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)](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.

In [1]:
```
# Import libraries
import pandas as pd
import numpy as np
from sklearn.manifold import TSNE

# Load the data
df = pd.read_csv('datasets/cosmetics.csv')

# Check the first five rows
display(df.sample(5))
#display(df.head(5)) Both will do

# Inspect the types of products
df.Label.value_counts()
```

	Label	Brand	Name	Price	Rank	Ingredients	Combination	Dry	Normal	Oily	Sensitive
467	Cleanser	SEPHORA COLLECTION	Cleansing Wipes - Cucumber - Anti-fatigue	7	4.4	No Info	0	0	0	0	0
250	Moisturizer	IT COSMETICS	Your Skin But Better CC+ Illumination Cream wi...	38	3.8	Visit the IT Cosmetics boutique	0	0	0	0	0
274	Moisturizer	BOBBI BROWN	Hydrating Face Cream Moisturizer	60	4.3	Water, Dimethicone, Butylene Glycol, Cetyl Alc...	0	0	0	0	0
1024	Face Mask	DIOR	Hydra Life Extra Plump Smooth Balm Mask	69	4.6	-Natural White Pine Oil: Comforts and soothes....	1	1	1	0	0
60	Moisturizer	OLEHENRIKSEN	C-Rush™ Brightening Gel Crème	44	4.6	Water, Glycerin, Dicaprylyl Carbonate, Propane...	1	1	1	1	1

Out[1]:
```
Moisturizer    298
Cleanser       281
Face Mask      266
Treatment      248
Eye cream      209
Sun protect    170
Name: Label, dtype: int64
```

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.

In [2]:
```
# Filter for moisturizers
moisturizers = df.loc[df['Label']=='Moisturizer']

# Filter for dry skin as well
moisturizers_dry = moisturizers.loc[moisturizers['Dry'] == 1]

# Reset index
moisturizers_dry = moisturizers_dry.reset_index(drop = True)
display(moisturizers_dry)
```

	Label	Brand	Name	Price	Rank	Ingredients	Combination	Dry	Normal	Oily	Sensitive
0	Moisturizer	LA MER	Crème de la Mer	175	4.1	Algae (Seaweed) Extract, Mineral Oil, Petrolat...	1	1	1	1	1
1	Moisturizer	SK-II	Facial Treatment Essence	179	4.1	Galactomyces Ferment Filtrate (Pitera), Butyle...	1	1	1	1	1
2	Moisturizer	DRUNK ELEPHANT	Protini™ Polypeptide Cream	68	4.4	Water, Dicaprylyl Carbonate, Glycerin, Ceteary...	1	1	1	1	0
3	Moisturizer	LA MER	The Moisturizing Soft Cream	175	3.8	Algae (Seaweed) Extract, Cyclopentasiloxane, P...	1	1	1	1	1
4	Moisturizer	IT COSMETICS	Your Skin But Better™ CC+™ Cream with SPF 50+	38	4.1	Water, Snail Secretion Filtrate, Phenyl Trimet...	1	1	1	1	1
5	Moisturizer	DRUNK ELEPHANT	Lala Retro™ Whipped Cream	60	4.2	Water, Glycerin, Caprylic/ Capric Triglyceride...	1	1	1	1	0
6	Moisturizer	DRUNK ELEPHANT	Virgin Marula Luxury Facial Oil	72	4.4	100% Unrefined Sclerocraya Birrea (Marula) Ker...	1	1	1	1	0
7	Moisturizer	KIEHL'S SINCE 1851	Ultra Facial Cream	29	4.4	Water, Glycerin, Cyclohexasiloxane, Squalane, ...	1	1	1	1	1
8	Moisturizer	KIEHL'S SINCE 1851	Midnight Recovery Concentrate	47	4.4	Caprylic/Capric Triglyceride Dicaprylyl Carbon...	1	1	1	1	1
9	Moisturizer	SUNDAY RILEY	Luna Sleeping Night Oil	105	4.1	Persea Gratissima (Extra Virgin, Cold Pressed ...	1	1	1	1	1
10	Moisturizer	FARMACY	Honeymoon Glow AHA Resurfacing Night Serum wit...	58	4.6	Water, Lactic Acid, Propanediol, Jojoba Esters...	1	1	1	1	1
11	Moisturizer	DRUNK ELEPHANT	The Littles™	90	4.4	Beste™ No.9 Jelly Cleanser: Water, Sodium Laur...	1	1	1	1	0
12	Moisturizer	FIRST AID BEAUTY	Ultra Repair® Cream Intense Hydration	30	4.6	Water, Stearic Acid, Glycerin, C12-15 Alkyl Be...	1	1	1	1	1
13	Moisturizer	CLINIQUE	Moisture Surge 72-Hour Auto-Replenishing Hydrator	39	4.4	Water , Dimethicone , Butylene Glycol , Glycer...	1	1	1	1	1
14	Moisturizer	SK-II	R.N.A. POWER Face Cream	230	4.3	Water, Glycerin, Galactomyces Ferment Filtrate...	0	1	1	0	1

