Pretend you are the Head of Marketing at Sephora. Corporate is championing a significant sales surge to make Q3 and Q4 of 2025 the company's most successful quarters to date. To accomplish this, you want to know what makeup products to highlight on Sephora.com–specifically, whether you feature celebrity-branded makeup products, such as Selena Gomez's Rare Beauty, or non-celebrity-branded makeup products, such as Charlotte Tilbury, on the homepage. It is highly likely that consumer sentiment towards celebrity-branded makeup products differs from non-celebrity makeup brands, especially because celebrity makeup brands are increasingly saturating the market. So, it is your mission is to assess whether celebrity makeup products have more polarized reviews, and use this insight to make a strategic decision as to what products to showcase on the homepage.

To do so, you will conduct sentiment analysis, using Python's VADER sentiment analysis package, on a dataset of 18 makeup products–9 celebrity-branded and 9 non-celebrity-branded. You will also build a logistic regression model to predict whether particular review keywords are more indicative of a celebrity versus non-celebrity branded makeup products. Please use the information and files found in https://github.com/gffiveash/DS4002_CS3. By analyzing these reviews, you'll uncover key trends that will help guide Sephora's product promotion strategy. Your deliverable will be a GitHub repository and a PDF document, outlining the process, select EDA visualizations, results, and conclusions/learnings. These results should detail whether or not celebrity brands are more likely to have more polarized reviews. This work will empower you and the Sephora team to optimize sales by selecting products for promotion based on customer sentiment in online reviews.