

# eCommerce Analytics- Project Report

## Table of Contents

<b><i>Problem Statement</i></b> .....	<b>2</b>
<b><i>Assumptions</i></b> .....	<b>2</b>
<b><i>Project Design</i></b> .....	<b>3</b>
<b><i>Analysis</i></b> .....	<b>4</b>
<b>Data issues</b> .....	<b>4</b>
<b>Price &amp; Ratings</b> .....	<b>5</b>
Inference based on Top 5 Brands summary stats .....	5
<b>Sponsored Listing</b> .....	<b>7</b>
<b><i>Insights &amp; Recommendations</i></b> .....	<b>8</b>
<b>Sneaker Brand Strategy by Business Lifecycle Phase</b> .....	<b>10</b>
<b><i>Overall Summary</i></b> .....	<b>11</b>

## Problem Statement

The scenario for this project is regarding a fictional new entrant (Company XYZ) into the eCommerce space in Singapore market. It seeks to kick-off retail operations by onboarding brands for Mens' shoes.

## Assumptions

The following assumptions were made for this test:

- XYZ is seeking to start retail operations in Singapore
- Competitor to monitor in the Singapore market will be amazon.sg
- Sneakers are the target category as it is a universal necessity
- Men are the target group for this project
- This project will only attempt to scrape 10 pages
- This application is only webscraping from 1 local machine

## Project Design

An initial investigation into amazon.sg webpages' structure was conducted:

- A manual verification via a "Continue shopping" button might appear before the actual amazon.sg page
- The brands and product name are nicely separated per product
- Page filters help to reduce total results returned
- The webscraper won't be able to detect the next button on an amazon.sg page until the page has been scrolled down to the bottom

Based on the above findings, webscraper (webscrape\_amazon.py) will be designed as such:

- Events:
  - 1) Pre-landing Page "Continue shopping" verification
  - 2) Arrival on Landing Page
  - 3) Execution of Search term
  - 4) Apply filters
  - 5) Scrape products
  - 6) Paginate
  - 7) Repeat steps 5 and 6
- A random sleep time will be executed between each event to mimic human-like browsing behaviour during scraping.
- Variables such as search term, website URL, pages to scrape and gender filters are externalized in a config.yaml file to enable reusability for other products of interest
- Type-hinting will be implemented to perform static analysis and error detection in code
- Logging will also be in place to allow better debugging and troubleshooting
- Scraped data will output in this format: <search\_term>\_YYYY-MM-DD\_HHMM.csv

## Analysis

Based on the 10 pages scraped, a total of 535 records were collected. This data was subsequently loaded into an analysis script (ETL & EDA.ipynb) for ETL and EDA purposes.

## Data issues

Raw data scraped from amazon.sg was subsequently cleaned up via

- Reducing to relevant columns only
- Removing any non-alphanumeric characters from 'brand' column
- Extracting only price from 'price' column
- Extracting only rating score from 'rating' column

