**Final Report of Traineeship Program 2024**

*On*

***“DATA ANLYTICS”***

******

**MED TOUR EASY**

**27th July 2024**

****** **ACKNOWLDEGMENTS**

I would like to express my sincere gratitude to my mentor at Med Tour Easy for their invaluable guidance and support throughout my 1-month traineeship in Data Analytics. The opportunity to work with such a dedicated team has been a remarkable experience.

I am particularly thankful to Med Tour Easy for providing this opportunity and for including the Google Data Analytics course as part of the internship, which greatly enhanced my learning. The project assigned to me was instrumental in developing my hands-on experience in data analytics. I deeply appreciate the support and encouragement from my mentors, which played a crucial role in my growth during this period.

The project is divided into two components:

1. **Building the recommendation system**

Objective: To build a recommendation system that helps users avoid allergens and prevent skin problems

1. **Data Analytics**

Objective: To analyse cosmetic product data to derive insights related to allergens and skin problems

1. **Recommendation System**

**ABSTRACT**

**Problem statement**

Consumers often face challenges when selecting cosmetic products due to the risk of allergic reactions from undisclosed allergens. Many individuals inadvertently use products containing allergens they are sensitive to, leading to severe skin issues. Despite the presence of ingredient lists on cosmetic packaging, interpreting these details is complex for non-experts.

**Solution**

To address this problem, it is essential to develop a recommendation system that helps users identify products free from allergens. This project focuses on creating such a recommendation system by analysing approximately 1472 cosmetic products and their ingredients to provide actionable insights

**Goal**

The goal is to enhance user safety and satisfaction by making informed product choices .

**1.1 About the company**

MedTourEasy, a global healthcare company, provides you the informational resources needed to evaluate your global options. MedTourEasy provides analytical solutions to our partner healthcare providers globally.

**1.2 About the Project**

The project focusses on building a simple recommendation system that helps compare similar cosmetic products and make informed decisions on purchase. The dataset had information like category, name of the product, brand, price, rank and ingredients, suitable skin types etc for 1472 products. Among these, to build this simple recommendation system, only one specific skin type (Dry) and one specific category of products (Moisturizer) were considered. And the recommendation system was built based on this sunset of data.

**METHODOLOGY**

2.1 Steps involved :

1. Tokenizing
2. Document-Term matrix
3. Cosmetic-Ingredient matrix
4. Dimensionality Reduction

and plotting in 2D

1. Adding hover tool
2. Comparison of similar products

**1.Tokenizing**

Each ingredient is assigned a number and stored in a python dictionary. Each ingredient is assigned a unique integer.

**2.Document-Term matrix(DTM)**

Creating a zero matrix , where the number of rows corresponds to the total number of products taken into consideration(190), and the number of columns corresponds to the total number of ingredients taken into consideration(2233).

**3.Cosmetic- Ingredient matrix**

Updating the DTM to create a Cosmetic -Ingredient matrix, in which each row represents a cosmetic product and each column represents an ingredient and if the ingredient is present in the cosmetic, then the value of the matrix at the position is 1, else 0. (1- present, 0- absent)

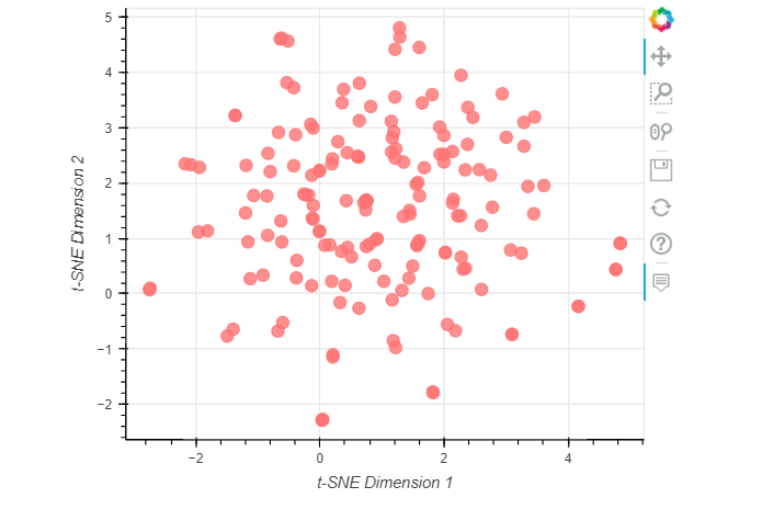
This matrix helps us identify all the ingredients present in a specific product.

