

Módulo: Fundamentos de Python

DATA
ANALÍTICA



Agenda

0. Reglas del Juego
1. ¿Qué es la ciencia de datos?
2. ¿Qué es Python? ¿Por qué Python es un lenguaje de programación popular para la ciencia de datos?
3. Introducción a Python
4. Introducción a pandas

Reglas del Juego

Mantener el micrófono apagado en caso de que no vayan a hablar.



**Nos encantaría verte.
Ten tu cámara encendida y
conozcámonos
virtualmente.**



**Preguntar en caso que
tengan dudas.**



**Disfruta de este espacio.
Desconecta del resto y
participa.**



**Por cada clase tendremos
10 min o 15 min de receso.**



Modo de Evaluación

20%

Evaluación continua:

Notebooks de ejercicios,
formularios de ejercicios o
tareas (challenges).

30%

Test de seguimiento:

Exámenes de 5 a 10 preguntas
que serán tomados al inicio de la
3ra, 5ta y 7ma sesión.

50%

Examen Final:

Examen donde evaluaremos los
aprendizajes obtenidos a lo largo del curso

**A VECES QUIERO CAMBIAR
LAS REGLAS DEL JUEGO**



**LUEGO RECUERDO QUE NO
SE PUEDE Y SE ME PASA.**

meme-generator.es



**¿Qué es la ciencia de
datos?**



Diversidad

Último

Revista

Ascender

Temas

podcasts

Video

Tienda

la gran idea

Datos y visuales

Selecciones de casos

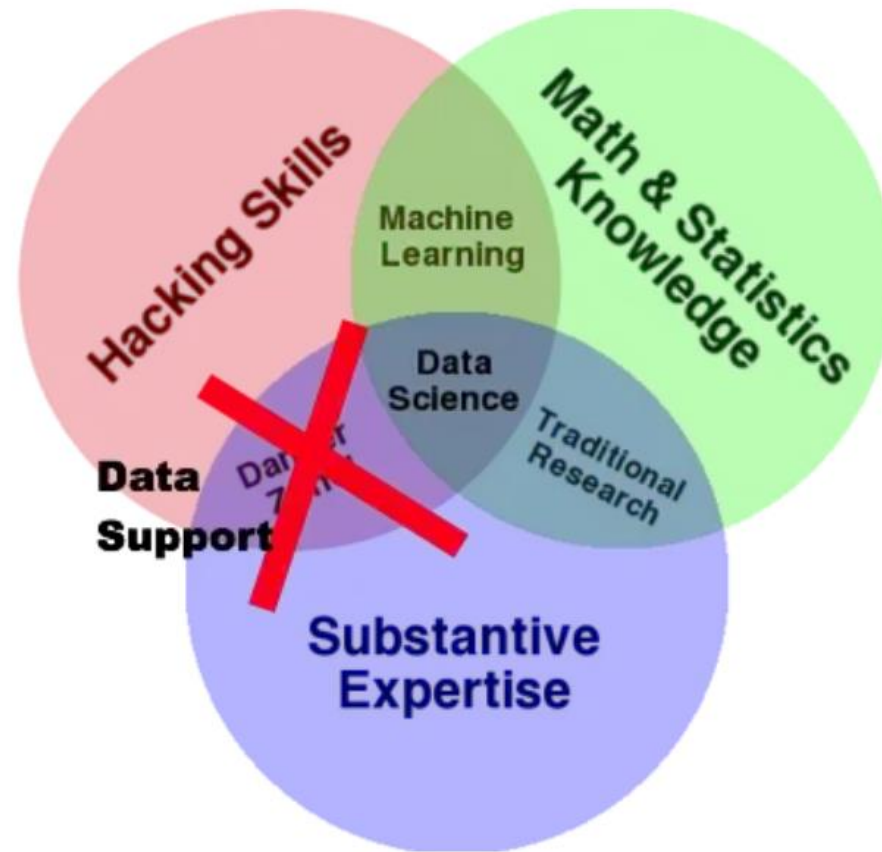
Analítica Y Ciencia De Datos

Científico de datos: el trabajo más sexy del siglo XXI

Conozca a las personas que pueden extraer tesoros de datos desordenados y no estructurados. por Thomas H. Davenport y dj patil

De la revista (octubre de 2012)

Ciencia de Datos



Ciencia de Datos

- Estudio de datos con el fin de **extraer información significativa** para empresas. Es un enfoque **multidisciplinario** que combina principios y prácticas del campo de las **matemáticas, la estadística, la inteligencia artificial y la ingeniería de computación** para analizar grandes cantidades de datos.
- Es un **conjunto de metodologías** para tomar **miles de formas de datos** que están disponibles y usarlos para **sacar conclusiones significativas**.

