

E-COMMERCE Data Analysis

www.hktvmall.com

Web Scraping Project

Sophia, Rainfield, Gemma, Michael



01. PROJECT AIMS



02. DATA COLLECTION



03. PREPROCESSING



04. ANALYSIS



05. CONCLUSIONS

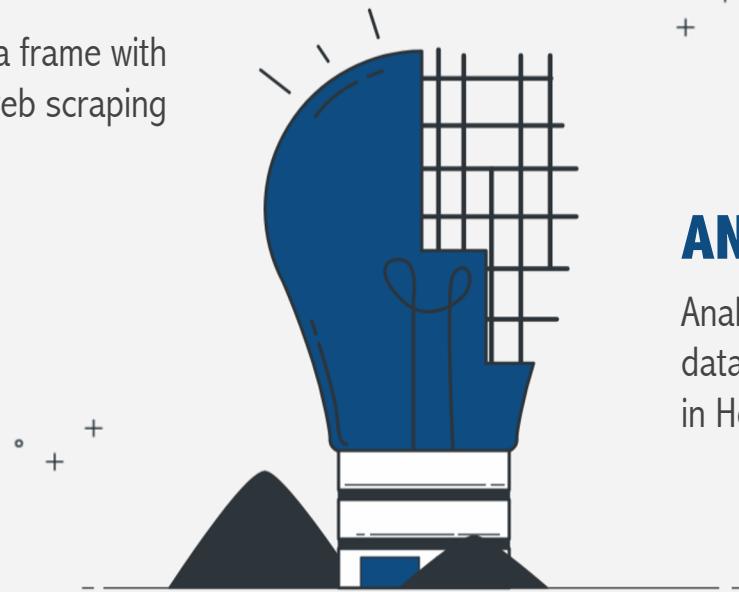




Project Aims

DATA

Constructing a data frame with information collected via web scraping



ANALYSIS

Analyzing the dataset and forming data-driven insights on a local market in Hong Kong

+

+

Business Value



+

Identifying new opportunities

Are there in-demand products that are under-supplied?

Are there new audiences to target?

+ +



Understanding Customer Needs

What do our customers want? How can we better cater to their needs?



. +



+

Determining Growth Factors

What are the biggest drivers of e-commerce sales?

+

+



Data Collection

```
df=pd.DataFrame()

URL= "https://www.hktvmall.com/hktv/en/main/search?page=0&q=%3Arelevance%3Azone%3Abeautyinhealth%3Astreet%3Amain%3A"
URL1="https://www.hktvmall.com/hktv/en/main/search?page="
URL2="&q=%3Arelevance%3Azone%3Abeautyinhealth%3Astreet%3Amain%3A"

driver = webdriver.Chrome(executable_path="/Applications/chromedriver")
driver.get(URL)
time.sleep(5)

html = driver.page_source
soup=BeautifulSoup(html, "html.parser")
URL_iter=0

while int(URL_iter)<867:

    for item in soup.find_all(class_="info-wrapper"):

        try:
            p_raw=item.find(class_="brand-product-name").text
            print("Raw Info:", p_raw)
        except:
            p_raw=str(np.nan)
            print("Raw Info:", p_raw)

        try:
            p_brand=re.match("(.*?)\s\-\s", p_raw).group(1)
            print("Product Brand:", p_brand)
        except:
            p_brand=p_raw
            print("Product Brand:", p_brand)
```

The screenshot shows the browser's developer tools with the 'Elements' tab selected. The page content is displayed in the main pane, and the element tree is shown on the right. The element being inspected is a `<div class="info-wrapper">`. The tree also includes other components like `upper-wrapper`, `brand-product-name`, `packing-spec`, `salesNumber-container`, `review-info`, `lower-wrapper`, `price-label`, `delivery-label-container`, `store-name-label`, and `bottom-button-wrapper`. A detailed breakdown of the `brand-product-name` element's style is shown in the bottom right, including margin, border, and padding properties.



