CREDIT ONE

Credit one is a company that rates customers but there is an increase in customer default rates.

The agenda is to ensure customers can/will pay their loans.

DATA CLEANING & ADDITIONAL COLUMNS

All the column was of object data types.

Needed to remove the duplicate rows

The default status column, sex column and Education columns

needed to be duplicated and populated with numeric data

```
 (default =1  ; non-default = 2)

df['DEFAULT']=df['DEFAULT STATUS'].map({'default': 1, 'not default' : 2})

#(Gender: male =1; female=2)

df['GENDER']= df['SEX'].map({'male':1, 'female': 2})

# Education 'graduate school': 1,'university': 2,'high school': 3, 'other' : 4

df['ED_LEVEL']= df['EDUCATION'].map({'graduate school': 1,'university': 2,'high school': 3, 'other' : 4})
```

Additional Columns

Additional column to store discretized CREDIT LIMIT data needed to be created.

AGE_BIN column is created to store discretized age

BILL_AMT_AVG was create to store the averages of six BILL_AMT columns

PAY_AMT_AVG was create to store the averages of six PAY_AMT columns

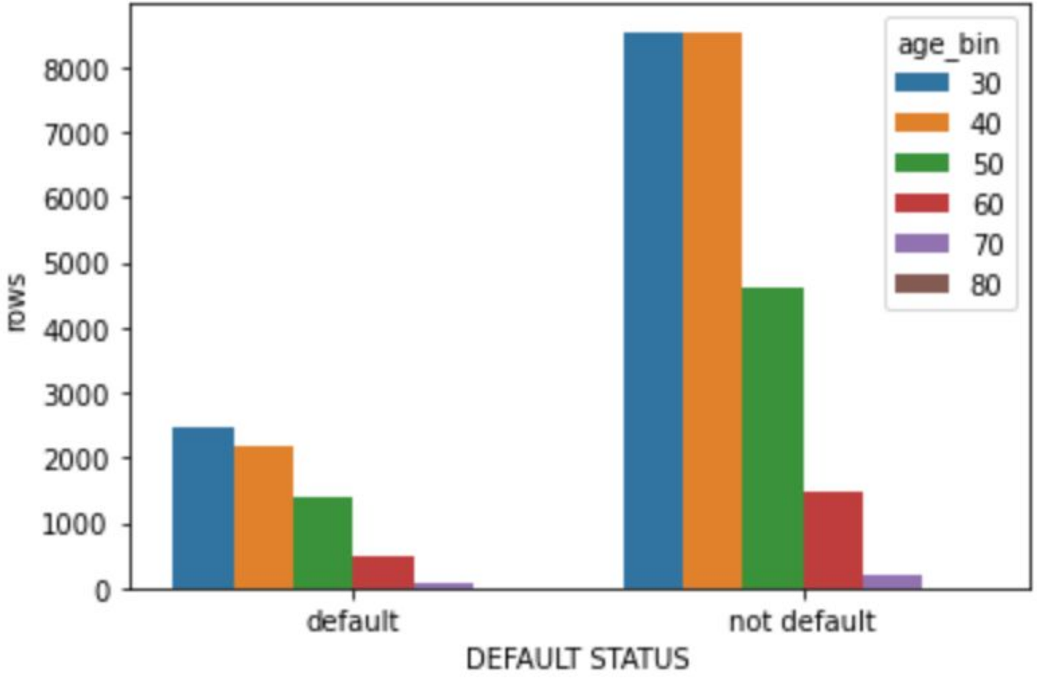
Additional columns were created to store the discretized data for the above two columns

BILL_DIFF_AVG column was created to check if it is has any correlation with default

