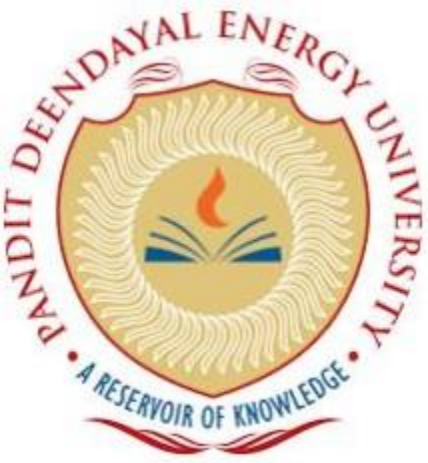
ASSIGNMENT-5

Advanced Python Lab

Name: Kashyap Maisuria

Roll no: 21BCP442D

DIVISION 3 (Group 6)



COMPUTER ENGINEERING

School of Technology,

Pandit Deendayal Energy University

❖ Data Analysis

- You are working for a large e-commerce platform, and your task is to perform customer segmentation based on their shopping behavior. You have access to a dataset containing information about customer transactions. The dataset is in a CSV format and contains the following columns:

- This problem requires you to use Pandas for data manipulation, NumPy for numerical operations, and potentially machine learning libraries for customer segmentation.

- It showcases the power of data analysis and segmentation for making data-driven decisions in e-commerce.

DataSet for the above analysis-
https://www.kaggle.com/datasets/puneetbhaya/online-retail/

1. CustomerID: Unique identifier for each customer.

2. TotalAmountSpent: The total amount spent by each customer on the platform.

3. TotalItemsPurchased: The total number of items purchased by each customer.

4. LastPurchaseDate: The date of the customer's most recent purchase.

5. AveragePurchaseValue: The average value of each customer's purchases.

Using NumPy and Pandas, your goal is to perform the following tasks:

1. Data Loading: Load the dataset into a Pandas DataFrame for analysis.

2. Data Cleaning: Check for missing values, duplicates, or any inconsistencies in the data. If found, clean the data appropriately.

3. Descriptive Statistics: Calculate basic statistics such as mean, median, and standard deviation of TotalAmountSpent and TotalItemsPurchased.

4. Customer Segmentation: Divide the customers into segments based on their shopping behavior. You can use techniques like K-means clustering or any other method you prefer. For example, you might create segments like "High Spenders," "Frequent Shoppers," and "Inactive Customers."

5. Visualization: Create visualizations (e.g., scatter plots, bar charts) to represent the different customer segments you've identified.

6. Customer Insights: Provide insights into each customer segment. What distinguishes one segment from another? How can the e-commerce platform tailor its marketing strategies for each segment?

7. Customer Engagement Recommendations: Based on your analysis, provide recommendations for the e-commerce platform on how to engage with each customer segment more effectively. For example, should they offer discounts, provide personalized product recommendations, or run targeted marketing campaigns?

Code: -

```
→import pandas as pd
 from sklearn.cluster import KMeans
 import matplotlib.pyplot as plt

 # Step 1: Data Loading
 df = pd.read_csv('/home/kali/Online Retail.csv')

 # Step 2: Data Cleaning
 print(df.isnull().sum())
 df = df.dropna()
 df = df.drop_duplicates()
 df = df[(df['Quantity'] > 0) & (df['UnitPrice'] > 0)]

 # Calculate Total Amount Spent
 df['TotalAmountSpent'] = df['Quantity'] * df['UnitPrice']


 # Step 3: Descriptive Statistics
 mean_amount_spent = df['TotalAmountSpent'].mean()
 median_amount_spent = df['TotalAmountSpent'].median()
 std_amount_spent = df['TotalAmountSpent'].std()

 mean_items_purchased = df['Quantity'].mean()
 median_items_purchased = df['Quantity'].median()
 std_items_purchased = df['Quantity'].std()

 print(f"Mean Total Amount Spent: {mean_amount_spent}")
 print(f"Median Total Amount Spent: {median_amount_spent}")
 print(f"Standard Deviation of Total Amount Spent: {std_amount_spent}")
```

```
print(f"Mean Total Items Purchased: {mean_items_purchased}")
print(f"Median Total Items Purchased: {median_items_purchased}")
print(f"Standard Deviation of Total Items Purchased: {std_items_purchased}")

# Step 4: Customer Segmentation
X = df[['TotalAmountSpent', 'Quantity']]
num_clusters = 3
kmeans = KMeans(n_clusters=num_clusters, random_state=0).fit(X)
df['Cluster'] = kmeans.labels_


# Step 5: Visualization
plt.scatter(df['TotalAmountSpent'], df['Quantity'], c=df['Cluster'], cmap='viridis')
plt.xlabel('Total Amount Spent')
plt.ylabel('Total Items Purchased')
plt.title('Customer Segmentation')
plt.show()

# Step 6: Customer Insights
cluster_stats = df.groupby('Cluster')[['TotalAmountSpent', 'Quantity']].mean()
print(cluster_stats)

# Step 7: Customer Engagement Recommendations
for cluster in range(num_clusters):
    cluster_data = df[df['Cluster'] == cluster]
```

