Lending Club Case Study

Vivek Pemawat
Virgillus Zita J

Problem Objective

A financial company aims to identify the key factors (or driver variables) that are strong indicators of loan default. Understanding these variables will help the company enhance its portfolio and risk assessment strategies, enabling better decision-making and risk management.

Solution

Perform a thorough Exploratory Data Analysis to understand the Loan Data set and prescribe insights from the understanding of the data.

Business Objective

The goal is to make informed decisions on incoming loan applications, determining whether to **approve** or **reject** them based on historical data and certain variables. The dataset at hand includes information on past loan applicants and their repayment behavior, indicating whether the loan was **defaulted** or not.

Dataset Details:

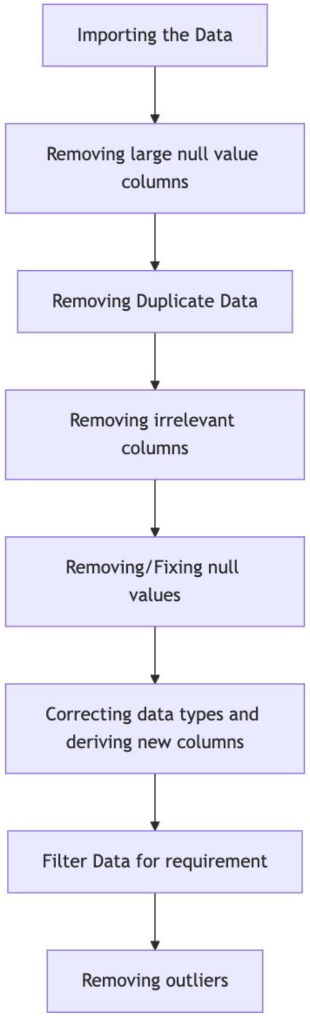
The dataset provides information exclusively on **approved loans**, with no data available for rejected applications. The loans are categorized into three main statuses:

1. **Fully Paid**: Loans that have been completely repaid by the borrower.
2. **Current**: Loans that are still active and are currently being repaid.
3. **Charged-Off**: Loans where the borrower has defaulted, and the lender has ceased collection efforts.

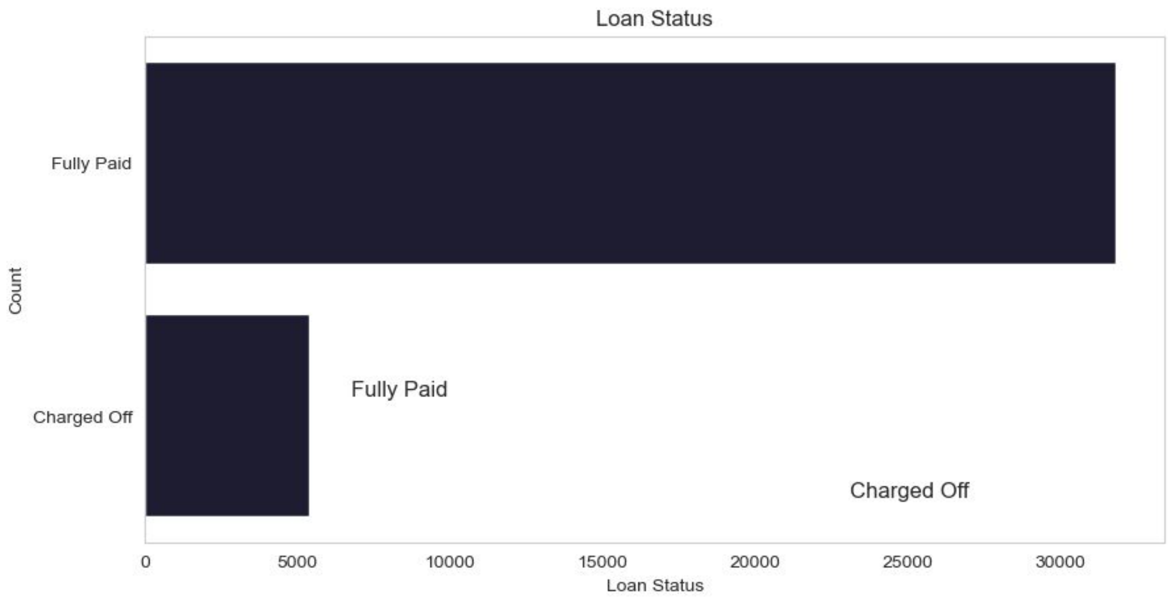
Benefits of Case study

- **Real-world application of EDA**: This case study demonstrates how EDA can be effectively used to address business problems, particularly in finance.
- **Introduction to risk analytics**: Gain a foundational understanding of how risk analytics is applied in the banking and financial sectors to manage and mitigate risks.
- **Minimizing financial loss**: Learn how data analysis can be applied to reduce losses in lending by identifying potential defaulters and adjusting lending practices accordingly.
- **Enhanced data visualization skills**: Improve your ability to visualize data and select the appropriate charts and graphs to draw meaningful conclusions from real-life financial data.

