

**Group Name:** Solo

**Name:** Armel Moumbe

**Email:** armel.moumbe@aivancity.education

**Country:** France

**College:** Aivancity school for Technology Business & Society

**Specialization:** Data Analyst

**NB: This project is done by me alone due to not having any group members. Thank you for your time and understanding.**

## **Problem description**

XYZ company is collecting the data customer using google forms/survey monkey and they have floated n number of forms on the web.

The company wants to create a pipeline which will collect all the data of these google forms/survey monkey and visualize the data in the dashboard. The company wants clean data and if there is any data issue present in the data then it should be treated by this pipeline (duplicate data or junk data).

## **Business understanding**

Business Understanding of the Fitness Industry with a Focus on Fitness Trackers

### **Market Overview**

The fitness industry encompasses a wide range of businesses, including gyms, personal training services, fitness apparel, nutritional supplements, and fitness technology. One of the fastest-growing segments within this industry is fitness technology, specifically fitness trackers, which will be the focus for the company.

### **Fitness Trackers: Definition and Types**

Fitness trackers are wearable devices that monitor and record various physical activities and health metrics. They come in different forms, such as wristbands, smartwatches, and clip-on devices. Key features typically include:

Step Counting, Heart Rate Monitoring, Sleep Tracking, Calorie Tracking, GPS Tracking and Activity Tracking which Logs various exercises such as running, cycling, swimming, and more.

## **Market Size and Growth**

The global fitness tracker market has seen significant growth over the past decade. Factors contributing to this growth include increased health awareness, technological advancements, and rising disposable incomes.

## **Key Players and Competitive Landscape**

The fitness tracker market is highly competitive, with several key players dominating the scene:

**Fitbit:** Known for a wide range of affordable and advanced fitness trackers.

**Apple:** Offers high-end smartwatches with comprehensive health tracking features.

**Garmin:** Specializes in fitness trackers with advanced GPS and outdoor activity tracking.

**Samsung:** Produces smartwatches and fitness bands with robust health monitoring capabilities.

**Xiaomi:** Provides cost-effective fitness bands with essential tracking features.

These companies continuously innovate to maintain their market positions, introducing new features, improving accuracy, and enhancing user experience.

## **Business Models**

Fitness tracker companies typically employ one or more of the following business models:

**Direct Sales:** Selling devices directly to consumers through online and offline retail channels.

**Subscription Services:** Offering premium services, such as personalized training plans and advanced analytics, on a subscription basis.

**Corporate Partnerships:** Collaborating with businesses to provide fitness trackers as part of corporate wellness programs.

**Health Insurance Partnerships:** Partnering with health insurers to offer fitness trackers to policyholders, encouraging healthier lifestyles.

# **Project lifecycle**

## **1. Initiation:**

- Define project objectives and scope.
- Identify stakeholders and their needs.
- Conduct a feasibility study if necessary.

## **2. Planning:**

- Develop a project plan outlining tasks, timelines, and resources.
- Create a detailed data acquisition plan using Google Forms/Survey Monkey.

## **3. Execution:**

### Data Acquisition:

1. Perform data acquisition using more than 2 Google Forms/Survey Monkey.
2. Collect all the forms data into a master form using a script.

### Data Processing:

3. Clean the data and perform dedup check.

### Data Visualization:

4. Visualize the data into a dashboard. Visualization can be based on the number of positive responses received location-wise or countrywide.

### Automation:

5. Create a batch which will run at specific times and dump the data into the master file.

## **4. Monitoring and Controlling:**

- Regularly review and monitor the progress of data acquisition and processing.
- Ensure data quality and consistency in the master form.
- Update the dashboard as new data is acquired and processed.

## **5. Closure:**

- Finalize and validate all deliverables.
- Document the challenges encountered during this implementation.
- Conduct a project review and gather lessons learned.
- Close the project formally and ensure all documentation is complete.

# Data Intake Report

Name: Data Collection Pipeline (Data Acquisition to Storytelling)

Report date: 18 July 2024

Internship Batch: 9572138

Version: 1.0

Data intake by: Armel MOUMBE

Data intake reviewer:

Data storage location: <https://github.com/m-armel/Data-glacier-Internship.git>

## Tabular data details: fitness consumer data

Total number of observations	30
Total number of files	
Total number of features	22
Base format of the file	.csv
Size of the data	12 KB

## Tabular data details: fitness analysis data

Total number of observations	545
Total number of files	
Total number of features	18

<b>Base format of the file</b>	.csv
<b>Size of the data</b>	163 KB

### Tabular data details: fitness trackers data

<b>Total number of observations</b>	610
<b>Total number of files</b>	
<b>Total number of features</b>	11
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	59 KB

### File Overview:

There are three files, fitness analysis, fitness consumer and fitness trackers. The first two files are files which contain information on the different customers and their fitness habits, who answered the surveys. The third one has information about the different types of fitness trackers the customers use, their color, brand, screen display and other features.

### File Content and Structure:

- **Columns and Data Types:** Boolean, string, decimal, integer.
- **Rows:** Each row in the files represents a unique individual.
- **Header:** The first row of these CSV files serves as a header, clearly labeling each column to denote the corresponding data field.

### Proposed Approach:

The approach is to use these files to create our master file, which will be used for our analysis. We'll start by analyzing and understanding the data of each file, what each feature represents and their relationship with one another.

### Assumptions:

The data type of each column is accurate.

The data recorded is true and has not been tampered with.

The data is not recent, so is not a current representation of the new demographic but can still be used to analyze some trends.

**GitHub Repo link:** <https://github.com/m-armel/Data-glacier-Internship.git>

