Lending Club Case Study

Vinayak and Vikas

Business Understanding

We work for a 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:

- 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

Business Objective and Goal

The data is shared to make decision after receiving the application based on Applicant profile. This involves Risk factors based on loss to the company, If a candidate repay the loan amount or not.

We need to understand if an applicant is defaulted or non-defaulted based on the data for past loan application. We need to find the person is likely to default based on we can define our below actions to avoid Risks.

- Denying loan
- Reducing loan Amount
- Lending at highest interest rate

We need to collect data based on consumer attribute and loan attributes.

Business Goal

Reduce the risk and make sure collect the data in to take decision on the loan application to avoid financial loss.

EDA Methodology

- Data Sourcing

- Data Cleaning

- Data Understanding

- Univariate Graphs

- Univariate Analysis

- Bivariate Graphs

- Bivariate Analysis

- Conclusion

Data Sourcing

- Source dataset is loan.csv(Private Data)
- Details about attributes and column definition defined under the Data_Dictionary.xlsx
- We are using the below tools to perform the case study
  - Google Collab
  - Python 3.0
  - Pandas 2.0.3
  - Seaborn 0.13.1
  - Matplotlib 3.7.1
- Data has the below stats:
  - 39717 entries
  - Total 111 columns
  - Involves float,int and object(string) data types
  - Data size is: 33.6+MB

Data Cleaning

Involves the below considerations:

- Sanity for duplicates
- Sanity for empty rows and columns
- Validation Missing values(NULL) and Handling
  - Removed the columns based on the NULL values counts
  - Fixed the NULL values
- Correct the data types
- Filtered the data
- Done pre-check and removed the outliers "annual_inc" based on IQR
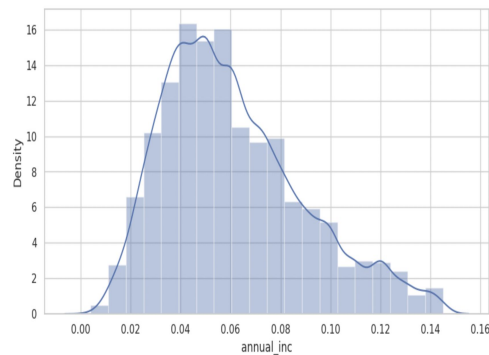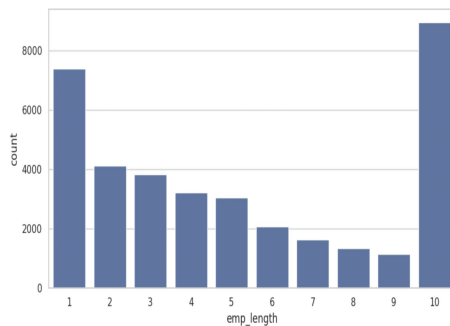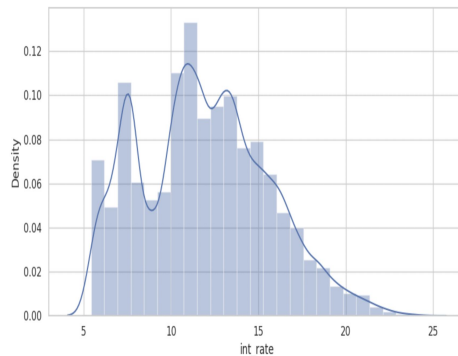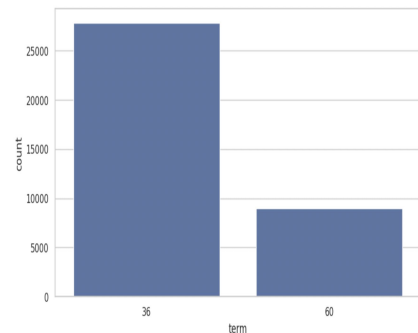- Created new columns

Data Understanding

- After filtering we need to validate the data.
- Segregated data based on data type.

| Numerical	Categorical
id	Grade
loan_amnt	Sub_Grade
term	Home_ownership
int_rate	Verification_status
emp_length	Loan_status
annual_inc	purpose
dti	addr_state
issue_d	pub_rec_bankruptcies

