SC 1015 Mini Project
Z139_Team8

William Johnson U2220859A
Soo Ying Xi U2220021D
Tay Zhi Xian U2220099F

Predicting Credit Card Approvals

01 Motivation

02 EDA

03 Decision-Tree -> Random-Forest + CV
Importance of predictors
Logistic Regression & Random-Forest + CV

04 Conclusion

Significance of Credit Cards

01

- Role of cash is decreasing
- Important to accurately assess applicants

Motivation

- Credit card companies can provide better customer services
- Risk assessment of high-risk applicants
- Help to automate approval process – increase efficiency

DataSet Overview

	Gender	Age	Debt	Married	BankCustomer	EducationLevel	Ethnicity	YearsEmployed	PriorDefault	Employed	CreditScore	DriversLicense	Citizen	Income	Approved	Approved_Status
0	b	30.83	0.000	u	g	w	v	1.250	t	t	1	f	g	0	+	Approved
1	a	58.67	4.460	u	g	q	h	3.040	t	t	6	f	g	560	+	Approved
2	a	24.50	0.500	u	g	q	h	1.500	t	f	0	f	g	824	+	Approved
3	b	27.83	1.540	u	g	w	v	3.750	t	t	5	t	g	3	+	Approved
4	b	20.17	5.625	u	g	w	v	1.710	t	f	0	f	s	0	+	Approved
5	b	32.08	4.000	u	g	m	v	2.500	t	f	0	t	g	0	+	Approved
6	b	33.17	1.040	u	g	r	h	6.500	t	f	0	t	g	31285	+	Approved
7	a	22.92	11.585	u	g	cc	v	0.040	t	f	0	f	g	1349	+	Approved
8	b	54.42	0.500	y	p	k	h	3.960	t	f	0	f	g	314	+	Approved
9	b	42.50	4.915	y	p	w	v	3.165	t	f	0	t	g	1442	+	Approved

Data Cleaning

Handling with:

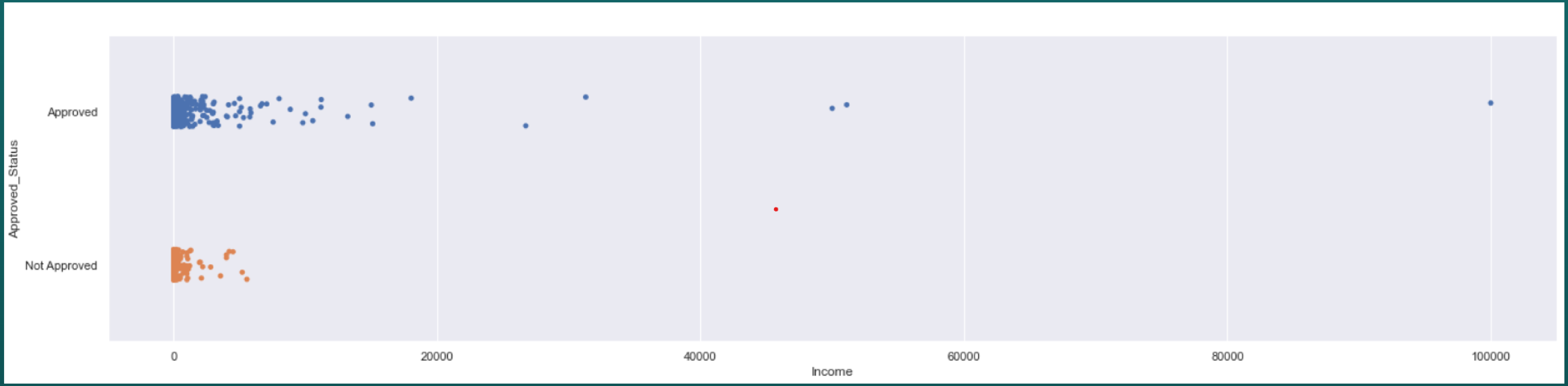
- NULL values – 1.7%
- ambiguous features – due to confidentiality (Education level & Ethnicity)

