**Dataset requirements**: Choose a dataset that is mostly ready to be analyzed. That is, it is already in the format of table data. The following requirements should be met:

1. The data includes categorical features (it can also include other forms of data, but must have categories)
2. The data must be 1,000 rows or larger
3. The data is **not** strictly image or text data
4. The dataset should have some prediction task associated with it (*i.e.*, labels to learn)

**Table Data Grading Rubric**

* Business Understanding (**15 points total**).
  + [**15 points**] In your own words, give an overview of the dataset. Describe the purpose of the data set you selected (i.e., why and how was this data collected in the first place?). What is the prediction task for your data and why are other third parties interested in the result? Once you begin modeling, how well would your prediction algorithm need to perform to be considered useful to these third parties?
    - Be specific and use your own words to describe the aspects of the data.
* Data Understanding (**30 points total**)
  + [**15 points**] 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 points**] Verify data quality: Explain any missing values or duplicate data. Visualize entries that are missing/complete for different attributes. 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).
* Data Visualization (**45 points total**)
  + [**20 points**] Visualize attribute distributions. Choose and visualize distributions for a subset of single 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. Visualize at least 5 attributes, at least one categorical and at least one numeric.
  + [**25 points**] Visualize relationships between a subset of attributes. Use whichever visualization method is appropriate for your data. Explain any interesting relationships. **Important:**Interpret the implications for each visualization. Visualize at least three subsets of the attributes.
* Exceptional Work (**10 points total**)
  + You have free reign to provide any additional analyses.
  + One idea (**required for 7000 level students**): implement dimensionality reduction using t-SNE, then visualize and interpret the results. Give an explanation of t-SNE dimensionality reduction methods.