Dataset Description:

1. *Fitbit\_flushed\_data\_cleaned.csv:*
   1. *Source:*
      * https://www.kaggle.com/datasets/gloriarc/fitbit-fitness-tracker-data-capstone-project
   2. *Data\_collection:* 
      * Data was collected from 30 consenting fitbit wears to track their daily activities for 30 days
   3. *Definitions:*
      * VeryActiveDistance - High Cardio Heartrate (150+bpm)
      * ModeratelyActiveDistance - Medium Cardio Heart Rate while walking (120 -150bpm)
      * LightActiveDistance - low/fat burn Heart Rate while walking (100 - 120bpm)
      * SedentaryActiveDistance - not in a cardio zone
      * VeryActiveMinutes - Minutes while in a high cardio zone (150+bpm)
      * ModeratelyActiveMinutes - Minutes while in a medium cardio zone (120 -150bpm)
      * LightActiveMinutes - Minutes while in a low cardio zone (100 - 120bpm)
      * SedentaryActiveMinutes - Heart rate at standard/average rate
   4. *Limitations:*
      * Had 90 rows in which data was insignificant due to charging or lack or wearing on a specific day
      * Dropped these rows from the data set
   5. *Plan for data usage:* 
      * Take a look into correlations amongst the data and build a linear regression for steps to calories burned
      * Then can use this linear regression to predict weight loss patterns based upon step count
2. *Fitbit\_mean\_data\_cleaned.csv (same as above but replaced missing values with mean of columns):*
3. *Step\_weight\_stats\_cleaned.csv:*
   1. *Source:* 
      * https://www.kaggle.com/datasets/nageshn/predict-weight-based-on-food-and-step-count
   2. *Data\_collection:* 
      * Individuals tracking of step count, weight, and calorie consumption (200+ days)
   3. *Definitions:*
   4. *Limitations:*
      * Only data from one person so can’t be considered to accurate
   5. *Plan for data usage:* 
      * See how model from dataset above predicts the weight loss metrics for this person

Data Dictionary:

Fitbit\_flushed\_data\_cleaned + Fitbit\_mean\_data\_cleaned:

Columns:

* Id (int) - number associated with person’s tracking stats
* ActivityDate (dateTime) - day month and year that the log took place
* ActivityDay (obj) - day of the week that the log took place
* TotalSteps (int64) - number of steps taken over course of day
* TrackerDistance (float64) - total distance over course of day
* LoggedDistance (float64) - total distance over course of day
* LoggedActivitiesDistance (float64) - Total distance for activities purposefully logged by user
* VeryActiveDistance (int64) - total distance while in a high cardio zone (150+bpm)
* ModeratelyActiveDistance (int64) - total distance while in a medium cardio zone (120 -150bpm)
* LightActivedDistance (int64) - total distance while in a low cardio zone (100 - 120bpm)
* SedentaryActiveDistance (int64) - total distance while not in a cardio zone
* VeryActiveMinutes (int64) - Minutes while in a high cardio zone (150+bpm)
* FairlyActiveMinutes (int64) - Minutes while in a medium cardio zone (120 -150bpm)
* LightlyActiveMinutes (int64) - Minutes while in a low cardio zone (100 - 120bpm)
* SedentaryMinutes (int64) - Minutes with non-elevated heart rate (inactive)
* Calories (int64) - Number of calories burned in day

step\_weight\_stats\_cleaned:

Columns:

* Id\_log (int) - Index Number
* Date (datetime) - Date of the log
* Gender (obj) - Gender type Male for all
* Height (int) - Height in centimeters
* Weight\_kg (float) - Total weight in kilograms
* Weight\_lbs (float) - Total weight in pounds
* Step\_Count (int) - Number of total steps taken in the day
* Calorie\_intake (int) - Number of calories consumed in the day