



Market Segmentation

A MACHINE LEARNING
MARKET SEGMENTATION PROJECT

Projected by : Bhagirathsinh jadeja



Introduction to Market Segmentation:

Content:

- Businesses may not satisfy all customers every time.
- People have different preferences; one product rarely satisfies everyone.
- Target marketing strategy involves dividing the market into segments and developing products/services for these segments.

Overview of the Dataset

1

Demographic Information:

- The dataset includes key demographic details such as age, gender, income, and occupation of the customers. These factors are crucial for understanding the composition of the customer base and identifying distinct market segments.

2

Purchase History:

- This data encompasses transaction records, including the frequency, type, and number of purchases made by each customer. Analyzing purchase history helps in identifying buying patterns and preferences, which are vital for effective segmentation.

3

Source of the Data & Number of Samples and Features:

- The dataset is sourced from the Kaggle community's customer segmentation data. It comprises a total of 8950 samples(rows) with 18 features(columns) each, providing a comprehensive view of customer demographics and purchasing behavior.

Datasets

Following is the Data Dictionary for Credit Card dataset:

1. **CUSTID** : Identification of Credit Cardholder (Categorical)
2. **BALANCE** : Balance amount left in their account to make purchases
3. **BALANCEFREQUENCY** : How frequently the Balance is updated, score between 0 and 1
(1 = frequently updated, 0 = not frequently updated)
4. **PURCHASES** : Amount of purchases made from the account
5. **ONEOFFPURCHASES** : Maximum purchase amount did in one-go
6. **INSTALLMENTSPURCHASES** : Amount of purchase done in installment
7. **CASH ADVANCE** : Cash in advance given by the user
8. **PURCHASESFREQUENCY** : How frequently the Purchases are being made score between 0 and 1
(1 = frequently purchased, 0 = not frequently purchased)
9. **ONEOFFPURCHASESFREQUENCY** : How frequently Purchases are happening in one-go
(1 = frequently purchased, 0 = not frequently purchased)
10. **PURCHASESINSTALLMENTSFREQUENCY** : How frequently purchases in installments are being done
(1 = frequently done, 0 = not frequently done)
11. **CASHADVANCEFREQUENCY** : How frequently the cash in advance being paid
12. **CASHADVANCETRX** : Number of Transactions made with "Cash in Advanced"
13. **PURCHASESTRX** : Number of purchase transactions made
14. **CREDIT LIMIT** : Limit of Credit Card for user
15. **PAYMENTS** : Amount of Payment done by the user
16. **MINIMUM_PAYMENTS** : Minimum amount of payments made by the user
17. **PRCFULLPAYMENT** : Percent of full payment paid by the user
18. **TENURE** : Tenure of credit card service for user



```
mirror_mod = modifier_ob.  
#set mirror object to mirror  
mirror_mod.mirror_object =  
operation == "MIRROR_X":  
mirror_mod.use_x = True  
mirror_mod.use_y = False  
mirror_mod.use_z = False  
operation == "MIRROR_Y":  
mirror_mod.use_x = False  
mirror_mod.use_y = True  
mirror_mod.use_z = False  
operation == "MIRROR_Z":  
mirror_mod.use_x = False  
mirror_mod.use_y = False  
mirror_mod.use_z = True  
  
#selection at the end -add  
mirror_ob.select= 1  
modifier_ob.select=1  
context.scene.objects.active  
("Selected" + str(modifier  
mirror_ob.select = 0  
= bpy.context.selected_obj  
data.objects[one.name].se  
print("please select exactly  
  
--- OPERATOR CLASSES ---  
  
types.Operator):  
X mirror to the selected  
object.mirror_mirror_x"  
mirror X"  
  
context):  
context.active_object is not
```

Data Collection: Collected data from Kaggle Community

<https://www.kaggle.com/datasets/mahnazarjmand/customer-segmentation>

TECH TOOLKITS USED

Python is the programming language which is used to code the project. Various algorithms used are coded in Python



Streamlit is an open-source framework for Machine learning and Data Science. It is used for prototype and deployment purpose of the models.



Jupyter notebook is the effective IDE used for coding in Python. It is very easy to use and widely used over the industry.

Algorithms used in this Project



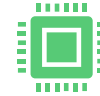
Project Steps



Data Collection: Gather data relevant to the segmentation bases.



Exploratory Data Analysis (EDA): Understand the dataset, identify patterns, and visualize data.



Feature Engineering: Create new features or modify existing ones to better represent the data.



Segmentation Model: Apply machine learning algorithms to segment the market.



