Loan Lending Club Case Study

October 2024

Risk Prediction Analysis

Predicting Loan Default Risks Using Univariate, Bivariate, Segmented, and Multivariate Analysis

Lending Club is a peer-to-peer lending company that provides loans to borrowers. As default risks increase, it's crucial for the company to identify factors that contribute to loan defaults to minimize losses. The company thus wants to predict whether a borrower will default on their loan, based on historical data of previous loans. Our objective is to develop a model to predict the risk of default and provide actionable insights into borrower profiles that are more prone to default.

Problem Statement

Our first objective is to identify key factors that influence loan default and create a predictive model to assess the likelihood of loan default based on these factors. Our secondary objectives include conducting exploratory analysis (univariate, bivariate, segmented) to uncover patterns and relationships, performing multivariate analysis to create a predictive model and providing recommendations for minimizing loan default risks.

Objective

The Lending Club dataset contains information on loans issued, including borrower characteristics, loan amount, interest rates, loan status, and more through the time period 2007 t0 2011.

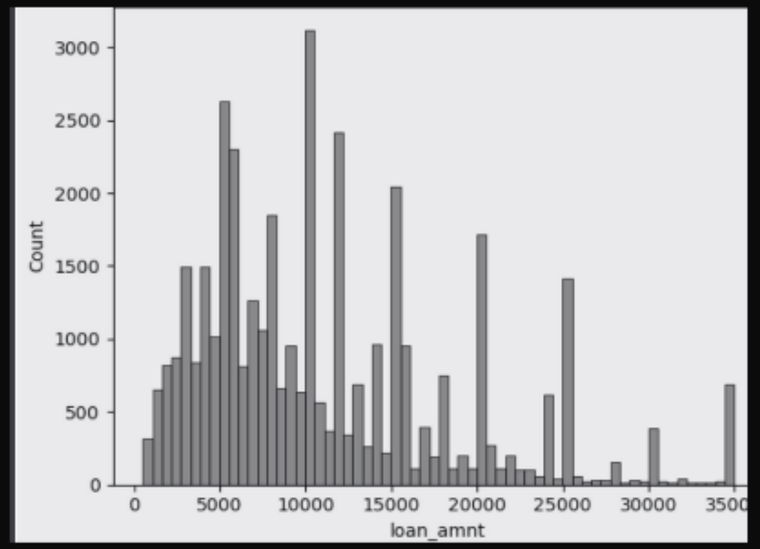
Data

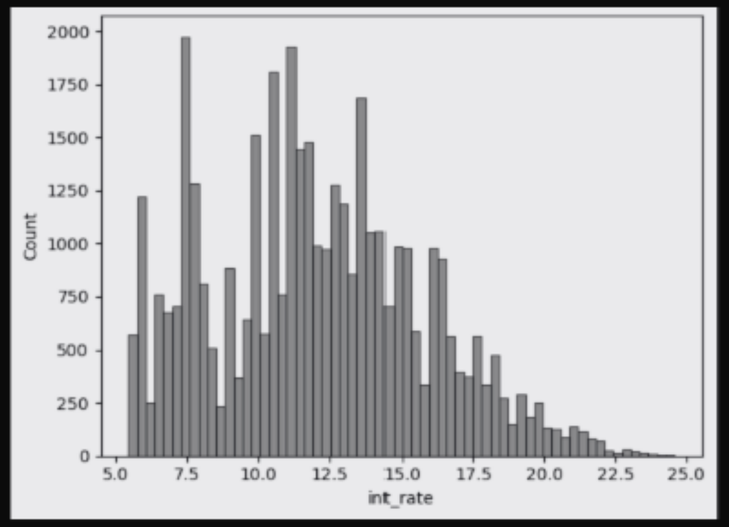
The methodology includes data preparation, where we loaded the dataset, conducted initial exploration, addressed missing values, and made adjustments to specific columns. We converted date fields, derived new columns, and dropped unnecessary ones. After preparing the final dataset, we performed univariate analysis on variables such as loan amount, interest rate, annual income, debt-to-income ratio, home ownership, and loan status. We also derived new metrics like the credit utilization rate, installment-to-income ratio, delinquency risk index, loan payment-to-principal ratio, and loan charge-off rate. Next, we conducted segmented analysis by employment length, income level, region, and loan purpose. In bivariate analysis, we explored relationships like credit utilization vs. loan status, installment-to-income ratio, and delinquency risk index. Lastly, we performed multivariate analysis to examine interactions between variables like loan amount, interest rate, and debt-to-income ratio.

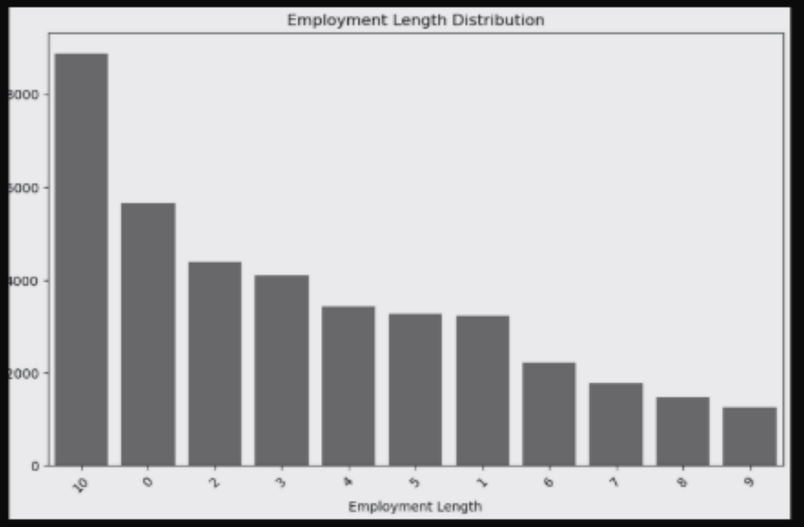
Methodology

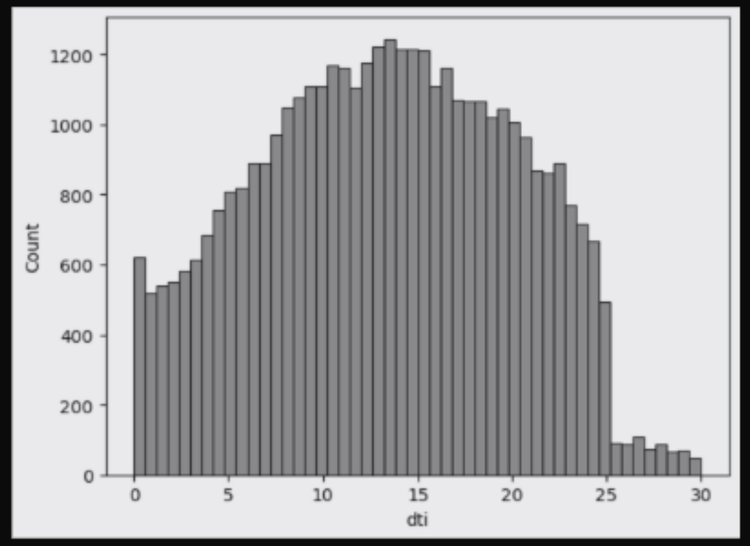
The majority of loan amounts ranged between $5,000 and $15,000, with most borrowers charged an interest rate around 7.5%. The Debt-to-Income Ratio (DTI) peaked at 15, suggesting moderate borrower indebtedness, though higher DTIs could signal potential risk. California had the highest number of loan applications, reflecting its large population and economic activity. Regarding homeownership, most borrowers were renters or had mortgaged properties, with a small portion owning their homes outright. Grade B loans were the most common, indicating moderate risk, and borrowers typically had around 10 years of employment, suggesting a stable employment background. Debt consolidation was the most frequent loan purpose, and most loans had a 36-month term, which is common for consumer loans.

