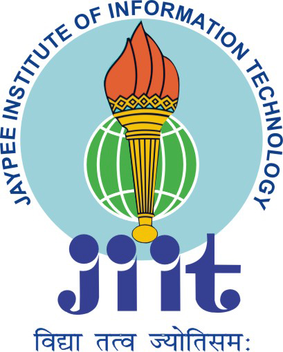
Jaypee Institute of Information Technology

Department of Computer Science & Engineering and Information Technology



GreenThread: Sustainable Fashion Tracker

#### Submitted to:

Prof. Pushp S. Mathur  
Dr. Amanpreet Kaur

#### Submitted By:

[ALOKIK GARG 22103302](mailto:22103302@mail.jiit.ac.in)

SHIVAM ARYAN 22103298

DAVID SHARMA 22103299

YASH GAUTAM 22103277

**COURSE**: Open-Source Software Lab (15B17CI575)

ODD 2024

# Introduction

## Problem Statement

The fashion industry is notorious for its significant environmental impact, contributing heavily to carbon emissions, excessive water usage, and waste production. This project, ***GreenThread: Sustainable Fashion Tracker***, addresses the need for transparency and accessible information regarding sustainable fashion brands. The current market lacks sufficient tools to empower consumers to make informed, eco-friendly choices. *GreenThread* seeks to fill this gap by providing users with accurate data on eco-conscious brands and guiding them towards more responsible consumption patterns.

## SDG Alignment

*GreenThread* directly aligns with **United Nations Sustainable Development Goal (SDG) 12**, which promotes responsible consumption and production. Specifically, SDG 12 emphasizes the importance of sustainable practices across industries, encouraging businesses and consumers to adopt environmentally-friendly habits. By tracking and promoting sustainable brands, *GreenThread* contributes to reducing the fashion industry's negative environmental impact, fostering responsible production, and encouraging consumers to shift toward sustainable alternatives.

## Objectives

The primary goal of GreenThread is to empower consumers by providing a platform that offers transparency into the environmental practices of fashion brands. The expected outcomes of the project include:

**Promoting Awareness**: Educating users about the sustainability of various fashion brands, leading to more informed purchasing decisions.

**Providing Data-Driven Insights**: Through brand recommendations and data visualizations, users can easily compare brands based on metrics like carbon footprint, water usage, and sustainability scores.

**Encouraging Responsible Consumption**: By guiding consumers toward eco-friendly brands, GreenThread contributes to reducing waste and promoting sustainable fashion practices.

**Collaboration and Community Engagement**: As an open-source project, GreenThread aims to foster collaboration within the community, encouraging developers and sustainability advocates to contribute to the platform’s continuous improvement.

By achieving these goals, GreenThread aspires to shift consumer behavior towards more eco-conscious choices, supporting the broader mission of reducing the environmental damage caused by the fashion industry.

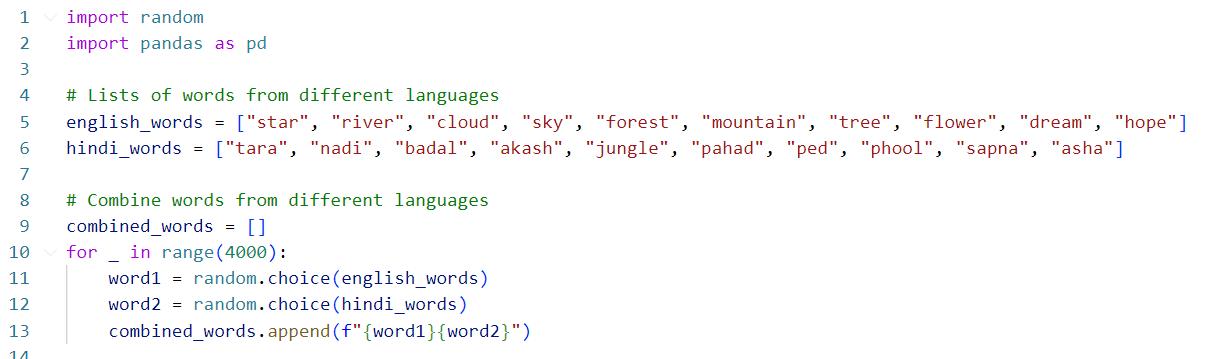
# Modelling/Implementation Details

## Sustainable Brands Database

* The sustainable brand database is the core feature of the GreenThread platform, containing detailed information on eco-friendly fashion brands. It consists of key attributes that reflect each brand's environmental footprint and sustainability efforts. The dataset contains 4,999 entries, each corresponding to a unique brand, with 25 columns describing various aspects of sustainability, including:
* **Brand Name and ID**: Unique identifiers for each brand.
* **Sustainability Metrics**: Carbon footprint (in metric tons), water usage (in litres), waste production (in kilograms), and sustainability score.
* **Certifications**: Whether the brand holds certifications for eco-friendly practices.
* **Recycling Programs**: Indicates if the brand offers recycling initiatives.
* **Market Trends and Material Types**: Describes the types of materials used and market trends for each brand.
* The database was pre-processed and enriched using Python libraries such as **pandas** for data manipulation, **scikit-learn** for scaling and encoding, and **MongoDB** for storage.

