

PERSONALIZED PRODUCT RECOMMENDATIONS FOR HOUSEHOLD SEGMENTS

**GROUP 13: ANUSHA BHAT, AARAV DEWANGAN,
BRUNA MEDEIROS, MAHIMA MASETTY**
Data Science for Algorithmic Marketing

March 11, 2025



Agenda

1

Business Problem

2

Data Description & Exploratory Data Analysis

3

Customer Segmentation

4

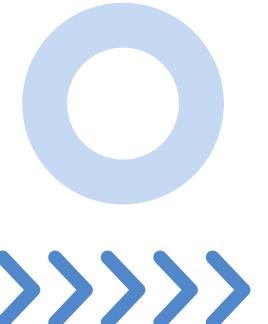
Association Rules

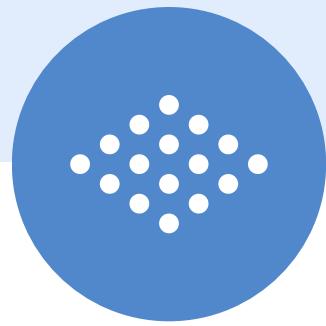
5

Collaborative Filtering

6

Recommendations & Future Steps





DATA DESCRIPTION AND EDA

Dunnhumby: A Complete Journey

Contains transaction-level data from a grocery store loyalty program, including customer purchases, discounts, and demographics.

The dataset consists of 3 main tables:

1

Transactions

~ 2.5 million rows

12 columns

One row per product bought

2

Products

~ 90 thousand rows

7 columns

One row per product

3

Demographics

801 rows

8 columns

One row per household

**BUSINESS
PROBLEM AND
GOAL**

The grocery retailer needs to personalize marketing and increase sales by understanding customer behavior. By segmenting shoppers, identifying product associations, and using collaborative filtering, we can help them optimize promotions, boost cross-selling, and improve customer retention.

Dunnhumby: A Complete Journey

Transactions

	Unique Values	
Quantity	0	to 90,000
Sales	0	to 840
Retail Discount	-180	to 3.99
Coupon Discount	-55.93	to 0
Coupon Match	-7.07	to 0
Discount		
Household ID		2,500
Basket ID		276,484
Product ID		92,339
Store ID		508
Day		1 - 711
Week		1 - 102

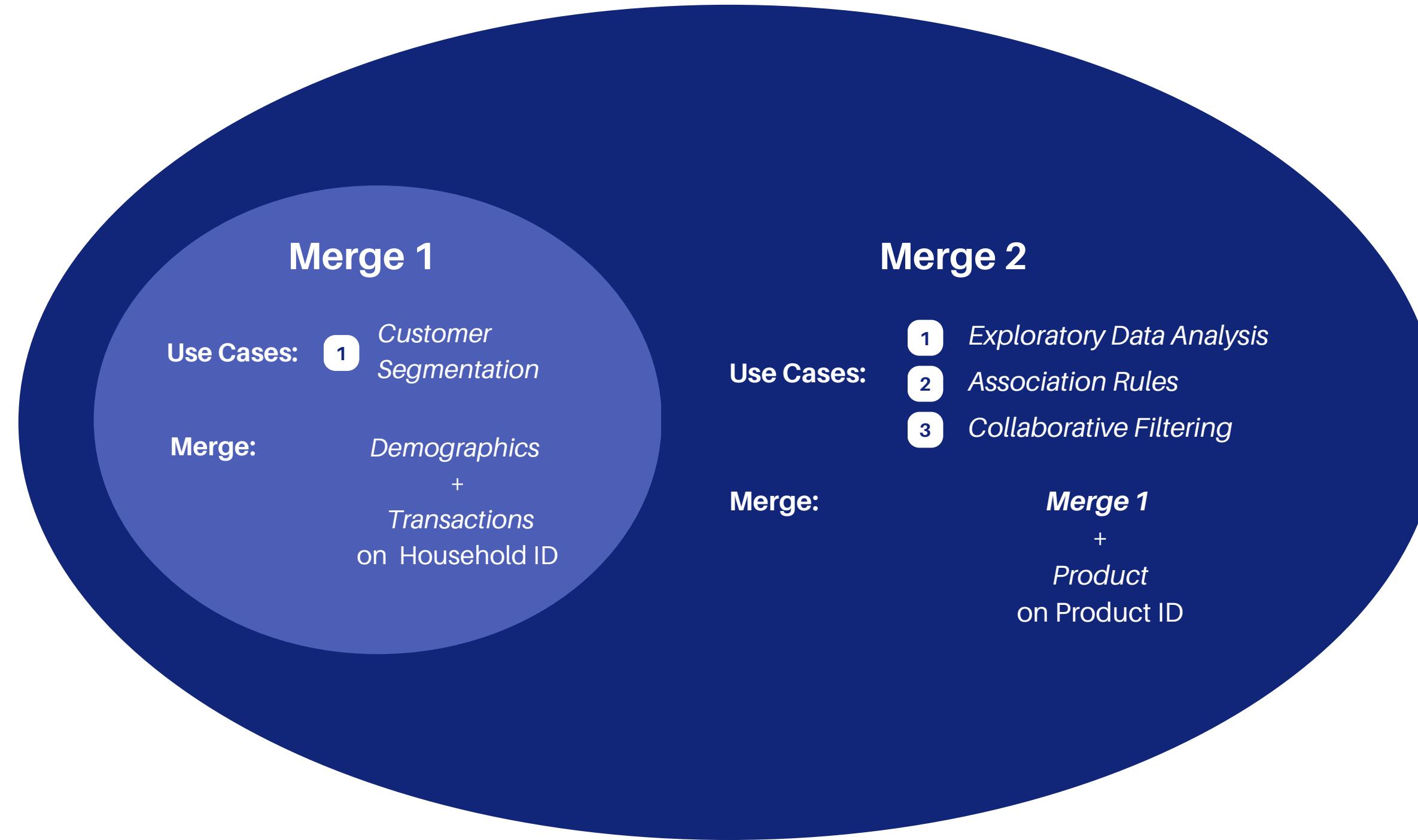
Products

	Unique Values	
Department		44
Brand	<i>Private or National</i>	
Commodity		308
Sub-Commodity		2,383
Product Size		4,345
Manufacturer ID		6,476
Product ID		92,353

Demographics

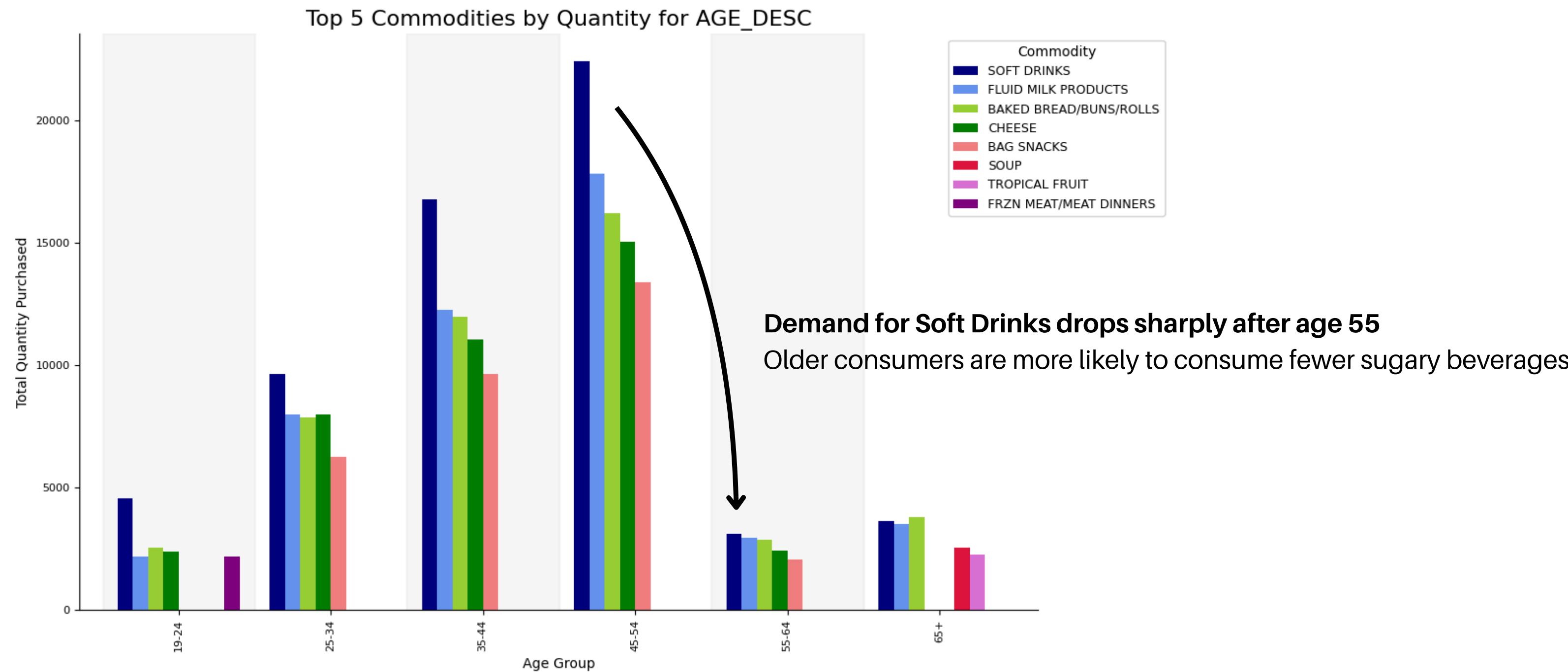
	Unique values	
Age range		6
Income range		12
Homeowner status		3
Marital status		3
Family/Household size		5
Household compositon		6
Number of children		4
Household ID		801

