Text Analytics Group Assignment

In this assignment you are an analytics consultant to a (i) brand manager, (ii) product manager and (iii) advertising manager. Your job is to give advice/insights to these individuals based on the analysis of social media conversations. The detailed tasks are described below.

Develop a crawler/scraper to fetch messages posted in Edmunds.com discussion forums. The crawler output should be a `.csv` file with the following columns: date, userid, and message (even though you will only use the messages in your analysis).

Fetch around 5,000 posts about cars from a General topics forum. Do NOT choose a forum dedicated to a particular brand or model. Instead, you can choose the General & Sedans categories and then select, for example, the Entry Level Luxury forum (https://forums.edmunds.com/discussion/2864/general/x/entry-level-luxury-performance-sedans). The idea is to have multiple brands and models being discussed without one of them being the focal point.

Once you fetch the data, find the top 10 brands from frequency counts. You will need to write a script to count the frequencies. **Be sure not to count a mention more than once per post, even if it is mentioned multiple times.** Replace **models** with **brands** so that from now on you have to deal with only brands and not models. You will need another script for this job. This step is meant to help simplify the analysis. A list of model and brand names (not exhaustive) are provided in a separate file (`models.csv`). Modify the list as needed.

Task A: Identify top 10 brands by frequency. From the posts, calculate lift ratios for associations between the brands. You will have to write a script to do this task. Show the brands on a multi-dimensional scaling (MDS) map.

Task B: What insights can you offer brand managers from your analysis in Task A? Choose two brands that you can offer the most interesting/useful insights for.

Task C: What are the 5 most frequently mentioned attributes of cars in the discussions? Note that the same attribute may be described by different words – e.g., pick-up and acceleration may both refer to a more general attribute, "performance". You have to make suitable replacements. Now pick the 5 most frequently mentioned brands. Which attributes are most strongly associated with which of these 5 brands?

Task D: What advice will you give to a (i) product manager, and (ii) marketing/advertising manager of these brands based on your analysis in Task C?

Task E: 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?

Provide the following details in your submission:

1. Which forum you chose (provide URL)
2. Which 10 brands you chose – provide the frequency table
3. Show all lift values in a table.
4. MDS map.
5. State the 5 attributes you chose (again, a table is good here).
6. For task E, provide all details of your analysis – e.g., how you measured "aspirational" and how you found the most aspirational brand.
7. Advice/insights based on your analysis for brand, product and advertising managers.

Your submission should include all scripts (and corresponding input files) as well as your answers to the questions above.