```python
import numpy as np
import pandas as pd
import os



from sklearn.datasets import load_iris
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans


iris=load_iris()
samples=iris.data
model=KMeans(n_clusters=3)
model.fit(samples)
labels=model.predict(samples)
print(labels)
```

```
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 2 2 2 2 0 2 2 2 2
 2 2 0 2 2 2 2 0 2 0 2 0 2 2 0 0 2 2 2 2 2 0 2 2 2 2 0 2 2 2 0 2 2 2 0 2 2 0 2
 2 0]
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
```

```python
xs=samples[:,0]
ys=samples[:,2]
plt.scatter(xs,ys,c=labels)
centroids=model.cluster_centers_
centroids_x=centroids[:,0]
centroids_y=centroids[:,2]
plt.scatter(centroids_x,centroids_y,marker='D',s=50)
plt.show()
```



```python
from google.colab import files
uploaded=files.upload()
```

Choose Files No file chosen   Upload widget is only available when the cell has been
executed in the current browser session. Please rerun this cell to enable.
Saving HousingData.csv to HousingData.csv

```python
import io


df = pd.read_csv(io.BytesIO(uploaded['HousingData.csv']))


print(df)
```

```
        CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD  TAX  \
0    0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900    1  296
1    0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671    2  242
2    0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671    2  242
3    0.03237   0.0   2.18   0.0  0.458  6.998  45.8  6.0622    3  222
4    0.06905   0.0   2.18   0.0  0.458  7.147  54.2  6.0622    3  222
..       ...   ...    ...   ...    ...    ...   ...     ...  ...  ...
501  0.06263   0.0  11.93   0.0  0.573  6.593  69.1  2.4786    1  273
502  0.04527   0.0  11.93   0.0  0.573  6.120  76.7  2.2875    1  273
```

```
503  0.06076    0.0  11.93    0.0  0.573  6.976   91.0  2.1675    1  273
504  0.10959    0.0  11.93    0.0  0.573  6.794   89.3  2.3889    1  273
505  0.04741    0.0  11.93    0.0  0.573  6.030    NaN  2.5050    1  273

       PTRATIO        B  LSTAT  MEDV
0         15.3   396.90   4.98  24.0
1         17.8   396.90   9.14  21.6
2         17.8   392.83   4.03  34.7
3         18.7   394.63   2.94  33.4
4         18.7   396.90    NaN  36.2
..         ...      ...    ...   ...
501       21.0   391.99    NaN  22.4
502       21.0   396.90   9.08  20.6
503       21.0   396.90   5.64  23.9
504       21.0   393.45   6.48  22.0
505       21.0   396.90   7.88  11.9

[506 rows x 14 columns]
```

```python
df_dict={
    0:'INDUS',
    1:'AGE',
    2:'TAX'
}
df_list=df['TAX'].map(df_dict).tolist()
```

```python
df=df.drop(['TAX'],axis=1)
df.head()
```

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3	18.7	396.90	NaN	36.2

```python
b=range(0,4)
PTRATIO =[]
```

```python
for k in b:
    model=KMeans(n_clusters=k)
```
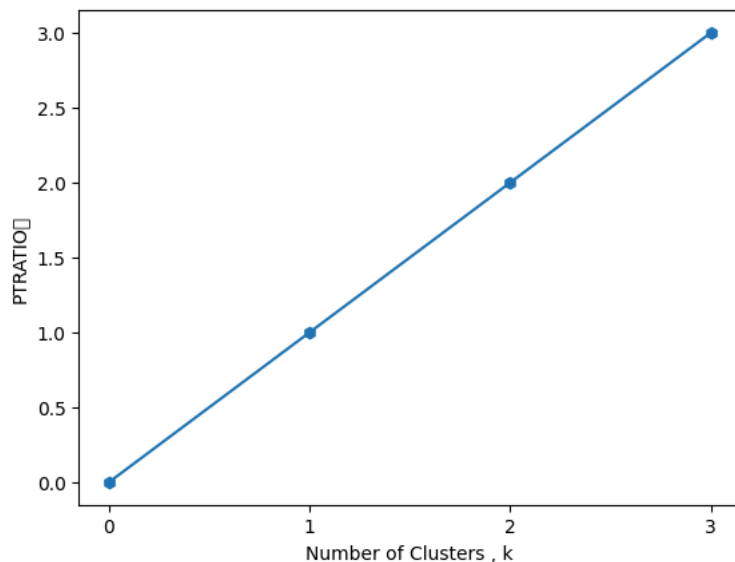
```python
plt.plot(b,'-h')
plt.xlabel('Number of Clusters , k')
plt.ylabel('PTRATIO ')
plt.xticks(b)
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 9 (  ) missing from current font.
  fig.canvas.print_figure(bytes_io, **kw)
```

```python
df.head()
```

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3	18.7	396.90	NaN	36.2

```python
model=KMeans(n_clusters=3)
centroids
```

```
array([[5.9016129 , 2.7483871 , 4.39354839, 1.43387097],
       [5.006     , 3.428     , 1.462     , 0.246     ],
       [6.85      , 3.07368421, 5.74210526, 2.07105263]])
```

```python
xs_ZN=df.iloc[:,0]
ys_AGE=df.iloc[:,-1]
centroids_xs_ZN=centroids[:,0]
centroids_ys_AGE=centroids[:,-1]
plt.scatter(xs_ZN,ys_AGE)
plt.scatter(centroids_xs_ZN,centroids_ys_AGE,c='red')
plt.show()
```