Data Mining Project on Sephora e-commerce data
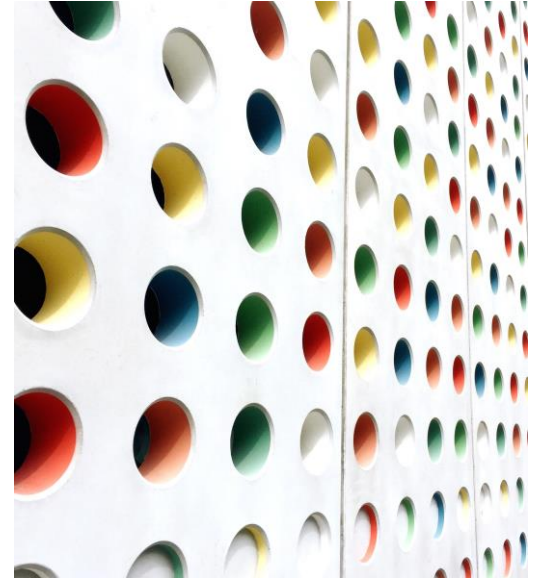
MUHAMMAD AULIA RENDY

USING PYTHON PROGRAMMING

About

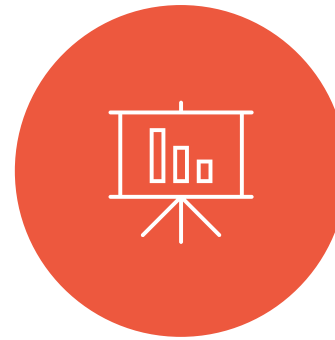This project was carried out with the aim of gaining business insights based on data from Sephora's e-commerce.

Outline

1. BUSINESS QUESTION

2. EXPLORATORY DATA ANALYSIS

3. GAINING INSIGHT

4. RECOMMENDATION

SEPHORA

Business question

What brands and product categories have the most transactions sold online?

How is the level of comparison between the number of transactions in the product category that is sold online and not?

Does the higher the product category with a high rating, the effect of a high number of transactions also through online sales?

Does the higher the number of product categories that visitors like, the effect of a high number of transactions also through online sales and how is the conversion rate?

Exploratory data analysis

Dataset Introduction

This dataset consists of 21 columns 9168 rows.

column	information
Id	The product ID at Sephora's website
brand	The brand of the product at Sephora's website
category	The category of the product at Sephora's website
name	The name of the product at Sephora's website
size	The size of the product
rating	The rating of the product
numberofreviews	The number of reviews of the product
love	The number of people loving the product
price	The price of the product
value_price	The value price of the product (for discounted products
URL	The URL link of the product
MarketingFlags	The Marketing Flags of the product from the website if they were exclusive or sold online only
MarketingFlags_content	The kinds of Marketing Flags of the product
options	The options available on the website for the product like colors and sizes
details	The details of the product available on the website
howtouse	The instructions of the product if available
ingredients	The ingredients of the product if available
online_only	If the product is sold online only
exclusive	If the product is sold exclusively on Sephora's website
limited_edition	If the product is limited edition
limitedtimeoffer	If the product has a limited time offer

SEPHORA

EDA (Missing value and duplicate check)

```
1 df.isnull().sum()
```

```
id                      0
brand                   0
category                0
name                    0
size                    0
rating                  0
number_of_reviews       0
love                    0
price                   0
value_price             0
URL                     0
MarketingFlags          0
MarketingFlags_content  0
options                 0
details                 0
how_to_use              0
ingredients             0
online_only             0
exclusive               0
limited_edition         0
limited_time_offer      0
dtype: int64
```

There are no missing value was found from every rows in this dataset

```
1 print(f'dataset dimension before duplication check: {df_sephora.shape[0]}')
2 df_sephora= df_sephora.drop_duplicates()
3 print(f'dataset dimension after duplication handling: {df_sephora.shape[0]}')
```

```
dataset dimension before duplication check: 9168
dataset dimension after duplication handling: 9168
```

