

# Lending Club Case-Study

*(2007-2011)*

**Pandi & Kishore (MLC39)**

## About Lending-Club Company:

- It is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- Borrowers can easily access lower interest rate loans through a fast online interface.

Source:<https://www.lendingclub.com>

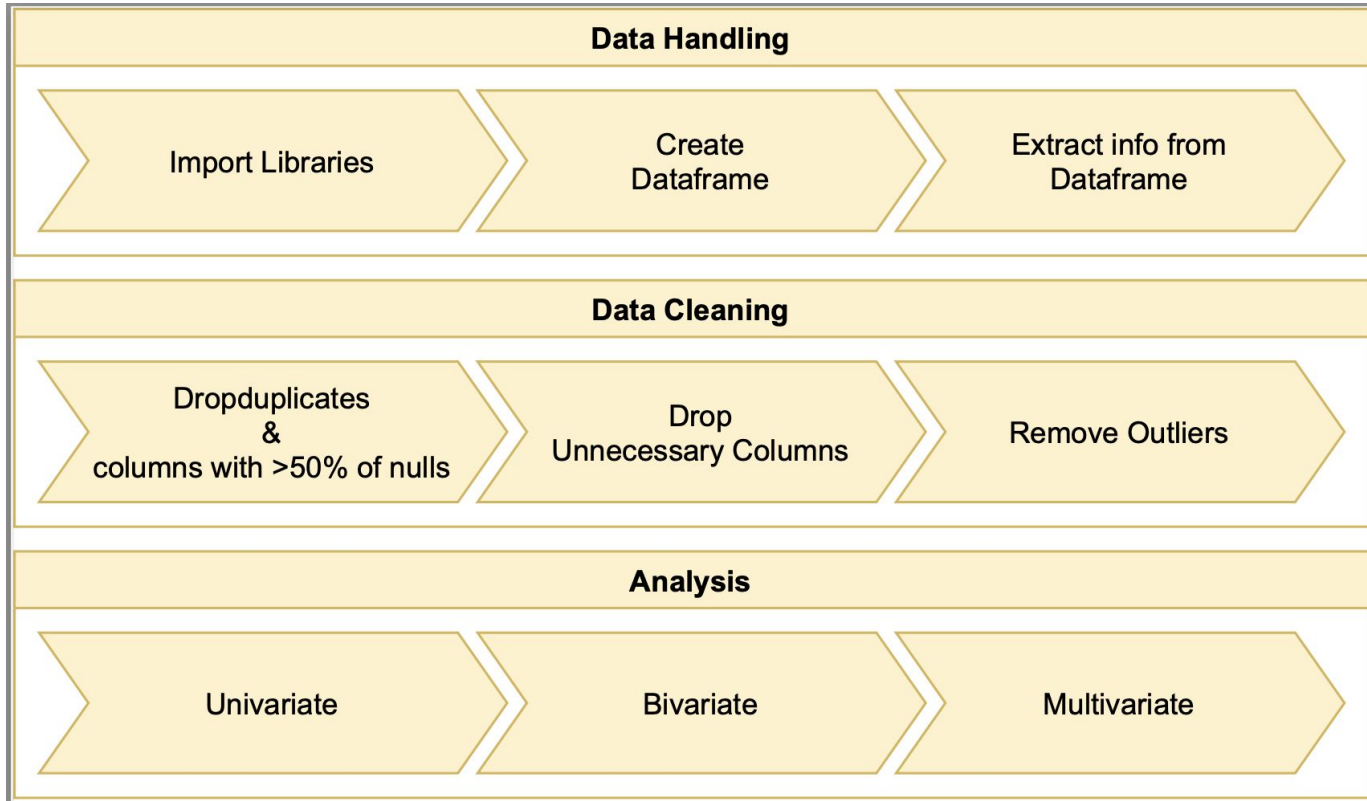
## Problem Statement:

Like most other lending companies, **lending loans to 'risky' applicants is the largest source of financial loss** (called credit loss). The credit loss is the amount of money lost by the lender when the **borrower refuses to pay or runs away with the money owed**.

## Expected Solution:

1. Identify these risky loan applicants
2. Understand the 'Driver Factors' behind 'Loan Default'

# Process flow:



# Data Handling:

Dataset information: “loan.csv”

