**Promotional Strategies Analysis**

Murtuja Afshar

Debug-Tech Task 4

June 27, 2025

Objective:

The objective of this project is to analyze competitors’ pricing strategies, product lines, and promotional tactics by collecting, cleaning, and analyzing product data from brands such as Uniqlo and Levi’s. This analysis aims to uncover patterns in pricing and product popularity to support strategic business decisions.

Data collection:

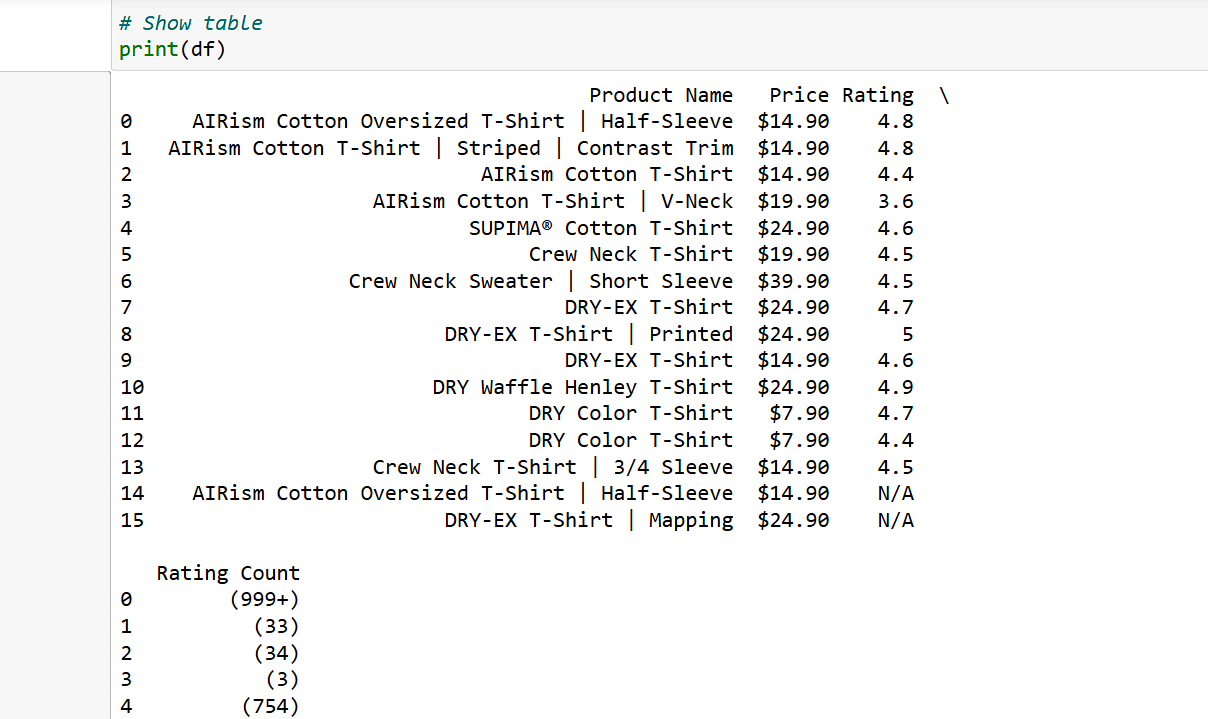
Product data for this project was **manually collected** from brand websites and structured into two datasets:

* **Uniqlo product listings:** including Product Name, Price, Rating, and Rating Count.
* **Levi’s product listings:** including Product Name, Price, Rating and Rating Count.

Uniqlo data:

To collect product data from **Uniqlo’s website**, a **Python-based Selenium web scraping script** was used.  
The script automated the following process:

* Opened the **Uniqlo Men’s T-shirts** category page using a **Chrome WebDriver**.
* Waited for the page to fully load dynamic content.
* Located product containers using **CSS selectors**.
* Extracted key product details:
  + **Product Name**
  + **Price**
  + **Rating**
  + **Rating Count**
* Handled missing values gracefully by assigning 'N/A' where necessary.
* Stored the collected data into a **Pandas DataFrame** for further cleaning and analysis.



Levi’s Data:

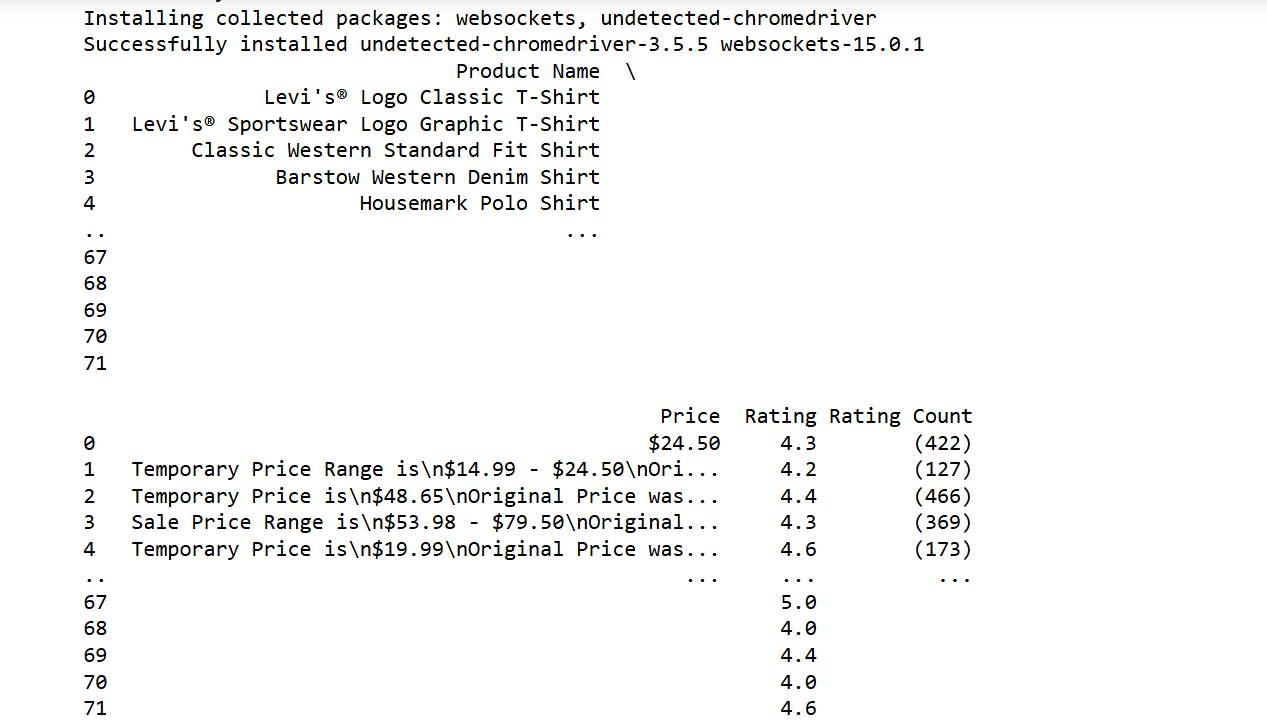
For collecting product data from the **Levi’s USA Men’s Shirts category page**, a **Python-based web scraping script using Undetected ChromeDriver (uc)** was implemented.  
**Undetected ChromeDriver** was used because Levi’s website employs bot detection measures that can block regular Selenium scrapers.

**Process:**

* Launched an **undetectable Chrome browser session** using undetected\_chromedriver.
* Navigated to the **Men’s Shirts** product page.
* Waited for dynamic content to load.
* Located product containers via **CSS selectors**.
* Extracted:
  + **Product Name**
  + **Price**
  + **Star Rating** (calculated from a CSS style width percentage converted to a 5-star scale)
  + **Rating Count**
* Managed missing or unavailable fields with 'N/A' placeholders.
* Compiled the data into a **Pandas DataFrame** for subsequent cleaning and analysis.

After scraping, the Levi’s product data required basic cleaning before analysis:

* **Replaced newline characters (\n) in the Price column** with a **| separator** to make multi-line price details readable in one line.
* **Cleaned up Rating Count values** by removing the enclosing parentheses.
* **Saved the cleaned dataset to a CSV file** for further analysis:



