****

**Universidad Politécnica de Tapachula**

“Innovación y Tecnología al servicio de la sociedad”

**Nombre del alumno – Matricula**

León Martínez Ángel Manuel– 203342

**Catedrático**

Karina Cancino Villatoro.

**Materia**

Minería de datos.

**Cuatrimestre / periodo escolar**

9ºA. (Abril – Agosto / 2023)

**Nombre de la práctica / proyecto**

ED1. Análisis de base datos

**Tipo de evidencia**

Investigación.

**Plan de estudios**

Ingeniería de Software

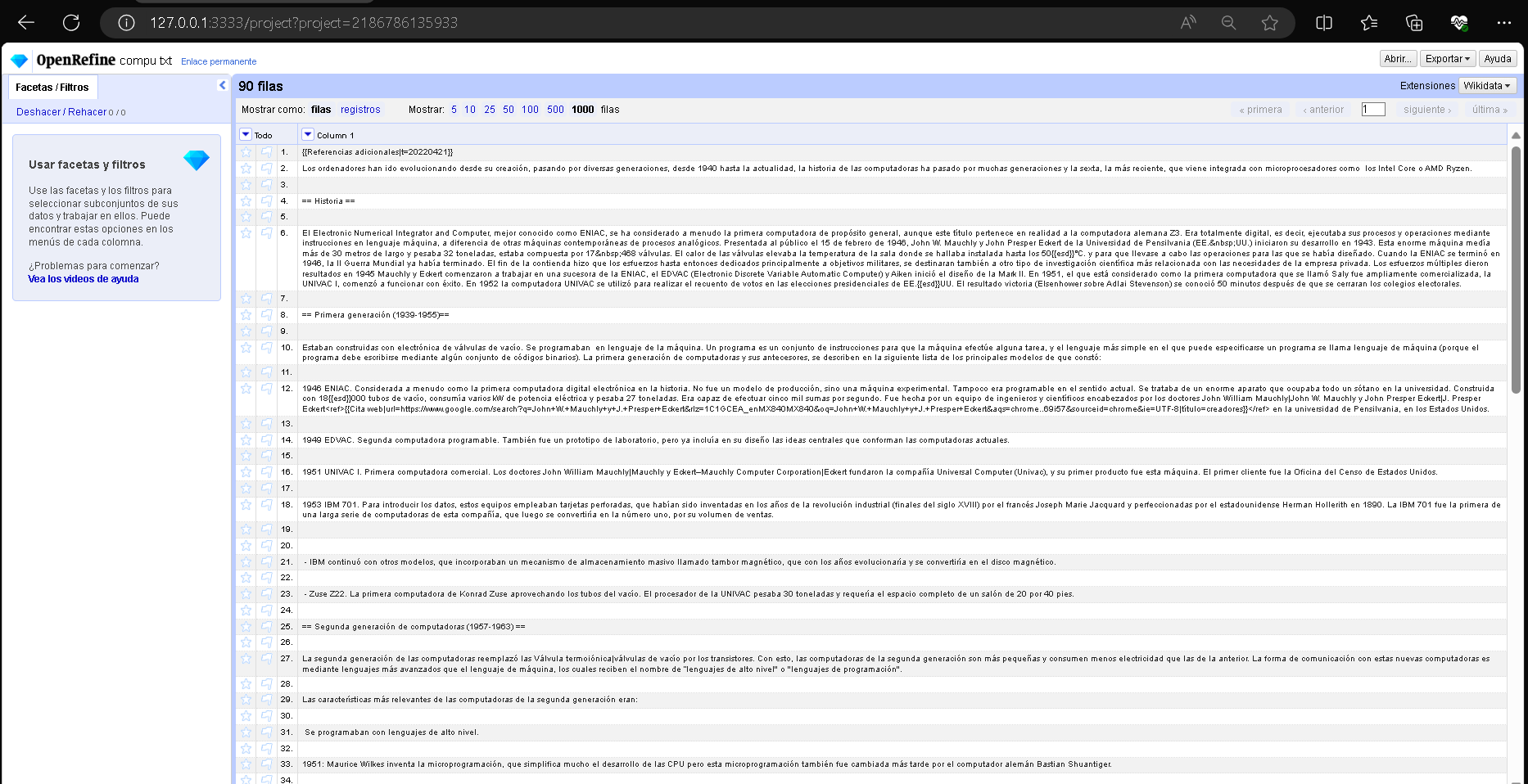
Tapachula, Chiapas; A 05 de Julio del 2023.

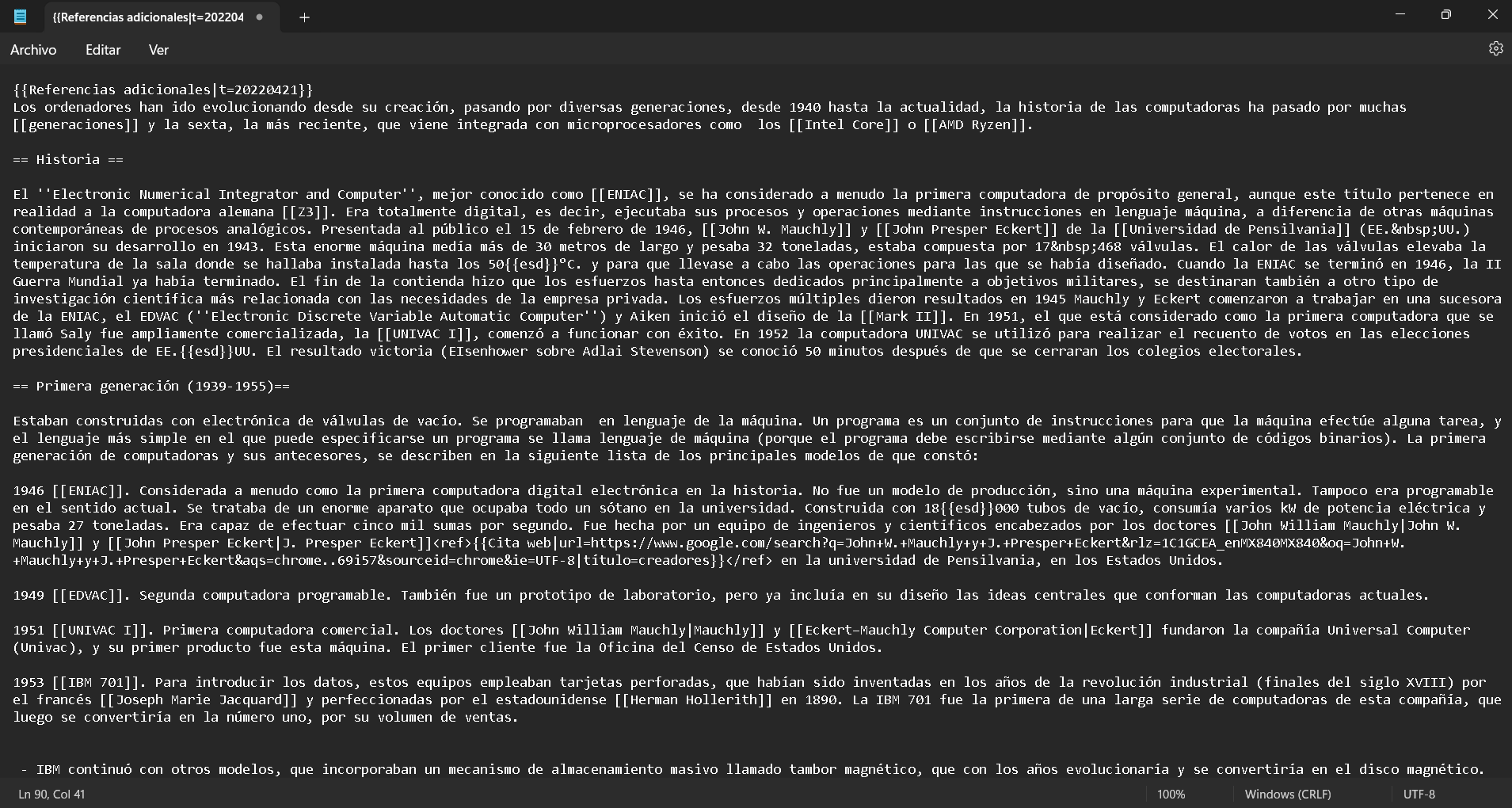
En este caso he decidido hablar de las generaciones de las computadoras como se aprecia



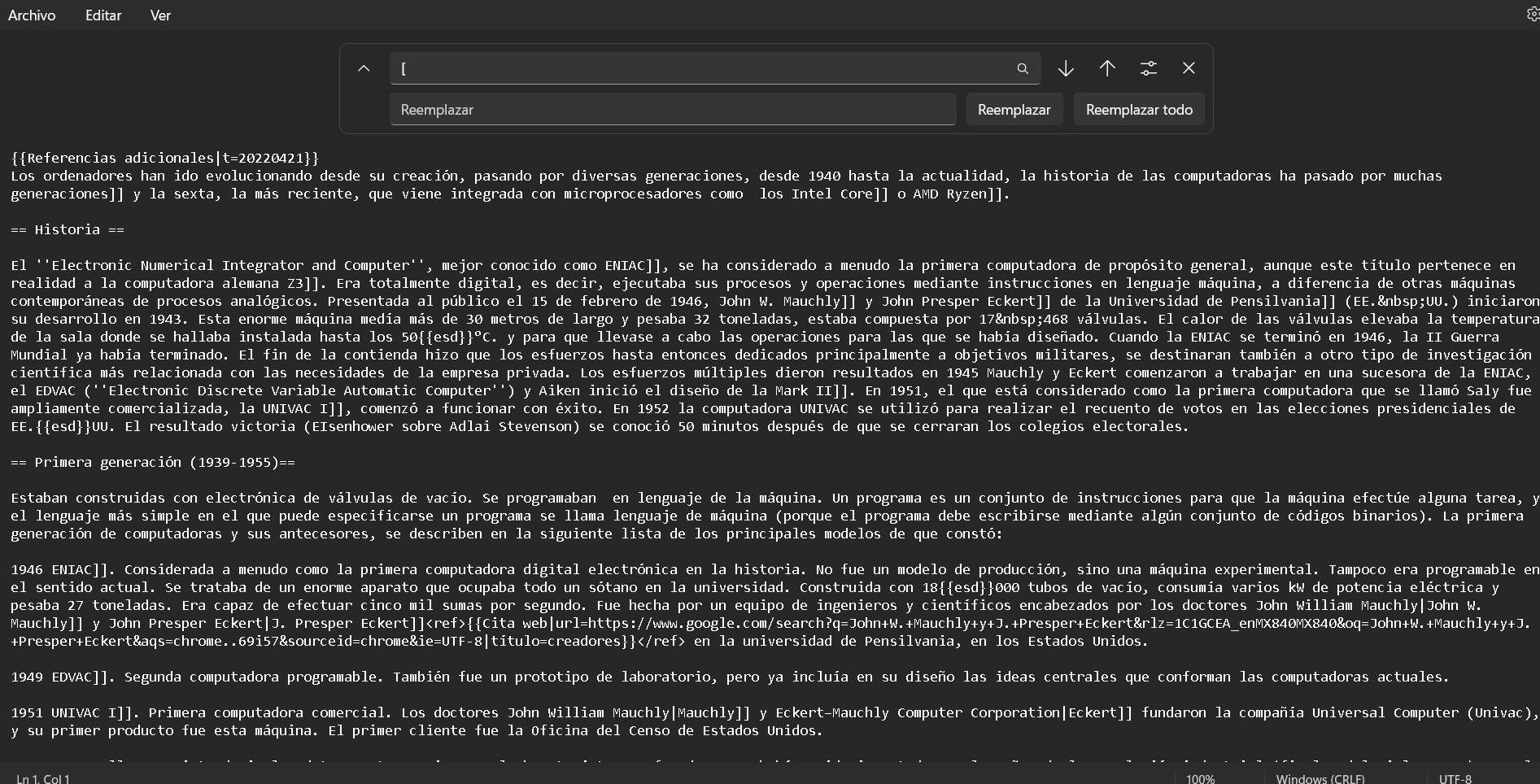
<https://es.wikipedia.org/w/index.php?title=Generaciones_de_computadoras&action=edit>

Enlace de formato a hablar en mi evidencia.

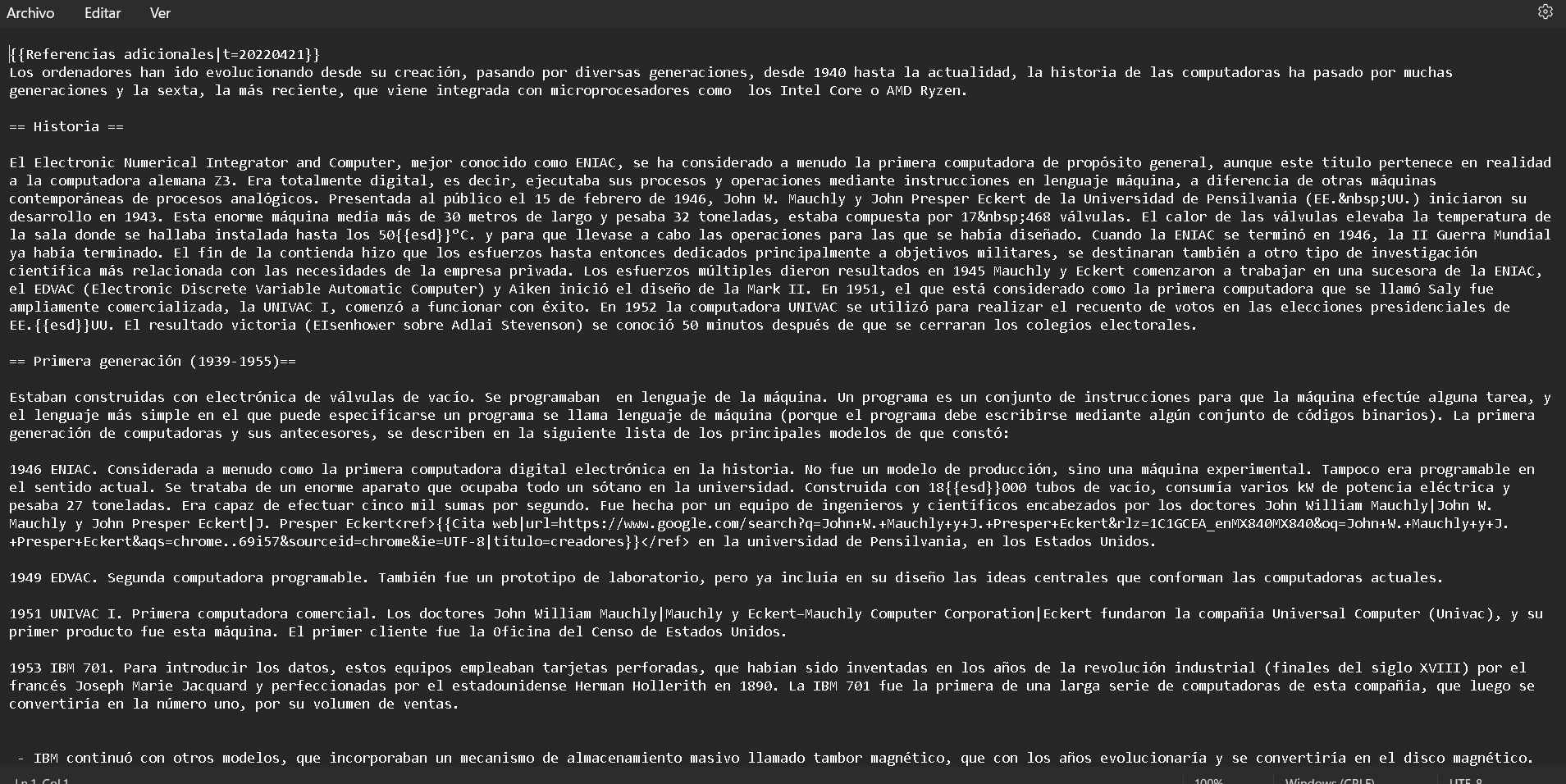




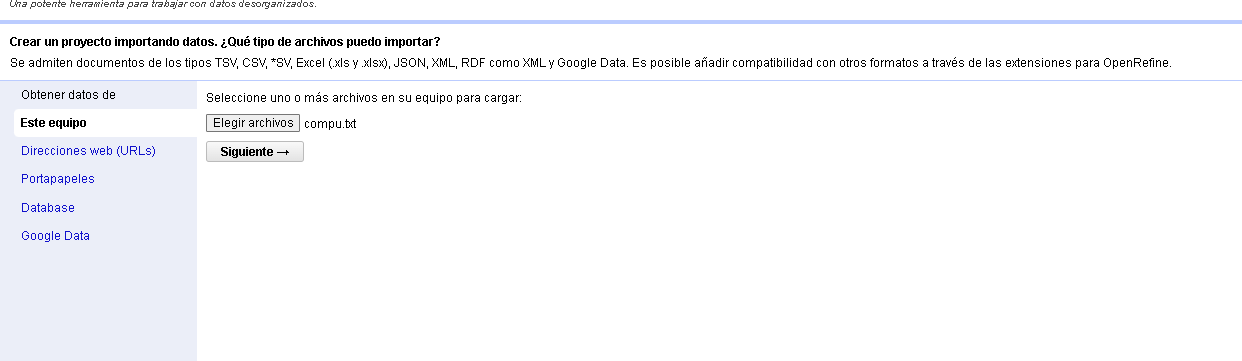
En este caso quitaremos estos signos que eliminare: [,],',”



