

---

# Project 1: Lending Club

Tecnológico de Monterrey Data Analytics  
Boot Camp

Francisco H. Muñoz Ceballos

Karen Reyes Villarreal

Daniel Bardán Lamadrid

David Tecuatl

---



# SCOPE

As the dedicated team supporting the Chief Operating Officer (COO), our main goal is to explore the nuances of loan distribution across different segments within our enterprise, the risk management opportunities, scalability and operations efficiency.

From understanding borrower profiles to evaluating loan performance in search of valuable insights essential for strategic decision-making.

---



- ¿What is Lending Club?

LendingClub is a peer-to-peer lending company, headquartered in San Francisco, California. It is the leading digital marketplace bank in the U.S., connecting borrowers with investors since 2007.

Lending Club enables borrowers to create unsecured personal loans between **\$1,000** and **\$40,000**. The standard loan period is three years. Investors can search and browse the loan listings and select loans that they want to invest based on the information supplied about the borrower, **amount of loan, loan grade, and loan purpose**. Investors make money from interest. Lending Club makes money by charging borrowers an origination fee and investors a service fee.

---

---

# The dataset

- **What can we tell from our dataset?**

With over 750,000 loans registered in this data set, we had access to complete loan data from 2016 to 2017, including the current loan status (Current, Late, Fully Paid, etc.), latest payment information, delinquency information, annual income, employment length, interest rates, etc.

Size of dataset = 759,338

<https://www.kaggle.com/datasets/husainsb/lendingclub-issued-loans>

---

---

# Insights from Lending Club Data

- Is there any trend of opportunity based on demographic or socioeconomic data?
  - How many users are there in each state? Popularity of each loan purpose, mean annual income in each state, loan amount by homeownership?
  - Are there factors that help to have a revenue growth model with lower risk of increasing revenue?
  - What's an approach to risk assessment and product categorization based on loan grades and sub-grades.
-