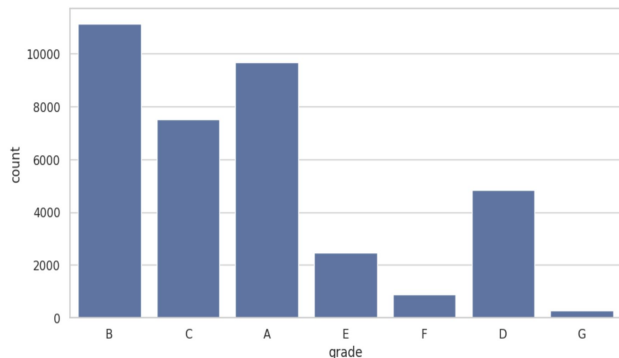# Univariate Graphs(Numerical)

- ~76% of borrowers are with "36 months" terms which is almost 3/4 of total application
- Observed that the most of the loan lended between 8% to 16%.
- 11.9% is the mean and highest counts of loan lended with
- Highest borrower are from more than 10 years of service.
- Majority of borrower has the less income(less than 0.6M) since there was a data skew in the above graph
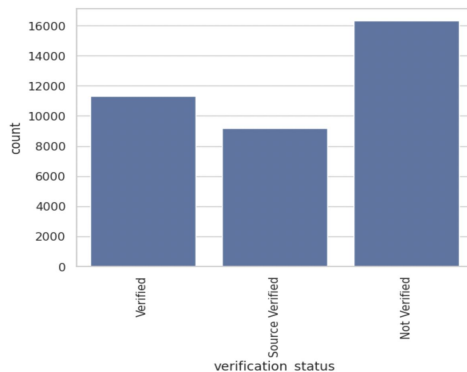
Univariate Graphs(Numerical)
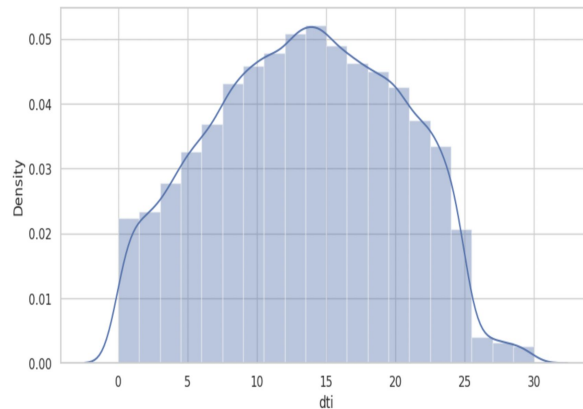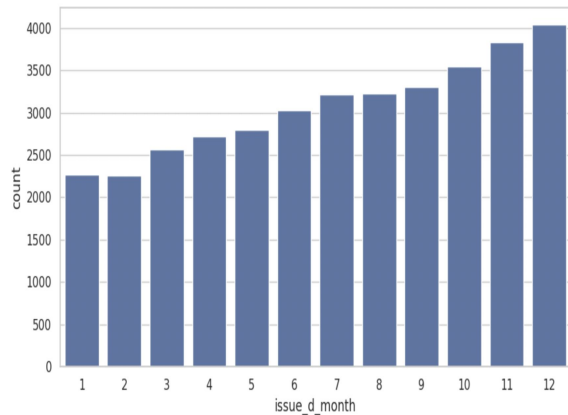
- Most borrower has high DTI means debts is comparatively high to salary
- Most of the borrowers have loan in 2011 year
- Every year borrowing loans are increasing
- There is an incremental growth for loan count for every month of the year

Univariate Graphs(Categorical)
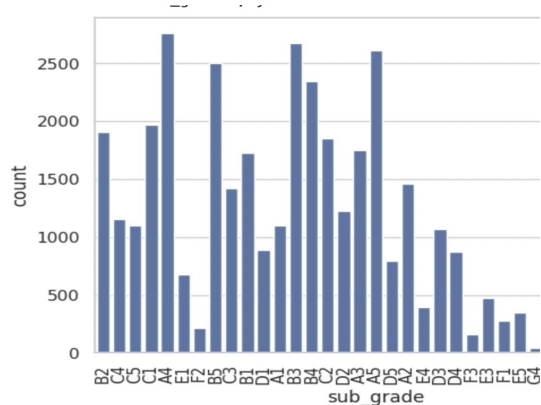
- Highest borrowers are from the grade B
- A4, B3 and A5 are the highest sub-categories who borrowed the loan
- Most of the borrowers are rent OR mortgage, Rent home ownership is the highest one.
- ~16000 borrowers are not verified
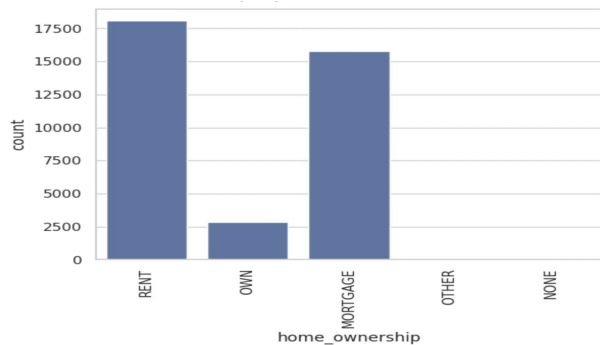
Univariate Graphs(Categorical)
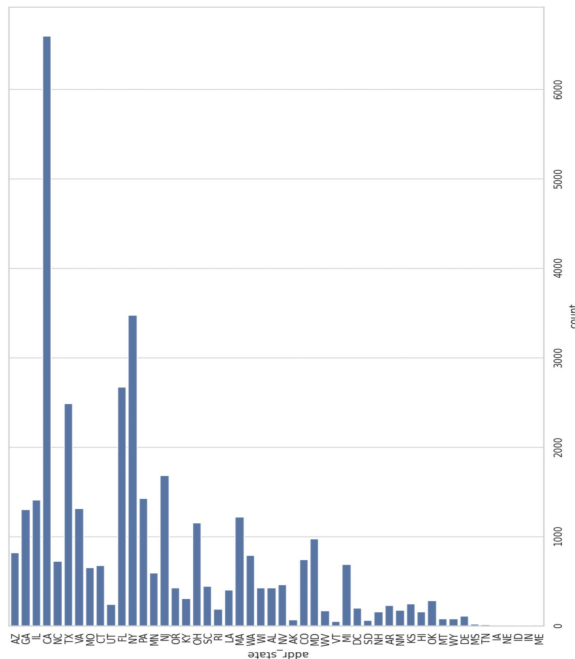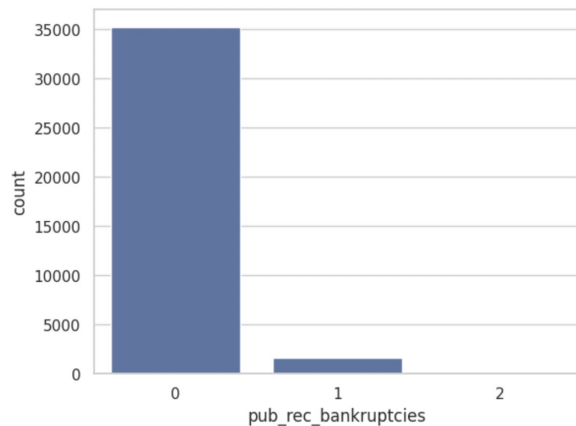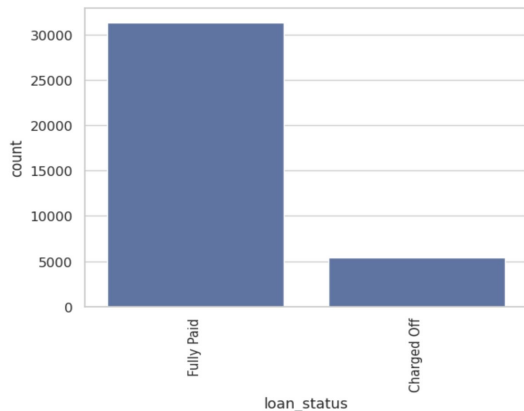
- ● ~85% count of loans has the loan_status as "Fully Paid" only ~15% of loans had status as "Charged off", So the default loans are very less compared to the paid "Paid"
- ● debt_consolidation is the most purpose to borrow the loan
- ● Majority of the borrowers have no record of Public Recorded Bankruptcy.

Univariate Analysis

- ~76% of borrowers are with "36 months" terms which is almost 3/4 of total application
- Observed that the most of the loan lended between 8% to 16%.
- 11.9% is the mean and highest counts of loan lended with
- Highest borrower are from more than 10 years of service.
- Majority of borrower has the less income(less than 0.6M) since there was a data skew in the above graph
- Most borrower has high DTI means debts is comparatively high to salary
- Most of the borrowers have loan in 2011 year
- Every year borrowing loans are increasing
- There is an incremental growth for loan count for every month of the year
- Highest borrowers are from the grade B
- A4, B3 and A5 are the highest sub-categories who borrowed the loan
- Most of the borrowers are rent OR mortgage, Rent home ownership is the highest one.
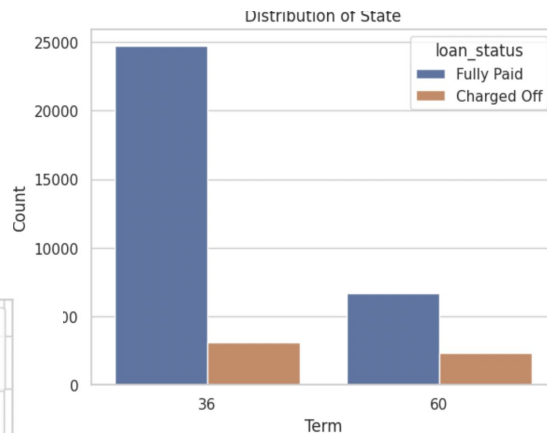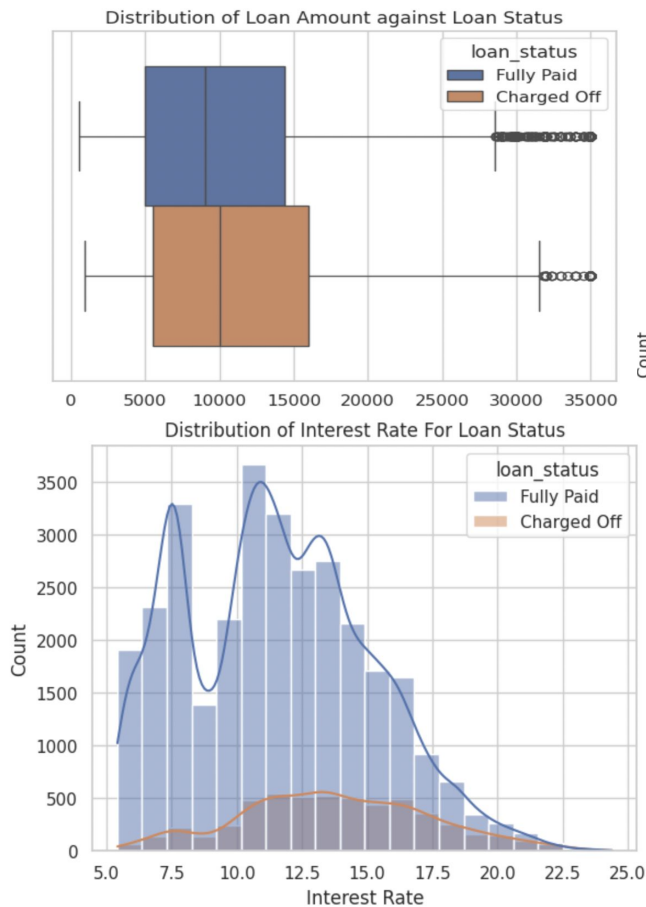- ~16000 borrowers are not verified
- ~85% count of loans has the loan_status as "Fully Paid" only ~15% of loans had status as "Charged off", So the default loans are very less compared to the paid "Paid"
- debt_consolidation is the most purpose to borrow the loan
- Majority of the borrowers have no record of Public Recorded Bankruptcy.

