

Predicting Skincare Product Price

Predicting pricing and popularity of skin care products via regression models

Crystal Han



Introduction

- How will skincare products rise based on several features?
- How will do skincare products fare in the current market and future market?
- Makeup is included within skincare products



Methodology

- Tools:
 - Python, Pandas, Numpy: Data Cleaning
 - Matplotlib, seaborn, BeautifulSoup
 - Selenium: Scrape sephora website
 - Statsmodels, scikit-learn: Create regression models

Data

- Data
 - Sephora skin care products data via web scraping with Selenium
 - 1000+ data points collected

ters



Glow Recipe
Watermelon Glow PHA + BHA Pore-Tight Toner

★★★★☆ 2.5K | Ask a question | ❤ 176.8K

\$34.00 or 4 interest-free payments of \$8.50 **Klarna** ⓘ

Size: 5.07 oz/ 150 mL

Standard size

5.07 oz/ 150 mL

Mini size

1.35 oz/ 40 mL

Get It Shipped

[Shipping & Returns](#)

☒ **Standard** Enjoy free shipping with code **FREESHIP**.



☐ **Same-Day Delivery** ⓘ Select to see availability for **92802** ▾



Buy Online & Pick Up ⓘ



☐ Select to see availability at **Disneyland** ▾

Add to Basket
for Standard Shipping





Features

- Continuous: Reviews, Number of Reviews, Number of Likes
- Categorical: Skin type, ingredients, brand,

About the Product

Item 2348431

ONLY AT SEPHORA

What it is: A bestselling, gentle PHA- and BHA-infused watermelon toner that hydrates skin and minimizes the appearance of pores.

Skin Type: Normal, Dry, Combination, and Oily

Skincare Concerns: Pores, Dryness, and Dullness

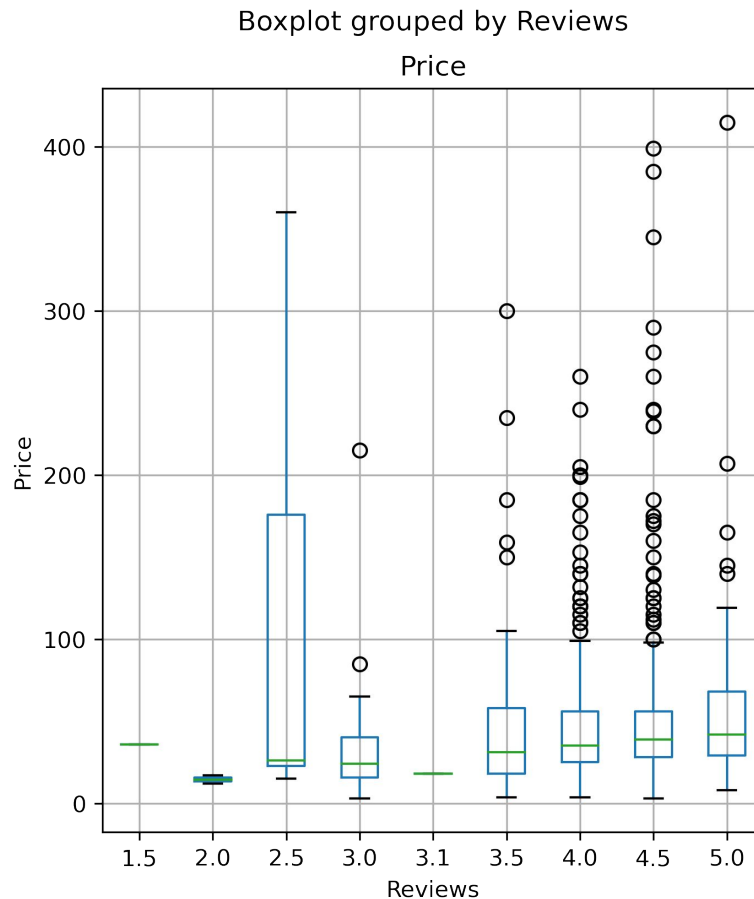
Formulation: Lightweight Liquid

[Show more](#)