There are no data duplication was found in this dataset
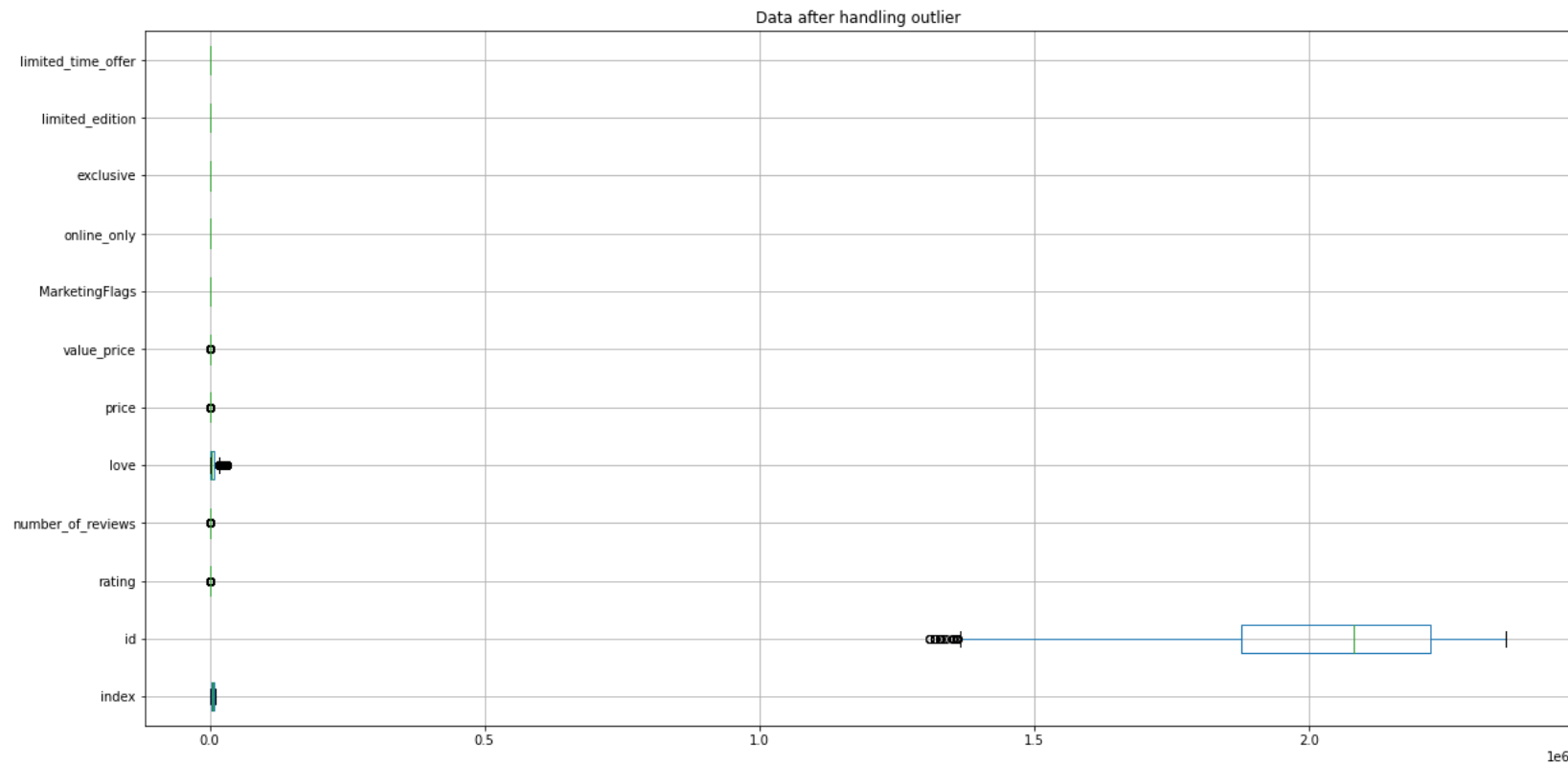
SEPHORA

EDA (Outlier check)