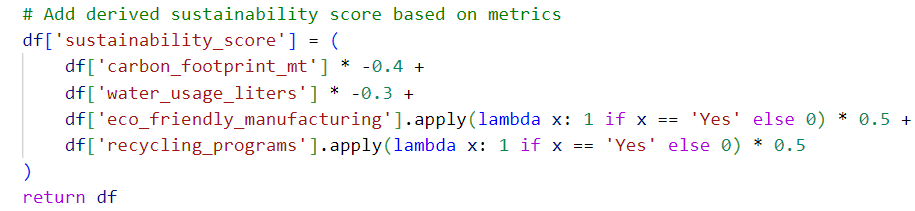
### Brand name Generation

* To replace generic brand names like "Brand\_1", "Brand\_2", etc., we implemented a Python-based name generator using word combinations from ten different languages, enhancing cultural diversity and creativity in the branding process.



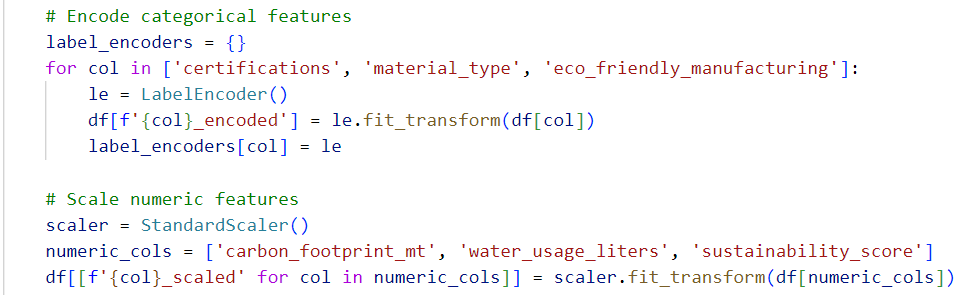
### Pre-processing

* The dataset underwent several preprocessing steps using Python's **pandas** library to clean and prepare the data for analysis:
* **Key Steps in Preprocessing:**
  + **Handling Missing Values**: Missing values were replaced with defaults where applicable.
  + For example, missing values in columns like "Certifications" were replaced with "None."
  + **Cleaning Text Fields**: Column names were normalized (e.g., removing spaces and converting to lowercase), and text fields were standardized.
  + **Handling Outliers**: Extreme outliers were removed to ensure data consistency.
  + **Calculating Sustainability Scores**: A derived column, "sustainability\_score", was created based on metrics like carbon footprint, water usage, and recycling programs.



### Feature Extraction

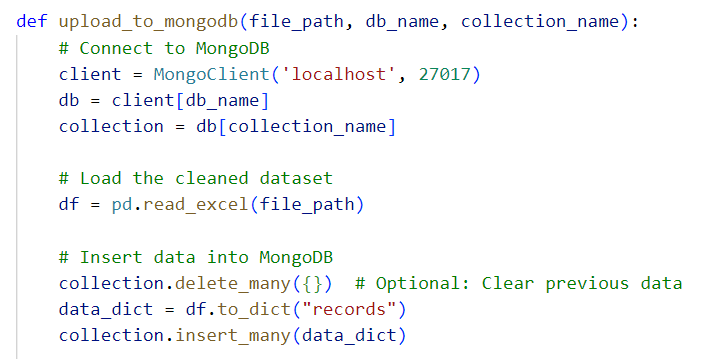
* We utilized the **sklearn.preprocessing** module to encode categorical variables and scale numeric features, ensuring the dataset was prepared for machine learning algorithms. This included encoding categorical features like material type, certifications, and market trends, as well as normalizing continuous variables such as carbon footprint and water usage.



* This process ensured that the dataset was well-prepared for downstream tasks such as machine learning or data visualization.

