**Text Analysis Homework 2: Edmunds Forums**

Daxi Cheng, Jianjie Zheng, Reece Wooten, Nick Scheifler

1. Forum Choice and URL:

Topic: General -- SubTopic: Sedans -- Forum: Midsize-Sedans-2.0

URL: <https://forums.edmunds.com/discussion/7526/general/x/midsize-sedans-2-0>

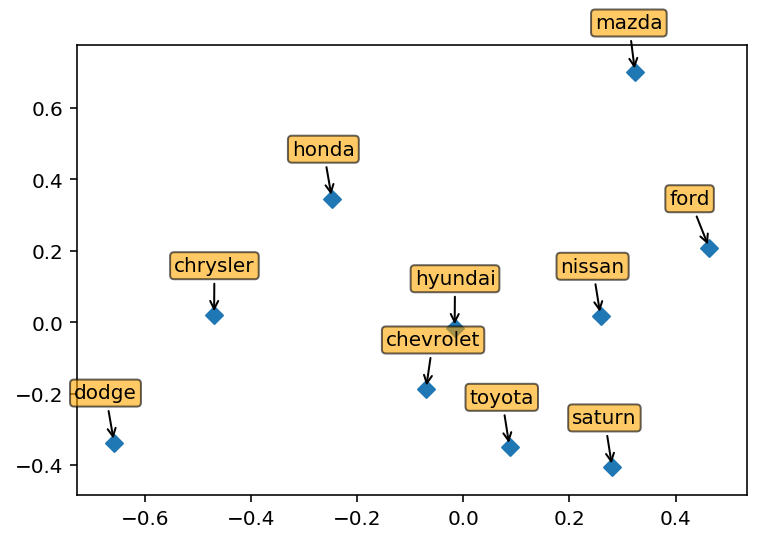
2. Top 10 Brands by Frequency:

|  |  |
| --- | --- |
| **Brand** | **Frequency** |
| Honda | 2777 |
| Toyota | 2358 |
| Ford | 1624 |
| Nissan | 858 |
| Hyundai | 853 |
| Mazda | 750 |
| Chevrolet | 408 |
| Chrysler | 334 |
| Saturn | 334 |
| Dodge | 290 |

3. All lift calculations in table form:

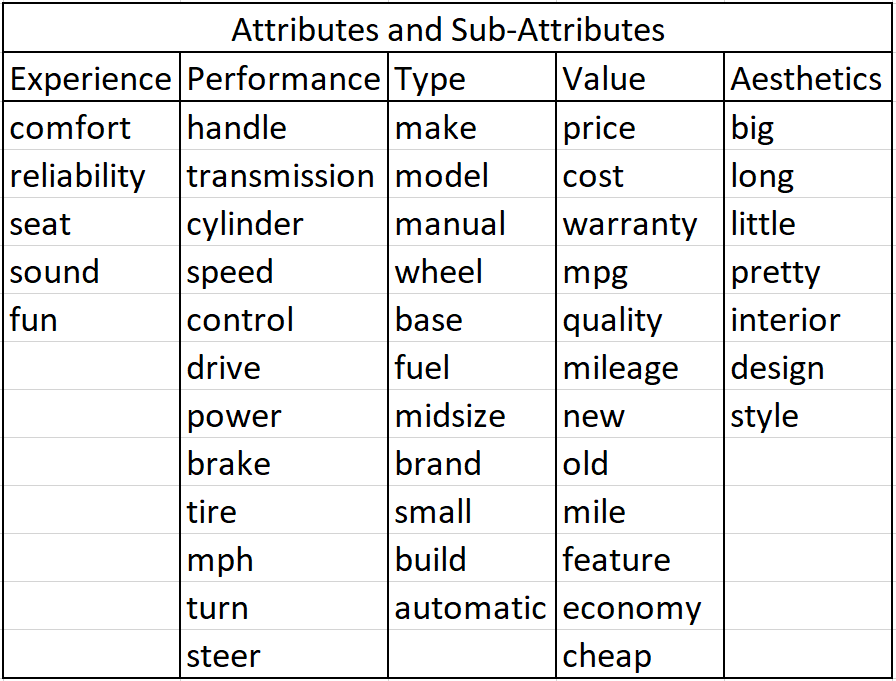
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **honda** | **toyota** | **ford** | **nissan** | **hyundai** | **mazda** | **chevrolet** | **chrysler** | **saturn** | **dodge** |
| **honda** | 2.17 | 1.35 | 1.19 | 1.32 | 1.43 | 1.14 | 1.31 | 1.37 | 1.33 | 1.18 |
| **toyota** | 1.35 | 2.56 | 1.3 | 1.46 | 1.39 | 1.19 | 1.57 | 1.37 | 1.55 | 1.35 |
| **ford** | 1.19 | 1.3 | 3.71 | 1.27 | 1.27 | 1.33 | 1.4 | 1.39 | 1.35 | 1.02 |
| **nissan** | 1.32 | 1.46 | 1.27 | 7.03 | 1.49 | 1.45 | 1.74 | 1.26 | 2.48 | 1.14 |
| **hyundai** | 1.43 | 1.39 | 1.27 | 1.49 | 7.07 | 1.23 | 1.59 | 1.93 | 1.63 | 1.44 |
| **mazda** | 1.14 | 1.19 | 1.33 | 1.45 | 1.23 | 8.04 | 1.22 | 0.84 | 0.94 | 1 |
| **chevrolet** | 1.31 | 1.57 | 1.4 | 1.74 | 1.59 | 1.22 | 14.78 | 2.21 | 4.6 | 1.38 |
| **chrysler** | 1.37 | 1.37 | 1.39 | 1.26 | 1.93 | 0.84 | 2.21 | 18.05 | 1.19 | 3.74 |
| **saturn** | 1.33 | 1.55 | 1.35 | 2.48 | 1.63 | 0.94 | 4.6 | 1.19 | 18.05 | 1.18 |
| **dodge** | 1.18 | 1.35 | 1.02 | 1.14 | 1.44 | 1 | 1.38 | 3.74 | 1.18 | 20.79 |

