Curso Power BI Básico

evaluación SUMATIVA módulo 4: Transformación y limpieza de datos en Power BI Desktop

Rodrigo Sotelo Valenzuela

2024

1. **INTRODUCCIÓN**

En el siguiente trabajo sumativo 4, se usará la misma base de datos descargada en el Módulo 2 para practicar con las distintas funciones y herramientas que nos proporciona Power Query para la transformación y limpieza de los datos.

En el **Anexo A**, se presenta todo lo realizado en el trabajo sumativo 2, esto es, el contexto de la dataset seleccionada, descripción y clasificación de las variables (columnas), creación de un repositorio en Github, proceso de importación de datos, transformación de datos en Power Query y configuración del parámetro de la ruta de acceso al archivo fuente de los datos.

En el **Anexo B**, se presenta todo lo realizado en el trabajo sumativo 3, esto es, la creación de visualizaciones y uso de distintos objetos que permiten presentar los datos de manera más efectiva.

1. **VISTA PREVIA DE LOS DATOS**

**Figura1:** Importación de datos y visualización en Power Query previo a la carga.

Pantalla de computadora

Descripción generada automáticamente con confianza media

1. **ELIMINAR FILAS DUPLICADAS**

Originalmente son 16.573 registros. Se elimina solo un registro.

**Figura2:** Eliminación de filas duplicadas a nivel tabla.

Pantalla de computadora

Descripción generada automáticamente con confianza media

1. **CAMBIAR TIPOS DE DATOS**

Se revisa variable por variable, como lo ha indicado la profesora en clases, para clasificarlas de forma correcta en Cuantitativas o Cualitativas. En el panel de “Vista” se añade, además, un visualizador de la calidad de las variables (columnas). Solo se cambió el tipo de variable “Rank” a Texto. No se visualizan celdas vacías.

**Figura3:** Cambio de tipo de la variable Rank.

Captura de pantalla de computadora

Descripción generada automáticamente

1. **QUITAR COLUMNAS INNECESARIAS**

Se eliminan 2 variables innecesarias: “Rank” y “Global\_Sales”. Originalmente se importaron 11 variables.

**Figura4:** Procesos de eliminación de variables innecesarias.

Una captura de pantalla de una computadora

Descripción generada automáticamente

Captura de pantalla de computadora

Descripción generada automáticamente

1. **RELLENAR VALORES FALTANTES**

Se analizó y no se detectaron valores faltantes en la base de datos.

1. **CREAR NUEVAS COLUMNAS O TABLAS**

En el panel Agregar columna, se selecciona la opción “columna personalizada” para agregar una columna nueva calculada a partir de la suma de las columnas NA\_Sales, EU\_Sales, JP\_Sales y Other\_Sales. La nueva columna se llamará **Global\_Sales\_Calc.** En total, quedan 10 variables para uso.

**Figura5:** Procesos de creación de una nueva columna.

Una captura de pantalla de una computadora

Descripción generada automáticamente Interfaz de usuario gráfica, Tabla, Excel

Descripción generada automáticamente

1. **GUARDAR Y COMPARTIR**

El trabajo en Power BI se guarda con el siguiente nombre: “Evaluación Sumativa4\_Rodrigo Sotelo.pbix”.

Además, se suben al repositorio de Github, detallado en el Anexo A, los tres archivos solicitados por la profesora.

ANEXO A

1. **CONTEXTO DE LA DATASET SELECCIONADA**

La base de datos que se seleccionó contiene una lista de videojuegos cuyas ventas son superiores a 100.000 copias, entre los años 1980 y 2020. Este archivo obtenido desde el sitio kaggle.com contiene 11 campos y un total de 16.598 registros. El nombre del archivo es “vgsales.csv”.

