PREDICTING THE PRICE OF SKINCARE PRODUCTS USING REGRESSION MODEL

JING LEE

APRIL 20, 2022

METIS PROJECT 2

WEB SCRAPING AND REGRESSION

Introduction

- Motivation
  - Skincare is a multibillion-dollar growing industry
  - The price for 1 oz cream range from $10 to $200+
  - What determine the pricing of products?
  - Budget brands vs luxury brands?
- Objective
  - Investigation the features that contribute to the pricing of products
  - Use a regression model to predict price of skincare products
- Goals
  - Consumers can decide what products to buy within their budget
  - New skincare companies can understand how products are priced

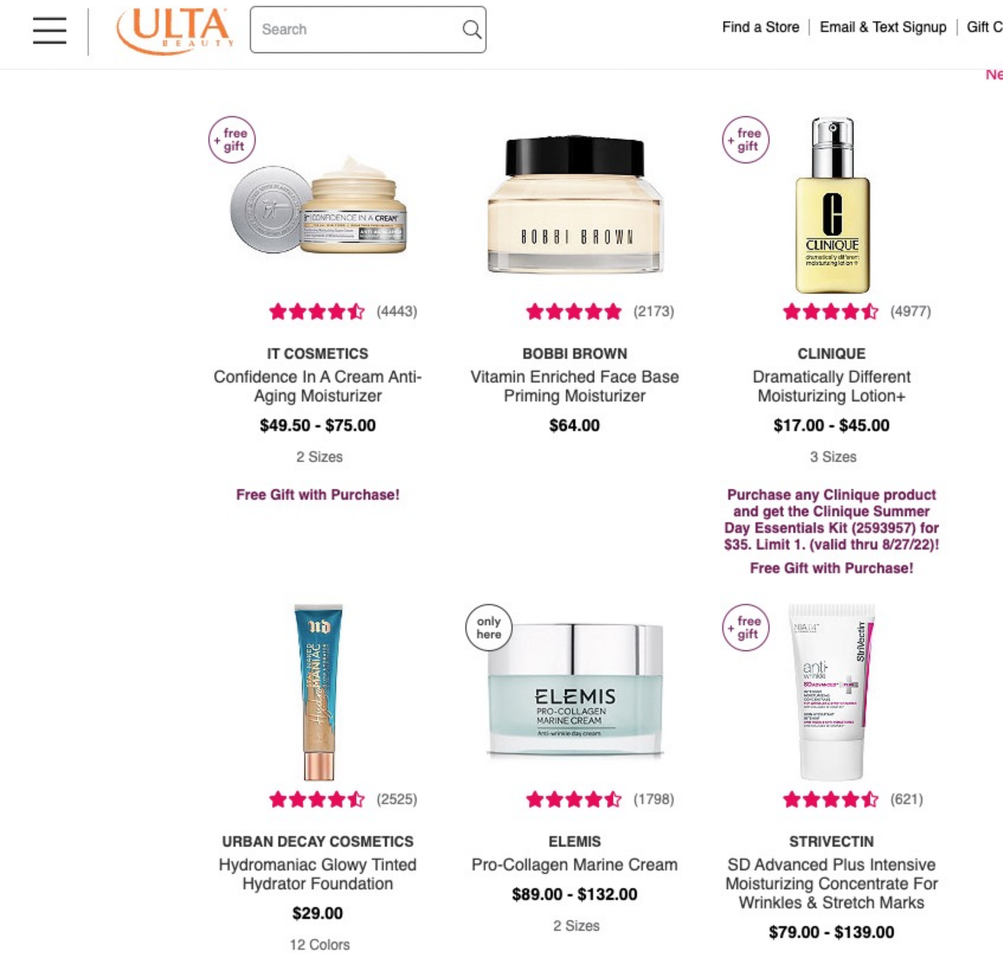
What goes into the pricing of products?

- $123 per 1 oz
- Parent company: L’oreal
- Luxury brand
- Rating: 4.6 stars
- Reviews: 3,919

- $17 for 1 oz
- Parent company: L’oreal
- Budget brand
- Rating 4.12 stars
- Reviews: 112

Methodology: data collection

- Web scraping data (~1847 products)
  - website: www.ulta.com
  - Major retailer of beauty products ranging from luxury to affordable brands



Search

Find a Store | Email & Text Signup | Gift C

Ne

free gift

(4443)

IT COSMETICS

Confidence In A Cream Anti-Aging Moisturizer

$49.50 - $75.00

2 Sizes

Free Gift with Purchase!

(2173)

BOBBI BROWN

Vitamin Enriched Face Base Priming Moisturizer

$64.00

free gift

(4977)

CLINIQUE

Dramatically Different Moisturizing Lotion+

$17.00 - $45.00

3 Sizes

Purchase any Clinique product and get the Clinique Summer Day Essentials Kit (2593957) for $35. Limit 1. (valid thru 8/27/22)!

Free Gift with Purchase!

(2525)

URBAN DECAY COSMETICS

Hydromaniac Glowy Tinted Hydrator Foundation

$29.00

12 Colors

only here

(1798)

ELEMIS

Pro-Collagen Marine Cream

$89.00 - $132.00

2 Sizes

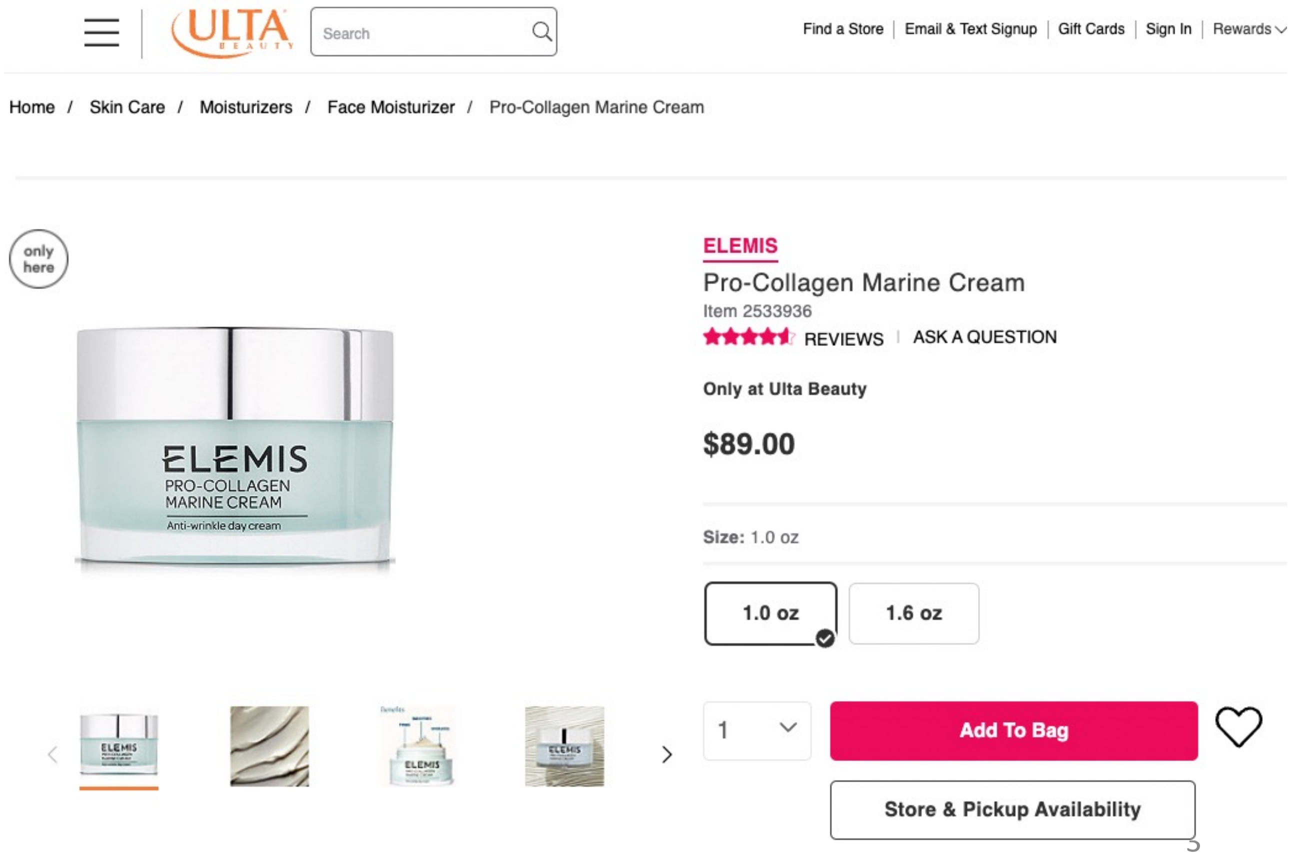
free gift

(621)

