

LENDING CLUB CASE STUDY

BY:

Yashwant Yadav and Vivek Sharma

DATA CLEANING

Columns with single unique values/ null values are deleted as they don't play much role in analysis. Like : ['pymnt_plan', 'initial_list_status', 'collections_12_mths_ex_med', 'policy_code', 'application_type', 'acc_now_delinq', 'chargeoff_within_12_mths', 'delinq_amnt', 'tax_liens']

Columns which don't contribute much to our analysis are also deleted. ["id", "member_id", "url"]

While analysing 'loan_status' we found that it was 'Current' as one of its value. Since a person who has already taken the loan won't play much role in our analysis, so we have dropped it as well.

The 'term' columns as empty spaces before and after it, so have stripped it for consistency.

DATA MANIPULATION

01

Column like 'issue_d' is having date values so we have converted it to datetime. We have further created two derived columns 'issue_year' and 'issue_month' from date for future analysis.

02

Columns 'int_rate' is having '%' symbol in the end, which is now stripped.

03

Columns like 'funded_amt_inv' / 'installment' / 'annual_inc' / 'dti' are rounded off to two decimal places.

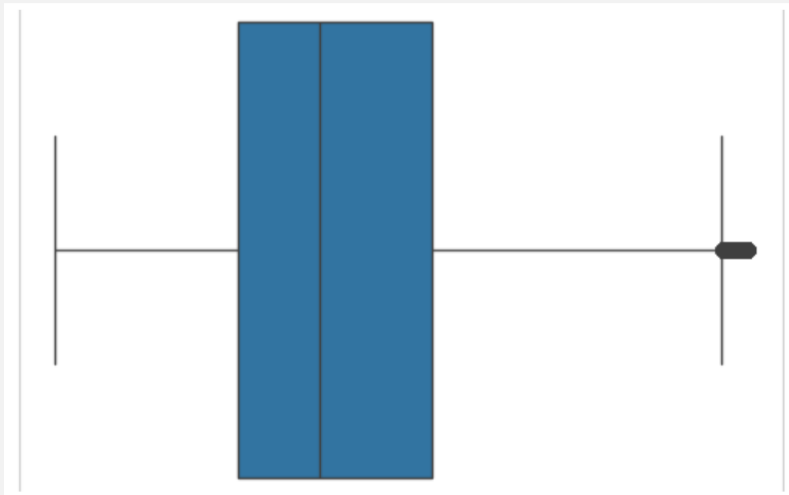
04

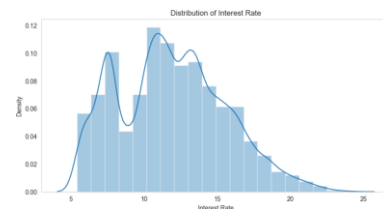
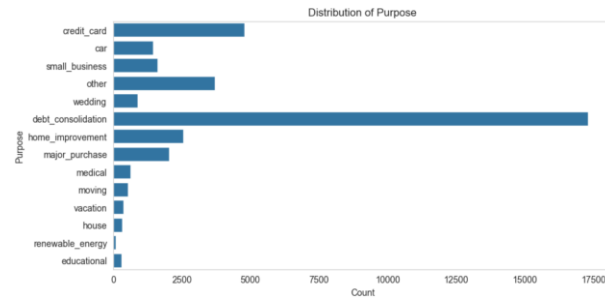
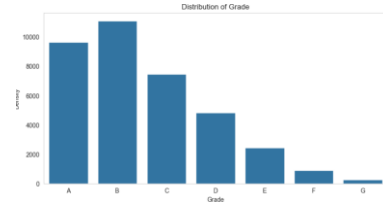
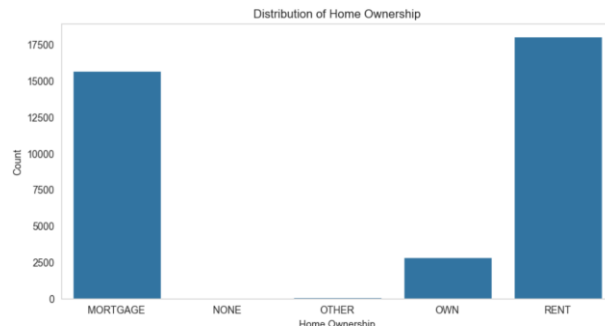
Columns like 'grade' / 'sub_grade' / 'home_ownership' / 'verification_status' are converted to category type so that we use them while label encoding the dataset in the future.



