

```
In [113.]: import pandas as pd
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt

In [114.]: df=pd.read_csv("income.csv")

In [115.]: df.head()

Out[117]:
```

	Name	Age	Income\$
0	Rob	27	70000
1	Michael	29	90000
2	Mohan	29	61000
3	Ismael	28	60000
4	Kory	42	150000

```


In [116.]: df=df.rename({"Income$":"Income"},axis=1) # change of column name

In [117.]: df.head()

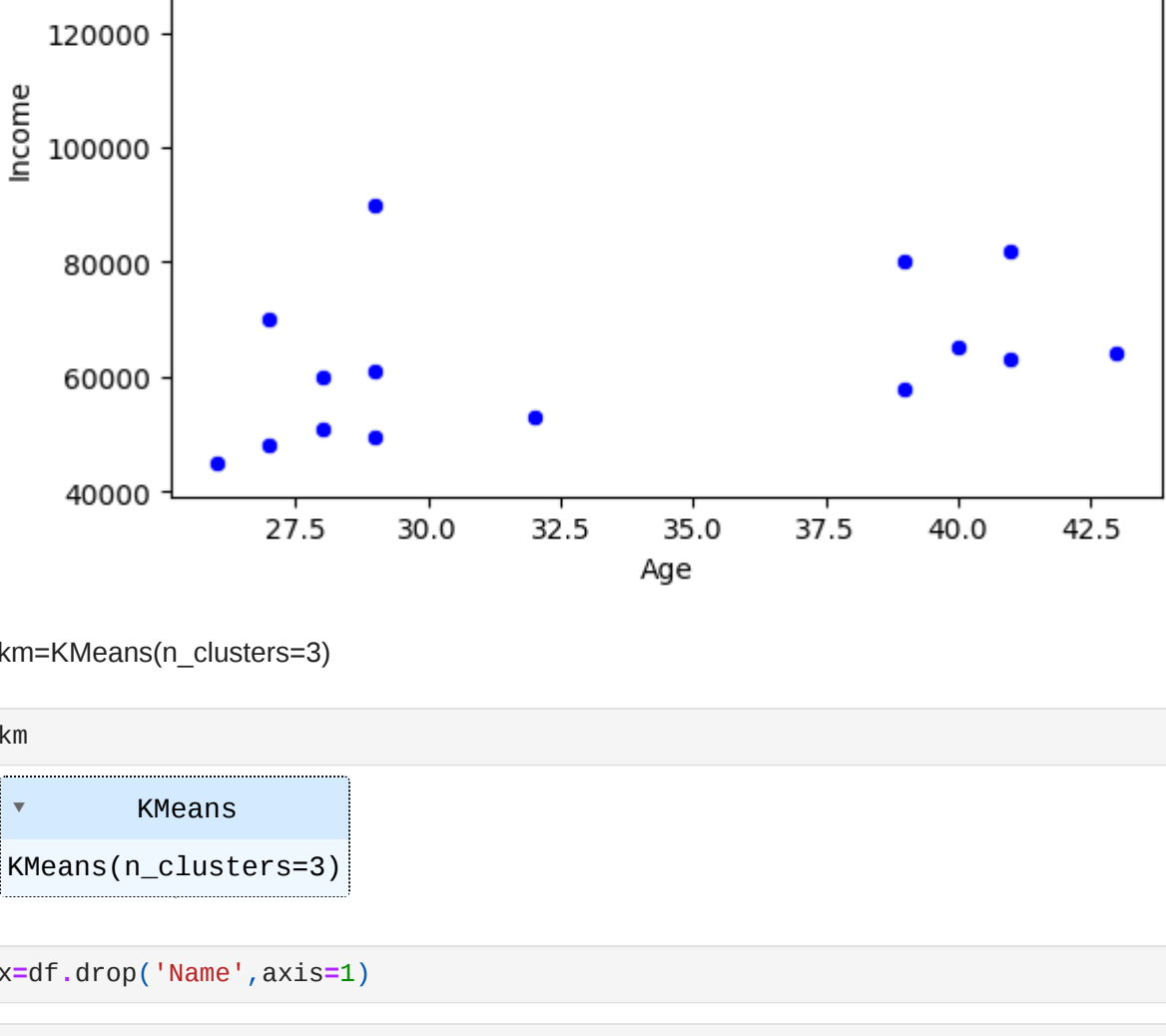
Out[117]:
```

	Name	Age	Income
0	Rob	27	70000
1	Michael	29	90000
2	Mohan	29	61000
3	Ismael	28	60000
4	Kory	42	150000

```


In [118.]: sns.scatterplot(x=df.Age,y=df.Income,color="blue")

Out[118.]: <AxesSubplot: xlabel='Age', ylabel='Income'>
```



```


In [119.]: km=KMeans(n_clusters=3)

Out[119.]:
```

	km
0	KMeans
1	KMeans(n_clusters=3)

```


In [120.]: x=df.drop("Name",axis=1)

In [121.]: x

Out[121.]:
```

	Age	Income
0	27	70000
1	29	90000
2	29	61000
3	28	60000
4	42	150000
5	39	150000
6	41	160000
7	38	162000
8	36	156000
9	35	130000
10	37	137000
11	26	45000
12	27	48000
13	28	52000
14	29	49000
15	32	53000
16	40	65000
17	41	63000
18	43	64000
19	39	80000
20	41	82000
21	39	59000

```


In [122.]: km.fit(x)

C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)

Out[122.]:
```

	km
0	KMeans
1	KMeans(n_clusters=3)

```


In [123.]: km.predict(x)

Out[123.]: array([0, 0, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 0, 8, 0, 2])

In [124.]: y=predicted=km.predict(x)

In [125.]: df["cluster"]=ypredicted

In [126.]: df

Out[126.]:
```

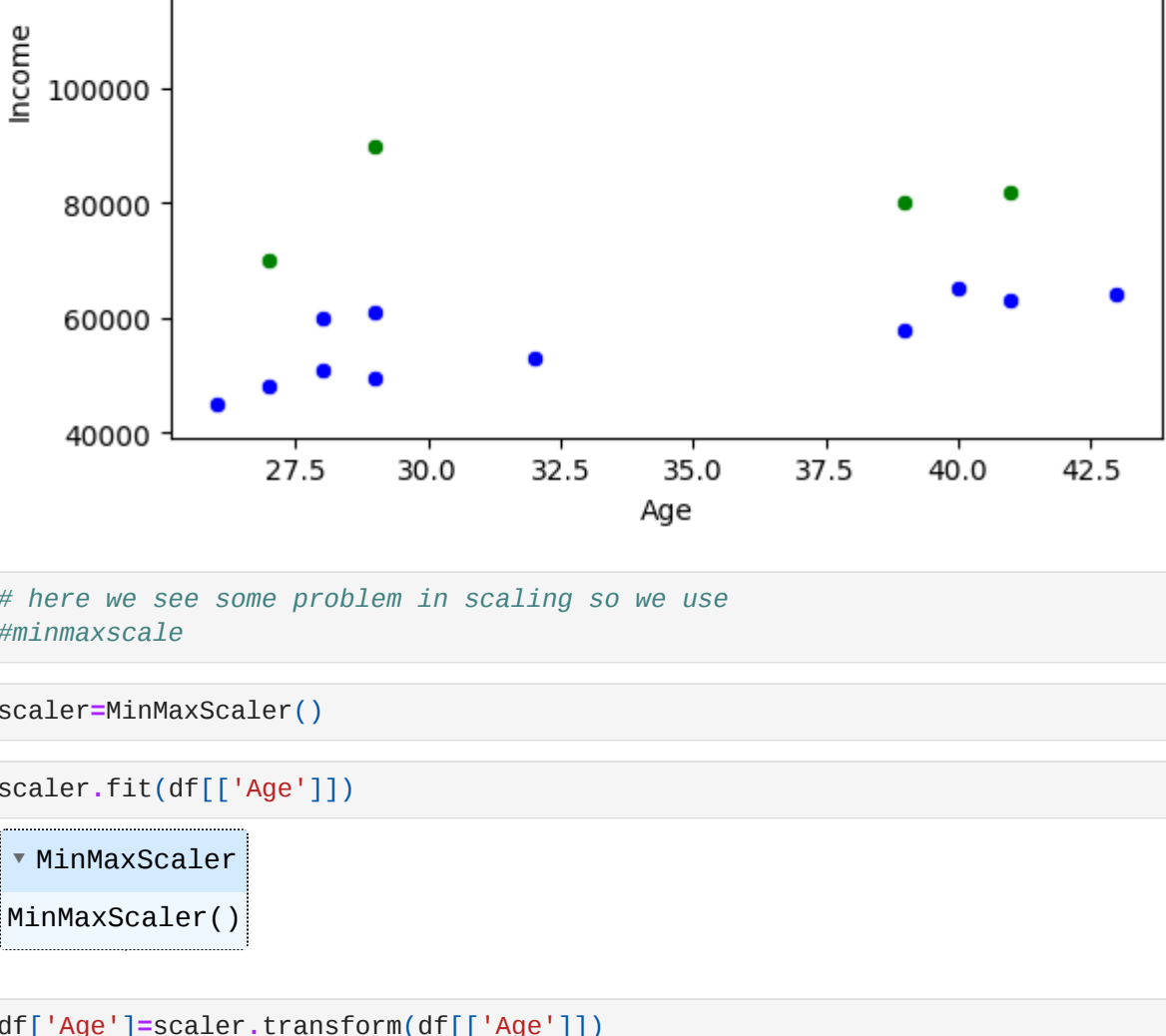
	Name	Age	Income	cluster
0	Rob	27	70000	0
1	Michael	29	90000	0
2	Mohan	29	61000	2
3	Ismael	28	60000	2
4	Kory	42	150000	1
5	Gautam	39	150000	1
6	David	40	160000	1
7	Andrea	39	162000	1
8	Brad	36	156000	1
9	Angelina	35	130000	1
10	Donald	37	137000	1
11	Tom	26	45000	2
12	Arnold	27	48000	2
13	Jared	28	51000	2
14	Stark	29	49000	2
15	Raulir	32	53000	2
16	Dipka	40	65000	2
17	Priyanka	41	63000	2
18	Nick	43	64000	2
19	Alia	39	80000	0
20	Sid	41	82000	0
21	Abdul	39	59000	2

```


In [127.]: df2=df[df["cluster"]==0]
df2=df[df["cluster"]==1]
df3=df[df["cluster"]==2]

In [128.]: sns.scatterplot(x=df1["Age"],y=df1["Income"],color="green")
sns.scatterplot(x=df2["Age"],y=df2["Income"],color="red")
sns.scatterplot(x=df3["Age"],y=df3["Income"],color="blue")

Out[128.]: <AxesSubplot: xlabel='Age', ylabel='Income'>
```



```


In [129.]: # here we see some problem in scaling so we use
MinMaxScaler

In [130.]: scaler=MinMaxScaler()

In [131.]: scaler.fit(df[["Age"]])

Out[131.]:
```