Columns	111
Rows	39717

Columns\_names

id,member\_id,loan\_amnt,funded\_amnt,funded\_amnt\_inv,term,int\_rate,installment,grade,sub\_grade,emp\_title,emp\_length,home\_ownership,annual\_inc,verification\_status,issue\_d,loan\_status,pymnt\_plan,url,desc,purpose,title,zip\_code,addr\_state,dti,delinq\_2yrs,earliest\_cr\_line,inq\_last\_6mths,mths\_since\_last\_delinq,mths\_since\_last\_record,open\_acc,pub\_rec,revol\_bal,revol\_util,total\_acc,initial\_list\_status,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,next\_pymnt\_d,last\_credit\_pull\_d,collections\_12\_mths\_ex\_med,mths\_since\_last\_major\_derog,policy\_code,application\_type,annual\_inc\_joint,dti\_joint,verification\_status\_joint,acc\_now\_delinq,tot\_coll\_amt,tot\_cur\_bal,open\_acc\_6m,open\_il\_6m,open\_il\_12m,open\_il\_24m,mths\_since\_rent\_il,total\_bal\_il,il\_util,open\_rv\_12m,open\_rv\_24m,max\_bal\_bc,all\_util,total\_rev\_hi\_lim,inq\_fi,total\_cu\_tl,inq\_last\_12m,acc\_open\_past\_24mths,avg\_cur\_bal,bc\_open\_to\_buy,bc\_util,chargeoff\_within\_12\_mths,delinq\_amnt,mo\_sin\_old\_il\_acct,mo\_sin\_old\_rev\_tl\_op,mo\_sin\_rent\_rev\_tl\_op,mo\_sin\_rent\_tl,mort\_acc,mths\_since\_recent\_bc,mths\_since\_recent\_bc\_dlq,mths\_since\_recent\_inq,mths\_since\_recent\_revol\_delinq,num\_accts\_ever\_120\_pd,num\_actv\_bc\_tl,num\_actv\_rev\_tl,num\_bc\_sats,num\_bc\_tl,num\_il\_tl,num\_op\_rev\_tl,num\_rev\_accts,num\_rev\_tl\_bal\_gt\_o,num\_sats,num\_tl\_120dpd\_2m,num\_tl\_30dpd,num\_tl\_90g\_dpd\_24m,num\_tl\_op\_past\_12m,pct\_tl\_nvr\_dlq,percent\_bc\_gt\_75,pub\_rec\_bankruptcies,tax\_liens,tot\_hi\_cred\_lim,total\_bal\_ex\_mort,total\_bc\_limit,total\_il\_high\_credit\_limit

# Data Cleaning:

## Dataset Columns with >50% null Values:

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

## Dataset Columns with Unnecessary Values:

```
'member_id', 'emp_title', 'pymnt_plan', 'url', 'desc', 'zip_code', 'delinq_2yrs', 'application_type'
```

- As these dataset columns are of no use, we can remove, it will not help the analysis rather degrades the quality of datasets.

## Observation:

Now, the final number of columns to analyze is: 46

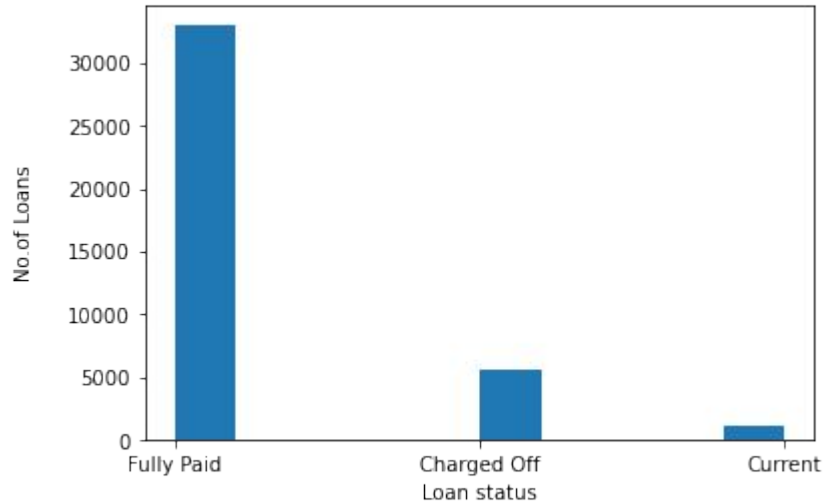
- Columns >50% null values are removed
- Unnecessary columns are removed
- Rows with duplicates are removed(if any)

