In [1]:
```
import pandas as pd
```

In [4]:
```
df=pd.read_excel("default of credit card clients.xls",header=1)
```

In [5]:
```
df.head()
```

Out[5]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	...	E
0	1	20000.0	2.0	2.0	1	24.0	2	2	-1	-1	...	
1	2	120000.0	2.0	2.0	2	26.0	-1	2	0	0	...	
2	3	90000.0	2.0	2.0	2	34.0	0	0	0	0	...	
3	4	50000.0	2.0	2.0	1	37.0	0	0	0	0	...	
4	5	50000.0	1.0	2.0	1	57.0	-1	0	-1	0	...	

5 rows × 25 columns

In [6]:
```
df.shape
```

Out[6]:
```
(30000, 25)
```

In [8]:
```
print((df == 0).sum().sum())
```

```
174353
```

In [9]:
```
df.dtypes
```

Out[9]:
```
ID                            int64
LIMIT_BAL                   float64
SEX                         float64
EDUCATION                   float64
MARRIAGE                      int64
AGE                         float64
PAY_0                         int64
PAY_2                         int64
PAY_3                         int64
PAY_4                         int64
PAY_5                         int64
PAY_6                         int64
BILL_AMT1                   float64
BILL_AMT2                   float64
BILL_AMT3                     int64
BILL_AMT4                   float64
BILL_AMT5                   float64
BILL_AMT6                   float64
PAY_AMT1                    float64
PAY_AMT2                      int64
PAY_AMT3                    float64
PAY_AMT4                      int64
PAY_AMT5                    float64
PAY_AMT6                      int64
default payment next month    int64
dtype: object
```

