

Group: Emma Keaveney

Name: Emma Keaveney

Email: [emma.keaveney@hotmail.com](mailto:emma.keaveney@hotmail.com)

Country: United Kingdom

University: Durham University

Specialization: Data Analyst

First of all check for missing data:

When data should be a certain set of values (E.g. Yes/No or Male/Female), if data does not match the expected input then replace with 'U' for unknown.

Do this for all the categorised data, then look at numerical data, and if data missing also replace with 'U' for unknown – we do not want to make any assumptions about what data to replace it with.

Outliers:

For numerical data we also look at the quartiles and std deviation to spot any outliers. A noticeable outlier for the 'Age' column was ages above 120 years as we know this isn't possible. Replace these ages with 'U', as again we don't want to make any assumptions about what the data should be, and change the deceases status for that entry from No to Yes.