

# **SkinWiz: A Personalized Skincare Recommender**

## **Group Members (A1 Group 2):**

Afifa Iqbal : afifai@andrew.cmu.edu

Lydia Abebe : labebe@andrew.cmu.edu

Rishika Dwaraghanath: rthiruve@andrew.cmu.edu

Xinyue Meng: xinyuem@andrew.cmu.edu

## **1. Project Overview**

This project implements a skincare recommendation system that provides product recommendations based on user input and weather data.

## **2. Prerequisites**

### **Required Python Version:**

- Python 3.8 (tested on Python 3.8.5)

### **Additional Python Modules:**

In addition to the standard Anaconda distribution, you need to manually install the following Python modules:

- requests – For handling HTTP requests.
- pandas – For data manipulation and analysis.
- arff – For handling .arff files
- beautifulsoup4 – For parsing HTML and XML documents during web scraping.
- lxml – For fast and efficient XML and HTML parsing.
- selenium – For automating web browser interaction (used in web scraping).
- tqdm – For displaying progress bars for loops.
- streamlit – For building web applications in Python.
- streamlit\_extras – A set of extensions and enhancements for Streamlit.
- scikit-learn – For machine learning algorithms and data processing.
- geopandas – For working with geospatial data to plot graphs in the project.
- contextily – For adding background maps to the geopandas visualizations.

You can manually install them by running the command:

```
Unset  
pip install <module_name>
```

They are also all listed in the **requirements.txt** file. You can install them all in one go using the command:

```
Unset  
pip install -r requirements.txt
```

## 3. Installing the Project

### Step 1: Download the project

Download the project files to your local machine.

### Step 2: Environment Setup

We recommend using a Python virtual environment. To create a new environment:

```
Unset  
python3 -m venv <env_name>  
source <env_name>/bin/activate
```

After activating the environment, install the required modules as mentioned in the prerequisites section.

**Note:** You do not require an API key to get weather data from [Open-Meteo](#).

## 4. Running the Project

### Step 1: Modify the config.py file

Make sure the **config.py** file is set up correctly with the following details:

1. Set the `download_dir` variable:

Set this variable to the full path of your system's default Downloads directory. This is where any downloaded files during the execution of the project will be saved.

Example:

```
Python
download_dir = '/Users/yourusername/Downloads' # For macOS/Linux
# OR
download_dir = 'C:\\\\Users\\\\yourusername\\\\Downloads' # For Windows
```

## 2. Modify the User-Agent field:

Update the User-Agent field in the HEADERS variable to reflect a valid User-Agent for your browser. This is crucial for scraping data.

You can find your current User-Agent by searching for "what is my User-Agent" in your browser, then copying and pasting it into the HEADERS variable.

Example:

```
Python
HEADERS = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36',
    'Accept-Language': 'en-US,en;q=0.9'
}
```

## Step 2: Running the Main Script

Run the project using Streamlit by executing the following command in your terminal:

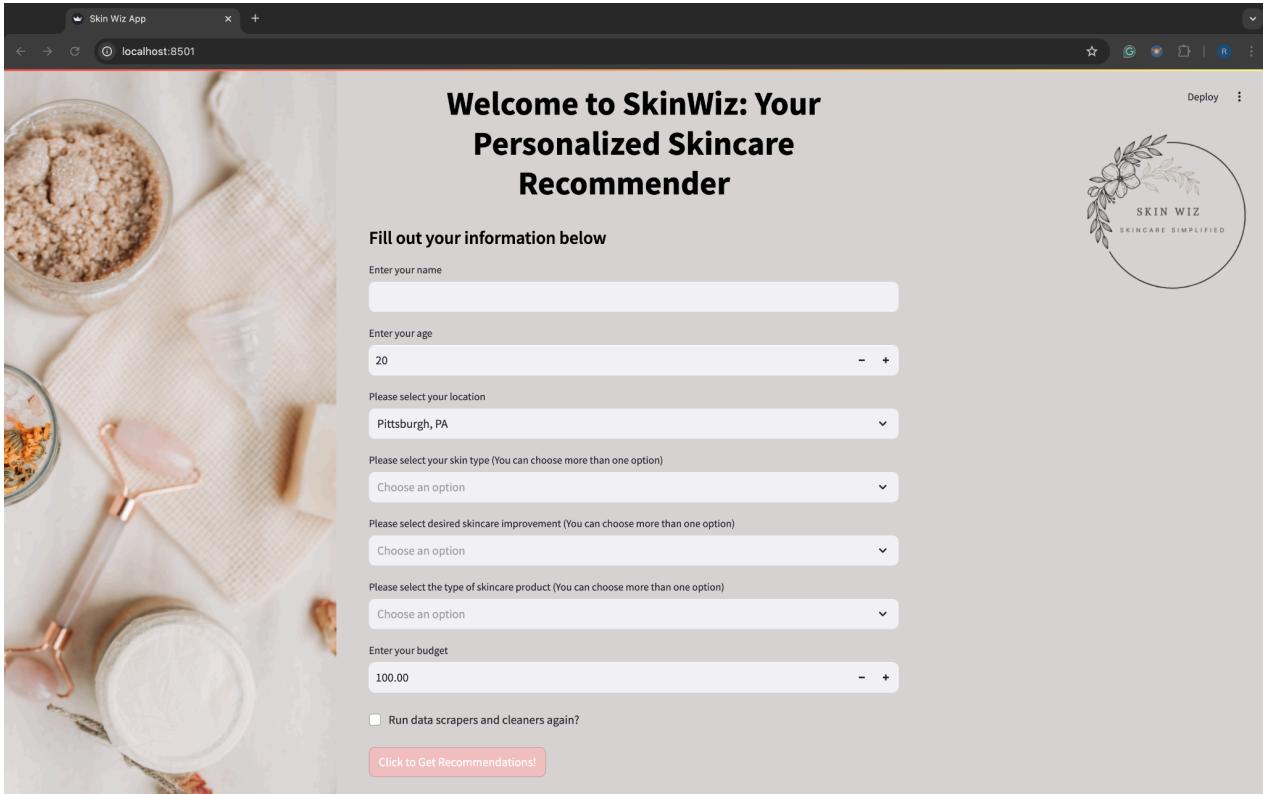
```
Python
streamlit run main.py
```

## Step 3: Accessing the Web Interface

Once the script is running, Streamlit will provide you with a local URL (typically <http://localhost:8501>) in the terminal. Open this URL in your web browser to interact with the application.

## 5. How to Use & Screenshots

### Web Interface:



The screenshot shows a Streamlit web application titled "Skin Wiz App" running on "localhost:8501". The main heading is "Welcome to SkinWiz: Your Personalized Skincare Recommender". Below it, a sub-instruction reads "Fill out your information below". The form consists of several input fields: "Enter your name" (text input), "Enter your age" (number input with a range from 20 to 60), "Please select your location" (dropdown menu showing "Pittsburgh, PA"), "Please select your skin type (You can choose more than one option)" (dropdown menu showing "Choose an option"), "Please select desired skincare improvement (You can choose more than one option)" (dropdown menu showing "Choose an option"), "Please select the type of skincare product (You can choose more than one option)" (dropdown menu showing "Choose an option"), "Enter your budget" (number input with a range from 100.00 to 1000.00), and a checkbox for "Run data scrapers and cleaners again?". At the bottom is a red button labeled "Click to Get Recommendations!". The background features a photograph of various skincare products like a pink roller massager, a jar of cream, and a bowl of granules.