Data outlier checking
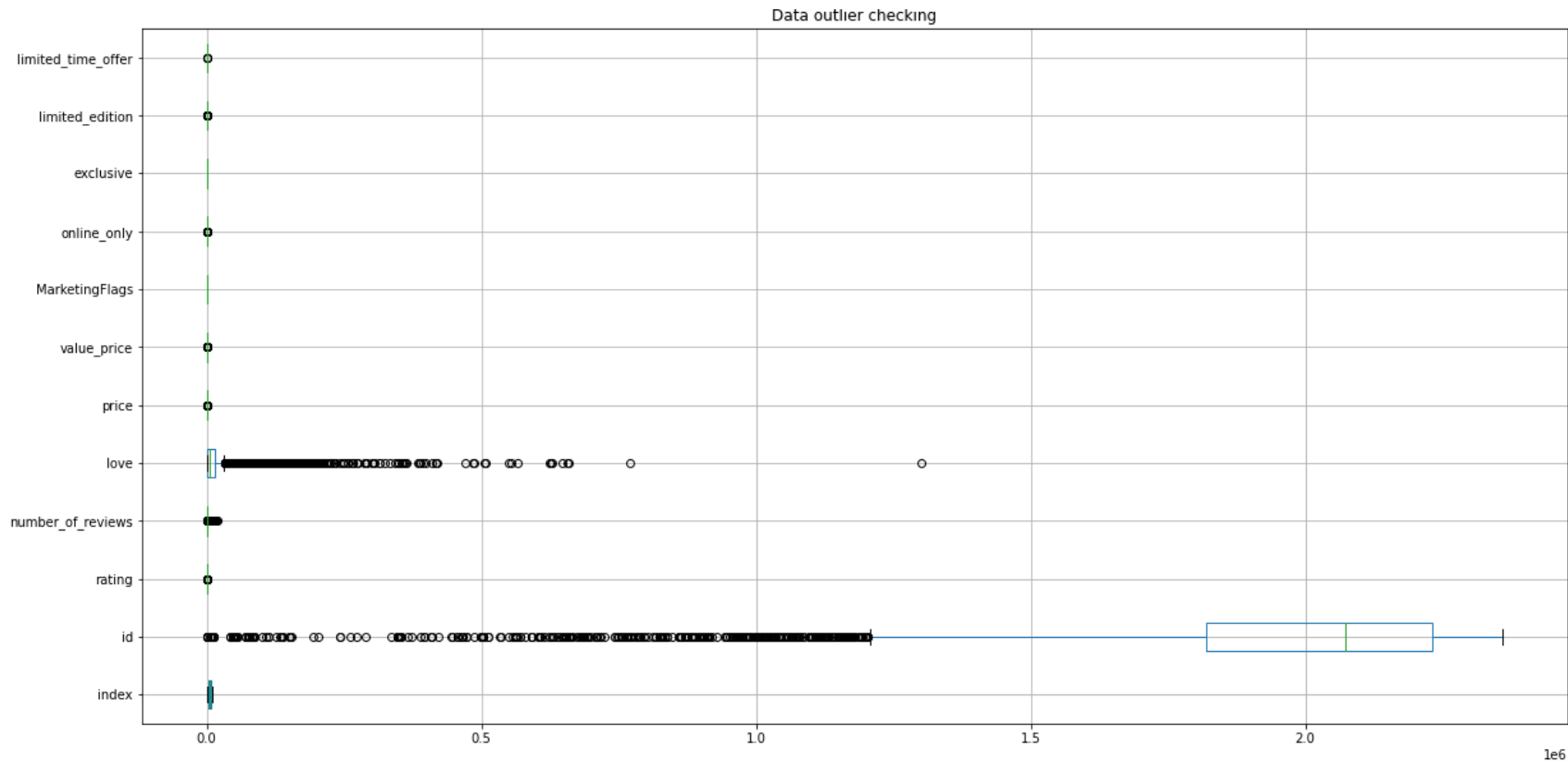
There are outliers with very extreme values in each column

Data that has extreme values can provide biased information, so it is necessary to deal with these values

SEPHORA

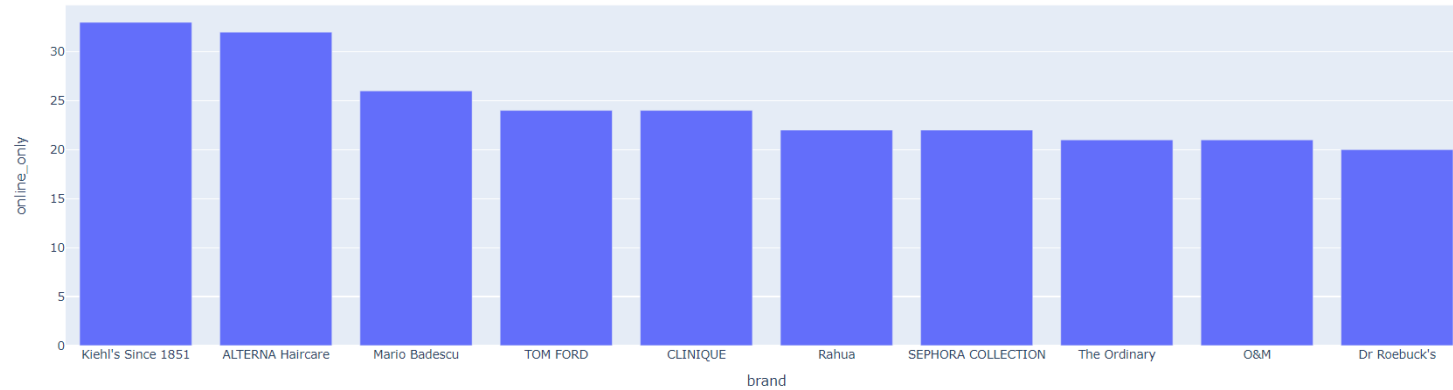# EDA (After handling Outlier)

Data after handling outlier

Handling outlier data is done by using the IQR method.

