

**upGrad**

Executive PG Programme in Machine Learning  
& AI - December 2021



**LendingClub**

LENDING CLUB

*DATA ANALYSIS by*

- *SANGHAMITRA DEY  
(Group Member)*
- *SHASHEESH RANE  
(Group Facilitator)*



We have been Assigned as a team to analyse past loan data from 2007 to 2011

Purpose: To Identify the cases and categorize the Loans having higher Charged-Off Rate based on Demographics and other financial factors

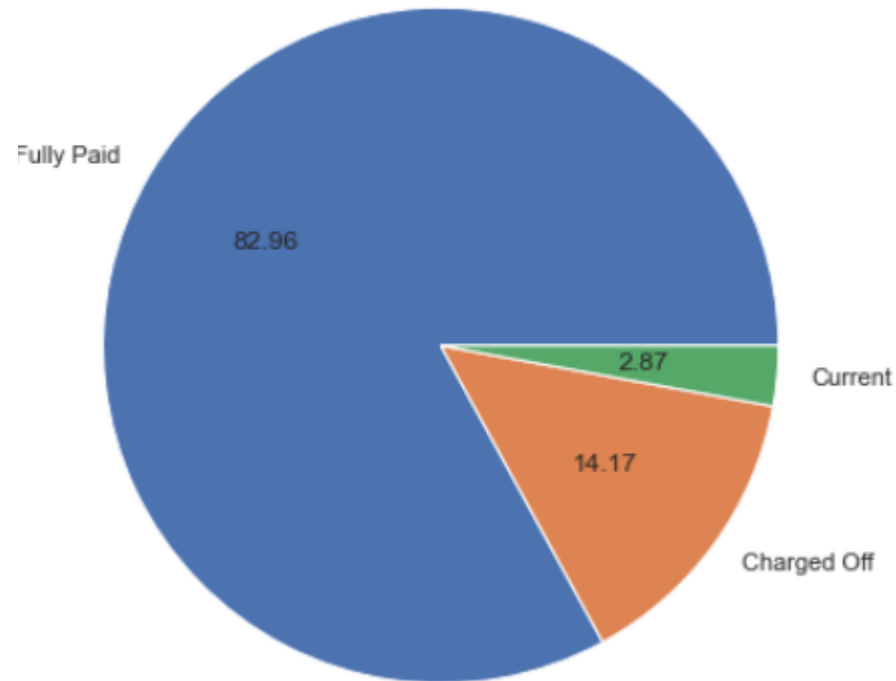
## BUSINESS UNDERSTANDING CONCEPTS

- Working for a consumer finance company which specializes in lending various types of loans to customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.
- Two Major types of risks are associated with the bank's decision: 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.
- The data given below contains the information about past loan applicants and whether they 'defaulted' or not. 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. We will try to use EDA to understand how consumer attributes and loan attributes influence the tendency of default.





OVERALL LOAN STATUS



HERE WE SEE ABOVE MAJORLY , of overall Data:

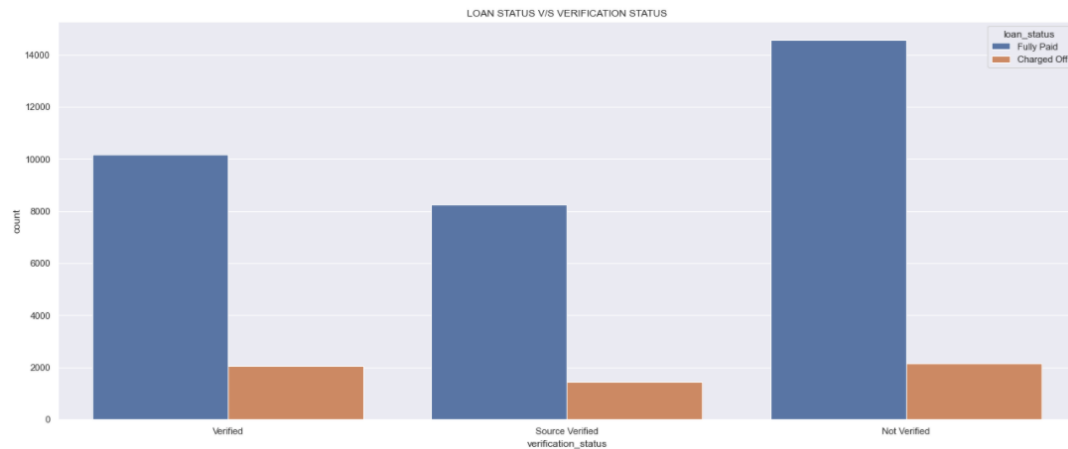
- **FULLY PAID** = 82.96%,
- **CHARGED OFF** = 14.17%
- **Current** = 2.87%

But From now on our major Focus will be Fully Paid and Charged Off. Since we need to calculate Risks

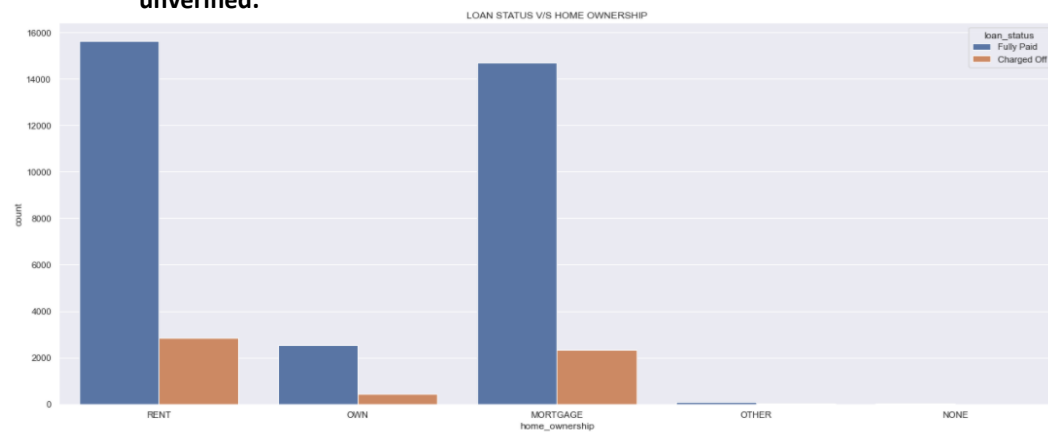
We are not going after customer behaviour variables because its already approved

