LENDING CLUB CASE STUDY

Submitted By:
Debi Mahapatra
Saswat Dash

Flow of Analysis

- Business Understanding & Problem Statement
- Data Preparation
- Univariate and Segmented Univariate Analysis
- Bivariate Analysis
- Conclusion/Observations
- Recommendations

Business Understanding & Problem Statement

Business Understanding

A Consumer Finance Company which specializes in lending various types of loans to various customers has to make a decision whenever it receives a loan request

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

1. **Loan accepted**
2. **Loan denied**

There are 3 possible scenarios when a loan gets approved:

1. Fully Paid
2. Charged Off
3. Current

Fully Paid: When the applicant fully pays the loan amount along with the interests

Charged Off: When the applicant has not paid the installments for a longer period

Current: When the applicant is still in the process of paying the installments

Problem Statement

Two **types of risks** are associated with the bank's decision:

1. If the applicant is **likely to repay the loan**, then not approving the loan results is a **loss of business** to the company
2. 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

The Consumer Finance Company has provided applicant's dataset from 2007 to 2011 for loans that were approved with details like whether the loan was charged off or fully paid.

The aim is to analysis all the variables in the given dataset and identify right patterns which indicate if a person is likely to fully pay the loan amount or will become a defaulter.

This in turn will help in taking the right actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Case Study Objective:

Identification of risky loan applicants using EDA is the aim of this case study

Data Preparation - I

Dropping Variables – I: (Data Based)

Drop the columns with:

with all empty columns

'mths_since_last_major_derog'	'mo_sin_rcnt_rev_tl_op'
'annual_inc_joint'	'mo_sin_rcnt_tl'
'dti_joint'	'mort_acc'
'verification_status_joint'	'mths_since_recent_bc'
'tot_coll_amt'	'mths_since_recent_bc_dlq'
'tot_cur_bal'	'mths_since_recent_inq'
'open_acc_6m'	'mths_since_recent_revol_delinq'
'open_il_6m'	'num_accts_ever_120_pd'
'open_il_12m'	'num_actv_bc_tl'
'open_il_24m'	'num_actv_rev_tl'
'mths_since_rcnt_il'	'num_bc_sats'
'total_bal_il'	'num_bc_tl'
'il_util'	'num_il_tl'
'open_rv_12m'	'num_op_rev_tl'
'open_rv_24m'	'num_rev_accts'
'max_bal_bc'	'num_rev_tl_bal_gt_0'
'all_util'	'num_sats'
'total_rev_hi_lim'	'num_tl_120dpd_2m'
'inq_fi'	'num_tl_30dpd'
'total_cu_tl'	'num_tl_90g_dpd_24m'
'inq_last_12m'	'num_tl_op_past_12m'
'acc_open_past_24mths'	'pct_tl_nvr_dlq'
'avg_cur_bal'	'percent_bc_gt_75'
'bc_open_to_buy'	'tot_hi_cred_lim'
'bc_util'	'total_bal_ex_mort'
'mo_sin_old_il_acct'	'total_bc_limit'
'mo_sin_old_rev_tl_op'	'total_il_high_credit_limit'

with only 1 distinct value

'pymnt_plan'
'initial_list_status'
'policy_code'
'chargeoff_within_12_mths'
'delinq_amnt'
'application_type'
'collections_12_mths_ex_med'
'acc_now_delinq'
'tax_liens'

Reasoning:

1. If a variable is completely empty, it is not required for analysis
2. If the variable has only 1 unique value, it is of no significance in our analysis

Dropping Variables – II: (Business Based)

revol_bal	**These variables are applicable only after the loan is approved.** **DROP THESE COLUMNS!!!**
revol_util	
out_prncp	
out_prncp_inv	
total_pymnt	
total_pymnt_inv	
total_rec_prncp	
total_rec_int	
total_rec_late_fee	
recoveries	
collection_recovery_fee	
last_pymnt_d	
last_pymnt_amnt	
last_credit_pull_d	

Reasoning:

1. Variables not available before the applicant's loan request is approved is of no use for the lender in it's decision making

DATA PREP CONTINUED IN NEXT SLIDE

Data Preparation - II

Missing Values

Columns & null counts

mths_since_last_record	36931
mths_since_last_delinq	25682
desc	12940
emp_title	2459
emp_length	1075
pub_rec_bankruptcies	697
title	11

Drop these columns. because of high percentage of null value

Drop it, we will be using purpose instead of title

Drop it, as emp title does not speak about profession type and has too many distinct values

Keep it as it is, these variables will be used for segmented univariate analysis in future

Remaining Columns For Analysis

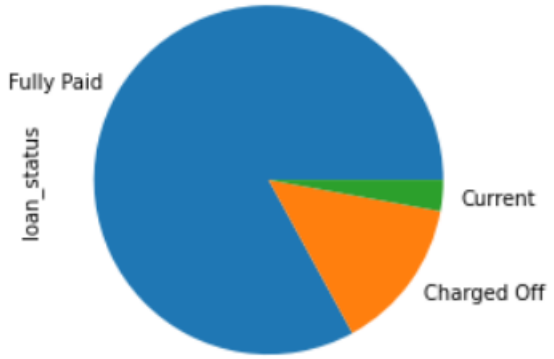
Column Name	Description
id	A unique LC assigned ID for the loan listing.
member_id	A unique LC assigned Id for the borrower member.
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
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.
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
int_rate	Interest Rate on the loan
installment	The monthly payment owed by the borrower if the loan originates.
grade	LC assigned loan grade
sub_grade	LC assigned loan subgrade
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
home_ownership	The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
annual_inc	The self-reported annual income provided by the borrower during registration.
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified
issue_d	The month which the loan was funded
loan_status	Current status of the loan
purpose	A category provided by the borrower for the loan request.
zip_code	The first 3 numbers of the zip code provided by the borrower in the loan application.
addr_state	The state provided by the borrower in the loan application
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
pub_rec_bankruptcies	Number of public record bankruptcies
open_acc	The number of open credit lines in the borrower's credit file.
pub_rec	Number of derogatory public records
total_acc	The total number of credit lines currently in the borrower's credit file
delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years

Note: From 111 columns we have reduced the number of columns to 26

DATA PREP CONTINUED IN NEXT SLIDE

Data Preparation - III

Remove Rows - I



Univariate analysis of Loan status

Observation :

The above pie chart shows there are 3 distinct values for loan status:

1. Charged Off
2. Fully Paid
3. Current

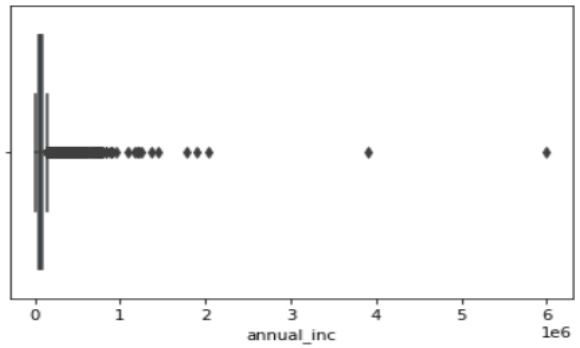
Loan Status Current does not have any significance in our analysis, Only dataset related to Fully Paid and Charged Off will help in our analysis

Action :

Drop the rows with loan status = "Current"

Outlier Detection - I

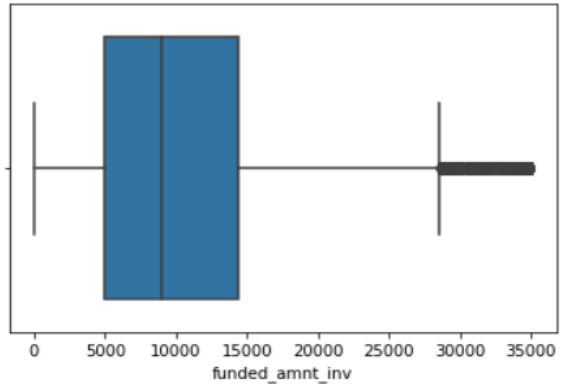
annual_inc



count	3.971700e+04
mean	6.896893e+04
std	6.379377e+04
min	4.000000e+03
25%	4.040400e+04
50%	5.900000e+04
75%	8.230000e+04
max	6.000000e+06

anything above 85 percentile is considered outlier (100K)

funded_amount_inv



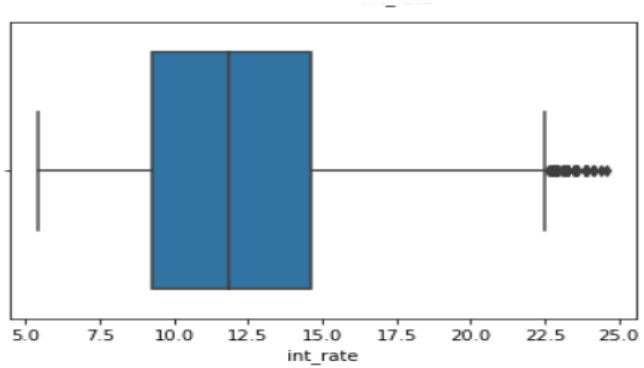
count	39717.000000
mean	10397.448868
std	7128.450439
min	0.000000
25%	5000.000000
50%	8975.000000
75%	14400.000000
max	35000.000000

anything above 95 percentile is considered outlier (25K)

Data Preparation - IV

Outlier Detection - II

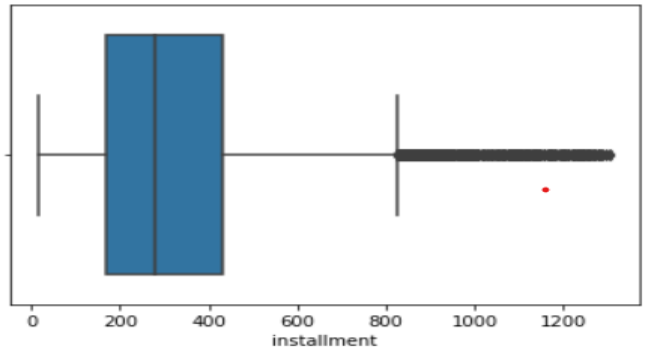
int_rate



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

anything above 95 percentile is considered outlier (20% approx.)

installment



count	39717.000000
mean	324.561922
std	208.874874
min	15.690000
25%	167.020000
50%	280.220000
75%	430.780000
max	1305.190000

anything above 95 percentile is considered outlier (800 approx.)

Data Cleaning - I

1. Check if the ints and floats are ints and floats respectively and not objects
2. Check for int columns with symbols like %, remove them special chars from percentage fields and make it a float
3. Check for date variables and convert to datetime

Before

	int_rate
0	10.65%
1	15.27%
2	15.96%
3	13.49%
4	12.69%

After

	int_rate
0	10.65
1	15.27
2	15.96
3	13.49
4	12.69

	issue_d
0	Dec-11
1	Dec-11
2	Dec-11
3	Dec-11

issue_d_month	issue_d_year
Dec	11
Dec	11
Dec	11
Dec	11
Dec	11

Before

10+ years	8879
< 1 year	4583
2 years	4388
3 years	4095
4 years	3436
5 years	3282
1 year	3240
6 years	2229
7 years	1773
8 years	1479
9 years	1258

After

10	8879
0	4583
2	4388
3	4095
4	3436
5	3282
1	3240
6	2229
7	1773
8	1479
9	1258
-1	1075

Removed <, +, years from the **emp_length** column

Data Preparation - IV

Derived Variable – (Date Based)

	issue_d	issue_d_month	issue_d_year	earliest_cr_line	earliest_cr_line_month	earliest_cr_line_year
0	Dec-11	Dec	11	Jan-85	Jan	85
1	Dec-11	Dec	11	Apr-99	Apr	99
2	Dec-11	Dec	11	Nov-01	Nov	01
3	Dec-11	Dec	11	Feb-96	Feb	96
4	Dec-11	Dec	11	Jan-96	Jan	96

Converting issue_d into separate columns like months and years

Converting earliest credit line into separate columns like months and years

Derived Variable – (Range Based)

	interest_category	int_rate
0	ir4[10.0-15.0]	10.65
1	ir5[15.0-20.0]	15.27
2	ir5[15.0-20.0]	15.96
3	ir4[10.0-15.0]	13.49
4	ir4[10.0-15.0]	12.69

	funded_amnt_inv_category	funded_amnt_inv
0	fa2[0.0-10.0]K	4975.0
1	fa2[0.0-10.0]K	2500.0
2	fa2[0.0-10.0]K	2400.0
3	fa3[10.0-20.0]K	10000.0
5	fa2[0.0-10.0]K	5000.0

	annual_inc_category	annual_inc
0	Inc3[20.0-40.0]K	24000.0
1	Inc3[20.0-40.0]K	30000.0
2	Inc2[0.0-20.0]K	12252.0
3	Inc4[40.0-60.0]K	49200.0
4	Inc6[80.0-100.0]K	80000.0

	monthly_installment_category	installment
0	Inst3[100.0-200.0]	162.87
1	Inst2[0.0-100.0]	59.83
2	Inst2[0.0-100.0]	84.33
3	Inst5[300.0-400.0]	339.31
4	Inst2[0.0-100.0]	67.79

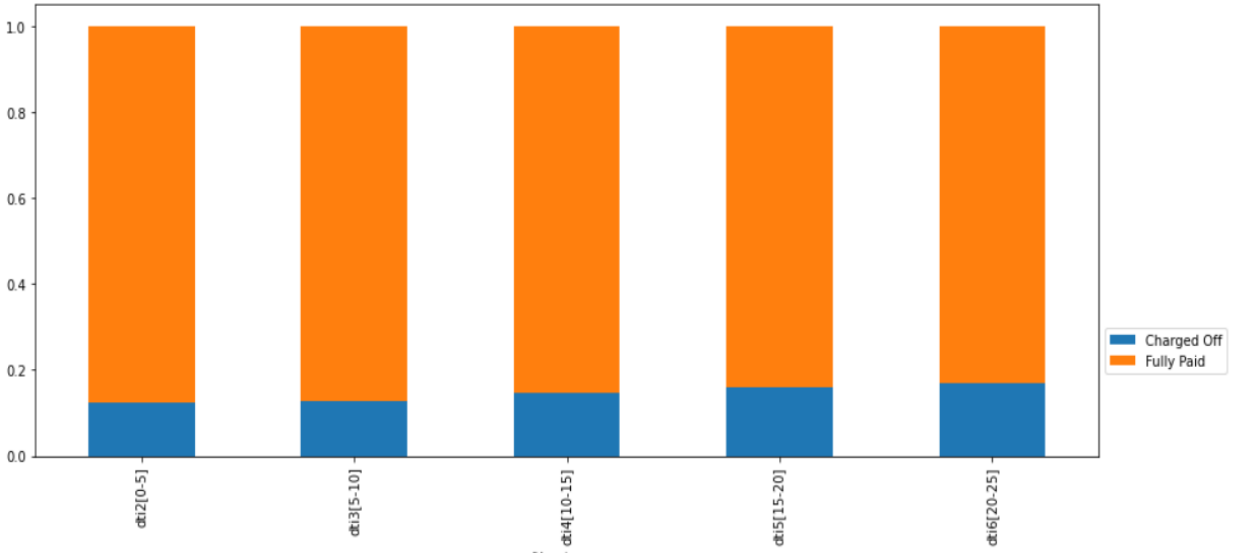
	dti_category	dti
0	dti7[25-30]	27.65
1	dti2[0-5]	1.00
2	dti3[5-10]	8.72
3	dti6[20-25]	20.00
4	dti5[15-20]	17.94

	open_acc_total_per_range	open_acc_total_per
0	ot5[30-40]	33.333333
1	ot9[70-80]	75.000000
2	ot4[20-30]	20.000000
3	ot4[20-30]	27.027027
4	ot5[30-40]	39.473684

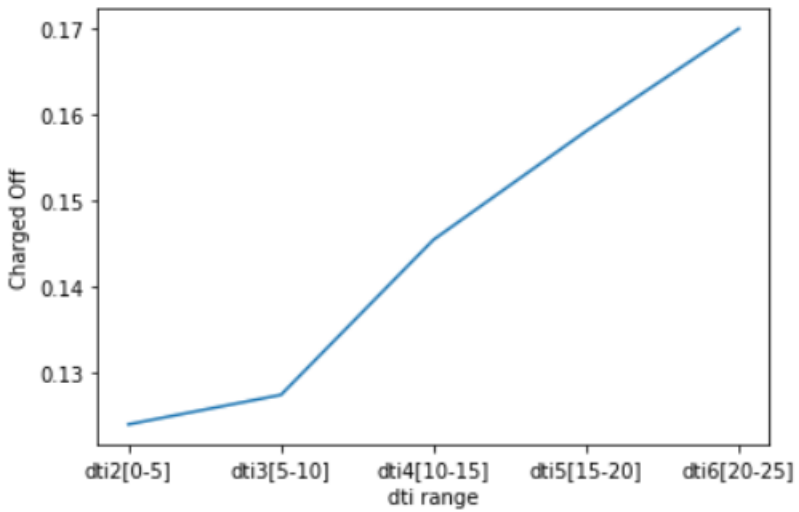
UNIVARIATE AND SEGMENTED UNIVARIATE ANALYSIS

DTI - Debt To Income ratio

<<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count and fully paid member counts for each dti bucket. This does not show the spike clearly, so refer the line chart for clarity



The above line chart depicts the rise in ratio (percentage) of charged off member count when it moves from a lower bucket to a higher bucket

What is DTI?

- The debt-to-income (DTI) ratio is the percentage of your gross monthly income that goes to paying your monthly debt payments and is used by lenders to determine your borrowing risk.

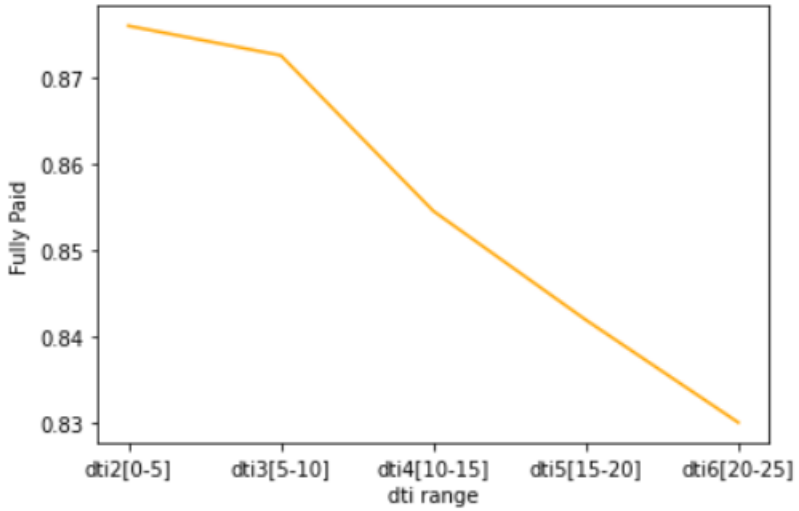
Conclusion:

- Lower the DTI value, lower the chance of defaulting, Higher the value of DTI higher the chances of defaulting

Recommendations:

- If the DTI value is high, lenders can think of either not approving the loan or think of reducing the loan amount, if other variables are not putting red flags
- If the DTI ratio is high, lenders can look at other variables more carefully like cost of living in a particular place, annual income trends over the years (up or down?)
- If the DTI ratio is low, lenders can approve the loan if there are no red flags from other driver variables

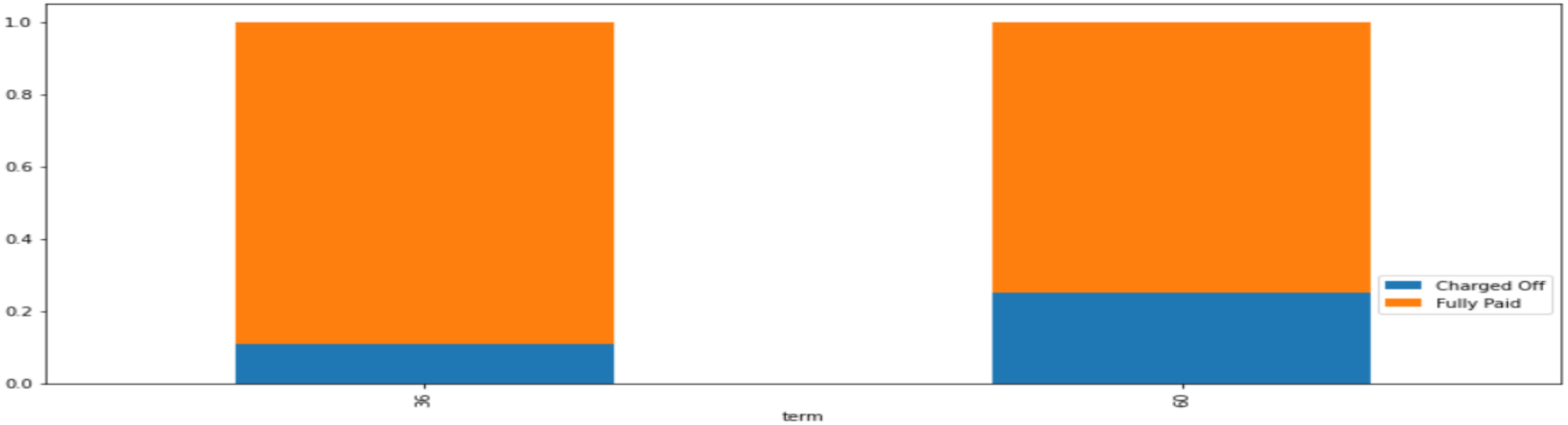
NOTE: DTI values has been put into multiple buckets (increment of 5) like [0-5],[5-10],[10-15] and so on.



The above line chart depicts the fall in ratio (percentage) of fully paid member count when it moves from a lower bucket to a higher bucket

Loan Terms

<<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count to fully paid member counts for each available loan term.

Loan Terms:

- The number of payments on the loan. Values are in months and can be either 36 or 60.

