**DATA INTAKE REPORT**

Name: **HealthCare Dataset**

Report date: 05/19/2024

Internship Batch: LISUM32

Version: 1.0

Data intake by: Vedant Wagh

Data intake reviewer:

Data storage location:

<https://drive.google.com/file/d/1P_oMc6gOBlhw6dY5PxaqxV2swdHMUooK/view>

Tabular data details:

|  |  |
| --- | --- |
| Total No.of Observations | 3,425 |
| Total No.of files | 2 (sheets in excel file) |
| Total No.of features | 72 |
| Base format of the file | Excel |
| Size of the data | 992 KB |

**Proposed Approach:**

* unique patient IDs to identify and delete duplicate records using deduplication validation (identification).
* Maintain the confidentiality of information by cross-checking several identifiers, such as patient ID and date of birth.

**Assumptions:**

* All items with an identical patient ID refer to the same person.
* Values that are missing can be handled via statistical techniques (mean/median) or model-based strategies.
* Standard preparation procedures can effectively address data quality issues such as outliers.