Img src: google

For

Group Facilitator:
Vernon Basil Mascarenhas

Group Member:
Anand Joshi

LendingClub

Introduction:

- We were asked by Lending club to try and figure out factors and other features of borrowers who apply for a loan in LC to minimize the number of defaulters which put into simple terms would imply approving and issuing loan to the borrowers who can repay the loan on time.
- The approach to the given business problem is to first identify useful features from the given dataset and disregard the rest
- Then we will proceed and analyze the dataset to remove certain rows of data which tends to give some logical errors and by this we mean to simply remove the rows of data which don't belong there
- After all the cleaning is done we visualize the data in the form of bar charts which we will see shortly in the upcoming slides and derive meaningful insights from it
- Finally we draw our conclusion and determine the factors resulting in borrowers defaulting and present the same in the last slide
- So without any further ado

Img src: google

acc_now_delinq	emp_length	last_fico_range_low	mths_since_recent_revol_delinq	open_il_24m	title
acc_open_past_24mths	emp_title	last_pymnt_amnt	next_pymnt_d	open_il_6m	tot_coll_amt
addr_state	fico_range_high	last_pymnt_d	num_accts_ever_120_pd	open_rv_12m	tot_cur_bal
all_util	fico_range_low	loan_amnt	num_actv_bc_tl	open_rv_24m	tot_hi_cred_lim
annual_inc	funded_amnt	loan_status	num_actv_rev_tl	out_prncp	total_acc
annual_inc_joint	funded_amnt_inv	max_bal_bc	num_bc_sats	out_prncp_inv	total_bal_ex_mort
application_type	grade	member_id	num_bc_tl	pct_tl_nvr_dlq	total_bal_il
avg_cur_bal	home_ownership	mo_sin_old_il_acct	num_il_tl	percent_bc_gt_75	total_bc_limit
bc_open_to_buy	id	mo_sin_old_rev_tl_op	num_op_rev_tl	policy_code	total_cu_tl
bc_util	il_util	mo_sin_rcnt_rev_tl_op	num_rev_accts	pub_rec	total_il_high_credit_limit
chargeoff_within_12_mths	initial_list_status	mo_sin_rcnt_tl	num_rev_tl_bal_gt_0	pub_rec_bankruptcies	total_pymnt
collection_recovery_fee	inq_fi	mort_acc	num_sats	purpose	total_pymnt_inv
collections_12_mths_ex_med	inq_last_12m	mths_since_last_delinq	num_tl_120dpd_2m	pymnt_plan	total_rec_int
delinq_2yrs	inq_last_6mths	mths_since_last_major_derog	num_tl_30dpd	recoveries	total_rec_late_fee
delinq_amnt	installment	mths_since_last_record	num_tl_90g_dpd_24m	revol_bal	total_rec_prncp
desc	int_rate	mths_since_rcnt_il	num_tl_op_past_12m	revol_util	total_rev_hi_lim
dti	issue_d	mths_since_recent_bc	open_acc	sub_grade	url
dti_joint	last_credit_pull_d	mths_since_recent_bc_dlq	open_acc_6m	tax_liens	verification_status
earliest_cr_line	last_fico_range_high	mths_since_recent_inq	open_il_12m	term	verified_status_joint
					zip_code

What we were given

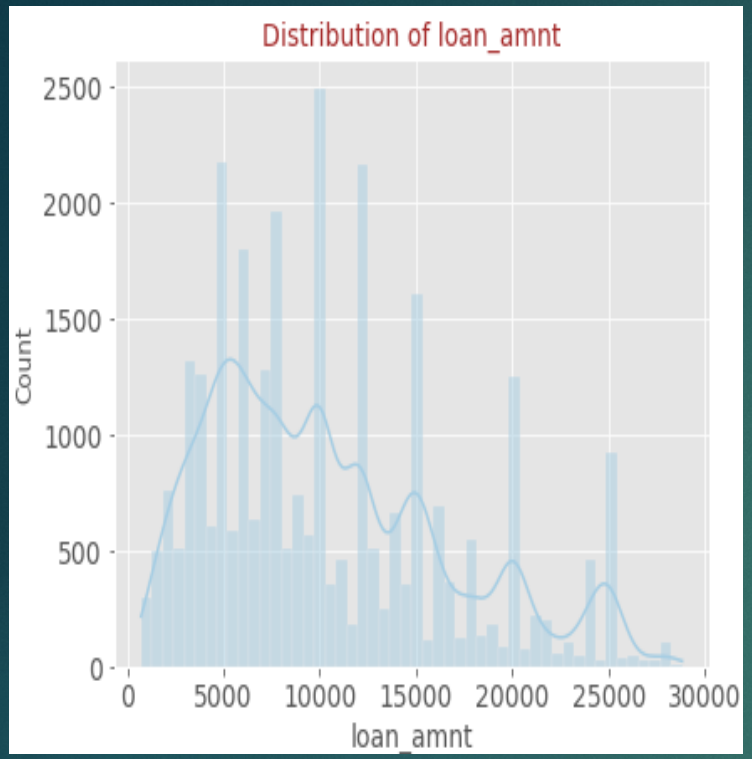
What we mainly used

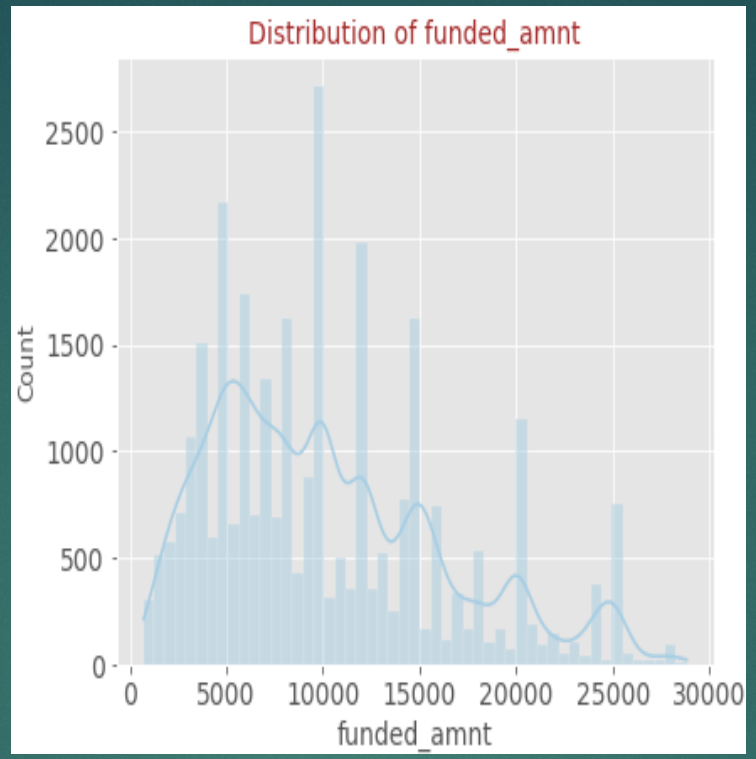
annual_inc
funded_amnt
funded_amnt_inv
grade
int_rate
loan_amnt
loan_status
pub_rec_bankruptcies
revol_util
term
verification_status

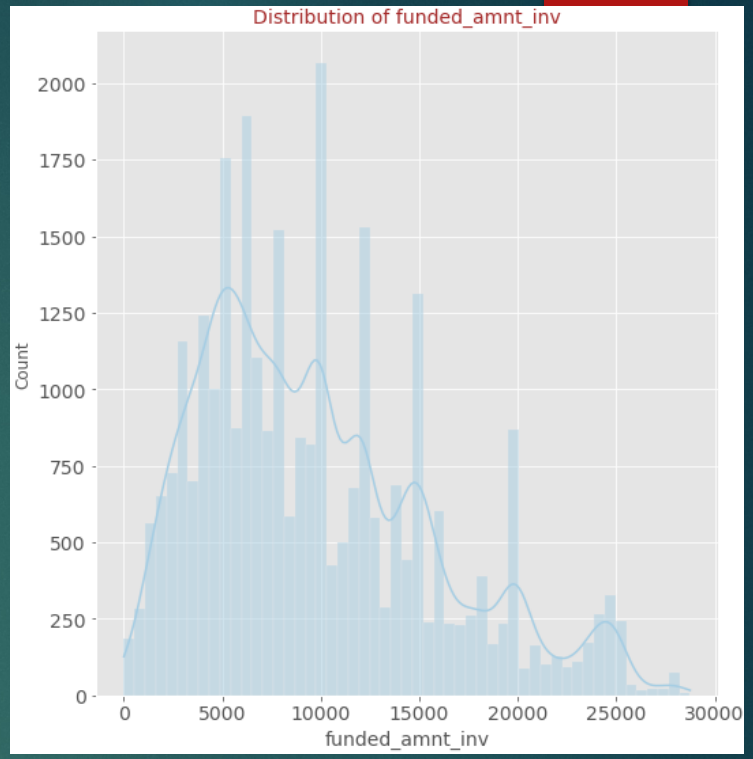
- First we loaded our dataset into our environment and got a over all feel of how the dataset actually looks like, we observed some discrepancies, type issues with certain columns
- We dealt all of this in our data cleaning phase which simply means that we removed or edited some information best suited for our dataset
- After this we began with some visualization to ensure that we do not have any hidden discrepancy or data which doesn't belong in our dataset
- We started with our univariate analysis and got the true hang of the dataset nothing extra ordinary was uncovered here
- After that came our bi variate analysis which helped us to observe how certain features which we saw in our previous slides behaved with each other, we did uncover some insights but in some cases we were required to introduce a third variable
- Over here in Multi variate analysis with three variables we saw how the features inter twined with each other and extracted al the features which to the best of our knowledge will help Lending club to increase their profits

Information about the features we used

annual_inc	The self-reported annual income provided by the borrower during registration.
funded_amnt	The total amount committed to that loan at that point in time.
funded_amnt_inv	The total amount committed by investors for that loan at that point in time.
grade	LC assigned loan grade
int_rate	Interest Rate on the loan
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
loan_status	Current status of the loan
pub_rec_bankruptcies	Number of public record bankruptcies
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified
purpose	A category provided by the borrower for the loan request.



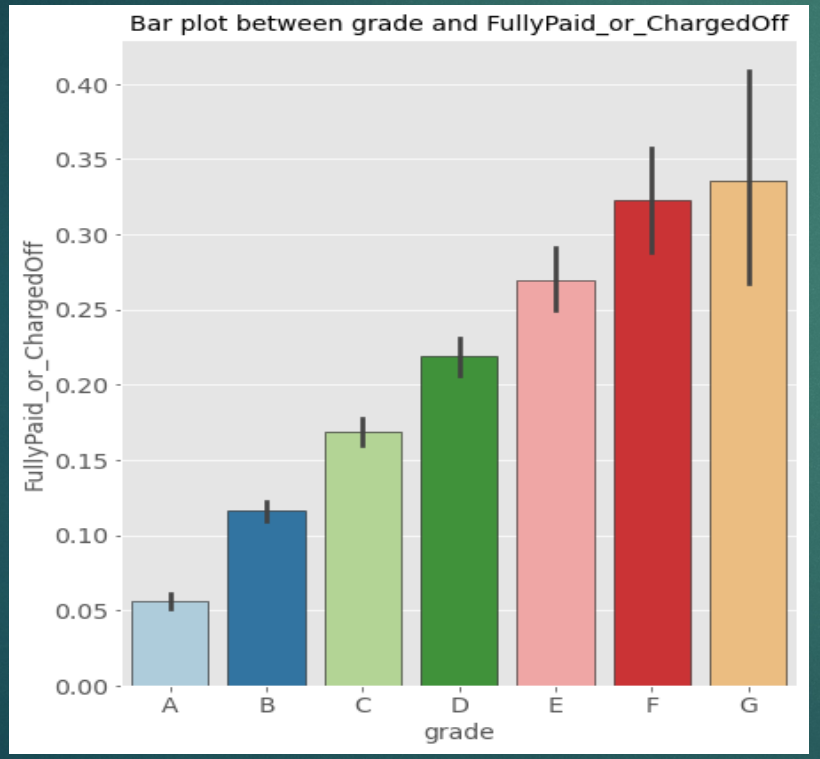


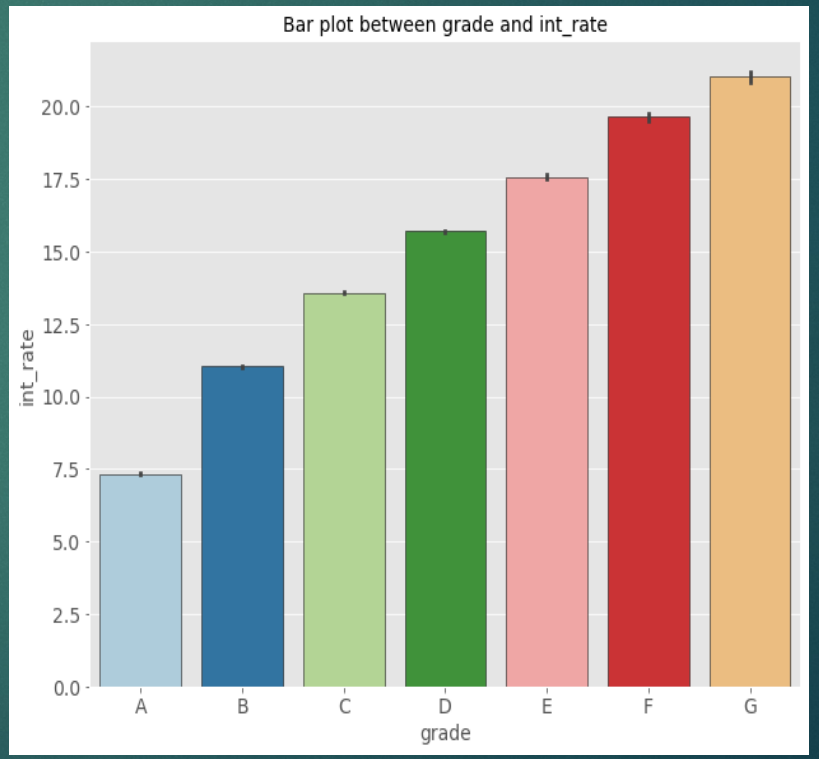


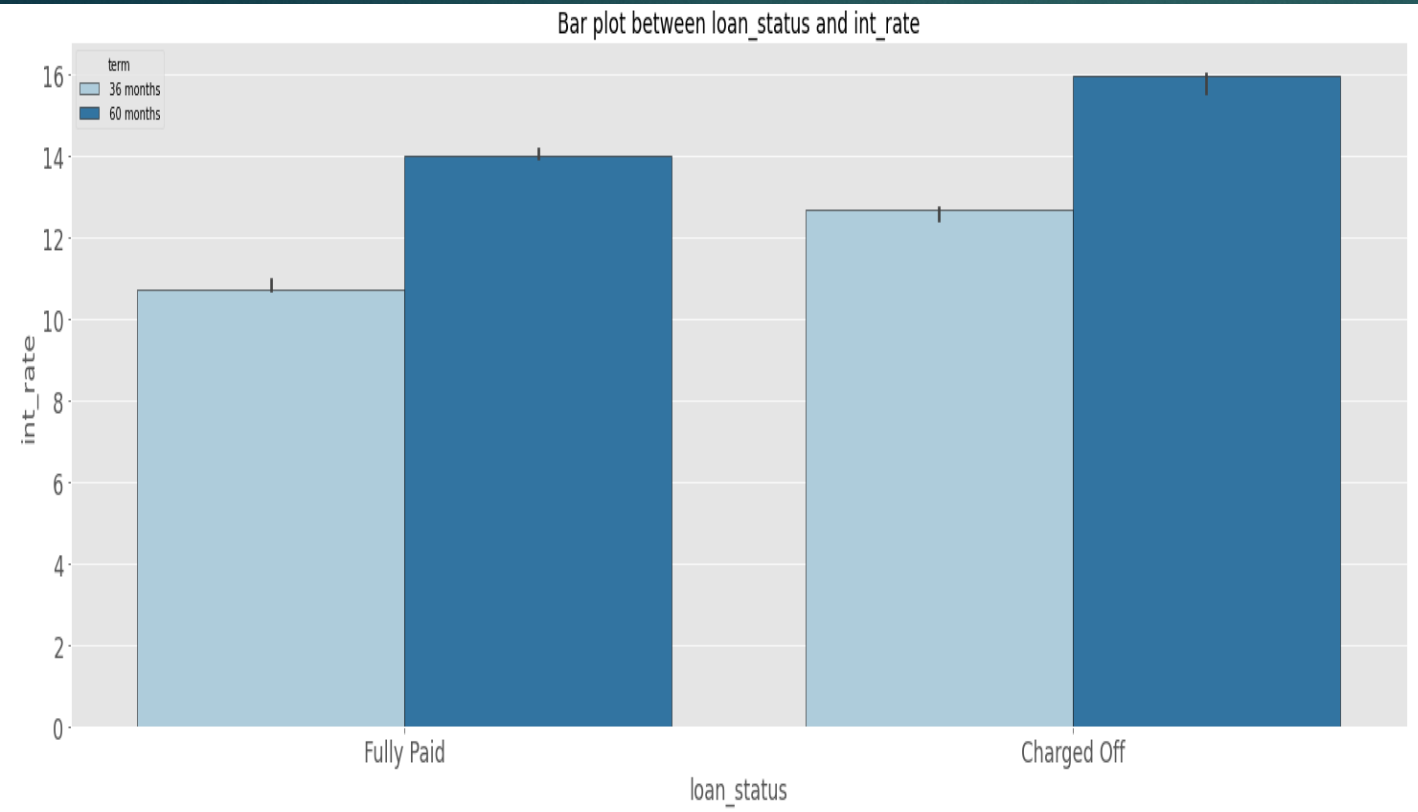
Before we begin let's take a look at the plots above, all of them are different plots but we can see that the distribution and the plot itself looks almost identical and thus we can make out that almost all of the requested loan amount was approved by Lending Club and the same amount was invested by investors

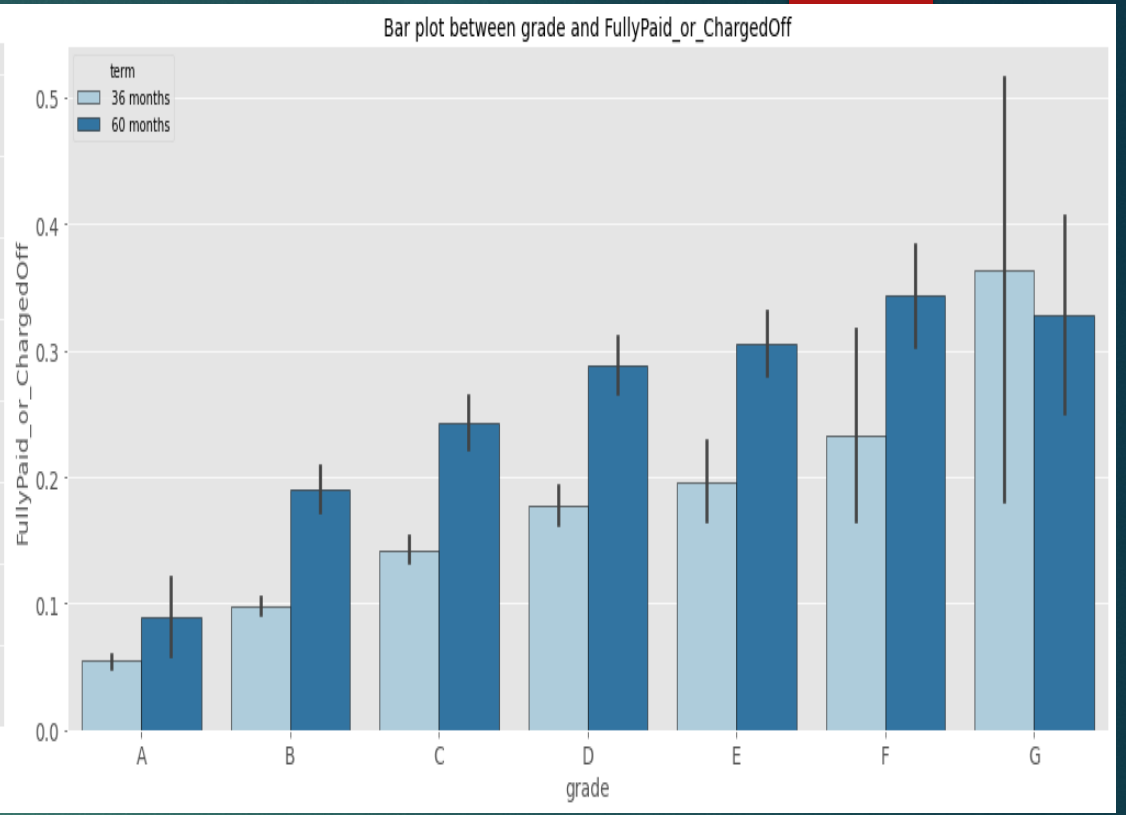
From the plots one can easily infer that higher the interest rates lower the grade one receives

Lower grades have a higher probability of defaulting since the interest rates for that grade can go beyond 20%, in the next slide we will see how it affects our borrowers