STRIVECTIN

SD Advanced Plus Intensive Moisturizing Concentrate For Wrinkles & Stretch Marks

$79.00 - $139.00



Search

Find a Store | Email & Text Signup | Gift Cards | Sign In | Rewards

Home / Skin Care / Moisturizers / Face Moisturizer / Pro-Collagen Marine Cream

only here

ELEMIS

Pro-Collagen Marine Cream

Item 2533936

REVIEWS | ASK A QUESTION

Only at Ulta Beauty

$89.00

Size: 1.0 oz

1.0 oz | 1.6 oz

1

Add To Bag

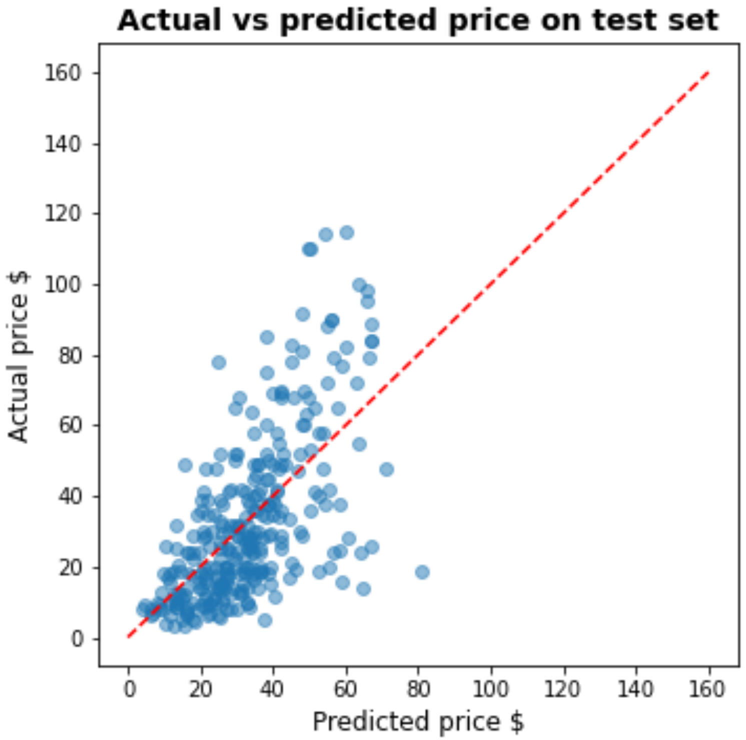
Store & Pickup Availability

Methodology: Tools

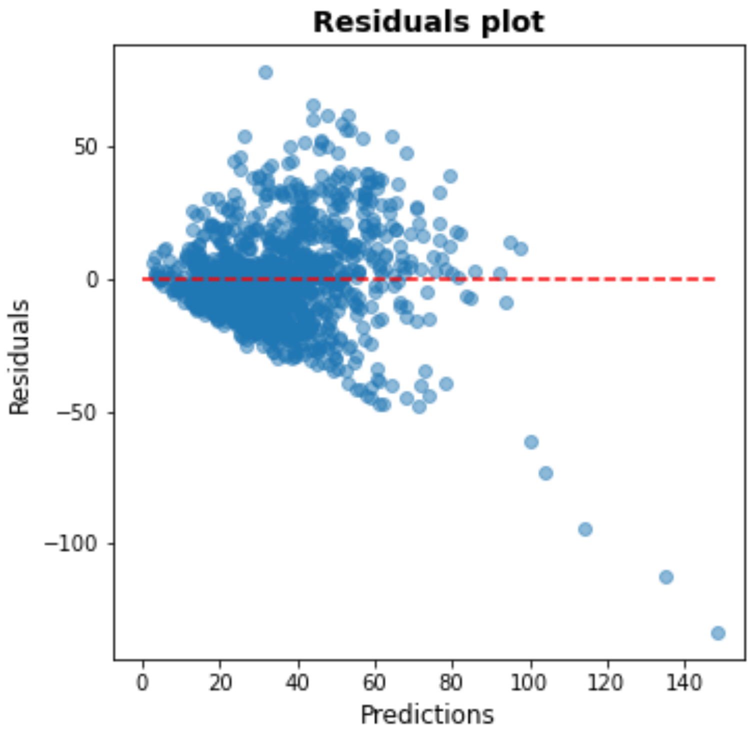
- Tools:
  - Requests/ BeautifulSoup, pandas, matplotlib, seaborn, statsmodels, sklearn
- Models:
  - Linear regression without regularization
  - Linear regression with regularization
    - Lasso
    - Ridge
    - Elastic net
  - Extensive feature engineering to improve model
- Metrics:
  - R^2 , mean absolute error (MAE)

Model selection – cross validation

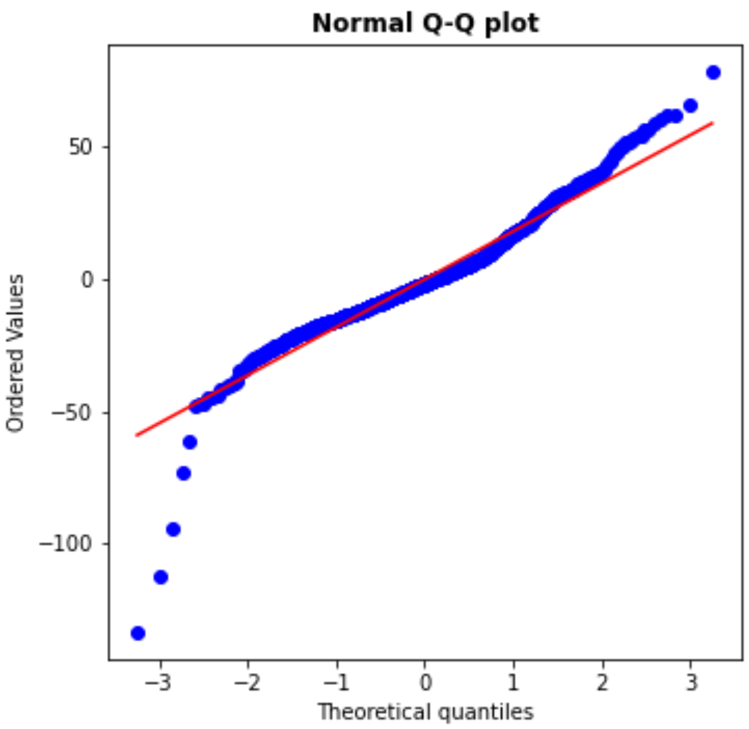
- Linear regression without regularization
  - Predictors: 10 → 17 and 1501 datapoints
- Mean train R^2 : 0.409 +/- 0.063
- Test R^2 : 0.413
- MAE: $13.66



