Project 2 Linear regression Proposal

I. Question/Need:

What is the question behind your analysis or model and what practical impact will your work have?
What are the most important factors that influence the rating of skincare products?

The beauty industry is a multibillion-dollar industry that is continuously growing since 2012 and is estimated to be $189.3 billion by 2025 (source: Statista.com). More people are beginning to use skincare at a younger age, and with so many products on the market, how could one decide what to buy based on their budget? By modeling the rating of skincare products with various product features, business can understand what customers are valuing in an item.

Who is your client and how will that client benefits from exploring this question or building this model/system?
The clients will be both businesses that sell beauty products and consumers. Businesses can use the model to understand what important to a positive rating and can sell products in ways to generate more positive reviews leading to more business. Consumers can understand how ratings of products are determined, and they can use this information to decide whether the rating reflects what is important to them to make better purchasing decisions.

II. Data Description:

What dataset(s) do you plan to use, and how will you obtain the data? Please include a link! (The link can be to the dataset you're downloading, the site you're scraping, etc.)
Web scraping on:
https://www.ulta.com/

Category: skincare, moisturizers (1,351 results)
https://www.ulta.com/skin-care-moisturizers?N=2796

What is an individual sample/unit of analysis in this project? In other words, what does one row or observation of the data represent?
Name of product, rating, reviews, and the features below

What characteristics/features do you expect to work with? In other words, what are your columns of interest?

1. Price (USD)
2. Volume (oz)
3. Price per oz
4. Number of reviews
5. Length of product name
6. Length of product description
7. Number of benefits
8. Number of key Ingredients (or length of ingredients list)
9. Number of times the word ‘clinical’/ ‘clinically’ appears
10. Number of pros and cons of the item based on reviews

11. Percentage of people who would recommend the product

If modeling, what will you predict as your target?
Target: rating

III. Tools:

How do you intend to meet the tools requirement of the project?
Web scraping: BeautifulSoup and Selenium
Linear regression: Python
Visualizations: Pandas and matplotlib

Are you planning in advance to need or use additional tools beyond those required?
I don't plan to need additional tools.

IV. MVP Goal:

What would a MVP Example look like for this project?
Scraping features from the website for over 1000 products. Then, perform EDA on the dataset followed by a linear regression model.