

Text Analytics Report

Contents

Data Extraction/Pre-processing Summary.....	3
Most Frequent Brands - Lift Values and MDS Visualization	3
Attributes Analysis	7
Aspiration Analysis	10
Bibliography	13

Data Extraction/Pre-processing Summary

Data was extracted from the link referenced below using the BeautifulSoup Package in Python. 5000 reviews were extracted. Following this, the comments were tokenized and put into lower case. The csv containing model/abbreviation and brand pairings was also put into lower case. The models.csv file was modified from its original state, by adding additional models to the existing brands, adding new brands that had not been mentioned (such as jaguar, jeep and bentley), as well as common car brand abbreviations/slang (bimmer, jag, L dog). A script was then run to replace any mention of a model in the comments with the associated brand from the models.csv file. Preprocessing was then continued on these 'brand only' comments by removing punctuation and symbols, replacing numbers with integers, removing stop words and lemmatizing. This left us comments that were in the 'bag of words' format, making the dataset leaner and easier to analyze.

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

Most Frequent Brands - Lift Values and MDS Visualization

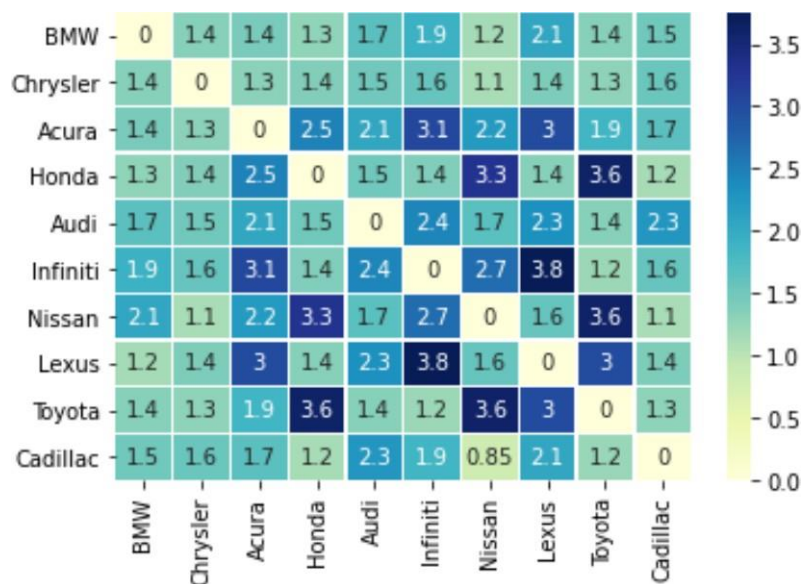
The number of comments in which a brand was mentioned is presented below.

Brands	count
bmw	1434
chrysler	1080
acura	540
honda	463
audi	405
infiniti	378
nissan	317
lexus	317
toyota	292
cadillac	261

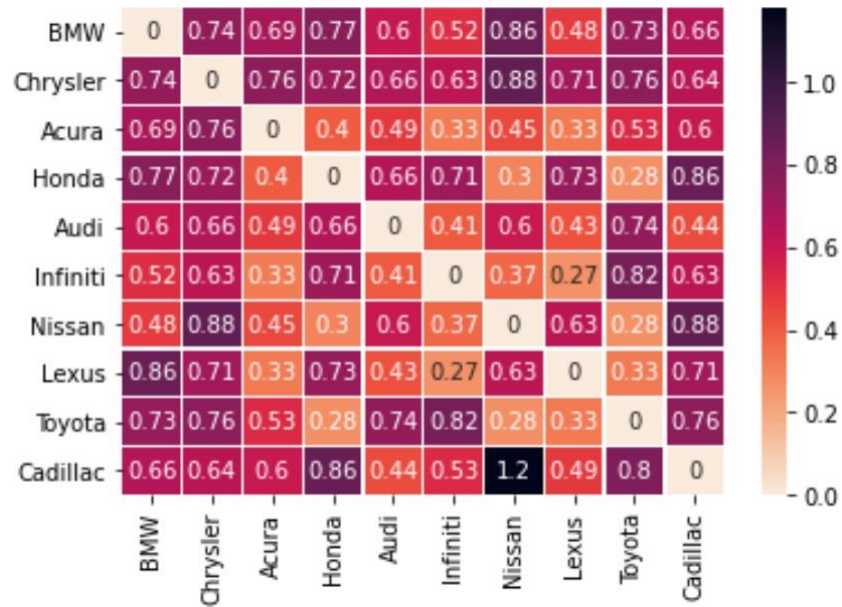
The number of times each of these popular brands were mentioned with the other popular brands was then counted and collected. The list for the first eight is presented below in tabular format.

