

Práctica Tercera Clase

Para esta práctica utilizaremos la herramienta OpenRefine. Se puede descargar desde <http://openrefine.org/> y corre en cualquier versión de sistema operativo. Para esta práctica ya contamos con la herramienta descargada en la PC (en el escritorio dentro de la carpeta openrefine-win-2.7-rc.2). La misma al ejecutarse abre una ventana de la línea de comandos y el explorador de internet por defecto, que es donde se usa la herramienta.

OpenRefine se define como una herramienta libre y abierta para trabajar con datos llenos de problemas.

Caso Lectores de suplemento Horizonte Agropecuario.

La editorial que imprime y distribuye el suplemento Horizonte Agropecuario, dispone su lista de lectores en una planilla Lectores.xlsx, la que utiliza para imprimir las etiquetas que luego son utilizadas en el franqueo de los envíos, en la misma a lo largo del tiempo se le han ido agregando filas únicamente. Cada fila se agrega al principio del listado por lo que no se controla si figura o no el destinatario en la lista. Solo se tiene en cuenta el código postal que se encuentra en otra hoja de la planilla para la carga del mismo, junto con la dirección y el nombre del destinatario.

Estructura de los datos.

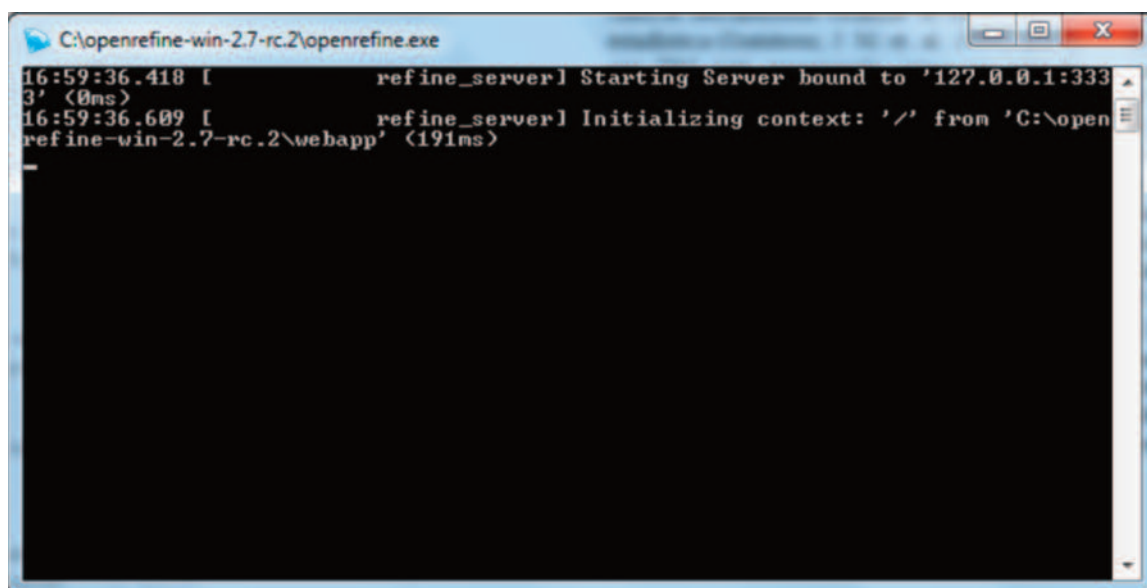
Lectores		Localidades	
Nombre	Texto	Código Postal	Numérico
Dirección	Texto	Localidad	Texto
Código Postal	Numérico	Provincia	Texto

- 1) Descargar el archivo Lectores.xlsx de la carpeta compartida del taller y abrir la planilla de Excel y realizar una vista rápida de los datos. ¿Se pueden encontrar algunos problemas de consistencia?

Toda la edición de los datos se realizará con la herramienta Open Refine, que realiza una importación de los datos y nos permite dejar el archivo original intacto cumpliendo una de las buenas prácticas del manejo de datos.

- 2) Ejecutar Open Refine, desde el idono en el escritorio. Al hacer doble clic sobre dicho icono se debe abrir un ventana de comandos en la que se ejecuta Open-Refine. De forma predeterminada, la ventana de comandos tiene un fondo negro:

Además se abrirá el explorador de internet que esté configurado por defecto.

