

Objective

To apply EDA techniques to a real-world problem, extract valuable insights, and present them in a business-centric manner through a clear and effective presentation.

Benefits of the case study:

- > Provides insight into how EDA is applied to real-world business challenges.
- > Enhances understanding of risk analytics in the banking and financial services industry.
- > Demonstrates how data can be leveraged to reduce financial losses in lending practices.
- > Strengthens knowledge of visualizations and the appropriate use of charts for real-world data analysis.

Domain Understanding

Consumer Finance Industry Overview

The consumer finance industry provides loans for personal, business, and medical purposes. Companies must assess applicants' risk profiles to make informed lending decisions.

Risk Analytics in Lending

Risk analytics helps predict the likelihood of loan defaults by analyzing historical data, minimizing financial losses for lenders.

Key Risks for Lenders

- -> Business Loss: Rejecting reliable borrowers results in lost business.
- -> Credit Loss: Approving risky borrowers leads to financial losses.

Loan Default Prediction

This analysis identifies patterns in borrower behavior that indicate default risk, allowing lenders to:

- -> Reject high-risk applicants.
- -> Adjust loan amounts or interest rates based on risk.

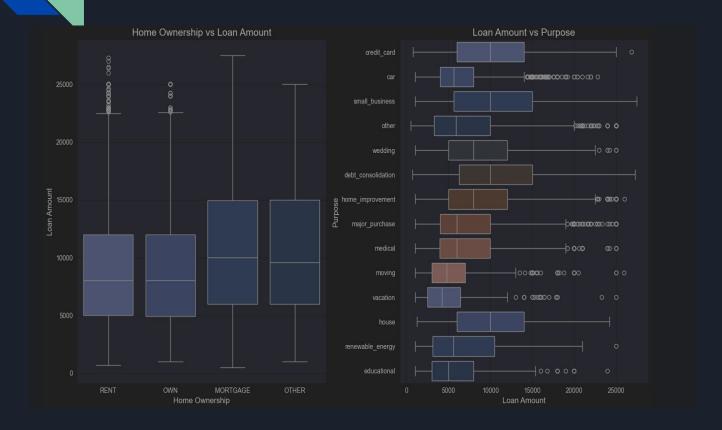
Focus on 'Charged-Off' Loans

'Charged-off' loans, where borrowers default, are the focus. Identifying factors behind these defaults helps improve risk assessment and reduce credit loss.

Key Steps in Exploratory Data Analysis (EDA)

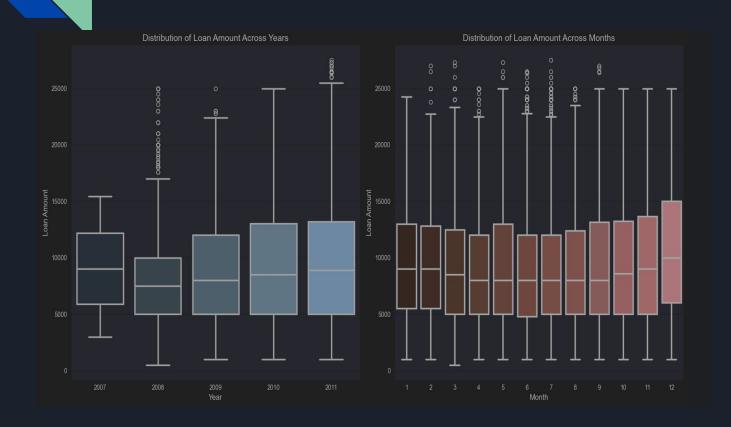
- Understanding the Dataset
- Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Outlier Detection
- Data Visualization
- Documenting Insights

Loan Amount Distribution by Home Ownership and Purpose



- 1. Home Ownership vs Loan Amount:
- Borrowers with mortgages apply for higher loan amounts (~15k USD), while renters and homeowners apply for smaller loans (~10k USD).
- Outliers exist in all categories for larger loans.
- 2. Loan Amount vs Purpose:
- Debt consolidation and credit card loans see the highest amounts (~15k USD).
- Categories like vacations and educational loans have much smaller amounts (~5k USD).
- Significant outliers appear for small business and major purchase loans.

