



CLASSIFICATION OF **STAR** CLIENTS

By *Amir Khoelilar*



INTRODUCTION

REVENUE OVERVIEW

- Website that connects people who want quick loans to lenders
- GetMoney.com make a profit from each successful loan approved lead.
- Lead = Person who opens an account and submits a loan.



INTRODUCTION

STAR & REGULAR LEADS

- Depending of the revenue made, the leads are classified as “Star” and “Regular”
- Star Client = Leads that make more than \$3
- Regular Client

PROBLEM: What are the feature & characteristics of STAR clients ?



INTRODUCTION

SOLUTION

- Classification models in order to recognize the best features that describe Star Leads

IMPACT & SUCCESS: Recognizing the key features of the Star leads of the website in order to

- Optimize marketing process
- Reduce funding waste



DATA & TOOLS

OVERVIEW

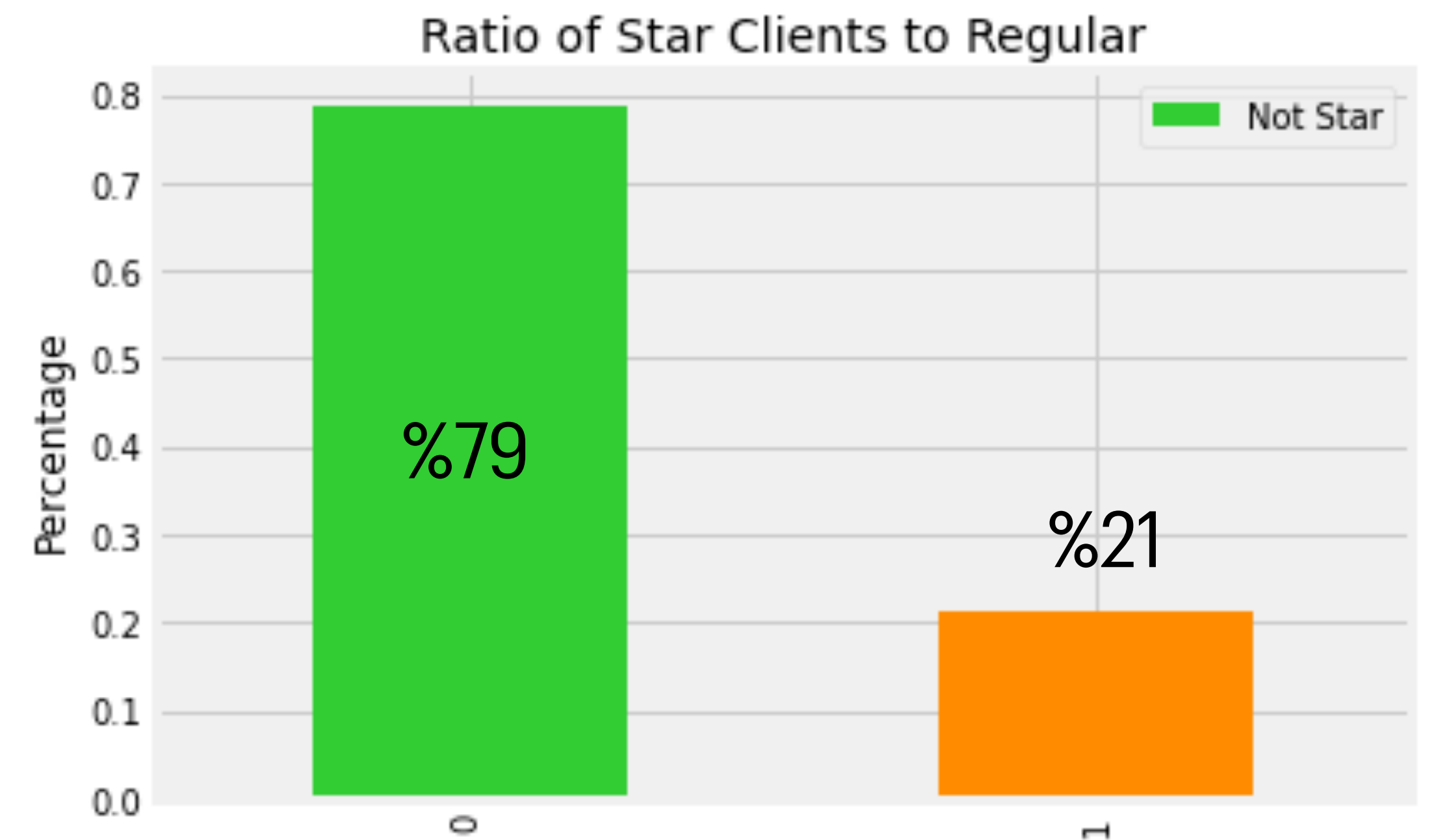
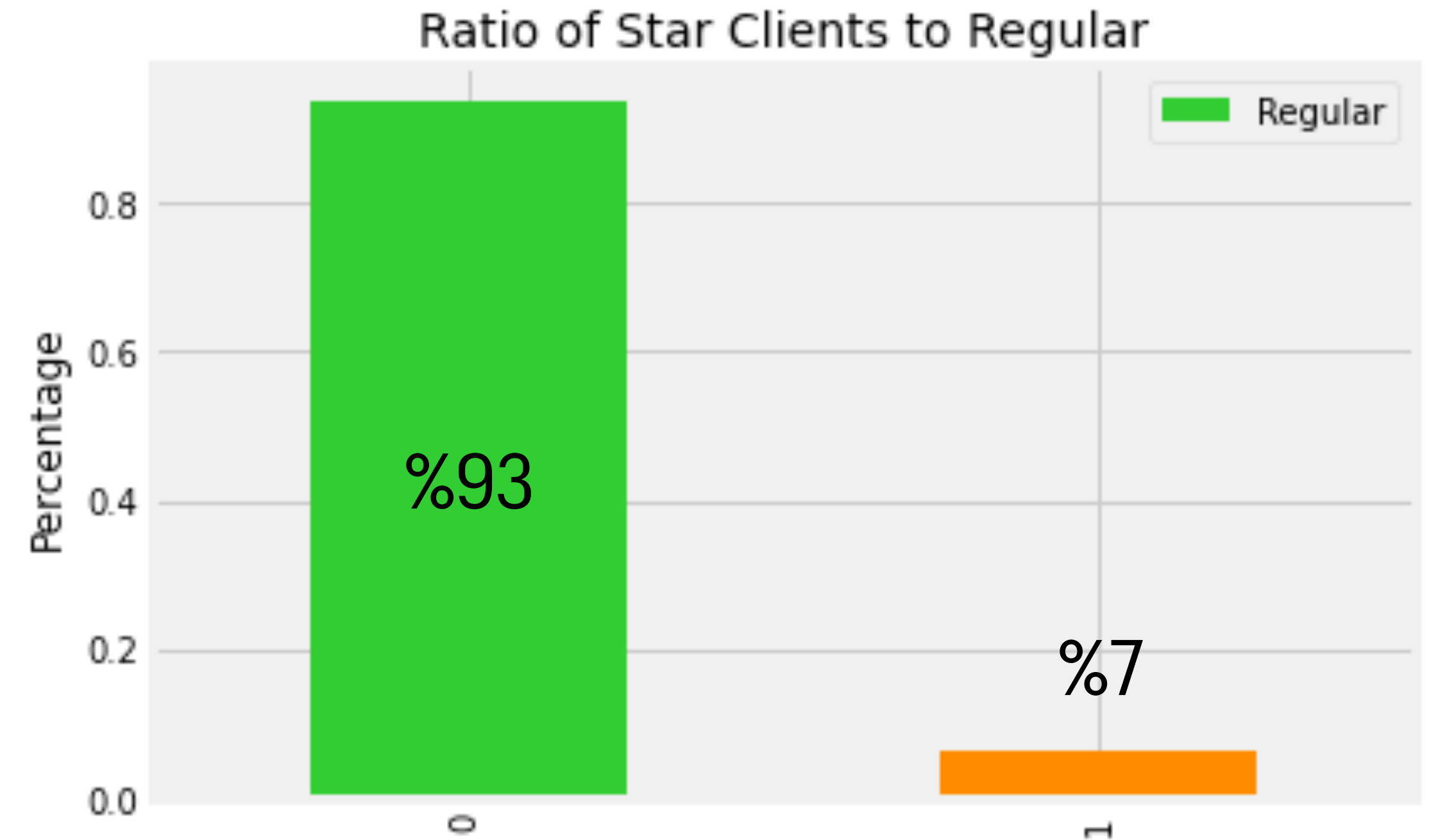
- Data - 6500 Rows & 15 Columns
- EDA
 - Google Sheets - Tableau Public
 - Pandas - Numpy
 - Tableau - Matplotlib - Seaborn
- Modeling
 - Scikit-learn



