A Report on

Myntra Menswear Sales Analysis
with an ecommerce dataset

Project Overview

To optimize sales strategies for the men's bottomwear category, this project utilized a full-stack data analysis approach. By processing the Myntra dataset through **MySQL** and **Python (Jupyter)**, raw sales figures were enriched with derived metrics to uncover hidden value drivers.

These insights were finalized in **Power BI**, creating a visual narrative that guides decision-making regarding pricing models, discount depth, and inventory prioritization.

Dataset Summary

File size: 52120 Rows and 7 Columns

Feature Groups:

Brand & product profile:

- brand_name, pants_description,

Pricing details:

- price, MRP, discount_percent

Customer base and brand credibility:

- ratings, number_of_ratings

Tools Used: Python(numpy, pandas, matplotlib, and seaborn), MySQL, PowerBI

Exploratory Data Analysis (EDA) & Feature Engineering

To transform raw sales figures into actionable metrics for analysis, I first performed a comprehensive Exploratory Data Analysis (EDA) using Pandas.

- **Data Ingestion & Inspection:**

  - Imported pandas and loaded the raw dataset

```
import pandas as pd
```

```
df = pd.read_csv("myntra_dataset.csv")
```

  - Conducted an initial audit of data dimensions (shape), data types, and summary statistics

```
df.shape
```

```
(52120, 7)
```

```
df.dtypes
```

```
brand_name            object
pants_description     object
price                float64
MRP                  float64
discount_percent     float64
ratings              float64
number_of_ratings    float64
dtype: object
```

```
df.describe()
```

	price	MRP	discount_percent	ratings	number_of_ratings
count	52120.000000	52120.000000	52120.000000	52120.000000	52120.000000
mean	1594.515445	3180.398438	1.648256	3.997794	91.568937
std	1495.972325	2201.883218	4.687529	0.420404	433.918513
min	337.000000	499.000000	0.020000	1.000000	5.000000
25%	989.000000	2499.000000	0.400000	3.800000	16.000000
50%	1439.000000	2999.000000	0.500000	4.000000	35.000000
75%	1829.000000	3499.000000	0.630000	4.200000	74.000000
max	54000.000000	72000.000000	64.000000	5.000000	30700.000000

- Performed null value assessment to ensure data integrity

```
df.isnull().sum()
```

```
brand_name           0
pants_description    0
price                0
MRP                  0
discount_percent     0
ratings              0
number_of_ratings    0
dtype: int64
```

- **Feature Engineering:** Four key operational metrics were derived to deep dive into further analysis:
  - **Effective Discount Percent:** To correlate price reductions with demand elasticity

```
# Creating column "effective_discount_percent"

df['effective_discount_percent'] = (
    (df['MRP'] - df['price']) / df['MRP']
) * 100
```

- **Rating Weighted:** To isolate trustworthy, high-performing products (balancing rating score vs. count)

```
# Creating column "rating_weighted"

df['rating_weighted'] = (
    df['ratings'] * np.log(df['number_of_ratings'] + 1)
)
```

- **Value for Money Score:** To guide pricing optimization and competitive positioning

```
# Creating column "value_for_money_score"

df['value_for_money_score'] = (
    df['ratings'] / (df['price'] / 1000)
)
```

- **Popularity Index:** To quantify overall market demand and product momentum

```
# Creating column "popularity_index"

df['popularity_index'] = (
    df['rating_weighted'] +
    df['effective_discount_percent']
)
```

- **Final Validation:** Re-evaluated the transformed dataset's shape and statistics to confirm the structure before exporting for visualization
- Finalized and stored the cleaned data for future use

```
df.to_csv('myntra_menswear.csv', index=False)
```

Exploratory Data Analysis (EDA) in final dataset

To establish a consistent analytical environment for the enriched dataset, the following setup procedures were executed:

- **Environment Setup:** Essential Python libraries were imported, followed by the ingestion of the final, processed dataset

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv("myntra_menswear.csv")
```

- **Data Integrity Check:** The dataset schema was validated to confirm the successful inclusion of all engineered features

```
df.columns
```

```
Index(['brand_name', 'pants_description', 'price', 'MRP', 'discount_percent', 'ratings', 'number_of_ratings', 'effective_discount_percent', 'rating_weighted',
       'value_for_money_score', 'popularity_index'],
      dtype='object')
```

- **Standardization:** Column names were refactored for enhanced readability and consistency

```
#converting MPR column name in lowercase
df.rename(columns={"MRP": "mrp"}, inplace=True)
```

```
#converting column name 'pants_description' to 'products'
df.rename(columns={"pants_description": "products"}, inplace=True)
```

- **Visualization Configuration:** Global plotting parameters were defined using the seaborn library (figure styles (whitegrid), dimensions (10x6), and context options to ensure uniform visualizations throughout the report)

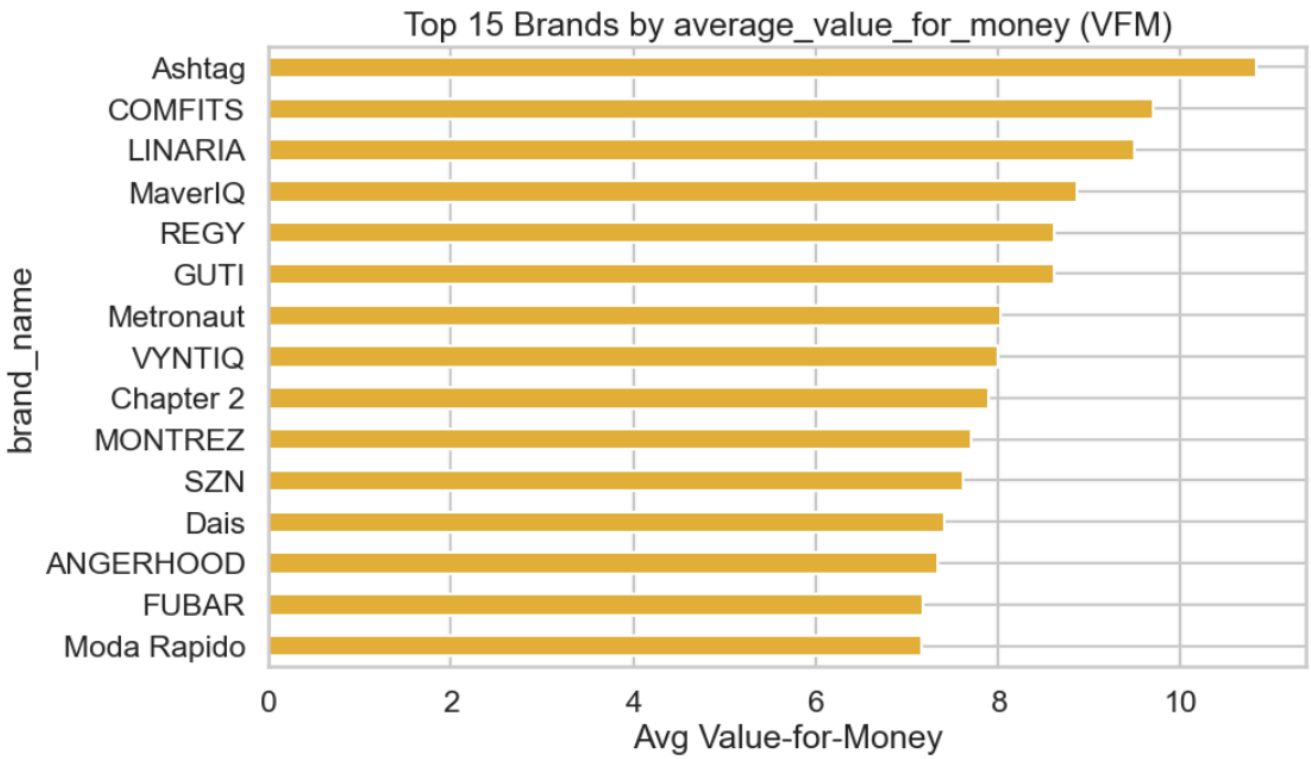
```
sns.set(style="whitegrid", context='talk')
plt.rcParams['figure.figsize'] = (10,6)
```

