

Final Project

ENHANCING CROSS-SELLING WITH MACHINE LEARNING ON PRODUCT RECOMMENDATION



Instructor: Assoc. Prof. Ho Trung Thanh, Ph.D.

Presented by: Group 03

Explore Now

GROUP MEMBER



Thai An

K224111481



Thuy Dung

K224111482



Dang Khoa

K224111487



Khanh Linh

K224111488



Tuyet Quan

K224111501



Table of contents

- 1 Business understanding
- 2 Data understanding & EDA
- 3 Model building & Data validation
- 4 Visualization
- 5 Conclusion & Future work

Business problems



Problem 01

Right product to right customer

- Identifying responsive customer groups based on preferences and demands.
- Aligning product recommendations with customer needs to ensure relevance.



Problem 02

Right timing

- Predicting customer needs and return periods.
- Poorly timed recommendations risk appearing irrelevant.



Problem 03

Right communication channel

- Ensuring consistent communication across multiple platforms.
- Building trust and customer loyalty is key to long-term success.

Business objectives

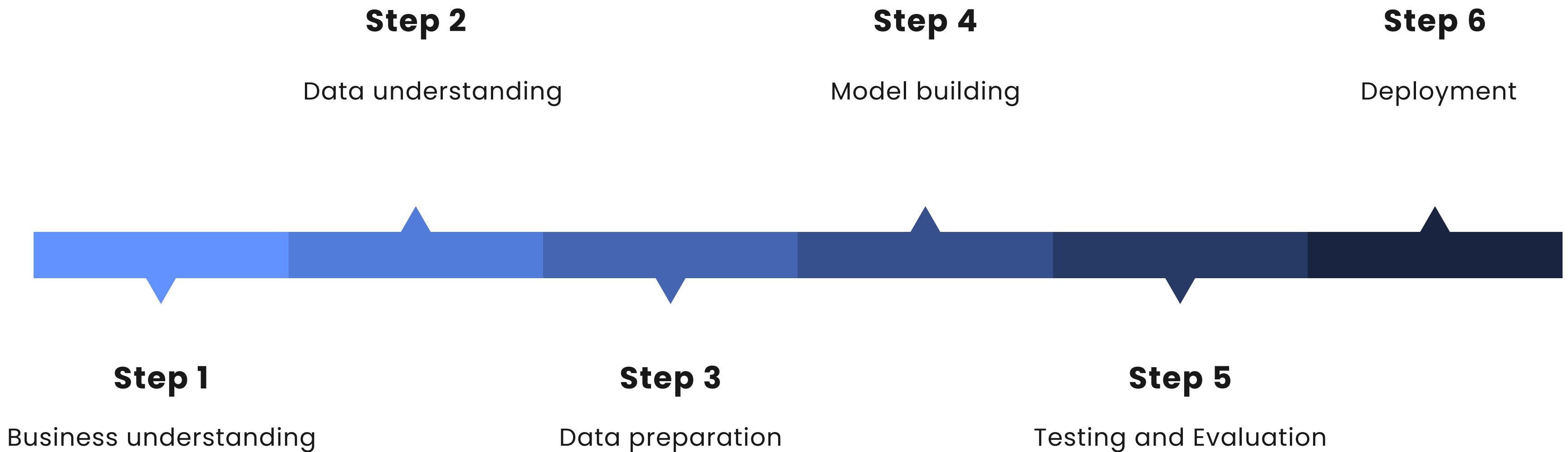
- Analyze past customer data to predict **sales patterns** and **operational results**.
- Identify customer behavior by spotting **engagement trends** and understanding **purchasing behaviors**.
- Improve retention and cross-selling plans by customizing marketing strategies and targeting customer groups based on cross-selling insights.
- Deliver practical insights that support **data-driven decisions** and promote **long-term business growth**.

Business questions

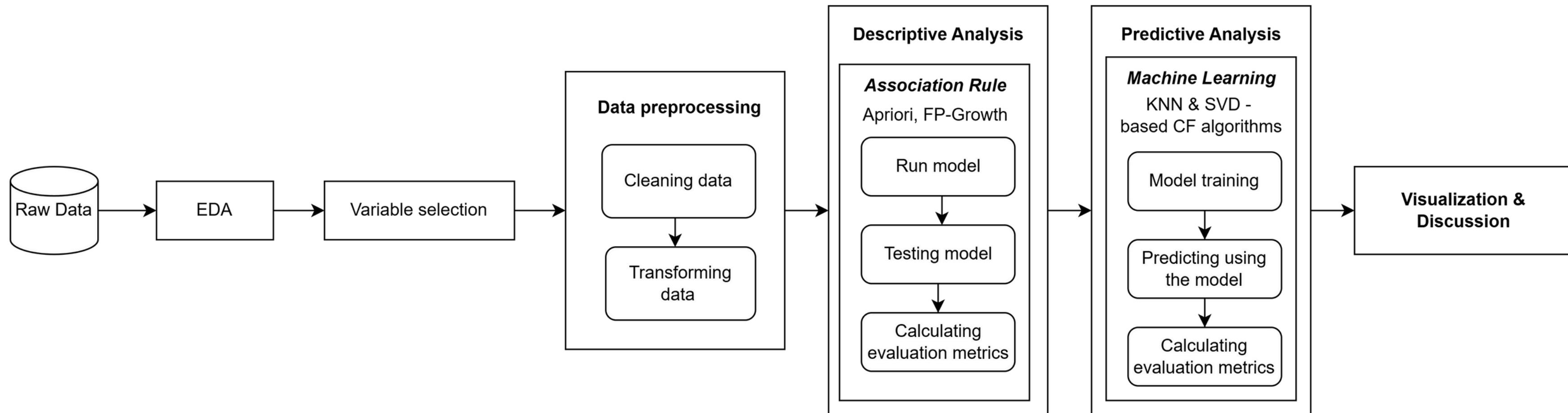
1. Which products or services should be recommended to customers?
2. What are the common characteristics of these product categories?
3. How can we predict and suggest relevant products to customers based on hidden patterns in their purchase behavior to increase cross-selling and sales?
4. Which product combinations are frequently purchased together?

Experimental method

The project uses the CRISP-DM process:



Data Preprocessing pipeline



Data Understanding



Dataset Overview

The dataset includes **transaction data** from **2,500 households**, with details on **purchases, demographics, and promotional campaigns**. The focus is on identifying frequently co-purchased products and customer purchasing patterns to optimize cross-selling strategies.

Key Data Types:



Product Data



Coupon Data



Transaction Data



Demographics



Goals

To identify **common traits** across product categories and **discover products frequently bought** together **answer business questions** and **aiding in personalized product recommendations** and cross-selling optimization.

EDA

Customer Behaviour

37.1 is average basket size of most households per transaction. With some transactions reaching up to 128 products.

\$25K is total spending of some households, making them outliers with significantly higher spending than most others in the dataset.

Demographic and Target Audience

35-54 is the range of age of largest customer group

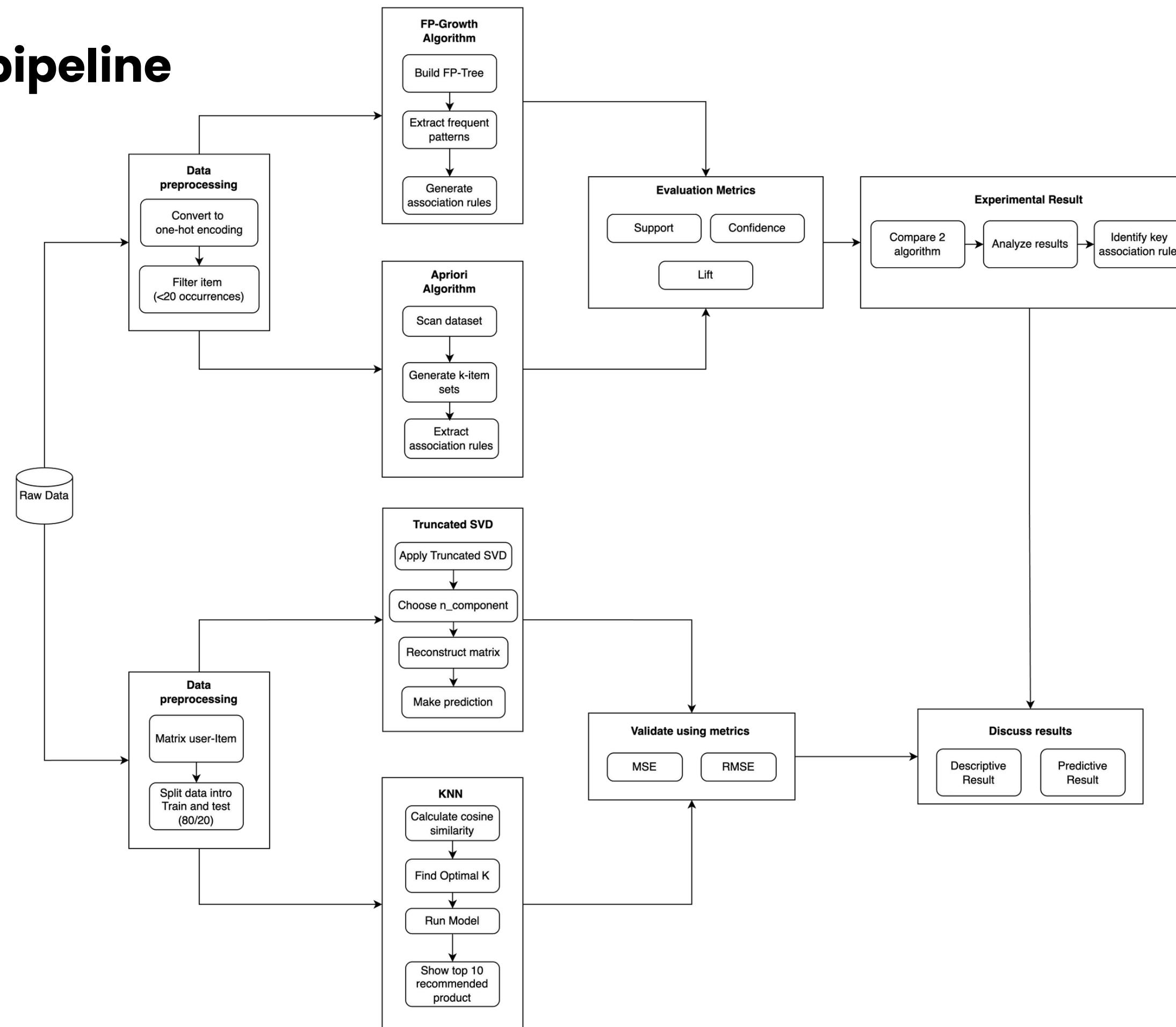
\$50K-\$74K is the income range of the majority of households in dataset

Correlation Insights

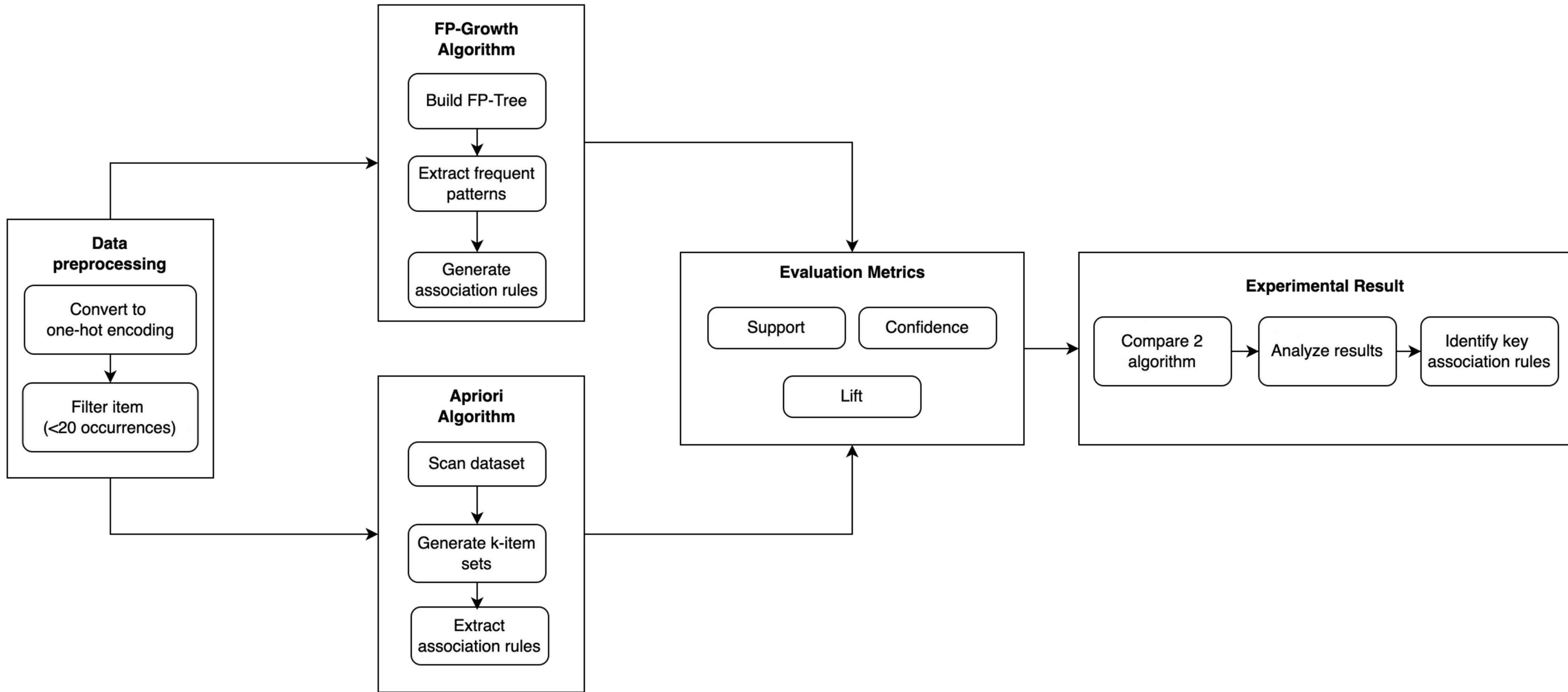
Strong correlation (0.71) between total campaigns and total redeemed coupons.

Total sales are positively correlated with the number of products purchased (0.63).

