



# Default of Credit Card Clients

YIMING PAN, LIN WANG, SHUJUN YE, YIMING YU, CHUN YUAN, JINGLIN ZHAO  
BUAN 6340 GROUP 3  
SPRING 2019

# History of Credit Card Reporting

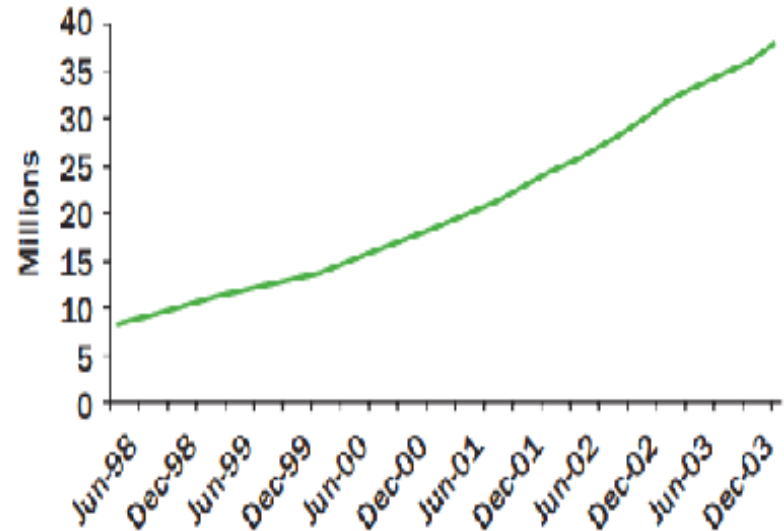
\$28.84 trillion global credit card market.

In 1973, First credit card was issued in Taiwan.

Has a rapid expansion from year 2000.

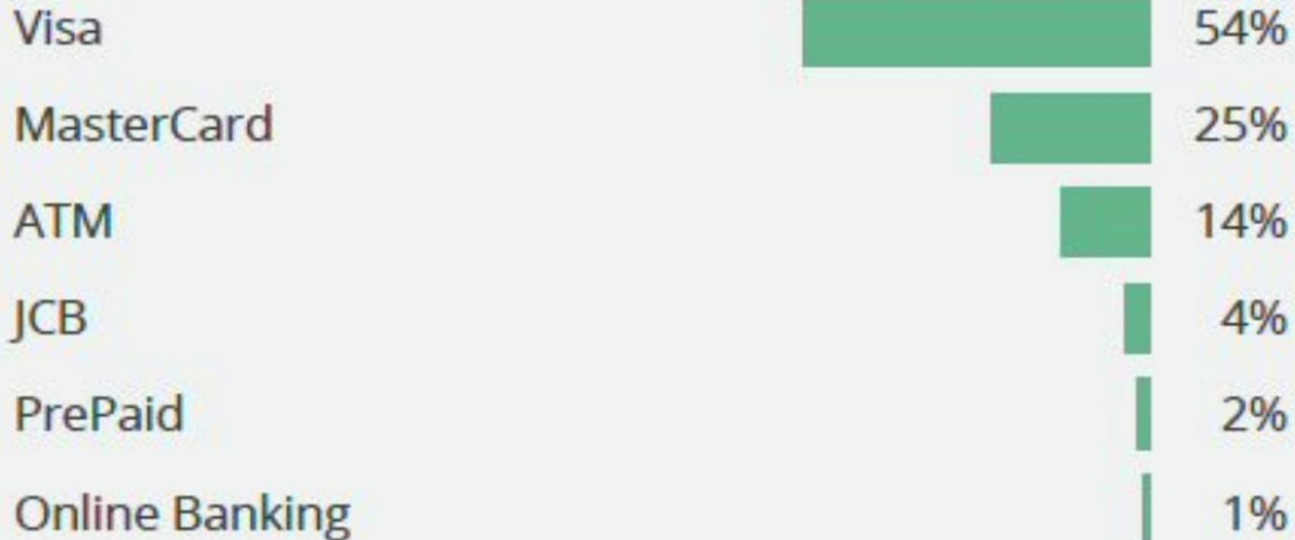


**Chart 1** Number of credit cards in Taiwan



# History of Credit Card Reporting

Top Online Payment Method in Taiwan



# History of Credit Card Reporting

- ↓ 49 credit card banks located in Taiwan.
- ↓ Over 45 million credit cards were issued in Taiwan.

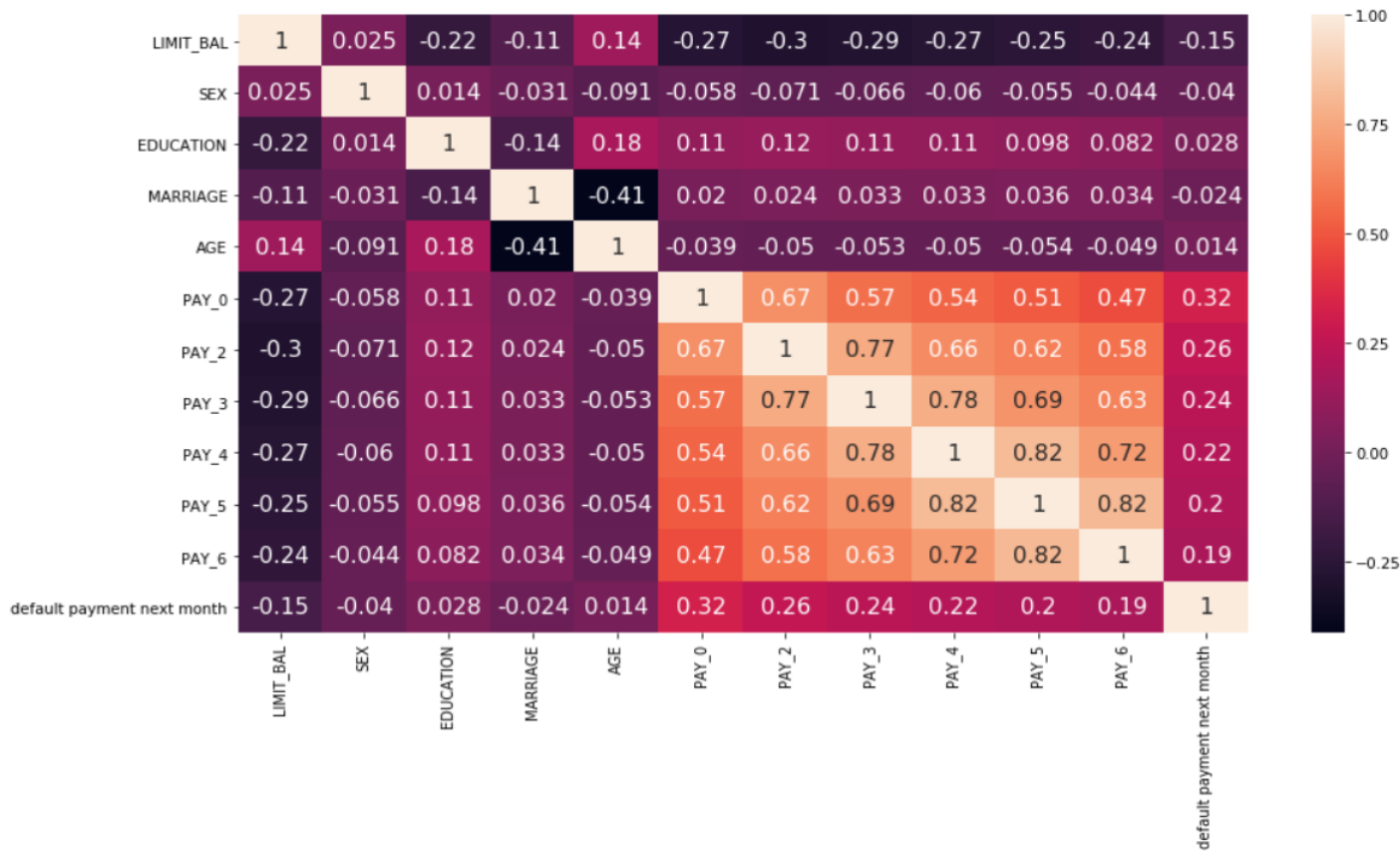
# Program Background

- ↓ Investigated credit card default issues.
- ↓ Explored the factors influencing the default of credit card clients.
- ↓ Determined how much each factor contribute to the default credit card clients.
- ↓ Proposed a model for predicting default of credit card clients

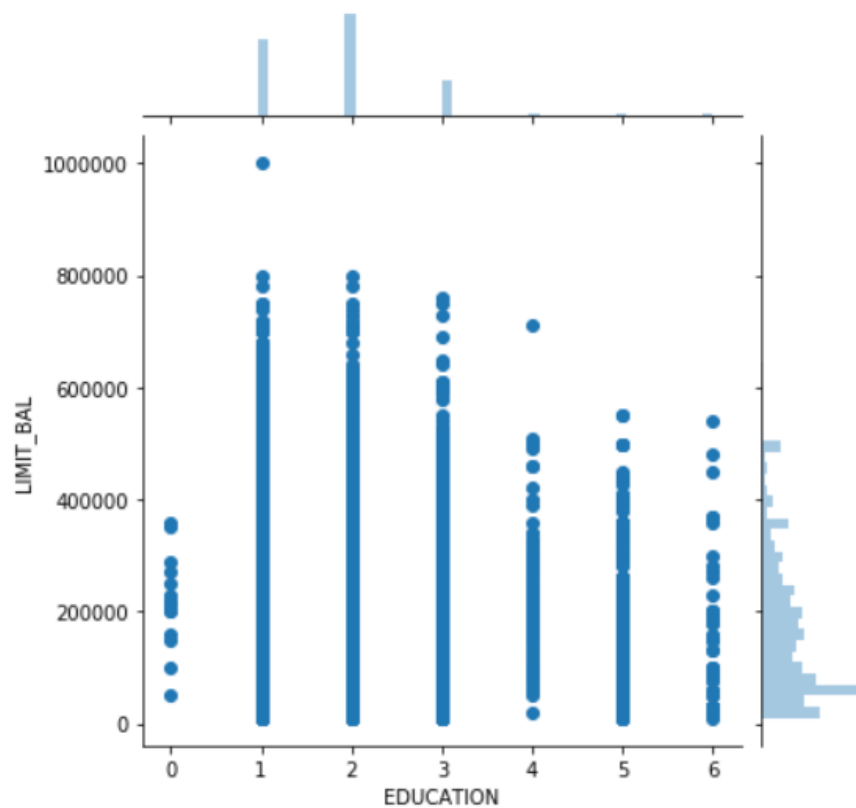
# Data Exploration

<b>ID</b>	ID of each client
<b>LIMIT_BAL</b>	Amount of given credit in NT dollars (includes individual and family/supplementary credit)
<b>SEX</b>	Gender (1=male, 2=female)
<b>EDUCATION</b>	(1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
<b>MARRIAGE</b>	Marital status (1=married, 2=single, 3=others)
<b>AGE</b>	Age in years
<b>PAY_0-6</b>	Repayment status in September-April, 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months,...)
<b>BILL_AMT1-6</b>	Amount of bill statement in September-April, 2005 (NT dollar)
<b>PAY_AMT1-6</b>	Amount of previous payment in September-April, 2005 (NT dollar)
<b>default.payment.next.month</b>	Default payment (1=yes, 0=no)