	Label	Brand	Name	Price	Rank	Ingredients	Combination	Dry	Normal	Oily	Sensitive
15	Moisturizer	LA MER	Crème de la Mer Mini	85	4.1	Algae (Seaweed) Extract, Mineral Oil, Petrolat...	1	1	1	1	1
16	Moisturizer	FRESH	Black Tea Firming Overnight Mask	92	4.1	Water, Glycerin, Butylene Glycol, Jojoba Ester...	1	1	1	0	0
17	Moisturizer	BELIF	The True Cream Moisturizing Bomb	38	4.6	Water, Glycerin, Cyclohexasiloxane, Hydrogenat...	0	1	1	0	0
18	Moisturizer	DRUNK ELEPHANT	Virgin Marula Luxury Facial Oil Mini	40	4.5	100% Unrefined Sclerocraya Birrea (Marula) Ker...	1	1	1	1	0
19	Moisturizer	ORIGINS	Dr. Andrew Weil For Origins™ Mega-Mushroom Rel...	34	4.4	Water, Butylene Glycol, PEG-4, Citrus Aurantiu...	1	1	1	1	1
20	Moisturizer	CLINIQUE	Dramatically Different Moisturizing Lotion+	28	3.9	Water , Mineral Oil/Paraffinum Liquidum/Huile ...	1	1	0	0	0
21	Moisturizer	SK-II	GenOptics Aura Essence Serum	240	4.1	Water, Galactomyces Ferment Filtrate (Pitera),...	1	1	1	1	1
22	Moisturizer	TATCHA	Pure One Step Camellia Cleansing Oil	48	4.5	Cetyl Ethylhexanoate, Oryza Sativa (Rice) Bran...	1	1	1	1	1
23	Moisturizer	OLEHENRIKSEN	Sheer Transformation® Perfecting Moisturizer	38	4.2	Visit the OLEHENRIKSEN boutique	1	1	1	1	1
24	Moisturizer	JOSIE MARAN	100 percent Pure Argan Oil	48	4.5	Organic Argania Spinosa (Argan) Kernel Oil*. *...	0	1	0	1	1
25	Moisturizer	IT COSMETICS	Your Skin But Better CC+ Cream Oil-Free Matte ...	38	3.9	Water, Dimethicone, Butylene Glycol Dicaprylat...	1	1	1	1	0
26	Moisturizer	LANEIGE	Water Sleeping Mask	25	4.4	Water, Butylene Glycol, Cyclopentasiloxane, Gl...	1	1	1	1	1
27	Moisturizer	LANEIGE	Water Bank Moisture Cream	35	4.4	Water, Glycerin, Butylene Glycol, Squalane, Di...	0	1	1	0	1
28	Moisturizer	SK-II	Facial Treatment Essence Mini	99	4.1	Galactomyces Ferment Filtrate (Pitera), Butyle...	1	1	1	1	1
29	Moisturizer	TATCHA	Luminous Dewy Skin Mist	48	4.0	Water, Glycerin, Squalane (Olive Origin), Cycl...	1	1	1	0	1
...	...	...	...	...	...	...	...	...	...	...	...

