

LENDING CLUB CASE STUDY

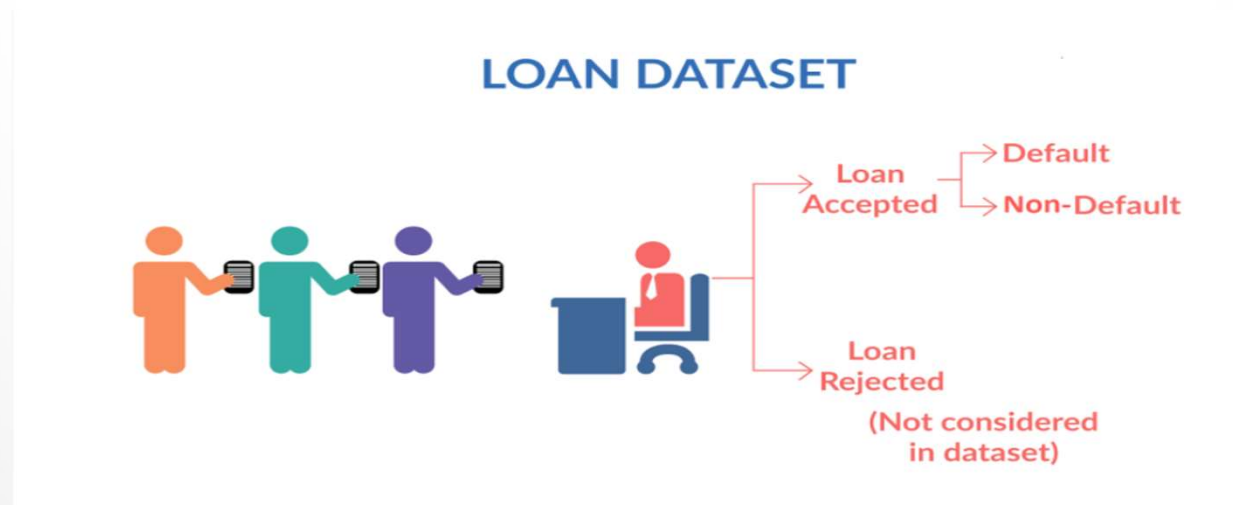
Submission

By - Divya Jayendrakumar Sodha



Problem Statement

- Lending club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures at lower interest rate loans through a fast online interface.
- Two types of risks are associated with any company's decision of lending loan
 - 1) If the applicant is likely to repay the loan, then not approving the loan results in 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(credit loss) for the company.
- Aim of this case study is to understand the driving factors (or driver variables) behind loan default (2nd type of risk), i.e. the variables which are strong indicators of default. Thus, the company can utilize this knowledge for its portfolio and risk assessment.



ASSUMPTIONS

- If the company approves the loan, there are 3 possible scenarios described below:
 - 1) *Fully paid*
 - 2) *Current*: Applicant is in the process of paying the instalments
 - 3) *Charged-off*: Applicant has not paid the instalments in due time for a long period of time
- On the basis of the understanding of problem statement and the data, assumptions are:
 - 1) All the rows with loan status as “Current” are to be removed

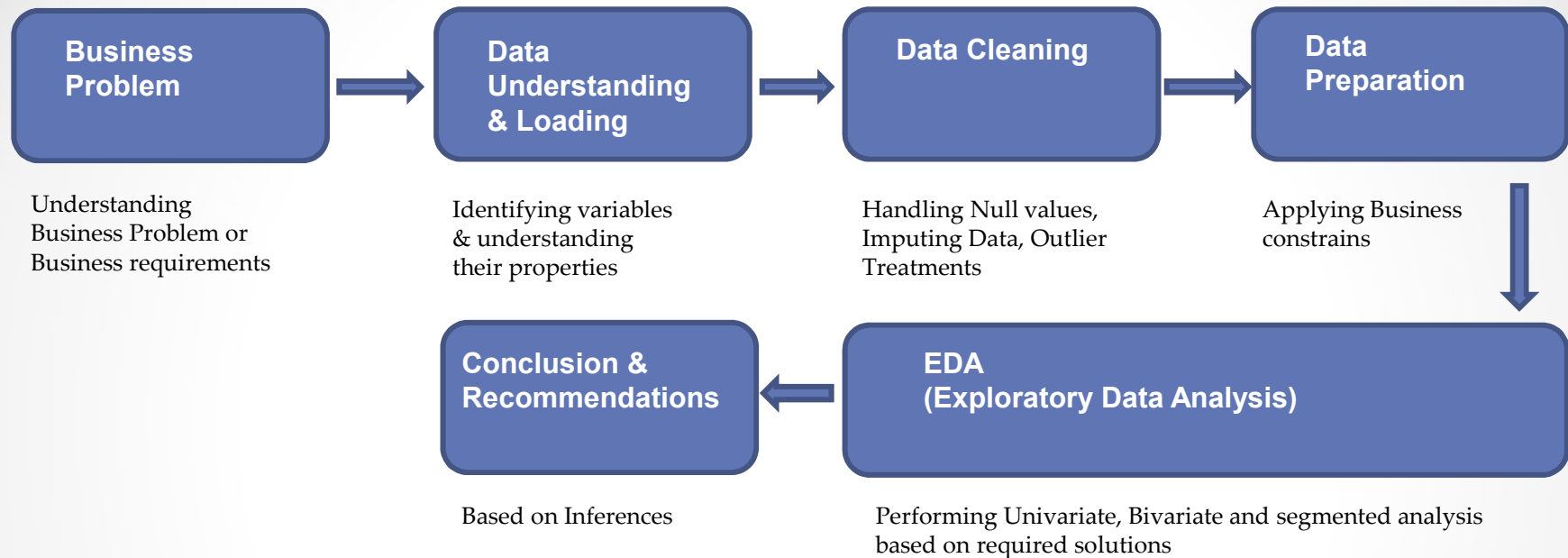
Reason – They are still on process, thus we can’t predict on the basis of that.
 - 2) All the columns with more than 50% of null values are to be removed.

Reason – Columns with higher percentages of null values, doesn’t lead to good insights.
 - 3) All the insights are drawn, taking into account the problem statement, data and the data dictionary provided. No assumptions are made on data apart from them.

Reason – We should rely on the information provided first for data understanding and should not make any assumptions randomly.

APPROACH

- The common approach followed in any EDA is given below:-



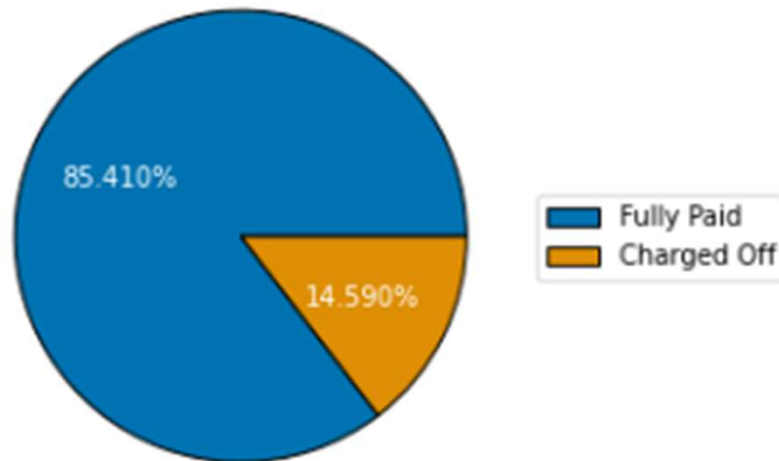
- Particularly, for this case study, mentioning approaches for Data Cleaning step:-

- 1) Dropped unnecessary Columns based on
 - a. Columns with id or text values.
 - b. Columns with more than 50% null values
 - c. Columns which are single valued or contains just null values.
 - d. Columns which are related to post approval feature and not useful now
 - e. Repeatative columns based on correlation.
- 2) Standardized values for "int_rate" and "emp_length"
- 3) Filled Null values with 'Unknown' for "emp_length"
- 4) Derived new columns "issue_month" and "issue_year" based on "issue_d" and formatted them as per requirement.
- 5) Removed outliers for "annual_inc" beyond 95 percentile.

ANALYSIS OF TARGET VARIABLE

- After Data Cleaning and Pre-Processing, we were left with 38445 rows and 18 columns
- Our target column is loan status
- Since, we have already filtered out current loan status, we are left with only charged off and fully paid loan status

LOAN STATUS DISTRIBUTION
(after removing "Current")

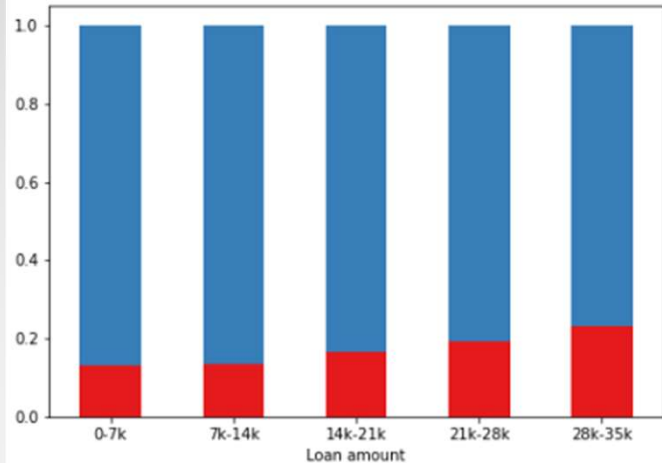


Count of Loan Status

loan_status	
Fully Paid	32836
Charged Off	5609

ANALYSIS ON LOAN AMOUNT & INTEREST RATE

COMPARISION OF LOAN STATUS WITH LOAN AMOUNT



loan_amnt_cats	0-7k	1883.00	12395.00	0.13
	7k-14k	1784.00	11488.00	0.13
	14k-21k	1149.00	5866.00	0.16
	21k-28k	534.00	2232.00	0.19
	28k-35k	259.00	855.00	0.23
		Charged Off	Fully Paid	Charged Off Ratio

loan_status

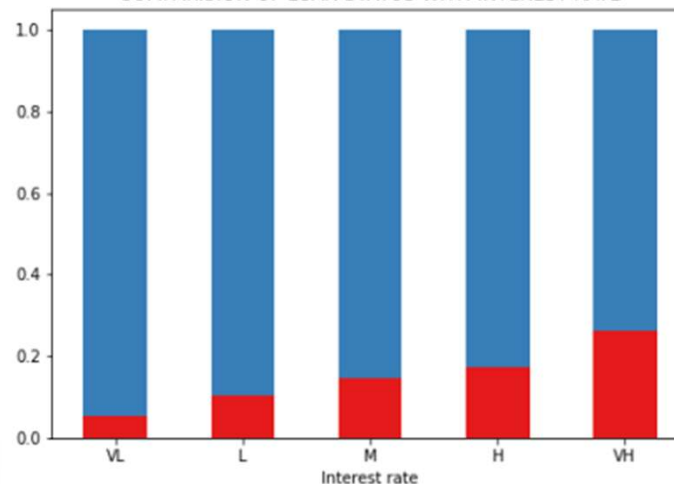
LOAN AMOUNT WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- Higher the loan amount, more likely the customers are to be defaulted

INTEREST RATE WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- At higher interest rate, customers are more likely to be defaulted

COMPARISION OF LOAN STATUS WITH INTEREST RATE

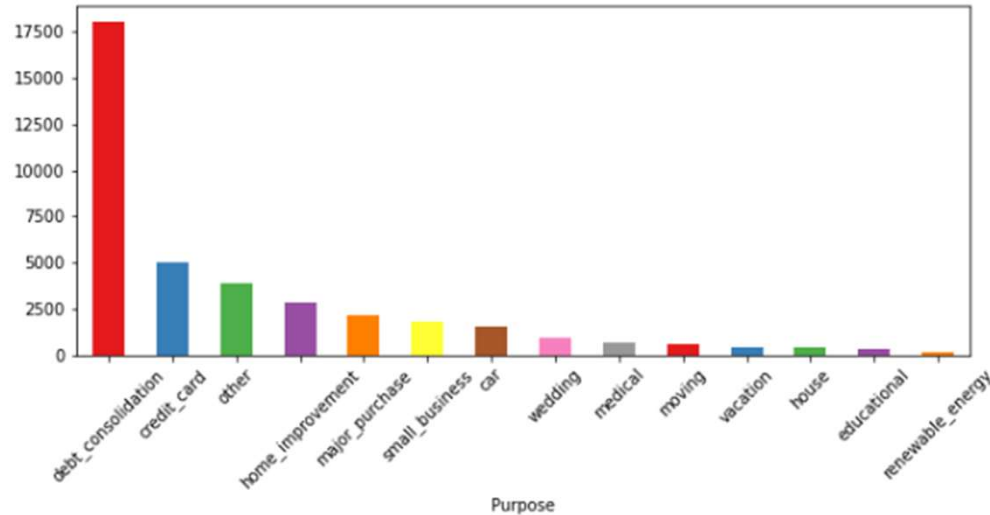


int_rate_cats	VL	429.00	7562.00	0.05
	L	853.00	7435.00	0.10
	M	1016.00	5894.00	0.15
	H	1309.00	6307.00	0.17
	VH	2002.00	5638.00	0.26
		Charged Off	Fully Paid	Charged Off Ratio

loan_status

ANALYSIS ON PURPOSE

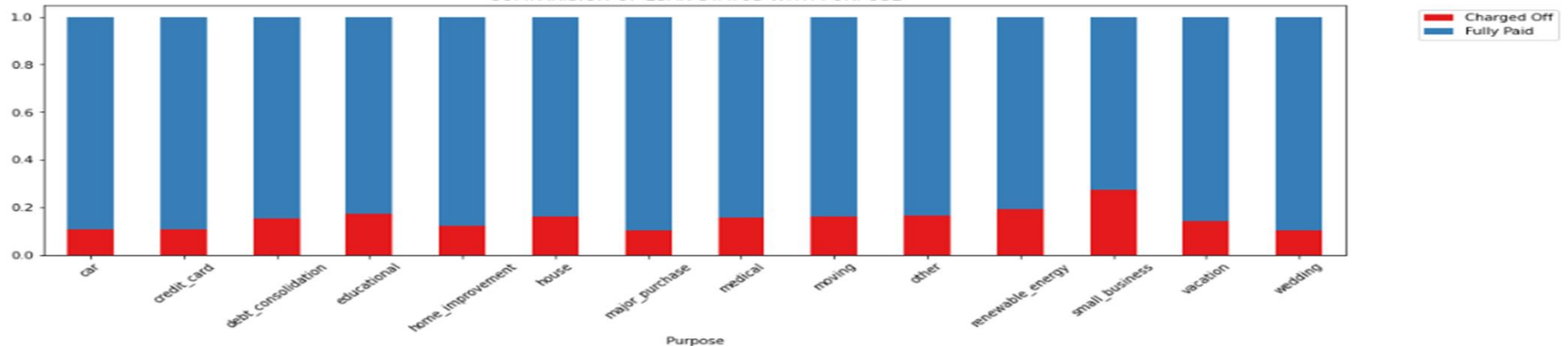
DISTRIBUTION OF LOAN PURPOSE



PURPOSE WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- **Small_business** owners are the highest defaulters, although the total count of debt consolidation is more
- **Renewable energy** and **educational** are also less in number, but are defaulting more

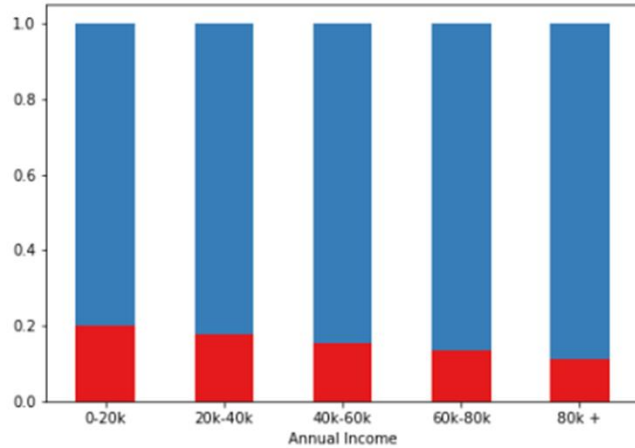
COMPARISON OF LOAN STATUS WITH PURPOSE



loan_status	purpose													
	car	credit_card	debt_consolidation	educational	home_improvement	house	major_purchase	medical	moving	other	renewable_energy	small_business	vacation	wedding
	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off	Charged Off
Charged Off	160.00	542.00	2763.00	56.00	342.00	59.00	220.00	106.00	91.00	630.00	19.00	473.00	53.00	95.00
Fully Paid	1339.00	4475.00	15246.00	268.00	2501.00	305.00	1921.00	572.00	481.00	3225.00	81.00	1273.00	322.00	827.00
Charged Off Ratio	0.11	0.11	0.15	0.17	0.12	0.16	0.10	0.16	0.16	0.16	0.19	0.27	0.14	0.10

ANALYSIS ON ANNUAL INCOME & TERM

COMPARISION OF LOAN STATUS WITH ANNUAL INCOME



annual_inc_cats	Charged Off	Fully Paid	Charged Off Ratio
0-20k	237.00	943.00	0.20
20k-40k	1514.00	7004.00	0.18
40k-60k	1729.00	9534.00	0.15
60k-80k	1024.00	6597.00	0.13
80k +	1105.00	8758.00	0.11

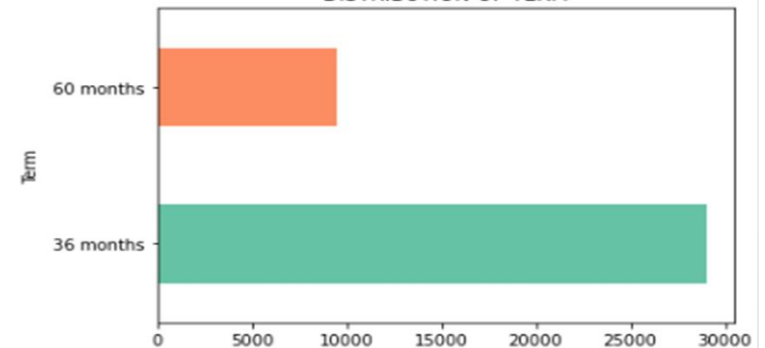
ANNUAL INCOME WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- Applicants with higher income default less

