ELLER COLLEGE OF MANAGEMENT
HSLopez School of Business Analytics

ANALYTICS REPORT

TO: JOHN MCCLINTOCK

FROM: JAKE MOORE

SUBJECT: CREDIT LIMIT AND DEFAULT ANALYSIS

DATE: NOVEMBER 12TH, 2025

Introduction

In this brief analysis, we examined a dataset of approximately 24,000 Taiwanese credit card customers to investigate two key issues: how education and demographic factors influence credit limits, and which customer characteristics affect the probability of defaulting on payments. The purpose of this analysis is to provide the company with a better understanding of credit behavior within their customers and default risk patterns so they can strengthen their existing credit policy and identify possible bias.

To address these two questions, we estimated two separate regression models- the Credit Limit Model and a Chance of Default Model. The goal of these models was to identify statistically meaningful relationships.

The results of our analysis concluded that certain demographic groups do receive different credit limits, and that repayment behavior and age show directional relationships with the risk of defaulting. However, both of our models proved to have very low predictive accuracy with limited usefulness for individual level predictions.

Based on these findings, we ultimately recommend that the company should collect more financial information on their customers such as income, credit score, missed payments, and debt-to-income before coming to any conclusions regarding the relationships. Although all variables are statistically significant, they only explained a small portion of overall outcomes. Finally, we recommend that the company pays extra attention when loaning to older customers as they exhibit higher than expected default rates.

Data Analysis

This section evaluates regression models including the Credit Limit Model and the Chance of Default Model to help us better understand how different demographic and financial variables can influence credit outcomes. Each model below includes the full estimated regression equation, coefficient interpretations, and an assessment of the model fit using the R^2 and standard error. This analysis also models two different example predictions using our models to demonstrate how each model performs on realistic cases. This section concludes with key insights, and recommendations for improving the models while acknowledging the potential limitations of their predictive power.

Credit Limit Model

Estimated Sample Regression Equation

$\widehat{Credit\ Limit} = 3352.01 + 363.15(Female_d) - 1141.66(Highschool_d) + 2319.98(Gradschool_d) - 2976.21(Single_d) + 67(Single_d * Age) + 51.15(Age) + \varepsilon$

<u>Model Fit</u>
Coefficient Interpretations
Female(d): Credit limit is $363.15 higher for females than for males, on average and all else constant.
High School(d): High school educated clients have credit limits $1,141.66 lower than university-educated clients, on average and all else constant.
Graduate School(d): Graduate-educated clients have credit limits $2,319.98 higher than university-educated clients on average and all else constant.
Single(d): At zero years old, credit limit would be $2,976.21 lower for single clients than for married clients, on average all else constant.
Age: For married clients, as age of client increases by 1 year, credit limit increases by $51.15, on average and all else constant
Single(d)*Age: For single clients, as age of client increases by 1 year, credit limit increases by $118.15, on average and all else constant.

R^2 Interpretation
The R^2 value of 0.1187 indicates that our model explains only 11.87% of the variability in credit limits, meaning we are 11.87% of the way toward perfectly predicting credit limits using this model. This suggests that our model has very low predictive power.

Standard Error Interpretation
The standard error of $4,029.25 represents the average difference between the observed and predicted values of credit limits. This further proves that this model's predictions are not reliable for individual predictions.

Example Prediction Using the Credit Limit Model
To see how the model could be used to predict, below is an example credit limit prediction for someone 45 years old, married, female, with a university degree:

$\widehat{Credit\ Limit} = 3352.01 + 363.15(1) - 1141.66(0) + 2319.98(0) - 2976.21(0) + 67(0 * 45) + 51.15(45)$
$\widehat{Credit\ Limit} = 3352.01 + 363.15 + 2301.75$
$\widehat{Credit\ Limit} = 6016.91$

Therefore, this model predicts a credit limit of $6,016.91, on average and all else constant. However, this prediction should be taken with caution as the model's standard error is very high at $4,029.25.

<u>**Chance of Default Model**</u>
Estimated Sample Regression Equation

$Prob(\widehat{Default} = 1) = 0.227 + 6.66E\text{-}06(AvgBill) - 0.030(Female_d) + 0.001(Age) - 1.68\text{-}E04(AvgPayment) + \varepsilon$

<u>Model Fit</u>

Coefficient Interpretations

Avg Bill: As average bill amount increases by $100, the chance of defaulting increases by 0.067 percentage points, on average and all else constant.

Female$_{(d)}$: Chance of defaulting is 2.99 percentage points lower for females than for males, on average and all else constant.

Age: As the client age increases by 1 year, the chance of defaulting increases by 0.1 percentage points, on average and all else constant.

Avg Payment: As average payment amount increases by $100, the chance of defaulting decreases by 1.68 percentage points, on average and all else constant.

R^2 Interpretation

The R^2 value of 0.0152 indicates that the model explains only 1.52% of the variability in the probability of default, meaning that we are 1.52% of the way toward perfectly predicting whether a customer will default using this model.

Standard Error Interpretation

The standard error of 0.4118 represents the average difference between the observed and predicted default probabilities, indicating that individual predictions using this model could be off by around 41.8 percentage points. This is a large standard error, and it confirms that the model is not very reliable for making individual-level predictions.

Example Prediction Using the Chance of Default Model

To see how the model could be used to predict, here is a prediction of the chance of defaulting for someone who is 35 years old, male, has an average bill amount of $850, and average payments of $760:

$Prob(\widehat{Default} = 1) = 0.227 + 6.66E\text{-}06(850) - 0.030(0) + 0.001(35) - 1.68\text{-}E04(760)$
$Prob(\widehat{Default} = 1) = 0.227 + 0.005661 - 0 + 0.035 - 0.12745$
$Prob(\widehat{Default} = 1) = 0.14021$ or 14.02%

Therefore, this model predicts a 14.02% chance of default for this customer, on average and all else constant. However, it is important to note the limitations of this prediction given the very low R^2 value of 1.52% (0.0152) and the large standard error of 0.4118.

<u>Recommendations & Insights</u>

Based on the results of our analysis, there does appear to be a gender-based difference in credit limits. Our analysis concluded that females receive credit limits that are $363.15 higher than males, on average and all else constant. This is surprising, as we found that men do have higher average bill amounts, and it seems intuitive that higher spending would correlate with higher limits. However, this relationship alone isn't conclusive enough to state possible bias. We recommend that the model includes other variables that could better evaluate the effect of gender on credit limits. For example, the implementation of income, credit score, current debt, and occupation in our model could help come to a more concrete solution.

Another key insight from this analysis is the limited accuracy of the Chance of Default model. The linear probability model is not reliable for predicting the chance of default for two reasons.

1. Very low R^2 (0.0152) – only explains 1.52% of variation in defaulting outcomes.
2. Large Standard Error (0.4118) – indicates that predictions using the model could be off by as much as 41 percentage points on average.

This means that our model is useful for identifying directional relationships but not for making reliable individual predictions. This is important as the purpose of this model was to predict the chance of default for individuals. We recommend improving this LPM model by adding more key financial information. For example, variables like payment history, missed payments, debt-to-income, and credit score.

Conclusion

In this analysis, we examined a dataset of approximately 24,000 Taiwanese credit card customers to determine how demographic and educational variables relate to credit limits and the chance of defaulting on payments. Our regression results proved that some demographic groups do receive different credit limits, and that repayment behavior and age have directional relationships with default risk.

However, both of our models had very low predictive accuracy. The Credit Limit Model only explained roughly 11% of the variation in credit limits, despite all the variables being statistically significant. The Chance of Default Model explained just 1.52% of default outcomes with a very large standard error.

Based on these findings, we recommend that the company collects more key financial information on their customers. Variables like income, credit score, debt-to-income, and missed payments could greatly improve these models and allow decision makers to make more decisions. Finally, we advise paying more attention to older customers as they tend to have higher-than-expected default rates.

Please feel free to contact me at jakemoore@arizona.edu if you have any questions or would like to discuss these recommendations in more detail

Technical Appendix

Figure 1 – Credit Model Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.344463625
R Square	0.1187
Adjusted R Square	0.1184
Standard Error	4029.2474
Observations	24289

ANOVA

	df	SS	MS	F	Significance F
Regression	6	53072948612.8629	8845491435.4771	544.8464	0
Residual	24282	394214262296.1250	16234834.9517		
Total	24288	447287210908.9880			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	3352.01306	186.57601	17.96594	1.04976E-71	2986.3126	3717.7136
Female (d)	363.154	53.39832	6.80085	1.06415E-11	258.4901	467.8181
High School (d)	-1141.6563	75.14602	-15.19251	6.87015E-52	-1288.9471	-994.3655
Grad School (d)	2319.9777	57.90556	40.06485	0	2206.4792	2433.4762
Single (d)	-2976.2146	235.00820	-12.66430	1.2178E-36	-3436.8452	-2515.5841
Single (d) * Age	67.0037	6.39897	10.47102	1.33091E-25	54.4613	79.5461
Age	51.1548	4.44282	11.51405	1.34598E-30	42.4466	59.8630

Jake Moore

Figure 2 – Chance of Defaulting Model Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.1231
R Square	0.0152
Adjusted R Square	0.0150
Standard Error	0.4118
Observations	24289

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	6.336E+01	1.584E+01	9.343E+01	5.367E-79
Residual	24284	4.117E+03	1.695E-01		
Total	24288	4.181E+03			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.2270	0.011481871	19.77031597	2.56307E-86	0.204495034	0.249505383
Avg Bill	6.660E-06	1.34395E-06	4.955718654	7.25481E-07	4.026E-06	9.29443E-06
Female (d)	-0.030	0.005436	-5.501417162	3.80551E-08	-0.040563653	-0.019252258
Age	0.001	0.000290	3.04020407	0.002366696	0.000313779	0.001452565
Avg Payment	-1.68E-04	9.2316E-06	-18.16339033	3.07501E-73	-0.000185772	-0.000149583

Jake Moore