CUSTOMER SEGMENTATION

Presented by: Ahmed Ayman Alhofy

BUSINESS PROBLEM

Wholesale customers have diverse purchasing behaviors → hard to target marketing or optimize logistics.

OBJECTIVE

Objective: Group customers into meaningful segments using unsupervised learning to:

1. Improve marketing personalization
2. Optimize inventory & delivery
3. Identify high-value or niche customer groups
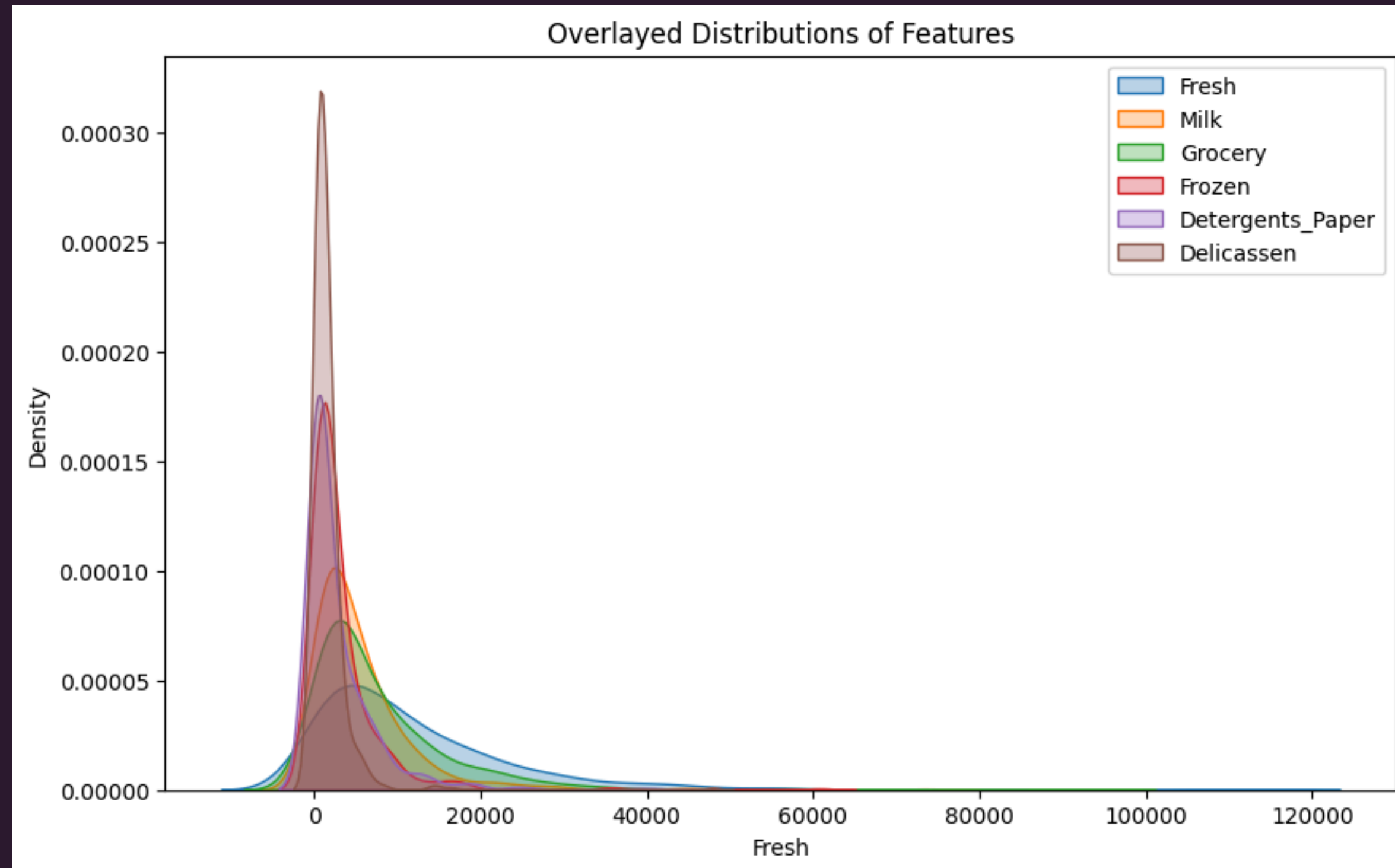
DATASET OVERVIEW

