Data Intake Report

Name: Deployment of Flask
Report date: June 20, 2022
Internship Batch: LISUM 10
Version:1.0
Data intake by: Jeeyeon Shon
Data intake reviewer: Jeeyeon Shon
Data storage location:
https://raw.githubusercontent.com/shonjeeyeon/DG_Week_4/main/drug200.csv
(Downloaded from: https://www.kaggle.com/datasets/prathamtripathi/drug-classification)

Tabular data details:

Drug200.csv

Total number of observations	200
Total number of files	1
Total number of features	6
Base format of the file	.csv
Size of the data	5.88KB

Proposed Approach:
- No duplicate rows were detected
- No patient identifiable information (e.g., name, address, date of birth) present in the dataset, so no additional de-identification steps are needed
- Dataset is small (200 rows), so undersampling is not desirable
- 4 features are categorical (Sex, BP, Cholesterol, and Drug): will need encode the data before establishing a model