

**Final Report of Traineeship Program 2024**

**On**

**“**[**Analysis of Chemical Components**](https://learningmanager.adobe.com/app/learner?accountId=127440&userId=25232229#/course/10040213/overview?lp_id=&cert_id=&ci_id=10727410&fromNotif=true)**”**

**23rd June 2024**



**MEDTOUREASY**



**ACKNOWLDEGMENTS**

I would also like to thank the team of MedTourEasy and my colleagues who made the working environment productive and very conducive. The traineeship opportunity that I had with MedTourEasy was a great change for learning and understanding the intricacies of the subject of Data Visualizations in Data Analytics; and also, for personal as well as professional development.

I am very obliged for having a chance to interact with so many professionals who guided me throughout the traineeship project and made it a great learning curve for me. Firstly, I express my deepest gratitude and special thanks to the Training & Development Team of MedTourEasy who gave me an opportunity to carry out my traineeship at their esteemed organization. Also, I express my thanks to the team for making me understand the details of the Data Analytics profile and training me in the same so that I can carry out the project properly and with maximum client satisfaction and also for spearing his valuable time in spite of his busy schedule.



**TABLE OF CONTENTS**

Acknowledgments .................................... i

Abstract .....................................................iii

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Topic** | **Page No.** |
| 1 | **Introduction** | 5 |
|  | 1.1 About the Company | 5 |
|  | 1.2 About the Project | 5 |
|  | 1.3 Objectives and Deliverables | 6 |
| 2 | **Methodology** | 7 |
|  | 2.1 Flow of the Project | 9 |
|  | 2.2 Software Used | 9 |
| 3 | **Implementation** | 10 |
|  | 3.1 Cosmetics, Chemicals | 13 |
|  | 3.2 One product, One type | 15 |
|  | 3.3 Tokenizing Ingredients | 16 |
|  | 3.4 Document-Term-Matrix | 17 |
|  | 3.5 Counter Function | 19 |
|  | 3.6 Cosmetic-Ingredient matrix | 20 |
|  | 3.7 t-SNE | 21 |
|  | 3.8 Bokeh | 22 |
|  | 3.9 Hover tool | 23 |
|  | 3.10 Mapping | 24 |
|  | 3.11 Comparison | 25 |
| 4 | **Observations** | 27 |
| 5 | **Conclusions & Future Scope** | 29 |
| 6 | **References** | 32 |



**ABSTRACT**

"In this project, we are focusing on the challenge of recommending suitable cosmetic products, particularly for individuals with sensitive skin, by developing a content-based recommendation system.

Our approach involves analyzing the ingredient lists of 1472 cosmetics available on Sephora using word embedding techniques. By employing t-SNE for dimension reduction and leveraging Bokeh for interactive visualization, we aim to create a tool that visually represents ingredient similarities among cosmetics.

This project aims to empower consumers with valuable insights into the chemical compositions of cosmetics, enabling them to make informed decisions tailored to their specific needs and preferences."

The cosmetics available on Sephora cover a wide range of products including skincare, makeup, haircare, fragrance, and beauty tools. Sephora offers products from various brands, both well-known and emerging, catering to different skin types, concerns, and preferences. The products range from moisturizers, serums, and cleansers in skincare to foundations, eyeshadows, lipsticks, and mascaras in makeup.

Additionally, Sephora carries haircare products like shampoos, conditioners, styling products, and treatments. Fragrances from popular brands and niche perfumers are also part of their collection. Beauty tools such as makeup brushes, sponges, and hair styling tools can also be found on Sephora's platform. Sephora regularly updates its inventory with new launches and trends in the beauty industry, providing customers with a diverse selection to choose from.



**I.INTRODUCTION**

**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**

In this project, we are focusing on developing a content-based recommendation system for cosmetic products, particularly for individuals with sensitive skin.

We are analyzing the ingredient lists of cosmetics available on Sephora using t-SNE for dimension reduction and Bokeh for interactive visualization. The goal is to create a tool that visually represents ingredient similarities among cosmetics, empowering consumers to make informed decisions tailored to their specific needs and preferences.

This project aims to provide valuable insights into the chemical compositions of cosmetics, enhancing the shopping experience for users seeking suitable products for their skin type and concerns.

Analyzing cosmetic products using t-SNE and Bokeh in machine learning involves a sophisticated process. t-SNE (t-Distributed Stochastic Neighbor Embedding) is a dimensionality reduction technique that can be used to visualize high-dimensional data in a lower-dimensional space while preserving the relationships between data points.

By applying t-SNE to the ingredient lists of cosmetic products, we can reduce the dimensionality of the data to create a visual representation that shows similarities and clusters among the products based on their ingredients.



**1.3 Objectives and Deliverables**

The objectives of this project include:

1. Developing a content-based recommendation system for cosmetic products.

2. Analyzing ingredient lists of cosmetics using t-SNE for dimension reduction.

3. Creating an interactive visualization tool using Bokeh to represent ingredient similarities.

4. Empowering consumers, especially those with sensitive skin, to make informed decisions about cosmetic products.

**The deliverables for this project will be:**

1. A functional content-based recommendation system for cosmetic products.

2. Visualization of ingredient similarities among cosmetics using t-SNE and Bokeh.

3. Interactive tools for users to explore and understand the chemical compositions of cosmetics.

4. Insights and recommendations for consumers to choose suitable cosmetic products based on their needs and preferences.



**II.METHODOLOGY**

**2.1 Flow of the Project:** The project followed the following steps to accomplish the desired objectives and deliverables. Each step has been explained in detail in the following section



**1. Data Collection:** Gather ingredient lists of cosmetic products from Sephora or relevant sources.

**2. Data Preprocessing:** Clean and preprocess the data to make it suitable for analysis.

**3. Feature Extraction:** Extract features from the ingredient lists for input into the t-SNE algorithm.

**4. Dimension Reduction:** Apply the t-SNE algorithm to reduce the dimensionality of the data and visualize ingredient similarities.