METHODOLOGY

DATA

- Data - 6500 Rows & 15 Columns
- Imbalanced data
 - Regular Clients 6105
 - Star Clients 405
- Oversampling (Train Data)
 - Regular Clients 3671
 - Star Clients 976
- **BEST METRIC:**
 - Model Comparison : ROC-AUC Curve
 - F1 Score - Focus on a higher score on Recall



METHODOLOGY

ROC AUC CURVE

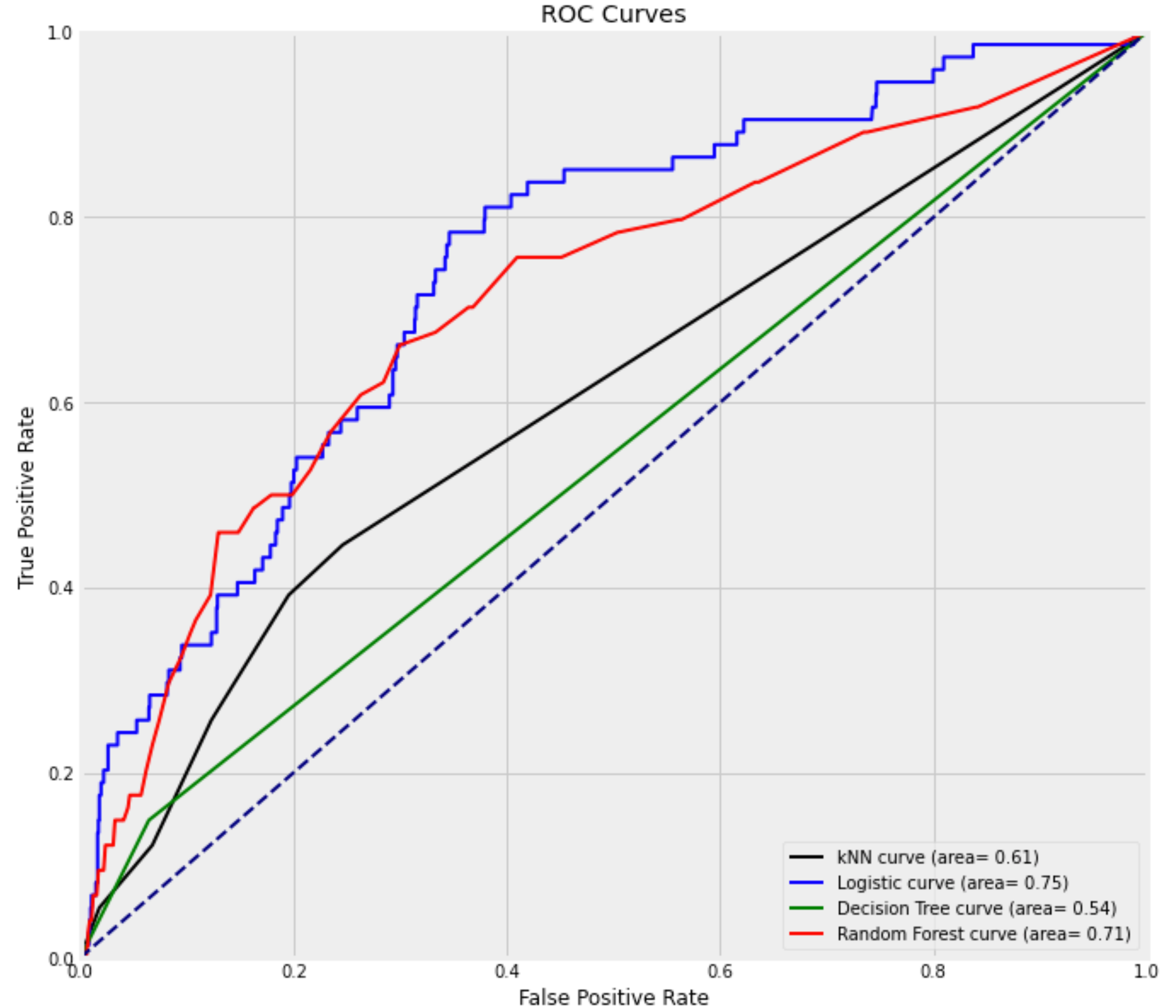
- Models

- 1. Logistic Regression (Area = .75)**

2. Random Forest (Area = .71)

3. KNN (Area = .61)

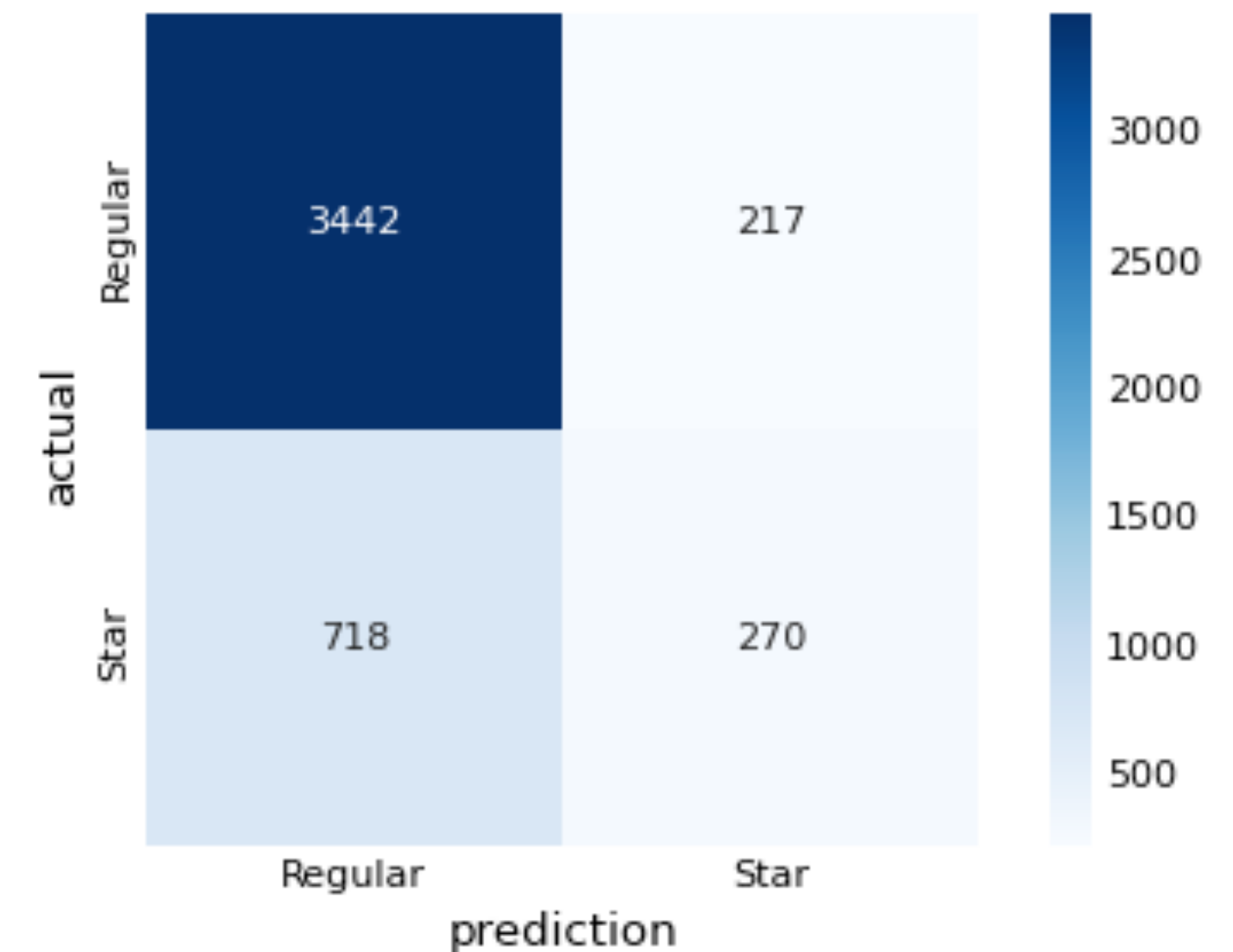
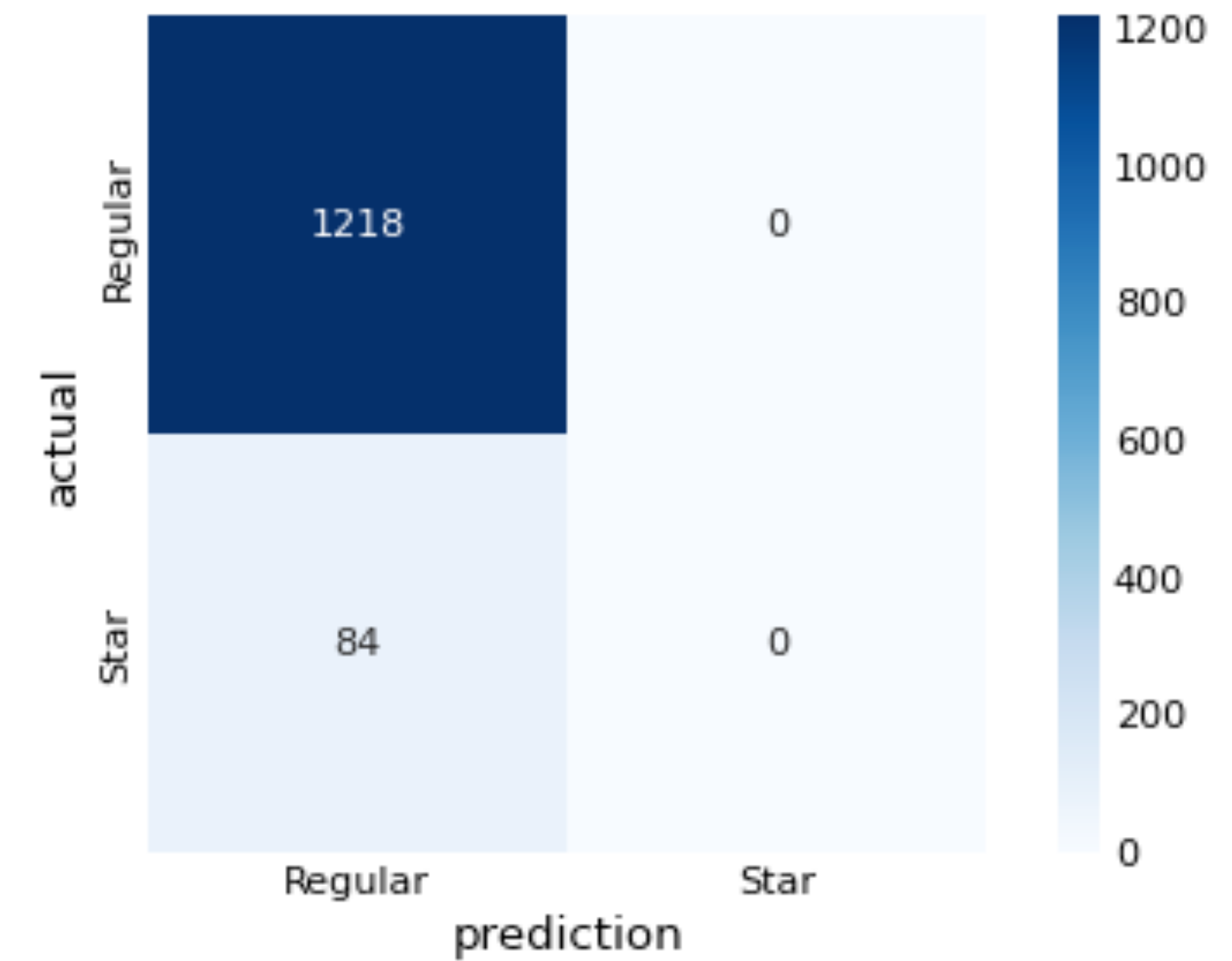
4. Decision Tree (Area = .54)