	MinMaxScaler
0	MinMaxScaler()

```


In [132.]: df["Age"]=scaler.transform(df[["Age"]])

In [133.]: df

Out[133.]:
```

	Name	Age	Income	cluster
0	Rob	0.058824	70000	0
1	Michael	0.176471	90000	0
2	Mohan	0.176471	61000	2
3	Ismael	0.176471	60000	2
4	Kory	0.941176	150000	1
5	Gautam	0.764706	150000	1
6	David	0.882353	160000	1
7	Andrea	0.765882	162000	1
8	Brad	0.588235	156000	1
9	Angelina	0.529412	130000	1
10	Donald	0.647059	137000	1
11	Tom	0.000000	45000	2
12	Arnold	0.058824	48000	2
13	Jared	0.117647	51000	2
14	Stark	0.176471	49000	2
15	Raulir	0.352941	53000	2
16	Dipka	0.823529	65000	2
17	Priyanka	0.882353	63000	2
18	Nick	1.000000	64000	2
19	Alia	0.764706	80000	0
20	Sid	0.882353	82000	0
21	Abdul	0.764706	59000	2

```


In [134.]: scaler.fit(df[["Income"]])

Out[134.]:
```

	MinMaxScaler
0	MinMaxScaler()

```


In [135.]: df["Income"]=scaler.transform(df[["Income"]])

In [140.]: df.drop("cluster",axis=1) # we drop the cluster bcz we need to train the model from start and find new cluster

Out[140.]:
```