**5. Interactive Visualization:** Utilize Bokeh to create interactive visualizations of the t-SNE output for user-friendly exploration.

**6. Model Development:** Build a content-based recommendation system using the analyzed data.

**7. Testing and Evaluation:** Test the recommendation system and visualization tools for functionality and accuracy.

**8. Deployment:** Deploy the recommendation system and visualization tools for user access.

**9. Presentation:** Present the findings, insights, and recommendations to stakeholders and users.



2.2 Software Used:

**Anaconda** is a [distribution](https://en.wikipedia.org/wiki/Software_distribution) of the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) and [R](https://en.wikipedia.org/wiki/R_(programming_language)) [programming languages](https://en.wikipedia.org/wiki/Programming_language) for [scientific computing](https://en.wikipedia.org/wiki/Scientific_computing) ([data science](https://en.wikipedia.org/wiki/Data_science), [machine learning](https://en.wikipedia.org/wiki/Machine_learning) applications, large-scale [data processing](https://en.wikipedia.org/wiki/Data_processing), [predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics), etc.), that aims to simplify [package management](https://en.wikipedia.org/wiki/Package_management) and [deployment](https://en.wikipedia.org/wiki/Deployment_environment). The distribution includes data-science packages suitable for [Windows](https://en.wikipedia.org/wiki/Microsoft_Windows), [Linux](https://en.wikipedia.org/wiki/Linux), and [macOS](https://en.wikipedia.org/wiki/MacOS). It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and [Travis Oliphant](https://en.wikipedia.org/wiki/Travis_Oliphant) in 2012.

**Installation of Anaconda Software**:

1. Download the [Anaconda installer](https://www.anaconda.com/downloads).
2. (Optional) Anaconda recommends verifying the integrity of the installer after downloading it.
3. Go to your Downloads folder and double-click the installer to launch. To prevent permission errors, do not launch the installer from the [Favorites folder](https://docs.anaconda.com/reference/troubleshooting/#distro-troubleshooting-favorites-folder).
4. Click Next.
5. Read the licensing terms and click I Agree.
6. It is recommended that you install for Just Me, which will install Anaconda Distribution to just the current user account. Only select an install for All Users if you need to install for all users’ accounts on the computer (which requires Windows Administrator privileges).
7. Click Next.
8. Select a destination folder to install Anaconda and click Next. Install Anaconda to a directory path that does not contain spaces or unicode characters.



**III. Implementation**

To implement machine learning using Jupyter Notebook, you can follow these steps:

1. **After Installing Anaconda Notebook**, Follow these steps:

2. **Open Jupyter Notebook:** Launch Jupyter Notebook from the Anaconda Navigator or by typing "jupyter notebook" in the command prompt.

3. **Create a New Notebook:** Click on "New" and select a Python notebook to start coding.

4. **Import Libraries:** Import necessary libraries like NumPy, Pandas, Scikit-learn for machine learning tasks.

5. **Data Preparation:** Load your dataset, preprocess data, handle missing values, and perform feature engineering.

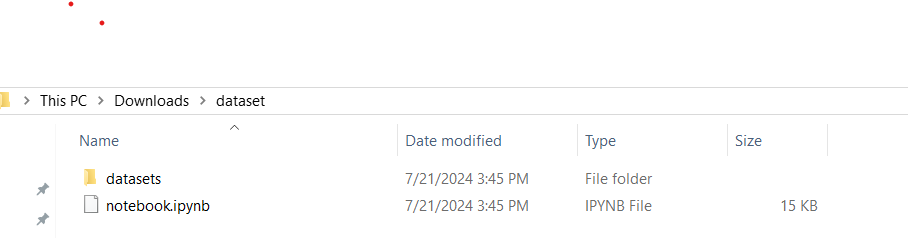
6. **Build Machine Learning Models:** Use Scikit-learn to create machine learning models such as regression, classification, clustering, etc.

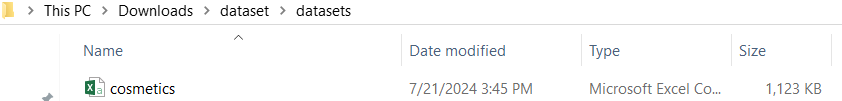
7. **Train and Evaluate Models:** Split the data into training and testing sets, train your models, make predictions, and evaluate model performance.

**8.Visualize Results:** Utilize libraries like Matplotlib or Seaborn to visualize data, model performance, and insights.

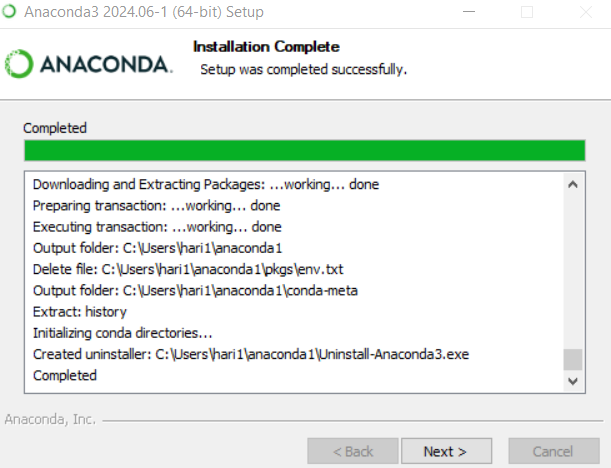


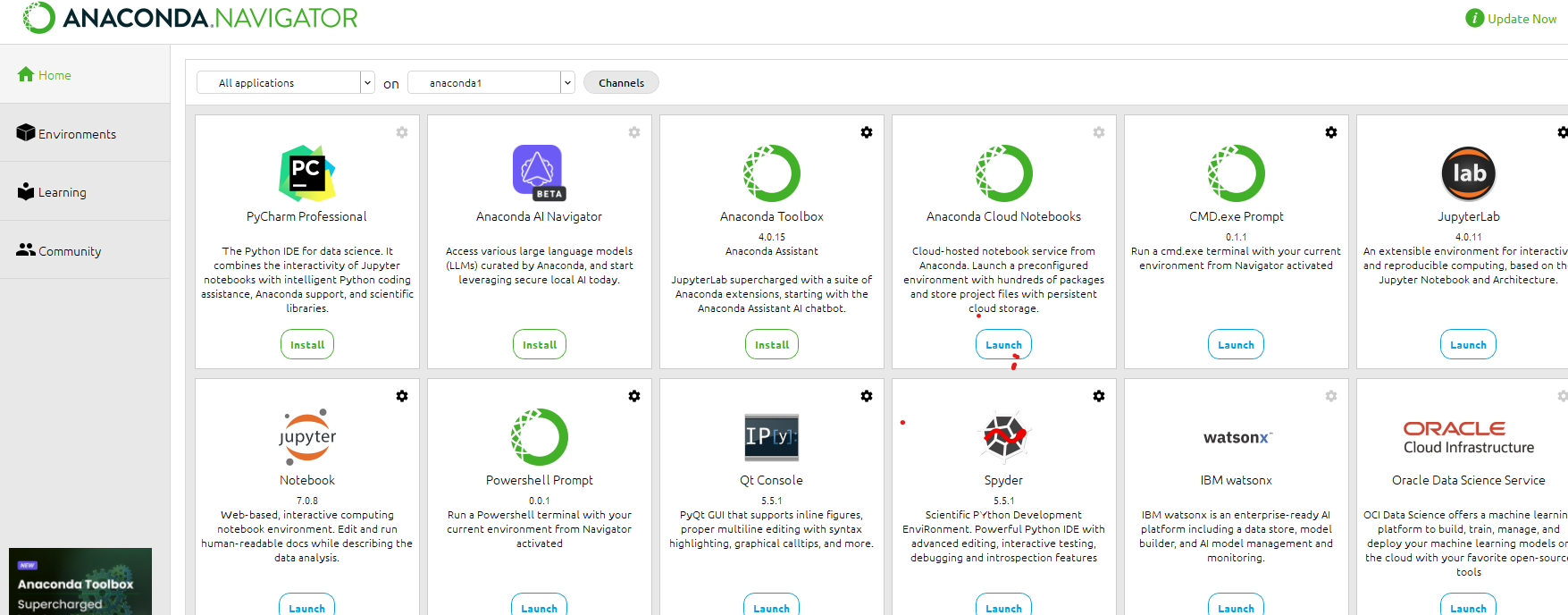
1. **Save and Share:** Save your Jupyter Notebook with the code, visualizations, and results. You can also share it with others for collaboration or presentation.













**3.1 Task 1-**

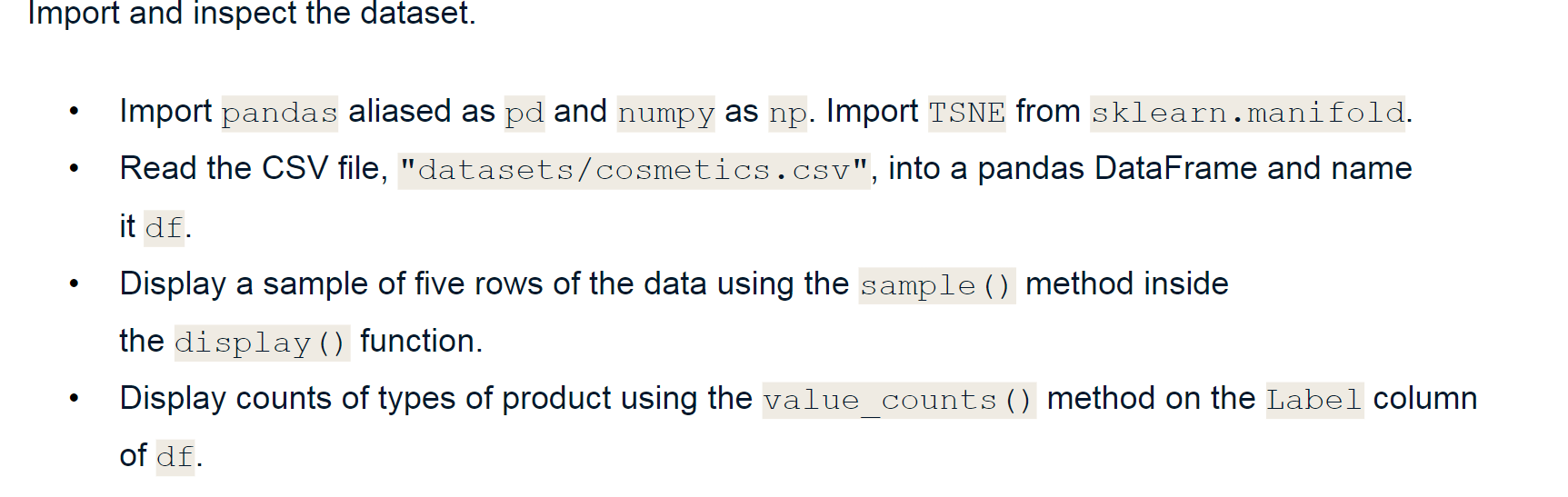
## Cosmetics, chemicals... it's complicated

Whenever I want to try a new cosmetic item, it's so difficult to choose. It's actually more than difficult. It's sometimes scary because new items that I've never tried end up giving me skin trouble.

We know the information we need is on the back of each product, but it's really hard to interpret those ingredient lists unless you're a chemist. You may be able to relate to this situation.

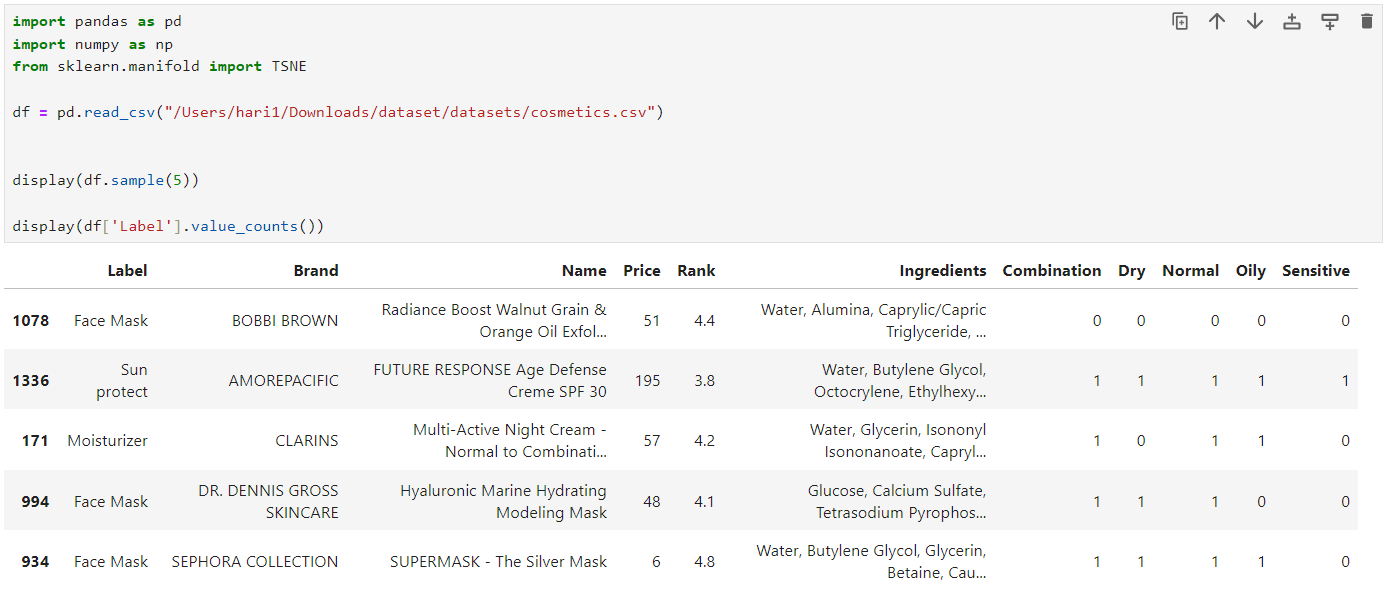
So instead of buying and hoping for the best, why don't we use data science to help us predict which products may be good fits for us?

In this notebook, we are going to create a content-based recommendation system where the 'content' will be the chemical components of cosmetics. Specifically, we will process ingredient lists for 1472 cosmetics on Sephora via [word embedding](https://en.wikipedia.org/wiki/Word_embedding), then visualize ingredient similarity using a machine learning method called t-SNE and an interactive visualization library called Bokeh. Let's inspect our data first.





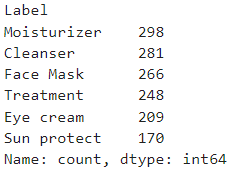
We have given a data set to find out the output, let’s proceed Step 1.



1.First, Import pandas as ps and numpy as np from sklearn.manifold.

2.Then, let us locate where the data set is located and mention it in parentices.

3.Run the code to find out the Name: count, dtype: int64





**3.2 Task 2-**

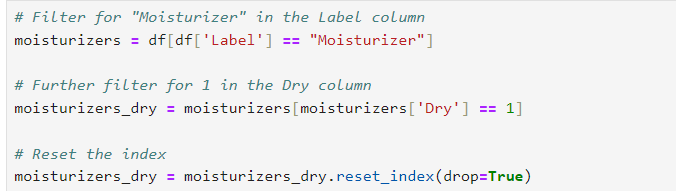
**Focus on one product category and one skin type**

There are six categories of product in our data (***moisturizers, cleansers, face masks, eye creams***, and ***sun protection***) and there are five different skin types (***combination, dry, normal, oily*** and ***sensitive***).

Because individuals have different product needs as well as different skin types, let's set up our workflow so its outputs (a t-SNE model and a visualization of that model) can be customized.

For the example in this notebook, let's focus in on moisturizers for those with dry skin by filtering the data accordingly.

Let us take **Product type as Moisturizers with Skin Type as Dry** and Run the Code.



From the given product variety, we are filtering Moisturizer that only be suitable for the skin type Dry.



**3.3 Task 3-**

## Tokenizing the ingredients

To get to our end goal of comparing ingredients in each product, we first need to do some preprocessing tasks and bookkeeping of the actual words in each product's ingredients list.

The first step will be tokenizing the list of ingredients in Ingredients column. After splitting them into tokens, we'll make a binary bag of words. Then we will create a dictionary with the tokens, ingredient\_idx, which will have the following format:

{ **"ingredient"**: index value, … }



**The index for decyl oleate is 25**



**3.4 Task 4-**

## Initializing a document-term matrix (DTM)

The next step is making a document-term matrix (DTM).

A Document Term Matrix (DTM) is a mathematical matrix that represents the frequency of terms that occur in a collection of documents.

Each row corresponds to a document, and each column corresponds to a term (word). The cells in the matrix contain the frequency of each term in the respective document. DTMs are commonly used in natural language processing and text mining tasks.

