	Ideal Score	Score	**Wid Sogata and James Tsai** Lending Club Instructor Notes
Total Points	100	70	
Exceptional Work		0	I think this has good potential for resubmission.
Describe the purpose of the data set you selected. Describe, why is this data important, and how do you know if you have mined useful knowledge from the data set? How would you measure the effectiveness of a good prediction algorithm? Be specific.	10	7	This is a good start, but lacks some critical information about what constitues a "good enough" model. Your classifier won't be perfect, so it makes sense to estimate the costs of confusions. For example, what is the cost of a false positive? A false negative? Which one is preferred?
Describe the meaning and type of data (scale, values, etc.) for each attribute in the data file.	10	9.5	Good. It would be nice to see your decription of some of the categorical attributes that you retained (like grade/ sub-grade).
Verify data quality: Explain any missing values, duplicate data, and outliers. Are those mistakes? How do you deal with these problems? Be specific.	15	14	What could you do about the DTI of 9999%? If it's a mistkae, you should probably throw away all those instances. It seems like you should find outliers like this that are likely transcription errors and throw out the data as needed.
Visualize appropriate statistics (e.g., range, mode, mean, median, variance,counts) for a subset of attributes. Describe anything meaningful you found from this or if you found something potentially interesting. Note: You can also use data from other sources for comparison. Explain why the statistics run are meaningful.	10	7	Good. These visuals exist, but there is very little explanation of them. Mostly theya re one sentence explanations of the type of plot with no conclusions/speculations.
Visualize the most interesting attributes (at least 5 attributes, your opinion on what is interesting). Important: Interpret the implications for each.	15	13	I am giving some credit here for the histograms throughout--though they are never really related back to this deliverable.
Explore relationships between attributes: Look at the attributes via scatter plots, correlation, cross tabulation, group-wise averages, etc. as appropriate. Explain any interesting relationships.	15	13	~Showing the cross tabulations on a log scale for the counts will likely make the bar charts more interprettable across your large dynamic ranges. ~Loved the boxplots, very informative. ~The number of loans grouped by purpose would work better as a bar chart than a table.
Identify and explain interesting relationships between features and the class you are trying to predict (i.e., relationships with variables and the target classification).	10	6.5	~I was not a fan of the table "heat map." Also, correlation compared to a categorical variable is a very odd metric to use, potentially not useful. ~Instead I think looking at only the loans not in good standing to see if they exhibited different groupings would be a much better inidicator.
Are there other features that could be added to the data or created from existing features? Which ones?	5	0	Skipped?