

DESAUTELS FACULTY OF MANAGEMENT

INSY 669 - Text Analytics



CAR FORUM TEXT ANALYSIS

Liu, Alice	261007356
Zhu, Diwei	260761307
Zhang, Ziye	260766101
Liang, Jiahua	260711529
Wang, Kexin	260787422
Wang, Yichen	260761601

February 2, 2022

1. Introduction

As noticed by many academics, user-generated content (UGC) is an abundant source of data that reflects more accurate customer attitudes. By applying text analytics tools to UGC data, key themes and user sentiments, as well as information about whether a certain brand is under the spotlight, can be detected and displayed.

In our case, we set our eyes to the Edmunds car forum. Based on the UGC collected from the forum, our objective is 1) to find out the most frequently mentioned car brands and car attributes, 2) to examine how the users associate the brands, and 3) to know in terms of which attributes they usually compare two brands. These findings that shed light on users' attitudes, when properly visualized and interpreted, will help form business insights that can be offered to product managers and marketing managers.

2. TASK A - Exploring Car Brands With Data Analytics Methods

2.1. Scraping the user review information

User review information is scrapped from the Edmund forum “Entry-level Luxury Performance Sedans”¹ with the help of python packages *request* and *BeautifulSoup*. We inspect the website and scrape user IDs, published dates and review contents, and encode and decode these pieces of information and store these reviews into a .csv file. This data serves as the basis of all our analyses. We focus our analysis on reviews between pages 200 and 310, which were written by users from September 2007 to April 2013 (Figure 1). We collected 5,500 comments from the source.

USER_ID	DATE	REVIEW
habitat1	Sep-07	The car doesn't appeal to me for a number of reasons (size, weight, bling
kdshapiro	Sep-07	"I think without the "buy American" loyalty,"I clearly have a buy American
fedlawman	Sep-07	"It is a matter of putting out a seriously competent competitor, however
designman	Sep-07	I expect to see this label soon:Outsourced in America
markcincinr	Sep-07	I do like the CTS, I will certainly consider one in early 2008, assuming they
jwaggoner	Sep-07	Has anyone sat in a 1 series yet? Im 6'1 and just wondering if i can fit in tl
pat	Sep-07	We are not here to talk about other members as you all well know.Is the
sevenfeet0	Sep-07	Agreed. This forum is supposed to be about things that most of us enjoy
cdnpinhead	Sep-07	". . .the 1976 Seville, the 1982 Cimarron, the 1989 Allante and the 1997 C

Figure 1. scrapped comments from the source

¹ <https://forums.edmunds.com/discussion/2864/general/x/entry-level-luxury-performance-sedans/l>

2.2. Word analysis

The package *nltk* helps break down sentences and turn them into separate words, thus tokenizing the reviews collected. In order to improve detection accuracy, an external car brand dataset from GitHub² is merged with the existing dataset so more models could be identified in the comment section. Our first step is, with the car model-brand dataset, to replace names of different models of cars with their brand names for the ease of counting in the later stages. Then, a similar searching process is done, but here all the brand names within a review are collected and added to a list, with all the duplicate brands eliminated. The number of mentionings of each brand is counted and the top 10 brands with highest mentioning frequency are shown in the figure below (Figure 2):

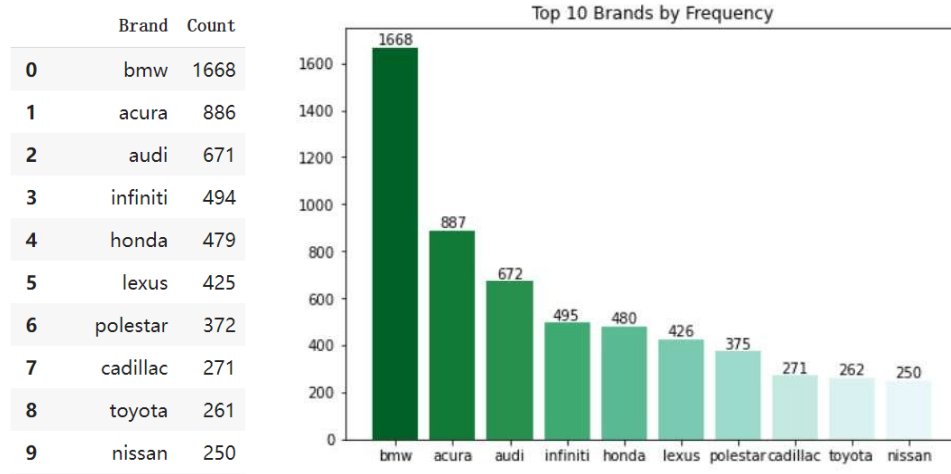


Figure 2. top ten car brands by frequency

2.3. Lift ratios of brand pairs

The lift of each pair of brands (A&B) among the 10 most frequently mentioned brands is calculated using the formula:

$$Lift(A, B) = N \times \frac{\#(A, B)}{\#(A) \times \#(B)}$$

where:

N = the total number of reviews

#(A, B) = the number of reviews mentioning both brand A and brand B

#(A) = the number of reviews mentioning brand A

#(B) = the number of reviews mentioning brand B

² <https://github.com/abhionlyone/us-car-models-data/blob/master/2022.csv>

The values needed for this formula can be obtained from the brand frequency count table and the table recording the brands mentioned in each review. The lift table is shown in Figure 3. The heatmap of the lift table is shown in Figure 4. In order to properly plot the MDS map, the dissimilarity between each pair of brands is calculated by $1/\text{Lift}$.

Brand	bmw	acura	audi	infiniti	honda	lexus	polestar	cadillac	toyota	nissan
Brand										
bmw	NaN	1.31621	1.781488	1.698950	1.209253	1.463180	1.714940	1.582048	1.208413	1.028964
acura	NaN	NaN	2.261074	2.355433	3.165504	2.067268	1.637276	1.556172	1.325576	1.414011
audi	NaN	NaN	NaN	2.348310	1.705419	2.478863	1.702690	2.899842	1.530972	1.211530
infiniti	NaN	NaN	NaN	NaN	1.551208	3.260895	1.748466	2.296440	1.781494	3.289487
honda	NaN	NaN	NaN	NaN	NaN	1.883167	2.139278	1.395549	3.543106	2.200400
lexus	NaN	NaN	NaN	NaN	NaN	NaN	1.274097	2.525449	4.682247	1.962798
polestar	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.515651	1.119796	1.173547
cadillac	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.936917	1.217934
toyota	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.115252
nissan	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure 3. Lift table for brand-brand association

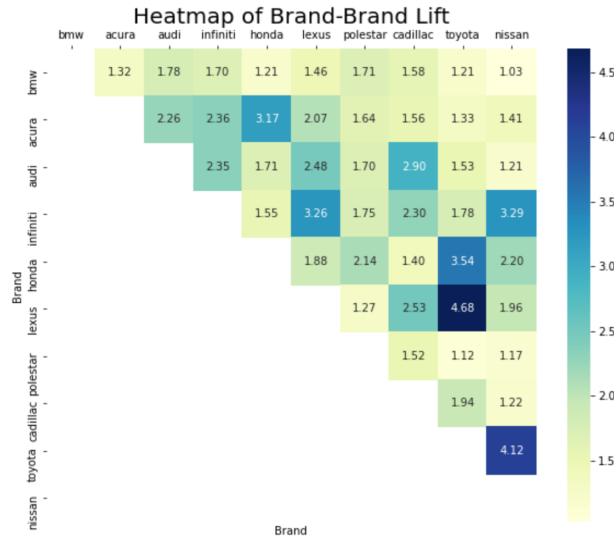


Figure 4. Heatmap for brand-brand Lift

2.4. Multidimensional scaling map

With the calculated lift ratios, we can plot the multidimensional scaling map (MDS map). We perform K Means clustering with the optimal number of clusters obtained from the Elbow

Method. As suggested by the Elbow Method, we split the top 10 car brands into 5 groups. Then, as shown in Figure 5, we plot a MDS map that displays the relative position of the top 10 brands.