Brand1	Brand2	comentions	Brand1	Brand2	comentions	Brand1	Brand2	comentions	Brand1	Brand2	comentions
bmw	bmw	1434	chrysler	bmw	419	acura	bmw	224	honda	bmw	172
bmw	chrysler	419	chrysler	chrysler	1080	acura	chrysler	154	honda	chrysler	138
bmw	acura	224	chrysler	acura	154	acura	acura	540	honda	acura	124
bmw	honda	172	chrysler	honda	138	acura	honda	124	honda	honda	463
bmw	audi	193	chrysler	audi	132	acura	audi	90	honda	audi	57
bmw	infiniti	209	chrysler	infiniti	130	acura	infiniti	125	honda	infiniti	49
bmw	nissan	106	chrysler	nissan	78	acura	nissan	76	honda	nissan	97
bmw	lexus	189	chrysler	lexus	97	acura	lexus	103	honda	lexus	40
bmw	toyota	114	chrysler	toyota	83	acura	toyota	59	honda	toyota	98
bmw	cadillac	114	chrysler	cadillac	88	acura	cadillac	47	honda	cadillac	28
Brand1	Brand2	comentions	Brand1	Brand2	comentions	Brand1	Brand2	comentions	Brand1	Brand2	comentions
audi	bmw	193	infiniti	bmw	209	nissan	bmw	106	lexus	bmw	189
audi	chrysler	132	infiniti	chrysler	130	nissan	chrysler	78	lexus	chrysler	97
audi	acura	90	infiniti	acura	125	nissan	acura	76	lexus	acura	103
audi	honda	57	infiniti	honda	49	nissan	honda	97	lexus	honda	40
audi	audi	405	infiniti	audi	75	nissan	audi	43	lexus	audi	60
audi	infiniti	75	infiniti	infiniti	378	nissan	infiniti	64	lexus	infiniti	90
audi	nissan	43	infiniti	nissan	64	nissan	nissan	317	lexus	nissan	32
audi	lexus	60	infiniti	lexus	90	nissan	lexus	32	lexus	lexus	317
audi	toyota	32	infiniti	toyota	27	nissan	toyota	67	lexus	toyota	56
audi	cadillac	48	infiniti	cadillac	37	nissan	cadillac	14	lexus	cadillac	34

