Credit Card Risk Control
Of
ABC Bank

By

Saheli Mukherjee

Problem Statement

The ABC Bank has over-issued its cash and credit card to increase its market share in 2005. They did not classify the applicants and issued credit cards even to the unqualified applicants. At the same time, most cardholders, irrespective of their repayment ability, overused credit card for consumption and accumulated heavy credit and cash – card debts. The crisis caused the blow to consumer finance confidence and it is a big challenge for both banks and cardholders.

The analysis focuses on estimating the probability of default to control the risks. It deals with the outstanding payment amount and other variables. The analysis comes under the backdrop of the 2008 financial crisis that hit the banking sector badly jeopardizing other sectors as well across different countries in the world.

Data Overview

The dataset belongs to the time period of April 2005 through September 2005. It has some technical terms that need to be understood clearly before framing the research questions and hypotheses.

- LIMIT_BAL: It includes the amount of the credit given to an individual and their family members.
- SEX: Sex is classified under male (1) and female (2).
- EDUCATION: The markers are as follows,

    1 = graduate school;

    2 = university;

    3 = high school;

    4 = others.
- MARRIAGE: 1, 2 and 3 signify the marital status to be married, single and other respectively.
- AGE: This is the age of the credit card holders.
- PAY_1 to PAY_6: It is the past monthly payment records where 0 means duly paid, 1 means that the payment was delayed for 1 month, 2 for 2 months, and, so on and so forth till the 8th marker. 9 does not exist in the data.
- BILL_AMT1 to BILL_AMT6: The amount of the bill statement represents the total amount owed on a credit card at a specific point in time. This balance is the amount that the cardholder is required to pay by the due date to avoid incurring interest charges or late fees. In the given dataset, the amounts are arranged for the months from April to September 2005 in descending

order. BILL_AMT1 represents the bill statement of September 2005 whereas BILL_AMT6 represents the same of April 2005.

- PAY_AMT1 to PAY_AMT6: The amount of the previous payment refers to the sum of money that a credit card holder has paid towards their outstanding balance in a previous billing cycle. It provides insight into the cardholder's payment history and financial behaviour. In the dataset, the amounts are given similar to the columns of bill statements i.e. the months from April to September 2005 are arranged in descending order.

- DEFAULT PAYMENT NEXT MONTH: Default payment next month refers to a binary indicator in credit risk analysis, particularly in the context of credit card payments. It signifies whether a credit card account is expected to default on its payment in the next billing cycle. Default (yes) is represented with 1 and non-default (no) with 0.

- Amount of outstanding payment (OP) is the amount obtained by subtracting the amount of previous payment (PP) from the amount of bill statement (BS) for each customer.

$$OP = BS - PP$$

Assumptions

Some assumptions have been made for the ease of analysis. Those are as follows:

- The PAY_0 column title has been renamed to PAY_1 aligned with other column titles, assuming that an incorrect title was entered.

- 1 marker is missing under PAY_5 and PAY_6. It is assumed that no payment was delayed for one month i.e. payment was either duly paid or was delayed for more than one month for the columns concerned.

- PAY_1 to PAY_6 does not contain the marker 9, because of which, it was assumed that no payment was delayed for 9 months and above.

- All the monetary values are in USD.

- The data belongs to the finance category where data deviations are commonly found. Therefore, it is assumed that no outliers exist for the sake of data integrity. Entire dataset has been considered for the analysis.

Research Questions

The following research questions have been framed in order to estimate the probability of default based on various factors.

- Compare the trends of bill statement and previous payment over the months.
- What are the distributions of outstanding payment amount with respect to sex, age, education and marriage?
- Is there any correlation between credit limit and outstanding amount?
- Does outstanding amount affect the default payment next month?
- If previous payment amount increased by 10% for each month, what changes would have been there with respect to outstanding amount?

Hypotheses

- Male customers have more outstanding amounts than the female ones.
- Older customers are more prone to having greater outstanding amount than their younger counterpart.
- Customers with less education level have more outstanding amount.
- Unmarried male customers are burdened with outstanding amount.
- There is a positive correlation between credit limit and outstanding amount i.e. higher credit limit leads to higher outstanding amount.
- Outstanding amount has no relationship with the default payment next month.

Analyses and Findings

The dataset has been cleaned before the final analysis. Some points to be noted are:

- 1 marker is missing under PAY_5 and PAY_6. It is assumed that no payment was delayed for one month i.e. payment was either duly paid or was delayed for more than one month for the columns concerned.
- EDUCATION column has three invalid values: 0, 5, 6 which have been omitted.
- MARRIAGE column has one invalid value: 0. Rows having the same value has been removed from the dataset.
  Total 29,601 datapoints have been retrieved after the data cleaning out of 30,000 samples given.

Trend analysis

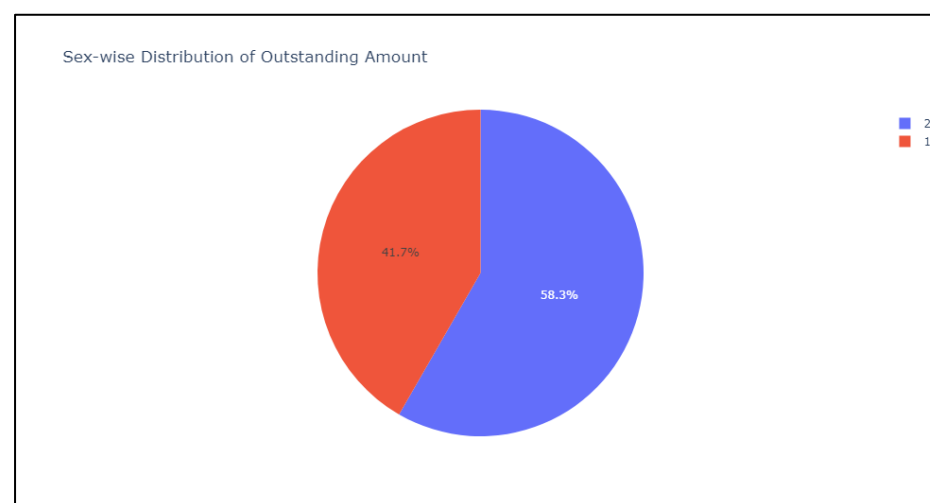The multiple line chart showcases the bill statement, previous amount and outstanding amount from the month of April 2005 to September 2005. The blue, green and red lines are representing bill statement, previous payment and outstanding amount over the given time period. Bill statement gradually rose from around $1.2B in April 2005 to $1.5B in May 2005. The payments remained more or less stable throughout the months suggesting that the customers were not able to or were not ready to pay their credit due. The previous payment amounts never went above $0.2B, August being the receiver of the highest amount amongst all. As a result, the outstanding amount followed the trend of the bill statement gradually increasing from one month to the other beginning from $1B to $1.3B (approx.).

The following analyses have been done based on the data points which have the outstanding amounts more than 0. Positive value of the outstanding payment renders the amount that a debtor owes the bank and the negative value means the debtor overpaid the bank.

Outstanding amount based on sex

1 represents male while 2 represents female. Interestingly, male customers had less credit dues (41.7%) while the females had 58.3%. It is clear from the chart that the female customers were affected more than the male counterparts.

Female debtors had less understanding

of the credit market that led them to default. Further analysis on education, marital status would throw light on this fact. ***Therefore, it cannot be concluded that the male customers had more outstanding amounts than the female ones.***

Relationship between age and outstanding amount

The age of the customers ranges between 21 and 79 years. Pearson's correlation gives the value of
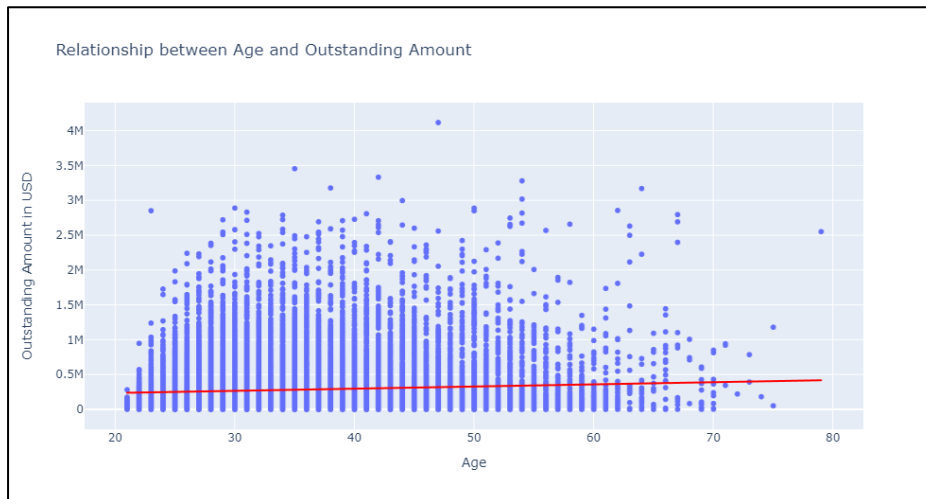
0.076 resulting in a weak positive relationship between the age of the customers and their outstanding amounts. The scatterplot shows the same pattern. ***Therefore, it can be said that the older customers are more prone to having greater outstanding amount than their younger counterpart although the relationship could be weak.***