As a result, the data of extreme value has been greatly reduced when compared to the previous data.

SEPHORA

Gaining insight from business question

SEPHORA

Brand with the most sales

Category with the most sales

It can be seen that the top 3 brands have the highest transactions based on those sold on the Sephora website.

However, Sephora which also has a brand is not the best selling brand in their own website.

The graph to the side shows the product categories that have the highest deals sold on the website. it can be seen that face serums, moisturizers and candles and home scents are in the top 3 highest transactions

What brands and product categories have the most transactions sold online?

SEPHORA

# How is the level of comparison between the number of transactions in the product category that is sold online or not?

Percentage sold by online and offline

■ 0
■ 1

23.5%

76.5%
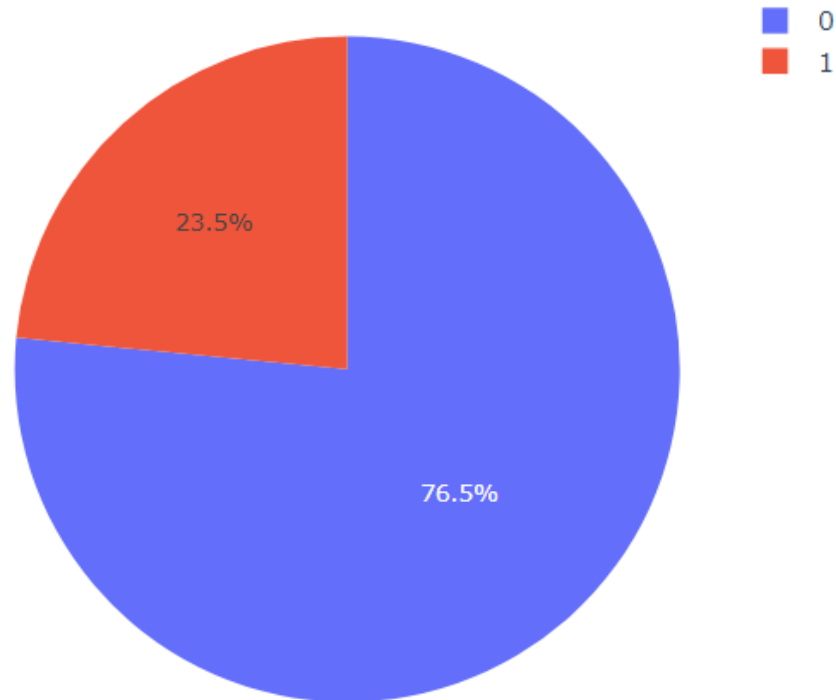
Based on the graph beside, it can be seen that the number of online transactions is still low when compared to non-online transactions that occur at Sephora

SEPHORA

Does the higher the product category with a high rating, affect of a high number of transactions also through online sales?