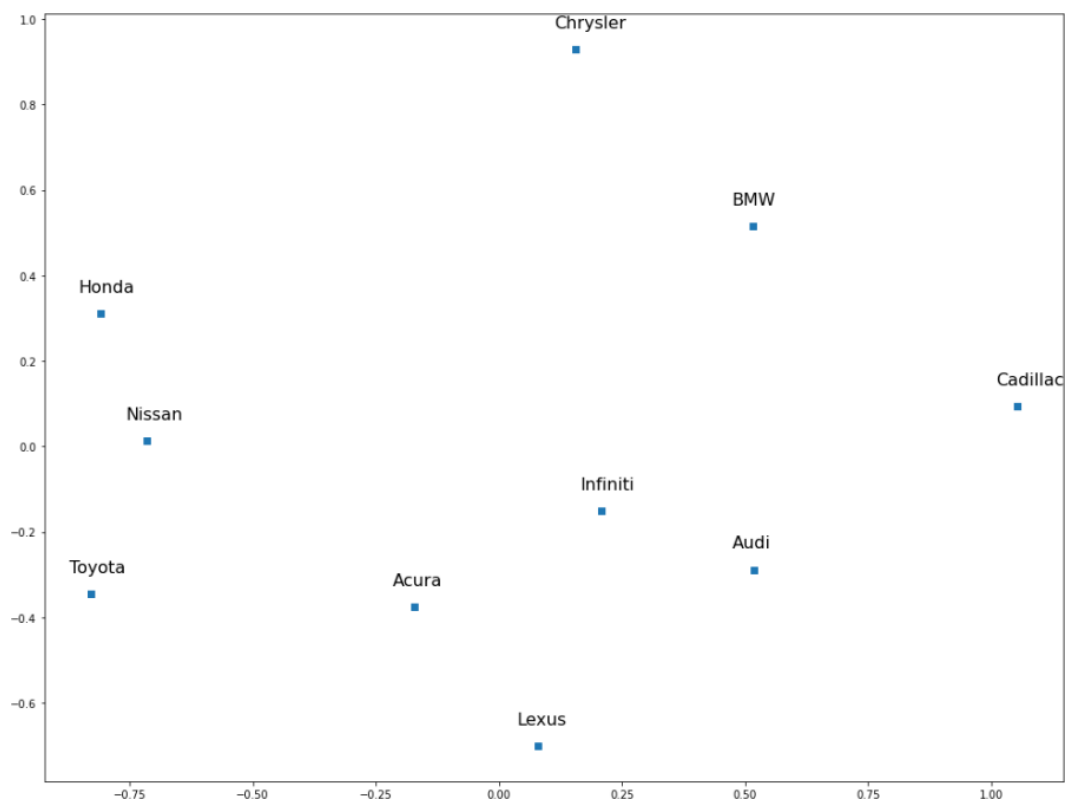
The corresponding lift value matrix was then derived and is visualized as a matrix below.



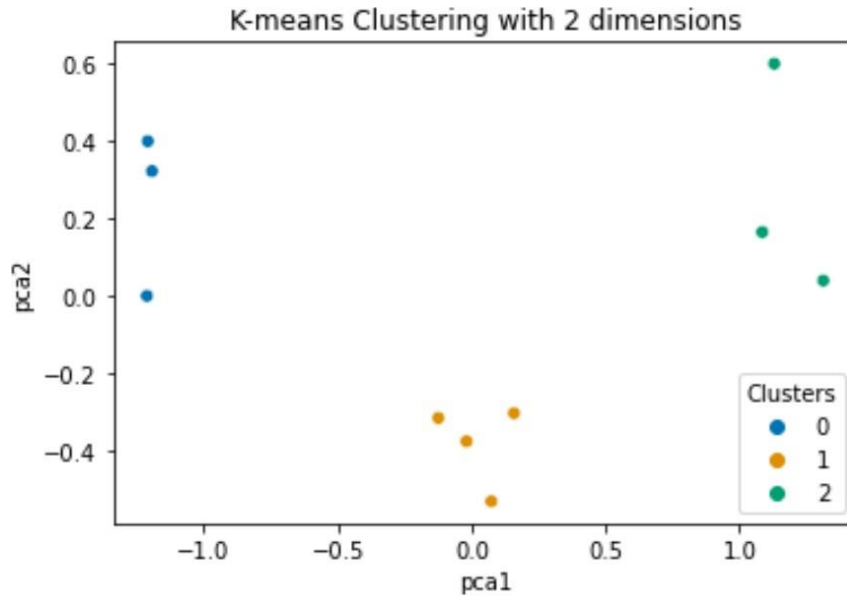
The corresponding Inverse lift value matrix was then derived and is visualized as a matrix below.



The MDS plot was then derived and is presented below.



Lastly, K-Means clustering was performed and presented in a 2D space using PCA technique for top 10 brands.



The below table displays the brands under each cluster and their inverse lift ratios.

INVERSE LIFT RATIOS											
Brands	BMW	Chrysler	Acura	Honda	Audi	Infiniti	Nissan	Lexus	Toyota	Cadillac	Clusters
Honda	0.77	0.72	0.40	-	0.66	0.71	0.30	0.73	0.28	0.86	0
Nissan	0.48	0.88	0.45	0.30	0.60	0.37	-	0.63	0.28	0.88	0
Toyota	0.73	0.76	0.53	0.28	0.74	0.82	0.28	0.33	-	0.76	0
Audi	0.60	0.66	0.49	0.66	-	0.41	0.60	0.43	0.74	0.44	1
Infiniti	0.52	0.63	0.33	0.71	0.41	-	0.37	0.27	0.82	0.63	1
Lexus	0.86	0.71	0.33	0.73	0.43	0.27	0.63	-	0.33	0.71	1
Acura	0.69	0.76	-	0.40	0.49	0.33	0.45	0.33	0.53	0.60	1
BMW	-	0.74	0.69	0.77	0.60	0.52	0.86	0.48	0.73	0.66	2
Chrysler	0.74	-	0.76	0.72	0.66	0.63	0.88	0.71	0.76	0.64	2
Cadillac	0.66	0.64	0.60	0.86	0.44	0.53	1.18	0.49	0.80	-	2

The MDS and K-means clustering plot reveals some interesting groupings: 1)Japanese

“Regular” Brands: Honda, Nissan, Toyota (cluster 0)

2) “Luxury” Brands: Acura, Infiniti, Audi, Lexus (cluster 1)

3) “Special Snowflakes” : Cadillac, Chrysler, BMW (cluster 2)

The first two groups are rather intuitive, especially when considering that Acura and Infiniti are the luxury divisions of Honda and Nissan. Whereas group 1 accounts for the more everyday “normal use” brand, group 2 consists of brands that are immediately associated with luxury (Infiniti, Lexus, Audi) and performance (Acura). However, the “Special Snowflakes” group (3) yields the most interesting insights for brand managers, so-named due to their unique positionings in the MDS plot.

Cadillac's position is distinctly separate from anything else, similar to its positioning in the *Mine Your Own Business* article discussed in class. It is closer to the Luxury group (2) than to the only other American brand (Chrysler). Cadillac went through a rebranding phase in the mid-2000s from "classic American" to "young luxury" (per the *Mine Your Own Business* authors), and the forum posts for this project are roughly from that era (2002-2006). A brand manager would be particularly interested in Cadillac's proximity to the other luxury brands over time to see if the rebranding was having any effect. Adding in some more everyday American brands would yield further insights, namely if Cadillac is markedly distinct from them in the MDS plot.

BMW is also worthy of note, and fans of "Bimmers" (forum spelling. . .) would be particularly excited by the distinct positioning confirming how special they are. A BMW brand manager would be reassured by the brand's separation from the Luxury group (2), as BMW's tend to be a mix of luxury and performance. Infiniti, Audi, and Lexus are generally more luxury-oriented, while Acura is more performance-oriented. The inclusion of Porsche and Jaguar, representing a higher-cost luxury/performance category in the MDS plot could provide an indicator to see if BMW falls somewhere in between those and group 2. This would confirm BMW's "hybrid" nature (equal mix of luxury/performance) and possibly suggest if a rebalancing is necessary from the brand manager's point of view.

Attributes Analysis

Attributes were first determined by singular occurrences per comment. These were then grouped into more general categories. A find-and-replace script was used to replace all occurrences of similar words with the general category per the below chart. A separate script was run to find specific words in sequence (front wheel drive, drive train), and these were replaced with the single word representing the category. The counts for categories, and similar words are listed below:

Category	Count	Category	Count	Category	Count	Category
Performance	1490	Price	1503	Drivetrain	1043	Sedan
Similar words		Similar words		Similar words		Similar words
Performance	636	Price	503	Fwd	319	N/A
Perform	102	Money	221	Awd	293	
Hp	316	Cheap	83	Rwd	378	
Horsepower	73	Expensive	123	Front wheel drive	12	
Acceleration	127	Value	221	Rear wheel drive	3	
Torque	236	Pricey	10	Drivetrain	38	
		Cost	227			
		Overprice	37			
		Cheaper	56			
		Costlier	2			
		Costly	5			
		Cheapest	3			
		inexpensive	12			

Once the find and replace script was run, the mentions per comment for each attribute were as follows:

Attribute	Count
performance	1159
price	1112
sedan	659
drivetrain	658
problem	499

The co-mentions were then calculated, which entailed finding comments where both the brand and the attribute were mentioned, resulting in the below numbers:

Brand	Attribute	Comentions
bmw	performance	466
bmw	price	479
bmw	sedan	257
bmw	drivetrain	205
bmw	problem	171
chrysler	performance	329
chrysler	price	334
chrysler	sedan	192
chrysler	drivetrain	184
chrysler	problem	176
acura	performance	183
acura	price	187
acura	sedan	134
acura	drivetrain	152
acura	problem	71
honda	performance	150
honda	price	162
honda	sedan	113
honda	drivetrain	93
honda	problem	77
audi	performance	126
audi	price	142
audi	sedan	86
audi	drivetrain	83
audi	problem	69

With the attribute and brand co-mentions count in hand, the lift ratios were derived. These can be observed in the chart below. The heat map coloring is done according to each attribute to communicate the intensity of respective life values:

LIFT	performance	price	sedan	drivetrain	problem
bmw	1.4019	1.5019	1.3598	1.3618	1.1949
chrysler	1.3142	1.3906	1.3488	1.2946	1.6329
acura	1.4620	1.5571	1.8828	2.1389	1.3174
honda	1.3976	1.5733	1.8517	1.5263	1.6664
audi	1.3422	1.5765	1.6111	1.5573	1.7071

The product manager is responsible for the success of a product and for improving it. As such, they would be interested in the general attributes; “performance” , “price” , and “problem”

. When analyzing this from the lens of an Audi product manager, the focus would be on the mention of “problem” and “price” terms, as these have the highest lift value among brands. “Problem” may hint to flaws in the production process, and further investigation into the specifics could bring to light the need for improvements or even recalls. The high lift value of “price” may be a result of the problems but could also indicate mispricing. If customers are mentioning high prices too often, improvements in production efficiency from “problems” could result in lowered costs and in turn lower prices. However, if a higher price is part of the product strategy, this could confirm its effectiveness.

The Acura product manager would be interested in the mention of “performance” terms, and if they are indeed positive, would affirm the manager’s success in producing a high- performing vehicle.

The “sedan” , and “drivetrain” attributes would be most applicable to the advertising manager. As the “sedan” attribute is most mentioned along with Acura and Honda, the advertising managers may want to focus advertising on sedan models to capitalize on their popularity. If advertising is already focused this way, it could indicate that their advertising is working well.