Data Collection

Website

HKTV mall

Category

Skincare

Method

Web scraping

API (not for public use and higher difficulty)

The screenshot shows the Chrome DevTools Elements tab with the DOM tree for a product page. The tree starts with a `<!-- /react-text -->` node, followed by a `</div>`, and then a `<div class="info-wrapper">`. Inside this wrapper are several other divs with classes like "upper-wrapper", "brand-product-name", "packing-spec", "salesNumber-container", "review-info", "lower-wrapper", "price-label", "delivery-label-container", and "store-name-label". A specific element, `...`, is highlighted with a blue selection bar. At the bottom of the DevTools, there are tabs for Styles, Event Listeners, DOM Breakpoints, Properties, and Accessibility. The Styles tab is currently selected. In the bottom right corner, there is a detailed view of the "brand-product-name" element's styling, showing margin, border, and padding properties.



Data Preprocessing

1. Cast object to specified dtype

Astype – Price, Review, Rating

2. Extract useful string

Regex – Brand, Product

3. Replace missing values

Try, Except – NaN, 0

4. Extra Columns

Revenue, Product Category, Gender, % Discount



Data Frame

	Platform	Product Category	Product Brand	Product Name	Original Price	Discount Price	Discount	% Discount	Price Range	Sales	Revenue	Rating	Review	Promo Tag	Gender
0	HKTV_Mall	Others	Kotex	Twin Pack	52	46	6	11.538462	0-100	40000.0	1840000.0	4.7	667.0	immediate shipment	Female
1	HKTV_Mall	Others	Curel	Intensive Moisture Care Foaming Wash 150ml (26...)	119	88	31	26.050420	0-100	3000.0	264000.0	4.6	99.0	2-day group shipment	Female
2	HKTV_Mall	Hydrating Mask	Eaoron	Hyaluronic Acid Collagen Hydrating Face Mask 2...	195	67	128	65.641026	0-100	20000.0	1340000.0	4.7	383.0	2-day group shipment	Female
...
50015	HKTV_Mall	Others	MANDOM	Fusain Facial wash, Mandom 炭深層洗面膏 (100g x3)	54	118	-64	0.000000	101-200	NaN	NaN	NaN	NaN	2-day group shipment	Male
50016	HKTV_Mall	Others	Clarins	Men Active Face Wash 125ml	300	198	102	34.000000	101-200	NaN	NaN	NaN	NaN	2-day group shipment	Male

50017 rows × 15 columns

Analysis



Male vs. Female



532

Male skincare
products listed



18,000+

Male skincare
products sold



49,485

Female skincare
products listed



3,600,000+

Female skincare
products sold

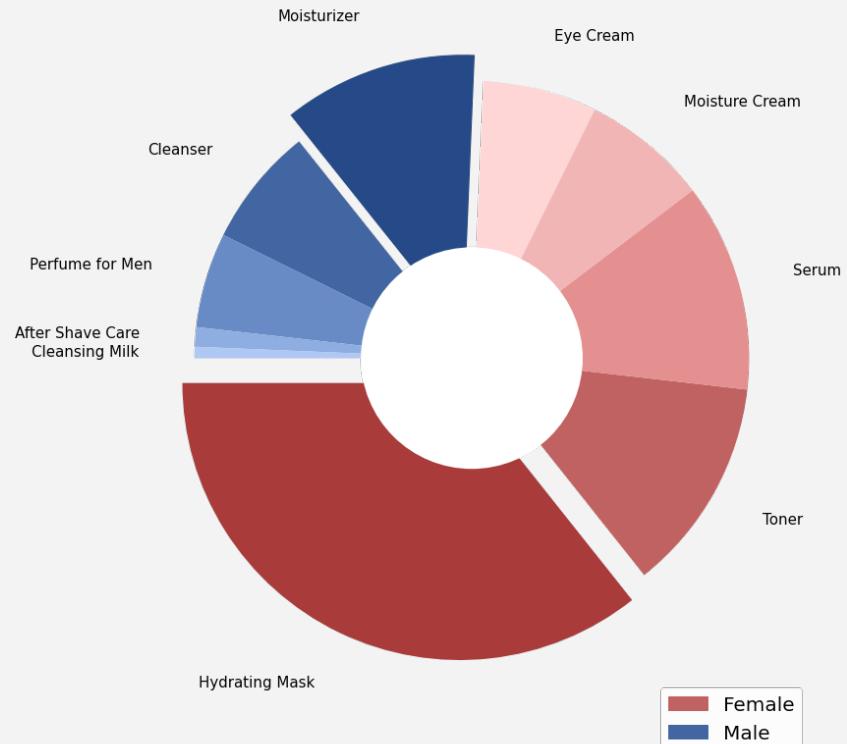
On average, female products were sold **twice** as much as male products



