

Name: Tahrir Kabir

Roll: BFH1825013F

Email: tahrirkabir321@gmail.com

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn
%matplotlib inline
```

```
In [6]: df = pd.read_csv('loan.csv')
df
```

```
Out[6]:
```

	Age	Loan(k)	House Price Index
0	25	40	135
1	35	60	256
2	45	80	231
3	20	20	267
4	35	120	139
5	52	18	150
6	23	95	127
7	40	62	216
8	60	100	139
9	48	220	250

```
In [7]: df.rename(columns={'Age': 'age', 'Loan(k)': 'loan', 'House Price Index': 'house price index'},in
place=True)
```

```
In [5]: df
```

```
Out[5]:
```

	age	loan	Default
0	25	40	N
1	35	60	N
2	45	80	N
3	20	20	N
4	35	120	N
5	52	18	N
6	23	95	N
7	40	62	Y
8	60	100	Y
9	48	220	Y

```
In [9]: #missing value testing
df.isnull().values.any()
```

```
Out[9]: False
```

```
In [10]: #Statistical Analysis
df.describe()
```

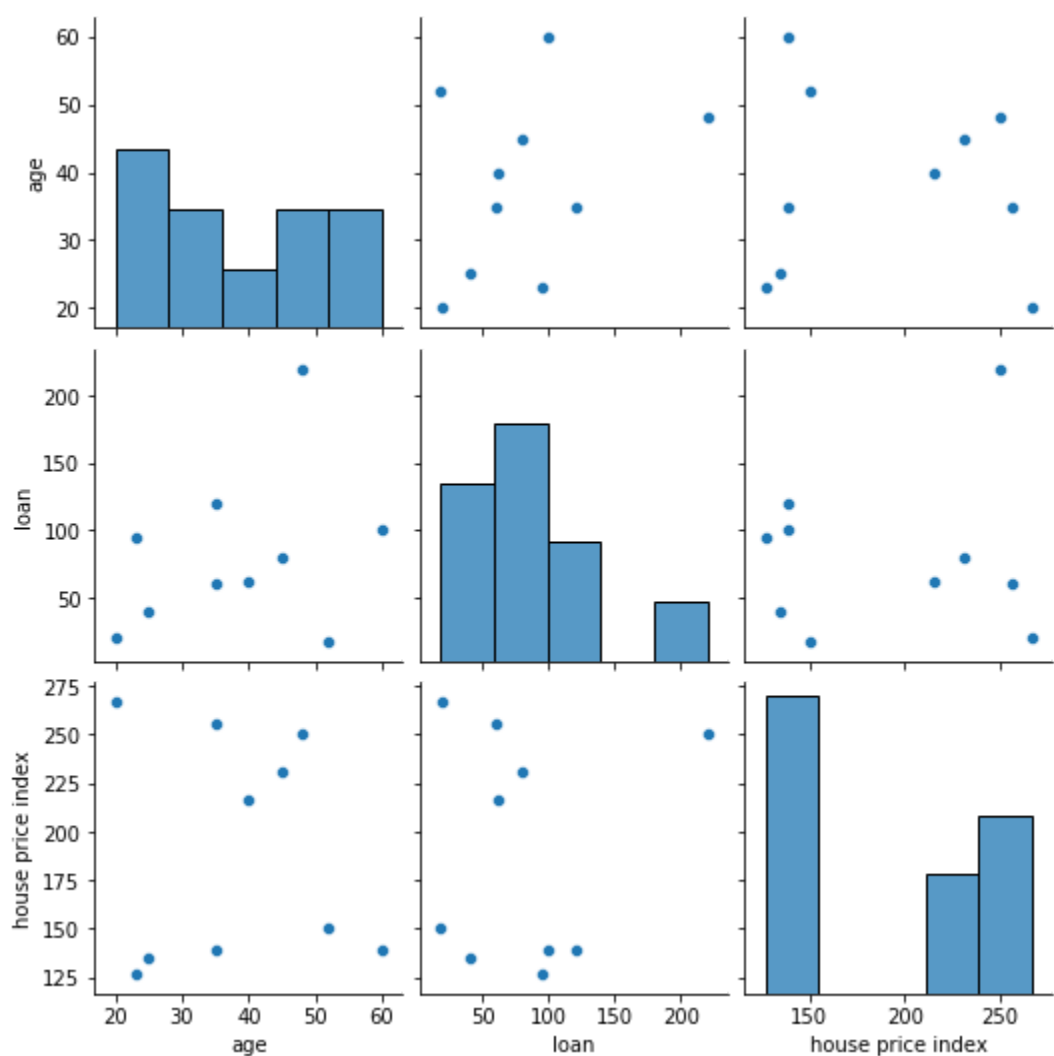
```
Out[10]:
```

	age	loan	house price index
count	10.000000	10.000000	10.000000
mean	38.300000	81.500000	191.000000
std	13.199747	59.226965	57.761964
min	20.000000	18.000000	127.000000
25%	27.500000	45.000000	139.000000
50%	37.500000	71.000000	183.000000
75%	47.250000	98.750000	245.250000
max	60.000000	220.000000	267.000000

## visualization by PairPlot

