A red and black rectangular sign with a heart and text

Description automatically generated

Andrea Plunkett Jackson

DAT 223 5-1 Assignment

Professor Usseglio, MSc

February 9, 2024

A diagram of data flow

Description automatically generated

**Characteristics of Data**

* The dataset was amassed by the Center for Disease Control (CDC) in 2014 using a quantitative discrete and continuous survey. Administered to more than 400,000 participants via the Behavioral Risk Factor Surveillance System (BRFSS), the leading continuously conducted health survey system in the world (CDC,2014). The BRFSS collects information pertaining to health-related risk behaviors, chronic health conditions, and the use of preventive services by the method of telephone interviews. Encompassing numerical and character data to query into household, health diagnosis, expenses, and habits.
* Insufficient data can potentially bring about deficient analytical consideration generating bias and erroneous decisions. The equivalent can be stated for less data than requisite. With inadequate data, statistical analysis becomes inaccurate restricting insight into patterns and risk assessment. Impacting the overall quality of the research.

**Data Provenance**

* Data Timeline (see illustration page 1)

1. Primary source of data, BRFSS telephone surveys completed by the CDC gathered from U.S. occupants in each state incorporating territories in 2014.
2. The initial storage was completed in 2017 by Heart Matters, the data was extracted from the BRFSS database and saved in CSV format. Housed securely on a hard drive in a locked room.
3. Upon expansion into Alabama, Heart Matters transported a copy of the data set and codebook using a flash drive to the newfound local.
4. Subsequently reaching Alabama in early 2018 via the IT system administrator it was directly imported into a SQL database after obtaining financing to store, manage, and analyze the data, converting from the raw CSV configuration to the structured database composition. Resulting in data loss due to unreadable characters within the file incompatible with SQL requirements.
5. New hire finalized in June 2018 for the position of data analyst at the Alabama location. Identification of data loss discovered after reviewing the master data set impacting the quality and validity.

* Access to Data

1. CDC
2. Heart Matters
3. IT system Administrator
4. Data Analyst

* Impact on quality

The provenance of data represents a fundamental part in revealing its quality and validity. By understanding the data origins, it adds to the trustworthiness and allowed for detection of mistakes and anomalies when importing was completed. Too many hands in the pot and not understanding the path the data has taken can cause security and interpretation.

**Data Management**

* Data Format

The data is presently within a relational SQL database file format called a Primary Data File (MDF) containing the schema and data. This information is stored on tables organized in rows and columns of records and attributes. Managed using a structured query language that allows the user to create and manipulate the data to customize extraction of applicable information for analysis.

* Legal and Ethical Requirements

Since the data acquired contains patient medical data one legal requirement would be HIPPA (Health Insurance Portability and Accountability Act), a federal law established to protect information from being disclosed without consent or knowledge (CDC,2022). Based on the data accumulated, concealing personal identifying information (PII) may be necessary. Contingent on state and local controls, further measures may be essential, and it is vital to determine protocols prior to handling of healthcare connected data. Ethically it is essential for transparency and confidentiality to be adhered to, ensuring privacy and equality of all research conducted.

References

Centers for Disease Control and Prevention. (2014, May 16). Behavioral risk factor surveillance system. Retrieved from https:// www.cdc.gov/brfss/about/index.htm

CDC. (2022, June 27). *Health insurance portability and accountability act of 1996 (HIPAA)*. Centers for Disease Control and Prevention. <https://www.cdc.gov/phlp/publications/topic/hipaa.html>

‌