

Project Guidelines

1. Choose a field of **your interest**:

- passion: sport, botanics
- medicin
- meteorology

or **work**

2. Where to get some data? Maybe you have it already.. If not, try:
search sites specific to the field:

- <https://sports-statistics.com/sports-data/>
- web scraping (e.g. job search sites)
- <https://www.kaggle.com/datasets>

3. Pose yourself a **question** that is of interest to you. (e.g. “can I cluster tennis players into groups?”, “can I cluster job offers?”, “can I estimate how many heart beats do I have in this ECG signal?” or “can I optimize revenue/customer churn? ”, “are chocolates sales related to temperature?”, “how do I want to classify these labelled images?”)

4. Develop a **plan of attack** with your data. What data I am using for my goal? Do I have to merge different sources? Am I doing Supervised or Unsupervised? Am I doing classification or regression? Do I have enough data to try neural networks to balance bias and variance ?

#hours : 30

Good luck!

aris.marcolongo@gmail.com
mykhailo.vladymyrov@unibe.ch