Problem Statement:

Delinquency rate on credit card debt for all commercial banks in U.S. was 2.54% in 2018. If banks can correctly predict which customer is going to default, then banks can take preventive measures (like reducing the credit limit, closing the account or proactively reaching out to the consumer for financial advice, etc.,) to reduce the number of defaults.

Data:

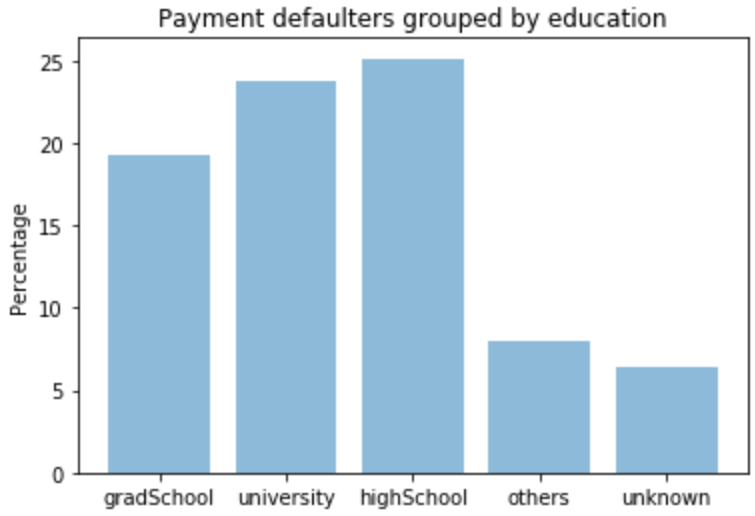
The data is available at https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset. This data is collection of customer data in a Taiwan bank. The data is donated by the Department of Information Management, Chung Hua University, Taiwan.

The data is checked for integrity, missing values and outliers and found that there is no missing value. I did not find any data integrity issues as well. Due to the nature of the financial data, it's not possible to determine the outliers. Almost every column can have wide range of values.

Exploratory Analysis:

Following observations are made about the data:

- Balance limit range from 10,000 to 1 million!
- Customers' age range from 21 to 79.
- In the dataset given, 28% defaulted on the payment. That's much larger than 2.54% delinquency rate on credit card debt for all commercial banks in U.S.
- Interestingly less educated people were more prompt in paying credit card bills!



- Customers who were late in payment in April 2005 were late in September 2005 as well. This may lead to early detection of default and influence bank staff to take preventive measures to reduce the delinquency rate and loss.

- Using pearson correlation coefficient, it's proven that the payment amount for previous month is correlated with repayment status for current month.
- The above finding is confirmed thru heatmap as well. We can conclude that repayment status is the most reliable indicator of defaulting on payment next month.