

Data Mining I: Introduction to Python



Python

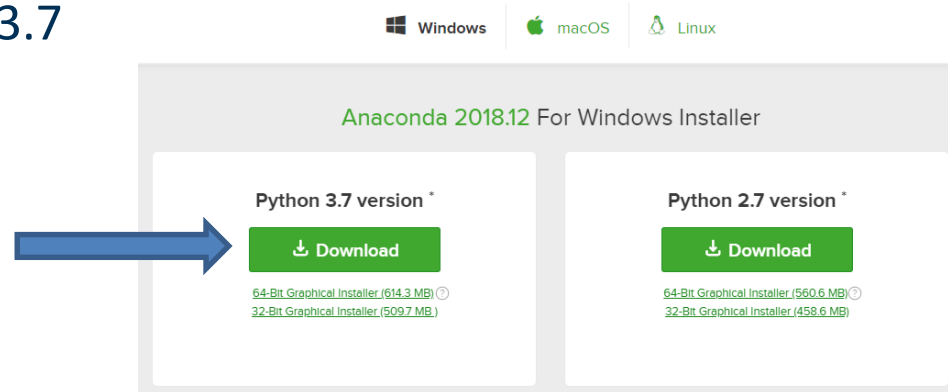
- Started in 1989 by Guido van Rossum
 - The name is a tribute to the British comedy group Monty Python
- Multi-paradigm programming language
 - object-oriented, structured, functional, aspect-oriented programming
 - even more supported by extensions
- Design goals
 - Be extensible, simple, and readable



Installation

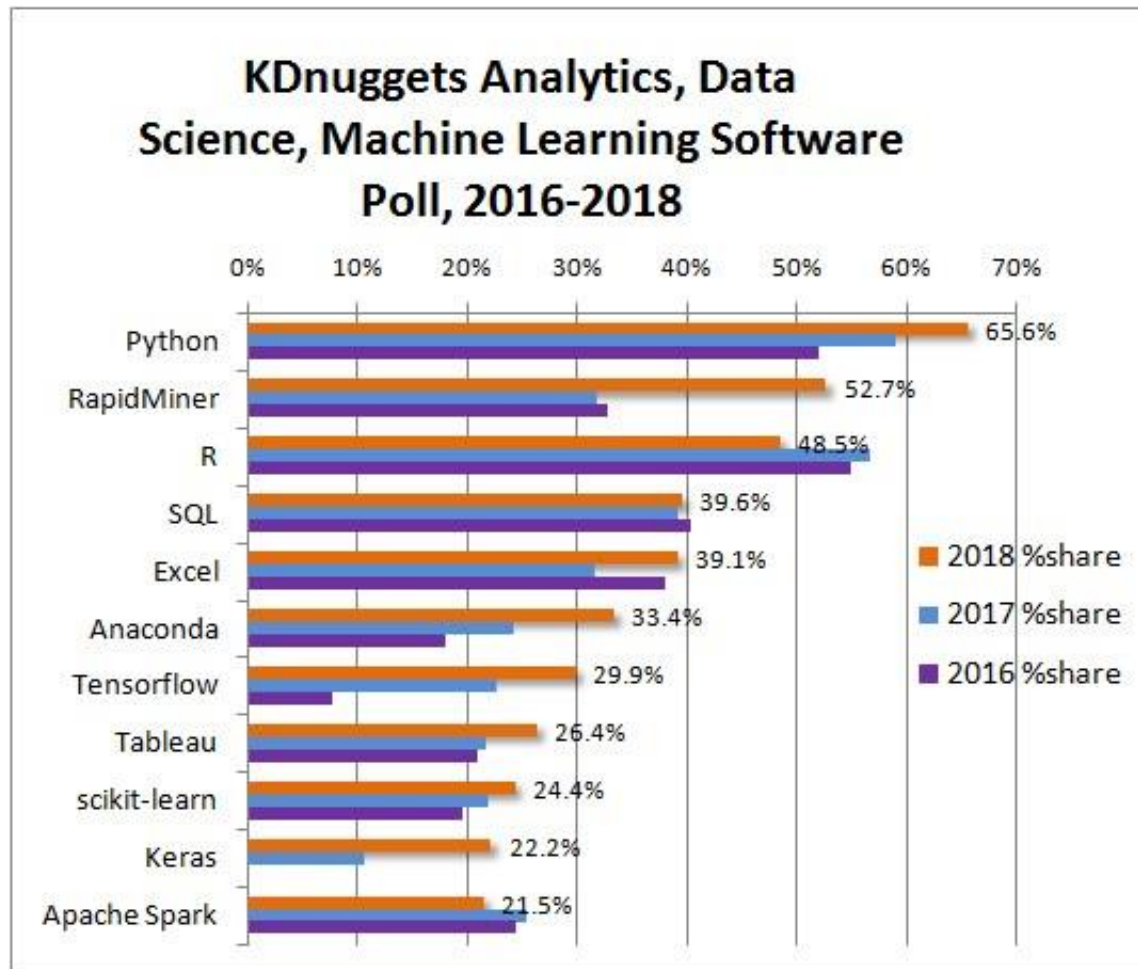


- Install Anaconda (Python Distribution)
 - <https://www.anaconda.com/download/>
 - Use Python 3.7



- 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>

Popularity



What does it look like?

```
> print("Hello, world!")  
Hello, world!
```

```
> 1 + 2  
3
```

```
> x = 5  
> 3*x  
15
```

```
> if x > 3:  
>   print("greater than 3")  
> else:  
>   print("less or equal 3")  
greater than 3
```

```
> list = [1,2,3]  
> for x in list:  
>   x*5  
5  
10  
15
```

How do I do it?

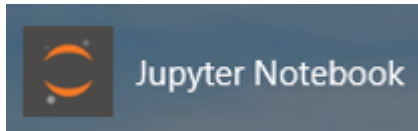
- After installation, you should have python on your path
 - Just type “python” in your command line to start it

```
(base) olember@wifo5-16:~$ python
Python 3.7.1 (default, Dec 14 2018, 19:28:38)
[GCC 7.3.0] :: Anaconda, Inc. on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Hello, world!")
Hello, world!
>>>
```

- In the exercises, we will use Jupyter Notebooks
 - Type “jupyter notebook” in your command line
 - More on the following slides!

Hands on

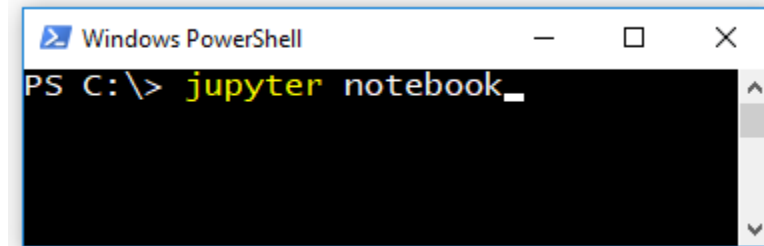
- Start Jupyter - Option 1 (Windows)



- Click on the *Jupyter Notebook* icon in the start menu
- The Jupyter Notebook App can access only files within its start-up folder (including any sub-folder)
 - default is your home folder (usually C:\Users\{username})
- 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

Hands on

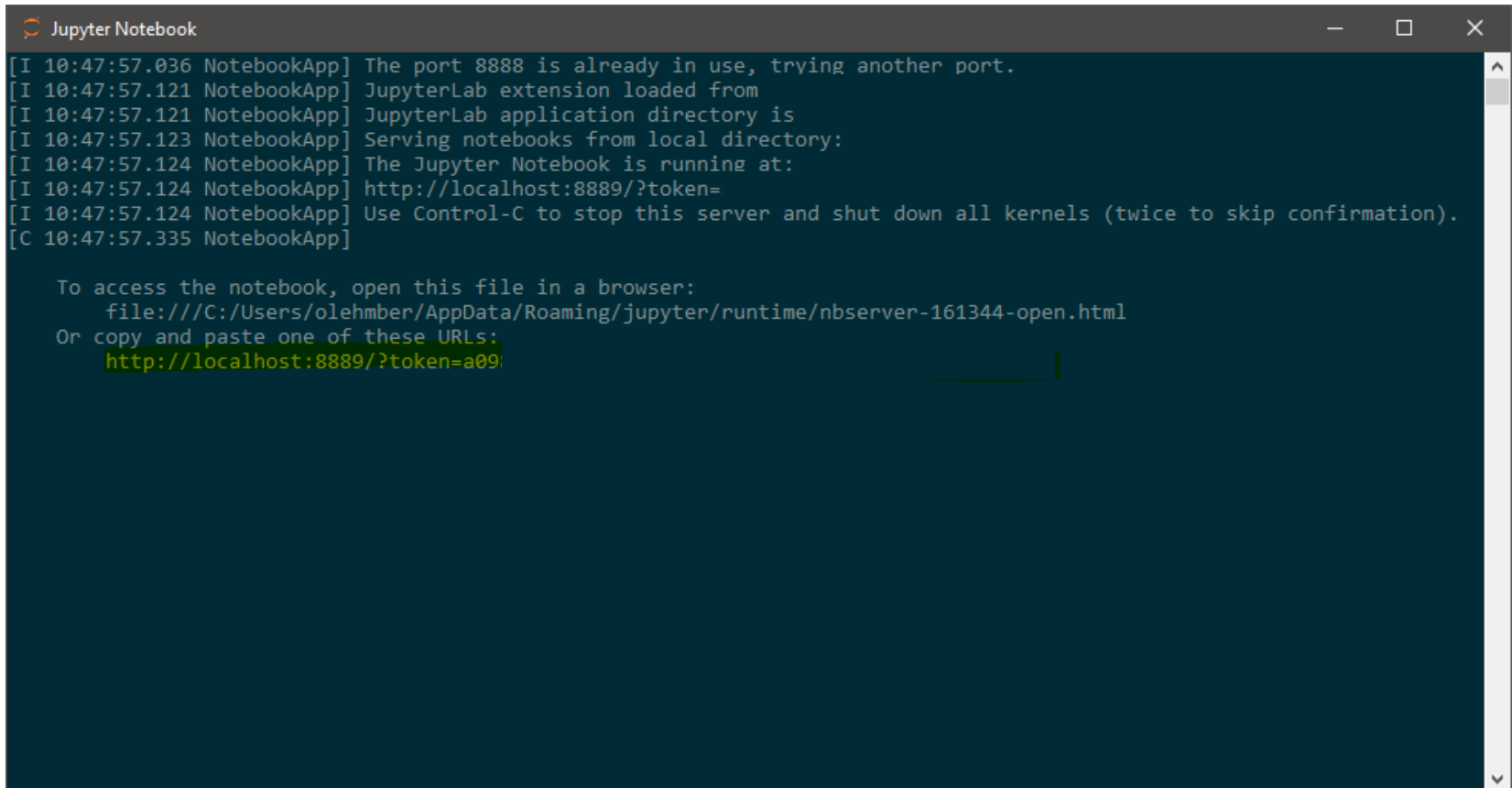
- Start Jupyter – Option 2 (Linux and Windows)
 - Run „jupyter notebook“ in command line
 - Navigate to the folder that you want to access before!
 - Or (Windows): “Shift+Right Click” in the corresponding folder and then “open command window/power shell here”



- 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

Hands on

- Open the URL on screen in your browser, if not already opened



```
Jupyter Notebook
[I 10:47:57.036 NotebookApp] The port 8888 is already in use, trying another port.
[I 10:47:57.121 NotebookApp] JupyterLab extension loaded from
[I 10:47:57.121 NotebookApp] JupyterLab application directory is
[I 10:47:57.123 NotebookApp] Serving notebooks from local directory:
[I 10:47:57.124 NotebookApp] The Jupyter Notebook is running at:
[I 10:47:57.124 NotebookApp] http://localhost:8889/?token=
[I 10:47:57.124 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 10:47:57.335 NotebookApp]

To access the notebook, open this file in a browser:
    file:///C:/Users/olehmber/AppData/Roaming/jupyter/runtime/nbserver-161344-open.html
Or copy and paste one of these URLs:
    http://localhost:8889/?token=a09
```

Jupyter Home Screen

- Startscreen in browser
 - like a file explorer

upload files to
notebook folder

Create new
files/notebooks



jupyter

Logout

Files Running Clusters

Select items to perform actions on them.

0

Untitled.ipynb Running seconds ago

untitled.txt seconds ago

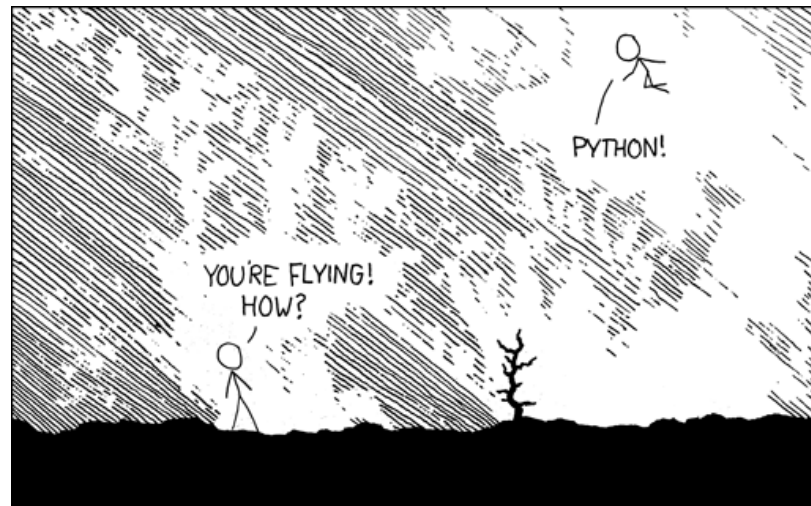
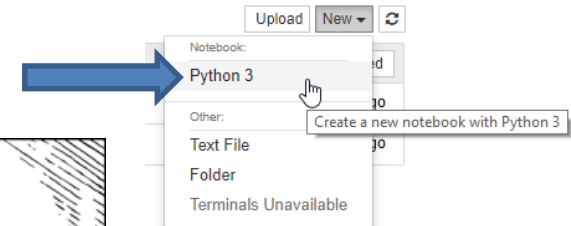
Upload New

Name Last Modified

file explorer

Now try it out

- Click in browser „New“ -> „Python 3“



<https://xkcd.com/353/>

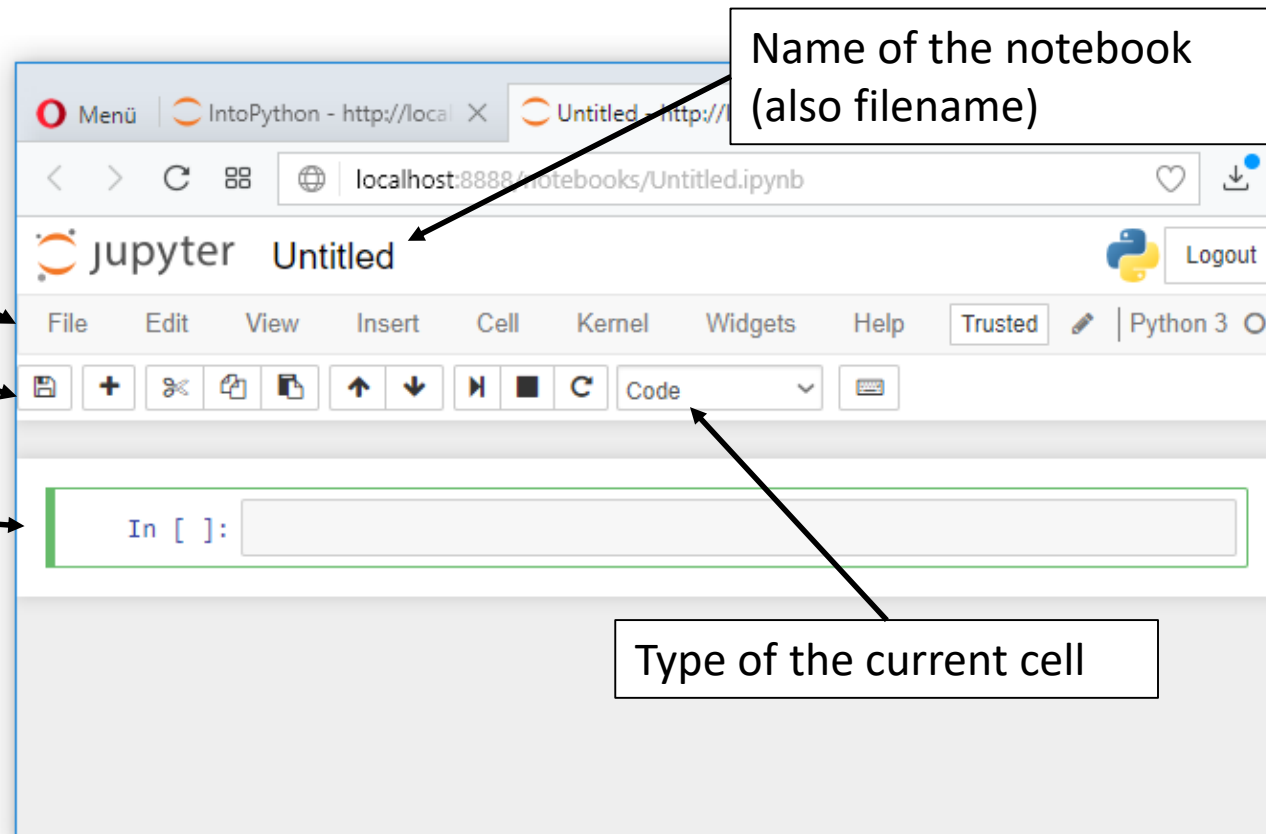
Jupyter Notebook

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

Menu Bar

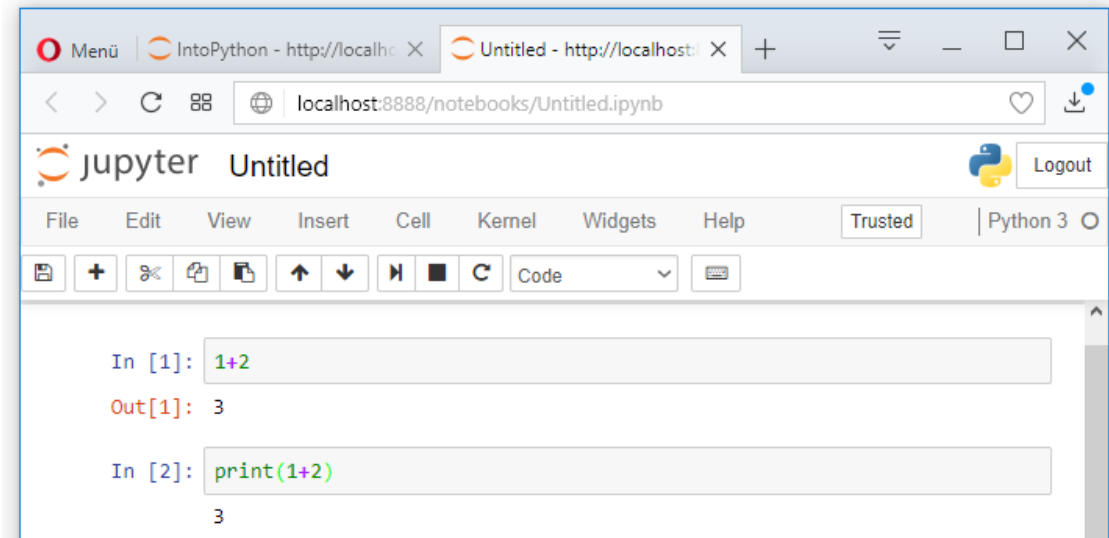
Shortcuts

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



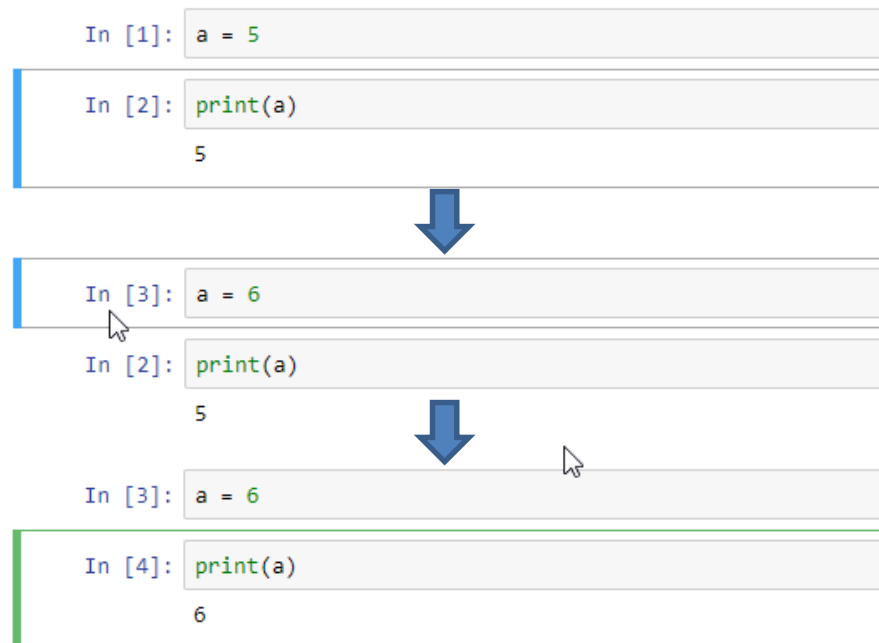
The screenshot shows a web browser window with two tabs: 'Menü' and 'IntoPython - http://localhost:8888'. The active tab is 'Untitled - http://localhost:8888'. The address bar shows 'localhost:8888/notebooks/Untitled.ipynb'. The Jupyter interface includes a 'Logout' button and a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. Below the menu bar is a toolbar with icons for file operations and execution. The main area displays two code cells. The first cell, labeled 'In [1]:', contains the code '1+2' and the output 'Out[1]: 3'. The second cell, labeled 'In [2]:', contains the code 'print(1+2)' and the output '3'.

```
In [1]: 1+2
Out[1]: 3

In [2]: print(1+2)
3
```

Jupyter Cells

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



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

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

Jupyter Cells

- Autocomplete by pressing <tab> when writing

```
In [2]: my_very_long_variable_with_hundreds_of_characters = 5  
my_second_very_long_variable_with_hundreds_of_characters = 6
```

```
In [ ]: my|  
my_second_very_long_variable_with_hundreds_of_characters  
my_very_long_variable_with_hundreds_of_characters
```

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

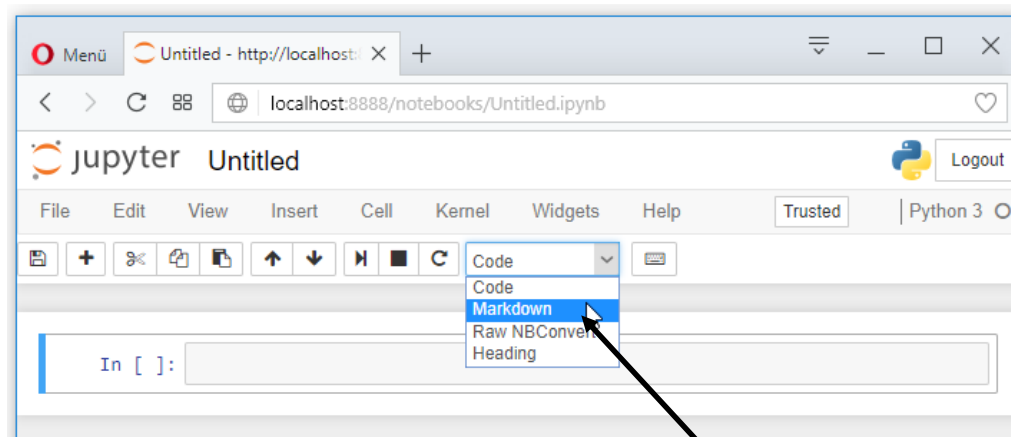
```
corr(datar, color_grades=5)
```

Signature:

```
corr(  
    ['data', "corr_method='spearman'", 'annot=False', 'mask=True', 'line_width=1', "line_  
    color='black'", 'color_grades=5', 'auto_sizing=True', "palette='default'", "style='asteti  
5 #from astetik import corr
```


Jupyter Cells

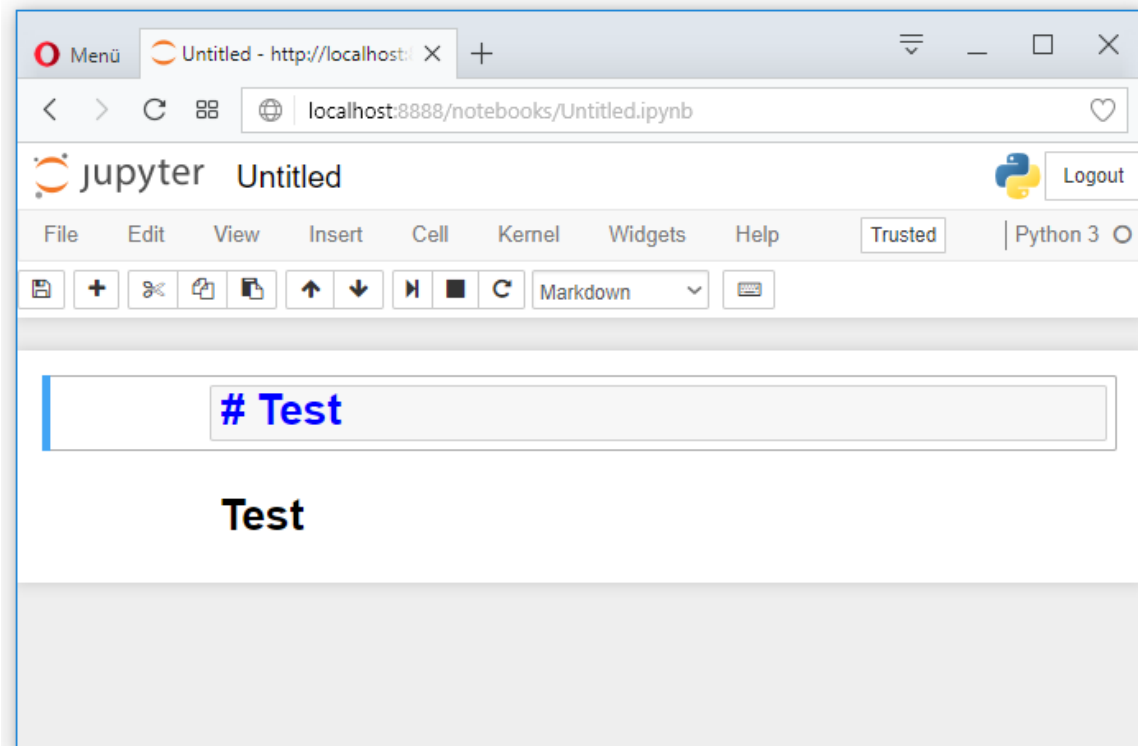
- What makes a notebook a notebook?
 - Markdown cells let you add documentation and notes
 - Create a new cell („Insert->Insert Cell Below“)
 - Change the type to Markdown



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

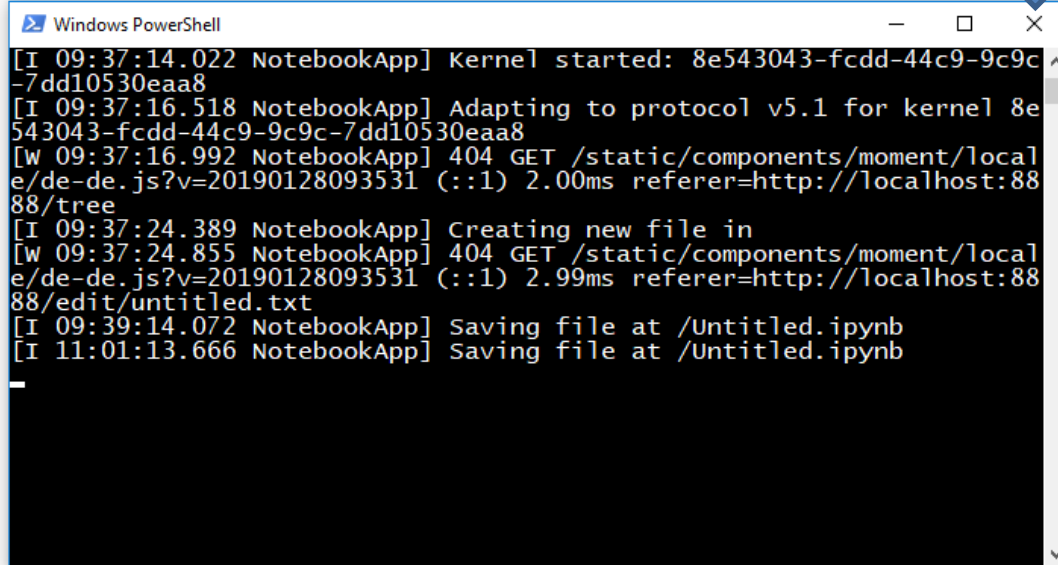

```
![Mannheim Image](https://www.uni-mannheim.de/1/00_UM_Dachmarke_DE_RGB.jpg)
```

- Quote

```
> This is a quotation
```

Shut down Jupyter

- Closing the browser (or the tab) will not close the Jupyter App
 - To completely shut it down you need to close the associated terminal
 - Or press “Ctrl” + “C”



```
Windows PowerShell
[I 09:37:14.022 NotebookApp] Kernel started: 8e543043-fcdd-44c9-9c9c-7dd10530eaa8
[I 09:37:16.518 NotebookApp] Adapting to protocol v5.1 for kernel 8e543043-fcdd-44c9-9c9c-7dd10530eaa8
[W 09:37:16.992 NotebookApp] 404 GET /static/components/moment/locale/de-de.js?v=20190128093531 (::1) 2.00ms referer=http://localhost:8888/tree
[I 09:37:24.389 NotebookApp] Creating new file in
[W 09:37:24.855 NotebookApp] 404 GET /static/components/moment/locale/de-de.js?v=20190128093531 (::1) 2.99ms referer=http://localhost:8888/edit/untitled.txt
[I 09:39:14.072 NotebookApp] Saving file at /Untitled.ipynb
[I 11:01:13.666 NotebookApp] Saving file at /Untitled.ipynb
```