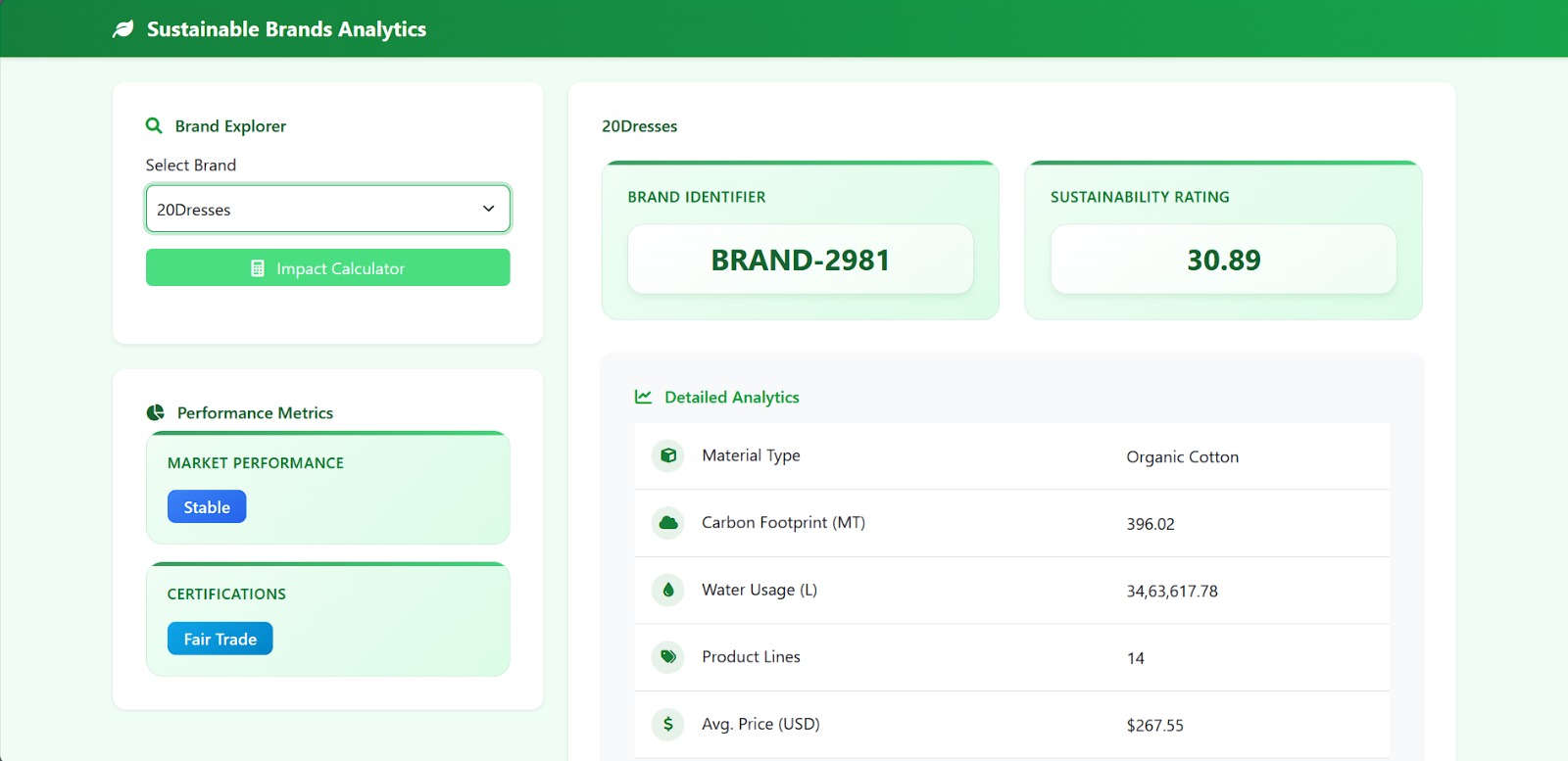
## Upload to MongoDB

* To store the processed data and make it accessible for the web application, we uploaded the dataset to **MongoDB** using the pymongo library. This step involved clearing any previous records in the database and inserting the new, preprocessed data.
* The MongoDB database now holds all 4,999 records from the sustainable brand dataset, making them accessible for further analysis and integration into the web application.



## Individual Brand Pages

* Each page displays key data such as:
  + **Sustainability Score**: A calculated score reflecting the brand's overall environmental impact.
  + **Carbon Footprint, Water Usage, Waste Production**: Detailed environmental metrics that give users a clear understanding of the brand’s sustainability efforts.
  + **Product Lines and Market Trend**: Insights into the brand’s product range and market position.
  + **Certifications**: Information about any eco-certifications the brand holds, highlighting their commitment to sustainable practices.



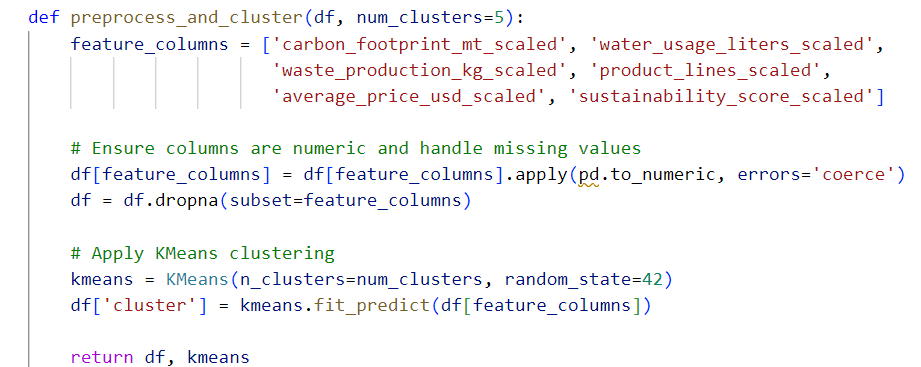
## Brand Recommendation Using scikit-learn and scipy

