



INTEREST RATE PREDICTOR

CAPSTONE PROJECT DATA SCIENCE CERTIFICATE

LENDING CLUB

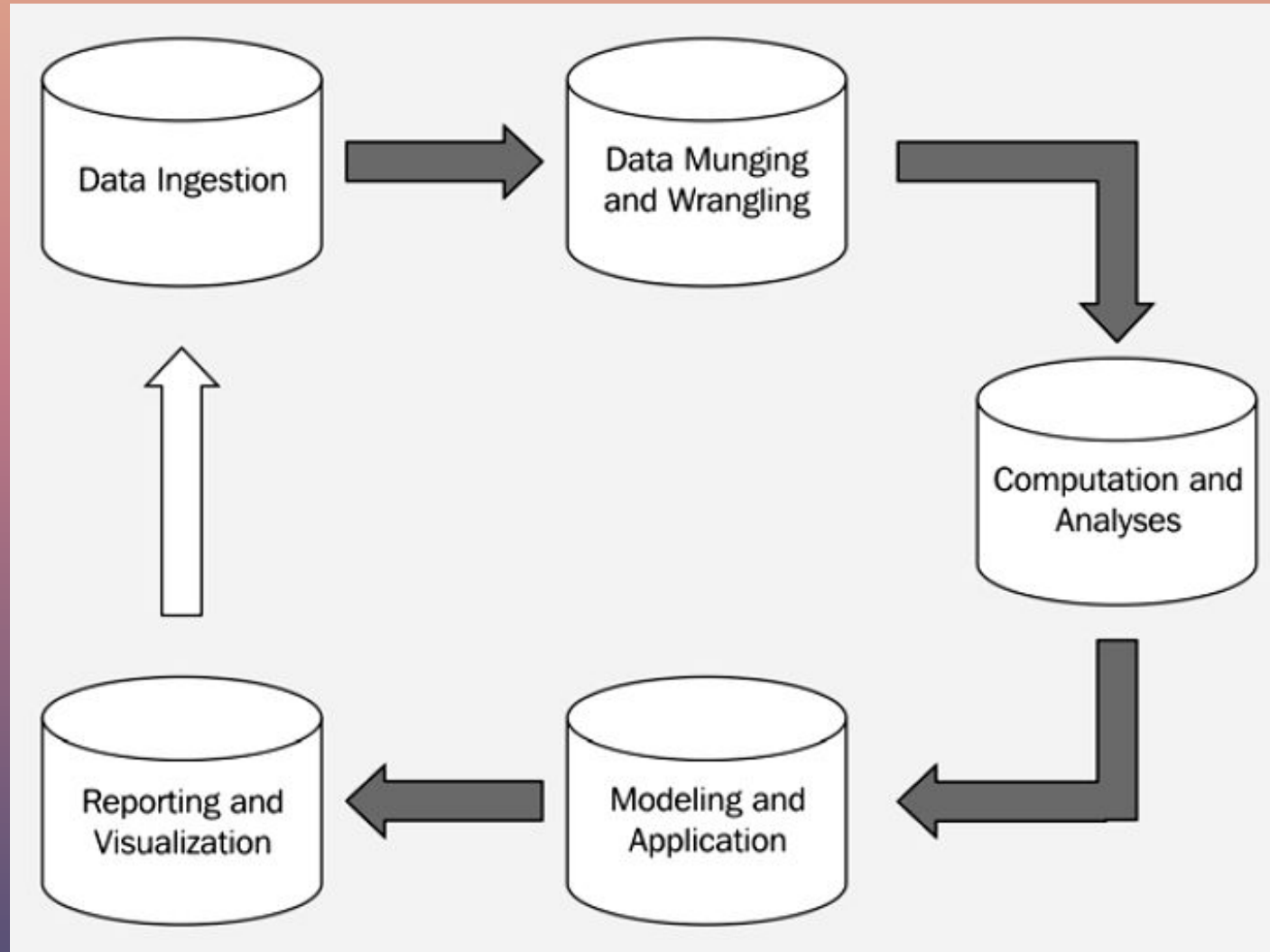
- Peer-to-peer lending, a growing method of debt financing, allows people to borrow and lend money without traditional financial institutions.
- LendingClub offers loans from \$1,000 to \$35,000 for individuals and loans from \$15,000 to \$300,000 for businesses.

HYPOTHESES:

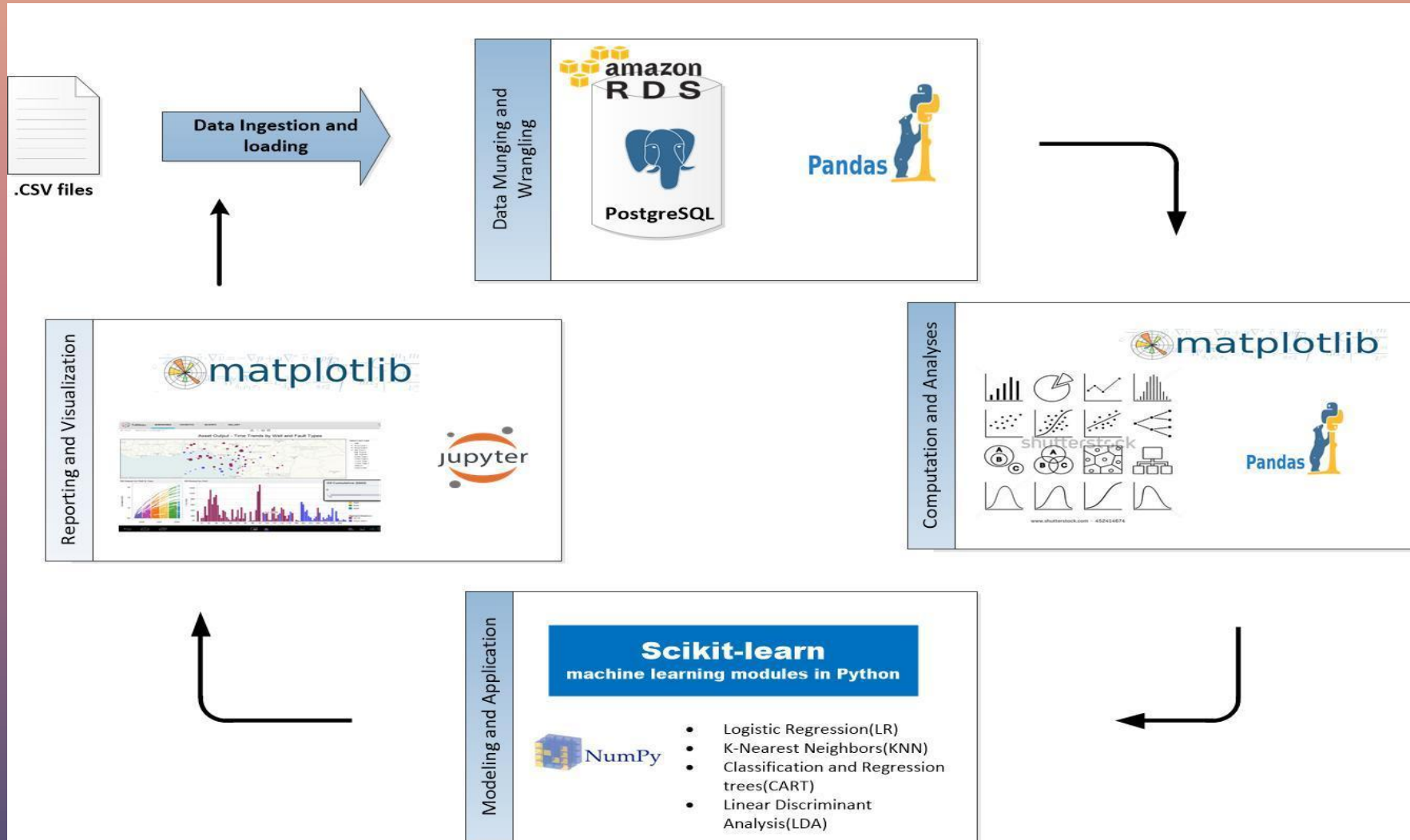
- What will be the loan interest rate for LendingClub customers given their profile and request?
- What variables help predict the ideal loan interest rate?
- Can we build a simple model with no more than 15 features and still reliably calculate an interest rate prediction?
- Would a classification model be more reliable than a regression model?



Data Science Pipeline



Design and Architecture



DATA INGESTION AND LOADING

- Loaded data from CSV files to tables in a PostgreSQL database
- Six CSV files, each containing a year's worth of data. We created six tables in the PostgreSQL database using SQL, one for each CSV file.
- Merged the six tables into one table for maximum efficiency.

DATA MUNGING AND WRANGLING

Data Verification: 235,629 instances and 153 features.

Removing irrelevant features: dropped joint application data, post-loan features and other unrelated features.

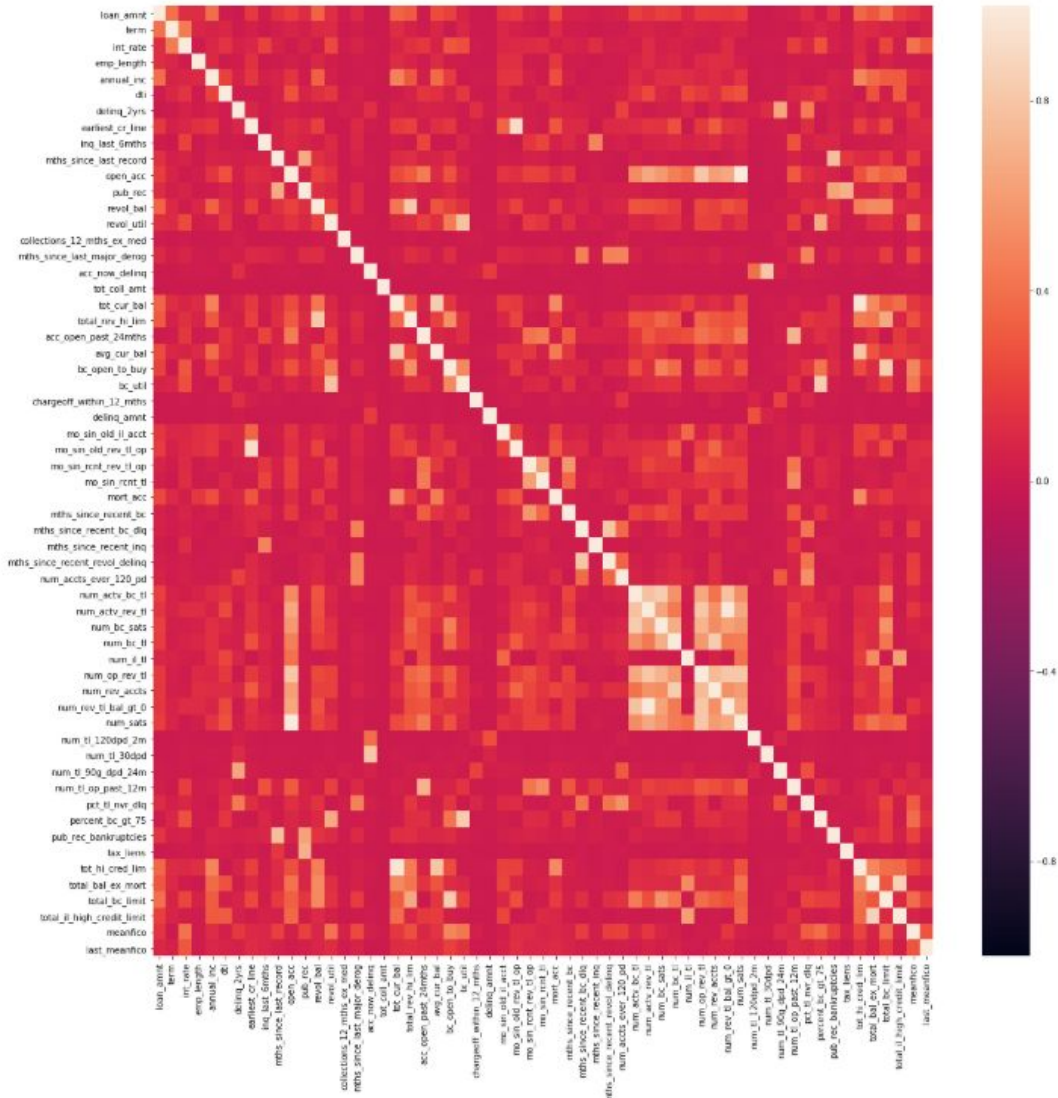
Addressing null values: replaced with zero or correspondent minimum value.

Formatting: converted variables to numeric by removing the string elements.

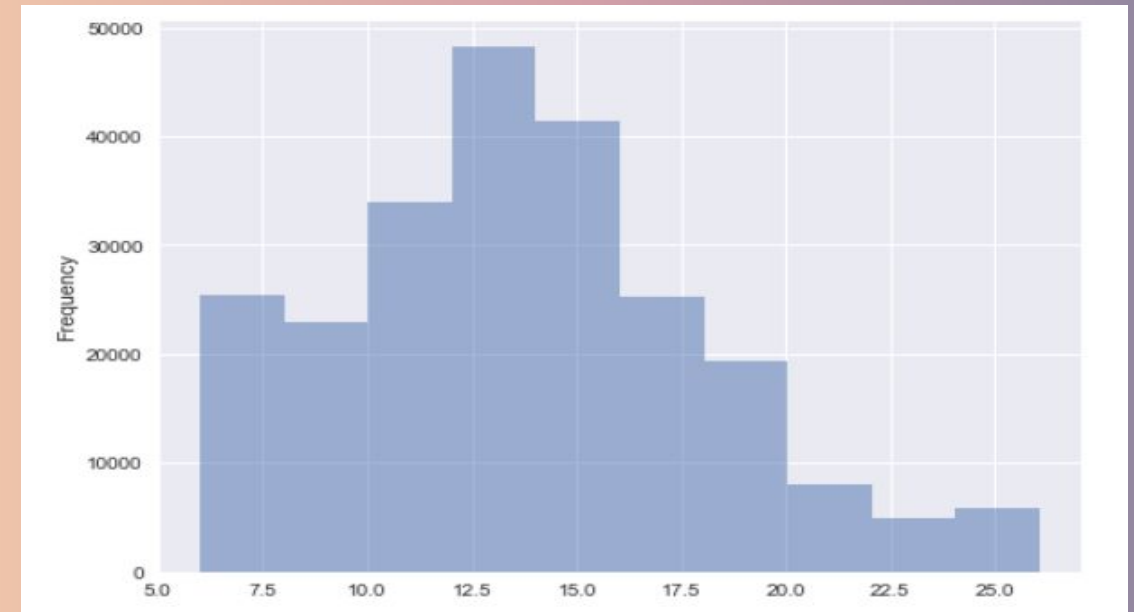


COMPUTATION AND ANALYSIS

CORRELATION HEATMAP



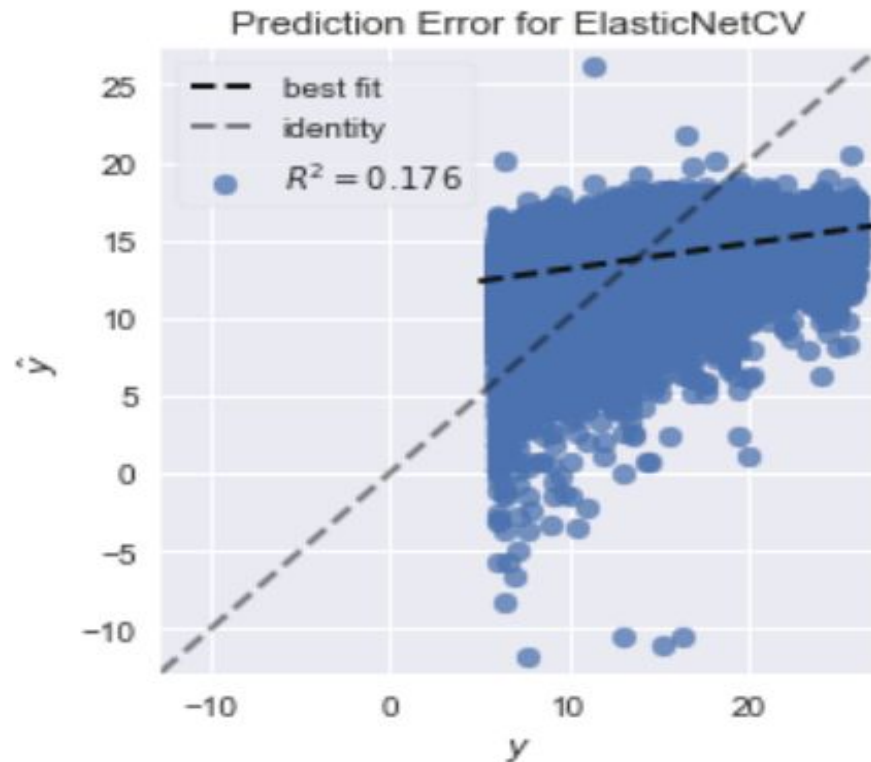
INTEREST RATE HISTOGRAM



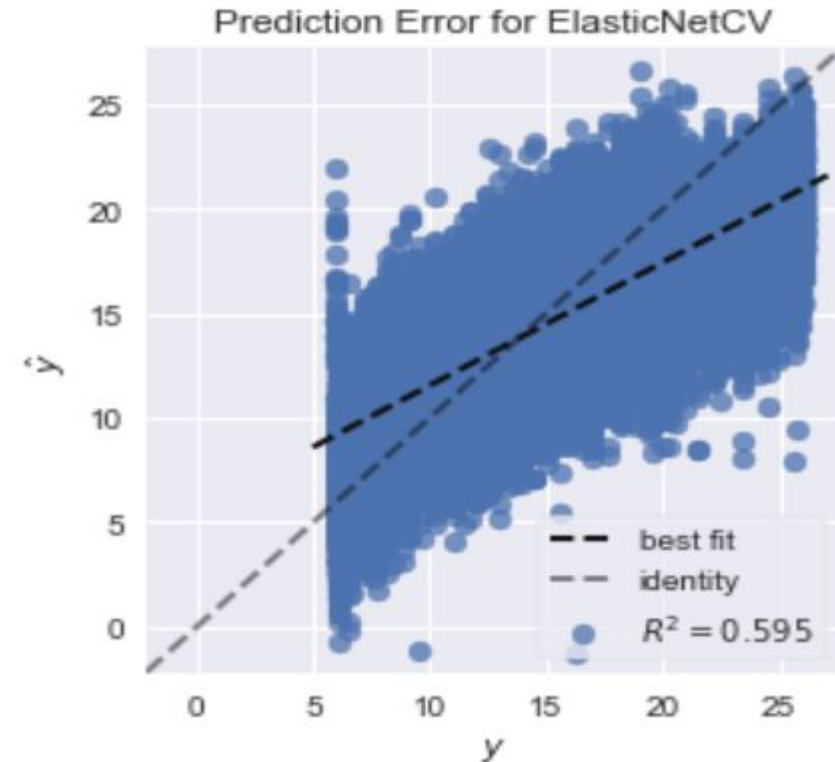
We evaluated data with categorical values to determine if it was significant and relevant to our model. We encoded the categorical values and added the significant data to our final model. This process increased our number of features from 69 to 131.

MODELING AND APPLICATION - Regression

ElasticNetCV after No Scaler



ElasticNetCV after StandardScaler



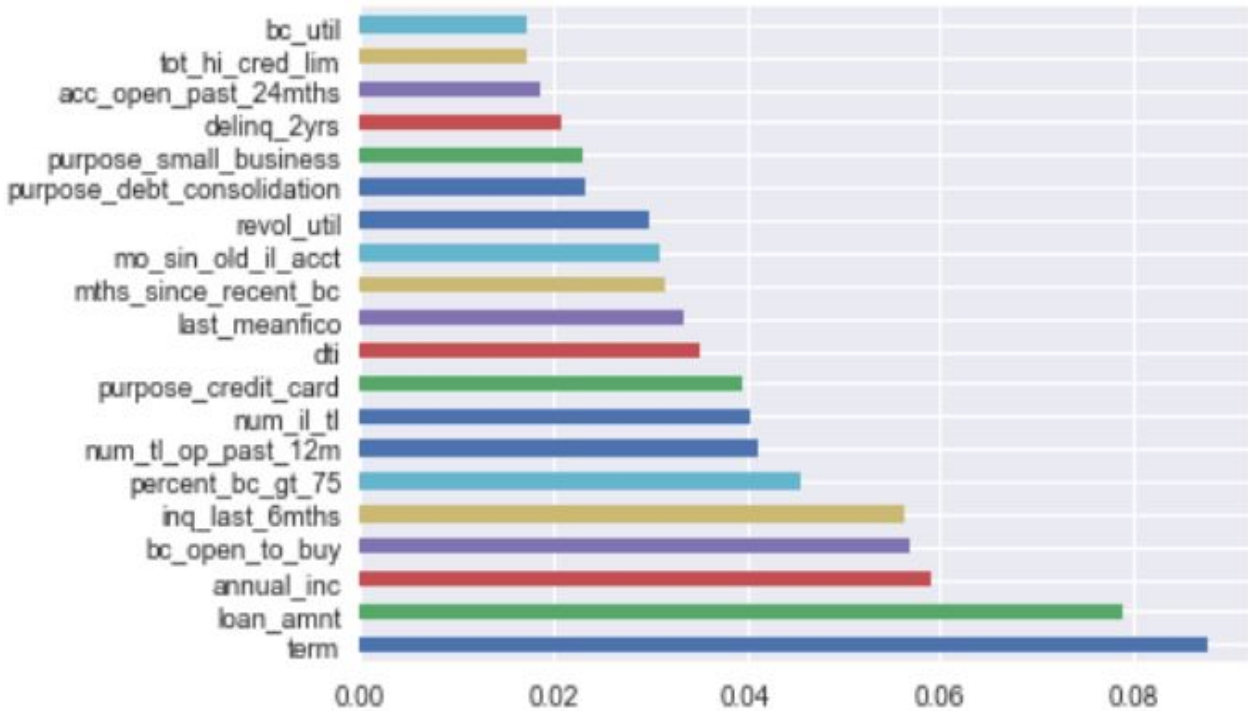
MODELING AND APPLICATION - Regression

MODEL	MEAN SQUARED ERROR	COEFFICIENT OF DETERMINATION
LinearRegression	7.579	0.595
Ridge	7.553	0.595
RidgeCV	7.741	0.586
LassoCV	9.027	0.519
ElasticNetCV	9.843	0.474
RandomForestRegressor	8.346	0.556
GradientBoostingRegressor	7.482	0.601



GRADIENT BOOSTING MODEL

Using the important feature plot for the gradient boosting model, we obtained the 20 most important features.



Testing Models

MODEL	MEAN SQUARED ERROR	COEFFICIENT OF DETERMINATION
GradientBoostingRegressor	7.482	0.601
GradientBoostingRegressor (GridSearch)	6.595	0.649
Secondary Gradient Boosting Model	7.90	0.582