Recommendations:

- Incase there are other variables showing red flags, then the lender can go for a shorter term considering factors like the current annual income is high or if one foresee a recession or rise in real estate price in the current locality

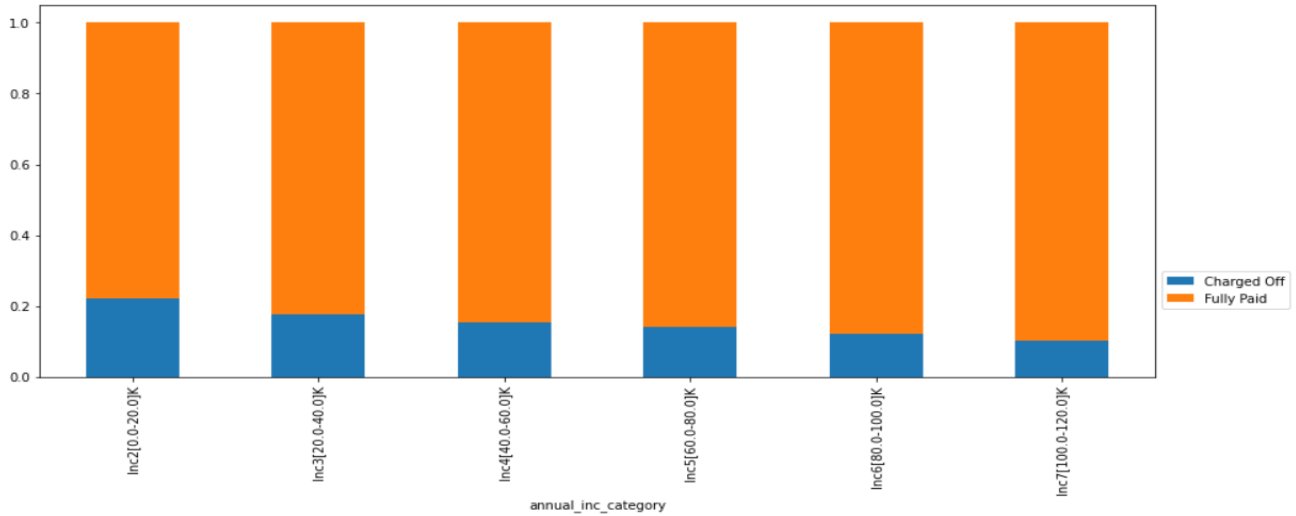
Conclusion:

- Lower the loan terms lower the chances of loan amount being charged off
- Higher the loan terms higher the chances of loan amount being charged off

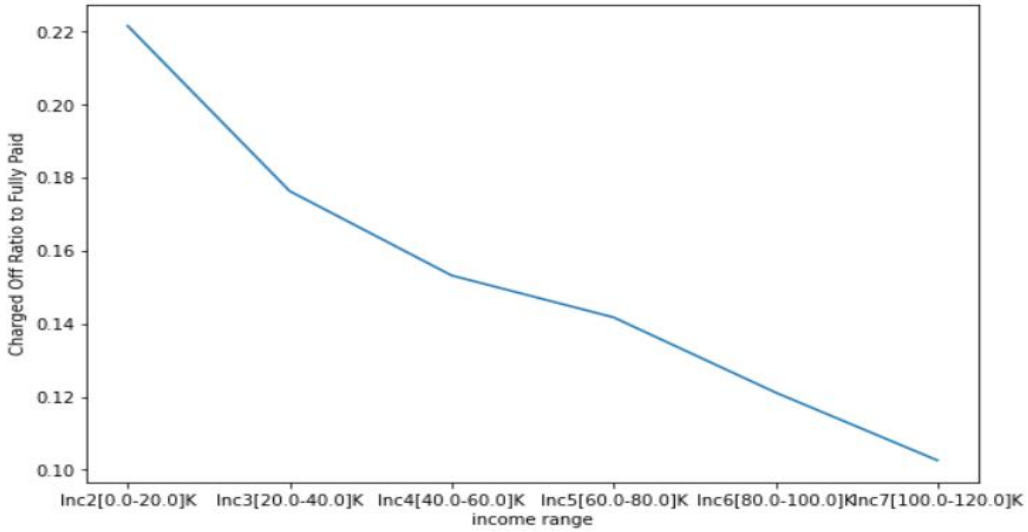
5 years is a bigger term in comparison to 3 years, lot of factors change in these extra 2 years. ***Recession might come in, loss of job, annual income might come down, relocating to a higher cost of living city, increase in home rents, inflation, change in government, new regulations*** are all possible reasons for increase in number of loan defaulter with higher loan terms

Annual Income

<<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count and fully paid member counts for each annual income bucket. This does not show the spike clearly, so refer the line chart for clarity



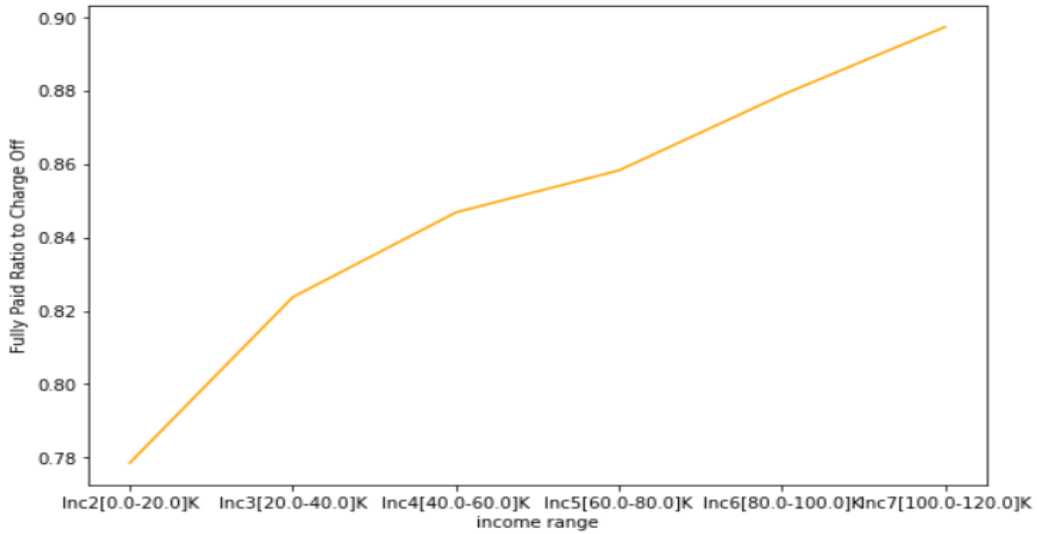
The above line chart depicts the fall in ratio (percentage) of charged off member count when it move from a lower income bucket to a higher income bucket

Annual Income:

- The self-reported annual income provided by the borrower during registration.

Conclusion:

- Lower the Annual Income range, higher the chance of defaulting, Higher the Annual Income range, lower the chances of defaulting
- Negative correlation detected



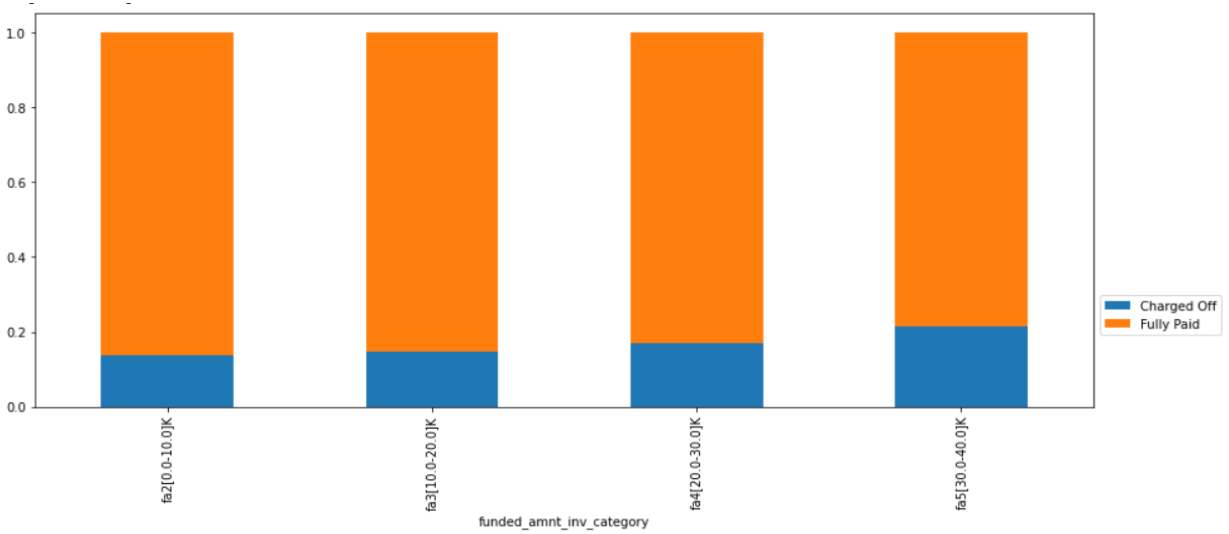
The above line chart depicts the rise in ratio (percentage) of fully paid member count when it moves from a lower income bucket to a higher income bucket

Recommendations:

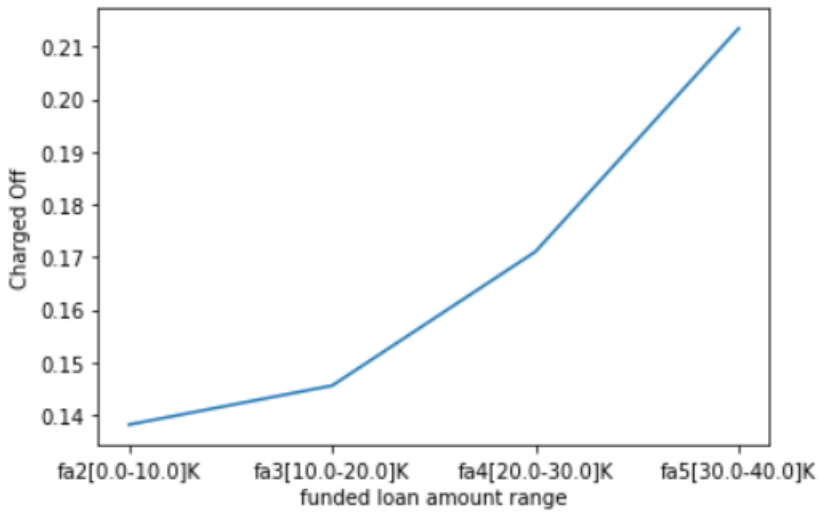
- Annual Income high definitely is a major factor for loan approval, but along with the **Annual Income**, do consider the **DTI and loan mount to Annual Income Ratio** as well

NOTE: Annual Income values have been put into multiple buckets (increment of 20K like [0-20]K,[20-40]K,[40-60]K and so on.

Loan Amount (Funded Amount Investor) <<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count and fully paid member counts for each loan amount bucket. This does not show the spike clearly, so refer the line chart for clarity



The above line chart depicts the fall in ratio (percentage) of charged off member count when it moves from a lower loan amount bucket to a higher loan amount bucket

Funded Amount Inv:

- The final loan amount that gets approved.

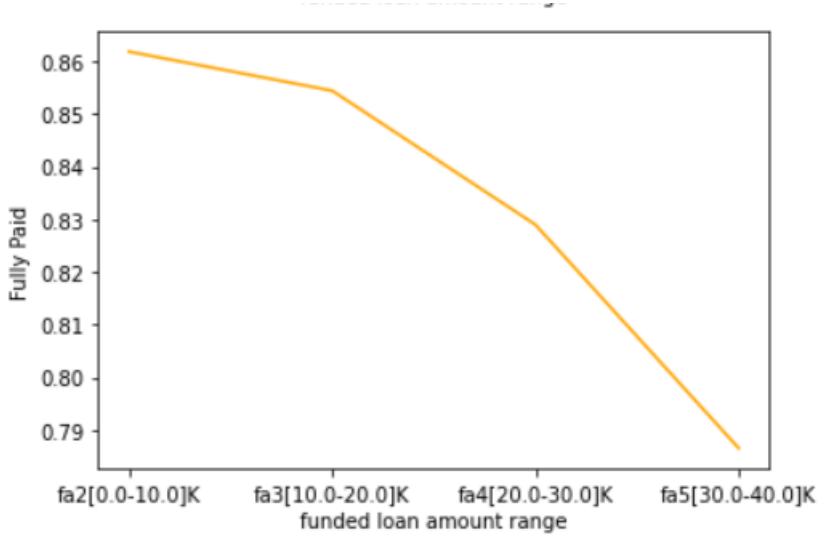
Conclusion:

- Higher the fund amount higher is the risk of defaulting

Recommendations:

- If there more red flags along with a high loan demand, then reject the loan amount or reduce the loan amount

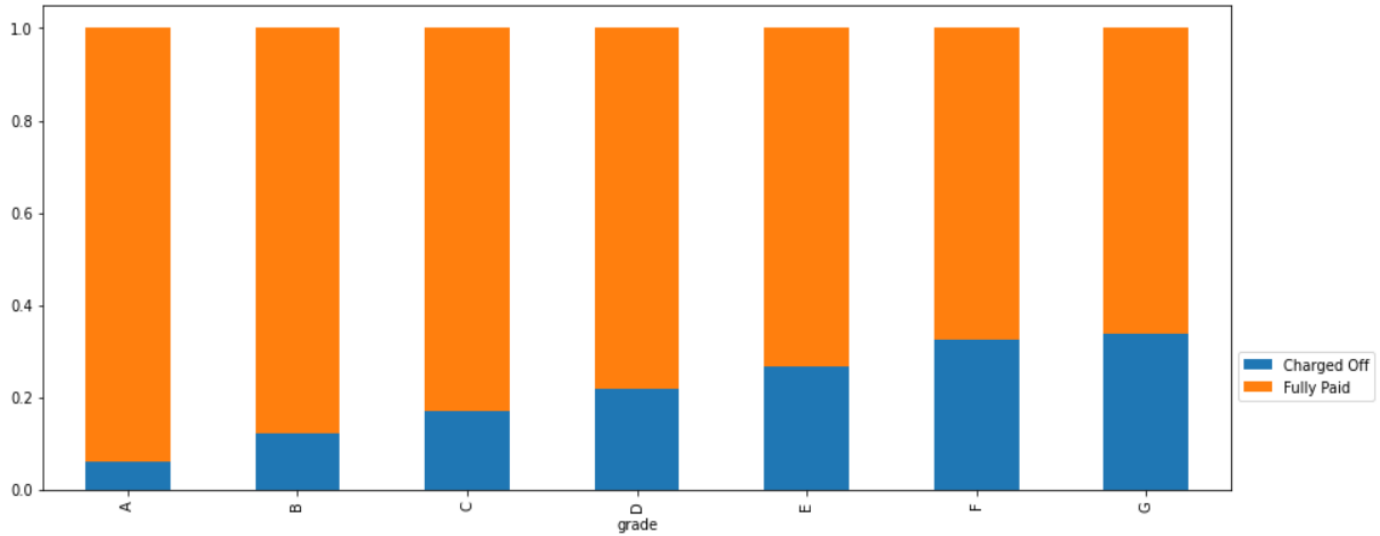
NOTE: loan amount values have been put into multiple buckets (increment of 10K like [0-10]K,[10-20]K,[20-30]K and so on.



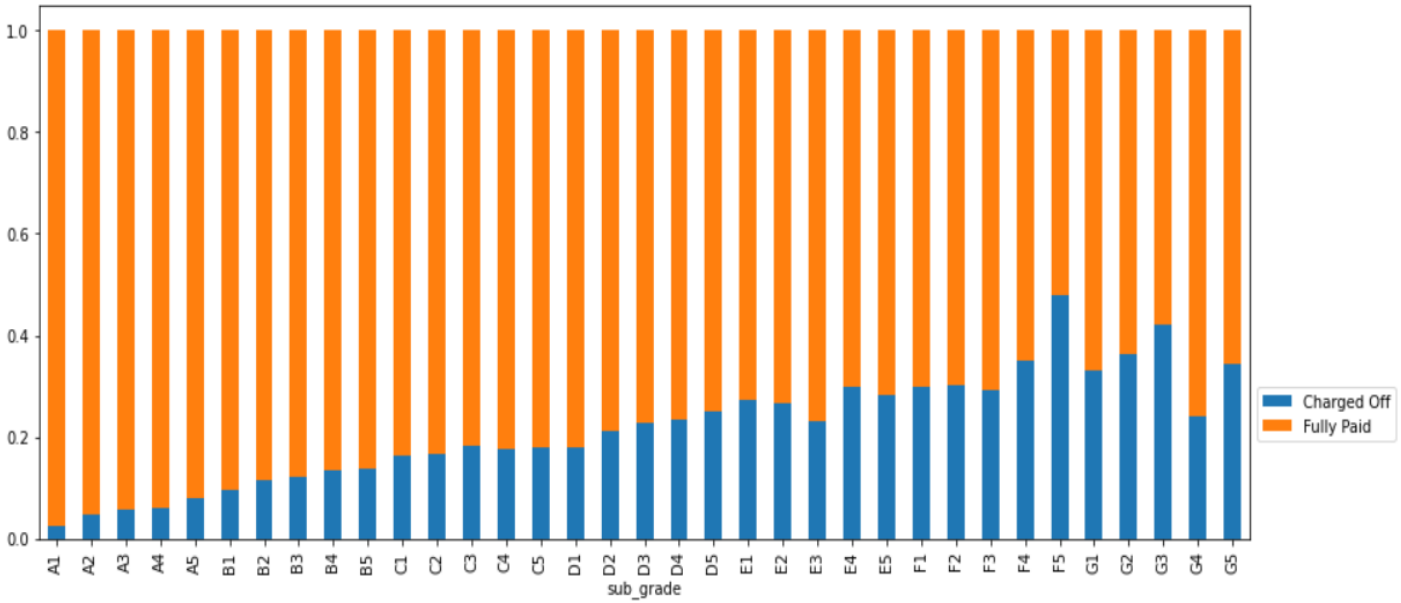
The above line chart depicts the rise in ratio (percentage) of fully paid member count when it moves from a lower loan amount bucket to a higher loan amount bucket

Grade and Sub-Grades

<<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count and fully paid member counts for each GRADE. The rise in chance od defaulting a loan is clearly visible with increase in grade



The stacked chart depicts the ratio of charged off member count and fully paid member counts for each SUB-GRADE.

Grade & Sub-Grade:

- LC assigned loan grade and sub-grade

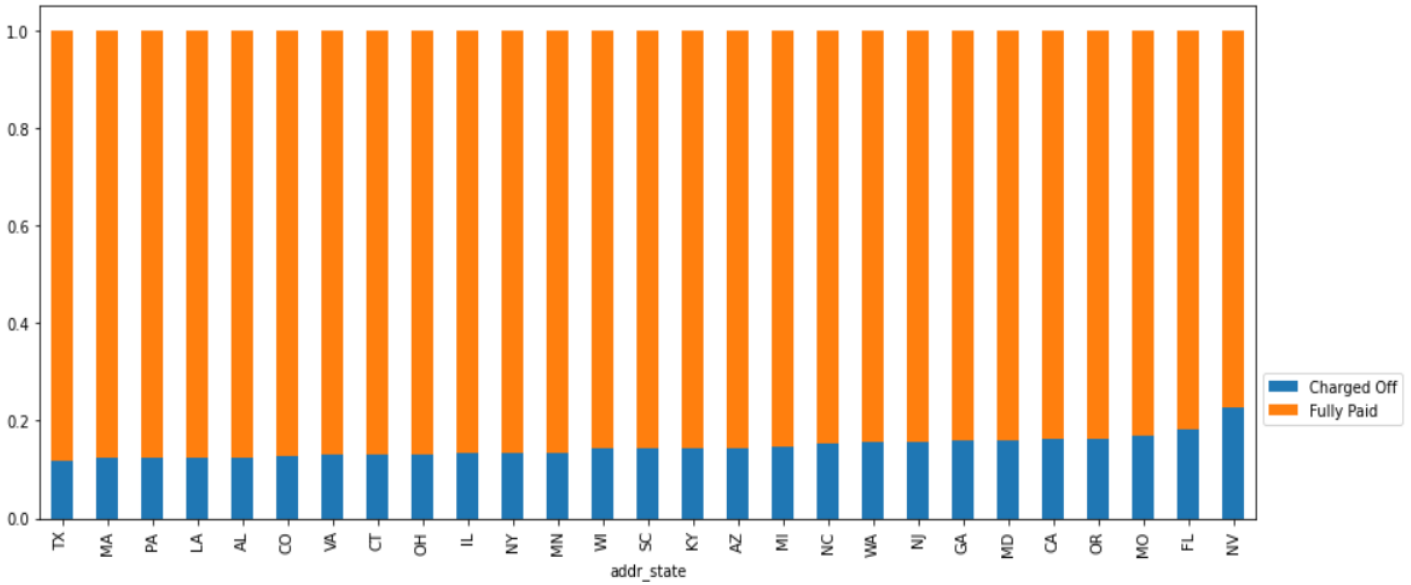
Conclusion:

- As we move from lower grade A to higher grade G, there is a good spike in the ratio of number of charged off members to number fully paid members
- Similarly, as we move from lower grade A to higher grade G, there is a good decline in the ratio of number fully paid members to number of charged off members
- So, chances of defaulting increases with higher grades and lower grades ensures higher possibilities of loan payment
- Similar trend is seen among sub grades within grades. Within a Grade, even if we move from 1 to 5 sub grade, the possibility of defaulting a loan is higher (**in maximum case**)

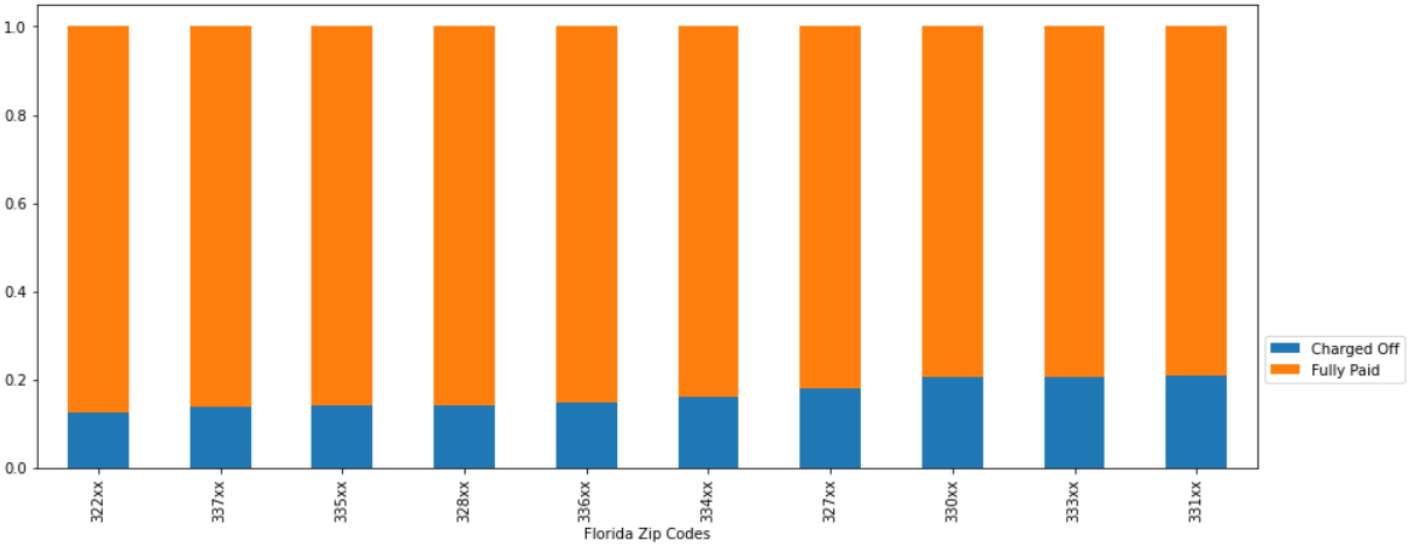
NOTE: In Chart 2…There are few exceptions where lower sub grades have higher ratio of charge off member counts specially in grade G. But Grade G has very few data size. That could be a possible reason

Address State, Zip Code and Cost of Living

<<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count and fully paid member counts for each STATE.



The stacked chart depicts the ratio of charged off member count and fully paid member counts for each Zip Code withing Florida.