Actual vs predicted price on test set



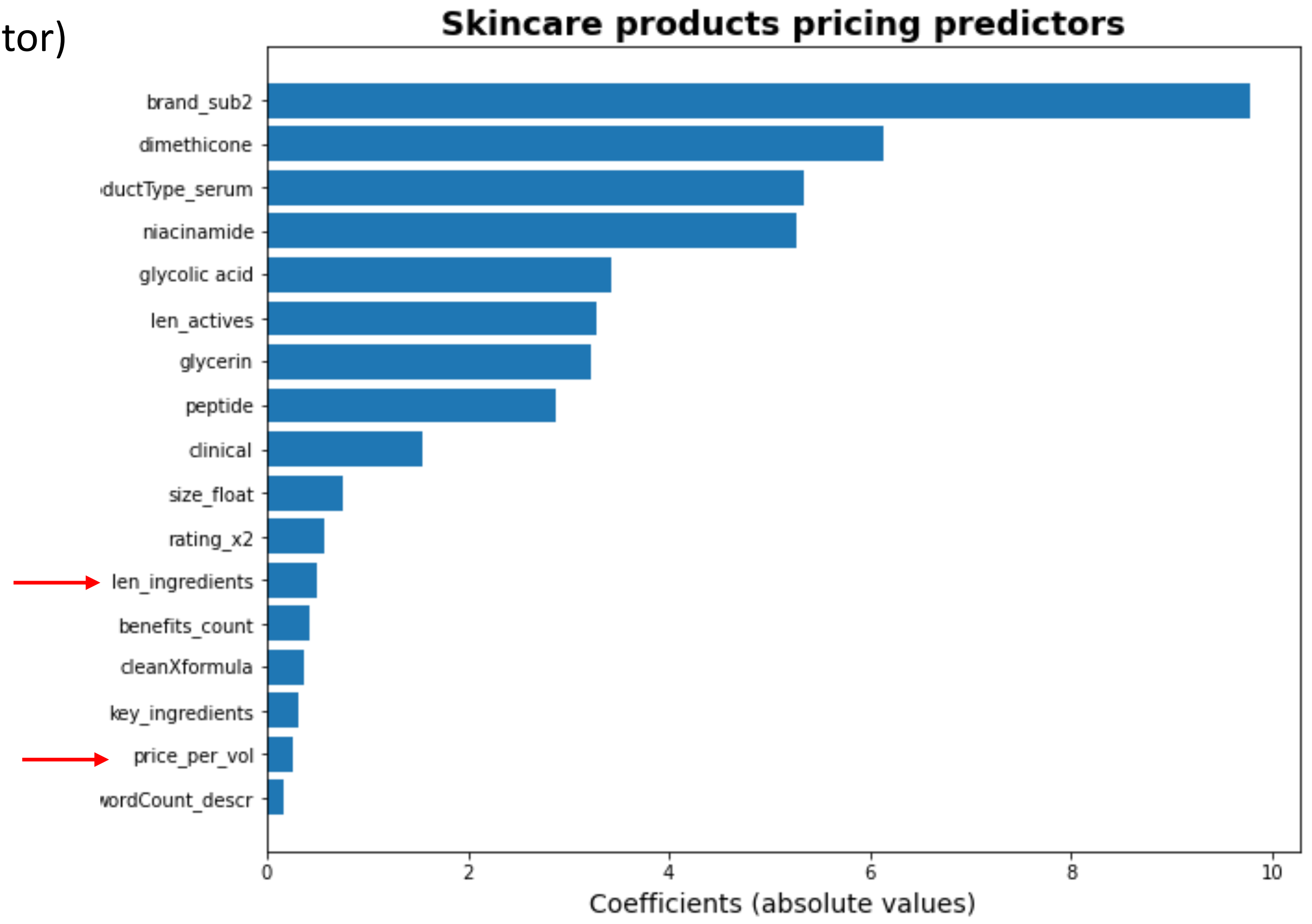
Residuals plot



Normal Q-Q plot

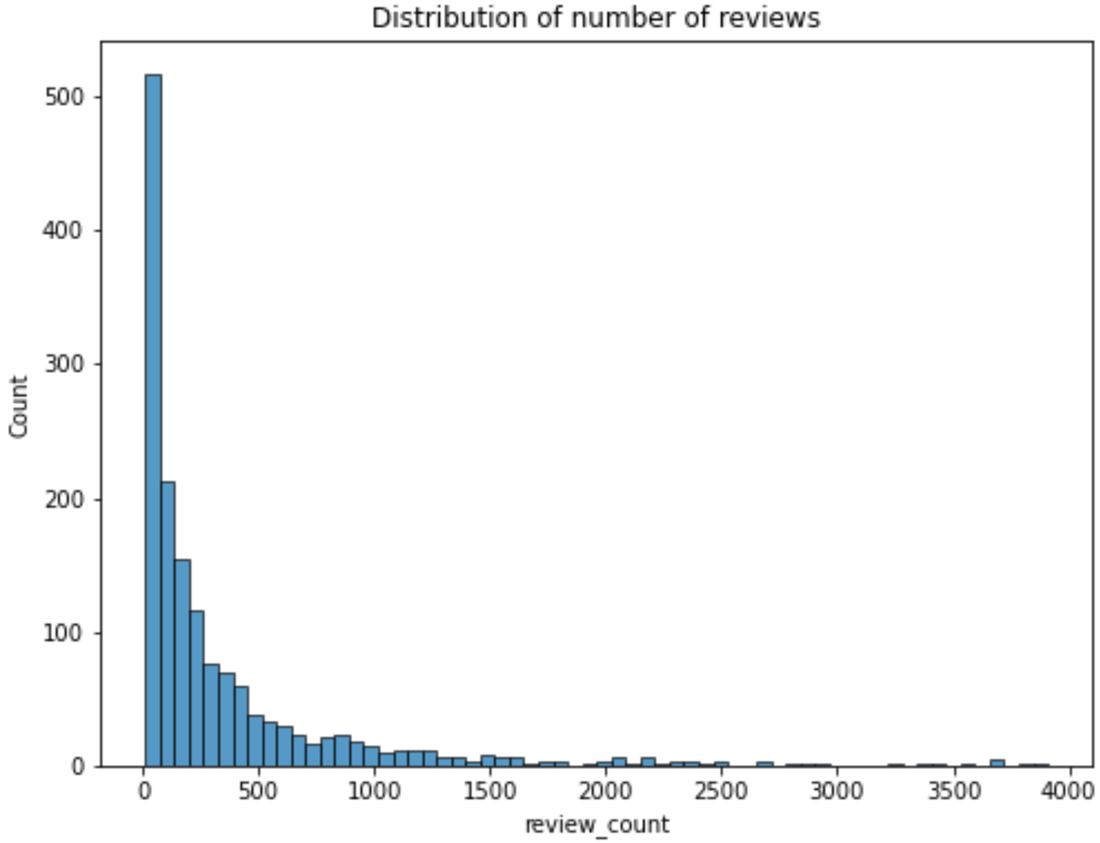
Predictors in the model

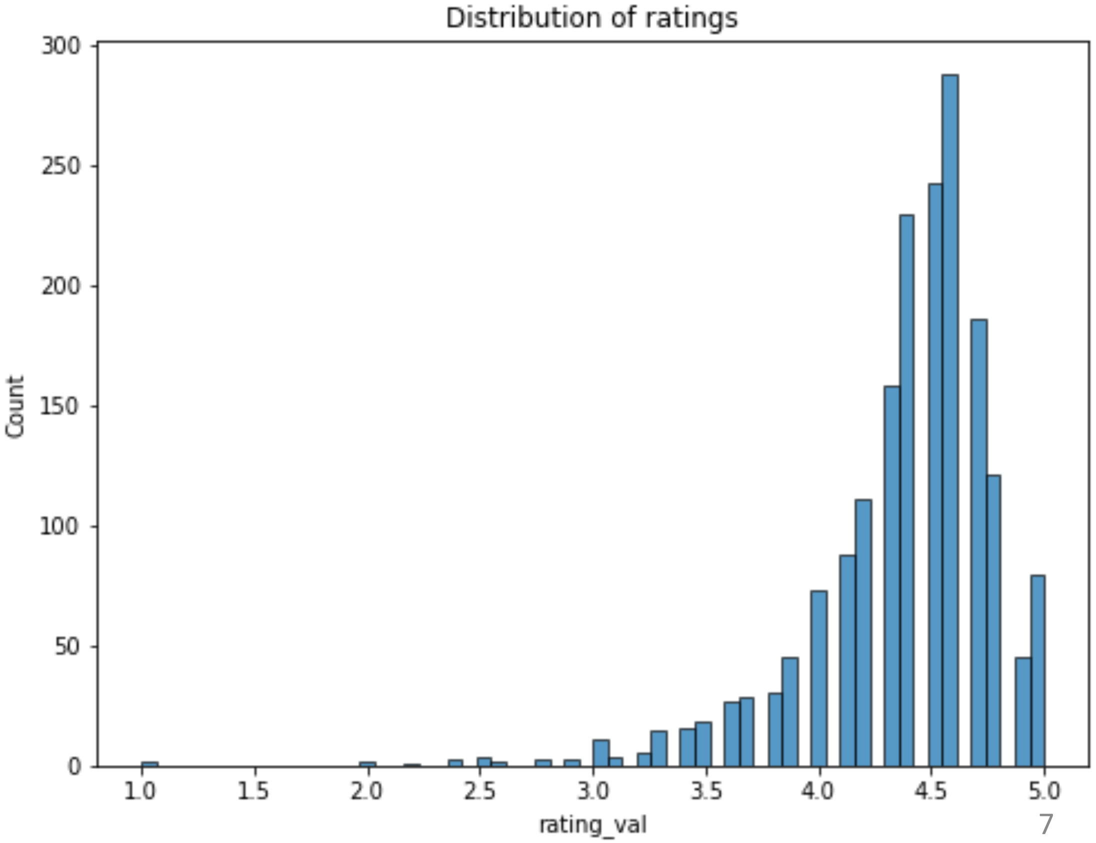
- Predictors correlation with price (predictor)
  - len_ingredients 0.44
  - price_per_vol 0.44



Other findings

- Most brand come have the same 7 major parent companies
- High ratings and low number of reviews and vice verse





Conclusions

- Predicting price with linear regression and the current features is not a good model
- Model has low prediction accuracy of ~ 41 %
- Price prediction is +/- $13.66 is high considering average products cost ~$35
- Missing important features such as investment in marketing which could heavily influence pricing and purchase power
- There is no clear predictor for products pricing

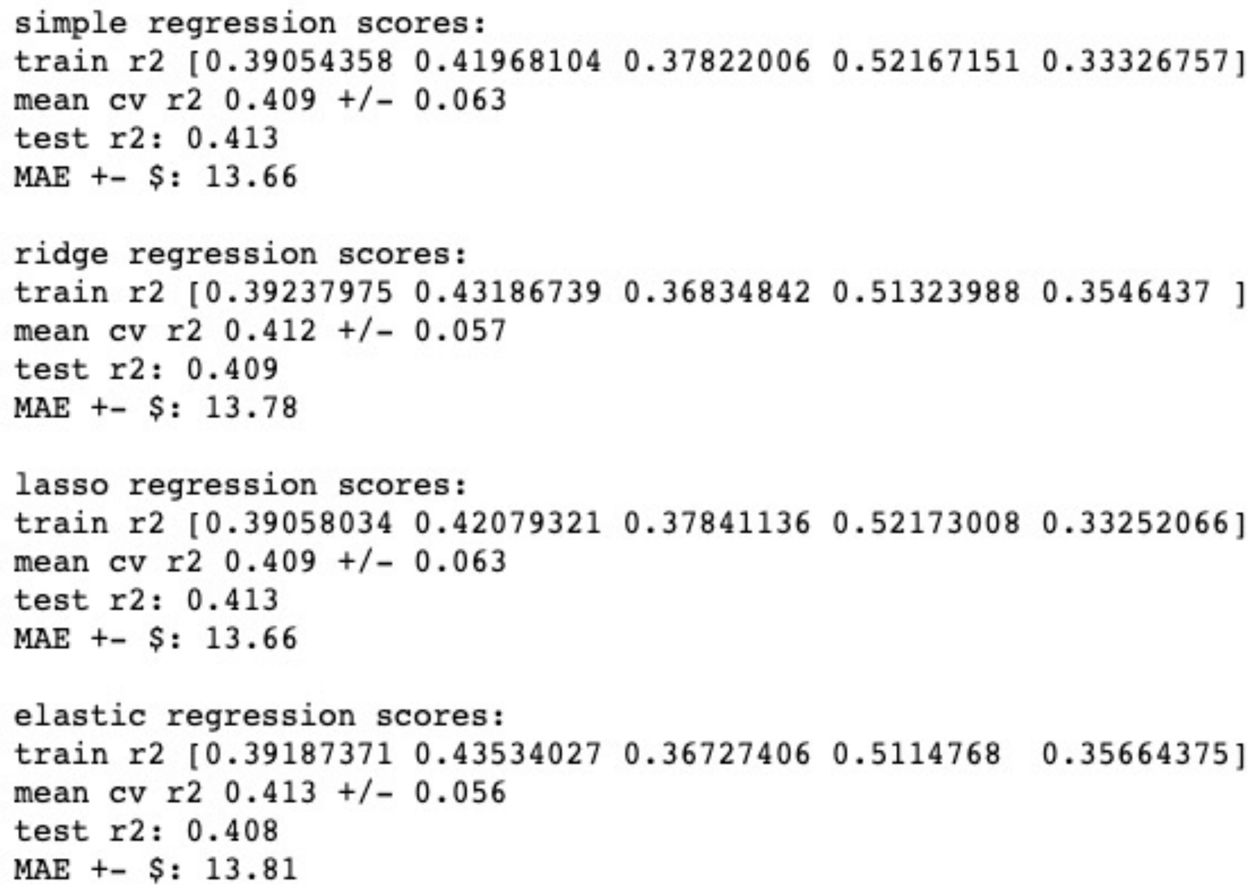
Future work

- Acquire data from other products sources
  - Sephora: luxury brands
  - Target: affordable brands
- In depth study of the actual ingredients in the product to determine pricing
  - use of classification model to improve prediction
- A feature to account for marketing influence

Thank you!!

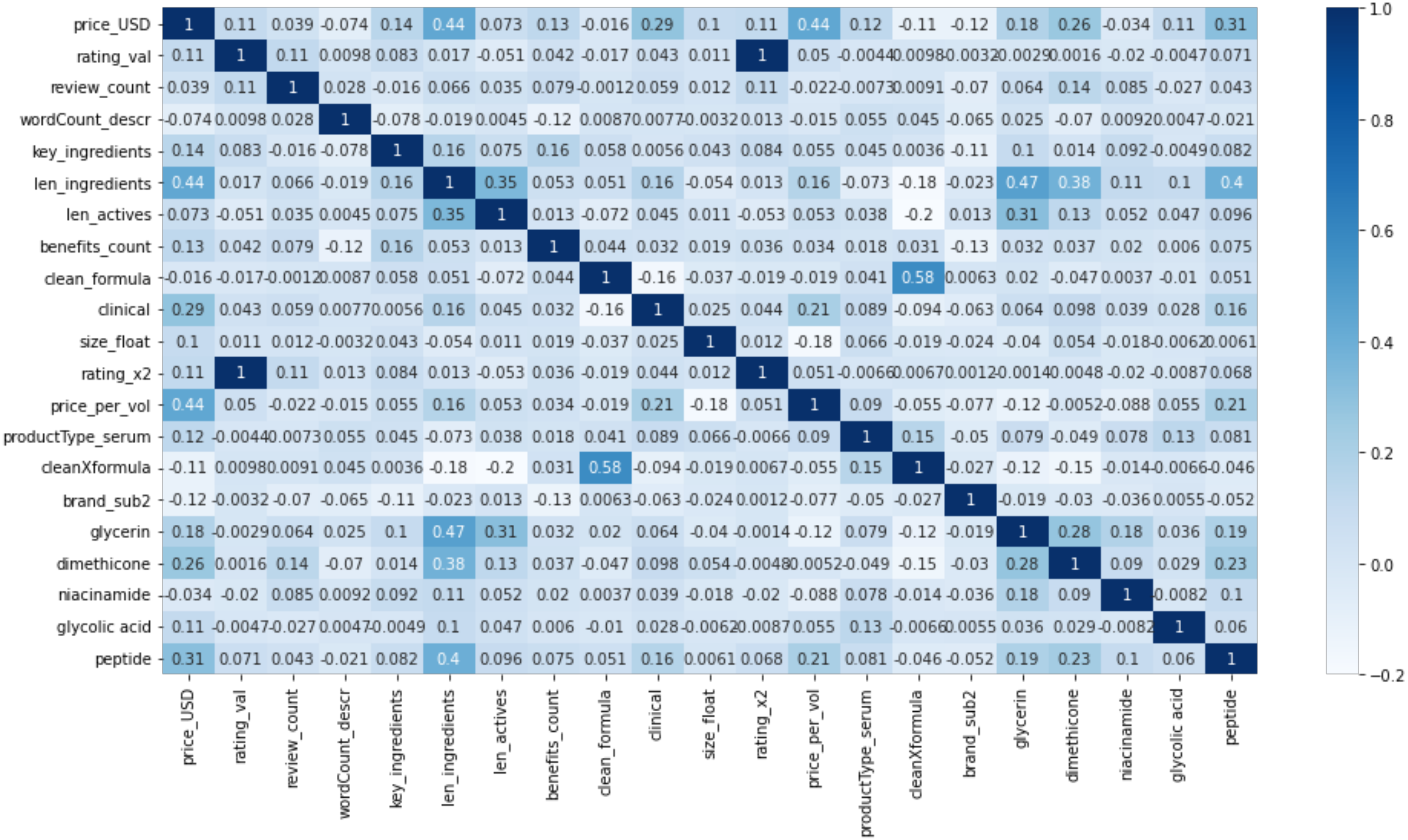
Questions?

