

EDA CASE STUDY

SUBMISSION

Group Submission

- Niranjan Ghadei
- Sandeep Nair
- Priyanka Navgire
- Kuldip Singh Chauhan

Gramener EDA Case Study: Credit Decision.

Lending Club is consumer finance company which specialises in lending various types of loans to urban 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 **types of risks** are associated with the bank's decision

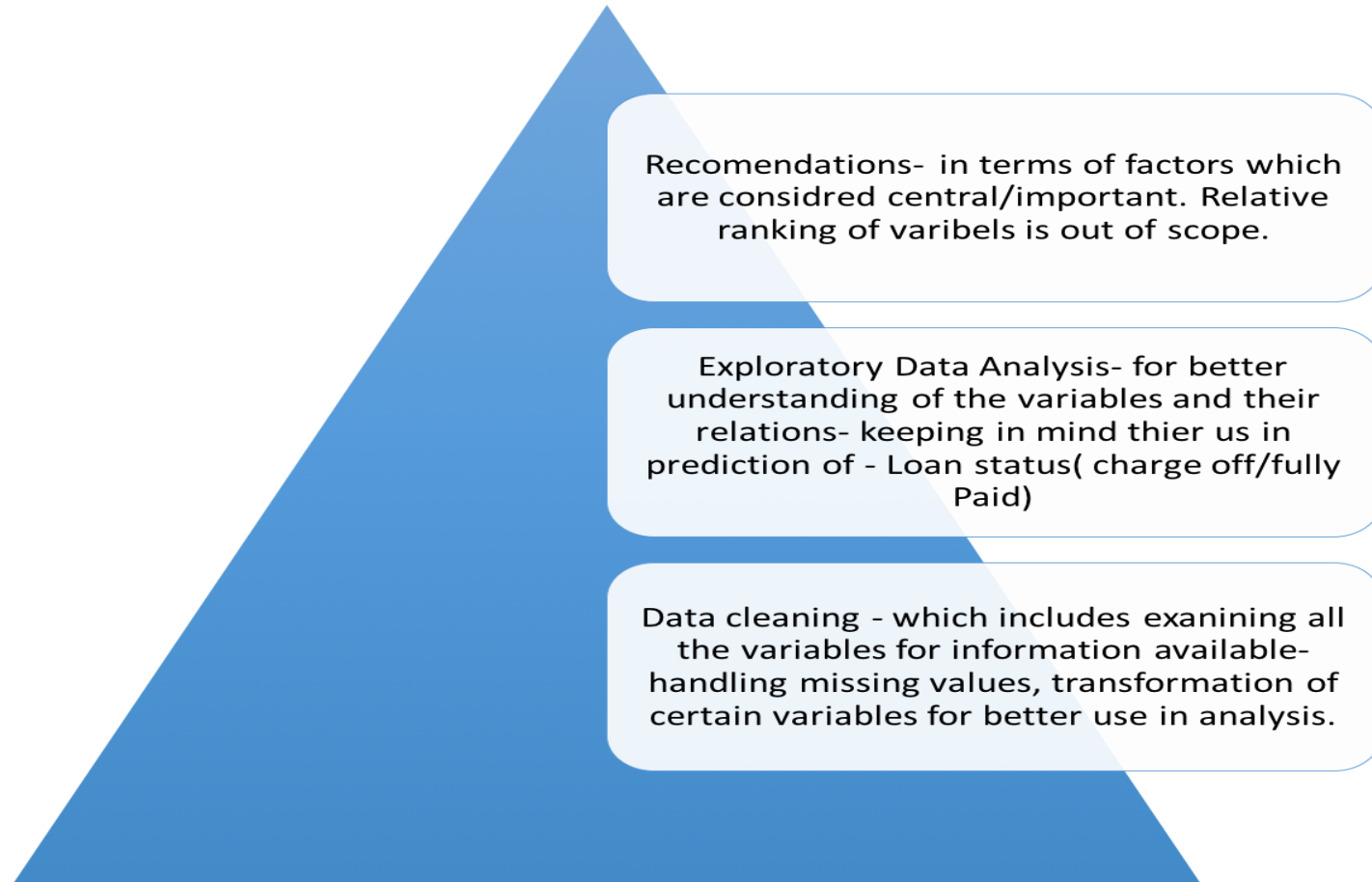
- Loss of business: through rejecting loan which should have been granted
- Credit Loss : through charge off/default from applications which should have been rejected

While there would be slippage in terms of credit loss – same needs to be governed keeping in mind the overall impact on business- very stringent norms can lead to lot of rejections and hence loss of business while loose norms can lead to high credit loss

Constraints

- We don't have reject data – which is also indicator of already in place criteria
- Lot of data with no meaningful use - >50% data is NA
- Lot of operational variables with little analytical use such as repeat identifiers, post delinquency variables like fee collected/principal collected/penal charges collected.

Problem solving methodology



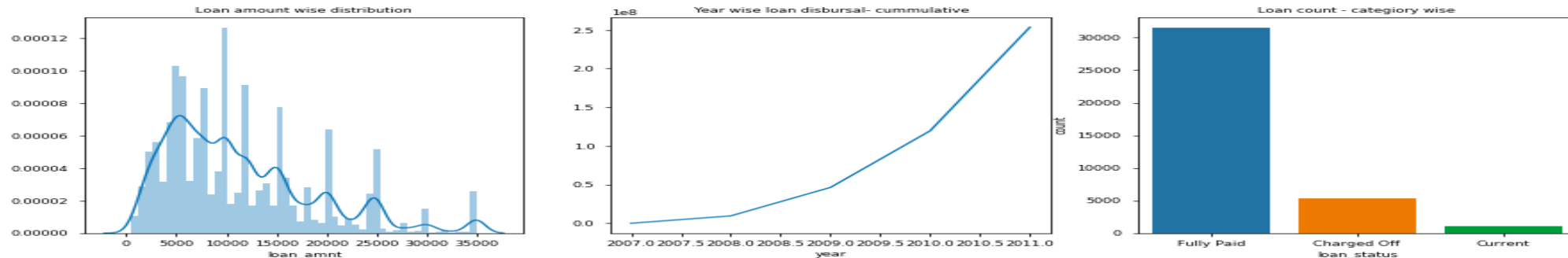
Data Cleaning and Missing value Handling.

1. 54 Columns with only NA as values- all deleted.
2. 37 variables identified where the utility of the variable is considered less or has values which are either one.
3. Missing values in rows since are low are removed.
4. Some of the variables are transformed for better use
 - Interest Rate
 - Utilisation of revolving facility.
 - Loan term

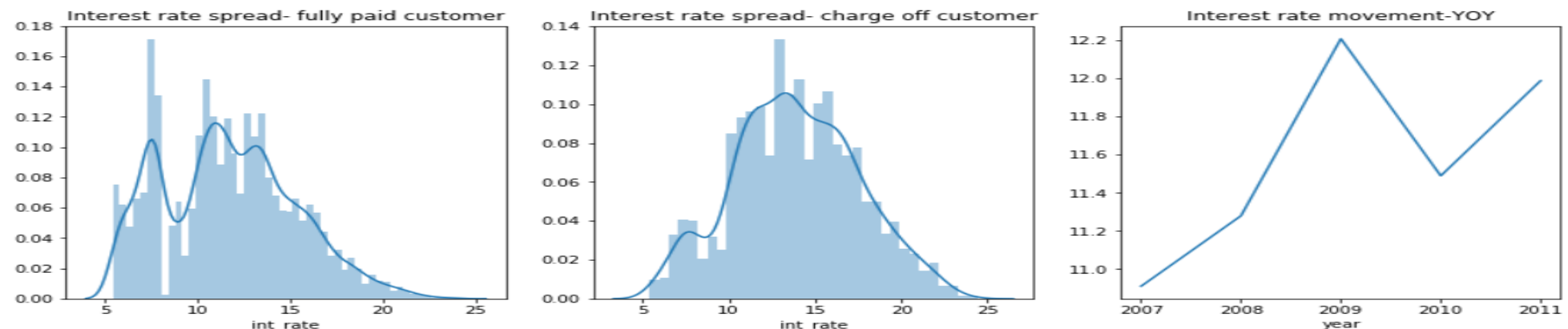
Univariate Analysis

Univariate Analysis

- **Loan Amount Distribution** - Loan amount majorly varies from 5600 to 15000 with median of 10000. Count plot for YOY loan disbursal- there is exponential growth in Loan disbursement.

