CUSTOMER SEGMENTATION REPORT

Data Overview and Summary

This report summarizes the findings of the **K-Means clustering** analysis performed on customer behavioral and demographic data, as detailed in the uploaded Colab Notebook. The goal of the project was to segment the customer base to enable more targeted and effective marketing strategies

1. Methodology and Data Overview

The analysis utilized the Customer_Behaviour.csv dataset, comprising 400 customer records with five key variables: User ID, Gender, Age, Estimated Salary, and Purchased (a binary target variable where 1 indicates a purchase).

Key Statistics (Overall): The total purchase rate across all customers was **35.75%** (143 out of 400). The average customer age was approximately 37.66 years, with an average estimated salary of $69,743.

Clustering Features: The K-Means algorithm was applied using the two continuous variables: **Age** and **Estimated Salary**.

Preprocessing: These features were scaled using a **Min Max Scaler** to prevent the magnitude of salary from disproportionately influencing the clustering distance.

Optimal K: The analysis was performed with an optimal cluster number (**K=4**).

1. Optimal Cluster Determination

The optimal number of clusters (K) was determined using the Elbow Method (Inertia) and the Silhouette Score.

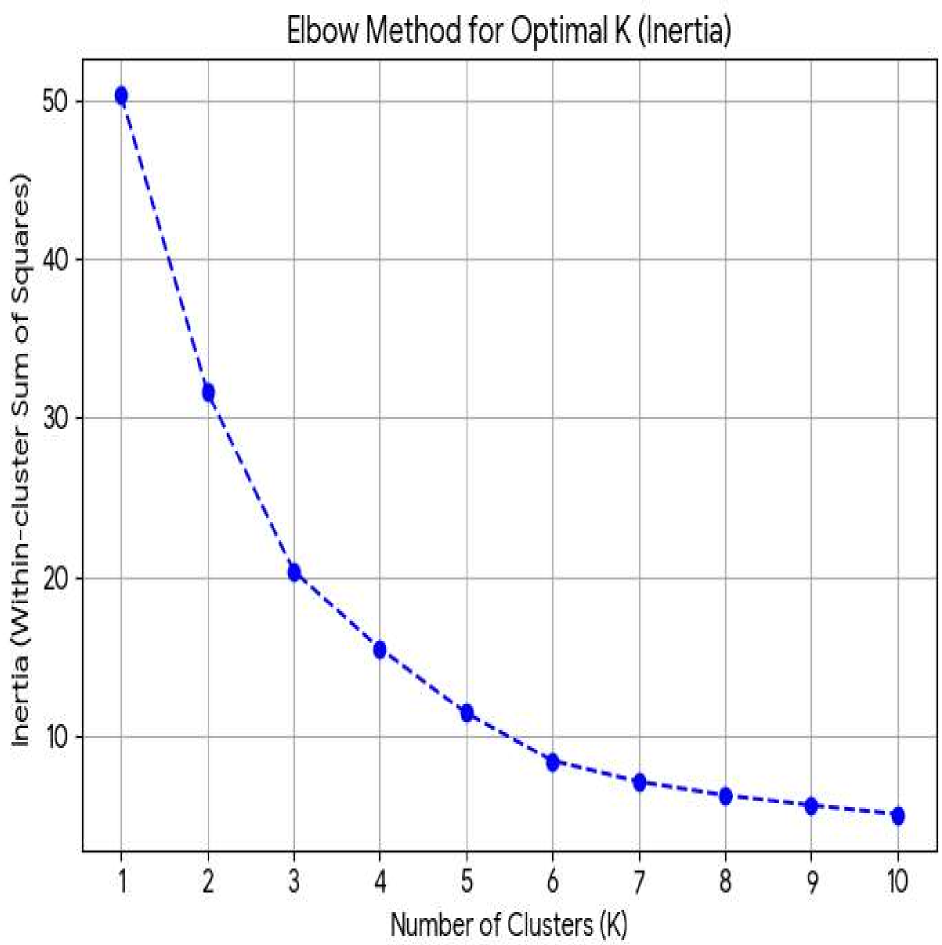
Elbow Method (Inertia)

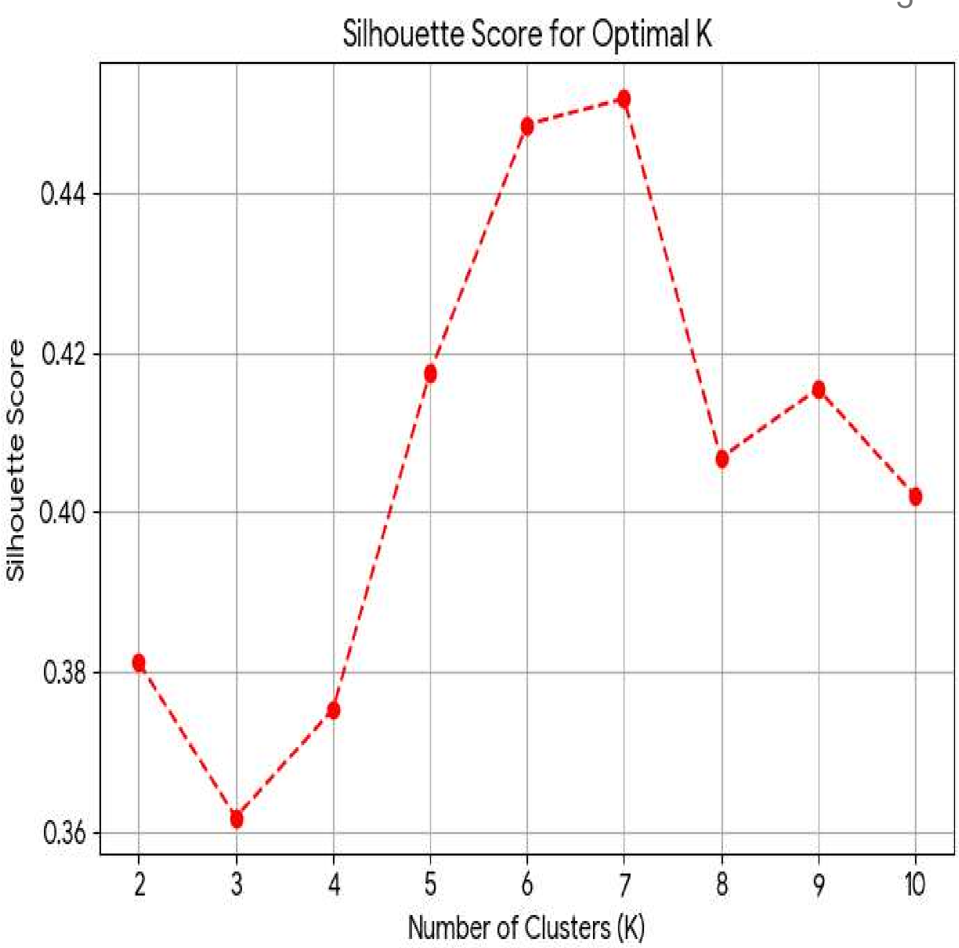
The Elbow Method plot shows the within-cluster sum of squares (Inertia) against the number of clusters. The "elbow" point, where the rate of decrease in Inertia sharply changes, suggests a reasonable choice for K. In this case, the elbow is most evident at K=4 or K=5.

Silhouette Score

The Silhouette Score measures how similar a data point is to its own cluster compared to other clusters. A higher score indicates better-defined clusters. The highest score (0.4518) was observed at K=7.

Given the trade-off between statistical performance (Silhouette Score) and practical interpretability, K=4 was chosen for the final clustering, as it represents a clear, interpretable segmentation based on the two input features (e.g., young/old and low/high salary segments) and is a strong candidate from the Elbow Method.


Elbow Method for Optimal K (Inertia)
Inertia (Within-cluster Sum of Squares)
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10
Number of Clusters (K)


Silhouette Score for Optimal K
Silhouette Score
0.44
0.42
0.40
0.38
0.36
2
3
4
5
6
7
8
9
10
Number of Clusters (K)

Data Normalization and Scaling

The numerical features, ***Age*** *and* ***Estimated Salary****, were scaled using the* ***Min Max Scaler****. This technique transforms the data so that all values fall within the range of [0,1], which is crucial for distance-based machine learning algorithms. The User ID, Gender, and the target variable Purchased were not scaled.*

2. K-Means Clustering and Visualization

K-Means clustering was performed using the chosen K=4. The resulting segments are visualized below using a scatter plot of the scaled features.
Since the clustering was performed on only two features
(Scaled Age and Scaled Estimated Salary), the data is already in 2D, making Principal Component Analysis (PCA) redundant for visualization.)

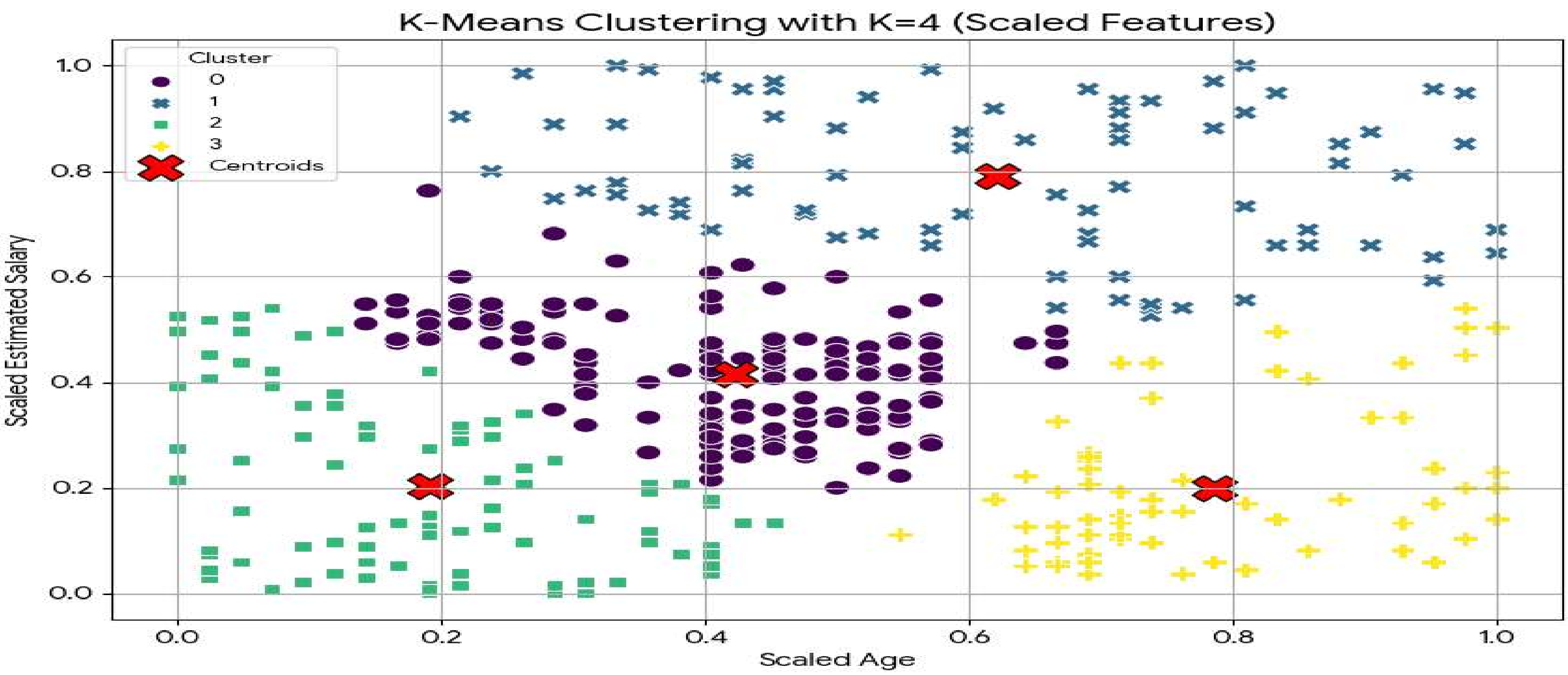
Cluster Segments

The scatter plot clearly shows four distinct segments, with the centroids marked by red 'X' s:

These segments align well with potential customer groups, where Cluster 2 ("Older & High Salary") and Cluster 1 ("Younger & High Salary") are typically target groups for certain products, and Cluster 2 strongly overlaps with the 'Purchased' group from the earlier EDA.

