Predicting Maximum Credit Before Loan Default

Phase 1 Presentation

Adam Raabe
Vishnu Vardhan Reddy Putha
Pranav Tanaji Ghadge

Background Information

When banks are choosing who is allowed to receive a loan, **banks are forced to predict whether or not an individual will be able to repay in full**.

In efforts to prevent their own losses, banks have developed strategies for mitigating their own risk, such as:

- Credit ratings
- FICO scores
- Analysis of the debt-to-income ratio

Significant research has gone into creating methods of classifying potential loan candidates into either "loan" or "defaulting". However, **often overlooked is the maximization of loan quantities for the defaulting lenders**.

Context: Federal Regulations

There are many federal regulations regarding lending, such as:

- Home Mortgage Disclosure Act (HMDA)
- Equal Credit Opportunity Act (ECOA)
- Fair Credit Reporting Act (FCRA)
- Office of Foreign Assets Control (OFAC) requirements (see part 501, subpart C, section 501.604)

The important takeaway here is that for each of these, lending institutions need to be able to legally explain their reasonings. With black-box algorithms, it is nearly impossible to describe what reasoning could go into someone being denied a loan.

Context: Similar Work

Predictive Modeling for Loan Defaults (Zhu, L): Binary classification using logistic regression, random forest, neural network, extreme gradient boost and ensemble.

Prediction and Analysis of Financial Default Loan Behavior Based on Machine Learning Model (Chen, H): Discusses using alternative weighting to compensate for unbalanced data to create more accurate models in the cases of linear regression and neural networks.

Explainable prediction of loan default based on machine learning models (Zhu, X et al., 2023): Discusses methods of interpreting the outputs of Logistic regression, decision tree, XGBoost, and LightGBM learning models.

Context: Relevant Discussion

Equifax Receives Utility Patent for Innovative NeuroDecision® Technology (2018): Equifax (a lending organization) advertises their movement to a transparent machine-learning model.

What percentage of your income should go towards your mortgage (Chase Bank): Among other things, describes basic important factors that a bank tells its customers it uses to determine mortgage approval.

How Is AI Transforming Lending and Loan Management? (Hovsepyan, T et al., 2023): Discusses how AI is used to facilitate the process of loan application and approval.

Goal of Project

This project hopes to create an algorithm that will allow users to:

(1) Determine whether an individual should be considered safe from default

(2) Determine what the maximum amount of credit he/she could receive before being unable to return the full quantity of the loan

If something similar were applied professionally, it could allow banks to capitalize on opportunities for loans that might otherwise be ignored as an overly-risky investment.

Data: Source and Background

Our data is taken from a Kaggle dataset originally scraped from a loan-providing online source called Lending Club. The data spans loans granted from 2007 to 2014.

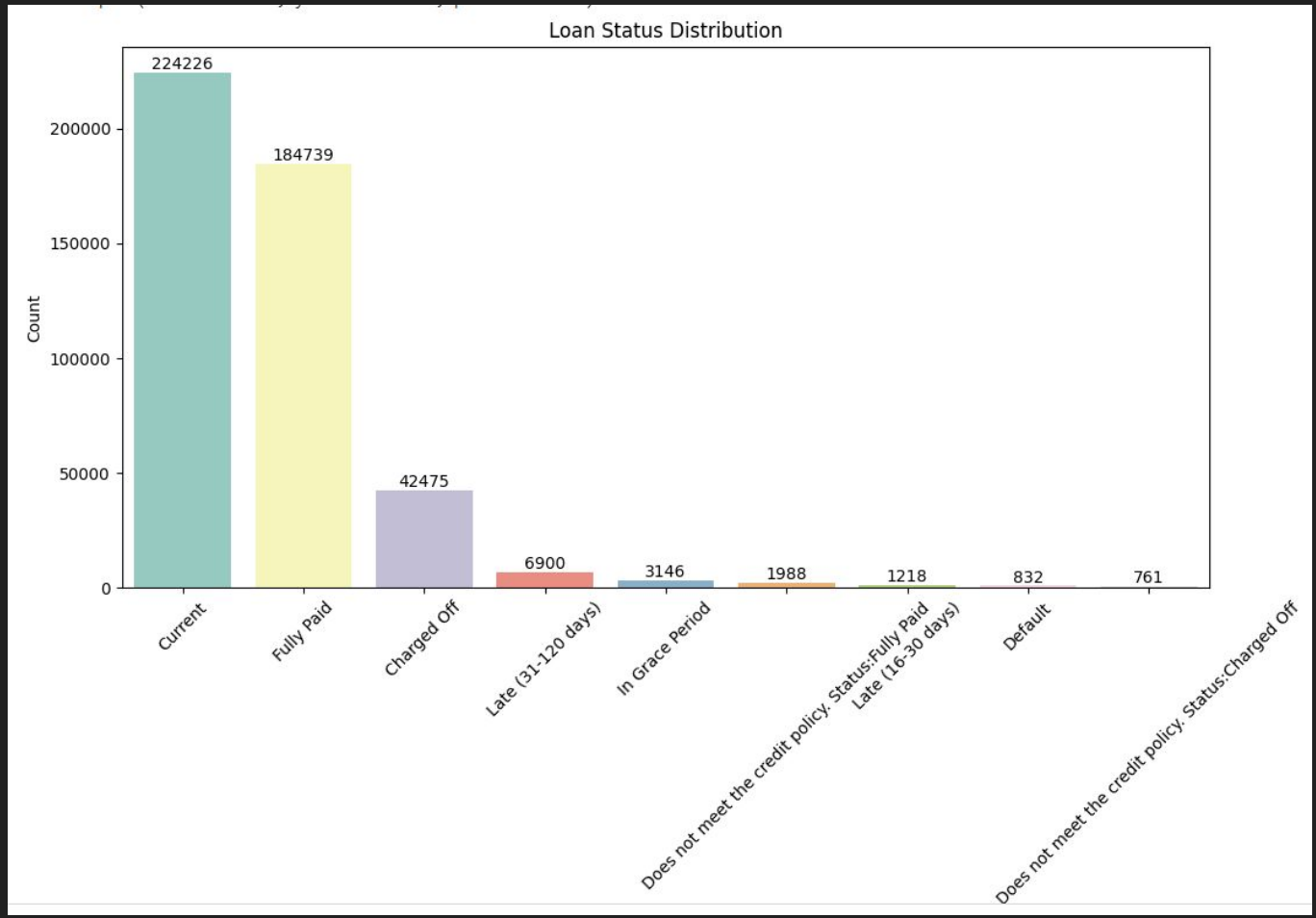
This data is **NOT** from credit card loans, even though some of the quantities for individual loans are equivalently small.

