Name(s) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Simulated Credit Score Data Case 2

Using the Credit Score Data Excel file available online, answer the following:

1. Basic Stats: What are the average, median and standard deviation of the credit scores?
2. Using the =percentile() function in Excel, determine and interpret which credit scores make up the 95th, 90th, 80th and 50th percentiles.
3. Draw a Histogram of the credit scores (possibly using a PivotTable). Try group sizes of 10, 25, 50 and 100 and see which one looks best. Print a standard column graph of the Histogram. Google images of the FICO Score distribution. Does the shape of your graph resemble (i.e., model) what the internet expects it to look like?
4. Sort the credit score column from smallest to largest. Scroll down the sheet and look at it. Does there appear to be a pattern in pink as you scroll from low to high credit scores?
5. Basic Stat: What percentage of loans are paid off and what percent default, in general? (According to this fake data – Sorry – If anyone knows where I can get real data like this, please let me know.)
6. Now, for the hard part. Choose appropriate ranges for credit scores, then determine the percentage of defaults for each range. The default percentage should decrease as the credit score increases, in general, but it may not be perfect. Choosing the correct number of categories is difficult: Too few categories and the trend may be hard to ascertain; too many categories will lead to small sample size concerns. **Explain** how you made your final choice on the ranges of data/number of categories (or classes).
7. Draw a scatter diagram with credit scores (think carefully about this part) as the x-axis and default percentage as the y-axis. Can you fit a line or curve to it? It shouldn’t be too linear but start there as a first approximation. Luckily, one of the built-in Excel curves should fit it pretty nicely. What is the equation of the line, with y = default percentage and x = credit score?
8. Assume that you can offer different interest rates on loans ranging from 2.5%, 3.5%, 4.5%, 5.5%, and 6.5%. You earn more with a higher rate, but people with good credit will shop around and find the lowest rate they can get. Further assume that for a particular interest rate *p*, you will earn *p*\*$100,000 on the loan *as long as it doesn’t default*. (A 4% loan will earn 4%\*$100,000 = $4,000.) If a loan does default, you will lose $100,000.
	1. Estimate the cutoff points where you will offer each interest rate. Assume that you will only offer a loan if the expected return on the loan is at least $2000 to cover the administrative costs of setting up the loan. How high does someone’s credit score need to be to get a loan at, say, 3.5%? These cutoff points can be determined either by looking carefully at the data or by doing some fancy algebra – good luck!
	2. At what credit score is a borrower too much of a risk for any loan?
	3. Assume you have some money to market your loan products. On which target group would it make the most sense to spend your advertising budget? Support your conclusions with analytic data, and if you need more data to make this decision, identify what it would be.
9. Summarize what you have learned in paragraph or two. What is missing from this model that would make it even better? List a couple of other realistic business decisions that could be analyzed similarly.