Lending Club
Case Study

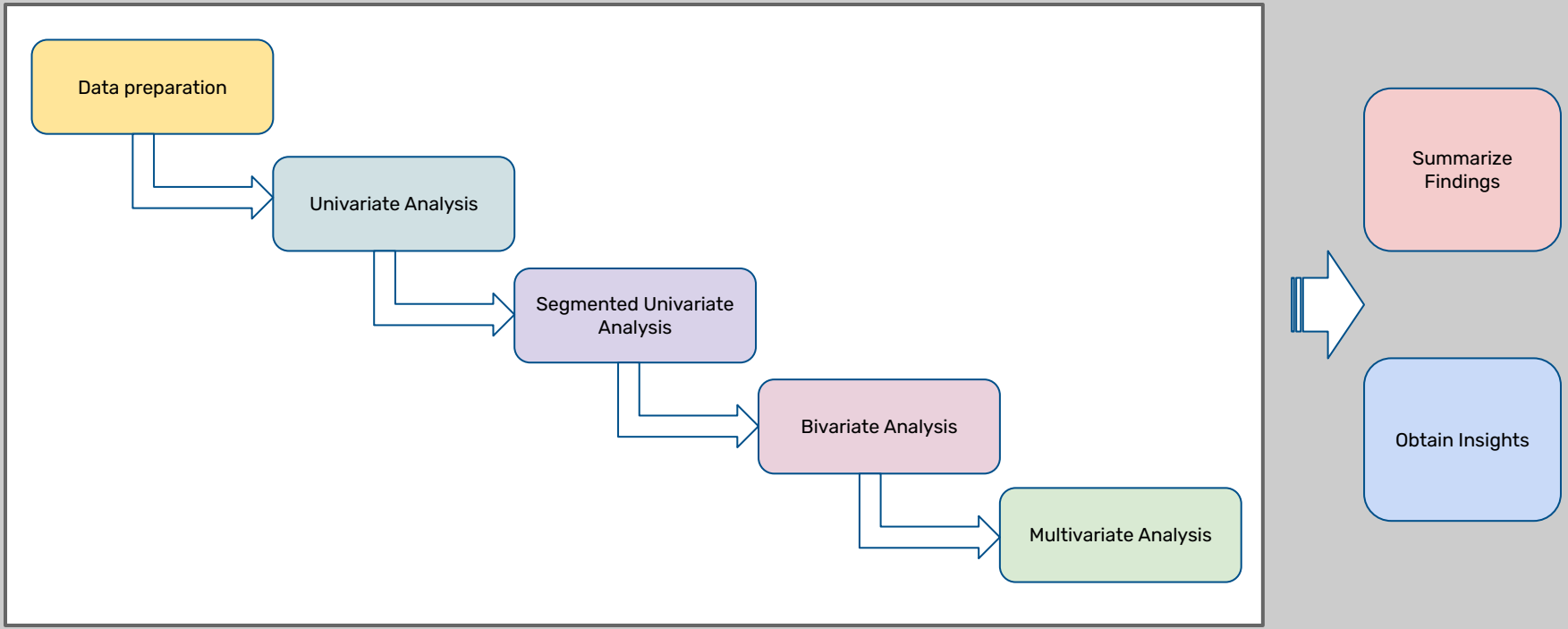
Deepesh Ramachandran Vijayalekshmi

Objective

To conduct Exploratory Data Analysis (EDA) on the Lending Club dataset, to identify the driving factors influencing loan defaults. This analysis aims to enhance the company's risk assessment capabilities, enabling more informed lending decisions that minimize financial losses.

By understanding the driving variables behind loan defaults, the company can better manage its portfolio and mitigate credit risk.

Approach


Data preparation
Univariate Analysis
Segmented Univariate Analysis
Bivariate Analysis
Multivariate Analysis
Summarize Findings
Obtain Insights

Top driving factors

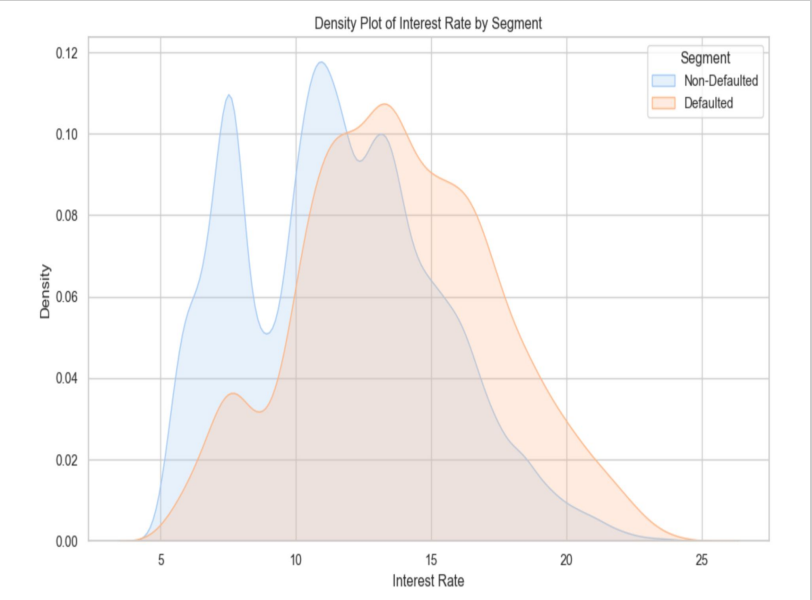
The primary factors contributing to loan defaults:

1. **Interest Rate**
2. **DTI - Debt to Income Ratio**
3. **LTI - Loan to Income Ratio**
4. **Grade**
5. **Sub Grade**
6. **Public Bankruptcies**
7. **Annual Income**
8. **Purpose**
9. **Loan Amount**

Interest Rate

Higher interest rates had a significantly higher rate of defaults

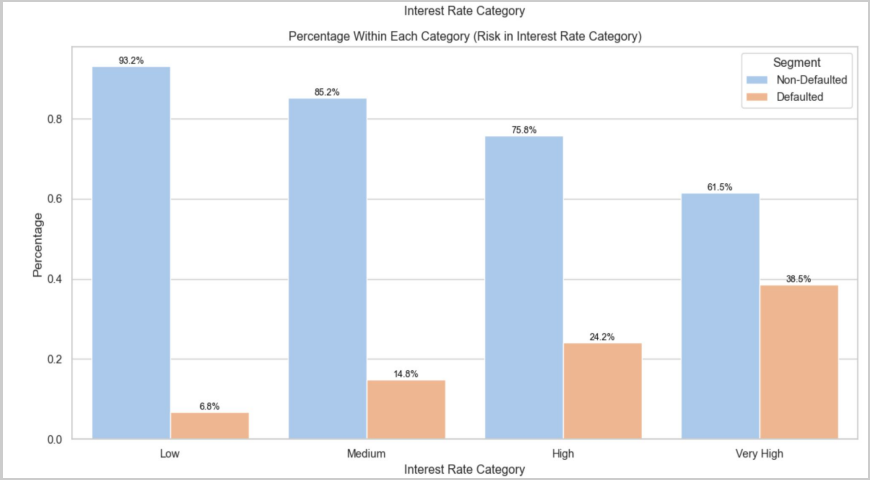
Borrowers with higher interest rates have a greater likelihood of default, as seen by the increased proportion of defaulted loans.



Interest Rate

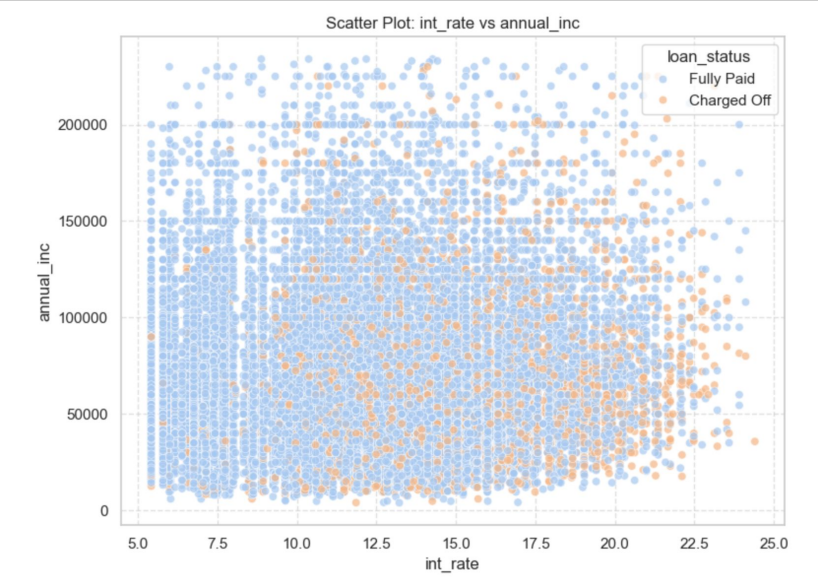
The proportion of defaulted loans increases almost linearly when the interest increases.

Higher interest category has higher defaults.



Interest Rate - Annual Income

Charged-off loans (orange dots) are denser in the lower income range when combined with higher interest rates (above 15%) .



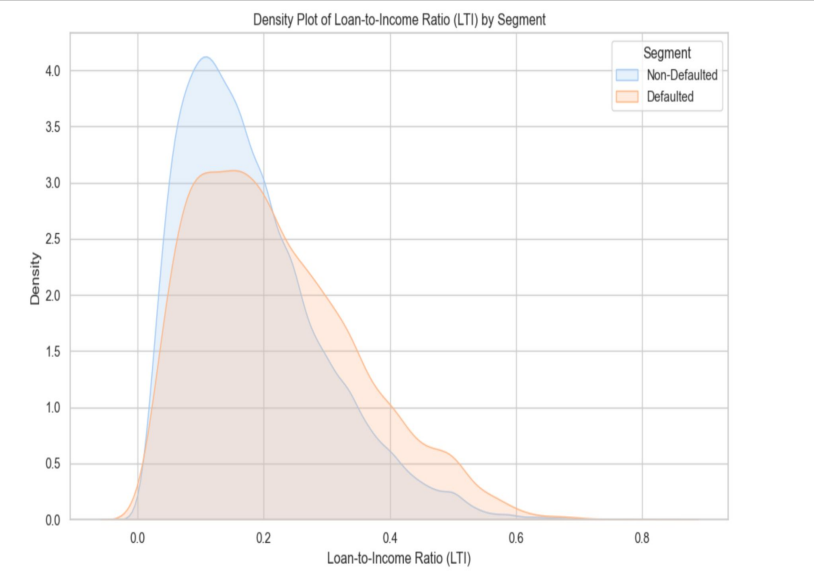
DTI - Loan to Income Ratio

For DTI values above approximately 14, the density for defaulted loans exceeds that of non-defaulted loans, indicating that borrowers with a higher DTI are more likely to default.

LTI - Loan to Income Ratio

The orange density curve (Defaulted) shifts slightly to the right compared to the blue curve (Non-Defaulted), indicating that borrowers with a higher Loan-to-Income ratio are more likely to default.

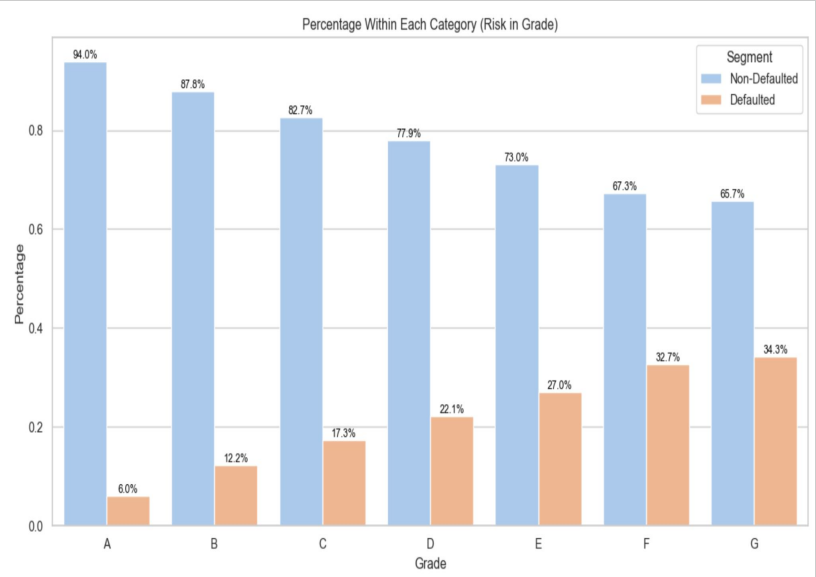
The Non-Defaulted loans peak at a slightly lower LTI value compared to the Defaulted loans, this suggests that lower LTIs are associated with lower default risks.



Grade

Risk of defaults increase towards lower grades. Grade A has lowest risk and G has the highest risk.

From the Percentage Within Each Category plot, almost a linear increase of default rate is seen towards lower grades.(orange bars)

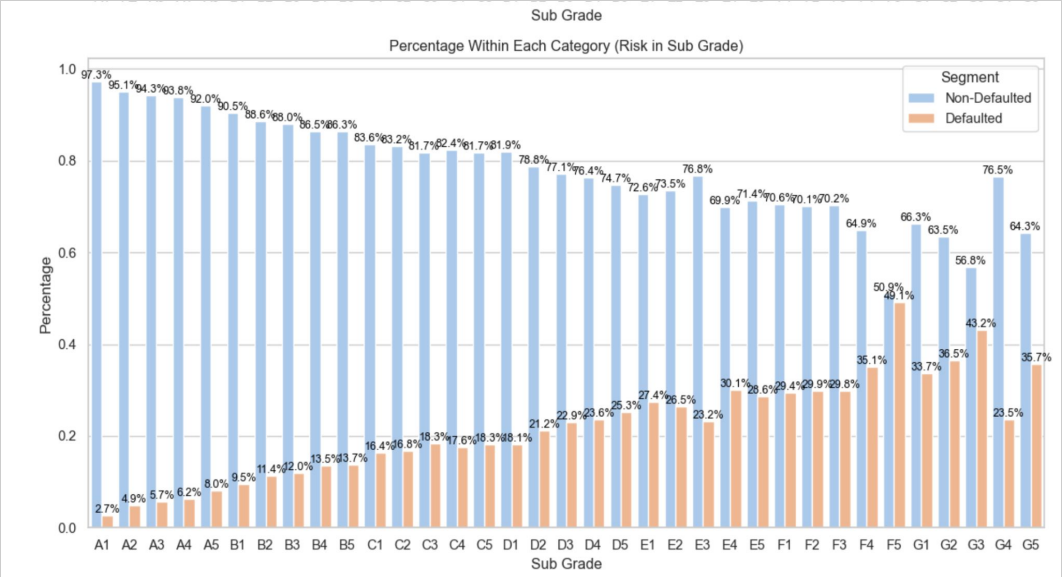


Sub Grade

The proportion of "Charged Off" loans (orange bars) increases in lower sub-grades, indicating higher risk associated with these categories.

There is a significantly higher default rate in F5 and G3

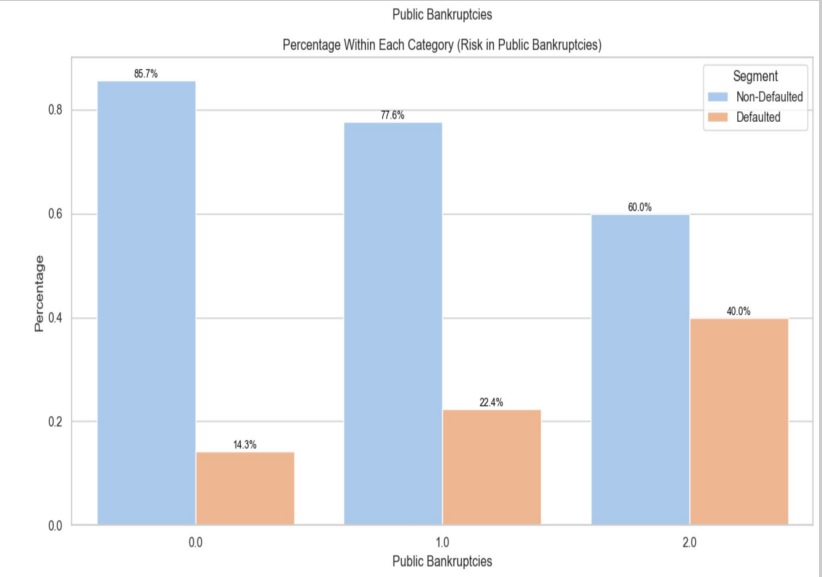
Note: The discrepancies on the right side (from F4 towards right) could suggest that the factors analyzed to allocate subgrades may need to be revised and updated



Public Record Bankruptcies

The percentage of defaulted loans increases with increase in number of bankruptcies showing an increasing risk.

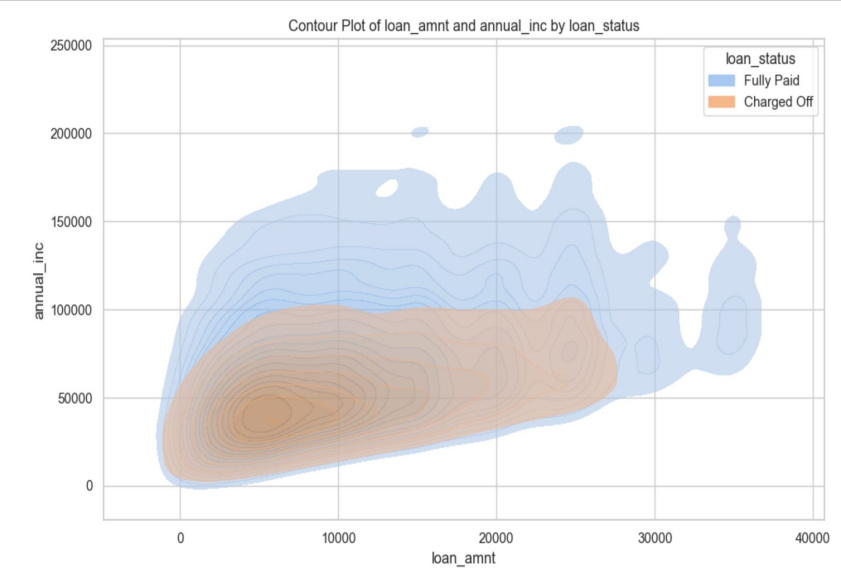
Borrowers with no history of bankruptcies have the highest non defaulted loans.



Annual Income

Lower incomes (< $60,000)are more densely concentrated within the Charged Off group, suggesting a higher likelihood of default in this segment.

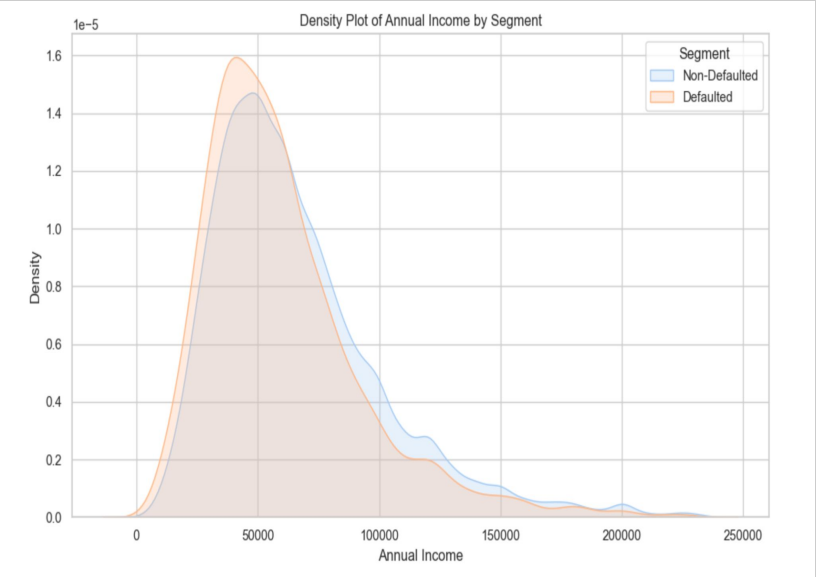
There are very few, if any, defaulted loans among borrowers with both high incomes and high loan amounts, indicating that this group is at a significantly lower risk of default.



Annual Income

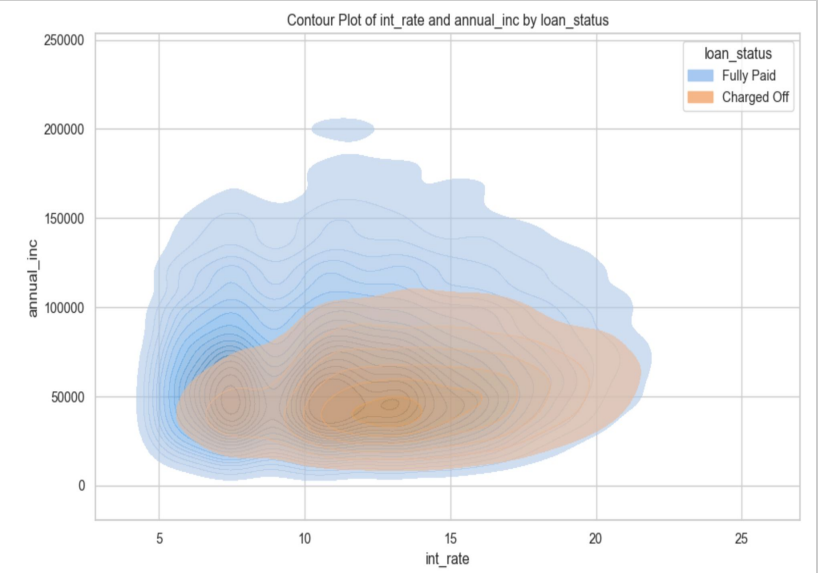
Borrowers with lower annual incomes are slightly more represented in the defaulted segment, as the orange curve peaks slightly higher at lower income levels compared to the blue curve.

