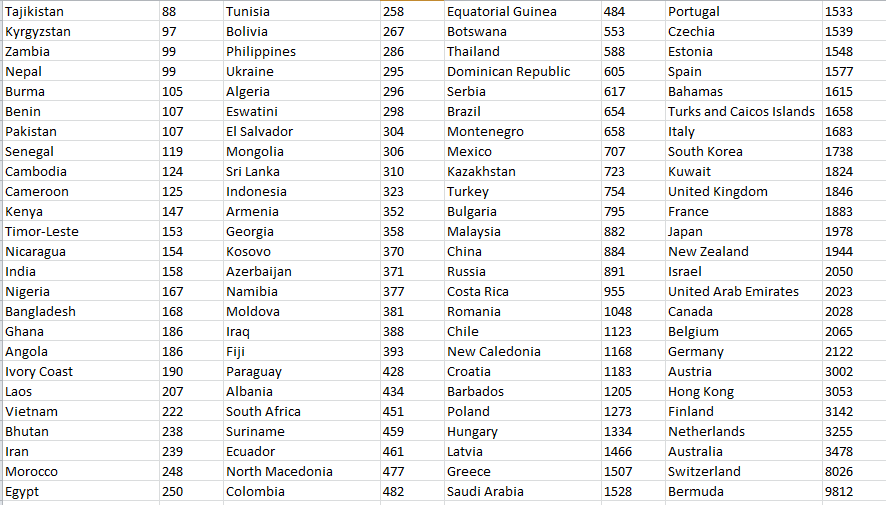
I created a large population size (100 entries) of monthly income earned in USD in different countries across the world. This is not a realistic data set, it has not been derived from any internet source and the amounts provided are not based on any facts but are example amounts I made up in order to answer the questions asked. I needed the amounts to be custom so that I could have outliers and two amounts that are more than double the value of the third highest amount.

Here is what my data set looks like:



What would I do if I found a couple of outliers in the data set of 100 values?

I have two outliers in this data set, which are Switzerland and Bermuda, being much higher than the other countries. In this context, I would double check to ensure this is not a typo, and read up on the country in question to verify that the monthly income there can be so much higher. If the country were something like Iran or Pakistan I would double check the survey itself as that cannot be correct, but as they are rich countries, it makes sense and is possible.

What would I do if I found two values that were twice the size of the third biggest value?

I would query both the higher two values and the third closest value that is less than half the value and ensure that this is a realistic relationship between the values. In this case, it is Bermuda and Switzerland that earns more than double the monthly salary in USD that Australia does, which is believable. Had it been something like Colombia earning more than double what Hong Kong or France does, I would have queried the source of information this data set is made up from.