category	rating	online_only
Face Serums	4.5	29
Moisturizers	4.0	27
Face Wash & Cleansers	4.5	21
Hair Styling Products	4.0	19
Face Masks	4.5	19
Shampoo	4.0	18
Hair Styling Products	4.5	18
Conditioner	4.0	18
Shampoo	3.5	17
Perfume	4.5	17
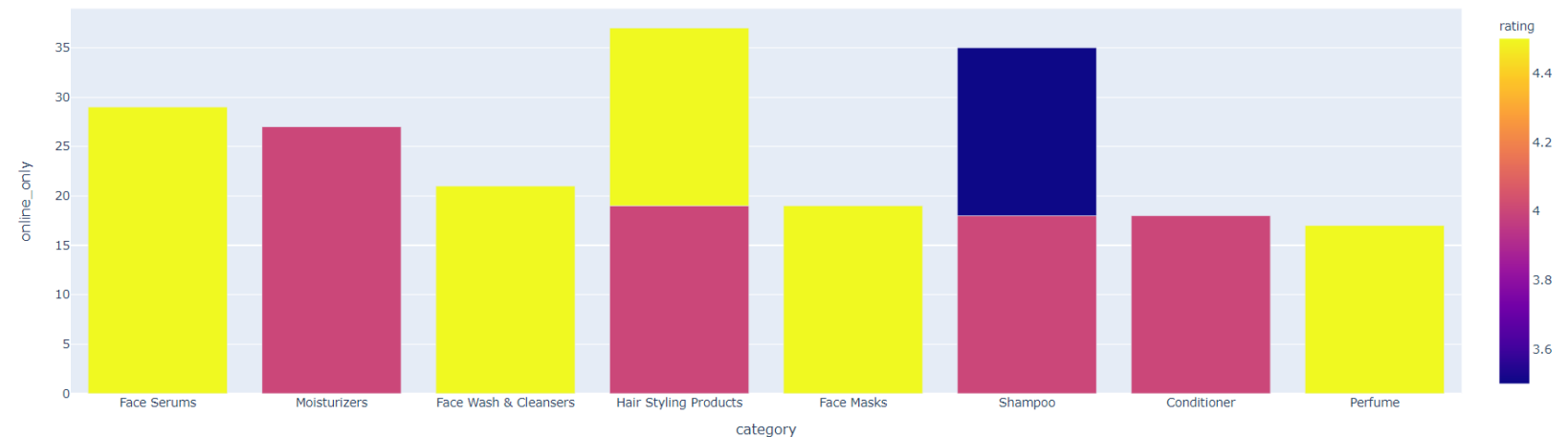
Based on the graph above, it can be seen that when a product has a high rating, it can affect the number of transactions sold. So, it can be concluded that the high rating of a product has a high correlation with the occurrence of sales transactions through online. So, if many products are given a low rating, it can have an effect on sales transactions.

SEPHORA

Does the higher the number of product categories that visitors like, affect of a high number of transactions also through online sales and how is the conversion rate?

category	online_only	love
Face Serums	66	385609
Moisturizers	61	151896
Candles & Home Scents	57	55920
Hair Styling Products	53	114960
Shampoo	52	107209
Face Wash & Cleansers	45	209121
Conditioner	41	57259
Beauty Supplements	37	67433
Body Lotions & Body Oils	35	129461
Face Masks	34	224311
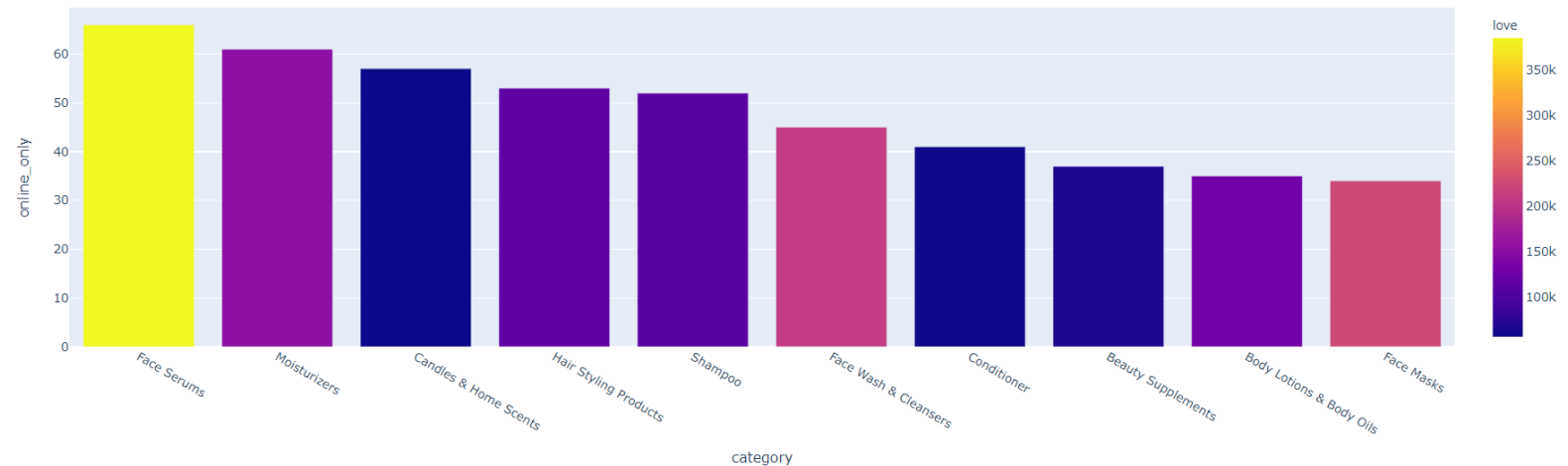
Based on the graph above, it can be seen that product categories with high transactions do not always have a high number of likes. So, the number of products that have high likes does not always trigger online sales transactions.

