

Herramientas para la GGVD

Tableau Prep Builder



RICARDO SINAÍ VARGAS KUMUL <170300227@ucaribe.edu.mx>

FERNANDO GOMEZ PERERA <160300129@ucaribe.edu.mx>

CALVIN ALBERTO LOPEZ ALVAREZ <170300226@ucaribe.edu.mx>

29 de enero del 2021

INTRODUCCIÓN	2
DESCRIPCIÓN DE LA HERRAMIENTA	4
Conectarse a los datos	4
Explorar los datos	5
Limpiar los datos	6
Combinar los datos	8
Ejecutar el flujo y generar resultados	10
VENTAJAS Y DESVENTAJAS DE LA HERRAMIENTA	12
Ventajas	12
Desventajas	13
DESARROLLO DEL EJEMPLO	14
Análisis del problema	14
Conjuntos de datos a usar	14
Desarrollo del ejemplo	15
CONCLUSIONES	25
REFERENCIAS	26

INTRODUCCIÓN

Tableau Prep Builder es una herramienta del conjunto de productos de Tableau, diseñada para preparar datos de forma fácil e intuitiva. Permite combinar datos, darles forma y limpiarlos para su análisis en Tableau o para su exportación a otro sitio o a un archivo.

Dentro de esta herramienta, es posible conectar datos desde diversos archivos, servidores o extracciones de Tableau para su manipulación. De igual forma, permite combinar distintas fuentes de datos entre sí para crear nuevos conjuntos de datos.

Es una herramienta tipo *drag and drop*, donde es posible añadir los pasos que componen al proceso de ETL de forma interactiva. Estos pasos incluyen operaciones comunes como filtrar los datos, renombrar las columnas, crear tablas dinámicas, unir filas y columnas, entre otras.

La composición de dichos pasos forman un *flujo*, el cual puede considerarse un grafo de tipo *DAG (Directed Acyclic Graph o Grafo Dirigido Acíclico)*¹, donde cada nodo es uno de los pasos que componen al proceso de ETL: Ya sea una *Entrada*, una *Salida* o una operación de *Transformación* (Limpieza, sumarización, unión, etc...). Cada nodo se representa visualmente en el *panel de flujo*.

De igual forma, Tableau Prep Builder lleva un seguimiento de cada operación para poder comprobar el trabajo y realizar cambios en cualquier punto del *flujo*³.

Un ejemplo se visualiza en la siguiente imagen:

The screenshot displays the Tableau Prep Builder interface. On the left, the 'Connections' pane lists three data sources: 'orders_south_2015.c...' (Text file), 'Orders_East.xlsx' (Microsoft Excel), and 'Orders_Central.csv' (Text file). The 'Tables' pane shows 'Orders_Central'. The main workspace shows a data flow: three input sources feed into three 'Clean' steps (Clean 1, Clean 2, Clean 3), which then feed into a 'Union 1' step, and finally into an 'Output' step. The 'Output' step is highlighted with a blue box. Below the workspace, the 'Output' window is open, showing a table of data. The table has columns: Table Names, File Paths, Sales, Quantity, Profit, Region, State, Row ID, Order ID, Order Date, and Sh. The data is organized into 47 rows, each representing a record from the 'Orders_Central.csv' file.

Table Names	File Paths	Sales	Quantity	Profit	Region	State	Row ID	Order ID	Order Date	Sh
Orders_Central.csv	null	68.81	5	-123.858	Central	Texas	15	US-2016-118983	11/22/2016, 12:00:00 AM	
Orders_Central.csv	null	2.544	3	-3.816	Central	Texas	16	US-2016-118983	11/22/2016, 12:00:00 AM	
Orders_Central.csv	null	665.88	6	13.3176	Central	Wisconsin	17	CA-2015-105893	11/11/2015, 12:00:00 AM	
Orders_Central.csv	null	19.46	7	5.0596	Central	Nebraska	22	CA-2017-137330	12/09/2017, 12:00:00 AM	
Orders_Central.csv	null	60.34	7	15.6884	Central	Nebraska	23	CA-2017-137330	12/09/2017, 12:00:00 AM	
Orders_Central.csv	null	29.472	3	9.9468	Central	Texas	35	CA-2018-107727	10/19/2018, 12:00:00 AM	
Orders_Central.csv	null	1,097.544	7	123.4737	Central	Texas	36	CA-2017-117590	12/08/2017, 12:00:00 AM	
Orders_Central.csv	null	190.92	5	-147.963	Central	Texas	37	CA-2017-117590	12/08/2017, 12:00:00 AM	
Orders_Central.csv	null	113.328	9	35.415	Central	Texas	38	CA-2016-117415	12/27/2016, 12:00:00 AM	
Orders_Central.csv	null	532.3992	3	-46.9764	Central	Texas	39	CA-2016-117415	12/27/2016, 12:00:00 AM	
Orders_Central.csv	null	212.058	3	-15.147	Central	Texas	40	CA-2016-117415	12/27/2016, 12:00:00 AM	
Orders_Central.csv	null	371.168	4	41.7564	Central	Texas	41	CA-2016-117415	12/27/2016, 12:00:00 AM	
Orders_Central.csv	null	147.168	4	16.5564	Central	Illinois	42	CA-2018-120999	09/10/2018, 12:00:00 AM	
Orders_Central.csv	null	45.98	2	19.7714	Central	Minnesota	45	CA-2017-118255	03/11/2017, 12:00:00 AM	
Orders_Central.csv	null	17.46	2	8.2062	Central	Minnesota	46	CA-2017-118255	03/11/2017, 12:00:00 AM	
Orders_Central.csv	null	211.96	4	8.4784	Central	Michigan	47	CA-2015-146703	10/20/2015, 12:00:00 AM	

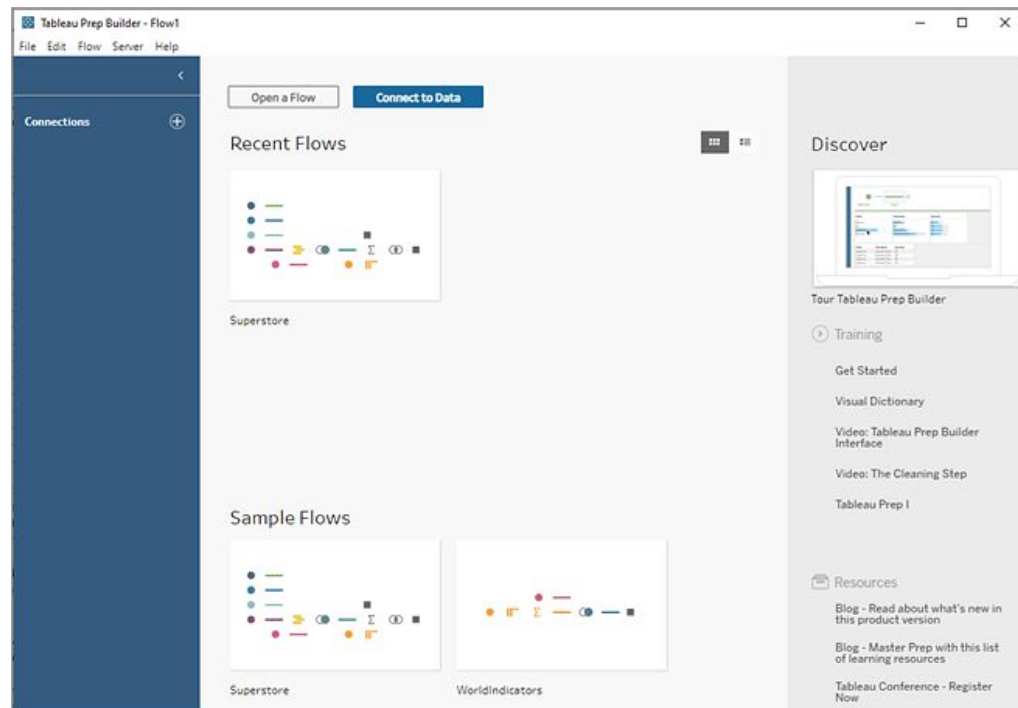
Cuando se termina de construir el *flujo*, Tableau Prep Builder permite ejecutarlo de forma simple para aplicar las operaciones a todos los conjuntos de datos.

