

Guía de aprendizaje de gobierno de datos:

Consumir los datos

La historia de la guía de aprendizaje corresponde a Golden Bank, que tiene varios departamentos que necesitan acceso a datos de alta calidad sobre hipotecas de cliente. Como analista de datos, tendrá que buscar y encontrar los datos correctos, comprender y proteger su contenido y, a continuación, prepararlos para que lo utilicen otros analistas de datos y científicos de datos.

Tarea 1: Comprender los activos de datos

Los activos de datos en los catálogos son mucho más que punteros a los datos. Contienen información sobre el formato y el significado de los datos y estadísticas sobre los valores de datos. Siga estos pasos para comprender el valor de los activos de datos:

1. En el **menú de navegación** , seleccione **Catálogos > Ver todos los catálogos**.
2. Abra el **Catálogo de aprobación de hipotecas**.

La sección de activos destacados muestra los activos **añadidos recientemente** y los activos **muy valorados** que los colaboradores del catálogo han valorado y revisado.

3. Pulse **Ocultar activos destacados** para cerrar esa sección.
4. Busque mortgage.
5. Pulse **HIPOTECAGE_APPLICANTS_TRUST** para ver ese activo de catálogo. La pestaña *Visión general* y el panel lateral proporcionan información básica sobre el activo como, por ejemplo, la descripción, una valoración, etiquetas, dónde se encuentra el activo, términos empresariales, clases de datos y artículos relacionados.
6. Pulse la pestaña **Perfil**. La información de perfil le ayuda a comprender el contenido, la calidad y la usabilidad de los datos.
7. Desplácese a la derecha para localizar la columna **ZIP_CODE** .
8. La clase de datos que se ha asignado automáticamente a la columna **ZIP_CODE** es *Entidad comercial y de gobierno*. Tenga en cuenta que la clase de datos asignada automáticamente puede variar. Puesto que los valores son códigos postales, puede reclasificar fácilmente esta

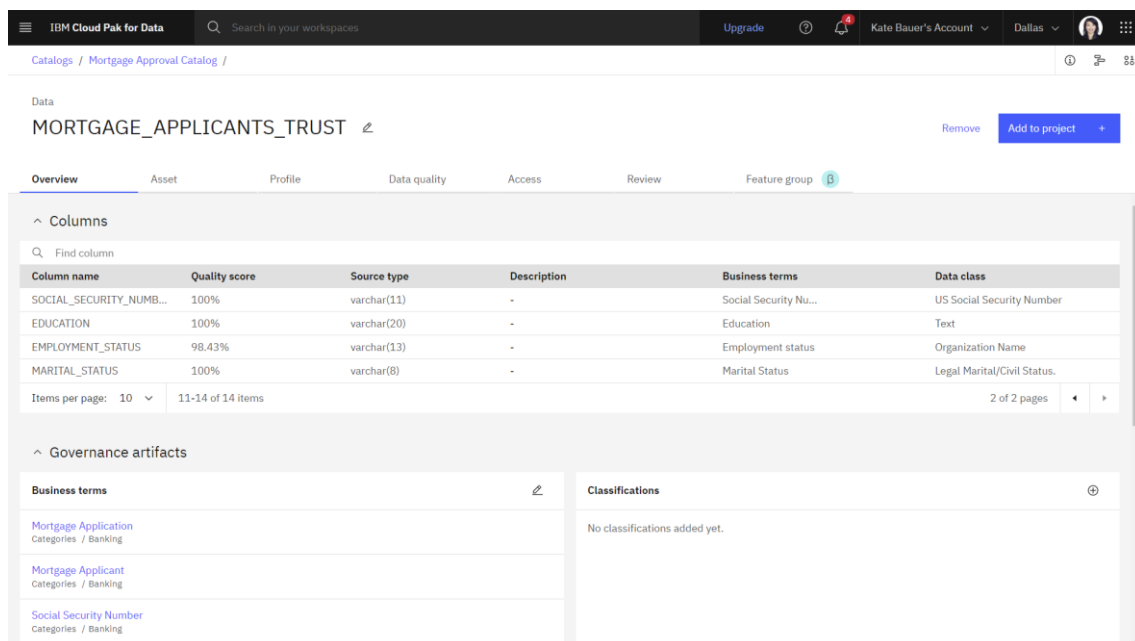
columna. Pulse la lista desplegable para ver otras clases de datos posibles y sus niveles de confianza. Seleccione **Código postal de EE. UU.**.

