Assignment:

Analysis of Lending Club Case Study

Date: 20.11.2024

Presented By:

1. Viraja Rama Gade
2. Neha Garg

Objective:

The Objective of this case study is to implement EDA techniques on a real world problem and understand the insights and present in a business first manner via presentation.

Problem Statement:

The company faces a dual risk when approving loans: losing business by rejecting reliable applicants or incurring financial loss by approving high-risk ones. The objective is to identify patterns of default based on borrowers' past loan histories.

Approach:

1. **Data Preprocessing:** **Cleaning the data**, removing irrelevant columns/variables, Correct inconsistencies.

1. **Conduct Exploratory Data Analysis** using **Univariate**, **Segmented** Univariate, **Multivariate Analysis**

2. **Interpret Results:**

   - Highlight high-risk variables (e.g., grades, interest rates, loan term).
   - Identify borrower segments needing stricter approval or monitoring.
   - Report temporal trends (e.g., year-on-year growth in loans and defaults).

4. Reporting and Visualization: Create visualizations for clarity:

- Bar charts and histograms: For distributions and categorical insights.
- Line graphs: For trend analysis (e.g., default trends over time).
- Heatmaps: For showing correlations.

5. Presentation of Key Findings from the Analysis

Analysis Summary:

Driving factors of Defaulting

Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:

1. DTI
2. Grades
3. Intrest Rates
4. Verification Status
5. Annual income
6. Pub_rec_bankruptcies

Consideration for Default

Borrowers who are considered as "Defaulter"

a. Borrowers who has 10+ years of experience.
b. Who has High debt to Income ratio (DTI).
c. Borrowers who has least grades like F, G & H are likely high risky customers.
d. Borrowers who has Annual Income between 2.5 lakhs to 6 lakhs, are likely to become default.
e. Borrowers who has declared the public Bankruptcy.

Executive Summary

High-Risk Factors

1. Loans with **low grades (C to G)**, higher **DTI ratios (>15%)**, and high-interest rates are more prone to defaults.
2. **Mortgage borrowers** and **Verified users** show higher default risks compared to other groups.
3. Borrowers with **10+ years of employment experience** represent a mixed risk-reward group.

Default Correlations:

1. Defaults are strongly correlated with higher loan amounts, longer loan terms, and borrowers with lower grades or verification inconsistencies.
2. **Economic conditions** (2007–2011) have exacerbated default rates, highlighting the need for adaptive credit policies.

Key Trends:

1. The majority of loans are under ₹1,50,000, with **36-month terms** being most popular.
2. Loan defaults are increasing year-on-year, particularly for **60-month terms** and high-interest loans (>15%).
3. **Debt Consolidation** is the dominant loan purpose, accounting for the highest number of both fully paid and defaulted loans.

Key Recommendations:

1. Prioritize tracking **Loan Amounts**, **Loan Term**, **Interest Rates**, **DTI Ratios**, and **Grades** for risk management.
2. Introduce strict approval criteria for high-risk segments, such as borrowers with low grades, high DTI, and long loan terms.
3. Monitor and review the performance of **Debt Consolidation loans** as they form a significant portion of the portfolio.

How Analysis of Key Variables changed as we moved across different Variat Analysis (1/3)

Variable	Univariate Analysis	Segmented Univariate	Bivariate/Multivariate Analysis
Loan Amount	75% of loans are below ₹1,50,000. - Loans above ₹2,00,000 are rare.	- Smaller loans are fully repaid more often. - Larger loans tend to default more frequently.	Larger loans with higher interest rates and lower grades significantly increase the risk of defaults.
Installments	- Most installments range between ₹100–₹450.	Installments are typically higher for defaulted loans compared to fully paid ones.	Higher installments, combined with higher grades or interest rates, show a noticeable increase in defaults.
Bankruptcy	- 93% of borrowers have no bankruptcy history.	Borrowers without bankruptcy history are less likely to default.	No public bankruptcy record correlates with higher chances of loan approval and full repayment.
Loan Status	- Fully Paid loans significantly outnumber defaults	- Default rates are higher among Verified users.	- Defaults are higher for loans with longer terms, lower grades, and higher interest rates.
Loan Term	- 36-month loans are more popular than 60-month loans.	- 36-month loans have higher fully paid rates. - 60-month loans show higher default rates.	60-month loans combined with low grades or high DTI ratios significantly increase default probabilities.