Features:
- Channel (1=Horeca, 2=Retail)
- Region (1-3)
- 6 product categories: Fresh, Milk, Grocery, Frozen, Detergents_Paper, Delicassen

Initial Stats:
- All numeric, no missing values
- Highly right-skewed distributions → required log-transform
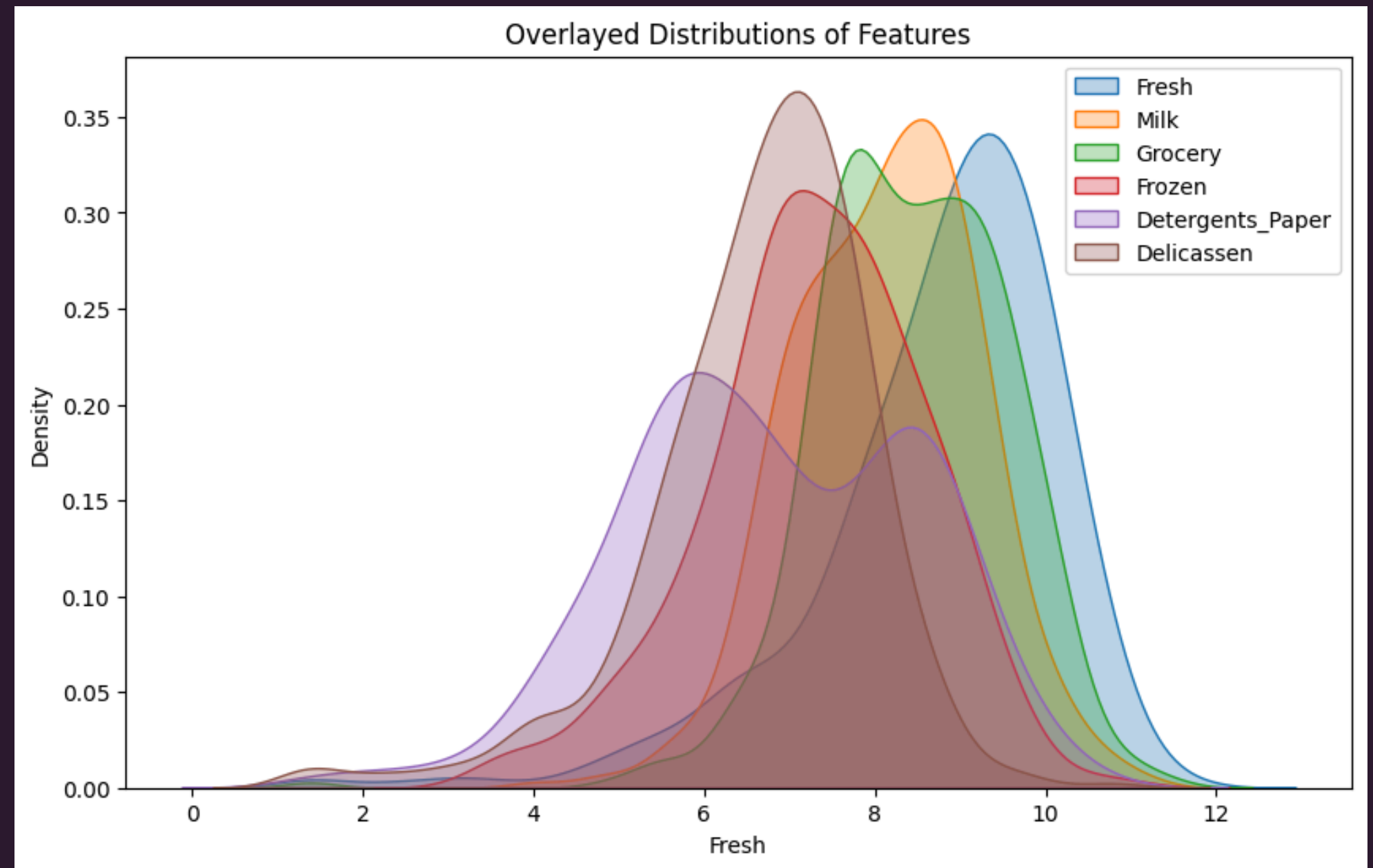
EXPLORATORY DATA ANALYSIS (EDA)

