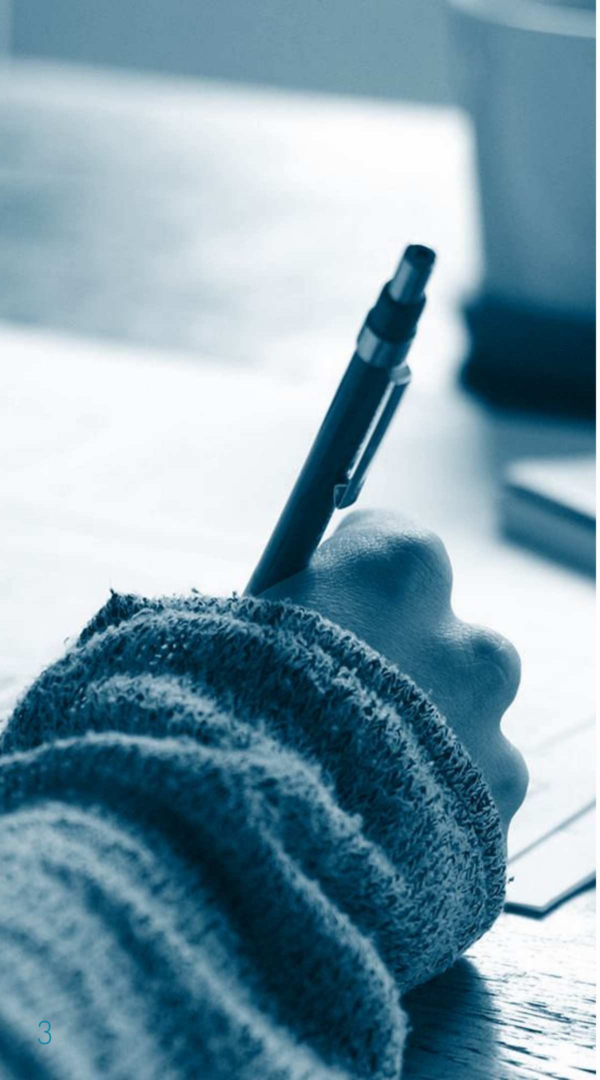
MIT - Applied Data Science Program

Marketing Campaign Customer Segmentation

Sunil Pradhan Sharma

Dec 2024

Table of Contents



Executive Summary

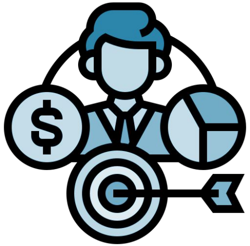
The marketing campaign analysis aimed to segment customers for targeted campaigns, enhancing marketing efficiency and ROI through data-driven insights.

Key Take-away

Customer Segmentation: Identified key customer segments using attributes like income, age, spending patterns, and campaign responses.

Behavioral Analysis: Analyzed spending on products, marketing engagement, and channel preferences.

Cluster Profiling: Defined distinct customer groups, highlighting differences in income, purchasing power, and responsiveness.



Most Important Findings:

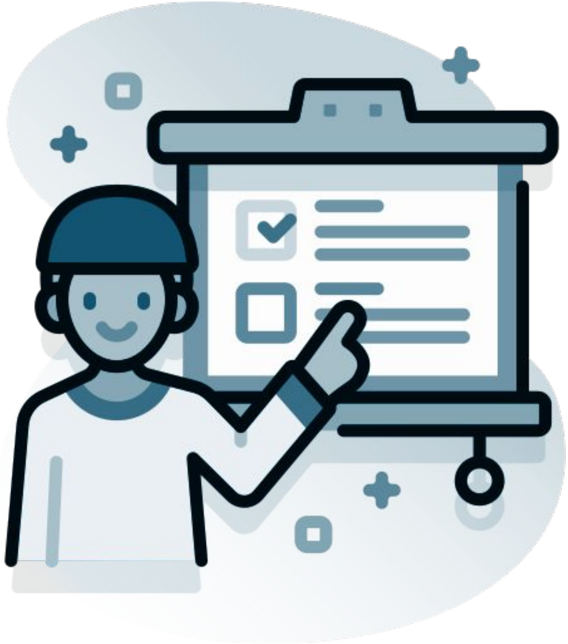
High-income customers spend more on premium products like wine and gold.

Families prioritize essential products, showing distinct spending patterns based on family size.

Campaign responses remain low, indicating the need for personalized marketing strategies.

Final Proposed Model Specifications:

- Optimal K for clustering determined as 3 using K-Means.
- Cluster characteristics:
  - Cluster 0: High-income, high-spending, premium product buyers.
  - Cluster 1: Moderate-income, family-oriented, moderate spenders.
  - Cluster 2: Low-income, low-spending, minimal campaign engagement.



Problem Statement

Problem Definition

- The challenge lies in businesses' inability to effectively engage diverse customer segments, leading to inefficient marketing strategies and suboptimal resource utilization.
- Without proper customer segmentation, marketing efforts remain broad and untargeted, resulting in reduced engagement and a lower return on investment. In an era where customers expect personalized experiences, the lack of segmentation prevents businesses from addressing specific needs, behaviors, and preferences.
- This gap in understanding customer dynamics hinders the ability to optimize marketing campaigns, drive meaningful customer engagement, and maximize revenue potential.