DATAFRAME COLUMNS

```
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   ID              30000 non-null  int64
 1   LIMIT_BAL       30000 non-null  int64
 2   SEX             30000 non-null  object
 3   EDUCATION       30000 non-null  object
 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  int64
 13  BILL_AMT2       30000 non-null  int64
 14  BILL_AMT3       30000 non-null  int64
 15  BILL_AMT4       30000 non-null  int64
 16  BILL_AMT5       30000 non-null  int64
 17  BILL_AMT6       30000 non-null  int64
 18  PAY_AMT1        30000 non-null  int64
 19  PAY_AMT2        30000 non-null  int64
 20  PAY_AMT3        30000 non-null  int64
 21  PAY_AMT4        30000 non-null  int64
 22  PAY_AMT5        30000 non-null  int64
 23  PAY_AMT6        30000 non-null  int64
 24  DEFAULT STATUS  30000 non-null  object
 25  DEFAULT         30000 non-null  int64
 26  GENDER          30000 non-null  int64
 27  ED_LEVEL        30000 non-null  int64
 28  age_bin         30000 non-null  category
 29  CREDIT_LIM      30000 non-null  category
dtypes: category(2), int64(25), object(3)
memory usage: 6.5+ MB
```

EDA FOR AGE AND DEFAULT



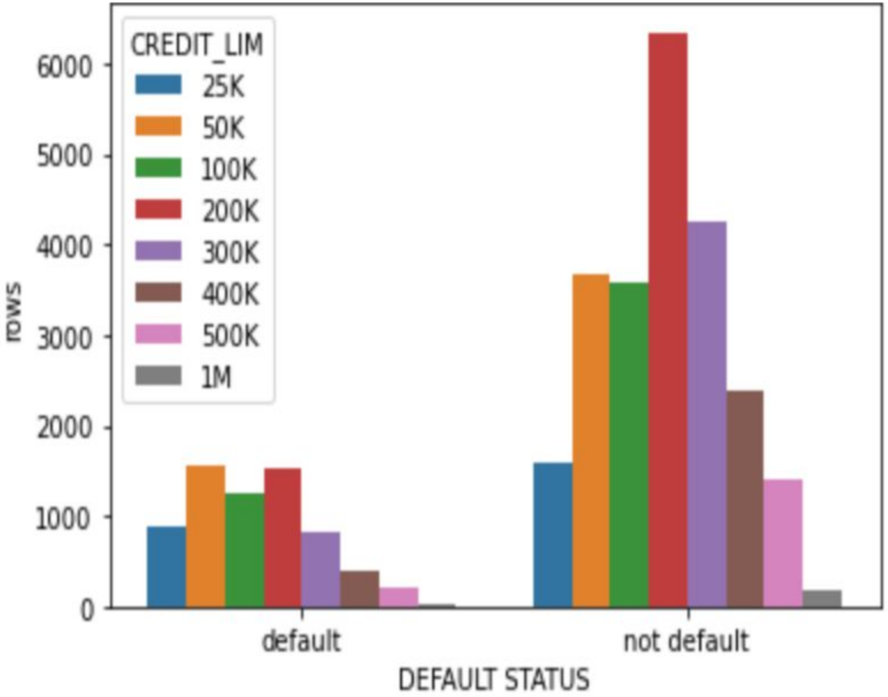
There was no relationship between age and default status. The 20-30 and 30-40 age group appears to be more in the non-default status because that age group has more more number of rows

EDA FOR AGE AND DEFAULT

	age_bin	DEFAULT STATUS	rows
0	30	default	2471
1	30	not default	8542
2	40	default	2189
3	40	not default	8524
4	50	default	1399
5	50	not default	4606
6	60	default	504
7	60	not default	1493
8	70	default	68
9	70	not default	189
10	80	default	5
11	80	not default	10

	age_bin
30	11013
40	10713
50	6005
60	1997
70	257
80	15

EDA FOR CREDIT LIMIT AND DEFAULT



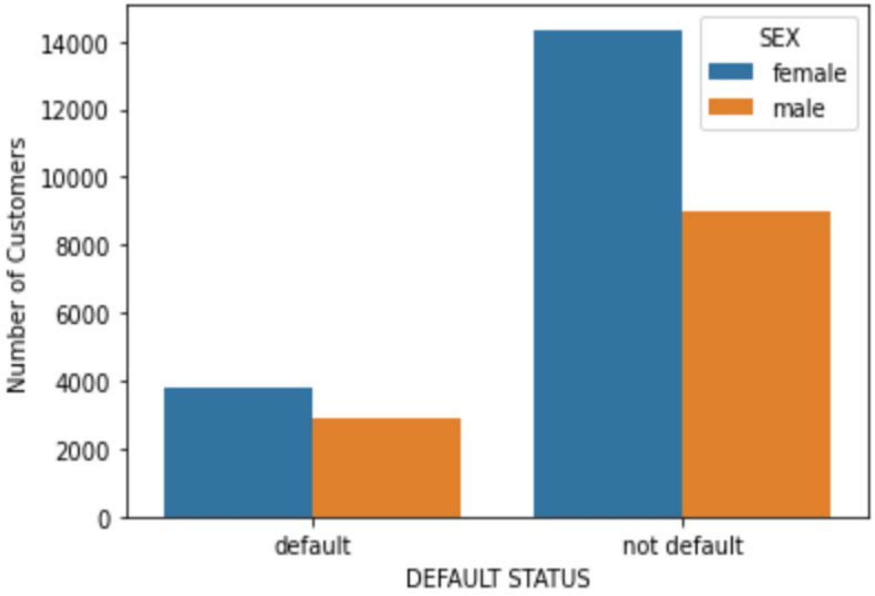
There was no relationship between age and default status.
There is more number of customers who fall between 100-200K followed by 25-50K and 50K-100K .

EDA FOR CREDIT LIMIT AND DEFAULT

	CREDIT_LIM	DEFAULT STATUS	rows
0	25K	default	895
1	25K	not default	1576
2	50K	default	1545
3	50K	not default	3660
4	100K	default	1244
5	100K	not default	3578
6	200K	default	1535
7	200K	not default	6345
8	300K	default	812
9	300K	not default	4247
10	400K	default	388
11	400K	not default	2371
12	500K	default	194
13	500K	not default	1404
14	1M	default	23
15	1M	not default	183

200K	0.262667
50K	0.173500
300K	0.168633
100K	0.160733
400K	0.091967
25K	0.082367
500K	0.053267
1M	0.006867

EDA FOR GENDER AND DEFAULT



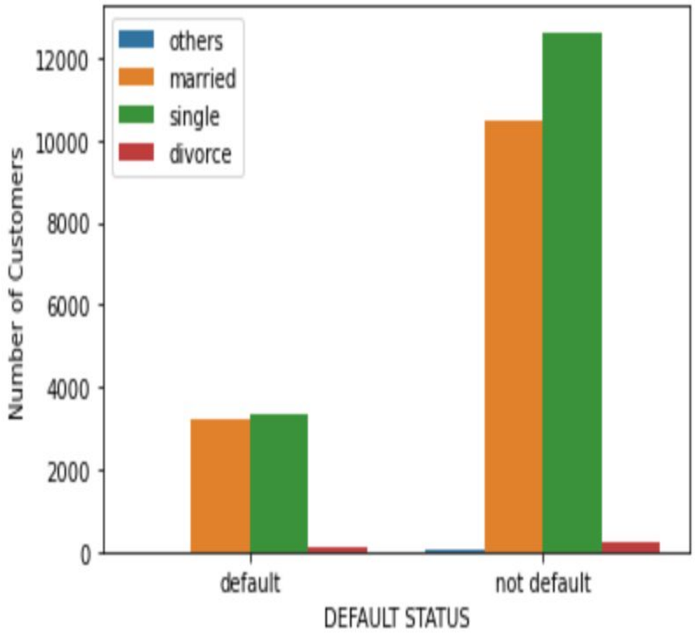
There was no relationship between gender and default status.
About 66% of the customers are women.So it appear that they are more in both the default and not default status

EDA FOR GENDER AND DEFAULT

	SEX	DEFAULT STATUS	Number of Customers
0	female	default	3763
1	female	not default	14349
2	male	default	2873
3	male	not default	9015

	SEX
female	18112
male	11888

EDA FOR MARRIAGE AND DEFAULT



There was no relationship between marriage status and default status.
There is more number of customers are married and single than others and divorced.