Evaluation and Interpretation: Evaluate the segmentation model and interpret the results to derive actionable insights



Deployment : To predict the cluster or Groups



EDA

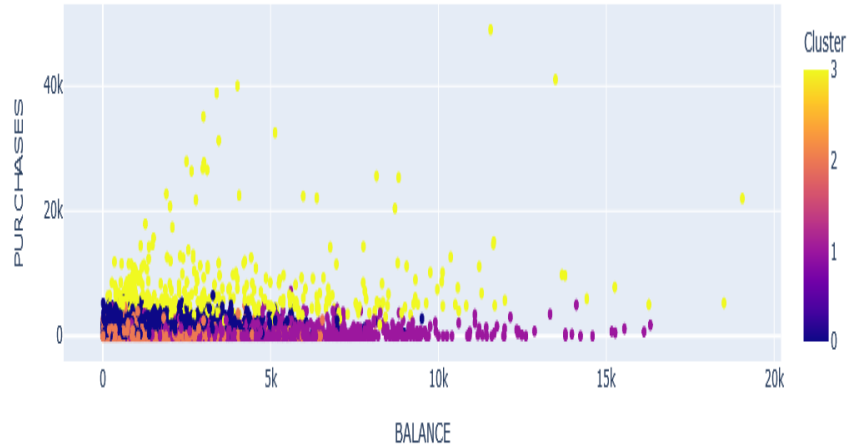
(Exploratory Data Analysis)

- **Handling Missing Values:**
 - Missing values in the dataset were filled using the mean value for 313 transactions. This method ensures that the dataset remains robust and accurate without losing significant information.
- **Normalization:**
 - Standard Scaler techniques were applied to normalize the data. This step scales the features to have a mean of zero and a standard deviation of one, which is essential for improving the performance and convergence speed of machine learning algorithms.

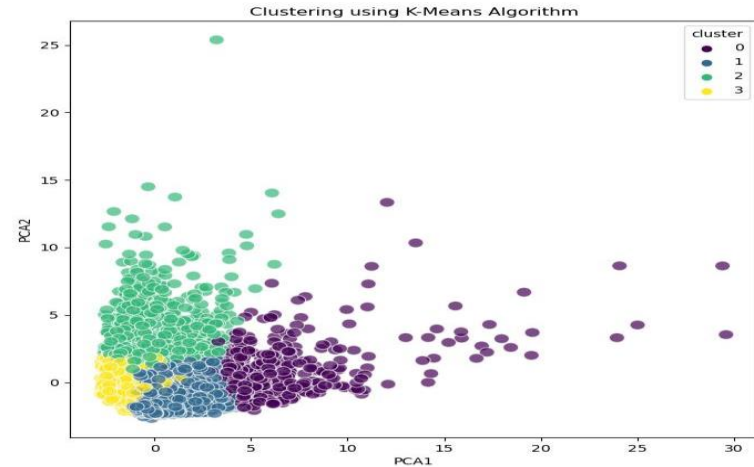
Applying PCA & K-means Clustering Algo result

