

Customer Segmentation & Shopping Behavior Analysis

This project was conducted as a client project provided by DataMites, focusing on customer segmentation and shopping behavior analysis using retail purchase data from the U.S. market.

1. Objective of the Project

The objective of this project was to analyze customer purchasing patterns, identify product categories where discounts should be applied, study spending behavior based on age, season, and location, and segment customers into meaningful groups to propose targeted marketing strategies.

2. Dataset Description

The dataset was sourced from a client-hosted MySQL database and contains approximately 3900 records. It includes attributes such as customer age, gender, product category, purchase amount, season, location, discount usage, payment method, and previous purchase history.

	Customer_ID	Age	Gender	Item_Purchased	Category	Purchase_Amount_USD	Location	Size	Color	Season	Review_Rating	
▶	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Y
	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Y
	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Y
	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Y
	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Y
	6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9	Y
	7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Y
	8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	3.2	Y
	9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver	Summer	2.6	Y
	10	57	Male	Handbag	Accessories	31	Missouri	M	Pink	Spring	4.8	Y

3. Tools & Technologies Used

- MySQL for data storage and extraction
- Python for data analysis
- Pandas & NumPy for data manipulation
- Matplotlib / Seaborn for visualization
- Scikit-learn for clustering analysis

4. Data Extraction & Preparation

The data was extracted from the MySQL database using SQL queries and loaded into Python using Pandas. Data preprocessing steps included handling missing values, removing duplicates, encoding categorical variables, and scaling numerical features to prepare the data for clustering.

5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis was performed to understand spending patterns across different customer segments. Discount analysis helped identify categories that responded positively to promotional offers. Spending behavior was also analyzed based on age groups, seasonal trends, and geographic locations.

Purchase_Amount_USD		
Category	Discount_Applied	Purchase_Amount_USD
Accessories	No	60.889527
	Yes	58.489871
Clothing	No	60.223658
	Yes	59.752394
Footwear	No	59.079412
	Yes	61.799228
Outerwear	No	58.655556
	Yes	55.319444

dtype: float64

Purchase_Amount_USD	
Season	Purchase_Amount_USD
Fall	60018
Spring	58679
Summer	55777
Winter	58607

dtype: int64

6. Customer Segmentation

K-Means clustering was applied to segment customers based on age, purchase amount, and previous purchase history. Three distinct clusters were identified: high-value customers, discount-sensitive customers, and loyal customers.

	Age	Purchase_Amount_USD	Cluster
0	55	53	0
1	19	64	1
2	50	73	0
3	21	90	2
4	45	49	0

7. Business Insights & Marketing Strategy

Based on the clustering results, tailored marketing strategies were proposed. High-value customers were recommended premium offers, discount-sensitive customers were targeted with promotional deals, and loyal customers were encouraged through loyalty rewards and subscription programs.

Cluster	Age	Purchase_Amount_USD	Previous_Purchases	Review_Rating
0	59.698917	58.569012	28.451962	3.735792
1	36.332817	59.765480	11.018576	3.748994
2	32.469027	61.326549	37.684071	3.769558

8. Conclusion

This project demonstrates an end-to-end data analytics workflow, from data extraction to business insight generation. The findings provide actionable recommendations that can help improve customer engagement, optimize discount strategies, and increase overall revenue.