Dunnhumby: A Complete Journey

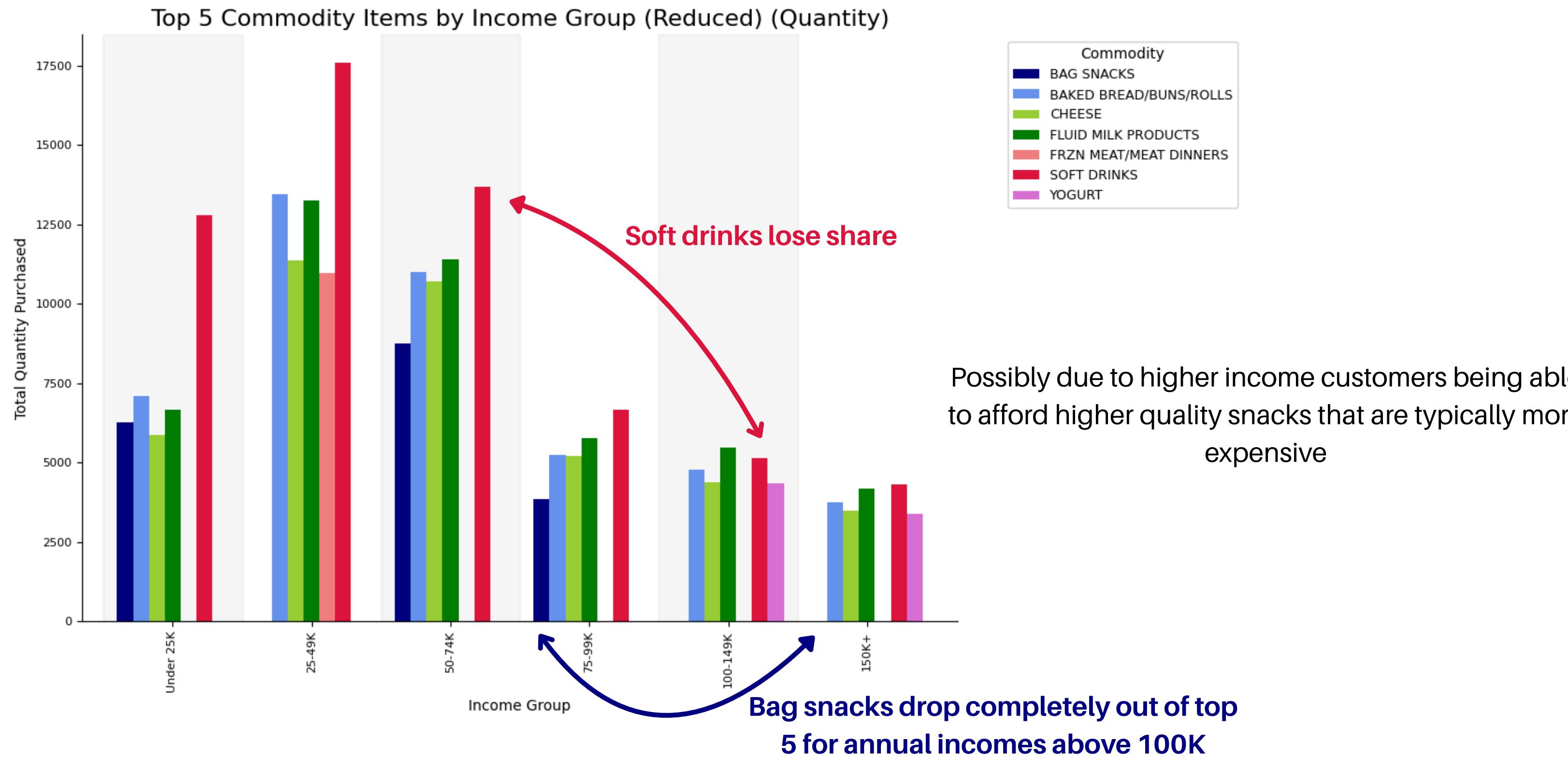


Milk, Soft drinks and Baked Breads/Buns are a Universal Commodity

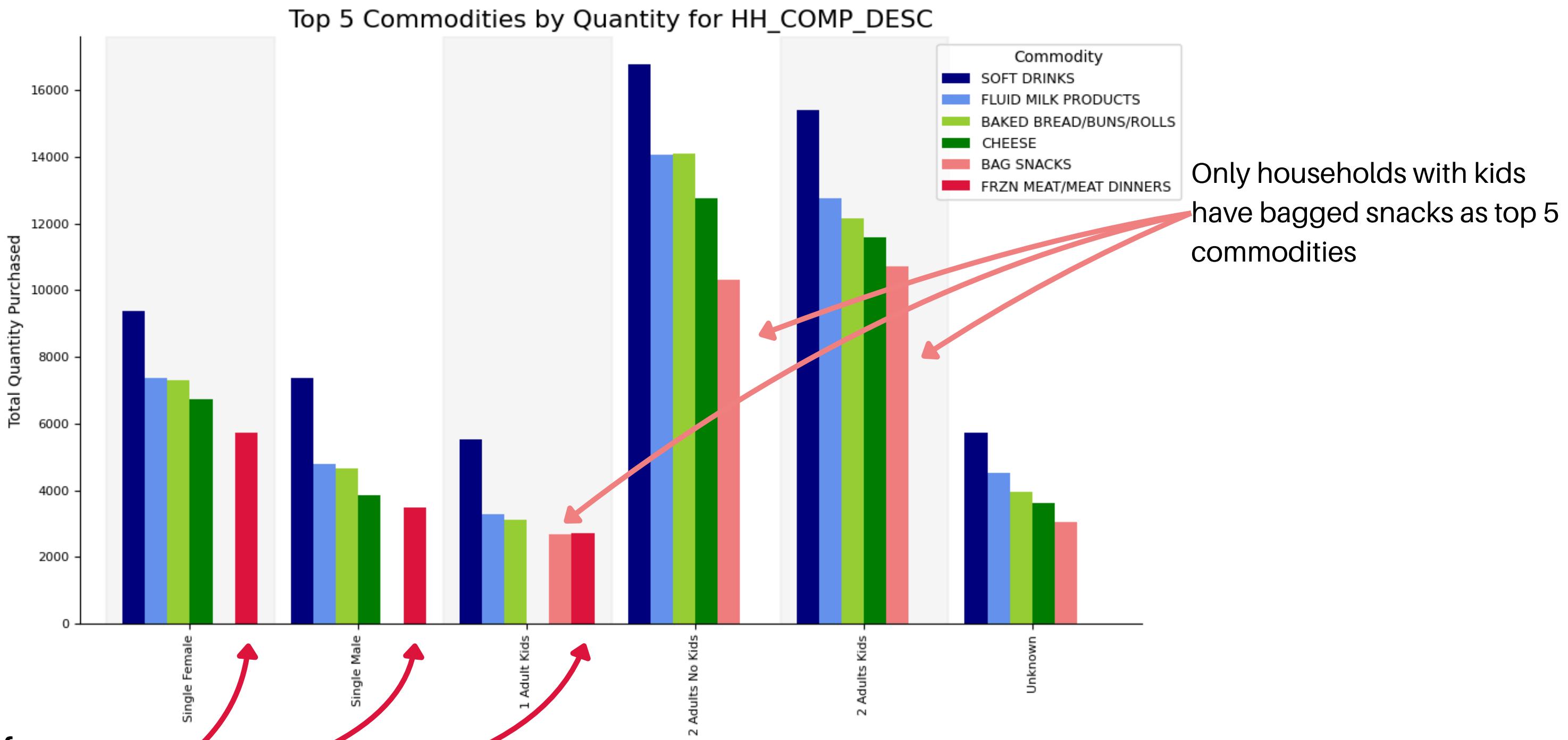
They appear in every demographic and are always demanded.



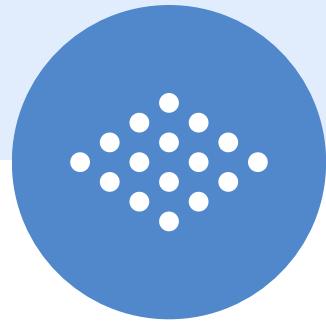
Soft drinks and Bag Snacks Are Popular Between Low-Middle Income Families



Households With Kids Demand More Bagged Snacks



Single households consume frozen meat/meat dinners
Possibly due to cost or convenience

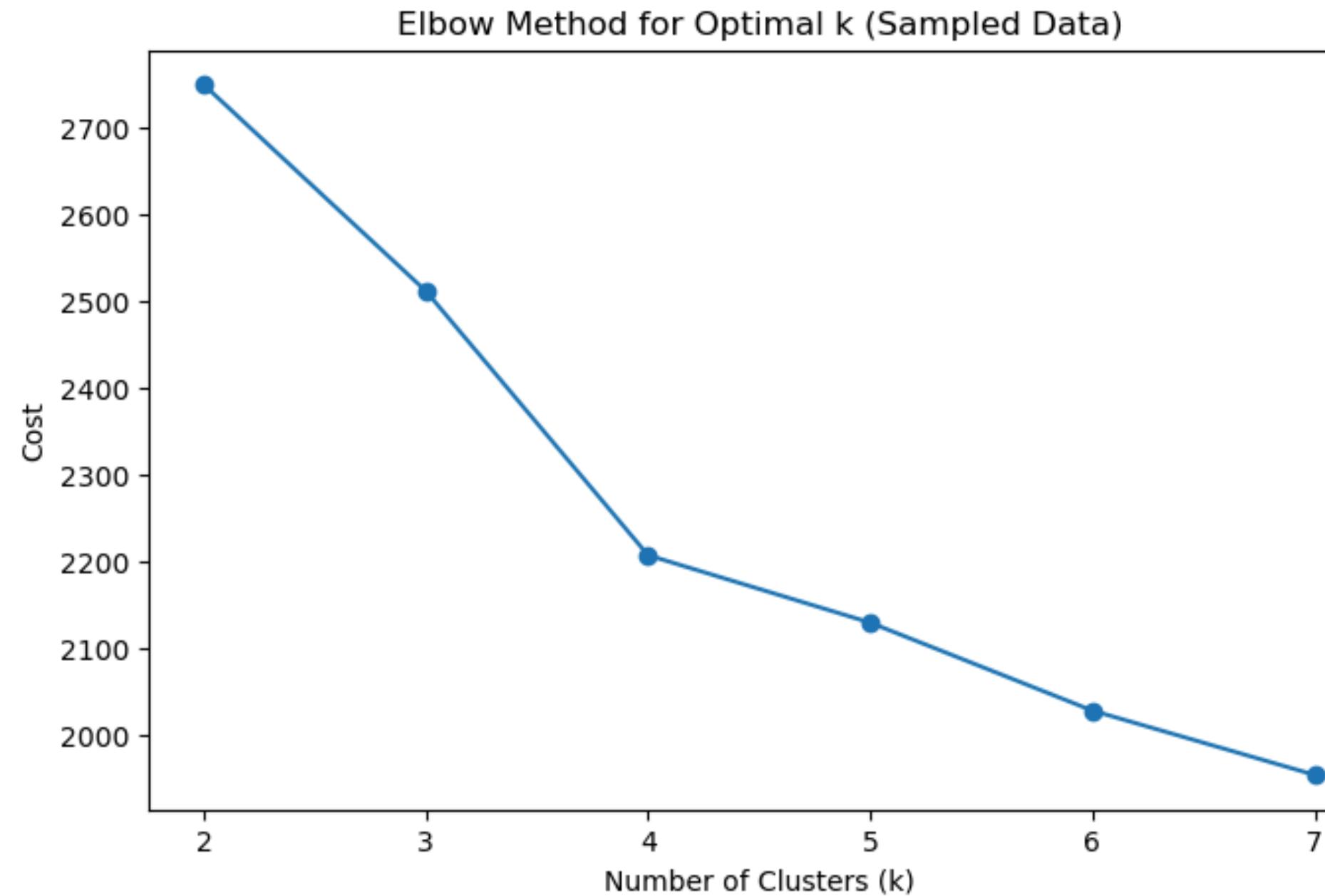


CUSTOMER SEGMENTATION



Segmenting Households Using K-Prototypes

- Our dataset, combining household and transaction data, contains both numerical and categorical features.
- K-prototypes effectively clusters mixed data by balancing Euclidean distance for numerical variables and Hamming distance for categorical ones, ensuring meaningful customer segmentation.



What is the Optimal K?

The elbow at $k=4$ marks where cost levels off, meaning more clusters diminishing returns in reducing the within-cluster variance.

Defining Target Personas for Each Segment

Cluster 0: Single, Low Spending Shoppers



Emily is a 45 year old independent bargain hunter who seeks affordable options that fit her modest income and single lifestyle.

- Age: 45-54 years
- Income: \$35K - \$49K (27.7%); \$50K - \$74K (18.5%)
- Purchase Volume: Low (76.6)
- Purchase Item Value: Low (\$2.85)
- Discount Reliance: Moderate (-\$0.45)
- Housing: Renters (72.7%)
- Household Type: Single female (48.9%) and male (26.2%)
- Household Size: Single person households (85.9%)
- Children: No children (95.1%)

Cluster 1: Young Families with Growing Needs



The Johnsons are a family in their late 30s who prioritize quality, quantity and discounts while managing a busy household with three members.

- Age: 35-44 years (46.8%)
- Income: \$75K - \$149K (40.5%)
- Purchase Volume: High (119.8)
- Purchase Item Value: Moderate (\$3.19)
- Discount Reliance: Moderate (-\$0.67)
- Housing: Homeowners (73.6%)
- Household Type: Nearly 80% are 2 adults with kids
- Household Size: 74.4% have 3 members; 16.2% have 4 members
- Children: 72.8% have at least 1 child

Defining Target Personas for Each Segment

Cluster 2: Large Price Conscious Households



The Marshalls are a middle-income family of homeowners aged 45-54, with a preference for moderate-priced items. They are less reliant on discounts, managing a large household with 3 or more children.

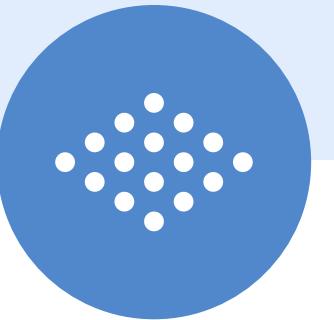
- Age: 45-54 years (41.8%)
- Income: Higher-income mix, 58.6% earn \$50K - \$149K
- Purchase Volume: Higher than average (90.4)
- Purchase Item Value: Moderate (\$3.07)
- Discount Reliance: Least reliant on discounts (-\$0.39)
- Housing: Homeowners (83.3%)
- Household Type: 72.6% are 2 adults with kids
- Household Size: 53.6% have 5+ members; 26.1% have 4 members
- Children: Very high proportion with 3+ children (55.8%)

Cluster 3: Wealthy Senior Empty Nesters



Walter and Mary are wealthy empty nesters in their late 40s and 50s, living in a two-person household. As homeowners with high spending power, they buy in large quantities but remain moderately discount-driven.

- Age: 45-54 years (42.4%)
- Income: Wealthier households, 44.9% earn \$75K - \$174K
- Purchase Volume: Highest among clusters (151.6)
- Purchase Item Value: Highest spending power (\$3.39)
- Discount Reliance: Moderate(-\$0.60)
- Housing: Highest proportion of homeowners (85.4%)
- Household Type: Mostly empty nesters (79.2% are 2 adults, no kids)
- Household Size: Predominantly 2-person households (89.8%)
- Children: Very few families with children (98.1% have none)



ASSOCIATION RULES



Association Rules for Cluster 0

Item 1	Item 2	Lift
Hamburger Buns	Hot Dog Buns	8.2
Sour Cream	Shredded Cheese	6.8
Dairy Case	White Bread	4.6
White Bread	Single Cheese	4.0
Bananas	Strawberries	3.9
Bananas	Dairy Case	3.8
Bananas	Adult Cereal	3.7
Milk	White Bread	3.7
Dairy Case	White Bread	3.7
Yogurt	Bananas	3.7



Emily prefers breakfast staples and dairy products, suggesting a preference for quick and nutritious morning meals with a focus on convenience and health. She also purchases both essential and indulgent products.

Marketing Strategies:

- **Bundle products:** create value packs featuring essential items that Emily frequently purchases (e.g. “Weekly grocery essentials”).
- **Discount offerings:** offer personalized discount or cashback for repeat purchases since Emily likes savings.

Association Rules for Cluster 1

Item 1	Item 2	Lift
Yogurt	Bananas	4.5
Bananas	Dairy Case	3.4
White Bread	Potato Chips	3.1
Shredded Cheese	Bananas	2.9
White Bread	Bananas	2.7
Potato Chips	Soft Drinks	2.7
Shredded Cheese	White Bread	2.7
Milk	Bananas	2.7
Snack Cake	Milk	2.6
Yogurt	Milk	2.6



The Johnsons prefer quick and balanced meals with a high preference for health snacks, fitting with their interest in premium products. They purchase a mix of everyday staples with occasional indulgent snacks.

Marketing Strategies:

- **In-store demo stations:** offer samples of healthy snacks and recipe cards to give the Johnsons family meal ideas and incentivize them to buy more groceries.
- **Limited time discounts:** offer short term discounts on healthy snacks to incentivize the Johnsons to make immediate purchases.

Association Rules for Cluster 2

Item 1	Item 2	Lift
Hamburgers	Hot Dog Buns	7.4
Strawberries	Bananas (fluid)	4.9
Bananas	Strawberries (fluid)	4.6
Bananas	Multigrain Bread	4.1
Bananas	Strawberries	3.7
Bananas	Eggs	3.7
Snack Cake	White Bread	3.7
Bananas	Shredded Cheese	3.6
Bananas	Adult Cereal	3.4
Bananas	Eggs (fluid)	3.3



The Marshalls prefer meal pairings and health snacks, often purchasing bananas and strawberries. They have a balanced diet with wholesome and convenient food choices fitting for a large family.

Marketing Strategies:

- **Family sized bundles:** create family-sized bundles of staple items that offer value and convenience for the Marshalls.
- **Meal planning solutions:** provide meal kits based on frequently bought items (e.g. milk, bananas, bread) to provide the large Marshall family with quick, easy meals that are moderately priced and convenient.

Association Rules for Cluster 3

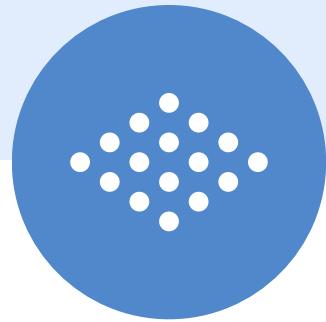
Item 1	Item 2	Lift
Adult Cereal	Milk	3.1
All Family Cereal	Milk	2.9
Bananas	Dairy Case	2.9
Soft Drinks	White Bread	2.8
Milk	Dairy Case	2.8
Milk	White Bread	2.6
Milk	Multigrain Bread	2.6
Eggs	Milk	2.6
Milk	Yogurt (individual pack)	2.5
Milk	Toilet Paper	2.4



Walter and Mary prefer staple, easy-to-prepare foods, for quick meals and tend to buy in bulk. They have a tendency to buy healthy and convenient snack options, making a mix of convenient and health-conscious choices.

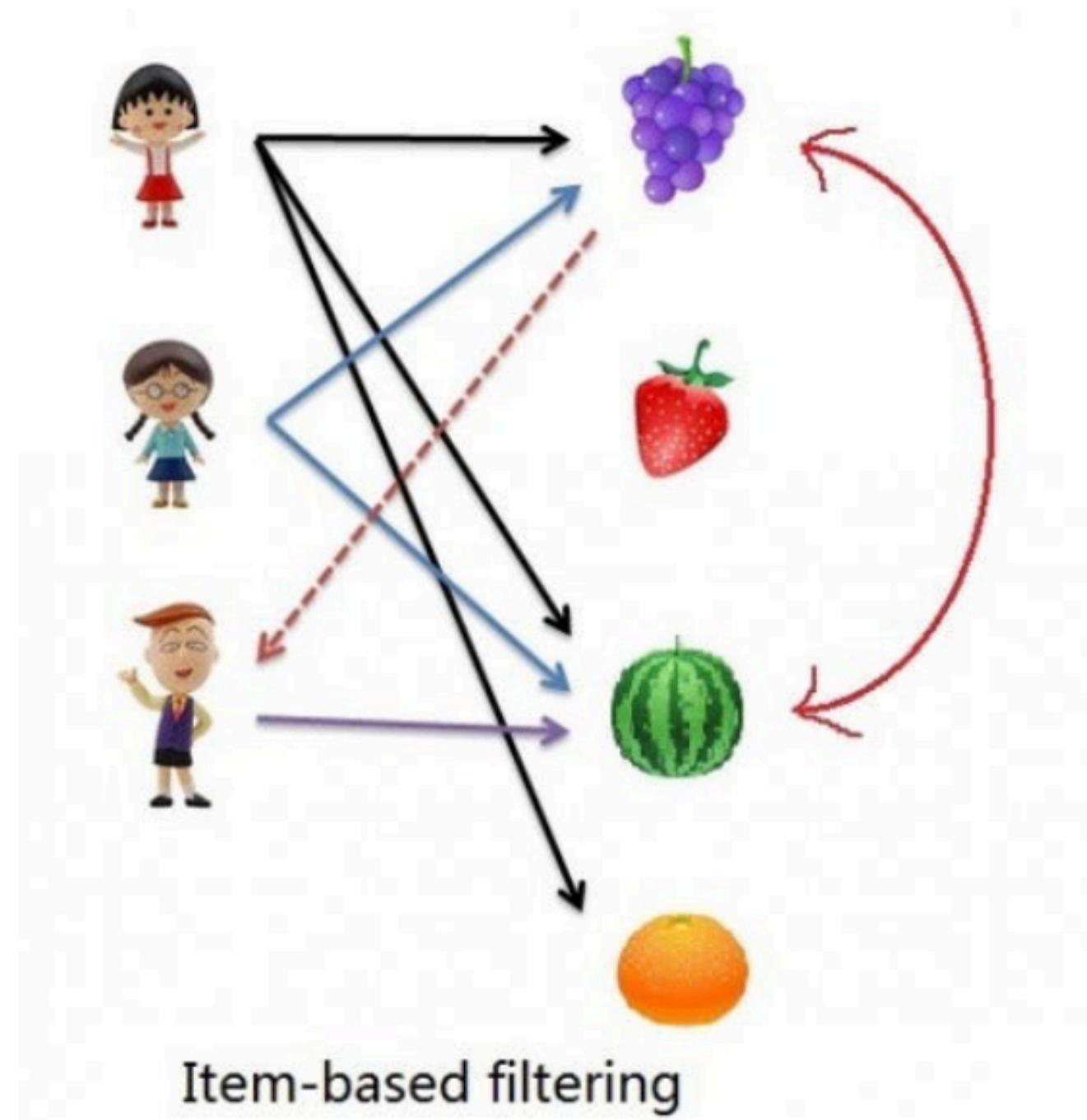
Marketing Strategies:

- **Premium convenience & health-oriented display:** set up displays in key areas of the store with health benefit signage to attract Walter and Mary to buy the healthier groceries.
- **Convenient bulk pickup and loyalty program:** offer bulk pickup for large orders to enhance the shopping experience for Walter and Mary and incentivize them to buy more bulk products by offering in-store rewards for repeat purchases.



COLLABORATIVE FILTERING

Collaborative Filtering Overview



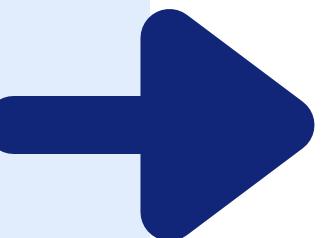
Steps:

1. Merged demographic data with transaction data, then merged that with product data.
2. For each cluster, created **User Product Interaction Matrix** based on quantity.
3. Applied **Singular Value Decomposition** to decompose the matrix into latent factors.
4. Predicted ratings for unseen products and generated top N recommendations.

Key Characteristics of CF:

- It relies on **user-item interactions** (e.g., purchase history, ratings).
- It does **not** use any information about the **attributes** of the items or users
- It assumes that users who interacted similarly in the past will have similar preferences in the future.
- We can apply this methodology to recommend products to households based on similar purchase habits from other households

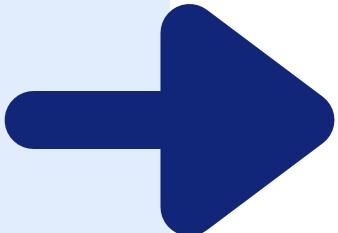
Recommendations For Emily



- 1 Pizza
- 2 Christmas lights
- 3 Stationary/School supplies
- 4 Soft drinks
- 5 Frozen meat/pre-made meat dinners

The focus on economy-focused, convenient, and affordable products matches the low spending power, single-person households, and shopping habits of people like Emily in Cluster 0.

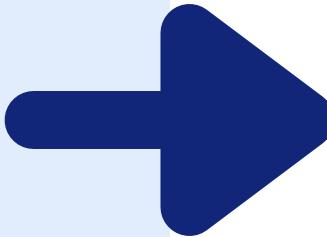
Recommendations for the Johnsons



- 1 Natural Beef
- 2 Kiwi
- 3 Refrigerated Speciality Rolls
- 4 Canned beans (baked w/ pork)
- 5 Seafood - tuna

The focus on family-friendly, high-quality, and convenient products matches the needs of young families with growing children like the Johnsons in Cluster 1.

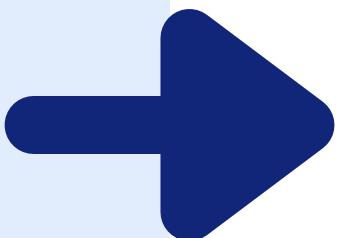
Recommendations for the Marshalls



- 1 Yogurt
- 2 Tuna
- 3 Frozen/Concentrated Orange Juice
- 4 Green Bell Peppers
- 5 Candy Green Bars

The focus on bulk-friendly, moderately priced, and convenient products matches the needs of large households with price-conscious buyers like the Marshalls in Cluster 2.

Recommendations for Walter and Mary



- 1 Strawberries
- 2 Lemon
- 3 Kiwi Fruit
- 4 Condensed Soup
- 5 Russet Potatoes

The focus on high-quality, fresh, and versatile products matches the needs of wealthy empty nesters and senior shoppers like Walter and Mary from Cluster 3.

Key Takeaways and Future Steps

Created Actionable Segments

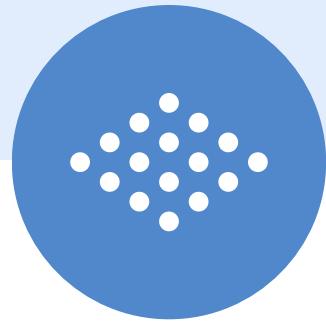
- K-prototypes clustering analysis reveals four distinct segments, each with unique shopping behaviors and preferences.
- These differences provide a strong foundation for targeted segmentation.

Enhanced Targeted Marketing

- Each segment exhibits distinct association rules and product recommendations, enabling personalized marketing strategies.
- Key approaches include tailored promotions, optimized product bundles, and strategic discounting to maximize engagement and sales.

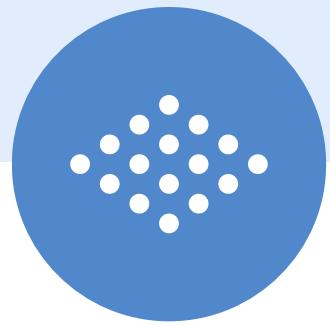
Future Steps

- Measuring the impact of digital marketing campaigns and the proposed strategies by tracking key performance metrics.
- Analyzing cross-price elasticity will further refine pricing and promotional tactics.



THANK YOU!





APPENDIX



Transactions

Summary

~ 2.5 million rows

12 columns

Additional Information

Each Basket ID uniquely identifies a purchasing occasion.

If 2 products are purchased, there will be **2 rows** dedicated to this purchase.

Each row will have the **same Basket and Household ID**, but different **Product ID**.

Quantity and **Sales** are specific to each product, not entire purchase basket.

Variables	Description	Unique Values
Quantity	Units of a product purchased	
Sales	Dollars retailer receives from sale	
Retail Discount*	Discount from retailer	
Coupon Discount*	Discount applied via manufacturer	
Coupon Match Discount*	Additional store discount	
Household ID	Household unique identifier	2,500
Basket ID	Basket unique identifier	276,484
Product ID	Product unique identifier	92,339
Store ID	Store unique identifier	508
Day	Day of transaction	1 - 711
Week	Week of transaction	1 - 102

*Negative values. Converted to positive for easier interpretation

Transactions

But some variables have extreme **outliers**...

Variables	Ranges	But 90% of values fall within	Ranges
Quantity	0 to 90,000		1 to 3
Sales	0 to 840		0.6 to 8
Retail Discount	\$(-180) to \$3.99		\$(-2.2) to \$0
Coupon Discount	\$(-55.93) to \$0		\$0 to \$0
Coupon Match Discount	\$(-7.07) to \$0		\$0 to \$0

Demographics

Summary

801 rows

8 columns

One row per household

Variables	Unique Values	Description
Age range	6	19 to 65+
Income range	12	under 15K to 250K+
Homeowner description	3	Homeowner, renter or unknown
Martial status	3	Married Single, and Unknown
Household size	5	1 to 5+
Household composition	6	ex: 1 Adult no kid, 2 Adult Kids
Number of kids	4	1 to 4+
Household ID	801	

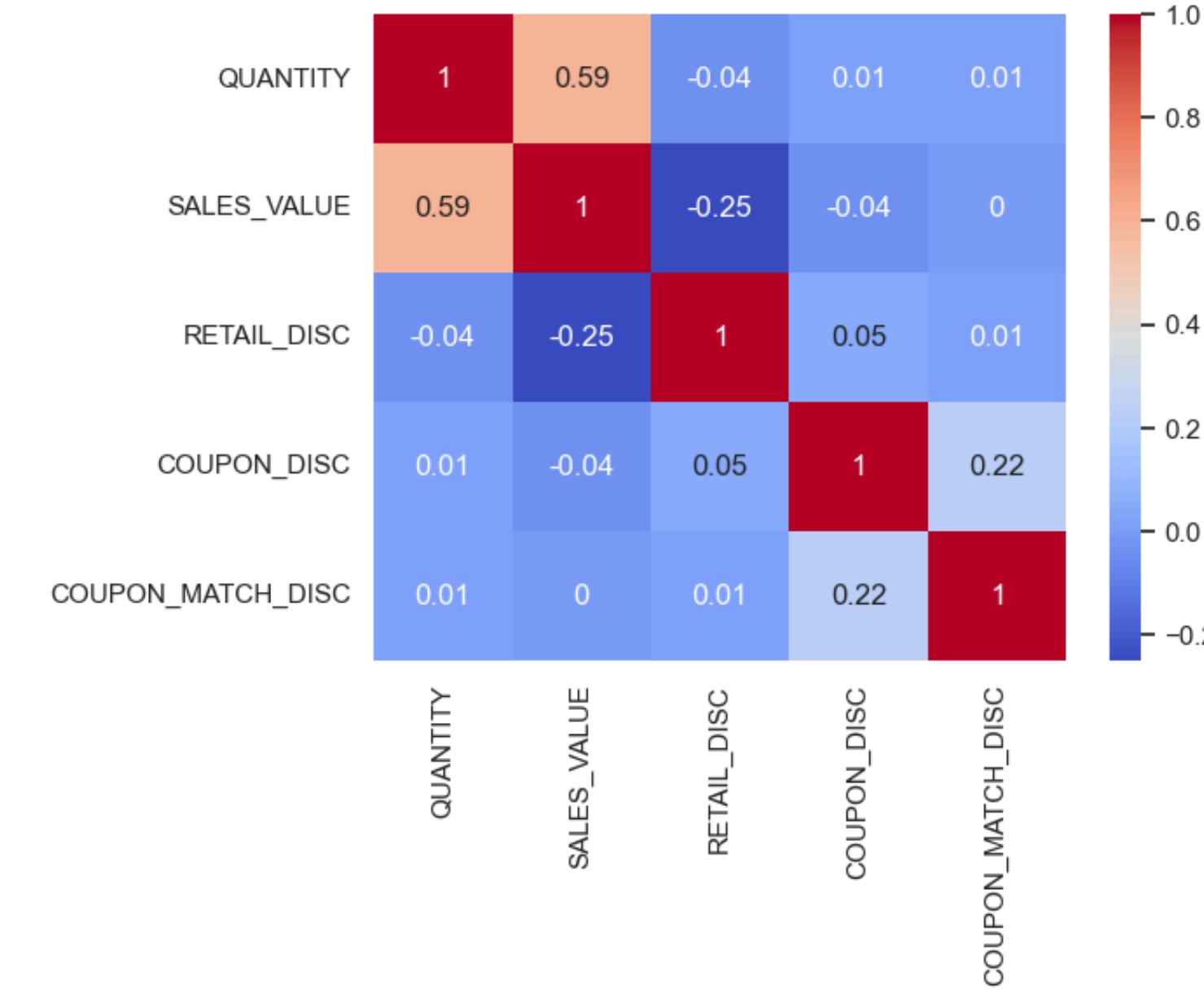
Insights

Majority of the surveyed clients:

- between 25 to 54 years old
- between \$35,000 and \$99,000 for income
- married or unknown
- homeowners
- 2 Adults without kids

Transactions

Correlation matrix confirms relationship between **sales** and **quantity**, but debunks hypothesis of strong relationship with **retail discounts**



Product

Summary

~ 90 thousand rows

7 columns

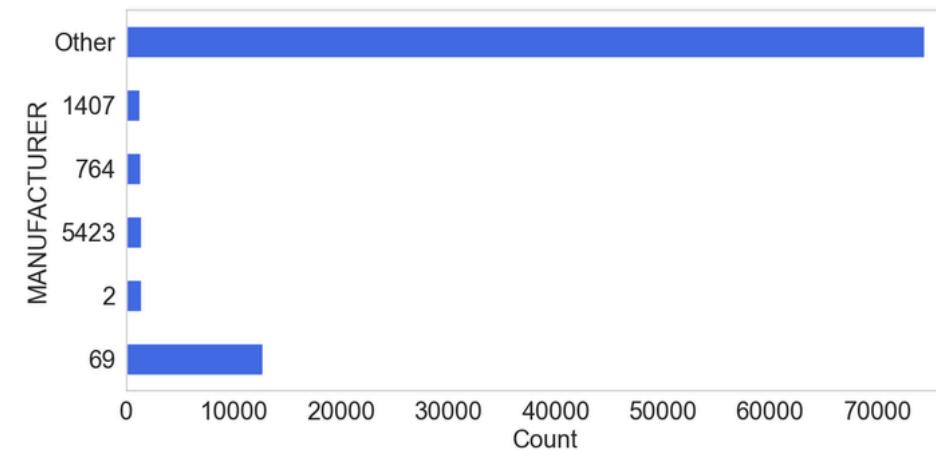
Each row represents one product

Every product ID shows up only once

Variables	Unique Values	Description
Department	44	
Brand	2	Private or National
Commodity	308	
Sub-Commodity	2,383	
Product size	4,345	
Manufacturer ID	6,476	
Product ID	92, 353	

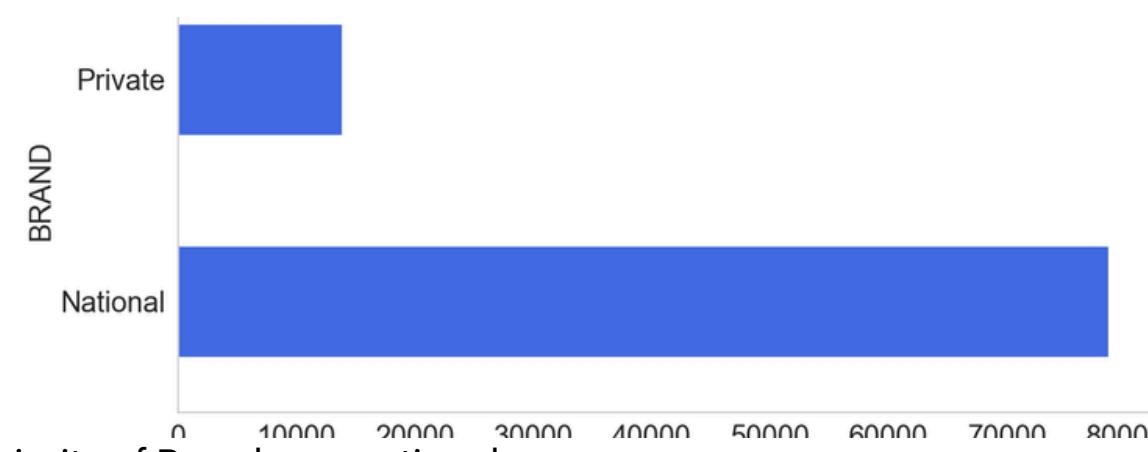
Product

Manufacturer ID Frequency



Concentrated manufacturer landscape with one dominant players and many smaller ones grouped as "Other".

Brand



Majority of Brands are national.

Top 5

Department

- Grocery
- Drug GM
- Produce
- Cosmetics
- Nutrition

Commodity

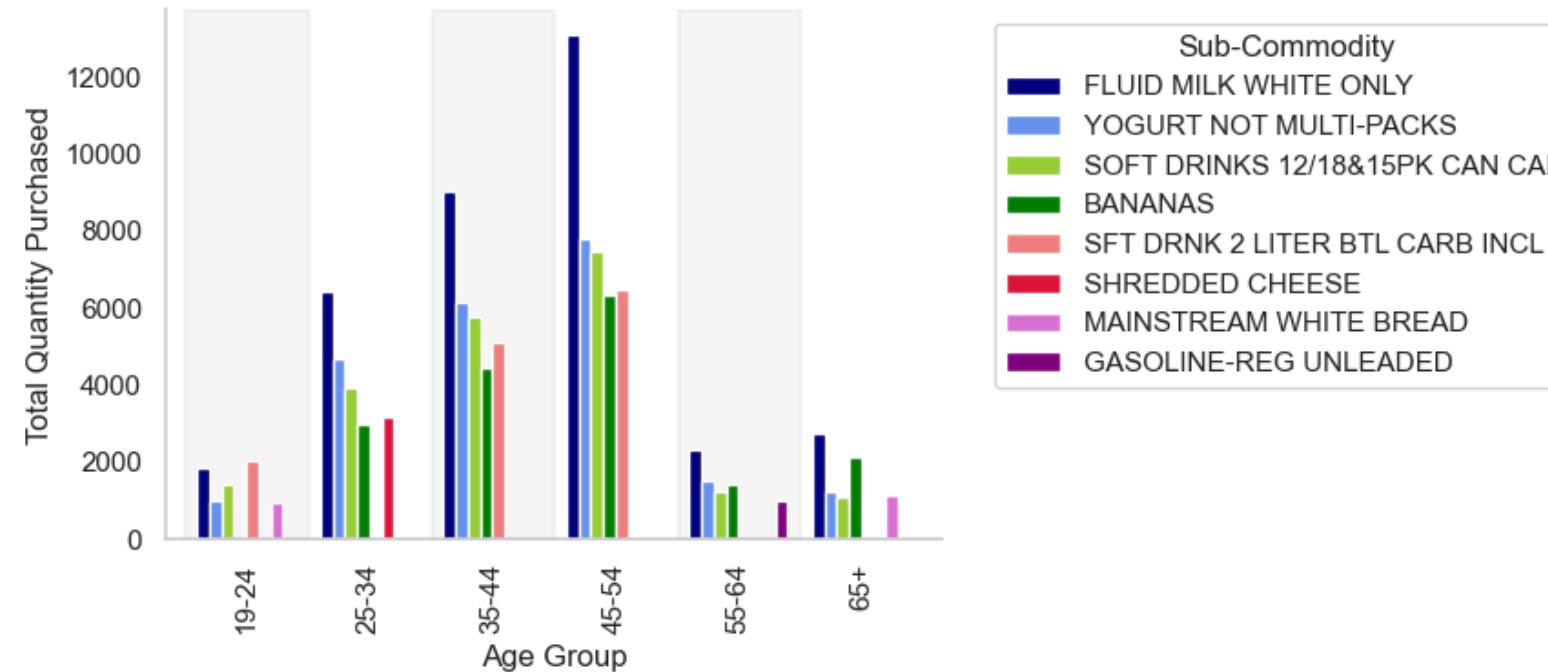
- Greeting cards/Wrapping/Party supply
- Candy
- Makeup and Treatment
- Hair Care
- Soft Drinks

Sub-Commodity

- Cards
- Beeralemalt Liquor
- Spices & Seasonings
- Gift-wrap
- Potato Chips

Sub-Commodity Preferences by Demographics

Top 5 Sub-Commodity for Age Group (Quantity)

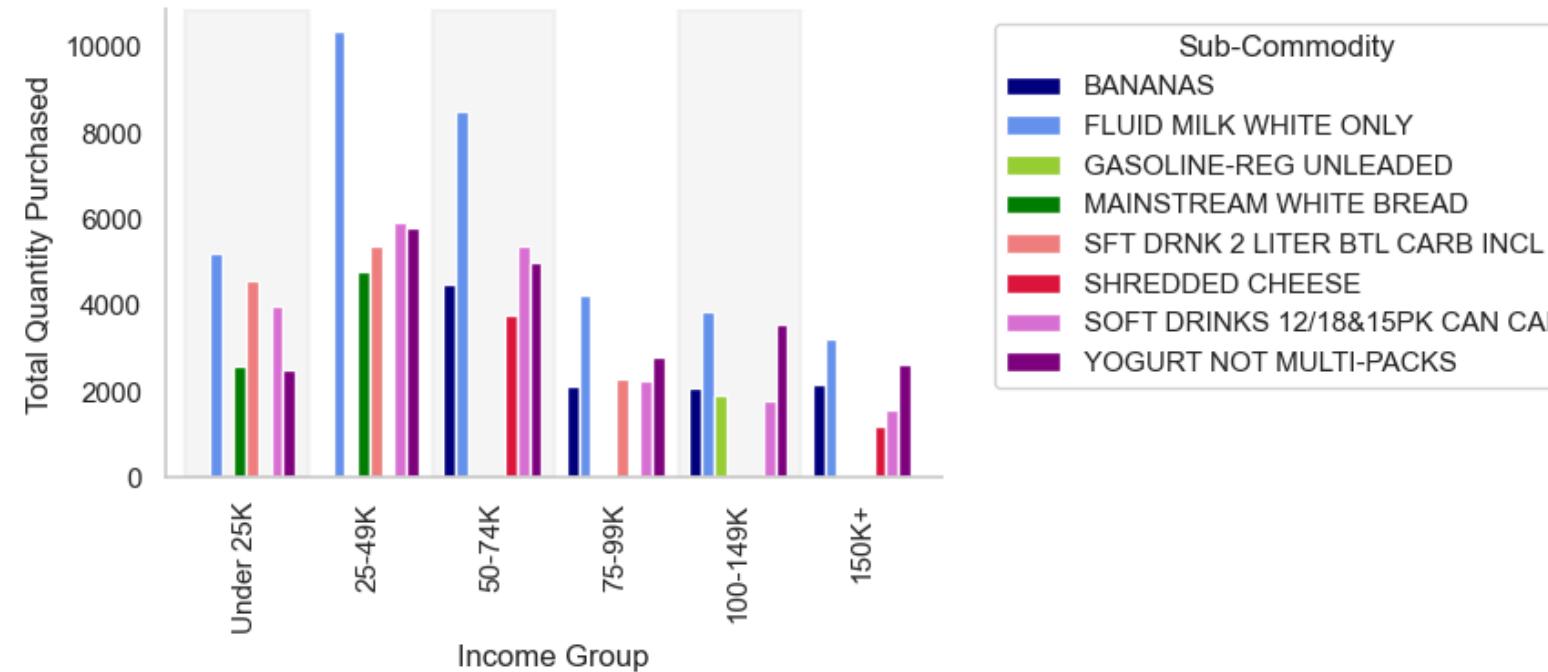


Insights

While **fluid milk** and **soft drinks** remain a household staple, **yogurt** and **bananas** grow to be household staples, as well.

Between all the **breads/buns/rolls**, **white bread** comes out as the most popular one.

Top 5 Sub-Commodity for Income Ranges (Quantity)

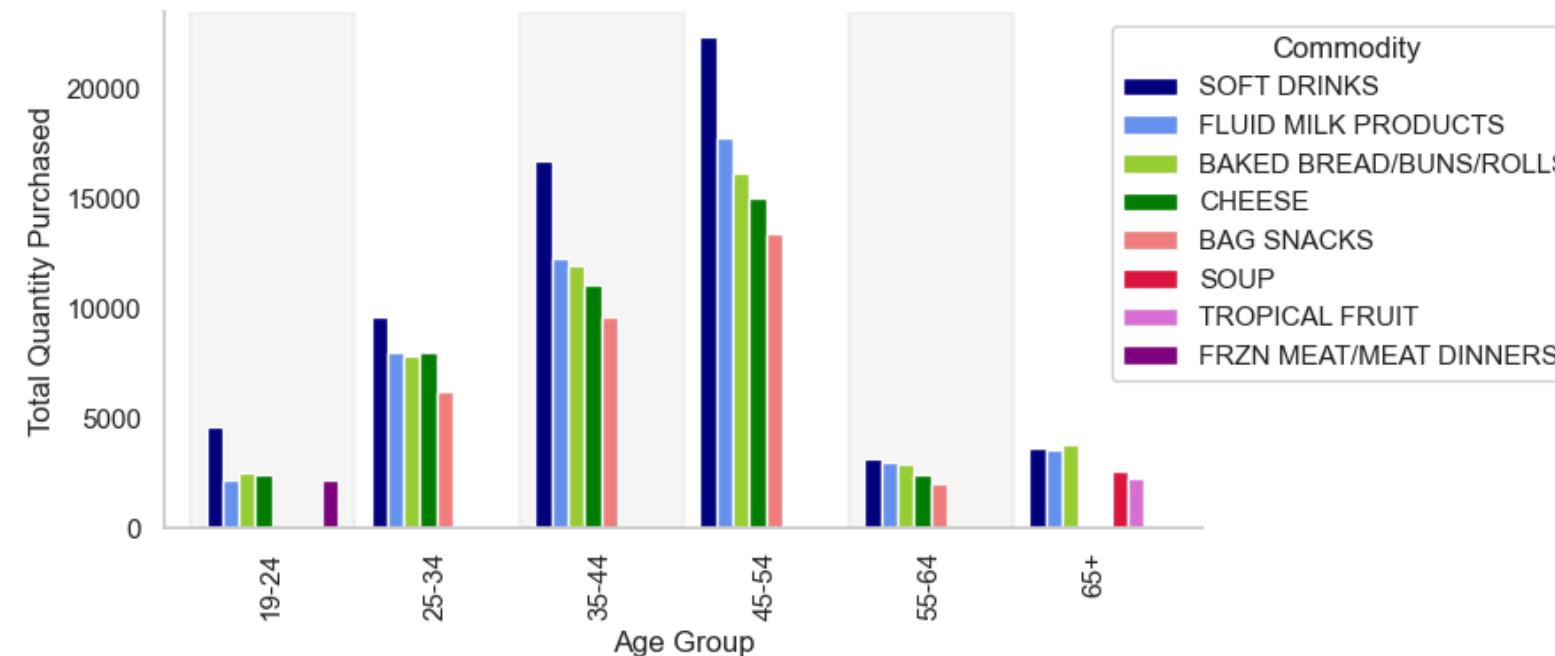


Bananas become prominent for **older age groups** (65+) and **higher incomes**, indicating a shift towards fresh produce.

Following same pattern in commodities, **soft drinks** (12/15 pk cans and 2L bottles) stay more popular between younger and low to middle-income demographics.

Commodity Preferences by Demographics

Top 5 Commodities by Age Group (Quantity)

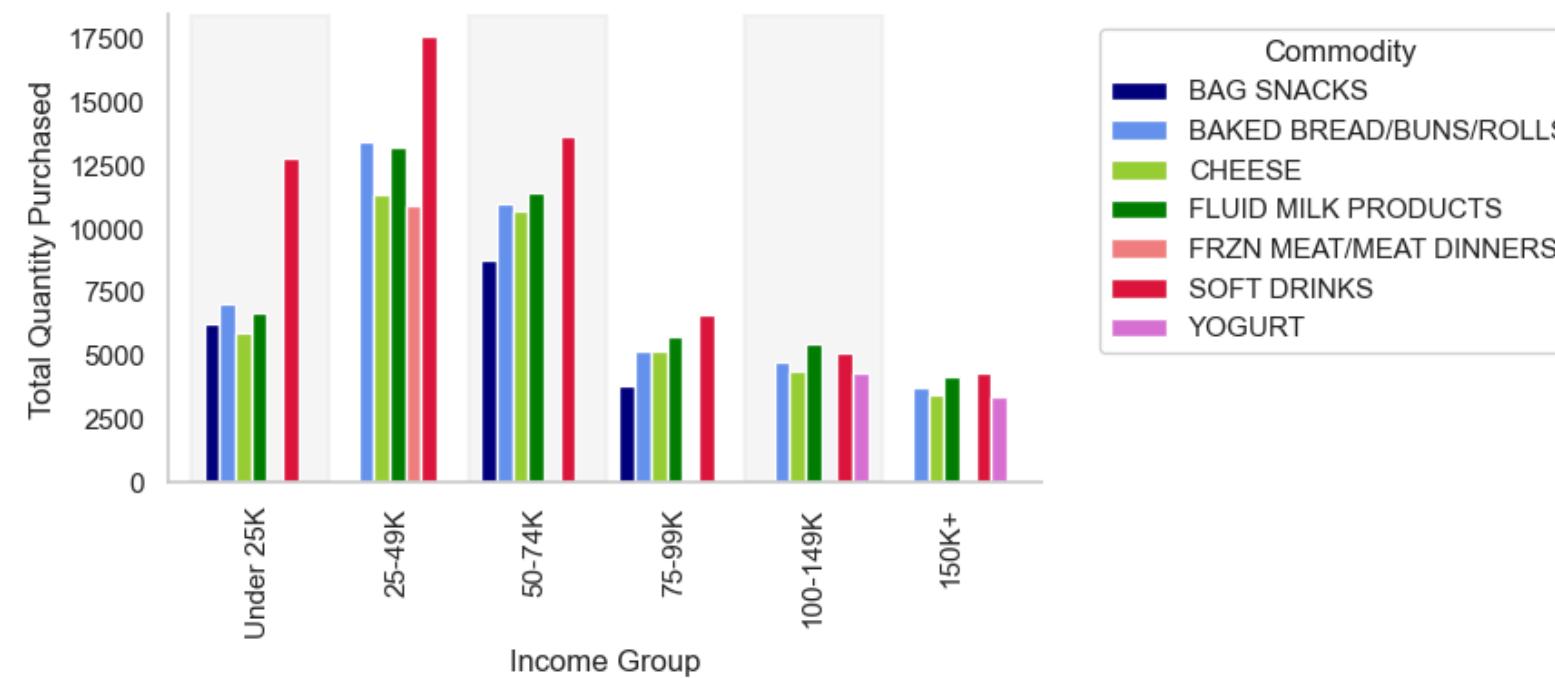


Insights

Fluid milk, soft drinks and baked breads/buns/rolls are an universal commodity. They appear in every demographic and are always demanded.

Demand for Soft Drinks drops sharply after age 55. Older consumers are more likely to consume fewer sugary beverages.

Top 5 Commodities for Income Ranges (Quantity)



Lower and middle-income families consume more soft drinks and bagged snacks, possibly due to cost, convenience, or younger consumer segments.