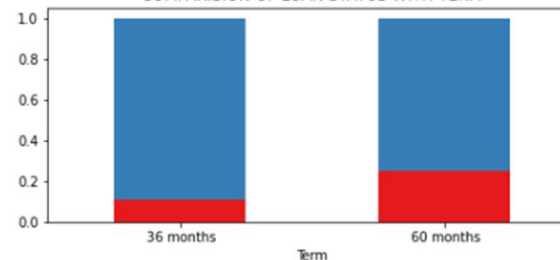
TERM WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- Total count of customers is higher for the term period of 36 months
- But, the rate of defaulters is more in case of 60 months term

DISTRIBUTION OF TERM



COMPARISION OF LOAN STATUS WITH TERM

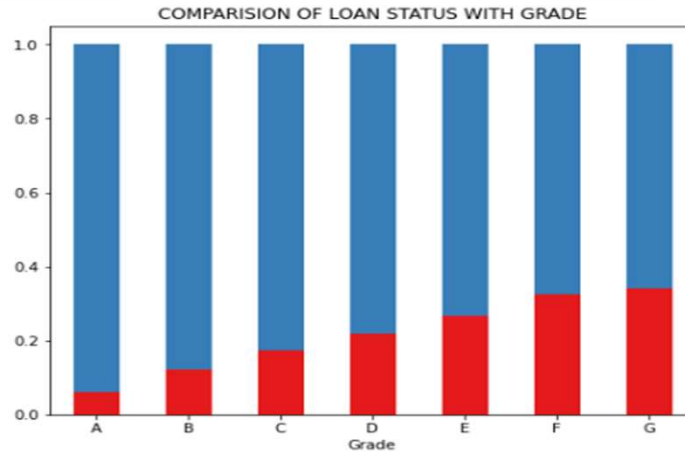


term	Charged Off	Fully Paid	Charged Off Ratio
36 months	3216.00	25783.00	0.11
60 months	2393.00	7053.00	0.25

ANALYSIS ON GRADE

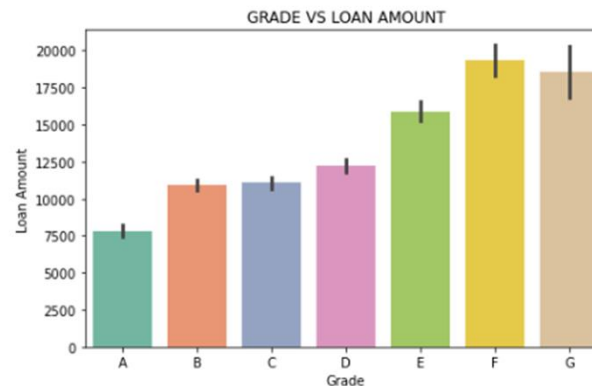
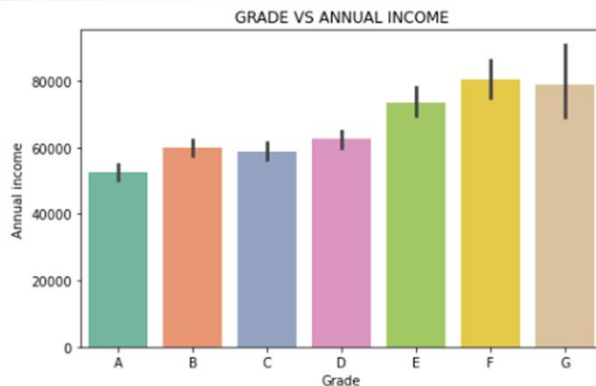
GRADE WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- The rate of defaulters increases from grade A to grade G
- Grade G has highest rate of defaulters



A	602.00	9418.00	0.06
B	1420.00	10217.00	0.12
C	1344.00	6461.00	0.17
D	1116.00	3955.00	0.22
E	710.00	1936.00	0.27
F	317.00	655.00	0.33
G	100.00	194.00	0.34

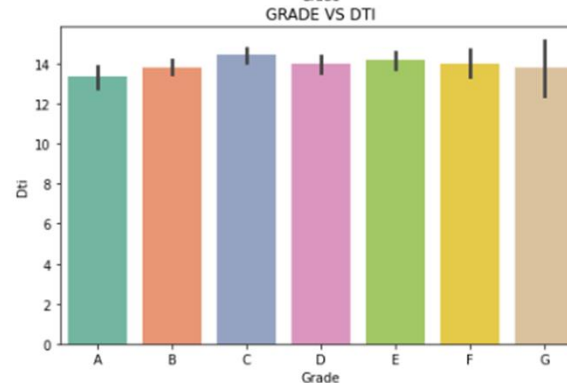
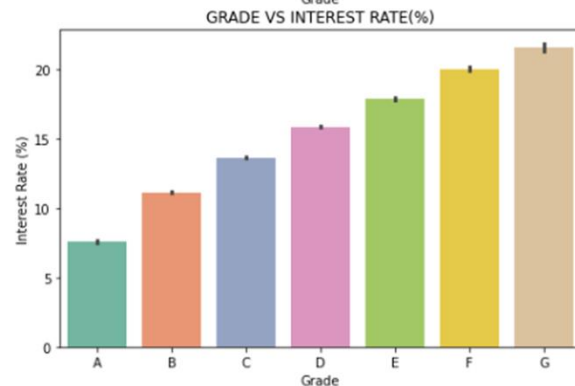
loan_status



