[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score
from sklearn.cluster import KMeans
```

[2]:
```
df = pd.read_csv('Canada_largest_companies.csv')
df.head()
```

[2]:

	Global Rank\t	Company\t	Sales ($billion)\t	Profits ($billion)\t	Assets ($billion)\t	Market Value ($billion)
0	50	Royal Bank of Canada	38.3	7.7	838.5	87.2
1	71	TD Bank Group	30.6	6.7	819.4	76.9
2	84	Bank of Nova Scotia	27.6	6.4	737.2	70.6
3	131	Bank of Montreal	20.9	4.1	542.9	41.0
4	142	Suncor Energy	38.8	2.8	76.8	47.3

[3]:
```
X = df.iloc[:,[2,5]].values
```

[4]:
```
model = KMeans(n_clusters = 3, random_state=0).fit(X)
print("Center is:\n", model.cluster_centers_)
pred_label = model.predict(X)
```

```
Center is:
 [[18.568      25.724     ]
 [33.825      70.5       ]
 [ 6.38888889  8.23888889]]
```

[6]:
```
plt.scatter(X[pred_label==0, 0], X[pred_label==0, 1], c = 'orange', label = 'Cluster 1', s = 80)
plt.scatter(X[pred_label==1, 0], X[pred_label==1, 1], c = 'blue', label = 'Cluster 2', s = 80)
plt.scatter(X[pred_label==2, 0], X[pred_label==2, 1], c = 'green', label = 'Cluster 3', s = 80)
plt.scatter(model.cluster_centers_[:,0], model.cluster_centers_[:,1], c = 'purple', label = 'Center', marker = 'D', s = 150)

plt.title('Kmeans Clustering')
plt.legend()
plt.show()
```