CMM201 - Programming Concepts for Business Analytics

- Dr. Carlos Moreno-Garcia
- Lecturer in Computing
- Placements and Electives Coordinator
- School of Computing Science and Digital Media

Aim of the Module

This module will introduce students to fundamental programming principles and concepts within the context of creating solutions for business analytics.

Learning Outcomes

On completion of this module, students are expected to be able to:

1. Critically appraise a range of programming languages and tools commonly used for data analytics (in this case, we will resort to the Python programming language).
2. Demonstrate a critical understanding of core programming techniques and concepts.
3. Use existing libraries and coding techniques to perform data management, data analysis and data visualization tasks.
4. Apply programming skills to business decision making problems.

Module Evaluation

Two courseworks outputs, each corresponding to **50%** of the final mark:

- First output: **7th November, 2019**
- Second output: **12th December, 2019**

The guidelines and submission instructions can be found here (http://campusmoodle.rgu.ac.uk/course/view.php?id=96406).

Industry Partnership/Engagement

Students are intended to learn basic programming knowledge and thus, the engagement with industry is expected to be developed in the follow-up module “CMM202 – Programming for Business Analytics”.

Bibliography

Books

- Python. Toby Donaldson, Peachpit Press, 2013.
- Python Essentials. Steven F. Lott, Packt Publishing Ltd, 2015.
- Think Python: How to Think Like a Computer Scientist. Allen Downey, O'Reilly Media, Inc., 2012.
- Fluent Python. Luciano Ramalho. O'Reilly Media, Inc., 2015.
- Python Cookbook: Recipes for Mastering Python 3. David Beazley, Brian K. Jones, O'Reilly Media, Inc., 10 May 2013.
- Introduction to Computing and Programming in Python: A Multimedia Approach. Mark Guzdial, Barbara Ericson. Pearson, 2016.
- Search the RGU Library in here (https://librarysearch.rgu.ac.uk/discovery/search?query=any,contains,python&tab=Everything&search_scope=MyInst_and_CI&vid=44RGU_INST:VU1&offset=0).

Podcasts

- https://talkpython.fm/ (https://talkpython.fm/)
- https://www.pythonpodcast.com/ (https://www.pythonpodcast.com/)
- https://testandcode.com/ (https://testandcode.com/)

Websites

- Python Documentation (https://docs.python.org/3/)
- StackOverflow (https://stackoverflow.com/questions/tagged/python)
  - Online community where people exchange ideas, doubts and code.
- Github (https://github.com/python) code repository.

Online Courses

- Datacamp (https://www.datacamp.com/courses/intro-to-python-for-data-science?utm_s
  utm_campaignid=805200711&utm_adgroupid=43370829484&utm_device=c&utm_key
  utm_matchtype=b&utm_network=g&utm_adpostion=1t1&utm_creative=19110499911
  414126611260&utm_loc_interest_ms=&utm_loc_physical_ms=9046834&
  gclid=CjwKCAjw7_rlBRBaEiwAc23rhjfi2Mw2qxQ1_zJmVEph39YX5t6HkFUTmp48oqIp
  (you can also download the mobile app to practice on the go)
- CodeInstitute (www.codeinstitute.net)
- LearnPython (https://www.learnpython.org/) (free course)
- EDX (https://www.edx.org/learn/python)
- Coursera (https://www.coursera.org/courses?query=python)
- CodeAcademy (https://www.codecademy.com/learn/learn-python)
- Udemy (https://www.udemy.com/python-for-finance-investment-fundamentals-data-ana
  Investment Fundamentals & Data Analytics.

Events in the City

- Aberdeen Data Meetup (1st Tuesday of each month).
  - Organised by Scotland Data Science & Technology Meetup group.
  - Attendance managed through Meetup (https://www.meetup.com/Scotland-Data-Science-Technology-Meetup/).
  - Discussion about latest news, projects and needs of the city, all disciplines welcome.
- Python Aberdeen Group (2nd Wednesday of each month).
  - Organised by Code the City Aberdeen (https://codethecity.org/).
    - They organise other events as well such as hackathons and workshops.
  - Attendance managed through Tito (https://ti.to/code-the-city/).

Why do you Need Python?

Source (https://www.datacamp.com/community/blog/why-your-company-needs-python-for-business-analytics)

Improves Work for Everyone

- Widely used top programming language
- Huge growing ecosystem due to its open source nature
- Almost every industry is on board

Is Replacing Excel

- Excel isn't scalable for modern business needs
- Allows collaboration

Descriptive Analytics and Dashboards

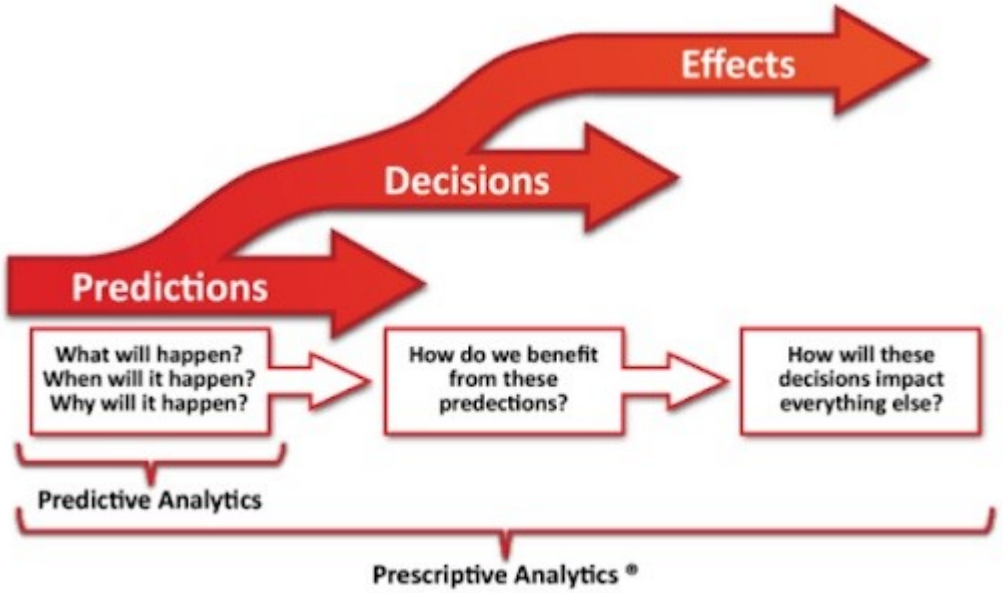
- Exploratory Data Analysis
- Manipulation of data
- Streamline work flows
- Creating visualisations

Machine Learning

- Predicting and classifying new data
- Recommender systems
- Can work with popular Google machine learning libraries (such as Tesseract and Tensorflow)

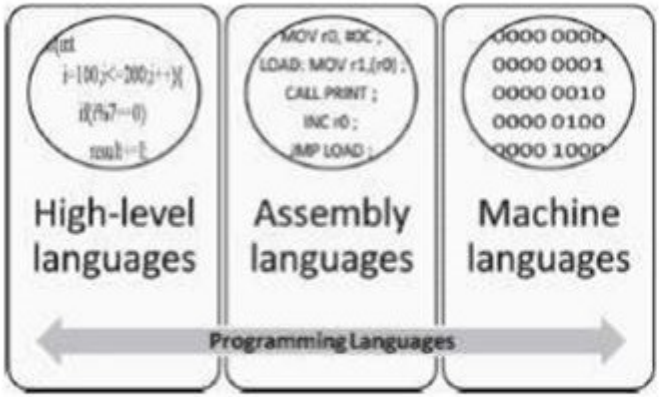
Predictive/Prescriptive Analytics

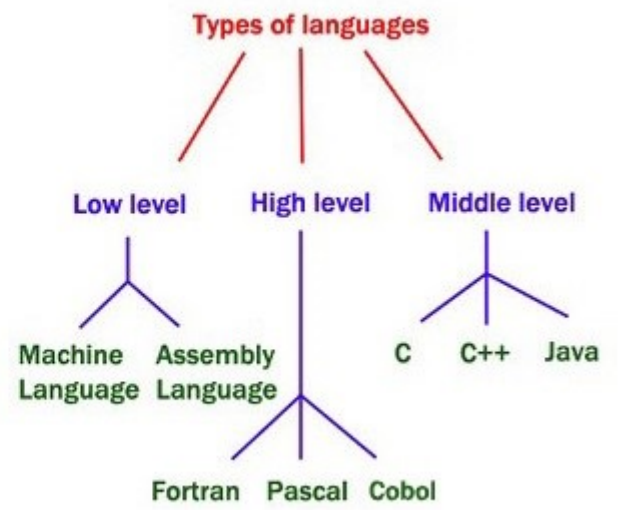
- Decision science
  - Anticipate what, when and why certain outcome will happen
  - What to do with information
- Deep learning to optimise outcomes



Fundamentals of Programming

Types of Programming Languages





Source 1 (http://4.bp.blogspot.com/-NvijJmjC13I/TmIbqlKKl8I/AAAAAAAAA3Q/mK4Nmy43en8/s1600/Untitled-1+%25281%2529.jpg) Source 2 (https://studyin24.com/wp-content/uploads/2018/12/Programming-language-types.jpg)

Advantages of High-level Programming Languages

- Programmer friendly.
- Easy to write, debug and maintain.
- Provide higher level of abstraction from machine languages.
- Machine independent language.
- Easy to learn.
- Less error prone.
- Results in better programming productivity.

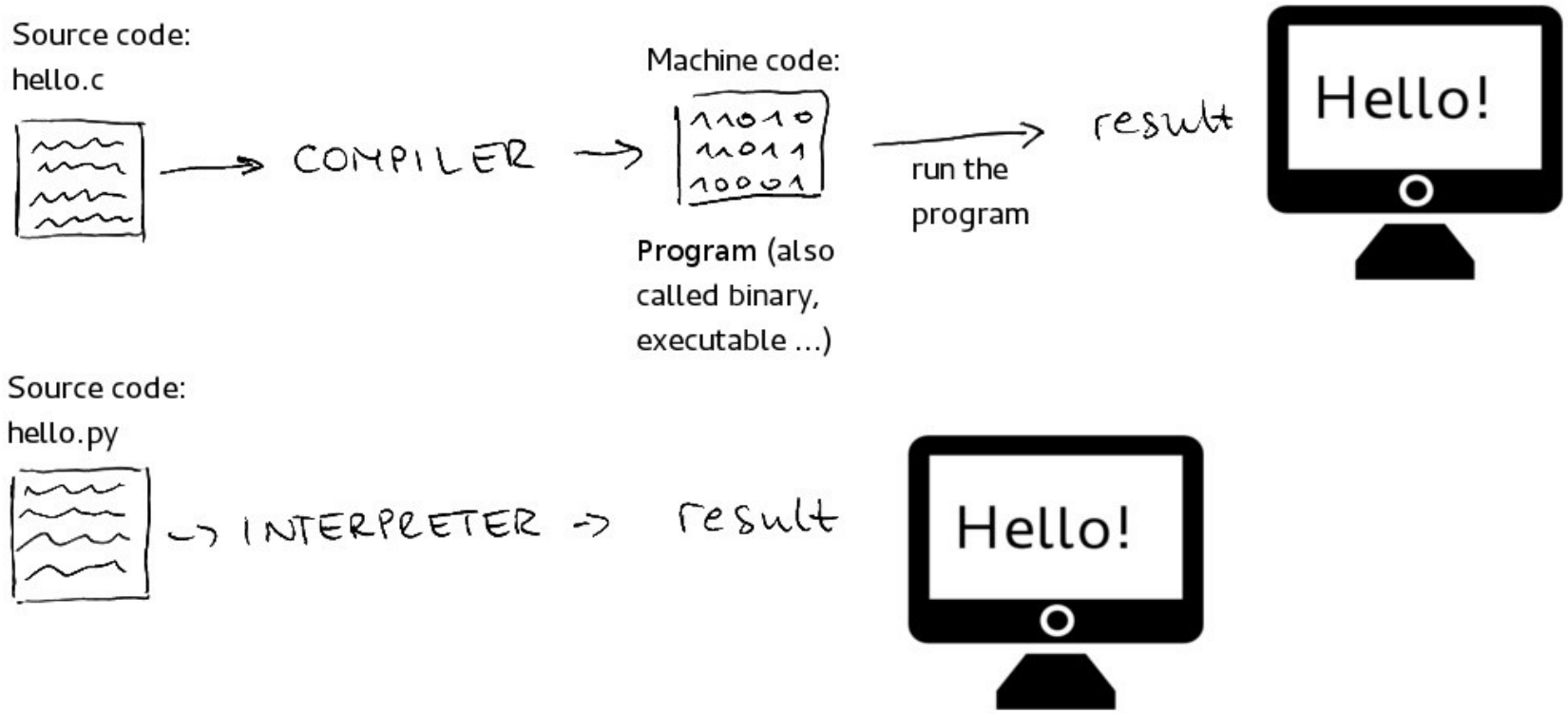
Compiled vs Interpreted Programming Languages

Compiled

- The high-level source code is translated to machine code using a compiler.
- Example: An addition + gets directly translated to the ADD instruction in the machine code.
- Examples: C, Fortran, COBOL, C++, and Java (compiled to bytecode).
- Advantages:
  - Ready to run.
  - Often faster.
  - Source code is kept private.

Interpreted

- Instructions are not directly executed, but read by another program.
- Instructions run freely without the need to compile them first!
- Examples: JavaScript, Perl, R, *Python*.
- Advantages:
  - Cross-platform (portability).
  - Simpler to test.
  - Display error as each instruction is run.



Source (https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwichoGF1KXkAhVOdhoKHebmAJwQjRx6BAgBEAQ&url=%2Furl%3Fsa%3Di%26rct%3Dj%26q%3D%26esrc%3Ds%26source%3Dimages%26cd%3D%26ved%3D%26url%3Dhttps%253A%252F%252Fmedium.com%252Ffrom-the-scratch%252Fstop-it-there-are-no-compiled-and-interpreted-languages-512f84756664%26psig%3DAOvVaw0CqS9Nmdo4wbc9J-p4WtL-%26ust%3D1567083827896505&psig=AOvVaw0CqS9Nmdo4wbc9J-p4WtL-&ust=1567083827896505)

Static vs Dynamic Programming Languages

- Static is designed to optimise *hardware* efficiency
- Dynamic is designed to optimise *programming* efficiency so that less code is used.
- In fact, dynamic languages are written using a static one.
    - Python is written in C!

Python

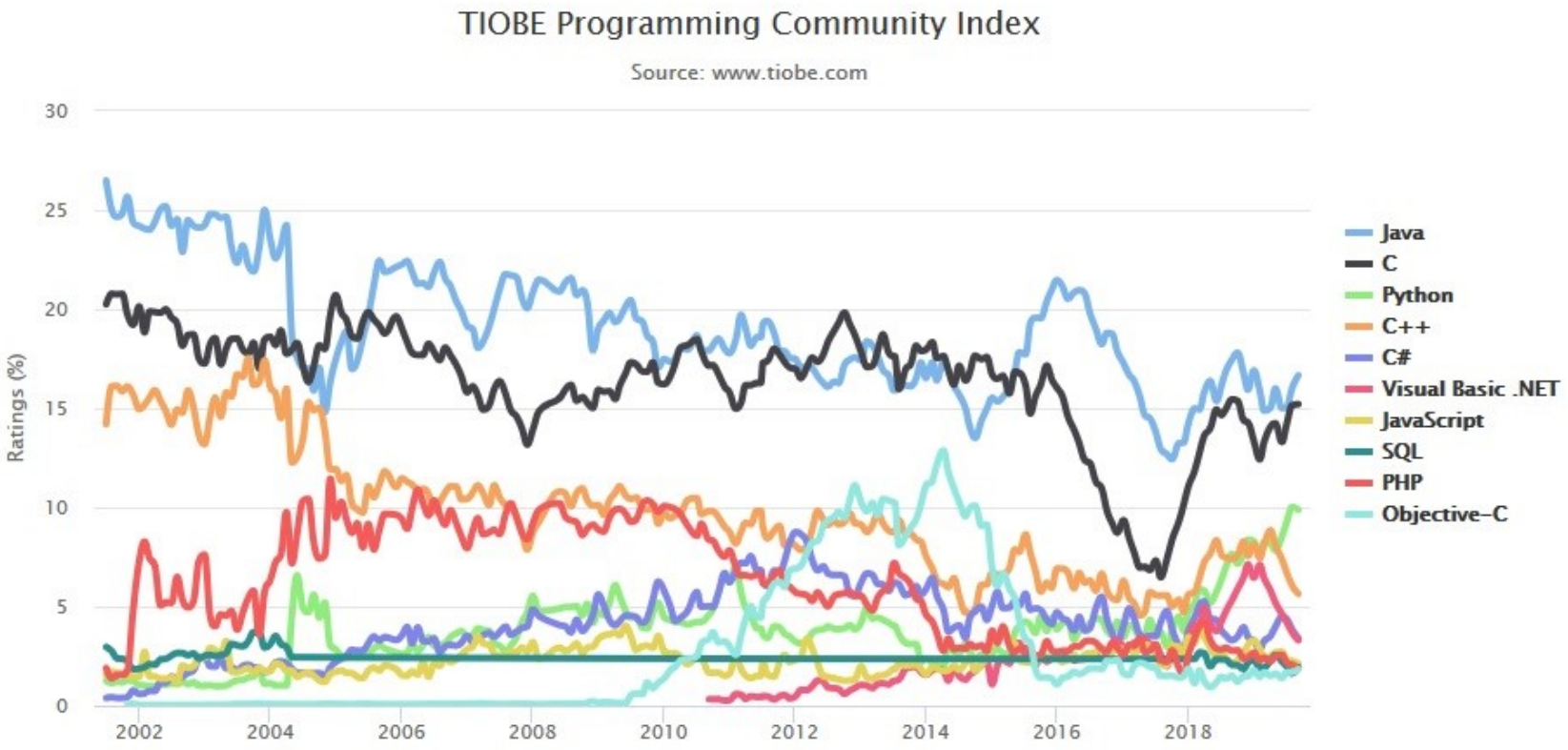
What is Python?

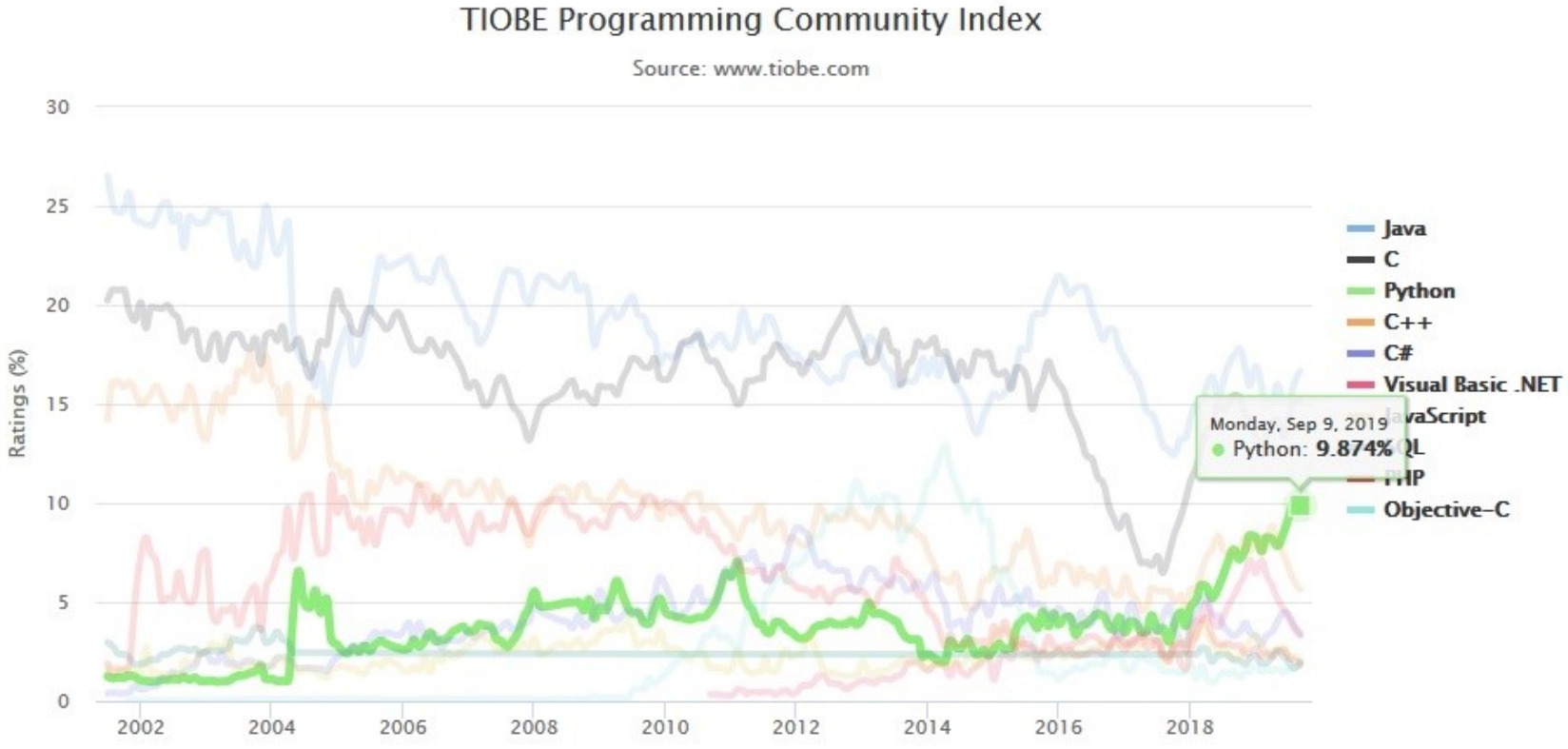
- Widely used **high-level**, **interpreted**, **dynamic** programming language.
- Emphasizes code readability.
- Its syntax allows programmers to express concepts in fewer lines of code.
- Similar to R and Matlab.

Some statistics

Popularity

Python is the third most popular programming language according to the TIOBE (https://www.tiobe.com/tiobe-index/) index, being the fastest growing one in this rubric for the current year.



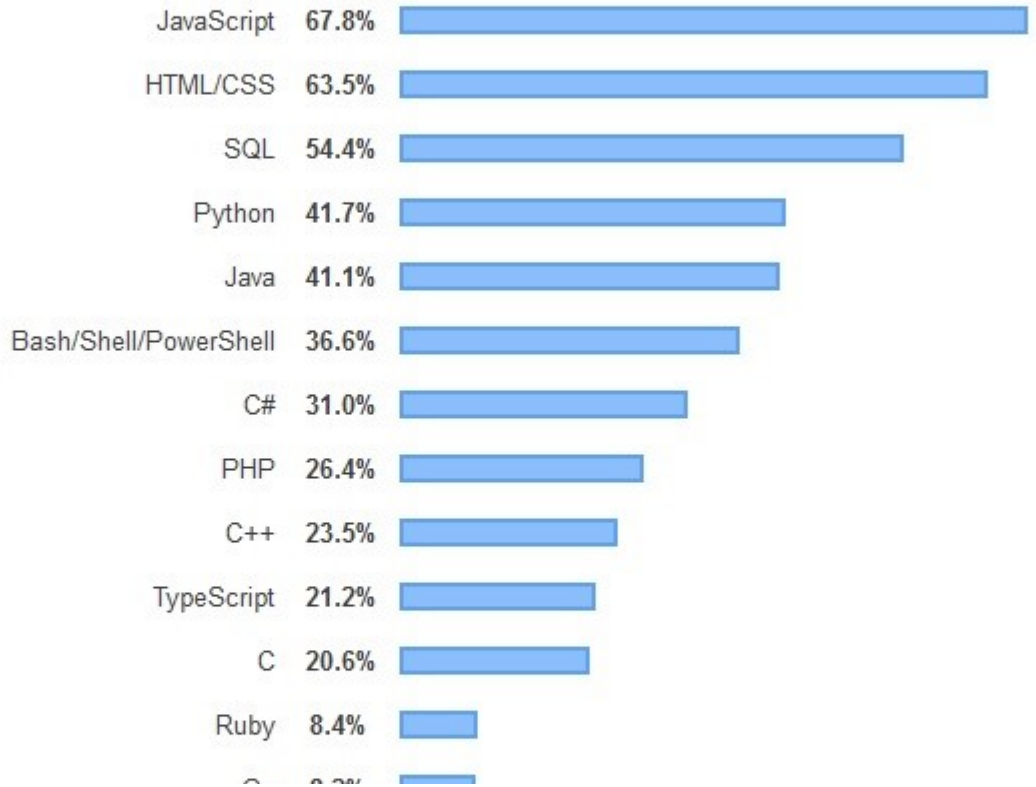

TIOBE Programming Community Index
Source: www.tiobe.com
Ratings (%)
30
25
20
15
10
5
0
2002
2004
2006
2008
2010
2012
2014
2016
2018
Java
C
Python
C++
C#
Visual Basic .NET
JavaScript
SQL
PHP
Objective-C
Monday, Sep 9, 2019
Python: 9.874%

According to the 2019 developer survey run by Stack overflow (https://insights.stackoverflow.com/survey/2019), Python is the 4th most popular programming language in the world, both for general public and for professional developers.

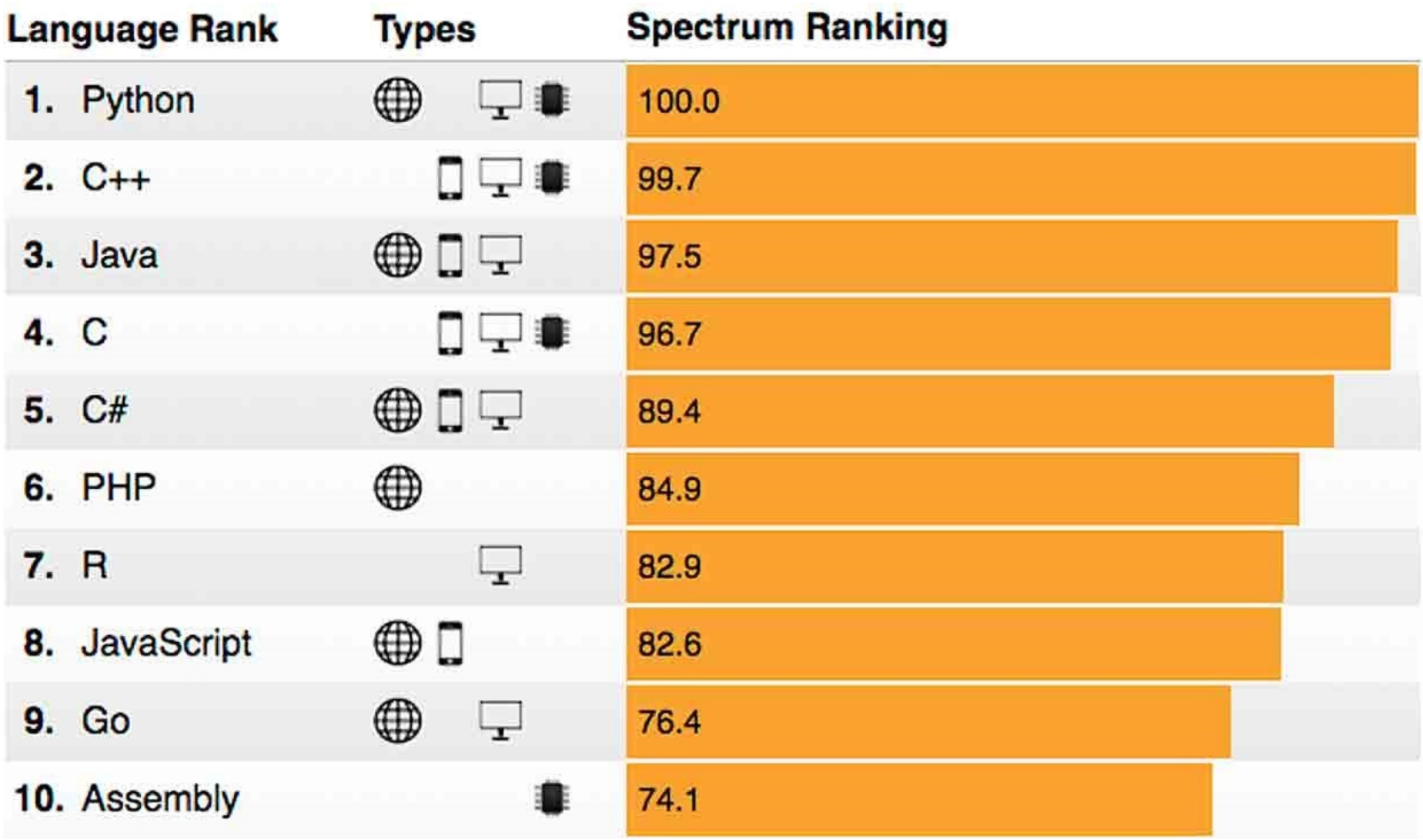
Most Popular Technologies

Programming, Scripting, and Markup Languages

All Respondents | Professional Developers



Python is currently the best ranked programming language according to the Institute of Electrical and Electronics Engineers (IEEE) (https://spectrum.ieee.org/at-work/innovation/the-2018-top-programming-languages).



Employabilty

Python is currently the language with the fastest growing rate of interest by employers according to Google Trends (https://medium.freecodecamp.org/best-programming-languages-to-learn-in-2018-ultimate-guide-bfc93e615b35).

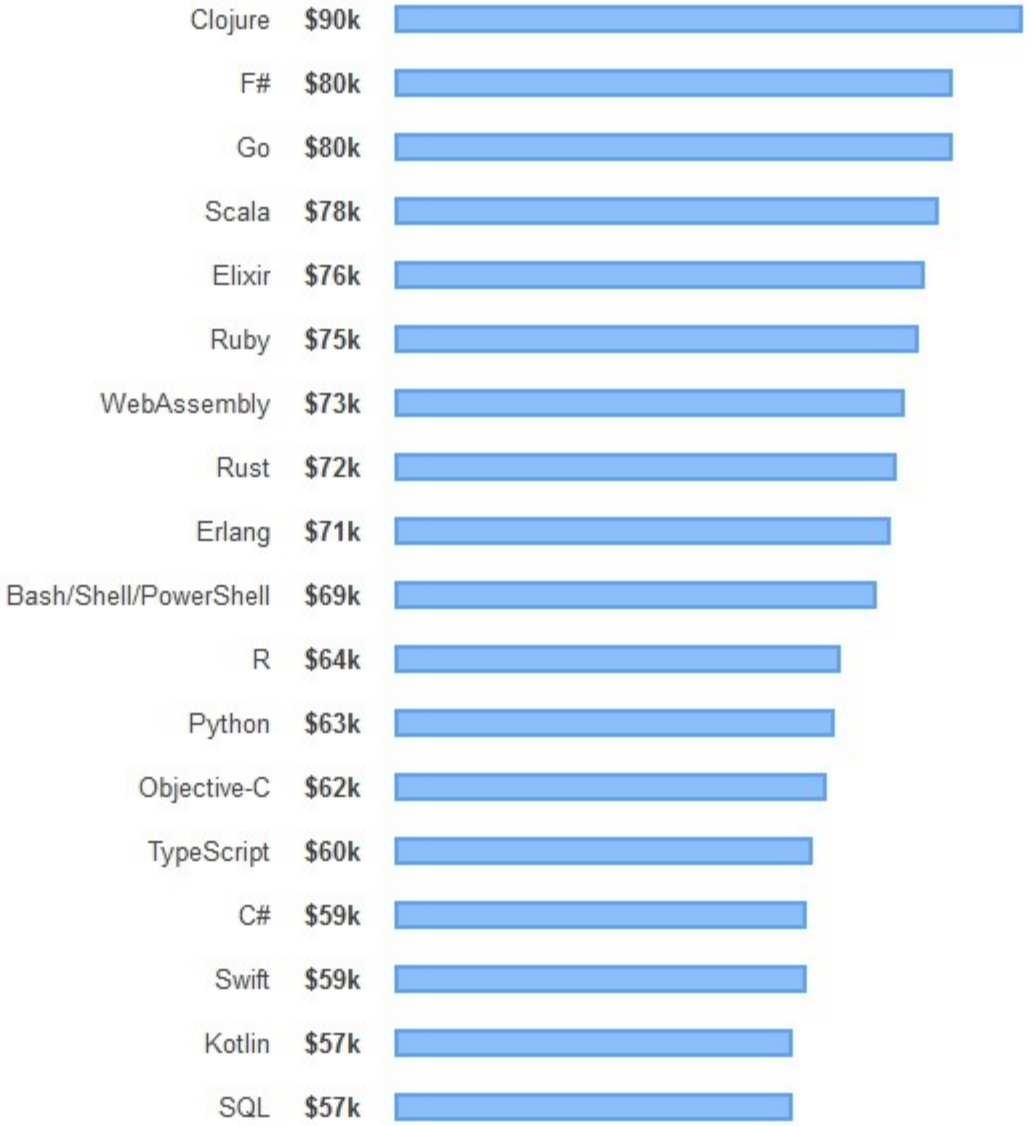
Fig. 8. Google trends interest over time

It is the 12th best paid language, but one of the fastest to adopt.

Top Paying Technologies

What Languages Are Associated with the Highest Salaries Worldwide?

Global | United States



SALARY BY LANGUAGE

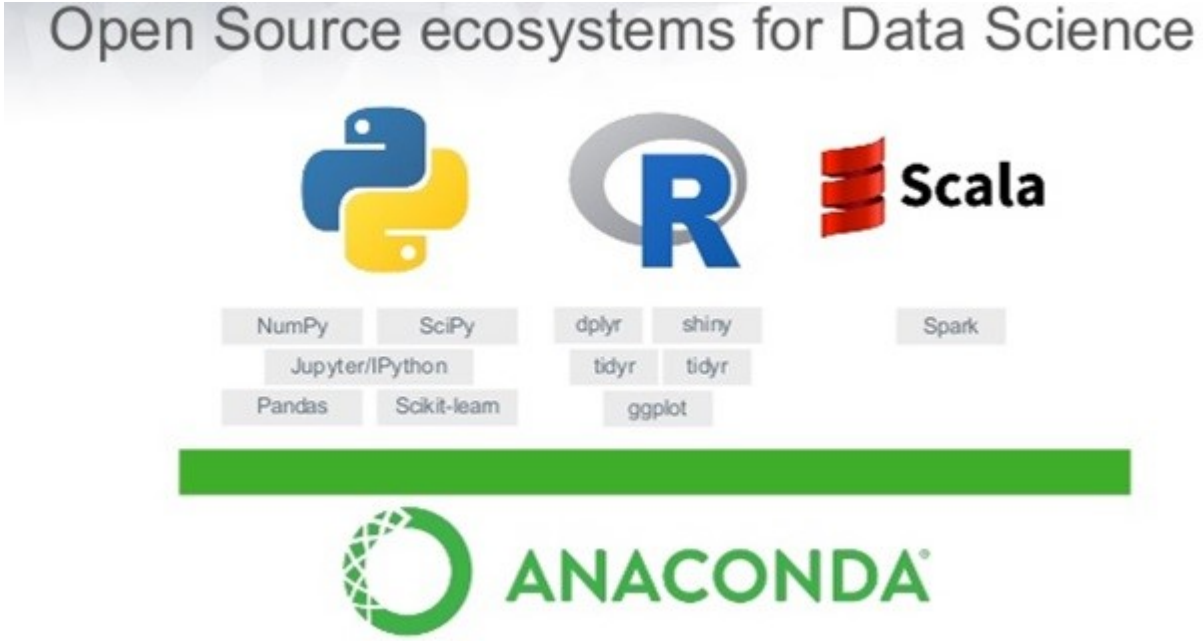
Installing Python

The long and hard way

1. Install Python (https://www.python.org (https://www.python.org)).
2. Install a Python Integrated Development Environment (IDE) such as IDLE (available when installing Python), Pycharm (https://www.jetbrains.com/pycharm/ (https://www.jetbrains.com/pycharm/)) or Spyder (https://pypi.org/project/spyder/ (https://pypi.org/project/spyder/)).
3. Install Jupytor Notebook (http://jupyter.org/ (http://jupyter.org/)).

The fast way: Anaconda Navigator

- Everything can be easily installed using a bundle called Anaconda Navigator (https://www.anaconda.com/download/).



How Does Python Look Like?

In its most simplistic state, Python acts like a calculator. You simply write one calculation, and Python gives you the answer!

In [ ]:
```
1+1
```

Moreover, you can also do some coding!

In [ ]:
```
x = 2
y = -1
z = x + y
print(z)
```

Notice the simplicity of the Python syntax in the sense that we do not need to define classes or use a complex and strict structure of parenthesis!

What else can I do in Python?

- Python is widely used in **data science**, as it contains a long list of packages that allow importing all kinds of data (i.e. images, sound files, video, spreadsheets, etc.)

- Once data has been imported, you can do some data pre-processing:
  - Visualising data of interest
  - Subsectioning rows/columns
  - Augmenting data artificially

- Furthermore, you can perform data analysis and statistics to:
  - Understand previous and new trends
  - Predict values of incoming new data
  - Cluster data

BONUS: In fact, this slideshow was done using one of the numerous Python tools that we have at hand!

- I used the *Jupyter Notebook* integrated development environment (IDE) with an extension found online called *Rise*.