EXPERINMENT-31

31. Scenario: You work as a data scientist for an e-commerce company that sells a wide range of products online. The company collects vast amounts of data about its customers, including their purchase history, browsing behavior, demographics, and more. The marketing team wants to understand their customer base better and improve their targeted marketing strategies. They have asked you to perform customer segmentation using clustering techniques to identify distinct groups
of customers with similar characteristics.
Question: Your task is to use Python and clustering algorithms to segment the customers into different groups based on their behavior and characteristics. The marketing team will use these segments to tailor their marketing campaigns and promotions effectively.

Code:

```
import pandas as pd
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
file = input("Enter CSV file name (e.g., all_customers.csv): ")
data = pd.read_csv(file)
print("\nColumns found in dataset:")
print(list(data.columns))
feature_input = input("\nEnter feature columns for clustering (comma separated): ")
features = [f.strip() for f in feature_input.split(",")]
X = data[features]
num_cols = X.select_dtypes(include=['int64', 'float64']).columns.tolist()
cat_cols = X.select_dtypes(include=['object']).columns.tolist()
print("\nNumeric columns:", num_cols)
print("Categorical columns:", cat_cols)
preprocessor = ColumnTransformer(
    transformers=[
        ("num", StandardScaler(), num_cols),
        ("cat", OneHotEncoder(handle_unknown="ignore"), cat_cols)
    ]
)
k = int(input("\nEnter number of clusters (k): "))
model = Pipeline([
```

```python
        ("pre", preprocessor),
        ("kmeans", KMeans(n_clusters=k, random_state=0))
])
model.fit(X)
labels = model.named_steps["kmeans"].labels_
data["segment"] = labels
print("\nCluster counts:")
print(data["segment"].value_counts())
pca = PCA(n_components=2)
X_preprocessed = model.named_steps["pre"].transform(X)
pca_result = pca.fit_transform(X_preprocessed)
plt.figure(figsize=(8, 6))
plt.scatter(pca_result[:, 0], pca_result[:, 1], c=labels, cmap="tab10", s=40)
plt.title("Customer Segments (PCA 2D Visualization)")
plt.xlabel("PCA Component 1")
plt.ylabel("PCA Component 2")
plt.grid(True, linestyle="--", alpha=0.4)
plt.show()
data.to_csv("customers_segmented.csv", index=False)
print("\nSegmented customer data saved as: customers_segmented.csv")
```

Output:

```
PS C:\Users\karan\OneDrive\Desktop\New folder (2)> python 31.py
Enter CSV file name (e.g., all_customers.csv): all_customers.csv

Columns found in dataset:
['age', 'annual_spend', 'visits', 'income_level', 'city_category']

Enter feature columns for clustering (comma separated): age, annual_spend, visits, city_category

Numeric columns: ['age', 'annual_spend', 'visits']
Categorical columns: ['city_category']
```

```
Enter number of clusters (k): 4

Cluster counts:
segment
2    82
0    74
3    72
1    72
Name: count, dtype: int64
```

Customer Segments (PCA 2D Visualization)