The relationship can be dissected further with two different approaches. Older debtors who were new in the credit market might have started defaulting due to their poor knowledge of how the market works. On the other hand, the debtors who were experienced enough in the market could not predict the fall coming. The precursor of the 2008 crisis might have caused the loss of jobs of many customers as well. Detailed data on customers, for example, their occupation status, number of dependent family members can provide more vivid picture in this regard. There is a scope for further research to understand the relationship between age of the customers and their outstanding amount.

Sex-wise distribution of percentage outstanding amount WRT education level

The markers 1, 2, 3 and 4 represent graduate school, university, high school and other education level. It is surprising to see that the customers having the university level education defaulted more, females (more than 50 percent), being the maximum in number. Graduate school also shows the high amount of outstanding payment and, in this case, male customers defaulted heavily. ***The data is quite surprising and cannot be inferred that the customers having less education level had more outstanding amount***.

The result could be attributed to the situation where customers, having higher education level, had
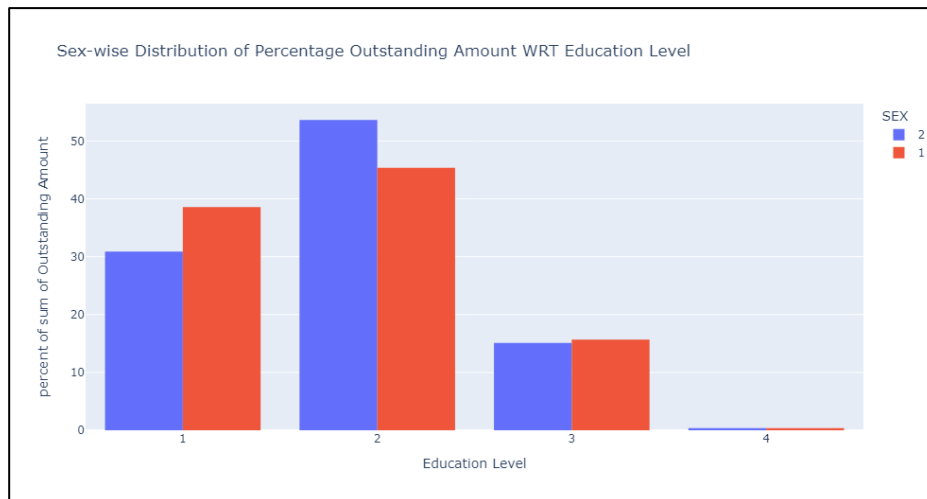

taken more credits for their consumption thinking that they could pay the amount. However, that overburdened them later on. Meanwhile, customers with less education level avoided depending on credit for consumption. Here, the importance of finance market awareness should not be underestimated. Females with university level education, belonging to the middle- and high-income groups of the society who have high consumption level, might have gone on spending money on consumer goods without the awareness of the coming financial crisis. Simply put, female customers with higher education level had enough money in hand to spend on consumer goods. They continued spending until and unless they started defaulting due to the approaching financial crisis. They could not repay the credit amount since they spent on consumption only.

