Dataset Description:

--------------------

1. Dataset Source:

- The original dataset was obtained from INFO 6105 Data Science Engineering Methods and Tools.

2. Initial Exploration:

- The dataset was initially explored using Pandas, and a copy, df\_copy, was created for further processing.

3. Remove Duplicates:

- The first column of the dataset was removed, and duplicate rows were dropped to ensure data cleanliness.

4. Data Encoding:

- Label encoding was applied to convert categorical columns ('Bacteria', 'Antimicrobial', 'MIC\_Interpretation') to numerical values.

5. Missing Value Imputation:

- IterativeImputer from scikit-learn was used to impute missing values in the dataset. The 'MIC\_Interpretation' column was rounded to the closest integer.

6. Binning:

- The 'MIC' column was binned using the qcut function from Pandas.

7. Outlier Detection:

- Outliers were detected using the IQR method for both 'MIC' and 'Patient\_Age' columns.

8. Outlier Removal:

- Rows containing outliers were removed from the dataset, resulting in the creation of filtered\_dataframe.

9. Normalization:

- Numeric columns ('MIC' and 'Patient\_Age') were normalized using StandardScaler from scikit-learn.

10. Index Reset:

- Index values were reset for both filtered\_dataframe and filtered\_df\_copy.

11. Resultant DataFrame:

- The final DataFrame, filtered\_df\_copy, was created by replacing 'Patient\_Age', 'MIC', and 'MIC\_Interpretation' columns with the normalized values from filtered\_dataframe.

12. Result File:

- The processed dataset, filtered\_df\_copy, was saved to a CSV file named "HW1\_1\_Result.csv".

--------------------

End of Description.