# Data Mining och Data Science Laboratory instructions

**Read all of these instructions before you get started.**

*View and interact with Jupyter notebooks and learn about data mining techniques.*

The laboratory sessions for this course are run in Jupyter. It is an environment that is web based, and allows you to do interactive programming inside a Web browser. Jupyter allows you to view and create computational notebooks, which are like Web pages that contain cells that are static content and cells that are Python code that you can run and view the output in the browser.

**Prior Knowledge**
Data Mining and Data Science lectures.

**Software Requirements**
The notebooks are available in the course GitHub repository:
https://github.com/UppsalaIM/2IS063

*On the lab computers*
Anaconda is a Python and R software distribution for data science that includes many pre-installed tools that we need to run the notebooks on a computer. For running on the lab computers, you will need only:

- Anaconda (Python 3.7 version)
- Graphviz binary installation.

These are already installed on the lab computers. You do not need to install the above listed software yourself.

*On the cloud using Binder*
Binder is a hosted service that takes care of all the installation for you, and runs the notebooks on the cloud. For running on the cloud using Binder we will need only:

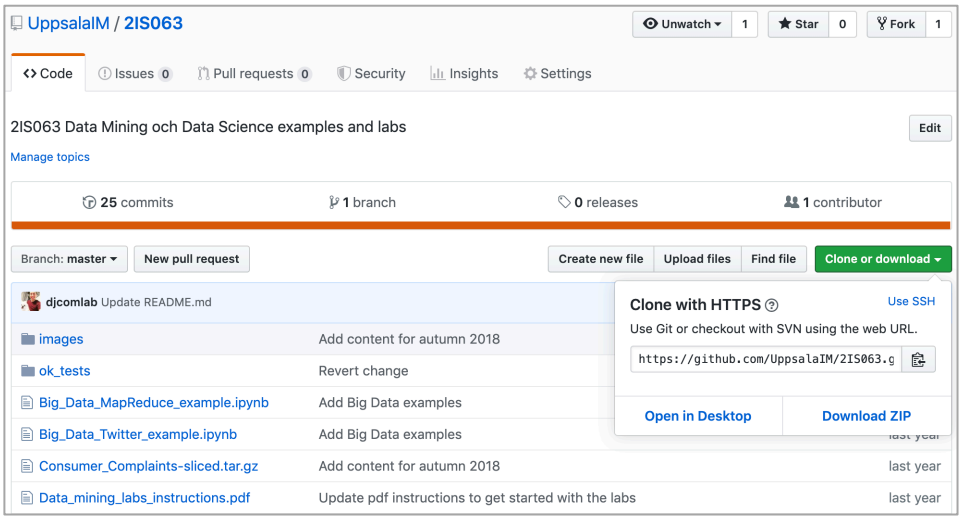
- A Web browser: Google Chrome or Mozilla Firefox.

*On your own computer*
Full installation instructions can be found in the `INSTALL.md` file contained in the GitHub repository.

**Getting started using the lab computers**
To run the notebooks, we will need to download the notebooks repository on to the lab computer we will use and install the required Python libraries that the notebooks will use. To do this:

1. Visit the GitHub source code repository for the Labs at https://github.com/UppsalaIM/2IS063

   The top of the page should look like this:




2. If you click on the green **`Clone or download`** button, it will open up a few options. Select **`Download ZIP`** to download the repository to a file called `2IS063-master.zip`

3. Locate the file in **`Downloads`** and right-click to select **`Extract All...`** and extract the archive to **`This PC → Documents`**

4. Find and open the application **`Anaconda Navigator (Anaconda3)`**

5. Select the **`Environments`** tab, then press the **`base (root)`** triangle ▶ to select **`Open Terminal`**, which will open up a command-line terminal.

6. Type into the command-line:

   ```
   pip install -r Documents\2IS063-master\requirements.txt
   ```

   and then press enter. If all went well, this installs the Python libraries we need for the labs to the `base (root)` environment of Anaconda.

7. Back in Anaconda Navigator, select the **`Home`** tab then **`Launch`** under Jupyter Notebook. Once launched, you should see your Windows home folder. You will need to then browse to the `Documents`, then the `2IS063-master` subfolders to find the lab notebooks in the Jupyter dashboard.