How Analysis of Key Variables changed as we moved across different Variat Analysis (2/3)

Variable	Univariate Analysis	Segmented Univariate	Bivariate/Multivariate Analysis
Loan Purpose	Debt Consolidation is the most common loan purpose.	- Debt Consolidation loans have both the highest fully paid and defaulted numbers.	- Default risks for Debt Consolidation loans are higher when paired with low grades and higher DTI ratios.
Default Trend		Defaults have been increasing year-on-year from 2007 to 2011.	The rising trend is amplified for loans with high interest rates and longer terms during this period
DTI Ratio		Loans with a DTI ratio of 10–15% have the highest number of defaults.	Higher DTI ratios (>15%) combined with high interest rates drastically increase default likelihood.
Verification Status	More than 50% of loans are issued to Verified or Source Verified borrowers.	Default rates are higher for Verified borrowers compared to Source Verified or Not Verified borrowers.	Verified borrowers with high DTI ratios and low grades show a significant risk of default.
Home Ownership	- Most borrowers do not own a house.	- Mortgage borrowers have higher charge-off rates compared to Renters or Owners.	- Mortgage borrowers with lower grades and higher interest rates are at greater risk of defaults.

How Analysis of Key Variables changed as we moved across different Variat Analysis (3/3)

Variable	Univariate Analysis	Segmented Univariate	Bivariate/Multivariate Analysis
Employment Experience	Borrowers with 10+ years of experience account for most loans.	Borrowers with 10+ years of experience are likely to both fully repay loans and default more frequently.	Default risks increase for 10+ years of experience when combined with higher loan amounts and low grades.
Interest Rate	Debt Consolidation is the most common loan purpose.	- Loans with higher interest rates (>15%) show increased default density.	Higher interest rates amplify default risks when paired with longer terms or low borrower grades.
Loan Grades	- Most loans are Grade A or B.	- Higher grades (A and B) have fewer defaults; Grades C to G exhibit more defaults.	- Grade G loans with high DTI ratios and high interest rates are extremely risky for defaults.
Loan Growth	Loan issuance doubled annually from 2007 to 2011.		- Default risks grew alongside loan issuance, especially for larger loans with higher interest rates during this period.

Key analysis of the variables is as follows: (1/2)

Loan Amount

Large loans (above ₹1,50,000) have higher default risks, especially when paired with high interest rates and low borrower grades.

Loan Term

Loans with a **60-month term** have a significantly higher likelihood of defaults compared to 36-month loans

Interest Rate

Loans with interest rates above **15%** are more likely to default, especially when combined with other risk factors like low grades or high DTI ratios.

Debt-to-Income (DTI) Ratio

Borrowers with a DTI ratio of 10-15% are at higher risk of default. This risk increases further for ratios exceeding 15%.

Loan Purpose

Debt Consolidation loans dominate both fully paid and defaulted loans. Monitoring this segment closely is critical, as it carries high volume and risk.

Loan Grades

Lower grades (C to G) are associated with significantly higher default risks.

Key analysis of the variables is as follows: (2/2)

Verification Status

Borrowers with Verified status default more frequently compared to those with Source Verified or Not Verified statuses.

Home Ownership

Mortgage borrowers show higher default rates compared to Renters or Owners

Employment Experience

Borrowers with 10+ years of experience have higher probabilities of both full repayment and default, depending on other factors.

Default Trend

Defaults have shown an increasing trend from 2007 to 2011. This indicates a need to tighten credit policies or track macroeconomic conditions influencing defaults

Key Findings across different variat analysis

Key Findings Across different variate Analysis (1/2)

Key Findings of Univariate Analysis:

1. Loan Amount:

75% of loans are below ₹1,50,000.

Loans exceeding ₹2,00,000 are rare.

Installments:

Most installment amounts range between ₹100 and ₹450.

3. Bankruptcy:

93% of borrowers have not declared bankruptcy, making them a lower-risk group.

4. Loan Status:

Fully Paid loans significantly outnumber Defaulted loans.

5. Loan Term:

36-month loans are more popular compared to 60-month loans.

Key Findings of Segmented Univariate Analysis

Loan Term:

36-month term: More users fully repay their loans.

60-month term: Higher proportion of defaults.

Loan Purpose:

Debt Consolidation is the most common purpose for both the highest number of fully paid and defaulted loans.

Default Trend:

Defaults show an increasing trend year-on-year from 2007 to 2011.

DTI Ratio:

Loans with a Debt-to-Income (DTI) ratio of 10–15% have a higher number of defaults.

Higher DTI ratios further increase the likelihood of defaults.

Verification Status:

Default rates are higher among "Verified" users compared to "Source Verified" or "Not Verified" borrowers.

Home Ownership

Mortgage borrowers exhibit higher charge-off rates.

Key Findings Across different variate Analysis (2/2)

Key Findings of Bivariate / Multivariate Analysis

1. Loan Grades:

Higher grades (A and B) have fewer defaults; Riskier grades (C to G) exhibit higher default rates.

2. Interest Rates:

Loans with interest rates above 15% show a noticeable increase in default density.

3. Bankruptcy:

Borrowers without public bankruptcy records are safer loan candidates.

4. Employment Experience:

Borrowers with 10+ years of experience account for the majority of loans.