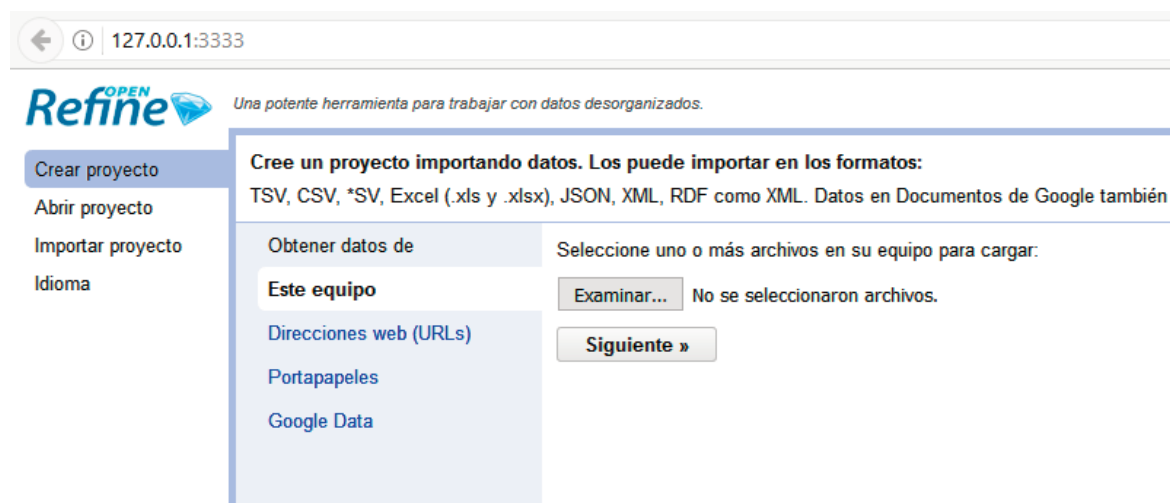


Dentro de la pagina en el explorador sobre el lado izquierdo hay un grupo de solapas: Crear Proyecto, Abrir Proyecto, Importar Proyecto e Idioma. Si al ejecutar el programa no aparece en la versión en español, ingresando en la solapa "Language Settings" y configurar la opción "Español".

Por defecto la herramienta se encuentra en la solapa Crear Proyecto, donde da una serie de alternativas, para esta práctica utilizaremos "Este Equipo"

3) Conformación del set de datos Lectores.

- 1) Dentro de la solapa "Crear Proyecto", en la opción "Este Equipo", hacer clic en el botón "Examinar" y buscar el archivo "Lectores.xlsx" descargado de la carpeta compartida.
- 2) Una vez encontrado el archivo, seleccionar y presionar el botón "Si-



guiente>>”. De esta forma pasa a la pantalla de preparación de la importación de los datos.

- 3) Seleccionar en “Hojas a Importar”, la hoja “Datos”. Configurar las siguientes

Una potente herramienta para trabajar con datos desorganizados.

Crear proyecto
Abrir proyecto
Importar proyecto
Idioma

« Inicio Configurar opciones de carga

Nombre del proyecto: Lectores.xlsx Crear proyecto »

Column 1	Column 2	Column 3
1.		
2.		
3.		
4.		
5.		
6.	Nombre	Direccion
7.	Bianchi, Raúl	Pellegrini 184
8.	Perdiguez, Enrique	Moreno 275
9.	Rodriguez, Juan Jesús	Coronel Plaza 563
10.	Cabral, Diego	C.C N° 79
11.	Barcia Adolfo	Sarmiento 965
12.	López Marcelino	Moreno S/N

Abrir archivo como

Excel files
JSON files
Line-based text files
CSV / TSV / separator-based files

Hojas a importar
☒ Datos 1476 rows
☐ Localidades 0 rows

☐ Ignorar primera(s) 0
☒ Seleccionar primera(s) 1
☐ Descartar primera(s) 0
☐ Cargar al menos 0

linea(s) al inicio del archivo
linea(s) para los nombres de las col
fila(s) de datos
fila(s) de datos

☐ Cargar filas en blanco
☐ Cargar celdas en blanco como nulas
☐ Cargar el origen del archivo (nombres, URLs) en cada fila

Actualizar vista previa

tes opciones, marcar la opción “Ignorar primeras”, colocar el número 6 en líneas al inicio del archivo. Por defecto aparece seleccionada la opción “Seleccionar Primeras” la cual no hay que modificar.

- 4) Sobre la esquina superior izquierda del cuadro editar el nombre del proyecto a “Lectores” y hacer clic en crear proyecto.

- 5) La siguiente pantalla es donde se lleva a cabo toda la manipulación de los datos. Observar que al lado de cada nombre de columna, se genera automáticamente un botón en forma de flecha (▼) donde se encuentran todas las opciones para manipular los datos de cada una de ellas.

Refine Lectores Enlace permanente

Facetas / Filtros Deshacer / Rehacer

1475 filas

Mostrar como: filas registrosMostrar: 5 10 25 50 filas

▼ Todo ▼ Nombre ▼ Direccion ▼ Codigo postal ▼ Column 4

1.	Bianchi, Raúl	Pellegrini 184	5600
2.	Perdiguez, Enrique	Moreno 275	5600
3.	Rodriguez, Juan Jesús	Coronel Plaza 563	5600
4.	Cabral, Diego	C.C N° 79	5600
5.	Barcia Adolfo	Sarmiento 965	5711
6.	López Marcelino	Moreno S/N	5711
7.	Arce, Orlando	Belgrano 235	5711
8.	Funes María Teresa	Barrio Estación. Calle 15	5711
9.	Consortio De Regantes.	Avda Colón S/N	5711
10.	Tissera J A	Avda Colón S/N	5711

Usar facetas y filtros

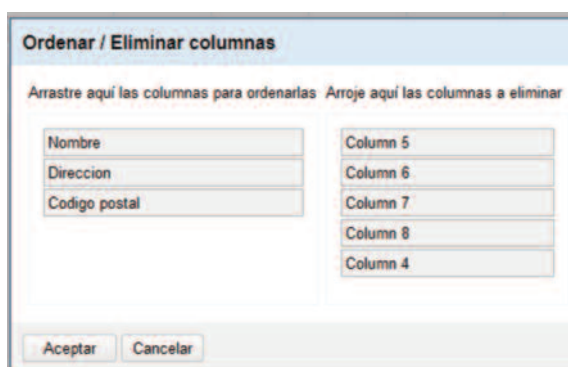
Use las facetas y los filtros para seleccionar subconjuntos de sus datos y trabajar en ellos. Puede encontrar estas opciones en los menús de cada columna.

¿Problemas para comenzar?
Vea los videos de ayuda

- 6) Eliminar las columnas sobrantes o sin datos, para ello hay que hacer clic en la flecha a la derecha de la columna "Todo", seleccionar la opción "Ordenar/Eliminar Columnas..." del menú "Editar Columnas".



- 7) A continuación en la ventana abierta para la edición de las columnas, arrastrar las columnas que se desean eliminar al panel de la derecha y en caso de querer alterar el orden de las columnas con datos se puede hacer dentro del mismo panel



- 8) Una vez que la planilla este con el formato deseado proceder a identificar valores de lectores similares y corregirlos, para ello sobre la col. de "Nombre" seleccionar del menú "Editar Celdas" la opción "Agrupar y editar..."



- 9) Dentro de la ventana que se despliega, observar que automáticamente la herramienta identifica los "Nombres" similares y los agrupa para su edición, la misma se realizará si se selecciona la opción "¿Unir?", en el cuadro de texto de la izquierda de cada grupo se podrá editar el contenido final de las celdas que se unan. Por último y luego de hacer todos los cambios deseados, hacer clic en "Unir seleccionados y Cerrar".

Agrupar y editar valores en la columna "Nombre"

Esta función le permite encontrar agrupaciones de diferentes valores que pueden ser representaciones alternativas de la misma cosa. Por ejemplo, "New York" y "new york" probablemente se refieren al mismo concepto, solo se presenta diferencia en la capitalización. De la misma manera "Gödel" y "Godel" probablemente se refieren a la misma persona. [Más información ...](#)

Método: **Colisión de llaves** Función: **Huella** 11 clusters Encontrado

Número de valores	Número de filas	Valores en la agrupación	¿Unir?	Nuevo valor de las celdas
2	2	<ul style="list-style-type: none">Richardson Carlos (1 rows)Richardson, Carlos (1 rows)	<input type="checkbox"/>	Richardson Carlos
2	3	<ul style="list-style-type: none">Mauricio Rausch (2 rows)Rausch, Mauricio (1 rows)	<input type="checkbox"/>	Mauricio Rausch
2	2	<ul style="list-style-type: none">Nestor Hernes (1 rows)Néstor Hernes (1 rows)	<input type="checkbox"/>	Nestor Hernes
2	2	<ul style="list-style-type: none">Weigel Esther (1 rows)Weigel, Esther (1 rows)	<input type="checkbox"/>	Weigel Esther
2	2	<ul style="list-style-type: none">Cerutti Dante (1 rows)Cerutti, Dante (1 rows)	<input type="checkbox"/>	Cerutti Dante
2	2	<ul style="list-style-type: none">Fernandez Juan Manuel (1 rows)Fernández Juan Manuel (1 rows)	<input type="checkbox"/>	Fernandez Juan Manuel
2	2	<ul style="list-style-type: none">Escuela Agrotecnica (1 rows)	<input type="checkbox"/>	Escuela Agrotecnica

Seleccionar todos Seleccionar ninguno Exportar agrupaciones Unir seleccionados y reagrupar Unir seleccionados y cerrar Cerrar

Filas en la agrupación: 2 — 3

Longitud promedio de los valores: 13 — 29

Varianza de los valores: 0 — 0.5

- 10) Ahora eliminar todos los registros que no cuentan con "Codigo Postal" ya que sin ese dato no se puede identificar a qué localidad pertenece el "Lector": En la columna "Codigo Postal", seleccionar del menú "Facetas", el sub menú "Facetas personalizadas" opción "Facetas por blanco". Esto mostrará en el panel de la izquierda las opciones "true" y "false", seleccionar la opción "false" para ver todos los registros que no cuentan con este dato. Luego en la columna "Todo", seleccionar el menú "Editar filas" opción "Eliminar todas las filas que encajen". De esta forma se eliminaran todas las filas sin "Codigo postal"
- 11) Por último identificar los registros duplicados, en cada columna del set de datos seleccionar del submenú "Facetas personalizadas" del menú "Facetas", la opción "Faceta por duplicados". De esta forma irá apareciendo en el panel de la izquierda dos opciones "true" y "false" para cada una, al finalizar la misma operación para todas las columnas, seleccionar la opción "true" de cada una de las ventanas del panel de la izquierda. Ahora en la

Taller de manejo de datos con herramientas libres

Nombre	Dirección	Código postal
Facetas	Faceta de texto	6300
Filtro de texto	Faceta numérica	6300
Editar celdas	Faceta de línea de tiempo	6319
Editar columnas	Faceta gráfica	6213
Transponer	Faceta de texto personalizada...	6319
Ordenar	Faceta numérica personalizada...	6389
Ver	Facetas personalizadas	Faceta por palabra
Cotejar	Emilio Mitre 790	Faceta por duplicados
Hernes, Nestor	13 De Caballería Oeste 2050	Faceta log. numérica
Hernes, Nestor	13 De Caballería Oeste 2050	Faceta log. numérica acotada-1
Intendencia Municipal		Faceta por longitud de texto
Intendencia Municipal		Faceta por longitud log. de texto
Intendencia Municipal		Faceta por caracteres Unicode
Intendencia Municipal		Faceta por error
José Mikolaszek	Unanue 1785	Faceta por blanco
José Mikolaszek	Unanue 1785	
Juan Pablo De La Iglesia	A. Illia 643	6300

columna "Código Postal" seleccionar la opción "Ordenar...", en la ventana emergente seleccionar "números" y "Aceptar" de esta forma nos quedan los registros duplicados por localidad. Aparece la palabra "Sort" sobre la

Refine Lectores [Enlace permanente](#)

Facetas / Filtros [Deshacer / Rehacer 4](#)

Actualizar Restablecer todos Remover todos

Nombre [cambiar invertir restaurar](#)

2 choices Ordenar por: A-Z conteo

false 381

true 98 [exclude](#)

