

Manual del usuario de Entity Analytics
para IBM SPSS Modeler 15



Nota: Antes de utilizar esta información y el producto que admite, lea la información general en Avisos el p. 66.

Esta edición se aplica a y a todas las versiones y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

Capturas de pantalla de productos de Adobe reimpresas con permiso de Adobe Systems Incorporated.

Capturas de pantalla de productos de Microsoft reimpresas con permiso de Microsoft Corporation.

Materiales bajo licencia: Propiedad de IBM

Derechos restringidos para los usuarios del gobierno de Estados Unidos: Uso, duplicación o revelación restringidos por GSA ADP Schedule Contract con IBM Corp.

Prefacio

IBM® SPSS® Modeler es el conjunto de programas de minería de datos de IBM Corp. orientado a las empresas. SPSS Modeler ayuda a las organizaciones a mejorar la relación con sus clientes y los ciudadanos a través de la comprensión profunda de los datos. Las organizaciones utilizan la comprensión que les ofrece SPSS Modeler para retener a los clientes más rentables, identificar las oportunidades de venta cruzada, atraer a nuevos clientes, detectar el fraude, reducir el riesgo y mejorar la prestación de servicios del gobierno.

La interfaz visual de SPSS Modeler invita a la pericia empresarial específica de los usuarios, lo que deriva en modelos predictivos más eficaces y la reducción del tiempo necesario para encontrar soluciones. SPSS Modeler ofrece muchas técnicas de modelado tales como pronósticos, clasificaciones, segmentación y algoritmos de detección de asociaciones. Una vez que se crean los modelos, IBM® SPSS® Modeler Solution Publisher permite su distribución en toda la empresa a los encargados de tomar las decisiones o a una base de datos.

Acerca de IBM Business Analytics

El software IBM Business Analytics ofrece información completa, coherente y precisa en la que los órganos de toma de decisiones confían para mejorar el rendimiento comercial. Un conjunto integral de [inteligencia empresarial](#), [análisis predictivo](#), [rendimiento comercial y gestión de estrategias](#), así como de [aplicaciones de análisis](#) le ofrece una información clara, inmediata e interactiva del rendimiento actual y la capacidad para predecir resultados futuros. En combinación con extensas soluciones sectoriales, prácticas probadas y servicios profesionales, las organizaciones de cualquier tamaño pueden conseguir el máximo de productividad y alcanzar mejores resultados.

Como parte de esta familia, el software de análisis predictivo de IBM SPSS ayuda a las organizaciones a predecir eventos futuros y actuar proactivamente según esa información para lograr mejores resultados comerciales. Los clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología de IBM SPSS como ventaja ante la competencia para atraer, retener y hacer crecer los clientes, reduciendo al mismo tiempo el fraude y reduciendo el riesgo. Al incorporar el software de IBM SPSS en sus operaciones diarias, las organizaciones se convierten en empresas predictivas, capaces de dirigir y automatizar decisiones para alcanzar los objetivos comerciales y lograr una ventaja considerable sobre la competencia. Para obtener más información o contactar con un representante, visite <http://www.ibm.com/spss>.

Asistencia técnica

La asistencia técnica está disponible para el mantenimiento de los clientes. Los clientes podrán ponerse en contacto con el servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de IBM Corp. o sobre la instalación en los entornos de hardware admitidos. Para ponerse en contacto con el servicio de asistencia, visite el IBM Corp. sitio Web en <http://www.ibm.com/support>. Prepárese para identificarse, identificar a su organización y su acuerdo de asistencia al solicitar asistencia.

Contenido

1	Análisis de entidades	1
	Acerca del análisis de entidades	1
	Análisis de entidades y análisis predictivo	2
2	Entity Analytics con IBM SPSS Modeler	4
	Uso de Entity Analytics con IBM SPSS Modeler	4
	Fase 1: leer los datos de fuente en SPSS Modeler	5
	Fase 2: crear el repositorio	6
	Fase 3: conectar SPSS Modeler al repositorio	7
	Fase 4: asignar los campos de entrada a funciones del repositorio	7
	Fase 5: exportar datos al repositorio y resolver conflictos	8
	Fase 6: Análisis las identidades resueltas	11
	Fase 7: resolver casos nuevos con el repositorio	11
	Fase 8: generar alertas	13
3	Tareas de análisis de entidades	14
	Acerca de las tareas	14
	Creación de un repositorio de entidad (nodo de exportación de EA)	14
	El repositorio de entidad	14
	Conexión con una fuente de datos	15
	Creación del repositorio	16
	Asignación de campos de entrada a funciones (nodo de exportación de EA)	19
	Visualización de las asignaciones de campos (nodo de exportación de EA)	21
	Configuración de un repositorio de entidad	22
	Visualización de las asignaciones de fuentes de datos	23
	Mantenimiento de las funciones del repositorio	24
	Adición o modificación de una función	27
	Mantenimiento de los tipos de entidad	28
	Establecimiento del umbral para coincidencia de entidades	31
	Reutilización de configuración de repositorio	31
	Guardar sus cambios de configuración	32
	Cerrar la ventana de configuración	32
	Análisis de las identidades resueltas [nodo de fuente Entity Analytics(EA)]	32
	Selección de una fuente de datos	33
	Cambio de nombre de campos de datos	33

Establecimiento de la información de tipos para campos de datos	34
Adición de nodos a la ruta	34
Comparación de nuevos casos con el repositorio (nodo Generación de análisis de entidad)	34
Asignación de campos de entrada a funciones (nodo Generación de análisis de entidad)	36
Visualización de las asignaciones de campos (nodo Generación de análisis de entidad)	37
Visualización de fuentes de datos (nodo Generación de análisis de entidad)	38
Resultados del nodo Generación de análisis de entidad	40
Uso de IBM SPSS Modeler Entity Analytics con otros productos IBM SPSS	40
Tareas administrativas	40
Configuración de asignaciones de puertos	41
Gestión de las credenciales de administrador para la base de datos del repositorio	41
Traslado del directorio de almacenamiento del repositorio.	42
Configuración de opciones de ruta para campos de fecha/hora y marca de tiempo	43
Ejecución de IBM SPSS Modeler Entity Analytics con el cliente de SPSS Modeler y SPSS Modeler Server en el mismo sistema Windows	43
Purga de un repositorio de entidad	44
Eliminación de un repositorio de entidad	44
Eliminación de un repositorio cuando no se puede conectar a él	45

4 El análisis de entidades en acción 47

Acerca de este ejemplo	47
El modelo original	47
Con análisis de entidades	51
Volcado de los datos de fuente en el repositorio	52
Lectura de las identidades resueltas	53
Comparación de los resultados del análisis de entidades con el modelo original.	61
Resumen	64

Apéndice

A Avisos 66

Índice 69

Análisis de entidades

Acerca del análisis de entidades

IBM® SPSS® Modeler Entity Analytics incorpora una dimensión completamente nueva al análisis predictivo de IBM® SPSS® Modeler. Mientras que el análisis predictivo trata de predecir comportamientos futuros a partir de datos del pasado, el análisis de entidades se centra en mejorar la coherencia de los datos actuales mediante la resolución de conflictos de identidades dentro de los propios registros. La identidad de un individuo, una organización, un objeto o cualquier otra entidad puede estar expuesta a ambigüedades. La resolución de identidades puede ser vital en diversos campos, entre los que se incluyen la gestión de la relación con el cliente, la detección de fraudes, la lucha contra el blanqueo de dinero y la seguridad nacional e internacional.

Imagínese que dispone de los siguientes registros de cliente de dos fuentes distintas, y que no está seguro de si hacen referencia a la misma o a distintas personas.

Fuente 1

N.º de registro: 70001
Nombre: Jon Smith
Dirección: 123 Main Street
Número de identificación fiscal: 555-00-1111
Permiso de conducir: 0001133107
Tarjeta de crédito: 10229127

Fuente 2

N.º de registro: 9103
Nombre: JOHNATHAN Smith
Fecha de nacimiento: 06/17/1934
Teléfono: 555-1212
Tarjeta de crédito: 10229128
Correo electrónico: jls@mail.com
Dirección IP: 9.50.18.77

No hay ninguna coincidencia exacta en los datos de los dos registros. Sin embargo, si introducimos una tercera fuente, encontramos algunos atributos en común.

Fuente 3

N.º de registro: 6251
Nombre: Jonathan Smith
Teléfono: 555-1212
Permiso de conducir: 0001133107
Tarjeta de crédito: 10229132

El número del permiso de conducir une los registros de las fuentes 1 y 3, mientras que el número de teléfono une las fuentes 2 y 3. Por tanto, podemos estar bastante seguros de que las tres fuentes se refieren a la misma persona.

Pero, ¿qué sucede si la distinción no es tan sencilla? Podría haber muy poca información sobre la que basar la decisión. Considere los dos registros siguientes.

Fuente 4

N.º de registro: S45286
Nombre: John T Smith Jr
Dirección: 456 Main Street
Teléfono: 703-555-2000
Fecha de nacimiento: 03/12/1984

N.º de registro: S45287
Nombre: John T Smith
Dirección: 456 Main Street
Teléfono: 703-555-2000
Permiso de conducir: 009900991

Evidentemente, este Sr. Smith no es el mismo de los registros anteriores: las diferencias son suficientes para descartarlo. Sin embargo, sigue habiendo un problema. Dos registros diferentes, de la misma fuente de datos, hacen referencia aparentemente a la misma persona. ¿Son registros duplicados? No podemos estar seguros a menos que encontremos otro registro relacionado que nos proporcione más información, quizás de una fuente distinta.

Fuente 5

N.º de registro: 769582-2
Nombre: John T Smith Sr
Dirección: 456 Main Street
Teléfono: 703-555-2000
Permiso de conducir: 009900991
Fecha de nacimiento: 06/25/1959

Esto resuelve el problema. Los dos registros de la fuente 4 no son duplicados; en realidad, son padre e hijo con el mismo nombre, que viven en la misma dirección y que tienen el mismo número de teléfono. Mediante un sistema manual, llevaría semanas encontrar el registro que resuelve el problema de identidad. Con un sistema de análisis de entidades automatizado, el tiempo de resolución se reduce considerablemente.

Análisis de entidades y análisis predictivo

Si todos sus datos procediesen de una única fuente de registros completa y sin ambigüedades, IBM® SPSS® Modeler lo tendría bastante fácil a la hora de resolver conflictos de identidad. Solo con el análisis predictivo, podría leer sus datos en SPSS Modeler, realizar sus procesos y obtener resultados fiables.

Sin embargo, las situaciones reales son bastante distintas. Normalmente, los datos no están completos, con frecuencia son ambiguos y están repartidos por diferentes fuentes de datos que registran muchos atributos distintos con campos que se solapan. Parte del valor del análisis de entidades reside en que reúne datos de fuentes distintas en una única área central de almacenamiento, conocida como **repositorio**. El sistema de análisis de entidades examina a continuación los datos minuciosamente para resolver conflictos, y añade un identificador único a los registros que provienen de la misma persona u organización.

La siguiente tabla ilustra las diferencias entre los dos tipos de análisis.

Tabla 1-1
Diferencias entre análisis predictivo y de entidades

Característica	Análisis predictivo	Análisis de entidades
Entrenamiento	Basado en conjuntos y rangos numéricos relativamente pequeños	Basado en conjuntos grandes: nombres, direcciones
Datos utilizados para el entrenamiento	Subconjunto (partición de entrenamiento)	Todos los datos
Generalización	Un algoritmo generaliza todos los datos de entrenamiento para generar un modelo conciso	Los datos permanecen en estructuras disponibles para la coincidencia de entidades y la detección de relaciones
Detección de fraude	Los registros se marcan como potencialmente fraudulentos si tienen características típicas de solicitudes fraudulentas	Los registros se marcan como potencialmente fraudulentos si guardan relación con registros fraudulentos o si proceden del mismo sujeto pero con identidades distintas

Entity Analytics con IBM SPSS Modeler

Uso de Entity Analytics con IBM SPSS Modeler

Es consciente de que puede tener problemas de identidad con sus datos. ¿Cómo puede ayudarle IBM® SPSS® Modeler Entity Analytics a solucionarlo? A continuación se sugiere un procedimiento, aunque es posible que tenga que hacer variaciones para ajustarlo a sus necesidades.

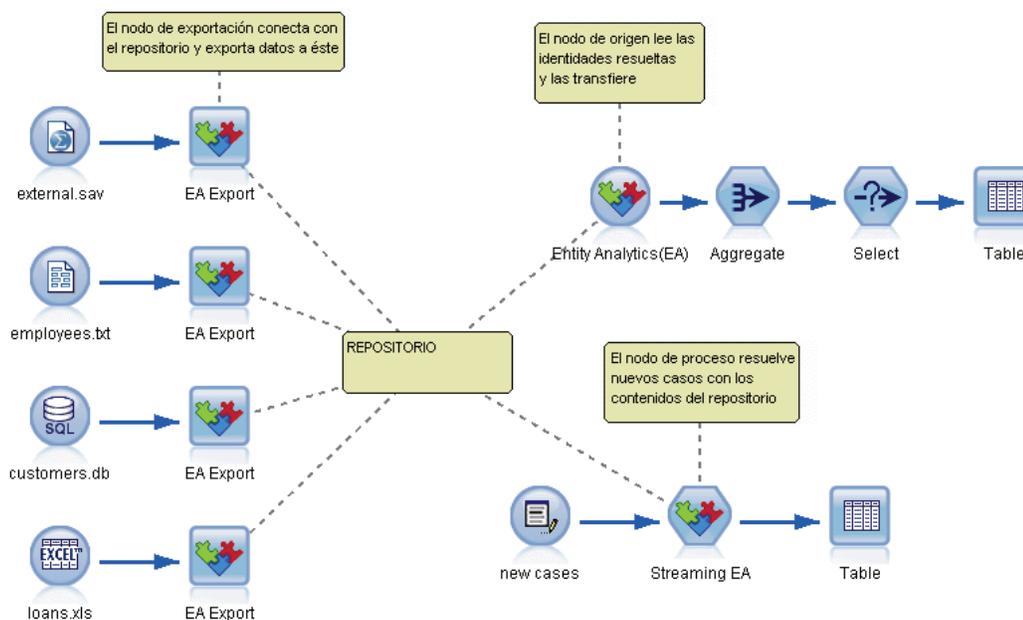
- leer los datos de fuente en IBM® SPSS® Modeler
- Cree un repositorio preparado para almacenar los datos
- conectar SPSS Modeler al repositorio
- Asigne los campos de datos a funciones del repositorio
- Exporte los datos al repositorio y resuelva las identidades
- Analice las identidades resueltas
- resolver casos nuevos con el repositorio
- Genere las alertas necesarias (lote o tiempo real)

En este punto, debe conocer algo del funcionamiento de SPSS Modeler. SPSS Modeler es una herramienta muy intuitiva, que se basa en la representación gráfica de una ruta de datos que fluyen a través de un número de nodos. Cada nodo representa una fase particular del flujo de trabajo.

SPSS Modeler ofrece una gran variedad de nodos, que cubren todas las funciones de minería de datos convencionales. IBM SPSS Modeler Entity Analytics añade nodos para utilizarlos específicamente en análisis de entidades. Estos son el nodo de exportación de EA, el nodo de fuente Entity Analytics(EA) y el nodo de proceso Generación de análisis de entidad.

La figura a continuación ilustra el proceso.

Figura 2-1
El proceso de análisis de entidades



Fase 1: leer los datos de fuente en SPSS Modeler

Su primera tarea es leer sus datos en SPSS Modeler a través de un nodo de fuente, representado en SPSS Modeler por un icono redondo.

Figura 2-2
nodo de fuente IBM SPSS Modeler



Los datos pueden estar en cualquier formato compatible con SPSS Modeler, como archivos de texto, tablas de bases de datos, hojas de cálculo, archivos XML, etc., pero cada formato requiere un nodo de fuente de SPSS Modeler correspondiente. La ilustración muestra un nodo de fuente Base de datos.

Cada archivo de fuente de datos debe tener un campo que identifique exclusivamente cada registro. Si una fuente de datos no tiene dicho campo, puede añadir uno fácilmente en SPSS Modeler. Si desea obtener más información, consulte el tema [Adición de un identificador de registro único](#) en el capítulo 3 el p. 15.

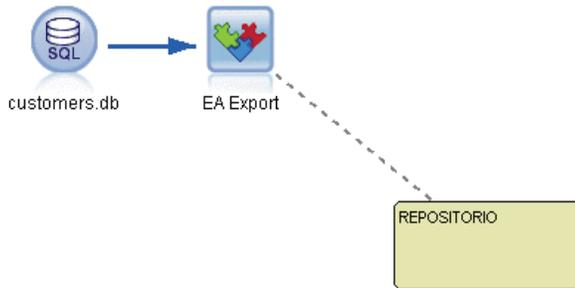
Si desea obtener más información, consulte el tema [Conexión con una fuente de datos](#) en el capítulo 3 el p. 15.

Fase 2: crear el repositorio

El punto central de sus esfuerzos de análisis de entidades es el repositorio, el área de almacenamiento central donde reúne todos sus registros de datos.

Para crear un repositorio, conecte la fuente de datos a un nodo de exportación de EA, representado por el icono cuadrado.

Figura 2-3
Conexión a un repositorio



Desde el nodo de exportación, puede crear un repositorio nuevo (o seleccionar uno existente) que esté preparado para recibir los datos exportados.

Figura 2-4
Creación de un repositorio



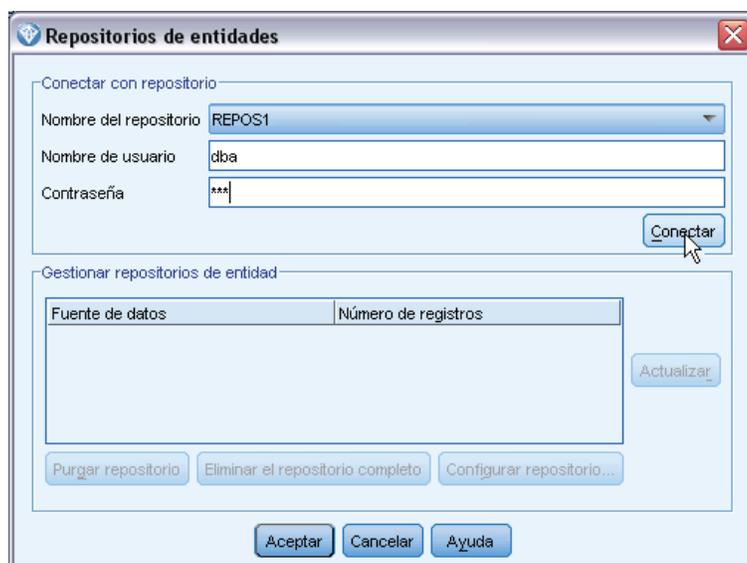
El proceso de creación de un repositorio se describe con detalle más adelante. Si desea obtener más información, consulte el tema [Creación de un repositorio de entidad \(nodo de exportación de EA\)](#) en el capítulo 3 el p. 14.

Una vez creado el repositorio, puede mantener su contenido de diversas formas. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) en el capítulo 3 el p. 22.

Fase 3: conectar SPSS Modeler al repositorio

Una vez creado el repositorio, conéctelo a la ruta de SPSS Modeler.

Figura 2-5
Conexión con el repositorio



Si desea obtener más información, consulte el tema [Opciones de repositorio de entidad](#) en el capítulo 3 el p. 18.

Fase 4: asignar los campos de entrada a funciones del repositorio

Las fuentes de datos pueden contener diversos tipos de información de la entidad. Ciertos tipos de información son comunes para la mayoría de fuentes de datos, mientras que otros son específicos de una fuente de datos particular. En un repositorio de entidad, estos diferentes tipos de información se llaman **funciones**. El repositorio ofrece un número de funciones estándar, pero también puede crear las suyas propias.