5. Home Ownership:

Most borrowers do not own a house, indicating reliance on income stability rather than collateral.

6. Verification Status:

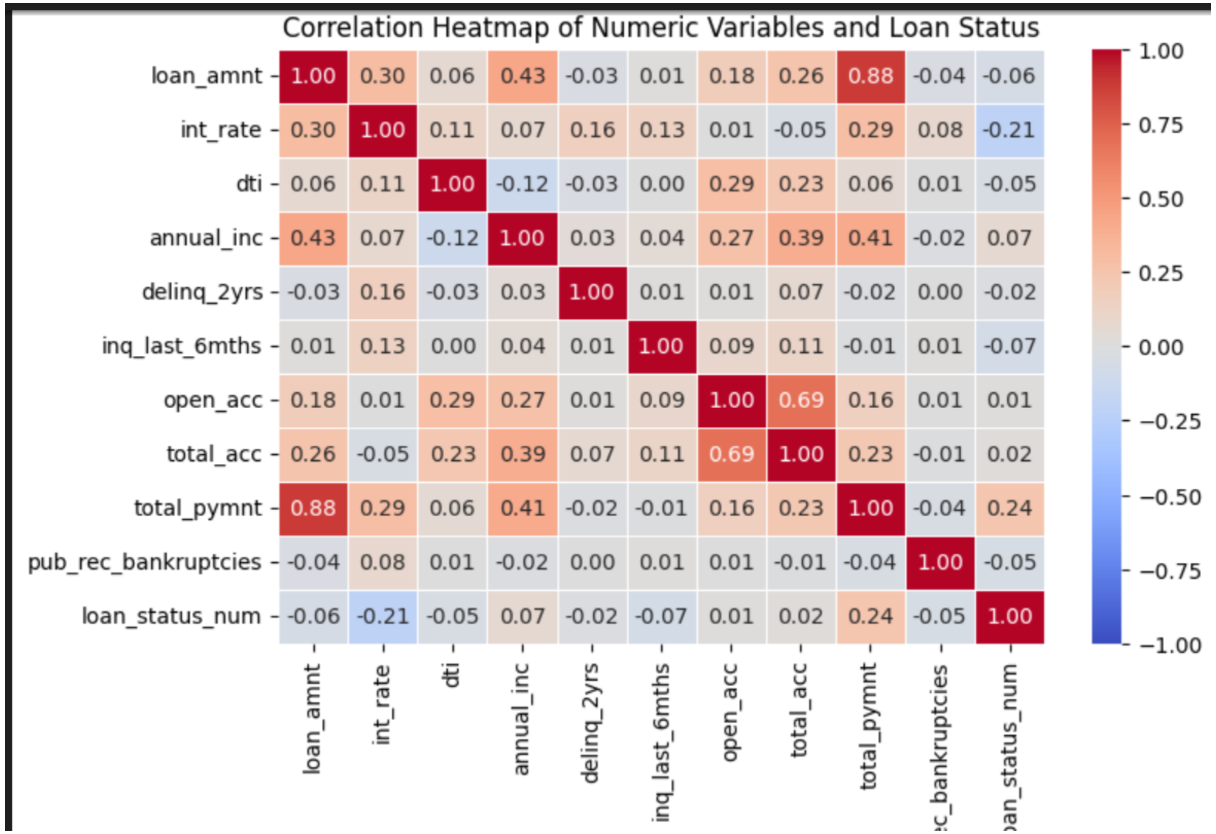
Over 50% of loans are issued to "Verified" or "Source Verified" users.

7. Loan Growth:

Loan issuance doubled annually from 2007 to 2011, indicating increased lending activity.

Some Key Graphical Analysis

Correlation Heatmap of Numeric Variables and Loan Status



1. **Loan Amount:**

   Strong positive correlation with total payment(0.88) which means higher loan amounts obviously have higher total payments; Weak negative correlation with Loan status. Slightly suggests that higher loan amounts may be less likely to default.

2. Interest Rate:

Negative correlation with loan status (-0.21) Highere interest rates are defaulting more

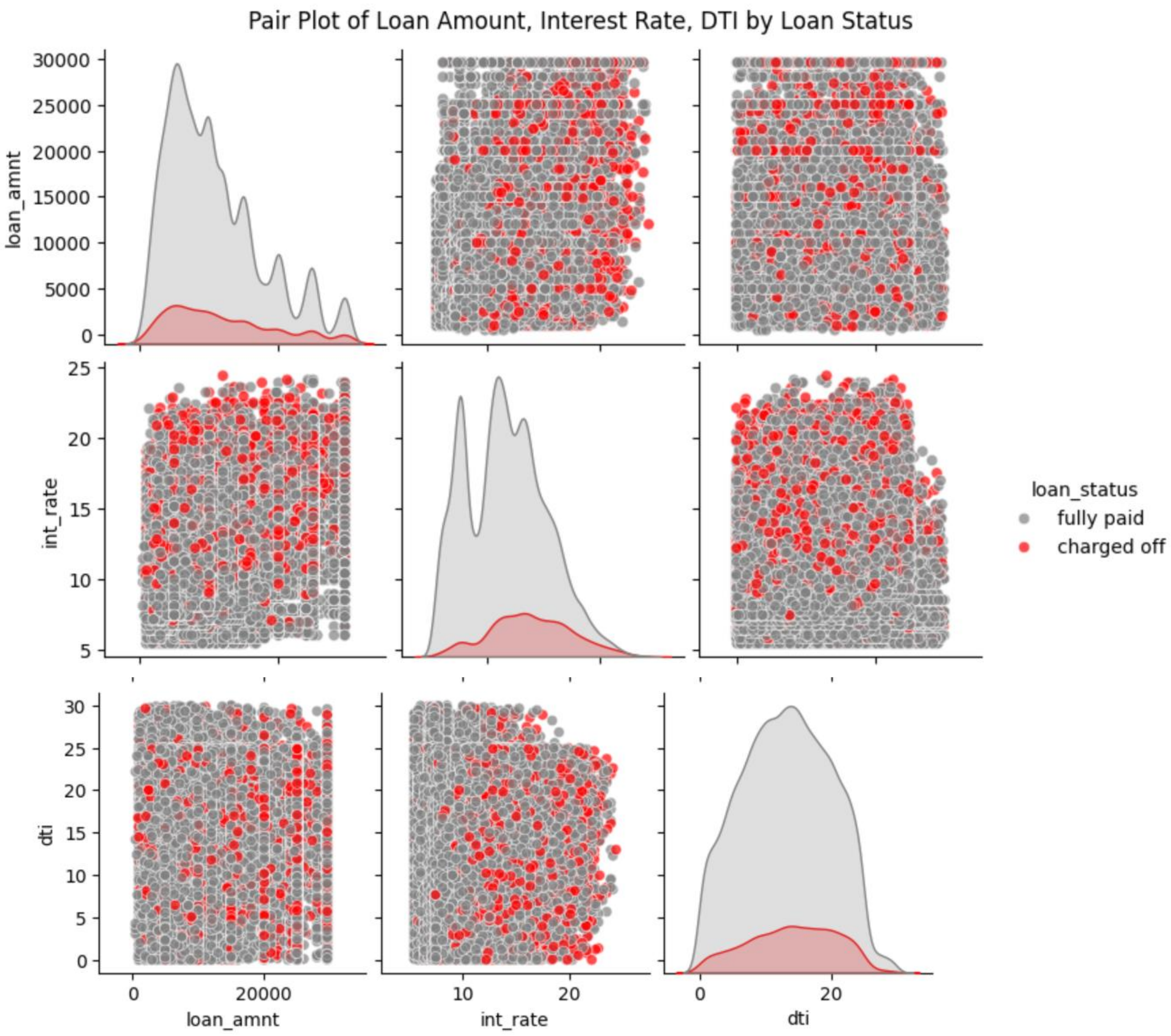
3. DTI: Very weak correlation with loan status (-.05) almost no impact on loan status directly

4. Annual income, no of open acc, public records have not much impact on loan status

5. Interst rates and loan amounts have decent correlation as expected conceptually.

Among all, interest rate (-0.21) is strongest, indicating high int_rate is a notable factor for default.

Pair Plot of Loan Amount, Interest Rate, DTI by Loan Status



1 Above plot Observations:

1. Loan Amount vs. Interest Rate:

Higher interest rates tend to be associated with smaller loan amounts.

Charged Off loans are more concentrated in the upper-interest regions, regardless of loan amount.

