

Error types. Description of cleaning

1) Abbreviation inconsistency

I renamed the inconsistent values to the complete name after selecting the text facet for the column.

E.g1: "Department of Agriculture" vs. "DoA" vs. "Agriculture Department"

E.g2: "Department of Commerce" vs. "DoC"

2) Inconsistent date format

Some dates format has "-" instead of "/" as a separator and year- day-month format instead of day/month/year. I changed date format to be day/month/year. I firstly isolated the time unit (year, day, month), then I concatenated them in the right order. After that, I replaced the separator.

E.g: In "Completion Date (B1)" column, row 2: 2012-30-09

3) Symbols occurrences

The "\$m)" symbol occurred multiple times. I identified the cells, and transformed them by removing the symbol using the formula: `value.replace("$m)", "").toNumber()*1000000`

E.g: In "Lifecycle Cost" column, row 151: 0.1 (\$m).

4) String or non-numeric values for numeric columns

Strings instead of numeric values or even blank cells. I selected the numeric facet and I ticked the non-numeric box in order to see all the non-numeric values. After this, I deleted all the "matching rows" and I removed them. I did the same for the blank values, too.

E.g: In "Project ID" column there were about 49 non-numeric values and 20 blank values. For example, rows 114- 128 had the value: " Access Management"

5) Incomplete and inconsistent names:

Ambiguous values due to the incomplete names.

E.g.2: Rows 580, 600: "the Department". Was too ambiguous, therefore I deleted these rows.

Also, multiple references to the same agency, but with different names

E.g.1: "Department of Health and Human Services" vs "Department of Health" - I renamed it to the complete name: "Department of Health and Human Services " after selecting the text facet for the column.

6) Spelling Mistakes

E.g: On "Agency Name" column, the spelling error was : "Department of Agraculture" instead of "Department of Agriculture". I renamed the cell value to correct the spelling mistake.