

Candy brand analysis

AI for Business Decisions Final project

Da Fang Lin & Hsueh Ning Chao



Agenda

Dataset Overview

Business Problem Definition

Data Preprocessing

Model Development and Prediction

Focal Brand and Competitor Comparison

Business Recommendations

Dataset overview

- **Resource:** The Complete Journey from dunnhumby.
- **Dataset Content:** Covers two years of shopping transactions for 2,500 frequent shopper households, with a total of 1,040,000 transaction records.
- **Product and Category:** Includes 92,300 products spanning 300 categories.
- **Dataset Structure:** There are 8 tables, including household-level transactions, consumer behavior(coupon), product details, marketing campaigns, and demography.

Category Selection

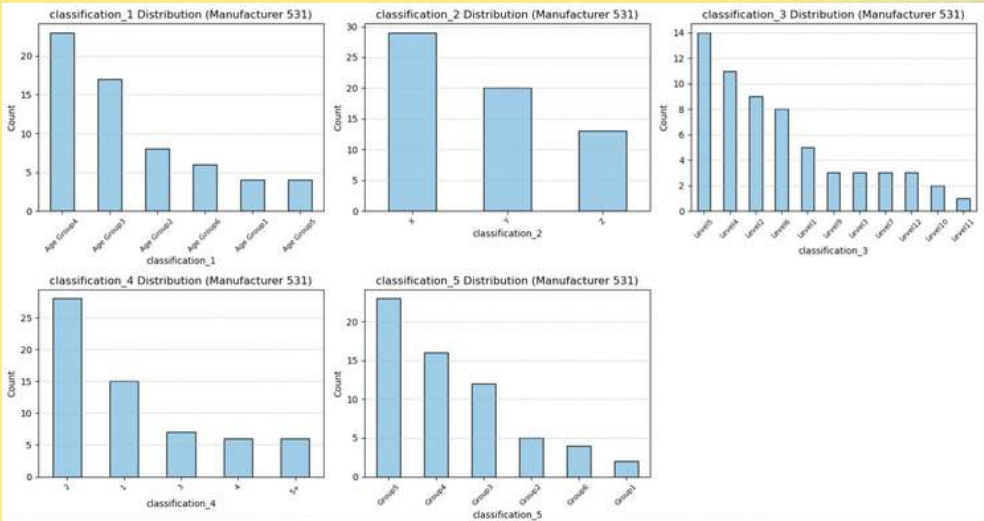
Across a total of 300 product categories, which category should we focus on?

CANDY - PACKAGED MANUFACTURER 693	686 data points
CANDY - PACKAGED MANUFACTURER 857	371 data points
CANDY - PACKAGED MANUFACTURER 531	169 data points
CANDY - PACKAGED MANUFACTURER 2172	152 data points
CANDY - PACKAGED MANUFACTURER 348	85 data points
⋮	⋮
SOFT DRINKS 2241	1 data points
SOFT DRINKS 2641	1 data points

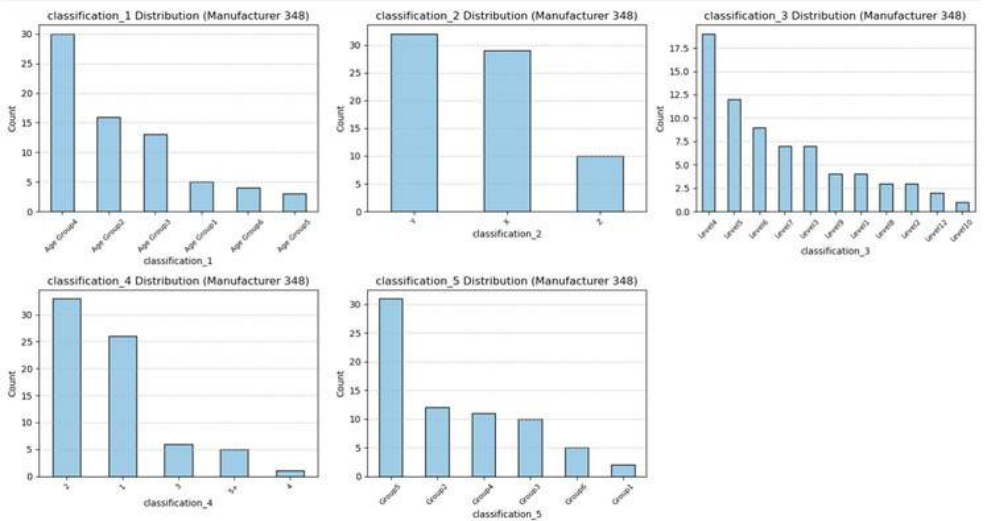
Focal Brand

In the Candy Packaged category, there are 115 brands. Which brand should we define as our focal brand?