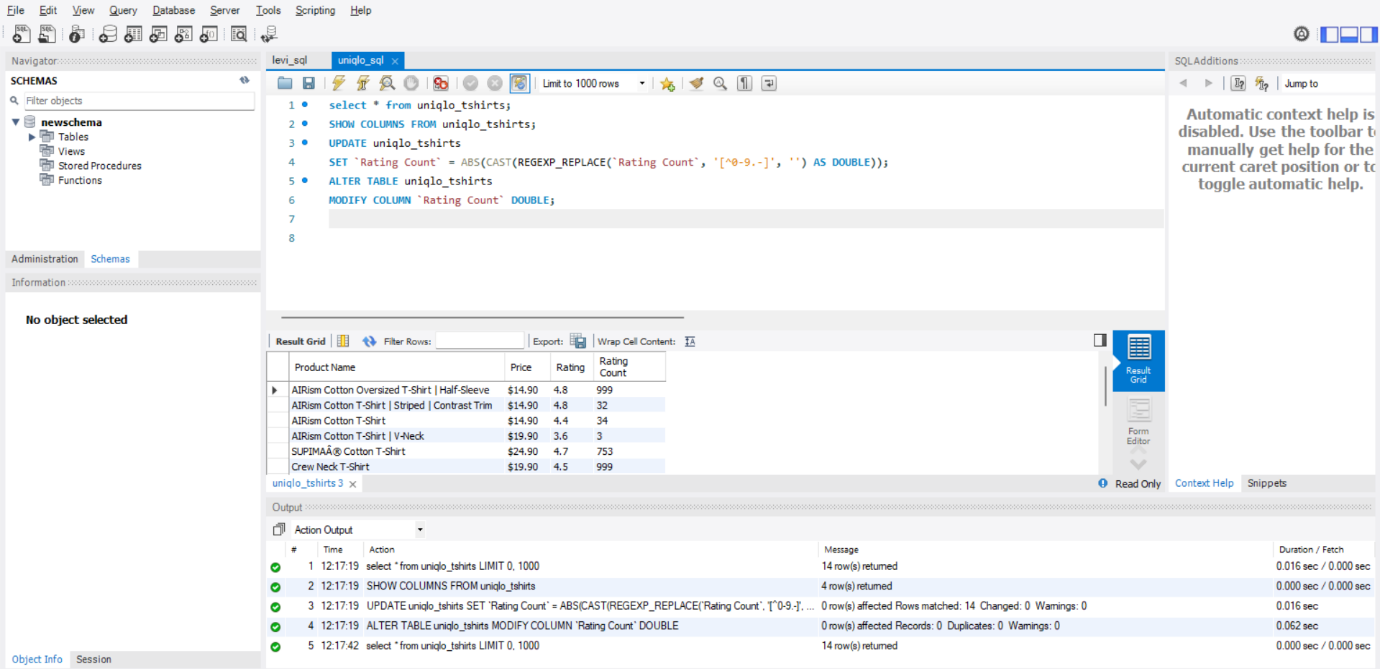
Data Cleaning (using MySQL):

Uniqlo Dataset:

After loading the raw Uniqlo product data into my MySQL table, I noticed that the Rating Count column had extra characters like parentheses and possibly other symbols mixed in with the numbers. To clean this up, I first ran a simple SELECT \* to view the data and a SHOW COLUMNS to confirm the data types.

Then, I used an UPDATE statement with REGEXP\_REPLACE to strip out anything that wasn’t a digit, a decimal, or a negative sign — basically cleaning out unwanted characters. I wrapped that in a CAST to convert the cleaned text into a numeric value, and then used ABS() to make sure all values were positive, just in case any stray negatives slipped in.

Once the data values were clean, I modified the column type from its original string format to a DOUBLE, so it could be treated as a proper numeric field for analysis later on.



Levi’s Dataset:

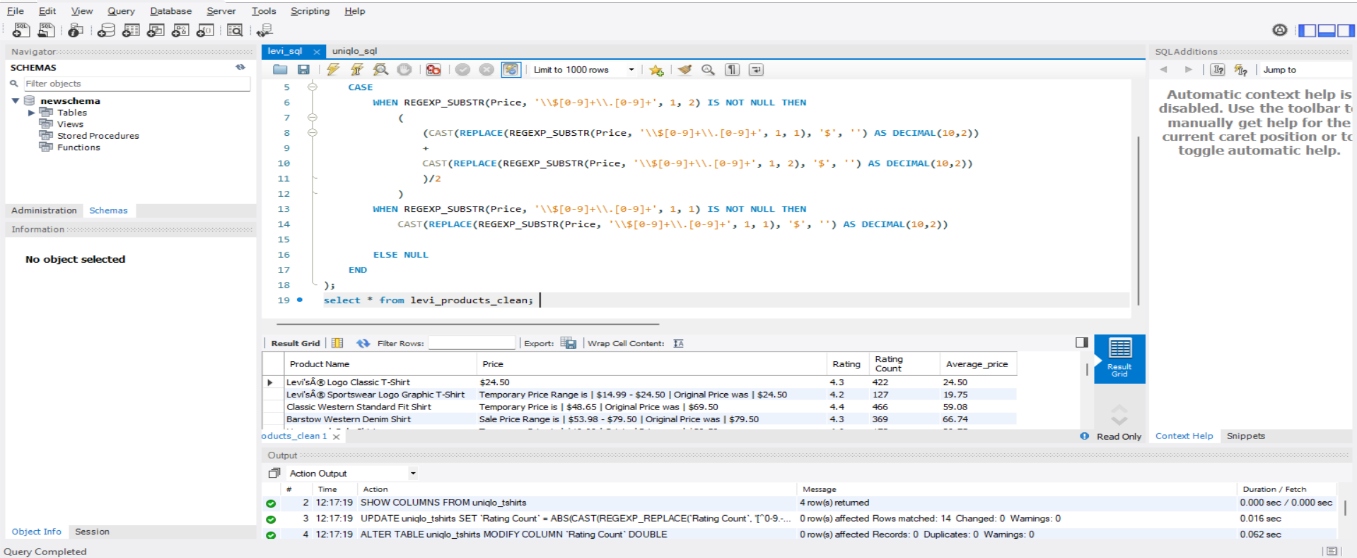
Once I imported the Levi’s product data into my MySQL table, I noticed the Price column was messy because some products had single prices while others had promotional ranges like $24.99 - $39.99. To prep this for analysis, I first ran a SELECT \* just to review the data.

Then, I added a new column called Average\_price to the table using ALTER TABLE, since I wanted a clean, numeric value to represent each product’s average price for analysis.

To populate this Average\_price field, I wrote an UPDATE statement using REGEXP\_SUBSTR to extract the first and second dollar values from the Price string:

* **If two prices were found** (like a range), I cleaned the dollar signs out and converted them to decimals, then averaged them.
* **If only one price was found**, I just extracted and converted that.
* **If no price matched the pattern**, I left it as NULL.

Finally, I ran another SELECT \* to confirm that the Average\_price column was properly filled with clean numeric values ready for clustering and visualizations.



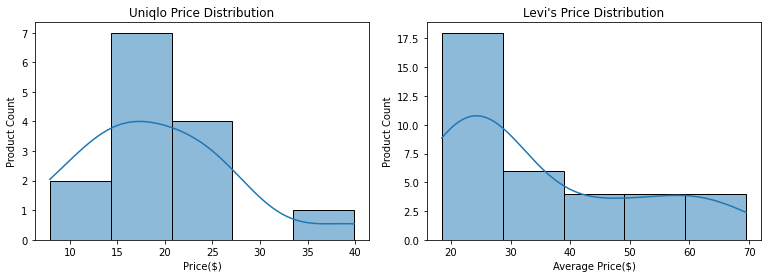
Data Anaysis:

Exploratory Data Analysis (EDA):

I start by importing pandas as pd because it provides powerful tools to read, manipulate, and analyze data easily.  
I use pd.read\_csv() to load the two datasets from their file paths into uniqlo\_df and levis\_df. This converts the CSV files into tabular DataFrames that I can work with in Python.  
By calling .head() on each DataFrame, I preview the first five rows to understand the structure and content of the data.  
In uniqlo\_df, the 'Price' column likely contains price values as strings including dollar signs (e.g., "$19.99"). I remove these dollar signs and commas by replacing them with an empty string using .replace('[\$,]', '', regex=True). Then, I convert the cleaned strings to floats with .astype(float). This way, I can perform numeric calculations on the prices.  
Using .info() on each DataFrame, I check the number of rows, column names, data types, and if there are any missing values. This helps me understand if the data loaded correctly and if further cleaning is needed.  
Finally, by calling .describe() on both DataFrames, I get summary statistics like count, mean, standard deviation, min, max, and quartiles for numeric columns. This gives me a quick overview of the distribution and range of the data.

