



Módulo 5. Aplicación Web de Ciencia de datos

LIVE | Data Science

SS2 | Preparar para aplicar

Bienvenida e Introducción

- Agenda
- Objetivos de la sesión
- Panorámica de la sesión
- Bienvenida
- Introducción

Agenda

Sesión Síncrona 1

Aprender

Trabajo asíncrono 1

Profundizar | Ruta de Aprendizaje

Sesión Síncrona 2

Preparar para Aplicar

Trabajo asíncrono 2

Aplicar en el trabajo | Reto



Objetivos particulares de la sesión



- **Manipularás información contenida** base de datos CSV para construir dataframes con memoria cache habilitada
- **Usarás las componentes de Streamlit para construir dashboards interactivos**
- **Generaras graficos y mapas para representar información de forma analitica para la toma de decisiones.**

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

Bienvenida | Dinámica

- Ingresar a la **url** que se te proporcione en el chat



Introducción



- El desconocimiento de la infraestructura de software y hardware para la creación de sitios web puede representar un reto importante para un científico de datos que tiene la necesidad de visualizar y caracterizar su información.
- Al término de esta sesión serás capaz de crear **crear dashboards interactivos profesionales** usando los recursos disponibles en **Streamlit** y Python

1 | Integrar

Aprender + Profundizar

- 1.1 Aprender + Profundizar
- 1.2 Conexión
- 1.3 Aprendizajes



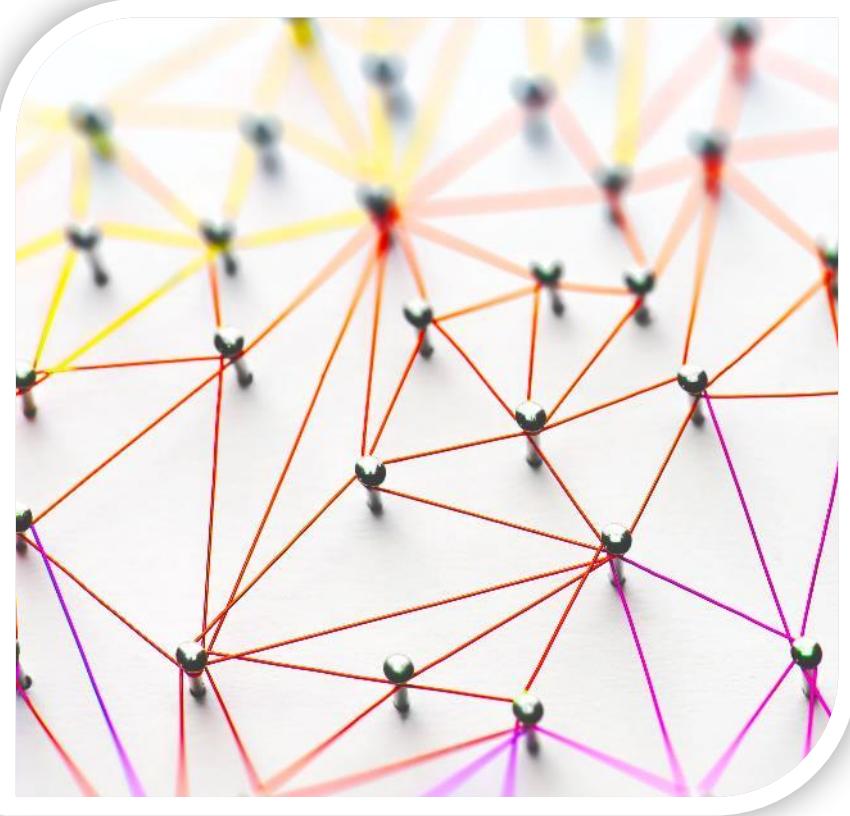
1
Integrar
Aprender + Profundizar

1. Integrar | Aprender + Profundizar



- En esta sesión asociaremos los temas de las sesiones previas para reflexionar:
- ¿Cómo **integro los conocimientos** adquiridos en mi quehacer diario para generar un dashboard?
- ¿Conozco la **información que manipulo** en mi trabajo, como puedo volverla un dataframe y visualizarlo de forma interactiva?
- ¿Será importante mostrar la información en forma gráfica para la toma de decisiones en mi empresa?

1. Integrar | Conexión



- Ingresa a <https://kahoot.it/> con el **codigo** que te proporcione en el chat
- Duración: **10 minutos**

1. Integrar | Aprendizajes



- Explorar **los mecanismos para la obtención de datos de manera eficiente**
- Crear y personalizar dashboards profesionales usando los componentes de **Streamlit**
- Emplear **gráficas** de diversos tipos para exploración de datos usando Streamlit y bibliotecas de Python
- Crear mapas que permitan visualizar base de datos con datos de latitud y longitud de puntos geográficos

2 | Práctica 1

Caso Netflix : Dashboard para visualizar datos de filmes

- 2.1 Contexto
- 2.2 Instrucciones
- 2.3 Trabajo en equipo
- 2.4 Integración del trabajo en equipo



Práctica

2.1 Práctica | Contexto

1629484330

- Netflix ofrece una base de datos de filmes gratuita para que científicos de datos pueden realizar aplicaciones
- Usaremos esta base de datos para construir un dashboard que nos permita visualizar filmes, realizar búsquedas, filtrar filmes por director entre otras funciones

2.2 Práctica | Instrucciones



- Descarga el archivo **movies.csv**
- Genera una libreta en **Google Colab** para esta práctica, cuya estructura esté basada en los análisis solicitados
- Lee, en un **dataframe**, todos los registros del archivo, usando una función con el atributo **@cache**, dicha función deberá recibir un parámetro para determinar el número de filmes a recuperar (500) por default

2.3 Práctica | Trabajo en equipo

Usando el **dataframe** anterior crear un dashboard con la siguiente funcionalidad:



1. Colocar un componente **checkbox** en el **sidebar** del dashboard que permita visualizar todos los filmes recuperados
1. En el sidebar del dashboard colocar un **componente de texto** y un **boton de comando** y realizar las siguientes operaciones:
 - Buscar los filmes por titulo
 - El tipo de coincidencia sera del tipo “que contenga”
 - Debera buscar filmes sin importar se escriban en minusculas o mayusculas

2.3 Práctica | Trabajo en equipo



4. En el sidebar colocar un componente **selectedbox** y un **boton de comando**, con las siguientes características:
 - Deberá mostrar un título y cargar la columna '**director**' del dataframe
 - El botón de comando deberá tomar el director seleccionado y llamar una función que filtre los filmes realizados por ese director

2.3 Práctica | Trabajo en equipo

Solo como sugerencia se espera un dashboard con la siguiente apariencia en la sección para **mostrar todos los filmes** :

The screenshot shows a Jupyter Notebook interface. On the left, there is a sidebar with the following elements:

- A checkbox labeled "Mostrar todos los filmes" which is checked.
- An input field labeled "Titulo del filme:" with a placeholder icon.
- A button labeled "Buscar filmes".
- A dropdown menu labeled "Seleccionar Director" with "James Cameron" selected.
- A button labeled "Filtrar director".

Netflix app

Done! (using st.cache)

Todos los filmes

	budget	company	country	director	ge
0	8,000,000.0000	Columbia Pictures Corpor...	USA	Rob Reiner	Ad
1	6,000,000.0000	Paramount Pictures	USA	John Hughes	Co
2	15,000,000.0000	Paramount Pictures	USA	Tony Scott	Ac
3	18,500,000.0000	Twentieth Century Fox Fil...	USA	James Cameron	Ac
4	9,000,000.0000	Walt Disney Pictures	USA	Randal Kleiser	Ad
5	6,000,000.0000	Hemdale	UK	Oliver Stone	Dr
6	25,000,000.0000	Henson Associates (HA)	UK	Jim Henson	Ad
7	6,000,000.0000	De Laurentiis Entertainm...	USA	David Lynch	Dr
8	9,000,000.0000	Paramount Pictures	USA	Howard Deutch	Co
9	15.000.000.0000	SLM Production Group	USA	David Cronenberg	Dr

2.3 Práctica | Trabajo en equipo

Solo como sugerencia se espera un dashboard con la siguiente apariencia en la sección de **buscar filmes**:

Mostrar todos los filmes

Titulo del filme :
STAR

Buscar filmes

Seleccionar Director
James Cameron

Filtrar director

Netflix app

Done! (using st.cache)

Total filmes mostrados : 5

	gross	name	rating	released	runtime
24	09,713,132.0000	Star Trek IV: The Voyage H...	PG	1986-11-26T00:00:00	119
622	111,987.0000	Stars and Bars	R	1988-03-18T00:00:00	94
697	52,210,049.0000	Star Trek V: The Final Fro...	PG	1989-06-09T00:00:00	107
739	8,143,225.0000	DeepStar Six	R	1989-01-13T00:00:00	105
818	67,878.0000	Brenda Starr	PG	1992-04-15T00:00:00	93