MODELING AND APPLICATION - Classification

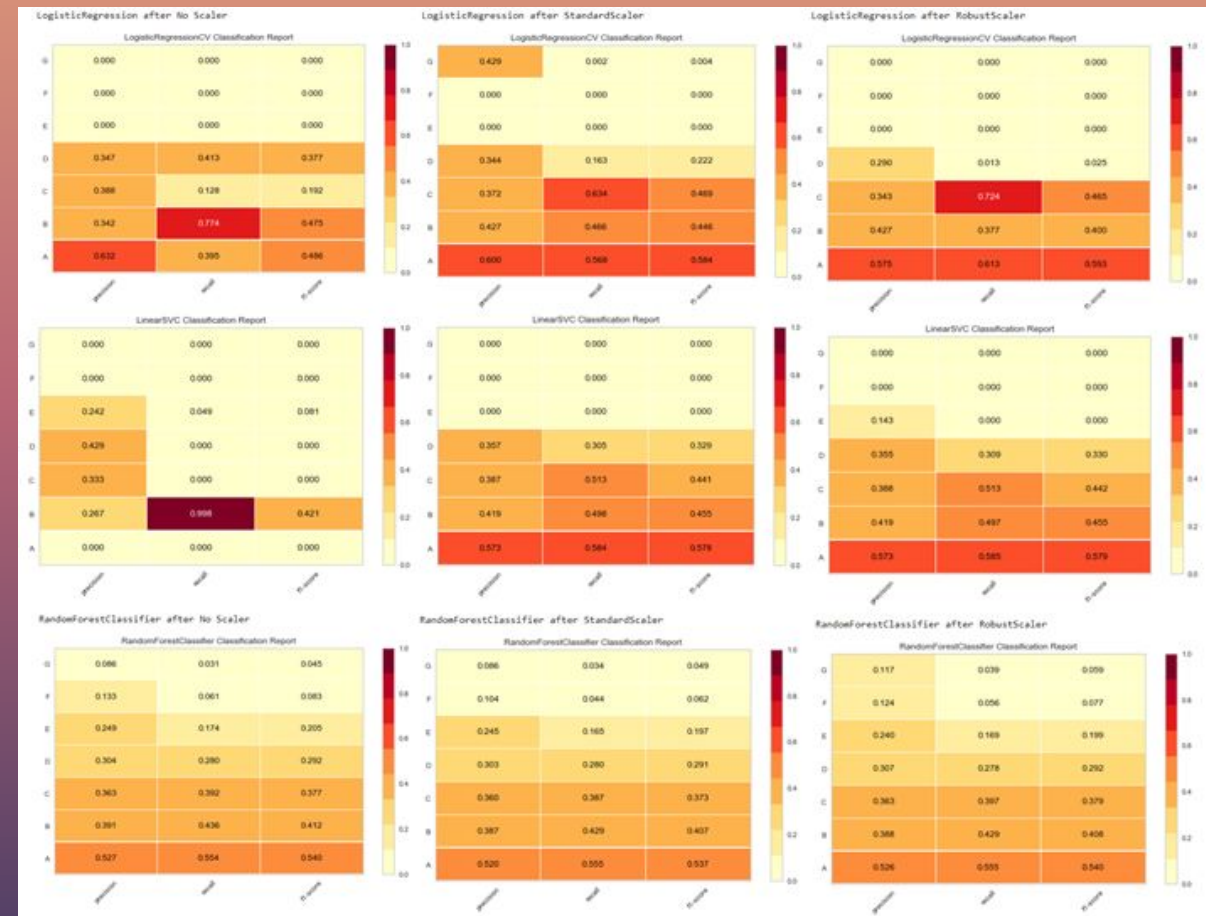
LENDING CLUB LOAN GRADES



Lower interest payments
Lower expected loan losses
 (fewer charge offs)
Lower expected returns
Lower expected volatility

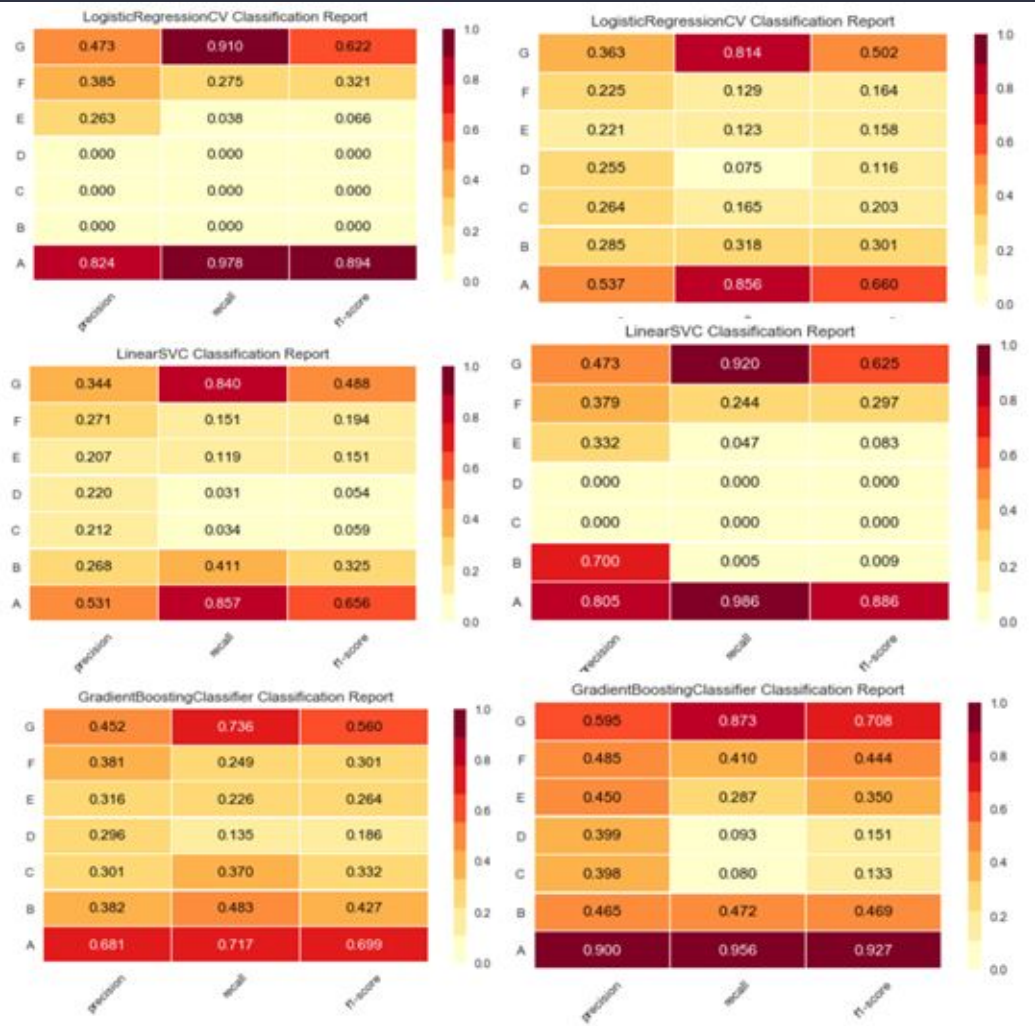
Higher interest payments
Higher expected loan losses
 (more charge offs)
Higher expected returns
Higher expected volatility