Una función de repositorio es un tipo de información individual que se puede utilizar con una fuente de datos de entidad. Algunas funciones (por ejemplo, Nombre, Apellidos, Fecha de nacimiento, etc.) se pueden utilizar con muchas fuentes de datos diferentes, mientras que otras funciones son específicas de una fuente de datos concreta. En términos generales, una función equivale a un campo de un registro de datos o a una columna de una tabla de base de datos.

Una vez creado el repositorio y conectado a él, designe un campo de sus datos de entrada como el campo de **clave única**, que se utilizará en los análisis subsiguientes. Asigne también los campos de datos de entrada a sus funciones correspondientes en el repositorio. Esto se hace para evitar duplicidades en el repositorio en casos en los que fuentes de datos distintas utilizan nombres distintos para campos que contienen el mismo tipo de información. El nodo de exportación de EA dispone de una tabla de asignaciones donde puede crear las asignaciones.

Figura 2-6
Asignación de campos de entrada a funciones

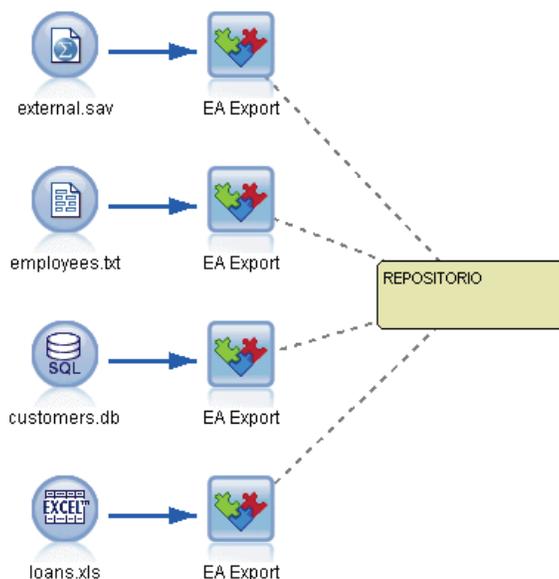
Campo	Vinculado a función	Uso
income		(Auto)
debtinc		(Auto)
creddebt		(Auto)
othdebt		(Auto)
default		(Auto)
preddef1		(Auto)
preddef2		(Auto)
preddef3		(Auto)
ID		(Auto)

Si desea obtener más información, consulte el tema [Asignación de campos de entrada a funciones \(nodo de exportación de EA\)](#) en el capítulo 3 el p. 19.

Fase 5: exportar datos al repositorio y resolver conflictos

Cada nodo de fuente de datos necesita su propio nodo de exportación de EA, de modo que si sus datos están dispersos en diferentes fuentes, puede que tenga una ruta como ésta.

Figura 2-7
Exportar datos al repositorio desde múltiples fuentes de datos



Puede elegir entre leer los registros desde una, algunas o todas las fuentes de datos si tiene más de una. El sistema Entity Analytics analiza los registros seleccionados y añade un campo de identificador denominado \$EA_ID a cada uno. Donde dos o más registros relacionados con identidades previamente ambiguas se pueden resolver, los identificadores añadidos a esos registros son únicos en todo el repositorio. El sistema también añade un campo para mostrar la fuente de datos desde la que se origina el registro.

Conecte cada nodo de fuente de datos a su propio nodo de exportación de EA, asigne los campos de entrada a las funciones de repositorio y, a continuación, ejecute la ruta para exportar datos desde SPSS Modeler al repositorio para resolver conflictos de identidades, en una sola operación. A modo de ilustración, imagine que tiene los siguientes registros en cuatro fuentes de datos distintas.

Datos externos

Nombre	Teléfono	Riesgo de crédito
Mike	555-1234	560
Joe	555-4567	780

Empleados

Nombre	Dirección	Teléfono
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

Cientes

Nombre	Dirección	Ahorros
Susan	1234 5th Street	\$1234
Joe	777 Oak Street	\$5

Préstamos

Nombre	Dirección	Teléfono	Préstamo
Sue	1234 5th Street	555-1234	\$10,000
Joseph	777 Oak Street	555-4567	\$50,000

Como hemos podido ver, usted exporta cada fuente de datos en el repositorio por turnos. Mientras lo hace, el repositorio actualiza la resolución de cada registro. En el repositorio, cada registro va precedido de un campo de identificador (llamado *SEA-ID*) y un campo de indicador de fuente (llamado *SEA-SRC*), que muestra la fuente de datos donde se originó el registro. Así que en nuestro ejemplo, una vez exportadas las cuatro fuentes de datos, el contenido del repositorio tiene este aspecto.

Tabla 2-1

Ejemplo de contenido del repositorio tras la fase de exportación

SEA-ID	SEA-SRC	Nombre	Teléfono	Dirección	Riesgo de crédito	Ahorros	Préstamo
1	Empleados	Michael	555-1234	1234 5th St			
1	External	Mike	555-1234		560		
2	Cientes	Joe		777 Oak St		\$5	
2	External	Joe	555-4567		780		
2	Préstamos	Joseph	555-4567	777 Oak St			\$50,000
3	Empleados	Fred	555-9876	543 1st Ave			
4	Cientes	Susan		1234 5th St		\$1234	
4	Préstamos	Sue	555-1234	1234 5th St			\$10,000

El sistema Entity Analytics ha determinado que *Mike* en el conjunto de datos *Externos* es la misma persona que *Michael* en el conjunto de datos *Empleados* basándose en el número de teléfono común, y le ha asignado el ID 1.

El caso de *Joe* en el conjunto de datos *Externos* es un poco más complicado. ¿Es la misma persona que *Joe* en *Cientes*? No se puede saber solo con esas dos fuentes de datos, pero hay una tercera fuente, *Préstamos*, en la que hay un *Joseph*. Ahora hay una coincidencia: El número de teléfono de *Joseph* es el mismo que el de *Joe* en el conjunto de datos *Externos*. Teniendo esto en cuenta, el sistema determina que son la misma persona y les otorga el identificador 2.

No hay registros múltiples para *Fred*, por lo que recibe el ID 3. *Susan* de *Cientes* es identificada como la misma persona que *Sue* de *Préstamos* puesto que tienen la misma dirección, así que recibe el ID 4.

Nota: este es un ejemplo de una coincidencia optimista con fines ilustrativos. Podría elegir un conjunto de reglas más pesimista, de modo que un simple nombre y un número de teléfono o una dirección por sí mismos no constituirían una coincidencia exacta y no se asignaría a ambos registros el mismo identificador.

Fase 6: Análisis las identidades resueltas

Una vez resueltos los conflictos de identidad en el repositorio, es posible continuar analizando y procesando los resultados. Por ejemplo, si la existencia de registros duplicados para la misma identidad le hace sospechar de posibles actividades fraudulentas, puede que quiera generar un informe con todas las duplicidades.

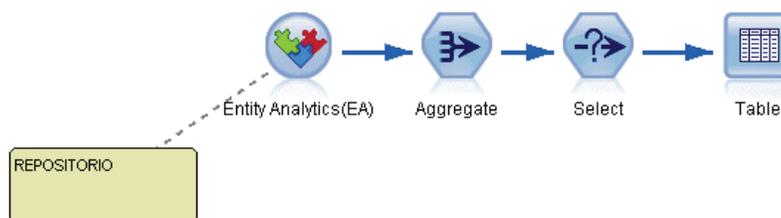
Lo primero es crear un nodo de fuente Entity Analytics(EA) y enlazarlo con el repositorio.

El resultado del nodo consta de los siguientes campos.

- El campo de identificador añadido por el sistema (*\$EA-ID* en el ejemplo de la fase 5)
- El campo de indicador de fuente añadido por el sistema (*\$EA-SRC* en el ejemplo de la fase 5)
- El campo de clave única designado en la fase 4

Para ver el resultado en SPSS Modeler, puede adjuntar un nodo de resultado de SPSS Modeler, como un nodo Tabla o un nodo Informe, y ejecutar esta parte de la ruta. Si necesita resumir el resultado, que podría ser muy extenso, puede incluir nodos de operaciones con registros, como Agregar o Seleccionar, como se muestra en la siguiente ilustración.

Figura 2-8
Ver el resultado



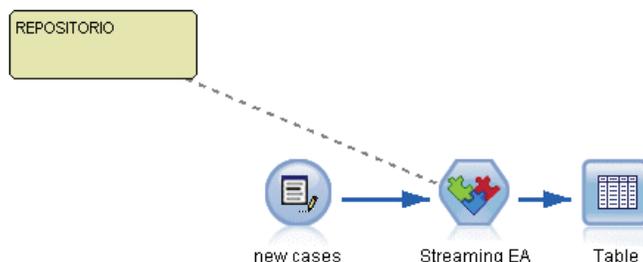
El nodo de fuente Entity Analytics(EA) se describe con detalle más adelante. Si desea obtener más información, consulte el tema [Análisis de las identidades resueltas \[nodo de fuente Entity Analytics\(EA\)\]](#) en el capítulo 3 el p. 32.

Fase 7: resolver casos nuevos con el repositorio

Ha resuelto los conflictos de identidad en todos los registros de sus fuentes de datos. ¿Pero qué sucede cuando quiere comparar un conjunto de nuevos registros para comprobar cómo se relacionan con lo que ya conoce y así obtener una mejor puntuación? Aquí entra en juego el nodo Generación de análisis de entidad.

En primer lugar, añada un nuevo nodo de fuente de datos de SPSS Modeler para leer los nuevos datos en la ruta. A continuación, conecte este nodo de fuente a un nodo Generación de análisis de entidad. Para ver el resultado, añada un nodo Tabla como antes, de forma que esta parte de la ruta tenga el siguiente aspecto.

Figura 2-9
Resolución de nuevos casos



Al ejecutar esta parte de la ruta, el nodo Generación de análisis de entidad lee cada registro nuevo y lo compara con el contenido del repositorio. Si encuentra registros coincidentes en el repositorio, el nodo Generación de análisis de entidad genera resultados de todos los registros coincidentes junto con el registro nuevo, al que añade los campos de ID e indicador de fuente. Si no se encuentra ninguna coincidencia, el nodo de proceso genera únicamente el registro nuevo, con los campos de ID e indicador de fuente añadidos.

Para ilustrar esto, imagine que el repositorio consta actualmente del contenido generado por el nodo de fuente Entity Analytics(EA). Consulte [Tabla 2-1](#) el p. 10.

A continuación, se reciben los siguientes registros nuevos. ¿Hacen referencia a alguien que ya conozcamos?

Tabla 2-2
Nuevos registros para almacenar

Nombre	Dirección	Teléfono	Préstamo
Suzan	1234 5th Street	555-1234	\$100,000
Mark	888 9th Ave	555-9999	\$60,000

Al comparar los nuevos datos con el contenido existente en el repositorio, el nodo Generación de análisis de entidad empareja el primer registro nuevo con la persona que tiene el identificador 4 en los registros existentes. Sin embargo, no hay coincidencias para el segundo registro nuevo, por lo que recibe un nuevo identificador único: 5.

El nodo Generación de análisis de entidad añade los campos de identificador e indicador de fuente, y genera los resultados de los nuevos registros junto con todos los registros coincidentes. El resultado será el siguiente.

Tabla 2-3
Resultados del nodo Generación de análisis de entidad

SEA-ID	SEA-SRC	Nombre	Teléfono	Dirección	Riesgo de crédito	Ahorros	Préstamo
4	Cliente	Susan		1234 5th St		\$1234	
4	Préstamo	Sue	555-1234	1234 5th St			\$10,000
4	Nuevo préstamo	Suzan	555-1234	1234 5th Street			\$100,000
5	Nuevo préstamo	Mark	555-9999	888 9th Ave			\$60,000

El identificador del análisis de entidades puede agregar el resultado y transferirlo a otros nodos inferiores para que sigan procesándolo.

El nodo Generación de análisis de entidad se describe con detalle más adelante.

Fase 8: generar alertas

De nuevo, la actividad potencialmente sospechosa puede volverse aparente. En este caso, la persona con el identificador 4 ya tiene un préstamo de 10.000 \$ y ahora solicita otro diez veces mayor, con un nombre ligeramente distinto. Evidentemente, es algo perfectamente aceptable y no tiene por qué haber intenciones fraudulentas. Sin embargo, si dicha actividad resulta sospechosa de acuerdo con las reglas de su empresa, puede que merezca la pena comprobarlo.

Podría, por ejemplo, adjuntar y ejecutar un nodo Tabla o un nodo Informe de SPSS Modeler e imprimir el contenido de la ventana de resultados para que alguien lo lea y genere alertas manualmente. Otra posibilidad sería analizar el resultado del nodo Generación de análisis de entidad con un modelo de evaluación de riesgos previamente creado con IBM® SPSS® Modeler, para conseguir un conjunto de resultados que reflejen las reglas de su empresa con mayor precisión. Asimismo, puede exportar el resultado a una base de datos o algún otro medio de resultados para su procesamiento. Con SPSS Modeler, dispone de una amplia gama de opciones que se ajustan a sus necesidades concretas.

Tareas de análisis de entidades

Acerca de las tareas

En esta sección se describen las siguientes tareas de análisis de entidades.

- Creación de un repositorio de entidad
- Configuración de un repositorio de entidad
- Análisis las identidades resueltas
- Resolución de casos nuevos con el repositorio de entidad
- Purga de un repositorio de entidad
- Eliminación de un repositorio de entidad
- Administración del análisis de entidades

Creación de un repositorio de entidad (nodo de exportación de EA)

El proceso de creación de un repositorio de entidad consta de las tareas siguientes.

1. Conectar con una fuente de datos. Si desea obtener más información, consulte el tema [Conexión con una fuente de datos](#) el p. 15.
2. Crear el repositorio. Si desea obtener más información, consulte el tema [Creación del repositorio](#) el p. 16.
3. Asignar campos de entrada en la fuente de datos a funciones del repositorio. Si desea obtener más información, consulte el tema [Asignación de campos de entrada a funciones \(nodo de exportación de EA\)](#) el p. 19.

Una vez creadas las asignaciones, puede mostrarlas en la fuente de datos actual o en todas las fuentes de datos conocidas en el repositorio. Si desea obtener más información, consulte el tema [Visualización de las asignaciones de campos \(nodo de exportación de EA\)](#) el p. 21.

El repositorio de entidad

El repositorio ofrece un área de almacenamiento central que actúa como caché de datos de toda la información de la entidad. Dado que el repositorio está activo, su estado es único, y por tanto no existe el concepto de versión para un repositorio de entidad. El repositorio almacena el estado actual de todos los datos de entrada, por lo que puede alcanzar un gran tamaño.

Los contenidos del repositorio se pueden mantener mediante una interfaz gráfica fácil de utilizar. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.

Nota: la versión de IBM® SPSS® Modeler Entity Analytics proporcionada con IBM® SPSS® Modeler Premium admite un único repositorio alojado en el producto IBM solidDB que se encuentra en el mismo paquete que IBM SPSS Modeler Entity Analytics. Con esta versión, debe eliminar el repositorio existente antes de crear uno nuevo. Tiene a su disposición una actualización con licencia independiente de IBM SPSS Modeler Entity Analytics (conocido como IBM SPSS Modeler Entity Analytics Unleashed) que permite que más de un repositorio coexistan en el mismo sistema; cada repositorio puede contener más de 10 millones de filas y utilizar más de cuatro núcleos de procesadores. Póngase en contacto con su representante local de asistencia técnica de IBM para obtener información detallada.

Conexión con una fuente de datos

En primer lugar, se leen los datos de fuente en SPSS Modeler mediante un nodo de fuente.

Para conectar con una fuente de datos

1. En la pestaña Fuentes, en la paleta de nodos de la parte inferior de la ventana principal de SPSS Modeler, pulse dos veces en el icono correspondiente al tipo de datos de fuente. De este modo se añade un nodo de fuente al lienzo de pantalla.
2. En el lienzo de pantalla, pulse dos veces en el icono para abrir su cuadro de diálogo.
3. En el campo Archivo, introduzca la ubicación y el nombre del archivo de datos de fuente.
4. Complete el resto del cuadro de diálogo según sea necesario (pulse en el botón Ayuda para más información) y pulse Aceptar.
5. Si el archivo de datos de fuente no tiene un campo que identifique exclusivamente cada registro, añada uno mediante el nodo Derivar. Si desea obtener más información, consulte el tema [Adición de un identificador de registro único](#) el p. 15.

Adición de un identificador de registro único

Cada archivo de fuente de datos introducido en el repositorio de entidad debe tener un campo que identifique exclusivamente cada registro. Si un archivo de fuente de datos no cuenta con dicho campo, puede añadirlo mediante el nodo Derivar de SPSS Modeler.

Para añadir un identificador de registro único a un archivo de fuente de datos

1. En el lienzo de pantalla, pulse en el nodo de fuente que añadió en la tarea anterior.
2. En la pestaña Operaciones con campos de la paleta de nodos, pulse dos veces en el icono Derivar para adjuntar un nodo Derivar al nodo de fuente.
3. En el lienzo de pantalla, pulse dos veces en el nodo Derivar para abrir su cuadro de diálogo.
4. En el campo Derivar, cambie el nombre por defecto por un nombre distintivo (p. ej.: ID) para el campo de identificador que está añadiendo.
5. Asegúrese de que el campo Derivar como está establecido como Fórmula

6. Establezca el tipo de Campo como Continuo.
7. En el cuadro de texto Fórmula, introduzca @INDEX y pulse en Aceptar.

Creación del repositorio

Es necesario crear un repositorio para almacenar todos los datos entrantes.

Para crear un repositorio

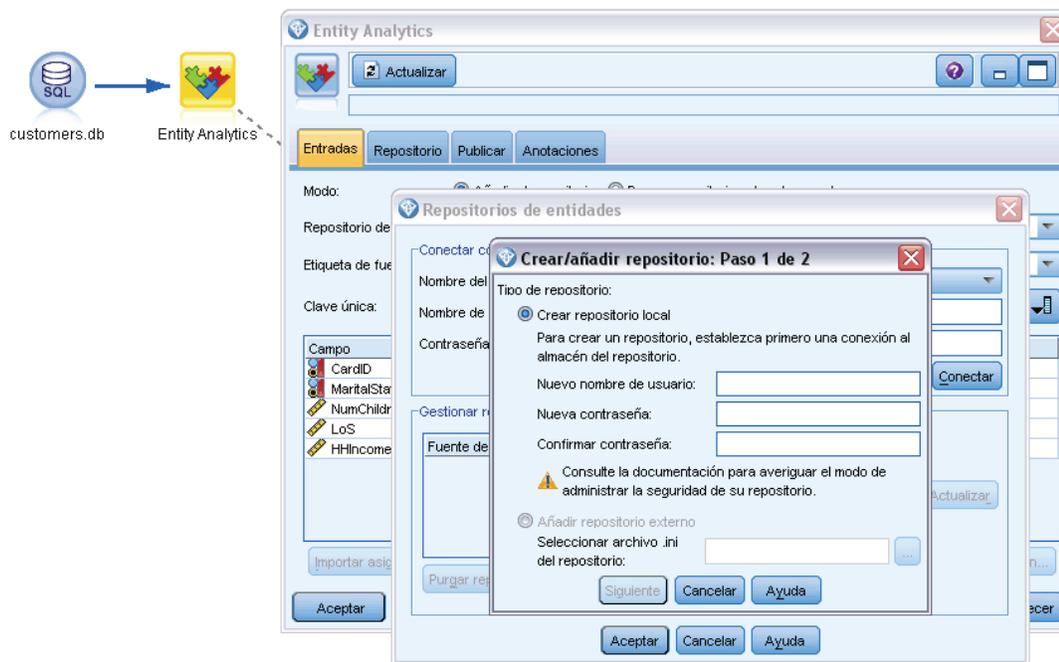
1. En la pestaña Exportar de la paleta de nodos de SPSS Modeler, ponga un nodo de exportación de EA en el lienzo de rutas.

Nota: si es la primera vez que crea un repositorio, utilice un nodo de exportación de EA y conéctelo al nodo de fuente de SPSS Modeler que contiene los datos que desea introducir en el repositorio (o al nodo Derivar, si ha añadido uno para obtener un campo de identificador único). Para conectar los nodos, haga lo siguiente.

- Pulse con el botón derecho en el nodo de fuente de SPSS Modeler.
 - Seleccione Conectar.
 - Pulse en el nodo de exportación de EA.
2. Pulse dos veces en el nodo de exportación de EA para abrir su cuadro de diálogo.
 3. Pulse en la lista del repositorio de entidad.
 4. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Repositorios de entidad.
 5. En el cuadro de diálogo Repositorios de entidad, pulse en el campo Nombre del repositorio.
 6. Elija <Crear repositorio nuevo...> para que se muestre el asistente de creación de repositorios.

Asistente de creación de repositorios

Figura 3-1
Asistente de creación de repositorios



Paso 1

Aquí debe seleccionar si desea crear un repositorio local, mediante el producto IBM solidDB que se encuentra en el mismo paquete que IBM SPSS Modeler Entity Analytics, o utilizar una base de datos externa para el repositorio.

Crear repositorio local. Especifique el nombre de usuario y la contraseña del administrador para la base de datos IBM solidDB que alojará el repositorio que está creando. Confirme la contraseña y pulse en Siguiente.

Nota: si posteriormente necesita cambiar las credenciales de administrador, puede hacerlo usando el editor de línea de comandos de la base de datos. Si desea obtener más información, consulte el tema [Gestión de las credenciales de administrador para la base de datos del repositorio](#) el p. 41.

Añadir repositorio externo. Utilice esta opción si desea utilizar una base de datos externa para alojar el repositorio. Escriba la ubicación del archivo .ini de la base de datos en el campo Seleccione el archivo .ini del repositorio y pulse en Siguiente.

Paso 2

Nombre de repositorio nuevo. Escriba un nombre no utilizado anteriormente para el nuevo repositorio.

Importar configuración de. (únicamente repositorio local) Si desea basar la configuración en la de un repositorio existente, elija aquí el repositorio; de lo contrario, elija Por defecto. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.

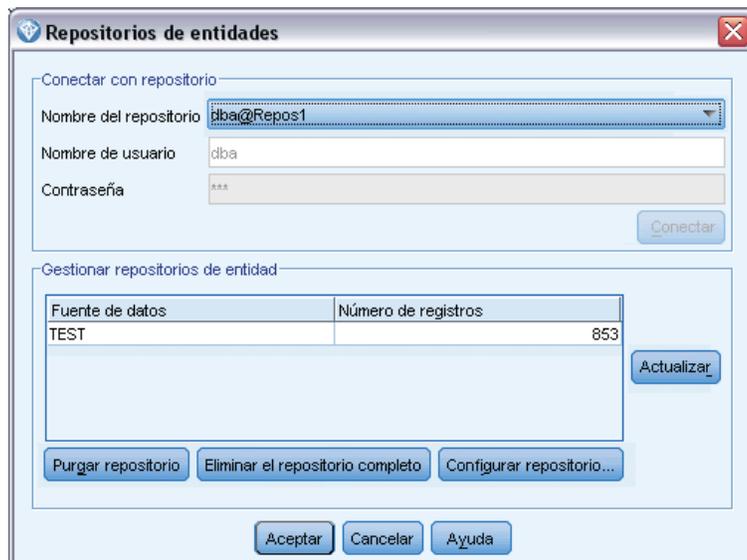
Si elige un repositorio existente, introduzca los detalles de conexión si difieren de los que introdujo en la pantalla anterior.

Pulse en Aceptar para crear el repositorio nuevo y que se muestre el cuadro de diálogo Casos de resolución de entidades, desde el que puede conectarse al repositorio.

Opciones de repositorio de entidad

El cuadro de diálogo Repositorios de entidad contiene varias opciones para crear, conectar, configurar y mantener un repositorio de entidad.

Figura 3-2
Opciones de repositorio de entidad



Conectar al repositorio. Utilice estas opciones para crear un nuevo repositorio de entidad o para conectar con uno existente.

- **Nombre del repositorio.** Muestra el repositorio de entidad actual si existe alguno. Si hay más de un repositorio, para elegir uno distinto selecciónelo de la lista.
Para crear un repositorio nuevo, seleccione <Crear repositorio nuevo...>. Se iniciará un asistente que le guiará a través del proceso de creación.
- **Nombre de usuario.** Introduzca su nombre de usuario para el repositorio seleccionado.
- **Contraseña.** La contraseña para este nombre de usuario.
- **Conectar.** Pulse para conectarse al repositorio actual.

Gestionar repositorios de entidad La tabla enumera las fuentes de datos cargadas en el repositorio actual (al que está conectado) y muestra el número de registros en cada fuente de datos.

- **Actualizar.** Actualiza la fuente de datos y el tamaño de información en la tabla (por ejemplo, si ha añadido una nueva fuente de datos o ha cambiado el tamaño de una existente).
- **Purgar repositorio.** Elimina todos los datos de fuente del repositorio, pero mantiene todos los detalles de configuración. Puede utilizar esta opción si la información de configuración aún le es útil pero desea eliminar todos los registros de datos del repositorio. Si desea obtener más información, consulte el tema [Purga de un repositorio de entidad](#) el p. 44.

Eliminar el repositorio completo. Elimina completamente los contenidos actuales del repositorio y los detalles de configuración. Si desea obtener más información, consulte el tema [Eliminación de un repositorio de entidad](#) el p. 44.

Configurar repositorio. Muestra una ventana donde puede configurar el repositorio actual. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.

Asignación de campos de entrada a funciones (nodo de exportación de EA)

El repositorio ofrece varias funciones predefinidas estándar. Fuentes de datos distintas pueden utilizar nombres de campo distintos (por ejemplo, Dirección1 o Datos de dirección 1) para tipos de información que corresponden a la misma función. Para evitar duplicidades, es necesario asignar campos de fuentes de datos entrantes a funciones del repositorio específicas. No es necesario asignar todos los campos en el conjunto de datos, solo aquellos que es probable que correspondan a la misma función en otro conjunto de datos.

Si una fuente de datos utiliza campos correspondientes a otros tipos de información no predefinidos en el repositorio, puede crear nuevas funciones en la ventana Configuración de repositorio. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.

Para asignar campos de entrada a funciones

1. Adjunte un nodo de exportación de EA a un nodo de fuente de datos en el lienzo de rutas. Cada nodo de fuente de datos que utilice deberá estar adjuntado a su propio nodo de exportación de EA.
2. Abra el nodo de exportación de EA para que se muestre la pestaña Entradas, que contiene opciones para asignar los campos de entrada. Si desea obtener más información, consulte el tema [Opciones de entrada del repositorio para asignación](#) el p. 20.
3. En el nodo de exportación de EA, seleccione la pestaña Repositorio para ver las asignaciones de la fuente de datos actual o, si utiliza más de una, de todas las fuentes de datos.
4. Para guardar un conjunto de asignaciones (por ejemplo, para usarlo con otro nodo de fuente de datos), pulse dos veces en Exportar asignación.

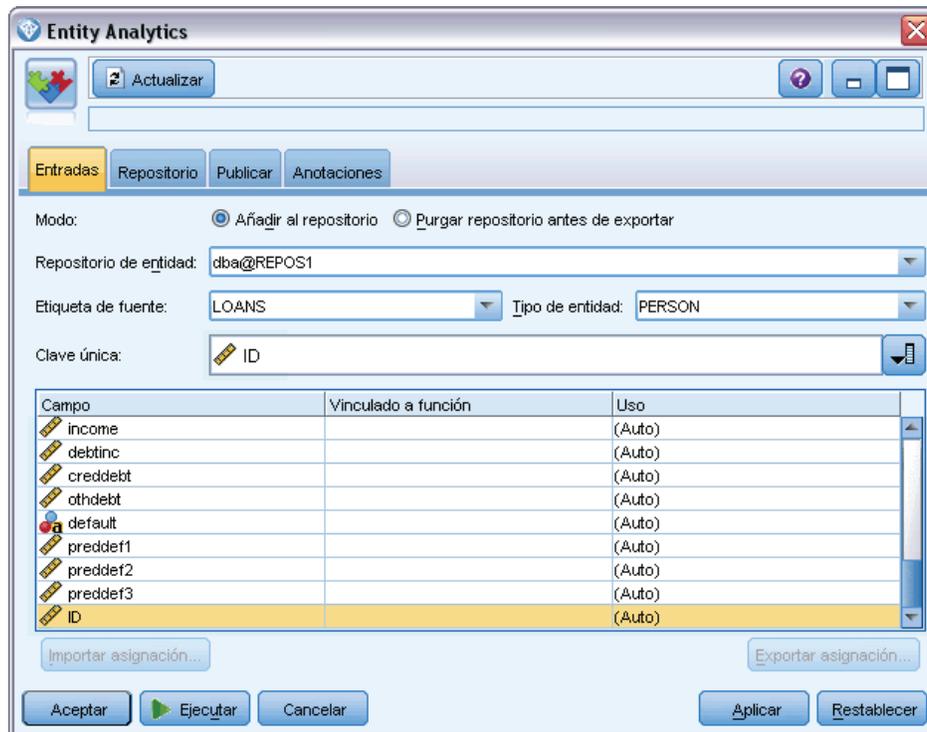
Cuando haya terminado de asignar el primer nodo de fuente de datos, repita el proceso con el resto de nodos de fuente de datos que desee utilizar.

Opciones de entrada del repositorio para asignación

La pestaña Entradas contiene las opciones para asignar campos de fuentes de datos a funciones del repositorio listas para exportar al repositorio. Cree las asignaciones en esta pestaña (opcionalmente, puede pulsar en la pestaña Repositorio para ver la asignación de otras fuentes de datos) y pulse en Ejecutar para exportar los datos al repositorio.

Si ya ha guardado un conjunto de asignaciones en un archivo XML, puede usarlas pulsando en Importar asignación.

Figura 3-3
Asignación de campos de entrada a funciones



Moda. Deje la selección predeterminada, Añadir al repositorio, si desea añadir los registros del archivo de fuente al contenido existente del repositorio. Si desea borrar el contenido del repositorio pero mantener la información de configuración antes de añadir los registros de fuente, seleccione Purgar repositorio antes de exportar.

Repositorio de entidad. Muestra el repositorio de entidad actual si existe alguno. Si hay más de un repositorio, para elegir uno distinto selecciónelo de la lista. Para crear un nuevo repositorio, elija <Examinar...> para ver un cuadro de diálogo en el que puede crear el repositorio. Si desea obtener más información, consulte el tema [Opciones de repositorio de entidad](#) el p. 18.

Etiqueta de fuente. Una lista de etiquetas que indica las fuentes de datos conocidas actualmente en el repositorio. Elija una de la lista o <Añadir nueva etiqueta de fuente...> para crear una etiqueta de una nueva fuente de datos.

Tipo de entidad. Una lista de tipos de entidades (es decir, conjuntos de funciones) definidos en el repositorio. Elija uno de la lista o <Añadir nuevo tipo de entidad...> para ver la ventana de configuración de repositorio, donde puede definir un nuevo tipo de entidad. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.

Clave única. (requerida) El campo de entrada que se utiliza para los identificadores únicos de los registros de datos.

Tabla de asignaciones. En esta tabla puede asignar cada campo de entrada a una función correspondiente en el repositorio. Si no existe ninguna función adecuada en el tipo de entidad seleccionado, puede crear una función nueva aquí.

- **Campo.** El conjunto de campos de entrada en la fuente de datos seleccionada. Cada campo tiene un icono que indica el nivel de medida (es decir, el tipo de datos) del campo.
- **Vinculado a función.** Para asignar un campo a una función, pulse dos veces en esta columna (o pulse la barra espaciadora) en la fila de campo y elija una función de la lista. Si no se muestra ninguna función adecuada, elija <Añadir nueva función...> para ver la ventana de configuración de repositorio, donde puede definir una función nueva para este tipo de entidad. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.
- **Uso.** Indica el contexto de un campo particular, en el que puede haber más de un contexto (por ejemplo, el número de teléfono de casa y el del trabajo). Hay tipos de uso predefinidos disponibles para las funciones ADDRESS y PHONE, y puede crear sus propios tipos de uso para todas las funciones. Para establecer un uso distinto al predeterminado (Auto), pulse en la fila que desea de esta columna y elija uno de los tipos de uso (si los hubiera) o pulse en <Añadir uso...> para crear uno nuevo. Si desea obtener más información, consulte el tema [Mantenimiento de los tipos de entidad](#) el p. 28.

Importar asignación. Importa un conjunto previamente exportado de asignaciones de campo a función desde un archivo XML externo. Puede resultar útil si tiene diferentes fuentes de datos con los mismos requisitos de asignación, ya que evita tener que redefinir las mismas asignaciones para fuentes distintas.

Exportar asignación. Exporta a un archivo XML externo el conjunto de asignaciones de campo a función que se muestran en la tabla de asignaciones.

Visualización de las asignaciones de campos (nodo de exportación de EA)

En la pestaña Repositorio, pulse en el botón Actualizar para ver las funciones del repositorio que tienen campos de entrada asociados a ellas. Puede ver las de la fuente de datos actual (la fuente controlada por el nodo de fuente asociado a este nodo de exportación) o de todas las fuentes de datos.

Figura 3-4
Visualización de las asignaciones de campos



Mostrar entradas para. Elija una opción para mostrar las asignaciones para la fuente de datos actual o para todas las fuentes de datos conocidas en el repositorio.

Actualizar. Actualiza la visualización de la opción de entrada seleccionada.

Funciones. Una lista de todas las funciones que tienen asignaciones en las fuentes de datos mostradas. Las funciones no asignadas no se muestran.

<Fuente de datos>. Cada columna enumera los campos asignados en una fuente de datos concreta para cada función para la que se ha definido una asignación.

Configuración de un repositorio de entidad

Puede mantener los contenidos del repositorio desde la ventana Configuración de repositorio, que ofrece una interfaz visual fácil de utilizar para el repositorio al completo.

si va a utilizar más de un repositorio con configuraciones idénticas o similares, puede crear una configuración básica y exportarla a un archivo que podrá importar desde otros repositorios. Si desea obtener más información, consulte el tema [Reutilización de configuración de repositorio](#) el p. 31.

Precaución: Si modifica y guarda la configuración de un repositorio que ya contiene datos, es muy recomendable que purgue el contenido del repositorio y vuelva a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

Para crear una configuración de repositorio

1. Abra cualquier nodo Entity Analytics.
2. Pulse en la lista del repositorio de entidad.
3. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Casos de resolución de entidades.
4. En el cuadro de diálogo Casos de resolución de entidades, pulse en la lista Nombre del repositorio.
5. Seleccione el repositorio para el que quiere crear la configuración.
6. Si no está conectado, introduzca el nombre de usuario de administrador y la contraseña y pulse en Conectar.
7. Cuando el botón Configurar repositorio esté activado, púlselo para ver la ventana Configuración de repositorio.
8. Cree los detalles de configuración como se explica en las secciones siguientes.

El panel de navegación en la parte izquierda de la ventana Configuración de repositorio contiene una estructura de árbol desde la que puede gestionar las distintas características del repositorio.

Tabla 3-1

Elementos principales de la ventana Configuración de repositorio

Sección	Descripción	
Fuentes de datos	Muestra las asignaciones de todas las fuentes de datos de las diversas funciones del repositorio.	Si desea obtener más información, consulte el tema Visualización de las asignaciones de fuentes de datos el p. 23.
Funciones	Crea una nueva función o duplica, edita o elimina una existente.	Si desea obtener más información, consulte el tema Mantenimiento de las funciones del repositorio el p. 24.
Tipos de entidad	Crea un nuevo tipo de entidad, o gestiona los existentes (duplicar, cambiar nombre, adjuntar o eliminar funciones, borrar).	Si desea obtener más información, consulte el tema Mantenimiento de los tipos de entidad el p. 28.
Reglas de resolución	Establecer umbral para coincidencia de entidades.	Si desea obtener más información, consulte el tema Establecimiento del umbral para coincidencia de entidades . el p. 31.

Visualización de las asignaciones de fuentes de datos

En la sección Fuentes de datos de la ventana Configuración de repositorio, la entrada Todas las fuentes muestra las asignaciones en formato de solo lectura de todas las fuentes de datos de las diversas funciones del repositorio.

Figura 3-5
Asignaciones de fuentes de datos en la ventana de configuración



Pulse en Actualizar para actualizar la lista si se han añadido nuevas fuentes de datos al repositorio.

Nota: no puede añadir una fuente de datos al repositorio aquí. Las fuentes de datos solo se pueden añadir creando un nodo de fuente de SPSS Modeler y conectándolo al nodo de exportación de Entity Analytics. Si desea obtener más información, consulte el tema [Conexión con una fuente de datos](#) el p. 15.

Mantenimiento de las funciones del repositorio

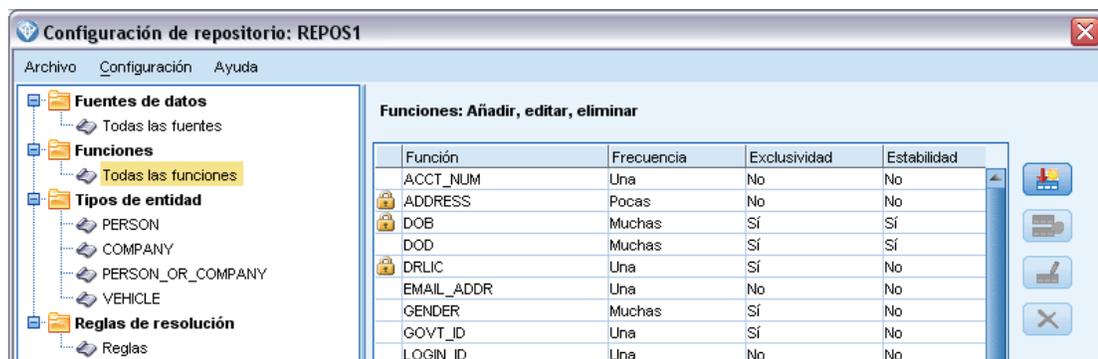
Una función de repositorio es un tipo de información individual que se puede utilizar con una fuente de datos de entidad. Algunas funciones (por ejemplo, Nombre, Apellidos, Fecha de nacimiento, etc.) se pueden utilizar con muchas fuentes de datos diferentes, mientras que otras funciones son específicas de una fuente de datos concreta. Una función puede contener uno o más elementos; en términos generales, una función equivale a un campo de un registro de datos o a una columna de una tabla de base de datos.

En la sección Funciones de la ventana Configuración de repositorio, la entrada Todas las funciones proporciona los medios para mantener todas las funciones del repositorio. Puede hacer lo siguiente.

- Crear una nueva función
- Duplicar una función existente (por ejemplo, para crear una función nueva basada en una existente)
- Modificar una función existente
- Eliminar una función existente

Puede encontrar las instrucciones de estas tareas más adelante en esta sección.

Figura 3-6
Funciones enumeradas en la ventana de configuración



La lista de funciones muestra todas las funciones que se han definido en este repositorio. Las columnas de la lista muestran las diversas propiedades que puede tener una función.

