

Project Report Template for

Customer Segmentation

(Mall Customers):

- ◆ **Title Page:**

- ◆ **Project Title:**

*Customer Segmentation using KMeans
(Mall Customers Dataset)*

- ◆ **NAME:** VISLAVATH HATHIRAM

- ◆ **Internship ID:** 5092

- ◆ **Internship Projects for Data Analytics
Internship at Skillbit Technologies**

- ◆ **Date:** Internship Start Date: 11/08/2025

- Internship End Date: 21/09/2025

- Task Submission Deadline: 21/09/2025

- Certificate Delivery Date: 23/09/2025

1. Introduction

Customer segmentation is the practice of dividing a customer base into groups of individuals with similar characteristics.

Companies use segmentation to:

- Improve marketing campaigns.
- Provide personalized offers.
- Increase customer retention and loyalty.

In this project, we applied KMeans clustering to the *Mall Customers dataset* to identify groups of customers based on Annual Income and Spending Score.

2. Objective

- To perform customer segmentation based on **Annual Income and Spending Score**.
- To identify customer groups with similar behavior.
- To provide business insights for targeted marketing strategies.

3. Dataset Description

- **Source:** Kaggle – Mall Customers dataset.
- **Rows:** 200 customers.
- **Features:**

1. CustomerID → Unique ID for each customer.
2. Gender → Male/Female.
3. Age → Customer age.
4. Annual Income (k\$) → Approx. yearly income.
5. Spending Score (1–100) → Rating assigned by the mall based on customer behavior & spending nature.

Why these features?

We mainly used **Annual Income & Spending Score** since they give the clearest segmentation of buying power & shopping habits.

4. Methodology (Workflow)

Step 1: Data Collection

- Load Mall Customers dataset.

Step 2: Data Preprocessing

- Handle missing values (if any).
- Select relevant features for clustering.

Step 3: Feature Selection

- Chose Annual Income and Spending Score for visualization and clustering.

Step 4: Applying K-Means Clustering

- Used Elbow Method to determine the optimal number of clusters.
- Applied KMeans algorithm with k=5.

Step 5: Visualization & Insights

- Created cluster visualization.
- Analyzed customer groups.

Workflow Diagram (Optional but Professional)

- A simple flowchart showing:
Data → Preprocessing → Clustering → Visualization → Insights → Business Strategy

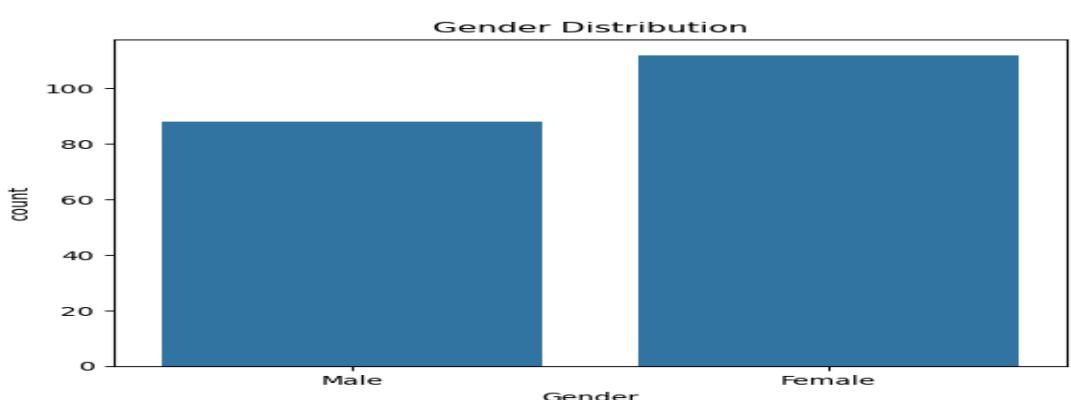
👉 This makes your methodology very clear in the report.