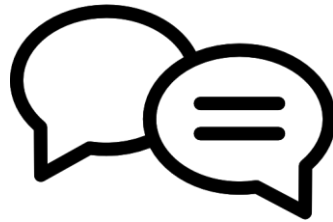
Ciencia de Datos: Ejemplo de diferentes formas de datos



Tabla en Excel



Tabla en SQL



Texto



Imágenes



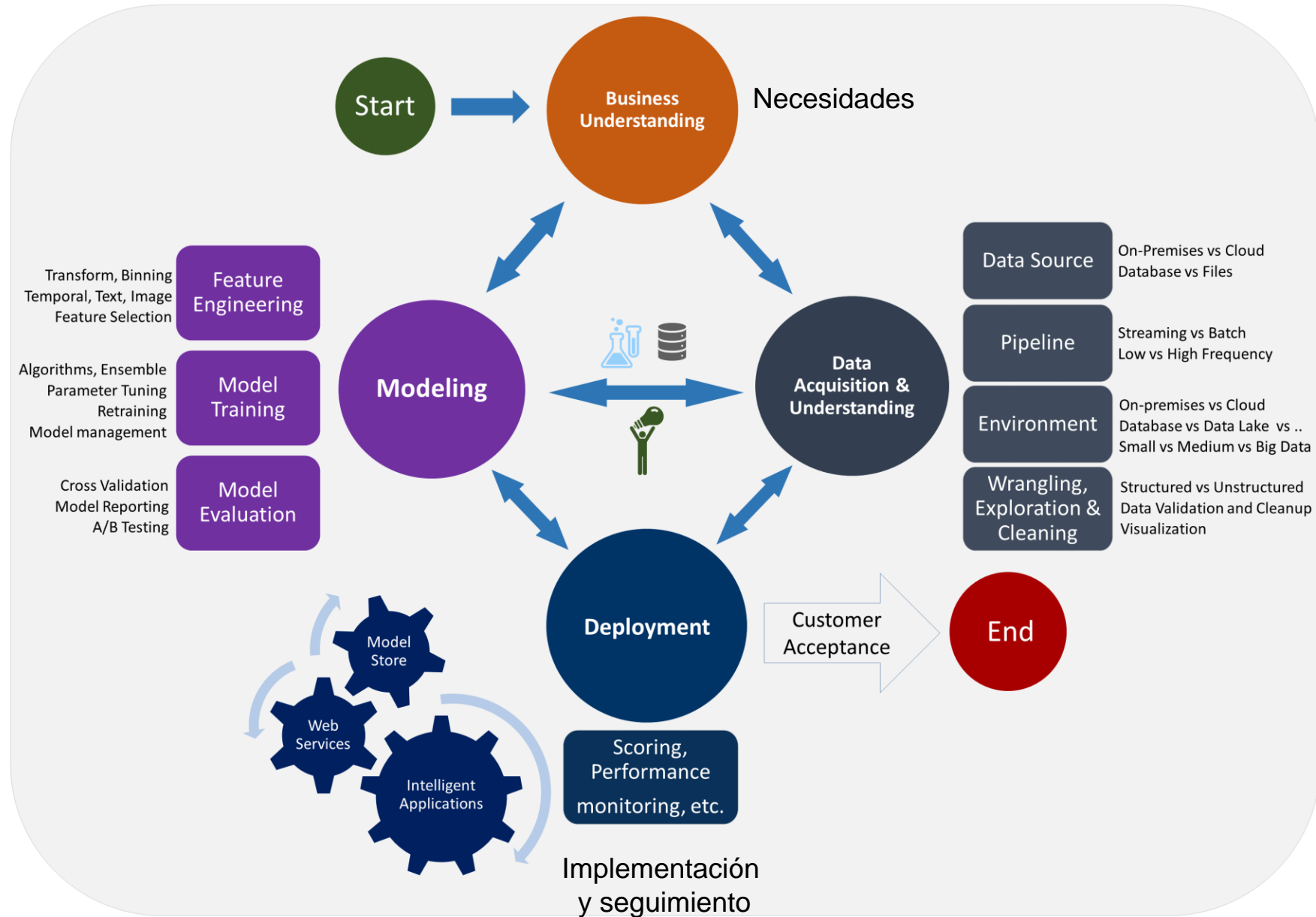
Videos

Data Science Lifecycle

Creación de variables

Entrenamiento

Evaluación



Fuentes de datos

Flujo de datos

Infraestructura tecnológica

Limpieza y exploración

¿Qué es Python?

Python es un lenguaje de programación que se suele utilizar para crear sitios web y software, automatizar tareas y realizar análisis de datos. Se trata de un lenguaje de propósito general.

Beneficios de usar Python

- Es gratuito
- Es fácil de usar
- Es altamente compatible
- Está orientado a objetos
- Tiene muchas bibliotecas
- Tiene estructuras de datos incorporadas
- Es ampliamente aplicable
- Aumenta la Velocidad y la Productividad

Introducción a Python

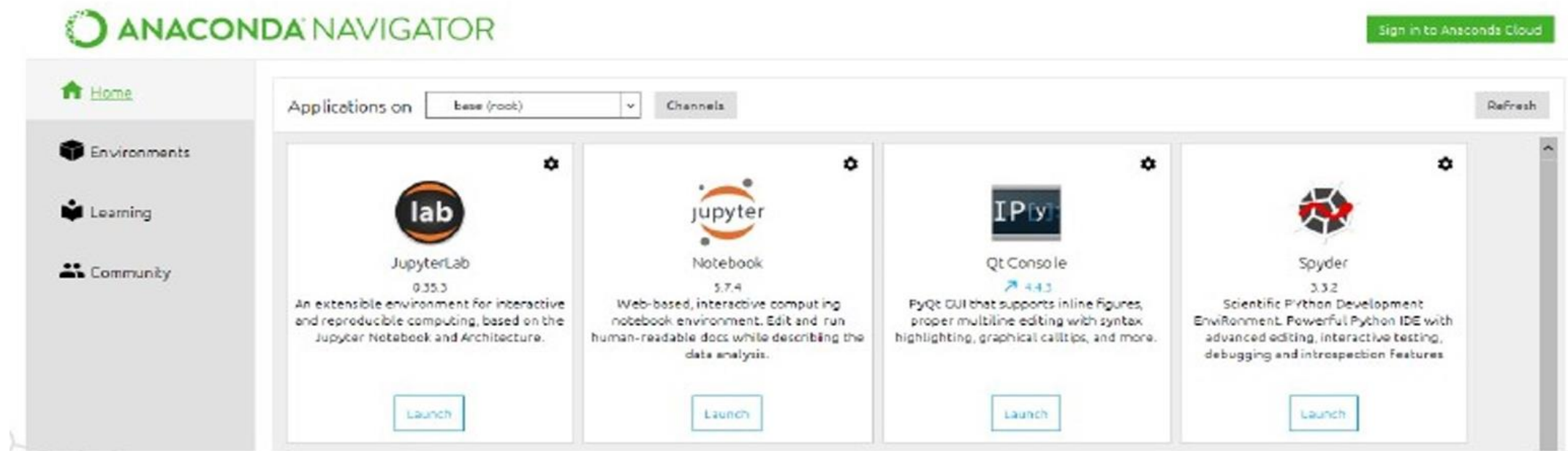
Conociendo la suite Anaconda



ANACONDA®

Conociendo la suite Anaconda

Anaconda es un suite de código abierto que contiene, entre sus principales aplicaciones, Jupyter Notebook y Spyder. Cuenta con +250 librerías instaladas ideales para proyectos en Data Science, además de un gestor para actualizar o instalar librerías.



Instalación de Anaconda

Data science technology for a better world.

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. Start working with thousands of open-source packages and libraries today.

Download 

For Windows


Python 3.9 • 64-Bit Graphical Installer • 621 MB

Get Additional Installers



<https://www.anaconda.com/>

Interfaces para trabajar con Python




JupyterLab

↗ 3.2.1

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch




Notebook

↗ 6.4.5

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



Spyder

5.1.5

Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Launch

Interfaces para trabajar con Python



Google Colaboratory

Challenge 1:

Elaborar 5 ejercicios (Python) de lo aprendido durante la clase de introducción a Python.

Enviarlos por correo con el asunto: Challenge 1 – Módulo Fundamentos de Python – [Apellidos y nombres]

Correo: team@dataanalitica.net

Los lenguajes de programación en mi
currículum viendo como agrego otro
después de haber aprendido
el Hola Mundo en él.



Introducción a Pandas



Se utiliza en casi cualquier análisis de datos realizado en Python.
Incluye estructuras de datos y operaciones para manipular tablas
numéricas y series temporales.

Tipos de Datos comúnmente usados en Pandas

int y float, estos datos serían numéricos, los primeros, como su nombre lo indican serían números enteros mientras que los float o flotantes, serían datos con decimales.

object, funciona de forma similar a la cadena de datos en Python.

datetime, que vendría siendo algo así como día y hora, es un tipo de datos muy útil para manejar series temporales de datos.

Reto en Clase:

Resuelve los 10 ejercicios del Notebook 3_Ejercicios Pandas.ipynb usando lo aprendido durante la clase de introducción a Pandas

Enviarlo al final de la sesión por correo con el asunto:
Ejercicios Pandas-Módulo Fundamentos de Python-
[Nombre y Apellidos]

Correo: team@dataanalitica.net

Challenge 2:

Elaborar 5 ejercicios (Python) de lo aprendido durante la clase de introducción a Pandas

Enviarlo por correo con el asunto: Challenge 2 – Módulo Fundamentos de Python – [Apellidos y nombres]

Correo: team@dataanalitica.net