

**School of Computer Science and Engineering**

**Master of Computer Applications (MCA)**

**Digital Assignment -PMCA507L(Machine Learning)**

*Skincare Recommendation Model*

**PROJECT REPORT**

BY : Stuti Gupta23MCA1002) &

Shivani Jain(23MCA1084)

*Abstract*— *This project report outlines the development of a skincare recommendation model and the implementation of this model . The report provides a comprehensive overview of the project's objectives, the problem statement, data sets used, tools employed, methodology, results, and includes relevant screenshots.*

Table of Content

|  |  |
| --- | --- |
| S.no | TOPIC |
| 1 | PROBLEM STATEMENT |
| 2 | INTRODUCTION |
| 3 | DATA SET USED |
| 4 | TOOLS USED |
| 6 | METHODOLGY |
| 7 | RESULTS |
| 8 | SCREENSHOTS |

**PROBLEM STATEMENT**

The problem at hand is to be able to get product recommendations based on an available product. The model aims to recommend similar products that have either same ingredients or serve the same purpose with different costs. The model should be able to analyze a user's input data, including information about their skin characteristics, lifestyle, and desired outcomes, and then recommend the most suitable skincare products from the variety of products.

The model should be trained on a large dataset containing information about different skincare products, their ingredients, and their suitability for different skin types and concerns. Additionally, the dataset should include user reviews, ratings, and feedback on various skincare products, which can help the model learn from real-world experiences.

**INTRODUCTION**

In today's world, the skincare industry is flooded with an overwhelming number of products, each claiming to address specific skin concerns and promising miraculous results. With so many options available, it can be incredibly challenging for individuals to navigate through the vast array of products and determine which ones are truly suitable for their unique skin types and needs. This is where machine learning can play a pivotal role in simplifying the process and providing personalized recommendations tailored to each individual's skin profile.

The proposed skincare recommendation model leverages the power of machine learning algorithms to analyze a user's skin characteristics, lifestyle factors, and personal preferences, and then suggests the most appropriate skincare products from a comprehensive database. By harnessing the capabilities of advanced data analysis and pattern recognition, this model aims to revolutionize the skincare industry, empowering users with tailored solutions that address their specific concerns while minimizing the guesswork and trial-and-error processes often associated with finding the right skincare regimen.

**DATA SET USED**

Our project starts with the collection of the skincare products data which we collected from Kaggle (<https://www.kaggle.com/datasets/eward96/skincare-products-clean-dataset> ):

Data columns include “product\_name”, ‘product\_url’, ‘product\_type’, ‘clean\_ingreds’ and ‘price’.

**TOOLS USED**

The development of the Skincare Recommendation Model involved the use of various tools and technologies:

*Data Processing & Feature Transformation:*

* pandas: A powerful data manipulation and analysis library for Python.
* numpy: A fundamental package for scientific computing in Python, providing support for large multi-dimensional arrays and matrices.
* re: A Python module that provides regular expression matching operations.
* sklearn.feature\_extraction.text: A utility from Scikit-learn to convert text data into a numerical feature representation using the TF-IDF (Term Frequency-Inverse Document Frequency) statistic.

*Analysis and Visualization:*

* matplotlib: A comprehensive library for creating static, animated, and interactive visualizations in Python.
* seaborn: A data visualization library based on matplotlib, providing a high-level interface for drawing attractive and informative statistical graphics.
* bokeh: A library for creating interactive, web-based visualizations from Python, with capabilities for linking, brushing, and annotations.

*Machine Learning Models:*

**Scikit-learn**

Implemented classifiers like Logistic Regression , TruncatedSVD , MLPClassifier , SVC, K-Nearest Neighbors, Decision Tree, Random Forest, and Gradient Boosting.

*Model Evaluation:*

**Scikit-learn**

Employed accuracy\_score metric for quantitative model performance assessment.

*Project Management:*

**Google Colab**

Cloud-based Jupyter environment for collaborative development and resource-efficient execution.

**METHODOLGY**

The methodology used in the model can be summarized in the following points:

**Data Preprocessing**

* Loading the dataset into a pandas DataFrame
* Handling missing values and cleaning the data
* Performing string operations to clean and preprocess the ingredient lists

**Feature Engineering**

* Creating a list of unique ingredients from the dataset
* Generating a one-hot encoded matrix (ingred\_matrix) to represent the presence or absence of each ingredient in each product

**Dimensionality Reduction**

* Using TruncatedSVD (Truncated Singular Value Decomposition) to reduce the feature space to 150 dimensions
* Applying t-SNE (t-Distributed Stochastic Neighbor Embedding) to further reduce the dimensions to 2 for visualization purposes

**Visualization**

* Creating an interactive bokeh plot to visualize the product similarities based on the reduced 2D space
* Allowing users to filter the plot by product type using a dropdown menu

**Data Analysis**

* Analyzing the product type distribution using a pie chart
* Visualizing the price distribution by product type using a box plot

**Brand Identification**

* Implementing a brand identification algorithm to extract brand names from product names

**Machine Learning Model Training**

* Converting the ingredient lists into a format suitable for machine learning
* Splitting the data into training and testing sets
* Training and evaluating various machine learning models (Decision Tree, Random Forest, SVM, KNN, MLP, Gradient Boosting, Naive Bayes, Logistic Regression) for product type classification

**Recommendation Engine**

* Implementing a content-based recommendation engine using cosine similarity
* Recommending similar products based on ingredient similarity, considering product type and brand
* Providing examples of recommendations for different input products.

The model covers a wide range of tasks, including data preprocessing, feature engineering, dimensionality reduction, visualization, data analysis, brand identification, machine learning model training and evaluation, and the development of a content-based recommendation engine using cosine similarity.

**RESULTS AND DISCUSSION**

The result of the model is to develop a content-based recommendation engine for skincare products. The recommendation engine should take the name of a skincare product as input and return several similar products based on the product's ingredients. The project implements an algorithm to extract brand names from the product names, which is useful for the recommendation engine to avoid recommending products from the same brand as the input product. The model trains and evaluates various machine learning models (Decision Tree, Random Forest, SVM, KNN, MLP, Gradient Boosting, Naive Bayes, Logistic Regression) for product type classification using the ingredient data. It defines a recommender function that takes the name of a product as input and recommends similar products based on their ingredient similarity. By implementing the recommender function and incorporating the necessary data preprocessing, dimensionality reduction, and brand identification steps, the code achieves the expected result of a content-based recommendation engine for skincare products based on ingredient similarity.

**CONCLUSION**

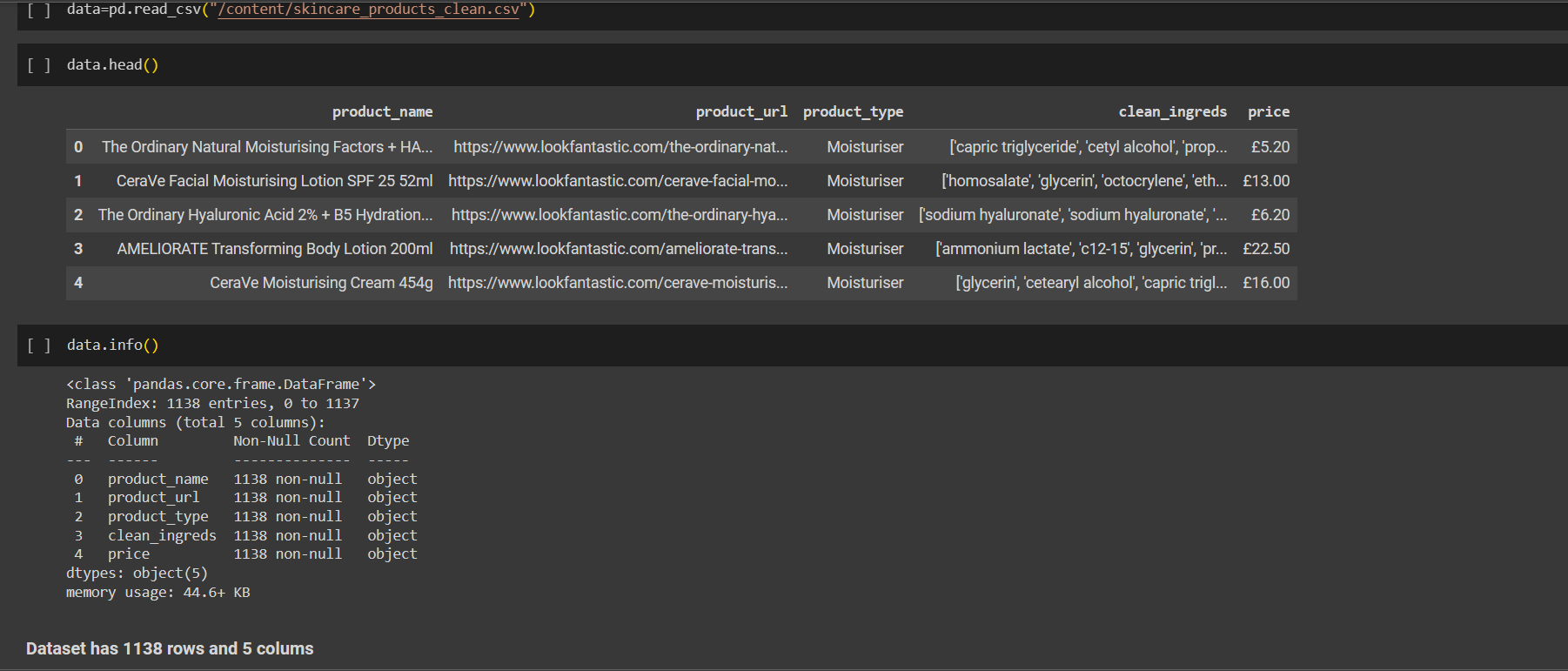
Our project holds potential to successfully implement a content-based recommendation engine for skincare products, leveraging ingredient data and cosine similarity to suggest relevant and similar products. Through dimensionality reduction, interactive visualizations, and careful data preprocessing, the project demonstrates the power of machine learning techniques in providing personalized recommendations tailored to individual needs. With further enhancements and integration into a user-friendly interface, this recommendation engine can significantly improve the shopping experience for skincare enthusiasts, enabling them to discover new products aligned with their preferences.

**SCREENSHOTS**

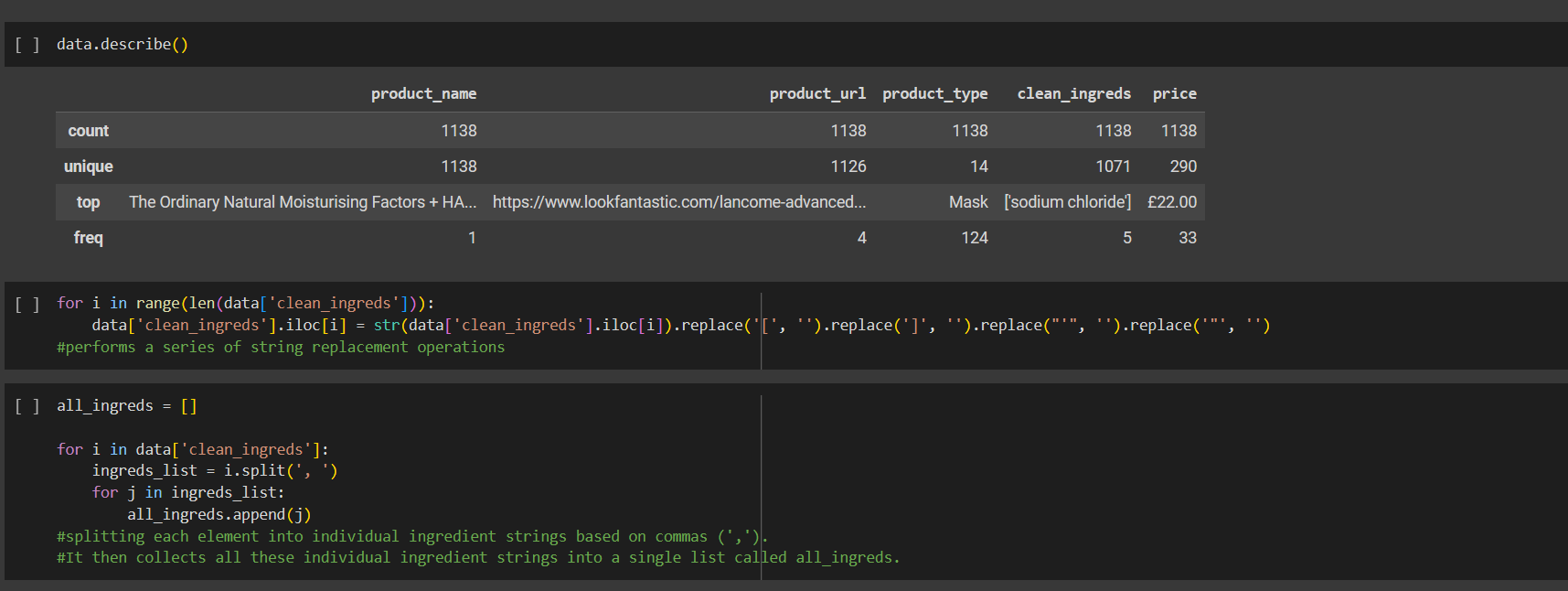
***->*** *Importing libraries*

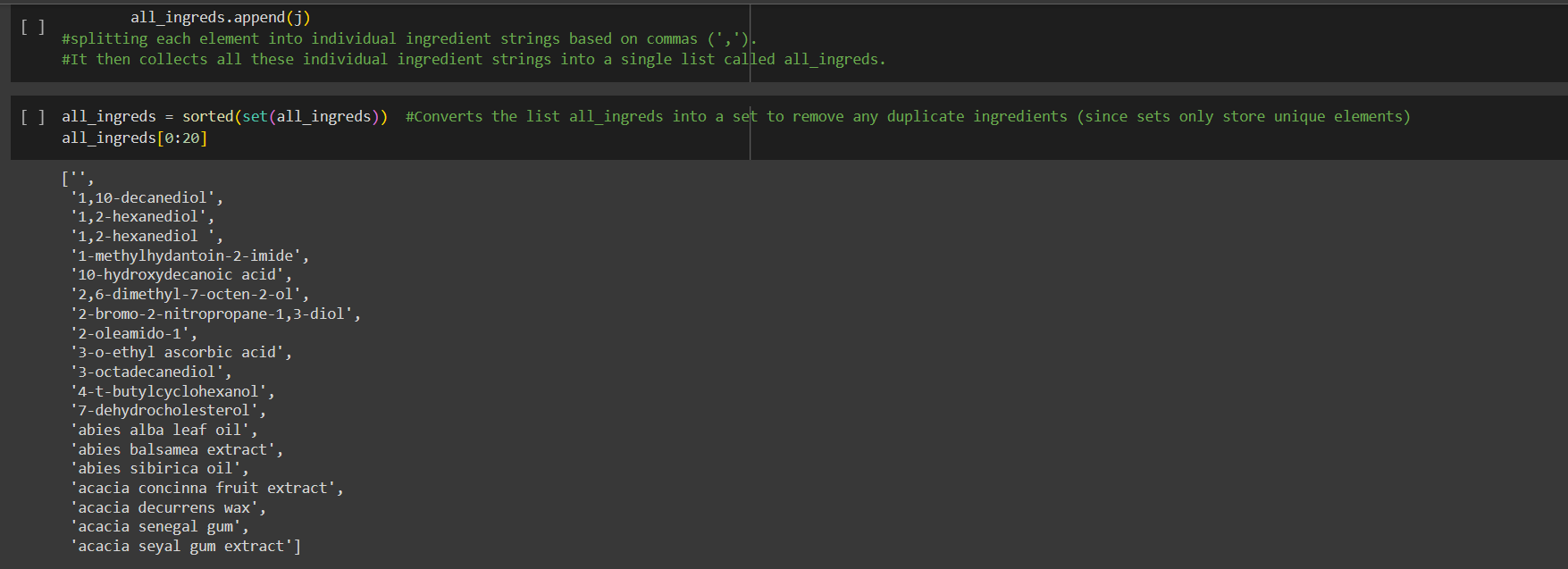
**

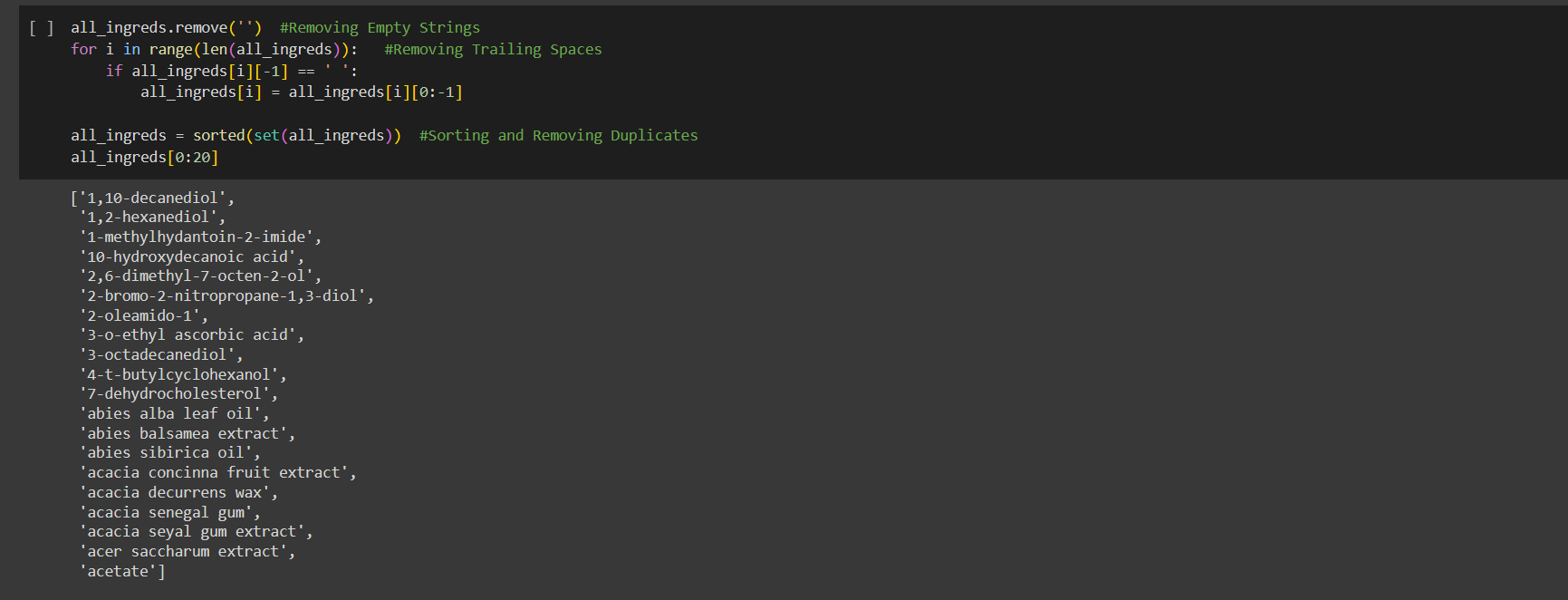
***->*** *Uploading dataset*

