Problem Identification
The efficacy of skincare products has always been a debated subject. With a vast variety of products available in the market, it is imperative to figure out whether a product lives up to its claims. How can we have a good idea if a skincare product is effective based on their ratings and their ingredients?

Objective:
Develop a model that can accurately predict the efficacy ratings of skincare products relying on their ingredients and user reviews. The model can be a reliable tool for consumers, helping them make informed decisions before buying a skincare product. It aims to ensure that customers get value for their money and achieve the desired results from the skincare products they use.

Data:
The data for this project will be sourced from online retailers like Sephora, which offers a wealth of information on ingredients used in different skincare products and their associated user ratings. A Kaggle dataset provided by Raghad Alharbi, which contains scraped data from Sephora's website, will be a primary resource.

The Sephora Website dataset encompasses:
21 features
9168 number of observations
Future efforts will be directed towards collecting additional data and expanding the product range beyond those available on Sephora.

Client:
The intended clients for this project are customers who regularly purchase skincare products. This includes men and women of all ages who wish to make informed decisions about their skincare products and aim to spend their money on effective products.
This project aims to increase customer satisfaction by providing a data-driven tool for choosing skincare products.

Approach:

1. Data collection: Collect data from online retailers like Sephora, including information on the ingredients used in different skincare products and their associated user ratings.
2. Data preprocessing and cleaning: Clean and preprocess the data to remove missing values, handle outliers, and convert categorical variables to numerical features.
3. Feature engineering: Extract relevant features from the data, such as ingredient composition, product type, and user sentiment.
4. Model selection: Select an appropriate machine learning algorithm that can predict skincare product efficacy ratings based on the available features.
5. Model training and evaluation: Train the model on a portion of the data and evaluate its performance on a holdout set using metrics such as accuracy, precision, and recall.

6. Model refinement and tuning: Refine the model by tweaking hyperparameters, selecting different features, and testing different algorithms to improve its performance.

Deliverables:
Code: Provide well-documented and organized code that can reproduce the data preprocessing, feature engineering, and model training process.

Report: Create a report that summarizes the project goals, methodology, and results. The report should include visualizations of the data, model performance metrics, and interpretations of the findings.

Presentation: Create a slide deck that presents the key findings and conclusions from the project. The presentation should be visually appealing and easy to follow.