```
In [11]: seaborn.pairplot(df[['age', 'loan', 'house price index']])
```

```
Out[11]: <seaborn.axisgrid.PairGrid at 0x1f5c0c87b88>
```



## Clustering

```
In [15]: import sklearn.cluster as cluster
```

```
In [16]: kmeans = cluster.KMeans(n_clusters=3)
```

```
In [18]: kmeans = kmeans.fit(df[['age', 'loan']])
```

```
In [19]: kmeans.cluster_centers_
```

```
Out[19]: array([[ 40.75,  98.75],
 [ 34.4 ,  40. ],
 [ 48. , 220. ]])
```

```
In [21]: df['loan_cluster']=kmeans.labels_
```

```
In [22]: df
```

```
Out[22]:
```

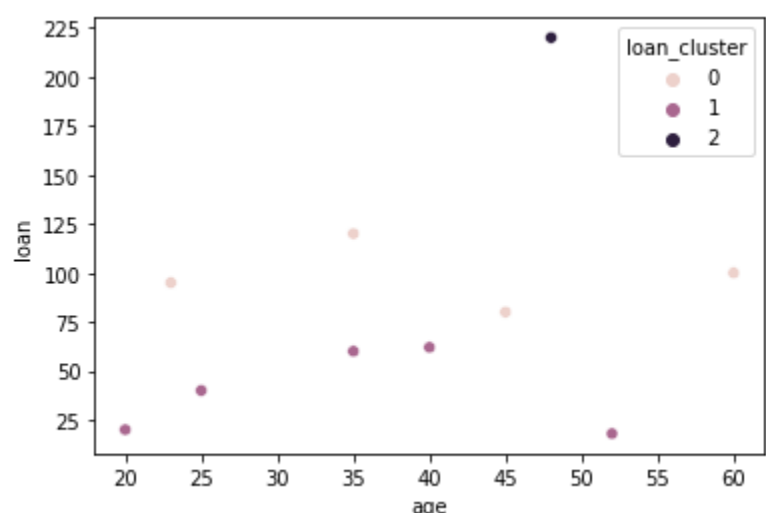
	age	loan	house price index	loan_cluster
0	25	40	135	1
1	35	60	256	1
2	45	80	231	0
3	20	20	267	1
4	35	120	139	0
5	52	18	150	1
6	23	95	127	0
7	40	62	216	1
8	60	100	139	0
9	48	220	250	2

```
In [23]: df['loan_cluster'].value_counts()
```

```
Out[23]: 1    5
0    4
2    1
Name: loan_cluster, dtype: int64
```

```
In [24]: #ScatterPlot
seaborn.scatterplot(x='age', y='loan', hue='loan_cluster', data=df)
```

```
Out[24]: <matplotlib.axes._subplots.AxesSubplot at 0x1f5c319f148>
```



tahrirkabir321@gmail.com

house@tahrirkabir321@gmail.com#2021

```
In [ ]:
```