

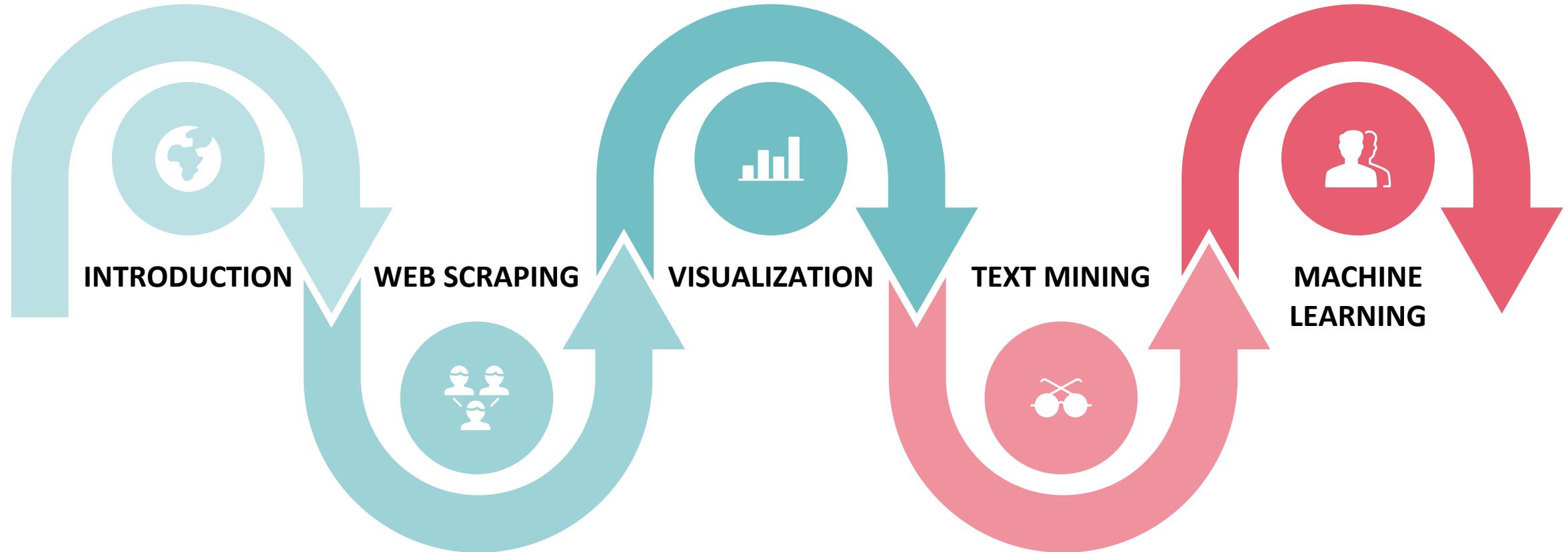
SECRET of SEPHORA



Shanghai Spicy Chicken Group

Chuwei Lyu
Xinyi Shu
Shuran Wang
Yijing Zhou

OVERVIEW



INTRODUCTION

- Holiday season motivates shopping
- 10,000+ products on Sephora.com. Hard to find ideal products

1. MOTIVATION

2. What is ALLURE?

Allure Best of Beauty Awards include each year's most amazing products in different categories

3. BUSINESS VALUE

- Provides insights for future allure list
- Provides men's product selection suggestions



Allure includes each year's most amazing products and brands

Search

SEPHORA

Hi, Shuran ▾
4,484 points • ROUGE

SHOP NEW BRANDS GIFTS SALE COMMUNITY HAPPENING IN STORE

MAKEUP SKINCARE HAIR TOOLS & BRUSHES FRAGRANCE

Just Arrived Bestsellers Value & Gift Sets Mini Size Makeup Palettes Clean Makeup

Eye Eye Palettes Mascara Eyeliner Eyebrow Eyeshadow Eye Primer Under-Eye Concealer False Eyelashes

Cheek Blush Bronzer Cheek Palettes

Brushes & Applicators Face Brushes Eye Brushes Lip Brushes

Featured Foundation Quiz Mascara Quiz Lip Quiz Clean Beauty Allure Best of Beauty

Allure Best of Beauty

DRUNK ELEPHANT Protini™ Polypeptide Moisturizer \$68.00 ★★★★☆

LANEIGE Lip Sleeping Mask \$20.00 ★★★★★

FOREO LUNA? mini 2 \$139.00 ★★★★★

KIEHL'S SINCE 1851 Ultra Facial Cream \$30.00 ★★★★★

A red arrow points from the "Allure Best of Beauty" section in the sidebar to the "Allure Best of Beauty" section in the "HAPPENING IN STORE" sidebar.

WEB SCRAPING

duct/all-nighter-long-lasting-makeup-setting-spray/p263504?cid2=products%20grid:p263504

The screenshot shows a product page on the Sephora website. At the top, there's a search bar, the Sephora logo, and a sign-in link. Below the header, a navigation menu includes categories like SHOP, NEW, BRANDS, GIFTS, SALE, COMMUNITY, and HAPPENING IN STORE. A breadcrumb trail shows the user has navigated from Makeup to Face to Setting Spray & Powder. The main product image is a black spray bottle labeled 'all nighter'. To its right, product details are displayed: brand (URBAN DECAY), name (All Nighter Long-Lasting Makeup Setting Spray), price (\$33.00), and a call-to-action button (ADD TO BASKET). A dropdown menu shows quantity (1) and a promotional offer (SPEND \$50 FOR FREE SHIPPING). The item number (ITEM 1900000) is also present. Below the product image, there are reviews: a 5-star rating with 9K reviews and 490K loves. The page also lists sizes: Standard Size - 4 oz / 118 mL, Value size (2 x 4 oz / 118 mL), and Standard size: Natural finish. A zoomed-in view of the product is shown with three small images: a close-up of the bottle, a video thumbnail, and a hand holding the bottle.

