Data Cleaning Project

Objective: Clean the given survey data and only include US values only by state.

Steps:

* Remove duplicates values
* Trim whitespaces in all columns
* Rename column names
* Use the “Proper” excel function to standardize the capitalization of all the data
* Use the find and replace feature to find common synonyms for the United States (US, USA, U.S.A etc) and standardized them as “United States”
* Filter by column for US values only and copy the results into a new sheet
* Remove unnecessary columns like “Country” and “If other Please ”
* Use the “If” function to set any missing values to 0 in the additional income column
* Create a “Total income ” column
* Go through each column and identify any blank values and see if I can infer the value
* Go through the State names to identify any errors and make sure everything is uniform
* Identify and correct errors in city column
* In the income column identify any outliers and decide if the values should be removed or not depending on the job title (example: A senior software developer making $150 or a School Principal making $600,000 )
* Check for any missing values or errors