5. Diagrams & Visualizations

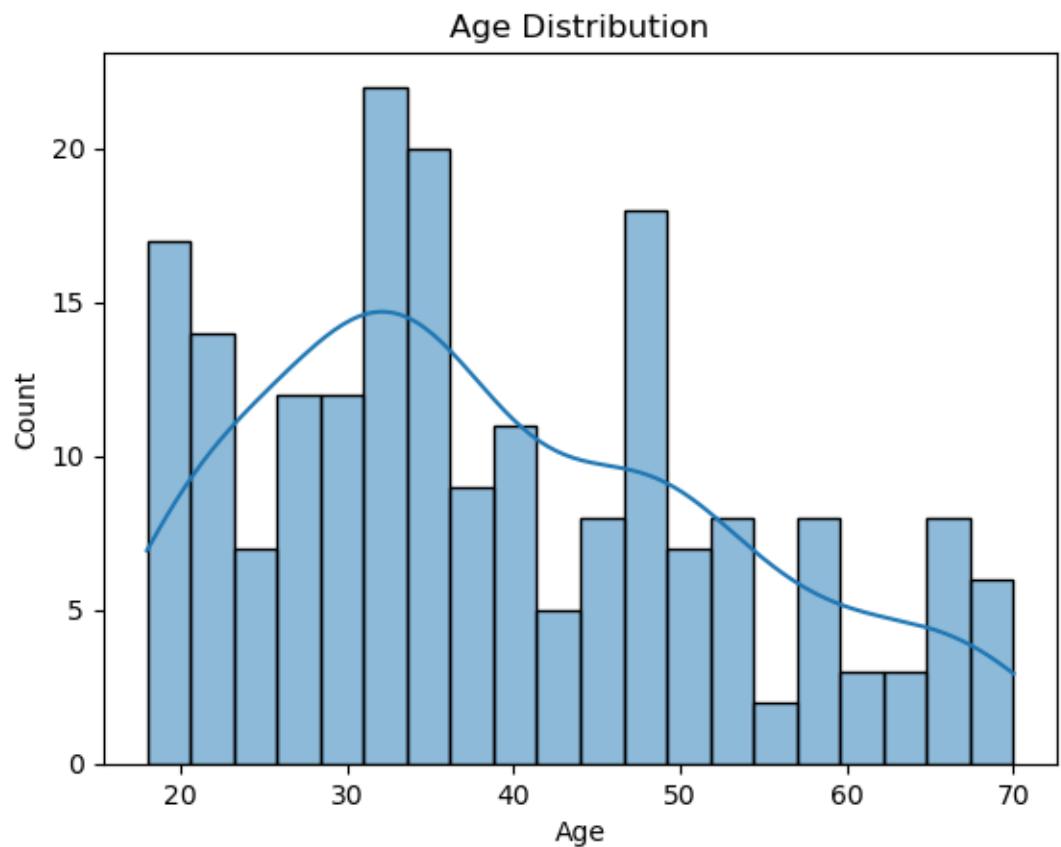
Recommended diagrams included in this project:

1. Dataset Overview

- Gender distribution (Bar chart)

