DLG Ecommerce Performance Analysis

Celine Ji
12/03/2024

TABLE OF CONTENTS - Deciphering Performance and Profitability

WHOA! Performance Spotlights

$ 20,280,463 Total Revenue

Over **3% CVR** for Electrics and Kids FMCG

01. A Holistic View of Site Performance - *October Data*

WHATS MATTER :

- AUV ($448.42) > ACV ($369.04), suggesting that users make more than one purchase on average, which could signal effective cross-selling strategies.
- With a strong AUV and ACV, even small improvements in the CVR (2.18%) could lead to substantial revenue growth.

STRATEGIC FOCUS

- Analyze **the impact of targeted campaigns** , such as holiday sales. Comparing CVR vs. non-campaign periods can pinpoint what strategies resonate with customers.
- Evaluate AUV against **user acquisition costs.** For example, if CAC exceeds AUV, it indicates inefficiencies in marketing efforts.
- Further **segment high-value users** , to invest in initiatives that encourage repeat purchases to amplify AUV, and bundling to incentive high ACV.

2.18%

Site Conversion Rate

CVR	Used Metric	Session Count
Measure the site's capability in monetizing visit session	Total Views	2,519,908
	Total Purchase	54,955

$ 448.42

Average User Value

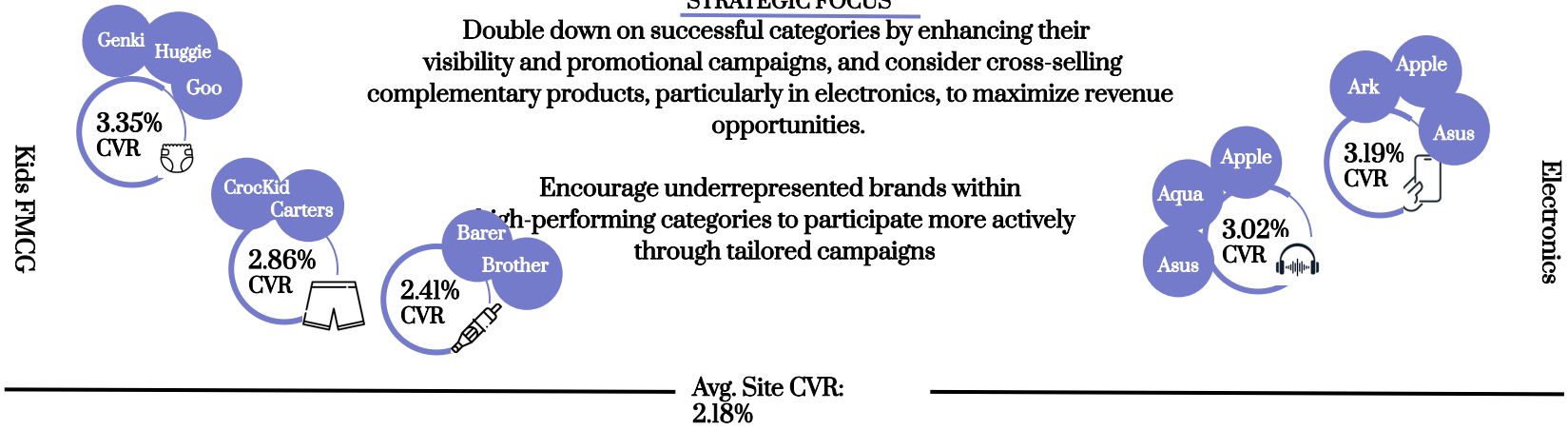
AUV	Used Metric	/
Understanding how much revenue each purchasing user contributes	Total Revenue	$ 20,280,463
	Total Unique Users	45,226

$ 369.0

Average Conversion Value

ACV	Used Metric	/
Reflect Avg. revenue per transaction	Total Revenue	$ 20,280,463
	Total Purchase	54,955

02. Conversion Rate By Category - *Winners and Opportunities*



STRATEGIC FOCUS

Double down on successful categories by enhancing their visibility and promotional campaigns, and consider cross-selling complementary products, particularly in electronics, to maximize revenue opportunities.