	Label	Brand	Name	Price	Rank	Ingredients	Combination	Dry	Normal	Oily	Sensitive
160	Moisturizer	SHISEIDO	Ibuki Beauty Sleeping Mask	40	4.4	Water, Dipropylene Glycol, Glycerin, Alcohol, ...	1	1	1	1	0
161	Moisturizer	LANCÔME	Rénergie Lift Multi-Action Sunscreen Broad Spe...	99	3.9	Water, Dimethicone, Glycerin, Hydrogenated Pol...	1	1	1	1	1
162	Moisturizer	BIOSSANCE	Squalane + Antioxidant Cleansing Oil	30	4.4	Caprylic/Capric Triglyceride, Polyglyceryl-2 C...	1	1	1	1	1
163	Moisturizer	ORIGINS	Plantscription™ Youth-Renewing Power Night Cream	60	4.4	Water, Caprylic/Capric Triglyceride, Myristyl ...	1	1	1	1	1
164	Moisturizer	LANCÔME	Énergie de Vie The Smoothing & Plumping Water-...	55	4.2	Visit the Lancôme boutique	1	1	1	1	1
165	Moisturizer	KATE SOMERVILLE	ExfoliKate Glow Moisturizer	65	4.6	Water, Cetyl Ricinoleate, Isostearyl Palmitate...	1	1	1	1	0
166	Moisturizer	J.ONE	Jelly Pack	42	4.3	Water, Polysorbate 80, PEG-150 Disterate, Niac...	1	1	1	1	1
167	Moisturizer	KATE SOMERVILLE	RetAsphere™ 2-in-1 Retinol Night Cream	85	4.3	Water, Ethylhexyl Palmitate, Diethylhexyl Carb...	1	1	1	1	0
168	Moisturizer	PERRICONE MD	High Potency Classics: Face Finishing & Firmin...	69	4.3	Water, Ethylhexyl Palmitate, Glycerin, Aleurit...	1	1	1	0	1
169	Moisturizer	JOSIE MARAN	Argan Infinity Cream Intensive Creamy Oil	28	4.5	-100 Percent Pure Argan Oil: Nourishes and pro...	1	1	1	0	1
170	Moisturizer	TATA HARPER	Repairative Moisturizer	110	3.4	*Ingredients from organic farming. **Clinical ...	0	1	0	0	0
171	Moisturizer	DR. JART+	Water Drop Hydrating Moisturizer	36	4.1	#NAME?	1	1	1	1	1
172	Moisturizer	HERBIVORE	Moon Fruit Superfruit Night Treatment	58	4.0	Organic Aloe Leaf Juice (Aloe Barbadensis), Re...	1	1	1	1	1
173	Moisturizer	LANCÔME	Visionnaire Advanced Multi-Correcting Cream	88	4.6	Water, Glycerin, Dicaprylyl Ether, Cyclohexasi...	1	1	1	1	1
174	Moisturizer	CAUDALIE	Premier Cru Rich Cream	140	4.5	Water, Coco-Caprylate/Caprate*, Butylene Glyco...	1	1	1	1	1

