Data Mining I: Introduction to Python

Installation

- Install Anaconda (Python Distribution)
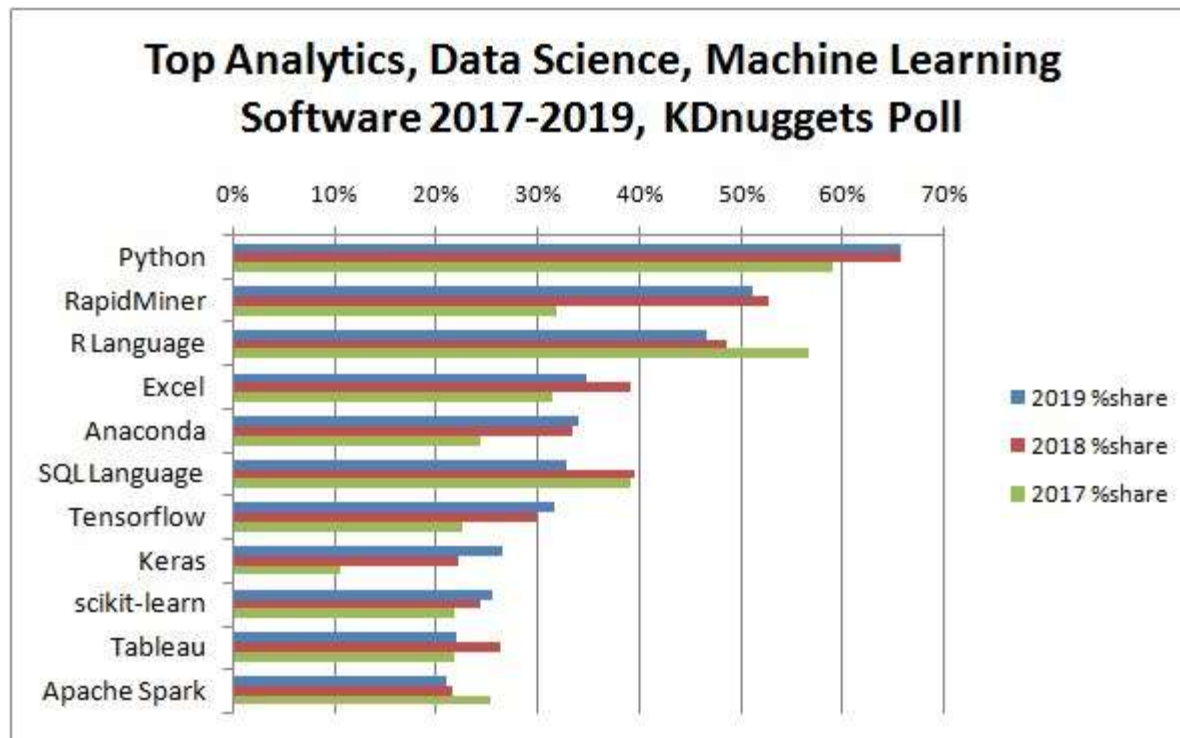 - https://www.anaconda.com/distribution/
  - Use Python 3.x

Anaconda Installers

Windows ⊞	MacOS	Linux △
Python 3.8	Python 3.8	Python 3.8
64-Bit Graphical Installer (466 MB)	64-Bit Graphical Installer (462 MB)	64-Bit (x86) Installer (550 MB)
32-Bit Graphical Installer (397 MB)	64-Bit Command Line Installer (454 MB)	64-Bit (Power8 and Power9) Installer (290 MB)

- If you don't have at least 3 GB disc space
  - Option 1 (better): Get a bigger disc!
  - Option 2: install miniconda
    - https://docs.conda.io/en/latest/miniconda.html

- Alternative: Use Google Colab

Python

- Started in 1989 by Guido van Rossum
  - The name is a tribute to the British comedy group Monty Python

- High-level, general-purpose programming language
  - Multi-paradigm: functional, imperative, object-oriented, reflective

