LENDING CLUB CASE STUDY

PROBLEM STATEMENT

When the company receives a loan application, the company has to make a decision for loan approval based on the # applicant’s profile

If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Loading the Lending Club Case Study

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_past_
0	1077501	1296599	5000	5000	4975	36 months	10.65%	162.87	B	B2	...	NaN	
1	1077430	1314167	2500	2500	2500	60 months	15.27%	59.83	C	C4	...	NaN	
2	1077175	1313524	2400	2400	2400	36 months	15.96%	84.33	C	C5	...	NaN	
3	1076863	1277178	10000	10000	10000	36 months	13.49%	339.31	C	C1	...	NaN	
4	1075358	1311748	3000	3000	3000	60 months	12.69%	67.79	B	B5	...	NaN	

5 rows × 111 columns

Observation: Data is loaded by means of Pandas Data Frame and it consists of 5 rows and 111 columns

INVESTIGATING NULL VALUES

OBSERVATION: It is inferred that 54 columns have null values

DATA CLEANING

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	collections_12_mths_ex_med	mths_sinc
0	1077501	1296599	5000	5000	4975	36 months	10.65%	162.87	B	B2	...	0	
1	1077430	1314167	2500	2500	2500	60 months	15.27%	59.83	C	C4	...	0	
2	1077175	1313524	2400	2400	2400	36 months	15.96%	84.33	C	C5	...	0	
3	1076863	1277178	10000	10000	10000	36 months	13.49%	339.31	C	C1	...	0	
4	1075358	1311748	3000	3000	3000	60 months	12.69%	67.79	B	B5	...	0	

5 rows × 59 columns

OBSERVATION: Now column with null values are removed and reduced to 59 columns

	id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	grade	sub_grade	emp_length	home_ownership	...	total_rec_prncp	total_rec_int	tot
0	1077501	5000	5000	4975	36 months	10.65%	B	B2	10+ years	RENT	...	5000	863.16	
1	1077430	2500	2500	2500	60 months	15.27%	C	C4	< 1 year	RENT	...	456.46	435.17	
2	1077175	2400	2400	2400	36 months	15.96%	C	C5	10+ years	RENT	...	2400	605.67	
3	1076863	10000	10000	10000	36 months	13.49%	C	C1	10+ years	RENT	...	10000	2214.92	
4	1075358	3000	3000	3000	60 months	12.69%	B	B5	1 year	RENT	...	2475.94	1037.39	

5 rows × 40 columns

OBSERVATION: Now unnecessary columns such as desc, mths_since_last_delinq, mths_since_last_record, next_pymnt_d, tot_hi_cred_lim, mths_since_last_major_derog, total_bal_ex_mort, total_bc_limit, total_il_high_credit_limit, member_id, url, emp_title, zip_code, tax_liens, application_type, policy_code, initial_list_status, installment, pymnt_plan are removed

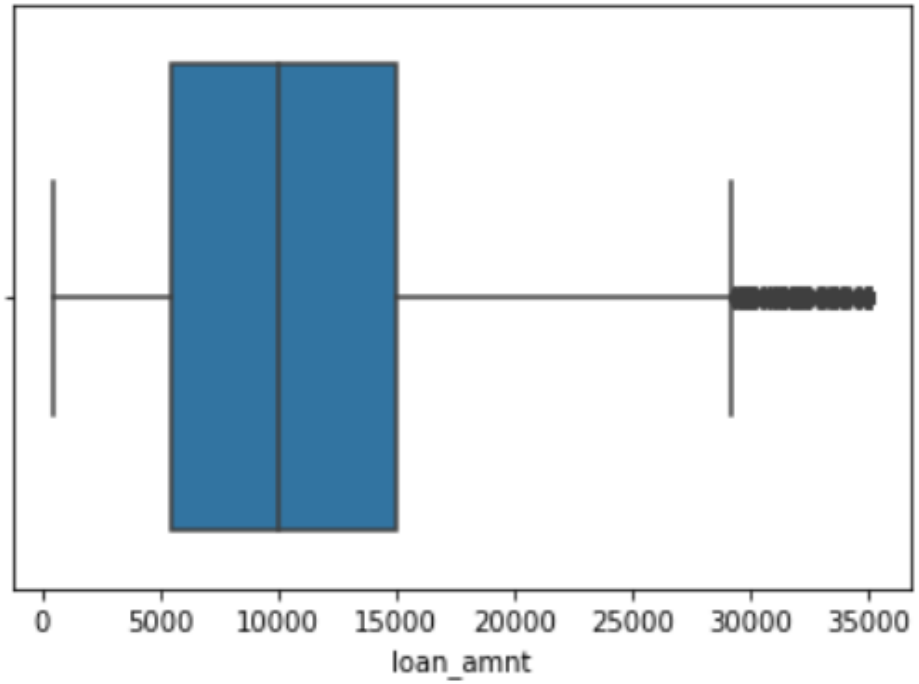
FILLING MISSING VALUES

	id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	grade	sub_grade	emp_length	home_ownership	...	total_rec_prncp	total_rec_int
0	1077501	5000	5000	4975	36 months	10.65%	B	B2	10	RENT	...	5000	863.16
1	1077430	2500	2500	2500	60 months	15.27%	C	C4	1	RENT	...	456.46	435.17

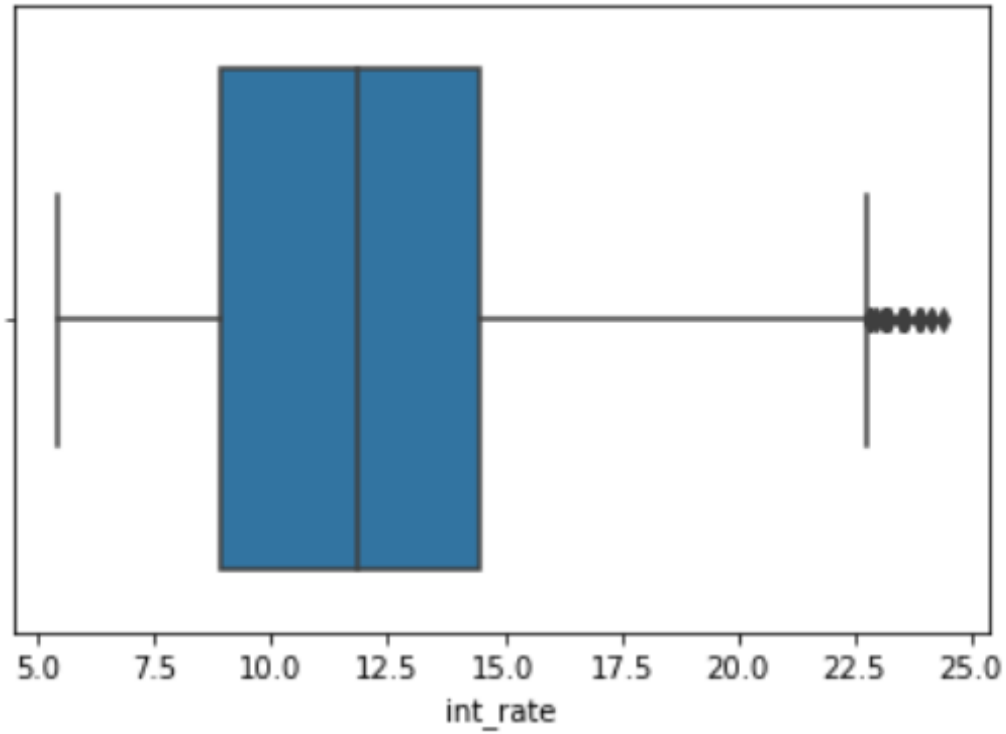
OBSERVATION: Now Null values in Employee Experience and Public Bankrupties has been replace with zero and Not Known values respectively

Extra character such as plus and signs such as < , > has been removed and these columns are converted to numeric

UNIVARIATE ANALYSIS



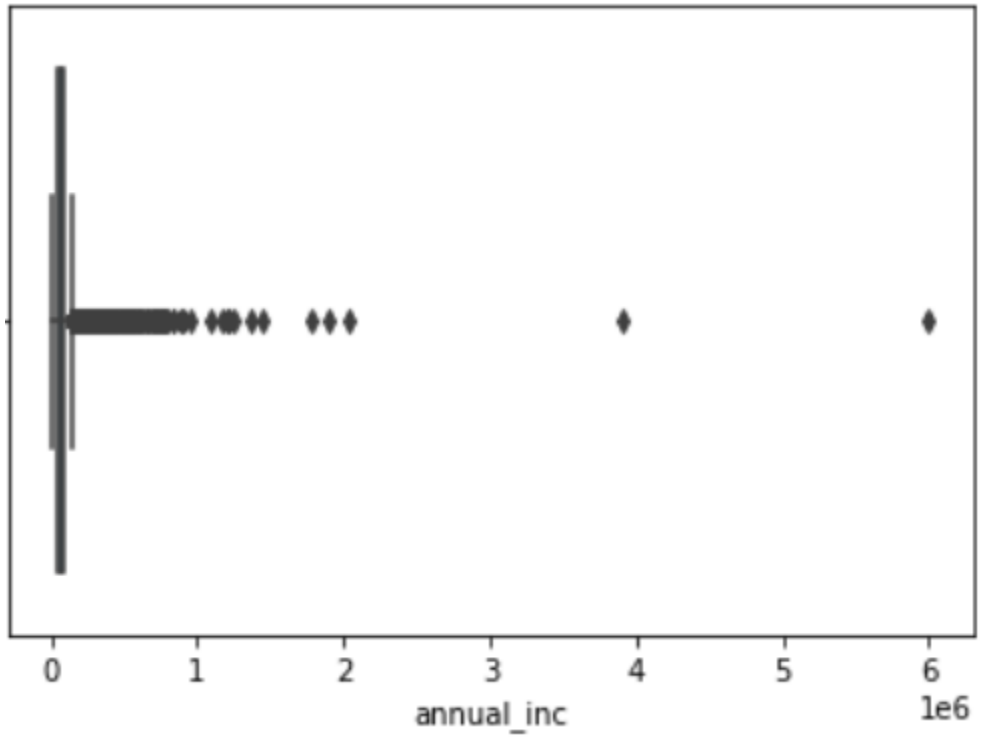
Observations : 75 % of loan amount lies below 15000 rupees



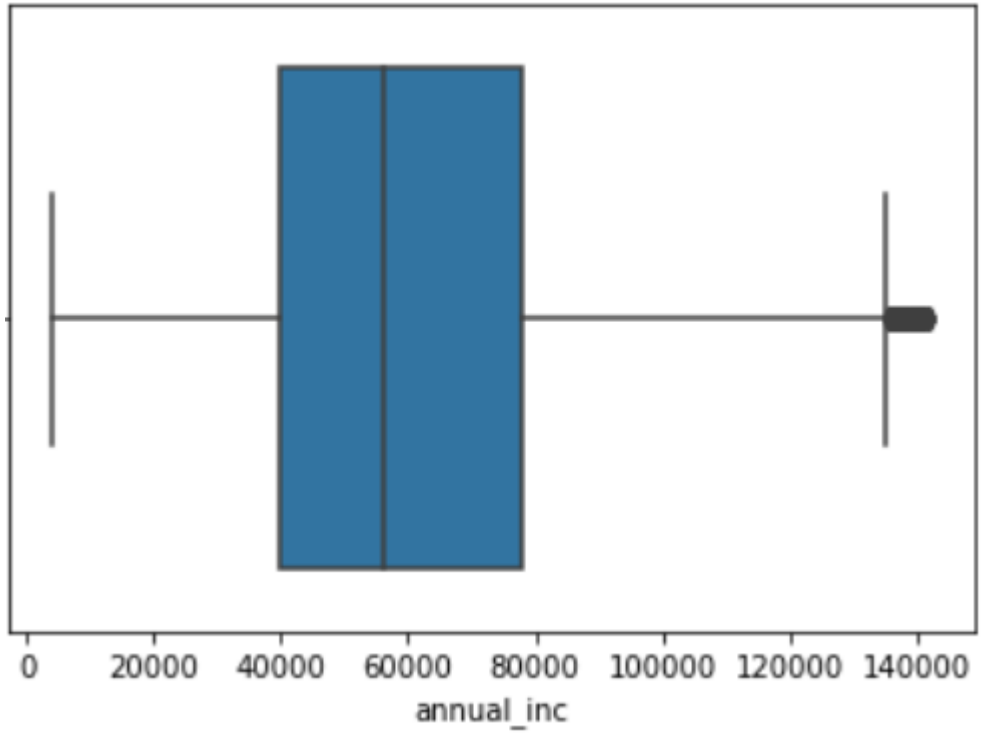
Observations : 75 % of loan amount lies below 15%

Outlier Treatment

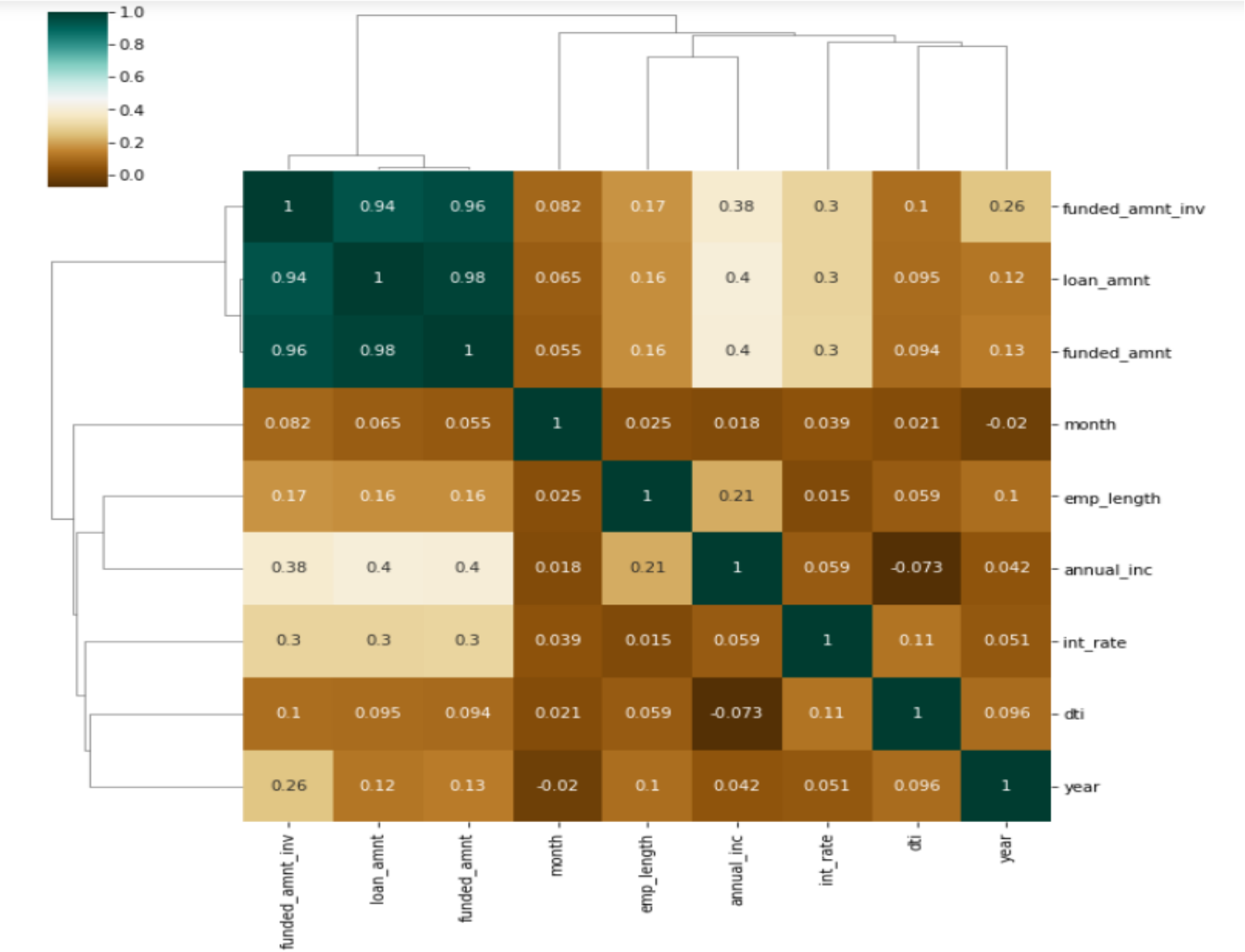
Before Removing Outliers



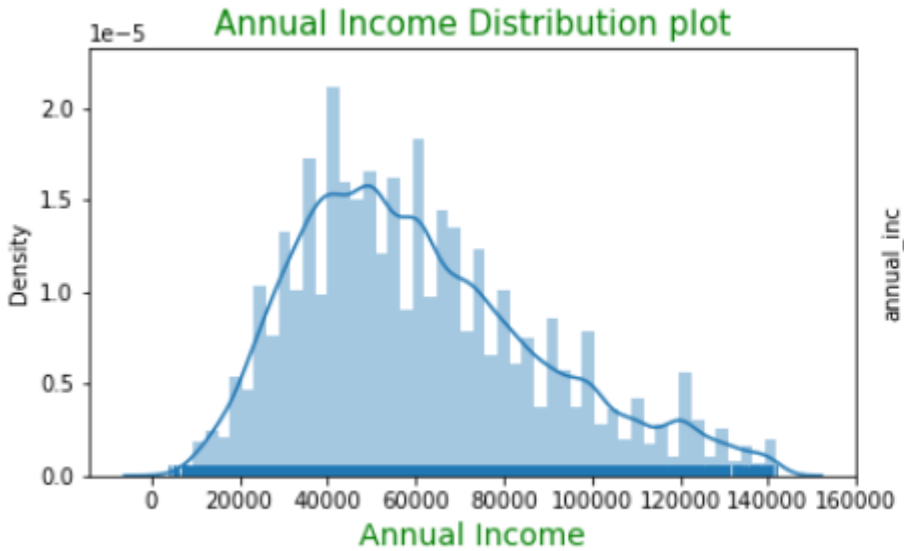
After Removing Outliers

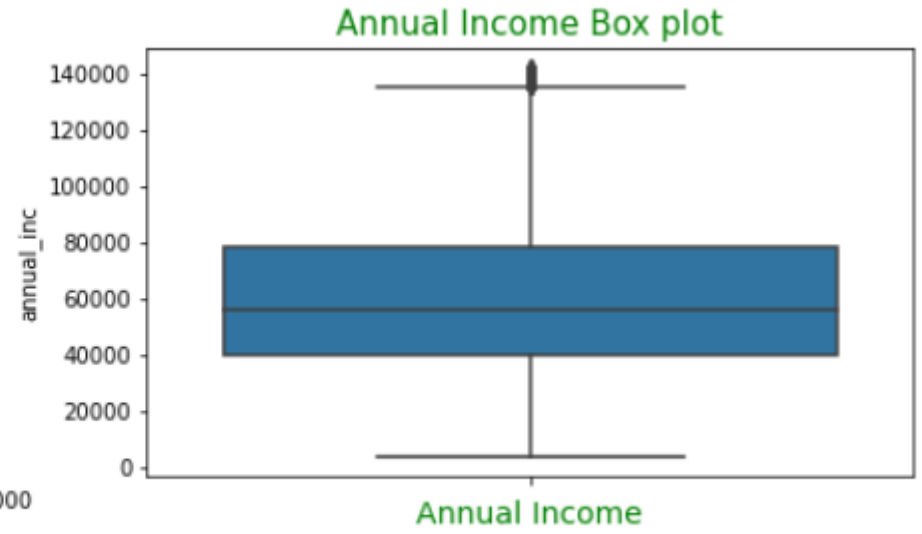


Observations : Here data above 95%(Quantile) of Annual Income is considered as outlier and removed



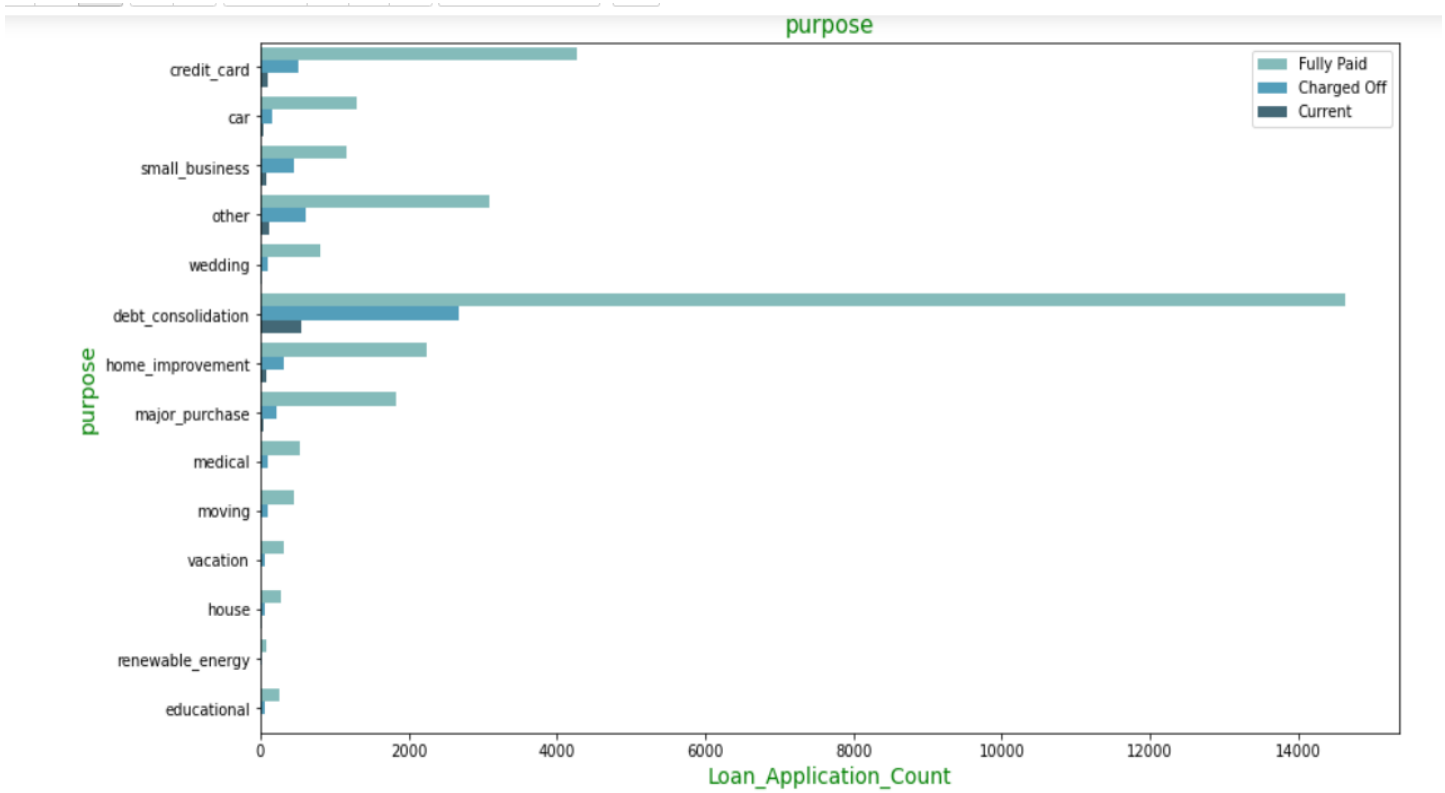
Observations : Here loan amount, investor amount and funding amount are strongly correlated. Income increases with work experience





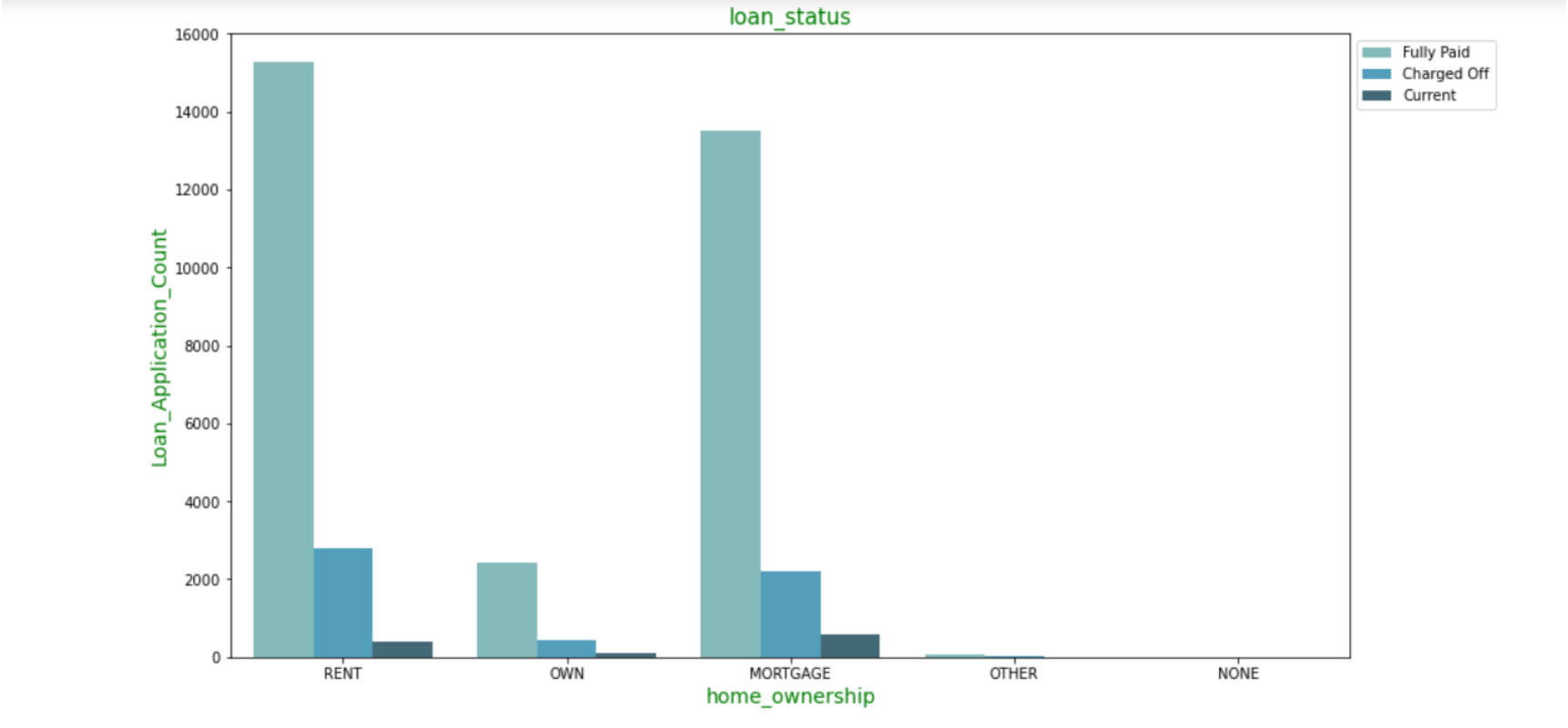
Observations : Here range of annual income is in the range of 40000 to 80000 and box plot shows better visulaisation than distribution plot

Univariate Analysis Categorical Variables



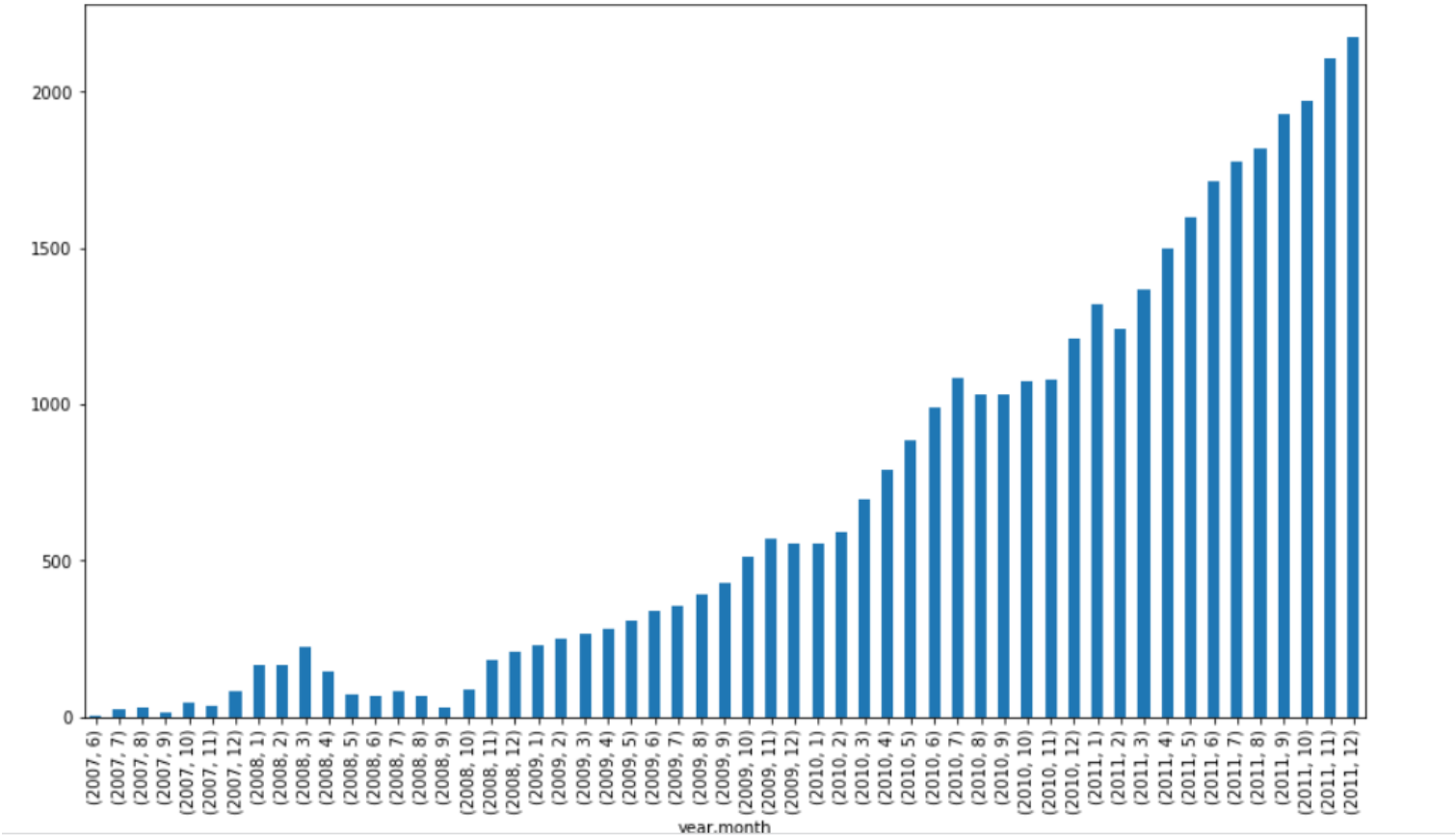
Observations : Below plot shows that most of the loans were taken for the purpose of debt consolidation & paying credit card bill.

Number of chraged off count also high too for these loans.



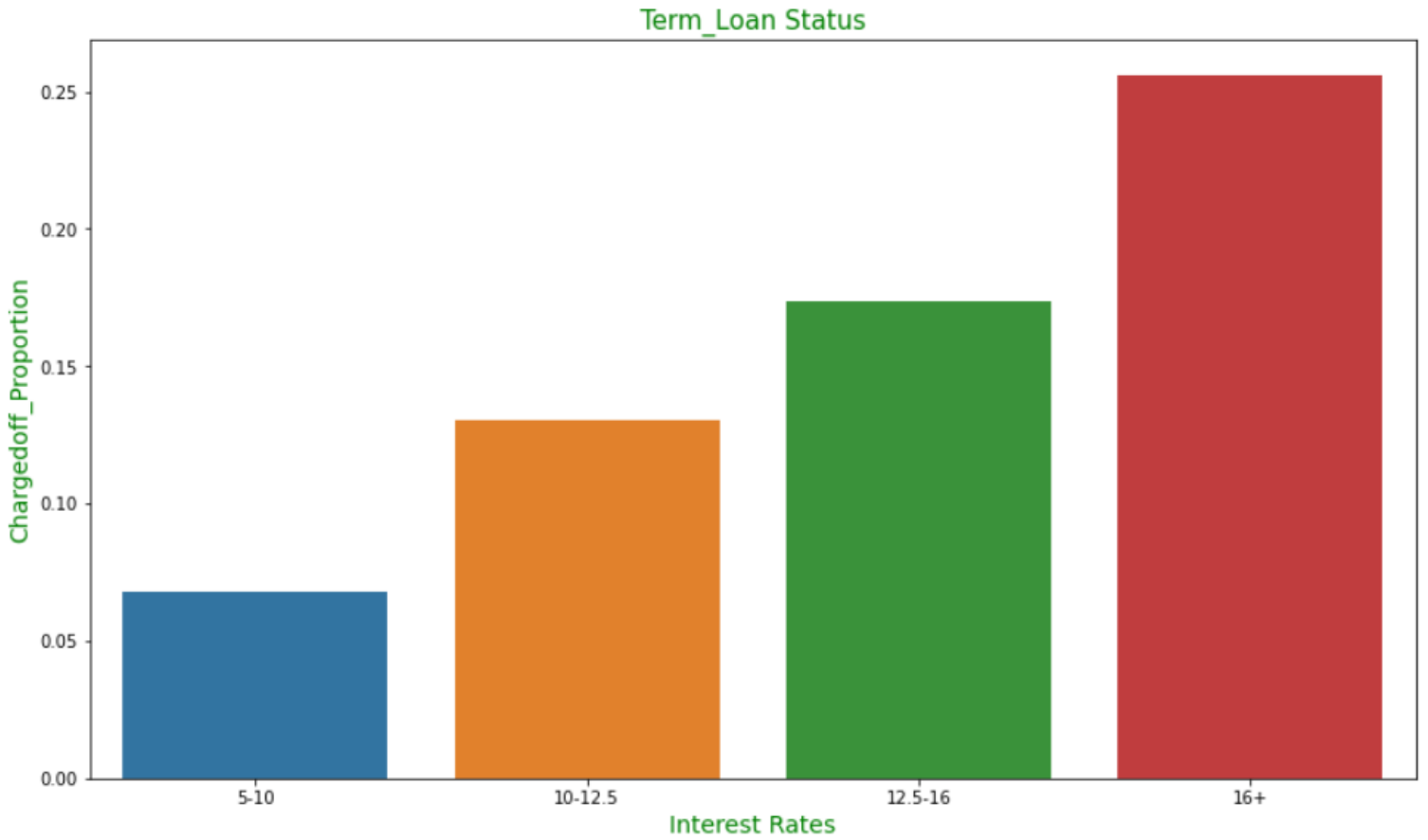
Observations : Plot shows that most of them are in rented house

Segmented Univariate Analysis



Observations : Loan Applicants are increasing every year. Loans from 2008(May-October) got dipped

Bivariate Analysis



Observations : Interest rates with more than 16+ has higher Charged off Propotions compared to other interest rates