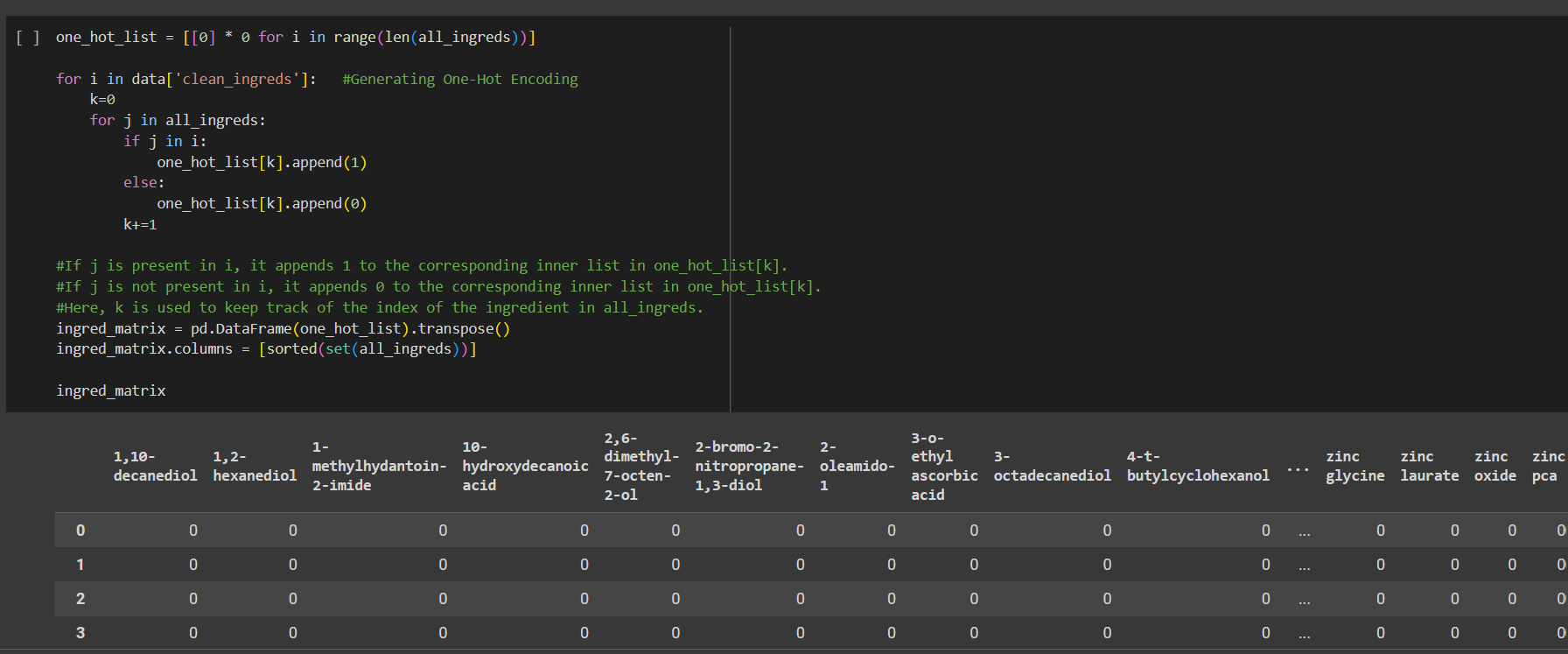
**

***->*** *Data Preprocessing*

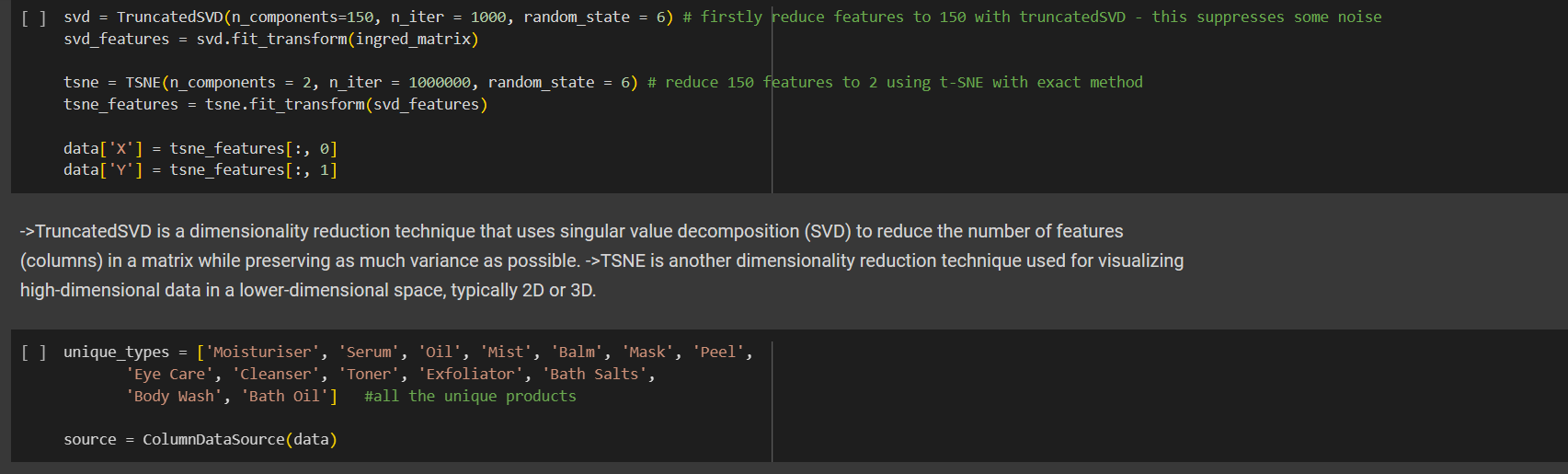
**

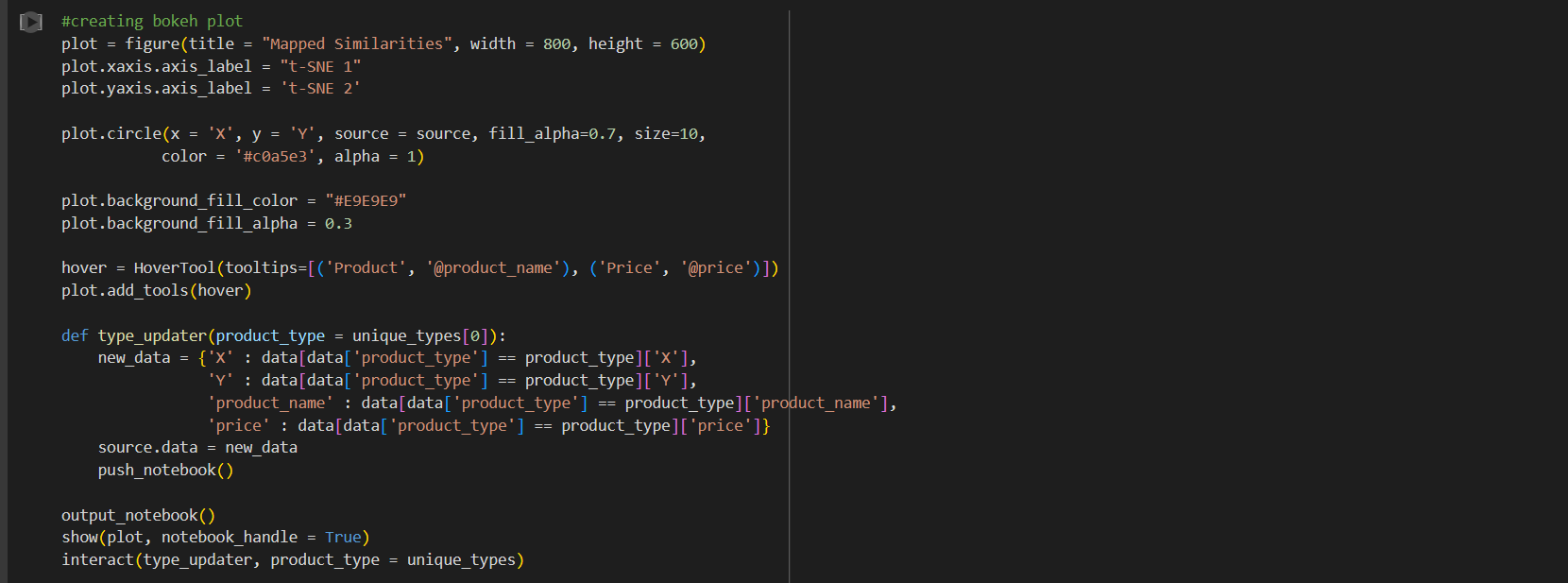
**

**

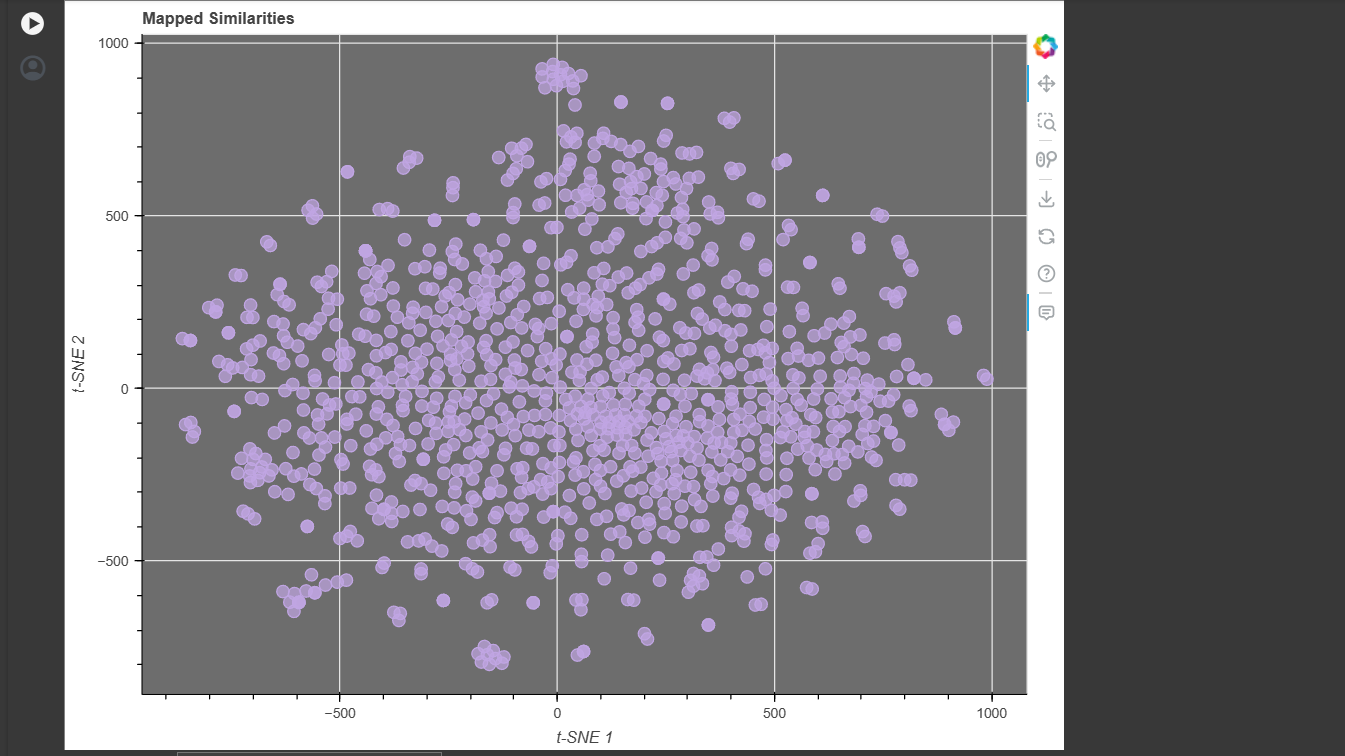
**

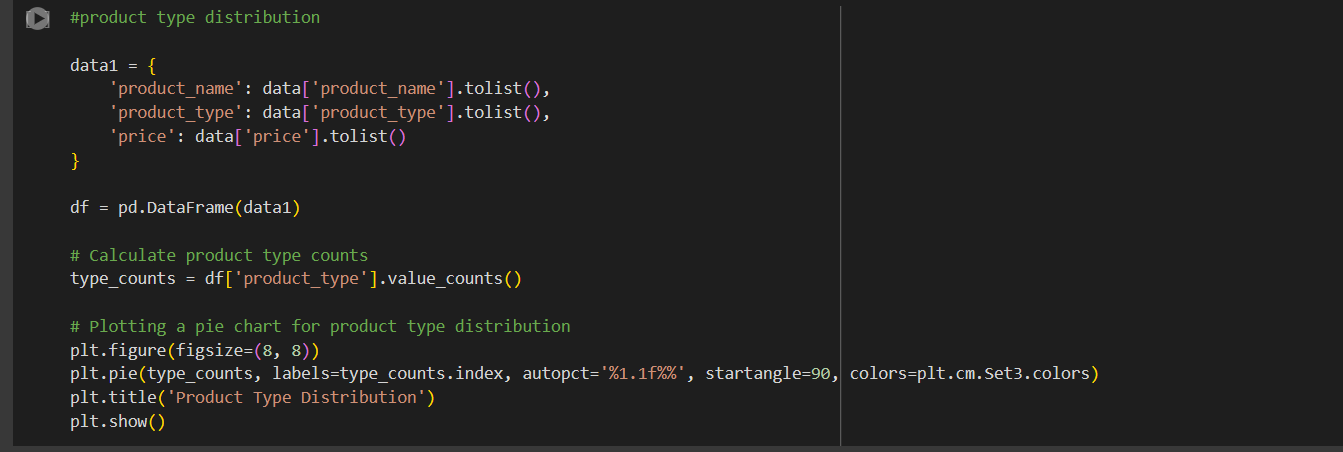
*-> Dimensionality Reduction*

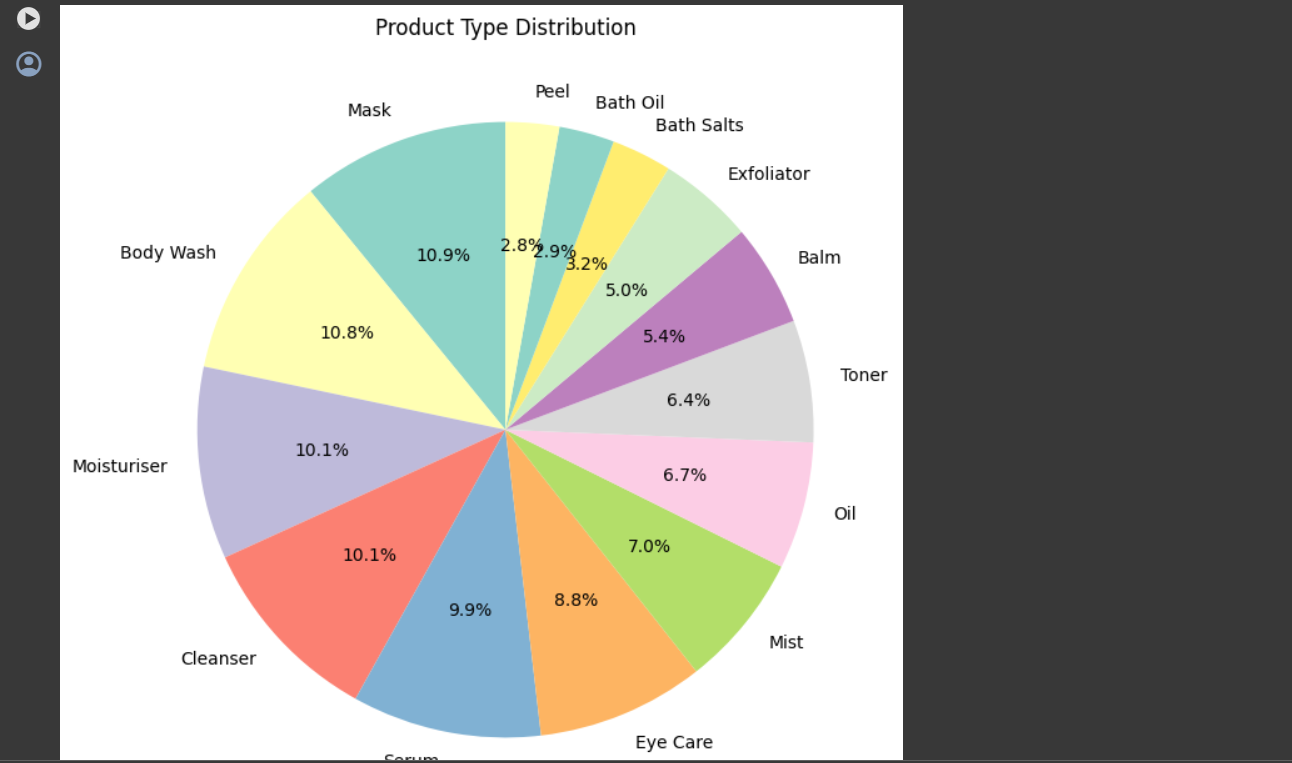
**

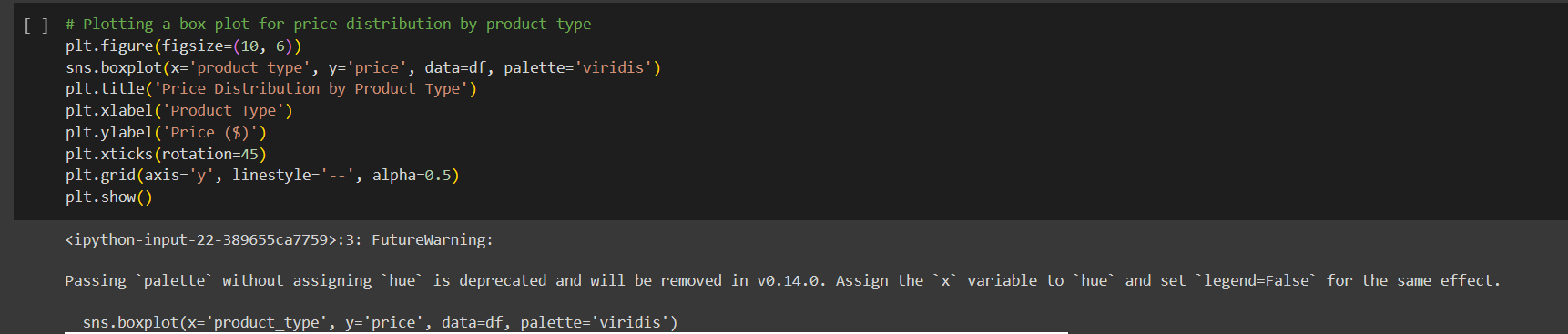
**

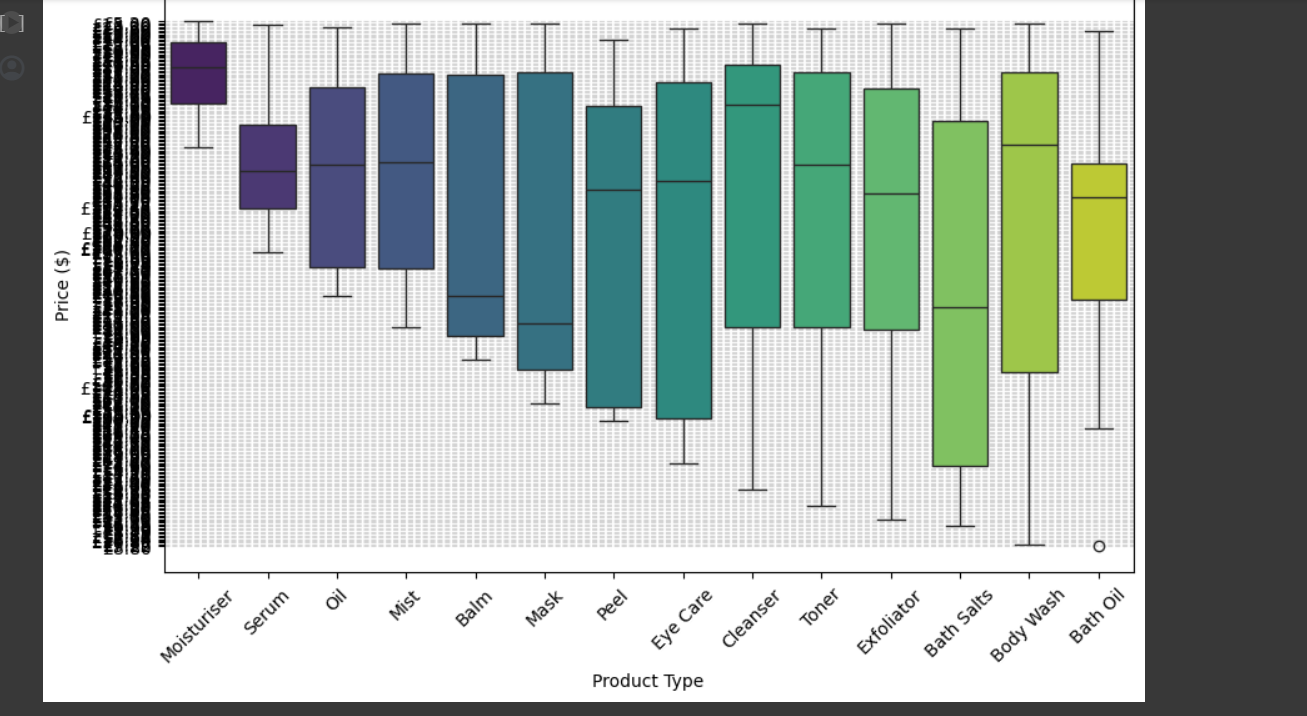
*-> Data Visualization*

**

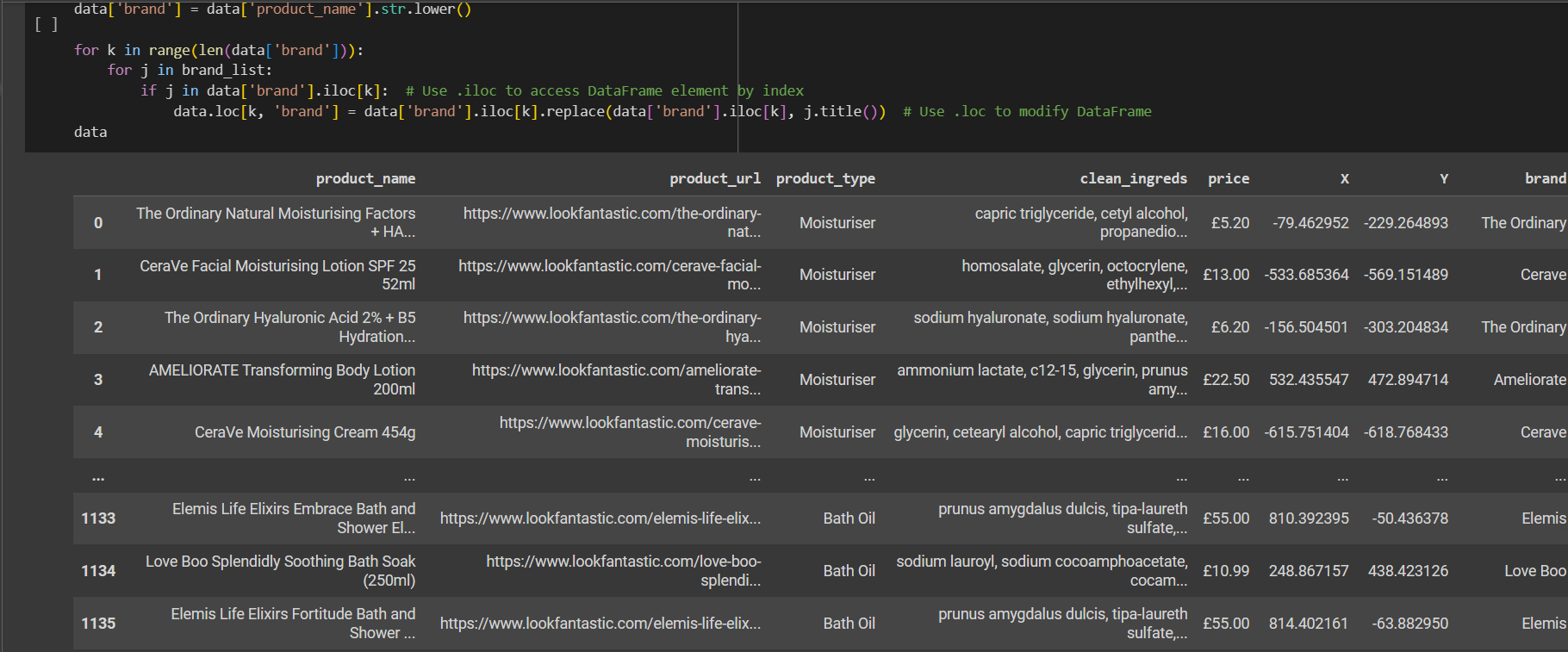
**

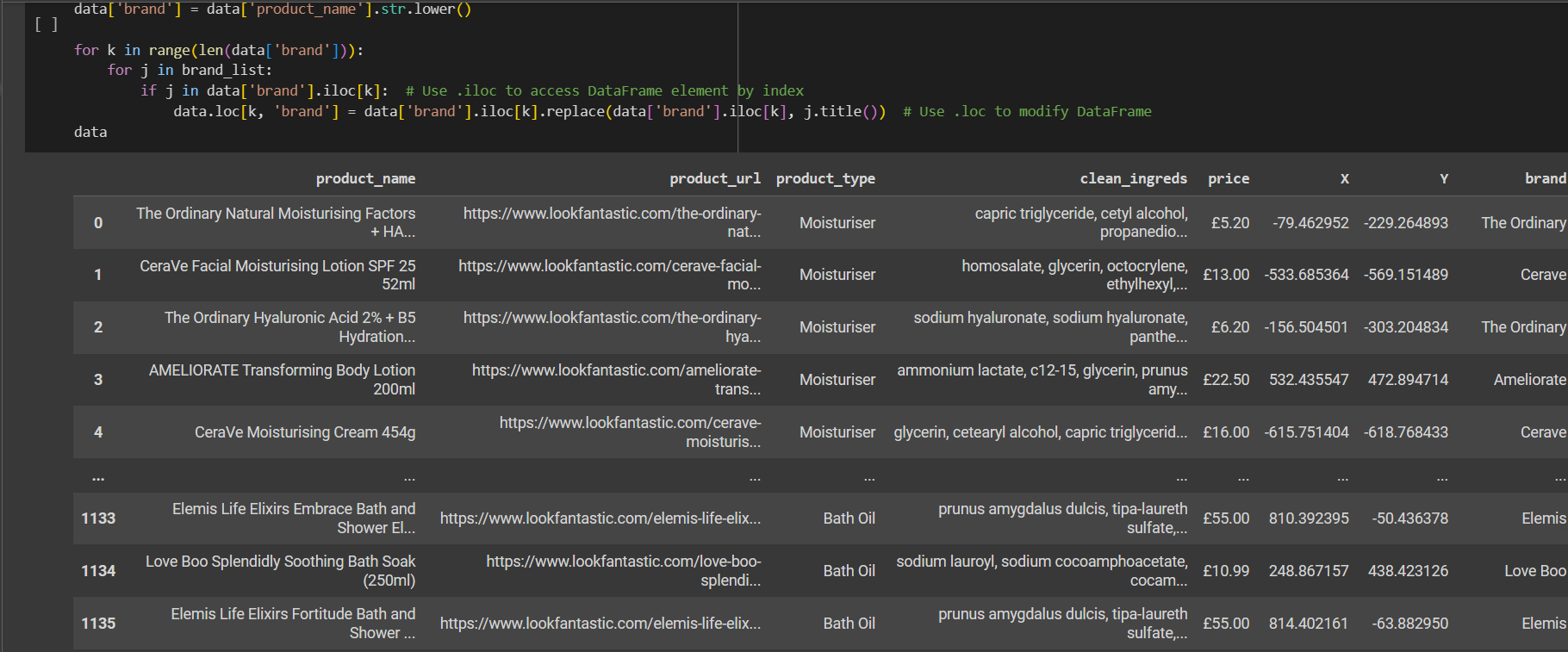
**

**

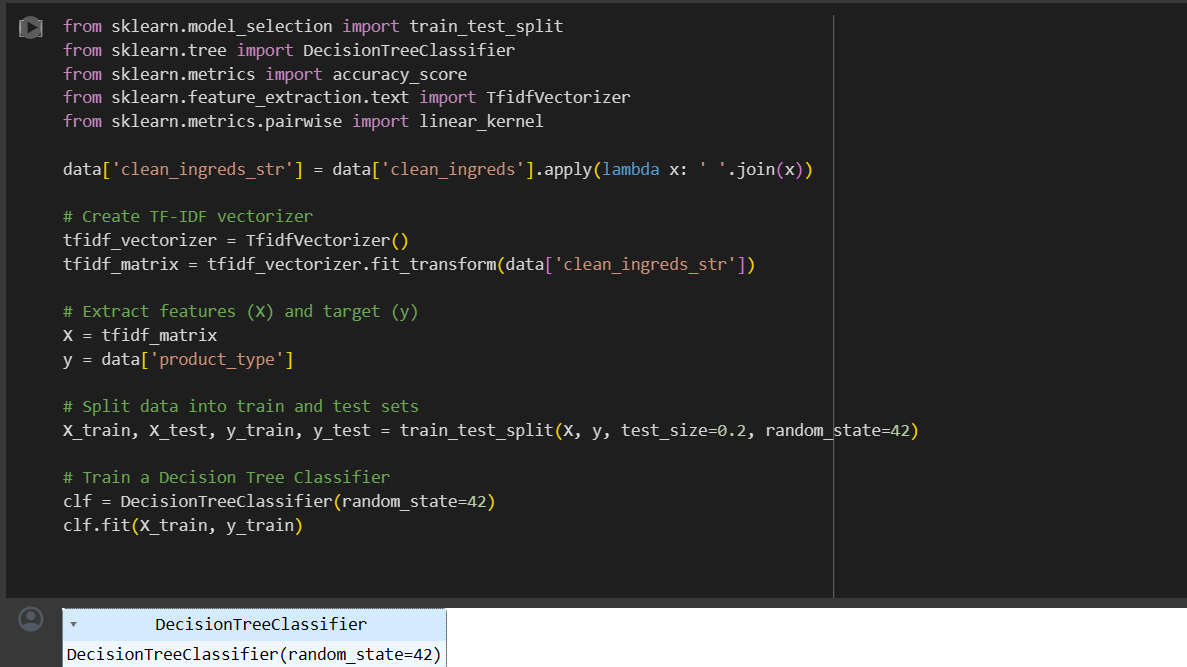
**

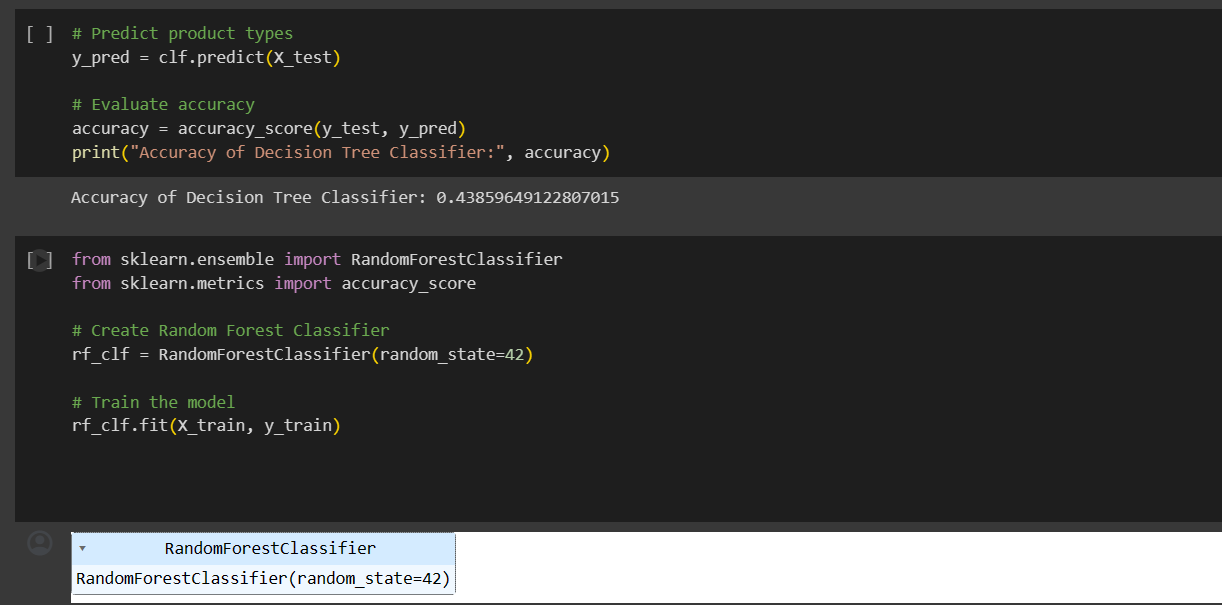