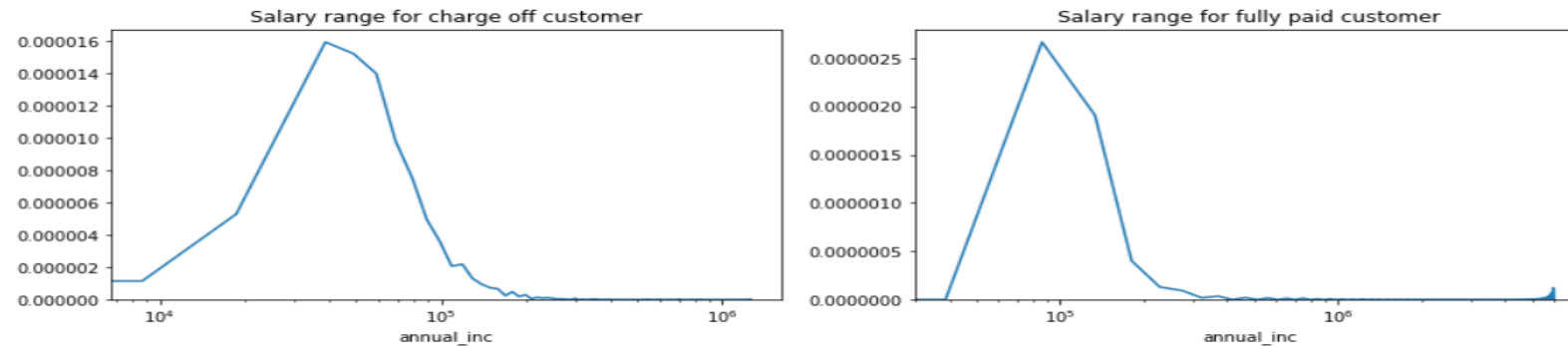


- **Interest Rate Distribution** - Pricing of the company seems right - given higher pricing for risky loans. Third plot is interest rate curve for time period - 2007 - 2011 for loan portfolio- with peak seen in 2009

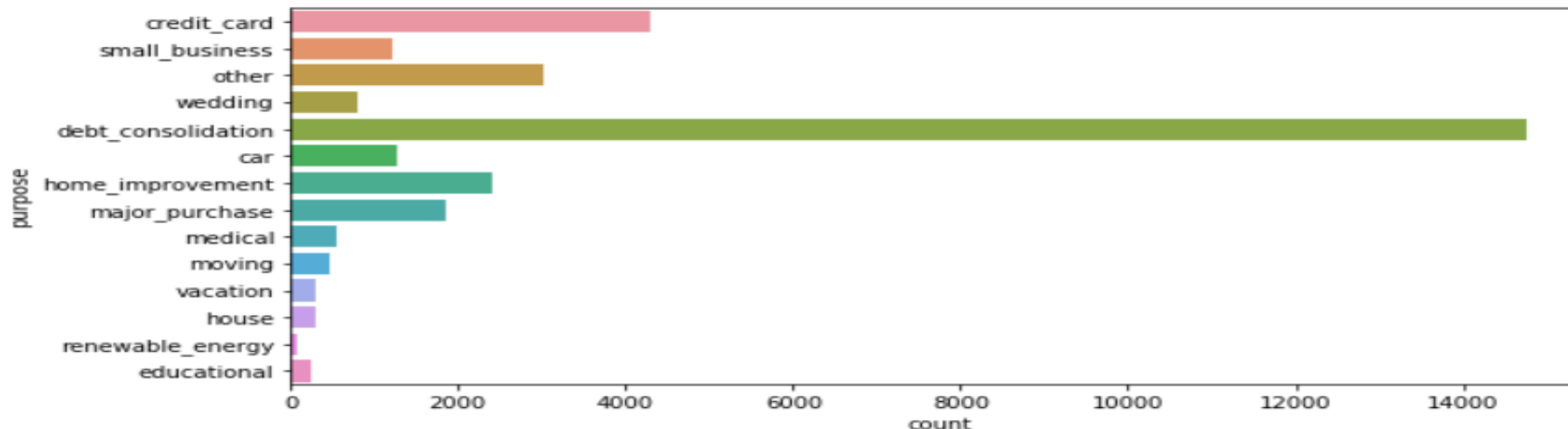


Univariate Analysis

- **Annual Income**- Annual income centers around USD 50K for charged off customer where as for fully paid same is around USD 100K.