## Final Columns:

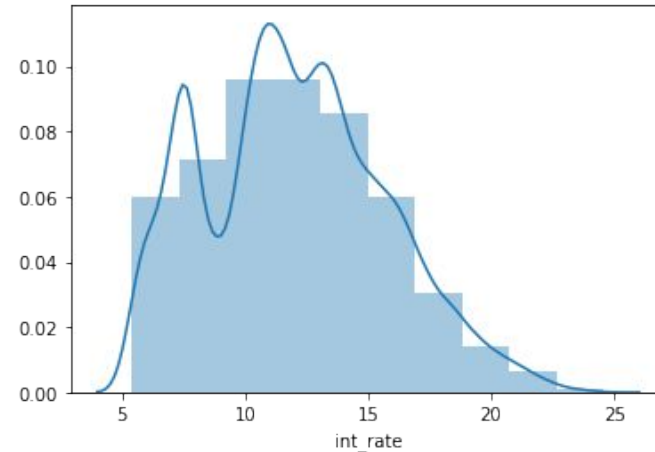
```
['id', 'loan_amnt', 'funded_amnt', 'funded_amnt_inv', 'term', 'int_rate', 'installment', 'grade', 'sub_grade',  
'emp_length', 'home_ownership', 'annual_inc', 'verification_status', 'issue_d', 'loan_status', 'purpose', 'title',  
'addr_state', 'dti', 'earliest_cr_line', 'inq_last_6mths', 'open_acc', 'pub_rec', 'revol_bal', 'revol_util', 'total_acc',  
'initial_list_status', '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', 'collections_12_mths_ex_med', 'policy_code', 'acc_now_delinq', 'chargeoff_within_12_mths',  
'delinq_amnt', 'pub_rec_bankruptcies', 'tax_liens', 'year', 'loan_amnt_bins', 'annual_inc_bins', 'int_rate_group']
```

# Analysis: Univariate

No.of Variables: ["loan\_status"],["int\_rate"],["loan\_amnt"],["Grade"],["purpose"]



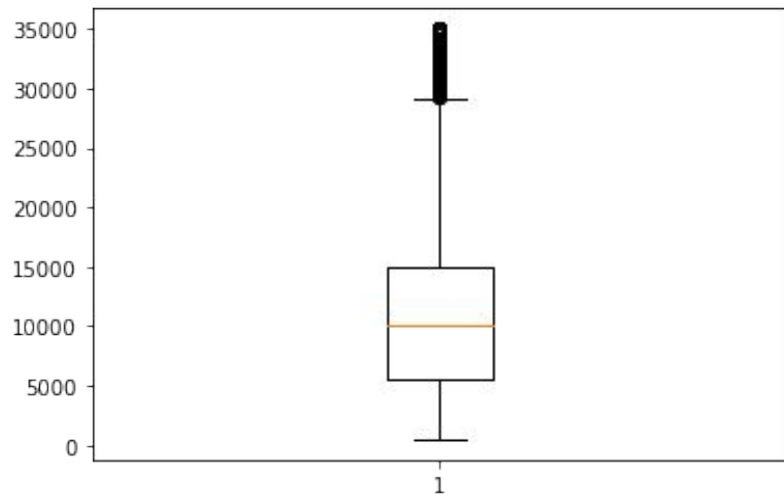
Loan\_status: break down/categories present



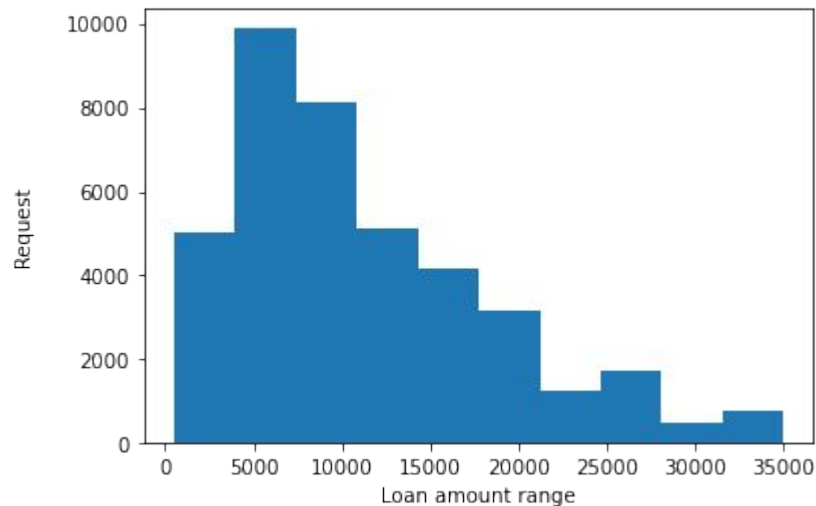
Int\_rate: 10-15% being high rate of int