The skincare market is still female-dominated in Hong Kong



Top 5 Categories – Male & Female

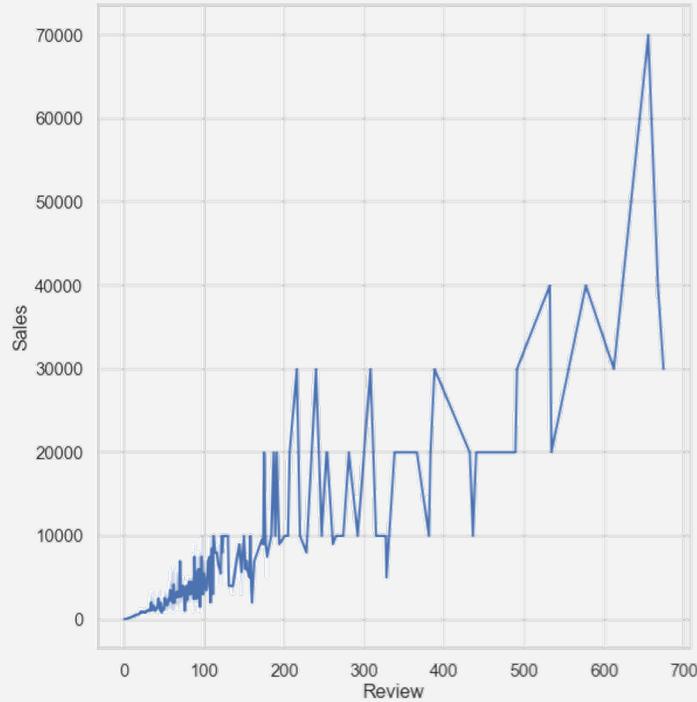


Top 5 Categories for Male and Female by proportion of Sales



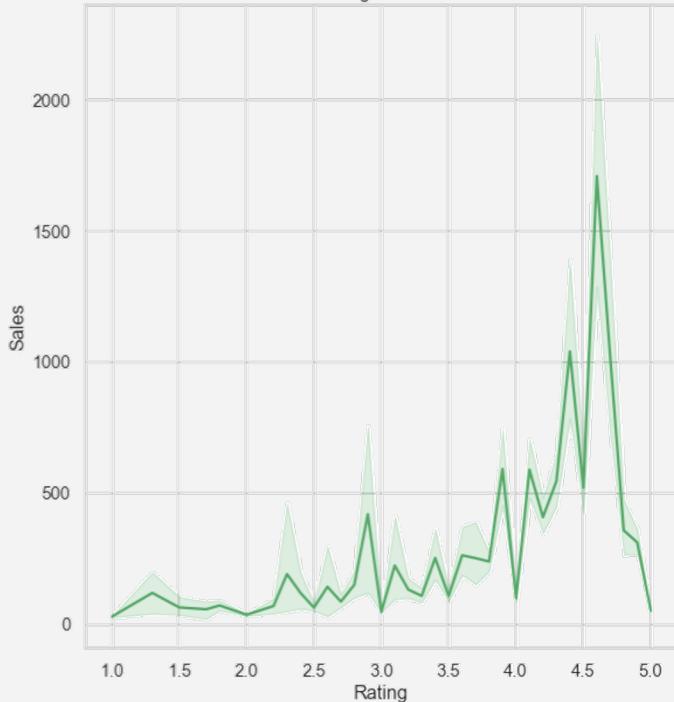
Reviews & Ratings

Review vs Sales



Number of reviews and number of sales are
strongly positively correlated

Rating vs Sales