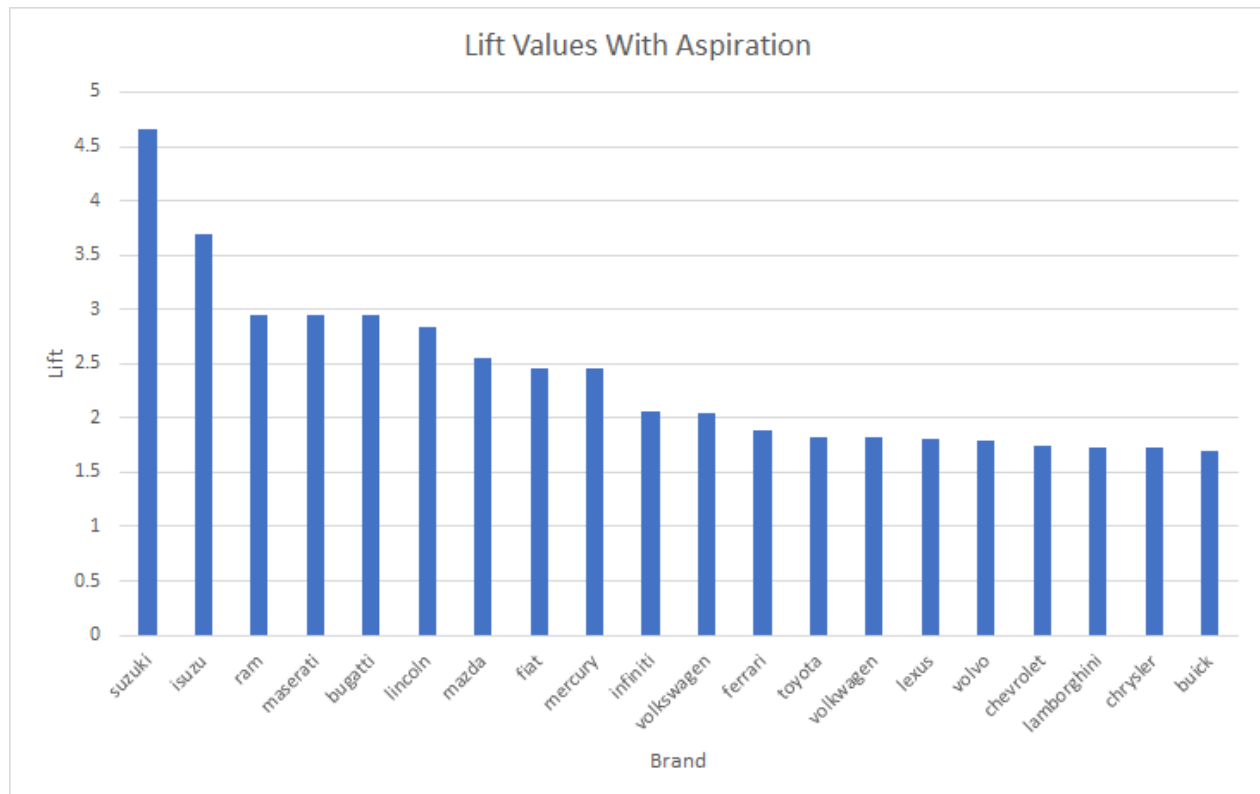
Similarly, an Acura advertising manager may be interested in hearing the prominence of “drivetrain” mentions. If the drivetrain is important for potential/current Acura owners, clearly mentioning drivetrain details in advertising would be key.

A marketing manager would also be interested in direct comparisons between brands in order to focus on where their product excels versus the competition. BMW’s performance/luxury hybrid nature could lend itself to comparisons between more distinct performance or luxury brands. As performance is most often mentioned in combination with Acura, while the price is most often mentioned with Audi, direct comparisons between these brand attributes would make for an effective marketing strategy.

Aspiration Analysis

According to “Lifestyle Brands: A Guide to Inspirational Marketing” , an aspirational brand is a consumer marketing term, which describes a brand where a large segment of its audience does not currently own the product due to economic reasons (Atwal et al, 2017). However, given the right economic conditions this segment would instead consist of transacting customers. In other words, these customers hold aspirations of owning a product from this brand one day (Atwal et al, 2017). The brands in question are often expensive, and purchase decisions by customers are made at an emotional level — often believed to enhance self-worth through the rise of social status (Atwal et al, 2017). This is evidenced with a 2013 survey from McKinsey & Company which identified that for 30% of Chinese middle-class consumers, primary motivations behind owning a premium car was attaining ‘socialstatus’ (Atwal et al, 2017).

