

In [ ]:

```
In [3]: #Importing Libraries
import pandas as pd
import matplotlib.pyplot as plt
#For Jupyter Notebooks to show the plots
%matplotlib inline
```

```
In [7]: from sklearn.datasets import load_iris
iris = load_iris()
```

```
In [11]: iris.keys()
```

```
Out[11]: dict_keys(['data', 'target', 'target_names', 'DESCR', 'feature_names', 'filename'])
```

```
In [14]: data = iris['data']
type(data)
```

```
Out[14]: numpy.ndarray
```

```
In [18]: df_iris = pd.DataFrame(data = iris['data'], columns=iris['feature_names'])
df_iris.head(5)
```

```
Out[18]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

Here we choose Sepal Length @ column 0 and Petal Length @ column 2

```
In [22]: X = df_iris[['sepal length (cm)', 'petal width (cm)']]
X.head(5)
```

```
Out[22]:
```

	sepal length (cm)	petal width (cm)
0	5.1	0.2
1	4.9	0.2
2	4.7	0.2
3	4.6	0.2
4	5.0	0.2

```
In [27]: from sklearn.cluster import KMeans  
         from sklearn.metrics import silhouette_score
```

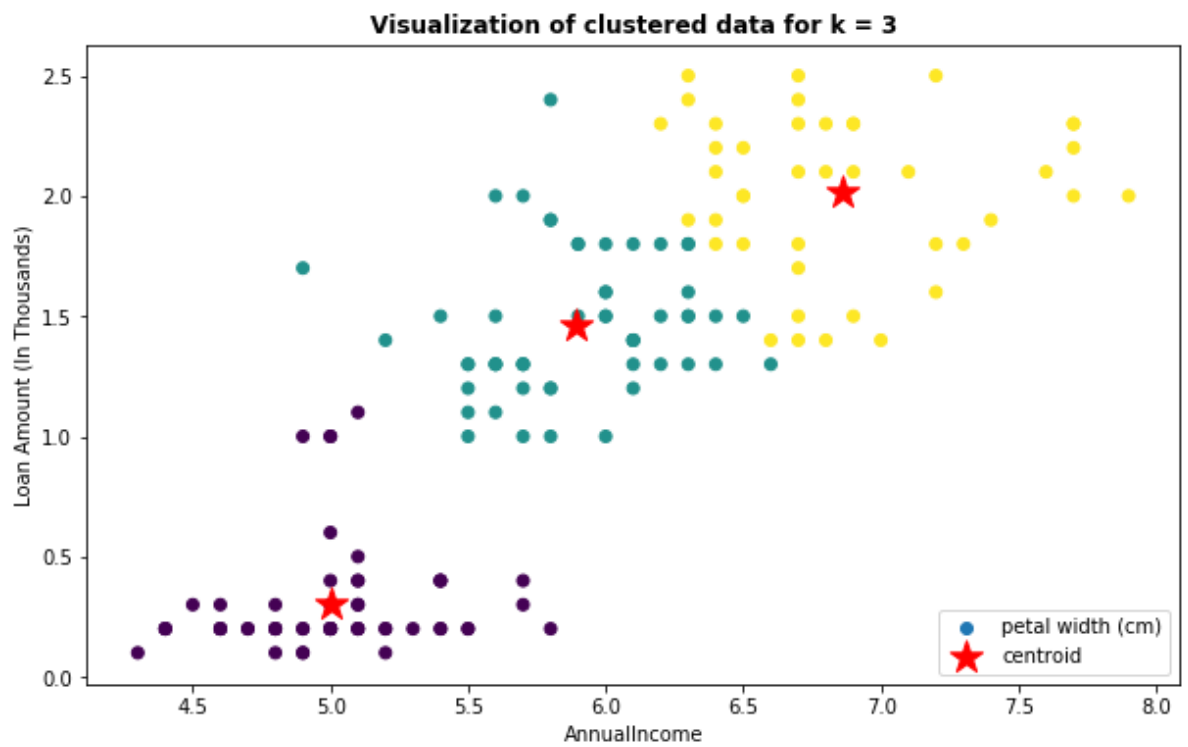
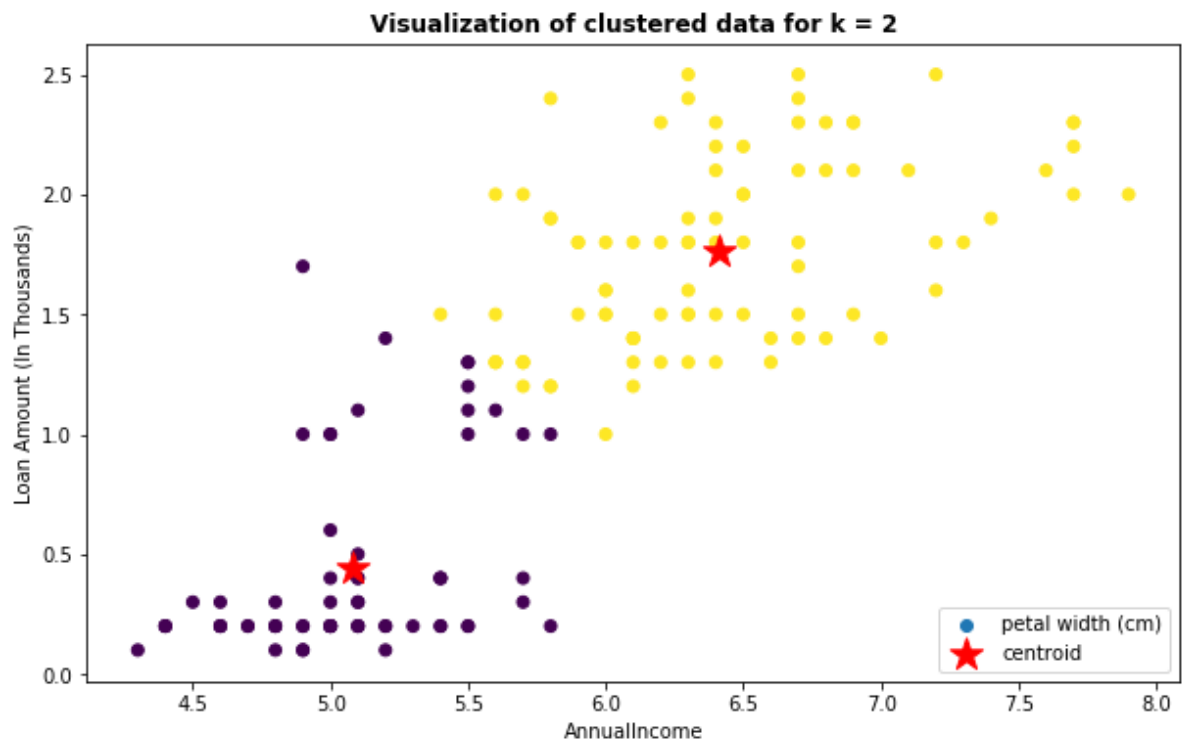
```

In [28]: for i, k in enumerate([2, 3]):
          fig, ax = plt.subplots(figsize=(10, 6))

          # Run the Kmeans algorithm
          km = KMeans(n_clusters=k)
          km.fit(X)
          labels = km.labels_
          centroids = km.cluster_centers_
          #print(centroids)

          # Get silhouette samples
          silhouette_vals = silhouette_score(X, labels)
          #print(silhouette_vals)
          # scatter plot
          plt.scatter(X['sepal length (cm)'], X['petal width (cm)'], c=labels)
          plt.scatter(centroids[:,0], centroids[:,1], marker='*', s=300, c='r', label='centroid')
          plt.xlabel('AnnualIncome')
          plt.ylabel('Loan Amount (In Thousands)')
          plt.title('Visualization of clustered data for k = {}'.format(k), fontweight='bold')
          plt.legend(loc =4)

```



In [ ]: