**Dataset Information of Default of Credit Card Clients**

**Source**: <https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients>

**Citation**: Yeh, I-Cheng. (2016). default of credit card clients. UCI Machine Learning Repository. <https://doi.org/10.24432/C55S3H>.

**Summary**: 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, 24 features + 1 label, 30k records in total.

**Demographic**:

1. *ID*: ID of each client – Numeric
2. *Sex*: Gender – Categorical
3. *Education* – Categorical
4. *Marriage* – Categorical
5. *Age* – Numeric

**Credit Card**:

1. *Limit\_Bal*: Given credit in NT dollars (family-based) – Numeric
2. *Pay\_0/2/3/4/5/6*: Payment status in 09/2005, 08/2005, 07/2005, 06/2005, 05/2005, 04/2005. – Categorical
3. *Bill\_Amt\_1/2/3/4/5/6*: Amount of bill statement in 09/2005, 08/2005, 07/2005, 06/2005, 05/2005, 04/2005. – Numeric
4. *Pay\_Amt\_1/2/3/4/5/6*: Amount of previous payment in 09/2005, 08/2005, 07/2005, 06/2005, 05/2005, 04/2005. – Numeric

**Label**:

1. default.payment.next.month

**Exploration Ideas**:

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?

**Data Pre-processing & Feature Selection**:

1. No missing values
2. Removed ID
3. Feature Selection – Correlation-based
4. 77.88% negative