Address State:

- State Code for the US states

Zip Code:

- First three characters of a zip code, can be useful in identifying a county or a city, if not the exact place

Conclusion:

- The first chart " address state vs Charged off/Fully Paid ratio chart" has **no trends** as such

BUT,

If we further drill down to Zip Codes, we can see the below trend
From the first chart let us consider Florida which is at higher risk of getting charged off
Then let us plot a ratio chart for all the zip codes in Florida.
Below is the observation:
322XX is Jacksonville FL, which is at a lower risk of defaulting
where as
331XX is Miami FL, which is at higher risk of defaulting
The cost of living in Miami, FL is 20.5% higher than in Jacksonville, FL
Employers in Miami, FL typically pay 2.7% more than employers in Jacksonville, FL for the same type of job

HIGHER COST OF LIVING PLACES ARE MORE LIKELY TO BE DEFAULTERS

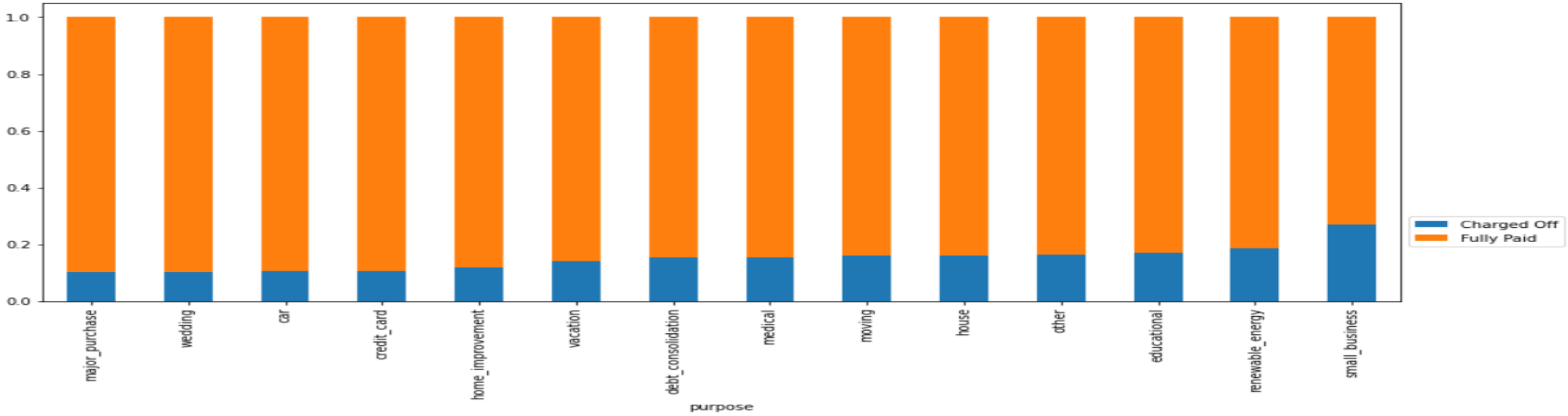
Recommendation:

Add a new dataset which has city and it's cost of living or a rank in terms of cost of living or some other metrics like average home rent in that county or city, which can give us some scope for analysis. Then map it against annual income.

NOTE: For chart1, we have excluded states with member counts < 300 and for chart 2 we have excluded zip codes with member count < 100

Purpose

<<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count to fully paid member counts for each available loan term.

Purpose of loan:

- A category provided by the borrower for the loan request.

Recommendations:

- In case the loans are for business purpose, have options of recovering back from assets like home owned by the loan requestor. Have a proper background check to find out how successful the requestor has been in previous businesses. Also deep digger into the business to find out the probability of the business getting successful. Figure out who are the competitors of that business. Figure out how niche is that business.

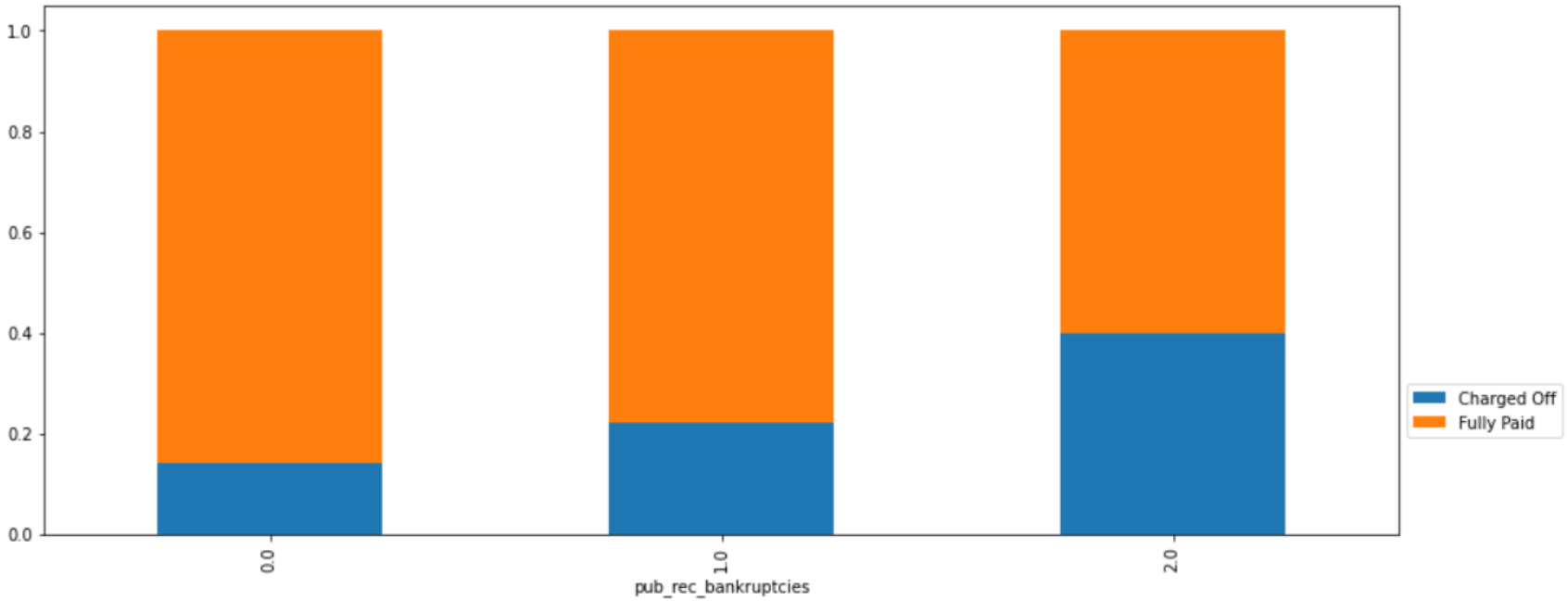
Conclusion:

Smaller business seems to be having a higher risk of getting charged off in comparison to loans taken for personal purposes like wedding and car

Business related loans are always at higher risks as loan payments are often related to the profit made

PUB REC BANKRUPTCIES

<<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count to fully paid member counts for each available loan term.

pub_rec_bankruptcies:

- Number of public record bankruptcies.
- Filing for bankruptcy is a **legal process that either reduces, restructures or eliminates your debts.**

Recommendations:

- For people who have a public record of bankruptcy filed on previous occasions of debt non clearance, check how much relaxation they received post filing of bankruptcy. If the restructured debt amount fits the new loan request (as long as any other driver variables are not giving any red flags), then go ahead with the loan approval else reject the loan request or lower the loan amount

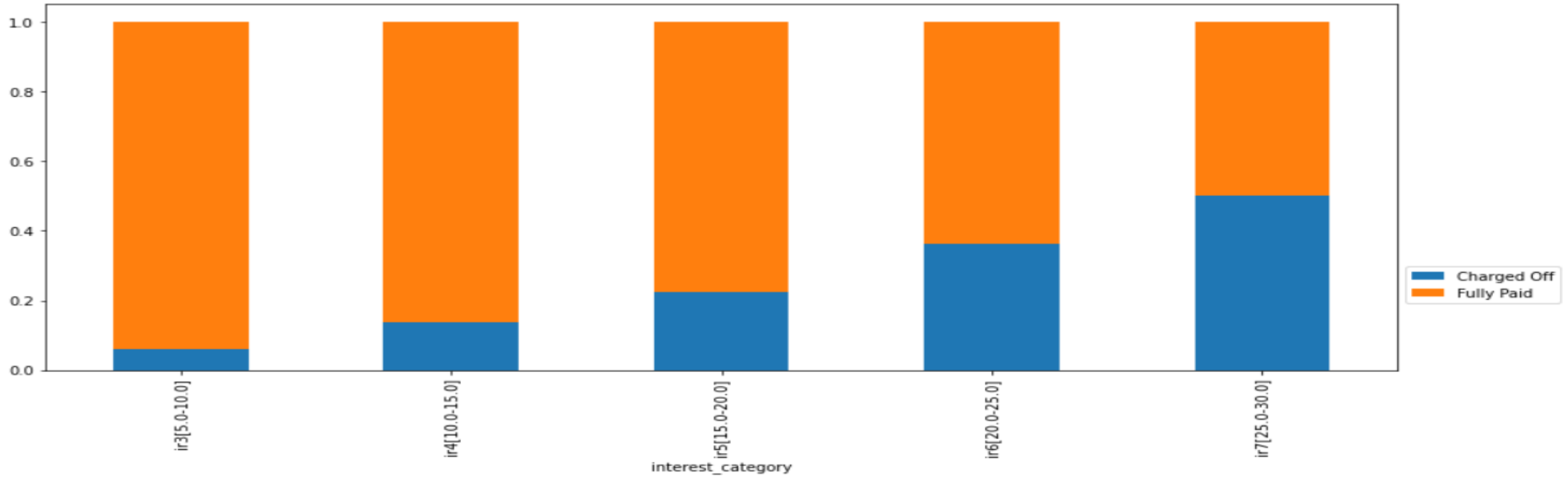
Conclusion:

People with 0 bankruptcies are lesser prone to loan defaulting than people with bankruptcies files. This variable clearly indicates that in previous occasions the loan requestors have failed to pay back the entire money and filed bankruptcy to reduce or eliminate the prior debts

So, clearly they have failed before and are likely to fail again based on the prior history of loan payment

INTEREST RATE

<<<<< DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio (percent) of charged off member count to fully paid member counts for each interest range

INTEREST RATE:

- Interest Rate on the loan
- **Interest Rates have categorized into multiple buckets like [5-10],[10-15],[15-20] and so on….**

Conclusion:

- The Lenders here are doing the correct thing by increasing the interest rates of risky loan requests in order to mitigate the risk of loan defaulting
- Average interest rates for small business is around 13% and it has appeared loans for small business are at higher risks of defaulting

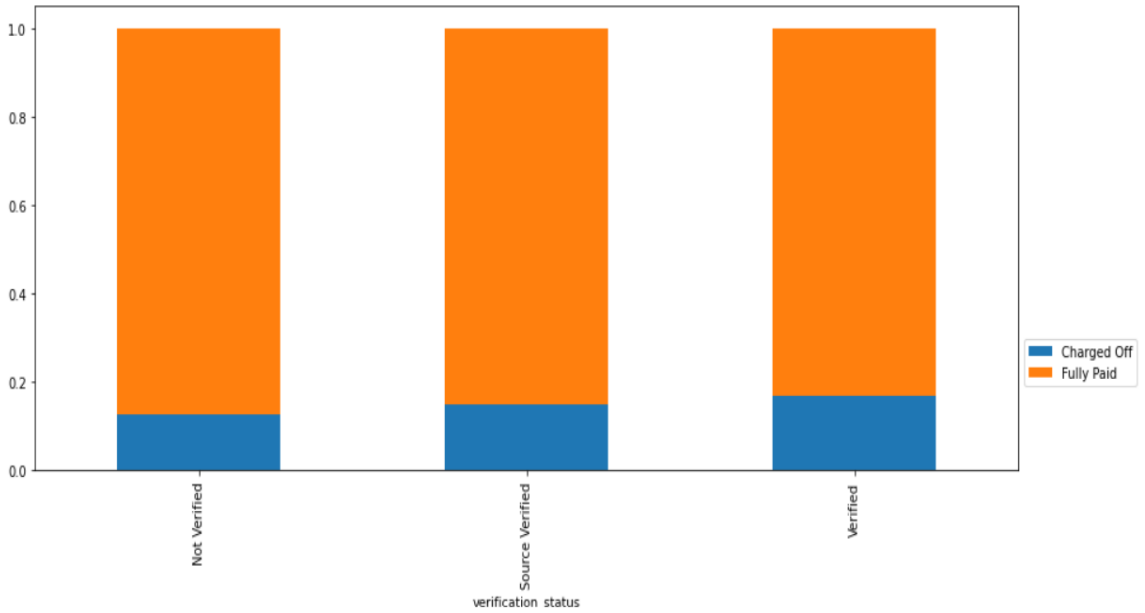
Recommendation:

- Further increase the interest rates for risky loans for purposes like small business (may be around 20 to 25% range)

purpose	average interest rates
car	10.530097
major_purchase	10.843064
vacation	10.856089
home_improvement	11.400265
renewable_energy	11.437573
medical	11.492121
moving	11.563619
educational	11.650831
credit_card	11.701173
other	11.815111
wedding	11.885544
house	12.228136
debt_consolidation	12.487890
small_business	13.009048

VERIFICATION STATUS

<<<<< NOT A DRIVER VARIABLE >>>>>



The stacked chart depicts the ratio of charged off member count to fully paid member counts for each available loan term. But the spike is not clear here

loan_status	Charged Off	Fully Paid
verification_status		
Not Verified	0.128310	0.871690
Source Verified	0.148186	0.851814
Verified	0.168032	0.831968

The above pivot chart clearly depicts the ratio of charged off member counts and fully paid member counts for different verification status

VERIFICATION STATUS:

- Indicates if income was verified by LC, not verified, or if the income source was verified

Recommendations:

- **FRAUD ALERT: AUDIT YOUR VERIFICATION INTERNAL DEPARTMENT/PARTNERS/EXTERNAL SOURCES, IF GUILTY CHANGE THEM**

Conclusion:

The above chart clearly shows the members with verified sources are more likely to default in comparison to non-verified sources, which is really strange as the purpose of verification is to reduce the number of defaulters

DRIVER VARIABLES FOR BIVARIATE ANALYSIS

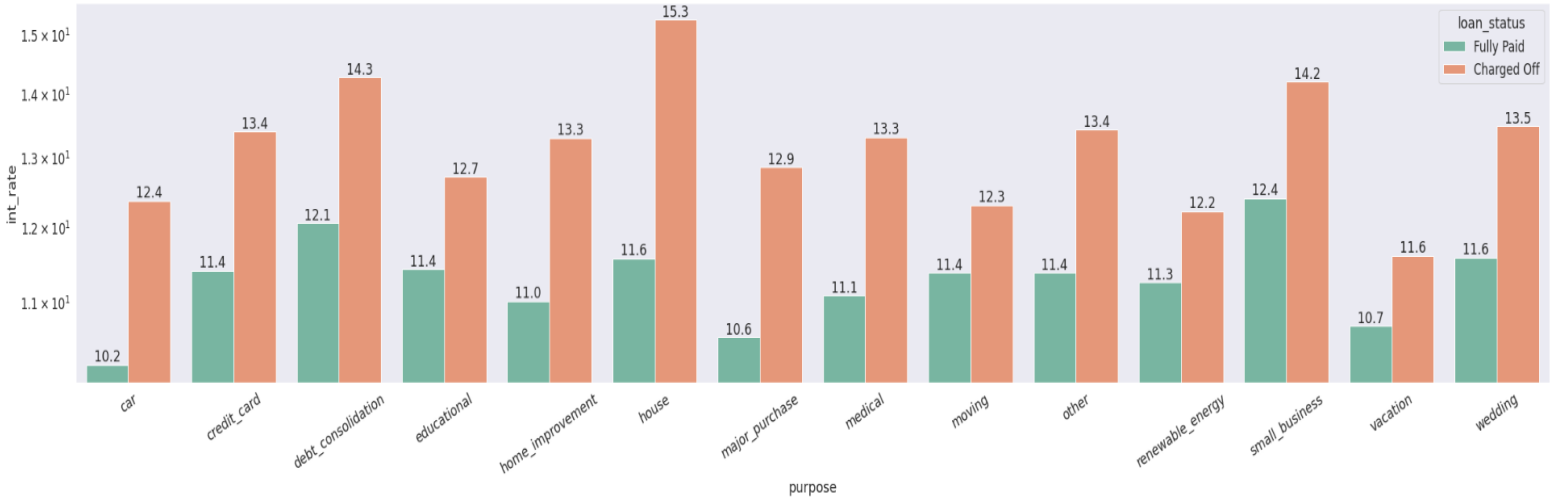
We concluded from the univariate analysis that the following columns are strong indicators to identify loan defaulters.

- DTI
- Interest rate
- Purpose
- Grade
- Loan amount/funded amount/funded amount by investor
- Term
- Annual income

Loan Status appears to be the key column.

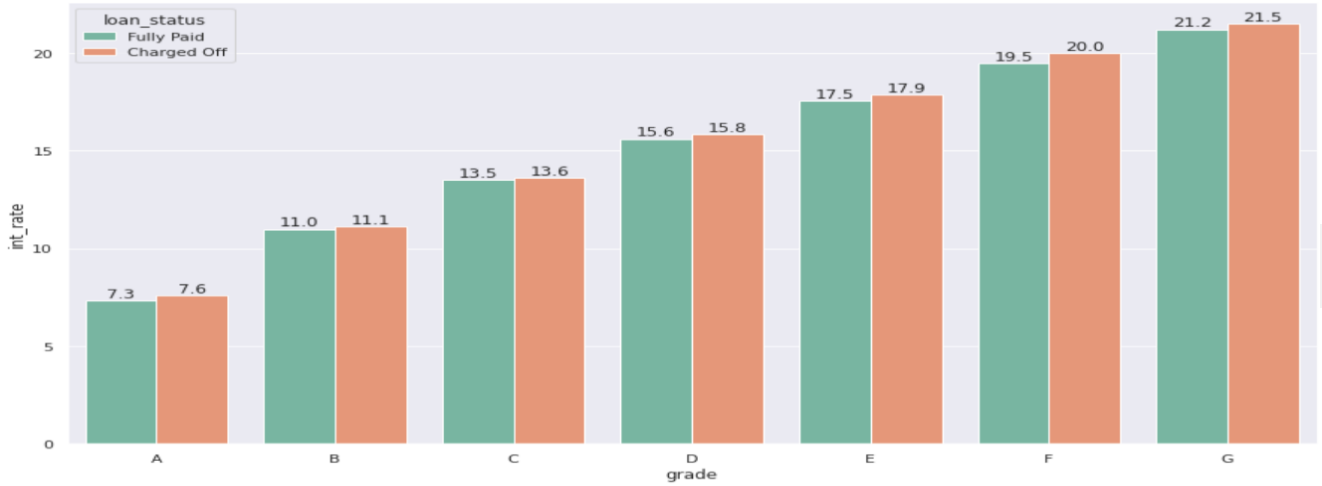
BIVARIATE ANALYSIS-I

Variables:
- Purpose
- Interest rate



Observations:
- Loan extended for home and small businesses are risky.
- Higher the loan interest, higher the charged offs.

Variables:
- Grade
- Interest rate

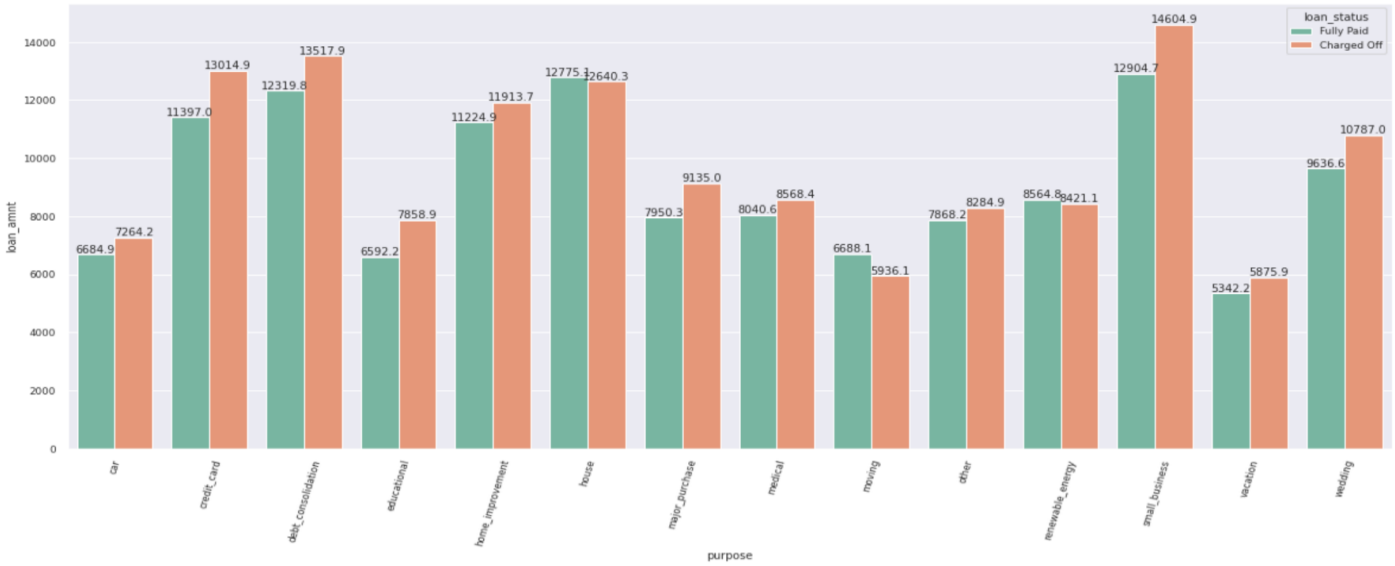


Observation:
- Grade and interest are highly correlated.

