NOVA IMS INFORMATION MANAGEMENT SCHOOL

# Machine Learning

## Practical Classes and Anaconda

Ricardo Santos
rcsantos@novaims.unl.pt

# Who am I?

- **Adjunct Lecturer at Nova IMS (2022-present day)**
  - Machine Learning
  - Text Mining
  - Data Mining I & Data Mining II

- **PhD Student**
  - Machine Learning Applications
  - Learning Analytics
  - Natural Language Processing

- **Academic Background**
  - MSc in Data Science and Advanced Analytics (2022)
  - BSc in Economics (2020)
  - BSc & Post-Grad in Biochemistry (2013)

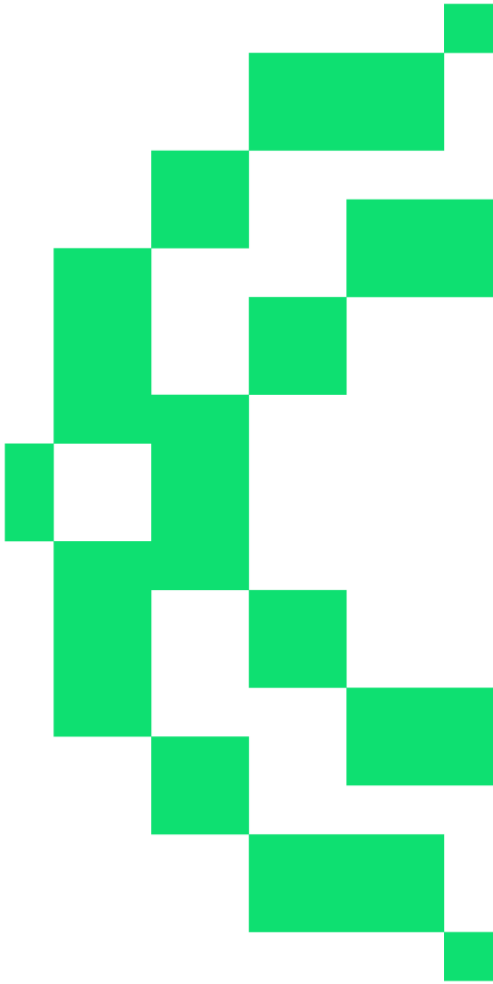
NOVA IMS Information Management School

1 2 9 0 UNIVERSIDADE D COIMBRA

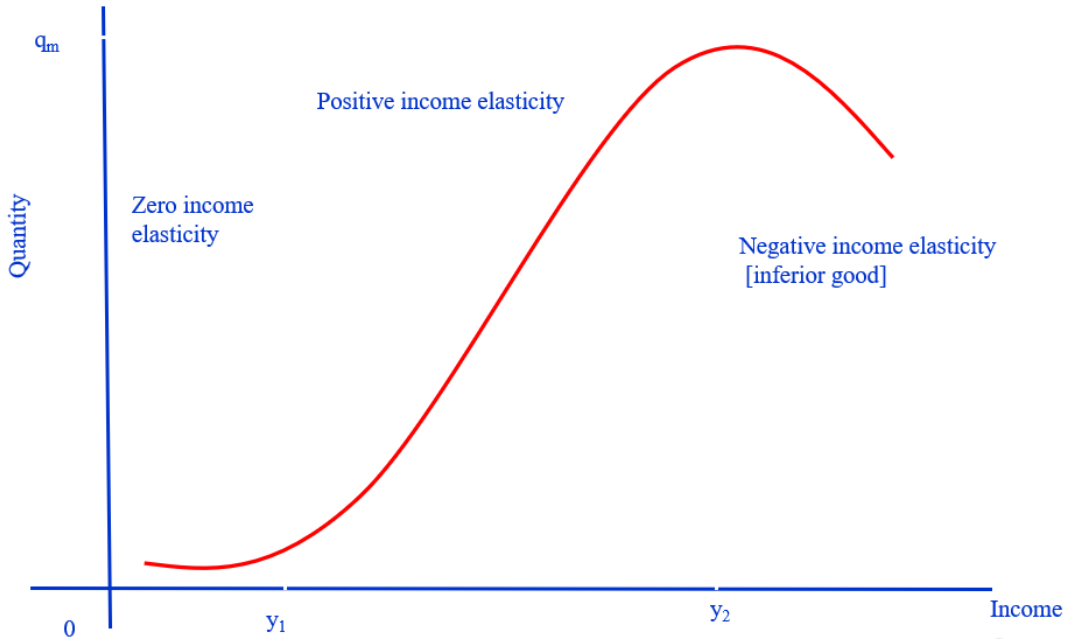
Office Hours: Thursdays 10h00-11h00

Office 138

# *What happened to your face?*



## Relationship between Income and Demand




Lipsey, R., & Chrystal, A. (2011). *Economics*. Oxford University Press, USA.

## Bacterial growth over time

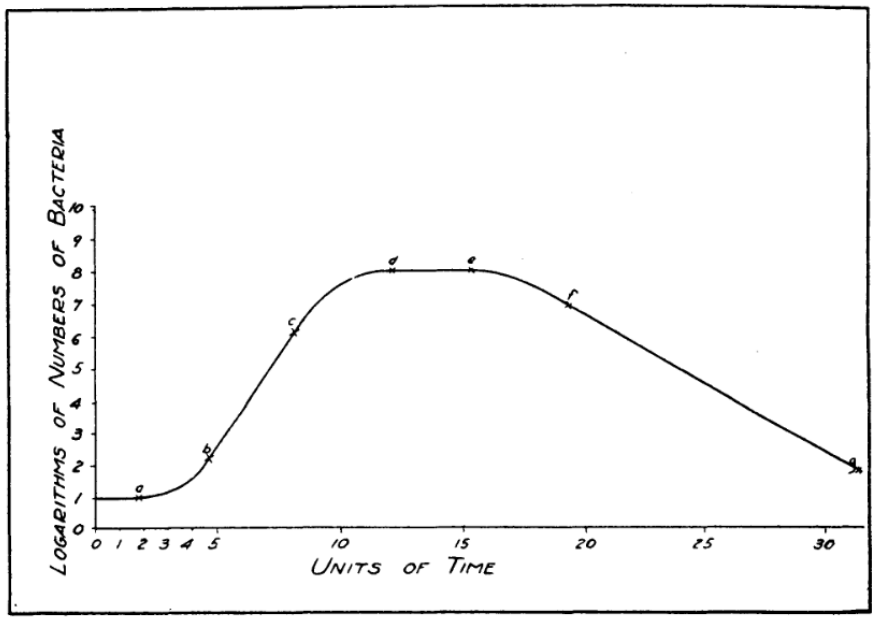
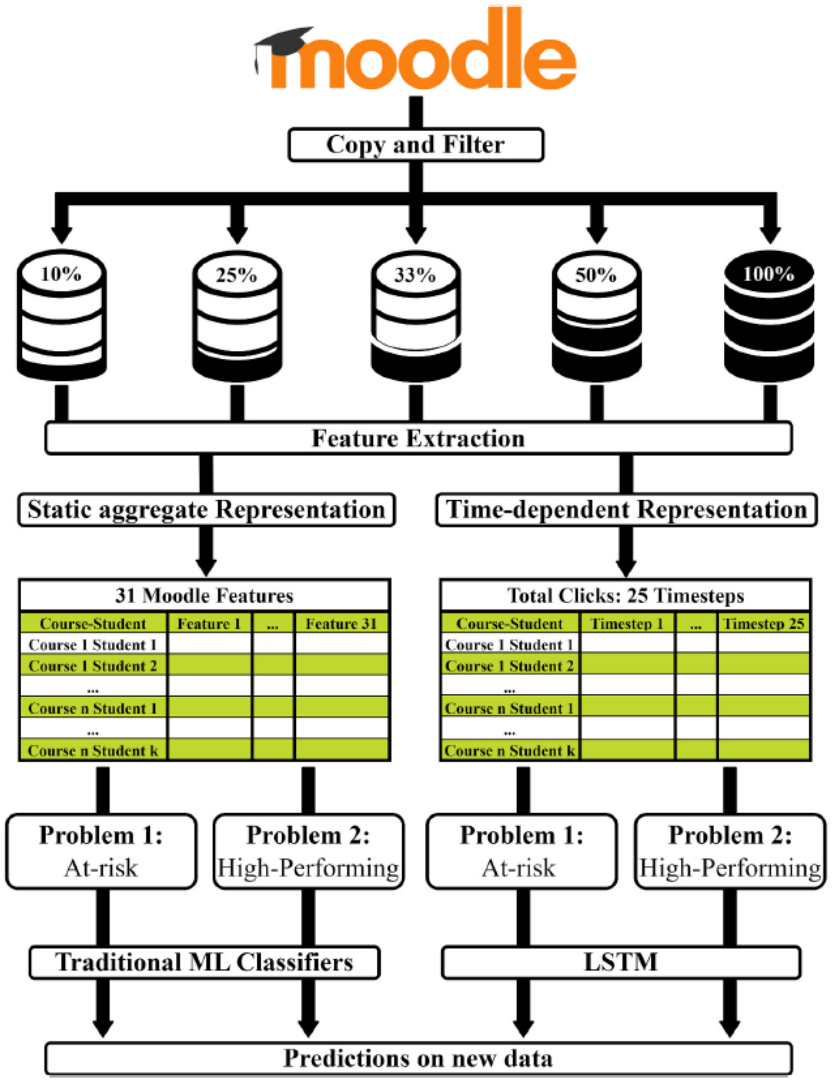



Chart 1.—Diagrammatic plot of logarithms of numbers of bacteria present in a culture.

Buchanan, R. E. (1918). Life Phases in a Bacterial Culture. *The Journal of Infectious Diseases*, *23*(2), 109–125.

# Main Research Focus: Machine Learning Applications in Higher Education

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | eventnam | componer | action | target | objecttable | userid | courseid | TStamp | CourseFullname | CourseShortname | cd_discip | startdate | enddate |
| 2 | \core\eve | core | viewed | course | NULL | 6563 | 1177 | 01/03/2021 10:56 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 3 | \core\eve | core | viewed | course | NULL | 6567 | 1177 | 02/03/2021 23:30 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 4 | \core\eve | core | viewed | course | NULL | 6135 | 1177 | 02/03/2021 23:32 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 5 | \core\eve | core | viewed | course | NULL | 5810 | 1177 | 03/03/2021 18:33 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 6 | \core\eve | core | viewed | course | NULL | 5834 | 1177 | 03/03/2021 19:12 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 7 | \core\eve | core | viewed | course | NULL | 6051 | 1177 | 04/03/2021 12:20 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 8 | \core\eve | core | viewed | course | NULL | 6051 | 1177 | 04/03/2021 12:20 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 9 | \mod_res | mod_reso | viewed | course_module | resource | 6051 | 1177 | 04/03/2021 12:20 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 10 | \core\eve | core | viewed | course | NULL | 6068 | 1177 | 04/03/2021 19:04 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 11 | \core\eve | core | viewed | course | NULL | 5760 | 1177 | 06/03/2021 10:46 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 12 | \core\eve | core | viewed | course | NULL | 6106 | 1177 | 06/03/2021 11:24 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 13 | \core\eve | core | viewed | course | NULL | 6149 | 1177 | 06/03/2021 15:32 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 14 | \mod_url | mod_url | viewed | course_module | url | 6149 | 1177 | 06/03/2021 15:37 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 15 | \core\eve | core | viewed | course | NULL | 6149 | 1177 | 06/03/2021 15:44 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 16 | \mod_res | mod_reso | viewed | course_module | resource | 6149 | 1177 | 06/03/2021 15:46 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 17 | \core\eve | core | viewed | course | NULL | 6149 | 1177 | 06/03/2021 15:46 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 18 | \core\eve | core | viewed | course | NULL | 6060 | 1177 | 07/03/2021 16:06 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 19 | \graderep | graderepc | viewed | grade_report | NULL | 6060 | 1177 | 07/03/2021 16:06 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 20 | \core\eve | core | viewed | course | NULL | 6060 | 1177 | 07/03/2021 17:07 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 21 | \core\eve | core | viewed | user_list | course | 6060 | 1177 | 07/03/2021 17:07 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 22 | \core\eve | core | viewed | course | NULL | 6060 | 1177 | 07/03/2021 17:07 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 23 | \core\eve | core | viewed | course | NULL | 6060 | 1177 | 07/03/2021 17:46 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 24 | \core\eve | core | viewed | course | NULL | 1584 | 1177 | 08/03/2021 10:21 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 25 | \core\eve | core | viewed | course | NULL | 6149 | 1177 | 08/03/2021 14:42 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 26 | \core\eve | core | viewed | course | NULL | 5846 | 1177 | 09/03/2021 16:33 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 27 | \core\eve | core | viewed | course | NULL | 6106 | 1177 | 09/03/2021 22:24 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 28 | \core\eve | core | viewed | course | NULL | 6671 | 1177 | 10/03/2021 16:03 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 29 | \core\eve | core | viewed | course | NULL | 6096 | 1177 | 11/03/2021 00:03 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 30 | \core\eve | core | viewed | course | NULL | 1574 | 1177 | 11/03/2021 01:09 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 31 | \core\eve | core | viewed | course | NULL | 1597 | 1177 | 11/03/2021 11:31 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 32 | \core\eve | core | viewed | course | NULL | 1584 | 1177 | 11/03/2021 15:35 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 33 | \core\eve | core | viewed | course | NULL | 6106 | 1177 | 11/03/2021 18:27 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |
| 34 | \core\eve | core | viewed | course | NULL | 6132 | 1177 | 12/03/2021 17:42 | MSI 2020/21 - Marketin | MSI-2020/21 | 200187 | 03/08/2020 15:02 | 30/08/2021 17:02 |



# Today's Agenda

01 Ground Rules & Grading

02 Tools

03 Setting up your environment

04 Initial Exercise

# 01.
# Ground Rules

# General Rules for Practical Classes

- 1 Practical Class per Week
- **In each week, practicals will focus on implementing the techniques covered in the theoretical classes.**
- Attendance in the practical classes is not strictly mandatory to complete the course.
- Please **attend the schedule you are assigned to** and register your presence using your card.

# Grading Rules - Project

***The practical component of the course will have one final group project.***

**Group Project (40% of Final Grade) – Due December 22nd at 18h00**

- Practical implementation of ML to solve a classification/regression problem
- You will receive a project specification, some labelled data (for training) and unlabelled data (for testing)
- You will need to follow the steps of a ML project to create a predictive model with your training data and use that model to make predictions with your test data.
- Evaluation is contingent on (i) quality of deliverables, (ii) performance in the discussion and (iii) peer evaluation on performance in the group.
- Form to select your group will be shared with you on Moodle shortly. **Please select a group before until the end of October 3rd (Each group can have between 3 or 4 members – no exceptions).**

**Disclaimers**

- The project grade is a one-time event (no second chances) and plays participation in all moments of evaluation is required to obtain approval in the course in both normal and resit seasons.
- The project discussion will take place during the exam season (specific date tbd but will be announced asap)
- If you get less than 5.00 (out of 20) in the project, you will automatically fail the course even if you do well on your exams.

# Grading Rules - Handout

***The practical component of the course will have one final group project.***

**Intermediate Deliverable (10%) – Due November 3rd**

- Group Submission of Project Code that should feature:
  - End-to-End implementation of a working ML pipeline for your project
  - At least one submission for the Project's Kaggle competition (more details on this later)

- A 2-page document containing:

  - A Schematic representation of your pipeline: main stages, techniques used at each stage.
  - A description of your main preprocessing and feature selection stages: what is done, to what variables and why.

# 02.
# Tools

# Why Python?

***Python is a programming language that allows you to work quickly and integrate systems more effectively.***

Widely used for general-purpose programming

Fastest growing programming language

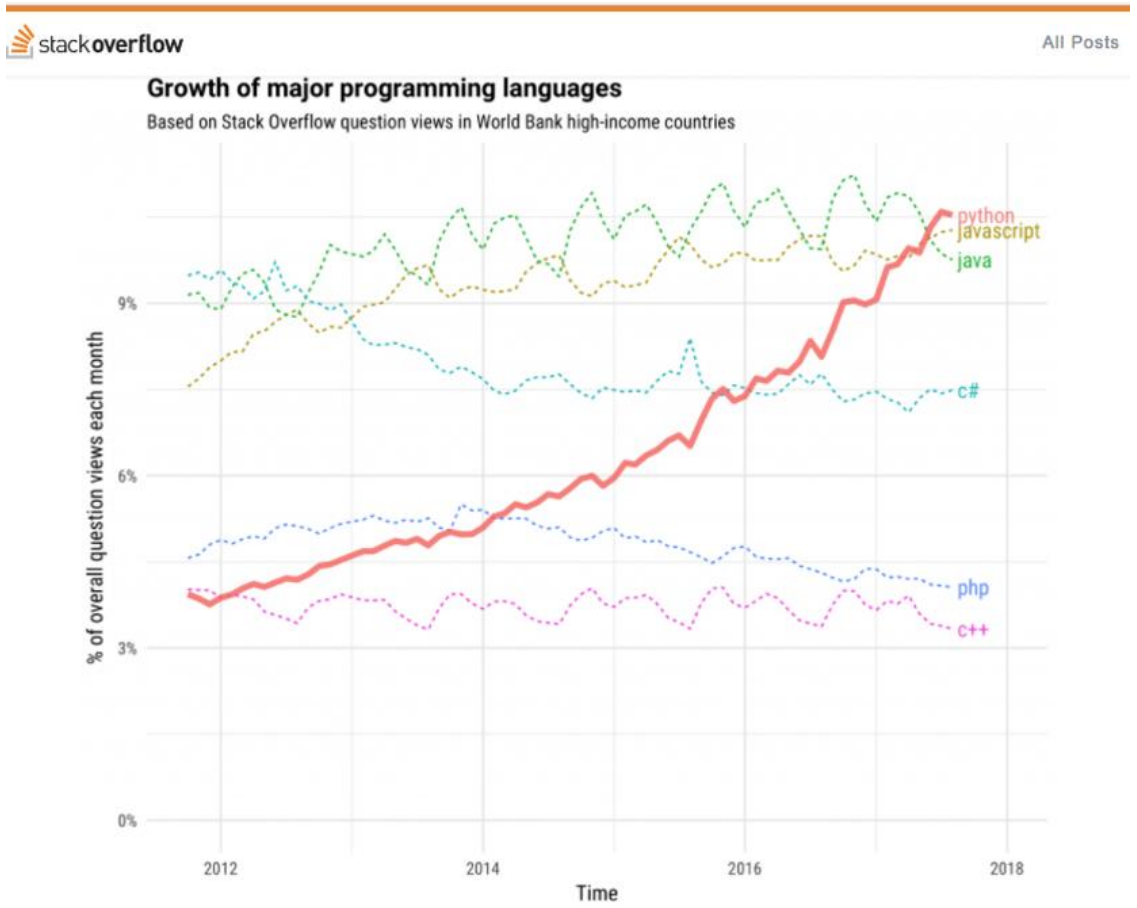
Easy to Read, Learn and Use

Hundreds of Libraries and Frameworks

Supportive Python Community

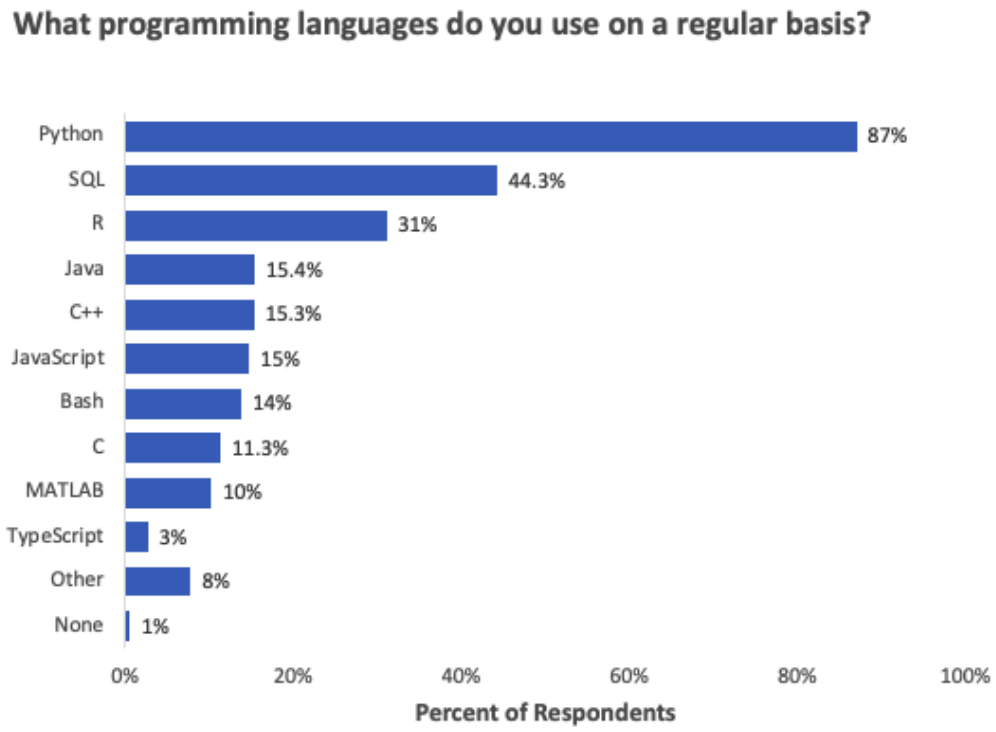
**Supported in Microsoft Excel & PowerBI**

**Big data**, **Machine Learning** and **Data Mining**




Source: Stack overflow

# Why Python?




Note: Data are from the 2019 Kaggle ML and Data Science Survey. You can learn more about the study here: https://www.kaggle.com/c/kaggle-survey-2019.
A total of 19717 respondents completed the survey; the percentages in the graph are based on a total of 14762 respondents who provided an answer to this question.

Source: https://businessoverbroadway.com/2020/06/29/usage-of-programming-languages-by-data-scientists-python-grows-while-r-weakens

pandas

scikit learn

NumPy

matplotlib

seaborn

TensorFlow

Keras

PyTorch

Transformers

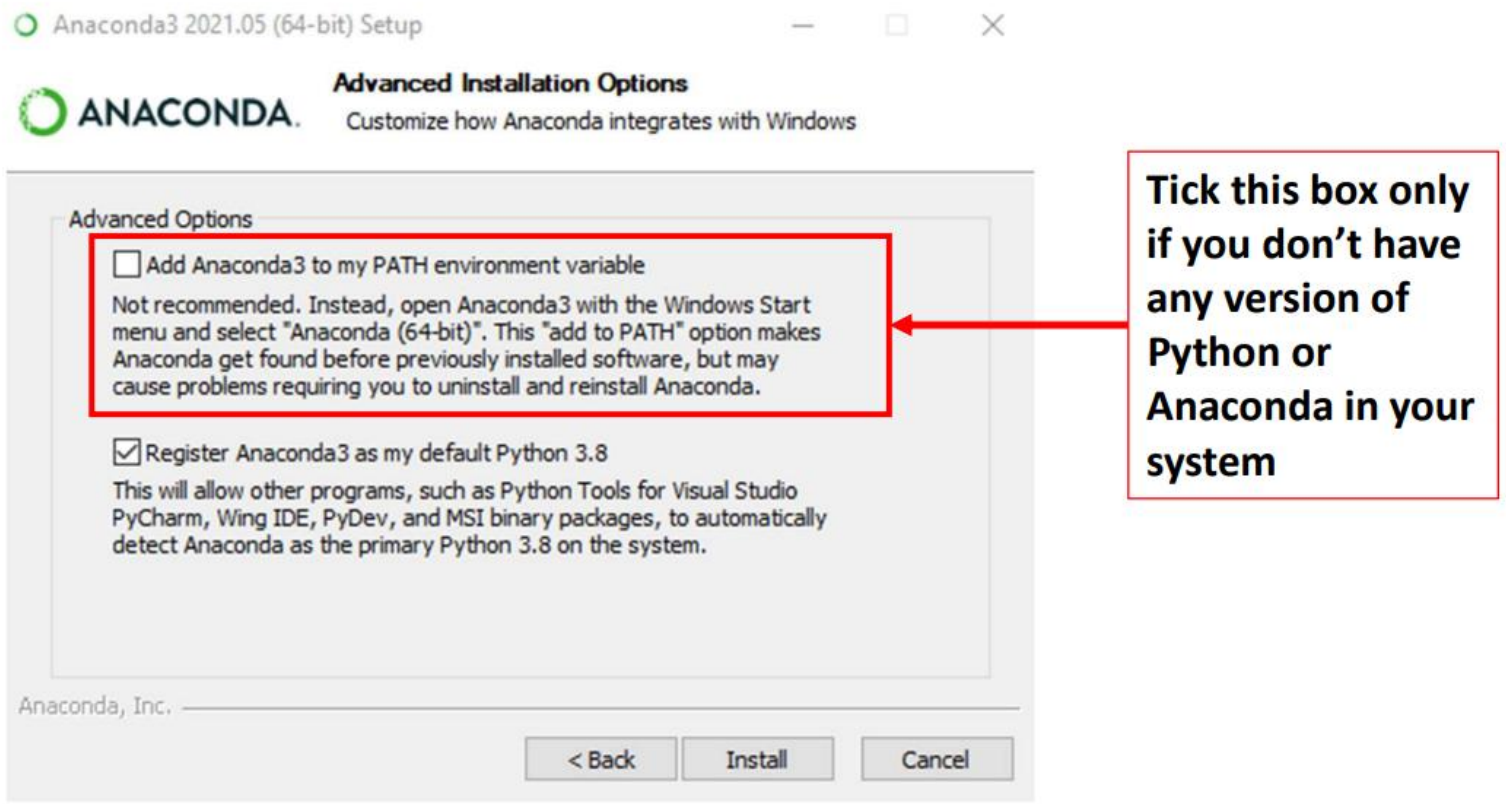
**We'll use the following Python packages often:**

- ***pandas*** – handles data analysis and manipulation (Excel on steroids)
- ***numpy*** – for intricate mathematical operations
- ***Scikit-learn*** –Machine Learning package that implements Machine Learning algorithms
- ***Matplotlib & seaborn*** – used for data visualization

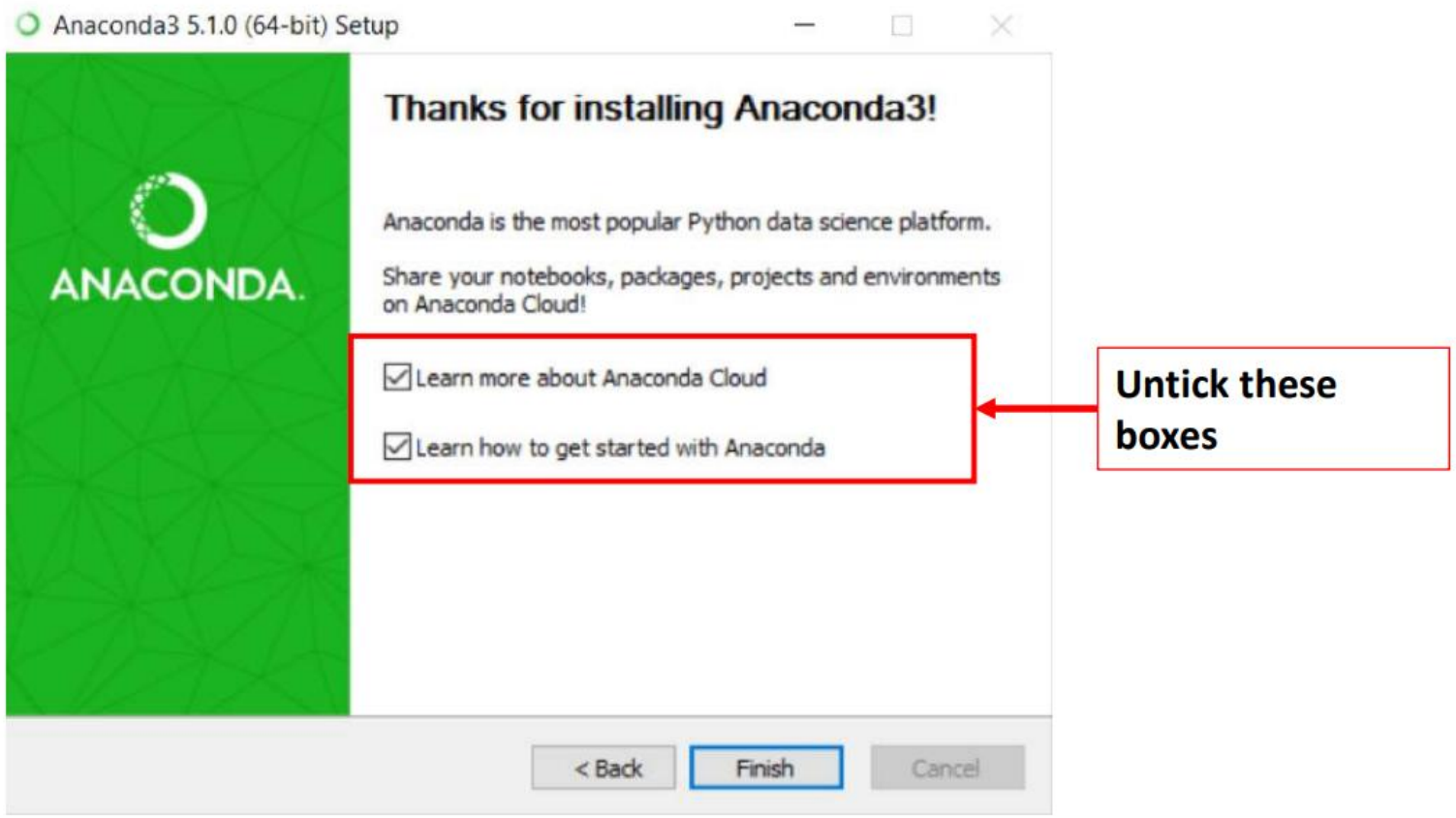
# 03.
# Setting up your environment

# 1. Installing Anaconda

1. Download and install Anaconda (https://www.anaconda.com)

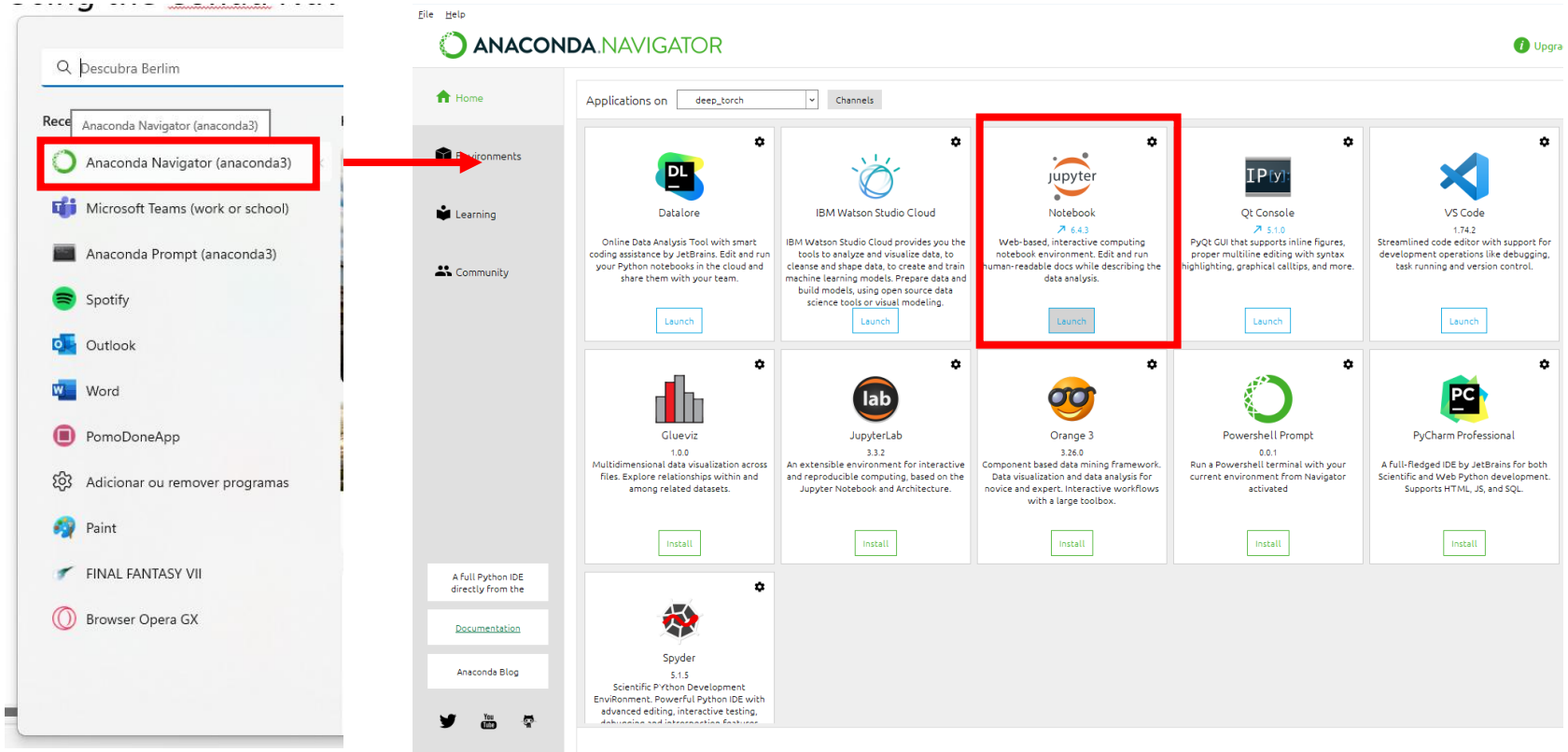


# 1. Installing Anaconda



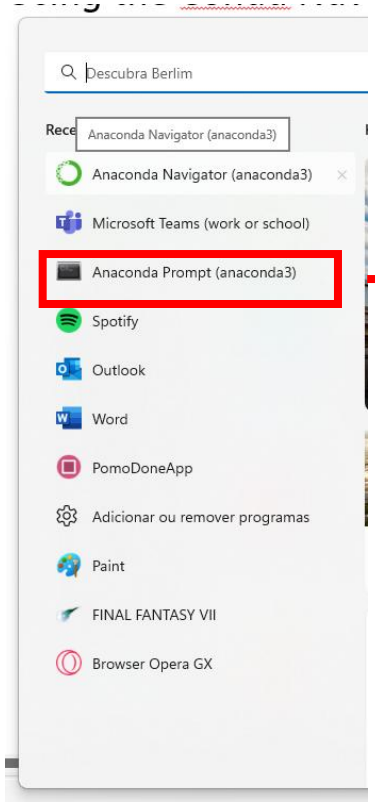
# 2. Opening Jupyter Notebook

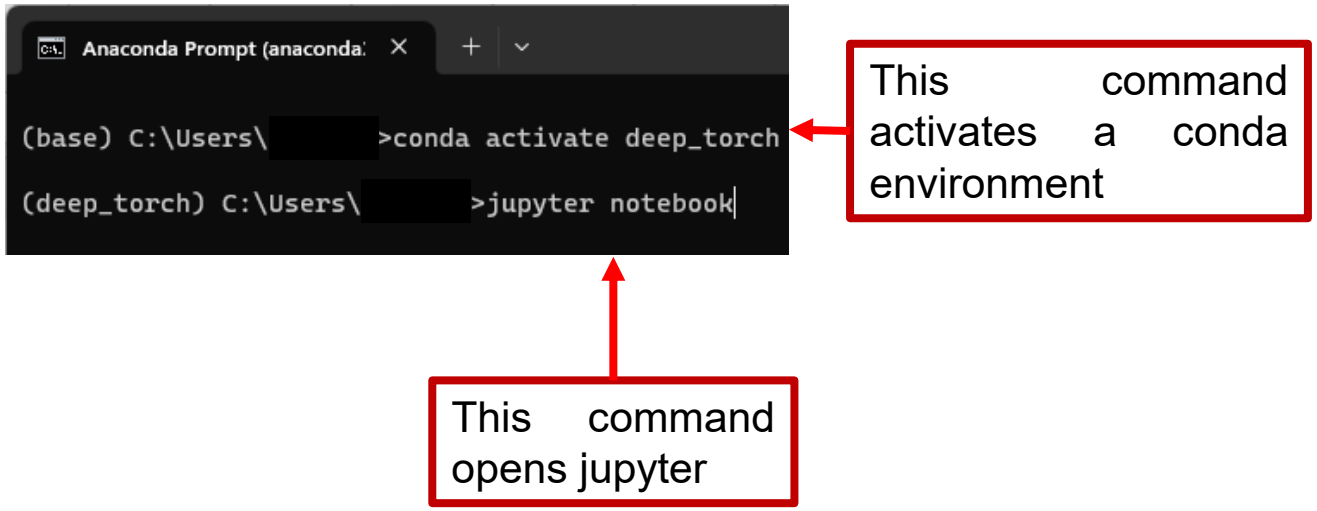
*a. Using the Anaconda Navigator*



# 2. Opening Jupyter Notebook

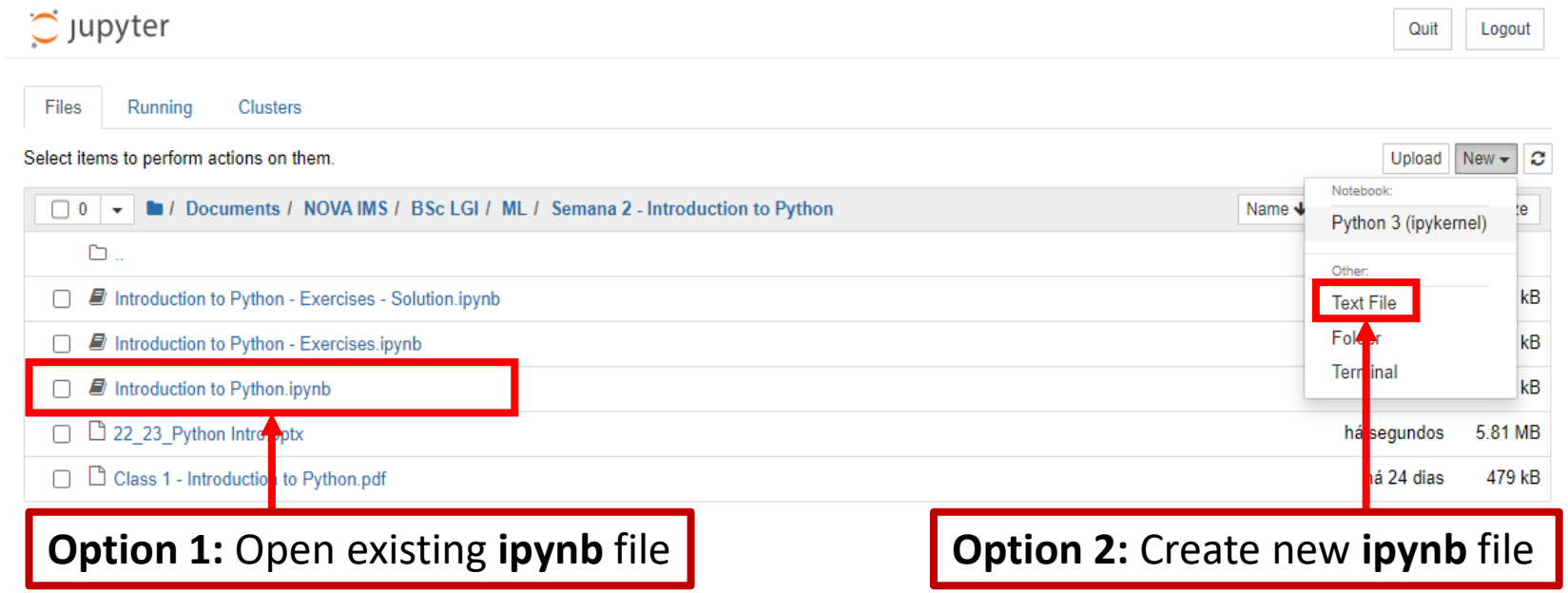
*b. Using the Anaconda Prompt*



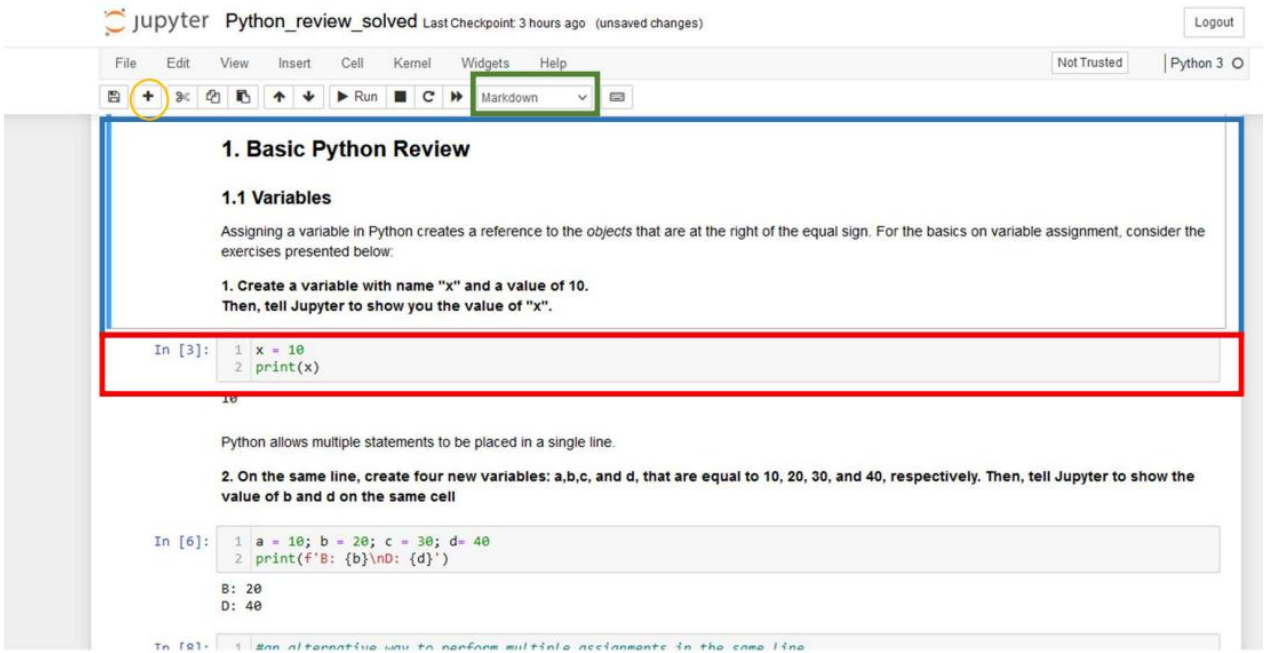




More information on setting a conda environment can be found here: https://docs.conda.io/projects/conda/en/latest/user-guide/getting-started.html#managing-environments

# 2. Opening Jupyter Notebook



# 2. Opening Jupyter Notebook




- Markdown cell
- Code cell
- Adds one cell below
- Switches between modes (Markdown/Code)

# 04
# Initial Exercise