	Review	
600	- if we have missing data, why is imputing with the mean or median not the best idea?	
	- we're underestimating the variance of the actual data	
	- we're potentially manually changing the distribution of our data	
800	Say you have a dataset that has 73 columns and 10,000 rows. This datset has that has no column label names, but you know your target variable. What are two things you could do to continue modeling your data without knowing anything about it.	
	You could technically create a correlation matrix and see what columns are correlated with th target	
	you could run pca on it and reduce it the the PC's that account for most of the variance in the model	
	you could use all the features and regularize with lasso (will reduce unhelpful features to zero)	
200	what is an interaction term?	
	I = col1 * col 2	
400	explain what stratified k fold does & how it works	
1000	when you use standard scaler, what exacly does it do to our data?	
	it makes the mean 0 and variance 1	