02

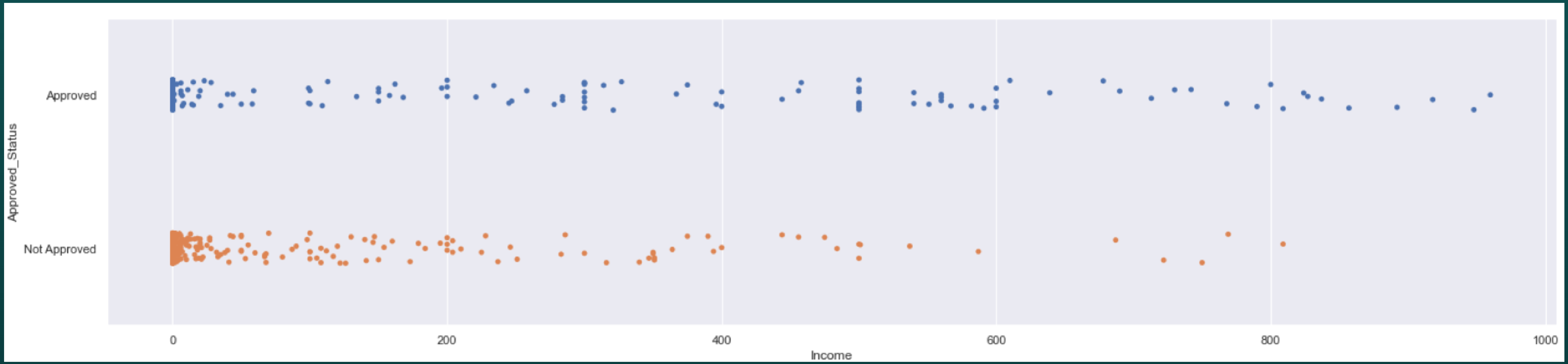
Exploratory Analysis

Numerical Data

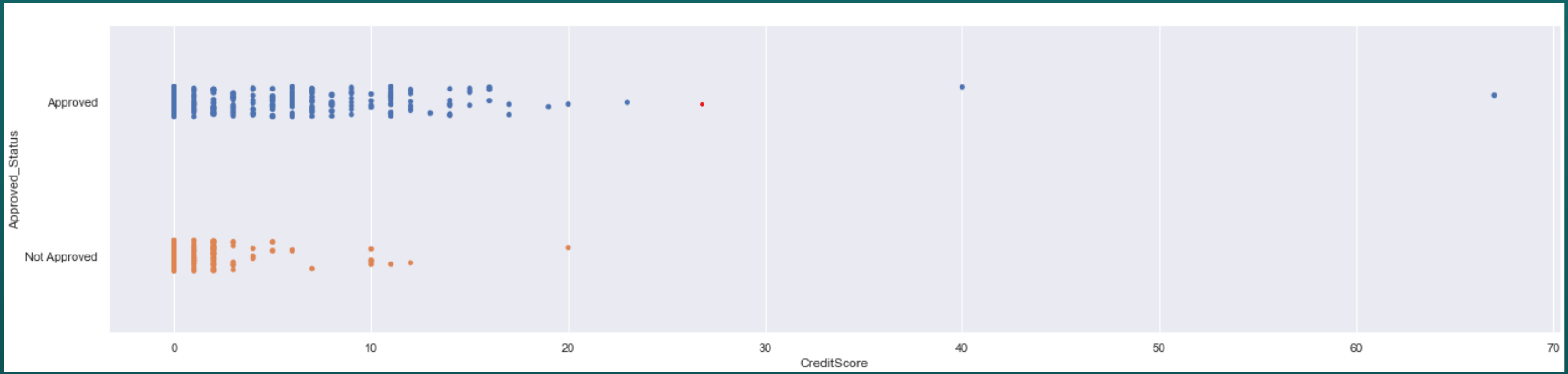
Income With Outliers



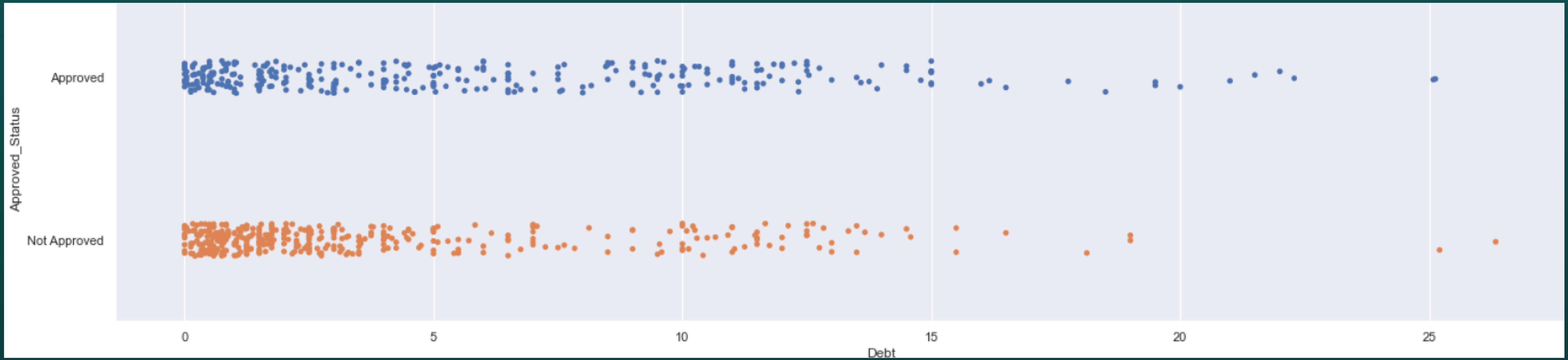
Income



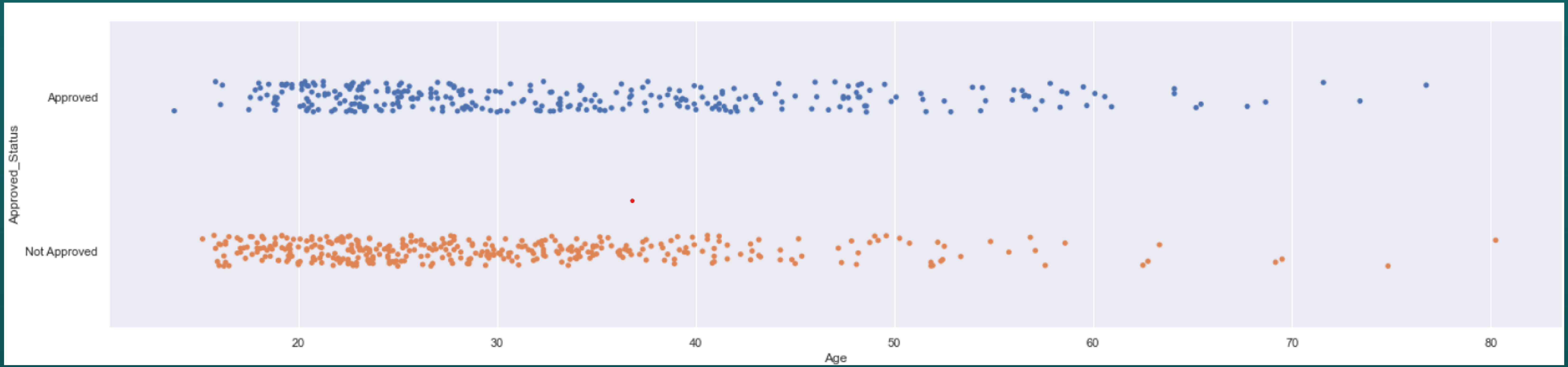
Credit Score