Sex-wise distribution of percentage outstanding amount WRT marital status

1 denotes married customers while 2 represents unmarried ones while others belong to 3. Therefore,
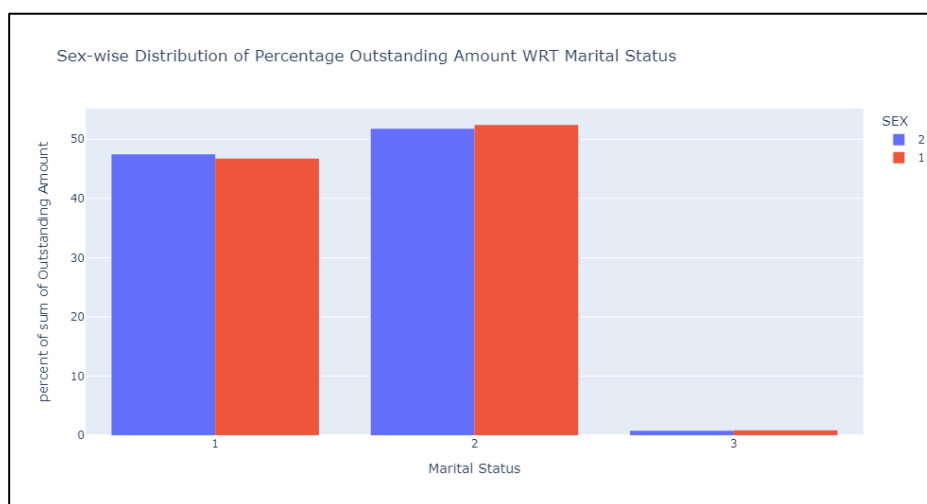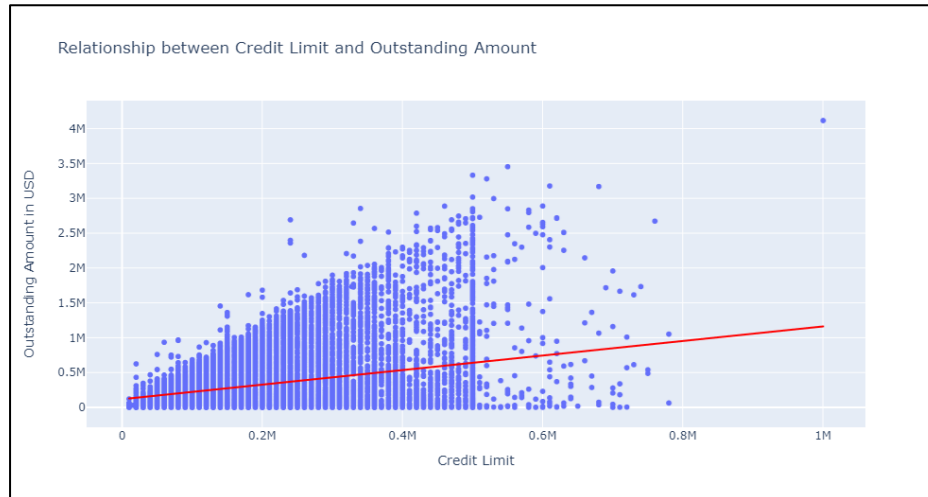

unmarried customers were overburdened with the outstanding payment and both male and female had more than 50 percent of the outstanding payment amount.

Therefore, unmarried male customers had the highest share in the outstanding payment. The next highest group includes the married customers where females were having slightly higher amount of the outstanding payment. It can be comprehended that the married customers were concerned about their family welfare and had taken less risk of using credit risk whereas the unmarried ones did not

calculate the risks too much. However, this must also be noticed that the difference between the married and the unmarried customers is negligible.

Correlation between credit limit and outstanding amount

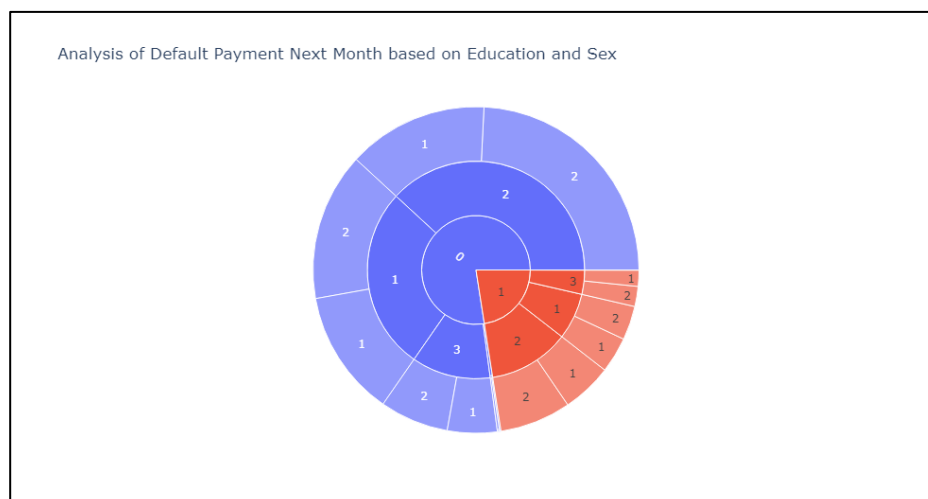The credit limit ranges between $0.1M and $1M. Pearson's correlation gives the value of 0.36 resulting in a moderate positive relationship between the credit limit and the outstanding amounts. The scatterplot shows the same pattern. ***Therefore, there is a positive correlation***

