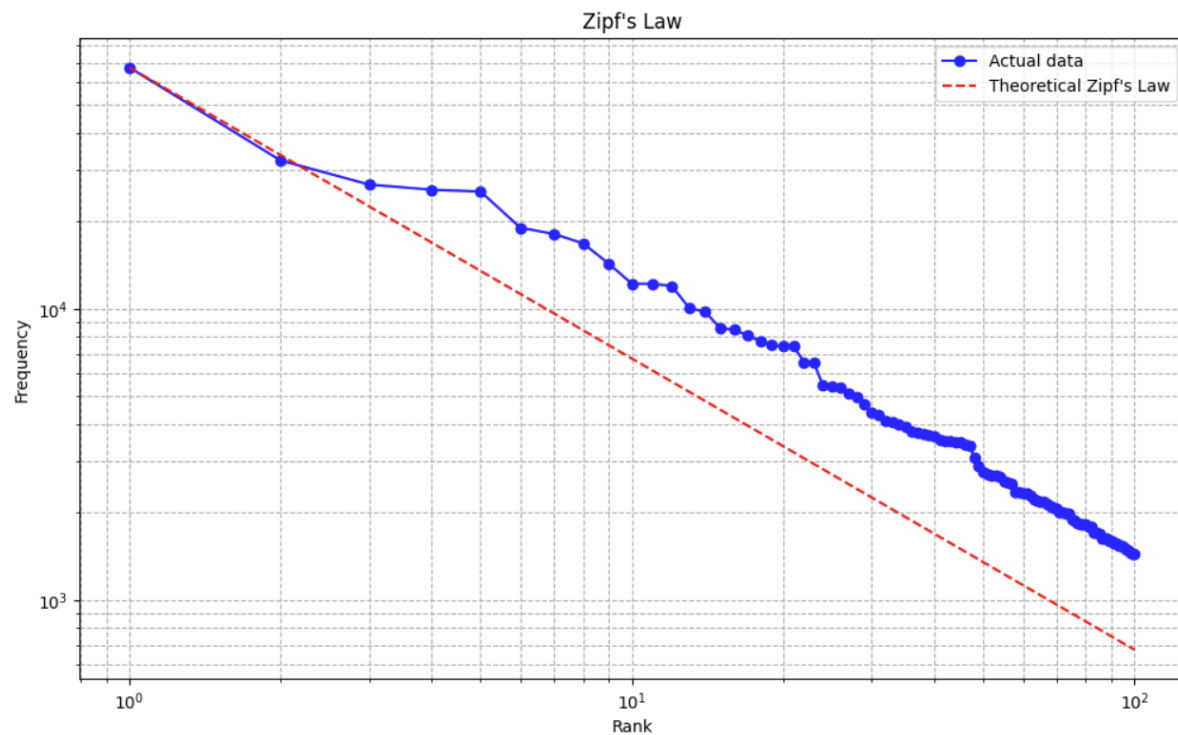


Task A:



Generally speaking, our data shows to follow the trend of Zipf's law seeing as the log frequency of our words in the data decreases as the log rank of said words increases. Moreover, we see the trend that few words occur frequently, a medium number of words occur at a medium frequency, and a large number of words occur very infrequently.

Task B: Top 10 Brands by frequency



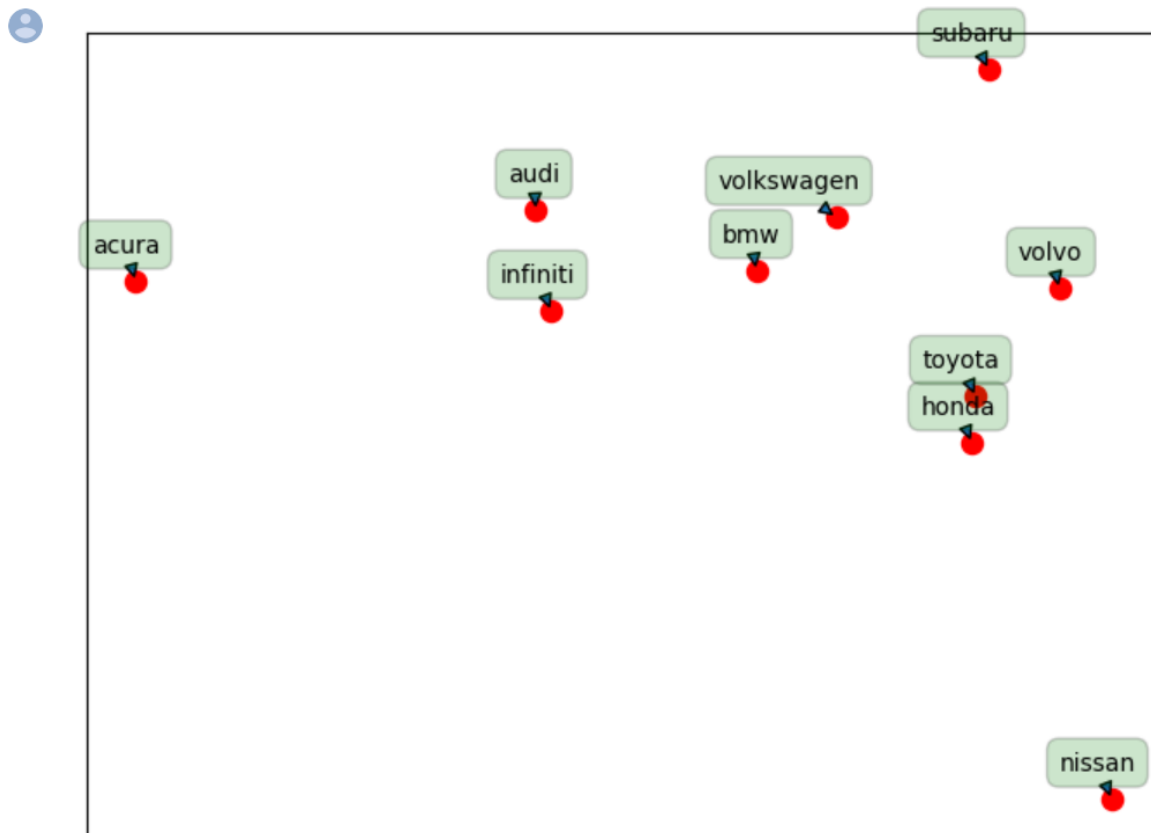
	brand	freq
0	bmw	8415
1	audi	3342
2	acura	3155
3	honda	2101
4	infiniti	1778
5	toyota	1440
6	subaru	1112
7	nissan	1076
8	volvo	863
9	volkswagen	825

Task C:



	brand2	acura	audi	bmw	honda	infiniti	nissan	subaru	toyota	volkswagen	volvo
brand1											
acura		6.772659	1.134732	0.676533	1.104682	1.845054	0.640393	1.618298	0.146391	0.646138	1.103342
audi		1.134732	6.542395	1.050921	0.322181	1.212216	0.327176	0.352309	0.093997	2.437040	1.656440
bmw		0.676533	1.050921	2.960225	0.345934	1.032031	0.347691	0.260859	0.135024	0.467747	0.522587
honda		1.104682	0.322181	0.345934	5.416853	0.392982	1.269890	0.234779	0.810462	0.619408	0.461945
infiniti		1.845054	1.212216	1.032031	0.392982	9.935936	1.292286	0.266992	0.160438	0.558805	1.002235
nissan		0.640393	0.327176	0.347691	1.269890	1.292286	14.751087	0.384546	0.529723	0.938982	0.585772
subaru		1.618298	0.352309	0.260859	0.234779	0.266992	0.384546	13.165816	0.060724	0.692849	0.720736
toyota		0.146391	0.093997	0.135024	0.810462	0.160438	0.529723	0.060724	0.939261	0.177931	0.104063
volkswagen		0.646138	2.437040	0.467747	0.619408	0.558805	0.938982	0.692849	0.177931	19.573490	1.139838
volvo		1.103342	1.656440	0.522587	0.461945	1.002235	0.585772	0.720736	0.104063	1.139838	22.983971

Task D:



Task E:

Task E:

What insights can you offer to your client from your analyses in Tasks C and D?

The analyses in Tasks C and D reveal information about the lift associations between brands. Firstly, the lift matrix returned the associative values of brands, with the higher numbers Signifying that commenters mentioned the brands together at a greater rate than the lower Numbers. For example, Honda and Audi have a low lift value of 0.322181 which seems logical Because the brands are not generally perceived as competitors. Conversely, Acura and Infiniti Have a high lift value of 1.845054, which makes sense as they can be seen as oppositional. Running the inverted lift returned a matrix characterized by the inverse lift values of the brands, Wherein a higher lift value represents dissimilarity. A low lift value in this matrix signals that the Brands are being talked about within the same comments, suggesting an oppositional or Aspirational relationship between the two. For example, the inverted lift between Volkswagen And Audi is 0.410334, which may be interpreted as commenters seeking to upgrade from

One brand to the other.

The MDS plot serves as a visual representation of the inverse lift associations, displaying some Weak clusters and some irregularities. Brands that may be seen as ‘utility’ or ‘value’ makes Are clustered, as noted with Toyota and Honda being situated together. Brands that may be Seen as ‘luxury’ or ‘high class’ tend to be clustered, as evidenced by the Audi, Infiniti, BMW, And Volkswagen group. Acura and Nissan deviated from their expected groups, Suggesting that commenters do not talk about them in reference to other brands. These analyses can be useful for clients as they can note which other brands commenters Frequently associate with their own brand, informing marketing, production, and pricing decisions.