DESCRIPCIÓN DE LA HERRAMIENTA

A continuación veremos de forma breve y concisa las características de Tableau Prep Builder usando la documentación oficial de la herramienta.

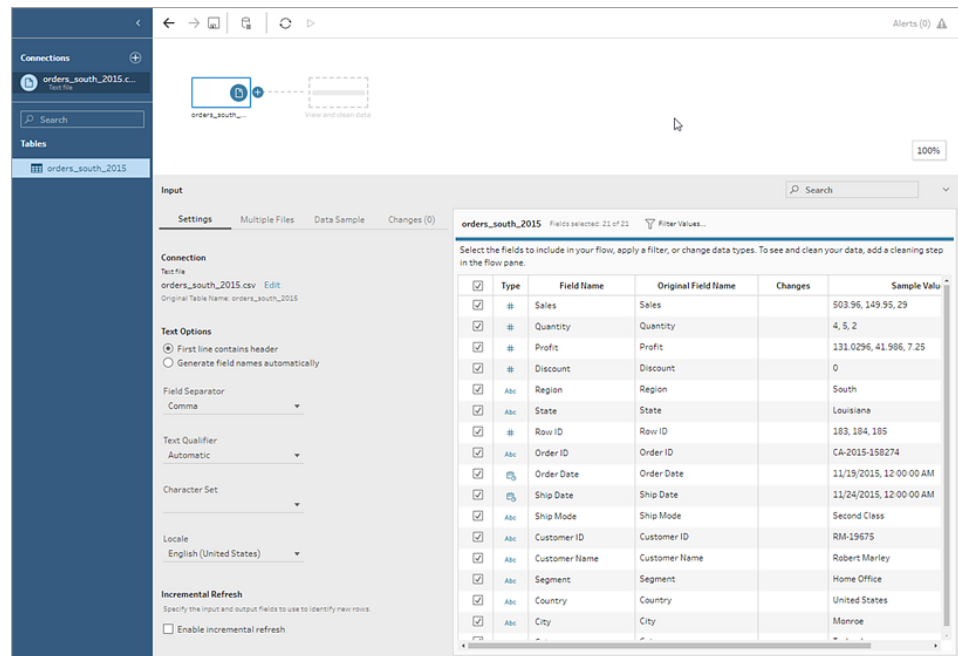
1. Conectarse a los datos

- a. Lo primero que se observa al abrir Tableau Prep Builder es una página de inicio con un panel **Conexiones**.



- b. hacemos clic en el botón Añadir conexión .

- c. El software permite agregar conexiones a datos de distintas fuentes, sean archivos o bases de datos externas. En este caso, buscamos un archivo en formato “.csv”, lo seleccionamos, lo abrimos y ¡listo!



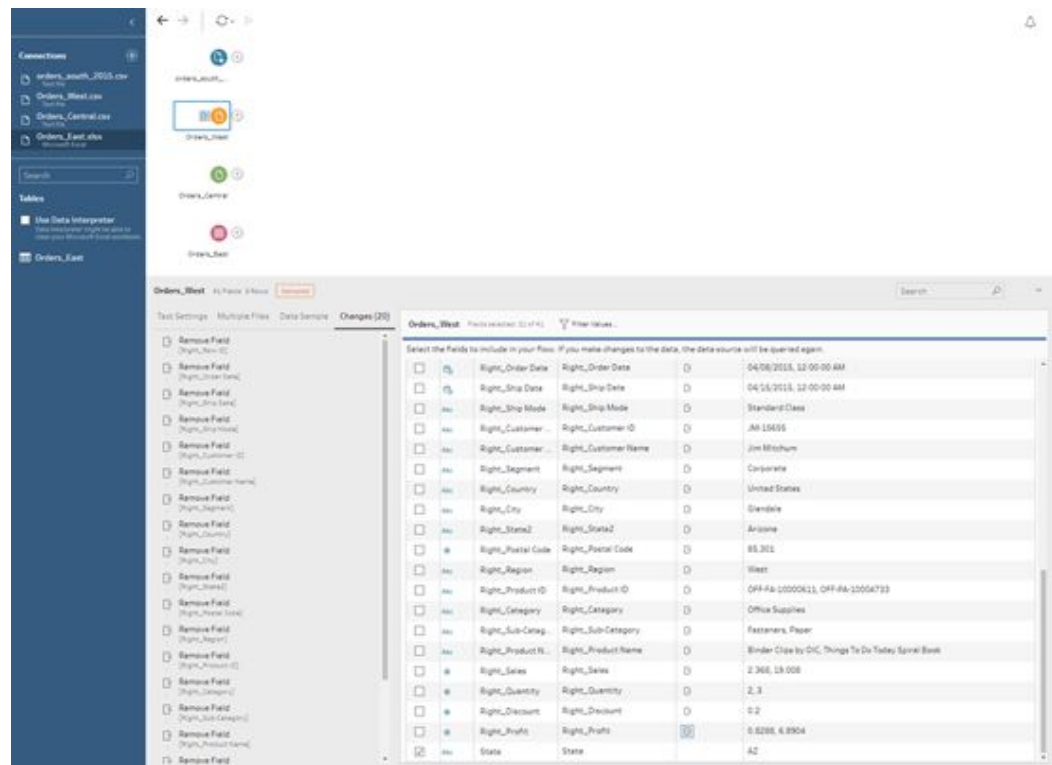
The screenshot shows the Tableau interface with the 'Input' pane open. The 'Input' pane displays a table of fields for the connection 'orders_south_2015.csv'. The table has columns: Type, Field Name, Original Field Name, Changes, and Sample Value. The fields listed are: Sales, Quantity, Profit, Discount, Region, State, Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, and City. The 'Flow Pane' is visible on the right side of the interface.

Type	Field Name	Original Field Name	Changes	Sample Value
✓	Sales	Sales		503.96, 149.95, 29
✓	Quantity	Quantity		4, 5, 2
✓	Profit	Profit		131.0296, 41.986, 7.25
✓	Discount	Discount		0
✓	Region	Region		South
✓	State	State		Louisiana
✓	Row ID	Row ID		183, 184, 185
✓	Order ID	Order ID		CA-2015-158274
✓	Order Date	Order Date		11/19/2015, 12:00:00 AM
✓	Ship Date	Ship Date		11/24/2015, 12:00:00 AM
✓	Ship Mode	Ship Mode		Second Class
✓	Customer ID	Customer ID		RM-19675
✓	Customer Name	Customer Name		Robert Marley
✓	Segment	Segment		Home Office
✓	Country	Country		United States
✓	City	City		Monroe

2. Explorar los datos


Como sugerencia, siempre es buena idea mirar los datos en busca de cualquier posible problema antes de conectarlos al programa.

- Después de conectarlos, se agregan al *panel de flujo* por medio de un paso de *Entrada*.
- Cuando se selecciona un paso de *Entrada* en el *panel de flujo*, aparece la configuración utilizada para introducir los datos, los campos incluidos y una vista previa de los valores.
- Este es un buen momento para decidir la cantidad de datos que desea incluir en el flujo, y eliminar o filtrar los campos no deseados. También se pueden cambiar los tipos de datos asignados incorrectamente.

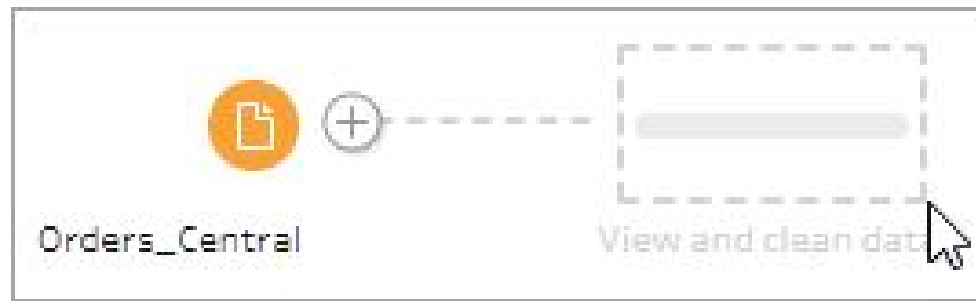


- d. Una vez identificados algunos problemas en los conjuntos de datos, el paso siguiente es examinar los datos más detenidamente, y limpiarlos para combinarlos y darles forma con el fin de generar un archivo de salida que pueda utilizarse en los análisis.