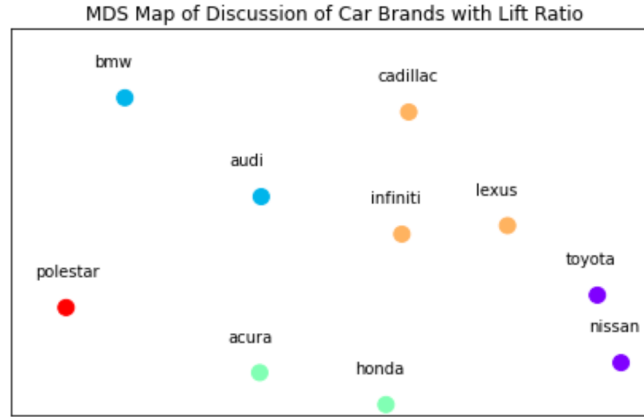


Figure 5. MDS Map of Discussion of Top 10 Car Brands

2.4. Supplemental sentiment analysis

In addition to the association between the brands, user sentiments associated with the brands are worth paying more attention to. If we do not measure user sentiments, it would be difficult to verify the reason why users link the two brands together – is it because both of them are preferred? or because the two brands are equally inferior? Therefore, in order to widen the basis for business insights generation, we implemented Sentiment Analysis with the *Afinn* package, which pre-defines a list of words with a sentiment value and calculates sentiment scores to each post according to the words it contains. A higher positive score of a review means a higher positive sentiment, and vice versa. Along with the generated sentiment scores for scraped forum reviews, we also dummified the mentioning of the top 10 brands of each post (1 denotes ‘mentioned’, 0 otherwise) and generated a dataset (Figure 6) for later analysis.

Sentiment score	Brand_Mentioned	Attribute_Mentioned	bmw	acura	audi	infiniti	honda	lexus	polestar	cadillac	toyota	nissan	performance	purchase	type	car parts	comfort
10	['cadillac', 'toyota']	['performance']	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0
18	['cadillac', 'audi', 'bmw']	['purchase', 'performance', 'comfort', 'car parts']	1	0	1	0	0	0	0	1	0	0	1	1	0	1	1
19	['cadillac', 'bmw', 'audi']	['performance', 'purchase']	1	0	1	0	0	0	0	1	0	0	1	1	0	0	0
3	['bmw', 'dodge', 'cadillac', 'ford']	['type', 'purchase']	1	0	0	0	0	0	1	0	0	0	1	1	1	0	0
-6	['volvo']	['performance']	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
2	['cadillac', 'lexus', 'bmw']	['performance']	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0
4	['bmw']	['type']	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
3	['bmw']	['type', 'car parts']	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
6	['cadillac']	['performance']	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
14	['cadillac', 'bmw', 'honda']	['performance', 'car parts', 'comfort']	1	0	0	0	1	0	0	1	0	0	1	0	0	1	1
9	['cadillac']	['comfort']	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
4	['cadillac', 'bmw', 'polestar', 'audi', 'infiniti']	['purchase', 'performance']	1	0	1	1	0	0	1	1	0	0	1	1	0	0	0
4	['bmw', 'cadillac']	['comfort']	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
16	['audi', 'lexus', 'cadillac']	['comfort']	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1
-3	['cadillac', 'bmw', 'infiniti']	['purchase', 'performance']	1	0	0	1	0	0	0	1	0	0	1	1	0	0	0
6	['bmw', 'audi', 'lexus', 'cadillac', 'infiniti', 'acura']	['comfort']	1	1	1	1	0	1	0	1	0	0	0	0	0	0	1
19	['audi', 'bmw', 'cadillac']	['performance', 'comfort', 'car parts']	1	0	1	0	0	0	0	1	0	0	1	0	0	1	1
11	['bmw', 'cadillac', 'acura']	['purchase', 'performance', 'car parts']	1	1	0	0	0	0	0	1	0	0	1	1	0	1	0
-29	['cadillac']	['car parts', 'performance', 'comfort']	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1
25	['cadillac', 'bmw']	['performance', 'comfort']	1	0	0	0	0	0	0	1	0	0	1	0	0	0	1
-28	['bmw']	['performance', 'comfort', 'car parts']	1	0	0	0	0	0	0	0	0	0	1	0	0	1	1
5	['cadillac']	['performance']	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0

Figure 6. generated dataset of sentiment scores and brand/attribute mentioning indicators

3. TASK B - Business Insights for Car Brands

3.1 MDS analysis

According to the heatmap of brand-brand lift values (Figure 4), we could observe the most associated brands (i.e., highest lift value)—brands that are always discussed together in the forum—are Lexus against Toyota. Furthermore, it appears that discussions around Toyota are heated, as many brands (such as Lexus, Nissan, Honda) have a high lift value with Toyota. Conversely, there is not much discussion between BMW and the other 9 brands in this forum as all lift values involving BMW are smaller than 2. The low values imply that BMW is less frequently compared with other brands than the comparison among those non-BMW brands. In fact, our later analysis in E reveals that the reason for the low lift score between BMW and others is attributed to the fact that BMW is the aspirational and benchmark brand to other entry level luxury cars. The heatmap also shows the least associated brands around which people compare and discuss are Nissan and BMW with a lift value of 1.03.

