```
import numpy as np
import pandas as pd
from sklearn.neighbors import KNeighborsClassifier
from scipy.stats import zscore
from sklearn.preprocessing import Imputer
from sklearn.metrics import accuracy_score
import seaborn as sns
import os
%matplotlib inline
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=9473

Enter your authorization code:
··········
Mounted at /content/drive
```

```
data = pd.read_csv('/content/drive/My Drive/8sept/project/Bank_Personal_Loan_Modelling-1.csv')
```

```
data
```

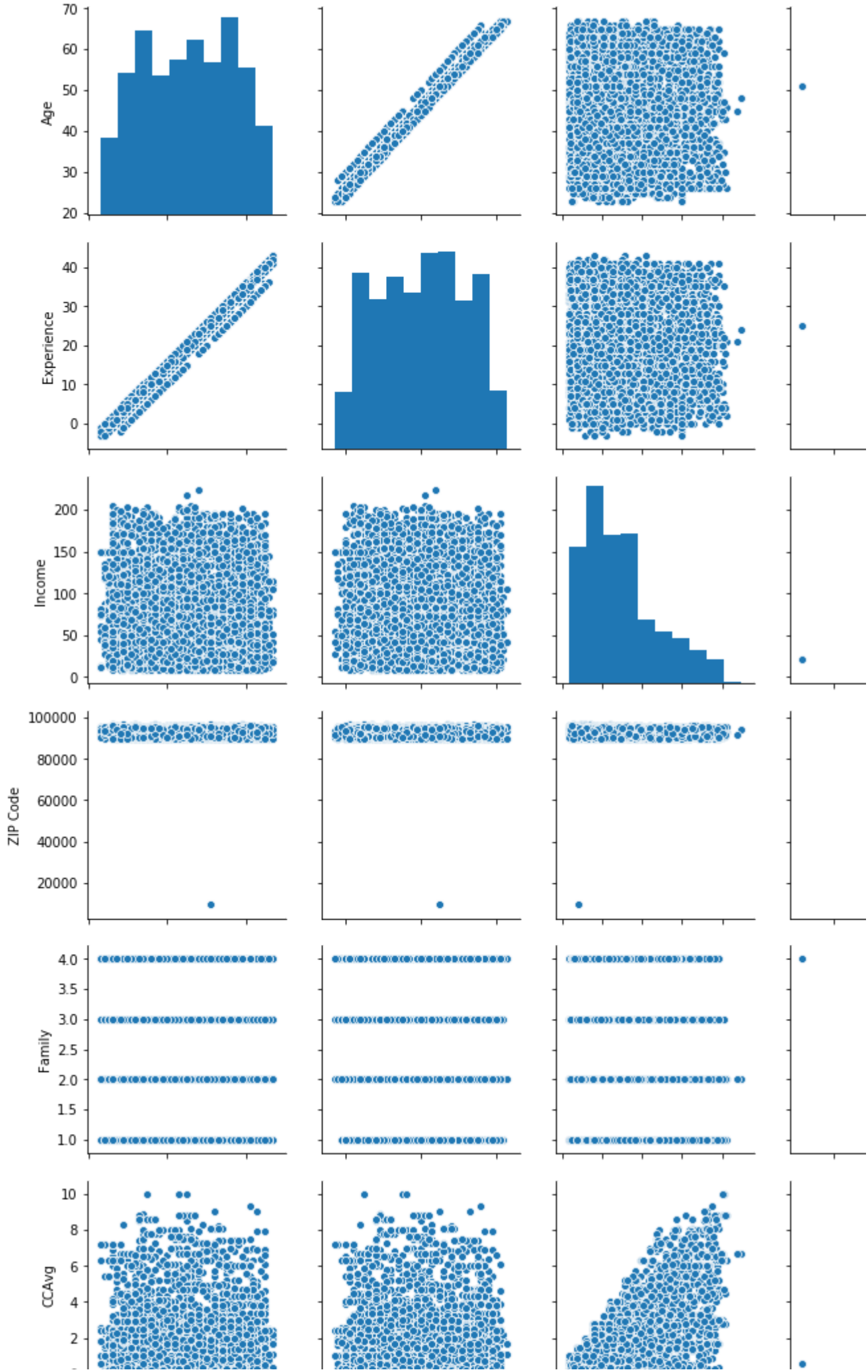
	ID	Age	Experience	Income	ZIP Code	Family	CCAvg	Education	Mortgage	Pe
0	1	25	1	49	91107	4	1.60	1	0	
1	2	45	19	34	90089	3	1.50	1	0	
2	3	39	15	11	94720	1	1.00	1	0	
3	4	35	9	100	94112	1	2.70	2	0	
4	5	35	8	45	91330	4	1.00	2	0	
5	6	37	13	29	92121	4	0.40	2	155	
6	7	53	27	72	91711	2	1.50	2	0	
7	8	50	24	22	93943	1	0.30	3	0	
