

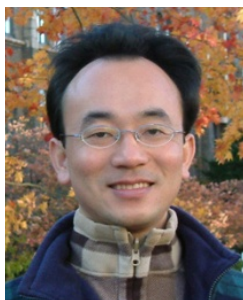
資料探勘 (Data Mining)

Python 資料探勘的基礎 (Foundations of Data Mining in Python)

1092DM03

MBA, IM, NTPU (M5026) (Spring 2021)

Tue 2, 3, 4 (9:10-12:00) (B8F40)



Min-Yuh Day

戴敏育

Associate Professor

副教授

Institute of Information Management, National Taipei University

國立臺北大學 資訊管理研究所

<https://web.ntpu.edu.tw/~myday>

2021-03-09



課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date) | 內容 (Subject/Topics) |
|-----------|------------|--|
| 1 | 2021/02/23 | 資料探勘介紹 (Introduction to data mining) |
| 2 | 2021/03/02 | ABC：人工智慧，大數據，雲端運算
(ABC: AI, Big Data, Cloud Computing) |
| 3 | 2021/03/09 | Python 資料探勘的基礎
(Foundations of Data Mining in Python) |
| 4 | 2021/03/16 | 資料科學與資料探勘：發現，分析，可視化和呈現數據
(Data Science and Data Mining:
Discovering, Analyzing, Visualizing and Presenting Data) |
| 5 | 2021/03/23 | 非監督學習：關聯分析，購物籃分析
(Unsupervised Learning: Association Analysis,
Market Basket Analysis) |
| 6 | 2021/03/30 | 資料探勘個案研究 I
(Case Study on Data Mining I) |

課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date) | 內容 (Subject/Topics) |
|-----------|------------|---|
| 7 | 2021/04/06 | 非監督學習：集群分析，行銷市場區隔
(Unsupervised Learning: Cluster Analysis, Market Segmentation) |
| 8 | 2021/04/13 | 監督學習：分類和預測
(Supervised Learning: Classification and Prediction) |
| 9 | 2021/04/20 | 期中報告 (Midterm Project Report) |
| 10 | 2021/04/27 | 監督學習：分類和預測
(Supervised Learning: Classification and Prediction) |
| 11 | 2021/05/04 | 機器學習和深度學習
(Machine Learning and Deep Learning) |
| 12 | 2021/05/11 | 卷積神經網絡
(Convolutional Neural Networks) |

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
13	2021/05/18	資料探勘個案研究 II (Case Study on Data Mining II)
14	2021/05/25	遞歸神經網絡 (Recurrent Neural Networks)
15	2021/06/01	強化學習 (Reinforcement Learning)
16	2021/06/08	社交網絡分析 (Social Network Analysis)
17	2021/06/15	期末報告 I (Final Project Report I)
18	2021/06/22	期末報告 II (Final Project Report II)

Foundations of Data Mining in Python

Outline

- **Foundations of Data Mining in Python**
 - **Python**
 - Programming language
 - **Numpy**
 - Scientific computing

Python Ecosystem for Data Science

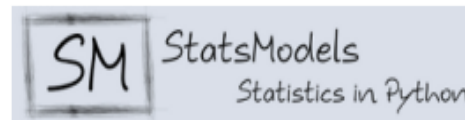
Few (~10) main libraries



Share code

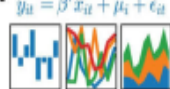


Information



Statistics

pandas



Structured data



SciPy

Scientific

matplotlib

Graphics



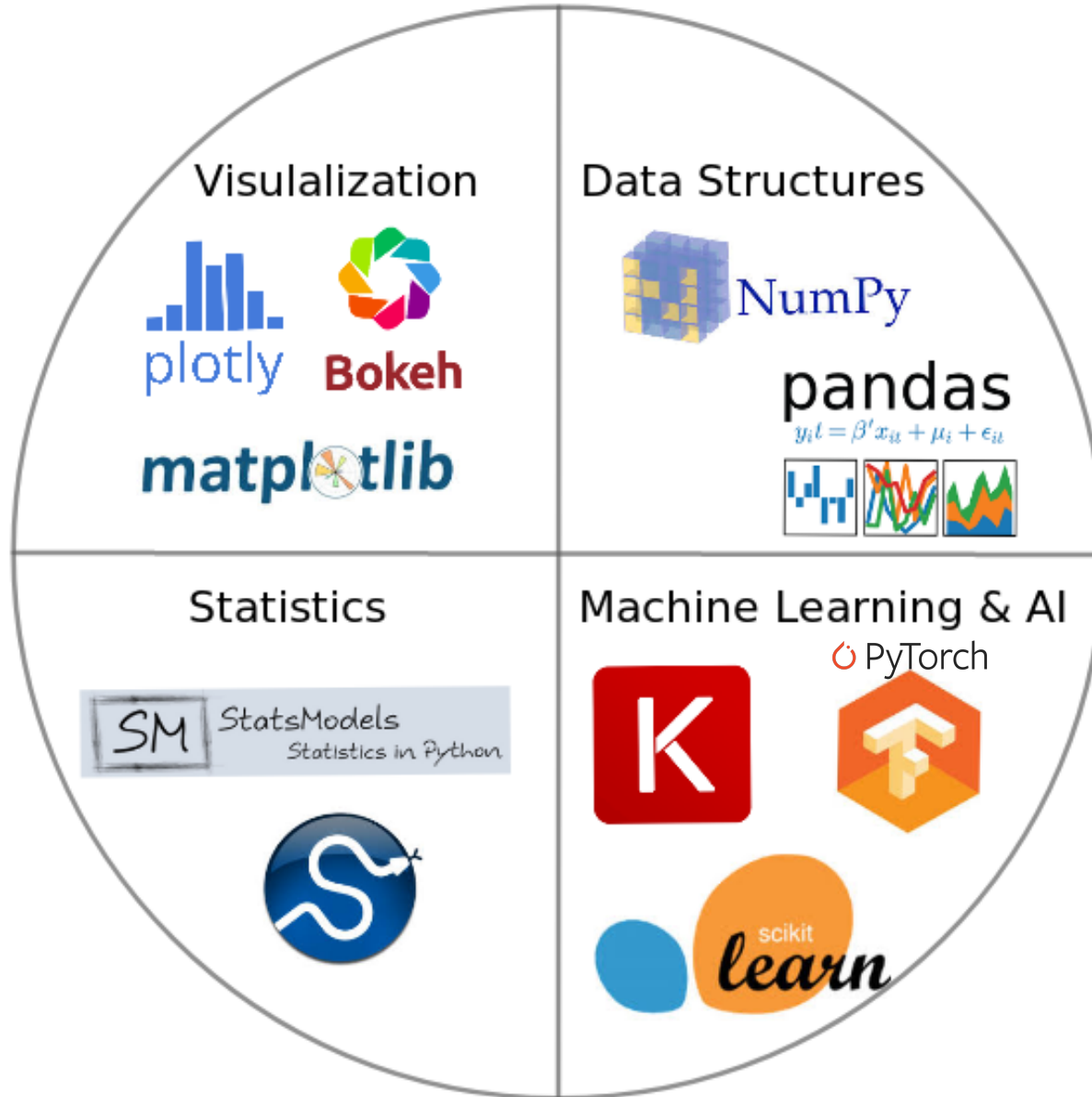
Lab book



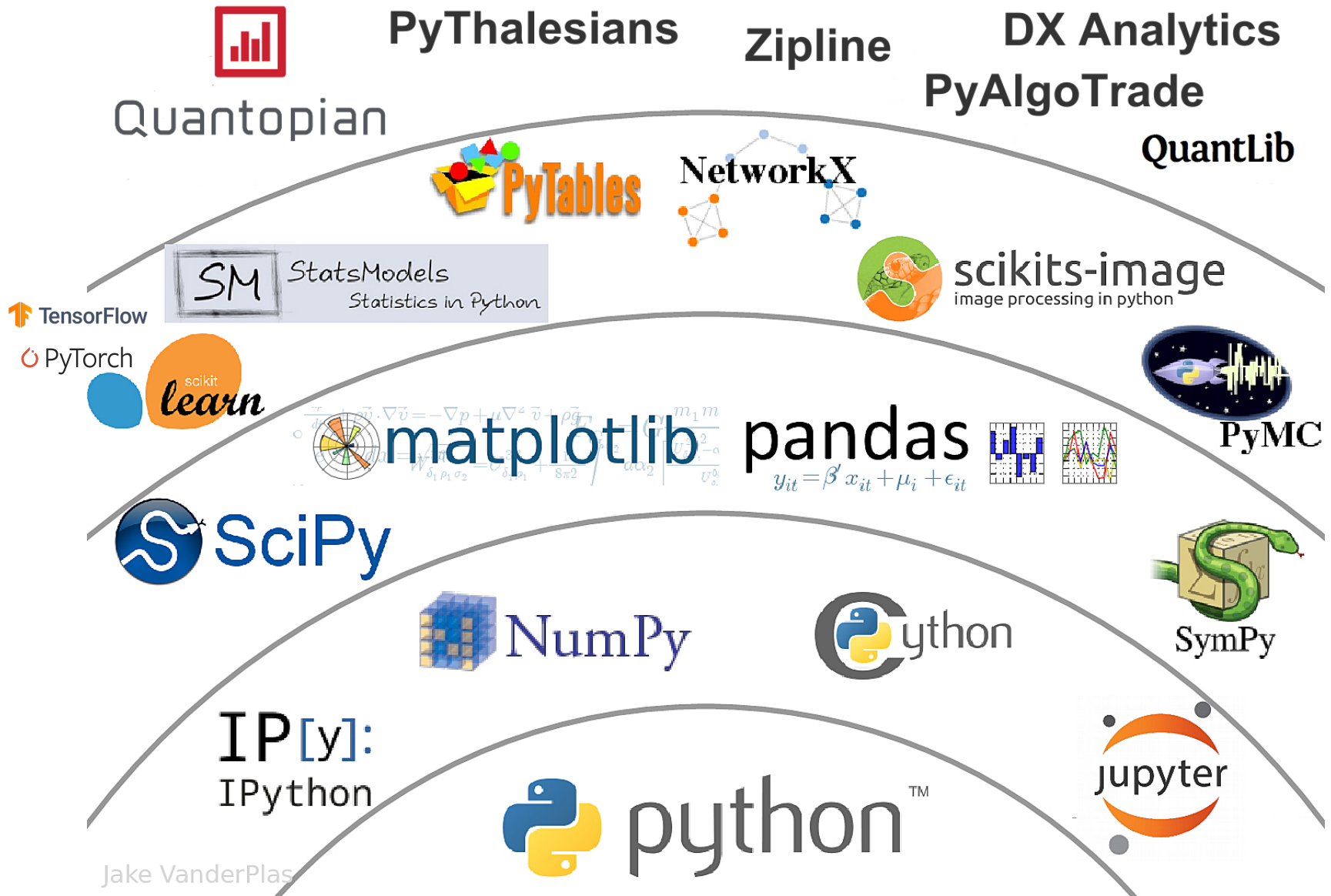
Numerical data (arrays)