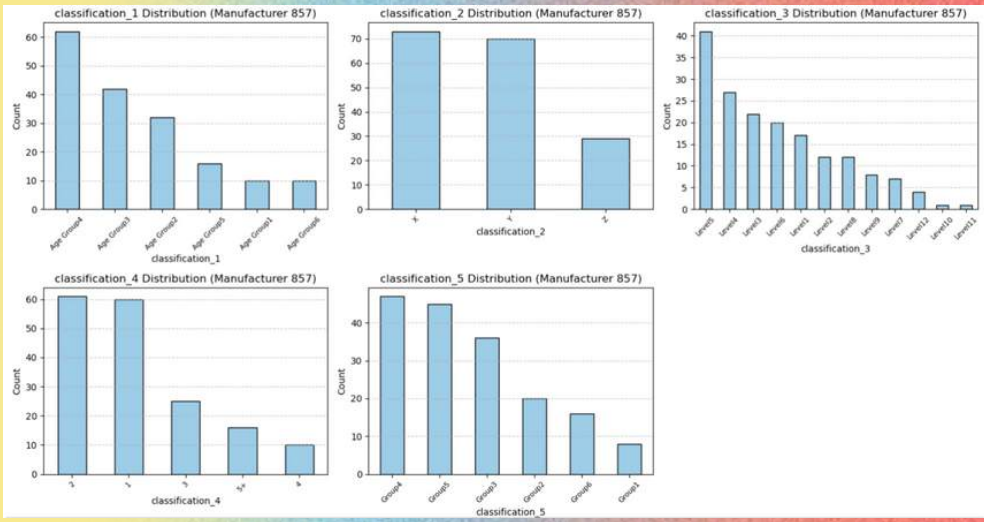
531



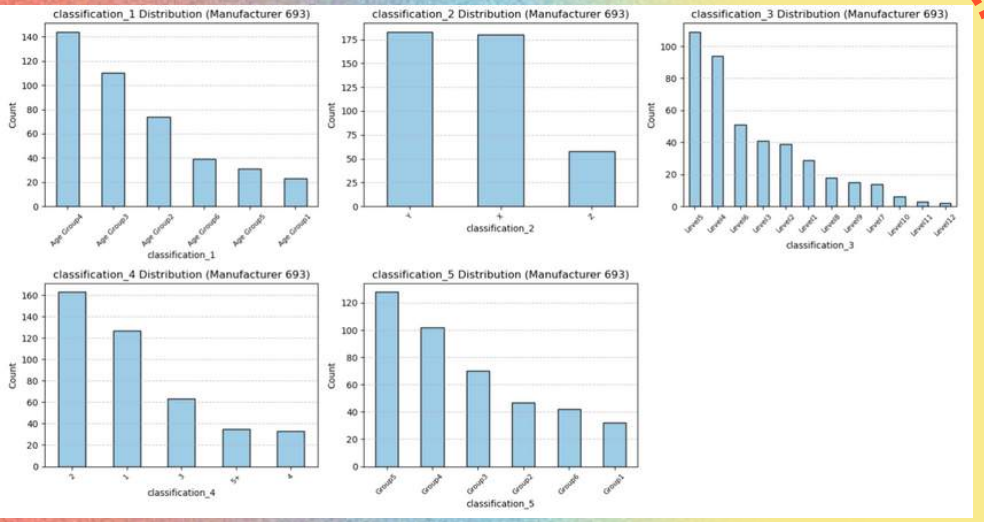
348



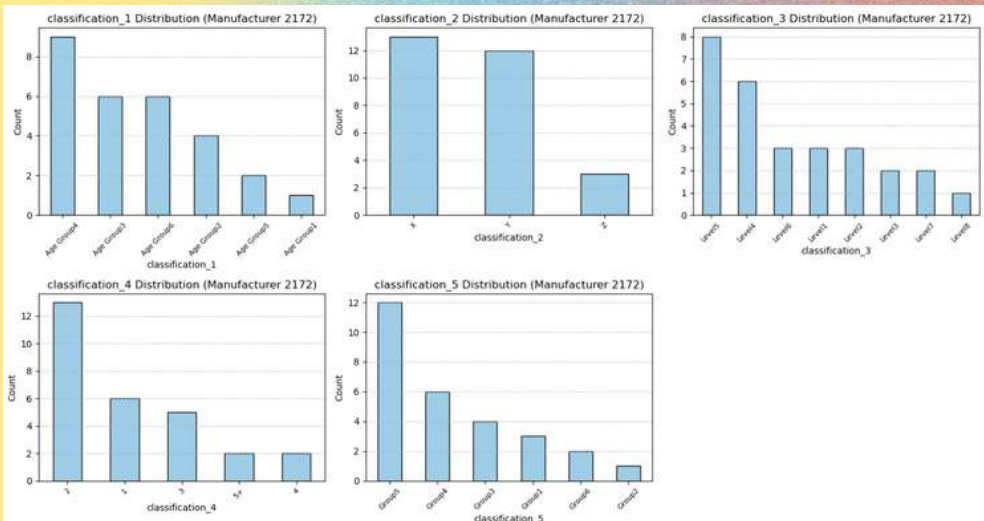
857



693

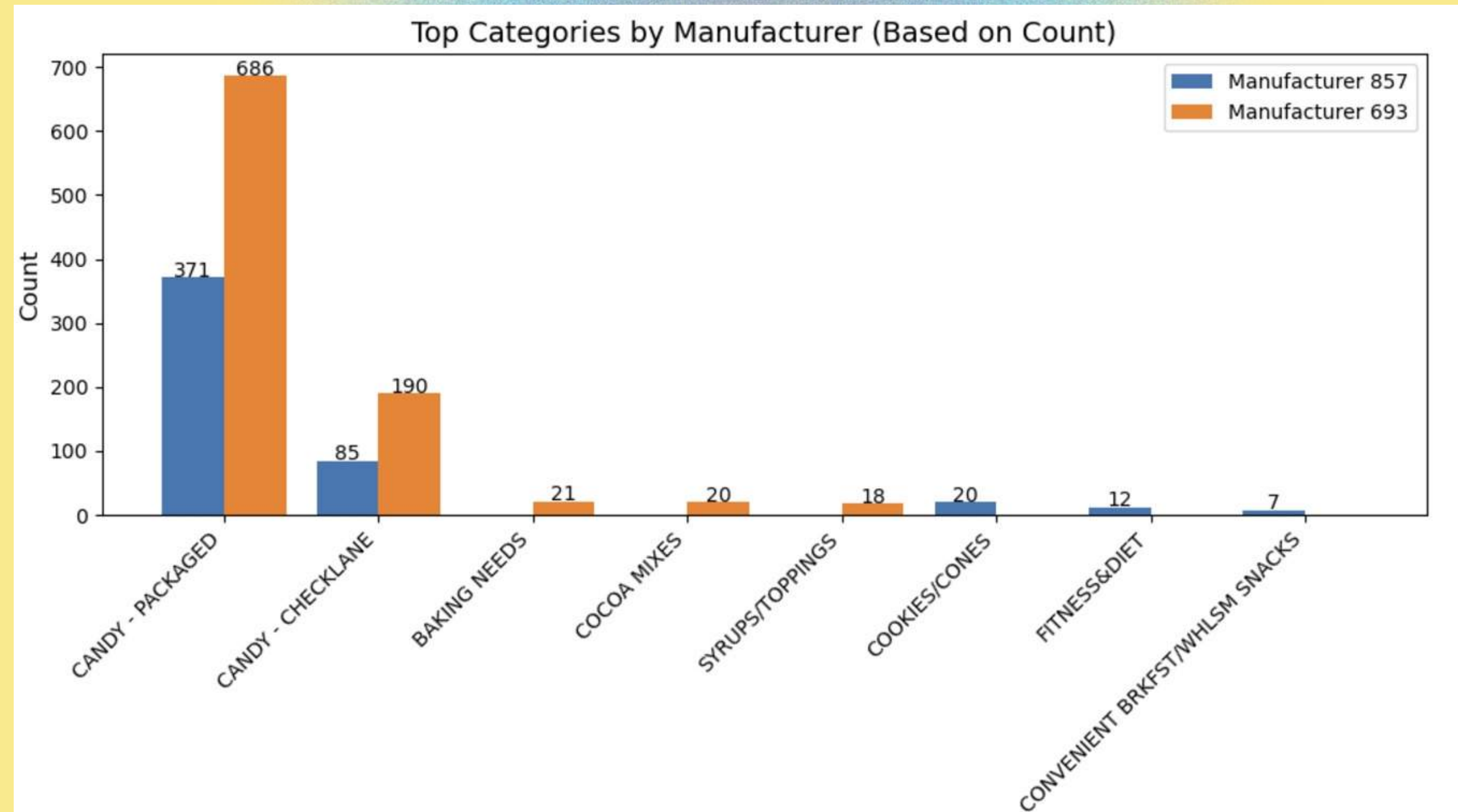


2172



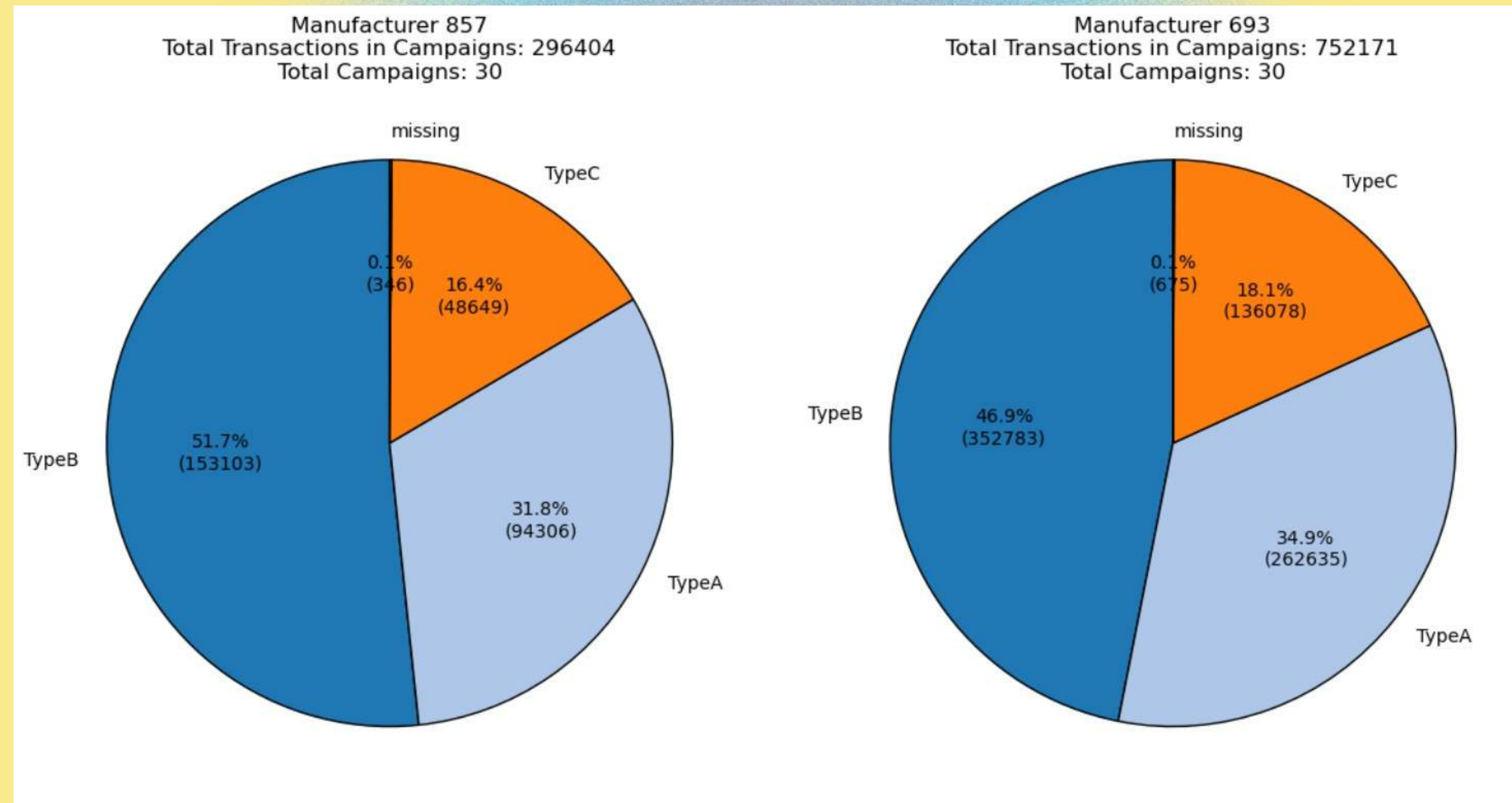
Current Status

What are the main products of the two brands?



Current Status

How many campaigns did each brand execute, and what are the proportions of their campaign types?



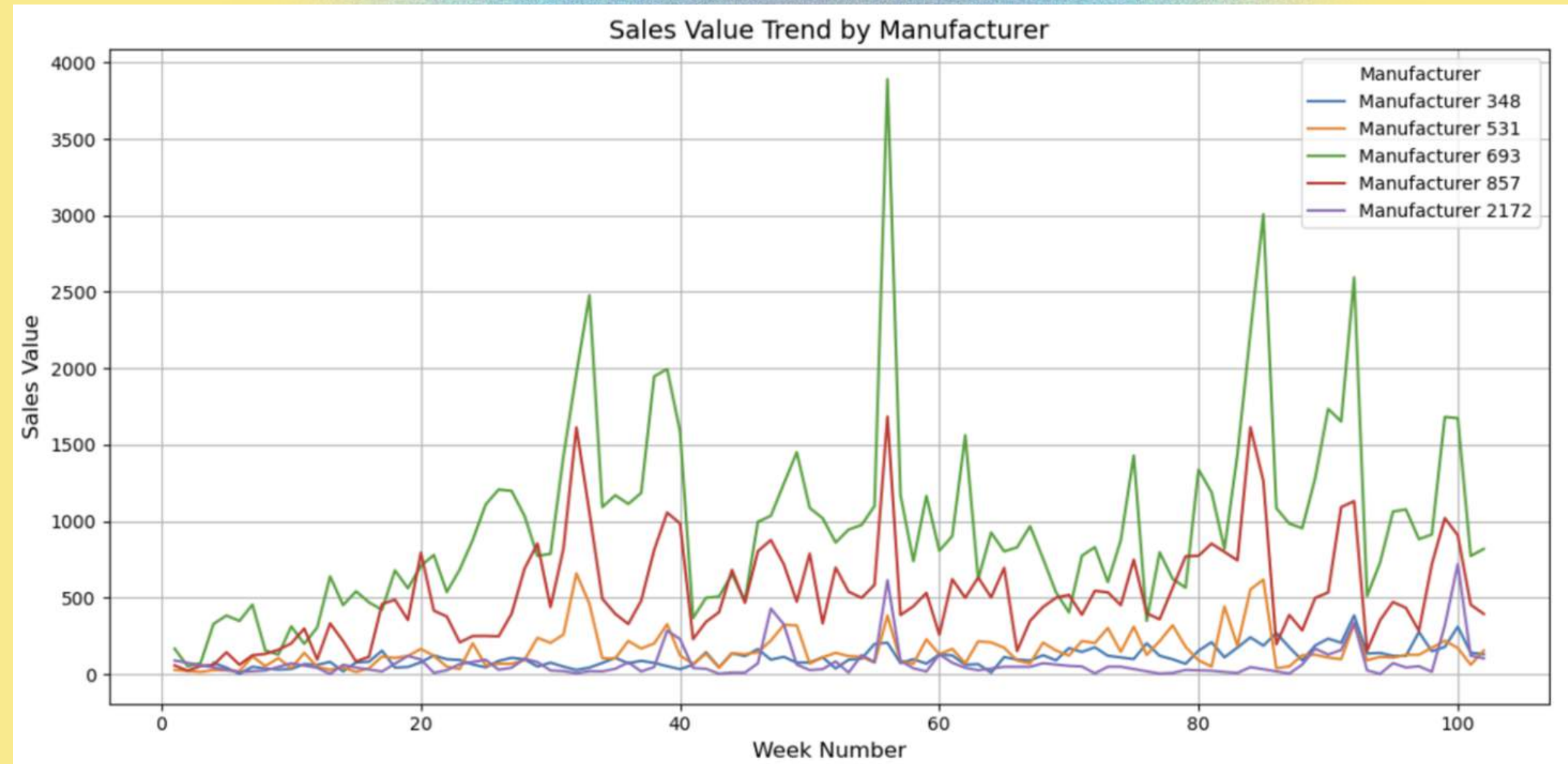
TypeA: Based on the customer's prior purchase behavior.

TypeB and TypeC: No selection criteria; all customers receive the same coupons.

Focal Brand

In the Candy Packaged category, there are 115 brands. Which brand should we define as our focal brand?

Sales Value across five brands



857 as our focal brand, 693 as our competitor.

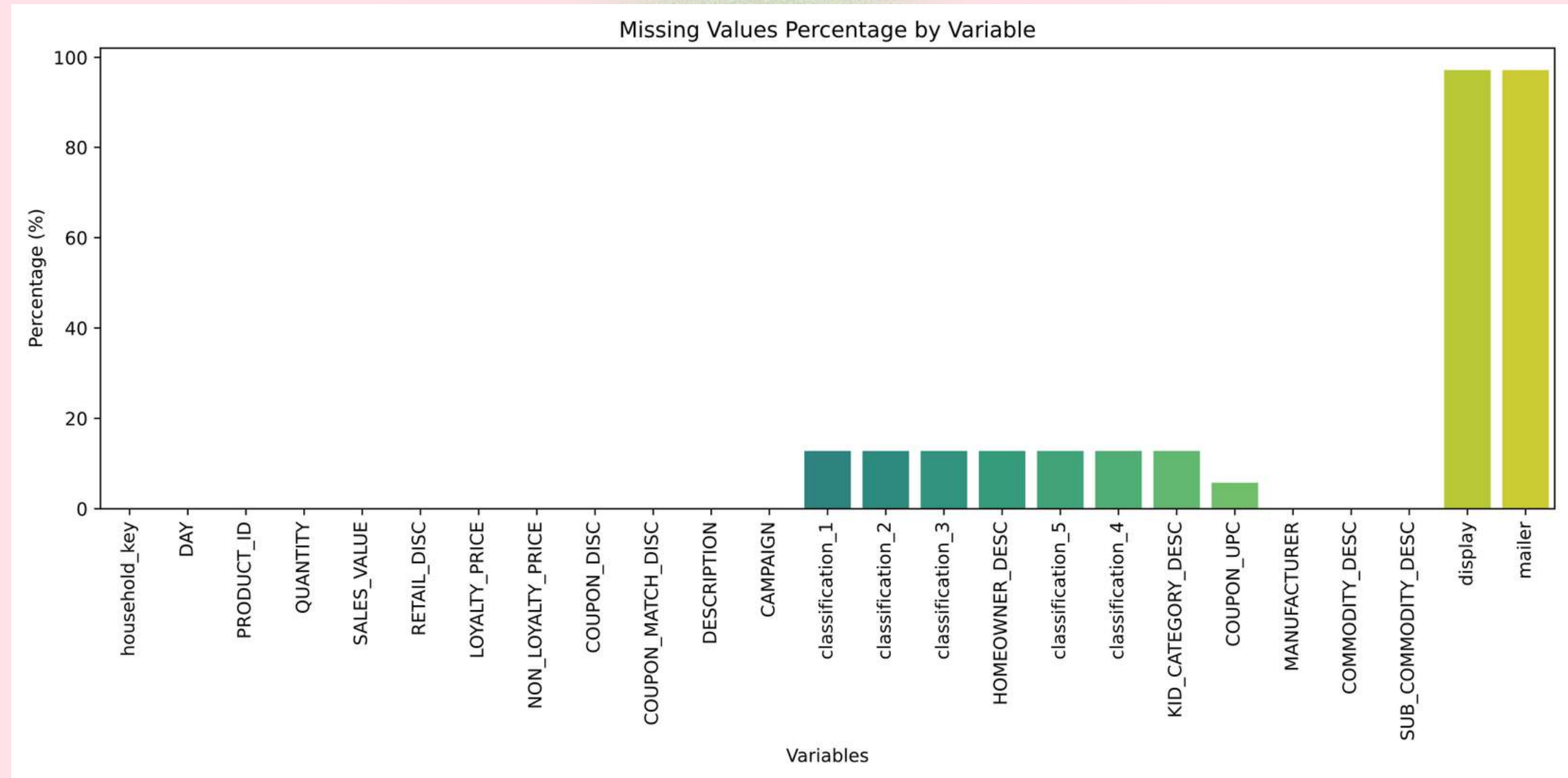
Business Problem

- The focal brand (857) and the competitor (693) have a similar consumer distribution.
- In terms of product categories, the two largest categories by sales share are also the same.
- The coupon campaign distribution strategy is also the same.

**However,
there is a significant difference in their sales performance.
How can we improve the sales of the focal brand?**

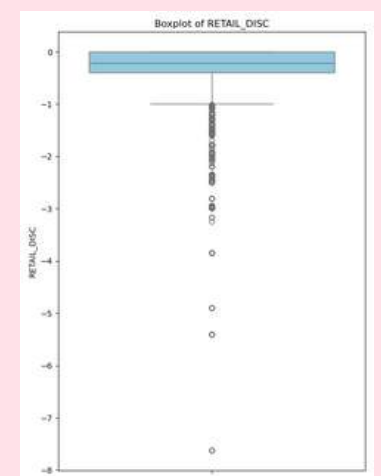
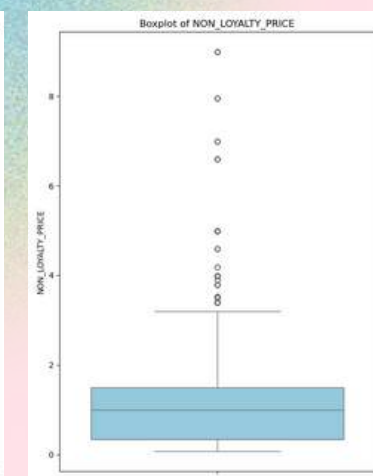
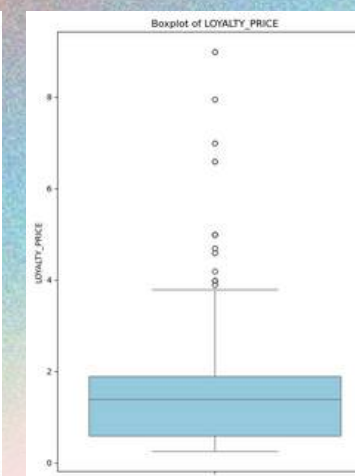
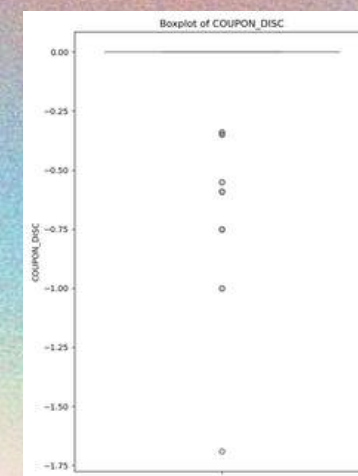
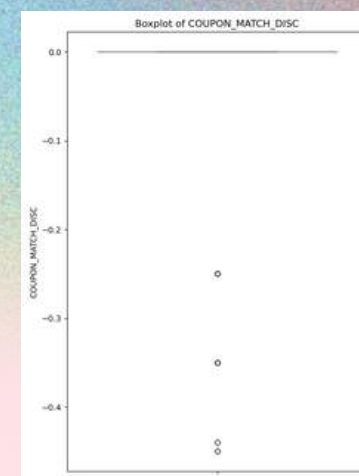
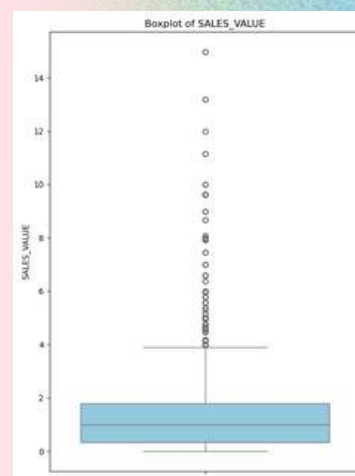
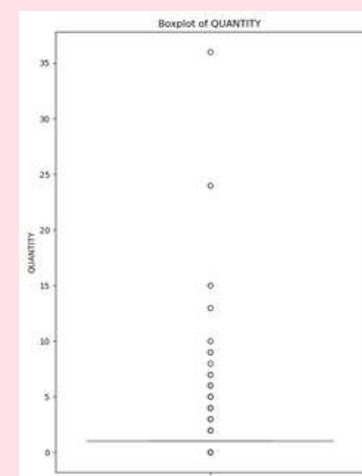
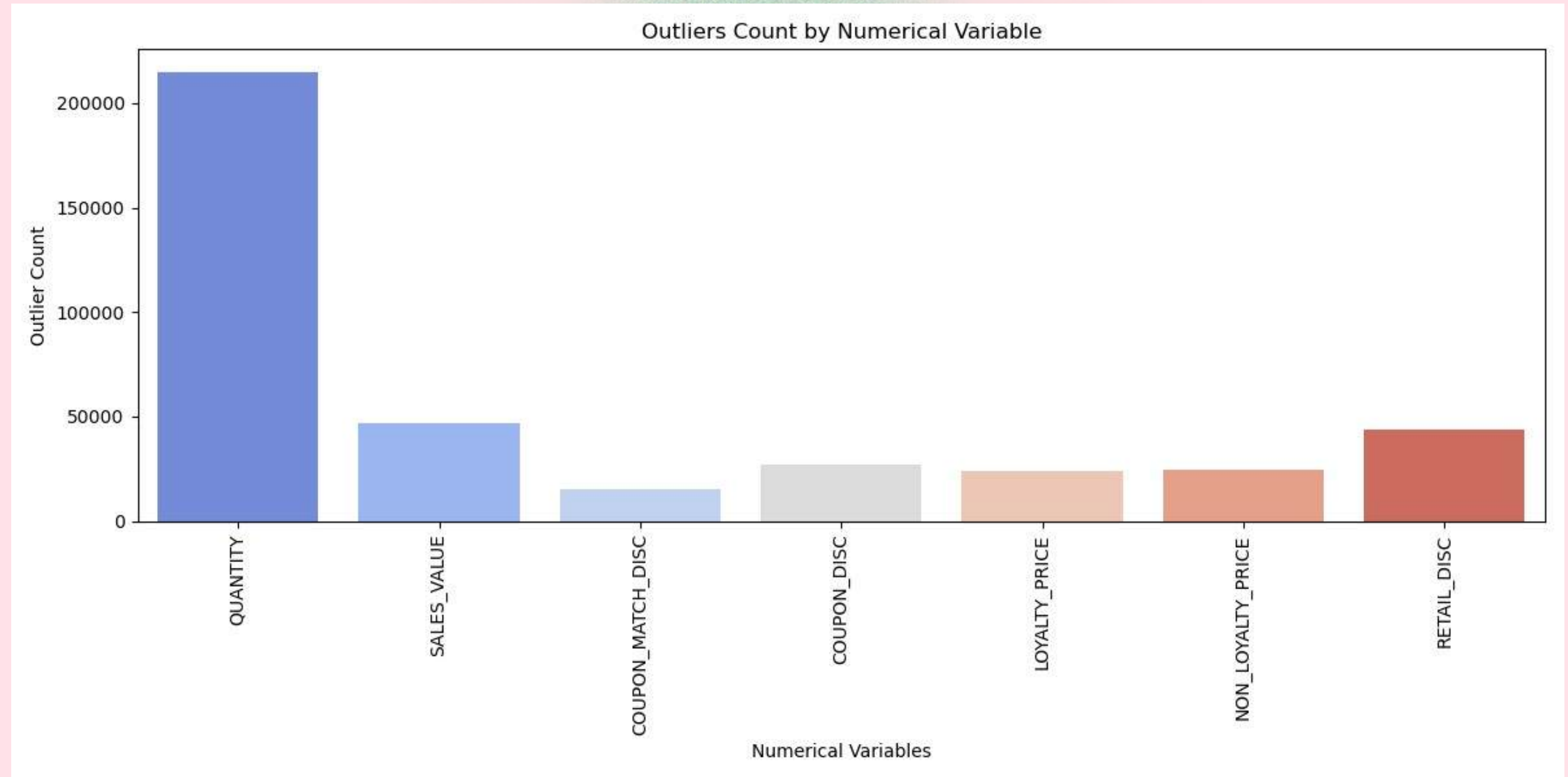
Data Preprocessing

Missing Value



- From **classification 1 to COUPON_UPC**, since the missing percentage around 12%, we decide create 'missing' as a new category for them.
- For **display & mailer**, since the missing percentage around 97%, we decide delete both variables.

Outliers

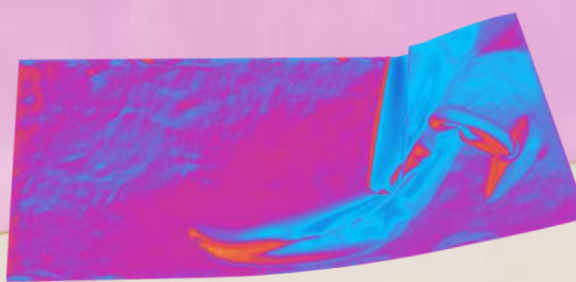


Model Development and Prediction

Model Selection

Categorical		Interval	ID
classification_1	CAMPAIGN	QUANTITY	COUPON_UPC
classification_2	DESCRIPTION	SALES_VALUE	household_key
classification_3	COMMODITY_DESC	RETAIL_DISC	PRODUCT_ID
classification_4	SUB_COMMODITY_DESC	COUPON_DISC	
classification_5	HOMEOWNER_DESC	LOYALTY_PRICE	Ordinal
MANUFACTURER	KID_CATEGORY_DESC	NON_LOYALTY_PRICE	DAY
		COUPON_MATCH_DISC	

Since we have many categorical variables, we choose to use one of the following methods: **Random Forest & Gradient Boosting (CatBoost)**.



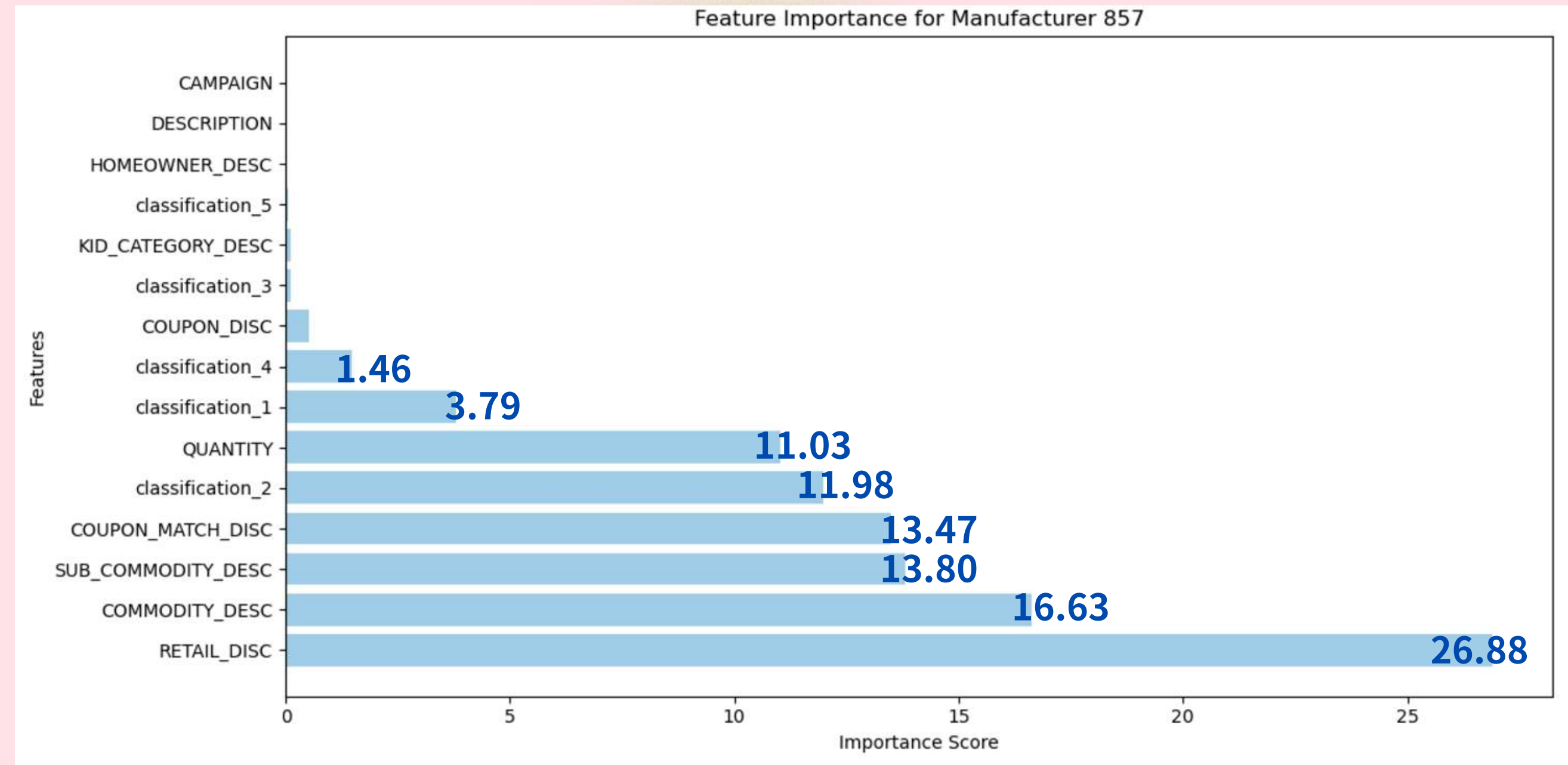
Focal brand
857



Competitor
693

Variable Importance

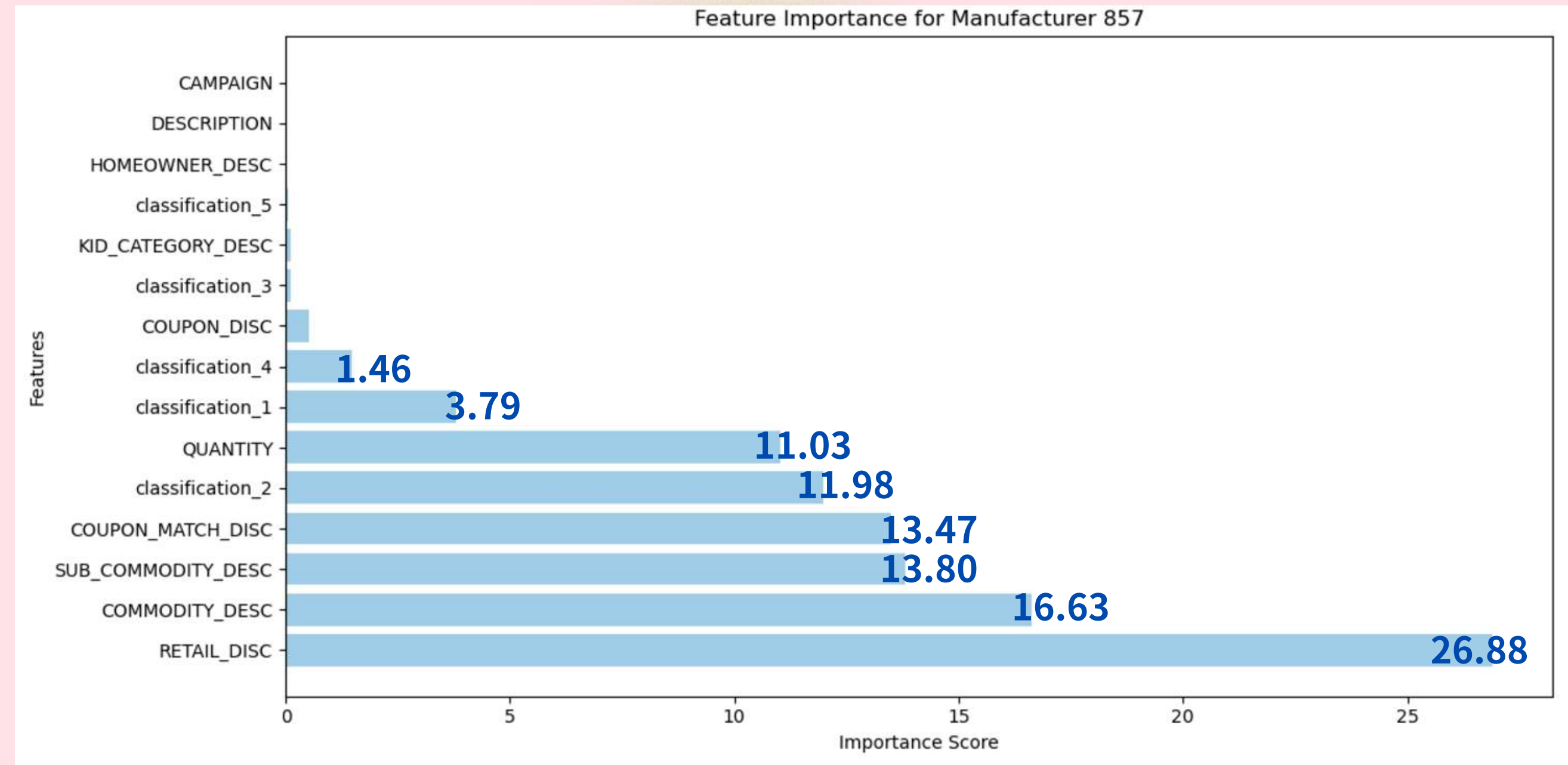
Top features for Focal brand (857)



- **RETAIL_DISC** - Retailer loyalty card discounts have a significant impact on sales for this competitor.
- **COMMODITY_DESC** - The product category plays a crucial role, similar to our brand(693), but with slightly less importance compared to retailer discounts.
- **SUB_COMMODITY_DESC & COUPON_MATCH_DISC** - Both also play notable roles in driving sales.

Variable Importance

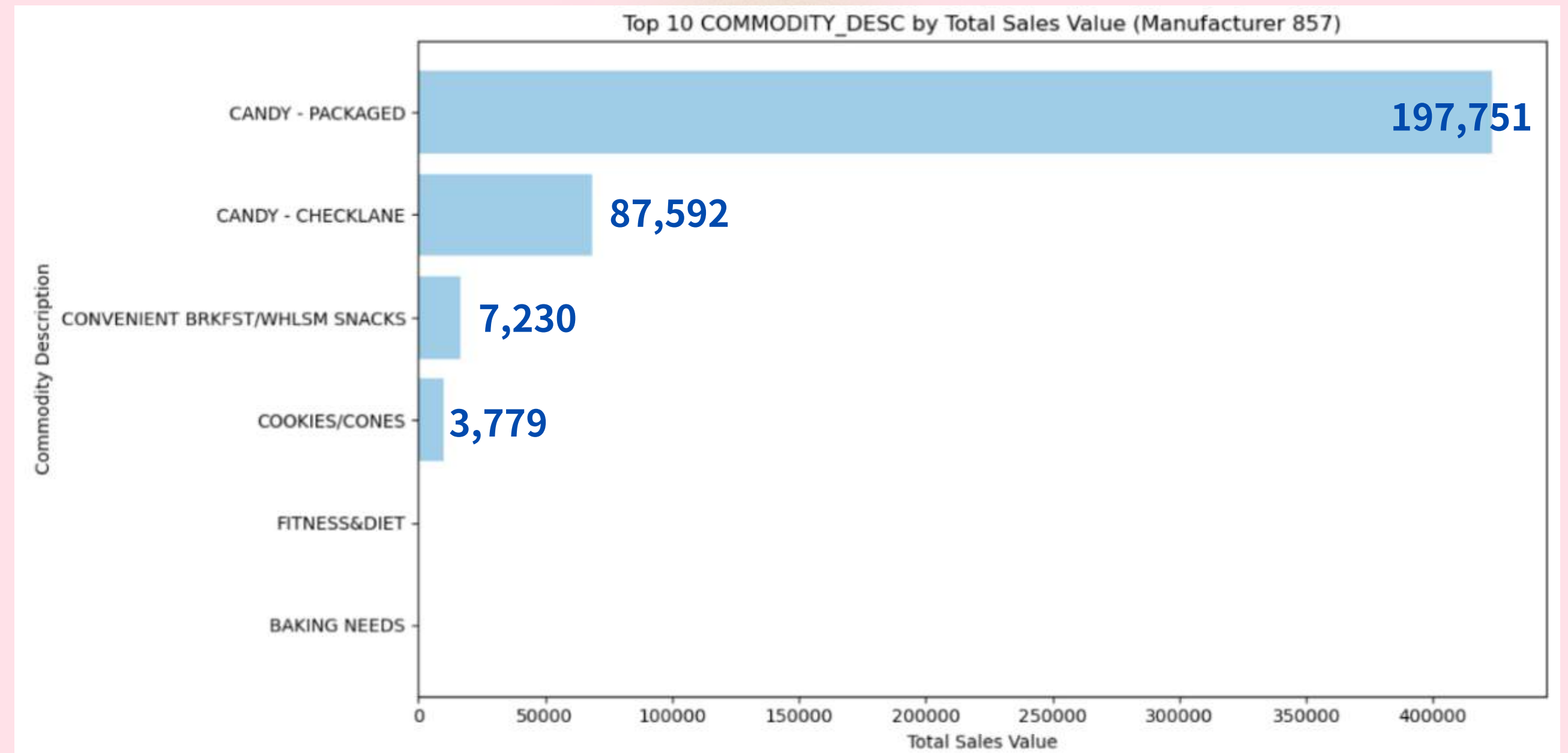
Moderate & low features for Focal brand (857)



- **QUANTITY** - The quantity purchased influences sales significantly. Multi-unit purchases or bundle sales can drive higher sales volumes.
- **classification_2 & classification_4 & classification_1** - Household demographic segments (e.g., classification_2) have a stronger impact compared to other classifications. Tailoring marketing strategies to specific household types may be more effective.

Detailed Analysis of Top Feature

Top 10 Commodity



Top-Contributing Commodity Category:

- **CANDY - PACKAGED:** significantly outperforms other categories, making it the primary driver of sales.

Second Strongest Category:

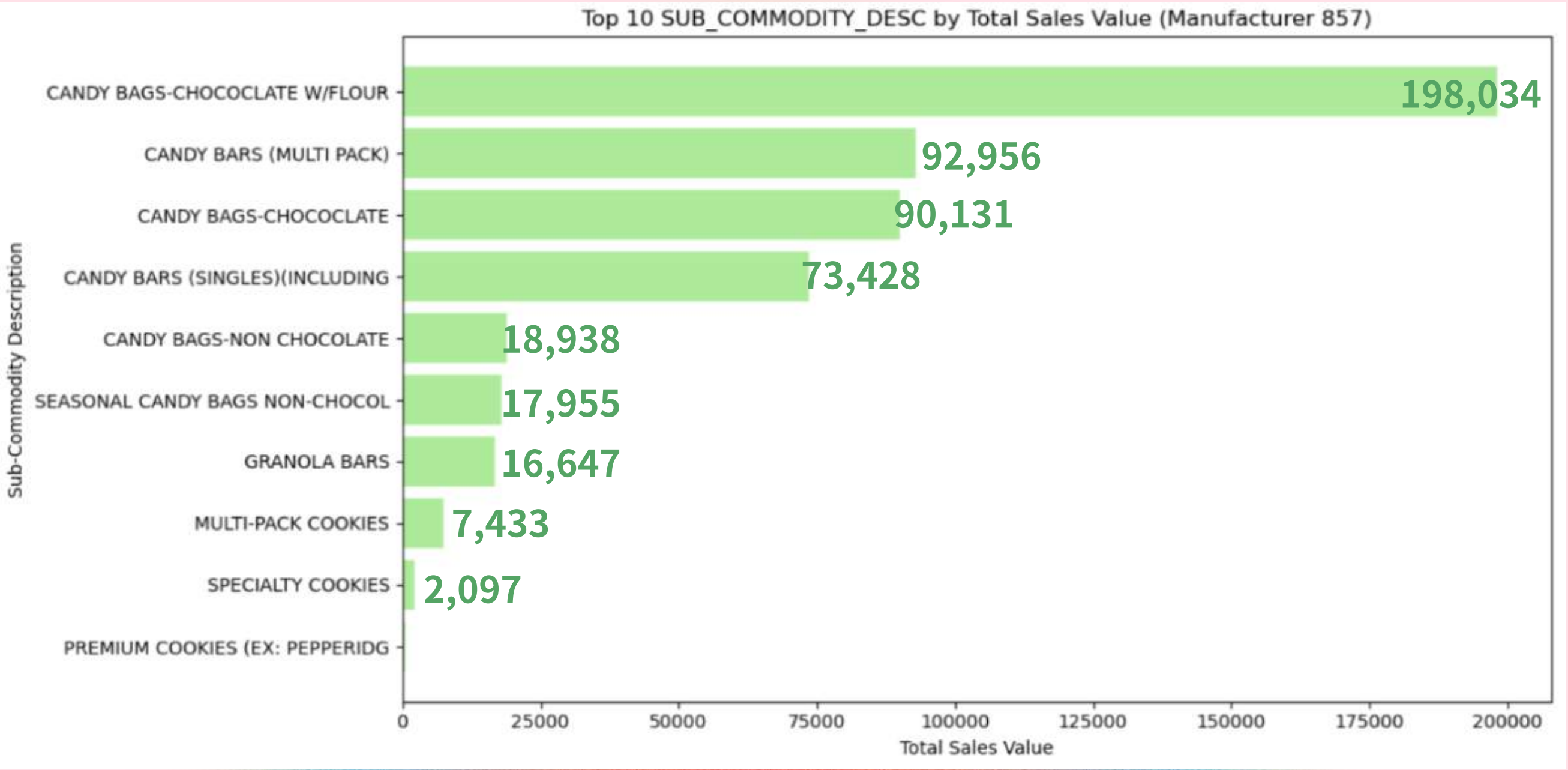
- **CANDY - CHECKLANE:** Checklane candy also shows strong sales, suggesting impulse purchases at checkout are effective.

Other Categories Have Low Sales:

- **CONVENIENT BRKFST/WHLSM SNACKS & COOKIES/CONES**

Detailed Analysis of Top Feature

Top 10 Sub-Commodity



Top Performing Category:

- **CANDY BAGS - CHOCOLATE W/FLOUR:** has the highest sales, far surpassing others.

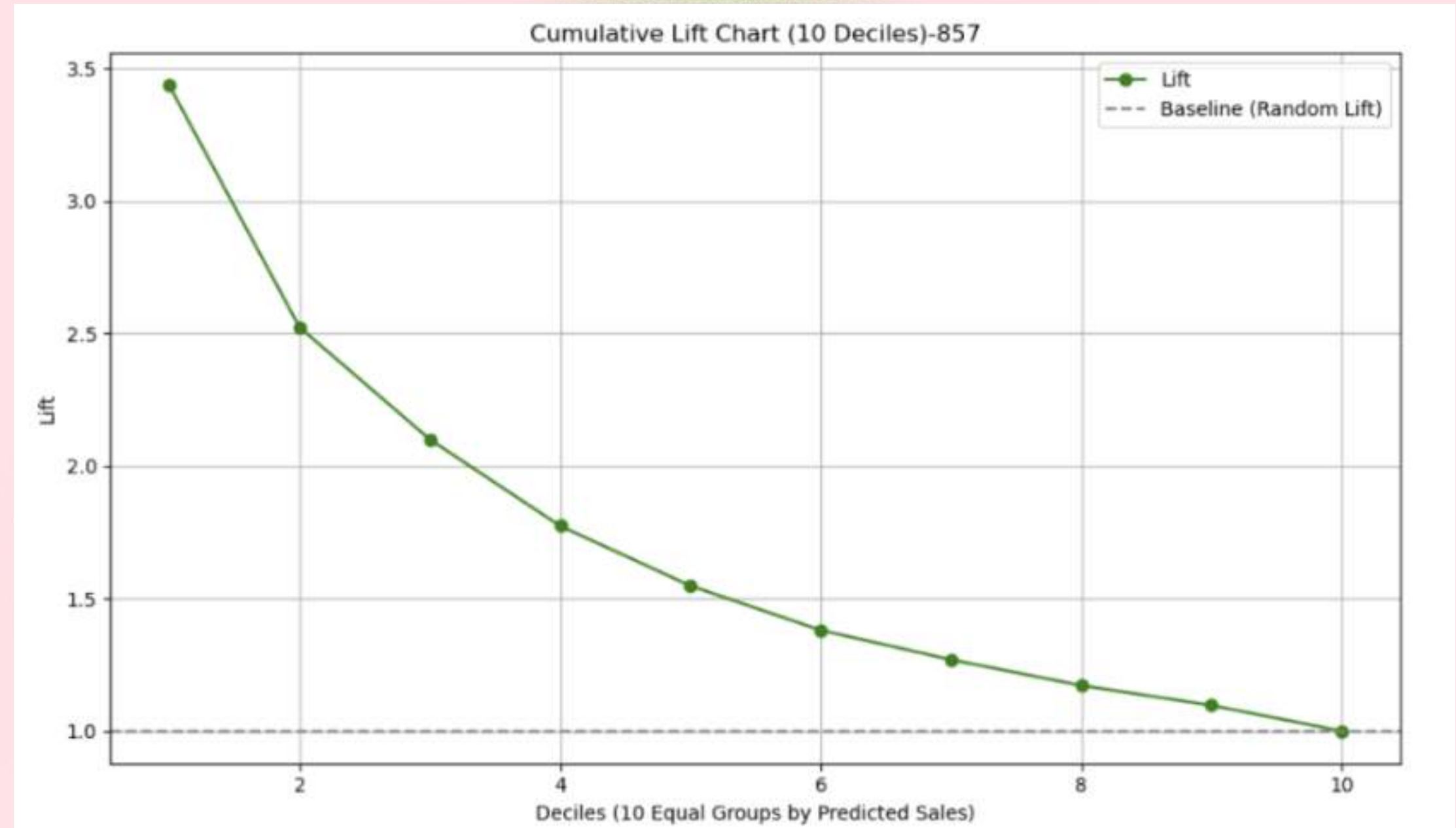
Mid-Tier Sub-Commodity Categories:

- **CANDY BARS (MULTI PACK) & CANDY BAGS - CHOCOLATE & CANDY BARS (SINGLES):** These three chocolate products exhibit steady and considerable sales, indicating continuous demand for different packaging formats of chocolate.

Non-Chocolate Candy and Other Categories Have Lower Sales

Detailed Analysis of Top Feature

Cumulative Lift Chart



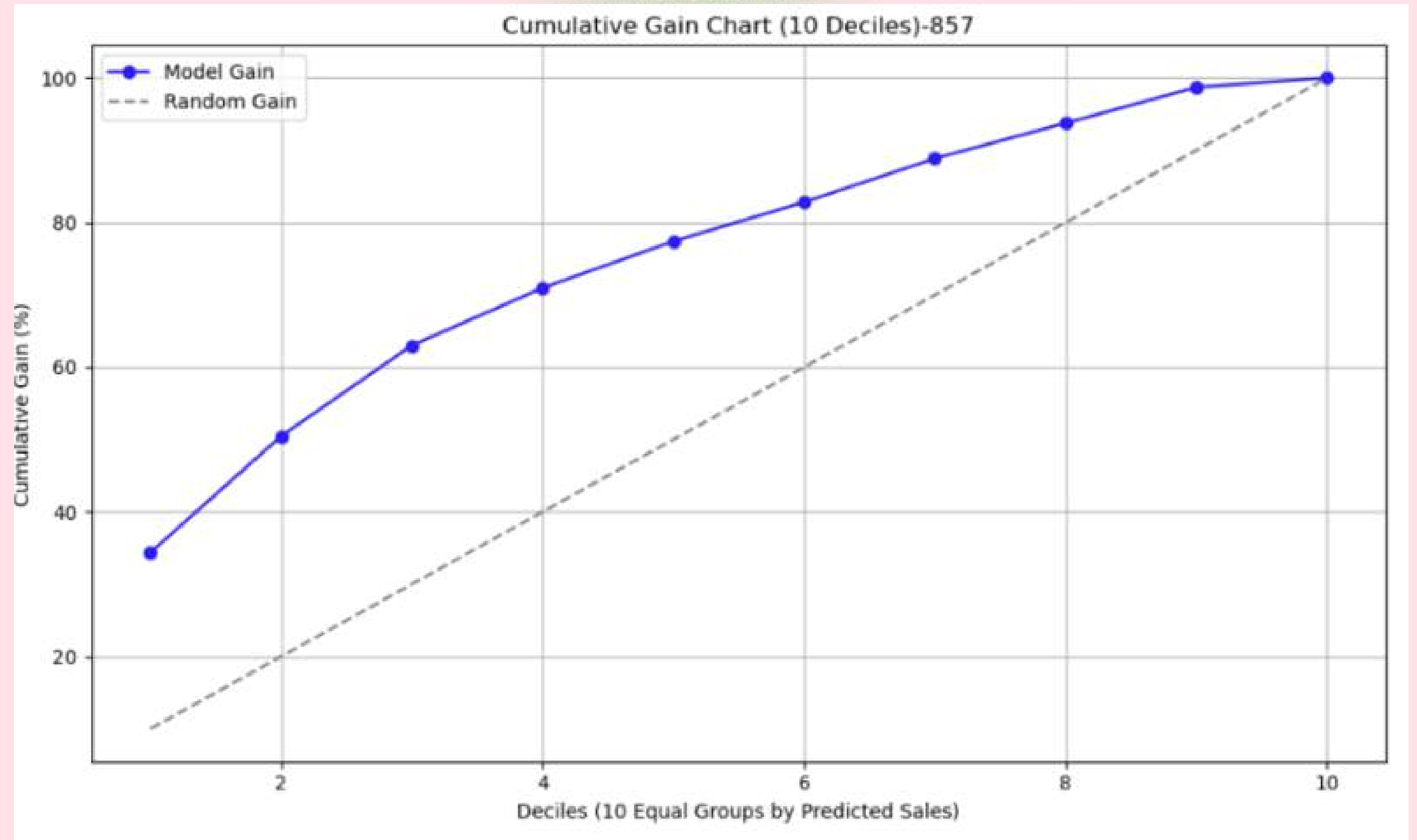
- The first 10% of the predictions contain sales values that are 3.5 times higher than what would be expected through random selection.
- The lift values are greater than 2 up until the third decile.

Stragety:

Focus on the top 30% of the deciles, as the lift values in these ranges are ≥ 2.0 , indicating that the model's predictive performance is clearly better than random guessing.

Detailed Analysis of Top Feature

Cumulative Gain Chart



- In the top 30% of the predictions, the cumulative gain reaches 70%.

Strategy:

Target top 30% groups for marketing campaigns, as they represent the highest sales potential.



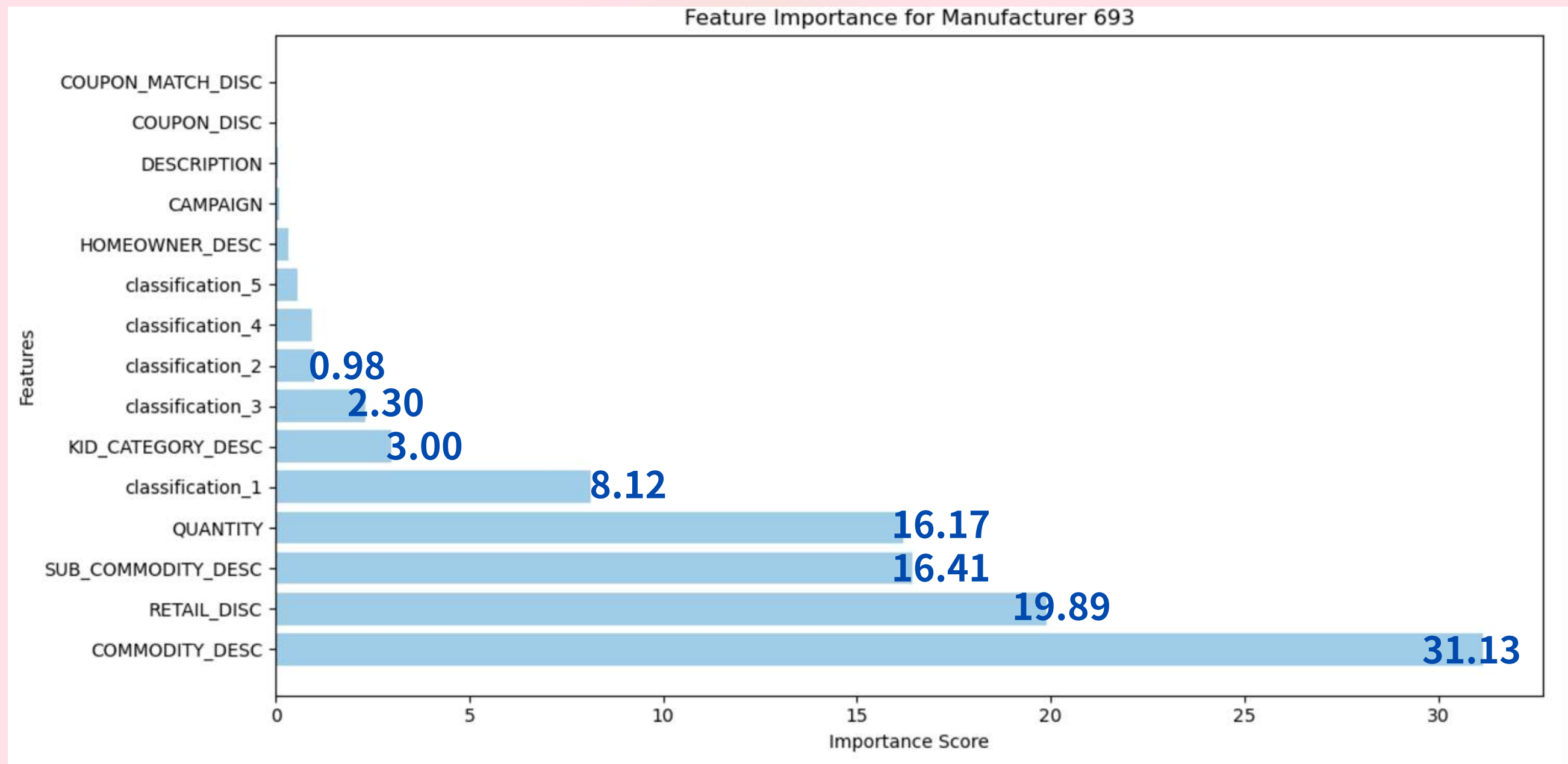
Focal brand
857



Competitor
693

Variable Importance

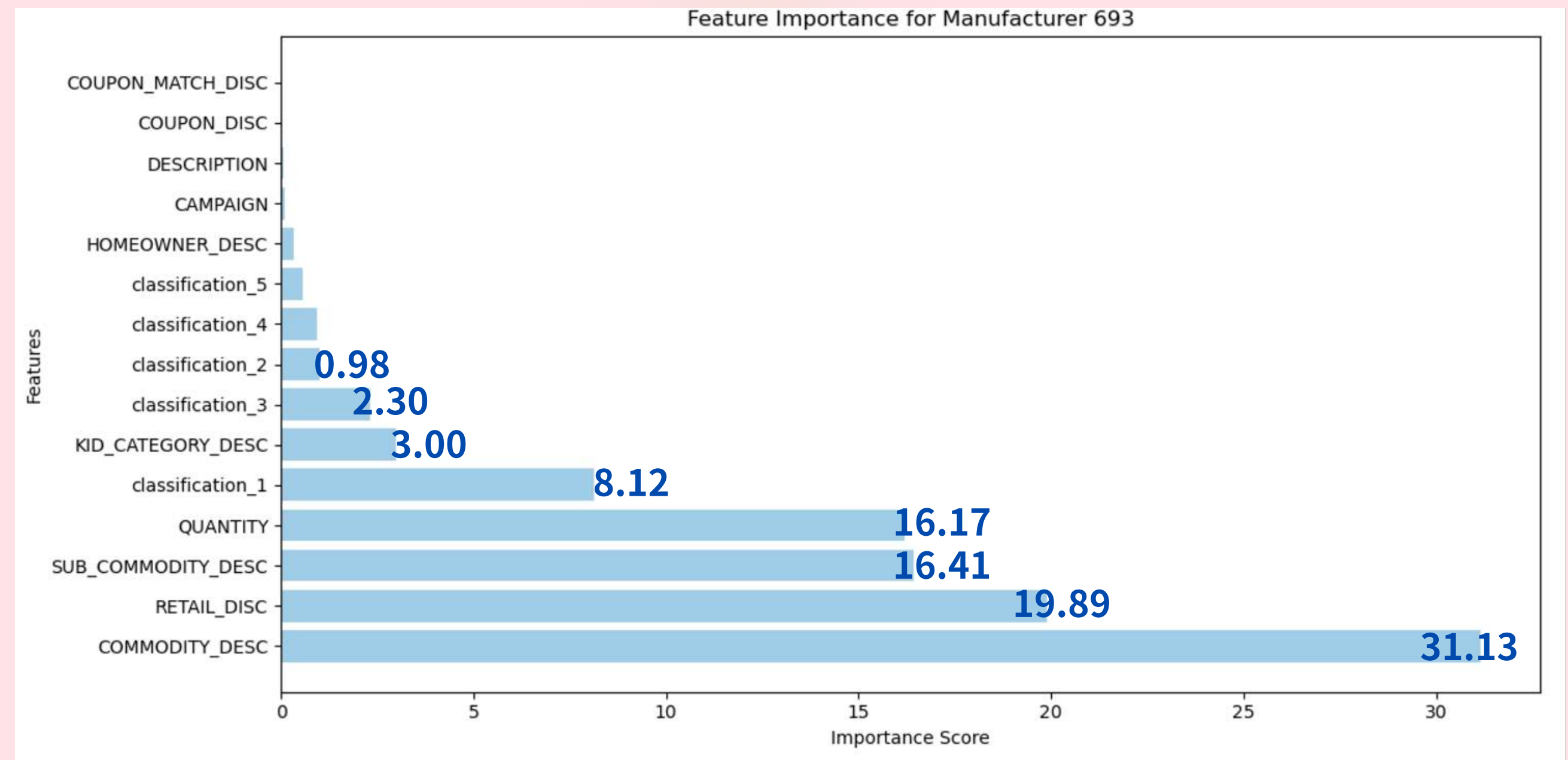
top features for competitor(693)



- **COMMODITY_DESC** - Product groupings have a strong correlation with sales performance, suggesting consistent purchasing patterns for certain categories.
- **RETAIL_DISC** - The retail discount significantly influences sales, indicating that consumers are highly responsive to promotional offers.
- **SUB_COMMODITY_DESC** - The lowest-level product category further refines which specific products drive sales, highlighting the importance of granular product differentiation.
- **QUANTITY** - As expected, more purchases naturally lead to higher sales volumes.

Variable Importance

Moderate & low features for competitor(693)



- **classification_1**(Group1 through to Group6.) - Household-level characteristics impact sales to a reasonable extent. Different household types or sizes likely exhibit varying purchasing behaviors.
- **KID_CATEGORY_DESC**, **HOMEOWNER_DESC**, **classification_3** (Level1 - 12), **classification_2**(X, Y, and Z), **classification_4** (1-5+), **classification_5** (Group 1-6)- These features have a relatively minor influence on sales, implying that factors like the number of children in a household or certain demographic classifications do not strongly impact purchasing behavior.

Comparison

Focal brand (857)

Competitor(693)

Top Features

RETAIL_DISC (26.88%)
COMMODITY_DESC (16.63%)

COMMODITY_DESC (31.13%)
RETAIL_DISC (19.89%)

857's customers are price-sensitive and responsive to loyalty programs, while 693's strategy focuses on category-driven sales.

Quantity Feature

QUANTITY (11.03%)

QUANTITY (16.17%)

- 693's customers purchase in higher quantities, whereas 857's customers tend to buy smaller quantities more frequently.
- 857's customers may prefer single or smaller purchases.

Demographic Features

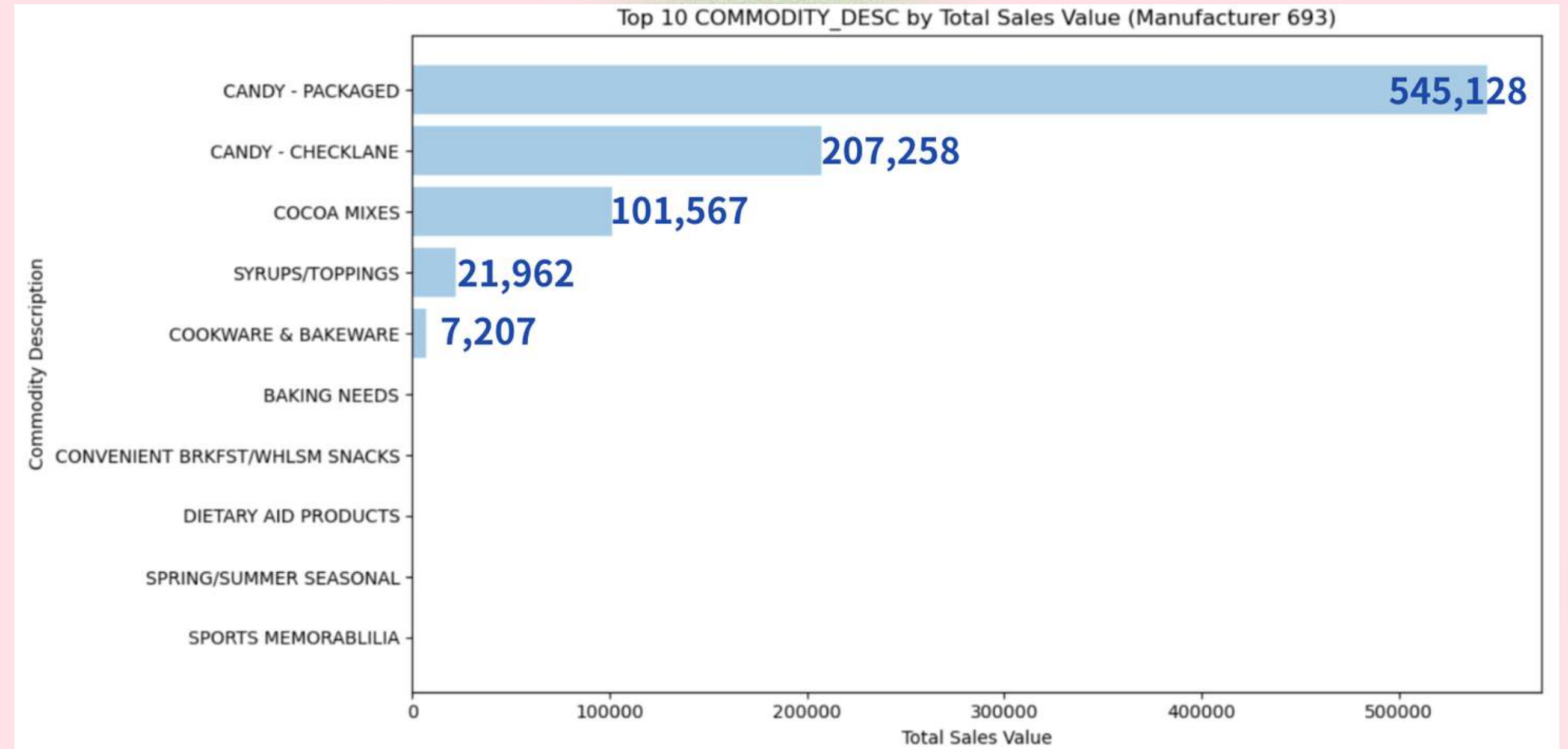
classification_1 (3.79%)

classification_1 (8.12%)

- 693 sees a stronger influence from demographic segmentation compared to 857.
- 693 may be more effective in tailoring its marketing to specific demographic groups.

Detailed Analysis of Top Feature

Top 10 Commodity



Top-Contributing Commodity Categories:

- **CANDY - PACKAGED & CANDY - CHECKLANE & COCOA MIXES**

Long-Tail Products:

- **SYRUPS/TOPPINGS & COOKWARE & BAKEWARE**

The top three categories together account for the majority of sales, indicating that sales are concentrated in a few key categories.

Detailed Analysis of Top Feature

How to Optimize Our Product Portfolio?

Focal brand (857)		Competitor(693)	
1.	CANDY - PACKAGED	1.	CANDY - PACKAGED
2.	CANDY - CHECKLANE	2.	CANDY - CHECKLANE
3.	CONVENIENT BRKFST/WHLSM SNACKS	3.	COCOA MIXES

Consistent Demand

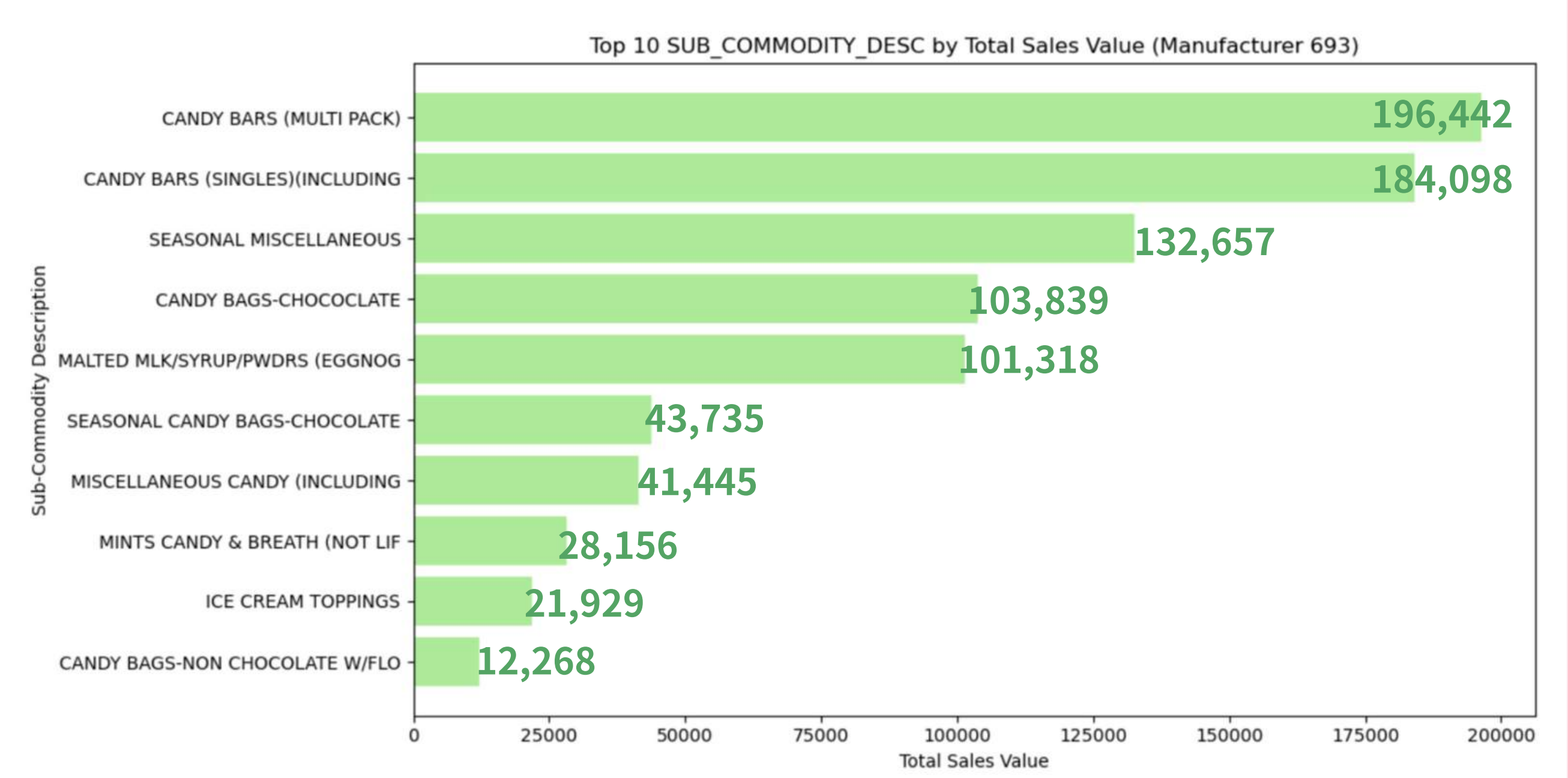
Both brands have the same top-selling categories suggests a widespread and consistent customer demand for packaged candy and checklane candy, not limited to just one brand.

Strong Competitiveness

This consistency also highlights the inherent competitiveness and profit potential of these categories, making them reliable investments for driving sales growth.

Detailed Analysis of Top Feature

Top 10 Sub-Commodity

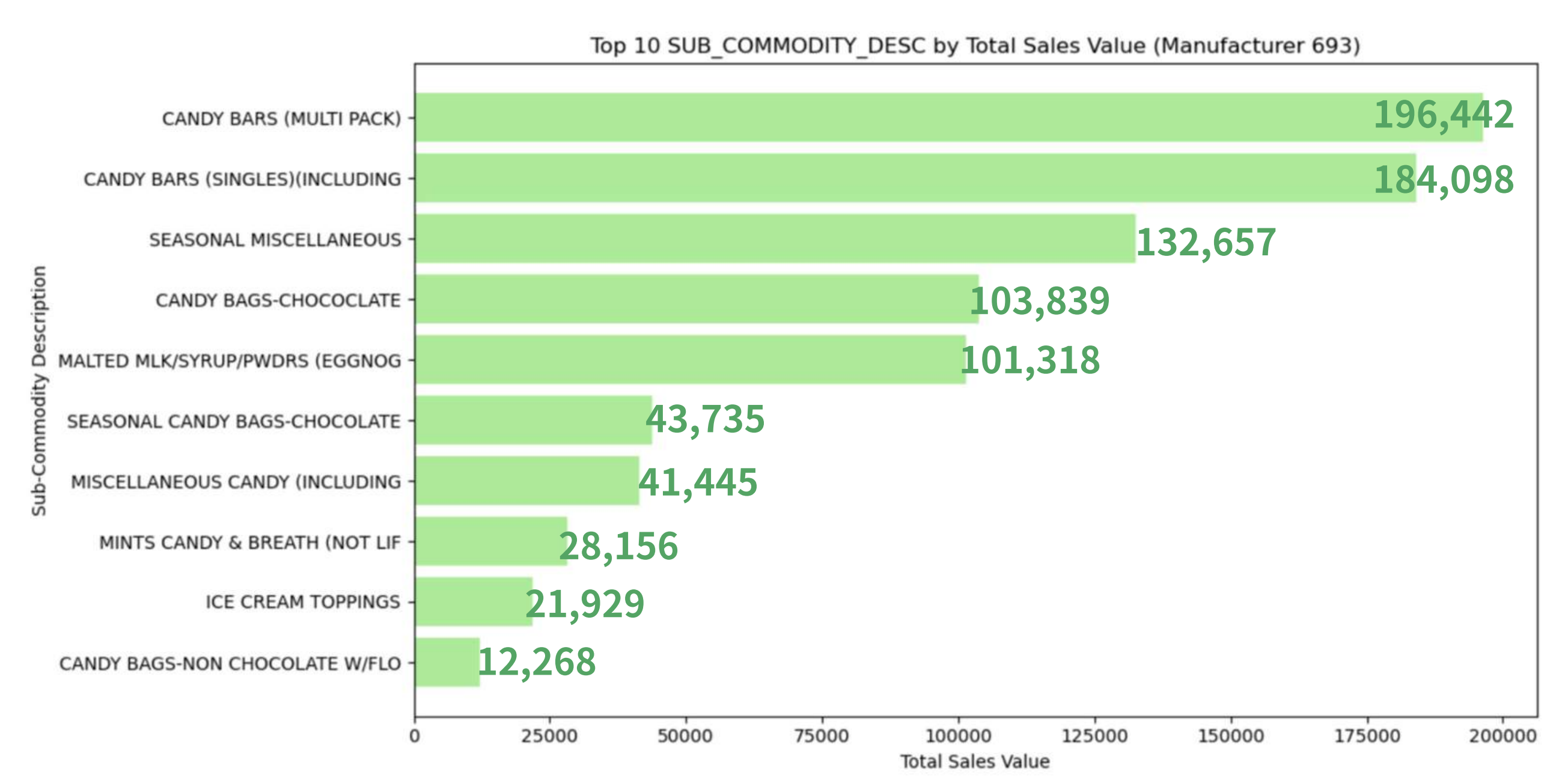


Top Performing Categories:

- **CANDY BARS (MULTI PACK)** - High demand for bulk-packaged candy bars.
- **CANDY BARS (SINGLES)** - Single-pack candy bars also play a significant role in driving sales.
- **SEASONAL MISCELLANEOUS** - Seasonal products contribute notably during specific periods.

Detailed Analysis of Top Feature

Top 10 Sub-Commodity Insights



Mid-Tier Sub-Commodity Categories:

- **CANDY BAGS - CHOCOLATE & MALTED MILK/SYRUP/POWDERS**

Lower-Tier Sub-Commodity Categories:

- **MINTS CANDY & BREATH & ICE CREAM TOPPINGS** - show lower sales but still have potential.

Detailed Analysis of Top Feature

How to Optimize Our Product Portfolio?

Shared Top-Performing Sub-Categories

Focal brand (857)

- 1. CANDY BAGS-CHOCOLATE W/FLOUR - 198,034
- 2. CANDY BARS (MULTI PACK) - 92,956
- 3. CANDY BAGS-CHOCOLATE - 90,131
- 4. CANDY BARS (SINGLES)(INCLUDING) - 73,428
- 5. CANDY BAGS-NON CHOCOLATE - 18,938

Competitor(693)

- 1. CANDY BARS (MULTI PACK) - 196,442
- 2. CANDY BARS (SINGLES)(INCLUDING - 184,098
- 3. SEASONAL MISCELLANEOUS - 132,657
- 4. CANDY BAGS-CHOCOLATE - 103,839
- 5. MALTED MLK/SYRUP/PWDRS (EGGNOG) - 101,318

CANDY BARS (MULTI PACK)

Competitor 693 significantly outperforms in this category.

- ➔ Invest more in multi-pack chocolate bars by introducing attractive promotions or bundle deals to capture more market share.

CANDY BAGS - CHOCOLATE

Both brands perform similarly, indicating stable demand.

- ➔ Maintain our current product line and consider adding new flavors or limited editions to attract more customers.

CANDY BARS (SINGLES)

Competitor 693 excels in single chocolate bars, suggesting strong market potential.

- ➔ Expand your single chocolate bar offerings and enhance checkout displays to encourage impulse buying.

Detailed Analysis of Top Feature

How to Optimize Our Product Portfolio?

Our Unique Strengths

Focal brand (857)

- 1. CANDY BAGS-CHOCOLATE W/FLOUR - 198,034
- 2. CANDY BARS (MULTI PACK) - 92,956
- 3. CANDY BAGS-CHOCOLATE - 90,131
- 4. CANDY BARS (SINGLES)(INCLUDING) - 73,428
- 5. CANDY BAGS-NON CHOCOLATE - 18,938

Competitor(693)

- 1. CANDY BARS (MULTI PACK) - 196,442
- 2. CANDY BARS (SINGLES)(INCLUDING - 184,098
- 3. SEASONAL MISCELLANEOUS - 132,657
- 4. CANDY BAGS-CHOCOLATE - 103,839
- 5. MALTED MLK/SYRUP/PWDRS (EGGNOG) - 101,318

CANDY BAGS - CHOCOLATE W/FLOUR

This is a clear strength for 857.

➡ Maintain strong inventory and increase promotional efforts to solidify market dominance.

CANDY BAGS - NON CHOCOLATE

While sales are lower, this category meets specific market needs.

➡ Consider launching new products or promotions to attract customers interested in non-chocolate options.

Business Recommendations

Suggstion

Leverage Retail Discounts

RETAIL_DISC (26.88%) is the most important feature for us.

Our customers are price-sensitive and respond well to loyalty programs and discounts.

- ➡ **Strengthen retailer partnerships** to offer exclusive discounts and loyalty program incentives.
- ➡ **Launch targeted discount campaigns** during peak shopping seasons to drive sales.

Encourage Bulk Purchases

QUANTITY (11.03%) is less influential for 857 compared to 693 (16.17%).

- ➡ Launch **bulk-buy promotions** to encourage customers to purchase in larger quantities.
- ➡ Consider **seasonal bundles** or limited-time offers to create urgency.

Suggstion

Optimize Product Categories

Top-performing categories

Packaged Candy and **Checklane Candy** are the top-performing categories for both brands, indicating consistent demand.

- ➡ Double down on these top categories by **expanding product variations** to differentiate from 693.
- ➡ **Enhance visibility** of these products in high-traffic areas, such as checkout counters.

Key Sub-Categories

1. Our top sub-category is **Chocolate Candy Bags with Flour**.
 - ➡ Maintain dominance in this sub-category through strong inventory and promotional efforts.
2. **Candy bags(non chocolate)** is a unique product for us, which 693 does not have an advantage in.
 - ➡ Introduce new flavors, healthier options, or fun shapes to **expand product variety**.
 - ➡ Focus on families, children, and consumers who prefer non-chocolate options, emphasizing the product's uniqueness.

Suggestion

Refine Targeting

The demographic feature has a very low impact , which may indicate that the current target audience definition is not precise enough.

- ➡ **Conduct market research** to understand which demographic characteristics are more likely to be interested in your products. Use these insights to design products tailored to them.
- ➡ **Perform A/B testing** with different family segments to identify which marketing messages or promotional strategies generate higher engagement and purchase rates.



**Thank
you!**