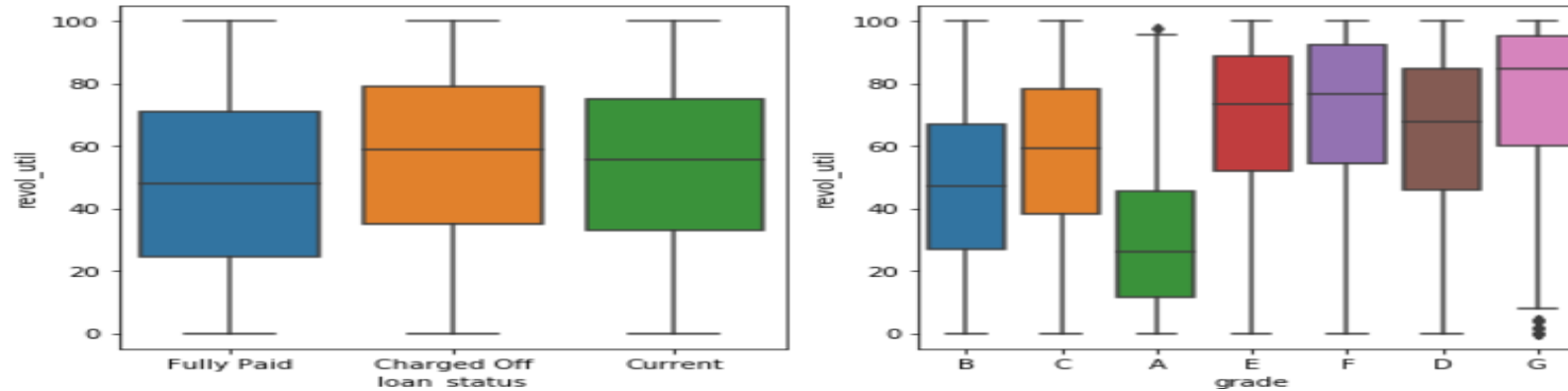


- **Loan Purpose** -Around 47% of the loans are distributed towards loan consolidations, followed by payments of credit cards. Given the risky loans - the interest rates are higher.

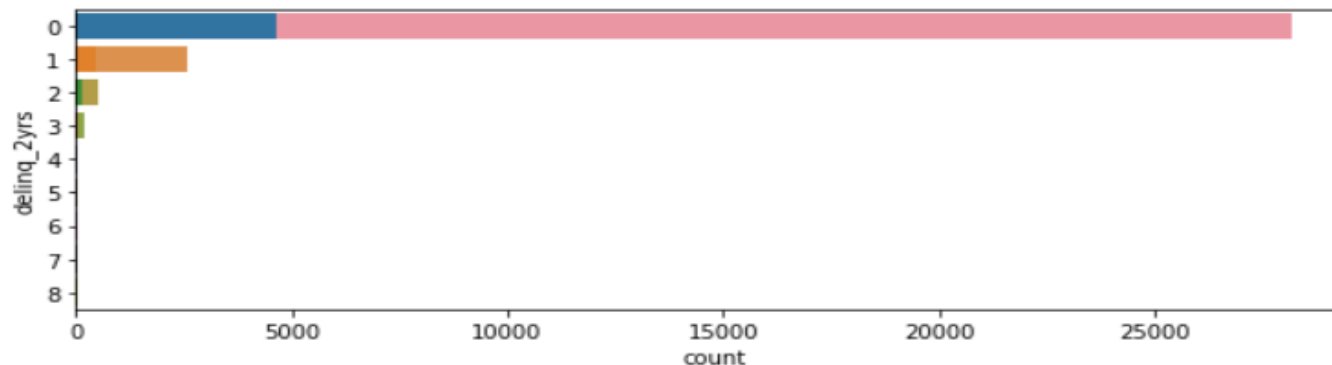


Bivariate Analysis

Loan Status V/S Revolving utilization-Loan status and percentage utilization of limits seem to be related ,given higher percentage utilization in charged off population. Internal Grading for the finance company further shows same by allocating lower grades to population with higher utilization.



