Assignment 1

Due: Sunday Sept 29 11:59pm

Goal: Our goal for this assignment is to familiarize ourselves with creating a Python **notebook**. A tangent goal is to learn how to create a **short video,** which you will need to do for your final project. Today, videos are a primary ways for distributing content, so it is an essential skill.

Description. Find an interesting dataset from the Internet. It is better if the dataset is relatively new, e.g., after year 2000, and it should not contain only a few datapoints. You can find many at Kaggle or at the UCI Machine Learning Repository. I would also urge you to find data about Switzerland! Many data about Switzerland can be found at https://opendata.swiss/en/.

The dataset may be in CSV, JSON, XLS or other format. Prepare a **Python notebook** that should answer in some narrative form the following:

Task 1:

Describe the dataset. Why did you choose it? What is it about? Put a **link** where one can download it. What is the **dimensionality** of the dataset (number of columns)? What is the **cardinality** of the dataset (number of rows/objects)?

To answer the above you have to read the readme file, and look at the data. For this assignment you can use Excel (or even Pandas if you are familiar with it).

If there is a direct link where one can download the dataset you can include some code as follows, so that one can download the dataset directly from the notebook!

```
!mkdir -p data
# Fetches the dataset and stores it under the folder data
!curl 'the web link here' -o data/my_dataset.csv
```

If the dataset is not too big, you may also upload with together with the notebook, in which case you will upload a zip file of the notebook and file.

Task 2:

What kind of **questions** would one be interested to ask/answer about this dataset? For example, predict something about…, etc.

Task 3:

Prepare a short video (max 1min) with the following content:
•Your name and student ID.
•What is the dataset about and a link where you found it.
•What questions one can ask on it?

The video should include voice-over and (optionally) the face of the narrator in the beginning ("my name is Y, I will present the dataset Z."). To create the video, you may use Camtasia or Quicktime

Player and iMovie on Mac. **Upload** the video in youtube and embed the link in the notebook (as in the Lab assignments).

```
%%HTML
<iframe width="560" height="340" src="https://www.youtube.com/
embed/inN8seMm7UI"></iframe>
```

In case you haven't done it before, there is a separate pdf file in Moodle that describes how to upload a video in YouTube and get its link.

Task 4:

4.1. In the slack channel of **week 1**, write your name and email, the name of the dataset, the link to your video.
4.2: Add also the same content in this file so that we have everything in one place (see below for a snapshot.

Data Mining and Machine Learning - Assignment 1

Below, fill in your name, email, short dataset description and link to your video.
(*If the table is filled, click on the table, and then on the "+" at the bottom of the table grid*)

#	Name	Email	Dataset	Video Link
1	Michalis Vlachos	michalis.vlachos@unil.ch	MovieLens-1M: users rating movies they have watched	https://www.youtube.com/watch?v=mfY6ET9a8HA
2				
3				
4				
5				
6				
7				

Grade (Total 10 Points): 5 points for creating a good narrative in your notebook, with proper links and description of the dataset. **5 points** for the video.