

Universidad Nacional Abierta y a Distancia
Vicerrectoría Académica y de Investigación
Guía única para el desarrollo del componente práctico del curso Gestión y
Almacenamiento de Datos 203008068

1. Información general del componente práctico.

Estrategia de aprendizaje: Aprendizaje basado en tareas
Tipo de curso: Metodológico
Momento de la evaluación: Intermedio
Puntaje máximo del componente: 150 puntos
Número de actividades del componente registradas en esta guía: 3
Con este componente se espera conseguir los siguientes resultados de aprendizaje: Valora procedimientos de tratamiento de base de datos SQL mediante técnicas analíticas para el aseguramiento de la eficiencia organizacional y la mitigación de los riesgos en cada uno de los contextos aplicados.

2. Descripción general actividad(es) del componente práctico.

Escenarios de componente práctico: Con Apoyo TIC	
Tipo de actividad: Independiente	
Número de actividad: 3	
Puntaje máximo de la actividad: 150 puntos	
La actividad inicia el: lunes, 2 de octubre de 2023	La actividad finaliza el: domingo, 29 de octubre de 2023
Los recursos con los que debe contar para el desarrollo de la actividad son los siguientes: <ul style="list-style-type: none"> • Internet • Computador con Microsoft Excel 	

La actividad consiste en:

Estimado Estudiante al finalizar la tarea 3, va a llevar a cabo una de las tareas más importantes y comunes para un científico de datos: La limpieza de datos.

- Revisar cada variable
- Comprobar la existencia de errores
- Identificar datos ausentes
- Realizar conversión de formatos si aplica
- Una vez preparado el dataset, llevarlo a una base de datos.

Debe tener los siguientes productos

- **Actividad 1:** Instalar OpenRefine para realizar la limpieza de datos. (revisar video tutorial publicado en foro de la tarea 3 de instalación y uso de OpenRefine)
- **Actividad 2:** Los datos para esta actividad están en el fichero data_tarea_03.cvs. Sobre dicho fichero llevarás a cabo la limpieza de datos. Existen al menos (6) seis errores que deben ser identificados y corregidos, con el fin de limpiar y homogenizar el dataset final. Para cumplir con esta actividad solicita:
 - a. Describe los problemas:** Para cada uno de los elementos susceptibles de ser limpiados indica:
 - Descripción del Problema: Indicando claramente:
 - Qué o cuál es el problema
 - Cómo se ha detectado.
 - Por qué que se trata de un problema

Forma de solucionarlo: Tener en cuenta que un mismo problema puede tener diferentes soluciones, razón por la cual debe seleccionar la solución(es) que considere más adecuada en cada caso. Es importante también justificar la elección de dicha solución. Es decir, explicar por qué el dataset final, tras aplicar la solución, es el más idóneo para ser tratado o es mejor que el dataset original

Importante: Cada problema detectado puede afectar a una o varias filas/columnas de sus datos. No debe centrarse en buscar problemas puntuales sino identificar problemas tipo a los que habrá que darles una solución. Por ejemplo, si encuentro un error del tipo "X", aplicar la solución "Y"

b. Proponer un formato xls oxlsx: Equivalente al CSV proporcionado.

Cada xls o xlsx debe tener toda la información contenida en la fila correspondiente del documento CSV. Aproveche esta conversión del formato (CVS a EXCEL) para eliminar los campos redundantes, los problemas del punto **a**, etc, Describe todos estos cambios o transformaciones para que el tutor conozca el proceso utilizado.

c. Proponer al menos dos mejoras en el conjunto de datos. Después de conocer el conjunto de datos utilizado, proponer dos o más mejoras para garantizar la calidad de los datos. Tener en cuenta que estas mejoras pueden afectar el procedimiento de captura de datos o a la estructura de la base de datos, por lo tanto, explicar la mejora como procedimiento para llevar a cabo.

• **Actividad 3:** Consolidado de la tarea en el Entorno de Evaluación:

- Entregar un informe con los elementos habituales (portada, índice, numeración de páginas y referencias bibliográficas con normas APA)

- Dar respuesta a los apartados planteados en la actividad 2. Cada apartado debe ir resuelto en secciones diferentes.
- El fichero xls oxlsx debe acompañar al informe propuesto y no debe contener ningún error de formato

Para el desarrollo de la actividad tenga en cuenta que:

En el entorno de Información inicial debe:

- Revisar la presentación del curso, aceptar las normas y condiciones para el desarrollo del curso y revisar la agenda del curso.

En el entorno de Aprendizaje debe:

- Revisar la lectura del de las referencias bibliográficas de la unidad 1 Base Datos en SQL
- Participar en el foro publicado pantallazo de la instalación de OpenRefine

En el entorno de Evaluación debe:

- Al finalizar la actividad entregará el informe en formato PDF y el fichero xls o xlsx

Evidencias de trabajo independiente:

Las evidencias de trabajo independiente para entregar son:

- El informe en formato PDF dando respuesta a los parámetros solicitados en la guía
- Propuesta del dataset final en formato EXCEL (xls o xlsx)

Evidencias de trabajo grupal:

En esta actividad no se requieren evidencias de trabajo grupal.

3. Lineamientos generales para la elaboración de las evidencias

Para evidencias elaboradas **de forma Independiente**, tenga en cuenta las siguientes orientaciones

1. Realice un reconocimiento general del curso y de cada uno de los entornos antes de abordar el desarrollo de las actividades.
2. Identifique los recursos y referentes de la unidad a la que corresponde la actividad.
3. Intervenga en el foro de discusión aplicando las normas de Netiqueta Virtual, evidenciando siempre respeto por las ideas de sus compañeros y del cuerpo docente.
4. Antes de entregar el producto solicitado revise que cumpla con todos los requerimientos que se señalaron en esta guía de actividades, rúbrica de evaluación y por parte del tutor en el foro de discusión.
5. No cometa fraudes, ni plagios ni actos que atenten contra el normal desarrollo académico de las actividades.

Tenga en cuenta que todos los productos escritos individuales o grupales deben cumplir con las normas de ortografía y con las condiciones de presentación que se hayan definido.

En cuanto al uso de referencias considere que el producto de esta actividad debe cumplir con las normas **APA**

En cualquier caso, cumpla con las normas de referenciación y evite el plagio académico, para ello puede apoyarse revisando sus productos escritos mediante la herramienta Turnitin que encuentra en el campus virtual.

Considere que En el acuerdo 029 del 13 de diciembre de 2013, artículo 99, se considera como faltas que atentan contra el orden académico, entre otras, las

siguientes: literal e) “El plagiar, es decir, presentar como de su propia autoría la totalidad o parte de una obra, trabajo, documento o invención realizado por otra persona. Implica también el uso de citas o referencias faltas, o proponer citad donde no haya coincidencia entre ella y la referencia” y literal f) “El reproducir, o copiar con fines de lucro, materiales educativos o resultados de productos de investigación, que cuentan con derechos intelectuales reservados para la Universidad.”

Las sanciones académicas a las que se enfrentará el estudiante son las siguientes:

- a) En los casos de fraude académico demostrado en el trabajo académico o evaluación respectiva, la calificación que se impondrá será de cero puntos sin perjuicio de la sanción disciplinaria correspondiente.
- b) En los casos relacionados con plagio demostrado en el trabajo académico cualquiera sea su naturaleza, la calificación que se impondrá será de cero puntos, sin perjuicio de la sanción disciplinaria correspondiente.

4. Formato de Rúbrica de evaluación

Tipo de actividad: Independiente	
Número de actividad: 3	
Momento de la evaluación: Intermedio	
La máxima puntuación posible es de 150 puntos	
Criterios	Desempeños
Primer criterio de evaluación: Describe los problemas Este criterio representa 70 puntos del total de 150 puntos de la actividad.	<p>Nivel alto: Para cada uno de los elementos limpiados: la descripción del problema es correcta y la solución aportada es correcta y es replicable. Si su trabajo se encuentra en este nivel puede obtener entre 40 puntos y 70 puntos</p> <p>Nivel medio: Algunos de los elementos limpiados: la descripción del problema es correcta y la solución aportada es correcta y es replicable. Si su trabajo se encuentra en este nivel puede obtener entre 10 puntos y 39 puntos</p> <p>Nivel bajo: Los elementos limpiados: la descripción del problema no es correcta y la solución aportada no es correcta y no replicable Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 9 puntos</p>
Segundo criterio de evaluación: Propuesta un formato xls o.xlsx. Este criterio representa 35 puntos del total de 150 puntos de la actividad	<p>Nivel alto: Se respeta el formato propuesto de todos los campos, en caso contrario se argumenta claramente los cambios y se han eliminado los campos redundantes Si su trabajo se encuentra en este nivel puede obtener entre 20 puntos y 35 puntos</p> <p>Nivel medio: Se respeta el formato propuesto de algunos campos, en caso contrario se argumenta algunos cambios y se han eliminado algunos campos redundantes Si su trabajo se encuentra en este nivel puede obtener entre 5 puntos y 19 puntos</p> <p>Nivel bajo: No respeta el formato propuesto de todos los campos, en caso contrario no se argumenta claramente los cambios y no se han eliminado los campos redundantes.</p>

	Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 4 puntos
Tercer criterio de evaluación: Proponer al menos dos mejoras en el conjunto de datos Este criterio representa 35 puntos del total de 150 puntos de la actividad	Nivel alto: La implicación de las mejoras propuestas están bien argumentadas y son viables Si su trabajo se encuentra en este nivel puede obtener entre 20 puntos y 35 puntos Nivel medio: La implicación de las mejoras propuestas pueden ser mejor argumentadas y son viables Si su trabajo se encuentra en este nivel puede obtener entre 5 puntos y 19 puntos Nivel bajo: Únicamente presenta una propuesta de mejora y la argumentación puede ser mejor y es viables / No presenta propuesta de mejora Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 4 puntos
Cuarto criterio de evaluación: Participación : Interactúa de forma oportuna, adecuada y respetuosa en el foro dejando evidencia de la actividad 1 de la tarea Este criterio representa 5 puntos del total de 150 puntos de la actividad	Nivel alto: El Estudiante interactúa de forma oportuna, adecuada y respetuosa en el foro dejando evidencia de la instalación de OpenRefine. Si su trabajo se encuentra en este nivel puede obtener entre 4 puntos y 5 puntos Nivel medio: El Estudiante interactúa faltando tres días para el cierre de la actividad de forma respetuosa en el foro dejando evidencia de la instalación de OpenRefine. Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 3 puntos Nivel bajo: El estudiante no deje evidencia en foro de la actividad de la evidencia de la instalación de OpenRefine. Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 puntos
Quinto criterio de evaluación: Presenta los documentos con los lineamientos	Nivel alto: Presenta los archivos, bien organizados, cumpliendo con todas las especificaciones solicitadas en la guía de actividades y entrega en el entorno de evaluación. Si su trabajo se encuentra en este nivel puede obtener entre 4 puntos y 5 puntos

<p>indicados en la guía de actividades</p> <p>Este criterio representa 5 puntos del total de 150 puntos de la actividad</p>	<p>Nivel Medio: Aunque presenta los archivos, estos presenta fallas en su organización, redacción y/o completitud, no sigue los lineamientos dados.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 3 puntos</p> <p>Nivel bajo: Los archivos presentados no sigue los lineamientos dados o no fue presentados en el entorno de evaluación</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 puntos</p>
--	--