

# Data Intake Report

Name: Healthcare – Persistency of a Drug

Report date: July 18, 2022

Internship Batch: LISUM 10

Version:<1.0>

Data intake by: Jeeyeon Shon

Data intake reviewer:

[https://drive.google.com/file/d/1P\\_oMc6gOBlhW6dY5PxaqxV2swdHMuooK/view](https://drive.google.com/file/d/1P_oMc6gOBlhW6dY5PxaqxV2swdHMuooK/view)

## Tabular data details:

Healthcare\_dataset.xlsx

<b>Total number of observations</b>	3,424
<b>Total number of files</b>	1
<b>Total number of features</b>	69
<b>Base format of the file</b>	.xlsx
<b>Size of the data</b>	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 duplicate observations
- The dataset has been deidentified already
- Most of the features are categorical; will need encoding to enable ML