BIVARIATE ANALYSIS-II

Variables:

- Purpose
- Loan Amount

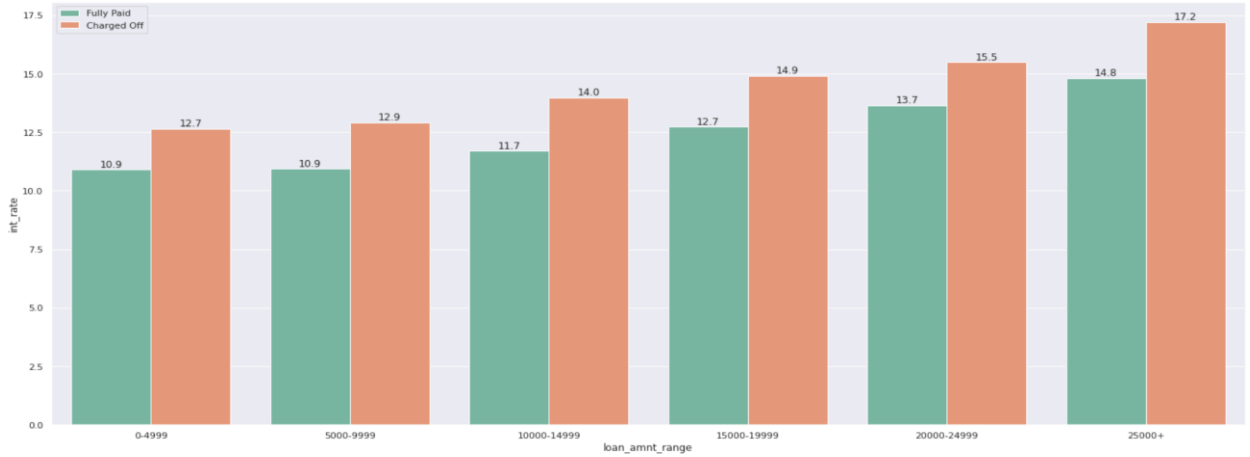


Observation:

- *Loan amount is highest for small businesses followed by debt consolidation*

Variables:

- Loan Amount
- Interest rate

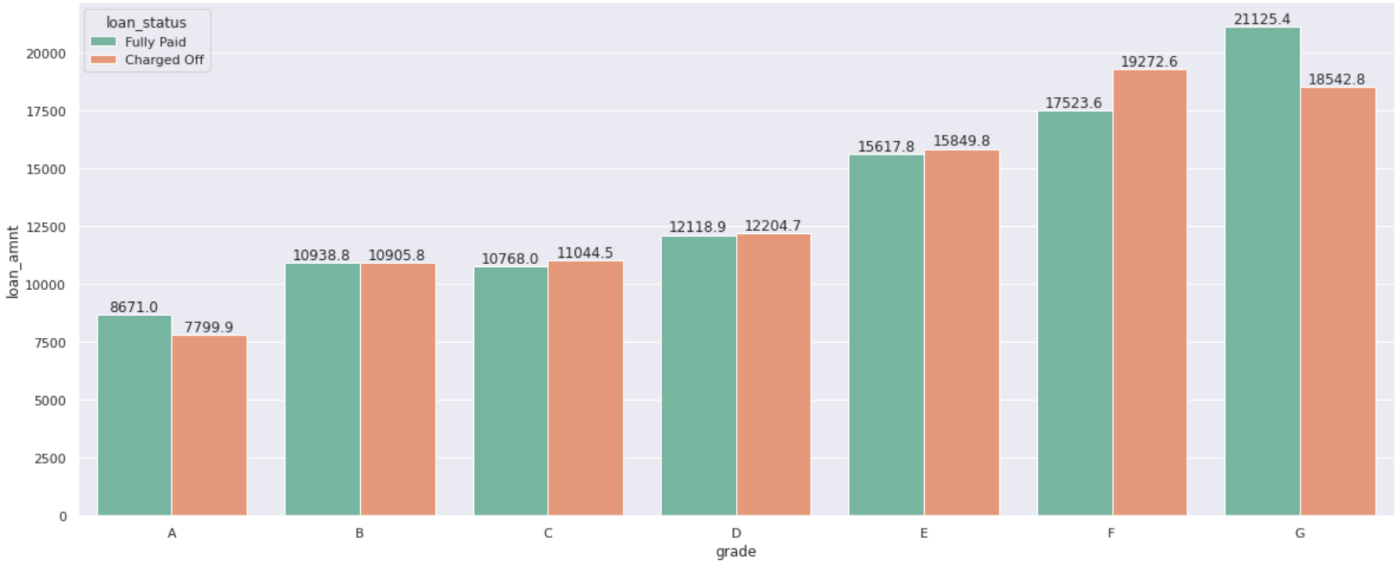


Observations:

- *Higher the loan amount higher the interest rate. The Lending Club wants to mitigate the risk of high charged-off numbers.*
- *Lending club should continue charging a high rate of interest on the higher loan amount.*

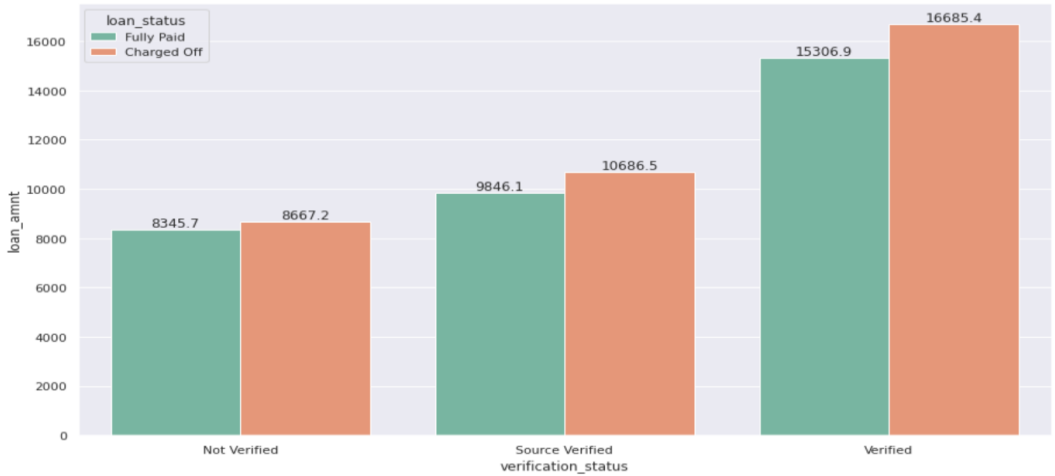
BIVARIATE ANALYSIS-III

Variables:
- Grade
- Loan Amount



Observations:
- *Lower grade customers have applied for a higher amount of loan.*
- *Lending club should take precautions while extending high loan amounts to F grade customers*

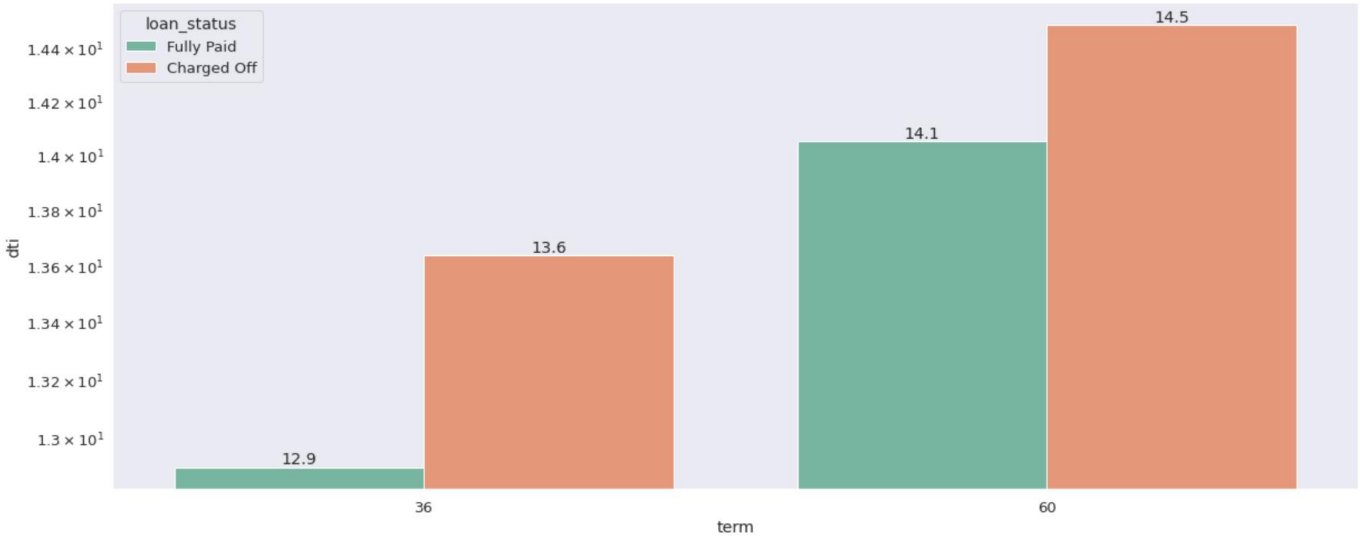
Variables:
- Verification Status
- Loan Amount



Observations:
- *Verified users apply for higher amount of loan and from above analysis.*
- *Higher amount of loan has higher chances of being charged off.*

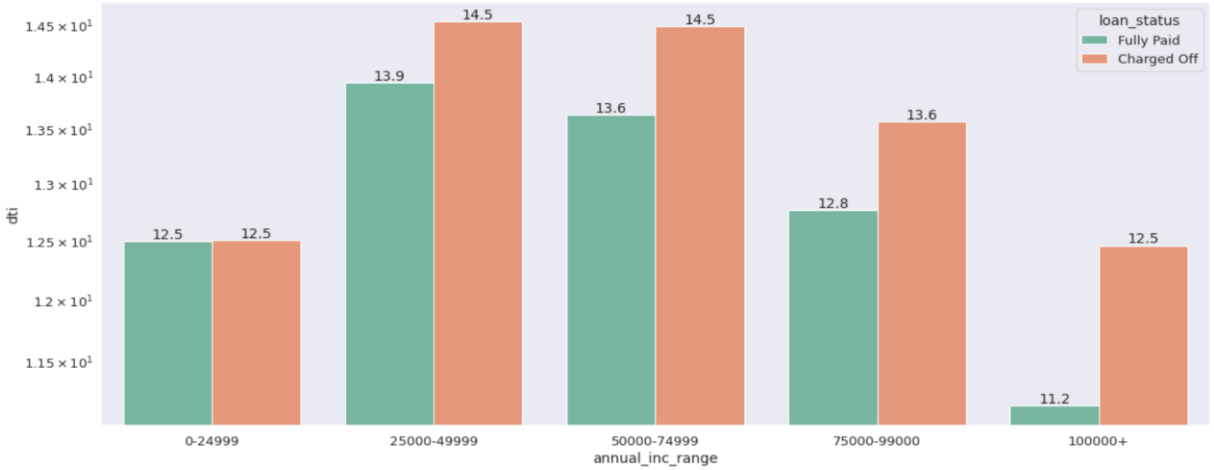
BIVARIATE ANALYSIS-IV

Variables:
- Term
- dti



Observations:
- *'dti' for fully paid customers who opted for 36 months term is ´significantly less compared to the charged-off customers.*
- *When the 'dti' is low the person is able to fully pay on the flip side when the 'dti' is high the person is always charged off.*
- *In the long term 'dti' is not playing that big role in determining fully paid vs charged off.*

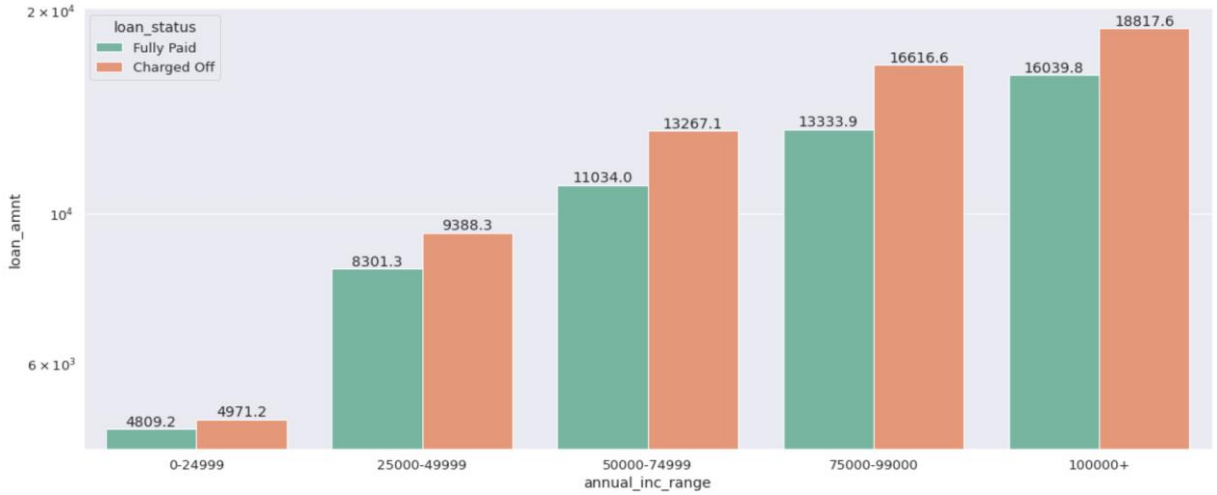
Variables:
- Annual Income
- dti



Observations:
- *For the lowest income range i.e. from 0 to 25k the dti of Fully paid and the dti of Charged off customers are equal which is opposite in other income groups, i.e. for rest all other income groups the dti of Fully paid customers is always lower than the dti of Charged off customers.*
- *For the highest income group i.e. 100000+, the gap in dti between Fully paid and Charged off is more as compared to rest other income groups.*