9. Pulse la pestaña **Activo** para ver una vista previa de los datos.

10. Vuelva a la pestaña **Visión general** para ver más metadatos sobre las columnas. En la lista de columnas, busque la columna *EMPLEMENT_STATUS* para ver los metadatos incluidos los términos empresariales asignados.

Comprueba tu progreso

La imagen siguiente muestra el activo HIPOTECAGE_APPLICANTS_TRUST en el catálogo. Ha explorado el tipo de información que IBM watsonx.data intelligence añade automáticamente a los activos de datos durante el enriquecimiento de metadatos. En la siguiente tarea, enriquecerá manualmente este activo de datos.



The screenshot shows the IBM Cloud Pak for Data interface. The top navigation bar includes the IBM logo, 'IBM Cloud Pak for Data', a search bar, and user account information. The main content area displays the 'MORTGAGE_APPLICANTS_TRUST' data asset. The 'Overview' tab is selected, showing a table of columns with their quality scores and source types. Below the table, there are sections for 'Business terms' and 'Classifications'.

Column name	Quality score	Source type	Description	Business terms	Data class
SOCIAL_SECURITY_NUMB...	100%	varchar(11)	-	Social Security Nu...	US Social Security Number
EDUCATION	100%	varchar(20)	-	Education	Text
EMPLOYMENT_STATUS	98.43%	varchar(13)	-	Employment status	Organization Name
MARITAL_STATUS	100%	varchar(8)	-	Marital Status	Legal Marital/Civil Status.

Items per page: 10 11-14 of 14 items 2 of 2 pages

Business terms

- Mortgage Application Categories / Banking
- Mortgage Applicant Categories / Banking
- Social Security Number Categories / Banking

Classifications

No classifications added yet.

Tarea 2: Enriquecer activos y crear relaciones

Puede hacer que los activos sean más valiosos añadiendo información a los mismos. Por ejemplo, puede añadir su opinión sobre el activo, actualizar propiedades de activo y crear relaciones para enlazar activos. Siga estos pasos para enriquecer activos y crear relaciones:

1. Para el activo de catálogo *MORTGAGE_APPLICANTS_TRUST* , pulse la pestaña **Revisar** . Valorar y comentar este activo para que otros puedan encontrar el activo fácilmente.
 - a. Seleccione **5 estrellas** para la valoración.
 - b. Para la revisión, copie y pegue el texto siguiente:
 - c. This contains high quality customer data from the mortgage system.
 - d. Pulse **Enviar**.
2. Pulse el separador **Visión general** .
3. Haga clic en el icono **Editar** junto al nombre del activo para editar el nombre del activo.
 - a. Cambie el nombre por:
 - b. MORTGAGE_APPLICANTS_TRUST_PROTECT
 - c. Haga clic en **Aplicar**.
4. En la sección *Descripción* del panel lateral derecho, haz clic en el icono **Añadir** .

Nota:

Si este activo tiene una descripción existente, verás un icono **Editar** en lugar de un icono *Añadir*.

- a. Copie y pegue la descripción siguiente:
 - b. Mortgage applicants from the Mortgage System
 - c. Haga clic en **Aplicar**.
5. Dado que este activo está relacionado con préstamos hipotecarios, junto a «*Condiciones comerciales*», haga clic en el icono **«Añadir»** o en el icono **«Editar»**.
 - a. En el campo *Buscar* , escriba loan.

Nota: No es necesario pulsar Intro después de escribir el término de búsqueda. Verá una lista de resultados inmediatamente después de escribir el término de búsqueda.

- b. Seleccione **Préstamo**.
- c. Pulse **Guardar**.