They are essential for tasks like sentiment analysis, text classification, and clustering. By converting text data into a DTM, you can analyze the frequency of words across documents, identify patterns, and extract valuable insights from the text data. It helps in quantifying and structuring text data for further analysis and machine learning tasks.

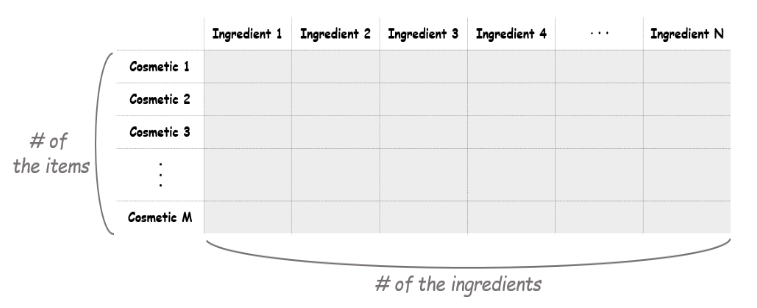
Here each cosmetic product will correspond to a document, and each chemical composition will correspond to a term. This means we can think of the matrix as a “cosmetic-ingredient” matrix.

The size of the matrix should be as the picture shown below.

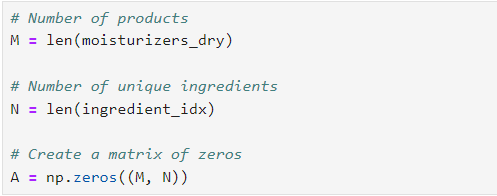
To create this matrix, we'll first make an empty matrix filled with zeros.

The length of the matrix is the total number of cosmetic products in the data. The width of the matrix is the total number of ingredients. After initializing this empty matrix, we'll fill it in the following tasks.





Here, we have taken Number of products-length from moisturizers dry, Number of unique ingredients-length from ingredient\_idx.





**3.5 Task 5 –**

## Creating a counter function

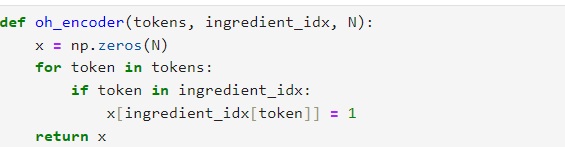
Before we can fill the matrix, let's create a function to count the tokens (i.e., an ingredients list) for each row. Creating a function to count tokens in text data is beneficial because it allows you to automate the process of tokenization and counting.

By encapsulating this functionality within a function, you can easily reuse it across different datasets or text sources without rewriting the code each time.

This saves time and effort, especially when working with large amounts of text data or when performing text analysis tasks repeatedly.

Additionally, a function provides modularity and organization to your code, making it easier to maintain and debug.

Our end goal is to fill the matrix with 1 or 0: if an ingredient is in a cosmetic, the value is 1. If not, it remains 0. The name of this function, oh\_encoder, will become clear next.

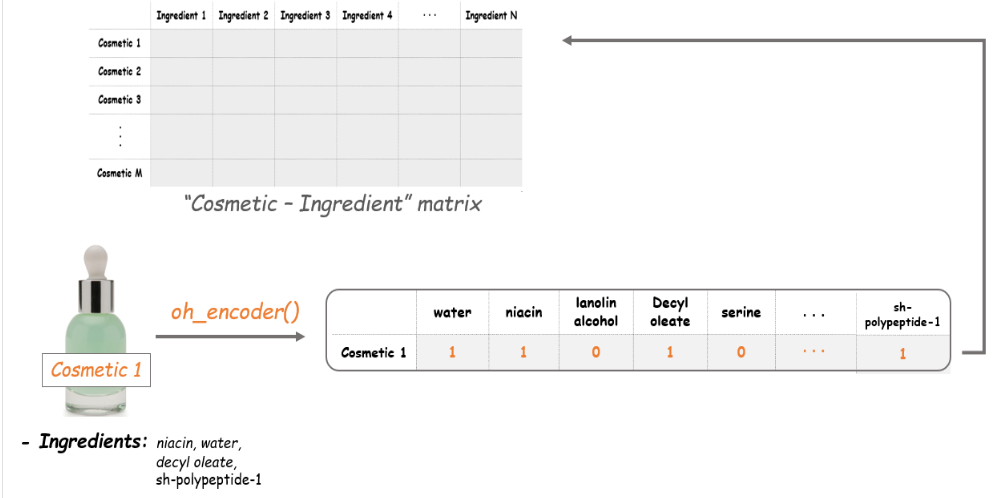




**3.6 Task 6 –**

## The Cosmetic-Ingredient matrix!

Now we'll apply the oh\_encoder () function to the tokens in corpus and set the values at each row of this matrix. So the result will tell us what ingredients each item is composed of. For example, if a cosmetic item contains water, niacin, decyl aleate and sh-polypeptide-1, the outcome of this item will be as follows.





 This is what we called one-hot encoding. By encoding each ingredient in the items, the Cosmetic-Ingredient matrix will be filled with binary values.



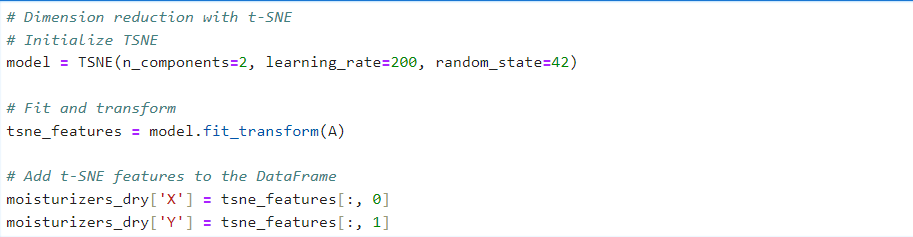
**3.7 Task 7 –**

## Dimension reduction with t-SNE

The dimensions of the existing matrix are (190, 2233), which means there are 2233 features in our data. For visualization, we should downsize this into two dimensions. We'll use t-SNE for reducing the dimension of the data here.