*->Brand Identification and listing*

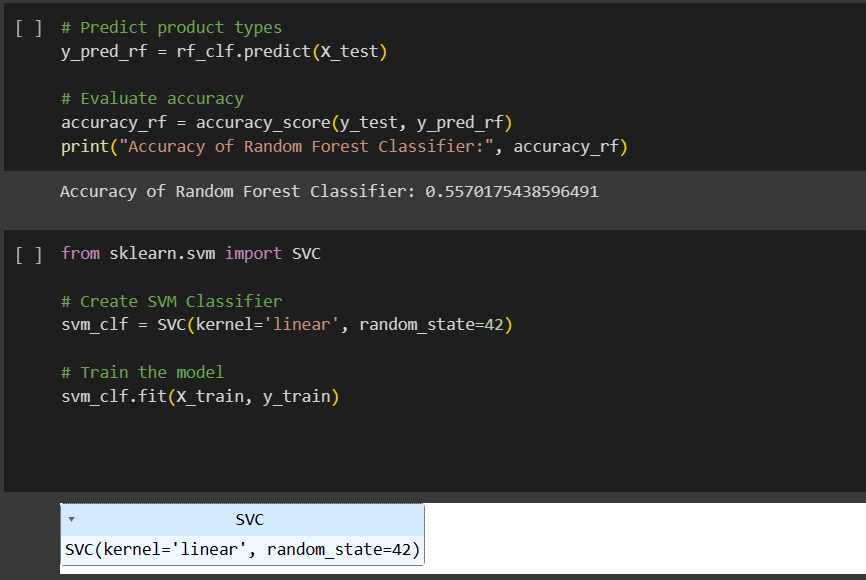
**

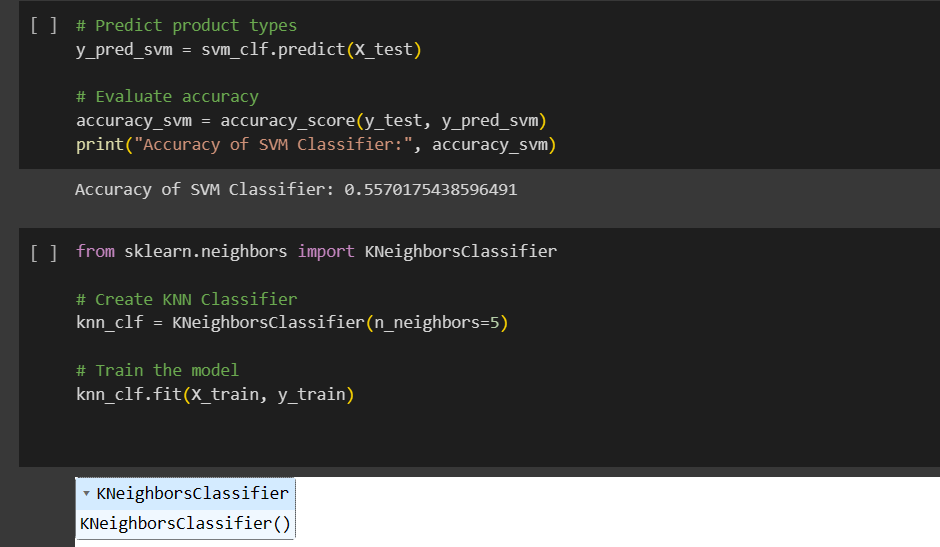
**

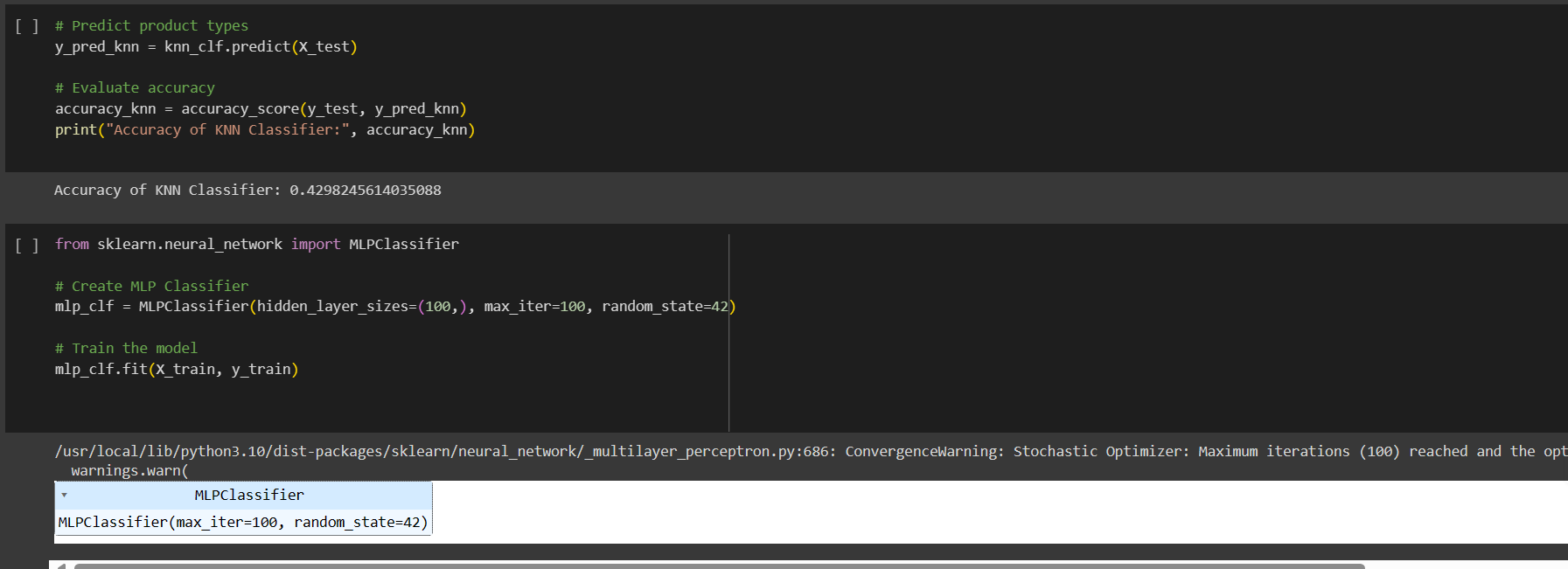
*->Implementing Algorithms*

