Through this analysis we identify the top 10 most frequently mentioned brands, the lift values indicating the association between them, and explore the most interesting insights from these results. Additionally, a similar analysis is undertaken for the association between the top 5 most mentioned brands and the 5 most frequently co-mentioned attributes. We focus on explaining insights that could inform the product and marketing departments of each of the 5 brands to improve their market sentiment.

Finally, we identify and discuss which of the op 5 brands is most closely viewed as an aspirational brand by the forum users based on the results of the above analysis.

# **1.0 Introduction and Industry Analysis**

The focus of this analysis will be on midsize sedans. The history of the midsize sedan dates to the late 1950’s, as it was introduced as a new size classification created by the US to distinguish itself from compact cars. Major players within the market at the time were AMC (American Motor Company), Ford, and Chevrolet (Severson, 2009). In today’s market, Ford remains a top manufacturer, although Asian brands like Honda and Toyota have largely taken over this market (Egan, 2021). Midsize sedans make up the second largest product segment at 41.6% of the US automobile industry, with the largest subsegments being compact and subcompact cars at 47.2% of the market (Egan, 2021).

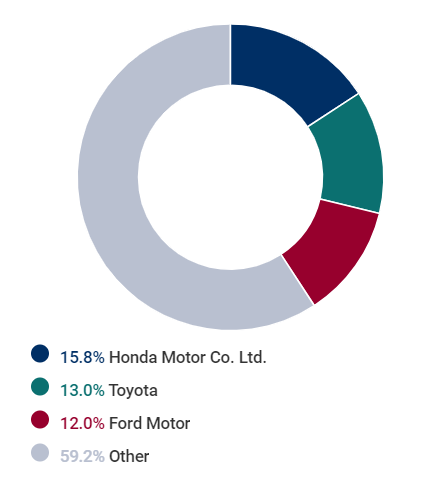
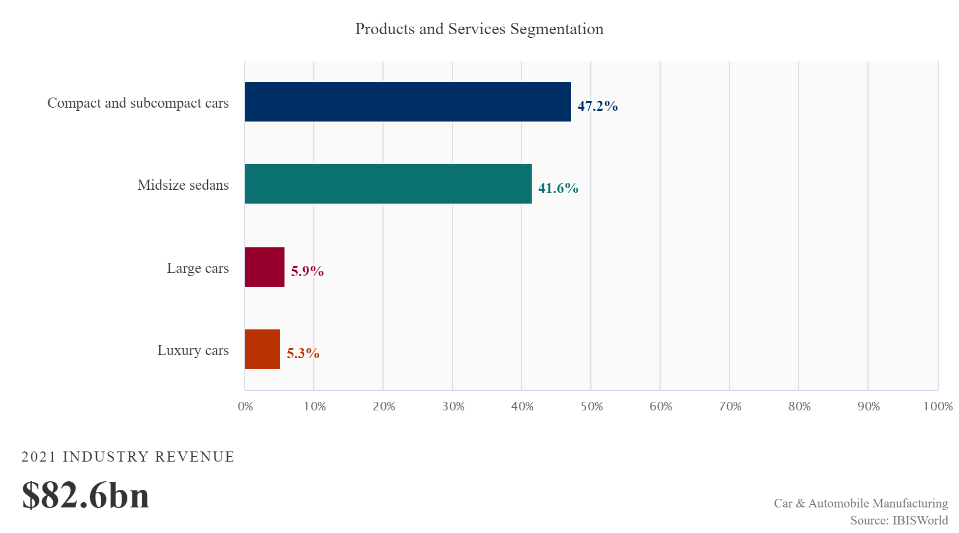
 

Figure 1 Brand and Product Segmentation Market Share within American Automotive Industry

Specifically for midsize sedans, Edmunds.com has released their top 10 midsize models for 2022. Within this list, Kia comes out on top, followed by Honda and Mazda (Best Sedans of 2022 and 2023 - Expert Reviews and Rankings, n.d.). Within the top 5 models, Kia is the only brand to show up more than once, further demonstrating that they are dedicated to producing some of the best midsize cars. With this ranking, it would be safe to assume that Kia would be used as a baseline comparison towards other manufacturers. What makes them the best will further be assessed in the next sections.