Terminado el proceso me queda de esta manera:



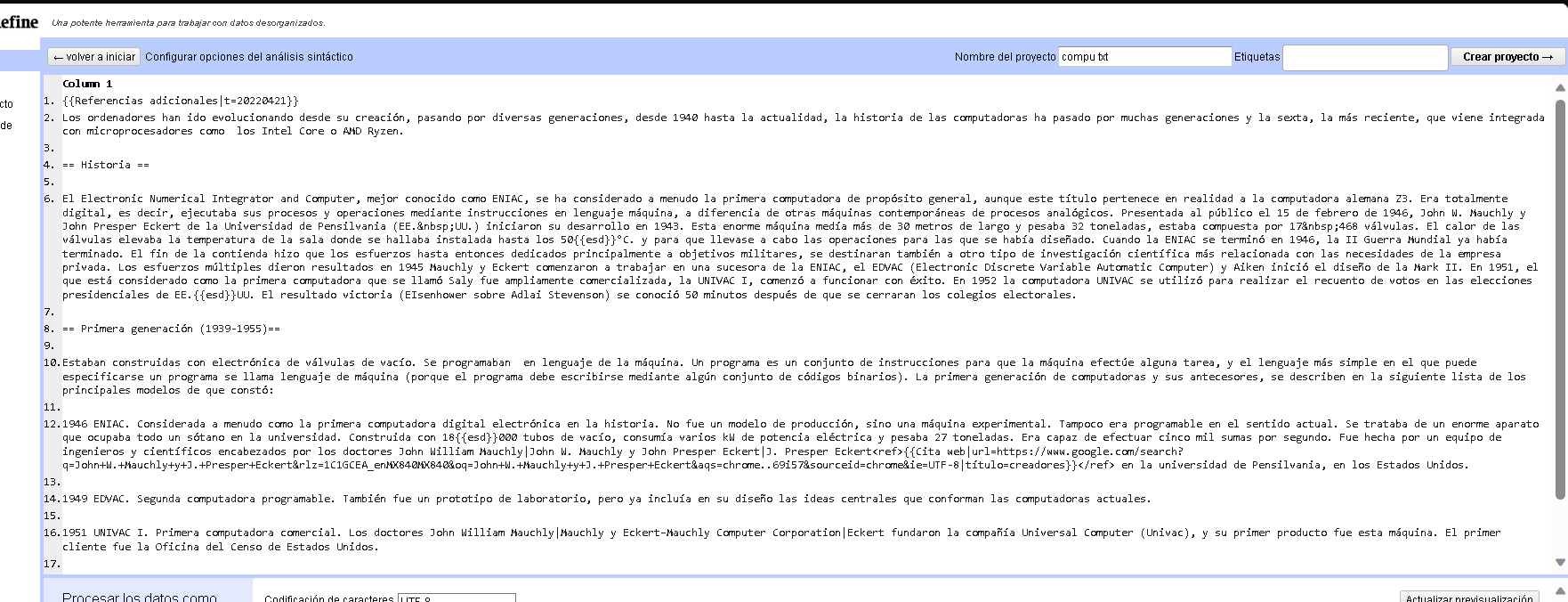
Elijo el archivo de texto que anteriormente remplace



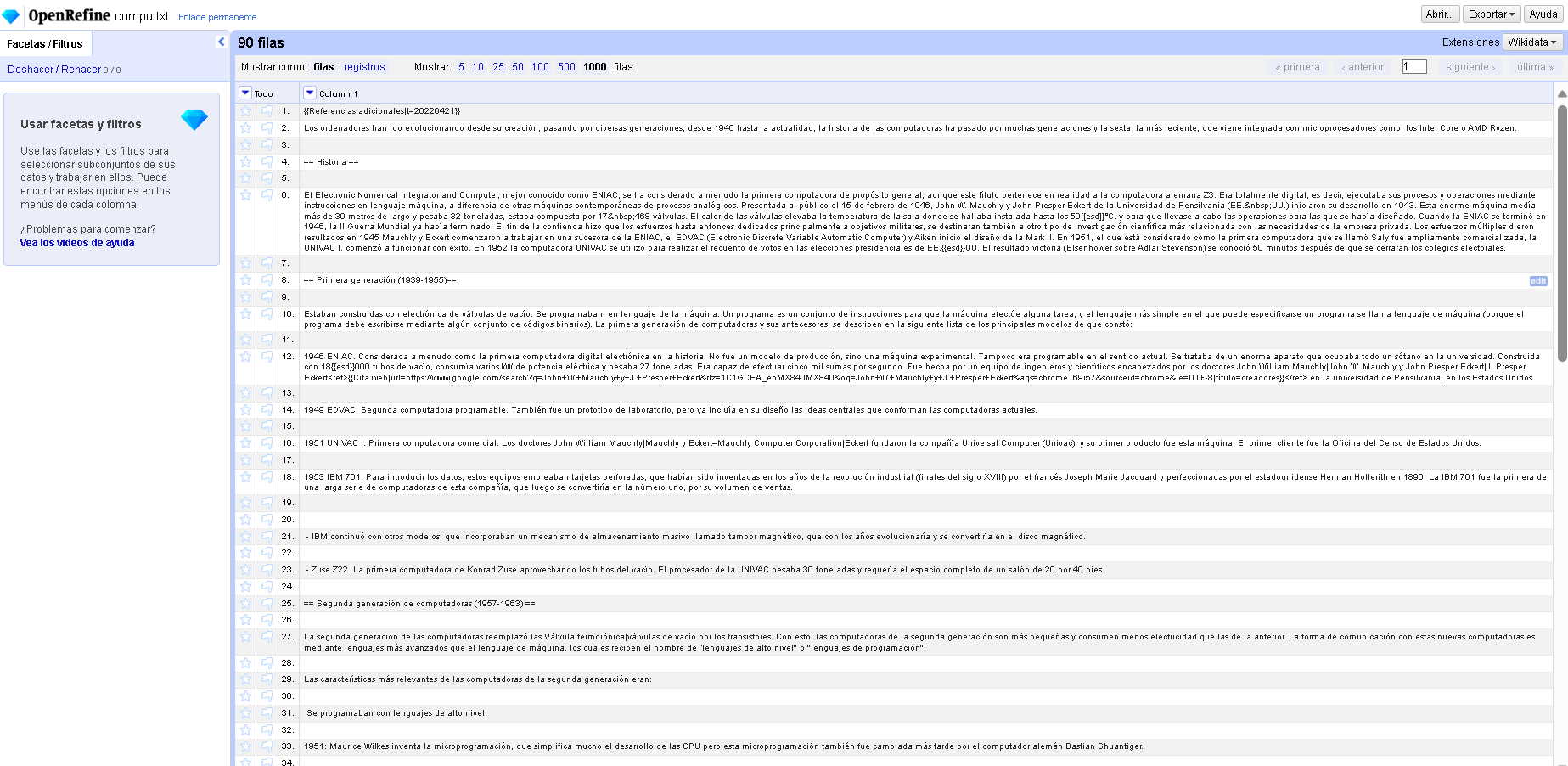
Seleccione siguiente para cargar mi txt



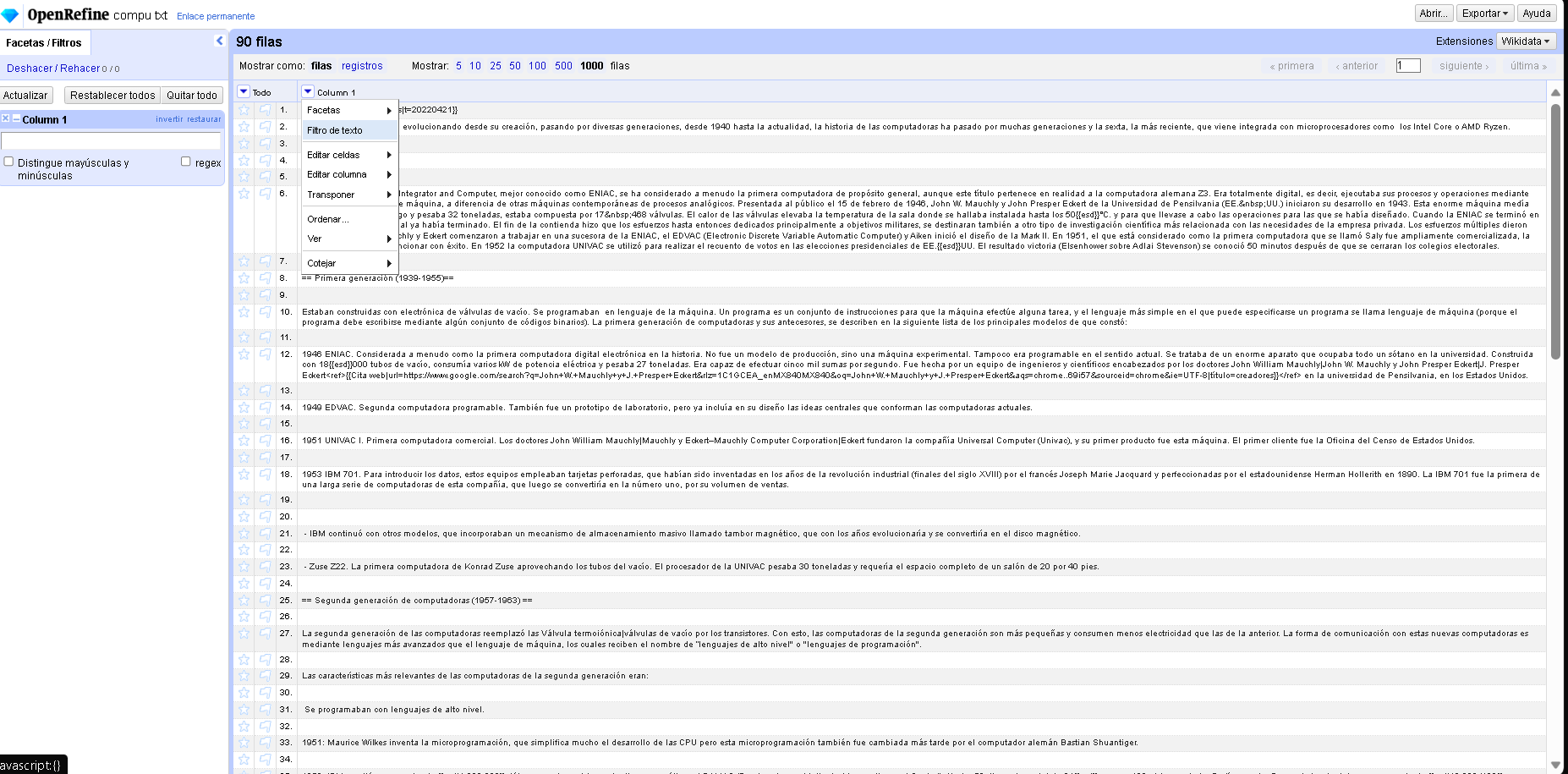
Seleccionare Archivos de texto con campos de anchura fija y me cargara mi texto que copie sin orden alguno tal y como estaba.

