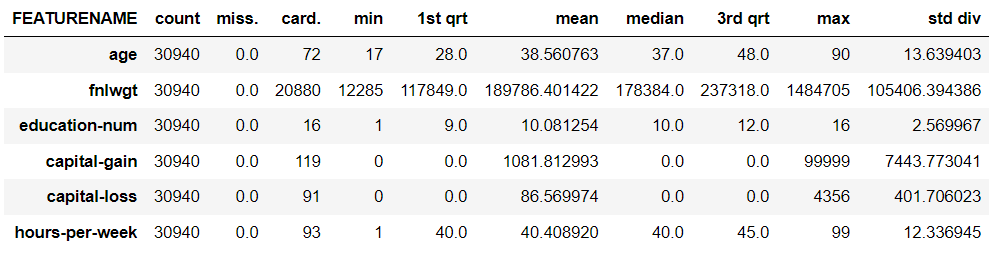
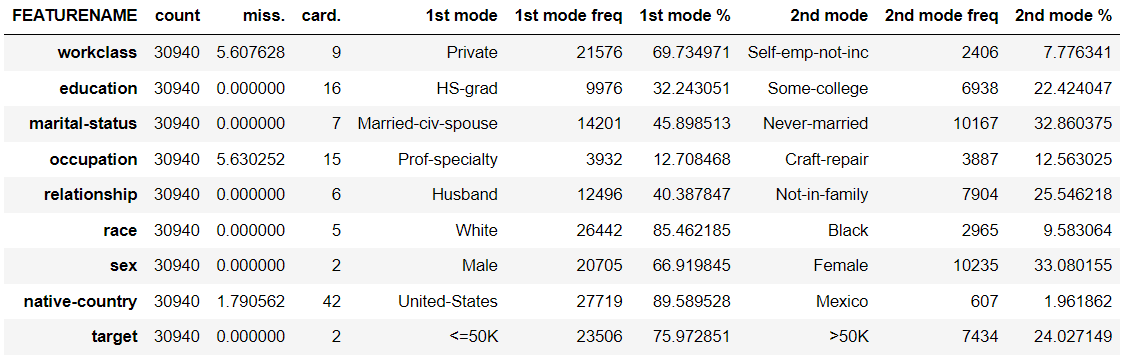
# Continuous Table



# Categorical Table



Above are the two tables that were produced from the data provided. In general, this is a good dataset. There are more than enough entries as seen in the count to be a good representation. For the size there are also every few missing values which again makes it a good representation. Certain info in the continuous table is not as useful and has far outliers such as the min / max of capital gain and not useful data such as the 1st and 3rd quartile of education number. Other sections such as the min / max of capital loss or hours per week are much more useful as they do not have as large outliers and standard deviations. Fnlwgt also has a large cardinality which will need to be considered if it is being used.

Like the continuous table the categorical table has very few missing values. There is also a large mode for most rows which could be very useful when working with the data and trying to reduce the amount of steps to teach an algorithm. Certain values such as the 1st and 2nd modes are not as useful for rows such as the sex as its usually going to be one of the two.

To use the data the most important steps to take before it is used in an algorithm would be to sort out the large cardinality of the fnlwgt rows. Depending on what the use is they might be fine as is but should be taken into account. The outliers for the capital gain should also be removed or handles if they are required. Occupations should be used carefully due to its low mode frequency. There will be many variable options and could increase the length of an algorithm greatly.