Customer segmentation Phase-4

Feature Engineering :

Feature engineering is the process that takes raw data and transforms it into features that can be used to create a predictive model using machine learning or statistical modeling, such as deep learning. Importing a dataset and libraries like numpy, pandas, scipy etc...

K-Means Clustering :

```
import numpy as
np  import pandas
as pd
import matplotlib.pyplot as
plt import seaborn as sns
dataset =
pd.read_csv('/kaggle/input/mall
-
customers/Mall_Customers.cs
vl) dataset.head()
```

Output :

| CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40

Visualization Program:

```python
kmeansmodel = KMeans(n_clusters =
5 , init = "k-means++",random_stat
e=0)
y_kmeans = kmeansmodel.fit_predict
(x)
```

```python
plt.scatter(x[y_kmeans==0,0],x[y_k
means==0,1],s=100,c="red",label =
"cluster 1")
plt.scatter(x[y_kmeans==1,0],x[y_k
means==1,1],s=100,c="blue",label =
"cluster 2")
plt.scatter(x[y_kmeans==2,0],x[y_k
means==2,1],s=100,c="green",label
= "cluster 3")
plt.scatter(x[y_kmeans==3,0],x[y_k
means==3,1],s=100,c="cyan",label =
"cluster 4")
plt.scatter(x[y_kmeans==4,0],x[y_k
means==4,1],s=100,c="magenta",labe
l = "cluster 5")
```

```python
plt.scatter(kmeans.cluster_centers
_[:,0],kmeans.cluster_centers_[:,
1],s=300,c="yellow",label="centroi
ds")
plt.title("clusters of customers")
plt.xlabel("Yıllık gelir")
plt.ylabel=("harcama skoru")
plt.legend()
plt.show()
```

Output :

clusters of customers
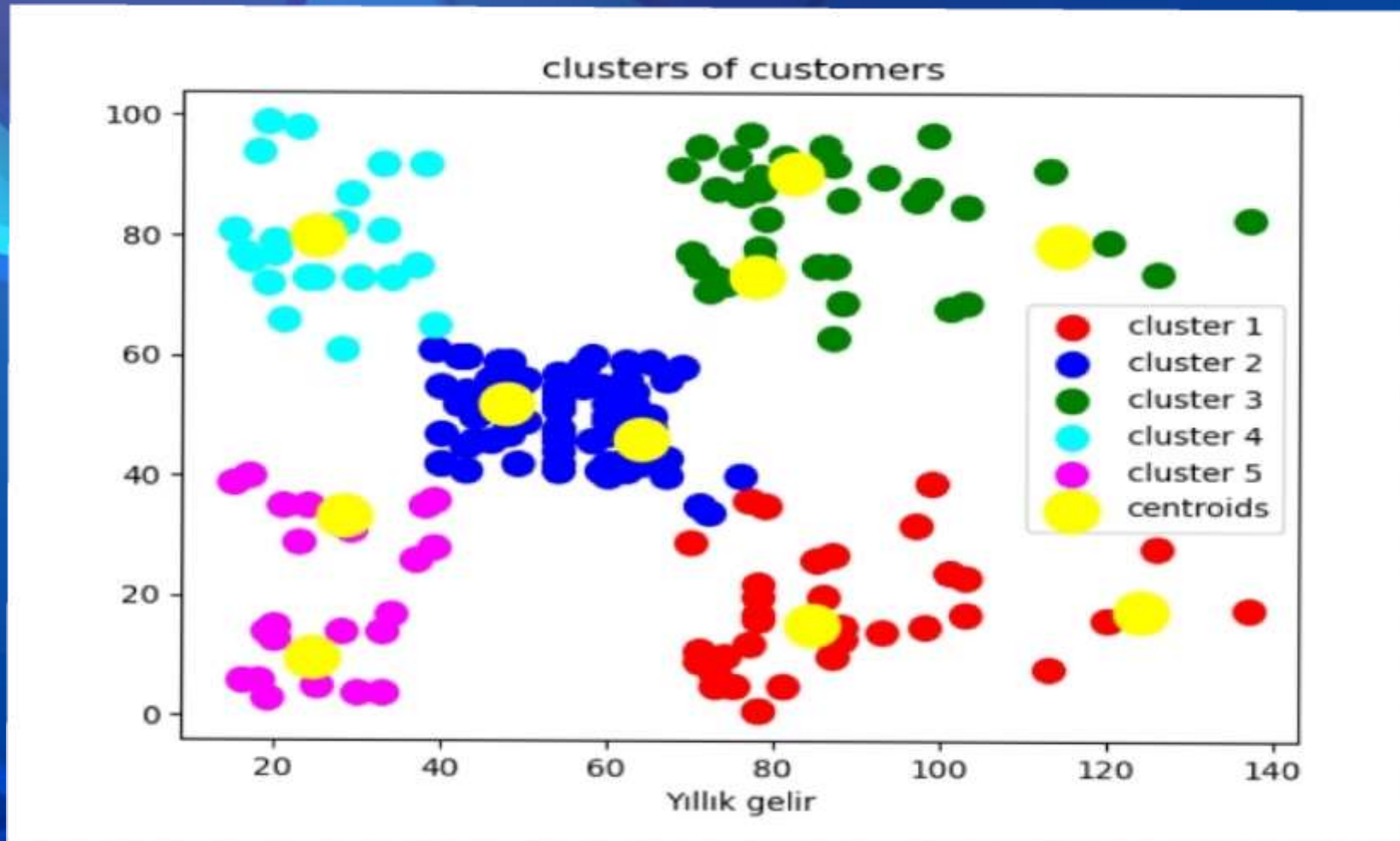
Interpretation :

Customer Segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer Segmentation can be a powerful means to identify unsatisfied customer needs. Using the above data companies can then outperform the competition by developing uniquely appealing products and services.

K Means Clustering Algorithm

1. Specify number of clusters K.

2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.

3. Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

Conclusion :

Using K-Means Clustering Method we get best solution for this project Customer segmentation