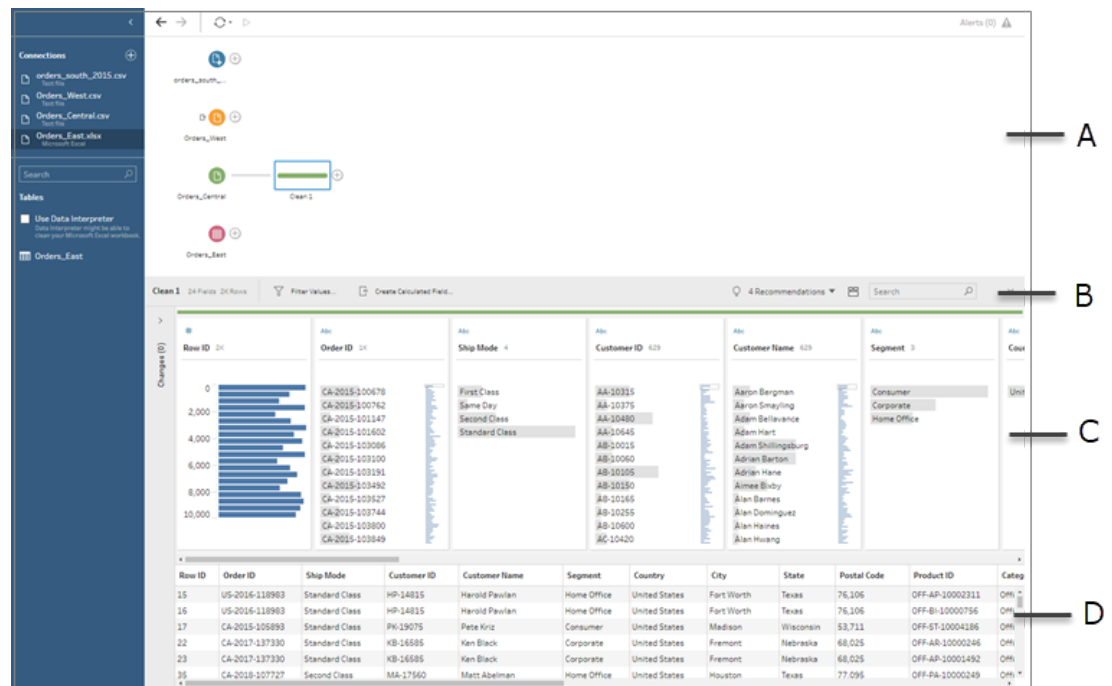
3. Limpiar los datos

- a. En Tableau Prep Builder, examinar y limpiar los datos es un proceso iterativo. Una vez definido el conjunto de datos con el que desea trabajar, el paso siguiente es examinar y tratar los datos mediante varias operaciones para limpiarlos, darles forma y combinarlos.
- b. Haz clic en el icono del signo más  y agrega un paso de limpieza. Dependiendo de la versión, esta opción del menú es **Añadir paso**,

Añadir paso de limpieza o Paso de limpieza.



- c. Cuando se añade un paso de limpieza al flujo, el espacio de trabajo cambia y es posible ver los detalles de los datos.



d.

A. Panel de flujo, B. Barra de herramientas, C. Panel de perfil, D. Cuadrícula de datos

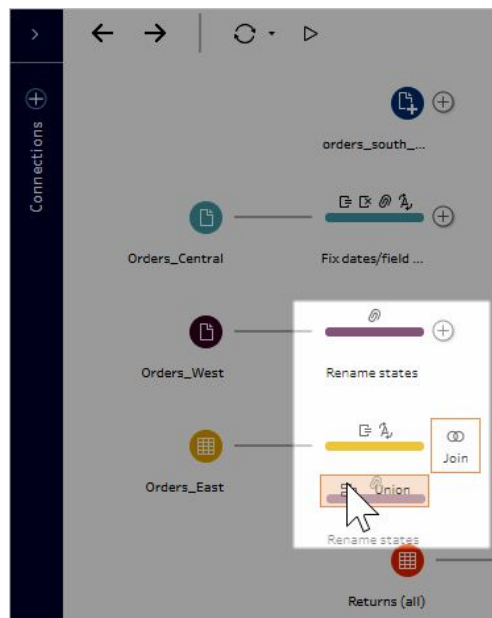
El espacio de trabajo ahora estará dividido en tres partes: el **panel de flujo**, el **panel de perfil** con una barra de herramientas y la cuadrícula **Datos**.

- e. Ahora estás listo para limpiar, examinar e iterar en los datos cargados. Hay múltiples formas de hacerlo y todo dependerá del proyecto y los tipos de datos que tenga.

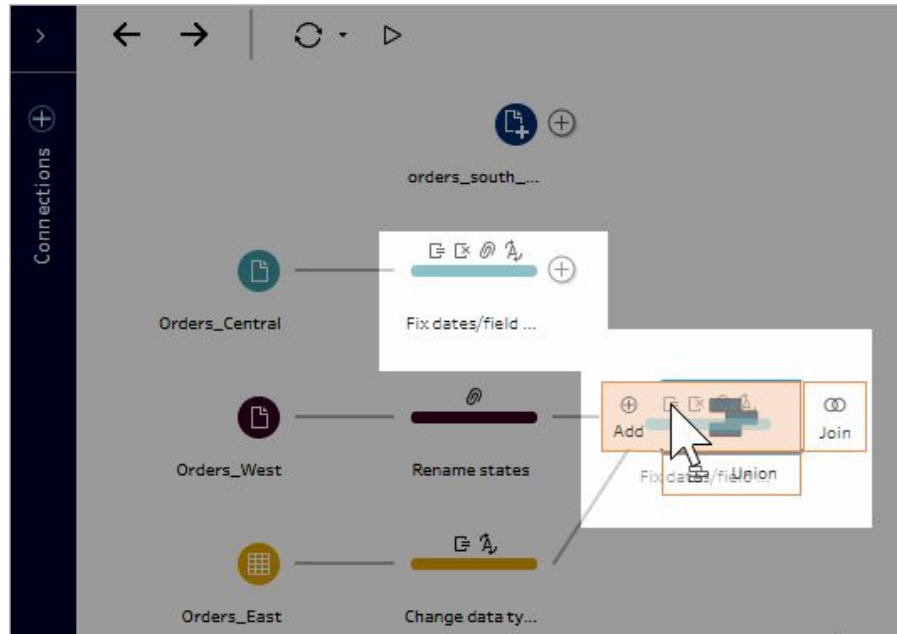
- f. Una vez limpiado un archivo, comprueba el resto de los archivos para detectar otros problemas que deban solucionarse.

4. Combinar los datos

- a. Una vez limpiados los archivos, por fin se pueden combinar.
- b. Debido a que todos los archivos tienen campos similares después de los esfuerzos de limpieza, para juntar todas las filas en una sola tabla, es necesario unirlos.
- c. Podemos unir las tablas principalmente por filas y/o columnas.
- d. En el **panel de flujo**, arrastra los pasos de limpieza que tenga al paso de limpieza final y suéltelo en la opción **Unión de filas**.



- e. Arrastre el siguiente paso de limpieza del flujo al nuevo paso **Unión de filas**, luego suéltelo en **Añadir** para agregarlo a la unión existente.




- f. Ahora, todos los archivos están combinados en una tabla individual. En el *panel de flujo*, seleccione el nuevo paso de **Unión de filas** para ver los resultados.