	Name	Age	Income
0	Rob	0.058824	0.213675
1	Michael	0.176471	0.384615
2	Mohan	0.176471	0.136752
3	Ismael	0.176471	0.128205
4	Kory	0.941176	0.897436
5	Gautam	0.764706	0.940171
6	David	0.882353	0.982906
7	Andrea	0.765882	1.000000
8	Brad	0.588235	0.947118
9	Angelina	0.529412	0.726496
10	Donald	0.647059	0.786325
11	Tom	0.000000	0.000000
12	Arnold	0.058824	0.025641
13	Jared	0.117647	0.051282
14	Stark	0.176471	0.038462
15	Raulir	0.352941	0.068376
16	Dipka	0.823529	0.170940
17	Priyanka	0.882353	0.153846
18	Nick	1.000000	0.162393
19	Alia	0.764706	0.299145
20	Sid	0.882353	0.316239
21	Abdul	0.764706	0.111111

```


In [143.]: df

Out[143.]:
```

	Name	Age	Income	cluster
0	Rob	0.058824	0.213675	0
1	Michael	0.176471	0.384615	0
2	Mohan	0.176471	0.136752	2
3	Ismael	0.176471	0.128205	2
4	Kory	0.941176	0.897436	1
5	Gautam	0.764706	0.940171	1
6	David	0.882353	0.982906	1
7	Andrea	0.765882	1.000000	1
8	Brad	0.588235	0.947118	1
9	Angelina	0.529412	0.726496	1
10	Donald	0.647059	0.786325	1
11	Tom	0.000000	0.000000	2
12	Arnold	0.058824	0.025641	2
13	Jared	0.117647	0.051282	2
14	Stark	0.176471	0.038462	2
15	Raulir	0.352941	0.068376	2
16	Dipka	0.823529	0.170940	2
17	Priyanka	0.882353	0.153846	2
18	Nick	1.000000	0.162393	2
19	Alia	0.764706	0.299145	0
20	Sid	0.882353	0.316239	0
21	Abdul	0.764706	0.111111	2

```


In [144.]: df.drop("cluster",axis="columns",inplace=True)

In [145.]: df

Out[145.]:
```

	Name	Age	Income
0	Rob	0.058824	0.213675
1	Michael	0.176471	0.384615
2	Mohan	0.176471	0.136752
3	Ismael	0.176471	0.128205
4	Kory	0.941176	0.897436
5	Gautam	0.764706	0.940171
6	David	0.882353	0.982906
7	Andrea	0.765882	1.000000
8	Brad	0.588235	0.947118
9	Angelina	0.529412	0.726496
10	Donald	0.647059	0.786325
11	Tom	0.000000	0.000000
12	Arnold	0.058824	0.025641
13	Jared	0.117647	0.051282
14	Stark	0.176471	0.038462
15	Raulir	0.352941	0.068376
16	Dipka	0.823529	0.170940
17	Priyanka	0.882353	0.153846
18	Nick	1.000000	0.162393
19	Alia	0.764706	0.299145
20	Sid	0.882353	0.316239
21	Abdul	0.764706	0.111111

```


In [146.]: ram=df[["Age","Income"]]

ram

In [147.]: ram

Out[147.]:
```

	Age	Income
0	0.058824	0.213675
1	0.176471	0.384615
2	0.176471	0.136752
3	0.176471	0.128205
4	0.941176	0.897436
5	0.764706	0.940171
6	0.882353	0.982906
7	0.765882	1.000000
8	0.588235	0.947118
9	0.529412	0.726496
10	0.647059	0.786325
11	0.000000	0.000000
12	0.058824	0.025641
13	0.117647	0.051282
14	0.176471	0.038462
15	0.352941	0.068376
16	0.823529	0.170940
17	0.882353	0.153846
18	1.000000	0.162393
19	0.764706	0.299145
20	0.882353	0.316239
21	0.764706	0.111111

```


In [148.]: km.fit(ram) # here we train model with new income and age

C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)

Out[148.]:
```

	km
0	KMeans
1	KMeans(n_clusters=3)

```


In [151.]: km.predict(ram)

Out[151.]: array([0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1])

In [159.]: y_pre=km.predict(ram)

In [160.]: df["cluster"]=y_pre

In [161.]: df

Out[161.]:
```