In [10]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30000 entries, 0 to 29999
Data columns (total 25 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   ID                          30000 non-null  int64  
 1   LIMIT_BAL                   29999 non-null  float64
 2   SEX                         29999 non-null  float64
 3   EDUCATION                   29999 non-null  float64
 4   MARRIAGE                    30000 non-null  int64  
 5   AGE                         29999 non-null  float64
 6   PAY_0                       30000 non-null  int64  
 7   PAY_2                       30000 non-null  int64  
 8   PAY_3                       30000 non-null  int64  
 9   PAY_4                       30000 non-null  int64  
 10  PAY_5                       30000 non-null  int64  
 11  PAY_6                       30000 non-null  int64  
 12  BILL_AMT1                   29998 non-null  float64
 13  BILL_AMT2                   29999 non-null  float64
 14  BILL_AMT3                   30000 non-null  int64  
 15  BILL_AMT4                   29999 non-null  float64
 16  BILL_AMT5                   29999 non-null  float64
 17  BILL_AMT6                   29999 non-null  float64
 18  PAY_AMT1                    29998 non-null  float64
 19  PAY_AMT2                    30000 non-null  int64  
 20  PAY_AMT3                    29999 non-null  float64
 21  PAY_AMT4                    30000 non-null  int64  
 22  PAY_AMT5                    29999 non-null  float64
 23  PAY_AMT6                    30000 non-null  int64  
 24  default payment next month  30000 non-null  int64  
dtypes: float64(12), int64(13)
memory usage: 5.7 MB
```

In [11]:
```
df.isnull().sum()
```

Out[11]:
```
ID                          0
LIMIT_BAL                   1
SEX                         1
EDUCATION                   1
MARRIAGE                    0
AGE                         1
PAY_0                       0
PAY_2                       0
PAY_3                       0
PAY_4                       0
PAY_5                       0
PAY_6                       0
BILL_AMT1                   2
BILL_AMT2                   1
BILL_AMT3                   0
BILL_AMT4                   1
BILL_AMT5                   1
BILL_AMT6                   1
PAY_AMT1                    2
PAY_AMT2                    0
PAY_AMT3                    1
PAY_AMT4                    0
PAY_AMT5                    1
PAY_AMT6                    0
default payment next month  0
dtype: int64
```

In [12]:
```
df.describe()
```

Out[12]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	A
count	30000.000000	29999.000000	29999.000000	29999.000000	30000.000000	29999.0000
mean	15000.500000	167485.238841	1.603753	1.853095	1.551867	35.4852
std	8660.398374	129749.727113	0.489125	0.790334	0.521970	9.2179
min	1.000000	10000.000000	1.000000	0.000000	0.000000	21.0000
25%	7500.750000	50000.000000	1.000000	1.000000	1.000000	28.0000
50%	15000.500000	140000.000000	2.000000	2.000000	2.000000	34.0000
75%	22500.250000	240000.000000	2.000000	2.000000	2.000000	41.0000
max	30000.000000	1000000.000000	2.000000	6.000000	3.000000	79.0000

8 rows × 25 columns

In [15]:
```
df['SEX']=df['SEX'].fillna(df['SEX'].mode()[0])
df['EDUCATION']=df['EDUCATION'].fillna(df['EDUCATION'].mode()[0])
df['AGE']=df['AGE'].fillna(df['AGE'].mean())
```

In [17]:
```
df.dropna()
```

Out[17]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_
0	1	20000.0	2.0	2.0	1	24.0	2	2	-1	
1	2	120000.0	2.0	2.0	2	26.0	-1	2	0	
2	3	90000.0	2.0	2.0	2	34.0	0	0	0	
3	4	50000.0	2.0	2.0	1	37.0	0	0	0	
4	5	50000.0	1.0	2.0	1	57.0	-1	0	-1	
...	...	...	...	...	...	...	...	...	...	
29995	29996	220000.0	1.0	3.0	1	39.0	0	0	0	
29996	29997	150000.0	1.0	3.0	2	43.0	-1	-1	-1	
29997	29998	30000.0	1.0	2.0	2	37.0	4	3	2	
29998	29999	80000.0	1.0	3.0	1	41.0	1	-1	0	
29999	30000	50000.0	1.0	2.0	1	46.0	0	0	0	

29989 rows × 25 columns

In [25]:
```
df_AGE = df[df['AGE'] > 30]
df_AGE
```

Out[25]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_
2	3	90000.0	2.0	2.0	2	34.0	0	0	0	
3	4	50000.0	2.0	2.0	1	37.0	0	0	0	
4	5	50000.0	1.0	2.0	1	57.0	-1	0	-1	
5	6	50000.0	1.0	1.0	2	37.0	0	0	0	
9	10	20000.0	1.0	3.0	2	35.0	-2	-2	-2	
...	...	...	...	...	...	...	...	...	...	
29995	29996	220000.0	1.0	3.0	1	39.0	0	0	0	
29996	29997	150000.0	1.0	3.0	2	43.0	-1	-1	-1	
29997	29998	30000.0	1.0	2.0	2	37.0	4	3	2	
29998	29999	80000.0	1.0	3.0	1	41.0	1	-1	0	
29999	30000	50000.0	1.0	2.0	1	46.0	0	0	0	

18987 rows × 26 columns

In [29]:
```
df_col=df[['AGE','SEX','EDUCATION','MARRIAGE','default payment next month']]
df_col
```

Out[29]:

	AGE	SEX	EDUCATION	MARRIAGE	default payment next month
0	24.0	2.0	2.0	1	1
1	26.0	2.0	2.0	2	1
2	34.0	2.0	2.0	2	0
3	37.0	2.0	2.0	1	0
4	57.0	1.0	2.0	1	0
...	...	...	...	...	...
29995	39.0	1.0	3.0	1	0
29996	43.0	1.0	3.0	2	0
29997	37.0	1.0	2.0	2	1
29998	41.0	1.0	3.0	1	1
29999	46.0	1.0	2.0	1	1

30000 rows × 5 columns

In [35]:
```
df_con3=df[(df['EDUCATION']==3 )& (df['default payment next month']==1)]
df_con3
```

Out[35]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_
50	51	70000.0	1.0	3.0	2	42.0	1	2	2	
60	61	500000.0	2.0	3.0	1	28.0	0	0	0	
82	83	60000.0	1.0	3.0	2	30.0	0	0	0	
103	104	50000.0	2.0	3.0	2	22.0	0	0	0	
120	121	50000.0	1.0	3.0	2	37.0	2	2	2	
...	...	...	...	...	...	...	...	...	...	
29919	29920	50000.0	1.0	3.0	1	37.0	-1	-1	2	
29929	29930	170000.0	1.0	3.0	1	46.0	-1	-1	-1	
29932	29933	160000.0	1.0	3.0	1	42.0	2	0	0	
29942	29943	130000.0	1.0	3.0	1	45.0	-1	-1	-1	
29998	29999	80000.0	1.0	3.0	1	41.0	1	-1	0	

1237 rows × 26 columns

In [ ]:
```
x=df.drop('default payment next month',axis=1)
y=df['default payment next month']
```

In [37]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,train_size=0.75,r
```

