Documentation for Cleaning Data-Set

Data Cleaning Documentation

Source of Data:

The Dataset was obtained from Kaggle, dataset made available thought Mobius and here's the link: https://www.kaggle.com/datasets/arashnic/fitbit

Step-by-step Documentation of Data Cleaning Process

- **Identify data quality issues:**
  - Missing Values:
    - All the cell has a value they change null cell to 0
  - Standardize:
    - Column Total_Distance and Tracker_distance have a value but not declared if the values are kilometer or anything for distance format.
    - Column Total_Distance and Tracker_distance Adjust decimal places to 2 for consistency.

- **Handle Missing Values:**
  - For the missing values they change it already to 0
- **Remove Duplicate record:**
  - No duplicate records were found in this dataset.
- **Correct data inconsistencies:**
  - For all distance column lot of decimal placement I change it to 2nd decimal only
- **Standardize data formats:**
  - Standardize column name:
    - Changing "namename" To "Name_Name".
- **Document changes:**
  - Importing Data-Set:
    - Utilized Power Query in Excel to import the raw dataset from the source to a new Excel workbook, maintaining the integrity of the raw dataset. A separate sheet was created to copy and paste the raw data for cleaning.
  - Inconsistencies:
    - The decimal placement of the all distance column
  - Standardize Title:
    - Changed column title from "namename" To "name_name".
  - Duplicate Record:
    - Do Duplicates records found in this dataset.

1. Opening new excel
2. Importing the dataset from the original data folder named “Fitabase Data 3.12.16-4.11.16” by using Power query in excel to import the files named “dailyActivity_merged” and create newssheet for the data to be copy.
3. Naming the new sheet as Fitabase Data 3.12.16-4.11.16 to know what is data on it.
4. Importing another dataset from the original data folder named “Fitabase Data 4.12.16-5.12.16” by using Power query in excel to import the files named “dailyActivity_merged” and create new sheet for the data to be copy.
5. Naming the new sheet as Fitabase Data 3.12.16-4.11.16 to know what is data on it.
6. Creating new sheet named Raw_merge to merge the 2 data set in one and to be ready to clean
7. First is i select all the values in the the sheet “Fitabase Data 3.12.16-4.11.16” and then copy then paste in to the Raw_marge sheets.
8. 2nd is i select all the values in the the sheet “Fitabase Data 4.12.16-5.12.16”” and then copy it then paste in to the last value vacant (row 459) in the Raw_marge sheets.
9. Saving the files with the name of the company with date and version “Bellabeat_22042024_v1 “

---

1. Select all then go to design panel upper then click “Convert to range”
2. Select all then click alt + h + O + I
3. Select all then click clear and clear format
4. Select a1 to change Id to ID
5. Select b1 to change ActivityDate to Activity_Date
6. Select c1 to change TotalSteps to Total_Steps
7. Select d1 to change TotalDistance to Total_Distance
8. Select e1 to change TrackerDistance to Tracker_Distance
9. Select f1 to change LoggedActivitiesDistance to Logged_Activities_Distance
10. Select g1 to change VeryActiveDistance to Very_Active_Distance
11. Select h1 to change ModeratelyActiveDistance to Moderately_Active_Distance
12. Select i1 to change LightActiveDistance to Light_Active_Distance
13. Select j1 to change SedentaryActiveDistance to Sedentary_Active_Distance
14. Select k1 to change VeryActiveMinutes to Very_Active_Minutes
15. Select l1 to change FairlyActiveMinutes to Fairly_Active_Minutes
16. Select m1 to change LightlyActiveMinutes to Lightly_Active_Minutes
17. Select n1 to change SedentaryMinutes to Sedentary_Minutes
18. Select all then click alt + h + O + I
19. Select all then click data ribbon then chick remove duplicates
20. Select D column then remove another decimal placement. Until 2nd decimal
21. Select E column then remove another decimal placement. Until 2nd decimal
22. Select F column then remove another decimal placement. Until 2nd decimal
23. Select G column then remove another decimal placement. Until 2nd decimal
24. Select H column then remove another decimal placement. Until 2nd decimal
25. Select I column then remove another decimal placement. Until 2nd decimal
26. Select J column then remove another decimal placement. Until 2nd decimal