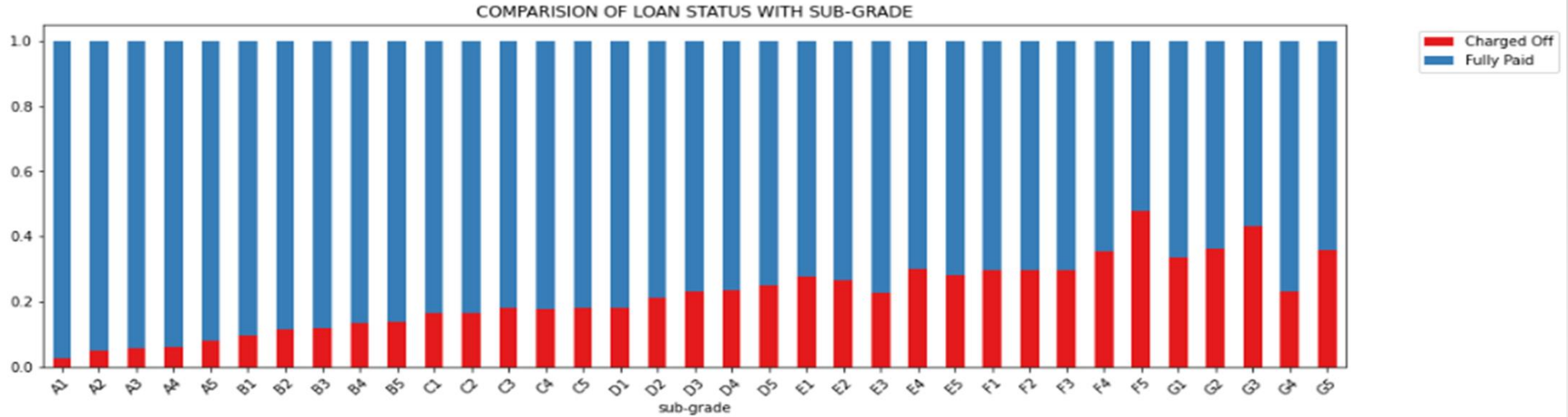
GRADE WITH RESPECT TO OTHER PARAMETERS (FOR CHARGED OFF LOAN STATUS)

- For Grade G, the highest possibility of default is when:-

- 1) Annual Income is between 60k to 80k
- 2) Loan Amount is between 17500 to 20000
- 3) Interest Rate is above 20%



ANALYSIS ON SUB-GRADE & ADDRESS STATE



SUB-GRADE WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- Rate of defaulters gradually increase from lower sub-grades to higher sub-grades
- Applicants with F5, G5, G3, G2 sub-grades are especially risky.

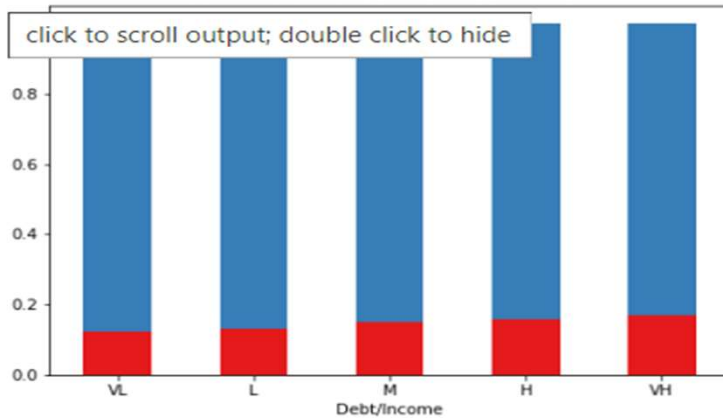


ADDRESS STATE WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- The top 3 states, having highest chances of getting default customers are NE, NV and SD

ANALYSIS ON DTI & EMPLOYEE LENGTH

COMPARISION OF LOAN STATUS WITH DTI



Charged Off
Fully Paid

VL	942.00	6749.00	0.12
L	1015.00	6674.00	0.13
M	1143.00	6554.00	0.15
H	1212.00	6474.00	0.16
VH	1297.00	6385.00	0.17
	Charged Off	Fully Paid	Charged Off Ratio
	loan_status		

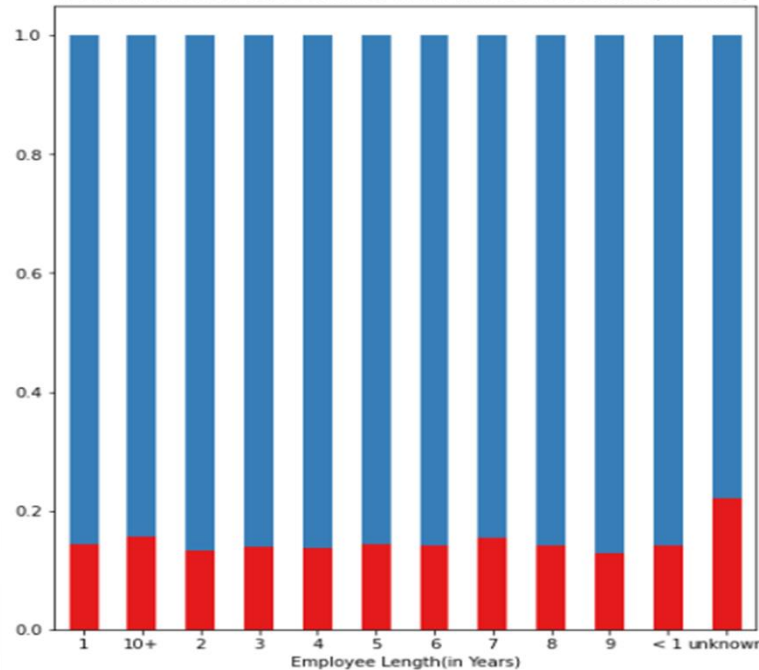
DTI WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- Rate of defaulters gradually increase from lower debt-to-income ratio to higher debt-to-income ratio

EMPLOYEE LENGTH WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- Customers with employment length 10+ years are having higher rate of defaults, if we ignore unknown values
- But, there is no pattern observed for other length of years

COMPARISION OF LOAN STATUS WITH EMPLOYEE LENGTH(in Years)