Why solving this problem matters ?

Enhanced Marketing Efficiency

Segmentation helps design targeted campaigns, optimizing resource allocation (time, budget, and effort) for each customer group.

Improved ROI

Personalized communication leads to higher engagement, with segmented campaigns showing 100% more clicks and 6-7x revenue growth.

Customer Satisfaction & Retention

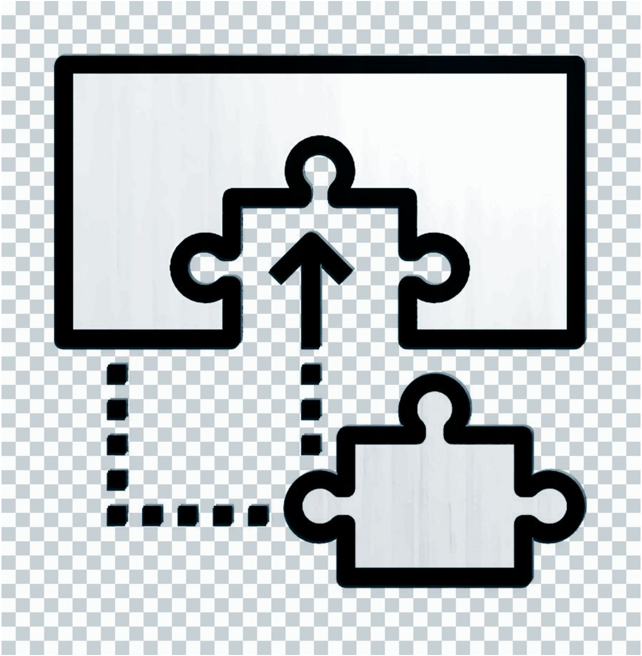
Tailored strategies meet individual needs, increasing satisfaction and fostering loyalty.

Strategic Insights for Decision Making

Segmentation provides insights to inform business strategies like product development, pricing, and customer service.

The Key Questions

- **Customer Profile**: What are the demographic, geographic, and behavioral characteristics of the customer base?
- **Customer Preferences:** What products, services, or features do different customer segments prefer, and how do these preferences vary across groups?
- **Customer Interaction:** What are the key engagement patterns across various marketing channels, and which channels are most effective for specific segments?
- **Customer Decision Drivers:** What are the key factors influencing customer purchase decisions, and how do motivations differ across segments?
- **Segment Value:** Which customer groups contribute most to revenue and profit, and what is the lifetime value of each segment?
- **Targeting Strategy:** What are the best strategies for engaging each segment, and which personalized offers or campaigns will resonate most?
- **Campaign Impact:** How do segmented campaigns perform compared to non-segmented campaigns in terms of conversion rates and ROI?
- **Resource Allocation:** How should marketing resources be optimized to maximize effectiveness across segments?



Solution

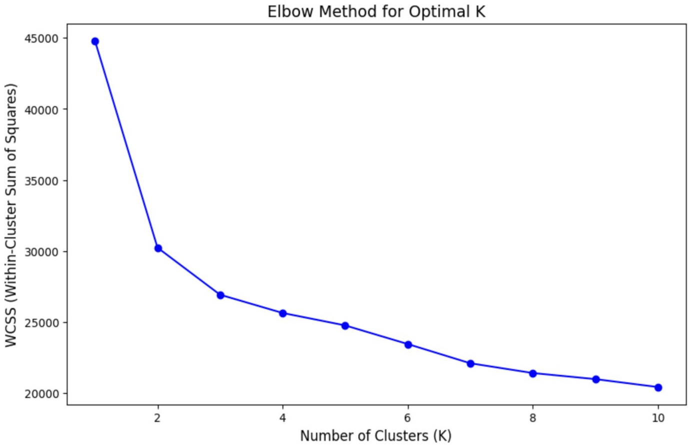
Final Proposed Solution Design

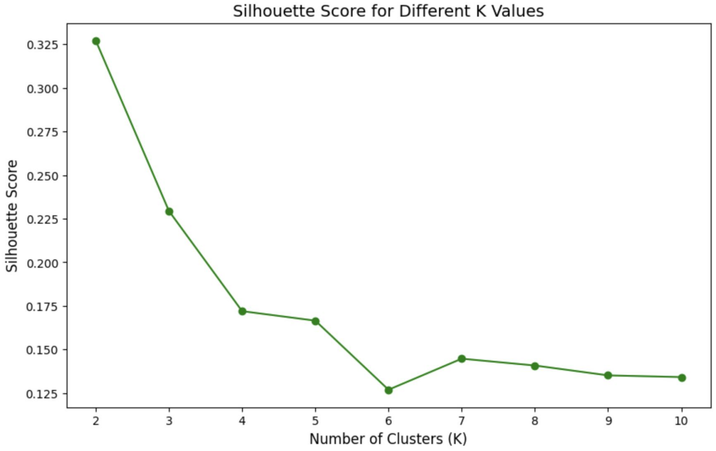
- **Data Preparation**:
  - Handled missing data and outliers.
  - Engineered features such as customer age, family size, and total spending.
- **Dimensionality Reduction**:
  - Applied PCA and t-SNE for data visualization and dimensionality reduction.
- **Customer Segmentation Using K-Means**:
  - Determined the optimal number of clusters as 3 using the Elbow and Silhouette methods.
  - Created Cluster Profiles:
    - **Cluster 0:** High-income, **Cluster 1:** Moderate-income, **Cluster 2:** Low-income
- **Cluster Profiling Insights:**
  - Income and family size strongly influence spending patterns.
  - High-spending customers prefer wine and gold products, while families prioritize food products.

