

# Lending Club Case Study

## Exploratory Data analysis

By Sagarika Tamavada & Dharmendra Singh Chouhan

# Executive Summary

## Data Background

- The data contains information about past loan applicants and whether they 'defaulted' or not.
- The dataset contains information starting from 2007 to 2011

## Objective

- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

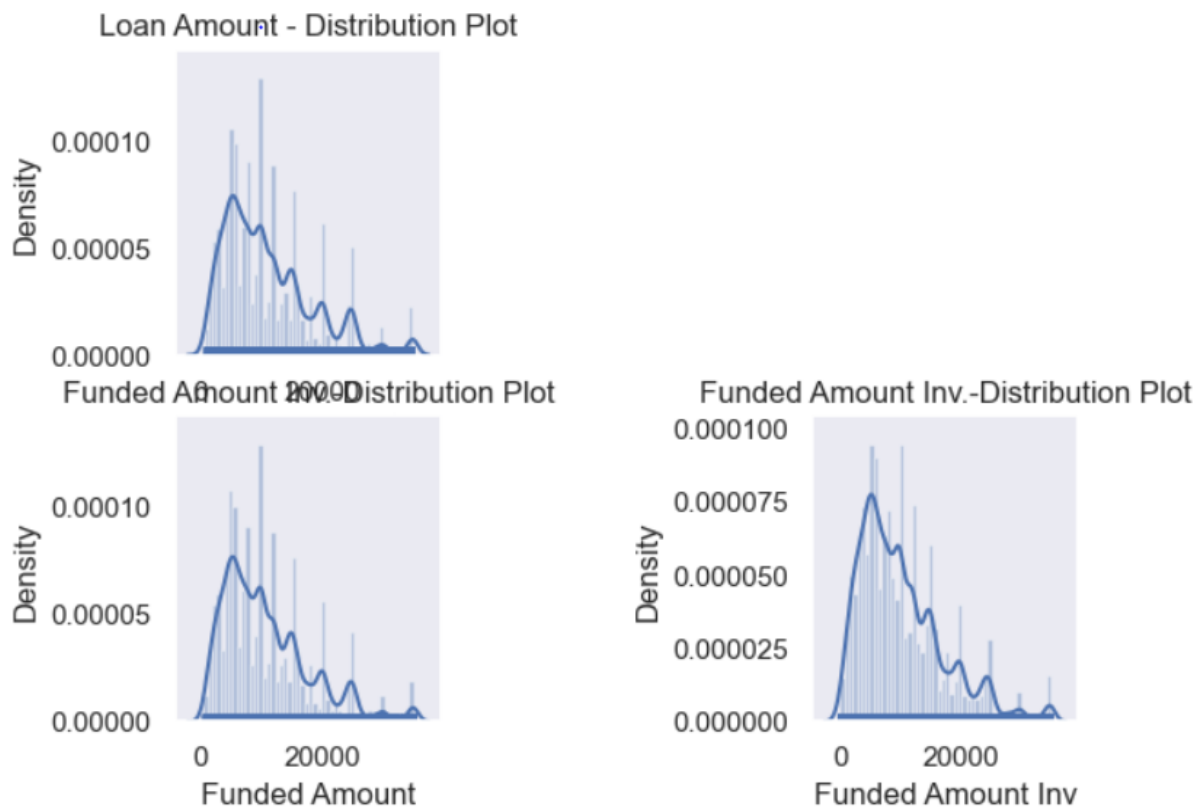
## Analysis

- Data Cleaning – Removing Columns with 100% null values, presence of extreme values, reclassifying existing column types
- Univariate, Bi-variate analysis

## Key Observations

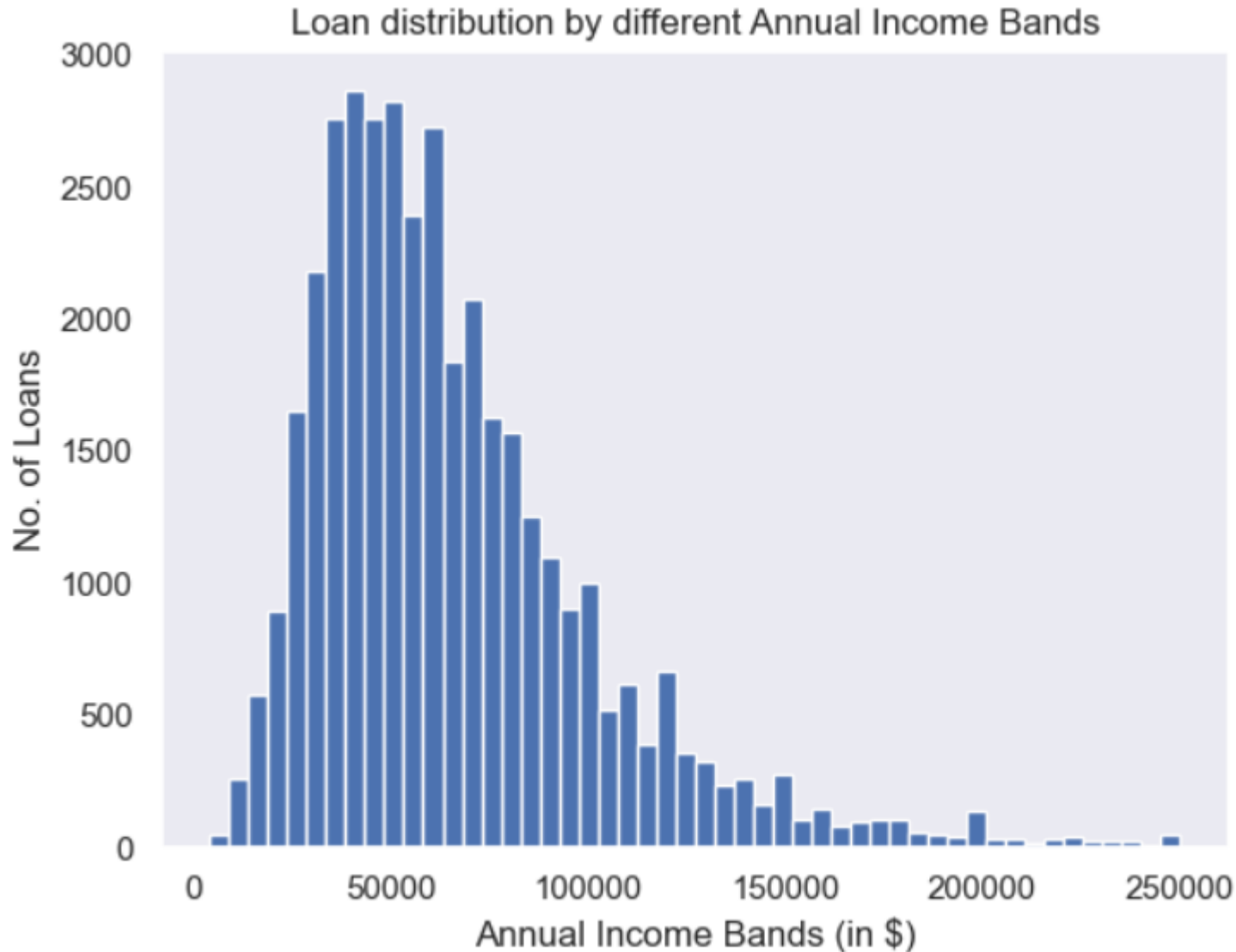
- Number of loans issued increased steadily by every year with a slight decrease in 2008.
- Of settled loans, 83% were Fully Paid and 14% were Charged Off.
- Income, Grade are key driving factors to identify charge-off accounts
- Majority of loans were from A, B, and C grade.
- There is an inverse relationship between interest rate and loan grade - lower grades(E,F,G) have higher interest rate.

# Univariate Analysis

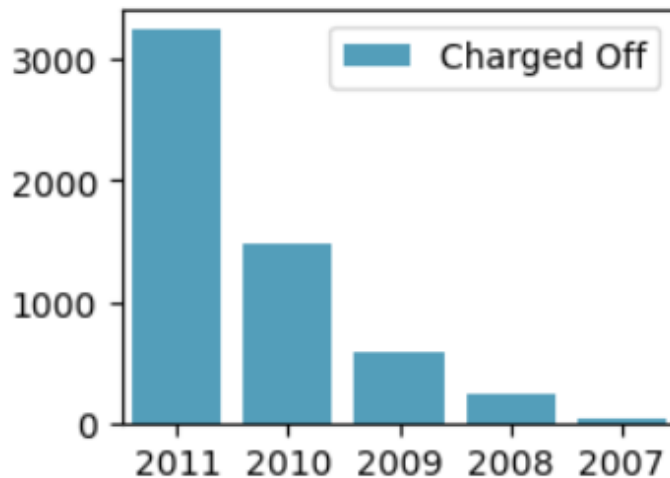


Distribution of loan amounts with Funded amount and Investor funded amount is same

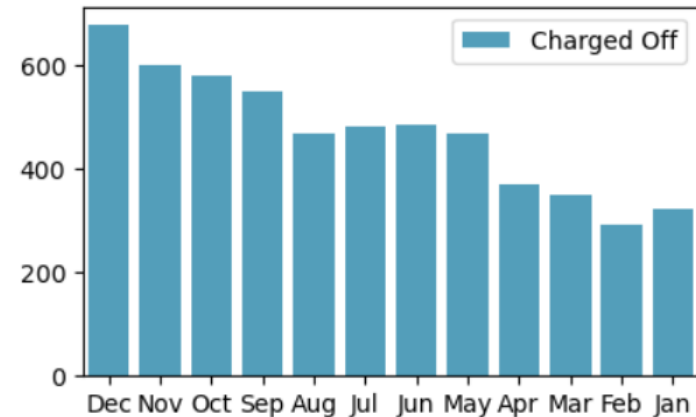
- ~95% of the loans are provided to the customers with annual income less than \$150k
- Most of the Annual Income are in the range 25000-75000



Loan Status vs issue year



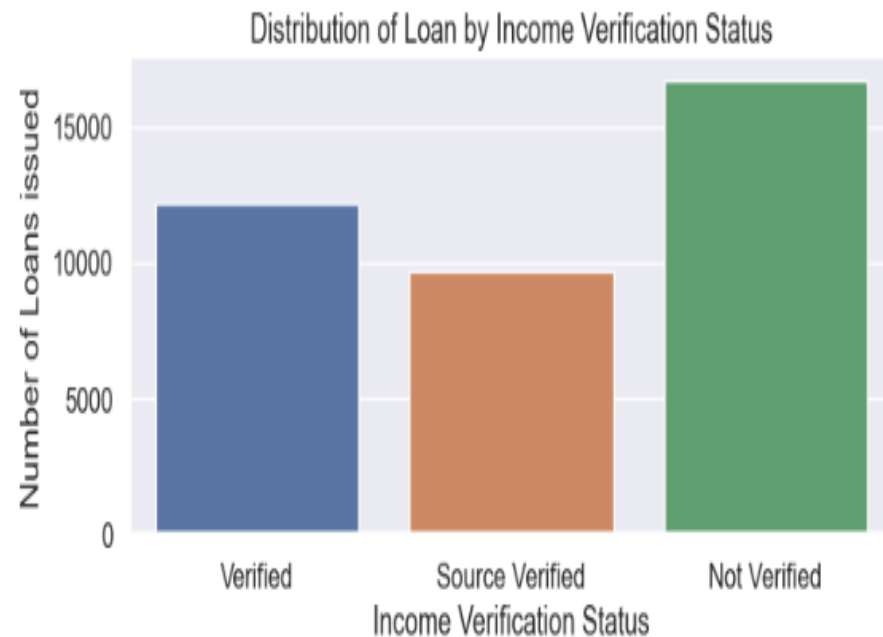
Loan Status vs issue month



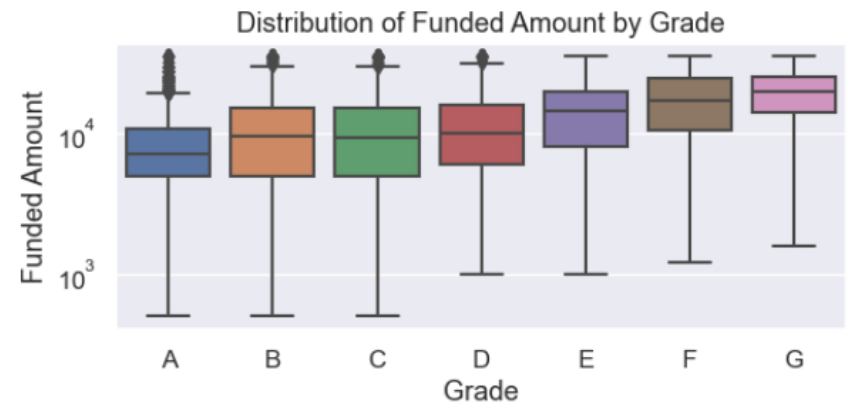
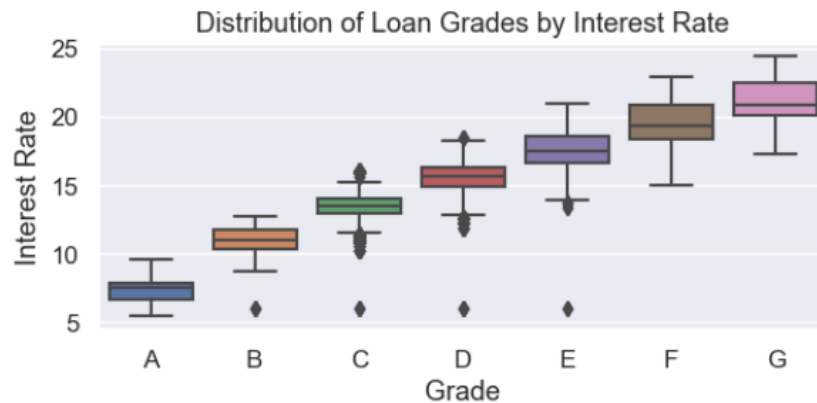
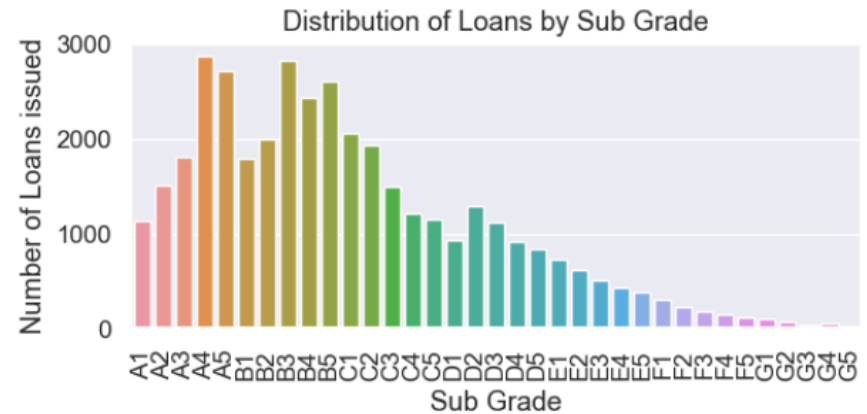
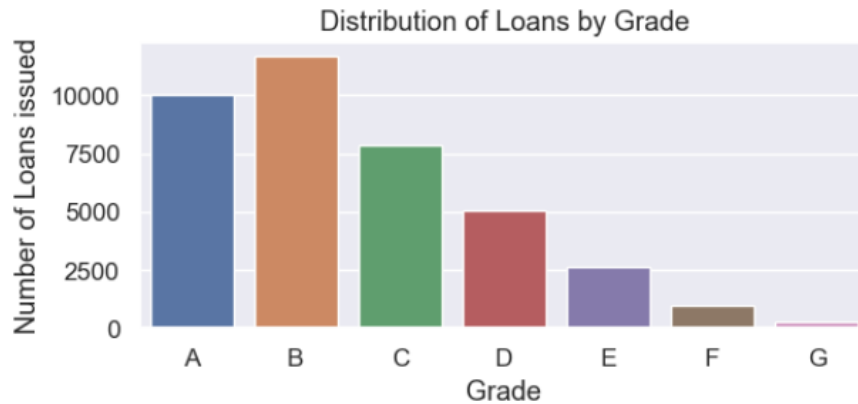
## Key Observations:

- It shows that greater number of loans that were charged off in month of Nov, Dec, the reasons could be Festivals (Thank's giving Day, Christmas and New Year).
- In 2011 Year, there was huge spike in charged off.
- Number of loans issued increased steadily with slight decrease in 2008 year.





- **Observations: (Unordered Categorical variables)**
  1. Number of Loans issued increases by the end of year
  2. Greater number of loans which have not verified the annual income.



- Observations: ( Ordered Categorical Variables)
- From above plots, it shows that greater number of loans were from B,A and C grade's and least from G grade.
- From Sub grades A4, B3 have greater number of loans.
- From 3rd plot, it shows that A,B,C grade loans have less interest rate and E,F,G have high interest rate. From 1st, 2nd plots there are greater number of loans from A,B,C grade(granularity check from sub-grades). It might be the reason that the loan Applicant's from A,B,C grades have better credit score and lower risk.
- From 4th plot, it shows that there are high funded amount in A,B,C and D grades as the Applicant's from these grades have better credit score and lower risk.



# Summary of Univariate Analysis

1. Number of loans issued increased steadily by every year with a slight decrease in 2008.
2. Of settled loans, 83% were Fully Paid and 14% were Charged Off.
3. Borrowers with own house and want to consolidate debt are not at much risk, but borrower with rent , mortgage and want to consolidate debt are high risk applicants.
4. Majority of loans were from A, B, and C grade.
5. There is an inverse relationship between interest rate and loan grade - lower grades have higher interest rate.

# Bivariate Analysis

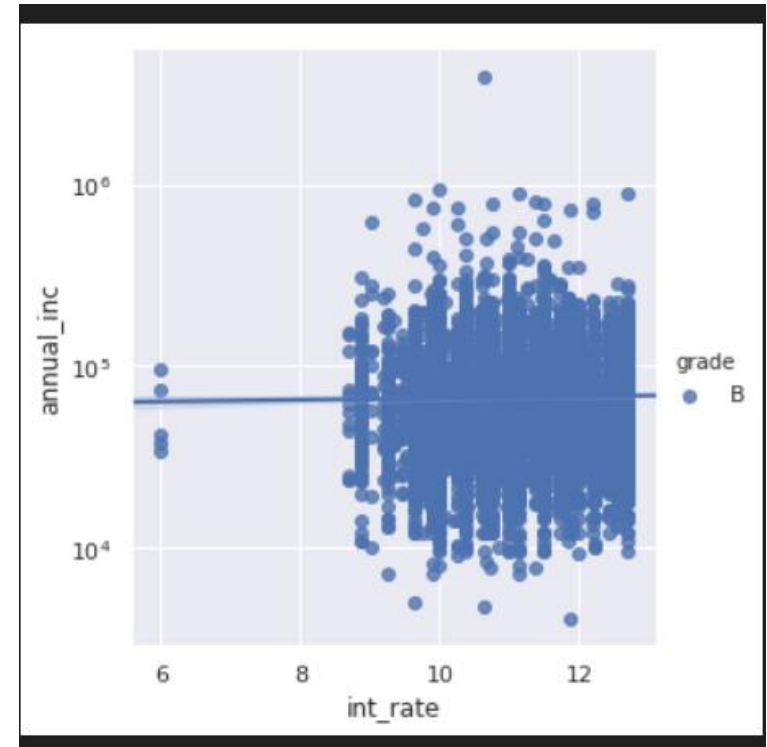
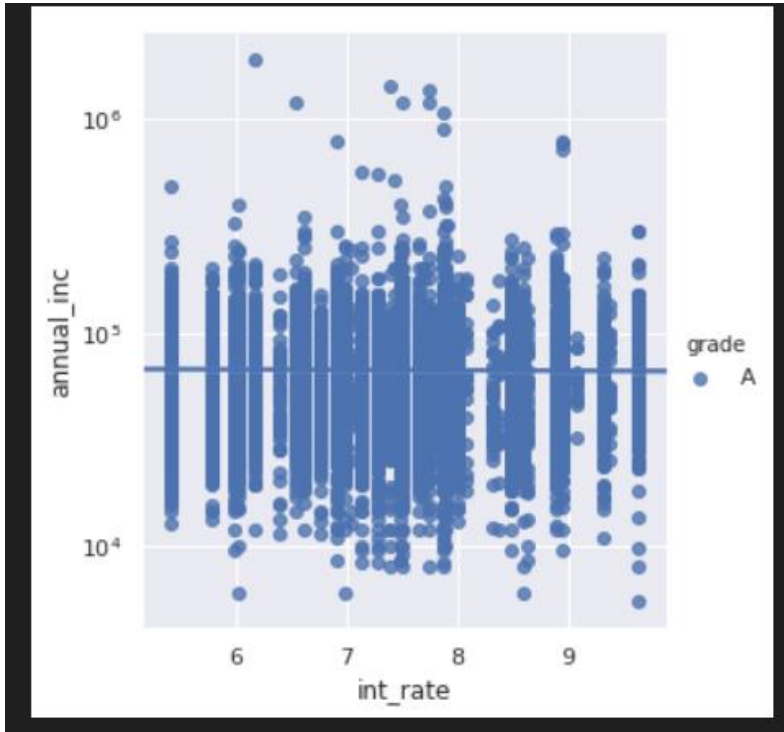
## Correlation Matrix

	id	loan_amnt	funded_amnt	funded_amnt_inv	term_months	int_rate	installment	emp_length	annual_inc	dti	...
id	1.000000	0.120614	0.131283	0.231603	0.176405	0.053936	0.076088	0.089470	0.005572	0.091785	...
loan_amnt	0.120614	1.000000	0.981790	0.937922	0.346650	0.301265	0.932260	0.156259	0.268999	0.062436	...
funded_amnt	0.131283	0.981790	1.000000	0.956172	0.324893	0.304930	0.958035	0.155893	0.264798	0.062194	...
funded_amnt_inv	0.231603	0.937922	0.956172	1.000000	0.343922	0.297473	0.905464	0.164291	0.251981	0.070663	...
term_months	0.176405	0.346650	0.324893	0.343922	1.000000	0.440206	0.090456	0.102642	0.043866	0.076182	...
int_rate	0.053936	0.301265	0.304930	0.297473	0.440206	1.000000	0.277203	0.011440	0.048899	0.110913	...
installment	0.076088	0.932260	0.958035	0.905464	0.090456	0.277203	1.000000	0.131151	0.267842	0.052038	...
emp_length	0.089470	0.156259	0.155893	0.164291	0.102642	0.011440	0.131151	1.000000	0.121229	0.048964	...
annual_inc	0.005572	0.268999	0.264798	0.251981	0.043866	0.048899	0.267842	0.121229	1.000000	-0.121530	...
dti	0.091785	0.062436	0.062194	0.070663	0.076182	0.110913	0.052038	0.048964	-0.121530	1.000000	...
delinq_2yrs	-0.008417	-0.031951	-0.031866	-0.038171	0.007270	0.158479	-0.019755	0.016091	0.022229	-0.033347	...
inq_last_6mths	-0.041021	0.012940	0.012857	-0.002800	0.047708	0.133362	0.011014	0.010937	0.035465	0.002178	...
open_acc	0.016256	0.177200	0.175682	0.162738	0.046162	0.006022	0.172893	0.099879	0.156927	0.287849	...
pub_rec	-0.017683	-0.049997	-0.050576	-0.051470	0.010266	0.098635	-0.045706	0.050631	-0.017864	-0.004742	...
revol_bal	0.001357	0.314022	0.306501	0.286265	0.066411	0.094179	0.309501	0.152821	0.277374	0.228090	...
total_acc	0.039902	0.256179	0.250551	0.242715	0.096168	-0.046539	0.229860	0.203638	0.234534	0.229119	...

- 1) There are no highly negative correlation between columns.
- 2) There are some highly correlated columns at top and middle left block.
- 3) There is 0.45 correlation between term\_months and interest rate(int\_rate).
- 4) There is 0.47 correlation between revol\_util and interest rate(int\_rate).
- 5) There is no much correlation between Annual Income, DTI with other columns.

# Categorical Variables:

Annual Income Vs Interest Rate distribution by each grade



## Observations:

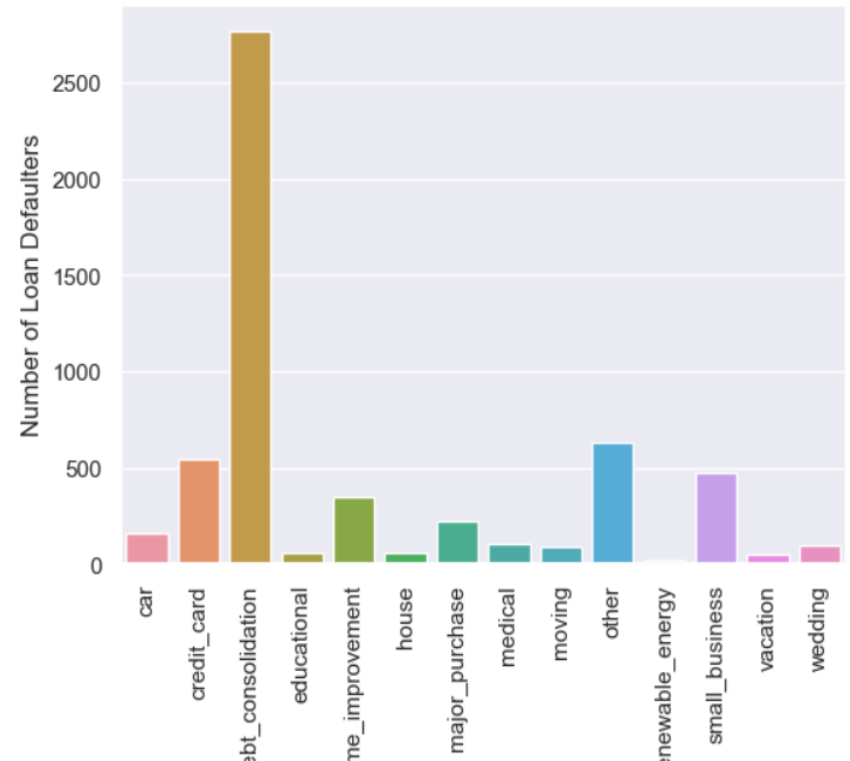
- 1) Grade A has borrowers with interest rate between 5-10%.
- 2) Grade B has borrowers with interest rate between 8-13%, but few are there with near 6%.

# Loan Default Analysis

Distribution of Loan Defaulters by Home Ownerships

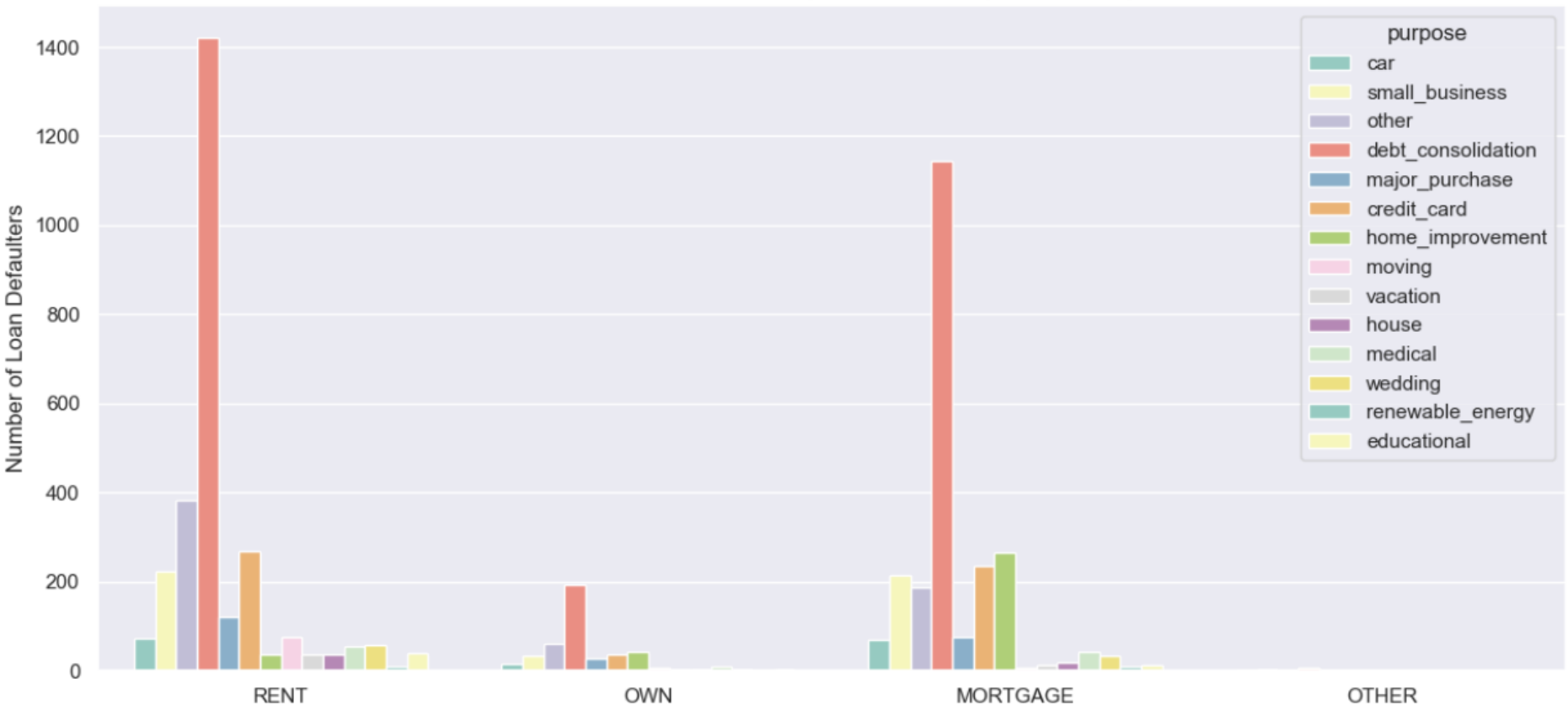


Distribution of Loan Defaulters by Loan Purpose

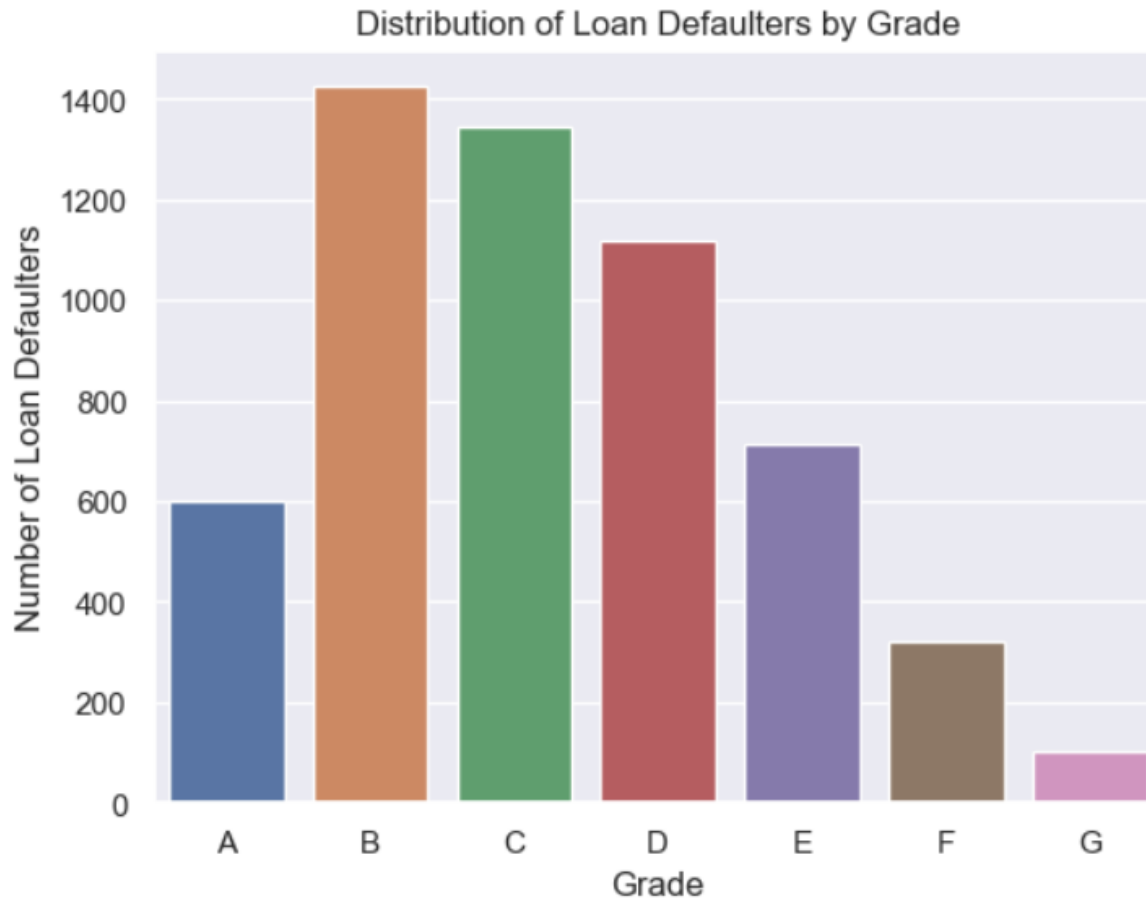


- 1) It shows there are more defaulters in RENT and MORTGAGE. let's check it in granular level.
- 2) There are more defaulters from 'debt\_consolidation','other', 'credit\_card' and 'small\_business'.

Distribution of Loan Defaulters by Home Ownerships & Loan Purpose



- From RENT category, there are more defaulters from 'debt\_consolidation', 'other', 'credit\_card' and 'small\_business'.
- From MORTGAGE category, there are more defaulters from 'debt\_consolidation', 'home\_improvement', 'credit\_card' and 'small\_business'.
- Overall, one should be careful with 'debt\_consolidation', 'credit\_card' and 'small\_business' loans when the borrowers don't have own house.



### Observations:

1. It shows there are more defaulters in B,C and D grades.
2. Grades F,G(more interest rate grades) are having less defaulters which is a good indicator.

# Summary of Bivariate Analysis

- **Bivariate Analysis on Continuous Variables:**

#Correlation Plot:

- a) There were no highly negative correlation between numerical columns.
- b) 0.45 correlation between term\_months and interest rate(int\_rate).
- c) 0.47 correlation between revol\_util and interest rate(int\_rate).
- d) No much correlation between Annual Income, DTI with other numerical columns.

- **Loan Defaulters Analysis:**

1. Distribution of Loan Defaulters by Loan Purpose: a) There are more defaulters from 'debt\_consolidation', 'other', 'credit\_card' and 'small\_business'.
2. Distribution of Loan Defaulters by Home Ownerships: a) By charts, it shows there are more defaulters in RENT and MORTGAGE.
3. Distribution of Loan Defaulters by Home Ownerships & Loan Purpose: a) There are more defaulters with 'debt\_consolidation', 'credit\_card' and 'small\_business' purpose from Rent, Mortgage category and hence, should be careful when borrowers don't have own house.
4. Distribution of Loan Defaulters by Grade: a) By charts, it shows there are more defaulters in B, C and D grades. b) Grades F, G (more interest rate grades) are having less defaulters which is a good indicator.

# Conclusion

- 1) Number of loans issued increased steadily by every year with a slight decrease in 2008.
- 2) Of settled loans, 83% were Fully Paid and 14% were Charged Off.
- 3) Borrowers with own house and the purpose of loan with consolidate debt, 'credit\_card' and 'small\_business' are not at much risk, but borrower with rent,mortgage are high risk applicants.
- 4) Majority of loans were from A, B, and C grade.
- 5) There is an inverse relationship between interest rate and loan grade - lower grades(E,F,G) have higher interest rate.
- 6) Overall, there are more defaulters from 'debt\_consolidation', 'others', 'credit\_card' and 'small\_business' purpose loans from all grades.