4. Multi-Dimensional Scaling Map



5. Top 5 Attribute Selection

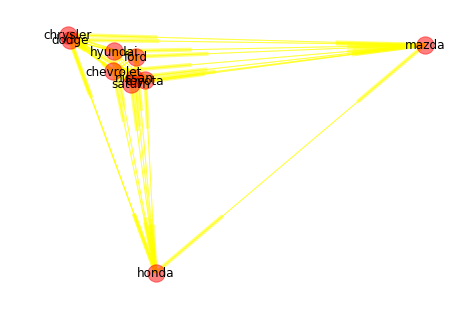
To decide on the 5 attributes, we sorted words by frequency and looked at the top 60-70 words that could be interpreted as attributes. We then pared that down to 52 terms and collapsed them into 5 umbrella attributes that each contain a number of sub-attributes. They are listed below:



6. For **task E**, provide all details of your analysis – e.g., how you measured “aspirational” and how you found the most aspirational brand.

We defined “aspirational” as having a high measure of centrality, conditioned on the occurrence of ideas related to purchasing. In other words, given that people are talking about purchasing, what brand is talked about the most in relation to other brands? We use the conditional lift of different brands conditional on people talking about purchasing decision as the similarity measure. Then we take the inverse of that as the distance to build a graph model to calculate centrality scores. Mazda scored the highest on this metric with a closeness centrality score of 0.98. Assuming that these co-occurrences are of positive sentiment, we can suppose that among these users in the midsize sedan market, Mazda is a brand of aspiration.

Here we have the graph plotted with Networkx with spectral layout. The graph is organized with the eigenvectors of Laplace matrix. We can see that Mazada (and Honda) really stand out.



7. Advice/insights based for brand, product and advertising managers.

The top 10 brands from the corpus were Honda, Toyota, Ford, Nissan, Hyundai, Mazda, Chevrolet, Chrysler, Saturn, Dodge.

The two brands with the highest lift were Chevy, and Saturn with a lift value of 4.6. This signifies that Chevy and Saturn are talked about a lot together in the posts. Also on the MDS graph Chevy and Saturn are relatively close together and look to be in the same cluster. This indicates that Chevy and Saturn compete in the same space when it comes to people comparing two cars. This information could be used to restructure how Chevy or Saturn competes in the automotive industry.

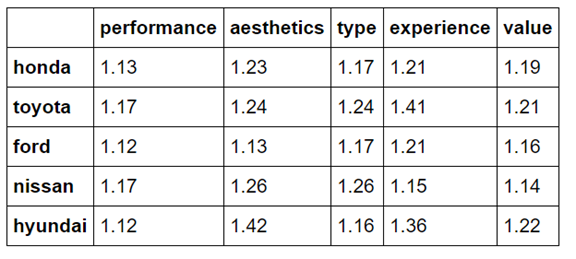
Given the MDS map, two other brands of interest are Hyundai and Dodge. Hyundai is talked about the most with other brands. This information can be used in an effort to convert conversations and top-of-mind associations regarding Hyundai to actual purchasing behavior. In contrast, Dodge is not talked about a lot with relation to other brands and a marketing effort can be taken to increase the popularity of Dodge to increase its conversation rate.

From the previous part of our project, we calculated the lift values for the top 10 brands to find the association between brands. For Task D, we are interested in finding the conditional lift values to determine which attributes will be useful for the product managers and marketing/ advertising managers.

In the following table, we calculated the conditional lift values for each of the top 5 brands given each of the attributes. Based on these values, we are able to suggest which attributes each brand is associated with the most.

For product managers, since they are more interested in exploring new products/functions of cars, they need to be more sensitive to what people are discussing. Therefore, if you look at the table below, a product manager should pay attention to the highest value in each of the attributes. In this case, he/she should explore toyota and nissan for performance, hyundai for aesthetics, nissan for type, toyota for experience and hyundai for value. So based on which attribute the manager is interested, manager should take a look into the brand with highest value and study/explore why people all talk about this brand. Then, he/she would be able to come out with some strategy to improve his/her product.

For marketing managers, we’d recommend looking at low conditional lift values for each attribute since it may indicate a relative weakness in brand perception. In this case, Ford and Hyundai have lowest value for performance, Ford has the lowest value in aesthetics, Hyundai has the lowest value in type, Nissan for experience and value. After identifying the right target, the manager can come up with some ways to improve brand image by comparing to other top brands within the attribute. The marketing manager could design a marketing campaign which attempts with an increase in conditional lift as a metric of success.



Regarding BMW: it turns out that BMW is indeed highly regarded as something resembling the “ultimate driving machine.” BMW achieves a lift score of **1.75 conditioned on “experience”** and **1.57 on “performance.”** This is significantly higher than the top-mentioned brands. Assuming that the sentiments are positive, we can reasonably conclude that users consider the BMW driving experience to be superior.