**Evaluación Sumativa 4**

**Modulo 3: Transformación y limpieza de datos en Power BI Desktop**

**Objetivo de la evaluación:** El objetivo de esta evaluación es aplicar los conocimientos adquiridos en el Módulo 4 para realizar tareas de transformación y limpieza de datos utilizando Power Query en Power BI Desktop. Los estudiantes deberán demostrar su habilidad para eliminar filas duplicadas, cambiar tipos de datos, quitar columnas innecesarias, rellenar valores faltantes, y crear nuevas columnas o tablas a partir de un dataset proporcionado.

**Instrucciones:**

1. **Descargar el Dataset:**
   * Buscar una data con necesidad de limpieza y transformaciones. Descarga el archivo en tu computadora.
2. **Abrir Power BI Desktop:**
   * Abre Power BI Desktop e importa el archivo de módulos anteriores.
3. **Panel de Consultas: Vista Previa de los Datos:**
   * Accede al panel de consultas y revisa la vista previa de los datos. Verifica que los datos se hayan importado correctamente.
4. **Eliminar Filas Duplicadas:**
   * Identifica y elimina cualquier fila duplicada en el dataset.
   * Pasos: Selecciona las columnas relevantes para la eliminación de duplicados y usa la opción "Eliminar duplicados".
5. **Cambiar Tipos de Datos:**
   * Cambia los tipos de datos de las columnas para que coincidan con la naturaleza de los datos. Por ejemplo, convierte columnas de texto a formato de fecha o numérico.
   * Pasos: Ajusta los tipos de datos desde el menú "Tipo de datos".
6. **Quitar Columnas Innecesarias:**
   * Elimina las columnas que no sean relevantes para el análisis. Por ejemplo, elimina columnas de identificadores internos que no aportan valor al análisis.
   * Pasos: Selecciona las columnas innecesarias y usa la opción "Eliminar columnas".
7. **Rellenar Valores Faltantes:**
   * Rellena los valores faltantes en el dataset utilizando métodos apropiados, como reemplazar con un valor promedio o un valor predeterminado.
   * Pasos: Usa las opciones "Rellenar hacia abajo" o "Rellenar hacia arriba", o reemplaza valores nulos manualmente.
8. **Crear Nuevas Columnas o Tablas:**
   * Crea al menos una nueva columna calculada o una nueva tabla basada en los datos existentes. Por ejemplo, crea una columna que calcule el total de ventas incluyendo impuestos.
   * Pasos: Usa la opción "Agregar columna" y aplica fórmulas para calcular nuevos valores.
9. **Guardar y Compartir:**
   * Guarda el archivo de Power BI con todas las transformaciones realizadas.
   * Sube el archivo a una plataforma de almacenamiento en la nube (GitHub) y comparte el enlace en la plataforma asignada.
   * Entregue un Word documentando paso a paso todas las modificaciones realizadas en Power BI.

**Criterio de Evaluación:**

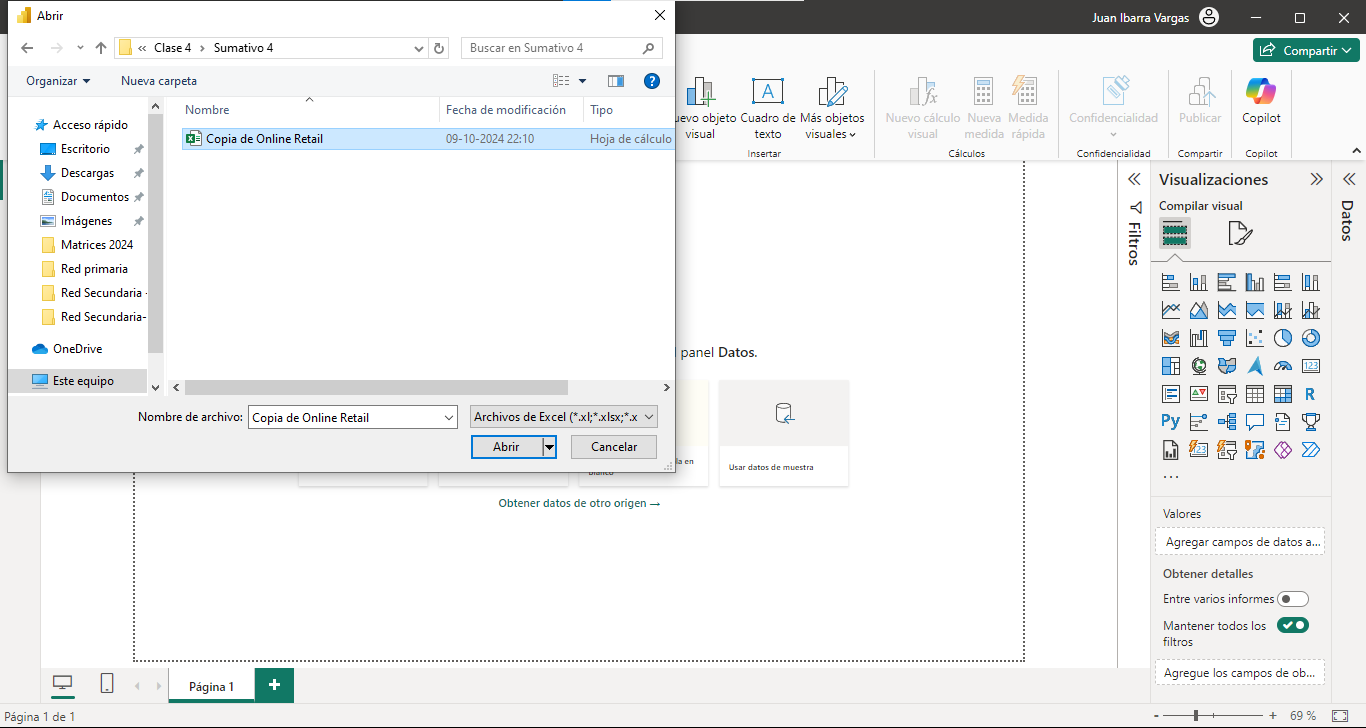
1. **Panel de Consultas: Vista Previa de los Datos (10 puntos)**
   * Verificación correcta de la vista previa de los datos y la estructura de los datos importados.
2. **Eliminar Filas Duplicadas (15 puntos)**
   * Identificación y eliminación efectiva de filas duplicadas, con evidencia clara de los cambios realizados.
3. **Cambiar Tipos de Datos (15 puntos)**
   * Ajuste adecuado de los tipos de datos en el dataset para asegurar la correcta visualización y análisis.
4. **Quitar Columnas Innecesarias (15 puntos)**
   * Eliminación efectiva de columnas no relevantes para el análisis, con evidencia de la reducción del conjunto de datos.
5. **Rellenar Valores Faltantes (15 puntos)**
   * Relleno adecuado de valores faltantes utilizando métodos apropiados, con evidencia de la corrección aplicada.
6. **Crear Nuevas Columnas o Tablas (15 puntos)**
   * Creación efectiva de nuevas columnas o tablas basadas en datos existentes, con cálculos correctos y relevantes.
7. **Guardar y Compartir (10 puntos)**
   * Archivo guardado correctamente y enlace compartido en la plataforma asignada.
8. **Calidad General del Trabajo (5 puntos)**
   * Coherencia y claridad en la aplicación de transformaciones, y precisión en la ejecución de las tareas.

**Total: 100 puntos**

**Notas Adicionales:**

* Asegúrate de revisar todas las transformaciones y ajustes antes de entregar el ejercicio.
* La precisión en la aplicación de las técnicas de transformación y limpieza de datos será evaluada en detalle.

¡Éxito en el desarrollo!



**Paso 2**: Buscamos la data descargada y la abrimos en Power-BI.

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

**Paso 3**: Una vez cargada la data, escogemos la pestaña que queremos ocupar (que en este caso nuestro archivo solo tiene una), y la previsualizamos en el navegador de Power-BI, donde podemos apreciar que se cargaron los datos correctamente. Luego apretamos el botón “transformar datos”, para pasar a limpiar la data.

Tabla

Descripción generada automáticamente

**Paso 4**: Como se aprecia en el screenshot, en la esquina superior izquierda de la tabla, se puede encontrar la opción de “quitar duplicados”, lo cual nos permitiría quitar de la tabla las filas que estén duplicadas.

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

**Paso 5**: Comenzamos la depuración de la data; la primera columna “VoiceNO” que estaba categorizada como “números y letras”, corresponde al número de factura, por lo que es una identificación y la colocamos como “texto”.

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

**Paso 5**: En el caso de la columna “StockCode”, inferimos que es un código de identificación por lo que también le cambiados el tipo, dejándola como “texto”.

Captura de pantalla de computadora

Descripción generada automáticamente

**Paso 5**: En el caso de la columna “UnitPrice”, que está como número decimal, la cambiamos a ”número decimal fijo” que es el formato utilizado cuando queremos categorizar los datos como moneda.

Interfaz de usuario gráfica, Tabla, Excel

Descripción generada automáticamente

**Paso 5**: En el caso de la columna CustomerID, se entiende que es un número identificador por lo que la colocamos como “texto”, finalizando así el paso 5.

Tabla, Excel

Descripción generada automáticamente

**Paso 6**: En este caso, nuestra data no posee columnas duplicadas, pero para efectos del ejercicio, duplicamos la última columna, para luego borrarla.

Tabla

Descripción generada automáticamente

**Paso 6**: Dando click derecho sobre el título de la columna, seleccionamos la opción de “Quitar”, eliminando así esta columna creada.

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

**Paso 7**: En la columna “CustomeID” vamos a reemplazar los valores que estén como “null” para dejarlas como “No conocido”. Para lo anterior, hacemos click derecho sobre el nombre de la columna y nos dirigimos a la opción de “reemplazar valores”.

Interfaz de usuario gráfica, Tabla, Excel

Descripción generada automáticamente

**Paso 7**: Podemos apreciar como reemplazó los valores, quedando ahora como “No Conocido”.

Interfaz de usuario gráfica, Tabla, Excel

Descripción generada automáticamente

**Paso 8**: Para agregar una columna lo haremos desde la vista tabla, para eso primero le damos al botón de “Cerrar y aplicar”.

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

Paso 8: Una vez cargados los datos, nos vamos a la vista de tabla y vamos al botón de “Agregar columna”.

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Paso 8: En la barra de fórmulas escribimos la operación que deseamos, en este caso que multiplique la columna Quantity por UnitPrice.

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

**Paso 8**: Vemos como nos creó una nueva columna donde se realiza la operación descrita anteriormente.

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

**Paso 8**: Luego le cambiamos el nombre a la columna y le ponemos “TOTAL”.

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Explicación: Podemos apreciar que existen cantidades negativas, todas ellas tienen un valor unitario equivalente a cero, por lo que en el total nos equivale también a cero y no nos afecta en nuestro análisis de datos