Comparison of Clustering Techniques

Technique	Optimal Clusters / Parameters	Strengths	Limitations
K-Means Clustering	K=3 (Elbow Method & Silhouette Score)	Works well with clearly separable clusters. Scalable for large datasets.	Sensitive to outliers. Requires pre-specification of the number of clusters.
K-Medoids Clustering	K=3 (Distortion Minimization)	Robust to outliers compared to K-Means.	Computationally more expensive for large datasets.
DBSCAN	eps=0.9, min_samples=10 (Silhouette Score)	Identifies clusters of arbitrary shapes. Handles noise and outliers effectively.	Performance highly dependent on hyperparameter selection.
Gaussian Mixture Model (GMM)	K=2 (BIC & AIC)	Assigns probabilities to cluster memberships. Handles overlapping clusters better than other techniques.	Computationally intensive for large datasets.

Optimal K





- Optimal K based on the Elbow Method: **3**
- Optimal K based on the Silhouette Score: **2**

Performance Analysis

Technique	Performance
K-Means Clustering	Performs well for clearly defined, separable clusters. The Elbow Method identified K=3 as optimal, and the clusters are interpretable. Sensitive to outliers, which may slightly affect its performance in noisy datasets.
K-Medoids Clustering	Robust to outliers and noise, providing better cluster stability than K-Means. Produces well-defined clusters with K=3, with a moderate Silhouette Score.
DBSCAN	Identifies clusters of arbitrary shapes and handles outliers effectively. Performance depends heavily on eps and min_samples values, which need careful tuning. Excels in datasets with irregular cluster shapes or noisy data.
Gaussian Mixture Model (GMM)	Handles overlapping clusters better by assigning probabilities to each cluster. Suitable for probabilistic clustering scenarios, but computationally expensive. Worked best with K=2 clusters as identified by Bayesian Information Criterion (BIC).

Why DBSCAN is the Best Solution to Adopt ?

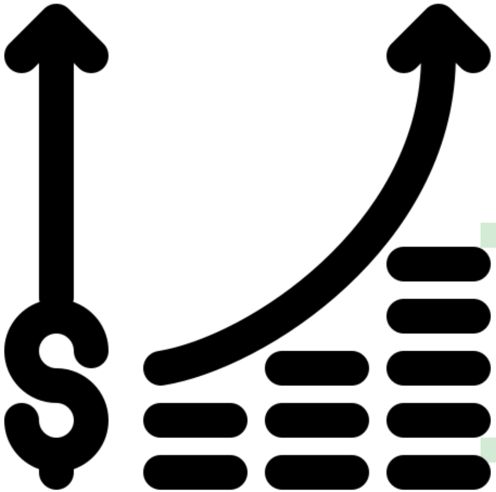
- **Robust to Outliers:**
  - DBSCAN is effective in identifying and isolating noise points (Cluster -1).
  - This ensures that anomalous customer behaviors do not adversely impact clustering results.
- **Handles Arbitrary Cluster Shapes:**
  - DBSCAN is not constrained to linear or spherical clusters, making it ideal for datasets with irregular cluster structures.
- **No Pre-Specified Number of Clusters:**
  - Unlike K-Means and K-Medoids, DBSCAN does not require pre-defining the number of clusters.
  - This makes it flexible for exploratory data analysis where the optimal number of clusters is not known in advance.
- **Meaningful Clusters:**
  - The clusters formed by DBSCAN align well with customer spending and channel usage patterns.
  - Low Spenders and High Spenders are distinctly identified, providing actionable segmentation.
- **Noise Handling:**
  - Outliers detected as noise can represent potential data errors or unusual customer behaviors.
  - Analyzing these outliers separately adds depth to customer insights.

Why DBSCAN is the Best Solution to Adopt ?

- **Silhouette Score:**
  DBSCAN achieved competitive silhouette scores for optimal eps and min_samples values, indicating well-defined clusters.
- **Interpretability:**
  Clusters formed by <u>DBSCAN align with business objectives and provide clear segmentation</u> for actionable strategies.

Why is this Valid Solution ?

- **Data-Driven Insights**:
  - Clustering revealed meaningful customer segments based on real purchasing and demographic data.
- **Personalized Marketing**:
  - Tailored campaigns for each cluster maximize ROI and enhance engagement.
- **Resource Optimization**:
  - Budget and marketing efforts can be focused on high-value customer groups.



- **Business Impact:**
  - Increased customer retention through loyalty programs.
  - Improved revenue by targeting premium customers with personalized offers.
  - Enhanced operational efficiency with strategic marketing investments.

Recommendations

Recommendations

- **Targeted Marketing Campaigns**:
  - Launch personalized campaigns based on customer clusters.
  - Use insights from Cluster 0 (high-income, premium spenders) for luxury promotions.
  - Develop family-oriented promotions targeting Cluster 1 (moderate-income families).
  - Offer discounts and budget-friendly deals for Cluster 2 (low-income, low-spending customers).
- **Resource Allocation**:
  - Focus marketing budgets on high-value clusters (Clusters 0 and 1).
  - Increase email and web campaigns targeting responsive segments.
- **Customer Engagement Strategies**:
  - Introduce loyalty programs for long-term customers in Clusters 0 and 1
  - Offer personalized incentives based on recency and past purchases..
- **Product Bundling & Cross-Selling:**
  - Bundle complementary products (e.g., wine and gold products) for Cluster 0.
  - Promote family-sized packs for Cluster 1.
- **Channel Optimization:**
  - Enhance digital channels for high-spending segments.
  - Use offline promotions and direct mail for less tech-savvy customers.

Expected Benefits and Costs

- **Expected Benefits**:
  - **Revenue Growth:** Estimated 15%-25% increase from targeted campaigns.
  - **Improved ROI:** Personalized offers reduce marketing waste, boosting engagement.
  - **Customer Retention:** Enhanced loyalty programs increase long-term retention.

  - Develop family-oriented promotions targeting Cluster 1 (moderate-income families).
  - Offer discounts and budget-friendly deals for Cluster 2 (low-income, low-spending customers).

- **Estimated Costs**:
  - **Marketing Campaign Costs: $50,000 - $100,000 annually**
  - **Data Infrastructure Costs: $25,000 annually** for maintaining a data warehouse and analytics tools.

Key Next-Steps

Steps for Stakeholders

* **Marketing Team:** Develop personalized campaigns for each cluster.

* **Data Analysts**: Refine data models using advanced clustering methods.

* **Product Development**: Design products aligned with identified customer needs.

* **Customer Service**: Offer tailored support for high-value customers.

+

Maximizing the Solution's Potential

*Use predictive analytics for campaign effectiveness forecasting.

*Implement multi-channel campaigns based on customer preferences.

Campaigns and other offerings that are personalized to the customers

Customers with inconsistent or rare purchasing habits

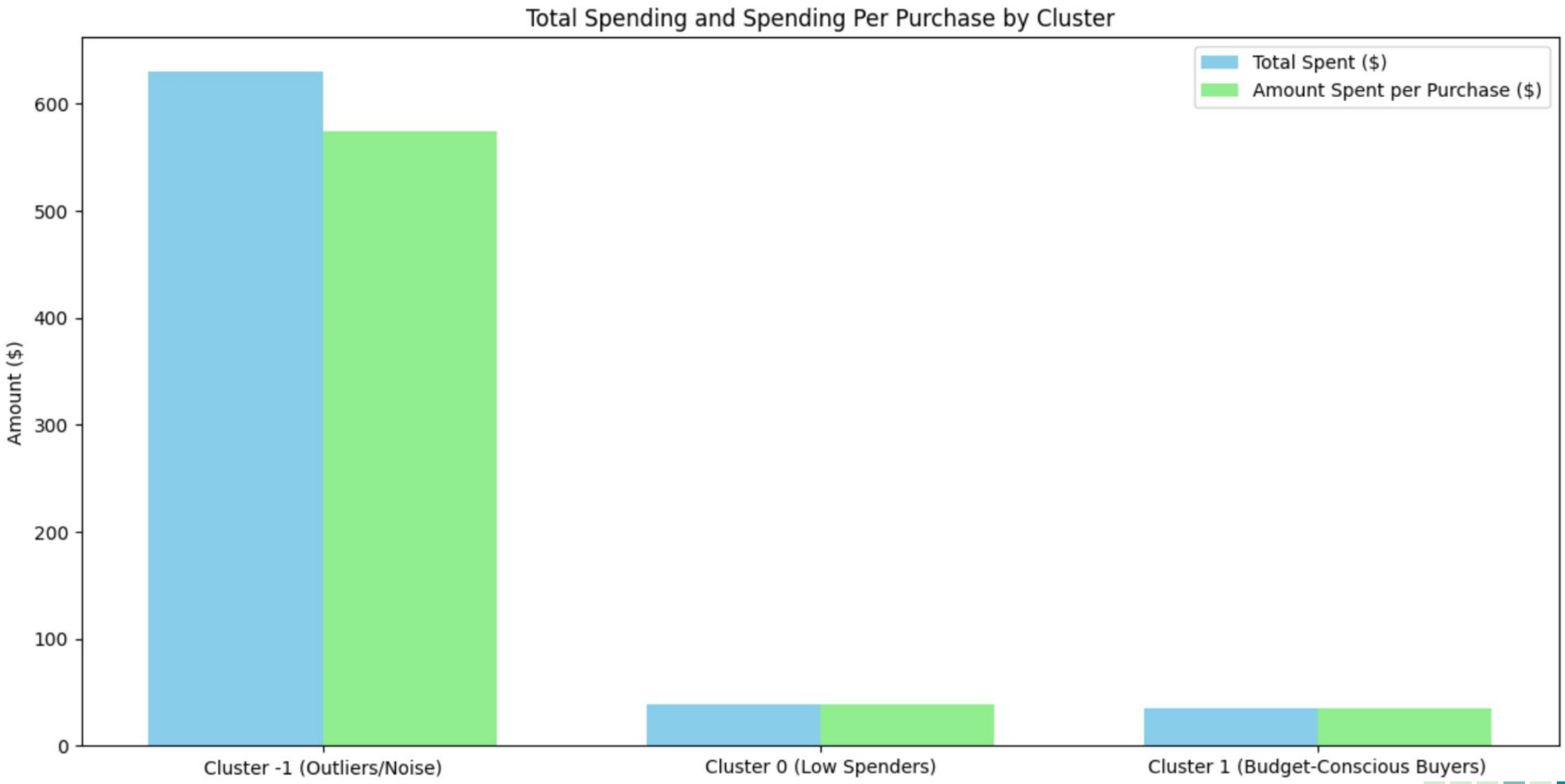
Cluster	Customer Profile	Campaigns and Offerings	Product Recommendations
Noise or Outliers (Cluster -1)	Categorized as outliers by DBSCAN, representing unusual behavior or anomalies. Customers with inconsistent or rare purchasing habits, requiring special attention.	Survey and Feedback Campaigns: Use surveys or incentivize customers to share feedback. Personalized Retargeting Offers: Target them with personalized offers to encourage consistent purchasing behavior. Exclusive Invitations: Offer VIP treatment, such as product previews or personalized shopping experiences.	Exclusive Invitations: Target outliers with personalized offers to turn them into more predictable customers

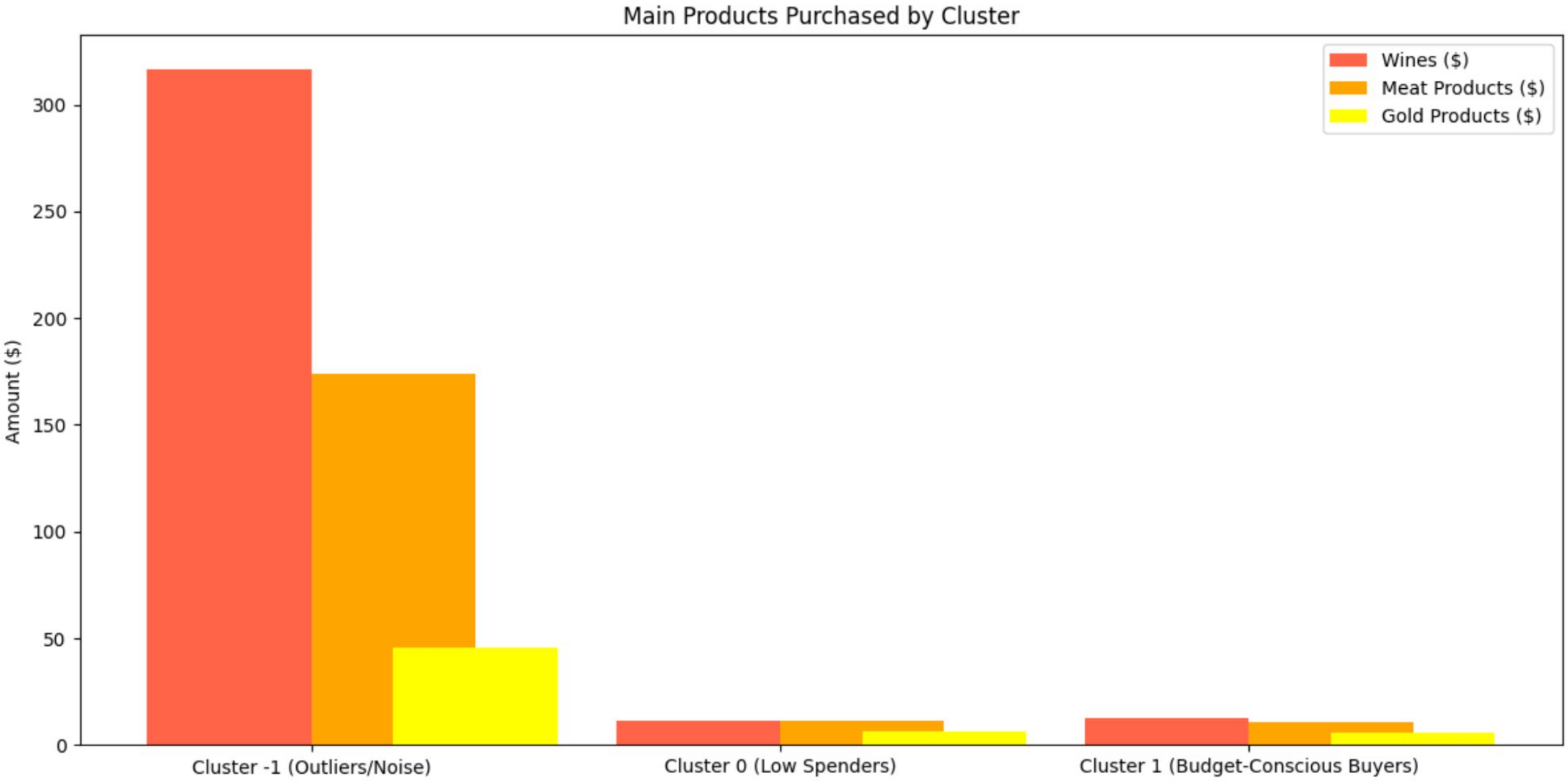
Customers with low spending habits and fewer transactions

Cluster	Customer Profile	Campaigns and Offerings	Product Recommendations
Low Spenders (Cluster 0)	Customers with low spending habits and fewer transactions. - Purchase basic items, engage more on the web, and make fewer purchases in catalogs or stores.	Budget-Friendly Bundles: Create bundles for wines, fruits, and meats, emphasizing discounts for purchasing multiple products. Discounted Offers for Web Engagement: Offer web-exclusive discounts or incentives for repeat purchases. Loyalty Programs: Offer gamified loyalty programs where they earn points for every purchase.	Entry-Level Product Recommendations: Recommend basic versions of wines, fruits, and meats. - First Purchase Discounts: Provide bundle discounts for affordable items.