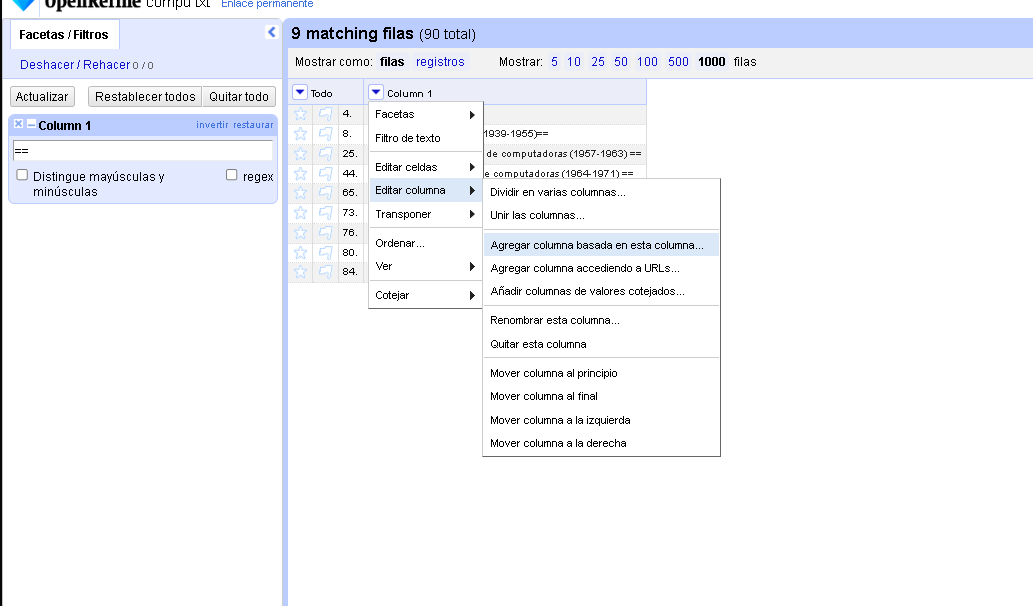


Seleccionamos crear proyecto y me tendría que cargar esto:

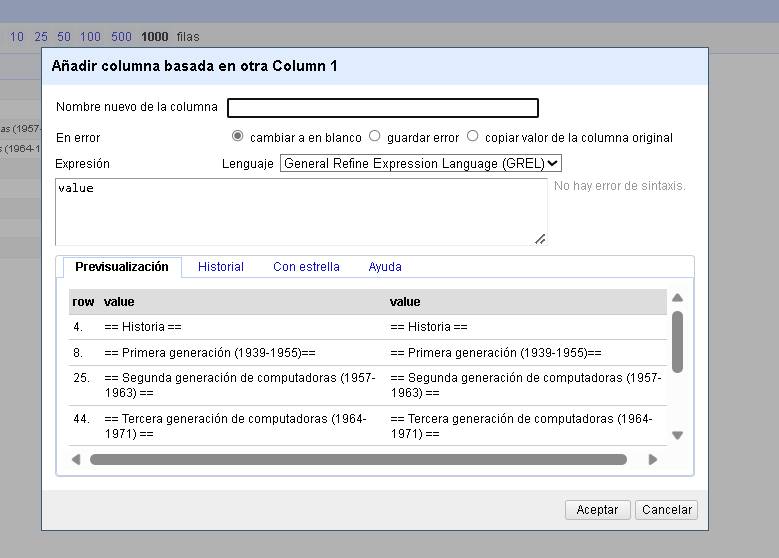


Seleccionare la columna 1 y seleccionamos filtro de texto en el lado izquierdo hare una búsqueda con **“==”** pues así se identifican todos los títulos

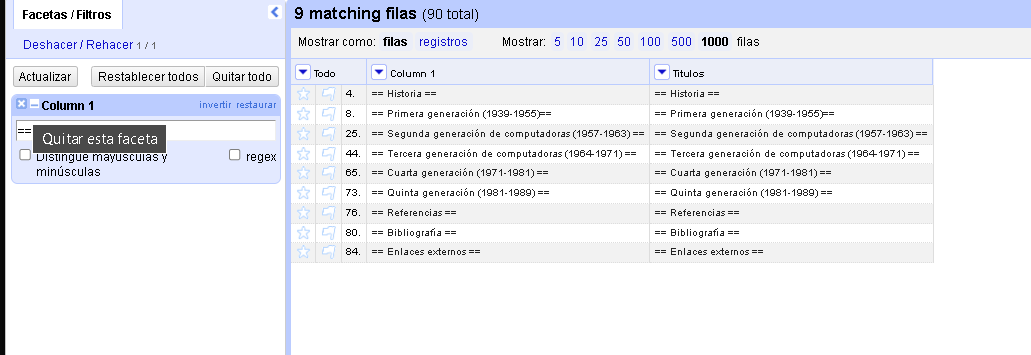




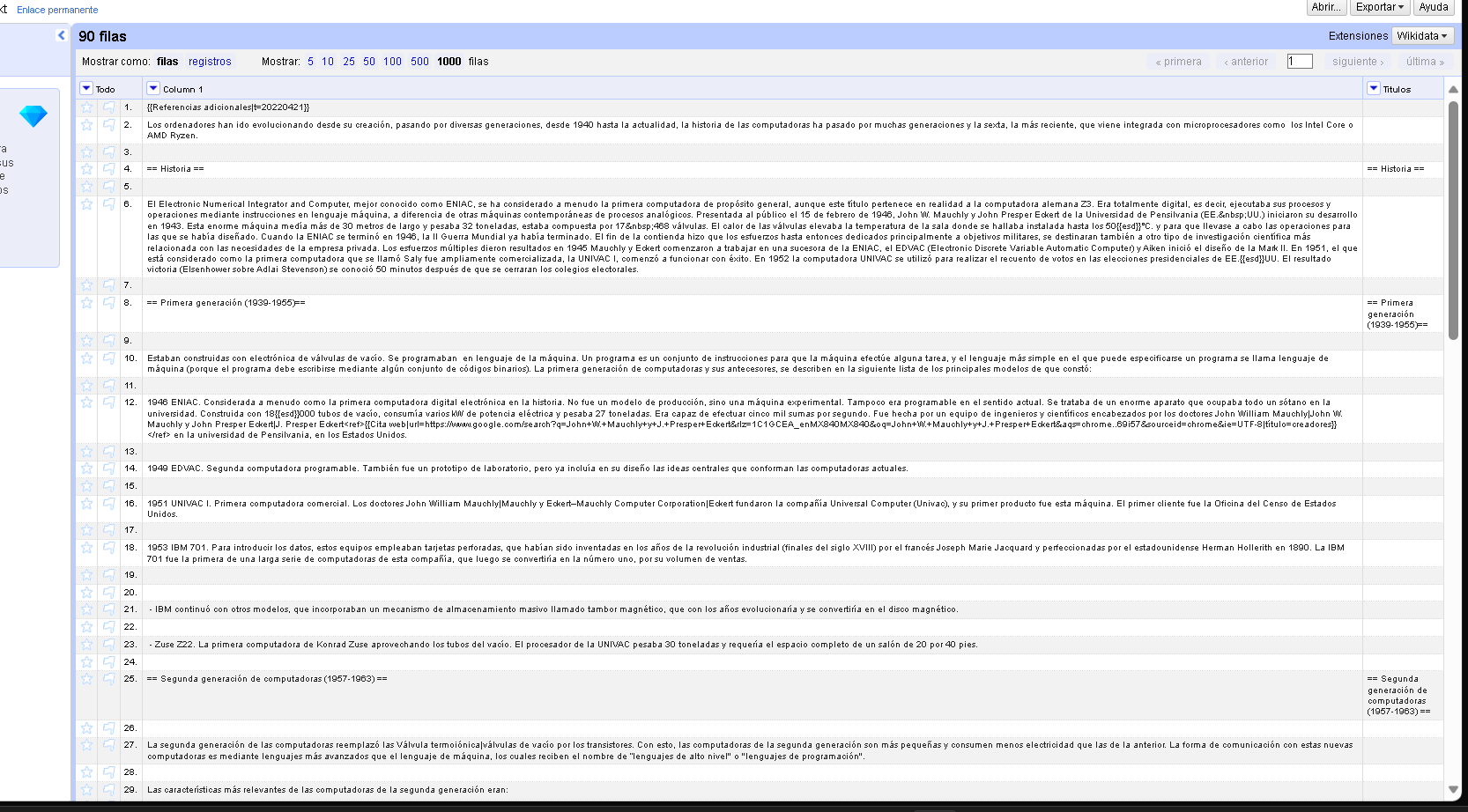
Nos aparecerán las columnas que es tiene el patrón de “**==**” por lo que seleccionare la columna de solamente los == agregare una columna basada en esta columna



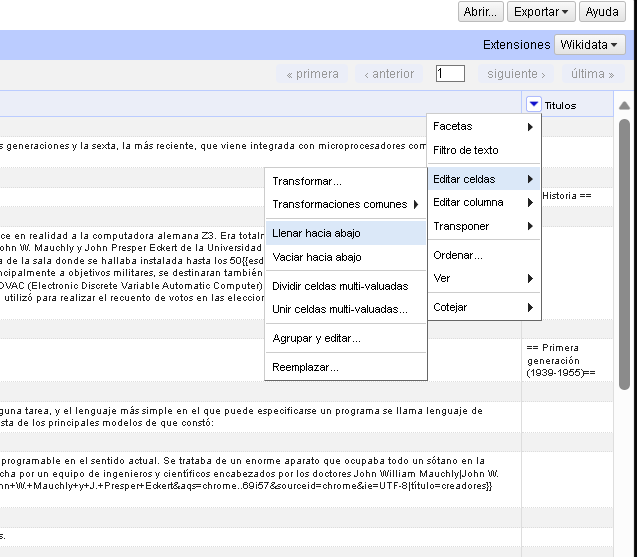
El nombre que le proporcionare será de **Títulos**



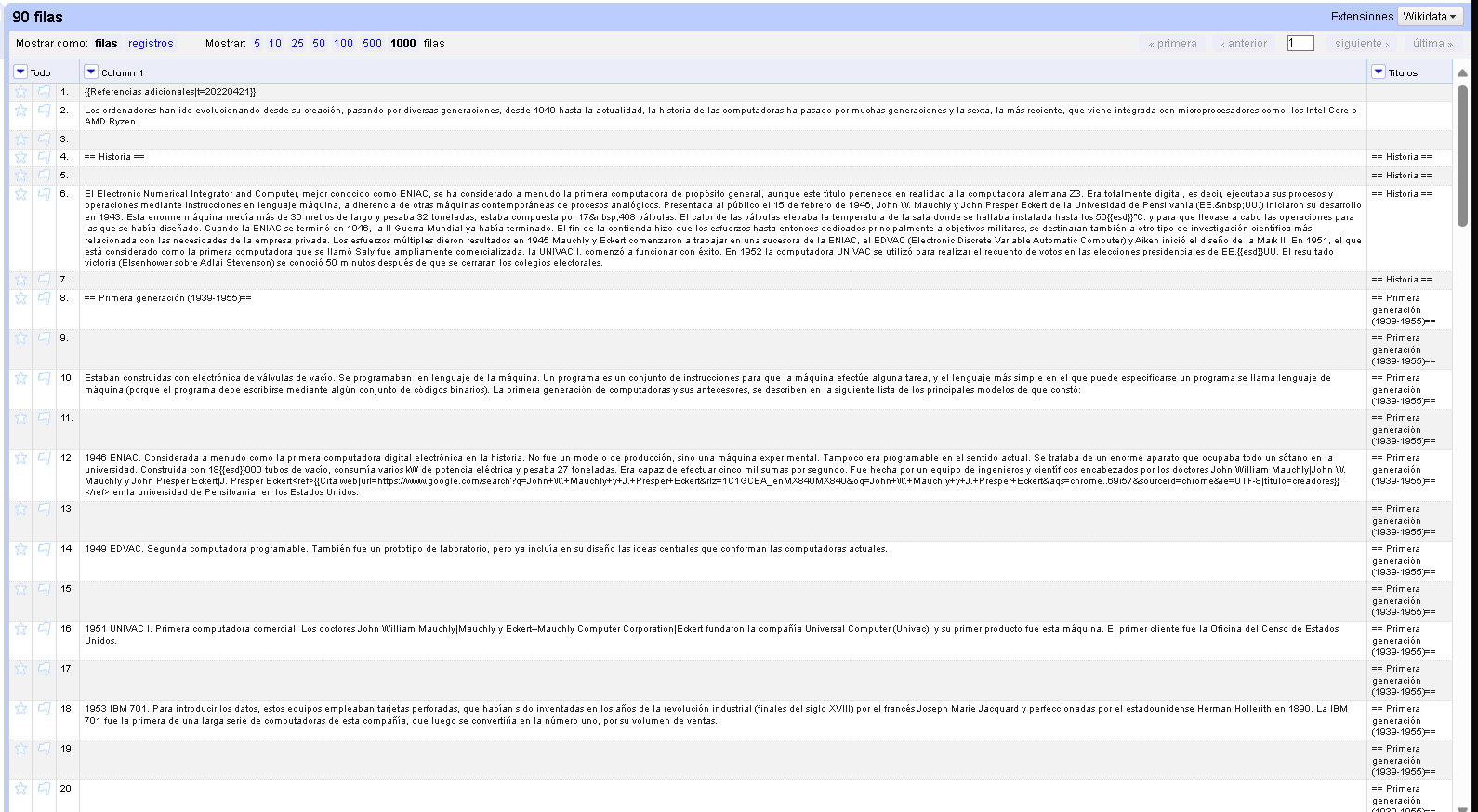
Quito el filtro para regresar a la normalidad



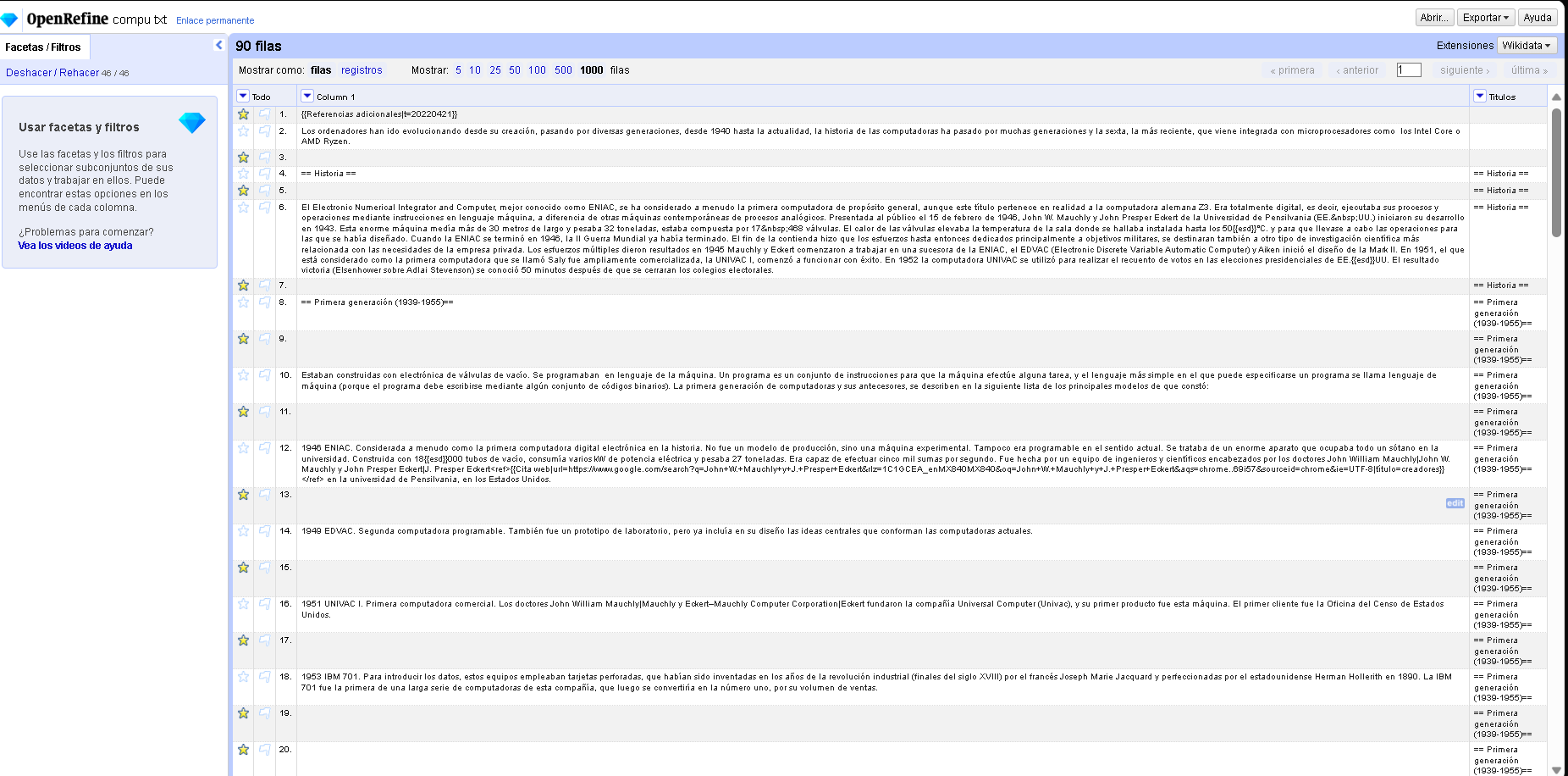
Ahora tengo una nueva columna llamada **Títulos** y esta cambiara cada que el titulo cambie acorde a la información



Para evitar los campos vacíos hare la siguiente opción la cual llenara los campos vacíos acorde a el cambio de Títulos que corresponda



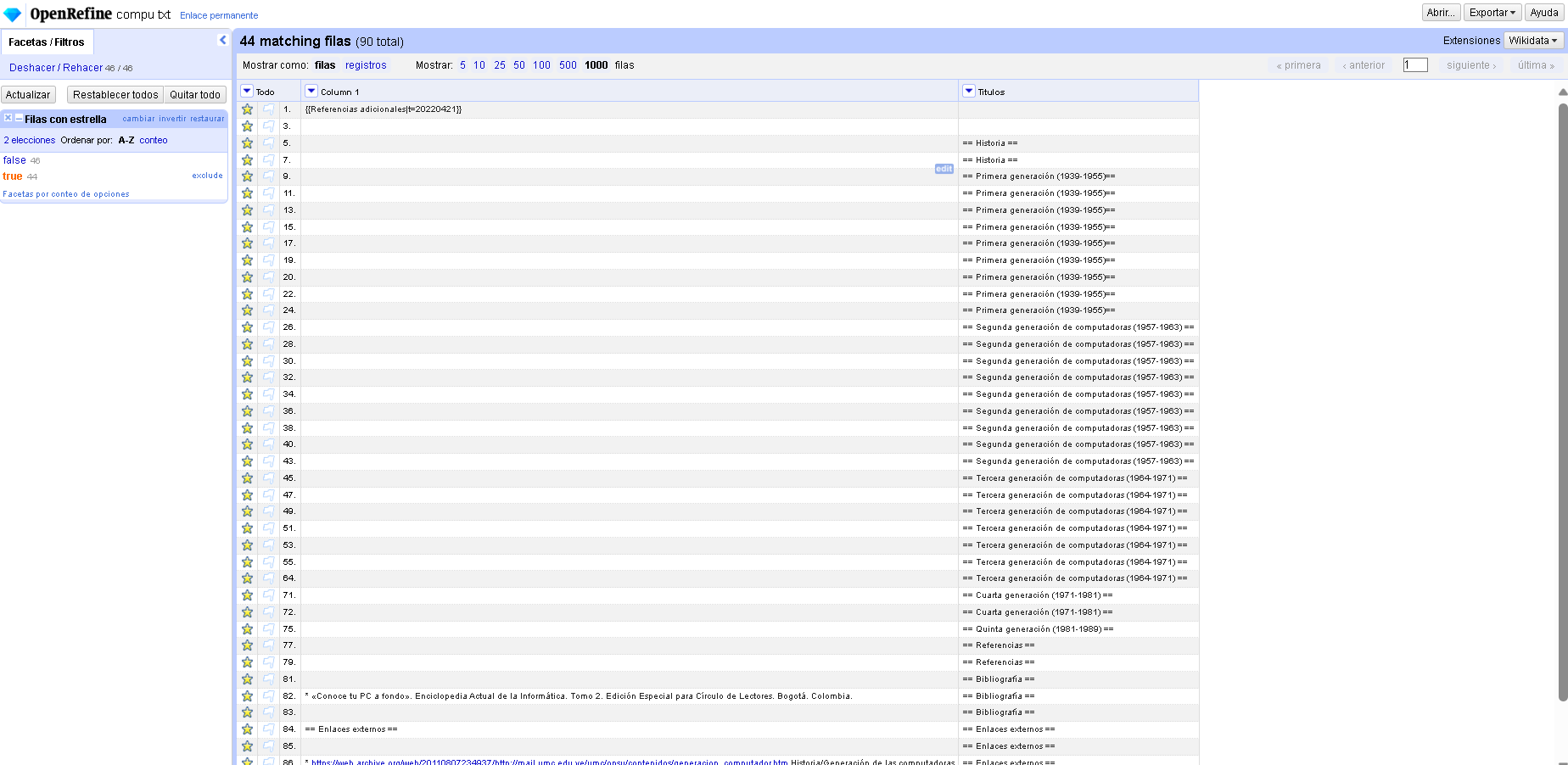
Me queda así

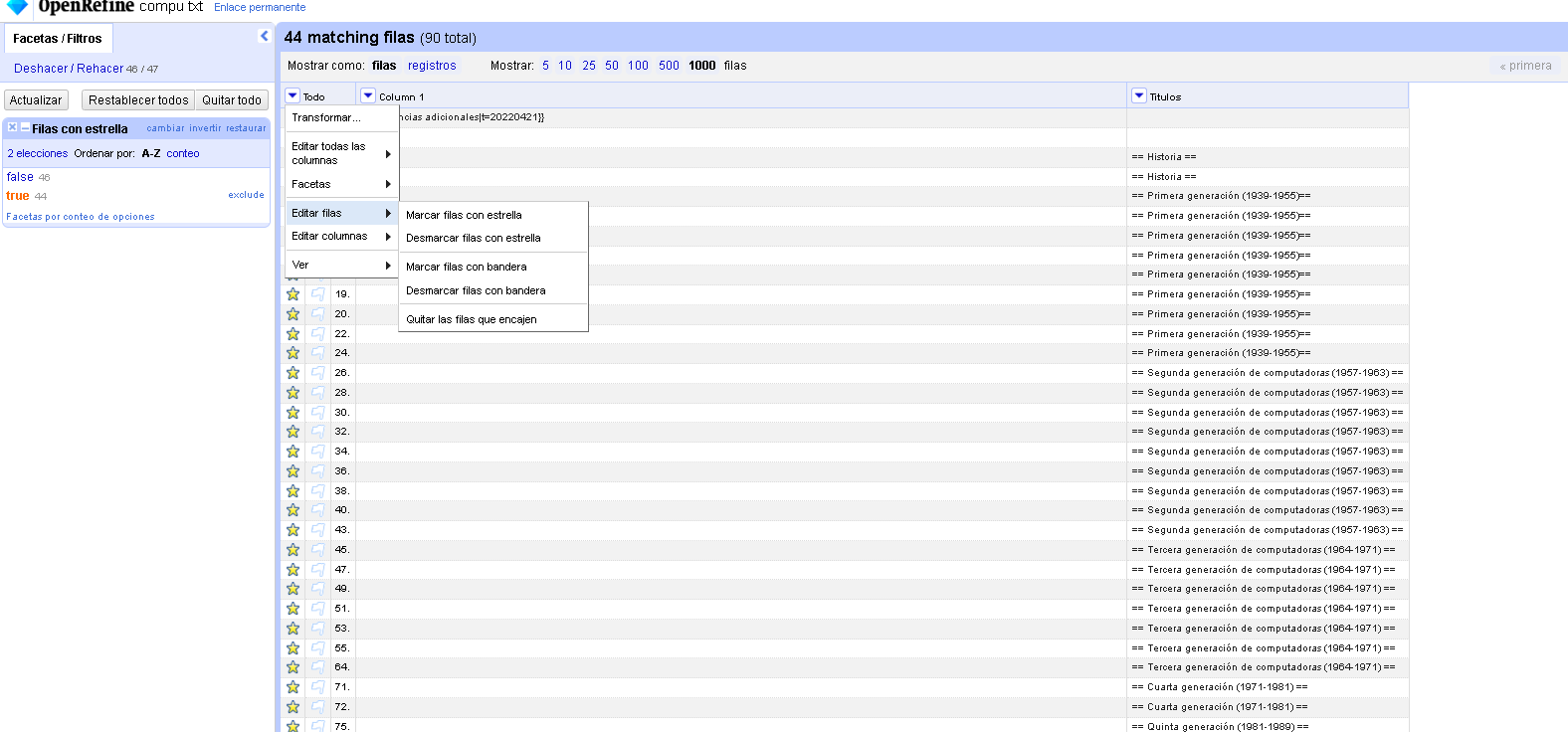


Seleccione las estrellas amarillas las cuales harán la función de eliminar las filas sin importación o información de utilidad, así como espacios blancos sin información.

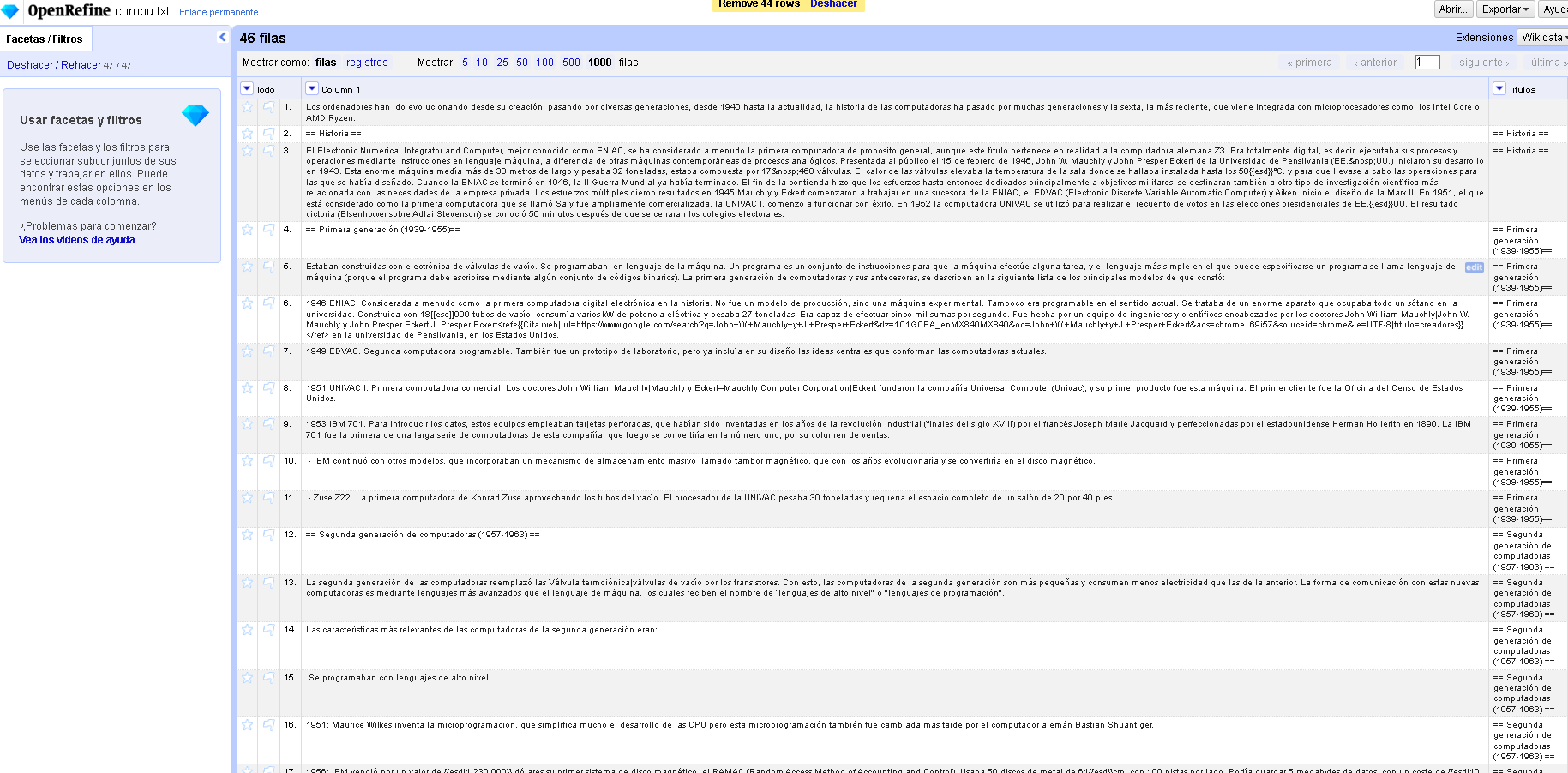


Separo lo seleccionado de la información que considero de importancia

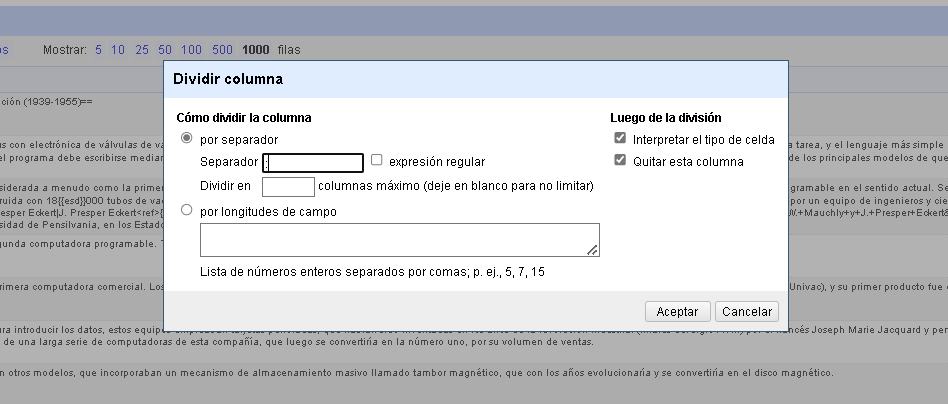




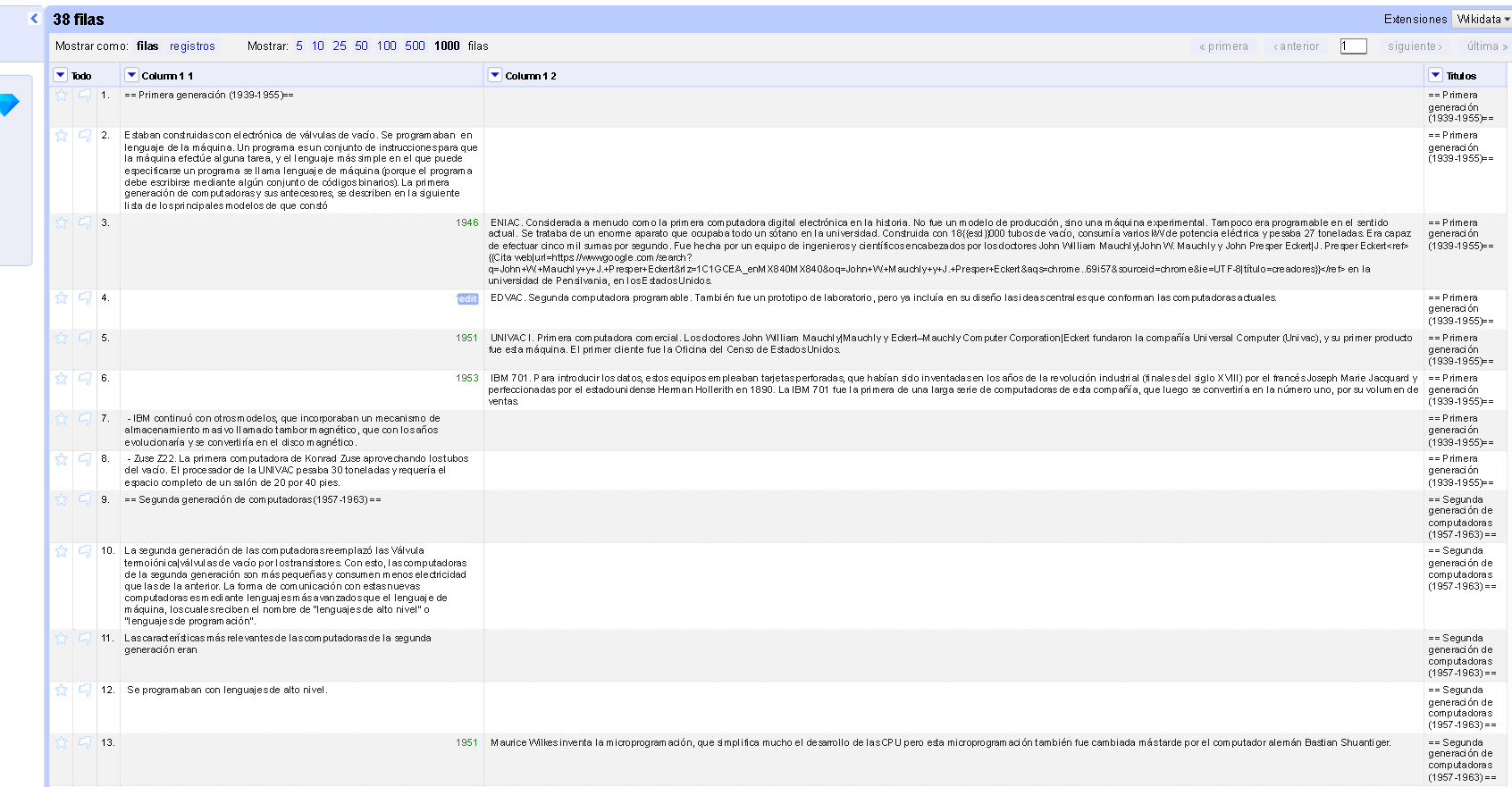
Hago lo siguiente para quitar las filas sin importacia para mi