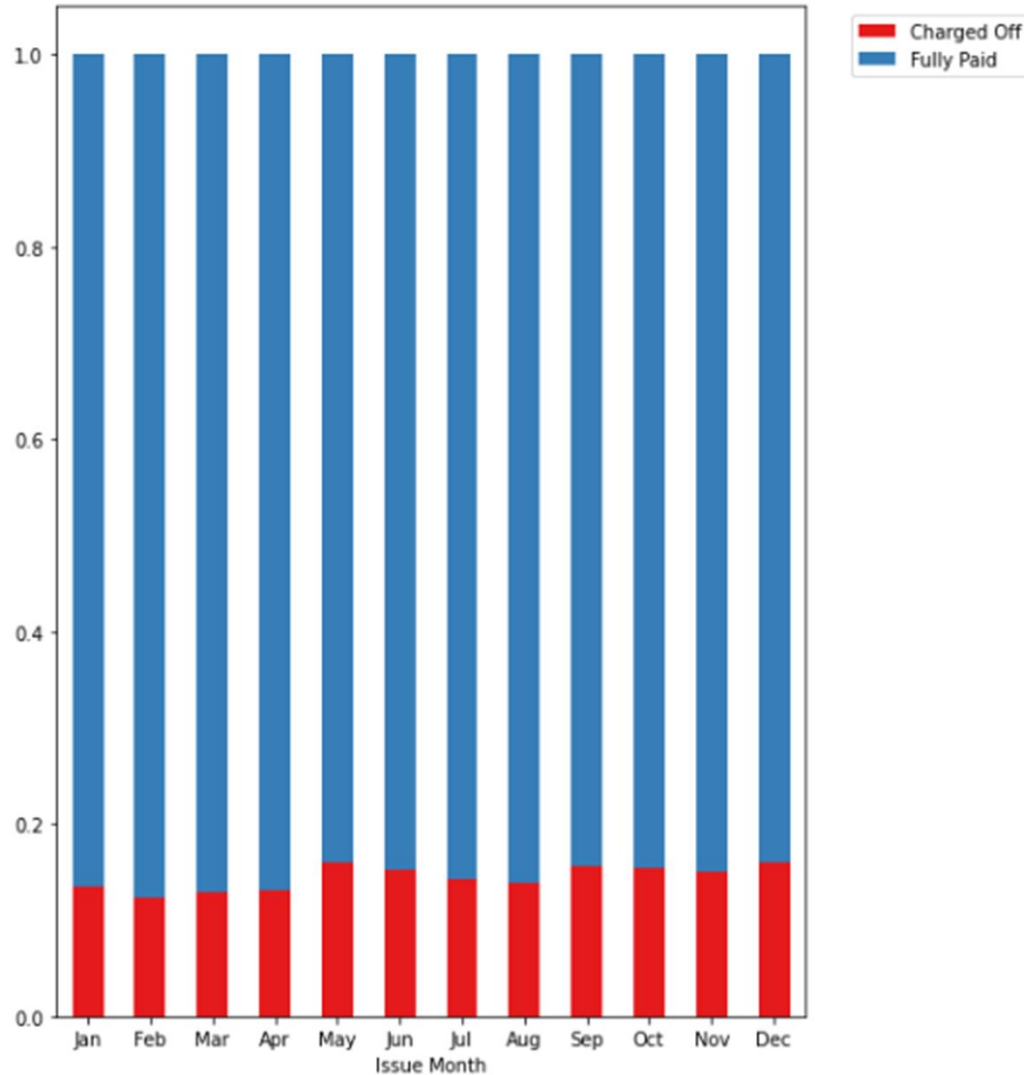


Charged Off
Fully Paid

1	455.00	2705.00	0.14
10+	1325.00	7117.00	0.16
2	565.00	3712.00	0.13
3	554.00	3444.00	0.14
4	458.00	2875.00	0.14
5	457.00	2727.00	0.14
6	306.00	1857.00	0.14
7	263.00	1443.00	0.15
8	203.00	1227.00	0.14
9	158.00	1063.00	0.13
< 1	638.00	3862.00	0.14
unknown	227.00	804.00	0.22
	Charged Off	Fully Paid	Charged Off Ratio
	loan_status		

ANALYSIS ON LOAN ISSUE MONTH

COMPARISON OF LOAN STATUS WITH LOAN ISSUE MONTH



Jan	319.00	2048.00	0.13
Feb	290.00	2062.00	0.12
Mar	347.00	2338.00	0.13
Apr	370.00	2454.00	0.13
May	463.00	2446.00	0.16
Jun	481.00	2685.00	0.15
Jul	477.00	2869.00	0.14
Aug	467.00	2914.00	0.14
Sep	546.00	2938.00	0.16
Oct	576.00	3161.00	0.15
Nov	597.00	3395.00	0.15
Dec	676.00	3526.00	0.16