# Analysis: Univariate

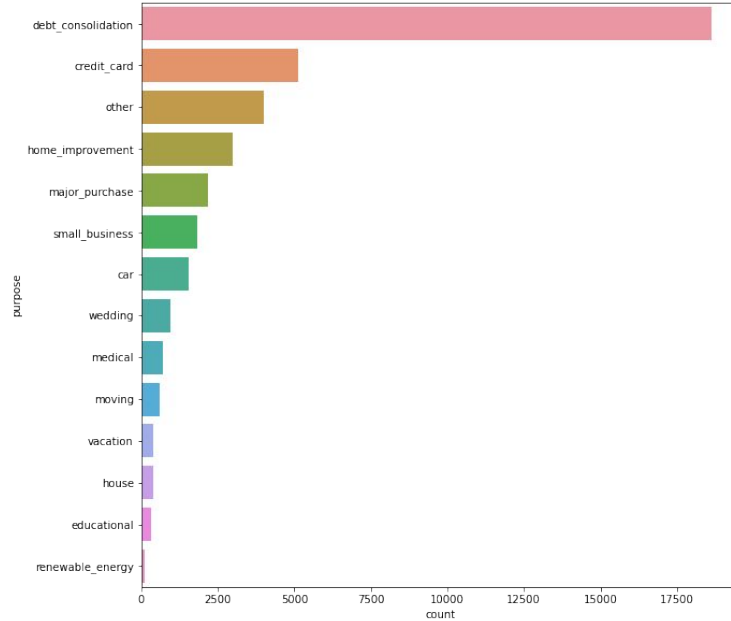


Loan\_amnt: Identification of outliers using box plot in loan amount

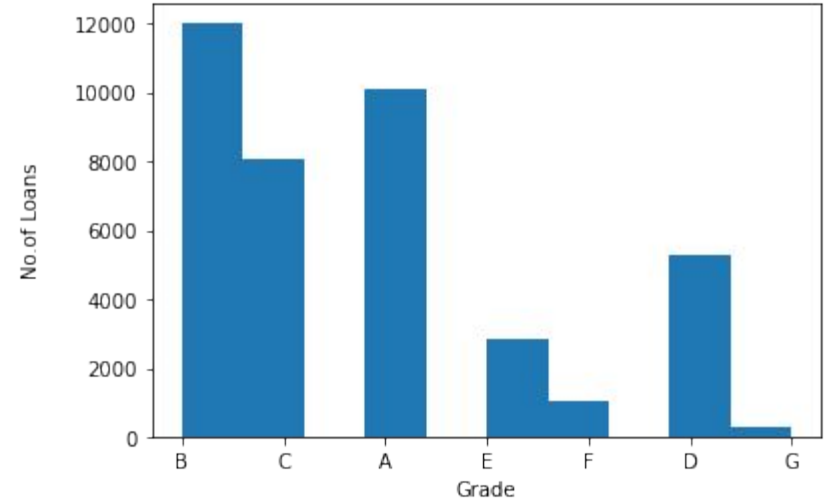


Loan\_amnt: Requested loan amount ranges high from 5000-15000

# Analysis: Univariate



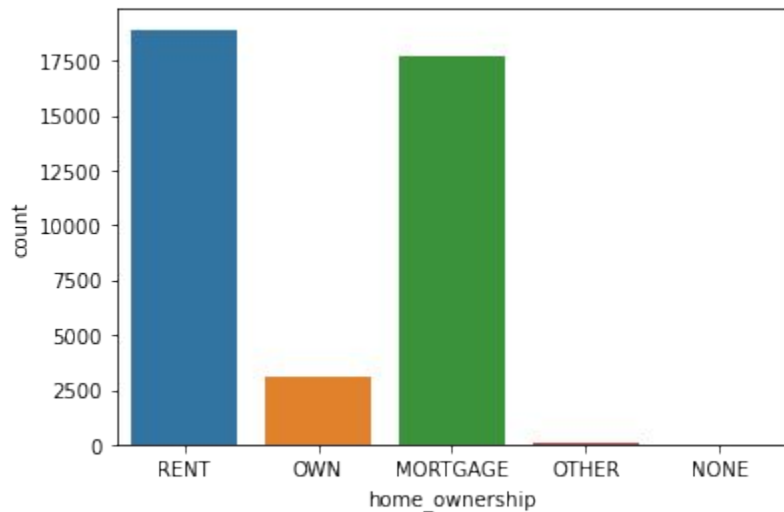
Purpose: Mostly debt consolidation and credit card being the reason for loan request



Grade: Grade B has more number of loans than others

# Analysis: Bivariate

Target Variables: ["home\_ownership"]