Loan Amount Distribution by Months and Year



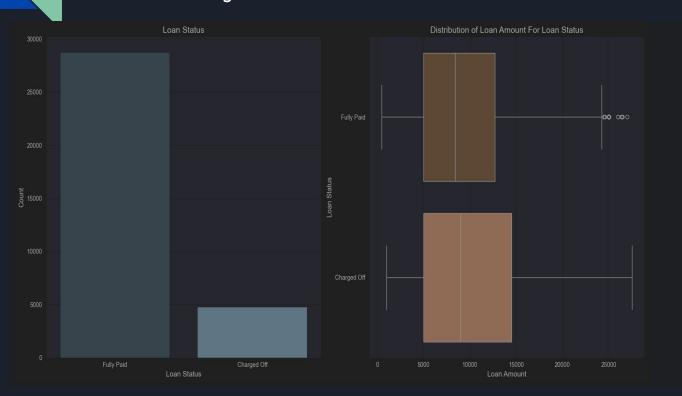
1. Distribution of Loan Amount Across Years:

- The median loan amount remained fairly stable from 2007 to 2011, with only slight fluctuations.
- In 2011, there are more high-value loan requests as indicated by outliers above \$25,000, suggesting a trend of borrowers seeking larger loans over time.

2. Distribution of Loan Amount Across Months:

- Loan amounts show consistent patterns across all months, with similar medians.
- Every month has some outliers with higher loan amounts, but there is no significant difference in median loan amounts throughout the year.

Loan Status Distribution and Loan Amount Variability



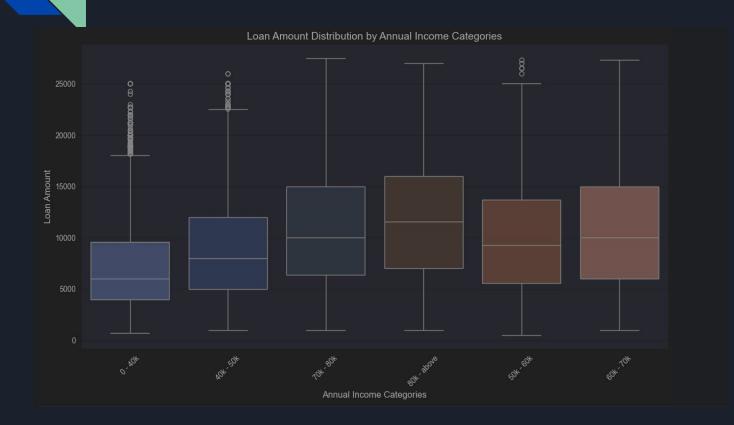
1. Loan Status:

- The majority of loans are categorized as Fully Paid, while a smaller portion of loans (around 14.5%) are Charged Off, indicating that a significant proportion of borrowers repay their loans.
- The stark difference in counts between Fully Paid and Charged Off suggests that defaulted loans represent a smaller but impactful subset of the total loans.

2. Loan Amount Distribution:

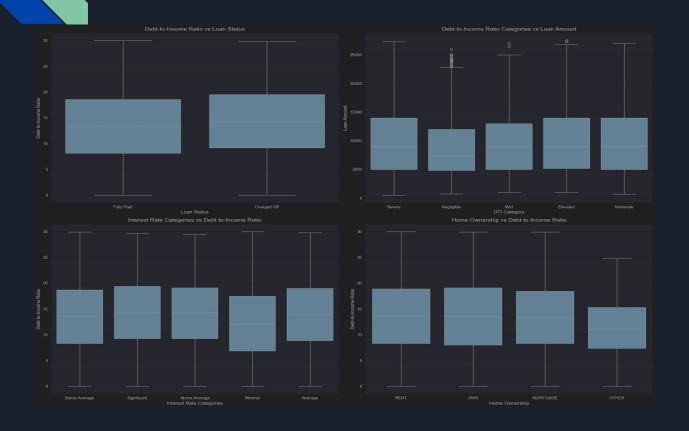
- Fully Paid loans tend to have a slightly higher median loan amount compared to Charged Off loans.
- While Charged Off loans exhibit a broader range in loan amounts, particularly on the upper end, they also show a higher number of outliers. This suggests that higher loan amounts may be associated with a greater risk of default.
- The box plot reveals that most loans, regardless of status, are clustered around similar loan amounts, though Charged Off loans have more variability.