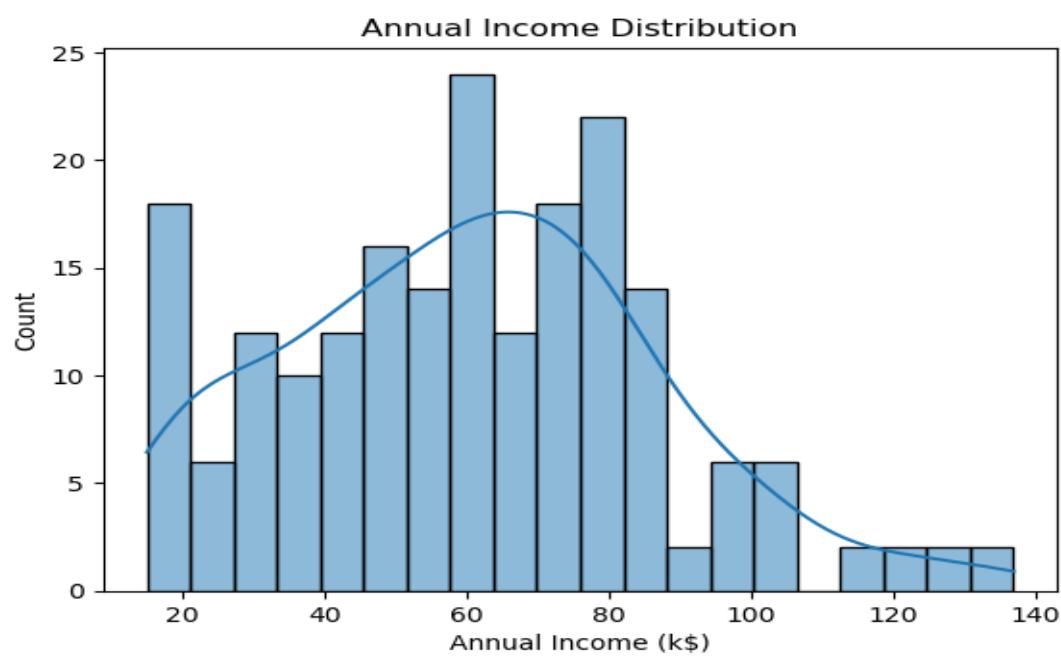


- Age distribution (Histogram)