Appendix

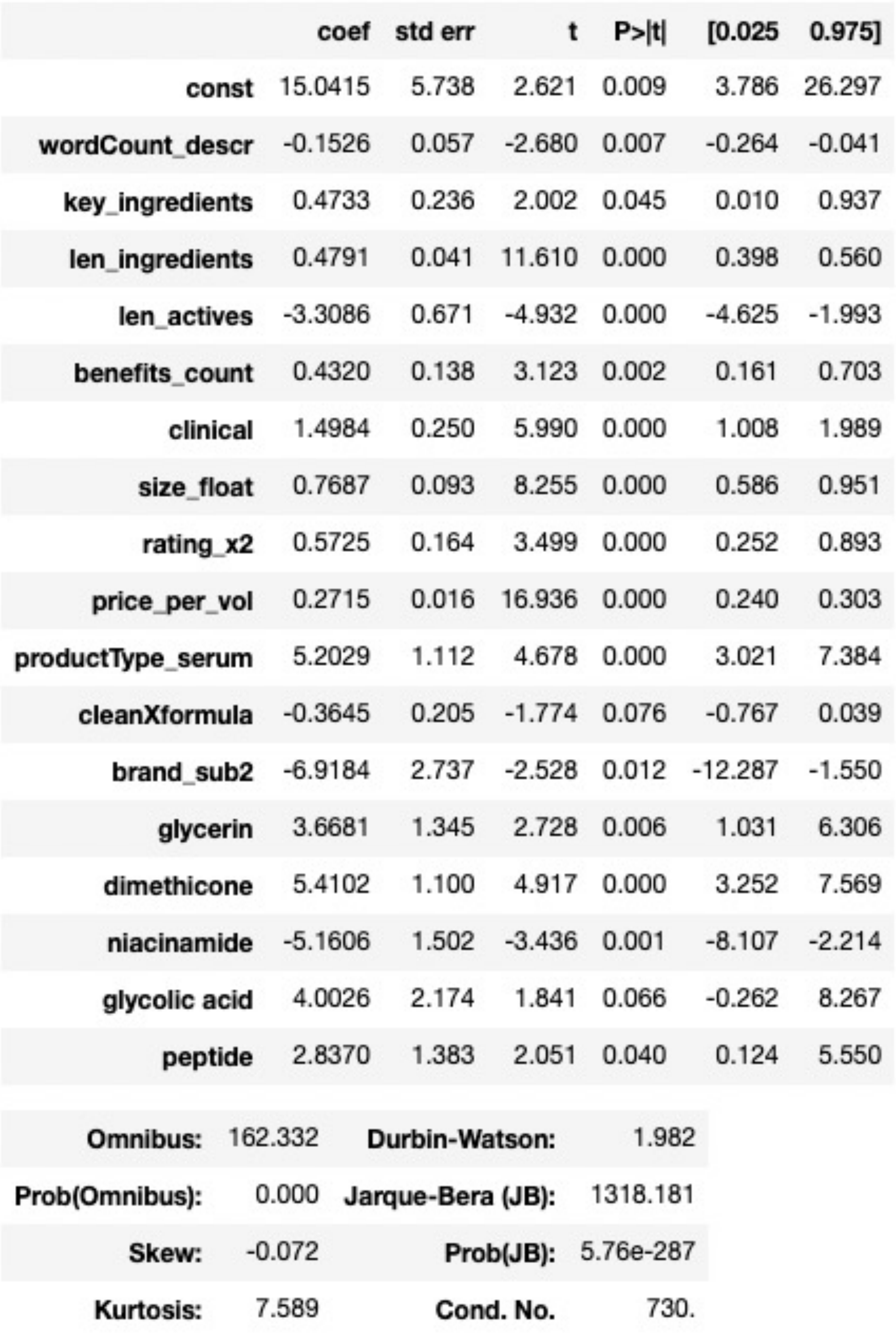


```
simple regression scores:
train r2 [0.39054358 0.41968104 0.37822006 0.52167151 0.33326757]
mean cv r2 0.409 +/- 0.063
test r2: 0.413
MAE +- $: 13.66

ridge regression scores:
train r2 [0.39237975 0.43186739 0.36834842 0.51323988 0.3546437 ]
mean cv r2 0.412 +/- 0.057
test r2: 0.409
MAE +- $: 13.78

lasso regression scores:
train r2 [0.39058034 0.42079321 0.37841136 0.52173008 0.33252066]
mean cv r2 0.409 +/- 0.063
test r2: 0.413
MAE +- $: 13.66

elastic regression scores:
train r2 [0.39187371 0.43534027 0.36727406 0.5114768  0.35664375]
mean cv r2 0.413 +/- 0.056
test r2: 0.408
MAE +- $: 13.81
```