- Design goals
  - Be extensible, simple, and readable

Popularity

Top Analytics, Data Science, Machine Learning Software 2017-2019, KDnuggets Poll

Source: https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html

Jupyter notebooks

- In the exercises, we will use Jupyter Notebooks
- The start procedure depends on your operating system
  - in general: run the **Anaconda Navigator** and click on **jupyter Notebook**
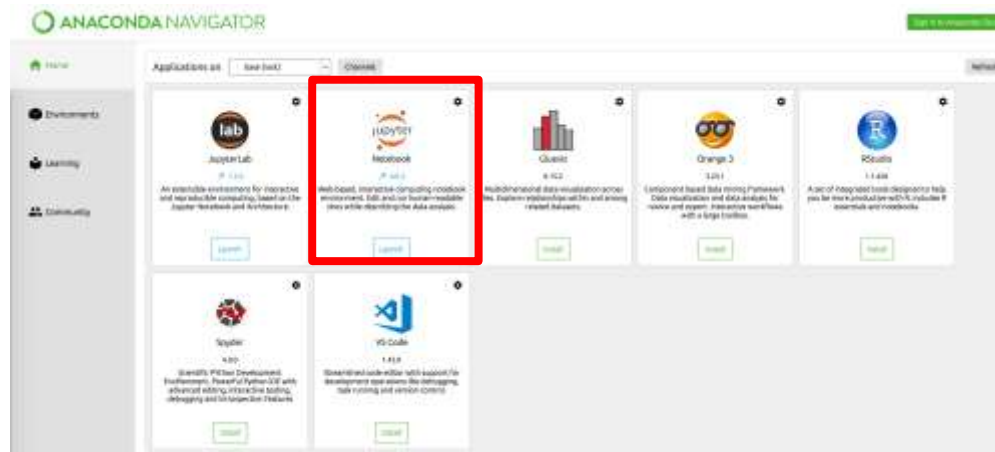
  - The Jupyter Notebook App can access only files within its start-up folder (including any sub-folder)
    - default is your home folder
      - Windows: usually C:\Users\{username}   Linux: /home/{username}

Jupyter notebooks

- Start Jupyter - Option 1 (Windows)

  - Click on the *Jupyter Notebook* icon in the start menu
  - To change this folder:
    - Copy the Jupyter Notebook launcher from the menu to the desktop.