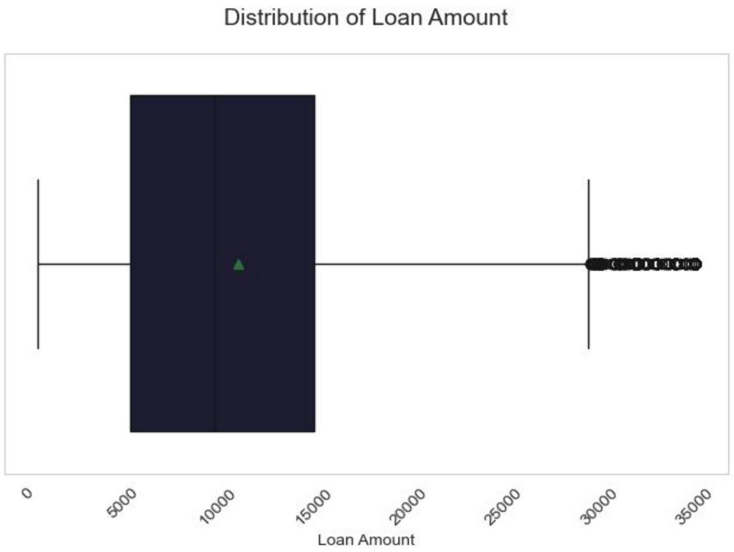
Data Clean-up and preparation process:



Loan Status and Amount

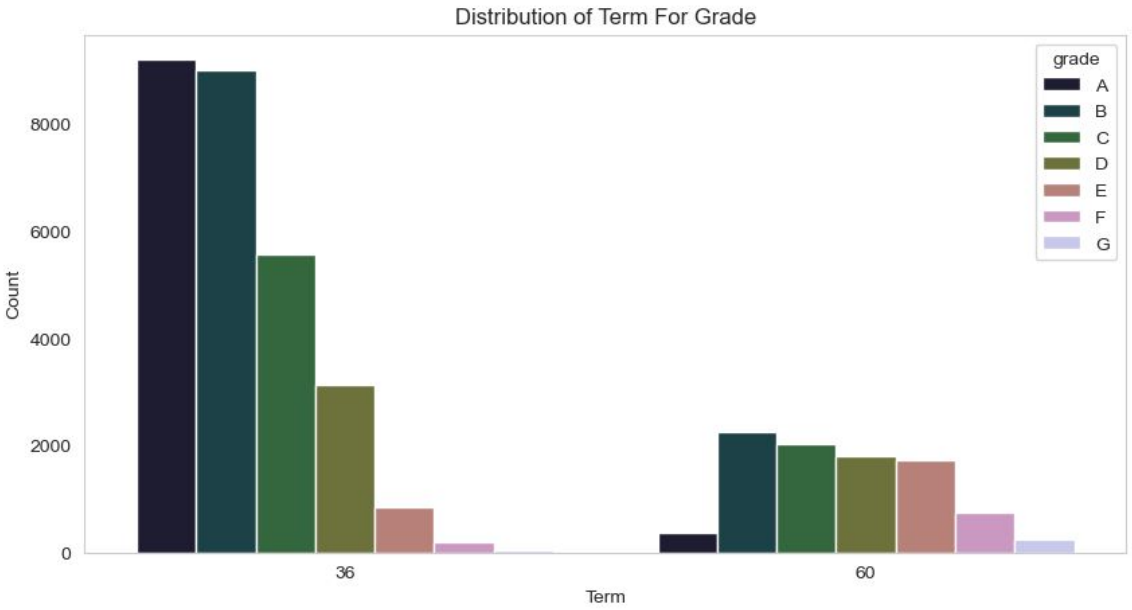


Loan Status Inference - Only 14.5% of loans (5130 out of 37173) resulted in charge-offs.

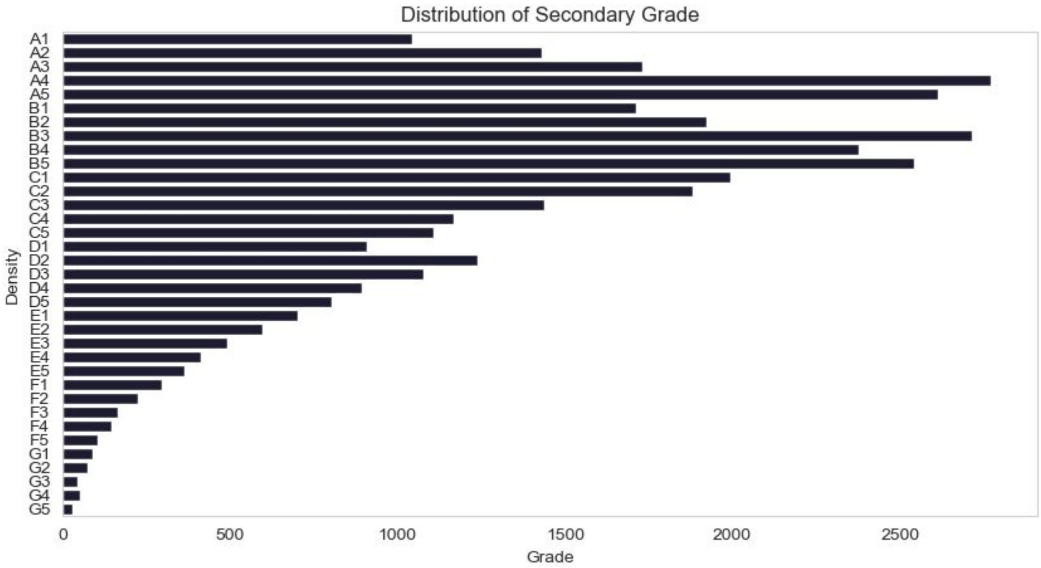


Loan Amount Inference: Loan amounts range from 500 to 35,000 with a median of 10,000. Most loans are relatively small, and the risk of default increases with larger loan amounts