RESULTS

BASELINE & OVERSAMPLED LR

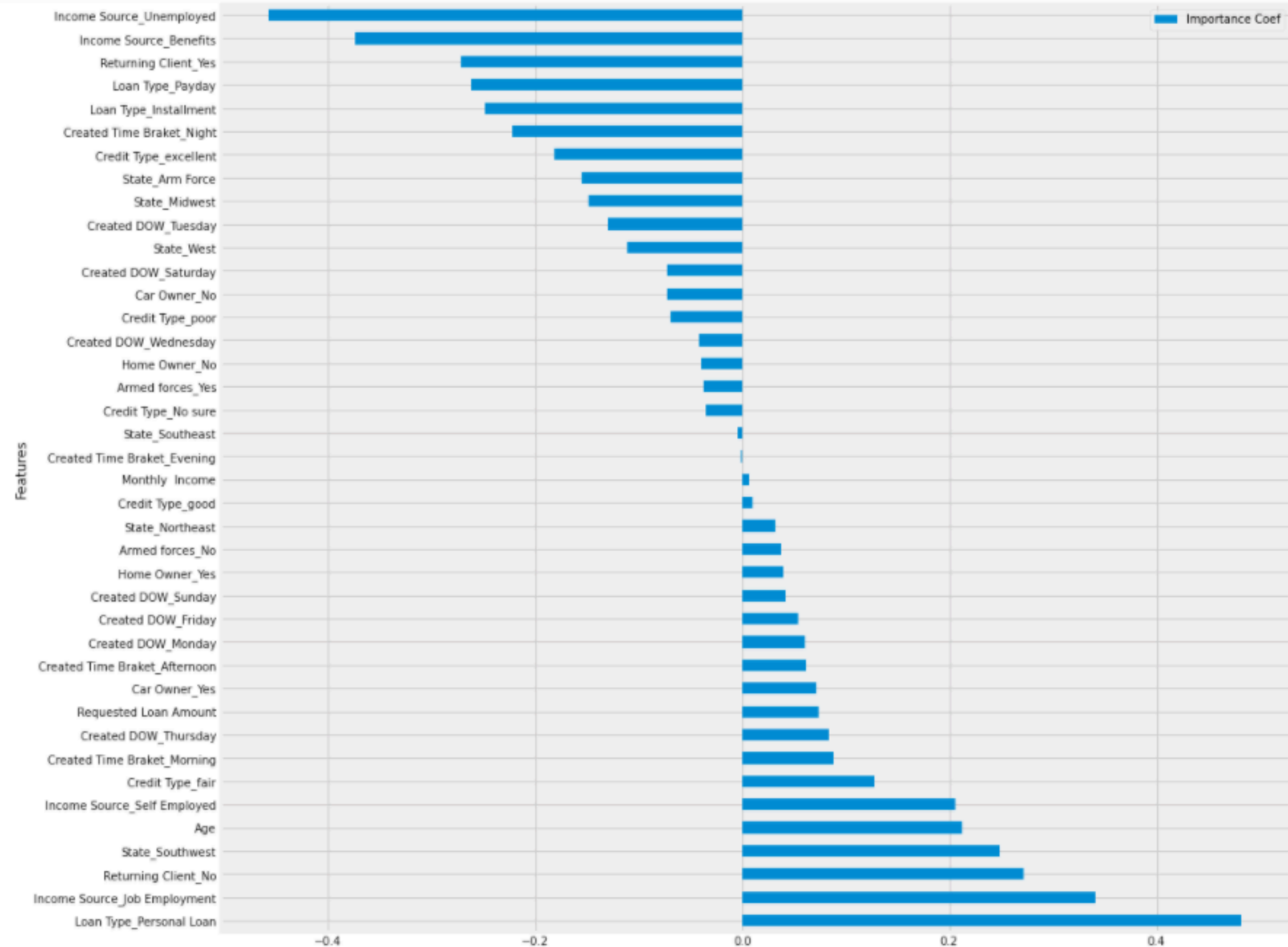
- Baseline Logistic Regression
 - F1 Score: 0.00 (Precision & Recall = 0.00)
 - ROC: 0.6456
- Logistic Regression (Validation)
 - F1 Score: .2466
 - ROC: 0.7455
- Logistic Regression (Test) - [Threshold = 0.14]
 - **F1 Score: 0.2682**
 - Precision: 0.1977
 - Recall: 0.4167



RESULTS

IMPORTANT FEATURES PLOT

- Important Features
 - Positive Coefficient Features
 - Negative Coefficient Features



RESULTS

IMPORTANT FEATURES

- Top Important Features
 - Positive Coef Features
 1. LoanType - Personal Loan
 2. Income Source - Job Employment & Self Employed
 3. State - Southwest [Arizona , New Mexico , Oklahoma , Texas]
 4. Age
 5. Credit Type - Fair
 - Negative Coef Features (NOT)
 6. Income Source - Unemployed
 7. Income Source - Benefits
 8. Credit Type - Excelent
 9. State - Midwest [North Dakota , South Dakota , Nebraska , Kansas , Minnesota , Iowa , Missouri , Illinois , Indiana , Wisconsin , Michigan , Ohio]