---

# DEBT VS debt

---



Moving and relocation

Vacation

Medical expenses

Business

Car financing

Other

Debt consolidation

Home improvement

Major purchase

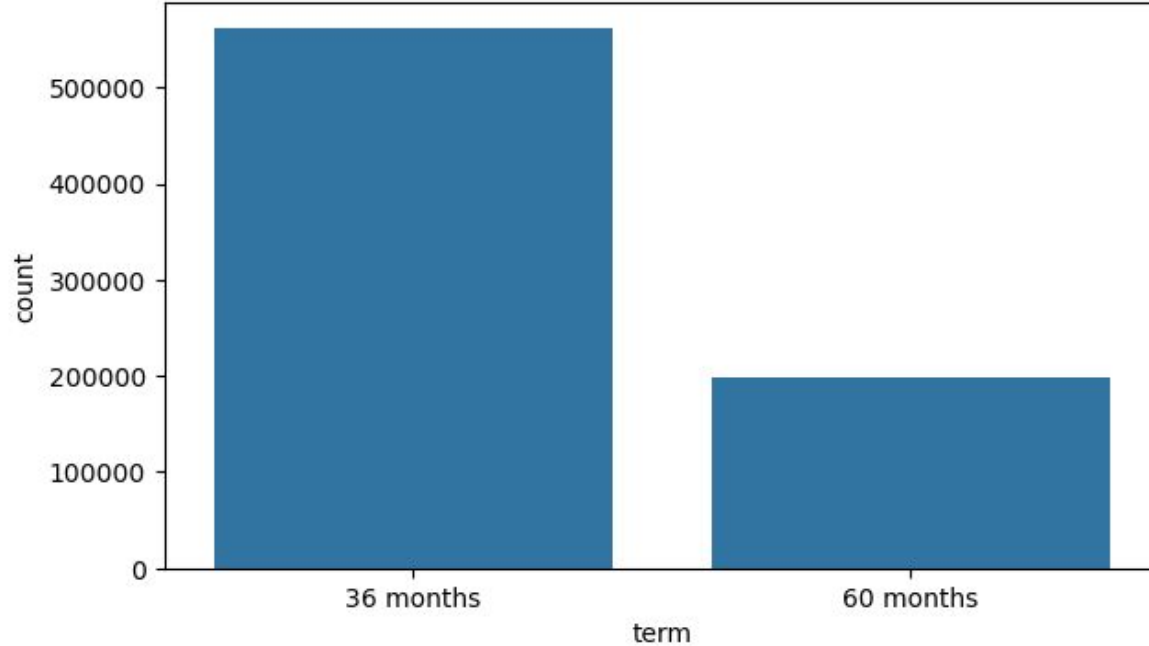
Home buying

Green loan

Credit card refinancing

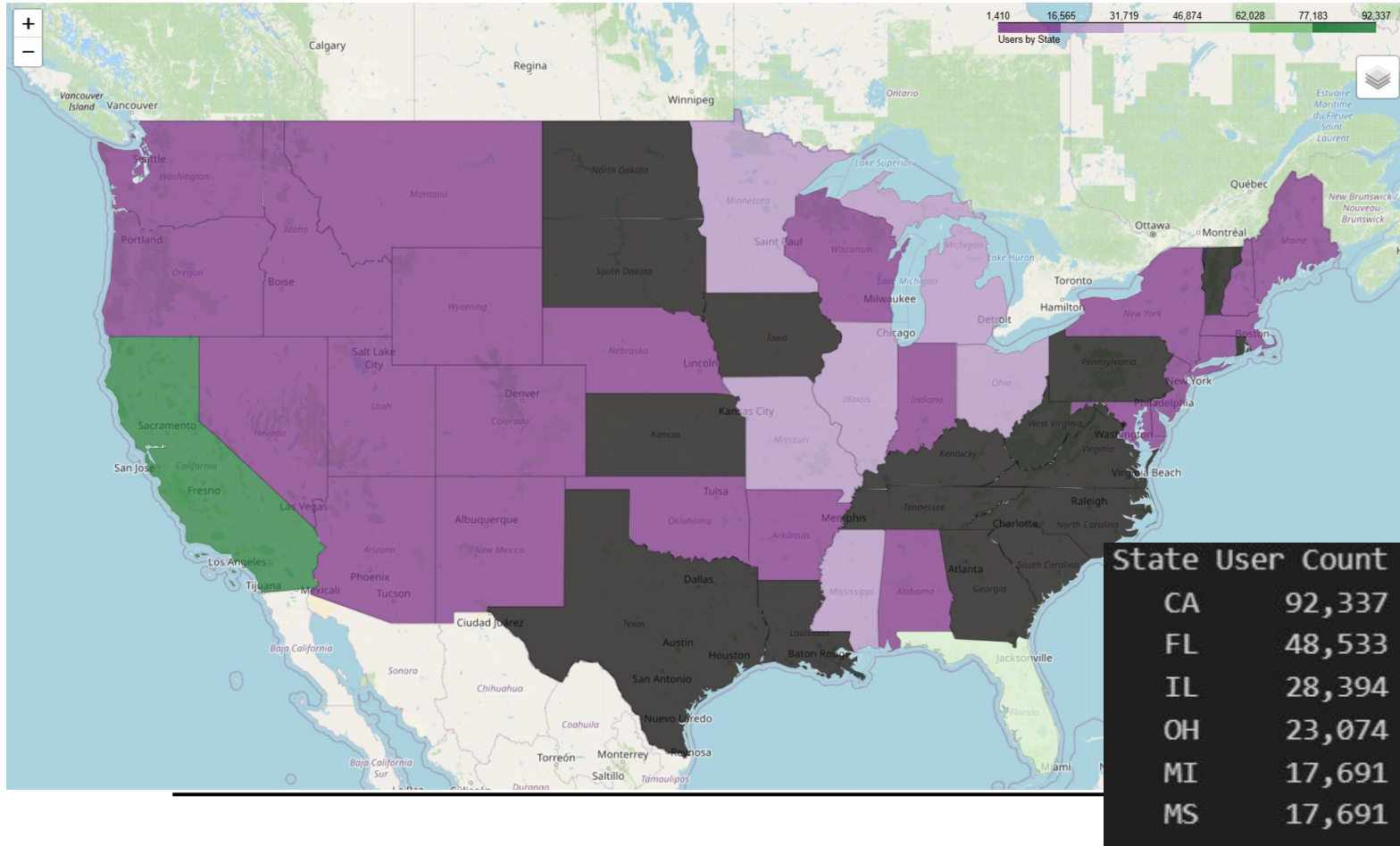
---

# Standard Loan Period



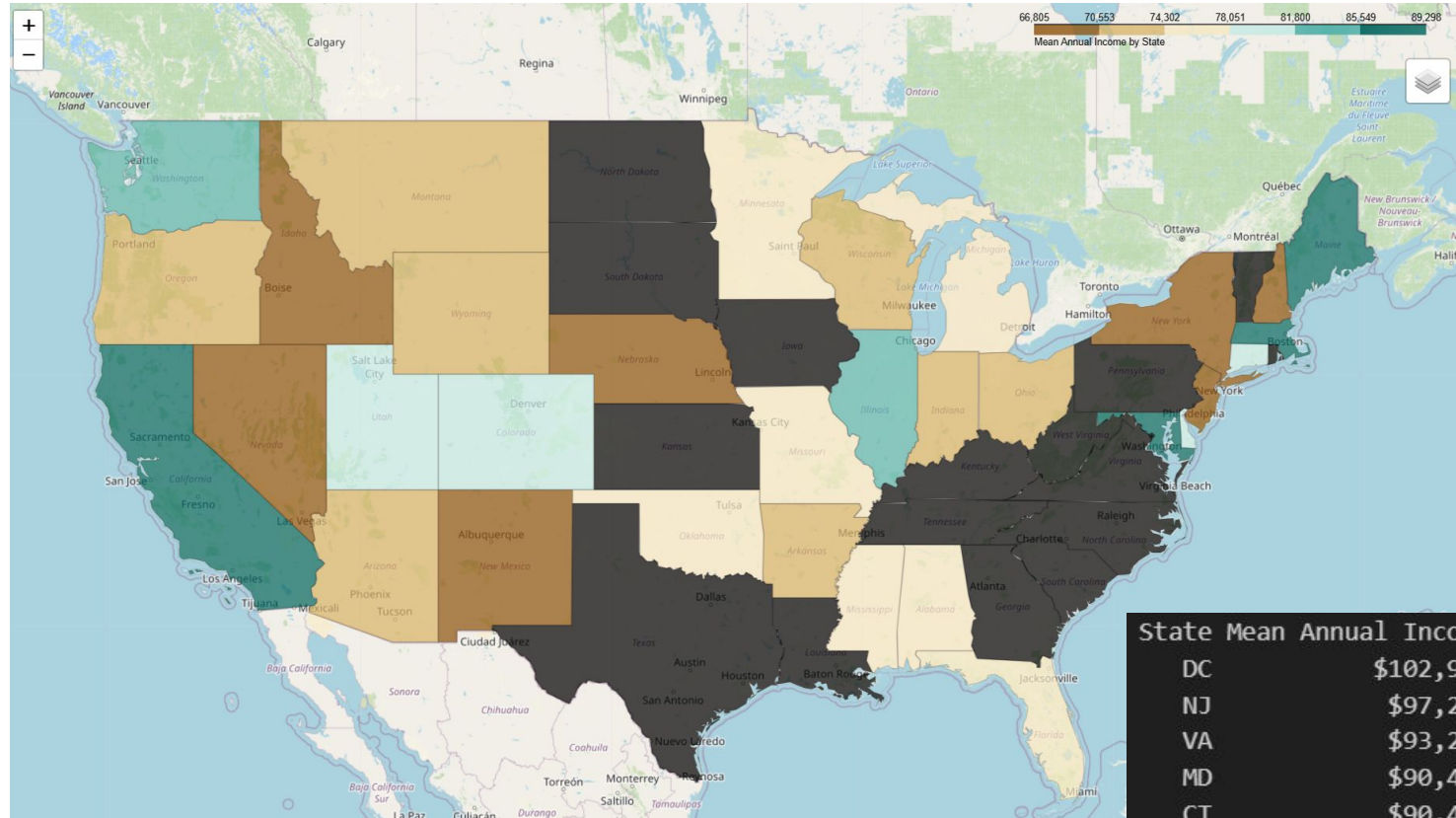
As mentioned LendingClub has most of its debt invested in a 36 month period. And only 26% in a 60 month period.