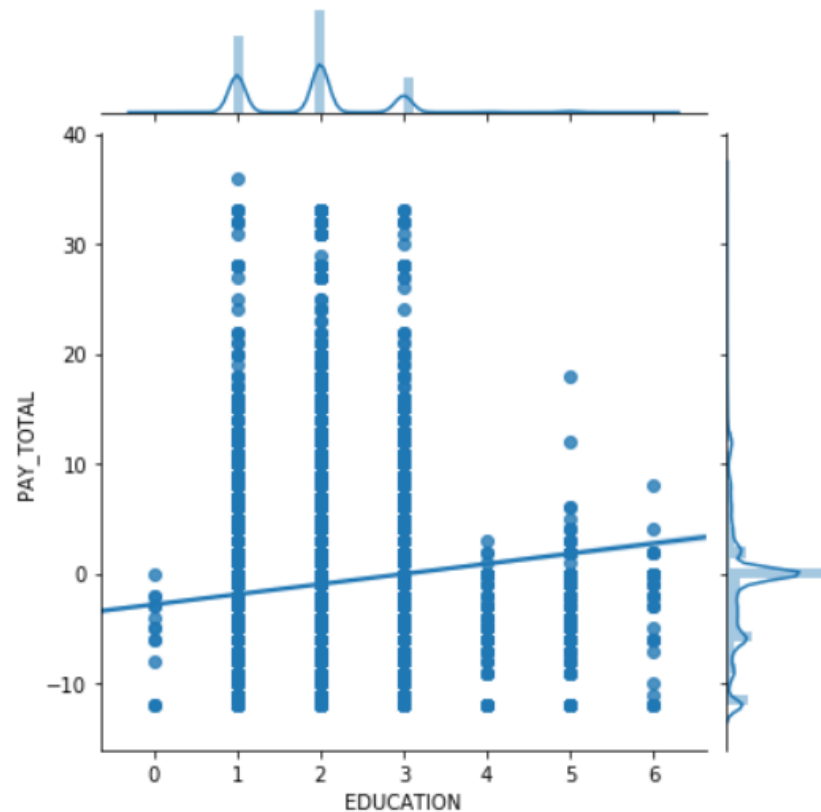
# Correlation Heatmap



## Education with Limit\_Bal

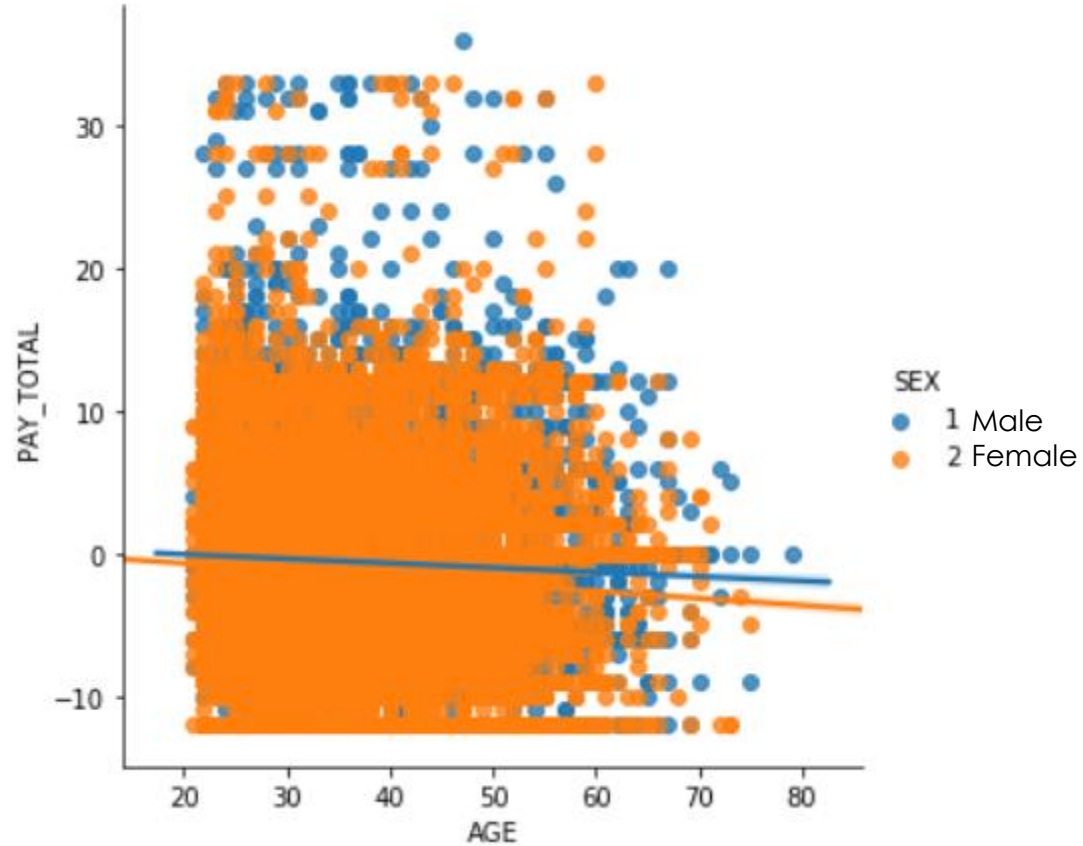


## Education with Pay\_Total

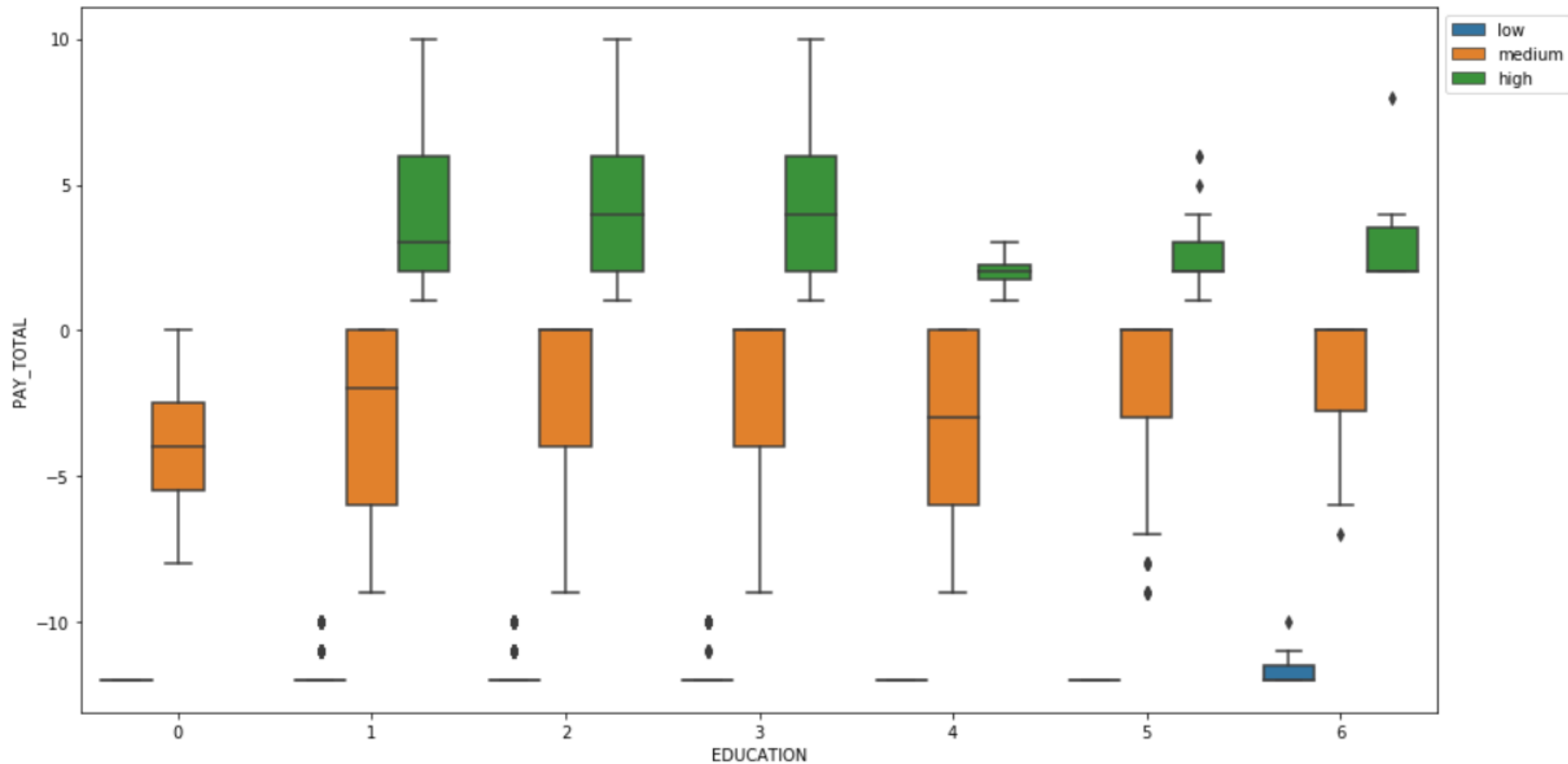




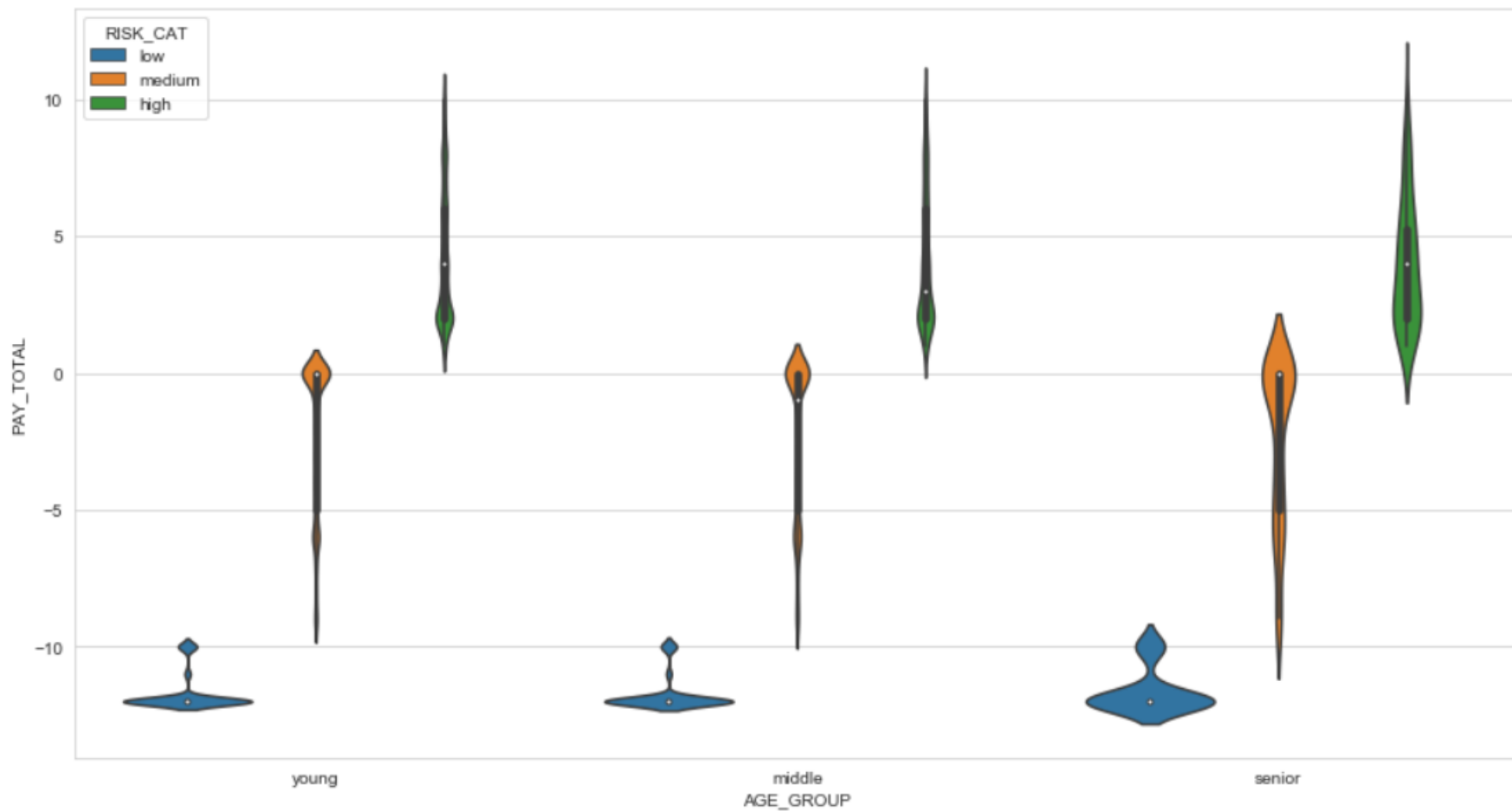
## Age/Sex with Pay\_Total



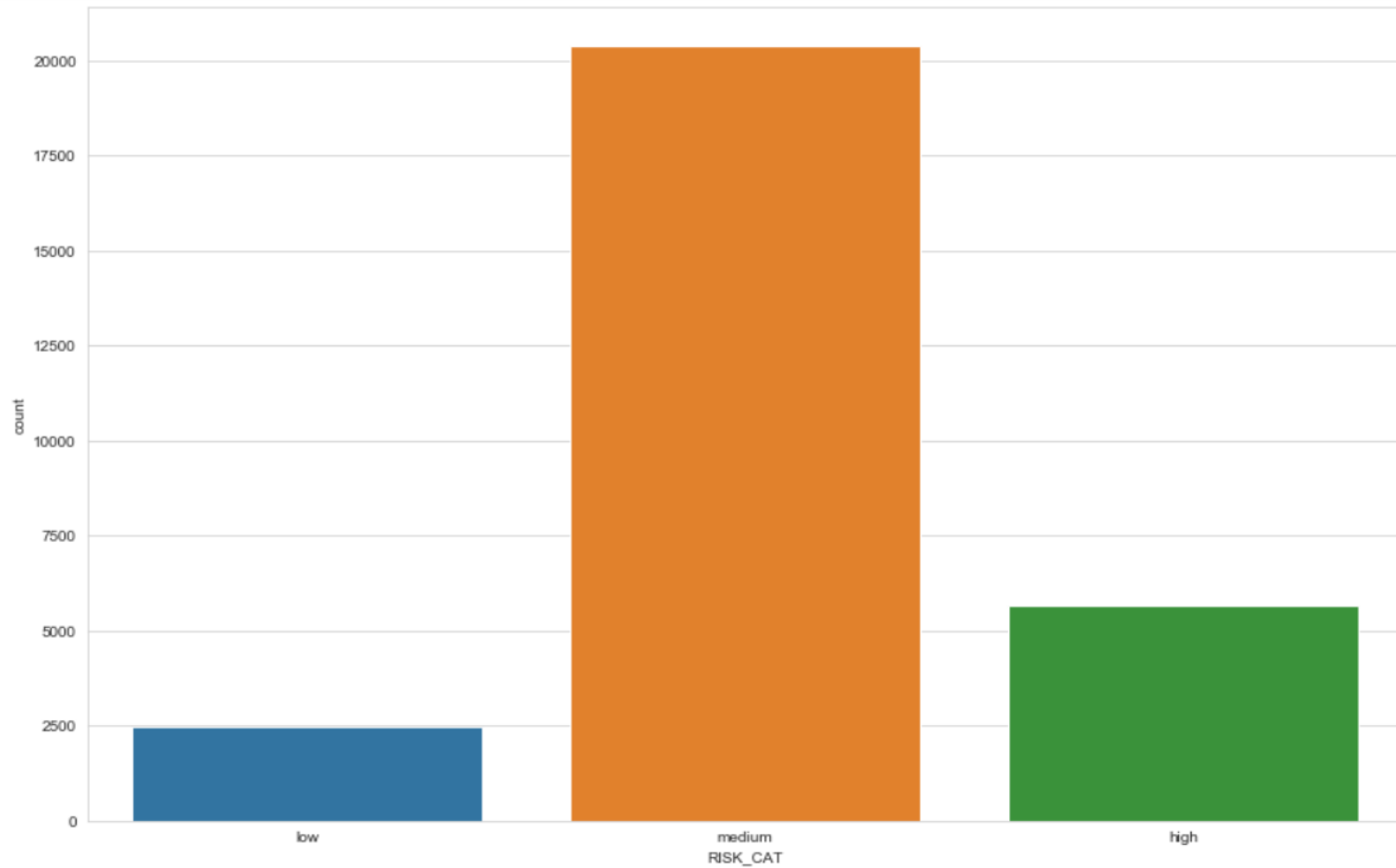
# Risk Group Distribution Towards Education



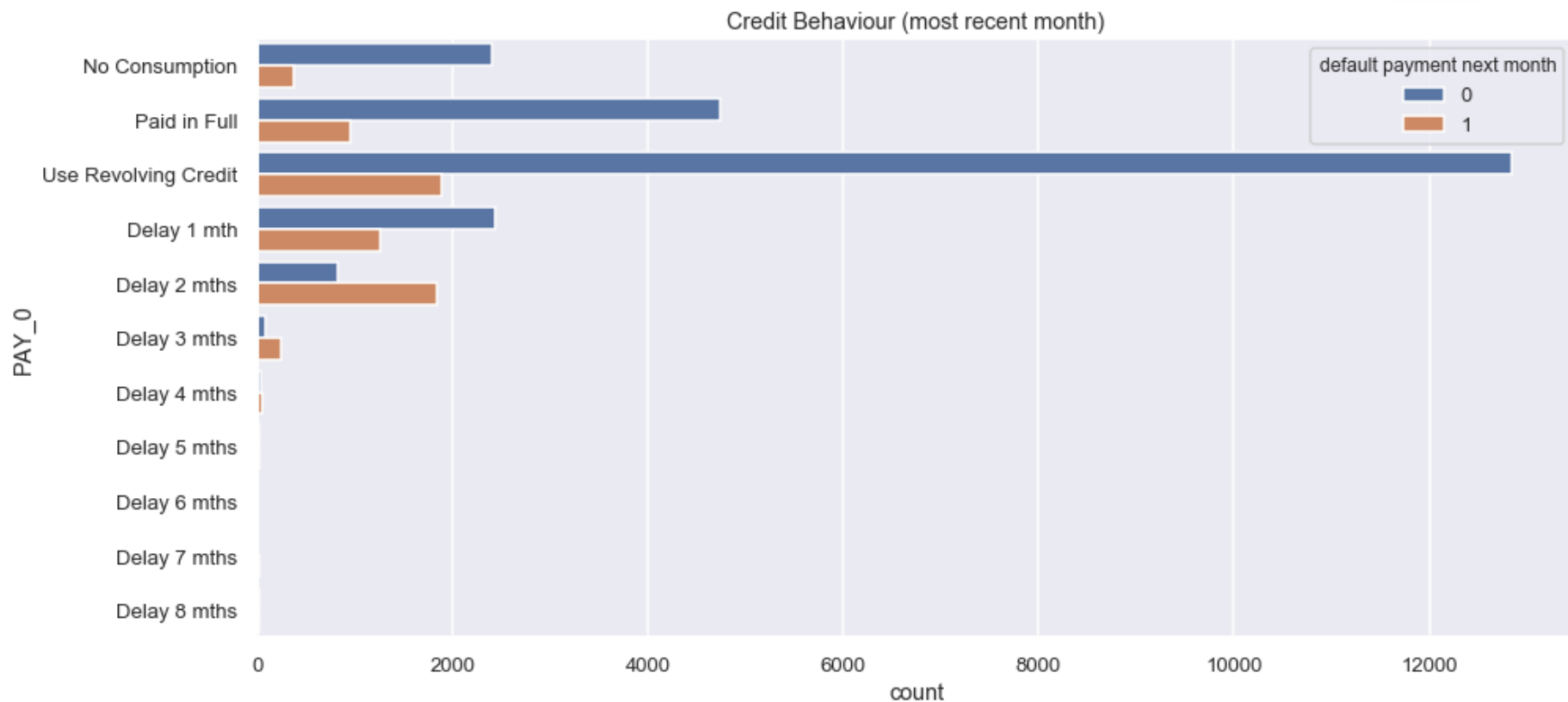
# Risk Group Distribution Towards Age



## Count of different risk group



# Default With Credit Behaviour



# Data Cleaning and Preprocessing

- ↓ Missing values
  - ↖ Zeros representing missing values
- ↓ Categorical values
  - ↖ Ordinal : EDUCATION, PAY\_n
  - ↖ Non-ordinal : MARRIAGE, SEX
- ↓ Unexpected values
  - ↖ Categorical values not provided in the data description
  - ↖ Combine? Keep?
- ↓ Train-test split

# Resampling



Why resampling?



Our dataset is imbalanced



Very low recall score

```
X_train_org['default'].value_counts()
```

```
0    18661
```

```
1     5339
```

```
Name: default, dtype: int64
```



Solution



Create 2 new training datasets:



Upsampled training data



Downsampled training data

```
1    18661
```

```
0    18661
```

```
Name: default, dtype: int64
```

```
1     5339
```

```
0     5339
```

```
Name: default, dtype: int64
```

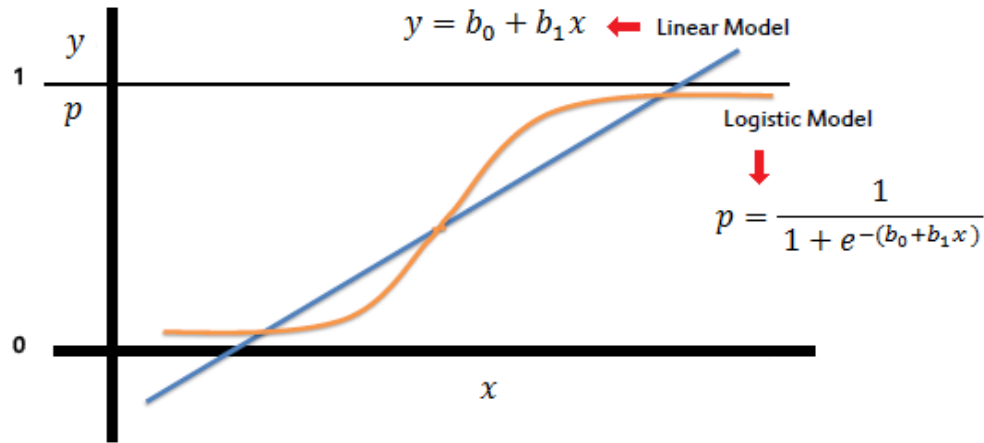
# Models and Analysis

Logistic Regression  
SVM  
Decision Tree  
Random Forest  
KNN  
Gradient Boosting Tree



# Models and Analysis

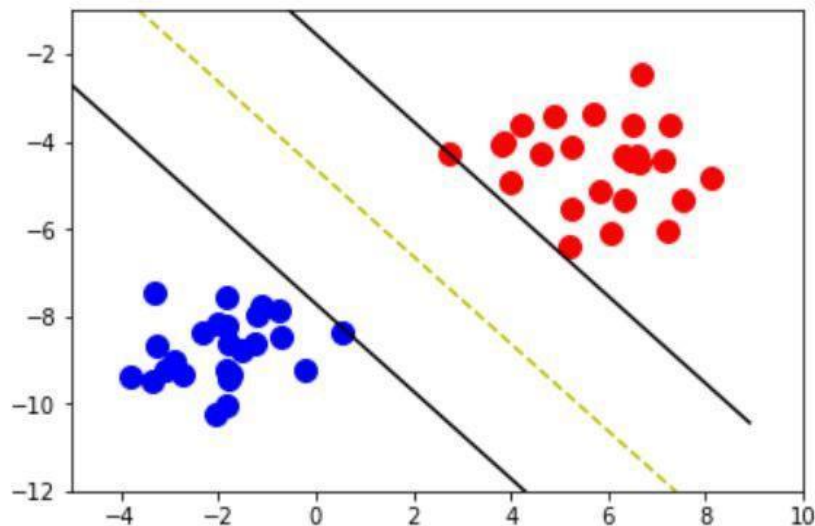
## Logistic Regression



- Simplest parametric model in classification
- Take the linear combination and apply a sigmoid function (logit)