CLASSIFICATION MODEL



MODELING AND APPLICATION

LENDING CLUB LOAN GRADES



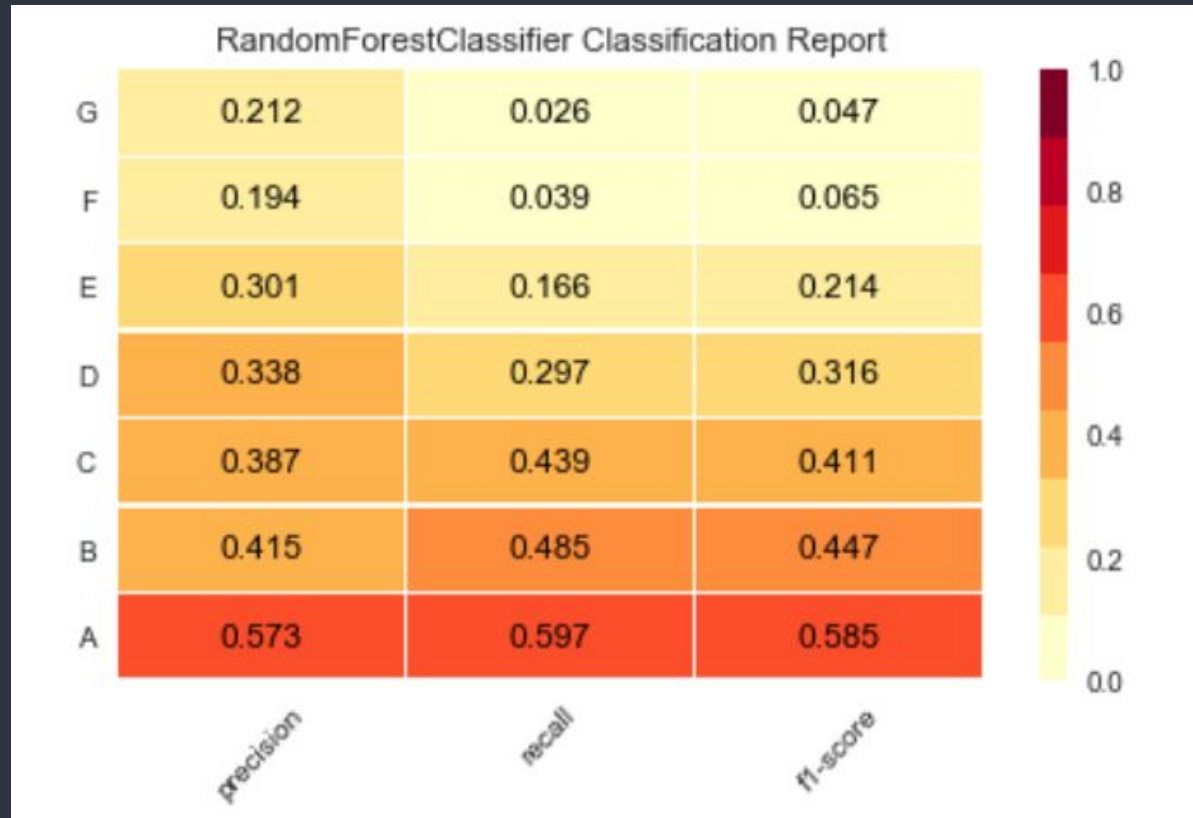
CLASSIFICATION MODEL

MODEL	PRECISION	RECALL	F1
LinearSVC	0.44	0.53	0.44
LogisticRegressionCV	0.43	0.53	0.44
RandomForestClassifier	0.78	0.79	0.78
GradientBoostingClassifier	0.58	0.60	0.57



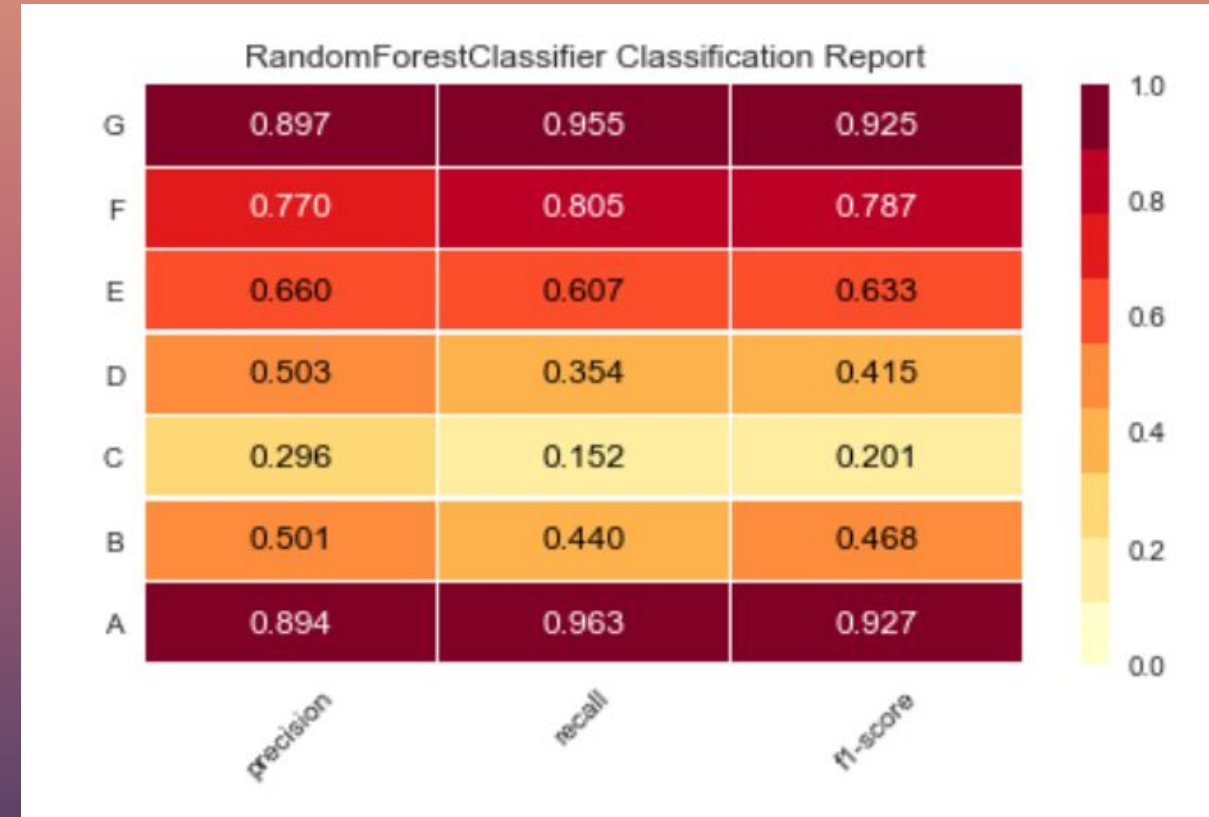
REPORTING AND VISUALIZATION

Before fix to imbalance



F1 score = 0.40

After fix to imbalance with SMOTEENN



F1 score = 0.78





CONCLUSION:

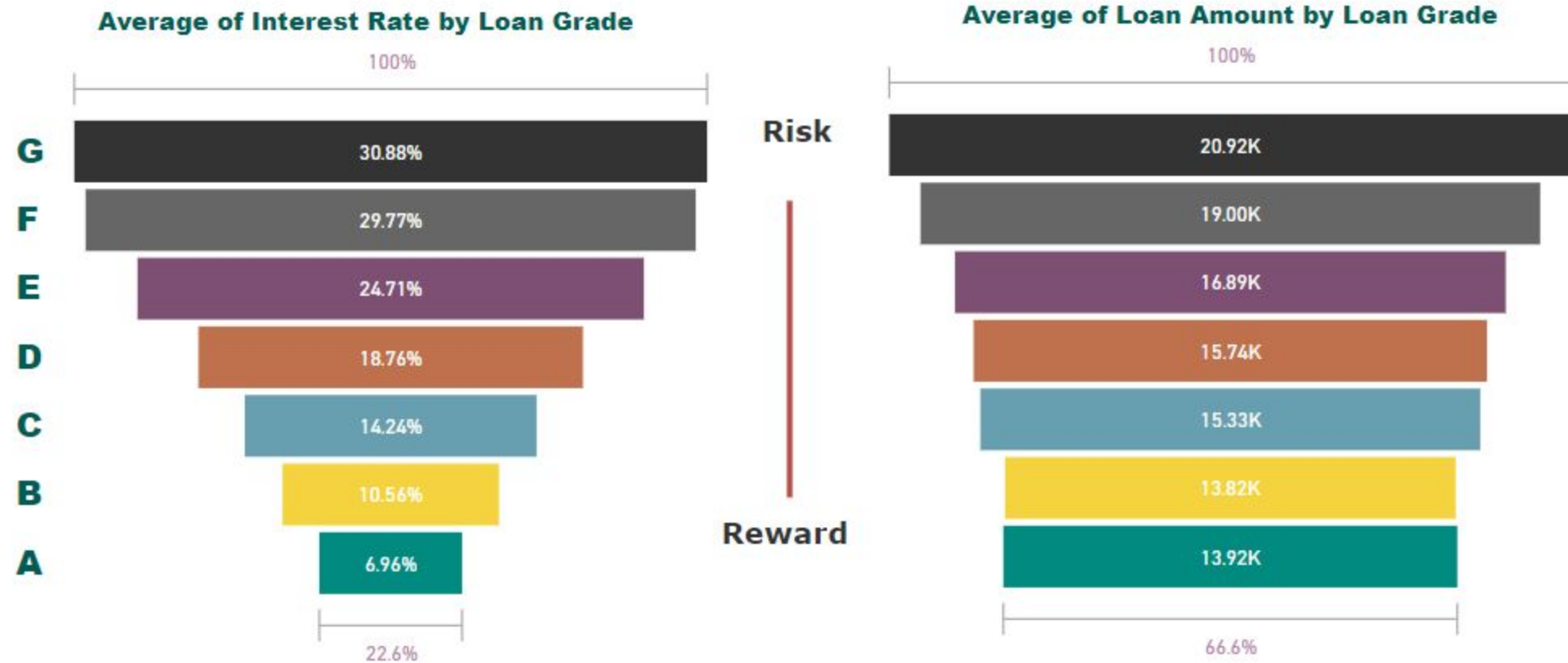
- We determined the key variables of the model are loan amount, term, employment length, home ownership, annual income, loan purpose, mean FICO, revolving line utilization rate, inquiries in the last six months, debt-to-income ratio, and delinquencies in the last two years.
- It is possible to create a simple, reliable model with classification and LendingClub's loan grade scale.

CHALLENGES:

- Inconsistency
- Hardware
- Constant changes of LendingClub's interest rate models



Lending Club Statistics



Interest Rate Predictor

Amount Requested:

Term (Loan Length):

Revolving Line Utilization Rate:

Credit inquiries in the last six months:

Debt to income ratio:

Delinquencies in last two years:

Annual Income:

Purpose:

Rent or Own Home:

Mean FICO:

Length of Employment (Years):

Predicted Interest Rate:

Data Product Prototype

- Utilizes the Flask framework
- Allows the user to input data and receive a predicted interest rate range



THANK YOU