6. Dado que este activo contiene información personal, junto a *Clasificaciones*, haga clic en el icono **Añadir** o en el icono **Editar** .
 - a. Seleccione **Información de identificación personal**.
 - b. Pulse **Guardar**.
7. Puesto que este activo está relacionado con otros activos de hipoteca, junto a *Elementos relacionados*, pulse **Añadir elementos relacionados > Añadir activos relacionados**.
 - a. Seleccione **Está relacionado con** y pulse **Siguiente**.
 - b. Seleccione los activos **CREDIT_SCORE** y **MORTGAGE_APPLICATION** y pulse **Añadir**.
8. Pulse **MORTGAGE_APPLICATION** para ver ese activo relacionado.

Comprueba tu progreso

La imagen siguiente muestra la pestaña Visión general para el activo HIPOTECAGE_APPLICANTS_TRUST_PROTECT en el catálogo. Ha hecho que estos activos sean más valiosos revisando, actualizando propiedades y añadiendo relaciones a los activos. En la siguiente tarea, añadirá el activo enriquecido a un proyecto.

The screenshot shows the IBM Cloud Pak for Data interface. The main header includes the IBM Cloud Pak for Data logo, a search bar, and navigation links like Buy, Renovations, Dallas, and DM. The breadcrumb trail indicates the current location: Catalogs / Mortgage Approval Catalog. The asset name is MORTGAGE_APPLICANTS_TRUST_PROTECT. The Overview tab is selected, showing governance artifacts and related assets. The related assets table lists MORTGAGE_APPLICATION and CREDIT_SCORE. The right sidebar provides detailed information about the asset, including its description, owner, privacy settings, format, and source.

Relationship	Asset name	Workspace	Asset type
Is related to	MORTGAGE_APPLICATION	Catalog \ Mortgage Approval Catalog	Data
Is related to	CREDIT_SCORE	Catalog \ Mortgage Approval Catalog	Data

Tarea 3: Añadir datos enriquecidos a un proyecto

El equipo de analistas de datos necesita los datos de los solicitantes de hipoteca en el proyecto de análisis de hipoteca para refinar, visualizar, analizar y utilizar como datos de entrenamiento para modelos. Siga estos pasos para añadir los datos enriquecidos a un proyecto:

1. Haga clic en **Catálogo de aprobación de hipotecas** en la ruta de navegación.

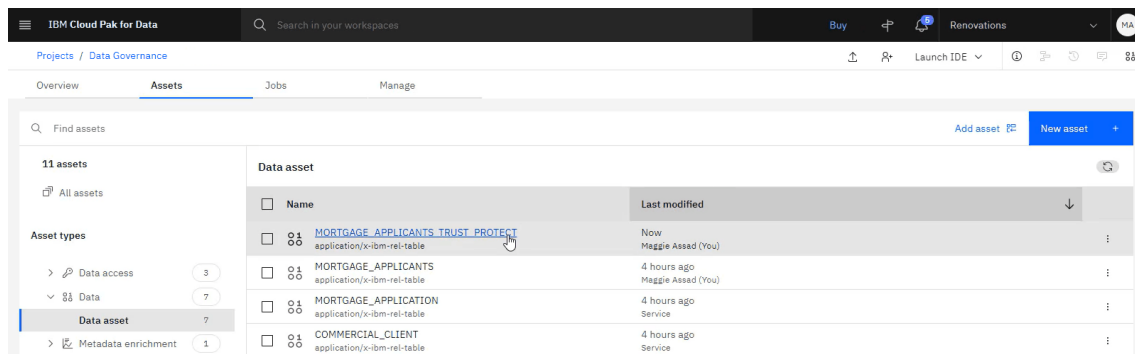


[Catalogs](#) / [Mortgage Approval Catalog](#)

2. Al final de la fila del activo del catálogo *MORTGAGE_APPLICANTS_TRUST_PROTECT*, haga clic en el menú **Desbordamiento** y seleccione **Añadir al proyecto**.
 - a. En la lista desplegable *Destino*, seleccione el proyecto **Gobierno de datos**.
 - b. Pulse **Añadir**.
3. Cuando se visualice la notificación, pulse **Ir al proyecto**. Si omite la notificación, entonces:
 - a. Haga clic en el **menú de navegación** y seleccione **Proyectos > Ver todos los proyectos**.
 - b. Pulse el proyecto **Gobierno de datos**.
4. En el proyecto, pulse el separador **Activos** para ver el activo de datos *HIPOTECAGE_APPLICANTS_TRUST_PROTECT*.

