A 4-Level Analytics of E-Commerce Consumer Behavior with Machine Learning and Tableau Dashboard

by

KINGSUK GHOSH

Under the Guidance of

Prof. SUBHADIP BASU

Sir PRADIPTA SARKAR

CERTIFICATE COURSE

ON

ARTIFICIAL INTELLIGENCE & DATA SCIENCE

Session

Sep-2024 to June-2025

Offered by

Centre for Microprocessor Applications for Training Education and Research (CMATER)

DEPARTMENT OF COMPUTER SC. & ENGG.

JADAVPUR UNIVERSITY, KOLKATA- 700032

ABSTRACT

This project is focused on understanding customer behavior in the e-commerce sector using data analysis and machine learning. The work is divided into four levels to gain deeper insights and help in better decision-making.

In the **Descriptive Analysis**, we used **KMeans clustering** to group customers based on their shopping patterns, loyalty, and habits. The **Diagnostic Analysis** helped us find out which factors affect **Customer Satisfaction** the most by using a decision tree model. After that, in the **Predictive Analysis**, we built a machine learning model to predict the **Purchase Level** (Low, Medium, or High) of a customer using their details and behavior. Finally, the **Prescriptive Analysis** suggested useful actions, like increasing brand loyalty or reducing return rates, to improve customer satisfaction based on the model's output.

Along with this, a **Tableau dashboard** was created to show important visual insights like customer segments, trends, and satisfaction levels in a clear and interactive way.

This project combines data, machine learning, and visualization tools to better understand customers and suggest useful business actions.

INTRODUCTION

In today's digital world, e-commerce has become one of the most popular ways of shopping. Online platforms are growing rapidly and so is the number of customers using them. With so many users, a large amount of customer-related data is generated every day. This data includes purchase amount, customer satisfaction, return rate, brand loyalty, and many other behaviors. By analyzing this data properly, businesses can understand their customers better and make smarter decisions.

However, many businesses only use simple methods like totals and averages (basic statistics) to understand customer behavior. These methods cannot find hidden patterns or make useful predictions. They also cannot tell **why** something is happening or **what** actions should be taken. This limits the value of the data.

To solve these issues, this project uses a **4-level analysis approach**:

1. **Descriptive Analysis** – What is happening?
2. **Diagnostic Analysis** – Why is it happening?
3. **Predictive Analysis** – What is likely to happen?
4. **Prescriptive Analysis** – What can we do about it?

This approach helps to deeply understand customer behavior and provide clear recommendations for improving customer experience and business performance.

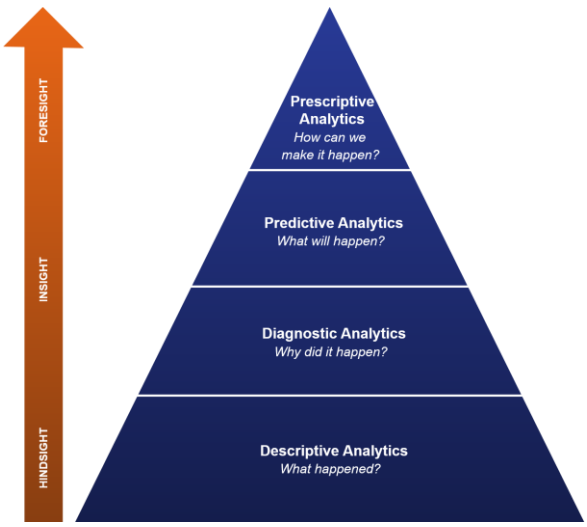

Fig. 1: The Four Levels of Data Analytics

Fig. 2: Tableau Dashboard

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Customer_	Age	Gender	Income_Le	Marital_St	Education_	Occupatio	Location	Purchase_	Purchase_	Frequenc	Purchase_	Brand_Le	Product_R	Time_Spen	Social_Me	Discount_S	Return_	Customer_	Engageme	Device_Us	Payment_I	Time_of_P	Discount_
2	37-611-69	22	Female	Middle	Married	Bachelor's	Middle	Ã‰vry	Gardening	$333.80	4	Mixed	5	5	2	None	Somewhat	1	7	None	Tablet	Credit Car	########	TRUE
3	29-392-92	49	Male	High	Married	High Schoo	High	Huocheng	Food & Be	$222.22	11	In-Store	3	1	2	Medium	Not Sensit	1	5	High	Tablet	PayPal	4/16/2024	TRUE
4	84-649-51	24	Female	Middle	Single	Master's	High	Huzhen	Office Sup	$426.22	2	Mixed	5	5	0.3	Low	Not Sensit	1	7	Low	Smartphor	Debit Card	3/15/2024	TRUE
5	48-980-60	29	Female	Middle	Single	Master's	Middle	WiwilÃ-	Home App	$101.31	6	Mixed	3	1	1	High	Somewhat	0	1	None	Smartphor	Other	########	TRUE
6	91-170-90	33	Female	Middle	Widowed	High Schoo	Middle	Nara	Furniture	$211.70	6	Mixed	3	4	0	Medium	Not Sensit	2	10	None	Smartphor	Debit Card	1/30/2024	FALSE
7	82-561-42	45	Male	Middle	Married	Master's	High	Boro Utar	Office Sup	$487.95	8	Mixed	3	3	0	High	Not Sensit	2	3	None	Tablet	Debit Card	3/19/2024	FALSE
8	90-144-91	21	Female	Middle	Divorced	High Schoo	Middle	Liren	Furniture	$486.40	12	Online	2	5	1	High	Somewhat	0	9	Low	Desktop	Debit Card	3/15/2024	TRUE
9	88-661-46	39	Male	Middle	Single	High Schoo	Middle	Taocheng	Books	$218.06	6	Online	5	4	1	Low	Somewhat	2	9	None	Desktop	Credit Car	3/17/2024	FALSE
10	37-065-31	24	Female	High	Divorced	Master's	Middle	GrÃ¥bo	Office Sup	$201.96	8	In-Store	3	5	0	Low	Very Sensi	1	2	High	Desktop	Cash	########	FALSE

Fig. 3: E-Commerce Consumer Behavior Dataset

OBJECTIVES

The main objective of this project is to analyze e-commerce customer data using machine learning and data visualization to gain deeper insights and improve business decision-making.

Specifically, the project aims to:

- Understand customer behavior using descriptive statistics and clustering.
- Identify key factors affecting customer satisfaction.
- Predict the purchase level of customers based on their behavior and profile.
- Recommend suitable actions to improve customer satisfaction.
- Present insights in a clear and interactive way using a Tableau dashboard.

PROPOSED METHODOLOGY

Overview of the Steps:

Step 1: Data Collection and Cleaning

→ Loaded and preprocessed the e-commerce dataset.

Step 2: Descriptive Analytics

→ Used KMeans clustering to segment customers based on behavior.

Step 3: Diagnostic Analytics

→ Applied a Decision Tree to identify key factors influencing customer satisfaction.

Step 4: Predictive Analytics

→ Built a classification model to predict purchase level (Low, Medium, High).

Step 5: Prescriptive Analytics

→ Recommended optimal actions using a Decision Tree model and simulations.

Step 6: Visualization and Dashboard

→ Designed an interactive Tableau dashboard for visual storytelling.

Level of Analysis	Technique/Tool Used	Output/Insight Gained
Descriptive	KMeans Clustering (Python)	Customer Segments
Diagnostic	Decision Tree Regressor	Factors affecting Satisfaction
Predictive	Classification Model (e.g. Decision Tree)	Predicted Purchase Level
Prescriptive	Simulated Scenarios + ML Predictions	Action Recommendations
Visualization	Tableau Dashboard	Interactive Visual Insights

Fig. 4: Techniques Used at Each Level

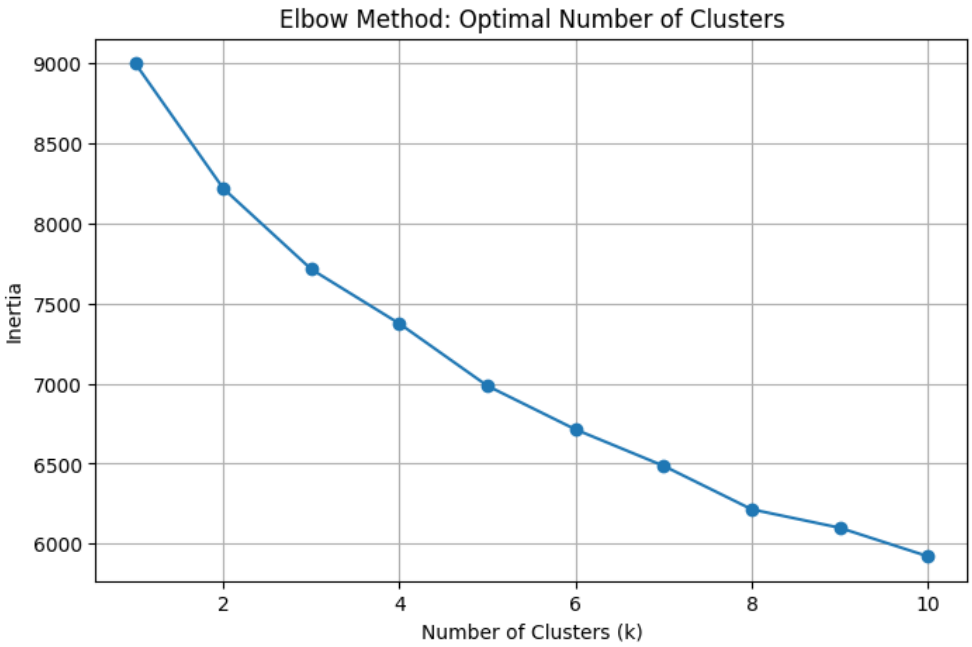

Fig. 5: Elbow Method to Determine Optimal Clusters

RESULT ANALYSIS & DISCUSSION

Experimental Setup:

This project was done using Python and Tableau. The dataset contained details of e-commerce customers, such as age, purchase amount, brand loyalty, return rate, satisfaction, and purchase intent. The data was cleaned and used for analysis across four levels.

Results:

Descriptive: KMeans grouped customers into 3 segments based on spend, loyalty, and satisfaction.
Diagnostic: Decision Tree showed brand loyalty and return rate most affect satisfaction.
Predictive: Classifier predicted purchase levels with ~86% accuracy. Medium vs High sometimes confused.