There is a decline in sales seen for products with ratings
above 4.7



Reviews & Ratings

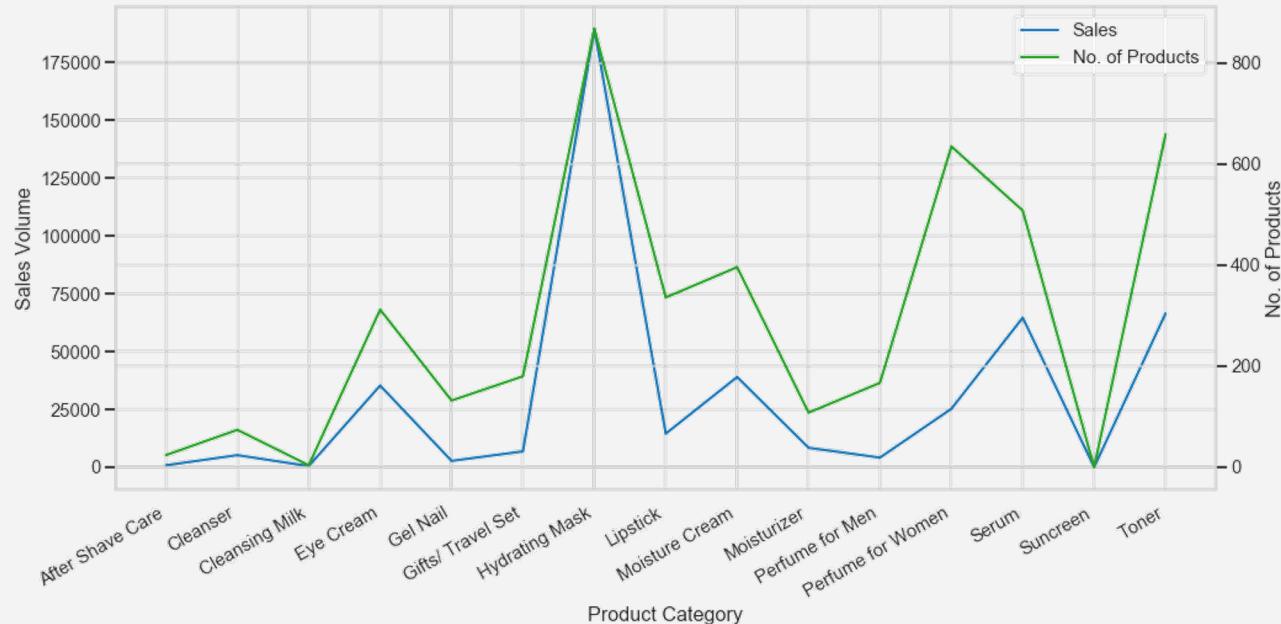


Correlation Heatmap for Reviews, Sales and Ratings



Sales vs Product Variety

Comparison of Sales & No. of Products by Product Category

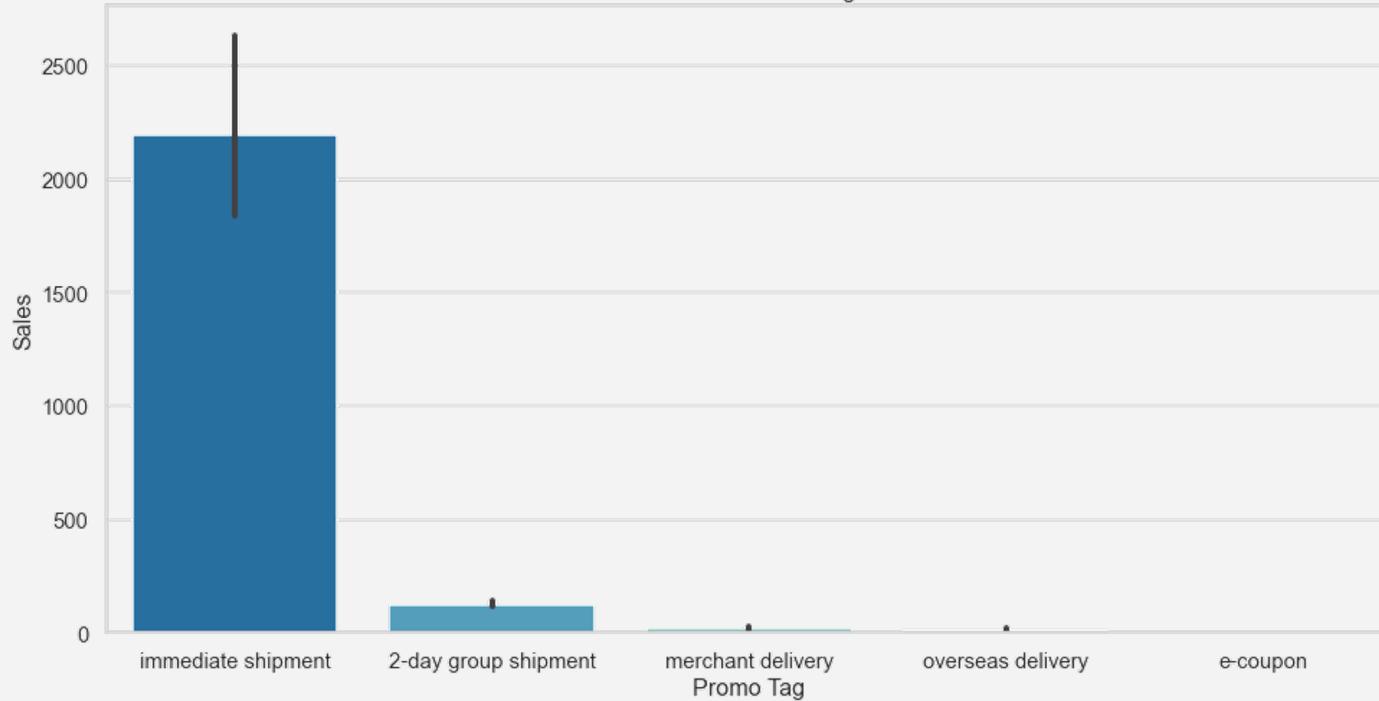


There are a **limited number** of Hydrating Mask products relative to its **high demand**



