

Lending Club Case Study

- Duc Luong
- Dnyaneshwari Chidrawar

Business Objective

- The company wants to understand the driving factors behind loan default
- The variables which are strong indicators of default.
- The company can utilise this knowledge for its portfolio and risk assessment.



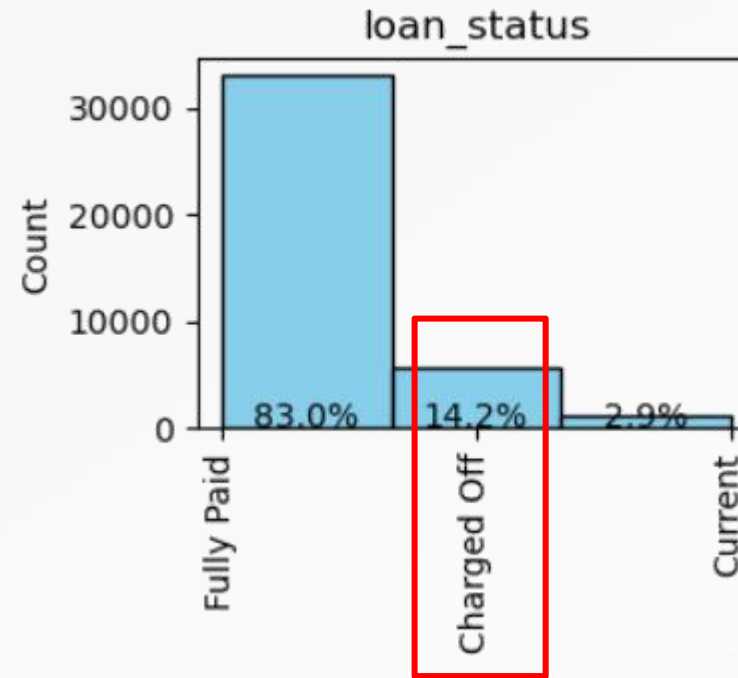
Data Cleaning

- Original data set contains 39717 rows and 111 columns
- 1. Dropped 54 **Empty Columns**
- 2. Dropped 9 columns with **Single Valued** throughout the rows
- 3. Dropped 3 columns which had **Unique value** for all rows
- 4. Dropped 2 fields which had more than 90% **missing values**
- 5. Dropped 4 fields which had most numbers at 0 and remaining were **outliers**
- 6. Dropped 14 **unnecessary fields** which wont help in analysis (e.g. "title", "zip_code")
- After cleaning data, our final data set contains only **25 columns/fields**

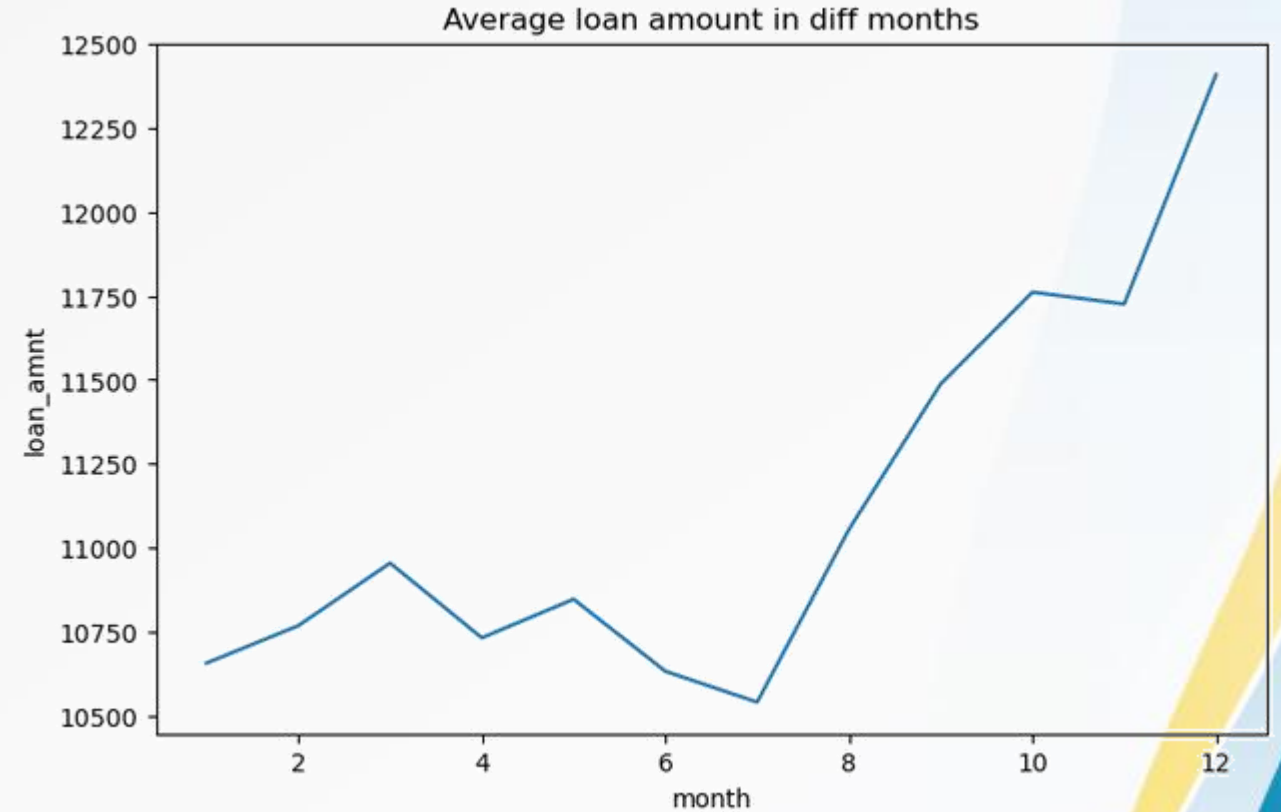
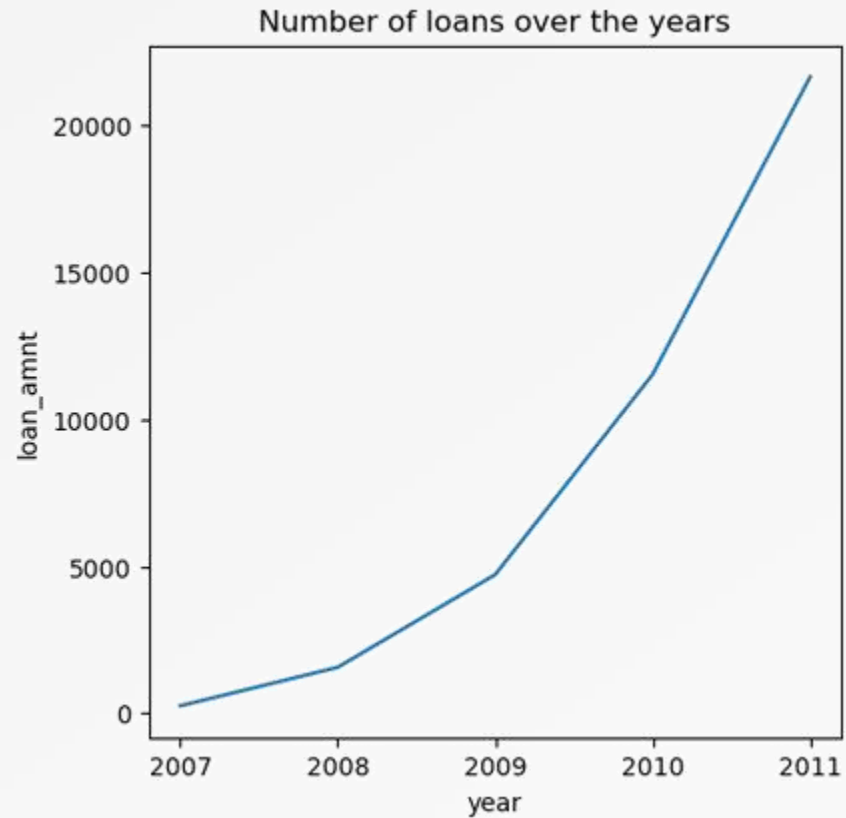


Loan Status

- **14.2%** of loans get **defaulted**.

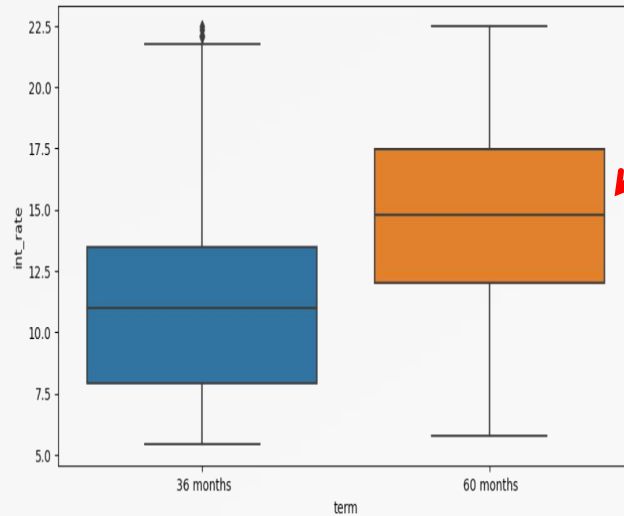
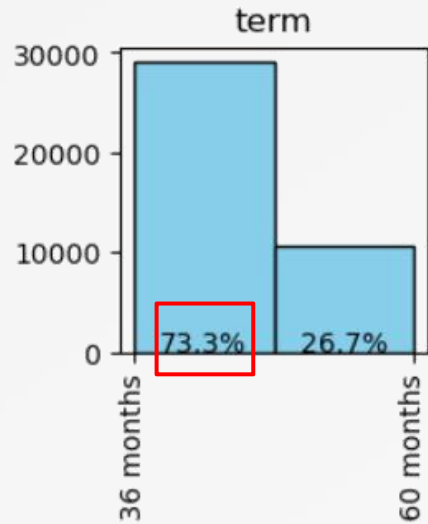


Loan vs Time Period



- Number of loan applicants are increasing over the period of time.
- Total borrowers in 2011 is highest i.e. 21K
- On an average loan amount given to the borrowers is high in 2nd half of the year

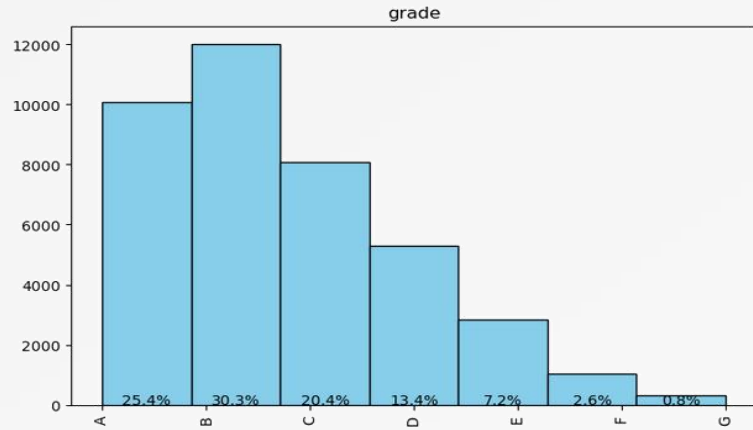
Term, Interest rate, Default Status



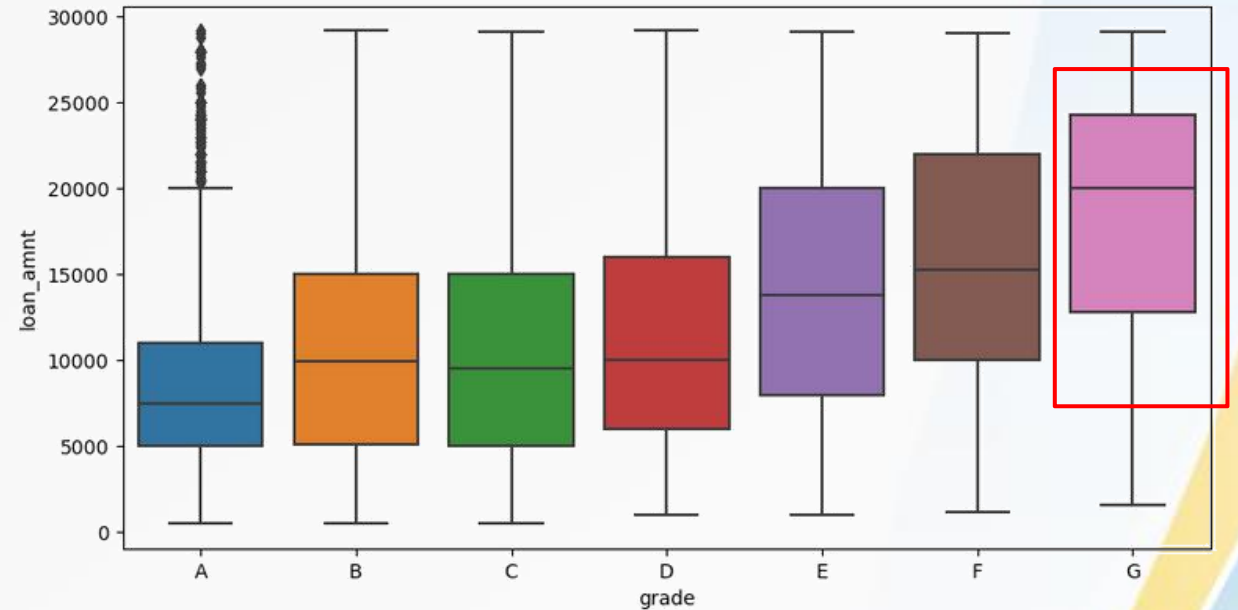
- 73% of loans have 36 months tenure where as only 26% loans have 60 months tenure.
- Loans with **60 month tenure** have higher interest rate than loans with 36 months
- Loans with 60 months tenure have **double chance (22.6%)** of getting defaulted than that of 36 months (11.1%)



Grade, Interest rate, Default Status



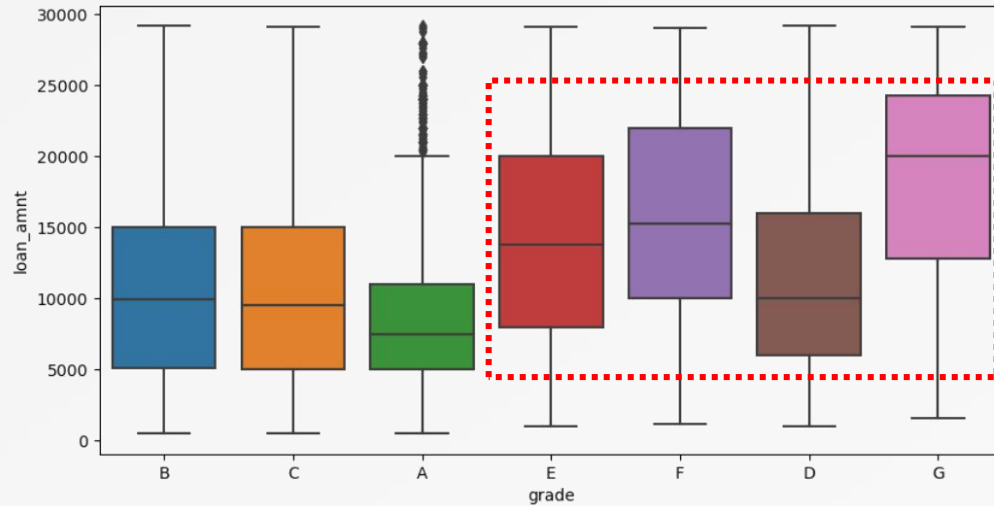
===== Defaulted Loan Status (in %) in field GRADE =====



- 75% loans are of grade A(25%), B(30%), C(20%)
- Interest rates and default rates are low for A, B, C grades.
- Interest rate is gradually increasing for the loans with grades A, B, C, D, E, F, G
- There is a 32% chance of G grade loans getting defaulted.
- **Higher the grade, higher is the interest rate, higher chance of getting defaulted**

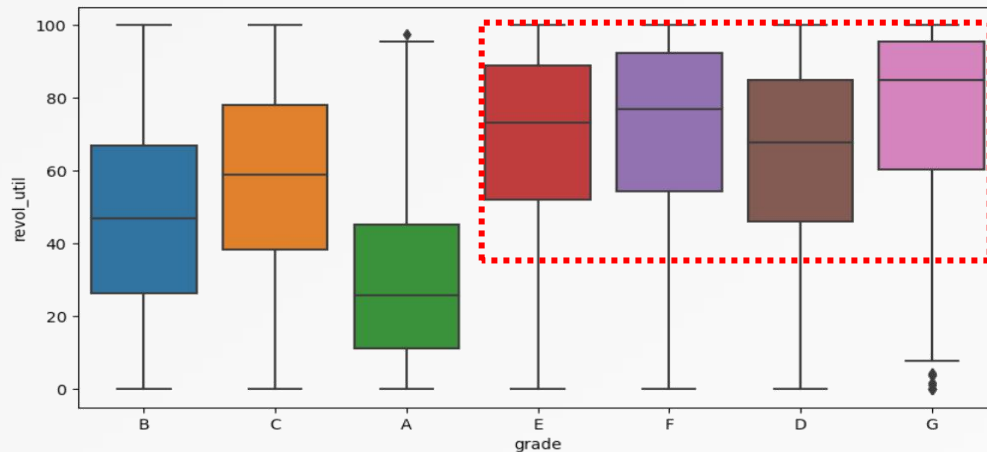
Loan amount, Grade & Revolving line utilization rate

field = loan_amnt Category = grade

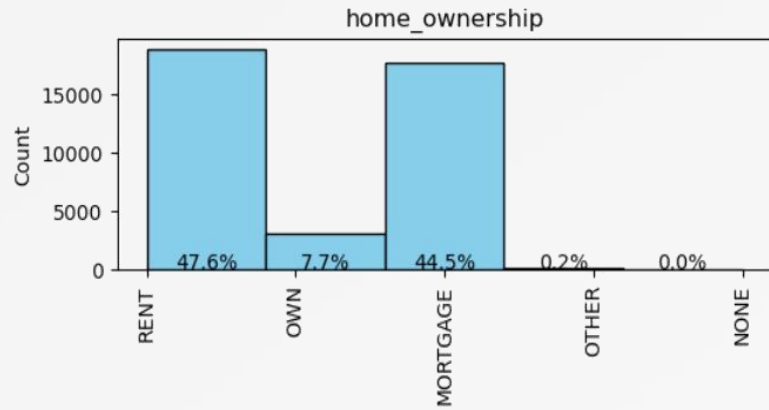


- Loan amount applied by the borrower for G grade loans is highest where as for A grade Loans is lowest.
- For upper grade loans possess **high loan amount** applied by borrowers - **loss is high** if a loan is defaulted
- Upper grade loans have higher line utilization rate and that correlates with higher % of defaults

field = revol_util Category = grade

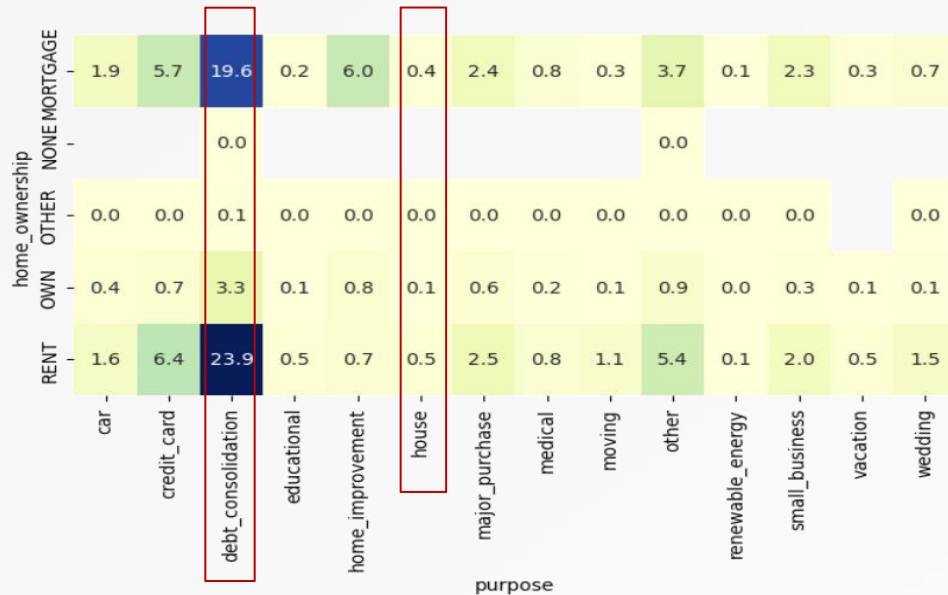


Home Ownership

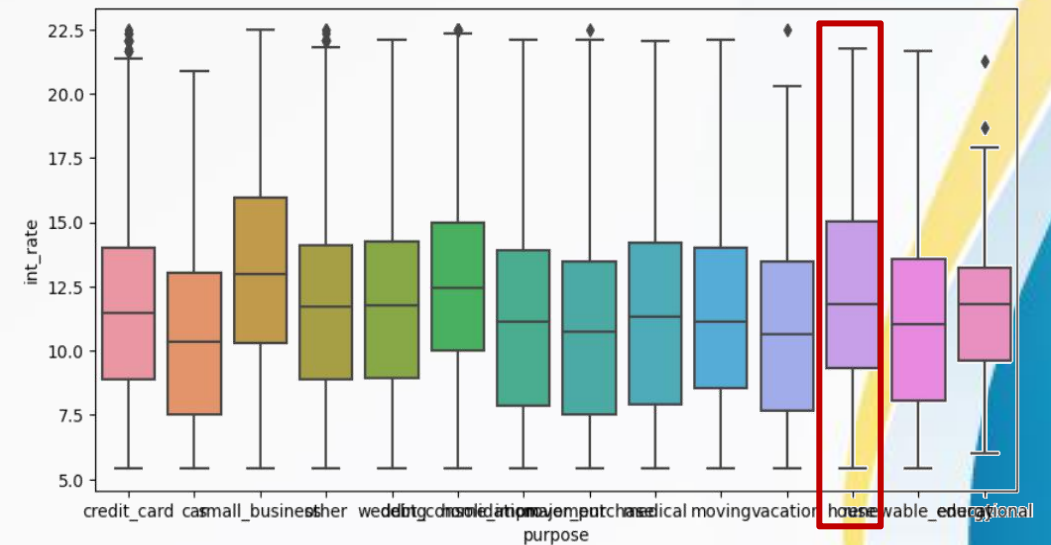


- 47% of the borrowers don't have house (paying rent) but only 1% of loans are home loans.
- home loans are not getting sold effectively. maybe because of higher interest rate
- And **47%** borrowers' purpose is '**debt_consolidation**'.

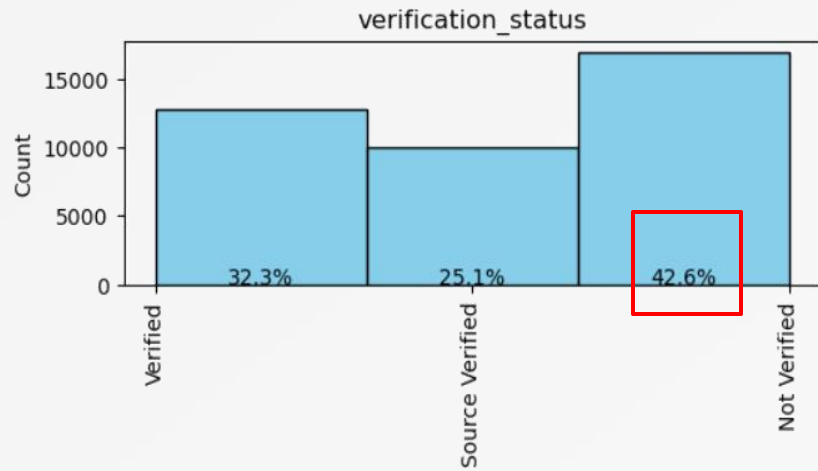
===== % of Loans applications distributed in home_ownership and purpose =====



field = int_rate Category = purpose



Verification Status

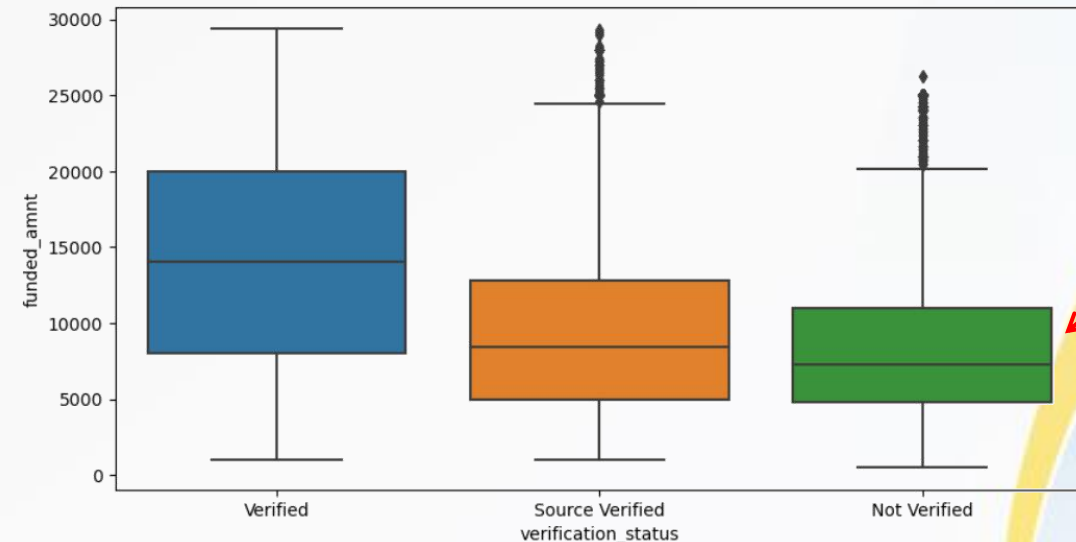


===== Defaulted Loan Status (in %) in field VERIFICATION_STATUS =====

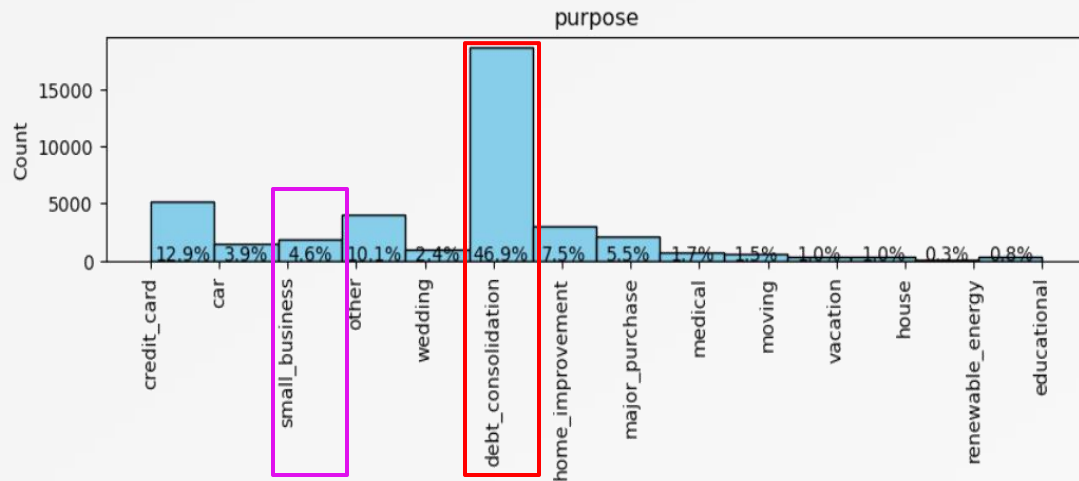


- **42% of the borrowers' income is 'not verified'** which is risky.
- Chance of non-verified loan getting defaulted is 12 %
- Total amount funded by the company to non-verified loans is less compared to verified loans

field = funded_amnt Category = verification_status

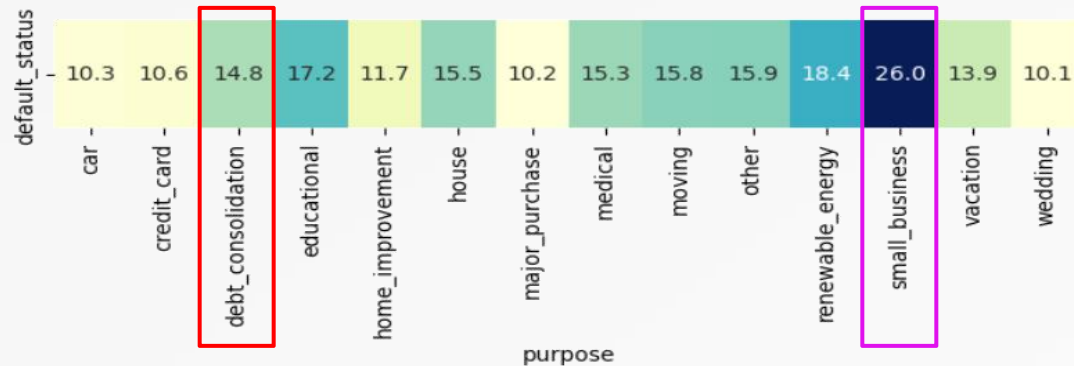


Purpose

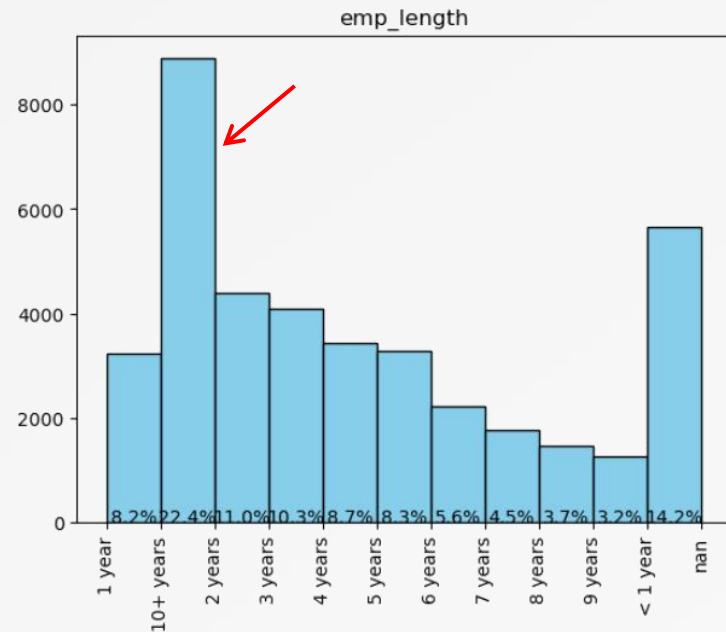


- **Debt_consolidation** is the most common purpose of the loan (47%) and has **14.8%** chances of getting default.
- Loans with small business purpose have highest chances (26%) of getting defaulted.
- Whereas loans for wedding, car or major purchase have lowest chances (only 10%) of getting defaulted.

===== Defaulted Loan Status (in %) in field PURPOSE =====

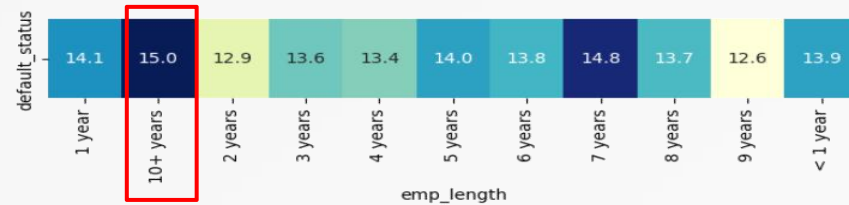


Employment Length



- Majority of the borrowers (22.4%) have 10+ years of employment length.
- There is a 15% chance of loan getting default when employment length of the borrower is 10+ years.

===== Defaulted Loan Status (in %) in field EMP_LENGTH =====

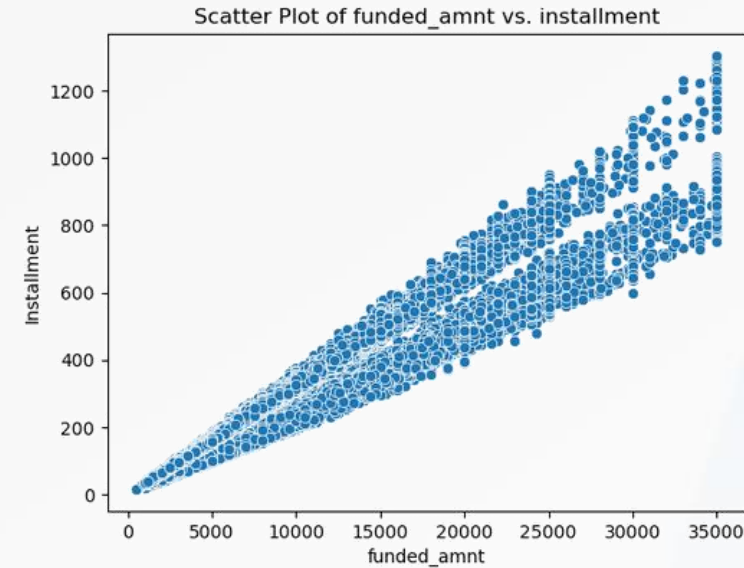
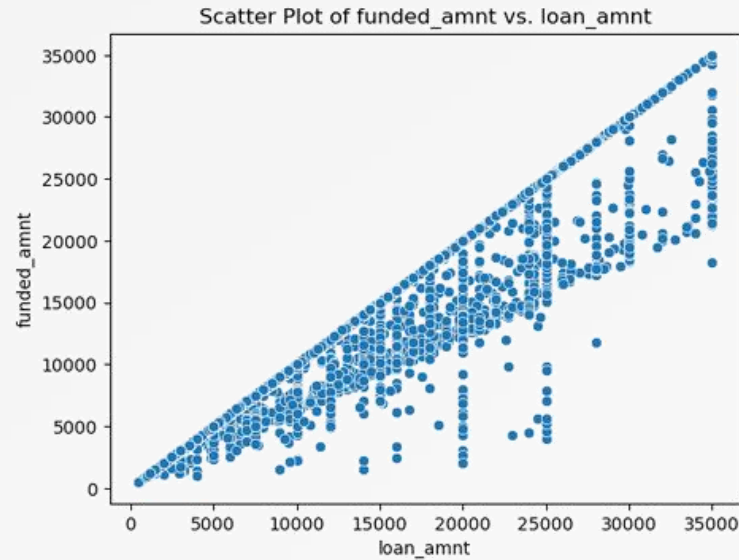


Correlation Between the fields

	loan_amnt	funded_amnt	funded_amnt_inv	int_rate	installment	annual_inc	dti	total_rec_pncp	revol_bal	revol_util	total_acc	open_acc
loan_amnt	1.00	0.98	0.94	0.31	0.93	0.27	0.07	0.85	0.32	0.07	0.26	0.18
funded_amnt	0.98	1.00	0.96	0.31	0.96	0.27	0.07	0.87	0.31	0.07	0.25	0.18
funded_amnt_inv	0.94	0.96	1.00	0.31	0.91	0.25	0.07	0.85	0.29	0.07	0.24	0.16
int_rate	0.31	0.31	0.31	1.00	0.28	0.05	0.11	0.19	0.10	0.47	-0.04	0.01
installment	0.93	0.96	0.91	0.28	1.00	0.27	0.05	0.85	0.31	0.10	0.23	0.17
annual_inc	0.27	0.27	0.25	0.05	0.27	1.00	-0.12	0.26	0.28	0.02	0.24	0.16
dti	0.07	0.07	0.07	0.11	0.05	-0.12	1.00	0.04	0.23	0.28	0.23	0.29
total_rec_pncp	0.85	0.87	0.85	0.19	0.85	0.26	0.04	1.00	0.28	0.03	0.23	0.16
revol_bal	0.32	0.31	0.29	0.10	0.31	0.28	0.23	0.28	1.00	0.30	0.31	0.29
revol_util	0.07	0.07	0.07	0.47	0.10	0.02	0.28	0.03	0.30	1.00	-0.07	-0.09
total_acc	0.26	0.25	0.24	-0.04	0.23	0.24	0.23	0.23	0.31	-0.07	1.00	0.69
open_acc	0.18	0.18	0.16	0.01	0.17	0.16	0.29	0.16	0.29	-0.09	0.69	1.00
	loan_amnt	funded_amnt	funded_amnt_inv	int_rate	installment	annual_inc	dti	total_rec_pncp	revol_bal	revol_util	total_acc	open_acc

- **Very high correlation** between **loan amount** applied by the borrower and **funded amount** committed by the investor. We can say that if loan application of the borrower gets passed by company, then he gets almost full amount of the loan applied.
- Certainly **loan amount**, monthly **installment** and **principal received** are **highly correlated**. Higher the loan amount, higher is monthly installment and higher is principal received vice versa.
- Number of **open credit lines** is also **highly** correlated with **total number of credit lines** in the borrower's credit file.
- **Revol_util, int_rate - Moderately high correlation**
 - Higher the amount of credit the borrower is using, higher is the risk of loan repayment. That's why higher is the interest rate, and vice-versa.

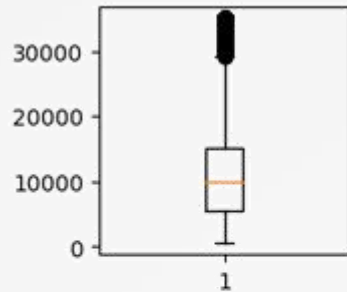
Loan Amount, Funded Amount, Installment



- Funded amount by the company is always less than or equal to loan amount applied by borrower.
- Funded amount by the company is highly correlated with monthly installment paid by the borrower.
- There is a **high** correlation between loan amount, funded amount and installment

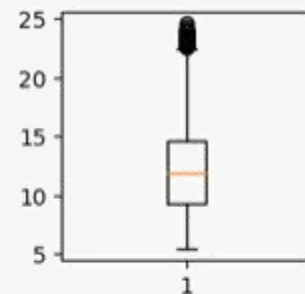
Loan Amount, Interest rate, Total Received principal

loan_amnt



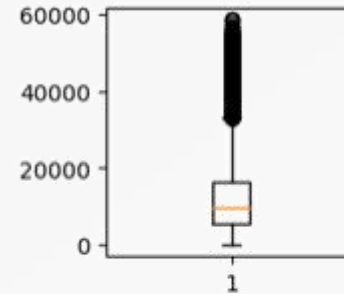
count	39717.000000
mean	11219.443815
std	7456.670694
min	500.000000
25%	5500.000000
50%	10000.000000
75%	15000.000000
max	35000.000000

int_rate



count	39717.000000
mean	12.021177
std	3.724825
min	5.420000
25%	9.250000
50%	11.860000
75%	14.590000
max	24.590000

total_pymnt



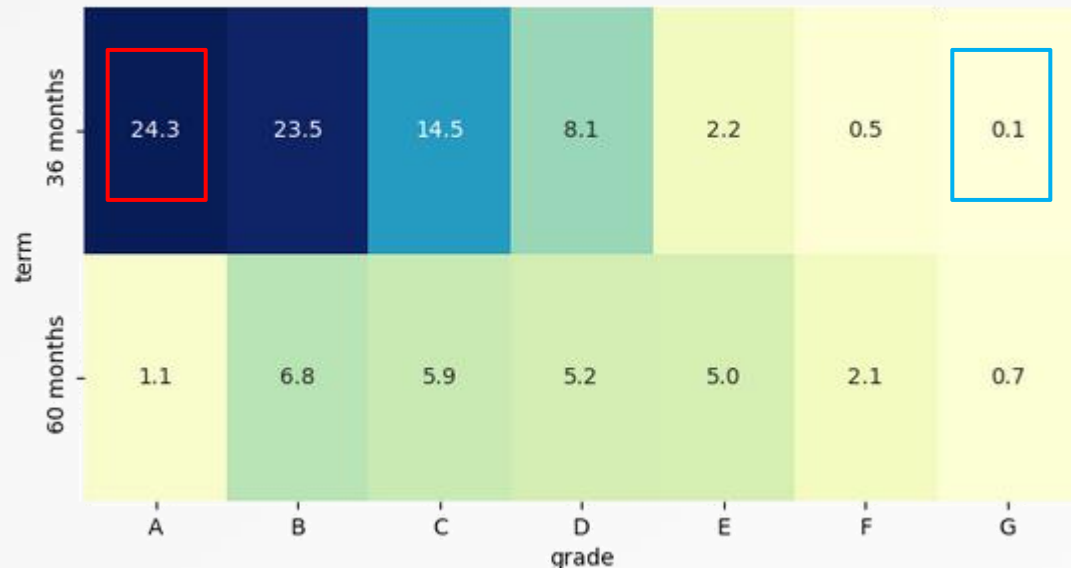
count	39717.000000
mean	12153.596544
std	9042.040766
min	0.000000
25%	5576.930000
50%	9899.640319
75%	16534.433040
max	58563.679930

- Avg loan amount applied by the borrower is 11.2K
- Avg interest rate for the loan is 12%
- Avg payment received by the LC is 12.2k. Ideally it should be $(11.2 * 1.12 =) 12.5k$ but because of some default loans, company is getting 12.2k payment.
- If the company reduces the default loans in future, then LC can make more profits.

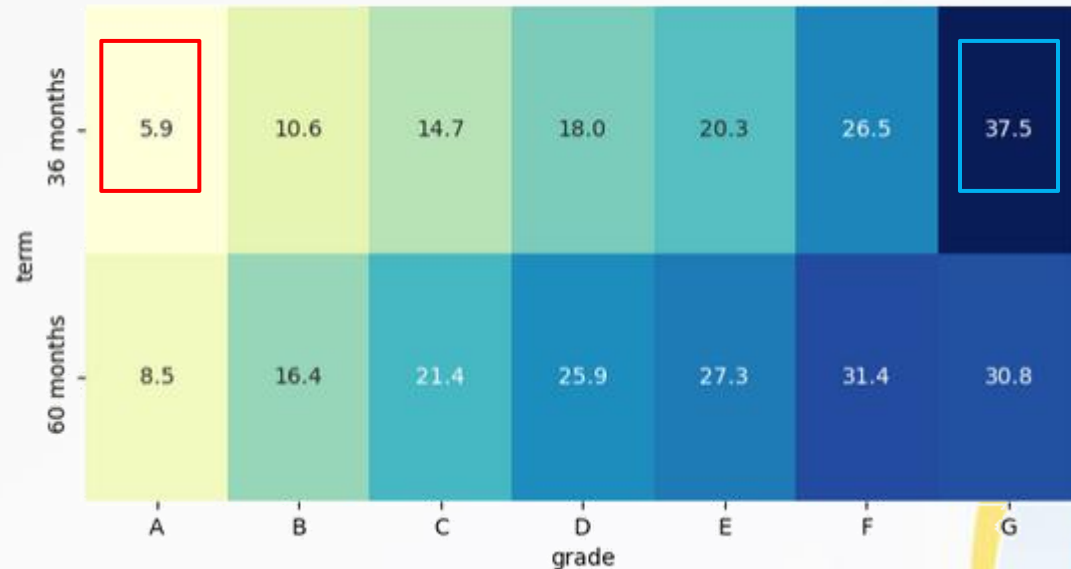
Impact of Term and Grade - Loan Status

- Highest number of loans (24%) have 36 months tenure with grade A and have lowest default rate (5.9%). This is very good for company.
- The company has only 0.1% of total loans with 36 months tenure and G grade. And these loans have highest risk of 37.5% of default rate.

===== % of Loans applications distributed in term and grade =====



===== Impact of TERM and GRADE together on LOAN STATUS =====



Impact of Emp_length and Verification Status - Loan Status

- Borrowers who have 10+ years of employment years and whose income is verified by lending company are highest 9%. And their default rate is maximum i.e. 17.2%.
- This is not a good situation for the company.
- This is the criteria where company should look into it.

% of Loans applications distributed in emp_length and verification_status

emp_length	Not Verified	Source Verified	Verified
1 year	3.9	2.1	2.2
10+ years	8.4	5.0	9.0
2 years	5.2	2.8	3.1
3 years	4.6	2.6	3.1
4 years	3.7	2.1	2.8
5 years	3.6	2.1	2.6
6 years	2.4	1.4	1.8
7 years	1.8	1.1	1.5
8 years	1.6	0.9	1.3
9 years	1.4	0.8	1.0
< 1 year	5.0	3.7	2.8

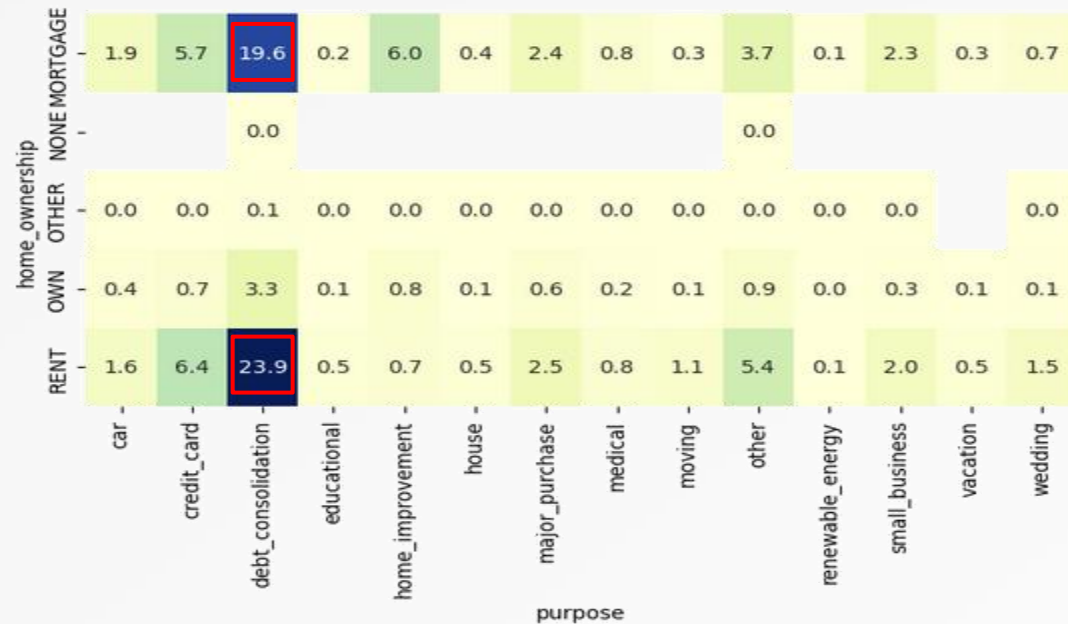
===== Impact of EMP_LENGTH and VERIFICATION_STATUS together on LOAN STATUS =====

emp_length	Not Verified	Source Verified	Verified
1 year	13.3	14.6	15.0
10+ years	12.6	14.9	17.2
2 years	11.2	14.2	14.6
3 years	11.6	15.7	14.6
4 years	12.0	13.5	15.3
5 years	12.6	14.1	15.8
6 years	11.5	14.8	15.8
7 years	13.5	15.6	15.9
8 years	12.2	11.3	17.3
9 years	11.6	13.8	12.9
< 1 year	13.8	12.9	15.6

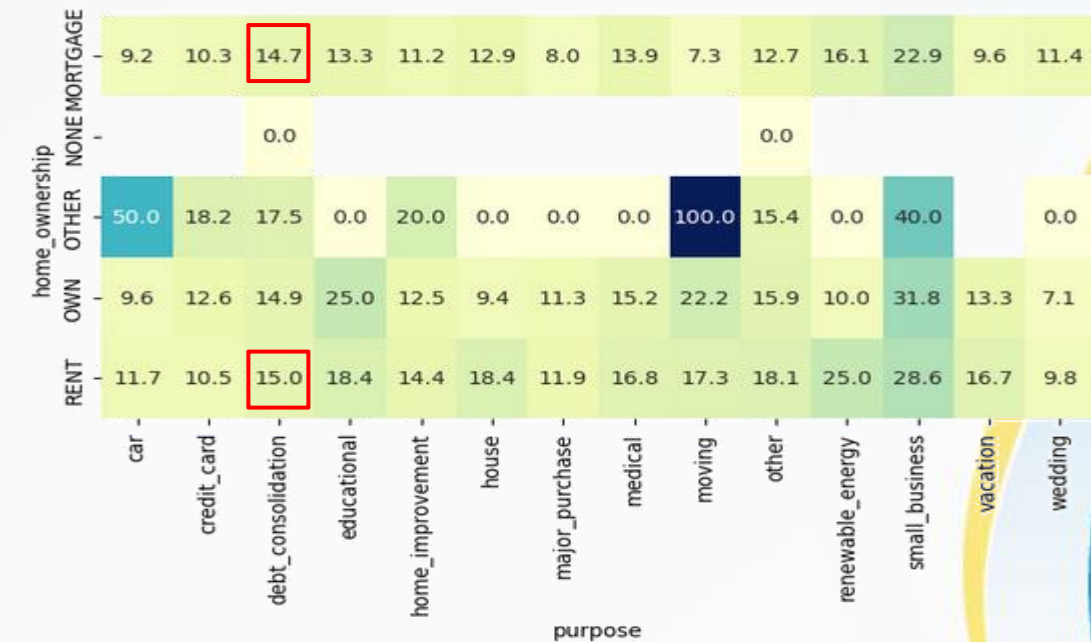
Impact of Home_ownership & Purpose - Loan Status

- High chunk of loans (23.9% + 19.6% = **43.5%**) are taken for debt_consolidation purpose and the borrowers of these loans stay in rented house or Mortgaged their house.
- And there is almost 15% chance of loan getting defaulted which is high and serious.

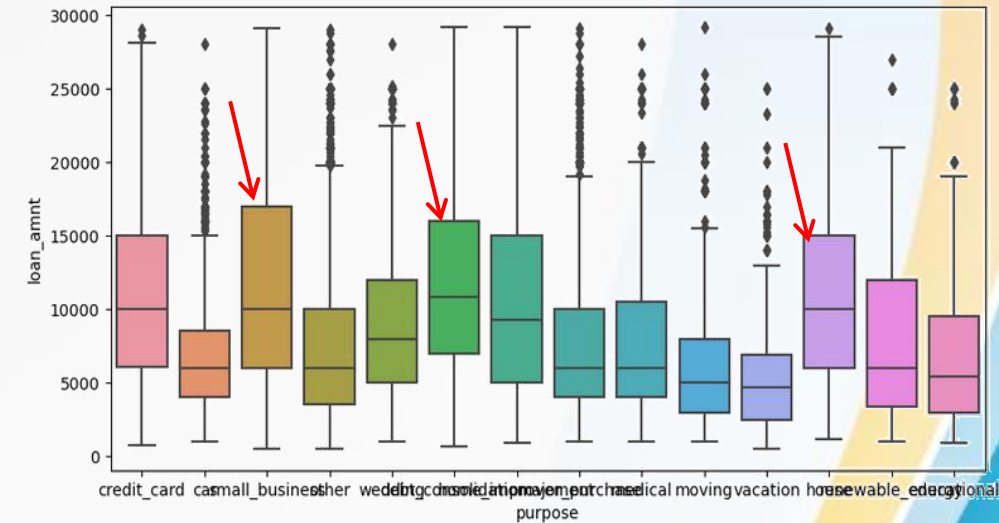
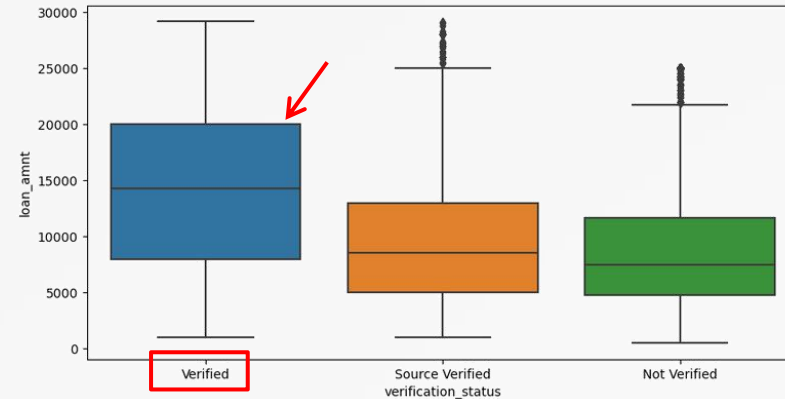
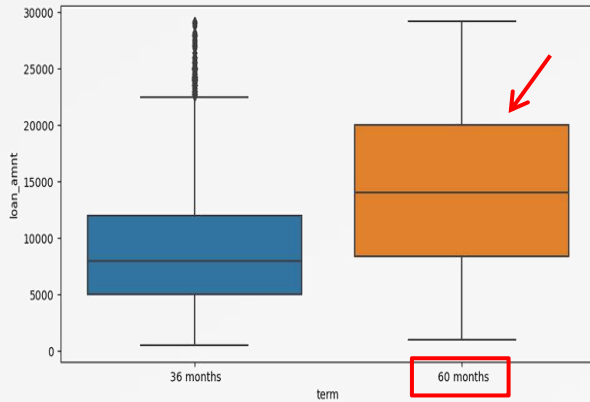
% of Loans applications distributed in home_ownership and purpose



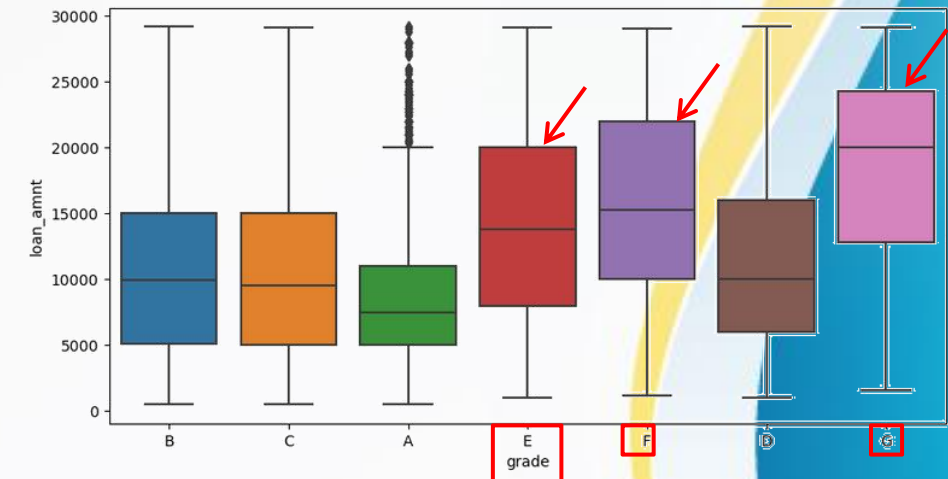
===== Impact of HOME_OWNERSHIP and PURPOSE together on LOAN STATUS =====



Loan Amount for different types of loans



- On an average loan amount is high for :
 - 60 month tenure loans
 - Verified borrower's income loans
 - upper grade loans
 - small business, house, debt consolidation purpose loans



Conclusion

On an average **14% of the loans gets defaulted.**

Driving Factors for Loan Default are:

- Lengthy tenure Loans - Chances of getting defaulted is double (22%) that of less tenure loans (11%).
- Upper Grade Loans - Higher the grade, higher is the interest rate, and high chances of getting default
- High Revol_util - Higher the amount of credit the borrower is using, higher is the risk of loan repayment.

Inferences :

- High loan amount is given to 60 months tenure loans, upper grade loans, small business loans where chance of getting defaulted is very high.
- Default rate for the borrowers, who have 10+ years of employment years and their income is also verified, is high
- For upper grade loans possess **high loan amount** applied by borrowers - **loss is high** if a loan is defaulted
- High Interest rates - Interest rate for 60 months tenure loans is higher than that of 36 months loan.