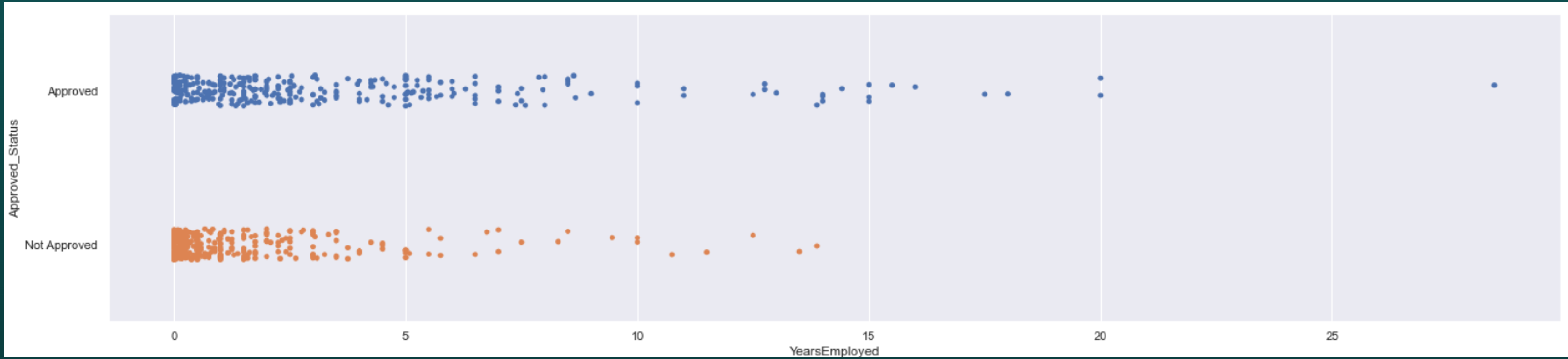
Debt



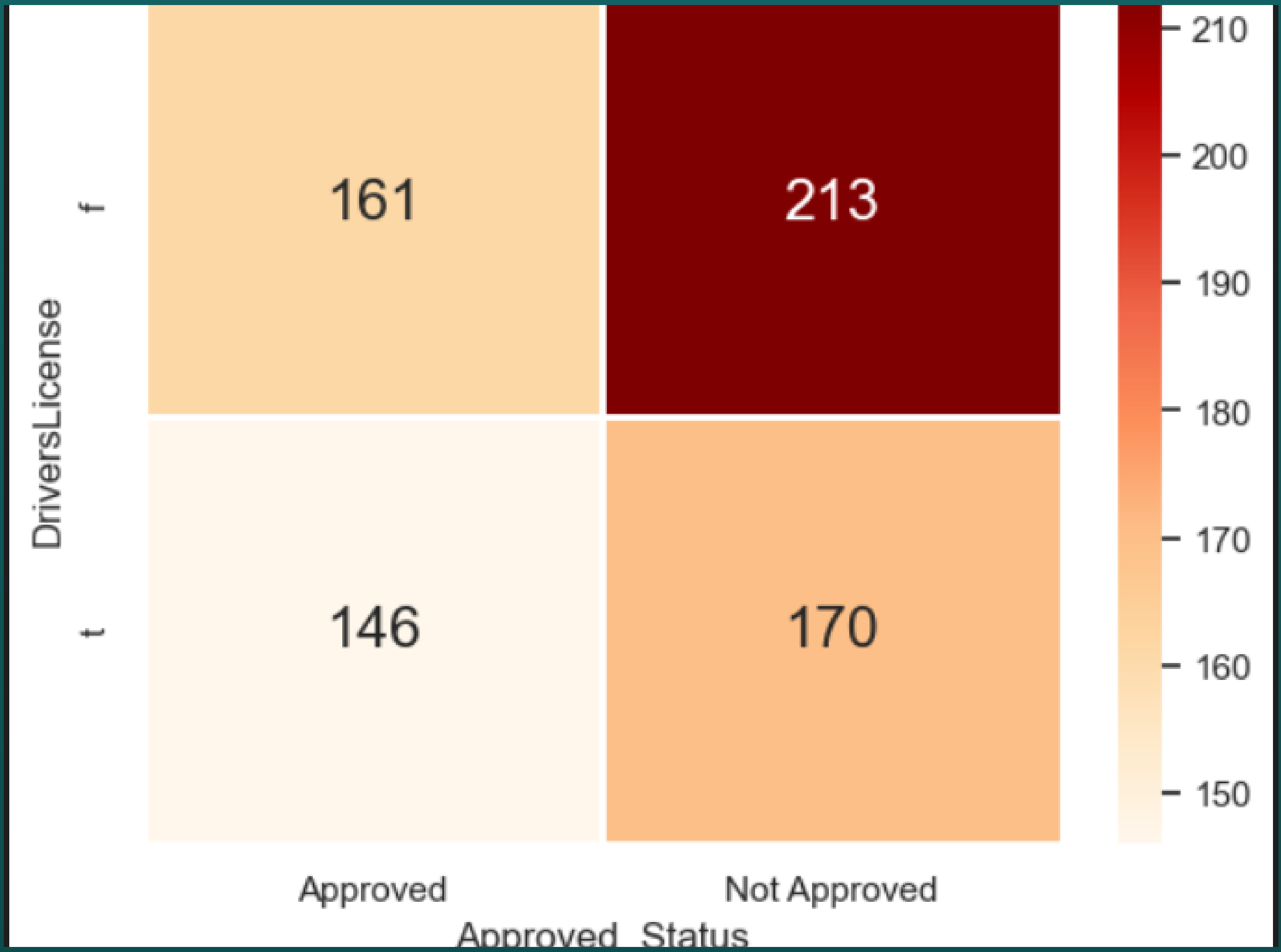
Age

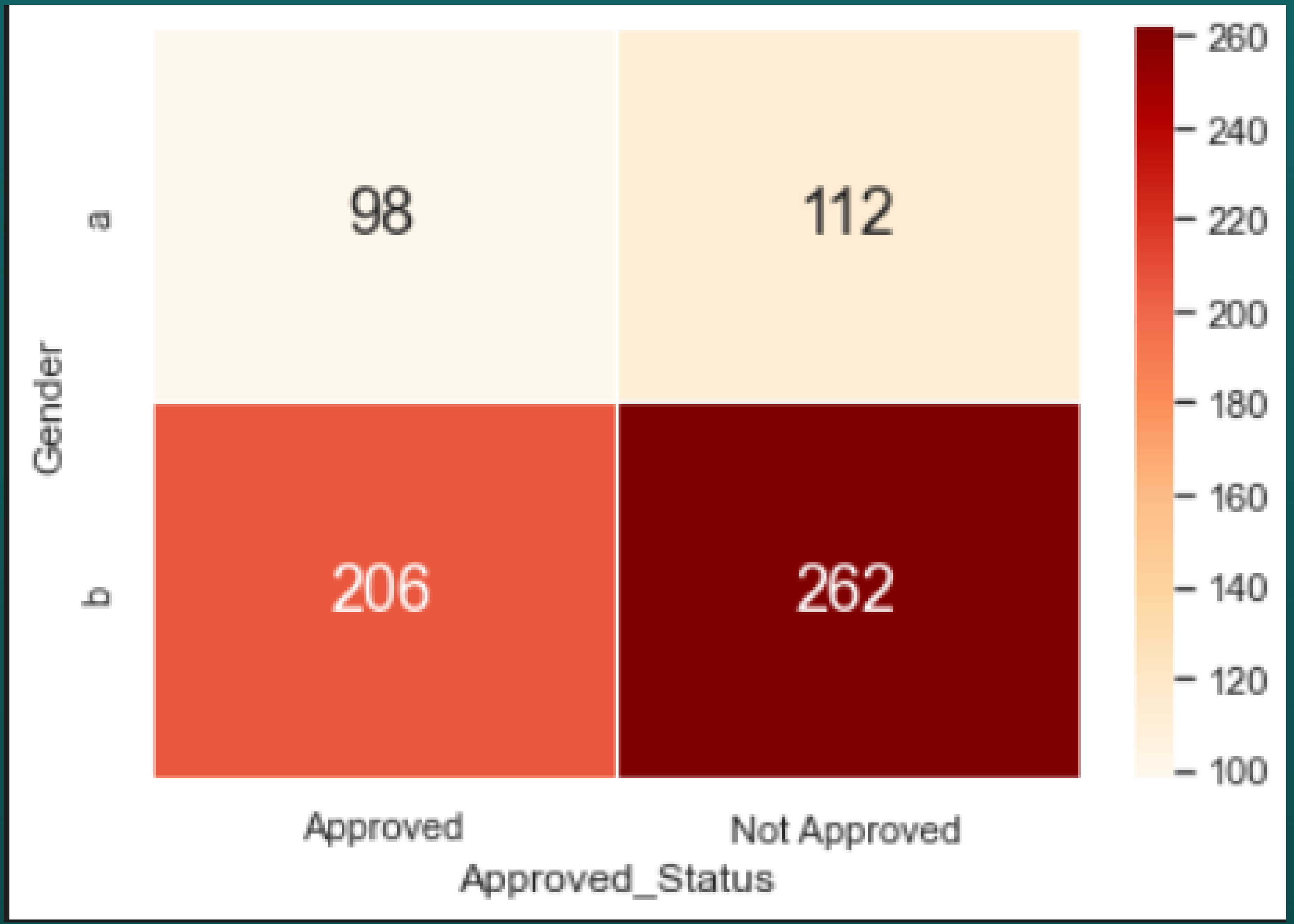


Years Employed



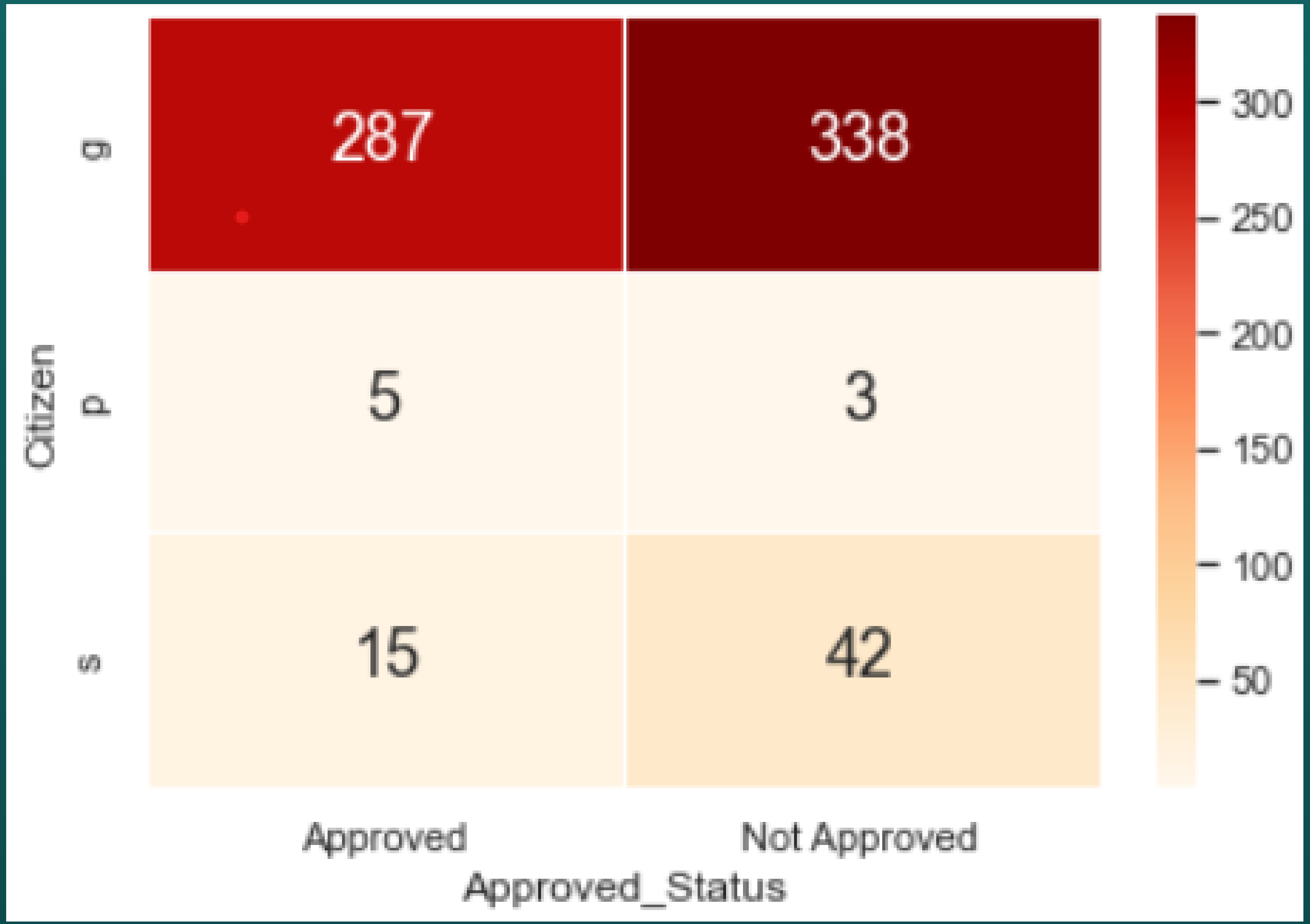
Categorical Data





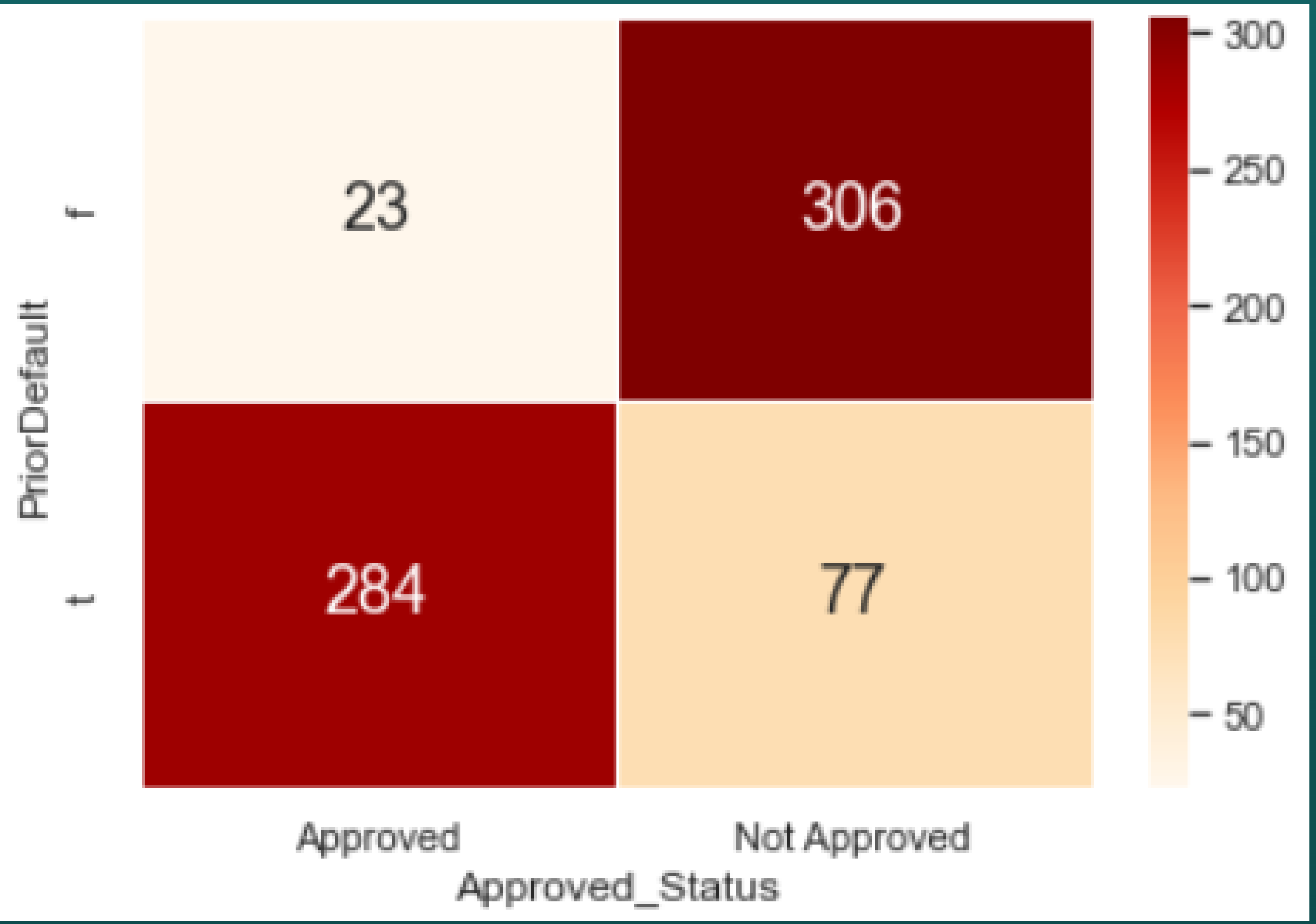