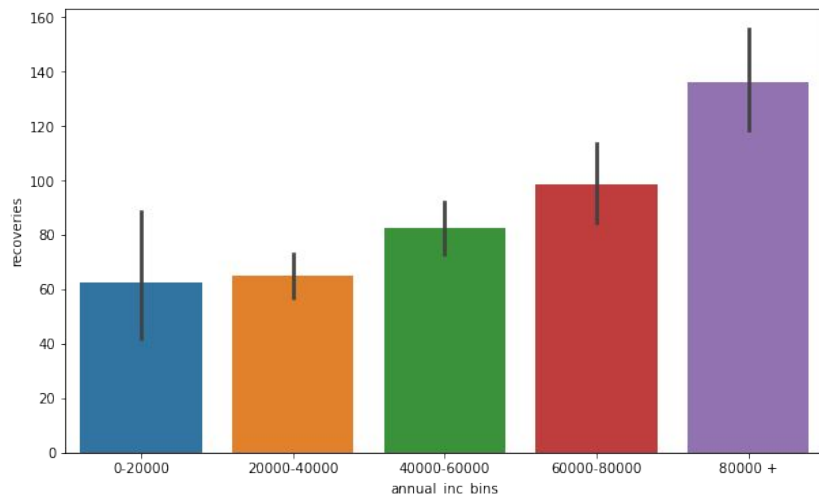


## Observation:

- Fully Paid - 32950, Charged Off - 5627, Current - 1140
  - Out of total loan sanction, 14% are likely to default the loan
- Max int.rate taken ranged from 10-15%
- People rarely go for >30,000 loan amount, and mostly take loan between 5000-15000
- Taken loans are mostly for Debt consolidation and to pay credit card bills
- Grade B - noticed the highest loan takers of all
- People who's 'home\_ownership' in mortgage and rent seems to take loans often

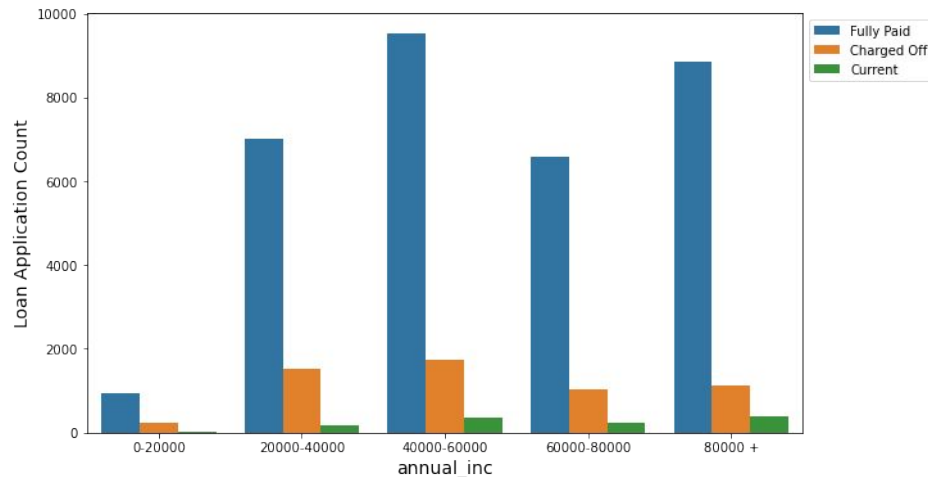
# Analysis: Bivariate

No.of Variables: ['annula\_inc'], ['loan\_amnt']



Recoveries are high where the annual income is high

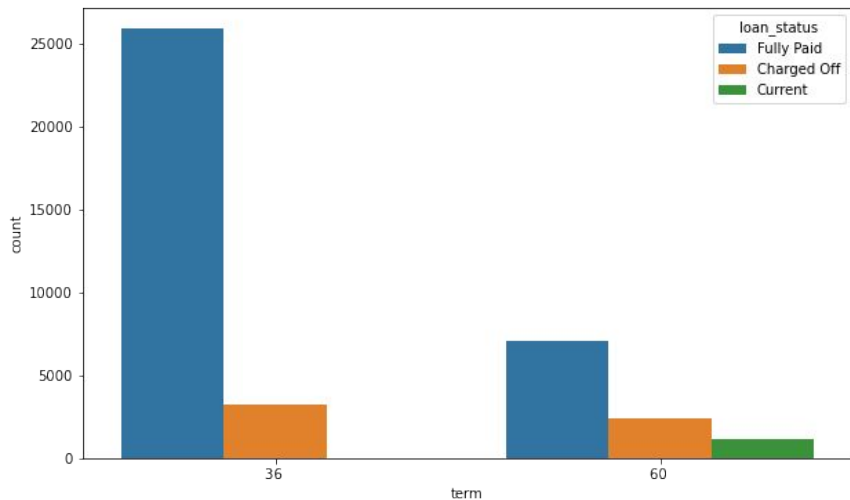
No.of Variables: ['annula\_inc'], ['loan\_Application\_count']