# Models and Analysis

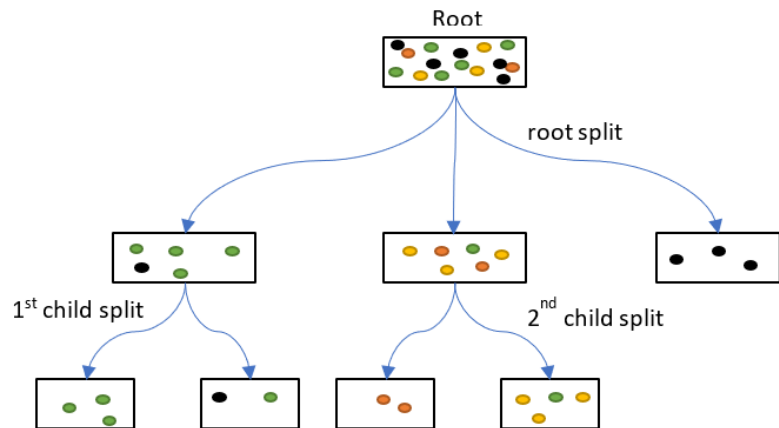
SVM



- Maximize the distance from the yellow line (decision boundary) that separates the data
- Black lines are support vectors that used to determine the decision boundary
- Can be used to classify non-linear relationship

# Models and Analysis

## Decision Tree

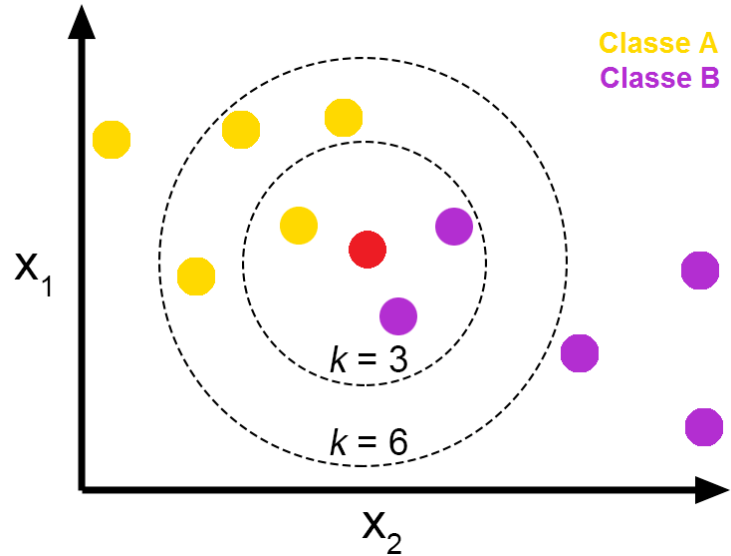


## Random Forest

- Ensemble method
- Construct multiple trees for prediction
- Prevent overfitting by including randomness

# Models and Analysis

KNN



- Learn based on how similar is a data from another
- Find the  $k$  nearest neighbors based on the distance and assign the class based on majority vote

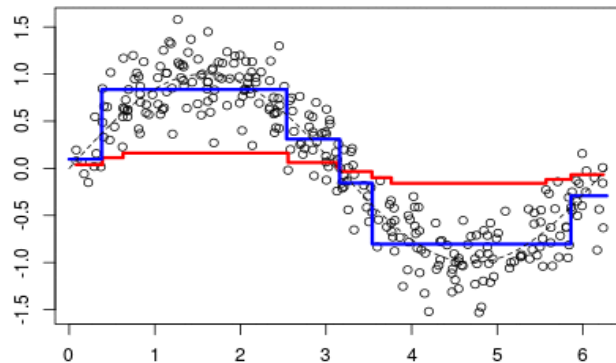
# Models and Analysis

## Gradient Boosting Tree

- Ensemble method
- Builds a sequence of decision tree models and each model learns errors from the previous model

### XG Boost

- A library for Gradient Boosting framework
- Powerful in implementing gradient boosting algorithm

