**Data segmentation:**

Data segmentation is a process of breaking down a dataset into discrete groups according to specific standards or attributes .

Benefits:

* Improves model performance
* It is easier to create strategies and solutions over certain segments.

Techniques :

* Supervised segmentation
* Unsupervised segmentation
* Semi-supervised segmentation

Like model training , in the process of segmenting we have to select relevant features.

The process is like :

1)Preprocessing the data.

2)Selecting relevant features

3)Standardizing or normalizing the features

4) fitting features to one appropriate segmentation model

5)Record those predictions from above model in separate column

6)cluster analysis is based on segments

**Example program:**

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

import matplotlib.pyplot as plt

import seaborn as sns

# Select relevant features

features = df[['tenure', 'MonthlyCharges', 'Contract\_Month-to-month', 'Contract\_One year', 'Contract\_Two year']]

# Normalize the data

scaler = StandardScaler()

features\_scaled = scaler.fit\_transform(features)

features\_scaled

n\_clusters = 3 # Example: You can choose a different number of clusters

kmeans = KMeans(n\_clusters=n\_clusters, random\_state=42)

df['Segment'] = kmeans.fit\_predict(features\_scaled)

cluster\_analysis = df.groupby('Segment')[['tenure', 'MonthlyCharges']].mean()

print(cluster\_analysis)

Output:

Segment tenure MonthlyCharges

0 56.735103 60.770413

1 18.036645 66.398490

2 42.044807 65.048608