Once you start the project via Streamlit, you will see a user interface with fields to input personal information like skin type, age, and budget. To get skincare recommendations:

1. Fill in your information.
2. Click the '**Click to Get Recommendations**' button.

### Scraping New Data:

If you would like to scrape fresh data from the internet, follow these steps:

1. Check the '**Run data scrapers and cleaners again?**' checkbox.
2. Click the '**Click to Get Recommendations**' button.

### IMPORTANT NOTE:

- When you check the '**Run data scrapers and cleaners again?**' option, the system scrapes the datasets using Selenium, which will be rerun for 2 out of our 3 datasets. For our 3rd dataset, we scrape Amazon data from ~1000 URLs, which takes approximately 4

hours to complete. This can be too cumbersome to wait for completion for a user running the application.

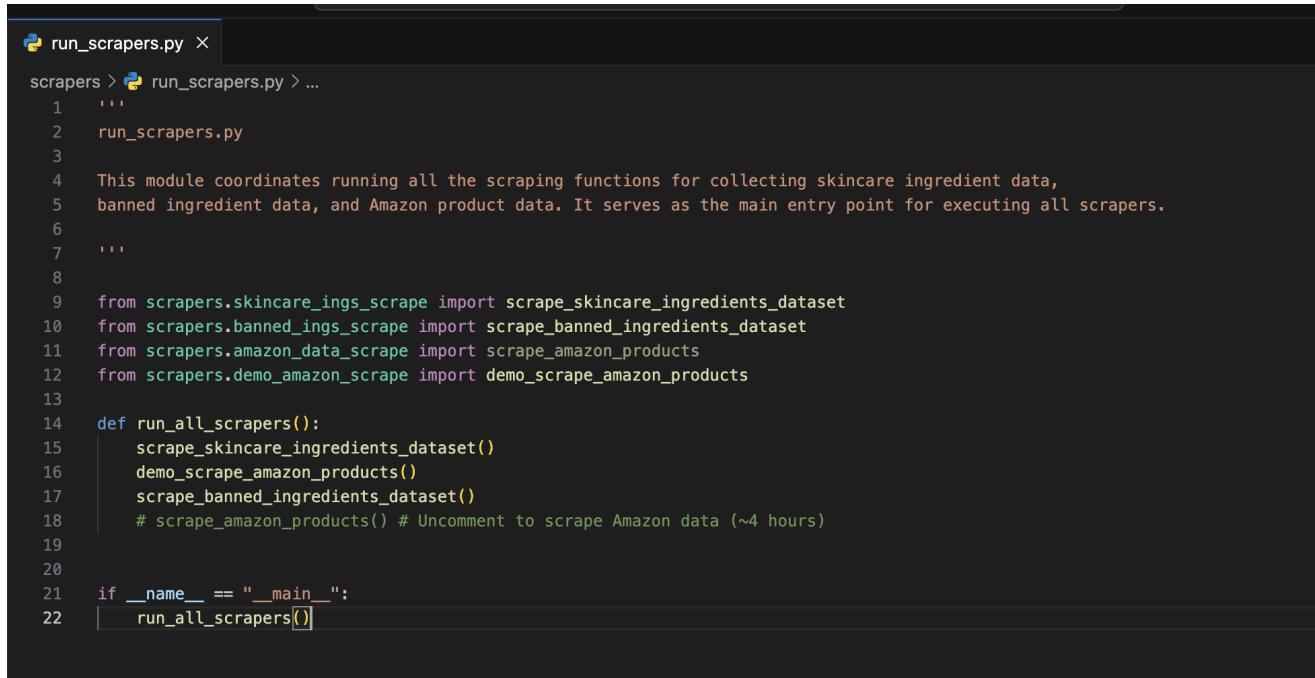
- So for project purposes, only 5 Amazon URLs will be scraped to provide a demo of how the scraping works. The scraped demo data is saved in **datasets/raw/demo\_amazon\_data.csv**.
- However, the **cleaning and merging process uses the full dataset** (**datasets/raw/amazon\_data.csv**), which contains the fully scraped data. You can compare the two files to see that the structure is the same, with the full dataset simply having more entries.

### Running the Full Amazon Scraping:

If you want to run the actual long scraping process (which takes several hours):

1. Open the **scrapers/run\_scrapers.py** file.
2. Uncomment the **line #18** as shown in the screenshot below.
3. Run the application as normal.

Here's a screenshot to show where you need to uncomment:



```
run_scrapers.py > run_scrapers.py > ...
1   ...
2   run_scrapers.py
3
4   This module coordinates running all the scraping functions for collecting skincare ingredient data,
5   banned ingredient data, and Amazon product data. It serves as the main entry point for executing all scrapers.
6
7   ...
8
9   from scrapers.skincare_ings_scrape import scrape_skincare_ingredients_dataset
10  from scrapers.banned_ings_scrape import scrape_banned_ingredients_dataset
11  from scrapers.amazon_data_scrape import scrape_amazon_products
12  from scrapers.demo_amazon_scrape import demo_scrape_amazon_products
13
14  def run_all_scrapers():
15      scrape_skincare_ingredients_dataset()
16      demo_scrape_amazon_products()
17      scrape_banned_ingredients_dataset()
18      # scrape_amazon_products() # Uncomment to scrape Amazon data (~4 hours)
19
20
21  if __name__ == "__main__":
22      run_all_scrapers()
```

Then, run the application as you normally would.

## Sample Output:

Skin Wiz App

localhost:8501

Hey Rishika! 😊

Thanks for using Skin Wiz! We've got some awesome skincare recommendations coming your way.

Recommended Products for Moisturizer:

Product	Price	Rating	Review
0 First Aid Beauty Ultra Repair Face Moisturizer with Sunscreen, SPF 30 - 2-in-1 Sun Pro	23.05	4.5	
1 Perricone MD High Potency Classics: Face Finishing & Firming Tinted Moisturizer Broa	67.15	4.5	
2 Water Drench Hyaluronic Cloud Cream Hydrating Moisturizer with Peter Thomas Roth	49	3.8	

You are from Los Angeles, CA!

City Location in America

(C) OpenStreetMap contributors

The current temperature is 26.4°C, and the maximum UV index is 6.1.

Recommended SPF for today: 30.

SKIN WIZ SKINCARE SIMPLIFIED

Skin Wiz App

localhost:8501

Recommended SPF for today: 30.

It's sunny out there, so don't forget your sunscreen! ☀️

Recommended Sunscreen Products:

Product	SPF	Price	Rating
0 First Aid Beauty Ultra Repair Face Moisturizer with Sunscreen, SPF 30 - 2-in-1 Sun Pro	30	23.05	4.5
2 Perricone MD High Potency Classics: Face Finishing & Firming Tinted Moisturizer Broa	30	67.15	4.5
1 Josie Maran Argan Oil Daily Facial Moisturizer with SPF 47 & SunBoost ATB - Lightwei	47	36	4.1

Check out some popular skincare statistics! ✨

Top 10 Most Highly Rated and Reviewed Products on Amazon

Rating: 4.8

Product Name

Review Count

Top 10 Most Highly Rated and Reviewed Products Rating: 4.8

Product Name	Rating	Review Count
First Aid Beauty Ultra Repair Cream Intense Hydration	5.0	4,800
Bioeffect Efg Serum	5.0	
Sephora Collection Cleansing Exfoliating Wipes	5.0	
La Mer Creme De La Mer	5.0	
Dermadoctor Calm Cool Corrected	5.0	
Lancer The Method: Cleanse	5.0	
Dr. Dennis Gross Skincare Alpha Beta Extra Strength Daily Peel	5.0	
Murad Hydro-Dynamic Ultimate Moisture	5.0	