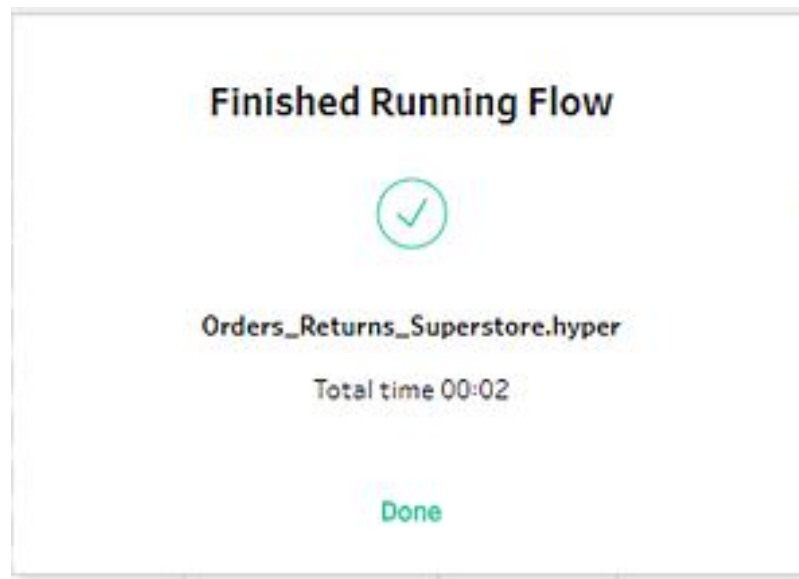
Table Names	File Paths	Product	Discounts	Category	City
Orders_Central.csv	\\...\\	"While you Were Out"	0	Furniture	Aberdeen
Orders_East.xlsx	\\...\\	#10 Gummed Flag (Whi...	0.1	Office Supplies	Ablene
Orders_West.csv	\\...\\	#10 24x36 Business En...	0.2	Technology	Albuquerque
		#10 4 1/8" x 9 1/2" Re...	0.3		Alexandria
		#10 4 1/8" x 9 1/2" Se...	0.32		Allen
		#10 4 1/8" x 9 1/2" P...	0.4		Allentown
		#6 1/4 Gummed Flag...	0.5		Altosona
		1.7 Cubic Foot Comps...	0.6		Amerville
		12 Colored Short Pencils	0.8		Braheum
		12 1/2 Diameter Rou...			Andover
		14 7/8 x 11 Blue Bar Co...			Ann Arbor

Como podrás observar, Tableau ha asociado automáticamente los campos del mismo nombre y tipo.

- g. También puede comprobar que los colores asignados a los pasos del flujo se usan en los perfiles de unión de filas para indicar la procedencia del campo y que también aparecen en la banda de colores en la parte superior de cada campo para indicar si el campo está presente en la tabla.

5. Ejecutar el flujo y generar resultados

- a. Los datos van adquiriendo un buen aspecto, y ya está preparado para generar el archivo de salida y analizarlo en Tableau u otro software.
- b. En el panel izquierdo del menú desplegable **Guardar salida en**, seleccionar **Archivo**, hacer clic en el botón **Examinar** y, en el cuadro de diálogo **Guardar extracción como**, escriba un nombre para el archivo y haga clic en **Aceptar**.
- c. En el campo **Tipo de salida**, seleccione un tipo de salida. Seleccione Extracción de datos de Tableau (.hyper) para Tableau Desktop o Valores separados por comas (.csv) si desea compartir la extracción con un tercero.
- d. En el panel Salida, haga clic en **Ejecutar flujo** o seleccione el botón **Ejecutar flujo**  en el panel del flujo para generar una salida.
- e. Cuando termine de ejecutarse el flujo, aparece un cuadro de diálogo de estado que indica si el flujo se ha ejecutado correctamente y el tiempo empleado para ello. Haga clic en **Listo** para cerrar el cuadro de diálogo.



El propósito de esta breve introducción a las características es mostrar la variedad de herramientas ETL que tiene el software. Omitimos varias que pueden ver y probar dentro de su página web².

VENTAJAS Y DESVENTAJAS DE LA HERRAMIENTA

Ventajas

- **Interfaz intuitiva de arrastrar y soltar**

Esta herramienta destaca por su extraordinaria facilidad de uso. Es lo suficientemente fácil para un usuario de negocios informal y, a la vez, tan poderosa como las herramientas de desarrollo.

- **Destaca en la visualización de los datos**

Su caso de uso más conveniente es el descubrimiento de datos estructurados a través de tablas, gráficos y otros tipos de visualización.

- **Soporte para conexión a más de 30 tipos de fuentes de datos.**

Por ejemplo, sistemas de datos en formatos de archivo (CSV, JSON, XML, MS Excel, etc.), sistemas de datos relacionales y no relacionales (PostgreSQL, MySQL, SQL Server, MongoDB, etc.) o sistemas en la nube (AWS, Oracle Cloud, Google BigQuery, Microsoft Azure). Así como una integración perfecta con muchas plataformas modernas de Big Data.

- **Poderosa colaboración comunitaria**

Tiene capacidad de colaboración en tiempo real que la convierte en una inversión valiosa para organizaciones comerciales y no comerciales por igual.

- **Compartir informes**

Hay varias formas de compartir los informes: publicándose en un servidor público de Tableau, a través del correo electrónico de Tableau Reader o cuando publicamos abiertamente el libro de trabajo de Tableau y damos acceso a cualquier persona que tenga un enlace.

- **Mezcla fuentes de datos.**

Tiene una característica especial de Data Blending donde puede conectarse a diferentes vistas de datos a la vez, crear extracciones, hacer transformaciones, unir y dividir datos, combinar diferentes fuentes de datos, crear grupos y conjuntos, etc.

- **Gran comunidad**

Una amplia biblioteca de materiales de video sobre la herramienta, cursos en línea o blogs de aprendizaje.

Desventajas

- **Capacitación** esencial o consultoría es necesaria para los análisis más complejos.
- Está limitado a sólo los **sistemas operativos** Windows o Mac.
- No está diseñado para **análisis de profundidad** en datos científicos y/o matemáticos en modo de análisis estadístico. (ej.: Minitab, SPSS, MatLab, etc.)
- En el manejo de grandes conjuntos de datos con conexiones en directo a las plataformas el **rendimiento** se puede degradar.
- Se requiere **preparación de datos** inicial
- Las características pueden parecer **demasiado especializadas y restrictivas**
- Aunque es excelente para fines analíticos, no puede reemplazar las aplicaciones de **informes financieros**
- No existe un concepto de **versiones** con el servidor

DESARROLLO DEL EJEMPLO

Se desarrollará un pequeño ejemplo, con el cual será posible demostrar el funcionamiento de este software.

Análisis del problema

El objetivo de este ejemplo es obtener un conjunto de datos con el cual sea posible analizar la *tasa de homicidios dolosos por arma de fuego* en México durante el 2020. Como contexto, la *tasa de homicidios* es un número real cuyo valor indica la razón del número de homicidios en un espacio geográfico por cada 100,000 habitantes, calculado anualmente.

Para obtenerlo, es necesario seguir la siguiente fórmula:

$$Tasa\ de\ homicidios = \frac{Total\ de\ homicidios}{Población\ total} \times 100,000$$

Para este ejemplo en concreto, se calculará la *tasa de homicidios* para los **homicidios dolosos por arma de fuego** por **Entidad (estado)** de México durante el año **2020**.

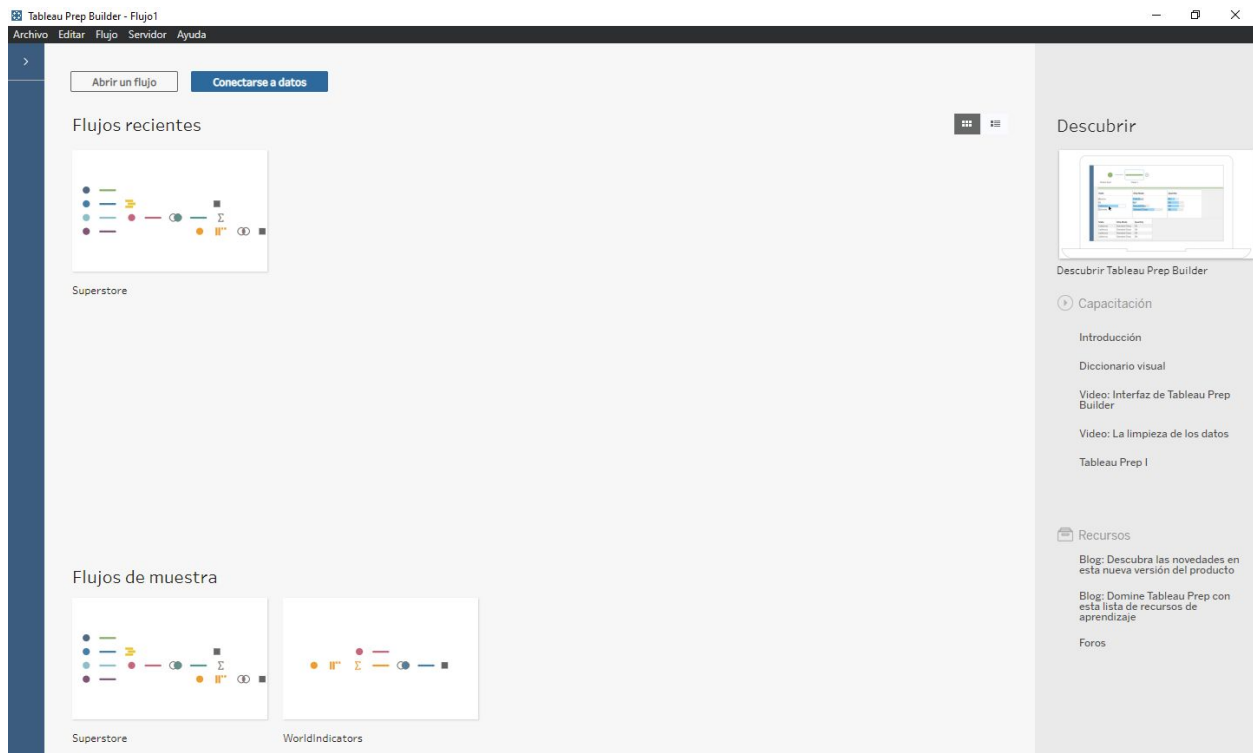
Conjuntos de datos a usar

- *Incidencia delictiva del fuero común por estado*, obtenido del *Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública*:
<https://www.gob.mx/sesnsp/acciones-y-programas/incidencia-delictiva-del-fuero-comun-nueva-metodologia?state=published>
 - Archivo en formato XLSX.
 - Este conjunto de datos proporciona los datos sobre los **homicidios dolosos por arma de fuego**.

- Los datos están desglosados por **Entidad** y por **mes**, para los años **2015-2020**.
- *Censo de Población y Vivienda 2020*, obtenido del *INEGI*:
<https://www.inegi.org.mx/programas/ccpv/2020/default.html#Tabulados>
 - Archivo en formato XLSX.
 - Este conjunto de datos proporciona los datos sobre el **total de población** de cada **Entidad** para el año **2020**.

Desarrollo del ejemplo

La primera vez que se abre el software, la pantalla principal se visualiza como en la siguiente imagen:



El software despliega elementos que denomina *flujos*: Un *flujo reciente* y dos *flujos de muestra*.

Como ya se explicó, el primer paso para crear un *flujo* es conectarse a un conjunto de datos. Para este ejemplo, es necesario conectarse a los 2 archivos XLSX descargados previamente.

Cada conjunto de datos genera un paso de *Entrada* al ser incluido dentro del *flujo*. Con los archivos de Excel, cada hoja dentro de cada archivo generaría una nueva *Entrada* al ser incluida. En este ejemplo, como los 2 archivos de Excel solamente tienen 1 hoja cada uno, solamente se generan 2 *Entradas*.

The screenshot shows the Tableau Prep Builder interface. On the left, the 'Conexiones' (Connections) pane lists two Microsoft Excel files: 'Estatad-Delitos-2015-...' and 'INEGI_Exporta_2021...'. Below this, the 'Tablas' (Tables) pane shows 'Hoja1'. The main workspace displays two data sources: 'Incidencia Delictiva 2020 (1)' and 'INEGI Censo 2020...'. The 'Entrada' (Input) step is selected, and the 'Configuración' (Configuration) tab is active. It shows the 'Incidencia Delictiva 2020 (1)' table with 19 fields. A table below lists the selected fields and their original names, with checkboxes for selection.

	Tipo	Nombre de campo	Nombre de campo original	Cambios	Valores de muestra
<input checked="" type="checkbox"/>	Abc	Año	Año	▼ ^	2020
<input type="checkbox"/>	Abc	Clave_Ent	Clave_Ent	☐ ^	1, 2, 3
<input checked="" type="checkbox"/>	Abc	Entidad	Entidad		Aguaascalientes, Baja California, Baja Cal..
<input type="checkbox"/>	Abc	Bien jurídico afe...	Bien jurídico afectado	☐	La vida y la Integridad corporal

Dentro del paso de *Entrada*, el software permite llevar a cabo ciertas operaciones sobre el conjunto de datos, como pueden ser:

- Renombrar columnas.
- Eliminar columnas.
- Filtrar los valores de cada columna.
- Cambiar el tipo de dato de cada columna.
- Definir si se desea trabajar sobre todo el conjunto de datos, o solamente una muestra.

Además, cada operación se registra en orden cronológico, de forma que es posible visualizar el orden en el que se llevaron a cabo junto con el estado en el que se encontraba el conjunto de datos después de llevar a cabo esa operación. También, permite eliminar alguna operación en específico sin afectar al resto

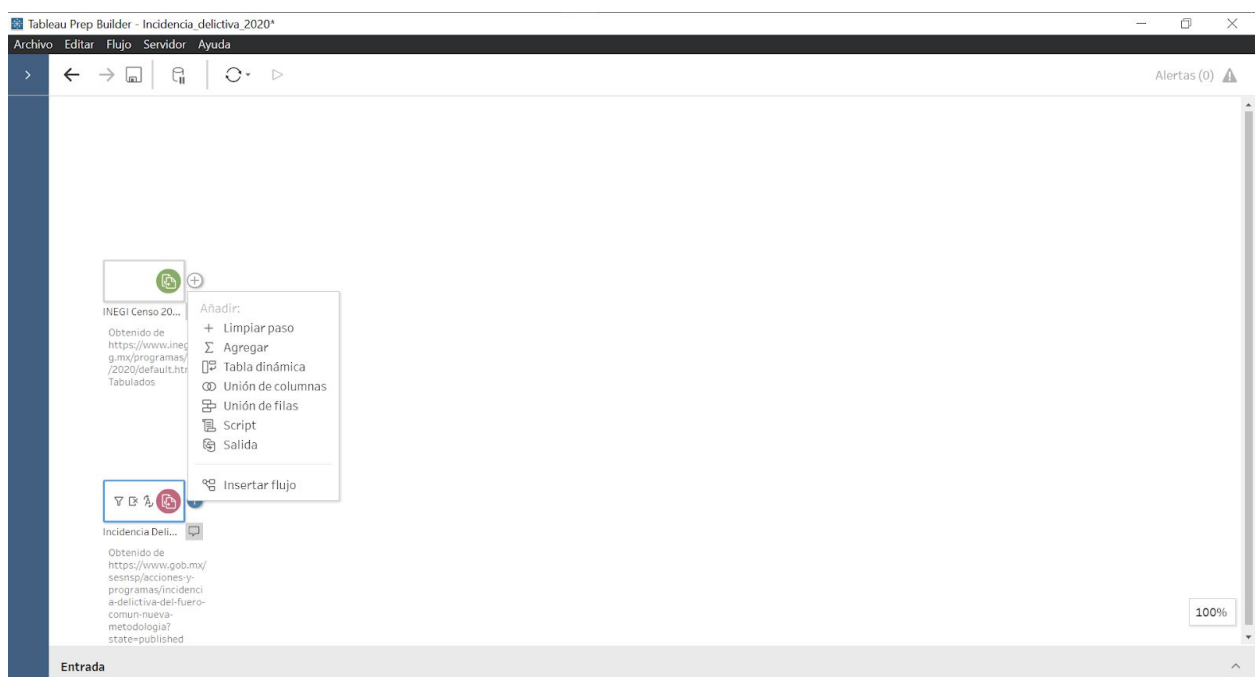
En este ejemplo, la *Entrada* de conjunto de datos de *Incidencia Delictiva* se filtra para quedarse solamente con los **homicidios dolosos** ocasionados **por arma de fuego**.



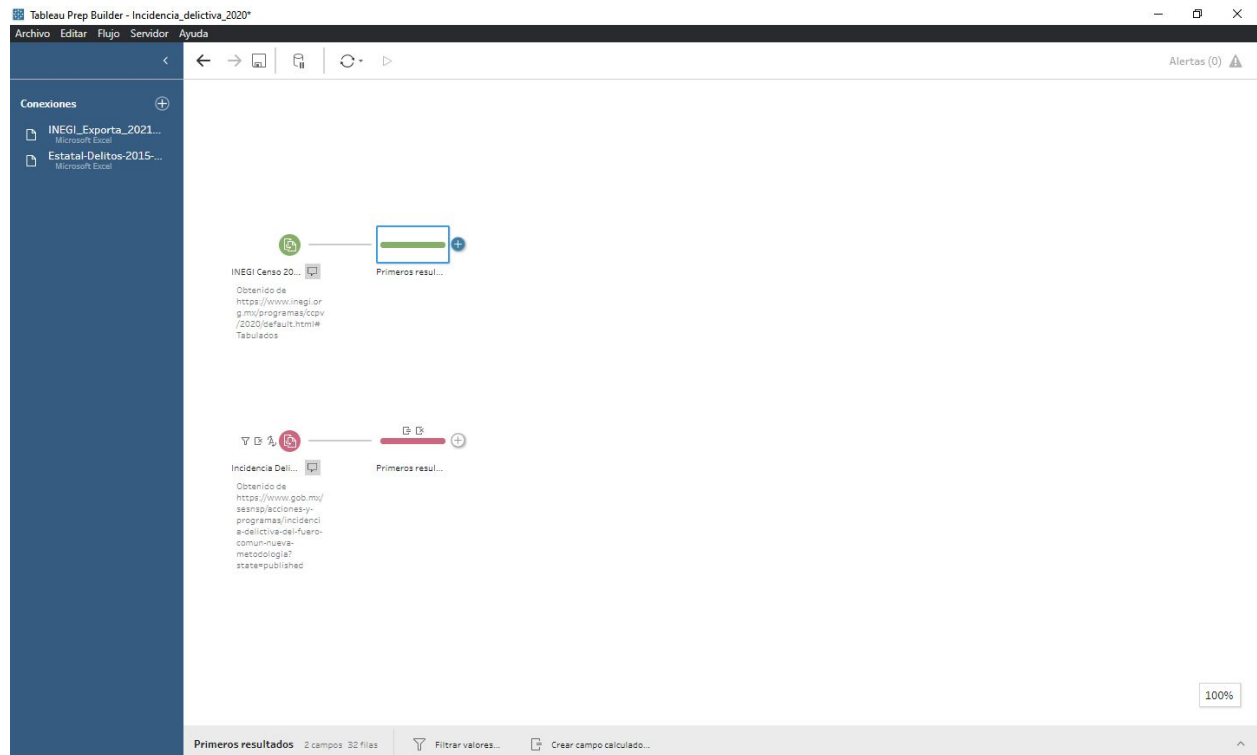
Como comentario extra, Tableau Prep Builder permite nombrar cada paso dentro del *flujo* y agregarle una descripción, para que sea más sencillo recordar o destacar su propósito dentro del flujo.

Para seguir con el proceso de ETL, es necesario agregar un nuevo paso. Como se visualiza en la siguiente imagen, el software permite agregar nuevos pasos para:

- Limpiar los datos.
- Agregar los datos (sumarlos, contarlos, etc...).
- Hacer que el conjunto de datos sea más ancho (con más columnas) o más largo (con más filas).
- Unirlo con otro conjunto de datos.
- Agregar un script de R o Python.
- Crear una *Salida*.
- O unir el *flujo* actual con otro *flujo* ya existente.



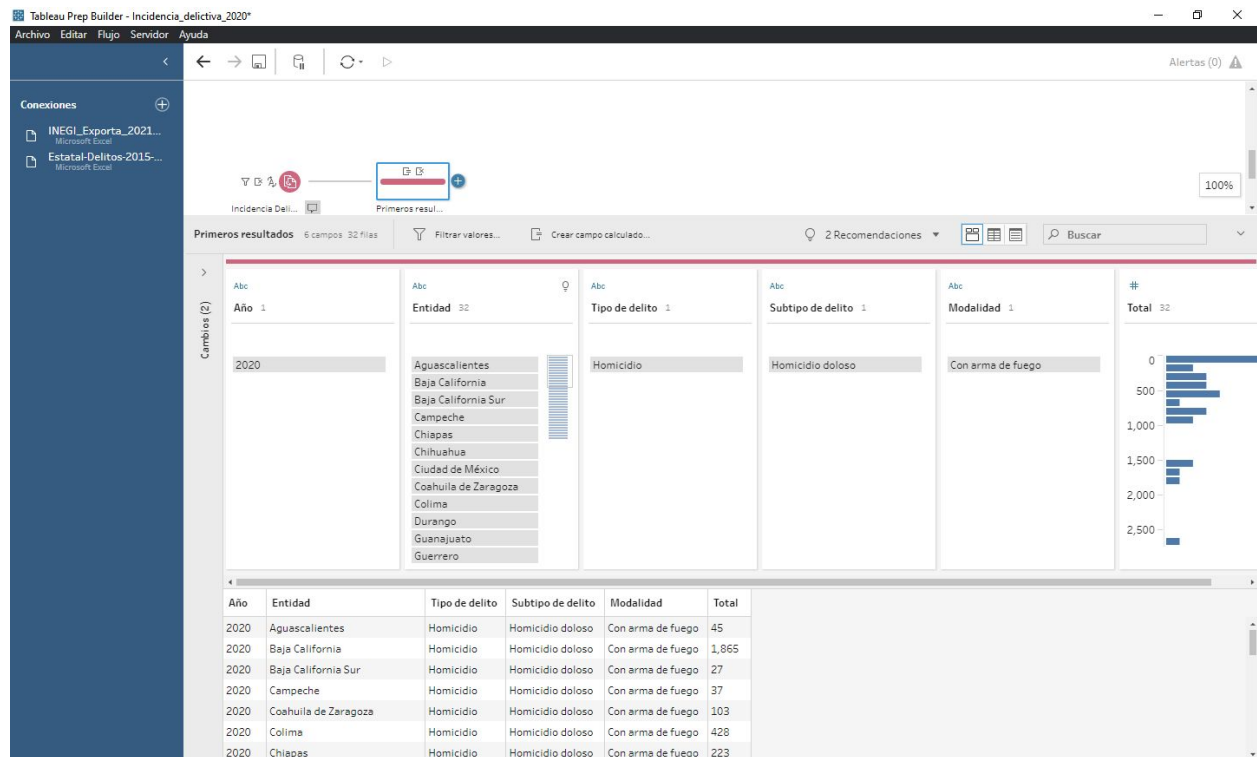
Para este ejemplo, se agrega un paso de limpieza para cada conjunto de datos.



Como su nombre lo indica, en este paso es posible llevar a cabo operaciones para limpiar el conjunto de datos. Algunas de estas operaciones pueden ser:

- Eliminar columnas.
- Filtrar los datos.
- Corregir algún registro o un conjunto de ellos.
- Aplicar algún tipo de operación sobre los distintos tipos de datos.
- Entre otras.

Para ello, Tableau Prep Builder permite explorar de forma sencilla el estado de las distintas columnas a partir de gráficos simples, como se visualiza en la siguiente imagen.



De igual forma, permite crear *campos calculados*, que son nuevas columnas que surgen a partir de aplicar algún tipo de operación sobre las columnas ya existentes.

En este ejemplo, se crea un nuevo *campo calculado* para el conjunto de datos de *Incidencia Delictiva*, el cual contendrá el **total de homicidios por entidad**, sumando las columnas de los meses. De igual forma, como las columnas individuales de los meses no sirven para nada más después de ser sumadas, se eliminan del conjunto de datos.

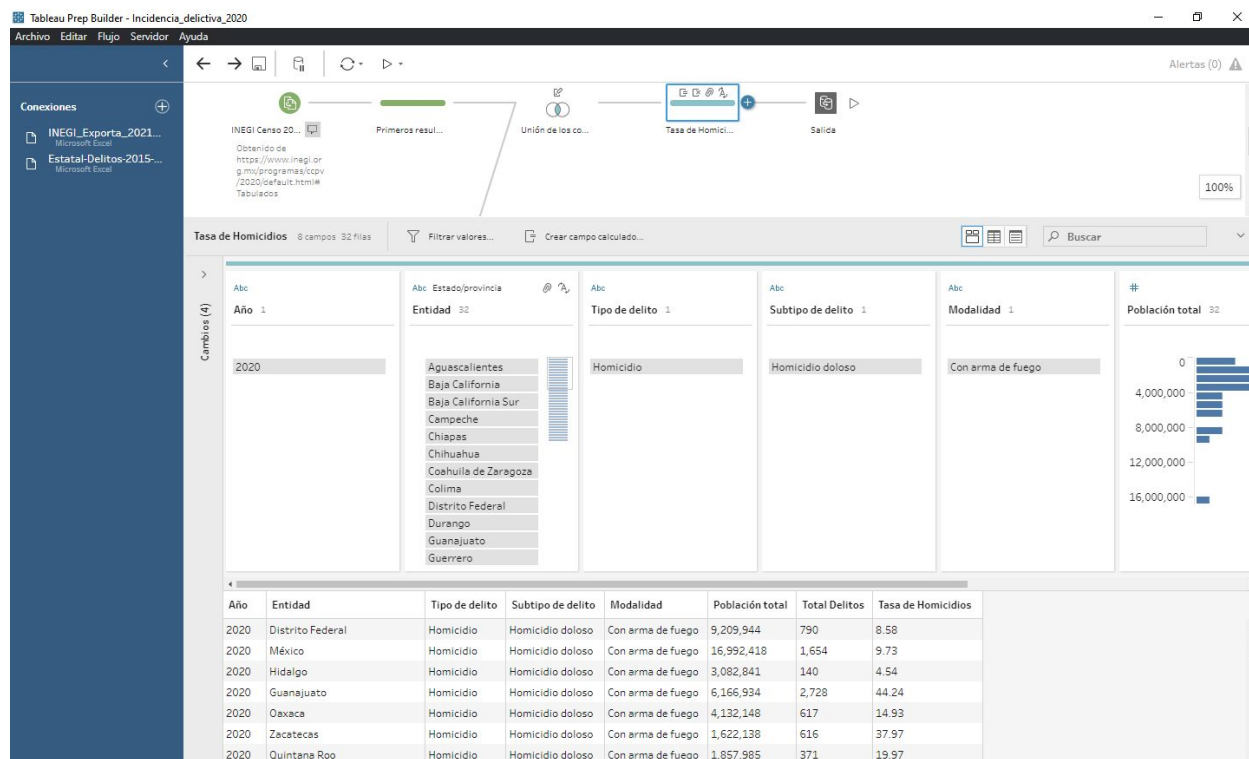
Con ambas fuentes de datos limpias, ahora es necesario unir sus columnas. Esto permitirá tener dentro de una misma tabla todos los datos necesarios para el futuro análisis.

En este ejemplo, ambas fuentes de datos originales se unen tomando como columna de unión la **Entidad**, manteniendo los valores coincidentes dentro de ambas tablas al mismo tiempo (lo que en SQL se conoce como *inner join*).

The screenshot shows the Tableau Prep Builder interface for a workflow titled 'Incidencia delictiva_2020'. The workflow consists of several steps: 'Conexiones', 'INEGI Censo 2010', 'Primeros resultados', 'Unión de los conjuntos de datos', 'Tasa de Homicidios', and 'Salida'. The 'Unión de los conjuntos de datos' step is currently selected, showing a configuration for an inner join on the 'Entidad' field. The 'Cláusulas de unión' section shows 'Entidad' as the join key. The 'Resultados de la unión' section displays a preview of the joined data, including columns for 'Año', 'Entidad', 'Tipo de delito', and 'Subtipo de delito'.

Año	Entidad	Tipo de delito	Subtipo de delito	Mod
2020	Aguaascalientes	Homicidio	Homicidio doloso	Co
2020	Baja California	Homicidio	Homicidio doloso	Co
2020	Baja California Sur	Homicidio	Homicidio doloso	Co
2020	Campeche	Homicidio	Homicidio doloso	Co
2020	Coahuila de Zaragoza	Homicidio	Homicidio doloso	Co

Antes de generar la salida, se incluye un paso de limpieza de datos más, para llevar a cabo algunas modificaciones adicionales, como convertir el nombre de la **Entidad** de cadena a tipo *Entidad/provincia* (tipo de dato espacial disponible en Tableau Prep Builder), o el cálculo de la *tasa de homicidios* a partir de la fórmula mencionada al principio del ejemplo.



Con el conjunto de datos preparado, es momento de exportarlo a partir de un paso de *Salida*.

Tableau Prep Builder permite exportar el conjunto de datos de 3 formas distintas:

- Como un archivo, ya sea un CSV o un archivo *hyper*, que es una extracción de datos para su uso dentro de otro software de Tableau.
- Como una fuente de datos publicada en Tableau Server o Tableau Online.
- Como una tabla dentro de una base de datos externa, como Amazon Redshift o Microsoft SQL Server.

Tableau Prep Builder - Incidencia delictiva_2020

Archivo Editar Flujo Servidor Ayuda

Conexiones

- INEGI_Exporta_2021... Microsoft Excel
- Estatad-Delitos-2015... Microsoft Excel

Flujo

INEGI Censo 20...
Obtenido de
<https://www.inegi.org.mx/programas/ccpv/2020/default.html#Tabulados>

Primeros resul...

Unión de los co...

Tasa de Homici...

Salida

Alertas (0)

100%

Salida 8 campos

Guardar salida en

Archivo

Examinar

Nombre

Ubicación

C:\...Fuentes de datos

Tipo de salida

Extracción de datos de Tableau (hyper)

Escribir opciones

Seleccione una opción para crear o actualizar la tabla de salida.

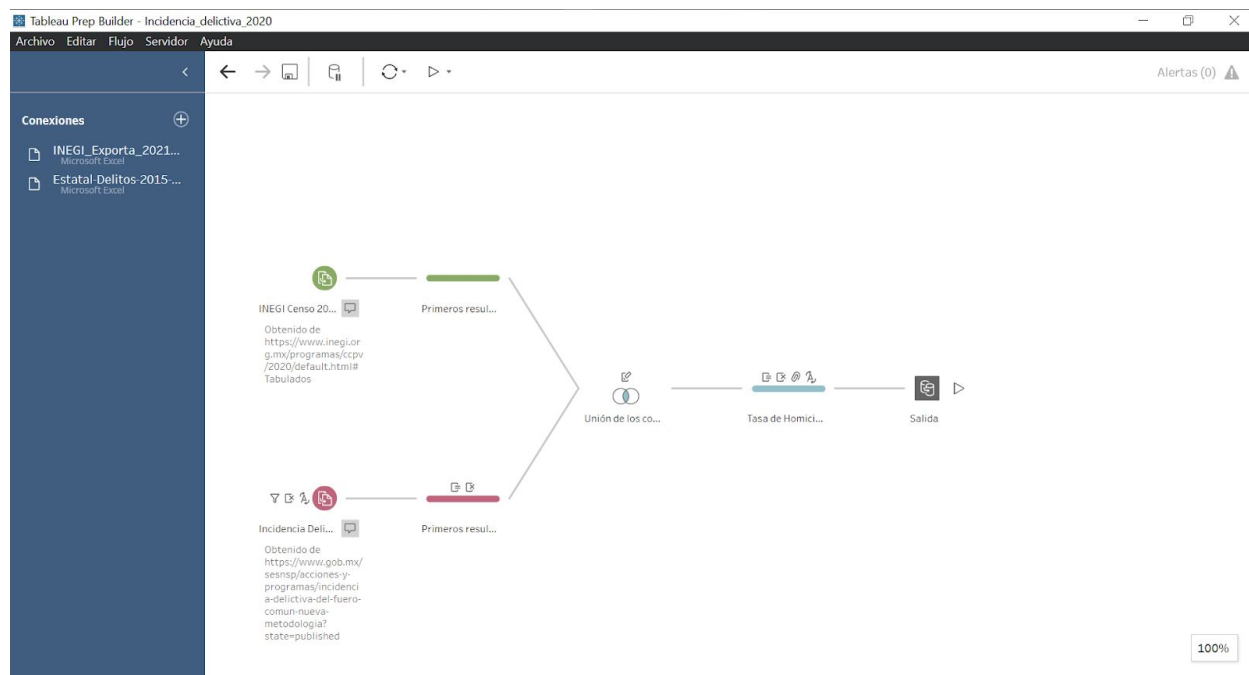
Actualización completa

Crear tabla

Ejecutar flujo

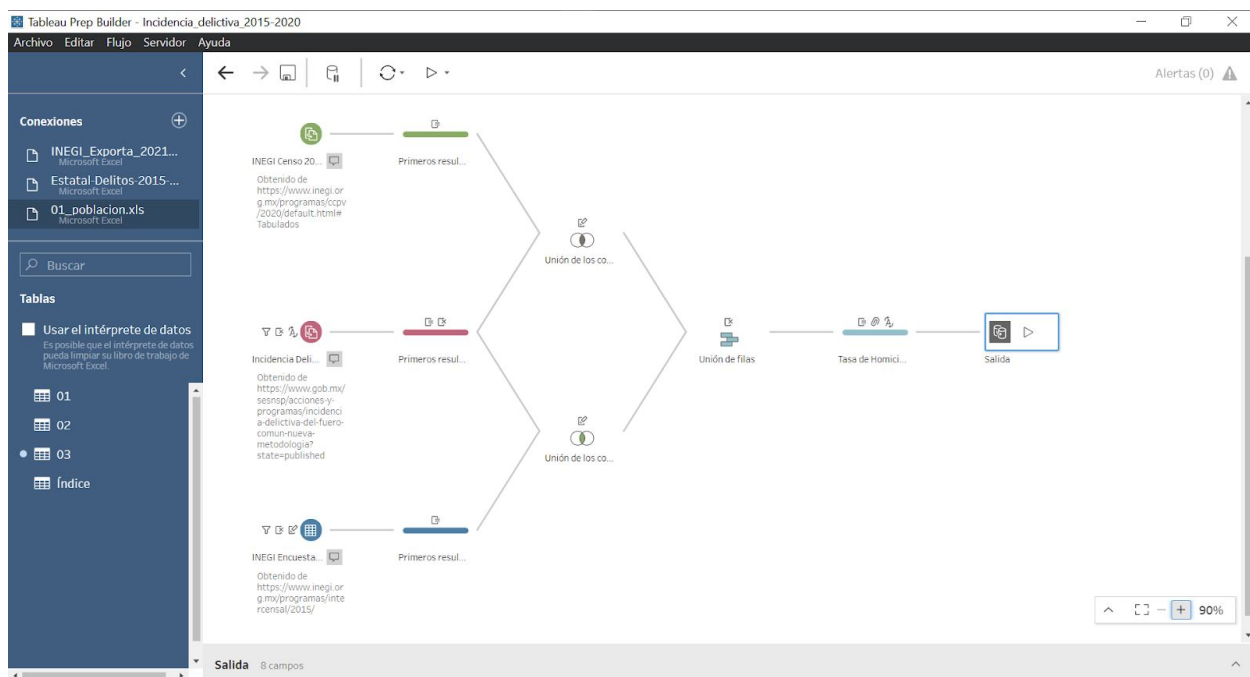
Año	Entidad	Tipo de delito	Subtipo de delito	Modalidad	Población total	Total Delitos	Tasa de Homicidios
2020	Distrito Federal	Homicidio	Homicidio doloso	Con arma de fuego	9,209,944	790	8.58
2020	México	Homicidio	Homicidio doloso	Con arma de fuego	16,992,418	1,654	9.73
2020	Hidalgo	Homicidio	Homicidio doloso	Con arma de fuego	3,082,841	140	4.54
2020	Guanajuato	Homicidio	Homicidio doloso	Con arma de fuego	6,166,934	2,728	44.24
2020	Oaxaca	Homicidio	Homicidio doloso	Con arma de fuego	4,132,148	617	14.93
2020	Zacatecas	Homicidio	Homicidio doloso	Con arma de fuego	1,622,138	616	37.97
2020	Quintana Roo	Homicidio	Homicidio doloso	Con arma de fuego	1,857,985	371	19.97
2020	Veracruz de Ignacio de la Llave	Homicidio	Homicidio doloso	Con arma de fuego	8,062,579	812	10.07
2020	Durango	Homicidio	Homicidio doloso	Con arma de fuego	1,832,650	73	3.98
2020	Chiapas	Homicidio	Homicidio doloso	Con arma de fuego	5,543,828	223	4.02
2020	Tamaulipas	Homicidio	Homicidio doloso	Con arma de fuego	3,527,735	323	9.16
2020	Baja California	Homicidio	Homicidio doloso	Con arma de fuego	3,769,020	1,865	49.48
2020	Nayarit	Homicidio	Homicidio doloso	Con arma de fuego	1,235,456	78	6.31
2020	Tlaxcala	Homicidio	Homicidio doloso	Con arma de fuego	1,342,977	53	3.95
2020	Baja California Sur	Homicidio	Homicidio doloso	Con arma de fuego	798,447	27	3.38
2020	Michoacán de Ocampo	Homicidio	Homicidio doloso	Con arma de fuego	4,748,846	1,583	33.33
2020	Querétaro	Homicidio	Homicidio doloso	Con arma de fuego	2,368,467	99	4.18
2020	Chihuahua	Homicidio	Homicidio doloso	Con arma de fuego	3,741,869	1,561	41.72
2020	Campeche	Homicidio	Homicidio doloso	Con arma de fuego	928,363	37	3.99
2020	Tabasco	Homicidio	Homicidio doloso	Con arma de fuego	2,402,598	357	14.86

El *flujo* completo se visualiza de la siguiente manera:



Para ejecutarlo, simplemente es necesario hacer click en el botón de *Play* en la salida. Así, Tableau Prep Builder ejecutará paso a paso el proceso de ETL completo.

Los *flujos* se vuelven cada vez más complejos si el análisis se compone de más pasos. Sin embargo, y a diferencia del código, visualizar todo el proceso de ETL como si fuera un grafo permite entender con mayor facilidad cada uno de los pasos que se han llevado a cabo para lograrlo. Como ejemplo, se puede comparar la complejidad del ejemplo anterior con un nuevo ejemplo donde se calcula la *tasa de homicidios* con las mismas condiciones, pero ahora calculando el valor para los años **2015-2020**, y tomando el total poblacional de los años **2015-2019** de la *Encuesta Intercensal INEGI 2015*.



CONCLUSIONES

Tableau Prep Builder, como herramienta ETL, ofrece una propuesta bastante intuitiva y poderosa, en comparación con otras herramientas clásicas como los lenguajes de programación. La empresa Tableau se ha esforzado en ofrecer a sus clientes, muchos de ellos alejados del área de la programación, una herramienta para preparar sus datos, la cual sirva como complemento a su software de analítica Tableau Desktop. Y de cara a sus clientes más especializados, también ofrece una herramienta que permite combinar sus conocimientos técnicos con las poderosas características que Tableau Prep Builder ya ofrece de por sí.

Es muy conocido en el mundo de la Ciencia de Datos que la de Limpieza de Datos es el paso más odiado y el que más tiempo consume dentro del proceso de Análisis de Datos. Por ello, Tableau Prep Builder ofrece también a este mercado una propuesta bastante interactiva y sencilla de usar, la cual permita agilizar el proceso de Limpieza de Datos para que sea menos tedioso y más sencillo de realizarlo.

REFERENCIAS

1. Barret, M. (2021). *An Introduction to Directed Acyclic Graphs*. Recuperado el 01 de febrero del 2021. Disponible en: <https://cran.r-project.org/web/packages/ggdag/vignettes/intro-to-dags.html>
2. Tableau. (2020). *Introducción a Tableau Prep Builder*. 31 Enero, 2021, de TABLEAU SOFTWARE LLC. Sitio web: https://help.tableau.com/current/prep/es-es/prep_get_started.htm
3. Tableau. (2020). *Acerca de Tableau Prep*. 31 Enero, 2021, de TABLEAU SOFTWARE LLC. Sitio web: https://help.tableau.com/current/prep/es-es/prep_about.htm#
4. Paradigmadigital.com. (2021). *Plataformas Business Intelligence: principales herramientas de uso*. Recuperado el 02 de febrero del 2021. Disponible en: <https://www.paradigmadigital.com/techbiz/comparativa-plataformas-business-intelligence-principales-herramientas/>