Based on the MDS diagram, we group the 10 brands into 5 clusters. As two of the established and well-known brands, Audi and BMW share similar characteristics and are often discussed against each other. Cadillac, Infiniti, and Lexus are often discussed together according to the MDS. Toyota and Nissan are more alike than the other brands, suggesting that they may target similar segments of customers. Acura and Honda are close to each other in the MDS map, because they are from the same automaker, with the former being more luxurious than the latter. Finally, Polestar seems to stand alone away from the other 9 brands on the MDS map.

3.2 Sentiment analysis

Figure 7 shows each brand's overall sentiment score sorted from the highest to the lowest. Interestingly, BMW is the least positive brand among the 10 popular brands in the forum from 2007 to 2013, whereas Audi is the brand which people are most positive about, followed by Cadillac, Lexus and Toyota. Overall, these top 10 brands all have a decent positive sentiment score, which means these brands are generally liked by people when being discussed.

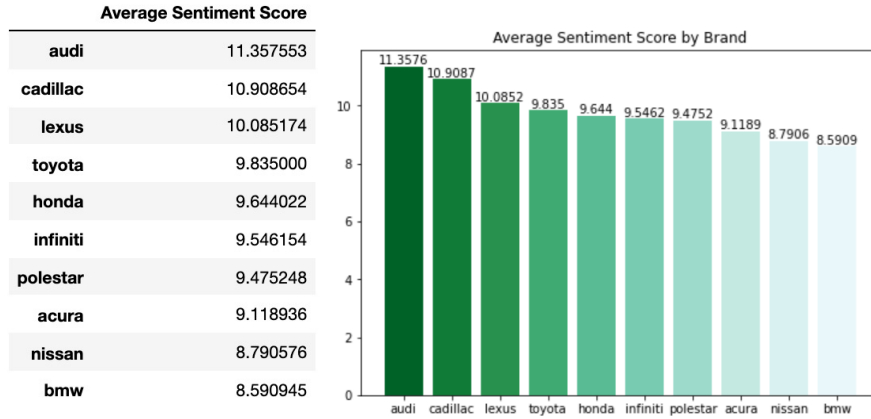


Figure 7. top ten car brands by sentiment score

We further implement a linear regression on sentiment scores using our top 10 brands along with the five categories of attributes. All these features have a significant relationship with sentiment scores since their p values are all below 5%. Intuitively, as these are the discussions on the car forum, we should have expected that people express their likes and dislikes concerning cars and their attributes. Note that the coefficients here cannot be interpreted as a causal effect on user sentiment, but rather reveal a correlation between each brand and sentiment when controlling for other brands and attributes.

A contradicted finding from the regression analysis is that while the average sentiment score of BMW is the lowest in the first chart, the coefficient of BMW has the second largest (only lower than Audi) relationship with positive sentiment among the 10 brands, i.e., on average, the sentiment score of the review increases by 2.42 when BMW is being discussed in the review. Among all, Acura is the brand that people express less degree of positiveness.

3.3 Two brands that we can offer the most interesting/useful insights

The two brands talked about here are Infiniti, a brand of Nissan, and Lexus, a brand of Toyota. Infiniti and Lexus are both the luxury division of its company building upon their middle-class best-selling brands Nissan and Toyota respectively. From the MDS visualization, Lexus and Infiniti belong to the same cluster and they are very close to each other, indicating a high association between the two. Furthermore, they are close to Toyota and Nissan in one direction whereas they are also close to BMW and Audi in the opposite direction. Being the luxury divisions of the companies that are primarily making middle-class cars, they act as the

transition from middle-end to high-end brands such as Audi and BMW, and they are more comparable to them in terms of price and performance.

From the regression analysis (Figure 8), reviewers generally have roughly the same sentiment towards Infiniti (a coefficient of 1.16) and Lexus (a coefficient of 1.37). Apparently, Lexus beats Infiniti by a small margin based on reviewers' attitudes and public opinion. However, Infiniti is more alike with well-known luxuries like Audi and BMW. For example, during 2007 – 2013, Infiniti offered the Q50S which was more similar to BMW 3 Series in terms of price, performance, and quality, compared to IS 350 which was a similar product by Lexus. While BMW created models with lower prices and downgraded quality to match broader targeted buyers, Infiniti and Lexus provided an upscale experience to people with higher quality and reliability cars.

By looking into these two brands back in 2007-2013, Lexus has a dominant power against Infiniti. The 2009 August Sales statistic³ shows that Lexus sold a total of 22,892 cars whereas Infiniti only sold 9,625 cars. Despite Lexus being less ambitious about approaching BMW, it created more distinct, higher quality, higher resale-value models compared to Infiniti, which attracted better reviews and a more positive attitude towards itself.