Facetas por conteo de opciones

Dirección [cambiar invertir restaurar](#)

2 choices Ordenar por: A-Z conteo

false 23

true 98 [exclude](#)

Facetas por conteo de opciones

Código postal [cambiar invertir restaurar](#)

1 choices Ordenar por: A-Z conteo

true 98 [exclude](#)

Facetas por conteo de opciones

98 matching filas (1360 total)

Mostrar como: [filas](#) [registros](#) [Mostrar: 5 10 25 50 filas](#)

	Todo	Nombre	Dirección	Código postal
☆	16.	Aguilera Lucas	A. Argentina 75 Dto "B"	6300
☆	17.	Aguilera Lucas	A. Argentina 75 Dto "B"	6300
☆	97.	Asociación Agrícola Ganadera	Chaco 401	6300
☆	98.	Asociación Agrícola Ganadera	Chaco 401	6300
☆	104.	Asociación Rural Realicó	Italia 1517	6200
☆	105.	Asociación Rural Realicó	Italia 1517	6200
☆	106.	Asociación Rural Realicó	Italia 1517	6200
☆	156.	Belcher, Antonio		6319
☆	341.	Cooperativa Agropecuaria		6203
☆	342.	Cooperativa Agropecuaria		6207
☆	448.	El Diario	25 De Mayo Y San Martín	6300
☆	449.	El Diario	25 De Mayo Y San Martín	6300
☆	467.	Escuela Agropecuaria		6319
☆	501.	Fernandez, Juan Manuel	España 144	6213
☆	502.	Fernandez, Juan Manuel	España 144	6213
☆	579.	García, Jorge		6319
☆	624.	Gracia, Gregorio		6389
☆	625.	Gracia, Gregorio		6389
☆	647.	Héctor García Casarota	Emilio Mitre 790	6300
☆	648.	Héctor García Casarota	Emilio Mitre 790	6300

fila de los nombre de columnas, seleccionar y elegir la opción "Ordenar filas permanentemente" así fijamos el orden dado. De esta forma podremos identificar los registros que realmente sean duplicados, marcar con una estrella y luego seleccionar el filtro por estrellas: menú "Facetas" opción "Facetas con estrellas". Luego en la columna "Todo", seleccionar el menú "Editar filas" opción "Eliminar todas las filas que encajen".

4) Conformación del set de datos Localidades

- 1) Dentro de la solapa Crear Proyecto, en la opción "Este Equipo", hacer click en el botón "Examinar" y buscar el archivo "Lectores.xlsx" descargado de la carpeta compartida.
- 2) Una vez encontrado el archivo, seleccionar y presionar el botón "Siguiente>>". De esta forma pasa a la pantalla de preparación de la importación de los datos.
- 3) Seleccionar en "Hojas a Importar", la hoja "Localidades". Por defecto aparece seleccionada la opción "Seleccionar Primeras" la cual no hay que modificar.
- 4) Sobre la esquina superior izquierda del cuadro editar el nombre del proyecto a "Localidades" y hacer click en crear proyecto.
- 5) En el caso de las "Localidades" no existen columnas de más, por lo que se dejan tal cual como están.
- 6) Para empezar con la edición de los datos realizamos un ordenamiento por provincias, seleccionando la opción "Ordenar..." del menú de esta columna. Seguido seleccionar la palabra "Sort" y luego la opción "Reordenar filas permanentemente". De esta forma logramos tener los datos ordenados por Provincias para una mejor edición.
- 7) En la columna "Provincia" seleccionar el "Agrupar y editar..." del menú "Editar celdas" de esta forma y como sucedió con los datos de "Lectores" la herramienta identifica los datos similares y los agrupará automáticamente, para poder elegir cuáles unir y definir el texto que llevará cada unión. Siguiendo con la edición de la columna "Provincia", seleccionar del menú "Facetas" la opción "Facetas de texto", sobre el margen izquierdo aparece un panel con las apariciones que cada uno de los datos de "Provincia". Dentro del panel de "Provincia", hacer clic sobre algunos de los nombres, esto genera un filtro y nos permite "editar" al seleccionar esta opción dentro del renglón del nombre. Cambiar los nombres "LP" por "La Pampa" y "SL" por "San Luis", de esta forma la herramienta automáticamente los agrupara con los otros datos que coincidan. Esto permite corregir errores de escritura, abreviaturas y demás que hacen que datos que significan una misma cosa se tengan en cuenta como distintos.

Taller de manejo de datos con herramientas libres

Refine Localidades [Enlace permanente](#)

Facetas / Filtros [Deshacer / Rehacer 2](#)

Actualizar Restablecer todos Remove todos

Provincia [cambiar](#)

8 choices Ordenar por: A-Z conteo [Agrupar](#)

Buenos Aires 6
Cordoba 1
La Pampa 47
LP 3
Mendoza 1
Rio Negro 1
San Luis 8
SL 2

Facetas por conteo de opciones

69 filas

Mostrar como: **filas** registrosMostrar: 5 10 25 50 filas Sort ▾

Todo	Codigo postal	Localidad	Provincia
1.	8187	Bordenave	Buenos Aires
2.	6400	Trenque Lauquen	Buenos Aires
3.	8000	Bahía Blanca	Buenos Aires
4.	8132	Médanos	Buenos Aires
5.	8136	Algarrobo	Buenos Aires
6.	8146	Mayor Buratovich	Buenos Aires
7.	5800	Río Cuarto	Cordoba
8.	6205	Ingeniero Luiggi	La Pampa
9.	6200	Realicó	La Pampa
10.	6203	E. Martini	La Pampa
11.	6203	Embajador Martini	La Pampa

- 8) Realizar similar tarea con los datos de "Localidad", seleccionar la opción "Faceta de texto" del menú "Facetas", aparece en el panel de la izquierda otra ventana, ahora con los nombres de las localidades agrupados por cantidad de apariciones. Para la edición de la misma se podrá utilizar el filtro de provincias, de esta forma se evitará eliminar localidad con el mismo nombre pero de diferentes "Provincias".