SEPHORA

how about the conversion rate?

category	online_only	love	conversion_rate
Face Serums	66	385609	0.017116
Moisturizers	61	151896	0.040159
Candles & Home Scents	57	55920	0.101931
Hair Styling Products	53	114960	0.046103
Shampoo	52	107209	0.048503
Face Wash & Cleansers	45	209121	0.021519
Conditioner	41	57259	0.071604
Beauty Supplements	37	67433	0.054869
Body Lotions & Body Oils	35	129461	0.027035
Face Masks	34	224311	0.015158

Product category conversion rate based on transactions sold against product category likes
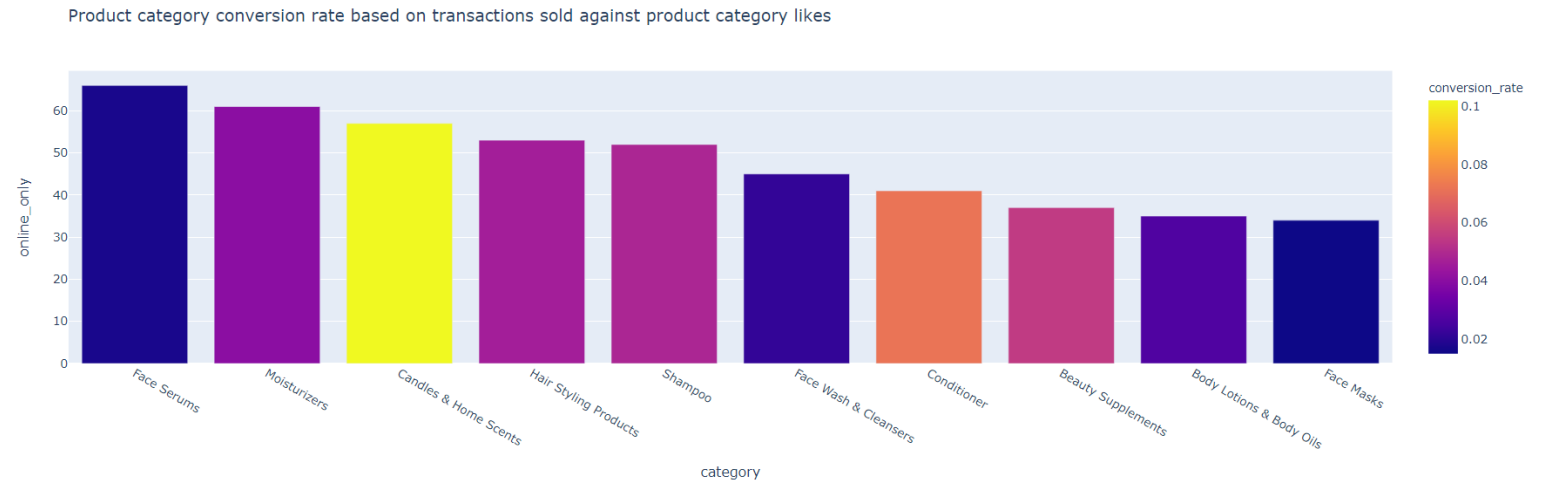
Based on the graph above, it can be seen that product categories that have a high number of transactions sold have a low conversion rate based on the number of likes products. This means that the number of products that are liked, does not trigger sales transactions

SEPHORA

Recommendation

1. The number of online sales transactions is still low at 23%. To overcome this, we recommend companies to increase product(website) awareness and give promo for online transactions, so that companies can bring in more new customers and increase online transactions.

2. Regarding the sephora brand that does not have the highest sales transaction, we recommend that the company should further increase sales of the sephora brand. So that way, Sephora is not too dependent on other brands, and can reduce sales costs if the Sephora brand contributes the most to sales.

Recommendation

3. The top 10 product categories with the highest sales transactions still have a low conversion rate of <1% based on the number of likes of a product. This indicates that the desire of a customer to make a purchase is still low. To overcome this, we recommend companies to increase product awareness and provide attractive promos, so that customers not only give likes to a product, but also convert them into purchases.

SEPHORA