OLS Regression Results						
=====						
Dep. Variable:	Sentiment score	R-squared:	0.197			
Model:	OLS	Adj. R-squared:	0.192			
Method:	Least Squares	F-statistic:	36.87			
Date:	Sun, 30 Jan 2022	Prob (F-statistic):	4.12e-96			
Time:	19:15:23	Log-Likelihood:	-8160.1			
No. Observations:	2267	AIC:	1.635e+04			
Df Residuals:	2251	BIC:	1.644e+04			
Df Model:	15					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-1.8714	0.479	-3.908	0.000	-2.810	-0.932
bmw	2.4197	0.388	6.239	0.000	1.659	3.180
acura	0.9897	0.441	2.247	0.025	0.126	1.854
audi	3.3816	0.457	7.396	0.000	2.485	4.278
infiniti	1.1617	0.510	2.280	0.023	0.162	2.161
honda	1.7405	0.532	3.274	0.001	0.698	2.783
lexus	1.3719	0.562	2.440	0.015	0.269	2.475
polestar	1.2470	0.558	2.235	0.025	0.153	2.341
cadillac	2.4008	0.655	3.665	0.000	1.116	3.685
toyota	1.5725	0.693	2.270	0.023	0.214	2.931
nissan	1.5705	0.692	2.269	0.023	0.213	2.928
performance	2.8587	0.387	7.387	0.000	2.100	3.618
purchase	3.6176	0.390	9.283	0.000	2.853	4.382
type	2.9074	0.397	7.325	0.000	2.129	3.686
car parts	0.9883	0.453	2.182	0.029	0.100	1.877
comfort	4.3417	0.494	8.794	0.000	3.374	5.310
=====						
Omnibus:	710.499	Durbin-Watson:	1.749			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6241.449			
Skew:	1.222	Prob(JB):	0.00			
Kurtosis:	10.752	Cond. No.	6.43			

Figure 8. Sentiment score regression model result

³ <https://www.goodcarbadcar.net/lexus-acura-infiniti/>

4. TASK C - Exploring Car Attributes With Data Analytics Methods

4.1. Generating car attribute list as reference

To label the mentions of attributes in reviews, we generated a list of attribute keywords based on the paper by Netzer et al. (2012)⁴ and on external resources, and then by experiential judgements assigned the keywords to 7 attribute categories—Performance, Comfort, Interior, Safety, Purchase, Type, and Car Parts. The list contains 122 different keywords. We listed the categories and keyword examples in Figure 9.

Attribute Category	Keywords examples
Performance	speed, acceleration, engine, power, RPM
Comfort	noise, heater, A/C, roomy
Interior	leather, panel, design
Safety	insurance, ABS, brake, airbag
Purchase	price, tax, discount, MSRP, used
Type	sport, compact, SUV, hybrid, commuter
Car Parts	sensor, door, wheel, battery

Figure 9. defined attribute categories and keywords examples

4.2. Word Analysis

With the attribute categories and keywords list, using the similar steps as the calculation of mentions frequency of car brands, we generated the numbers of mentions of each attribute category across the 5,500 reviews. As shown in Figure 10, the top 5 frequently mentioned attributes are sorted and listed.

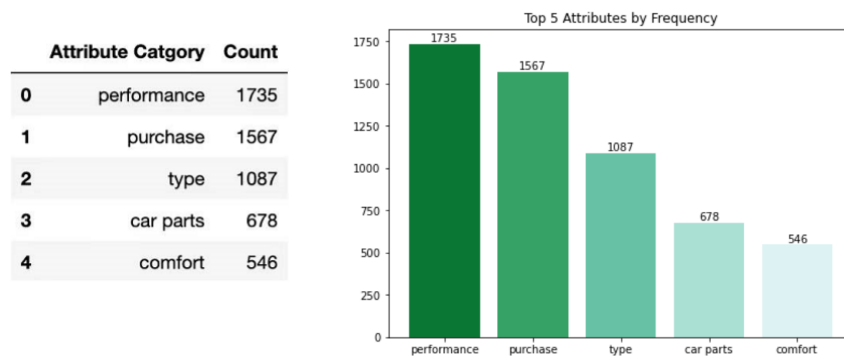


Figure 10. top 5 attributes by mentioning frequency

⁴ Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3), 521–543. <https://doi.org/10.1287/mksc.1120.0713>

4.3. Calculating brand-attribute lift ratio

To find out which attributes are most strongly associated with which of these 5 most frequently mentioned brands, the lift value of each pair of brands and attributes is calculated following the same calculation process as for the brand-brand association except that the A and B in the formula become each of the top 5 brands and top 5 attributes (generalized to “Word”). The lift table of brand-attribute association is shown in Figure 11. Note that the lift of each pair of the top 5 brands and each pair of the top 5 attributes are also included in the table as a byproduct of the way we calculate lift which can be ignored. A heatmap for the brand-attribute lift table is shown in Figure 12. From the table and heatmap, we can identify the attributes that are most strongly associated with each of these 5 brands (Figure 13).