Distribution of Loan amount by Annual Income Bins



- Loan amounts tend to increase slightly as annual income rises, but the variance in loan amounts remains high across all income categories.
- Applicants with an annual income between 40k and 80k generally apply for higher loan amounts, as seen by the medians being higher than in the lower income groups.
- The highest-income group (80k and above) has a wide range of loan amounts, suggesting that even high-income earners apply for varying loan amounts.
- There are outliers in almost every category, particularly in the lower income ranges (0-40k), where some applicants are requesting significantly higher loan amounts despite their income level.
- Overall, loan amounts are relatively stable across income groups, indicating that income alone is not the sole factor influencing loan amount requests.

Debt-to-Income Ratio Analysis: Impact of Loan Status, Loan Amount, Interest Rate, and Home Ownership



- Debt-to-Income Ratio vs Loan Status:

 Observation: Loans that are "Charged Off" have a higher median debt-to-income (DTI) ratio compared to "Fully Paid" loans.
- Risk Assessment: Higher DTI ratio is associated with a higher likelihood of loan default, making it a critical factor in risk evaluation. Borrowers with high DTI should be assessed more carefully.
- 2. Debt-to-Income Categories vs Loan Amount:
- Observation: The loan amount remains relatively consistent across different DTI categories, although those in the "Severe" category seem to request slightly higher loan amounts.
- Risk Assessment: Borrowers with severe DTI requesting larger loan amounts represent a higher risk for default. It's important to carefully evaluate loan approvals in this category.
- 3. Interest Rate Categories vs Debt-to-Income Ratio:
- Observation: There isn't a significant variation in DTI across interest rate categories, though loans with higher interest rates slightly correspond to higher DTI.
- Risk Assessment: Borrowers with higher interest rates and DTIs should be flagged for additional scrutiny since they may present a higher risk of default.
- Home Ownership vs Debt-to-Income Ratio:
 Observation: Renters tend to have higher
 DTI ratios compared to those who own homes or have mortgages.
- Risk Assessment: Renters represent a higher risk, and loans to this group should be monitored closely. Implementing stricter evaluation criteria for renters could mitigate potential defaults.

Loan Status Distribution by Key Factors: Home Ownership, Purpose, Term, and Grade



1. Home Ownership:

- Renters have a relatively higher rate of charged-off loans at 15.1%.
- Those with a mortgage have the lowest charge-off rate at 13.2%.
- The "Other" category has a higher charge-off rate at 18.0%.

2. Loan Purpose:

- Small business loans have the highest charge-off rate at 26.8%, making it a riskier loan purpose.
- Loans for vacation and weddings have the lowest charge-off rates, below 10%, indicating a lower risk.
- Debt consolidation, which has a large number of applicants, sees a charge-off rate of 14.9%, an average risk.

3. Loan Term:

- 60-month term loans show a higher charge-off percentage at 25.4%, suggesting that longer-term loans may carry more risk.
- 36-month loans have a lower charge-off rate of 11.0%, indicating lower risk for shorter terms.

4 Loan Grade:

- Loans in Grade A have the lowest charge-off rate at 6.2%, implying that higher-grade loans are safer.
- Grades E, F, and G have significantly higher charge-off rates, with Grade G being the riskiest at 39.4%.

5. Verification Status:

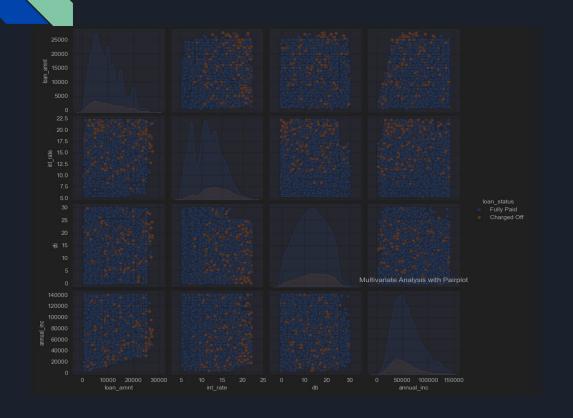
- Not verified loans have a higher charge-off rate at 16.6%.
- Verified loans have a slightly lower charge-off rate at 12.7%, which implies that verification can reduce the risk but does not eliminate it.