Drivers License

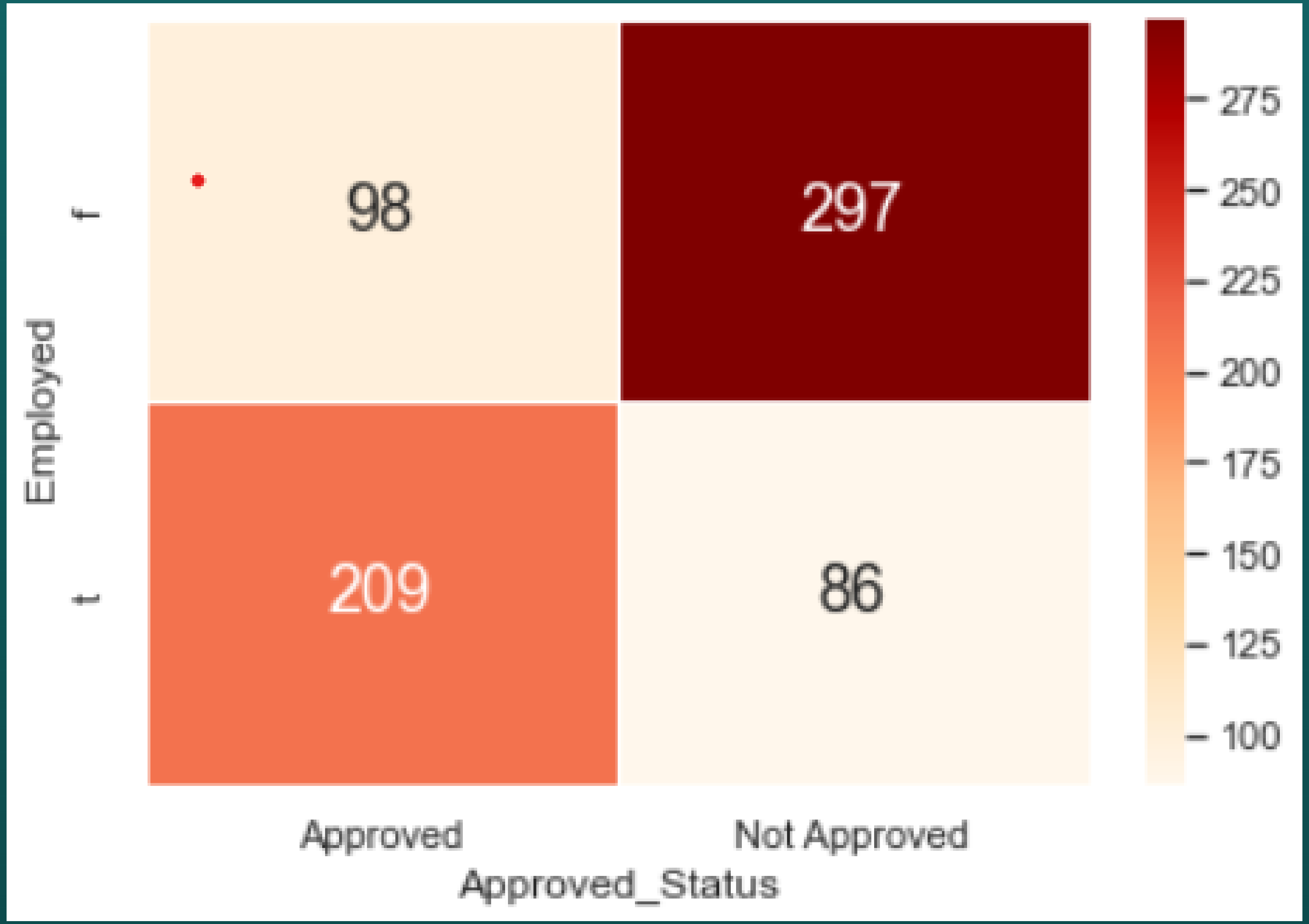
Gender



Citizen

Marital Status





PriorDefault

Employed

Model Training

- From a data set with a balanced number of values in the response variable, we tried to train some classification models to get better predictions.

- After a few attempts of model training using different approaches, we managed to improve the accuracy from about 0.81 to 0.89 on average.

- The highest attempt is with an accuracy of 0.91, with low false positive and low false negative rate.