EDA FOR MARRIAGE AND DEFAULT

0-others,1-married,2-single, 3- divorce

	MARRIAGE	DEFAULT STATUS	Number of Customers
0	0	default	5
1	0	not default	49
2	1	default	3206
3	1	not default	10453
4	2	default	3341
5	2	not default	12623
6	3	default	84
7	3	not default	239

0-others,1-married,2-single, 3- divorce

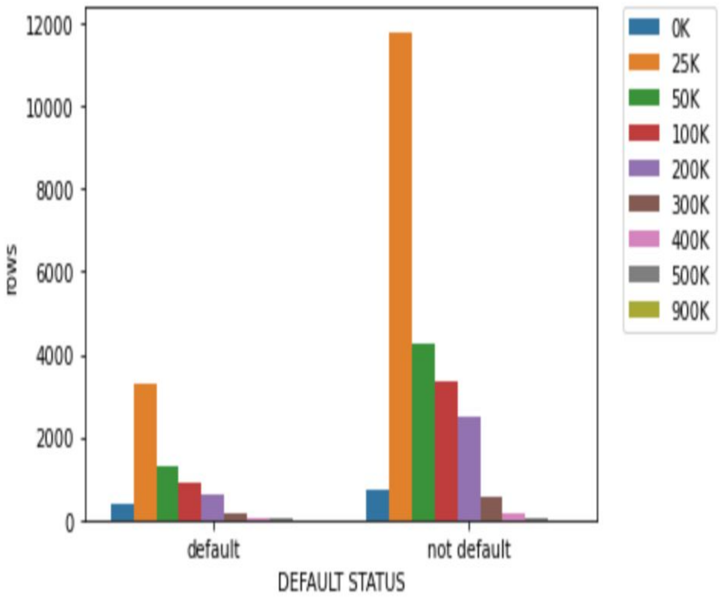
	MARRIAGE
0	54
1	13659
2	15964
3	323

EDA PAY STATUS & DEFAULT

The EDA output revealed that there are more customers who use

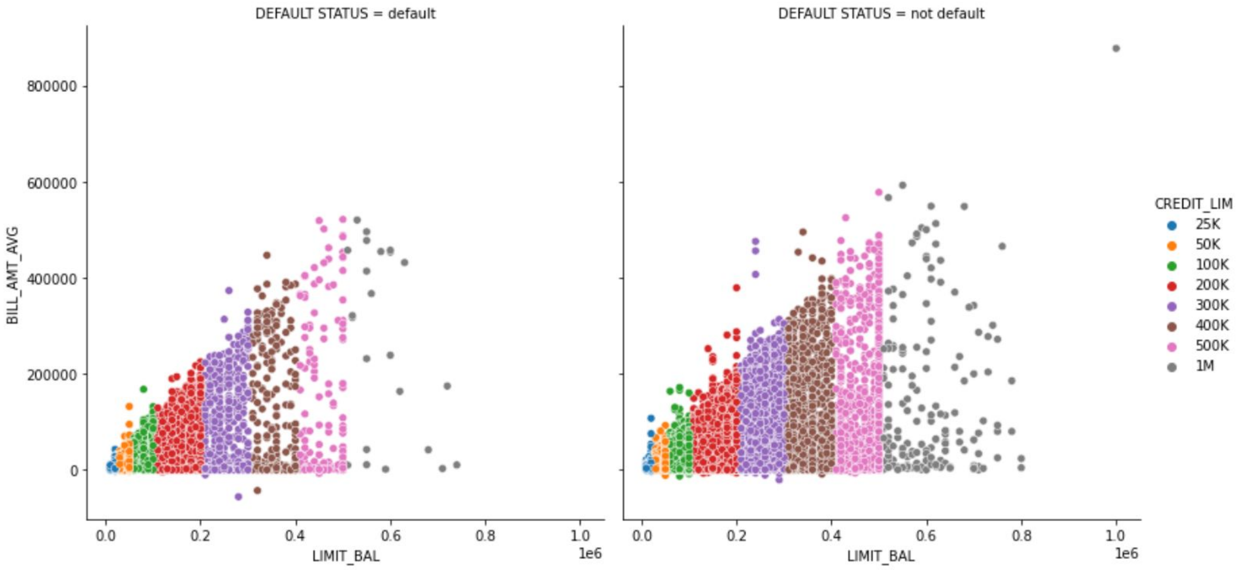
- Revolving credit followed by, (status 0)
- no consumption and then by, (status -2)
- paid in full. (status -1).

