CAS Applied Data Science, University of Bern, S. Haug

**Assignment Work Instructions for Module 1**

The assignment work for Module 1 is a Conceptual Design Report for a Data Science project you would like to perform. Ideally during the CAS with some dataset you find interesting. Please use or orient yourself according to the template below. Remove this page for the submission version. Key information is the following.

**Language:** English or German

**Deadline:** To be defined in class

**Deliverables M1:** Conceptual Design Report, (this document), GitHub Repository, including also Jupyter Notebook and poster from Module 2.

**Expected effort and length:** About 30 hours, minimum 5 pages

**Further formal quality requirements:**

* All references to be listed in corresponding section and cited with number in text
* All tables and figures to have numbered legends with short explanations (tables above, figures below) and be referenced in text (Figure 1: blablan, Table 1: blblbl).
* Figures to be as self explanatory as possible, e.g. plots with at least axis labeling including units.

**Lionel Perret**

Route de Diesse 8

2518 Nods

Lionel.Perret.76@gmail.com

+41 76 2311093

*Data Science Project*

**Prediction of Avalanche danger level in the Davos region with meteo data from the Weissfluhjoch station**

Conceptual Design Report

*16 October 2020*

# Abstract

This data science project analyses the data set of13’918 avalanches, recorded in the Davos area from 1998 to 2019 and combine it with the weather data of the Weissfluhjoch station, above Davos. The avalanche data set includes information on the size, the type and the number of avalanches. For each day, when avalanches are recorded, the official avalanche danger level of that day is also available in this database.

From the snow fall, wind and temperature data from the weather dataset, a “New Snow Problem” binary variable is created, in order to investigate if more avalanches were recorded on the days when this “New snow problem” condition is present.

The final objective of this project is to investigate, the degree of accuary of a prediction of the avalanche danger level on a particular day, with the help of the meteo data parameter available in the Weissfluhjoch data set.

# Table of Contents

[Table of Contents 2](#_Toc52612048)

[1 Project Objectives 4](#_Toc52612049)

[1.1 ADS Project - Module 1 - Objectives 4](#_Toc52612050)

[1.2 ADS Project - Module 2 – Objectives 5](#_Toc52612051)

[1.3 ADS Project – Module 3 – Objectives 5](#_Toc52612052)

[2 Methods 6](#_Toc52612053)

[Infrastructure 6](#_Toc52612054)

[Tools 6](#_Toc52612055)

[Python Libraries 6](#_Toc52612056)

[3 Data 7](#_Toc52612057)

[*Data Set 1 -* Weissfluhjoch Meteorological and snowpack measurement 7](#_Toc52612058)

[*Data Set 2 -* Snow avalanche data Davos, Switzerland, 1999-2019 8](#_Toc52612059)

[4 Metadata 10](#_Toc52612060)

[5 Data Quality 11](#_Toc52612061)

[6 Data Flow 12](#_Toc52612062)

[7 Data Model 13](#_Toc52612063)

[Conceptual Data Model 13](#_Toc52612064)

[Logical Data Model 13](#_Toc52612065)

[Physical Data Model 13](#_Toc52612066)

[8 Risks 14](#_Toc52612067)

[9 Preliminary Studies 15](#_Toc52612068)

[9.1 Plots done for Module 1 15](#_Toc52612069)

[9.1.1 Weissfluhjoch Meteo data 15](#_Toc52612070)

[9.1.2 Davos avalanche data 16](#_Toc52612071)

[9.2 Plots and Report for Module 2 of ADS CAS 18](#_Toc52612072)

[9.2.1 Overview and Concept 19](#_Toc52612073)

[9.2.2 Objective 1 19](#_Toc52612074)

[9.2.3 Objective 2 20](#_Toc52612075)

[9.2.4 Project Module 2 Conclusion 20](#_Toc52612076)

[9.2.5 Project Module 3: Outlook and open questions 21](#_Toc52612077)

[10. Conclusions 21](#_Toc52612078)

[11. Acknowledgements 22](#_Toc52612079)

[12. References and Bibliography 22](#_Toc52612080)

[13. Illustrations-Graphs List 22](#_Toc52612081)

# Project Objectives

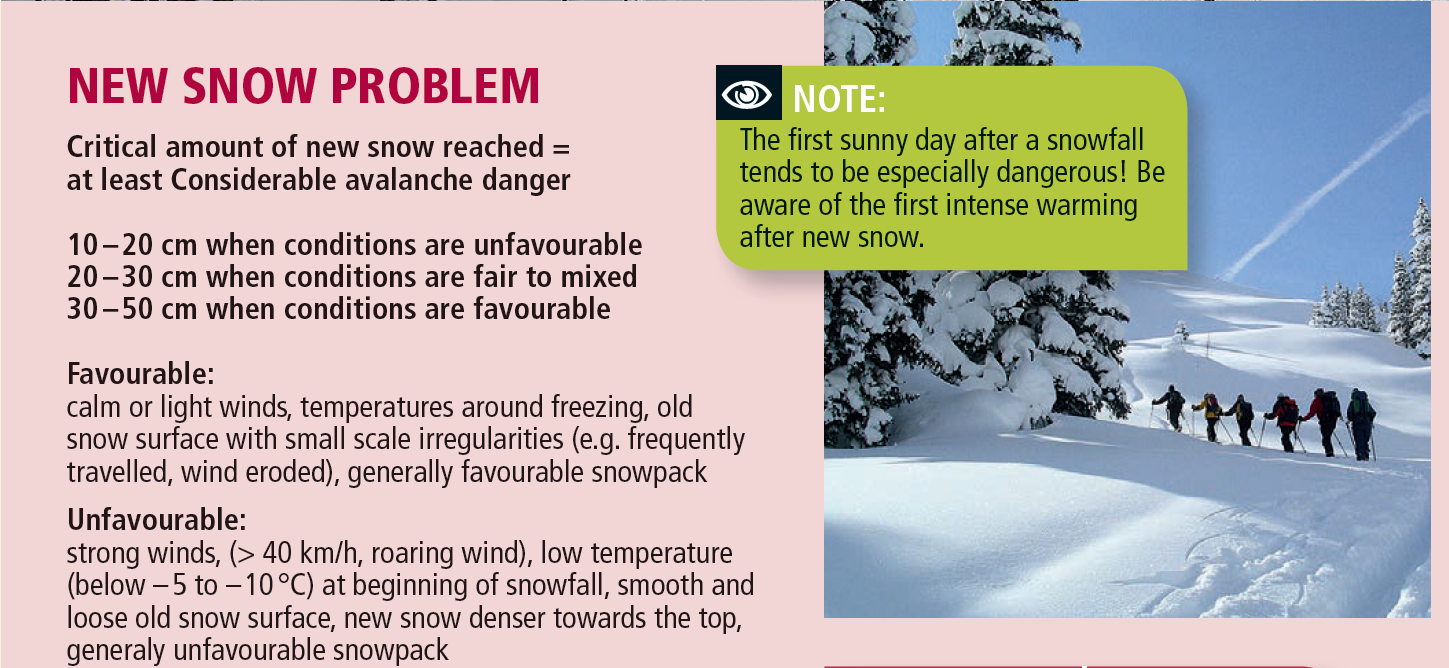
## ADS Project - Module 1 - Objectives

The main activity for my module 1 project will be first to get familiar with the 2 datasets and after to achieve the following goals:

* Pre-process the avalanche dataset and start to describe it statistically the avalanches that were recorded: Number of avalanches, type, size, causes, …
* Pre-process the meteo dataset and create a binary variable “New Snow Problem” conditions for each day:

1 = “New Snow Problem” is present on the day = “New Snow Problem” conditions

0 = “New Snow Problem” is not present on the day = “Other Snow” conditions



"Favorable and unfavorable avalanche condition after new snow" – SLF [1]

***Note:***

*The above recommendation* “Critical amount of new snow reached = at least **Considerable** avalanche danger”refers to the European avalanche danger level, more information in picture below.

|  |  |  |
| --- | --- | --- |
| **Level** | **Characteristics** |  |
| **5**  **Very high** | **Extraordinary avalanche situation**  Numerous very large and extremely large natural avalanches can be expected. These can reach roads and settlements in the valley. |
| **4**  **High** | **Very critical avalanche situation**  Natural and often very large avalanches are likely. Avalanches can easily be triggered on many steep slopes. Remote triggering is typical. Whumpf sounds and shooting cracks occur frequently. |
| **3**  **Considerable** | **Critical avalanche situation**  Whumpf sounds and shooting cracks are typical. Avalanches can easily be triggered, particularly on steep slopes with the aspect and elevation indicated in the avalanche bulletin. Natural avalanches and remote triggering can occur. |
| **2**  **Moderate** | **Mostly favorable avalanche situation**  Warning signs can occur in isolated cases. Avalanches can be triggered in particular on very steep slopes with the aspect and elevation indicated in the avalanche bulletin. Relatively large natural avalanches are not to be expected. |
| **1**  **low** | **Generally, favorable avalanche situation**  No warning signs present. Avalanches can only be triggered in isolated cases, in particular on extremely steep slopes. |

"Definition of European Avalanche Danger Level" – SLF [2]

## ADS Project - Module 2 – Objectives

Following the pre-processing of the datasets, I will combine those two datasets and my objectives for the module 2 project are:

* Analysis of number of avalanches per day with regards to avalanche danger level

Here I want to analyse if the the number of avalanche per day are increasing and how they are increasing.

* Influence of new snow and other meteo parameter on the number of avalanches per day, by using the binary variable created in module 1 project and by comparing the 2 distributions

1 = “New Snow Problem” is present on the day = “New Snow Problem” conditions

0 = “New Snow Problem” is not present on the day = “Other Snow” conditions

* Prepare the data and check the possibility to use the obtained data for prediction of the avalanche danger level with the meteo parameter

## ADS Project – Module 3 – Objectives

The final goal of my project is to be able to predict the avalanche danger level in an area, with the help of the weather data from the past days.

First I would like to train the model in using the combined available weather data of the Weissfluhjoch (time periode from the avaialbe period and check if how good is my prediction for the real avalanche danger level of winter 2020-2021.

# 2 Methods

## Infrastructure

* Microsoft Surface Book 2 with Intel Core i5-8350U, 8Go, 256Go, SSD
* Windows 10 Professional, Version 10.0.19041

## Tools

* Jupyter Notebook (Version 6.0.0): Interactive computing product based on the platform Anaconda 2019.07 (Python 3.7 Version [6]). Data analysis is performed in Python 3

## Python Libraries

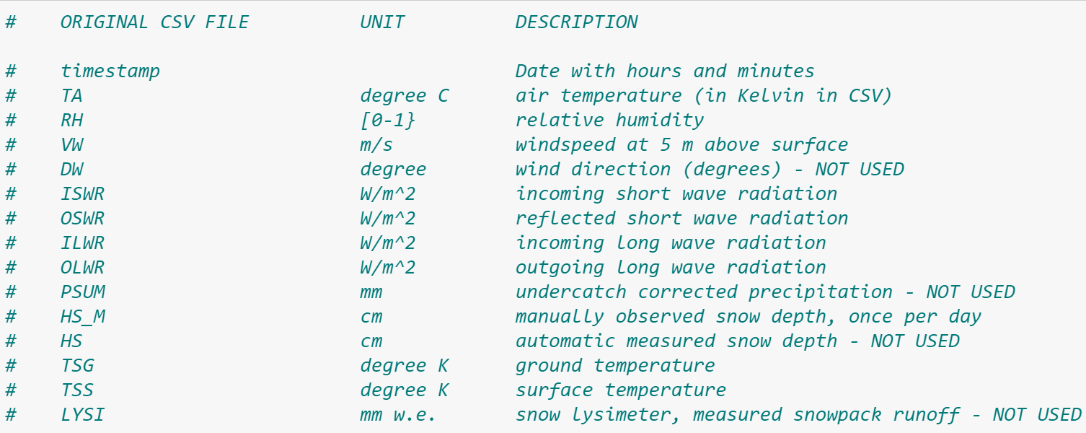
* pandas: Library and tool for dealing with data structures, data tables, manipulating data.
* numpy: Library and tool for calculating with data and arrays.
* matplotlib: Library and tool for the visualization of data in diagrams.
* scipy: Library and tool for mathematical calculations (scipy.stats: probability distributions, normality tests, regression etc.)

# 3 Data

## *Data Set 1* Weissfluhjoch Meteorological and snowpack measurement 1999-2017

*Description*: This dataset includes standard meteorological and snowpack measurements from the automatic weather station at Weissfluhjoch, Davos, Switzerland

**Reference [2]**



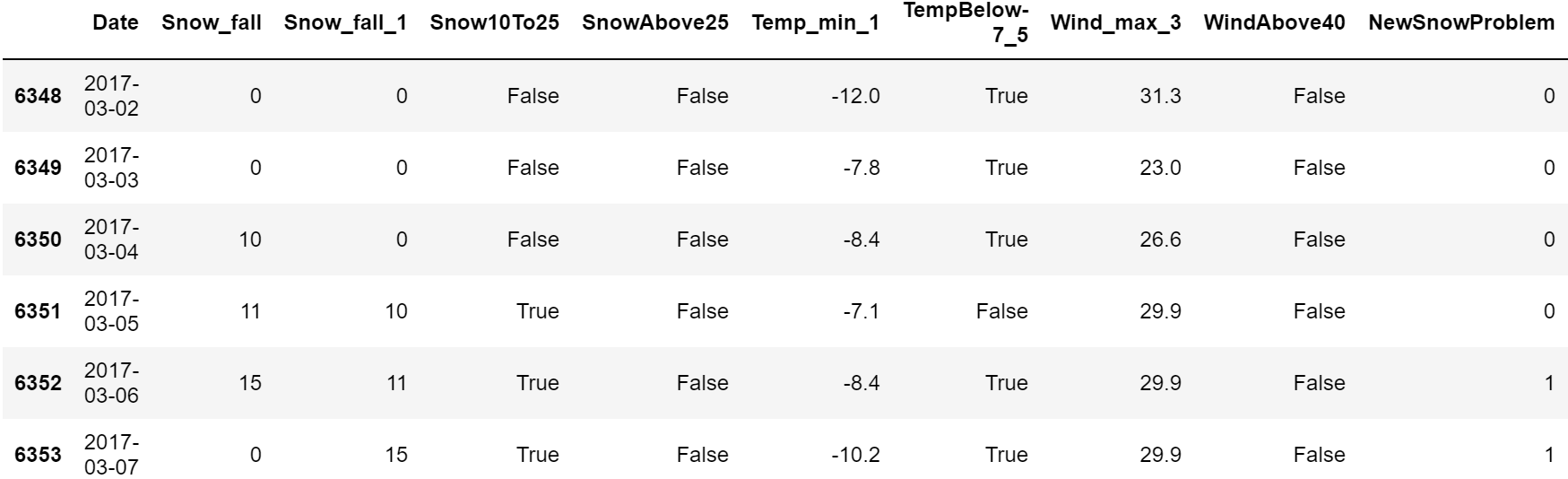
*Column*

*names of*

*original csv*

*file*

In the *Meteo.ipynb* Jupyter notebook (refer to Data model chapter below for description of all the files used), I have processed the above data-set and have created additional columns to calculate the binary variable “NewSnowProblem”:



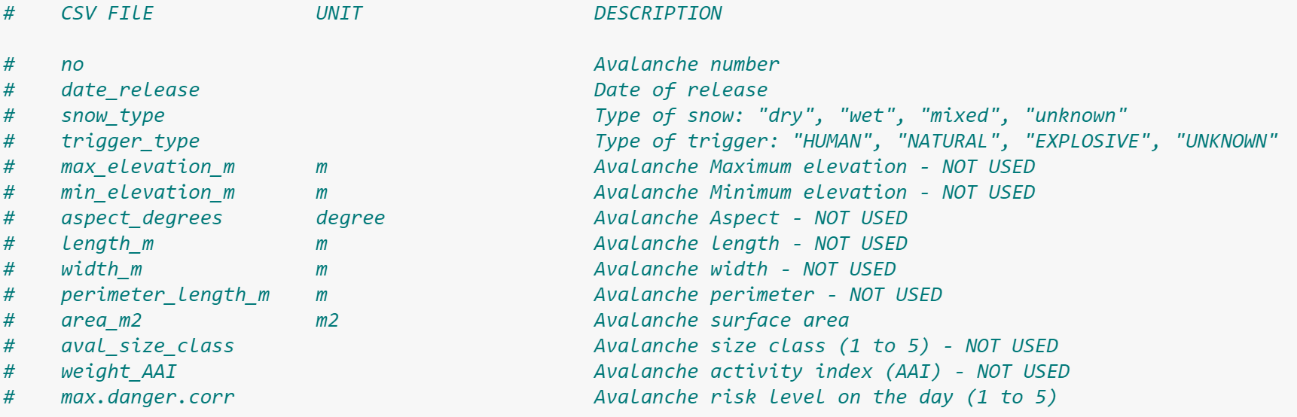
At the end of the *Meteo.ipynb* Jupyter notebook, I save the obtain data in *Meteo.csv*. This file has the following columns:



## *Data Set 2* Snow avalanche data Davos, Switzerland, 1999-2019

*Description*: This dataset includes observations about all avalanches recorded in the region of Davos, Switzerland, during the winters 1998-1999 to 2018-2019 - 21 years – total 13918 avalanches

**Reference [3]**

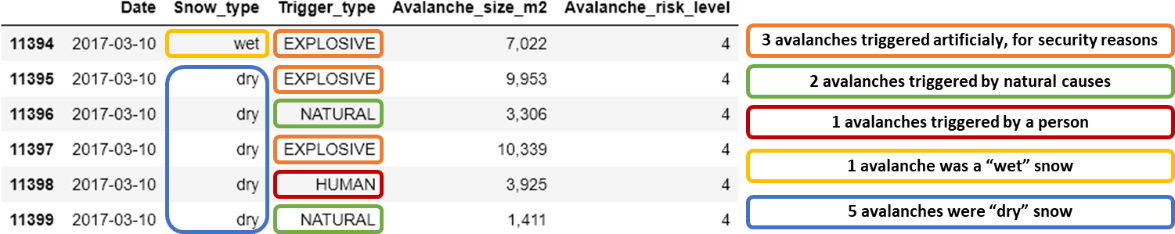


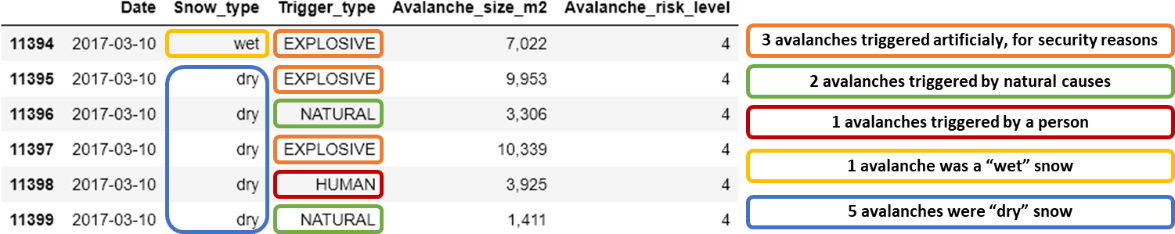
*Column Names*

*Original*

*CSV file*

*Snow Type* column:

* **“dry” avalanche: Type of avalanche where the snow has not melt. Can be an avalanche with fresh powder snow or also a “slab” avalanche. “Slabs” are caused by the effect of the wind on the snow.
* “wet” avalanche: In this avalanche, the snow as melt because of the sun and/or temperature. This type of avalanche is more frequent in the second part of the winter.



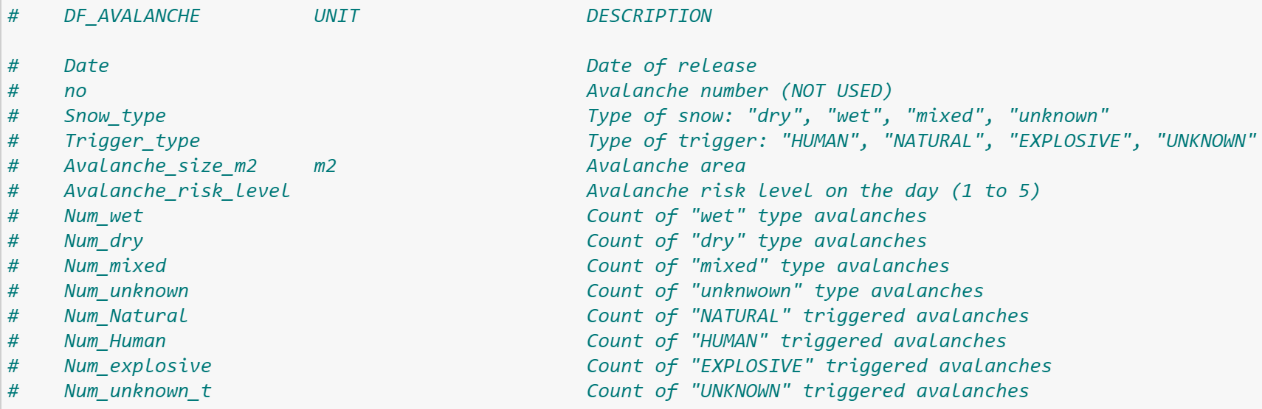
*Trigger Type* column

* “EXPLOSIVE”: avalanche triggered artificially, for security reasons
* “NATURAL”: avalanche triggered by natural causes
* “HUMAN”: avalanche triggered by a person unintentionally
* “UNKNOW”: Triggered cause of the avalanche is unknown

*Date* column:

in the original data frame, there is one line per recorded avalanche, so there is often more than one row per day. I have done some processing in order to get all information on the number of avalanche per day and per type/trigger causes of avalanches 🡺 columns “Count” in picture below.

In the *Avalanches.ipynb* Jupyter notebook, I save the obtain data in *Avalanches.csv*. This file has the following columns



*Column*

*Names*

*Avalanches*

*CSV file*

*Data Set 3:* **Combined avalanche and meteo data – used in analysis**

*Description*:

*Columns:*

*df\_xxx*

*(processed*

*Data frame)*

# 4 Metadata

Explanation of the variables used in the Jupyter Notebook ProjectM2.ipynb analysis file is available on github under:

In this directory

* *Meteo\_metadata.txt* contains the metadata information about the meteo variables found in the *Meteo.csv* file
* *Avalanches\_metadata.txt* contains the metadata information about the avalanches variables found in the *Avalanches.csv* file

# 5 Data Quality

What are the quality requirements you have to reach your analysis goal (precision ...)?

Are they met? If not, do you expect a significant impact on your results,

One of the big incertitude that I have in starting this project, is the “quality” of the weather information. I have no doubt that the weather measurement from the Weisfluhjoch station are precise, but those measurement are from one location in the davos area. All the recorded avalanche are close to this location, but are spread out in an area that is about 17 square km around davos and the Weissfluhjoch (refer to illustration on the next page)

Snow Fall or wind for example can be quite different a few km away and at another altitude than the measuring station.

On the data quality also, no gps position of the avalanche are available in the data set. This could have been a good information to check in which local condition (steepness of the slope) where the avalanche occurred.

Any measures to improve the quality?

Data requirement: snow fall at the place of avalanche, not available,…

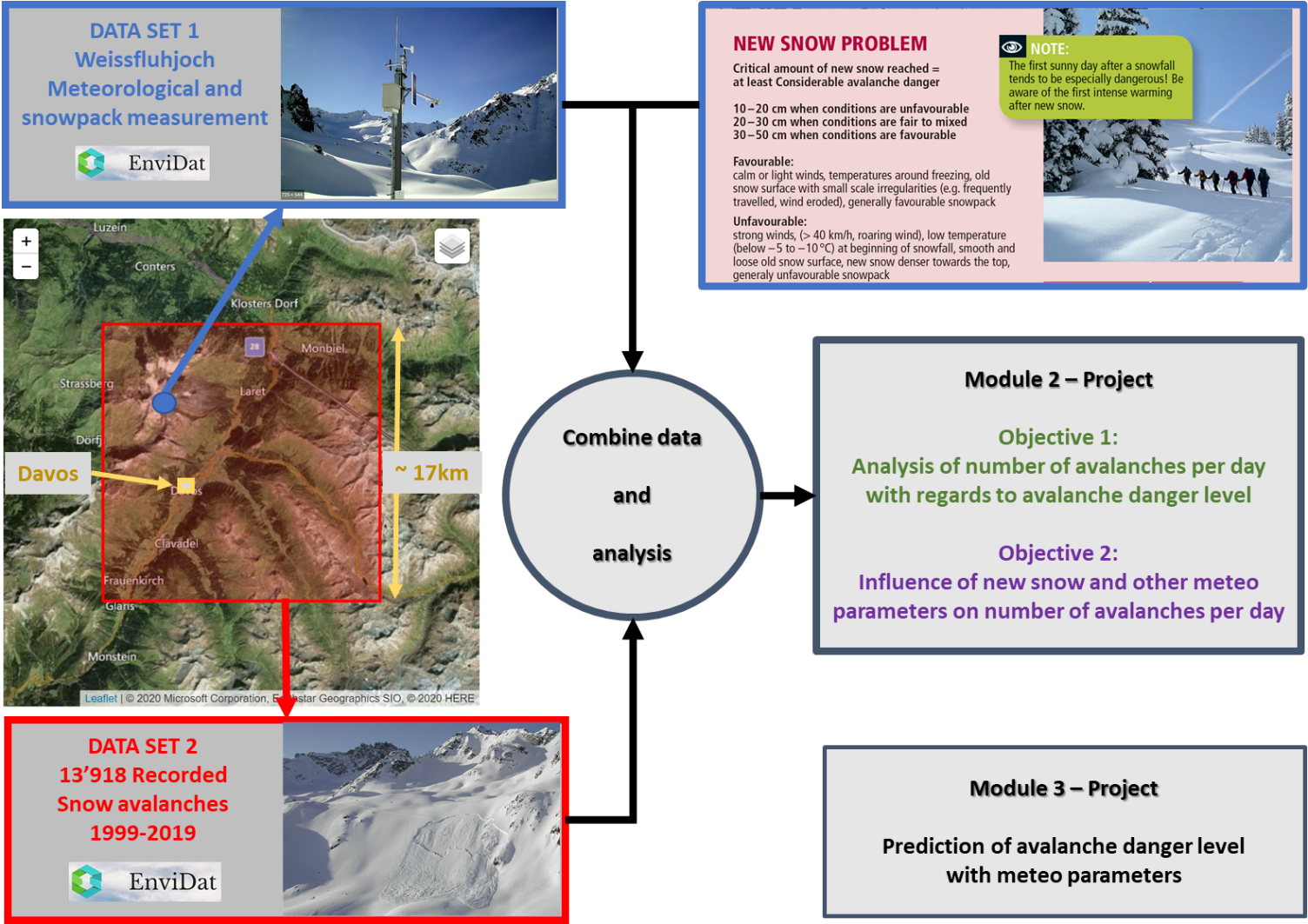
Data complete, precision

Data quality requirement

Number of avalanche recorded is quite high, but this contains all different type of avalanches (caused by explosive, unknown,….) so

# 6 Data Flow

The following 2 data sets found on “EnviDat” website ([www.envidat.ch](http://www.envidat.ch)) are used for this project. Those data sets are combined and analyzed for my personal project.



# 7 Data Model

## Concept

## Logical

## 

## Physical

The need on infrastructure is not important and all project will be done on my personal laptop.

# 8 Risks

What can go wrong?

When this and that goes wrong, what counter measures do you have?

What will be the impact on the quality of the aimed output, project time schedule, project cost ?

* Time management, Back-up, Personal issue

# 9 Preliminary Studies

## 9.1 Plots done for Module 1

### 9.1.1 Weissfluhjoch Meteo data

#### Mean Temperature

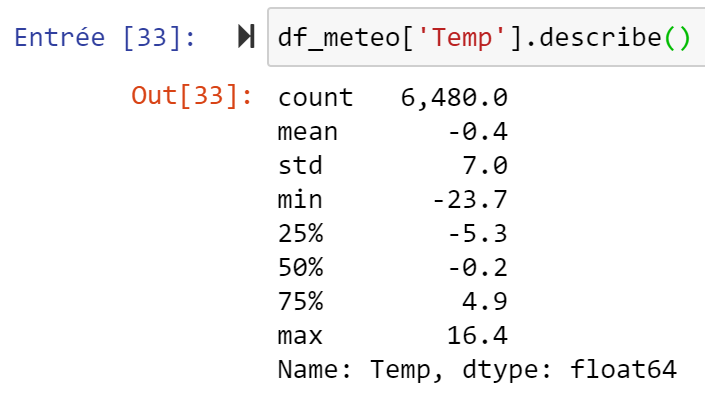
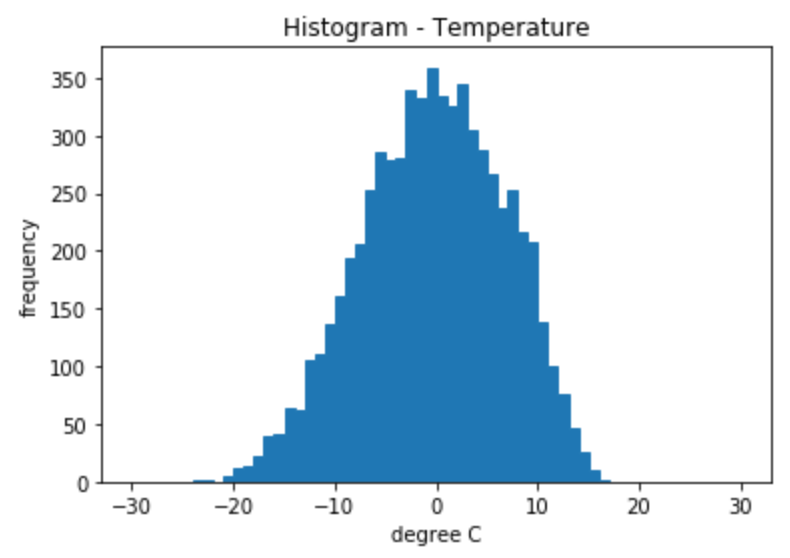
 

Figure No. [1] / Mean Temperature - Weissfluhjoch weather station

#### Sum of snow fall of the last 3 days

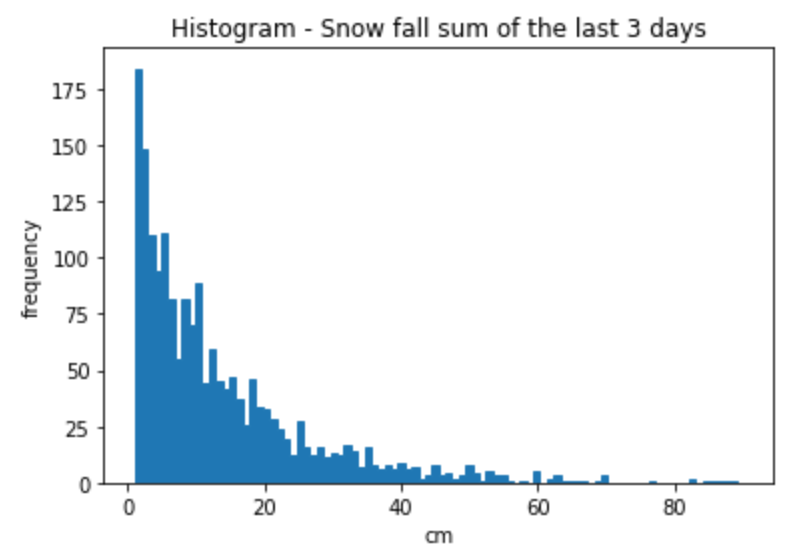
 

Figure No. [2] / Sum of snow fall of last 3 days - Weissfluhjoch weather station

#### Mean wind

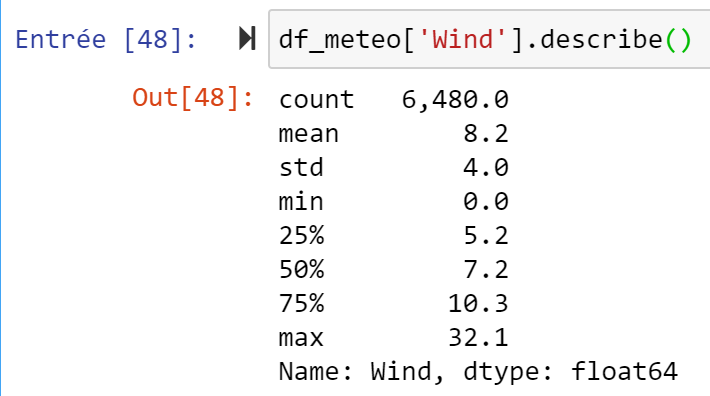
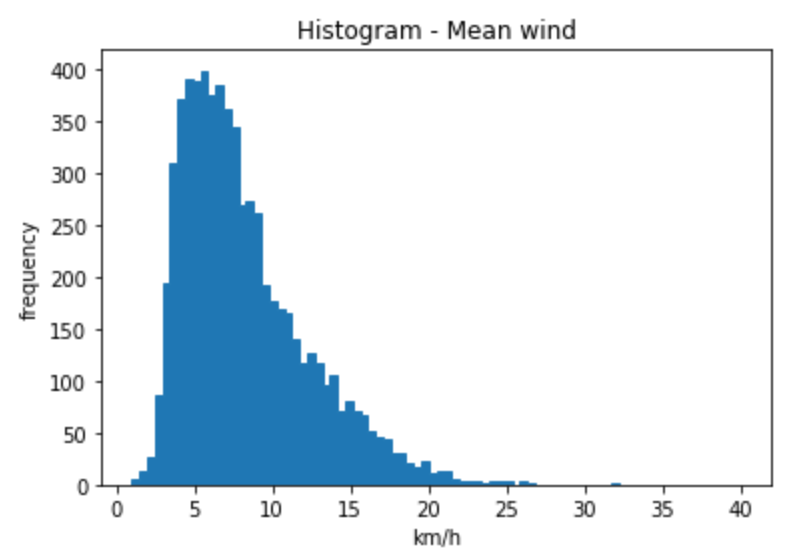
 