# Hispanics for the loans





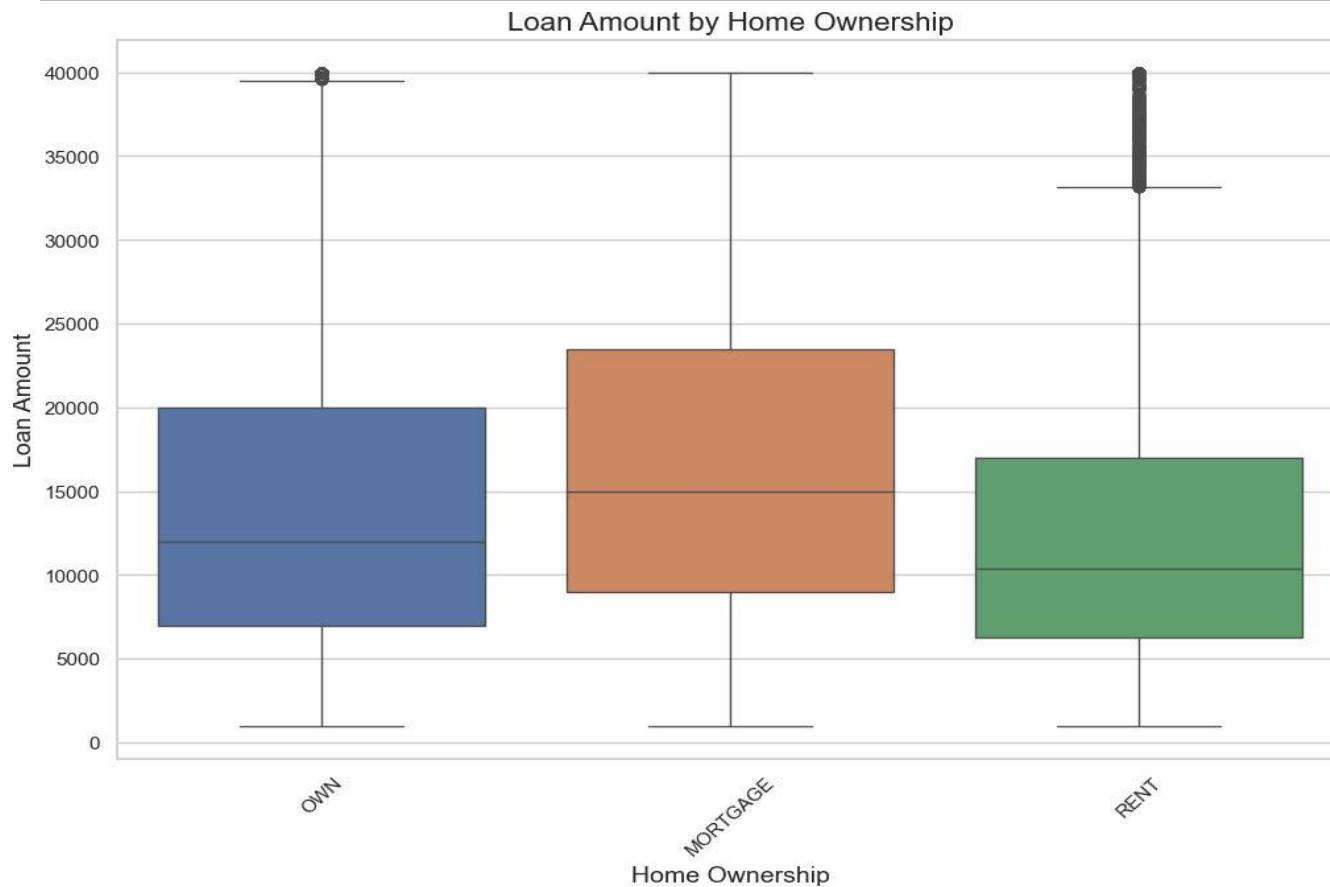
East  
coast  
for  
the  
win



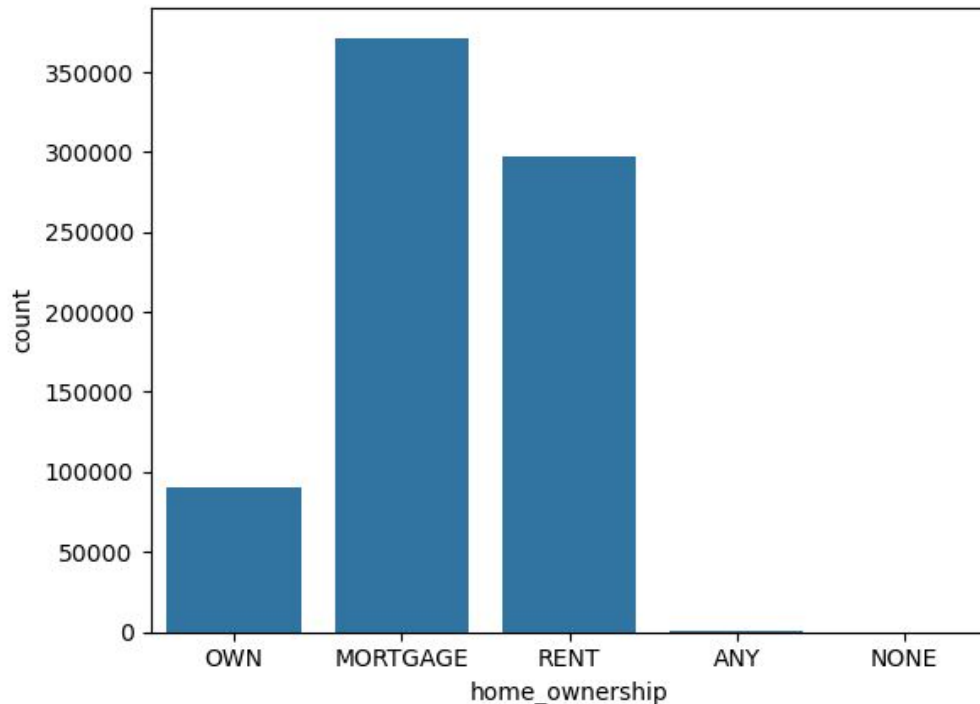
### State Mean Annual Income

DC	\$102,965
NJ	\$97,221
VA	\$93,211
MD	\$90,437
CT	\$90,401
CA	\$89,298

More  
money,  
  
more  
rent

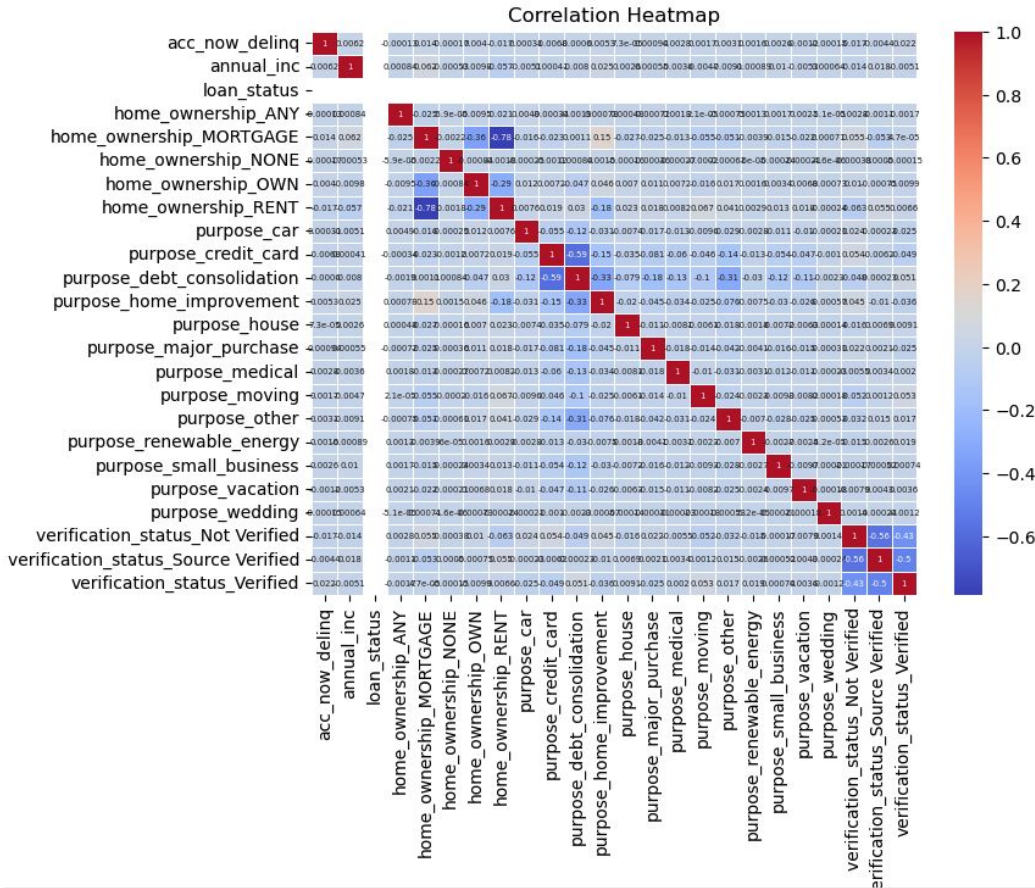


# Home ownership Count



When trying to better understand our client base, we noticed an important factor of them and it is that a big part of them (39.5%) don't own a house and also don't have a mortgage in place. Leaving room for the commercial and risk department to assess the opportunities of those clients to access a mortgage.

# When analyzing if there was any correlation between Delinquency and other variables we did not found a clear one



We analyzed variables such as:

1. Delinquency vs:
  - a. Annual income.
  - b. If they were home owners.
  - c. If they had a Mortgage.
  - d. If they rented.
  - e. Credit purpose.
  - f. If their credit was verified or not.
  - g. Their debt consolidation.

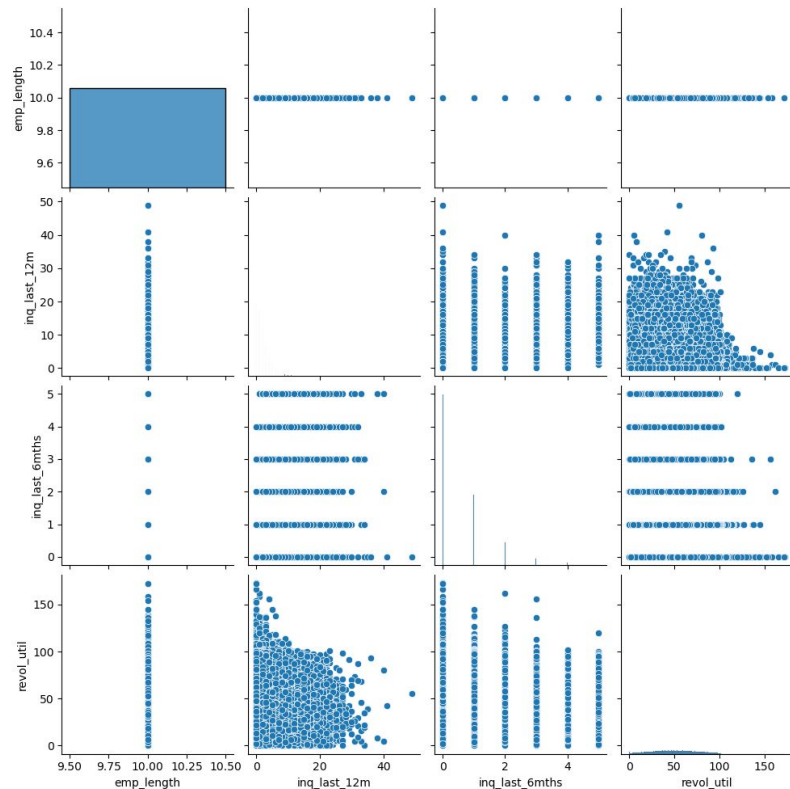
When we made an analysis by state to see if there were any states above average in delinquency we found 4

Acc Now Delinq Distribution by State



We found states like California, Texas, NY & Florida to be above, the question is whether that's because they have large populations or their ratio is similar to the other states.

# Most of the population who have asked for a loan in the last 12 months have a uniform use of their credit lines (**revol\_util**)



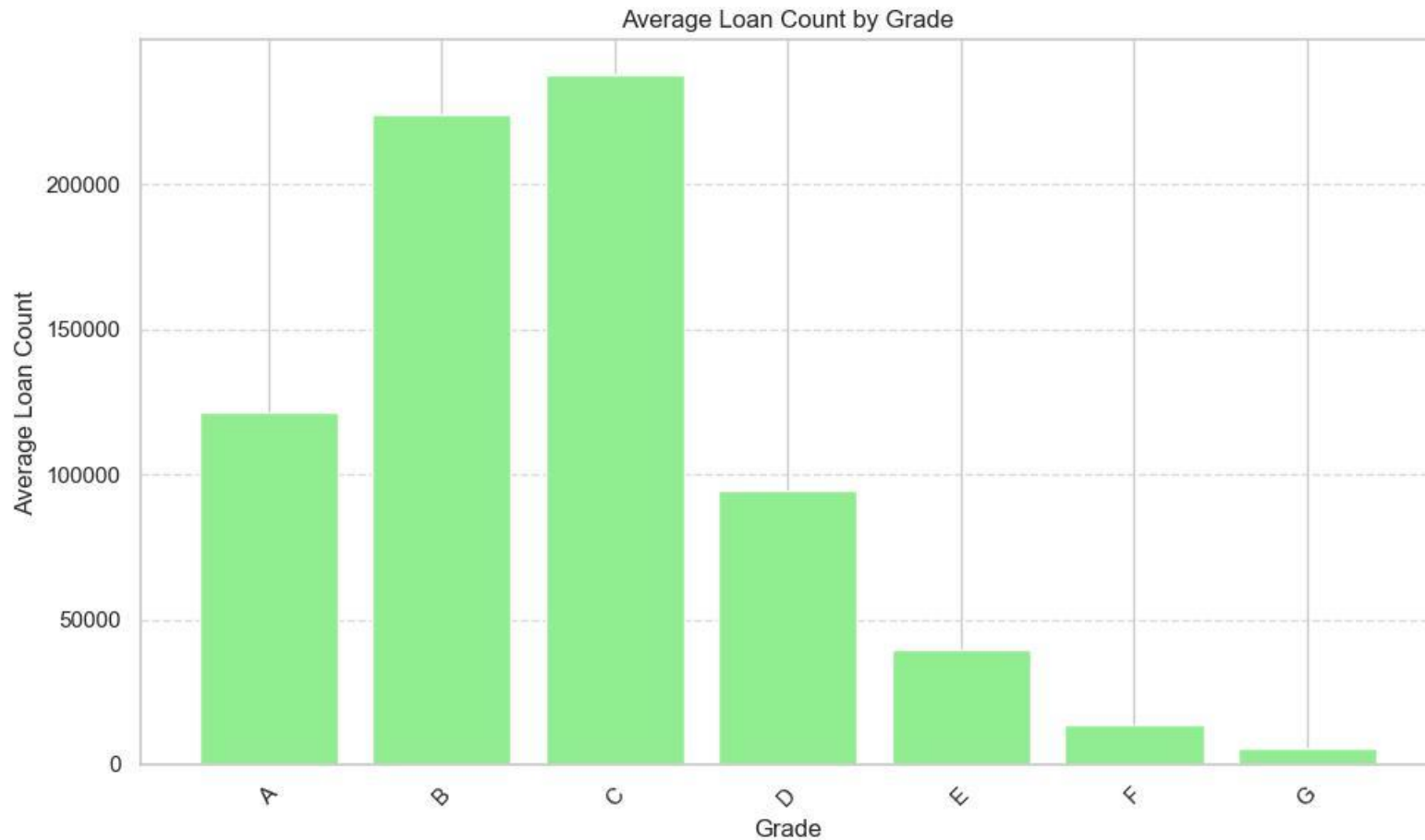
There is an abnormal behavior in employment length that we need to address since the data shown only shows ranges from 9.50 to 10.50