8	9	35	10	81	90089	3	0.60	2	104	
9	10	34	9	180	93023	1	8.90	3	0	
10	11	65	39	105	94710	4	2.40	3	0	
11	12	29	5	45	90277	3	0.10	2	0	
12	13	48	23	114	93106	2	3.80	3	0	
13	14	59	32	40	94920	4	2.50	2	0	
14	15	67	41	112	91741	1	2.00	1	0	
15	16	60	30	22	95054	1	1.50	3	0	
16	17	38	14	130	95010	4	4.70	3	134	
17	18	42	18	81	94305	4	2.40	1	0	
18	19	46	21	193	91604	2	8.10	3	0	
19	20	55	28	21	94720	1	0.50	2	0	
20	21	56	31	25	94015	4	0.90	2	111	
21	22	57	27	63	90095	3	2.00	3	0	
22	23	29	5	62	90277	1	1.20	1	260	
23	24	44	18	43	91320	2	0.70	1	163	
24	25	36	11	152	95521	2	3.90	1	159	
25	26	43	19	29	94305	3	0.50	1	97	
26	27	40	16	83	95064	4	0.20	3	0	
27	28	46	20	158	90064	1	2.40	1	0	
28	29	56	30	48	94539	1	2.20	3	0	
29	30	38	13	119	94104	1	3.30	2	0	
...	...	...	...	...	...	...	...	...	...	...
4970	4971	37	13	95	95821	2	1.70	2	0	
4971	4972	58	28	73	90024	1	1.40	3	0	

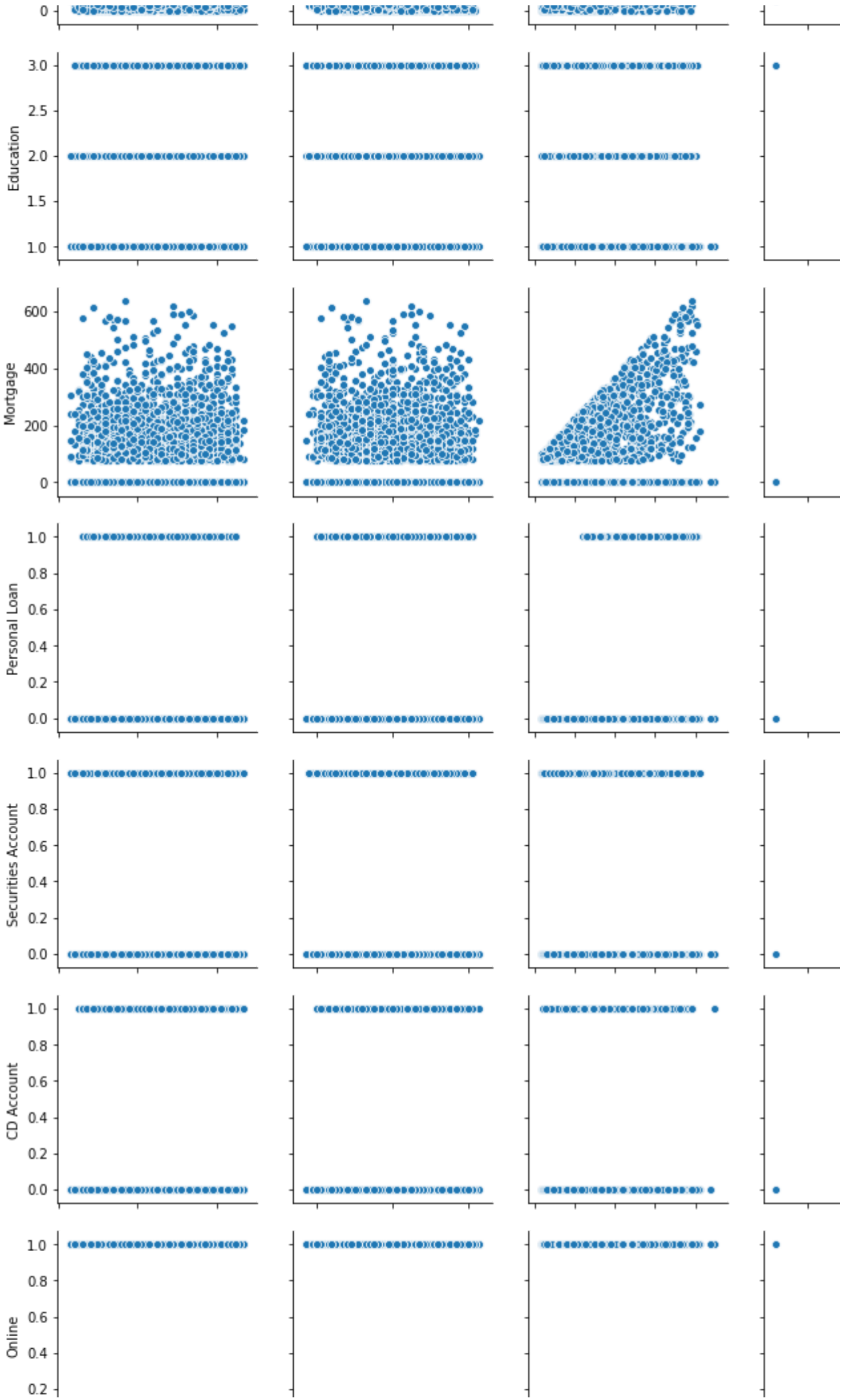
4971	4972	58	28	73	90024	1	1.40	3	0
4972	4973	58	32	41	93401	3	2.20	3	148
4973	4974	31	1	68	95045	4	4.00	3	0
4974	4975	59	33	64	92867	4	1.70	2	0
4975	4976	38	11	29	95207	4	1.00	2	0
4976	4977	29	5	31	95039	1	1.80	2	0
4977	4978	40	15	54	90266	3	0.80	3	0
4978	4979	57	27	63	90210	4	2.00	3	0
4979	4980	50	26	92	90740	1	2.60	2	213
4980	4981	29	5	135	95762	3	5.30	1	0
4981	4982	34	9	195	90266	2	3.00	1	122
4982	4983	36	10	45	95126	4	0.20	1	0
4983	4984	51	26	72	95370	1	2.90	1	0
4984	4985	27	1	98	94043	4	2.30	3	0
4985	4986	48	23	30	94720	3	1.70	2	162
4986	4987	32	6	78	95825	1	2.90	3	0
4987	4988	48	23	43	93943	3	1.70	2	159
4988	4989	34	8	85	95134	1	2.50	1	136
4989	4990	24	0	38	93555	1	1.00	3	0
4990	4991	55	25	58	95023	4	2.00	3	219
4991	4992	51	25	92	91330	1	1.90	2	100
4992	4993	30	5	13	90037	4	0.50	3	0
4993	4994	45	21	218	91801	2	6.67	1	0
4994	4995	64	40	75	94588	3	2.00	3	0
4995	4996	29	3	40	92697	1	1.90	3	0
4996	4997	30	4	15	92037	4	0.40	1	85
4997	4998	63	39	24	93023	2	0.30	3	0
4998	4999	65	40	49	90034	3	0.50	3	0

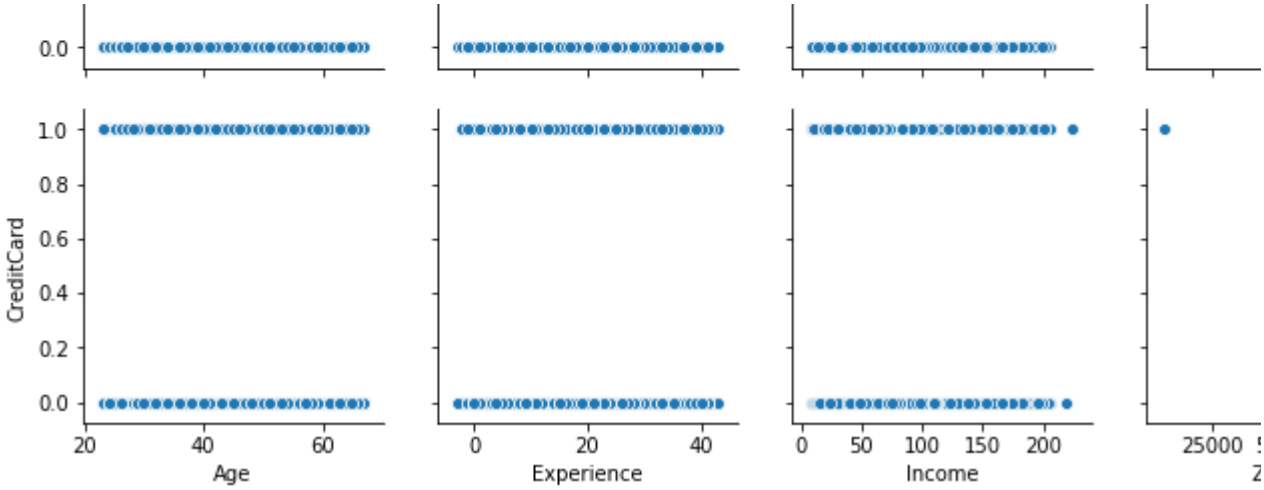
```
data.describe()
```

	ID	Age	Experience	Income	ZIP Code	Family	
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5(
mean	2500.500000	45.338400	20.104600	73.774200	93152.503000	2.396400	
std	1443.520003	11.463166	11.467954	46.033729	2121.852197	1.147663	

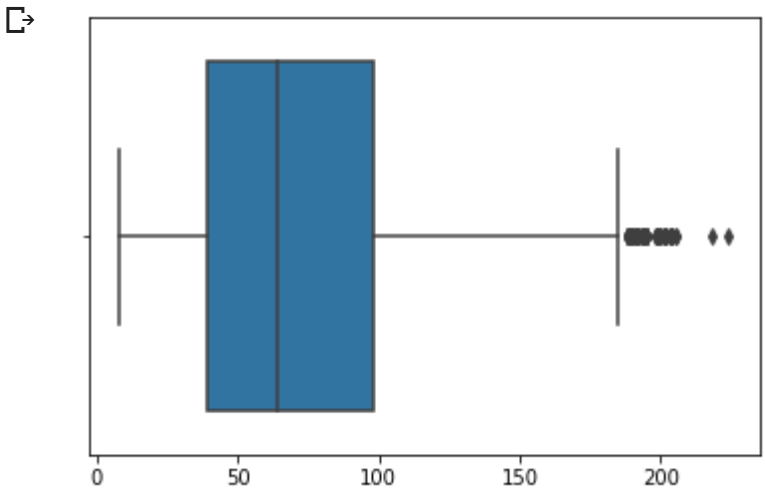
```
plt = sns.pairplot(data[['Age','Experience','Income','ZIP Code','Family','CCAvg' ,'Education' , '
```


Age
Experience
Income
ZIP Code
Family
CCAvg

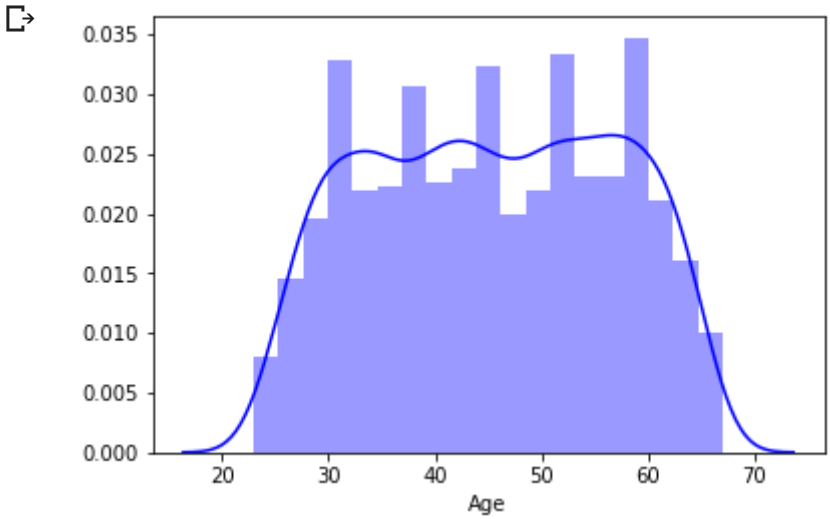

Education
Mortgage
Personal Loan
Securities Account
CD Account
Online




```
plt = sns.boxplot(data[['Income']])
```



```
plt=sns.distplot(data['Age'], color = 'b')
```

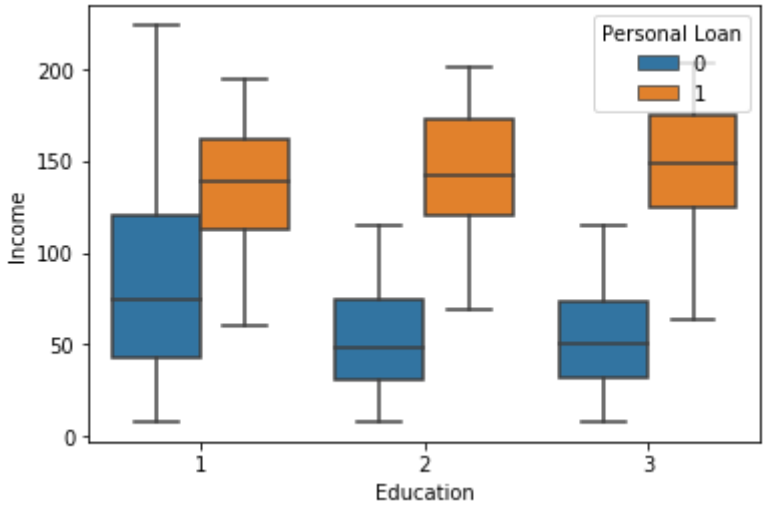



```
#Observation

#Most of the customers age is in  range of 30 to 60 yrs and their experience ranging  of 5 to 35
```

```
sns.boxplot(x="Education", y="Income", hue="Personal Loan", data=data)
```

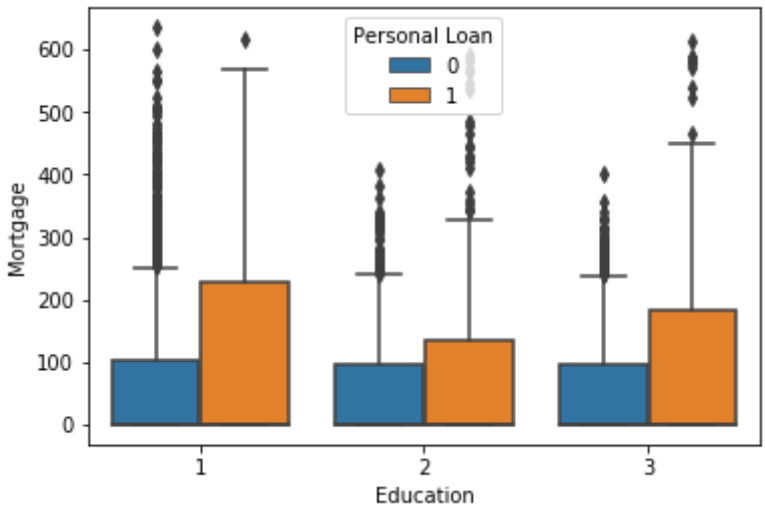
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6fc2a59358>
```




```
# here people with education leavel 1 have highh income but people who go for personal loans have

sns.boxplot(x="Education", y='Mortgage', hue="Personal Loan", data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6fc2999da0>
```



```
# people who have taken load are having high mortguage

sns.countplot(x="Family", data=data,hue="Personal Loan")
```

```
sns.countplot(x="Securities Account", data=data, hue="Personal Loan")
```

```
sns.countplot(x="CD Account", data=data,hue="Personal Loan")
```

```
sns.countplot(x="CreditCard", data=data,hue="Personal Loan")
```

```
import pandas as pd
import numpy as np
import matplotlib as mp
import seaborn as sns
%matplotlib inline
sns.set(style="ticks")

from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import BaggingClassifier, RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier

from scipy.stats import zscore
```

```
from matplotlib import pyplot as plt
plt.figure(figsize=(25, 25))
ax = sns.heatmap(data.corr(), vmax=.8, square=True, fmt='.2f', annot=True, linecolor='white', lin
plt.title('Correlation')
plt.show()
```

```
#Age and Experience is highly corelated
#Income and CCAvg also corelated
```

```
from sklearn.model_selection import train_test_split
train_set, test_set = train_test_split(data.drop(['Experience' ,'ID' ,'CCAvg'], axis=1), test_siz
```

```
train_set.describe()
```

```
test_set.describe()
```

```
from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import RandomForestClassifier
train_labels = train_set.pop("Personal Loan")
test_labels = test_set.pop("Personal Loan")


#DecisionTreeClassifier


dt_model = DecisionTreeClassifier(criterion = 'entropy' , max_depth = 3)


dt_model.fit(train_set, train_labels)
```

```
dt_model.score(test_set , test_labels)
```

```
#Naive Bayes


naive_model = GaussianNB()
naive_model.fit(train_set, train_labels)
```