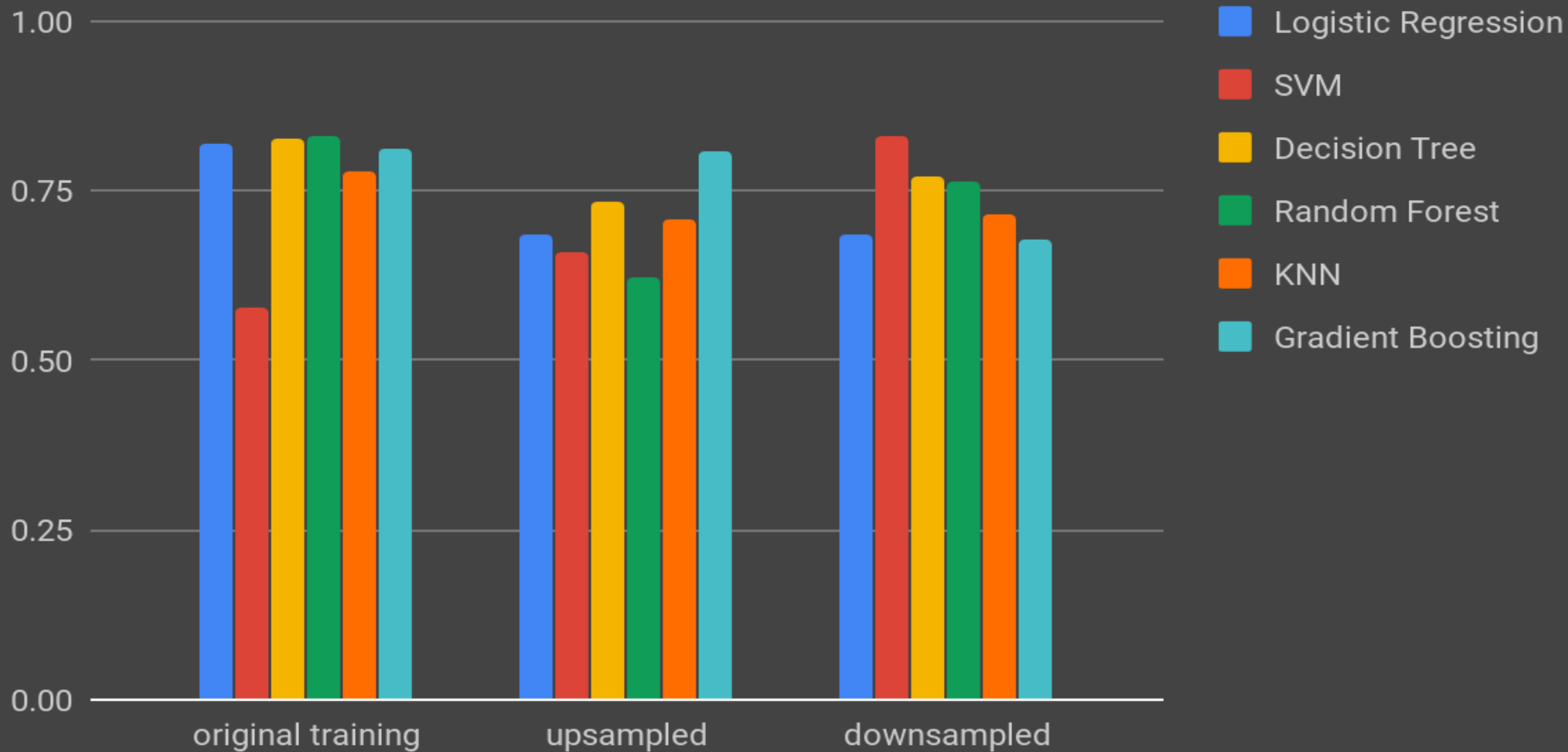


<https://freakonometrics.hypotheses.org/527>

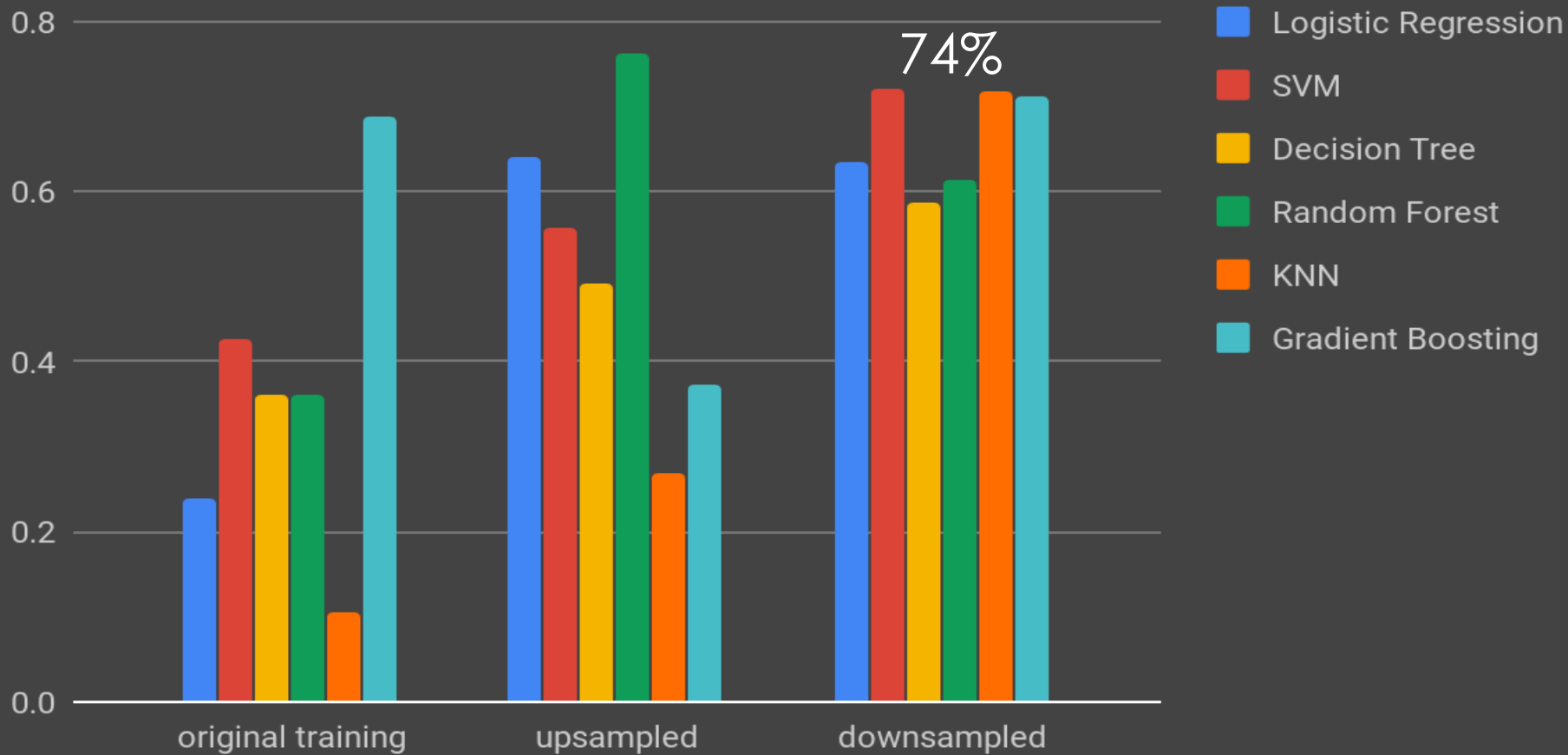


# **Models Performance**

# Accuracy



# Recall





# Model Selection

Gradient Boosting

KNN

SVM

# For further learning

# Other Model Ensembling Methods

# Recommendations



Focus on clients using revolving credits, users only make minimum payments every month

Push notification of payment remind to those clients who are predicted to be with late payment



Check your credit card accounts frequently

Check your credit report at least once per year to monitor credit history

# Reference

Ismail, S., Amin, H., Shayeri, S. F., & Hashim, N. (2014). Determinants of Attitude towards Credit Card Usage. *Jurnal Pengurusan*, 41, 145–154. Retrieved from <http://search.ebscohost.com.libproxy.utdallas.edu/login.aspx?direct=true&db=bth&AN=99747154&site=ehost-live>

Lee, Y.-H., & Huang, Y.-L. (2011). Do you have credit cards? The expansion of the credit card market in Taiwan. *Applied Economics Letters*, 18(17), 1639–1644. <https://doi-org.libproxy.utdallas.edu/10.1080/13504851.2011.556586>