---

## 4. Which products have the greatest revenue growth potential and low risk of defaulting or delaying payments?

- **Revenue Analysis: Identifying Titles with Highest Paid Interest Income**  
Analysis of all loans to identify those generating the highest revenue through paid interest.
-

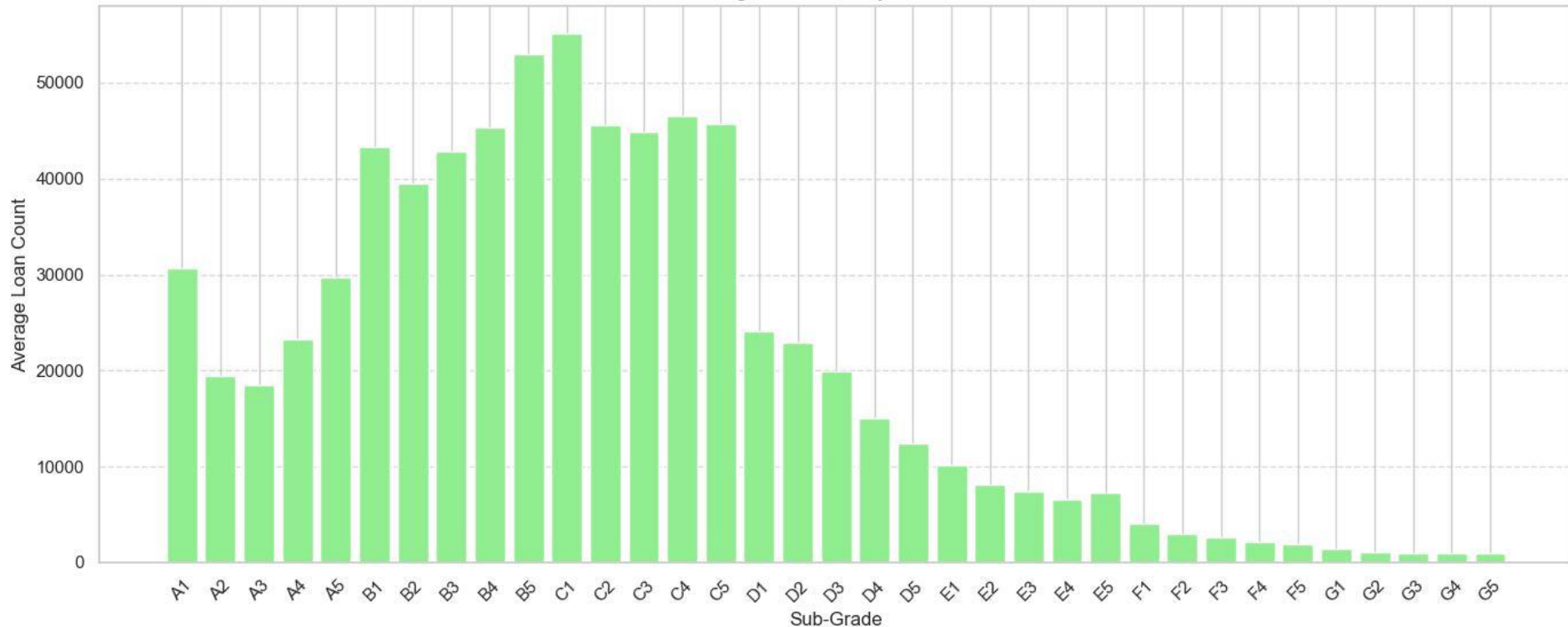
# — Average Loan Count by Grade





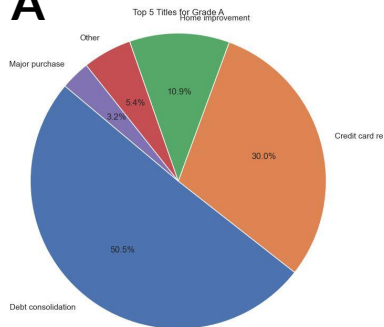
# — Average Loan Count by Sub-Grade

Average Loan Count by Sub-Grade

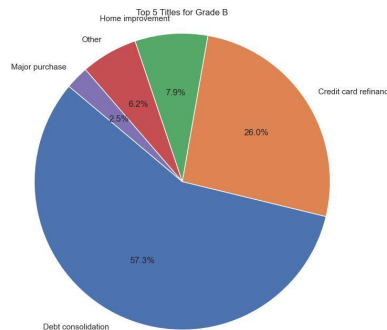


# — Top Products by Grade

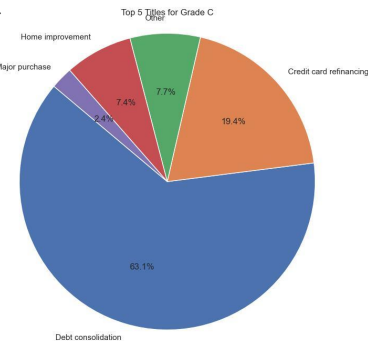
**A**



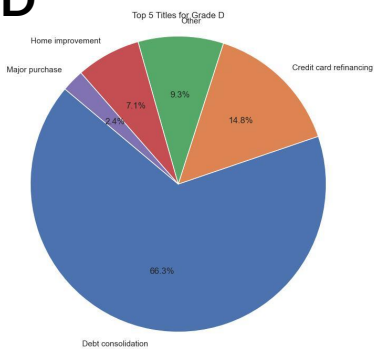
**B**



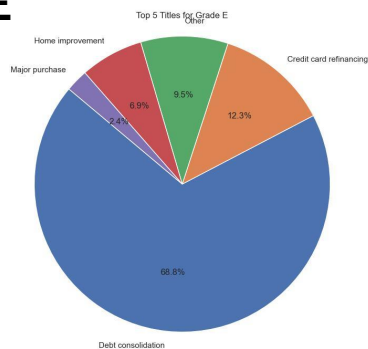
**C**



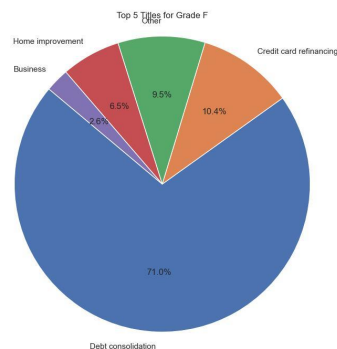
**D**



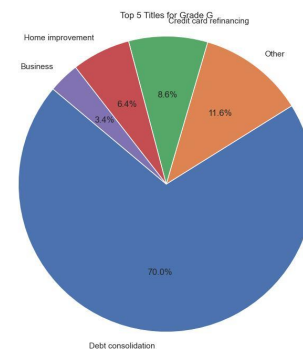
**E**



**F**



**G**

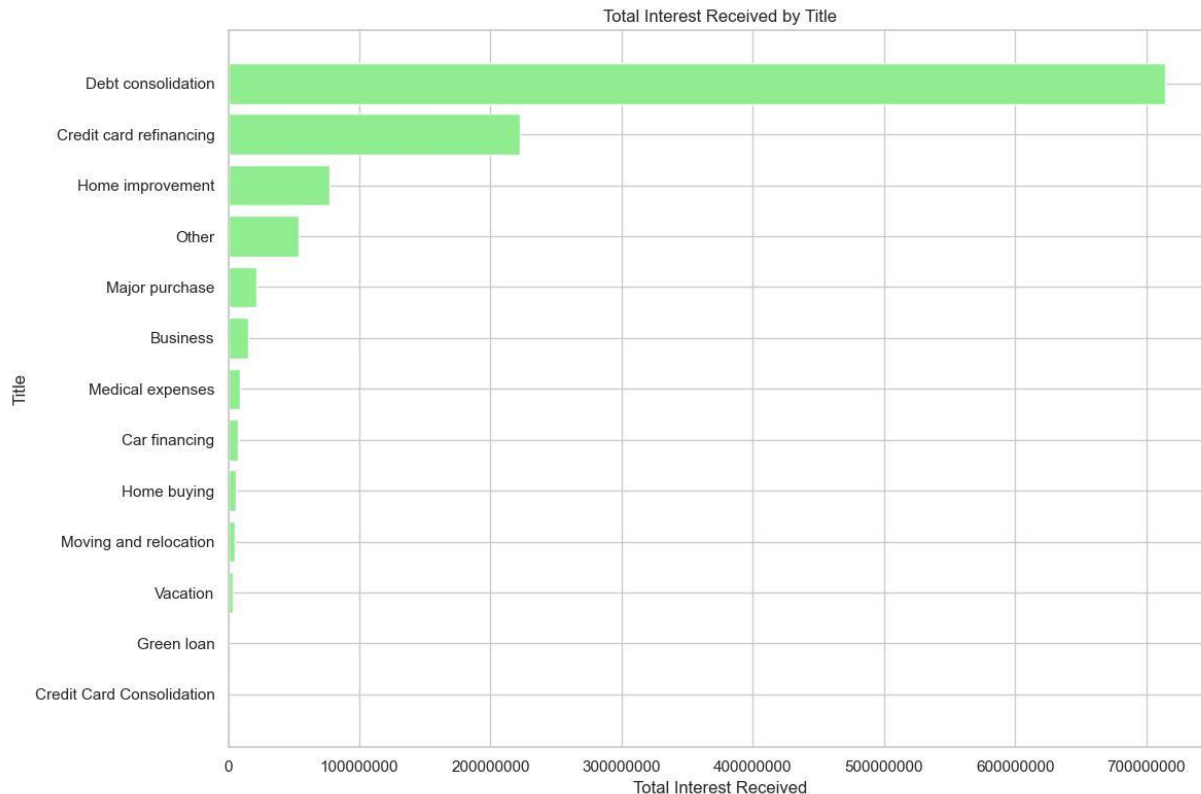


# Total Interest Received by Product

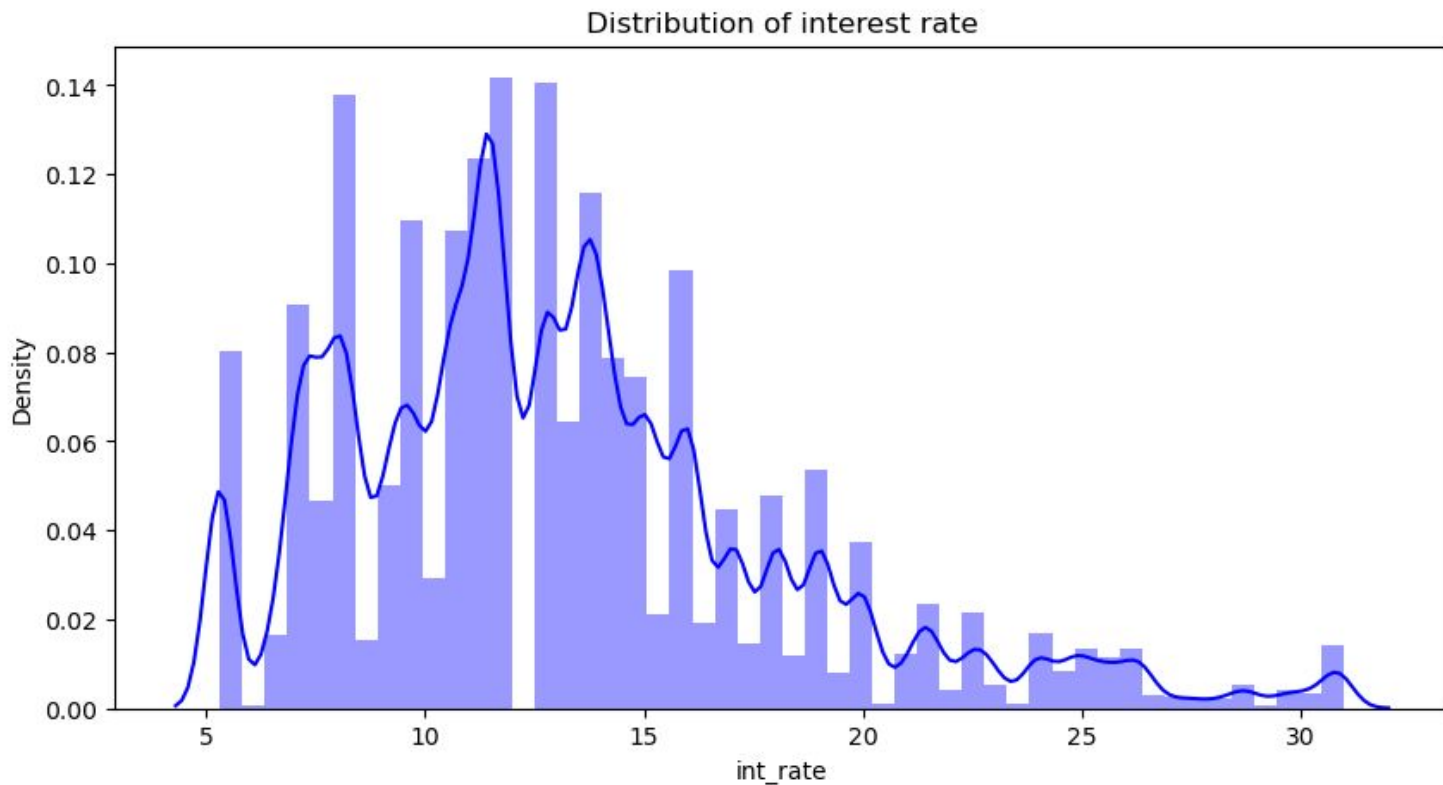
**Debt Consolidation:** This category has the highest total interest received, indicating that it's the most profitable loan purpose for Lending Club. (over \$700 million US dollars)

**Credit Card Refinancing and Home improvement** display the subsequent most profitable products for Lending Club. (\$222 million and \$76 million US dollars respectively)

**Vacation and Green Loan:** These categories have the shortest bars, suggesting they generate the least amount of interest. (\$3.5 million and \$570,000 dollars respectively)



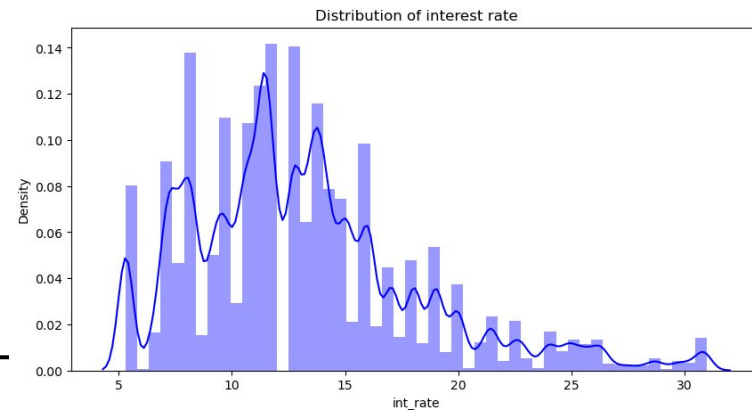
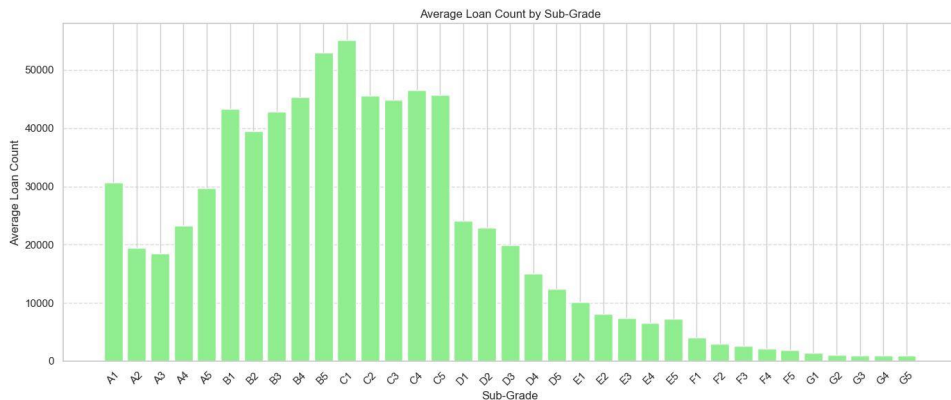
# Loan count by subgrade and Interest rate distribution



# Loan count by subgrade and Interest rate distribution

A direct correlation happens when analyzing the count of clients by subgrade and the interest rate distribution. At first glimpse totally different variables but with an identical graph due to the nature of its contents.