Grade and Sub-Grade

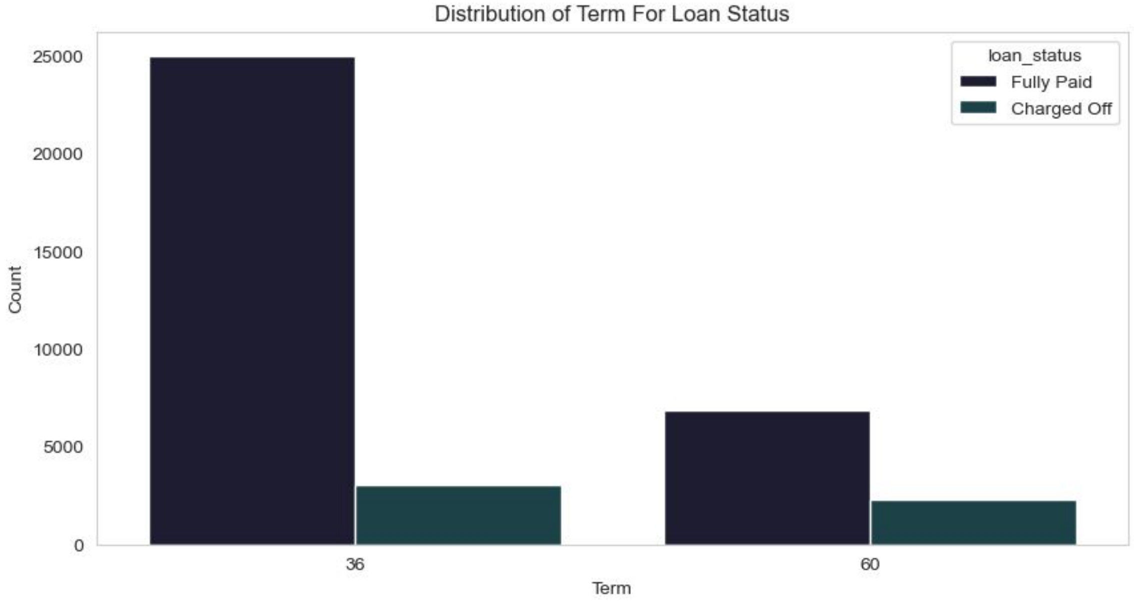


Grade Inference: The loan approved are majorly of higher grade as they are of low risk thus low chance of defaulting. 60 month term loans have larger number of lower grade loans with high risk.



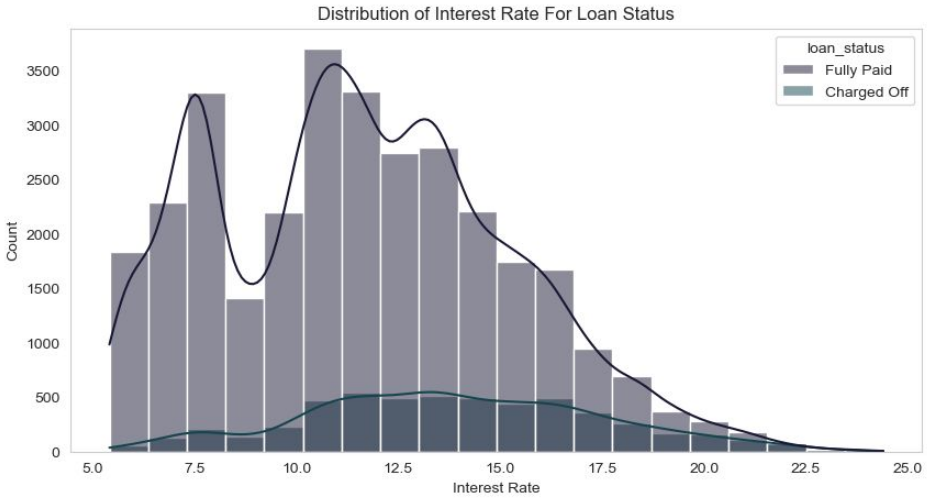
Sub Grade Inference: This provides more insight that the loans within grade are more skewed towards lowered sub grades.

Term and Interest Rate

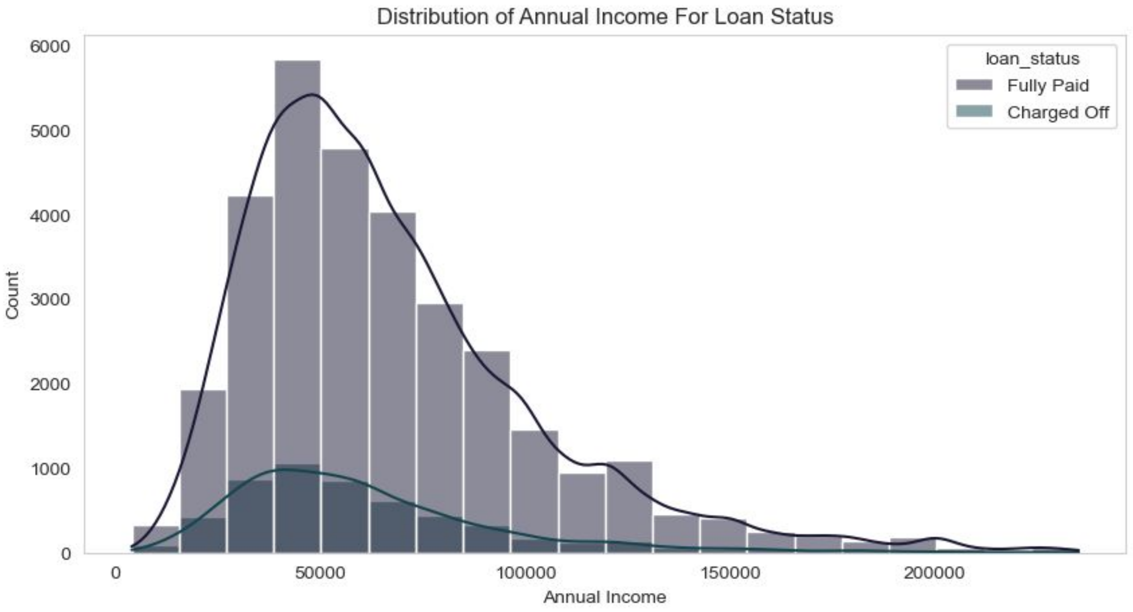


Loan Term Inference: Loans with a 36-month term are more common and have a higher success rate compared to 60-month loans, which are more likely to default.

