

Módulo I

CIENCIAS DE DATOS.

**Ciencias de datos
para mentes
inquieta**



Orientaciones para los/las mentores/as



Autoridades

Jefe de Gobierno
Horacio Rodríguez Larreta

Ministra de Educación
María Soledad Acuña

Jefe de Gabinete
Luis Bullrich

Subsecretario de Carrera Docente
y Formación Técnica Profesional
Manuel Vidal

Subsecretaria de Coordinación
Pedagógica y Equidad Educativa
María Lucía Feced Abal

Subsecretaria Agencia de Aprendizaje
a lo Largo de la Vida
Eugenia Cortona

Subsecretario de Gestión Económico
Financiera y Administración de Recursos
Sebastián Tomaghelli

Subsecretario de Tecnología
Educativa y Sustentabilidad
Santiago Andrés



==
Ciencias de datos
para mentes
inquietas

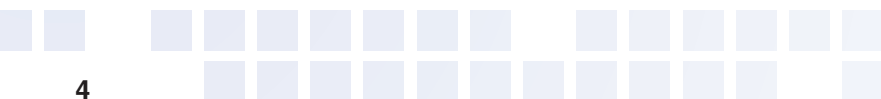
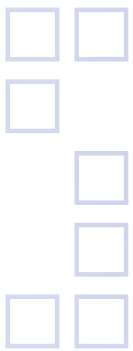
Módulo I

CIENCIAS DE DATOS.