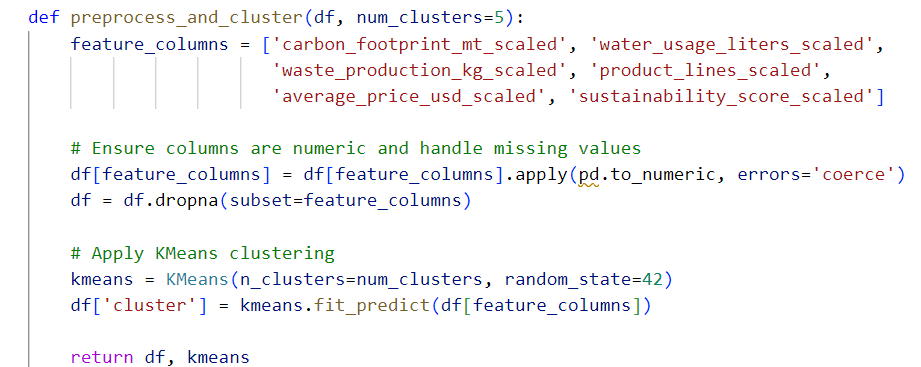
* The *GreenThread* project includes a robust recommendation system that identifies similar sustainable brands based on key features such as carbon footprint, water usage, and sustainability score. The process uses **KMeans clustering** from the scikit-learn library and distance metrics from scipy to suggest brands that are similar to a selected one.

### Overview of the Process

#### Data Loading: The system loads the preprocessed dataset containing scaled values of key sustainability metrics.

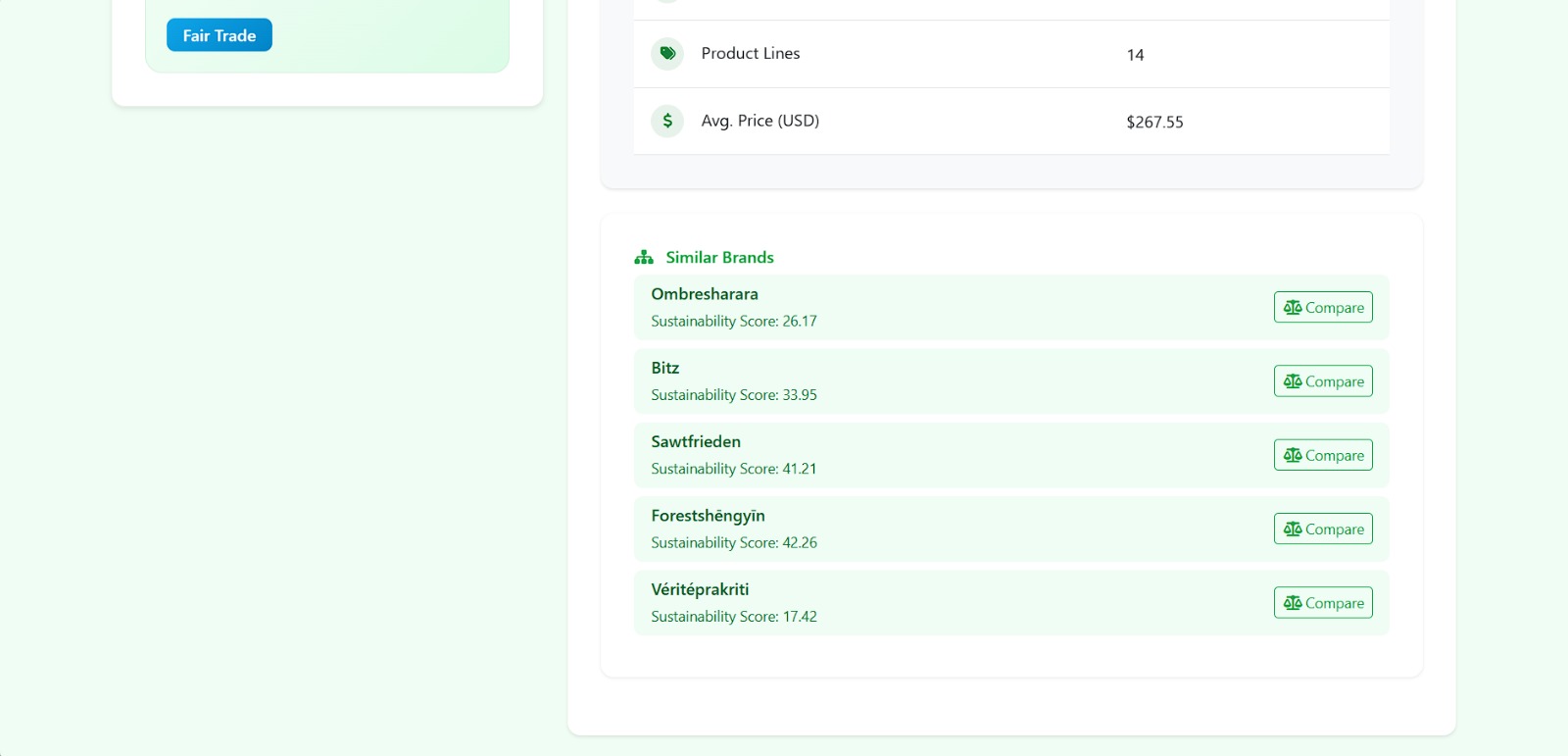
#### Preprocessing and Clustering: Brands are clustered based on their scaled features using the KMeans algorithm. This clustering groups brands with similar sustainability profiles.





