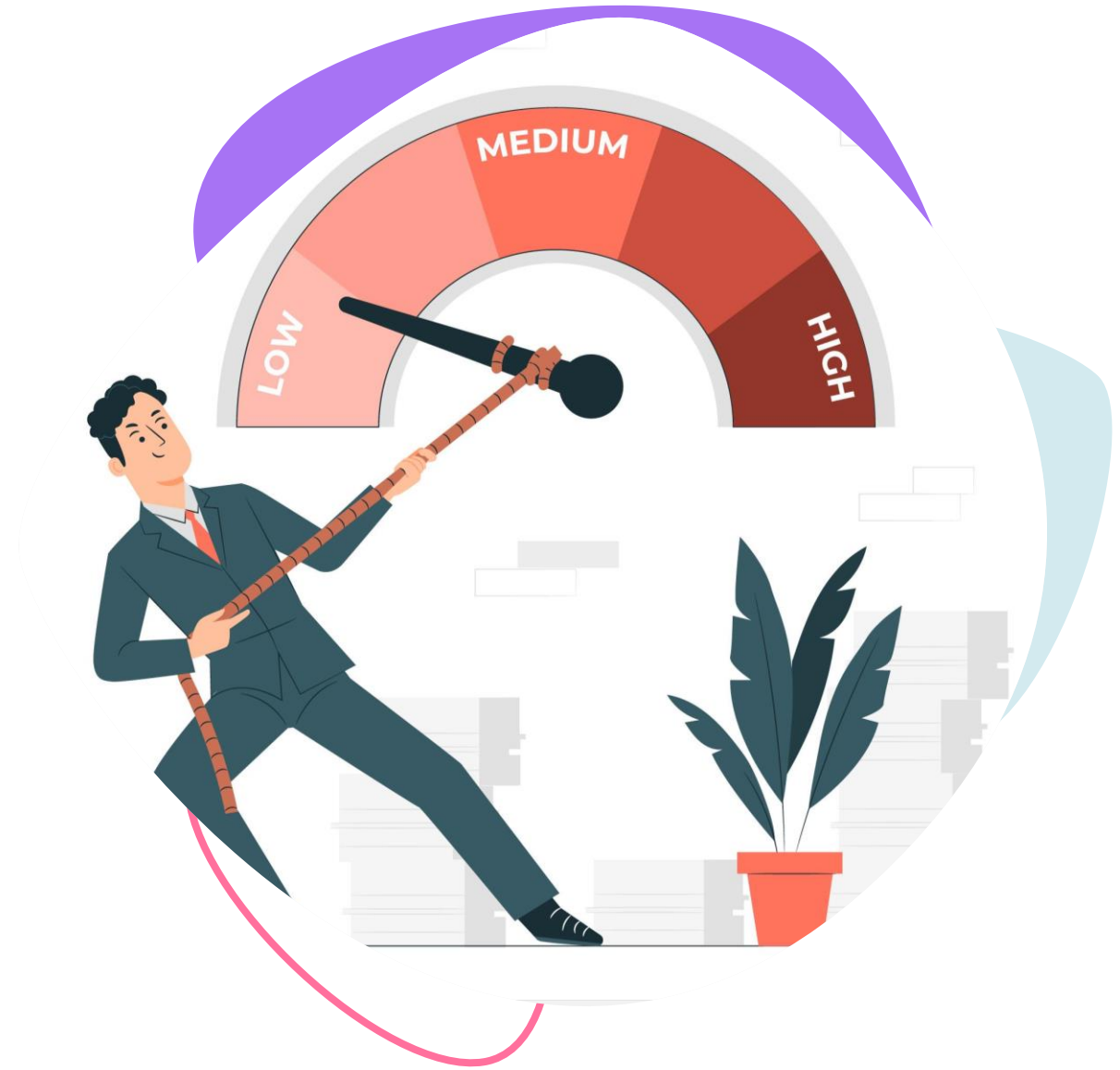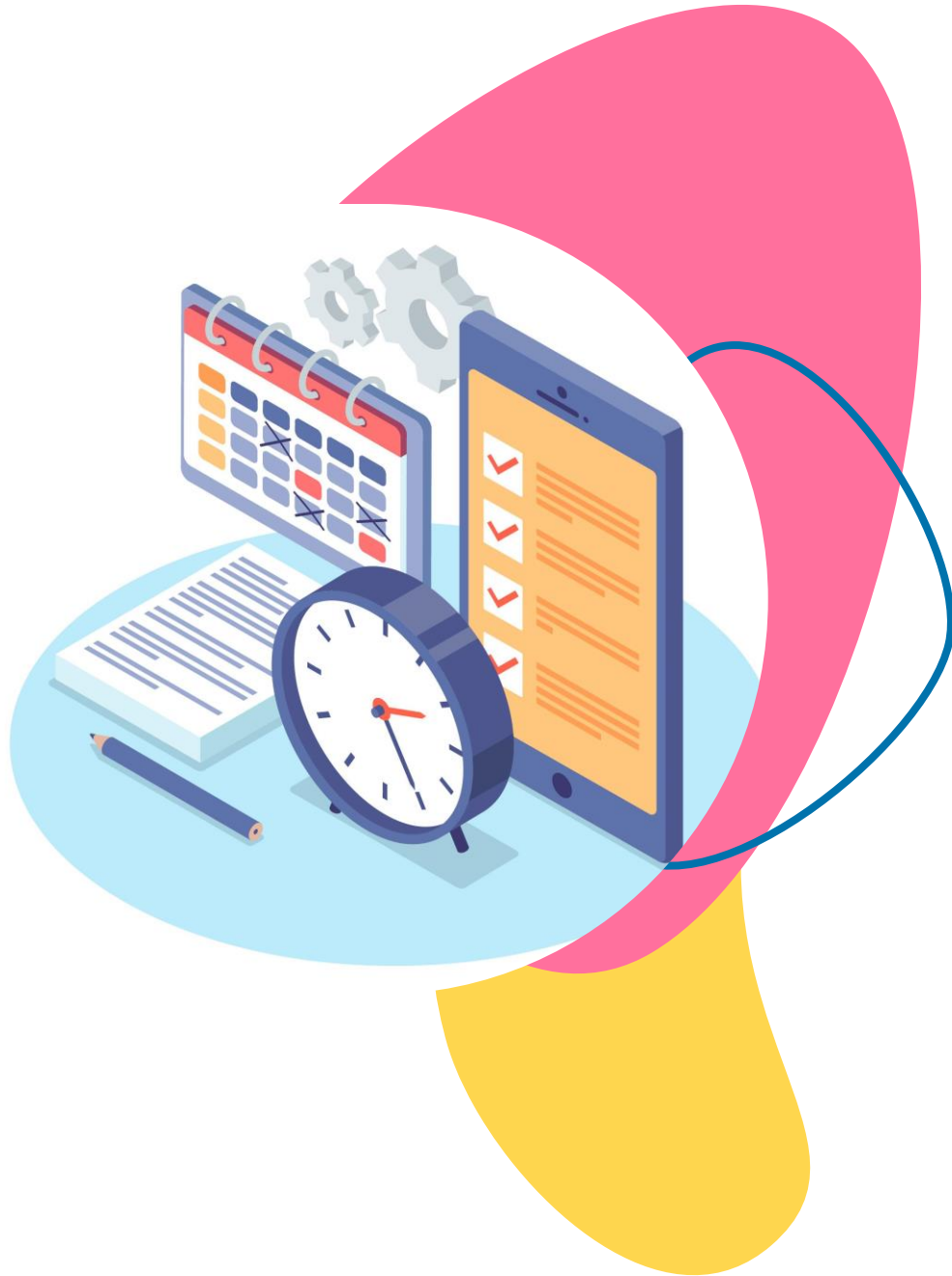# Credit Score Classification