Encourage underrepresented brands within high-performing categories to participate more actively through tailored campaigns

WHAT

These could represent niche products, or there may be challenges with inadequate product positioning, pricing, or visibility.

STRATEGIC FOCUS

Analyze their relevance to the audience and evaluate whether their products need better promotion, competitive pricing, or improved positioning to increase engagement.

03. Top Category Pairs with High Affinity - *view - cart- purchase*

Methodology

Pairwise Conditional Probability

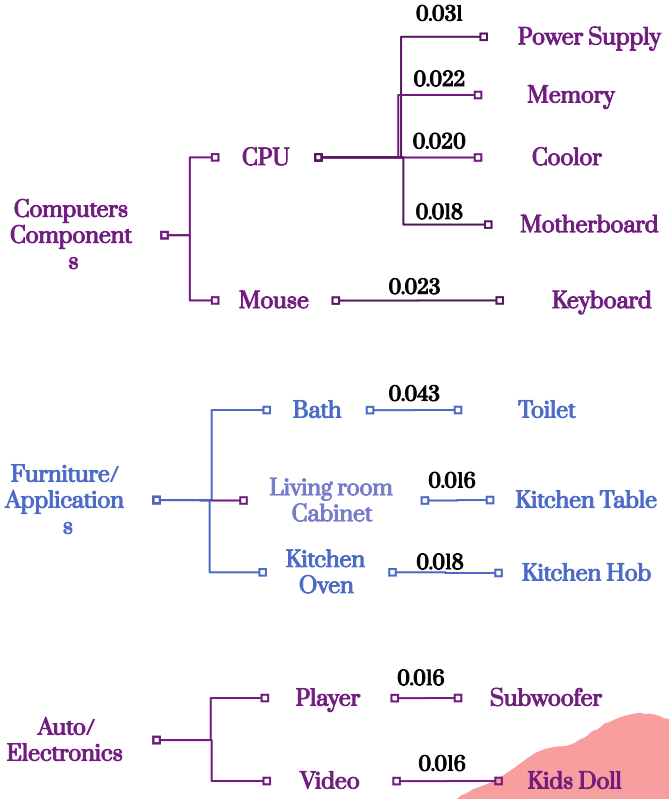
$$P(A \mid B) = P(A \cap B) / P(B)$$

$$P(B \mid A) = P(A \cap B) / P(A)$$

$$\text{Affinity Score} = 2 * P(A|B) * P(B|A) / (P(A|B) + P(B|A))$$

Weighted Affinity Score Across Events

- Purchase- Conversion
- Cart- Purchase Intent
- View- Exploratory Stage



04. Underperforming Categories and Brands

How Discontinue Product is Defined

Products with Below Avg. Purchase-to-View CVR, or Contributes Little to Revenue

Products with Below 30% Quantile of Total Interaction Counts (view, cart, purchase)

Next Steps

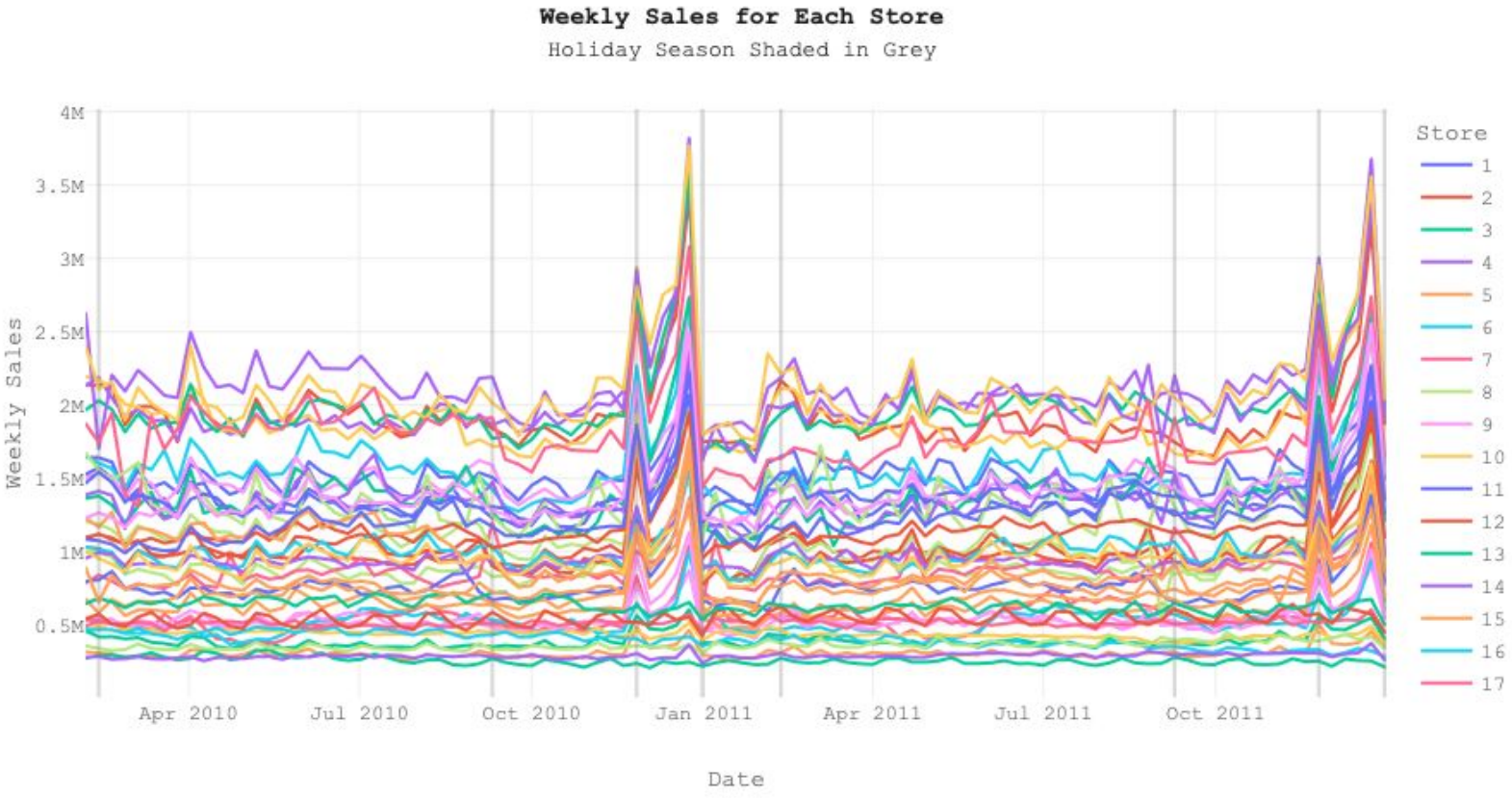
- Conduct a deeper evaluation incorporating cost of goods, returns, and warehousing expenses.
- Focus on cross-selling opportunities for low-performing but strategically valuable products. Bundle with high-performing and trending products to increase overall value.