Cluster	Count
0	158
1	81
2	95
3	66

The K-Means clustering analysis was performed on the scaled Age and Estimated Salary features.



Visual Analysis Insights

The visualizations provide a clear picture of the data distributions and the relationship between features and the target variable (Purchased).

Age and Estimated Salary Distributions: Both Age and Estimated Salary distributions are relatively normal, with no extreme outliers visible in the histograms.

Purchase Count by Gender: The purchase rate appears similar for both Male and Female customers, suggesting gender alone may not be the strongest predictor of purchase behavior.

Age vs. Estimated Salary (Colored by Purchase):

Customers who Did Not Purchase (0) are predominantly found in the lower-right area of the plot (younger people with a mix of salaries) and the lower-left (older people with low salaries).

Customers who Purchased (1) are mostly concentrated in the upper-right area of the plot (older people with higher salaries), and a small cluster of older people with lower salaries.

This indicates that higher Age and higher Estimated Salary are positively correlated with a higher probability of purchase.

Marketing Recommendations

Based on the K-Means clustering of customer **Age** and **Estimated Salary**, the customer base has been segmented into four distinct groups.

Cluster Profiles and Common Traits

The clusters are defined by the mean characteristics of their members:

Cluster	Count	Avg. Age	Avg. Salary	Purchase Rate	Primary Trait
1	81	44.1	$121,654	**86.4%**	**High-Value, Affluent Buyers**
3	66	51.0	$41,955	**83.3%**	**Older, Value Buyers**
0	158	35.8	$71,146	11.4%	**Mid-Segment Majority**
2	95	26.0	$42,453	**0.0%**	**Budget-Conscious Youth**

Cluster	% Female	% Male
1	60.5%	39.5%
3	54.5%	45.5%
0	45.6%	54.4%
2	49.5%	50.5%

Marketing Strategies for Each Group

Cluster	***Profile Name***	***Common Traits***	***Recommendation***
1	High-Value, Affluent Buyers	≈44 years old, High Income, High Purchase Rate. Slightly Female bias.	Retention & Upsell: This is the most valuable segment. Focus on personalized, premium product offerings, exclusive bundles, and a top-tier loyalty program to ensure retention and maximize Lifetime Value (LTV).
3	Older, Value Buyers	≈51 years old, Low Income, High Purchase Rate.	Affordability & Trust: This group highly values the product despite a lower salary. Market stability, reliability, and value. Promote clear, simplified payment plans, seasonal discounts, and customer testimonials to build trust.
0	Mid-Segment Majority	≈36 years old, Moderate Income, Very Low Purchase Rate. Largest group.	Nurturing & Education: This group is the largest potential audience. Use educational content, limited-time offers, and mid-range pricing to move them toward conversion. Focus on clear ROI and product differentiation to overcome initial hesitancy.
2	Budget-Conscious Youth	≈26 years old, Low Income, Zero Purchase Rate.	Future Investment & Free Value: Do not spend heavily on immediate conversion. Focus on building brand awareness and loyalty through social media engagement, content marketing, and freemium or low-cost entry products. These are your buyers of tomorrow.

3. Conclusion and Recommendations

The segmentation model provides a clear, actionable framework for targeted marketing by classifying customers based on their propensity to purchase, age, and estimated salary.

Marketing Recommendations:

Prioritize Clusters 1 & 3: These two groups account for the vast majority of purchases (86.4% and 83.3% purchase rates, respectively). Marketing spend should be heavily weighted towards these high-conversion segments.
Affluent Strategy (Cluster 1): Focus on quality, convenience, and status.
Value Strategy (Cluster 3): Focus on sales, bundled offers, and product utility to appeal to their price-conscious yet high-purchasing behavior. Given the high female majority, campaigns can be tailored accordingly.
Efficiency Savings (Cluster 2): Reduce spending on direct sales campaigns aimed at this segment, as their zero purchase rate indicates a low probability of conversion.

Thank you