

Credit Card Default

```
In [25]: #This project is on credit card default
```

```
In [93]: !date
Sat 13 May 2023 12:51:54 PM EDT
```

```
In [94]: from IPython import display
display.Image("/home/harohutch77/pictures/cc.gif",height='330',width='500')
```

```
Out[94]: <IPython.core.display.Image object>
```

```
In [4]: # imports from libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import cufflinks as cf
```

```
In [6]: # reading the csv file using a path to the file
ccdf = pd.read_csv('/home/harohutch77/python/python_master/UCI_Credit_Card.csv')
ccdf
```

```
Out[6]:
```

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	BILL_AMT4	BILL_AMT5	BILL_AMT6	P	
0	1	20000.0	2	2	1	24	2	2	-1	-1	0.0	0.0	0.0	0.0
1	2	120000.0	2	2	2	26	-1	2	0	0	3272.0	3455.0	3261.0	0.0
2	3	90000.0	2	2	2	34	0	0	0	0	14331.0	14948.0	15549.0	0.0
3	4	50000.0	2	2	1	37	0	0	0	0	28314.0	28959.0	29547.0	0.0
4	5	50000.0	1	2	1	57	-1	0	-1	0	20940.0	19146.0	19131.0	0.0
...	0.0
29995	29996	220000.0	1	3	1	39	0	0	0	0	88004.0	31237.0	15980.0	0.0
29996	29997	150000.0	1	3	2	43	-1	-1	-1	-1	8979.0	5190.0	0.0	0.0
29997	29998	30000.0	1	2	2	37	4	3	2	1	20878.0	20582.0	19357.0	0.0
29998	29999	80000.0	1	3	1	41	1	-1	0	0	52774.0	11855.0	48944.0	0.0
29999	30000	50000.0	1	2	1	46	0	0	0	0	36535.0	32428.0	15313.0	0.0

30000 rows × 25 columns

```
In [7]: ccdf.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       30000 non-null   float64
 2   SEX             30000 non-null   int64  
 3   EDUCATION       30000 non-null   int64  
 4   MARRIAGE        30000 non-null   int64  
 5   AGE             30000 non-null   int64  
 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       30000 non-null   float64
 13  BILL_AMT2       30000 non-null   float64
 14  BILL_AMT3       30000 non-null   float64
 15  BILL_AMT4       30000 non-null   float64
 16  BILL_AMT5       30000 non-null   float64
 17  BILL_AMT6       30000 non-null   float64
 18  PAY_AMT1        30000 non-null   float64
 19  PAY_AMT2        30000 non-null   float64
 20  PAY_AMT3        30000 non-null   float64
 21  PAY_AMT4        30000 non-null   float64
 22  PAY_AMT5        30000 non-null   float64
 23  PAY_AMT6        30000 non-null   float64
 24  default.payment.next.month 30000 non-null   int64 
dtypes: float64(13), int64(12)
memory usage: 5.7 MB
```

Statistics on Credit Card Default dataframe - csv file

```
In [93]: #here is the description of the dataset starting with count to max
ccdf.describe().transpose()
```

```
Out[93]:
```