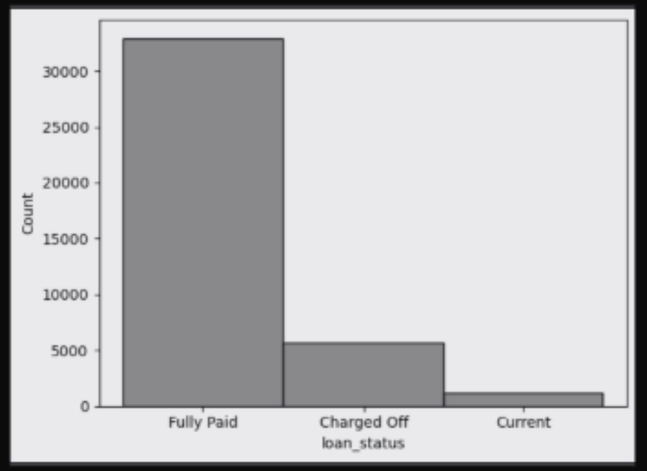
Univariate Analysis

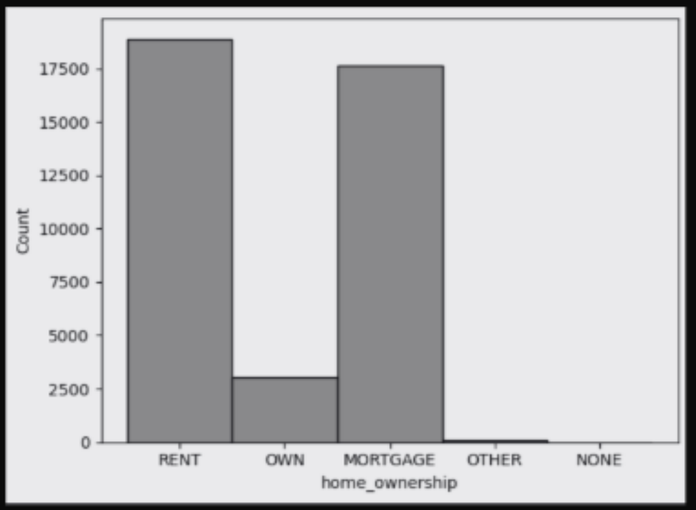

Count
loan_amnt


Count
int_rate


Employment Length Distribution
Employment Length

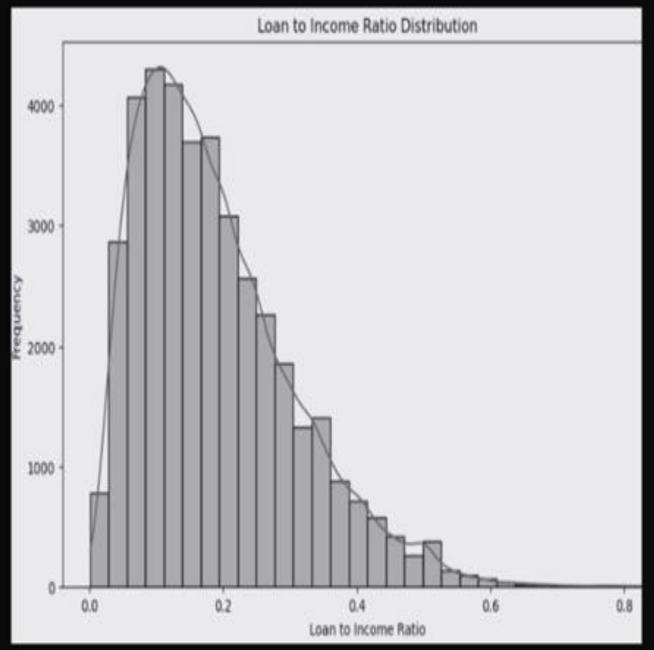

Count
dti

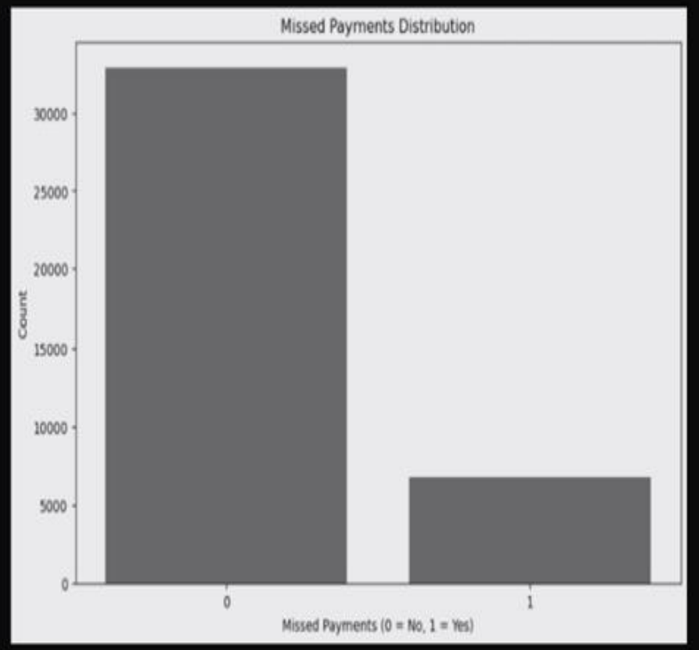

Count
Fully Paid
Charged Off
Current
loan_status

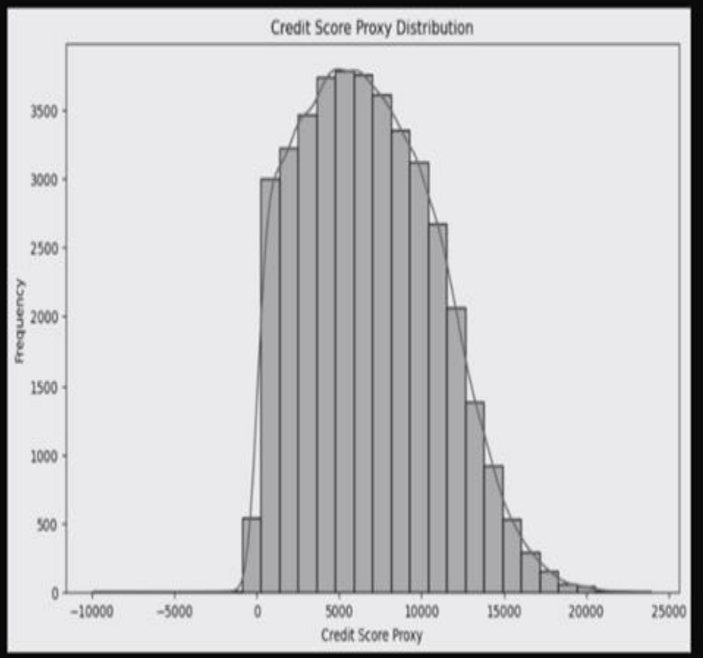

Count
RENT
OWN
MORTGAGE
OTHER
NONE
home_ownership

New metrics were derived to deepen insights into default risks. The Loan-to-Income Ratio followed a bell-shaped distribution, with most values between 0 and 0.2, indicating that borrowers generally did not overextend relative to income. A FICO Proxy Score was created using variables like delinquency and revolving credit, with the highest concentration between 5,000 and 10,000. The Missed Payments metric, based on loan status and late fees, showed that most borrowers did not miss payments, though those who did were heavily skewed toward charged-off loans. The Credit Utilization Rate was around 10,000 for fully paid loans, compared to under 2,000 for charged-off loans, suggesting higher credit usage for borrowers who fully repaid. The Installment-to-Income Ratio was 0.19 for charged-off loans, higher than the 0.16 for fully paid loans, indicating a heavier financial burden on defaulting borrowers. Additionally, the Loan Charge-Off Rate was highest for Grade F loans, establishing a clear link between low credit grades and higher likelihood of default.

Derived Metrices

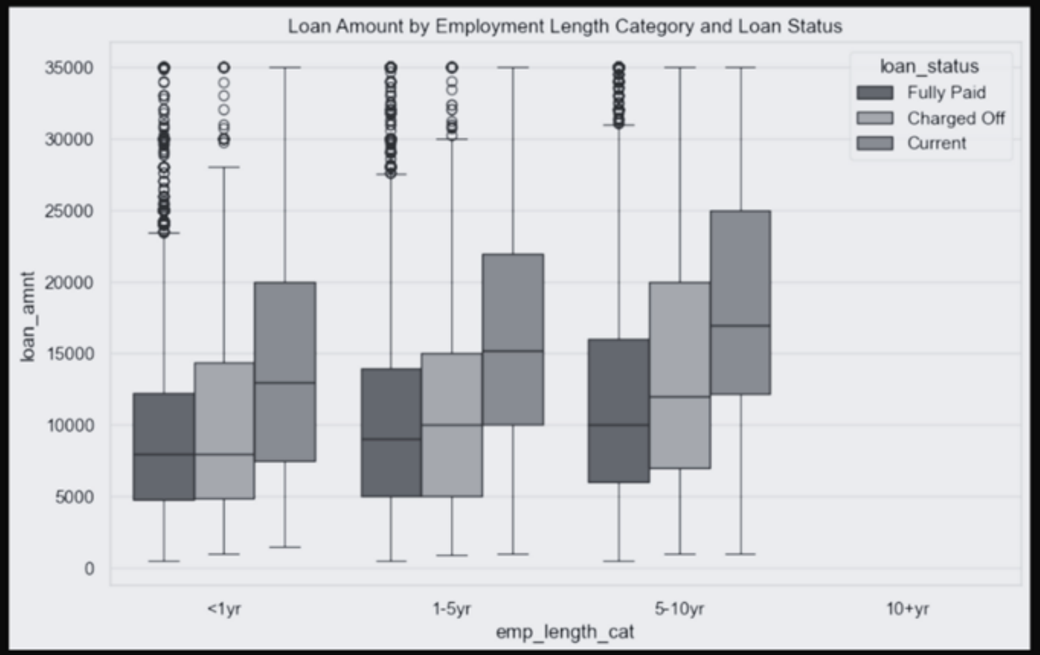

Loan to Income Ratio Distribution
Frequency
Loan to Income Ratio
0
1000
2000
3000
4000
0.0
0.2
0.4
0.6
0.8

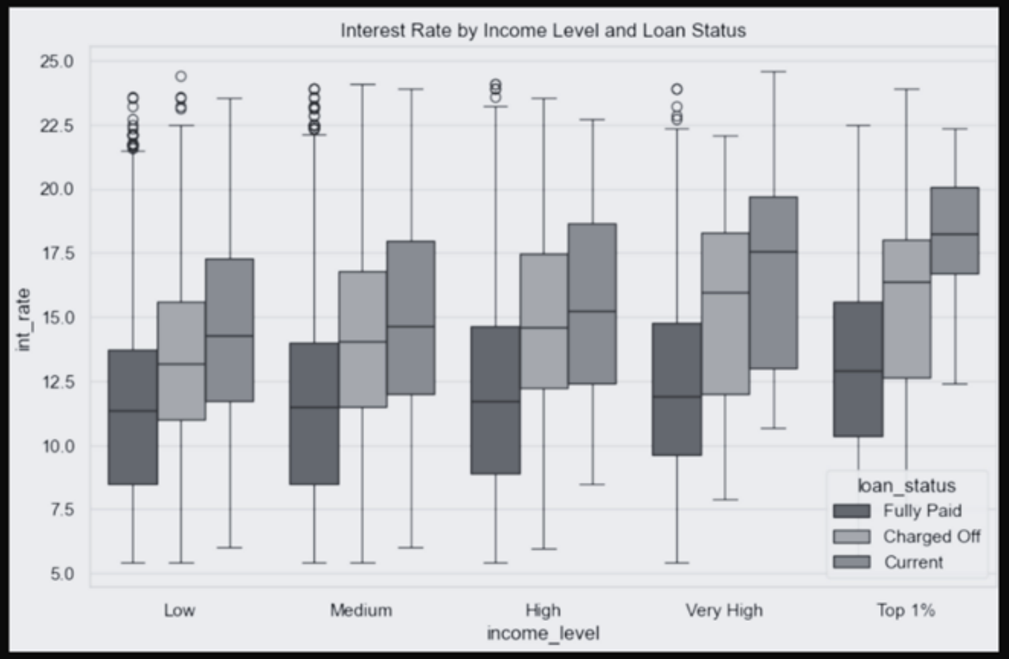

Missed Payments Distribution
Count
Missed Payments (0 = No, 1 = Yes)
0
5000
10000
15000
20000
25000
30000
0
1

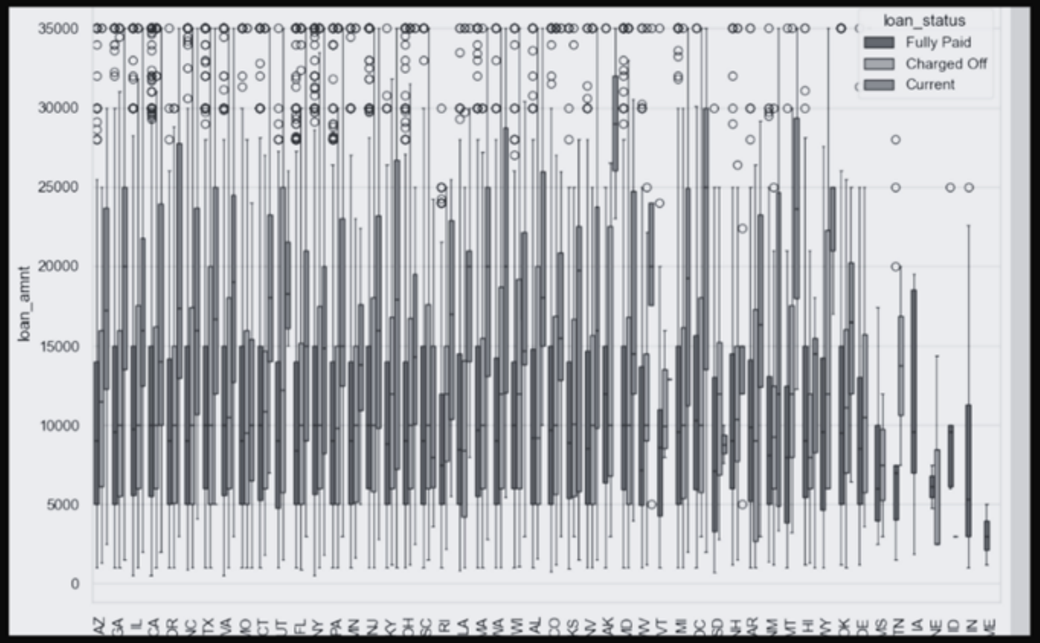

Credit Score Proxy Distribution
Frequency
Credit Score Proxy
0
500
1000
1500
2000
2500
3000
3500
−10000
−5000
0
5000
10000
15000
20000
25000

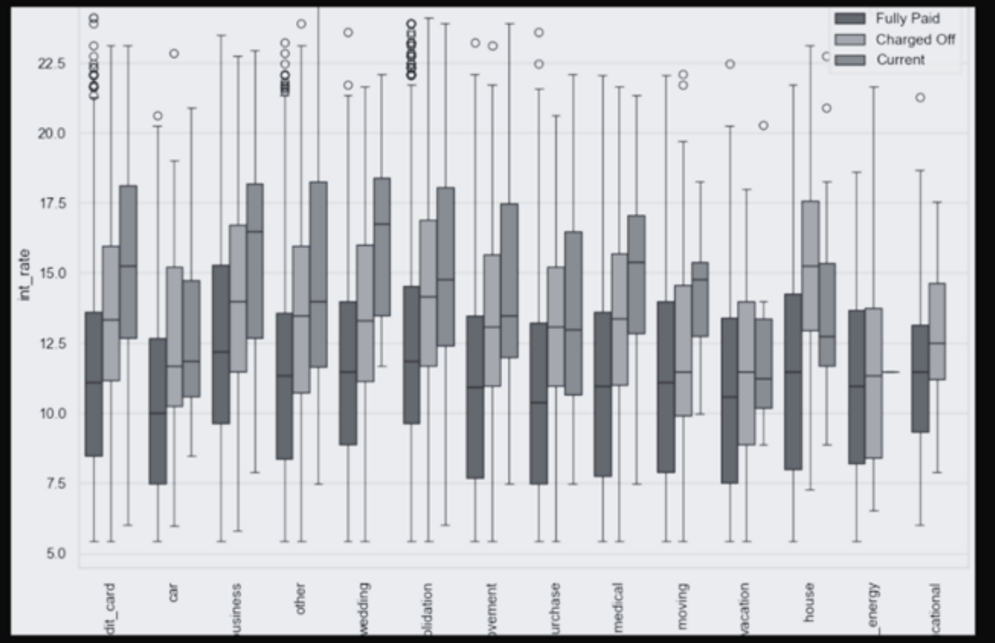
Segmentation of the data revealed that loan amounts were highest for current loans, followed by charged-off and then fully paid loans. Among high-income borrowers, those in the top 1% had the largest charged-off loans, around $25,000, while medium- and low-income groups had loans between $10,000 and $20,000. The interest rates for charged-off loans were notably higher, particularly for current loans (17.5%) and charged-off loans (16%). Utah, Wyoming, and Alaska had the highest loan amounts for charged-off loans, often associated with higher interest rates, emphasizing geographic risk variations. The most common loan purpose for charged-off borrowers was debt consolidation, which consistently had the highest loan amounts. Charged-off loans for home-related purposes carried the highest interest rates (17.5%).