2.3 Práctica | Trabajo en equipo

Solo como sugerencia se espera un dashboard con la siguiente apariencia en la sección de **filtrar por director** :

Mostrar todos los filmes

Título del filme :

Buscar filmes

Seleccionar Director

Filtrar director

Netflix app

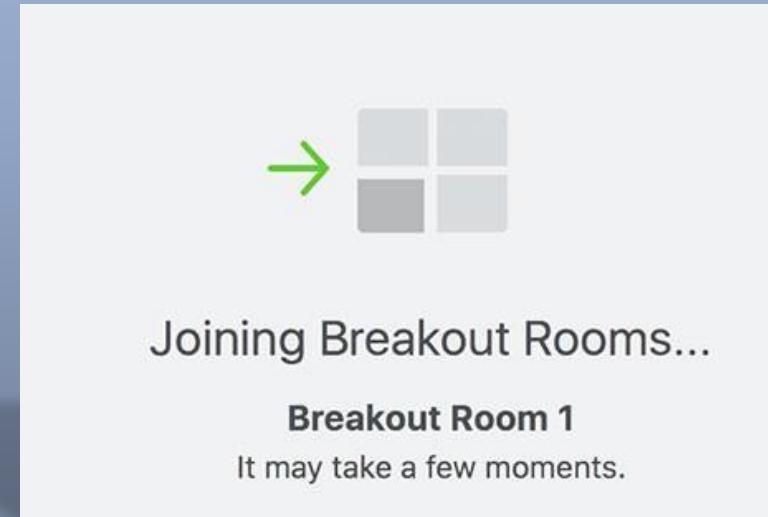
Done! (using st.cache)

Total filmes : 4

	budget	company	country	director	genre	
5	6,000,000.0000	Hemdale	UK	Oliver Stone	Drama	
236	15,000,000.0000	Twentieth Century Fox Fil...	USA	Oliver Stone	Crime	
552	4,000,000.0000	Cineplex-Odeon Films	USA	Oliver Stone	Drama	
684	14,000,000.0000	Ixtlan	USA	Oliver Stone	Biography	

Zoom | BreakOut Rooms (BOR)

1. Al unirte,
aparecerá un
mensaje
1. Despues estarás
con un grupo
pequeño de
compañeros



2.4 Práctica | Integración del trabajo en equipo



- ¿Cuál consideras fue el grado de **dificultad** de la práctica?
- ¿Qué **controles** ocupaste con más frecuencia?
- ¿Te fue posible **usar cache en** algunas de los funciones generadas?
- ¿Qué **reportes** gráficos podrías generar con la base de datos proporcionada?

Práctica

2

2 | Práctica 2

Explorando el uso de bicicletas en la ciudad de nueva york

- 2.1 Contexto
- 2.2 Instrucciones
- 2.3 Trabajo en equipo
- 2.4 Integración del trabajo en equipo

2.1 Práctica | Contexto

245340823

Citi Bike el proyecto de bicicletas compartidas más grande de Estados Unidos, con más de 20,000 bicicletas y alrededor de 1,300 estaciones alrededor de Manhattan, Brooklyn, Queens, the Bronx y Jersey City.

Existe una base de datos con la información de los recorridos reales que realizan las personas, usaremos este dataset para construir un dashboard que nos permita visualizar esos recorridos, graficar el uso de bicicletas por hora, pintar en mapas las estaciones de bicicletas con más demanda por hora del día, entre otros.

1. Se tomará el **mes de septiembre del 2021**, recuperado del portal <https://ride.citibikenyc.com/system-data>.

2.2 Práctica | Instrucciones



- Descarga el archivo `citibike-tripdata.csv`
- Genera una libreta en **Google Colab** para esta práctica, cuya estructura esté basada en los análisis solicitados
- Lee, en un **dataframe** n registros del archivo, usando una función con atributo `@cache`, enviar un parametro para determinar el numero de renglones a recuperar (500 default)

2.3 Práctica | Trabajo en equipo



Crear un dashboard con las siguientes características:

1. Crear un **sidebar**
1. En el sidebar crear los siguientes componentes
 - Un **checkbox** para visualizar todos los registros recuperados de la base de datos
 - Gráfica de barras para visualizar el número de recorridos por hora

2.3 Práctica | Trabajo en equipo



3. En el sidebar crear un control **slider** para seleccionar la hora del dia del campo **“started_at”**

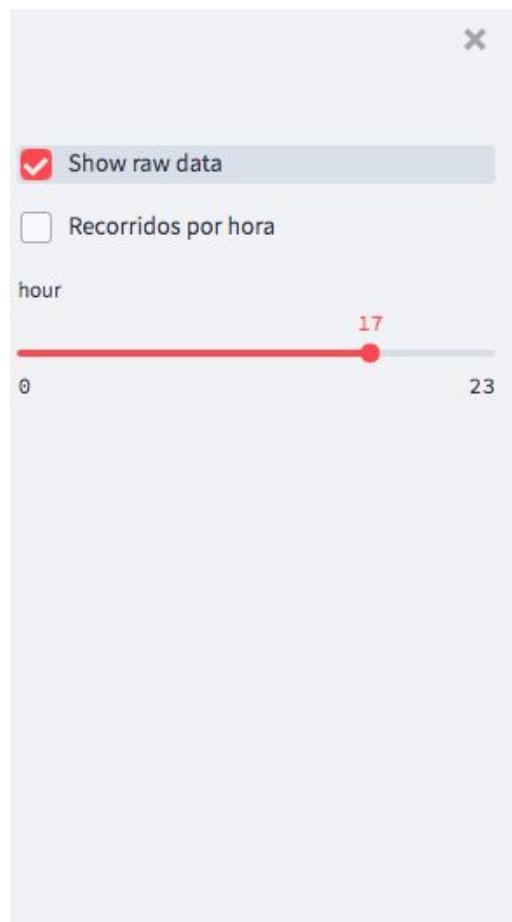
3. Renombrar el campo **started_lat** a lat.

3. Renombrar el campo **started_lng** a lon

3. usando la funcion **st.map(dataframe)** visualizar un mapa para mostrar las puntos gps de las estaciones donde iniciaron los recorridos por hora del día (tomada del control slider)

2.3 Práctica | Trabajo en equipo

Solo como sugerencia se espera un dashboard con la siguiente apariencia en la sección **mostrar todos los datos**:



Cicle Rides in NYC

Done! (using st.cache)

Raw data

	ride_id	rideable_type	started_at	ended_at	sta
0	228F2E2FEBFBF843	classic_bike	2021-09-11T16:03:10	2021-09-11 16:20:57	Je
1	DCDBDC0A5178B...	classic_bike	2021-09-14T17:27:50	2021-09-14 17:34:18	Je
2	DD7AD7B81588187B	docked_bike	2021-09-04T20:29:34	2021-09-04 20:41:17	12
3	4D3CFEEB6E3E514B	classic_bike	2021-09-16T17:26:37	2021-09-16 17:32:48	Je
4	D05FC6DDE1112F08	classic_bike	2021-09-16T18:05:35	2021-09-16 18:16:02	Je
5	8769F362BE51A110	classic_bike	2021-09-13T08:15:09	2021-09-13 08:25:07	Je
6	55E3308E4A04F30A	classic_bike	2021-09-19T13:38:13	2021-09-19 13:43:03	12
7	A6804335147F2379	classic_bike	2021-09-25T18:40:52	2021-09-25 18:45:23	12
8	8E22087D9E5ABC3E	classic_bike	2021-09-25T16:09:26	2021-09-25 16:15:01	12
9	228BF949DD76A3A5	classic_bike	2021-09-25T20:33:38	2021-09-25 20:40:55	12

2.3 Práctica | Trabajo en equipo

Solo como sugerencia se espera un dashboard con la siguiente apariencia:



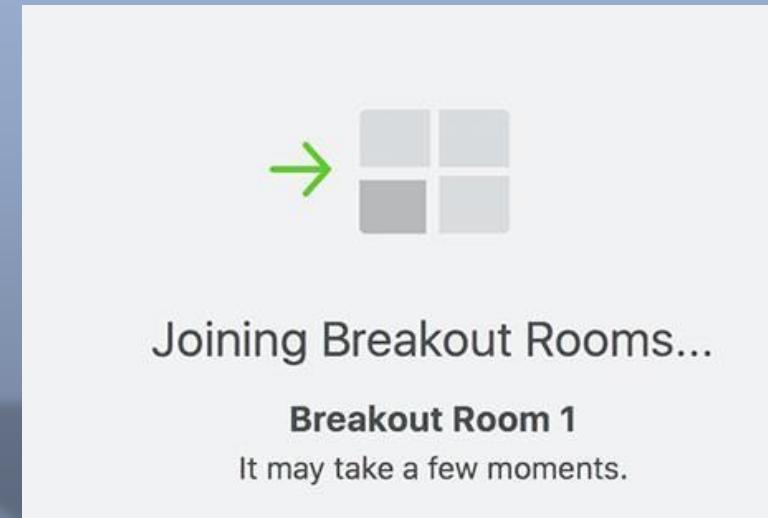
2.3 Práctica | Trabajo en equipo

Solo como sugerencia se espera un dashboard con la siguiente apariencia:



Zoom | BreakOut Rooms (BOR)

1. Al unirte,
aparecerá un
mensaje
1. Despues estarás
con un grupo
pequeño de
compañeros



2.4 Práctica | Integración del trabajo en equipo



- ¿Te fue posible **concluir** todos los ejercicios de la práctica?
- ¿Por qué tuvimos que renombrar 2 columnas para que se llamaran lat y lon?
- ¿Qué **otra información** importante podrías visualizar gráficamente en el mapa con los datos proporcionados?
- ¿Qué elementos de la práctica podrías incorporar en los **reportes** que generas en tu organización?

3 | Reflexión General

Ciclo de prácticas

- 3.1 Tips y reflexiones
- 3.2 Resultados y hallazgos



Reflexión

3.1 Reflexión General: Resultados y Hallazgos obtenidos de las prácticas



- A través de los ejercicios prácticos pudiste constatar que ya tienes los conocimientos para **construir y visualizar** aplicaciones web que accesen a grandes volúmenes de información.
- Los dashboard creados te dan cierta experiencia para mejorar tus habilidades para participar en proyecto de ciencia de datos como desarrollador de aplicaciones web.
- ¿Cómo integrar esta experiencia en los proyectos de tu organización?

3.1 Reflexión General: Tips y Reflexiones de la práctica



- Estudia objetivamente el proyecto y las necesidades reales que pide tu cliente.
- Familiarízate siempre con tus datos, conoce su **estructura y contenido** antes de empezar a trabajar con el proyecto.
- Divide el problema en entregables funcionales
- Construye prototipos simples que funcionen y puedan mostrarse en juntas para recibir retroalimentacion
- Utiliza la mayor parte de los controles y funciones disponibles en Streamlit y bibliotecas de Python para hacer mas facil tu trabajo.



Plan de aplicación

4

4 | Plan de aplicación

Análisis de deserción de empleados



Sesión Síncrona 1
Aprender



Trabajo asíncrono 1
Profundizar | Ruta de Aprendizaje



Sesión Síncrona 2
Preparar para Aplicar



Trabajo asíncrono 2
Aplicar en el trabajo | Reto

Plan de aplicación: Reto



- Para el reto continuarás trabajando con el archivo de empleados provisto en el Hackathon **HackerEarth 2020**.
- Como ya conoces su estructura, aplicarás las **técnicas de desarrollo** aprendidas en esta sesión.
- **Crearas** una dashboard que utilice dataframes de esta base de datos , previa **agrupación y filtrado** de la información en los casos necesarios.
- Con lo anterior podrás tener una visión global de los datos y sacar **conclusiones significativas**.

Plan de aplicación: Reto | Rúbrica



Criterio	Puntaje
Dashboard con controles aplicables a cada sección	10
Funciones para importar y/o filtrar datos usando las funciones de memoria caché.	10
Gráficas y mapas para la visualización de los datos	10
Interpretación y análisis de resultados	5
TOTAL	35 puntos

Cierre Sesión Sincrónica 2

- Integración final de aprendizajes
- Reflexión final
- Concepto clave | TakeAway



Cierre de sesión

Cierre | Sesión Sincrónica 2



- En esta sesión completaste dos prácticas que reforzaron tus conocimientos para:
 - **Explorar** la información de un único o varios archivos de datos CSV usando funciones en Python
 - Usar la memoria cache para evitar tráfico y maximizar recursos
 - Crear **dashboards** interactivos con los controles mas comunes de Streamlit
 - Agregar gráficas y mapas a las aplicaciones

El siguiente paso es replicar estos análisis en la información que manipulas como parte de tus responsabilidades laborales.

Cierre: Concepto clave

Actualmente todas las disciplinas necesitan aplicaciones de frontend web y/o móvil para visualizar sus datos; pues con ello se aporta significado a la información. Las aplicaciones web te permitirán generar nuevas oportunidades, solucionar problemas sustanciales y ser un apoyo para toma de decisiones



Tecnológico de Monterrey | 2020

Prohibida la reproducción total o parcial de esta
obra sin expresa autorización del Tecnológico
de Monterrey

Gracias | Programas LIVE