

Data Analysis with Python

Paolo G. Hilado
paolo@ptcsitesresearch.com



What's in store for us?

- * Introduction to the Python Programming Language
- * Requisites for Analyzing Data using Python
- * Refresher on Data Levels of Measurement
- * Requisites for Inferential Analysis
- * Orientation of the Spyder IDE
- * Data Management
- * Exploratory Data Analysis
- * Descriptive Analysis
- * Inferential Analysis



The Python Programming Language

paolo@ptcsitesresearch.com



Photo Source: wallpapercave.com



GUIDO VAN ROSSUM
Creator of Python

Photo Source: TFIR: Open Source & Emerging Technologies

Python as an Open Source

paolo@ptcsitesresearch.com



Having a
source code freely
available for
possible
modification and
redistribution.

Merriam-Webster. (n.d.). Open-source. In *Merriam-Webster.com dictionary*. Retrieved October 3, 2022, from <https://www.merriam-webster.com/dictionary/open-source>

Python as a General Programming Language

paolo@ptcsitesresearch.com

PYTHON PROJECT IDEAS

STARTER

1 2 3 4 5
6 7 8 9 0

NUMBER GUESSING



SNAKE GAME



STORY GENERATOR



STOPWATCH

INTERMEDIATE



TYPING SPEED APP



EMAIL AUTOMATION



PAINT APP



WHATSAPP AUTOMATION

ADVANCED



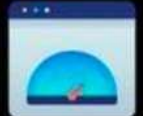
CHAT APP



FACE DETECTION



YOUTUBE CLONE



DATA DASHBOARD

EXPERT



INSTAGRAM BOT



TINDER BOT



SUDOKU SOLVER



REAL ESTATE
PRICE PREDICTION

Python as a General Programming Language

paolo@ptcsitesresearch.com



Python as a High-Level Language

paolo@ptcsitesresearch.com



C++

```
#INCLUDE <Iostream>

using namespace std;

int main ()
{
    cout<<"Hello World";

    return 0;}
```



