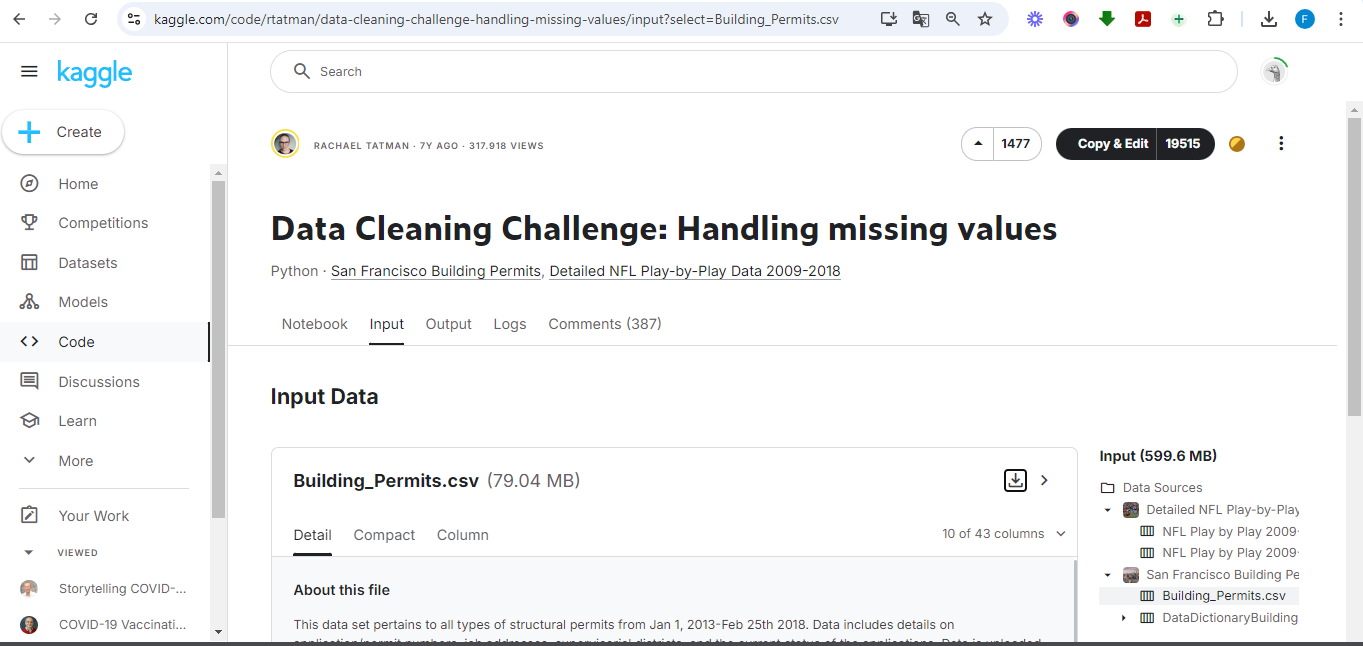
**EVALUACIÓN SUMATIVA NÚMERO 4**

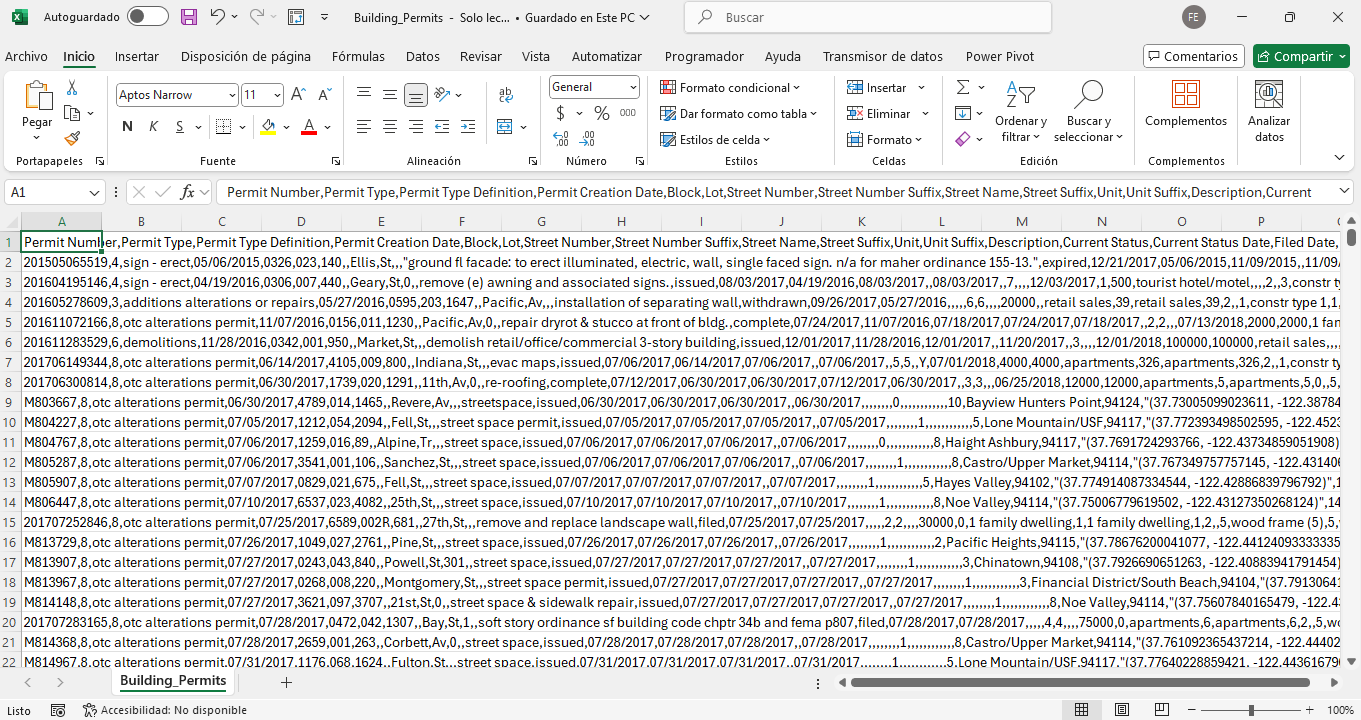
**INTEGRANTE: FLAVIO EICHIN CAMPOS.**

1. **Descargar la base de Datos**

* Buscar una data con necesidad de limpieza y transformaciones. Descarga el archivo en tu computadora.



Se buscó en la plataforma Kaggle.com dataset que tuviesen retos de limpieza de datos. Al respecto escogí el archivo Building\_Permits.csv.



Posteriormente, realicé la transformación del archivo delimitado separado por comas utilizando el Menú Datos/Texto en columnas de MS Excel.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Utilizando el asistente para convertir texto en columnas en el paso 1 de 3, se elije el tipo de dato “Delimitados”.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Luego, en el paso 2 de 3 del asistente, se marca el separador “Coma”.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Finalmente, en el paso 3 de 3 del asistente, se verifican el formato de los datos de las columnas.

Interfaz de usuario gráfica

Descripción generada automáticamente

En este apartado se les da formato a los siguientes campos:

* Permit Number: Texto.
* Permit Type: Texto.
* Permit Type Definition: Texto.
