*There are some questions that ask for business implications. Do make sure that your answers are based only on the numerical answers that have been calculated. The main purpose of this assignment is to understand how to base business decisions on data. Avoid the temptation to expand to your conclusions beyond what your data analysis has demonstrated.*

1. A company manager says that the average balance on their credit cards is $500. Do you think that this assertion is justified? Use a one-sample t-test to draw your conclusion.
2. Is there a difference between men and women as far as average balance is concerned? Use a two-sample t-test to draw your conclusion.
3. Is there a difference between students and non-students as far as average balance is concerned? Use a two-sample t-test to draw your conclusion.
4. It is generally assumed that if there are more credit cards then the balance on the cards will be more. Based on this dataset, do you think this is true? Calculate a correlation coefficient and show a scatter plot to support your answer.
5. Examine whether the following demographic variables influence balance: (a) age, (b) years of education, (c) marital status. For age and years of education, use scatter plots to depict their relationship with balance and calculate the correlation coefficient. For the relationship between marital status and balance, use a two-sample t-test to draw your conclusion
6. “Ethnicity of the cardholder does not matter as far a balance is concerned.” Carry out an analysis of variance (ANOVA) and discuss whether this statement is supported by the data or not.
7. A general principle that credit card companies often follow is to assign a higher credit limit to people with a higher credit rating. Does the data show that this principle is being followed?
8. Run a simple linear regression of balance on the credit limit. (Here credit limit is the X and the balance is the Y). Report the coefficients and the R-squared. Show a scatter plot.
9. Run a simple linear regression of balance (Y) on credit rating (X). Report the coefficients and R-squared. Show a scatter plot
10. Consider your findings in questions 8-9. Discuss business mechanisms to increase or decrease the balance on credit cards. Try to quantify your answers.In this context, focus on possible specific strategies using variables in Q8 and Q9 that the business could adopt to increase the balance on credit cards
11. The credit limit is provided as a consolidated amount for all the credit cards the cardholder has. Run a multiple linear regression of Balance (Y) on Limit and Cards as two X variables. Report the coefficients. Discuss the effect on the balance of (a) increasing the credit limit on the same number of cards and (b) increasing the number of cards without altering the total credit limit.
12. Run a simple linear regression equation with Income as X and Balance as Y. Report the coefficients. Is the coefficient of Income significantly different from zero? What does this say about the effect of income on balance?
13. Based on the equation derived in question 12, what is the estimated balance for a person with an income of USD 100k per year?
14. Based on the dataset, explore the relationship between credit card balance (Y) and (a) Income (b) Age (c) Education (c) Limit, and (d) Rating as X variables? Estimate a multiple linear regression model and report the statistical significance of each of these variables.