Python

```
print ("Hello World")
```

Lots of Useful Libraries for Data Analysis

pandas

numpy

scipy

statsmodels

paolo@ptcsitesresearch.com

researchpy

pingouin

seaborn

matplotlib

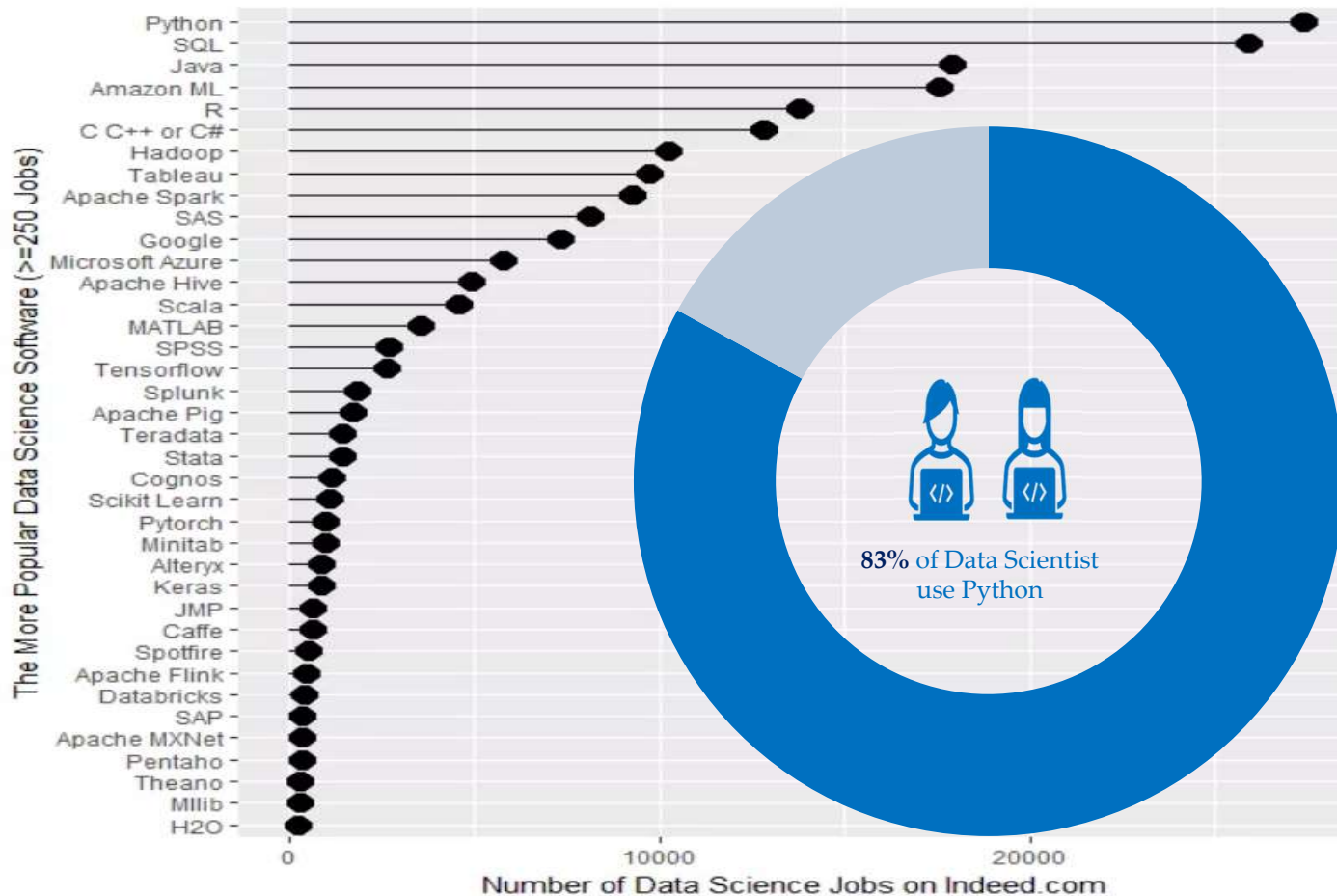


Figure 1a. Number of data science jobs for the more popular software.

Source: Muenchen, B. (2019). Analyzing the World of Analytics, Kdnuggets News

Python's Popularity in Data Science

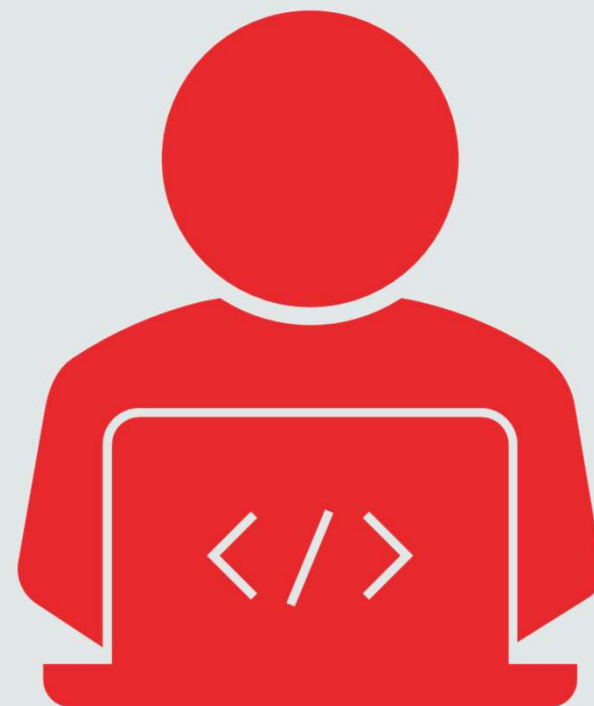
Source: KDNuggets survey, 2019

paolo@ptcsitesresearch.com

Valuable for Data Science

paolo@ptcsitesresearch.com

Handles processes such as data tidying, data analysis, visualizations, machine learning, deep learning, and natural language processing.



From User to Creator

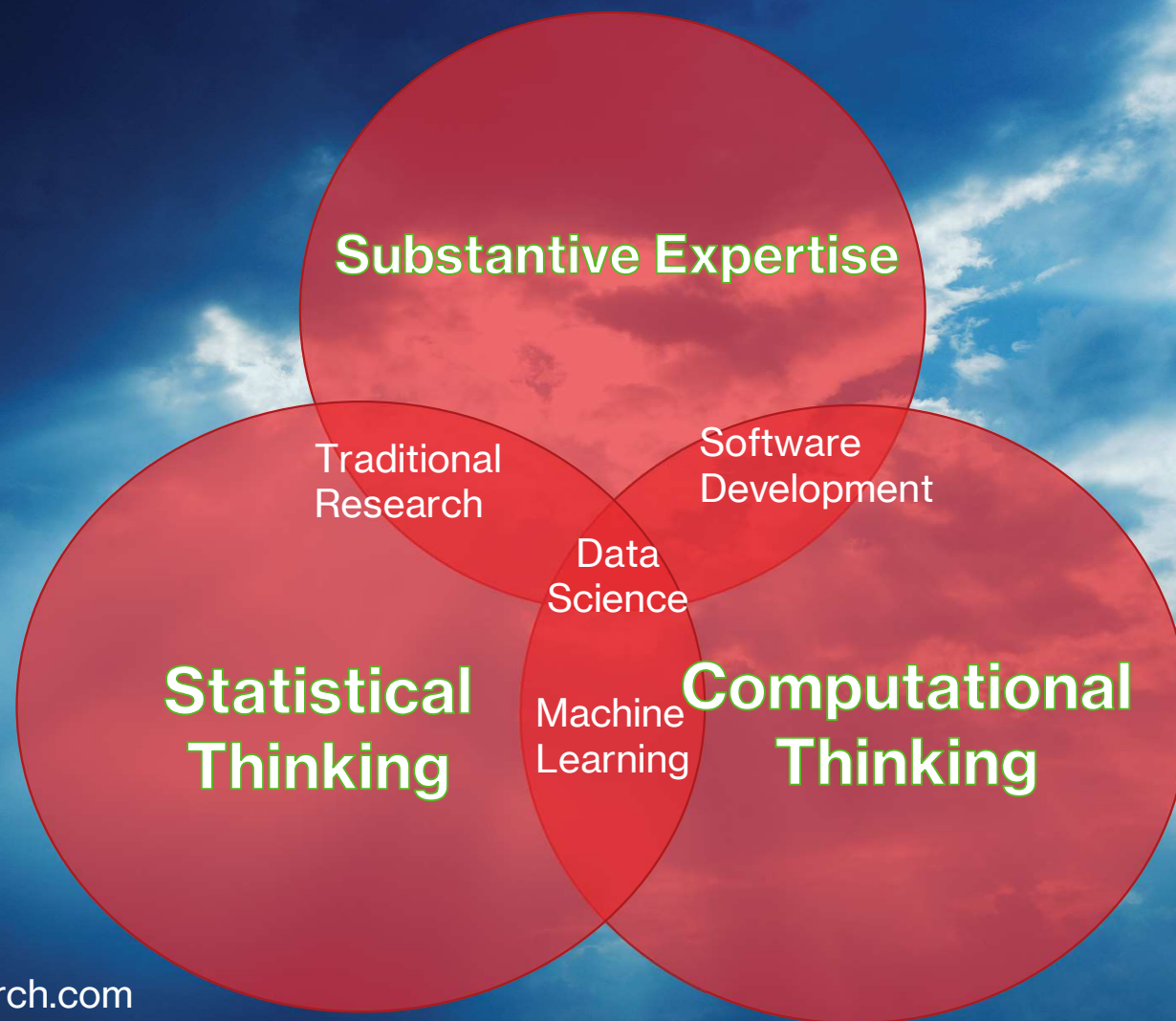
paolo@ptcsitesresearch.com

```
object to mirror
mirror_mod.mirror_object =
operation == "MIRROR_X":
mirror_mod.use_x = True
mirror_mod.use_y = False
mirror_mod.use_z = False
operation == "MIRROR_Y":
mirror_mod.use_x = False
mirror_mod.use_y = True
mirror_mod.use_z = False
operation == "MIRROR_Z":
mirror_mod.use_x = False
mirror_mod.use_y = False
mirror_mod.use_z = True

selection at the end -add
mirror_ob.select= 1
mirror_ob.select=1
context.scene.objects.active
("Selected" + str(modifier_
mirror_ob.select = 0
= bpy.context.selected_object
data.objects[one.name].select
print("please select exactly

-- OPERATOR CLASSES -----

types.Operator):
X mirror to the selected
object.mirror_mirror_x"
mirror X"
```



paolo@ptcsitesresearch.com

Requisites for Analyzing Data with Python

Data Levels of Measurement



Categorical Variables

paolo@ptcsitesresearch.com

Nominal – numbers merely represent an element

e.g., 1 – Male, 2 – Female

Ordinal – numbers inform us of ranking

e.g., 1 – Small, 2 – Medium, 3 – Large

Continuous Variables

Ratio Scale – Units are equally spaced
e.g. height

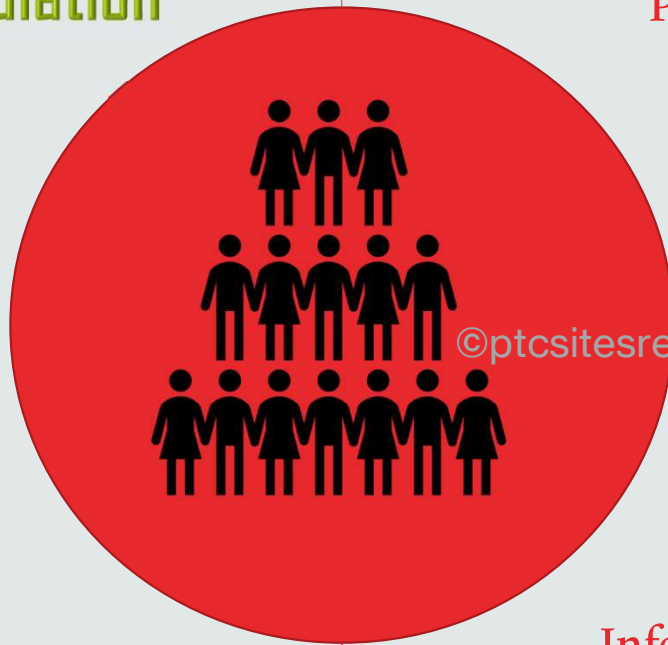
Interval Scale - Units are equally spaced but there is no true zero
e.g. temperature

Requisites of Inferential Analysis



Making an Inference...

Population



Probability

Descriptive Stats



Sample

Inferential Stats

Hypothesis Testing

Interesting!

paolo@ptcsitesresearch.com

A	B	C	A	B	C	A	B	C
	GWA (Male)	GWA (Female)		GWA (Male)	GWA (Female)		GWA (Male)	GWA (Female)
	88	90		88	90		88	90
	82	82		82	82		82	82
	73	85		73	85		73	85
	72	88		72	88		72	88
	99	75		99	75		99	75
	82	72		82	72		82	72
	94	84		94	84		94	84
	86	82		86	82		86	82
	78	91		78	91		78	91
	90	94		90	94		90	94
Population Mean	84.4	84.3	Population Mean	84.4	84.3	Population Mean	84.4	84.3
Sample Mean (1)			Sample Mean (1)	81.4	86.2	Sample Mean (1)	81.4	86.2
Sample Mean (2)			Sample Mean (2)			Sample Mean (2)	77.4	84.4
Sample Mean (3)			Sample Mean (3)			Sample Mean (3)		



Integrated Development Environment

Integrated Development Environment

- * An application that is instrumental in creating applications
- * Usually includes a code editor, interpreter, and debugger in a graphical user interface

paolo@ptcsitesresearch.com



The Spyder IDE



Spyder

Open-source
IDE popular for
python

Interactive
execution allows
to run line, file
and etc.

Convenience in
managing
working
directory

paolo@ptcsitesresearch.com

Tab-completion
assist when
coding

Provides real
time
introspection

Spyder

Spyder (Python 3.9)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\ACER\Documents\Pyspark Projects

C:\Users\ACER\spyder-py3\temp.py

temp.py

1

I

Source Console Object

Usage

Here you can get help of any object by pressing **Ctrl+H** in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in *Preferences > Help*.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console 1/A

```
Python 3.9.7 (default, Sep 16 2021, 16:59:28) [MSC v.1916 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.
```

```
IPython 7.29.0 -- An enhanced Interactive Python.
```

```
In [19]:
```

IPython console History

LSP Python: ready conda (Python 3.9.7) Line 1, Col 1 ASCII CRLF RW Mem 38%

Let us go hands-on!