[T-distributed Stochastic Neighbor Embedding (t-SNE)](https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding) is a nonlinear dimensionality reduction technique that is well-suited for embedding high-dimensional data for visualization in a low-dimensional space of two or three dimensions.

Specifically, this technique can reduce the dimension of data while keeping the similarities between the instances. This enables us to make a plot on the coordinate plane, which can be said as vectorising. All of these cosmetic items in our data will be vectorized into two-dimensional coordinates, and the distances between the points will indicate the similarities between the items.

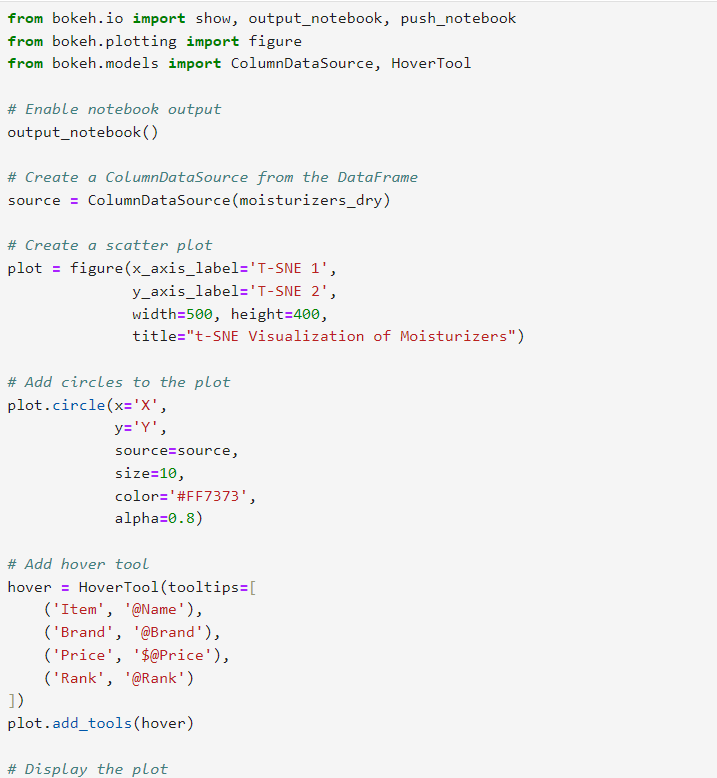




**3.8 Task 8 –**

**Let's map the items with Bokeh**

We are now ready to start creating our plot. With the t-SNE values, we can plot all our items on the coordinate plane. And the coolest part here is that it will also show us the name, the brand, the price and the rank of each item. Let's make a scatter plot using Bokeh and add a hover tool to show that information.





**3.9 Task 9 –**

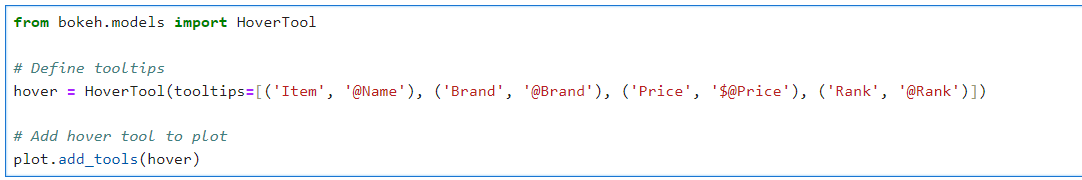
## Adding a hover tool

Why don't we add a hover tool? The hover tool is used in data visualization to provide interactive features when hovering over data points.

It allows users to view additional information or details about specific data points by simply hovering the cursor over them.

This interactive feature enhances the user experience by enabling quick access to relevant information without cluttering the visualizations. It's commonly used in tools like Bokeh and Plotly for creating interactive and informative data visualizations.

Adding a hover tool allows us to check the information of each item whenever the cursor is directly over a glyph. We'll add tooltips with each product's name, brand, price, and rank (i.e., rating).



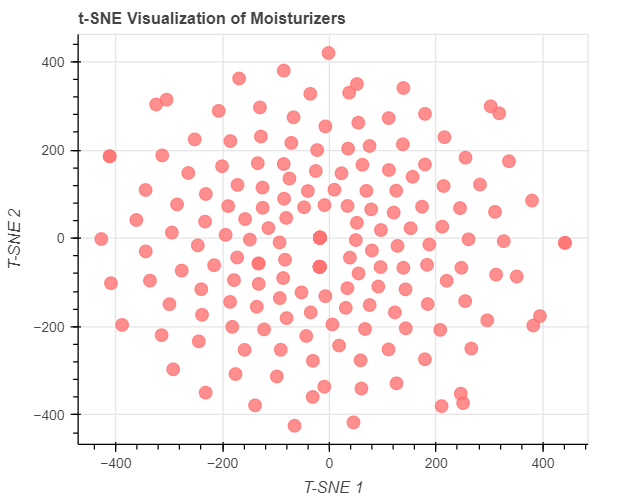


**3.10 Task 10 –**

## Mapping the cosmetic items

Finally, it's show time! Let's see how the map we've made looks like. Each point on the plot corresponds to the cosmetic items.

Then what do the axes mean here? The axes of a t-SNE plot aren't easily interpretable in terms of the original data. Like mentioned above, t-SNE is a visualizing technique to plot high-dimensional data in a low-dimensional space. Therefore, it's not desirable to interpret a t-SNE plot quantitatively.

Instead, what we can get from this map is the distance between the points (which items are close and which are far apart). The closer the distance between the two items is, the more similar the composition they have. Therefore, this enables us to compare the items without having any chemistry background.