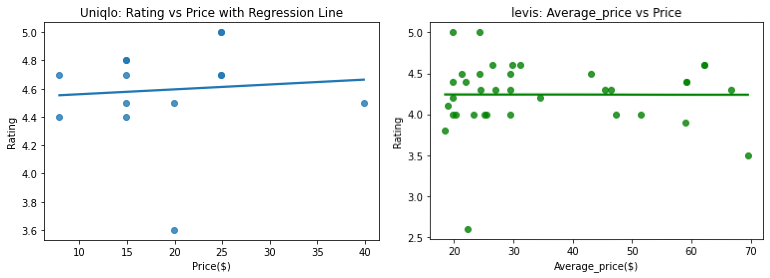
I looked at the price distributions for Uniqlo and Levi’s products using histograms.

* A histogram breaks down prices into groups (bins) and shows how many products fall into each price range.
* I chose 5 bins to see broad price ranges.
* Adding a smooth curve (KDE) helps visualize the overall shape of the price distribution.
* This helped me understand whether prices cluster around certain values or are spread out.



I explored the relationship between product price and customer rating with regression plots.

* I plotted prices on the x-axis and ratings on the y-axis to see if higher prices correspond to better or worse ratings.
* The regression line shows the overall trend, ignoring noise from individual points.
* This way, I could check if there is a positive, negative, or no correlation between price and rating for each brand.



I analyzed promotions in the Levi’s dataset by creating a new “Promotion” column.

* I checked each product’s price description to see if it mentioned “Temporary” or “Sale.”
* If yes, I labeled it as a promotional price; otherwise, it’s regular.
* Then, I counted how many products were on promotion versus regular prices using a count plot.
* This visualization shows the balance between discounted and regular-priced items in the Levi’s dataset.



K-Means Custering:

I prepared the data for clustering.

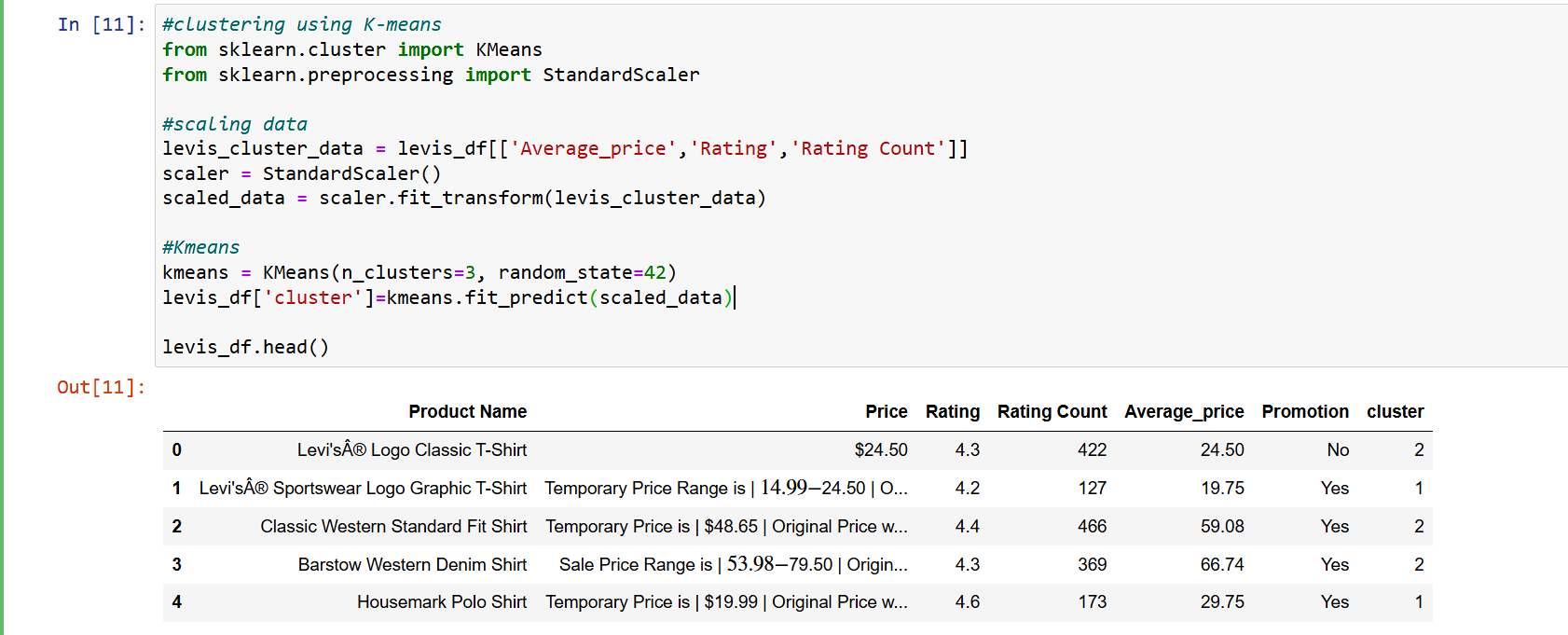
* I selected specific numeric columns: for Levi’s, *Average Price*, *Rating*, and *Rating Count*; for Uniqlo, *Price*, *Rating*, and *Rating Count*.
* These columns were chosen because clustering works best on quantitative variables that can meaningfully define groups.