	Label	Brand	Name	Price	Rank	Ingredients	Combination	Dry	Normal	Oily	Sensitive
175	Moisturizer	LANCÔME	Hydra Zen Anti-Stress Gel Moisturizer	48	4.6	Water, Glycerin, Cyclohexasiloxane, Cetearyl E...	1	1	1	1	1
176	Moisturizer	FIRST AID BEAUTY	5 in 1 Face Cream SPF 30	40	3.9	-Zinc Oxide (OTC): Filters harmful UV rays. -A...	1	1	1	1	1
177	Moisturizer	HERBIVORE	Rose Hibiscus Coconut Water Hydrating Face Mist	32	4.1	Water, Aloe Barbadensis Leaf Water, Rosa Damas...	1	1	1	1	1
178	Moisturizer	TATA HARPER	Retinoic Nutrient Face Oil	48	4.3	Visit the Tata Harper boutique	1	1	1	1	1
179	Moisturizer	CLINIQUE	Limited Edition Dramatically Different Moistur...	39	0.0	Water, Mineral Oil/Paraffinum Liquidum/Huile M...	1	1	0	0	0
180	Moisturizer	PHILOSOPHY	Hope In A Jar	39	4.0	Visit the philosophy boutique	1	1	1	1	1
181	Moisturizer	KIEHL'S SINCE 1851	Ultra Facial Cream SPF 30	29	4.2	Water, Glycerin, Squalane, Dimethicone, Peg-10...	1	1	1	1	1
182	Moisturizer	DIOR	Capture Youth Age-Delay Advanced Crème	95	3.9	Visit the Dior boutique	1	1	1	1	1
183	Moisturizer	GLAMGLOW	DREAMDUO™ Overnight Transforming Treatment	59	4.2	Water, Dimethicone, Isohexadecane, Glycerin, B...	1	1	1	1	0
184	Moisturizer	ALGENIST	POWER Recharging Night Pressed Serum	95	4.4	Cocos Nucifera (Coconut) Water, Water (Aqua, E...	1	1	1	1	1
185	Moisturizer	KIEHL'S SINCE 1851	Ultra Facial Deep Moisture Balm	29	4.7	Water, Glycerin, Shea Butter, Glyceryl Stearat...	0	1	1	0	0
186	Moisturizer	SHISEIDO	White Lucent All Day Brightener Broad Spectrum...	62	4.6	Water, Sd Alcohol 40-B, Dimethicone, Dipropyle...	1	1	1	0	0
187	Moisturizer	SATURDAY SKIN	Featherweight Daily Moisturizing Cream	49	4.6	Water, Butylene Glycol, Ethylhexyl Palmitate, ...	1	1	1	1	1
188	Moisturizer	KATE SOMERVILLE	Goat Milk Moisturizing Cream	65	4.1	Water, Ethylhexyl Palmitate, Myristyl Myristat...	1	1	1	1	1
189	Moisturizer	GO-TO	Face Hero	34	4.8	Almond Oil, Jojoba Oil, Macadamia Oil, Brazil ...	1	1	1	1	1

190 rows × 11 columns

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:

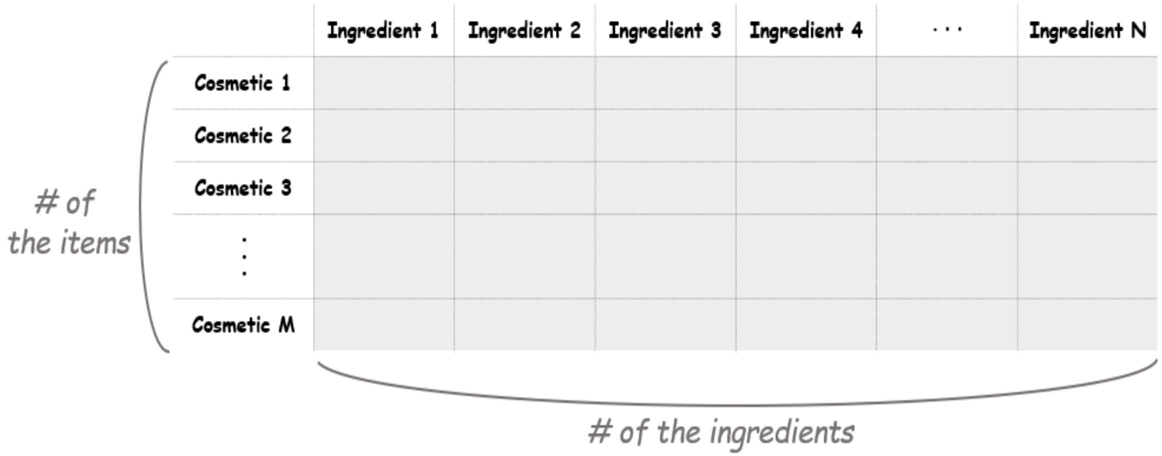
{ ***"ingredient"***: index value, ... }

In [3]:
```
# Initialize dictionary, list, and initial index
ingredient_idx = {}
corpus = []
idx = 0

# For loop for tokenization
for i in range(len(moisturizers_dry)):
    ingredients = moisturizers_dry['Ingredients'][i]
    ingredients_lower = ingredients.lower()
    tokens = ingredients_lower.split(', ')
    corpus.append(tokens)
    for ingredient in tokens:
        if ingredient not in ingredient_idx:
            ingredient_idx[ingredient] = idx
            idx += 1

# Check the result
print("The index for decyl oleate is", ingredient_idx['decyl oleate'])
```

```
The index for decyl oleate is 25
```

4. Initializing a document-term matrix (DTM)

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

In [4]:
```
# Get the number of items and tokens
M = len(moisturizers_dry)
N = len(ingredient_idx)

# Initialize a matrix of zeros
A = np.zeros((M,N))
```

5. Creating a counter function

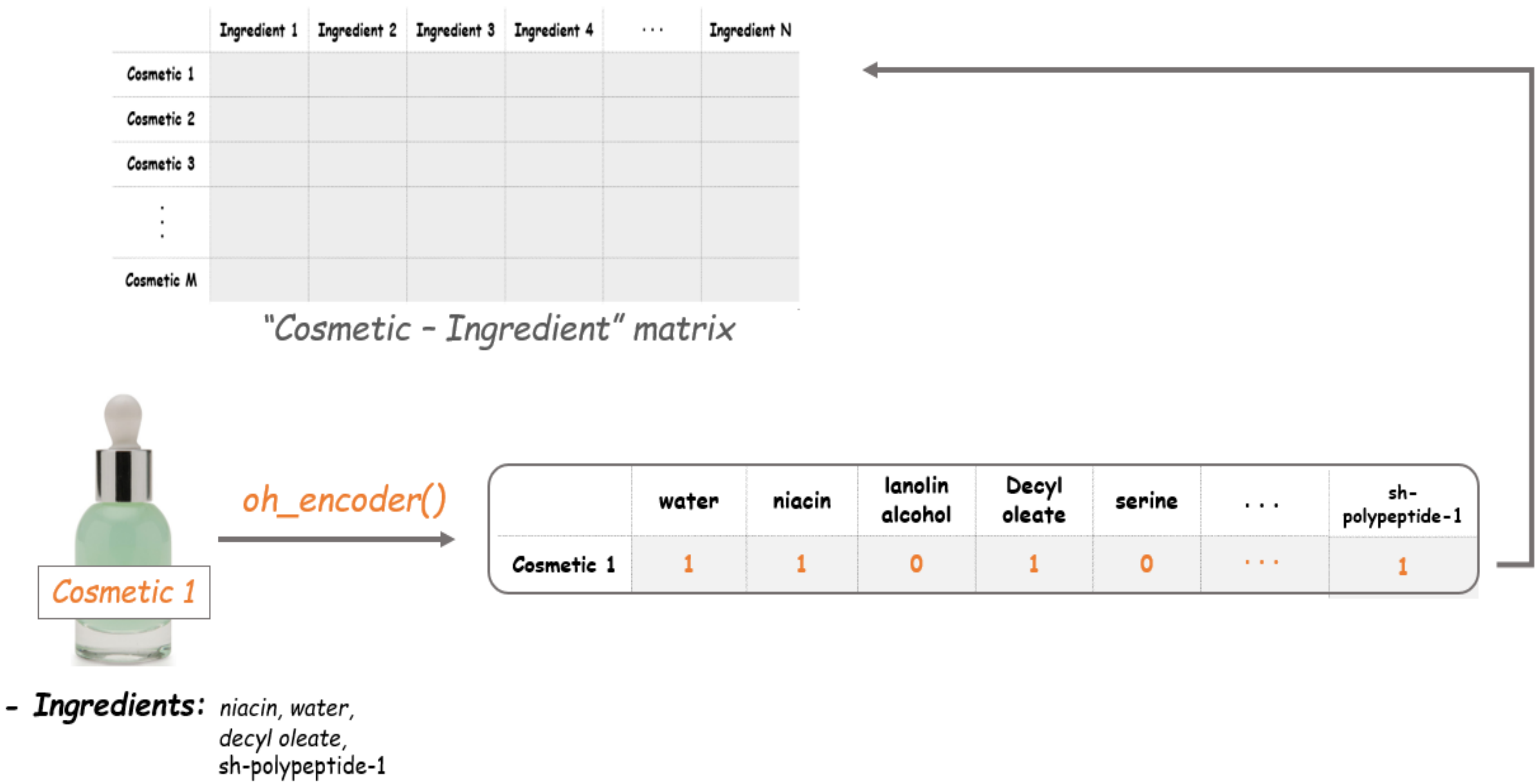
Before we can fill the matrix, let's create a function to count the tokens (i.e., an ingredients list) for each row. Our end goal is to fill the matrix with 1 or 0: if an ingredient is in a cosmetic, the value is 1. If not, it remains 0. The name of this function, `oh_encoder`, will become clear next.

