1. Remove blank rows

A screenshot of a computer

Description automatically generated

* 1. Convert ID column to numeric and create a numeric faucet. Remove all matching rows (5blank rows).

A screenshot of a data

Description automatically generated

1. Remove duplicates.
   1. Faucet ID column by duplicate.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

* 1. Then faucet by star. Remove all matching rows (4 rows).

A screenshot of a computer

Description automatically generated

1. Fix Hospital Ownership column.
   1. Separate column by “ – ”

A screenshot of a computer

Description automatically generated

* 1. Cluster groups with the same name

A screenshot of a phone

Description automatically generated

* 1. Merge and cluster

A screenshot of a computer

Description automatically generated

* 1. A couple cannot be discovered by cluster: government/government state and voluntary non profit/ voluntary non-profit. Manually edit the names

A screenshot of a phone

Description automatically generated -> A screenshot of a medical survey

Description automatically generated

* 1. Remove all matching rows with “Not Available” (2 rows)

1. Outliers in “Hospital Overall Rating”
   1. Convert the column to number
   2. Facet by number

A screenshot of a survey

Description automatically generated

* 1. Remove all matching rows for the outlier (1 row)

A screenshot of a computer

Description automatically generated

1. Consistent values
   1. I faucet every column by text and reviewed if any special characters/inconsistency exist. All columns look good
2. Summary:
   1. The columns cleaned and type of cleaning:
      1. Removed blank rows from facility ID(5)
      2. Removed duplicate rows from facility ID (4)
      3. Removed uninformative rows from Hospital ownership (2)
      4. Removed outlier from Hospital Overall Rating (1)
   2. Outlier column:
      1. Hospital overall rating
   3. Total number of rows in cleaned dataset:
      1. 5412