    - Right click on the new launcher and change the Target field, change %USERPROFILE% to the full path of the folder which will contain all the notebooks.
    - Use the *Jupyter Notebook* desktop launcher to start the notebook

Jupyter notebooks

- Start Jupyter – Option 2 (Linux and Windows)
  - Run „jupyter notebook" in command line
    - Navigate to folder that you want to access before!
    - Or (Windows):
      1. Open the Anaconda Promt
      2. Copy the path to your folder
      3. Write "cd " and then make a **right** mouse click into the terminal folder to insert the copied path
      1. then type "jupyter notebook" + enter
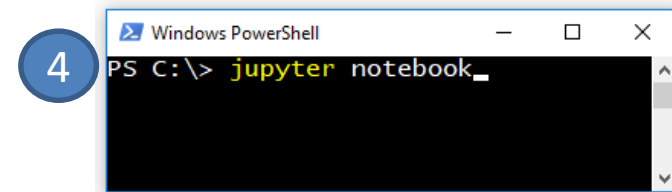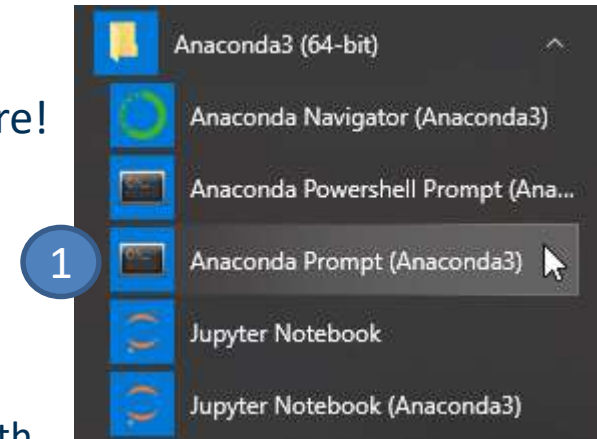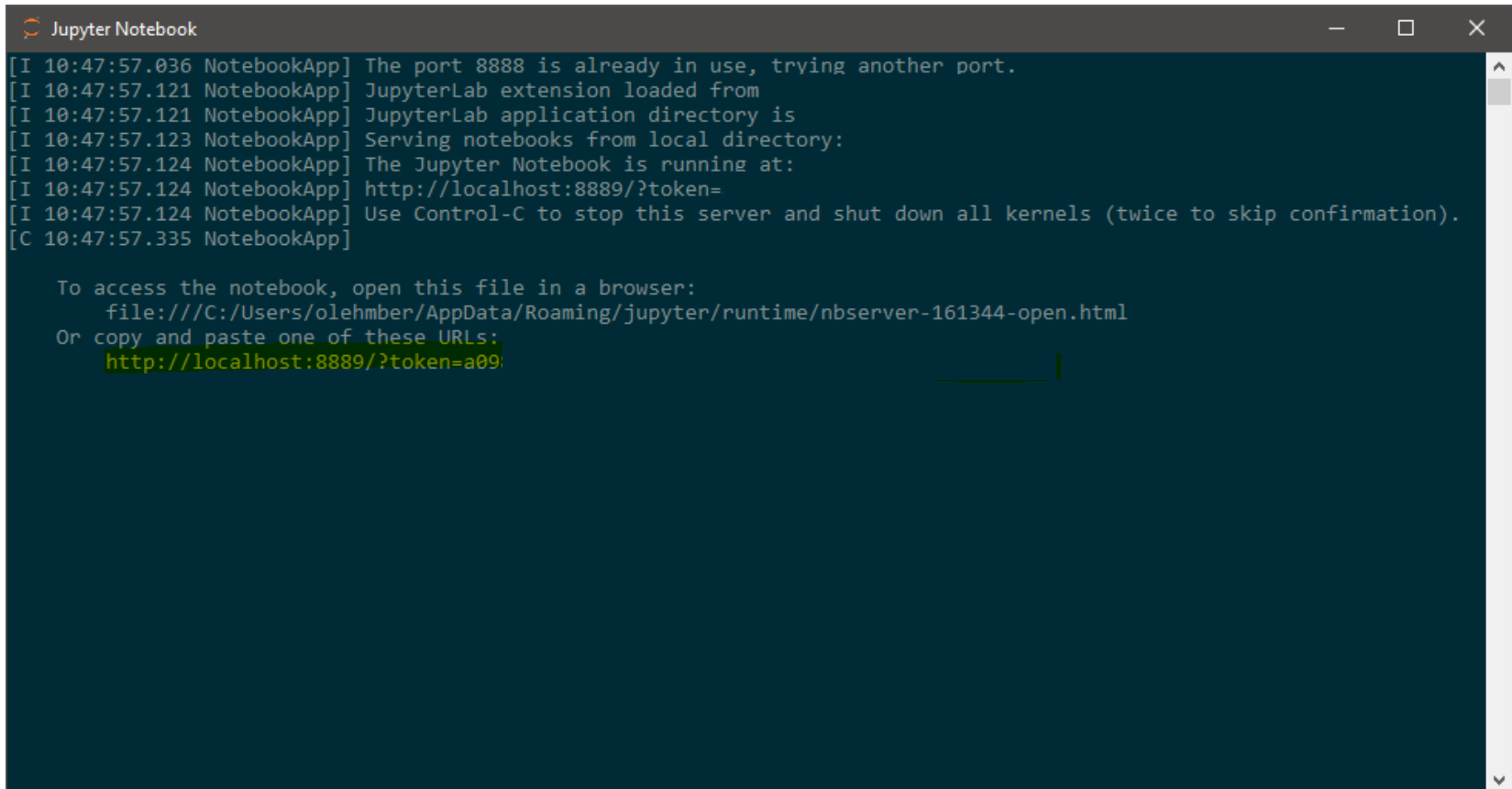
- Start Jupyter – Option 3 (Mac OS)
  - Click on spotlight, type "terminal" to open a terminal window
  - Enter the startup folder by typing "cd /some_folder_name".
  - Type "jupyter notebook" to launch the Jupyter Notebook App