Interest Rate Inference: There's a strong correlation between interest rates and loan demand. Loans are most sought-after at lower interest rates (5-15%), and the risk of default grows with higher rates.

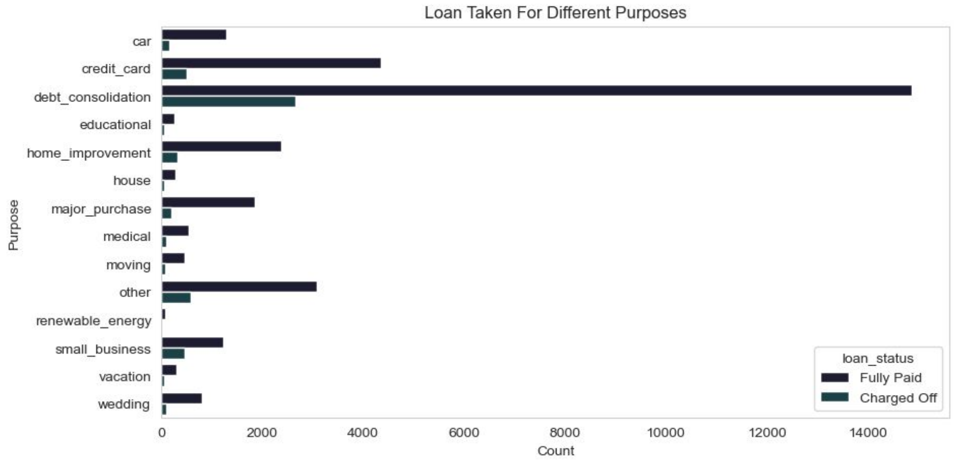


Annual Income & Purpose

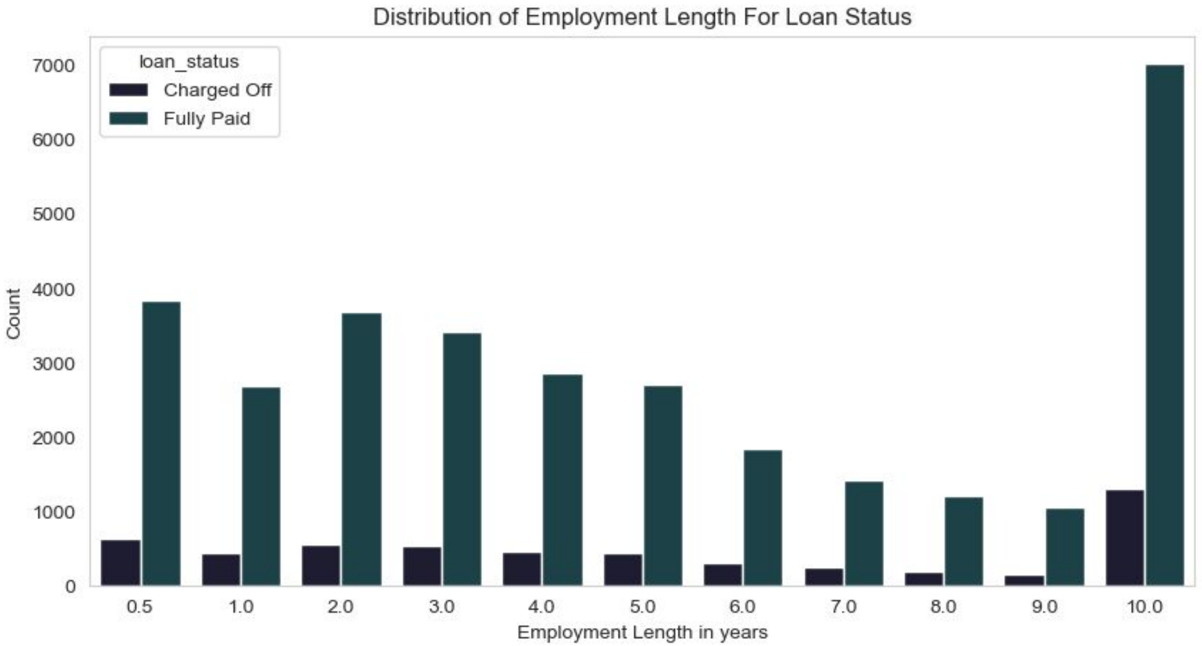


- **Annual Income Inference:** The Majority of clients have low annual income compared to rest and income lower than 50k has higher chance of defaulting.

- **Purpose Inference:** Loans are taken mostly for debt consolidation followed by credit card payment. Whereas the debt consolidation has highest fully paid loan but also has highest defaulted loans as well.