2. Interest Rate vs. DTI:

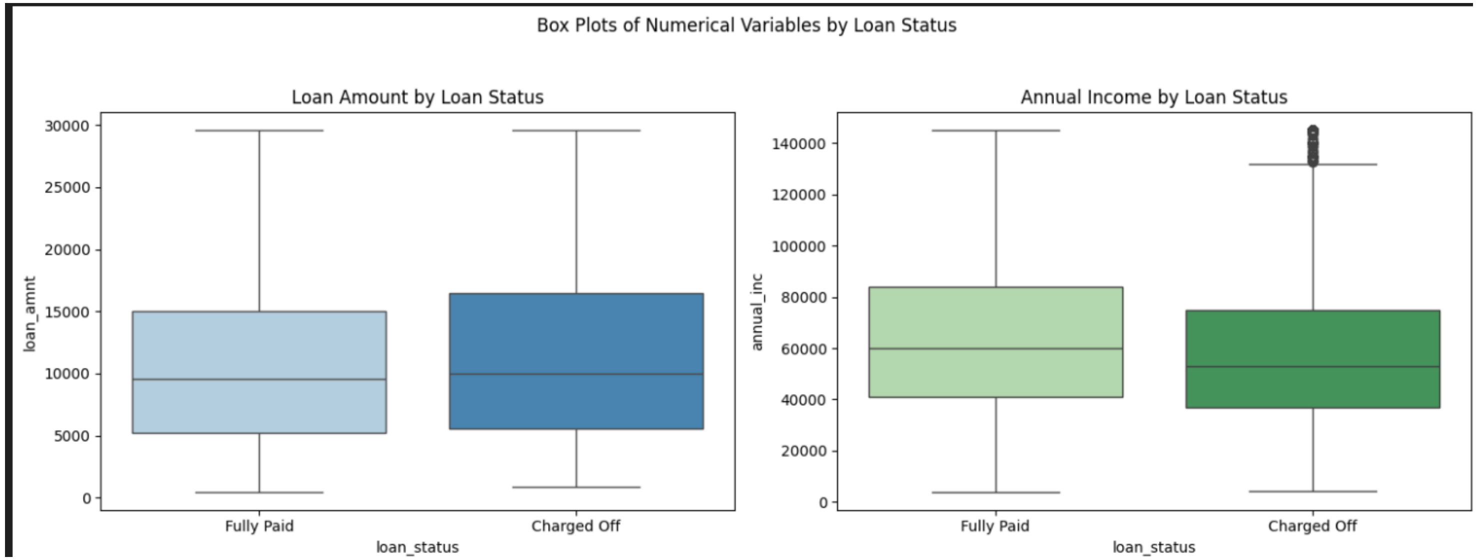
Similar to the first plot, higher interest rates combined with moderately high DTI values (>15) show an increased likelihood of loans being Charged Off.

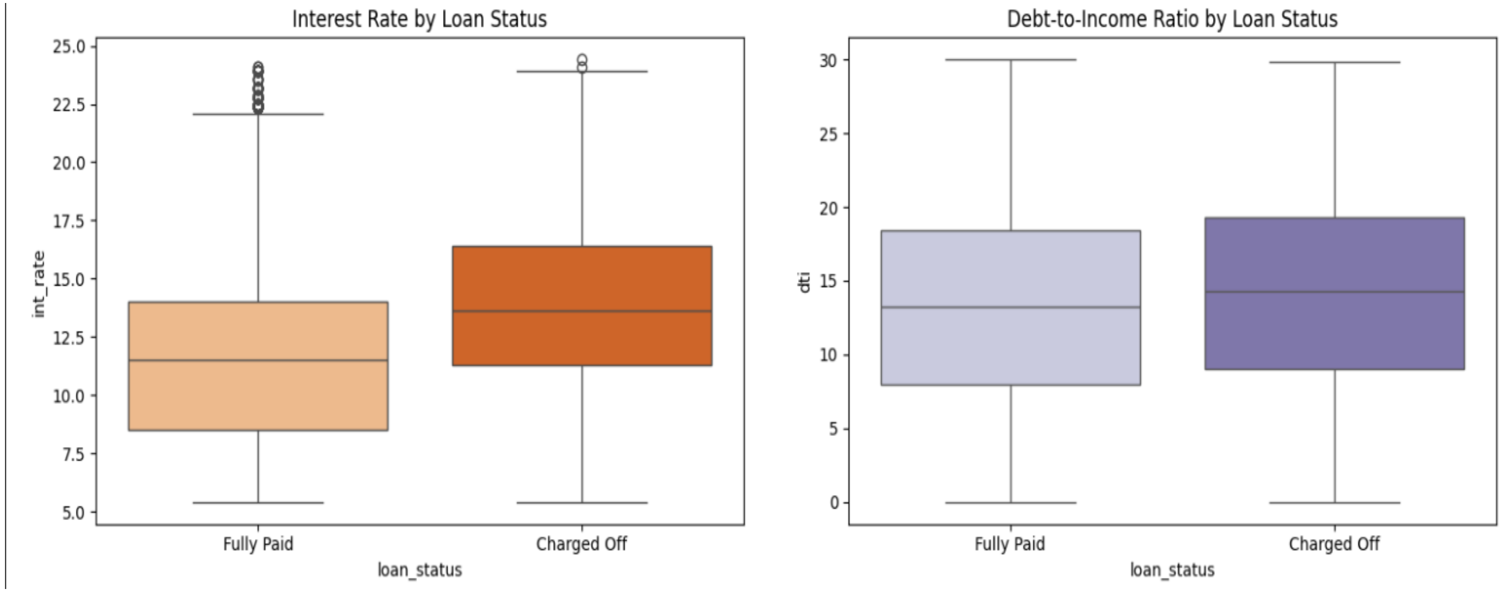
3. Note:

Interest Rate is a Key Predictor: Higher interest rates are strongly associated with loan defaults (Charged Off loans).