Comprueba tu progreso

La imagen siguiente muestra el activo *HIPOTECAGE_APPLICANTS_TRUST_PROTECT* en el proyecto. Ahora está preparado para visualizar los datos.



Tarea 4: Visualizar los datos

Es necesario limpiar y refinar los datos de los solicitantes de hipoteca para que estén listos para sus herramientas y modelos analíticos. Una forma rápida y fácil de determinar cómo debe configurarse es visualizar los datos en Data Refinery. La visualización se basa en las primeras 5.000 filas de los datos. Siga estos pasos para visualizar los datos:

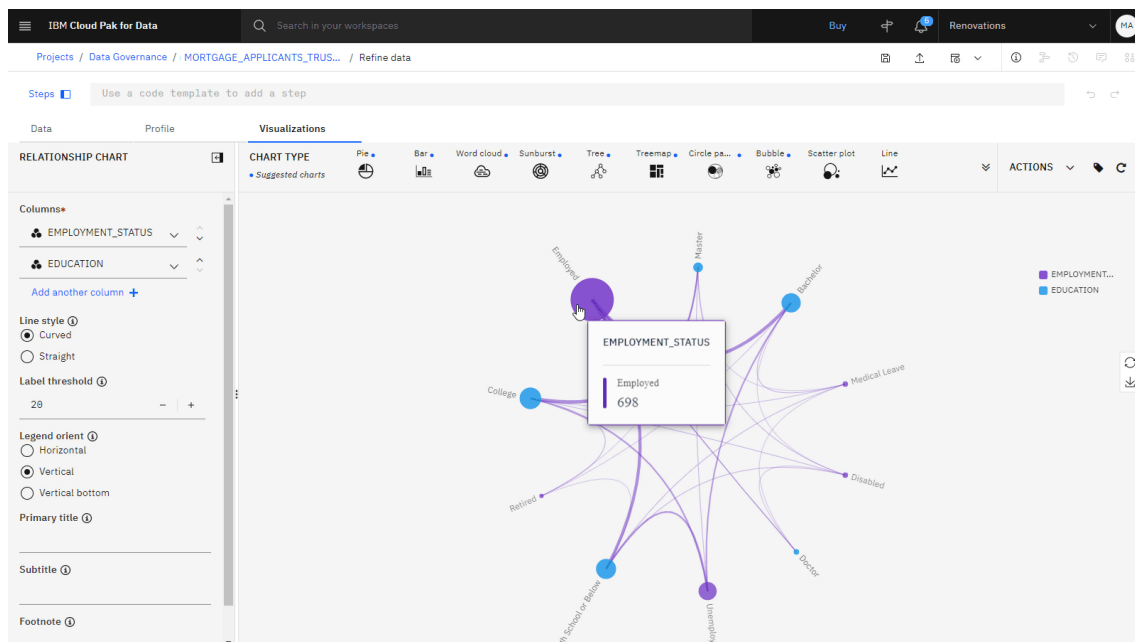
1. Pulse el activo de datos **HIPOTECAGE_APPLICANTS_TRUST_PROTECT** para obtener una vista previa de los datos.
2. Pulse **Preparar datos** para abrir el activo de datos en Data Refinery y espere a que los datos se lean y procesen.
3. En el panel *Acerca de este activo*, pulse la **X** para cerrar el panel.
4. En el panel *Pasos*, pulse la **X** para cerrar el panel.
5. Pulse la pestaña **Visualizaciones**.
6. Para la *Columna a visualizar*, seleccione **EMPLEMENT_STATUS**.
7. Pulse **Visualizar datos**. La herramienta selecciona un gráfico circular como el mejor tipo de gráfico para esta columna, que muestra la distribución de solicitantes por estado de empleo. Observe los tipos de gráfico sugeridos que se indican mediante un punto azul junto a la barra, la nube de palabras y la ráfaga solar.
8. Para el *Tipo de gráfico*, seleccione el tipo de gráfico **Burbuja**. El gráfico de burbujas es una forma fácil de visualizar rápidamente la distribución de valores en un conjunto de datos determinado.
9. En el desplegable *Tipo de gráfico*, seleccione el tipo de gráfico **Relación**.
10. Este tipo de gráfico necesita dos columnas. Seleccione estas columnas:

- a. Para la primera columna, seleccione **EMPLEMENT_STATUS**.
- b. Pulse **Añadir otra columna**.
- c. Para la segunda *Columna*, seleccione **EDUCATION**.