```
pd.set_option('display.max_columns', 120)
pd.set_option('display.width', 180)
```

Conducting EDA to Answer Key Business Questions

01_Identification of the Brands that provide the Highest overall Value-for-money

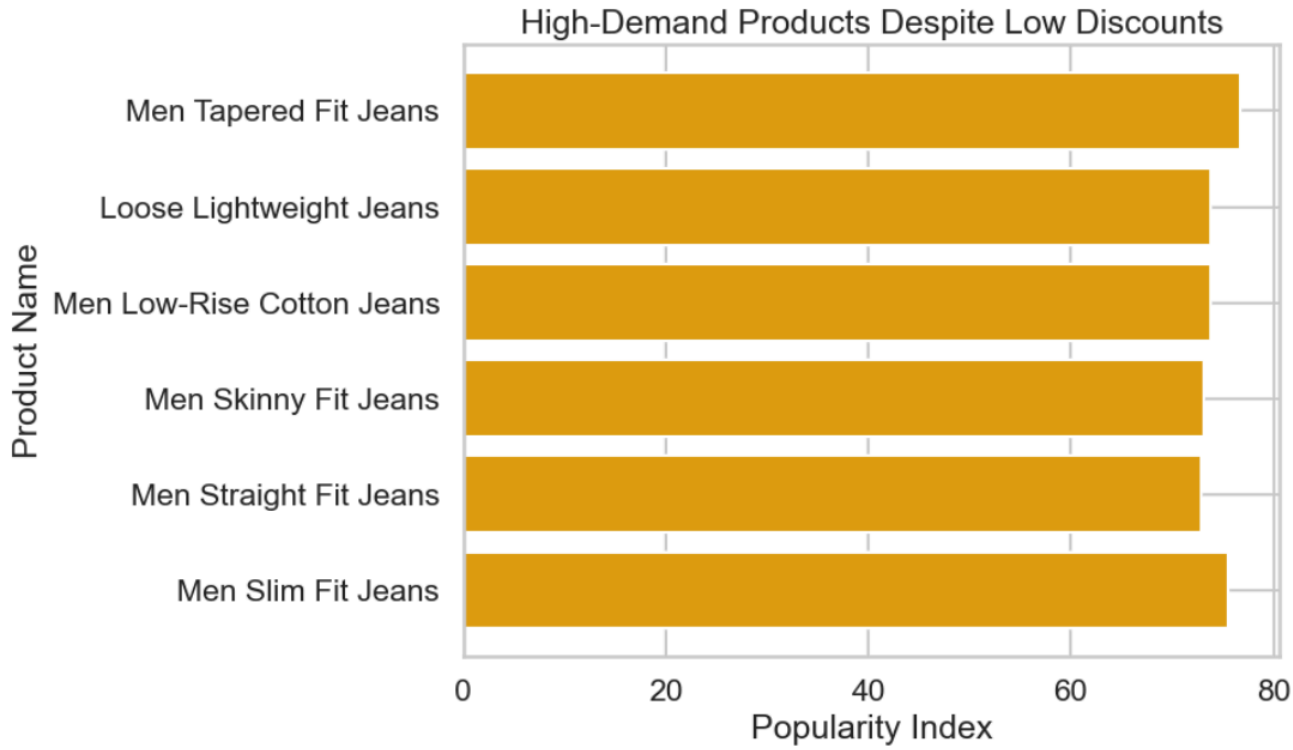
brand_name	mean	median	count
Ashtag	10.840108	10.840108	1
COMFITS	9.703202	10.000000	489
LINARIA	9.499433	9.829060	811
MaverIQ	8.867684	8.750000	146
REGY	8.617234	8.617234	1
GUTI	8.617234	8.617234	1
Metronaut	8.029163	8.400541	112
VYNTIQ	7.996946	8.016032	21
Chapter 2	7.899807	7.899807	1
MONTREZ	7.711443	7.835821	3
SZN	7.619412	8.030819	186
Dais	7.414830	7.915832	10
ANGERHOOD	7.343406	7.414830	64
FUBAR	7.181373	7.515658	14
Moda Rapido	7.156099	7.692308	171



Although **Ashtag** holds the highest mean **Value-for-Money score (10.84)**, it serves as an outlier based on a single review. The most reliable top performers are **LINARIA** and **COMFITS**, which combine strong scores with substantial review counts **(811 and 489, respectively)**. Unlike low-volume brands like REGY, these two offer statistically robust proof of consistent high value for customers.

02_Products exhibiting High Demand despite offering Low Discounts

	products	brand_name	price	effective_discount_percent	popularity_index
41183	Men Tapered Fit Jeans	HIGHLANDER	1063.0	44.023170	76.782261
12208	Men Slim Fit Jeans	WROGN	1455.0	44.016930	75.573265
45997	Men Slim Fit Jeans	WROGN	1679.0	44.014672	75.381697
46046	Men Slim Fit Jeans	WROGN	1679.0	44.014672	75.381697
95	Men Slim Fit Jeans	Flying Machine	1829.0	39.013004	75.208856
28387	Men Slim Fit Jeans	Flying Machine	1549.0	38.015206	73.964493
11077	Loose Lightweight Jeans	H&M	1511.0	44.016302	73.798124
11076	Men Slim Fit Jeans	WROGN	1511.0	44.016302	73.798124

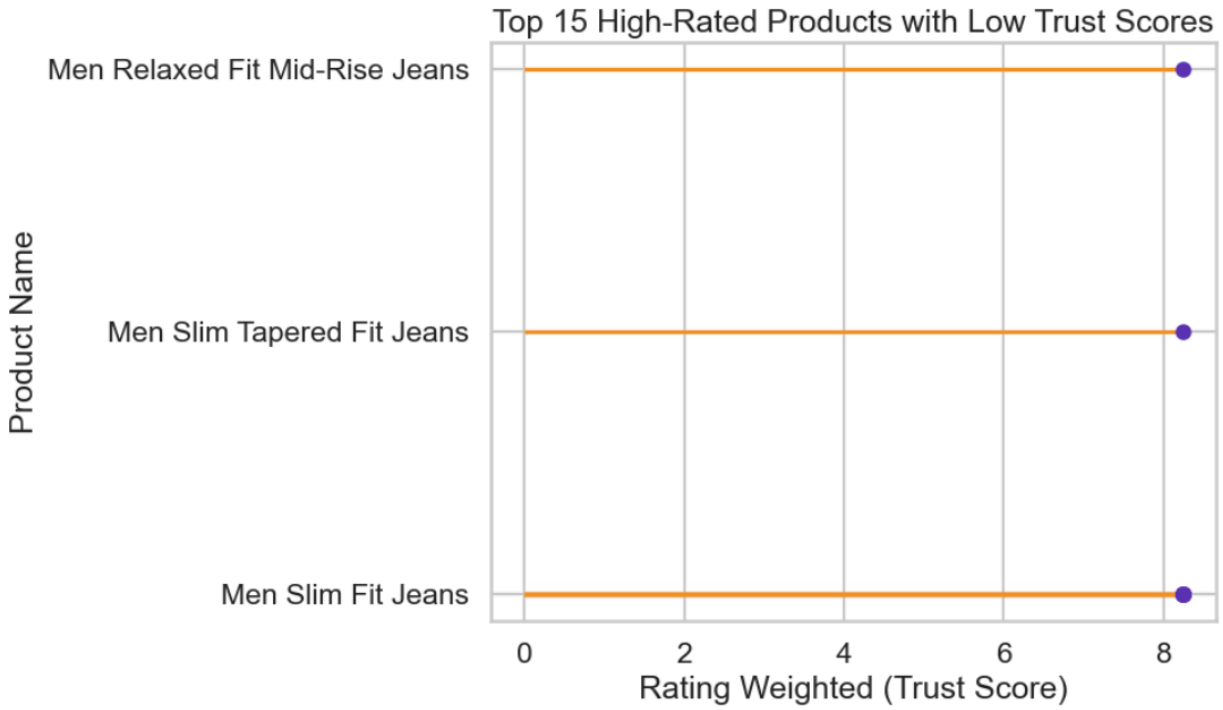


These insights show products that remain highly popular despite offering low effective discounts. **WROGN** and **Flying Machine** appear most often, showing **strong customer demand even at higher prices**.

The **top item is Highlander's Men Tapered Fit Jeans**, **with a popularity index of 76.78 and a price of 1063.** This pattern suggests solid brand loyalty and pricing power, useful for inventory and sales planning.

03_ Products having High Ratings but Low Trust

	products	brand_name	ratings	number_of_ratings	rating_weighted
1	Men Slim Fit Jeans	Flying Machine	4.6	5.0	8.242094
201	Men Slim Fit Jeans	Flying Machine	4.6	5.0	8.242094
251	Men Slim Fit Jeans	Flying Machine	4.6	5.0	8.242094
292	Men Slim Tapered Fit Jeans	WROGN	4.6	5.0	8.242094
301	Men Slim Fit Jeans	Flying Machine	4.6	5.0	8.242094

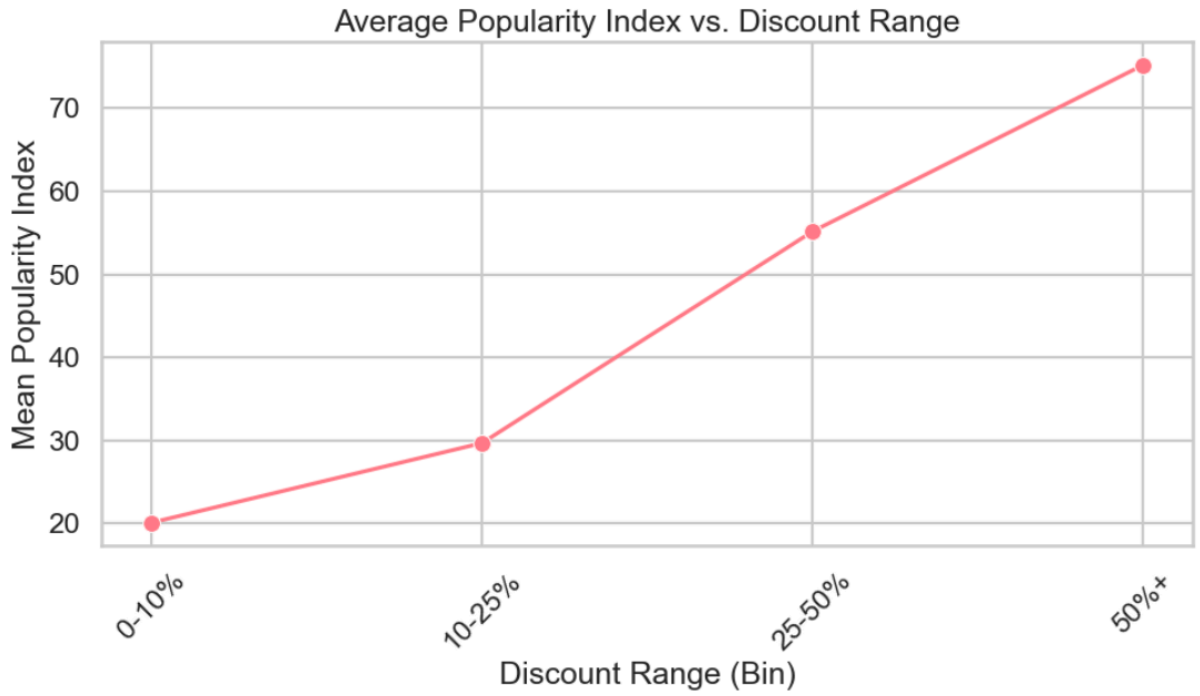


This insight highlights products with a **high customer rating of 4.6 but statistically a very low weighted rating of 8.24, signalling low trust**. The list is dominated by Flying Machine, especially Men Slim Fit Jeans.

The gap comes from a small number of perfect 5.0 reviews, making these ratings unstable. These items appear to be new or low-volume products that haven't gathered enough reviews to build reliable trust.

04_Optimal Discount Range that maximizes Product Popularity across brands

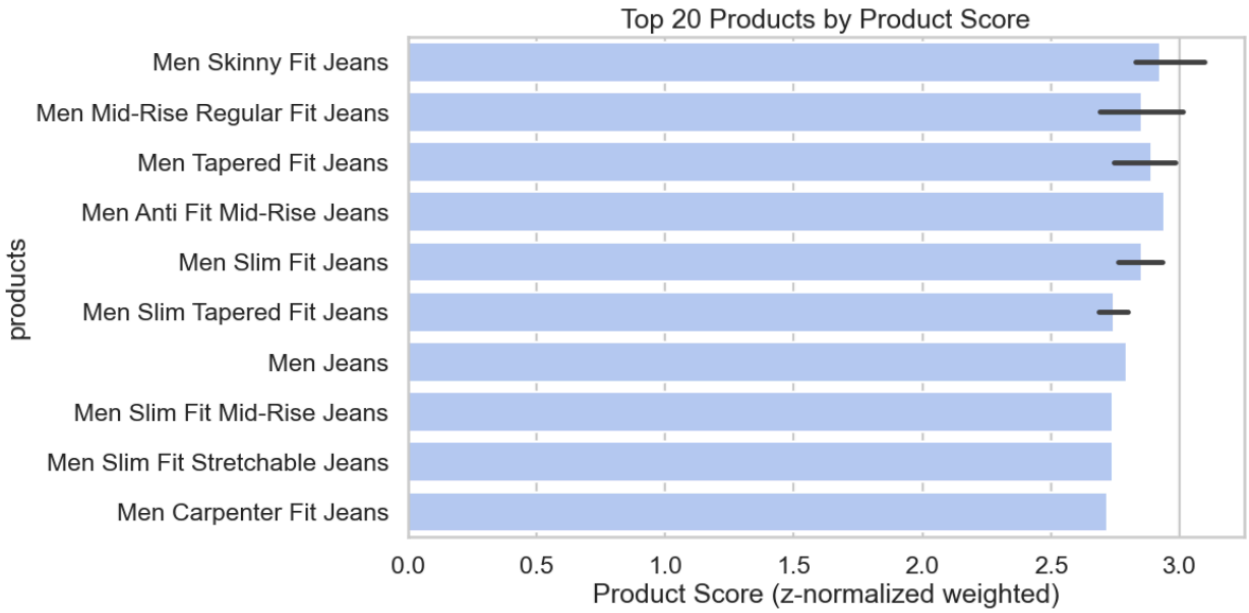
	discount_bin	count	mean	median
4	50%+	28831	75.227129	75.266677
3	25-50%	21087	55.176650	56.293920
2	10-25%	1460	29.683746	29.795676
1	0-10%	742	20.093077	20.040273
0	0%	0	NaN	NaN



The analysis shows that **higher discounts strongly boost product popularity**. The most effective range is 50 percent or more, which reaches a mean popularity score of 75.23. Popularity rises steadily with discount levels, from 20.09 at 0–10 percent to 55.18 at 25–50 percent. This points to deep discounts as the best driver of demand for men's bottomwear and there's no items priced at a true zero percent discount.

05_products that are best suited for promotion during sales campaigns based on their product score drivers

	products	brand_name	price	ratings	number_of_ratings	product_score
28991	Men Skinny Fit Jeans	Roadster	479.0	3.9	22600.0	3.094104
39367	Men Mid-Rise Regular Fit Jeans	United Colors of Benetton	449.0	3.8	4000.0	3.012864
39366	Men Tapered Fit Jeans	LOCOMOTIVE	449.0	3.8	4000.0	3.012864
44496	Men Anti Fit Mid-Rise Jeans	United Colors of Benetton	449.0	3.6	6000.0	2.938917
44514	Men Slim Fit Jeans	LOCOMOTIVE	449.0	3.6	6000.0	2.938917

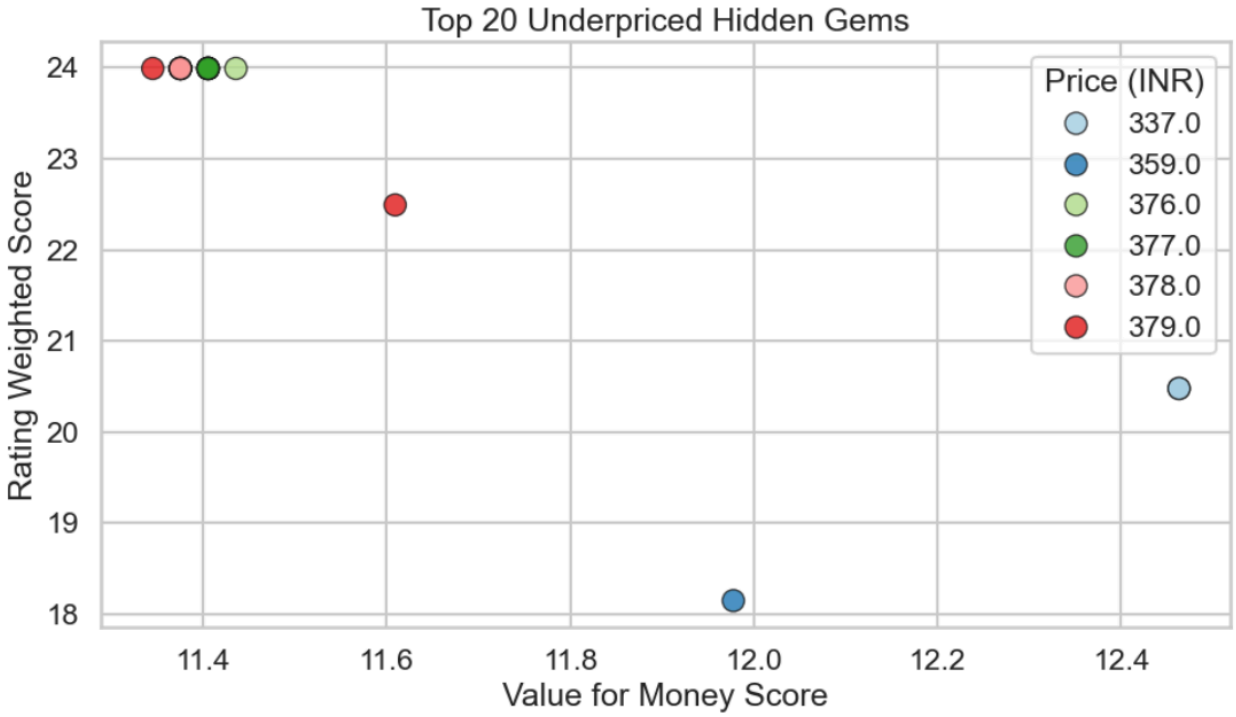


The highlights the top 20 products ranked by product score, which places more weight on popularity and value than on discounts. **These items are strong candidates for sales promotions**.

The **top performers are Roadster's Men Skinny Fit Jeans** (3.09) and **United Colors of Benetton's Men Mid-Rise Regular Fit Jeans** (3.01). Their high scores come from strong demand and solid value, making them reliable choices to feature in campaigns.

06_Identifying Underpriced Hidden Gems

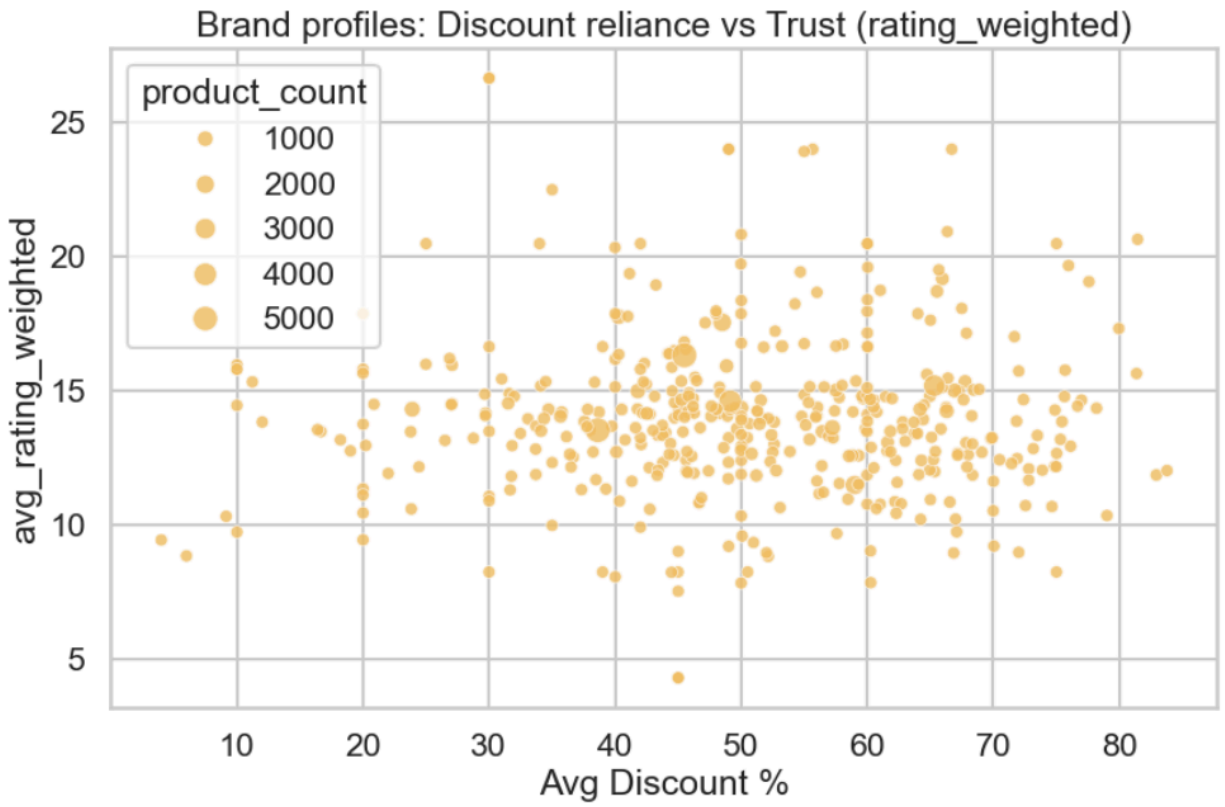
	products	brand_name	price	ratings	number_of_ratings	value_for_money_score	rating_weighted
24183	Classic Slash Knee Jeans	COMFITS	337.0	4.2	130.0	12.462908	20.475829
24184	Men Slim Fit Mid-Rise Jeans	United Colors of Benetton	337.0	4.2	130.0	12.462908	20.475829
46495	Men Relaxed Fit Jeans	Roadster	359.0	4.3	67.0	11.977716	18.143883
14541	Mid-Rise Slim Fit Jeans	LINARIA	379.0	4.4	165.0	11.609499	22.492746
22976	Men Tapered Fit Mid-Rise Jeans	SZN	376.0	4.3	264.0	11.436170	23.992838
14670	Men Stretchable Jeans	LINARIA	377.0	4.3	264.0	11.405836	23.992838
21327	Men Slim Fit Jeans	LINARIA	377.0	4.3	264.0	11.405836	23.992838
21727	Men Mid-Rise Slim Fit Jeans	LINARIA	377.0	4.3	264.0	11.405836	23.992838



This highlights the **top 20 underpriced hidden gems**—products that score high on value and trust while being priced below the median. **LINARIA and COMFITS appear most often, showing strong value at lower price points**. The top items include COMFITS' Classic Slash Knee Jeans and UCB's Men Slim Fit Mid-Rise Jeans, both priced at 337. These products offer excellent affordability and performance, making them ideal for budget-focused promotions.

07_Brands that rely heavily on discounts vs. those which rely on product quality and popularity

	brand_name	avg_discount	avg_rating_weighted	avg_popularity	product_count
0	7 For All Mankind	45.387487	13.462992	58.850479	13
1	7OUNCE	55.685228	23.992838	79.678067	3
2	AD By Arvind	38.729692	14.198874	52.928566	123
3	ADBUCKS	56.022409	14.055753	70.078162	8
4	ADIDAS Originals	35.006418	9.977016	44.983434	2



This analysis shows a sample of the Brand Profile table, which helps classify brands by their sales approach. **Some brands rely heavily on discounts, like prochain and ADBUCKS, which offer high average discounts but only moderate quality or trust scores**. Others lean on quality and brand strength, such as ADIDAS Originals and 7 For All Mankind, which keep discounts lower while maintaining steady popularity and stronger overall performance.

Business Insights and Recommendations

- **Value-for-Money Leaders**
    - Brands like LINARIA and COMFITS consistently deliver strong value backed by substantial customer feedback. They should be prioritized for visibility, collaborations, and inclusion in value-focused campaigns

- **High-Demand, Low-Discount Products**
    - Brands such as WROGN and Flying Machine continue to sell well without deep discounts. These items show strong pricing power and can be positioned as margin-positive products during regular sales cycles

- **High Rating but Low Trust Items**
    - Some products show high ratings but have too few reviews to be considered reliable. These items should be monitored rather than promoted, and encouraged to gather more verified feedback before making stocking or marketing decisions

- **Discount Range that Maximizes Popularity**
    - Customer engagement increases sharply with higher discount brackets. Deep discounts perform best during major campaigns. This range can be used strategically to clear inventory and boost traffic

- **Best Products for Sales Campaigns**
    - Products that combine popularity, quality, and value stand out as high-impact choices for promotional placement. These are ideal for campaign banners, recommendations, and converting high-intent traffic

- **Underpriced Hidden Gems**
    - Several products, especially from LINARIA and COMFITS, offer strong quality at lower price points. Featuring these in curated collections helps attract budget-conscious customers and improves overall basket value

- **Brand Strategy Segmentation**
    - Some brands rely heavily on discounts to compete, while others perform well with modest markdowns. Understanding which brands fall into price-driven vs value-driven categories helps shape discount strategy, partnership plans, and campaign design