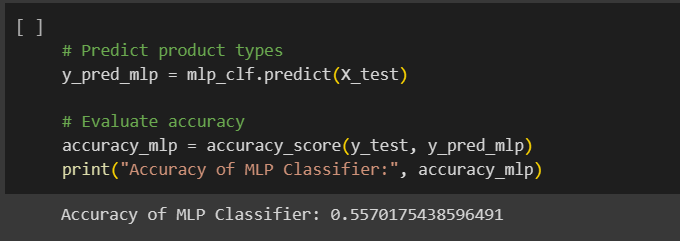
**

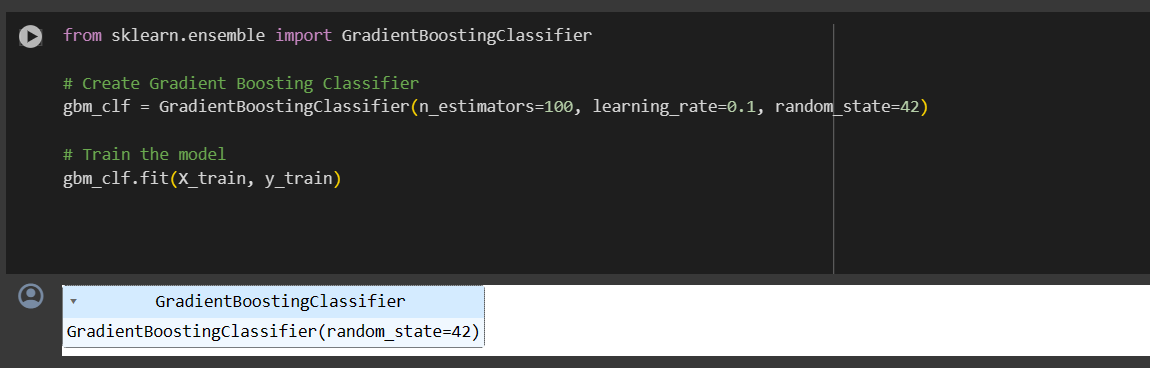
**

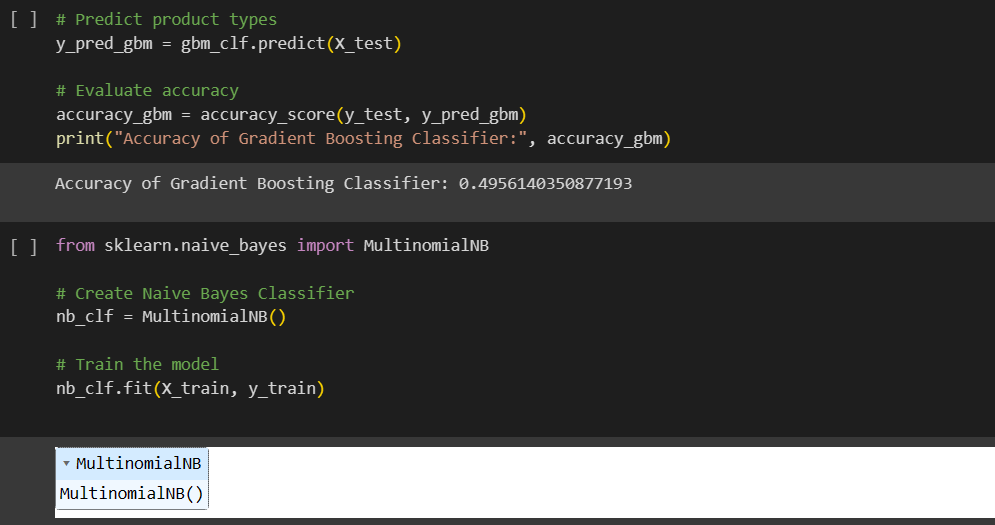
**

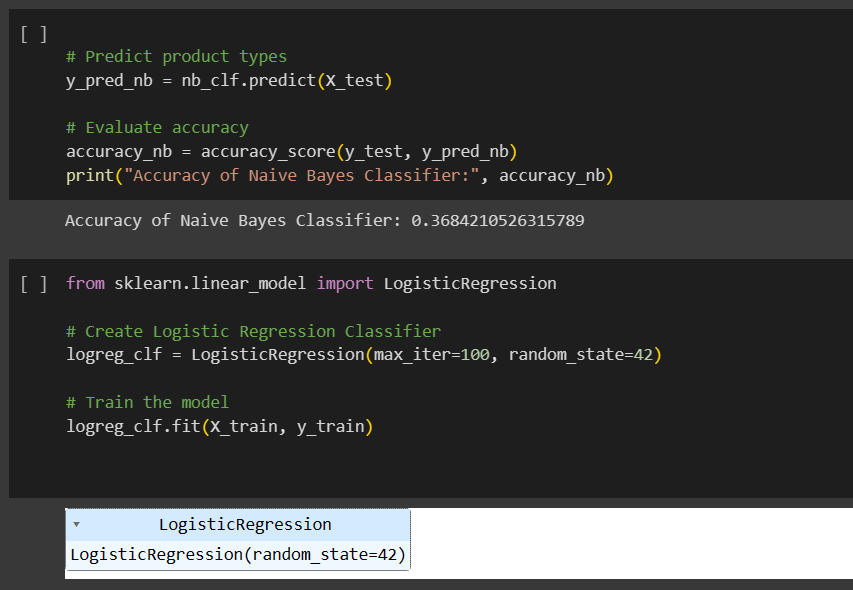
**

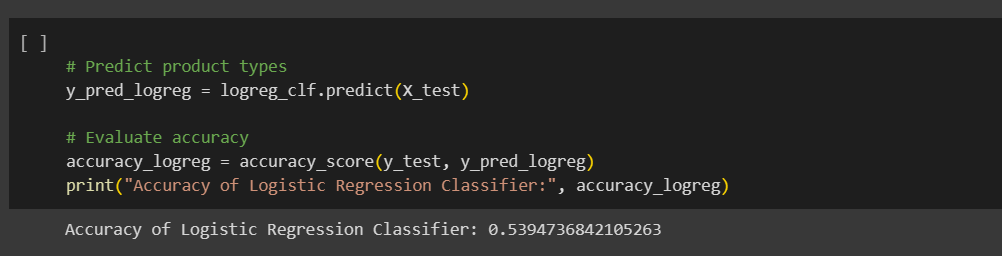