Segmented Analysis


Loan Amount by Employment Length Category and Loan Status
loan_amnt
emp_length_cat
<1yr
1-5yr
5-10yr
10+yr
loan_status
Fully Paid
Charged Off
Current

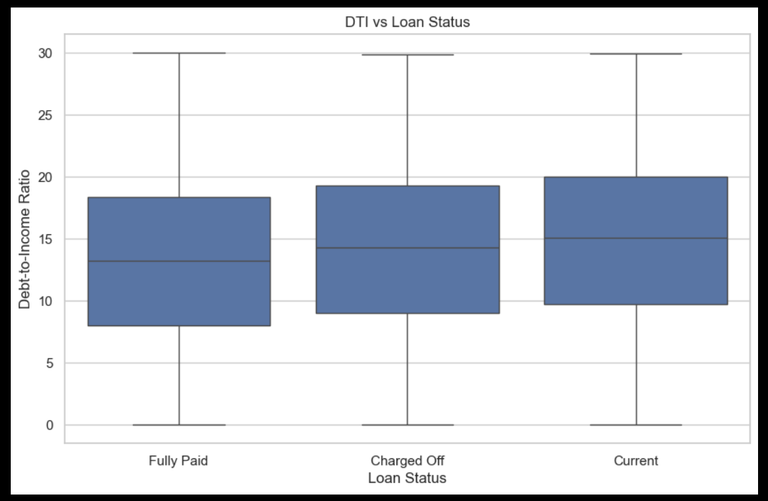

Interest Rate by Income Level and Loan Status
int_rate
income_level
Low
Medium
High
Very High
Top 1%
loan_status
Fully Paid
Charged Off
Current

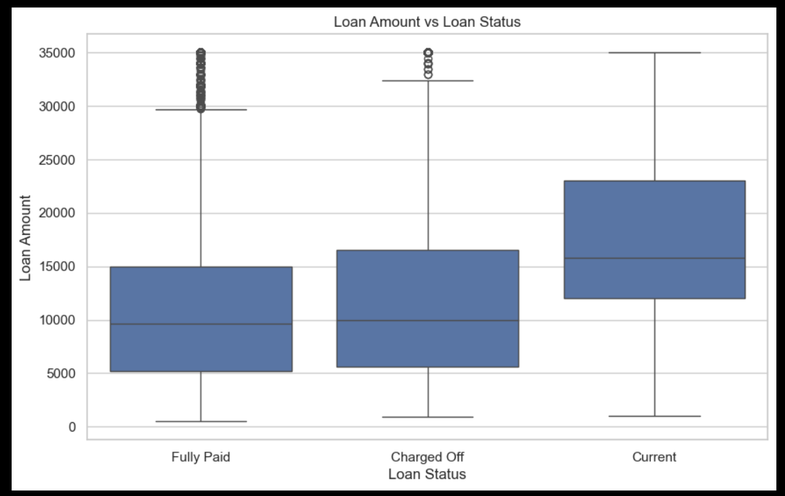

loan_amnt
loan_status
Fully Paid
Charged Off
Current

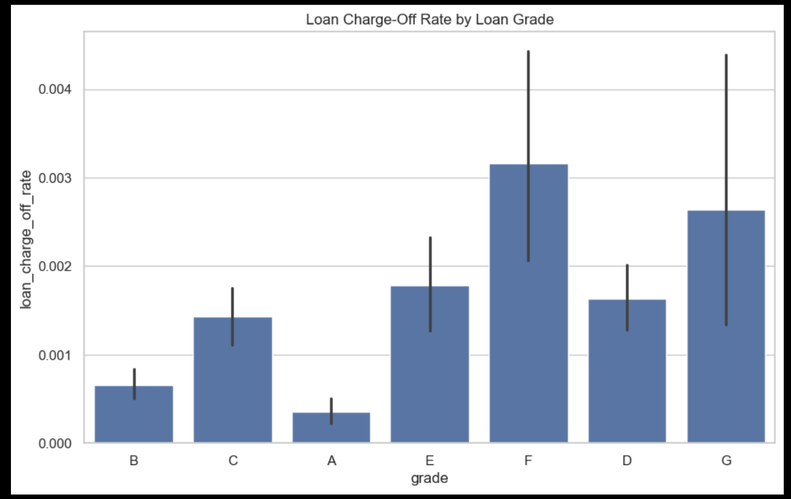

int_rate
Fully Paid
Charged Off
Current
car
other
house
medical
moving
vacation

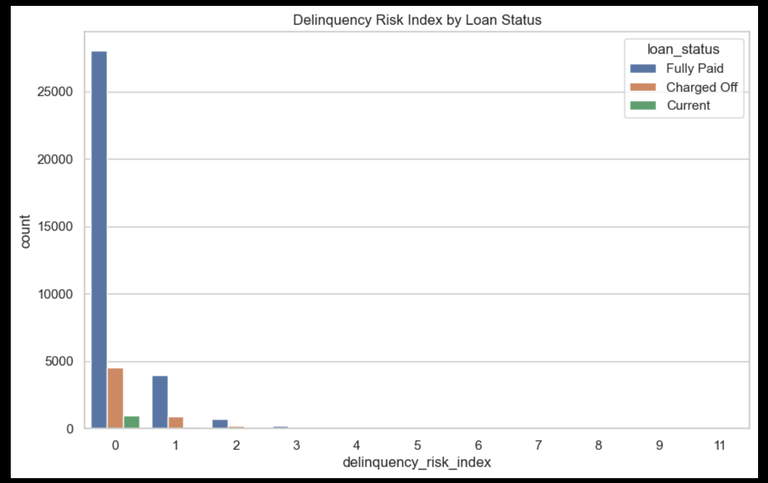
Bivariate analysis identified key risk factors. Interest rates were significantly higher for current borrowers and charged-off loans, signaling that borrowers who struggled with repayments were charged at risk-adjusted higher rates. Employment length was a strong indicator of default risk, with borrowers having shorter employment histories (around 1 year) more likely to default. This trend suggests that job stability plays a critical role in loan repayment success. While the DTI ratio did not show large variations across loan statuses, lower credit grades consistently had higher DTI ratios, particularly Grade F and B. Homeownership was more common among fully paid borrowers, while charged-off borrowers predominantly rented, indicating that homeownership may correlate with financial stability. Borrowers with 36-month loan terms generally fared better, with higher default rates linked to longer loan terms.