**4. Dimensionality Reduction and Plotting**

Since there are huge number of columns, it’s difficult to visualize this data. So, t-SNE(t-Distributed Stochastic Neighbor Embedding), a part of Bokeh library in python, is used. It's a dimensionality reduction technique often used for

visualizing high-dimensional data by mapping it to a lower-dimensional space while preserving the data's structure.

It plots the information as a scatter plot, where each data point represents a specific cosmetic product.



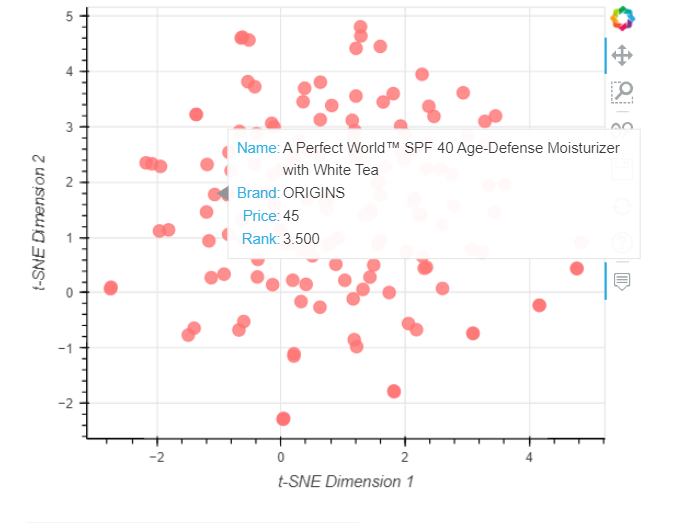
Each data point in the above plot , created using Bokeh, represents a cosmetic product

**5. Adding Hover Tool**

 Adding a hover tool allows us to check the information of each item whenever the cursor is directly over a data point.

Tooltips with each product's name, brand, price, and rank are added , to enhance user experience.

In the below diagram, you can see the pointer hovered over a specific product, and the product details getting displayed.



**6. Comparison on products**

The more closer , the more similar the products are in their composition and other features.

On hovering, it’s easy to find similar products , make comparisons on their price and rating, and make informed purchase decisions. It’s also useful in knowing the ingredients of each cosmetic product.

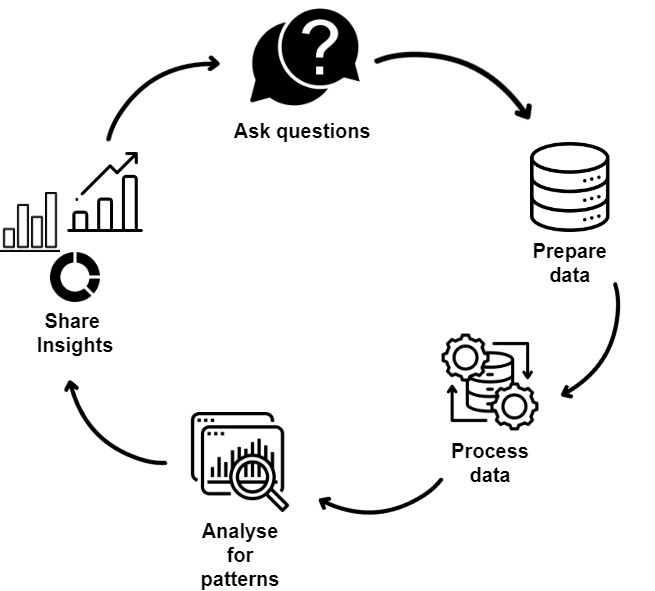
2.2 Language and Platform Used

* Python programming language
* IDE: Jupyter notebook, Libraries: t-SNE and Bokeh

**II .Data Analytics**

**Overview**

* The given dataset was analysed ,and a number of **patterns** and **relationships** were found.
* These insights were designed to inform and enhance the decision-making process for choosing the right products.
* The analysis provides valuable information to guide users towards safer and more effective cosmetic choices.