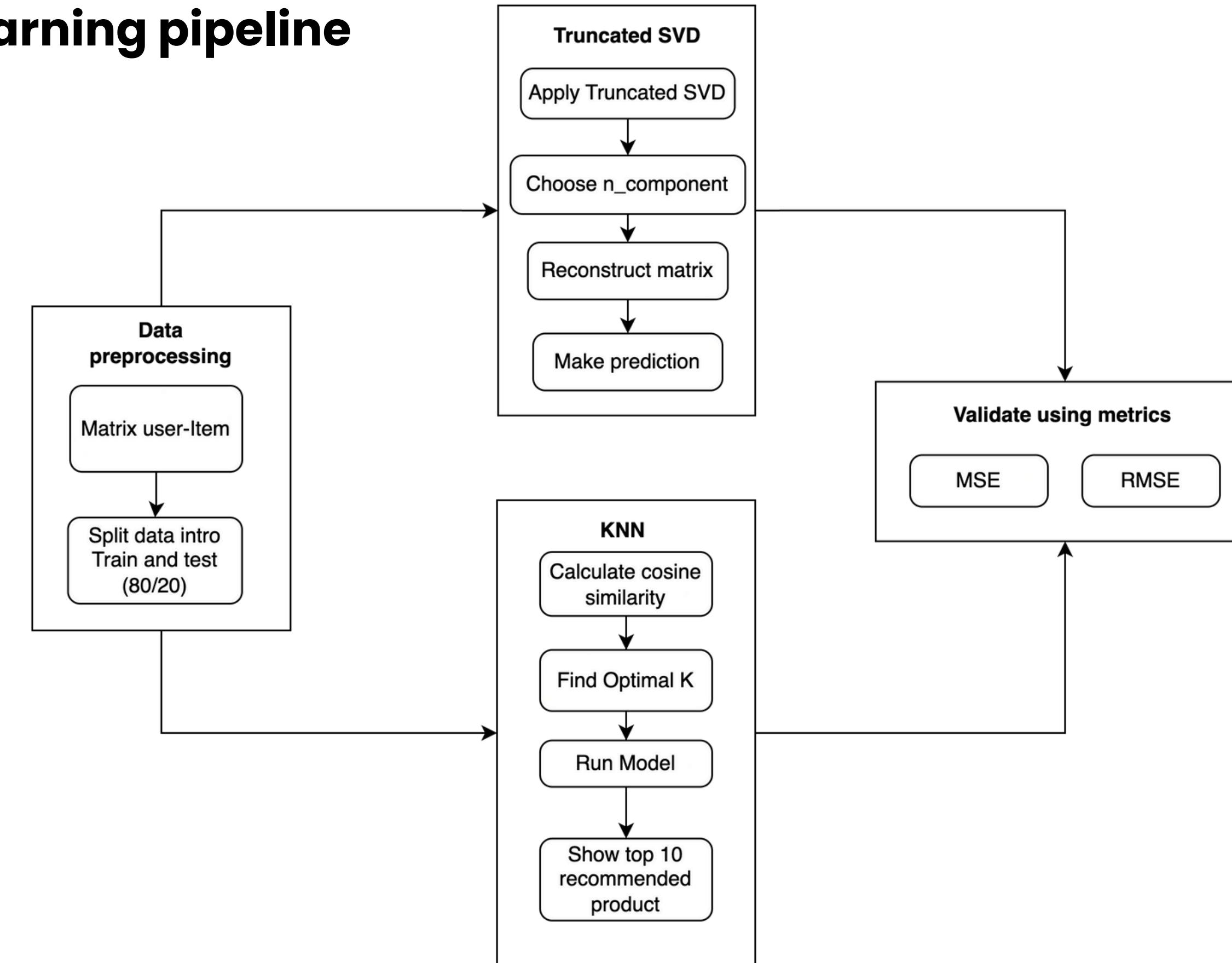
Overall model pipeline



Association Rule pipeline



Machine Learning pipeline



Validation

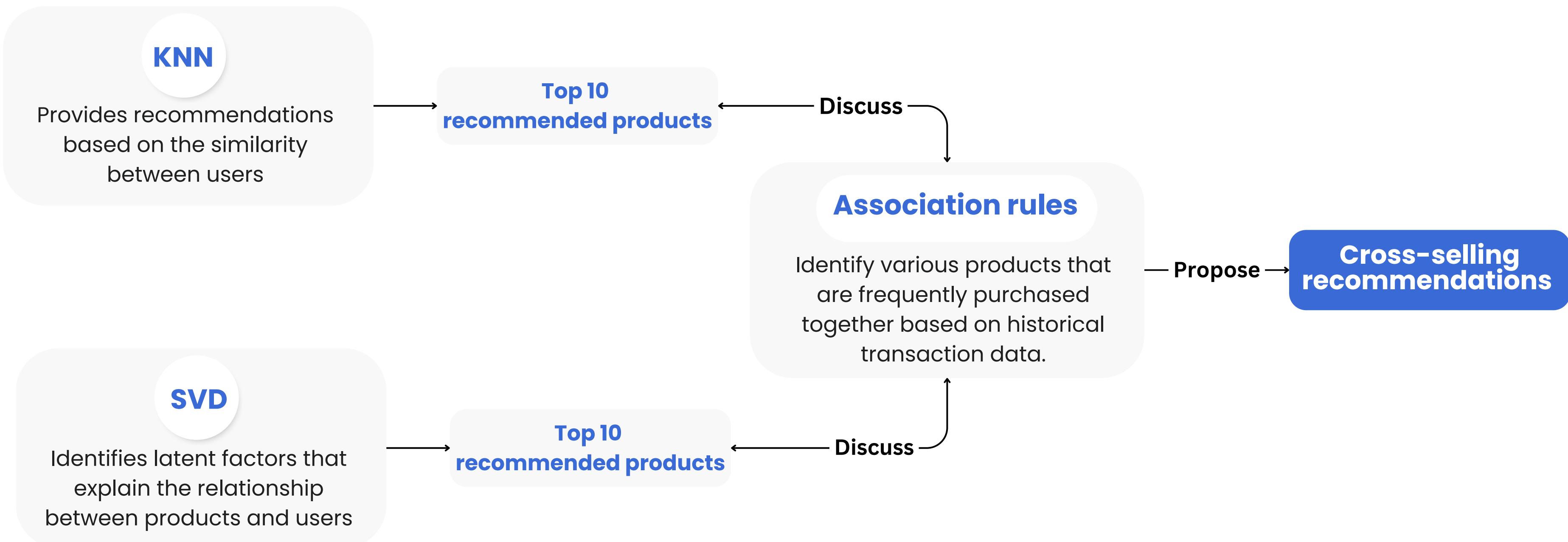
	Apriori	FP-Growth
Number of subsets generated	446	446
Execution time	37.9s	846.2s
Mean Support	0.009093	0.009093
Mean Lift	0.555446	0.555446
Mean Confidence	3.250797	3.250797

Table. Apriori and FP-Growth performance

	KNN	SVD
RMSE	0.639	0.416
MSE	0.40	0.17

Table. KNN and SVD validation

Results Discussion



Integrating KNN and SVD with association rules enhances cross-selling **by recommending both popular items and less common products that match individual preferences.**

Results Discussion

Recommended products for household 1366 using KNN

SUB_COMMODITY_DESC	RECOMMENDATION_SCORE
SOFT DRINKS 12/18&15PK CAN CAR	40.307592
GARDEN PLUS	26.738700
FLUID MILK WHITE ONLY	20.752423
POTATO CHIPS	8.380786
TUNA	7.981701
TUNA	7.981701
MUSHROOMS CND & GLASS	7.981701
MEAT: LUNCHMEAT BULK	7.183531
MUSHROOMS CND & GLASS	7.183531
LEAN	7.183531

Recommended products for household 1366 using SVD

SUB_COMMODITY_DESC	PREDICTED_QUANTITY
VALUE GLASS WINE	23.939786
SPECIALTY CRACKERS	20.766364
DRY DOG FOOD PREMIUM (ALPO/PUR)	15.457166
BEERALEMALT LIQUORS	8.780089
SECURITY & SAFETY	8.637327
LUNCH COMBO	8.253565
SOFT DRINKS 20PK&24PK CAN CARB	6.707387
VIENNA SAUSAGE	6.396062
VITAMIN - INDIVIDUAL	6.164891
CRACKERS	5.776277

The results differ, except for 'SOFT DRINKS 12/18&15PK CAN CAR'. This is due to **KNN recommending based on user similarity**, while **SVD uncovers hidden patterns and latent features**.

Association algorithms like Apriori and FP-Growth find frequently bought pairs, but **they don't account for individual preferences**

By combining KNN, SVD, and association rules, we can create a **personalized cross-selling strategy**, recommending products tailored to both **general patterns and individual preferences**.

Sales Overview Dashboard

8.06M

Total Sales

276K

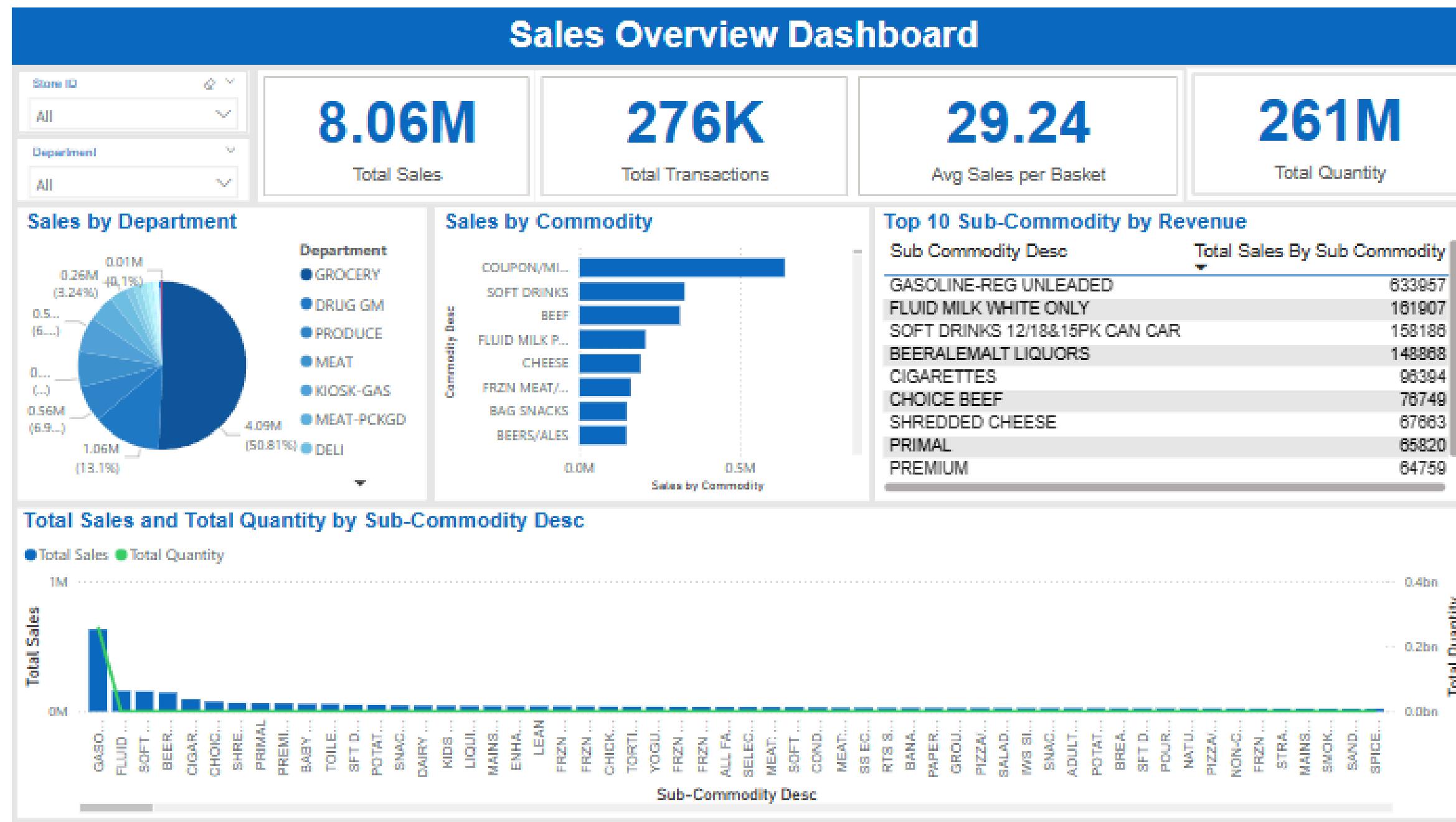
Transaction

29.24

Avg Sales/Basket

261M

Total Quantity



Sales by Department

50.8%

GROCERY sales in the Total Sales

MEAT and **DELI** categories follow after with **13.1%** and **6%** respectively

Sales by Commodity

'COUPON/MISC' dominates the Sales followed by 'SOFT DRINKS', 'BEEF', and 'FLUID MILK' => Staple items and daily essentials drive the bulk of sales.

'FRZ MEAT', 'CHEESE', and 'BEER/ALES' show comparatively lower sales => Reassessment in terms of stock and sales strategy focus.

Sales Overview Dashboard

8.06M

Total Sales

276K

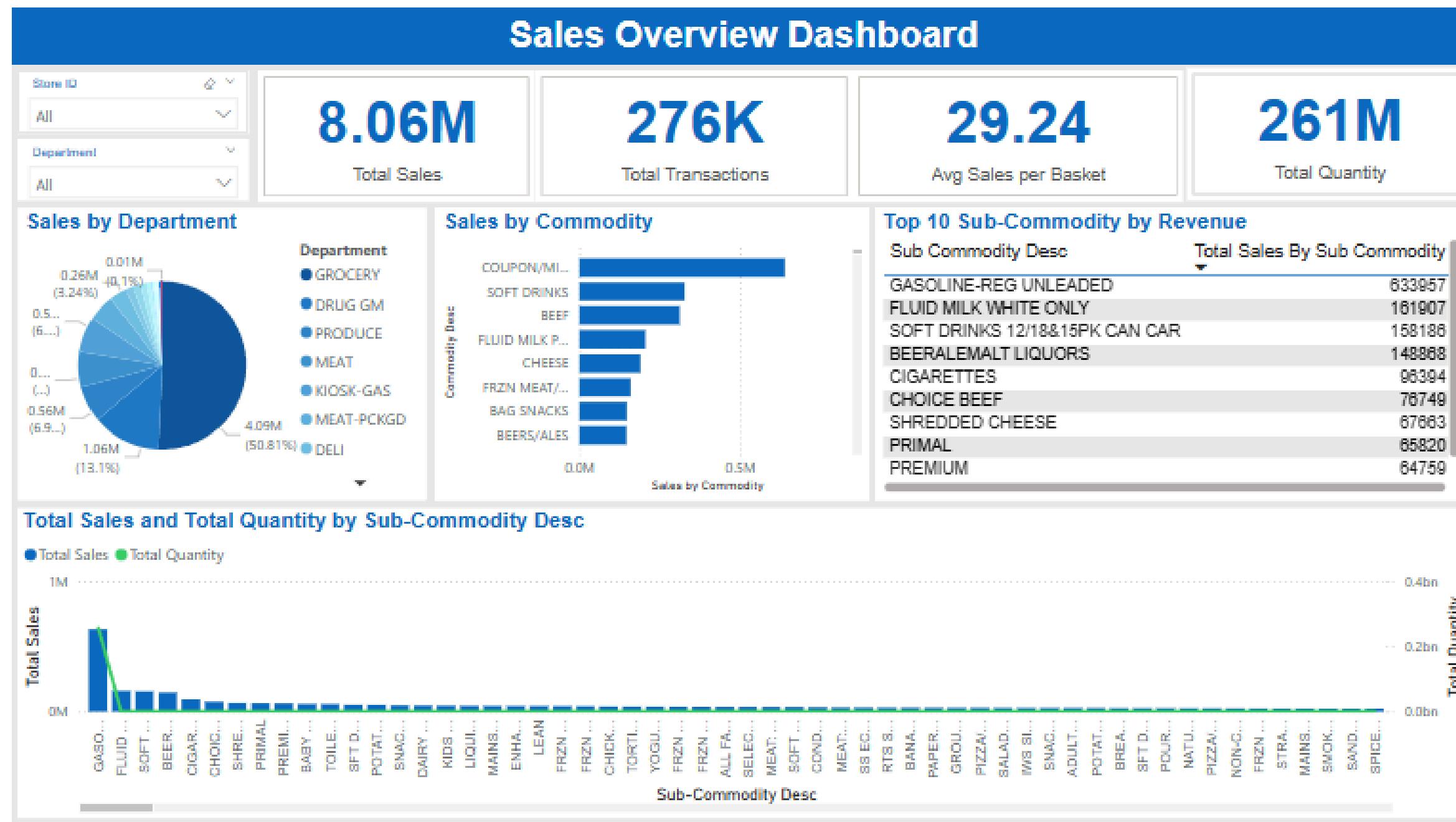
Transaction

29.24

Avg Sales/Basket

261M

Total Quantity



Top 10 Sub-Commodities by Revenue

'GASOLINE-REG UNLEADED' leads significantly with the highest revenue. Essential consumed items such as 'FLUID MILK WHITE ONLY', 'SOFT DRINKS 12/18&15PK CAN CAR', and 'BEER/ALE MALT LIQUORS' also contribute significantly.

Total Sales vs. Total Quantity

A significant spike in sales for 'GASOLINE-REG UNLEADED' that despite lower quantities compared to other products, dominates in revenue

'FLUID WHITE MILK ONLY' also stands out as having both high revenue and high volume, => Essential role in daily consumer purchases.

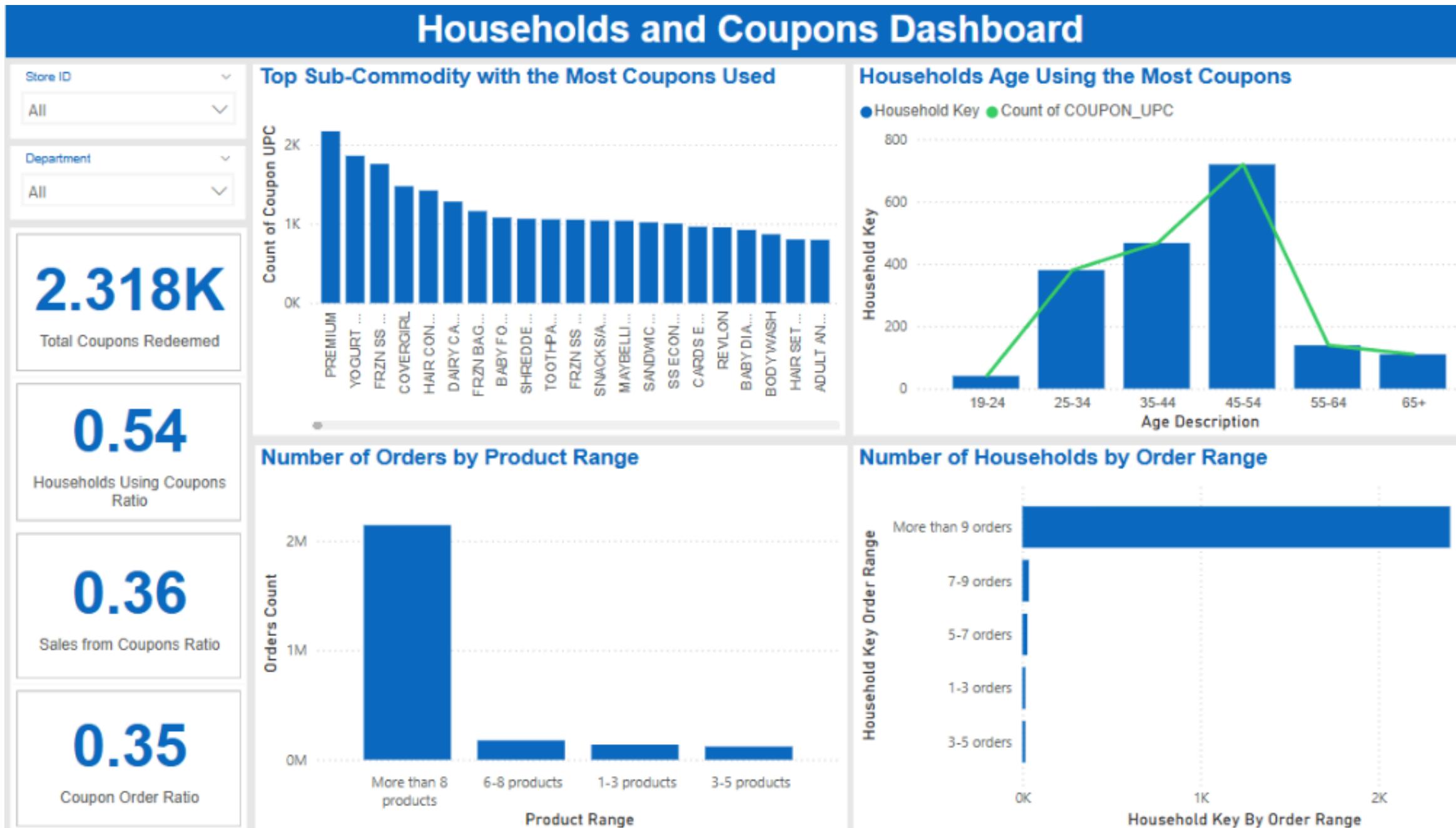
Households and Coupons Dashboard

2318K Total Coupon Redeemed

54% Household Using Coupons

36% Sales from Coupons

35% Coupon Order



Top Sub-Commodities used Coupons

Dairy products, frozen foods, and personal care items are the top sub-commodities where most coupons are used. However the sales of all coupons play **only 36% of total sales**, barely reaching a half.

Meanwhile, approximately 54% of households using coupons show a moderate level of participation in the promotions offered.

=> The problem here might be that **the variety of coupons (2.318K coupons) did not show effectiveness**. Instead, the company can focus on products popular to customers such as: gasoline, fluid milk, etc.

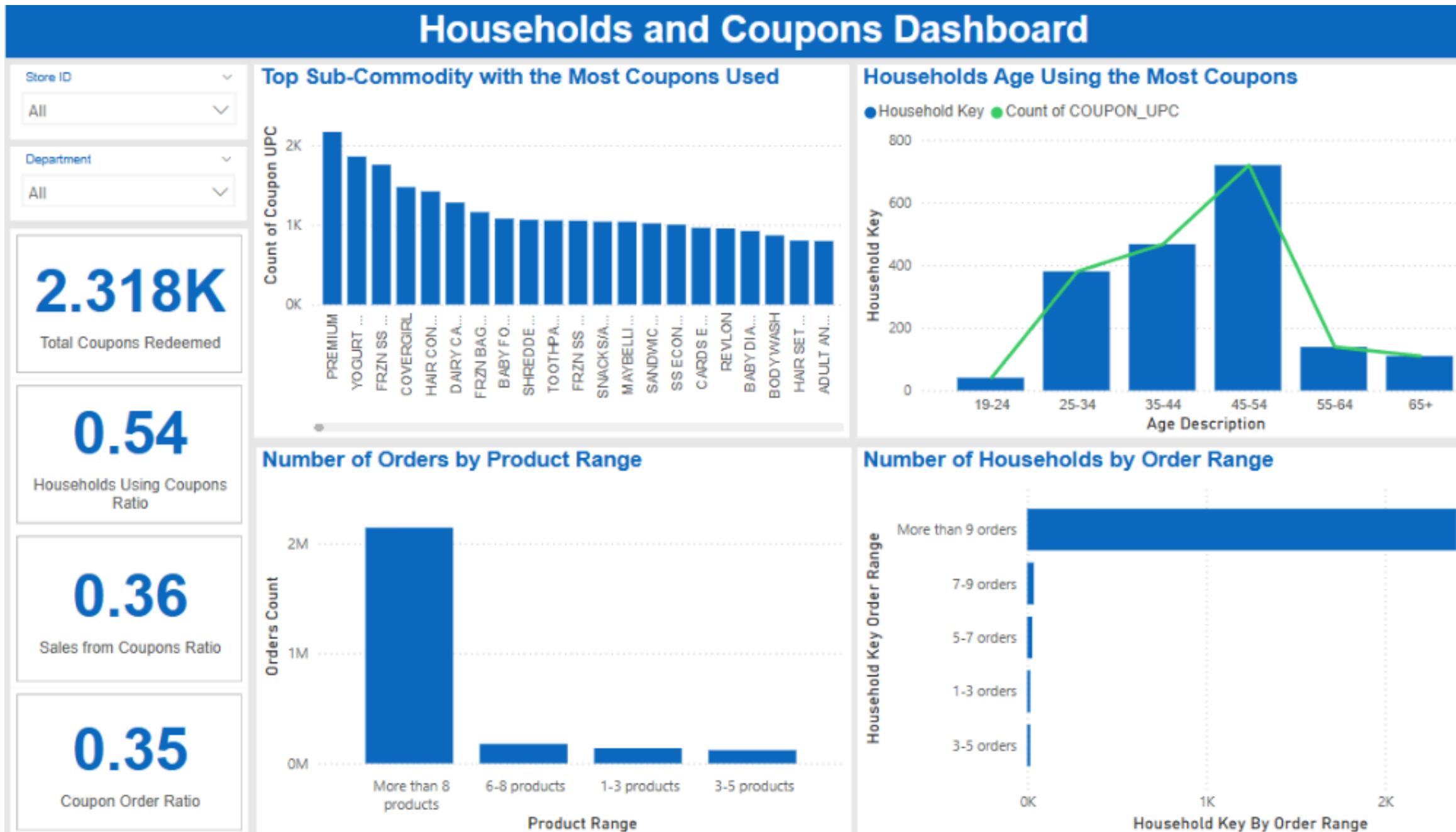
Households and Coupons Dashboard

2318K Total Coupon Redeemed

54% Household Using Coupons

36% Sales from Coupons

35% Coupon Order



Households Age Using the most Coupons

The most active coupon users fall within the 35-54 age range, **peaking age 45-54** => Represents working adults who are managing household budgets, making them **ideal targets for tailored marketing**.

Number of Orders by Product

The majority of customers buy **more than 8 products per order**, while the number of orders with **only 1-3 products is very low**.

=> Customers tend to **buy many products at once**, creating great opportunities for cross-selling strategies and promotional combos.

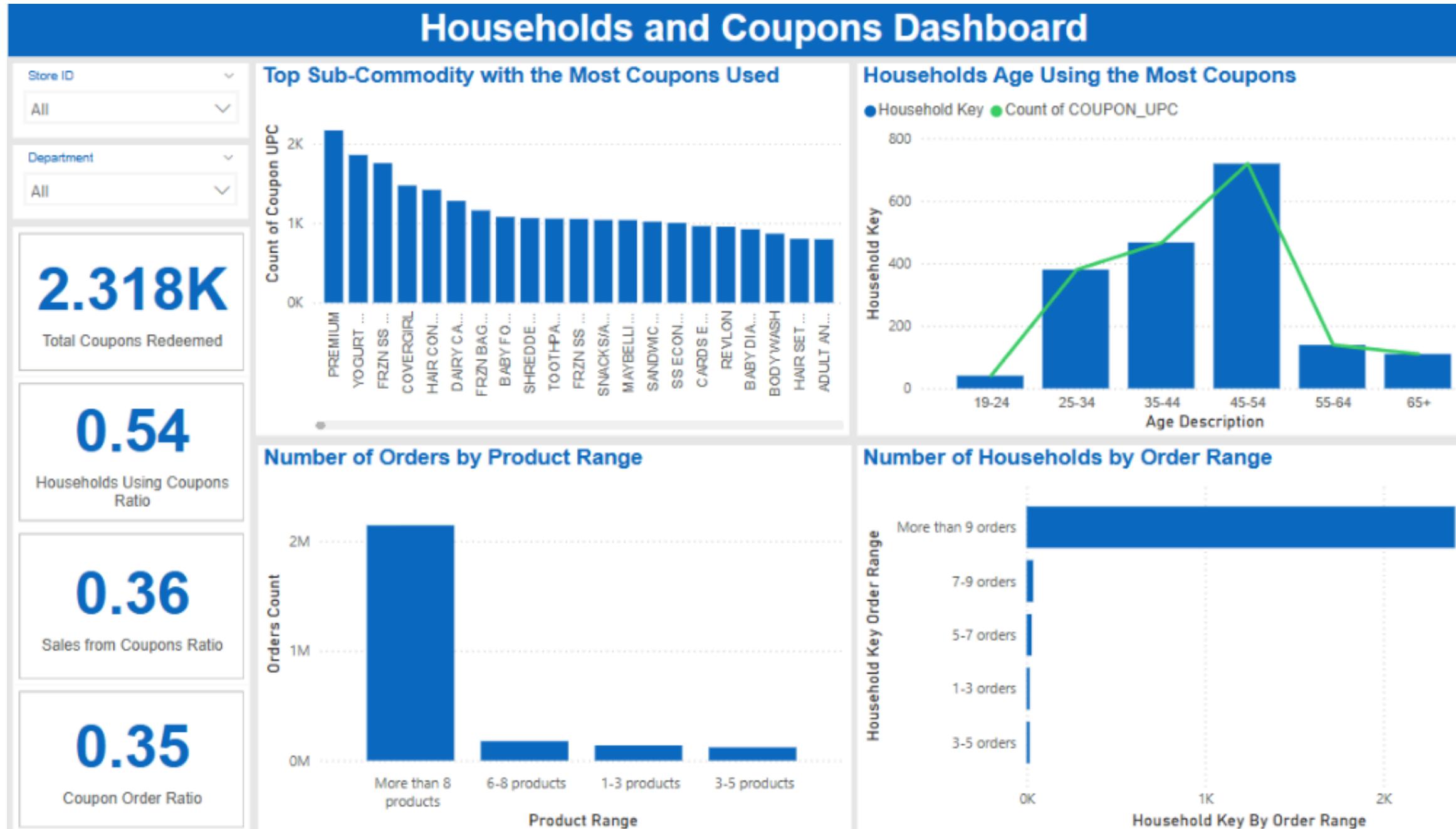
Households and Coupons Dashboard

2318K Total Coupon Redeemed

54% Household Using Coupons

36% Sales from Coupons

35% Coupon Order



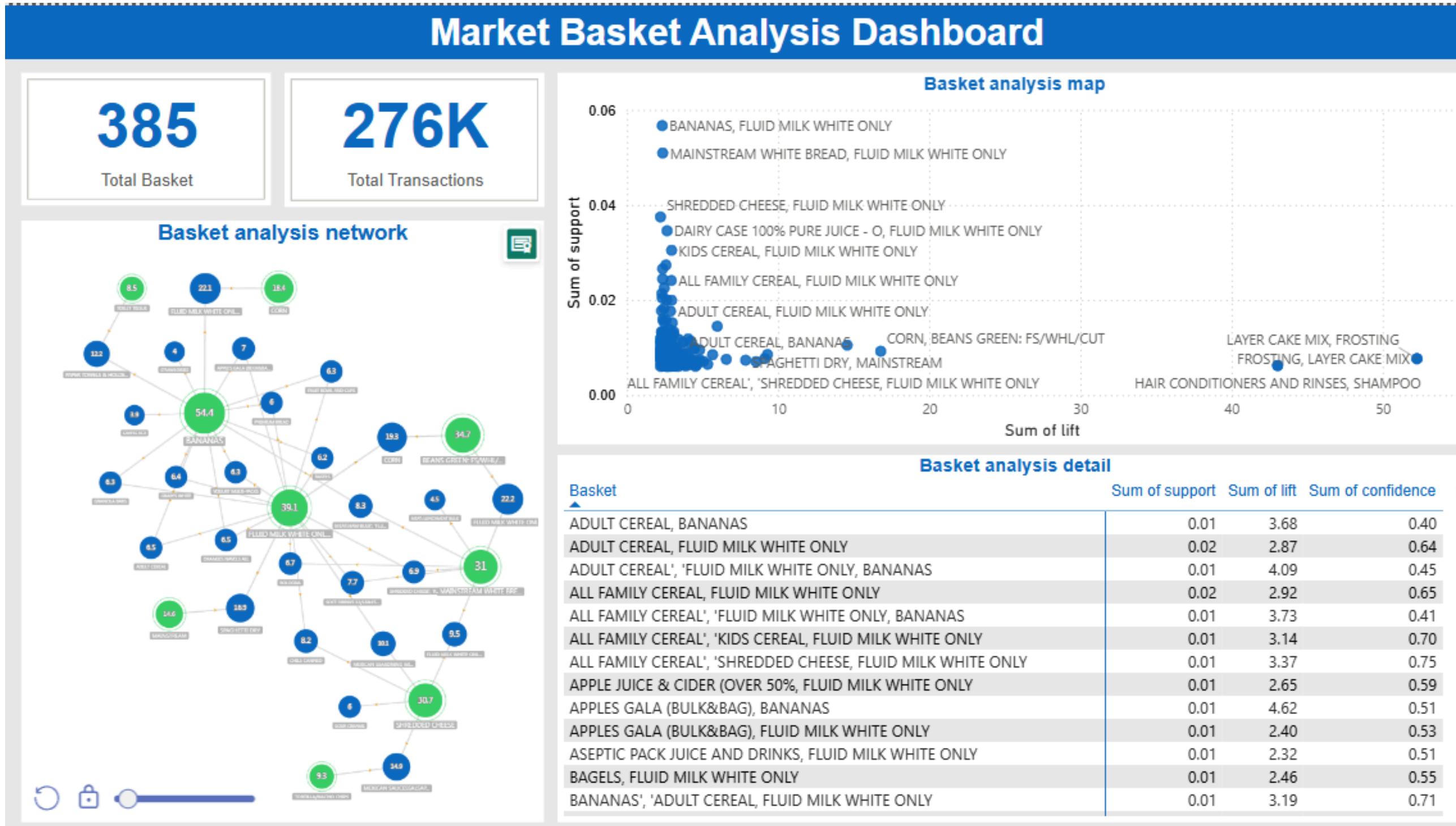
Number of Households by Order

The majority of customers have more than 9 orders, demonstrating **a high rate of loyal customers**.

Businesses can implement **loyalty programs** by publishing more coupons relating to their often bought products.

They can also **promote remarketing (retargeting)** the previously sold products to **customers** who buy less to increase transaction frequency.

Market Basket Analysis Dashboard

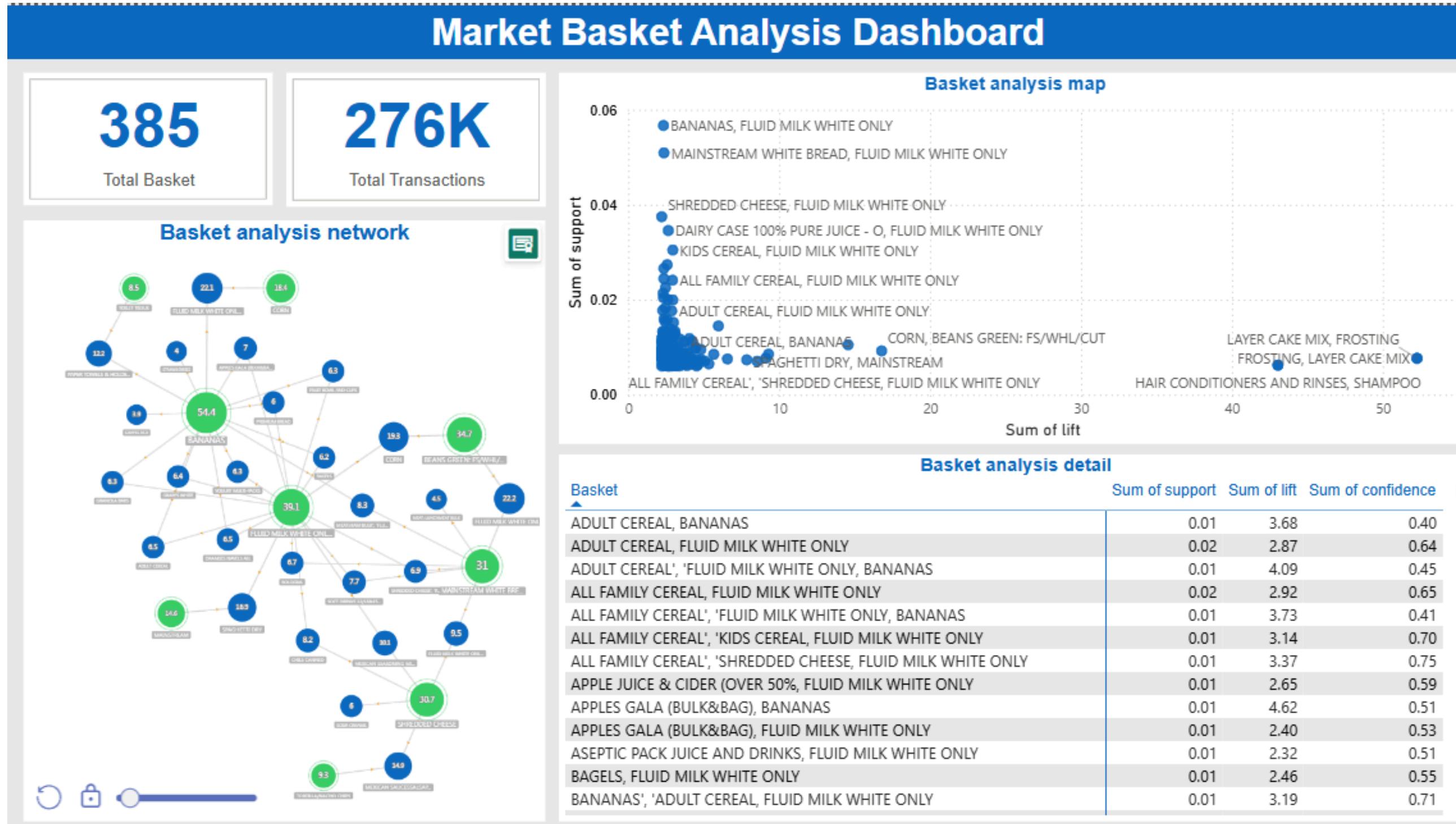


Basket Analysis

The pairing of **bananas with fluid milk** shows a high sum of support and moderate lift meaning that it is a common and reasonably associated purchase

More unique combinations, such as adult cereal with bananas and green beans, **display lower support but higher lift**, pointing to **niche but strong complementary** purchasing patterns possibly linked to specific customer segments or promotional activities.

Market Basket Analysis Dashboard

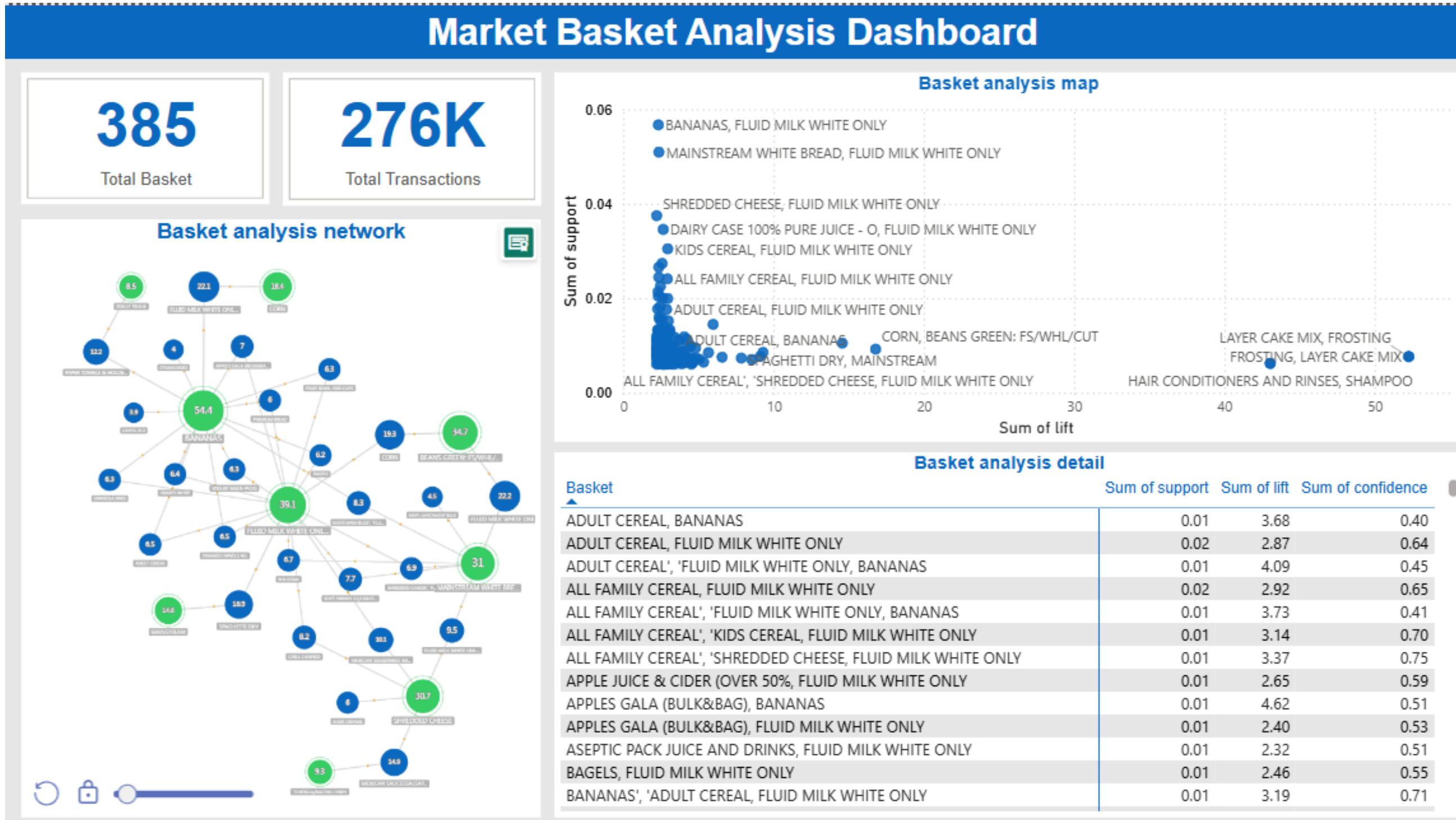


Basket Analysis

Combinations like layer cake mix with frosting and hair conditioners with shampoo, **while less frequent in support, exhibit the highest lift values**

=> This exceptionally strong association indicates that the purchase of one significantly triggers the purchase of the other, suggesting that cross-merchandising or joint promotions could be particularly effective.

Market Basket Analysis Dashboard



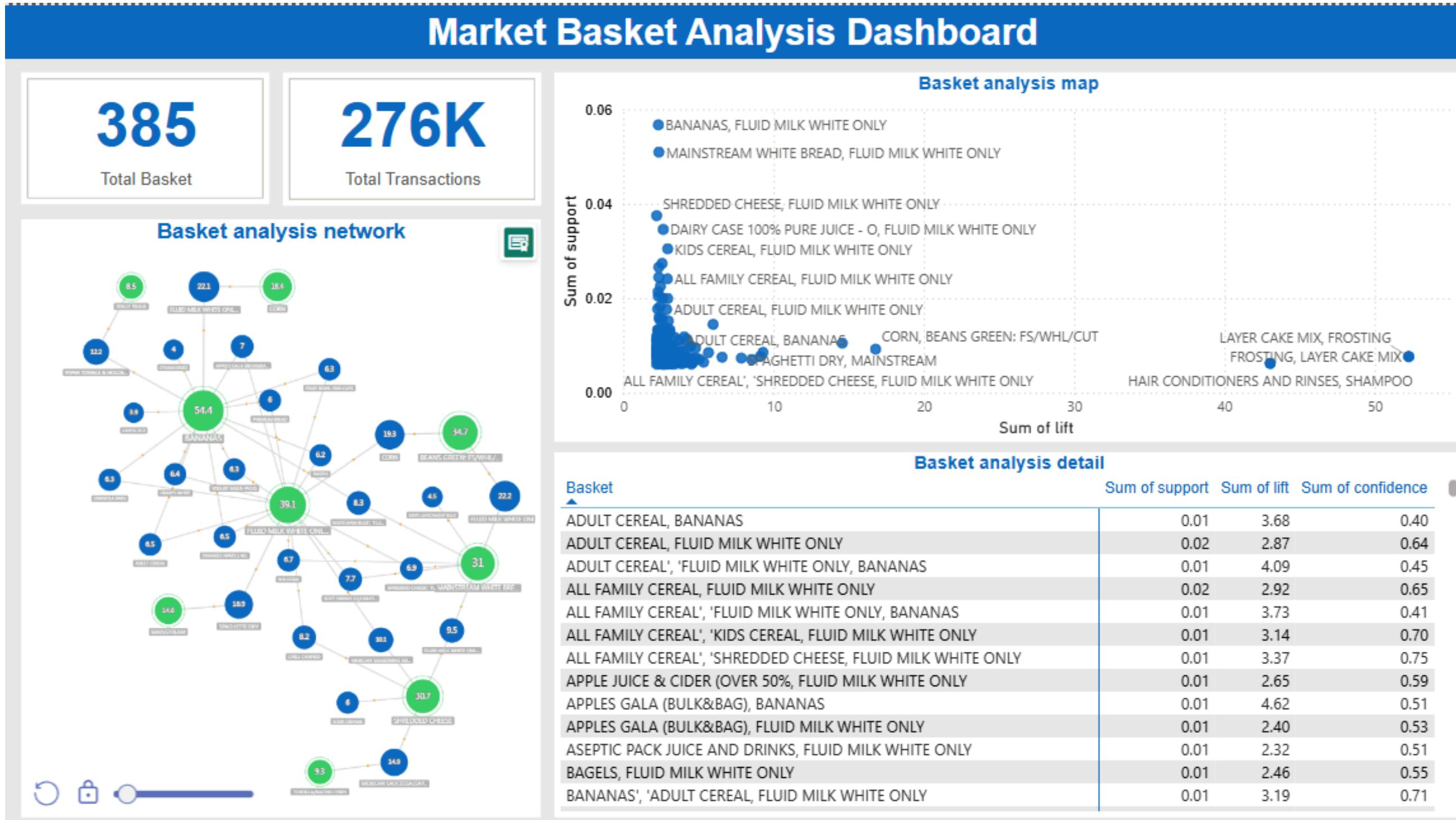
Basket Analysis

Targeted promotions can leverage these primary - secondary relationships.

For instance, offering a discount on bread and bananas when purchased with milk could increase overall basket size and encourage customers to buy more than they initially planned.

Sending personalized emails or app notifications suggesting secondary items based on previous primary item purchases.

Market Basket Analysis Dashboard



Basket Analysis

Implementing dynamic pricing strategies that adjust prices in real-time based on the combination of primary and secondary items.

Creating bundle offers for combinations with strong links, especially during checkout, can also be promoted effectively both in-store and online.

Online platforms can employ algorithms to recommend high-lift items together when one is added to a shopping cart.

Conclusion & Future work



Completed Achievements

- Addressed and solved key business questions
- Identified product associations using Apriori and FP-Growth
- Improved recommendations with KNN and Truncated SVD.



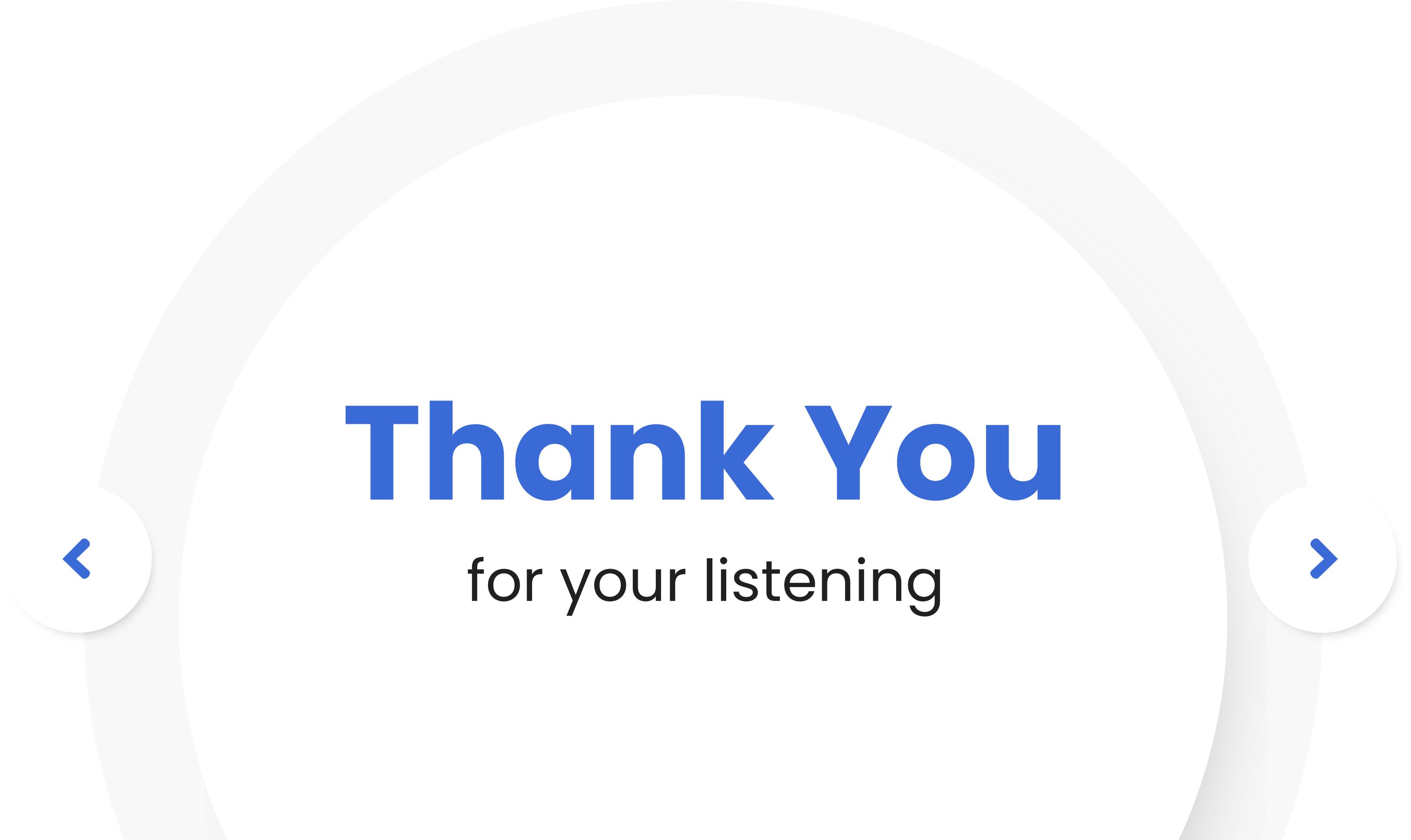
Research Gaps

- Lack of time-based data
- Data noise and missing values that complicate preprocessing.
- Limited complexity in current recommendation models, needing further refinement.
- Absence of external data for more comprehensive recommendations.



Future Works

- Integrating time-based data and hybrid recommendation systems.
- Incorporating sentiment analysis and real-time data for dynamic recommendations.
- Exploring customer retention models for better long-term cross-selling strategies.



Thank You

for your listening