CONCLUSION

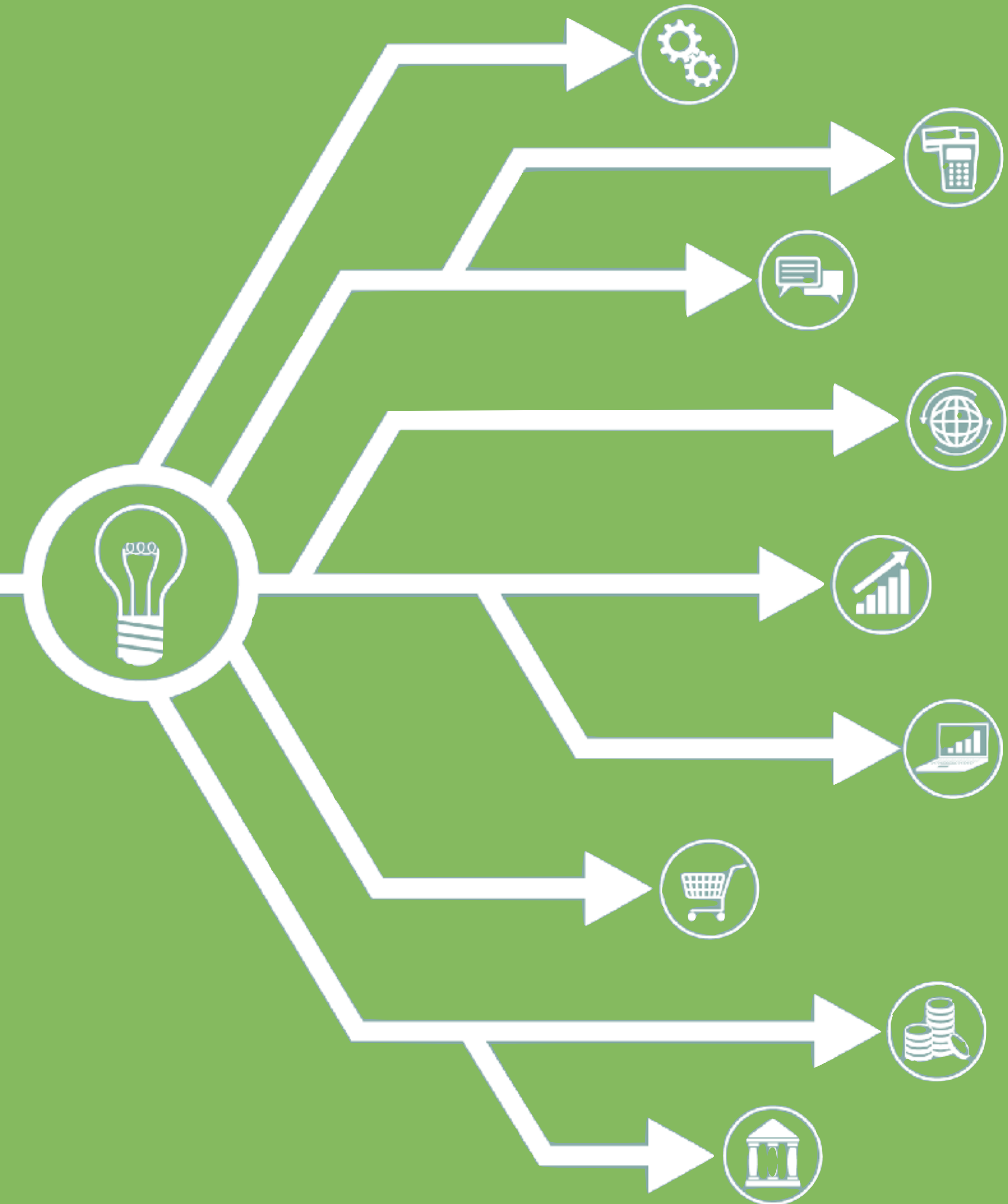
FINAL INTAKES & RECOMMENDATIONS

Recommendations for the GetMoney.com board:

- Star Client Potential features
- By emphasizing on the recommended features during purchasing data
- Collect better **REVENUE** from acquired data



FUTURE WORK



- DATA IMBALANC
 - OVERSAMPLING AGAIN
 - UNDERSAMPLING
 - WEIGHTED SAMPLING
- DIFFERENT MODELS
 - NAIVE BAYES
 - XGBOOST
- FEATURE ENGINEERING
 - CREATED MORE FEATURES

APPENDIX

FINAL INTAKES & RECOMMENDATIONS FROM BUSINESS EDA

- Focus marketing on the people with credit types of **Good & Excellent**
- Focus marketing in regions with **Higher Income** and **Higher Age Bracket** (50's & 60's)



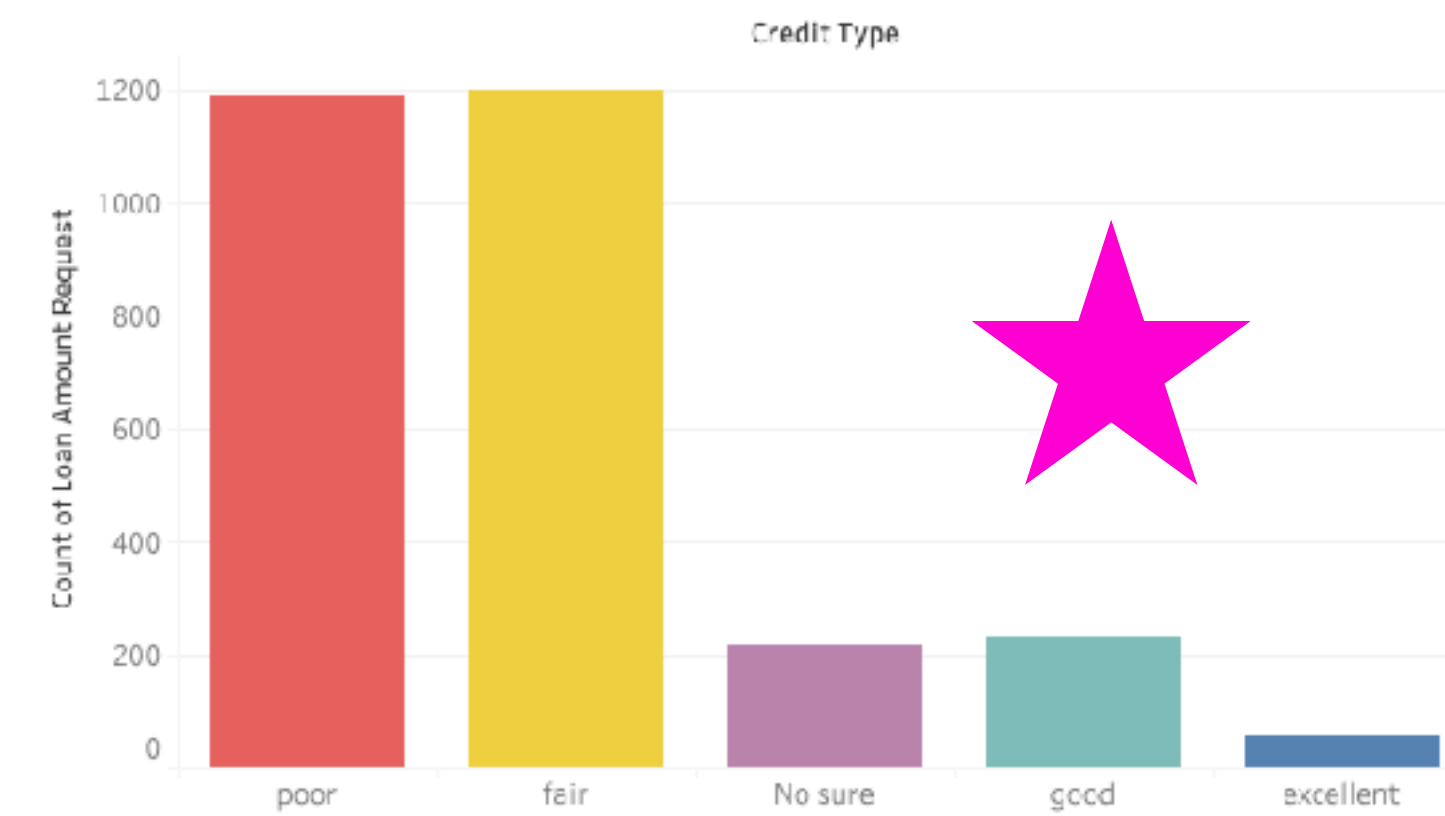
ATTENTION:THE CLASSIFICATION FEATURE IMPORTANCE WOULD
ALMOST RECOMMEND DIFFERENT CHARACTERISTICS AND FEATURES FOR POTENTIAL CLIENTS
FOR OPTIMIZATION IN MARKETING PLANS