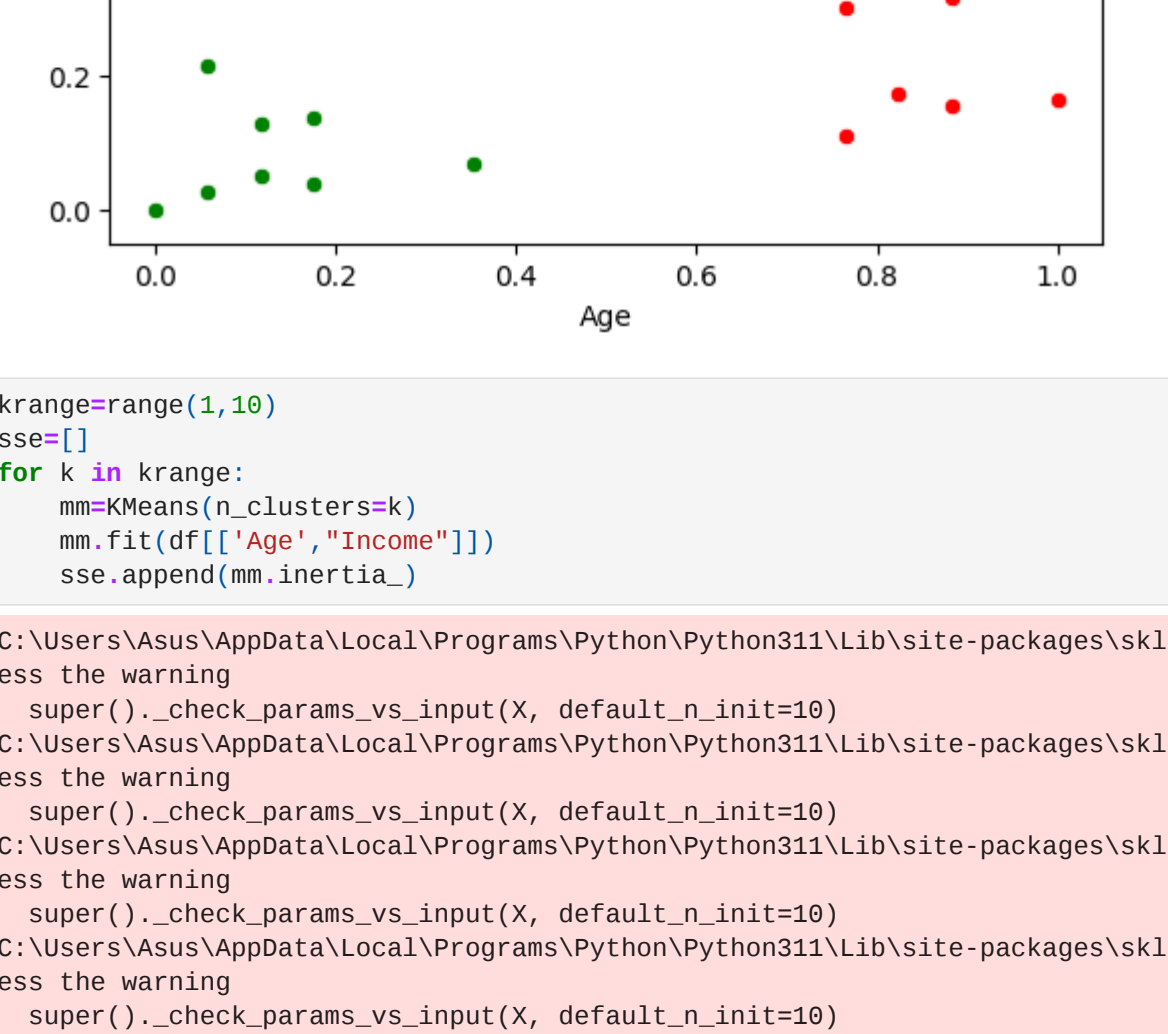
	Name	Age	Income	cluster
0	Rob	0.058824	0.213675	0
1	Michael	0.176471	0.384615	0
2	Mohan	0.176471	0.136752	0
3	Ismael	0.176471	0.128205	0
4	Kory	0.941176	0.897436	2
5	Gautam	0.764706	0.940171	2
6	David	0.882353	0.982906	2
7	Andrea	0.765882	1.000000	2
8	Brad	0.588235	0.947118	2
9	Angelina	0.529412	0.726496	2
10	Donald	0.647059	0.786325	2
11	Tom	0.000000	0.000000	0
12	Arnold	0.058824	0.025641	0
13	Jared	0.117647	0.051282	0
14	Stark	0.176471	0.038462	0
15	Raulir	0.352941	0.068376	0
16	Dipka	0.823529	0.170940	1
17	Priyanka	0.882353	0.153846	1
18	Nick	1.000000	0.162393	1
19	Alia	0.764706	0.299145	1
20	Sid	0.882353	0.316239	1
21	Abdul	0.764706	0.111111	1

```


In [162.]: df2=df[df["cluster"]==0]
df2=df[df["cluster"]==1]
df3=df[df["cluster"]==2]

