

DATA SCIENCE

Bienvenid@s a Acámica

DS-BUE-8/Clase 01



Hoja de ruta de hoy

19:00	Presentación del equipo
19:20	Sobre la metodología Acámica.
19:50	Presentación personal: "Encuentra tu Pareja" Nos aproximamos a presentar la carrera entre todos/as
20:50	Break
21:00	Presentación Carrera
21:45	Reconstrucción y cierre

DS-BUE-8/Clase 01



ACÁMICA

ACAMICA
talentodelfuturo



ACÁMICA PRESENCIAL

**FORMAMOS EL
TALENTO DEL FUTURO**

¿QUIÉNES
SOMOS?

ACÁMICA PRESENCIAL



A grayscale photograph of students in a classroom. One student is pointing at a laptop screen while others look on. Papers and a pen are on the table. A blue square is on the left side of the image.

METODOLOGÍA: Blended Learning



PROGRESO /
RECURSOS



COMUNICACIÓN

A group of people are working at a wooden table. In the foreground, a person in a red shirt is writing in a notebook. To their left, another person is using a laptop. On the table, there are two glasses of iced drinks, a pair of glasses, and several notebooks. The background shows other people working. A semi-transparent dark overlay covers the bottom half of the image, and a blue square is on the left side.

PROYECTOS

DATA SCIENCE

Actividad de presentación
personal:
Encuentra tu pareja!



Actividad de presentación de la carrera: ¿Qué es Data Science?





¿Qué es Data Science?



¿Qué es Data Science?

Recogida de la información



No siempre se tiene la información inmediatamente disponible en una base de datos (u hoja de cálculo) y **su obtención es parte del proceso**. La recogida de la información puede realizarse mediante colocación de **sensores, scrapeo de la web, peticiones a APIs, formularios...** o de cualquier forma que la creatividad nos permita.

Preprocesado



Una vez se tiene la información que se considera necesaria para resolver nuestro modelo, se le debe aplicar un **preprocesado**. Es decir, sin perder información de valor, dar vueltas a la información que ya tenemos para **prepararla para la siguiente fase**. El objetivo es representar la información en un formato que consiga reducir el coste de cómputo y optimizar los resultados de los algoritmos.



¿Qué es Data Science?

Entrenamiento o machine learning



Entrenar el modelo significa **alimentar algoritmos de machine learning** con nuestros datos.

Los algoritmos de machine learning (o aprendizaje automático) son capaces de **predecir y clasificar información** nueva, a raíz de haber sido entrenados con información pasada.

Testing



Por lo general, no existe un algoritmo mejor que otro, algunos rinden mejor en unos casos y otros, en otros casos. Y la manera de hallar la mejor solución es **probándolos uno a uno**, con diferentes configuraciones, hasta encontrar el mejor para nuestro caso. Se sigue una metodología experimental **prueba-error** porque, a priori, es muy difícil adivinar qué tipo de algoritmo y configuración va a tener mejores resultados. Solo los mejores profesionales en machine learning pueden tener una vaga idea de qué tipo de algoritmos pueden funcionar mejor con un set de datos determinado. Pero aún así **tendrían que probar varias opciones y testearlas**.



¿Qué es Data Science?

Visualización e interpretación de datos



Finalmente, para poder comunicarlo a terceros, se tiene que representar de alguna forma el conocimiento obtenido. Se tiene que hallar la manera más **limpia e intuitiva de visualizar los resultados** y, con ayuda de diferentes softwares, crear estas visualizaciones.



¿Qué es Data Science?

Ejemplos de aplicaciones:

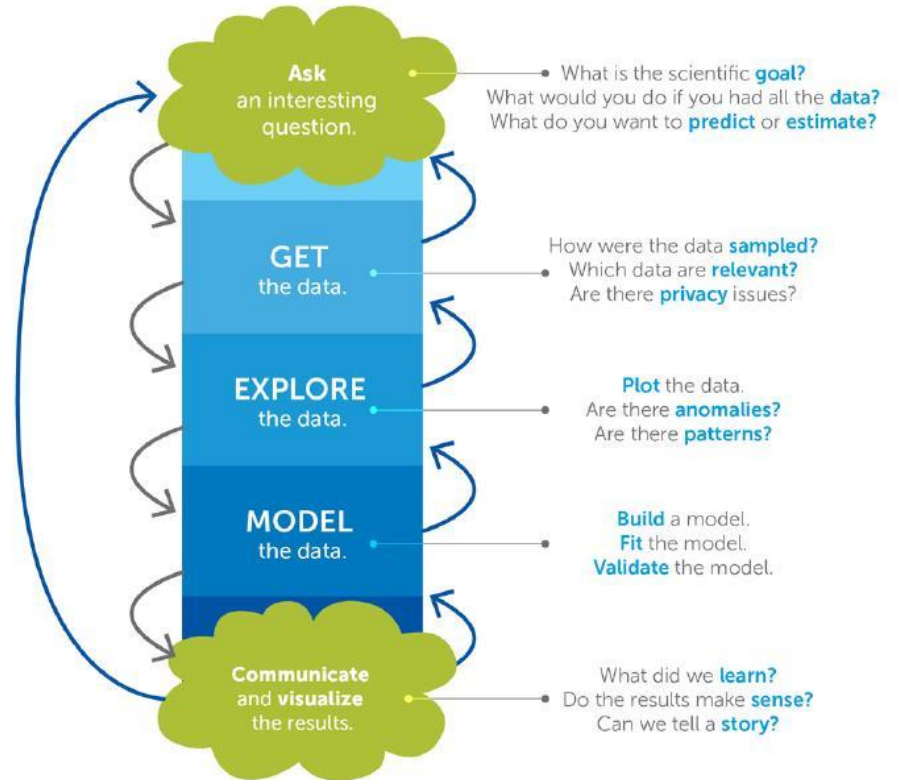
- Chatbots
- Face recognition
- Detección de fraude bancario
- Detección de enfermedades cardíacas
- Publicidad dirigida
- Análisis de opinión a gran escala
- ...



¿Qué es Data Science?

Miembros de la Familia

- Machine Learning
- Big Data
- Data Mining

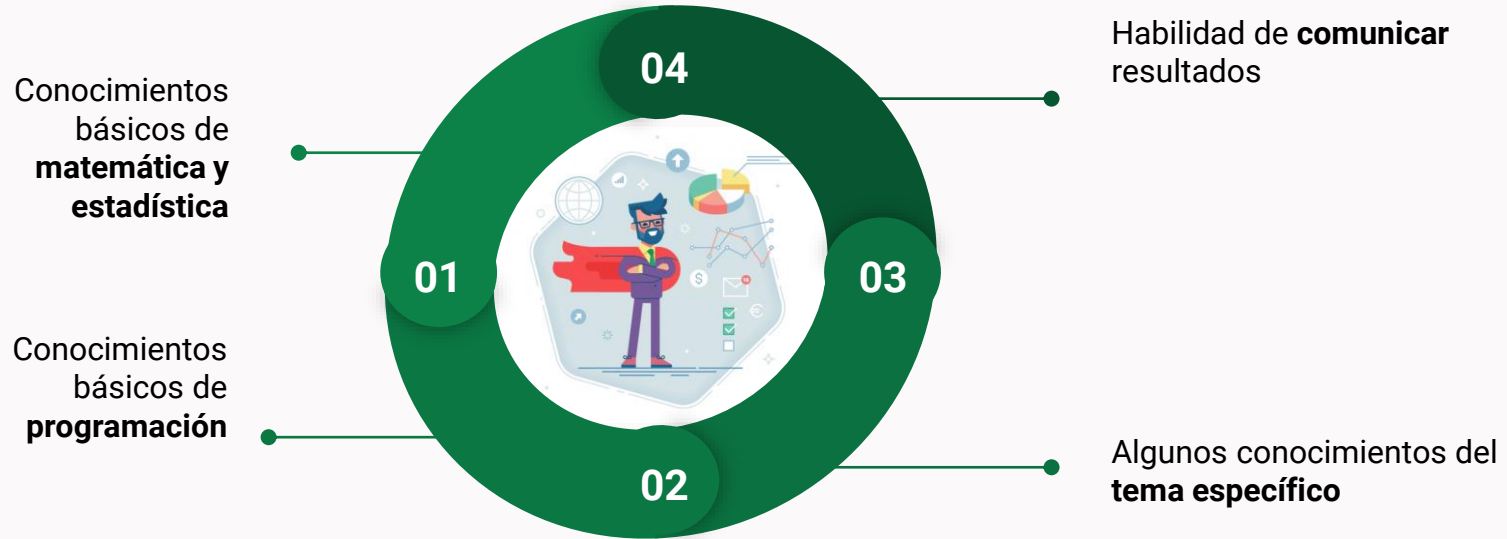


Skillset de un data scientist

¿Qué HABILIDADES BLANDAS y que
HABILIDADES DURAS pensás que implica ser un
Data Scientist?



Skillset de un data scientist



Contenidos de la Carrera y Cronograma



Conocimientos ¿previos?

- Probabilidad y Estadística (<https://seeing-theory.brown.edu/es.html#firstPage>)
- Programación (cualquier lenguaje)
- Uso de la terminal (Windows, Mac, Ubuntu)
- [Recomendación] Entender algo de inglés

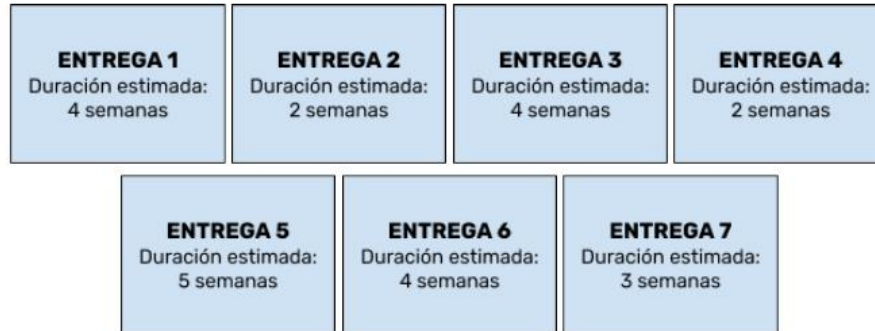
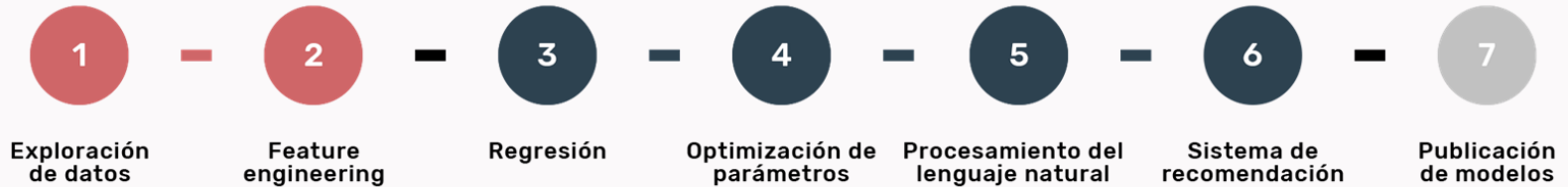


Entregas y Proyectos

ADQUISICIÓN Y EXPLORACIÓN

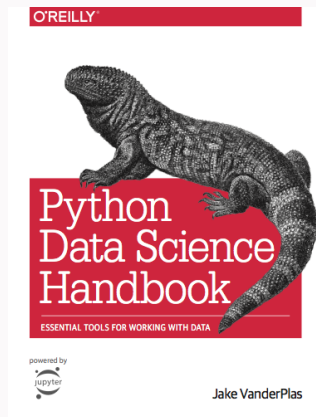
MODELADO

DEPLOY



Recursos y Bibliografía

- Videos de la plataforma de Acámica
- Introducción a python: <https://learnxinyminutes.com/docs/python3/>
- Introducción a algunas de las librerías (y un poco más):
<https://jakevdp.github.io/PythonDataScienceHandbook/>
- Más en <https://github.com/acamica/biblio-ds>



Herramientas





- Alto nivel, fácil de usar
 - Rápido y eficientes
 - Gran comunidad online
 - Librerías para prácticamente todo lo que uno quiera hacer
-
- **Importante:** Python 3!
 - **Instalación recomendada:** Anaconda o Miniconda.

```
[1]: print('Hola Mundo')
```

Hola Mundo

```
[2]: for i in range(5):  
      print(i)
```

0

1

2

3

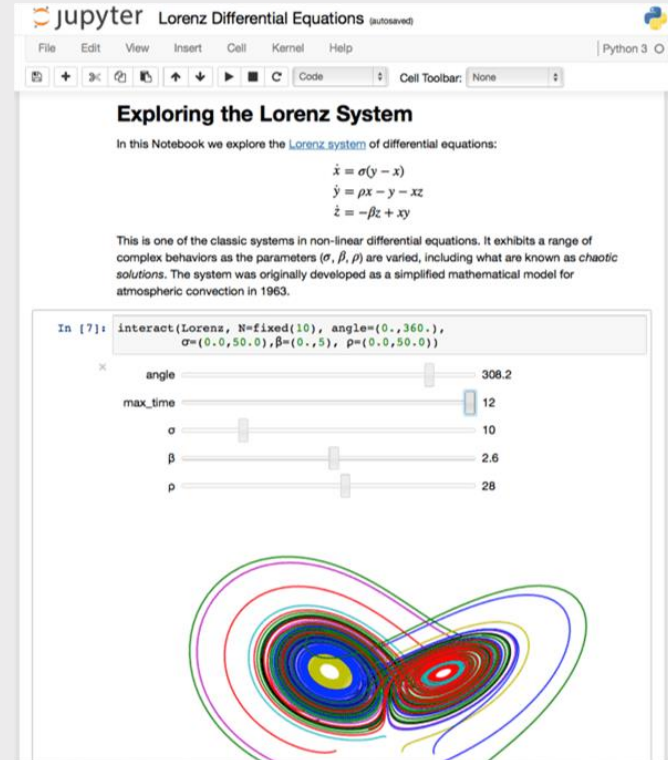
4





Entorno de trabajo

<https://jupyter.org/>





NumPy nos provee de:

- Estructuras de datos
- Operaciones eficientes sobre los datos

Es el corazón del ecosistema de Data Science.

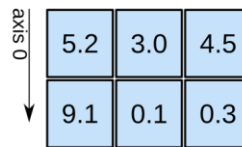
www.numpy.org

1D array



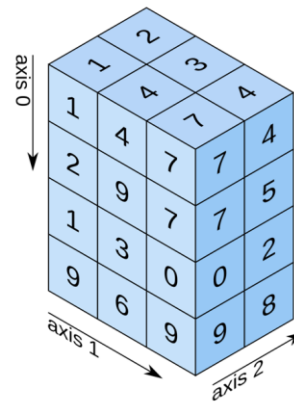
shape: (4,)

2D array



shape: (2, 3)

3D array



shape: (4, 3, 2)



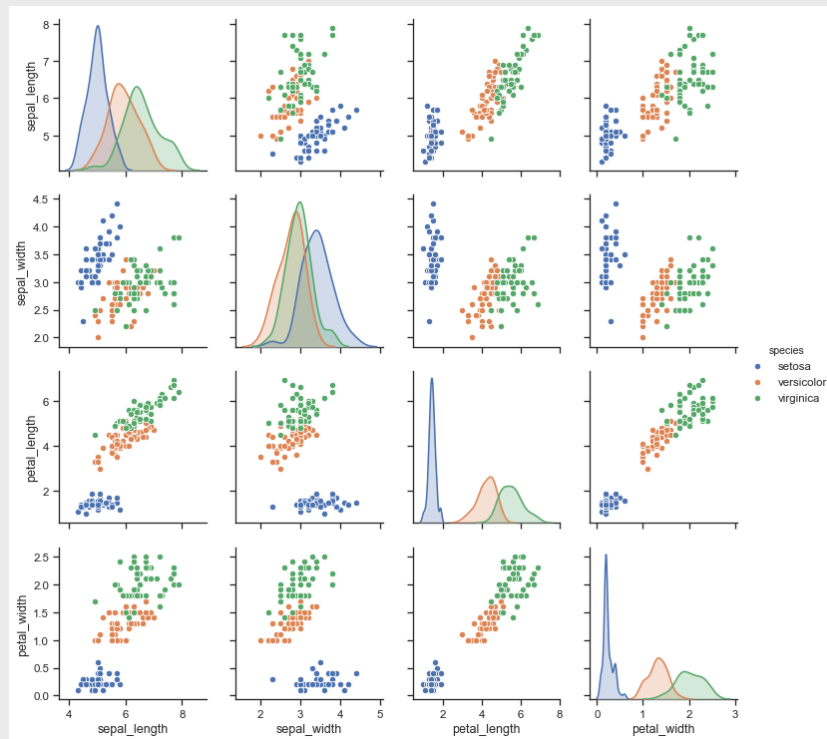


seaborn

¡Lindos Gráficos!

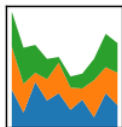
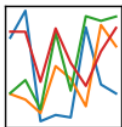
<https://matplotlib.org/>

<https://seaborn.pydata.org/>



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Librería para trabajar con estructuras de datos

<https://jupyter.org/>

	sepal length	sepal width	petal length	petal width	target
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa



Y la más importante



|

Buscar con Google

Me siento con suerte

Ofrecido por Google en: [Español \(Latinoamérica\)](#)



Para el próximo encuentro...



Para la próxima...

- Instalar una distribución de Python con Anaconda (sean libres de probar miniconda si lo desean) «««« Tenemos tutorial »»»»
- Lograr abrir un entorno de trabajo en Jupyter