In [38]:
```
x_train
```

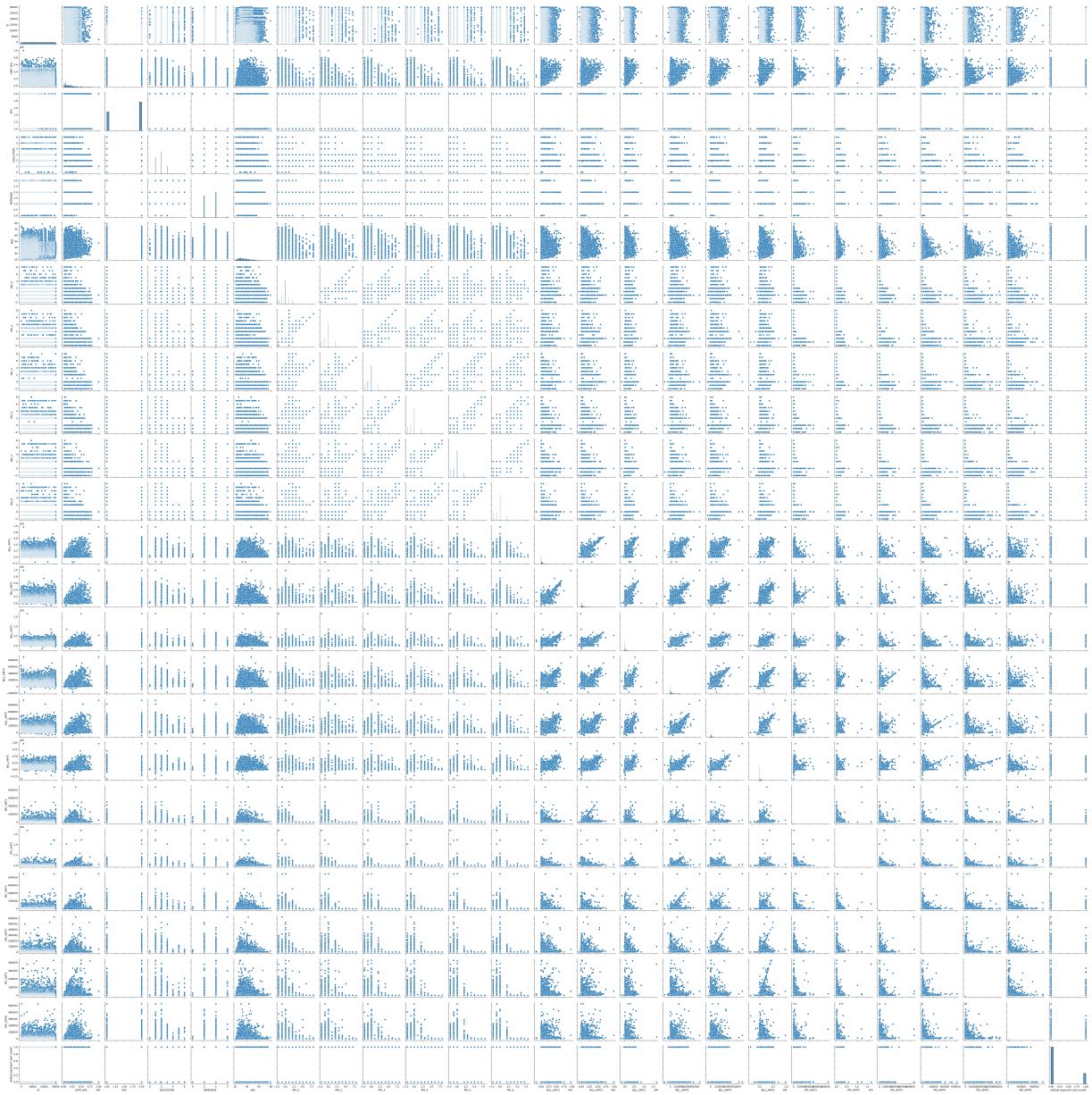
Out[38]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_
21177	21178	300000.0	1.0	3.0	2	31.0	0	0	0	
23942	23943	20000.0	1.0	2.0	2	24.0	0	0	0	
1247	1248	90000.0	2.0	2.0	2	35.0	-1	-1	-1	
23622	23623	300000.0	2.0	2.0	1	40.0	1	-2	-2	
28454	28455	70000.0	2.0	2.0	2	36.0	0	0	0	
...	...	...	...	...	...	...	...	...	...	
29802	29803	50000.0	1.0	2.0	2	32.0	0	0	0	
5390	5391	200000.0	1.0	1.0	2	37.0	2	2	2	
860	861	50000.0	1.0	1.0	2	26.0	-2	-2	-2	
15795	15796	70000.0	2.0	2.0	2	25.0	0	0	0	
23654	23655	160000.0	2.0	2.0	1	36.0	-2	-2	-2	

22500 rows × 25 columns

In [40]:

```
import matplotlib.pyplot as plt
import seaborn as sns
sns.pairplot(df)
plt.show()
```



In [ ]: