1. Download the dataset here: <https://www.kaggle.com/sohier/calcofi> and create one DataFrame for the bottle file and another for the cast file. Perform preprocessing on this dataset. You do not need to do all six types of preprocessing we discussed. For this dataset, 2-3 is all you need.
2. Create a correlation matrix to see how the features are correlated. Describe your findings. Pay particular attention to what is correlated with water temperature.
3. Perform linear regression on the dataset to see if you can predict water temperature based on salinity.
4. Find your mean squared error and R2 values. Be sure to indicate what these tell you. You can import these from sklearn.metrics .
5. Create a final plot of the relationship between water temperature and salinity. Be sure to show the values from the dataset in a scatter plot with a trend line that shows the predicted temperature values.
6. Do some research on polynomial regression. Describe what it is any how it works in markdown. This article is useful <https://machinelearningmastery.com/polynomial-features-transforms-for-machine-learning/>
7. Perform polynomial regression on this same dataset. Train and validate your model as you did when you used linear regression. Mean squared error and R2 both apply here.
8. Create a final plot of the relationship between water temperature and salinity. Be sure to show the values from the dataset in a scatter plot with a trend line that shows the polynomial regression predicted temperature values. The resulting trend line should be curved.