11. Con el gráfico *Relación* , puede seleccionar puntos finales para ver las relaciones. Por ejemplo, puede ver el estado de empleo de los solicitantes por nivel de educación.

Comprueba tu progreso

La imagen siguiente muestra el activo HIPOTECAGE_APPLICANTS_TRUST_PROTECT visualizado en Data Refinery. Ahora está preparado para limpiar los datos.



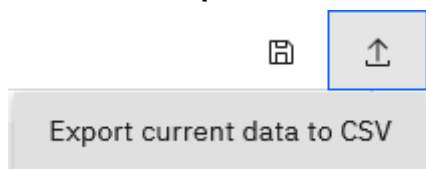
Tarea 5: Preparar los datos para la analítica y la inteligencia artificial

No puede procesar a los solicitantes sin un número de seguridad social, por lo que necesita revisar los datos y eliminar a los solicitantes sin números de seguridad social. Para preparar los datos de HIPOTECAGE_APPLICANTS_TRUST_PROTECT, deberá:

- Vea la frecuencia de valores en la columna *Social_Security_Number*.
- Filtre los candidatos con valores perdidos de la columna *Social_Security_Number*.

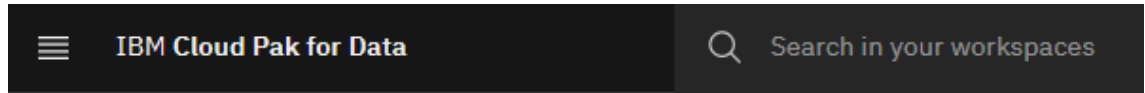
Siga estos pasos para preparar los datos:

1. En Data Refinery, pulse el separador **Perfil** .
2. Desplácese a la derecha para localizar la columna *Social_Security_Number* . Observe que faltan varios valores.
3. Pulse la pestaña **Datos** para filtrar estos registros. En la barra de estado de la parte inferior de la pantalla, Data Refinery indica que el *CONJUNTO DE DATOS COMPLETO* es de 1101 filas.
4. Si el panel *Pasos* no está visible, pulse **Pasos** para abrir el panel.
5. Pulse **Paso nuevo**.
 - a. En la sección *Borrar* , seleccione **Filtro**.
 - b. En el campo *Columna* , seleccione la columna **Social_Security_Number** .
 - c. En el campo *Operador* , seleccione **No está vacío**.
 - d. Haga clic en **Aplicar**. Observe que en la barra de estado de la parte inferior de la pantalla, Data Refinery ahora indica que *FULL DATA SET* es de 1000 filas porque las filas con números de seguridad social que faltan se filtran. Observe que se muestra un paso nuevo en el panel *Pasos* que muestra la operación *Filtro* .
6. Pulse la pestaña **Perfil**.
7. Desplácese a la derecha para localizar la columna *Social_Security_Number* . Observe que los valores perdidos han desaparecido.
8. En la barra de herramientas, haz clic en el icono **Guardar**.
9. En la barra de herramientas, haga clic en el icono **Exportar** y seleccione **Exportar datos actuales a CSV**.



- a. Guarde **MORTGAGE_APPLICANTS_TRUST_PROTECT_shaped.csv** en una carpeta local.
- b. Vaya a esa carpeta y abra el archivo CSV, que contiene 1000 filas y a ningún candidato le falta el número de la seguridad social.

10. Vuelve a Cloud Pak for Data y haz clic en el proyecto **de gobernanza de datos** en la ruta de navegación.



Projects / Data Governance

11. Pulse **Todos los activos** y localice el nuevo activo de flujo Data Refinery con el nombre **MORTGAGE_APPLICANTS_TRUST_PROTECT_flow**.

Sugerencia: Puede guardar el conjunto de datos refinado en el proyecto o en un origen de datos externo, como por ejemplo la instancia de Db2 Warehouse donde se almacenan los conjuntos de datos originales. Para obtener más información, consulte [Creación de trabajos en Data Refinery](#).

