Elise Kline

Professor Minkoff

Data Journalism

02/12/22

**Data Diary**

1. **For the first cleaned column:**

I added a column next to the condition assessment to insert the new cleaned data. I cleaned the data by filtering out capitalization and the “not rated” text. I did this by going to edit and replacing all. For the “not rated” I replaced all with no new term and with all caps terms I replaced all with the lowercase spellings. I did that for all the terms that had all caps including “satisfactory,” “unsatisfactory,” “poor,” and “fair.” I also highlighted the cleaned column to easier point out what I did.

1. **For the second cleaned column:**

For the second column cleaning, I added a new column next to the city column. I cleaned this data again by filtering out the “none” text words and the text that had a comma and SC written next to the city. Since we know this is a South Carolina data set we don’t need the state in the city column, so I went through all the text and filtered out all the ones with comma SC next to it and replaced them with just the city. I also left in the text like “intersection” or “secondary road” or “dam” as I will have to go back to my original sources to double check if these qualify for the city column. Just like the last column, I highlighted the cleaned one so that it is easier to spot.

1. **For the third cleaned column**

For the third column cleaning, I chose the county column. Here I cleaned the variations from capitalizations and non-capitalizations. So, for example I changed “CHESTERFIELD” to “Chesterfield” so they would all be the same. I did this through the same process of going to edit, replace, replace all. I also found during the cleaning process to beware of adding a space after a word, otherwise it won’t recognize the text to replace. I noticed on the county column that not all of the all caps terms showed up on the drop down menu, so I had to scroll through to make sure I didn’t miss anything and I found that there were a few all caps texts that I had to go back and filter out/ replace with lowercase.