Google Sheets Change log

Daily Files
- Created fitabase_daily_data spreadsheet with the individual sheets comprised of daily data in the "Fitabase Data 4.12.16-5.12.16" folder
  - Raw Data copies remain in project folder

- Renamed each individual sheet to have consistent naming style and for clarity

- Changed "ActivityDate" format in daily_sleep to match rest of sheets

- Sorted ActivityDate column in both directions to ensure all dates fell within the correct range and they all did

- Checked length of all values in Id column, in all sheets, to be sure all Id's are 10 characters. Used LEN function and conditional formatting to highlight any values != 10

- Switched the dates in daily_sleep and daily_activities_merged to strings, to better work with CONCATENATE function

- Concatenated "Id" and ActivityDate columns to become Unique_Id (in column A) in daily_sleep and daily_activities_merged with the following function:
    =CONCATENATE(B2,C2)
    # Deleted Id and ActivityDate in daily_sleep

- Used Unique_Id to bring the sleep data into the daily_activities_merged sheet with the following VLOOKUPS:
    =IFERROR(VLOOKUP(A2, daily_sleep!$A$2:$D, 2, false), 0)
    =IFERROR(VLOOKUP(A2, daily_sleep!$A$2:$D, 3, false), 0)
    =IFERROR(VLOOKUP(A2, daily_sleep!$A$2:$D, 4, false), 0)
    # the IFERROR function is to prevent #N/A outputs and replaces them with 0
    # 0's must be accounted for in analysis

- Copied and special pasted values only for TotalSleepRecords, TotalMinutesAsleep, and TotalTimeInBed in daily_activities_merged

- Used Unique_Id column in daily_sleep to recreate Id (column A) and ActivityDate (column B) with the following functions:
    =LEFT(C2, 10)   # 10 being the number of characters in each Id
    =RIGHT(C2, (LEN(C2) - 10))   # It takes the length of the Unique_Id and subtracts the amount of characters in each Id to account for dates having different lengths (e.g. 4/30/2016 vs. 5/1/2016)
    # Unique_Id was kept in daily_activities_merged to keep a primary key in case it is needed later.

- Copy and special pasted values only for Id and ActivityDate in daily_sleep and deleted Unique_Id column

- Switched the ActivityDate columns in daily_sleep and daily_activities_merged, from being formatted as strings, back to dates

- Added DayOfWeek column next to date to be able to compare FitBit use by day of week with:
    =CHOOSE(weekday(C2), "Sun", "Mon", "Tue", "Wed", "Thu", "Fri", "Sat")

- Copy and Special pasted values in DayOfWeek column

- Formatted distance, time, step, and calorie measurements in daily_activities_merged for improved readability

- Cleared excess whitespace in all sheets

- Created pivot table to show number of data points per user (COUNTA).
    # There are 33 users and the Kaggle dataset says the data is for a group of 30 users.
    # After examining the Id column there are no obvious errors. Could be different devices
for one user

- Removed duplicate columns, columns with all 0 values, and columns with insufficient or unusable data in daily_activities_merged

- Removed 79 rows with 1440 (total minutes in a day) sedentary minutes under the assumption that the device was turned on, but not being worn. Rows decreased from 940 to 861. Did not affect total number of users.

- Removed 5 more rows with zero steps and 0 distance travelled under the assumption that the FitBit was not used by the user on that day. Rows decreased from 861 to 856. Did not affect total number of users.

- Created TotalMinutesWorn by adding VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, and SedentaryMinutes. Will be used to gauge how often users wear their device for the entirety of a day

- Copy and special pasted values for TotalMinutesWorn

- Sorted TotalMinutes Worn in descending order to make sure no data was recorded for more minutes than there are in a day

- Created a scatter plot comparing TotalSteps and TotalDistance to check for positive correlation and look for outliers in the dataset. Everything looked good.

- Created pivot table to show COUNTA for Unique_Id. The table showed that no Id and ActivityDate were combined more than once and there are not duplicate rows.

daily_activities_merged Spreadsheet Analysis
- Created a pivot table with average steps taken by day of the week and overall
    # Average steps of all data is 8,329
    # In descending order Sat(8,979), Tue(8,927), Mon(8,488), AVERAGE(8,329)
Wed(8,191), Thurs(8,185), Fri(7,821), Sun(7,669)

- Edited pivot table to show COUNTA of Unique_Id. Since the data begins on a Tuesday and ends on a Thursday, I filtered out the last three days of data to have an equal number of every day of the week. Shows which days of the week had the largest number of FitBit uses.
    # Friday had most uses and Sunday and Monday tied for the least uses

- Edited pivot table to show Average of TotalMinutesAsleep for users with 1 sleep record for the day, to get average amount of sleep recorded overnight. 0 records would indicate the user didn't wear the device while sleeping. 2 or more would skew TotalMinutesAsleep average because of naps.
    # Sunday had most sleep. Saturday had least sleep.

Average sleep for days with sleep data was 413 minutes, indicating the average user does not get the recommended amount of sleep.

- Edited pivot table to check DayOfWeek vs average of TotalMinutesWorn. Nothing stands out as useful.

- Edited pivot table to look compare Id vs TotalSleepRecords and average of TotalMinutesAsleep on days with 1 sleep record.
    # 24 of 33 users recorded sleep data at least once
    # 13 of 33 users recorded sleep data 22 or more times, 11 of 33 had between 1 and 15 sleep records, and 9 of 33 with 0 sleep records
    # Minor positive correlation between tracking sleep and average TotalMinutesAsleep

- Created a chart comparing VeryActiveMinutes and TotalMinutesAsleep for users on days with 1 sleep record recorded
    #The users with the most and fewest minutes of sleep tended to have the least amount of minutes asleep
    #The users who recorded the most VeryActiveMinutes tended to average around 400 minutes asleep

- Created the same chart as above with FairlyActiveMinutes instead of VeryActiveMinutes
    # Again users with the most and least sleep had among the least amount of FailyActiveMinutes

-

Hourly Data files
- Imported hourlyIntensities_merged into Google Sheets and renamed to hourly_intensities to match format of other sheets

- Changed ActivityHour, in hourly_intensities, date format to: 9/26/2008 15:59:00
    # The date time format in the raw data could not be imported into Big Query

- Imported hourlySteps_merged into Google Sheets and renamed to hourly_steps to match format of other sheets

- Changed ActivityHour, in hourlySteps_merged, date format to: 9/26/2008 15:59:00
    # The date time format in the raw data could not be imported into Big Query

- Downloaded the updated hourly_intensities and hourly_steps to the "Fitabase Working data" folder locally for uploading, cleaning, and preparing data in SQL