In [163.]: sns.scatterplot(x=df1["Age"],y=df1["Income"],color="green")
sns.scatterplot(x=df2["Age"],y=df2["Income"],color="red")
sns.scatterplot(x=df3["Age"],y=df3["Income"],color="blue")

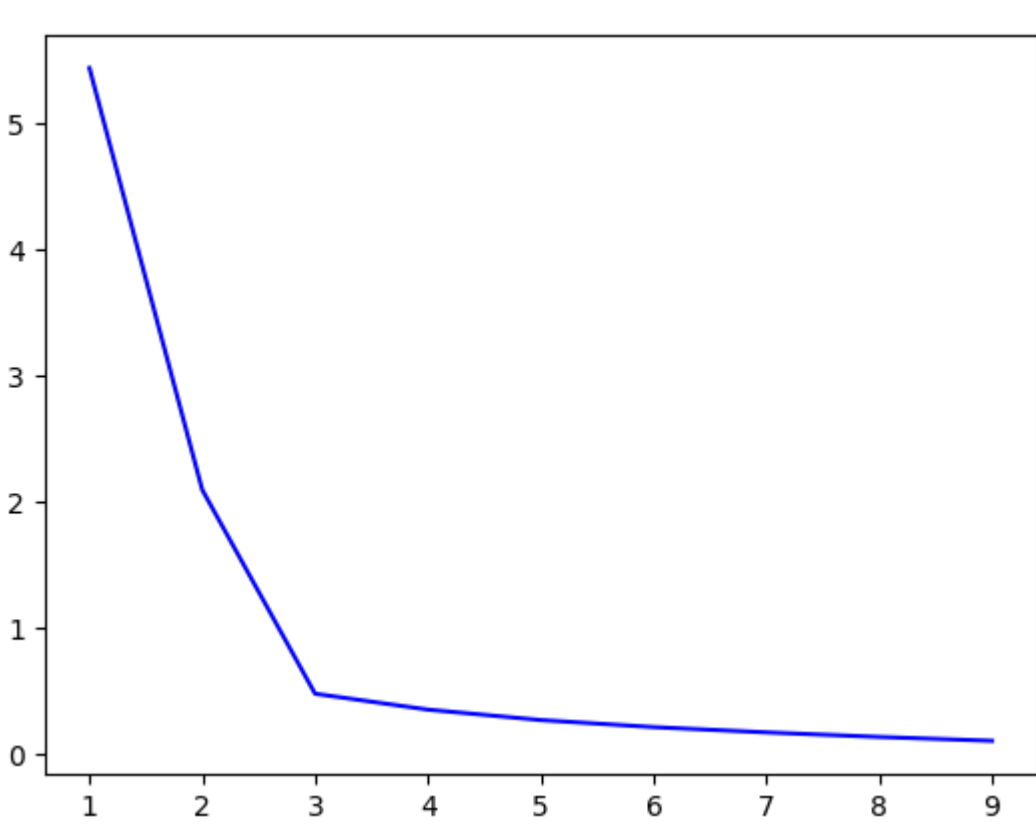
Out[163.]: <AxesSubplot: xlabel='Age', ylabel='Income'>
```



```


In [169.]: k=range(1,18)
sse=[]
for k in krange:
    mm=KMeans(n_clusters=k)
    mm.fit(df[["Age","Income"]])
    sse.append(mm.inertia_)

C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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 warning
  super().__check_params_vs_input(X, default_n_init=10)
C:\Users\Axius\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: FutureWarning: The default value of 'n_init' will change from 10 to 'auto' in 1
```



In []: # here the elbow is at 3 so the proper value of k would be 3