EDA AVG BILL AMOUNT VS DEFAULT

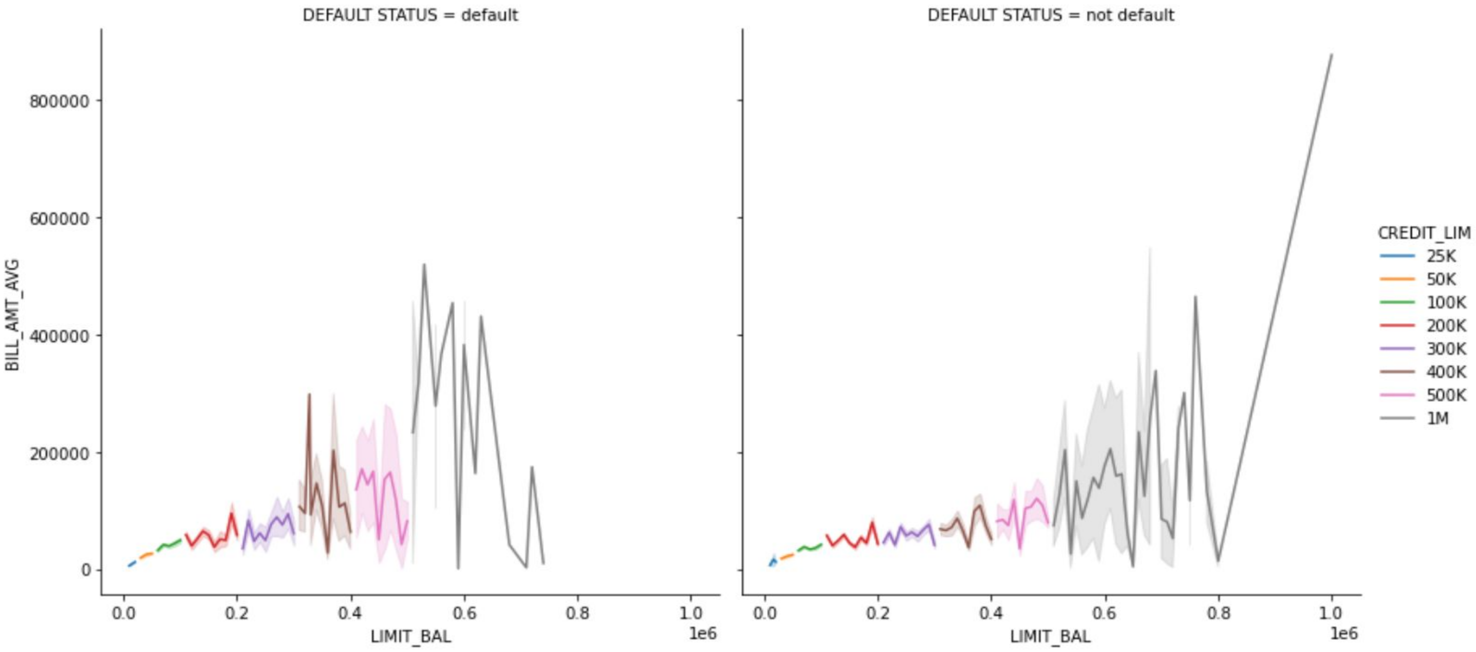


	BILL_AMT_AVG_BIN	DEFAULT STATUS	rows
0	0K	default	368
1	0K	not default	703
2	25K	default	3272
3	25K	not default	11784
4	50K	default	1300
5	50K	not default	4268
6	100K	default	872
7	100K	not default	3336
8	200K	default	588
9	200K	not default	2473
10	300K	default	154
11	300K	not default	580
12	400K	default	56
13	400K	not default	159
14	500K	default	22
15	500K	not default	52
16	900K	default	4
17	900K	not default	9

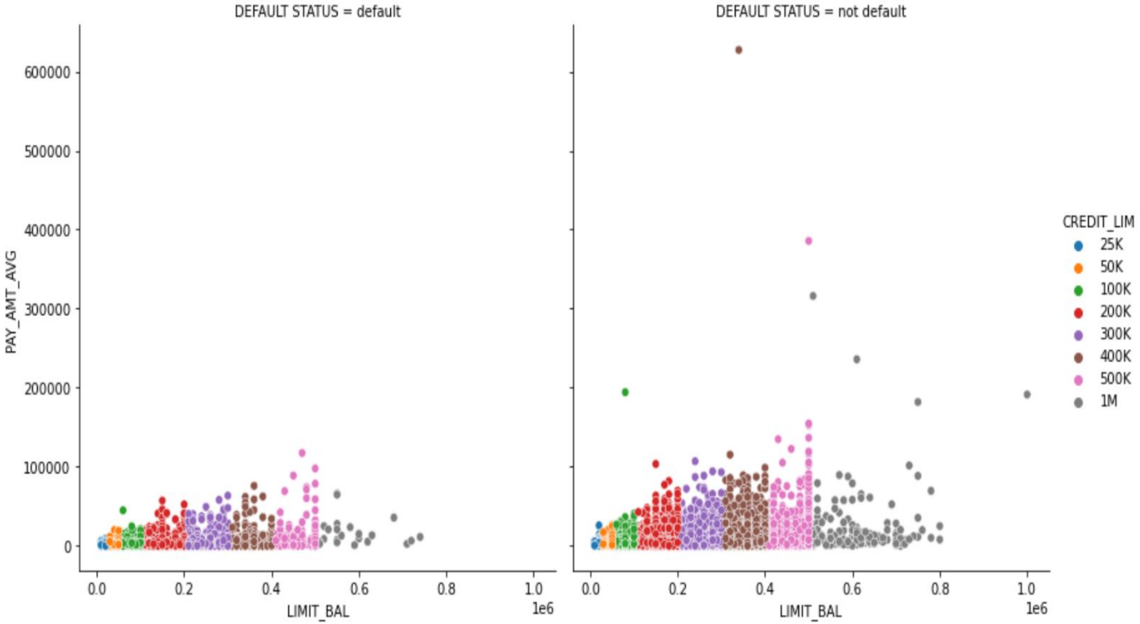
EDA AVG BILL AMOUNT VS CREDIT LMT AND DEFAULT STATUS



EDA AVG BILL AMOUNT VS CREDIT LMT AND DEFAULT STATUS

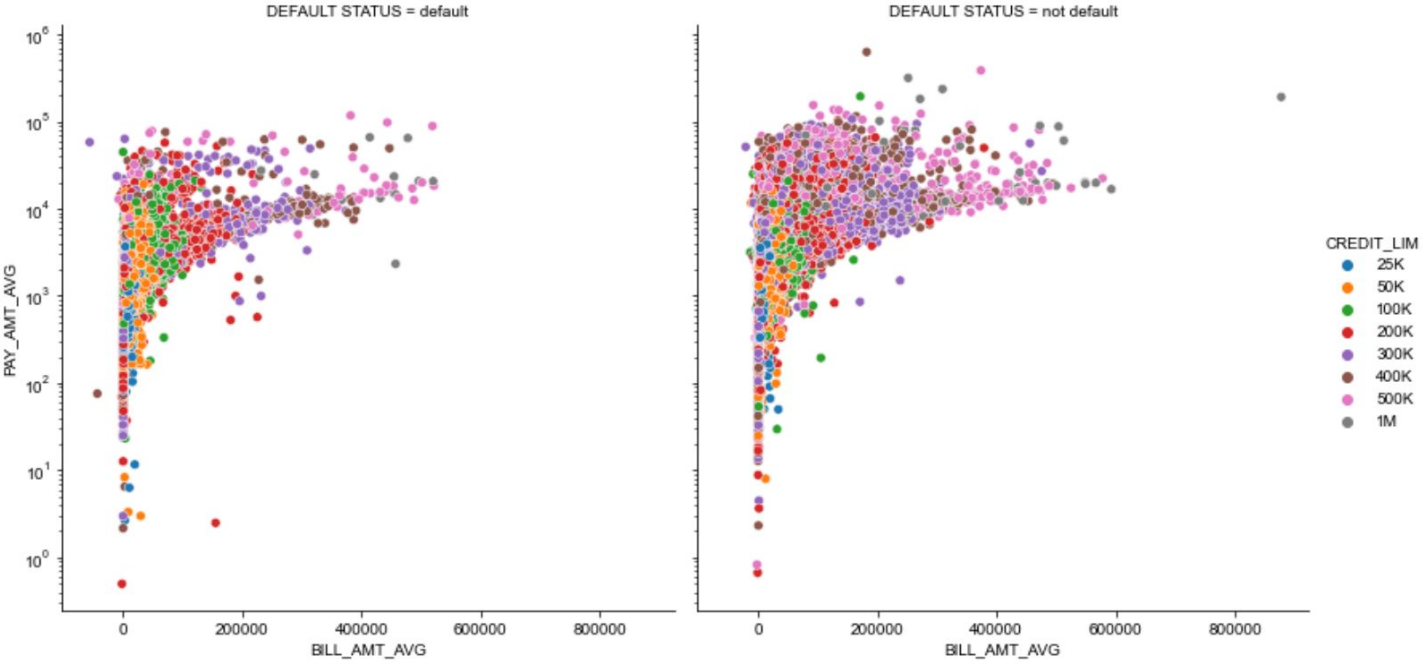