* Permit Creation Date: Texto.
* Block: Texto.
* Lot: Texto.
* Street Number: Texto.
* Street Number Suffix: Texto.
* Street Name: Texto.
* Street Suffix: Texto.
* Unit: Texto.
* Unit Suffix: Texto.
* Description: Texto.
* Current Status: Texto.
* Current Status Date: Texto.
* Filed Date: Texto.
* Issued Date: Texto.
* Completes Date: Texto.
* First Construction Document Date: Texto.
* Structural Notifications: Texto.
* Number of existing stories: General.
* Number of proposed stories: General.
* Voluntary Soft-Story retorfit: General.
* Fire Only Permit: General.
* Permit Expiration Date: Texto.
* Estimated Cost: General.
* Revised Cost: General.
* Existing Use: Texto.
* Existing Units: General.
* Proposed Use: Texto.
* Proposed Units: General.
* Plansets: Texto.
* TIDF Compliance: Texto.
* Existing Construction Type: Texto.
* Existing Construction Type Description: Texto.
* Proposed Construction Type: Texto:
* Proposed Construction Type Description: Texto.
* Site Permit: Texto.
* Supervisor District: Texto.
* Neighborhood – Analysis Boundaries: Texto.
* Zipcode: Texto.
* Location: Texto.
* Record: Texto.

De acuerdo a lo visto en las clases, los campos de fechas se definieron con formato de texto, para unificar el criterio MM/DD/AAAA.