between credit limit and outstanding payment amount i.e. higher credit limit leads to higher outstanding amount. The bank did not check on the capability of the debtors to repay and high credit limit led to the uncontrolled expenditure on consumer goods that was hard, for the bank, to recover from the defaulters.

Analysis of default payment next month

The centre-most ring represents the probable default payment for the next month. Although the probability of default was less (since number of noes were more than yeses) this analysis focuses on the share of the default payment (1). The second ring of the sunburst diagram

shows that the customers with university level education had a more chance to default next month and the third ring indicates that both male and female credit card users would have had the equal probability of failing to repay the bill. ***Therefore, the outstanding amount does not affect the default***

payment next month. This diagram is completely in sync with the previous analysis and a specific set of characteristics of the defaulters could be identified. All the analyses regarding sex, education level and marital status draw such inference that unmarried female customers having higher education level were more prone to defaulting. They had taken the risk of the credit market without apprehending the future financial market crisis. Moreover, using credits on consumption made the situation worse since such expenditure did not return anything in terms of monetary gains.

Effect of 10 percentage increase in previous payment amount on outstanding amount

10 percent increase in previous payment would result in 9 percent decrease in the outstanding
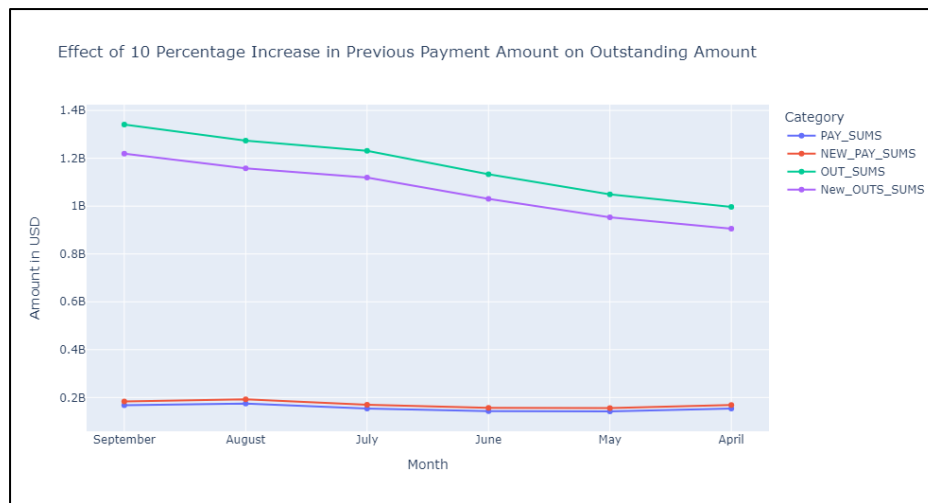
amount. The figure presents the amounts under consideration in USD. Interestingly, it is vivid that a small increase in previous payment amount would decrease the outstanding amount to a significant level. Therefore, the more the previous payment amount increases, the greater the outstanding amount would decrease. This inference can be used to select the ideal potential customers in the future who would have enough flexibility to increase their monthly payment amount to a considerable extent.