Jupyter notebooks

- A local server is started
- Open the URL on screen in your browser, if not already opened

Jupyter Home Screen

- Startscreen in browser
  - like a file explorer

upload files to notebook folder

Create new files/notebooks

file explorer

Now try it out

- Click in browser „New" -> „Python 3"
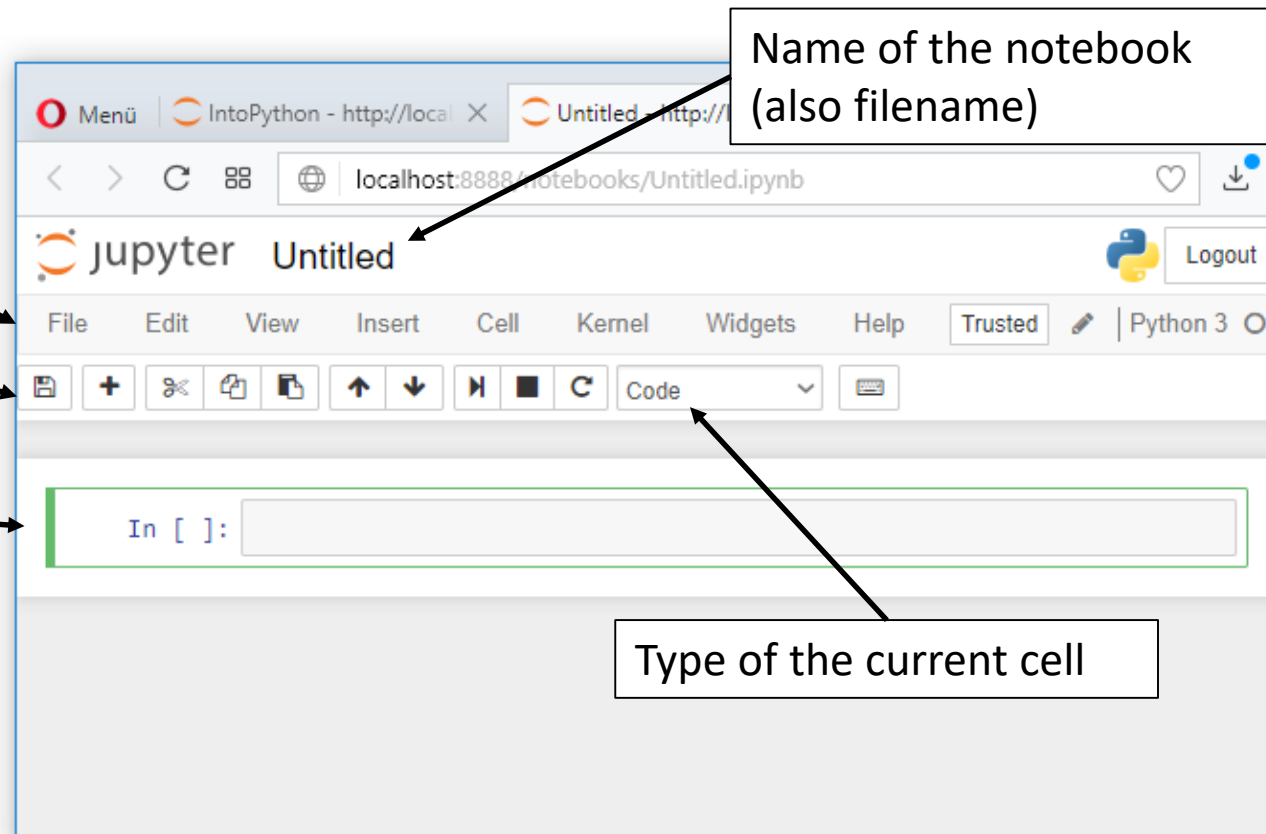
https://xkcd.com/353/
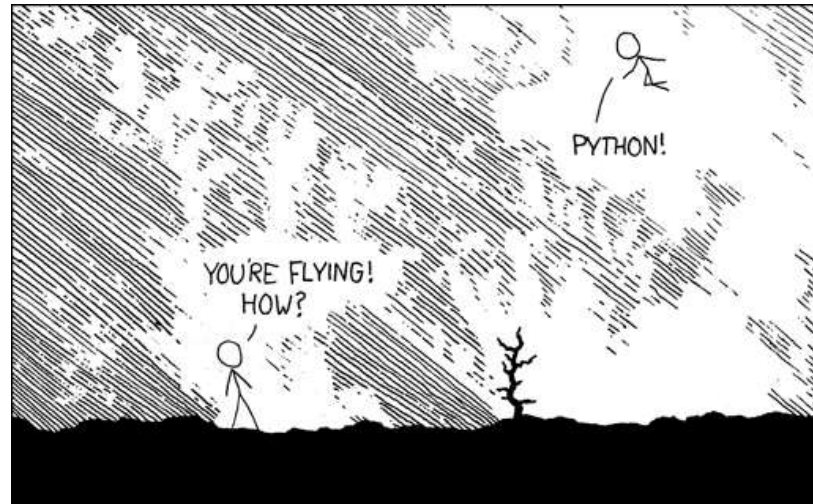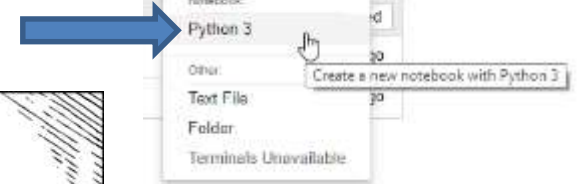
Jupyter Notebook

- Every notebook is composed of cells
  - Cells contain a specific type of content
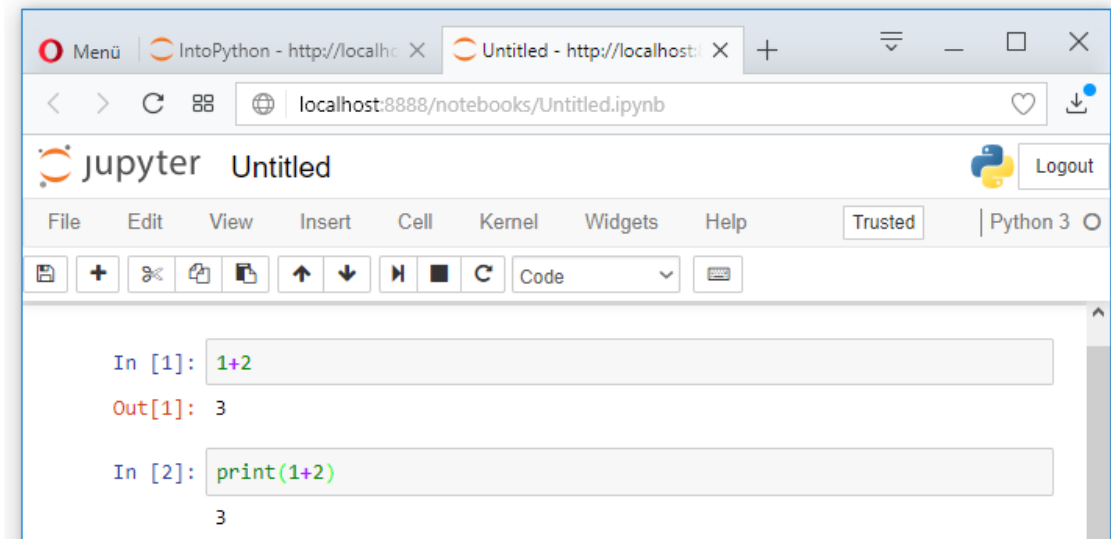 - markdown cells (for documentation and structure)
  - code cells