Word	bmw	acura	audi	infiniti	honda	performance	purchase	type	car parts	comfort
bmw	NaN	1.130099	1.491664	1.438849	1.051034	1.385462	1.506644	1.410555	1.522234	1.413155
acura	NaN	NaN	1.873121	1.979206	2.480271	1.386664	1.357264	1.705614	1.618761	1.294647
audi	NaN	NaN	NaN	1.984127	1.551649	1.500103	1.598254	1.679070	1.617590	1.948696
infiniti	NaN	NaN	NaN	NaN	1.412037	1.479347	1.659221	1.655934	1.704359	1.261701
honda	NaN	NaN	NaN	NaN	NaN	1.459534	1.477079	1.676058	1.723820	1.594933
performance	NaN	NaN	NaN	NaN	NaN	NaN	1.171312	1.376498	1.458774	1.457284
purchase	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.204406	1.257970	1.304960
type	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.888096	1.390054
car parts	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.005738
comfort	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure 11. Lift table for brand-attribute association

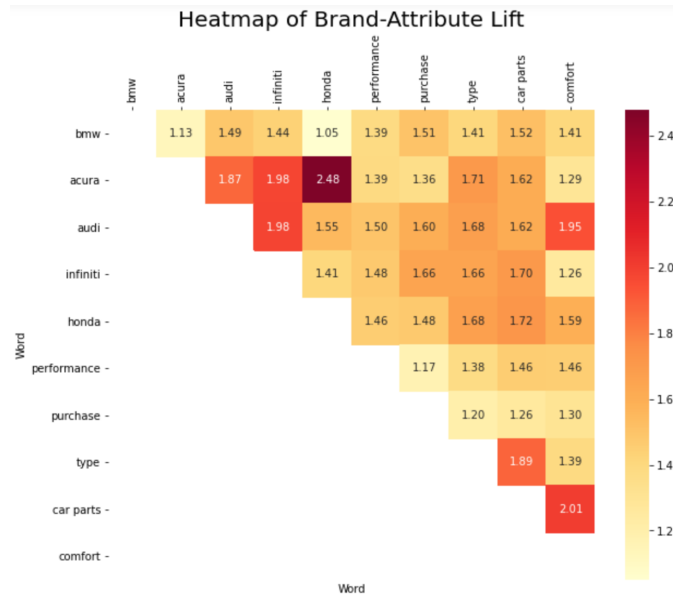


Figure 12. Heatmap for brand-attribute Lift

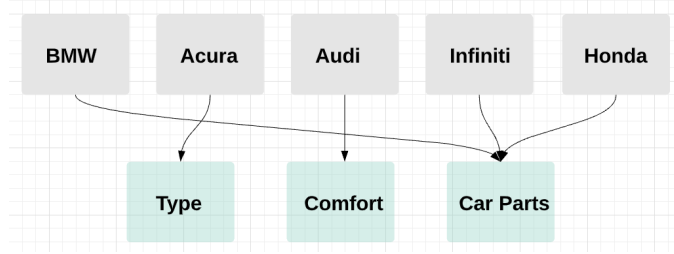


Figure 13. Most strongly associated attribute of each top 5 brands

4.4. MDS mapping

We also adopt similar methods for analyzing the top 5 attributes in the top 5 brands. The Elbow Method suggests us split the brands and attributes into 3 groups. However, while we are running trials on the number of clusters, we find that the result is intuitively more explainable when the number of clusters equals 2. We then proceed to perform K Means clustering, as shown in Figure 14.

