

# Data Mining

Intro. Python

**Santiago Alonso-Díaz**

Tecnológico de Monterrey  
EGADE, Business School

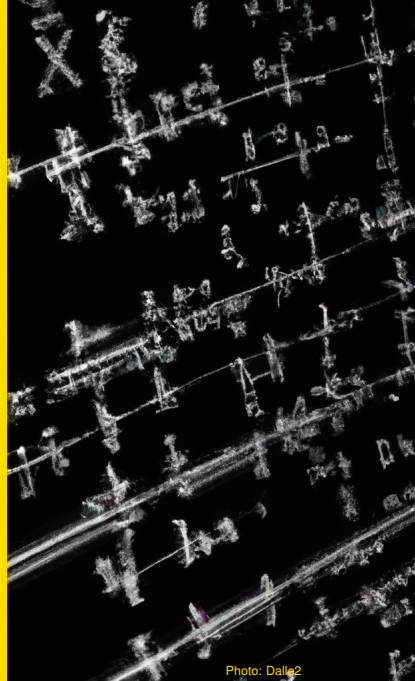


Photo: Dalle2

Why should someone interested in businesses learn/understand code?

- Excel is great but ...
- Data science
- Execute non-trivial models and analyses

# Table of contents

## 1 Python basics

## 2 Python data objects

- Numpy
- Pandas
- Polars

## 3 References

# Python basics

# Basics

A Jupyter notebook premier on Google's Collab

Open: Python \ Python Tutorial \ python-intro - Eng.ipynb

# Python data objects

# Numpy

Loosely based on [scipy.org](https://scipy.org) lectures

## What is Numpy?

Python package to build and compute on **arrays** of any dimension (e.g. 1D vectors, 2D matrices, 3D coordinates, 4D fMRI data).

# Numpy

In Python do:

## 1D array

```
a = np.array([0, 1, 2, 3])  
print(a, a.ndim, a.shape)
```

## 2D array

```
b = np.array([[0, 1], [2, 3]])  
print(b, b.ndim, b.shape)
```

## 3D array

```
c = np.array([[[1,2], [2,3]], [[3,4], [4,5]]])  
print(c, c.ndim, c.shape)
```



# Numpy

Not just real numbers:

## Complex numbers

```
a = np.array([1+2j, 3+4j, 5+6*1j])
```

## Booleans

```
b = np.array([True, False, False, True])
```

## Strings

```
c = np.array(['Bonjour', 'Hello', 'Hallo'])
```

# Numpy

Numpy has many routines/functions. Do in Python:

```
a = np.arange(1, 9, 2)
b = np.linspace(0, 1, 6)
c = np.ones((3, 3))
d = np.zeros((2, 2))
e = np.random.rand(4)
f = np.random.randn(4,4)
g = np.sqrt(33)
h = np.sort([3,2,1])
i = np.round(np.pi, 4)
```

# Numpy

## Linear algebra functions:

### Dot product

```
a = [[1, 0], [0, 1]]
```

```
b = [[4, 1], [2, 2]]
```

```
np.dot(a, b)
```

### Eigenvalues

```
matrix = np.array([[1,2],[3,4]])
```

```
np.linalg.eigvals(matrix)
```

# Numpy

## Indexing:

```
>>> a[0, 3:5]
```

```
array([3, 4])
```

```
>>> a[4:, 4:]
```

```
array([[44, 45 ],  
       [54, 55]])
```

```
>>> a[:, 2]
```

```
a([2, 12, 22, 32, 42, 52])
```

```
>>> a[2::2, ::2]
```

```
array([[20, 22, 24],  
       [40, 42, 44]])
```

0	1	2	3	4	5
10	11	12	13	14	15
20	21	22	23	24	25
30	31	32	33	34	35
40	41	42	43	44	45
50	51	52	53	54	55

Source: <https://lectures.scientific-python.org/>

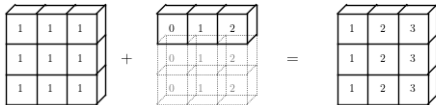
# Numpy

## Broadcasting:

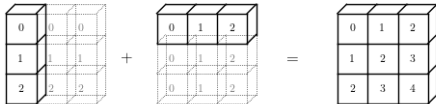
`np.arange(3) + 5`



`np.ones((3, 3)) + np.arange(3)`



`np.arange(3).reshape((3, 1)) + np.arange(3)`



Source: <https://numpy.org/doc/stable/user/basics.broadcasting.html>

# Pandas

[Check Molin tutorial on Github](#)

## What is Pandas?

Python package based on Numpy but with more data science functionalities.

# Polars

[Check Polars documentation](#)

## What is Polars?

Python package inspired in Pandas but faster.

# References