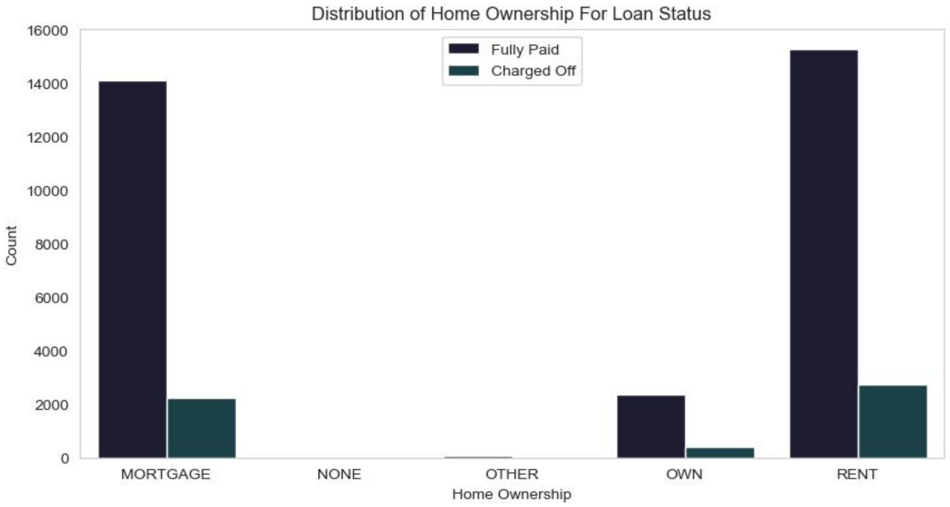


Employment Length & Homeownership

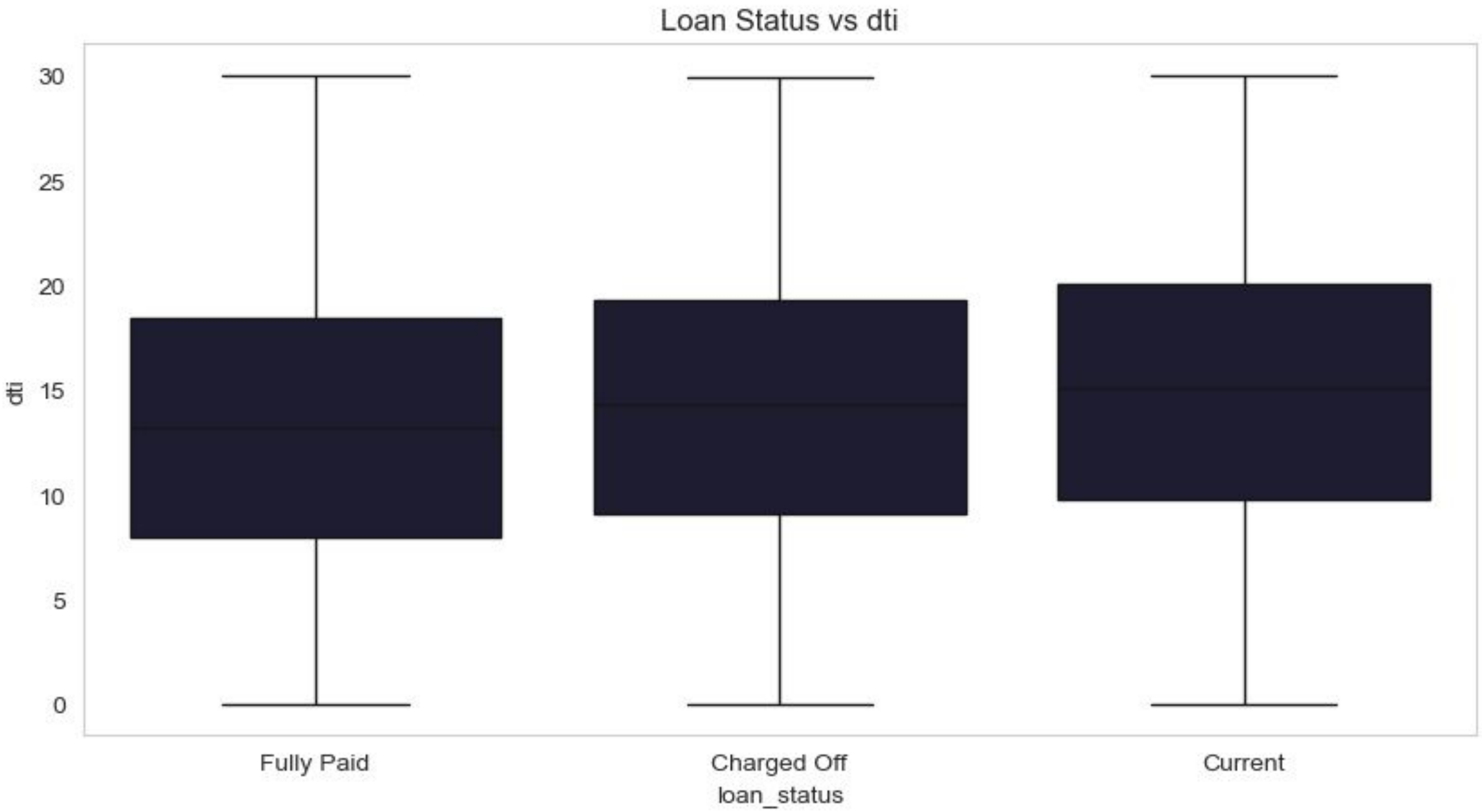


Employment Length Inference: Majority of clients have 10+ years of experience and has highest number of defaulted loan.

- **Home Ownership Inference:** Majority of clients are lacking ownership of any property and are on rent or mortgage and have a higher chance of defaulting.