In [5]:

```
# Define the oh_encoder function
def oh_encoder(tokens):
    x = np.zeros(N)
    for ingredient in tokens:
        # Get the index for each ingredient
        idx = ingredient_idx[ingredient]
        # Put 1 at the corresponding indices
        x[idx] = 1
    return x
```

6. The Cosmetic-Ingredient matrix!

Now we'll apply the `oh_encoder()` functon 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.

In [6]:
```
# Make a document-term matrix
i = 0
for tokens in corpus:
    A[i, :] = oh_encoder(tokens)
    i += 1
```

7. Dimension reduction with t-SNE

The dimensions of the existing matrix is (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 vectorizing. 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.

In [7]:
```
# Dimension reduction with t-SNE
model = TSNE(n_components=2, learning_rate=200, random_state=42)
tsne_features = model.fit_transform(A)

# Make X, Y columns
moisturizers_dry['X'] = tsne_features[: ,0]
moisturizers_dry['Y'] = tsne_features[: ,1]
print(moisturizers_dry['X'], moisturizers_dry['Y'])
```

```
0     -0.072740
1      0.242541
2     -0.432800
3     -0.883814
4     -3.103213
5     -3.462883
6     -0.293262
7     -1.828106
8     -0.322829
9     -0.306376
10    -1.715162
11    -3.614759
12    -2.425915
13     0.722821
14     0.552942
15    -0.074804
16    -0.101659
17     1.466658
18    -0.288054
19     0.859572
20     0.428914
21     1.071283
22     1.450490
23    -0.295611
24    -0.296942
25    -3.091204
26    -2.760984
27    -2.859019
28     0.241052
29     2.388048
         ...
160   -1.102251
161   -3.437738
162    0.156861
163    3.170805
164   -0.292089
165   -1.547426
166   -1.892699
167   -0.708245
168   -1.123780
169   -0.291880
```

```
170   -0.291021
171   -0.291711
172    0.505489
173    0.242755
174    1.589120
175    0.469557
176   -0.281507
177   -0.802881
178   -0.286835
179    0.120966
180   -0.287903
181   -3.286143
182   -0.320299
183   -0.395039
184    1.606050
185   -1.096746
186    1.331309
187   -1.872760
188   -1.536723
189   -0.014644
Name: X, Length: 190, dtype: float32 0      3.230559
1     -1.743190
2     -1.650751
3      3.570323
4      2.308803
5     -0.086115
6     -0.290012
7      1.416506
8      0.624228
9      1.159293
10    -1.025219
11    -0.083801
12    -0.410152
13     1.829232
14    -2.138117
15     3.228830
16    -2.142008
17    -0.285083
18    -0.286764
19    -1.662301
20     0.346917
21    -3.537667
```

```
22     1.296166
23    -0.287168
24    -0.286210
25     2.308036
26     0.629061
27     0.964677
28    -1.743162
29     1.446533
         ...
160    0.440849
161   -1.352856
162    0.887035
163   -2.248709
164   -0.288890
165   -0.404199
166    0.049710
167   -1.799106
168    1.496665
169   -0.119543
170   -0.285994
171   -0.289339
172    0.248210
173   -1.242982
174    2.724004
175   -0.547537
176   -0.289492
177    0.100561
178   -0.288532
179   -1.216989
180   -0.289935
181   -1.285314
182   -0.287969
183   -3.111079
184   -0.621609
185    0.652184
186   -1.595261
187   -0.372652
188   -0.363564
189    1.110812
Name: Y, Length: 190, dtype: float32
```

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. Note that we won't display the plot yet as we will make some more additions to it.