More than 99.9% of products are suggested for discontinuation

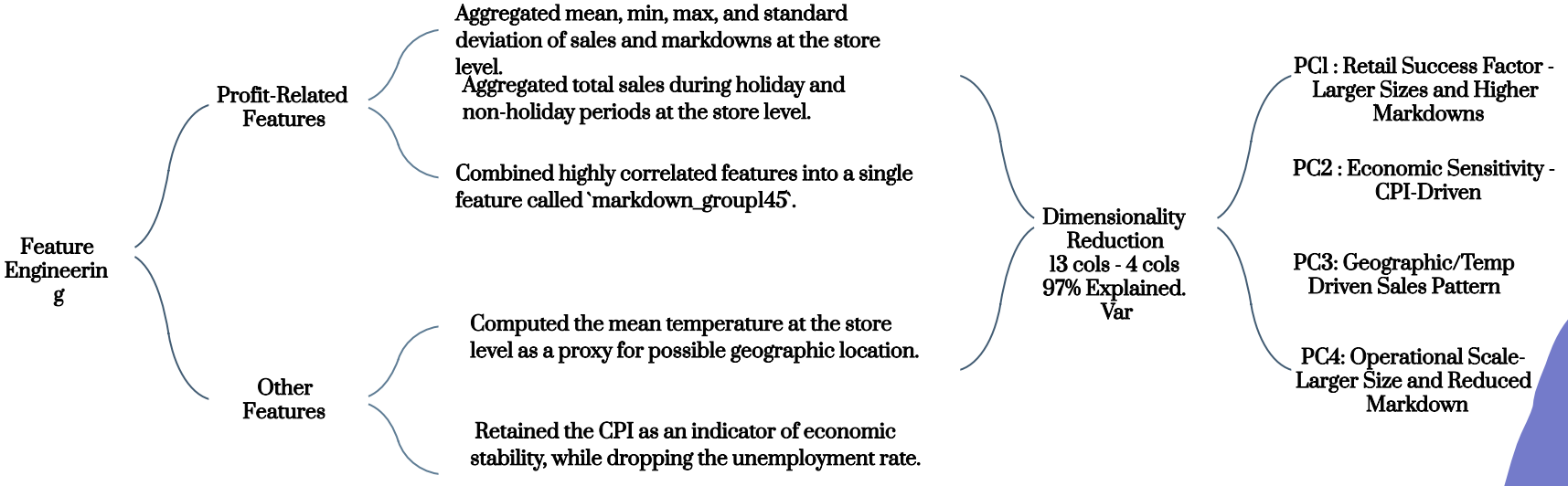
category_code	discontinued_count	discontinuation_ratio
apparel.belt	12.0	0.857143
apparel.shorts	16.0	0.800000
sport.tennis	70.0	0.752688
stationery.cartrige	166.0	0.751131
apparel.skirt	19.0	0.730769
construction.components.faucet	461.0	0.722571
sport.ski	76.0	0.716981
apparel.underwear	638.0	0.663202
apparel.shoes.sandals	231.0	0.658120
apparel.jeans	38.0	0.644068

brand	discontinued_count	discontinuation_ratio
omoikiri	23.0	1.0
adamas	9.0	1.0
dab	9.0	1.0
kress	9.0	1.0
egg	8.0	1.0
nanga	8.0	1.0
maak	6.0	1.0
baboo	5.0	1.0
barer	5.0	1.0
lange	5.0	1.0

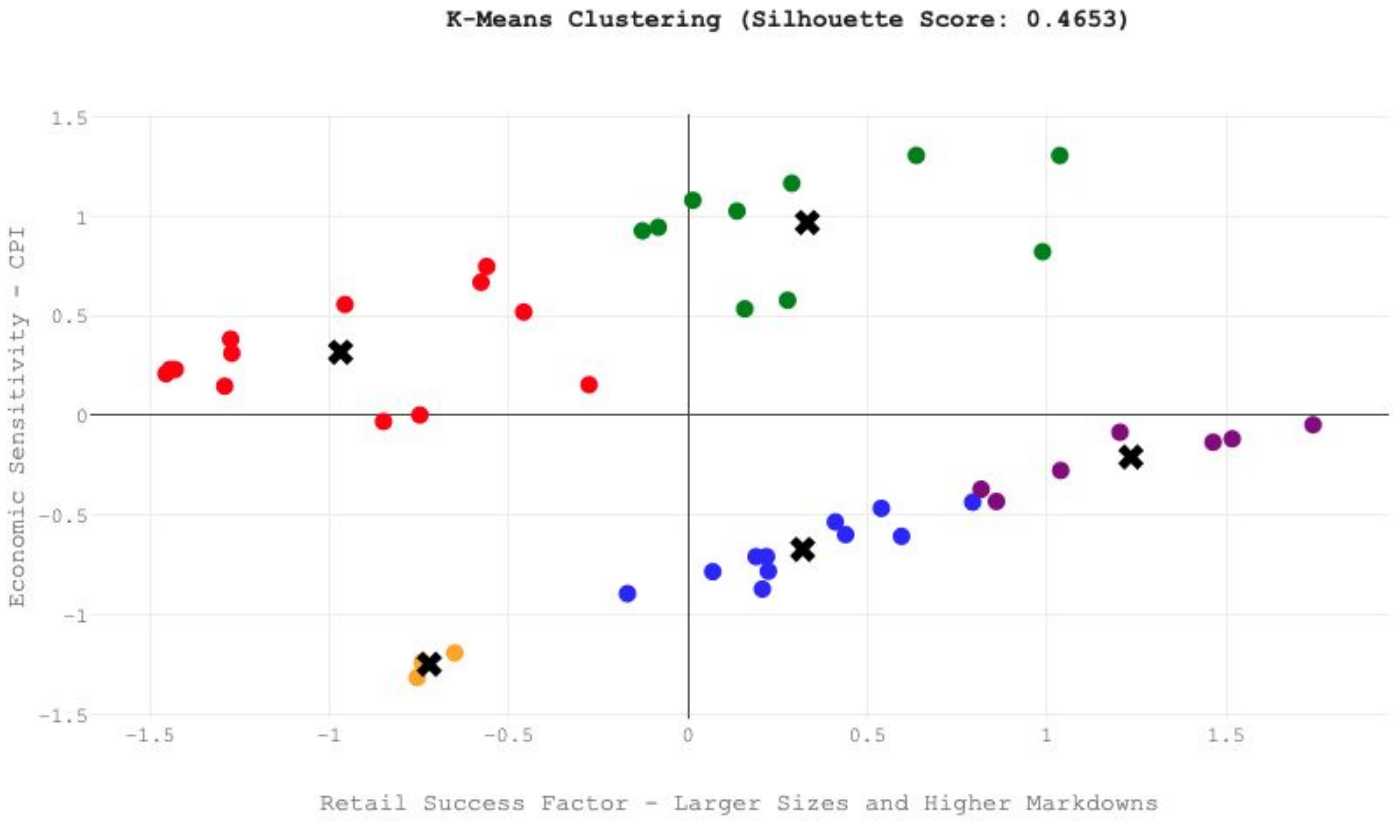
05. K-Means Clustering Analysis for Store Optimization - *EDA*