To know whether the loans accepted will not be a defaulter

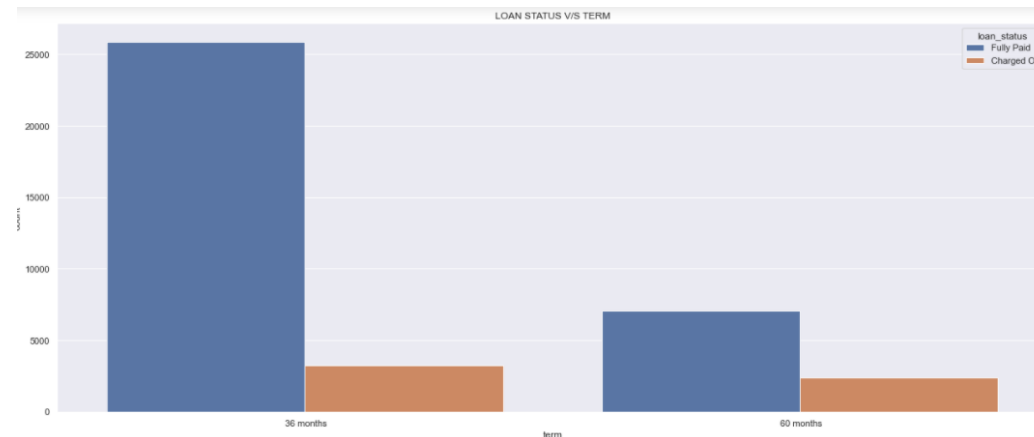
## FEW INSIGHTS:



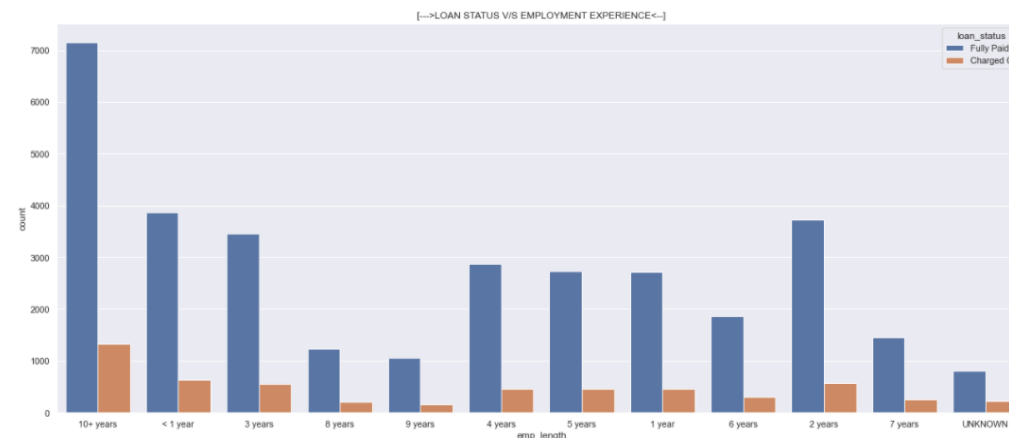
a. Being Verified or un-verified doesn't have an effect on Charged-off count, in fact the cases of Fully Paid are much better in unverified.



c. Home owners are at the least risk of Being Charged-Off, amongst rented & mortgage.

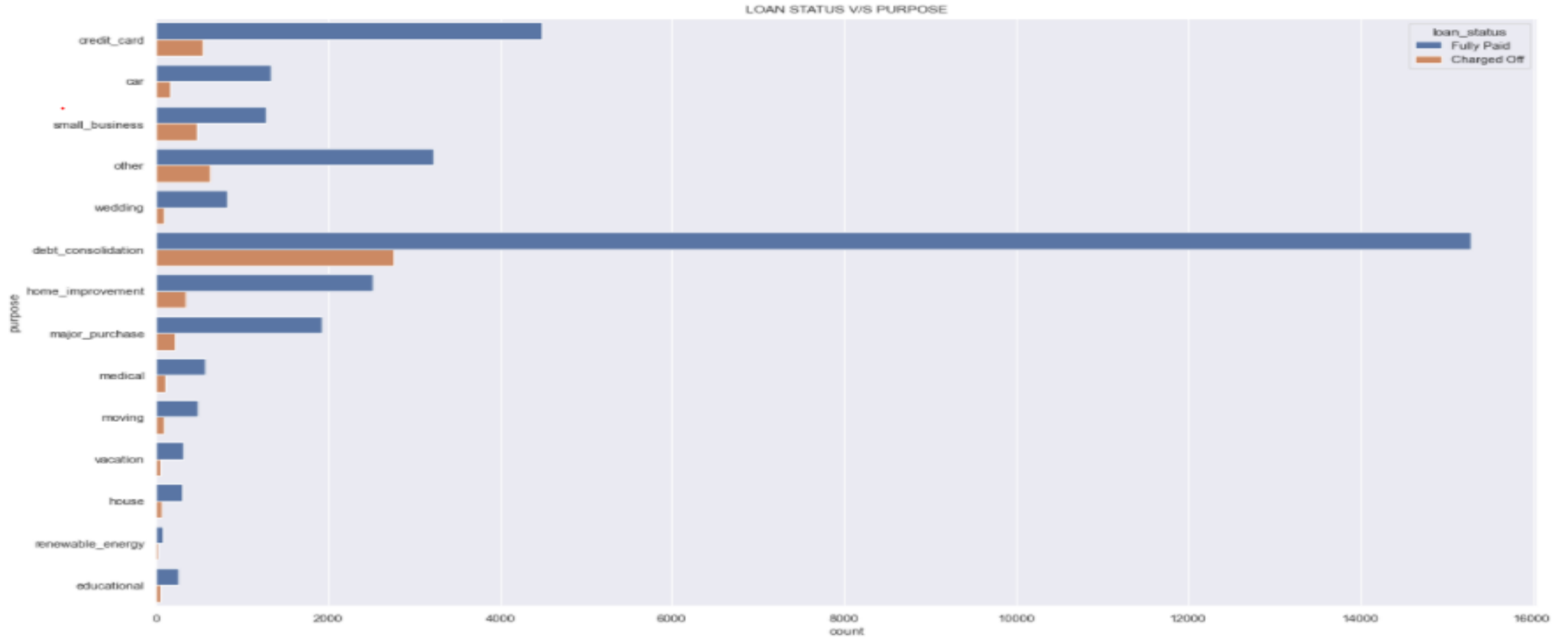


b. As expected , shorter term of 36 months has more cases of Charged-off.



d. From 6,7,8,9 years of experience have lesser count of Charged-off cases, and 10+ years have the highest Fully paid counts

# PURPOSE OF LOANS



- Highest risk cases come from debt\_consolidation, credit card, other have an effect on Charged-off count.
- Lowest Risks of Charged-off seem to come from education, renewable energy, vacation.

# GETTING A LOOK INTO THE Funded Amount by Investors:

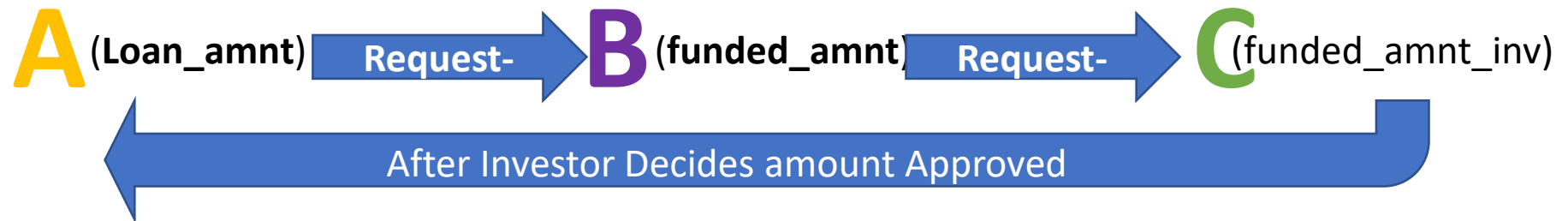
Our Logical Inference from the Data is :

A = Applicant

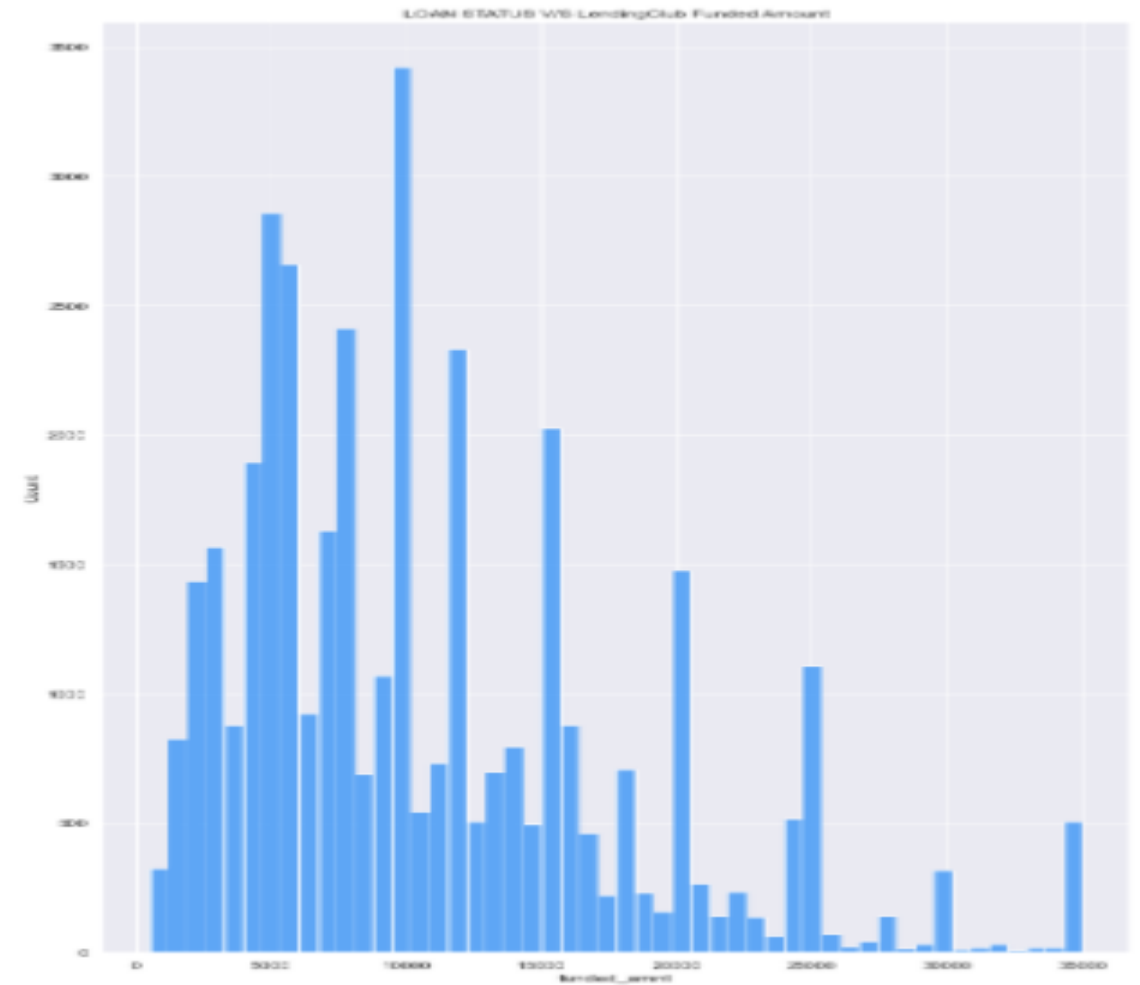
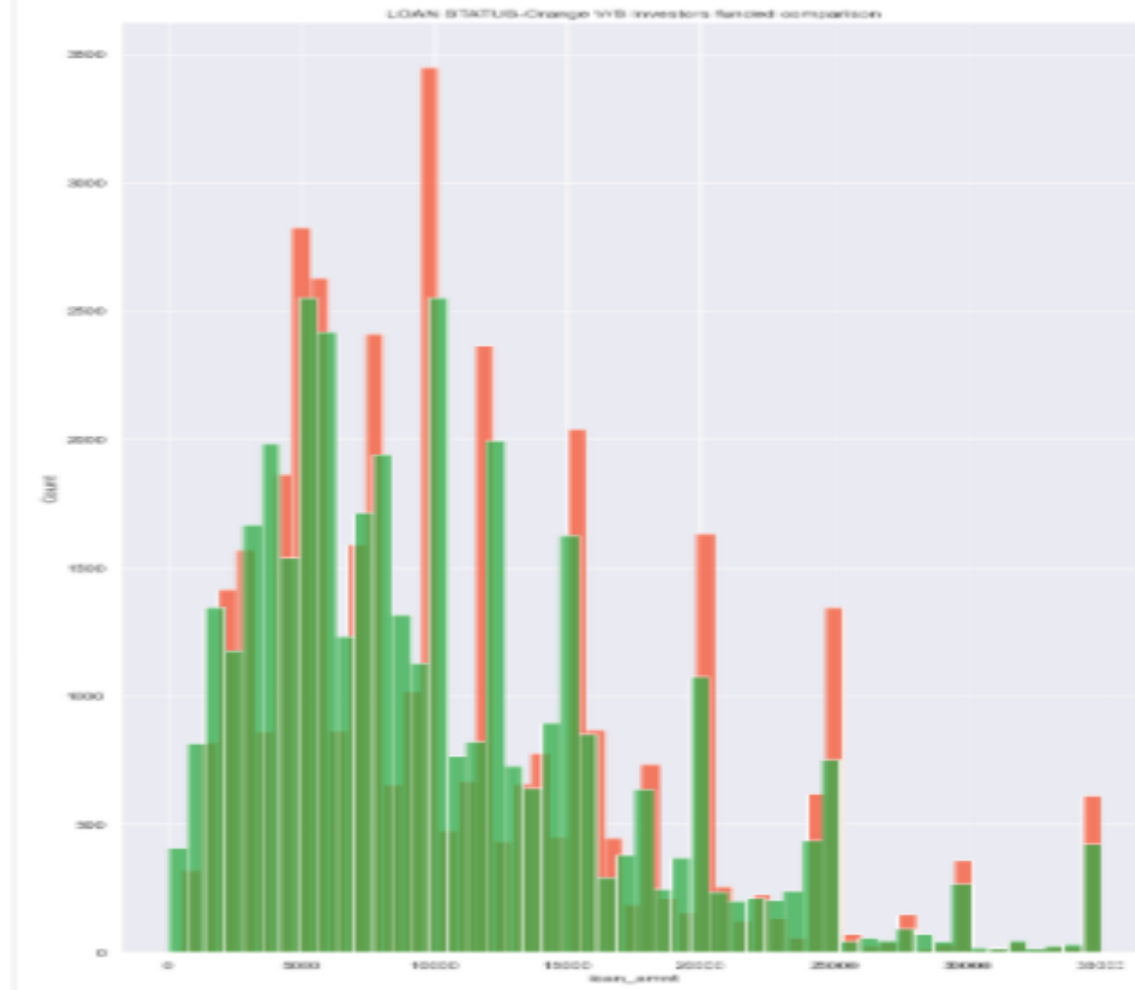
B = Lending Club