Zeyad El-Sayed Usf

Agenda

- Introduction
- Dataset Overview .
- Exploratory Data Analysis.
- Data Pre-processing and Problems.
- Modeling and Results.
- Conclusion.

Introduction

The Credit Score Classification project aimed to develop a robust and accurate system for classifying credit scores.

Credit score classification plays a vital role in assessing an individual's creditworthiness and determining their eligibility for loans, credit cards, and other financial services.

It is a crucial component of risk assessment for financial institutions, helping them make informed decisions while minimizing the risk of defaults.

Exploratory Data Analysis

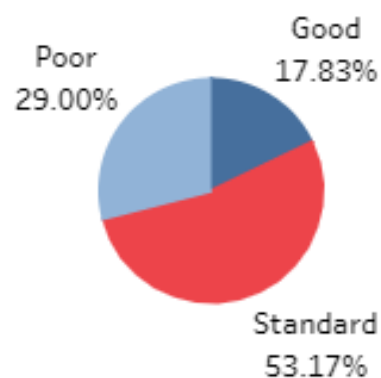- In Tableau : https://public.tableau.com/app/profile/zeyad.usf/viz/CreditScoreEDA/CreditScoreEDA2

NO.Recordes

100,000

NO.Unique

12,500

CreditScore

Poor 29.00%

Good 17.83%

Standard 53.17%

Occupations

Occupation	Credit Score		
	Good	Poor	Standard
Accountant	236	363	753
Architect	240	344	757
Developer	244	338	745
Doctor	232	328	743
Engineer	251	366	748
Entrepreneur	230	355	715
Journalist	250	325	717
Lawyer	250	364	805
Manager	235	335	727
Mechanic	220	367	752
Media_Manager	253	316	744
Musician	227	323	704
Scientist	235	366	758
Teacher	234	357	732
Writer	182	356	700

Months

Credit Score / Month

Good — January 1,975, August 2,413

Poor — February 3,539, June 3,728

Standard — March 6,965, June 6,410
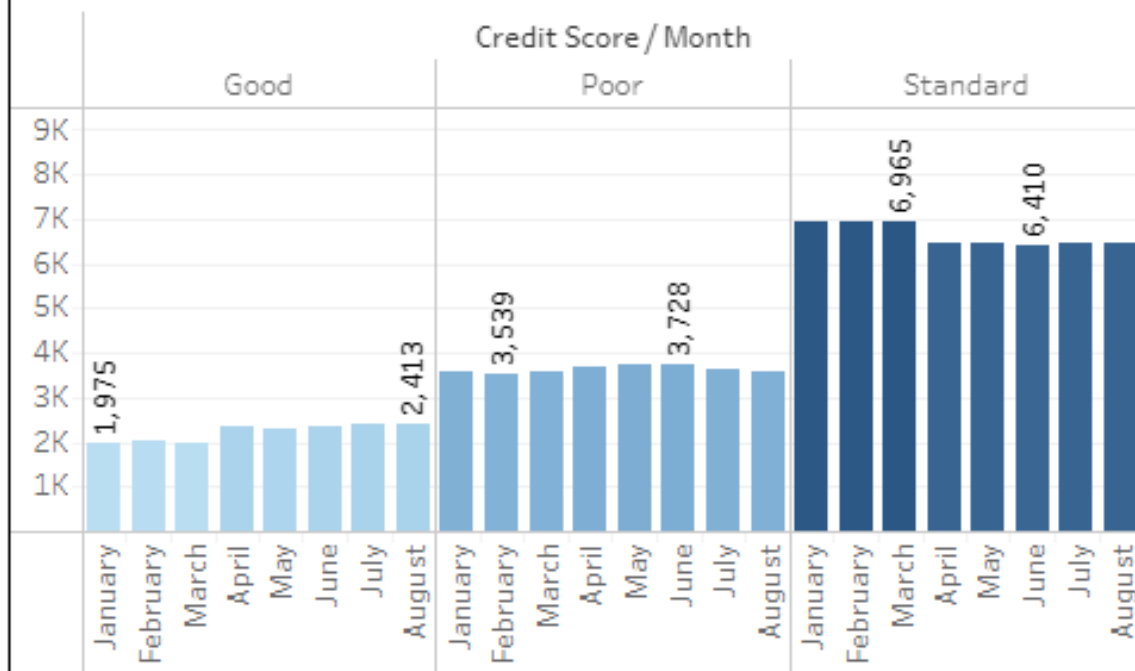
1,975 — 6,965

Avg.Age

Credit Score	
Good	36.389
Poor	31.182
Standard	33.426

Avg.Annual Income

Credit Score	
Good	65,204
Poor	40,585
Standard	50,987

Credit Max

Credit Mix / Credit Score

Bad — Good 121, Poor 2,351, Standa.. 2,494

Good — Good 2,497, Poor 890, Standa.. 3,164

Standard — Good 901, Poor 1,962, Standa.. 5,442

Payment Behaviour

Credit Score
- Good
- Poor
- Standard

Segment	Standard	Poor	Good
High spen Large value	4,809	2,037	1,984
High spent Medium value	5,828	2,675	2,174
High spent Small value	4,519	2,310	1,615
Low spent Large value	4,195	2,236	1,447
Low spent Medium value	5,006	2,846	1,721
Low spent Small value	6,765	4,033	1,934