Multivariate Correlation Heatmap



- 1. Loan Amount and Interest Rate Interaction (0.88):
- Strong correlation: larger loans with higher interest rates increase total loan costs.
- 2. Loan Amount and Annual Income (0.33):
- Moderate correlation: higher income borrowers tend to take larger loans.
- 3. Loan Amount and Interest Rate (0.20):
- Weak correlation: bigger loans usually have slightly higher interest rates.
- 4. Interest Rate and Loan Status (0.21):
- Weak correlation: loans with higher interest rates are slightly more likely to default.
- 5. Debt-to-Income (DTI) and Loan Status (0.04):
- Very weak correlation: DTI alone is not a strong predictor of loan defaults.
- 6. Overall:
- No single numeric variable strongly predicts loan default, suggesting a mix of factors, both numeric and categorical, influence defaults.

Multivariate Analysis with Pairplot



- 1. Loan Amount: Larger loans tend to have more charged-off loans compared to smaller loans.
- 2. Interest Rate: Higher interest rates are associated with more charged-off loans.
- 3. Debt-to-Income Ratio (DTI): No clear pattern between DTI and loan status, indicating it may not strongly predict defaults.
- 4. Annual Income: Lower income borrowers have more charged-off loans, while higher income borrowers tend to fully repay loans.

Final Insights

- 1. Loan Amount, Interest Rates, and Income
- Loan Amount vs Income: Borrowers with higher annual income tend to take larger loans. This suggests that income is a significant factor when deciding loan amounts, though it doesn't directly indicate risk.
- Loan Amount vs Interest Rates: Larger loans generally have slightly higher interest rates, but the correlation is weak. However, borrowers who take out larger loans may face more difficulty in repayment, especially when paired with higher interest rates.
- Interest Rates and Default Risk: Loans with higher interest rates show a slight increase in default risk. Lenders may want to focus on reducing defaults by better assessing higher-interest loans, as these tend to be more vulnerable to defaults.

2. Debt-to-Income Ratio (DTI)

- Despite common assumptions, the DTI ratio alone doesn't strongly correlate with loan defaults, indicating that other factors, such as employment history, loan grade, and interest rates, might play a more significant role in assessing risk.
- Borrowers with severe DTI still request larger loan amounts, increasing their risk of default. While DTI is not a standalone indicator, combining it with other metrics could enhance risk prediction.

3. Loan Status and Grade

- Grade G has the highest charge-off percentage (39.4%), while Grade A has the lowest (6.2%). As loan grade worsens, the likelihood of default increases, making loan grade one of the strongest indicators of risk.
- Longer loan terms (60 months) show higher charge-off percentages compared to shorter terms (36 months), with 25.4% default risk for longer loans, indicating higher risk with extended repayment periods.

4. Verification Status

- Not Verified loans have a higher default rate (16.6%) compared to Verified loans (12.7%). Verification plays a key role in mitigating risk, and focusing on thorough verification processes could reduce default rates.

5. Purpose of Loan

- Small business loans show the highest charge-off rate (26.8%), suggesting that loans given for small business purposes carry a higher risk. In contrast, personal loans for debt consolidation or vacations show lower default risk, implying a safer loan portfolio.

Additional Final Insights:

- Combining Factors: The heatmap and pairplot analyses reveal that loan defaults aren't driven by any single variable. Factors like loan amount, interest rates, loan purpose, and verification status interact in complex ways to affect the likelihood of default.
- Income and Loan Size: While borrowers with higher incomes tend to take out larger loans, these loans don't necessarily carry a higher risk. However, combining income with DTI and loan grade provides a more nuanced picture of risk.
- Interest Rates and Risk: Higher interest rates correlate with a higher likelihood of default, suggesting that borrowers with riskier profiles (indicated by higher rates) should be carefully monitored.
- Categorical Variables are Key: Variables like loan grade, verification status, and loan purpose are stronger indicators of loan defaults than simple numeric measures like DTI or annual income. A focus on these categorical variables in risk models could yield better risk predictions.

Conclusion:

Lending Club can use these insights to refine its risk assessment model, focusing on a combination of loan grade, interest rates, loan term, and verification status. This approach will help mitigate the risk of defaults and improve overall loan portfolio performance.