Curso   
Power BI Básico

Módulo3:  
Exploración de Datos Básica

Autor: Mauro Chávez Pérez

Fecha: 07.OCT.2024

Notas para la revisión:

De la página 1 a la 11, este informe da cuenta del Módulo 2.

[De la página 12 a la 19, este informe da cuenta del Módulo 3](#Marca2).

1. Contexto de la dataset seleccionada

El conjunto de datos seleccionados, Student Perfomance Factors, está asociado con la información sobre el desempeño de los estudiantes y los factores que contribuyen. Proporciona una descripción general completa de varios factores que afectan el desempeño de los estudiantes en los exámenes. Incluye información sobre hábitos de estudio, asistencia, participación de los padres y otros aspectos que influyen en el éxito académico.

1. Variables y descripción

| **Variable / Attribute** | **Descripción / Description** |
| --- | --- |
| **Hours\_Studied** | Number of hours spent studying per week. |
| **Attendance** | Percentage of classes attended. |
| **Parental\_Involvement** | Level of parental involvement in the student's education (Low, Medium, High). |
| **Access\_to\_Resources** | Availability of educational resources (Low, Medium, High). |
| **Extracurricular\_Activities** | Participation in extracurricular activities (Yes, No). |
| **Sleep\_Hours** | Average number of hours of sleep per night. |
| **Previous\_Scores** | Scores from previous exams. |
| **Motivation\_Level** | Student's level of motivation (Low, Medium, High). |
| **Internet\_Access** | Availability of internet access (Yes, No). |
| **Tutoring\_Sessions** | Number of tutoring sessions attended per month. |
| **Family\_Income** | Family income level (Low, Medium, High). |
| **Teacher\_Quality** | Quality of the teachers (Low, Medium, High). |
| **School\_Type** | Type of school attended (Public, Private). |
| **Peer\_Influence** | Influence of peers on academic performance (Positive, Neutral, Negative). |
| **Physical\_Activity** | Average number of hours of physical activity per week. |
| **Learning\_Disabilities** | Presence of learning disabilities (Yes, No). |
| **Parental\_Education\_Level** | Highest education level of parents (High School, College, Postgraduate). |
| **Distance\_from\_Home** | Distance from home to school (Near, Moderate, Far). |
| **Gender** | Gender of the student (Male, Female). |
| **Exam\_Score** | Final exam score. |

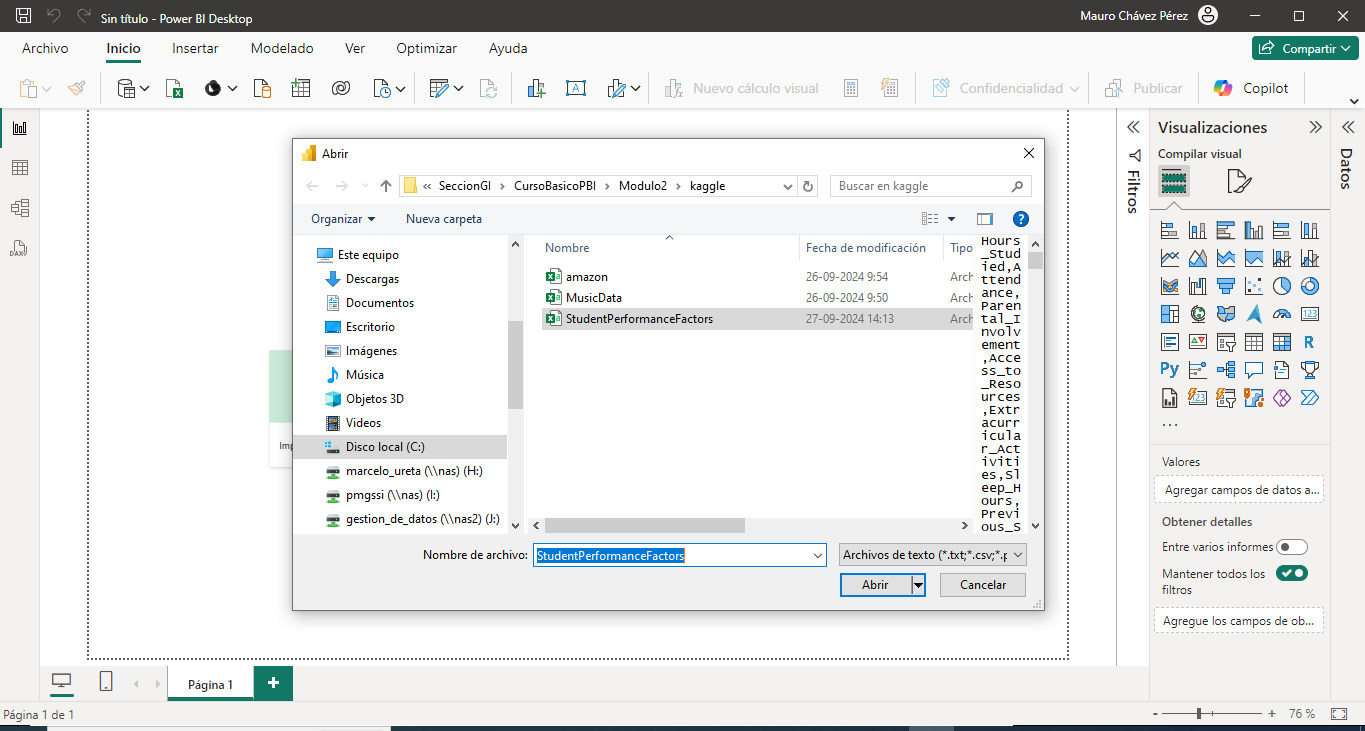
1. Enlace a GitHub

El enlace al repositorio en GitHub, donde está alojado el dataset Student Perfomance Factors es el siguiente:

<https://github.com/mgchavezp/Power-BI-Basico.git>

1. Capturas de pantalla

Importación de datos (Accesa localmente el dataset obtenido en Kaggle)



Importación datos (Muestra la vista del archivo una vez seleccionado)

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Revisión y Transformación

Se revisaron todas las columnas en cuanto a los formatos se refiere, como también los eventuales errores y celdas vacías o en blanco.

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

Transformación de celdas en blanco en columna Teacher\_Quality.

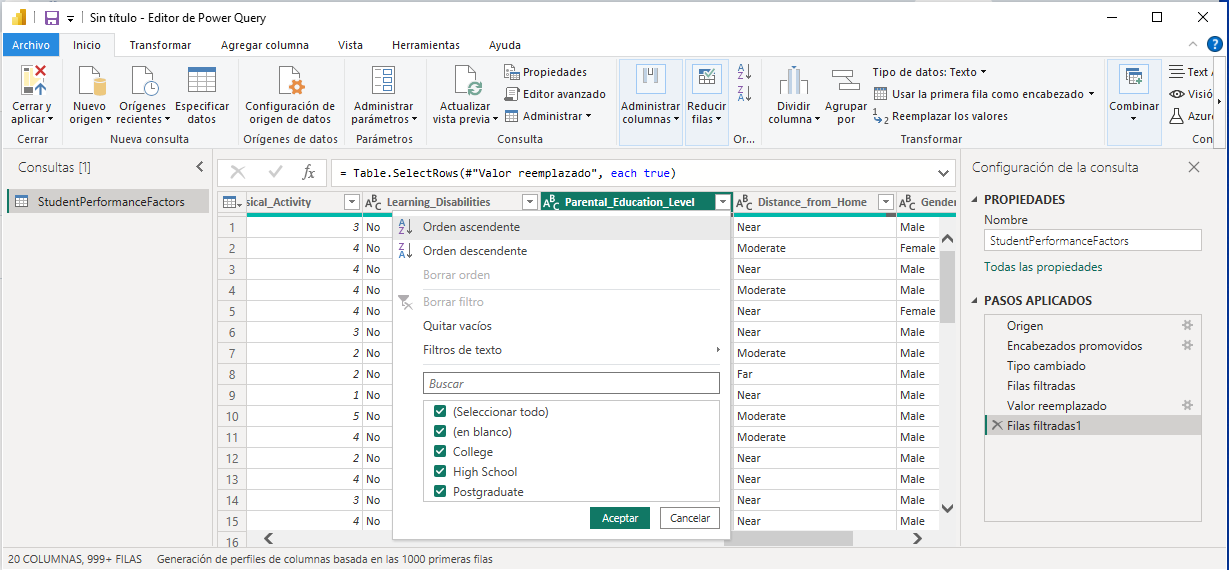
Interfaz de usuario gráfica, Tabla, Excel

