1. Perform combined over and undersampling on the diabetes dataset (use SMOTEENN). Explain how combined sampling works.
2. Perform logistic regression with the new data from question 1.
   1. Comment on the performance of combined sampling vs the other approaches we have used for the diabetes dataset.
3. Create an ROC Curve for the model <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html> and calculate the AUC.
   1. BRIEFLY explain what the ROC Curve shows and what the resulting AUC means.
4. What is outlier detection? Why is it useful? What methods can you use for outlier detection?
5. Perform a linear SVM to predict credit approval (last column) using this dataset: <https://archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+Approval%29> . Make sure you look at the accompanying document that describes the data in the dat file. You will need to either convert this data to another file type or import the dat file to python.

You can use this code, but otherwise you follow standard practices we have already used many times:

**from** **sklearn.svm** **import** SVC

classifier = SVC(kernel='linear')

* 1. How did the SVM model perform?

1. What kinds of jobs in data are you most interested in? Do some research on what is out there. Write about your thoughts in under 400 words.