Figure No. [3] / Mean wind - Weissfluhjoch weather station

### 9.1.2 Davos avalanche data

#### Size of avalanches

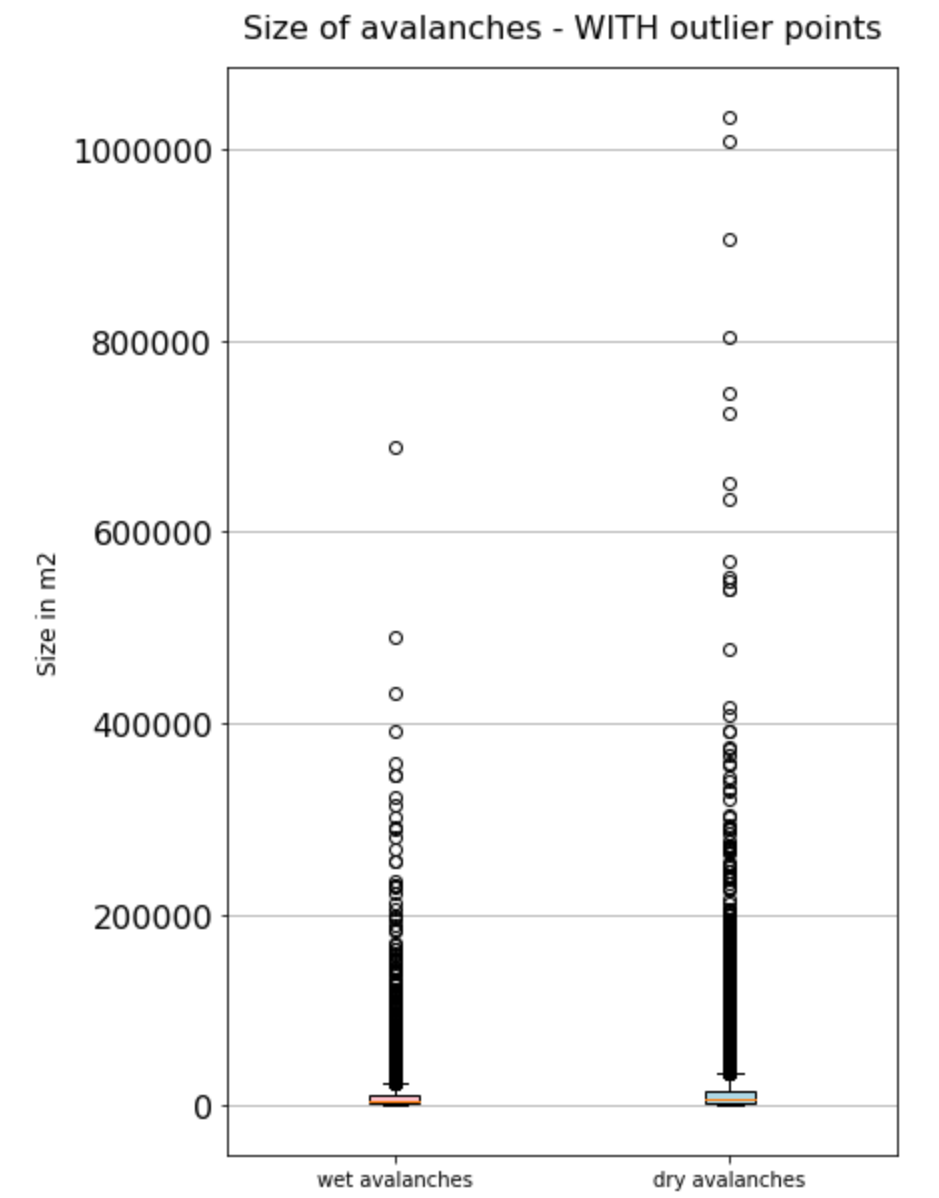
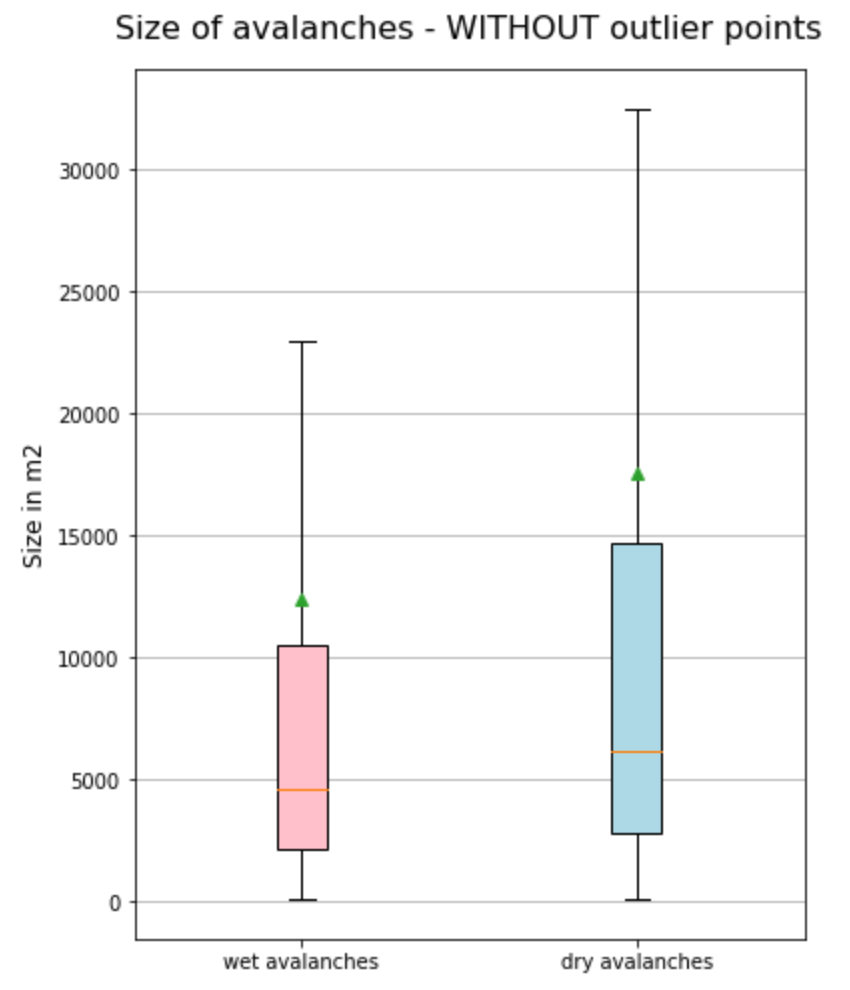
 

Figure No. [4] / Avalanche Size for “Wet” and “Dry” avalanches

It is interesting to see there are some outliers points far from the means (represented with the green triangles)

The avalanche size can be quite big, in 2 cases, above 1’000’000 square meters.

The distribution of the avalanche size has a similar shape for wet and dry avalanches

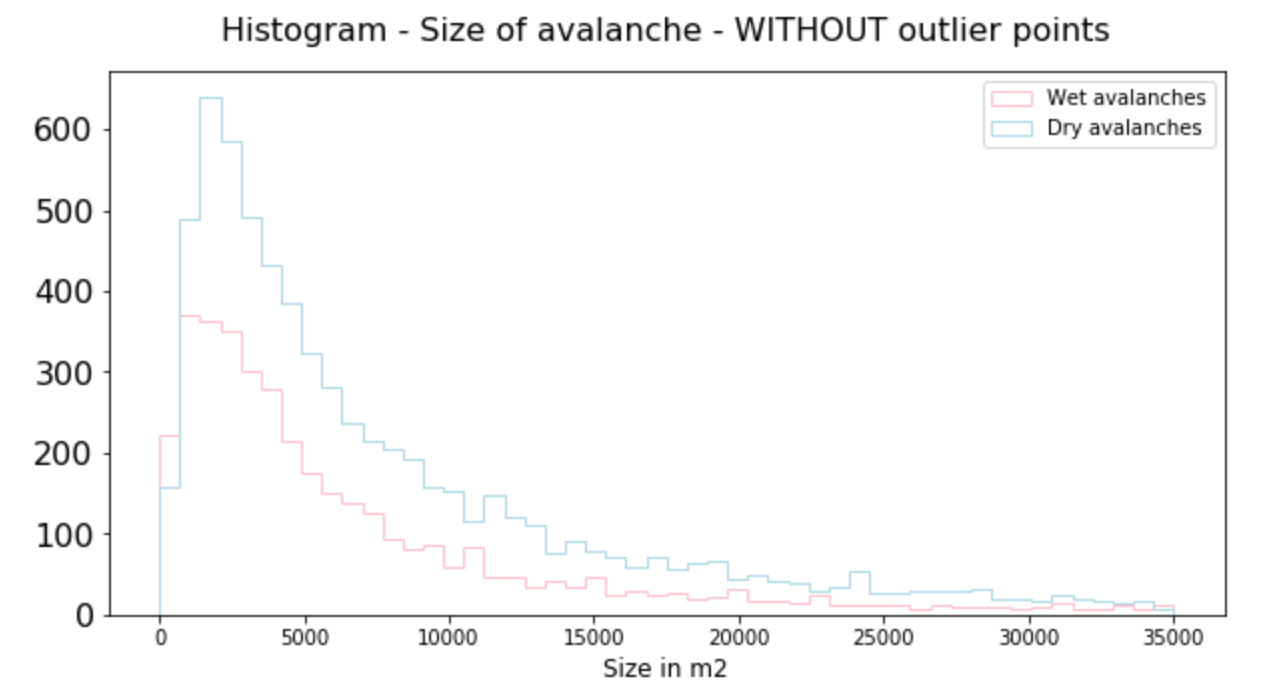


Figure No. [5] / Histogram – Size of Avalanche – WITHOUT outlier points

#### Number of days with “dry” avalanches recorded

New snow avalanches are always “dry” type avalanche, so by filtering this type of avalanche and counting the number of days we obtain:

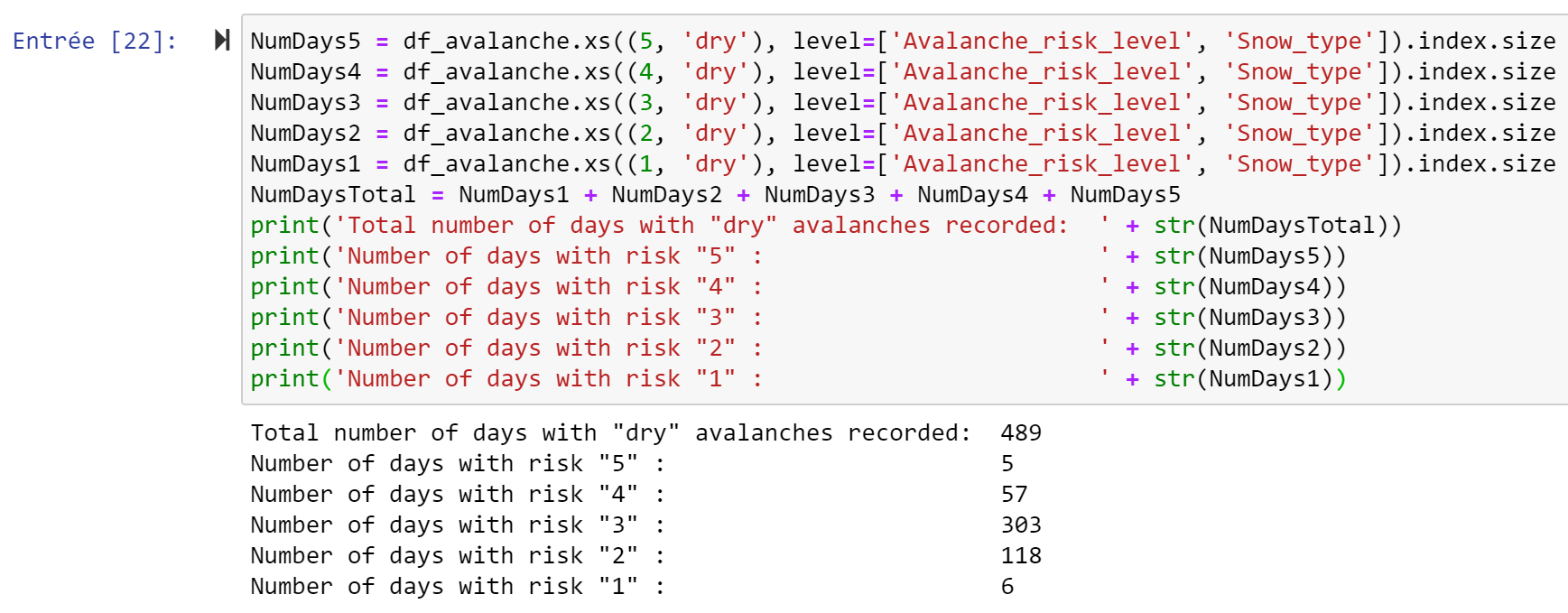
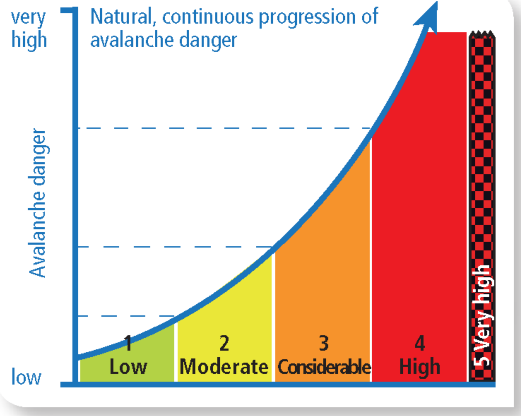


Figure No. [6] / Number of days with “dry” type avalanches recorded



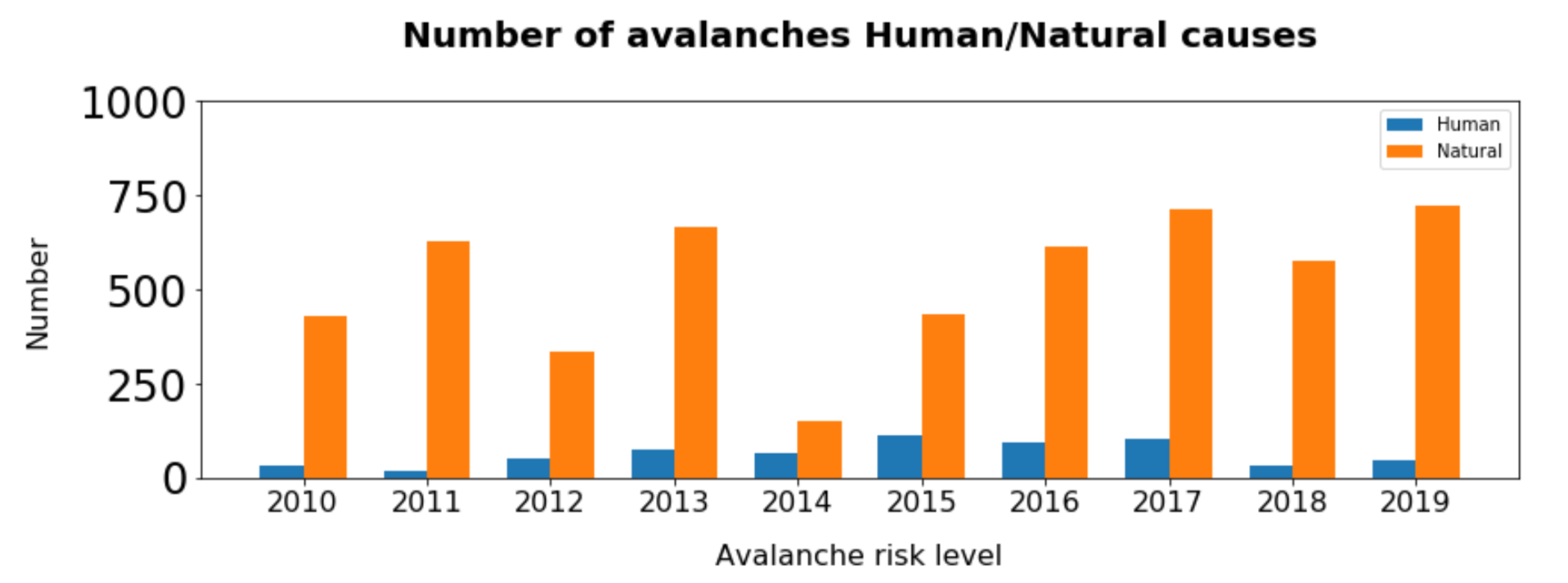
🡺There is only 5 days where “dry” avalanches have been recorded with risk level “1”. This is normal, as risk “1” means very low danger of avalanche (refer to picture …)

🡺There is only 6 days where “dry” avalanches have been recorded with risk level “5”. This is normal as risk level “5” happen only very rarerly

**🡺For the rest of this analysis and for module 2 analysis, the risk levels “1” and “5” are not considered and are removed.**

**Figure No. [7]** European avalanche danger level **Ref. [1]**

#### Number of avalanches caused by “HUMAN” and with natural causes in the past 10 years



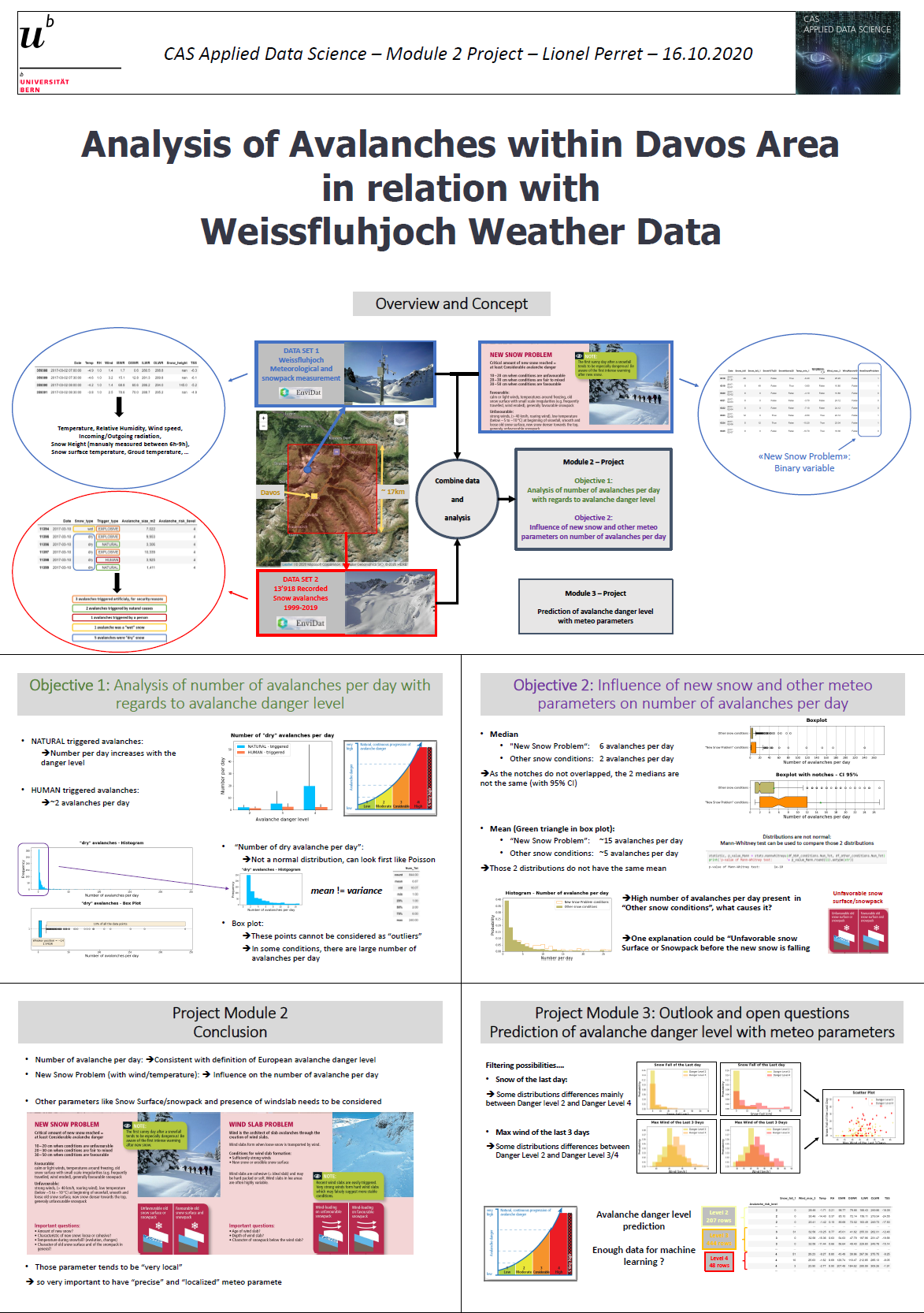
Over the years, the number of Natural – Human avalanche are fluctuating, and in the past 2 years, the number of “HUMAN” caused avalanches is decreasing.

## 9.2 Plots and Report for Module 2 of ADS CAS

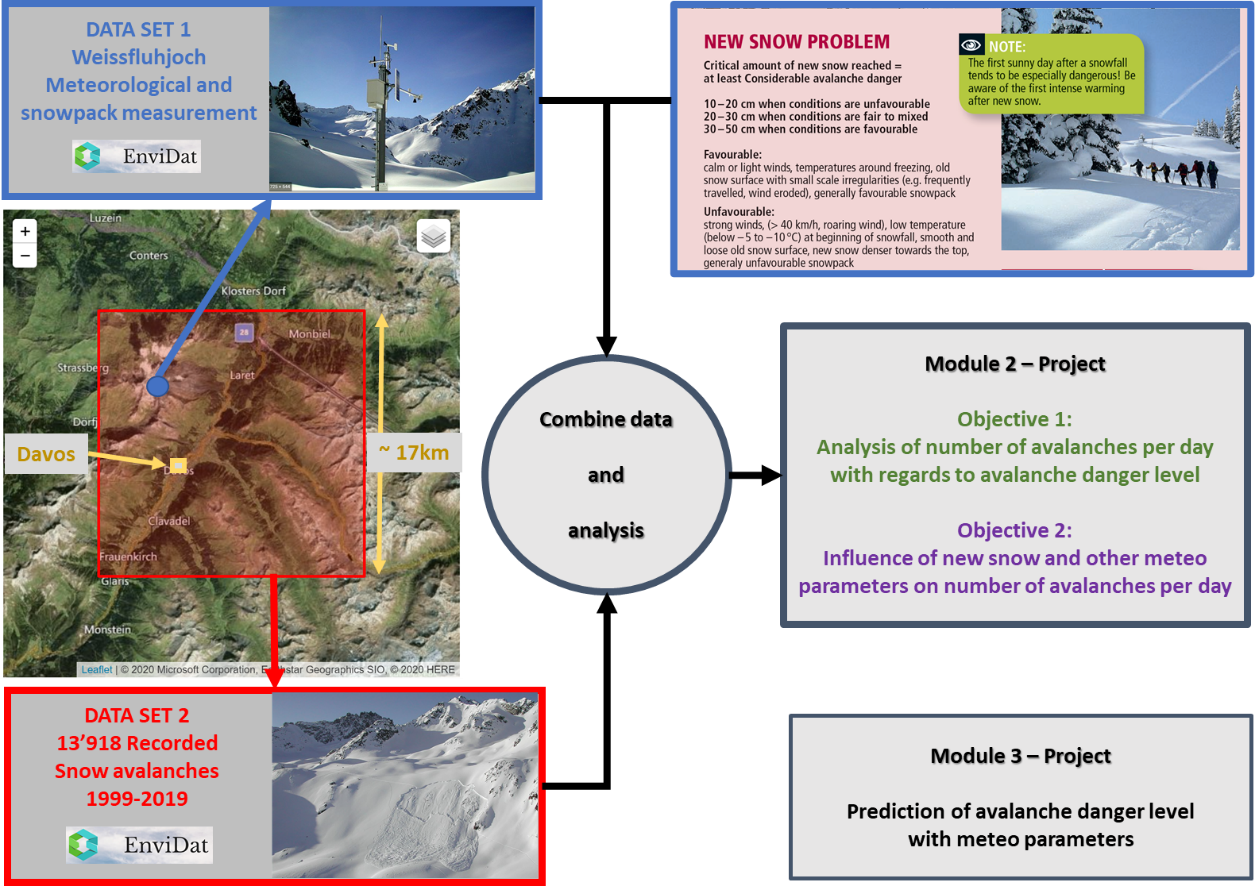
Module 2 Jupyter Notebook is available under:

Pdf of A1 format poster available under:

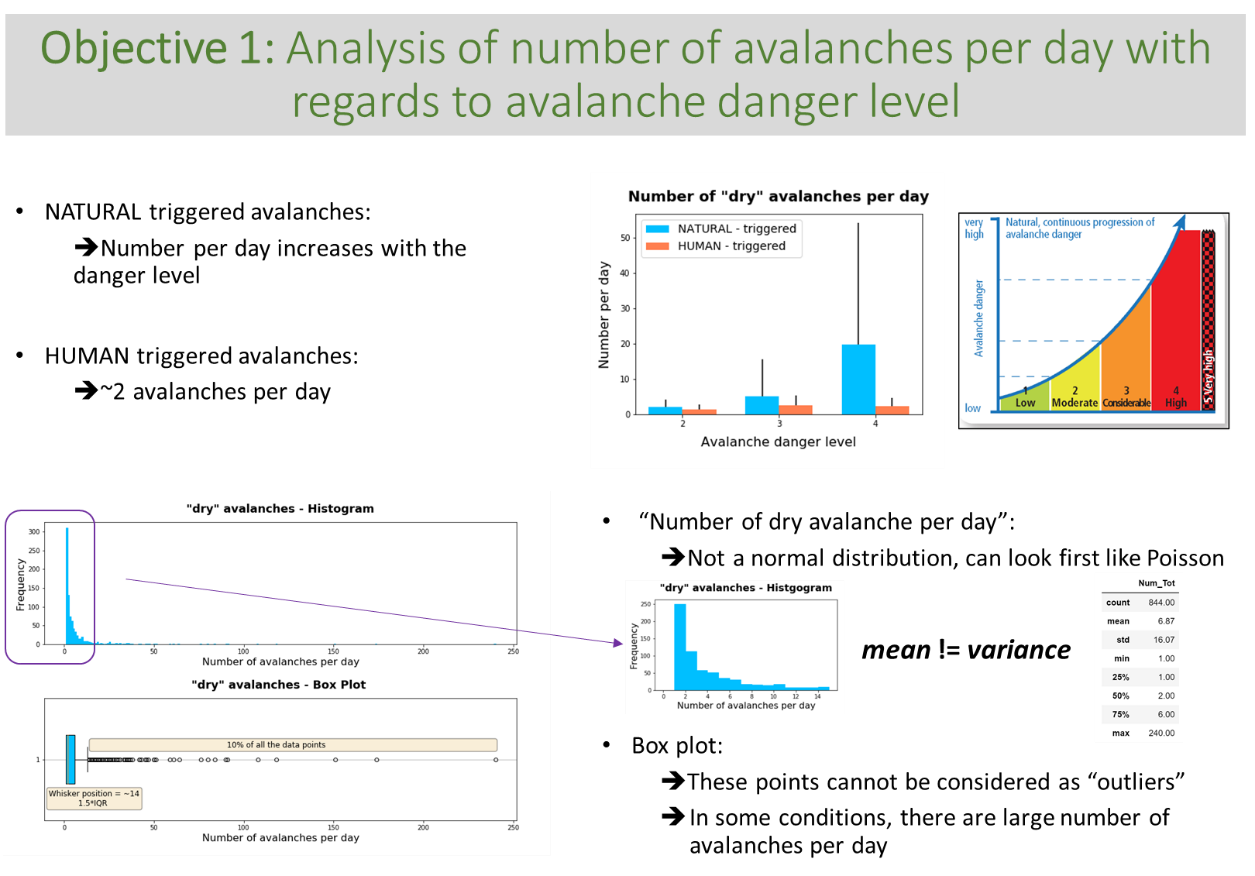
Different parts of the poster (powerpoint slides), pls refer to next pages