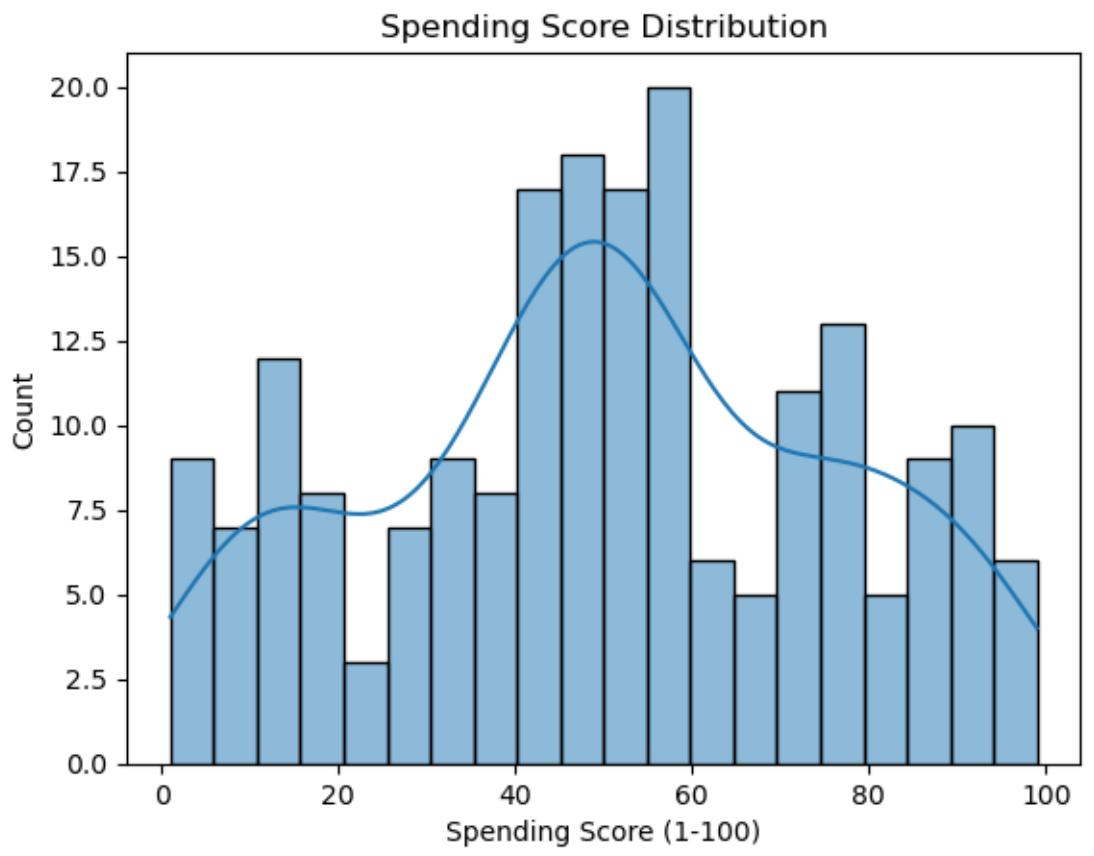
2. Feature Analysis

- Annual Income (Boxplot/Histogram)



◦

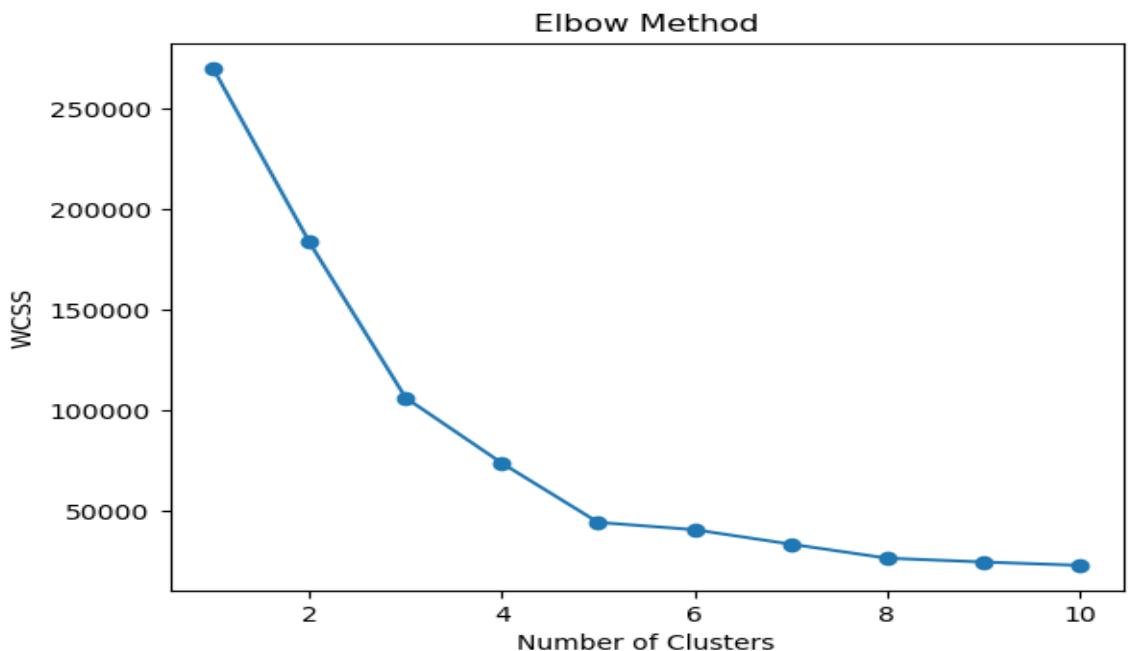
- Spending Score (Histogram)