Promotions

Sales vs. Promo Tag

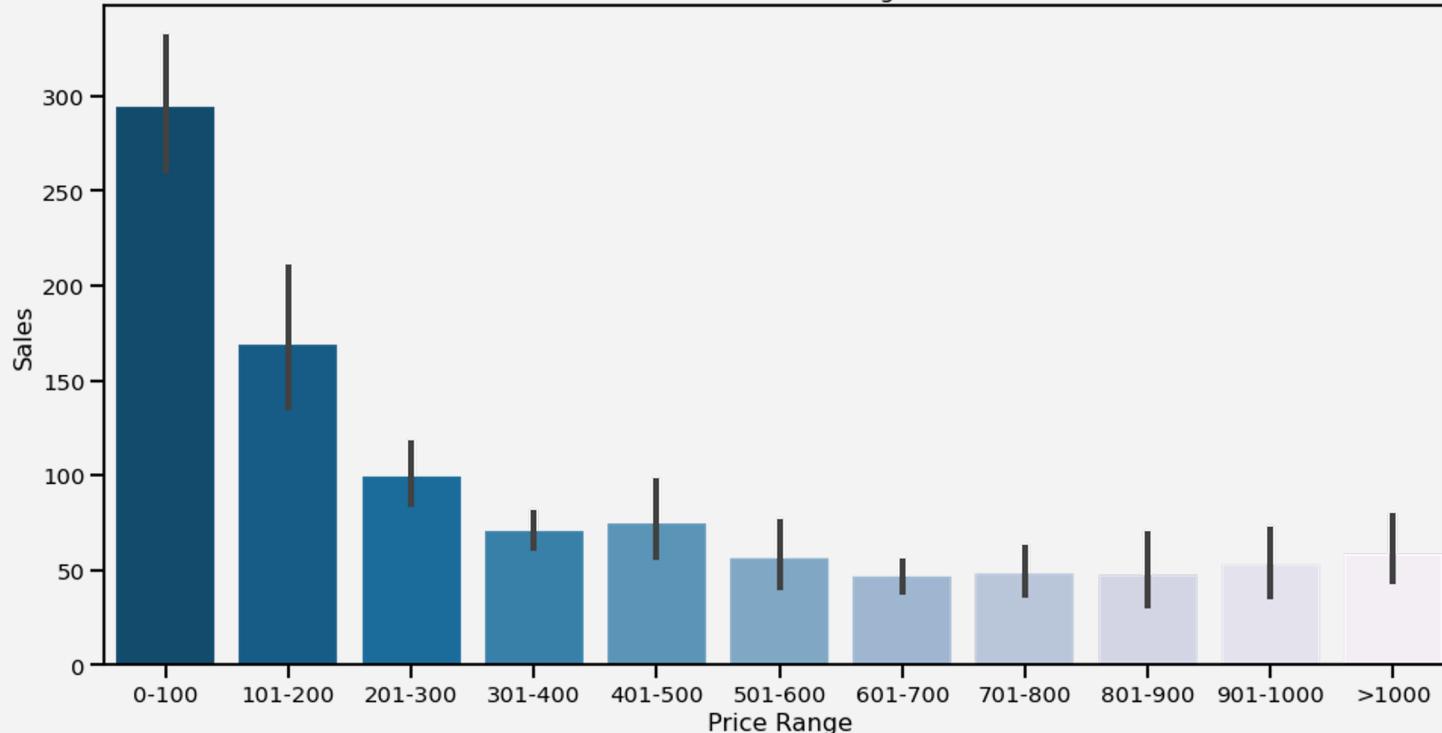


Customers value **immediate delivery**



Pricing

Sales vs. Price Range