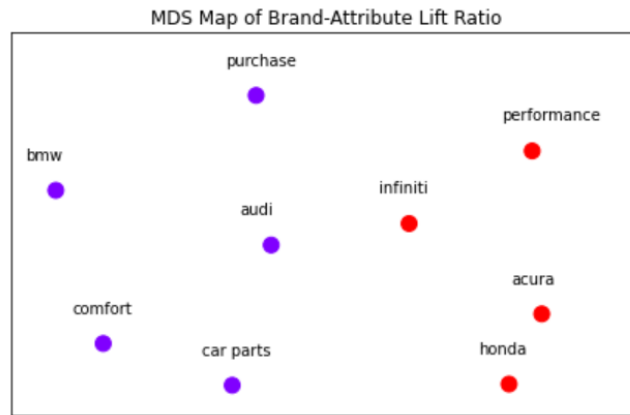


Figure 14. MDS Map of Top 5 Car Brands and Top 5 Most Mentioned Attributes

4.5. Supplemental sentiment analysis

Sentiment analysis is applied to car attributes as well, similar to the process in Question 2(e), where we dummified the top 5 attributes (1 means mentioned, 0 otherwise) at the same time. With the help of the Afinn package, users' sentiments towards each attribute could be identified. By combining this information with lift scores calculated, more business insights could be obtained about the relationship between each brand and attributes in terms of positive and negative attitudes.

5. TASK D - Car Attributes-related Business Insights

5.1. Production Insights

From Figure 15, we can see that, among the top 5 brands, the order of the 5 attributes stays the same: from performance, purchase, type, car parts, to comfort. It shows that the feature importance of each attribute in customers' consideration is unchangeable no matter which brands. Therefore, we provide a general suggestion to the product manager: make improvements on the current models and future R&D based on the order of the top 5 attributes. Performance is the priority, but there's no right or wrong way to measure it. Customers have their own criteria, so targeting customers, making segmentation towards different needs (speed, tight seat, handling and brake) is crucial before designing the final draft. The multiple combinations of customers' preferences can influence their choices towards various model types such as SUV, Sedan, Coupe and so on. Then, the product manager could consider choosing car parts based on what type of qualification so as to control production cost.

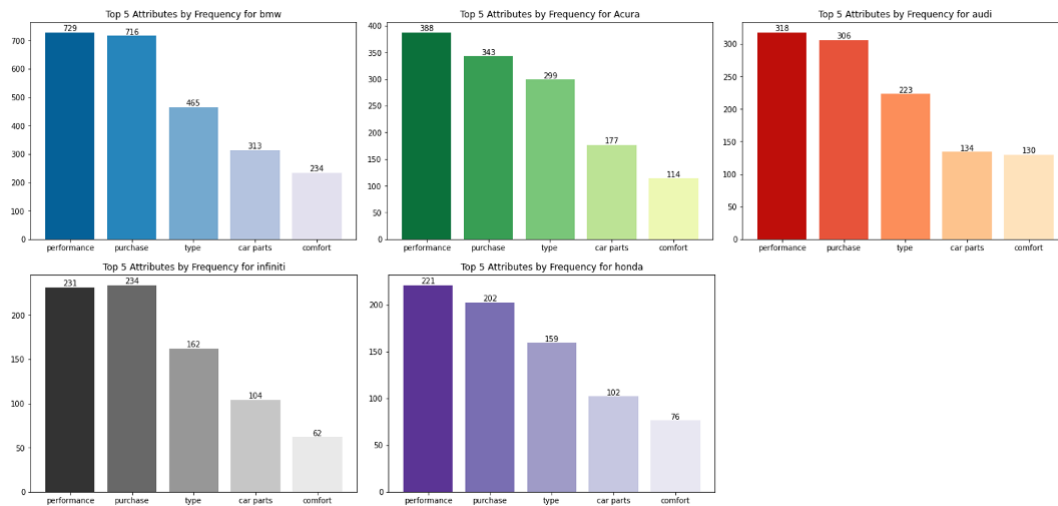


Figure 15. Top 5 Attributes by Frequency for Top 5 brands

5.2. Marketing/Advertising Insights

From a marketing/advertising perspective, we see that customers, based on their reviews on the forum, show keen interests in a car's performance and purchase. Among the top 5 discussed attributes, the frequency of mentioning performance and purchase greatly exceed others. The conveying message we obtained from this result is that when looking for a car, customers want to first see how fast it can run, how quickly it can accelerate and how much fuel

it costs. Then, they will take a look at the car's price and make considerations, potentially related to "whether I can purchase this car or not" or "whether this car's performance matches its price", etc. However, a car's performance and price possess an inverse relationship. Obviously, high-performance cars are not economically-friendly. Therefore, when marketing a car model, managers should mainly consider highlighting its advantages regarding either performance (such as speed, power, acceleration) or purchase (such as price, luxuriousness, mileage). By doing this, the advertisement can quickly target the right customers and capture their interests.