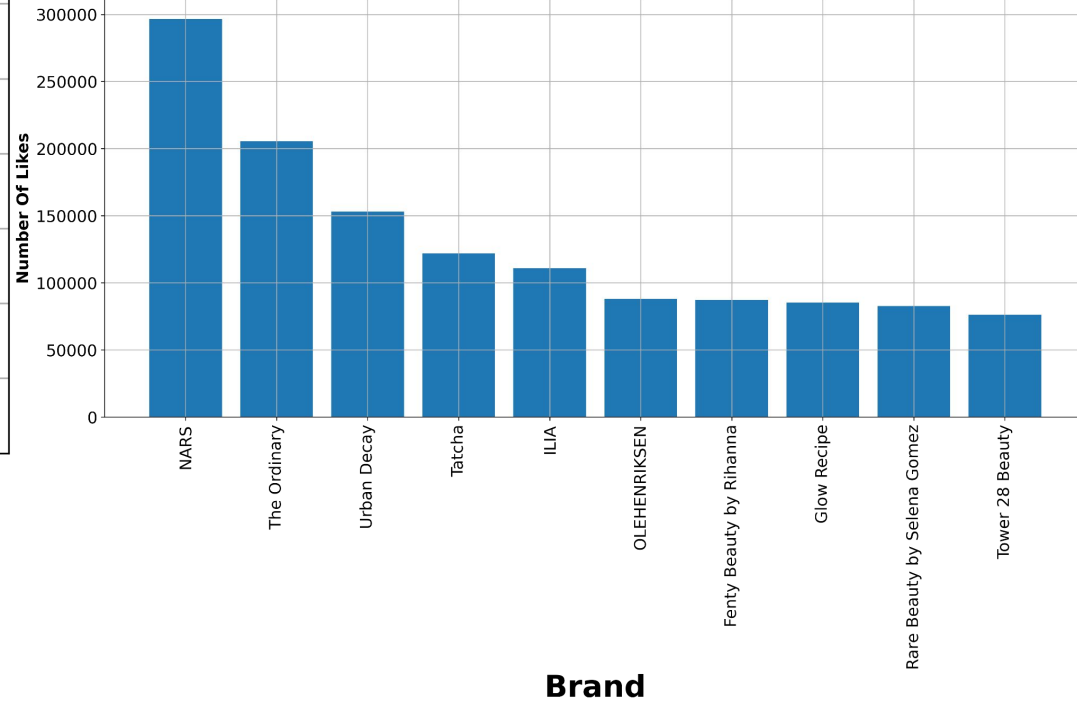
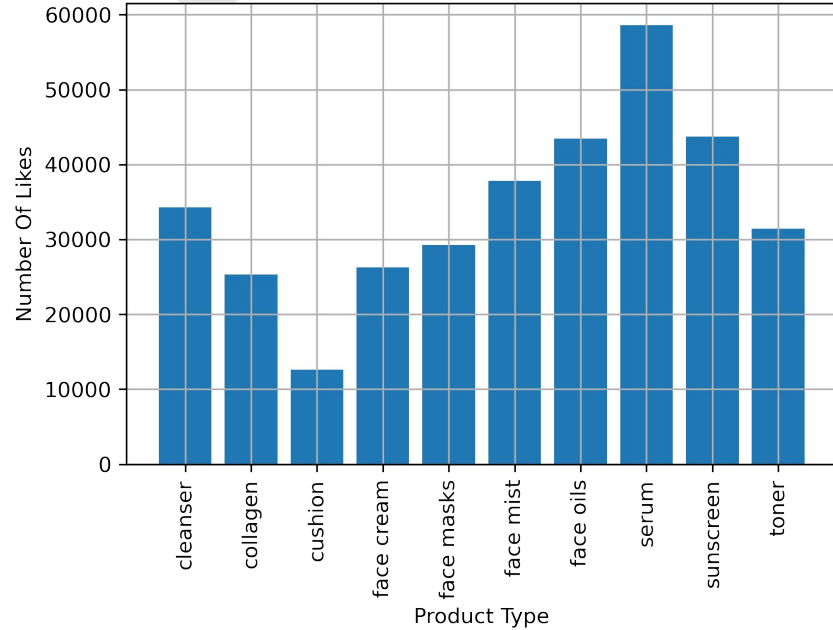
Ingredients

Reviews vs Price

- Relationship between Price and Reviews
- Higher price not necessarily good review

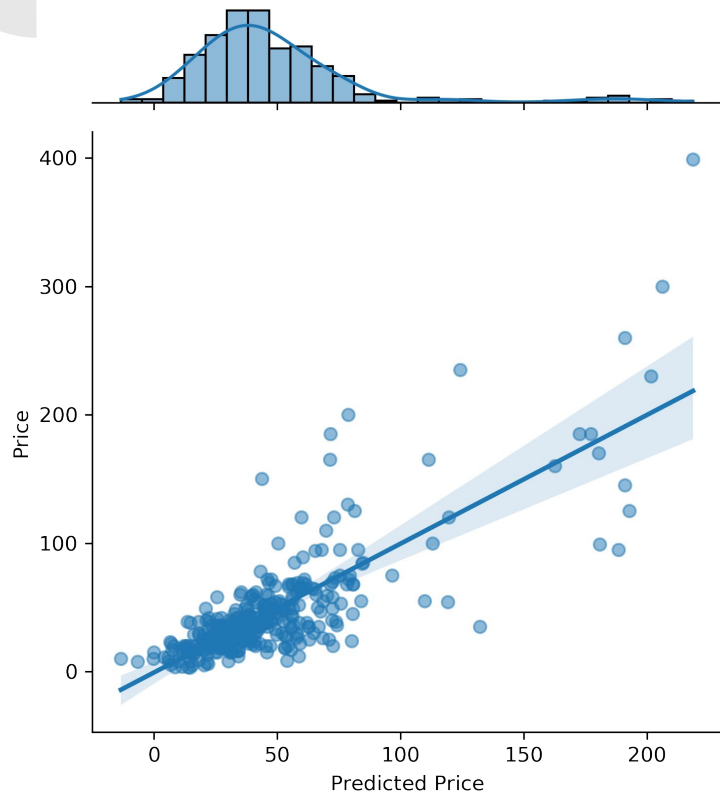


Number of Like vs Product Type and Brand

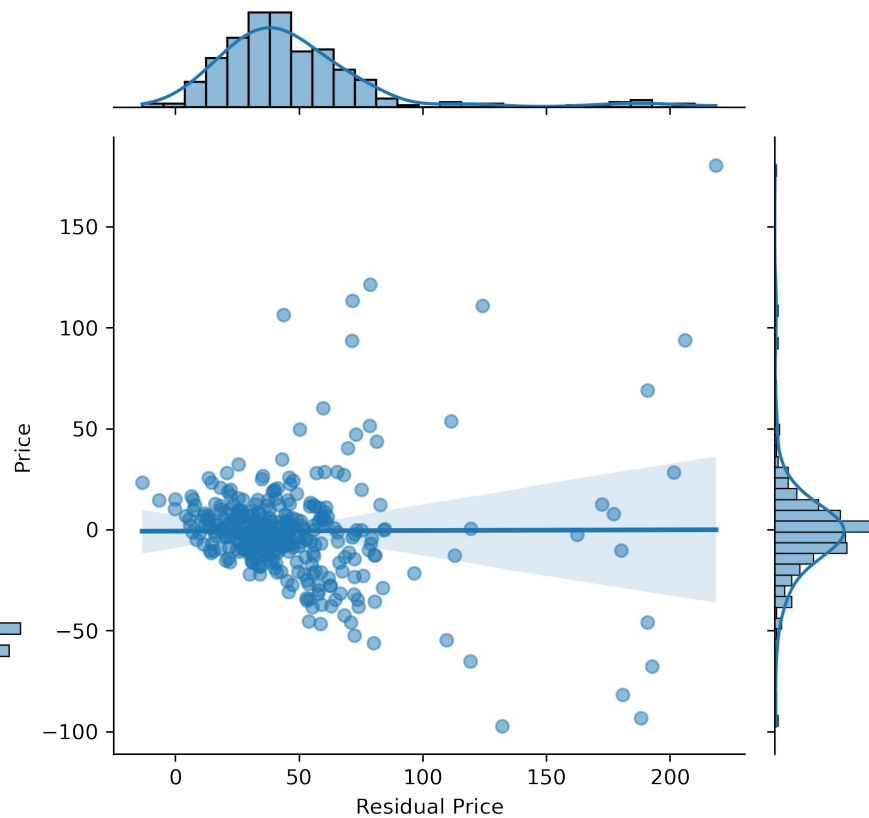


Lasso Regression

Observed Price vs Predicted Price



Residual Price vs Predicted Price

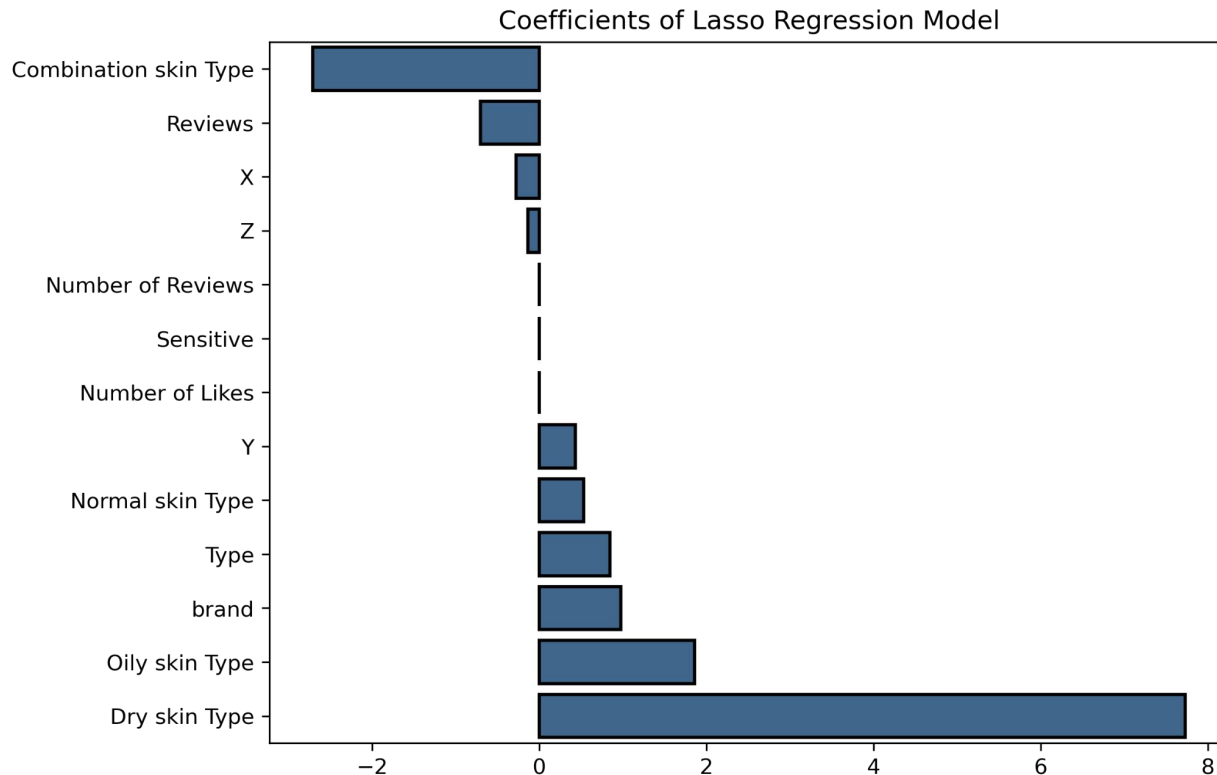




R² Values of Regression Models

Train/Validation/Test	R ²	Cross Validation/Test	R ²	Mean Absolute Error
Linear Regression	0.6362	Linear Regression	0.6024	16.3508
Ridge Regression	0.6362	Ridge Regression	0.6030	16.3413
Lasso Regularization	0.6363	Lasso Regularization	0.6024	16.3398
Polynomial Regression	0.6362	Polynomial Regression	0.6020	16.2660

Features Affecting Price





Future Work

- Gathering skin care product information from other websites such as Ulta, OliveYoung, YesStyle, etc.
- Getting more numerical features and clinical data
- Comparison between top products, skincare vs makeup products
- Rank ingredients using ordinal encoding



Conclusion

- Used Lasso Regression because of Mean Absolute Error
- Price range of most skincare products: ~\$15.00 to ~\$100.00



Appendix

- 1000+ data points on skincare products:
 - Toner
 - Cleansers
 - Facial Cream
 - Facial Sheet masks
 - Sunscreen
 - Cushions
- Data cleaning: Products with no reviews will be considered as 5 stars so as not to assume they are bad products
- Reasons for selecting products:
 - Used in skincare steps
 - Skin suitability based on types: Normal, Oily, Combination, Sensitive
 - Makeup as a daily skincare product as well
- Ingredients:
 - Retinol, Ascorbic Acid, Niacinamide, Ceramide, Squalane, Hyaluronic Acid, Vitamin E, Titanium Dioxide, Glycolic Acid, Sodium Acetate, Sodium Benzoate, Sodium Hyaluronate, Panthenol

Thank you

