

Find Issues with Data

For the attached dataset on patient information for a medical clinic; identify the issues with each of the fields(variables) listed below (eg: missing values, non-standard values, incorrect/outlier values)

- Date:
- Age:
- Gender:
- Address:
- Problematic Area:
- Type of Problem:
- Any Treatment Before:
- Sources:
- Name of Doctor:

Definition

- Missing Values are those which are blank on a cell.
- Non standard values are those which are not in the usual format or most popular format of the given data. Ex. Age should be in number format, date format should be same.
- Incorrect/outlier values are those which are out of the range of the possible values. Ex. Age 150 for a person, date as 2021//05/10, in case of address spelling mistake.

Analysis of Date Column

- The date column has values of date but there is **not a standard date format followed**. For example, some date is in A.D. and some are in B.S.
- There is only one row with missing value.
- Some of **outliers and errors** in date column are

26-03-3078

27//03/2078

30/03/20778

31-03-3078

16/04/2078/

Analysis of Age Column

- The age column has values of age which are not standard like 16Month, there are 57 rows with such a column.
- The number of Missing Values is 53.

- The numbers of outliers and errors are 7.

Analysis of Gender Column

- All values are in standard format except four values are missing.

Analysis of Address Column

- The address column has address which are not in standard format like number of district are 161 and remaining address are name of particular place like Thimi, Tansen , Kalanki.
- Number of outliers and error are 4.

Address
Boudha
Baudha
B aneshwo
KTM

- Number of missing values are 7

Analysis of Problematic Column

- There are 23 missing values in problematic column.

Analysis of types of problem Column

- There are 3 rows which are not in standard format.

Type of problem
Acne Acne Scar
Rases + Hairfall
Acne Acne Scar

- Numbers of missing values are 322.

Analysis of Any Treatment Before Column

- Number of missing values are 101.

Analysis of Sources Column

- 111 Missing values are appear there.

Analysis of Name of Doctor Column

- Number of missing values are 51.