Descripción generada automáticamente

Una captura de pantalla de una computadora

Descripción generada automáticamente

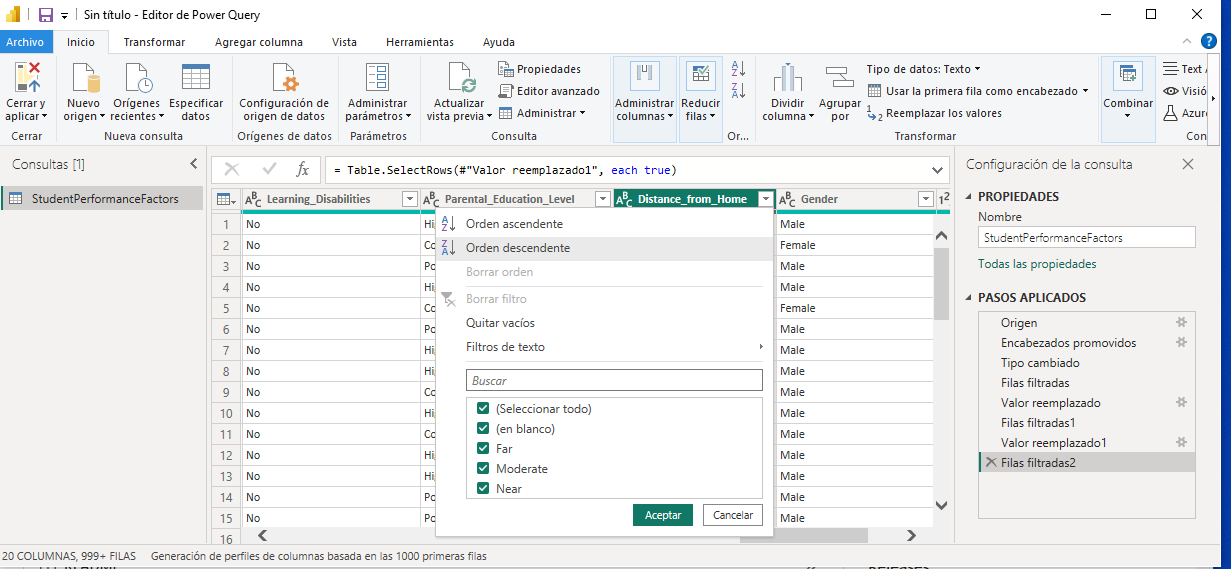
Transformación de celdas en blanco en columna Parental\_Education\_Level.



Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Revisión de celdas en blanco en columna Distance\_from\_Home



Transformación de celdas en blanco por dato válido. Todas las celdas en blanco del dataset quedaron con el dato “Sin información”.