APPENDIX

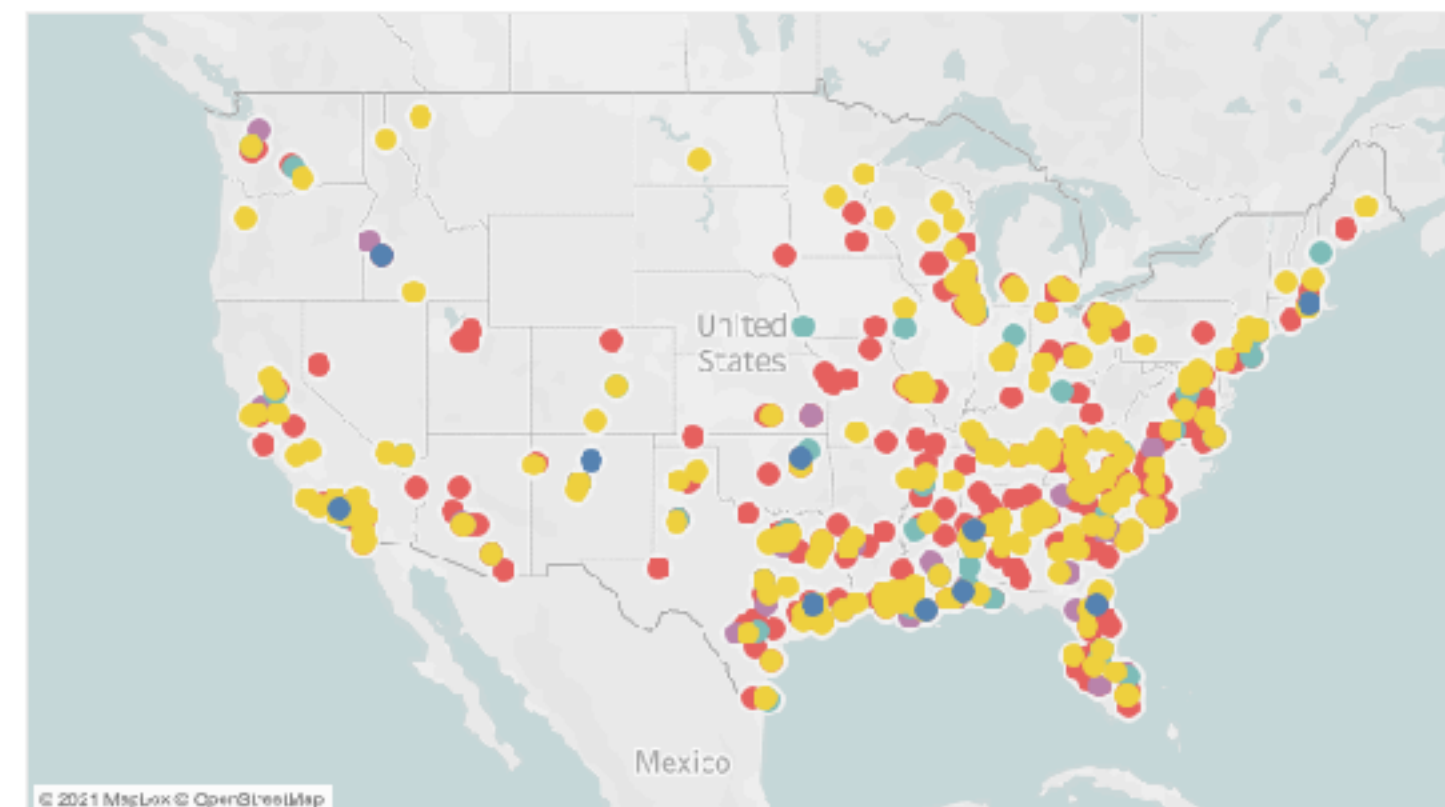
SOLUTION & CHARTS

Leads X City X Credit Type

Number of Leads X Credit Type

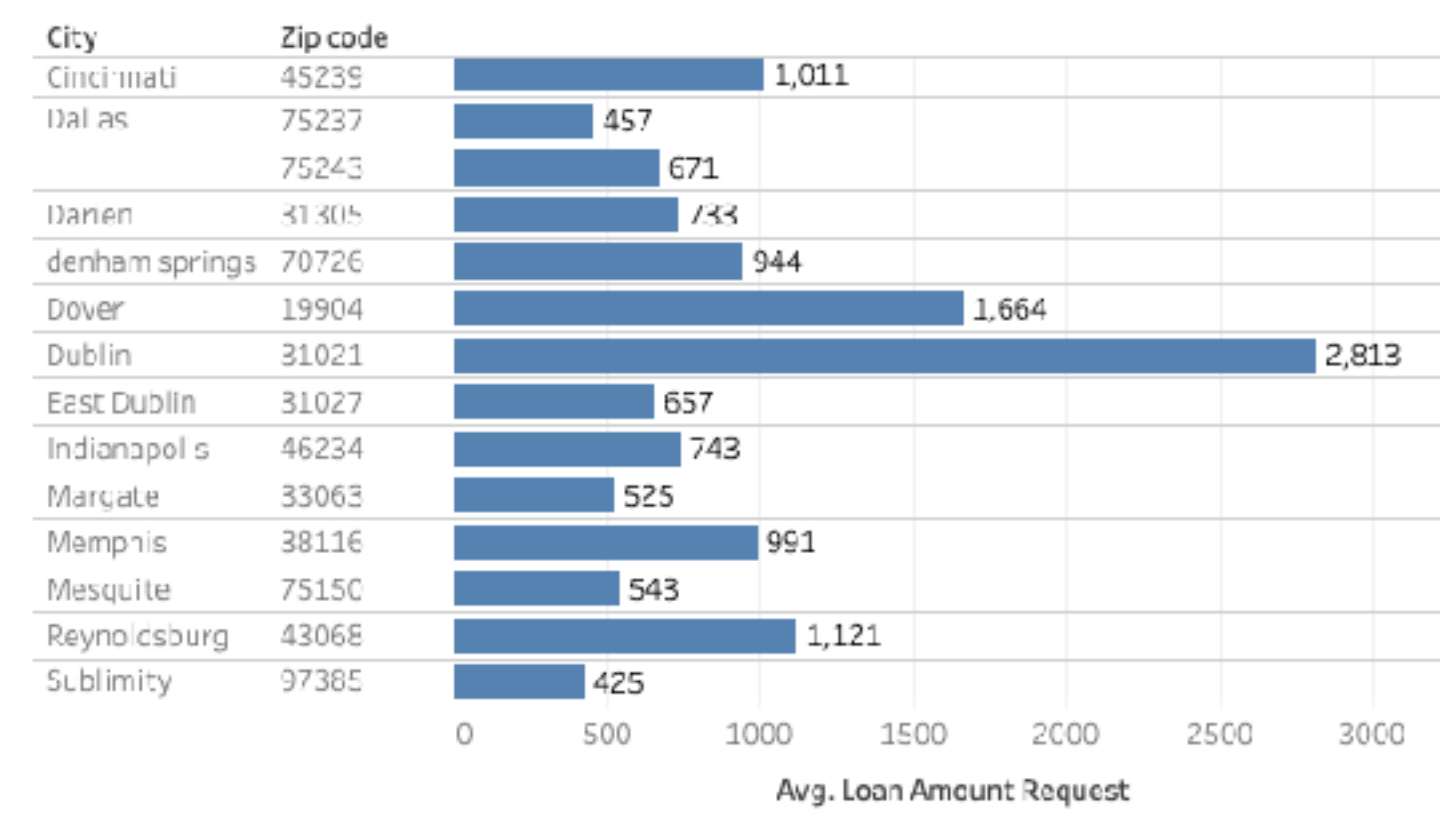


Credit Types X City