## Data Overview

For the purpose of our analysis, we will be using a forum from Edmunds called “[Midsize Sedans 2.0](https://forums.edmunds.com/discussion/7526/general/x/midsize-sedans-2-0)”. Considering that Edmunds is a website that focuses on cars for the US market, we can safely assume that most users within the forum are in the US and are representative of the sentiment of American the market towards these brands. This forum contains over 27,100 comments from 2007-2021. Our analysis focuses on the comments from 2013-2016 (inclusive), as this was the most recent period of heightened activity on the forum, allowing for the collection of sufficient forum discussions within a condensed time frame. The smaller time window serves to minimize the impact of time variation in consumer preference and forum activity on the results of the analysis, approximating a snapshot in time of market brand perception and top-of-mind association. The total number of comments captured within this range is 7,782.

## Data Preprocessing

To identify the top 10 most frequently mentioned brands, data preprocessing was performed using tokenization, position tagging, and lemmatization to facilitate a binary count of words mentioned in each comment. A modified version of the car model and brand database found on [BackApp](https://www.back4app.com/database/back4app/car-make-model-dataset/all-cars-by-model-and-by-make-and-by-year) was used to group the occurrences of car models with their respective brands, resulting in a binary count of brand mentions in each comment.

# **2.0 Top 10 Brands and The Associations Between Them**

Based on the output of the data pre-processing, the following frequency table was obtained for the 10 most frequently mentions brands:

|  |  |  |
| --- | --- | --- |
| **Brand** | **# Of mentions** | **Country/Continent** |
| Honda | 2833 | Japanese (Asia) |
| Ford | 2342 | American (NA) |
| Infiniti | 2114 | Japanese (Asia) |
| Kia | 1724 | South Korean (Asia) |
| Toyota | 1503 | Japanese (Asia) |
| Acura | 1437 | Japanese (Asia) |
| Hyundai | 1241 | South Korean (Asia) |
| Mazda | 1136 | Japanese (Asia) |
| Chevrolet | 1046 | American (NA) |
| Volkswagen | 810 | German (EU) |

Table 1 Top 10 Brands from Midsize Sedans 2.0 Forum

Aligned with the market-share information, we can see how Asian manufacturers have become the dominant brands within the American market, making up 7 of the top 10 brands mentioned within the forum. This is due to the impact of the 2008 depression, where many American brands faced bankruptcy, further allowing brands from the Asian market to enter into the market and expand their operations within North America (Stoll, 2015). Preferences from American consumers shifted towards fuel efficiency (due to high fuel prices at the time), advancements in technology, and overall quality improvements that which Asian brands were offering (MotorBiscuit.com, 2019).

When it comes to comparing these competitors with each other, we can use the following lift matrix and MDS plots to understand brand positioning views within the forum:

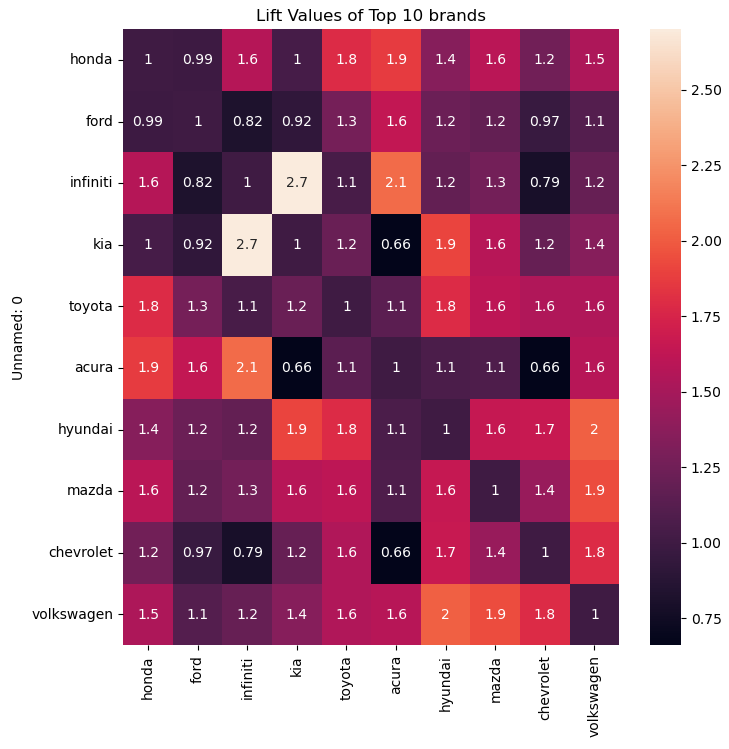


Figure 2 Lift Matrix for Top 10 Brands

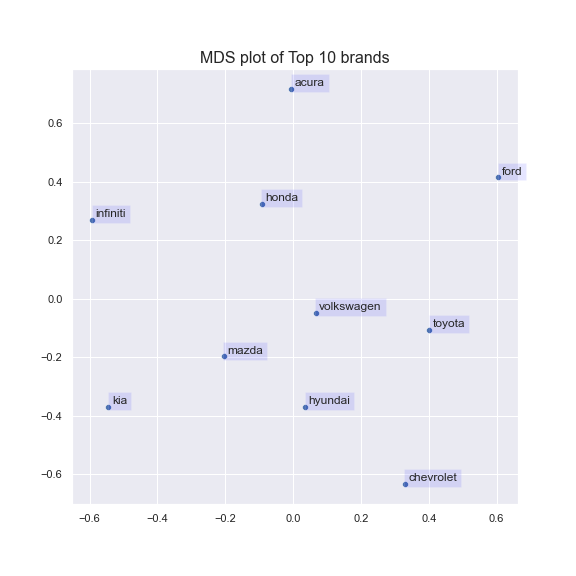


Figure 3 MDS Plot for Top 10 Brands

The most surprising result from these plots is Kia and Infiniti being mentioned together with the highest likelihood. With Kia having an affordable brand positioning, and Infiniti a luxury position, one may wonder how the two managed to end up in the same conversations.

When taking a closer look at Kia, their lift value compared to other luxury brands like Acura is very low. Not surprisingly, they are mentioned together with Hyundai with high likelihood (who owns approximately 34% of Kia) (Hyundai Versus Kia - which is the best?, 2020). Although both companies use similar platforms and components when manufacturing their cars, the designs and engineering are separate, which could lead to consumers having more conversations between the two brands when it comes to comparisons.

When looking at Infiniti, their likelihood of being in the same conversations with other affordable brands like Ford and Chevrolet is very low. Not surprisingly, the likelihood of Infiniti and Acura being in the same conversation is very high, seeing that they are both luxury brands and target similar consumers. Unlike Kia and Hyundai, Infiniti’s parent company, Nissan (Guy, 2020), didn’t make it in the top 10 list.

There may be multiple reasons for the unexpected association between the Kia and Infiniti brands. Since Kia is the top brand for midsize sedans on Edmunds, consumers in the forum may be using it as a baseline comparison and have positive sentiment for Kia in terms of quality and price, making it so that having an Infiniti is not worth the price tag (bargain deal). This association could be an indication that Infiniti’s branding strategy was not be successful in setting it apart from what many may view as the baseline brand. Infiniti’s marketing department may benefit from deeper analysis into this association to assess the points of comparison in the users’ minds and what changes can be made to the product/ marketing campaign to distance Infiniti from these brands.

# **3.0 Top 5 Attributes and Their Associations with The Top 5 Brands**

## Additional Data Processing

To identify which car attributes were most frequently mentioned, filtered out stop words and brand names from our processed data and extracted the top 1000 most frequent tokens. The tokens were further examined to identify and group words that were related to car attributes. This grouping was hen applied to the dataset in a similar manner to the grouping of car models by brand.

## Results

The attribute processing resulted with 13 different attribute groups: age, build, performance, comfort, issues, price/purchase, size, aesthetic, safety, luxury, emotion, friends & family, and condition. The top 5 most frequently mentioned attributes from the analysis are age, build, price/purchase, aesthetic, and luxury. The following is a frequency table that shows the attribute mentions across all brands:

|  |  |  |
| --- | --- | --- |
| **Attributes** | **Criteria** | **Frequency** |
| Age | Includes mentions of model/ purchase year | 5907 |
| Build | Includes vehicle components such as engine, suspension, wheels, etc. | 5125 |
| Price/Purchase | Includes mentions of leasing, price, warranty, and discounts | 3758 |
| Aesthetic | Includes any attributes such as design, paint colours, and look | 3385 |
| Luxury | Includes non-standard vehicle components such as leather and sunroofs | 3290 |

Figure 4 Top 5 Attributes

From these results, we can see that discussion around the age and build of the car are most popular. This is not surprising, considering that when comparisons are made amongst brands within the forum, owners will most likely mention the year the model belongs to (ex: 2013 Honda Accord), as well as components of the car (ex: engine, steering, etc.). This is also inline with the user profile of most people posting on Edmund’s, i.e., car enthusiasts tend to care about the model year and car components as they are knowledgeable enough to differentiate between cars on that basis.

