**User Generated Content Analytics Assignment 1**

**New Due Data (09/16 by 11:59 p.m. on Canvas)**

In this assignment you have been hired as an analytics consultant by JD power and Associates, who wants to perform a competitive analysis of the entry level luxury car market in the USA. Your job is to give advice/insights to these individuals based on the analysis of social media conversations. The detailed tasks are described below.

Scrape Edmunds.com discussion forums to fetch about 5,000 posts about cars from the “Entry Level Luxury Sedans” forum. Use Web Scraper to scrape the data. Follow the Zoom recording on scaping.

1. Once you fetch the data, find 10 most popular (i.e., most frequently mentioned) brands. To do this, use the python script wordfreq.py on pythonanywhere.com, which creates an output file with words and their frequencies.

**Task A:** For the 10 brands you have chosen, calculate the lift ratio for associations between the brands. Use the python script lift.py to calculate lifts.

Show the brands on a multi-dimensional scaling (MDS) map. Use mds.py for this purpose.

**What insights can you offer from your analysis in Task A? Be very specific in your response, and provide insights as bullet points and not as a long paragraph.**

**Task B:** Pick 5 most frequently mentioned attributes from the posts. Which brands (choose the 5 most frequently mentioned instead of 10) have the highest association with these attributes? E.g., Audi with style, BMW with performance, etc. (these are just examples, the reality can be very different).

**What insights can you offer based on your analysis in Task B? Again, be very specific in your response.**

Note: Given that this assignment has to be completed before we learn sentiment analysis, I am NOT asking for sentiments regarding these brands and attributes. For the time being, you can assume a **positive sentiment in all cases.**

**Task C:** Which is the most **aspirational** brand among the top 5? How did you find the answer? **Show all steps.**

**Provide the following details in your write-up (I am not setting a strict page limit, but 3-4 pages, single-spaced, 12 font size should be sufficient even with screenshots)**

1. Names of group members
2. Which forum you chose (provide URL)
3. Which 10 brands you chose and why – provide # mentions (a summary table is good here)
4. Show all lift values in a table.
5. Show the MDS map (put a screenshot in your report)
6. Insights from Task A
7. State the attributes and the basis of their selection (again, a summary table with frequencies is good here).
8. Insights from Task B
9. Show all details of “aspirational brand” analysis.

Submit a word or pdf file with the above information. Write your name inside the document. Do not submit any CSV or other data files.