issue_month

loan_status

Charged Off

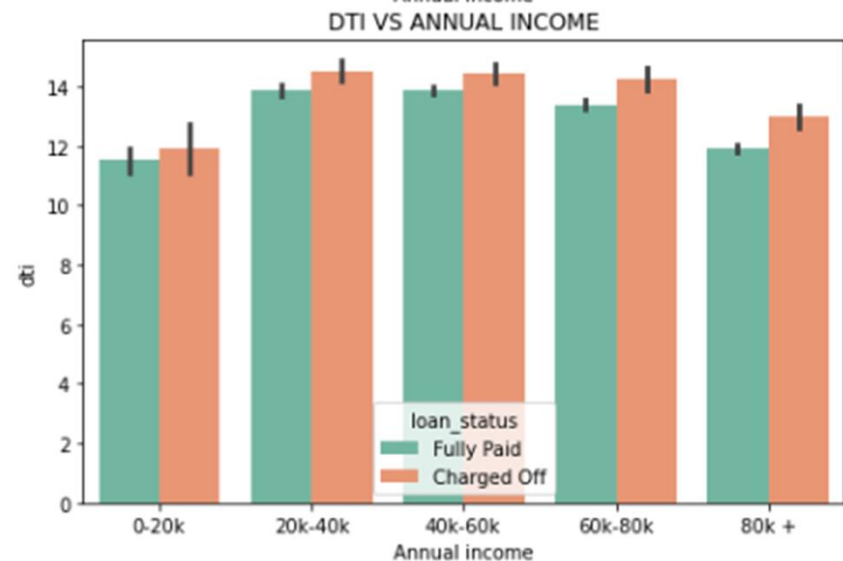
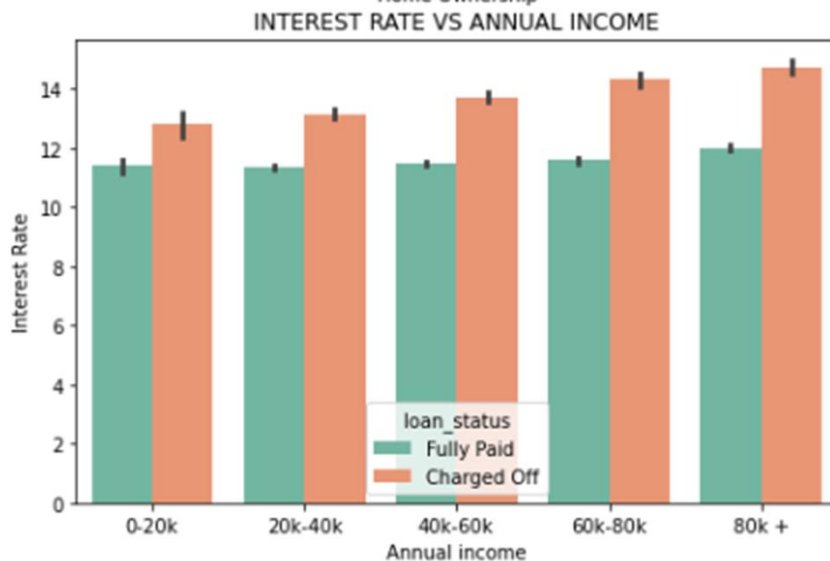
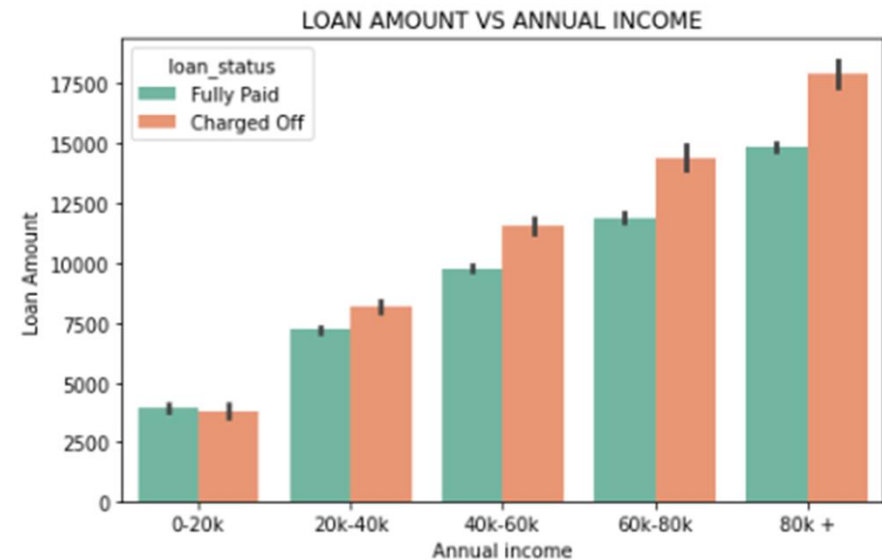
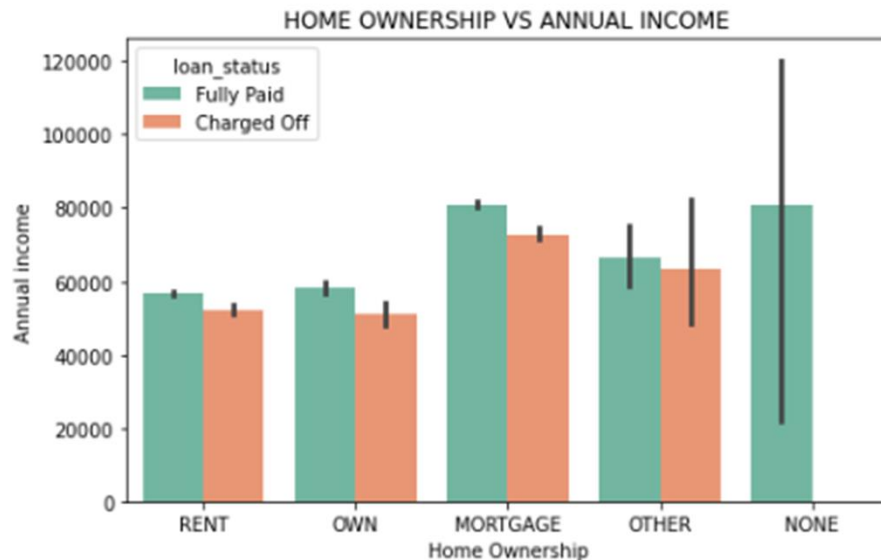
Fully Paid

Charged Off Ratio

**LOAN ISSUE MONTH
WITH RESPECT TO
CHARGED OFF LOAN
PROPORTION**

- There is more possibility of defaulters in 1st quarter of the year compared to 3rd and 4th quarters

MULTIVARIATE ANALYSIS ON ANNUAL INCOME



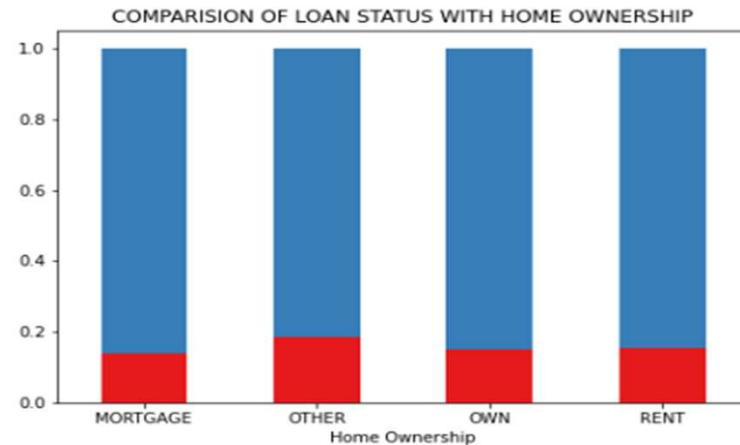
ANNUAL INCOME WITH RESPECT TO LOAN STATUS VS OTHER PARAMETERS

- Across all the income groups, "Loan Amount", "Interest Rate" and "Debt-to-Income Ratio" are higher for the customers who defaulted

ANALYSIS ON HOME OWNERSHIP & VERIFICATION STATUS

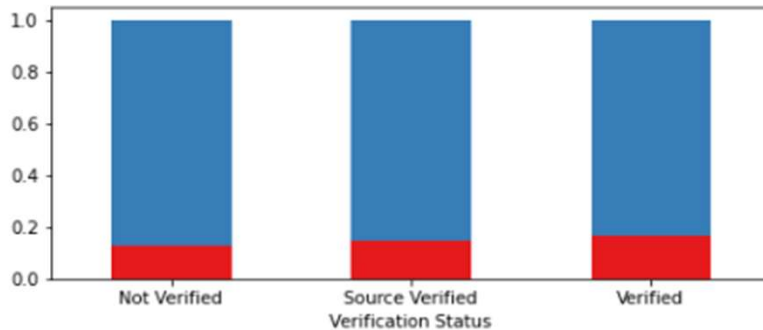
HOME OWNERSHIP WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- There is not much variation in ratio of defaulters with respect to non-defaulters
- Doesn't seem to be a good driving factor



home_ownership	MORTGAGE	2317.00	14616.00	0.14
	NONE	0.00	3.00	0.00
	OTHER	18.00	79.00	0.19
	OWN	440.00	2525.00	0.15
	RENT	2834.00	15613.00	0.15
		Charged Off	Fully Paid	Charged Off Ratio
		loan_status		