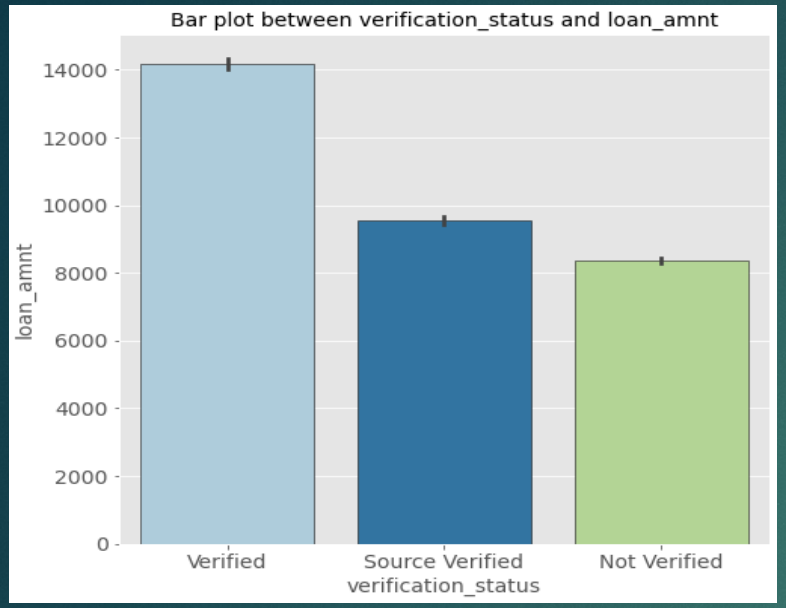


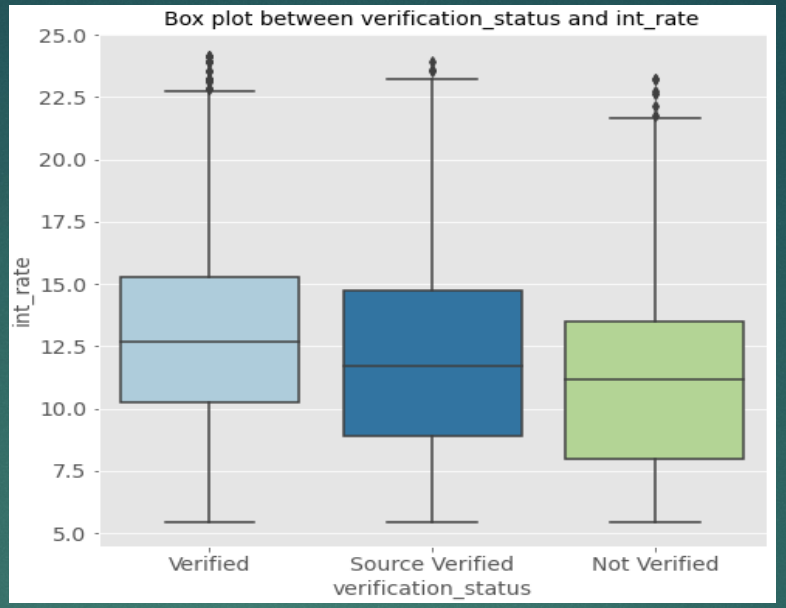


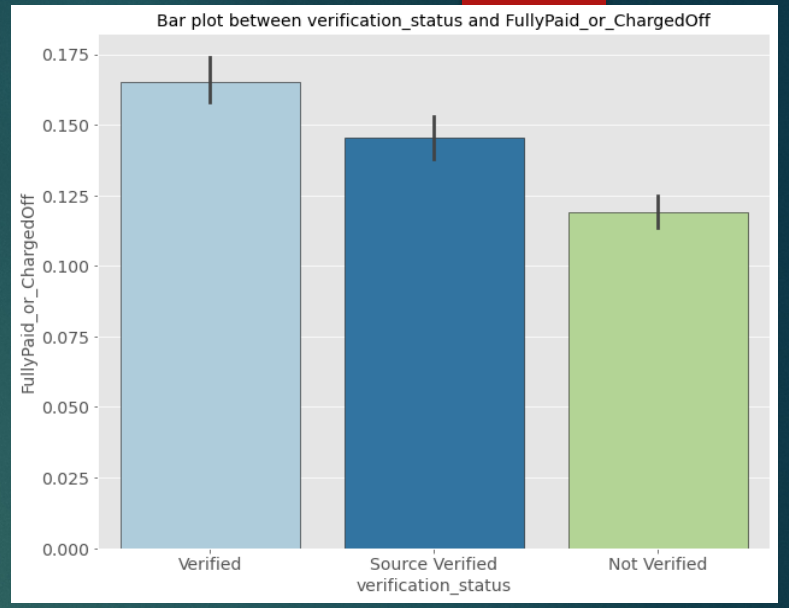




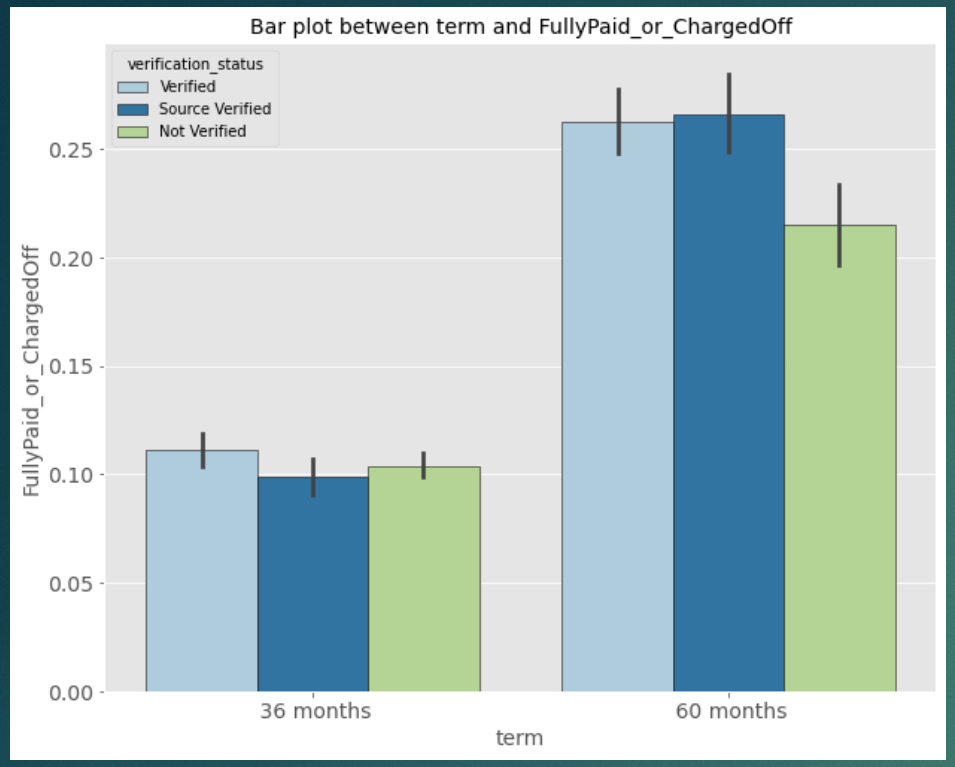
- We further added another component called “Term” and “Loan status” and “grade” to see how the interest rates differ
- We deduce that for borrower opting to repay the loan in 36 months on an average should be charged an interest rate of less than 12% and less than 14% for those who opt for 60 months to minimize the risk of borrowers defaulting
- We also observe that the majority of the borrowers default if they opt for 60 months.



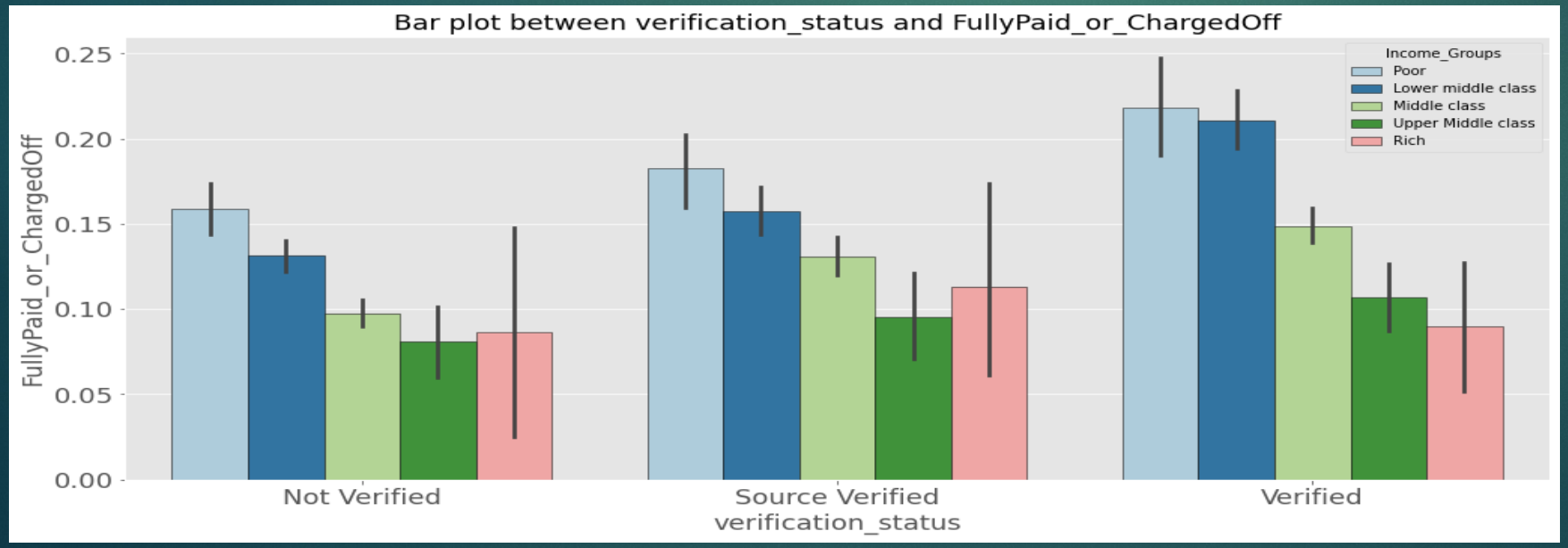


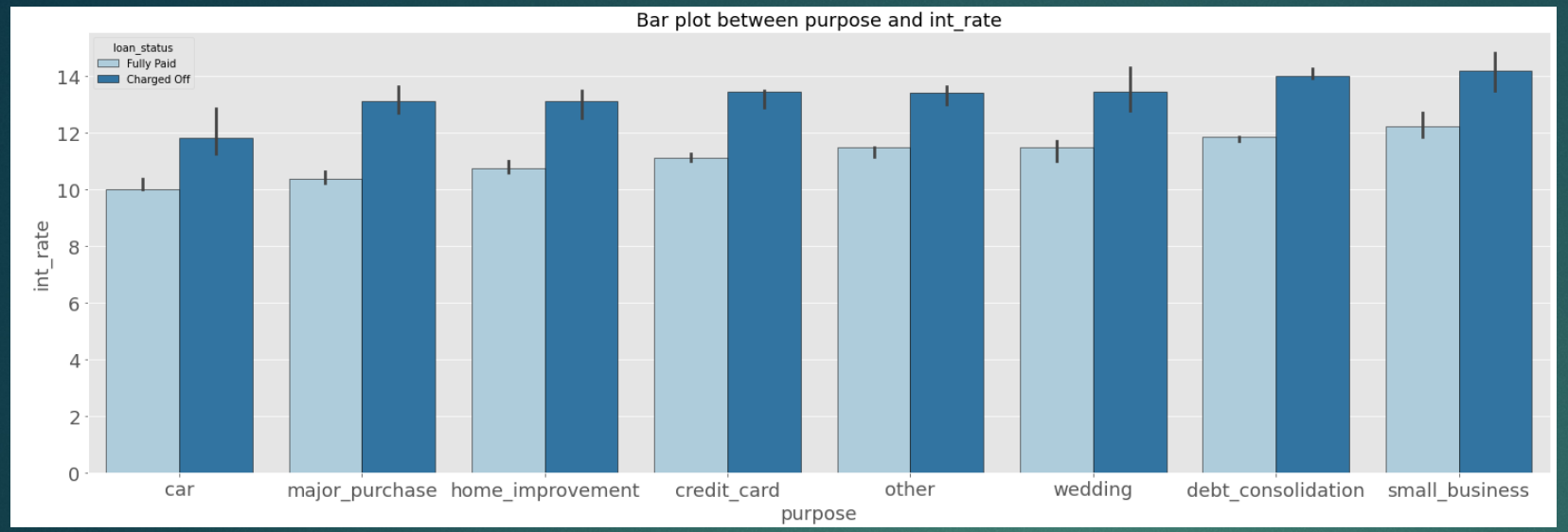


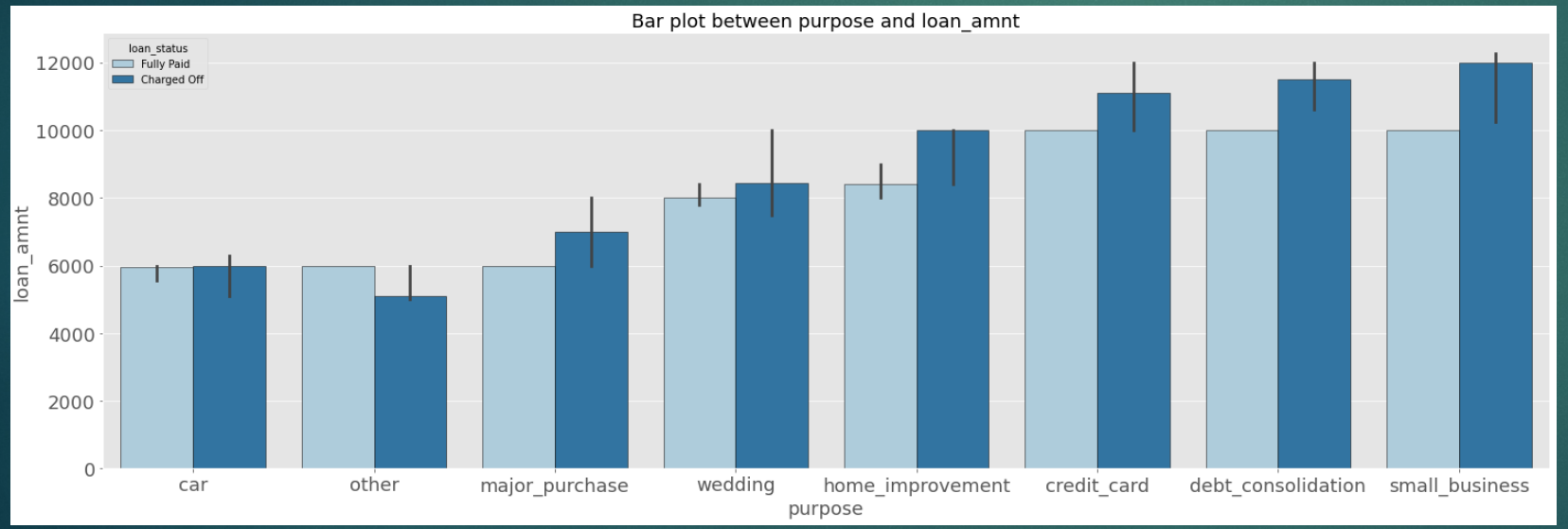
- The average loan amount requested was the highest for borrowers belonging to the category where their income source is “Verified”
- In the second plot we can see that they also have been charged with high interest rates; and in the final plot they risk defaulting more often than the other categories
- Lending club must carefully investigate this trend of lending high amounts of loan with asome of the high interest rates to borrower belonging to “Verified” category



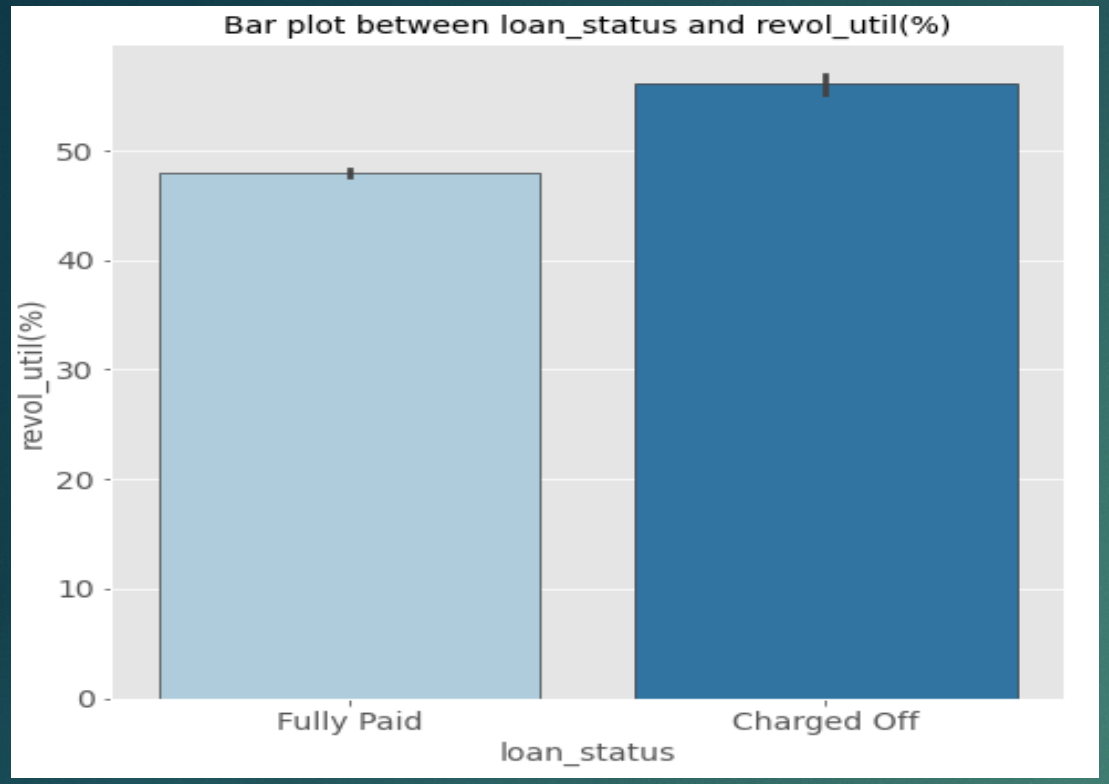
- Over here as well we see the same trend that borrowers opting or 60 months have a higher chance to default across different categories of Income Verification i.e verification_status in the plot
- The second plot depicts the power to repay the loan across various Income Groups the most prone to default are the borrowers belonging to Poor and Middle Class
- This says that the borrower having a higher income has a low risk of defaulting