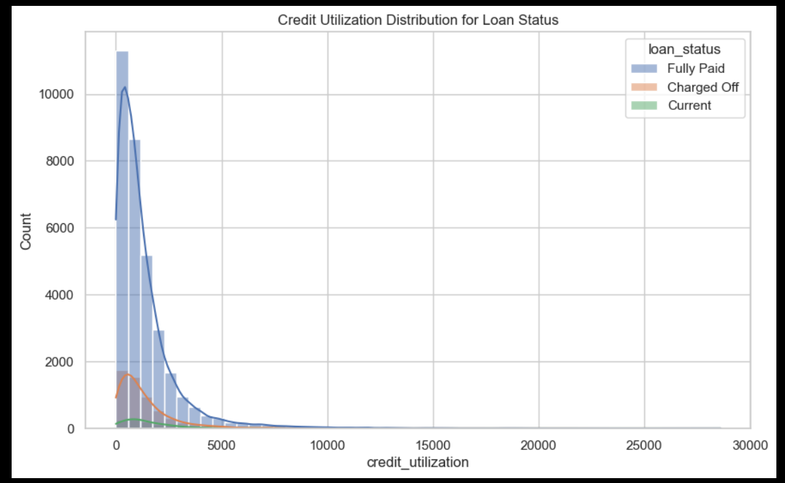
Bivariate Analysis

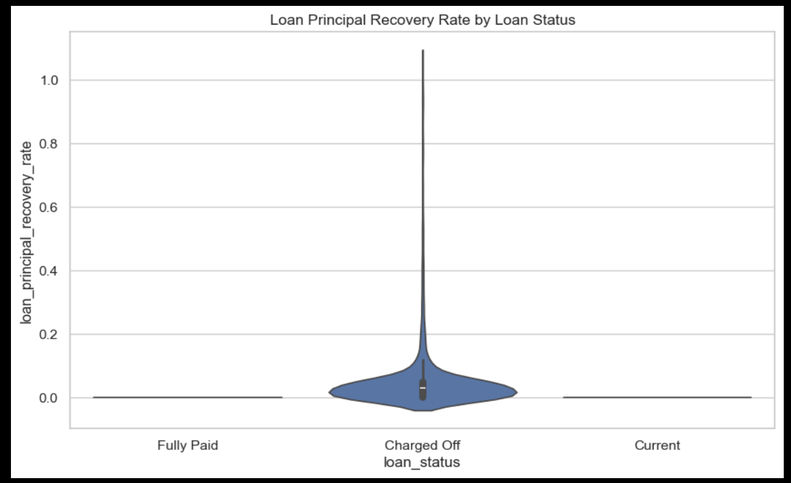

DTI vs Loan Status
Debt-to-Income Ratio
0
5
10
15
20
25
30
Fully Paid
Charged Off
Current
Loan Status


Loan Amount vs Loan Status
Loan Amount
0
5000
10000
15000
20000
25000
30000
35000
Fully Paid
Charged Off
Current
Loan Status


Loan Charge-Off Rate by Loan Grade
loan_charge_off_rate
0.000
0.001
0.002
0.003
0.004
B
C
A
E
F
D
G
grade

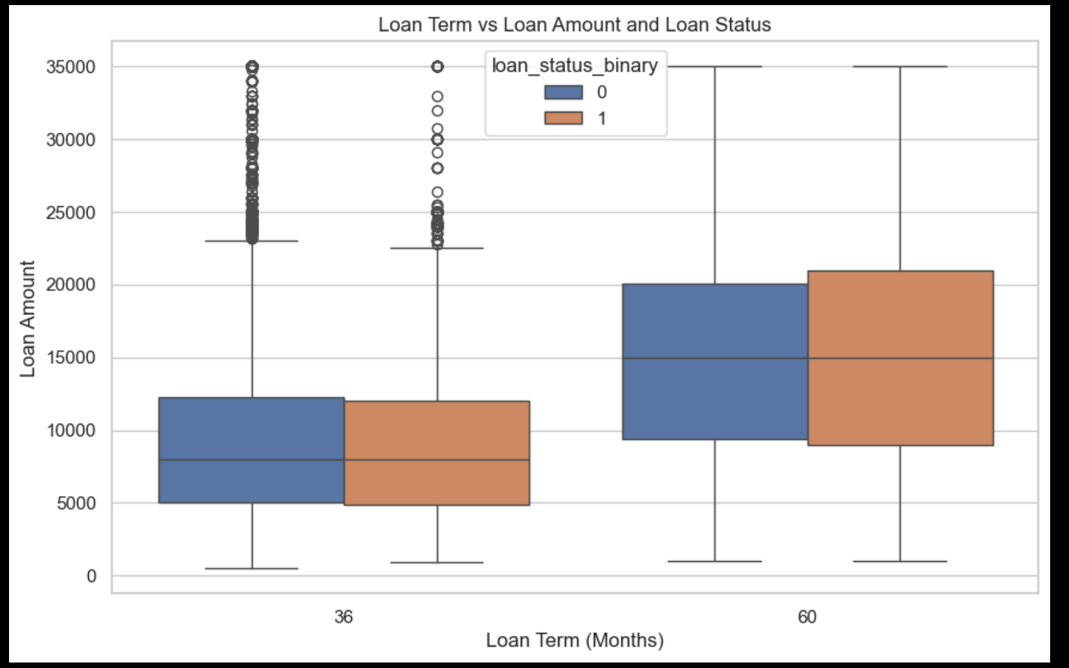

Delinquency Risk Index by Loan Status
loan_status
Fully Paid
Charged Off
Current
count
0
5000
10000
15000
20000
25000
0
1
2
3
4
5
6
7
8
9
11
delinquency_risk_index