05. K-Means Clustering Analysis for Store Optimization - *Feature Engineering*



05. K-Means Clustering Analysis for Store Optimization - *Modeling*



05. K-Means Clustering Analysis for Store Optimization - *Cluster Profile*

Cluster 0: Moderate Performers with Cost-Efficiency Opportunities

Achieving moderate sales (approx. $1.29M)
Revenue performance depends heavily on markdown strategies.
Relatively large store sizes (around 122K sq. ft.).
The lower CPI (132.9) suggests low cost of living opportunities for further improvement.

Strategy Focus
- Optimize markdown ROI
- Consider strategies to enhance profitability by focusing on operational costs.

Cluster 1: Low Performers with Cost-Efficiency Opportunities

Generating lower sales ($688.6K)
Lower markdown spending ($151.9K).
Located in higher CPI regions (around 203.4), might be inflating operational costs and reducing profitability.

Strategy Focus
- Drive Operational Efficiency to mitigate costs
- Optimize Markdown pricing Strategy to drive profitability

Cluster 2: Top Performers Driving Revenue Growth

Highest revenue generators ($1.59M) with significant markdown investments ($287.1K). While their large store sizes (around 195K sq. ft.) and high CPI (205.92) increases costs, the sales performance more than drawbacks.

Strategy Focus
- Prioritized for expansion and marketing efforts.
- Monitor Markdown Spending to ensure it doesn't erode profitability is critical.

Cluster 3: Low Performers with Limited Markdown Reliance

Lowest average weekly sales ($401.7K)
A relatively small markdown budget ($43.9K).
Lower CPI (128.98) indicates operational affordability, but the low sales performance remains a concern.

Strategy Focus
- May benefit from optimization or a reevaluation of resource allocation. Consider testing localized promotions or repurposing underperforming stores.
- Explore potential reasons for low sales and identify ways to drive traffic.

Cluster 4: Niche Markets with HIgh Operating Costs

Lowest average weekly sales ($401.7K)
A relatively small markdown budget ($43.9K).
Lower CPI (128.98) indicates operational affordability, but the low sales performance remains a concern.

Strategy Focus
- Consider shifting marketing strategies to focus on higher-margin products that can offset the high operating costs.
- Reallocate resources to optimize profitability without relying heavily on markdowns.