SEPHORA

Hi, Beautiful ▾
Sign In or Register

SHOP NEW BRANDS GIFTS SALE COMMUNITY HAPPENING IN STORE

Makeup > Face > Setting Spray & Powder

URBAN DECAY
All Nighter Long-Lasting Makeup Setting Spray

\$33.00 SPEND \$50 FOR FREE SHIPPING

1 ADD TO BASKET

ITEM 1900000

★★★★★ 9K reviews | ❤ 490K loves

SIZE: Standard Size - 4 oz / 118 mL

Value size

2 x 4 oz / 118 mL

Standard size: Natural finish

Standard Size - 4 oz / 118 mL

Mini size: Natural finish

Mini Size - 1 oz / 30 mL

Roll over or click image to zoom in

allure BEST OF BEAUTY AWARD WINNER 2019

9572 reviews

Ratings & Reviews

Write a review

5 stars 6638

4 stars 1579

3 stars 521

2 stars 361

1 star 473

4.4 / 5 stars

Show reviews from my Beauty Matches ?

Filter ▾ Sort by: Most Helpful ▾

colorsmaela

★★★★★

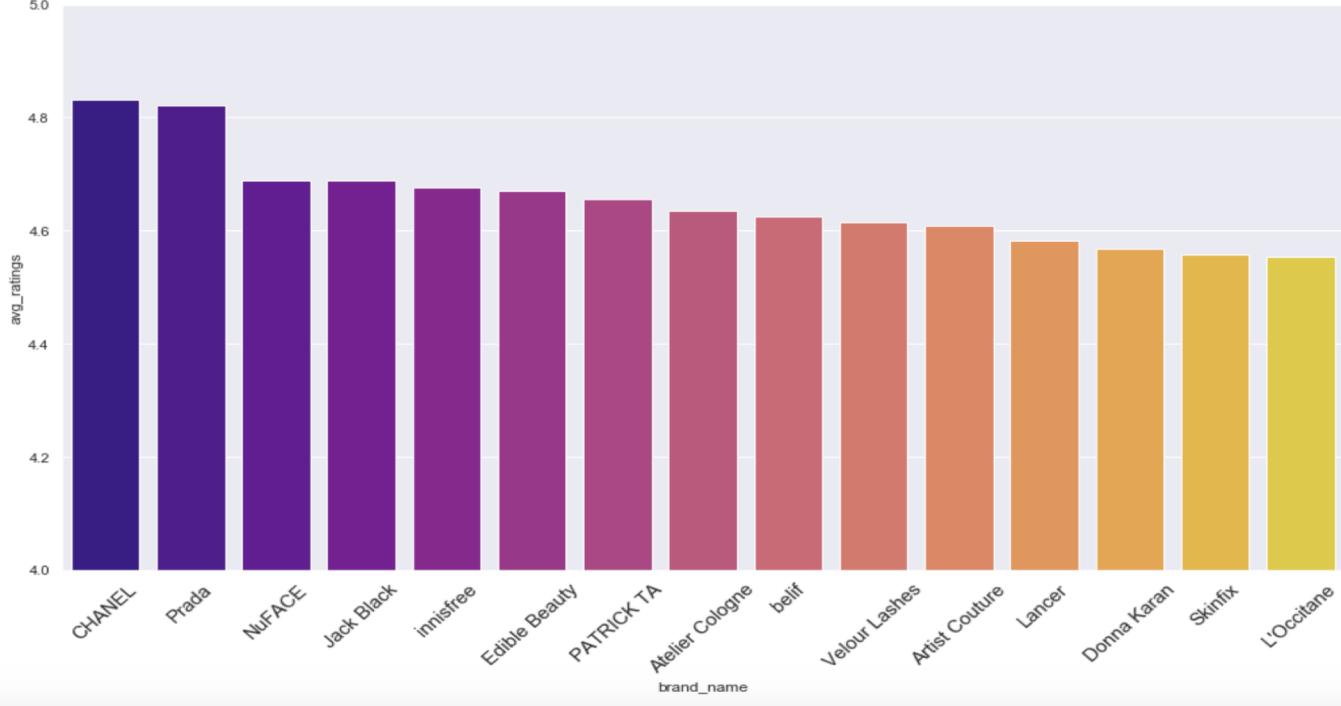
DONT SLEEP ON THIS

THIS SURVIVED ALL OF MY EMOTIONAL THERAPY SESSIONS. LASTED THROUGH DADDY ISSUES, LOW SELF ESTEEM AND ALL OTHER CRYING FACTORS. BAWLED MY EYES OUT BUT MY FACE STILL BEAT, HIGHLIGHT STILL POPPIN, CONTOUR STILL SHARP AS A KNIFE. 10/10 WOULD RECOMMEND.

NOT HELPFUL (77) HELPFUL (1670)