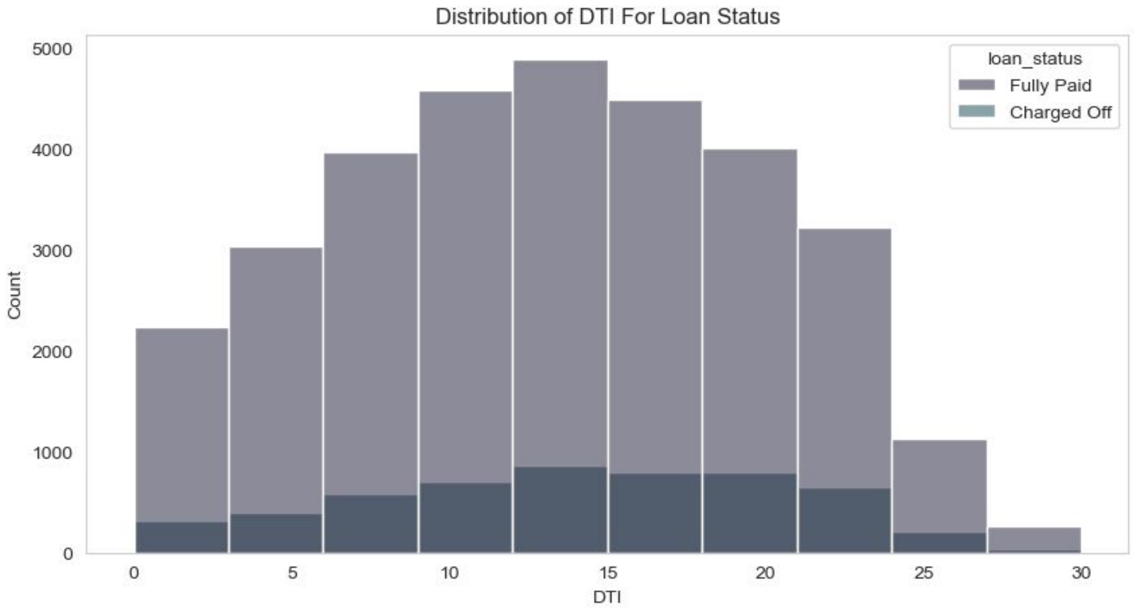


DTI vs Loan Status



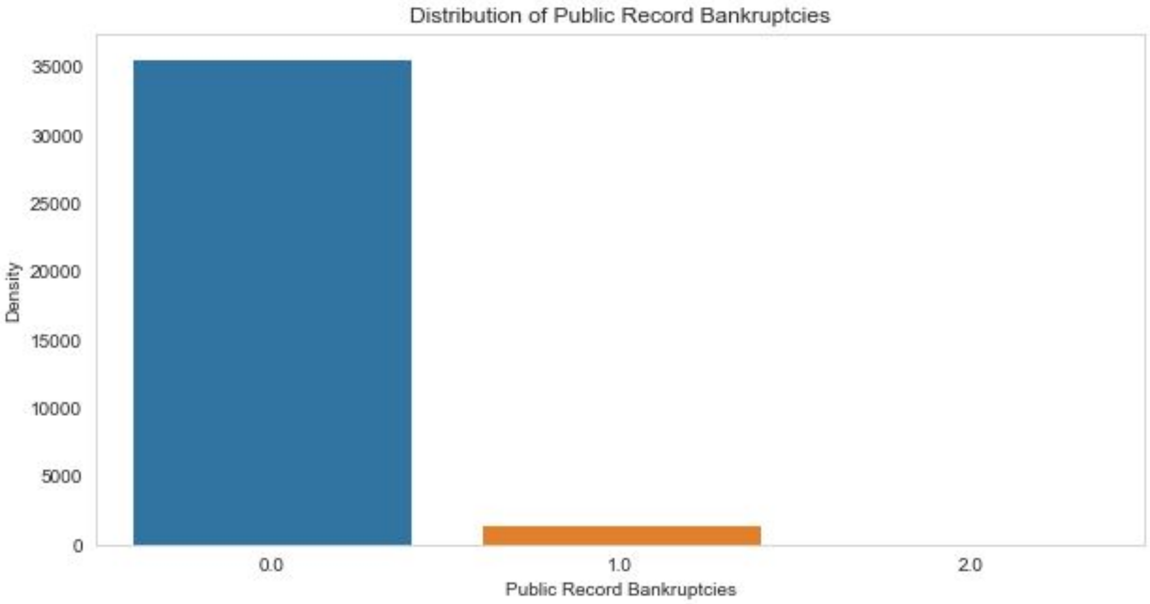
Dti ranges from 7 to 20 for all 3 categories of Loan Status

