FLP **Fintec**
Financial Tecnology

Predicting Loan Risk

From Borrower and Loan Characteristics

Asif Shazad | Jose Traboulsi | Karla Lopez | Seyhr Waqas

FLP **Fintec**
Financial Tecnology

Executive Summary

This project develops a predictive machine learning model leveraging borrower and loan data to help lenders assess loan eligibility and riskiness.

Primary Focus

Enhance the approval process and the pricing of loan risk (i.e. setting interest rates and loan terms to manage risk exposure).

Stay competitive in the market

As the lending landscape evolves, companies must adapt by rapidly assessing borrower risk to make timely, data-driven lending decisions.

FLP **Fintec**
Financial Tecnology

Key Goals

Innovative Financial Services

- **Enhanced Loan Risk Prediction Accuracy**
  - Identify a loan as low or high risk based on a combination of borrower and loan characteristics.
- **Fair and Precise Loan Pricing**
  - Once risk level is determined, lenders can review results and data to price loans.
- **Data-Driven Portfolio Management**
  - Informed credit decisions help monitor loan portfolio risk .

Fair and Precise Loan Pricing

Once risk level is determined, lenders can review results and data to price loans.

Data-Driven Portfolio Management

Informed credit decisions help monitor loan portfolio risk.

FLP **Fintec Financial Tecnology**

Project Steps

Phases of Execution from Start to Finish

Clean Dataset

- Identify key columns to keep for our model
- Understand the distribution of features
- Apply feature engineering where applicable
- Identify a target

Classification Model Training & Evaluation

Train and optimize classification models (Random Forest, XGBoost, Neural Networks).

Class Imbalance Correction Methods

Handle class imbalance using SMOTE and class weighting.

Continuous Model Improvement

Track model performance and make adjustments to models based on findings to improve predictive power.

FLP **Fintec**
Financial Tecnology

Dataset

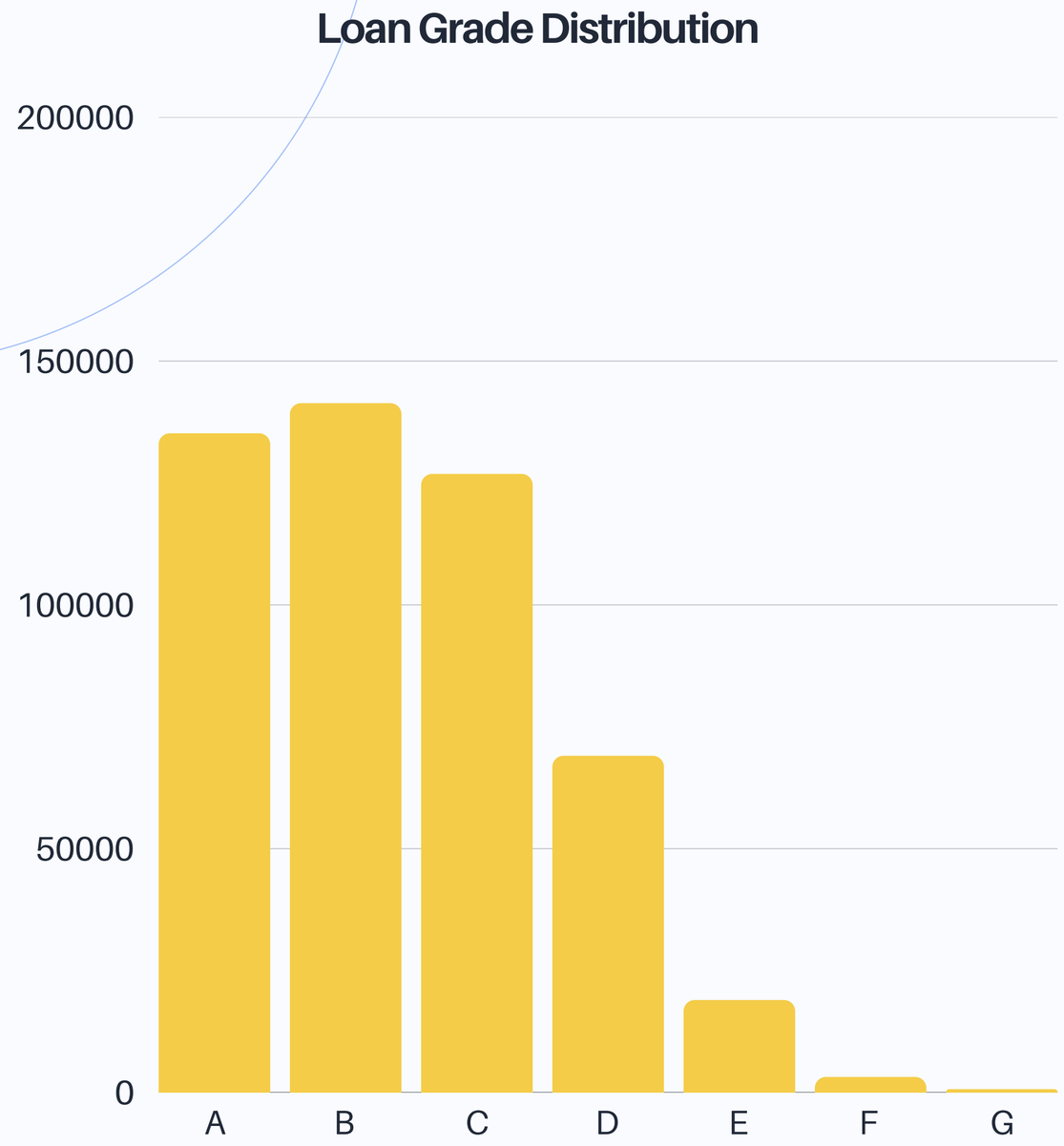
Raw Data Summary

Key borrower features include income, loan amount, debt-to-income ratio (DTI), employment status, default history, credit limit, and other financial indicators relevant to credit risk assessment.

- Loan features:
  - Loan purpose
  - Loan grade
  - Interest rate
  - Individual/joint, and more.

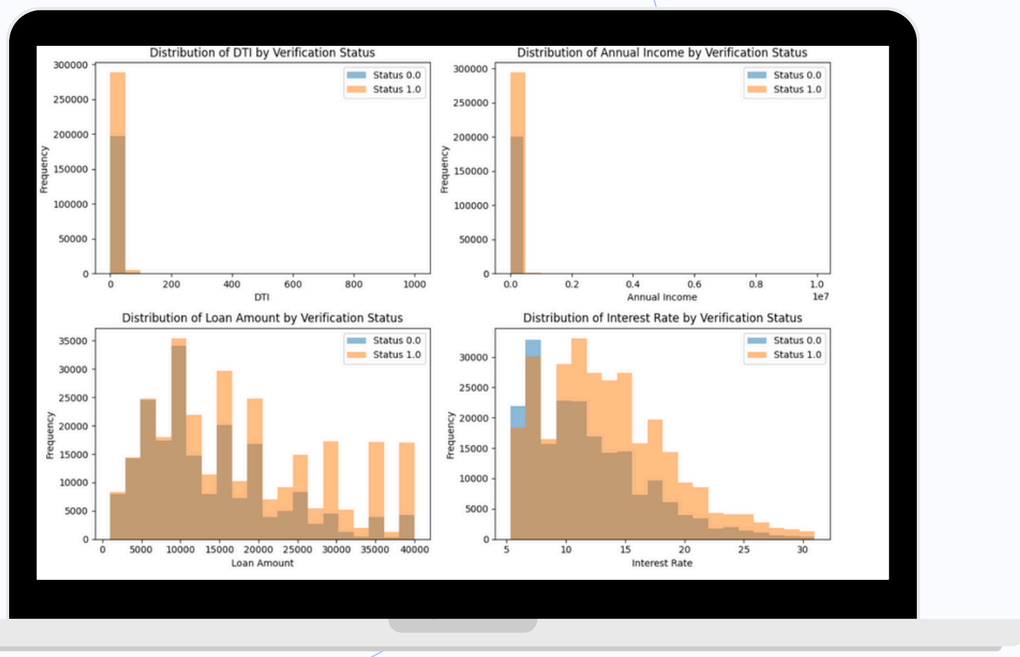
File size used : Lending Club via Kaggle (US)