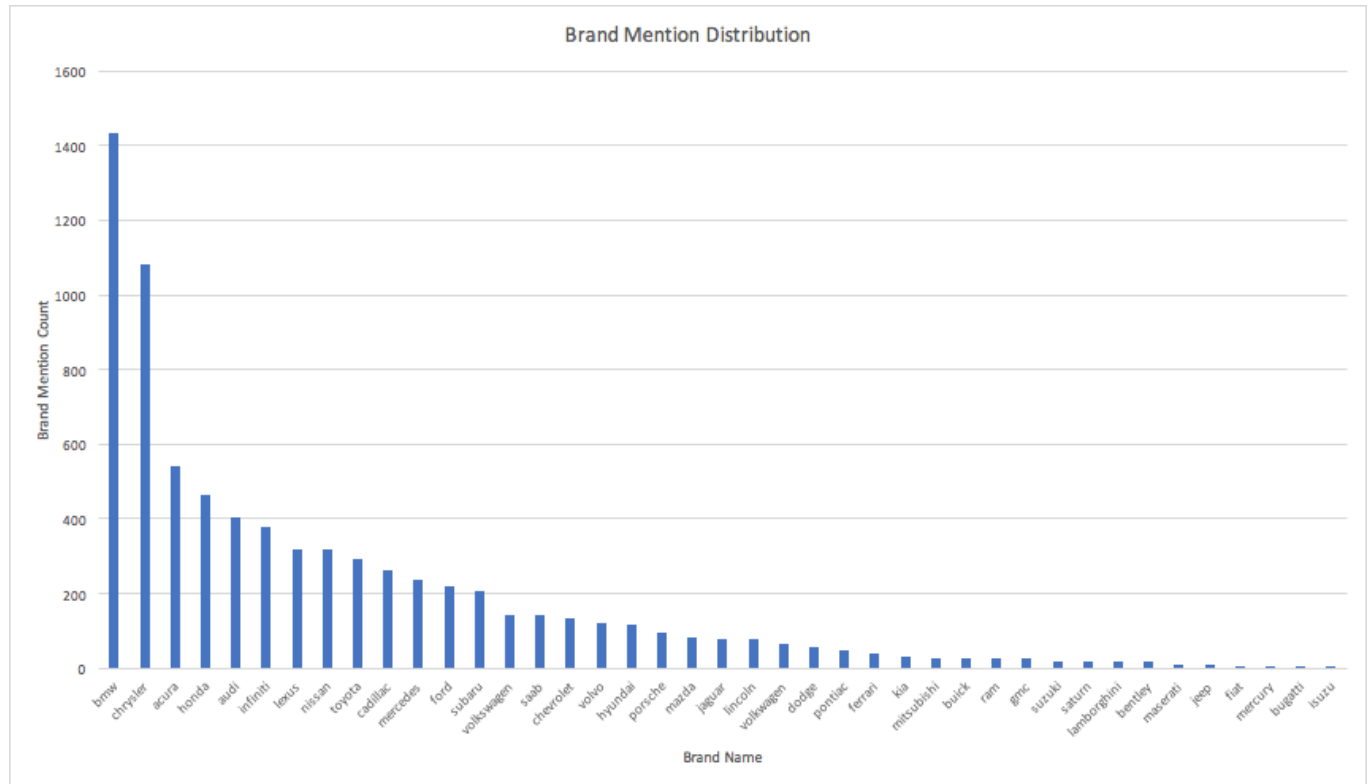
Given this knowledge, our team tested common words that communicated aspirations of purchasing such as “wish” , “dream” , “someday” and “hope” . We also tested words that communicated social status such as “prestige” , “craftsmanship” , “famous” , and “premium” . These words were grouped under the umbrella term “aspiration” , and a find-and-replace script was run on the comments. The resulting lifts of the top 20 brands and “aspiration” are displayed in the graph below.



One limitation of our existing strategy is that brands that appear sparsely in the comments can have very high lift values. This is evidenced by Suzuki, which has the highest lift value of all brands. While this initially leads us to believe that Suzuki is the most aspirational brand, the fact is that the brand has 6 co-mentions and 19 overall brand mentions, possibly resulting in a deceptively high lift value. This issue with small brand representation is also the case with other brands such as Isuzu, Bugatti, and Mercury, which all have less than 10 overall mentions.

In addition to this, our sample dataset had varying representation within brands in general. For example, the brand with the greatest representation in our sample was BMW, which had 1434 mentions, with only one more brand, Chrysler, with greater than 1000 mentions. On the other hand, the brand with the lowest representation was Isuzu with only 4 mentions. Overall, 23 out of 41 brands had less than 100 mentions. The exact distribution can be observed in the graph below. This extreme range in brand representation seems to result in misleading lift values for the scarcely represented brands. It is suspected that a more complete analysis on aspirational brands and their respective lift values would require a

dataset with less imbalance among brands. Alongside this, we could further refine the quality of such an analysis by having a larger dataset. This would assure that the scarce occurrences are not anomalies.



Bibliography

Atwal, G., & Bryson, D. (2017). *Luxury Brands in China and India*. Palgrave Macmillan.

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>