All spending features are heavily right-skewed → large spenders distort averages.
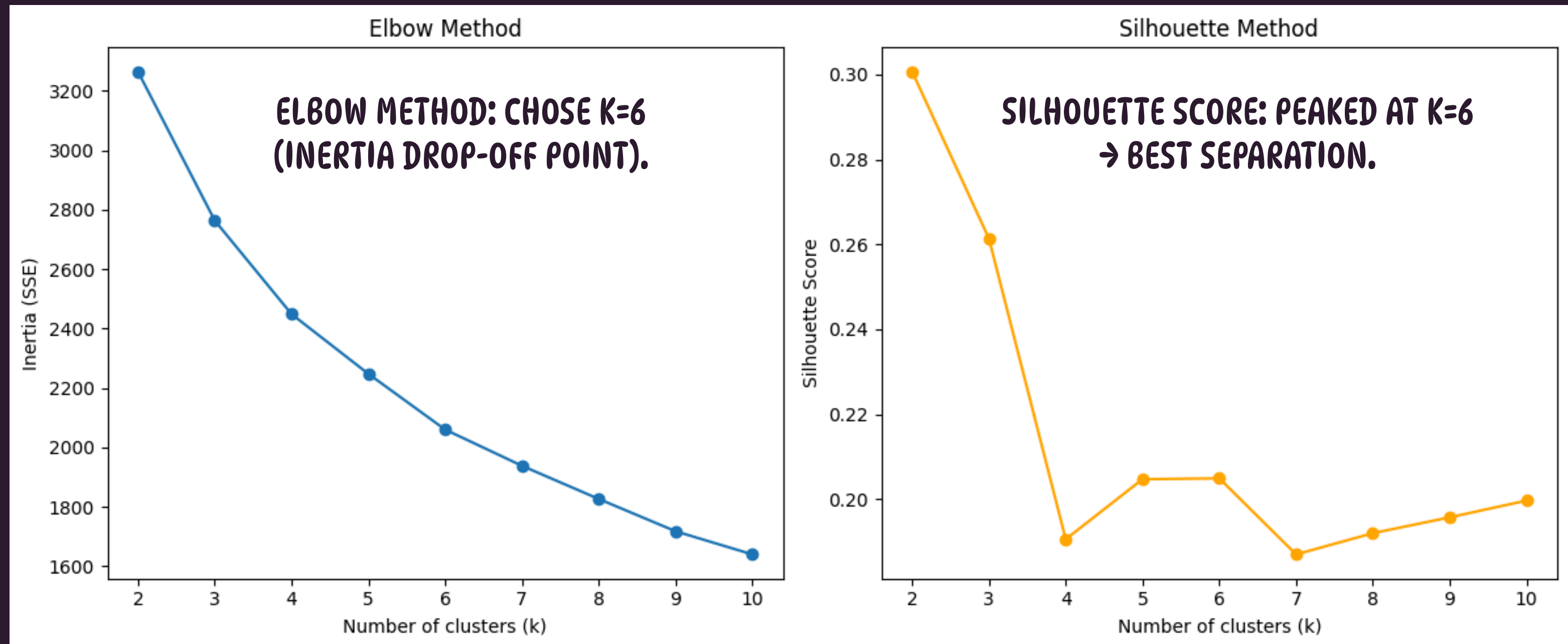
EXPLORATORY DATA ANALYSIS (EDA)

Action Taken: Applied np.log1p to normalize distributions.

LOG-TRANSFORM MADE FEATURES MORE GAUSSIAN-LIKE → BETTER FOR DISTANCE-BASED CLUSTERING.

Overlayed Distributions of Features

MODEL SELECTION - K-MEANS
(ELBOW & SILHOUETTE)

ELBOW METHOD: CHOSE K=6 (INERTIA DROP-OFF POINT).

SILHOUETTE SCORE: PEAKED AT K=6 → BEST SEPARATION.