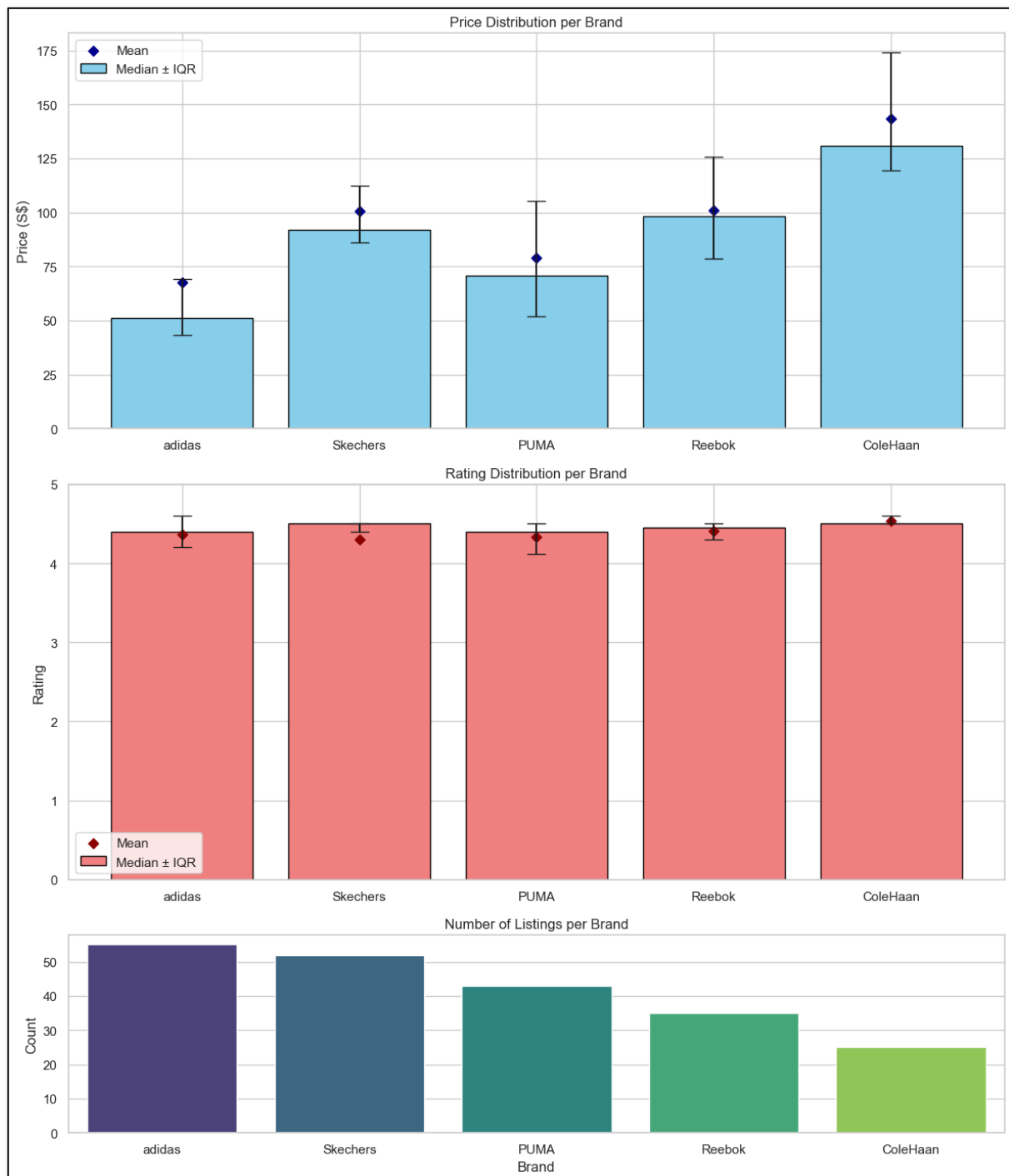
Before data cleansing:

	prodTypeID	brand	title	price	rating	sponsored	link
0	B08PKD3ZYL	Nike	Women's Offcourt Trainers	S\$89.00	4.5 out of 5 stars	Yes	<a href="https://www.amazon.sg/sspa/click?ie=UTF8&amp;spc=M...">https://www.amazon.sg/sspa/click?ie=UTF8&amp;spc=M...</a>
1	B0C5WSQHPX	VANZACK	Sneakers for Girls Espadrille Sneakers Women C...	S\$7.87	NaN	Yes	<a href="https://www.amazon.sg/sspa/click?ie=UTF8&amp;spc=M...">https://www.amazon.sg/sspa/click?ie=UTF8&amp;spc=M...</a>
2	B09VT15Z95	adidas	Grand Court TD Lifestyle Sneakers	S\$54.75	4.3 out of 5 stars	No	<a href="https://www.amazon.sg/Adidas-Grand-Sneakers-Fo...">https://www.amazon.sg/Adidas-Grand-Sneakers-Fo...</a>
3	B09VSZKVJ1	adidas	Advancourt Base Lifestyle LIT48 Men's Sneakers	S\$43.33	4.1 out of 5 stars	No	<a href="https://www.amazon.sg/Adidas-EE7690-Sneakers-f...">https://www.amazon.sg/Adidas-EE7690-Sneakers-f...</a>
4	B0B691P2Y4	adidas(アディダス)	Advan Court Base Lifestyle Men's Sneakers	S\$43.20	4.2 out of 5 stars	No	<a href="https://www.amazon.sg/Adidas-EE7690-Sneakers-F...">https://www.amazon.sg/Adidas-EE7690-Sneakers-F...</a>