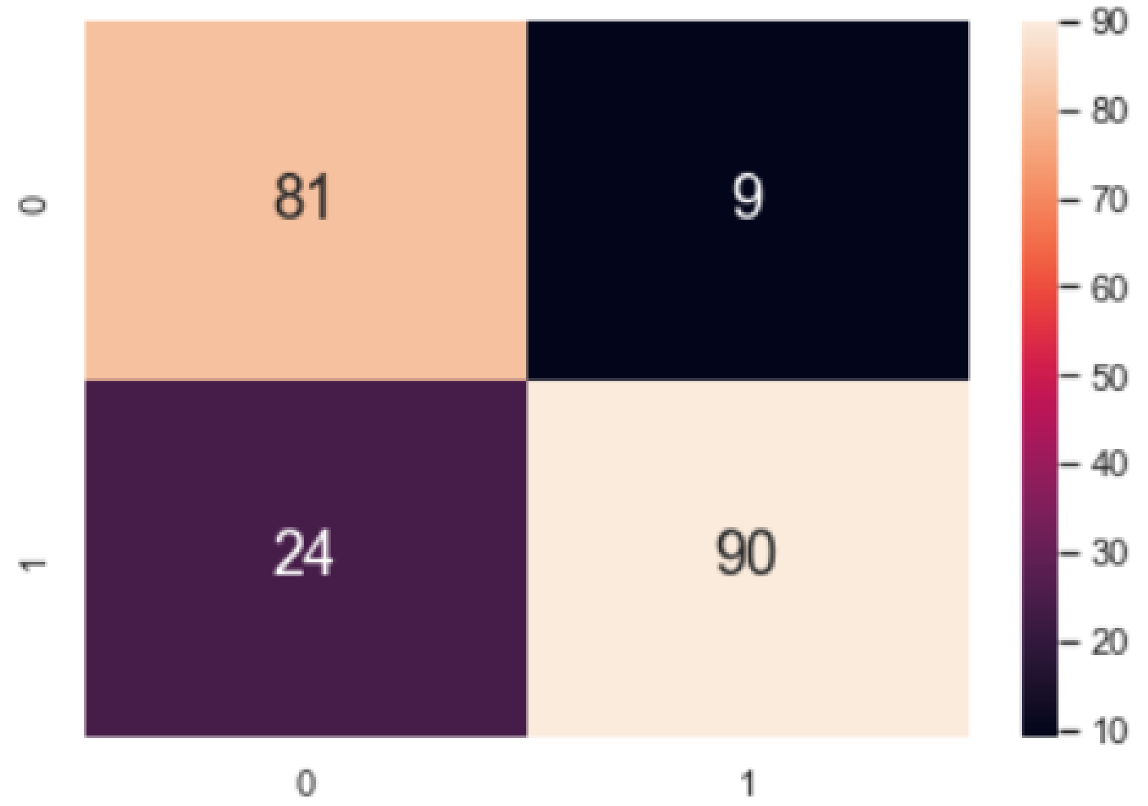
Attempt 1: Decision Tree

```
Test Data
Accuracy  :       0.8382352941176471

TPR Test :        0.7894736842105263
TNR Test :        0.9

FPR Test :        0.1
FNR Test :        0.21052631578947367

Precision: 0.7714
Recall: 0.9000
F-score: 0.8308
```



Attempt 2: Random Forest Classifier (Cross-Validation)

```
Train Accuracy: 0.865424430641822
 Test Accuracy: 0.8309178743961353
```

Current accuracy on the test set = 0.83

We try to improve this!

```
TPR Test :        0.7818181818181819
TNR Test :        0.8865979381443299

FPR Test :        0.1134020618556701
FNR Test :        0.2181818181818181 7

Precision: 0.7818
Recall: 0.8866
F-score: 0.8309
```

