






Data Cleaning:

To begin with, I uploaded the data into BigQuery for cleaning and analysis. One thing that happened was uploading the sleepDay and weightLogInfo errors. This was due to the fact that the date column had AM and PM indicators in the columns. After reviewing the data, the time measurements were not a factor in consideration when reviewing. As such the data type was changed to date within google sheets then exported back to CSV. This allows me to then import all data back into BigQuery.

Reviewing the data shows that it is in the correct types for each column at first look.

	dailyActivityMerged	 QUERY ▾	 SHARE	 COPY
<div>SCHEMADETAILSPREVIEW</div>				
<div> Filter Enter property name or value</div>				
<input type="checkbox"/>	Field name	Type	Mode	Collation
<input type="checkbox"/>	Id	INTEGER	NULLABLE	
<input type="checkbox"/>	ActivityDate	DATE	NULLABLE	
<input type="checkbox"/>	TotalSteps	INTEGER	NULLABLE	
<input type="checkbox"/>	TotalDistance	FLOAT	NULLABLE	
<input type="checkbox"/>	TrackerDistance	FLOAT	NULLABLE	
<input type="checkbox"/>	LoggedActivitiesDistance	FLOAT	NULLABLE	
<input type="checkbox"/>	VeryActiveDistance	FLOAT	NULLABLE	
<input type="checkbox"/>	ModeratelyActiveDistance	FLOAT	NULLABLE	
<input type="checkbox"/>	LightActiveDistance	FLOAT	NULLABLE	
<input type="checkbox"/>	SedentaryActiveDistance	FLOAT	NULLABLE	
<input type="checkbox"/>	VeryActiveMinutes	INTEGER	NULLABLE	
<input type="checkbox"/>	FairlyActiveMinutes	INTEGER	NULLABLE	
<input type="checkbox"/>	LightlyActiveMinutes	INTEGER	NULLABLE	
<input type="checkbox"/>	SedentaryMinutes	INTEGER	NULLABLE	
<input type="checkbox"/>	Calories	INTEGER	NULLABLE	

sleepDay QUERY SHARE

SCHEMA

DETAILS

PREVIEW

Filter Enter property name or value

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	Id	INTEGER	NULLABLE
<input type="checkbox"/>	SleepDay	DATE	NULLABLE
<input type="checkbox"/>	TotalSleepRecords	INTEGER	NULLABLE
<input type="checkbox"/>	TotalMinutesAsleep	INTEGER	NULLABLE
<input type="checkbox"/>	TotalTimeInBed	INTEGER	NULLABLE

weightLogInfo QUERY SHARE

SCHEMA

DETAILS

PREVIEW

Filter Enter property name or value

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	Id	INTEGER	NULLABLE
<input type="checkbox"/>	Date	DATE	NULLABLE
<input type="checkbox"/>	WeightKg	FLOAT	NULLABLE
<input type="checkbox"/>	WeightPounds	FLOAT	NULLABLE
<input type="checkbox"/>	Fat	INTEGER	NULLABLE
<input type="checkbox"/>	BMI	FLOAT	NULLABLE
<input type="checkbox"/>	IsManualReport	BOOLEAN	NULLABLE
<input type="checkbox"/>	LogId	INTEGER	NULLABLE

Now that the data is imported into BigQuery we can then start querying and cleaning. The code that will be listed here is what I think will be major parts of the code; however, a full code sheet can be found on github.

One thing that is found is that the only table that has null values are in the weightLogInfo where the column FAT is null. Since this is data that is missing from the dataset that was provided as well as not a column that we will be mainly using this is alright and we can continue forward.

Another aspect to consider is that since the data is in long format there are duplicates in the ID column for each table.