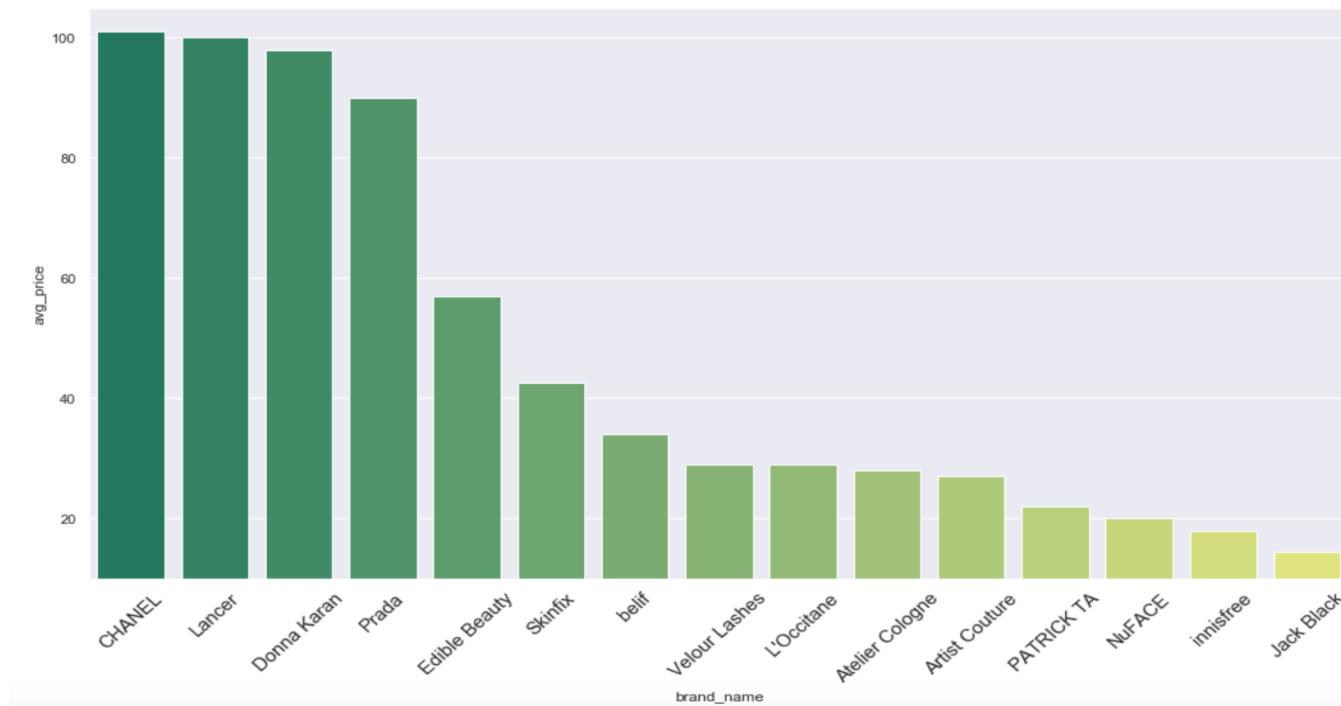


VISUALIZATION



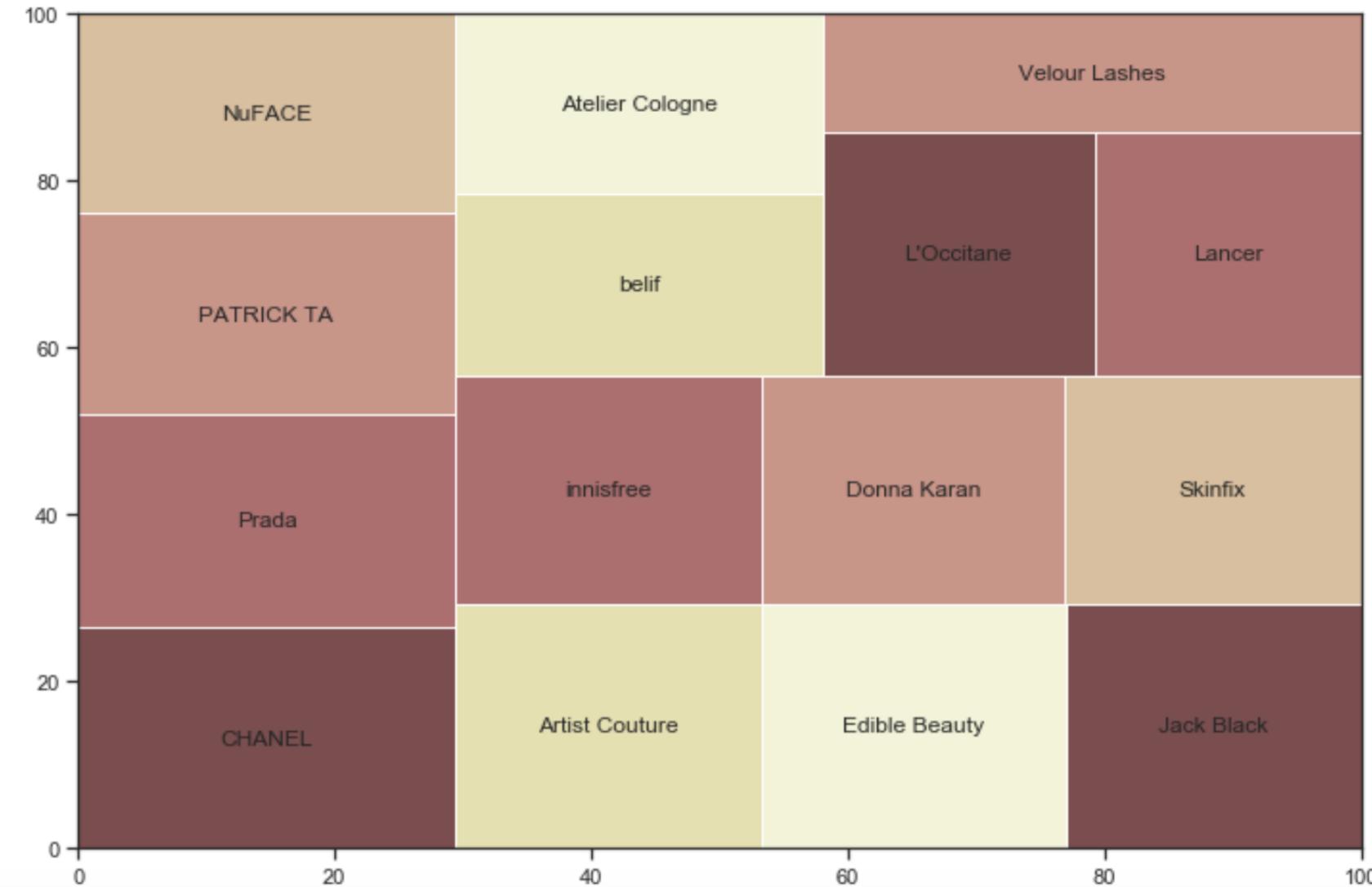
After remove products with less than 100 reviews:

15 Brands with highest average ratings



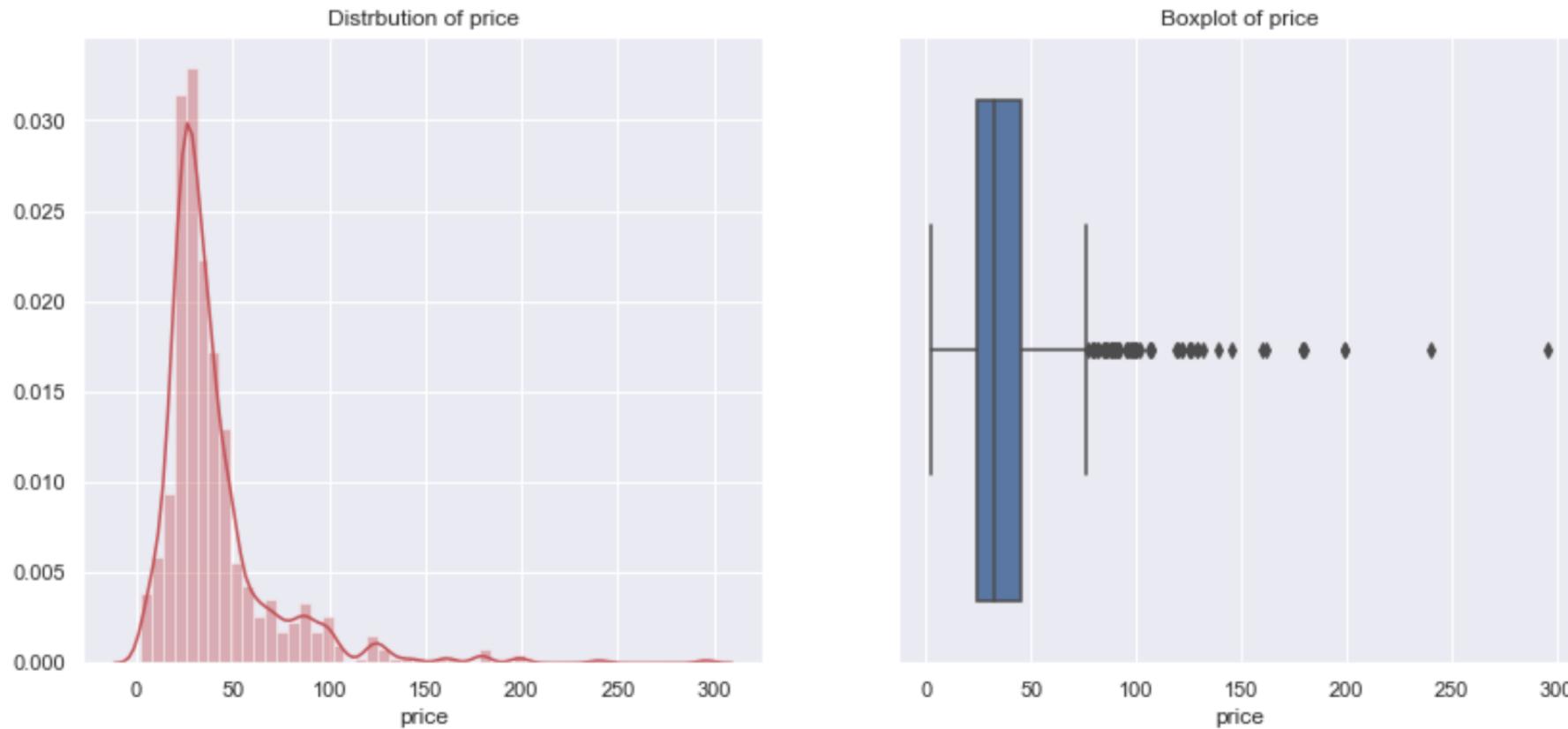
Among those 15 Brands, plot their average prices

VISUALIZATION



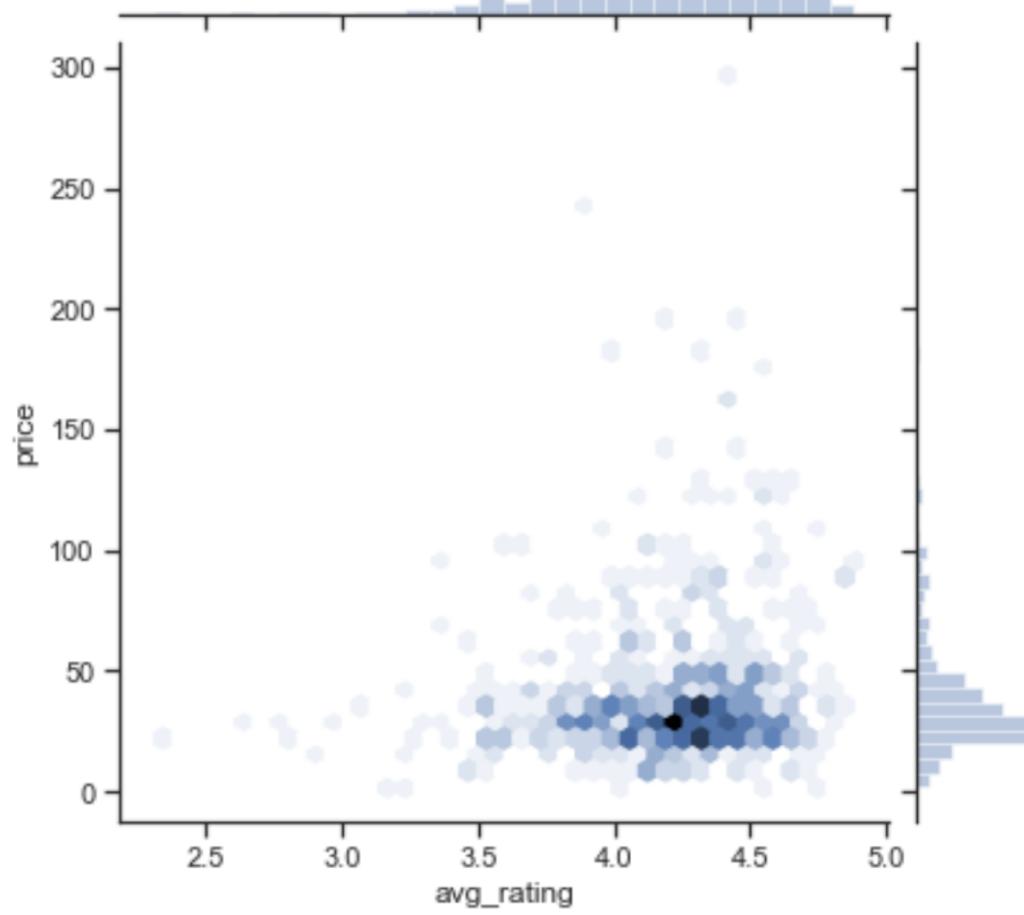
Top 15 brands with their corresponding 5-star percentages