Link: https://www.kaggle.com/datasets/gregorut/videogamesales

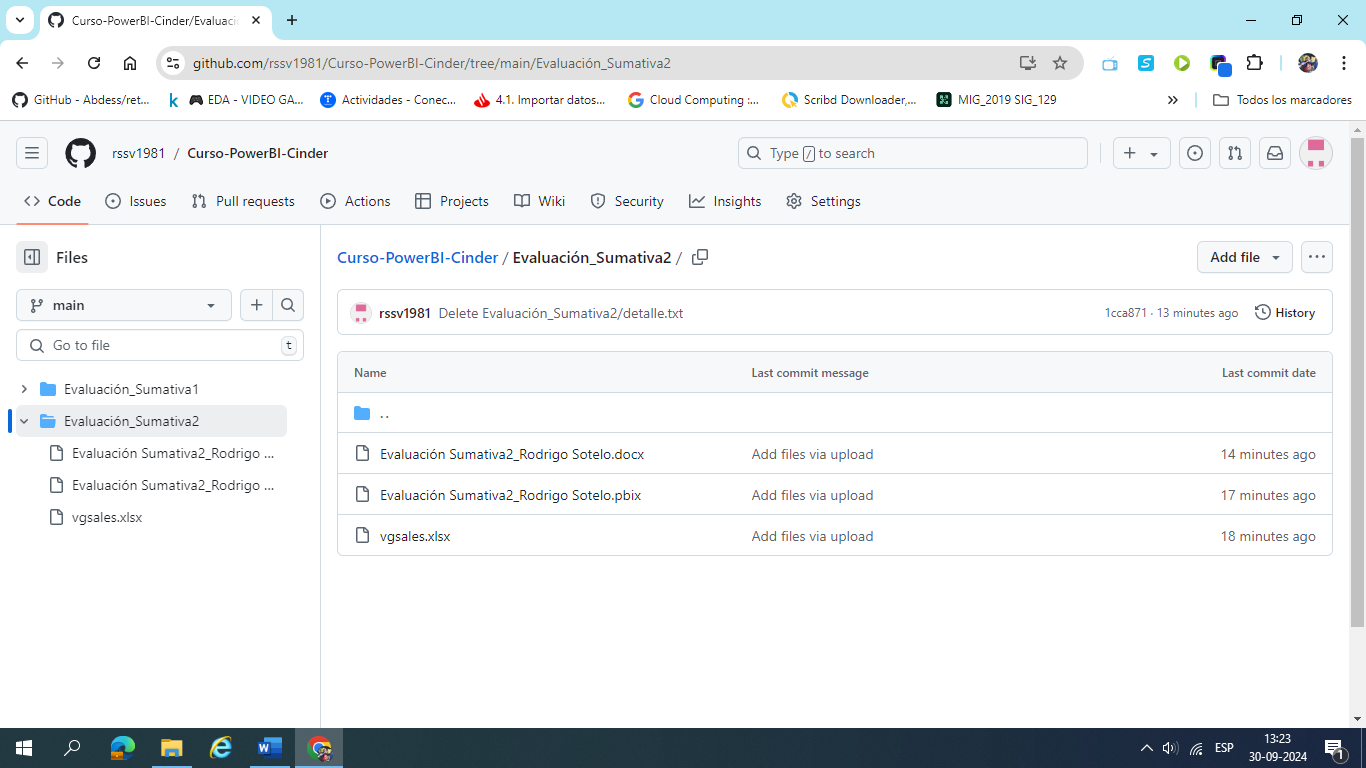
1. **VARIABLES Y DESCRIPCIÓN**

La base de datos contiene las siguientes 11 variables (campos), 6 cualitativas y 5 cuantitativas:

* **Rank**: Número correlativo que ordena los juegos de acuerdo con el campo Global\_Sales de forma descendente. (Cualitativa)
* **Name**: Nombre del juego. (Cualitativa)
* **Platform**: Nombre de la plataforma de lanzamiento (i.e. PC, PS4, etc.). (Cualitativa)
* **Year**: Año de lanzamiento del juego. (Cualitativa)
* **Genre**: Género del juego. (Cualitativa)
* **Publisher**: Empresa editora del juego. (Cualitativa)
* **NA\_Sales**: Ventas en Norte América (en millones). (Cuantitativa)
* **EU\_Sales**: Ventas en Europa (en millones). (Cuantitativa)
* **JP\_Sales**: Ventas en Japón (en millones). (Cuantitativa)
* **Other\_Sales**: Ventas en el resto del mundo (en millones). (Cuantitativa)
* **Global\_Sales**: Total de ventas en el mundo (en millones). (Cuantitativa)

1. **ENLACE A GITHUB Y PANTALLAZO DE REPOSITORIO CREADO**

<https://github.com/rssv1981/Curso-PowerBI-Cinder.git>



1. **PROCESO DE IMPORTACIÓN DE DATOS, CONFIGURACIÓN DEL PARÁMETRO, Y LA TRANSFORMACIÓN DE DATOS EN POWER QUERY.**

* Paso1: Proceso de transformación, limpieza y formateo de datos en Excel.

**Figura1:** El archivo es descargado en formato csv, y se abre con excel.Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

**Figura2:** Se procede a separar cada variable en una columna con la función “Texto en columnas”.

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

**Figura3:** Se transforman los valores de las variables Cuantitativas a decimales (con coma).

Tabla

Descripción generada automáticamente

**Figura4:** Se detectan juegos sin información (campos vacíos) en el resto de las variables y se toma la decisión de eliminarlos de la base de datos. Para el análisis quedan 16.572 registros.

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

**Figura5:** Se procede a formatear como “Texto” y “Número” a los campos cualitativos y los cuantitativos respectivamente. Los numéricos se formatean con dos decimales. Se guarda el archivo con la extensión .xlsx.

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

* Paso2: Proceso de importación del archivo Excel en Power BI, transformación y carga.

**Figura6:** Importación del archivo Excel “vgsales.xlsx” en Power BI.

Captura de pantalla de computadora

Descripción generada automáticamente

**Figura7:** Proceso de transformación de datos en Power Query previo a la carga.

Imagen de la pantalla de un computador

Descripción generada automáticamente con confianza baja

Solo el campo “Rank” tuvo que convertirse a formato texto.

**Figura8:** Proceso de carga de datos al presionar el botón “Cerrar y Aplicar”.

Una captura de pantalla de una computadora

Descripción generada automáticamente

* Paso3: Creación de visualizaciones y configuración de parámetro.

**Figura9:** Creación de visualizaciones (tablas y gráficos de barra) en la página Tablero1 del Power BI.

Captura de pantalla de computadora

Descripción generada automáticamente

**Figura10:** Configuración del parámetro llamado “Ruta” que contiene la ruta original del archivo fuente de los datos.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

**Figura11:** Configuración de la ruta origen de la base de datos mediante el parámetro creado llamado “Ruta”.

Interfaz de usuario gráfica, Aplicación, Tabla, Word

Descripción generada automáticamente

ANEXO B

1. **PANEL DE CONSULTAS**

El proceso de visualización, limpieza de los datos y de la correcta clasificación de las variables se realizó en el trabajo sumativo 2. Ver Anexo A.

1. **TIPOS DE DATOS Y FORMATOS**

El proceso de visualización, limpieza de los datos y de la correcta clasificación de las variables se realizó en el trabajo sumativo 2. Ver Anexo

1. **CREACIÓN DE TABLAS RESÚMENES Y FILTROS**

En la página **Tablero1** de Power BI se crean tablas y gráficos de barras para visualizar y comparar el género más vendido de juegos en Europa, Japón y a nivel mundial. Luego se aplica un filtro a toda la página para solo visualizar el comportamiento del año 2015 y los géneros con más ventas.

**Figura1:** Creación de visualizaciones (tablas y gráficos de barra).

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

**Figura2:** Creación de filtro para visualizar solo el comportamiento del año 2015.

Captura de pantalla de computadora

Descripción generada automáticamente

1. **USO DE TARJETAS**

En la página **Tablero2** se crean 5 tarjetas que varían su valor (Suma) según los valores de los campos (año, género y plataforma) seleccionados. Las tarjetas muestran: ventas de Europa, ventas de Norte América, ventas de Japón, Otras Ventas y ventas a nivel global.

**Figura3:** Creación de la tarjeta que muestra el total (Suma) de ventas a nivel global.

Interfaz de usuario gráfica

Descripción generada automáticamente

En el panel de visualizaciones del objeto, se coloca el campo “Global\_Sales” en el apartado de Campos.

1. **USO DE MATRICES**

En la misma página **Tablero2** se crea 1 matriz que muestra las ventas (en millones) por género y años, variando según los valores de los campos (año y plataforma) seleccionados.

**Figura4:** Creación de matriz de ventas a nivel global, por género y años.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

En el panel de visualizaciones del objeto, se coloca el campo “Genre” en el apartado de Filas, “Year” en Columnas y “Global\_Sales” en Valores.

1. **USO DE MEDIDOR**

En la misma página **Tablero2** se crean 3 medidores que muestran las ventas globales (en millones) de los años 2014, 2015 y 2016, variando según los valores de los campos (plataforma y género) seleccionados.

**Figura5:** Creación de medidor de las ventas a nivel global del año 2014.Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

En el panel de visualizaciones del objeto, se coloca el campo “Global\_Sales” en el apartado de Valor y en el apartado Valor Máximo. Además, en los filtros del objeto visual se coloca la variable “Year” y se selecciona el año 2014.

1. **USO DE SEGMENTADOR**

En la misma página **Tablero2** se crean 2 segmentadores de datos que muestran los posibles valores seleccionables de las variables “Year” y “Platform”. Estos permiten filtrar los datos de ventas y les dan dinamismo a los objetos creados en la hoja.

**Figura5:** Creación de segmentador de datos de la variables “Platform”.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

En el panel de visualizaciones del objeto, se coloca el campo “Platform” en el apartado Campo.

1. **GUARDAR Y COMPARTIR**

El trabajo en Power BI se guarda con el siguiente nombre: “Evaluación Sumativa3\_Rodrigo Sotelo.pbix”.

Además, se suben al repositorio de Github, detallado en el Anexo A, los tres archivos solicitados por la profesora.