THANK YOU

Github Repo For Detailed Work

05. Appendix - Data Assumptions & Cleaning

Data Source: [DLG_Inc E-commerce Data]

- The e-commerce dataset is assumed to accurately capture all data points without omissions or placeholder values (e.g., 'unknown', 'none', 'na', '', 'n/a'). Validation checks were performed to identify and address unexpected or default-like values.

- Rows with identical values across all columns are assumed to be duplicates.. A total of 462 duplicated rows were identified and removed. Cross-referencing with a backup source may be required to validate this assumption.

- Each product_id is uniquely associated with a single, correct brand. Any inconsistencies in the brand labeling for the same product_id are assumed to be errors. To resolve these inconsistencies, the most frequently occurring brand value for each product_id was assumed to be the correct one.

- A user's session can involve multiple actions, such as viewing several products, adding items to the cart, or making a purchase. But if a user purchases multiple items within one session, that may indicate a possible error or unusual behavior (purchase count>4 are flagged, require future check).

- Long sessions with only one action could be flagged for potential issues (e.g., the session might have timed out or was recorded incorrectly). Upon checking, no records shown such behavior by session_id.

Data Source: [DLG_B&M Data]

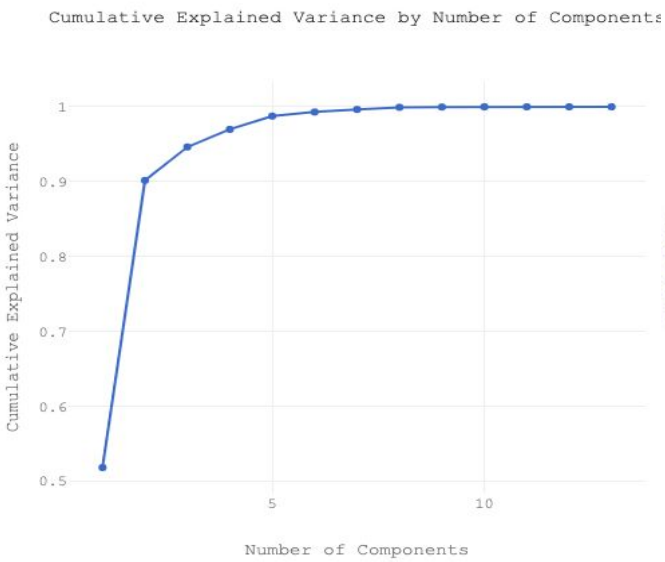
- The MarkDownl, MarkDown2, MarkDown3, MarkDown4, and MarkDown5 columns likely represent different price markdowns applied to products over time or during different promotional periods.

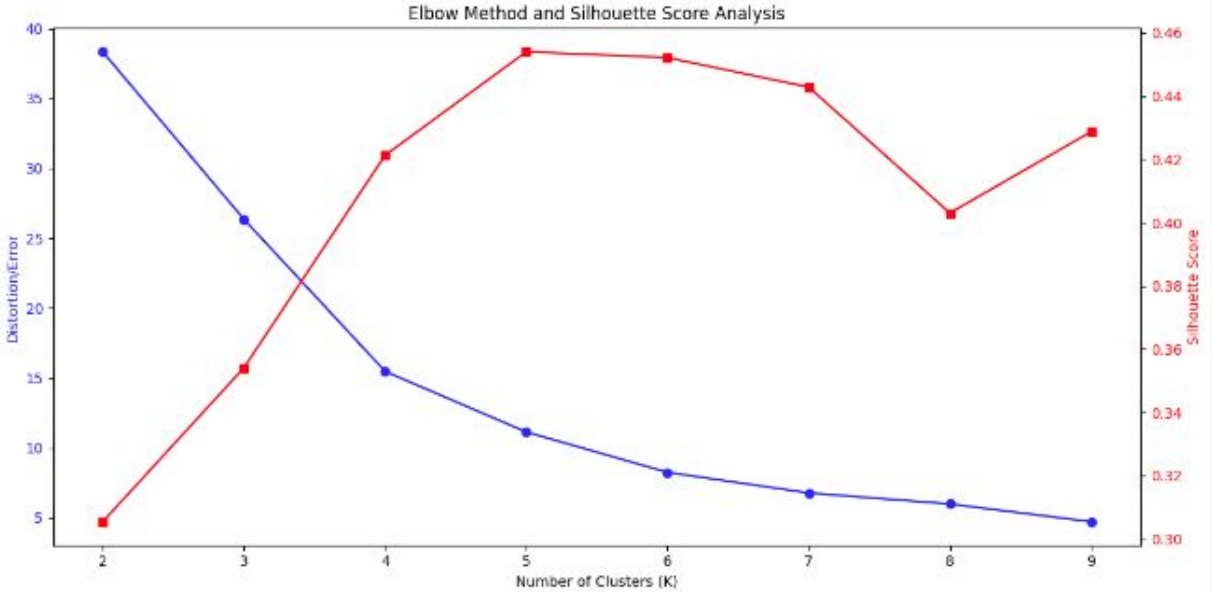
Clustering requires a unified dataset that combines all relevant features (e.g., sales, markdowns, and store attributes). Ensuring overlapping date ranges ensures that the combined dataset reflects valid, simultaneous observations across both datasets.

```
Date range in features_df:
2010-02-05 to 2013-07-26
Date range in historic_sales_df:
2010-02-05 to 2011-12-30
```