2. Loan Amount May Play a Role: While defaults occur across all loan amounts, there may be slightly higher risks for lower loan amounts combined with higher interest rates

Box Plots of Numerical Variables by Loan Status





As seen here:

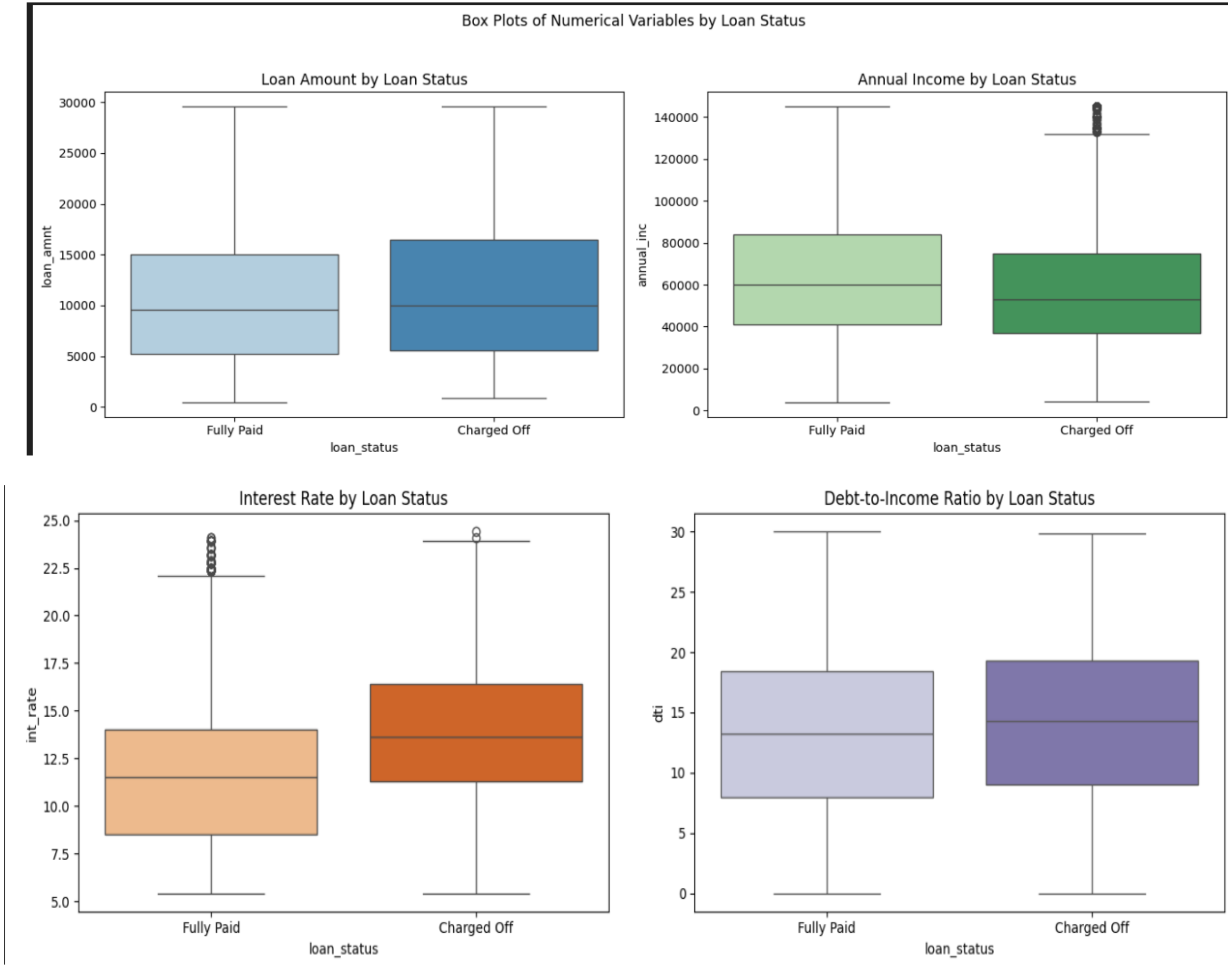
income and interest rates appear to have strong correlations with loan status.

Borrowers with low income and high interest rates are more likely to default.

