THE DEFAULTRIX

What is Defaultrix?

Content

Exploratory Data Analysis

Getting to know our customer

Pre-Processing

Briefly explanation about data processing

Model Result

How well the model performed?

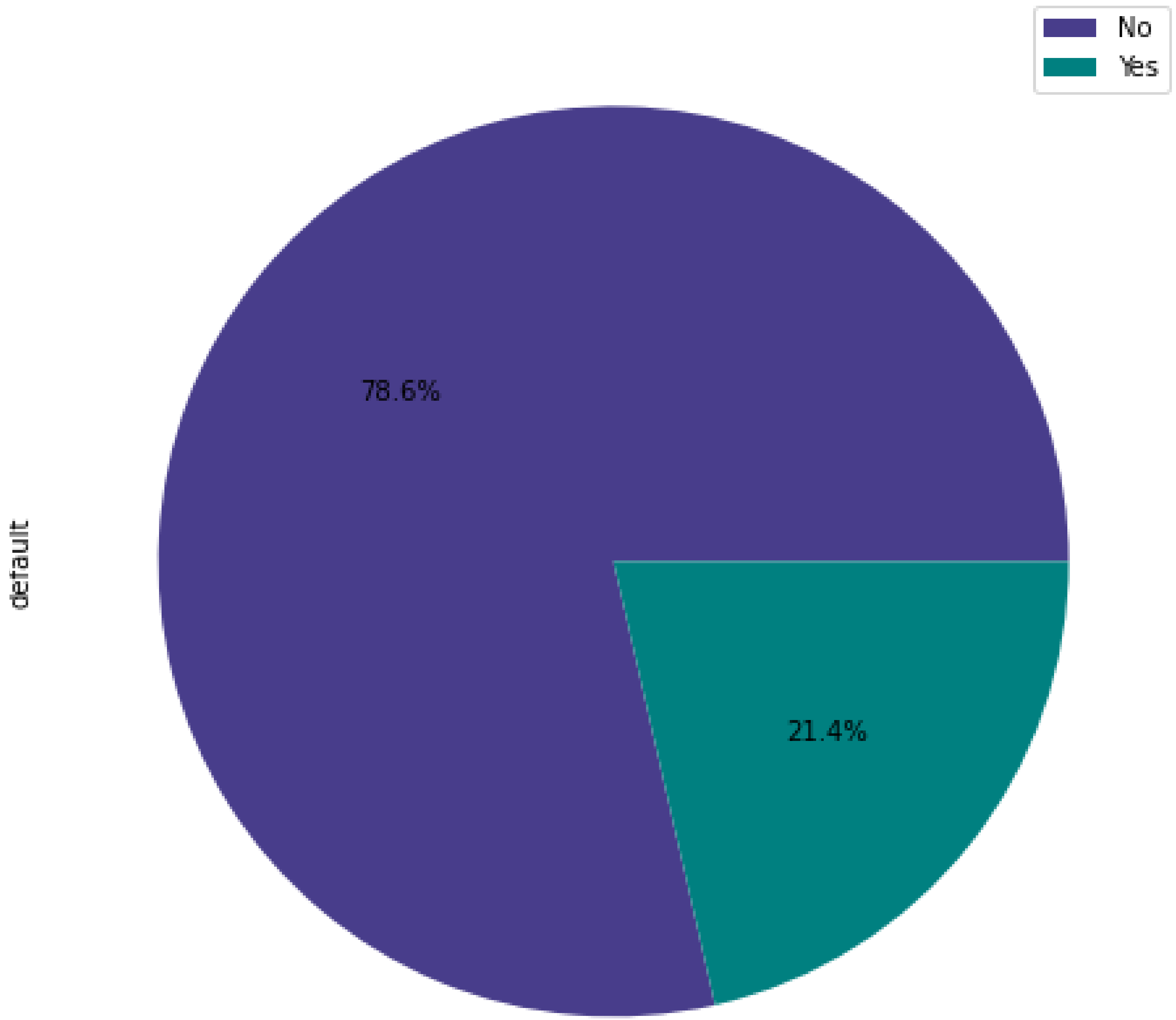
Next Steps

Our plan and suggestion for improvement

About The Data

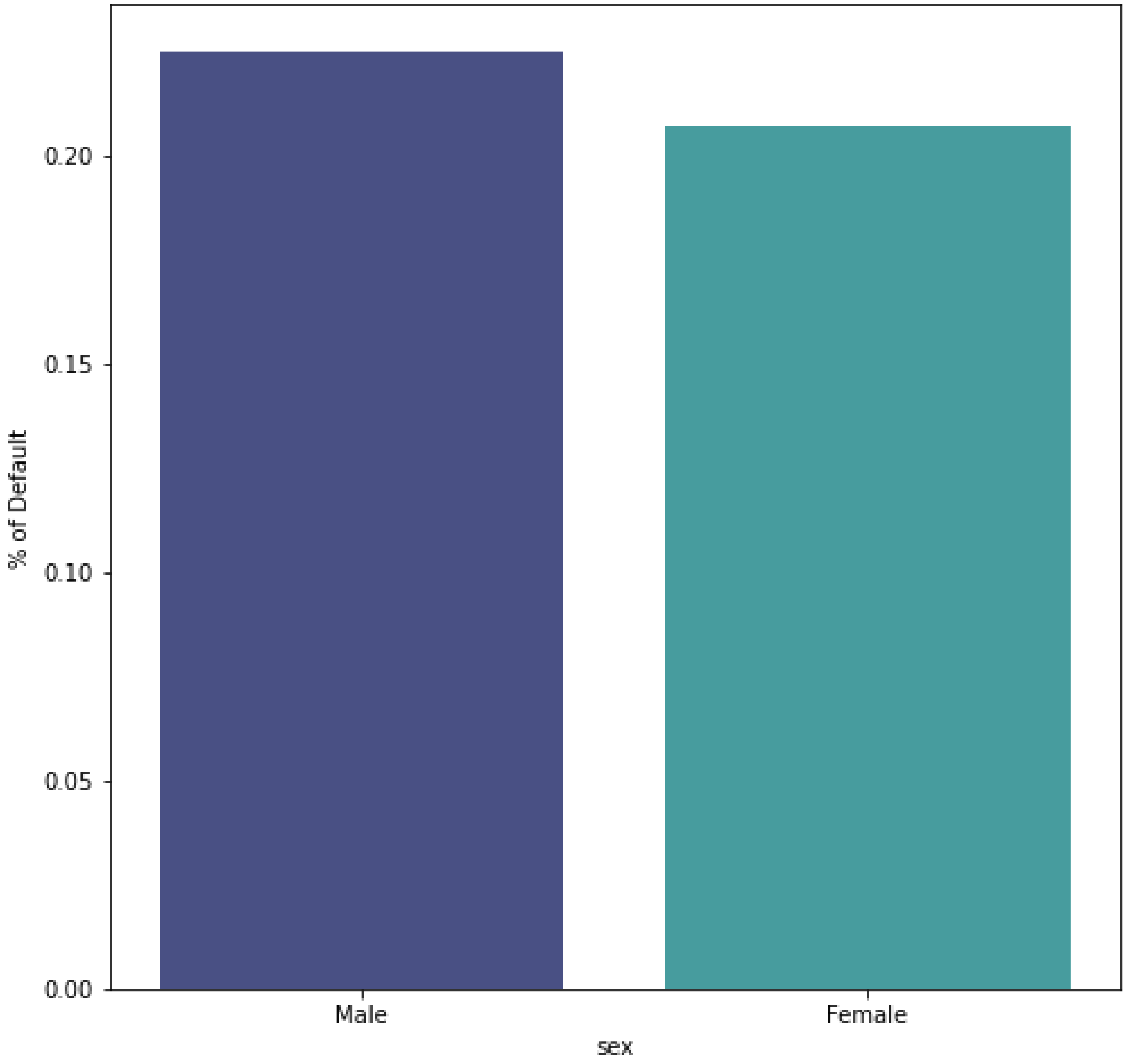
limit_balance	sex	education_level	marital_status	age	pay_0	pay_2	pay_3	pay_4	pay_5	pay_6	bill_amt_1	bill_amt_2	bill_amt_3
50000	1	1	2	39	0	0	0	0	0	0	47174	47974	48630
110000	2	1	2	29	0	0	0	0	0	0	48088	45980	44231
270000	1	1	2	36	0	0	0	2	0	0	78630	68921	46512
130000	1	1	1	45	0	0	0	0	0	0	58180	59134	61156
50000	1	1	2	24	0	0	0	0	0	0	42058	35340	22110
20000	1	1	2	29	0	0	2	0	0	0	14897	17512	16926
220000	1	1	2	38	0	0	0	0	0	0	209044	211453	217237
50000	1	1	1	42	0	0	0	0	0	0	49887	49515	38680
170000	1	1	1	41	0	0	0	0	0	0	149941	68912	72741
50000	2	1	2	24	0	0	2	0	0	0	52227	55264	52028