F:



	attr	freq
0	performance	2116
1	luxury	1888
2	price	1848
3	engine	1408
4	interior	1352
5	sport	1187
6	tires	913
7	class	950
8	handling	1060
9	power	1064



(half a page)

G:

(around a page) (summary for each brand)

From our frequency table, we can see which attributes were the most common from our list of overall words. This list was obtained by scanning the top 100 words for adjectives and general descriptors of cars. In doing so, the top five most used attribution words were performance, luxury, price, engine, and interior. The rest of our top 10 attributes are in the table above.

To see which attributes were most strongly correlated with the top 10 brands, we calculated the lift values between the top 10 attributes and the top 10 brands. To better visualize the lift values, we created a heat map that showcases which lift values were particularly high. In doing so, we can see that our highest lift value is between the brand Volvo and the attribute luxury. This came as a bit of a surprise to our group seeing as we initially did not associate Volvo as the highest luxury brand. However, this gives us lots of insight moving forward to be able to recommend Volvo to customers seeking a luxurious vehicle.

In addition, we saw somewhat higher lift values for the brand Audi and the attribute interior. This lets us know that customers may generally like their Audi cars for their nice interior, again showing us what cars could be recommended to people seeking this attribute.

It is also worth noting which brands have very low lift in terms of the attributes. Generally speaking, the top ten attributes are quite positive words. For instance, attributes like luxury and performance are generally associated with good cars and are seen as good qualities. Knowing this, we also see that the brand Toyota has very low lift values for all attributes. This means that people are not generally using these positive words to describe their Toyota's. Moreover, people may describe Toyota in terms of their longevity and reliability (or perhaps with more negative words), but since they were not the most popular attributes in our data, the lifts for Toyota are quite low.

Lastly, we saw Infiniti as having high lift values for the attributes interior and luxury. This came as a bit of a surprise since many people in our group do not see as many Infinitis in the road these days, however, this does show us that people who do own these cars value them potentially for their luxurious interiors.

Moreover, this attribute and brand analysis gives us insight as to what qualities people search for in their cars, those being the attributes from the table above. However, seeing as lift values for luxury, interior, class, and handling are all generally high across multiple brands, we can see that these attributes are the most sought after or highly regarded. Moreover, for businesses, it may be useful to use these traits to describe or market their vehicles seeing as these are the attributes that show up frequently and people value in their respective cars.

H:

attribute	aspire
brand	
acura	1.031112
audi	1.041471
bmw	1.055005
honda	0.818811
infiniti	1.279139
nissan	0.963782
subaru	0.877309
toyota	0.353665
volkswagen	1.111346
volvo	1.166847

Our process for finding aspirational brands first consisted of our group coming up with synonyms to the word “want”. This is because when people talk about car ownership, they generally talk about “wanting” a car. However, when it comes to aspirational brands, people may also use terms like “dream” or “desire” or “hope” when they describe aspiring to own a vehicle. Moreover, we created a dictionary of words of multiple tenses that fit this profile. Once we did this, we proceeded to attribute these words to the word “aspire”, essentially replacing all the different aspirational words with the word aspire. Once we did this, we calculated the lift between our top ten brands and the word “aspire” (which now represents many aspirational words). In doing so, we calculated the lift values above for the top 10 brands.

In terms of finding “the most aspirational brand” there does not seem to be one brand that has a drastically higher lift value to aspirational terms than another. We can see that Infiniti has the

highest lift value in comparison with aspiration at 1.28, however, it more generally seems like there is a cluster of brands that are more aspirational than another cluster. Moreover, the brands Acura, Audi, BMW, Infiniti, Volkswagon, and Volvo seem to have higher lifts, and the brands Honda, Nissan, Subaru and Toyota seem to have the lowest lifts. Seeing this, there seems to be an association between our higher lift valued brands and our attributes concerning luxury. Whereas our group of lower lifts is less correlated with cars concerning class or luxury. Moreover, this shows us that brands seen as “aspirational” are luxury brands of cars. This makes sense seeing as many people “dream” of having really nice cars, and this analysis speaks to this phenomenon.

Business Impact:

Task E:

What insights can you offer to your client from your analyses in Tasks C and D?

Task F:

What are 5 most frequently mentioned attributes or features of cars in the discussions? Which

attributes are most strongly associated with which of these 5 brands? There is no need to plot the

plot for this question.

Task G:

What advice will you give to your client from Task F?

Task H:

Which is the most

aspirational

brand in your data in terms of people actually wanting to buy or

own? Describe your analysis. What are the business implications for this brand