Dropped Type Feature since no relevant information revealed by it

05. Appendix - Selecting the most suitable N for PCA & clusters





05. Appendix - Analyzing PCA components loadings

```
Loadings of the first principal component (PC1)
Weekly_Sales_mean        0.301987
Weekly_Sales_min         0.286946
Weekly_Sales_max         0.308772
Weekly_Sales_std         0.295667
Temperature_mean        -0.086061
MarkDown_group145_sum    0.294837
MarkDown2_sum            0.301976
MarkDown3_sum            0.259133
CPI_mean                -0.259031
CPI_min                 -0.260179
CPI_max                 -0.258494
Size_first               0.311374
Total_Markdown           0.304084
Name: PC1, dtype: float64
```

```
Loadings of the second principal component (PC2)
Weekly_Sales_mean        0.172692
Weekly_Sales_min         0.186908
Weekly_Sales_max         0.163357
Weekly_Sales_std         0.132626
Temperature_mean         0.087948
MarkDown_group145_sum    0.115551
MarkDown2_sum            0.148464
MarkDown3_sum            0.098228
CPI_mean                 0.508971
CPI_min                  0.510629
CPI_max                  0.508250
Size_first               0.209877
Total_Markdown           0.126612
Name: PC2, dtype: float64
```

```
Loadings of the third principal component (PC3)
Weekly_Sales_mean        0.118689
Weekly_Sales_min         0.180384
Weekly_Sales_max         0.046893
Weekly_Sales_std        -0.023674
Temperature_mean         0.948420
MarkDown_group145_sum    0.038870
MarkDown2_sum            0.011673
MarkDown3_sum           -0.146096
CPI_mean                -0.071007
CPI_min                 -0.069970
CPI_max                 -0.067970
Size_first              -0.111781
Total_Markdown          -0.032717
Name: PC3, dtype: float64
```

```
Loadings of the fourth principal component (PC4)
Weekly_Sales_mean       -0.019124
Weekly_Sales_min         0.049500
Weekly_Sales_max        -0.144140
Weekly_Sales_std        -0.297616
Temperature_mean         0.017463
MarkDown_group145_sum    0.168595
MarkDown2_sum           -0.232228
MarkDown3_sum           -0.359747
CPI_mean                -0.031610
CPI_min                 -0.032333
CPI_max                 -0.031283
Size_first               0.810680
Total_Markdown          -0.124618
Name: PC4, dtype: float64
```

	Retail Success Factor - Larger Sizes and Higher Markdowns	Economic Sensitivity - CPI	Climate Influence – Temperature-Driven Sales Patterns	Operational Scale – Larger Store Sizes with Reduced Markdown
0	-0.083084	0.945349	0.180553	0.030112
1	0.636193	1.306152	0.195650	-0.060658
2	-1.275479	0.383321	0.126219	-0.223829
3	1.463638	-0.133800	0.241659	0.037691
4	-1.272157	0.312914	0.077948	-0.196145

05. Appendix - clusters remapping to original dataset demo.