○

3. Elbow Method

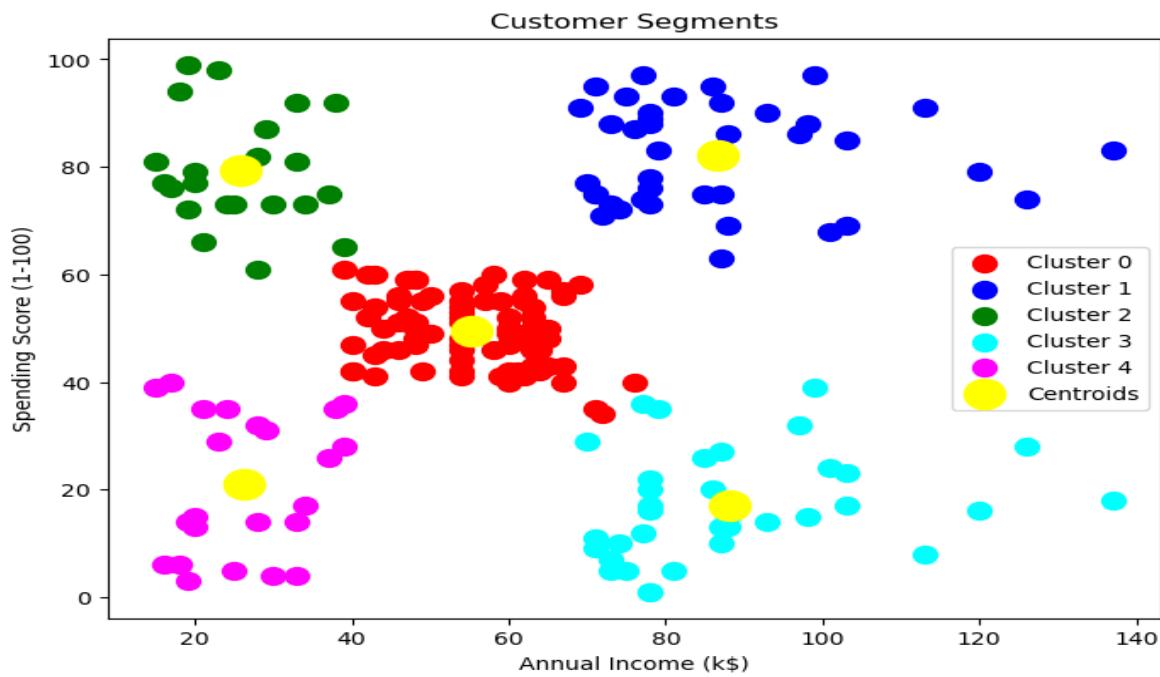
- Plot of WCSS vs k (to choose optimal clusters).



○

4. Cluster Visualization (Main Result)

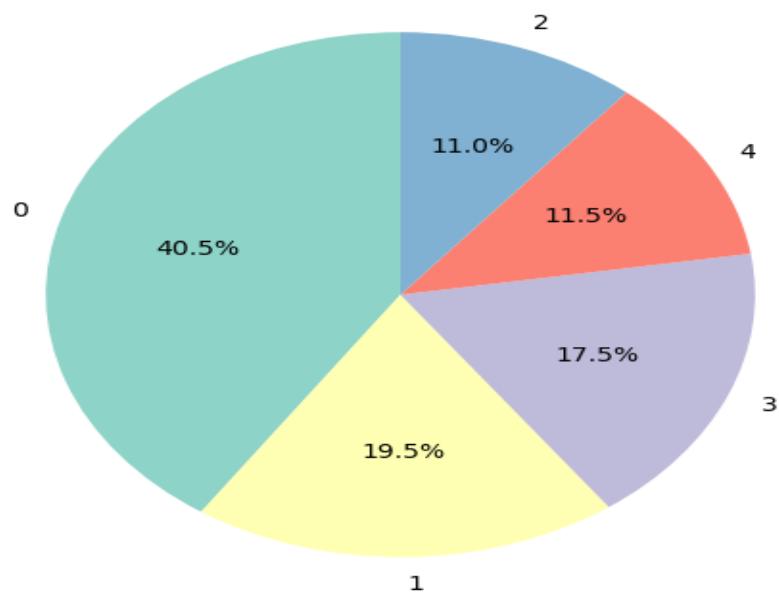
- 2D scatter plot of Annual Income vs Spending Score with 5 clusters.



5. Cluster Distribution

- Pie chart of customers per cluster.