Loan Status and Delinquency_2yrs-Delinquency within 2 yrs/ or late payments is a critical factor for loan approval. Number of loans approved decrease exponentially with increase in late payments. Also the approved loans appears to be more risky. **Loan Status V/S Public records of Bankruptcy**- Number of public records appears to be a criteria factor for loan appraisal- given drastic drop in the mentioned categories. Number of “charge off“ is very high with instances of public derogatory.

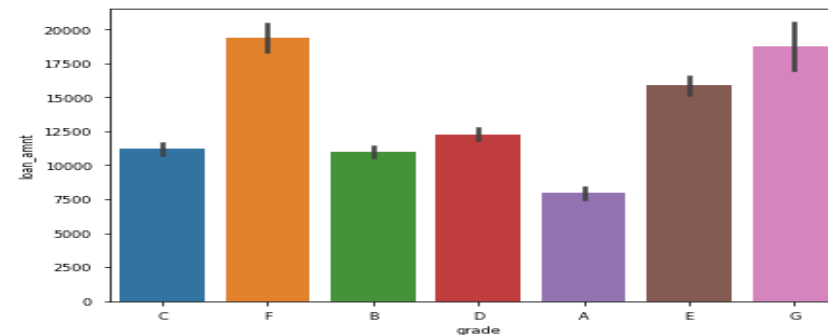
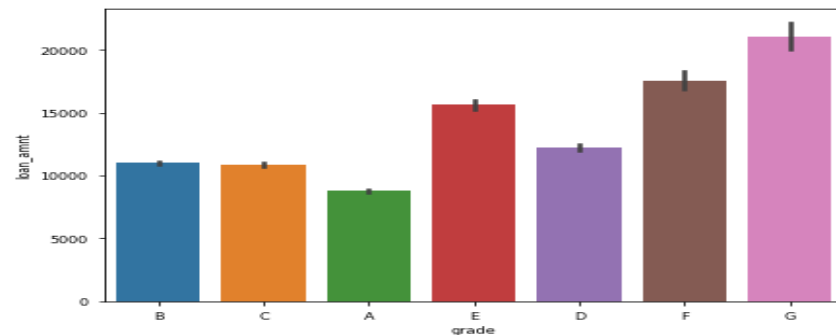


- **Loan Status vs Purpose for all loan status- Small business** shows **30% charged** off loans which is maximum among all categories

loan_status	Charged Off	Current	Fully Paid
purpose			
car	151	49	1278
credit_card	501	101	4298
debt_consolidation	2623	567	14750
educational	49	0	247
home_improvement	319	94	2415
house	58	14	291
major_purchase	206	35	1849
medical	99	12	546
moving	81	7	460
other	578	120	3017
renewable_energy	18	1	76
small_business	444	73	1212
vacation	50	4	291
wedding	89	21	804

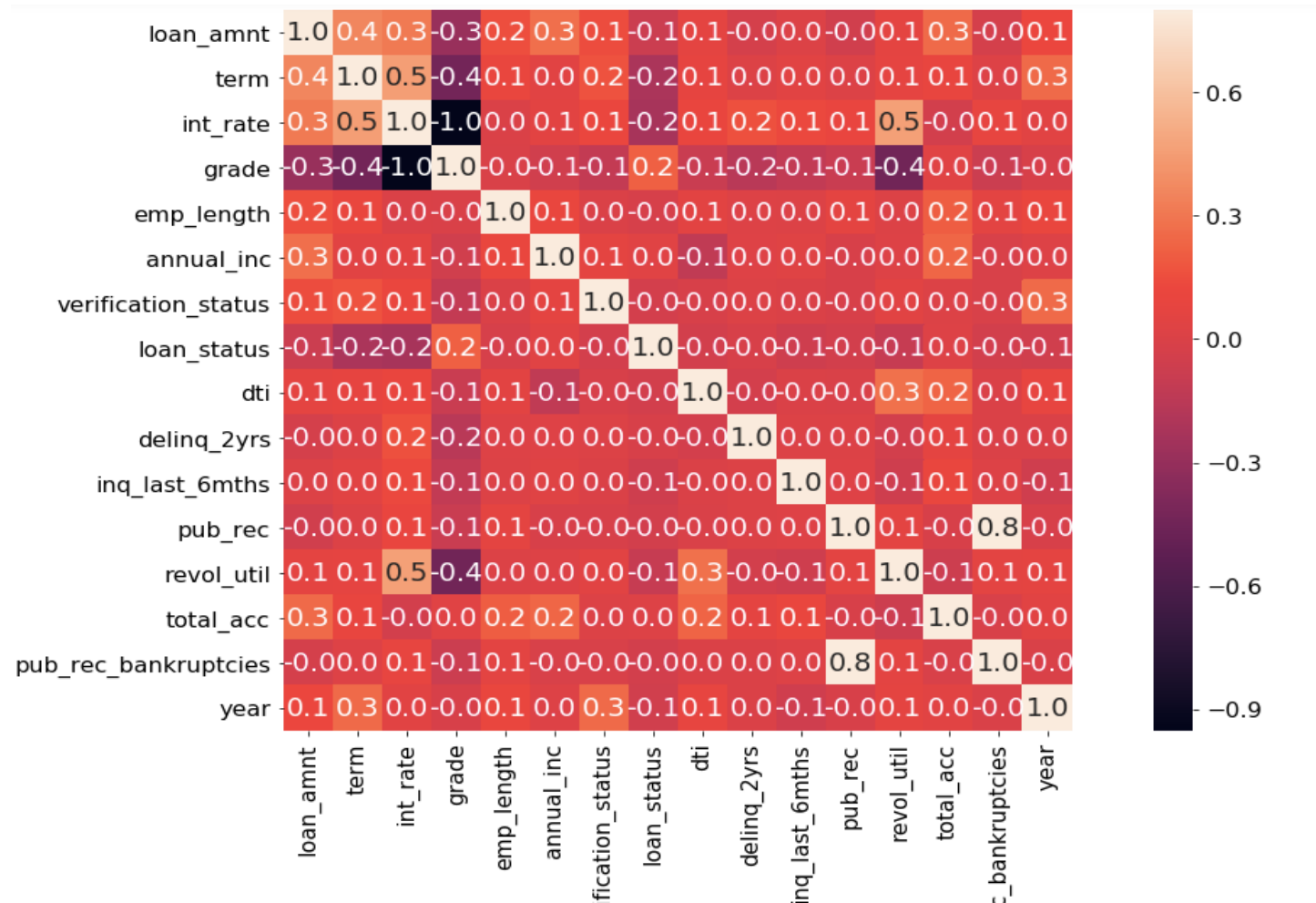
Loan Status vs Grade Relation—Loan status and Grade has a direct relationship with higher charged off observed in lower grades. The loan portfolio shows higher concentration in lower graded loans. (Targeting of high risk customers looks like business strategy - given Higher interest rate opportunity

grade	A	B	C	D	E	F	G
loan_status							
Charged Off	552	1326	1251	1049	686	304	98
Current	35	327	257	216	175	71	17
Fully Paid	8966	9823	6223	3803	1888	635	196



Correlation HeatMap

Correlation indicating pair wise comparison



CONCLUSION

1. Consumer lending company is targeting higher risk borrower – given higher concentration in lower grades and higher Interest rates.
2. Small business shows 30% charged off loans which is maximum among all categories.
3. Average loans size has been increasing over years.
4. There seems to be concentration of charge off in lower annual income category – however provided debt to income ratio doesn't give conclusive evidence of same.
5. Pairwise correlation heat map doesn't provide conclusive insights which could be used as criteria
6. Five major Indicators of default
 - **Grade** - Loan type with debt consolidation and credit card payment confirms trend of targeting high risk borrowers – hence strong relationship between grade and Interest rate- where grade is lower for high risk customers.
 - **Delinquency within 2 yrs, Public records and Public records of bankruptcy**- are three critical factors for loan approval. Number of loans approved decrease exponentially with increase in late payments/bankruptcy instances.
 - **Utilization of revolving limits** is directly linked with lower grades – and are generally higher in charge off loans.