**3.11 Task 11 –**

## Comparing two products

Since there are so many cosmetics and so many ingredients, the plot doesn't have many super obvious patterns that simpler t-SNE plots can have ([example](https://campus.datacamp.com/courses/unsupervised-learning-in-python/visualization-with-hierarchical-clustering-and-t-sne?ex=10)).

Our plot requires some digging to find insights, but that's okay!

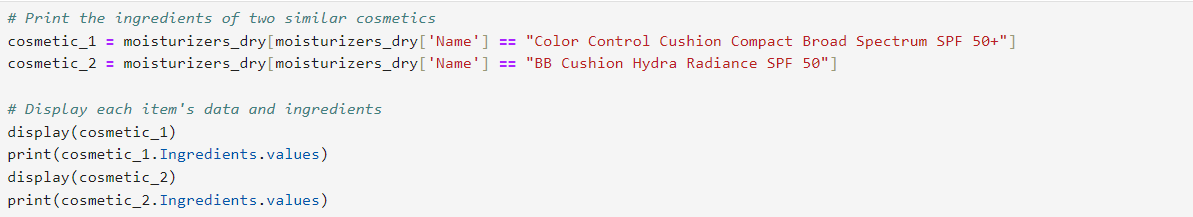
Say we enjoyed a specific product, there's an increased chance we'd enjoy another product that is similar in chemical composition.

Say we enjoyed AmorePacific's [Color Control Cushion Compact Broad Spectrum SPF 50+](https://www.sephora.com/product/color-control-cushion-compact-broad-spectrum-spf-50-P378121). We could find this product on the plot and see if a similar product(s) exist. And it turns out it does!

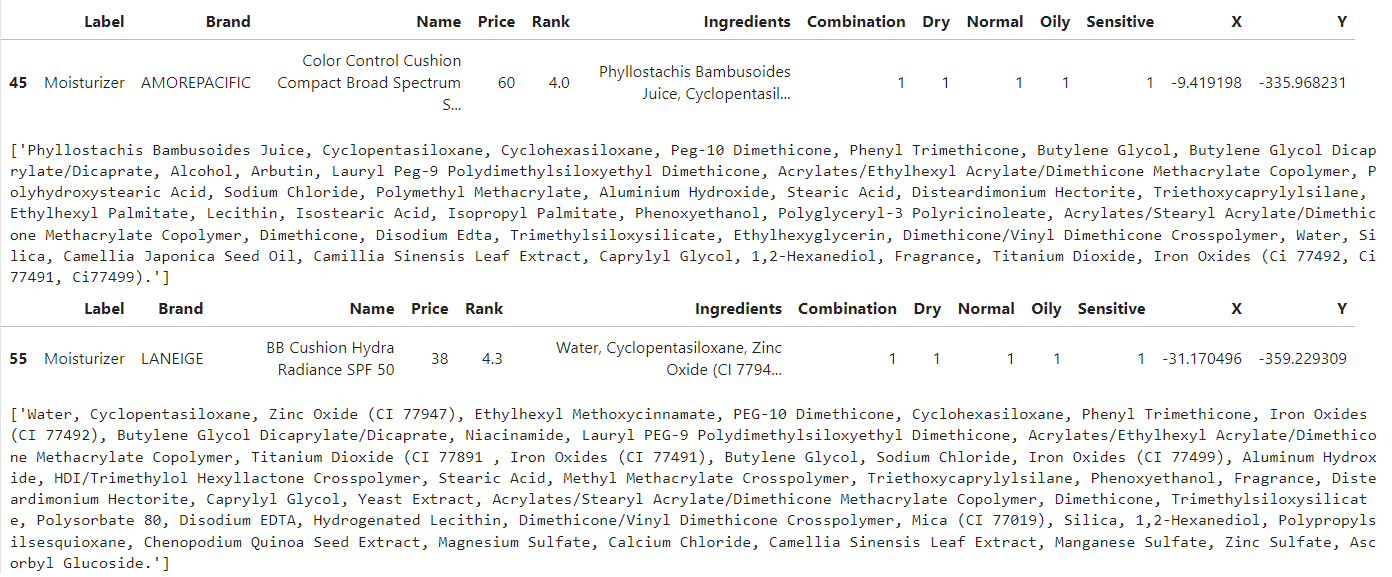
If we look at the points furthest left on the plot, we see LANEIGE's [BB Cushion Hydra Radiance SPF 50](https://www.sephora.com/product/bb-cushion-hydra-radiance-P420676) essentially overlaps with the AmorePacific product.

By looking at the ingredients, we can visually confirm the compositions of the products are similar (*though it is difficult to do, which is why we did this analysis in the first place!*), plus LANEIGE's version is $22 cheaper and actually has higher ratings.

It's not perfect, but it's useful. In real life, we can actually use our little ingredient-based recommendation engine help us make educated cosmetic purchase choices.









**IV. OBSERVATIONS**

The observation from comparing the two moisturizers, Amorepacific Color Control Cushion Compact Broad Spectrum SPF 60 and Laneige BB Cushion Hydra Radiance SPF 50, reveals differences in their key characteristics.

1. Price: The Amorepacific moisturizer is priced at $60, while Laneige's is $38.4, indicating a notable price variation between the two products.  
  
2. Rank: Amorepacific is rated higher at 4.0 compared to Laneige's 3.4, suggesting that the former is perceived to be of superior quality or effectiveness.  
  
3. Ingredients: Both moisturizers contain a range of ingredients such as cyclopentasiloxane, various oxides, and butylene glycol, but they also differ in the specific formulations and additional components included.  
  
4. Skin Type Compatibility: The Amorepacific moisturizer is suitable for dry, normal, oily, and sensitive skin types, while Laneige's product is also formulated for all skin types except sensitive, indicating a potential limitation for those with sensitive skin.  
  