Name of the notebook (also filename)

Menu Bar

Shortcuts

Cell

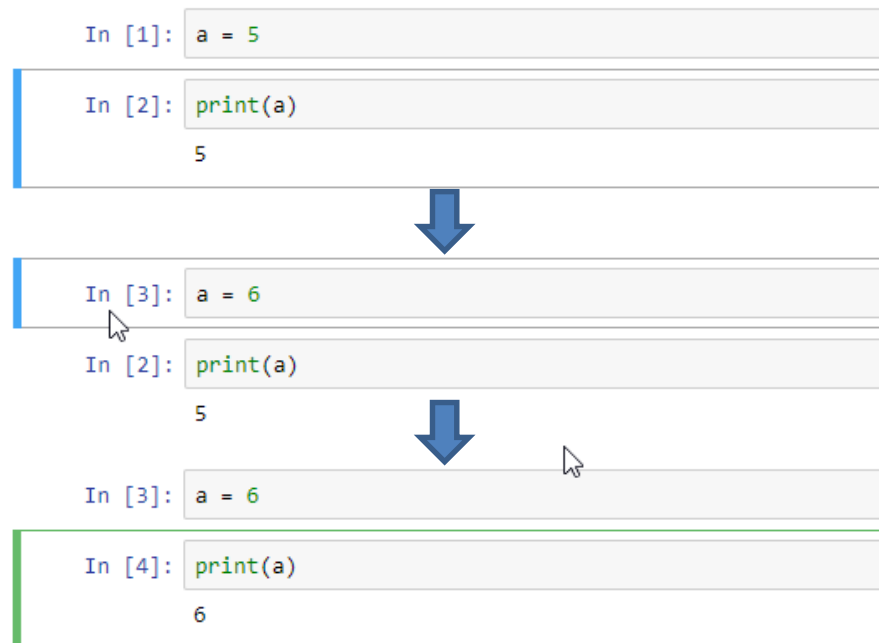Type of the current cell

Jupyter Cells

- Code cell:
  - You can type python code (because you created a python notebook)
    - Hit „Ctrl + Enter" to run the code
    - Hit „Shift + Enter" to run it and create a new cell
    - Try it and type 1 + 2
  - The output is shown below the cell

Jupyter Cells

- Each „code cell" can be reevaluated (indicated by a number)
  - All previous results / variables are stored (like in R workspace)

```
In [1]:  a = 5

In [2]:  print(a)
         5
```

Change code
and run again
[1] -> [3]

```
In [3]:  a = 6

In [2]:  print(a)
         5
```

run next cell
again [2] -> [4]

```
In [3]:  a = 6

In [4]:  print(a)
         6
```

Jupyter Cells

- Autocomplete by pressing <tab> when writing

- Signature of function by pressing <shift>+<tab>

Jupyter Cells

- What makes a notebook a notebook?
  - Markdown cells let you add documentation and notes
  - Create a new cell („Insert->Insert Cell Below")
  - Change the type to Markdown
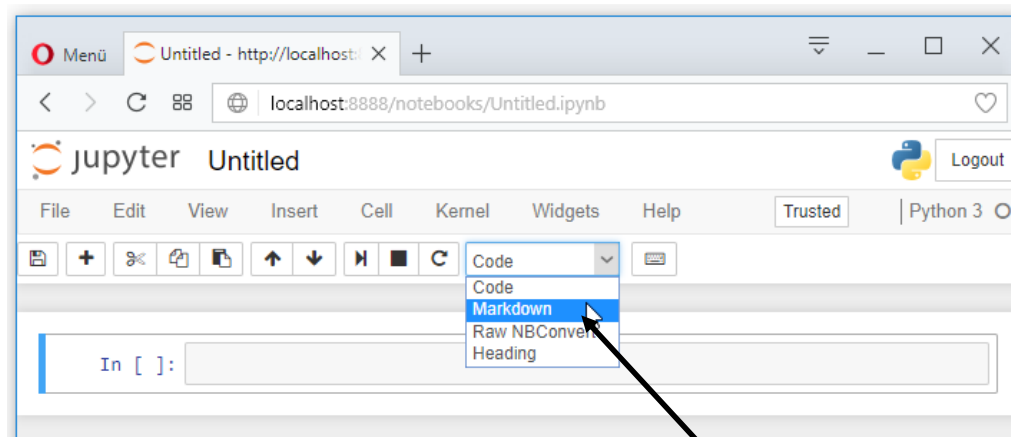
Type of the current cell

Jupyter Cells

- What makes a notebook a notebook?
  - Type „# Test" which creates a heading (add more „#" for smaller headline)
    - Whitespace after #
  - Evaluate the cell and see the result
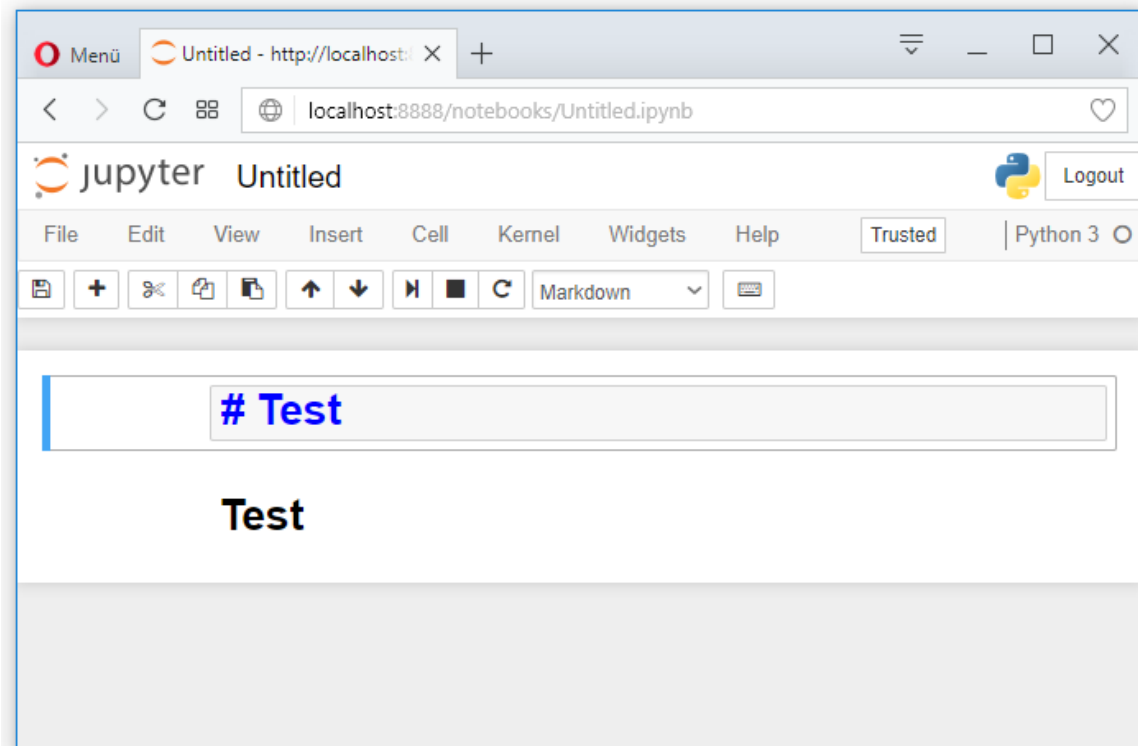
Jupyter Cells - Markdown

- Different possibilities to structure

  - Header

    ```
    # H1
    ## H2
    ### H3
    ```

  - Unordered List (use "*", "+", or "−" in front)

    ```
    -   Item
    -   Item
    ```

  - Ordered list

    ```
    1. Item one
    2. Item two
    ```

  - Links

    ```
    [link to google](https://www.google.com)
    ```
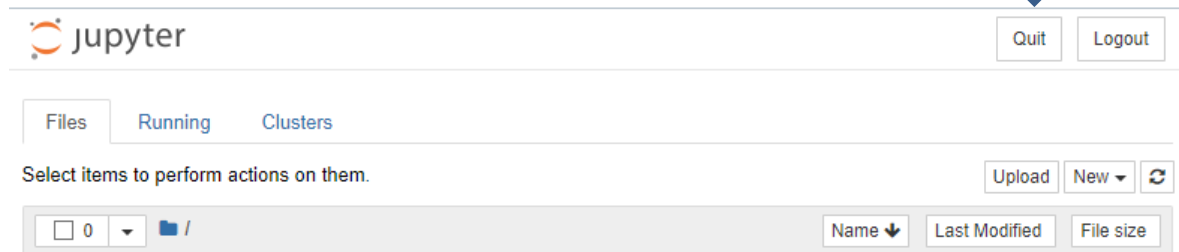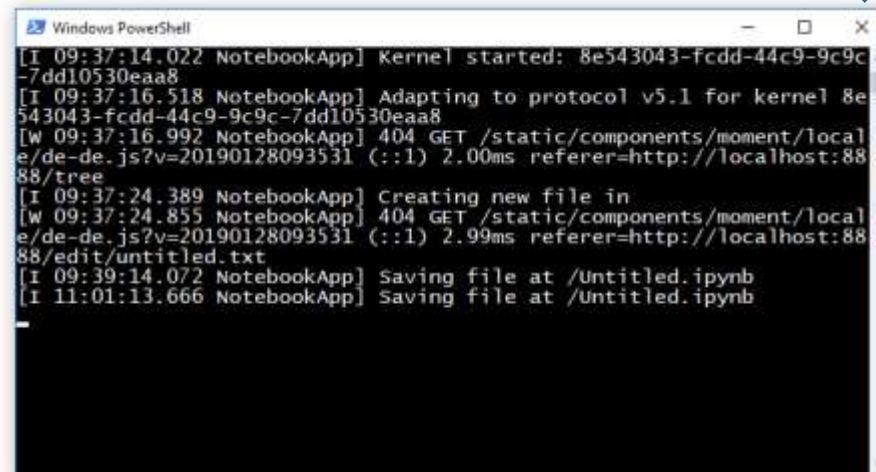
  - Image

    ```
    ![Mannheim Image](https://www.uni-mannheim.de/1/00_UM_Dachmarke_DE_RGB.jpg)
    ```

  - Quote

    ```
    > This is a quotation
    ```

Shut down Jupyter

- Closing the browser (or the tab) will not close the Jupyter server
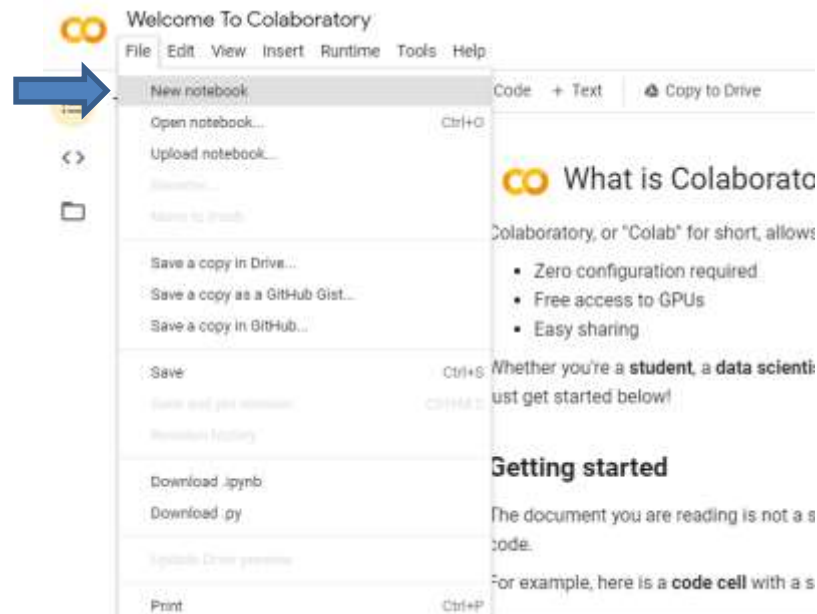- Option 1: click on **Quit** in the jupyter homepage

- Option 2: close the associated terminal or press "Ctrl" + "C"

