In [1]:

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

In [2]:

```
df = pd.read_csv('customer.csv')
```

In [3]:

```
backup_df = df
```

In [4]:

```
df.head()
```

Out[4]:

	CustomerID	Gender	Age	Annual Income ($)	Spending Score (1-100)	Profession	Work Experience	Family Size	Graduate
0	1	Male	19	15000	39	Healthcare	1	4	Ye
1	2	Male	21	35000	81	Engineer	3	3	Ye
2	3	Female	20	86000	6	Engineer	1	1	N
3	4	Female	23	59000	77	Lawyer	0	2	N
4	5	Female	31	38000	40	Entertainment	2	6	N

In [5]:

```
#df.drop(['Profession', 'Graduated','Gender'], axis=1, inplace=True)
```

In [6]:

```
df.shape
```

Out[6]:

```
(2000, 9)
```

In [7]:

```
# Select features for clustering
features = ["Age", "Annual Income ($)"]
```

In [9]:

```
# Perform k-means clustering
kmeans = KMeans(n_clusters = 2)
kmeans.fit(df[features])
```

```
D:\Softwares\Coding\Anaconda\lib\site-packages\sklearn\cluster\_kmeans.py:
870: 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
  warnings.warn(
```

Out[9]:

```
KMeans
KMeans(n_clusters=2)
```

In [10]:

```
# Visualising the results
plt.scatter(df["Age"], df["Annual Income ($)"], c=kmeans.labels_)
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1])
plt.xlabel("Age")
plt.ylabel("Annual Income ($)")
plt.title("Customer Clustering")
plt.show()
```

In [ ]:

In [ ]: