

		Jiang, Pan, Yu	
		Team: HYJHHP	
		Instructor Notes	
		Score	Score Points:08.30
Total Points	Ideal Score	Score	Score Points:08.30
Total Percentage	Ideal Score	Score	Score Points:08.30
Exceptional Work: 7000 implement dimensionality reduction, then visualize and interpret the results.	10	5	Good start on the PCA, but much more can be analyzed and explained.
Give an overview of the dataset. Describe the purpose of the data set you selected (i.e., why was this data collected in the first place?). Why is this data important and how will you know if you have gathered useful knowledge from the dataset? Be specific and use your own words to describe the aspects of the data.	15	7	This is a good start but leaves a lot of information out. There is not description of how the data was collected. You should define what the exact use case of your prediction model would be, how others will use it, and how you use that information to guide about how well it needs to perform. This will also help to guide what evaluation criteria is best. Why is accuracy good for attrition and why 80%? These conclusions should be backed by a discussion.
Load the dataset and appropriately define data types. What data type should be used to represent each data attribute? Discuss the attributes collected in the dataset. For datasets with a large number of attributes, only discuss a subset of relevant attributes.	15	15	
Verify data quality: Explain any missing values or duplicate data. Are those mistakes? Why do these quality issues exist in the data? How do you deal with these problems? Give justifications for your methods (elimination or imputation).	15	15	nice
Choose and visualize distributions for a subset of attributes. Choose any appropriate visualization such as histograms, kernel density estimation, box plots, etc. Describe anything meaningful or potentially interesting you discover from these visualizations. Note: You can also use data from other sources to bolster visualizations.	20	19	High quality. Add a bit more in terms of the observatiosn from the visuals.
Visualize relationships between interesting attributes. Use whichever visualization method is appropriate for your data. Explain any interesting relationships. Important: Interpret the implications for each visualization. (required for 7000 level students): implement dimensionality reduction, then visualize and interpret the results.	25	22	Nice. Need some more discussion of the conclusions from eh visuals, but this eas good.