



Bryce Davis
Team Leader
Statistics



Ted Woodsides
Statistics



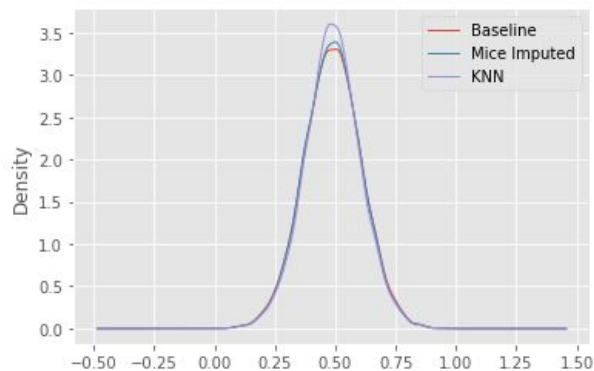
Taylor Last
Statistics



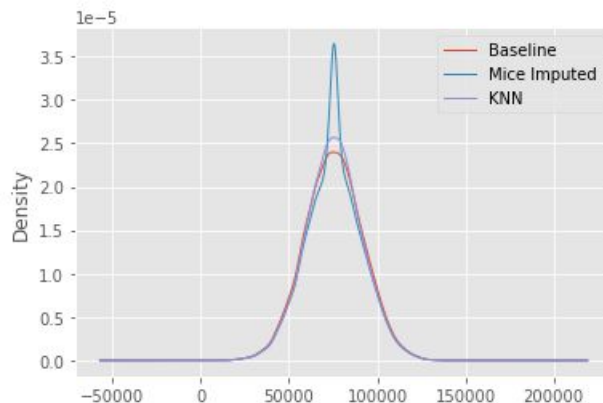
Anderson Molter
Computer Science

Missing Values

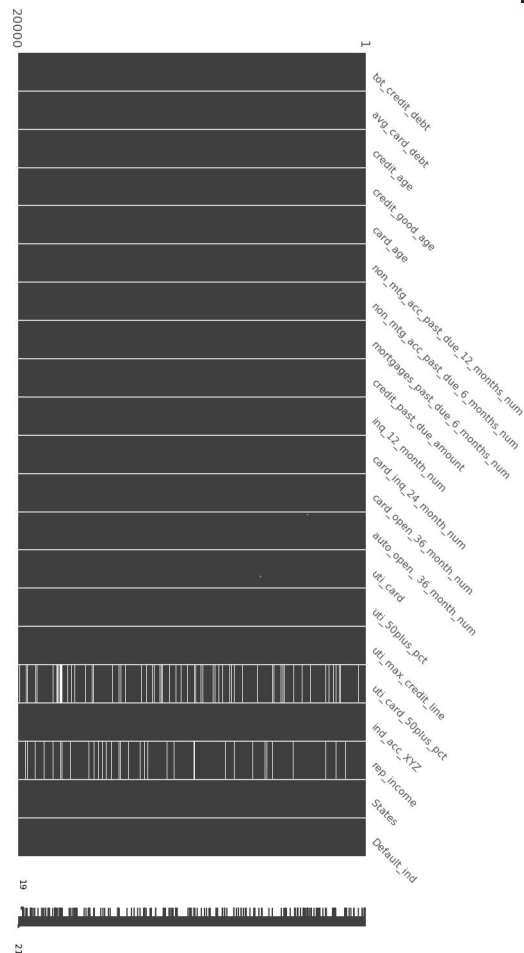
- Nearly complete dataset, other than reported income and percentage of open credit cards with over 50% utilization
- MICE
 - Predictive Mean Matching estimator
 - Card Utilization
- KNN
 - Reported Income



