UNIVERSIDAD EAFIT

# Científico de Datos

Aliados:

Microsoft

idata
Advanced analytics for business

Vigilada Mineducación

UNIVERSIDAD EAFIT

# INTRODUCCIÓN A LA CIENCIA DE DATOS

Aliados:

Microsoft

idata
Advanced analytics for business

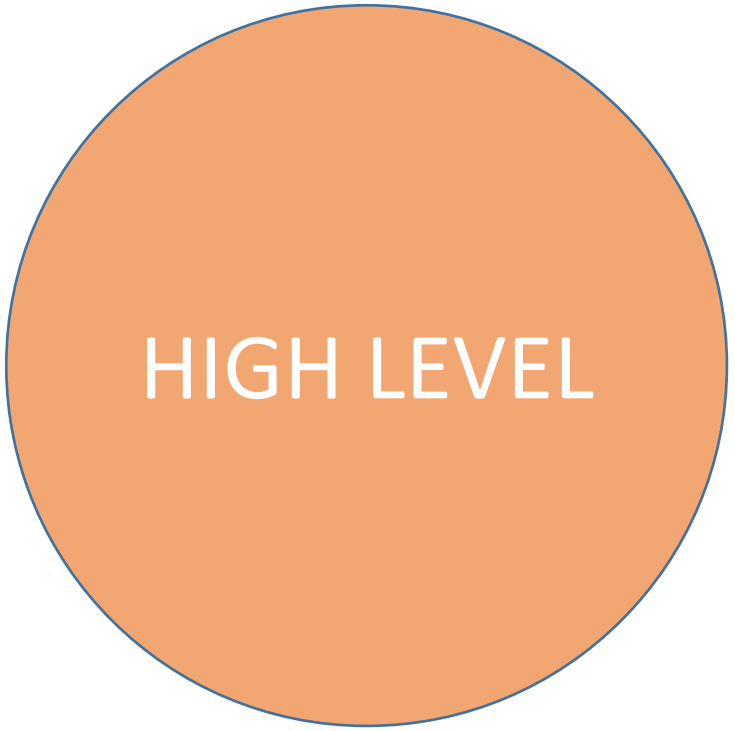
Vigilada Mineducación

# WHAT IS A PROGRAMMING LANGUAGE ?

"An artificial language designed to express computations that can be carried out by machines such as computers"

UNIVERSIDAD EAFIT

# TYPES OF PROGRAMMING LANGUAGES

# LOW LEVEL


## Machine language

- A low-level language comprised of binary digits (ones and zeros)
- Only language that computers can understand

Example: **10110000 01100001**

## Assembly Language

- Designed to be readable by humans
- Type of programming language that translates high-level languages into machine language

Example: **MOV AL, 61h** (asigna el valor hexadecimal 61 al registro «AL»)

# HIGH LEVEL

## Compiled Language

- The code is translated into machine language previously and then it is executed

C, C++, Erlang, Haskell, Rust, and Go ...

## Interpreted Language

- As the language is being translated into machine code, it is executed

PHP, Ruby, Python, JavaScript ...

High-level language allows you to write codes using languages we know

# SOME MODERN PROGRAMMING LANGUAGES

**Python**
Data science, machine learning, and artificial intelligence

**Java**
Mostly used for enterprise-level and Android software development

**Ruby**
Web application development

**JavaScript**
Rapid and productive web development

**C++**
System software development, microcontrollers

**C#**
Game development, web forms, and web applications development.

UNIVERSIDAD EAFIT

# PYTHON

- Multi- purpose (Web, GUI, Data Science, etc…)
- Object Oriented
- Interpreted
- Dynamically typed
- Focus on readability and productivity

# WHO USE PYTHON ?


## ALL KIND OF PROFESSIONALS

Economists, developers and engineers use it to automate processes, transform data , perform calculations, etc..

## MAJOR ORGANIZATIONS

Google, Pinterest, Instagram, YouTube, DropBox, NASA, ESRI

## SPECIALIZATIONS AND INDUSTRIES

Database Access, scientific and numeric computing, desktop graphical user interfaces, web development

# WHAT MAKES LEARNING IT IMPORTANT?

- Open Source
- Simple to learn
- Easily to read
- RAD
- Big community

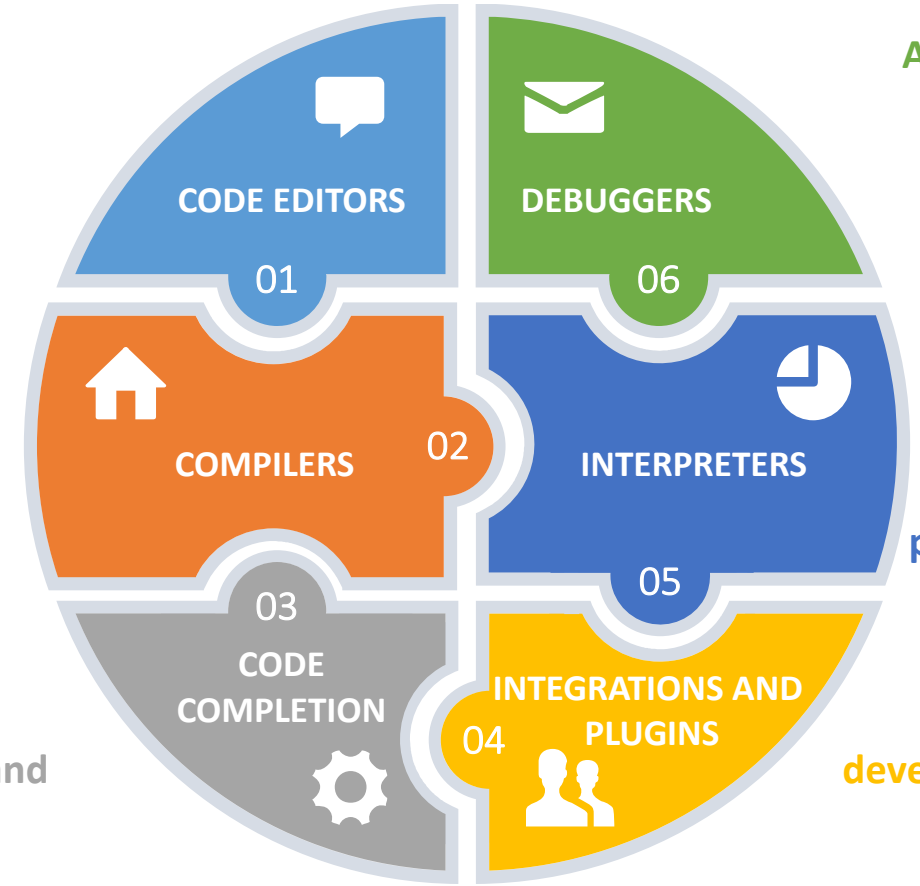
# PROGRAMMING SOFTWARES

IDE

Designed to write and manipulate source code

Translate programming language into a form machines can process.

Assist programmers by intelligently identifying and inserting common code components.




Assist users in identifying and remedying errors within source code.

Loads the code entered and translates the instructions so that the program can be executed.

Tools will improve development workflows and productivity.

UNIVERSIDAD EAFIT

# DATA TYPES

See Jupyter Notebook

UNIVERSIDAD EAFIT

# DATA STRUCTURES

See Jupyter Notebook

UNIVERSIDAD EAFIT

# LOOPS

See Jupyter Notebook

UNIVERSIDAD EAFIT

# FUNCTIONS

See Jupyter Notebook

UNIVERSIDAD EAFIT

# OBJECTS

See Jupyter Notebook

# Numpy


- **Mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more**
- **Brings the computational power of languages like C and Fortran to Python**
- **Click Here & Put Your Option**
  Lorem ipsum dolor sit amet, consectetur elit.

https://numpy.org/doc/stable/user/quickstart.html

# Pandas


- Data analysis

https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html

# Matplotlib


- Create static, animated, and interactive visualizations

https://matplotlib.org/2.0.2/gallery.html - showcase

# Scikit Learn


- Predictive data analysis
- Built on NumPy, SciPy, and matplotlib

https://scikit-learn.org/stable/

# Tensorflow


- Develop and train ML models
- Used for complex models

# Shap


- **Approach to explain the output of any machine learning model**

UNIVERSIDAD EAFIT

# Q&A

Aliados:

Microsoft

idata Advanced analytics for business

# Workshop #1

https://github.com/santyjara/Data-science-workshops

# Lab capitulo #1

1. Leer el Data set del laboratorio
2. Encontrar las columnas con mas del 40% de valores nulos
3. Identificar que todos los tipos de columnas sean los correctos

UNIVERSIDAD EAFIT

# ¡Gracias!

Aliados:

Microsoft

idata
Advanced analytics for business

Vigilada Mineducación