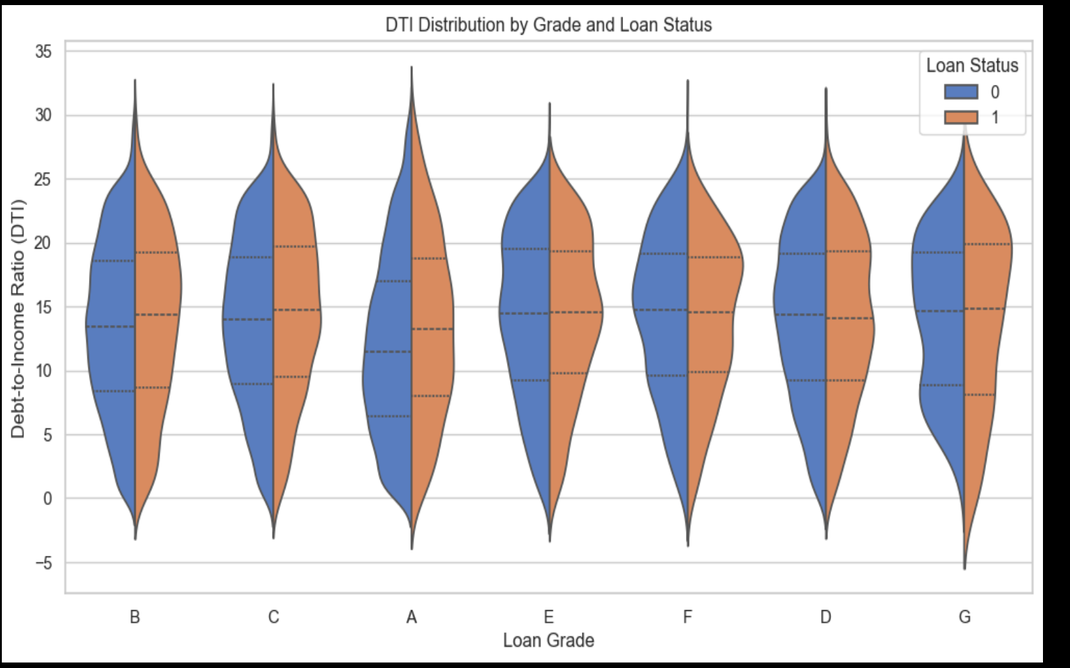

Credit Utilization Distribution for Loan Status
loan_status
Fully Paid
Charged Off
Current
Count
0
2000
4000
6000
8000
10000
0
5000
10000
15000
20000
25000
30000
credit_utilization

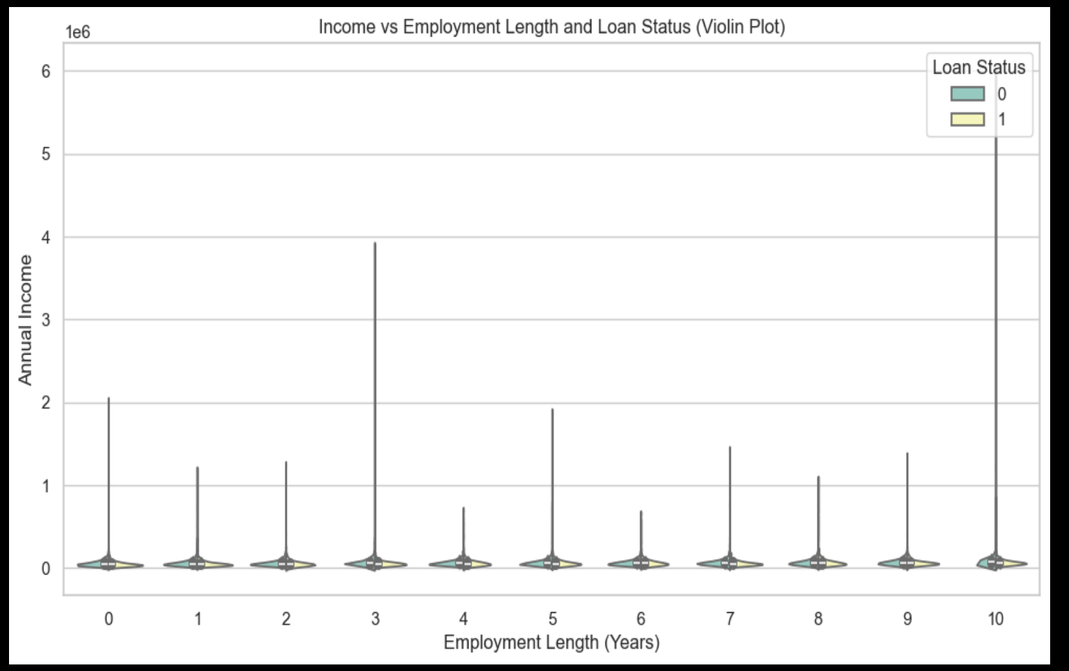

Loan Principal Recovery Rate by Loan Status
loan_principal_recovery_rate
0.0
0.2
0.4
0.6
0.8
1.0
Fully Paid
Charged Off
Current
loan_status

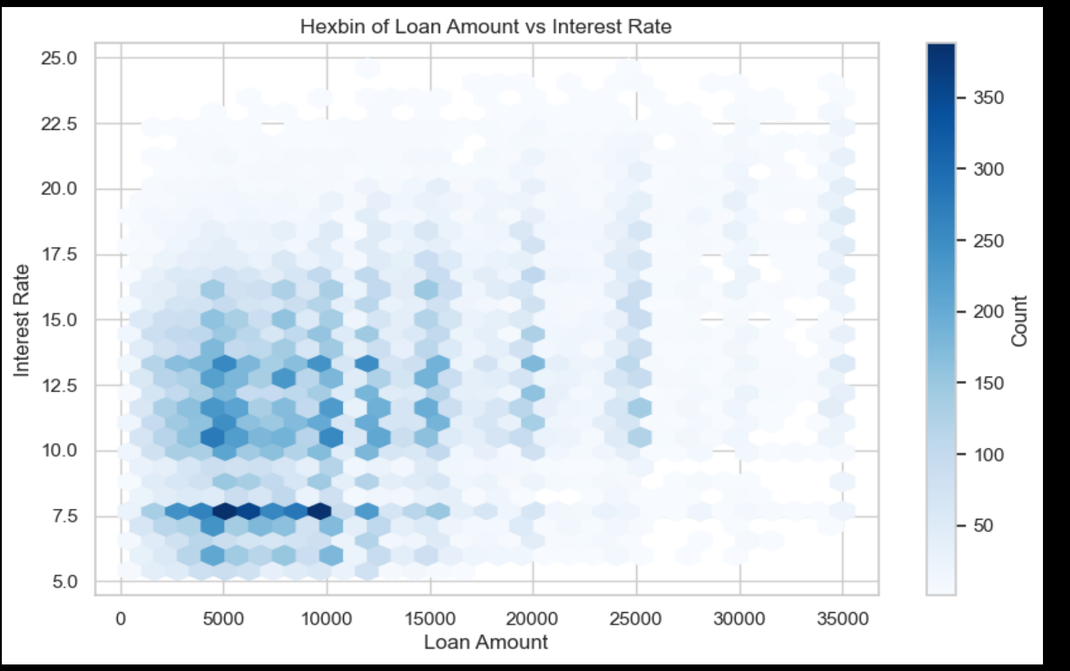
The multivariate analysis further revealed that loan amounts between $5,000 and $10,000, paired with an interest rate of 7.5%, were most common among borrowers. However, charged-off loans were more frequently associated with shorter employment lengths, typically around 1 year, and DTI ratios peaking at 35 for lower grades (F and B), indicating a strong likelihood of default. Fully paid loans, by contrast, had DTIs below 20, suggesting a more conservative financial profile. The Loan Payment to Principal Ratio showed that 75% of borrowers with charged-off loans had a ratio between 0.75 and 1, reinforcing that these borrowers had difficulty paying beyond the principal amount. The Delinquency Risk Index highlighted that charged-off loans had a higher risk score, helping identify borrowers with high default risk early on. Loan Charge-Off Rates for Grade F loans were the highest, underscoring the importance of credit grade as a risk factor.