What is surprising from this frequency table is that attributes like issues did not make it within the top attributes. Forums like this one tend to be used by consumers to express problems and their dissatisfaction with brands. It could be that many of these brands mentioned within the forum do not face many issues, showing a true testament to their durability and quality. Additionally, this may indicate that most posters on Edmund’s may not have hands-on experience with the brands in question and are speculating based on marketing and product information, for example, a user posting his opinion on an upcoming car purchase. This furthers the hypothesis that this data is more geared towards overall market sentiment and not brand customer sentiment.

For our top 5 brands (Honda, Ford, Infiniti, Kia, and Toyota), we get the following frequency and lift matrices associated with the top 5 attributes:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Age** | **Build** | **Price/Purchase** | **Aesthetic** | **Luxury** |
| **Honda** | 2443 | 2246 | 1627 | 1447 | 1459 |
| **Ford** | 2008 | 1771 | 1176 | 1123 | 1274 |
| **Infiniti** | 2041 | 2006 | 1195 | 1098 | 888 |
| **Kia** | 1581 | 1493 | 997 | 952 | 757 |
| **Toyota** | 1248 | 1145 | 955 | 822 | 752 |

Table 2 Frequency Table of Top 5 Brands with the Top 5 Attributes

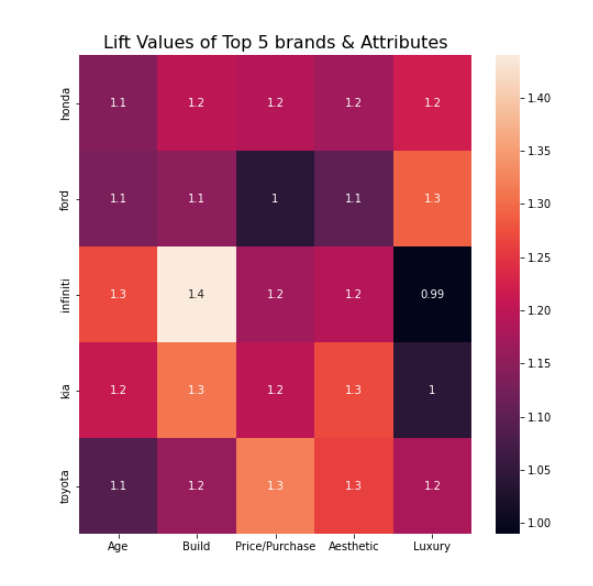


Figure 5 Lift Values for Top 5 Brands and Attributes

In terms of strong association between the attributes and the brands, we can see that age and build are most associated with Infiniti, price and purchase is associated with Toyota, Kia is associated with aesthetic, and Ford for luxury. It is quite surprising to see Ford have the highest likelihood of mentions with luxury, and Infiniti having the least. As the analysis does not include user sentiment, it is unclear if these associations are positive, for example Ford is not positioned as a luxury brand, the association may be attributed to lack of luxury being a key differentiating factor between Ford and some of its competing brands.

Further study incorporating sentiment analysis of these attribute/brand co-mentions would help shed a brighter light on market sentiment. Comparing lift scores of attribute/brand co-mentions when only one brand is mentioned with lift scores when at least two brands are mentioned, may shed light on which attributes are viewed as brand definers and which are brand differentiators.

# **4.0 Managerial Advice**

With this newfound information gathered from our analyses, we have a better understanding of what people are talking about when it comes to midsize sedans. In terms of competitors, comparing affordable vs expensive/luxury brands won’t be of any benefit. This is largely because the top 5 brands, let alone the top 10, are dominated by brands that are typically positioned as “affordable”. This could be because those who use the forum are considered as “average Americans” that will most likely be familiar and using brands that are targeted towards them, such as Kia, Honda, Toyota, and Ford (affordability positioning).

Marketing managers from these brands should focus their advertising strategies towards the “average American” target market, as they make up most users within the forum and who’s insights are largely based on their views and opinions of the brands. These marketing managers should also position themselves against each other, since they seem to be the most well-known brands for midsize sedans and that which consumers are discussing the most.