Customer Distribution by Cluster



6.Workflow Diagram (Optional)

- Flow: Data → Preprocessing → KMeans → Clusters → Insights

6. Results

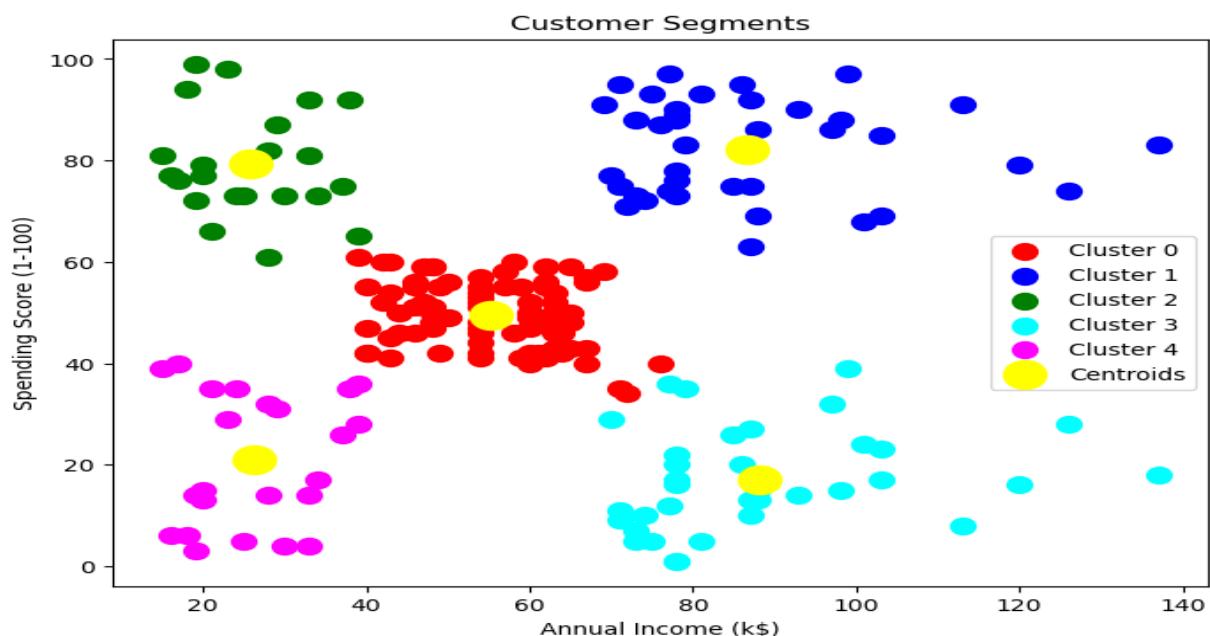
Cluster 0 (Careful Spenders): Low income, low spending.

