Data Selection Summary: Sephora Website

1) **Source of Data:**

- The dataset was sourced from Kaggle. This dataset provides a comprehensive list of products available on the Sephora website, including various attributes related to each product.
    - https://www.kaggle.com/datasets/raghadalharbi/all-products-available-on-sephora-website

2) **How many data files? What are the relationships among various data files?**

- The dataset consists of a single data file in CSV format. The file contains information about various products available on Sephora's website, including attributes such as product ID, name, category, price, ratings, number of reviews, and additional features like ingredients and brand names. Since there is only one file, the relationships among various data elements will be established through the attributes contained within that file.

3) **How many tables do you anticipate in your database?**

- Based on the attributes present in the dataset, I anticipate creating approximately 6-7 tables in the database. The proposed tables are as follows:
    - **Product Table:** Contains primary product details (e.g., product ID, name, description, price, and category).
    - **Brand Table:** Lists unique brands, linked to products by brand ID (e.g., brand ID and brand name).
    - **Category Table:** Contains unique product categories, linked to products by category ID (e.g., category ID and category name).
    - **Review Table:** Captures user reviews, ratings, and the relationship to specific products (e.g., review ID, product ID, user rating, and review text).
    - **Ingredient Table:** Contains information on product ingredients, possibly linked to products by product ID.
    - **Price Table:** Tracks price history over time, linking product ID with historical price data.
    - **Marketing Table:** Captures marketing flags, sales, and special offers (e.g., marketing ID, product ID, marketing flags, online only, limited offer, exclusive).
- These tables will allow for efficient organization of the data and support complex queries related to product characteristics, brand performance, customer feedback, and pricing strategies.

** brand ID, category ID, etc. do not currently exist in the table and will need to be created*