

TALLER CIENCIA DE DATOS



IDATHA



SAP next-gen »

HERRAMIENTAS



Jupyter Notebooks

1. Es un ambiente interactivo, en el cual podemos combinar código (junto con su ejecución y resultados), texto enriquecido, tablas, entre otros.
2. Por ello, lo utilizaremos para algunos de los contenidos del curso. Además, los estudiantes deberán trabajar en grupos en un proyecto final, pudiendo entregar parte de los resultados en formato Jupyter Notebook.



Google Colab

1. Colaboratory es un entorno de notebook de Jupyter gratuito que no requiere configuración y se ejecuta completamente en la nube.
2. Con Colaboratory, puedes escribir y ejecutar código, guardar y compartir análisis, y acceder a recursos informáticos potentes, todo gratis en tu navegador.



Anaconda

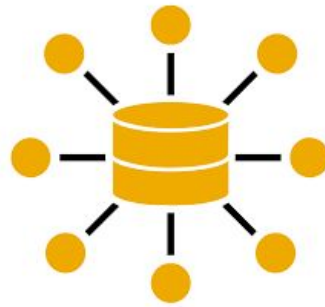
1. Esencialmente es una distribución libre y abierta de los lenguajes Python y R, junto con herramientas que facilitan la configuración y despliegue de un ambiente virtual con todas sus dependencias (librerías, etc.).
2. Es ampliamente utilizado en el área de la ciencia de datos y además es compatible con Linux y Windows.
3. Podemos utilizarlo como gestor para la instalación de Python/R, Jupyter Notebooks y todas las librerías que vayamos a necesitar.



SAP Services

SAP cuenta con varias herramientas para el procesamiento de datos en el contexto de DS:

1. SAP Data Hub
<https://www.sap.com/products/data-hub.html>
2. SAP Leonardo Machine Learning Foundation
<https://www.sap.com/products/machine-learning-foundation.html>
3. SAP Data Intelligence
<https://www.sap.com/products/data-intelligence.html>



PROBLEMAS...



ANÁLISIS DATOS COMPRAS



Los datos de compras estatales contienen información de todas las compras y adjudicaciones realizadas por organismos dependientes del estado. Se puede revisar casos que incumplan con el TOCAF, agregando datos por Inciso/Unidad Ejecutora, investigar tipos de productos o rubros en los que más dinero se gasta, etc.

Links:

- <https://www.gub.uy/agencia-compras-contrataciones-estado/>
- <https://catalogodatos.gub.uy/dataset/acce-compras-estatales>
- <https://catalogodatos.gub.uy/dataset/acce-datos-de-compras-en-linea>

Ejemplos interesantes de uso de estos datos:

- <https://transparenciapresupuestaria.opp.gub.uy/inicio/presupuesto-nacional/a-d%C3%B3nde-van-nuestros-impuestos>
- <http://cuentasclaras.uy/#/home>

ENCUESTAS CONTINUAS DE HOGARES



Los datos de las encuestas continuas de hogares contienen microdatos referentes a niveles de ingresos, educación, calidad de vivienda, entre otros, de una muestra representativa de la sociedad.

En particular en el catálogo de datos abiertos, están los datos de encuestas para los últimos siete años, con los que se pueden realizar una gran variedad de análisis e incluso cruzar con otros data sets de su interés.

Links:

- <https://catalogodatos.gub.uy/dataset?q=encuesta+continua>
- <https://catalogodatos.gub.uy/dataset/ine-encuesta-continua-de-hogares-ano-2017>
- <http://www.ine.gub.uy/encuesta-continua-de-hogares1>

INFRACCIONES DE TRÁNSITO



También en el catálogo de datos abiertos, se cuenta con datos de infracciones de tránsito levantadas en particular por la IMM. Hay datos disponibles para los últimos diez años aproximadamente.

Puede ser muy interesante cruzar estos datos con datos de siniestralidad tanto de la IMM como de otros organismos, como la UNASEV. ¿Están relacionadas las multas por infracciones con los siniestros?

Links:

- <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-multas-transito>
- https://catalogodatos.gub.uy/dataset/unasev-fallecidos_siniestros_transito

ACCIDENTES DE TRÁNSITO



Del catálogo de datos abiertos, se puede extraer datos de accidentes de tránsito para la ciudad de Montevideo. A su vez, otros organismos como la UNASEV y el Mides también publican datos estadísticos sobre siniestralidad. ¿Qué patrones se pueden identificar? ¿Influye la edad, sexo o tipo de vehículo en la cantidad de siniestros? ¿Existe una relación entre las campañas de la UNASEV/IMM y el aumento/disminución? ¿Qué pasa con las multas de tránsito, están relacionadas a los accidentes?

Links:

- https://catalogodatos.gub.uy/dataset/unasev-fallecidos_siniestros_transito
- https://catalogodatos.gub.uy/dataset/unasev-fallecidos_siniestros_transito
- <https://catalogodatos.gub.uy/dataset/mides-indicador-10488>

ANÁLISIS DATOS AMAZONAS



Desde que cobró dominio público la situación ambiental en la selva Amazonas a raíz de los incendios forestales de los últimos meses, predominan los análisis y visualizaciones sobre el estado actual de la selva, los focos activos y la deforestación.

Puede resultar interesante analizar algunos de los datos disponibles como los datos abiertos de Brasil, NASA, entre otros y cruzar estos datos para extraer información sobre la situación actual en la Amazonia. Existe una correlación entre la deforestación y las exportaciones de Brasil. Por ejemplo qué actividades económicas se ven beneficiadas por esto y de cuánto dinero estamos hablando. ¿Qué sucede con las personas desplazadas? ¿Y el impacto ambiental?

Links:

- <https://www.kaggle.com/gustavomodelli/forest-fires-in-brazil>
- <https://infoamazonia.org/en/data/>
- <https://colombia.infoamazonia.org/>

REAL ESTATE



Un problema cotidiano y muy interesante de analizar es el mercado inmobiliario. ¿Cómo está el precio del m2 actualmente? ¿En qué barrio vale más? ¿Varía según tipo de propiedad, tecnología constructiva, etc? ¿Y con los alquileres que sucede? Qué relación existe entre el valor de los alquileres por barrio y los ingresos de las personas que allí viven.

Links:

- <http://www.ine.gub.uy/encuesta-continua-de-hogares1>
- <https://www.kaggle.com/ppicardo/real-estate-offers-in-montevideo-uruguay>
- API Mercado Libre <https://developers.mercadolibre.com.uy/>

FAKE NEWS (problema abierto)



No hace mucho, las redes sociales empezaron a sufrir inundaciones de noticias falsas. Estas noticias pueden influir en la opinión pública y de esta forma incidir en sucesos trascendentales como los resultados de una elección presidencial. Por ello grandes tecnológicos como Twitter y Facebook comenzaron una carrera contrarreloj, con el objetivo de solucionar este problema.

¿Cómo determino si una noticia es falsa o no? ¿Cómo automatizamos el proceso de categorización para que un algoritmo pueda hacerlo?

Links:

- <http://www.fakenewschallenge.org/>
- <https://www.kaggle.com/c/fake-news/overview>

SAP Next Gen



The SAP Next-Gen program is a purpose driven innovation university and community for the SAP ecosystem enabling companies, partners and universities to connect and innovate with purpose linked to the 17 UN Sustainable Development Goals (SDG). Reimagine the future of industries and experience management with exponential technologies. Seed in disruptive innovation with startups. Build skills for digital futures. Showcase thought leadership.

<https://www.sap.com/corporate/en/company/innovation/next-gen-innovation-platform.html>

SDGs: <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>



1 NO POVERTY

A white icon of a family consisting of two adults and two children, standing together.

2 ZERO HUNGER

A white icon of a bowl with three wavy lines above it, representing steam or food.

3 GOOD HEALTH AND WELL-BEING

A white icon of a heart with a pulse line passing through it.

4 QUALITY EDUCATION

A white icon of an open book with a pencil standing next to it.

5 GENDER EQUALITY

A white icon of a female symbol with an equals sign inside the circle.

6 CLEAN WATER AND SANITATION

A white icon of a water drop falling into a glass.

7 AFFORDABLE AND CLEAN ENERGY

A white icon of a sun with a power button symbol in the center.

8 DECENT WORK AND ECONOMIC GROWTH

A white icon of a bar chart with an upward-pointing arrow.

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

A white icon of three stacked cubes.

10 REDUCED INEQUALITIES

A white icon of a double-headed arrow with three horizontal bars in the center.

11 SUSTAINABLE CITIES AND COMMUNITIES

A white icon of several city buildings of different heights.

12 RESPONSIBLE CONSUMPTION AND PRODUCTION

A white icon of a circular arrow, representing a cycle or recycling.

13 CLIMATE ACTION

A white icon of an eye with the Earth as the pupil.

14 LIFE BELOW WATER

A white icon of a fish swimming above three wavy lines representing water.

15 LIFE ON LAND

A white icon of a tree with two birds flying above it.

16 PEACE, JUSTICE AND STRONG INSTITUTIONS

A white icon of a dove holding an olive branch, standing next to a gavel.

17 PARTNERSHIPS FOR THE GOALS

A white icon of four interlocking circles.

Cómo Seguimos?

1. Armado de Grupos
 - a. moodlex.um.edu.uy
 - b. Equipos de 4 estudiantes
 - c. Multidisciplinarios: al menos 1 informático, 1 economista y 1 otros
2. Enviar correo a eviotti@idatha.com con **Nombre y Correo** de cada uno de los integrantes del equipo.
3. Clases monitoreo -> día y hora a definir



TALLER CIENCIA DE DATOS

Sebastián García

sgarcia@idatha.com

@dsgarcia

Emiliano Viotti

eviotti@idatha.com

@Efviodo

Manu Reynaert

mreynaert@idatha.com

Bitly <http://bit.ly/idatha-ds-course>

Long <https://github.com/efviodo/idatha-data-science-course>

