Data Intake Report

Name: Persistency of Drug
Report date: July 18, 2022
Internship Batch: LISUM 10
Version:<1.0>
Data intake by: Jeeyeon Shon
Data intake reviewer:
Data storage location:
https://github.com/shonjeeyeon/DG_Week_7/blob/main/Healthcare_dataset.xlsx

Tabular data details:
Healthcare_dataset.xlsx

Total number of observations	3424
Total number of files	1
Total number of features	68
Base format of the file	.xslx
Size of the data	898 KB

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

- Remove the “Feature Description” sheet and save the data to .csv
- Ptid can be used to identify and remove duplicates observations
- The dataset has been deidentified already
- Most of the features are categorical; will need encoding to enable ML