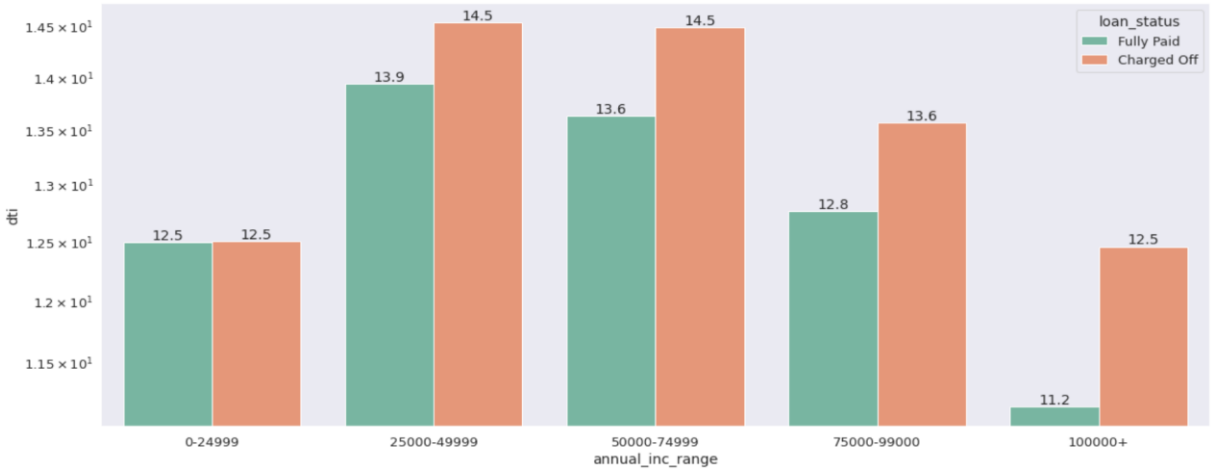
BIVARIATE ANALYSIS-V

Variables:
- Annual Income
- Loan Amount



Observation:
- *If the annual income is low then there is very less difference between fully paid and charged off loans if funded amount is also less. That means less chance of loans getting charged off.*

Variables:
- Annual Income
- dti

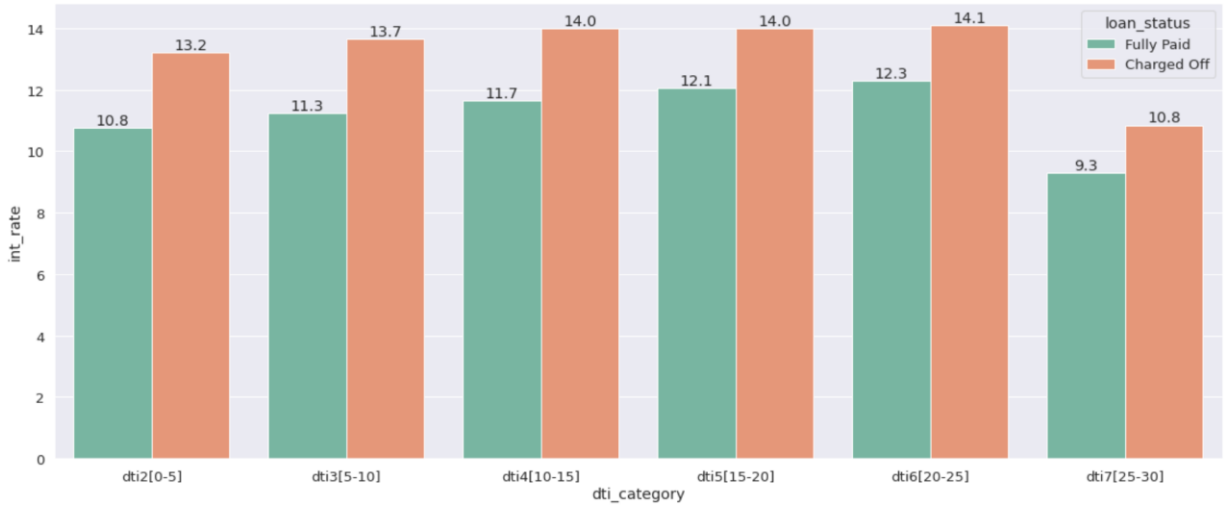


Observations:
- *For the lowest income range i.e. from 0 to 25k the dti of Fully paid and the dti of Charged off customers are equal which is opposite in other income groups, i.e. for rest all other income groups the dti of Fully paid customers is always lower than the dti of Charged off customers.*
- *For the highest income group i.e. 100000+, the gap in dti between Fully paid and Charged off is more as compared to rest other income groups.*

BIVARIATE ANALYSIS-VI

Variables:

- dti category
- Interest Rate

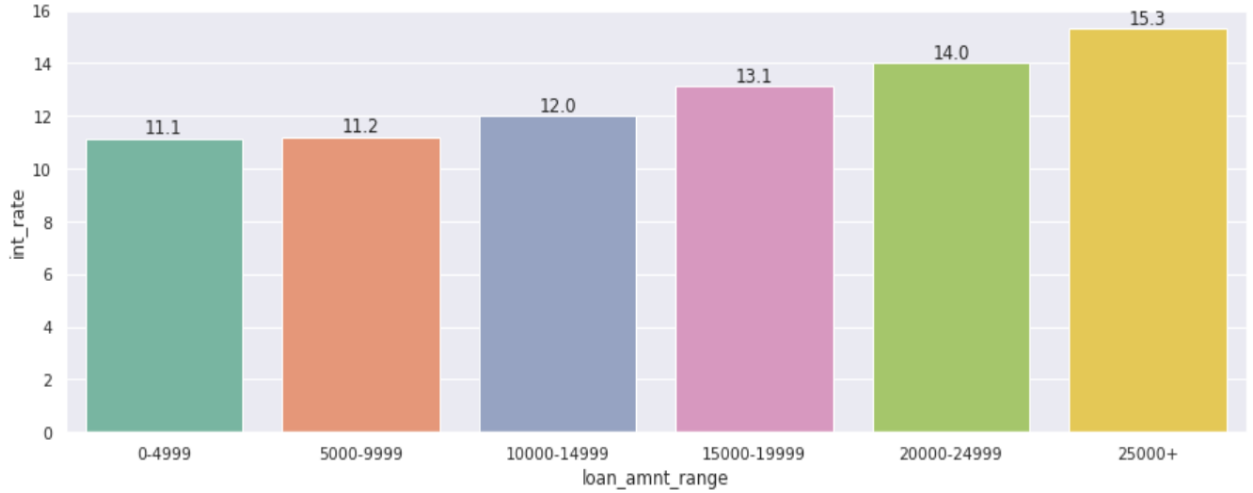


Observation:

- *Lending club should charge higher rate of interest to customers with dti greater then 2 5%*

Variables:

- Loan Amount Range
- Interest rate



Observation:

- *High interest rate is changed for high amount of loan*

CONCLUSIONS

DTI:
Lower the DTI value, lower the chance of defaulting, Higher the value of DTI higher the chances of defaulting

Loan Terms:
Lower the loan terms lower the chances of loan amount being charged off
Higher the loan terms higher the chances of loan amount being charged off

5 years is a bigger term in comparison to 3 years, lot of factors change in these extra 2 years. ***Recession might come in, loss of job, annual income might come down, relocating to a higher cost of living city, increase in home rents, inflation, change in government, new regulations*** are all possible reasons for increase in number of loan defaulter with higher loan terms

Verification Status:
The above chart clearly shows the members with verified sources are more likely to default in comparison to non-verified sources, which is really strange as the purpose of verification is to reduce the number of defaulters

Interest Rates:
The Lenders here are doing the correct thing by increasing the interest rates of risky loan requests in order to mitigate the risk of loan defaulting
Average interest rates for small business is around 13% and it has appeared loans for small business are at higher risks of defaulting

Public Bankruptcies
People with 0 bankruptcies are lesser prone to loan defaulting than people with bankruptcies files. This variable clearly indicates that in previous occasions the loan requestors have failed to pay back the entire money and filed bankruptcy to reduce or eliminate the prior debts
So, clearly they have failed before and are likely to fail again based on the prior history of loan payment

Loan Amount:
Higher the fund amount requested higher is the risk of defaulting

Purpose:
Smaller business seems to be having a higher risk of getting charged off in comparison to loans taken for personal purposes like wedding and car

Business related loans are always at higher risks as loan payments are often related to the profit made

State Code, Zip Code & Cost of Living:
The first chart " address state vs Charged off/Fully Paid ratio chart" has **no trends** as such

But, If we further drill down to Zip Codes, we can see the below trend
From the first chart let us consider Florida which is at higher risk of getting charged off
Then let us plot a ratio chart for all the zip codes in Florida.
Below is the observation:
322XX is Jacksonville FL, which is at a lower risk of defaulting
where as
331XX is Miami FL, which is at higher risk of defaulting
The cost of living in Miami, FL is 20.5% higher than in Jacksonville, FL Employers in Miami, FL typically pay 2.7% more than employers in Jacksonville, FL for the same type of job

HIGHER COST OF LIVING PLACES ARE MORE LIKELY TO BE DEFAULTERS

Grades & Sub-Grades:
As we move from lower grade A to higher grade G, chances of defaulting increases
Similar trend is seen among sub grades within grades. Within a Grade, even if we move from 1 to 5 sub grade, the possibility of defaulting a loan is higher (**in maximum case**)

Annual Income:
Lower the Annual Income range, higher the chance of defaulting, Higher the Annual Income range, lower the chances of defaulting
Negative correlation detected

RECOMMENDATIONS

DTI:
If the DTI value is high, lenders can think of either not approving the loan or think of reducing the loan amount, if other variables are not putting red flags
If the DTI ratio is high, lenders can look at other variables more carefully like cost of living in a particular place, annual income trends over the years (up or down?)

Loan Terms:
Incase there are other variables showing red flags, then the lender can go for a shorter term considering factors like the current annual income is high or if one foresee a recession or rise in real estate price in the current locality

Verification Status:
FRAUD ALERT: AUDIT YOUR VERIFICATION INTERNAL DEPARTMENT/PARTNERS/EXTERNAL SOURCES, IF GUILTY CHANGE THEM

Interest Rates:
Annual Income high definitely is a major factor for loan approval, but along with the **Annual Income**, do consider the **DTI and loan mount to Annual Income Ratio** as well

Public Bankruptcies
For people who have a public record of bankruptcy filed on previous occasions of debt non clearance, check how much relaxation they received post filing of bankruptcy. If the restructured debt amount fits the new loan request (as long as any other driver variables are not giving any red flags), then go ahead with the loan approval else reject the loan request or lower the loan amount

Purpose:
In case the loans are for business purpose, have options of recovering back from assets like home owned by the loan requestor. Have a proper background check to find out how successful the requestor has been in previous businesses. Also deep digger into the business to find out the probability of the business getting successful. Figure out who are the competitors of that business. Figure out how niche is that business.

State Code, Zip Code & Cost of Living:
Add a new dataset which has city and it's cost of living or a rank in terms of cost of living or some other metrics like average home rent in that county or city, which can give us some scope for analysis. Then map it against annual income.

Annual Income:
Annual Income high definitely is a major factor for loan approval, but along with the **Annual Income**, do consider the **DTI and loan mount to Annual Income Ratio** as well