

Data Intake Report

Name: Healthcare Drug Persistence ML Project

Report date: 11/19/22

Internship Batch: LISUM14

Version: 1.0

Data intake by: Pedram Doroudchi

Data intake reviewer:

Data storage location:

Tabular data details:

“healthcare_dataset.csv”

Total number of observations	3,424
Total number of files	1
Total number of features	69
Base format of the file	.csv
Size of the data	913 KB

Proposed Approach:

- Used pandas duplicated() method to find zero duplicate rows in dataset
- Used pandas isnull() method to detect zero missing values in dataset
- Using assumptions given in variable description spreadsheet from original .xlsx dataset