Multivariate Analysis


Loan Term vs Loan Amount and Loan Status
loan_status_binary
0
1
Loan Amount
35000
30000
25000
20000
15000
10000
5000
0
36
60
Loan Term (Months)


DTI Distribution by Grade and Loan Status
Loan Status
0
1
Debt-to-Income Ratio (DTI)
35
30
25
20
15
10
5
0
−5
B
C
A
E
F
D
G
Loan Grade


Income vs Employment Length and Loan Status (Violin Plot)
1e6
Loan Status
0
1
Annual Income
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
10
Employment Length (Years)


Hexbin of Loan Amount vs Interest Rate
Interest Rate
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
0
5000
10000
15000
20000
25000
30000
35000
Loan Amount
Count
350
300
250
200
150
100
50

Driving Factors

Table 1: Driving Factors (1-5)

Driving Factor	Importance & Relation to Default
Home Ownership	A critical indicator of credit risk; borrowers who rent are statistically more likely to default compared to homeowners, as homeownership often signifies greater financial stability.
Income Level	Lower income thresholds (e.g., < $40,000) correlate with increased default probabilities, highlighting the significance of income stability in loan repayment capacity.
Employment Tenure	Short employment duration (less than 1 year) is associated with higher default rates, suggesting that longer employment histories are indicative of financial reliability and stability.
Loan Purpose	Loans for debt consolidation are frequently linked to borrowers experiencing financial hardship, thereby heightening the risk of default.
Interest Rates	Elevated interest rates (e.g., > 17.5%) are associated with increased default risk, necessitating careful assessment of borrowers' ability to manage higher repayment obligations.

Table 2: Driving Factors (6-10)

Driving Factor	Importance & Relation to Default
Loan Amount	Larger loan amounts (e.g., > $10,000) are correlated with heightened default risk, particularly among lower-income borrowers who may struggle to meet repayment obligations.
Debt-to-Income Ratio (DTI)	A high DTI ratio (e.g., > 35%) suggests that a substantial portion of a borrower's income is allocated to debt repayments, increasing the likelihood of default and indicating financial strain.
Geographical Factors	Certain regions, such as Utah, Alaska, and Wyoming, exhibit elevated default rates, implying that regional economic conditions and local market factors must be considered in credit risk assessments.
Credit Utilization	Higher credit utilization ratios (e.g., > 30%) may indicate financial distress, which is a precursor to increased default risk, necessitating close monitoring of borrowers' credit behavior.
Loan Term	Extended loan terms (e.g., 60 months) can result in lower monthly payment obligations but may lead to higher cumulative interest costs, adversely affecting a borrower's ability to repay over time.

Recommendations

Table 1: Recommendations 1-5

Recommendation	Policy	Reasoning
Enhanced Risk-Based Underwriting	Implement advanced machine learning algorithms to analyze historical loan data and borrower behavior.	Identifies non-obvious risk factors that traditional models may miss, enabling more accurate risk assessment and reducing defaults.
Stricter Lending Criteria for High-Risk Borrowers	Require a minimum down payment of 20% and collateral for borrowers with prior delinquencies.	Historical delinquencies are strongly correlated with increased risk; tightening standards minimizes exposure to high-risk borrowers.
Real-Time Credit Monitoring	Establish a system to track borrowers' credit utilization and overall financial health continuously.	Enables proactive risk management, allowing intervention when borrowers show signs of financial distress, reducing potential defaults.
Debt-to-Income (DTI) Ratio Threshold	Set a maximum DTI ratio of 40% for loan approvals.	High DTI indicates potential over-leverage and correlates significantly with repayment issues, making it critical for loan viability.
Loan Amount Capping Based on Income	Limit loan amounts to a maximum of 35% of the borrower's annual income.	Ensures loans remain manageable, particularly for lower-income individuals, reducing the risk of default due to excessive debt.

Table 2: Recommendations 6-10

Recommendation	Policy	Reasoning
Dynamic Interest Rate Adjustments	Offer interest rate reductions for borrowers with low credit utilization and high credit scores.	Rewards responsible credit behavior, encouraging low-risk behavior that aligns with favorable loan performance indicators.
Employment Stability Requirement	Give preference to borrowers with at least three years of stable employment history.	A stable employment history is positively correlated with loan repayment success, serving as a critical factor in risk assessment.
Risk Segmentation Framework	Develop a framework to classify applicants into low, medium, and high-risk categories.	Tailored loan terms based on risk categories allow for more strategic lending, minimizing potential losses.
Incentives for Early Repayment	Implement incentives such as interest rate reductions or fee waivers for early loan repayments.	Encourages early repayment, enhancing cash flow for the lender and mitigating default risk.
Use of Predictive Analytics for Loan Monitoring	Integrate predictive analytics tools to assess ongoing borrower risk profiles and identify potential defaults.	Utilizes data-driven insights to make informed decisions and take preventive measures, reducing credit loss.

Thank You