Refine Localidades [Enlace permanente](#)

Facetas / Filtros [Deshacer / Rehacer 5](#)

Actualizar Restablecer todos Remove todos

Provincia [cambiar](#) [invertir](#) [restaurar](#)

6 choices Ordenar por: A-Z conteo [Agrupar](#)

Buenos Aires 6
Cordoba 1
La Pampa 50 [exclude](#)
Mendoza 1
Rio Negro 1
San Luis 10

Facetas por conteo de opciones

Localidad [cambiar](#)

45 choices Ordenar por: A-Z conteo [Agrupar](#)

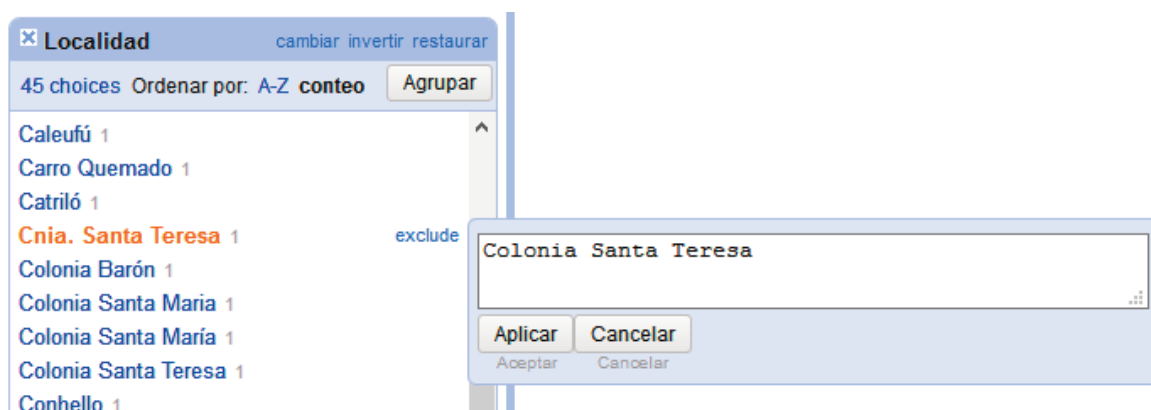
Ataliva Roca 2
Doblas 2
Parera 2
Rancul 2
Santa Rosa 2

50 matching filas (69 total)

Mostrar como: **filas** registrosMostrar: 5 10 25 50 filas

Todo	Codigo postal	Localidad	Provincia
8.	6309	Alpachiri	La Pampa
9.	6207	Alta Italia	La Pampa
10.	6326	Anguil	La Pampa
11.	6301	Ataliva Roca	La Pampa
12.	6308	Ataliva Roca	La Pampa
13.	8204	Bernasconi	La Pampa
14.	6387	Caleufú	La Pampa
15.	6319	Carro Quemado	La Pampa
16.	6330	Catrilo	La Pampa
17.	6312	Cnia. Santa Teresa	La Pampa
18.	6315	Colonia Barón	La Pampa
19.	8214	Colonia Santa Maria	La Pampa
20.	8214	Colonia Santa Maria	La Pampa
21.	6312	Colonia Santa Teresa	La Pampa
22.	6381	Conhello	La Pampa
23.	6305	Doblas	La Pampa
24.	6306	Doblas	La Pampa

Con el filtro de "Provincia" y de "Localidades", comenzar a controlar la cantidad de apariciones de las mismas, pudiendo editar tanto en el panel de la izquierda como en las celdas donde se encuentran los datos.



Edición en el panel de Filtro (pasando el cursor por el renglón de la localidad se observan las opciones "editar" e "include")



Edición en celda (al momento de editar de esta forma en necesario configurar el tipo de dato de la celda)

Después de realizar todas estas modificaciones, realizar el ordenamiento de nuevo, seleccionar en la columna "Provincia" la opción "Ordenar...", hacer lo mismo con "Localidad" de esta forma quedan en filas consecutivas las "Filas duplicadas". Seleccionando la palabras "Sort" sobre los nombres de las columnas y luego la opción "Reordenar filas permanentemente". Para eliminar dichas filas, seleccionar en la columna "Localidad" la opción "Vaciar hacia abajo" del menú "Editar celdas". Así se borran los datos de "Localidad" duplicados, para eliminar las filas con celdas vacías en la columna "Localidad", seleccionar en esta la opción "Faceta por blanco", que se encuentra en el sub menú "Facetas personalizadas" del menú "Facetas". Esto muestra en el panel de la izquierda las opciones "true" y "false" para "Localidad", seleccionar "true" y así se muestran todas las filas sin dato de

Taller de manejo de datos con herramientas libres

Refine Localidades Enlace permanente

Facetas / Filtros Deshacer / Rehacer 18

Actualizar Restablecer todos Remover todos

Provincia cambiar
6 choices Ordenar por: A-Z conteo Agrupar

Buenos Aires 6
Cordoba 1
La Pampa 50
Mendoza 1
Rio Negro 1
San Luis 10
Facetas por conteo de opciones

Localidad cambiar
56 choices Ordenar por: A-Z conteo Agrupar

Santa Rosa 3
Villa Mercedes 3
Ataliva Roca 2
Colonia Santa María 2
Colonia Santa Teresa 2
Doblas 2
Embajador Martini 2
General Acha 2
Monte Nievas 2
Parera 2
Rancul 2
Algarrobo 1
Alpachiri 1
Alta Italia 1
Anchorena 1
Anguil 1
Bahía Blanca 1
Bernasconi 1
Bordenave 1

69 filas

Mostrar como: filas registros Mostrar: 5 10 25 50 filas Sort ▾

Todo	Codigo postal	Localidad	Provincia
1.	8136	Facetas	Buenos Aires
2.	8000	Filtro de texto	Buenos Aires
3.	8187	Editar celdas	Buenos Aires
4.	8146	Editar columnas	Buenos Aires
5.	8132	Transponer	Buenos Aires
6.	6400	Ordenar...	Buenos Aires
7.	5800	Ver	Buenos Aires
8.	6309	Cotejar	Buenos Aires
9.	6207	Ataliva Roca	La Pampa
10.	6326	Ataliva Roca	La Pampa
11.	6308	Bernasconi	La Pampa
12.	6308	Caleufú	La Pampa
13.	8204	Carro Quemado	La Pampa
14.	6387	Catriló	La Pampa
15.	6319	Colonia Santa Teresa	La Pampa
16.	6330	Colonia Barón	La Pampa
17.	6312	Colonia Santa María	La Pampa
18.	6315	Colonia Santa María	La Pampa
19.	8214	Colonia Santa María	La Pampa
20.	8214	Colonia Santa Teresa	La Pampa
21.	6312	Conhelo	La Pampa
22.	6381	Doblas	La Pampa
23.	6305	Doblas	La Pampa
24.	6305	Embajador Martini	La Pampa
25.	6203	Eduardo Castex	La Pampa
26.	6380	Embajador Martini	La Pampa
27.	6203	General Acha	La Pampa
28.	8200	General Pico	La Pampa
29.	6360	General San Martín	La Pampa
30.	8206	General Acha	La Pampa
31.	8200	General Acha	La Pampa

"Localidad". En la columna "Todo" seleccionar "Eliminar todas las filas que encajan" del menú "Editar filas".

Como último paso para cerrar la edición de este set de datos, cerrar todas las ventanas en el panel de la izquierda. De esta forma aparecerán los datos que quedaron. Observar que siguen existiendo datos duplicados, esta última edición la realizaremos marcando los registros repetidos que visualmente encontremos. Para ello en la columna "Localidad" seleccionar la opción "Faceta de texto" en el menú "Facetas", vemos que el panel de la izquierda nos muestra nuevamente las localidades dentro de las que vemos algunas con 2 apariciones, ordenamos por "conteo" y seleccionamos con la opción "include" de cada renglón las que tengan más de una aparición. Ahora con el listado de las repetidas marcar con una estrella los registros a eliminar.

Taller de manejo de **datos** con **herramientas libres**

Refine Localidades [Enlace permanente](#)

Facetas / Filtros [Deshacer / Rehacer 25](#)

Actualizar Restablecer todos Remover todos

Localidad cambiar invertir restaurar

56 choices Ordenar por: A-Z conteo Agrupar

- Colonia Santa Teresa 2 exclude
- Embajador Martini 2 exclude
- General Acha 2 exclude
- Monte Nieves 2 exclude
- Algarrobo 1
- Alpachiri 1
- Alta Italia 1

8 matching filas (60 total)

Mostrar como: **filas** registrosMostrar: 5 10 25 50 filas

Todo	Codigo postal	Localidad	Provincia
☆	16.	6312	Colonia Santa Teresa La Pampa
★	19.	6312	Colonia Santa Teresa La Pampa
☆	22.	6203	Embajador Martini La Pampa
★	24.	6203	Embajador Martini La Pampa
☆	25.	8200	General Acha La Pampa
★	28.	8200	General Acha La Pampa
☆	36.	6383	Monte Nieves La Pampa
★	39.	6383	Monte Nieves La Pampa

Seleccionar en la columna "Todo", la opción "Faceta por estrellas" en el menú "Facetas", aparece la ventana en el panel izquierdo, con las opciones "true" y "false", seleccionar "true" en dicha ventana. Luego seleccionar la opción "Eliminar todas las filas que encajan" del menú "Editar filas" de la columna "Todo". Cerramos nuevamente todas las ventanas del panel izquierdo.

5) Unificar los set de datos Lectores y Localidades.

- 1) Para comenzar con la última parte de la consolidación de los datos del caso de estudio abrir nuevamente el set de datos Lectores, con el botón "Abrir..." que se encuentra en el rincón superior derecho.

Refine Localidades [Enlace permanente](#) [Abrir...](#) [Exportar](#) [Ayuda](#)

Facetas / Filtros [Deshacer / Rehacer 26](#)

56 filas Extensiones:

Mostrar como: **filas** registrosMostrar: 5 10 25 50 filas « primera < anterior 1 - 50 siguiente > última »

Todo	Codigo postal	Localidad	Provincia
☆	1.	8136	Algarrobo Buenos Aires
☆	2.	8000	Bahia Blanca Buenos Aires
☆	3.	8187	Bordenave Buenos Aires
☆	4.	8146	Mayor Buratovich Buenos Aires
☆	5.	8132	Médanos Buenos Aires
☆	6.	6400	Trenque Lauquen Buenos Aires
☆	7.	5800	Rio Cuarto Cordoba
☆	8.	6309	Alpachiri La Pampa
☆	9.	6207	Alta Italia La Pampa
☆	10.	6326	Anguil La Pampa
☆	11.	6308	Ataliva Roca La Pampa
☆	12.	8204	Bernasconi La Pampa
☆	13.	6387	Caleufú La Pampa
☆	14.	6319	Carro Quemado La Pampa

Usar facetas y filtros

Use las facetas y los filtros para seleccionar subconjuntos de sus datos y trabajar en ellos. Puede encontrar estas opciones en los menús de cada columna.

¿Problemas para comenzar?
[Vea los videos de ayuda](#)

- 2) Hacer click en la solapa de la izquierda "Abrir proyecto", esto nos muestra todos los proyectos que se generaron. Seleccionar el proyecto "Lectores" para abrirlo nuevamente.

Taller de manejo de **datos** con **herramientas libres**

Refine OPEN

Una potente herramienta para trabajar con datos desorganizados.

Crear proyecto

Abrir proyecto

Importar proyecto

Idioma

Última modificación	Nombre
hoy 3:20 PM	Localidades
hoy 0:10 AM	Lectores
3 días atrás	Lectores
3 días atrás	Lectores
4 días atrás	Localidades
4 días atrás	Lectores.xlsx
4 días atrás	Lectores
4 días atrás	lectores horizonte
4 días atrás	Plano Anguil Trad LA PAMPA 2013 2014
4 días atrás	Datos Prueba OpenRefine

- 3) Ahora agregar nuevas columnas "Localidad" y "Provincia" al proyecto actual desde el proyecto "Localidades". Para esto usaremos la opción "Agregar columna basada en esta columna..." en el menú "Editar columnas" de la columna "Codigo Postal".

Refine OPEN Lectores Enlace permanente

Facetas / Filtros Deshacer / Rehacer 1307 filas

Mostrar como: **filas registros** Mostrar: 5 10 25 50 filas

Todo	Nombre	Dirección	Código postal
1. Abdala, Adalberto			
2. Abel Reñones	Pueyrredón 33		
3. Abraham, Idilio	Juan B. Justo 249		
4. Abrate Jorge	B 551, Viv. Mz 2110,		
5. Abrate Nicolás	Remedios De Escala		
6. Abt Hnos.			
7. Abt Pedro	San Martín Calle 13 C		
8. Acevedo Gladys Noemi	Jujuy 351		
9. Agrícola Bonansea	25 De Mayo 903		
10. Agrícola Villa Mercedes	Av. Aviador Origone		
11. Agroinsuños Avda Del Centenar			5773
12. Agronomía Los Caldenes			6313
13. Agropecuaria Tricol S.A.			6389
14. Agrosistemas. At. Ing. David Los	Av. Pte Perón 1879		5730
15. Agrotec S.A.	Av. S. Marzo 2185		6300
16. Aguilera Lucas	A. Argentina 75 Dto		6300
17. Aguilera Miguel	Don Bosco 768		8200
18. Aimar, Alcides	Belgrano Norte 620		6300
19. Aimar, Carlos			6214
20. Aimar, Eduardo			6214
21. Aimar, Fabián	A. Kardec 957		6300
22. Aimar, José	Mario Scofet S/N		6385
23. Aimar, Juan			6214

Usar facetas y filtros

Use las facetas y los filtros para seleccionar subconjuntos de sus datos y trabajar en ellos. Puede encontrar estas opciones en los menús de cada columna.

¿Problemas para comenzar?
Vea los videos de ayuda

Facetas

- Facetas
- Filtro de texto
- Editar celdas
- Editar columnas
 - Dividir en varias columnas...
 - Agregar columna basada en esta columna...
 - Agregar columna accediendo a URLs...
 - Renombrar esta columna
 - Eliminar esta columna
 - Mover columna al principio
 - Mover columna al final
 - Mover columna a la izquierda
 - Mover columna a la derecha
- Transponer
- Ordenar...
- Ver
- Cotejar

Al seleccionar esta opción se abre una ventana emergente en donde configurar opciones que nos va a permitir agregar nuevas columnas. Dentro de las configuraciones, completar el nombre de la nueva columna, en el primer caso sería "Localidad", dejar seleccionado "cambiar a en blanco", el lenguaje a utilizar será "GREL" como está configurado y en el cuadro "Expresión" completar con la expresión:

cell.cross("Localidades","Codigo postal").cells["Localidad"].value[0]

Agregar columna basada en la columna Código postal

Nuevo nombre de la columna:

☒ cambiar a en blanco ☐ guardar error ☐ copiar valor de la columna original

Expresión: No hay error de sintaxis.

Lenguaje:

Vista previa Historial Con estrella Ayuda

row	value	cell.cross("Localidades","Codigo postal").cells["Localidad"].value[0]
1.	6317	Luan Toro
2.	8000	Bahía Blanca
3.	6387	Caleufú
4.	5730	Villa Mercedes
5.	5730	Villa Mercedes
6.	8214	Colonia Santa Maria

Aceptar Cancelar

En el cuadro inferior se muestra una vista previa de la columna agregada.

Realizar la misma secuencia para agregar la columna de "Provincia", también basada en el "Codigo postal". La expresión que se debe utilizar para la nueva columna es la siguiente:

cell.cross("Localidades","Codigo postal").cells["Provincia"].value[0]

Como así también se debe tener en cuenta que el nombre de la columna debe ser "Provincia".

Luego de estos pasos contamos con el listado de lectores sin duplicados y con el dato de Localidad y Provincia.

Refine Lectores [Enlace permanente](#)

Facetas / Filtros Deshacer / Rehacer 65 **1307 filas**

Mostrar como: **filas** registrosMostrar: 5 10 25 50 filas

▼ Todo	▼ Nombre	▼ Direccion	▼ Codigo postal	▼ Localidad	▼ Provincia
☆	1. Abdala, Adalberto		6317	Luan Toro	La Pampa
☆	2. Abel Reñones	Pueyrredón 33	8000	Bahía Blanca	Buenos Aires
☆	3. Abraham, Idilio	Juan B. Justo 249	6387	Caleufú	La Pampa
☆	4. Abrate Jorge	B 551, Viv Mz 2110, Casa 25	5730	Villa Mercedes	San Luis
☆	5. Abrate Nicolás	Remedios De Escalada 104	5730	Villa Mercedes	San Luis
☆	6. Abt Hnos.		8214	Colonia Santa Maria	La Pampa
☆	7. Abt Pedro	San Martín Calle 13 Calle 15		Colonia Santa Maria	La Pampa
☆	8. Acevedo Gladys Noemi	Jujuy 351	6301	Ataliva Roca	La Pampa
☆	9. Agrícola Bonansea	25 De Mayo 903	5730	Villa Mercedes	San Luis
☆	10. Agrícola Villa Mercedes	Av. Aviador Origone 161	5730	Villa Mercedes	San Luis

Usar facetas y filtros

Use las facetas y los filtros para seleccionar subconjuntos de sus datos y trabajar en ellos. Puede encontrar estas opciones en los menús de cada columna.

¿Problemas para comenzar?
[Vea los videos de ayuda](#)

6) Registro de todos los cambios realizados

- 1) Open Refine lleva un registro de todos los cambios y acciones que realizamos sobre los datos trabajados en el proyecto. Si es necesario volver algún paso atrás o tener el historial de acciones realizadas sobre los datos, la herramienta los documenta por nosotros. Se pueden ver haciendo clic sobre la opción Deshacer/Rehacer:

Refine OPEN lectores horizonte.xlsx [Enlace permanente](#)

Facetas / Filtros **Deshacer / Rehacer 6**

Extraer... Aplicar...

Filtrar:

- 0. Create project
- 1. Mass edit 3 cells in column Provincia
- 2. Mass edit 3 cells in column Provincia
- 3. Mass edit 3 cells in column Provincia
- 4. Mass edit 1161 cells in column Provincia
- 5. Mass edit 1162 cells in column Provincia
- 6. Reorder columns**

1550 filas

Mostrar como: **filas** registros Mostrar: 5 10 25 50 filas « **primera** »

▼	▼	▼	▼	▼
Todo	Nombre	Direccion	Co	
★	1.	Dagnino, Ernesto	Zona Rural Rucanelo	
★	2.	Corral Hector O.	Zona Rural - Q. Quemu	
★	3.	De La Berrera Lucio	Zona Rural - Q. Quemu	
★	4.	El Verdadero Paraíso Srl.	Zona Rural	
★	5.	Viotti Horacio	Zona Rural	
★	6.	Prato Hugo	Zona Rural	
★	7.	Parodi, Hector	Zona Rural	
★	8.	Pardo, Oscar Y Hnos.	Zona Rural	
★	9.	Doratto Hnos.	Zona Rural	
★	10.	Boriolo, Valentin	Zona Rural	

7) Guardando los datos trabajados

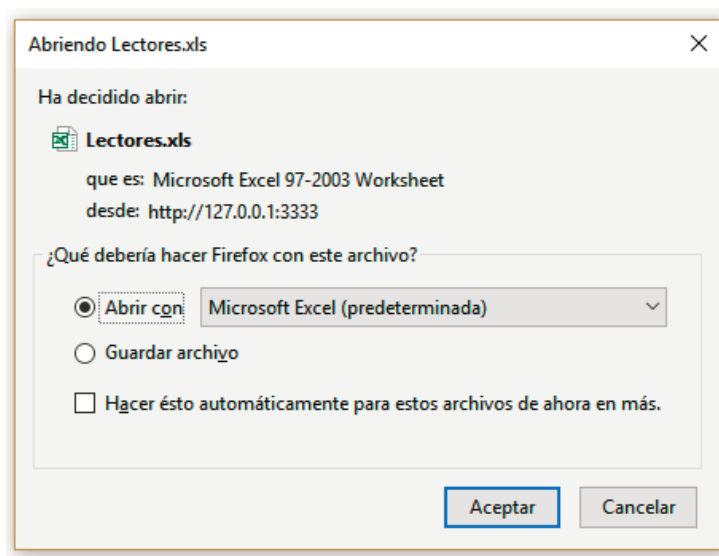
- 1) Para finalizar y obtener un archivo con los datos ya limpios y trabajados como necesitamos debemos hacer clic en "Exportar" y allí seleccionar el formato que más nos convenga.

Abrir... **Exportar ▼** Ayuda

- Exportar proyecto
- Delimitado por tabulaciones
- Delimitado por comas
- Tabla HTML
- Excel (.xls)
- Excel en XML (.xlsx)
- Hoja de cálculo ODF
- Triple loader (Freebase)
- MQLWrite (Freebase)
- Configurar exportación ...
- Plantilla ...

Hacer clic en el botón “Exportar” que se encuentra en la esquina superior derecha de la pantalla, allí visualizar todas las opciones de exportación junto con los formatos en los que se pueden exportar los datos, en este caso vamos a seleccionar el formato “Excel en XML (.xlsx)”.

Al seleccionar la opción se abrirá otra ventana en el explorador:



En esta nueva ventana seleccionar la opción “Guardar como” así poder elegir la ubicación del archivo final.