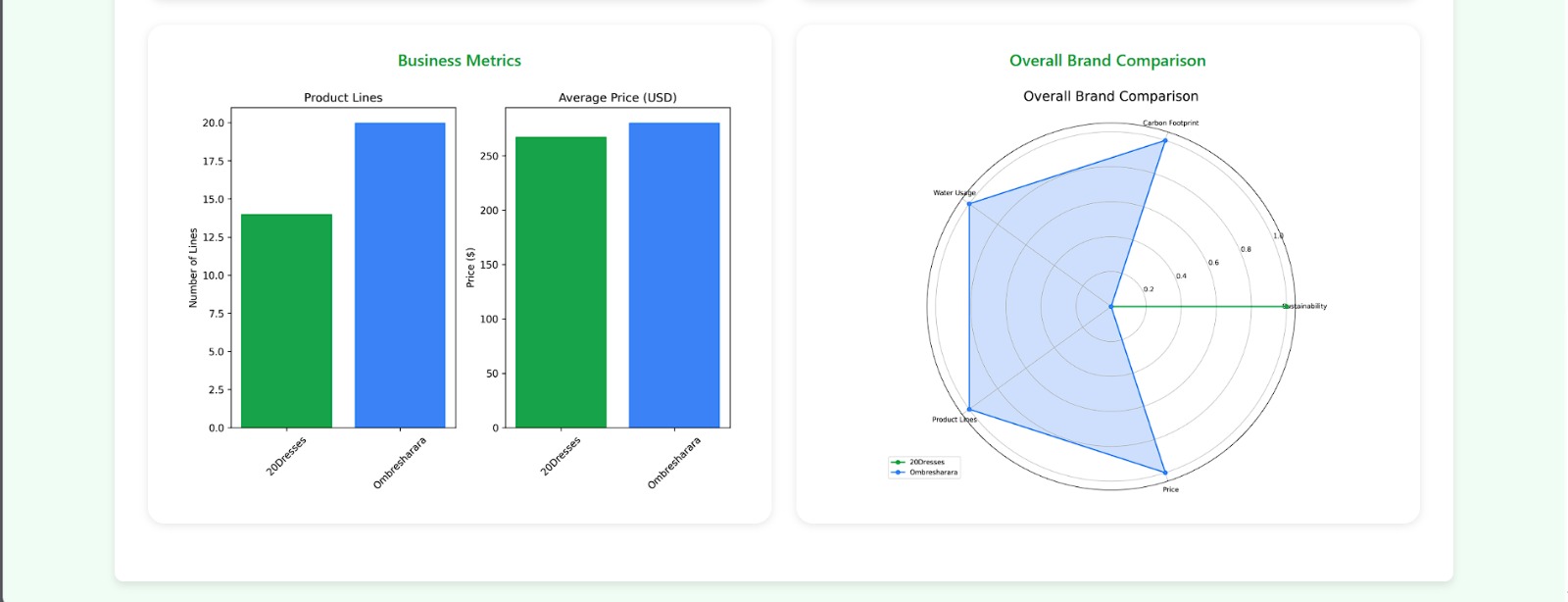
#### Similarity Calculation: For a given brand, the system identifies brands within the same cluster by calculating pairwise Euclidean distances between feature vectors. The closest brands are then recommended.