When looking at the attributes and their association to each of these brands, consumers seem to associate Kia the most with its build and aesthetic, Honda and Ford with build and luxury, and Toyota for its price and aesthetic. To establish the polarity of these associations, the lift between the “Issues” attribute and the top 5 attributes has been calculated for each brand. The results are summarized in the following table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Age** | **Build** | **Price/Purchase** | **Aesthetic** | **Luxury** |
| **Honda - Issues** | 1.03 | 1.08 | 1.147 | 1.232 | 1.192 |
| **Ford - Issues** | 1.01 | 1.11 | 1.267 | 1.368 | 1.166 |
| **Infiniti - Issues** | 1.01 | 1 | 1.177 | 1.197 | 1.301 |
| **Kia - Issues** | 1.02 | 1.06 | 1.196 | 1.125 | 1.25 |
| **Toyota - Issues** | 1.05 | 1.07 | 1.112 | 1.27 | 1.228 |

Table 3: Lift Values for Top 5 Brands and “Issues” for each brand

A high lift value for a specific attribute and brand pair implies that for that specific brand, the attribute in question is often mentioned with issues. In this case, the association would be negative. A low lift value would imply the opposite. In other words, the attribute is usually mentioned with fewer issues which would imply a positive association.

As previously mentioned, Kia is associated with build and aesthetic the most. The lift of both attributes with “Issues” for Kia is the lowest, implying that out of all four brands, Kia is perceived by users of the forum as having the best build and aesthetic qualities. Kia’s pain points seem to be the price paid with a lift between Issues and Price/Purchase of 1.196, making it the second highest lift value for the price attribute. Product managers at Kia should therefore focus on this aspect when releasing new midsize sedans. It is quite interesting how consumers perceive Kia as having good aesthetics and build but are uncomfortable with their prices. This might be because, while the models look aesthetically pleasing, there might exist issues within the build that do not justify the price paid. Kia product managers should investigate why their cars have those issues in the first place and improve build quality accordingly.

Honda and Ford are both strongly associated with build and luxury. However, while Honda is perceived as having more issues with its luxurious attributes and less with its build, Ford is perceived the opposite way. The pain point of both these companies is their aesthetic, and so, product managers should focus on that aspect to improve their product perception. It is interesting to note that Ford has the strongest associations of issues with build, aesthetic, and price. While users of the forum tend to associate Ford as a luxury brand, their prices might be too high for their engine quality and aesthetic. Ford should investigate those aspects to improve their competitive positioning.

Finally, Toyota is strongly associated with price and aesthetic. The lift between issues and price for Toyota is the lowest, meaning that users of the forum see this brand as the best “bang for their buck”. On the other hand, while we might have originally thought that Toyota was perceived as an aesthetically pleasing brand, the data shows that it is the second brand that has the most issues with respect to this attribute. Product managers at Toyota should therefore improve the aesthetic of their cars if they want to compete with the four other brands.

Among these four brands, build and aesthetic seem to be the most prominent attributes and most valued by these customers. Given this information, when introducing new midsize sedans into the market or new features on existing models, a product manager must focus their vision of the product development process on the aesthetic and build, as these are valued the most by customers within the forum, further meeting their “wants” and expectations. The marketing manager should tailor their advertisements on these attributes to ensure that the message towards the customers addresses these wants and values in purchasing a midsize sedan.

# **5.0 Aspirational Brand Analysis**

An aspirational brand is defined as “a brand of goods that people believe is of good quality and that will make them feel successful if they own it” (Aspirational Brand, n.d.). When translating this towards cars, an aspirational brand is often times associated with luxury brands (i.e. BMW, Mercedes, Lexus, Infiniti, etc.). This is because the idea of aspirational and luxury encompasses anything of a “higher value”, that is high quality build, high performance, very comfortable, and high price tag.

The brand that is the most aspirational according to our dataset should be one of the five most mentioned brands previously identified. This is because, with aspiration comes willingness to purchase. Customers that are willing to purchase will spend a fair amount of time researching a product on forums and asking for more information. This means that a brand that is mentioned a lot in a forum has a higher probability of being an aspirational brand. As our definition of aspirational brand includes luxury attributes, we have decided to look at the association between luxury and aesthetic with each of the five brands. By taking the average lift values of luxury and aesthetic with brand we get the following table.

|  |  |
| --- | --- |
|  | **Average Lift (Luxury + Aesthetic)** |
| **Toyota** | 1.220 |
| **Honda** | 1.196 |
| **Ford** | 1.195 |
| **Kia** | 1.154 |
| **Infiniti** | 1.094 |

Table 4: Lift Values for Top 5 Brands for “Aesthetic” and “Luxury” for each brand

The table implies that Toyota is seen as the most aspirational brand for users in this forum. The average lift values of luxury and aesthetic with issues for each brand have also been calculated to ensure the average lift value for Toyota is not associated the most with issues. In fact, an aspirational brand should not be one with many issues. As we can see from the table below, the average lift of luxury and aesthetic with issues for Toyota is average, implying that these attributes are perceived more positively for that brand than others.