C = Investors

Here WE SEE,



At point C the Final Approval Funded Amount is decided by the investors to Pass the Loan to the A (Applicant)

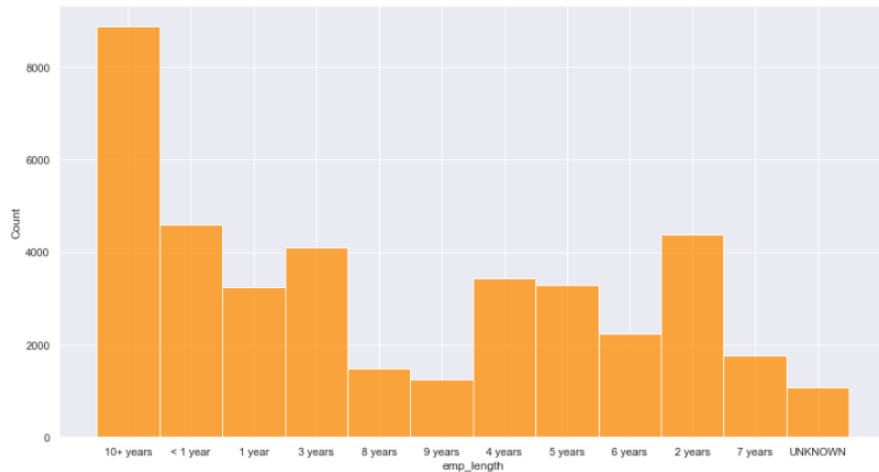


- When we compare against the A(Applicant) v/s C(Investors Funded), We can make a conclusion that the Investors are keen on giving lesser funds than required to the borrowers.

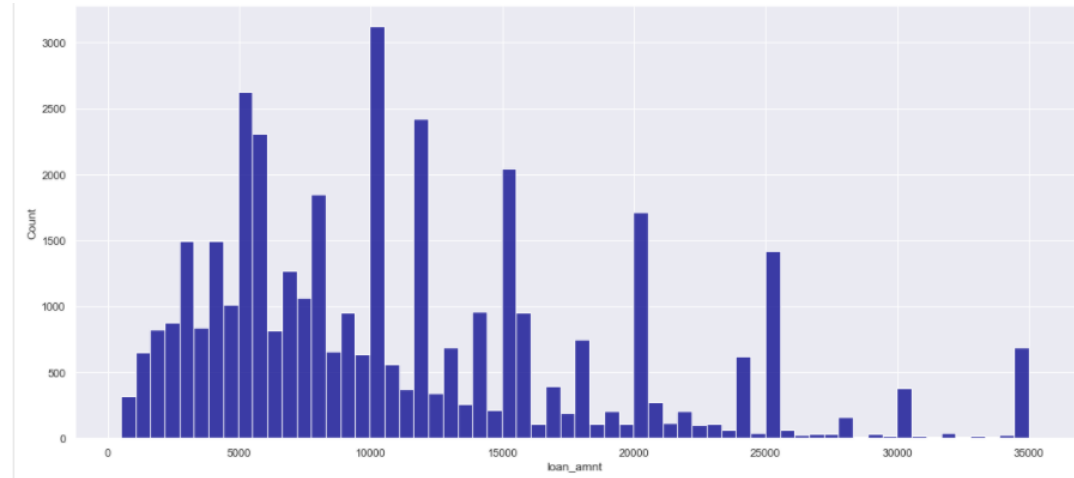
Here we can make a deduction by checking the Loan Amount v/s charged off and see which range the investors can approve of loans of particular amount. So they can avoid risk of charged off Applications.