****

Analytics Life-cycle

**Key Findings**

1.Category of product most likely to contain allergens

* The United States **Food and Drug Administration (FDA)** lists allergens found in cosmetic products. Using this list, I created a list of the most common allergens in cosmetics and filtered out products containing these allergens.
* The list of allergens considered, includes

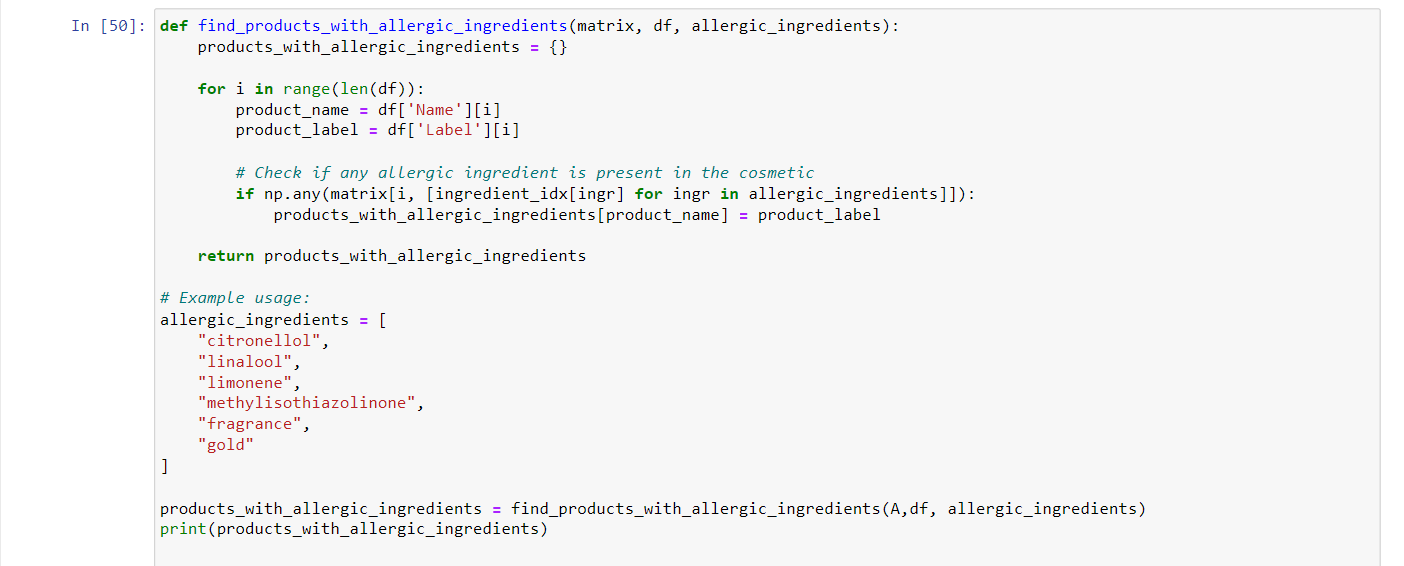
1. citronellol
2. linalool
3. limonene

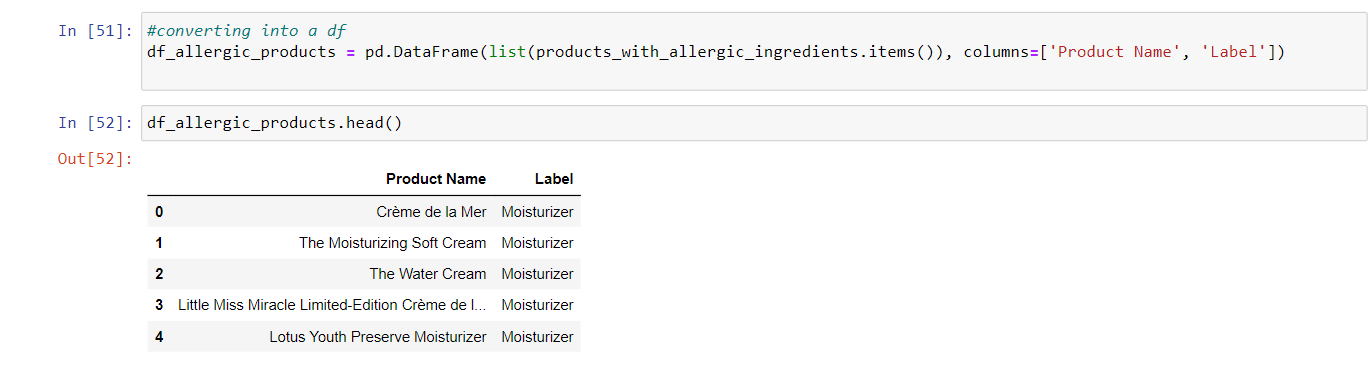
4.methylisothiazolinone

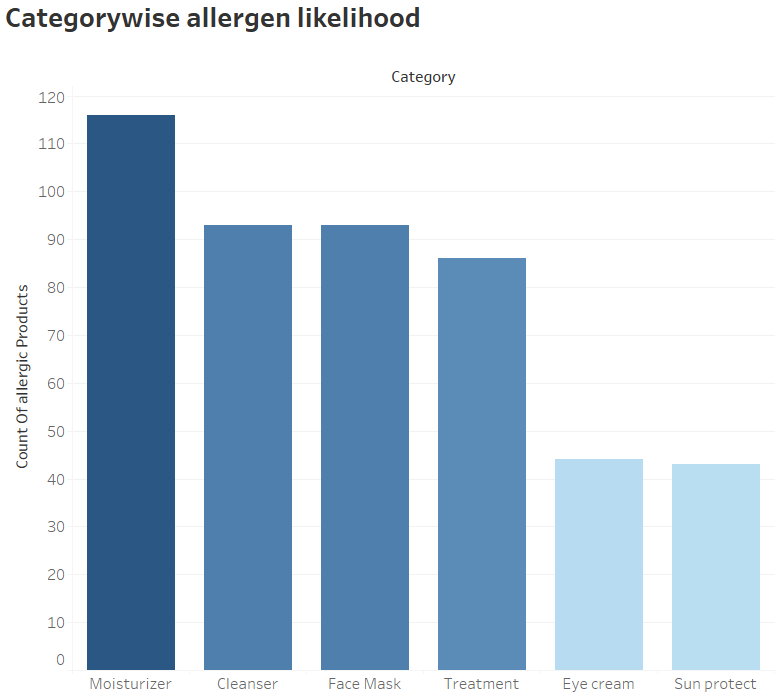
5.fragrance

6.gold

* On analysing the category of each product, it was found that the category that is most likely to have allergens is **Moisturizers.**



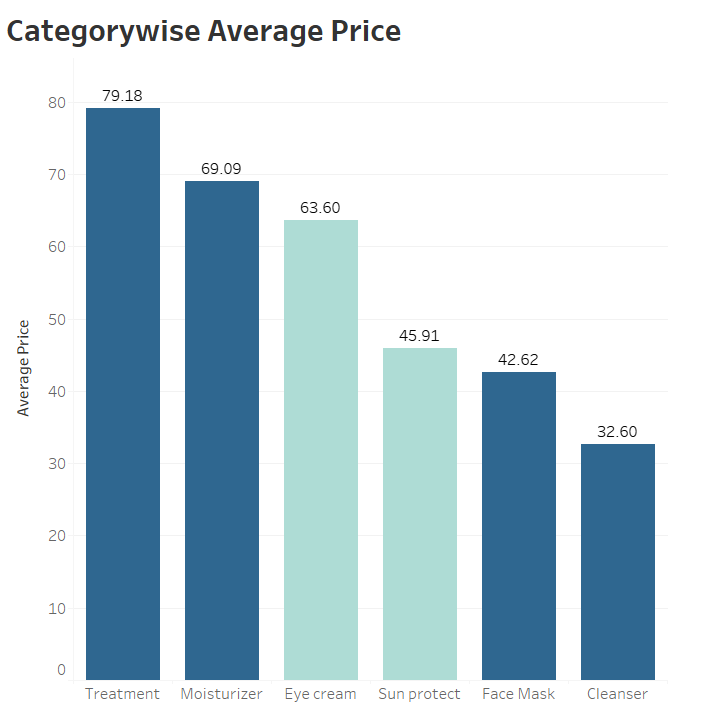
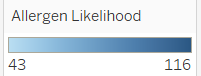




2.Average price vs Allergen Likelihood

On analysing the average price of each category it was found that

* The products that are most likely to contain allergic ingredients are either **highly expensive** or **very** **less expensive** .
* This contrasts with products less likely to contain allergens, which generally fall within a more **moderate price range.**

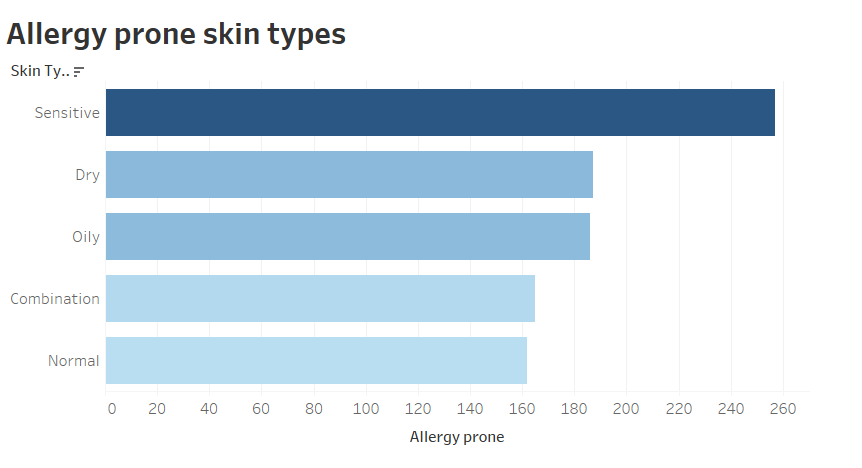
****

This relationship between price and allergen likelihood suggests the following:

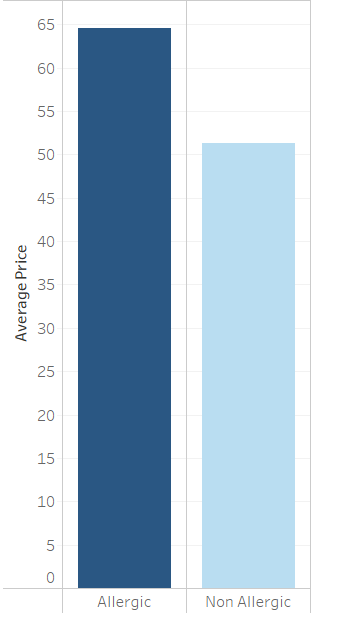
* **High-Cost Products:** These products may use premium ingredients or complex formulations, which could include rare or potent allergens, contributing to their higher cost.
* **Low-Cost Products:** These products might utilize basic or commonly used ingredients, some of which could be allergens due to their widespread use or lower-quality standards.

3.Allergy prone skin types

* Considering only the products containing allergic ingredient and their recommendations about skin types, it was found that,
* **Sensitive skin** is the most common skin type for which these products are not recommended. Other skin types that are also prone to allergies include dry and oily skin.



4.Average price comparison of Allergic and Non allergic products



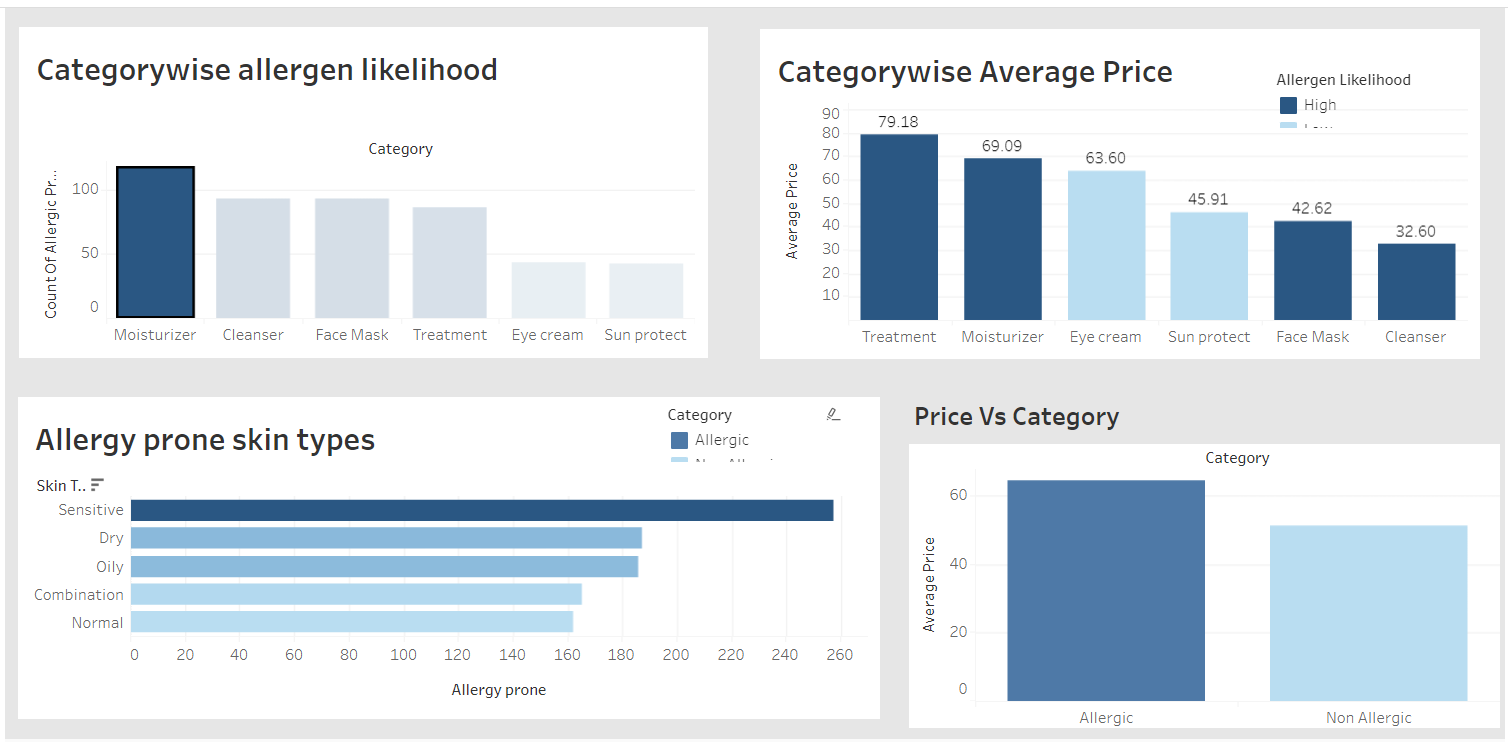
* The analysis has shown that products more prone to allergens tend to be significantly **more expensive** compared to those with lower allergen risk.

These are the key findings from the data analysis. The insights derived are valuable for making informed decisions when purchasing cosmetics. Some of the recommendations are listed below

**Recommendations**

* **Moisturizers:** These products are most likely to contain allergens, so it’s important to be cautious when purchasing them.
* **High-Cost Products:** Expensive cosmetics may contain rare or potent allergens. Exercise caution with such products.
* **Mid-Range Products:** Opting for products in a reasonable price range can be a safer choice, as they are less likely to contain high-risk allergens.
* **Sensitive Skin:** Individuals with sensitive skin should take extra care, as this skin type is particularly prone to allergies.

**Dashboard**

****

**REFERENCES**

The following websites have been referred to obtain the list of common allergens present in cosmetics.

U.S Food and Drug Administration: <https://www.fda.gov/cosmetics/cosmetic-ingredients/allergens-cosmetics>