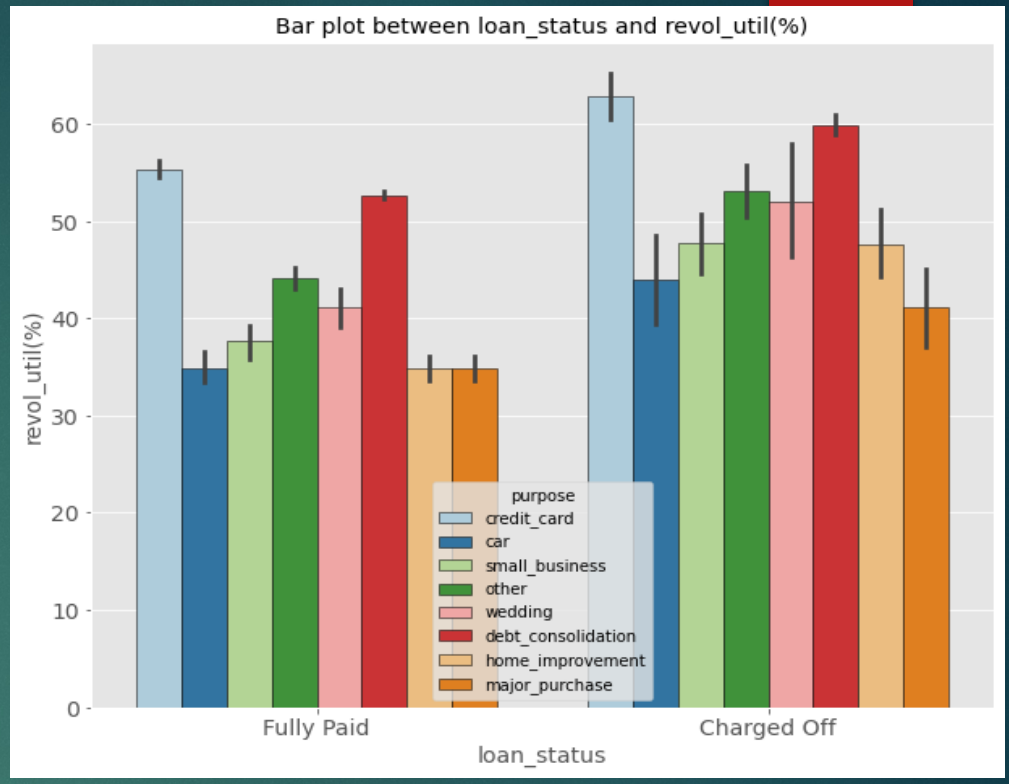




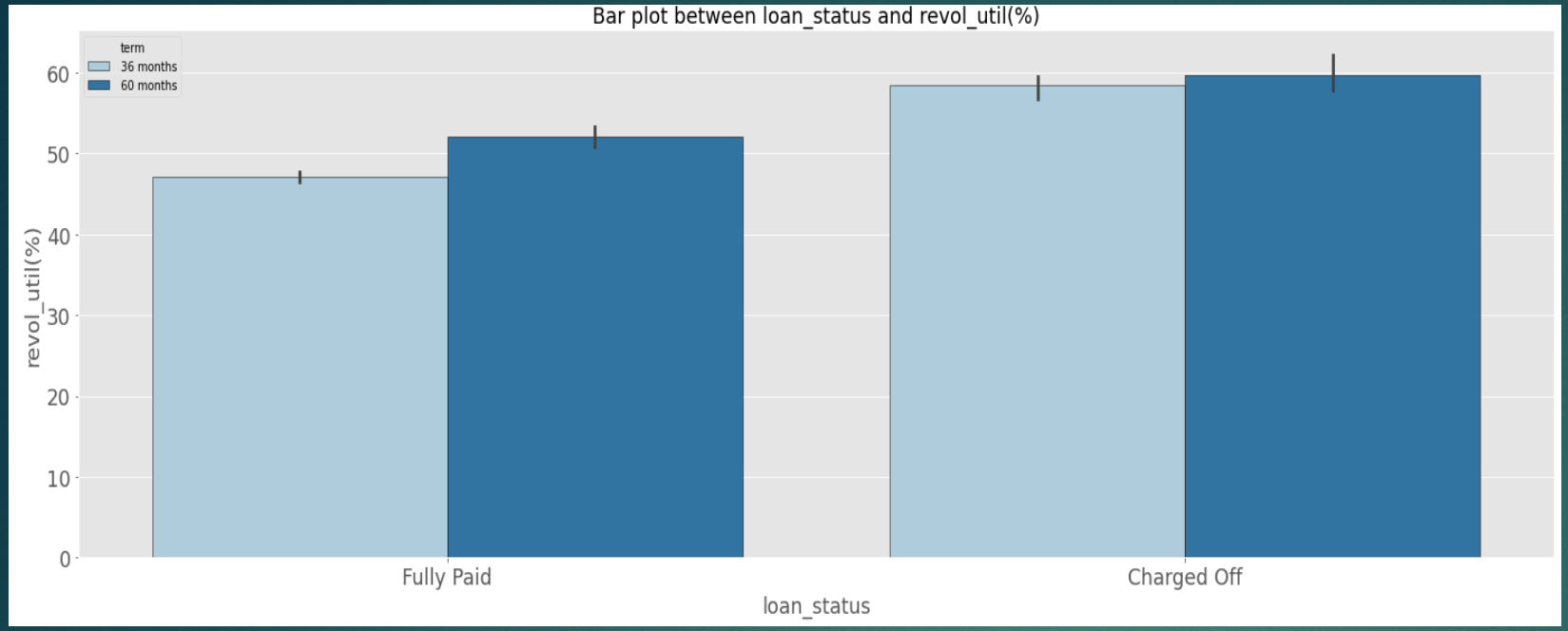


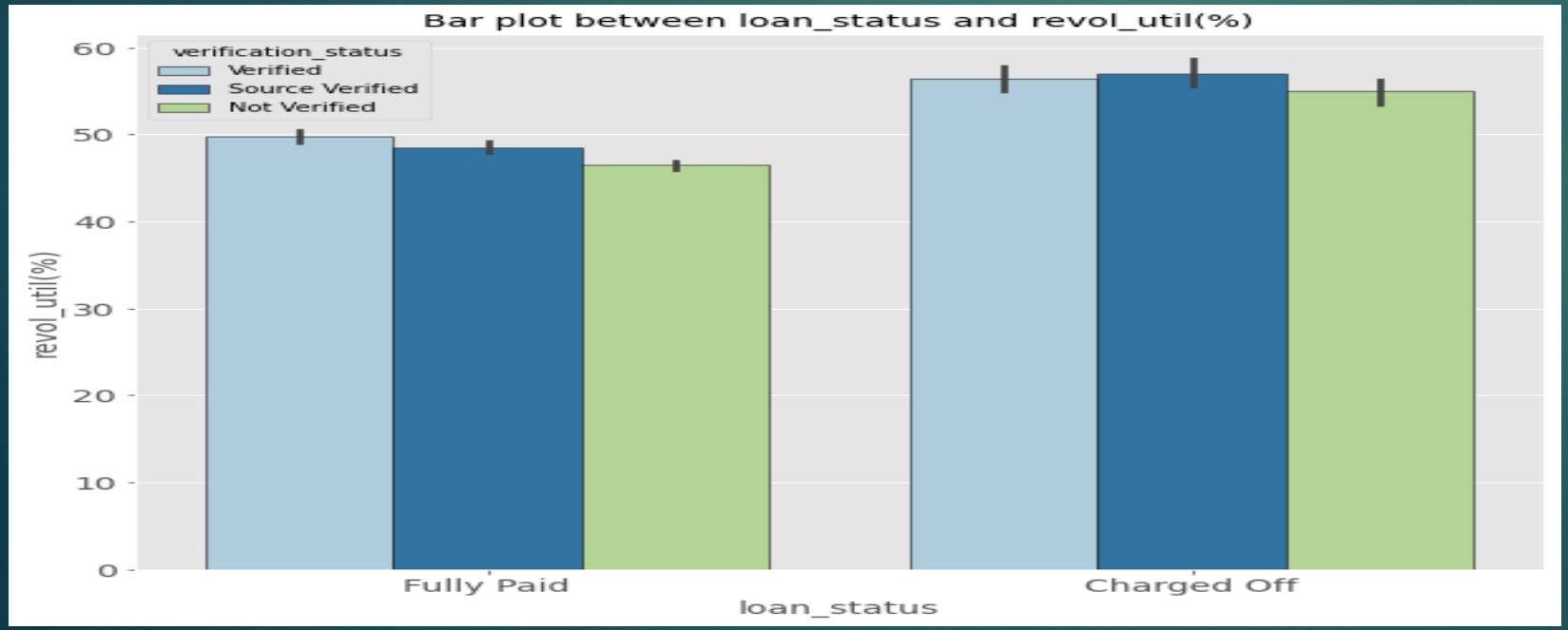
- On an average the interest rate for those who defaulted is 12% and up touching up to 14%
- The riskiest investments are for *Credit card, debt consolidation and small business*
- *Higher the loan amount and higher the interest rates maketh the perfect recipe for an increase in defaulter ratio*