Bivariate Graphs

Distribution of Loan Amount against Loan Status
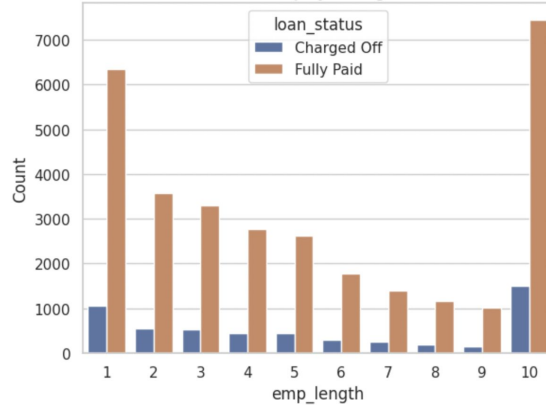
Distribution of State

Distribution of Interest Rate For Loan Status

- Difference in the 75 Percentile indicates that large amount of loan can cause the defaulting
- Borrowers has 36 Months terms is more safer and less chance to be defaulter. Compares to 60 Months term
- Defaulters increased from 10% interest rate and then decreased after 17% of rate interest.

Bivariate Graphs

Distribution of employee length of service

Disctibution of employee based on dti

Disctibution of employee based on annual income
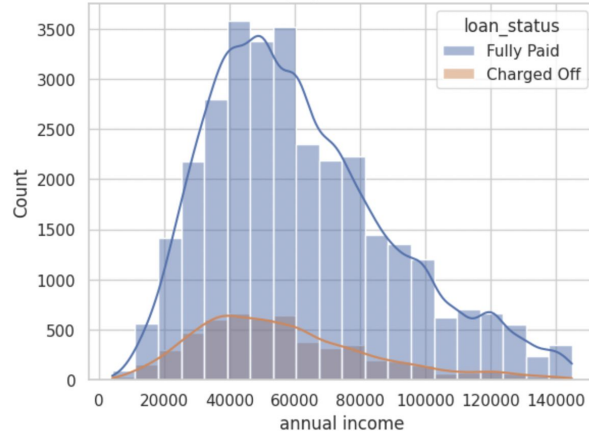
- Employee less than 1 and higher than 10 years of service length has the high number of defaulters
- Income between 40000 to 60000 has the high number of charged off and that can leads high number of defaulters.
- After 60000 we see that number of defaulter decreased
- We can see that the loans in DTI ratio 12-17 have higher number of defaulted loan but higher dti has higher chance of defaulting.

Bivariate Graphs
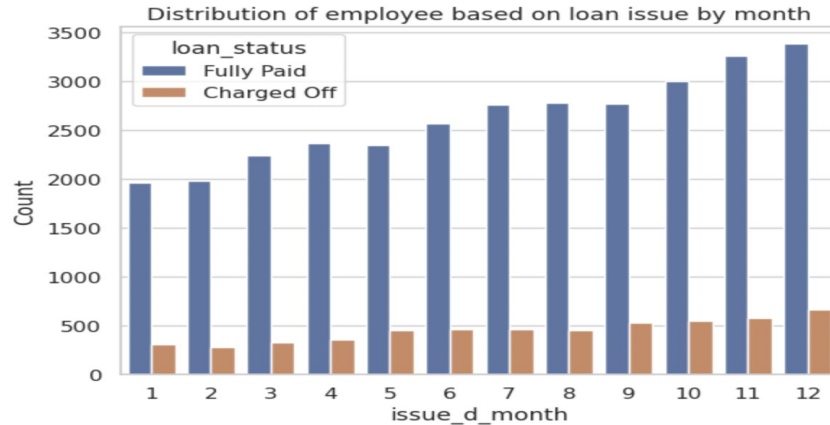
Distribution of employee based on loan issue by month

