Zara has bought a fitness tracker that can track her steps and distance she's moved (whether by walking, running, cycling, etc.). She hopes this is the start of a journey to better all-around health and fitness. She’s even taken up cycling! Since you have expressed your interest in working with data, Zara has asked you to help her on this journey. She wants you to use your data skills to add value to her fitness data.

What data can I extract from the app to follow Zara’s fitness journey?

##### **Data Pipeline**

##### **Data Sources and Raw Data**

Building the first stage of the data pipeline, which shows the data sources and the raw data you’re dealing with.

* app records the **number of steps** and Zara’s **resting heart rate**

##### **Operations and Information**

The next stage in the data pipeline shows the operations you can perform on this data to transform it into useful information.

##### **Destination**

Then you can see the **destination** (i.e., where this data might go next after processing). Zara might want to share her successes with her friends!

**Datasource** :Fitness Trackers’s app on Zara’s phone.

**Raw Data :**Number of steps, Resting heart rate

**Operations :**Create fitness report in the phone app

**Information :**Fitness achievements

**Destination(Insight) :**Share Zara’s achievements with zara’s friends

##### **Adding More Data to the Pipeline**

Sure! Let’s ask your health-conscious friend to start recording other details about her health, particularly any worrying symptoms she has.

Zara starts to record this data in a spreadsheet that looks like this:

| **Date** | **Symptoms** |
| --- | --- |
| February 01, 2022 |  |
| February 02, 2022 | Headache |
| February 03, 2022 | Fatigue |
| February 04, 2022 |  |
| … | … |

How would you add this new data source to the data pipeline above?

**Datasource** :Fitness Trackers’s app on Zara’s phone.

:Health log spreadsheet on Zara’s computer

**Raw Data :**Number of steps, Resting heart rate

:Symptoms, Cycled

**Operations :**Combine

**Information :**Fitness Report (showing improvements in resting heart

rate with exercise)

:Health Report(showing symptoms and heart

measurements over time)

**Destination(Insight) :**Fitness Plan (Fitness Instructor)

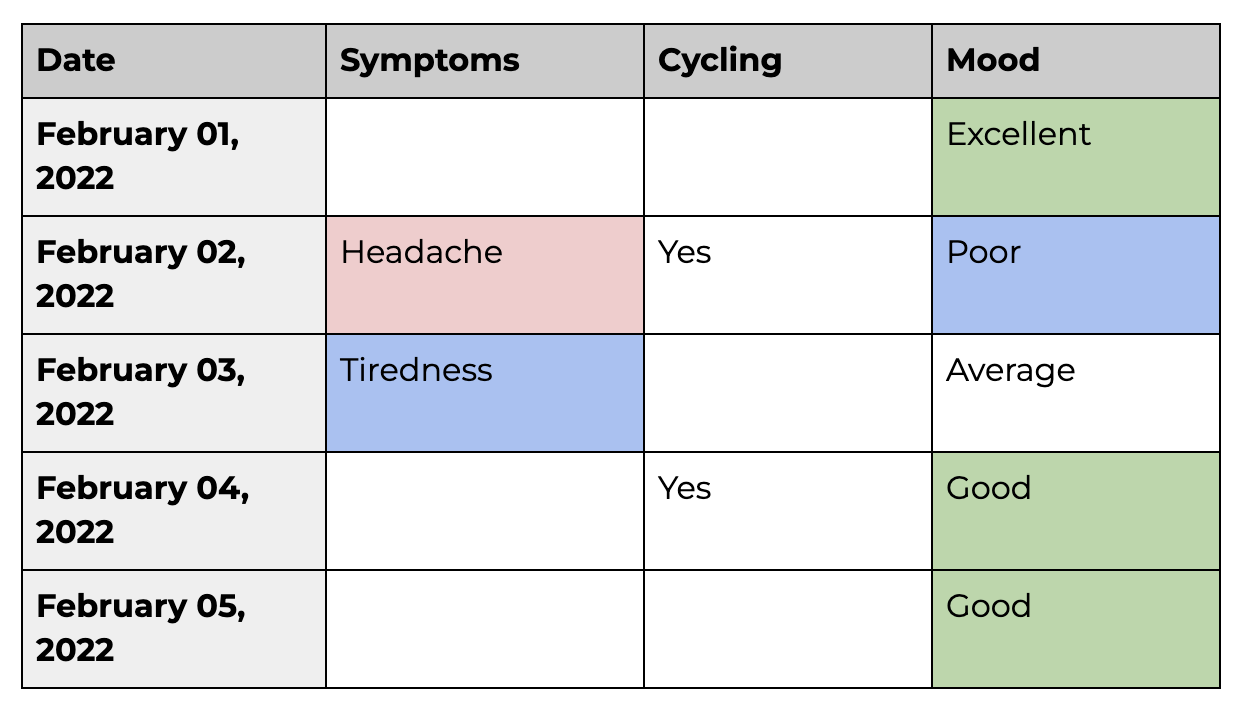
:Diagnostic chart (Doctor)

**Potential Actions** :Optimize Zara’s exercise regime

:New diagnosis and treatment

:Do nothing! Zara is reassured by her doctor.

But in Zara’s case, we can take data from her tracker app and health log spreadsheet, right?



***Zara's Fitness Jurney Project1***

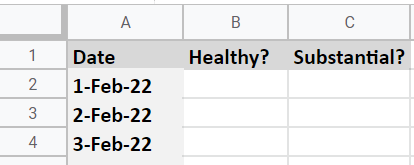
**Data from the fitness app**

Zara has exported the data from her fitness app for February to a file called **fitness\_tracker.xls**. You can download the file [here](https://course.oc-static.com/courses/7725096/fitness_tracker.zip).

**Health log spreadsheet**

Zara has created a spreadsheet named **health\_log.xlsx**, containing her health log. You can download the file [here](https://course.oc-static.com/courses/7725096/health_log.zip).

Zara has downloaded the food photos from her phone and stored them [here](https://course.oc-static.com/courses/7725096/Images-for-food_log.zip). She is just taking a picture of her main meal each day. Can you help her by turning this unstructured data into structured data? Create a new spreadsheet containing three columns: date, healthiness of the meal and how substantial it is. The first few rows should look like this:



create a **consolidated spreadsheet**, combining the data from all three sources. This spreadsheet will help you build a more complex view of Zara’s lifestyle choices, and start analyzing their impact on her health and fitness goals. In addition, Zara will provide you with some new data sources that you’ll be able to combine.

Step-by-step, you will:

1. **combine** the Fitness Tracker, Health Log, and Food Log data into a consolidated spreadsheet.
2. download some new data about Zara’s bank transactions.
3. **group and aggregate** the bank transactions, so you have one total transaction per day.
4. **filter** the bank transactions to just February.
5. **clean** up any issues with the data.

Zara likes to spend money because it lifts her mood. Her mother disagrees and thinks spending money doesn't really help. In an attempt to disprove her mother’s theory, Zara wants to collect some data about her spending habits to see if she can find a pattern linking her mood and spending. She downloads her bank transactions:

Let’s first combine the data in the [Fitness Tracker](https://course.oc-static.com/courses/7725096/fitness_tracker.zip), [Health Log](https://course.oc-static.com/courses/7725096/health_log.zip), and [Food Log](https://course.oc-static.com/courses/7725096/food_log.zip) files:

