Lauren Berryman

Prof. Minkoff

Data Journalism

12 February 2022

**SC Data Diary**

To clean the data in the “Condition Assessment” column, I first added a column to the right. I named this new column “condition\_assessment\_formatted.” I pasted the data from the left column into the right column. I inspected the data. Then, I selected this column (BT) on my sheet, and clicked edit, find and then replace. I wanted to filter out the “Not rated” data, so I typed in “Not rated” and replaced with nothing. I made sure “Find entire cells only” was selected and then clicked replace all. To make my data look more consistent, I formatted the cells to sentence case style rather than all caps. Therefore, I typed in “POOR” and replaced it with “Poor” and did the same for “FAIR,” “SATISFACTORY” and “UNSATISFACTORY.”

To clean the data in the “City” column, I first added a column to the right. I named this new column “city\_formatted.” I pasted the data from the left column into the right column. I inspected the data. Then, I selected this column (N) on my sheet, and clicked edit, find and then replace. I wanted to filter out the “NONE” data, so I typed it in the top box and replaced with nothing. I made sure “Find entire cells only” was selected and then clicked replace all. I also cleaned my data to make it more consistent. I changed the ones like “CLEMSON, SC” to just “CLEMSON” since it is in the city column and this data is about SC. I also found a “CLEMSON SC” record and changed to “CLEMSON” to make more uniform. I went through and found similar instances, like “GREAT FALLS, SC,” “NEWBURY, SC,” etc. and changed to just city name. There are still records that say “SECONDARY ROW” and “CHEROKEE FALLS DAM” which seem questionable. Thus, I would go back to my original source and fact-check that these are or are not cities. If they’re not, I would filter these out or replace with the correct city.

To clean the data in the “County” column, I first added a column to the right. I named this new column “county\_formatted.” I inspected this data. Then, I pasted the data from the left column into the right column. I selected this column (K) on my sheet, and clicked edit, find and then replace. I wanted to make the records more consistent by changing them all to sentence case. Most were already written in this format, but some were all caps. I followed the same steps as above and changed names like “CHESTERFIELD” to “Chesterfield,” “DARLINGTON” to “Darlington,” etc. A few have two counties listed, like “Fairfield, Lancaster” so I would go back to my source to clarify if it’s one county or both.