- Month 1, 5 and 9 Has the high charged off comparatively similarity Fully paid in the other months
- Month 2,4,7 and 8 has less charged off comparatively similarity Fully paid in the other months
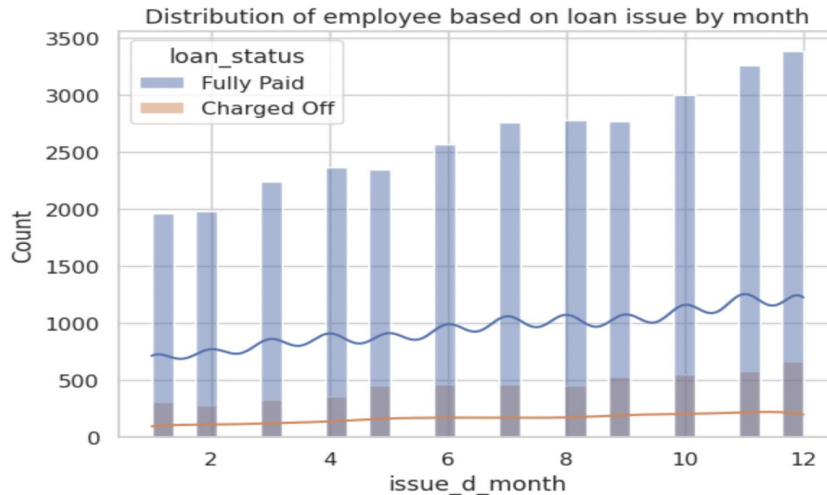
Bivariate Graphs

Distribution of employee based on grade

Distribution of employee based on purpose
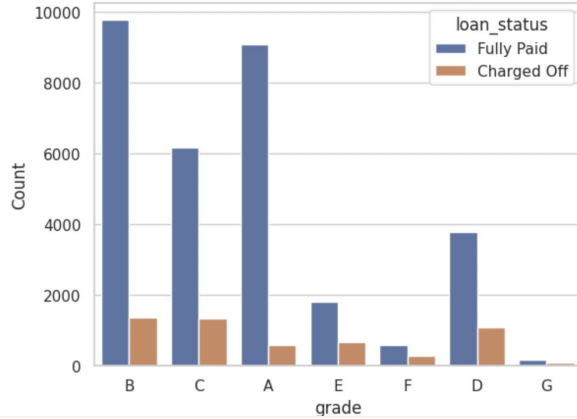
Distribution of employee based on sub-grade

- Category A and B is good to consider since the charger off is comparably less.
- Category E and F is not consider since the charger off is comparably high.
  a. Fully_Paid: A1,A2,A3,A4 and A5 is Good Comparatively B1 ,B2, B3, B4 and B5
  b. Defaulters: E and F still bad, however E is having good count compares to F
- Along with debt_consolidation car, credit_card are other good purpose