Price Distribution of all the Products

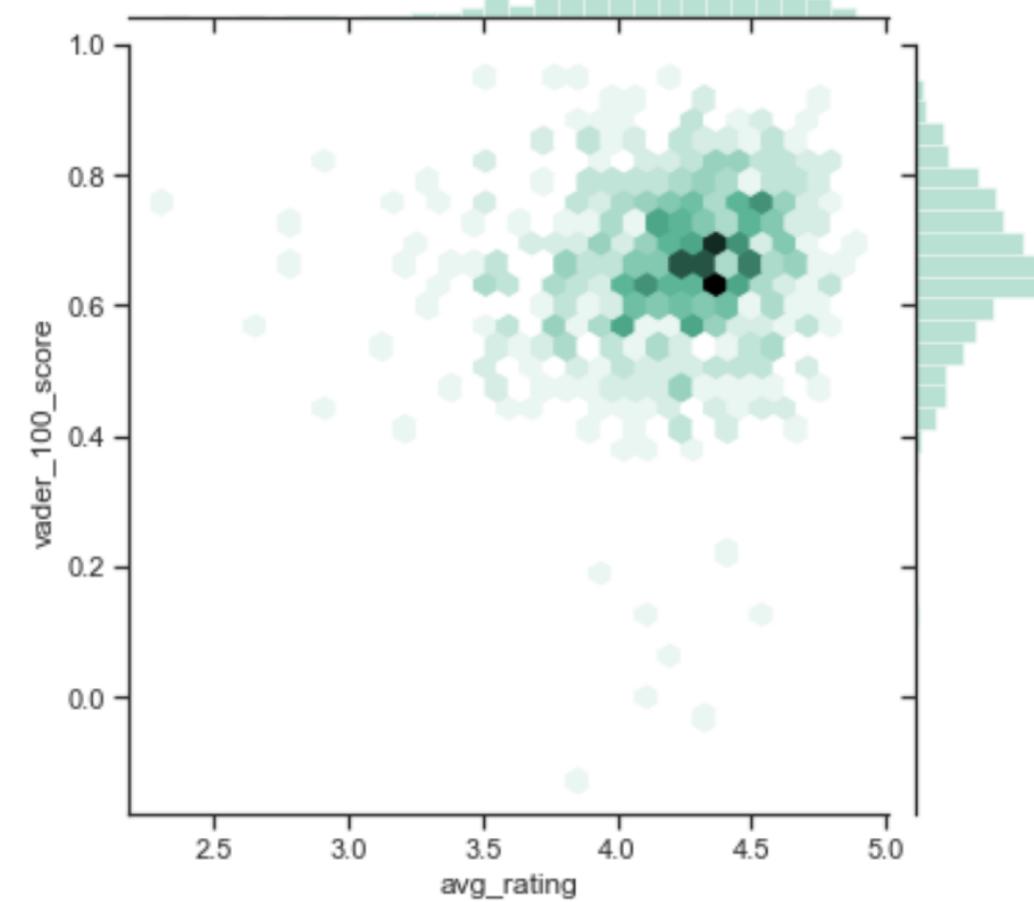


- Mean price: around 35 dollars
- Lots of outliers

VISUALIZATION