1. **Abrir Power BI Desktop**

* Abre Power BI Desktop e importa la base de datos descargada**.**

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

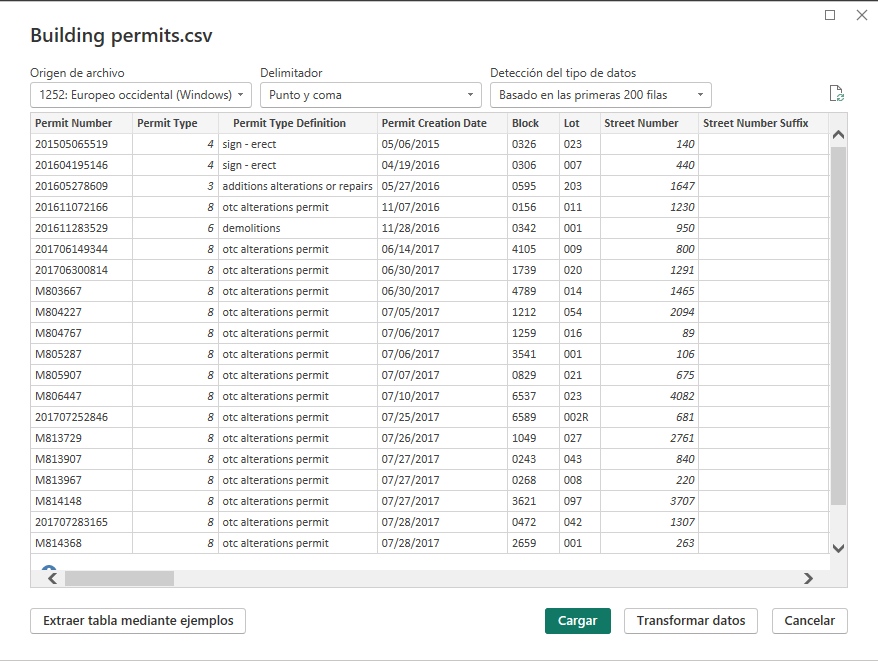
Luego se importa la base de datos descargada y previamente trabajado según el punto 1 del presente trabajo.

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

1. **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.

****

Se verifica que los datos fueron importados de manera correcta.

Interfaz de usuario gráfica, Tabla, Excel

Descripción generada automáticamente

1. **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".

En la base de datos, no se observan columnas duplicadas.

Interfaz de usuario gráfica

Descripción generada automáticamente

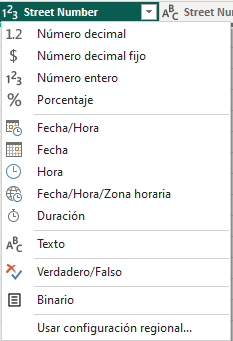
1. **Cambiar Tipo de Datos:**

* 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".

Se observa que el campo Permit Type, Street Number y Plansets, Existing Construction Type, Proposed Construction Type, Supervisor District y Zipcode, se cargaron como Número Entero, y se cambian a Tipo de dato Texto.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente



Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

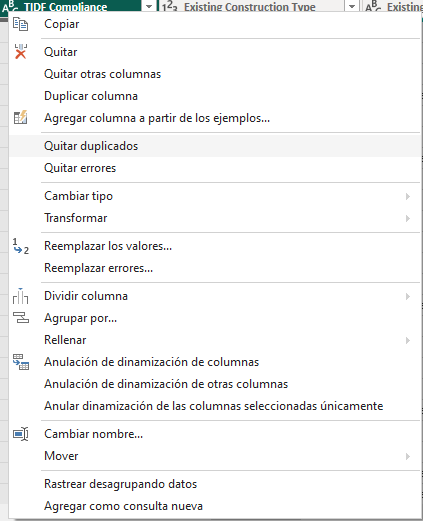
Descripción generada automáticamente

1. **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".

De la base de datos, que tiene 198.901 observaciones, se puede determinar que los siguientes campos no aportan valor al análisis**:**

* TIDF Compiance, posee 3 observaciones solamente.



1. **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.

El campo Unit, posee datos null. Como no se tiene certeza del significado del campo, ya que posee número aleatorios, se reemplaza por indeterminado. Además, se cambiará el formato de la variable por Texto.

**Interfaz de usuario gráfica, Texto

Descripción generada automáticamente**

**Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media**

El campo Current Status, posee valores en blanco. Se reemplazará por Indeterminado.

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

El campo Issued Date posee valores en blanco. Se reemplazará por Indeterminado.

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

**Imagen de la pantalla de un celular con letras

Descripción generada automáticamente con confianza media**

El campo Completed Date posee valores en blanco. Se reemplazará por Indeterminado.

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

**Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media**

El campo First Construction Document Date posee valores en blanco. Se reemplazará por Indeterminado.

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

**Tabla

Descripción generada automáticamente**

El campo Number of Existing Stories posee valores null. Se reemplazará por 0. Lo mismo se realizara con el campo Number of Proposed Stories, Estimated Cost y Revised Cost.

**Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente**

**Imagen que contiene Gráfico

Descripción generada automáticamente**

**Interfaz de usuario gráfica

Descripción generada automáticamente**

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

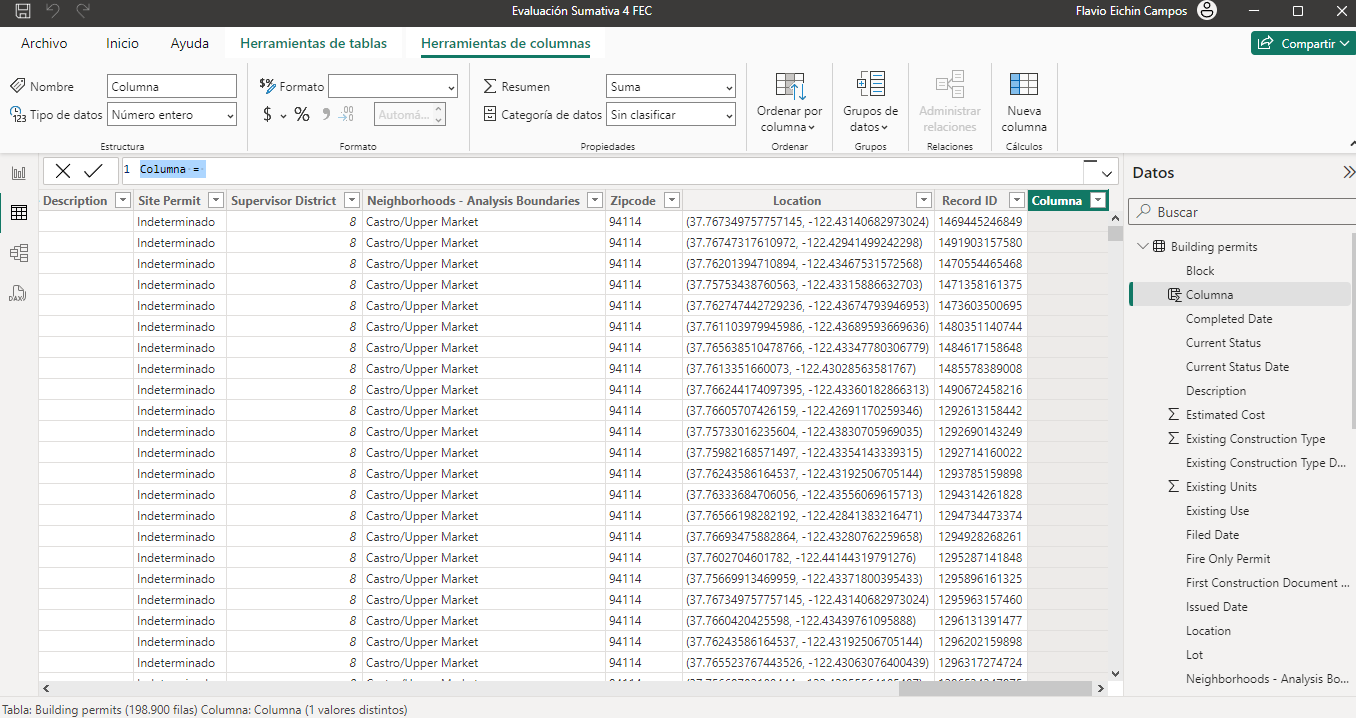
**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

En resumen, la gran cantidad de campos de formato texto, en el cual habían valores en blanco, se cambiaron por indeterminado y los campos de formato número, en el cual habían valores null, se cambiaron por 0.

1. **Crea 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.

****

Generé una nueva columna a partir de la diferencia entre el Estimated Cost y el Revised Cost. Ya que existían valores 0, el hecho de calcular una diferencia porcentual, arrojaría valores indeterminados.

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

Esta columna creada se nombró como Diferencia de costos.

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

1. **Guardar y compartir:**

* Guardar el archivo en Power BI con todas las transformaciones realizadas.
* Sube el archivo a una plataforma de almacenamiento GitHub y comparte el enlace en la plataforma asignada.
* Entregue un Word documentando paso a paso todas las modificaciones realizadas en Power BI.

<https://github.com/flavioeichin/sumativa-4>