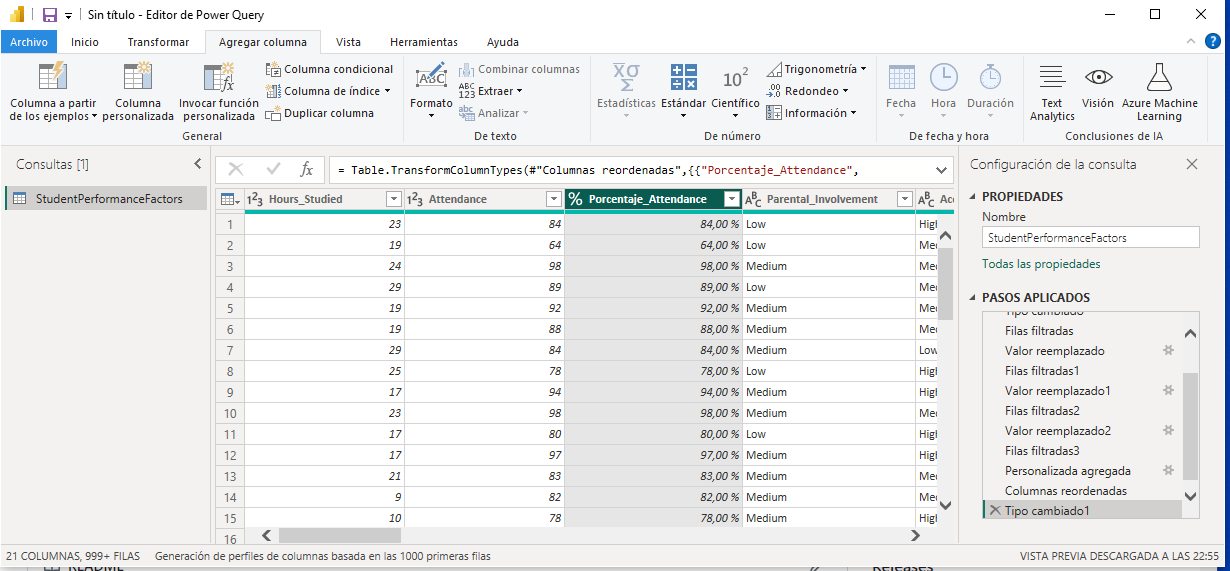
Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

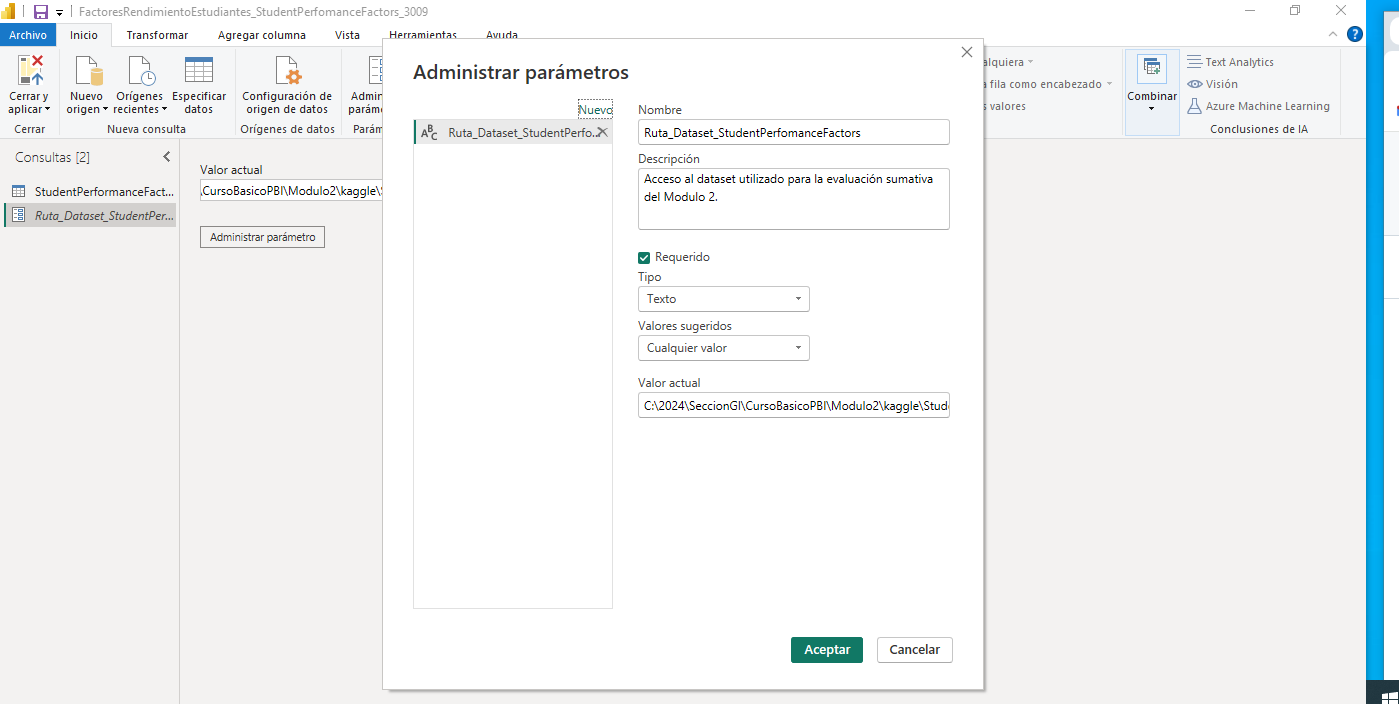
Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

Transformación de la columna Attendance, que se encuentra con valores en formato entero a valores en formato porcentaje (%), para su uso más directo en posibles cálculos posteriores.