Card Utilization

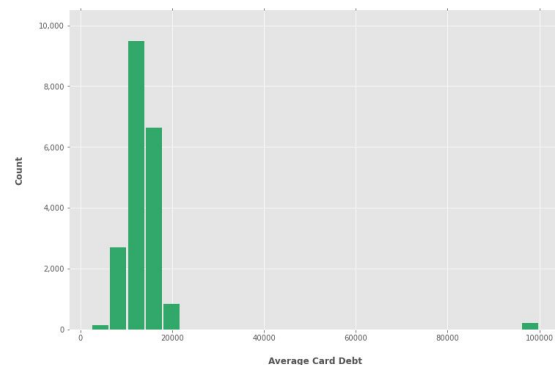
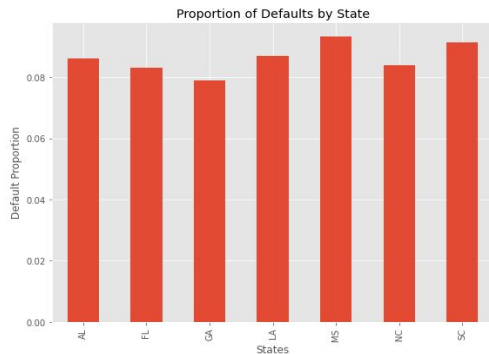
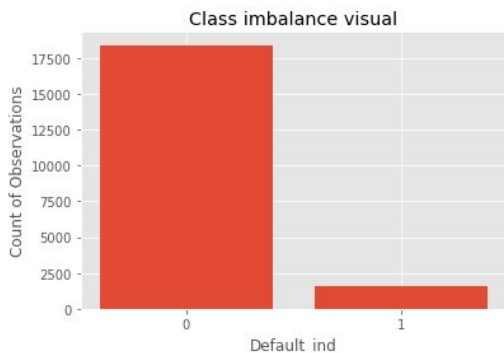
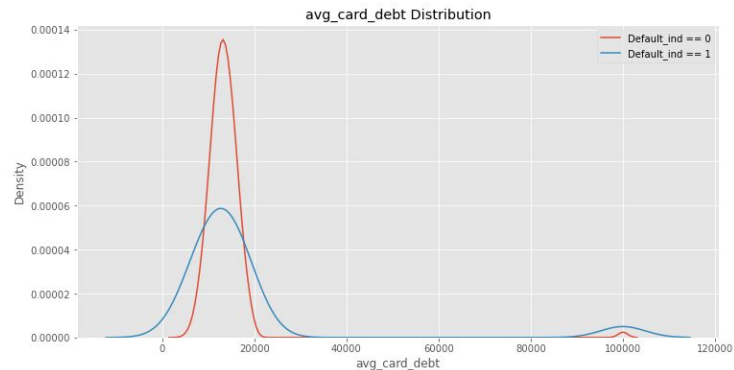


Reported Income



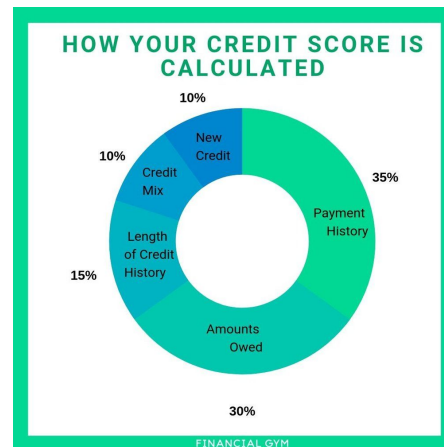
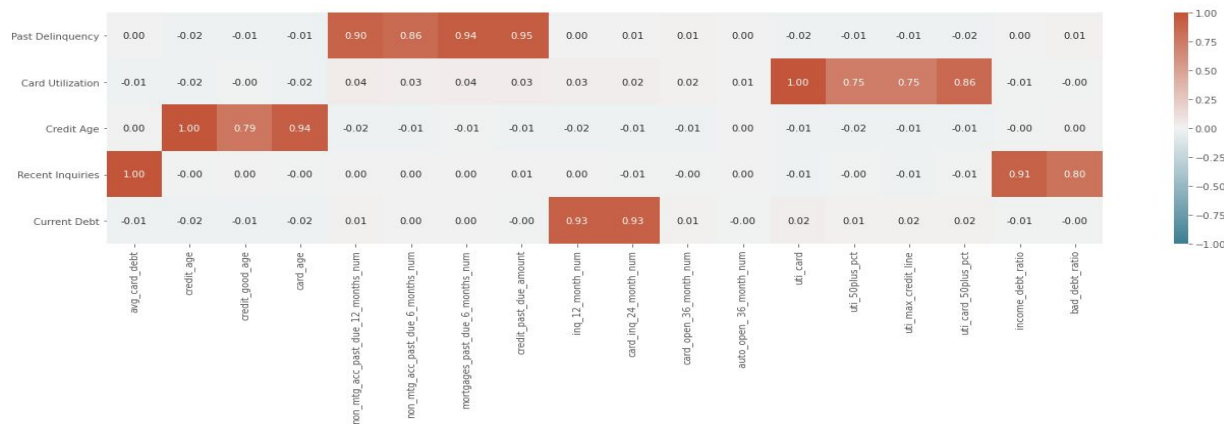
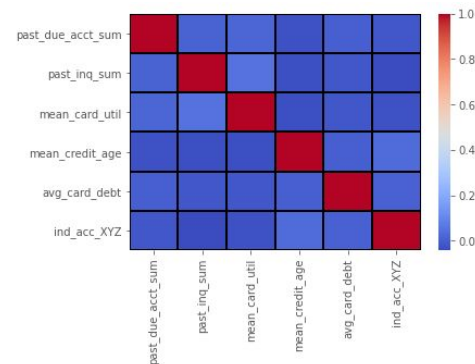
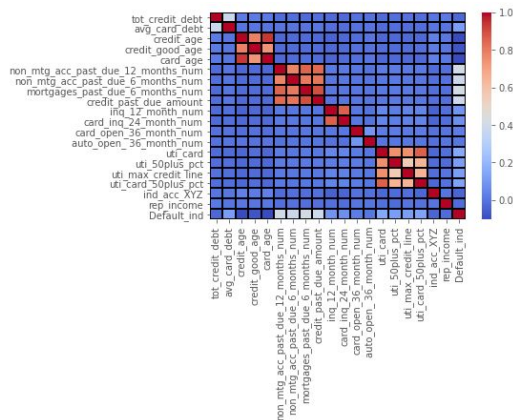
Important EDA

- Large class imbalance
 - 92% did not default
- States not relevant
 - Between state variance = .00002
- Influential outliers, but meaningful
 - 45% of 99999's default



Variable Manipulation

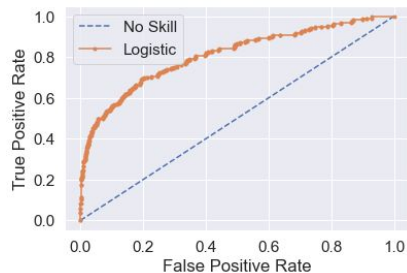
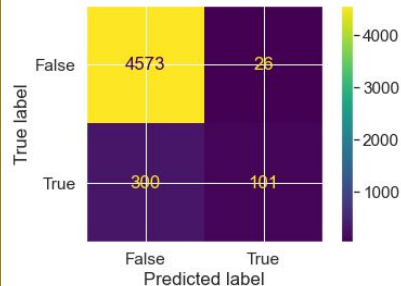
- New Variables
 - Bad_debt_ratio
 - Income_debt_ratio
- Factor Analysis
- FICO Score connection



Logistic Regression

- Manually tuned decision boundary
- Easy to interpret factored coefficients
- **Overall Accuracy = 93.5%**
- **AUCROC = 0.81**
- **F1-Score = 0.48**

	Coef	log_coef	effect
Past Delinquency	0.746302	2.109186	110.918635
Card Utilization	0.718956	2.052289	105.228911
Credit Age	-0.425868	0.653203	-34.679744
Debt	0.286103	1.331230	33.122966
Recent Inquiries	0.348755	1.417302	41.730205

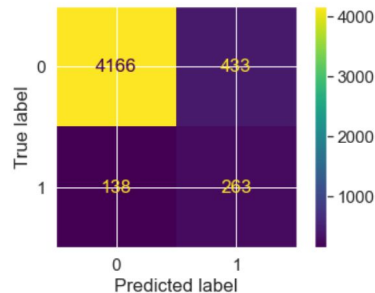
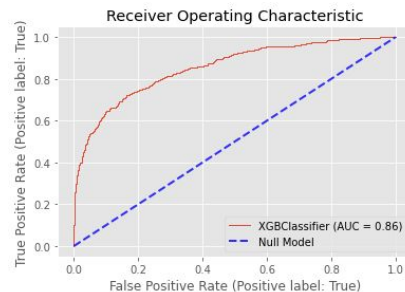


Gradient Boosting

- More resilient to class-imbalance
 - Random Forest, Neural Net
- Extensive list of hyper parameters
- Bayesian Hyperparameter Optimization
- “Black Box”
- **Overall Accuracy = 88.5%**
- **AUCROC = 0.86**
- **F1-Score = 0.89**

base_score	0.5126
gamma	6.1425
learning_rate	0.0408
max_depth	35.3378

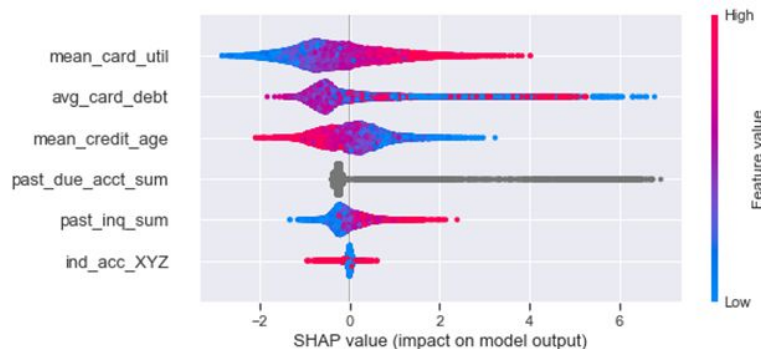
min_child_weight	8.5961
reg_alpha	0.0778
reg_lambda	0.0847
scale_pos_weight	21.411



Do existing Bank Customers have advantage?

How to explain results to a defaulted customer?

SHAP Values



LIME Explanation

Prediction probabilities

Not Default 0.01
Default 0.99

Not Default

Default

past_due_acct_sum > ... 0.40
mean_card_util > 0.57 0.21
mean_credit_age > 2... 0.11
past_inq_sum > 8.00 0.06
avg_card_debt <= 113... 0.04
ind_acc_XYZ <= 0.00 0.02





Thank you!