annual income between 20000 to 60000 has more charged off loans

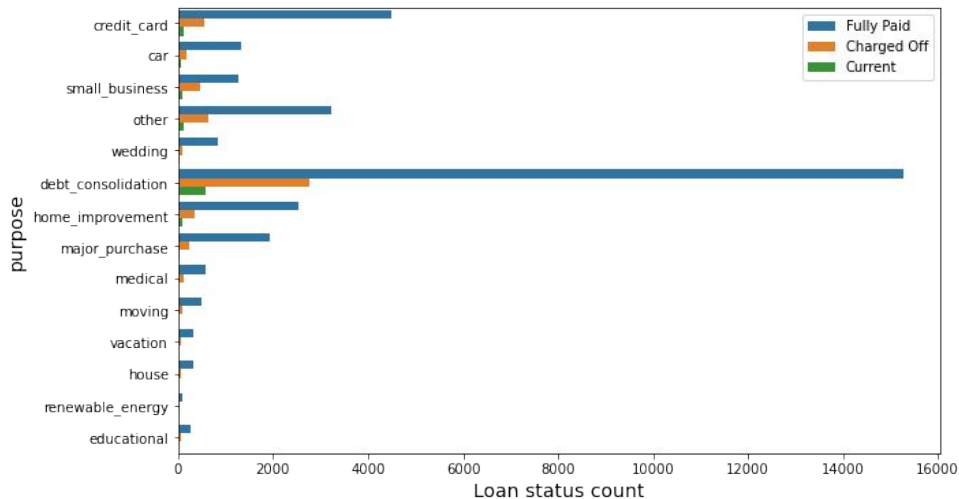
# Analysis: Bivariate

No.of Variables: ["term"],["loan\_status"]



36 month term has more charged off

No.of Variables: ["purpose"],["loan\_status"]



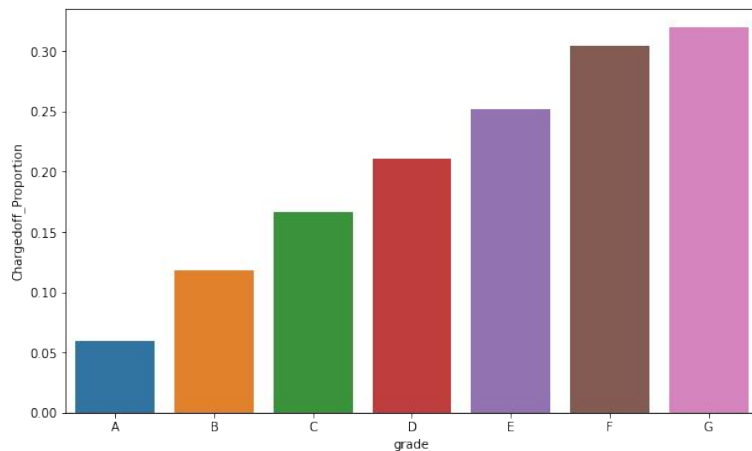
Debt consolidation and credit card purpose loan has the more charged off.

## Observation:

- People who have high annual income most likely to repay the loan
- People likely to default when their annual income ranges between 20,000 to 60,000
- People who opt for 36 month term are most likely to default
- People who take loan for their debt and to pay their credit card are most likely to default

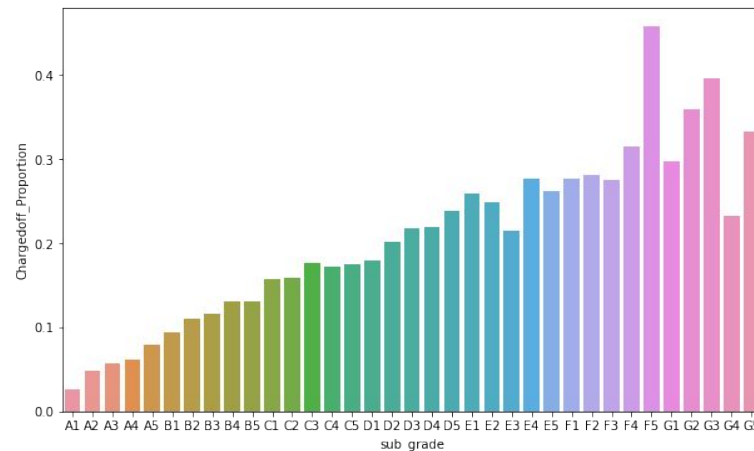
# Analysis: Multivariate

No.of Variables:  
["gradevsloan"], ["chargedoff\_proportion"]



charged off is linearly increased towards the grade A to G

No.of Variables: ["purpose"], ["loan\_status"]

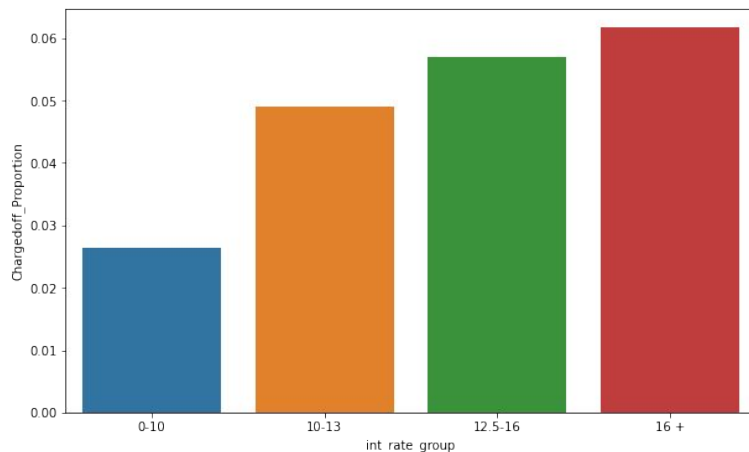


F and G has high charged off values



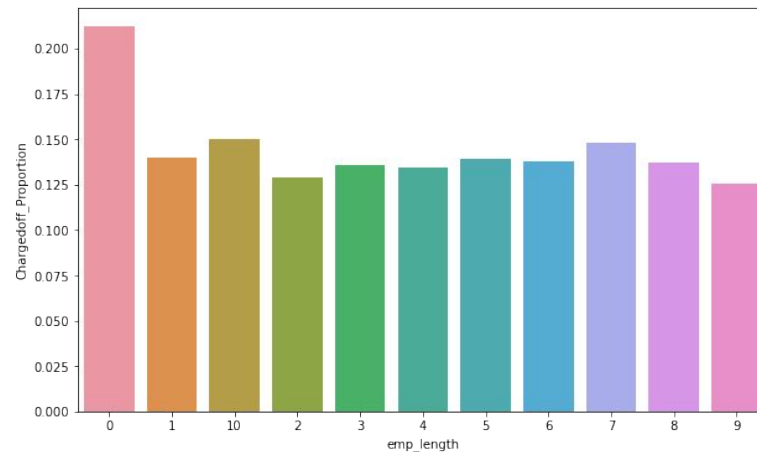
# Analysis: Multivariate

No.of Variables: ["term"], ["loan\_status"]



Charged off is very high when the interest rate is high.

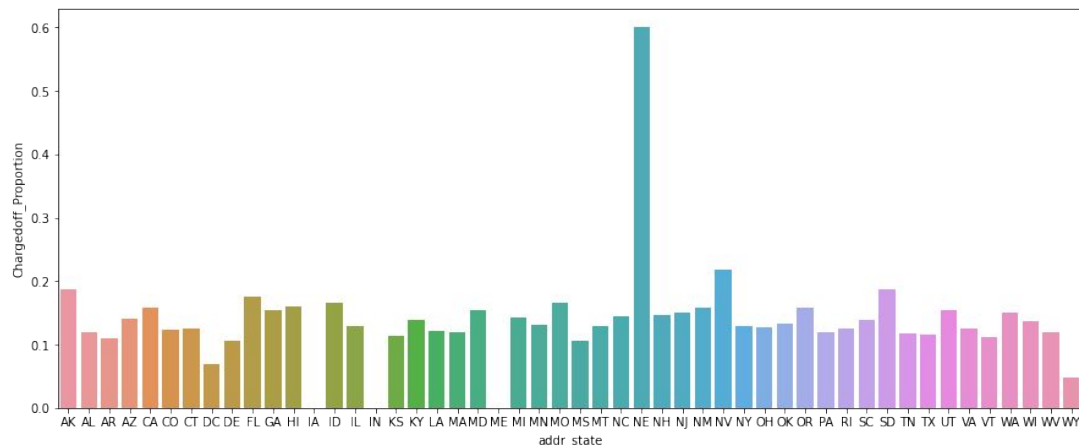
No.of Variables: ["purpose"], ["loan\_status"]



Charged off is high when there is no employment.

# Analysis: Multivariate

No.of Variables: ["term"],["loan\_status"]

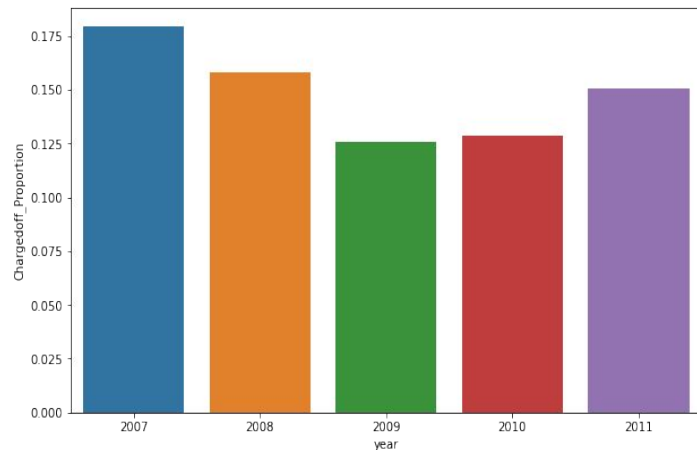


-NE looks high charged off values however it has very low total loan application.

-NV, FL and AK has more application and charged off also high relatively.