Definición del parámetro como ruta de acceso al dataset utilizado para la evaluación sumativa. Parámetro se denominó Ruta\_Dataset\_StudentPerfomanceFactors



Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

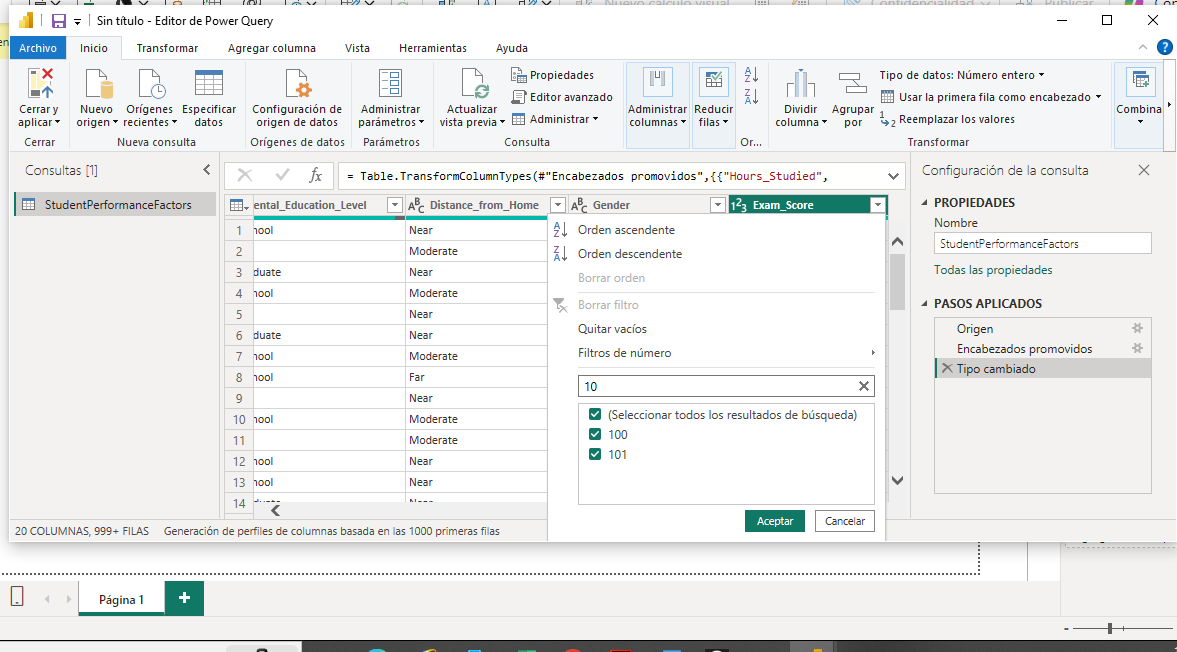
**A partir de aquí comienza el desarrollo de las actividades del Módulo 3 – Exploración de Datos Básica.**

Las actividades 1, 2 y 3 se entienden desarrolladas en el Módulo 2.

El Dataset utilizado es el mismo del módulo 2, StudentPerformanceFactors.csv

**Actividad 4. Tipos de datos y formatos:**

Si bien la revisión y ajustes de los formatos y tipos de datos se desarrollaron en el modulo anterior, para este módulo, se procedió a la eliminación de una fila que tenía el dato correspondiente a la nota del examen (Exam\_Score) mayor que 100, suponiendo que el máximo posible es 100.



Se ordenó de menor a mayor, para dejar la fila con la nota 101 como última fila y se procedió a eliminarla.

Una captura de pantalla de una red social

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

**Actividad 5. Creación de tablas resúmenes y filtros**

Tablas de Resumen, para Promedio de Notas por Calidad de Profesor y Género; y para Promedio de Notas por Género.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Filtros para las Horas de Estudio (Hours\_Studied) y Actividad Física (Physical\_Activity).

Interfaz de usuario gráfica, Aplicación, Excel

Descripción generada automáticamente

**Actividad 6. Uso de tarjetas:**

Se generaron dos tarjetas, una para indicar el total de estudiantes y la otra para el promedio de las notas. Para ambos casos se requirió realizar cálculos previos los que quedan registrada con nombres similares en el modelo. Para el total de personas se eligió la columna Gender, ya que no tenía errores ni celdas que corregir.

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

Actividad 7. Uso de Matrices:

Se elaboró una matriz para mostrar diferentes promedios de factores y género

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Actividad 8. Uso de Medidor:

Se incorporó un medidor para el campo Promedio de Asistencia, previamente calculado.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Actividad 8. Uso de Segmentador:

Se incorporó un segmentador para el campo Exam\_Score.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Actividad 11. (Bonus Track) Se intentó elaborar un histograma de los datos del campo Exam\_Score y su relación con el número de personas por rango.

Para esta actividad se realizaron los siguientes pasos

1. Creación de un gráfico de barras apiladas con Exam\_Score en el eje X y el Total de Personas en el eje Y.

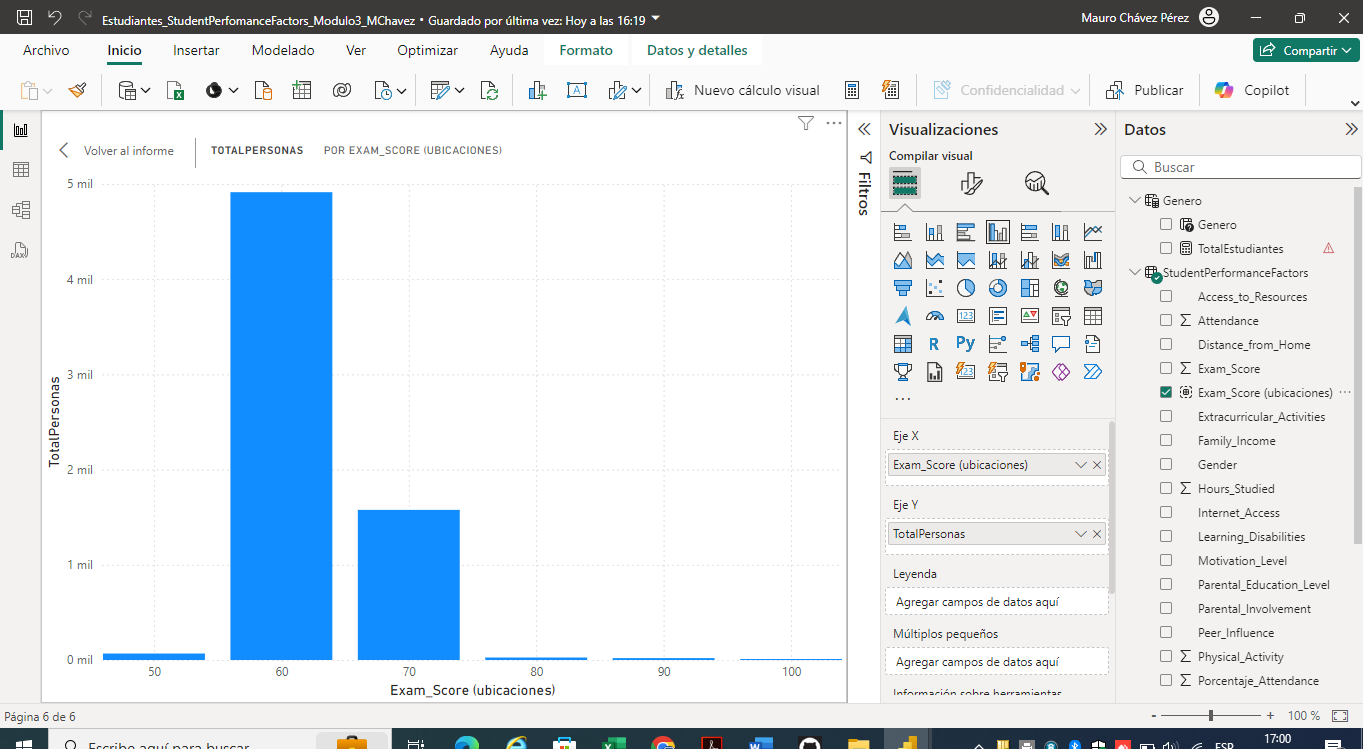
Gráfico, Histograma

Descripción generada automáticamente

1. Se discretizó el eje X con valores de 10 en 10 con el nombre de Grupo Exam\_Score (ubicaciones) y se reemplazó este Grupo en el eje X

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente



1. Se agregó una columna condicional para establecer los rangos de notas, la cual debe quedar con formato texto

Una captura de pantalla de una computadora

Descripción generada automáticamente

Interfaz de usuario gráfica, Tabla, Excel

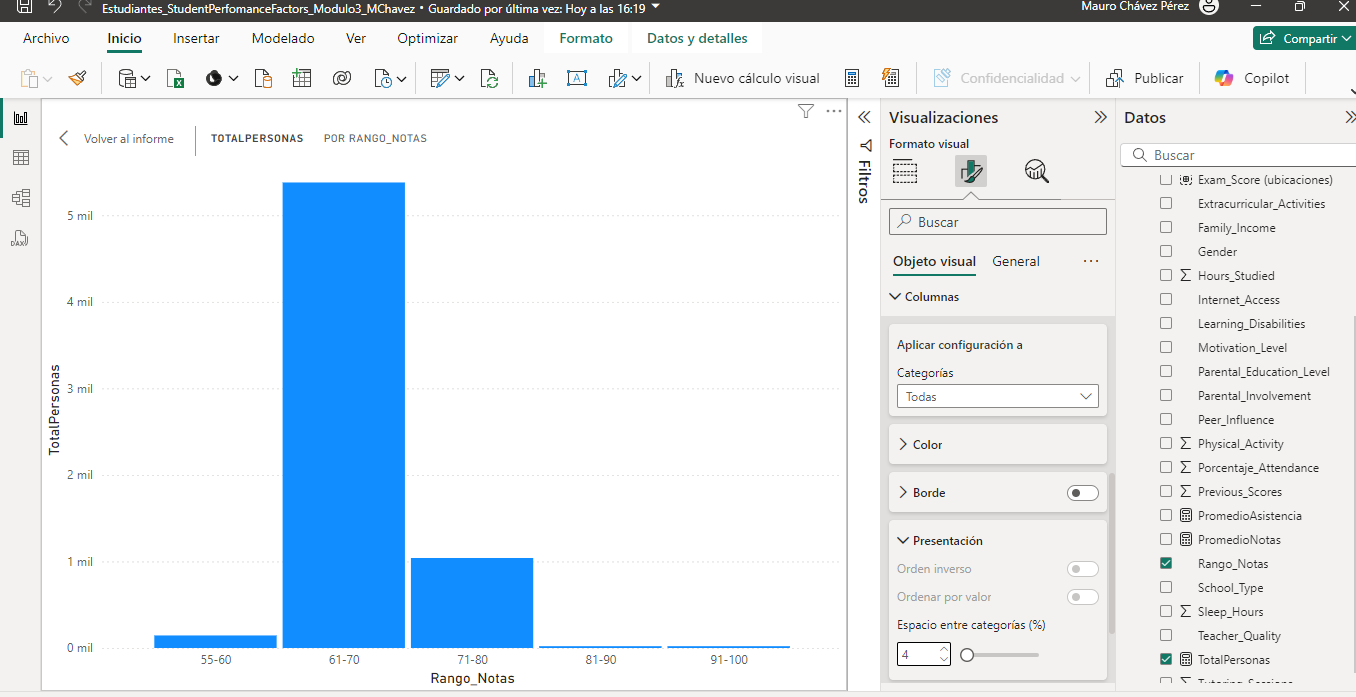
Descripción generada automáticamente

1. La columna Rango\_Notas debe reemplazar a lo que hay en el eje X

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. A continuación, se ordena el eje X y se acomodan las columnas con sus etiquetas de datos.



Gráfico

Descripción generada automáticamente