In [8]:
```
from bokeh.io import show, output_notebook, push_notebook
from bokeh.plotting import figure
from bokeh.models import ColumnDataSource, HoverTool
output_notebook()

# Make a source and a scatter plot
source = ColumnDataSource(moisturizers_dry)
plot = figure(x_axis_label = 'T-SNE 1',
              y_axis_label = 'T-SNE 2',
              width = 500, height = 400)
plot.circle(x = 'X',
    y = 'Y',
    source = source,
    size = 10, color = '#FF7373', alpha = .8)
```

(https://bokeh.pydata.org) BokehJS 1.2.0 successfully loaded.

Out[8]: **GlyphRenderer**(id = '1038', …)

9. Adding a hover tool

Why don't we add a hover tool? 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).

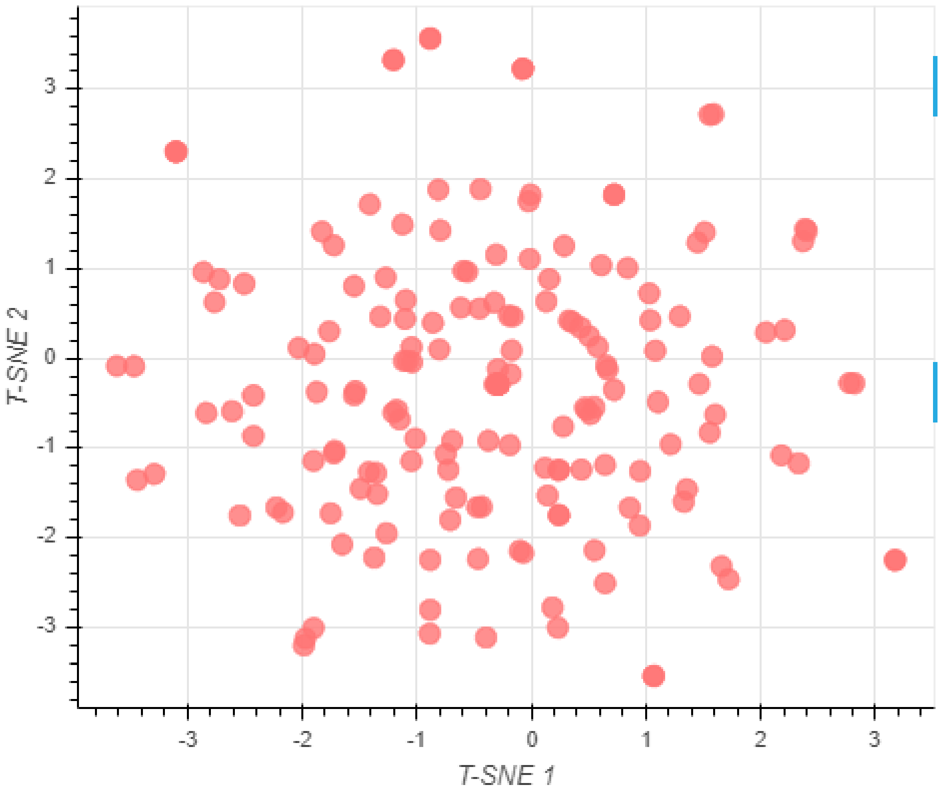
In [9]:
```
# Create a HoverTool object
hover = HoverTool(tooltips = [('Item', '@Name'),
                              ('Brand', '@Brand'),
                              ('Price', '$@Price'),
                              ('Rank', '@Rank')])
plot.add_tools(hover)
```

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.

In [12]:
```
# Plot the map

show(plot)
```



(http

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.