After data cleansing:

	prodTypeID	brand	title	price	rating	sponsored
0	B08PKD3ZYL	Nike	Women's Offcourt Trainers	89.00	4.5	Yes
1	B0C5WSQHPX	VANZACK	Sneakers for Girls Espadrille Sneakers Women C...	7.87	NaN	Yes
2	B09VT15Z95	adidas	Grand Court TD Lifestyle Sneakers	54.75	4.3	No
3	B09VSZKVJ1	adidas	Advancourt Base Lifestyle LIT48 Men's Sneakers	43.33	4.1	No
4	B0B691P2Y4	adidas	Advan Court Base Lifestyle Men's Sneakers	43.20	4.2	No

## Price & Ratings



## Inference based on Top 5 Brands summary stats

### Price Insights

- **ColeHaan** is clearly the premium brand here, with the highest average price (\$143.43), median price (\$131.04), and wide upper quartile (\$174.25). This indicates they target higher-end sneakers.

- **Reebok** and **Skechers** also tend toward higher prices with means over \$100. Their quartiles suggest more consistency but still pricey models.
- **PUMA** sits in the mid-range price bracket (mean about \$79), with a notably wide spread between 25th and 75th percentiles (\$51.86 to \$105.24), indicating more variety in product price tiers.
- **adidas** has the lowest average price (\$67.79) among the top 5, with a relatively tight interquartile range (\$43.26 to \$69.11), suggesting most products are budget or mid-range priced.

### *Rating Insights*

- **ColeHaan** has the highest average rating (4.54) and also the highest lower quartile rating (4.50), showing strong, consistently positive customer feedback.
- **Reebok** follows with a similarly high rating mean of 4.41, with the ratings tightly clustered between 4.30 and 4.5.
- **PUMA**, **adidas**, and **Skechers** have similar mean ratings in the low 4.3s to 4.36, indicating overall good but slightly less stellar customer satisfaction than ColeHaan and Reebok.
- The narrow interquartile ranges for ratings across all brands indicate customer ratings are fairly consistent within each brand.

### *Volume & Popularity*

- **adidas** has the highest number of products listed (count 55), closely followed by **Skechers** (52), then **PUMA** (43), **Reebok** (35), and finally **ColeHaan** (25).
- Higher counts combined with moderate prices (e.g., adidas, Skechers) suggest these brands focus on volume and a broad product offering.
- Brands like ColeHaan and Reebok, with fewer products but higher prices and ratings, may be positioning themselves as premium or niche options.

### Summary of the Top 5 listed sneaker brands on amazon.sg

- adidas: Budget-friendly, high variety, solid ratings.
- Skechers & Reebok: Higher price range with solid customer satisfaction.
- PUMA: Mid-range prices with more price variability.
- ColeHaan: Premium brand with consistently high ratings but fewer products.

## Sponsored Listing

The Top 3 brands that utilized sponsored listing are (in order): SAFIGLE, VANZACK, NIKE

```
cleaned_df[cleaned_df[['brand', 'sponsored']]['sponsored'] == "Yes"]['brand'].value_counts()

brand
SAFIGLE    19
VANZACK    12
Nike       10
FRCOLOR     9
Lscsdf      2
AB00FAN     1
ORFOFE      1
Yardwe      1
Name: count, dtype: int64
```

## Insights & Recommendations

Insight	Explanation	Implication	Company XYZ Strategy
Opportunity in Budget-to-Mid Market	adidas has <b>the most listings, tight price range</b> , and solid ratings.	There's strong demand for <b>affordable, reliable sneakers</b> - ideal for mass market appeal.	<ul style="list-style-type: none"> <li>Target this <b>high-volume, mid-range</b> segment.</li> <li>Launch with <b>adidas and Skechers SKUs</b> (or their direct competitors).</li> <li><b>Highlight affordability + reliability</b> in marketing.</li> <li>Use this as a foot-in-the-door before introducing premium lines.</li> </ul>
White Space in Premium Segment	ColeHaan stands out as a <b>premium brand</b> with <b>few products but very high ratings</b> .	Singapore's market includes a <b>niche premium buyer segment</b> looking for quality over quantity.	<ul style="list-style-type: none"> <li>Secure <b>exclusive partnerships</b> with premium labels like ColeHaan.</li> <li>Position XYZ as a <b>curated destination</b> for discerning sneaker buyers.</li> <li>Bundle with fast delivery and return guarantees to reduce purchase friction.</li> </ul>
Data-Led Category Expansion	PUMA has a <b>wide price spread</b> , indicating product tiering flexibility.	Consumers may be receptive to <b>tiered pricing</b> within a single brand.	<ul style="list-style-type: none"> <li>Emulate this flexibility in-house (e.g., through private labels or price-tiered SKUs).</li> <li>Optimize <b>cross-selling</b> by offering entry-level and premium sneakers within the same brand family.</li> </ul>
Brand Strategy & Stocking Decisions	adidas and Skechers dominate in volume.	They are likely <b>easier to stock, better-known</b> , and <b>lower risk</b> .	<ul style="list-style-type: none"> <li>Prioritize volume brands for <b>initial inventory</b>.</li> <li>Test premium brands in <b>limited drops or flash sales</b>.</li> <li>Let user feedback and CTR/sales data guide future SKU depth.</li> </ul>



