# Objective

Skincare shopping can be overwhelming, especially for people with dry or sensitive skin. Many cosmetic products list long, complex ingredients that users often do not understand. This project aims to simplify the skincare selection process by using a content-based filtering system. We use natural language processing and visualization techniques to recommend and compare cosmetic products based on their ingredients.

# Dataset Description

The dataset contains 1,472 cosmetic products obtained from Sephora. Each entry includes:

* Product Name
* Product Type (Label)
* Skin Type
* Ingredient List

The ingredient list is the core feature of this analysis. Products are filtered to include only moisturizers for dry skin.

# Methodology

1. Data Filtering: Focused on moisturizers for dry skin by filtering on the 'Label' and 'Skin Type' columns.
2. Text Pre-processing: Tokenized the ingredient list, converted to lowercase, and created a unique vocabulary.
3. Vectorization: Created a binary document-term matrix representing which ingredients appear in which products.
4. Dimensionality Reduction: Applied t-SNE to reduce high-dimensional ingredient vectors into 2D for

visualization.

1. Interactive Visualization: Used Bokeh to plot and interactively explore product clusters.
2. Comparison Tool: Developed functionality to compare two products' ingredient profiles for similarity or difference.

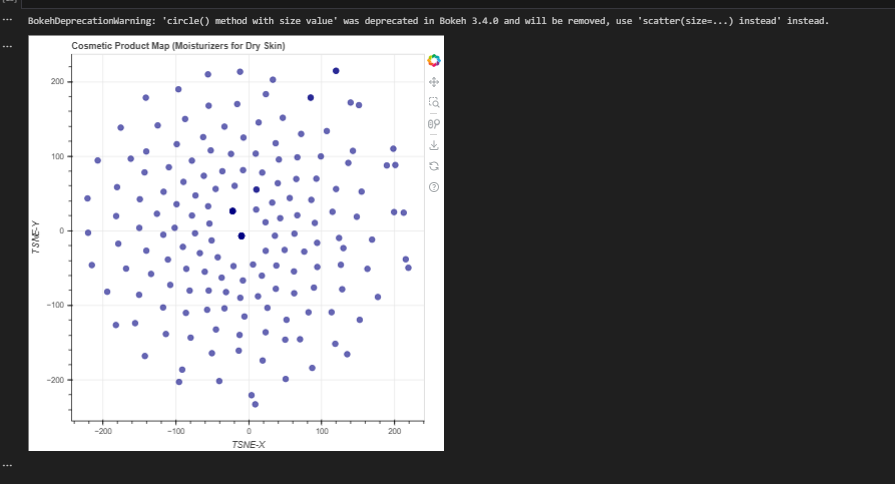
# Technologies Used

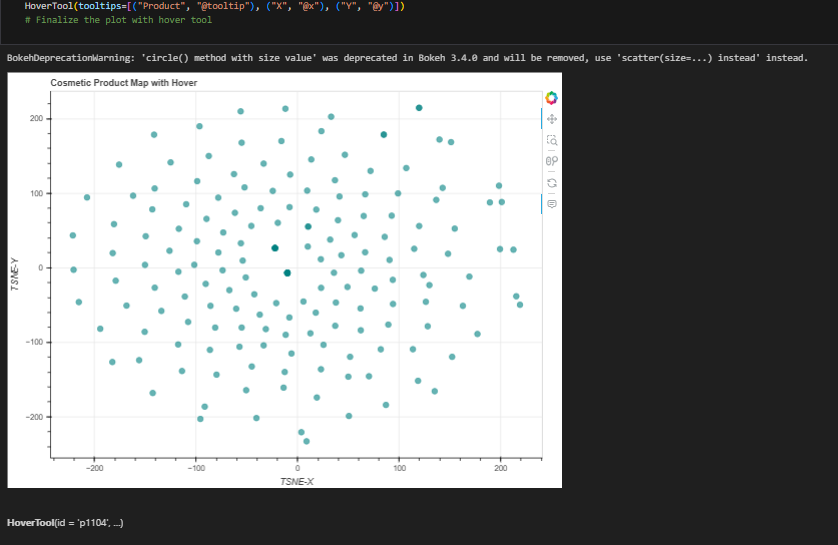
* Python
* Pandas, NumPy
* Scikit-learn (t-SNE)
* Bokeh (interactive plotting)
* Jupyter Notebook / VS Code

# Visualizations

The visual output includes:

* A 2D scatter plot of products based on ingredient similarity
* Color-coded categories/brands for better differentiation
* Hover tooltips for product names