For higher income levels (above $100,000), the density of defaulted loans diminishes more rapidly compared to non-defaulted loans, indicating that borrowers with higher incomes are less likely to default.



Annual Income - Interest Rate

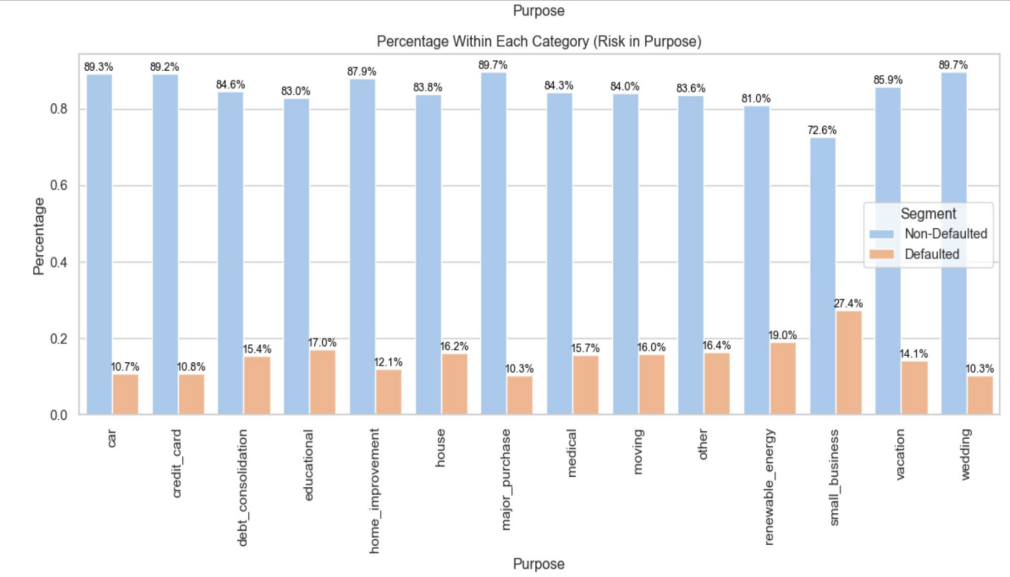
The combination of low income and high interest rates is a critical risk factor for defaults.



Purpose

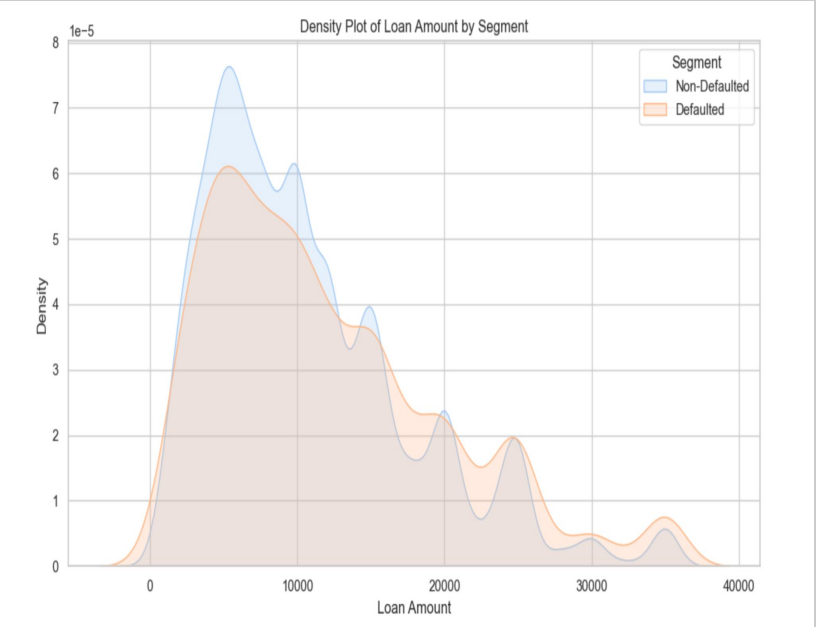
Small business purpose, has proportionately highest number of defaulted loans

Renewable energy comes next in the percentage of defaulted loans



Loan Amount

Higher loan amounts had a higher rate of defaults. The right side of the density curve shows higher default rates(orange) for higher loan amounts.



Recommendations

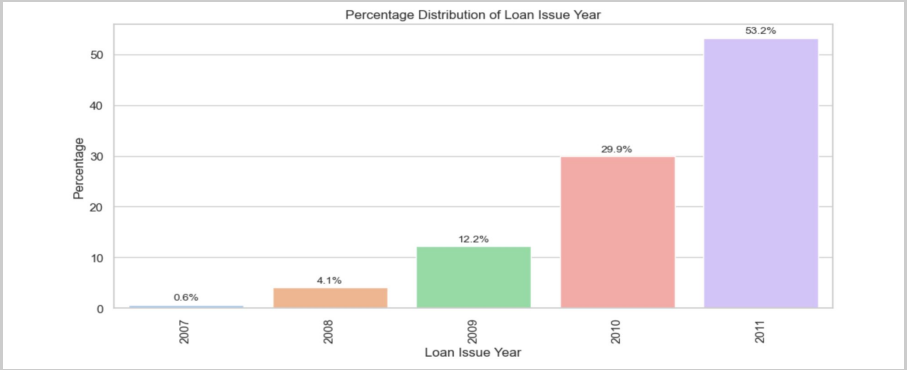
Recommendations

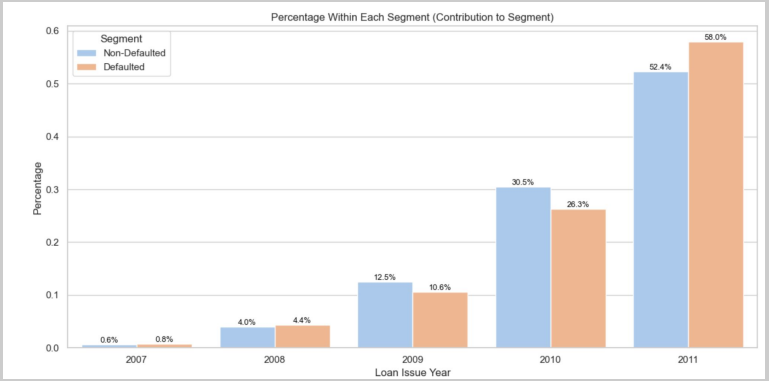
1. Prefer lower DTI and LTI.
2. Avoid significantly high interest rates.
3. Prefer higher grades for loan approvals and reduce volume of lower grade loans.
4. Avoid loans to borrowers with public bankruptcies record greater than 1.
5. Prefer borrowers with no public bankruptcies records.
6. Prefer higher income borrowers.
7. Limit higher loan amounts to high income category only and encourage shorter term loans.
8. Research on root causes for high default rates in purposes - Small Business and Renewable Energy, and mitigate the risk on loans with those purposes.
9. Research on why high default rate happened in Verified Status loans and implement corrections.

Appendix

Additional observations are included for further insights.

Observation #1

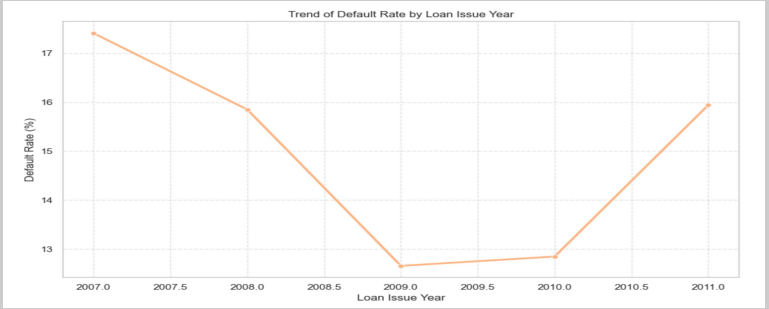




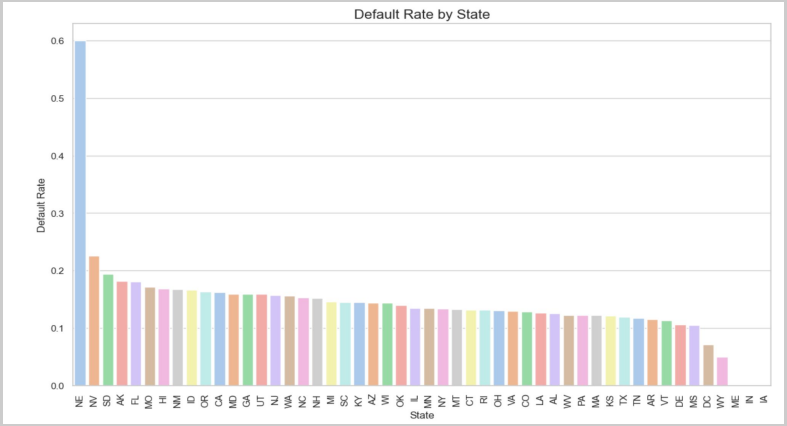
Trend shows exponential increase in loans given every year.

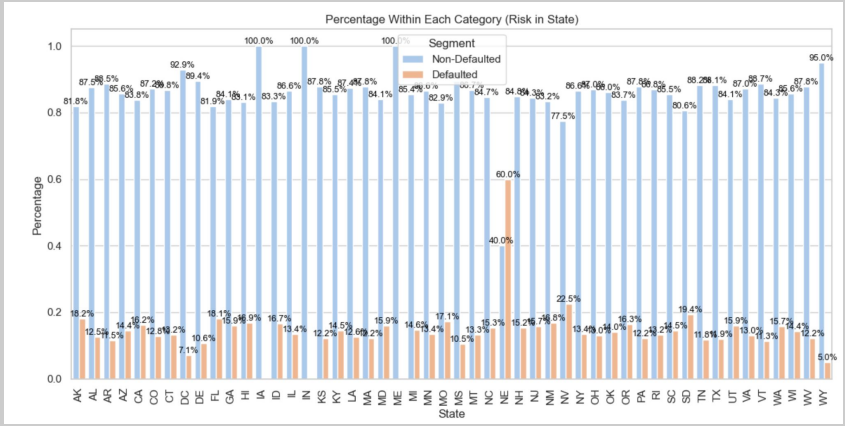
But there is also a similar trend (plots on right side) on increasing default rates.

This indicates a need for more stringent loan approval criteria and risk mitigation strategies.



Observation #2



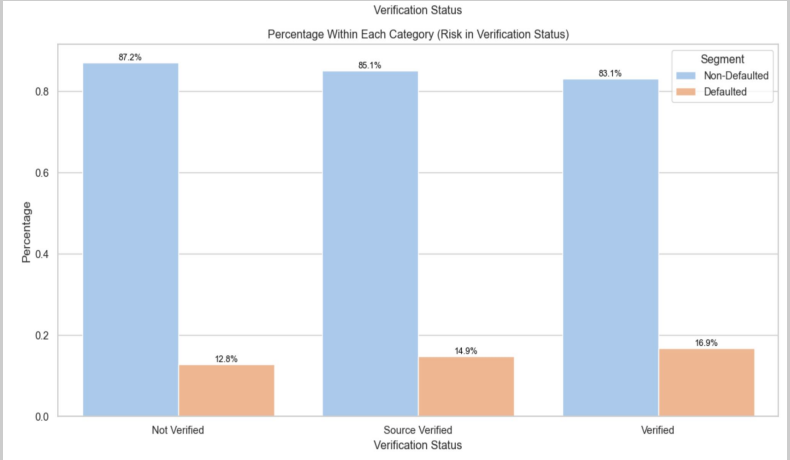


NE (Nebraska) has the highest default rate, followed by NV (Nevada), SD (South Dakota), and AK (Alaska).
These states represent higher financial risk for loans, possibly due to economic conditions or borrower profiles

States such as ME (Maine), IN (Indiana), and IA (Iowa) have the lowest default rates, indicating relatively stable repayment patterns.These states could be safer regions for loan disbursement.

Observation #3

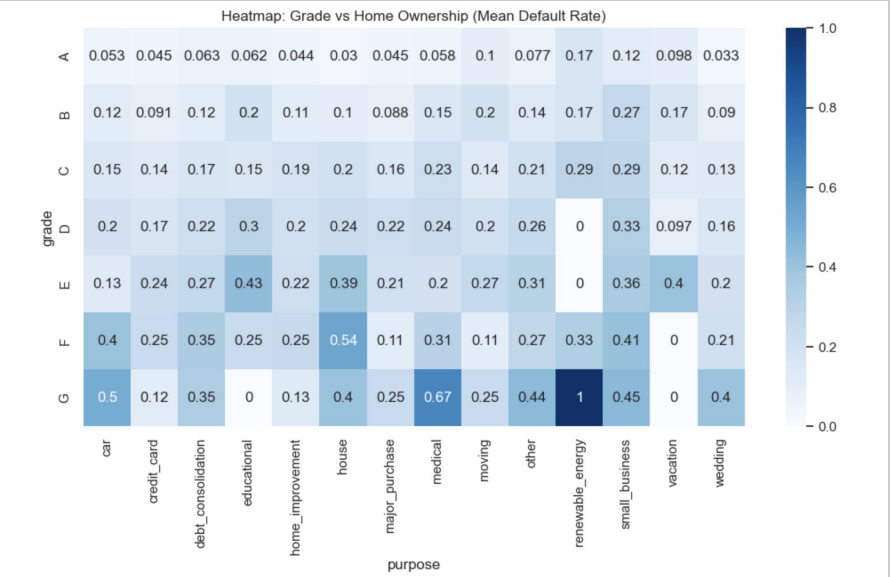
Contrary to expectations, the proportion of Charged Off loans (orange bars) is actually higher in the Verified and Source Verified categories compared to the Not Verified category.



Observation #4

Renewable Energy Loans show extreme default rates in grade G (100%), suggesting significant risk associated with loans for this purpose in the lowest grades.

This may be due to a small sample size, but worth analyzing further

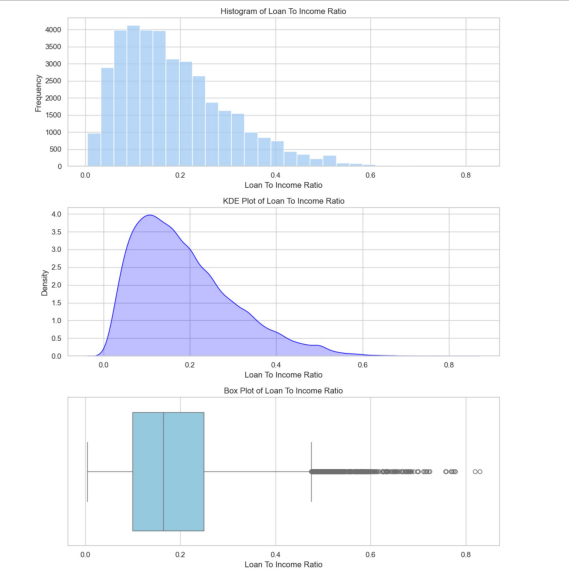


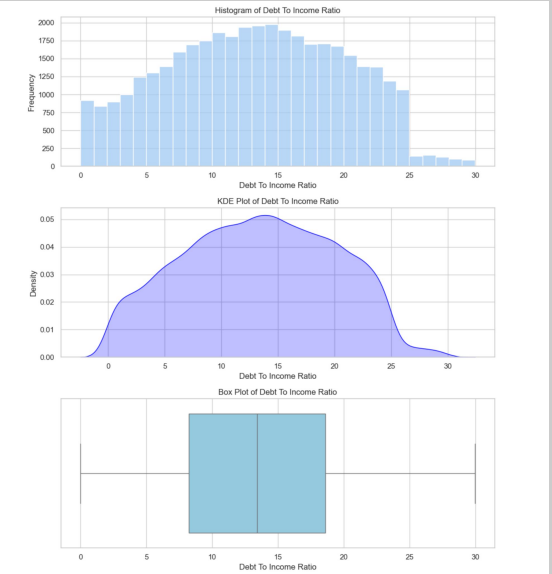
Observation #5

The right skew in the LTI chart shows that majority of the borrowers have low lti ratios, which means most borrowers are taking loans that represent a relatively small proportion of their income from the perspective of the
specific loan amount.