- ● EL ECOSISTEMA
DE LOS DATOS



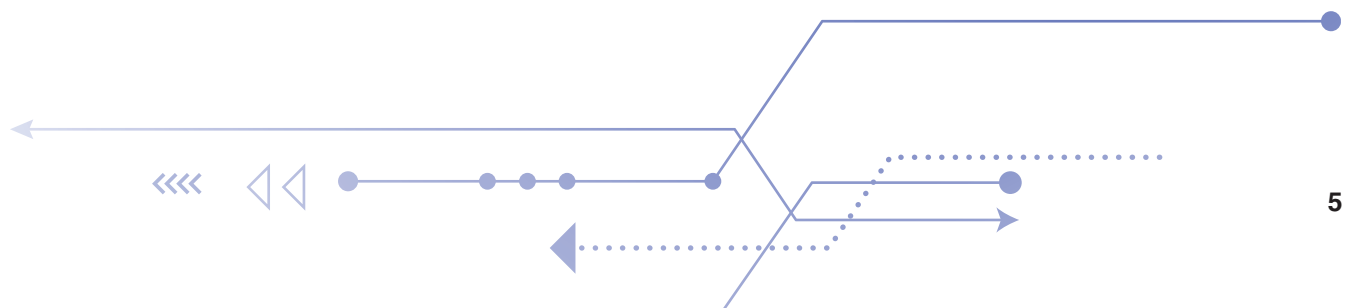
Orientaciones para los/las mentores/as





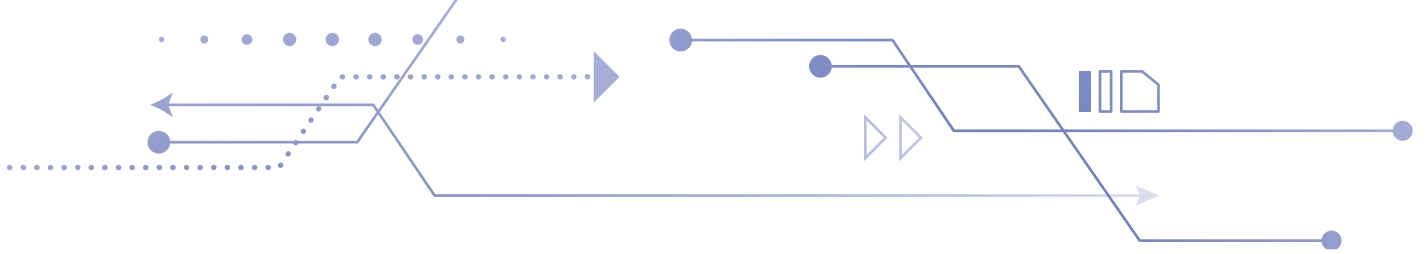
Índice

| | | |
|------------|--|-----------|
| 01. | Sobre el módulo | 7 |
| 02. | Semana 1 | 8 |
| | Objetivos | 8 |
| | Encuentro sincrónico | 9 |
| | Actividad 1: Recuperando saberes previos | 9 |
| | Actividad 2: Conociendo el entorno y la modalidad de trabajo | 9 |
| | Actividad 3: Montando un entorno de trabajo en Python | 10 |
| | Materiales necesarios | 10 |
| | Encuentro asincrónico | 10 |
| 03. | Semana 2 | 10 |
| | Objetivos | 10 |
| | Encuentro sincrónico | 10 |
| | Actividad 1: Aplicaciones de la Ciencia de Datos en la vida real | 10 |
| | Actividad 2: Python: listas, diccionarios, arreglos | 11 |
| | Actividad 3: Presentación del proyecto del módulo I: Ejercicios | 11 |
| | Materiales necesarios | 11 |
| | Encuentro asincrónico | 11 |
| 04. | Semana 3 | 11 |
| | Objetivos | 11 |
| | Encuentro sincrónico | 12 |
| | Actividad 1: Narrando los datos | 12 |
| | Actividad 2: Estructuras de control en programación | 13 |
| | Materiales necesarios | 13 |
| | Encuentro asincrónico | 13 |



Índice

| | | |
|------------|--|-----------|
| 05. | Semana 4 | 14 |
| | Objetivos | 14 |
| | Encuentro sincrónico | 14 |
| | Actividad 1: Bases de datos relacionales: SQL | 14 |
| | Actividad 2: Bases de datos no relacionales: No SQL | 15 |
| | Actividad 3: Introducción a Pandas | 15 |
| | Materiales necesarios | 15 |
| | Encuentro asincrónico | 15 |
| | | |
| 06. | Semana 5 | 16 |
| | Objetivos | 16 |
| | Encuentro sincrónico | 16 |
| | Actividad 1: El juicio a la ciencia de datos | 16 |
| | Actividad 2: Pandas - Filtrar y agrupar | 17 |
| | Materiales necesarios | 17 |
| | Encuentro asincrónico | 17 |
| | | |
| 07. | Semana 6 | 17 |
| | Objetivos | 17 |
| | Encuentro sincrónico | 17 |
| | Actividad 1: Puesta en común del primer módulo | 17 |
| | Actividad 2: Introducción a visualizaciones | 17 |
| | Actividad extra: Introducción al Pensamiento Computacional | 17 |
| | Materiales necesarios | 18 |
| | Encuentro asincrónico | 18 |

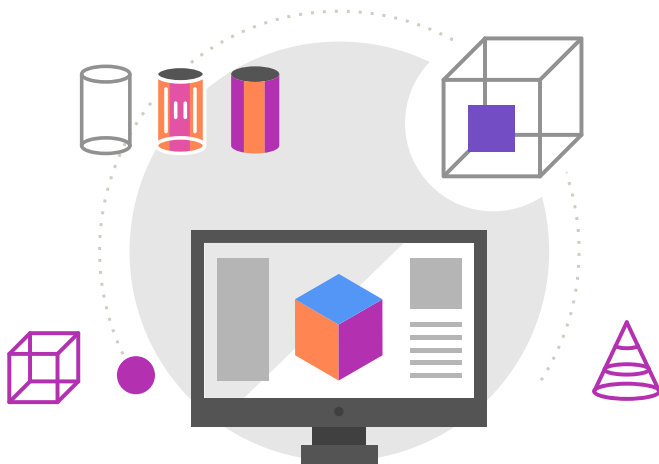


Sobre el módulo

Hoy más que nunca estamos más conectados con personas y dispositivos, tenemos acceso a más redes y servicios, y sin duda consumimos y producimos mayores cantidades de datos e información. Por lo que requerimos contar con las habilidades, conocimientos, experiencias y técnicas de los científicos de datos para procesar, analizar y visualizar de formas más inteligentes los datos en información, promoviendo así, más y mejores conocimientos de nuestra realidad en sus contextos.

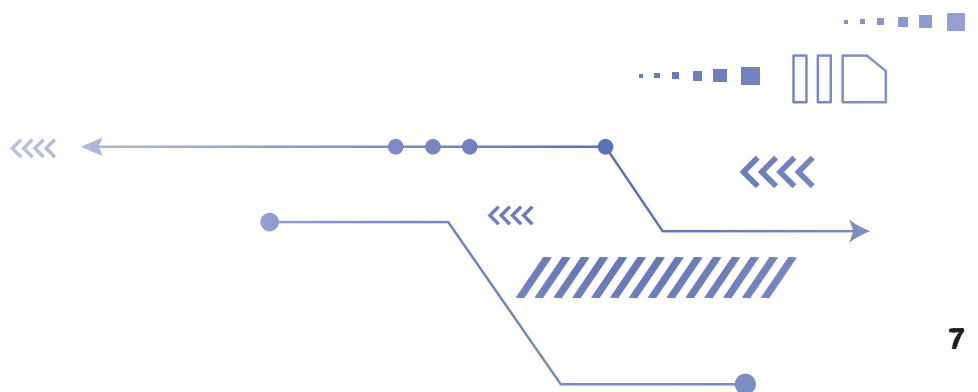
Este proyecto introduce a los/las estudiantes en los conceptos básicos para comprender el objetivo de realizar el análisis de datos y comenzar a realizar sus primeros modelados comprendiendo de qué modo un analista puede trabajar con un modelo tomado de la realidad.

El propósito final es despertar en los/las estudiantes el interés por el valor oculto e intangible de los datos y la programación orientada a datos. Comprender el ecosistema de datos abiertos, sin el cual es imposible pasar de la publicación a la generación de valor social.



Este primer proyecto permite a los/las estudiantes poder responder las preguntas más usuales que las personas se hacen al hablar de datos:

¿Qué son los datos? ¿Para qué se utilizan y cómo? ¿Para qué buscar relaciones entre los datos y cómo empezar a modelar las mismas? ¿Cómo se logra esto mediante la programación?



Saberes y competencias que se pondrán en juego:

Interpretación y abstracción de datos.

Transformación de datos en información.

Modelado y relación de datos.

Selección, análisis e interpretación de datos.

Producción creativa singular y plural.

Valoración y participación en la construcción de saberes. Se comparte y produce en colaboración con otros, conocimientos, ideas y creaciones, de diversa índole. Construcción de argumentos basados en razonamientos coherentes y lógicos.

Eje conceptual:

El mundo conectado: donde se generan los datos. - Ciencia de datos vs. analítica de datos e inteligencia artificial.- Datos e información - *Python* para científicos de datos.

La evaluación será continua en lo individual valorando lo actitudinal, procedimental y cognoscitivo en los/las estudiantes mediante matrices de evaluación. Se acompañará a los/las estudiantes desde lo virtual sincrónica y asincrónicamente detectando las dificultades en el proceso de adquisición de conocimientos y las habilidades para el trabajo de campo supervisando las fuentes de los/las estudiantes.

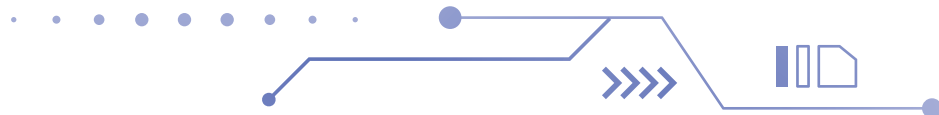
01. Organización de la propuesta de enseñanza por semana



Semana 1

Objetivos

- Conocer a los/las estudiantes
- Realizar un diagnóstico grupal sobre conocimientos previos e intereses
- Dar un panorama integral del trayecto de ciencia de datos



- Presentar la plataforma virtual
- Presentar *Google Colab* para trabajar con *Python* (primer acercamiento Hola Mundo)

Actividad 1

Recuperando saberes previos

Encuentro sincrónico

En este encuentro, los/las estudiantes se presentarán diciendo su nombre, año de la secundaria, de qué escuela son y por qué les dieron ganas de inscribirse en este curso.

Se realizará una indagación sobre qué pre-conceptos y conocimientos previos traen los/las estudiantes sobre el mundo de los datos y se irá confeccionando un listado de ejemplos de su uso en la vida cotidiana.

Preguntas sugeridas:

¿Qué conocen de ciencia de datos y de bases de datos?

¿De qué hablamos cuando hablamos de ciencia de datos?

¿Y cuando hablamos de análisis de datos?

¿Cuáles son las diferencias entre ciencia de datos, análisis de datos e inteligencia artificial?

¿Qué es un dato? ¿Cuál es la diferencia entre dato, información, hecho y evidencia?

Se hará una revisión de la lista a la luz de todas las participaciones y de lo conversado. Luego se elaborará con los/las estudiantes un cuadro con las diferencias más importantes entre ciencias de datos, análisis de datos, *machine learning*, *deep learning* e inteligencia artificial.

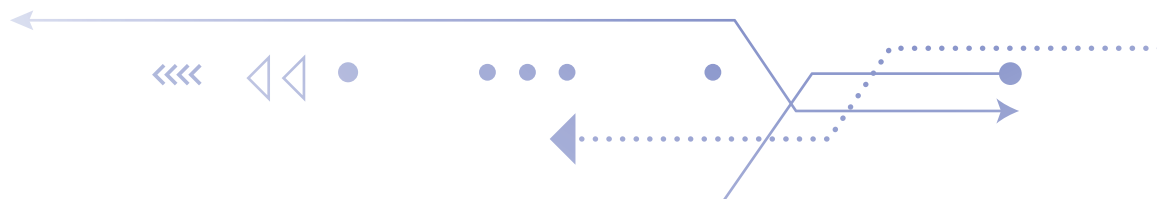
Actividad 2

Conociendo el entorno y la modalidad de trabajo

Terminada la actividad de presentación de los/las estudiantes, se describe el trayecto del año, la metodología que se aplicará (aprendizaje basado en problemas y proyectos) y la modalidad de evaluación.

Luego de todo esto se procederá a la presentación y navegación de la plataforma virtual, y se darán las pautas para el trabajo asincrónico.

Por último se comentará cómo va a funcionar el espacio de tutorías asincrónicas (*Open Lab*) y decirles que habrá algunas tareas de entre semana para que vayan investigando y que es importante poder realizar, ya que forman parte de los contenidos curriculares del curso.



Actividad 3

Montando un entorno de trabajo en Python

Presentación de *colab*. Cómo ingresar, cómo cargar la notebook 1, y empezar a trabajar con los comandos básicos de *Python*. Contar la historia de “Hola Mundo” para todos y todas.

Materiales necesarios:

- > Plataforma virtual abierta y con ejemplos claros de los espacios (navegar antes por la misma para poder responder preguntas que puedan llegar a surgir)
- > Notebook 1, intro a Colab y bases de Python (correrla antes, fijarse que esté todo bien)
- > Espacio online para registrar la lluvia de ideas ([Miro](#))
- > Anotador personal donde ir registrando los nombres y datos de los/las estudiantes
- > PPT Presentación Ciencia de Datos

Actividad

Nuestros primeros comandos en Python

Encuentro asincrónico

- Seguir completando la notebook de inicio a *Python*.
- *Investigar sobre casos que conozcan de aplicación de Ciencia de Datos en la vida cotidiana.*



Semana 2

Objetivos

- Entender diferentes usos y aplicaciones de la ciencia de datos en el mundo real que tengan impacto positivo en la vida de las personas
- Avanzar en el aprendizaje programación en Python: listas, diccionarios, arreglos, tipos de datos

Actividad 1

Aplicaciones de la Ciencia de Datos en la vida real

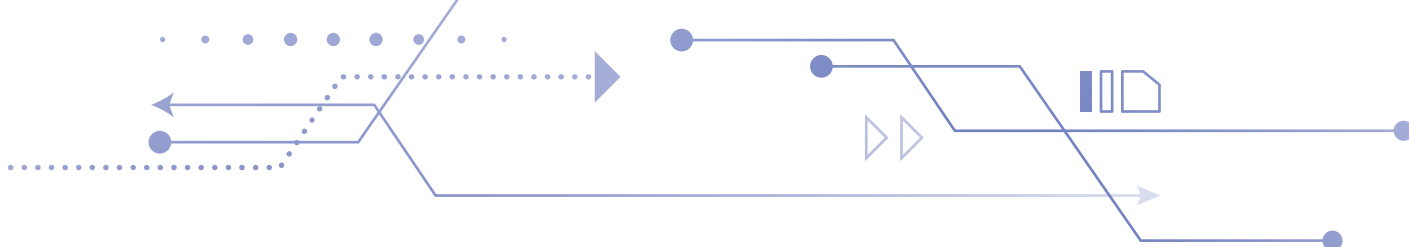
Encuentro sincrónico

La idea de esta actividad será trabajar sobre aplicaciones y casos de éxito de la ciencia de datos: bioinformática, investigación espacial, clima, aeronáutica, juegos, robots que aprenden, medicina. La idea es derribar preconceptos.

Mostrando diversos ejemplos en forma de presentación iremos debatiendo e investigando sobre los mismos y pondremos a cada ejemplo asociado con la forma de aprender y predecir que se utiliza (redes neuronales, regresiones, árboles de decisión, etc.).

Resultado: un documento colaborativo con diversas aplicaciones y grupos que trabajan con ciencia de datos desde diversas disciplinas.





Actividad 2

Python: listas, diccionarios, arreglos

Se comenzará a explicar la notebook 2 de *Python*: listas, conjuntos, diccionarios y tuplas.

Actividad 3

Presentación del Proyecto del Módulo: Ejercicios

Presentación de Proyecto de Módulo: Ejercicios

Se mostrará a los/las estudiantes el Proyecto de Módulo que consta de una Guía de Ejercicios que deberán realizar clase a clase en el espacio asincrónico para practicar lo visto en el encuentro sincrónico. Asimismo se mostrará la guía de Ejercicios optativa para los/las estudiantes que deseen realizar más ejercicios de práctica.

Materiales necesarios:

- > PPT Casos de éxito
- > Notebook 2 - Tipos de datos
- > Notebook Proyecto de Módulo
- > Notebook Ejercicios optativos

Encuentro asincrónico

Seguiremos trabajando con la notebook de listas, diccionarios y arreglos, que explica cada una de las diferencias entre estas estructuras. Lo que se propone es que los/las estudiantes recorran la notebook 2 y resuelvan los ejercicios al final.

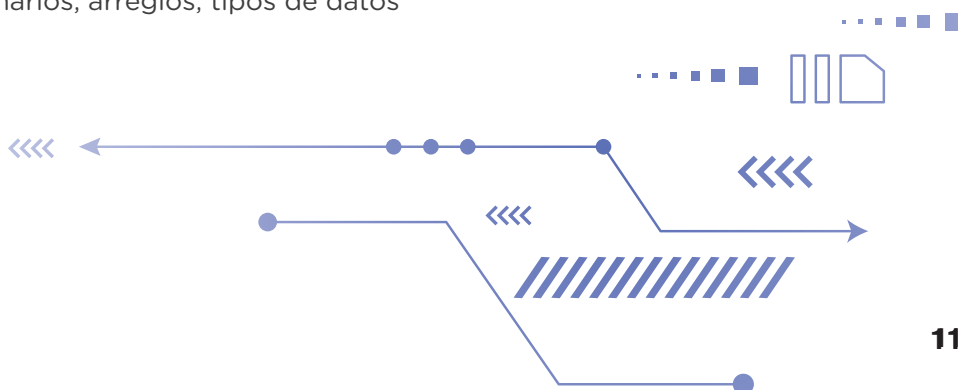
Asimismo deberán realizar los Ejercicios correspondientes del Proyecto del Módulo y plantear las dudas.



Semana 3

Objetivos

- Introducir el concepto de storytelling en ciencia de datos
- Trabajar con booleanos y estructuras de control en Python
- Despejar dudas sobre programación en *Python*: listas, diccionarios, arreglos, tipos de datos



Actividad 1

Narrando los datos

Encuentro sincrónico

En esta clase vamos a ver el concepto de *storytelling* como narración de datos. Esta es una metodología para comunicar información, adaptada a una audiencia específica, con una narrativa convincente. Son los últimos pasos de un análisis de datos y posiblemente el aspecto más importante, ya que es el que logra tener cierto impacto si lo hacemos bien.

Como seres humanos, históricamente nos juntamos alrededor del fuego para compartir historias. Hay muchas teorías que sugieren incluso que la narración fue la plataforma de lanzamiento principal para la transmisión de conocimiento entre grandes grupos de personas, lo que formó las culturas tal como las conocemos hoy y permitió que se transmitan a través de generaciones.

Ahora, con tantos datos disponibles, sólo la narración de datos puede poner una perspectiva humana en el mundo cada vez más complejo y rápidamente cambiante de la era digital.

Es por eso que un proyecto no debe ser un simple reporte o informe de los resultados obtenidos, sino un medio para contar una historia con datos, un medio para poder interpelar a la audiencia.

Vamos a trabajar sobre la creatividad y veremos diferentes plataformas para poder realizar una presentación *online* (pueden ser sugeridas por los/las estudiantes).

Contaremos que la narración de datos se compone de tres partes:



Ciencia de datos: este campo de especialización es el campo interdisciplinario de las ciencias, que extrae conocimientos de los datos, haciéndolos fácilmente disponibles. Este apasionante campo ha realizado cambios significativos en nuestra vida diaria en las últimas dos décadas.

Los científicos de datos a menudo son expertos en recopilar y entregar datos, pero carecen de las habilidades para transmitir una verdadera comprensión de las oportunidades ocultas en los datos entregados.



Visualizaciones: la transformación de datos en gráficos implica que podamos visualizar nuestros datos como nunca antes. Sin embargo, las visualizaciones de datos por sí solas tienen limitaciones. Proporcionan una mirada superficial de los datos, sin el contexto necesario para explicar qué hay detrás.



Narrativa: la tercera parte y, en cierto modo, la más vital de una historia de datos es la narrativa. La narrativa usa el lenguaje en un formato que se adapta a nuestras necesidades particulares, aumentando nuestra comprensión completa de la nueva información. Una narrativa es un vehículo clave para transmitir conocimientos, y las visualizaciones y los datos son puntos que apoyan a la narrativa dándole veracidad.

Actividad 2

Estructuras de control en programación

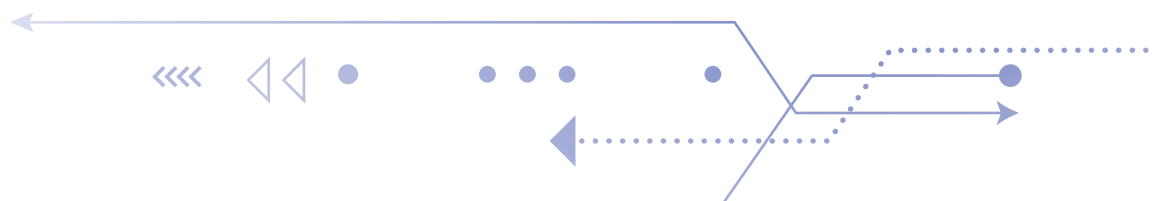
Se continuará con *Python*, esta vez se introducirá la idea de *booleanos* y de estructuras de control condicional mediante la notebook. Considerando la complejidad de este tema, es aconsejable trabajar de a poco y a una velocidad que les resulte cómodo a los/las estudiantes. Si es necesario, se puede continuar trabajando en el aprendizaje de condicionales en el momento asincrónico de la clase.

Materiales necesarios:

- > PPT de *storytelling*
- > Notebook 3 - Estructuras de control

Encuentro asincrónico

La propuesta es seguir trabajando sobre el tema que vimos en la clase y que los/las estudiantes planteen las dudas que surgan. Deberán completar los Ejercicios del Proyecto correspondientes a estructuras de control.





Semana 4

Objetivos

- Aprender sobre los sets de datos
- Responder las siguientes preguntas: ¿Qué son las bases de datos? ¿Para qué sirven?
- Conocer *SQL* y *NoSQL*
- Conocer la Librería Pandas

Actividad 1

Bases de datos
relacionales: *SQL*

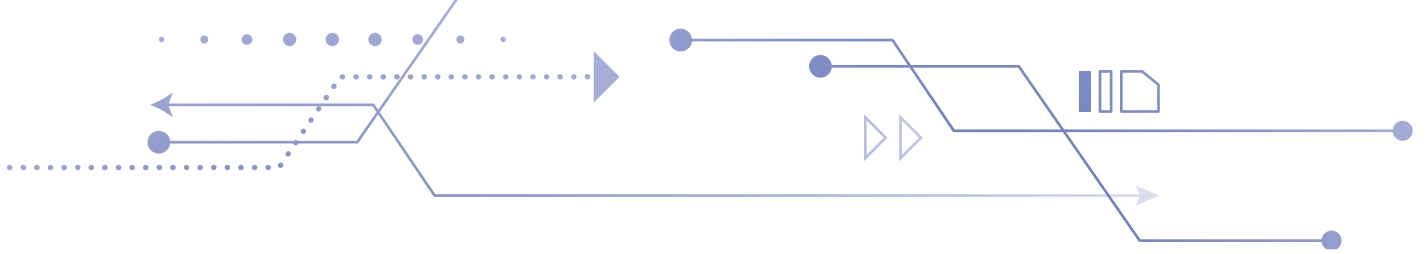
Encuentro sincrónico

¿Qué vamos a enseñar de *SQL* y *NoSQL*?