Heatmap



	price_USD	rating_val	review_count	wordCount_descr	key_ingredients	len_ingredients	len_actives	benefits_count	clean_formula	clinical	size_float	rating_x2	price_per_vol	productType_serum	cleanXformula	brand_sub2	glycerin	dimethicone	niacinamide	glycolic acid	peptide
price_USD	1	0.11	0.039	-0.074	0.14	0.44	0.073	0.13	-0.016	0.29	0.1	0.11	0.44	0.12	-0.11	-0.12	0.18	0.26	-0.034	0.11	0.31
rating_val	0.11	1	0.11	0.0098	0.083	0.017	-0.051	0.042	-0.017	0.043	0.011	1	0.05	-0.0044	0.0098	-0.0032	-0.0029	0.0016	-0.02	-0.0047	0.071
review_count	0.039	0.11	1	0.028	-0.016	0.066	0.035	0.079	-0.0012	0.059	0.012	0.11	-0.022	-0.0073	0.0091	-0.07	0.064	0.14	0.085	-0.027	0.043
wordCount_descr	-0.074	0.0098	0.028	1	-0.078	-0.019	0.0045	-0.12	0.0087	0.0077	-0.0032	0.013	-0.015	0.055	0.045	-0.065	0.025	-0.07	0.0092	0.0047	-0.021
key_ingredients	0.14	0.083	-0.016	-0.078	1	0.16	0.075	0.16	0.058	0.0056	0.043	0.084	0.055	0.045	0.0036	-0.11	0.1	0.014	0.092	-0.0049	0.082
len_ingredients	0.44	0.017	0.066	-0.019	0.16	1	0.35	0.053	0.051	0.16	-0.054	0.013	0.16	-0.073	-0.18	-0.023	0.47	0.38	0.11	0.1	0.4
len_actives	0.073	-0.051	0.035	0.0045	0.075	0.35	1	0.013	-0.072	0.045	0.011	-0.053	0.053	0.038	-0.2	0.013	0.31	0.13	0.052	0.047	0.096
benefits_count	0.13	0.042	0.079	-0.12	0.16	0.053	0.013	1	0.044	0.032	0.019	0.036	0.034	0.018	0.031	-0.13	0.032	0.037	0.02	0.006	0.075
clean_formula	-0.016	-0.017	-0.0012	0.0087	0.058	0.051	-0.072	0.044	1	-0.16	-0.037	-0.019	-0.019	0.041	0.58	0.0063	0.02	-0.047	0.0037	-0.01	0.051
clinical	0.29	0.043	0.059	0.0077	0.0056	0.16	0.045	0.032	-0.16	1	0.025	0.044	0.21	0.089	-0.094	-0.063	0.064	0.098	0.039	0.028	0.16
size_float	0.1	0.011	0.012	-0.0032	0.043	-0.054	0.011	0.019	-0.037	0.025	1	0.012	-0.18	0.066	-0.019	-0.024	-0.04	0.054	-0.018	-0.0062	0.0061
rating_x2	0.11	1	0.11	0.013	0.084	0.013	-0.053	0.036	-0.019	0.044	0.012	1	0.051	-0.0066	0.0067	0.0012	-0.0014	-0.0048	-0.02	-0.0087	0.068
price_per_vol	0.44	0.05	-0.022	-0.015	0.055	0.16	0.053	0.034	-0.019	0.21	-0.18	0.051	1	0.09	-0.055	-0.077	-0.12	-0.0052	-0.088	0.055	0.21
productType_serum	0.12	-0.0044	-0.0073	0.055	0.045	-0.073	0.038	0.018	0.041	0.089	0.066	-0.0066	0.09	1	0.15	-0.05	0.079	-0.049	0.078	0.13	0.081
cleanXformula	-0.11	0.0098	0.0091	0.045	0.0036	-0.18	-0.2	0.031	0.58	-0.094	-0.019	0.0067	-0.055	0.15	1	-0.027	-0.12	-0.15	-0.014	-0.0066	-0.046
brand_sub2	-0.12	-0.0032	-0.07	-0.065	-0.11	-0.023	0.013	-0.13	0.0063	-0.063	-0.024	0.0012	-0.077	-0.05	-0.027	1	-0.019	-0.03	-0.036	0.0055	-0.052
glycerin	0.18	-0.0029	0.064	0.025	0.1	0.47	0.31	0.032	0.02	0.064	-0.04	-0.0014	-0.12	0.079	-0.12	-0.019	1	0.28	0.18	0.036	0.19
dimethicone	0.26	0.0016	0.14	-0.07	0.014	0.38	0.13	0.037	-0.047	0.098	0.054	-0.0048	-0.0052	-0.049	-0.15	-0.03	0.28	1	0.09	0.029	0.23
niacinamide	-0.034	-0.02	0.085	0.0092	0.092	0.11	0.052	0.02	0.0037	0.039	-0.018	-0.02	-0.088	0.078	-0.014	-0.036	0.18	0.09	1	-0.0082	0.1
glycolic acid	0.11	-0.0047	-0.027	0.0047	-0.0049	0.1	0.047	0.006	-0.01	0.028	-0.0062	-0.0087	0.055	0.13	-0.0066	0.0055	0.036	0.029	-0.0082	1	0.06
peptide	0.31	0.071	0.043	-0.021	0.082	0.4	0.096	0.075	0.051	0.16	0.0061	0.068	0.21	0.081	-0.046	-0.052	0.19	0.23	0.1	0.06	1



	coef	std err	t	P>|t|	[0.025	0.975]
const	15.0415	5.738	2.621	0.009	3.786	26.297
wordCount_descr	-0.1526	0.057	-2.680	0.007	-0.264	-0.041
key_ingredients	0.4733	0.236	2.002	0.045	0.010	0.937
len_ingredients	0.4791	0.041	11.610	0.000	0.398	0.560
len_actives	-3.3086	0.671	-4.932	0.000	-4.625	-1.993
benefits_count	0.4320	0.138	3.123	0.002	0.161	0.703
clinical	1.4984	0.250	5.990	0.000	1.008	1.989
size_float	0.7687	0.093	8.255	0.000	0.586	0.951
rating_x2	0.5725	0.164	3.499	0.000	0.252	0.893
price_per_vol	0.2715	0.016	16.936	0.000	0.240	0.303
productType_serum	5.2029	1.112	4.678	0.000	3.021	7.384
cleanXformula	-0.3645	0.205	-1.774	0.076	-0.767	0.039
brand_sub2	-6.9184	2.737	-2.528	0.012	-12.287	-1.550
glycerin	3.6681	1.345	2.728	0.006	1.031	6.306
dimethicone	5.4102	1.100	4.917	0.000	3.252	7.569
niacinamide	-5.1606	1.502	-3.436	0.001	-8.107	-2.214
glycolic acid	4.0026	2.174	1.841	0.066	-0.262	8.267
peptide	2.8370	1.383	2.051	0.040	0.124	5.550

Omnibus:	162.332	Durbin-Watson:	1.982
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1318.181
Skew:	-0.072	Prob(JB):	5.76e-287
Kurtosis:	7.589	Cond. No.	730.