Función. El nombre de la función. Un símbolo de candado junto al nombre de una función indica que la función está bloqueada. Las funciones que están bloqueadas no se pueden eliminar ni duplicar, ni se pueden guardar cambios en ellas.

Frecuencia. Indica el número de identidades que pueden tener el mismo valor en esta función. Los valores válidos son Ninguna, Nombre, Una (p. ej., el número de pasaporte), Pocas (p. ej., la dirección) o Muchas (p. ej., la fecha de nacimiento).

Exclusividad. Indica que una entidad debe tener normalmente solo un tipo de función. Por ejemplo, una fecha de nacimiento o un número de identidad nacional tendría el valor Sí aquí, mientras que una dirección o número de tarjeta de crédito tendría el valor No (dado que una entidad puede tener varias direcciones o tarjetas de crédito).

Estabilidad. Indica el valor de estabilidad de esta función (es decir, si es *poco probable* que cambie durante la vida útil de una entidad). Por ejemplo, una función de fecha de nacimiento tendría el valor Sí, ya que nunca cambia, pero una función de dirección tendría No, ya que es bastante probable que cambie y, por lo tanto, es menos estable. *Nota:* el valor de género normalmente es estable a lo largo de la vida pero, dado que con frecuencia se especifica de manera incorrecta debido a datos erróneos, la configuración por defecto le da el valor No.

Para crear una nueva función

1. Realice una de las siguientes acciones.
 - Pulse en el botón Crear nueva función (el botón superior en la parte derecha de la pantalla).
 - Pulse con el botón derecho en Todas las funciones en el panel de navegación a la izquierda de la pantalla, y elija Nueva función.
2. Complete el cuadro de diálogo Añadir o modificar función. Si desea obtener más información, consulte el tema [Adición o modificación de una función](#) el p. 27.

Para duplicar una función existente

1. En la columna Función de la tabla en la parte derecha de la pantalla, seleccione la función que quiere duplicar.
2. Pulse en el botón Duplicar función seleccionada (el segundo botón en la parte derecha de la pantalla).
3. Complete el cuadro de diálogo Añadir o modificar función. Si desea obtener más información, consulte el tema [Adición o modificación de una función](#) el p. 27.

Para modificar una función existente

Precaución: si modifica o elimina una función o un elemento de función cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

1. En la columna Función de la tabla en la parte derecha de la pantalla, seleccione la función que quiere modificar. *Nota:* puede modificar solamente las funciones que haya creado, no las proporcionadas por el sistema.
2. Pulse en el botón Modificar función seleccionada (el tercer botón en la parte derecha de la pantalla).
3. Complete el cuadro de diálogo Añadir o modificar función. Si desea obtener más información, consulte el tema [Adición o modificación de una función](#) el p. 27.

Para eliminar una función existente

Precaución: si modifica o elimina una función o un elemento de función cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

1. En la columna Función de la tabla en la parte derecha de la pantalla, seleccione la función que quiere eliminar. *Nota:* puede eliminar solamente las funciones que haya creado, no las proporcionadas por el sistema.
2. Realice una de las siguientes acciones.
 - Pulse en el botón Eliminar función seleccionada (el botón inferior en la parte derecha de la pantalla).
 - Pulse con el botón derecho en Todas las funciones en el panel de navegación a la izquierda de la pantalla, y elija Eliminar.
3. Pulse en Continuar para confirmar la eliminación de la función. *Nota:* esta eliminación no se puede deshacer.

Adición o modificación de una función

Precaución: si modifica o elimina una función o un elemento de función cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

En el cuadro de diálogo Añadir o modificar función, puede crear una nueva función de repositorio, o duplicar o modificar una función existente.

Figura 3-7
Modificación de una función

Elemento	Descripción	Tipo de dato	Buscar	Comparar	Descriptivo
1 ADDR_TYPE	Address t...	Cadena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2 FULL_ADDR	Full address	Cadena	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 ADDR1	Addr line 1	Cadena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4 ADDR2	Addr line 2	Cadena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5 ADDR3	Addr line 3	Cadena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6 CITY	City	Cadena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7 STATE	State	Cadena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8 POSTAL_...	Postal code	Cadena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
9 COUNTRY	Country	Cadena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Tipo de función. Una etiqueta que indica el tipo de información con el que está relacionada la función. La etiqueta constituye la primera parte del identificador de la función.

Descripción. Un texto breve que describe el tipo de función, con fines informativos únicamente.

Frecuencia. Indica el número de identidades que pueden tener el mismo valor en esta función. Los valores válidos son Ninguna, Nombre, Una (p. ej., el número de pasaporte), Pocas (p. ej., la dirección) o Muchas (p. ej., la fecha de nacimiento).

Exclusividad. Indica que una entidad debe tener normalmente solo un tipo de función. Por ejemplo, una fecha de nacimiento o un número de identidad nacional tendría el valor Sí aquí, mientras que una dirección o número de tarjeta de crédito tendría el valor No (dado que una entidad puede tener varias direcciones o tarjetas de crédito).

Estabilidad. Indica el valor de estabilidad de esta función (es decir, si es *poco probable* que cambie durante la vida útil de una entidad). Por ejemplo, una función de fecha de nacimiento tendría el valor Sí, ya que nunca cambia, pero una función de dirección tendría No, ya que es bastante probable que cambie y, por lo tanto, es menos estable. *Nota:* el valor de género normalmente es

estable a lo largo de la vida pero, dado que con frecuencia se especifica de manera incorrecta debido a datos erróneos, la configuración por defecto le da el valor No.

Persistir historial. Si se establece como Sí (valor predeterminado), preserva el historial de cambios del valor de campo. Esto es de utilidad, por ejemplo, en el caso del campo de dirección de un cliente. En este caso, conservar el historial de cambios podría mejorar la coincidencia si un cliente ha cambiado de dirección un par de veces. Sin embargo, para un campo como el saldo de una cuenta, debería establecerlo como No para no conservar cantidades de datos excesivas.

Tabla de elementos. Una lista de los elementos que comprende esta función.

- **Elemento.** Nombre del elemento.
- **Descripción.** Una descripción breve de lo que proporciona el elemento.
- **Tipo de datos.** El tipo de los datos que se pueden utilizar para este elemento.
- **Buscar.** Si se selecciona, significa que este elemento de la función puede buscarse por sí mismo o en combinación con otros elementos de la búsqueda etiquetada para la función. Por ejemplo, supongamos que una función PASSPORT tiene dos elementos: ID_NUM y COUNTRY. Si quiere poder buscar únicamente en ID_NUM, debe seleccionar su cuadro Buscar y cancelar la selección del cuadro Buscar de COUNTRY. Los modificadores de la función, como país o estado o fecha de emisión y caducidad, no formarán parte de la búsqueda.
- **Comparar.** Indica qué elementos de la función se enviarán a la rutina de comparación para la resolución de entidades. Siguiendo con el ejemplo de la función PASSPORT con los elementos ID_NUM y COUNTRY, ya que puede que COUNTRY no se proporcione siempre, únicamente ID_NUM debería enviarse a la rutina de comparación.
- **Descriptivo.** Indica qué elementos de la función deberían aparecer en su descripción interna. En el ejemplo anterior de una función PASSPORT con los elementos ID_NUM y COUNTRY, si quiere que ambos aparezcan en la descripción de la función, debe seleccionar la casilla Descriptivo de ambos. Pero supongamos que la función tiene un elemento adicional: ISSUE_DT. Debe cancelar la selección de su casilla Descriptivo para asegurarse de que no aparezca en la descripción de la función.

Botón Añadir nuevo elemento. Añade una fila nueva a la tabla de elementos para poder definir un elemento nuevo.

Botón Eliminar elemento. Elimina una fila seleccionada de la tabla de elementos. Esta acción no se puede deshacer.

Precaución: si modifica o elimina una función o un elemento de función cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

Mantenimiento de los tipos de entidad

Un **tipo de entidad** es un conjunto nombrado de funciones de repositorio que pertenecen al mismo grupo de manera lógica. Por ejemplo, un tipo de entidad pensado para su uso con un conjunto de datos de clientes puede constar de funciones como Nombre, Fecha de nacimiento, Sexo, Dirección, Número de teléfono, etc.

El repositorio de IBM SPSS Modeler Entity Analytics se proporciona con un conjunto estándar de tipos de entidad y puede añadir sus propios tipos de entidad.

La sección Tipos de entidad de la ventana Configuración de repositorio enumera los distintos tipos de entidad que se han creado. Puede hacer lo siguiente.

- Crear un nuevo tipo de entidad
- Duplicar un tipo de entidad existente (por ejemplo, para crear un tipo de entidad nuevo basado en uno existente)
- Adjuntar funciones a un tipo de entidad
- Eliminar funciones de un tipo de entidad
- Cambiar nombre a un tipo de entidad
- Eliminar un tipo de entidad

Figura 3-8

Enumeración de funciones que pertenecen a un tipo de entidad



Tipo de entidad. El nombre del tipo de entidad seleccionado.

Función. La lista de funciones válidas que comprende este tipo de entidad.

Tipo de uso. (Opcional) Indica distintos contextos en los que se podría usar esta función. Pulse dos veces en esta columna para añadir o modificar un tipo de uso, separando los tipos de uso con una coma y un espacio. Los valores que especifique aquí definen los valores mostrados en el nodo de exportación de EA o el nodo Generación de análisis de entidad cuando un usuario pulsa en la columna Uso para una función en la pestaña Entradas. Si desea obtener más información, consulte el tema [Opciones de entrada del repositorio para asignación](#) el p. 20.

Para crear un nuevo tipo de entidad

1. Pulse con el botón derecho en Tipos de entidad en el panel de navegación a la izquierda de la pantalla.
2. Elija Nuevo tipo de entidad.
3. Introduzca un nombre no utilizado anteriormente para el tipo de entidad y pulse en Aceptar.

4. Adjunte funciones al tipo de entidad (ver sección siguiente).

Para adjuntar funciones a un tipo de entidad

1. Seleccione el tipo de entidad en el panel de navegación a la izquierda de la pantalla.
2. Pulse en el botón Adjuntar función (el botón superior en la parte derecha de la pantalla).
3. De la lista de funciones disponibles, elija una o varias (deje pulsada la tecla Ctrl mientras pulsa en las distintas funciones) y pulse en Aceptar.

Para eliminar funciones de un tipo de entidad

1. Seleccione el tipo de entidad en el panel de navegación a la izquierda de la pantalla.
2. Seleccione una o más funciones de la tabla de funciones adjuntadas en la parte derecha de la pantalla. Mantenga pulsada la tecla Ctrl para elegir más de una función.
3. Pulse en el botón Desvincular función (el botón inferior en la parte derecha de la pantalla).

Para duplicar un tipo de entidad existente

1. En el panel de navegación de la parte izquierda de la pantalla, pulse con el botón derecho en el tipo de entidad que desea duplicar.
2. Elija Duplicar tipo de entidad.
3. Introduzca un nombre no utilizado anteriormente para el nuevo tipo de entidad y pulse en Aceptar.
4. Adjunte o desvincule funciones del tipo de entidad según sea necesario (ver instrucciones anteriores).

Para cambiar el nombre a un tipo de entidad

Precaución: si modifica o elimina una función o un elemento de función cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

1. En el panel de navegación de la parte izquierda de la pantalla, pulse con el botón derecho en el tipo de entidad al que desea cambiar el nombre.
2. Elija Cambiar nombre.
3. Introduzca el nombre nuevo para el tipo de entidad y pulse en Aceptar.

Para eliminar un tipo de entidad

Precaución: si modifica o elimina una función o un elemento de función cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

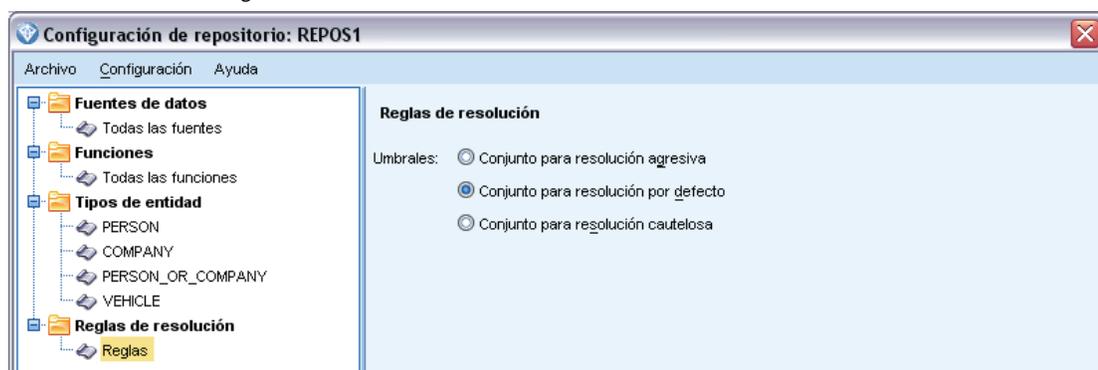
1. En el panel de navegación de la parte izquierda de la pantalla, pulse con el botón derecho en el tipo de entidad que desea eliminar.

2. Seleccione Eliminar.
3. Pulse en Aceptar para confirmar la eliminación del tipo de entidad. *Precaución:* esta eliminación no se puede deshacer.

Establecimiento del umbral para coincidencia de entidades.

En la sección Reglas de resolución de la ventana Configuración de repositorio, puede elegir el umbral al que se producirá la coincidencia de entidades.

Figura 3-9
Establecimiento de reglas de resolución de entidades



Al crear el repositorio, la coincidencia está definida en el umbral por defecto.

Elija Conjunto para resolución agresiva si no encuentra suficientes coincidencias en sus registros para llevar a cabo una resolución de entidad.

Elija Conjunto para resolución por defecto para regresar al umbral por defecto desde una de las otras configuraciones.

Elija Conjunto para resolución cautelosa si recupera demasiadas coincidencias.

Reutilización de configuración de repositorio

Si ya ha creado una configuración y desea utilizarla en otro repositorio, puede exportar la configuración existente a un archivo XML e importarlo desde el otro repositorio (destino).

Para reutilizar una configuración existente

1. Abra la ventana Configuración de repositorio del repositorio cuya configuración desea utilizar. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.
2. En el menú de esa ventana, elija Configuración > Exportar configuración.

3. En el cuadro de diálogo Guardar como, elija el nombre y la ubicación del archivo XML de exportación.
4. Abra la ventana Configuración de repositorio del repositorio de destino.
5. En el menú de esa ventana, elija Configuración > Importar configuración.
6. En el cuadro de diálogo Abrir, elija el nombre y la ubicación del archivo XML exportado anteriormente y pulse en Abrir.

Guardar sus cambios de configuración

Para guardar los cambios de configuración

En el menú de la ventana Configuración de repositorio, elija File > Guardar.

Cerrar la ventana de configuración

Para salir de la ventana de configuración

En el menú de la ventana Configuración de repositorio, elija File > Salir.

Si hay cambios sin guardar en la configuración, pulse en Aceptar para guardar los cambios y salir, o en Cancelar para salir sin guardar.

Análisis de las identidades resueltas [nodo de fuente Entity Analytics(EA)]

Con al menos una entrada de fuente de datos en el repositorio, puede utilizar el nodo de fuente Entity Analytics(EA) para pasar las identidades resueltas a otros nodos IBM® SPSS® Modeler para un posterior análisis o procesamiento, como crear un informe que enumere las identidades resueltas.

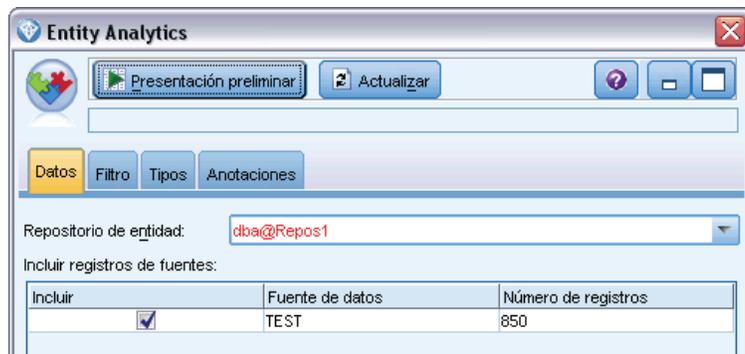
Para analizar las identidades resueltas

1. Añada un nodo de fuente Entity Analytics(EA) a una ruta.
2. Abra el nodo de fuente Entity Analytics(EA).
3. En la pestaña Datos, seleccione el repositorio de entidad y una o más de sus fuentes de datos de entrada (pulse en Actualizar para actualizar los recuentos de registros). Si desea obtener más información, consulte el tema [Selección de una fuente de datos](#) el p. 33.
4. Añada más nodos a la ruta para realizar los procesos que desee. Si desea obtener más información, consulte el tema [Adición de nodos a la ruta](#) el p. 34.

Selección de una fuente de datos

En la pestaña Datos, seleccione al menos una fuente de datos en el repositorio en el que se realizarán los próximos procesos. Pulse en Actualizar para actualizar los recuentos de registros de las fuentes de datos enumeradas.

Figura 3-10
Selección de una fuente de datos en el nodo Fuente



Repositorio de entidad. Muestra el repositorio de entidad actual si existe alguno. Si hay más de un repositorio, para elegir uno distinto selecciónelo de la lista. Para crear un nuevo repositorio, elija <Examinar...> para ver un cuadro de diálogo en el que puede crear el repositorio. Si desea obtener más información, consulte el tema [Opciones de repositorio de entidad](#) el p. 18.

Incluir registros de fuentes. Esta tabla enumera las distintas fuentes de datos que se han introducido en el repositorio, junto con el número de registros en cada fuente. Seleccione la casilla de verificación Incluir de aquellas fuentes de datos que desee utilizar para realizar análisis y procesos posteriores.

Cambio de nombre de campos de datos

Puede usar la pestaña Filtro para cambiar el nombre de los campos de identidades resueltas que pasan hacia abajo para procesos posteriores. Puede que quiera cambiar el nombre de un campo de identidad resuelta, por ejemplo, para mantener la compatibilidad del nombre de campo en la fusión hacia abajo con otro conjunto de datos.

Los campos con sus nombres originales son los siguientes.

Tabla 3-2
Campos de identidades resueltas

Campo	Descripción
\$EA-ID	Identificador de entidad
\$EA-SRC	Etiqueta de fuente que identifica la fuente de datos donde se crearon los registros
\$EA-KEY	Campo designado como clave única en un archivo de fuente de datos

Nota: aunque también puede usar la pestaña Filtro para excluir campos, no debería hacerlo ya que los campos de identidades resueltas son el mínimo necesario para el procesamiento de análisis de entidades.

Establecimiento de la información de tipos para campos de datos

En la pestaña Tipos, puede ver o cambiar diversas propiedades de los campos de identidades resueltas que pasan hacia abajo para procesos posteriores.

Las propiedades que puede cambiar son las mismas que las de la pestaña Tipos de un nodo Tipo de SPSS Modeler normal, y son las siguientes.

Tabla 3-3
Propiedades de Tipo para campos

Propiedad	Descripción
Medida	El nivel de medida (es decir, el tipo de datos), usado para describir características de los datos en el campo.
Valores	Proporciona opciones para la lectura de valores de datos desde el conjunto de datos.
Perdidos	Se utiliza para especificar cómo se tratan los valores perdidos de un campo.
Compruebe	Opciones de validación para asegurar que los valores de campo se ajustan a los valores o rangos especificados.
Rol	Especifica el modo en que se usará el campo si se transfieren los datos a un nodo de modelado o un nugget de modelo.