DTI ratio & Bankruptcy

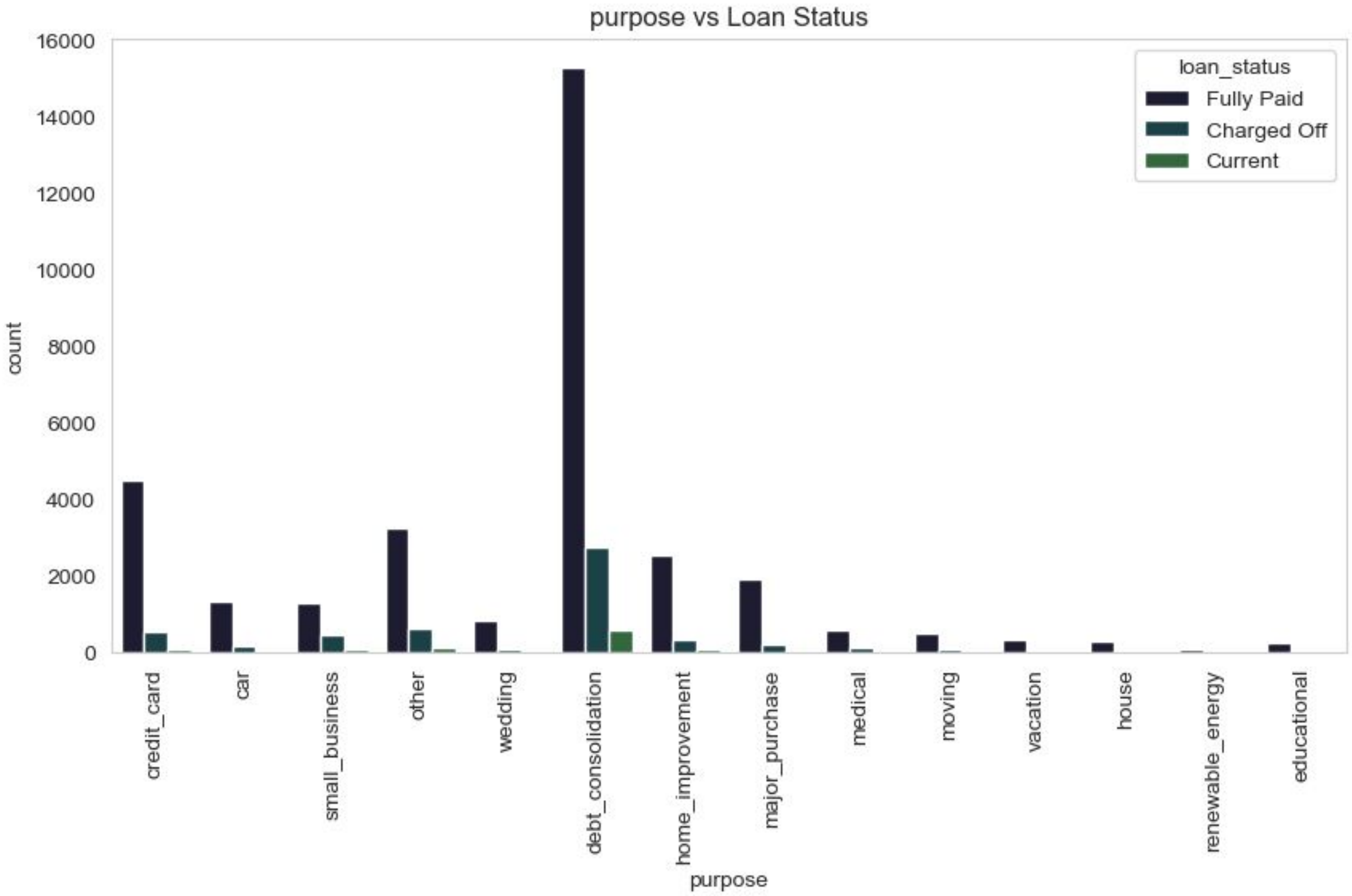


- **DTI Inference:** A large number of clients face financial challenges, as evidenced by their high debt-to-income ratios.

- **Public Recorded Bankruptcy Inference:** The majority of clients have not filed for bankruptcy.

