# DIY DATA WORKSHOPS

## Introduction to Data Analytics

Lead Facilitator: **Reina Reyes, Ph.D.**

January 17, 2020 - Python PH

# Today's Schedule

| | | |
|---|---|---|
| **AM Lab** | 9:30-10:15 | Introduction to the Workshop & Set-up |
| | 10:15-11:00 | *Lesson 1:* Introduction to Python & Jupyter Notebook |
| | 11:00-12:00 | *Lesson 2:* Exploring the Forbes 100 Celebrity List |
| | 12:00-1:00 | Lunch |
| **PM Lab** | 1:00-1:20 | *Lesson 3:* Exploring Celebrity Wikipedia Edits |
| | 1:20-2:00 | *Lesson 4*: Wikipedia Metrics - Correlation Analysis |
| | 2:00-3:00 | *Lesson 5:* Celebrity Edits Time Series Charts |
| **Class Project** | 3:00-4:00 | Class Project – Teamwork |
| | 4:00-5:00 | Class Project Presentations |

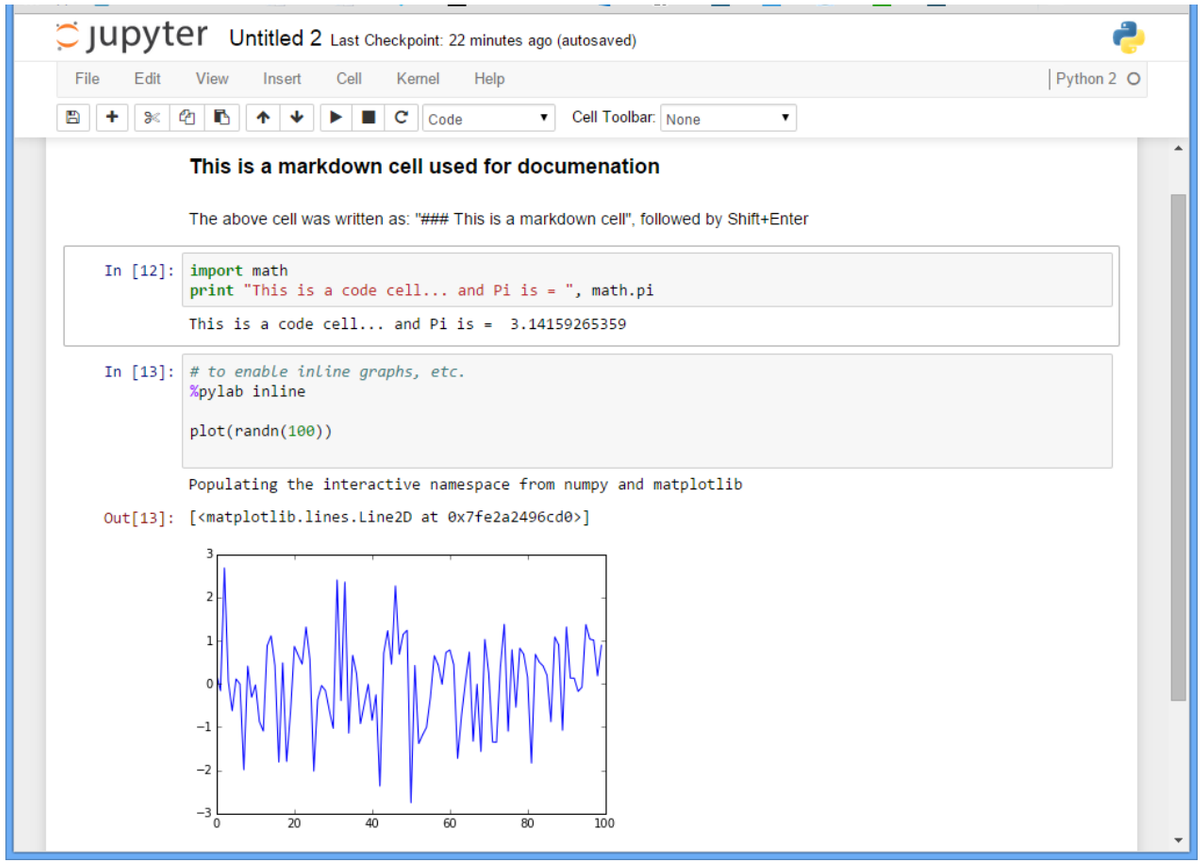
# *Why* DIY Data?

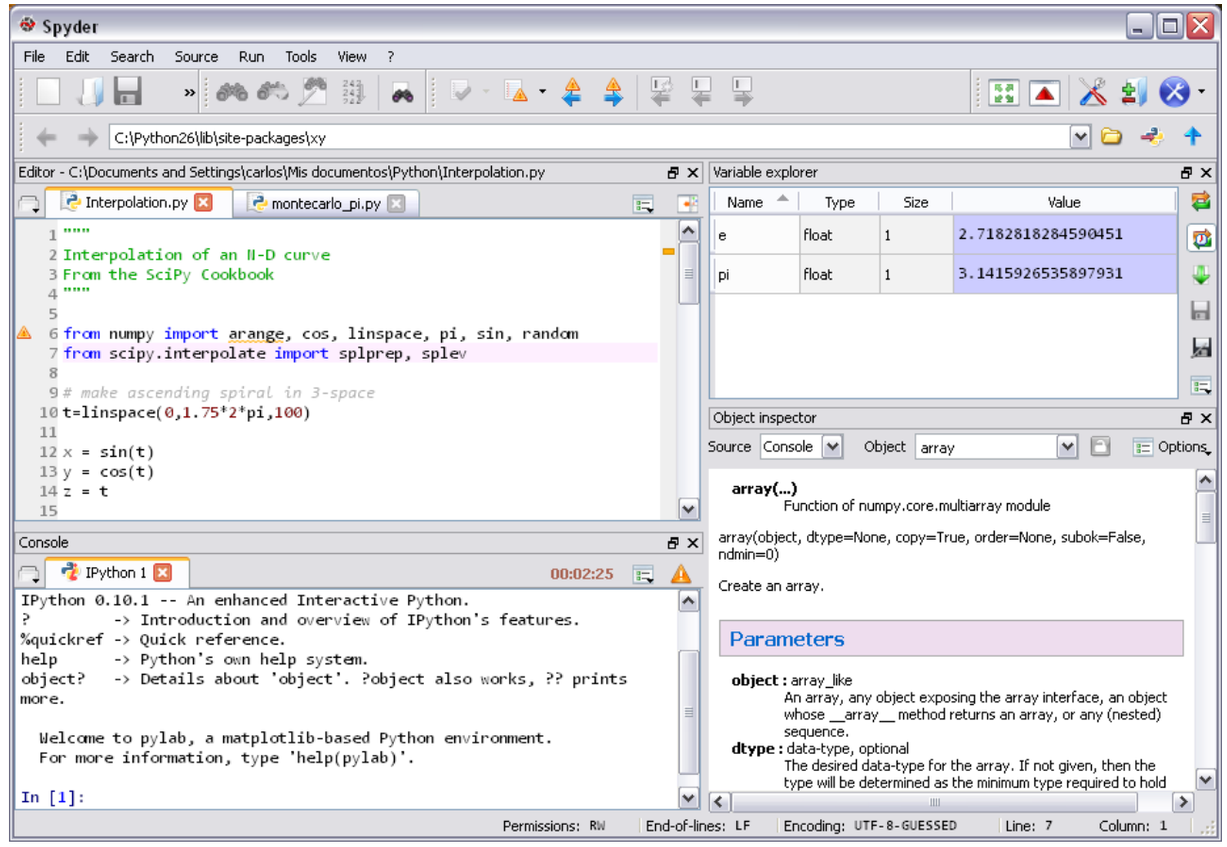
# *Why* DIY Data?

# *Why* Python?

# *Why* Jupyter?

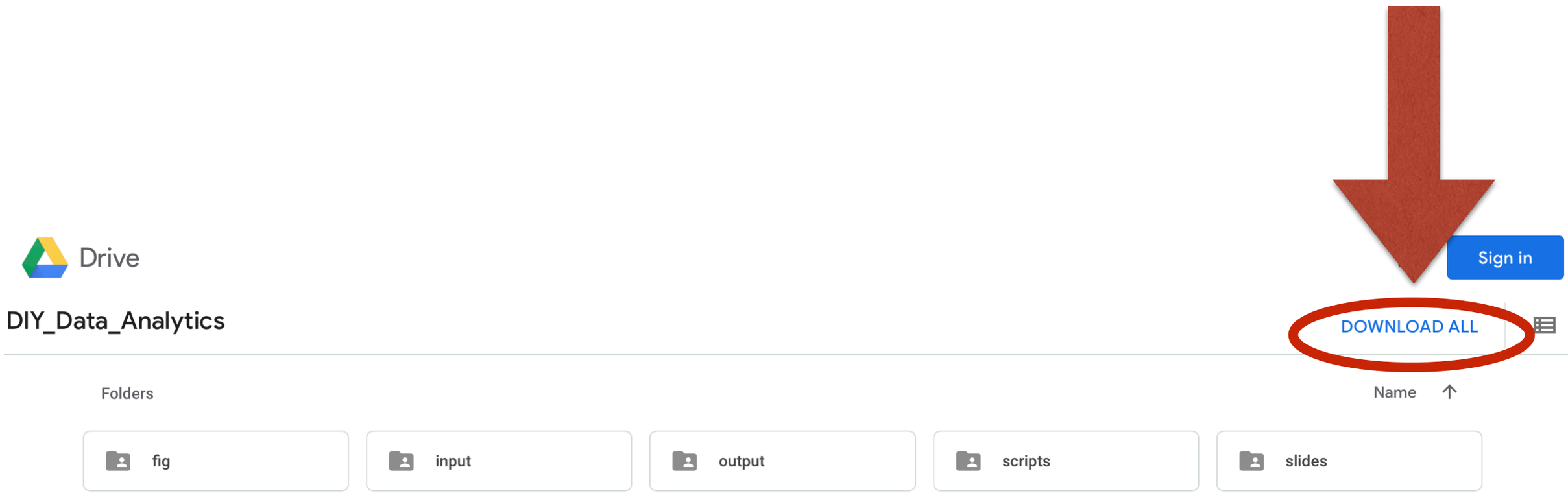
# Jupyter & Spyder



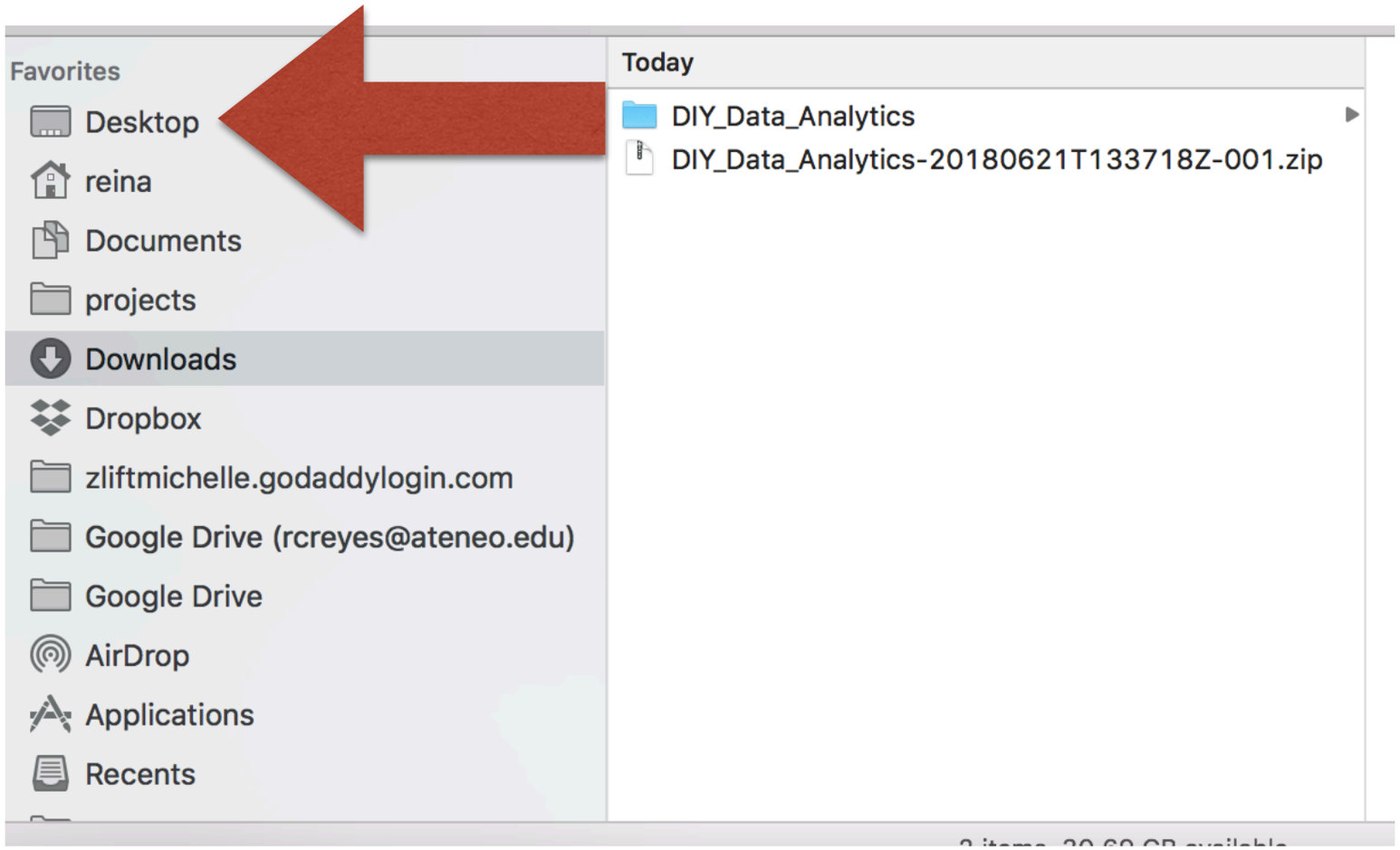



# Let’s get started!

# **Step 1:** Download: https://tinyurl.com/diy-data-analytics



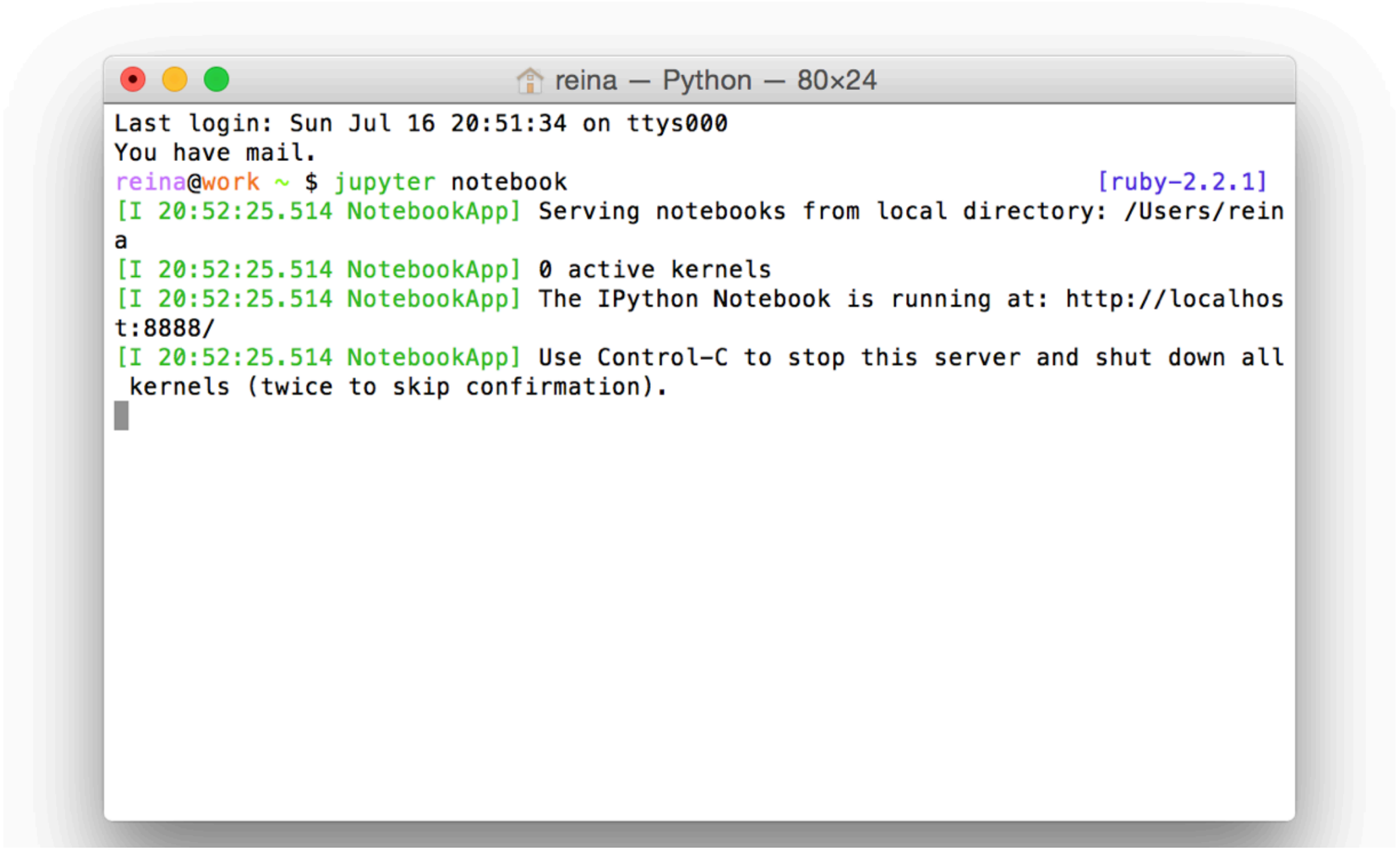
# **Step 2:** Unzip folder & move it to your Desktop



# **Step 3:** Launch Jupyter Notebook

*For Windows:* Go to the Start Menu and click on the Jupyter Notebook icon (installed by Anaconda)

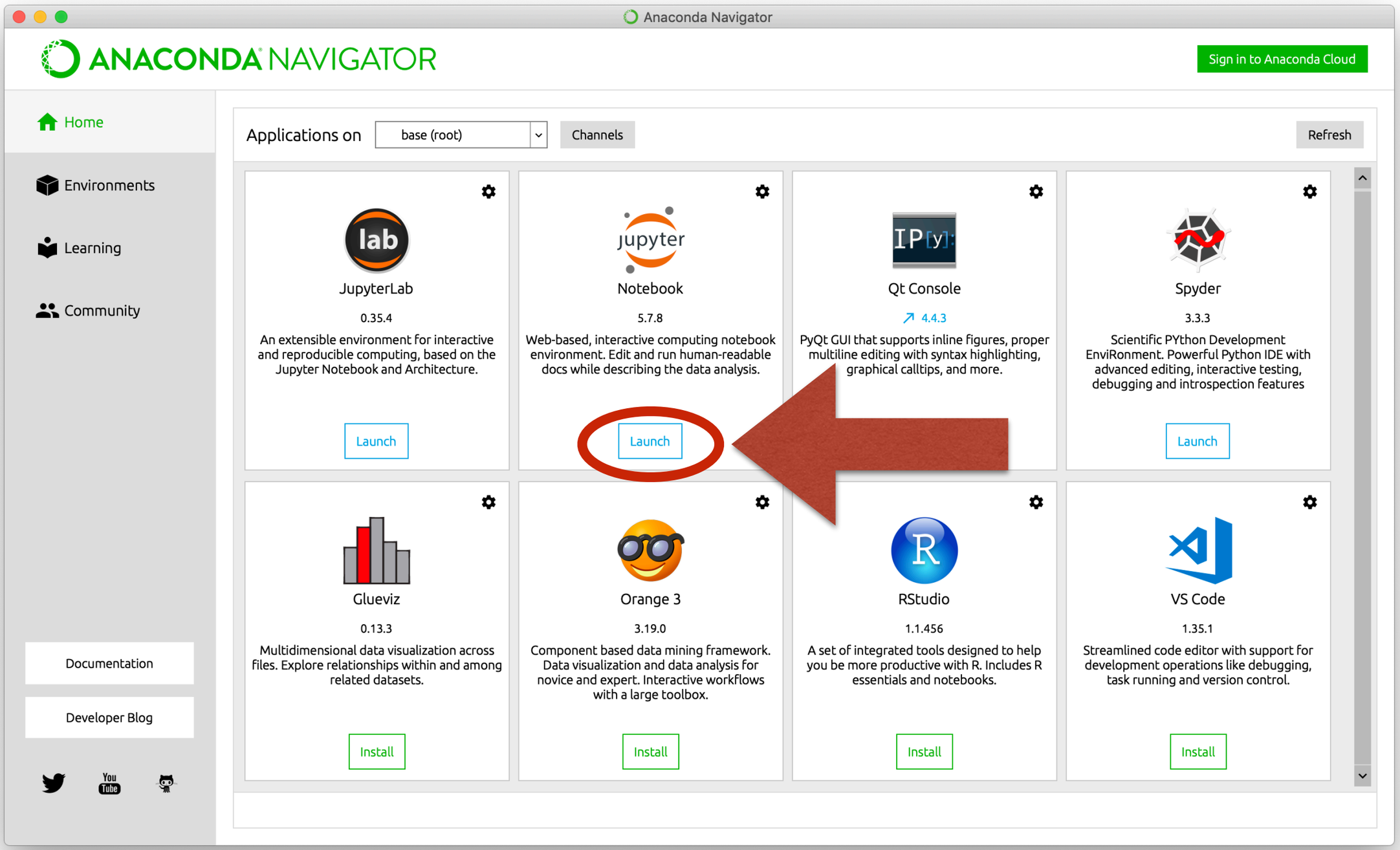
*For Mac OS:* Open your Terminal, type: "jupyter notebook" and hit enter. Your default web browser window should open. The terminal window should look like this screenshot:



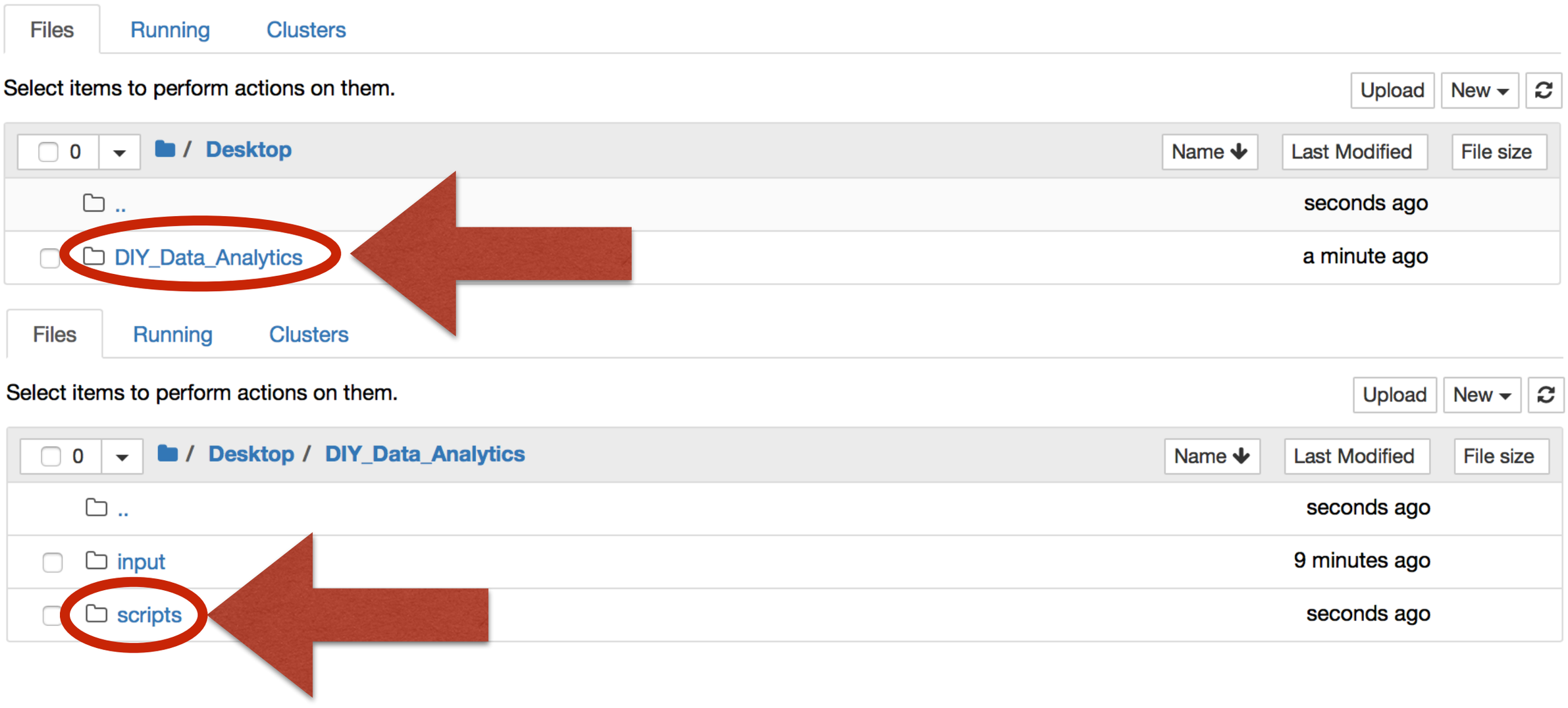
# **Step 3:** Launch Jupyter Notebook

Alternative for Mac OS:

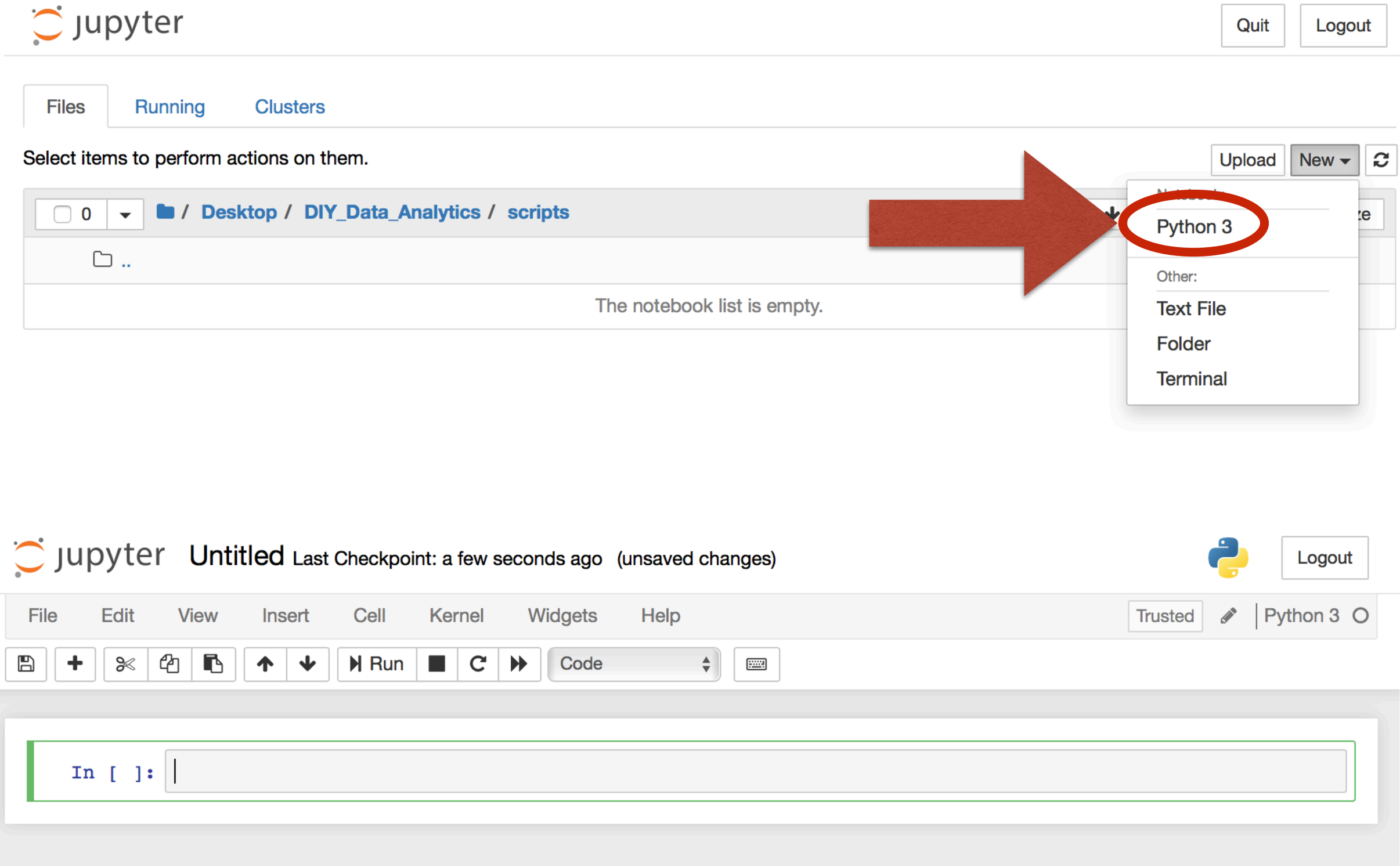
1. Go to Spotlight and type Anaconda-Navigator
2. Click “Launch” button for Jupiter Notebook



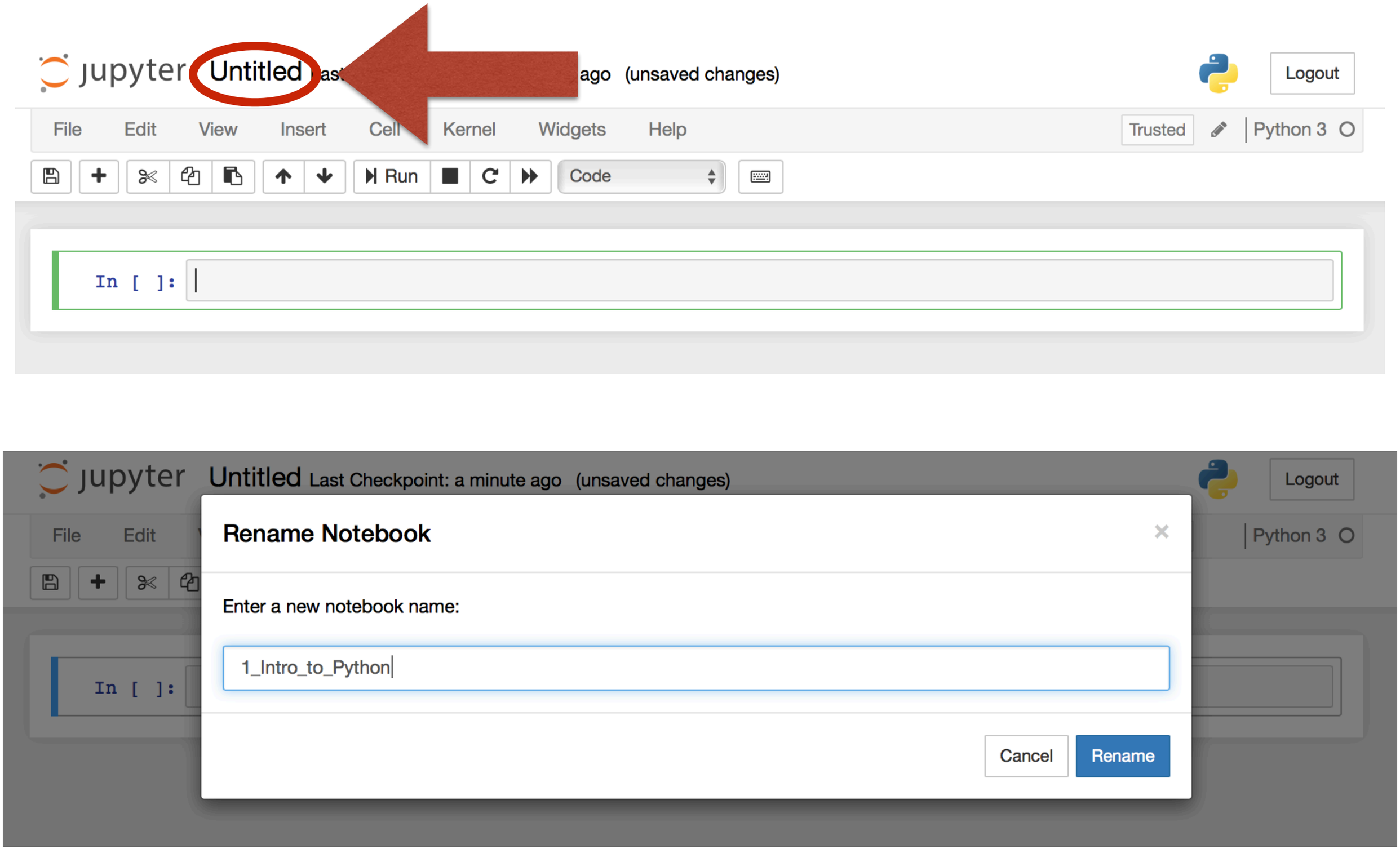
# **Step 4:** Go to **Desktop > DIY_Data_Analytics > scripts**

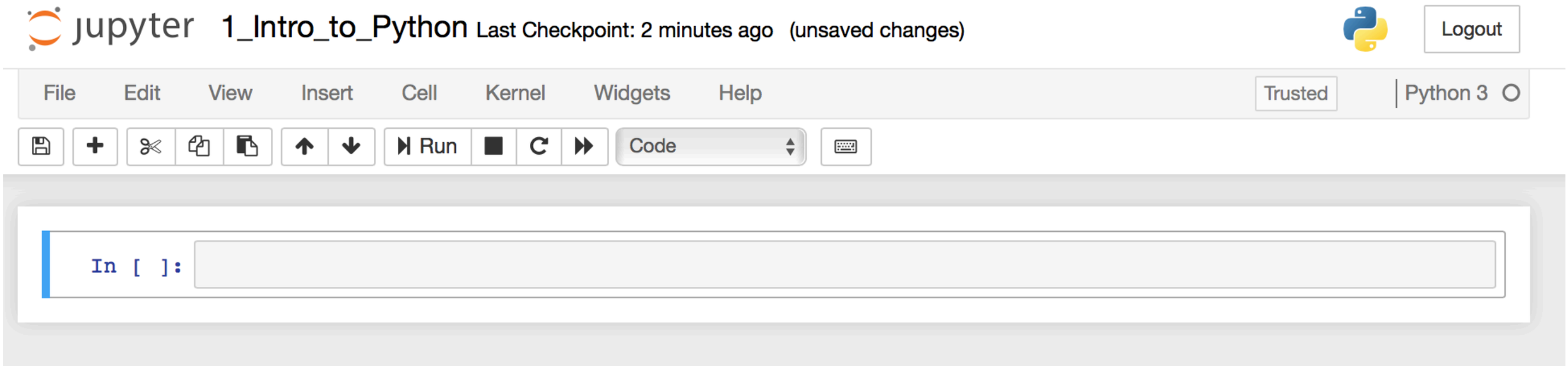
# **Step 4:** Open new Jupyter notebook



# **Step 5:** Rename Jupyter notebook



# This is your Jupyter notebook:



# **Lesson 1:**
# Introduction to Python & Jupyter Notebook

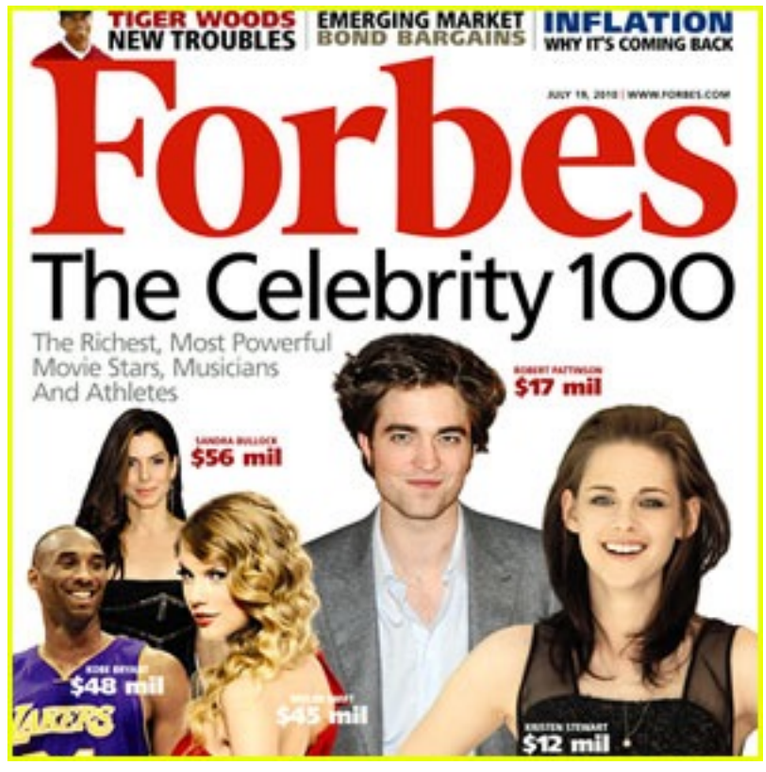
Refer to handout: **1_Intro_to_Python**

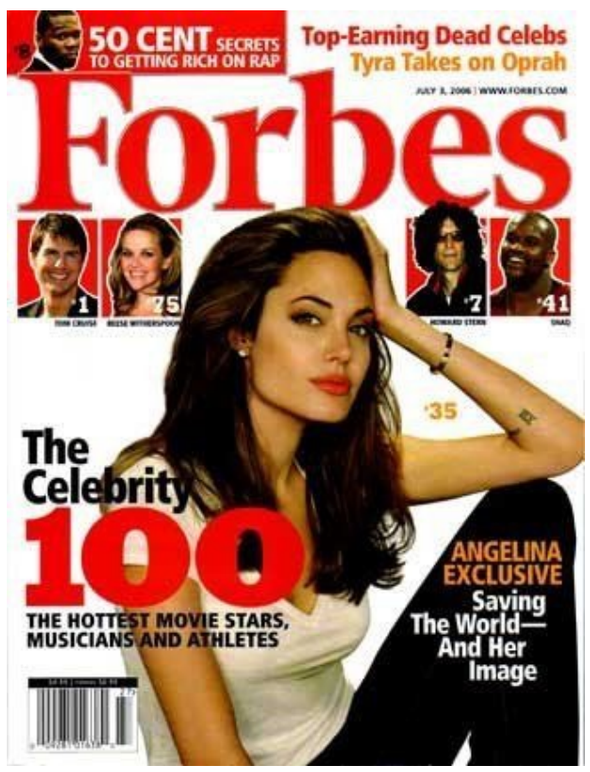
# **Lesson 2:**
# Forbes Celebrity 100 List

Refer to handout:
**2_Explore_Forbes_Celebrity_List**


Forbes
CELEBRITY
100


HORROR SHOW
ELI ROTH'S HOSTEL FRANCHISE
Millionaire Exodus at Google
Will You Pay $7 for Gas?
Short REITs Now
JULY 2, 2007 | WWW.FORBES.COM
Forbes
THE
Celebrity100
THE HOTTEST MOVIE STARS, MUSICIANS & ATHLETES
GISELE BUNDCHEN MODEL $33 MIL.
JAY-Z HIP HOP IMPRESARIO $83 MIL.
TIGER WOODS ATHLETE $100 MIL.
NICOLE KIDMAN ACTRESS $28 MIL.
RACHAEL RAY CHEF $16 MIL.
JOHNNY DEPP ACTOR $92 MILLION


TIGER WOODS NEW TROUBLES
EMERGING MARKET BOND BARGAINS
INFLATION WHY IT'S COMING BACK
JULY 19, 2010 | WWW.FORBES.COM
Forbes
The Celebrity 100
The Richest, Most Powerful Movie Stars, Musicians And Athletes
ROBERT PATTINSON $17 mil
SANDRA BULLOCK $56 mil
$48 mil
$45 mil
KRISTEN STEWART $12 mil


50 CENT SECRETS TO GETTING RICH ON RAP
Top-Earning Dead Celebs
Tyra Takes on Oprah
JULY 3, 2006 | WWW.FORBES.COM
Forbes
'1
'75
'7
'41
'35
The Celebrity 100
THE HOTTEST MOVIE STARS, MUSICIANS AND ATHLETES
ANGELINA EXCLUSIVE
Saving The World— And Her Image

# Forbes Celebrity 100



## *Forbes* Celebrity 100

From Wikipedia, the free encyclopedia

**Celebrity 100** is an annual list compiled and published by *Forbes* magazine since 1999. The purpose is to list the world's 100 highest-paid celebrities.



### Celebrity 100 lists [ edit ]

Below is the top 10 for each year since the list's inception.

#### 1990s [ edit ]

**1999**[1]

| No. | Recipient | Career |
|---|---|---|
| 1 | [Michael Jordan]] | Sportsperson (Basketball) |
| 2 | Oprah Winfrey | Television personality |

https://en.wikipedia.org/wiki/Forbes_Celebrity_100

# Forbes Celebrity 100

forbes_celebrity_100_w_title

Search Sheet

Home Insert Page Layout Formulas Data Review View

A2 | 1999

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Year | Rank | Recipient | Country | Career | Tied | Title |
| 2 | 1999 | 1 | Michael Jordan | United States | Sportsperson (Basketball) | 0 | Michael_Jordan |
| 3 | 1999 | 2 | Oprah Winfrey | United States | Television Personality | 0 | Oprah_Winfrey |
| 4 | 1999 | 3 | Leonardo DiCaprio | United States | Actor | 0 | Leonardo_DiCaprio |
| 5 | 1999 | 4 | Jerry Seinfeld | United States | Actor | 0 | Jerry_Seinfeld |
| 6 | 1999 | 5 | Steven Spielberg | United States | Filmmaker | 0 | Steven_Spielberg |
| 7 | 1999 | 6 | Spice Girls | United Kingdom | Musicians | 0 | Spice_Girls |
| 8 | 1999 | 7 | Harrison Ford | United States | Actor | 0 | Harrison_Ford |
| 9 | 1999 | 8 | Robin Williams | United States | Actor | 0 | Robin_Williams |
| 10 | 1999 | 9 | Celine Dion | Canada | Musician | 0 | Celine_Dion |
| 11 | 1999 | 10 | The Rolling Stones | United Kingdom | Musicians | 0 | The_Rolling_Stones |
| 12 | 2000 | 1 | Julia Roberts | United States | Actor | 0 | Julia_Roberts |
| 13 | 2000 | 2 | George Lucas | United States | Filmmaker | 0 | George_Lucas |
| 14 | 2000 | 3 | Oprah Winfrey | United States | Television Personality | 0 | Oprah_Winfrey |
| 15 | 2000 | 4 | Tom Hanks | United States | Actor | 0 | Tom_Hanks |
| 16 | 2000 | 5 | Michael Jordan | United States | Sportsperson (Basketball) | 0 | Michael_Jordan |
| 17 | 2000 | 6 | The Rolling Stones | United Kingdom | Musicians | 0 | The_Rolling_Stones |
| 18 | 2000 | 7 | Tiger Woods | United States | Sportsperson (Golf) | 0 | Tiger_Woods |
| 19 | 2000 | 8 | Backstreet Boys | United States | Musicians | 0 | Backstreet_Boys |
| 20 | 2000 | 9 | Cher | United States | Musician | 0 | Cher |
| 21 | 2000 | 10 | Steven Spielberg | United States | Filmmaker | 0 | Steven_Spielberg |
| 22 | 2001 | 1 | Tom Cruise | United States | Actor | 0 | Tom_Cruise |
| 23 | 2001 | 2 | Tiger Woods | United States | Sportsperson (Golf) | 0 | Tiger_Woods |
| 24 | 2001 | 3 | The Beatles | United Kingdom | Musicians | 0 | The_Beatles |

forbes_celebrity_100_w_title

Ready 150%

input/forbes_celebrity_list.csv

# **Lesson 3:**
# Wikipedia Edits

Refer to handout:
**3_Explore_Wikipedia_Edits**

# Wikipedia Page







Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

Create a book

# Oprah Winfrey

From Wikipedia, the free encyclopedia

*Not to be confused with Orpah.*
*"Oprah" redirects here. For the talk show, see The Oprah Winfrey Show.*

**Oprah Winfrey** (born **Orpah Gail Winfrey**;[1] January 29, 1954) is an American media proprietor, talk show host, actress, producer, and philanthropist. She is best known for her talk show *The Oprah Winfrey Show*, which was the highest-rated television program of its kind in history and was nationally syndicated from 1986 to 2011 in Chicago, Illinois.[6] Dubbed the "Queen of All Media",[7] she was the richest African American of the 20th century[8][9] and North America's first multi-billionaire black person[10], and has been ranked the greatest black philanthropist in American history.[11][12] Several assessments rank her as the most influential woman in the world.[13][14]

Winfrey was born into poverty in rural Mississippi to a teenage single mother and later raised in an inner-city Milwaukee neighborhood. She has stated that she was molested during her childhood and early teens and became pregnant at 14; her son died in infancy.[15] Sent to live with the man she calls her father, Vernon Winfrey, a barber in Tennessee, she landed a job in radio while still in high school and began co-anchoring the local evening news at the age of 19. Her emotional ad-lib delivery eventually got her transferred to the daytime talk show arena, and after boosting a third-rated local Chicago talk show to first place,[16] she launched her own production company and became internationally syndicated.

Credited with creating a more intimate confessional form of media communication,[17] she is thought to have popularized and revolutionized[17][18] the tabloid talk show genre pioneered by Phil Donahue,[17] through which, according to a Yale study, she broke 20th-century taboos and allowed LGBT people to enter the

**Oprah Winfrey**

Oprah Winfrey in October 2014

| | |
|---|---|
| **Born** | Orpah Gail Winfrey[1]<br>January 29, 1954 (age 64) |

https://en.wikipedia.org/wiki/Oprah_Winfrey

# Wikipedia Edits

Article | Talk | Read | View source | View history | Search Wikipedia

## Oprah Winfrey: Revision history

Help

View logs for this page

Search for revisions

From year (and earlier): 2018 From month (and earlier): all Tag filter: Show

For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary.

External tools: Revision history statistics · Revision history search · Edits by user · Number of watchers · Page view statistics · Fix dead links

(cur) = difference from current version, (prev) = difference from preceding version, **m** = minor edit, → = section edit, ← = automatic edit summary

(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- (cur | prev) 04:06, 14 June 2018 SennaNiks (talk | contribs) . . (122,009 bytes) **(+541)** . . *(information added regarding new series)* *(Tag: Visual edit)*
- (cur | prev) 13:22, 11 June 2018 Ser Amantio di Nicolao (talk | contribs) **m** . . (121,468 bytes) (-32) . . *(Removing from Category:American actresses using Cat-a-lot)*
- (cur | prev) 13:15, 8 June 2018 Rivertorch (talk | contribs) . . (121,500 bytes) (0) . . *(Undid revision 844939355 by FutureLitigator (talk) per source)* *(Tag: Undo)*
- (cur | prev) 06:45, 8 June 2018 FutureLitigator (talk | contribs) . . (121,500 bytes) (0) . . *(Typo. 1968 to 1986)* *(Tags: Mobile edit, Mobile web edit)*
- (cur | prev) 21:58, 5 June 2018 Michaelhurwicz (talk | contribs) **m** . . (121,500 bytes) (0) . . *(→Romantic history)*
- (cur | prev) 21:55, 5 June 2018 Michaelhurwicz (talk | contribs) . . (121,500 bytes) (+38) . . *(→Homes)*
- (cur | prev) 19:43, 4 June 2018 JustAMuggle (talk | contribs) . . (121,462 bytes) (+1) . . *(Copy edit.)*
- (cur | prev) 23:55, 31 May 2018 Leschnei (talk | contribs) . . (121,461 bytes) (+24) . . *(→As actress: Lincoln to Lincoln (1992 TV Movie))*
- (cur | prev) 03:15, 29 May 2018 TypoBoy (talk | contribs) **m** . . (121,437 bytes) (+6) . . *(Supply missing conjunction)*

# Wikipedia Edits

| title | parentid | revid | timestamp | user | userid | size |
|---|---|---|---|---|---|---|
| Oprah_Winfrey | 845388698 | 845787664 | 2018-06-14T04:06:37Z | SennaNiks | 33247211 | 122009 |
| Oprah_Winfrey | 844973393 | 845388698 | 2018-06-11T13:22:13Z | Ser Amantio di Nicolao | 753665 | 121468 |
| Oprah_Winfrey | 844939355 | 844973393 | 2018-06-08T13:15:39Z | Rivertorch | 1938168 | 121500 |
| Oprah_Winfrey | 844597107 | 844939355 | 2018-06-08T06:45:01Z | FutureLitigator | 27024836 | 121500 |
| Oprah_Winfrey | 844596756 | 844597107 | 2018-06-05T21:58:29Z | Michaelhurwicz | 24636444 | 121500 |
| Oprah_Winfrey | 844416252 | 844596756 | 2018-06-05T21:55:31Z | Michaelhurwicz | 24636444 | 121500 |
| Oprah_Winfrey | 843863095 | 844416252 | 2018-06-04T19:43:23Z | JustAMuggle | 17618788 | 121462 |
| Oprah_Winfrey | 843429951 | 843863095 | 2018-05-31T23:55:43Z | Leschnei | 27335766 | 121461 |
| Oprah_Winfrey | 843142172 | 843429951 | 2018-05-29T03:15:28Z | TypoBoy | 5337685 | 121437 |
| Oprah_Winfrey | 843123137 | 843142172 | 2018-05-27T04:56:01Z | Rivertorch | 1938168 | 121431 |
| Oprah_Winfrey | 843122973 | 843123137 | 2018-05-27T01:25:46Z | CleverConservative1 | 30361015 | 121440 |
| Oprah_Winfrey | 843121836 | 843122973 | 2018-05-27T01:24:23Z | CleverConservative1 | 30361015 | 121436 |
| Oprah_Winfrey | 843121157 | 843121836 | 2018-05-27T01:14:53Z | CleverConservative1 | 30361015 | 121431 |
| Oprah_Winfrey | 843120271 | 843121157 | 2018-05-27T01:08:56Z | CleverConservative1 | 30361015 | 121430 |
| Oprah_Winfrey | 843119950 | 843120271 | 2018-05-27T01:01:03Z | CleverConservative1 | 30361015 | 121430 |
| Oprah_Winfrey | 843115960 | 843119950 | 2018-05-27T00:58:26Z | CleverConservative1 | 30361015 | 121429 |
| Oprah_Winfrey | 843114414 | 843115960 | 2018-05-27T00:24:19Z | CleverConservative1 | 30361015 | 121428 |
| Oprah_Winfrey | 842604284 | 843114414 | 2018-05-27T00:09:34Z | CleverConservative1 | 30361015 | 121427 |

input/wikipedia_edits.csv

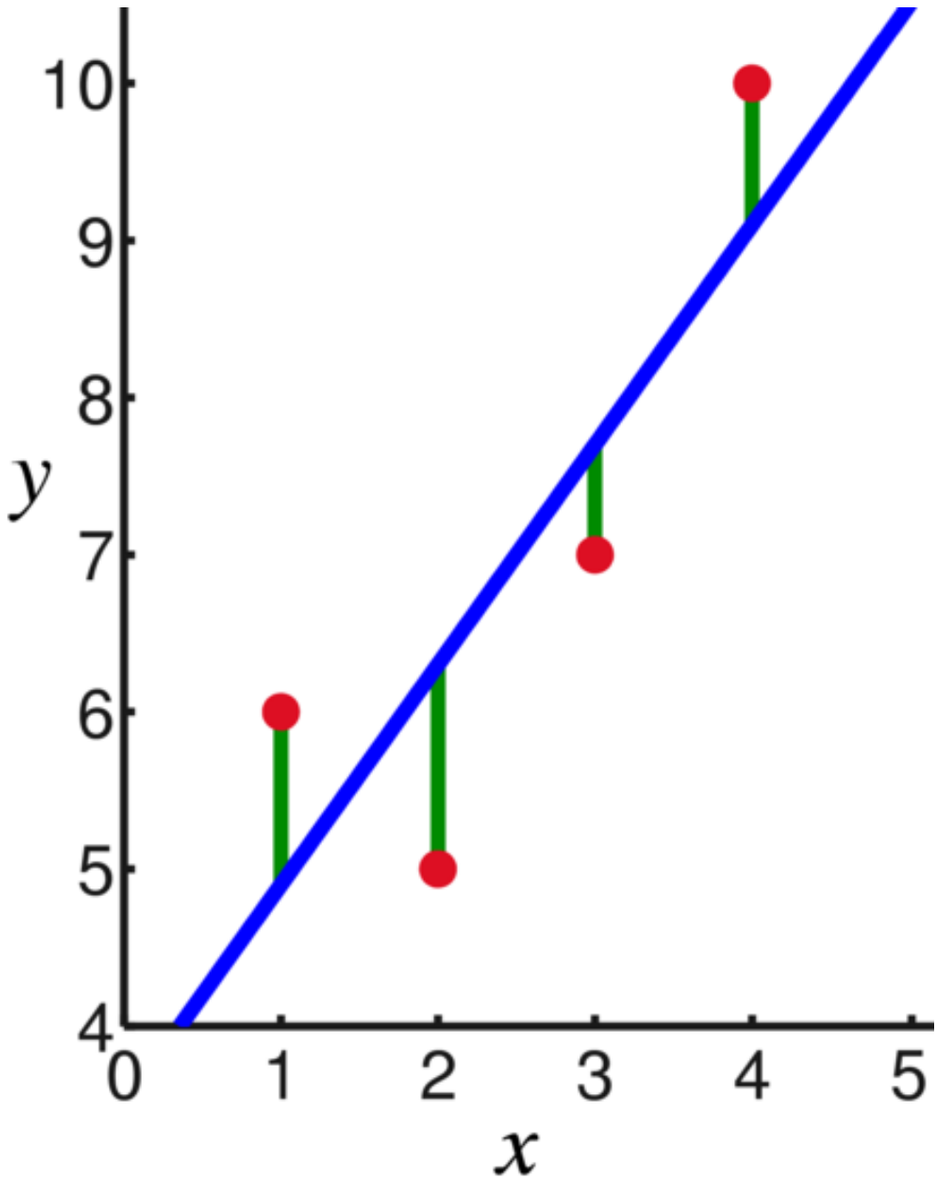
# **Lesson 4:**
# Wikipedia Metrics - Correlation Analysis

Refer to handouts:

**4a_Wikipedia_Metrics_Table**

**4b_Wikipedia_Metrics_Plots**

# Linear Regression




Linear model:

$$y = mx + b$$

slope

intercept

Find *m* and *b* that minimises:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

predicted value

actual value

# **Lesson 5:**
# Celebrity Edits Time Series

Refer to handout:
**5_Celebrity_Edits_Time_Series**

# Class Project

- Group into 3's and choose a team name! Create a notebook with your Team Name as title

- Formulate at least 3 questions about the Forbes 100 celebrities that can be answered by our datasets

- Analyze the data to answer your Top 3 Questions. Discuss and interpret the results.

- Email your notebook to diydataworkshops@gmail.com

- Present your notebook to the class for 5 minutes/team