

# Clustering K-Means

In [1]:

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

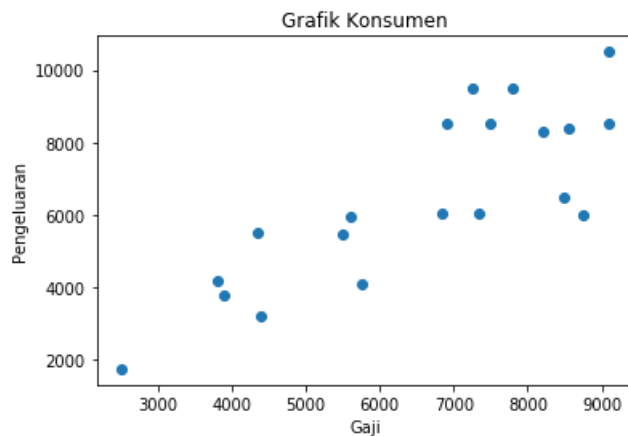
file = 'konsumen.xlsx'
dataframe = pd.read_excel(file)
predata = dataframe.drop('no', axis=1)
print(predata)
```

	gaji	pengeluaran
0	2500	1750
1	3800	4200
2	3900	3800
3	4350	5500
4	4400	3200
5	5500	5450
6	5600	5950
7	5750	4100
8	6850	6050
9	6900	8500
10	7250	9500
11	7350	6050
12	7500	8500
13	7800	9500
14	8200	8300
15	8500	6500
16	8550	8400
17	8750	6000
18	9100	10500
19	9100	8500

In [2]:

```
dataset = np.asarray(predata)

plt.scatter(dataset[:, 0], dataset[:, 1])
plt.xlabel("Gaji")
plt.ylabel("Pengeluaran")
plt.title("Grafik Konsumen")
plt.show()
```



In [3]:

```
kmeans = KMeans(n_clusters=2)
kmeans.fit(dataset)

print(kmeans.cluster_centers_)
```

```
print(kmeans.labels_)
```

```
[[7987.5  8025.  ]  
 [4475.   4243.75]]  
[1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0]
```

In [23]:

```
plt.scatter(dataset[:, 0], dataset[:, 1], c=kmeans.labels_, cmap='rainbow')  
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], color='black', marker='*',  
            s=200)  
plt.xlabel('Gaji')  
plt.ylabel('Pengeluaran')  
plt.title('Grafik Konsumen')  
plt.show()
```