1.1 GB & 2M+



FLP **Fintec**
Financial Tecnology

EDA & Dataset Challenges



Profiling, Cleaning, and Preprocessing Requirements

The dataset presents several preprocessing challenges:

- Large file size—over 1.1 GB
- Numerous null values and 'overlapping' columns
- Inconsistent labeling and lack of standardized categories

And importantly...

grade	avg_loan_amnt	avg_int_rate
A	15775.350651368206	7.08
B	16196.126339617304	10.92
C	16177.99546708711	14.71
D	15905.81713640182	19.49
E	15305.560976896297	25.18
F	19027.409448818897	29.48
G	19741.50521609538	30.82

grade	avg_int_rate	prev_avg_int_rate	int_rate_diff	int_rate_diff_from_a
A	7.081353114806082	7.081353114806082	0.0	0.0
B	10.91878513068961	7.081353114806082	3.837432015883528	3.837432015883528
C	14.711079463935045	10.91878513068961	3.792294333245435	7.629726349128963
D	19.489089737273535	14.711079463935045	4.77801027333849	12.407736622467453
E	25.17508914442214	19.489089737273535	5.685999407148607	18.09373602961606
F	29.482948031496313	25.17508914442214	4.307858887074172	22.401594916690232
G	30.81675111773485	29.482948031496313	1.333803086238536	23.73539800292877

FLP **Fintec Financial Tecnology**

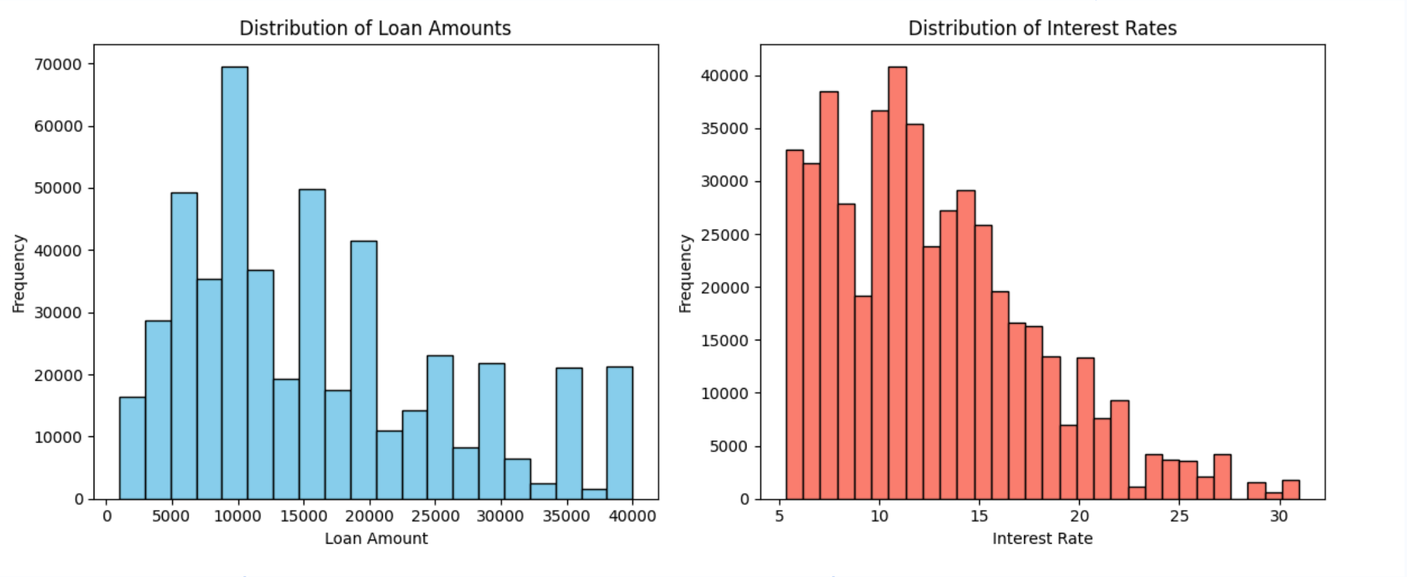
1. Imbalanced Classes:

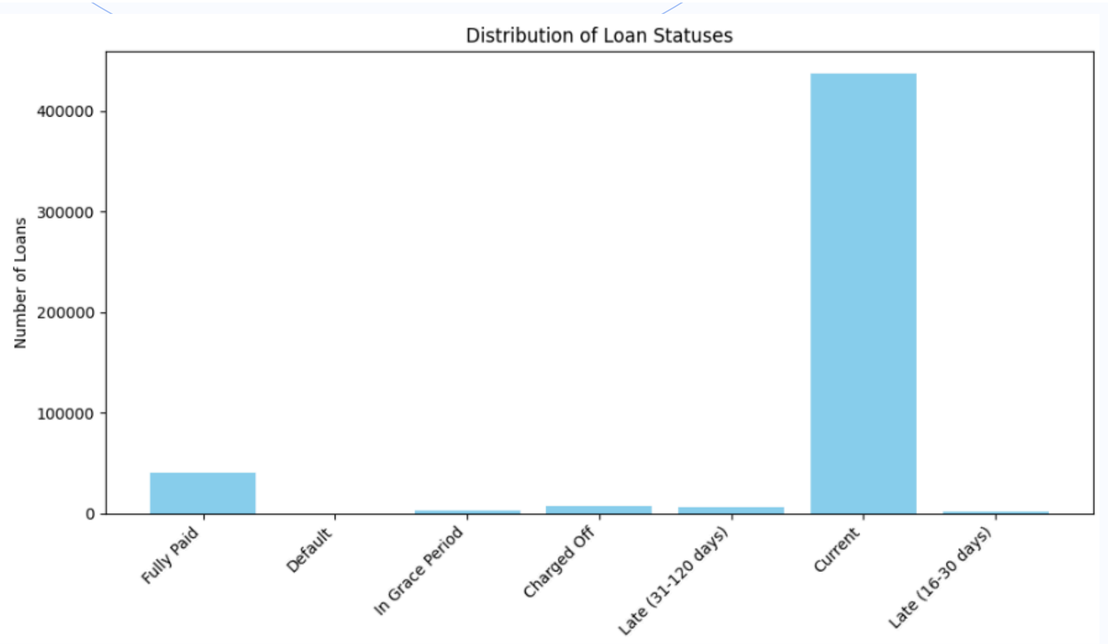
- Significant concentration of loans in lower-risk categories—specifically Grades A to C.
- Vast majority of loans are labeled as “Current” or “Fully Paid,”

2. Redundancy in Risk Definition:

- There is a circular relationship between loan grade, interest rate, and risk.

These factors can bias the model toward predicting the dominant classes, reducing its sensitivity to high-risk cases.





FLP **Fintec**
Financial Tecnology

AWS RDS PostgreSQL Deployment

- Deployed a PostgreSQL database using AWS RDS Free Tier
- Data normalized to 2NF and stored in 4 relational tables to ensure data integrity and reduce redundancy
- Table structure:
  - borrowers, loans, credit_history, delinquency_info
- Uploaded cleaned dataset using Python libraries psycopg2 and SQLAlchemy
- Enabled cloud-based data access from notebooks

FLP **Fintec**
Financial Tecnology

Models Explored

for Loan Risk Prediction

Random Forest Accuracy

74%

Keras Sequential Neural Network Accuracy

73%

Logistic Regression Accuracy

71%

Perceptron Accuracy

59%

FLP **Fintec Financial Tecnology**

Random Forest Optimization

Using RandomizedSearchCV

- Converted multi-class grades to binary target (0 = Low, 1 = High)
- Scaled features using StandardScaler
- Applied SMOTE to balance class distribution in training data
- Tuned hyperparameters using RandomizedSearchCV (n_estimators, max_depth, etc etc
- Saved model in .pkl
- Integrated with Gradio app for real-time prediction Some inconsistencies remain in high- risk loan classification

```
Random Forest Tuned Model Accuracy: 0.7390
Classification Report:
              precision    recall  f1-score   support

    Low Risk       0.72      0.75      0.74    219317
   High Risk       0.76      0.73      0.74    232817

    accuracy                           0.74    452134
   macro avg       0.74      0.74      0.74    452134
weighted avg       0.74      0.74      0.74    452134
```

FLP **Fintec**
Financial Tecnology

Thank You So Much

Questions?

Github Repo:

https://github.com/karlaflopezm/Fintech_Loan_Prediction