00a_JupyterIntroduction

May 17, 2022

1 Python for Data Science & Analysis

1.1 Notes 1.0, Jupyter

Objectives
* Start jupyter notebook
* Create a new notebook
* Rename a notebook
* Add a text cell
* Add a code cell
* Run cells

1.2 What is Anaconda?

Anaconda is a collection of python packages, programms and other software for data analysis & data science.

1.3 What is jupyter notebook?

A notebook is a documentation or commentary-first style of programming, where the key elements are explanations & comments. The code is less important, and often only small amounts.

1.4 How do I start Anaconda?

Start Menu > Anaconda Navigator

1.5 How do I start jupyter notebook?

In Anaconda Navigator, press `LAUNCH` underneath the jupyter notebook icon...

1.6 How do I use the file browser?

Jupyter notebook starts with a file browser that displays files on your system.

eg., browse to your documents folder in jupyter.

1.7 How do I create notebooks?

In your documents folder, go to the `NEW` button, and press `Python3`.

Jupyter will create a new notebook called `Untitled`.

1.8 How do I rename a notebook?

Click on the title and a renaming box will appear, choose, eg., `MyNotes`.

1.9 How do I add and modify cells?

1.9.1 command mode: ESC (blue bar)

- `arrows` move around notebook
- `m` – markdown / text cell
- `y` – code cell / python cell
- `a` – insert cell above
- `b` – insert cell below
- `dd` – delete cell
- `z` – undo delete
- `CTRL+ENTER` to run
- `SHITT+ENTER` to run & move one cell below

1.10 How do I edit the contents of cells?

1.10.1 edit mode: ENTER (green bar)

- type
- `arrows` to move around text
- `CTRL+ENTER` run cell
- `SHIFT+ENTER` run cell & move one cell below

1.11 How do I add python code to a notebook?

Add a cell (eg., press `b`) and press `ENTER` to edit...

```
[1]: print(2 ** 2)
```

```
4
```

1.12 How do I run a python code cell?

Press `CTRL + ENTER` or the RUN button

1.13 What happens when I run a code cell in jupyter?

Jupyter always `print()`s the **last line...**

```
[2]: 5
```

```
[2]: 5
```

In an ordinary python program, you have `print()` everything you wish to show on the screen.

1.14 How do I add a text cell?

Add a cell, change type to markdown

or press `m` in COMMAND MODE.

1.15 How do I get help?

- `command?`
- `command??`
- `help(command)`

1.16 How do I get live feedback on my code?

`SHIFT-TAB` provides live autocomplete...

```
[3]: import pandas as pd
     import seaborn as sns
```

`pd.Da`

1.17 How do I find the arguments of a function?

```
[4]: import statistics
     statistics.mean?
```

1.18 How do I stop jupyter from printing the last line of a cell?

Use ;

```
[5]: 5
```

```
[5]: 5
```

```
[6]: 5;
```

1.19 EXTRA: How do I see the source code of a function?

```
[7]: import statistics
     statistics.mean??
```

1.20 How do I run command line programs within jupyter?

On unix system, `ls`:

```
[8]: !ls
```

```
LIVE_01_JupyterReview.ipynb
LIVE_312_SQL_CSV.ipynb
LIVE_31_SQL.ipynb
```

```
LIVE_32_Spark.ipynb
LIVE_331_Pandas.ipynb
LIVE_411_EDA_Pandas.ipynb
LIVE_412_EDA_Visuals.ipynb
LIVE_413_EDA_Plotly.ipynb
LIVE_42_Text.ipynb
NHS_Workbook.ipynb
QAZNHS2_Overview.ipynb
datasets
images
notes
reference
titanic.db
workbook
```

1.21 How do I run command line programs within jupyter?

On windows, `dir`:

[9]: `!dir`

```
 Volume in drive D is KUN-D
 Volume Serial Number is EA49-7B93

 Directory of D:\DATA\QA\20201029 - znhspython\20201116 - Data Analysis - KH

16/11/2020  11:03    <DIR>          .
16/11/2020  11:03    <DIR>          ..
16/11/2020  11:02    <DIR>          .ipynb_checkpoints
16/11/2020  10:53    <DIR>          datasets
16/11/2020  10:53    <DIR>          images
16/11/2020  11:03            13,318 LIVE_01_JupyterReview.ipynb
12/11/2020  12:03             7,841 LIVE_312_SQL_CSV.ipynb
12/11/2020  11:48            34,406 LIVE_31_SQL.ipynb
16/11/2020  09:50            23,555 LIVE_32_Spark.ipynb
13/11/2020  10:46            69,103 LIVE_331_Pandas.ipynb
13/11/2020  11:18            50,595 LIVE_411_EDA_Pandas.ipynb
13/11/2020  11:45           197,591 LIVE_412_EDA_Visuals.ipynb
13/11/2020  14:34         5,021,192 LIVE_413_EDA_Plotly.ipynb
13/11/2020  16:05            28,835 LIVE_42_Text.ipynb
13/11/2020  16:05            52,313 NHS_Workbook.ipynb
16/11/2020  10:53    <DIR>          notes
16/11/2020  09:48             7,637 QAZNHS2_Overview.ipynb
16/11/2020  10:53    <DIR>          reference
12/11/2020  11:55            12,288 titanic.db
16/11/2020  10:53    <DIR>          workbook
              12 File(s)      5,518,674 bytes
               8 Dir(s)  80,285,663,232 bytes free
```

```
[10]: !head datasets/titanic.csv
```

survived,pclass,sex,age,sibsp,parch,fare,embarked,class,who,adult_male,deck,emba
rk_town,alive,alone
0,3,male,22.0,1,0,7.25,S,Third,man,True,,Southampton,no,False
1,1,female,38.0,1,0,71.2833,C,First,woman,False,C,Cherbourg,yes,False
1,3,female,26.0,0,0,7.925,S,Third,woman,False,,Southampton,yes,True
1,1,female,35.0,1,0,53.1,S,First,woman,False,C,Southampton,yes,False
0,3,male,35.0,0,0,8.05,S,Third,man,True,,Southampton,no,True
0,3,male,,0,0,8.4583,Q,Third,man,True,,Queenstown,no,True
0,1,male,54.0,0,0,51.8625,S,First,man,True,E,Southampton,no,True
0,3,male,2.0,3,1,21.075,S,Third,child,False,,Southampton,no,False
1,3,female,27.0,0,2,11.1333,S,Third,woman,False,,Southampton,yes,False

1.22 How do I install python packages with jupyter?

```
[11]: !pip install plotly pyspark
```

Requirement already satisfied: plotly in c:\users\kunal\anaconda3\lib\site-
packages (4.10.0)
Collecting pyspark
  Downloading pyspark-3.0.1.tar.gz (204.2 MB)
Requirement already satisfied: six in c:\users\kunal\anaconda3\lib\site-packages
(from plotly) (1.14.0)
Requirement already satisfied: retrying>=1.3.3 in
c:\users\kunal\anaconda3\lib\site-packages (from plotly) (1.3.3)
Collecting py4j==0.10.9
  Downloading py4j-0.10.9-py2.py3-none-any.whl (198 kB)
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py): started
  Building wheel for pyspark (setup.py): still running…
  Building wheel for pyspark (setup.py): finished with status 'done'
  Created wheel for pyspark: filename=pyspark-3.0.1-py2.py3-none-any.whl
size=204612248
sha256=1b2290e8f6b057db0a550acca13fe3f96234f9ee0333cab325e9cf9ffca1a6b7
  Stored in directory: c:\users\kunal\appdata\local\pip\cache\wheels\32\d6\81\b2
26c0fb07391610277bc3e3fcc50b24bae5040b71d9dba360
Successfully built pyspark
Installing collected packages: py4j, pyspark
Successfully installed py4j-0.10.9 pyspark-3.0.1

1.23 Exercise (5 min)

Q1

- Download any courseware to your documents folder

Q2

- Create a new notebook
- Add a text cell (`b`, `m`)
  - set it to contain `# My Notes`
  - run it with **CTRL+ENTER**
- Add a code cell (`b`)
  - containing `print(2 ** 2)`
  - run with **CTRL+ENTER**

EXTRA: Q3

- try `help()` on:
  - str.upper
  - len
- try `?` on:
  - str.upper
  - str.replace
  - …
- try `dir()` on
  - "Michael"