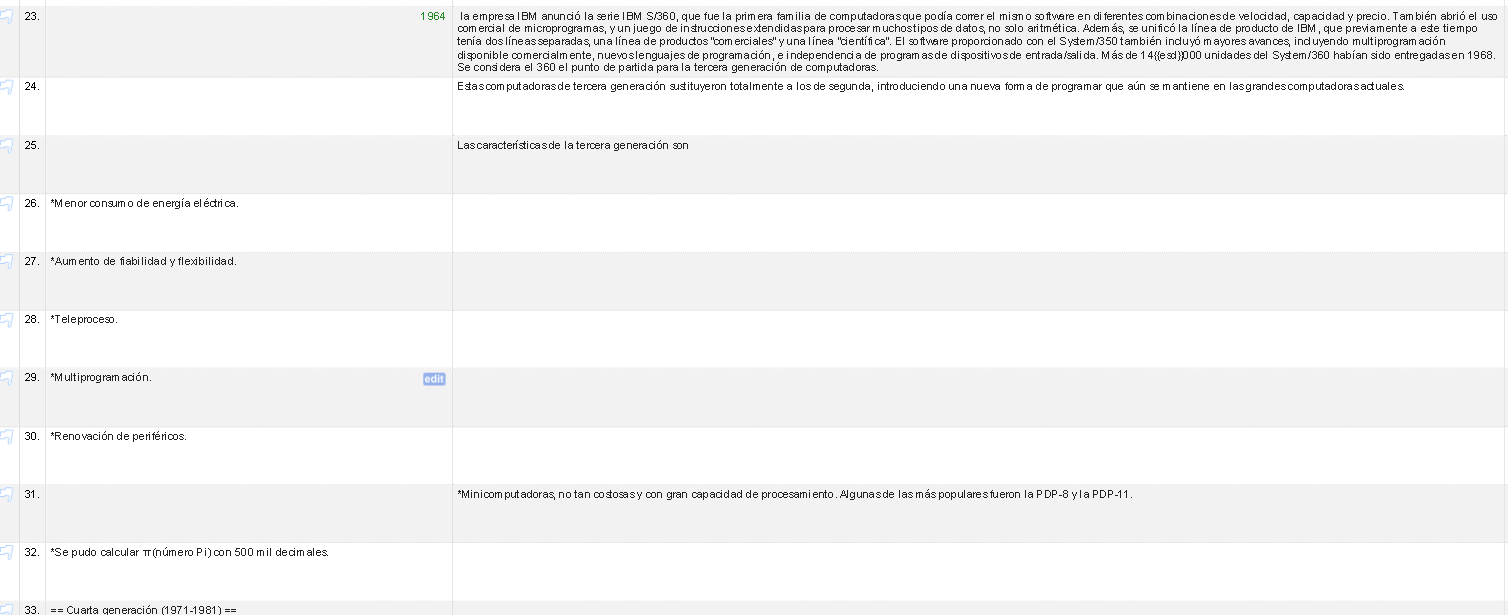
Disminuye la cantidad de filas que tengo



En este caso habra menos filas puesto que elimine unas que aun me salian en blanco, dividido las columnas por : pues esta dividen las fechas



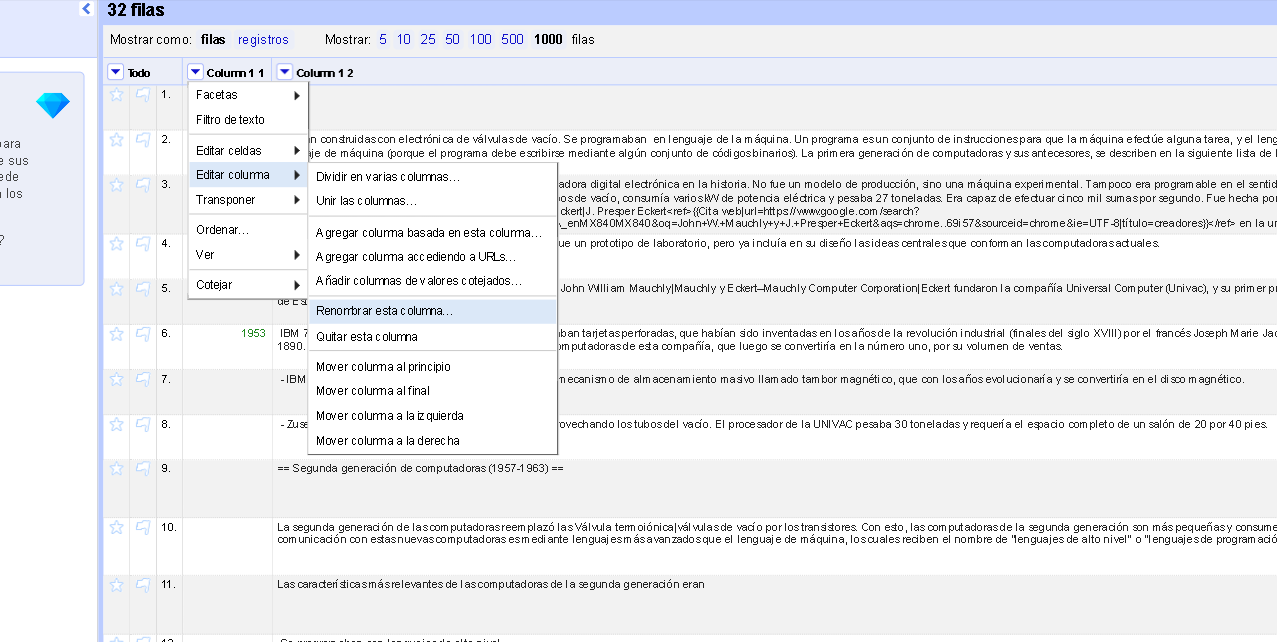
Queda asi por lo que manualmente tendre que hacer unos ajustes



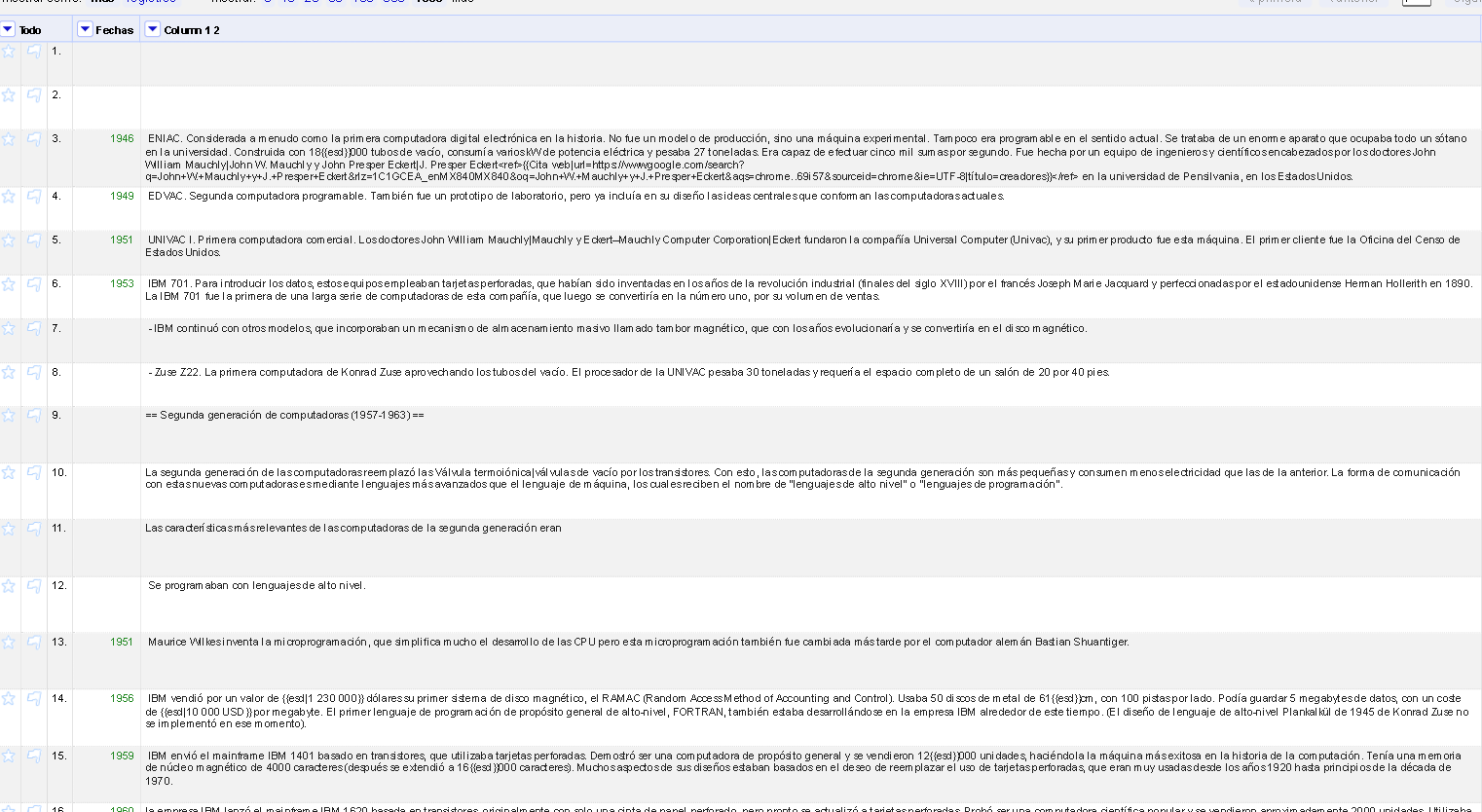
Como se puede apreciar dice las caracteristicas de la 3 generacion son y estan separadas por lo que se tendra que juntar todo estos en donde va y se eliminaran los espacios en blanco dejando asi menos filas



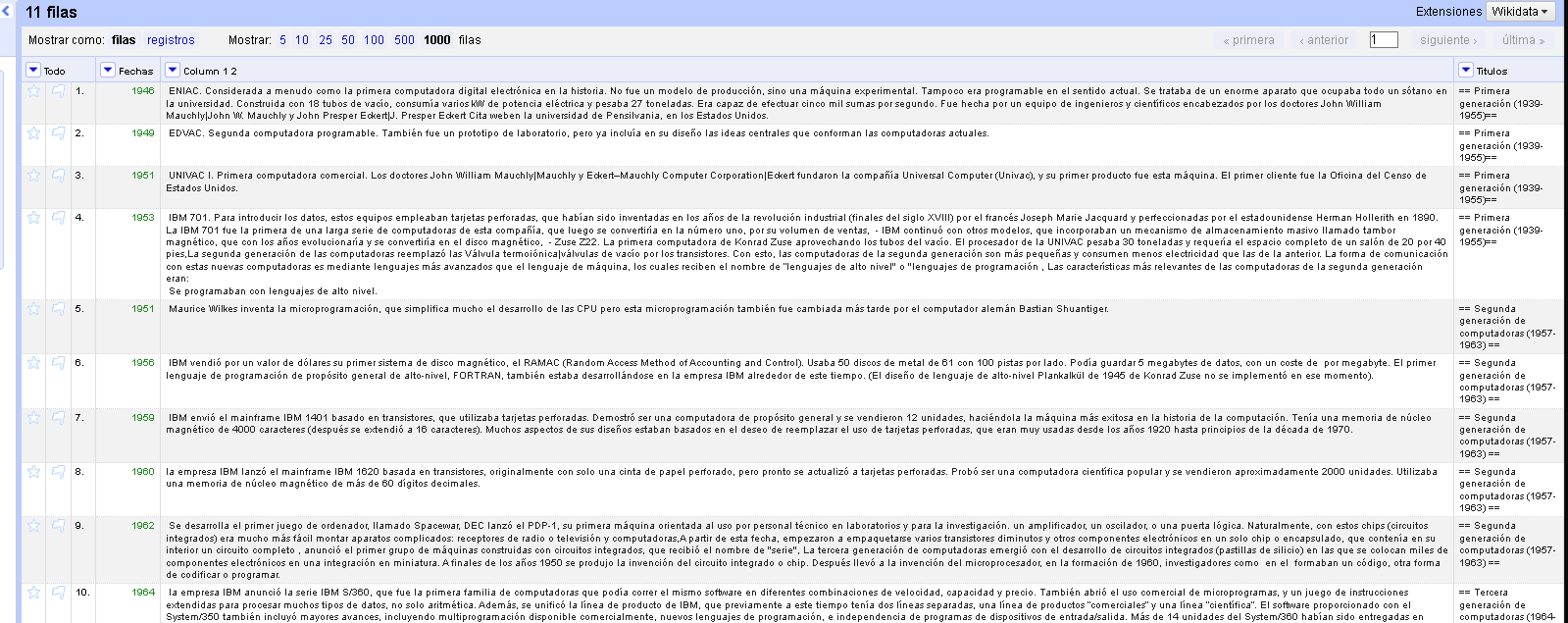
Quedaron filas sin informacion se eliminan