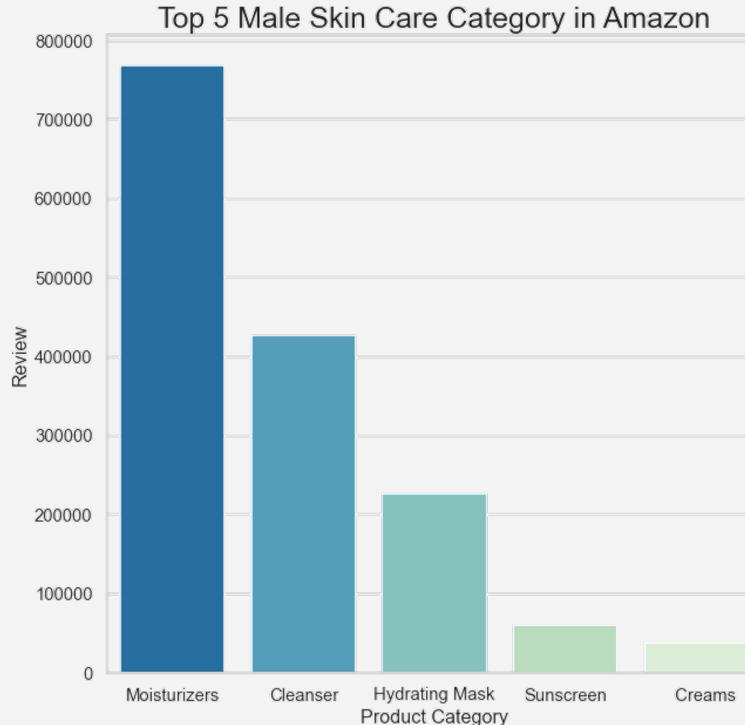
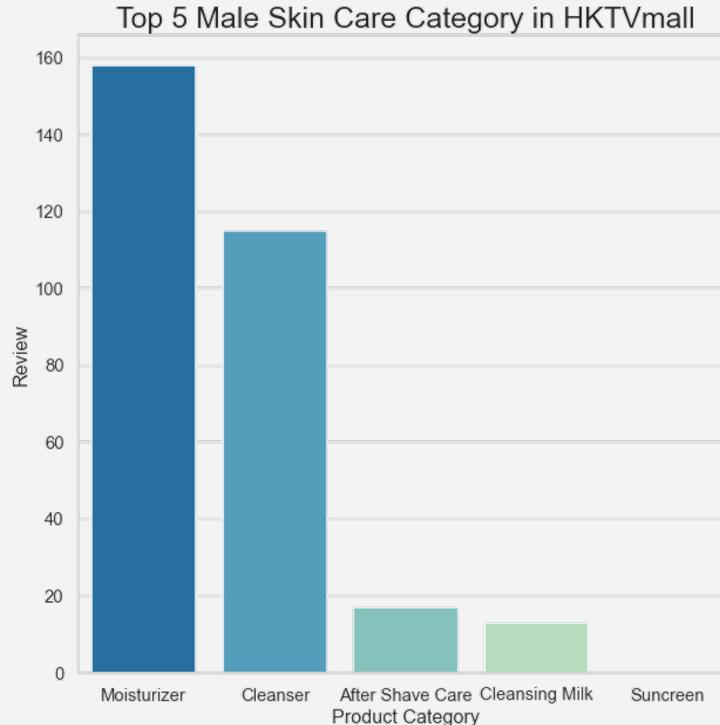
Low price elasticity above a certain price i.e. \$300

