**Comparing Cosmetics by Ingredients**

Process ingredient lists for cosmetics on Sephora then visualize similarity using t-SNE and Bokeh.

#### Project Description

Buying new cosmetic products is difficult. It can even be scary for those who have sensitive skin and are prone to skin trouble. The information needed to alleviate this problem is on the back of each product, but it's tought to interpret those ingredient lists unless you have a background in chemistry.

Instead of buying and hoping for the best, we can use data science to help us predict which products may be good fits for us. In this Project, you are going to create a content-based recommendation system where the 'content' will be the chemical components of cosmetics. Specifically, you 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.

This Project lets you apply the skills from [Manipulating DataFrames with pandas](https://www.datacamp.com/courses/manipulating-dataframes-with-pandas), Chapter 1 of [Dimensionality Reduction in Python](https://www.datacamp.com/courses/dimensionality-reduction-in-python), and [Interactive Data Visualization with Bokeh](https://www.datacamp.com/courses/interactive-data-visualization-with-bokeh). This Project also includes the concepts of natural language processing and word embedding, which you can learn about in [Natural Language Processing Fundamentals in Python](https://www.datacamp.com/courses/natural-language-processing-fundamentals-in-python).

#### Project Tasks

* 1 Cosmetics, chemicals... it's complicated
* 2 Focus on one product category and one skin type
* 3 Tokenizing the ingredients
* 4 Initializing a document-term matrix (DTM)
* 5 Creating a counter function
* 6 The Cosmetic-Ingredient matrix!
* 7 Dimension reduction with t-SNE
* 8 Let's map the items with Bokeh
* 9 Adding a hover tool
* 10 Mapping the cosmetic items
* 11 Comparing two products