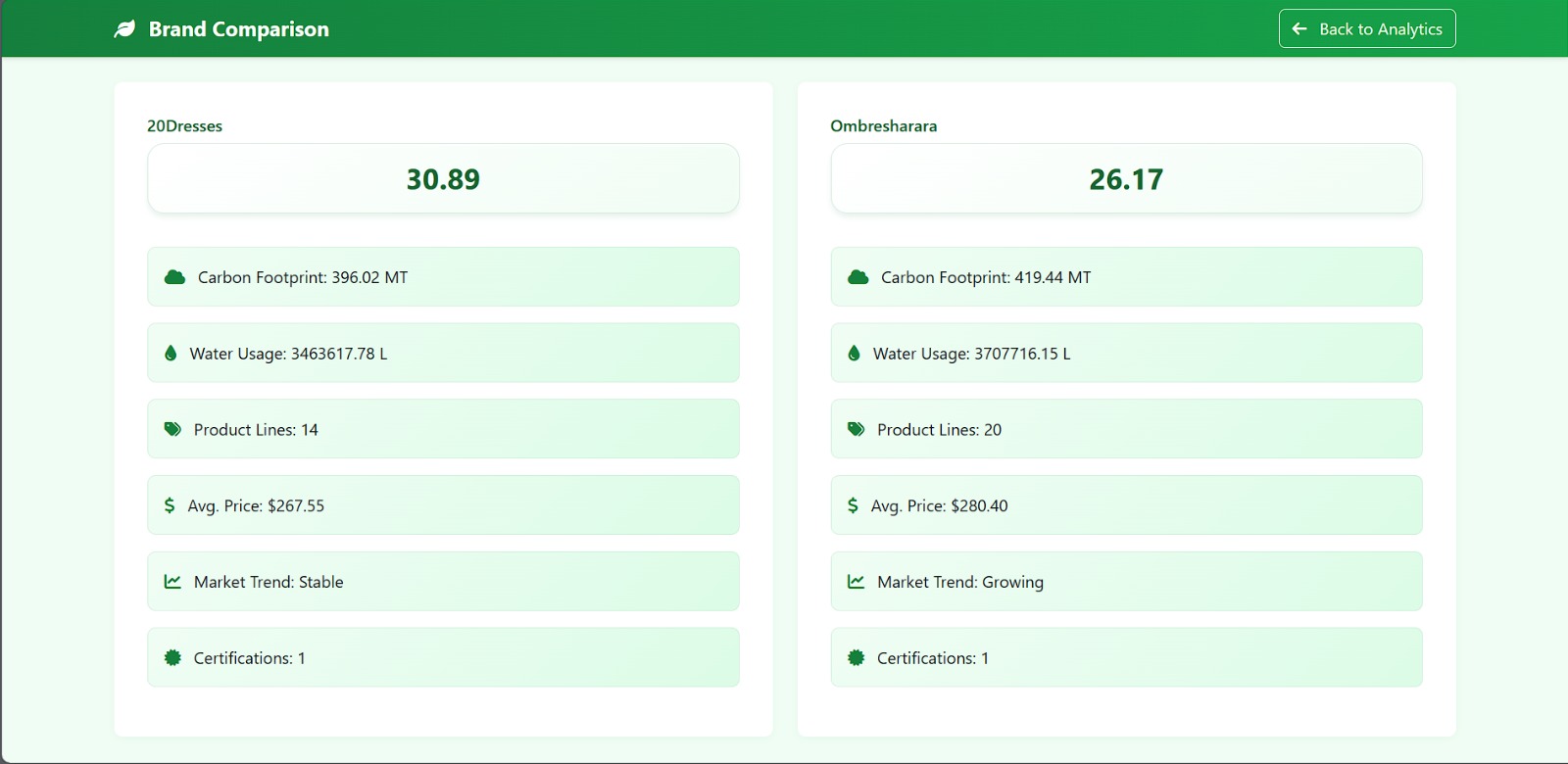
#### Visualization: The system generates a bar chart to visually represent how similar the recommended brands are to the selected brand.





#### Brand Comparison Cards:

* Each brand's metrics are displayed in separate Bootstrap cards. These cards include:
* Sustainability Score: A visually highlighted metric.
* Carbon Footprint, Water Usage, Product Lines, Average Price, Market Trend, and Certifications.

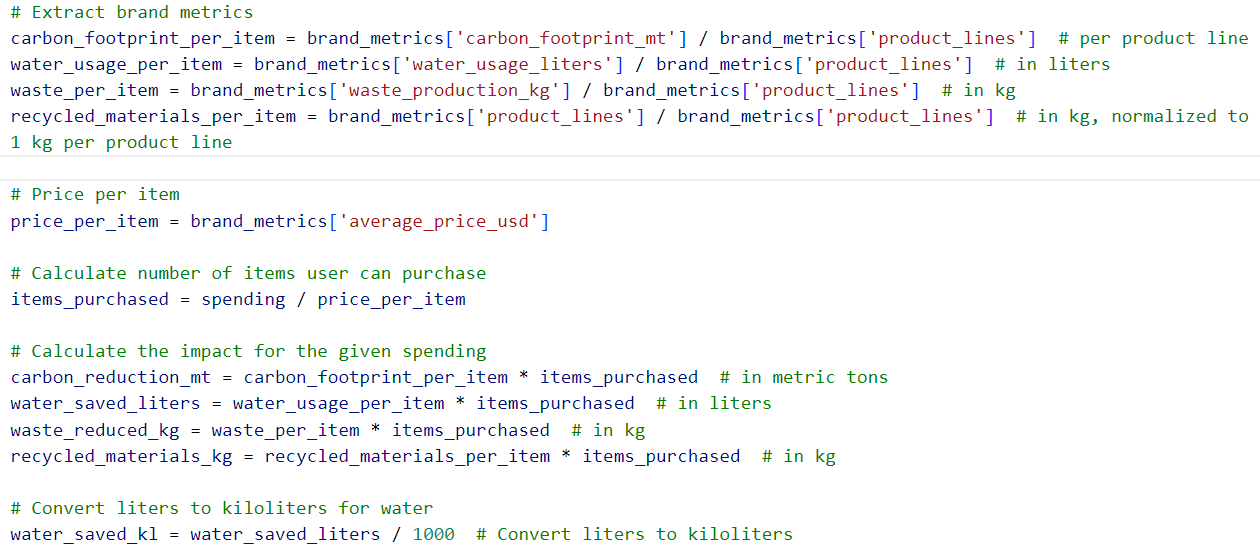


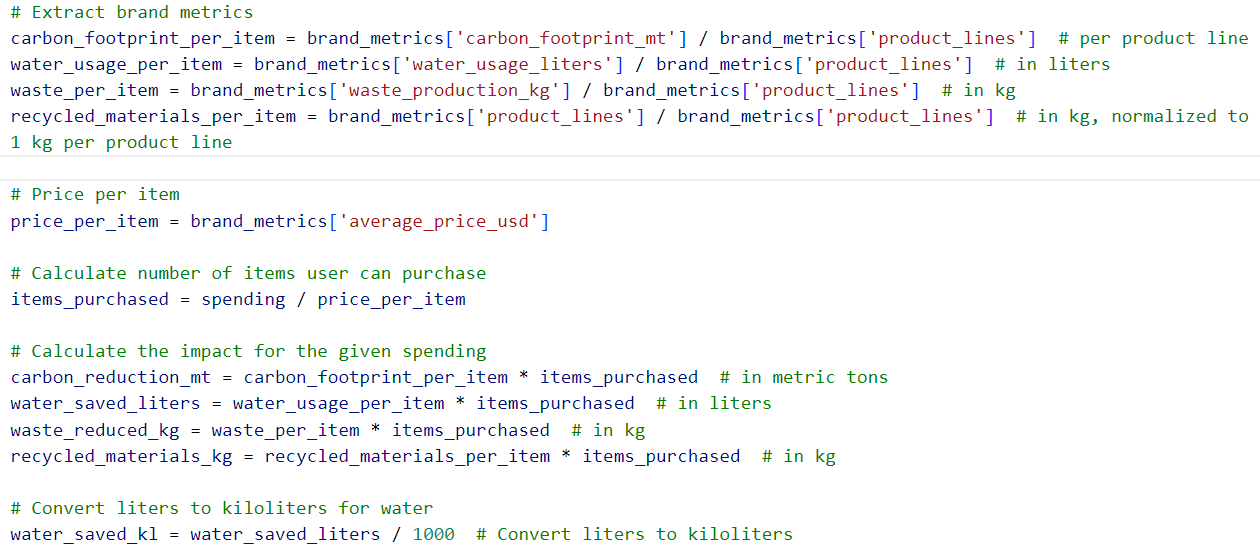
## Impact Calculator

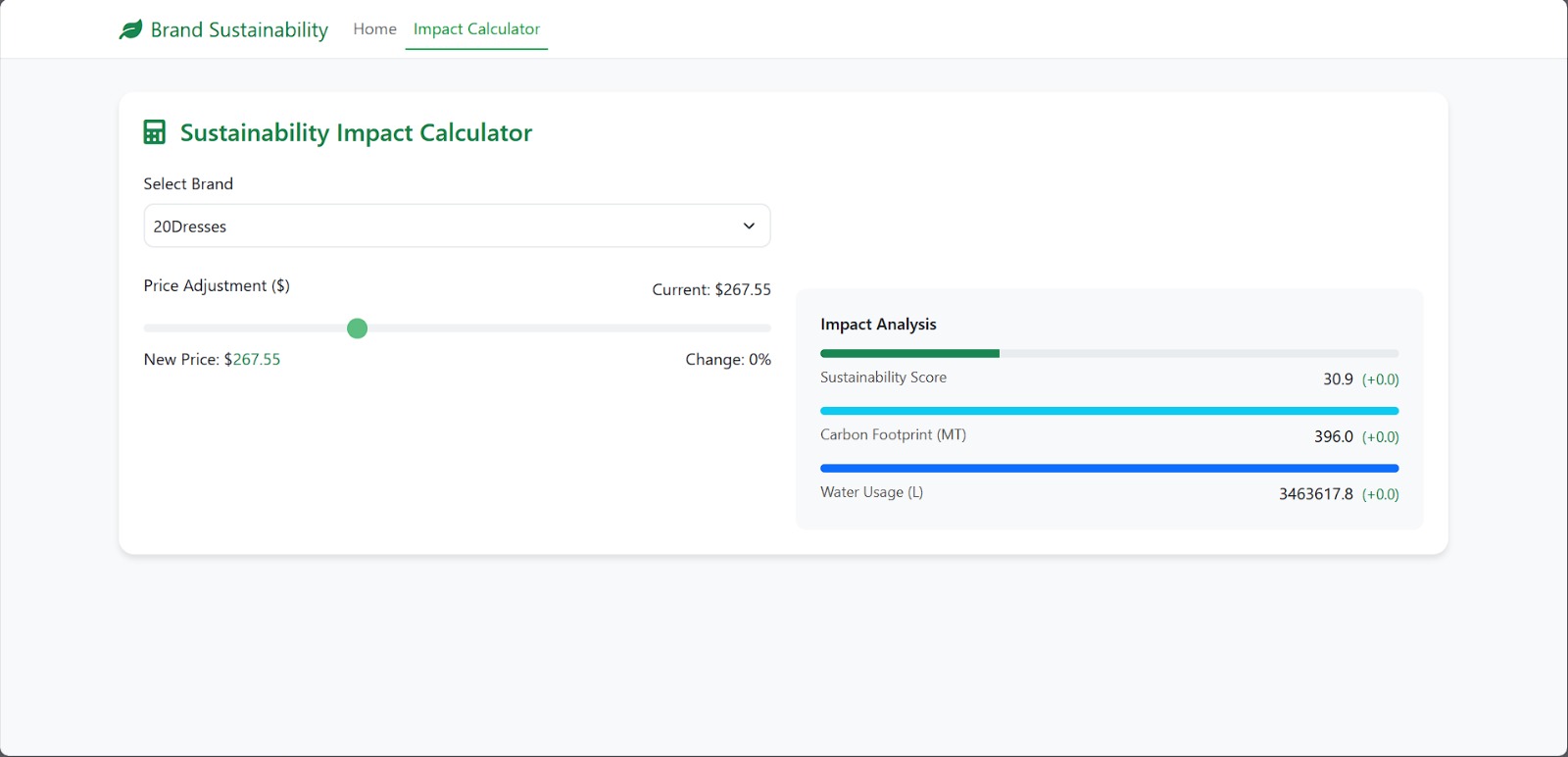
* The *Impact Calculator* allows users to estimate and compare the environmental impact of different brands based on factors like **carbon footprint**, **water usage**, and **waste production**. This tool provides users with insights into how their brand choices affect the environment.

