

viu
.es

2023 - 2024



ACTIVIDAD GUIADA

Máster en Big Data y Data Science

01MBID – Fundamentos de la Tecnología Big data

Curso 2024 – Ed. Octubre – Grupo A-B

viu

Universidad
Internacional
de Valencia

Actividad Guiada

DESCRIPCIÓN	
Introducción	Extracción, procesamiento y visualización de datos. Primer contacto con una base de datos NoSQL.
Objetivo	<ol style="list-style-type: none"> 1.- Crear una base de datos usando MongoDB. 2.- Crear una colección con cuentas de twitter y tweets a partir del dataset proporcionado. 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.
Trabajo previo	<ol style="list-style-type: none"> 1.- Revisar los tutoriales de creación y uso de herramientas para el manejo de MongoDB, Compass y Colab. 2.- Material Campus.
Metodología	El profesor brindará los conocimientos suficientes relacionados a la creación, uso y configuración de las herramientas necesarias para realizar la actividad.
Tareas para el PORTAFOLIO	<p>1.- Crear una base de datos MongoDB: utilizar el servicio cloud MongoDB Atlas (https://www.mongodb.com/)</p> <p>2.- Crear las colecciones e ingestar los datos:</p> <ol style="list-style-type: none"> 2.1.- Cargar los datasets proporcionados por el profesor en formato JSON: 1) json de cuentas de twitter y 2) json de tweets. 2.2.- Cargar la colección de ejemplo de MongoDB Atlas con datos de geolocalización (sample_geospatial >> shipwrecks). <p>3.- Crear una cuenta en Google Colab.</p> <p>4.- Utilizando como base el script en Python profesor, realizar los cambios necesarios para:</p> <ol style="list-style-type: none"> 4.1.- En la colección de twitters, añadir los campos <i>amigos</i> y <i>tweets enviados</i>: cargar los datos correspondientes mediante consulta mongodb + código python. 4.2.- En la colección de tweets, calcular la antigüedad del tweet en función de la fecha actual considerando antigüedad 0 el día de hoy y sumando +1 por cada día transcurrido. Este nuevo campo se llamará <i>Frescura</i>. 4.3.- En la colección de tweets, calcular la antigüedad del tweet relativa con la fecha de creación de la cuenta. Considerando antigüedad 0 si fue enviado el mismo día de creación de la cuenta y sumando +1 por cada día transcurrido desde entonces en función de la fecha del tweet. Este nuevo campo se llamará <i>Madurez</i>. <p>IMPORTANTE: DEJAR ÚNICAMENTE EL CÓDIGO NECESARIO PARA LA ACTIVIDAD, LIMPIAR EL RESTO DE CÓDIGO INNECESARIO PROPORCIONADO, COMENTAR EL CÓDIGO ADECUADAMENTE.</p>

	<p>5.- Cuadro de mandos (MongoDB Charts):</p> <p>5.1.- Cuenta de Twitter con mayor cantidad de <i>tweets enviados</i>. Es un dato disponible dentro de los datos de cada tweet (<i>Developer Platform – Docs – Object Model</i>).</p> <p>5.2.- Cuenta de Twitter con mayor cantidad de <i>amigos</i> (cuentas a las que sigue el usuario) (<i>Developer Platform – Docs – Object Model</i>).</p> <p>5.3.- Número total de Tweets.</p> <p>5.4.- Número total de Tweets por cuenta.</p> <p>5.5.- Cuentas de <i>Twitters</i> versus <i>amigos</i> ordenado de mayor a menor.</p> <p>5.6.- Cuentas de <i>Twitters</i> versus <i>tweets enviados</i> ordenado de mayor a menor.</p> <p>5.7.- <i>Tweets</i> versus día semana.</p> <p>5.8.- Ordenar los Tweets según su Madurez. Indicar también que cuenta de Twitter es la más Madura.</p> <p>5.9.- Ordenar los Tweets según su Frescura. Indicar también que cuenta de Twitter es la más Fresca.</p> <p>5.10.- Visualizar en el mapa mundial todos los pecios hundidos hasta 20 metros de profundidad. Pintando en verde los hundidos hasta 10 metros, amarillo hasta 15 metros y en rojo hasta 20 metros.</p>
<p>Forma de entrega</p>	<p>1.- Archivo en formato PDF: longitud máxima de 20 páginas (flexibilidad)</p> <p>1.1.- Formato de la Universidad, con portada (datos de la asignatura, datos del alumno). Utilizar plantilla proporcionada (o bien propia en Latex o similar que tenga el mismo aspecto/formato que la plantilla proporcionada).</p> <p>1.2.- Capturas de pantalla de la configuración realizada, además de una explicación.</p> <p>1.3.- Código: incluir el código final, preferentemente en formato texto y/o como capturas. Incluir explicaciones ya sea como comentario en el propio código y/o explicaciones adicionales.</p> <p>1.4.- Charts: Captura de pantalla de cada chart, junto con captura de la configuración y breve explicación.</p> <p>2.- Código: formato IPYNB (notebook de Google Colab)</p> <p>2.1.- Adjuntar a parte el archivo Python ejecutable en Google Colab (.ipynb).</p> <p>IMPORTANTE:</p> <p>NO SUBIR LOS ARCHIVOS ANTERIORES COMO .ZIP,</p> <p>DEBEN SUBIRSE DE FORMA SEPARADA E INDEPENDIENTE.</p>

Fecha de entrega	
1ª Convocatoria	19/05/2024 hasta las 23:59 (hora de Valencia, España)
2ª Convocatoria	25/10/2024 hasta las 23:59 (hora de Valencia, España)

RÚBRICA DE EVALUACIÓN DE LA ACTIVIDAD GUIADA				
	Estilo/Formato (10%)	Carga de datos (20%)	Programación (30%)	Visualización (40%)
Muy competente 10 - 9	Redacción impecable, con estructura definida, estilo formal y sin faltas ortográficas.	Datos cargados, incluyendo la manipulación de los mismos	Todo funciona correctamente. Configuración correcta, código bien comentado.	Todos los Charts correctos.
Competente 8,9 - 7	Redacción coherente, estilo informal o alguna falta ortográfica.		Pequeños errores de programación y/o de configuración.	Falta algún chart o errores menores en los mismos.
Aceptable 6,9 - 5	Redacción muy corta, con alguna incoherencia y alguna falta ortográfica.	Datos cargados, sin la manipulación de los mismos	Errores de programación y/o de configuración.	Falta bastantes charts o errores en los mismos.
Aún no Competente 4,9 - 0	Redacción incoherente gramaticalmente con muchas faltas ortográficas.	Sin datos	Actividad no se puede validar por errores de programación y/o de configuración.	Se presentan menos de la mitad de los charts o con errores graves.