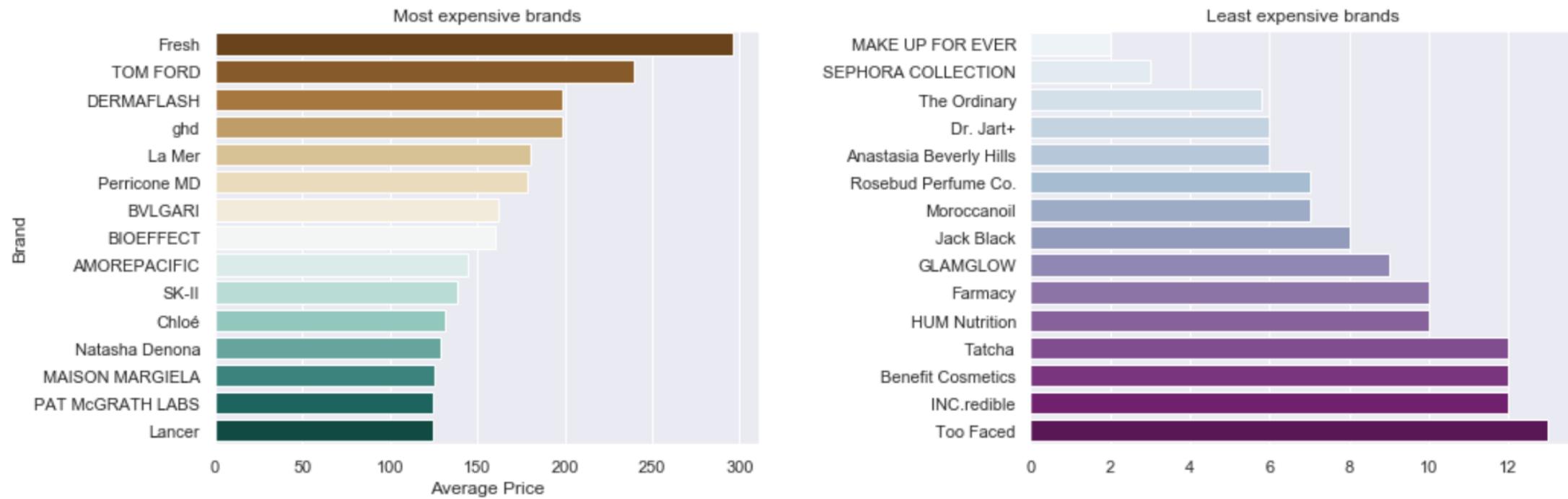


Relation between price and average ratings



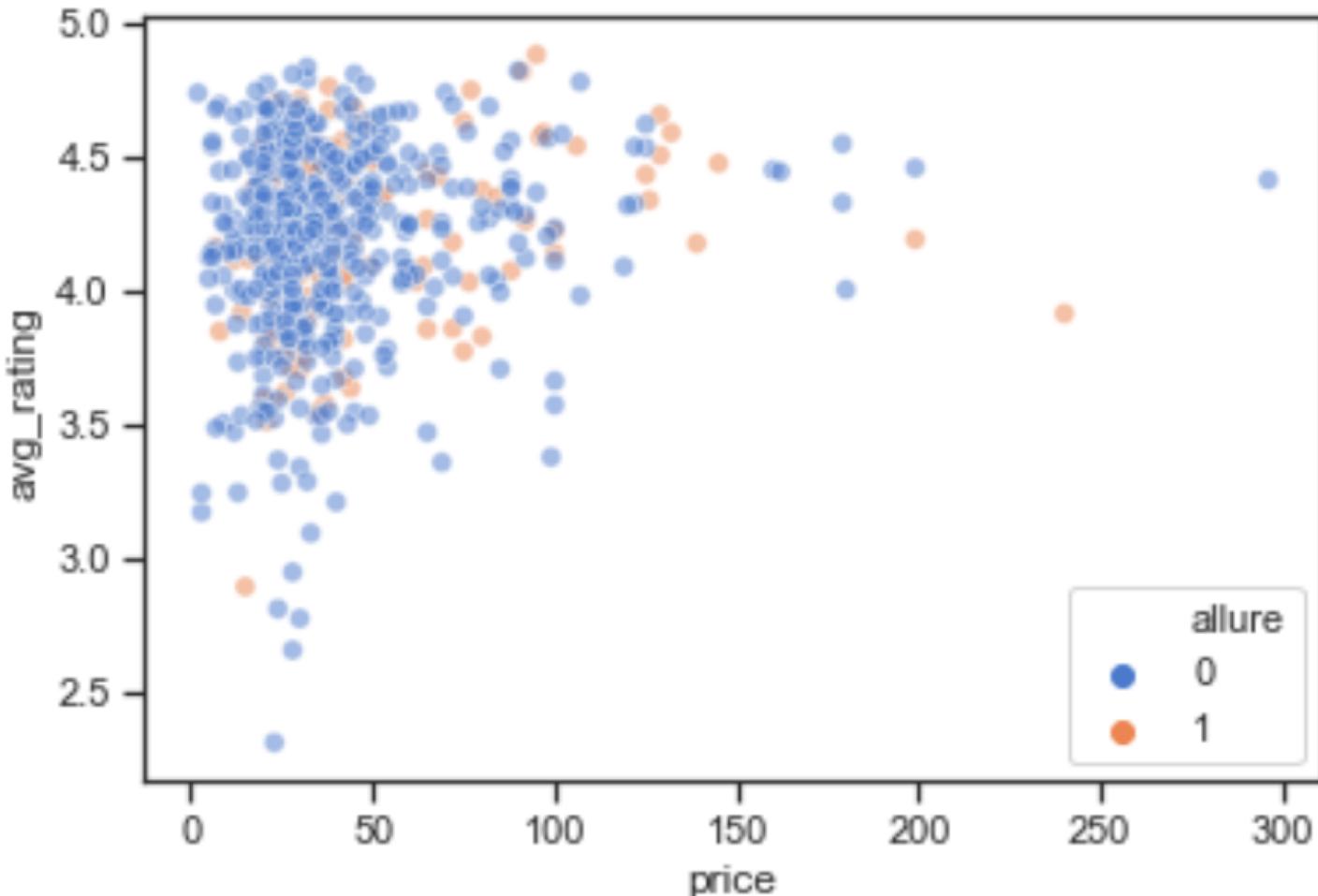
Relation between Vader Score and average ratings