EDA AVG PAY AMOUNT VS CREDIT LMT AND DEFAULT STATUS

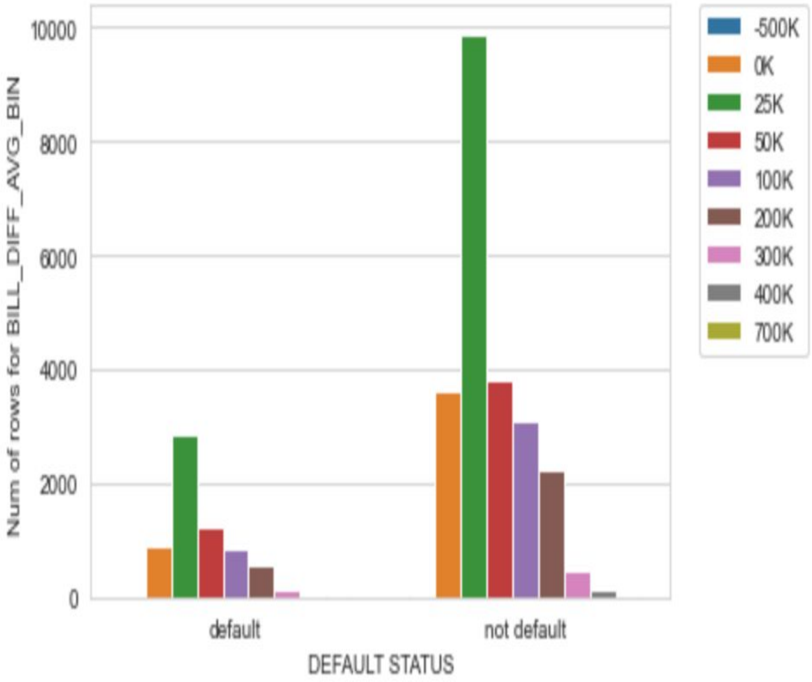


	PAY_AMT_AVG
count	30000.000000
mean	5275.232108
std	10137.946323
min	0.000000
25%	1113.290000
50%	2397.165000
75%	5583.915000
max	627344.330000

