## **About Dataset**

## **Dataset Information**

This dataset contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card clients in Taiwan from April 2005 to September 2005.

## **Content**

There are 25 variables:

* ****ID****: ID of each client
* ****LIMIT\_BAL****: Amount of given credit in NT dollars (includes individual and family/supplementary credit
* ****SEX****: Gender (1=male, 2=female)
* ****EDUCATION****: (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
* ****MARRIAGE****: Marital status (1=married, 2=single, 3=others)
* ****AGE****: Age in years
* ****PAY\_0****: Repayment status in September, 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months, … 8=payment delay for eight months, 9=payment delay for nine months and above)
* ****PAY\_2****: Repayment status in August, 2005 (scale same as above)
* ****PAY\_3****: Repayment status in July, 2005 (scale same as above)
* ****PAY\_4****: Repayment status in June, 2005 (scale same as above)
* ****PAY\_5****: Repayment status in May, 2005 (scale same as above)
* ****PAY\_6****: Repayment status in April, 2005 (scale same as above)
* ****BILL\_AMT1****: Amount of bill statement in September, 2005 (NT dollar)
* ****BILL\_AMT2****: Amount of bill statement in August, 2005 (NT dollar)
* ****BILL\_AMT3****: Amount of bill statement in July, 2005 (NT dollar)
* ****BILL\_AMT4****: Amount of bill statement in June, 2005 (NT dollar)
* ****BILL\_AMT5****: Amount of bill statement in May, 2005 (NT dollar)
* ****BILL\_AMT6****: Amount of bill statement in April, 2005 (NT dollar)
* ****PAY\_AMT1****: Amount of previous payment in September, 2005 (NT dollar)
* ****PAY\_AMT2****: Amount of previous payment in August, 2005 (NT dollar)
* ****PAY\_AMT3****: Amount of previous payment in July, 2005 (NT dollar)
* ****PAY\_AMT4****: Amount of previous payment in June, 2005 (NT dollar)
* ****PAY\_AMT5****: Amount of previous payment in May, 2005 (NT dollar)
* ****PAY\_AMT6****: Amount of previous payment in April, 2005 (NT dollar)
* ****default.payment.next.month****: Default payment (1=yes, 0=no)

## **Inspiration**

Some ideas for exploration:

1. How does the probability of default payment vary by categories of different demographic variables?
2. Which variables are the strongest predictors of default payment?

## **Acknowledgements**

Any publications based on this dataset should acknowledge the following:

Lichman, M. (2013). UCI Machine Learning Repository [[http://archive.ics.uci.edu/ml]](http://archive.ics.uci.edu/ml]" \t "/Users/sarojrai/Library/Containers/com.kingsoft.wpsoffice.mac.global/Data/tmp/wps-sarojrai/x/_blank). Irvine, CA: University of California, School of Information and Computer Science.

The original dataset can be found [here](https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients" \t "/Users/sarojrai/Library/Containers/com.kingsoft.wpsoffice.mac.global/Data/tmp/wps-sarojrai/x/_blank) at the UCI Machine Learning Repository.