Payment of Min Amount

NM
4,699
2,557
1,661
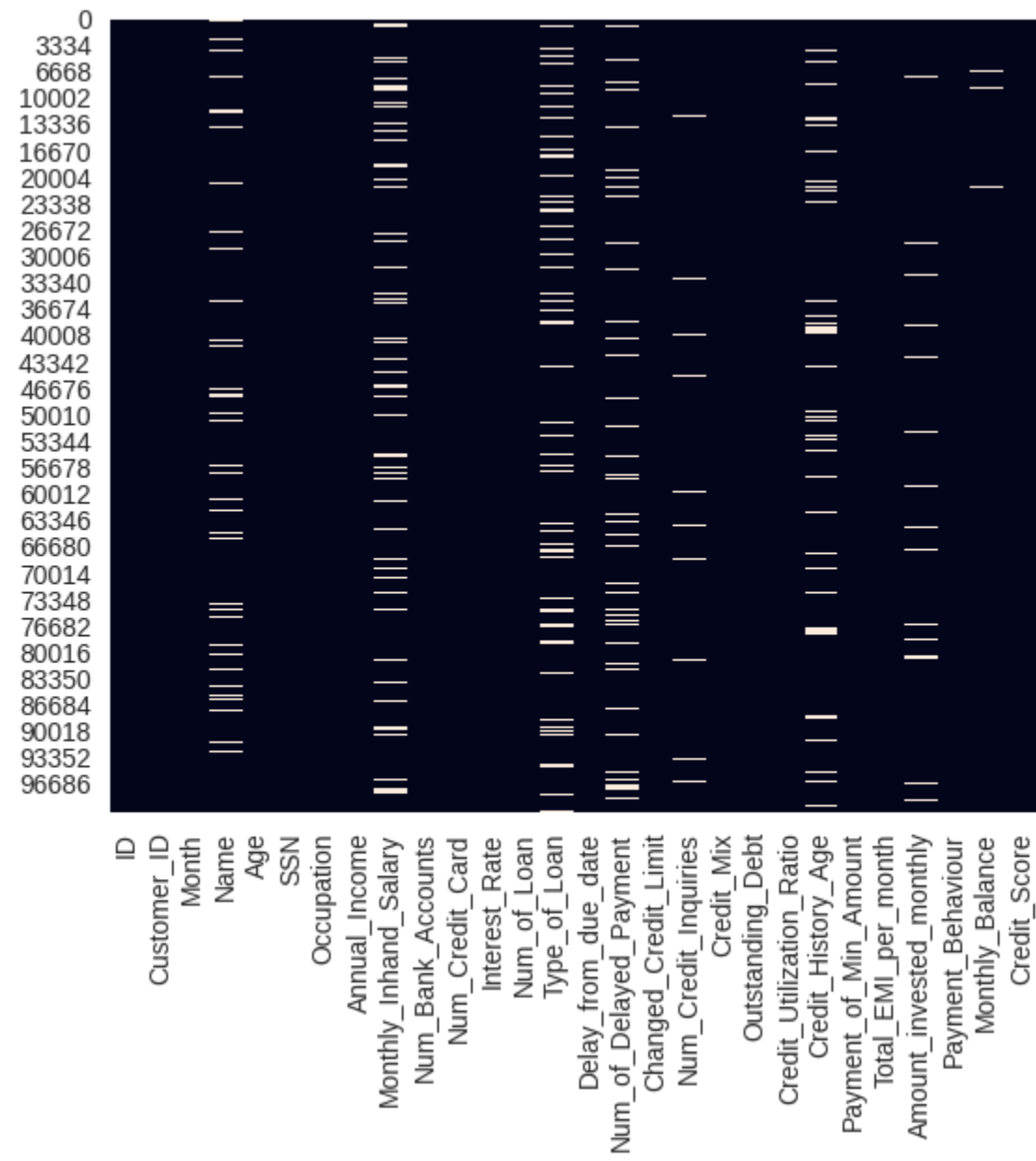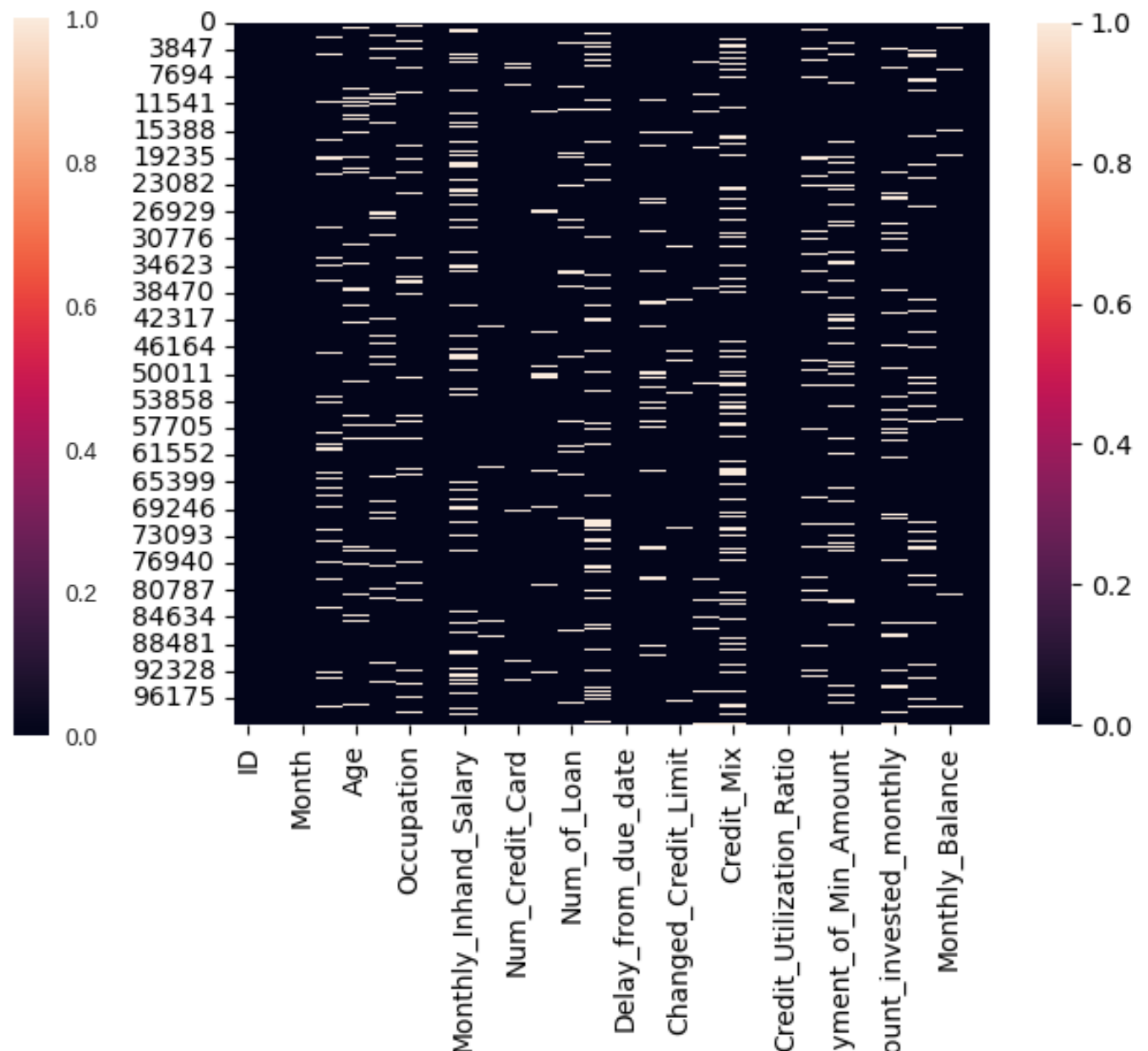
No
9,007
3,641
4,405

Yes
15,535
7,740
5,147

Data Pre-processing and Problems.

Missing 'NaN' Values In data

Missing 'NaN' and Wrong Values In data

Name

Sarah McBridec   Sarah McBridec
Sarah McBridec   Sarah McBridec
NaN   →   Sarah McBridec
Sarah McBridec   Sarah McBridec
Sarah McBridec   Sarah McBridec
Sarah McBridec   Sarah McBridec
Sarah McBridec   Sarah McBridec
Sarah McBridec   Sarah McBridec

Customer ID = CUS_0x8600

Age

30   30
30   30
30   → 30
181   30
30   30
30   30
31   31
31   31
31   31

Customer ID = CUS_0x33d2

Credit History Age

0 Years and 0 Months → 21 Years and 1 Months
21 Years and 2 Months   21 Years and 2 Months
21 Years and 3 Months   21 Years and 3 Months
21 Years and 4 Months   21 Years and 4 Months
21 Years and 5 Months   21 Years and 5 Months
21 Years and 6 Months   21 Years and 6 Months
21 Years and 7 Months   21 Years and 7 Months
0 Years and 0 Months → 21 Years and 8 Months

Customer ID = CUS_0xf16

- Type of Loan (random choice)

- Amount invested monthly (mean-min)

- Changed Credit Limit (mode)
- Num Credit Inquiries
- Payment Behaviour
- Num of Delayed Payment