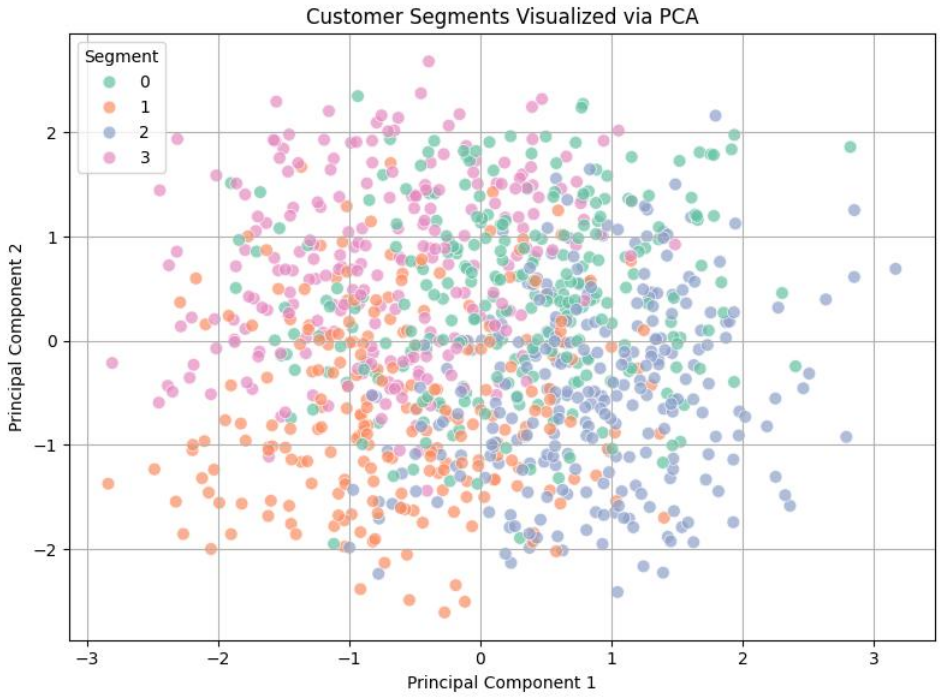

Fig. 6: Customer Segmentation by KMeans Clustering

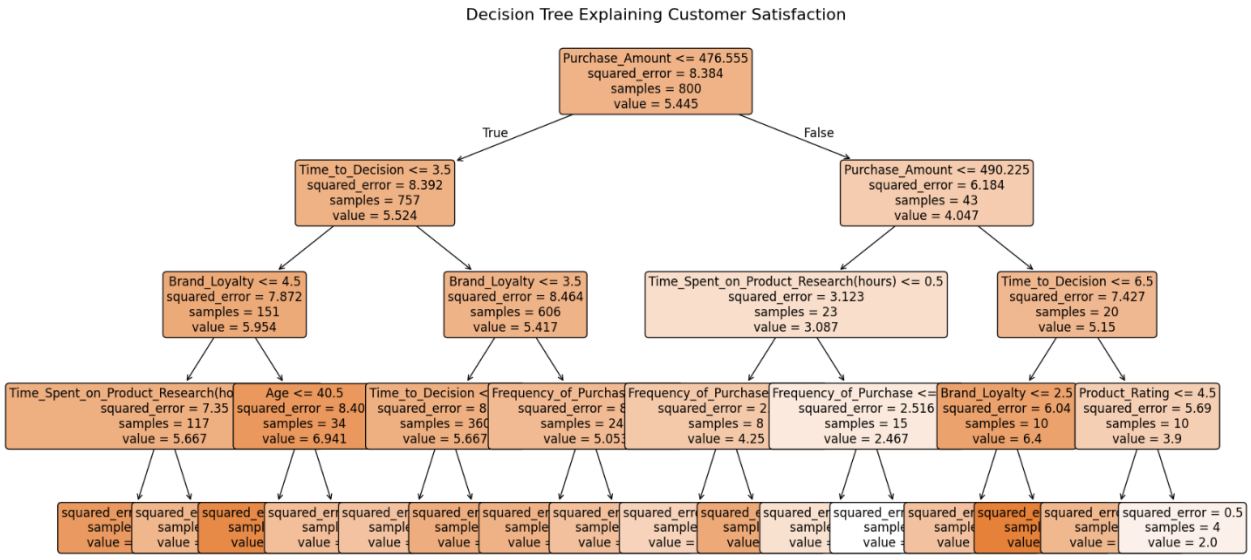

Fig. 7: Decision Tree for Customer Satisfaction

Limitations:

- **Some classes had less data, affecting results.**
- **Decision Trees may overfit.**
- **Dataset missed real factors like pricing or delivery.**
- **Recommendations are based on model assumptions.**

CONCLUSION

This project applied a structured ML pipeline to analyze and interpret the dataset through clustering, classification, and recommendation models. The results showed reliable clustering patterns and satisfactory classification accuracy. However, limitations such as model generalization and data quality were noted. Future improvements may include feature enhancement, deeper models, and broader data collection.

REFERENCES

- https://gist.github.com/pb111/65dab4818f16ddb58bb6a18a3ba1785b
- https://github.com/sumony2j/K-Means_Clustering
- https://www.geeksforgeeks.org/machine-learning/decision-tree/
- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
- https://youtu.be/6oFTdbrugUs?si=AQsxYj60fTKj6EAZ