2.2 Problem 2

Load the auto-mpg sample dataset into Python using a Pandas dataframe. The horsepower feature has a few missing values with a? - replace these with a NaN from NumPy, and calculate summary statistics for each numerical column.

How do the summary statistics vary when excluding the NaNs, vs. imputing them with the mean (Hint: Use an Imputer from Scikit) - can we do better than just using the overall sample mean?

**Answer:**