**

**

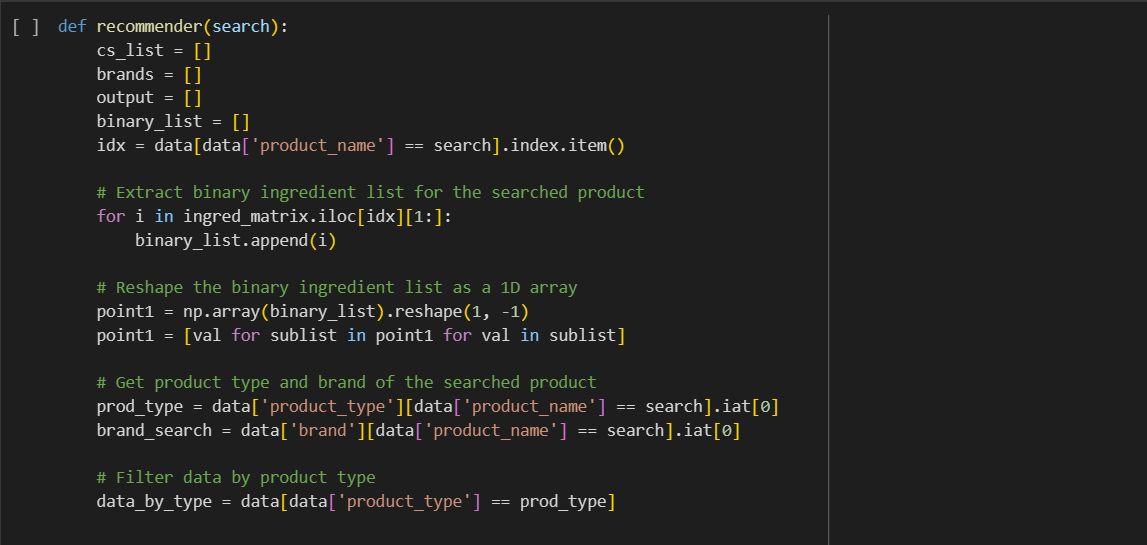
**

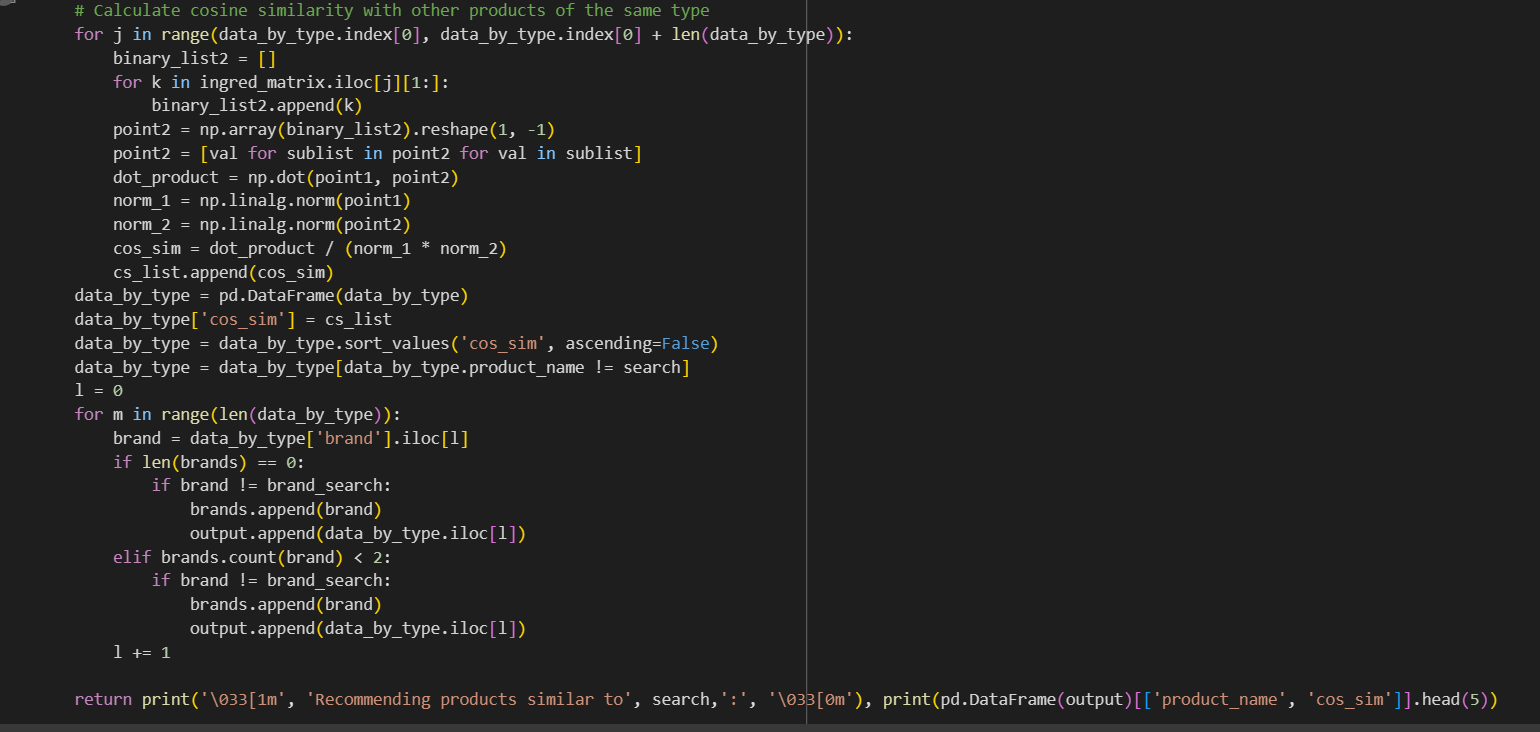
**

**

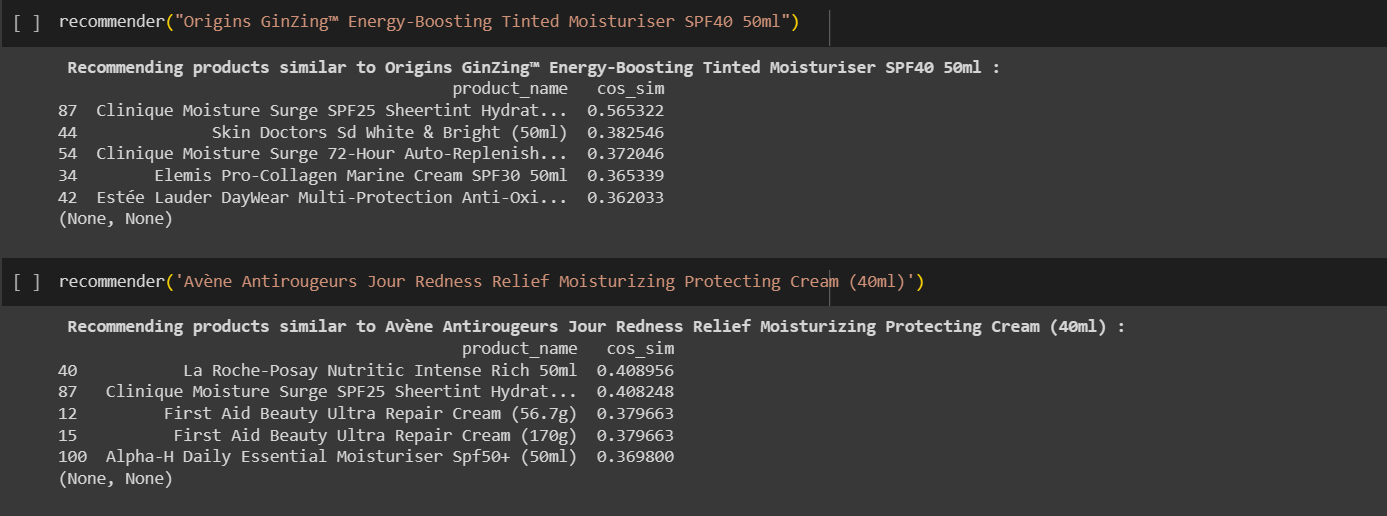
**

*->Recommendation Function*

**

**

*->Results*

**

**

