

# Stage\_2\_Prepare

Jakub Jędrych

2024-04-11

In this stage we need to answer to some questions about our data. Its important to know what data is that ? Does this data ROCCC? From what source is that data and who are responsible for optimizing and manage this data (data engineer)?

This is really simple but important stage because we need to ask more and more questions if we do not so our data can include a lot of bias and i can be really make hard to create professional process of analysis a data

- **Where is the data stored?**

FitBit Fitness Tracker Data (CC0: Public Domain) is stored in Kaggle

- **How is the data organized? Is it in long or wide format?**

Data have structured format tables and rows. Its a long format because of repeated values in columns like Id column (measured data for example each hour)

- **Are there issues with bias or credibility in this data? Does this data ROCCC?**

Data was created by Mobius.

Metadata about our data says:

*These datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Individual reports can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences.*