The reason of this graphs to be that similar is that the allocation of the highest interest rates go to the riskier clients. From D1 to G5 and that is exactly the same range where we as a company wouldn't invest.



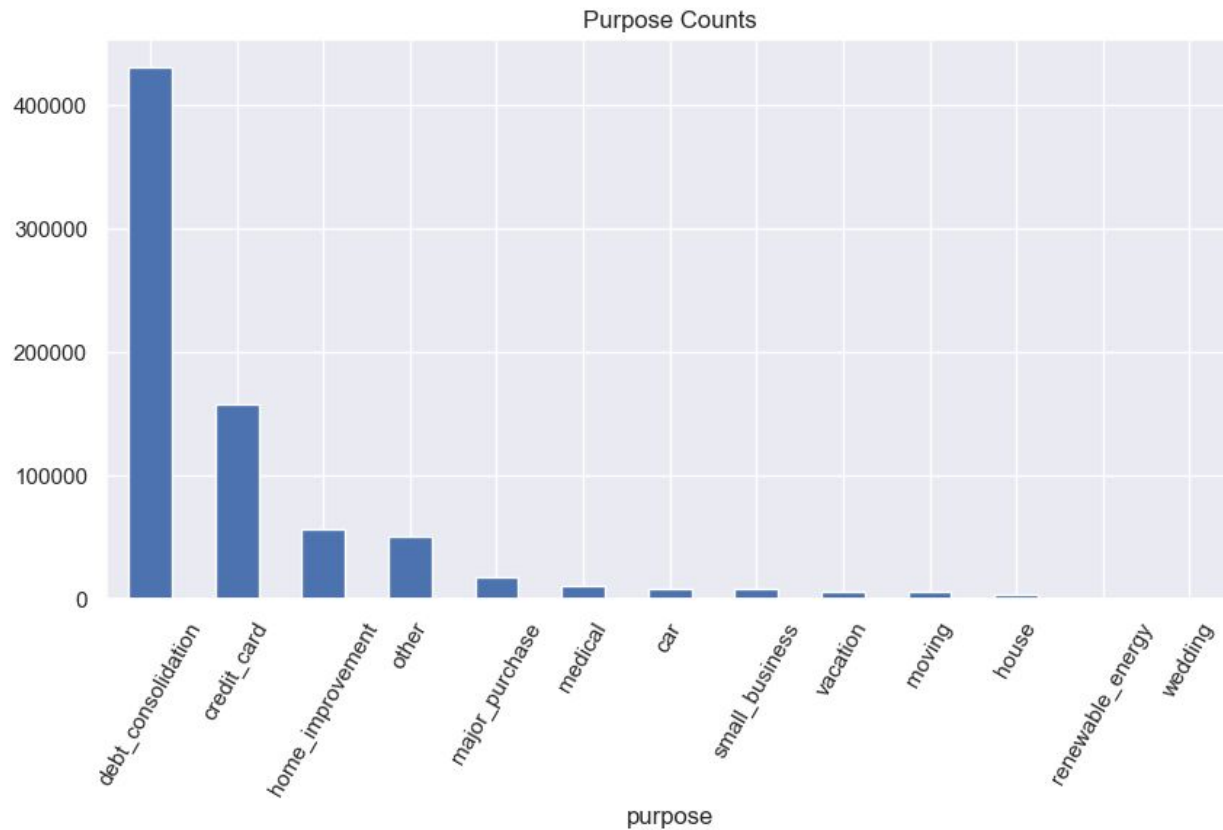
---

- **Risk Assessment and Product Categorization Based on Loan Grades and Sub-grades**

Based on the loan grades and sub-grades, determine the corresponding products for each category and identify those with the lowest risk.

---

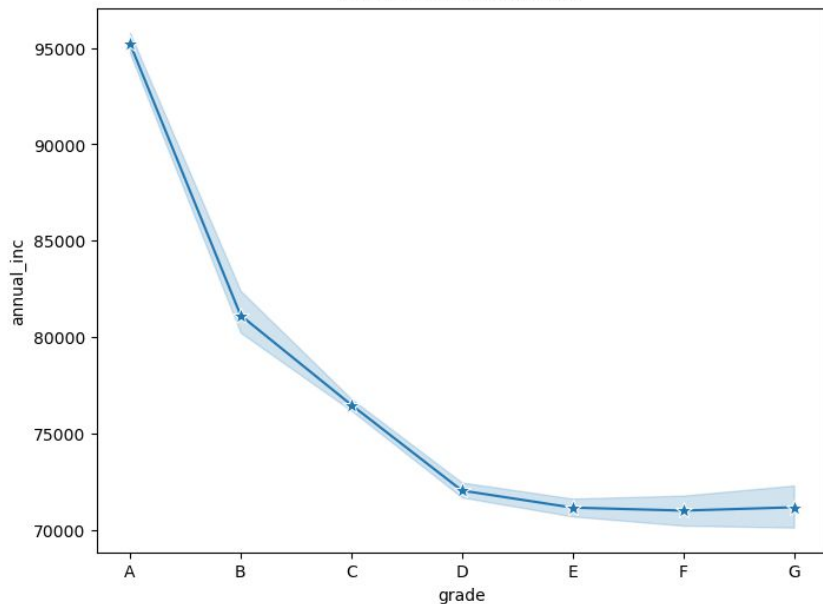
# ¿Why are they asking for a loan?



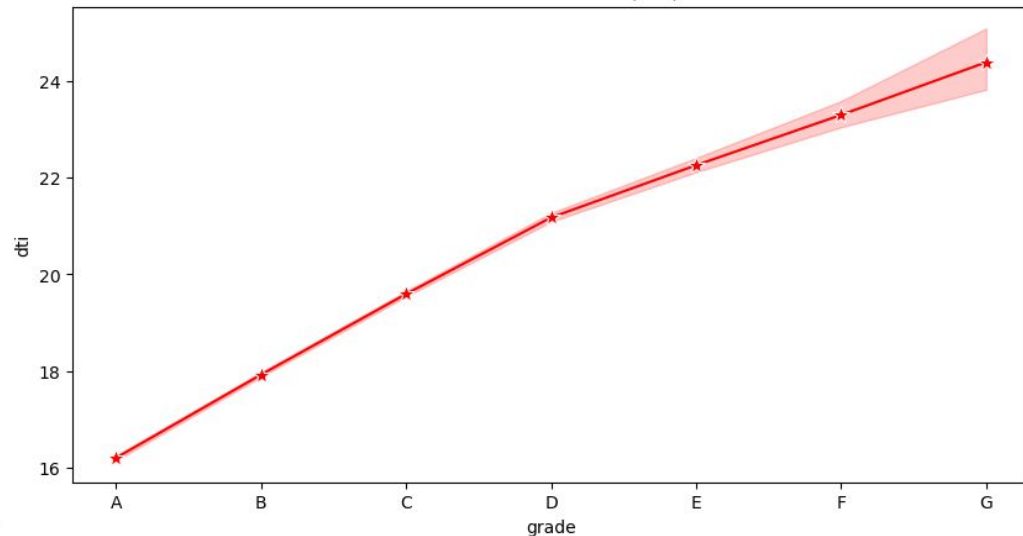
Another important data extracted from our Dataset is what our clients use the given loan for. For this, the vast majority uses its approved loan for Debt Consolidation and Credit Card Repayment.