Output :

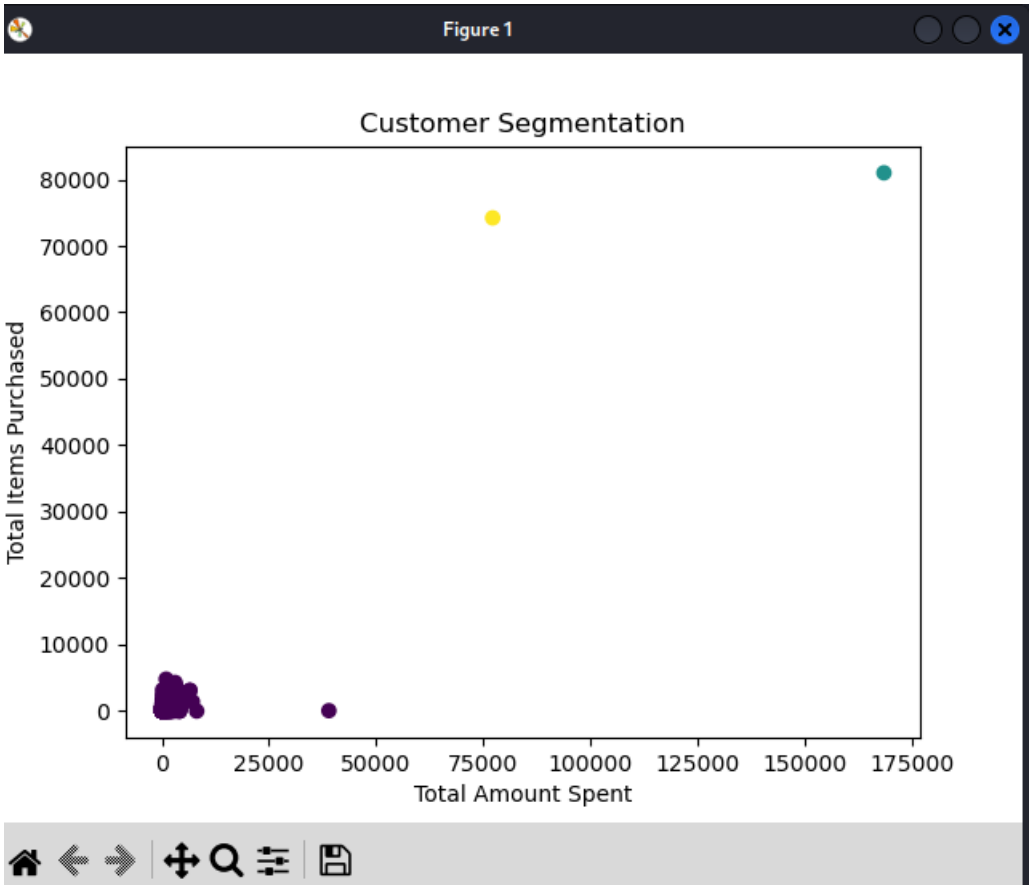
1. Online Retail.csv

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/2010 8:26	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/2010 8:26	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNION JACK	6	12/1/2010 8:28	1.85	17850	United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	12/1/2010 8:28	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/2010 8:34	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/1/2010 8:34	2.1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/1/2010 8:34	2.1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	12/1/2010 8:34	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MUG COSY	6	12/1/2010 8:34	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/1/2010 8:34	4.25	13047	United Kingdom

```
┌──(kali㉿kali)-[~/Desktop/pythn/pydata_analysis]
└─$ python3 pr5.py
InvoiceNo            0
StockCode            0
Description       1454
Quantity             0
InvoiceDate          0
UnitPrice            0
CustomerID      135080
Country              0
dtype: int64
```

```
Mean Total Amount Spent: 22.6314997351614
Median Total Amount Spent: 12.45
```

```
Standard Deviation of Total Amount Spent: 311.09922433483325
Mean Total Items Purchased: 13.1197019547126
Median Total Items Purchased: 6.0
Standard Deviation of Total Items Purchased: 180.49283198920438
```



Cluster		
0	22.00605	12.724521
1	168469.60000	80995.000000
2	77183.60000	74215.000000