

viu  
.es

2024 - 2025



# ACTIVIDAD GUIADA

Máster en Big Data y Data Science

01MBID – Fundamentos de la Tecnología Big data

Curso 2024 – Ed. Octubre – Grupo A

# 1 . Descripción general

## Actividades Guiadas

DESCRIPCIÓN	
<b>Introducción</b>	Extracción, procesamiento y visualización de datos. Primer contacto con una base de datos NoSQL.
<b>Objetivo</b>	1.- Crear una base de datos usando MongoDB. 2.- Crear una colección con cuentas de twitter y tweets a partir de los datasets proporcionados. 3.- Cargar el ejemplo de datos geolocalizados que incorpora MongoDB Atlas. 4.- Consultar los datos y Manipular los datos 5.- Visualizar los datos utilizando MongoDB Charts.
<b>Trabajo previo</b>	1.- Revisar los tutoriales de creación y uso de herramientas para el manejo de MongoDB, Compass y Colab. 2.- Material Campus. 3.- Revisar la documentación de Twitter: <i>Developer Platform – Docs – Object Model</i>
<b>Metodología</b>	El profesor brindará los conocimientos suficientes relacionados a la creación, uso y configuración de las herramientas necesarias para realizar la actividad.
<b>Tarea para el e-portfolio</b>	1.- Crear una base de datos MongoDB en la nube usando MongoDB Atlas ( <a href="https://www.mongodb.com/">https://www.mongodb.com/</a> ) 2.- Crear las colecciones 2.1.- Cargar los datasets de cuentas de twitter y tweets proporcionados por el profesor. 2.2.- Cargar la colección de ejemplo geolocalizada de MongoDB Atlas (sample_geospatial -> shipwrecks). 3.- Crear una cuenta en Google Colab. 4.- Utilizando como base el script en Python profesor, realizar los cambios necesarios para: 4.1.- En la colección de cuentas de <i>twitter</i> , tener los campos <i>amigos</i> y <i>tweets enviados</i> , cargar los datos correspondientes mediante consulta <code>mongodb + código python</code> .

## DESCRIPCIÓN

4.2.- En la colección de tweets, calcular la antigüedad para cada tweet en función de la fecha actual considerando antigüedad 0 el día de hoy y sumando +1 por cada día transcurrido. Incluir en el mismo documento del tweet un nuevo campo que se llamará **antigüedad\_dias** con esa antigüedad calculada.

4.3.- En la colección de tweets, calcular la antigüedad de cada tweet relativa con la fecha de creación de la cuenta más antigua de la colección actual de datos. Considerando antigüedad 0 si fue enviado el mismo día de creación y sumando +1 por cada día transcurrido desde entonces en función de la fecha del tweet. Incluir en el mismo documento del tweet un nuevo campo se llamará **frescura\_relativa\_dias**.

### 4.- Cuadro de mandos (MongoDB Charts)

4.1.- Crear un chart que muestre la cuenta de Twitter con mayor cantidad de *tweets enviados*.

4.2.- Crear un chart que muestre la cuenta de Twitter con mayor cantidad de *amigos* (cuentas a las que sigue el usuario).

4.3.- Crear un chart que muestre el número total de Tweets en la Colección.

4.4.- Crear un chart que muestre el número total de Tweets por cuenta.

4.5.- Crear un chart que muestre las cuentas de *Twitters* versus *amigos* ordenado de mayor a menor.

4.6.- Crear un chart que muestre las cuentas de *Twitters* versus *tweets enviados* ordenado de mayor a menor.

4.7.- Crear un chart que muestre los *Tweets* versus día semana en el que fueron enviados/creados.

4.8.- Crear un chart que muestre ordenados todos los Tweets según su **antigüedad\_dias**. Dado que la versión gratuita tiene limitaciones de la cantidad de datos a mostrar, mostrar dos charts: uno ordenado de mayor a menor y otro de menor a mayor.

4.9.- Crear un chart que muestre ordenados todos los Tweets según su **frescura\_relativa\_dias**. Dado que la versión gratuita tiene limitaciones de la cantidad de datos a mostrar, mostrar dos charts: uno ordenado de mayor a menor y otro de menor a mayor.

4.10.- Crear un chart para visualizar en el mapa mundial todos los pecios hundidos hasta 50 metros de profundidad. Pintando en verde los hundidos hasta los 10 metros, amarillo hasta los 20 metros y en rojo hasta los 50 metros.

DESCRIPCIÓN	
<b>Forma de entrega</b>	<p>1.- Archivo <b>PDF</b>:</p> <p>1.1.- Formato de la Universidad, portada (datos de la asignatura, datos del alumno). Utilizar plantilla proporcionada.</p> <p>1.2.- Capturas de pantalla de la configuración realizada, además de una explicación breve.</p> <p>1.3.- Charts: Captura de pantalla, más captura de la configuración y breve explicación.</p> <p>2.- Código:</p> <p>2.1.- Archivo Python.</p>
Fecha de entrega	
1ª Convocatoria	<b>10/11/2024 hasta las 23:59 (hora de Valencia, España)</b>
2ª Convocatoria	<b>23/03/2025 hasta las 23:59 (hora de Valencia, España)</b>

RÚBRICA DE EVALUACIÓN DE LAS ACTIVIDADES GUIADAS				
Laboratorio informático 25% + Resolución de problemas 20% ( 45% sobre nota final)				
	Estilo/Formato (5%)	Configuración (5%)	Carga de datos (15%)	Visualización (75%)
<b>Muy competente</b>  10 - 9	Redacción impecable, con estructura definida, estilo formal y sin faltas ortográficas.	Todo funciona correctamente.	Datos cargados, incluyendo la manipulación de los mismos	Todos los Charts correctos.
<b>Competente</b>  8,9 - 7	Redacción coherente, estilo informal o alguna falta ortográfica.	Pequeños errores de configuración.		Falta algún chart o errores menores en los mismos.
<b>Aceptable</b>  6,9 - 5	Redacción muy corta, con alguna incoherencia y alguna falta ortográfica.	Errores de configuración.	Datos cargados, sin la manipulación de los mismos	Falta bastantes charts o errores en los mismos.
<b>Aún no Competente</b>  4,9 - 0	Redacción incoherente gramaticalmente con muchas faltas ortográficas.	Actividad no se puede validar por errores de configuración.	Sin datos	Se presentan menos de la mitad de los charts o con errores graves.