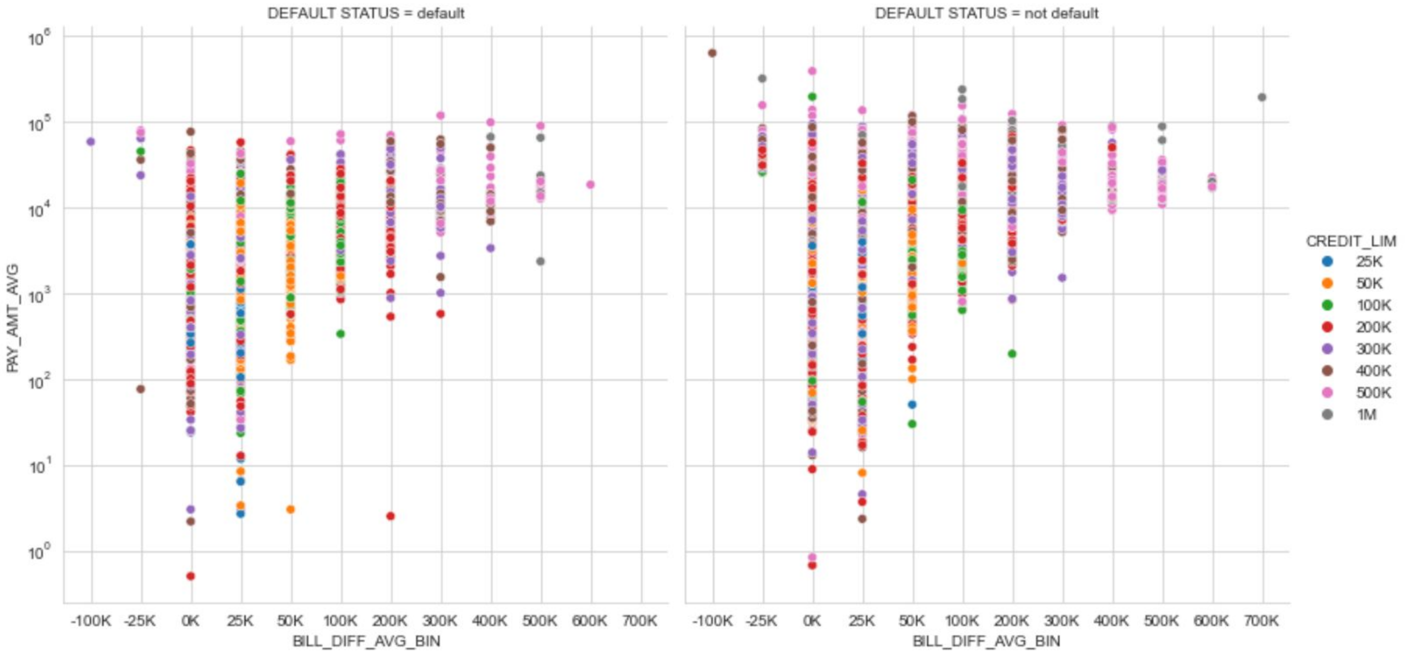
EDA AVG BILL AMOUNT VS AVG PAY AMOUNT AND DEFAULT STATUS



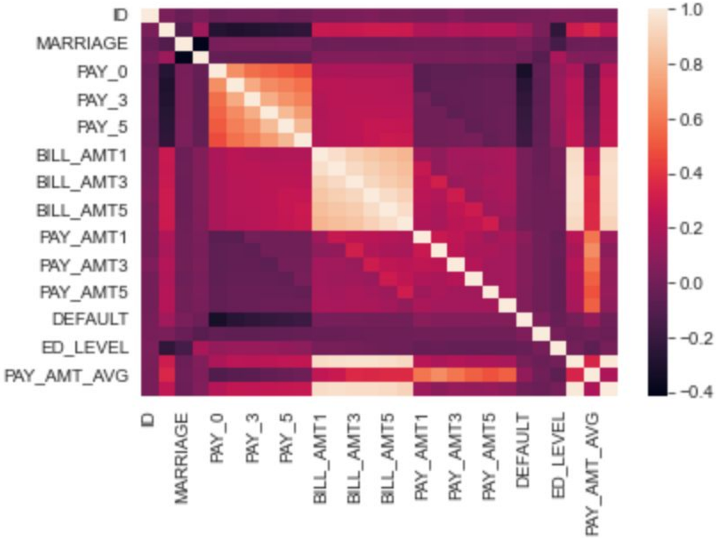
EDA BILL DIFF AVG BIN AND DEFAULT STATUS



EDA BILL DIFF AVG BIN AND PAY AMT AVG & DEFAULT STATUS



CORRELATION MAP



CORRELATION

- LIMIT_BAL is positively correlated with BILL_AMT_AVG, PAY_AMT_AVG,BILL_DIFF_AVG with 30%, 35%, 25% respectively

- LIMIT_BAL and DEFAULT is correlated 15%

- Payment status columns next to each other seems to be positively correlated at or above 0.75

  PAY_0 AND PAY_1; PAY_1 AND PAY_2; PAY_3 AND PAY_4; PAY_5 and PAY_6

RECOMMENDATION AND CONCLUSION

- Looks like the payment status are the only columns that can be used to predict the subsequent payment status.
- Instead of waiting for 8 months, Credit one can take actions within 2-3 months.