Adición de nodos a la ruta

Puede añadir diversos nodos de SPSS Modeler a la ruta para realizar operaciones de análisis o procesos de los resultados del nodo de fuente Entity Analytics(EA). Por ejemplo, puede añadir uno o más de los siguientes.

- Nodo Agregar para resumir los resultados, que pueden ser muy voluminosos
- Nodo Seleccionar para seleccionar un subconjunto de los resultados
- Nodo Tabla para ver el resultado desde el nodo de fuente Entity Analytics(EA)
- Nodo Informe para imprimir los resultados en un informe
- Nodo de exportación de SPSS Modeler para exportar el resultado en un formato diferente, como una hoja de cálculo o base de datos

Si desea más información, consulte las secciones sobre los nodos Operaciones con registros, Resultado y Exportar en la guía Nodos de fuente, proceso y resultado de *IBM SPSS Modeler*.

Comparación de nuevos casos con el repositorio (nodo Generación de análisis de entidad)

Una vez realizada alguna resolución de identidad en el repositorio, puede utilizar el nodo Generación de análisis de entidad para comparar casos nuevos que encuentre a partir de entonces con el contenido del repositorio. Este nodo procesa registros desde una nueva fuente de datos,

los compara con las entidades resueltas que ya se encuentran en el repositorio y pasa cualquier registro coincidente para su posterior procesamiento. Se puede establecer que las coincidencias sean exactas o relacionadas más libremente con entidades existentes.

Al igual que el nodo de exportación de EA, el nodo Generación de análisis de entidad toma un único nodo de fuente de SPSS Modeler como entrada. Sin embargo, el nodo Generación de análisis de entidad difiere en lo siguiente. Mientras que el nodo de exportación genera registros para todas las entidades relacionadas con sus registros de entrada, el nodo Generación de análisis de entidad únicamente genera registros para aquellas entidades relacionadas con las entidades ya resueltas en el repositorio. Si desea obtener más información, consulte el tema [Resultados del nodo Generación de análisis de entidad](#) el p. 40.

Para comparar casos nuevos con el repositorio

1. Conéctese a la fuente de datos que contiene los nuevos registros que desee comparar con las entidades existentes. Si desea obtener más información, consulte el tema [Conexión con una fuente de datos](#) el p. 15.
2. En la pestaña Operaciones con registros, adjunte un nodo Generación de análisis de entidad al nodo de fuente de datos.
3. Pulse dos veces en el nodo de exportación de Entity Analytics para abrir su cuadro de diálogo.
4. Pulse en la lista del repositorio de entidad.
5. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Repositorios de entidad.
6. En el cuadro de diálogo Repositorios de entidad, pulse en el campo Nombre del repositorio.
7. Pulse en el nombre del repositorio que desea usar.
8. Especifique el nombre de usuario y contraseña de este repositorio y pulse en Conectar. Pulse en Aceptar cuando el repositorio esté conectado.
9. En el cuadro de diálogo Generación de análisis de entidad, seleccione el tipo de entidad que desea usar. Si desea obtener más información, consulte el tema [Mantenimiento de los tipos de entidad](#) el p. 28.
10. Asigne los campos de entrada en la fuente de datos a funciones del repositorio. Si desea obtener más información, consulte el tema [Asignación de campos de entrada a funciones \(nodo Generación de análisis de entidad\)](#) el p. 36.
11. De manera opcional, puede pulsar en la pestaña Repositorio para ver una tabla que muestra las asignaciones para esta fuente de datos (o todas las fuentes de datos que se han asignado). Si desea obtener más información, consulte el tema [Visualización de las asignaciones de campos \(nodo Generación de análisis de entidad\)](#) el p. 37.
12. Pulse en la pestaña Ver para ver detalles de las distintas fuentes de datos introducidas en el repositorio y establezca el criterio de selección para recuperar entidades existentes. Si desea obtener más información, consulte el tema [Visualización de fuentes de datos \(nodo Generación de análisis de entidad\)](#) el p. 38.

13. Pulse en Aceptar cuando haya configurado el nodo correctamente.
14. Conecte un nodo Tabla al nodo Generación de análisis de entidad y ejecute la ruta.

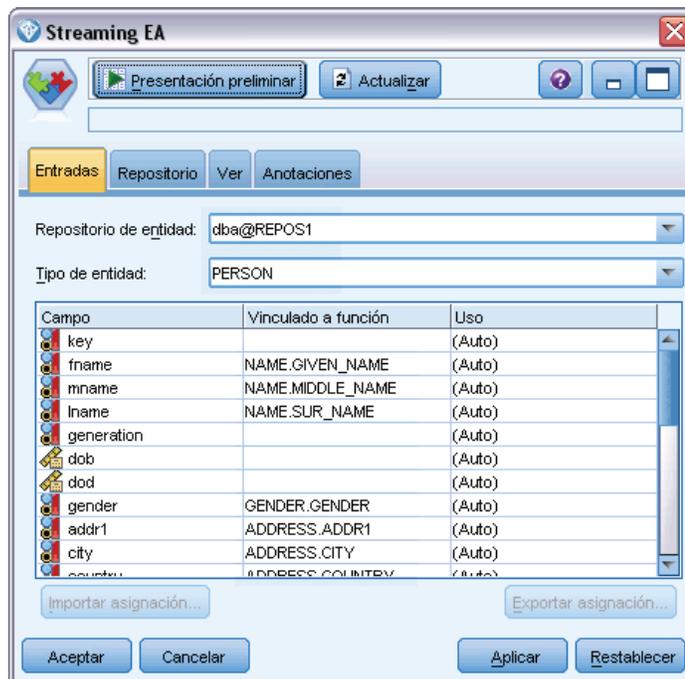
La ventana de resultados del nodo Tabla enumera todas las entidades recuperadas que coinciden con los nuevos registros de la fuente de datos. Los campos de resultados tienen el prefijo \$EA-añadido. Si desea obtener más información, consulte el tema [Resultados del nodo Generación de análisis de entidad](#) el p. 40.

Asignación de campos de entrada a funciones (nodo Generación de análisis de entidad)

La pestaña Entradas contiene las opciones para asignar campos de la entrada de este nodo a las funciones del repositorio. Cree las asignaciones en esta pestaña, o la pestaña Ver para ver detalles de todas las fuentes de datos en el repositorio y, a continuación, pulse en Aceptar.

Si ya ha guardado un conjunto de asignaciones en un archivo XML, puede usarlas pulsando en Importar asignación.

Figura 3-11
Asignación de campos de nuevos registros a funciones en el repositorio



Repositorio de entidad. Muestra el repositorio de entidad actual si existe alguno. Si hay más de un repositorio, para elegir uno distinto selecciónelo de la lista. Para crear un nuevo repositorio, elija <Examinar...> para ver un cuadro de diálogo en el que puede crear el repositorio. Si desea obtener más información, consulte el tema [Opciones de repositorio de entidad](#) el p. 18.

Tipo de entidad. Una lista de tipos de entidades (es decir, conjuntos de funciones) definidos en el repositorio. Elija uno de la lista o <Añadir nuevo tipo de entidad...> para ver la ventana de configuración de repositorio, donde puede definir un nuevo tipo de entidad. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.

Tabla de asignaciones. En esta tabla puede asignar cada campo de entrada a una función correspondiente en el repositorio. Si no existe ninguna función adecuada en el tipo de entidad seleccionado, puede crear una función nueva aquí.

- **Campo.** El conjunto de campos de entrada en la fuente de datos seleccionada. Cada campo tiene un icono que indica el nivel de medida (es decir, el tipo de datos) del campo.
- **Vinculado a función.** Para asignar un campo a una función, pulse dos veces en esta columna (o pulse la barra espaciadora) en la fila de campo y elija una función de la lista. Si no se muestra ninguna función adecuada, elija <Añadir nueva función...> para ver la ventana de configuración de repositorio, donde puede definir una función nueva para este tipo de entidad. Si desea obtener más información, consulte el tema [Configuración de un repositorio de entidad](#) el p. 22.
- **Uso.** Indica el contexto de un campo particular, en el que puede haber más de un contexto (por ejemplo, el número de teléfono de casa y el del trabajo). Si desea obtener más información, consulte el tema [Mantenimiento de los tipos de entidad](#) el p. 28.

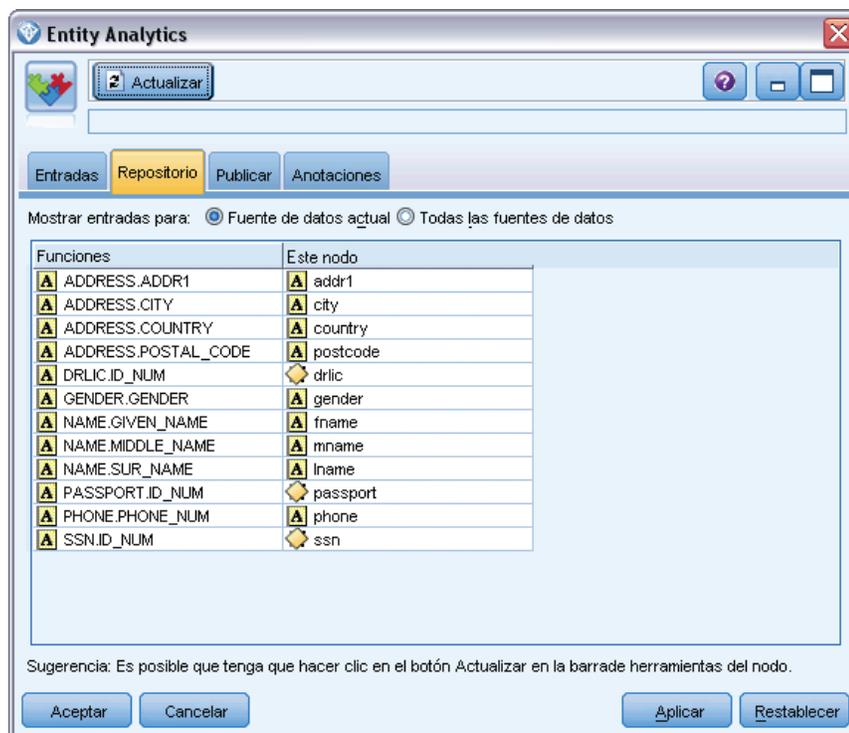
Importar asignación. Importa un conjunto previamente exportado de asignaciones de campo a función desde un archivo XML externo. Puede resultar útil si tiene diferentes fuentes de datos con los mismos requisitos de asignación, ya que evita tener que redefinir las mismas asignaciones para fuentes distintas.

Exportar asignación. Exporta a un archivo XML externo el conjunto de asignaciones de campo a función que se muestran en la tabla de asignaciones.

Visualización de las asignaciones de campos (nodo Generación de análisis de entidad)

En la pestaña Repositorio, pulse en Actualizar para ver las funciones del repositorio que tienen campos de entrada asociados a ellas. Puede ver las de la fuente de datos actual (la fuente controlada por el nodo de fuente asociado a este nodo de exportación) o de todas las fuentes de datos.

Figura 3-12
Visualización de las asignaciones de campos



Mostrar entradas para. Elija una opción para mostrar las asignaciones para la fuente de datos actual o para todas las fuentes de datos conocidas en el repositorio.

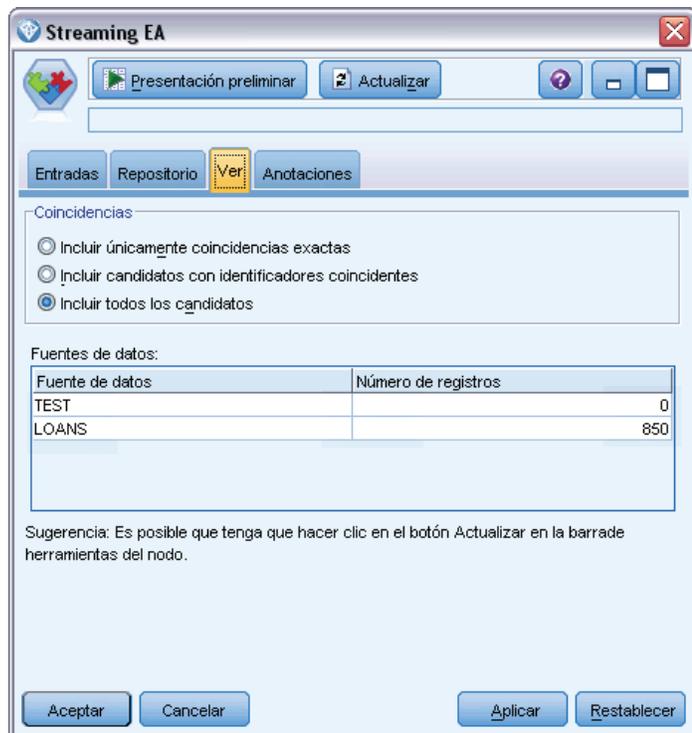
Funciones. Una lista de todas las funciones que tienen asignaciones en las fuentes de datos mostradas. Las funciones no asignadas no se muestran.

Este nodo. Cada columna enumera los campos asignados en una fuente de datos concreta para cada función para la que se ha definido una asignación.

Visualización de fuentes de datos (nodo Generación de análisis de entidad)

En la pestaña Ver puede ver detalles de diversas fuentes de datos que se han introducido en el repositorio. Estas son las fuentes de datos con las que se procesan las entradas este nodo, para buscar y recuperar entidades coincidentes. Pulse en Actualizar para actualizar los recuentos de registros.

Figura 3-13
Visualización de los detalles de fuentes de datos



Coincidencias. Estas opciones especifican la precisión con que la información de asignaciones de campo a función que ha especificado en la pestaña Entradas se hará coincidir con los registros de candidato (es decir, todos los contenidos del repositorio). Cuanto más preciso sea el criterio de búsqueda, menos entidades serán devueltas.

- **Incluir solo resultados exactos.** Éste es el criterio de coincidencia más preciso y con el que se seleccionan resultados en menos registros. Utilice esta opción cuando desee que se devuelvan únicamente estas entidades consideradas como coincidencias exactas.
- **Incluir a candidatos con identificadores coincidentes.** Utilice esta configuración cuando desee que se devuelvan las entidades coincidentes y las entidades que compartan los mismos identificadores (aquellos cuyas funciones se hayan configurado con un valor de frecuencia de Uno; por ejemplo, números de identificación fiscal o números de tarjeta de crédito coincidentes, etc.).
- **Incluir a todos los candidatos.** Utilice esta opción si desea ver el número de entidades más amplio posible en el repositorio con funciones compartidas. Este es el criterio de coincidencia menos preciso y con el que se seleccionan resultados en más registros. Esta opción devuelve coincidencias exactas y entidades que comparten casi cualquier función (normalmente aquellas con un valor de frecuencia de Uno o Pocos). Por ejemplo, se incluirán las entidades con el mismo número de identificación fiscal y las entidades con direcciones similares.

Fuentes de datos. Esta tabla enumera las etiquetas de fuente de diferentes fuentes de datos que se han cargado en el repositorio, junto con el número de registros en cada fuente.

Resultados del nodo Generación de análisis de entidad

Los resultados del nodo Generación de análisis de entidad constan de los siguientes campos para cada registro que se recupera.

Campo	Descripción
<i>Campo1</i> [, <i>Campo2</i> [, ... <i>CampoN</i>]]	Campos de la fuente de datos que contiene los nuevos registros.
\$EA-ID	Identificador de entidad para este registro en el repositorio.
\$EA-SC	Campo que indica la cercanía de la coincidencia entre este registro y una entidad observada en el repositorio; un valor de 1,0 (mala coincidencia) a 10,0 (buena coincidencia).
\$EA-SRC	Etiqueta de fuente que identifica la fuente de datos donde se creó este registro.
\$EA-KEY	Valor de la clave única de este registro en el archivo de fuente de datos.
\$EA-Función1[, \$EA-Función2[, ... \$EA-FunciónN]]	Valores de las funciones asignadas para este registro en el repositorio.

Uso de IBM SPSS Modeler Entity Analytics con otros productos IBM SPSS

Los instaladores pueden permitirle utilizar IBM SPSS Modeler Entity Analytics con los siguientes productos:

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch para Windows
- IBM SPSS Modeler Solution Publisher

Deberá ejecutar estos instaladores antes de poder utilizar las funciones de IBM SPSS Modeler Entity Analytics con estos productos. Si desea obtener más información, consulte la guía de *instalación de IBM SPSS Modeler Premium*.

Tareas administrativas

Para los repositorios creados en Entity Analytics, se crea un nuevo servicio de base de datos usando el producto IBM solidDB. Hay algunas tareas administrativas asociadas a solidDB; estas tareas las realiza normalmente el administrador de la base de datos o del sistema. Son las siguientes:

- Configuración de asignaciones de puertos
- Gestión de las credenciales de administrador para la base de datos del repositorio

Otras tareas administrativas que quizás deban aplicarse a todos los repositorios son las siguientes:

- Traslado del directorio de almacenamiento del repositorio
- Configuración de opciones de ruta para campos de fecha/hora y marca de tiempo
- Ejecución de IBM SPSS Modeler Entity Analytics con el cliente de SPSS Modeler y SPSS Modeler Server en el mismo sistema Windows

- Purga de un repositorio de entidad
- Eliminación de un repositorio de entidad
- Eliminación de un repositorio cuando no se puede conectar a él

Configuración de asignaciones de puertos

A cada servicio de base de datos solidDB se le debe asignar un puerto que no se puede asignar a otros servicios que se ejecuten en la máquina. Los servicios de bases de datos residen en la misma máquina en la que se ejecuta IBM® SPSS® Modeler Server (o, si IBM® SPSS® Modeler se utiliza sin una conexión a SPSS Modeler Server, la máquina en la que se ejecuta IBM® SPSS® Modeler).

Por defecto, Entity Analytics asigna puertos en el rango de 1320 a 1520, comenzando por el puerto 1320 para el primer repositorio creado. En caso de conflicto, puede configurar la asignación de los puertos modificando el archivo: *<ruta de instalación del servidor de modeler>/ext/bin/pasw.entityanalytics/ea.cfg* y estableciendo valores apropiados para los ajustes *min_port* y *max_port*. Los contenidos por defecto de este archivo se muestran a continuación:

```
# port range configuration for entity analytics

#

# this port range controls which ports SolidDB databases

# (created to store Entity Analytics Repositories in)

# may use. Configure this if the default port range will

# introduce a conflict on your system.

#

# default min_port = 1320

# default max_port = 1520

min_port, 1320

max_port, 1520
```

Gestión de las credenciales de administrador para la base de datos del repositorio

El nombre de usuario y contraseña del administrador para la base de datos solidDB que aloja un repositorio de entidad se definen al crear el repositorio. Si conoce las credenciales actuales, puede cambiarlas mediante el editor de SQL de solidDB.

Para iniciar el editor de SQL de solidDB

1. En un equipo cliente, abra una ventana de símbolo de sistema.
2. Escriba:

```
cd directorio_instalación_modeler\ext\bin\pasw.entityanalytics\solidDB\bin
```

donde *directorio_instalación_modeler* es el directorio de instalación de SPSS Modeler.

3. Escriba:

```
solsql -c "C:\Documents and Settings\All Users\Application  
Data\IBM\SPSS\Modeler\versión\EA\repositories\nombre_repositorio
```

donde *versión* es el número de versión de la instalación de SPSS Modeler y *nombre_repositorio* es el nombre del repositorio.

4. En el mensaje que se muestra, escriba el nombre de usuario y contraseña de administrador de base de datos actuales para que se muestre el símbolo de sistema `solsql>`.

Para cambiar la contraseña de administrador de base de datos

1. En el símbolo de sistema `solsql>`, escriba:

```
alter user nombre_de_usuario identified by contraseña;  
commit work;
```

donde *nombre_de_usuario* es el nombre de usuario del administrador de base de datos actual y *contraseña* es la nueva contraseña.