We can see that the LendingClub and Investors are very well synchronized in terms of decision making for issuing the loans.,





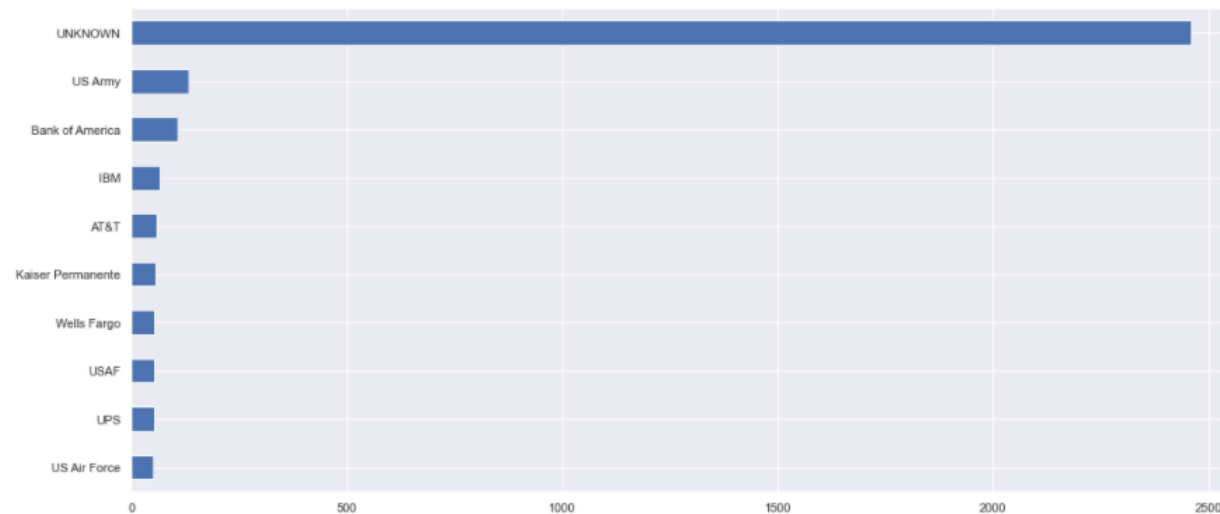
Here we see that highest amount loans are availed by people of work experience +10 Years, > 8000 People, 2nd best is < 1 year



Also the count of the People availing loans(>3000) taken around 10000 are the highest.

## Demographics insights:

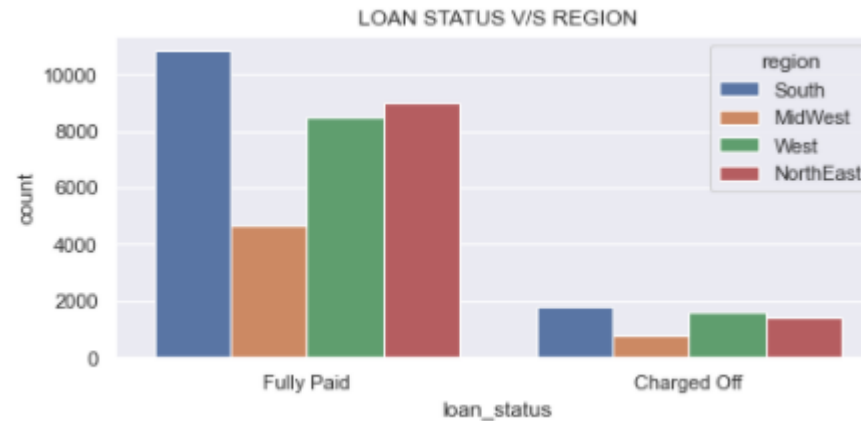
- Employment Period
- Loan Amount
- Companies



Here we see that highest number of loans are availed by people from

- US ARMY ,
- Bank of America ,
- IBM,
- AT & T

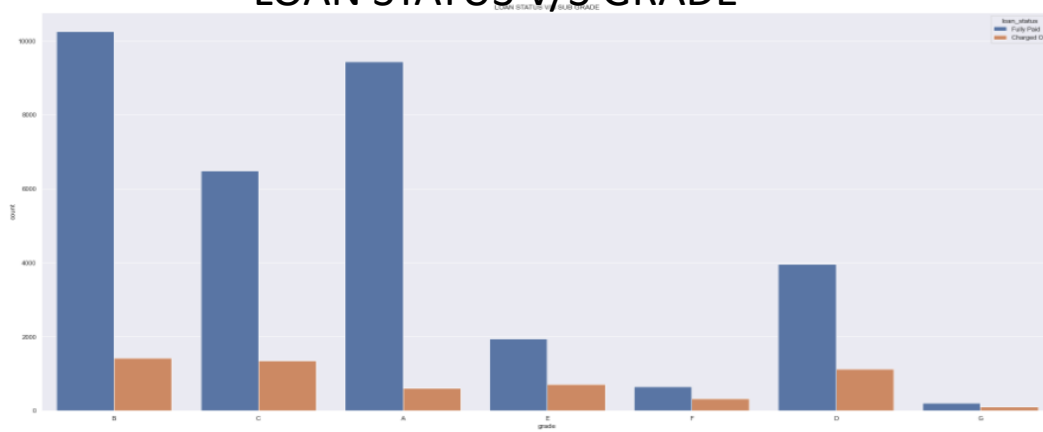
are in the TOP 4#  
BUSINESS DRIVER : Loans can be marketed easily to the TOP 10 companies