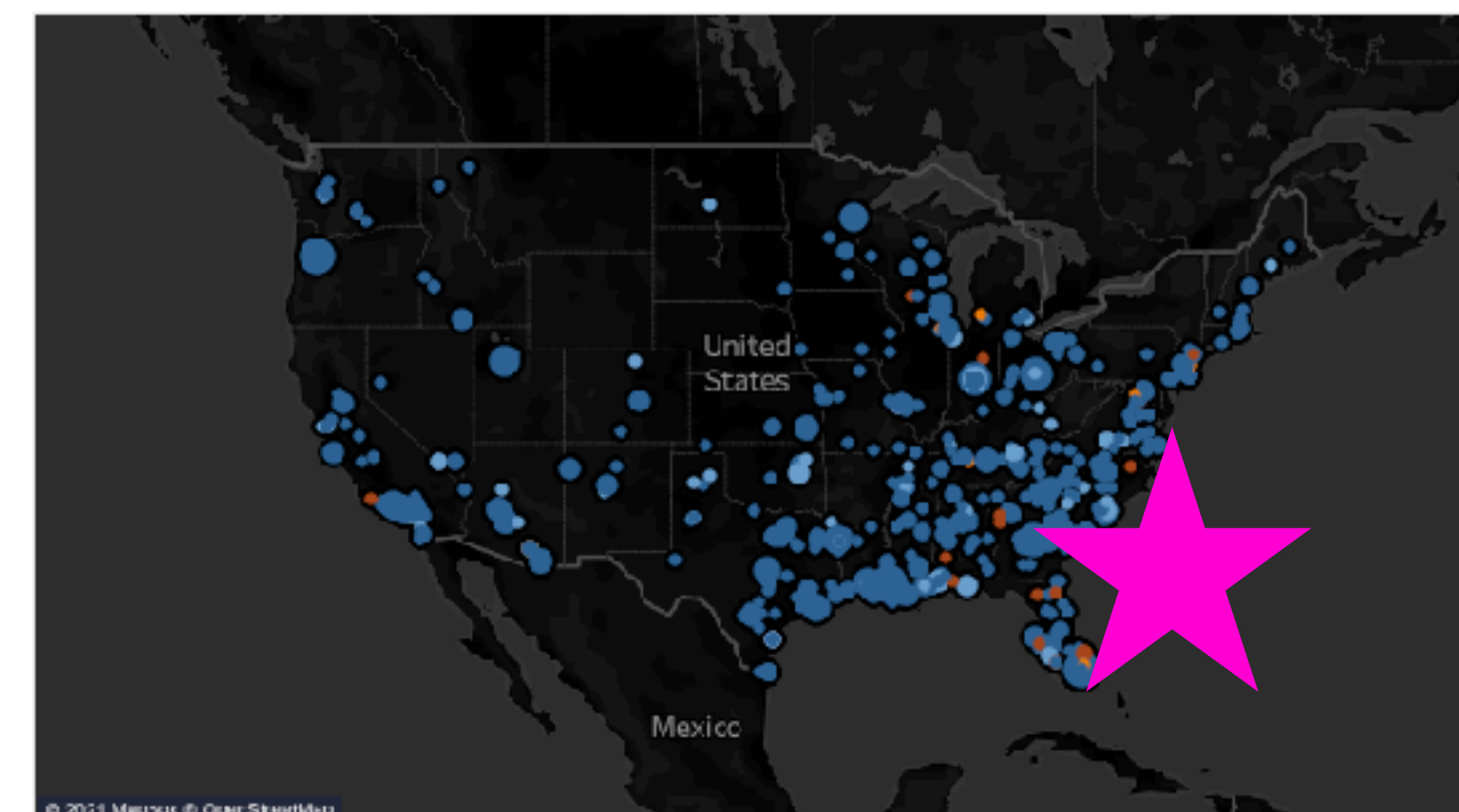


GMDC | City X Zipcode X Leads

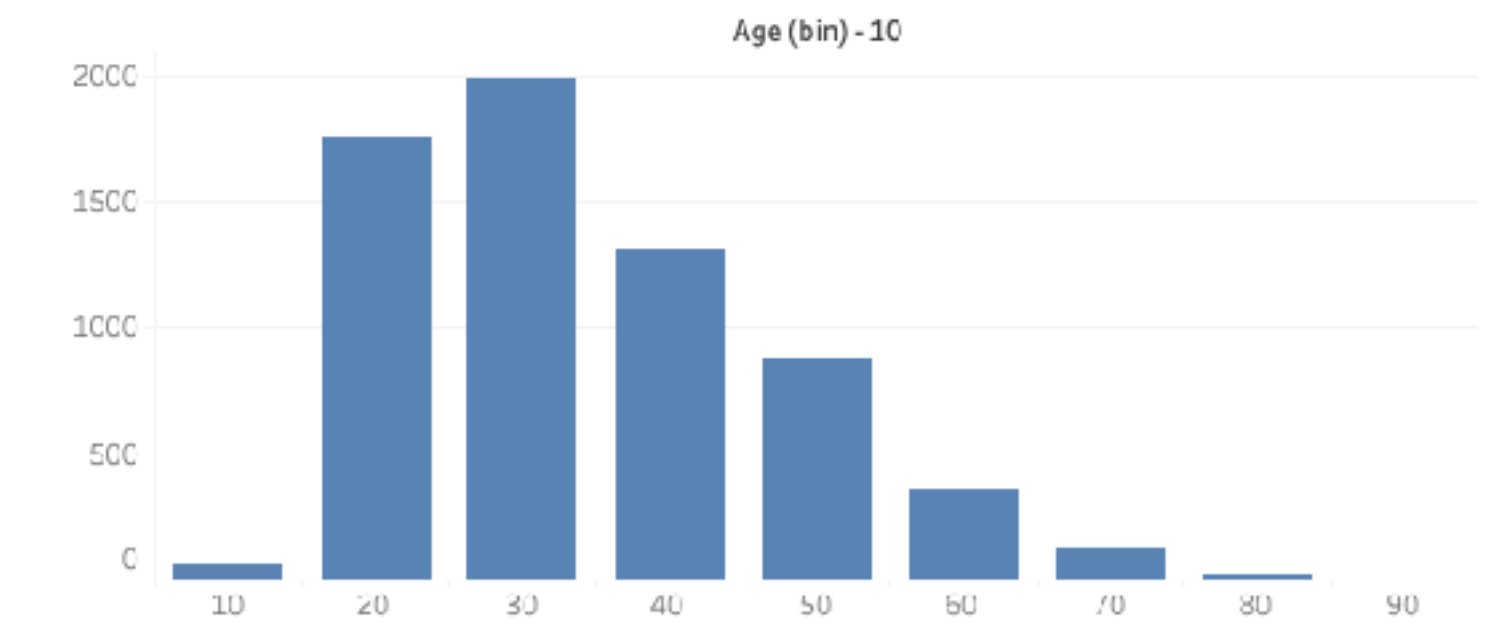
Average Loan Amount Requested X (City - Zipcode) - Bar Chart



Average Loan Requested X (City - Zipcode)