2. Salga para cerrar el editor.
3. Reinicie el cliente de SPSS Modeler.

Para realizar otras tareas administrativas

Si necesita información para hacer otras tareas administrativas relacionadas con la base de datos solidDB, consulte la documentación para la versión apropiada de IBM solidDB en <http://publib.boulder.ibm.com/>.

Traslado del directorio de almacenamiento del repositorio

De manera predeterminada, los archivos del repositorio se almacenan en un directorio denominado *EA* en las siguientes ubicaciones:

- C:\Documents and Settings\All Users\ApplicationData\IBM\SPSS\Modeler\versión\EA (sistemas Windows)
- *directorio_de_instalación_de_modeler*/ext/bin/pasw.entityanalytics/EA (sistemas UNIX)

Como los archivos utilizados para almacenar el repositorio pueden llegar a ser muy grandes, puede que necesite moverlos a un disco o partición diferente para tener más espacio disponible. Para ello, haga lo siguiente.

1. Salga de SPSS Modeler.
2. Mueva el directorio *EA* desde su ubicación original (indicada anteriormente) a una nueva ubicación. Por ejemplo, en Windows puede que desee moverlo a una nueva ubicación como *F:\data\EA*.
3. Edite el archivo *<ruta de instalación del servidor de modeler>/ext/bin/pasw.entityanalytics/ea.cfg* para añadir la siguiente opción:

directorio_de_datos_del_repositorio, nueva_ubicación

donde *nueva_ubicación* es el directorio al que ha movido el directorio EA, por ejemplo, *F:\data\EA*.

Configuración de opciones de ruta para campos de fecha/hora y marca de tiempo

Si sus datos de fuente incluyen campos que contienen datos de fecha/hora o marca de tiempo, asegúrese de establecer las opciones de ruta correspondientes con un formato reconocido por IBM SPSS Modeler Entity Analytics.

Para establecer el formato de opción de ruta

1. En el menú principal de SPSS Modeler, seleccione:
Herramientas > Propiedades de ruta > Opciones.
2. Seleccione Fecha/hora.
3. Establezca Formato de fecha como AAAA-MM-DD.
4. Establezca Formato de hora como HH:MM:SS.
5. Pulse en Aceptar.

Ejecución de IBM SPSS Modeler Entity Analytics con el cliente de SPSS Modeler y SPSS Modeler Server en el mismo sistema Windows

Si ha instalado IBM SPSS Modeler Entity Analytics en el cliente de SPSS Modeler y SPSS Modeler Server en el mismo sistema Windows, tanto el cliente como el servidor compartirán el mismo repositorio de manera predeterminada. Si desea que utilicen repositorios separados, deberá editar el archivo de configuración *ea.cfg* en **uno** de los sistemas y configurarlo para que utilice un rango de puertos y una carpeta de repositorio diferentes.

Nota: en particular, si utiliza un cliente de SPSS Modeler de 32 bits y un SPSS Modeler Server de 64 bits (o viceversa), deberá realizar este procedimiento.

1. Abra el archivo *<ruta de instalación [del servidor] de modeler>/ext/bin/pasw.entityanalytics/ea.cfg* para su edición.
2. Cambie los ajustes *min_port* y *max_port* para utilizar puertos diferentes desde el otro sistema. Si desea obtener más información, consulte el tema [Configuración de asignaciones de puertos](#) el p. 41.

3. Cambie el ajuste `repository_data_directory` para utilizar un directorio diferente desde el otro sistema.
4. Guarde y cierre el archivo `ea.cfg`.

Purga de un repositorio de entidad

Si quiere eliminar los registros de datos de un repositorio de entidad, pero mantener la información de configuración, utilice la opción de purga del repositorio.

Para purgar un repositorio

1. Abra un nodo Entity Analytics.
2. Pulse en la lista del repositorio de entidad.
3. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Casos de resolución de entidades.
4. En el cuadro de diálogo Casos de resolución de entidades, pulse en la lista Nombre del repositorio.
5. Seleccione el repositorio que desea purgar.
6. Si no está conectado, introduzca el nombre de usuario de administrador y la contraseña y pulse en Conectar.
7. Cuando el botón Purgar repositorio esté activado, púselo.
8. Pulse en Purgar repositorio en el cuadro de diálogo Purgar fuentes de datos para confirmar la purga del repositorio.

Eliminación de un repositorio de entidad

Si ya no necesita un repositorio, puede eliminarlo completamente.

Precaución: la acción hace exactamente lo que dice. **Esta acción no se puede deshacer.** Si no está seguro, utilice el botón Purgar para eliminar todos los datos de fuente. Así no se eliminar la configuración del repositorio. Si desea obtener más información, consulte el tema [Purga de un repositorio de entidad](#) el p. 44.

Nota: el siguiente procedimiento asume que puede conectarse al repositorio desde SPSS Modeler y que conoce el nombre de usuario y la contraseña de administrador de la base de datos donde se aloja el repositorio. Si no es el caso, siga el procedimiento para eliminar un repositorio cuando no se puede conectar a él. Si desea obtener más información, consulte el tema [Eliminación de un repositorio cuando no se puede conectar a él](#) el p. 45.

Para eliminar un repositorio

1. Abra un nodo Entity Analytics.
2. Pulse en la lista del repositorio de entidad.

3. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Casos de resolución de entidades.
4. En el cuadro de diálogo Casos de resolución de entidades, pulse en la lista Nombre del repositorio.
5. Seleccione el repositorio que desea eliminar.
6. Si no está conectado, introduzca el nombre de usuario de administrador y la contraseña y pulse en Conectar.
7. Cuando el botón Eliminar el repositorio completo esté activado, púselo.
8. Pulse en Eliminar para confirmar la eliminación del repositorio.
9. Pulse en Aceptar para aceptar que se ha eliminado correctamente.

Eliminación de un repositorio cuando no se puede conectar a él

Use el siguiente procedimiento si desea eliminar un repositorio de entidad pero no puede conectarse al mismo, ya sea por problemas de conectividad con SPSS Modeler o porque no recuerda el nombre de usuario o la contraseña.

Realice este procedimiento en el equipo donde se aloja la base de datos del repositorio.

Sistemas de Windows

1. Abra una ventana de símbolo de sistema.
2. Escriba:

```
cd directorio_instalación_modeler  
cd ext\bin\pasw.entityanalytics\tools  
delete_repository.bat nombre_repositorio
```

donde *directorio_instalación_modeler* es el directorio de instalación de SPSS Modeler y *nombre_repositorio* es el nombre del repositorio.

3. Continúe desde “Completar el procedimiento” (más adelante en esta sección).

Sistemas UNIX

1. Abra una shell.
2. Escriba:

```
cd directorio_instalación_servidor_modeler  
cd ext/bin/pasw.entityanalytics/tools  
./delete_repository.sh nombre_repositorio
```

donde *directorio_instalación_servidor_modeler* es el directorio de instalación de SPSS Modeler Server y *nombre_repositorio* es el nombre del repositorio.

Completar el procedimiento (todos los sistemas)

1. En el mensaje que se muestra, confirme la eliminación del repositorio especificando Y.

2. Una vez eliminado el repositorio, verá el mensaje:

Info - please remove repository files from the following directory:

(Información: elimine los archivos del repositorio del siguiente directorio:)

ruta_directorio

(Nota: puede que sea necesario reiniciar primero antes de eliminar los archivos del repositorio)

3. Elimine el directorio que tenga el mismo nombre que el repositorio que ha eliminado. Si no puede eliminar el directorio, reinicie el equipo e inténtelo de nuevo.

El análisis de entidades en acción

Acerca de este ejemplo

En este ejemplo veremos cómo el análisis de entidades puede mejorar aún más los resultados de por sí impresionantes de IBM® SPSS® Modeler.

Este ejemplo usa la ruta *loan_entity_analytics.str*, que hace referencia al archivo de datos *loan_applications.csv*. Estos archivos están disponibles en el directorio *Demos* de cualquier instalación de SPSS Modeler que tenga instalado también IBM® SPSS® Modeler Entity Analytics. Puede acceder al directorio *Demos* desde el grupo de programas de SPSS Modeler en el menú Inicio de Windows. El archivo *loan_entity_analytics.str* se encuentra en el directorio *Entity_Analytics*.

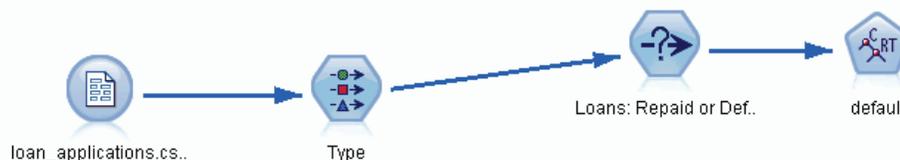
Nota: Antes de que pueda ejecutar esta ruta de ejemplo, necesita crear un repositorio en su sistema. Hágalo antes de continuar con este ejemplo. Si desea obtener más información, consulte el tema [Creación del repositorio](#) en el capítulo 3 el p. 16.

Comencemos por una situación conocida: a los ejecutivos de un banco les interesa saber si es probable que los clientes con una solicitud de préstamo pendiente no puedan pagar dichos préstamos. El departamento de TI del banco utiliza SPSS Modeler desde hace tiempo, por lo que su personal ya ha creado una ruta y ha generado un modelo predictivo a partir de los datos existentes acerca de 700 préstamos concedidos en el pasado. Puede que estos préstamos hayan sido devueltos o no por los clientes.

El modelo original

Así generó el personal del banco su modelo y ésta es la información que obtuvieron.

Figura 4-1
Ruta inicial con nodo de modelado

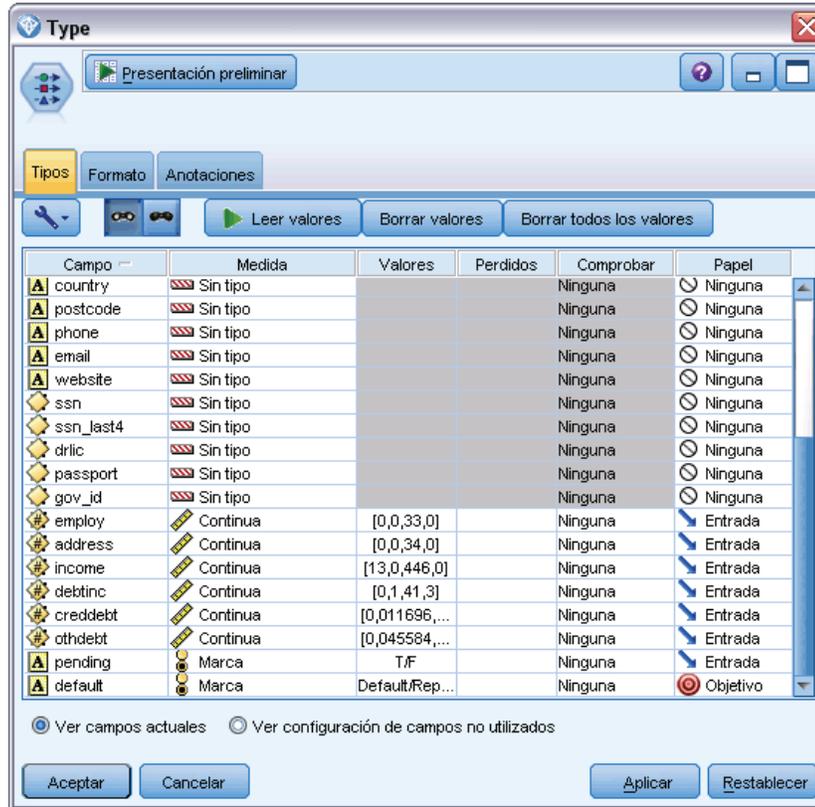


Además de detalles de préstamos anteriores, el conjunto de datos *loan_applications.csv* incluye detalles de 150 clientes cuyas solicitudes de préstamos aún están pendientes, lo que daría un total de 850 registros.

No todos los campos del conjunto de datos resultan útiles en la predicción (los campos de nombre, por ejemplo, se pueden ignorar). El nodo Tipo filtra los campos que se deben ignorar estableciendo su rol como Ninguno. Los campos que se van a usar en la predicción tienen su rol

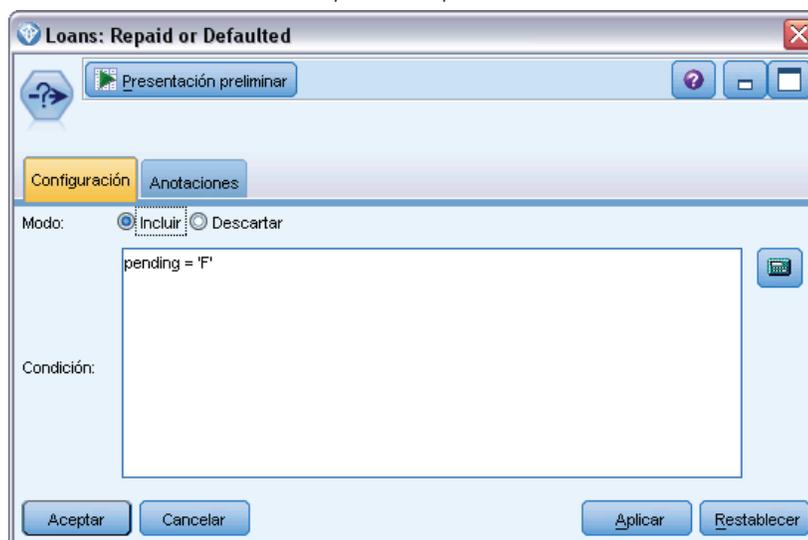
establecido como Entrada, y el campo cuyo valor está tratando de predecir el modelo tiene el rol establecido como Objetivo.

Figura 4-2
Roles de los campos establecidos en el nodo Tipo



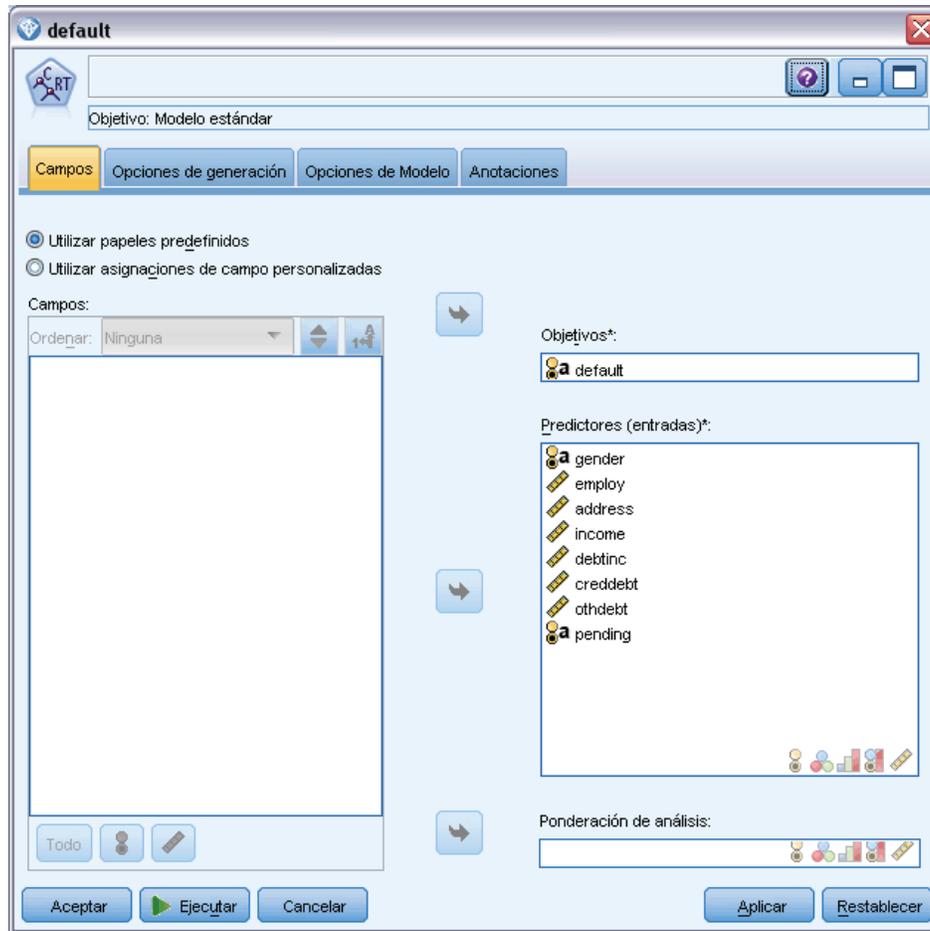
Dado que el modelo debe realizar las predicciones basándose tan solo en datos pasados, la ruta incluye un nodo Seleccionar que incluye solo los préstamos *no* marcados como Pendientes, y por tanto descarta los 150 préstamos pendientes.

Figura 4-3
Descarte de las solicitudes de préstamo pendientes



Con los préstamos pendientes descartados, solo los 700 préstamos restantes, tanto los que se devolvieron como los que no, pasaron al nodo de modelado. El banco podría haber utilizado varios algoritmos de SPSS Modeler para generar un buen modelo. En este caso, han utilizado un nodo Árbol C&R, que se empleará para generar un modelo que prediga quiénes son posibles morosos basándose en el comportamiento anterior de los clientes del banco.

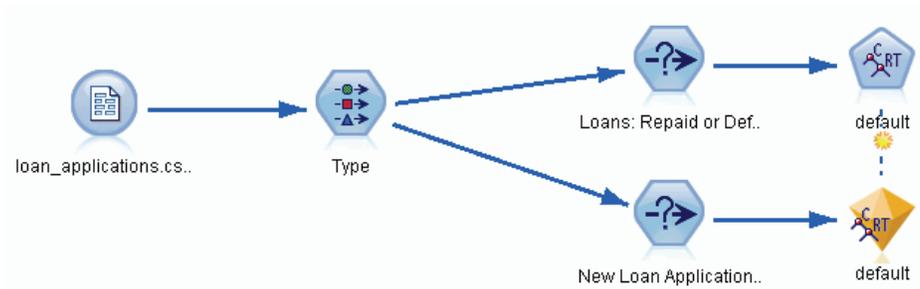
Figura 4-4
Asignación de campos objetivos y predictores



Los campos usados para hacer la predicción son designados como campos predictores, y el campo cuyo valor intenta predecir el modelo, impago en este caso, se establece como campo objetivo, como se definieron anteriormente en el nodo Tipo.

Al ejecutar esta ruta se genera un nugget de modelo, que contiene el modelo creado a partir de los campos predictores.

Figura 4-5
Ruta con nugget de modelo añadido



Ahora, el analista del banco puede utilizar el modelo para comenzar a predecir si es probable que los clientes con pagos pendientes no paguen. Utilizando el conjunto de datos original, el analista inserta un nodo Seleccionar que esta vez incluye solamente los 150 registros de préstamos marcados como pendientes, en lugar de descartarlos. El analista pasa estos registros directamente al modelo, y añade un nodo Distribución que representa visualmente las predicciones del modelo.

Figura 4-6

Ruta de selección de solicitudes de préstamos nuevas con nodo Distribución añadido



El nodo Distribución muestra la distribución de los valores del campo *\$R-default* en el modelo. Al ejecutarlo, el nodo Árbol C&R añade este campo al modelo de datos. El campo contiene la predicción sobre si el nuevo solicitante devolverá o no el préstamo, y posteriormente utilizaremos este campo para comparar el efecto de añadir nuevos análisis de entidades.