## More Demographics insights

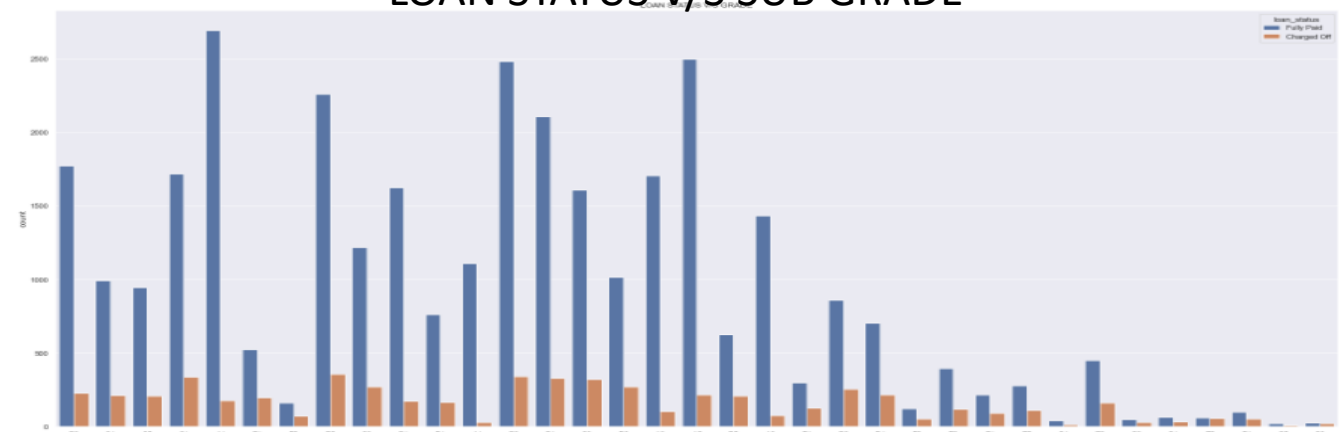
The South Region has the Most number of Loans Available, having the most number of Charged offs.  
The Number of Charged - offs gets higher in the West > Northeast

### LOAN STATUS V/S GRADE



Based on Above Graph we can infer that Grades B,C, A are availing max loans and repaying them, whereas Grade D , is the 2nd Highest Charged Off

### LOAN STATUS V/S SUB GRADE



Based on Above Graphs we can infer that Sub-Grades B5,B3, C1 are Highest Charged Off.  
TOP #4 Grades who are Fully paid are A4, B5, B3, A5  
there's a clear trend of higher probability of charge-off as the subgrade lowers.

# Our Hypothesis

Business Driver Variables :

**LOAN AMOUNT**

**INTEREST RATE**

**ANNUAL INCOME**

**DEBT TO INCOME(DTI)**

**TERM**

We have selected these based on our understanding of Payment Capacity of the loan:

e.g. if Loan amount/Interest rate is high then chances are payment may default

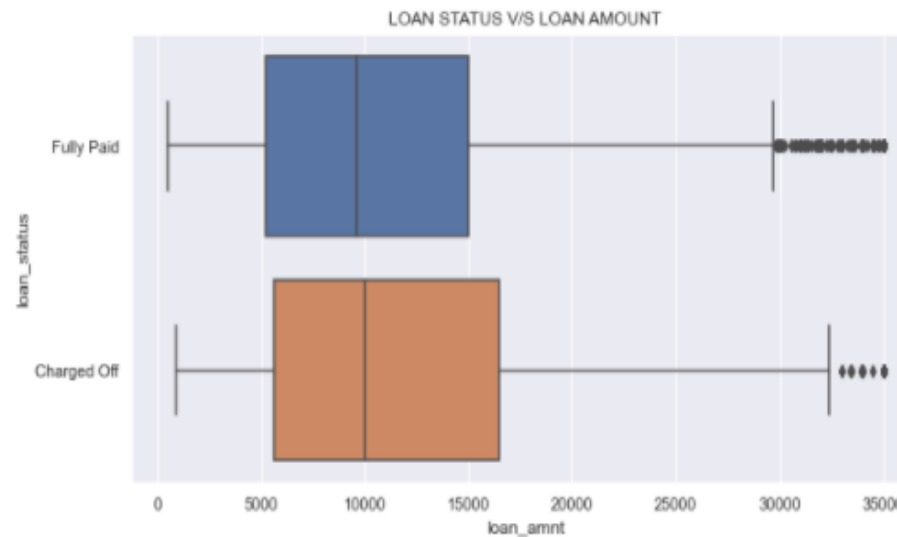
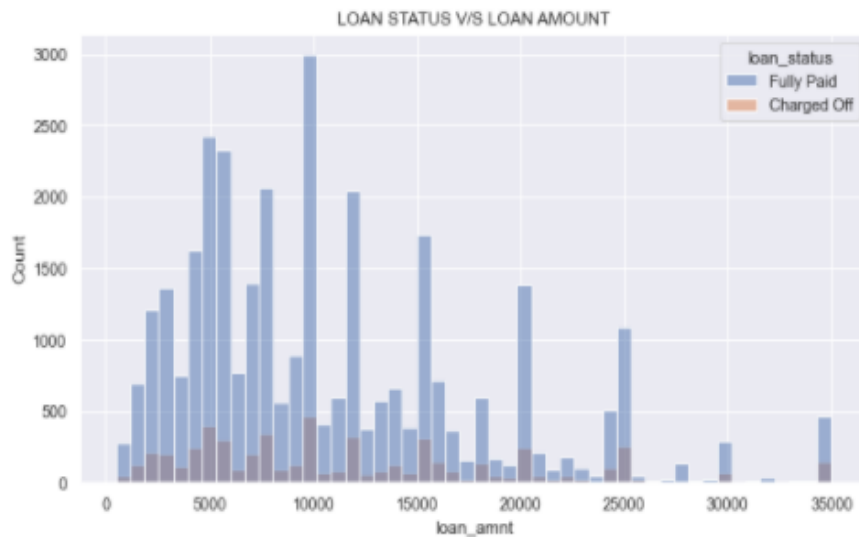
One more reason is : For the LendingClub the above may seem ordered Numerical Verified Values when issuing the loan