# Repayment factors

Grade vs Annual income

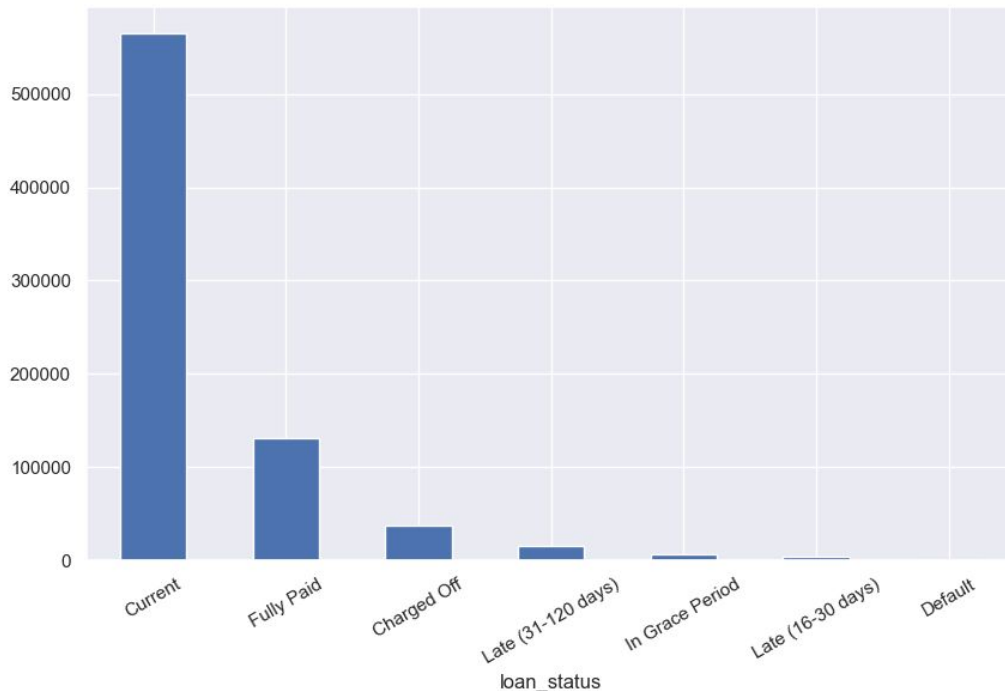


Grade vs Debt-to-Income (DTI) Ratio



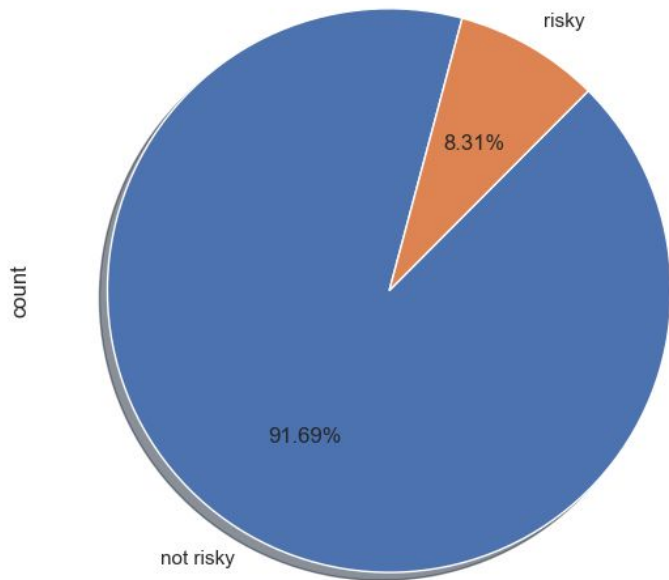


# Loan Status



Loan status is an important parameter to fully understand Portfolios Health and the followed strategy of loan allocation. From this data it is safe to assume that although Charged Off loans are the third biggest group, Fully Paid overcompensates the losses in a 4 to 1 ratio. To prevent this number from growing it would be beneficial to start a collections campaign for those Loans in the Late (31 to 120) status.

# Loan Status



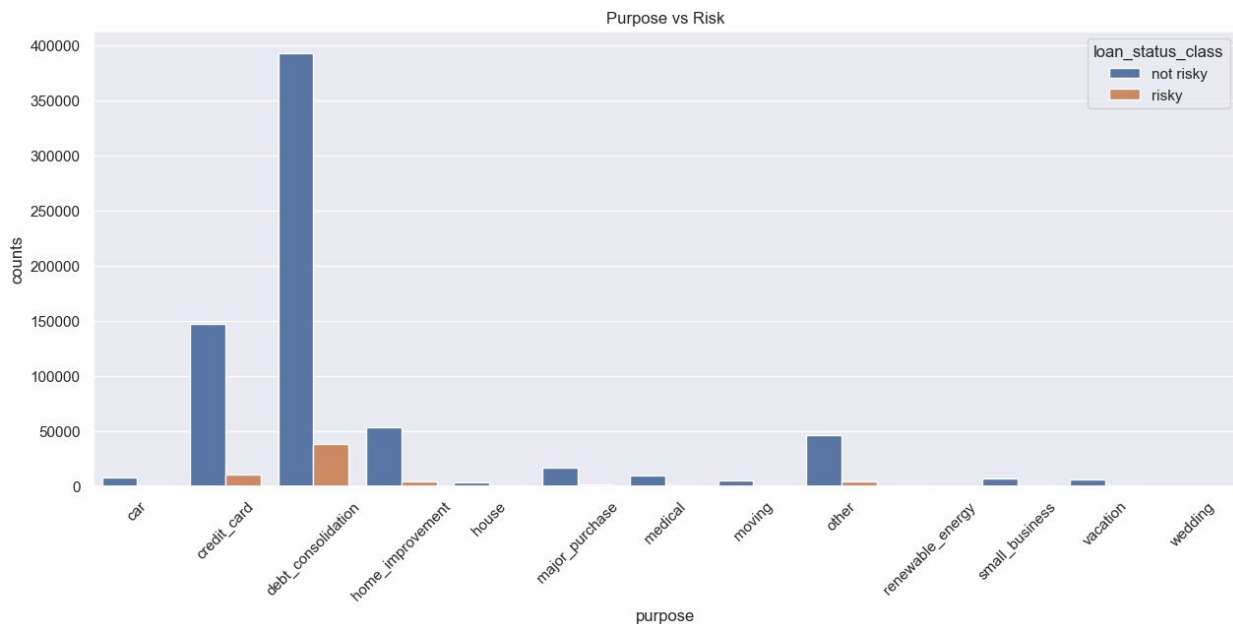
## Risky:

- Charged Off
- Late
- In Grace Period
- Default

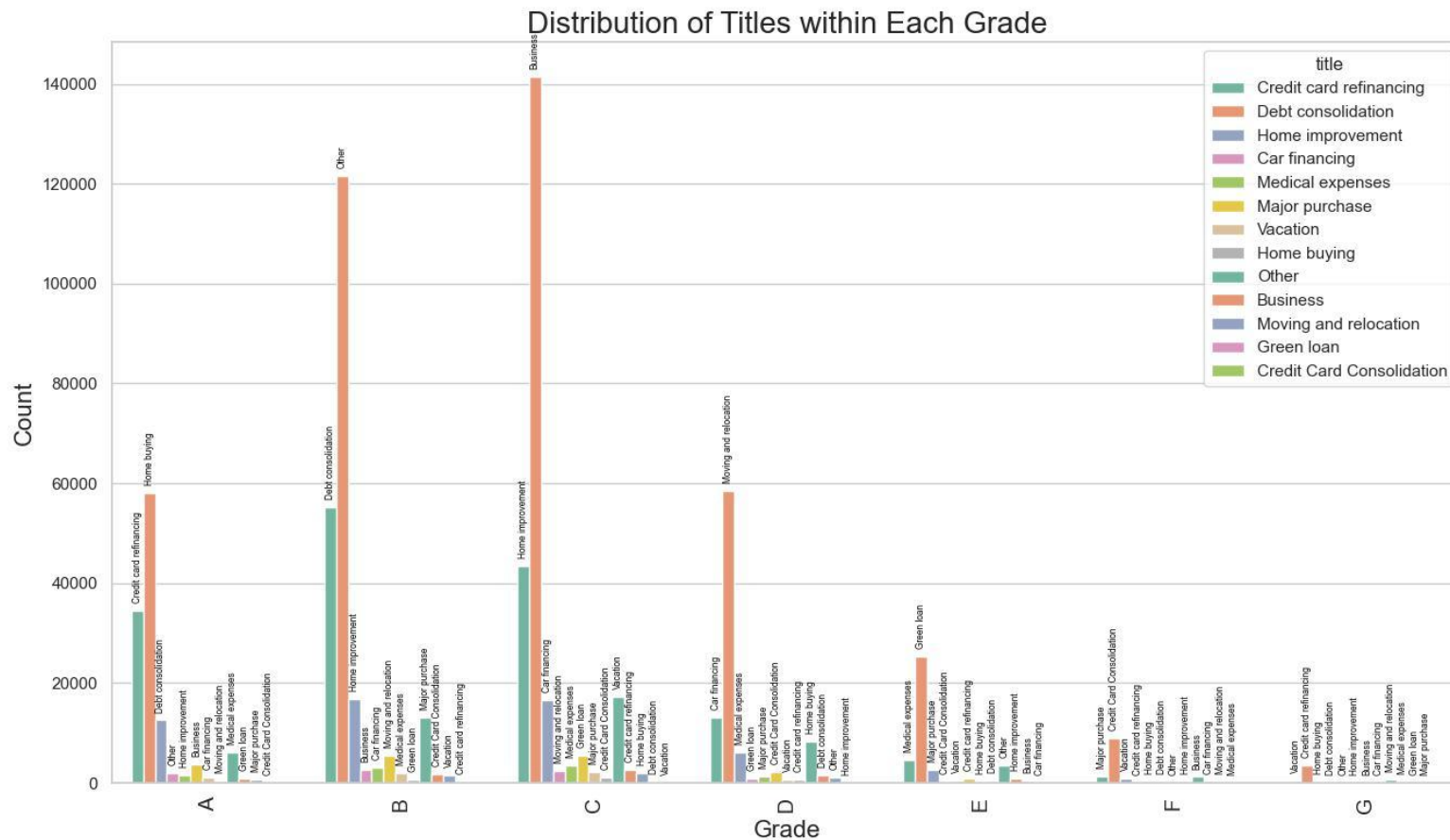
## Not Risky:

- Current
- Fully Paid

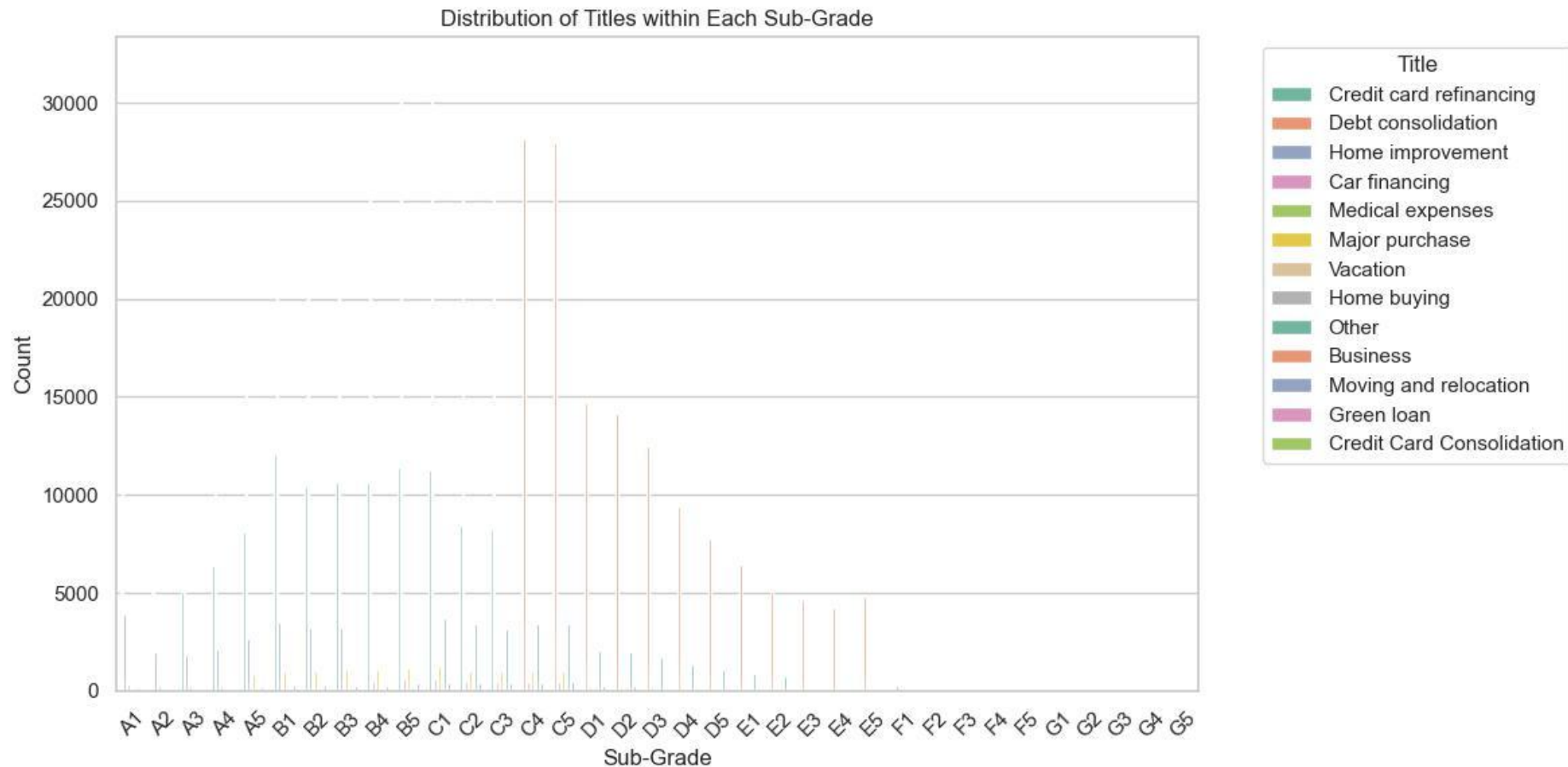
# Purpose and Risk Sentiment



In order to better understand the purpose of the loan and to match it with the Risk Sentiment, we found that our targeted products are the ones generating the most Risk. This gives us the upper hand to adjust the strategy and seek for other options to allocate debt.



# — Distribution of products within each Sub-Grade

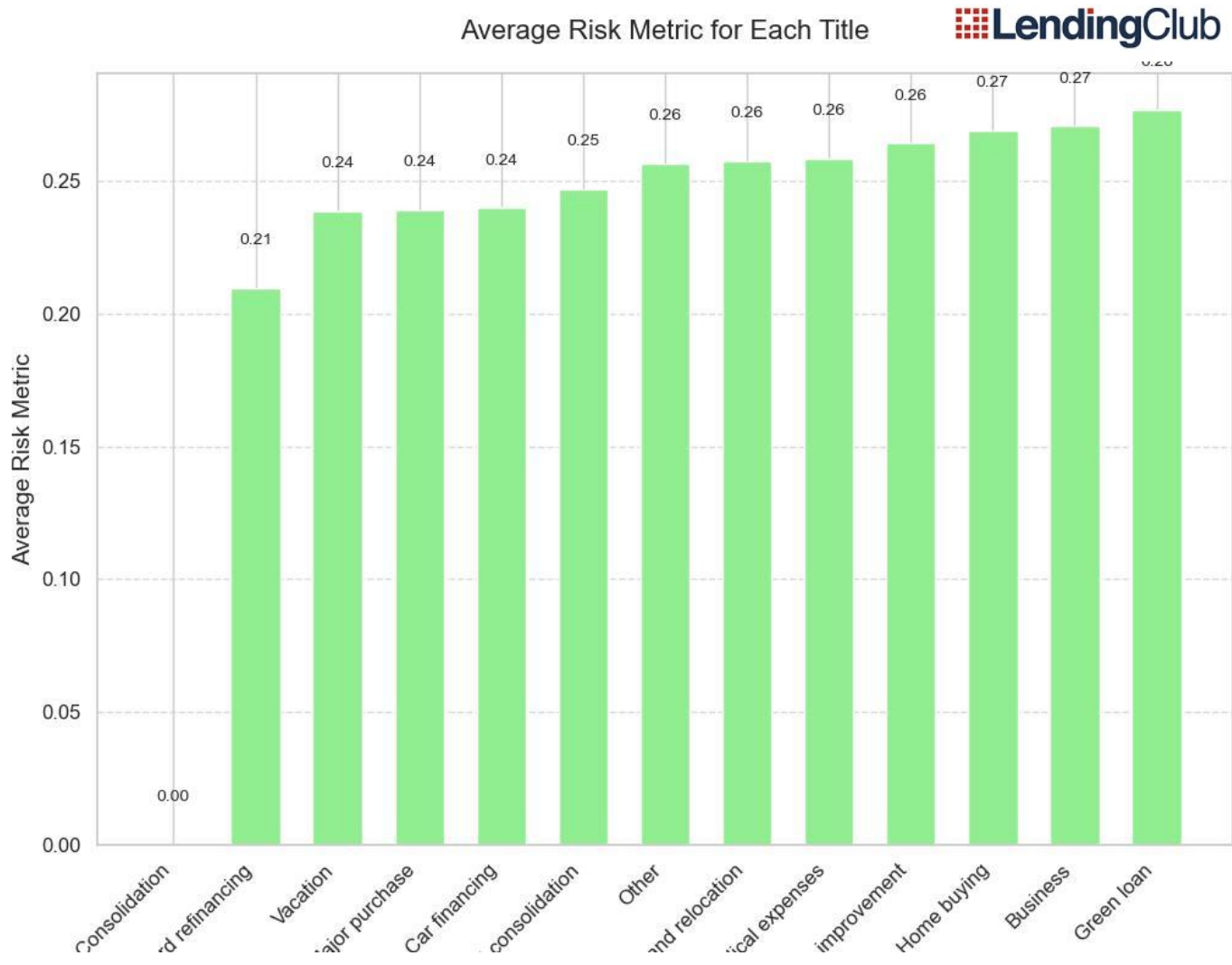


## — Most prominent products within each Grade and Sub-Grade

- **Debt Consolidation and Credit Card Refinancing:** These titles appear to have the highest counts across all grades, indicating that they are the most common purposes for loans across all credit grades. (417,193 and 152, 48 loans respectively in total)
  - **Home Improvement:** This title also appears frequently across all grades, summing 55,898 loans in total.
  - **Other Titles:** The remaining titles (such as Vacation, Green loan, Home buying/relocating, Moving and relocation, Car financing, Medical expenses, Major purchases, Businesses, Other) have varying counts across different grades. Being the most prominent "Major purchase" (17,939 loans), "Medical expenses", (10,128 loans), and "Car financing" (8,589 loans).
-

# Average Risk Metric for Each Product

These values represent the average risk metric for each loan purpose, which is calculated based on factors like [accounts now delinquent](#), [delinquencies in the past two years](#), and [late loan status](#).





# Major findings

**We believe that this analysis now opens more accurate questions to be handed to the revenue team to further explore opportunities of revenue growth.**

1. Are there any opportunities with segments who are given the major interest rates? do they represent a major threat?
  2. Shall we consider looking at some states with lower delinquency to seek for segments who are less riskier?
  3. Can we actually build a prevention algorithm to prevent delinquency in riskier segments and place more loans?
  4. Is debt consolidation an opportunity or a threat?
-



---

**Thank you!**

---