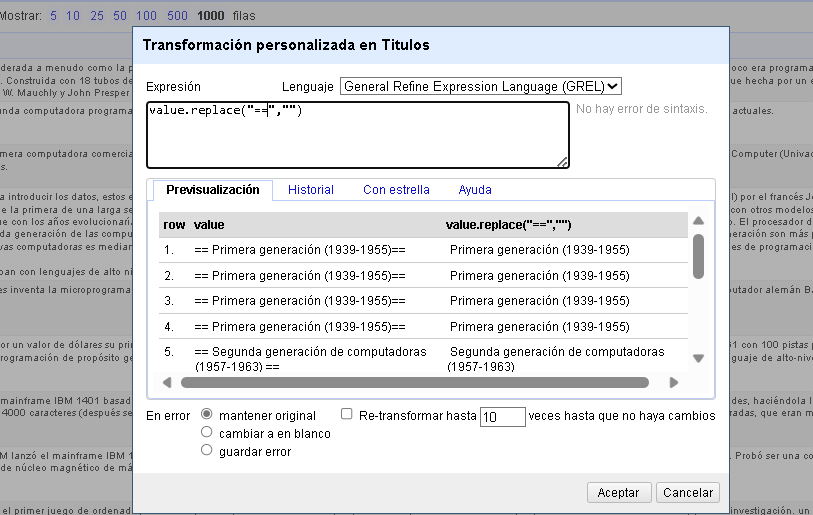


Nombramos la columna como fechas

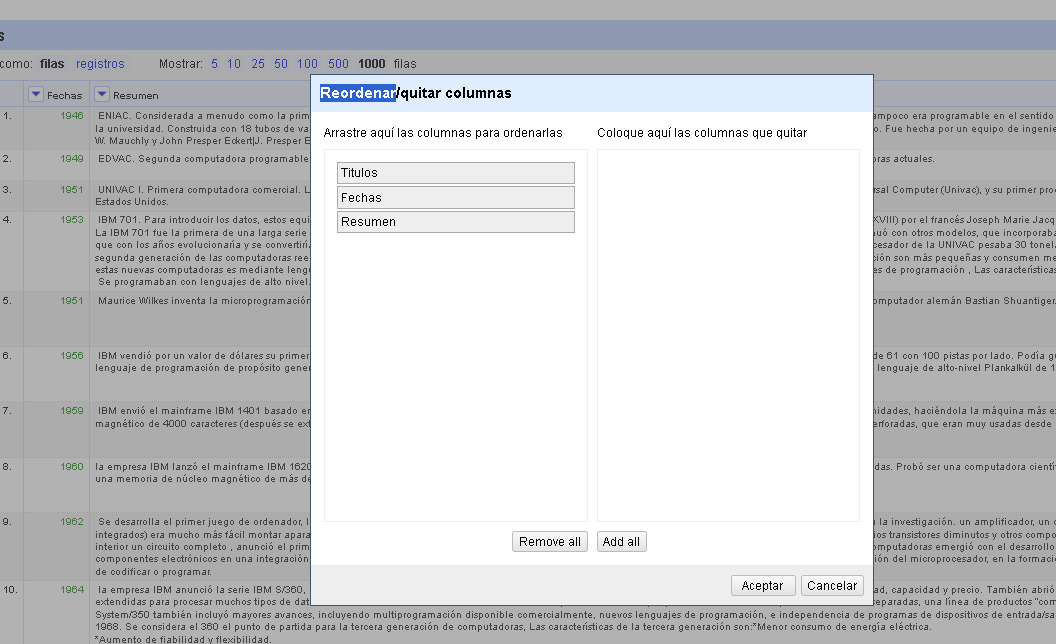


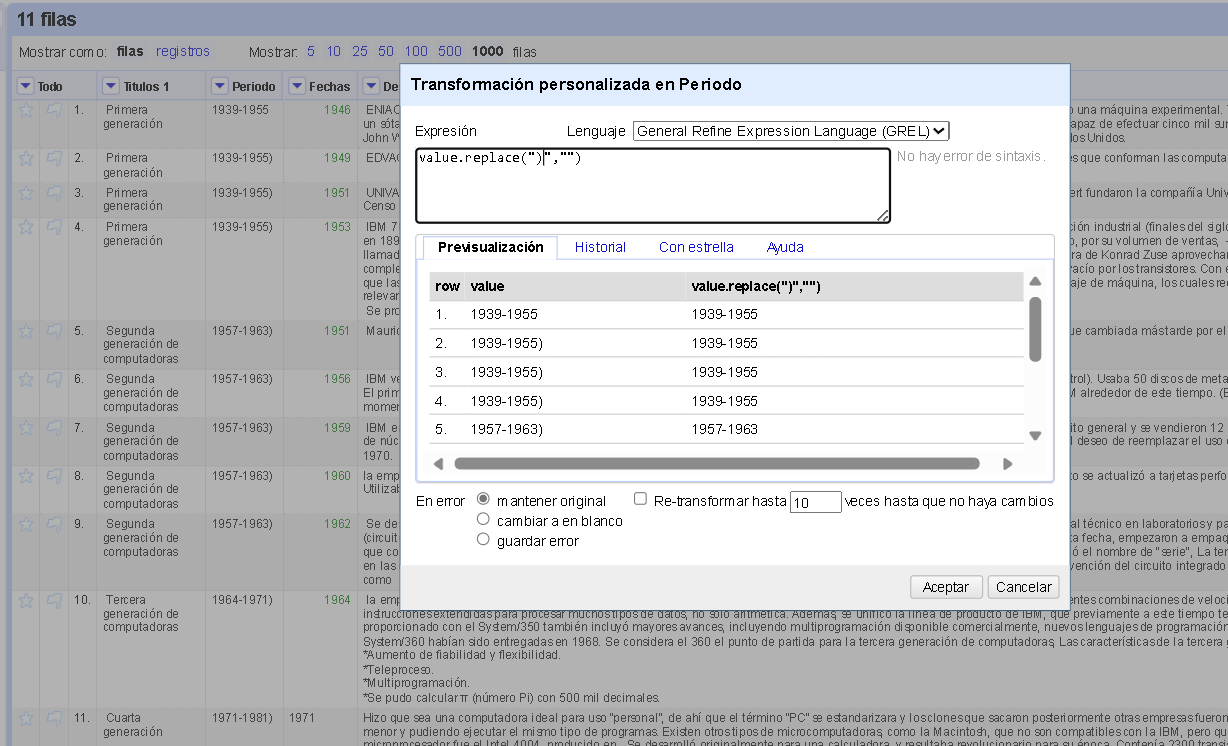
Eliminare informacion no util en la descripcion como enlaces sin informacion y junatare todo acorde a las fechas





Quitare los simbolos del titulo





Mientras hacia el proceso he considerado eliminar el titulo pues es algo que no podría investigar o sacar información precisa en una base de datos y lo he dejado de esta manera

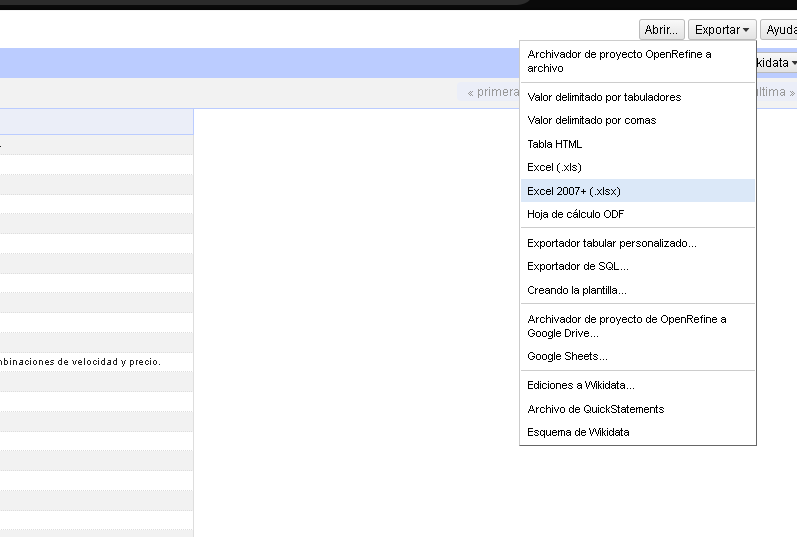


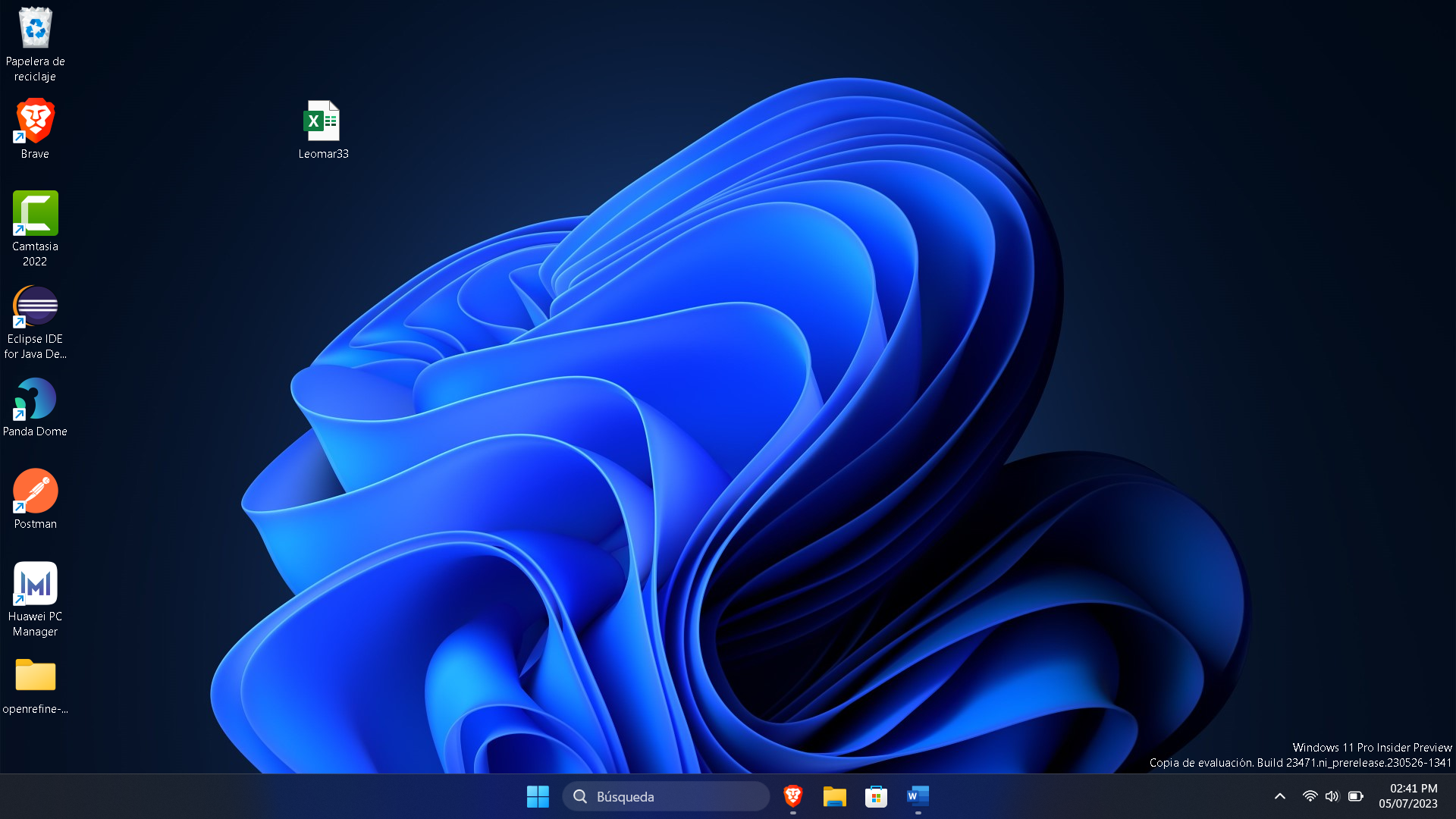
Así queda el texto de esta manera el cual ya está organizado

ERROR DESPUES DE CARGAR EL CSV DECIDI LIMPIAR DE MANERA MANUAL MI BASE DE DATOS ASI COMO BORRAR LA COLUMNA PERIODO

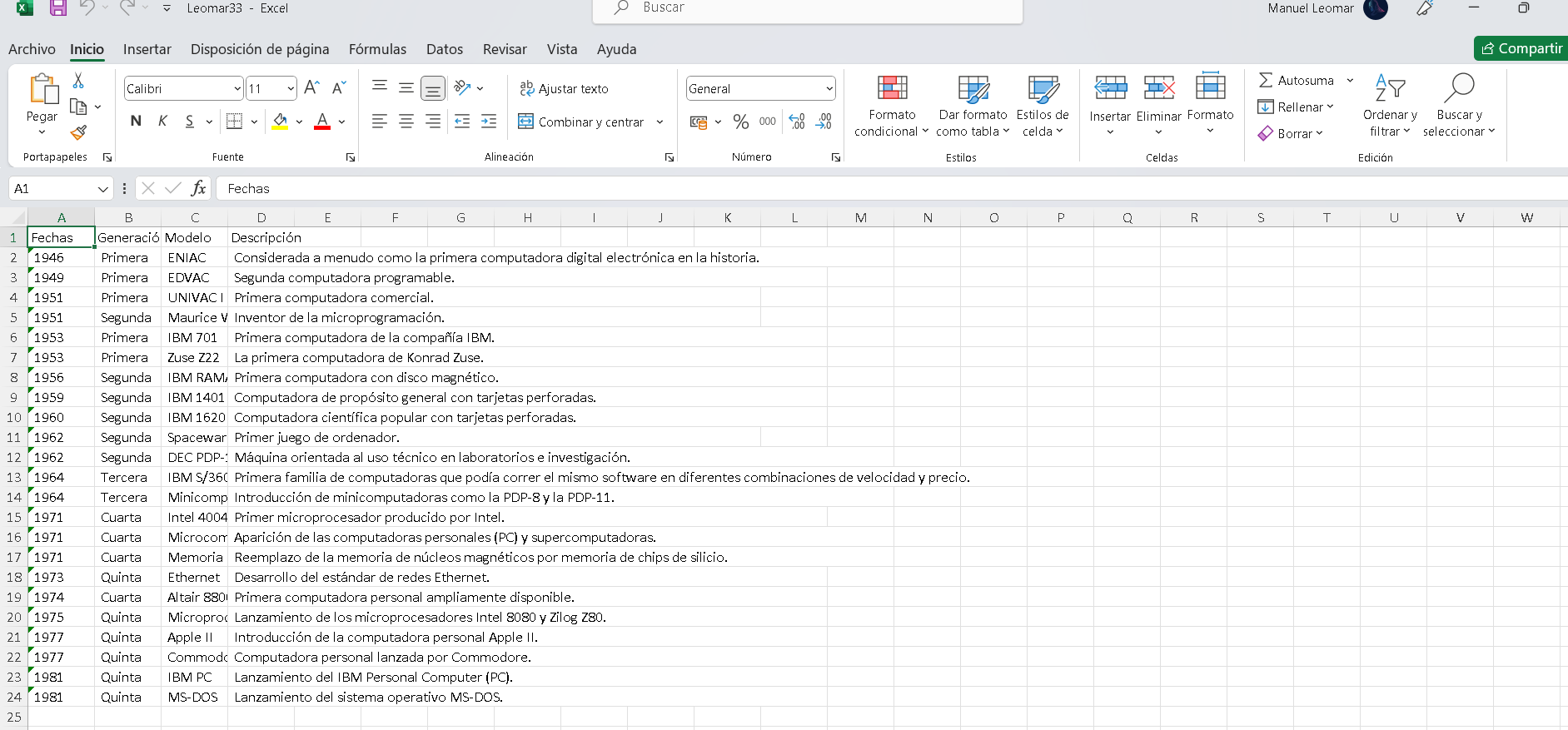


Exporte en (xlsx) Excel 2007+

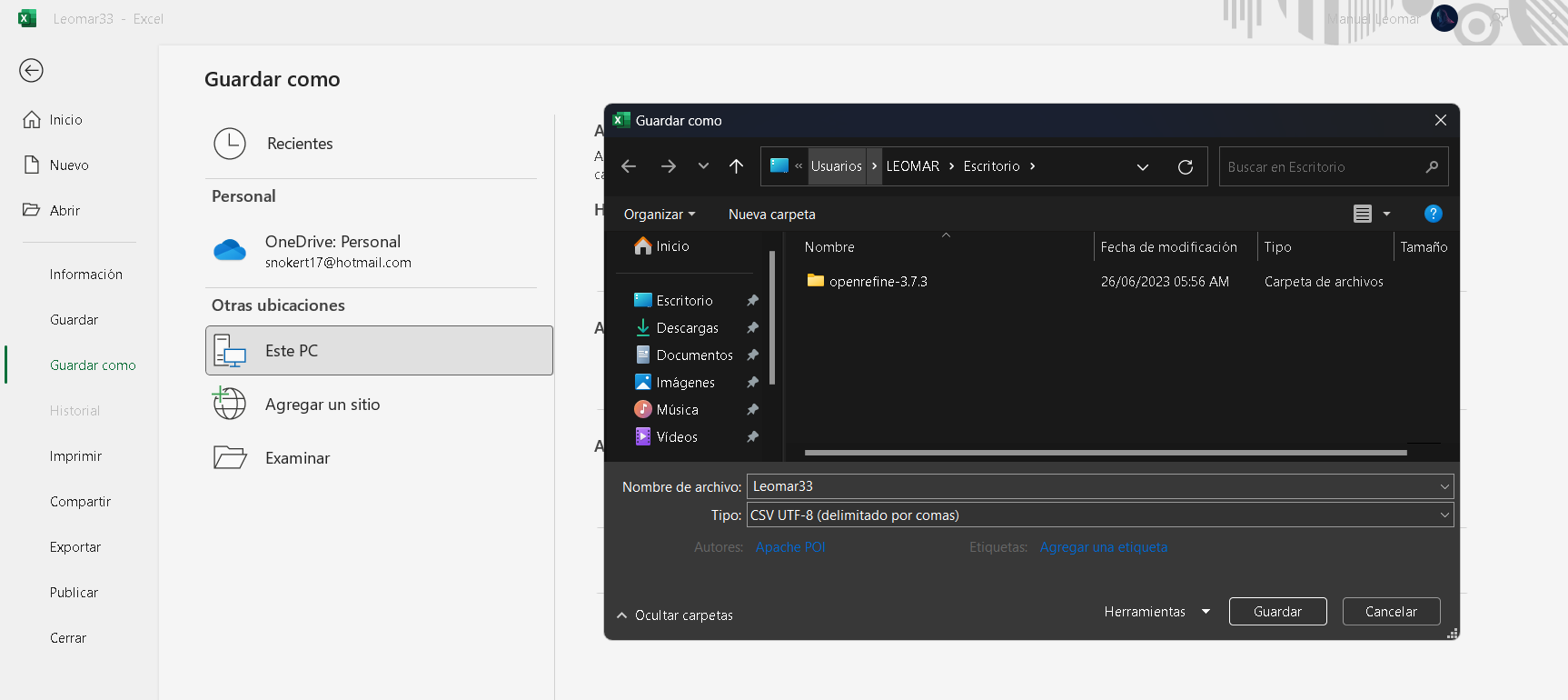




En este caso yo le cambie el nombre, pero se les descargara con el nombre que le pusieron a crear el proyecto



Aquí estarán mis tablas de **Openrifine** pero en Excel



Tengo que guardar mi Excel, pero en CSV delimitado por comas para que kaggle lo pueda leer sin problema alguno



Quedara así.

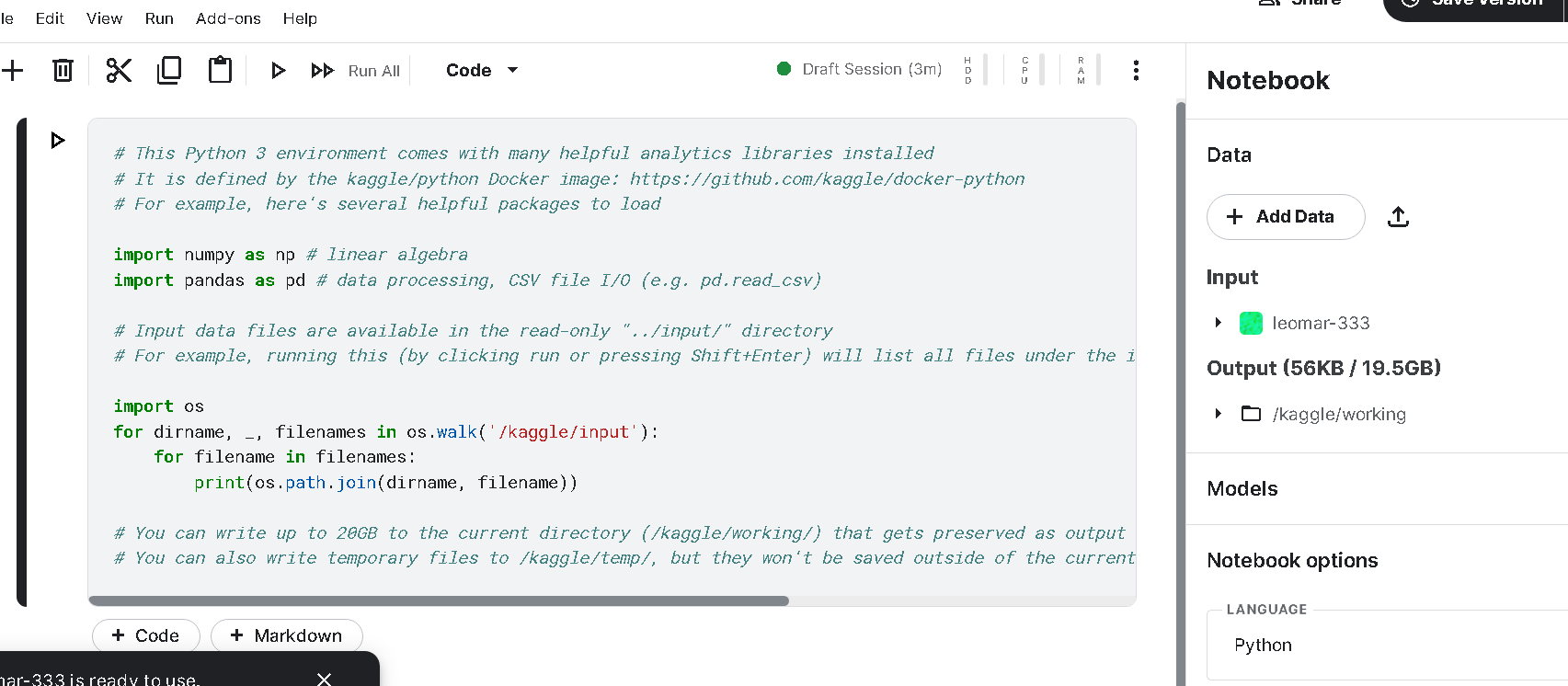


Creare un nuevo libro en el cual pondré mi CSV que cree anteriormente



Yo cambiare el nombre del libro, así como también la encenderé

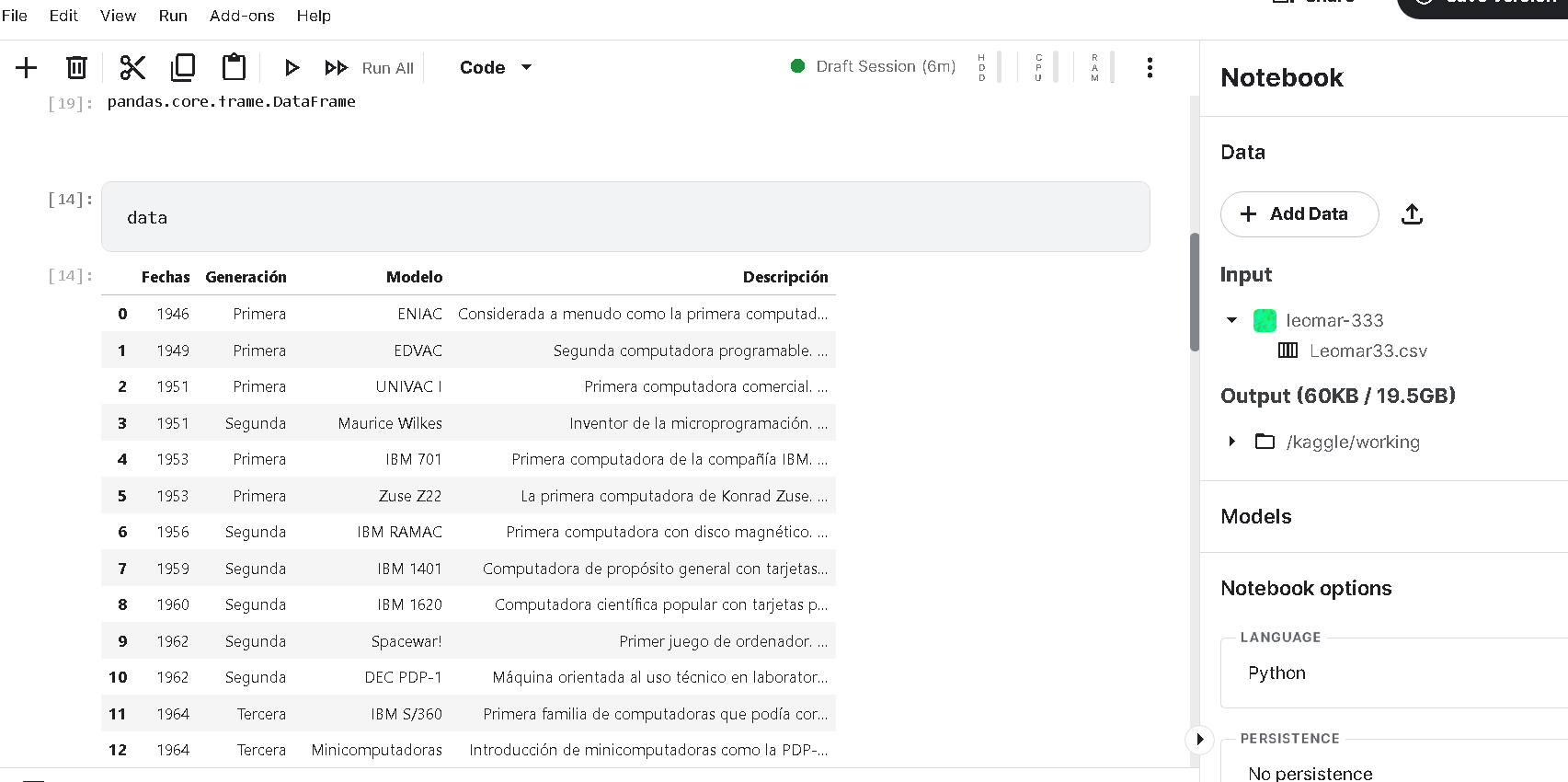


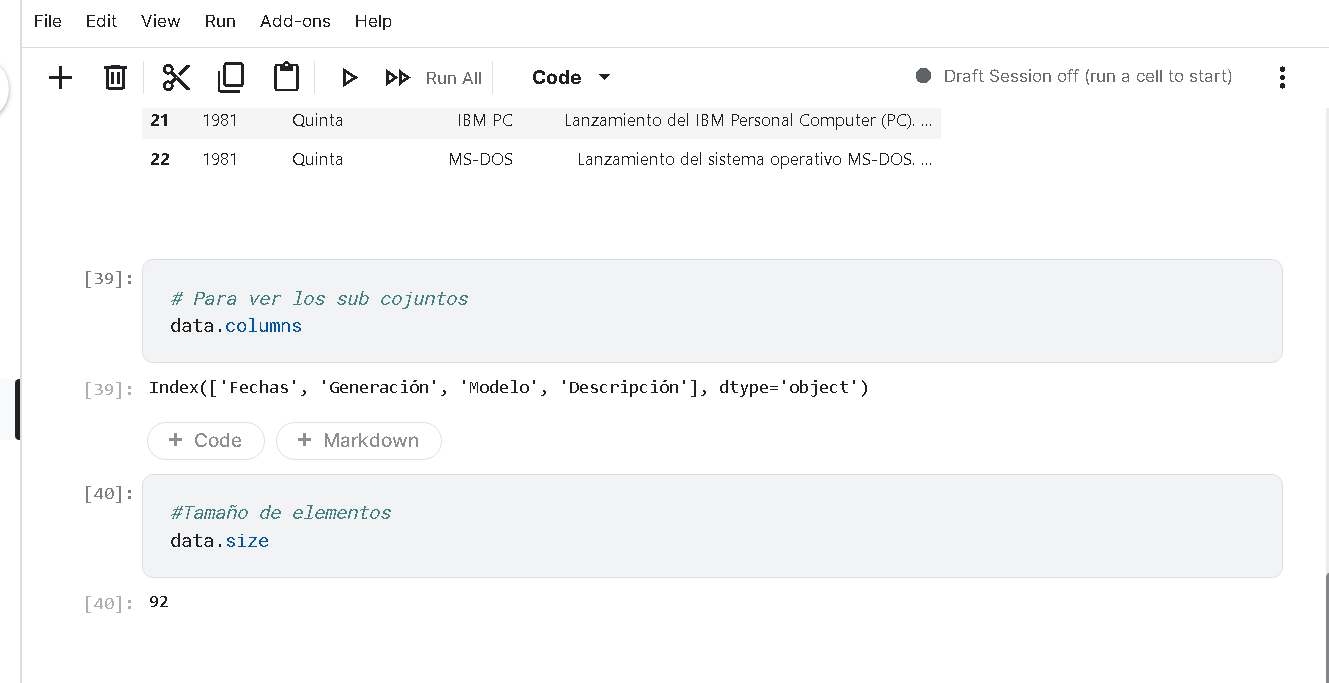


Ya que cargo nuestra base de datos, la ejecutamos para que descargue la información

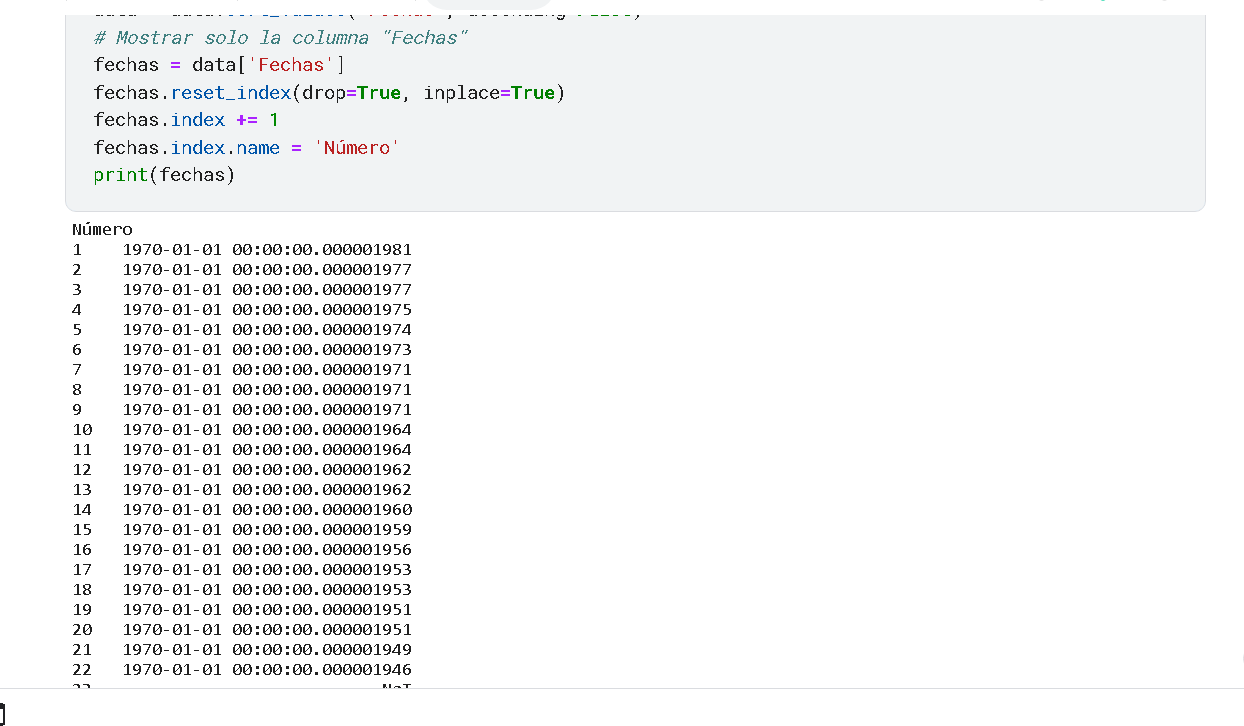


Creamos un objeto de la clase Dataframe usando nuestra base de datos y ponemos nuestra dirección

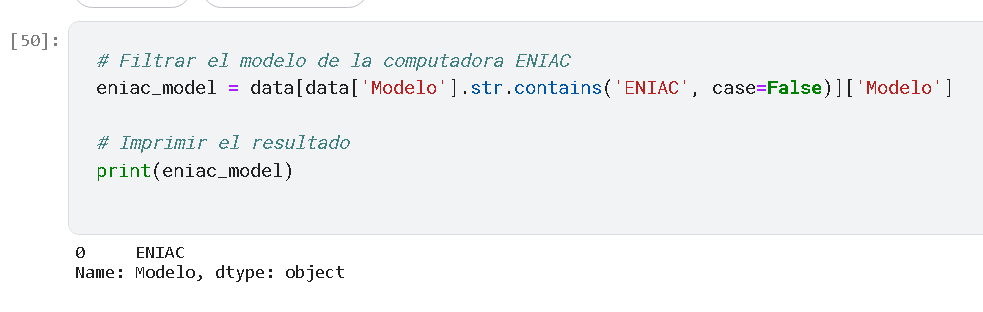




Para ver el nombre de las columnas y ver la cantidad de datos con size



Hago una consulta para las fechas de mayor a menor



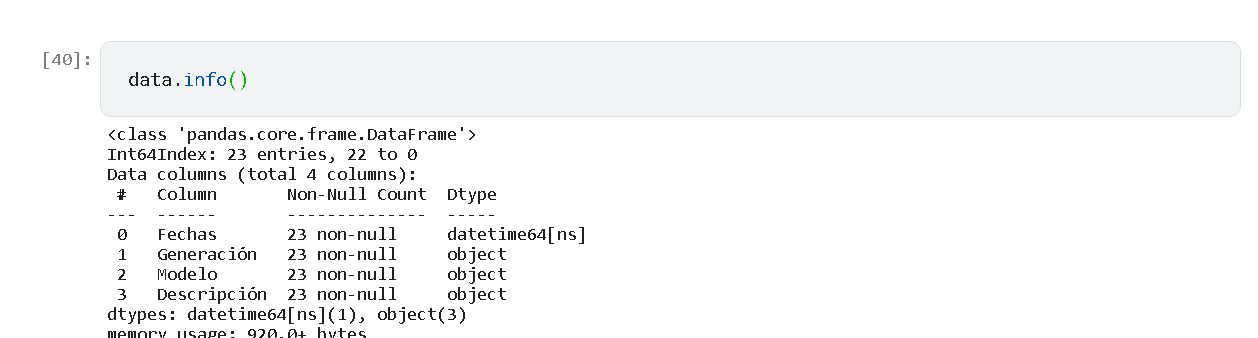
Consulto un modelo especifico de computadora y este me arroja el modelo

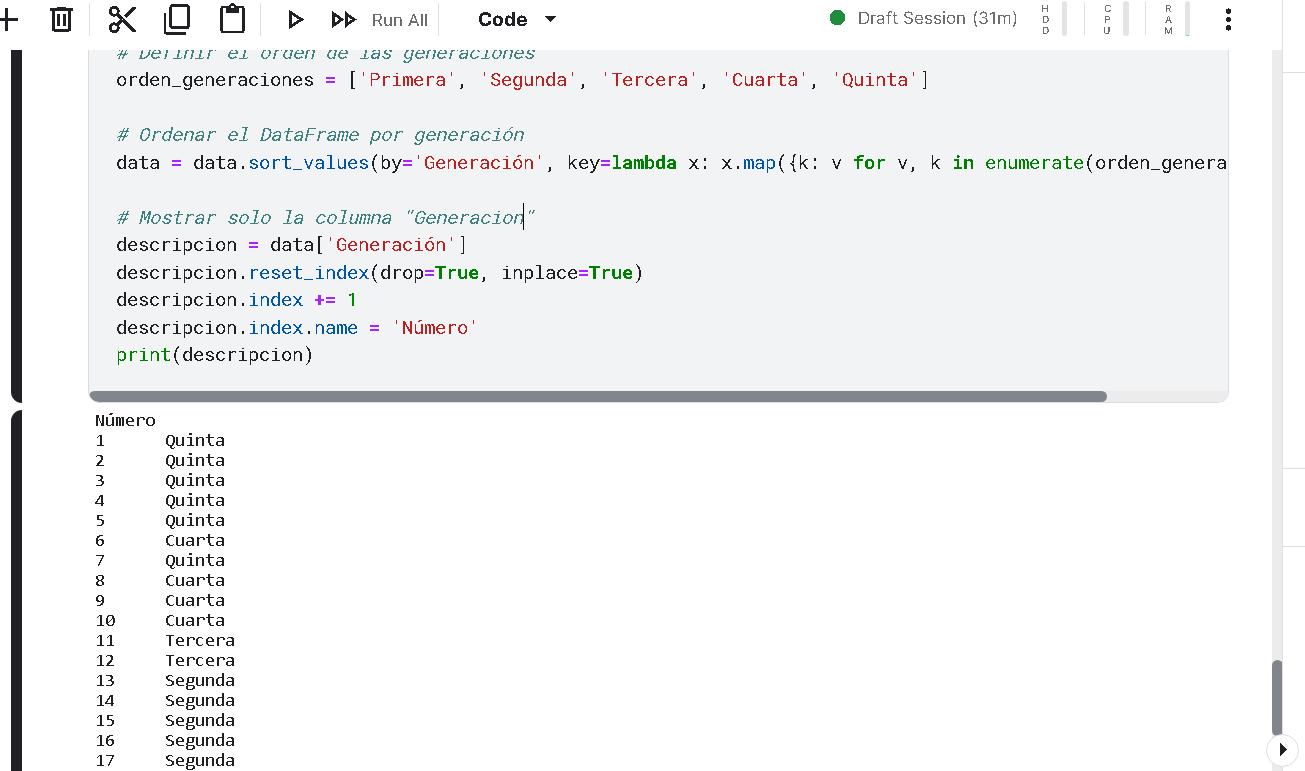


Consulta de los modelos de computadora de segunda generación

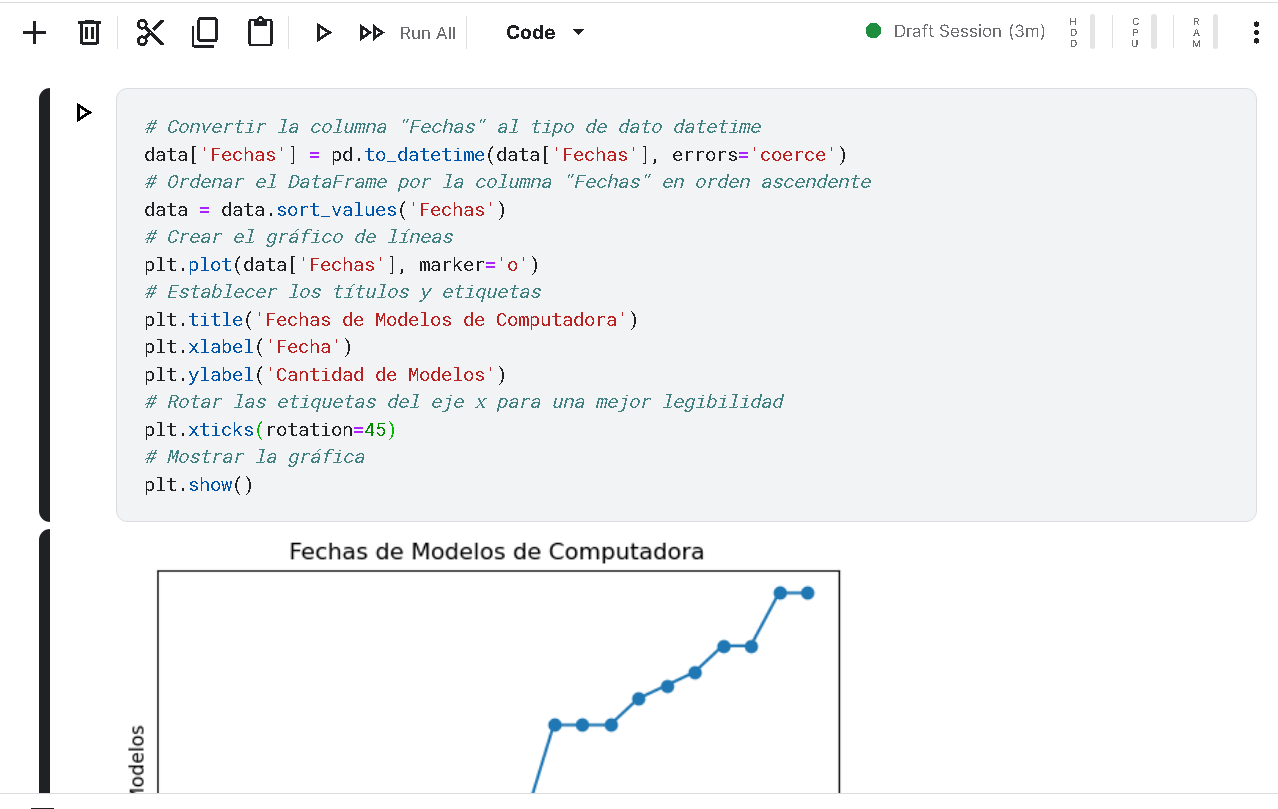


Enlista solamente la columna modelos, NOTA: los que dicen microprocesadores no fueron quitados por que al quitarlos dejaba de cargar el CSV son 3 filas.

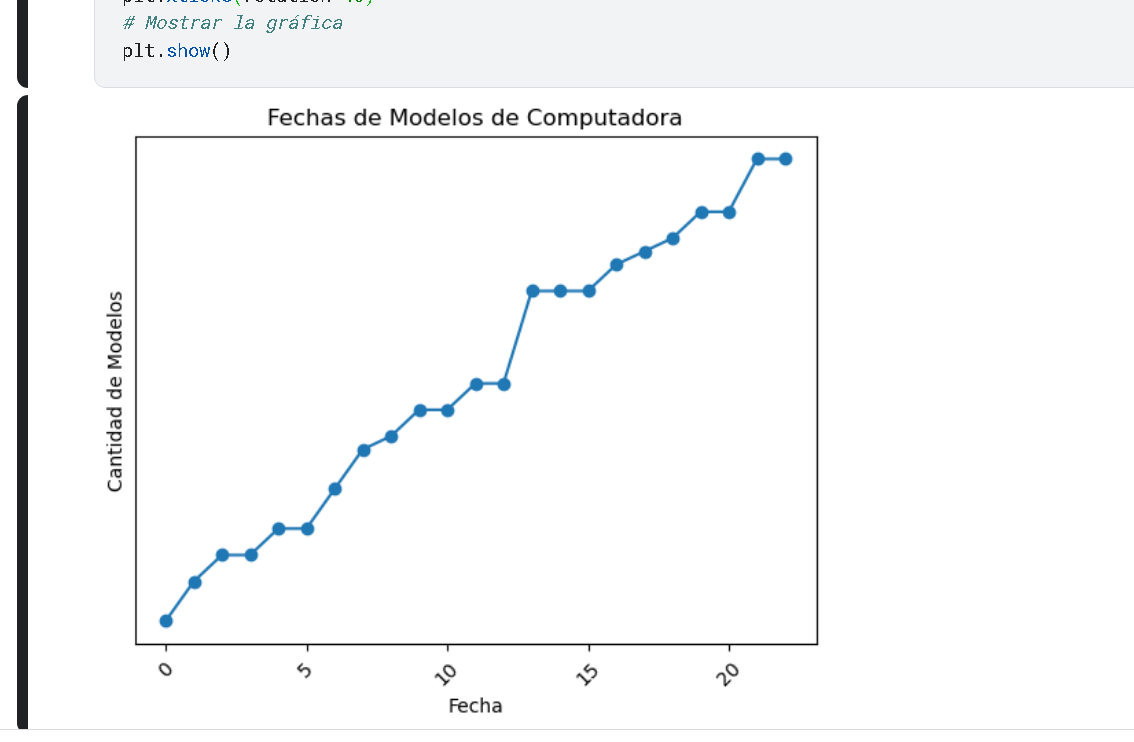




Mostrar en orden las generaciones de computadoras de la mas alta a la más baja.



Grafica para visualizar las fechas en un gráfico, puedes utilizar un gráfico de líneas para mostrar la evolución de las fechas a lo largo del tiempo.



De esta manera logro hacer mi investigación sobre la base de datos sacadas de un texto wiki donde puedo apreciar las generaciones de las computadoras y como con el tiempo han ido evolucionando.

Las generaciones de las computadoras representan diferentes etapas en el desarrollo de la tecnología informática, cada una marcada por avances significativos en términos de hardware, software y capacidad de procesamiento.

**Objetivo de usar OpenRefine:**

El objetivo de openrefine es poder limpiar base de datos y corregir errores comunes en los conjuntos de datos, como valores incorrectos, inconsistentes o duplicados. También ofrece herramientas para estandarizar formatos, eliminar espacios en blanco y corregir errores de formato.

En este caso como lo mencione la en el texto pasado cada una de las cosas mencionadas se las he aplicado al texto wiki con el objetivo de limpiar mi texto y para posteriormente pasarlo a un CSV el cual podre usar para un análisis de base de datos.

**Objetivo de usar Kaggle:**

En mi proyecto de investigación sobre las generaciones de las computadoras, decidí utilizar Kaggle como plataforma para compartir, colaborar y obtener retroalimentación en mi trabajo.

A través de Kaggle, pude compartir mis hallazgos con la comunidad de científicos de datos y entusiastas de la tecnología, recibir reconocimiento por mi trabajo, colaborar con otros investigadores interesados en el tema, participar en competencias relacionadas y acceder a recursos adicionales para enriquecer mi investigación.

Utilizar Kaggle fue una experiencia valiosa que me permitió mejorar mi proyecto y obtener perspectivas adicionales sobre las generaciones de las computadoras.

**Objetivo de esta investigación:**

Es apreciar las fechas, modelos, generación y descripción de las computadoras pero no hacerlo en forma de investigación si no poder vivir la experiencia de lo que es un proceso de limpieza de base de datos así como la estructuración de estas.