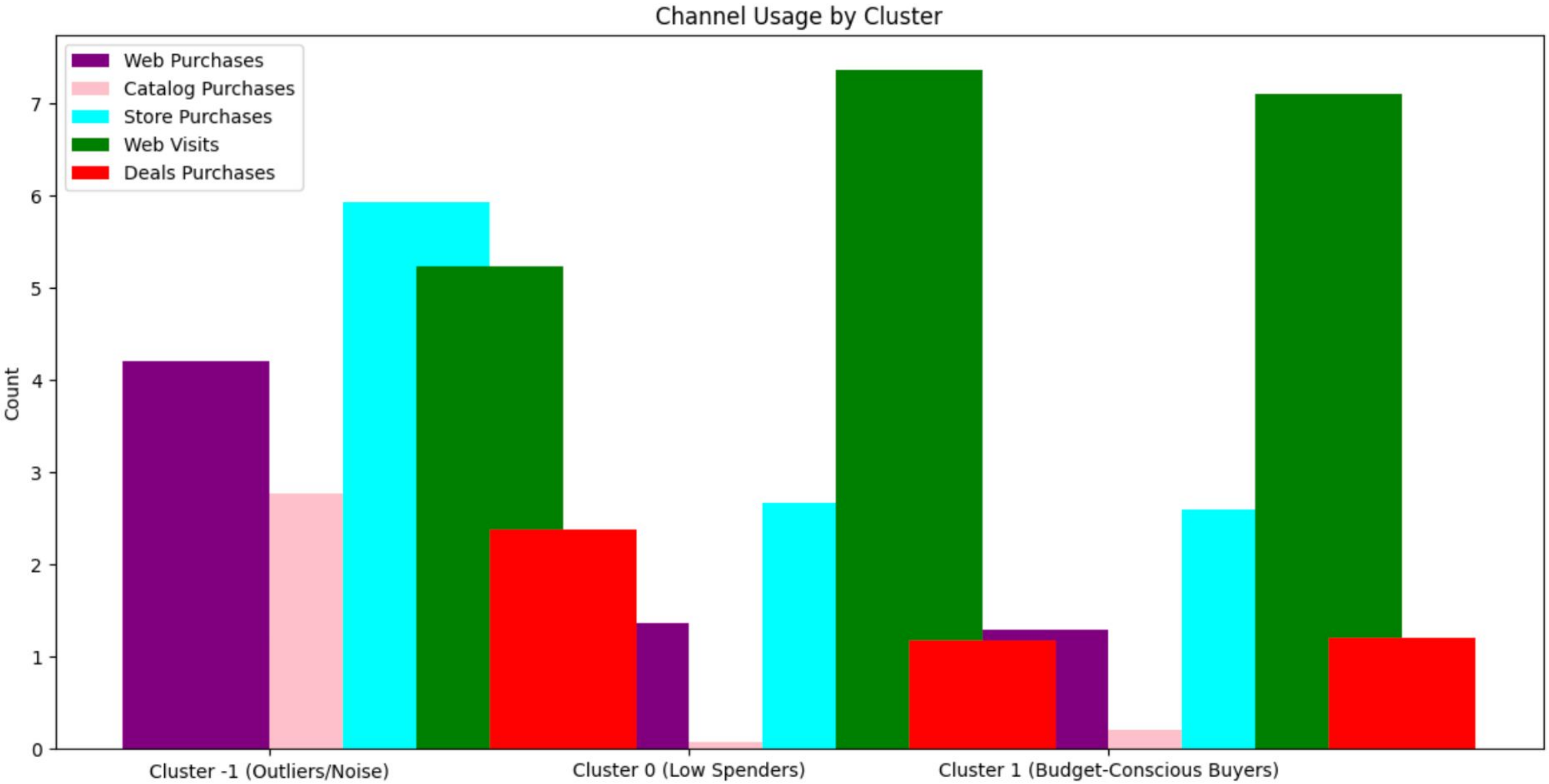
High-value spenders purchasing

Cluster	Customer Profile	Campaigns and Offerings	Product Recommendations
High Spenders (Cluster 1)	High-value spenders purchasing premium products (wines, meats, gold). - Engage in various channels (online, catalog, and in-store), showing a preference for luxury products.	Exclusive VIP Campaigns: Target with premium offerings like wine tastings, personalized shopping, and early access to new product lines. - Premium Product Bundles: Create high-value bundles with wines, gold products, and gourmet meats, offering discounts for purchasing together. - Loyalty and Membership Programs: Offer VIP memberships with personalized services.	Luxury Product Recommendations: Provide tailored recommendations for rare wines, gourmet meats, and gold products. - Personalized Catalogs: Send catalogs featuring only the most relevant high-end products based on their preferences.