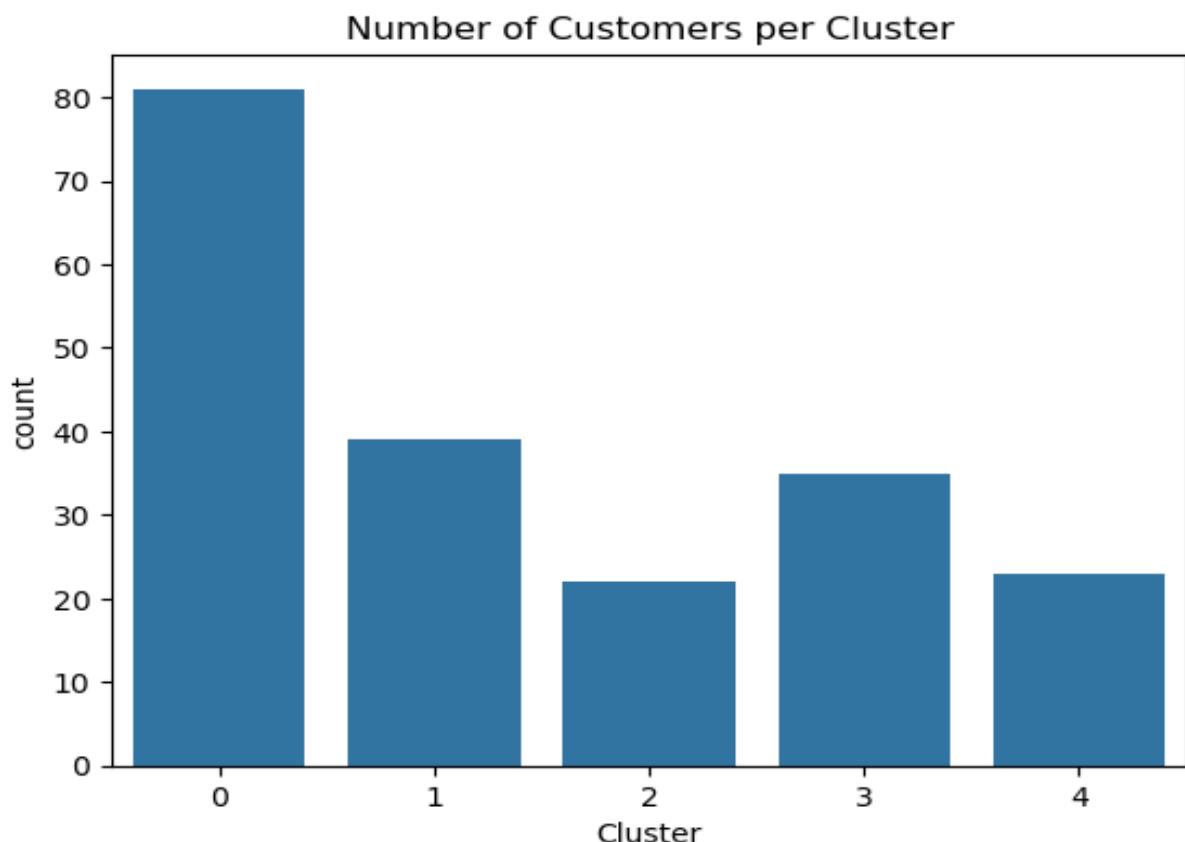
Cluster 1 (High Value Customers): High income, high spending – most profitable.

Cluster 2 (Impulsive Buyers): Low income, high spending – target with budget-friendly offers.

Cluster 3 (Cautious Wealthy): High income, low spending – potential for premium marketing.

Cluster 4 (Middle Class Customers): Average income & spending.





Visualization Outputs (Jupyter Notebook):

- Scatterplot of clusters in 2D.
- Pairplots showing relationship between age, income, and spending score.
- Elbow curve (for selecting k).

7. Business Impact

By identifying these clusters, a business can:

- Cluster 2: Reward with loyalty programs.

- Cluster 3: Offer discounts & bundle deals.
- Cluster 4: Target with luxury or premium services.
- Cluster 1: Reduce marketing spend (low ROI).

This helps the mall save costs and increase revenue.

8. Tools & Technologies

- **Programming Language:** Python
- **Libraries:** Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn
- **Environment:** Jupyter Notebook (Anaconda)

9. Future Scope

- Include more features (Age, Gender) for deeper analysis.
- Apply advanced clustering techniques (DBSCAN, Hierarchical Clustering).
- Integrate machine learning models for predictive customer segmentation.

10. Conclusion

This project successfully segmented mall customers into five groups using K-Means clustering. The insights can help businesses in designing targeted strategies to increase sales and customer satisfaction.

11. References

- **Kaggle Mall Dataset:**
[https://www.kaggle.com/datasets/shwetabh123/
mall-customers](https://www.kaggle.com/datasets/shwetabh123/mall-customers)
- Scikit-learn Documentation (KMeans, Silhouette
Score)
- Matplotlib & Seaborn Docs