Short Summary of changes

The main task was to ensure data cleaning and preprocessing techniques to maintain the integrity and usability of datasets.



The missing values were addressed by exploring various methods in Excel and/or Python.



Standardizing data to normalize feature scales, which is vital for algorithms sensitive to magnitude was also covered



Duplicate records were identified and removed to prevent skewed analysis.



Inconsistent formats, especially in date/time fields, were resolved using functions like pd.to_datetime() to ensure uniformity.



Common challenges such as inconsistent types, encoding issues, and schema mismatches were highlighted, along with strategies to assess data quality through completeness, consistency, and validity checks.