I standardized the data.

* Since the values in these columns are on different scales (like prices in dollars and ratings out of 5), I used StandardScaler to convert them to the same scale — where each feature has a mean of 0 and a standard deviation of 1.
* This ensures that no one variable dominates the clustering process due to its scale.

I applied K-Means clustering.

* I chose to create 3 clusters to group similar products together based on their pricing, rating, and rating count behavior.
* K-Means assigns each product to one of these clusters based on how similar it is to the cluster’s center.
* After clustering, I stored the cluster number for each product in a new column called ‘cluster’ in both Levi’s and Uniqlo datasets.



**Cluster Visualization**

To understand how the clusters behave, I visualized them with strip plots.

* I plotted product *prices on the x-axis* and *ratings or rating counts on the y-axis*.
* Each point represents a product, and its cluster group is indicated by color.
* This makes it easy to see how products group together — for example, whether higher-priced products tend to have better ratings, or if certain clusters have more customer feedback.

I used two strip plots for each brand:

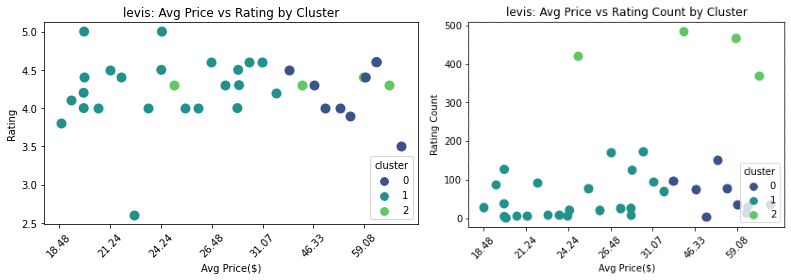
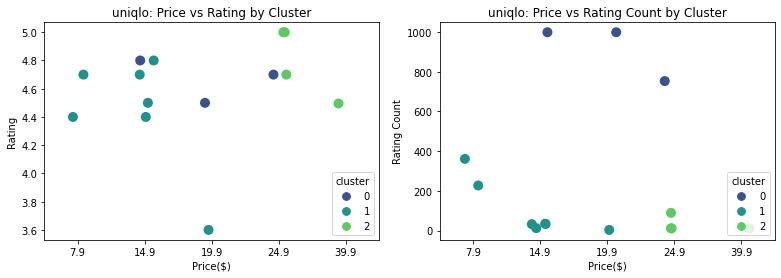
* One showing how *Price (or Average Price) relates to Rating*, by cluster.
* Another showing how *Price (or Average Price) relates to Rating Count*, by cluster.

I added jitter to the points to avoid overlap and made them larger for clarity.

* Legends and axis labels help interpret which cluster is which and what each axis represents.

Finally, for Levi’s plots, I rotated the x-axis labels and adjusted the layout.

* This was necessary because price values might be long or close together, and rotating them improves readability.



Conclusion:

The exploratory data analysis (EDA) and clustering results provide a detailed understanding of Uniqlo and Levi’s pricing, product lines, and customer segments, using clusters as the core separation criterion. For Uniqlo, clustering reveals three distinct groups: Cluster 0, dominated by mid-range priced items ($14.9-$19.9) with high rating counts (up to 1000) and ratings around 4.4-4.6, indicating a strong base of affordable, popular products like T-shirts; Cluster 1, spanning $7.9-$24.9 with moderate ratings (4.2-4.6) and review counts (200-600), suggesting a broader accessible range; and Cluster 2, featuring higher-priced items ($24.9-$39.9) with slightly better ratings (4.6-4.8) and lower review counts (200), pointing to premium offerings. Price distribution confirms a focus on affordability, with most products under $25, aligning with Uniqlo’s strategy of quality basics. For Levi’s, clustering identifies Cluster 0 with mid-range average prices ($18.48-$46.48) and high review counts (100-300) with stable ratings (4.0-4.5), reflecting a core segment of moderately priced denim and casual wear; Cluster 1, covering $21.24-$59.08 with ratings (3.5-4.5) and counts (50-200), indicating a mix of standard and premium products; and Cluster 2 with fewer points and mixed ratings, suggesting niche or promotional items. The price range spans $18.48-$59.08, with a notable peak at $20-$30, and a significant portion of products on promotion (25 vs. 5 non-promotional), highlighting a strategy balancing regular and discounted offerings.

Regarding promotional strategies, Uniqlo employs a robust approach leveraging email newsletters, discounts, and bundle deals. The brand offers a $10 off coupon for first-time email subscribers on orders of $75 or more, with additional incentives like 20-30% seasonal discounts on items like HEATTECH and Ultra Light Down, as noted on their website and corroborated by sources like couponfollow.com and uniqlo.com. Bundle deals, such as multibuy offers on essentials like socks, are promoted online, encouraging bulk purchases. Uniqlo also uses the app and social media for exclusive promotions, including limited-time sales during Black Friday and Cyber Monday, enhancing customer engagement. Levi’s focuses on email newsletters and discounts, offering 10% off the first order plus free shipping for subscribers, with additional 15% discounts for students, military, and first responders via the Levi’s RedTab program, as seen on levi.com and couponfollow.com. Promotional strategies include frequent sales (e.g., up to 30% off) and bundle-like deals through the RedTab membership, which provides birthday rewards and free hemming. However, specific bundle deals are less emphasized compared to Uniqlo.

Suggestions:

**Target Cluster-Specific Promotions**:

* For Uniqlo, focus on Cluster 0 ($14.9-$19.9), which has high rating counts (up to 1000) and ratings (4.4-4.6). Offer bundle deals (e.g., buy 2 T-shirts, get 20% off) via email newsletters to capitalize on this popular segment, aligning with Uniqlo’s current multibuy strategy (uniqlo.com). For Cluster 2 ($24.9-$39.9), introduce premium-focused promotions like limited-time 15% discounts on higher-rated items to boost engagement, given their lower review counts.
* For Levi’s, target Cluster 0 ($18.48-$46.48) with promotions like a 10% off + free shipping deal for first-time subscribers (levi.com), leveraging its high review counts (100-300). For Cluster 1 ($21.24-$59.08), offer seasonal sales (e.g., 30% off denim) to attract the broader price range, building on their existing discount frequency.

**Incorporate Competitive Promotional Tactics**:

* Emulate Uniqlo’s successful email and app-based campaigns, such as the $10 off first order for $75+ (couponfollow.com), to build a subscriber base. Add exclusive app-only bundle deals to enhance customer retention.
* Adopt Levi’s RedTab membership perks (e.g., birthday rewards, free hemming) and extend student/military discounts (15% off, levi.com) to new segments, increasing loyalty.

**Leverage Seasonal and Urgency Tactics**:

* Use time-sensitive promotions during peak seasons (e.g., Black Friday, Cyber Monday for Uniqlo; levi.com notes similar sales), as these drive urgency and align with current strategies.

**Derive Insights for Product Line Expansion**:

* Uniqlo’s Cluster 2 suggests potential for expanding premium lines (e.g., innovative fabrics), supported by higher ratings (4.6-4.8).
* Levi’s stable ratings across clusters indicate an opportunity to diversify denim offerings within the $20-$40 range (Cluster 0), where engagement is highest.

**Monitor and Adjust with Data**:

* Regularly update cluster analysis with real-time web data (e.g., via scraping tools) to track price and promotion shifts, ensuring insights remain relevant.
* Conduct A/B testing on discount levels (e.g., 10% vs. 20%) to optimize promotional impact, as suggested by data-driven marketing practices.

References:

 Uniqlo promotional details: [Women's, Men's and Kids' Clothing & Accessories | UNIQLO US](https://www.uniqlo.com/us/en/), [Coupon Codes in Real-Time - CouponFollow](https://couponfollow.com/).

 Levi’s promotional strategies: [End of Season Sale - 50% Off Select Styles | Levi's® US](https://www.levi.com/US/en_US/), [Coupon Codes in Real-Time - CouponFollow](https://couponfollow.com/)