- Monthly Balance (mean)
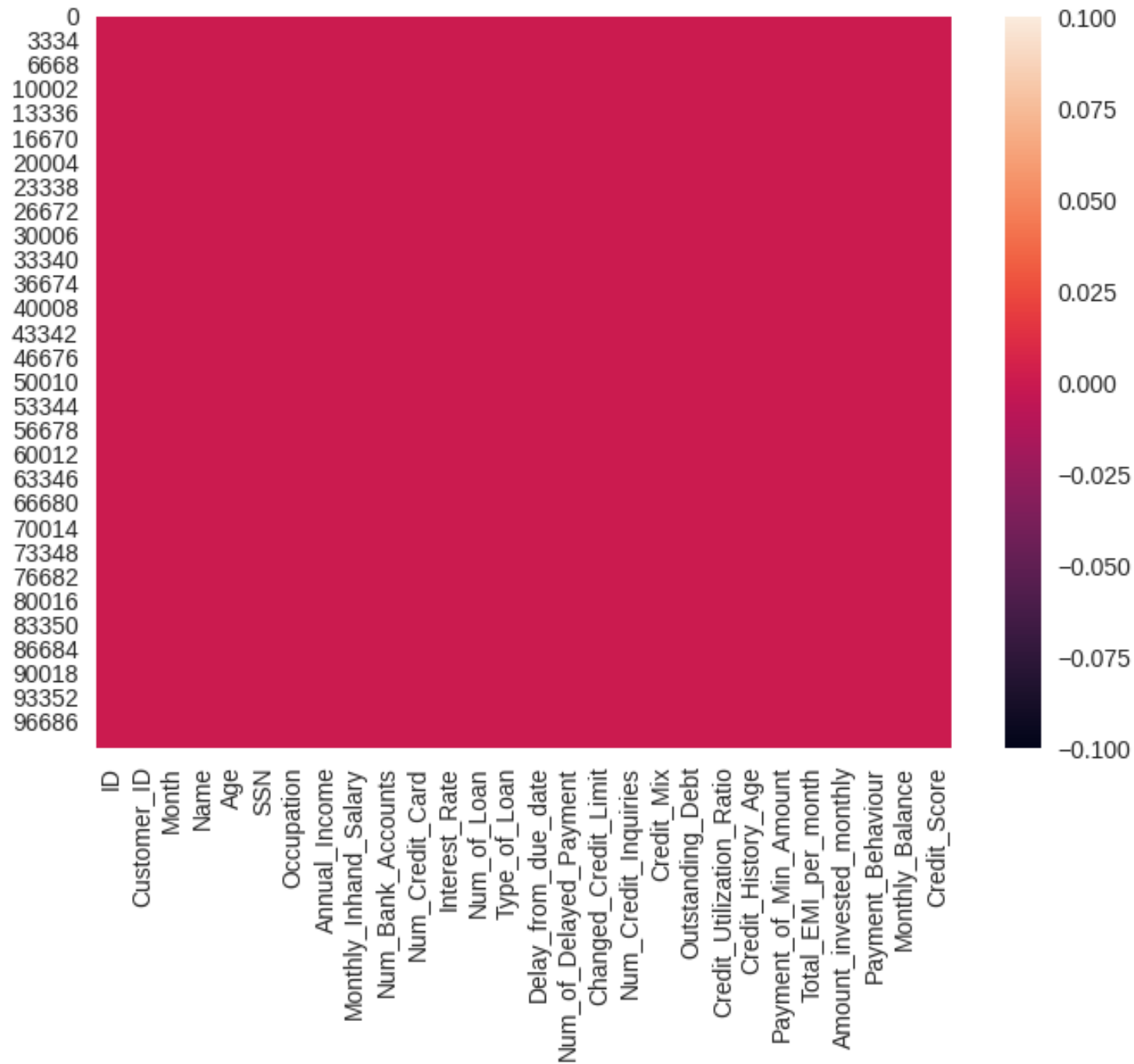
Models

- **PyCaret**
  - Extra Trees Classifier
  - Random Forest Classifier
  - Extreme Gradient Boosting
  - K Neighbors Classifier
 - Decision Tree Classifier
- **RandomForest**

PyCaret

PyCaret

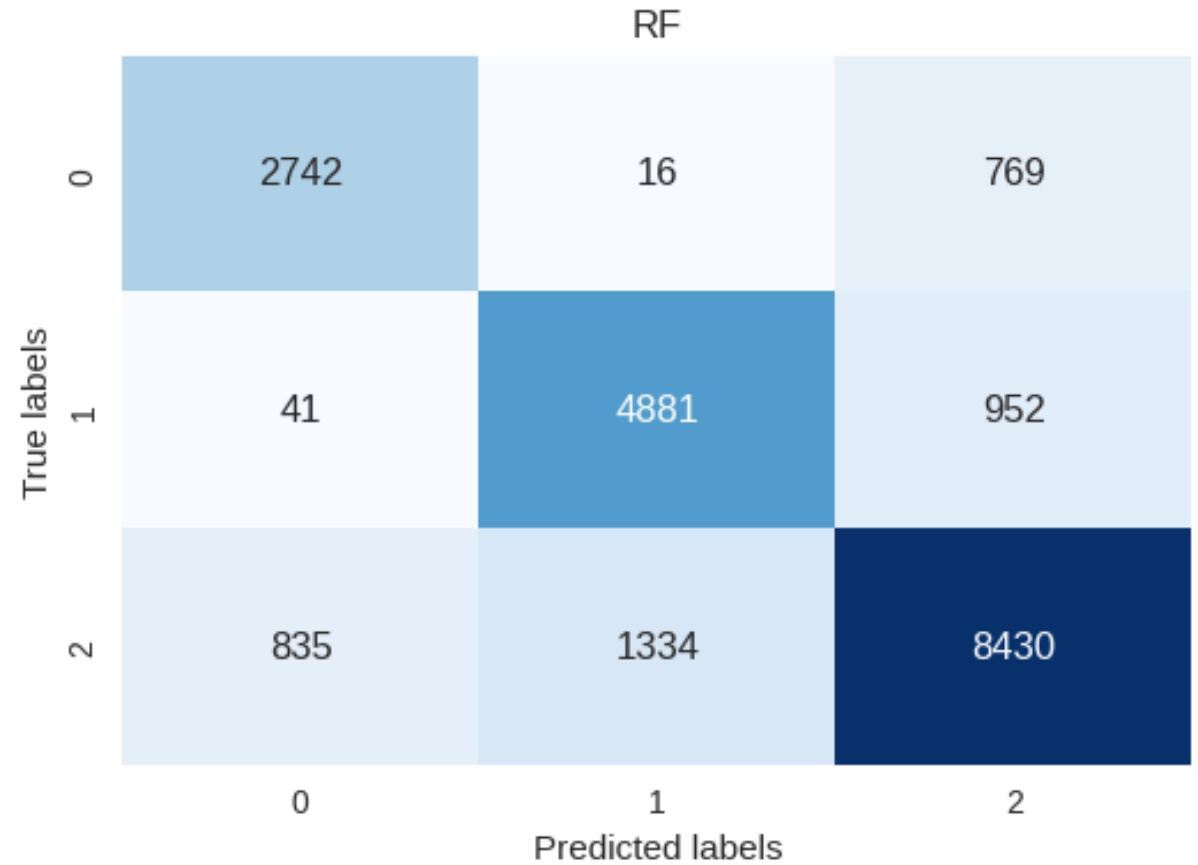|	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
et	Extra Trees Classifier	0.8664	0.9506	0.8664	0.8654	0.8654	0.7997	0.8001	15.8010
rf	Random Forest Classifier	0.8641	0.9541	0.8641	0.8633	0.8630	0.7962	0.7969	36.8500
xgboost	Extreme Gradient Boosting	0.8303	0.9467	0.8303	0.8300	0.8294	0.7455	0.7463	73.2510
knn	K Neighbors Classifier	0.8200	0.9270	0.8200	0.8249	0.8119	0.7300	0.7385	9.0000
dt	Decision Tree Classifier	0.7891	0.8418	0.7891	0.7890	0.7890	0.6837	0.6837	2.2940