Bivariate Analysis

- Difference in the 75 Percentile indicates that large amount of loan can cause the defaulting
- Borrowers has 36 Months terms is more safer and less chance to be defaulter. Compares to 60 Months term
- Defaulters increased from 10% interest rate and then decreased after 17% of rate interest.
- Employee less than 1 and higher than 10 years of service length has the high number of defaulters
- Income between 40000 to 60000 has the high number of charged off and that can leads high number of defaulters.
- After 60000 we see that number of defaulter decreased
- We can see that the loans in DTI ratio 12-17 have higher number of defaulted loan but higher dti has higher chance of defaulting.
- Month 1, 5 and 9 Has the high charged off comparatively similarity Fully paid in the other months
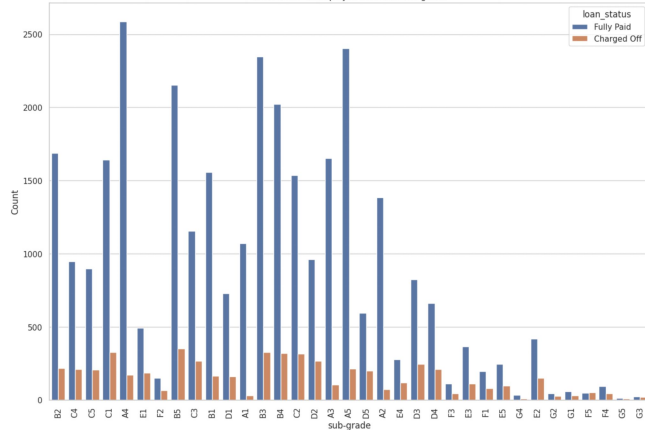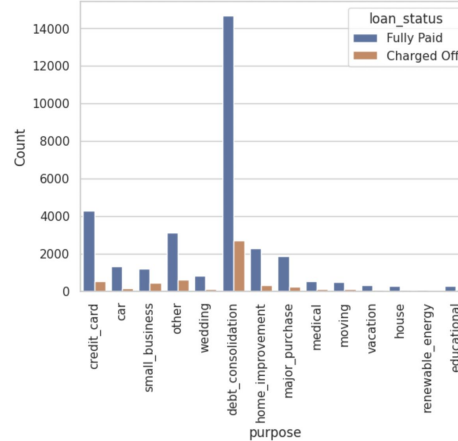- Month 2,4,7 and 8 has less charged off comparatively similarity Fully paid in the other months
- Category A and B is good to consider since the charger off is comparably less.
- Category E and F is not consider since the charger off is comparably high.
  a. Fully_Paid: A1,A2,A3,A4 and A5 is Good Comparatively B1 ,B2, B3, B4 and B5
  b. Defaulters: E and F still bad, however E is having good count compares to F
- Along with debt_consolidation car, credit_card are other good purpose

Conclusion

Loan Amount: Larger loan amounts significantly increase the likelihood of defaulting. Higher-value loans require careful assessment of borrower capacity and stringent risk management practices.

Loan Term: Borrowers opting for longer loan terms, such as 60 months, exhibit higher default rates compared to those choosing shorter terms like 36 months. Encouraging borrowers to opt for shorter terms can mitigate default risk.

Interest Rate: There's a critical threshold between 10-17% interest rates where the risk of default is heightened. Loans with rates above 10% show increased default rates, peaking before declining again after 17%. Monitoring and managing loans within this range is crucial to mitigate default risks.

Employment Tenure: Borrowers with extreme employment tenures, either very new or very experienced, show higher default rates. Targeted risk assessment and mitigation strategies are needed for these groups.

Conclusion ctnd.

Income: Lower-income borrowers, especially those earning between 40,000 and 60,000, exhibit a higher risk of default. Targeted risk management strategies are necessary for this income group.

Debt-to-Income (DTI) Ratio: Borrowers with higher DTI ratios are at an increased risk of defaulting. DTI with 12-17 have higher number of defaulted loan but higher dti has higher chance of defaulting.

Seasonal Trends: There are seasonal trends in loan borrowing and repayment behavior. Understanding these patterns can help anticipate potential default risks during certain months.

- Month 1, 5 and 9 Has the high charged off comparatively similarity Fully paid in the other months
- Month 2,4,7 and 8 has less charged off comparatively similarity Fully paid in the other months

Loan Grade and Sub-categories: Grades A and B show favorable performance with lower charge-off rates, while categories E and F demonstrate comparatively higher charge-off rates. Sub-categories A1-A5 and B1-B5 exhibit better repayment records compared to E1-E5 and F1-F5
.

Loan Purpose: Debt_consolidation is the primary purpose for borrowing loans. Understanding borrower motivations can help assess their ability to repay.

Conclusion ctnd.

Overall Default Rate: While around 15% of loans have been charged off, the majority (approximately 85%) have been fully paid. Despite this favorable trend, careful risk assessment and management remain essential.

Verification: Unverified borrowers may present higher default risks due to potential inaccuracies in their financial profiles, necessitating additional risk management measures.

Public Recorded Bankruptcy: The majority of borrowers and defaulters having no record of public recorded bankruptcy suggests a lower prevalence of extreme financial distress among borrowers. This may indicate a lower overall risk of default compared to populations with higher bankruptcy rates.

Considering these factors collectively and integrating them into the lending decision-making process can help mitigate default risks and ensure responsible lending practices.

Thank You