Repeating the same with normalization of feature

- We must see that the values of data don’t vary much in sale and magnitude.
- Ex: The age can range from 18 to 70, but the income range:1lac to 9lac
- The gap in the two are huge, which may effect the model preparation
- To treat them equally we need to do scaling of features

Here features are normalized

Repeating the same with normalization of feature

```
from sklearn.preprocessing import StandardScaler
```

```
scaler = StandardScaler()
scaled_customers_df = scaler.fit_transform(customers_df[["age",
                      "income"]])
scaled_customers_df[0:5]
```

```
array([[ 1.3701637,    0.09718548],
       [-1.3791283,    0.90602749],
       [ 1.10388844,   0.51405021],
       [ 0.23849387,   -1.27162408],
       [-0.35396857,   -1.32762083]])
```

Here features are normalized

Create Clusters Using Normalized Feature Set

```
from sklearn.cluster import KMeans
clusters_new = KMeans(3, random_state=42)
clusters_new.fit(scaled_customers_df)
customers_df["clusterid_new"] = clusters_new.labels_
```

```
#Display the sample data after clustering operation
customers_df[0:5]
```

	income	age	clusterid	clusterid_new
0	41100.0	48.75	2	0
1	54100.0	28.10	0	1
2	47800.0	46.75	2	0
3	19100.0	40.25	1	2
4	18200.0	35.80	1	2

Observe that previous cluster and this clustering Differs, it is due to normalization of features

Scatter plot after normalization

```
#Plotting the customers with their segments
sn.lmplot(  data=customers_df,x="age",y="income",hue="clusterid_new");
plt.title( "Fig 3: Customer Segments Based on Income and Age with clusterid_new");
```