The next three most frequently mentioned attributes are type, car parts and comfort. We don't recommend managers to take a significant amount of consideration for these attributes. However, their importance should still not be overlooked in constructing the optimal marketing strategy. This is because our study is based on data extracted from a vehicle forum, consisting of both direct customers who need purchase recommendations and autofans who owns product knowledge but little interest in purchasing. It is possible that autofans tend to focus more on the performance of a car and omit other factors such as driver's experience, interior design, etc.. Therefore, our suggestions for the marketing/advertising team is to address these factors to the extent so that enough information is conveyed but the main selling points regarding performance and purchase are not compromised.

6. TASK E - Identifying and Analyzing The Aspirational Brand

Albeit having no consistent definition, the aspirational brands can be conceptualized as those currently unaffordable "dream brands" for which individuals possess desires to purchase upon reaching a higher income and/or social class (Trocchia et al., 2015)⁵. To identify the most aspirational brand, we realize the need to compare the purchasing intentions of users by examining the *Purchase*-brand lift ratios. Additionally, with the help of a further interpretation of brand-brand lift ratios and sentiment analysis results, we can locate the most aspirational brand based on multiple attests.

6.1. Aspirational Brand Identification

It is interesting to observe from the brand mentioning frequency table and the brand-brand lift table that BMW has the most mentions in the reviews (1668 times) but the

⁵ Trocchia, P. J., Saine, R. Q., & Luckett, M. G. (2015). I've Wanted A BMW Since I Was A Kid: An Exploratory Analysis Of The Aspirational Brand. *Journal of Applied Business Research (JABR)*, 31(1), 331. <https://doi.org/10.19030/jabr.v31i1.9011>

brand-brand lift ratios between BMW and other brands are relatively low. The lift formula implies that the occurrence of BMW itself in the reviews is high while the co-occurrence of BMW and other brands is low, which means that people always mentioned BMW but less frequently compared it to other brands. On the other hand, the brand-brand lift ratios among the Japanese luxury brands such as Lexus and Infiniti, as well as their parent brands Toyota and Nissan are relatively higher, which suggests that people compare Japanese luxury brands and their parent brands more frequently. A possible reason behind it could be that BMW was always at the top of the global luxury sedan list and people always liked to talk about BMW with luxury car discussion. However, due to the generally higher price of German luxury brands like BMW, people with less purchasing capabilities but still are interested in luxury cars, especially the users that are posting comments in this Entry-level Luxury Sedan forum page, prefer cheaper alternatives that are at a similar level. In this sense, Japanese luxury brands could be a good consideration. In addition, from the brand-attribute lift table, we can observe that the two German luxury brands, BMW and Audi, and the Japanese luxury brand Infiniti are strongly associated with the Purchase attribute (lift ratio 1.51, 1.60, 1.66 respectively). This suggests that people frequently mention those brands when they discuss prices. According to the regression model based on sentiment analysis, BMW demonstrates the second high positive coefficient to sentiment scores (2.4197), indicating an overall positive user sentiment related to the brand. Integrating the reasonings, we therefore conclude that BMW can be identified as the most aspirational brand among the brands discussed on the selected forum pages.

6.2. Brand Business Application

Our analysis of BMW's brand image finds that BMW puts a high priority on performance. A genuine BMW is always driver-focused, dedicated only to the driver. Therefore, we believe the way BMW seeks the best overall solution for their customers, from the engineering to the marketing side, makes it become an aspirational brand. BMW recommends their engineers create every detail to better serve drivers, and it also with perseverance keeps investing in creating flagship stores to maximize customers' experience, to let customers try test drive and be impressed by the comfort. Therefore, for those customers who have yet to aspire to purchase a particular brand or BMW, the flagship strategy will attract potential customers to try.

Even though BMW is the world's No.1 premium car company that owns Rolls Royce, MINI cooper, and comprehensive model series which attracts its exposure audience wishes to own it by promotion, but for economic reasons cannot. It's not easy to keep high performance but low purchase level, and keep customer satisfaction. However, we find that BMW tries to provide entry-luxury models like 3-series and e-series that maintain their brand strategy with good performance and driver experience but are still relatively cheap compared with the other models. By proper promotions from BMW owner, lover, marketing teams and mass media, BMW could easily express their brand and attract more potential customers from their technological advantage, design and how it is meant to be for the driver-experience premium people. We believe the best business application towards BMW is maintaining their current strategy, keeping its aspirational brand through marketers and customers, and turn exposure customer who with certain pursue into the real BMW owner.