

Part I:

Instructions: Answer each discussion question in your own words. You may use posted resources or other online resources to answer the questions (**cite your sources**). Thoroughly explain your responses with a minimum of one paragraph (3-5 sentences) in length. Be thoughtful. We will discuss the answers in class.

1. Why is it important to have a data destruction policy as part of a data retention plan?
2. Go to an organization (large) of your choice and include a link to their data retention plan. What do you notice about? Describe any elements that you find interesting?
3. What are some of the regulatory regimes that have to be considered when developing a data retention plan?
4. Why is storytelling and narrative so important in communication analysis results?
5. Describe some of the types of ensemble methods in machine learning. What are the advantages and disadvantages of each type?

Part II:

Instructions: Use the attached dataset (**marketing_hw8.xlsx**) to complete the following tasks in Python. Report your answers to the questions on this homework sheet. Include your Jupyter notebook file along with your homework submission, saved as a PDF. Make sure that any graphs you create are quality graphs with a legend (as needed), axis titles, a descriptive title, appropriate ranges (bar graphs start at 0, etc.). They should be able to stand on their own. You may need to relabel some elements (such as replacing 0s and 1s with string labels).

Use Python to answer the following:

1. Select a classification model of your choice as your simple classifier. Model the Has Tried variables. Prepare the data as needed. Include appropriate metrics and visualizations for your base model.
2. Create a bagging model (it can be a general bagging model, or random forest). Include appropriate assessments.
3. Create a boosting model (adaboost, gradient boost, XGBoost are examples). Include appropriate assessments.
4. Create a stacking model. Include appropriate assessments.
5. Compare the results of your ensemble models to the simple classifier. Compare the performances of the various models.