OUTLIER ANALYSIS

- 'annual_inc' was seeing having varied variety of values, So we removed values after the 95 percentile.

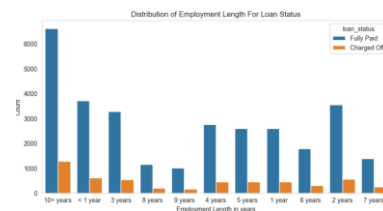
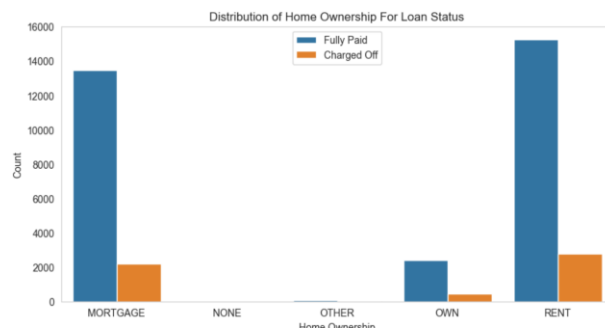
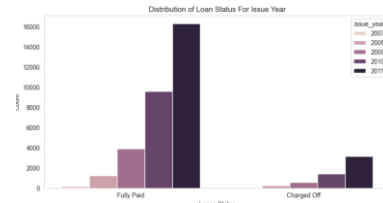
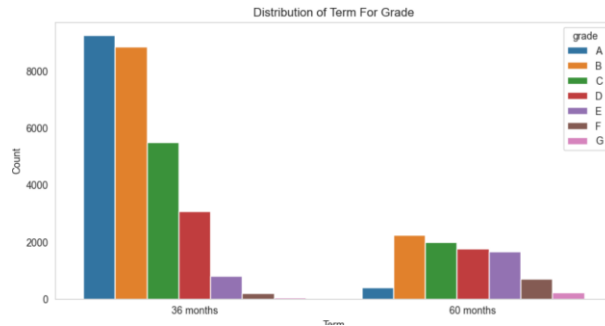




UNIVARIATE ANALYSIS

- Analysis of loan status : Fully Paid = 85% & Charged off = 15%
- Loan amount varies from 500 to 35000
- Most than half of the loan taken for 3 years compared to 5 years.
- Most of the interest percentage lies between 5-10 and 10-15 percent.
- A large amount of loans are with grade 'A' and 'B' compared to rest showing most loans are high grade loans.
- A large amount of loans are with lower subgrade in grade 'A' and 'B' compared to rest showing most loans are high grade loans. This matches with the Grade distribution.
- Majority of borrowers have working experience greater than 10 years.
- Majority of borrowers don't posses property and are on mortgage or rent.
- About 50% of the borrowers are verified by the company or have source verified.
- Annual Income shows left skewed normal distribution thus we can say that the majority of burrowers have very low annual income compared to rest.
- A large percentage of loans are taken for debt consolidation followed by credit card.
- Majority of the borrowers have very large debt compared to the income registerd, concentrated in the 10-15 DTI ratio.
- Majority of the borrowers have no record of Public Recorded Bankruptcy.
- Majority of the loans are given in last quarter of the year.