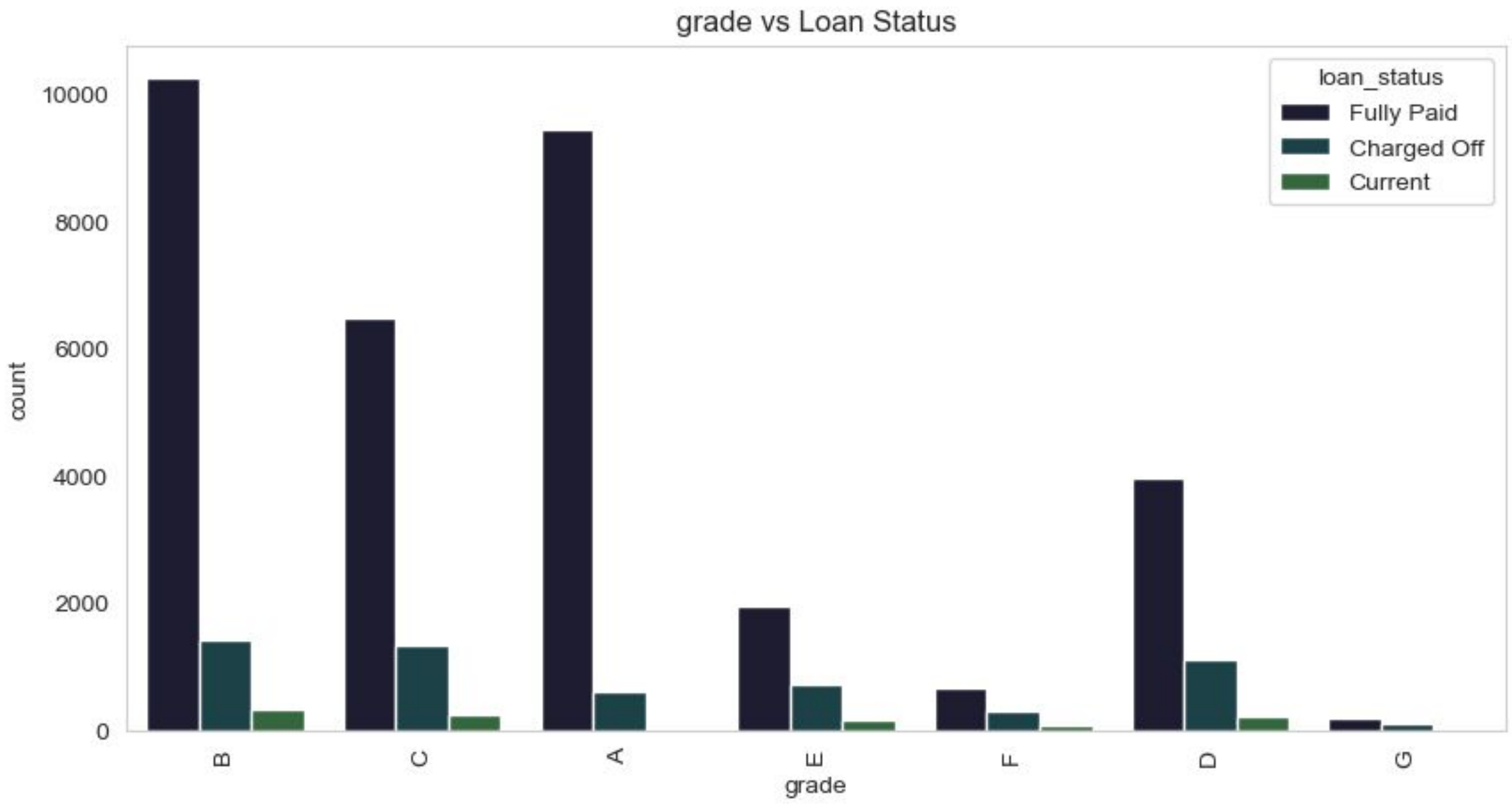


Loan Status vs Purpose



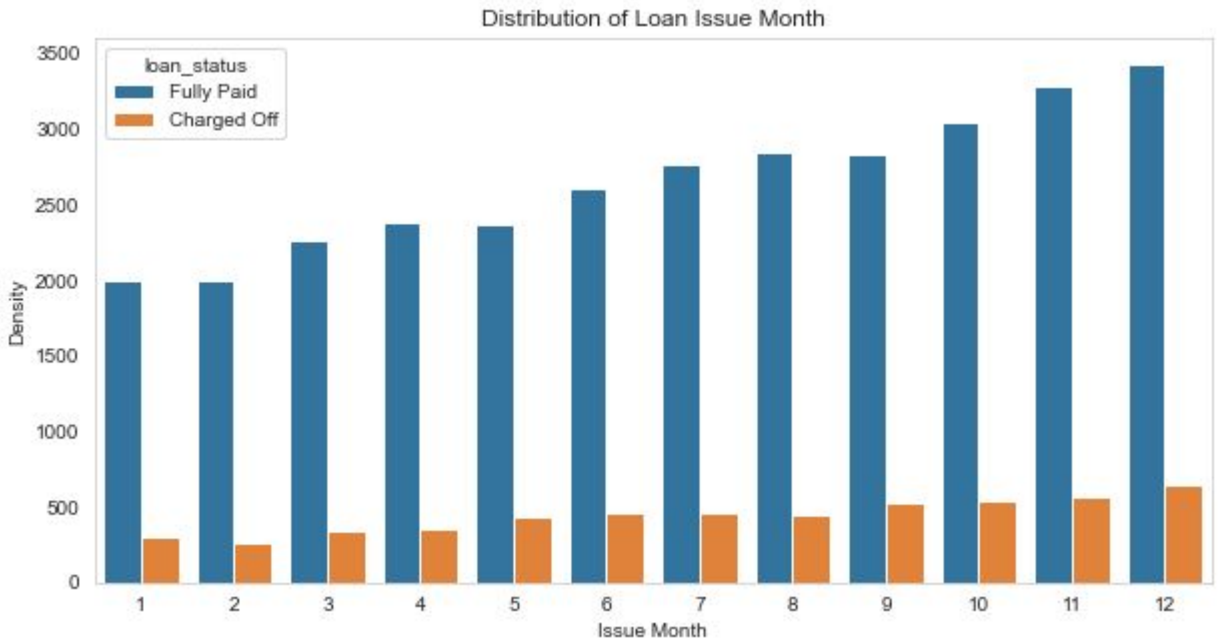
Debt_consolidation seems to be the most likeliest category for repaying the loans followed by credit_card

Grade vs Loan Status



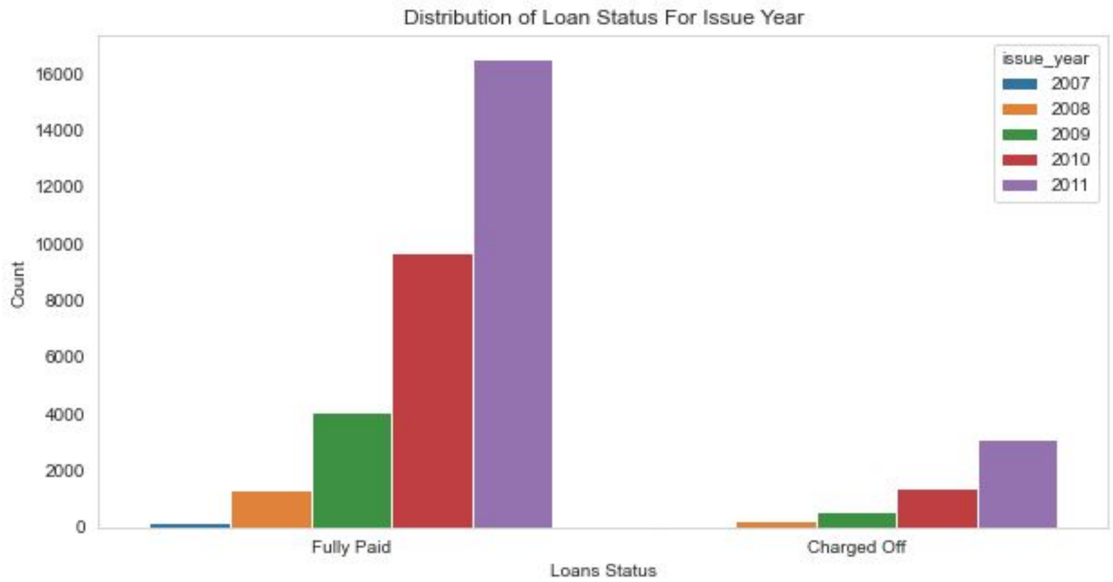
Lower grades are more likely to pay the loans fully than Higher grades

Loan Trend over years



We see a gradual increase in loan taken through the year, with lesser defaulting rate in April ,August, December quarter wise and better more late in year.

With each passing year loan taken are increasing exponentially which indicate we are seeing large increase in DTI ratio and decrease in defaulting rate.



Observations & Recommendations

Key factors that can help predict the likelihood of default and reduce credit losses:

Debt-to-Income (DTI) ratio.
Loan grades.
Verification status.
Annual income.
Public record of bankruptcy.
Other considerations for identifying high-risk 'defaults':

Borrowers from areas other than large urban cities (e.g., California, New York, Texas, Florida).
Borrowers with annual incomes between
100,000.
Borrowers with a history of public-recorded bankruptcy.
Borrowers with lower loan grades (e.g., grades E, F, G), which represent higher risk.
Borrowers with very high debt-to-income (DTI) ratios.
Borrowers with 10+ years of work experience.