VISUALIZATION



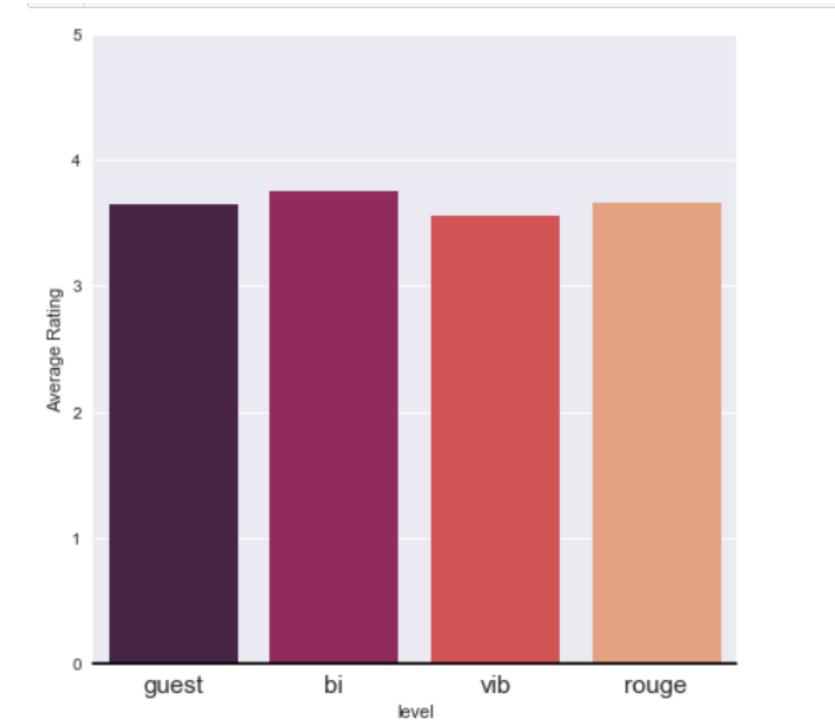
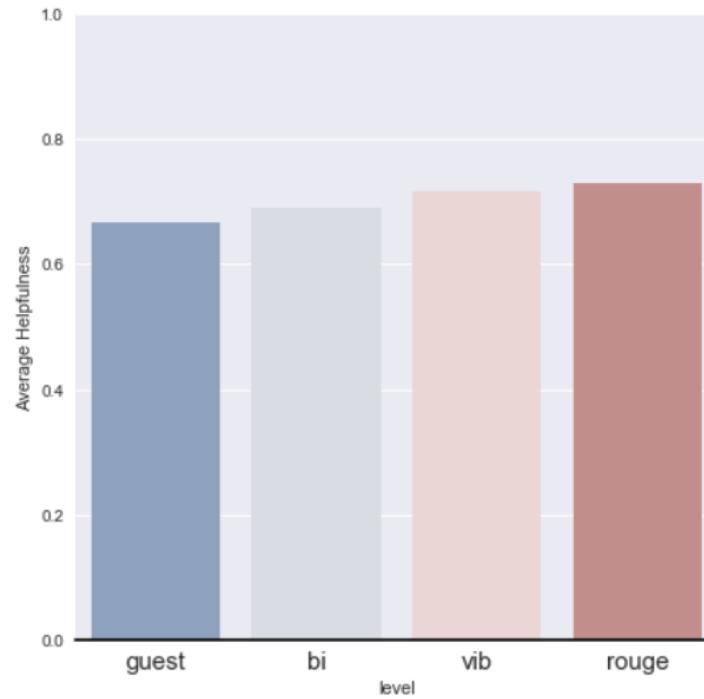
Most Expensive and least expensive brands

VISUALIZATION



Rating & Price Comparison for Both Allure and Non-Allure Products

VISUALIZATION



Different Membership with their Helpfulness & Ratings

TEXT MINING

Sentiment Analysis:

- Calculate Vader score for each product
- Used compound score, range from -1 to +1
- Represents the probability of positive sentiment in text

0.902161

'Pleasantly Surprised'

"This set is amazing! A fantastic formula that isn't sticky at all, yet lasts a good length of time.

-0.223200

'Poor Quality'

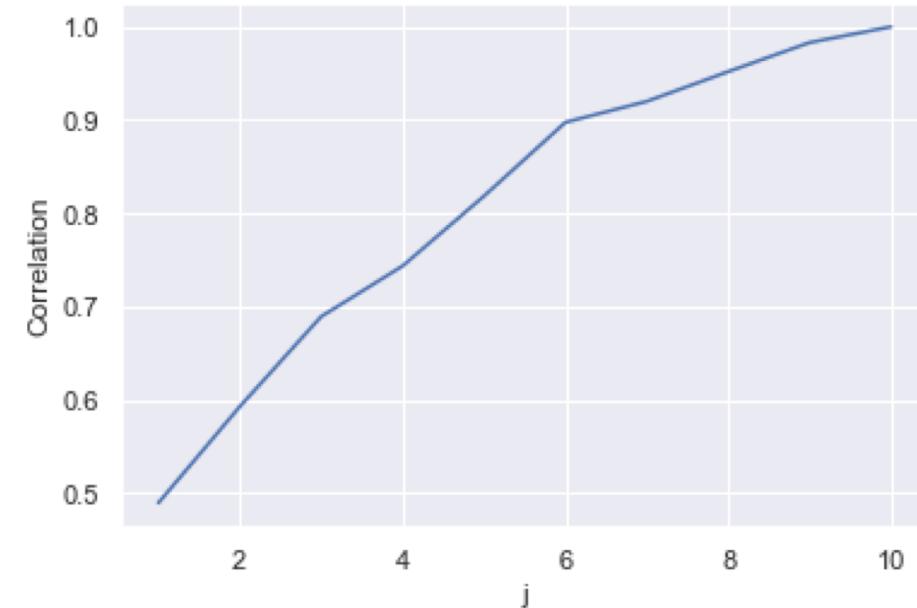
'This diffuser should not be sold, you can only smell it when you're on top of it, **VERY** disappointed'

1	percentage_df	10% percentage	20% percentage	30% percentage	40% percentage	50% percentage	60% percentage	70% percentage	80% percentage	90% percentage	100% percentage
0	0.719739	0.711500	0.706347	0.707237	0.709310	0.707229	0.702496	0.696181	0.697227	0.695798	
1	0.342933	0.556762	0.602789	0.598319	0.587488	0.619823	0.601098	0.587871	0.582425	0.574662	
2	0.500520	0.500335	0.519446	0.519964	0.530518	0.529445	0.537543	0.538444	0.541083	0.545100	
3	0.495921	0.494757	0.556329	0.558353	0.534384	0.551915	0.565614	0.550202	0.556781	0.568927	
4	0.511706	0.649016	0.674553	0.689288	0.704955	0.711570	0.718154	0.721910	0.725638	0.729269	
5	0.665600	0.621225	0.655555	0.679580	0.689123	0.699649	0.704461	0.700578	0.696238	0.682117	
6	0.822914	0.753916	0.688079	0.697787	0.704185	0.710442	0.716446	0.722878	0.736286	0.743622	
7	0.563039	0.574542	0.609664	0.612343	0.612888	0.617632	0.617370	0.617589	0.619183	0.621797	
8	0.573757	0.556634	0.582450	0.511659	0.499645	0.482852	0.487294	0.504067	0.513643	0.516609	
9	0.809356	0.755059	0.775439	0.779120	0.778789	0.779445	0.785710	0.777403	0.781314	0.783790	