Al ejecutar esta parte de la ruta, los resultados del nodo Distribución indican al analista que es previsible que 137 de los 150 solicitantes nuevos devuelvan sus préstamos. Se prevé que los 13 restantes no pagarán sus préstamos, por lo que el analista probablemente recomendará al banco que rechace sus solicitudes.

Figura 4-7

Resultados del nodo Distribución sin análisis de entidades

Distribución de \$R-default			
Valor	Proporción	%	Recuento
Default		8.67	13
Repaid		91.33	137

Con análisis de entidades

Veamos a continuación si la situación mejora añadiendo el análisis de entidades a la ecuación. Imagine que usted es un especialista en análisis de entidades, y que el banco ha contratado sus servicios para investigar posibles entradas fraudulentas en el registro de clientes de los datos de fuente. Puede que haya registros duplicados por errores en la entrada de datos, pero también es posible que el solicitante de un préstamo esté intentando falsear su identidad. En cualquier caso, el banco debe saber la verdad.

Para el propósito de este ejemplo, asumiremos que ya se ha creado un repositorio de entidad. Si desea obtener más información, consulte el tema [Creación del repositorio](#) en el capítulo 3 el p. 16.

Volcado de los datos de fuente en el repositorio

En primer lugar, añada un nodo de exportación de EA al nodo de fuente de datos para que pueda exportar los datos de fuente al repositorio de la entidad.

Figura 4-8

Asignación de un nodo de exportación de EA al nodo de fuente de datos



Antes de poder exportar los datos, debe asignar los campos de la fuente de datos a funciones del repositorio de la entidad. Esto es necesario porque diferentes fuentes de datos pueden utilizar nombres de campos distintos para el mismo tipo de información. El repositorio de entidad proporciona un conjunto estándar de tipos de información (conocido como “funciones”) para evitar duplicidades.

Figura 4-9

Asignación de campos a funciones

Campo	Vinculado a función	Uso
key		(Auto)
fname	NAME.GIVEN_NAME	(Auto)
mname	NAME.MIDDLE_NAME	(Auto)
lname	NAME.SUR_NAME	(Auto)
generation	NAME.NAME_GEN	(Auto)
dob	DOB.DOB	(Auto)
dod		(Auto)
gender	GENDER.GENDER	(Auto)
addr1	ADDRESS.ADDR1	(Auto)

En el nodo de exportación de EA, configure los detalles acerca del repositorio: los detalles de conexión, la etiqueta de fuente (para identificar la fuente de datos: TEST en este caso), el tipo de entidad (el conjunto de funciones que estamos utilizando, denominado PERSON) y el campo de clave única (para identificar cada registro de forma exclusiva). En este caso, utilice el campo clave como clave única.

Ahora puede establecer las asignaciones. En el conjunto de funciones que está utilizando, hay algunas que se corresponden con los campos *nombre*, *segundo nombre*, *apellidos*, *generación*, *dob* (*fecha de nacimiento*), *sexo*, *dirección 1*, *ciudad*, *país*, *código postal*, *teléfono*, *correo electrónico*, *ssn* (*número de la seguridad social*), *drlic* y *pasaporte*.

Establezca en primer lugar la asignación para *nombre*. Pulse dos veces en la columna Vinculado a función en la tabla en la fila *nombre*, desplácese hacia abajo hasta la entrada NAME.GIVEN_NAME y pulse en ella para crear la asignación.

Ahora asigne los campos restantes que tienen funciones correspondientes, de modo que todo el conjunto de asignaciones tenga el siguiente aspecto.

Tabla 4-1
Campos asignados a funciones de repositorio

Campo	Vinculado a función
<i>nombre</i>	NAME.GIVEN_NAME
<i>segundo nombre</i>	NAME.MIDDLE_NAME
<i>apellidos</i>	NAME.SUR_NAME
<i>generación</i>	NAME.NAME_GEN
<i>dob</i> (<i>fecha de nacimiento</i>)	DOB.DOB
<i>sexo</i>	GENDER.GENDER
<i>dirección 1</i>	ADDRESS.ADDR1
<i>ciudad</i>	ADDRESS.CITY
<i>país</i>	ADDRESS.COUNTRY
<i>código postal</i>	ADDRESS.POSTAL_CODE
<i>teléfono</i>	PHONE.PHONE_NUM
<i>correo electrónico</i>	EMAIL_ADDR.ADDR
<i>ssn</i> (<i>número de seguridad social</i>)	SSN.ID_NUM
<i>drlic</i>	DRLIC.ID_NUM
<i>pasaporte</i>	PASSPORT.ID_NUM

Pulse en Ejecutar para exportar los datos al repositorio. Este proceso tardará un poco, así que cuando se cierre el cuadro de diálogo Comentarios de la ejecución, la exportación se habrá completado.

Lectura de las identidades resueltas

Al exportar los datos al repositorio, el sistema de análisis de entidades comienza a resolver posibles conflictos de identidad, asignando un identificador único de entidad, que posteriormente verá como el campo *\$EA-ID*. (Nota: No es el mismo que el campo de clave única en el nodo de exportación de EA; dicho campo se utiliza para identificar de forma exclusiva registros de fuente de datos.)

El primer paso para leer las identidades resueltas consiste en añadir un nodo de fuente Entity Analytics(EA) a la ruta. Este nodo de fuente no debe estar conectado a nada en esta fase.

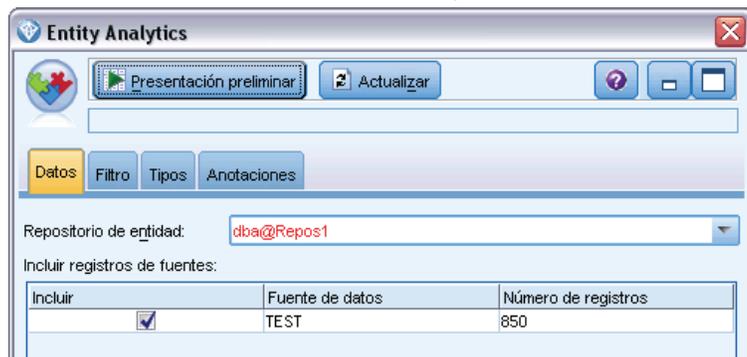
Figura 4-10
Nodo de fuente Entity Analytics(EA)



Entity Analytics(EA)

Abra el nodo de fuente Entity Analytics(EA) y configure los detalles del Repositorio de entidad. Se mostrará una lista con las fuentes de datos exportadas al repositorio: en este caso solo hay una.

Figura 4-11
Selección de una fuente de datos en el repositorio



Marque la casilla de verificación de la fuente de datos TEST y pulse en Aceptar.

Veamos lo que el sistema de análisis de entidades ha hecho con los datos. Adjunte un nodo Tabla al nodo de fuente Entity Analytics(EA), abra el nodo Tabla y pulse en Ejecutar para ver la ventana de resultados del nodo Tabla.

Figura 4-12
Resultados del nodo Tabla

	\$EA-ID	\$EA-SRC	\$EA-KEY
1	1001	TEST	REF00000001
2	1002	TEST	REF00000002
3	1003	TEST	REF00000003
4	1004	TEST	REF00000004
5	1005	TEST	REF00000005
6	1006	TEST	REF00000006
7	1007	TEST	REF00000007
8	1008	TEST	REF00000008
9	1009	TEST	REF00000009
10	1010	TEST	REF00000010
11	1011	TEST	REF00000011
12	1012	TEST	REF00000012
13	1013	TEST	REF00000013
14	1014	TEST	REF00000014
15	1015	TEST	REF00000015
16	1016	TEST	REF00000016
17	1017	TEST	REF00000017
18	1018	TEST	REF00000018
19	1019	TEST	REF00000019
20	1020	TEST	REF00000020

Solo uno de los campos resulta familiar: el que tiene la etiqueta *\$EA-KEY*. En verdad, es el campo *clave* de los datos de fuente, el cual aparece aquí porque usted lo eligió como campo de clave única en el nodo de exportación de EA.

Sin embargo, el sistema ha añadido otros dos campos. El campo *\$EA-ID* es el identificador único, no de los registros de fuente, sino de las identidades resueltas. La diferencia se ve en un instante. El campo *\$EA-SRC* identifica de dónde proceden los datos: pone TEST porque ésa es la etiqueta de fuente que usted le asignó en el nodo de exportación de EA.

¿Qué ha ocurrido con el resto de campos en los datos de fuente? No se preocupe, siguen en el repositorio: lo que ocurre es que, por razones de rendimiento, el nodo de fuente Entity Analytics(EA) pasa solamente el conjunto mínimo de campos hacia abajo para su posterior procesamiento.

A continuación, desplácese hacia abajo en los resultados del nodo Tabla hasta la fila 385.

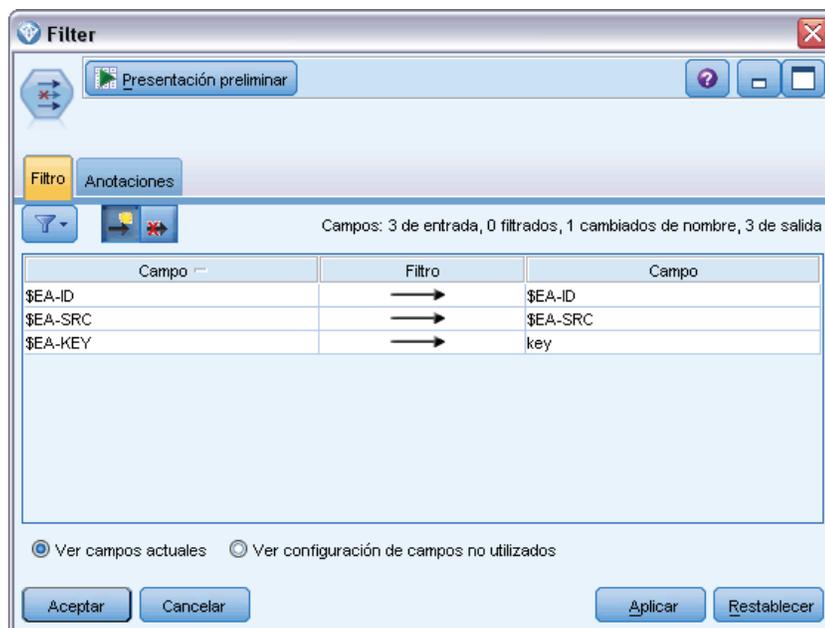
Figura 4-13
Diferencias entre filas de resultados de Tabla y números de \$EA-ID

	\$EA-ID	\$EA-SRC	\$EA-KEY
375	1375	TEST	REF0000375
377	1377	TEST	REF0000377
378	1378	TEST	REF0000378
379	1379	TEST	REF0000379
380	1380	TEST	REF0000380
381	1381	TEST	REF0000381
382	1382	TEST	REF0000382
383	1383	TEST	REF0000383
384	1384	TEST	REF0000384
385	1045	TEST	REF0000385
386	1386	TEST	REF0000386
387	1387	TEST	REF0000387
388	1388	TEST	REF0000388
389	1389	TEST	REF0000389
390	1390	TEST	REF0000390
391	1391	TEST	REF0000861
392	1392	TEST	REF0000392
393	1393	TEST	REF0000393
394	1394	TEST	REF0000394
395	1395	TEST	REF0000395

Observe cómo el número de *\$EA-ID* parece que no pertenece a esta secuencia. El sistema de análisis de entidades ha determinado que el registro REF00000385 hace referencia a la persona identificada como entidad 45, que también tiene el registro REF00000045. Si se sigue bajando en los resultados, hay más números que no siguen la secuencia, por ejemplo en las filas 485, 517, 520, etc. Será mejor que lo analicemos detenidamente.

En primer lugar, destacaremos el hecho de que el campo *\$EA-KEY* contiene los datos del campo *clave* en los datos de fuente, cambiándole el nombre a *clave*. Adjunte un nodo Filtro al nodo de fuente Entity Analytics(EA) y abra el nodo Filtro. Pulse dos veces en la cadena *\$EA-KEY* en la segunda columna Campo y escriba *key*.

Figura 4-14
Cambio de nombre del campo \$EA-KEY



Pulse en Aceptar para cerrar el nodo Filtro.

Ahora es necesario ordenar los ID de entidad de *\$EA-ID* en orden ascendente. Conecte un nodo Clasificar al nodo Filtro. Abra el nodo Clasificar, pulse en el botón superior junto a la tabla Ordenar por, seleccione *\$EA-ID* y pulse en Aceptar.

Figura 4-15
Clasificación de los ID de entidad en orden ascendente



Deje el orden de clasificación como Ascendente y pulse en Aceptar.

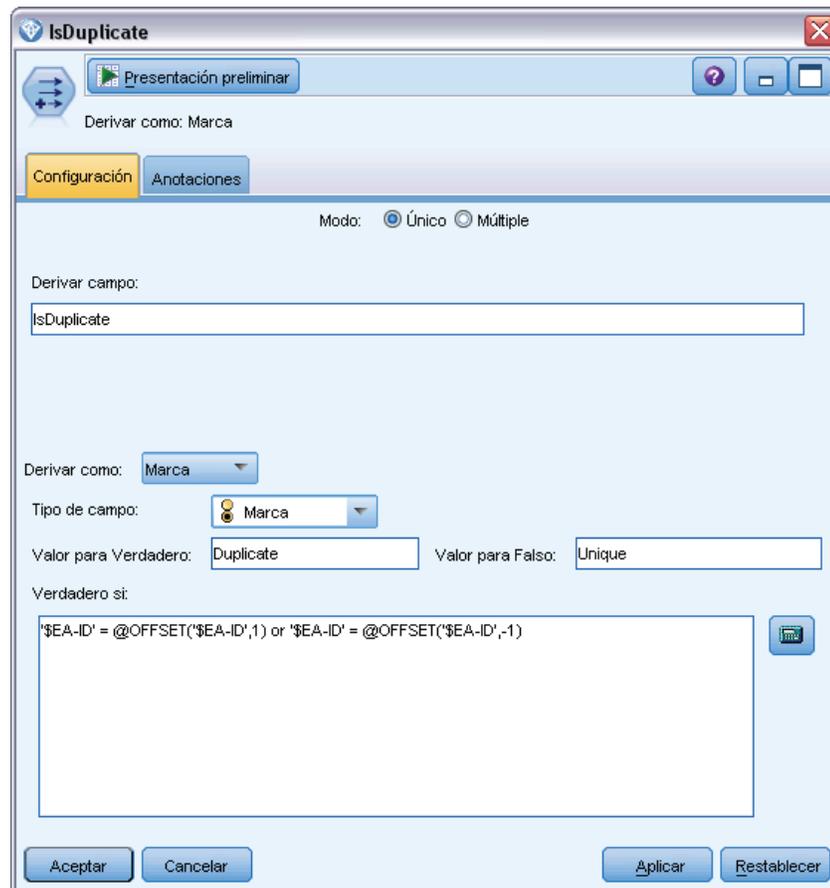
Ahora necesita crear un campo adicional que indique si un registro es único o está duplicado. Conecte un nodo Derivar al nodo Clasificar. Abra el nodo Derivar y establezca el nombre del campo Derivar en `IsDuplicate`. En la lista Derivar como, elija `Marca`, que además establece el tipo de Campo como `Marca`. Establezca el campo del valor Verdadero como `Duplicado` y el campo del valor Falso como `Único`.

Para encontrar registros duplicados, utilizará una función de secuencia especial denominada `@OFFSET`, que acompaña a SPSS Modeler.

Escriba lo siguiente en el campo If:

```
'$EA-ID' = @OFFSET('$EA-ID',1) o '$EA-ID' = @OFFSET('$EA-ID',-1))
```

Figura 4-16
Establecimiento de la condición en el nodo Derivar



Una vez clasificados los ID de entidad en orden ascendente, la función `@OFFSET` comprueba si dos ID de entidad adyacentes son idénticos, en cuyo caso los dos registros son duplicados. En ese caso, el valor *Es duplicado* se establece como *Duplicado*; de lo contrario se establece como *Único*.

Pulse en `Aceptar` para cerrar el nodo.

Para ver el efecto del nodo Derivar, adjunte un nodo Tabla al nodo Derivar, abra el nodo Tabla y pulse en Ejecutar. Desplácese hacia abajo en la ventana de resultados del nodo Tabla hasta la fila 45.

Figura 4-17
Resultados del nodo Derivar

	\$EA-ID	\$EA-SRC	key	IsDuplicate
39	1039	TEST	REF00000039	Unique
40	1040	TEST	REF00000040	Unique
41	1041	TEST	REF00000041	Unique
42	1042	TEST	REF00000042	Unique
43	1043	TEST	REF00000043	Unique
44	1044	TEST	REF00000044	Unique
45	1045	TEST	REF00000045	Duplicate
46	1045	TEST	REF00000385	Duplicate
47	1046	TEST	REF00000046	Unique
48	1047	TEST	REF00000047	Unique
49	1048	TEST	REF00000048	Unique
50	1049	TEST	REF00000049	Unique
51	1050	TEST	REF00000050	Unique
52	1051	TEST	REF00000051	Unique
53	1052	TEST	REF00000052	Unique
54	1053	TEST	REF00000053	Unique
55	1054	TEST	REF00000054	Unique
56	1055	TEST	REF00000055	Unique
57	1056	TEST	REF00000056	Unique
58	1057	TEST	REF00000057	Unique

Recuerde que hemos visto los resultados directamente desde el nodo de fuente Entity Analytics(EA). El sistema ya ha identificado que ese registro REF00000385 hace referencia a la misma persona que la entidad 45. Ahora hemos ido un poco más lejos y hemos marcado que los registros REF00000045 y REF00000385 son duplicados, ya que ambos hacen referencia a la entidad 45.

Siga bajando en la ventana de resultados y podrá ir viendo el resto de registros marcados como duplicados.

Para obtener un informe que enumere los registros duplicados, conecte un nodo Informe (en la pestaña Resultados de la paleta de nodos) al nodo Derivar *Es duplicado*. Abra el nodo Informe, copie el siguiente texto en el campo de entrada de la pestaña Plantilla y pulse en Ejecutar.

```
<html>
<h1>Lista de registros de clientes duplicados.

<h2>Este informe se generó: [@TODAY]

<h2>tabla de
<Registros duplicados>
<tr>
```

```
<td>ID de entidad</td>
<td>Clave</td>
</tr>

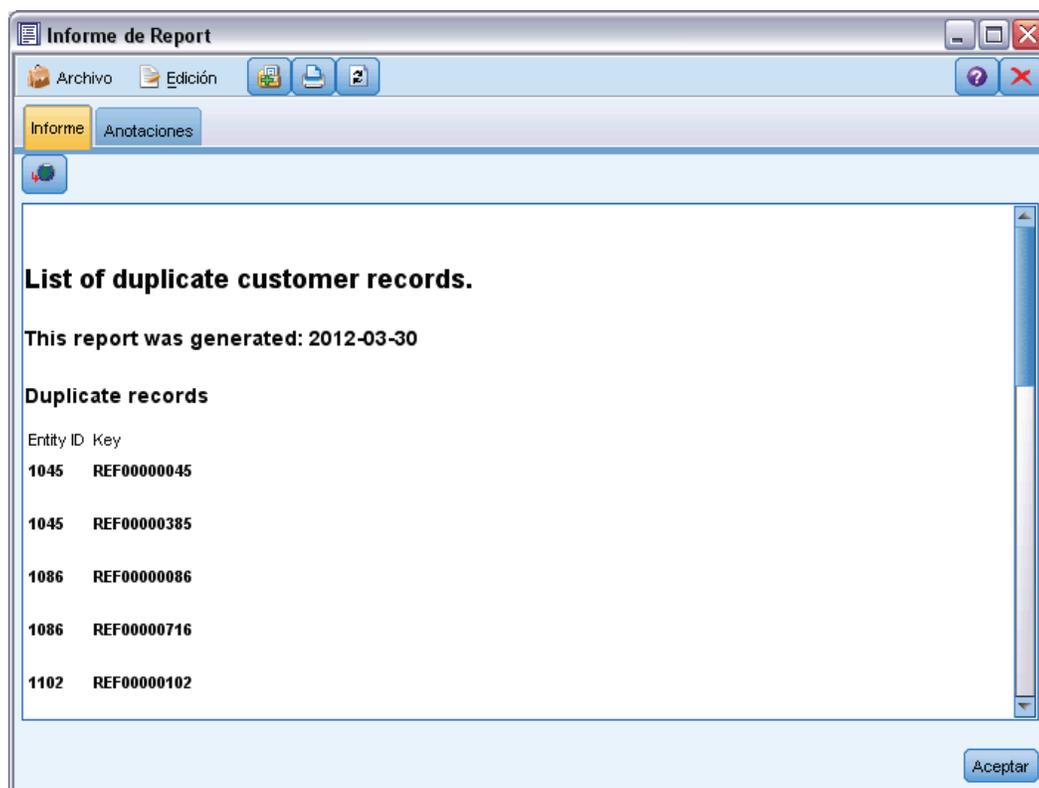
#WHERE IsDuplicate = "Duplicate"

<tr>
<td>['$EA-ID']</td>
<td>[key]</td>
</tr>
#
</table>

</html>
```

Esto genera el siguiente resultado.