***In the following 5 slides we will discuss each business driver Variable***

# Loan Status V/S LOAN AMOUNT

Business Driver Variable

## LOAN AMOUNT



### Insight:

Higher Loan amount ranges

➤ 16500 (Charged Off) and

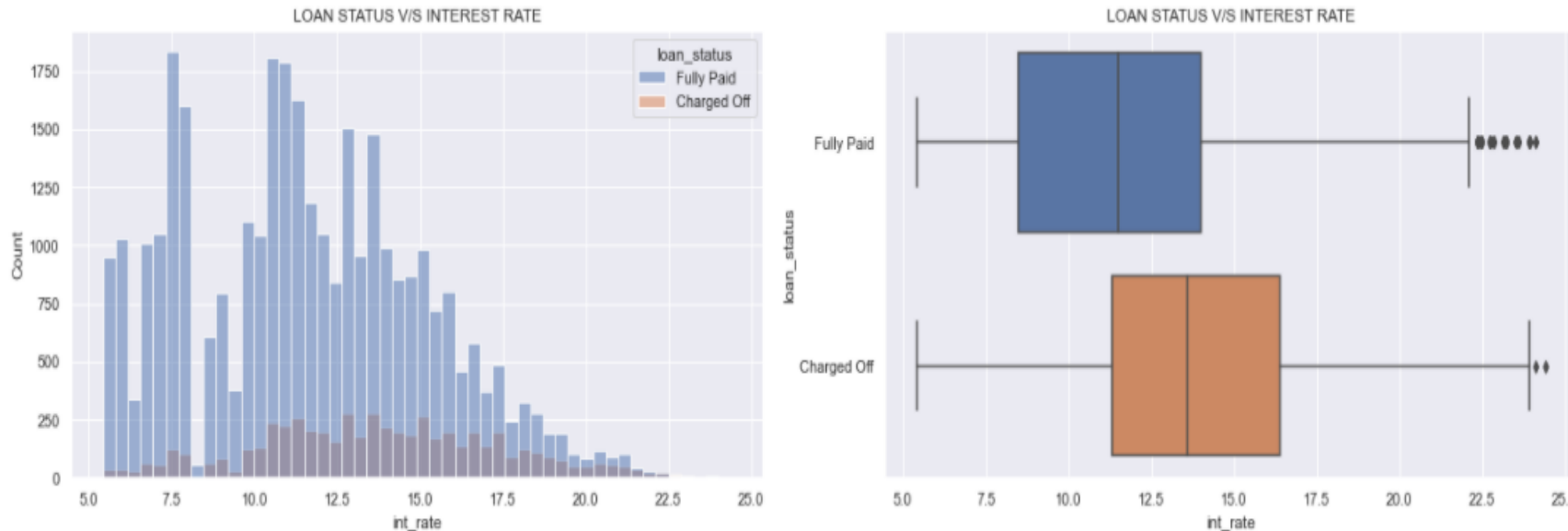
➤ Fully Paid is > 15000,

Highest Charged Off counts are in range 1000 -> charged off <- 15000



	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	5627.0	12104.385108	8085.732038	900.0	5600.0	10000.0	16500.0	35000.0
Fully Paid	32950.0	10866.455994	7199.629493	500.0	5200.0	9600.0	15000.0	35000.0

# Loan Status V/S INTEREST RATE



Business Driver Variable

## INTEREST RATE

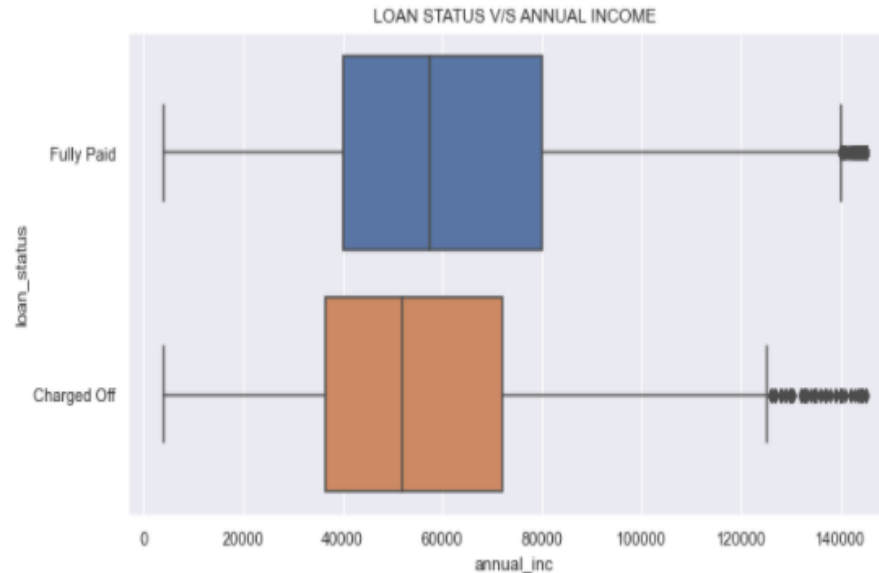
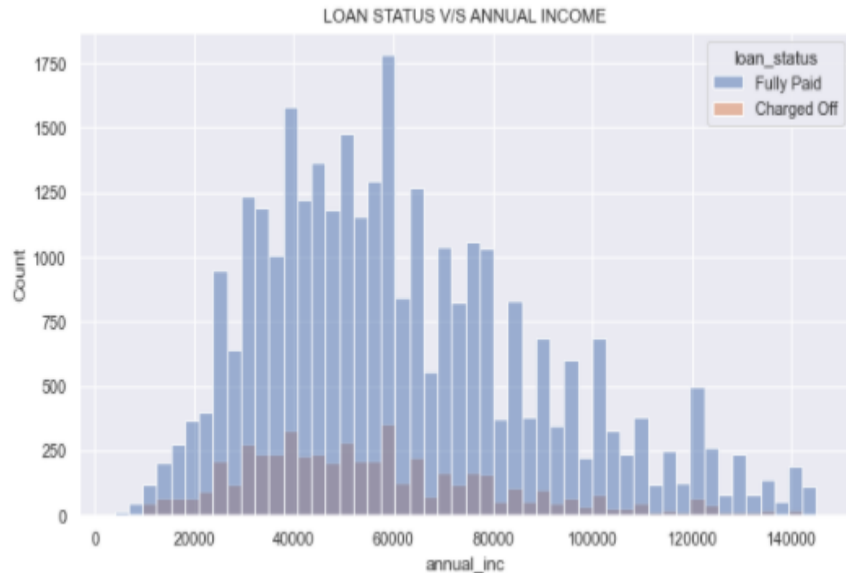


### Insight:

- Here we see that the Interest Rate between 10 to 20% has the maximum charge offs.
- So the lending club can decide to increase or decrease the interest rate to promote or deter the applications.

	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	5627.0	13.820432	3.654413	5.42	11.31	13.61	16.40	24.40
Fully Paid	32950.0	11.609761	3.599964	5.42	8.49	11.49	13.99	24.11

# Loan Status V/S ANNUAL INCOME



Business Driver Variable

## ANNUAL INCOME



### Insight:

- We see a high volume of fully paid between 30000 to 60000.
- Same range has higher number of defaulters too. but as the loan amount increases the charged off cases reduce beyond 11000.