Loan amount and dti has less impact

Annexures

Box Plots of Numerical Variables by Loan Status



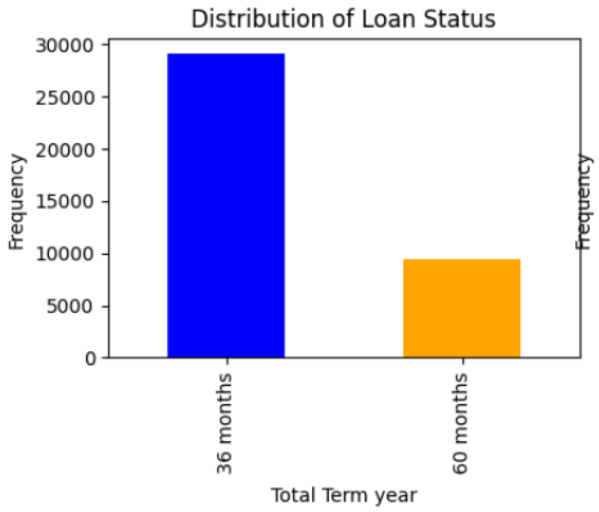
As seen here:

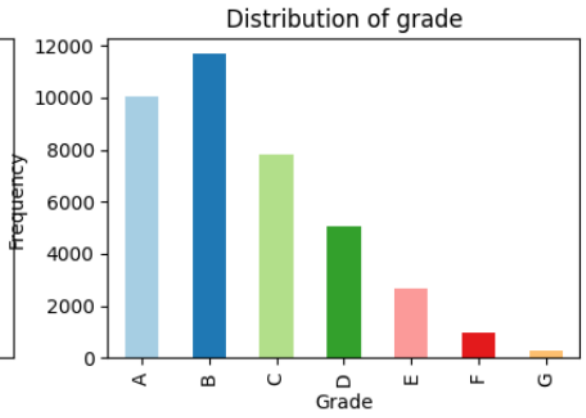
income and interest rates appear to have strong correlations with loan status.

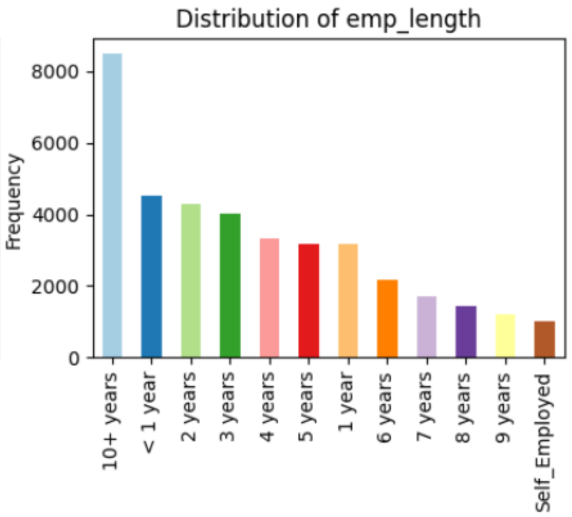
Borrowers with low income and high interest rates are more likely to default.

Loan amount and dti has less impact

Other Key Graphs:

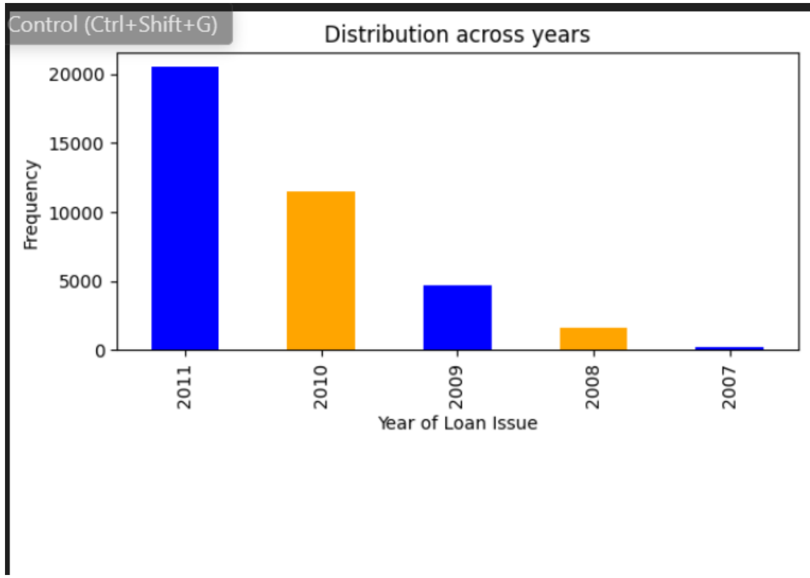


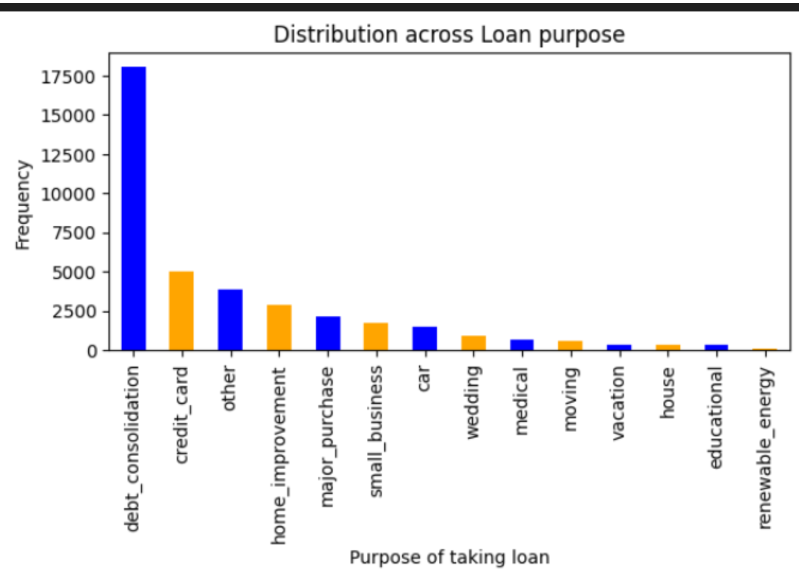




Observation :

1. 36 Months loan term has been taken by most of the people where 60 months is less
2. Maximum loans were taken by people who has 10+ years of experience.
3. Most loans has grade A & B compare to others. Means higher grades loans were taken.

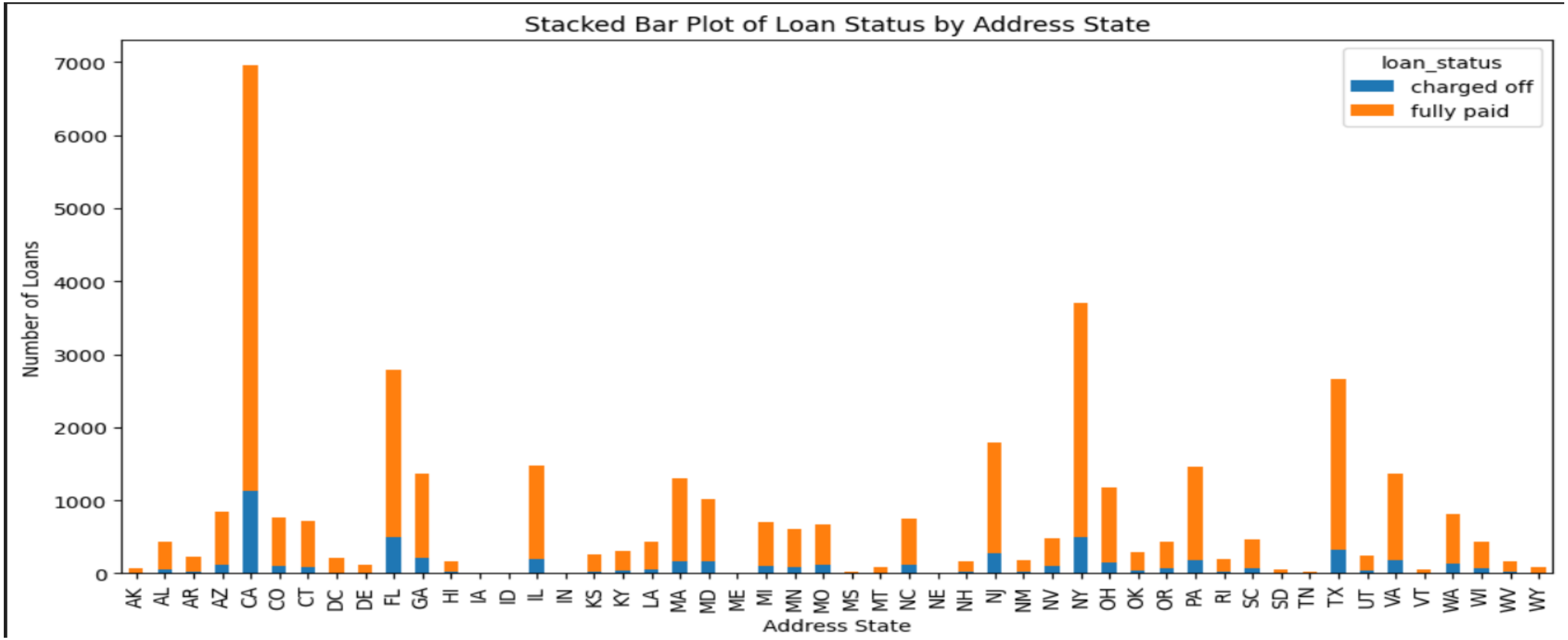




Observation :

1. Loan issue is getting doubled each year from 2007 to 2011.
2. More number of people took loan due to debt_consolidation.

Other Key Graphs:



Observation :

1. **California has highest count with fully paid as well as charged off. How ever the proportion seems to be close to consistent across most states.**

2. **California, Florida, Texas, New York have high loan volumes. These states might have more factors contributing to defaults.**