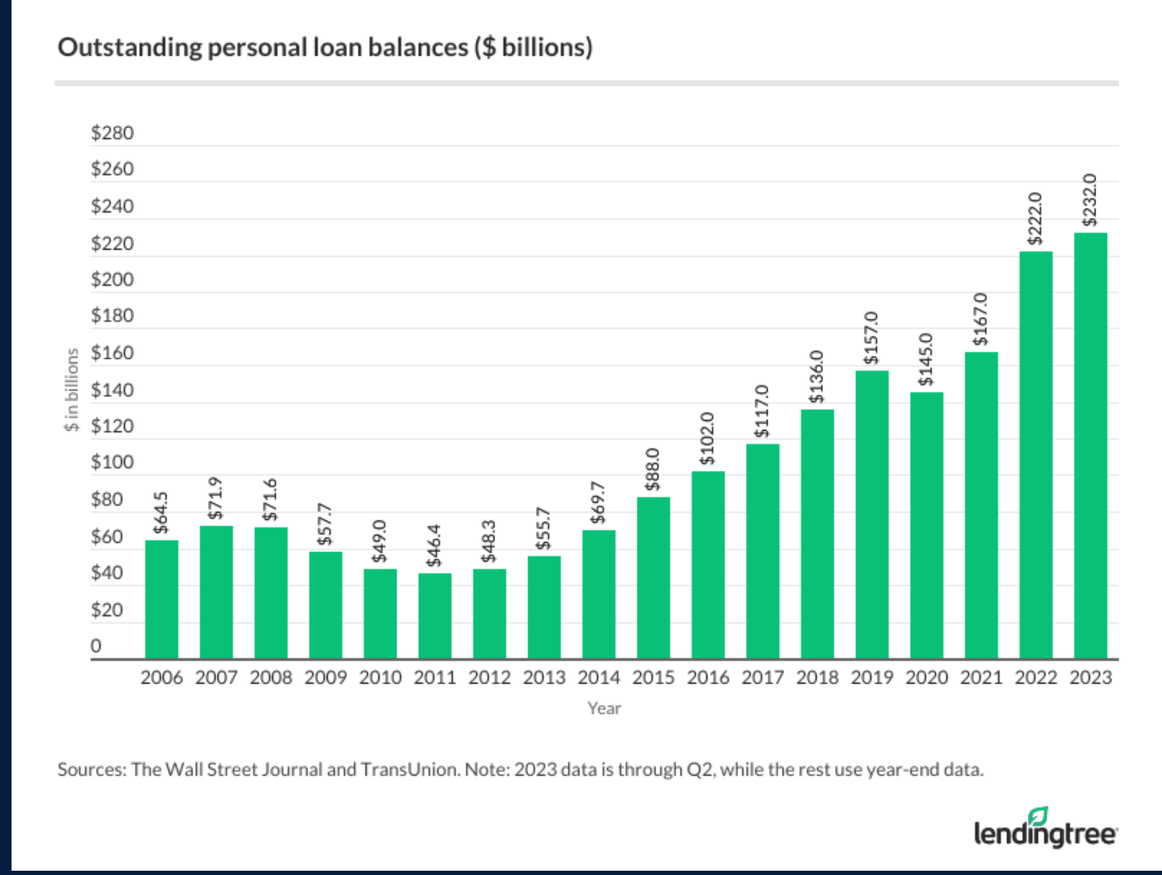
LOAN EVALUATION

BIG IDEA

How might we use ML to identify patterns and predictors to more efficiently predict a loan applicants credit worthiness, overcoming the information asymmetry within peer to peer (P2P) loan markets.



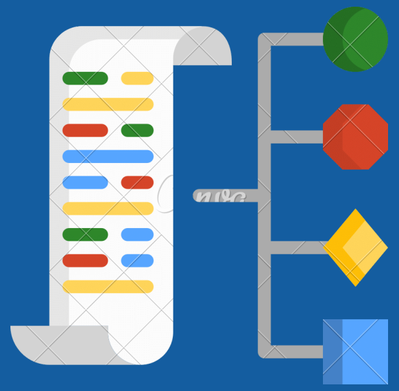
THE WHY

- Allocate capital more efficiently
- Cut out grey metrics
- Increased transparency
- Decreased evaluation time
- Decreased cost in evaluation process