Local Relevance	<p>Singapore has:</p> <ul style="list-style-type: none"> <li>A large <b>brand-conscious but price-sensitive middle class</b>.</li> <li>A rising fitness and lifestyle culture (especially post-COVID).</li> </ul>	Brands offering <b>value and comfort (like Skechers)</b> will likely perform well.	<b>Differentiate with faster delivery, smart pricing, and dynamic inventory</b> from its logistics core.
Sponsored Listings: Aggressive vs. Organic Visibility	Brands with High Organic Ratings + Low Sponsorship (e.g., ColeHaan, Reebok)	These brands <b>earn trust through quality</b> , not ads.	<b>Partner with them as underdog premium brands</b> and position them with <b>trust-centric messaging</b> (e.g., "Highly rated by real users, no gimmicks").
	Brands Heavily Relying on Sponsored Listings (e.g., SAFIGLE, VANZACK)	<b>Risk of over-reliance on ads</b> to drive traffic.	<ul style="list-style-type: none"> <li><b>Vet product quality before listing.</b></li> <li>Use them to <b>test ad auction performance</b>, but not anchor the sneaker category on them.</li> <li>Consider them as <b>low-cost filler inventory</b>, not brand leaders.</li> </ul>
	<p>NIKE as a Special Case</p> <p>Although a top sponsor, NIKE is a <b>global brand with strong baseline demand</b>.</p>	<p>Sponsorship may signal:</p> <ul style="list-style-type: none"> <li>Market testing in Singapore.</li> <li>Boosting specific models.</li> </ul>	<b>Prioritize NIKE</b> , especially for flagship SKUs, and <b>mirror their sponsored strategy</b> in its own marketplace to learn bidding dynamics.

## Sneaker Brand Strategy by Business Lifecycle Phase

Business Phase	Objective	Recommended Brands	Strategy
<b>Entry Point</b>	Rapidly acquire users and gain market share	adidas, Skechers, PUMA	Offer high-volume, mid-priced brands with good ratings. Focus on <b>variety, value, and trust</b> .
<b>Premium Position</b>	Capture the upper-income and fashion-conscious segment	ColeHaan, Reebok, NIKE	Highlight <b>quality and brand equity</b> ; use influencer seeding and premium landing pages.
<b>Product Tiering</b>	Serve multiple segments across price points	adidas, PUMA, NIKE	Showcase products in <b>budget, mid, and premium</b> ranges to maximize basket diversity.
<b>Brand Focus</b>	Partner closely with high-performing, brand-consistent sellers	Skechers, Reebok, ColeHaan	Build <b>long-term seller relationships</b> with strong brands and high average ratings.
<b>Differentiation</b>	Stand out from competitors (e.g., Lazada, Zalora)	NIKE (sponsored), ColeHaan (premium), SAFIGLE (aggressive ads)	Leverage <b>exclusive drops, ad partnerships, and premium packaging or shipping experiences</b> .

## Overall Summary

- **adidas** is ideal for **Entry Point** and **Tiering** due to its budget-friendly catalog and volume.
- **ColeHaan** is perfect for **Premium Positioning** and **Brand Focus**—fewer SKUs, high ratings.
- **NIKE** fits several roles, including **Differentiation** and **Premium**, especially with its aggressive sponsored listings.
- **Skechers** offers stable quality for **Entry**, but is also a good brand for **Focus** with its wide appeal.
- **SAFIGLE** is only suitable if XYZ wants to experiment with **Differentiation via ad partnerships**, though the product trust is low.