   You should now be ready to run the labs on the lab computers. If everything was set up correctly, you should only need to do the above steps *once*.

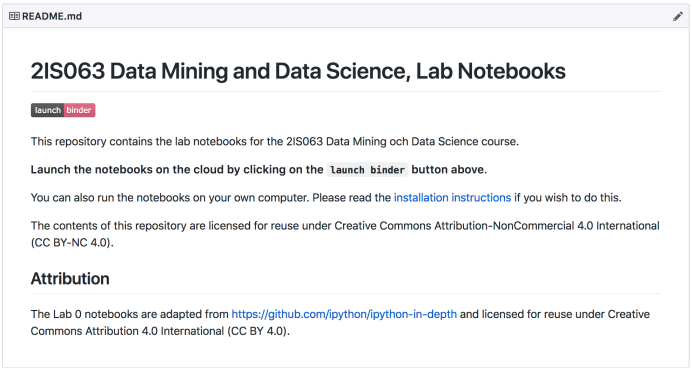
   Now, skip to the section below *Do the exercises for each lab notebook*.

### Getting started using Binder

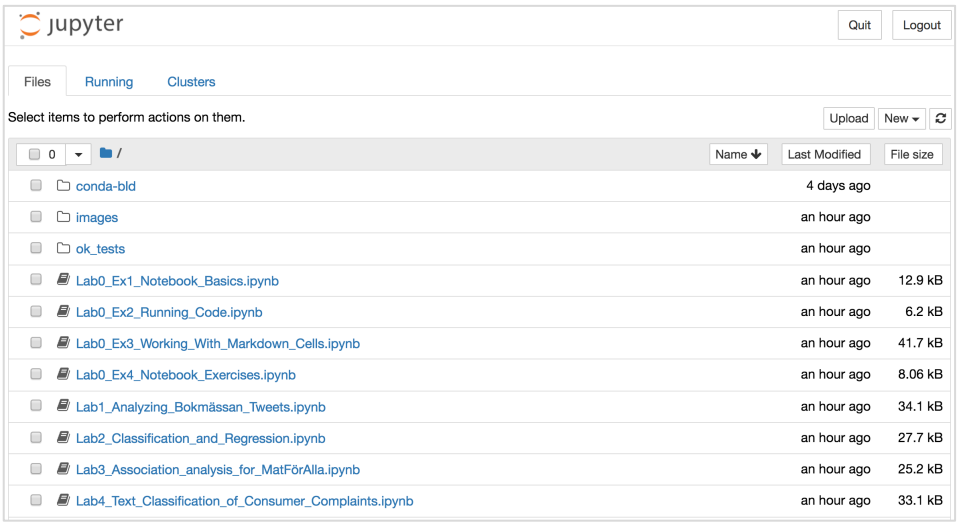
An alternative way to run the notebooks is open the notebooks repository on the cloud using Binder. To do this:

1. Visit the GitHub source code repository for the Labs at https://github.com/UppsalaIM/2IS063

   Scroll down and you should see some content that looks like this:




2. Click on the **`launch binder`** button and wait. Once launched, you should see the repository contents in the Jupyter dashboard.




## Do the exercises for each lab notebook

To launch a specific notebook, click on any of the **`.ipynb`** files in the dashboard.

1. If you have not already done so, make sure to familiarize yourself with the Jupyter environment by doing the Lab 0 exercises 1-4

   i.e. Open up, read and complete `Lab0_Ex1_Notebook_Basics.ipynb`, `Lab0_Ex2...`, `Lab0_Ex3...` etc.

2. Once you are familiar with the Jupyter environment, open the Lab notebook corresponding to the session and complete the questions in the notebook.

   e.g. During Lab 1, open `Lab1_Analyzing_Bökmassan_Tweets.ipynb`, etc.

3. When you are finished, get the attention of a teaching assistant or the lecturer to have a brief discussion about the questions you answered in the notebook.

**If you are stuck, do not hesitate to ask for help from one of the teaching assistants (Evelina and Martin) or the lecturer (David)! Good luck!**