But DTI charts showed that majority had higher debts in total, compared to their income.

Default risk could be brought down if loans are given to borrowers with lesser DTI.

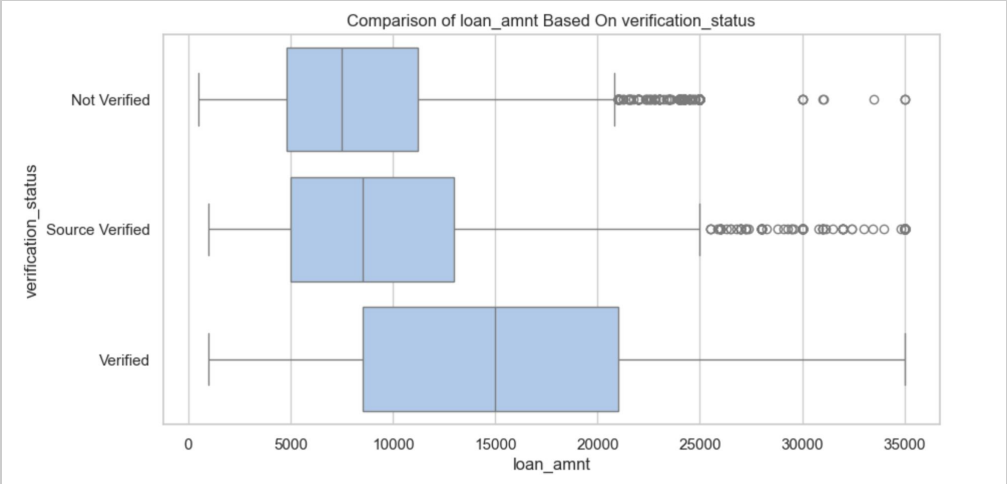




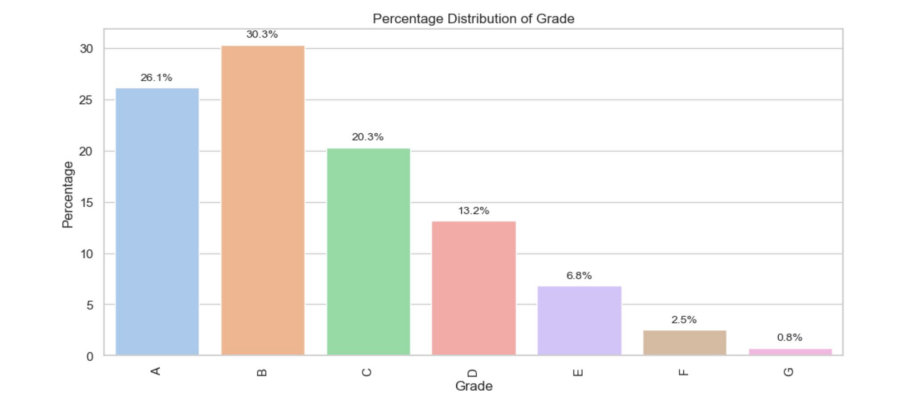
Observation #6

More outliers are present in the "Not Verified" and "Source Verified" categories, especially for larger loan amounts.

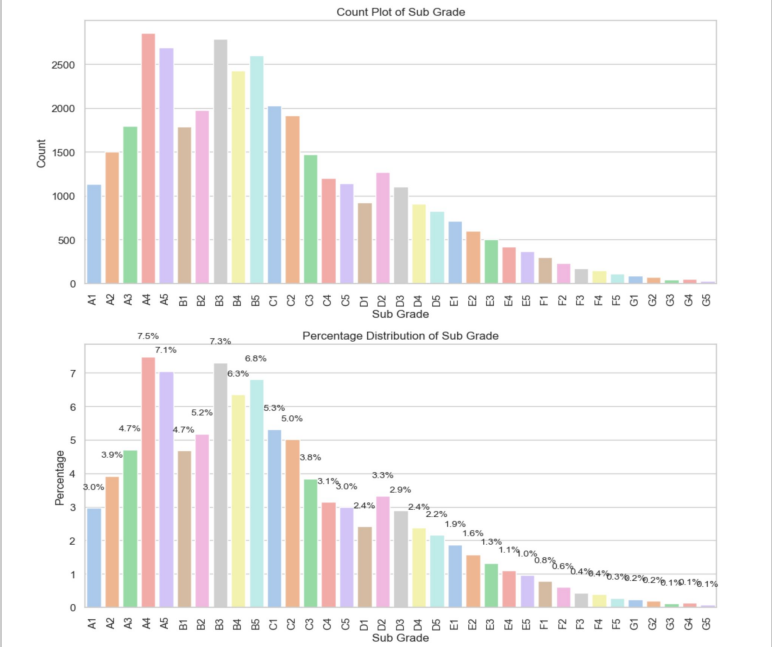
It is worth investigating the root cause of these outliers.



Observation #7



As per the Grade and Sub Grade plots, A Grade loans are lesser compared to B.
Focusing more on A grade loans will reduce risk.



Data preparation summary

1. Columns with >50% missing values were removed from the dataset.
2. Columns with identical values across all rows were excluded.
3. Columns related to post loan related attributes were removed from the dataset.
4. Missing values (Null/NaN) in pub_rec_bankruptcies were assumed to represent no bankruptcies and replaced with 0.
5. Loans with a ‘Current’ status were excluded from the analysis as their default status is unknown and irrelevant to the study.
6. Significant outliers were capped at the 99th percentile to minimize their impact.
7. Missing values in other columns were imputed using industry-standard techniques.

Thanks