Comprueba tu progreso

La imagen siguiente muestra el archivo **MORTGAGE_APPLICANTS_TRUST_PROTECT_shaped.csv** que ha refinado en Data Refinery. Este conjunto de datos contiene la información sobre los solicitantes de hipoteca que proporcionaron un número de seguridad social.

ID	NAME	STREET_ADDRESS	CITY	STATE	STATE_ZIP	EMAIL_ADDRESS	PHONE_NUMBER	GENDER	SOCIAL_SECURITY	EDUCATION	EMPLOYMENT	MARITAL_STATUS
1	Hillary Murphy	4356 Buckland Hills Dr	North Hills	California	CA	43485.Hillary.Murphy@lion.ca	937-255-8778	Female	457-12-7640	Doctor	Disabled	Divorced
2	Branson Crist	6095 Connecticut Avenue	Tahoe City	California	CA	43251.Branson.Crist@juanita.me	845-512-1137	Male	591-81-1843	High School or Below	Unemployed	Married
3	Emma Friesen	639 Toringford Street	Ashburn	Virginia	VA	43016.Emma.Friesen@jace.biz	630-893-4159	Male	479-51-0416	College	Retired	Single
4	Julia Gerhold	3161 Connal Promenade Plwy	Ocoee	Florida	FL	42840.Julia.Gerhold@jace.biz	651-215-0230	Male	386-68-7811	Master	Employed	Married
5	Reese Wunsch	1474 West Bypass	Westwood	New Jersey	NJ	42782.Reese.Wunsch@loraine.ca	251-920-0881	Male	220-76-6498	High School or Below	Employed	Married
6	Zita Keeling	1818 Momeans Ave	Eagle River	Alaska	AK	42606.Zita.Keeling@sidrick.net	541-971-7757	Male	672-65-1323	College	Unemployed	Single
7	Glen Marks	9881 Douglas Avenue	Burlington	Vermont	VT	42171.Glen.Marks@harmony.biz	228-872-5055	Male	408-02-0954	High School or Below	Unemployed	Divorced
8	Rene Leannon	10340 Spring Am SW	Madison	Wisconsin	WI	42137.Rene.Leannon@kevin.info	914-298-2798	Male	367-25-8238	Master	Employed	Married
9	Halle Casin	60509 Aaron Arnone Drive	Oak Lawn	Illinois	IL	41961.Halle.Casin@bertha.co.uk	480-862-8317	Male	104-21-2766	Bachelor	Unemployed	Single
10	Dennis Dooley	930 South McKenzie St	Highlands Ranch	Colorado	CO	41843.Dennis.Dooley@xavier.tv	405-372-7114	Male	303-85-3800	College	Employed	Single
11	Tate Murazik	3525 Fort Morgan Road	Plainview	Texas	TX	41667.Tate.Murazik@eldred.biz	918-292-9609	Male	374-87-7016	High School or Below	Employed	Married
12	Broderick Hauck	270717 Hwy 135	New Rochelle	New York	NY	41609.Broderick.Hauck@floyd.co.uk	714-620-3592	Female	814-45-7554	College	Unemployed	Single
13	Joanie King	60947 Red Farmer Drive	Lakewood	Colorado	CO	41613.Joanie.King@donnell.ca	351-533-1340	Male	217-66-0602	Bachelor	Unemployed	Single
14	Jerrell Eichmann	56421 Hwy 78 East	Reston	Virginia	VA	41158.Jerrell.Eichmann@tricia.biz	972-510-7295	Male	394-17-0450	High School or Below	Unemployed	Married
15	Chance Adams	57785 Rangeline Service Rd	Lakewood	California	CA	40964.Chance.Adams@fatima.me	528-505-4385	Female	690-41-9539	High School or Below	Disabled	Married
16	Drew Keebler	4095 Ann St	Denver	Colorado	CO	40788.Drew.Keebler@loyce.ca	516-688-6066	Female	899-82-0628	High School or Below	Unemployed	Married
17	Rickey Smitham	5138 West Avalon Ave	Myrtle Beach	South Carolina	SC	40729.Rickey.Smitham@yessenia.org	412-340-3823	Female	180-67-1839	Bachelor	Employed	Divorced