	count	mean	std	min	25%	50%	75%	max
ID	30000.0	15000.500000	8660.398374	1.0	7500.75	15000.5	22500.25	30000.0
LIMIT_BAL	30000.0	167484.322667	129747.661567	10000.0	50000.00	140000.0	240000.0	1000000.0
SEX	30000.0	1.603733	0.489129	1.0	1.00	2.0	2.00	2.0
EDUCATION	30000.0	1.853133	0.790349	0.0	1.00	2.0	2.00	6.0
MARRIAGE	30000.0	1.551867	0.521970	0.0	1.00	2.0	2.00	3.0
AGE	30000.0	35.485500	9.217904	21.0	28.00	34.0	41.00	79.0
PAY_0	30000.0	-0.016700	1.123802	-2.0	-1.00	0.0	0.00	8.0
PAY_2	30000.0	-0.133767	1.197186	-2.0	-1.00	0.0	0.00	8.0
PAY_3	30000.0	-0.166200	1.196668	-2.0	-1.00	0.0	0.00	8.0
PAY_4	30000.0	-0.220667	1.169139	-2.0	-1.00	0.0	0.00	8.0
PAY_5	30000.0	-0.266200	1.133187	-2.0	-1.00	0.0	0.00	8.0
PAY_6	30000.0	-0.291100	1.149988	-2.0	-1.00	0.0	0.00	8.0
BILL_AMT1	30000.0	51223.330900	73635.860576	-165580.0	3558.75	22381.5	67091.00	964511.0
BILL_AMT2	30000.0	49179.075167	71173.768783	-69777.0	2984.75	21200.0	64006.25	983931.0
BILL_AMT3	30000.0	47013.154800	69349.387427	-157264.0	2666.25	20088.5	60164.75	1664089.0
BILL_AMT4	30000.0	43262.948967	64332.856134	-17000.0	2326.75	19052.0	54506.0	891586.0
BILL_AMT5	30000.0	40311.409067	60797.155770	-81334.0	1763.00	18104.5	50190.50	927171.0
BILL_AMT6	30000.0	38871.760400	59554.107537	-33960.0	1256.00	17071.0	49198.25	961664.0
PAY_AMT1	30000.0	5663.580500	15653.280354	0.0	1000.00	2100.0	5006.00	873552.0
PAY_AMT2	30000.0	5921.163500	23040.870402	0.0	833.00	2009.0	5000.00	1684259.0
PAY_AMT3	30000.0	5225.681500	17606.961470	0.0	390.00	1800.0	4505.00	896040.0
PAY_AMT4	30000.0	4826.076867	15666.159744	0.0	296.00	1500.0	4013.25	621000.0
PAY_AMT5	30000.0	4799.387633	15278.305679	0.0	252.50	1500.0	4031.50	426529.0
PAY_AMT6	30000.0	5215.502567	17777.465775	0.0	117.75	1500.0	4000.00	528666.0
default.payment.next.month	30000.0	0.221200	0.415062	0.0	0.00	0.0	0.00	1.0

```
In [ ]:
```

```
In [62]: ccdf.mean()
```

```
Out[62]:
```

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	default.payment.next.month
15000.000000	167484.322667	1.603733	1.853133	1.551867	35.485500	-0.016700	-0.133767	-0.166200	-0.220667	51223.330900	49179.075167	47013.154800	43262.948967	40311.409067	38871.760400	0.221200

```
In [66]: ccdf.std()
```

```
Out[66]:
```

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	default.payment.next.month
8660.398374	129747.661567	0.489129	0.790349	0.521970	9.217904	1.123802	1.197186	1.196668	1.169139	73635.860576	71173.768783	69349.387427	64332.856134	60797.155770	59554.107537	0.415062

```
In [ ]:
```

```
In [62]: ccdf.mean()
```

```
Out[62]:
```

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	default.payment.next.month
15000.000000	167484.322667	1.603733	1.853133	1.551867	35.485500	-0.016700	-0.133767	-0.166200	-0.220667	51223.330900	49179.075167	47013.154800	43262.948967	40311.409067	38871.760400	0.221200

```
In [ ]:
```

```
In [66]: ccdf.std()
```

```
Out[66]:
```

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	default.payment.next.month
8660.398374	129747.661567	0.489129	0.790349	0.521970	9.217904	1.123802	1.197186	1.196668	1.169139	73635.860576	71173.768783	69349.387427	64332.856134	60797.155770	59554.107537	0.415062

```
In [ ]:
```

```
#different ways to bring up your data - many ways with csv
#ccdf[ID]==29991][['AGE','SEX']]
#ccdf[ccdf['default.payment.next.month']==1]['ID']
```

```
In [72]: ccdf[['LIMIT_BAL','SEX','AGE','EDUCATION','MARRIAGE','default.payment.next.month']].count()
```

```
Out[72]:
```

LIMIT_BAL	SEX	AGE	EDUCATION	MARRIAGE	default.payment.next.month
30000	30000	30000	30000	30000	30000

```
In [5]: #how many people are going to be in default next month 6636
ccdf[ccdf['default.payment.next.month']==1].count()
```

```
Out[5]:
```

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	default.payment.next.month
6636	129747.661567	0.489129	0.790349	0.521970	9.217904	-0.016700	-0.133767	-0.166200	-0.220667	51223.330900	49179.075167	47013.154800	43262.948967	40311.409067	38871.760400	0.221200

```
In [5]: #how many people are going to be in default next month 6636
ccdf[ccdf['default.payment.next.month']==1].count()
```

```
Out[5]:
```

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AM
----	-----------	-----	-----------	----------	-----	-------	-------	-------	-------	-----------	-----------	-----------	-----------	-----------	---------