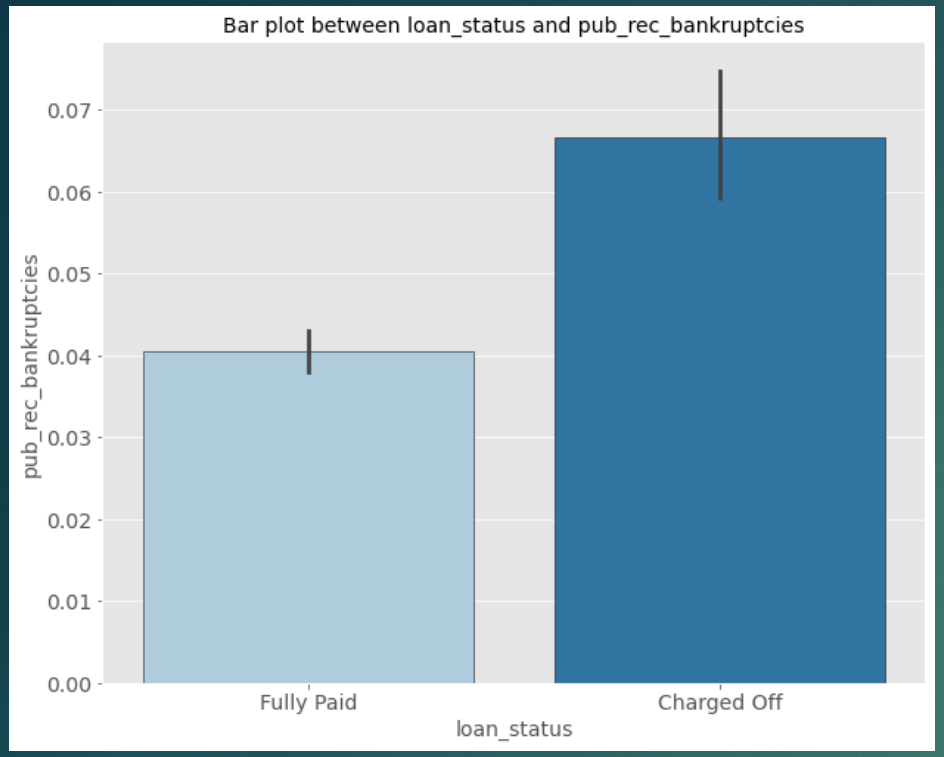


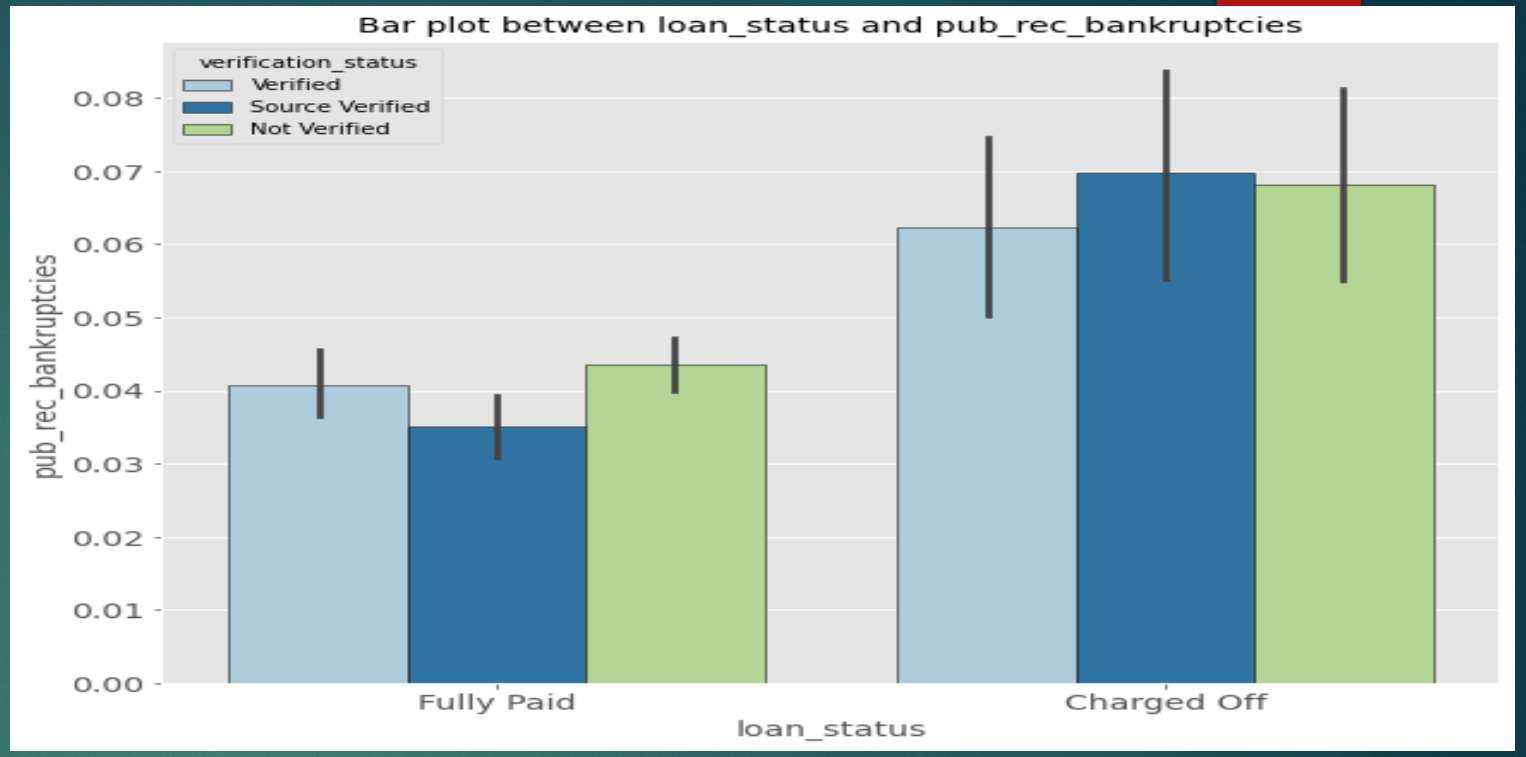
- Revolving Line utilization rate or simply put the credit utilization rate; lesser the rate higher the chances are that the borrower will not default
- It is independent of any categories and only conveys that if the average utilization rate is less or equal to 50 the borrower will most likely wont default, the next slide will complement our current analysis



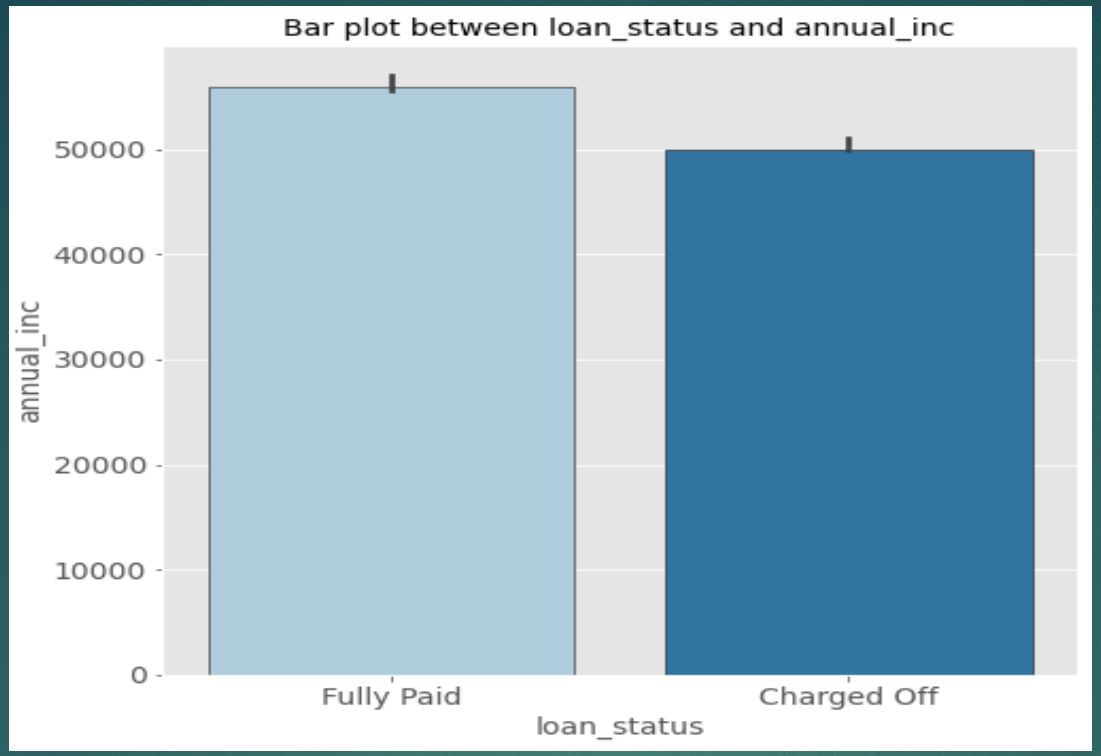


- Credit utilization really doesn't depend on any other factors even for income source verification status and term

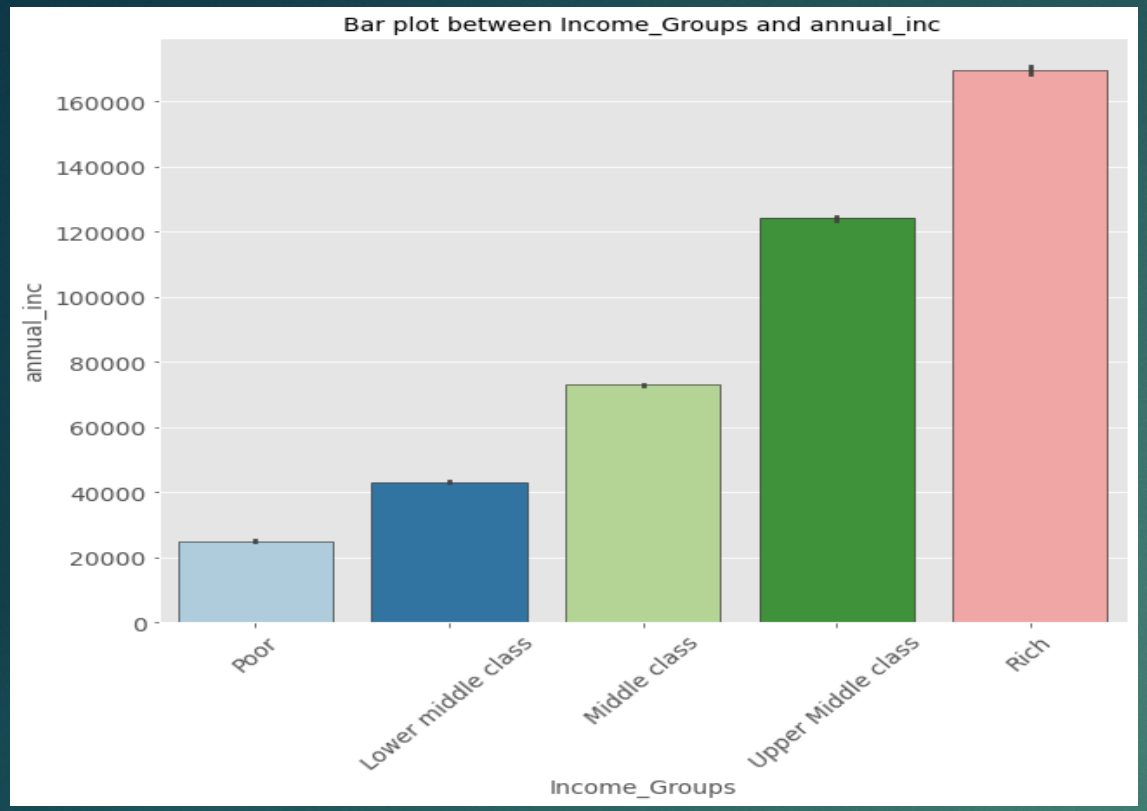


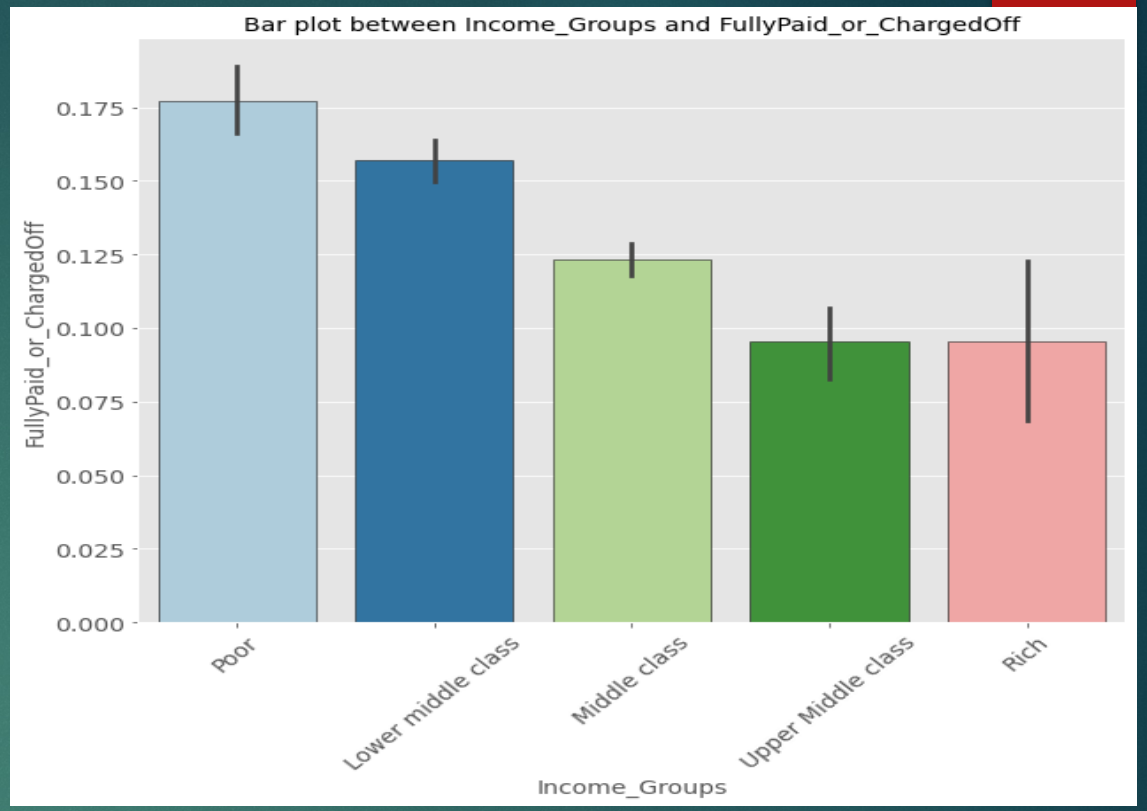


- Borrowers who have a history of being bankrupt end up defaulting more often than those who don't



Borrowers to earn more annually tend to default less which means that since their income is on the higher side they are able to repay in full and Lending club should focus on such borrowers keeping in mind the other factors we discovered in our previous slides





- These plots confirms the fact that people belonging to Middle class and up tend to default less and Lending club can issue or approve them loan based on factors we and insights we obtained from this entire analysis

CONCLUSION

- Lending club already does a great job with keeping the defaulter ratio under check
- We saw the interest rates with various other variables like purpose, verification status, and loan status, we can conclude that for every category the sweet spot is 12% in general and if we consider the variable term then on an average borrowers dont default if the interest rate doesnt exceed 12% for a period of 36 months and 14% for a period of 60 months across every category
- Also the revolving line of credit the ideal rate for which the borrower is not likely to default is less than 50%, beyond 50% the borrower is more likely to default
- LC can either decrease the requested loan amount or observe their annual income if it falls anywhere between middle to upper class then LC can think of approving their loan based on revolving line of credit, they should also take a quick loo at their interest rates as well cause it plays an important factor
- Borrowers who request loan for small business, debt consolidation or credit card must be given extra attention cause they have the highest risk of defaulting, LC must investigate and dig even deeper in these categories so they can understand the complexity of the situation and demand more information
- Borrowers who have been bankrupt in the past have a higher defaulter ratio and Lc must take into account their annual income, revolving line of credit, how much interest will they be paying and based on these factors LC can either approve the entire loan amount requested if the borrower has a higher probabilty of repaying and if not LC can always reduce the amount of loan approved while keeping the interest same therbey reducing the risk of the borrower defaulting
- LC must not be biased towards borrowers belonging to verified category of verification status, it seems pretty clear that higher amount of loan has been approved for such borrowers even though their credit utilization rate was well beyond 50%, with more than 12% interest charged and them opting to repay the loan in 60 months

- In order to ensure that LC is issuing loan to the right people care must be taken that:
- 1) On an average the interest rate shouldnt exceed more than 12% for those who have opted for 36 months and 14% for those who have opted for 60 months if at all a high interest rate is justified care must be taken that the borrower belongs to Upper Middle or atleast Middle
- 2) They must identitfy the same when borrowers request loan to repay credit card bills, consolidate debt or small businesses
- 3) The median loan amount on which the borrower wont default is around 9 grand and anything beyond that the borrower has a higher probabilty of defaulting, but there is more to it like how much is the borrower earning, how well is his utilization rate, the purpose of the loan, the term which the borrower prefers to opt affect
- 4) LC can delve deeper if the loan amount and the interest is on the higher end when a person applied for a loans