Primero partiremos explicando qué tipo de datos podemos levantar desde una notebook, hablaremos de: *csv*, *de xls*, *json*, etc. Pero que estos archivos en sí no son bases de datos, para que haya una base de datos tiene que haber un modelo relacional o no relacional. Diremos que *SQL* es un lenguaje de dominio específico utilizado para comunicarse con sistemas de bases de datos relacionales.

La idea es dar un
panorama general
sobre la forma
de almacenar y
consultar datos en
bases relacionales
y no relacionales.





Actividad 2

Bases de datos
no relacionales:
No SQL

Se mostrará *MongoDB* como ejemplo de bases no relacionales. El modelado de datos en las bases de datos *NoSQL* es un área de reciente estudio y no tiene aún un soporte teórico como sí lo tienen las bases de datos relacionales.

Se puede hablar sobre la idea de colecciones y de diferentes tipos de organización de los datos mostrando ejemplos sencillos.

El sentido es despertar la curiosidad de los/las estudiantes para que sigan explorando por su cuenta estas herramientas, pero no serán las centrales del curso.

Actividad 3

Introducción a Pandas

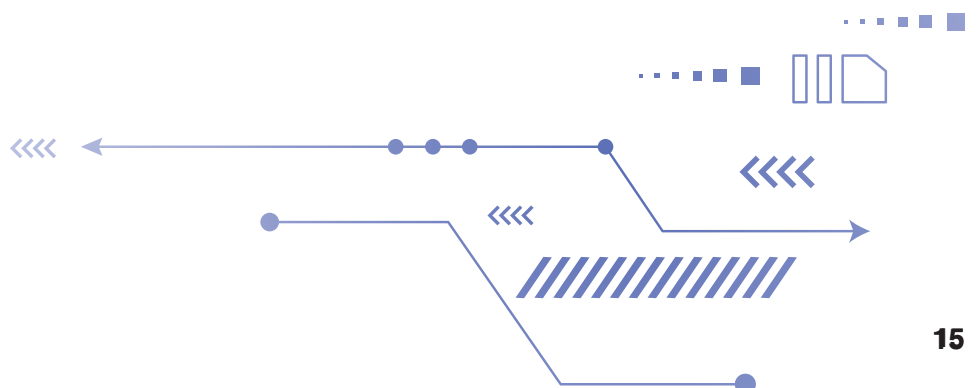
En la segunda parte de este encuentro se trabaará con una Notebook de introducción a la librería Pandas utilizando un dataset de encuentros educativos ambientales extraídos de la web del Gobierno de la Ciudad de Buenos Aires. Es un dataset pequeño y sencillo para poder comenzar a conocer como explorar un dataset con Python. El objetivo es explorar los valores de cada columna para entender mejor cómo están conformadas y procederemos a documentar cada columna con sus posibles valores en un archivo Excel.

Materiales necesarios:

- > PPT Bases de datos
- > Notebook Introducción a Pandas
- > Excel de Metadata

Encuentro asincrónico

En esta actividad los/las estudiantes realizaran los Ejercicios del Proyecto correspondientes a la introducción a Pandas y consultaran con el/la mentor/a las dudas que surgan.





Semana 5

Objetivos

- Seguir profundizando conocimientos de *Python*
- Reconocer la importancia del trabajo con datos y los riesgos de trabajar desde un lugar irresponsable con los mismos
- Identificar problemas de sesgos e inteligencia artificial
- Comprender el concepto de ética de datos
- Reflexionar colectivamente sobre los diversos problemas vinculados a la privacidad
- Ejercer Ciudadanía Digital

Actividad 1

El juicio a la ciencia de datos

Encuentro sincrónico

Parte 1: Investigación

Nos repartiremos en subgrupos para conformar mesas de trabajo que tendrán 15 minutos para investigar sobre diferentes problemas reales vinculados a sesgos en la ciencia de datos (el modelo de contratación de *Amazon* que sólo contrataba hombres, el uso de inteligencia artificial para decidir sentencias, reconocimiento facial que no detecta rostros de personas afrodescendientes, entre otras).

Parte 2: Armado de argumentos

Una vez que terminen la pequeña investigación y tengan un poco más de información al respecto se dividirá al grupo en dos sub-grupos que deberán organizarse para ser las dos partes de un juicio sobre el uso de datos personales para la ciencia de datos. Un grupo deberá elaborar argumentos a favor y el otro argumentos en contra.

La idea es que los argumentos sean lógicos y fundamentados. Cada grupo tendrá 20 minutos para elaborar su estrategia.

El objetivo de esta actividad no es que haya un equipo ganador y otro perdedor, sino mostrar la complejidad a la que nos enfrentamos a la hora de pensar en estos temas. Por un lado, pueden ser técnicas que ayudan mucho a la humanidad y por otro, pueden resultar mecanismos de control y de exposición de ciertas personas según los intereses de quien tiene el poder de los datos.

Parte 3: El juicio

Se realizarán las presentaciones orales en defensa o en contra de la utilización de datos personales en ciencia de datos. El/la mentor/a hará de jurado y deberá moderar el debate.



Actividad 2

Pandas: filtrar y agrupar

Luego de la exposición y del debate se comenzará a explicar la notebook sobre como se realizan filtros y agrupaciones con la librería Pandas, continuaremos utilizando el dataset de Encuentros educativos ambientales.

Materiales necesarios:

- > PPT Sesgos y peligros de la inteligencia artificial
- > Notebook 5 - Pandas: filtrar y agrupar

Encuentro asincrónico

La idea de este encuentro es seguir trabajando sobre lo que hicimos en clase,. Deberán completar todos los ejercicios del proyecto y plantear las dudas que surjan.



Semana 6

Objetivos

- Realizar feedback a los/las compañeros/as en este primer módulo
- Hacer un breve resumen y un repaso de lo aprendido hasta el momento
- Relevar qué conceptos hay que reforzar y practicar con más ejemplos concretos
- Realizar una introducción a Visualizaciones
- Introducir el concepto de Pensamiento Computacional

Actividad 1

Repaso y feedback sobre lo visto hasta el momento

Encuentro sincrónico

Se realizará un intercambio con los/las estudiantes sobre lo visto hasta el momento y un cierre de este primer módulo (puede utilizarle como soporte una encuesta de Zoom).

Se realizará un repaso de lo aprendido con la plataforma Kahoot para poder observar los conocimientos incorporados y lo que sea necesario repasar.

Actividad 2

Introducción a visualizaciones

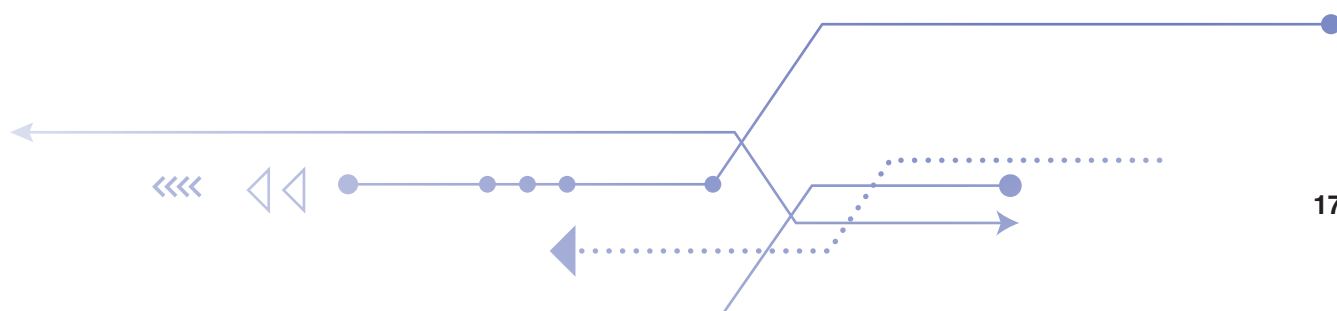
Utilizando una Notebook se mostrará una introducción a visualizaciones con las librerías Seaborn y Matplotlib.

Actividad extra

Introducción al Pensamiento Computacional y sus habilidades

Comenzaremos a explicar el concepto de Pensamiento Computacional y sus diferentes habilidades:

- Algoritmos
- Descomposición en subproblemas
- Representación de la información
- Búsqueda de patrones



Materiales necesarios:

- > Kahoot para realizar repaso
- > Notebook Visualizaciones
- > PPT Pensamiento Computacional

Actividad

Entrega final de los
proyectos

Encuentro asincrónico

Aprovecharemos la última clase asincrónica para resolver cualquier problema o duda de las y los estudiantes para finalizar el proyecto del módulo. Una vez terminado, subiremos todos los proyectos a la plataforma virtual.