PADALA NAVYANTH REDDY (21BCE3372)

In [34]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [35]:
```
df=pd.read_csv("Mall_Customers.csv")
df.head()
```

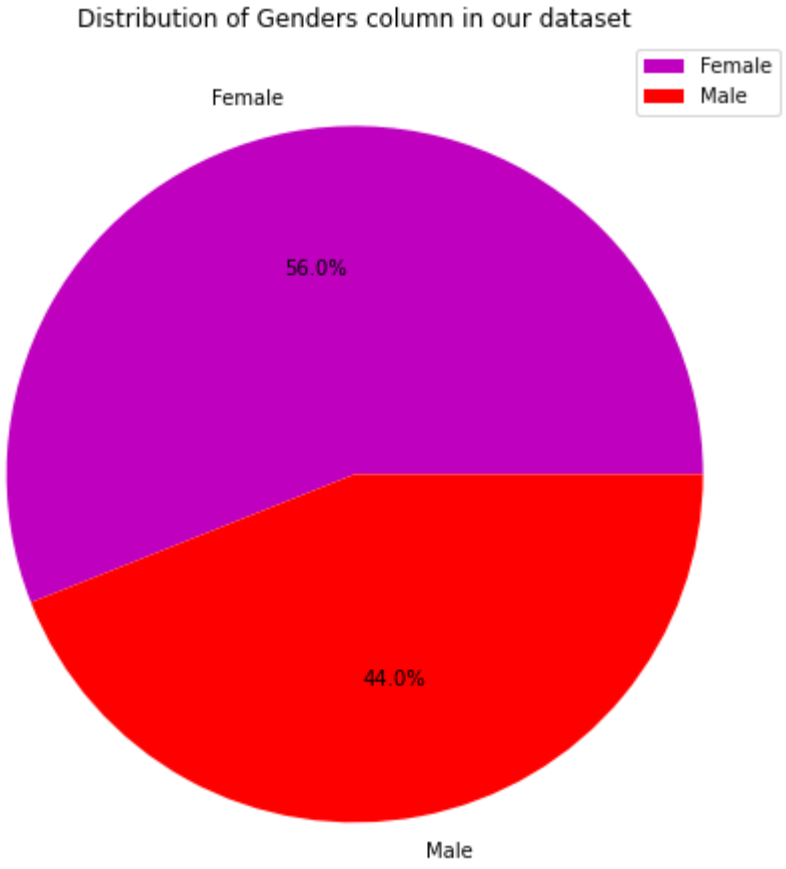
Out[35]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

In [40]:
```
df.isnull().sum()
```

Out[40]:
```
CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
dtype: int64
```

In [42]:
```
a=['Female','Male']
colors=['m','r']
plt.figure(figsize=(12,8))
plt.pie(df['Gender'].value_counts(),labels=a,colors=colors,autopct = "%1.1f
plt.title("Distribution of Genders column in our dataset")
plt.legend()
plt.show()
```



In [45]:
```
from sklearn.cluster import KMeans
```

In [46]:
```
new_df=df[['Annual Income (k$)', 'Spending Score (1-100)']]
new_df.head()
```

Out[46]:

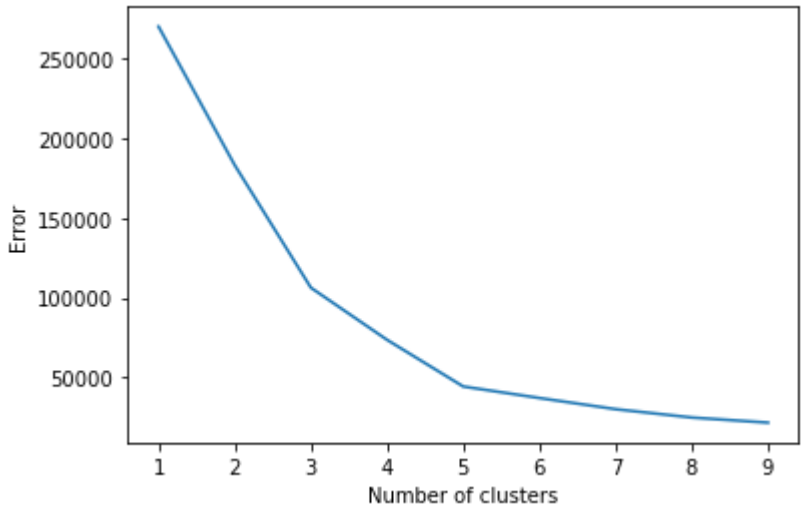
	Annual Income (k$)	Spending Score (1-100)
0	15	39
1	15	81
2	16	6
3	16	77
4	17	40

In [47]:
```
error= []
for i in range(1, 10):
    km = KMeans(n_clusters =  i)
    km.fit(new_df)
    error.append(km.inertia_)
```

```
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
```

In [49]:
```
plt.xlabel('Number of clusters')
plt.ylabel("Error")
plt.plot(range(1,10), error)
```

Out[49]:
```
[<matplotlib.lines.Line2D at 0x7fdb4e5edae0>]
```



In [50]:
```
km = KMeans(n_clusters = 5)
pred = km.fit_predict(new_df)
pred
```

```
/home/syam/.local/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:
1416: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warnin
g
  super()._check_params_vs_input(X, default_n_init=10)
```

Out[50]:
```
array([4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3,
       4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 0,
       4, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 1, 0, 1, 2, 1, 2, 1,
       0, 1, 2, 1, 2, 1, 2, 1, 2, 1, 0, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
       2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
       2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
       2, 1], dtype=int32)
```

In [51]:
```
df['Cluster'] = pred
df
```

Out[51]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	Cluster
0	1	Male	19	15	39	4
1	2	Male	21	15	81	3
2	3	Female	20	16	6	4
3	4	Female	23	16	77	3
4	5	Female	31	17	40	4
...	...	...	...	...	...	...
195	196	Female	35	120	79	1
196	197	Female	45	126	28	2
197	198	Male	32	126	74	1
198	199	Male	32	137	18	2
199	200	Male	30	137	83	1

200 rows × 6 columns

In [66]:
```
# Test the model with random observation

km.predict([[60,79]])
```

```
/home/syam/.local/lib/python3.10/site-packages/sklearn/base.py:465: UserWa
rning: X does not have valid feature names, but KMeans was fitted with fea
ture names
  warnings.warn(
```

Out[66]:
```
array([4], dtype=int32)
```