Total Spending and Spending Per Purchase by Cluster
Total Spent ($)
Amount Spent per Purchase ($)
Amount ($)
600
500
400
300
200
100
0
Cluster -1 (Outliers/Noise)
Cluster 0 (Low Spenders)
Cluster 1 (Budget-Conscious Buyers)



Main Products Purchased by Cluster
Wines ($)
Meat Products ($)
Gold Products ($)
Amount ($)
300
250
200
150
100
50
0
Cluster -1 (Outliers/Noise)
Cluster 0 (Low Spenders)
Cluster 1 (Budget-Conscious Buyers)


Channel Usage by Cluster
Web Purchases
Catalog Purchases
Store Purchases
Web Visits
Deals Purchases
Count
0
1
2
3
4
5
6
7
Cluster -1 (Outliers/Noise)
Cluster 0 (Low Spenders)
Cluster 1 (Budget-Conscious Buyers)

THANK YOU !

Any questions?

You can find me at:

sharmasunilpradhan@gmail.com

https://www.linkedin.com/in/sunil-p-sharma/

Appendix

Data Background and Observations

- **Shape:** The dataset contains 2,240 rows and 27 columns, detailing customer information and marketing campaign interactions.
- **Data Types:** The data includes 23 int64 columns, 1 float64, and 3 categorical (object) columns (e.g., Education, Marital_Status, Dt_Customer).
- **Missing Values:** 24 missing values (2.2%) in the Income column.
- **Statistical Summary:** Numerical: Income ranges from 0 to 1493, with a mean of 67.2k. Other attributes (e.g., MntWines, MntMeatProducts) range from 0 to 1725.
- Categorical: Insights into customer education levels and marital status.
- **Initial Observations:**
    - **Outliers:** Extreme values in income and spending columns, which may require further analysis.
    - **Spending:** Skewed distribution in spending across categories like MntWines and MntGoldProds.
    - **Campaign Responses:** Low engagement in campaign response variables (AcceptedCmp1 to AcceptedCmp5).

Key Patterns & Insights

- **Customer Behavior:**

  There is a noticeable variation in how customers engage with different types of campaigns. For instance, most customers have not responded to campaigns (AcceptedCmp1 to AcceptedCmp5), while a small portion has engaged significantly.

- **Income and Spending Patterns:**

  The dataset reveals high variability in customer income and spending, which could be crucial for targeting high-value customers in marketing strategies.

- **Outliers:**

  Income and spending columns like MntWines and MntMeatProducts contain significant outliers that may skew analysis and require data transformation (e.g., capping or log transformation).

Data Treatments and Preprocessing Required

Handling Missing Values:

The Income column contains 24 missing values. These can be imputed with the median or mean to avoid bias, or treated with more advanced imputation methods like regression or k-NN if deemed necessary.

Outlier Detection and Treatment:

Extreme values in numerical columns such as Income, MntWines, and MntGoldProds need to be identified and addressed, potentially through capping, transformation (e.g., log transformation), or removal, depending on their impact on the analysis.

Feature Scaling:

Numerical features such as Income, MntWines, and NumWebPurchases should be normalized or standardized to ensure equal contribution in models and prevent skewed results due to differing magnitudes.

Data Treatments and Preprocessing Required (Contd ...)

Date Feature Transformation:

The Dt_Customer column, representing the customer's registration date, should be converted into a more usable format, such as extracting the customer's age or calculating the recency of their engagement with the business.

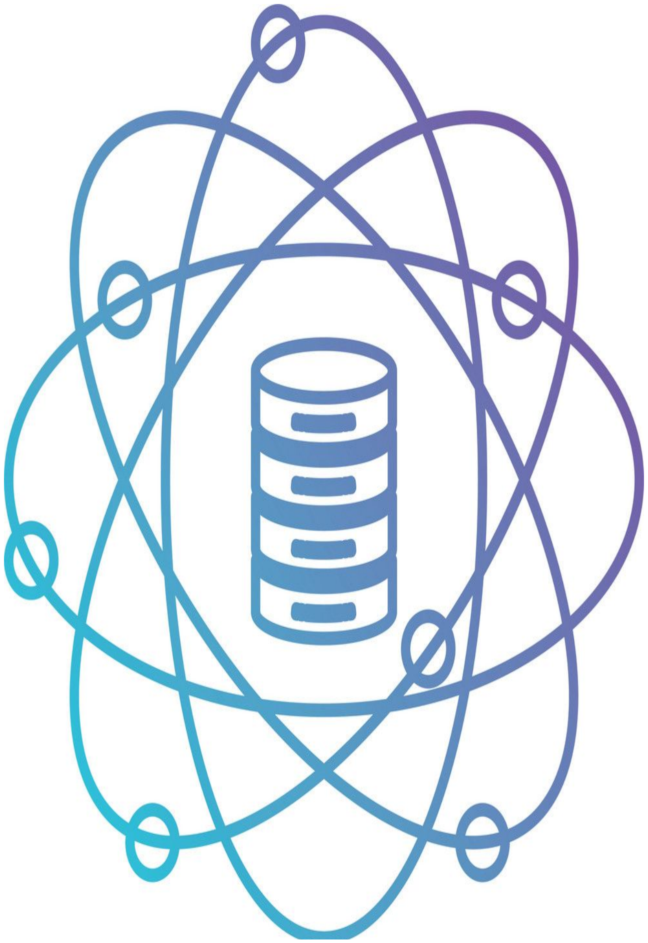
Variable Engineering:

Age: Derived from the Year_Birth column (e.g., 2024 - Year_Birth).