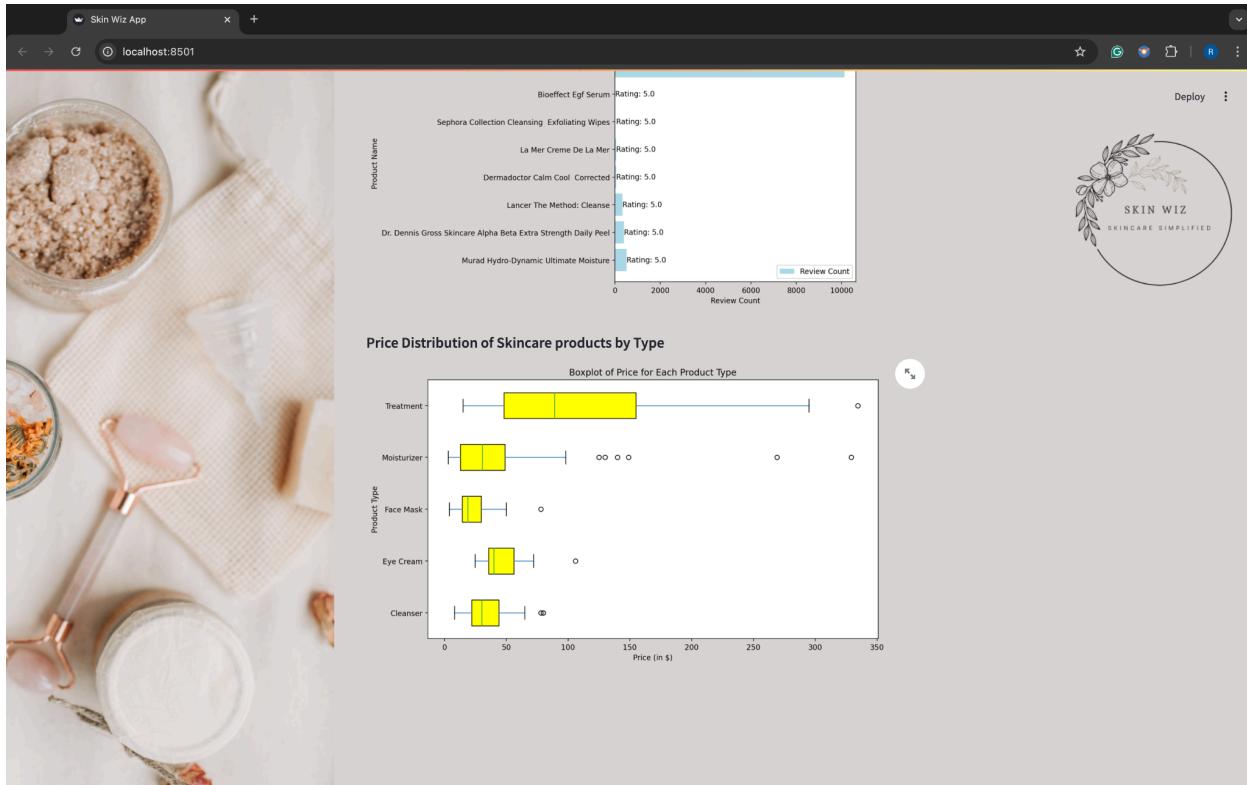
Price Distribution of Skincare products by Type

Boxplot of Price for Each Product Type

I

Boxplot of Price for Each Product Type

I



## 6. Troubleshooting

### Common Issues:

- Error When Rescraping Data:**
  - The website where the Banned ingredients dataset is scraped from may often crash or load slowly, causing an error.
  - Solution:** Rerun until the site responds.
- Permission Error (Selenium):**
  - Web scraping requires the correct web driver in your system's PATH.
  - Solution:** Ensure the driver (e.g., ChromeDriver) is installed and accessible.
- ModuleNotFoundError:**
  - Missing Python modules can cause this error.
  - Solution:** Install required modules using pip.

## 7. Demo Video

[Skinwiz Demo Video](#)