Extra Trees Classifier

ExtraTreesClassifier Classification Report
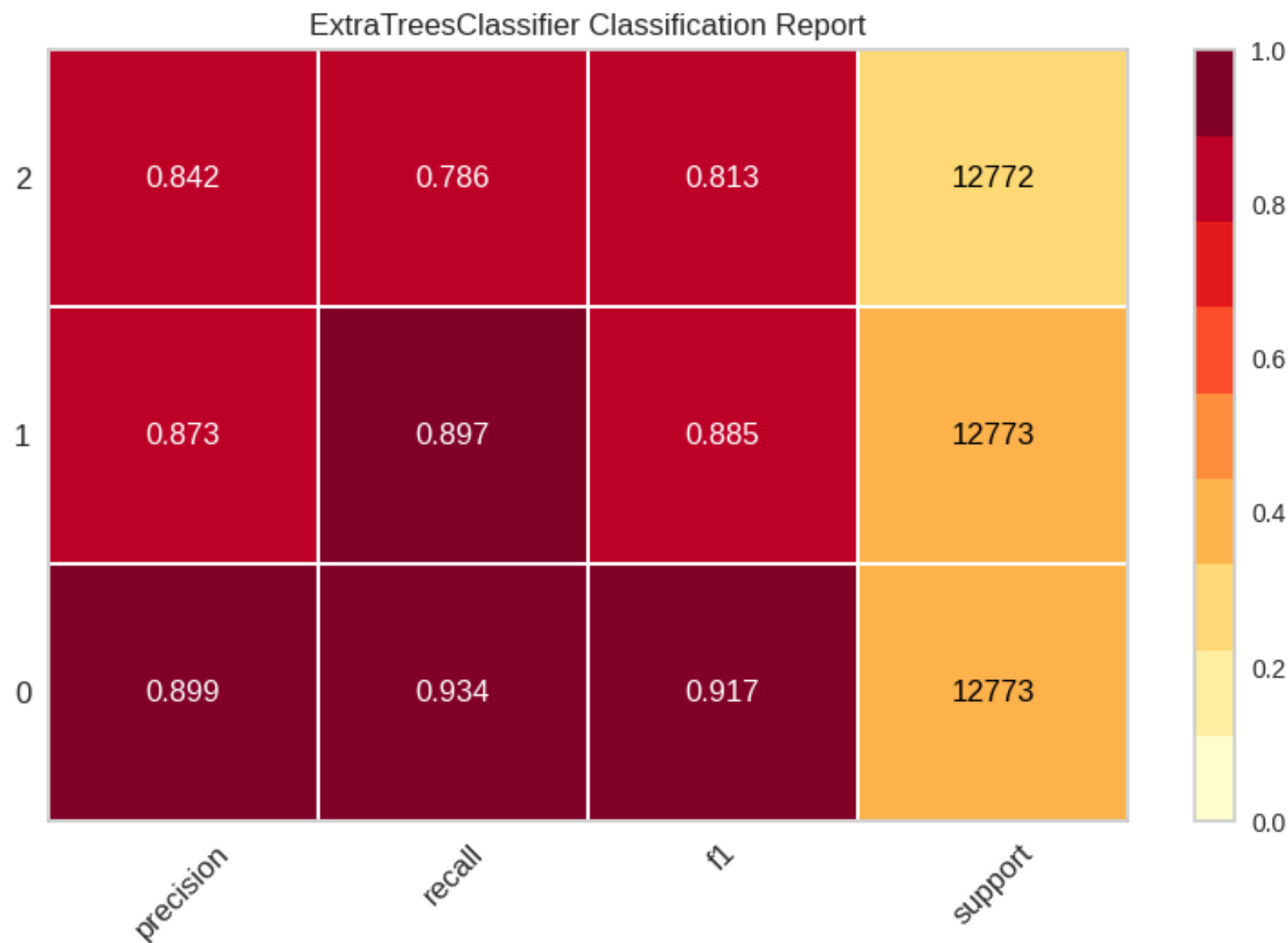
ExtraTreesClassifier Confusion Matrix
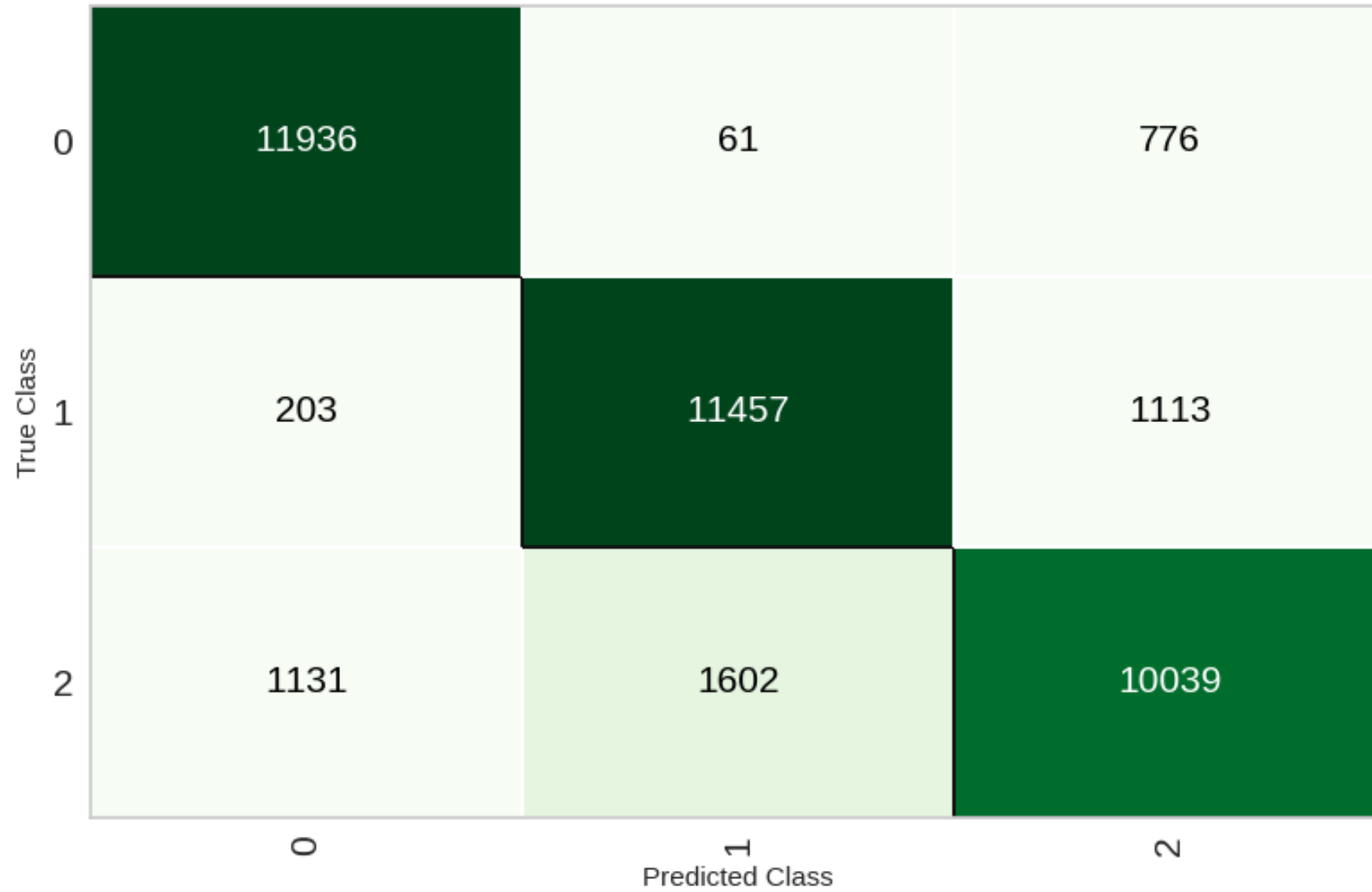
Random Forest

```
           precision    recall  f1-score   support

        0       0.76      0.78      0.77      3527
        1       0.78      0.83      0.81      5874
        2       0.83      0.80      0.81     10599

 accuracy                           0.80     20000
macro avg        0.79      0.80      0.80     20000
weighted avg     0.80      0.80      0.80     20000
```

Random Forest Model

RF

True labels	Predicted 0	Predicted 1	Predicted 2
0	2742	16	769
1	41	4881	952
2	835	1334	8430

Predicted labels

Thank you

My Accounts

Git Hub ● https://github.com/zeyadusf

Linkedin ● https://www.linkedin.com/in/zeyadusf/

Kaggel ● https://www.kaggle.com/zeyadusf