Inferential analyses

Financial data are prone to having outliers and no exception is there for the current dataset. Removing
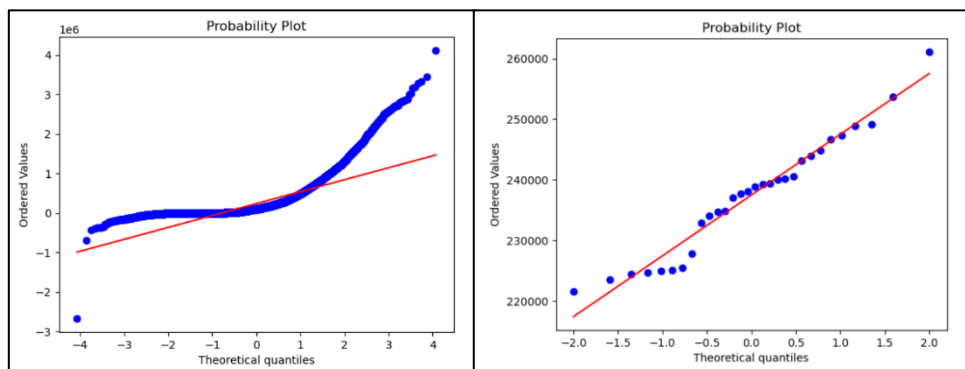
those data points from the dataset, sometimes, erases important insights.

Therefore, all 29,601 data points, in the dataset given, have been taken into account for the analysis. That has been considered to be the population size since there are all the information of the debtors

concerned. From the Q-Q plot on the left-hand-side, it is evident that the dataset does not follow the normal distribution. Therefore, sampling distribution of the mean with the sample size of 30 (Q-Q plot on the right-hand-side) has been considered, as per the central limit theorem, for further analyses using z-statistic. The outstanding amounts parameter of the population range from the negative to the positive values signifying both debits as well as credits on the customers' part. The mean of sampling distribution ranges between USD 0.20 million and USD 0.28 million only inferring that the customers have positive outstanding amount i.e. the dataset has more credits rather than debits.

- **Some Inferences**
  - It can be said, with 95% confidence level, that the outstanding amount of the debtors has a narrow margin of USD 237465.23 and USD 237469.15. Therefore, the chance to default in repaying the credit increased when the outstanding amount of a debtor entered this interval.
  - Considering the sex of the debtors, with 10% stratified random sampling performed on the population), the outstanding amount of the male debtors, at 95% significance level, lies between USD 252704.42 and 252708.34 while that of the female creditors lies between USD 221627.96 and USD 221631.88. Therefore, the default interval of the male is higher than their counterpart although they occupy the lesser share in the total outstanding amount.
  - The probable ranges of the outstanding amount for the married and single debtors have also been calculated at 95% significance level. It is found that the outstanding amount range is at lower level for the single customers compared to the married ones. It is interesting to note that although married customers planned better about their credit expenses than the single ones and they have less share in the total outstanding amount, they actually have a higher level of outstanding amount, on an average.

Marital Status	Lower Level of Outstanding Amount (in USD)	Upper Level of Outstanding Amount (in USD)
Married	234010.18	234014.11
Single	227217.61	227221.53

  - The age has been divided into two groups i.e. below or equal to 50 yrs. and above 50 yrs. The 95% significance level suggests that the customers of the first group defaulted within the interval of USD 221602.84 and USD 221606.76 while the other group defaulted at USD 269463.20; up to USD 269467.12.

Suggestions

- The probability of default is higher in case of female customers with higher level of education, basically, belonging to the middle- and high-income group.

- Unmarried customers show the higher probability to default in repaying the credit bill, however, they, on an average, have lower outstanding amount than their counterparts.

- Age has a moderately weak correlation with the outstanding amount. Therefore, the bank can put a less weightage on this factor while considering a potential customer.

- Bank must take care of its credit card limit since there is a moderate correlation between it and the outstanding amount.

- Bank cannot become risk-averse and its potential customers mostly belong to the middle- and high-income group itself. Therefore, it can organize awareness campaigns to keep the customers aware of its credit card policies, looming risks, pros and cons of using a credit card for consumption. Participation of the female customers must be encouraged.

- Thorough analysis of the credit card market needs to be done to monitor the health of the market and take the actions required.

- Customers having enough earning from one source or having multiple sources of income would have enough flexibility to increase their monthly payment amount to a considerable extent. Such customers could be prioritised.

- Customers' outstanding amount mostly ranges between USD 237465.23 and USD 237469.15. The bank must become more alert when a debtor's outstanding amount enters this figure.