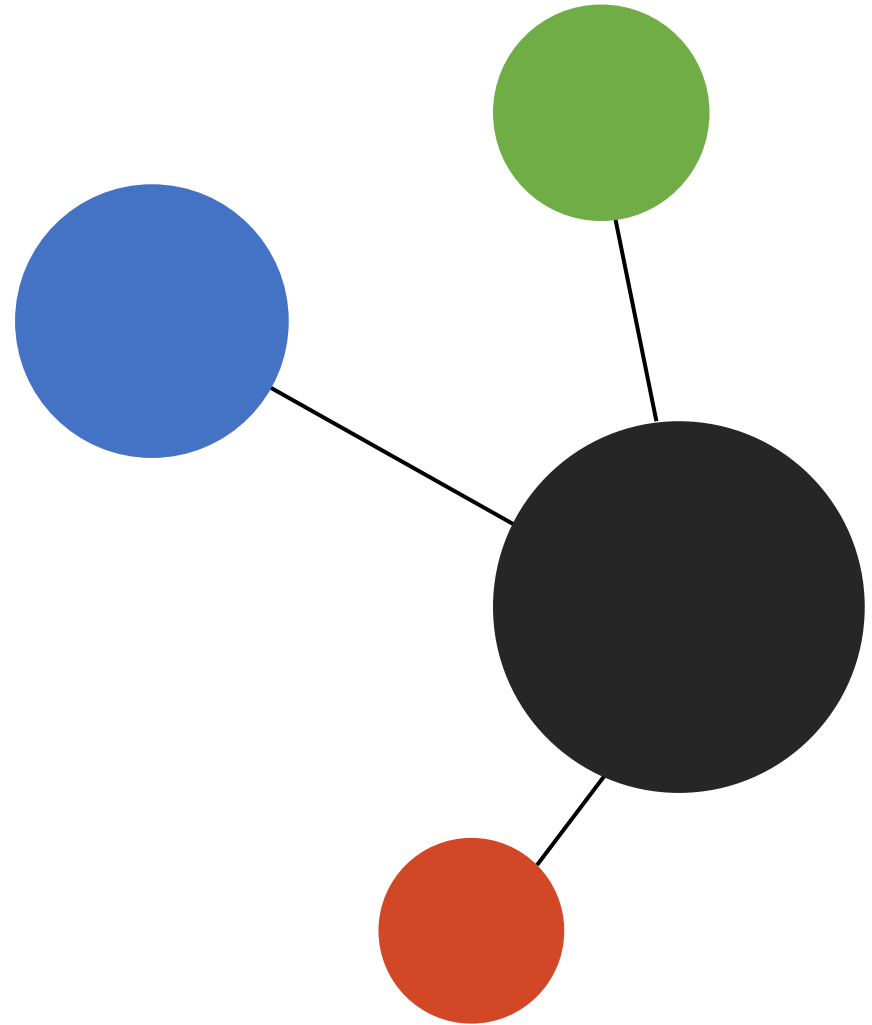


# Lending club case study

Presenter: Aditya Karnik



# Agenda

---

- Introduction
- Loan data exploration
  - Data overview
  - Data preparation and cleanup
- Exploratory Data analysis:
  - Univariate analysis
  - Bivariate analysis
  - Multivariate analysis
- Results and discussions
- Summary and Conclusion

# Introduction and problem statement

---

**Lending Club** is one of the largest online marketplace where people can apply for the loan at lower interest rate. Company wants to implement low risk approach while approving loans but at the same time does not want to lose the business.

This case study aims at providing insights by analysing historical data for approved loans which contains information about the charged off (defaulted) loans.





# Loan Data Descriptions

- This is a csv file having **39717 rows and 111 columns**
- This dataset contains Different categorical columns which are order categorical and unordered categorical variables and numerical variables.
- Columns with all NA values: 54
- Columns with NA values more than 30% but not all null:  
['desc', 'mths\_since\_last\_delinq', 'mths\_since\_last\_record', 'next\_pymnt\_d']
- Other columns with Null values:

emp_title	2459
emp_length	1075
title	11
revol_util	50
last_pymnt_d	71
last_credit_pull_d	2
collections_12_mths_ex_med	56
chargeoff_within_12_mths	56
pub_rec_bankruptcies	697
tax_liens	39



## Important Categorical variables:

- 1) Grades and subgrades
- 2) Issue date
- 3) Employment duration
- 4) Home ownership status
- 5) Loan verification status
- 6) Loan Purpose
- 7) Address state
- 8) Loan status
- 9) Derogatory public record

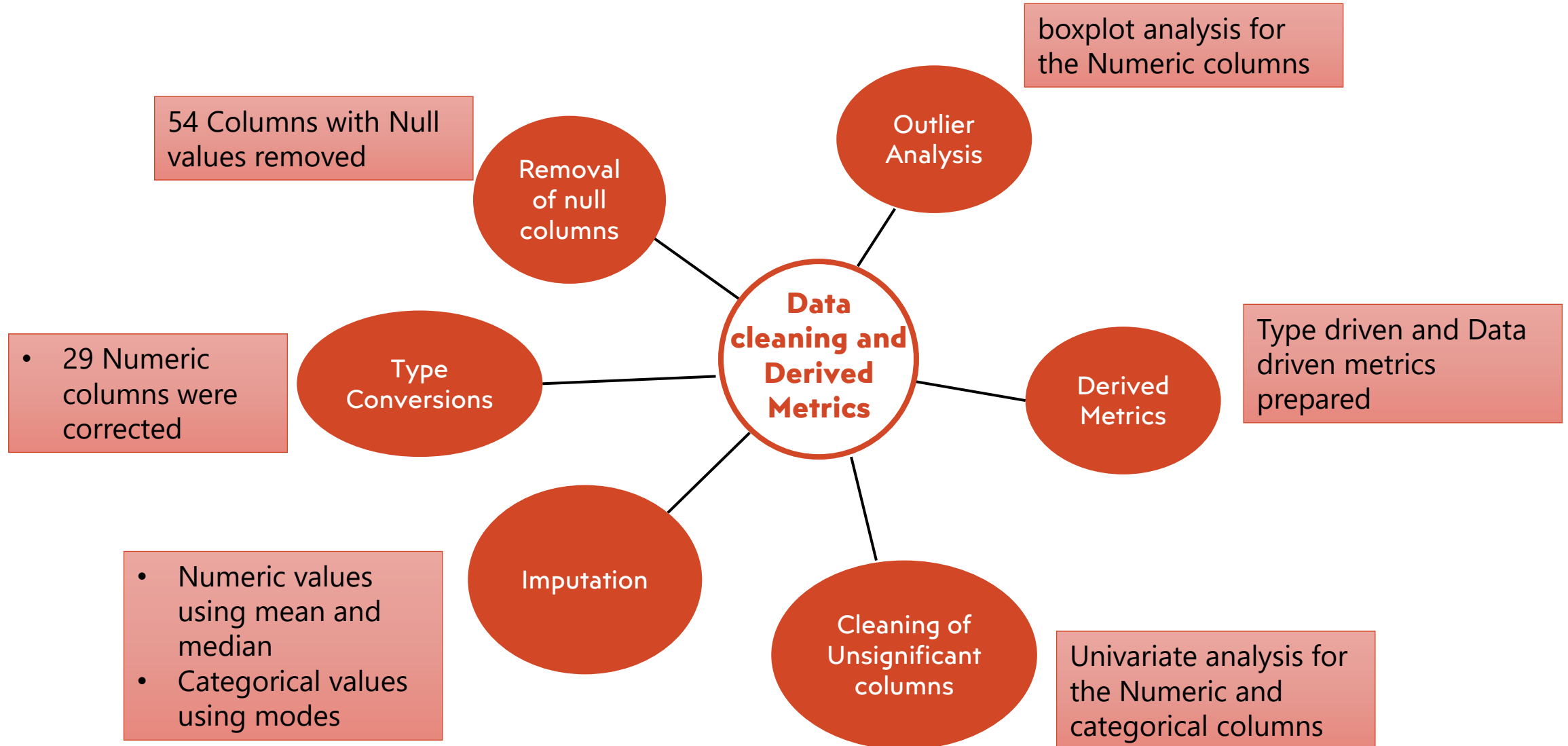


## Important facts column:

- 1) Debt to income ratio
- 2) Annual income
- 3) Loan amount
- 4) Recoveries

# Data cleaning and Derived metrics

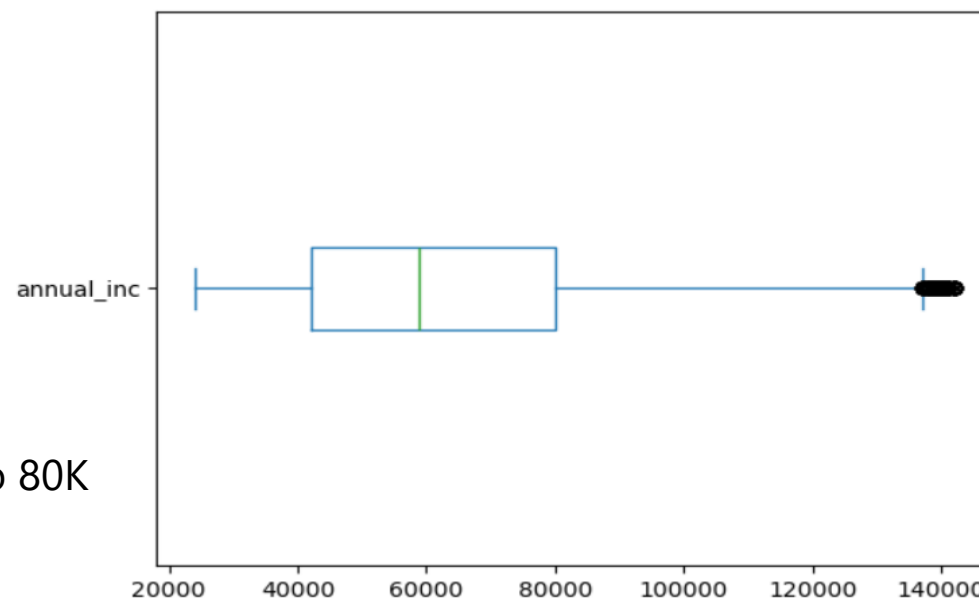
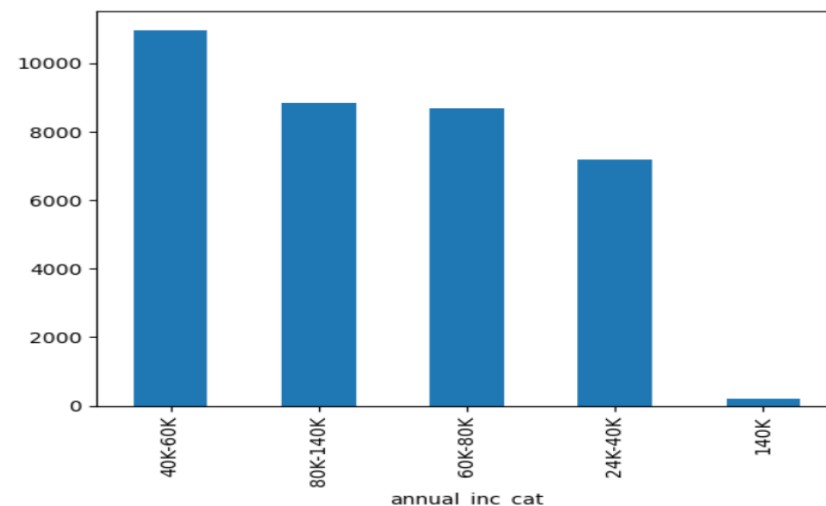
---



# Annual income

```
count    35940.000000
mean     63240.818464
std      26722.628739
min       24000.000000
25%      42000.000000
50%      58898.000000
75%      80000.000000
max      142000.000000
```

**Note:** Annual income is ranging between 20K to 80K



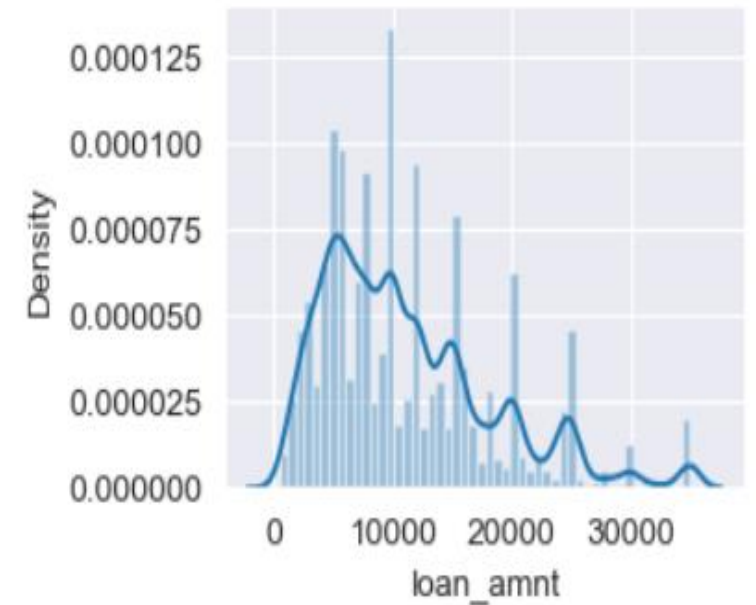
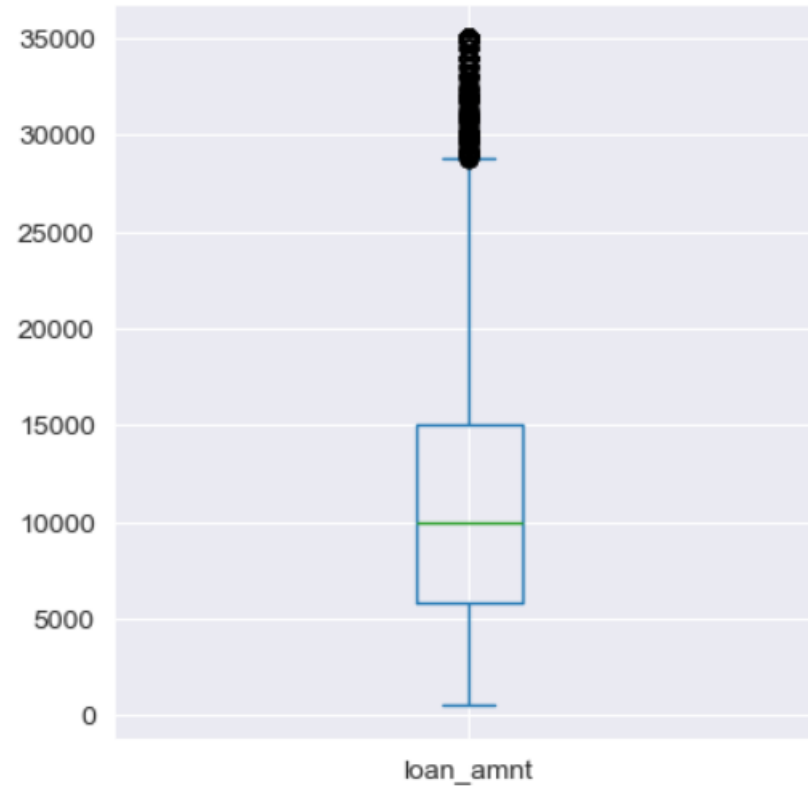
# Loan Amount

---

Descriptive analysis for loan\_amnt

count	35940.000000
mean	11170.496661
std	7162.391367
min	500.000000
25%	5825.000000
50%	10000.000000
75%	15000.000000
max	35000.000000

Name: loan\_amnt, dtype: float64



**Note:** Maximum loan amounts are ranging between 5K to 15K

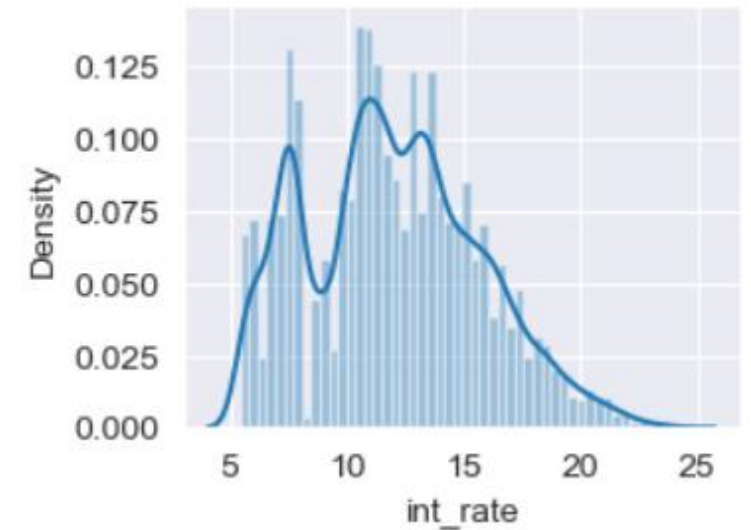
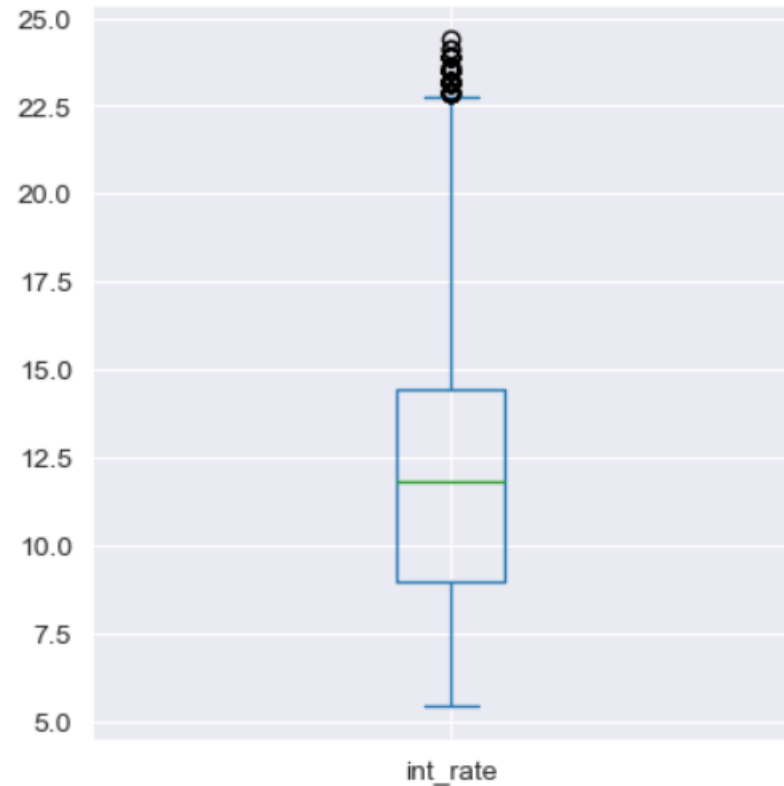
# Interest Rate

---

Descriptive analysis for int\_rate

```
count    35940.000000
mean      11.984632
std       3.712117
min       5.420000
25%       8.940000
50%      11.830000
75%      14.460000
max      24.400000
```

Name: int\_rate, dtype: float64



**Note:** Maximum interest rates are ranging between 8% to 15%  
But there are interest rates **higher than 20% till 24.4%**

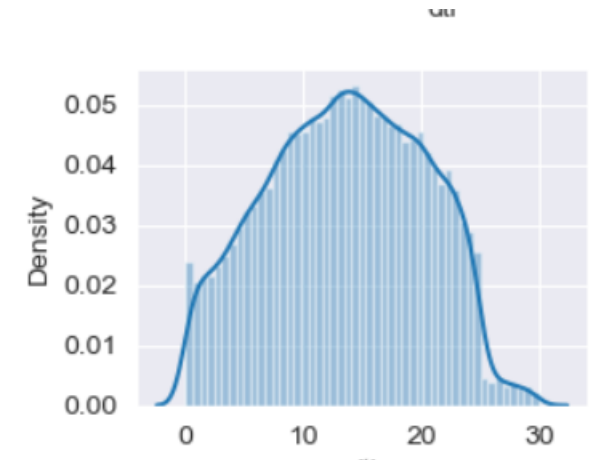
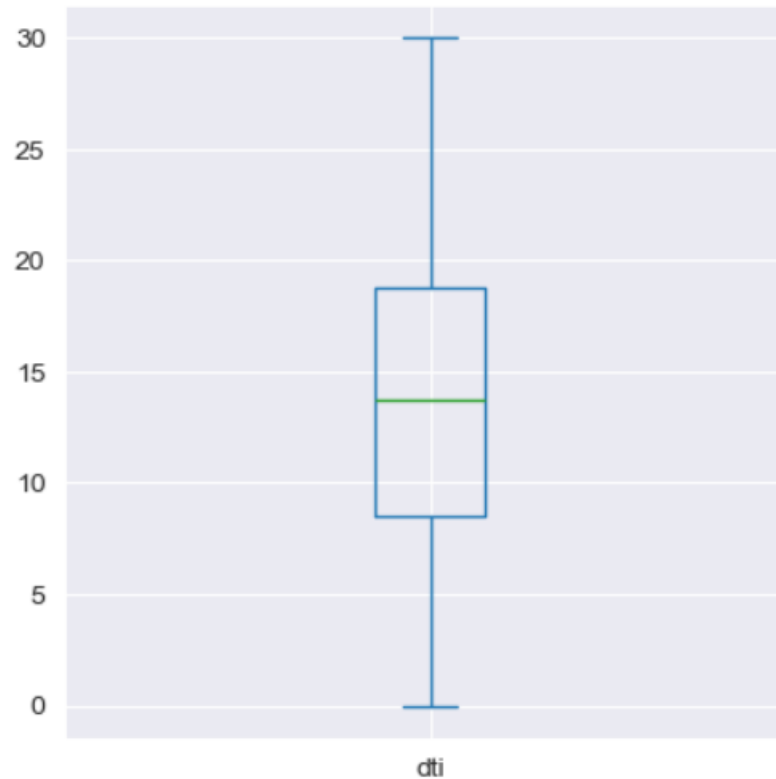


# Debt to Income ratio

---

Descriptive analysis for dti

```
count    35940.000000
mean      13.555674
std        6.640927
min        0.000000
25%        8.500000
50%       13.700000
75%       18.800000
max       29.990000
Name: dti, dtype: float64
```



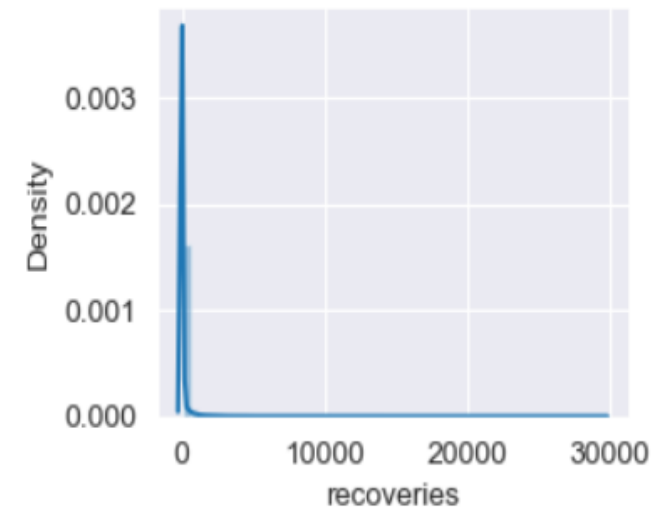
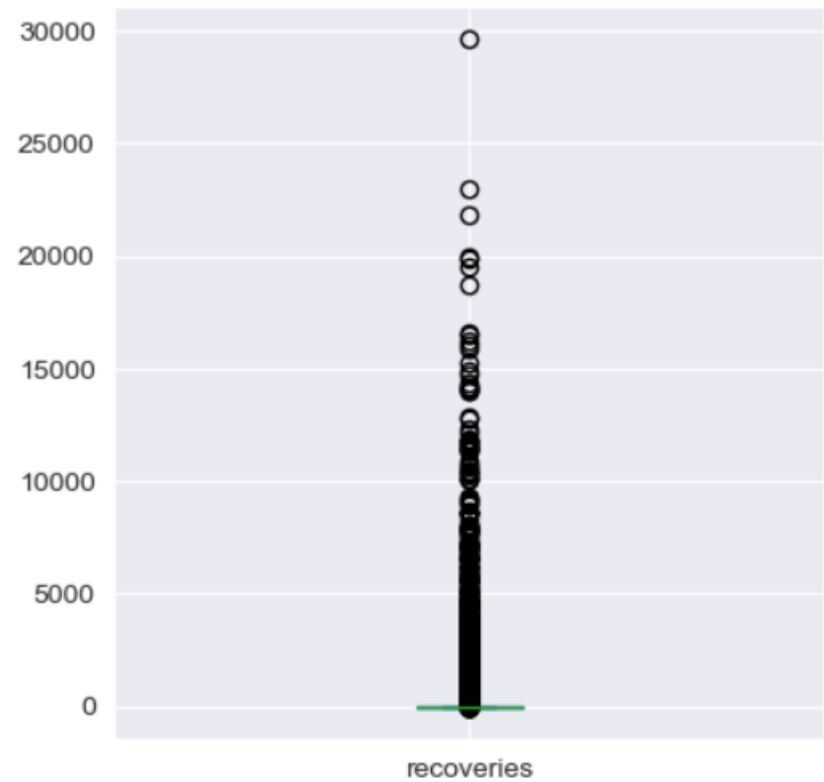
**Note:** Maximum debt to income are ranging between 8 to 19.  
But there are DTI **higher than 20, whereas some of them are 30**

# Recoveries

---

Descriptive analysis for recoveries

count	35940.000000
mean	94.276378
std	682.737432
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	29623.350000

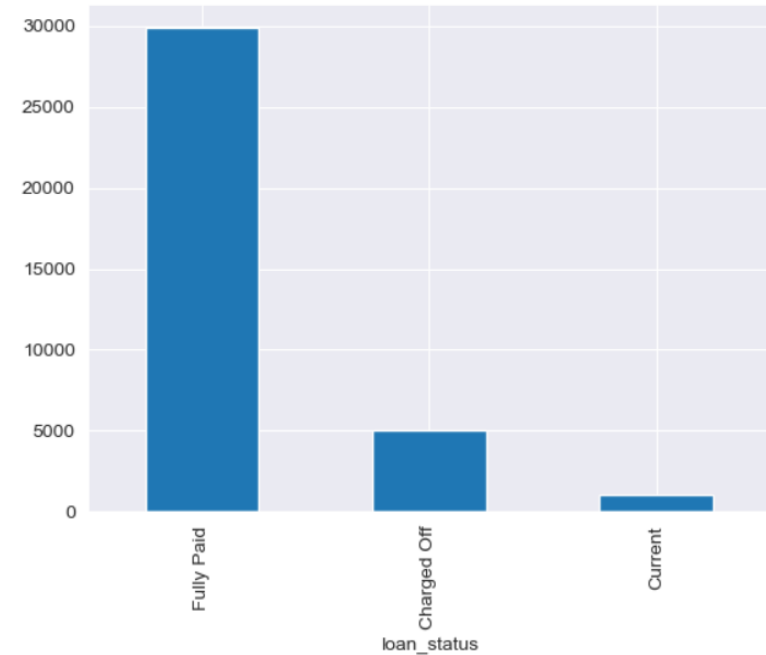
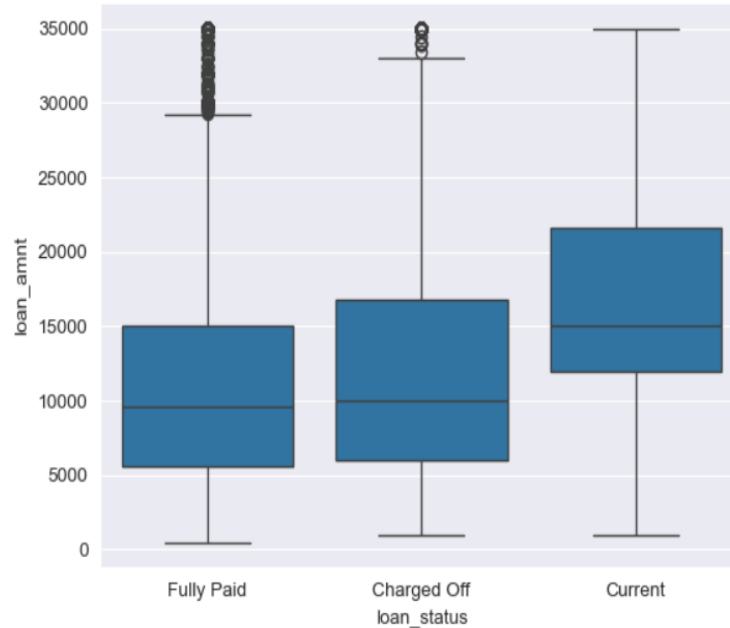


**Note:** No significant recoveries of the amounts

# Loan status and Loan amounts

---

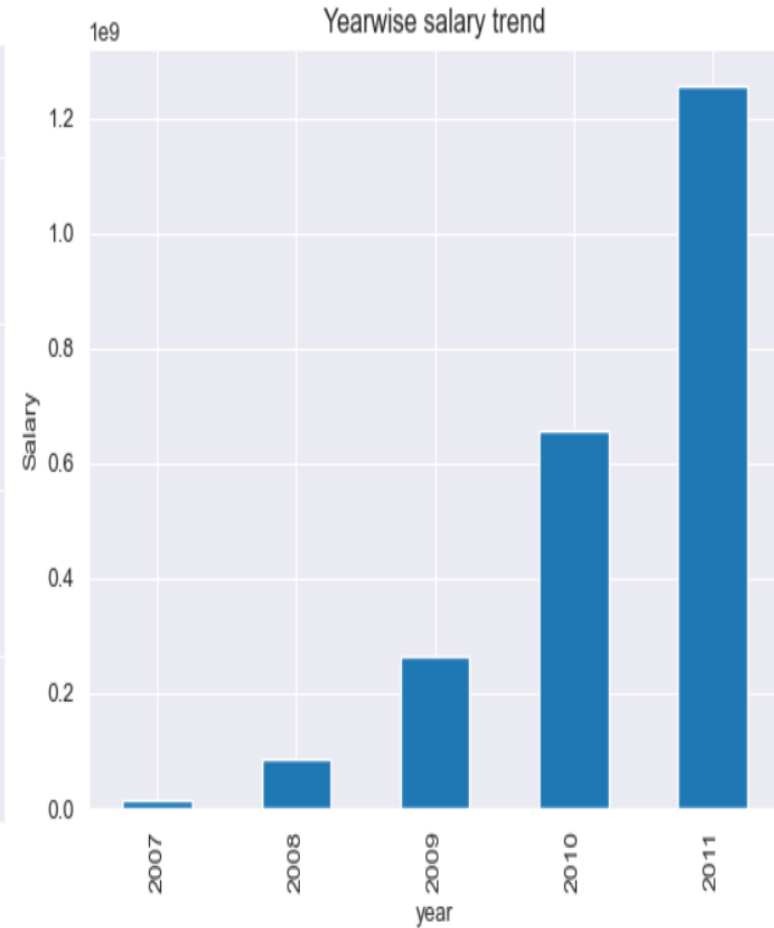
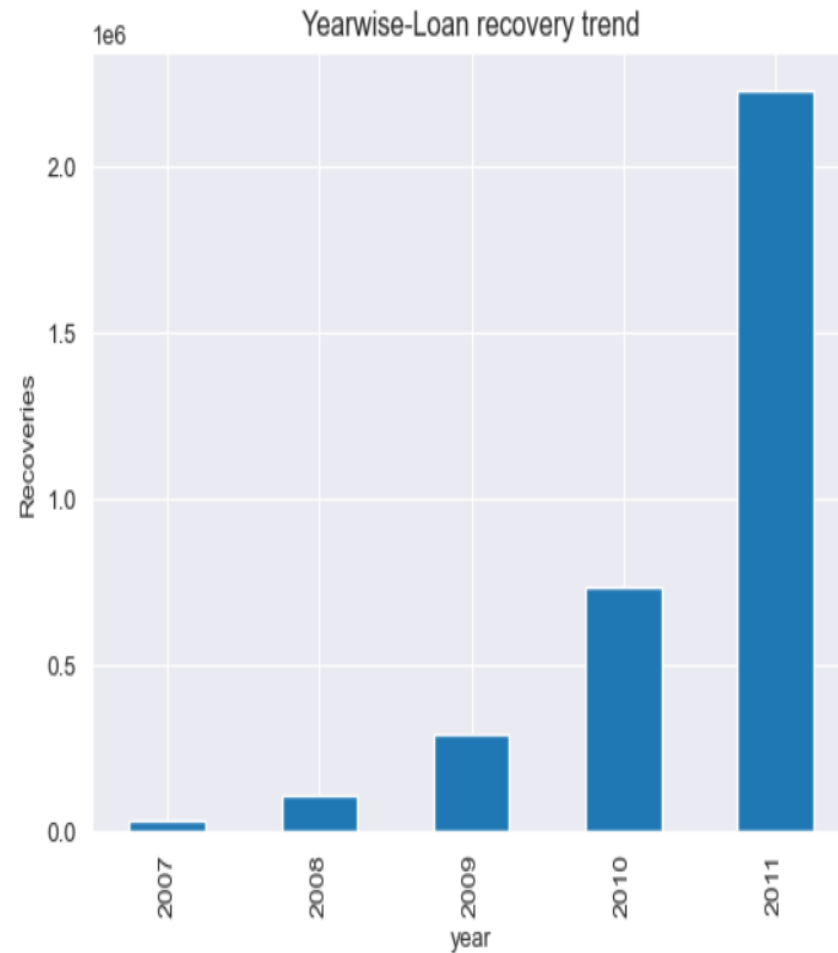
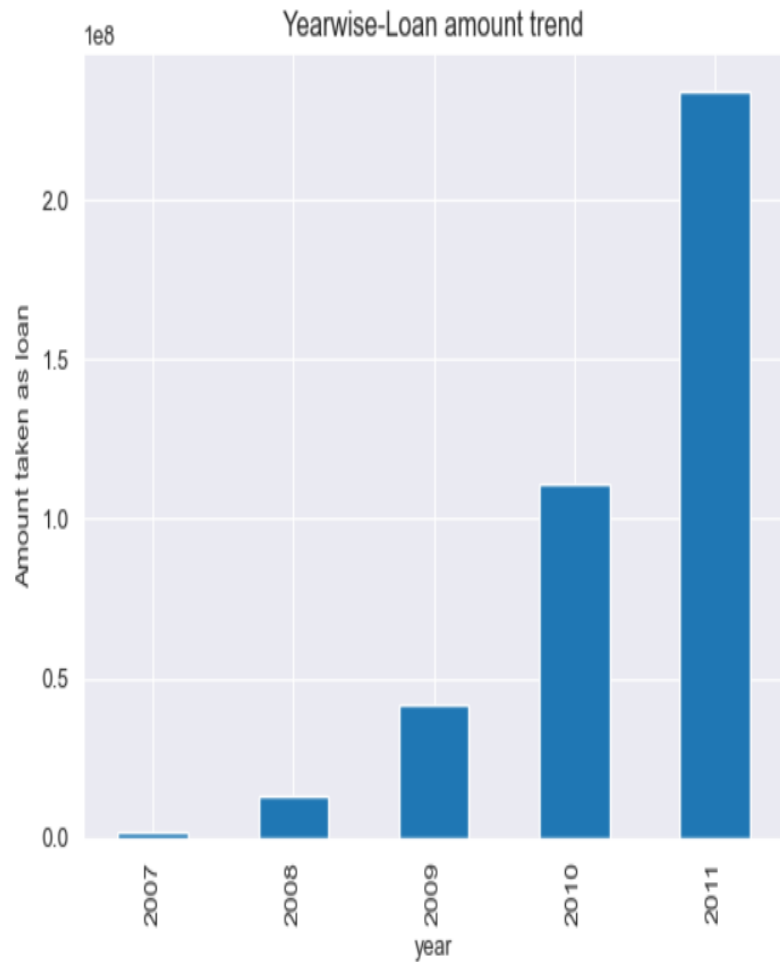
```
loan_status
Charged Off    62450075
Current        17249200
Fully Paid     321768375
Name: loan_amnt, dtype: int64
```



**Note:** Maximum defaulted loans ranged between 6K to 17K loan amounts. Whereas the fully paid loans are also ranging between the same. Count of applicants repaying the loan is more than the defaulted ones.

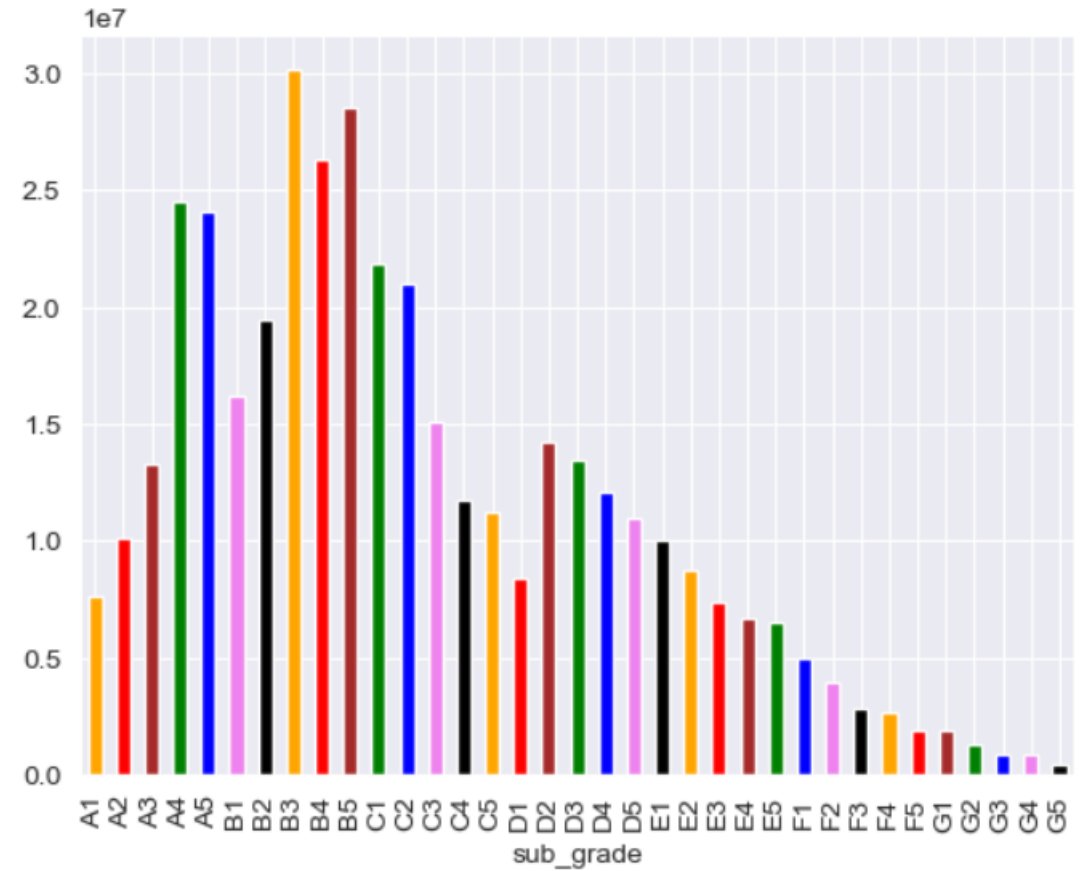
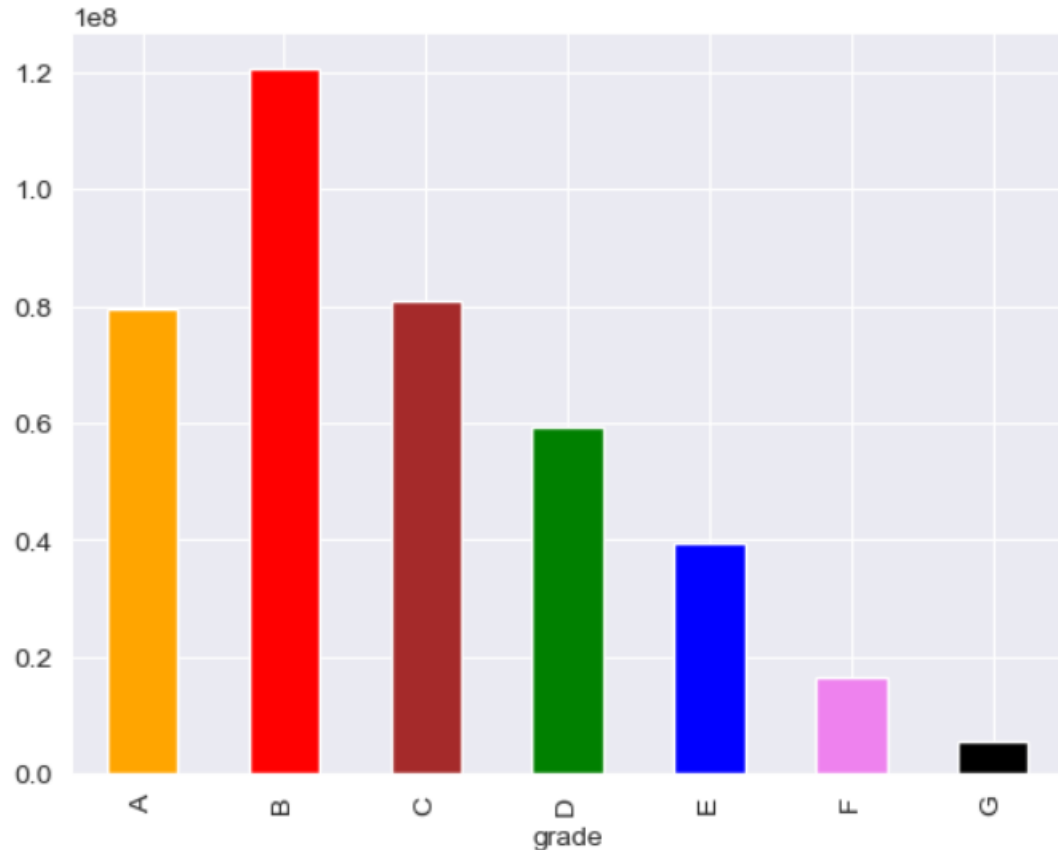
# Year Vs Loan amount, Recoveries and Annual income

---



**Note:** As the purchasing power(income) is increasing with year. In 2011 it increased significantly. There is trend of taking more loans. There are some recoveries also seen but the amounts are insignificant.

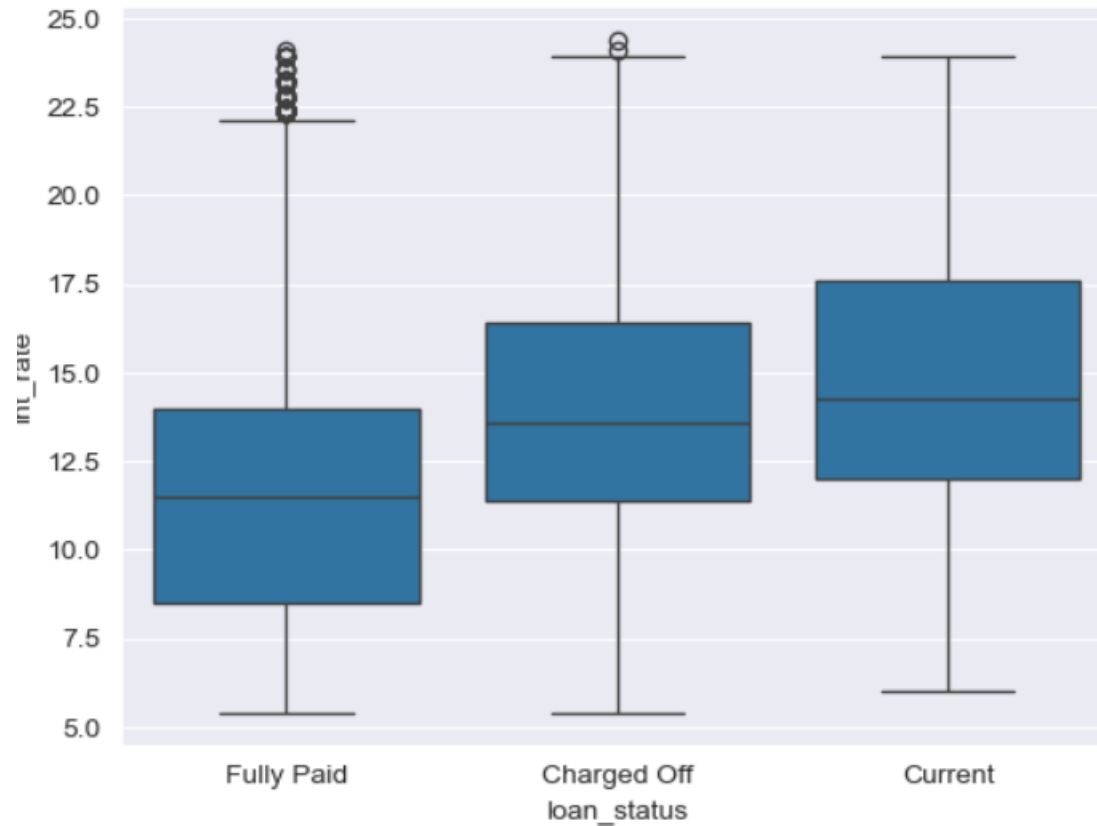
# Grade Vs loan amounts



**Note:** Grade B loans are applied more and then Grade C and Grade A. Subgrades A4, A5, B3-B5, C1-C3 and D2 are more popular.

# Interest Rates and Loan status

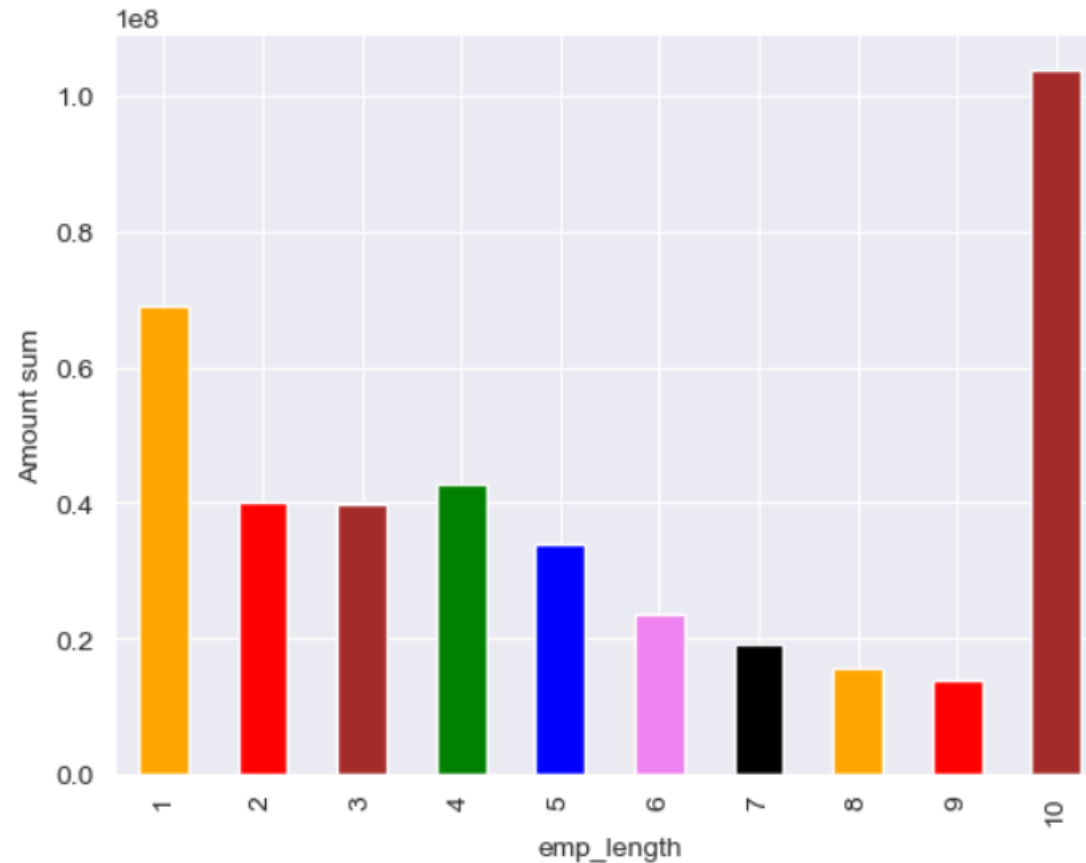
---



**Note:** The loans between 12% interest rates to 16% interest rates are more Prone to be defaulted compared to lower interest rates.

# Employee Length Vs loan amount

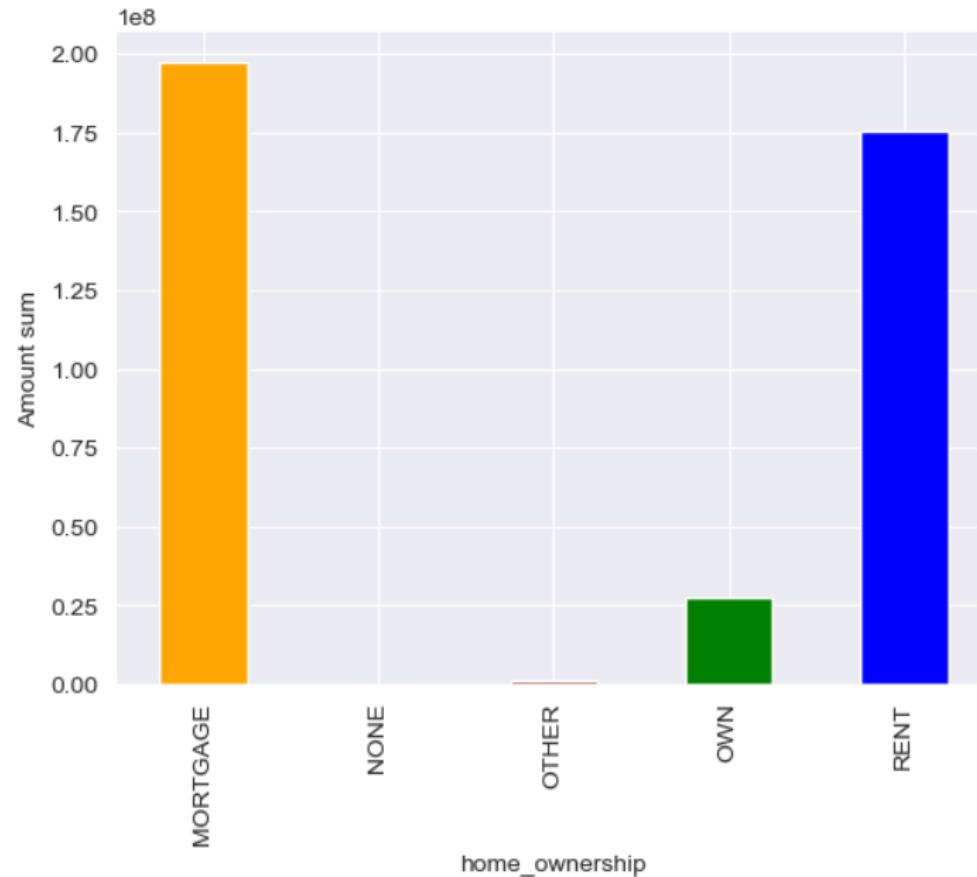
---



**Note:** Employees having experience more than 10 years and less than 1 year take more loans compared to others.

# Home ownership Vs Loan Amount

---

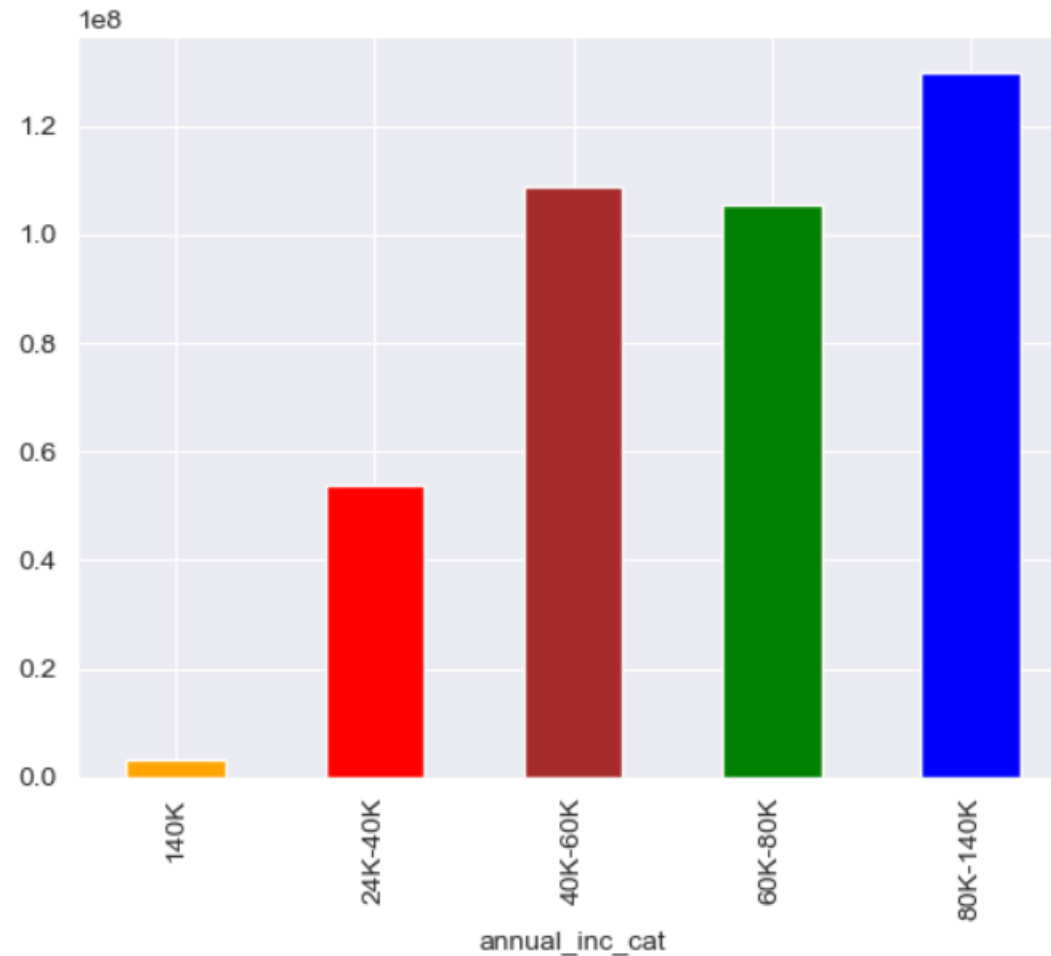


**Note:** People with Mortgaged home and Rented home are more likely to take loans.



# Annual income Vs Loan amount

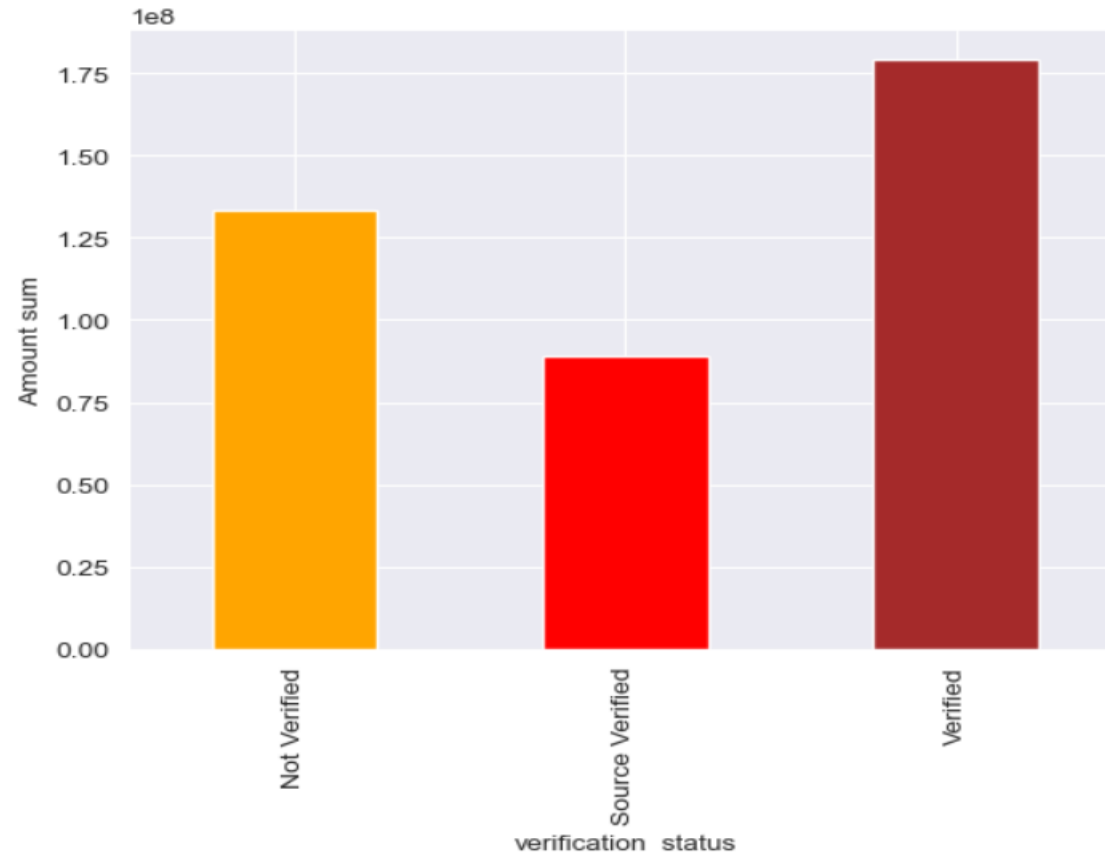
---



**Note:** Applicants with Annual income more than 40K contribute to major loan portfolio

# Verification Status Vs Loan Amount

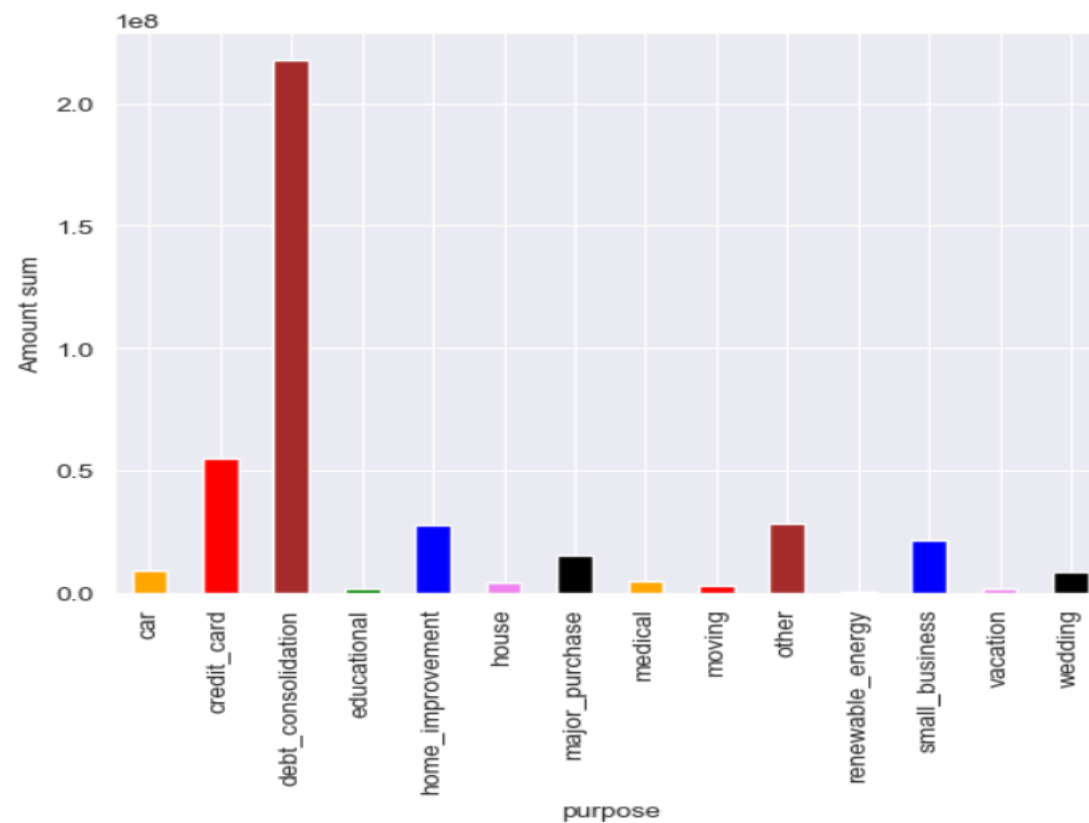
---



**Note:** Applicants whose verification is done have taken more loans.

# Purpose Vs Loan amounts

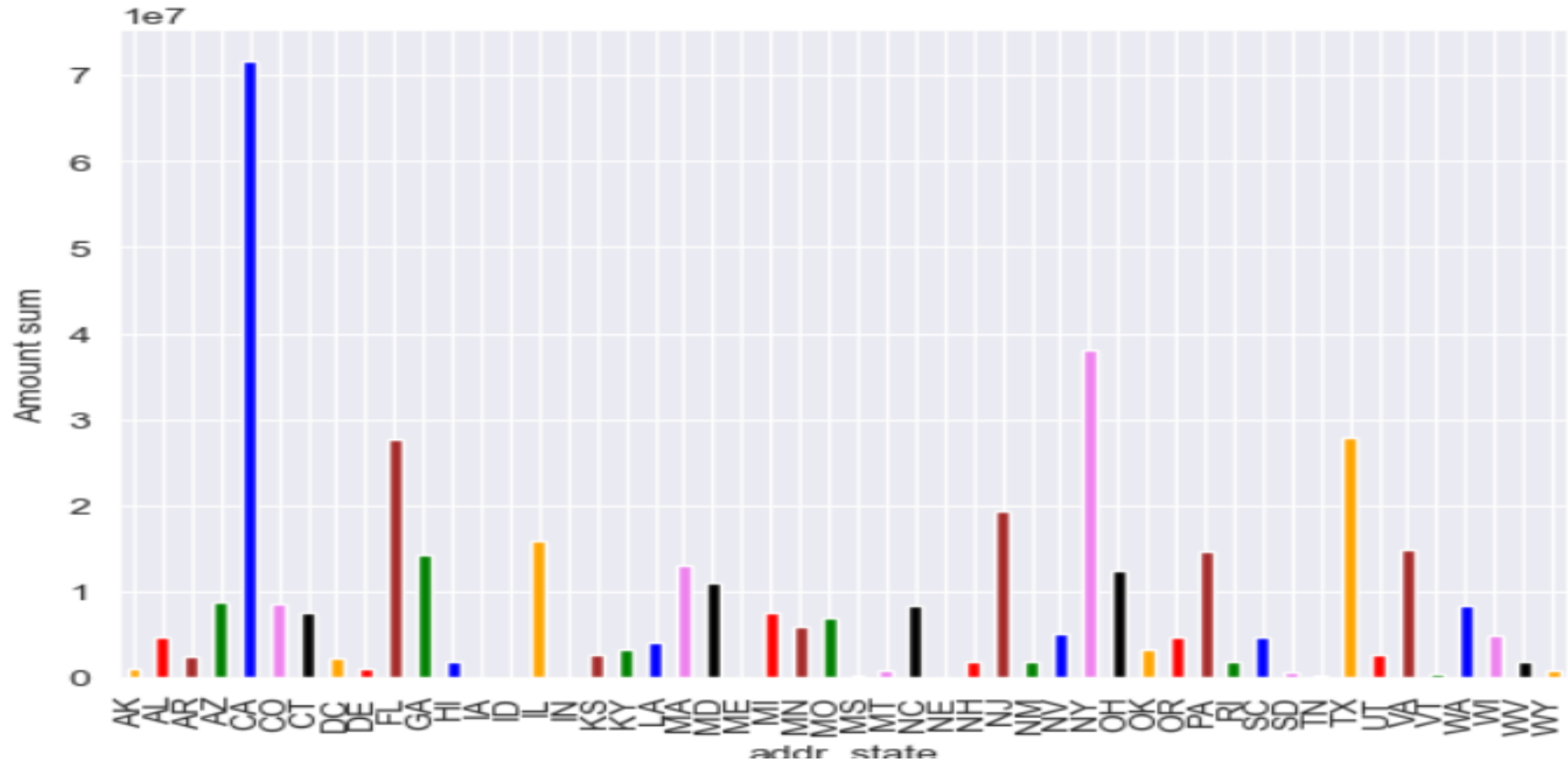
---



**Note:** Applicants take more loan for Debt consolidation

# Address State Vs Loan amount

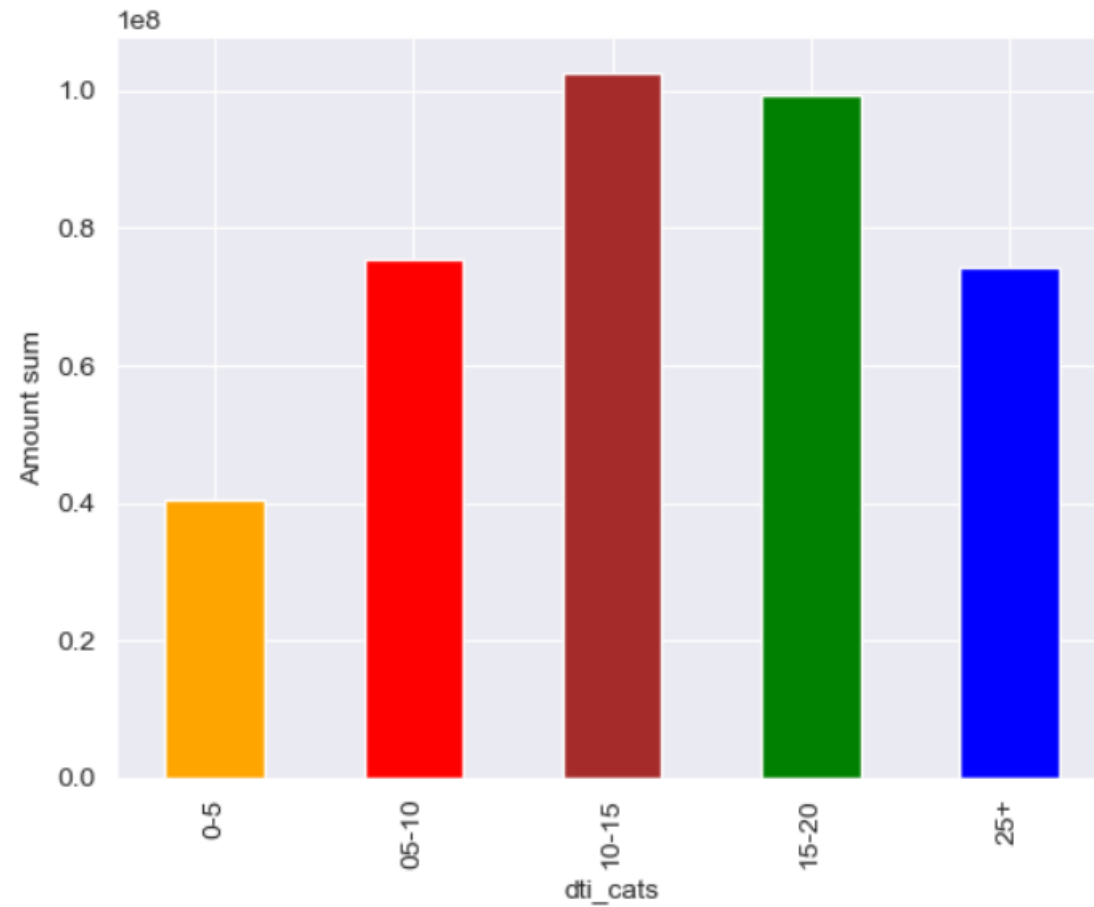
---



**Note:** Applicants from State CA, FL, NY, TX take more loan than other states

# Debt to Income Vs Loan amounts

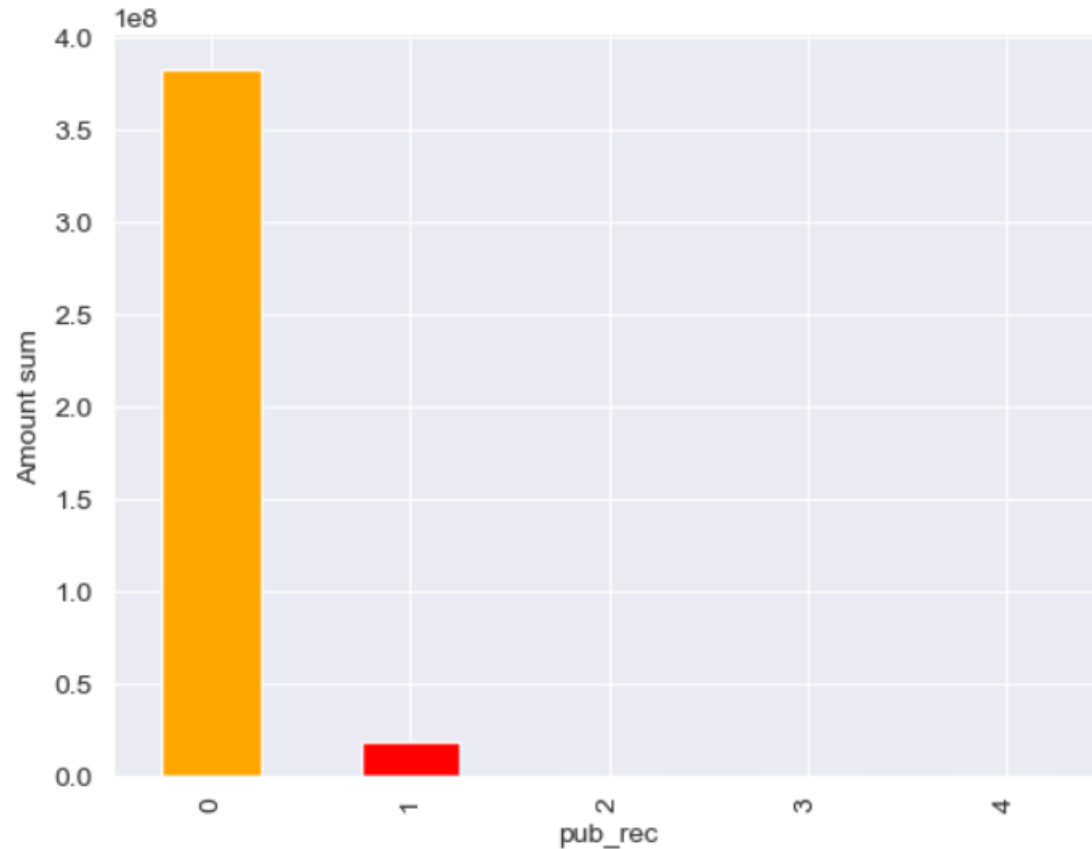
---



**Note:** Applicants with DTI 10 to 20 take more loans

# Derogatory public record Vs Loan amount

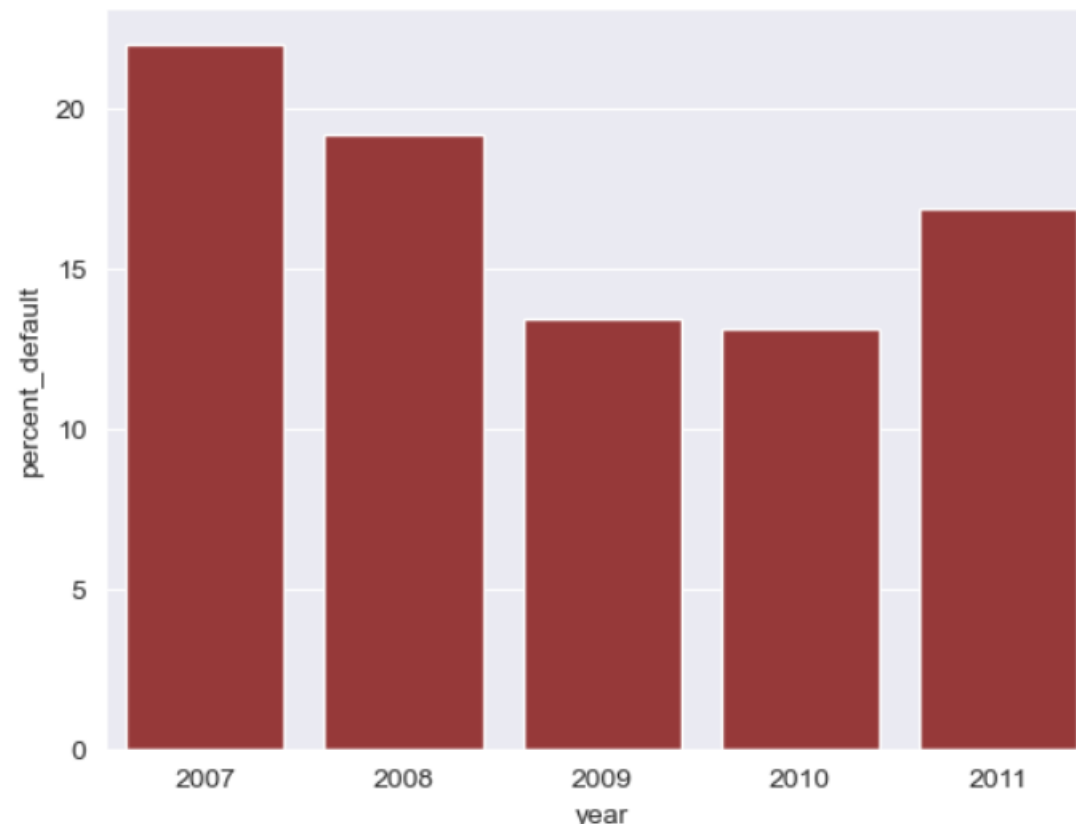
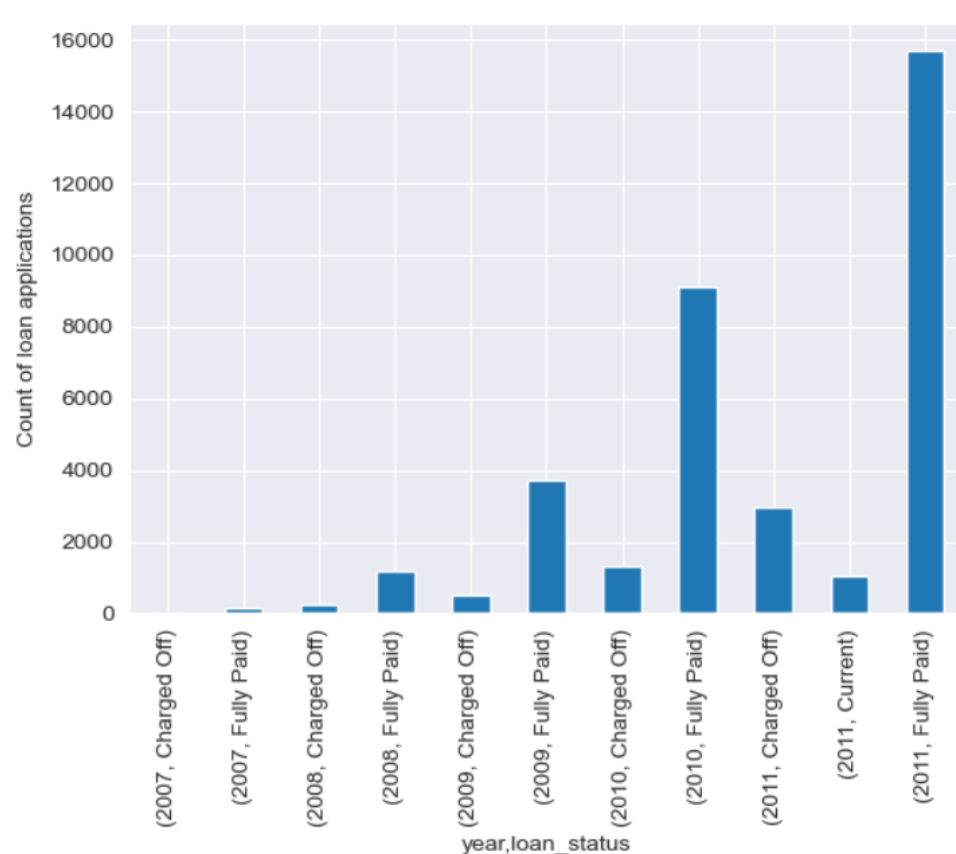
---



**Note:** Applicants with 0 derogatory records take loans, with 3 or 4 derogatory record do not take loans.

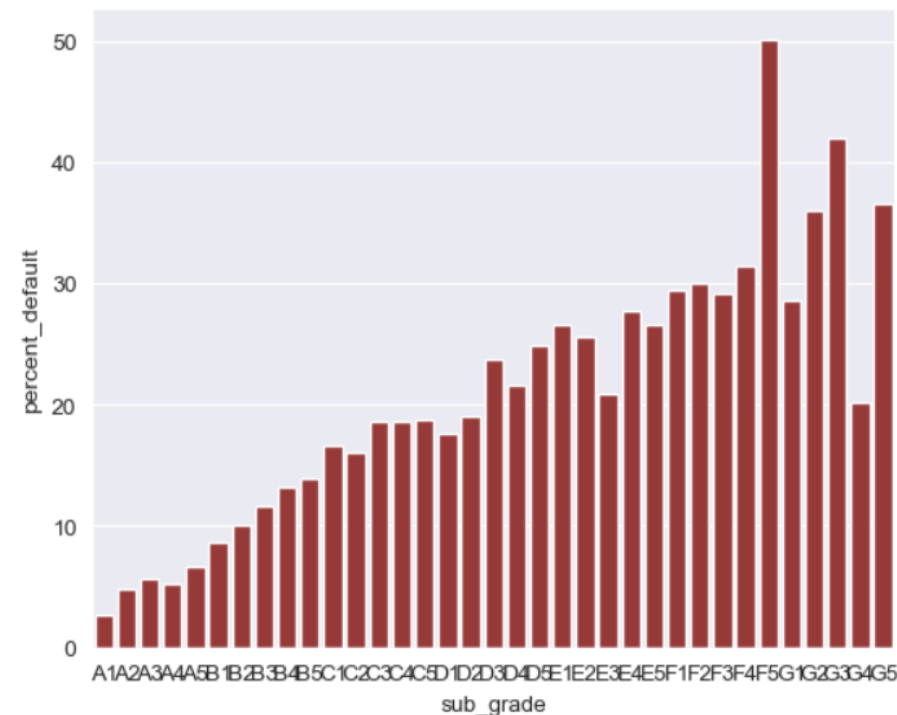
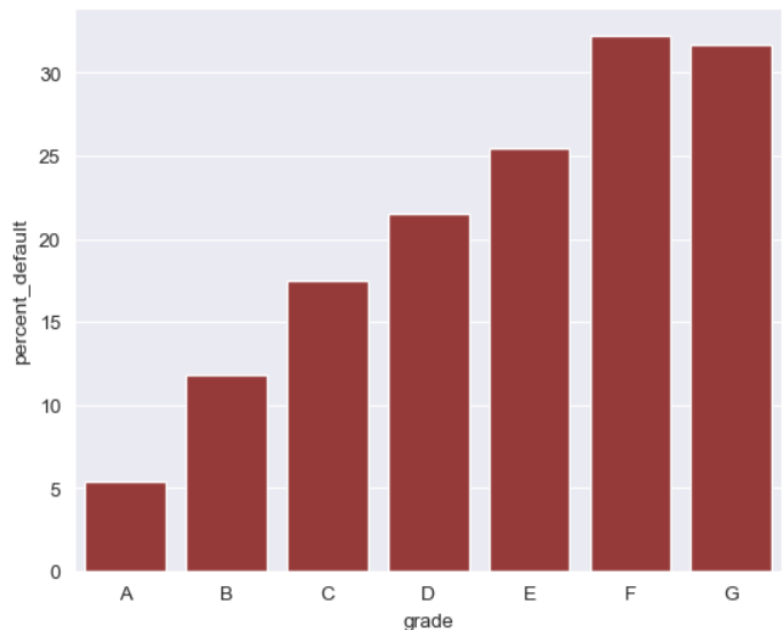
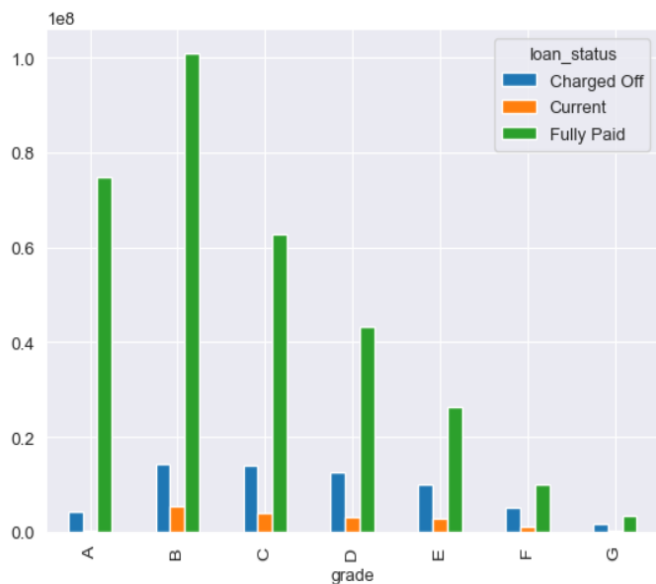
# Year, Loan Status Vs Loan amounts

---



**Note:** Defaulted amounts are increasing every year along with the fully paid loans. Percentage trend decreased till 2009 then it stagnated in 2010 and again increased in 2011. This signifies that As individuals take on more debt, it becomes harder to manage their finances even after increased income. This can lead to missed payments and ultimately, defaults.

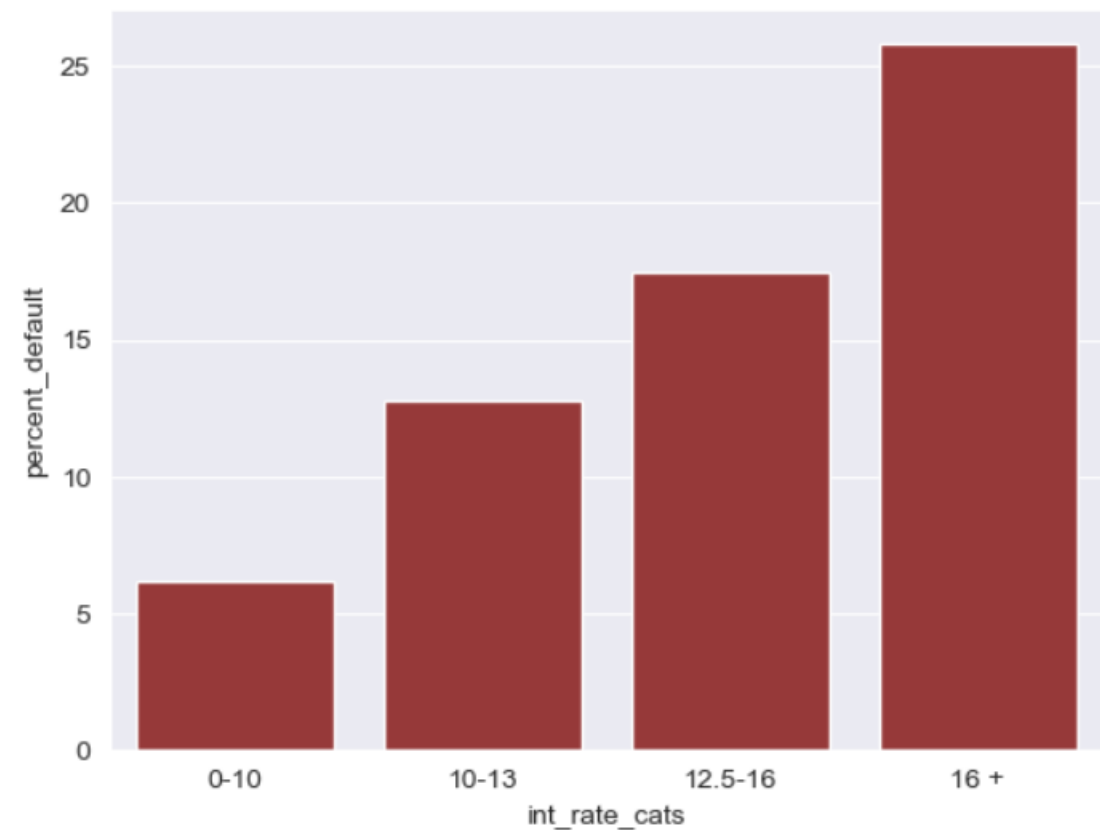
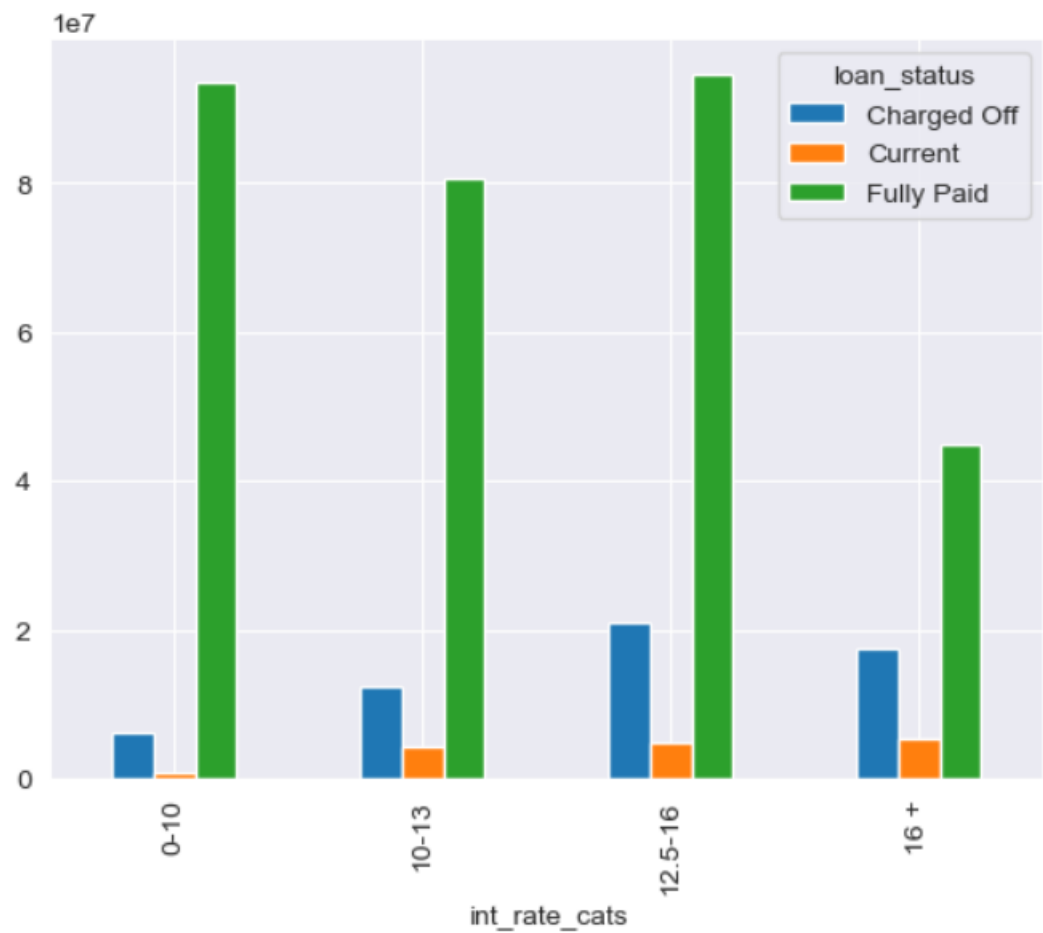
# Grade, Loan Status Vs Loan amount



**Note:** Grade B, C and D have more charged off amounts but Grades E,F and G are more likely to default having default possibility more than 25%. Especially F5, G2,G3 and G5 are more likely to default.



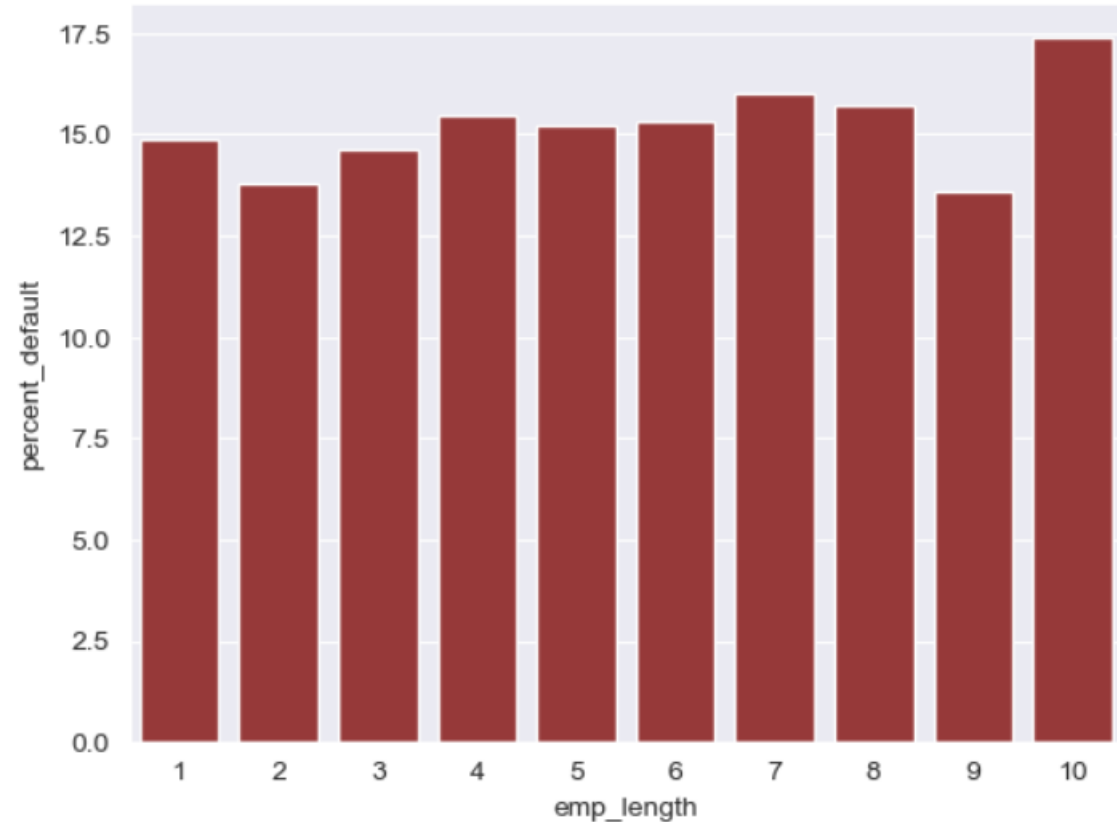
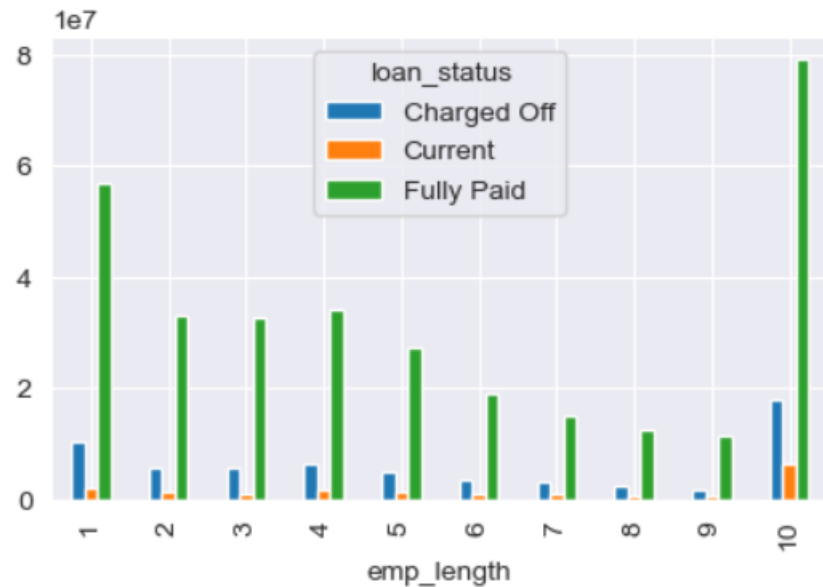
# Interest Rates, Loan Status Vs Loan amounts



**Note:** Higher the interest rates more is the possibility of the default.

# Employment Length, Loan status and Loan amounts

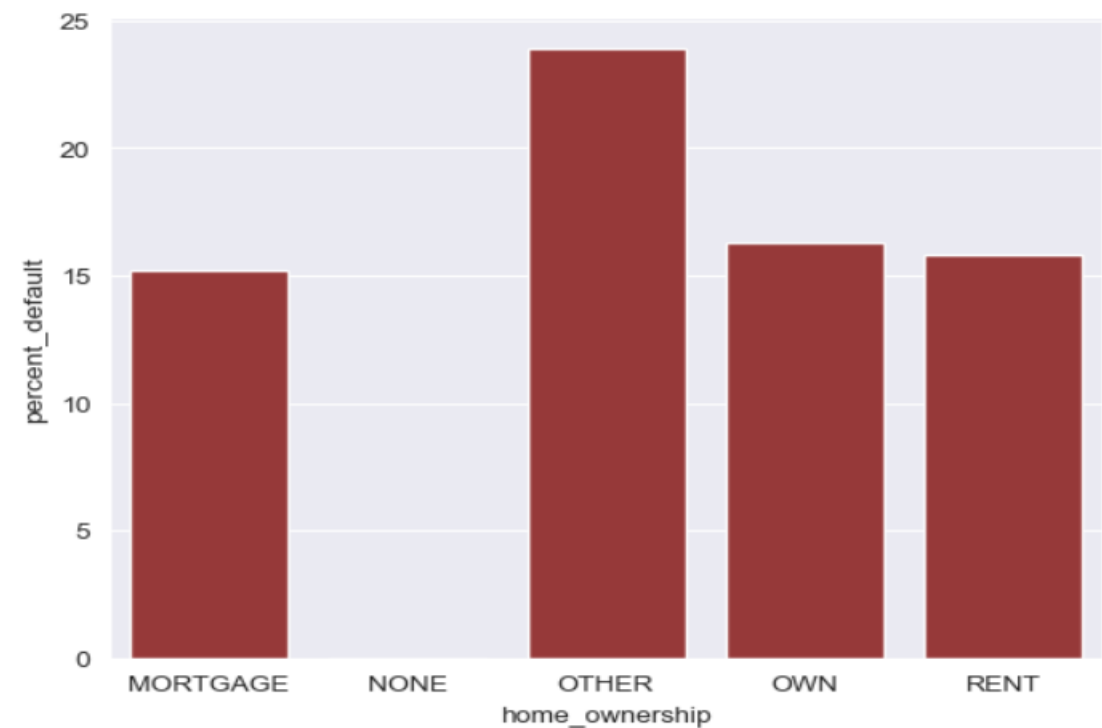
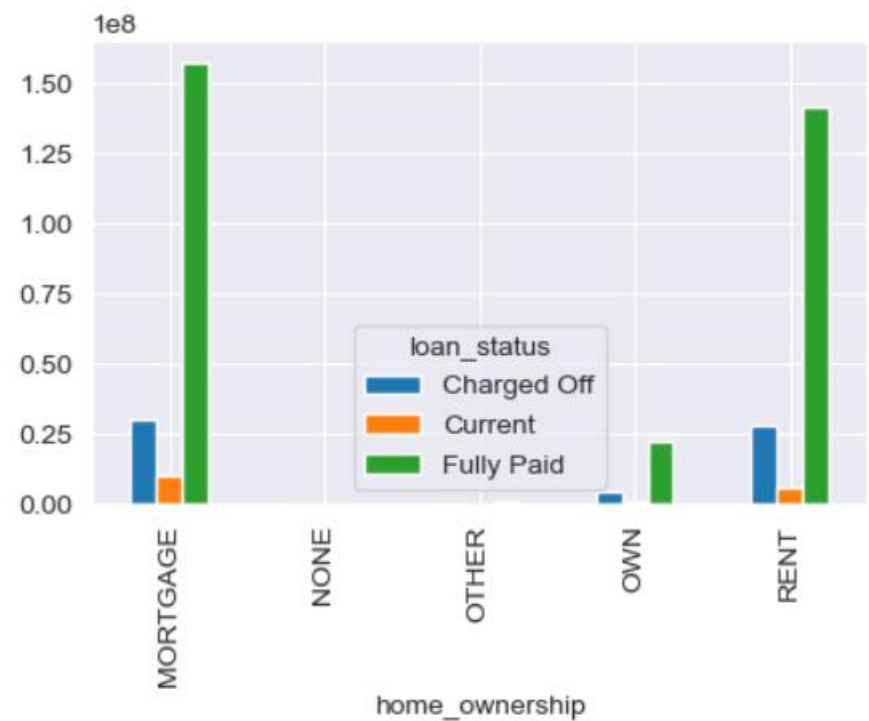
---



**Note:** Overall for all the experiences on an average loan defaults are between 12.5 to 15 percent but for employees having experience more than 10 years are having higher percentage of defaulting.

# Home ownership, Loan Status Vs Loan amount

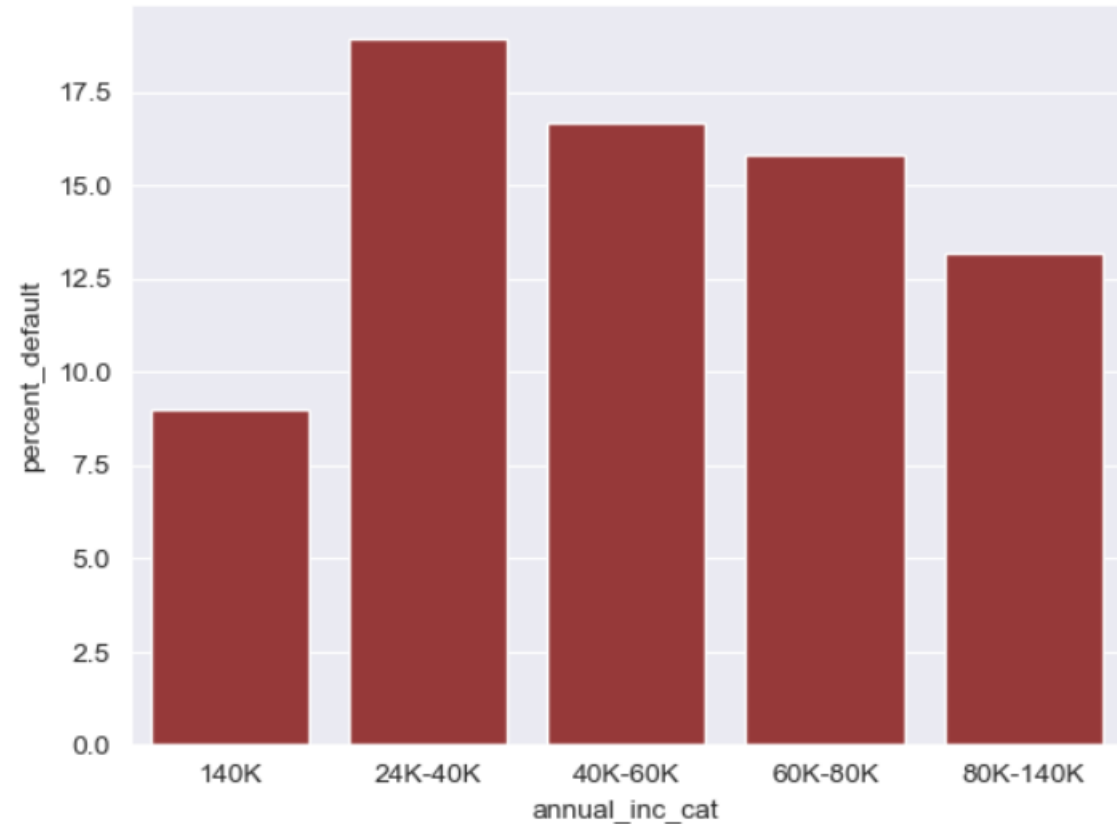
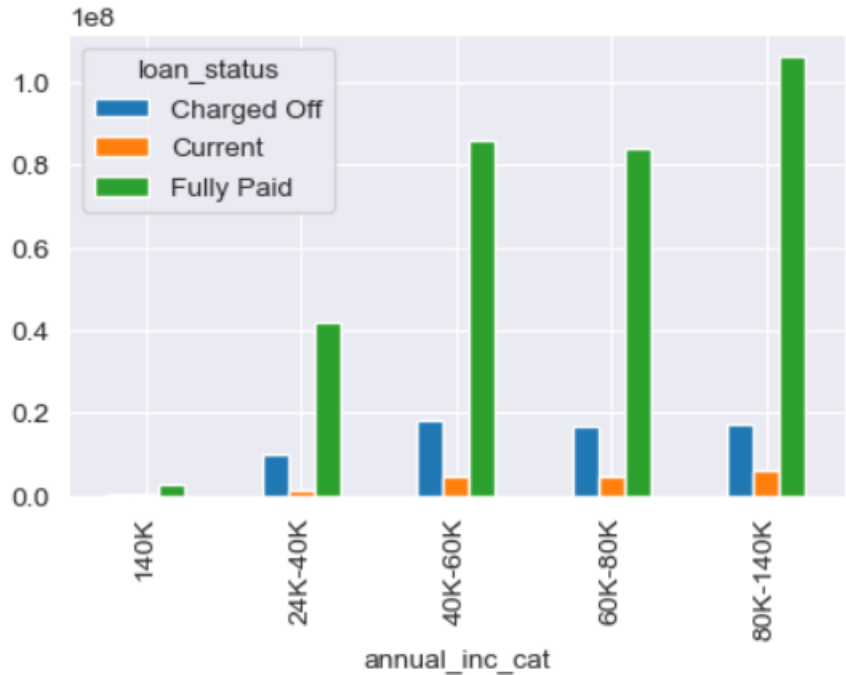
---



**Note:** People with Mortgaged home and Rented home are more likely to take loans and considering the percentages they contribute to significant defaults

# Annual income, Loan Status Vs Loan Amount

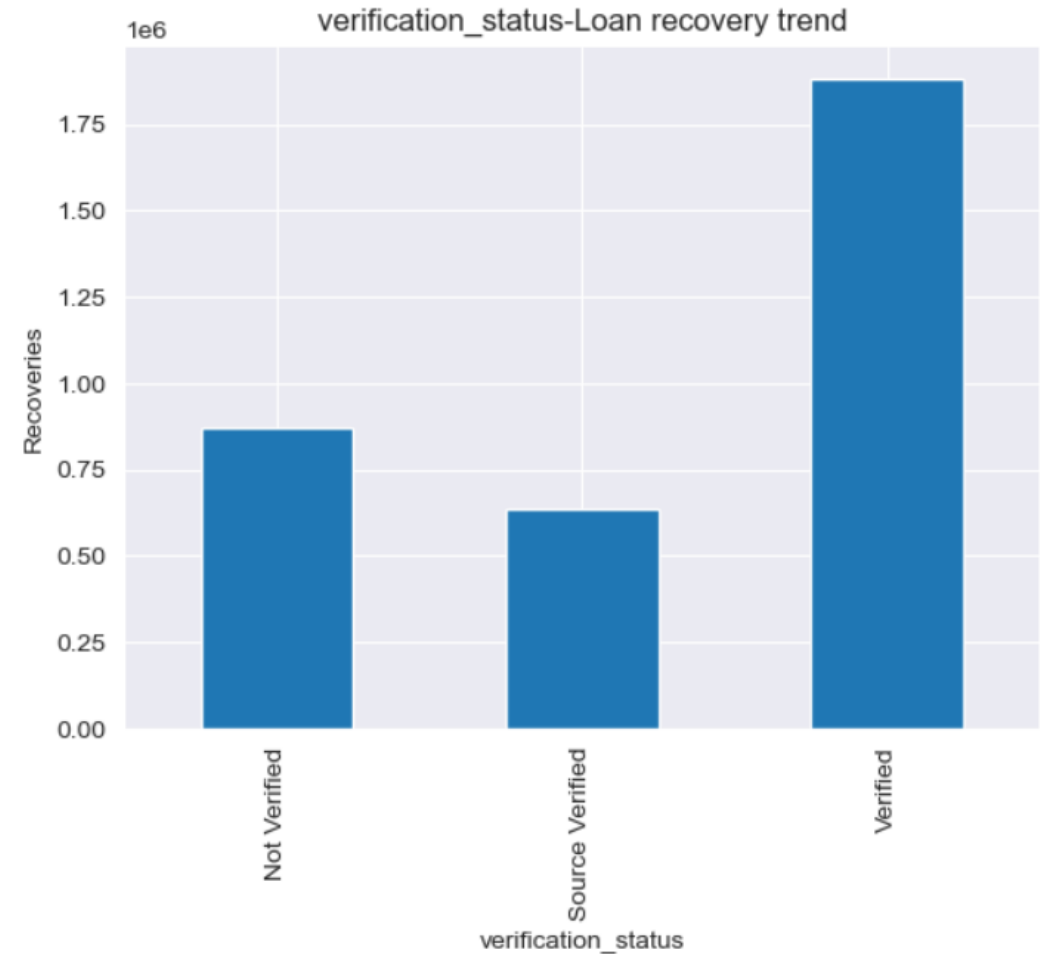
---



**Note:** Applicants with lower income are more likely to default. 24K to 40K income have more percentage of default

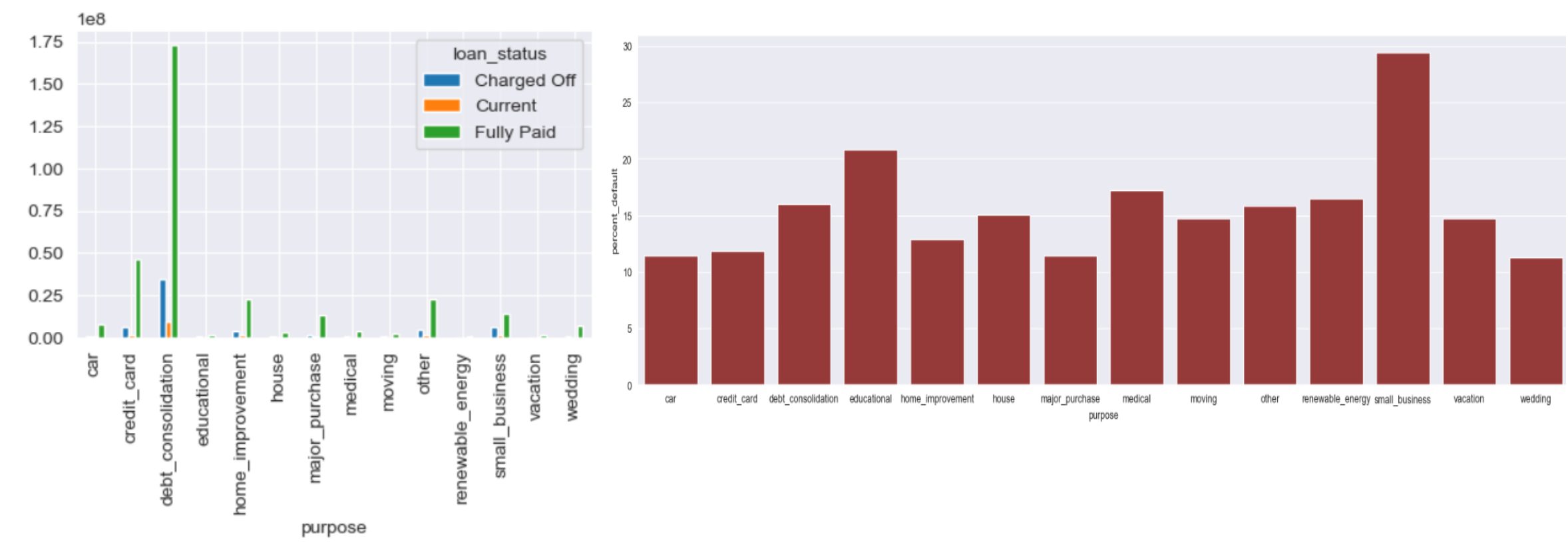
# Verification status, Loan Status Vs Loan Amount, Recoveries

---



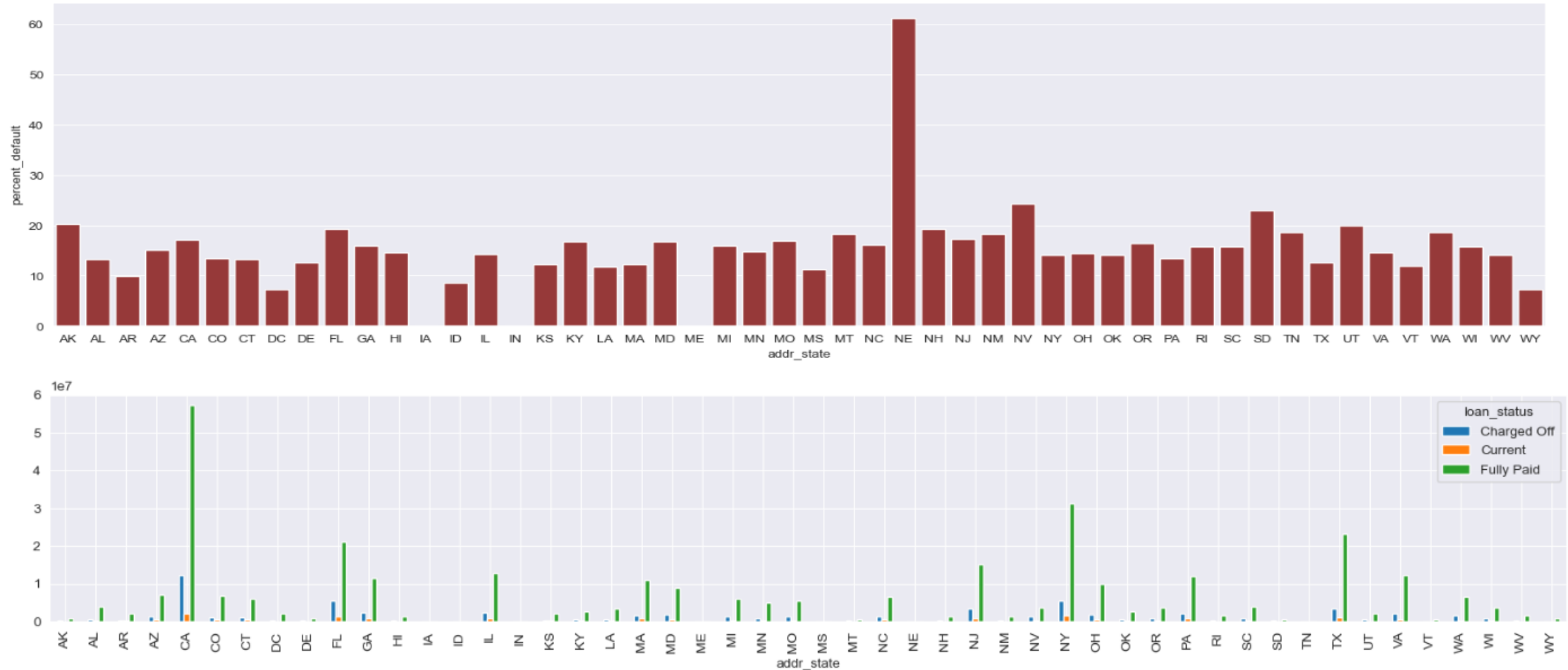
**Note:** Verified applicants are more likely to default, Even though recovery is found it is very insignificant considering the processed loan amount

# Purpose, Loan Status Vs Loan Amount



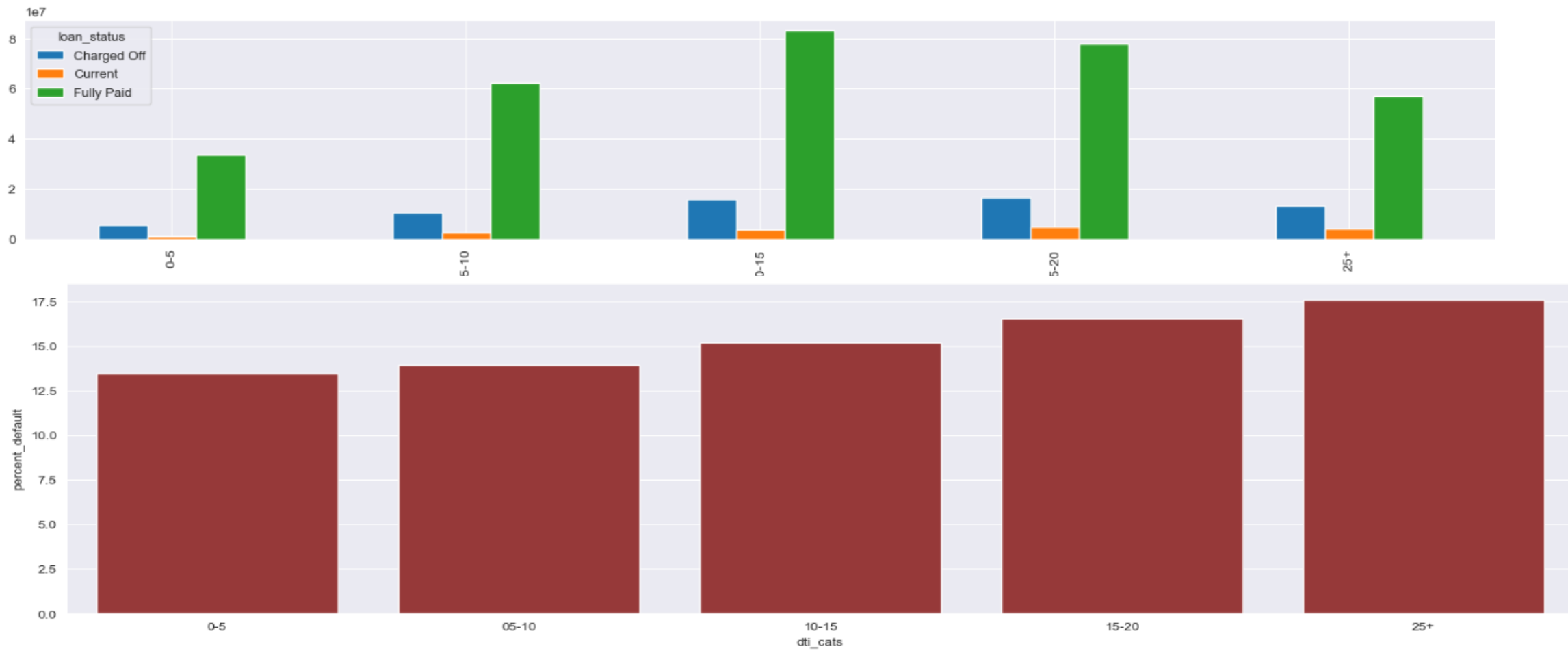
**Note:** Applicants taking loan for debt consolidation are highest contributor of the defaulters. Other than them, loans taken for small Businesses, Renewable energy, Other, medical and Educational purpose have higher default percentages.

# Address state, Loan Status Vs



**Note:** FL is one of the major contributor to the loan amounts, Applicants from FL are having more defaulting tendencies.

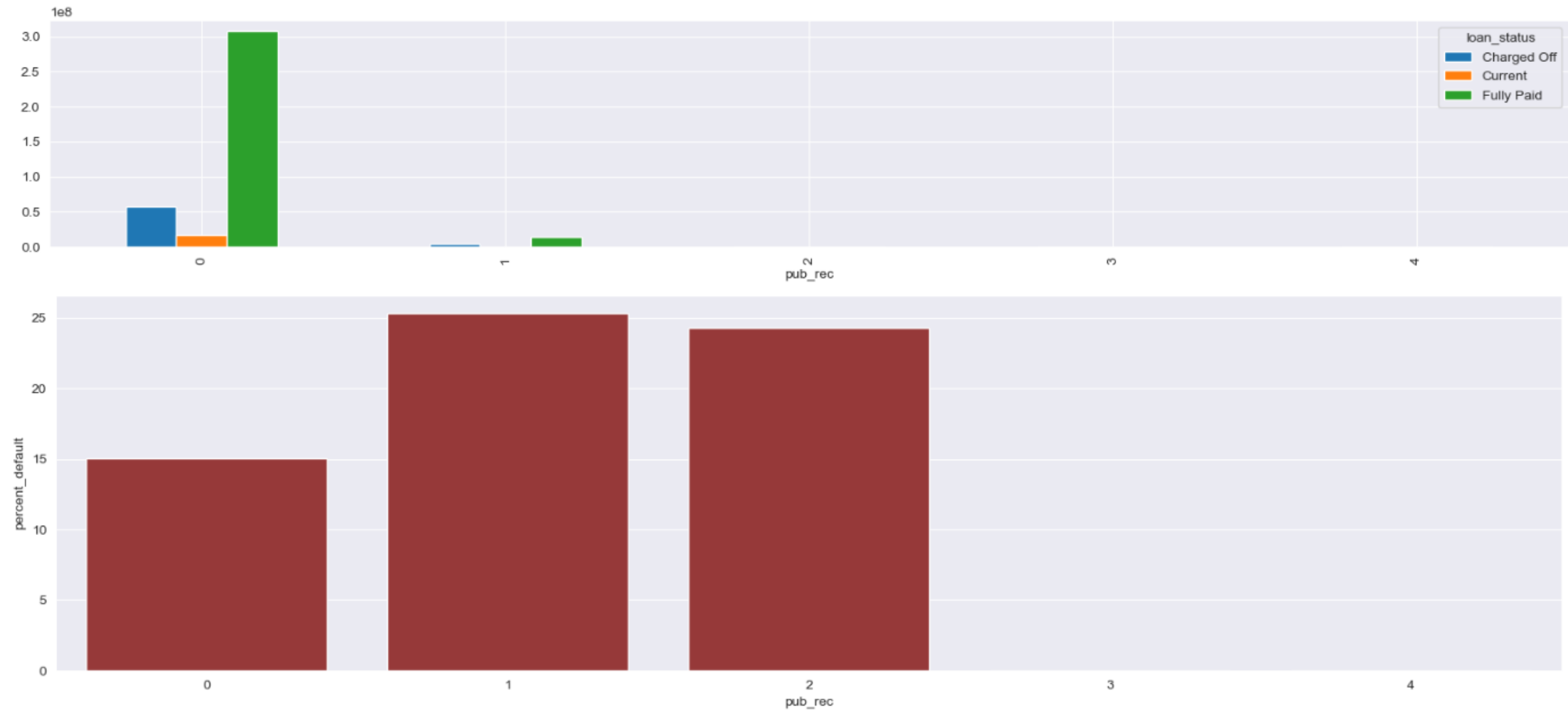
# DTI, Loan status Vs Loan amounts



**Note:** As the DTI increases, tendency to default increases.



# Derogatory public record, Loan Status Vs Loan Amounts



**Note:** Applicants with 1 or two derogatory records have more tendency to default.

# Interest rate category, Loan status, grade Vs Loan Amount(Sum and Average), Count

int_rate_cats	sum				mean				count			
	loan_amnt				loan_amnt				loan_amnt			
	0-10	10-12.5	12.5-16	16 +	0-10	10-12.5	12.5-16	16 +	0-10	10-12.5	12.5-16	16 +
grade												
A	4321175	0	0	0	8061.893657	NaN	NaN	NaN	536	0	0	0
B	1866300	11280925	1136800	0	9285.074627	11326.229920	13066.666667	NaN	201	996	87	0
C	0	1064500	13012550	32800	NaN	9948.598131	11566.711111	16400.000000	0	107	1125	2
D	0	114375	6385775	6211625	NaN	9531.250000	11302.256637	14479.312354	0	12	565	429
E	0	0	516500	8861025	NaN	NaN	12912.500000	15683.230088	0	0	40	565
F	0	0	2125	2253700	NaN	NaN	2125.000000	17203.816794	0	0	1	131
G	0	0	0	100750	NaN	NaN	NaN	14392.857143	0	0	0	7

# Annual income category, Loan status, grade Vs Loan Amount(Sum and Average), Count

annual_inc_cats	grade	sum					mean					count				
		loan_amnt					loan_amnt					loan_amnt				
		0-20000	20000-40000	40000-60000	60000-80000	80000+	20000-40000	40000-60000	60000-80000	80000+	0-20000	20000-40000	40000-60000	60000-80000	80000+	
A	0	1201275	1593150	831225	695525	6636.878453	8005.778894	9339.606742	10380.970149		0	181	199	89	67	
B	0	3329800	4760200	3066425	3127600	8242.079208	11121.962617	12265.700000	15483.168317		0	404	428	250	202	
C	0	3146125	4700725	3053925	3209075	8572.547684	10931.918605	13453.414097	15281.309524		0	367	430	227	210	
D	0	2200300	3976025	3119275	3416175	8495.366795	11591.909621	14996.514423	17429.464286		0	259	343	208	196	
E	0	1235400	3156650	2605625	3026575	9430.534351	14348.409091	17725.340136	22091.788321		0	131	220	147	137	
F	0	404700	1429325	1631700	1822025	10937.837838	16242.329545	20144.444444	24621.959459		0	37	88	81	74	
G	0	200725	325175	448950	736400	11151.388889	15484.523810	20406.818182	23754.838710		0	18	21	22	31	

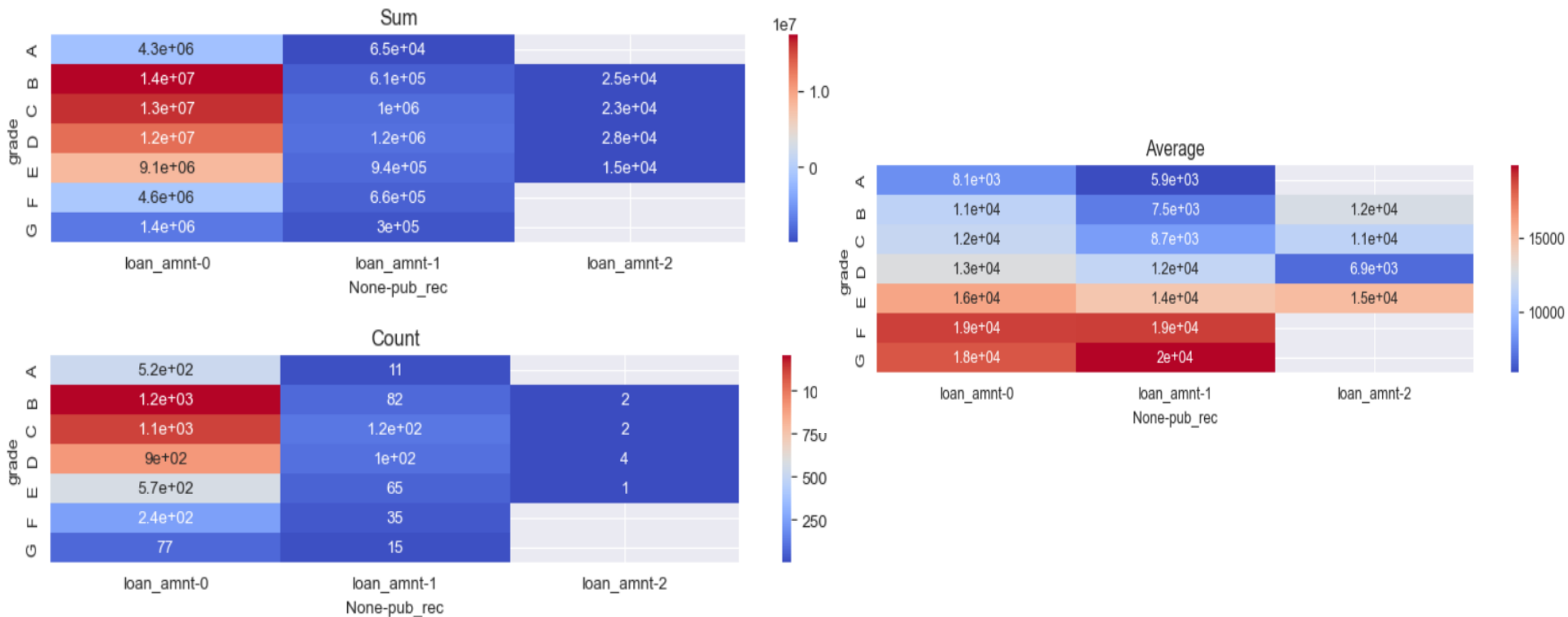
# Loan amount category, Loan status, grade Vs Loan Amount(Sum and Average), Count

loan_amnt_cats	sum					mean					co			
	loan_amnt					loan_amnt					loan_ai			
	0-7000	7000-14000	14000-21000	21000-28000	28000 +	0-7000	7000-14000	14000-21000	21000-28000	28000 +	0-7000	7000-14000	14000-21000	21000-28000
grade														
A	1182100	2092000	746525	240550	60000	4361.992620	10106.280193	16228.804348	24055.000000	30000.000000	271	207	46	10
B	1971200	4598950	4588325	2137100	988450	4390.200445	10242.650334	17120.615672	24285.227273	32948.333333	449	449	268	88
C	1860875	4235725	4475625	2370400	1167225	4368.251174	10231.219807	17147.988506	24187.755102	33349.285714	426	414	261	98
D	1205575	3666750	4294650	2286650	1258150	4532.236842	10213.788301	17178.600000	24587.634409	33109.210526	266	359	250	93
E	541050	1975475	3020475	2408975	2078275	4585.169492	10736.277174	17560.901163	24581.377551	32988.492063	118	184	172	98
F	146925	563950	1497025	1758650	1321200	4897.500000	10640.566038	17612.058824	24425.694444	33030.000000	30	53	85	72
G	54500	215300	499450	504600	437400	4954.545455	11331.578947	17837.500000	24028.571429	33646.153846	11	19	28	21

# Loan amount category, DTI category, Loan status Vs Loan Amount(Sum and Average), Count

annual_inc_cats	sum					mean					count			
	loan_amnt					loan_amnt					loan_amnt			
	0-20000	20000-40000	40000-60000	60000-80000	80000+	20000-40000	40000-60000	60000-80000	80000+	0-20000	20000-40000	40000-60000	60000-80000	80000+
dti_cats														
0-5	0	958100	1656150	948150	1865675	6653.472222	9304.213483	11154.705882	16807.882883	0	144	178	85	111
05-10	0	1831625	2871350	2622350	3223875	8176.897321	10835.283019	13587.305699	17713.598901	0	224	265	193	182
10-15	0	2830900	4844375	3814250	4141400	8657.186544	11589.413876	14285.580524	16972.950820	0	327	418	267	244
15-20	0	2930675	5491875	4108100	3919125	8594.354839	12070.054945	15328.731343	17418.333333	0	341	455	268	225
25+	0	2879900	4666775	2917875	2606225	8780.182927	12345.965608	15040.592784	18225.349650	0	328	378	194	143

# Public Record, Loan status, grade Vs Loan Amount(Sum and Average), Count



# Observations Based on Multivariate analysis

---

- **Interest rate, Grade, Loan Status Vs Loan amount, counts**

- 1) Grade A is taking loan with interest rates 0 to 10 only
- 2) Maximum defaults are in B, C and D
- 3) In B the one taking interest rate 10-12.5 have more defaulters
- 4) In C the one taking interest rate 12.5-16 have more defaulters
- 5) For 16+ interest rates grade D and E have more defaulters

- **Annual income, Grade, Loan Status Vs Loan amount, counts**

- 1) Grade B,C and D have more defaulters for annual income 20K-60K
- 2) Grade E have more defaulters with annual income 40K to 60K

- **Loan amounts categories, Grade, Loan Status Vs Loan amount, counts**

- 1) Grade B,C and D have more defaulters with loan amount 0-21K
- 2) Grade E have more defaulters with 7K-21K

- **Public record, Grade, Loan Status Vs Loan amount, counts**

- 1) Grade F and G have in general high average in the defaulter, but having 0 or 1 public record among them is having high average amount defaulters.

- **Home ownership, Grade, Loan Status Vs Loan amount, counts**

- 1) Mortgaged home ownership is more prone to be defaulted

- **Purpose, Grade, Loan Status Vs Loan amount, counts**

- 1) Grade B,C,D and E have maximum defaulter with the purpose of debt\_consolidation

- **State, Grade, Loan Status Vs Loan amount, counts**

- 1) Grade A have max defaulters in states: CA, NJ
- 2) NY, FL, NJ and CA have maximum defaulters for A to E categories

- **DTI, Grade, Loan Status Vs Loan amount, counts**

- 1) For grade B,C,D,E maximum defaulters are in 10-20 dti
- 2) For grade A maximum defaulters are in 5-15 dti

- **Sub-Grades, Loan Status Vs Loan amount, counts**

- 1) A4 and A5 have more defaulters
- 2) B3,B4 and B5 have more defaulters
- 3) C1, C2, C3 have more defaulters
- 4) F1, E1,E2 have more defaulters
- 5) D2 has more tendency to default

# Conclusions and suggestions

---

1. **Grades B,C and D:** Grades B, C, and D attract a significant portion of loan applicants, presenting a higher risk of default. Thorough risk assessment is crucial for these grades, particularly for subgrades B3-B5, C1-C3, as well as D2.
2. **Annual Income 20K to 60K and DTI:** Applicants with annual incomes between 20,000 and 60,000 are more likely to apply for loans in grades B, C, and D, placing them within a higher-risk portfolio. These applicants often exhibit higher debt-to-income (DTI) ratios, typically ranging from 10 to 25+ or more. Careful monitoring of these factors is essential.
3. **Reject loans with high amount for grades F and G:** Grades F and G do not have many loan applicants but the loan applications are of high amounts. Very high amounts for these grades need to be rejected as they are more likely to be defaulted.
4. **Reduction in interest rates:** Loan applications with higher interest rates are more likely to be defaulted. Lending club should consider reduction in the interest rates for reducing number of defaults.



# Conclusions and Suggestions

---

5. **Purpose and derogatory record:** Loan applicant taking loan for debt consolidation should be properly verified. Also, applicant with zero and one derogatory public record tend to default as well. Risk should be assessed for them as well.

6. **Improvement in the Recovery and Evaluation of covering Asset:** From the Analysed details the recovery of the amount is not significant. Company needs to review the evaluation of the assets used as guarantee or cover in case of default.

7. **Increase in the defaulting percentage due to significant increase in income in 2011:** In 2011 the income has increased significantly which lead to higher loan applications but it also increased the defaulted loans.

As individuals take on more debt, it becomes harder to manage their finances. This can lead to missed payments and ultimately, defaults.

8. **People with Rented house or Mortgage house:** People with Rented house or Mortgaged house are more likely to defaulted as more operational cost leads to the missing payout for loans.

---



Thank  
You!