Figura 4-18
Resultado del nodo Informe



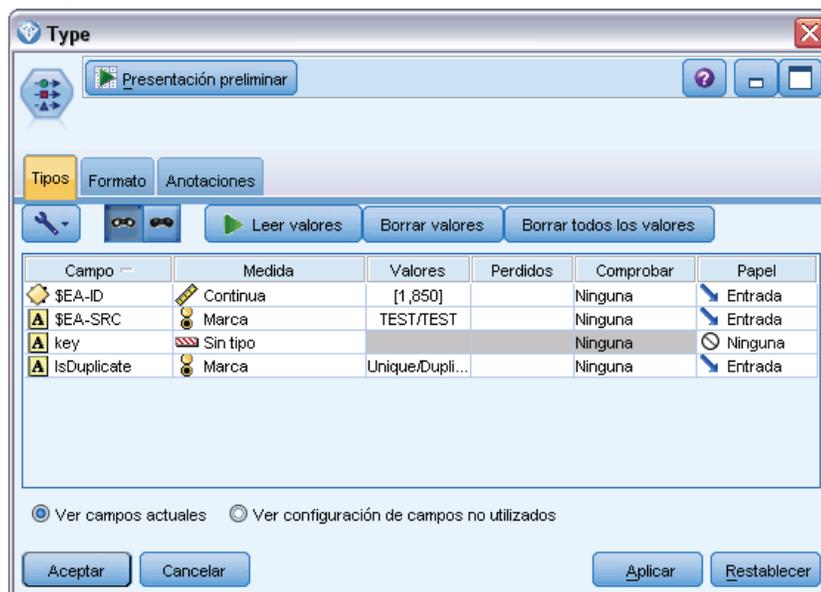
El informe utiliza formato HTML en este caso, aunque también puede utilizar XML o ASCII.

Comparación de los resultados del análisis de entidades con el modelo original

La última fase de este ejemplo consiste en comprobar si al añadir el análisis de entidades se ve afectado el pronóstico original del banco. Recuerde que el modelo original predijo que 13 de los 150 solicitantes no pagarían. Va a utilizar un nodo Fundir para fundir los resultados del modelo con información acerca de los registros duplicados del análisis de entidades y comprobar si cambia la predicción.

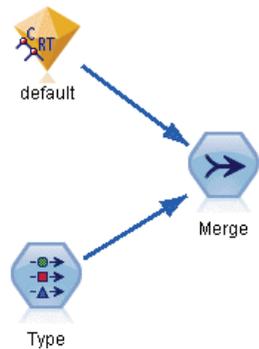
En primer lugar, asegúrese de que los nuevos campos añadidos por el análisis de entidades incluyen los tipos de datos correctos, o los *niveles de medida*, como se conocen en SPSS Modeler. Adjunte un nodo Tipo al nodo Derivar Es duplicado, abra el nodo Tipo y pulse en el botón Leer valores.

Figura 4-19
Configuración del nodo Tipo



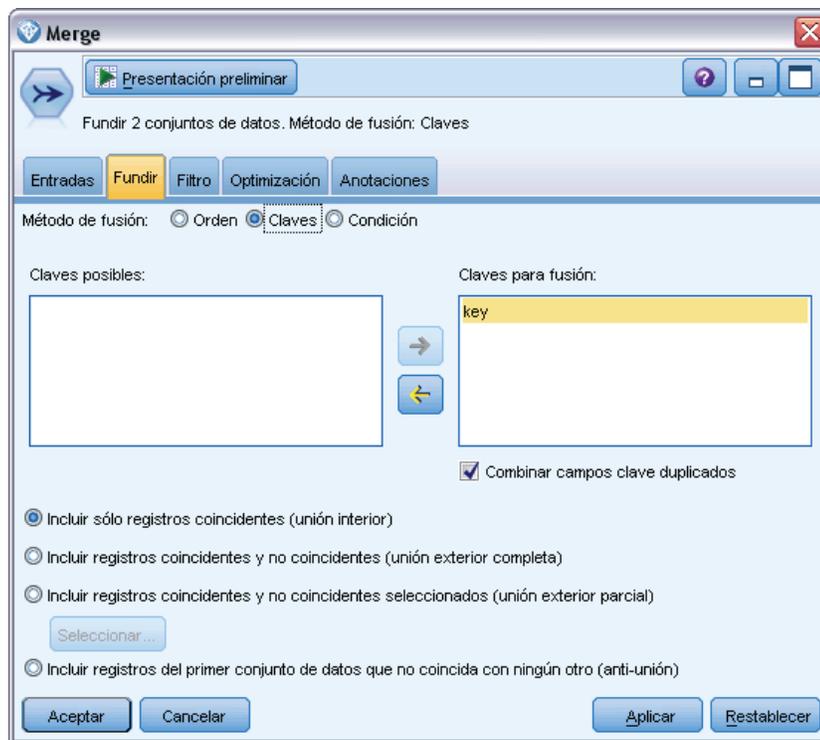
Ahora puede añadir el nodo Fundir. Conéctelo al nodo Tipo y al nugget dorado que contiene el modelo original. Para ello, pulse con el botón derecho en el nugget dorado, elija Conectar y pulse en el nodo Fundir, que ahora debería tener dos flechas de entrada.

Figura 4-20
Entradas al nodo Fundir



Abra el nodo Fundir, en Método de fusión elija Claves, y pulse el botón de la flecha derecha para mover el campo clave de Claves posibles a Claves para fusión y después pulse en Aceptar.

Figura 4-21
Especificación del campo clave para la operación de fusión



Ya está todo prácticamente preparado para realizar la comparación. No obstante, si adjuntase un nodo Distribución y lo ejecutase en este punto, no notaría ninguna diferencia con respecto a la predicción original. Pese a que la ruta funde ahora los resultados del nugget de modelo original con los nuevos campos creados por el análisis de entidades, el propio campo de predicción (*\$R-default*) en el modelo de datos no se ha actualizado con la nueva información.

Para ello, utilice un nodo Rellenar, que puede sustituir valores de campos. Conecte un nodo Rellenar al nodo Fundir y abra el nodo Rellenar.

Pulse en el botón superior a la derecha de Rellenar campos, desplácese hacia la parte inferior de la lista, elija \$R-default y pulse en Aceptar. Este es el campo cuyos valores se cambiarán si se cumple la condición especificada en el resto del cuadro de diálogo.

Para especificar la condición, asegúrese de que Reemplazar está fijado en En función de una condición, y en el campo Condición introduzca:

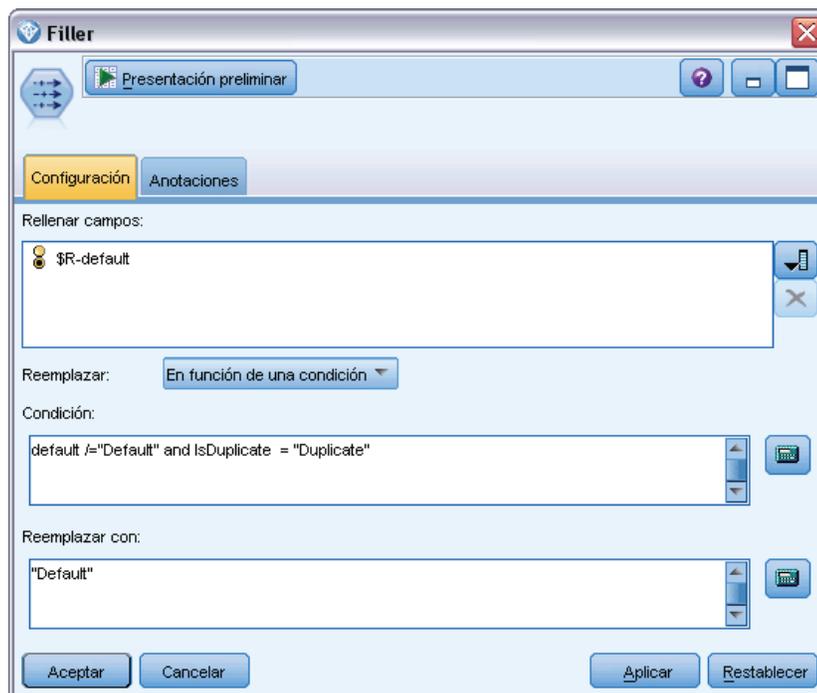
```
default != "default" and IsDuplicate = "Duplicate"
```

En el campo Reemplazar con, introduzca:

```
"default"
```

Figura 4-22

Especificación de la condición para reemplazar valores de campo



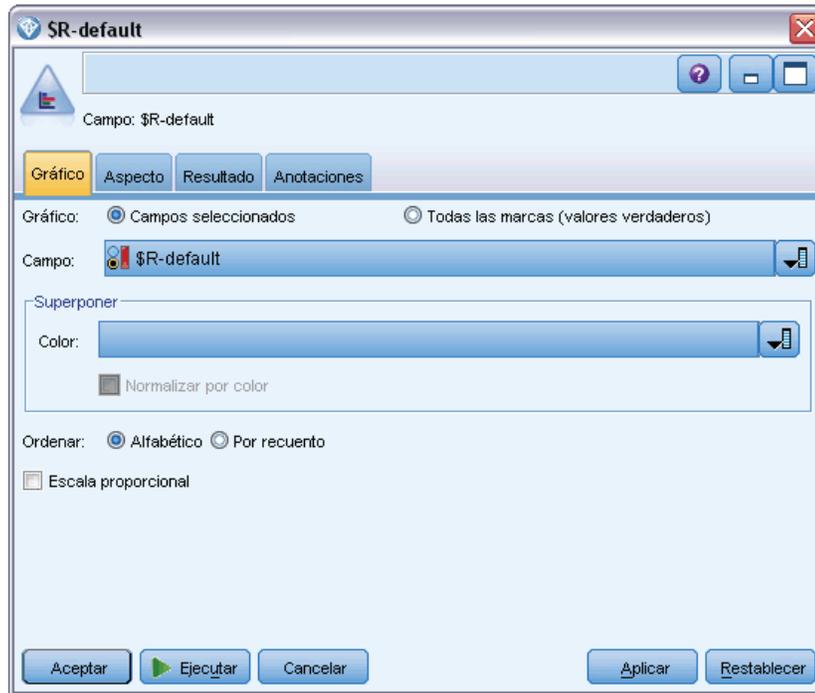
Esta configuración requiere una breve explicación. La condición establece que, en los registros en que el valor del campo *por defecto* del conjunto de datos original no coincida con el valor por defecto y el registro haya sido marcado como duplicado, el valor del campo *\$R-default* en el modelo se establezca por defecto.

El campo *\$R-default* es el campo del modelo que contiene la predicción de si es probable que un cliente no devuelva el préstamo. De este modo, los clientes con registros duplicados se añaden al modelo como morosos en potencia.

Pulse en Aceptar para cerrar el nodo Rellenar.

Finalmente, ya puede ver la diferencia que supone el análisis de entidades. En la paleta Gráficos, conecte un nodo Distribución al nodo Rellenar y abra el nodo Distribución. Pulse en la lista Campo y seleccione \$R-default.

Figura 4-23
Configuración del nodo Distribución



Pulse en Ejecutar para generar el gráfico de la nueva predicción.

Figura 4-24
Resultados del nodo Distribución tras el análisis de entidades



Ahora hay 16 solicitudes de riesgo en lugar de 13. Estas solicitudes adicionales podrían haber resultado muy costosas de no haber pagado, de modo que puede demostrar gráficamente al banco las ventajas de añadir el análisis de entidades a sus operaciones de evaluación de riesgos.

Resumen

Este ejemplo ha mostrado cómo se puede eliminar la duplicidad de registros en los datos de personas u organizaciones mediante el análisis de entidades, lo que proporciona una mejora en la calidad de las predicciones.

Nota: se recomienda eliminar los registros duplicados antes de llevar a cabo cualquier otro proceso. Puede continuar este proceso utilizando el nodo de preparación automática de datos (ADP) para analizar sus datos e identificar los valores fijos, filtrar los campos problemáticos o que no serán útiles, derivar nuevos atributos cuando sea necesario y mejorar el rendimiento mediante técnicas de muestreo inteligente.

La combinación del análisis de entidades y la preparación automática de datos le garantiza que está trabajando con datos lo más limpios posible.

Avisos

Esta información se ha desarrollado para los productos y servicios ofrecidos en todo el mundo.

Puede que IBM no ofrezca los productos, los servicios o las características de los que se habla en este documento en otros países. Consulte a su representante local de IBM para obtener información acerca de los productos y servicios que está disponibles actualmente en su zona. Toda referencia que se haga de un producto, programa o servicio de IBM no implica que sólo se deba utilizar ese producto, programa o servicio de IBM. En su lugar, puede utilizarse todo producto, programa o servicio con funcionalidades equivalentes que no infrinjan los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o aplicaciones de patentes pendientes que cubren el asunto descrito en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, Estados Unidos

Si tiene alguna pregunta sobre la licencia relacionada con la información del juego de caracteres de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus preguntas por escrito a:

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

El párrafo siguiente no se aplica a los Reino Unido o cualquier otro país donde tales disposiciones son incompatibles con la legislación local: INTERNATIONAL BUSINESS MACHINES PROPORCIONA ESTA PUBLICACIÓN “TAL CUAL” SIN GARANTÍA DE NINGÚN TIPO, YA SEA EXPRESA O IMPLÍCITA, INCLUYENDO, PERO NO LIMITADA A, LAS GARANTÍAS IMPLÍCITAS DE NO INFRACCIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN PROPÓSITO PARTICULAR. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones , por lo tanto , esta declaración no se aplique a usted.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente, se efectúan cambios en la información aquí y estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar mejoras y/o cambios en los productos y/o los programas descritos en esta publicación en cualquier momento sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información sólo es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales en esos sitios Web no forman parte del material de este producto de IBM y el uso de estos sitios Web es bajo su propio riesgo.

IBM puede utilizar cualquier información que le suministre en cualquier forma que considere adecuada, sin incurrir en ninguna obligación para usted.

Los licenciatarios de este programa que deseen tener información sobre el mismo con el objetivo de habilitar: (i) el intercambio de información entre programas creados independientemente y otros programas (incluyendo este) y (ii) el uso común de la información que se ha intercambiado, deben ponerse en contacto con:

IBM Software Group, a la atención de: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

IBM proporciona el programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible para el mismo bajo los términos de IBM Customer Agreement (Acuerdo de cliente de IBM), IBM International Program License Agreement (Acuerdo de licencia de programa internacional de IBM) o cualquier acuerdo equivalente entre las partes.

Cualquier dato de rendimiento mencionado aquí ha sido determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en otros entornos operativos pueden variar de forma significativa. Es posible que algunas medidas se hayan realizado en sistemas en desarrollo y no existe ninguna garantía de que estas medidas sean las mismas en los sistemas comerciales. Además, es posible que algunas medidas hayan sido estimadas a través de extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben consultar los datos que corresponden a su entorno específico.

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otras fuentes disponibles públicamente. IBM no ha probado estos productos y no puede confirmar la precisión de su rendimiento, su compatibilidad o cualquier otra reclamación relacionada con productos que no sean de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

Todas las declaraciones sobre el futuro del rumbo y la intención de IBM están sujetas a cambio o retirada sin previo aviso y representan únicamente metas y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos lo máximo posible, los ejemplos incluyen los nombres de las personas, empresas, marcas y productos. Todos esos nombres son ficticios y cualquier parecido con los nombres y direcciones utilizados por una empresa real es pura coincidencia.

Si está viendo esta información en copia electrónica, es posible que las fotografías y las ilustraciones en color no aparezcan.

Marcas comerciales

IBM, el logotipo de IBM, ibm.com y SPSS son marcas comerciales de IBM Corporation, registradas en muchas jurisdicciones de todo el mundo. Existe una lista actualizada de marcas comerciales de IBM en Internet en <http://www.ibm.com/legal/copytrade.shtml>.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas comerciales y los logotipos basados en Java son marcas comerciales de Sun Microsystems, Inc. en Estados Unidos, otros países o ambos.

Otros productos y nombres de servicio pueden ser marcas comerciales de IBM u otras empresas.



Índice

- análisis de entidades
 - comparado con análisis predictivo, 2
 - definido, 1
 - uso con IBM SPSS Modeler, 4
- asignaciones de puertos
 - configuración para análisis de entidades, 41
- avisos legales, 66

- cambio del nombre
 - campos de datos para análisis de entidades, 33
- campos de correspondencia
 - a las funciones del repositorio de entidad, 7, 19–21, 23, 37
- claves únicas
 - análisis de entidades, 7
 - repositorio de entidad, 20
- coincidencia de entidades, establecimiento de umbral, 31
- configuración
 - repositorio de entidad, 22, 31–32

- eliminación
 - repositorio de entidad, 44–45
- etiquetas de fuente
 - repositorio de entidad, 20
- exportación
 - datos a un repositorio de entidad, 8

- fuente de datos, selección para análisis de entidades, 33
- fuentes de datos
 - que conectan con el análisis de entidades, 5, 15
 - visualización del análisis de entidades, 18, 38
- funciones
 - repositorio de entidad, 7, 19–21, 23–24, 27, 37

- identidades resueltas, análisis, 32
- información de tipos, establecimiento para análisis de entidades, 34

- marcas comerciales, 67

- nodos
 - adición a una ruta de análisis de entidades, 34
- nodos de exportación
 - Entity Analytics, 8
- nodos de fuente
 - Entity Analytics, 11

- purga
 - repositorio de entidad, 44

- repositorio
 - administración del análisis de entidades, 40
 - análisis de entidades, 6–8, 14, 16–24, 27–28, 31–32, 37, 44–45
 - repositorio de entidad, 14
 - conexión con IBM SPSS Modeler, 7
 - configuración, 22, 31–32
 - configuración de asignaciones de puertos, 41
 - creación, 6, 14, 16–17
 - eliminación, 44–45
 - funciones, 27
 - mantenimiento, 24
 - opciones, 18
 - purga, 44
 - tareas administrativas, 40
 - resolución de identidades, análisis de entidades, 8
 - resolver identidades, análisis de entidades, 8

- tipos de entidad
 - análisis de entidades, 28
 - repositorio de entidad, 20
- tipos de uso, análisis de entidades, 28