TEXT MINING

Sentiment Analysis:

- Calculate correlations
- 30% of the reviews usually reflects enough information
- 60% of the reviews has a 0.9 correlation with total reviews



1 | percentage_df

	10% percentage	20% percentage	30% percentage	40% percentage	50% percentage	60% percentage	70% percentage	80% percentage	90% percentage	100% percentage
0	0.719739	0.711500	0.706347	0.707237	0.709310	0.707229	0.702496	0.696181	0.697227	0.695798
1	0.342933	0.556762	0.602789	0.598319	0.587488	0.619823	0.601098	0.587871	0.582425	0.574662
2	0.500520	0.500335	0.519446	0.519964	0.530518	0.529445	0.537543	0.538444	0.541083	0.545100
3	0.495921	0.494757	0.556329	0.558353	0.534384	0.551915	0.565614	0.550202	0.556781	0.568927
4	0.511706	0.649016	0.674553	0.689288	0.704955	0.711570	0.718154	0.721910	0.725638	0.729269
5	0.665600	0.621225	0.655555	0.679580	0.689123	0.699649	0.704461	0.700578	0.696238	0.682117
6	0.822914	0.753916	0.688079	0.697787	0.704185	0.710442	0.716446	0.722878	0.736286	0.743622
7	0.563039	0.574542	0.609664	0.612343	0.612888	0.617632	0.617370	0.617589	0.619183	0.621797
8	0.573757	0.556634	0.582450	0.511659	0.499645	0.482852	0.487294	0.504067	0.513643	0.516609
9	0.809356	0.755059	0.775439	0.779120	0.778789	0.779445	0.785710	0.777403	0.781314	0.783790

MODELLING - Classification

Data:

240 allure products (1) vs. **240** non-allure products (0)

Features:

loves, price, total_review, avg_rating, recommended_count, vader_100_score,
onestar_pct, twostar_pct, threestar_pct, fourstar_pct, fivestar_pct

Preparation:

- shuffling
- normalization (MinMaxScaler)
- train-test split (30%)

Modelling - Version I

Models:

- Logistic Regression
- Decision Tree
- Random Forest
- Bagging
- Neural Network

Hyperparameter Tuning:

- GridSearchCV(cv=5)

Evaluation Metrics:

- Accuracy
- AUC

Model	Accuracy	AUC
Logistic Regression	0.58	0.58
Decision Tree	0.60	0.61
Random Forest	0.65	0.65
Bagging	0.62	0.62
Neural Network	0.61	0.61

How can we improve?

Modelling - Add data & oversample

1. Enlarge our dataset

240 allure products (1) vs. **1200** non-allure products (0)

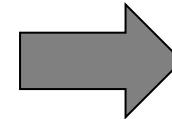
1. Deal with imbalanced dataset

Synthetic Minority Over-sampling Technique (SMOTE)

```
In [156]: 1 from imblearn.over_sampling import SMOTE
In [157]: 1 pd.Series(np.ravel(y_train)).value_counts() #before SMOTE
Out[157]: 0    841
           1    167
          dtype: int64
In [158]: 1 sm = SMOTE(random_state=12, ratio = 1.0)
           2 x_train, y_train = sm.fit_sample(x_train, y_train)
In [159]: 1 pd.Series(np.ravel(y_train)).value_counts() #after SMOTE
Out[159]: 1    841
           0    841
          dtype: int64
```

Modelling - Version II

Model	Accuracy	AUC
Logistic Regression	0.58	0.58
Decision Tree	0.60	0.61
Random Forest	0.65	0.65
Bagging	0.62	0.62
Neural Network	0.61	0.61



Model	Accuracy	AUC
Logistic Regression	0.71	0.62
Decision Tree	0.70	0.64
Random Forest	0.78	0.66
Bagging	0.74	0.63
Neural Network	0.74	0.65

240 non-allure

1200 non-allure

Modelling – Prediction for MEN

Our choice:

Men's Fragrance (exclusively for men)

98 products → **20** potential products

	brand_name	product_name
0	Prada	Luna Rossa
1	CLINIQUE	Clinique Happy for Men
2	Viktor&Rolf	Spicebomb
3	BURBERRY	Brit Splash
4	Dior	Sauvage Eau de Parfum
5	TOM FORD	Grey Vetiver Eau de Parfum
6	Armani Beauty	Acqua di Gio Absolu
7	TOM FORD	Noir Extreme
8	TOM FORD	Noir Anthracite
9	John Varvatos	Dark Rebel
10	Mugler	A*MEN
11	Gucci	Guilty Intense Pour Homme
12	Dior	Sauvage Eau de Toilette
13	Versace	Eros
14	DOLCE&GABBANA	Light Blue Pour Homme Eau de Toilette
15	HERMÈS	Terre d'Hermes Eau Intense Vétiver Eau de Parfum
16	TOM FORD	FOR MEN
17	Calvin Klein	ETERNITY FOR MEN
18	MAISON MARGIELA	'REPLICA' Jazz Club
19	Armani Beauty	Acqua Di Gio Profumo

Modelling - Validation

Fragrance List	Total	Sephora sells	We include
Trendspotter	35	21	9
Gearhungry	20	10	4
GQ	7	2	1
Bazaar	21	8	3

Reference:

- <https://www.thetrendspotter.net/18-best-smelling-irresistible-cologne-men/>
- <https://www.gearhungry.com/best-mens-colognes/>
- <https://www.gq.com/story/best-cologne-2019>
- <https://www.harpersbazaar.com/uk/beauty/fragrance/g25710/best-male-fragrances/>

Dior Sauvage Eau de Parfum:
appears 4 times



Conclusion & Future Work

Classification problem to predict potential allure products

- **Add more data and features**
 - **Validate model using other websites**
-



THANK YOU