18	Matilde Hegmann	3051 Pelham Plwy	Manchester	New Hampshire	NH	40553.Matilde.Hegmann@phyllis.org	405-212-1411	Male	455-97-7384	Bachelor	Employed	Married
19	Sophia Boyer	17951 Hwy 14 East	Webster	Texas	TX	40118.Sophia.Boyer@billie.info	325-212-6229	Male	531-39-8976	College	Employed	Married
20	Estrella Weimann	58360 Us 231 South	Columbus	Ohio	OH	40084.Estrella.Weimann@lon.net	303-898-0485	Male	626-34-0623	High School or Below	Employed	Married
21	Hershel Beer	60909 Brodton Avenue	San Diego	California	CA	39908.Hershel.Beer@vivila.org	856-655-1234	Female	279-42-6939	College	Unemployed	Single
22	Estevan Stracke	6242 Hartford Avenue	Cherry Hill	New Jersey	NJ	39649.Estevan.Stracke@victoria.us	501-734-9981	Male	400-74-8270	High School or Below	Unemployed	Married
23	Mable Hirthe	1733 Fairhaven Commons Way	Cedar Park	Texas	TX	39673.Mable.Hirthe@derek.net	516-752-4601	Male	855-57-1844	High School or Below	Unemployed	Single
24	Savanna Hand	5335 Soojian Dr	Salmon Creek	Washington	WA	39439.Savanna.Hand@linor.info	579-592-5306	Male	317-42-6094	High School or Below	Unemployed	Divorced
25	Emilee Leuschke	6337 Main Street	Minneapolis	Minnesota	MN	39204.Emilee.Leuschke@gmaila.name	614-403-3230	Male	367-20-2500	Bachelor	Employed	Married
26	Alexane Weber	5071 Colony Place	Rogers	Arkansas	AR	39028.Alexane.Weber@stephanie.me	732-335-1617	Female	879-11-8599	High School or Below	Unemployed	Married
27	Elouise Hickley	24520 Fall River Avenue	Gaffney	South Carolina	SC	38911.Elouise.Hickley@effery.org	602-391-5153	Male	582-27-7754	Bachelor	Employed	Married
28	Dorothea Leannon	1060 Palmer Road	Wilmington	North Carolina	NC	38735.Dorothea.Leannon@kiera.name	970-356-1041	Male	857-83-5404	High School or Below	Employed	Single
29	Josefa Reilly	4493 Rt 59	Phoenix	Arizona	AZ	38676.Josefa.Reilly@bianco.tv	773-120-8671	Male	269-47-8882	Master	Employed	Divorced
30	June Durgan	30393 Veterans Memorial Drive	Syracuse	New York	NY	38500.June.Durgan@frida.biz	954-975-9429	Female	356-78-3824	High School or Below	Employed	Single
31	Stephania Fadel	8504 Brewerton Rd	Littleton	Colorado	CO	38266.Stephania.Fadel@charity.us	440-442-6907	Male	385-65-7822	Doctor	Employed	Married
32	Anasley Bogisch	27358 Basile Rowe	Chino Hills	California	CA	38031.Anasley.Bogisch@grady.io	212-851-7140	Female	297-81-4190	Master	Employed	Divorced
33	Abel Feest	56160 Lakeville Road	Council Bluffs	Iowa	IA	37855.Abel.Feest@yah.co.uk	434-927-8028	Female	865-56-1661	College	Employed	Single
34	Rachael Erdman	4632 Rt 9W	Massillon	Ohio	OH	37797.Rachael.Erdman@quinton.info	614-691-8594	Male	644-60-8226	High School or Below	Employed	Divorced
35	Camyille Renner	55560 State Route 36	St. Charles	Illinois	IL	37621.Camyille.Renner@maryjane.org	919-697-1003	Male	747-09-7764	Bachelor	Employed	Married

Como analista de datos de Golden Bank, ha aprendido a buscar y encontrar los datos correctos, comprender y confiar en su contenido y, a continuación, prepararlo para que lo utilicen otros analistas de datos y científicos de datos.