Credit Rating Prediction:

I. What independent (predictor) variables are relevant to predicting credit score? Create a table with three columns for predictor, expected sign of effect, and a one-sentence rationale for effect. Note that all variables provided in the data set may not influence credit score. If some variables are not relevant, explain why they should not influence credit score.

Credit Score: A Credit Score, typically between 300 – 850 is a three-digit number which represents one's likelihood t pay their bills on time. It is based on credit history which includes number of open accounts, history of repayment, debt level etc. (Created by Fair Isaac Corporation also known as FICO for short).

Factor Effects:
Positive Effect: If there is a direct proportionality (+X then +y / -X then -y) between the predictor variable (X) and the response variable (y), we can say there is a positive effect.

Negative Effect: If there is an inverse proportionality (+X then -y / -X then +y) between the predictor variable (X) and the response variable (y), we can say there is a negative effect.

Relevant Factors		
Predictor	**Effect**	**Explanation**
Income	Positive	Income is one of the important factors as it dictates the expenditure capacity of a person.
Limit	Positive	Credit limit and score are directly proportional. So, having a higher limit means higher spending and the ability to repay amount on time.
Cards	Positive/Negative	Having more cards can increase total available limit, which can contribute a lot to the credit score. Inverse is also true.
Age	Positive	Credit age is considered vital in predicting credit score. An older person will have a higher credit age than a younger person.
Student	Negative	Student status usually no effect on credit rating but ongoing/unpaid Student Loans can have a negative impact.
Balance	Positive	Credit card balances have direct impact on the credit scores. Higher balance means more capacity to take debt.
Irrelevant Factors		
Predictor	**Effect**	**Explanation**
ID	No Effect	Used for sorting
Gender	No Effect	Due to legal restrictions, gender is not factored in while calculating credit score.
Education	No Effect	Education is not considered when calculating the credit score.
Married	No Effect	Most credit bureaus do not include marital status in their records.
Ethnicity	No Effect	Credit scores are race neutral.

II. Run TWO alternative regression models in R to predict credit score. Copy and paste the R code for the two models and the combined output of both models using stargazer.

```
#1st linear model
m1 <- lm(Rating ~ Gender + Education + Married + Ethnicity, data = cr)

#2nd linear model
m2 <- lm(Rating ~ Income + Cards + Age + Student + Balance, data = cr)
```

```
#Stargazer
stargazer(m1, m2, type='text', single.row = TRUE)
```

```
## ===============================================================
##                                Dependent variable:
##                          -------------------------------------
##                                       Rating
##                             (1)                 (2)
## -------------------------------------------------------------
## GenderMale              -2.896 (15.548)
## Education               -1.578 (2.493)
## MarriedYes              13.987 (16.072)
## EthnicityCaucasian      -11.923 (19.169)
## EthnicityHispanic       -21.860 (22.071)
## Income                                      2.064*** (0.041)
## Cards                                        -0.439 (0.920)
## Age                                          0.125* (0.074)
## StudentYes                                 -98.422*** (4.345)
## Balance                                      0.234*** (0.003)
## Constant                380.502*** (38.490) 144.096*** (5.096)
## -------------------------------------------------------------
## Observations                  400                 400
## R2                           0.005               0.974
## Adjusted R2                  -0.008               0.974
## Residual Std. Error (df = 394)  155.319           24.965
## F Statistic (df = 5; 394)     0.390            2,986.402***
## ===============================================================
## Note:                            *p<0.1; **p<0.05; ***p<0.01
```

```
stargazer(m1, m2, type='text', ci=TRUE, ci.level=0.95, single.row = TRUE)
```

```
## ==========================================================================================
##                                         Dependent variable:
##                          ---------------------------------------------------------------
##                                                Rating
##                                   (1)                            (2)
## ----------------------------------------------------------------------------------------
## GenderMale              -2.896 (-33.369, 27.578)
## Education               -1.578 (-6.464, 3.308)
## MarriedYes              13.987 (-17.513, 45.488)
## EthnicityCaucasian      -11.923 (-49.493, 25.647)
## EthnicityHispanic       -21.860 (-65.119, 21.398)
## Income                                            2.064*** (1.983, 2.145)
## Cards                                              -0.439 (-2.242, 1.364)
## Age                                                0.125* (-0.020, 0.270)
## StudentYes                                       -98.422*** (-106.938, -89.905)
## Balance                                            0.234*** (0.228, 0.240)
## Constant                380.502*** (305.064, 455.940) 144.096*** (134.108, 154.083)
## ----------------------------------------------------------------------------------------
## Observations                      400                            400
## R2                               0.005                          0.974
## Adjusted R2                      -0.008                         0.974
## Residual Std. Error (df = 394)   155.319                        24.965
## F Statistic (df = 5; 394)        0.390                      2,986.402***
## ==========================================================================================
## Note:                                             *p<0.1; **p<0.05; ***p<0.01
```

III. Based on your analysis, answer the following questions:

- **What is the marginal effect of income on credit score?**
  The effect of Independent variable on the Dependent variable is called as Marginal effect (Also known as the β – Coeff). It is calculated as $\frac{\Delta y}{\Delta x}$. Income has a positive effect on rating. So, an increase in income by $1000 could lead to an increase in credit rating by 2.064. Standard error (0.041) in this case is low as well which means our population mean

- **Does having more credit cards help or hurt credit score? By how much?**
  For each addition of a credit card, the credit rating could decrease by 0.43. But the standard error is 0.920, which means the credit score can vary between -2.24 and 1.36 (At 95% Confidence Interval at 2SD). Based on this, we cannot say if the number of credit cards has a positive or negative impact.

- **Does maintaining credit card balances help or hurt credit score? By how much?**
  For an increase in credit balance by $1 the credit rating increases by 0.234. So, credit balance helps the credit score.

- **Is there a racial bias in credit score, i.e., do African-Americans or Hispanic people have less credit score than Caucasian people, if all other factors are equal? If so, by how much?**
  The Hispanic people have lower credit score than the Caucasian people by 9.93 points. But as the standard error is high, the racial data is statistically insignificant.

- **Is there a gender bias in credit score, i.e., do females have lower (or higher) credit score than males? If so, by how much.**
  No, gender bias has no effect on credit score. Based on our model, male's credit score is -2.8 points lower than female's credit scores.