EXECUTION

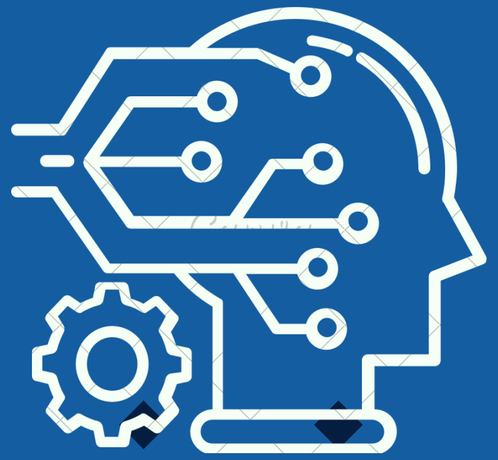
1. Clean data, remove extraneous variables , optimize the dataframe

2. Perform feature engineering and EDA



3. Perform benchmark logistic regression model

4. Train and evaluate more complex models such as KNN, XGBoost, decision trees



IMPACT

Practical Impact

- Improved consumer confidence and trust
- Economic Growth
- Risk mitigation
- Operational Efficiency

Market size

- $16 billion in P2P loans made
- Global personal loans forecasted to be 620 Billion in 2032

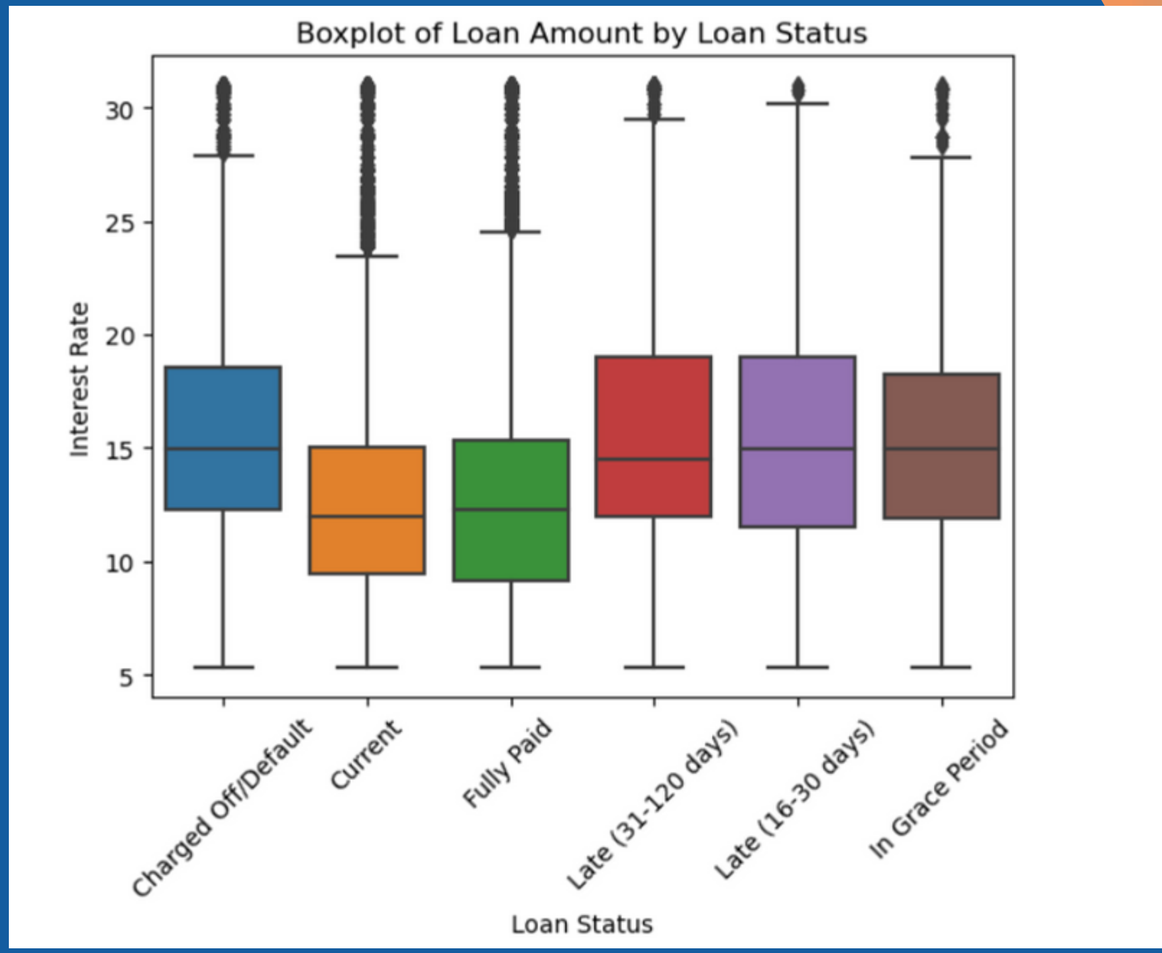
LendingClub

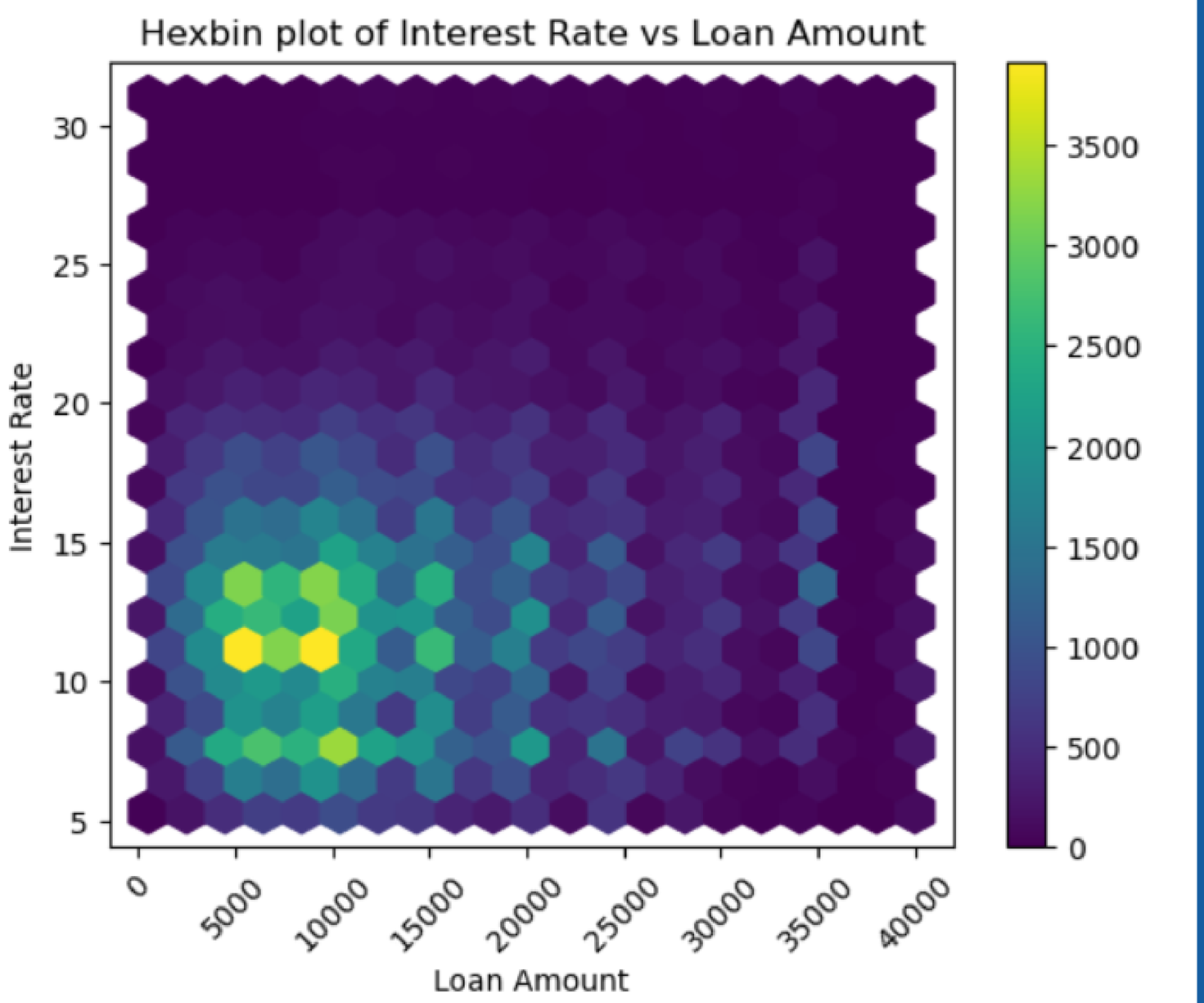
THE DATASET

Real-World Data-set from a US based P2P lending platform, spanning 11 years

Potential Issues

- Dataset size
- Changes in data recording
- Extraneous columns needed for special loan conditions
- Columns used for internal record keeping
- Verification status
- Loans that were in progress when the data was collected
- Model Benchmarking





FINDING

- Clustering of loans in the $5000 to $10000 USD
- High interest rate for smaller loans

NEXT STEPS

Complete data Cleaning and EDA

More managable, space optimized dataset

Finish grouping loans with uncertain features

- Loan to income ratio
- Loan Purpose one hot encoding
- Employment length to numeric

Fit a benchmark logistic model