bill_amt_4	bill_amt_5	bill_amt_6	pay_amt_1	pay_amt_2	pay_amt_3	pay_amt_4	pay_amt_5	pay_amt_6	default_payment_next_month
50803	30789	15874	1800	2000	3000	2000	2000	2000	0
32489	26354	20221	2000	2010	3000	3000	3000	1000	0
40335	37165	22156	10076	4018	14	2051	2000	0	0
62377	63832	65099	2886	2908	2129	2354	2366	2291	0
19837	19855	20151	1367	1606	692	709	721	692	0
17368	17959	19023	3170	0	1000	1000	3000	0	0
198681	202479	206221	7705	9656	7189	7404	7490	7517	0
29664	29057	29083	1850	1507	1100	1200	1116	2900	0
76149	84474	92400	3200	6000	5000	10000	10000	780	0
32140	32216	31482	5350	0	1085	2000	1226	2415	0



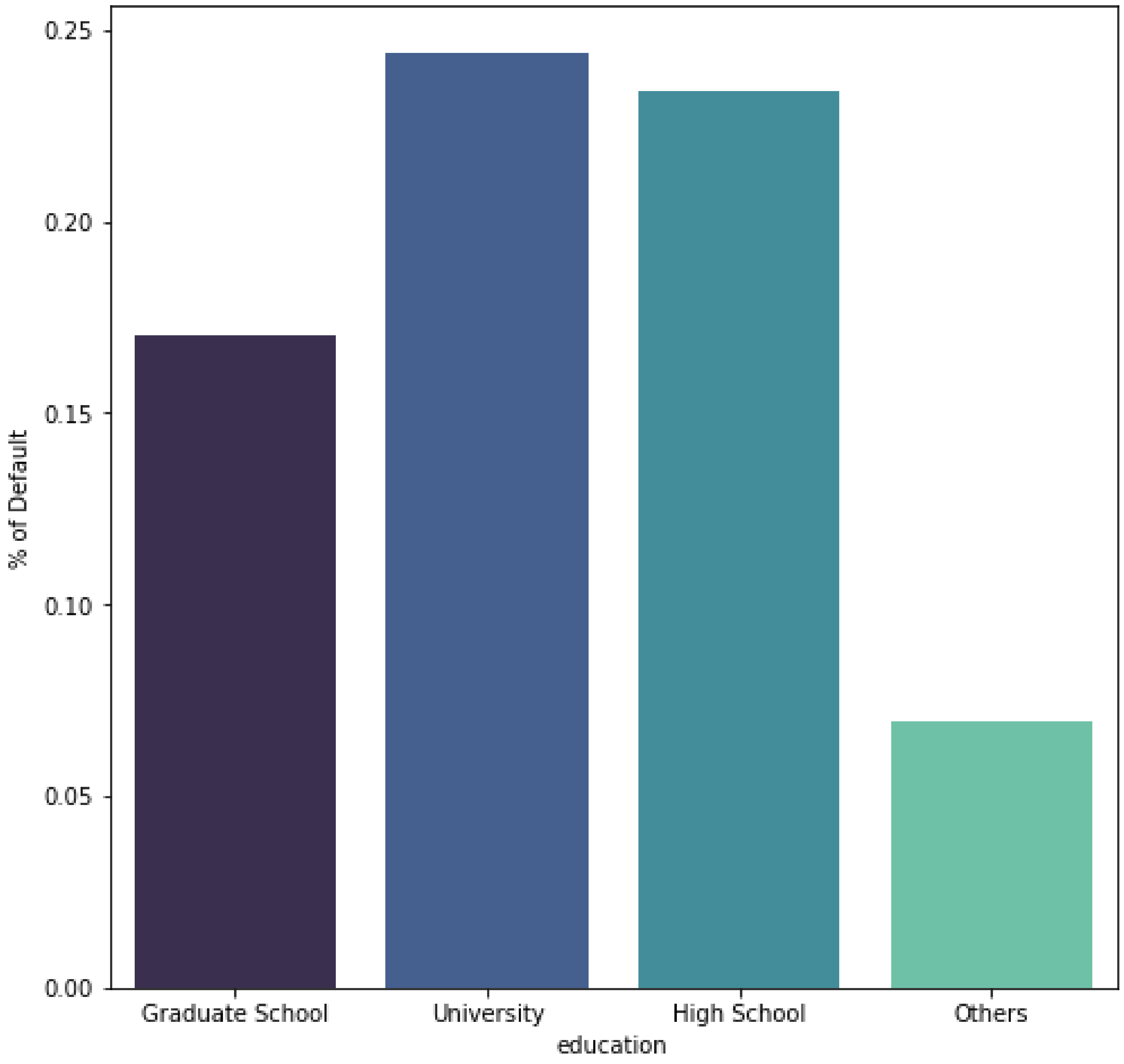
Default Customer

21% of our customer defaulting their credit card payment.



