

ANÁLISIS DE DATOS

FACULTAD DE INGENIERÍA Y NEGOCIOS

(ACI777)

Melissa Alegría Arcos, PhD.
malegriaa@udla.cl

Agosto 2021

Descripción del programa

Modalidad : Presencial – E-Support

Horario : Miércoles 15:30 a 18:50

Distribución de horas de dedicación:

- Laboratorio: 54
- Trabajo Personal 126

La asignatura ANÁLISIS DE DATOS tiene por meta formativa que los estudiantes aprendan los fundamentos de análisis basado en datos para resolver problemas organizacionales y apoyar la toma de decisiones en base a la generación de modelos descriptivos y predictivos a partir de datos.

- Esta asignatura cuenta con un aula virtual e-support, la que será fuente oficial de material transversal para la asignatura.
<https://udla.blackboard.com>
- Material se sibirá también a github de la asignatura.

Descripción del programa

Contenido:

1) Introducción al análisis de datos

- Repaso Python y herramientas
- Datos, Limpieza Datos y Pandas.
- Estadística Descriptiva, Predictiva

2) Modelos Analíticos Descriptivos

- Análisis Exploratorio de datos
- Analítica para toma decisiones.

3) Aprendizaje No Supervisado

- Reducción de dimensionalidad
- Análisis de Componentes Principales (PCA)
- Clustering

4) Aprendizaje Supervisado

- Regresión Lineal y Regresión Logística
- Árboles de Decisiones
- Máquinas de Soporte Vectorial (SVM), redes neuronales...etc.

Evaluaciones

En la próxima clase se darán a conocer las fechas

Régimen	Ponderación	Componente	% Componente	Subcomponente	% Subcomponente
TODOS	24	EXAMEN	35	EXAMEN	100
		CATEDRA	50	CATEDRA 1	50
				CATEDRA 2	50
		EJERCICIO	15	EJERCICIO 1	25
				EJERCICIO 2	25
				EJERCICIO 3	25
				EJERCICIO 4	25

Sobre eximición:

El examen de esta asignatura es obligatorio para todo estudiante que la curse, por lo que no contempla eximición bajo ninguna circunstancia.

- El **plagio** y toda acción u omisión que vaya contra la ética, el reglamento y la normativa vigente será sancionada de acuerdo al Reglamento del alumno y normativa Institucional vigente. En particular, el plagio será sancionado con nota mínima 1,0 en la actividad evaluativa. Sin perjuicio de lo anterior y dependiendo de la falta cometida por el estudiante, el caso será evaluado por la Dir. de Escuela y autoridades FINE para determinar si corresponde la reprobación inmediata de la asignatura con nota mínima 1,0 como nota final de la asignatura, de acuerdo a la normativa y reglamento vigente.

Profesora a cargo



Educación:

- Ingeniero en Bioinformática, Universidad de Talca.
- PhD en Ciencias m/ Biofísica y Biología Computacional, Universidad de Valparaíso.

Posiciones:

- 2012-2015 Asistente de Investigación. Centro de Bioinformática y Biología Integrativa, UNAB
- 2017- 2019 Ingeniero en Bioinformática en uBiome Chile. Drug Design Group.
- 2019 -2021 Founder & CEO DrugDevs SpA. Start-up dedicada a Drug Repurposing . TSF-10 (start-up Chile, Corfo).
- 2021 Académico Investigador. Facultad de Ingeniería y Negocios, Universidad de Las Américas.

Pasantías:

- Instituto de Bioinformática Europeo (EBI), grupo de la Dr. Janet Thornton, directora del EBI en Reino Unido. (2 meses)
- Octubre-Diciembre 2016 (2 meses) Grupo de la Dra. Patricia Babbitt., Universidad de California en San Francisco. Tema: Aprendizaje y perfeccionamiento en SSN (Redes de Similitud de Secuencia).

Innovación:

Octubre 2015: Winner of the 2nd Competition R&D – “APLICA TU IDEA”, Santiago, Chile. → Reingeniería de enzimas.
Septiembre 2019 : DrugDevs, ganador TSF10 start-up Chile. (4 meses y posterior extensión) → Drug repurposing.
Junio 2021: Founder Grupo Lidev (Semilla Inicia Corfo) → Algoritmos e IA para generar productos.

Presentación con los estudiantes



Introducción

Análisis de datos.

También llamada analítica de datos, tiene como propósito examinar un conjunto o grandes cantidades de datos con el fin de encontrar tendencia o relaciones sobre determinada información para predecir, tomar decisiones o simplemente realizar un análisis exploratorio sobre un determinado tema.



Introducción



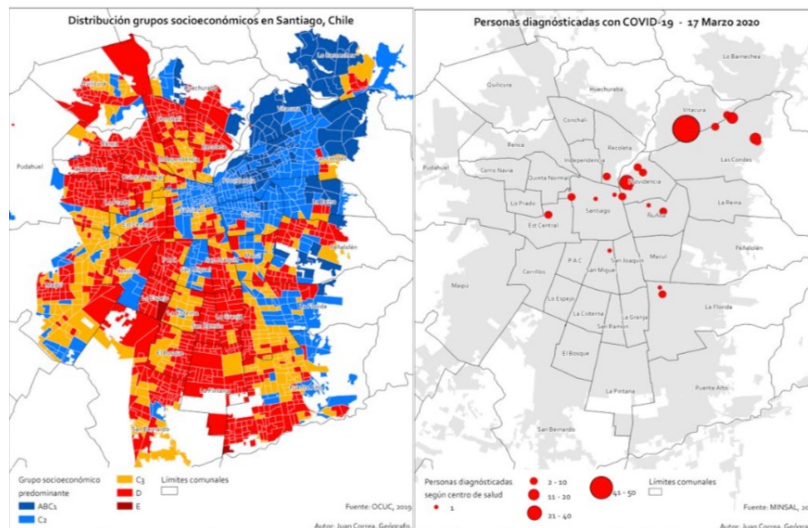
Análisis de Mercado



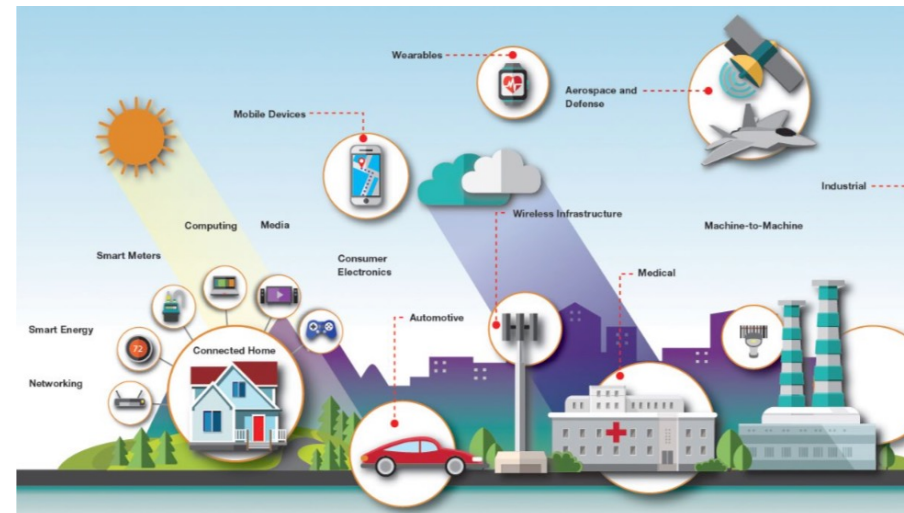
Seguridad y Control



Académicos

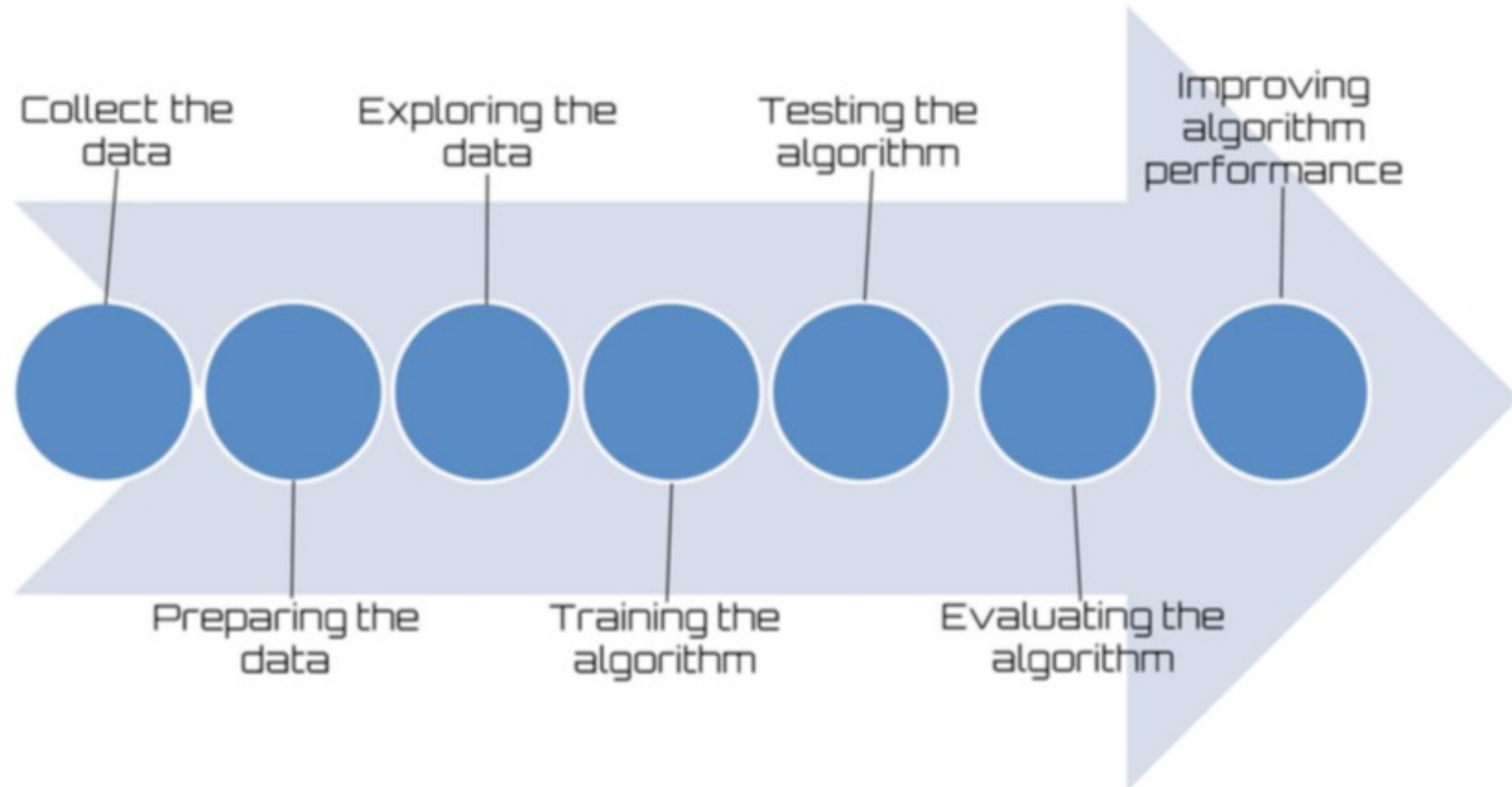


Análisis Georreferenciados



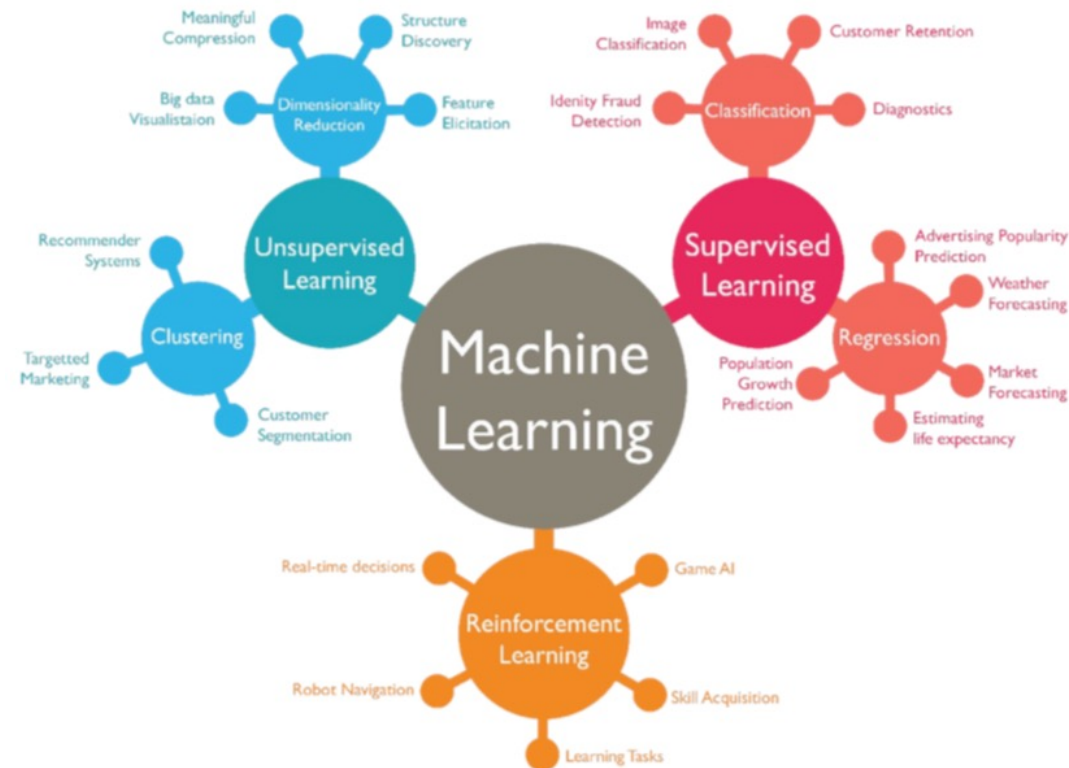
Cuidades Inteligentes y el IoT

Proceso del análisis de Datos



Machine Learning

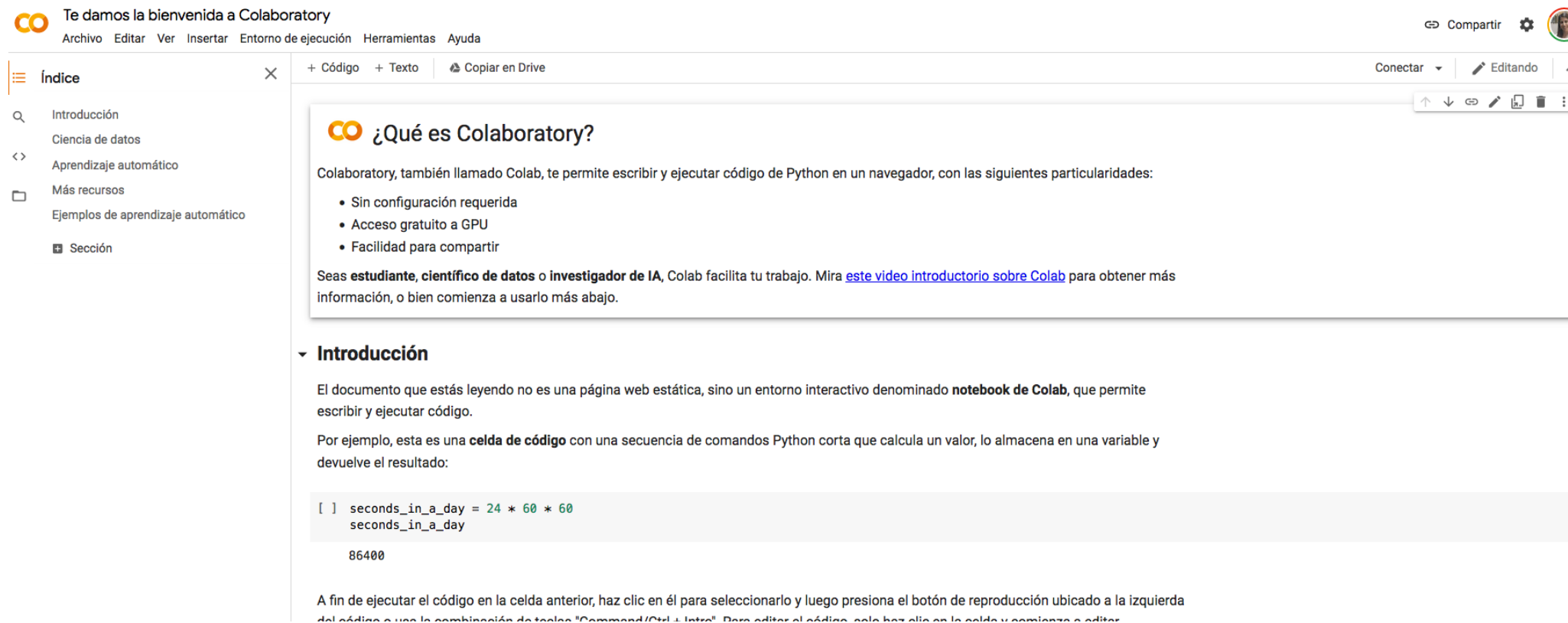
El aprendizaje automático es el estudio científico de algoritmos y modelos estadísticos que los sistemas informáticos utilizan para realizar una tarea específica sin utilizar instrucciones explícitas, basándose en patrones e inferencias. Se considera un subconjunto de la inteligencia artificial



Requerimientos para el curso

Crear una cuenta gratuita en google Colaboratory

<https://colab.research.google.com/notebooks/welcome.ipynb>



The screenshot shows the Google Colaboratory interface. At the top, there's a header with the Colab logo and the text "Te damos la bienvenida a Colaboratory". Below this is a navigation bar with options like "Archivo", "Editar", "Ver", "Insertar", "Entorno de ejecución", "Herramientas", and "Ayuda". On the right side of the header, there are icons for "Compartir", "Configuración", and a user profile. Below the header, there's a sidebar on the left with an "Índice" section containing links to "Introducción", "Ciencia de datos", "Aprendizaje automático", "Más recursos", "Ejemplos de aprendizaje automático", and "Sección". The main content area displays a welcome message titled "¿Qué es Colaboratory?". It explains that Colab allows writing and running Python code in a browser. It lists three key features: no configuration required, free GPU access, and ease of sharing. It also mentions that it's suitable for students, data scientists, and AI researchers, and provides a link to an introductory video. Below this, there's a section titled "Introducción" which explains that the document is an interactive notebook. It gives an example of a code cell that calculates the number of seconds in a day (24 * 60 * 60), resulting in 86400. At the bottom, it instructs the user on how to execute the code by clicking on the code cell and then the play button on the left.

Te damos la bienvenida a Colaboratory

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

Compartir Configuración Perfil

Conectar Editando

Índice

- Introducción
- Ciencia de datos
- Aprendizaje automático
- Más recursos
- Ejemplos de aprendizaje automático
- Sección

¿Qué es Colaboratory?

Colaboratory, también llamado Colab, te permite escribir y ejecutar código de Python en un navegador, con las siguientes particularidades:

- Sin configuración requerida
- Acceso gratuito a GPU
- Facilidad para compartir

Seas **estudiante**, **científico de datos** o **investigador de IA**, Colab facilita tu trabajo. Mira [este video introductorio sobre Colab](#) para obtener más información, o bien comienza a usarlo más abajo.

Introducción

El documento que estás leyendo no es una página web estática, sino un entorno interactivo denominado **notebook de Colab**, que permite escribir y ejecutar código.

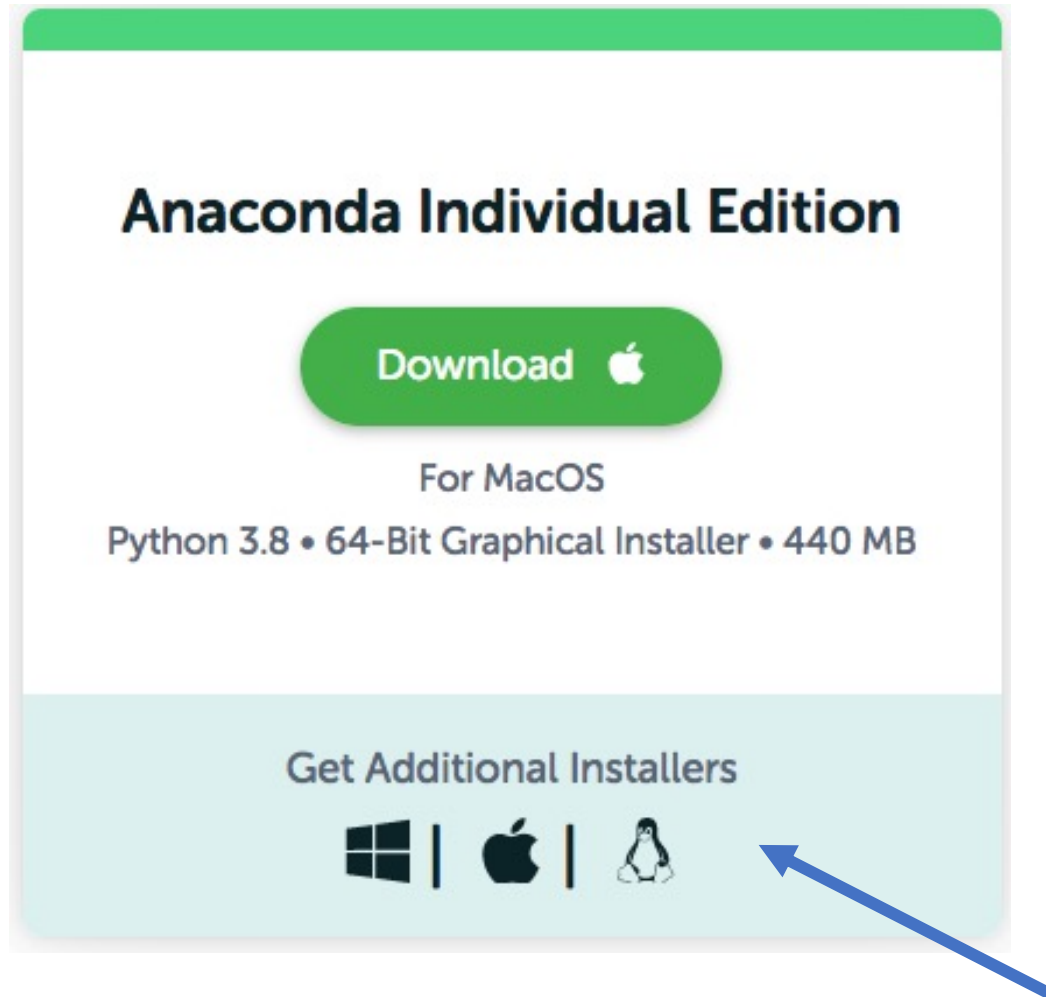
Por ejemplo, esta es una **celda de código** con una secuencia de comandos Python corta que calcula un valor, lo almacena en una variable y devuelve el resultado:

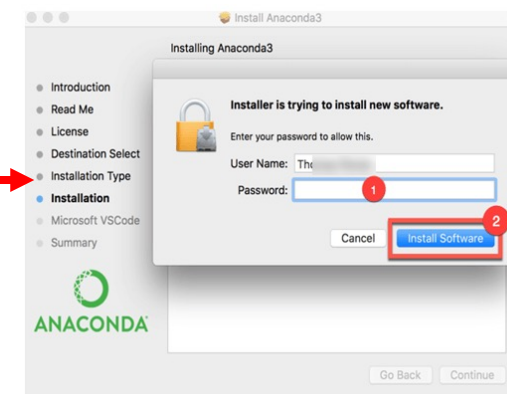
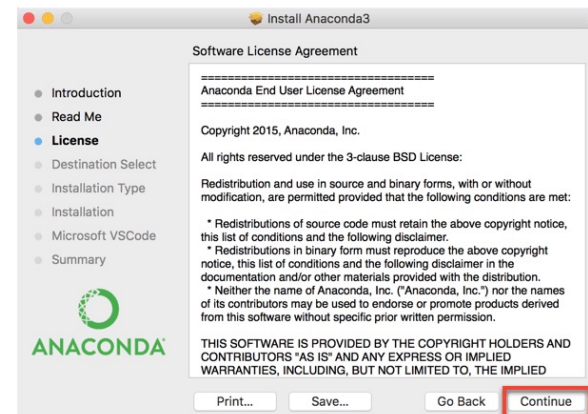
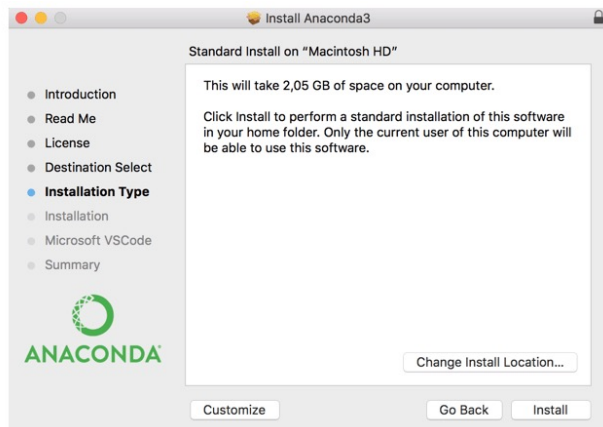
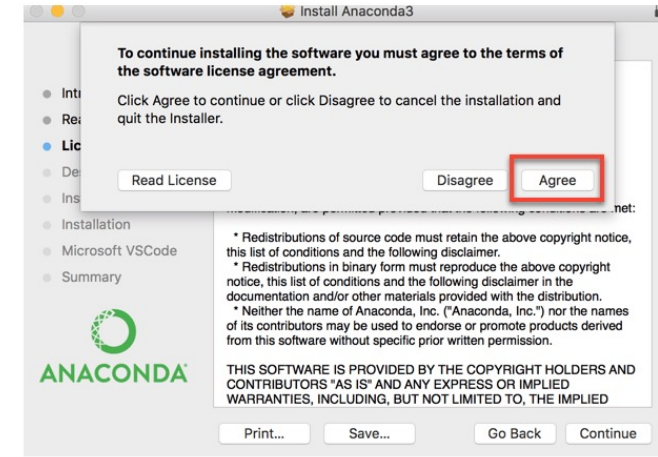
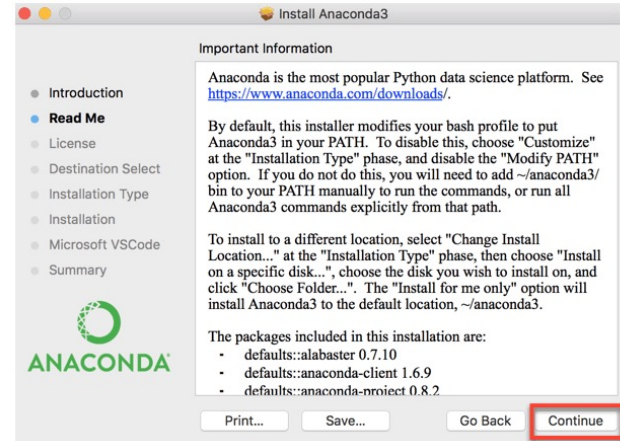
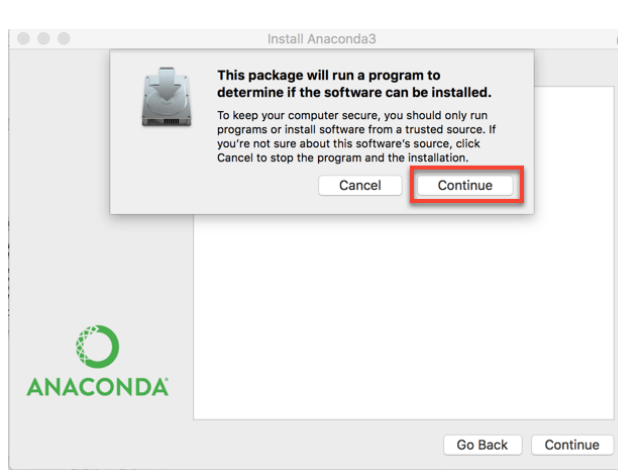
```
[ ] seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day

86400
```

A fin de ejecutar el código en la celda anterior, haz clic en él para seleccionarlo y luego presiona el botón de reproducción ubicado a la izquierda del código o usa la combinación de teclas "Command/Ctrl + Enter". Para editar el código, solo haz clic en la celda y comienza a editar.

Si se desea tener instalado Python de forma local





Windows

****Verificar que su nombre de usuario en Windows no tenga espacios ni caracteres como tildes, comas, apóstrofes o eñes. Esto debido a que se pueden generar errores de codificación.**

Ejemplos:

- `C:\Users\Melissa` : *Correcto*
- `C:\Users\Noé` : **Incorrecto**
- `C:\Users\Melissa Alegría` : **Incorrecto**

El siguiente es instalar la app en nuestro sistema. (Deberá tener permisos de Administrador si instala para todos los usuarios).

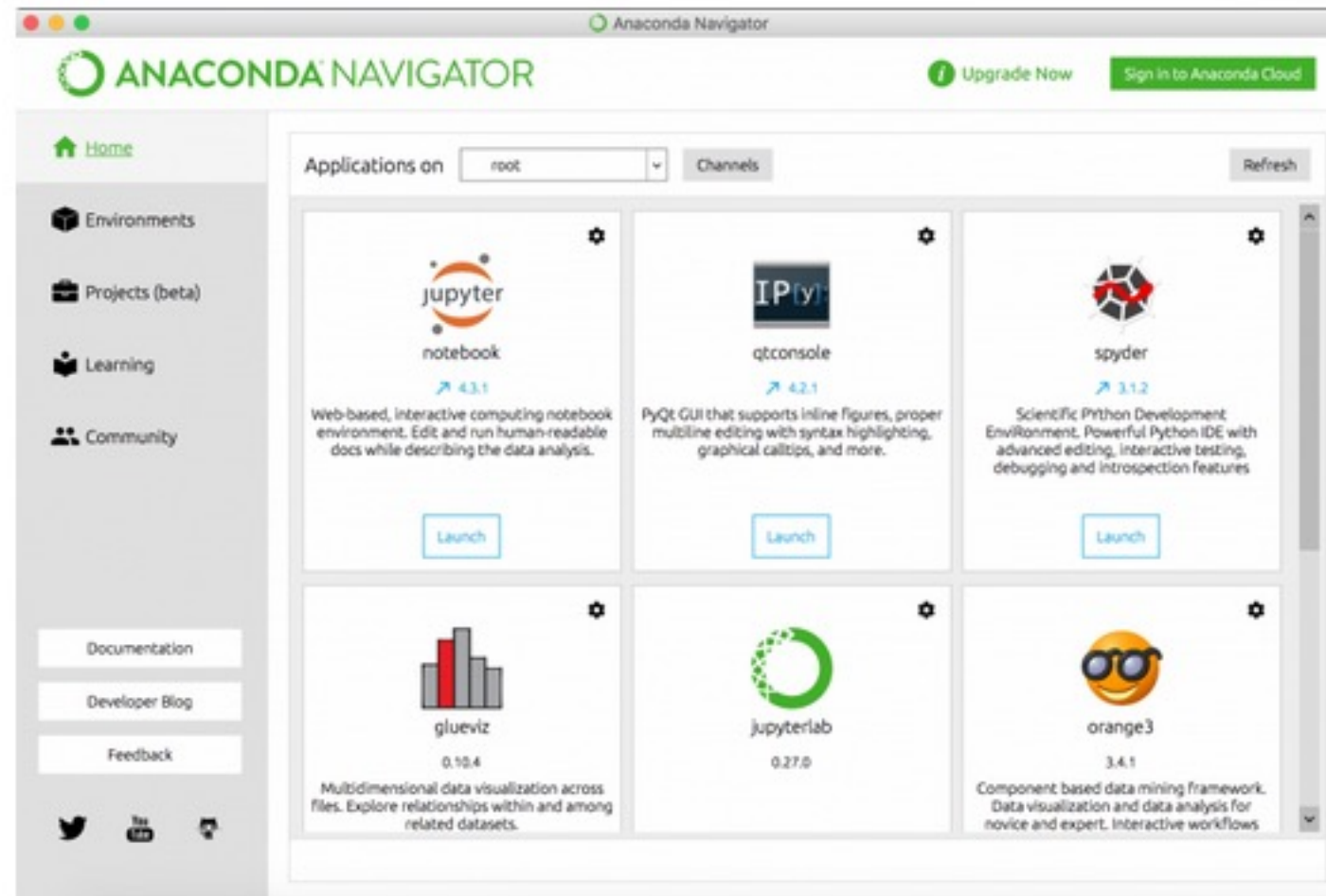
Ejecutamos el archivo que descargamos haciendo doble click.

Se abrirá un “Típico Wizard” de instalación.

Seguiremos los pasos, podemos seleccionar instalación sólo para nuestro usuario, seleccionar la ruta en disco donde instalaremos y listo.

Iniciar y Actualizar Anaconda

Anaconda viene con una suite de herramientas gráficas llamada “Anaconda Navigator”. Cuando la abrimos se podrá visualizar una pantalla como esta:



Requerimientos para el curso

Comprobar la instalación abrimos una Terminal de Mac/Linux/Ubuntu o la Linea de Comandos de Windows.

Escribimos:

>conda -V

```
c(base) MacBook-Pro-de-Melissa:~ melissaalegria$ conda -V
conda 4.10.1
```

>python -V

```
melissaalegria — -bash — 80x24
(base) MacBook-Pro-de-Melissa:~ melissaalegria$ python -V
Python 3.8.5
```

>conda update conda *#debemos poner 'y' para actualizar y se descargarán.*

>conda update anaconda

Requerimientos para el curso

Comprobar la instalación este bien.

<https://raw.githubusercontent.com/jbagnato/machine-learning/master/versiones.py>

```
1 # scipy
2 import scipy
3 print('scipy: %s' % scipy.__version__)
4 # numpy
5 import numpy
6 print('numpy: %s' % numpy.__version__)
7 # matplotlib
8 import matplotlib
9 print('matplotlib: %s' % matplotlib.__version__)
10 # pandas
11 import pandas
12 print('pandas: %s' % pandas.__version__)
13 # statsmodels
14 import statsmodels
15 print('statsmodels: %s' % statsmodels.__version__)
16 # scikit-learn
17 import sklearn
18 print('sklearn: %s' % sklearn.__version__)
```

Luego ejecutamos:

```
>python versiones.py
```

Requerimientos para el curso

Actualizar librería scikit-learn

Para actualizar esta librería muy usada en Machine Learning ejecutaremos los siguientes comandos:

```
> conda install scikit-learn. # instalar scikit-learn
```

```
> conda update scikit-learn #Debemos confirmar la actualización poniendo 'y' en la terminal.
```

Volvemos a ejecutar el script versiones.py

```
> python versiones.py
```

Instalación Jupyter Notebook o Jupyter Lab

Jupyter Notebook es un entorno de trabajo interactivo web que permite desarrollar código en Python de manera dinámica, a la vez que integrar en un mismo documento tanto bloques de código como texto, gráficas o imágenes. Te permite editar y ejecutar documentos de notebook a través de cualquier navegador web, e incluso dentro de algunas IDE.

Instalación y Uso de Jupyter Notebook



<https://jupyter.org/install>

Instalación Jupyter Notebook o Jupyter Lab

<https://jupyter.org/install>

Installation with mamba or conda

JupyterLab can be installed with `mamba` and `conda`:

```
mamba install -c conda-forge jupyterlab
```

or

```
conda install -c conda-forge jupyterlab
```

Note: If you have not installed mamba or conda yet, you can get started with the [miniforge](#) distribution.

Installation with pip

If you use `pip`, you can install it with:

```
pip install jupyterlab
```

If installing using `pip install --user`, you must add the user-level `bin` directory to your `PATH` environment variable in order to launch `jupyter lab`. If you are using a Unix derivative (FreeBSD, GNU / Linux, OS X), you can achieve this by using `export PATH="$HOME/.local/bin:$PATH"` command.

Run JupyterLab

Once installed, launch JupyterLab with:

```
jupyter-lab
```


Instalación Jupyter Notebook o Jupyter Lab

<https://jupyter.org/install>

Getting started with the classic Jupyter Notebook

Installation with mamba or conda

The classic notebook can be installed with **mamba** and **conda** :

```
mamba install -c conda-forge notebook
```

or

```
conda install -c conda-forge notebook
```

Installation with pip

If you use **pip** , you can install it with:

```
pip install notebook
```

Congratulations, you have installed Jupyter Notebook! To run the notebook, run the following command at the Terminal (Mac/Linux) or Command Prompt (Windows):

```
jupyter notebook
```

Verificar la instalación

```
notebooks — -bash — 80x24
(base) MacBook-Pro-de-Melissa:notebooks melissaalegria$ jupyter-notebook
```

[Quit](#)[Logout](#)[Files](#)[Running](#)[Clusters](#)

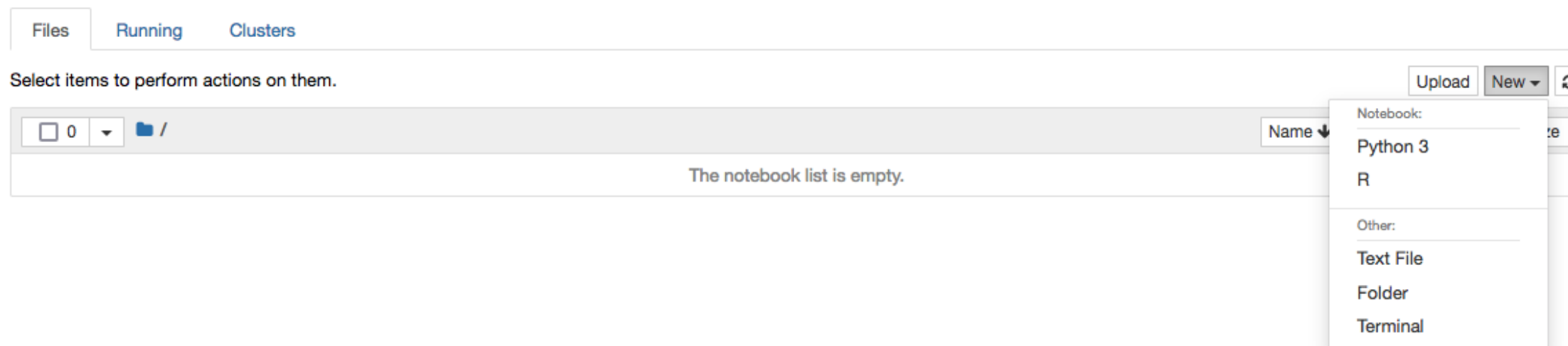
Select items to perform actions on them.

[Upload](#)[New ▾](#)☐ 0 ▾  /[Name ▾](#)[Last Modified](#)[File size](#)

The notebook list is empty.

Verificar la instalación

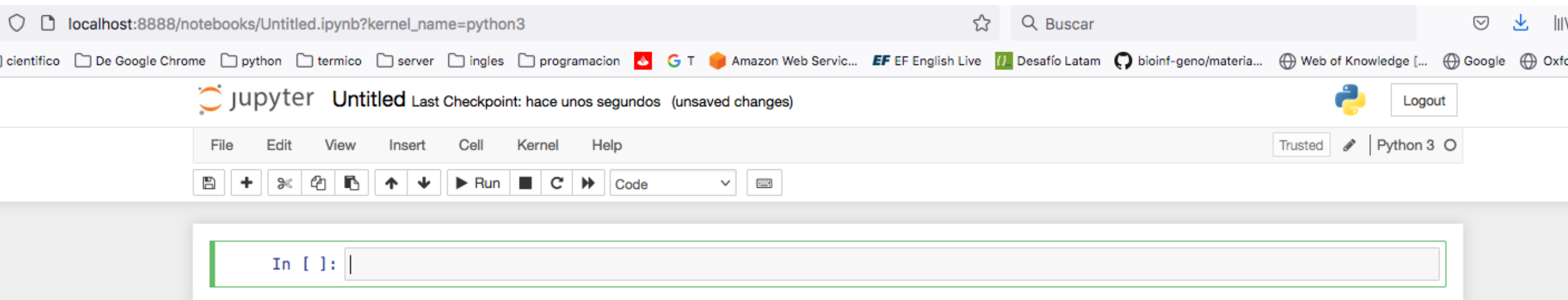
Crear un nuevo proyecto en New--> Python3



The screenshot shows the JupyterLab interface. At the top, there are three tabs: 'Files', 'Running', and 'Clusters'. Below the tabs, a message says 'Select items to perform actions on them.' The main area is a file browser showing a folder icon and a path '/'. A dropdown menu is open, showing options for creating a new project. The 'New' button is highlighted, and the dropdown menu lists 'Notebook:' with 'Python 3' and 'R', and 'Other:' with 'Text File', 'Folder', and 'Terminal'. The text 'The notebook list is empty.' is visible in the background.

Verificar la instalación

Deben visualizar un nuevo notebook:



localhost:8888/notebooks/Untitled.ipynb?kernel_name=python3

Buscar

científico De Google Chrome python termico server ingles programacion

jupyter **Untitled** Last Checkpoint: hace unos segundos (unsaved changes)

Trusted Python 3

File Edit View Insert Cell Kernel Help

Run

In []: |

Verificar la instalación

Ejecutar estos comandos

Clase 1 Análisis de Datos

```
In [1]: print ("Clase 1 de Analisis de Datos")
```

```
Clase 1 de Analisis de Datos
```

```
In [6]: name = input("Cuál es tu nombre:\n")
```

```
Cuál es tu nombre:  
Melissa
```

```
In [7]: print("Hola %s" % name)
```

```
Hola Melissa
```

Verificar la instalación

Ejecutar estos comandos

Operaciones matemáticas

```
In [9]: 0+1
```

```
Out[9]: 1
```

```
In [10]: 1+1
```

```
Out[10]: 2
```

```
In [15]: 50+100
```

```
Out[15]: 150
```

```
In [17]: a = 100  
b = 50  
c = a+b  
print (c)
```

```
150
```

Verificar la instalación

Importando librería matemática

```
In [18]: import math
```

```
In [20]: variable = 60  
c = 100  
d = math.sqrt(100)
```

imprimiendo variables

```
In [22]: print(variable + 100)  
print(d)
```

```
160  
10.0
```

```
In [23]: print (d + c)
```

```
110.0
```

```
In [24]: # la potencia = 2 elevado a 4  
print (2**4)
```

```
16
```

```
In [26]: #dividir  
print(100/60)
```

```
1.6666666666666667
```

Verificar la instalación

```
In [29]: # variables de texto
nombre = 'Carlos'
apellido = 'Santana'
# Interpolación
print("Mi nombre es {} {}".format(nombre, apellido))
```

Mi nombre es Carlos Santana

Verificar la instalación

Verificando instalación de Pandas

Librería orientada a la manipulación y limpieza de estructuras de datos, que nos permite trabajar fácilmente con tablas de datos.

Combina el alto desempeño de operaciones vectorizadas de numpy con las manipulaciones flexibles de hojas de cálculo y bases de datos relacionales.

- Se añade como "import pandas as pd".

```
In [30]: import pandas as pd
```

Dataset con esos tipos de datos: <https://www.kaggle.com/lepchenkov/usedcarscatalog>

```
In [32]: df = pd.read_csv('cars.csv')
df
```

```
Out[32]:
```


Verificar la instalación

	manufacturer_name	model_name	transmission	color	odometer_value	year_produced	engine_fuel	engine_has_gas	engine_type	engine_capacity	...
0	Subaru	Outback	automatic	silver	190000	2010	gasoline	False	gasoline	2.5	...
1	Subaru	Outback	automatic	blue	290000	2002	gasoline	False	gasoline	3.0	...
2	Subaru	Forester	automatic	red	402000	2001	gasoline	False	gasoline	2.5	...
3	Subaru	Impreza	mechanical	blue	10000	1999	gasoline	False	gasoline	3.0	...
4	Subaru	Legacy	automatic	black	280000	2001	gasoline	False	gasoline	2.5	...
...
38526	Chrysler	300	automatic	silver	290000	2000	gasoline	False	gasoline	3.5	...
38527	Chrysler	PT Cruiser	mechanical	blue	321000	2004	diesel	False	diesel	2.2	...
38528	Chrysler	300	automatic	blue	777957	2000	gasoline	False	gasoline	3.5	...
38529	Chrysler	PT Cruiser	mechanical	black	20000	2001	gasoline	False	gasoline	2.0	...
38530	Chrysler	Voyager	automatic	silver	297729	2000	gasoline	False	gasoline	2.4	...

38531 rows x 30 columns

Finalmente, suplementar esto con lo que esta en Github y realizar las actividades que allí se suben

Gracias por su atención
Mail de contacto malegriaa@udla.cl