# Insights and Findings

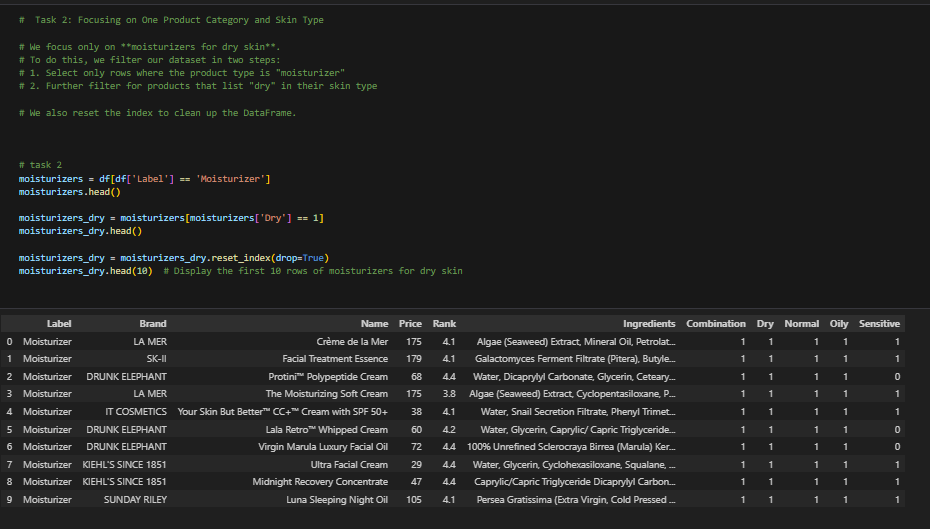
* Ingredients such as water, glycerin, and dimethicone are found frequently.
* The t-SNE plot clearly grouped similar products together based on ingredients.
* The comparison function allowed us to identify which ingredients were shared or unique to any twoproducts.

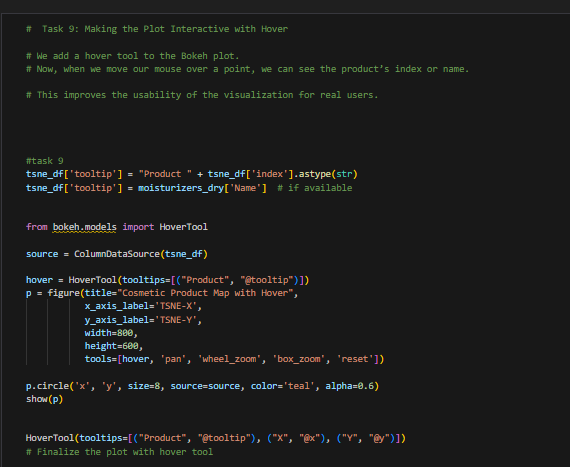
# Conclusion

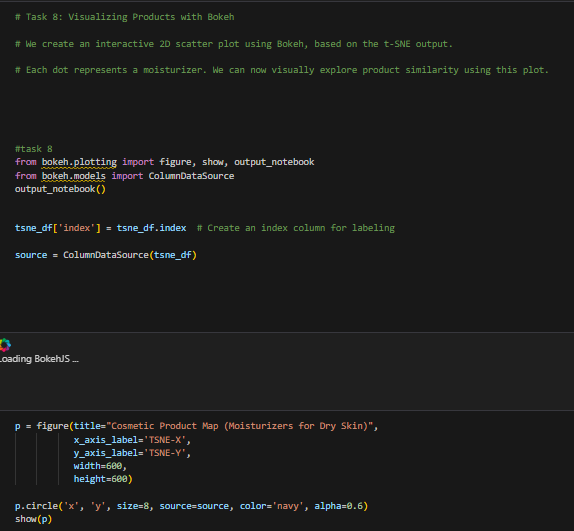
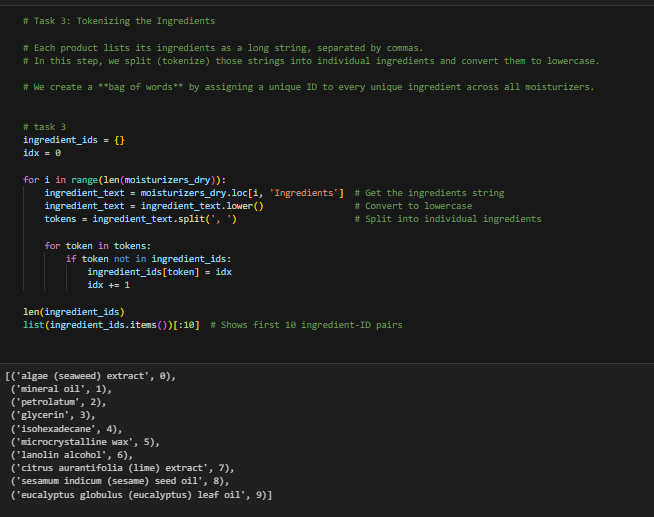
This project successfully demonstrates how data science can help simplify cosmetic choices. By analyzing ingredient compositions and visualizing similarity, users can make more informed skincare decisions. Future extensions could include user-based filtering, sentiment analysis on reviews, or deployment as a web

application.

# Appendix





* 
* Code snippets
* Additional screenshots
* Ingredient frequency tables or comparison results