Data Intake Report

Name: Project Week7
Report date: Nov 16th, 2023
Internship Batch: LISUM26
Version: 1.0
Data intake by: Krishna Ratna Deepika Haripuram
Data intake reviewer: Data Glacier
Data storage location:
https://drive.google.com/file/d/1P_oMc6gOBlhw6dY5PxaqxV2swdHMUooK/view

Tabular data details:

Total number of observations	3424
Total number of files	1
Total number of features	69
Base format of the file	.xlsx
Size of the data	902 KB

Proposed Approach

Dedup Validation (Identification) Approach:

A systematic approach was taken to ensure data quality:

1. Automated Duplicate Detection: Applied algorithms to identify duplicate records across patient data fields.
2. Manual Verification: Conducted manual checks on flagged duplicates for accuracy.
3. Data Consistency Checks: Ensured consistency and logical data ranges across all variables.

Assumptions for Data Quality Analysis

Key assumptions made during the analysis included:

1. Comprehensive Data: The dataset was assumed to cover all relevant factors influencing drug persistency.
2. Data Representativeness: Assumed the data accurately represented broader patient behaviors beyond the specific data collection period.
3. Accuracy in Data Entry: Presumed the dataset was free from significant entry errors, ensuring reliability in observed patterns.

This approach and these assumptions were vital in maintaining the focus and integrity of the analysis, specifically tailored to understanding drug persistency.