	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	5431.0	56650.282105	26889.449133	4080.0	36400.0	51996.0	72000.0	145000.0
Fully Paid	31384.0	62008.670381	28375.140167	4000.0	40000.0	57496.0	80000.0	145000.0

# Loan Status V/S (DTI) DEBT TO INCOME

Business Driver Variable

## DTI-(Debt to Income)



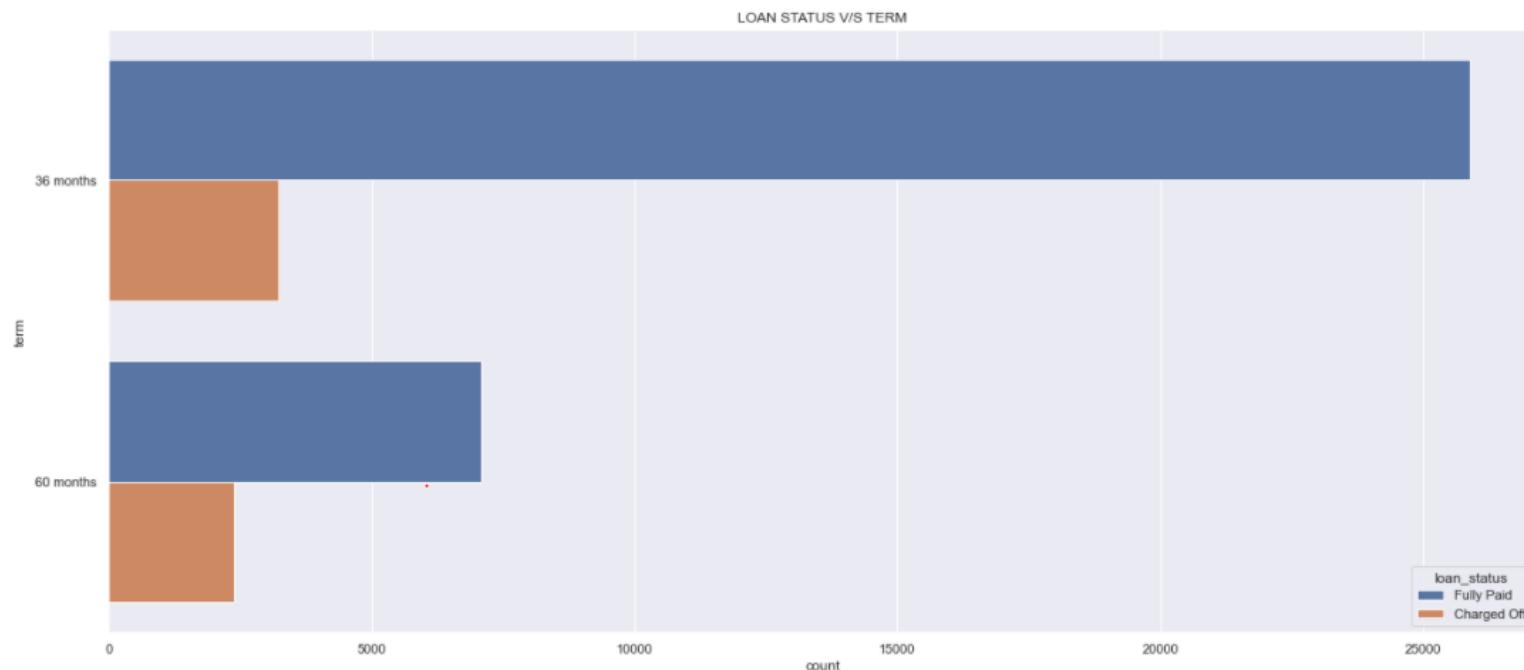
### Insight:

- As per Above Graphs the **Charged off Status is higher in DTIs ranging from 10 to 25**, Should be considered when issuing loans.
- The Average DTI mean = 14, and Max = 29.85 , for Charged Off



	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	5431.0	14.102541	6.584022	0.0	9.18	14.40	19.40	29.85
Fully Paid	31384.0	13.322548	6.667198	0.0	8.19	13.42	18.58	29.99

## Loan Status V/S TERM

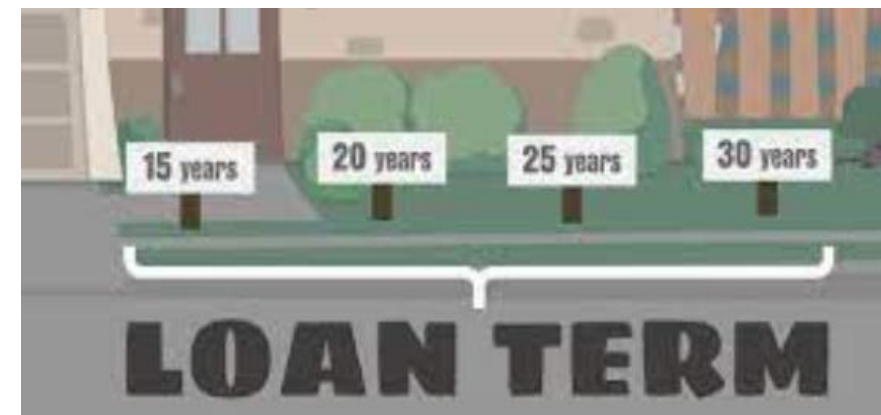


### Insight:

*Here we see that for 36 months duration the cases are higher for Charged-off  
But the Fully paid cases are also Higher, Its a High risk high gain scenario.*

Business Driver Variable

# TERM



	count	unique	top	freq
loan_status				
Charged Off	5431	2	36 months	3120
Fully Paid	31384	2	36 months	24717





- ❖ LendingClub should try giving long term loans , lower amount at lower interest rate, to parties who can pay off or have higher reliability based on certain demographic.
- ❖ Home Owners, Education seekers can will avail of this loan, attractive rates of interest can be given to increase business around > 10000 to 12000 USD.
- ❖ Higher interest rate should be given to segments who can afford to pay on time , like Home Owners, Grade-A,B,C, Work experience 7,8,9 years
- ❖ Education, Vacation and renewable energy are lower risk purposed where the LendingClub can actually provide attractive Schemes and good Payment EMIs to attract more customers
- ❖ LendingClub need to be careful ,Grade D Applicants, they have moderately average Charged-Off count
- ❖ Same goes for West Zone Applicants region wise.
- ❖ US ARMY , Bank of America , IBM, AT & T are in the TOP 4# BUSINESS DRIVER : Loans can be marketed easily to the TOP 10 companies



# Thank you

*Smt. Sanghamitra Dey*  
*Shri. Shasheesh Rane*