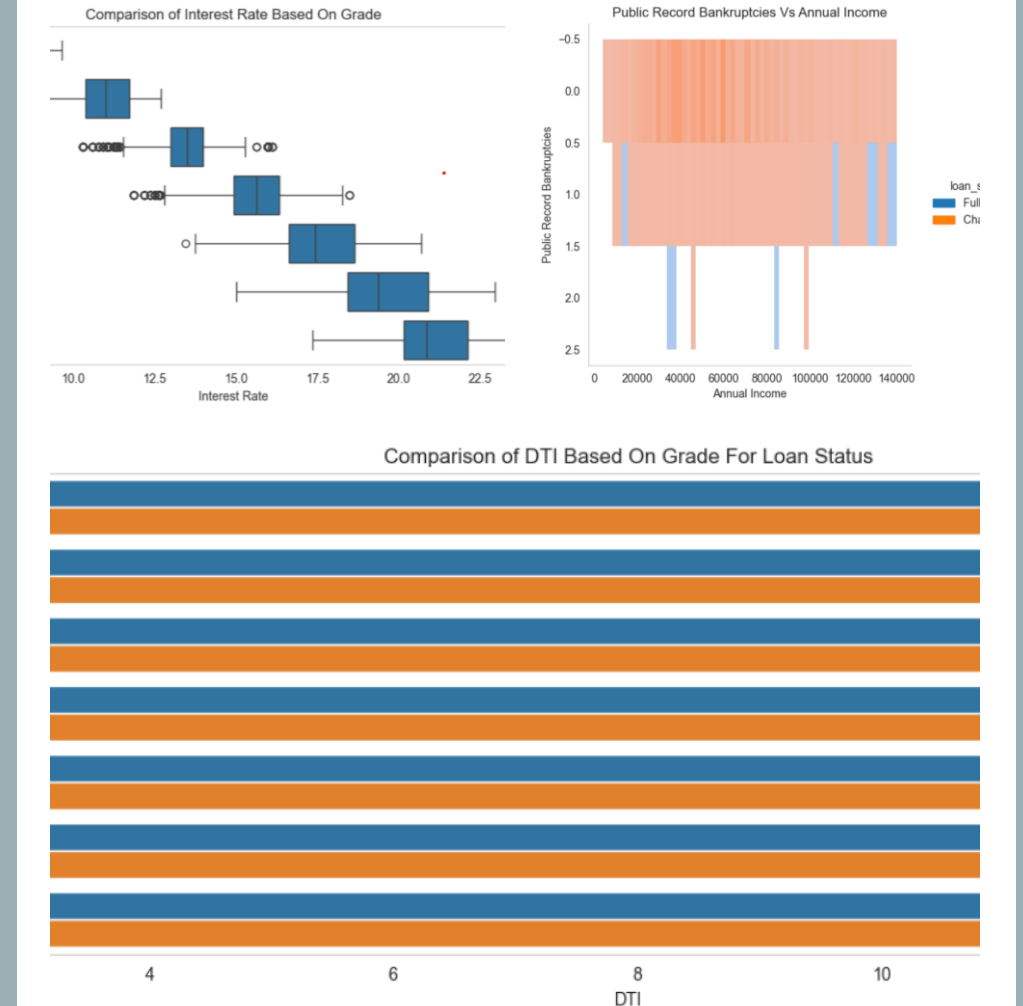
SEGMENTED UNIVARIATE ANALYSIS



- Debt Consolidation is the most popular loan purpose and has highest number of fully paid loan and defaulted loan.
- The mean and 25% percentile are same for both but we see larger 75% percentile in the defaulted loan which indicate large amount of loan has higher chance of defaulting.
- The 60 month term has higher chance of defaulting than 36 month term whereas the 36 month term has higher chance of fully paid loan.
- The loans in 36 month term majorly consist of grade A and B loans whereas the loans in 60 month term mostly consist of grade B, C and D loans.
- The Loan Status varies with DTI ratio, we can see that the loans in DTI ratio 10-15 have higher number of defaulted loan but higher dti has higher chance of defaulting.
- The Defaulted loan are lower for the burrowers which own their property compared to on mortgage or rent.
- Burrowers with less 50000 annual income are more likely to default and higher annual income are less likely to default.
- The Fully paid loan are increasing exponentially with the time compared to defaulted loan.
- The Employees with 10+ years of experience are likely to default and have higher chance of fully paying the loan.

BIVARIATE ANALYSIS

- The Grade represent risk factor thus we can say interest rate increases with the risk.
- The Grade A which is lowest risk also has lowest DTI ratio which we can say that higher grade has lower rate of default.
- The borrowers are mostly having no record of Public Recorded Bankruptcy and are safe choice for loan issue.



RECOMMENDATIONS

- - Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:
 - 1. DTI
 - 2. Grades
 - 3. Verification Status
 - 4. Annual income
 - 5. Pub_rec_bankruptcies
- - Other considerations for 'defaults' :
 - 1. Borrowers not from large urban cities like california, new york, texas, florida etc.
 - 2. Borrowers having annual income in the range 50000-100000.
 - 3. Borrowers having Public Recorded Bankruptcy.
 - 4. Borrowers with least grades like E,F,G which indicates high risk.
 - 5. Borrowers with very high Debt to Income value.
 - 6. Borrowers with working experience 10+ years.