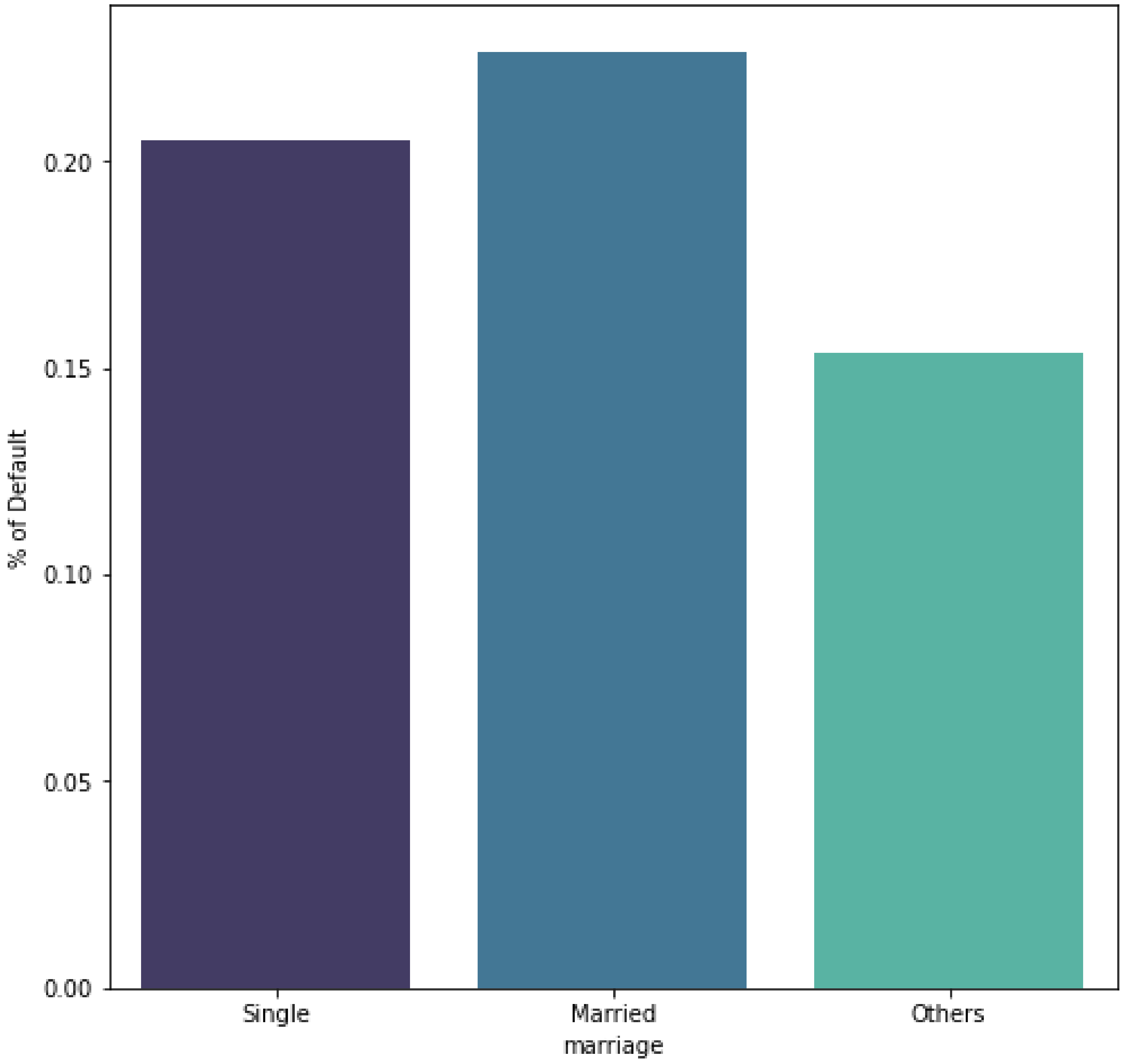
Based on Gender

Men are more likely than women to default the payment



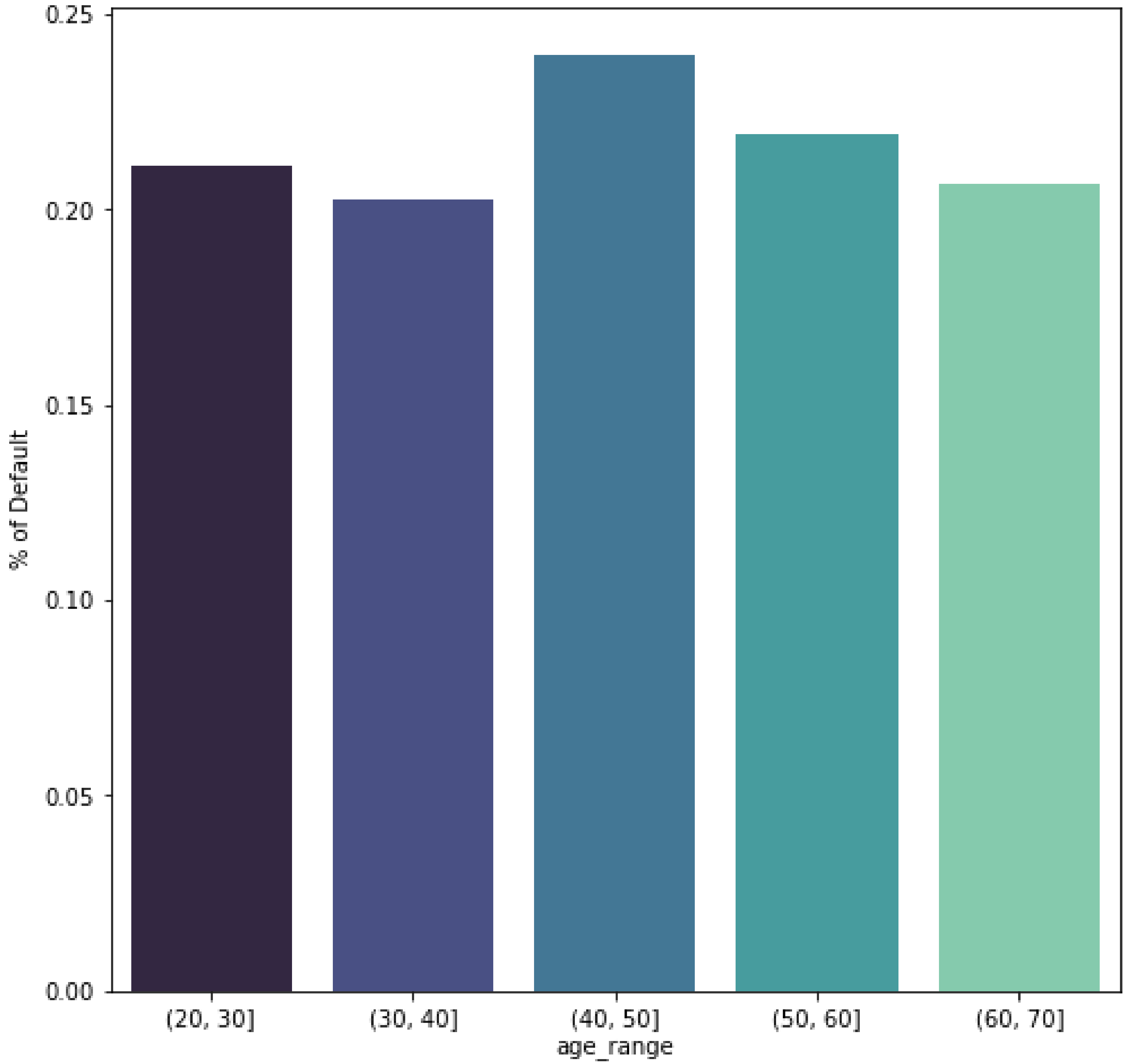
Based on Education

Customer with university and high school degree are more likely than others to default their payment