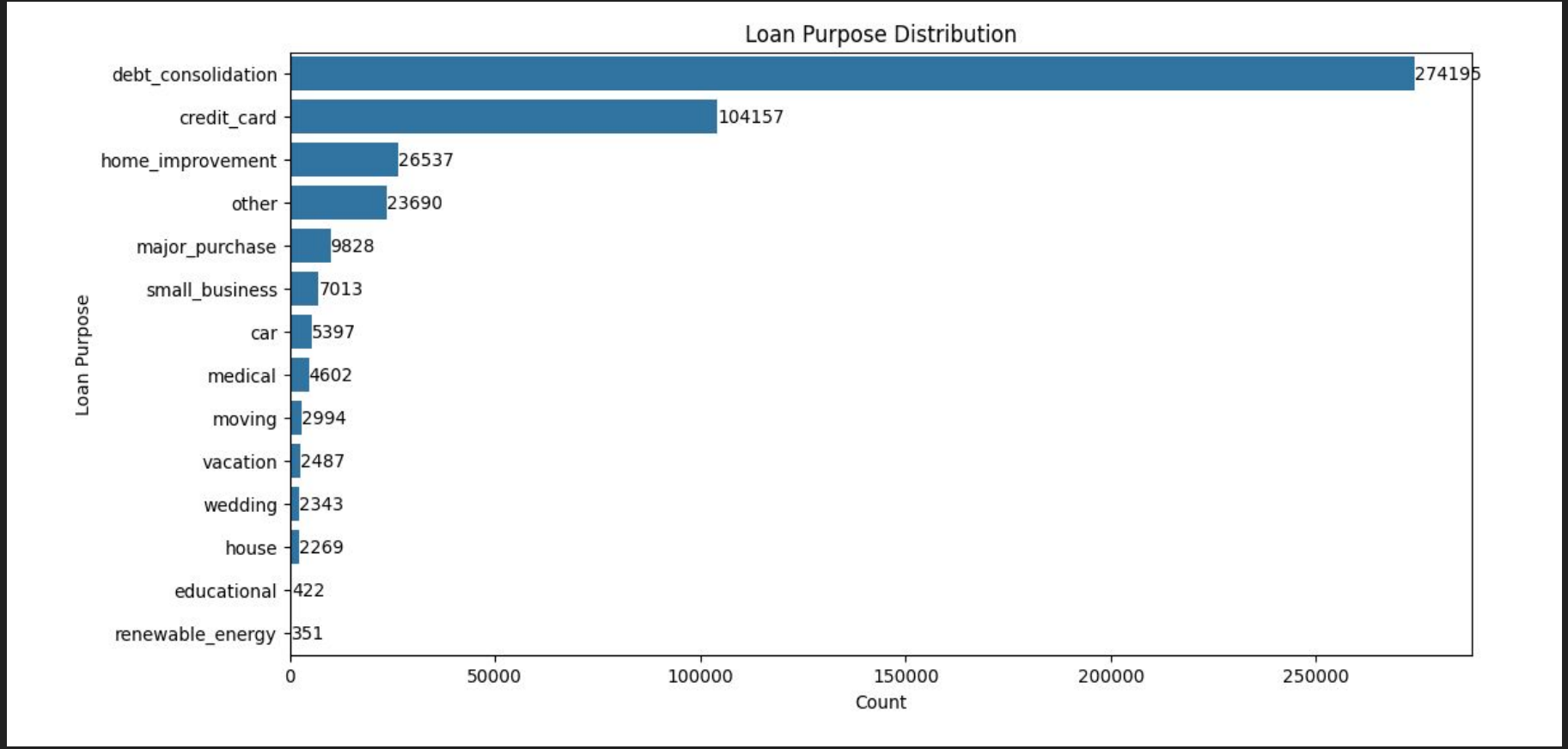
Lending Club has since stopped providing similar information publicly.

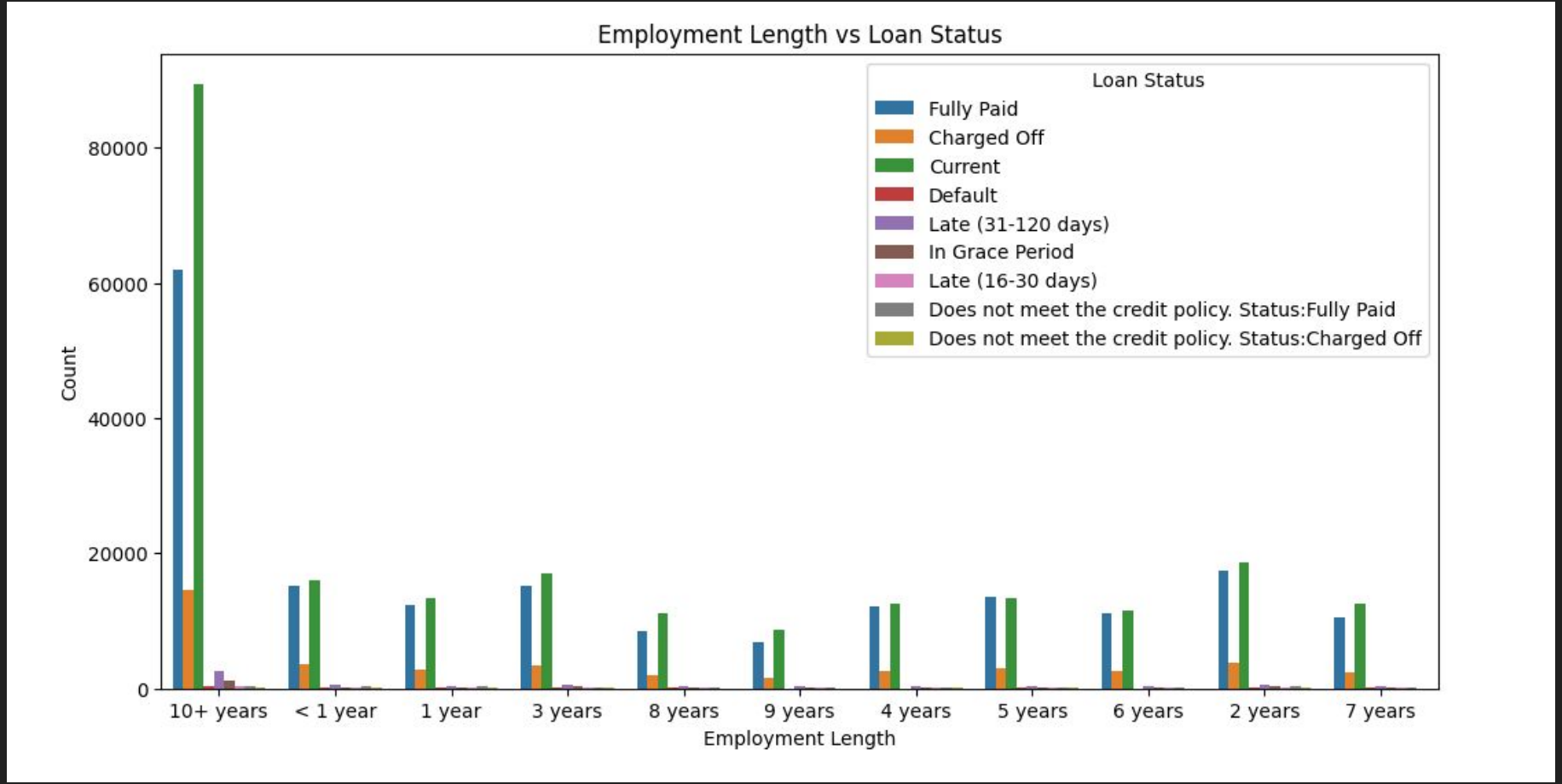
The dataset is composed of 466,000 records that contain information about 75 columns.

Data: Key Features

- Loan Amount
- Funded Amount
- Term
- Installment
- Grade
- SubGrade
- Employment Length
- Home Ownership
- Verification Status
- Purpose
- Payment Plan
- Title
- DTI ratio
- Months Since Last Record
- Total Payment
- Loan Status


Loan Status Distribution
Count
224226
184739
42475
6900
3146
1988
1218
832
761
0
50000
100000
150000
200000
Current
Fully Paid
Charged Off
Late (31-120 days)
In Grace Period
Does not meet the credit policy. Status:Fully Paid
Late (16-30 days)
Default
Does not meet the credit policy. Status:Charged Off


Loan Purpose Distribution
debt_consolidation
credit_card
home_improvement
other
major_purchase
small_business
car
medical
moving
vacation
wedding
house
educational
renewable_energy
274195
104157
26537
23690
9828
7013
5397
4602
2994
2487
2343
2269
422
351
Loan Purpose
0
50000
100000
150000
200000
250000
Count


Employment Length vs Loan Status
Loan Status
Fully Paid
Charged Off
Current
Default
Late (31-120 days)
In Grace Period
Late (16-30 days)
Does not meet the credit policy. Status:Fully Paid
Does not meet the credit policy. Status:Charged Off
Count
0
20000
40000
60000
80000
10+ years
< 1 year
1 year
3 years
8 years
9 years
4 years
5 years
6 years
2 years
7 years
Employment Length