# **Data Wrangling**

**1. What kind of cleaning steps did you perform?**

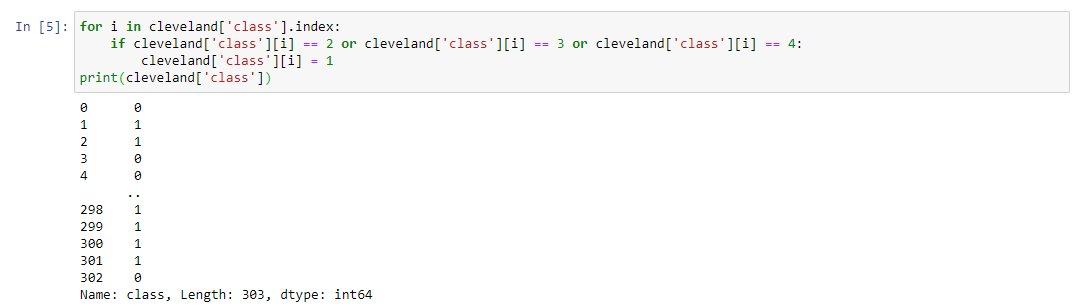
The data looks good and hence no cleaning required as there are no missing values nor there are any text values present in the dataset. Hence no dummy variables required. All the features are of the type float64 except for 'ca' and 'thal' which are object types. I'll concentrate on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0). So, I'll be replacing values 2,3,4 with the value 1.

**2. How did you deal with missing values, if any?**

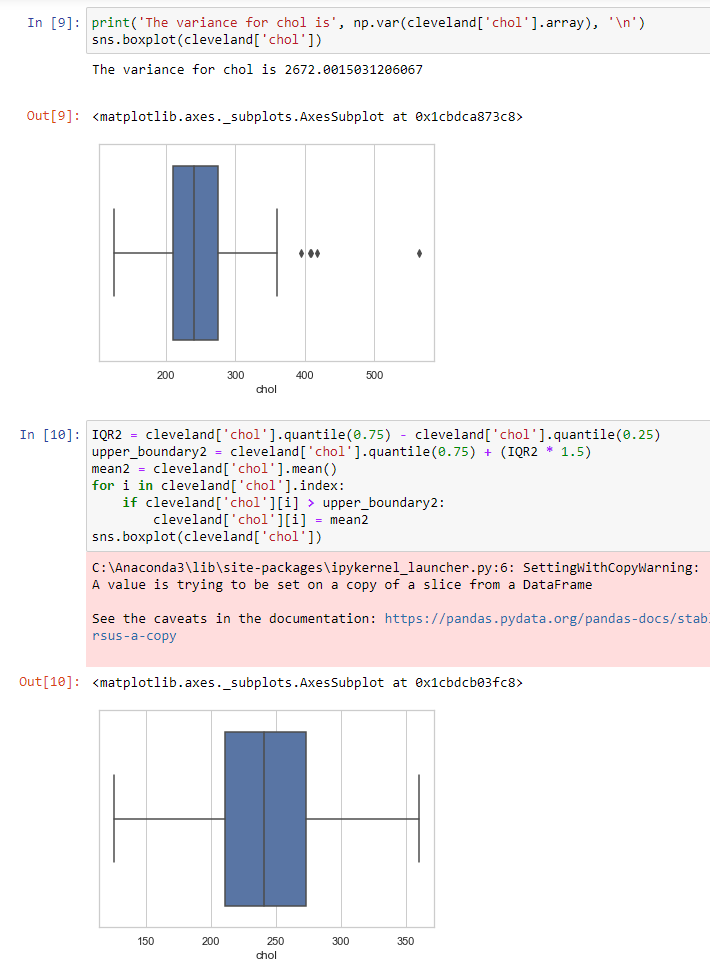
There are no missing values in the dataset.

**3. Were there outliers, and how did you handle them?**

To detect the outliers we plot the boxplot for the quantitative features 'age', 'trestbps', 'chol', 'thalach' and 'oldpeak' as below. No chance of outliers for categorical features.









We can see that features 'trestbps', 'chol', 'thalach' and 'oldpeak' have outliers. We have dealt with the outliers by replacing them with mean.