-These state needs additional review to approve the loan

No.of Variables: ["purpose"],["loan\_status"]

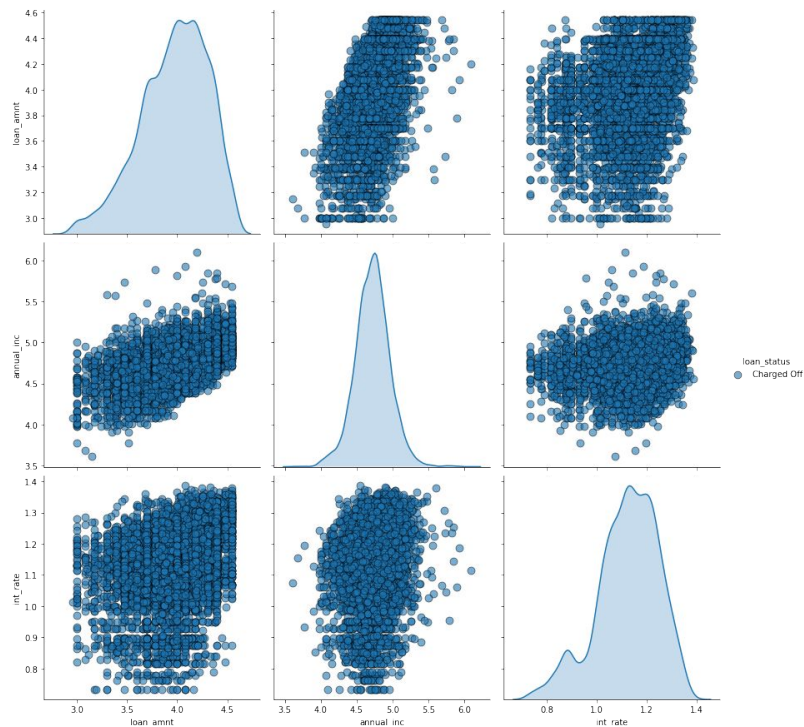


-Even though the charged off proportion is high in 2007 the total loan application is very less.

-Comparing to other years 2011 has more loan application and charged off also very high

# Analysis: Multivariate-Stack

No. of Variables: ["chargedoff\_proportion"] on ["loan\_amt"], ["annual\_inc"], ["int\_rate"]



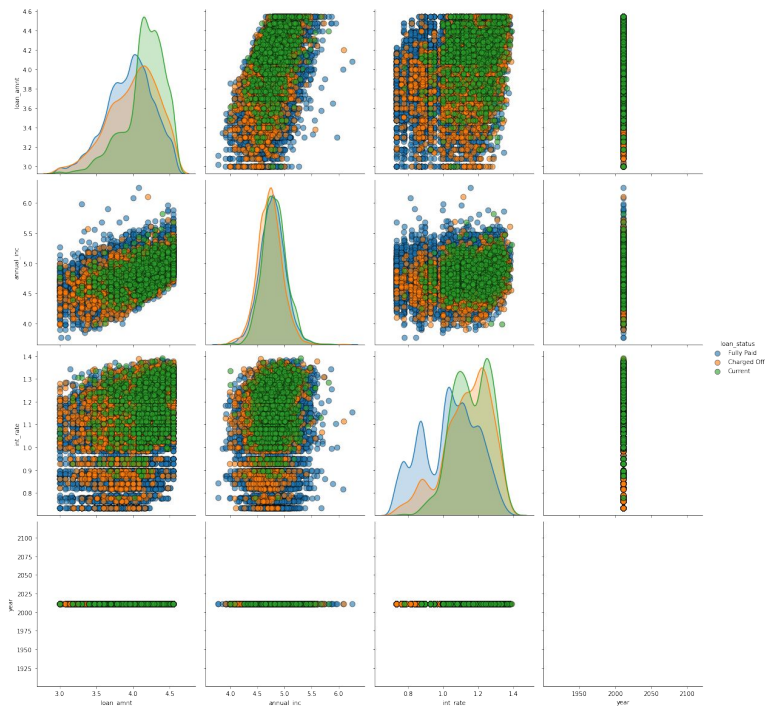
Annual income positively correlates with loan amount

Interest rate positively correlates with loan amount

Interest rate positively correlated with charge off proportion

# Analysis: Multivariate-Stack

No.of Variables:[“loan\_status”] on [“loan\_amt”],[“annual\_inc”],[“int\_rate”],[“year”]



Interest rate is increase when the loan amount is increase

Higher charged off when there is an interest rate increase

# Conclusion:

- People with bad credit history (Grade - G) most likely to default
  - Suggestion, Loan amount slab can be gradually reduced as per the grade
- People who fall in Grade F & G are most likely to default due to their bad credit score
- People default when they are not employed
- People default when the interest is high
- People from NE most likely to default
  - Considering the population being less, recommend to have lowest interest rates for them
- People from NV, FL and AK most likely to default
- 2008 - where charged off was high (may be due to any financial bane)
- People with lower income below 60,000 applying loan with higher amount with higher interest are most likely to default

Driving factors: ["purpose"], ["int\_rate"], ["annual\_inc"], ["funded\_amnt"], ["term"], ["grade"], ["addr\_state"]