### Core Functionality:

* **User Input**: The calculator takes user input (e.g., the number of products purchased, average price, or specific sustainability metrics) for one or more brands.
* **Impact Calculation**: Based on the selected brand(s) and user inputs, the tool calculates the environmental impact in terms of:
  + **Carbon Footprint**: Total emissions generated by purchasing from the brand.
  + **Water Usage**: The volume of water consumed.
  + **Waste Production**: Amount of waste produced.
* **Comparison**: The results are compared across multiple brands, allowing users to make informed, environmentally-conscious decisions.







# Tools and Technologies Used

## PYTHON AND its Libraries

* Python was the primary programming language used throughout the project, with its rich ecosystem of libraries supporting various functionalities, from data processing to web development.

### Pandas - Handling missing values, scaling features, and preparing the dataset for analysis.

### Scikit-learn - Machine learning, specifically KMeans clustering for recommendations.

* Grouping brands based on sustainability metrics.

### Matplotlib and seaborn - Visualizing sustainability scores, trends, and comparisons across brands.

### Scipy - Calculating Euclidean distances to recommend brands within clusters.

### Numpy - Managing feature vectors for distance calculations.

### Pymongo - Database management, connecting to MongoDB.

### Flask- Web framework for building the application.

## Front-End Technologies – HTML, CSS, JavaScript

# Conclusions and Future Work

## Conclusion

* The GreenThread project has made significant strides in promoting sustainability within the fashion industry, directly contributing to **Sustainable Development Goal (SDG) 12: Responsible Consumption and Production.** By enabling consumers to access detailed, data-driven insights into the sustainability practices of individual fashion brands, the platform empowers them to make more eco-conscious purchasing decisions.
* The project’s achievements include:
  + **Brand Recommendation System**: Gbrands based on similar sustainability metrics, making it easier for consumers to identify environmentally-friendly alternatives.
  + **Impact Calculator**: Allowing users to calculate the environmental impact of purchasing decisions, based on factors like carbon footprint, water usage, and waste production.
  + **Individual Brand Pages**: This ensures that users have the information needed to evaluate a brand's overall environmental impact.
  + **Data Visualizations**: Allow users to easily compare sustainability trends across brands, helping them make more informed choices.
* By utilizing **Flask, pandas, scikit-learn, pymongo**, and other Python libraries, the platform combines data processing, machine learning, and a responsive web interface to create a seamless user experience.
* **Additionally, by following open-source principles, GreenThread invites contributions from the development community, fostering continuous growth and improvement.**

## Future Works

While the current platform is functional, there are several areas for future development:

1. **Enhanced Data Collection**:
   * **API Integrations**: Integrating fashion industry APIs could provide even more accurate, real-time data on sustainable fashion brands.
2. **Advanced Analytics and Machine Learning**:
   * **Personalized Brand Recommendations**: Enhancing the recommendation system with advanced algorithms, such as **collaborative filtering** or **deep learning**, could help tailor brand suggestions based on user preferences.
   * **Predictive Analytics**: Implementing predictive models to forecast future sustainability trends based on current brand practices.
3. **Expanded Impact Metrics**:
   * Including additional sustainability metrics such as **social responsibility** or **labor practices**, along with **supply chain transparency**, would provide users with a more holistic view of a brand’s environmental and social impact.
   * **Lifecycle Assessments**: Integrating lifecycle assessments would give users a complete picture of a product's environmental footprint from production to disposal.
4. **Individual Brand Pages Enhancements**:
   * Include **user reviews** and **ratings**, adding an interactive and dynamic aspect to brand evaluation.
   * **Detailed Comparisons**: Allow users to compare multiple brands side-by-side on their individual pages, visualizing metrics such as carbon footprint, water usage, and sustainability score in a more detailed manner.
   * **Integration of Interactive Charts**: Providing more interactive, real-time charts that update based on user input.
5. **Open-Source Collaboration**:
   * Encouraging more **open-source contributions** from developers, data scientists, and sustainability advocates can drive further innovation and improvements in the platform..
6. **Scalability and Deployment**:
   * **Cloud Deployment**
   * **Mobile App**