In [11]:
```
# Print the ingredients of two similar cosmetics
cosmetic_1 = moisturizers_dry[moisturizers_dry['Name'] == "Color Control Cushion Compact Broad Spectrum SPF 50+"]
cosmetic_2 = moisturizers_dry[moisturizers_dry['Name'] == "BB Cushion Hydra Radiance SPF 50"]

# Display each item's data and ingredients
display(cosmetic_1)
print(cosmetic_1.Ingredients.values)
display(cosmetic_2)
print(cosmetic_2.Ingredients.values)
```

	Label	Brand	Name	Price	Rank	Ingredients	Combination	Dry	Normal	Oily	Sensitive	X	Y
45	Moisturizer	AMOREPACIFIC	Color Control Cushion Compact Broad Spectrum S...	60	4.0	Phyllostachis Bambusoides Juice, Cyclopentasil...	1	1	1	1	1	2.775364	-0.274434

```
['Phyllostachis Bambusoides Juice, Cyclopentasiloxane, Cyclohexasiloxane, Peg-10 Dimethicone, Phenyl Trimethicone, Bu
tylene Glycol, Butylene Glycol Dicaprylate/Dicaprate, Alcohol, Arbutin, Lauryl Peg-9 Polydimethylsiloxyethyl Dimethic
one, Acrylates/Ethylhexyl Acrylate/Dimethicone Methacrylate Copolymer, Polyhydroxystearic Acid, Sodium Chloride, Poly
methyl Methacrylate, Aluminium Hydroxide, Stearic Acid, Disteardimonium Hectorite, Triethoxycaprylylsilane, Ethylhexy
l Palmitate, Lecithin, Isostearic Acid, Isopropyl Palmitate, Phenoxyethanol, Polyglyceryl-3 Polyricinoleate, Acrylate
s/Stearyl Acrylate/Dimethicone Methacrylate Copolymer, Dimethicone, Disodium Edta, Trimethylsiloxysilicate, Ethylhexy
glycerin, Dimethicone/Vinyl Dimethicone Crosspolymer, Water, Silica, Camellia Japonica Seed Oil, Camillia Sinensis Le
af Extract, Caprylyl Glycol, 1,2-Hexanediol, Fragrance, Titanium Dioxide, Iron Oxides (Ci 77492, Ci 77491, Ci7749
9).']
```

	Label	Brand	Name	Price	Rank	Ingredients	Combination	Dry	Normal	Oily	Sensitive	X	Y
55	Moisturizer	LANEIGE	BB Cushion Hydra Radiance SPF 50	38	4.3	Water, Cyclopentasiloxane, Zinc Oxide (CI 7794...	1	1	1	1	1	2.814905	-0.277909

```
['Water, Cyclopentasiloxane, Zinc Oxide (CI 77947), Ethylhexyl Methoxycinnamate, PEG-10 Dimethicone, Cyclohexasiloxan
e, Phenyl Trimethicone, Iron Oxides (CI 77492), Butylene Glycol Dicaprylate/Dicaprate, Niacinamide, Lauryl PEG-9 Poly
dimethylsiloxyethyl Dimethicone, Acrylates/Ethylhexyl Acrylate/Dimethicone Methacrylate Copolymer, Titanium Dioxide
(CI 77891 , Iron Oxides (CI 77491), Butylene Glycol, Sodium Chloride, Iron Oxides (CI 77499), Aluminum Hydroxide, HD
I/Trimethylol Hexyllactone Crosspolymer, Stearic Acid, Methyl Methacrylate Crosspolymer, Triethoxycaprylylsilane, Phe
noxyethanol, Fragrance, Disteardimonium Hectorite, Caprylyl Glycol, Yeast Extract, Acrylates/Stearyl Acrylate/Dimethi
cone Methacrylate Copolymer, Dimethicone, Trimethylsiloxysilicate, Polysorbate 80, Disodium EDTA, Hydrogenated Lecith
in, Dimethicone/Vinyl Dimethicone Crosspolymer, Mica (CI 77019), Silica, 1,2-Hexanediol, Polypropylsilsesquioxane, Ch
enopodium Quinoa Seed Extract, Magnesium Sulfate, Calcium Chloride, Camellia Sinensis Leaf Extract, Manganese Sulfat
e, Zinc Sulfate, Ascorbyl Glucoside.']
```

In [ ]:

In [ ]: