UNIVERSITY OF PORTSMOUTH

Python for Data Analysis Introduction (Week 1)

Atefeh Khazaei

atefeh.khazaei@port.ac.uk

What is this unit about?

- The aim of this unit is to improve your skills on some data analysis operations using Python programming.

1. Manipulating, pre-processing, cleaning datasets
2. Applying advanced libraries
3. Visualising the dataset
4. Creating statistics about dataset

What you will get after this unit?

- You will able to analyse any given dataset using advanced python packages
- Also, you will able to produce meaningful statistics for deep understanding of your data

How we will do our lectures?

- First hour will be theoretical knowledge; I will give you some background and examples of the methods and functions.
- Second hour will be the practical hour; I will give you some example codes and datasets and you can apply what you learn from the first hour on the datasets.

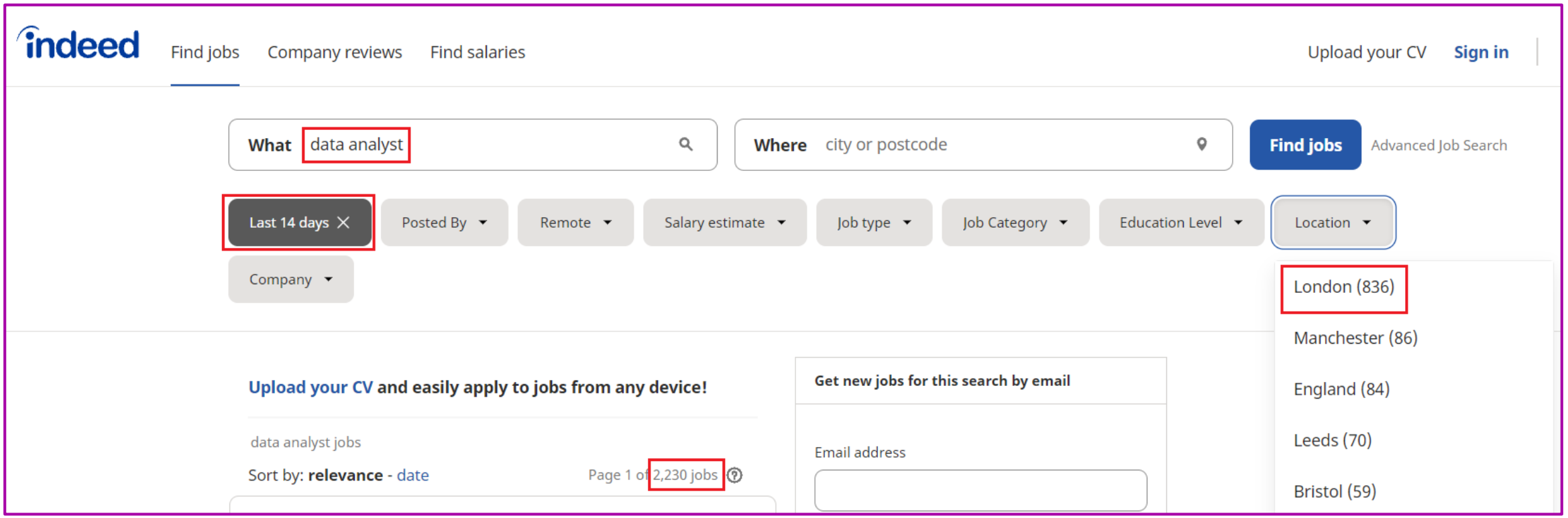
Data Analysis

- Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information and knowledge, and supporting decision-making.
- Data analysis has multiple approaches, diverse techniques (under a variety of names), and is used in different business, science, and social science domains.
- In today's world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

Why Data Analysis?

- The Explosive Growth of Data
  - Data collection and data availability
    - Automated data collection tools, database systems, Web, computerized society
- We are living in the ~~information~~ **DATA** age
- The world is data rich but information and knowledge poor.
- Data is power!!

Data Analysis Job Vacancies



https://uk.indeed.com/

Data Analysis Job Vacancies (cont.)

Top Growing Careers in the U.S.

Occupation	2019 Median Pay (Yearly)	Growth Rate (2019-2029)
1. Wind turbine service technicians	$52,910	61%
2. Nurse practitioners	$109,820	52%
3. Solar photovoltaic installers	$44,890	51%
4. Statisticians	$91,160	35%
5. Occupational therapy assistants	$61,510	35%
6. Home health & personal care aides	$25,280	34%
7. Physical therapist assistants	$58,790	33%
8. Medical & health services managers	$100,980	32%
9. Physician assistants	$112,260	31%
10. Information security analysts	$99,730	31%
11. Data scientists & mathematical science occupations	$94,280	31%

Howmuch.net

Python Programming

- Python is a high-level general-purpose programming language.
- Its design emphasizes code readability with its use of significant indentation.
- In recent years new Python libraries are introduced with many useful functions and methods
  - Numpy
  - Pandas
  - scikit-learn
- They made attractive python for data analysis

Outcome of First Week

- Preparing coding environment
- Understanding basics of Python programming

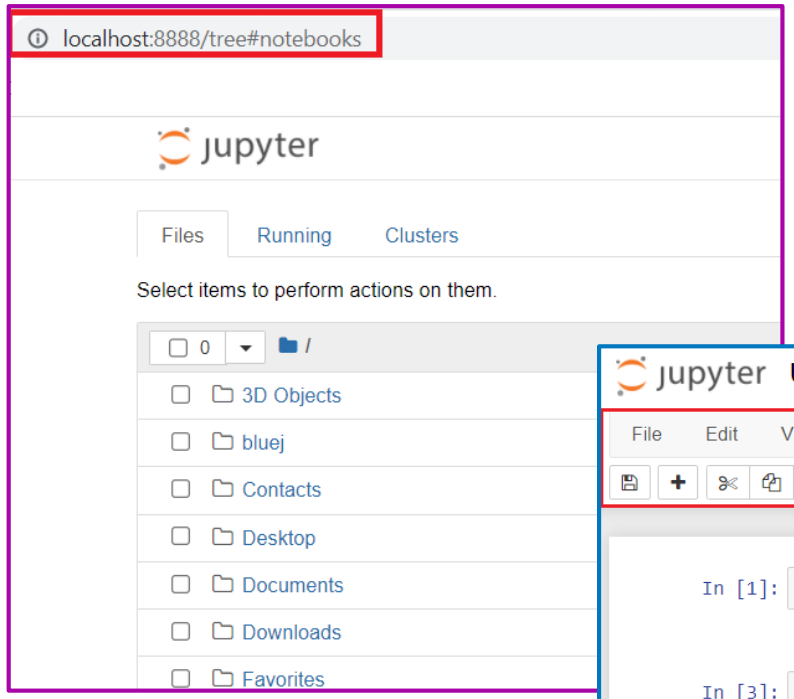
Installation and Setup

Anaconda

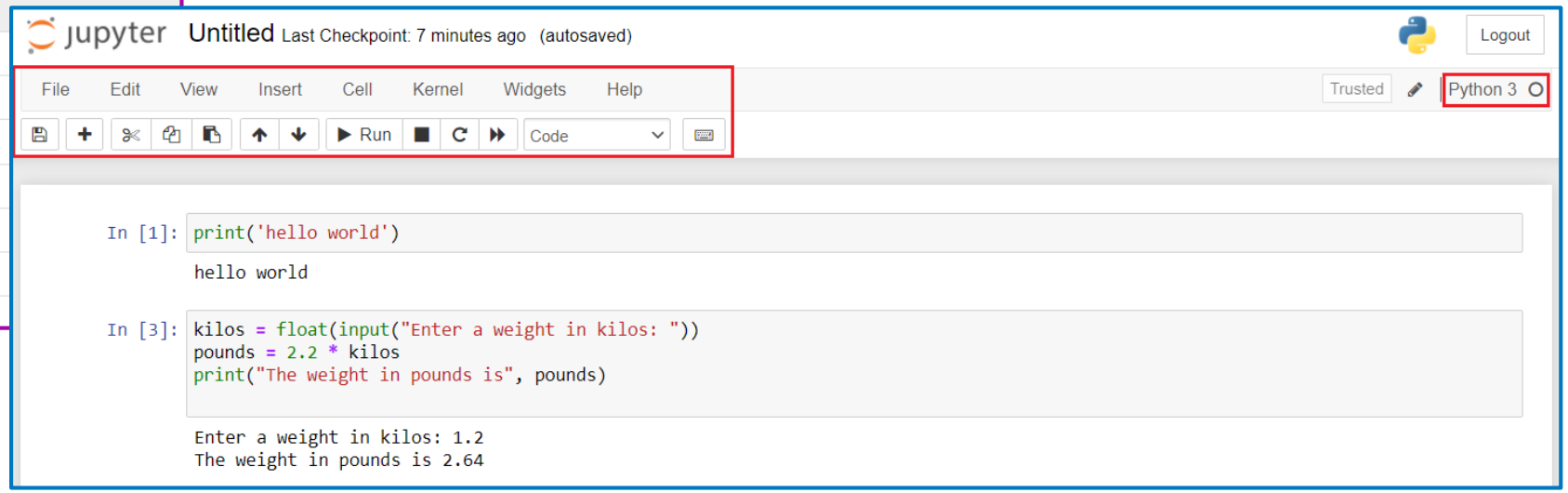
- Anaconda development environment
  - Anaconda is a distribution of the Python and R programming languages for scientific computing that aims to simplify package management and deployment.
- For all operating system users please check:
  - https://www.anaconda.com/products/individual
- Python 3. versions will be used.

Installation and Setup (cont.)

Jupyter



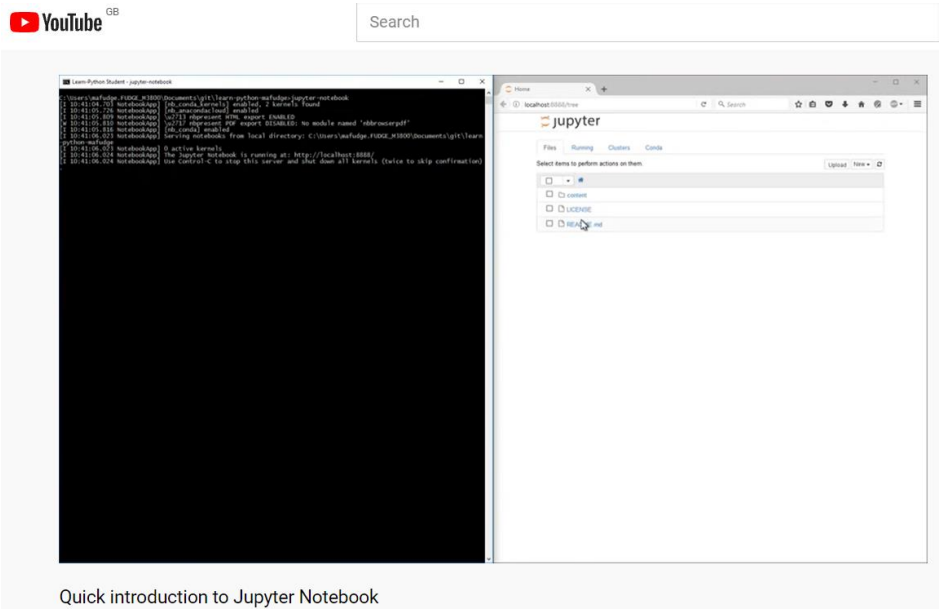
- You can start the notebook from command line
  jupyter notebook
- Jupyter notebook is opened in your browser on default port 8888.
- You may also specify a port manually.
  jupyter notebook --port 9999



Installation and Setup (cont.)

Jupyter

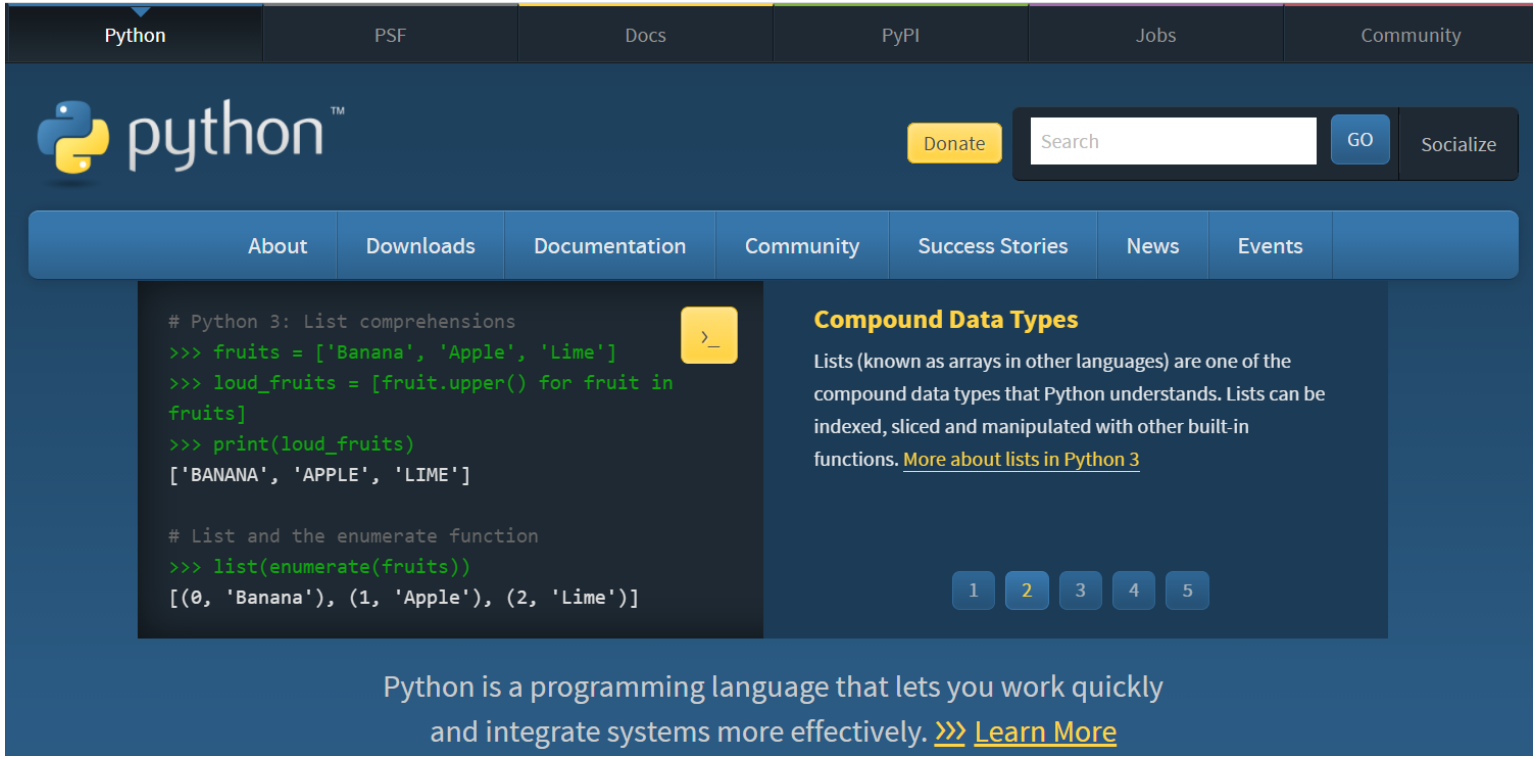
- Watch this video as an introduction to Jupyter
  - Quick introduction to Jupyter Notebook
  - https://www.youtube.com/watch?v=jZ952vChhuI



Installation and Setup (cont.)

Other options

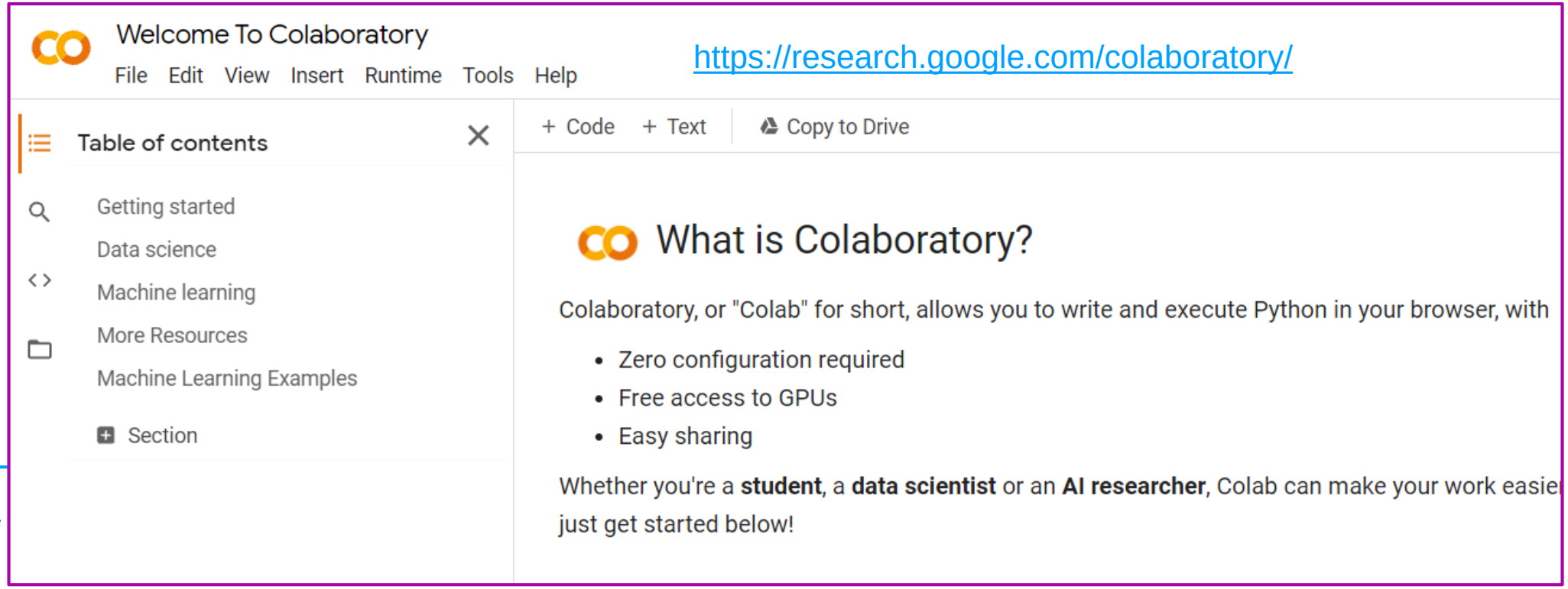
- The official home of the Python Programming Language: https://www.python.org/



Installation and Setup (cont.)

Other options

- Colaboratory, or “Colab” for short, is a product from Google Research.
- Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.



Python IDE Options

- An IDE (Integrated Development Environment) is a software application used by developers for creating programs.
- IDEs are meant to make the developer's job easier by combining tools that are necessary during software development.
  1. PyCharm
  2. Visual Studio Code
  3. Jupyter Notebook
  4. Spyder

Testing & Basics

- You can put this command after you open your Jupyter notebook to see if everything is working

```
print('Hi World')
```

Testing & Basics (cont.)

- Write a weight converter program that
  - Transforms a weight measured in kilos (kilograms) into an equivalent pounds.

- In other words:
  - User input: a weight measured in kilos
  - Output to screen: a weight measured in pounds, equivalent the input

Testing & Basics (cont.)

- The following equation relates kilograms and pounds:
  - pounds = 2.2 × kilos
    - 1 kilo = 2.2 pounds
    - 10 kilos = 22 pounds

Testing & Basics (cont.)

- A reasonable algorithm for our task, written in English, is:

1. Obtain a kilos value from the user
2. Calculate a pounds value using = 2.2 × kilos
3. Output the pounds value to screen

Testing & Basics (cont.)

- Implementation of weight converter in python can be like:

```
kilos = float(input("Enter a weight in kilos: "))
pounds = 2.2 * kilos
print("The weight in pounds is", pounds)
```

Program concepts

Statements & Variables

- Each line of the program is called a **command** or **statement**.
- A **variable** denotes a part of computer memory where value is stored.
- Variables have names in the program
- A statement in the program may:
  - Create a new variable
  - Use the value of a variable
  - Change the value of a variable

```
kilos = float(input("Enter a weight in kilos: "))
pounds = 2.2 * kilos
print("The weight in pounds is", pounds)
```

Program concepts (cont.)

Assignment statements

- An assignment statement is used to assign a value variable.
- Assignment statements are executed in two steps:
  1. Evaluate the expression on right hand side (i.e. find out its value)
  2. Assign this value to the variable on left hand side.
- If the variable on left hand side doesn't yet exist, then it is created.
- Otherwise (if the variable already exists), its old value is replaced.

```
kilos = float(input("Enter a weight in kilos: "))
pounds = 2.2 * kilos
print("The weight in pounds is", pounds)
```

Program concepts (cont.)

Numeric & String Values

```
kilos = float(input("Enter a weight in kilos: "))
pounds = 2.2 * kilos
print("The weight in pounds is", pounds)
```

- 2.2 is numeric
- "Enter a weight in kilos: “ is string
- In string values, we can use double or single quotes, but can’t mix them;
  - So: "hello" and ’hello’ are OK,
  - Whereas: "hello’ is not!

Program concepts (cont.)

Arithmetic Expressions

- Python allows standard arithmetic expressions to be formed from +, -, * (multiplication), /, and brackets (and).
- An expression is evaluated to give a value.
- We will see more arithmetic operations in next lecture

Program concepts (cont.)

Built-in Functions

- A built-in function is an algorithm that part of the Python language, and can be accessed by using its name.
- The input built-in function:
  - Displays a prompt on the screen;
  - Waits for the user to enter a value;
  - Gives us the value that user entered.

```
kilos = float(input("Enter a weight in kilos: "))
pounds = 2.2 * kilos
print("The weight in pounds is", pounds)
```

- The print built-in function displays information on the screen.

Practical Session

- Install required tools on your laptop
  - Anaconda
  - Python 3
  - Jupyter
- Run our simple sample programs