|  |  |
| --- | --- |
|  | **Average Lift (Luxury + Aesthetic)** |
| **Ford - Issues** | 1.267 |
| **Infiniti - Issues** | 1.249 |
| **Toyota - Issues** | 1.249 |
| **Honda - Issues** | 1.212 |
| **Kia - Issues** | 1.1875 |

Table 5: Lift Values for Top 5 Brands Issues for “Aesthetic” and “Luxury” for each brand

Therefore, we identify Toyota as our most aspirational brand for the time frame 2013 - 2016. Being an aspirational brand, Toyota is setting the example of aesthetically pleasing luxurious midsize sedans on the market. This not only means that Toyota has a huge influence on that segment of the market but also, that it is the benchmark that every competitor aspires to overthrow. Toyota therefore needs to ensure that every car model they release has as few issues as possible. From our previous analysis, the most common issues that Toyota has are surprisingly associated to its luxury and aesthetic attributes. This would mean that even if there are issues with respect to those attributes, consumers still have a positive and strong opinion of the brand. This is a good indicator that Toyota’s brand image and equity remain strong and unhinged despite the issues that their car models might have, which further confirms that Toyota is an aspirational brand in consumers’ minds. Toyota should still resolve those issues and ensure the future long-term viability of the brand. Further text analysis on consumers’ posts should be done to identify what issues and for what car model should be addressed by Toyota.

# **Conclusion**

Analysis of the mention frequency of each brand in the forum asserted a strong dominance of Asian brands in the American market with up to 70% of mentions being associated with an Asian car brand. Through our analysis of brand co-mentions, we revealed a surprising association between Kia (an affordable brand) and Infiniti (a luxury brand). This result should be concerning to managers at Infiniti as a top-of mind association between the two brands maybe an indication of failed brand positioning or subpar product quality perception. We recommend further analysis into the nature of the association.

Our review of the most frequently mentioned attributes indicated that users are primarily concerned with the age, build, price, aesthetic, and luxury accommodations when discussing different car brands. Build and aesthetic were mentioned most frequently, indicating that these are the aspects users care most about and should be the focus of product and marketing strategy for any new models, especially when targeting a broader segment in the American market.

In our search for an aspirational brand, we identified Toyota as the best fit. This was largely based on the prevalence on the brand in user comments and its alignment with aspirational attributes such as luxury and aesthetic. Our analysis also revealed that Toyota has a few branding challenges when it comes to user perception that may threaten its long-term standing as an aspirational brand.

**Works Cited**

Aspirational Brand. (n.d.). In Cambridge Dictionary. https://dictionary.cambridge.org/dictionary/english/aspirational-brand

Best Sedans of 2022 and 2023 - Expert Reviews and Rankings. (n.d.). Retrieved from Edmunds website: <https://www.edmunds.com/sedan/>

Egan, S. (2021, May). Car & Automobile Manufacturing in the US: 33611A. IBISWorld. Retrieved Jan 30, 2022, from <https://my-ibisworld-com.proxy3.library.mcgill.ca/us/en/industry/33611a/about>

Guy, M. (2020, August 14). Wait, so who makes that? A breakdown of which automakers own which brands. Driving. Retrieved January 30, 2022, from <https://driving.ca/features/feature-story/wait-so-who-makes-that-a-breakdown-of-which-automakers-own-which-brands>

Hyundai Versus Kia - which is the best? (2020, December 17). Myers Barrhaven Hyundai. Retrieved January 30, 2022, from <https://www.myersbarrhavenhyundai.com/hyundai-versus-kia-which-is-the-best/#:%7E:text=In%20fact%2C%20Hyundai%20Motor%20Co,own%20engineering%20and%20design%20studios>.

MotorBiscuit.com. (2019, October 8). *10 Reasons Why We Buy Japanese Instead of American Cars*. MotorBiscuit. Retrieved January 30, 2022, from https://www.motorbiscuit.com/reasons-why-we-buy-japanese-instead-of-american-cars/

Severson, A. (2009, December 25). What’s a Matador? The AMC Matador, Rebel, and Classic > Ate Up With Motor. Ate up with Motor. <https://ateupwithmotor.com/model-histories/amc-matador/>

Stoll, J. D. (2015, June 24). Asian Car Makers Out-Earn U.S. Rivals. WSJ. Retrieved January 30, 2022, from https://www.wsj.com/articles/asian-car-makers-out-earn-u-s-rivals-1435078736