COMPARISON OF LOAN STATUS WITH VERIFICATION STATUS



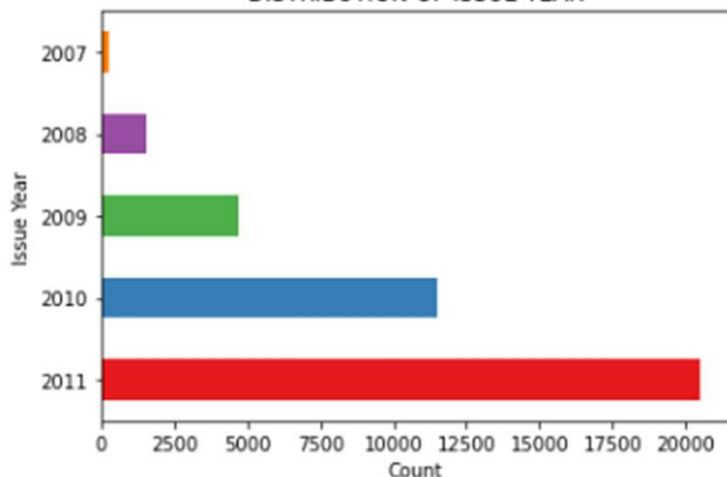
verification_status	Not Verified	2135.00	14515.00	0.13
	Source Verified	1429.00	8202.00	0.15
	Verified	2045.00	10119.00	0.17
		Charged Off	Fully Paid	Charged Off Ratio
		loan_status		

VERIFICATION STATUS WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- Less variation in ratio of defaulters with respect to non-defaulters and lots of not-verified status.
- Sounds to be not a good driving factor

ANALYSIS ON LOAN ISSUE YEAR

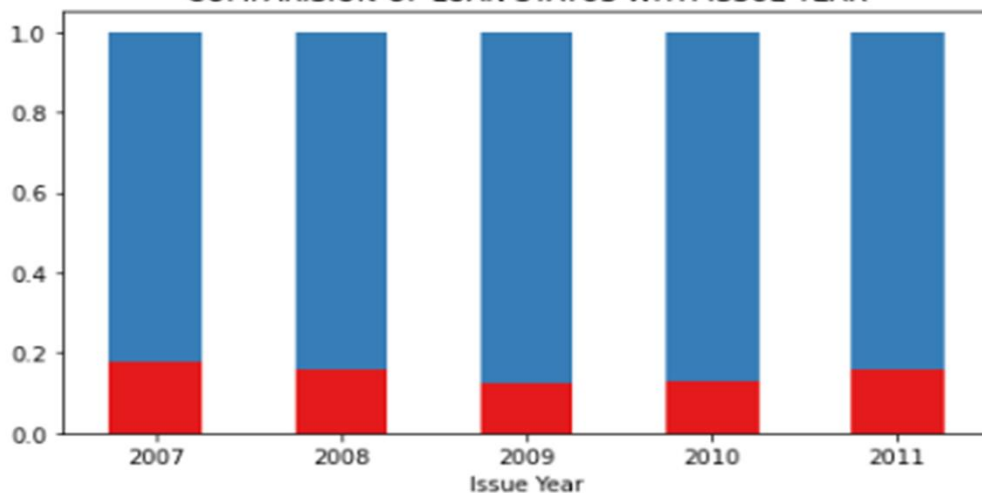
DISTRIBUTION OF ISSUE YEAR



LOAN ISSUE YEAR WITH RESPECT TO CHARGED OFF LOAN PROPORTION

- The count of loan issued increases from 2007 to 2011
- The higher rate of defaulters were in year 2009 and 2010, and we can't guess the reason
- Seems not a good driving factor

COMPARISION OF LOAN STATUS WITH ISSUE YEAR



Charged Off
Fully Paid

issue_year	loan_status		Charged Off Ratio
	Charged Off	Fully Paid	
2007	44.00	205.00	0.18
2008	246.00	1310.00	0.16
2009	593.00	4103.00	0.13
2010	1475.00	10007.00	0.13
2011	3251.00	17211.00	0.16

CONCLUSION

- Higher the loan amount, more likely the customers are to be defaulted
- At higher interest rate, customers are more likely to be defaulted
- Small_business owners are the highest defaulters
- People with higher income default less
- The rate of defaulters increases from grade A to grade G
- The customers with higher sub-grades are 'risky' applicants
- The rate of defaulters is more in case of 60 months term
- The rate of defaulters gradually increase from lower debt-to-income ratio to higher debt-to-income ratio
- The customers with employment length 10+ years are having higher rate of defaulters, if we ignore unknown values. But, there is no pattern observed for other length of years
- The top 3 states, having highest chances of getting default customers are NE, NV and SD
- There is more possibility of defaulters in 1st quarter of the year compared to 3rd and 4th quarters
- Across all the income groups, "Loan Amount", "Interest Rate" and "Debt-to-Income Ratio" are higher for the customers who defaulted

DRIVING FACTORS: -

loan_amnt, int_rate, purpose, annual_inc, grade, sub_grade, term, dti, address_state, issue_month

NON-DRIVING FACTORS: -

home_ownership, verification_status, issue_year, emp_length

RECOMMENDATIONS

Lending Club should be careful while giving loan in following scenarios:-

- Applicants with grade G, especially when –
 - 1) Annual Income is between 60k to 80k
 - 2) Loan Amount is between 17500 to 20000
 - 3) Interest Rate is above 20%
- Applicants with higher sub-grades, especially F5, G5, G3, G2
- Applicants with lower income
- Applicants with higher loan amount
- Applicants with loan term period of 60 months
- Applicants are small business owners
- Applicants with high debt-to-income ratio
- Applicant is from state NE