5. SPF Protection: Amorepacific offers broad-spectrum protection with SPF 60, whereas Laneige provides SPF 50 protection, highlighting a slight variance in sun protection levels.  
  
Overall, the observation showcases that while both moisturizers offer similar benefits and function as color control cushions, there are distinctions in their price, ranking, ingredients, skin type compatibility, and SPF protection, allowing consumers to make an informed choice based on their specific needs and preferences.









****



**V.CONCLUSION AND FUTURE SCOPE**

Based on the data provided, here are the conclusions and future scope for these two moisturizers:

**Conclusion**:

**1. Product Comparison:**

- AMOREPACIFIC Color Control Cushion Compact Broad Spectrum SPF 50:

- Price: $60

- Rank: 4.0

- Ingredients: Contains a mix of silicones (e.g., Cyclopentasiloxane, Dimethicone), UV filters (e.g., Titanium Dioxide, Iron Oxides), and botanical extracts (e.g., Camellia Japonica Seed Oil, Camellia Sinensis Leaf Extract). It also includes alcohol, which may be a concern for sensitive skin types.

- Target Skin Types: Suitable for all skin types (Combination, Dry, Normal, Oily, Sensitive) according to the data.

- **LANEIGE BB Cushion Hydra Radiance SPF 50**:

- Price: $38

- Rank: 4.3

- Ingredients: Includes UV filters (e.g., Zinc Oxide, Titanium Dioxide), silicones (e.g., Cyclopentasiloxane, Dimethicone), and botanical extracts (e.g., Camellia Sinensis Leaf Extract). It also contains niacinamide, which is beneficial for brightening and improving skin texture.

- Target Skin Types: Also suitable for all skin types (Combination, Dry, Normal, Oily, Sensitive) according to the data.



**2. Price and Value:**

- LANEIGE offers a lower price point ($38) compared to \*\*AMOREPACIFIC\*\* ($60) while maintaining a slightly higher rank (4.3 vs. 4.0). This indicates that the LANEIGE product may offer better value for money, given its lower price and higher customer satisfaction.

**3. Ingredients Analysis:**

- Both products contain similar core ingredients like silicones and UV filters, which help in providing a smooth application and sun protection.

- LANEIGE includes niacinamide, which has additional benefits such as reducing hyperpigmentation and improving skin texture, potentially giving it an edge over \*\*AMOREPACIFIC\*\* in terms of skincare benefits.

**4. Suitability for Sensitive Skin:**

- Both products list sensitive skin as a target category, but \*\*AMOREPACIFIC\*\* includes alcohol, which might irritate very sensitive skin. \*\*LANEIGE\*\* does not list alcohol as a major ingredient, possibly making it a better option for those with sensitive skin.

**Future Scope:**

**1. Ingredient Innovations:**

- Future products could incorporate advanced, skin-friendly ingredients that enhance efficacy without causing irritation. This could include more natural extracts, peptides, or other innovative compounds that improve skin health while reducing potential irritants.

**2. Formulation Improvements:**

- Given the presence of alcohol in the \*\*AMOREPACIFIC\*\* product, future formulations could aim to reduce or eliminate alcohol to cater to sensitive skin types. Similarly, improvements could focus on enhancing moisturizing and anti-aging properties.



**3. Consumer Preferences:**

- Analyzing consumer feedback and preferences more deeply can guide product development. For instance, formulations could be adjusted based on the popularity of certain ingredients or the specific needs of different skin types.

**4. Price Point Analysis:**

- Companies could explore offering similar high-quality products at competitive prices to attract cost-conscious consumers. Comparing price versus effectiveness can drive product positioning and marketing strategies.

**5. Sustainability:**

- There is a growing trend towards sustainability in the beauty industry. Future products could focus on eco-friendly packaging, cruelty-free testing, and sustainable sourcing of ingredients.

**6. Personalization:**

- As skincare becomes more personalized, future products could offer customizable options based on individual skin assessments, improving efficacy and customer satisfaction.

In summary, while both \*\*AMOREPACIFIC\*\* and \*\*LANEIGE\*\* offer high-quality moisturizers with broad-spectrum sun protection and beneficial ingredients, \*\*LANEIGE\*\* may offer better value and suitability for sensitive skin due to its lower price and lack of alcohol. Future developments could focus on ingredient innovation, formulation improvements, consumer preferences, price competitiveness, sustainability, and personalization.

**VI.REFERENCES**

**Data Collection:**

The following websites have been referred to obtain the input data and statistics:

<https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML>

<https://www.w3schools.com/python/python_ml_getting_started.asp>

<https://www.ibm.com/topics/machine-learning>

<https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>

<https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding#:~:text=t%2Ddistributed%20stochastic%20neighbor%20embedding%20(t%2DSNE)%20is,two%20or%20three%2Ddimensional%20map>.

<https://docs.bokeh.org/en/latest/>

<https://discourse.holoviz.org/t/how-to-set-up-hovertool-for-multiple-columns-with-hvplot/4294>

<https://towardsdatascience.com/learning-product-similarity-in-e-commerce-using-a-supervised-approach-525d734afd99>

**Software References:**

The following websites have been referred for Anaconda installation and Jupyter notebook:

<https://docs.anaconda.com/anaconda/install/windows/>

<https://www.star.nesdis.noaa.gov/atmospheric-composition-training/software_anaconda_install.php>

<https://www.anaconda.com/download>

<https://clouds.eos.ubc.ca/~phil/docs/problem_solving/01-Orientation/01.03-Installing-Anaconda-on-Windows.html>

<https://clouds.eos.ubc.ca/~phil/docs/problem_solving/01-Orientation/01.03-Installing-Anaconda-on-Windows.html>

<https://www.dataquest.io/blog/jupyter-notebook-tutorial/>

<https://docs.jupyter.org/en/latest/install/notebook-classic.html>