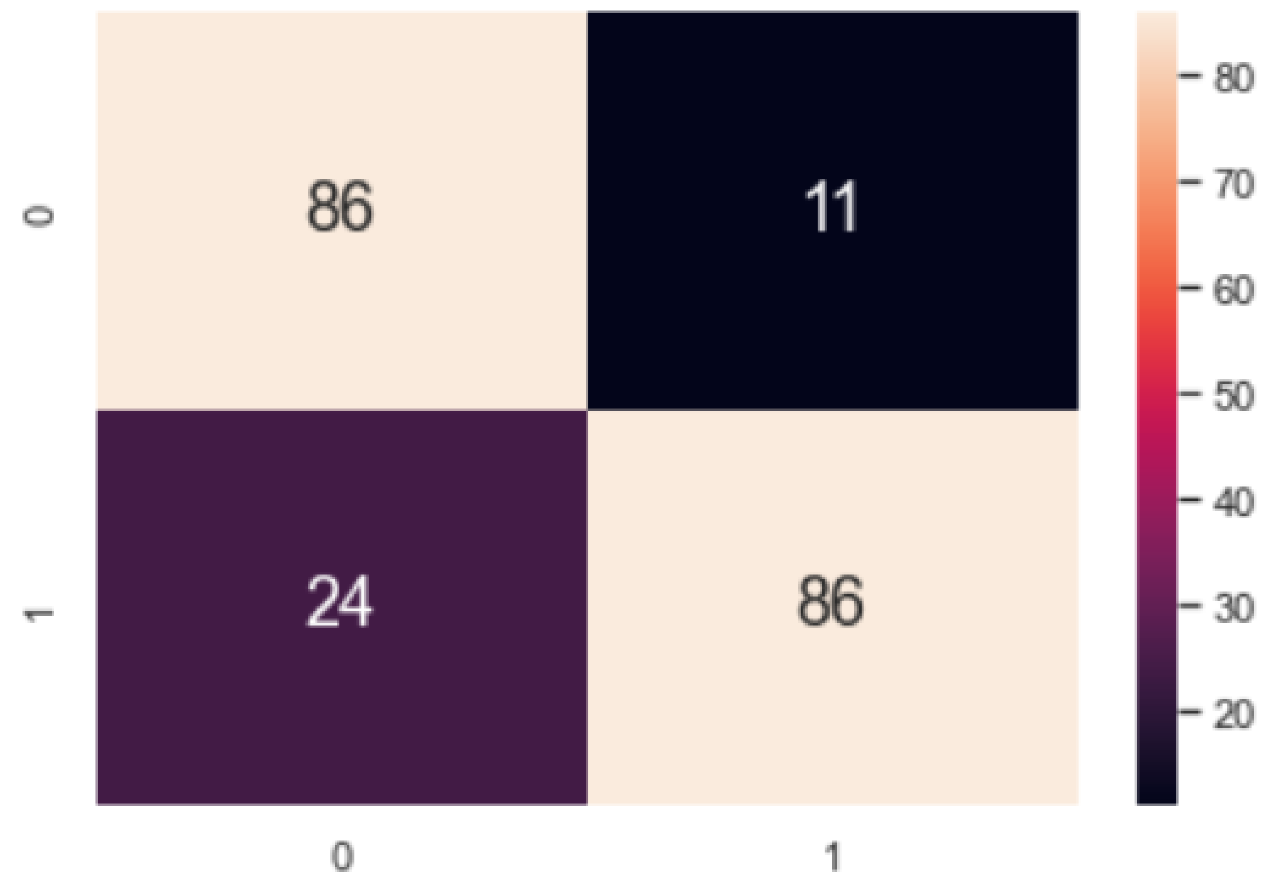
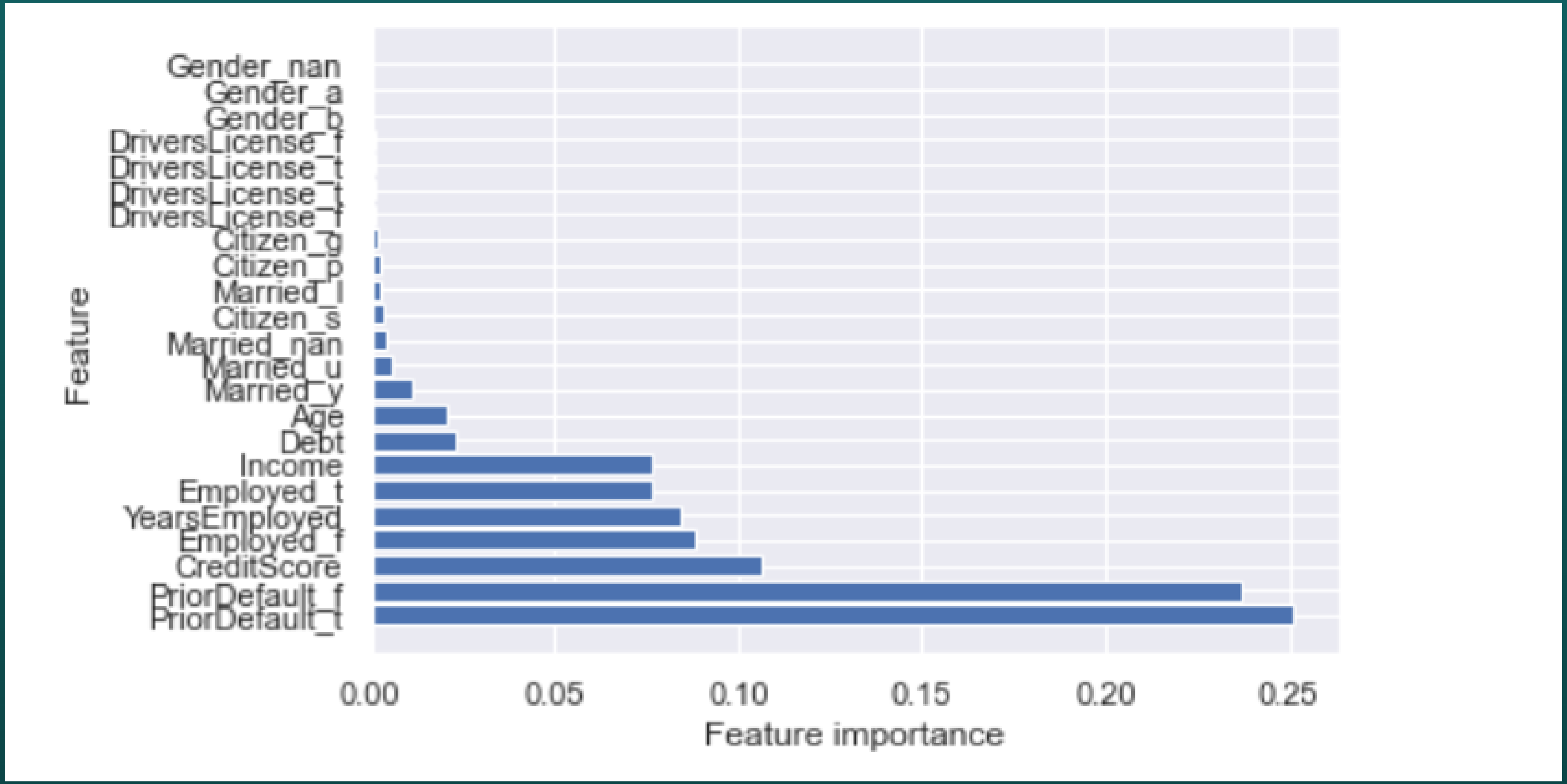


Figure out Features with Higher Importance



From the graph, we extract some of the more important features:

i. Numerical Predictors: Debt, Income, Credit_Score, Years_Employed,

ii. Categorical Predictors: Employed, Prior_Default

Attempt 3:
Logistic Regression with the extracted predictors

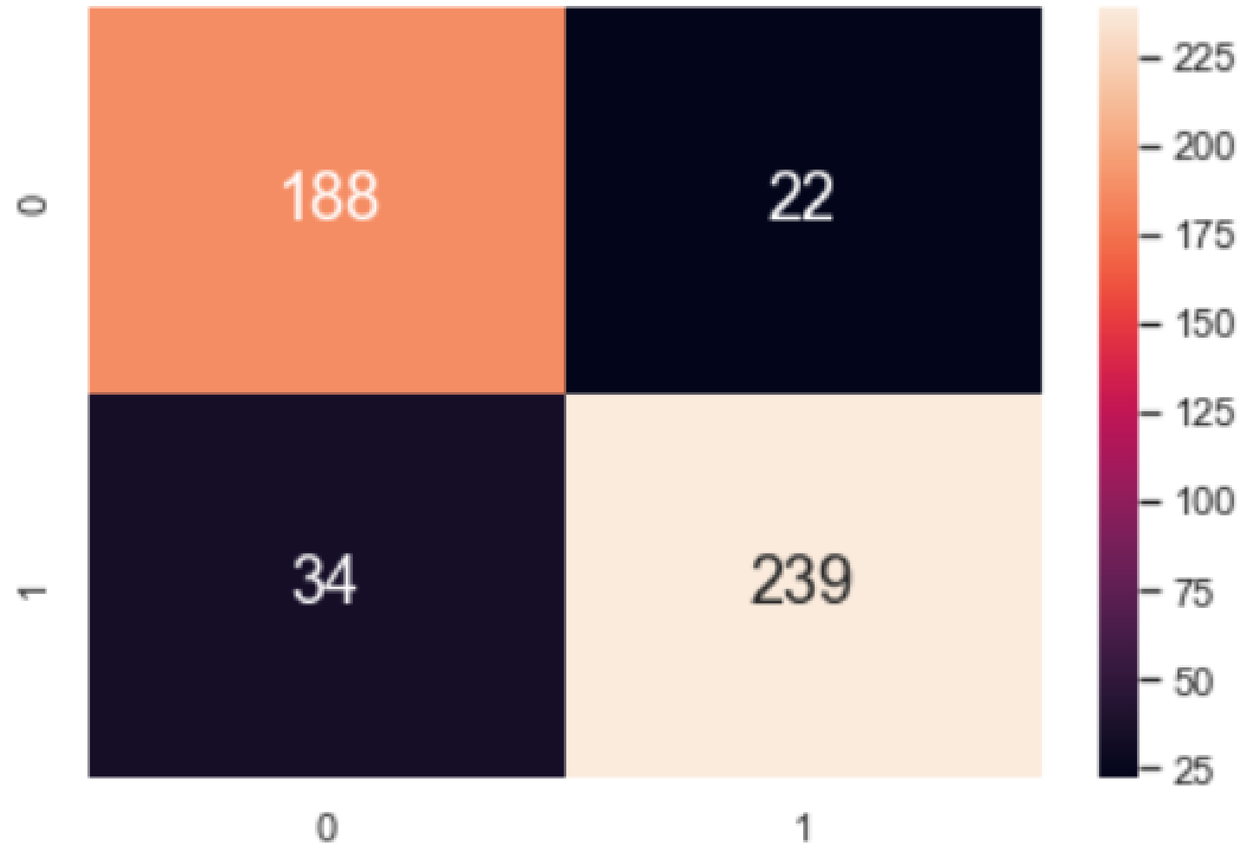
```
Train Data
Accuracy  :      0.8840579710144928

Test Data
Accuracy  :      0.8502415458937198
```

```
TPR Train :      0.8754578754578755
TNR Train :      0.8952380952380953

FPR Train :      0.10476190476190476
FNR Train :      0.12454212454212454

Precision: 0.8468
Recall: 0.8952
F-score: 0.8704
```



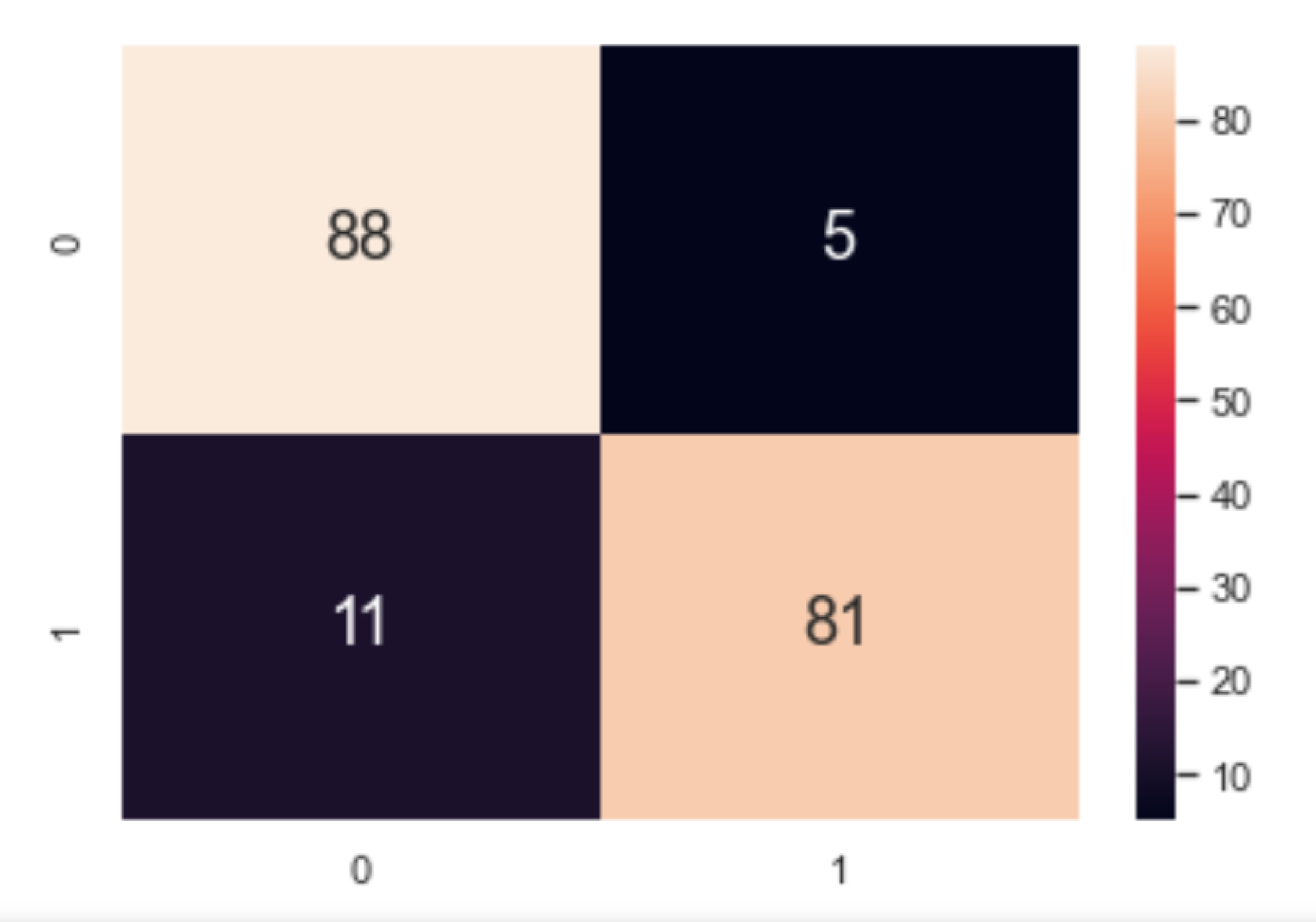
Attempt 4: Random Forest Classifier (cross-validation) using the extracted predictors

```
Test Data
Accuracy  :      0.9135135135135135
```

```
TPR Test :       0.8804347826086957
TNR Test :       0.946236559139785

FPR Test :       0.053763440860215055
FNR Test :       0.11956521739130435

Precision: 0.8889
Recall: 0.9462
F-score: 0.9167
```



Conclusion

From our classification model, a credit card application is likely to be approved when:

- debt <= 2.35K,
- income >= 397/month
- being employed
- years of employed of at least 5 years is prefferred,
- no record of prior default (active/passive cancellation of credit card)
- credit score > 2.5