Clusters of Customers

Before

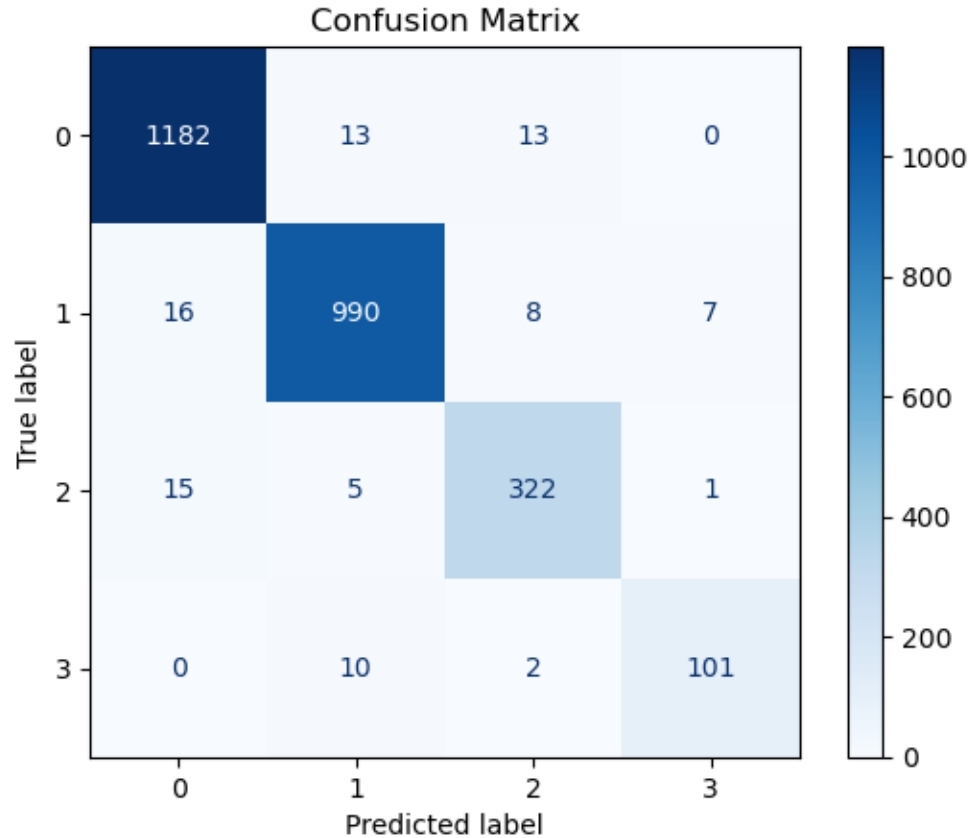


After



Accuracy

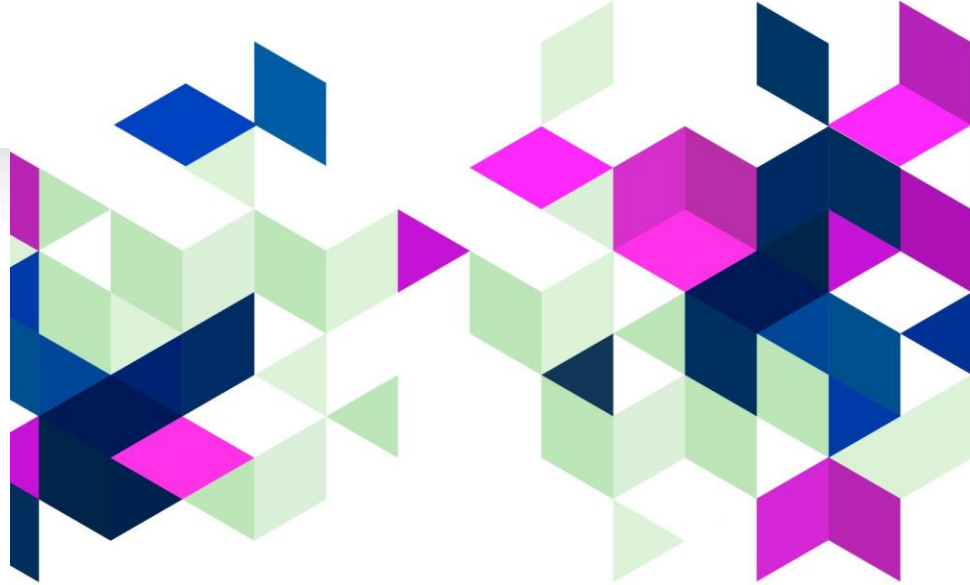
- Testing All Algo We used model **Gradient Boosting Classifier** algorithm which gave us 97% accuracy



0.9664804469273743 % Accuracy

Deployment

We have created a Streamlit Application based on this clustering technique, where we are taking the customer details & identifying which cluster the customer belongs to.



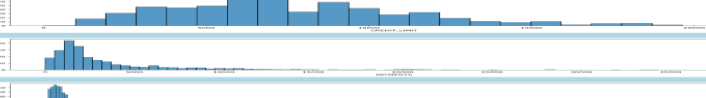
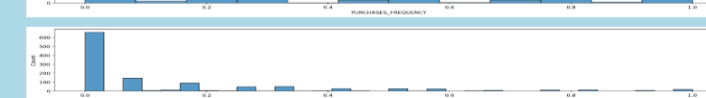
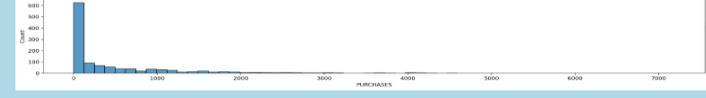
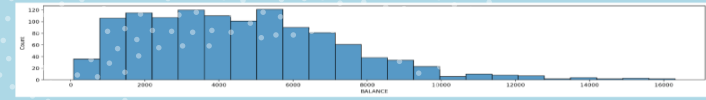
Market Segment Prediction

Balance	<input type="text" value="0.000000"/>	-- *
Balance Frequency	<input type="text" value="0.000000"/>	-- *
Purchases	<input type="text" value="0.00"/>	-- *
OneOff Purchases	<input type="text" value="0.00"/>	-- *
Installments Purchases	<input type="text" value="0.00"/>	-- *
Cash Advance	<input type="text" value="0.000000"/>	-- *
Purchases Frequency	<input type="text" value="0.000000"/>	-- *
OneOff Purchases Frequency	<input type="text" value="0.000000"/>	-- *
Purchases Installments Frequency	<input type="text" value="0.000000"/>	-- *
Cash Advance Frequency	<input type="text" value="0.000000"/>	-- *
Cash Advance Tix	<input type="text" value="0"/>	-- *
Purchases Tix	<input type="text" value="0"/>	-- *
Credit Limit	<input type="text" value="0.0"/>	-- *
Payments	<input type="text" value="0.000000"/>	-- *
Minimum Payments	<input type="text" value="0.000000"/>	-- *
PRC Full Payment	<input type="text" value="0.000000"/>	-- *
Tenure	<input type="text" value="0"/>	-- *
<input type="button" value="Submit"/>		

Streamlit library input

Output Predicts the result Data Belongs to Cluster 1

Data Belongs to Cluster 1



THANK
YOU