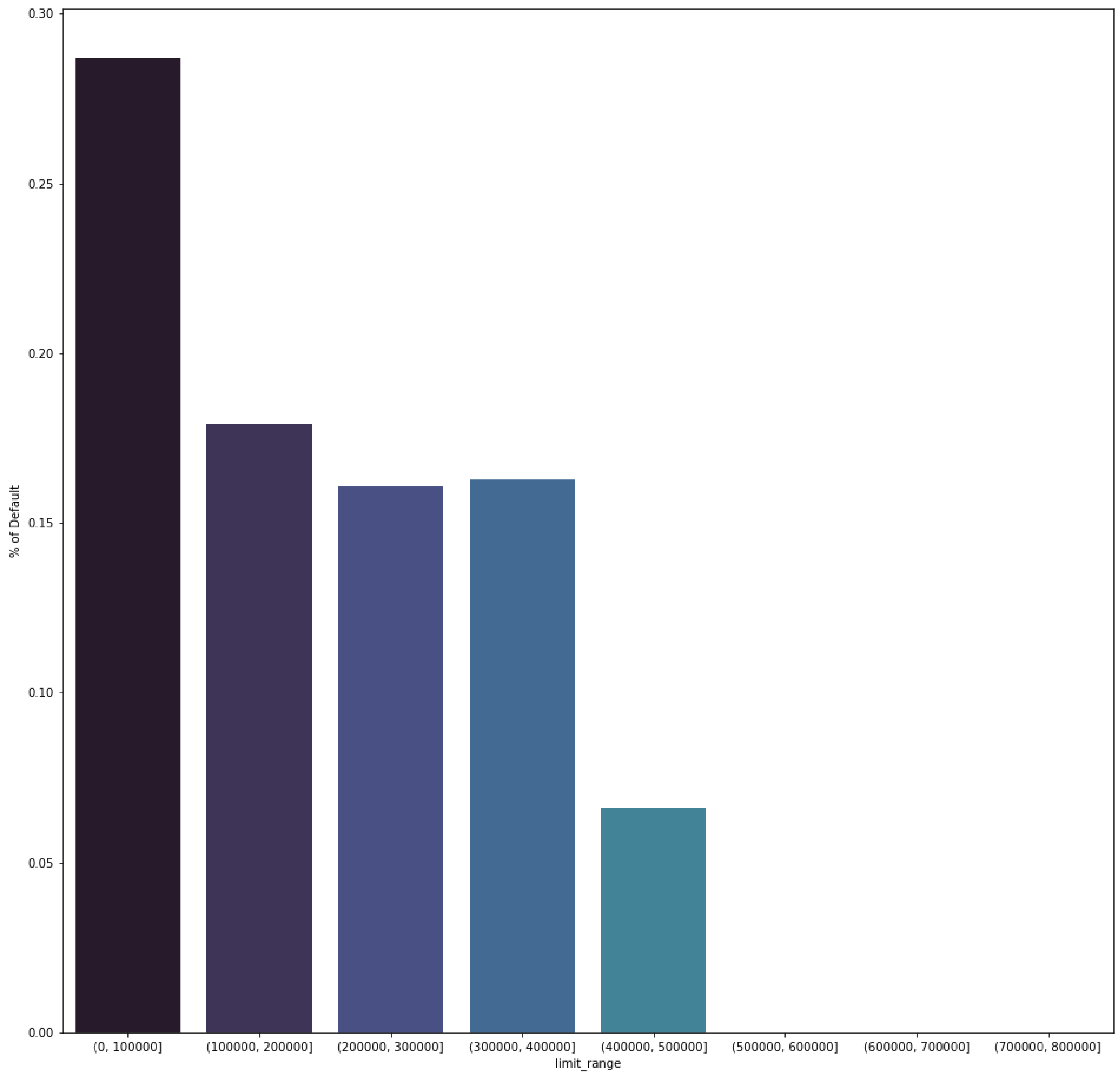
Based on Status

Married customers are more likely than others to default their payment



Based on Age

Customer with age between 41-50 are more likely to default



Based on Credit Limit

Customer with low limit are more likely to default

Data Pre-Processing

Splitting Train and Test

Split dataset to train and test with 85:15 proportion

Feature Selection

Select feature for the model based on correlation value and personal judgement

Handling Outliers

Cleaning the data from extreme value

Feature Scaling

Equalize the range of numerical column to help model train better

The Machine Learning Models

- Logistic Regression
- Suppor Vector Machine
- Decision Tree
- Random Forest
- K-Nearest Neighbours
- Naive Bayes
- Gradient Boosting

Model Evaluation Metric

Model	Train Accuracy	Test Accuaracy	Recall 0 Test	Recall 1 Test
Logistic Regression	0.83 ± 0.02	0.86 ± 0.06	0.97	0.40
Support Vector Machine	0.84 ± 0.02	0.82 ± 0.04	0.96	0.46
Decision Tree	0.75 ± 0.02	0.77 ± 0.10	0.82	0.45
Random Forest	0.81 ± 0.02	0.82 ± 0.06	0.92	0.42
K-Nearest Neighbours	0.81 ± 0.01	0.84 ± 0.06	0.94	0.42
Naive Bayes	0.81 ± 0.03	0.82 ± 0.06	0.91	0.49
Gradient Boosting	0.83 ± 0.02	0.83 ± 0.04	0.95	0.45

Model Improvement

After we choose one model that have better performance,. Then we make improvements by tuning the model's hyperparameters using GridSearchCV function

Metric	Before	After
Accuracy	0.846	0.851
Precision	0.86 & 0.73	0.87 & 0.74
Recall	0.95 & 0.45	0.95 & 0.47
F1 Score	0.91 & 0.56	0.91 & 0.58

Overall Summary

- The performance of Gradient Boosting model is still far from expectations, especially in accurately predicting default customers (Recall value : 0.47)
- Mistakes in predicting default customers to be non-default will lead to huge financial loss and will harm the company
- For non-default customer predictions, the model's performance is quite good (Precision value : 0.87).
- For now this model maybe useful for the marketing team. It can help identify our non-default customer and offer them another products we have.

Next Steps

Try resampling the target data to make them balance. Maybe it will improve the model performance.

Perform hyperparameter tuning for other models.

Using PCA for dimensional reduction/feature selection.

For more details on this project, please visit:
https://github.com/H8-Assignments-Bay/p1---ftds-001-hck--ml1-dimitriasta