Customer Engagement: Deriving a new variable for the time since the last purchase or engagement based on Recency and other related features could provide valuable insights into customer loyalty.

Skewed Distributions:

Features with highly skewed distributions, such as spending amounts (MntWines, MntMeatProducts), may benefit from transformations to reduce skew and normalize their distribution, which can improve the performance of predictive models.



Building Data Models

Technique	Overview	Why Explore This Technique	Considerations	Best for ?
K-Means Clustering	Partitions data into a predefined number of clusters based on feature similarity. Each cluster has a centroid that minimizes within-cluster variance.	K-Means is effective when you expect well-separated customer groups with clear boundaries, based on customer demographics and spending patterns.	Sensitive to initial centroids and outliers. Assumes spherical clusters, which may not work well for complex patterns.	Segmenting customers into distinct groups, such as targeting different spending behaviors or engagement levels.
Gaussian Mixture Models (GMM)	A probabilistic model that assumes the data is generated from a mixture of Gaussian distributions. Provides soft clustering, allowing a data point to belong to multiple clusters.	GMM is more flexible than K-Means and can model elliptical or overlapping clusters. It‚Äôs useful for customer segmentation where the boundaries between segments are not clear-cut.	Requires careful tuning of the number of components (clusters). May be computationally more expensive than K-Means.	More complex segmentation where customers may exhibit overlapping characteristics, such as mixed behavior or engagement patterns.

Technique	Overview	Why Explore This Technique	Considerations	Best for ?
K-Medoids Clustering	Similar to K-Means but uses actual data points (medoids) as the cluster centers. Works well with datasets containing outliers.	K-Medoids is more robust to outliers compared to K-Means, making it suitable for customer data with extreme or noisy values, such as very high-income customers.	Computationally more expensive than K-Means. Sensitive to the initial choice of medoids.	Datasets with outliers or noisy data, like customers with extreme spending or engagement behaviors.
Principal Component Analysis (PCA)	A dimensionality reduction technique that transforms data into a smaller set of uncorrelated components while retaining most of the variance.	PCA reduces the complexity of the data by extracting the most important features, which is useful when you have many features (e.g., income, spending patterns) that may be highly correlated.	Reduces dimensionality but may result in loss of some information. Requires standardization of features for optimal performance.	Reducing feature space before clustering, making the clustering process more efficient and interpretable. Could be followed by K-Means or GMM clustering for improved results.

Technique	Overview	Why Explore This Technique	Considerations	Best for ?
t-SNE	A non-linear dimensionality reduction technique that focuses on preserving local structures of the data. Typically used for visualization.	t-SNE is ideal for visualizing complex high-dimensional data in 2D or 3D, helping to uncover patterns or clusters in the data that may not be obvious in high-dimensional space.	Computationally intensive for large datasets. Results are more useful for visualization and not for actual clustering.	Visualizing the results of clustering algorithms like K-Means or GMM to see if customer segments are well separated in lower dimensions.
Silhouette Score	A metric used to evaluate the quality of clustering by measuring how close each point in one cluster is to the points in the neighboring clusters.	Silhouette Score helps determine the optimal number of clusters by measuring how well-separated the clusters are. It validates the clustering structure, ensuring clusters are well-formed.	Higher scores indicate better-defined clusters, but interpretation can be subjective. May require multiple cluster numbers for comparison.	Evaluating clustering results and determining the optimal number of clusters in K-Means, GMM, or K-Medoids.

Technique	Overview	Why Explore This Technique	Considerations	Best for ?
Elbow Method	A method for determining the optimal number of clusters by plotting the Within-Cluster Sum of Squares (WCSS) for different numbers of clusters.	The Elbow Method helps identify the point at which adding more clusters no longer significantly improves the model, ensuring the model isn't overfitted with unnecessary clusters.	The 'elbow' point may not always be obvious. Requires running K-Means clustering for multiple cluster sizes.	Selecting the optimal number of clusters before applying K-Means clustering, ensuring the chosen number is not too high or too low.

Technique	Overview	Why Explore This Technique	Considerations	Best for ?
DBSCAN	DBSCAN is a density-based clustering algorithm that groups together points closely packed in a region while marking points in sparse regions as outliers (noise). The algorithm works by defining a neighborhood based on a radius (eps) and a minimum number of points (min_samples), identifying core points, and expanding clusters by connecting core points and their neighbors iteratively.	Robustness to Noise: Capable of identifying and excluding outliers. Arbitrary Cluster Shapes: Can identify non-linear clusters. No Pre-Specified Number of Clusters: Determines clusters based on density. Improved Interpretability: Isolates outliers and forms distinct clusters.	Hyperparameter Sensitivity: Sensitive to \`eps\` and \`min_samples\` selection. Scalability: Computationally expensive for large datasets. Sparse Data Issues: Struggles with sparse data. Distance Metric: The choice of distance metric affects results.	Noisy Datasets: Identifies and excludes noise. Irregular Cluster Shapes: Works with non-linear and non-spherical clusters. Exploratory Data Analysis: Ideal when the number of clusters is unknown. Customer Segmentation: Useful for identifying customer groups and outliers.