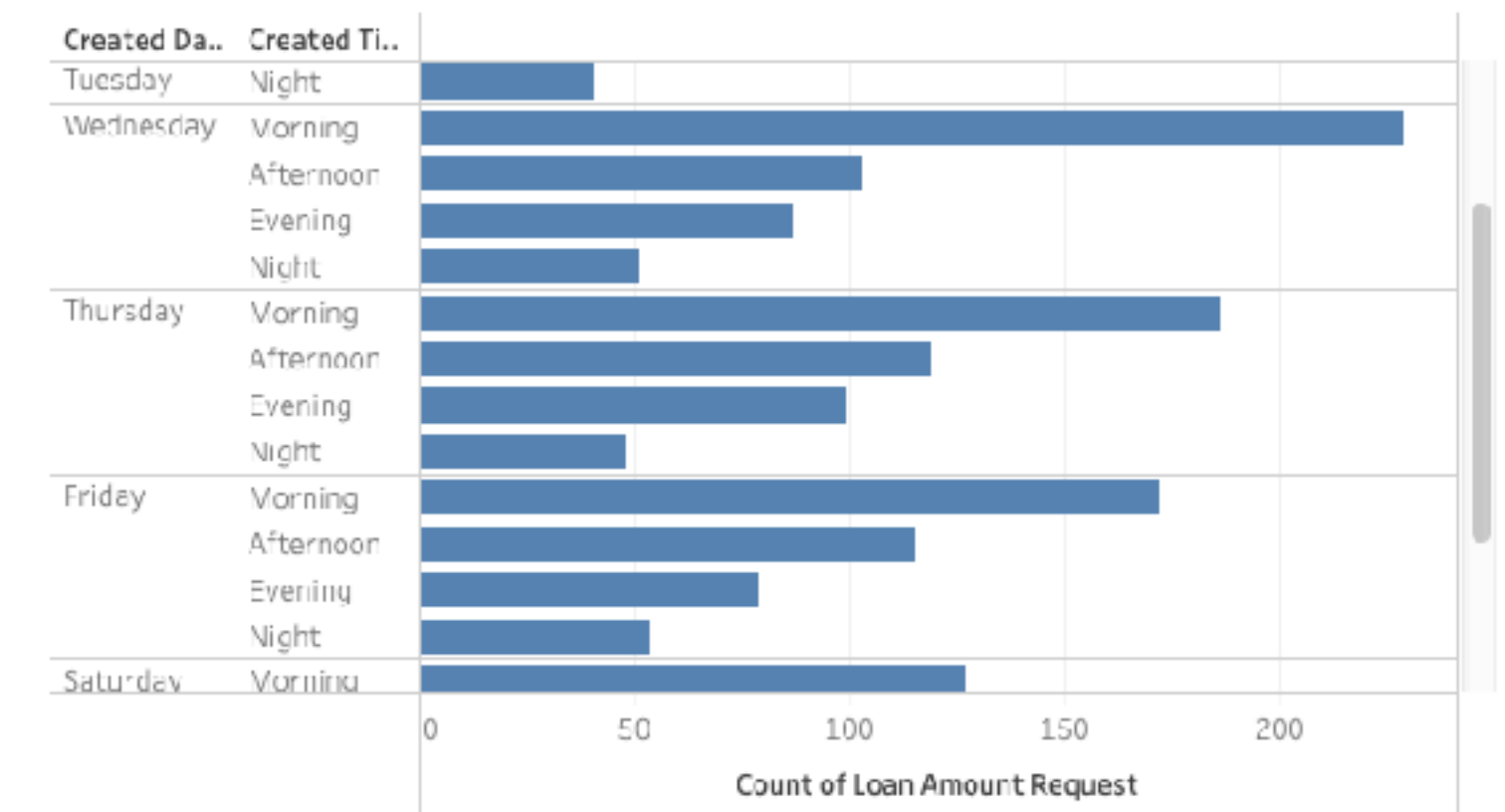


Lead Count X Age



GMDC | Day of Week X Time X Age Brakets X Leads

Website Attraction Day and Time X Leads



Average Loan Amount | Number of Loan Requests X Age