Google Colab

- Runs in the cloud on Google Servers and free to use with Google account

- Uses jupyter notebooks with a modified interface, which are saved in your google drive
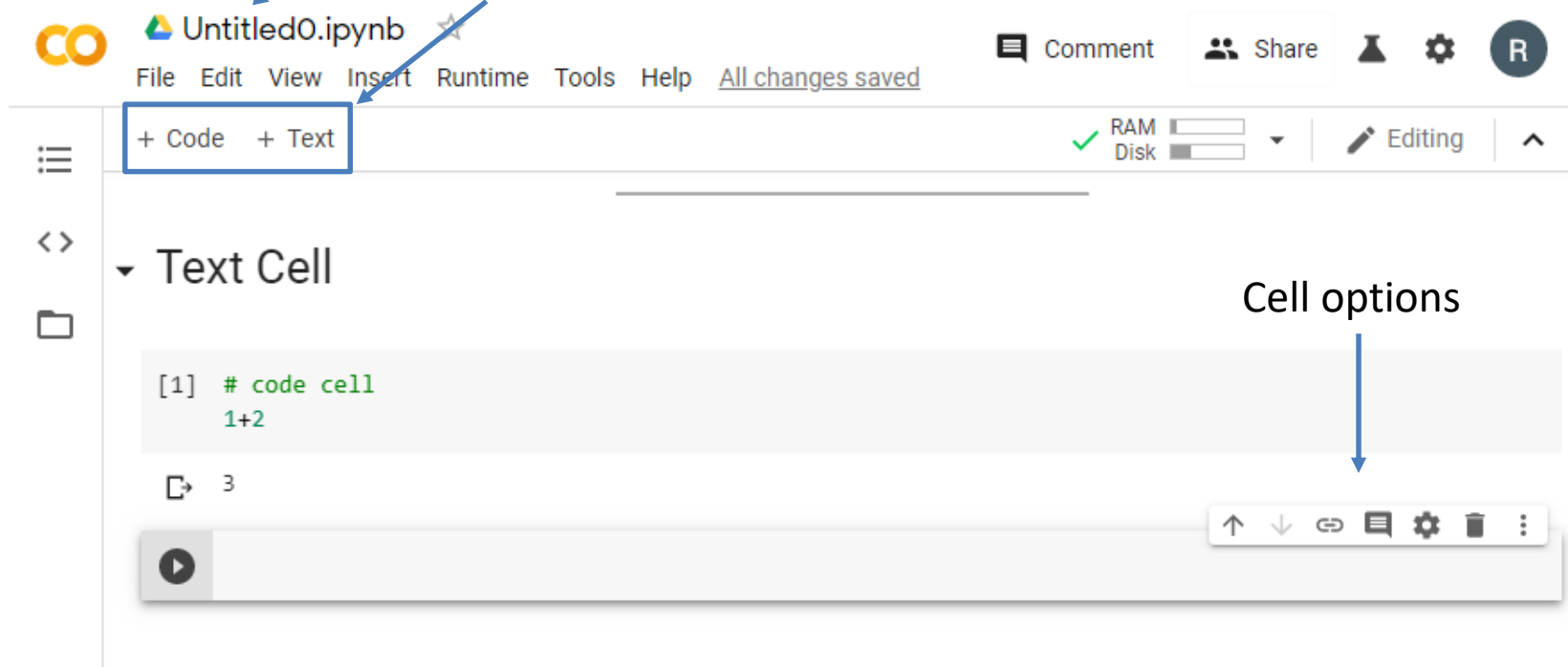
Create new notebook

Google Colab Layout

- Code and Text cells
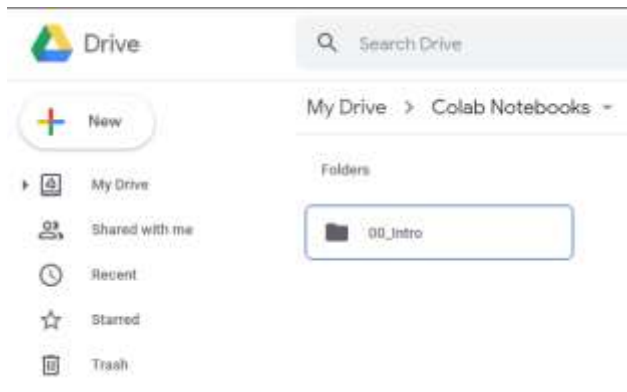- Shortcuts like Shift+Enter work the same way
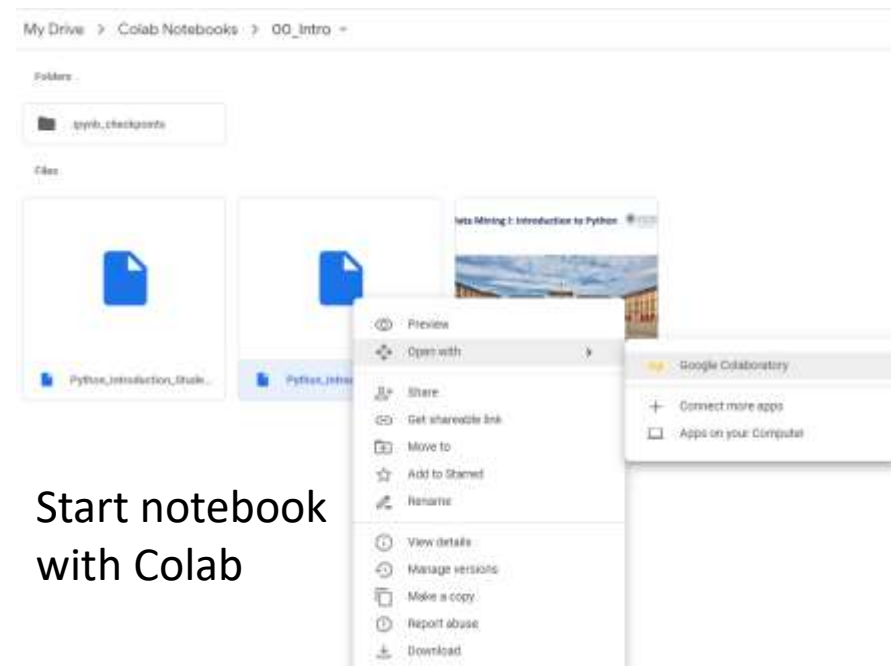

Change name of notebook

Insert code or text cell

Cell options

Google Colab

- Saves notebooks to folder "Colab Notebooks" on your Google Drive by default

- To facilitate exercises and have stuff "just work": upload/unzip exercise material in this folder

File Location

Start notebook with Colab