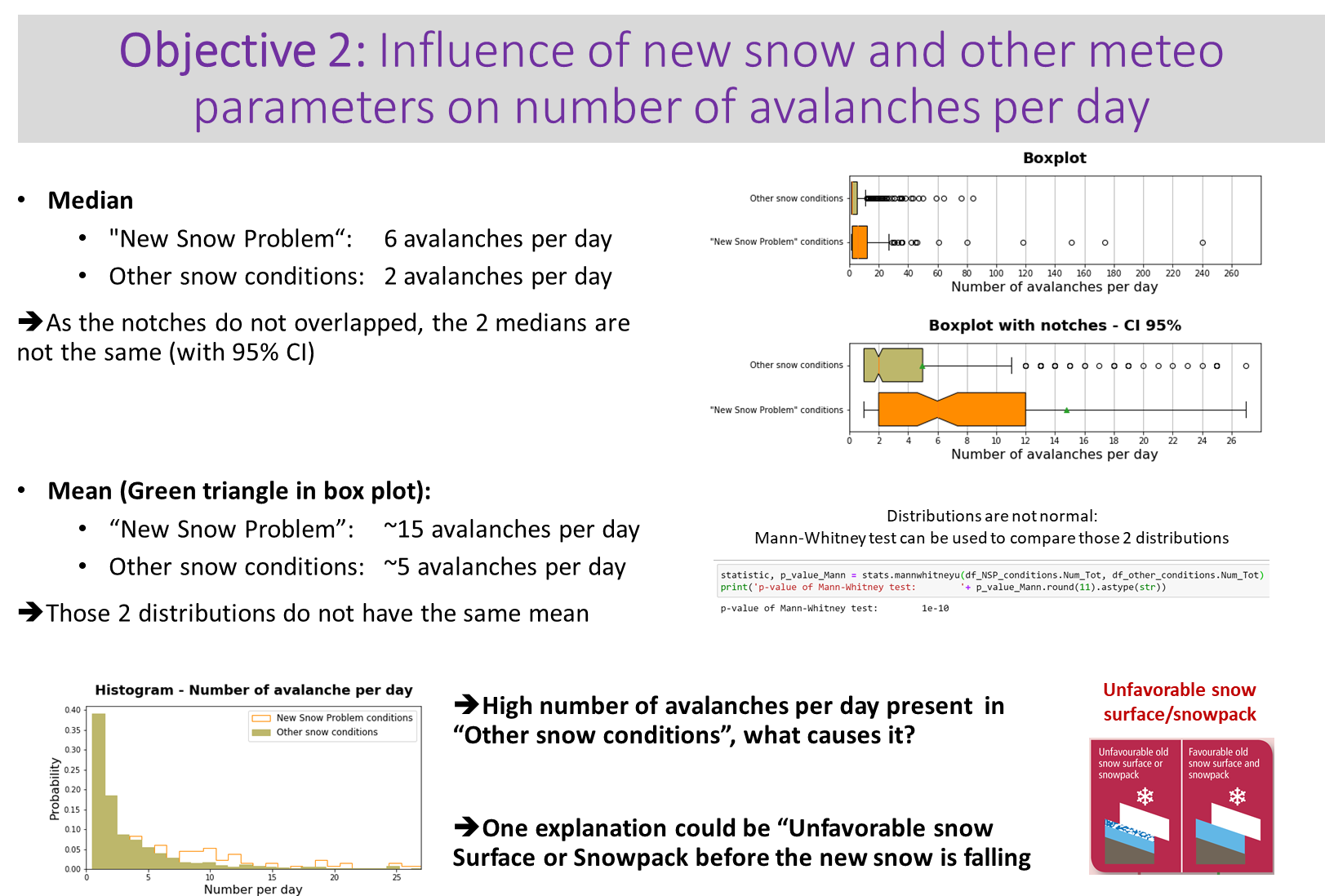
### 9.2.1 Overview and Concept



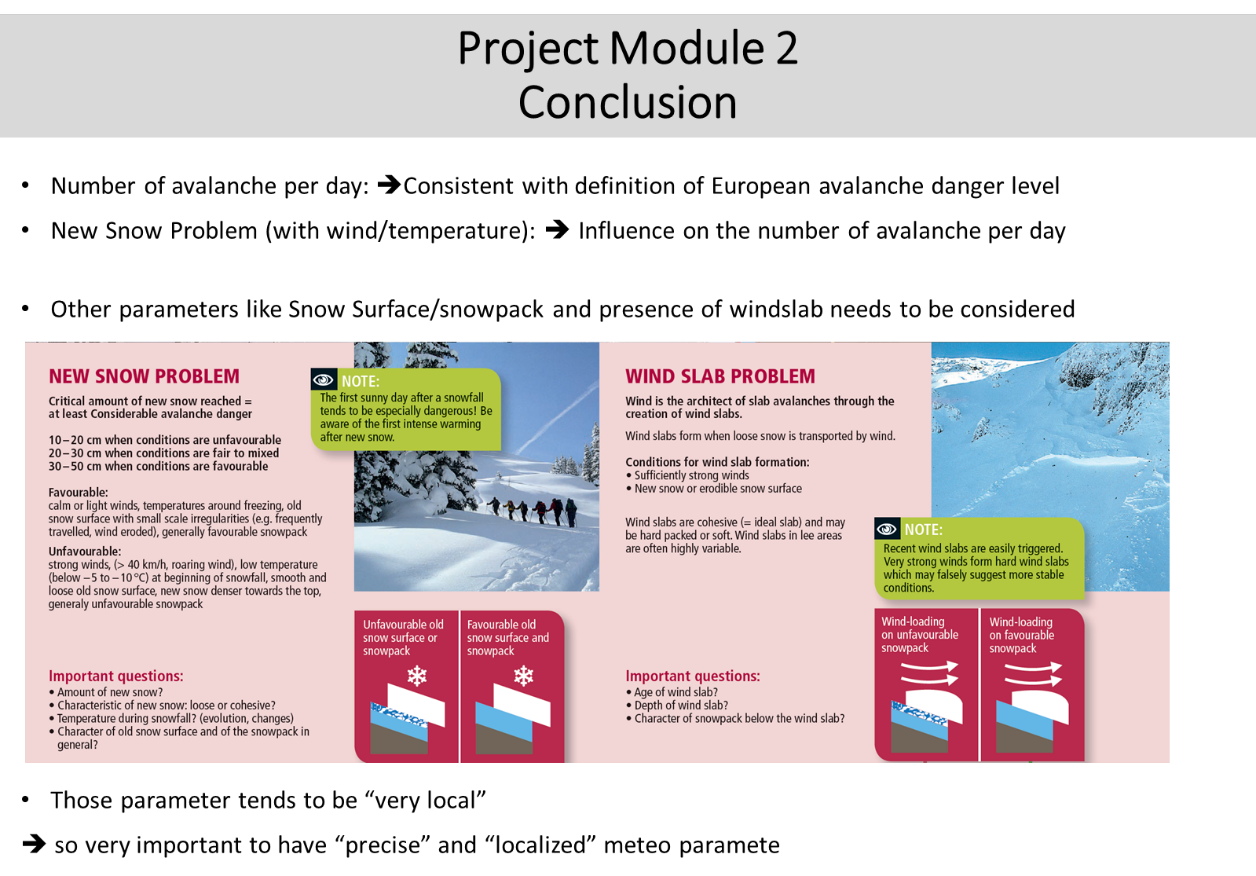
### 9.2.2 Objective 1



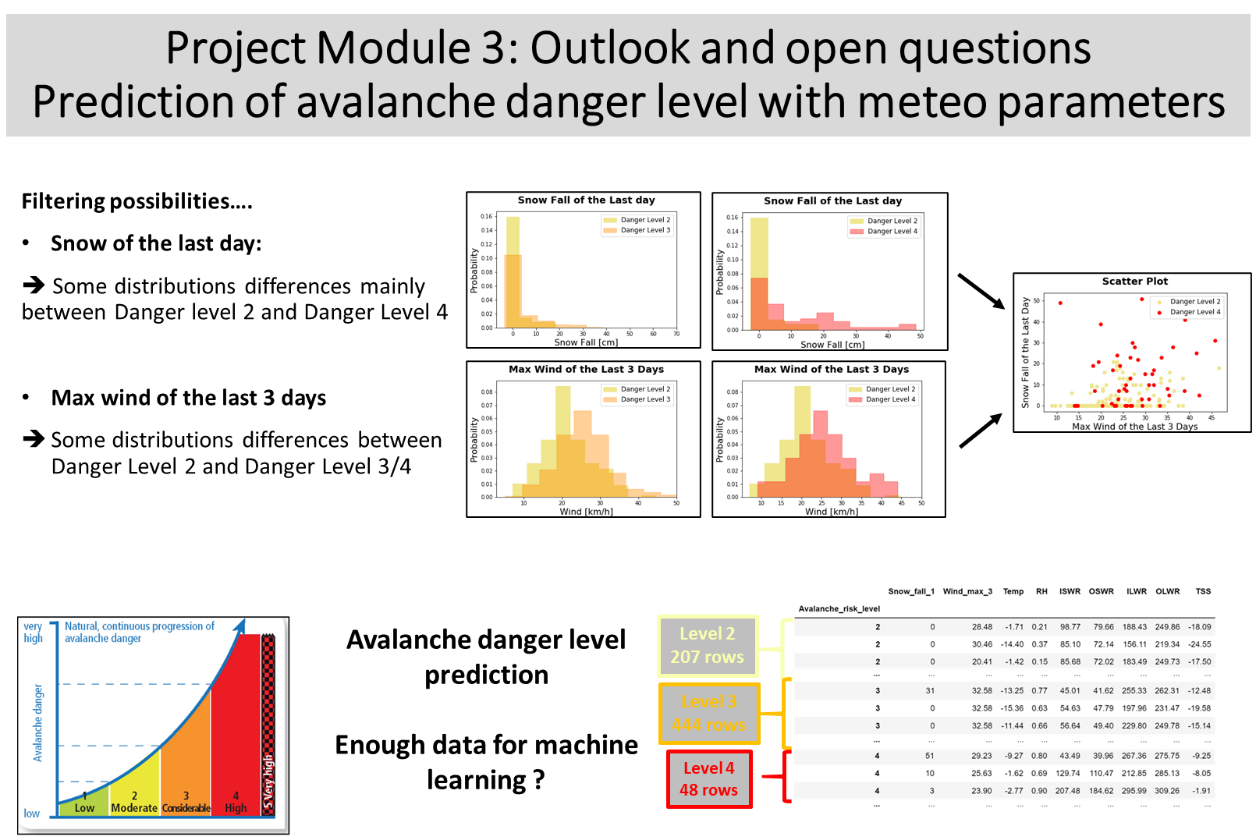
### 9.2.3 Objective 2



### 9.2.4 Project Module 2 Conclusion



### 9.2.5 Project Module 3: Outlook and open questions



# 10. Conclusions

….

# 11. Acknowledgements

I would like to thank the Institute for Snow and Avalanche Research – SLF (<https://www.slf.ch/en/index.html>)

as the information available on its website is very accurate, well presented and interesting. This was of great help not only for this study, but also for gathering knowledge and to go in the mountains with a safe and responsible approach.

# 12. References and Bibliography

**[1]** **Leaflet**  **“Caution Avalanches”**

Published by the SLF, Institute for Snow and Avalanche Research, Davos

<https://www.slf.ch/en/publicationssearch/books-and-brochures.html>

**[2]** **Data Set** ***WFJ\_MOD: Meteorological and snowpack measurements from***

***Weissfluhjoch, Davos, Switzerland*** <https://www.envidat.ch/#/metadata/10-16904-1>

Author: *Nander Weber*

**[3]** **Data Set** ***Snow avalanche data Davos, Switzerland, 1999-2019***

<https://www.envidat.ch/#/metadata/snow-avalanche-data-davos>

Authors: *Jürg Schweizer; Christoph Mitterer; Frank Techel; Andreas Stoffel; Benjamin Reuter*

[] Diverse Illustration: Web site of the Snow

# 13. Illustrations-Graphs List

[] https://www.slf.ch/en/avalanche-bulletin-and-snow-situation/about-the-avalanche-bulletin/danger-levels.html