PROCESO DE GESTIÓN DE FORMACIÓN PROFESIONAL INTEGRAL

FORMATO GUÍA DE APRENDIZAJE

1. IDENTIFICACIÓN DE LA GUIA DE APRENDIZAJE.

- Denominación del Programa de Formación: programacion para analitica de datos
- código del programa de formación: 228117
- nombre del proyecto : aplicar buenas practicas para preparar, limpiar, refinar y explorar grandes volumenes de datos en el sector productivo
- fase del proyecto : fase 4. adquisición de datos
- actividad de proyecto : 1.1 recolectar datos
- competencia : 210601026 - procesar datos de acuerdo con procedimiento técnico y metodología estadística
- resultados de aprendizaje alcanzar: 584007 - 2.realizar el proceso de limpieza de datos de acuerdo con la herramientas informáticas seleccionada.

2. PRESENTACIÓN



Apreciado Aprendiz:

La limpieza de datos es un proceso fundamental en el análisis de datos que implica la identificación y corrección de errores o inconsistencias en un conjunto de datos. Este proceso es crucial porque los datos sucios pueden llevar a conclusiones erróneas y afectar la calidad de los resultados en cualquier análisis posterior.

Importancia de la Limpieza de Datos

Mejora la Calidad de los Datos: Asegura que los datos sean precisos y confiables.

Facilita el Análisis: Un conjunto de datos limpio permite un análisis más efectivo y eficiente.

Aumenta la Credibilidad: Los resultados obtenidos de datos limpios son más creíbles y pueden ser utilizados para la toma de decisiones.

Apreciado Aprendiz, le invitamos a contestar las siguientes preguntas y compartir sus apreciaciones con sus compañeros e instructor.

2.1 ¿Qué conoces por limpieza de datos ?,¿Cuándo aplicas la limpieza de datos en tu vida cotidiana?,¿Conoces aplicaciones para limpieza de datos? Y ¿Cómo seria el mundo sin limpiar los datos?

ERES UN APRENDIZ AUTÓNOMO, QUE PUEDES APRENDER A TRAVÉS DE ESTRATEGIAS COMO ESTA. ¡ANIMATE Y PARTICIPA!

Animo y éxitos.

3. FORMULACIÓN DE LAS ACTIVIDADES DE APRENDIZAJE

3.1 Actividades de Reflexión Inicial

3.1.1. Participo en una charla introductoria con el Instructor, para analizar y meditar sobre las actividades que vamos a desarrollar, propongo ideas y estrategias para consecución de los resultados de aprendizaje sobre la limpieza de datos , observo los videos,

Material de Apoyo: limpieza de datos dos-limpieza de datos no estructurados-limpieza de datos

3.1.2. Expongo mi concepto sobre el desarrollo de las actividades del proyecto, los problemas presentados hasta el momento, propongo ideas para mejorar.

3.1.3 Diseña un mapa mental que responda la siguiente pregunta.¿por qué crees que la limpieza de datos es importante en el análisis ?

3.2 ACTIVIDADES DE CONTEXTUALIZACIÓN E IDENTIFICACIÓN DE CONOCIMIENTOS NECESARIOS PARA EL APRENDIZAJE

- Realiza un diagrama apartir del siguiente documento , donde identificaras los aspectos mas relevantes de la limpieza de datos

Material de apoyo: 02.Análisis de la calidad del dato-1.pdf

- Realizar el taller sobre modelos matemáticos

Material de apoyo: taller modelo matemáticos.pdf

- **Realiza la base de datos utilizando la aplicación Access**

Material de apoyo: base de datos- colegio.pdf

Debe presentar el archivo que genera Access y el tutorial con evidencias del proceso realizado.

3.3 ACTIVIDADES DE APROPIACIÓN DEL CONOCIMIENTO (CONCEPTUALIZACIÓN Y TEORIZACIÓN).

En este punto de aprendizaje daras inicio al desarrollo y conectextualizacion del proceso de limpieza de datos con el siguiente ejercicio practico utilizando la aplicación Excel y el tutorial.

- Descargar el el siguiente material de apoyo:
  - ✓ datos-fallecidos-en-siniestros-de-trnsito-2013.csv

- ✓ metadatos_fallecidos_siniestros_transito.json
- ✓ tutorial-limpieza-y-anc3a1lisis-de-datos-bc3a1sico-hassel-fallas1.pdf

sigue el tutorial de limpieza y evidencia tu proceso tomando pantallazos, realiza los 32 puntos propuestos

terminado el proceso , responde tus preguntas frente al análisis realizado

- **¿Cuántos fallecidos en total son ?**
- **¿Cuántos campos encontraste nulos en la base de datos?**
- **¿Cuál fue la ciudad donde mas accidentes ocurrienron ?**
- **¿ que funciones utilizase para hacer la limpieza ?**
- **¿Que te parecio lo mas difícil del ejericio realizar una descripción y por que.?**

Realiza un mapa mental con el glosario de términos de esta guía , los conceptos los deben colocar con sus propias palabras de la interpretación que realicen.

3.3.1 ACTIVIDADES DE APROPIACIÓN DEL CONOCIMIENTO

Actividad Práctica con OpenRefine.

- **Utilizar el programa de limpieza de datos OpenRefine y realizar el tutorial.**

Material de Apoyo: practica-open-refine.pdf

Presentar evidencias atravez de pantallazos hasta la pagina 20 del tutorial del proceso realizado

Esta aplicación ocupa recursos del equipo como memoria RAM, si tu equipo no cuenta con esos recursos minimos, para la evidencia de esta actividad debes documentar acerca del funcionamiento de esta aplicación y que funcionales tiene y para que sirven, de forma general colocalas.

Material de Apoyo: **guia-para-la-limpieza-de-datos-sobre-biodiversidad-con-openrefine.es.PDF**

3.4 ACTIVIDADES DE TRANSFERENCIA DEL CONOCIMIENTO

- Desarrolla las siguiente funciones en Excel de limpieza de datos (propone los ejercicios y entrega evidencias del proceso)

1. Mayúsculas y minúsculas

Es común encontrar en una misma columna un uso indistinto de mayúsculas y minúsculas. Las siguientes fórmulas te pueden ayudar a homologar su uso.

MINUSC	Convierte todas las mayúsculas en letras minúsculas	MINUSC(“HOLA”) = hola
MAYUSC	Pone el texto en letras mayúsculas	MAYUSC(“hola”) = HOLA
NOMPROPIO	Cambia las primeras letras de cada palabra en mayúsculas	NOMPROPIO(“eScUELA de dATOS”) = Escuela De Datos

2. Espacios de más

ESPACIO	Elimina los espacios del texto, excepto el espacio normal que se deja entre palabras	ESPACIO(“H O L A MUNDO”) = HOLA MUNDO

3. Extraer y concatenar

EXTRAE	Extrae una o varias letras. Es necesario indicar la posición inicial y el número de caracteres que deseamos extraer	EXTRAE(«HOLA»,2,3) = OLA
CONCATENAR	Una de las funciones de texto, para unir dos o más cadenas de texto en una sola	CONCATENAR(«Bienvenido a»,» «,»Excel Total») = «Bienvenido a Excel Total»

4. Corregir fechas

FECHA	Convierte un texto con formato de fecha en formato AÑO/MES/DÍA	FECHA(1990; 05; 19) = 1990/05/19

4. PLANTEAMIENTO DE EVIDENCIAS DE APRENDIZAJE PARA LA EVALUACIÓN EN EL PROCESO FORMATIVO.

Fase del proyecto formativo	Actividad del proyecto formativo	Actividad de Aprendizaje	Evidencias de Aprendizaje	Criterios de Evaluación	Técnicas e Instrumentos de Evaluación
EJECUCION	APLICAR TECNICAS PARA LA INTEGRACION DE DATOS MEDIANTE EL SOFTWARE REQUERIDO	**realizar guía de aprendizaje**	**Evidencias de Producto:** Resolución de la actividades guía de aprendizaje y entregas pdf. Desarrollo ejercicios practicos limpieza de datos, archivo formato accbd-csv	interpreta requerimientos para el tratamiento de datos según la necesidad del cliente. analiza la muestra según los componentes del análisis de datos para identificar los tipos de datos. identifica las características de big data para su contextualización de acuerdo al entorno. selecciona la herramienta para la limpieza de datos de acuerdo a la muestra y requerimiento del usuario. clasifica y ordena las variables según	**lista de chequeo** Respuesta a preguntas. Observación.

				los requerimientos	

5. GLOSARIO DE TÉRMINOS

Datos Faltantes: Valores que no están presentes en un conjunto de datos, lo que puede afectar el análisis.

Duplicados: Registros que aparecen más de una vez en un conjunto de datos, lo que puede llevar a interpretaciones erróneas.

Normalización: Proceso de transformar los datos a un formato estándar, como uniformizar el formato de fechas o texto.

Inconsistencias: Datos que no son coherentes entre sí, como diferentes formas de escribir el mismo valor (por ejemplo, "Sí" y "si").

Transformación de Datos: Modificación de los datos para que sean más adecuados para el análisis, que puede incluir normalización, agregación, etc.

Validación de Datos: Proceso de verificar que los datos son precisos y cumplen con ciertos criterios antes de ser utilizados en análisis.

Outliers (Valores Atípicos): Datos que se desvían significativamente del resto del conjunto y pueden influir negativamente en los resultados del análisis.

Formato de Datos: La manera en que se presenta la información, como texto, número, fecha, etc. La limpieza puede requerir la conversión entre formatos.

Errores Tipográficos: Errores en la escritura que pueden causar problemas en el análisis de datos, como nombres mal escritos.

Facetas: Herramientas de filtrado que permiten explorar y segmentar datos en OpenRefine y otras plataformas.

Conjunto de Datos: Colección de datos organizados en filas y columnas, comúnmente en formato tabular.

Extracción, Transformación y Carga (ETL): Proceso de tomar datos de diversas fuentes, transformarlos para su uso y cargarlos en un sistema de almacenamiento.

Data Wrangling: Proceso de limpiar y transformar datos para hacerlos más útiles para el análisis.

Análisis Descriptivo: Método que resume y describe las características de un conjunto de datos.

Metadatos: Datos que describen otros datos, proporcionando información sobre su origen, formato y calidad.

Filtrado: Proceso de seleccionar un subconjunto de datos que cumple con ciertos criterios.

Limpieza de Datos: Proceso de identificar y corregir errores en los datos para mejorar su calidad y utilidad.

Herramientas de Limpieza de Datos: Software o aplicaciones (como OpenRefine, Excel, Python) que facilitan el proceso de limpieza de datos.

Análisis Exploratorio de Datos (EDA): Técnica utilizada para analizar conjuntos de datos y resumir sus principales características, a menudo utilizando visualizaciones.

Consolidación de Datos: Proceso de combinar datos de múltiples fuentes en un solo conjunto coherente.

6. REFERENTES BILBIOGRÁFICOS

- Data Cleaning: Problems and Current Approaches"
  Autores: A. K. Jain, R. C. Jain
  Descripción: Este libro aborda los problemas comunes en la limpieza de datos y presenta enfoques actuales para solucionarlos.
- "Data Wrangling with pandas"
  Autor: Jacqueline Kazil, Katharine Jarmul
  Descripción: Una guía práctica sobre cómo utilizar la biblioteca Pandas de Python para limpiar y transformar datos.

7. CONTROL DEL DOCUMENTO

	Nombre	Cargo	Dependencia	Fecha
Autor (es)	**Edgar Vernaza arboleda**	**INSTRUCTOR**	**COMERCIO Y SERVICIOS**	**1/08/2024**

8. CONTROL DE CAMBIOS (diligenciar únicamente si realiza ajustes a la guía)

	Nombre	Cargo	Dependencia	Fecha	Razón del Cambio
Autor (es)					