Hong Kong vs. Worldwide?





HKTVMall vs. Amazon



Top-selling categories are **very similar** between Hong Kong and globally



Key Insights

+

01

Female products are purchased **twice** as often as male products

+

02

There is a decline in sales seen for products with ratings **above** 4.7

+

03

There are a **limited number** of Hydrating Mask products relative to its **high demand**

+

04

Low price elasticity above a certain price i.e. \$300

+

05

Customers value **immediate delivery**

+

06

Top-selling categories are **very similar** between Hong Kong and globally

Business Recommendations

1. Prioritize immediate shipment of products

- Ensure majority of products are available for immediate delivery
- Increase promotion of ‘immediate shipment’ products

2. Increase number of hydrating mask products

- There is an insufficient variety of masks relative to its high demand

3. Recommend products with ratings up to a maximum of 4.7

- Similar findings in recent study

\ Variables: Average Star Ratings

4. Five Stars is “Too Good to be True”

More stars equals more sales, right? It seems fairly intuitive that higher ratings would lead to higher sales. Our research found that this is true—but only up to a point.

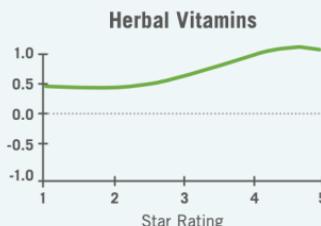
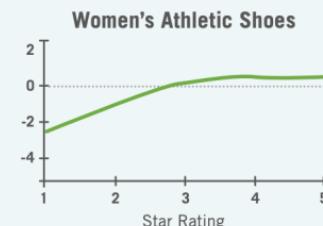
Across product categories, we found that purchase likelihood typically peaks at ratings in the 4.0 - 4.7 range, and then begins to decrease as ratings approach 5.0. Put differently, products with an average star rating in the 4.7 - 5.0 range are less likely to be purchased than those in the 4.2 - 4.7 range.

This suggests that shoppers see ratings at the far end of the spectrum as “too good to be true.”

FINDING THE SWEET SPOT

The optimal star rating varied across product categories. But what is most interesting to note is that in no category was the optimal rating 5.0.

Star Ratings' Impact on Purchase Probability



Maslowska, E., Malthouse, E.C., and Bernritter, S. (2017). “Too Good to be True: The Role of Online Reviews’ Features in Probability to Buy.” International Journal of Advertising, 36(1) 142-163.



Challenges & Next Steps

Challenges

- No sales data on Amazon
- Amazon product list too large
- Errors when web scraping
- Majority of products are under 'Others' category
- Choosing the right graphs to represent data

Next Steps

- Categorise the products to be more accurate
- Explore other websites
- Explore other markets outside of skincare



Questions



◦ + +