FINAL CHOICE: K-MEANS WITH 6 CLUSTERS FOR INTERPRETABILITY & PERFORMANCE.
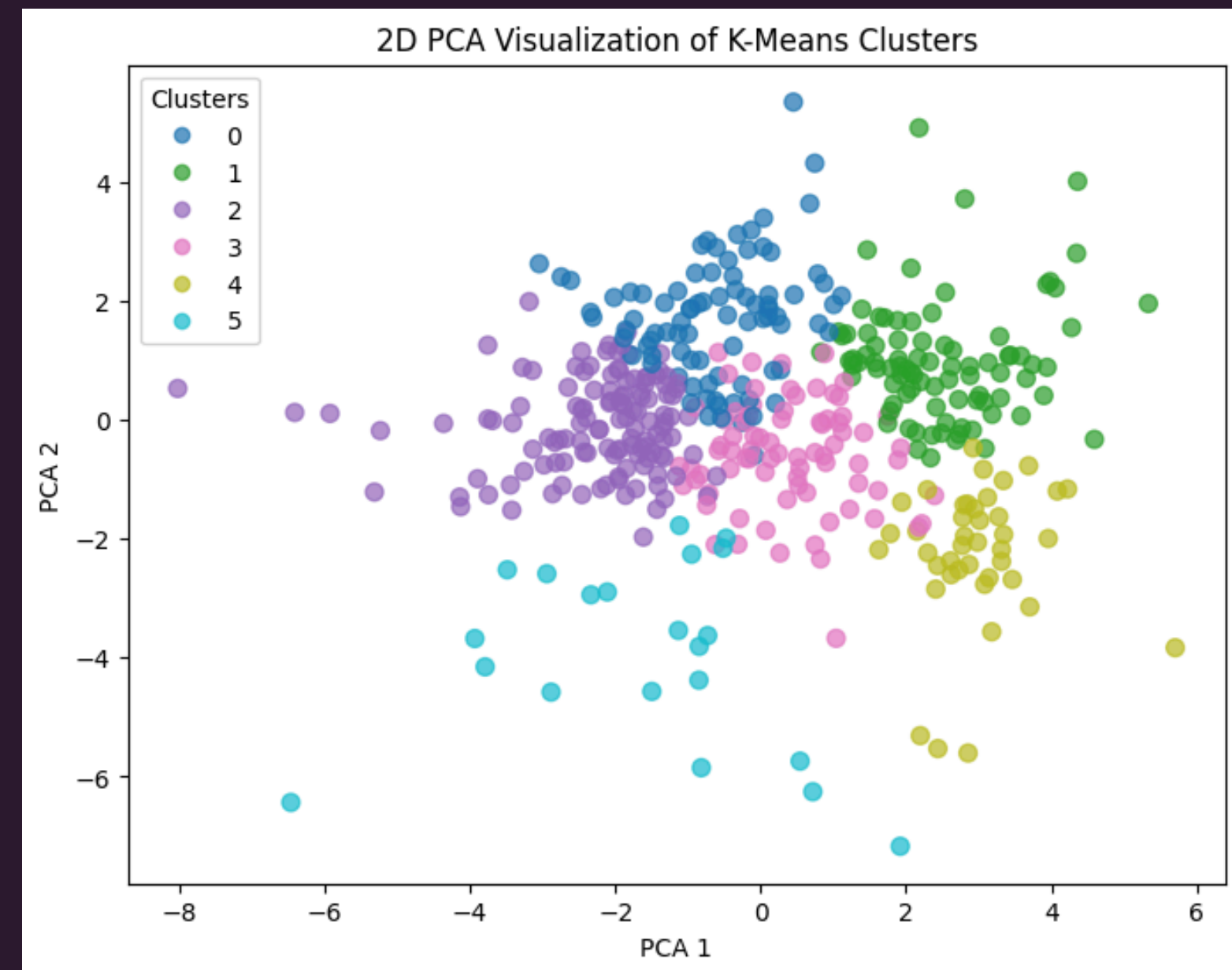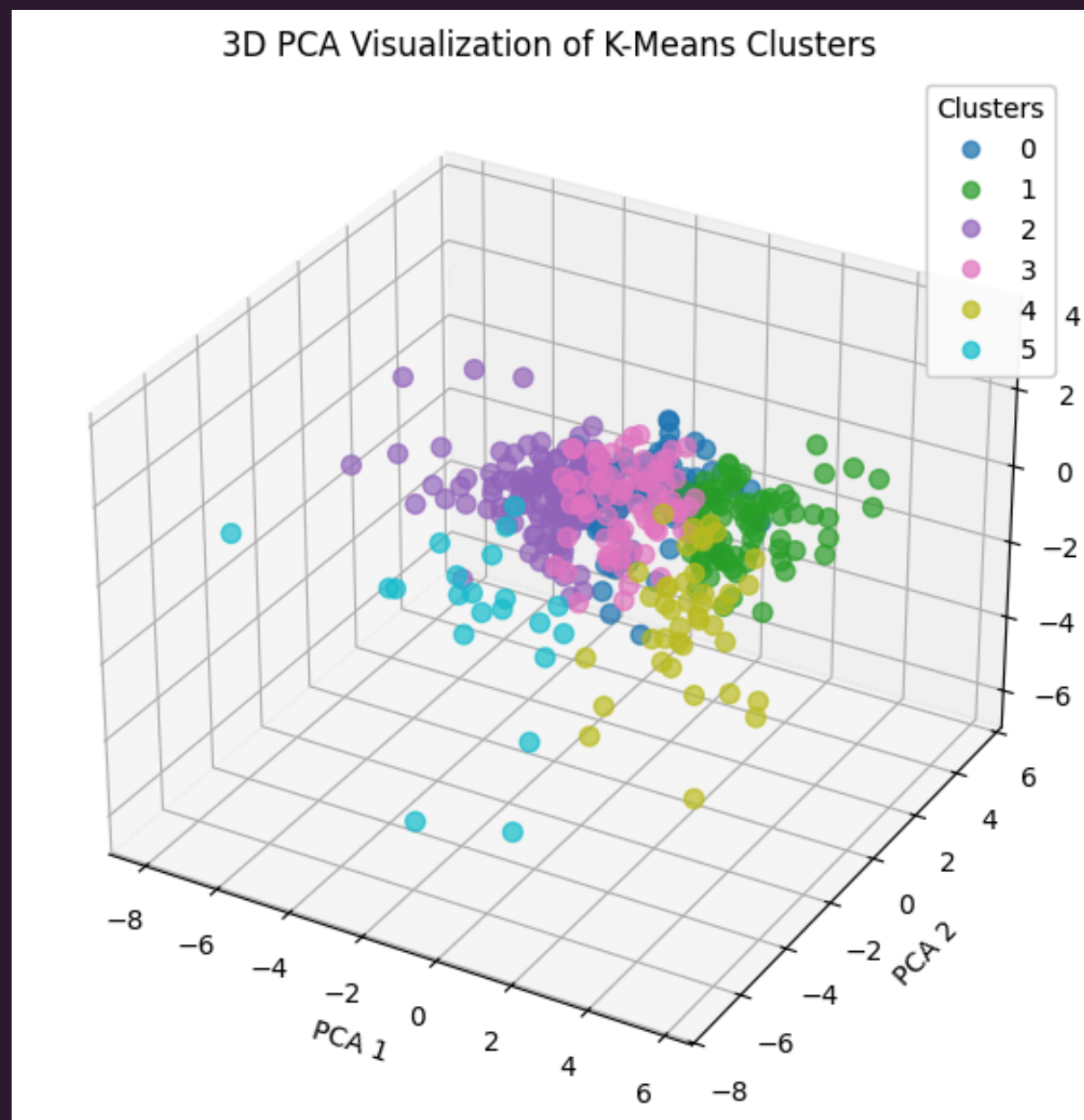
MODEL SELECTION - K-MEANS

We chose K-Means over other algorithms because it provided the best balance of interpretability, performance, and cluster separation - with a clear elbow point at k=6 and the highest Silhouette Score among tested values, making it ideal for deriving actionable customer segments.
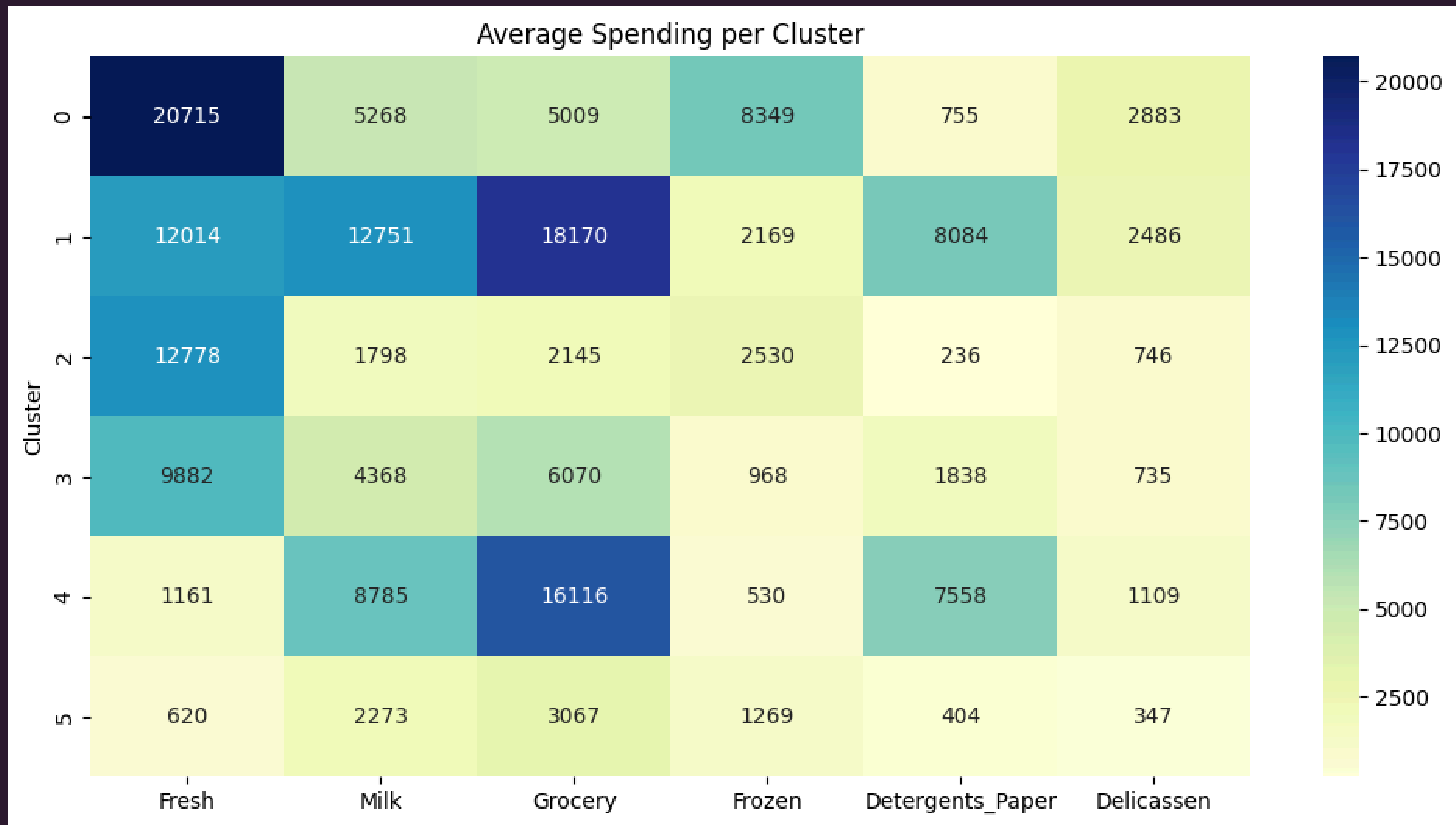
VISUALIZATION - PCA PLOTS

2D & 3D PCA TO VISUALIZE CLUSTER SEPARATION.
INSIGHT: CLUSTERS ARE REASONABLY SEPARATED IN REDUCED SPACE → VALIDATES CLUSTERING QUALITY.

AVERAGE SPENDING PER CLUSTER

Average Spending per Cluster

Cluster	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
0	20715	5268	5009	8349	755	2883
1	12014	12751	18170	2169	8084	2486
2	12778	1798	2145	2530	236	746
3	9882	4368	6070	968	1838	735
4	1161	8785	16116	530	7558	1109
5	620	2273	3067	1269	404	347

CUSTOMER SEGMENTS

1. FRESH & FROZEN – MEAL PREPPERS / SMALL BUSINESSES
2. BULK BUYERS – RETAILERS, RESTAURANTS, LARGE FAMILIES
3. FRESH-FIRST – PRODUCE-FOCUSED BUYERS
4. HOUSEHOLD ESSENTIALS – CLEANING-FOCUSED HOUSEHOLDS
5. GROCERY & MILK – DAIRY & PACKAGED GOODS SHOPPERS
6. LOW-VOLUME – INFREQUENT / SINGLE SHOPPERS

BUSINESS RECOMMENDATIONS

- Marketing: Tailor email campaigns & discounts per segment.

- Inventory: Stock more fresh for Clusters 0 & 2; more detergents for Cluster 3.

- Customer Support: Prioritize Cluster 1 (high spenders).

GITHUB LINK

github.com/AhmedAyman4/konectainternship/tree/main/Task-4

THANK
YOU