Python Ecosystem for Data Science



The Quant Finance PyData Stack





Python

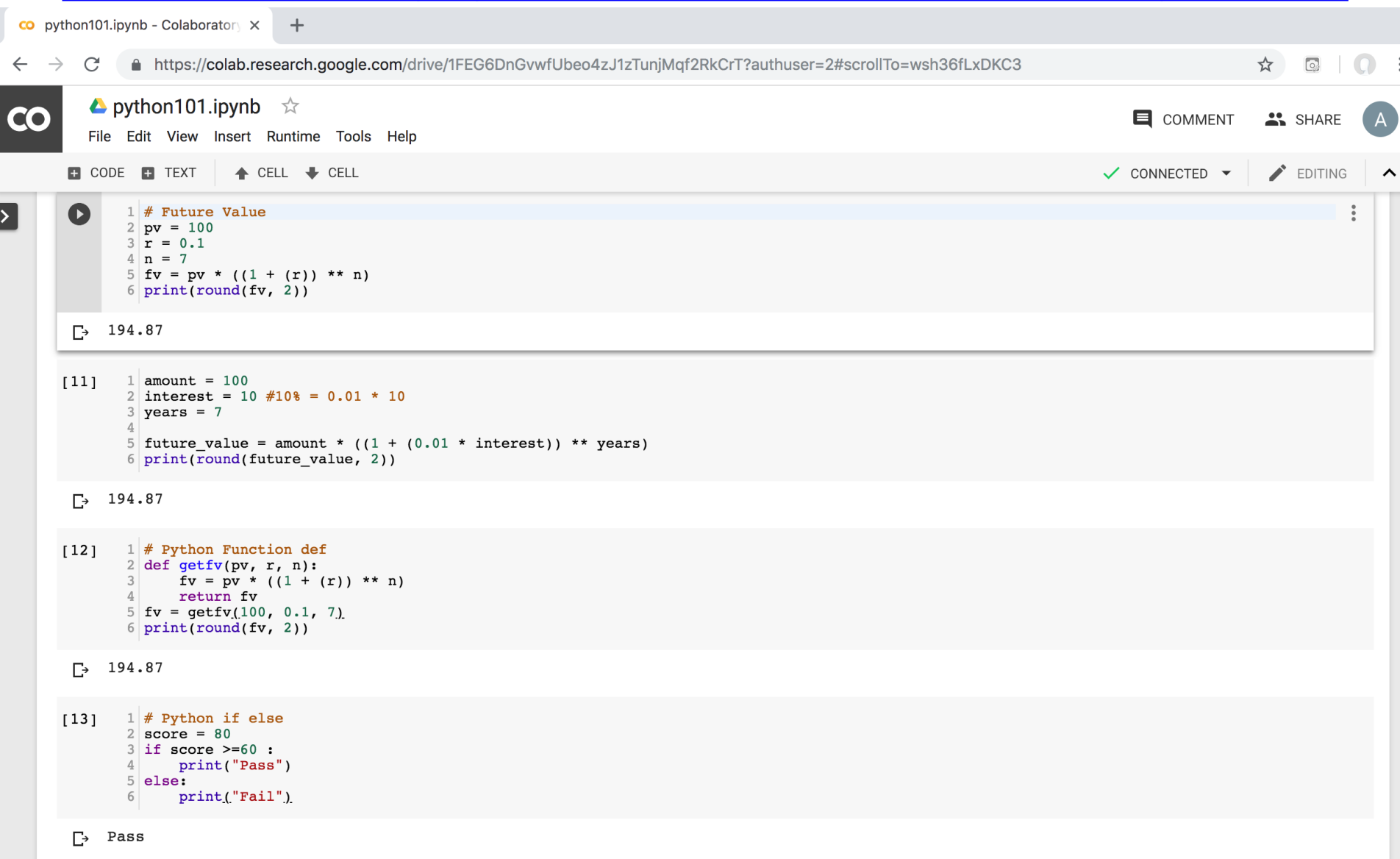
Python is an
interpreted,
object-oriented,
high-level
programming language
with
dynamic semantics.

Google Colab

The screenshot shows the Google Colab web interface. At the top, the browser address bar displays the URL <https://colab.research.google.com/notebooks/welcome.ipynb>. The main header includes the Colab logo, the text "Hello, Colaboratory", and a menu with options: File, Edit, View, Insert, Runtime, Tools, and Help. On the right side of the header, there is a "SHARE" button and a user profile picture. Below the header, a toolbar contains buttons for "+ CODE", "+ TEXT", "↑ CELL", "↓ CELL", and "COPY TO DRIVE". On the far right of the toolbar, there are "CONNECT" and "EDITING" options. A left-hand sidebar contains a "Table of contents" section with links to "Getting Started", "Highlighted Features", "TensorFlow execution", "GitHub", "Visualization", "Forms", "Examples", and "Local runtime support". The main content area features a large "Welcome to Colaboratory!" message with the Colab logo and a link to the "FAQ". Below this is a "Getting Started" section with a bulleted list of links: "Overview of Colaboratory", "Loading and saving data: Local files, Drive, Sheets, Google Cloud Storage", "Importing libraries and installing dependencies", "Using Google Cloud BigQuery", "Forms, Charts, Markdown, & Widgets", "TensorFlow with GPU", and "Machine Learning Crash Course: Intro to Pandas & First Steps with TensorFlow". A "Highlighted Features" section is partially visible, starting with a "Seedbank" subsection that says "Looking for Colab notebooks to learn from? Check out [Seedbank](#), a place to discover interactive machine learning examples." Below that, the "TensorFlow execution" subsection begins with the text "Colaboratory allows you to execute TensorFlow code in your browser with a single click. The example below adds two matrices." followed by a mathematical equation:
$$\begin{bmatrix} 1. & 1. & 1. \end{bmatrix} + \begin{bmatrix} 1. & 2. & 3. \end{bmatrix} = \begin{bmatrix} 2. & 3. & 4. \end{bmatrix}$$

Python in Google Colab (Python101)

<https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT>



python101.ipynb - Colaboratory

File Edit View Insert Runtime Tools Help

COMMENT SHARE

CONNECTED EDITING

```
1 # Future Value
2 pv = 100
3 r = 0.1
4 n = 7
5 fv = pv * ((1 + (r)) ** n)
6 print(round(fv, 2))
```

194.87

```
[11] 1 amount = 100
2 interest = 10 #10% = 0.01 * 10
3 years = 7
4
5 future_value = amount * ((1 + (0.01 * interest)) ** years)
6 print(round(future_value, 2))
```

194.87

```
[12] 1 # Python Function def
2 def getfv(pv, r, n):
3     fv = pv * ((1 + (r)) ** n)
4     return fv
5 fv = getfv(100, 0.1, 7)
6 print(round(fv, 2))
```

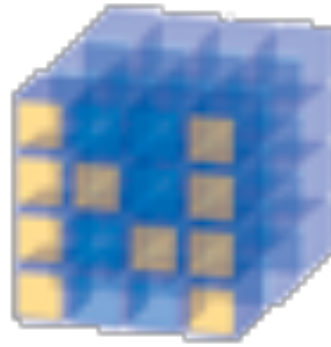
194.87

```
[13] 1 # Python if else
2 score = 80
3 if score >=60 :
4     print("Pass")
5 else:
6     print("Fail").
```

Pass

<https://tinyurl.com/aintpupython101>

NumPy



NumPy
Base

**N-dimensional array
package**

Python

matplotlib

matplotlib

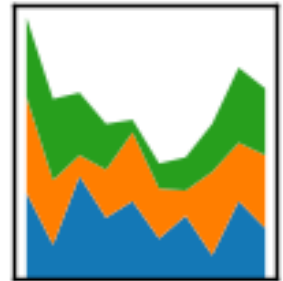
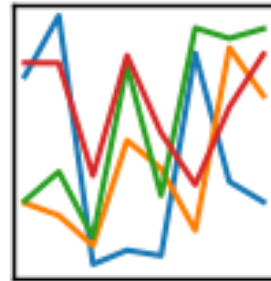
The logo for Matplotlib, which is a circular plot with a white background and a blue border. Inside the circle, there are several colored segments (orange, yellow, green, blue) radiating from the center, resembling a pie chart or a polar plot.

Python

Pandas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Connect Google Colab in Google Drive

The image shows a browser window with the Google Drive interface. The address bar shows the URL `https://drive.google.com/drive/u/2/my-drive`. The main header includes the Drive logo, a search bar, and navigation icons. The left sidebar contains a 'New' button and a list of navigation options: My Drive, Computers, Shared with me, Recent, Starred, Trash, Backups, and Storage. The 'My Drive' option is highlighted with a red dashed box. A dropdown menu is open from 'My Drive', listing options: New folder..., Upload files..., Upload folder..., Google Docs, Google Sheets, Google Slides, and More. The 'More' option is also highlighted with a red dashed box. A second dropdown menu is open from 'More', listing: Google Forms, Google Drawings, Google My Maps, Google Sites, and Connect more apps. The 'Connect more apps' option is highlighted with a red dashed box. The main content area shows a 'Files' section with a table of file management options: Store safely, Sync seamlessly, Access anywhere, and Share easily. A storage usage bar at the bottom left indicates '0 bytes of 15 GB used' and a 'Get Backup and Sync for Mac' notification is visible.

Google Colab








My Drive - Google Drive x +

https://drive.google.com/drive/u/2/my-drive

Drive Search Drive

Connect apps to Drive

All colab x

 ZIP Extractor Extract ZIP files to Google Drive Extraction complete. View extracted files Share Extract another  Test.zip ZIP Extractor 307,585 users	 LUMIN PDF The fast and simple PDF Viewer box	 cloudconvert CloudConvert 373,161 users
 Sejda Merge PDF - Split PDF - Sejda.com ★★★★★ (1106)	 DocHub Edit, Send & Sign PDFs DocHub - Edit and Sign PDF Docu... 2,131,600 users	 Google Forms Google Forms 4,803,614 users

Get Backup and Sync for Mac

Access anywhere
Share easily

Name ↑

Google Colab

My Drive - Google Drive x +

https://drive.google.com/drive/u/2/my-drive

Drive Search Drive

New

My Drive

Computers

Shared with me

Recent

Starred

Trash

Backups

Storage

0 bytes of 15 GB used
[UPGRADE STORAGE](#)

Get Backup and Sync for Mac


Access anywhere

Share easily

Connect apps to Drive

All

colab



Colaboratory
offered by <https://colab.research.google.com>

A data analysis tool that combines code, output, and descriptive text into one collaborative document.

+ CONNECT

Productivity
★★★★★ (195)

Name ↑

Connect Colaboratory to Google Drive

The screenshot shows a web browser window with the Google Drive interface. The address bar displays "https://drive.google.com/drive/u/2/my-drive". The main content area is a "Connect apps to Drive" dialog box. At the top of the dialog, there is a search bar containing "colab" and a dropdown menu set to "All". Below the search bar, a notification card for Colaboratory is displayed. The card features the Colaboratory logo (two yellow circles with "CO" inside), a green checkmark, and the text "Colaboratory was connected to Google Drive." Below this, there is a checked checkbox and the text "Make Colaboratory the default app for files it can open". A blue "OK" button with a red dashed border is located at the bottom right of the notification card. To the right of the notification card, there is a "RATE IT" button with a star icon, followed by the text "Productivity" and a star rating of five stars with "(195)" next to it. The background of the dialog box shows a list of files, with the top item partially visible, showing a green checkmark and the Colaboratory logo. The Google Drive sidebar is visible on the left, showing options like "New", "My Drive", "Computers", "Shared with me", "Recent", "Starred", "Trash", "Backups", and "Storage". The storage status at the bottom left indicates "0 bytes of 15 GB used" and a link to "UPGRADE STORAGE".

Google Colab

The image shows a browser window with the Google Drive interface. The address bar shows the URL `https://drive.google.com/drive/u/2/my-drive`. The 'New' button is highlighted with a red dashed box, and its dropdown menu is open. The 'More' option in the dropdown is also highlighted with a red dashed box. The 'More' dropdown is open, showing options like Google Forms, Google Drawings, Google My Maps, Google Sites, and Colaboratory. The 'Colaboratory' option is highlighted with a red dashed box. The interface includes a search bar, navigation icons, and a storage status indicator showing 0 bytes of 15 GB used.

My Drive - Google Drive

Search Drive

My Drive

Quick Access

New

- My Drive
- Computer
- Shared with me
- Recent
- Starred
- Trash
- Backups

New folder...

Upload files...

Upload folder...

Google Docs

Google Sheets

Google Slides

More

- Google Forms
- Google Drawings
- Google My Maps
- Google Sites
- Colaboratory

Connect more apps

Name ↑

Storage

0 bytes of 15 GB used

UPGRADE STORAGE

Get Backup and Sync for Mac

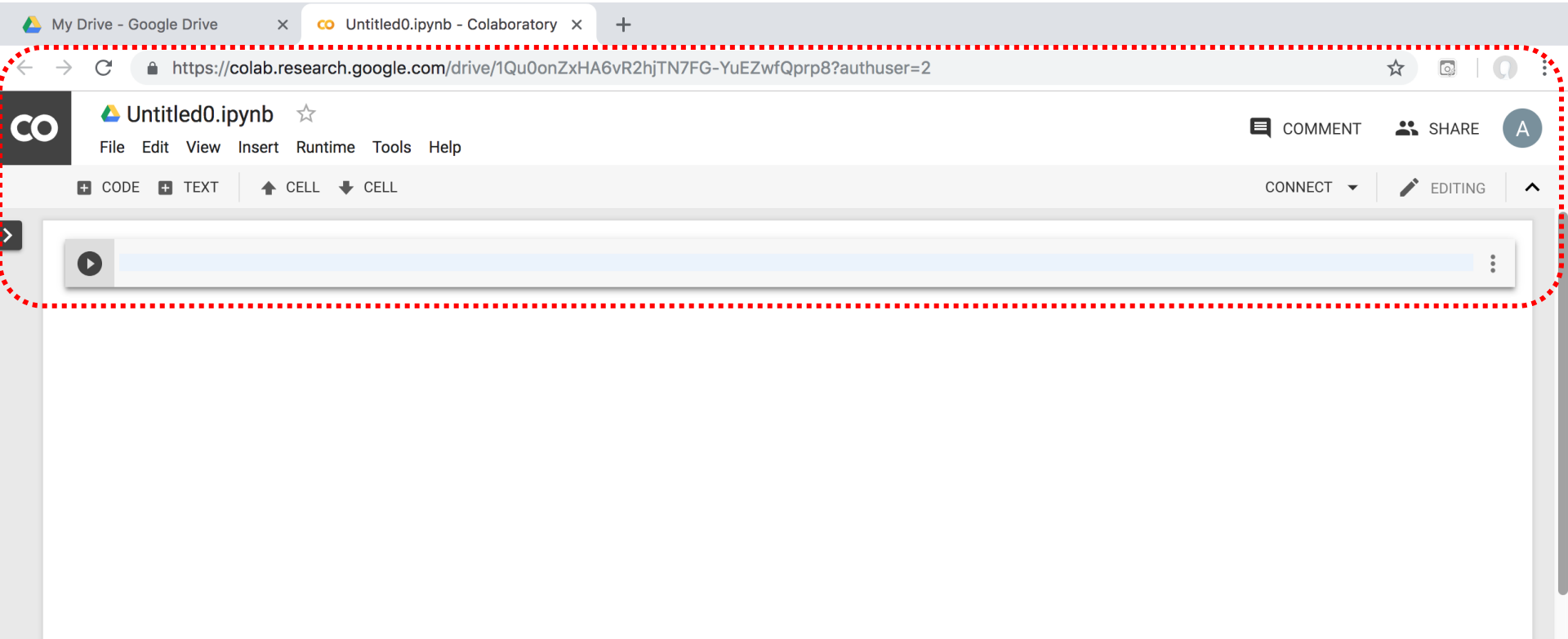
Store safely

Sync seamlessly

Access anywhere

Share easily

Google Colab



Google Colab

The image shows a browser window with two tabs: "My Drive - Google Drive" and "Untitled0.ipynb - Colaboratory". The address bar shows the URL: <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2>. The main content area displays the Google Colab interface for a notebook titled "Untitled0.ipynb". The menu bar includes "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". The "Runtime" menu is open, showing the following options:

- Run all (⌘/Ctrl+F9)
- Run before (⌘/Ctrl+F8)
- Run the focused cell (⌘/Ctrl+Enter)
- Run selection (⌘/Ctrl+Shift+Enter)
- Run after (⌘/Ctrl+F10)
- Interrupt execution (⌘/Ctrl+M I)
- Restart runtime... (⌘/Ctrl+M .)
- Restart and run all...
- Reset all runtimes...
- Change runtime type
- Manage sessions

The "Runtime" menu title and the "Change runtime type" option are highlighted with red dashed boxes. The interface also shows a "CONNECT" button and an "EDITING" mode indicator.

Run Jupyter Notebook Python3 GPU Google Colab

The image shows a web browser window with two tabs: "My Drive - Google Drive" and "Untitled0.ipynb - Colaboratory". The address bar shows the URL: <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2>. The page title is "Untitled0.ipynb". The navigation menu includes "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". Below the menu are buttons for "+ CODE", "+ TEXT", "↑ CELL", and "↓ CELL". On the right side, there are "COMMENT", "SHARE", "CONNECT", and "EDITING" options. A "Notebook settings" dialog box is open in the center, containing the following elements:

- Notebook settings**
- Runtime type: Python 3
- Hardware accelerator: GPU
- Omit code cell output when saving this notebook
- CANCEL SAVE

Google Colab Python Hello World

```
print('Hello World')
```

The screenshot displays the Google Colaboratory web interface. At the top, the browser tab is labeled "Untitled0.ipynb - Colaboratory". The address bar shows the URL: <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2#scrollTo=6s-m3sER8G1u>. The main header area includes the "Untitled0.ipynb" title, a star icon, and navigation links for "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". On the right side of the header, there are buttons for "COMMENT", "SHARE", and a user profile icon labeled "A". Below the header, a toolbar contains options for "+ CODE", "+ TEXT", "↑ CELL", and "↓ CELL". The status bar at the bottom right indicates "CONNECTED" with a green checkmark and "EDITING" with a pencil icon. The central workspace features a code cell with a play button icon on the left and a vertical ellipsis menu on the right. The code cell contains the Python code `print('Hello World')`. Below the code cell, the output is displayed as "Hello World" with a copy icon to its left.



Anaconda
The Most Popular
Python
Data Science Platform

Download Anaconda



[Documentation](#) [Blog](#) [Contact](#) [Anaconda Cloud](#) [Q](#)

[What is Anaconda?](#) [Products](#) [Support](#) [Community](#) [Resources](#) [About](#)

[Download](#)

[Don't Miss AnacondaCon Apr 8-11 Austin TX!](#)

Download Anaconda Distribution

Version 5.1 | Release Date: February 15, 2018

Download For:   

High-Performance Distribution

Easily install 1,000+ [data science packages](#)

Package Management

Manage packages, dependencies and environments with [conda](#)

Portal to Data Science

Uncover insights in your data and create interactive visualizations



[Anaconda 5.1 For macOS Installer](#)

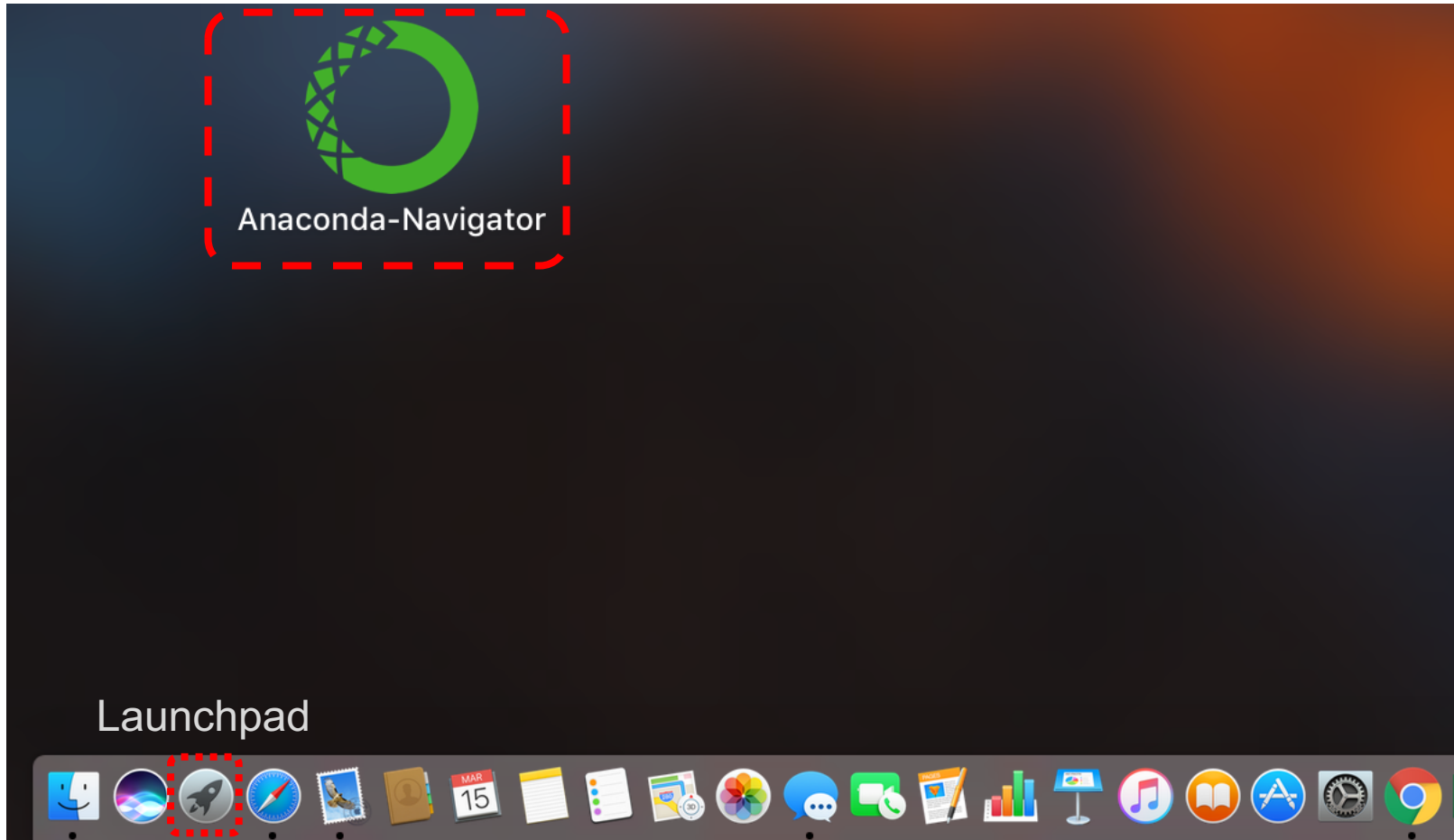
<https://www.anaconda.com/download>



Python

HelloWorld

Anaconda-Navigator



Anaconda Navigator

Anaconda Navigator

ANACONDA NAVIGATOR

Sign in to Anaconda Cloud

Home

Environments

Learning

Community

Documentation

Developer Blog

Feedback



Applications on

base (root)

Channels

Refresh



jupyterlab

0.31.5

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch



notebook

5.4.0

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



qtconsole

4.3.1

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Launch



spyder

3.2.6

Scientific PYTHON Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Launch

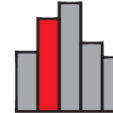


vscode

1.22.2

Streamlined code editor with support for development operations like debugging, task running and version control.

Launch



glueviz

0.12.4

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Install

Jupyter Notebook

The screenshot shows a web browser window with the Jupyter Notebook interface. The browser's address bar displays the URL `localhost:8888/tree/Documents/Data/BDA`. The Jupyter logo and a "Logout" button are visible in the top left and right corners, respectively. Below the navigation tabs ("Files", "Running", "Clusters"), there are buttons for "Upload", "New", and a refresh icon. A message "Select items to perform actions on them." is displayed above the file list. The file list shows a folder named `..` with a "seconds ago" timestamp. A red dashed box highlights the file list area, which also contains the message "The notebook list is empty." at the bottom.

Home x

localhost:8888/tree/Documents/Data/BDA

jupyter Logout

Files Running Clusters

Select items to perform actions on them. Upload New ↕ ↻

0 ▾ / Documents / Data / BDA Name ↓ Last Modified

.. seconds ago

The notebook list is empty.

Jupyter Notebook

New Python 3

Home x Untitled x

localhost:8888/tree/Documents/Data/BDA

jupyter Logout

Files Running Clusters

Select items to perform actions on them.

Upload New

0 / Documents / Data / BDA

..

The notebook list is empty.

Notebook:
Python 3

Other:
Text File
Folder
Terminal

print("hello, world")

The image shows a Jupyter Notebook interface in a web browser. The browser tabs are labeled 'Home' and 'HelloWorld'. The address bar shows the URL 'localhost:8888/notebooks/Documents/Data/BDA/HelloWorld.ipynb'. The Jupyter logo and 'HelloWorld (autosaved)' are visible in the top left. A 'Logout' button is in the top right. Below the browser is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. To the right of the menu bar are 'Trusted' and 'Python 3' indicators. A toolbar contains icons for file operations and a 'Run' button, which is highlighted with a red box. Below the toolbar is a code cell with the text 'In [1]: print("hello, world")' and its output 'hello, world'. The code line is also highlighted with a red box. Below the code cell is an empty input field labeled 'In []:'.

```
from platform import python_version
print("Python Version:", python_version())
```

The screenshot shows a web browser window with a Jupyter Notebook. The browser tabs include 'Home' and 'HelloWorld'. The address bar shows 'localhost:8888/notebooks/Documents/Data/BDA/HelloWorld.ipynb'. The Jupyter interface includes a header with the 'jupyter HelloWorld' logo, a 'Logout' button, and a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, Help. A 'Trusted' status indicator and 'Python 3' are also visible. Below the menu bar is a toolbar with icons for file operations and execution. The notebook content consists of three input cells:

- In [1]:** `print("hello, world")`
Output: `hello, world`
- In [2]:** `from platform import python_version`
`print("Python Version:", python_version())`
Output: `Python Version: 3.6.5`
- In []:** (Empty cell)



Python

Programming



Python Fiddle

Python Cloud IDE | Python Fiddle x

pythonfiddle.com

Run Reset Share Import Login Language

Python Fiddle Python Cloud IDE

G+1 2.6k

Examples

- Chaining comparison operators
- Decorators
- Creating generators objects
- Enumerate
- Function closure
- Lex tokenizer
- Step argument in slice operators
- For Else
- Verbose regular expressions
- In-place value swapping
- Function argument unpacking

Packages

Hotkeys

```
1 print("Hello Python Fiddle")
2
```

Title:

Description:

Tags:

A comma-separated list of tags.

Save

Hello Python Fiddle

Text input and output

```
print("Hello World")
```

```
print("Hello World\nThis is a message")
```

```
x = 3  
print(x)
```

```
x = 2  
y = 3  
print(x, ' ', y)
```

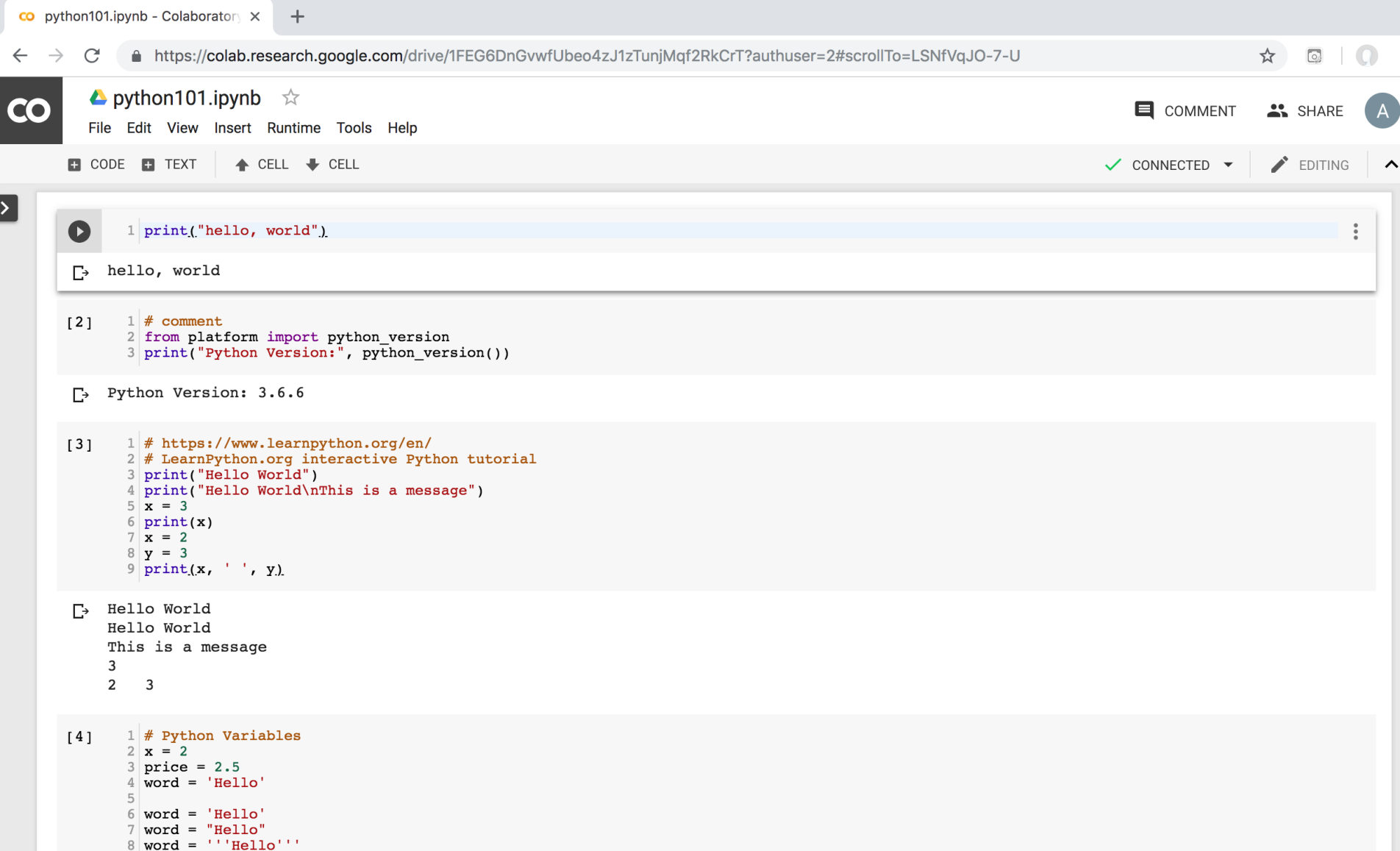
```
name = input("Enter a name: ")
```

```
x = int(input("What is x? "))
```

```
x = float(input("Write a number "))
```

Python in Google Colab

<https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT>



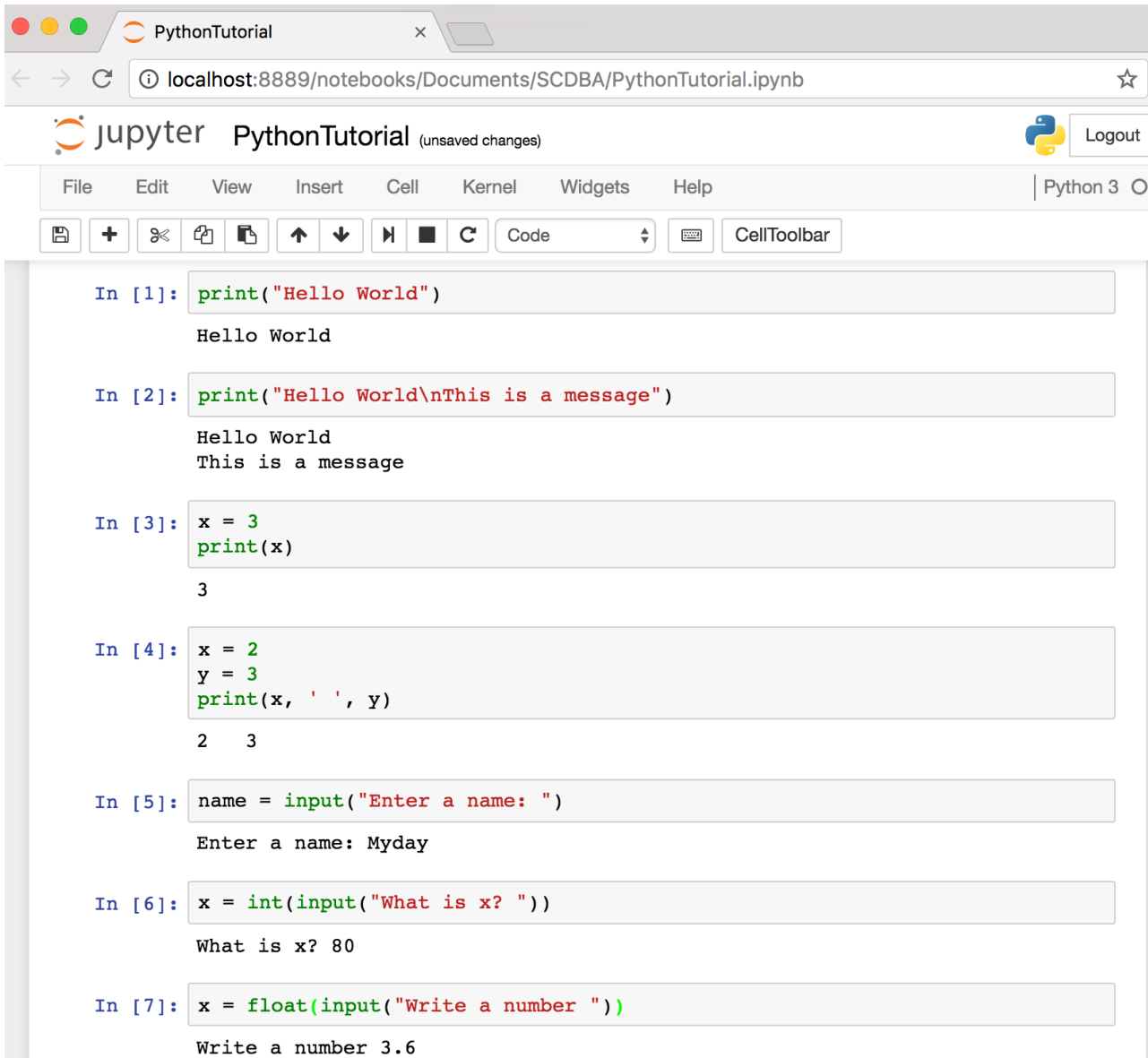
The screenshot shows a Google Colab notebook interface. The browser address bar displays the URL: <https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT?authuser=2#scrollTo=LSNfVqJO-7-U>. The notebook title is "python101.ipynb". The interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help), a toolbar with options for CODE, TEXT, and CELL, and a status bar showing "CONNECTED" and "EDITING".

The notebook contains four code cells:

- Cell 1:** `print("hello, world").` Output: `hello, world`
- Cell 2:** `# comment`, `from platform import python_version`, `print("Python Version:", python_version())`. Output: `Python Version: 3.6.6`
- Cell 3:** `# https://www.learnpython.org/en/`, `# LearnPython.org interactive Python tutorial`, `print("Hello World")`, `print("Hello World\nThis is a message")`, `x = 3`, `print(x)`, `x = 2`, `y = 3`, `print(x, ' ', y).` Output: `Hello World`, `Hello World`, `This is a message`, `3`, `2 3`
- Cell 4:** `# Python Variables`, `x = 2`, `price = 2.5`, `word = 'Hello'`, `word = 'Hello'`, `word = "Hello"`, `word = '''Hello'''`

<https://tinyurl.com/aintpuython101>

Text input and output



The screenshot shows a Jupyter Notebook window titled "PythonTutorial" with the URL `localhost:8889/notebooks/Documents/SCDBA/PythonTutorial.ipynb`. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations and execution. The notebook contains seven code cells, each with its input code and the resulting output:

```
In [1]: print("Hello World")
Hello World
```

```
In [2]: print("Hello World\nThis is a message")
Hello World
This is a message
```

```
In [3]: x = 3
print(x)
3
```

```
In [4]: x = 2
y = 3
print(x, ' ', y)
2  3
```

```
In [5]: name = input("Enter a name: ")
Enter a name: Myday
```

```
In [6]: x = int(input("What is x? "))
What is x? 80
```

```
In [7]: x = float(input("Write a number "))
Write a number 3.6
```


Variables

```
x = 2  
price = 2.5  
word = 'Hello'
```

```
word = 'Hello'  
word = "Hello"  
word = '''Hello'''
```

```
x = 2  
x = x + 1  
x = 5
```

Python Basic Operators

```
print('7 + 2 =', 7 + 2)
print('7 - 2 =', 7 - 2)
print('7 * 2 =', 7 * 2)
print('7 / 2 =', 7 / 2)
print('7 // 2 =', 7 // 2)
print('7 % 2 =', 7 % 2)
print('7 ** 2 =', 7 ** 2)
```

```
print('7 + 2 =', 7 + 2)
print('7 - 2 =', 7 - 2)
print('7 * 2 =', 7 * 2)
print('7 / 2 =', 7 / 2)
print('7 // 2 =', 7 // 2)
print('7 % 2 =', 7 % 2)
print('7 ** 2 =', 7 ** 2)
```

```
7 + 2 = 9
7 - 2 = 5
7 * 2 = 14
7 / 2 = 3.5
7 // 2 = 3
7 % 2 = 1
7 ** 2 = 49
```

BMI Calculator in Python

```
height_cm = float(input("Enter your height in cm: "))
weight_kg = float(input("Enter your weight in kg: "))

height_m = height_cm/100
BMI = (weight_kg/(height_m**2))

print("Your BMI is: " + str(round(BMI,1)))
```

BMI Calculator in Python

```
In [1]: height_cm = float(input("Enter your height in cm: "))
weight_kg = float(input("Enter your weight in kg: "))

height_m = height_cm/100
BMI = (weight_kg/(height_m**2))

print("Your BMI is: " + str(round(BMI,1)))
```

```
Enter your height in cm: 170
Enter your weight in kg: 60
Your BMI is: 20.8
```

```
In [ ]:
```

Future value
of a specified
principal amount,
rate of interest, and
a number of years

Future Value (FV)

```
# How much is your $100 worth after 7 years?  
print(100 * 1.1 ** 7)  
# output = 194.87
```

```
print(100 * 1.1 ** 7)
```

```
194.871710000000012
```

Future Value (FV)

```
pv = 100  
r = 0.1  
n = 7
```

```
fv = pv * ((1 + (r)) ** n)  
print(round(fv, 2))
```

```
pv = 100  
r = 0.1  
n = 7  
  
fv = pv * ((1 + (r)) ** n)  
print(round(fv, 2))
```

```
194.87
```

Future Value (FV)

```
amount = 100
interest = 10 #10% = 0.01 * 10
years = 7

future_value = amount * ((1 + (0.01 * interest)) ** years)
print(round(future_value, 2))
```

```
amount = 100
interest = 10 #10% = 0.01 * 10
years = 7

future_value = amount * ((1 + (0.01 * interest)) ** years)
print(round(future_value, 2))
```

194.87

if statements

> greater than
< smaller than
== equals
!= is not

```
score = 80
if score >=60 :
    print("Pass")
else:
    print("Fail")
```

Pass

```
score = 80
if score >=60 :
    print("Pass")
else:
    print("Fail")
```

if elif else

```
score = 90
grade = ""
if score >= 90:
    grade = "A"
elif score >= 80:
    grade = "B"
elif score >= 70:
    grade = "C"
elif score >= 60:
    grade = "D"
else:
    grade = "E"
print(grade)
# grade = "A"
```

A

<http://pythontutor.com/visualize.html>
<https://goo.gl/E6w5ph>

for loops

```
for i in range(1,11):  
    print(i)
```

```
1  
2  
3  
4  
5  
6  
7  
8  
9  
10
```

for loops

```
for i in range(1,10):  
    for j in range(1,10):  
        print(i, ' * ', j, ' = ', i*j)
```

```
9 * 1 = 9  
9 * 2 = 18  
9 * 3 = 27  
9 * 4 = 36  
9 * 5 = 45  
9 * 6 = 54  
9 * 7 = 63  
9 * 8 = 72  
9 * 9 = 81
```

while loops

```
age = 10

while age < 20:
    print(age)
    age = age + 1
```

```
10
11
12
13
14
15
16
17
18
19
```

def Functions

```
def convertCMtoM(xcm) :  
    m = xcm/100  
    return m
```

```
cm = 180  
m = convertCMtoM(cm)  
print(str(m))
```

1.8

Lists []

```
x = [60, 70, 80, 90]
print(len(x))
print(x[0])
print(x[1])
print(x[-1])
```

4

60

70

90

Tuples ()

A **tuple** in Python is a collection that **cannot be modified**.

A tuple is defined using **parenthesis**.

```
x = (10, 20, 30, 40, 50)
print(x[0])
print(x[1])
print(x[2])
print(x[-1])
```

```
10
20
30
50
```


Dictionary {key : value}

```
k = { 'EN' : 'English' , 'FR' : 'French' }  
print(k[ 'EN' ])
```

Dictionary

'EN' → 'English'

'FR' → 'French'

English

Sets {}

```
animals = {'cat', 'dog'}
```

```
animals = {'cat', 'dog'}
print('cat' in animals) # Check if an element is in a set; prints "True"
print('fish' in animals) # prints "False"
animals.add('fish') # Add an element to a set
print('fish' in animals) # Prints "True"
print(len(animals)) # Number of elements in a set; prints "3"
animals.add('cat') # Adding an element that is already in the set does nothing
print(len(animals)) # Prints "3"
animals.remove('cat') # Remove an element from a set
print(len(animals)) # Prints "2"
```

```
True
False
True
3
3
2
```

```
animals = {'cat', 'dog'}
print('cat' in animals)
print('fish' in animals)
animals.add('fish')
print('fish' in animals)
print(len(animals))
animals.add('cat')
print(len(animals))
animals.remove('cat')
print(len(animals))
```

File Input / Output

```
with open('myfile.txt', 'w') as file:  
    file.write('Hello World\nThis is Python File Input Output')  
  
with open('myfile.txt', 'r') as file:  
    text = file.read()  
print(text)
```

```
with open('myfile.txt', 'w') as file:  
    file.write('Hello World\nThis is Python File Input Output')
```

```
with open('myfile.txt', 'r') as file:  
    text = file.read()  
print(text)
```

```
Hello World  
This is Python File Input Output
```

```
text
```

```
'Hello World\nThis is Python File Input Output'
```

File Input / Output

```
with open('myfile.txt', 'a+') as file:  
    file.write('\n' + 'New line')
```

```
with open('myfile.txt', 'r') as file:  
    text = file.read()  
print(text)
```

```
with open('myfile.txt', 'a+') as file:  
    file.write('\n' + 'New line')
```

```
with open('myfile.txt', 'r') as file:  
    text = file.read()  
print(text)
```

```
Hello World  
This is Python File Input Output  
New line
```

try except finally

try:

```
file = open("myfile.txt")  
#file = open("myfile.txt", 'w')  
file.write("Python write file")  
print("file saved")
```

except:

```
print("Exception file Error")
```

finally:

```
file.close()  
print("finally process")
```

Exception file Error

finally process

class

```
class Person:
    def __init__(self, name, age):
        self.name = name
        self.age = age

    def myfunc(self):
        print("Hello my name is " + self.name)

p1 = Person("Alan", 20)
p1.myfunc()
print(p1.name)
print(p1.age)
```

```
Hello my name is Alan
Alan
20
```

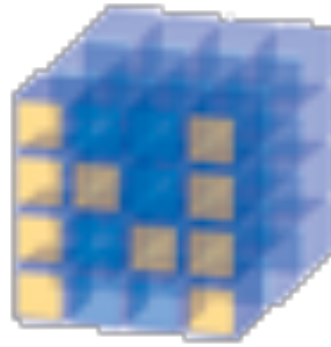
Big Data Analytics

with

Numpy

in Python

NumPy



NumPy
Base

**N-dimensional array
package**

NumPy
is the
fundamental package
for
scientific computing
with **Python.**



NumPy

- NumPy provides a **multidimensional array** object to store homogenous or heterogeneous data; it also provides **optimized functions/methods** to operate on this array object.

NumPy



NumPy

Scipy.org

NumPy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

NumPy is licensed under the *BSD license*, enabling reuse with few restrictions.

Getting Started

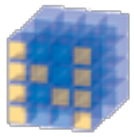
- [Getting NumPy](#)
- [Installing the SciPy Stack](#)
- [NumPy and SciPy documentation page](#)
- [NumPy Tutorial](#)
- [NumPy for MATLAB® Users](#)
- [NumPy functions by category](#)
- [NumPy Mailing List](#)

For more information on the SciPy Stack (for which NumPy provides the fundamental array data structure), see scipy.org.

About NumPy

[License](#)

[Old array packages](#)



NumPy

NumPy ndarray

One-dimensional Array (1-D Array)

0	1			n-1
1	2	3	4	5

Two-dimensional Array (2-D Array)

	0	1			n-1
0	1	2	3	4	5
1	6	7	8	9	10
	11	12	13	14	15
m-1	16	17	18	19	20



NumPy

```
v = list(range(1, 6))
```

```
v
```

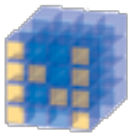
```
2 * v
```

```
import numpy as np
```

```
v = np.arange(1, 6)
```

```
v
```

```
2 * v
```



NumPy

Base

N-dimensional
array package

```
1 v = list(range(1, 6))  
2 v
```

```
[1, 2, 3, 4, 5]
```

```
1 2 * v
```

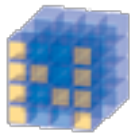
```
[1, 2, 3, 4, 5, 1, 2, 3, 4, 5]
```

```
1 import numpy as np  
2 v = np.arange(1, 6)  
3 v
```

```
array([1, 2, 3, 4, 5])
```

```
1 2 * v
```

```
array([ 2,  4,  6,  8, 10])
```



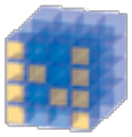
NumPy

NumPy Create Array

```
import numpy as np  
a = np.array([1, 2, 3])  
b = np.array([4, 5, 6])  
c = a * b  
c
```

```
import numpy as np  
a = np.array([1, 2, 3])  
b = np.array([4, 5, 6])  
c = a * b  
c
```

```
array([ 4, 10, 18])
```



NumPy

NumPy

```
1 import numpy as np
2
3 a = np.zeros((2,2)) # Create an array of all zeros
4 print(a)           # Prints "[[ 0.  0.]
5                     #           [ 0.  0.]]"
6
7 b = np.ones((1,2)) # Create an array of all ones
8 print(b)           # Prints "[[ 1.  1.]]"
9
10 c = np.full((2,2), 7) # Create a constant array
11 print(c)            # Prints "[[ 7.  7.]
12                    #           [ 7.  7.]]"
13
14 d = np.eye(2)       # Create a 2x2 identity matrix
15 print(d)           # Prints "[[ 1.  0.]
16                    #           [ 0.  1.]]"
17
18 e = np.random.random((2,2)) # Create an array filled with random values
19 print(e)            # Might print "[[ 0.91940167  0.08143941]
20                    #           [ 0.68744134  0.87236687]]"
```

```
[[0.  0.]
 [0.  0.]]
[[1.  1.]]
[[7 7]
 [7 7]]
[[1.  0.]
 [0.  1.]]
[[0.66258211 0.65552598]
 [0.00429934 0.21695824]]
```


Numpy Quickstart Tutorial

Quickstart tutorial

Prerequisites

Before reading this tutorial you should know a bit of Python. If you would like to refresh your memory, take a look at the [Python tutorial](#).

If you wish to work the examples in this tutorial, you must also have some software installed on your computer. Please see <http://scipy.org/install.html> for instructions.

The Basics

NumPy's main object is the homogeneous multidimensional array. It is a table of elements (usually numbers), all of the same type, indexed by a tuple of positive integers. In NumPy dimensions are called *axes*. The number of axes is *rank*.

For example, the coordinates of a point in 3D space `[1, 2, 1]` is an array of rank 1, because it has one axis. That axis has a length of 3. In the example pictured below, the array has rank 2 (it is 2-dimensional). The first dimension (axis) has a length of 2, the second dimension has a length of 3.

```
[[ 1., 0., 0.],  
 [ 0., 1., 2.]]
```

NumPy's array class is called `ndarray`. It is also known by the alias `array`. Note that `numpy.array` is not the same as the Standard Python Library class `array.array`, which only handles one-dimensional arrays and offers less functionality. The more important attributes of an `ndarray` object are:

`ndarray.ndim`

the number of axes (dimensions) of the array. In the Python world, the number of dimensions is referred to as *rank*.

`ndarray.shape`

Table Of Contents

- Quickstart tutorial
 - Prerequisites
 - The Basics
 - An example
 - Array Creation
 - Printing Arrays
 - Basic Operations
 - Universal Functions
 - Indexing, Slicing and Iterating
 - Shape Manipulation
 - Changing the shape of an array
 - Stacking together different arrays
 - Splitting one array into several smaller ones
 - Copies and Views
 - No Copy at All
 - View or Shallow Copy
 - Deep Copy
 - Functions and Methods Overview
 - Less Basic
 - Broadcasting rules
 - Fancy indexing and index tricks
 - Indexing with Arrays of

```
import numpy as np
```

```
a = np.arange(15).reshape(3, 5)
```

```
a.shape
```

```
a.ndim
```

```
a.dtype.name
```

```
import numpy as np  
a = np.arange(15).reshape(3, 5)  
a
```

```
array([[ 0,  1,  2,  3,  4],  
       [ 5,  6,  7,  8,  9],  
       [10, 11, 12, 13, 14]])
```

```
print(a.shape)
```

```
(3, 5)
```

```
a.ndim
```

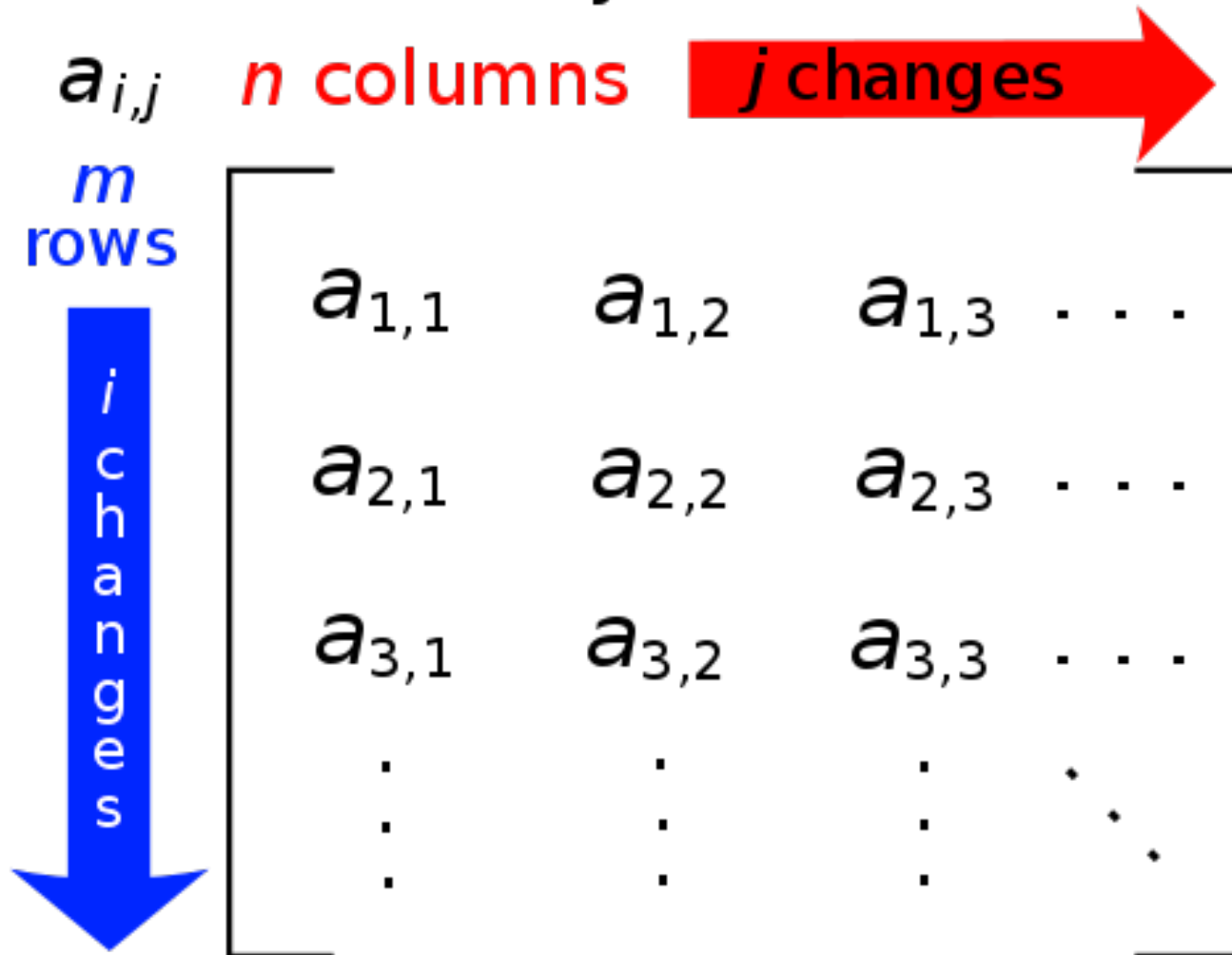
```
2
```

```
a.dtype.name
```

```
'int64'
```

Matrix

m -by- n matrix



NumPy ndarray: Multidimensional Array Object

NumPy ndarray

One-dimensional Array (1-D Array)

0	1			n-1
1	2	3	4	5

Two-dimensional Array (2-D Array)

	0	1		n-1	
0	1	2	3	4	5
1	6	7	8	9	10
	11	12	13	14	15
m-1	16	17	18	19	20

```
import numpy as np
a = np.array([1,2,3,4,5])
```

One-dimensional Array (1-D Array)

0	1			n-1
1	2	3	4	5

```
a = np.array([1,2,3,4,5])
a
```

```
array([1, 2, 3, 4, 5])
```

```
a = np.array([ [1,2,3,4,5],[6,7,8,9,10],[11,12,13,14,15],[16,17,18,19,20] ] )
```

Two-dimensional Array (2-D Array)

	0	1		n-1	
0	1	2	3	4	5
1	6	7	8	9	10
	11	12	13	14	15
m-1	16	17	18	19	20

```
a = np.array([[1,2,3,4,5],[6,7,8,9,10],[11,12,13,14,15],[16,17,18,19,20]])  
a
```

```
array([[ 1,  2,  3,  4,  5],  
       [ 6,  7,  8,  9, 10],  
       [11, 12, 13, 14, 15],  
       [16, 17, 18, 19, 20]])
```

```
import numpy as np
a = np.array([[0, 1, 2, 3],
              [10, 11, 12, 13],
              [20, 21, 22, 23]])
a
```

0	1	2	3
10	11	12	13
20	21	22	23


```
a = np.array ([[0, 1, 2, 3], [10, 11, 12, 13], [20, 21, 22, 23]])
```

```
a = np.array([[0, 1, 2, 3], [10, 11, 12, 13], [20, 21, 22, 23]])  
a
```

```
array([[ 0,  1,  2,  3],  
       [10, 11, 12, 13],  
       [20, 21, 22, 23]])
```

```
print(a.ndim)
```

```
2
```

```
print(a.shape)
```

```
(3, 4)
```

0	1	2	3
10	11	12	13
20	21	22	23

NumPy Basics: Arrays and Vectorized Computation

NumPy Array

axis 1

0

1

2

0

0,0

0,1

0,2

axis 0

1

1,0

1,1

1,2

2

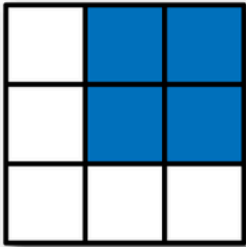
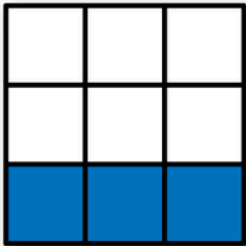
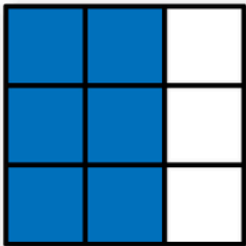
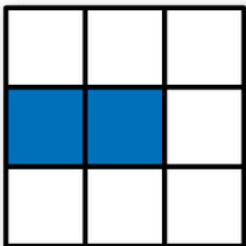
2,0

2,1

2,2

	0	1	2
0	0,0	0,1	0,2
1	1,0	1,1	1,2
2	2,0	2,1	2,2

Numpy Array


	Expression	Shape
	<code>arr[:2, 1:]</code>	<code>(2, 2)</code>
	<code>arr[2]</code> <code>arr[2, :]</code> <code>arr[2:, :]</code>	<code>(3,)</code> <code>(3,)</code> <code>(1, 3)</code>
	<code>arr[:, :2]</code>	<code>(3, 2)</code>
	<code>arr[1, :2]</code> <code>arr[1:2, :2]</code>	<code>(2,)</code> <code>(1, 2)</code>
















Wes McKinney (2017), "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", 2nd Edition, O'Reilly Media.

Materials and IPython notebooks for "Python for Data Analysis" by Wes McKinney, published by O'Reilly Media

52 commits 2 branches 0 releases 6 contributors

Branch: 2nd-edition ▾ New pull request Find file Clone or download ▾

 **betatim** committed with **wesm** Add requirements (#71) Latest commit ea47998 5 days ago

 datasets	Add Kaggle titanic dataset	5 months ago
 examples	Remove sex column from tips dataset	4 months ago
 .gitignore	Add gitignore	2 years ago
 COPYING	Use MIT license for code examples	a month ago
 README.md	Add launch in Azure Notebooks button (#70)	19 days ago
 appa.ipynb	Make more cells markdown instead of raw	a month ago
 ch02.ipynb	Make more cells markdown instead of raw	a month ago
 ch03.ipynb	Make more cells markdown instead of raw	a month ago
 ch04.ipynb	Convert all notebooks to v4 format	a month ago
 ch05.ipynb	Make more cells markdown instead of raw	a month ago
 ch06.ipynb	Make more cells markdown instead of raw	a month ago
 ch07.ipynb	Convert all notebooks to v4 format	a month ago
 ch08.ipynb	Make more cells markdown instead of raw	a month ago
 ch09.ipynb	Make more cells markdown instead of raw	a month ago
 ch10.ipynb	Make more cells markdown instead of raw	a month ago

<https://github.com/wesm/pydata-book>


Wes McKinney (2017), "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", 2nd Edition, O'Reilly Media.




wesm / pydata-book Watch 640 Star 4,031 Fork 4,348

[Code](#) [Issues 2](#) [Pull requests 0](#) [Projects 0](#) [Insights](#)

Branch: 2nd-edition **pydata-book / ch04.ipynb** Find file Copy path

wesm Convert all notebooks to v4 format c2780a0 on Sep 27

2 contributors 

1857 lines (1856 sloc) | 32.6 KB Raw Blame History   

NumPy Basics: Arrays and

```
In [ ]: import numpy as np
np.random.seed(12345)
import matplotlib.pyplot as plt
plt.rc('figure', figsize=(10, 6))
np.set_printoptions(precision=4, suppress=True)
```

```
In [ ]: import numpy as np
my_arr = np.arange(1000000)
my_list = list(range(1000000))
```

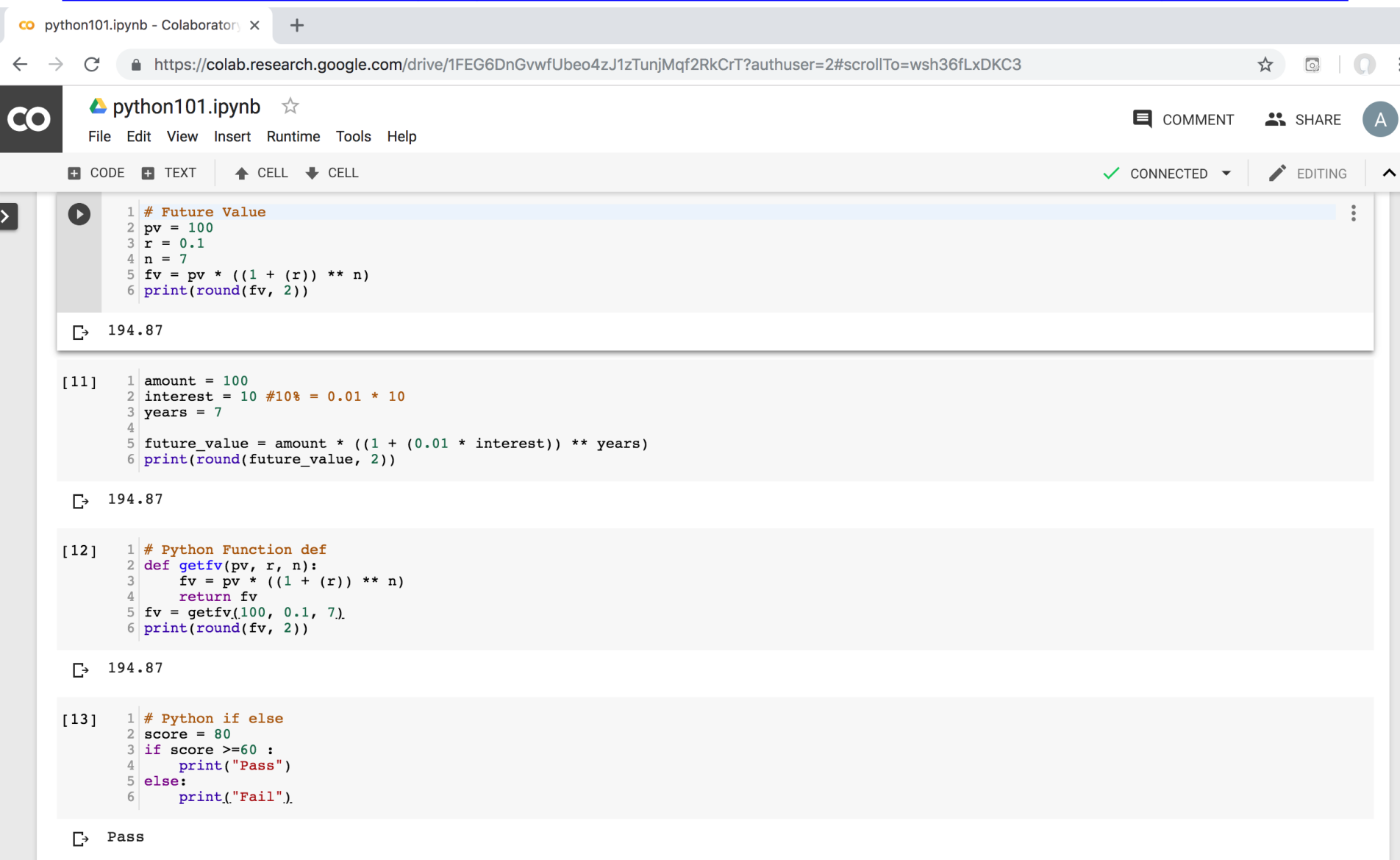
```
In [ ]: %time for _ in range(10): my_arr2 = my_arr * 2
%time for _ in range(10): my_list2 = [x * 2 for x in my_list]
```

The NumPy ndarray: A Multidimensional Array Object

```
In [ ]: import numpy as np
# Generate some random data
data = np.random.randn(2, 3)
data
```

Python in Google Colab (Python101)

<https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT>



python101.ipynb - Colaboratory

File Edit View Insert Runtime Tools Help

COMMENT SHARE

CONNECTED EDITING

```
1 # Future Value
2 pv = 100
3 r = 0.1
4 n = 7
5 fv = pv * ((1 + (r)) ** n)
6 print(round(fv, 2))
```

194.87

```
[11] 1 amount = 100
2 interest = 10 #10% = 0.01 * 10
3 years = 7
4
5 future_value = amount * ((1 + (0.01 * interest)) ** years)
6 print(round(future_value, 2))
```

194.87

```
[12] 1 # Python Function def
2 def getfv(pv, r, n):
3     fv = pv * ((1 + (r)) ** n)
4     return fv
5 fv = getfv(100, 0.1, 7)
6 print(round(fv, 2))
```

194.87

```
[13] 1 # Python if else
2 score = 80
3 if score >=60 :
4     print("Pass")
5 else:
6     print("Fail").
```

Pass

<https://tinyurl.com/aintpuppython101>

Summary

- **Foundations of Data Mining in Python**
 - **Python**
 - Programming language
 - **Numpy**
 - Scientific computing

References

- Robert Layton (2017), Learning Data Mining with Python - Second Edition, Packt Publishing.
- Wes McKinney (2017), "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", 2nd Edition, O'Reilly Media.
<https://github.com/wesm/pydata-book>
- Ties de Kok (2017), Learn Python for Research,
<https://github.com/TiesdeKok/LearnPythonforResearch>
- Avinash Jain (2017), Introduction To Python Programming, Udemy,
<https://www.udemy.com/pythonforbeginnersintro/>
- Python Programming, <https://pythonprogramming.net/>
- Python, <https://www.python.org/>
- Python Programming Language, <http://pythonprogramminglanguage.com/>
- Numpy, <http://www.numpy.org/>
- Pandas, <http://pandas.pydata.org/>
- Skikit-learn, <http://scikit-learn.org/>
- Min-Yuh Day (2021), Python 101, <https://tinyurl.com/aintpupython101>