

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 *IMPLEMENT_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. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and various account and workspace options. Below the header, the main area displays the 'MORTGAGE_APPLICANTS_TRUST' asset. The asset card includes a 'Remove' button and an 'Add to project' button. A 'Feature group' section is also visible. The 'Overview' tab is selected, showing a table of columns with details like name, quality score, source type, description, business terms, and data class. Below the table, sections for 'Governance artifacts' (Business terms and Classifications) are shown, both of which are currently empty. At the bottom, there are pagination controls for 'Items per page' and '2 of 2 pages'.

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 with the following details:

- Header:** IBM Cloud Pak for Data, Search in your workspaces, Buy, Renovations, Dallas, and a user profile icon.
- Breadcrumbs:** Catalogs / Mortgage Approval Catalog.
- Asset Overview:**
 - Data:** MORTGAGE_APPLICANTS_TRUST_PROTECT
 - Actions:** Remove, Add to project (+)
- Sections:**
 - Overview:** Contains tabs for Asset, Access, Review, and Profile.
 - Governance artifacts:** Business terms, Social Security Number, Loan.
 - Classifications:** Personally Identifiable Information.
 - Related assets:**

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
 - Show:** Any relationship, Add asset +
 - Items per page:** 25, 1–2 of 2 items, 1 of 1 pages, navigation arrows.
- Right Panel (About this asset):**
 - Description:** Mortgage applicants from the Mortgage System.
 - Asset owner:** Dina Maroni, sricher+dmaroni@mail.test.us.ibm.com
 - Privacy:** Public
 - Format:** application/x-ibm-rel-table
 - Asset details:** Size: -, Columns: 14, Rows: -
 - Source:** Connection: Data Fabric Trial - Db2 Warehouse, Source type: IBM Db2 Warehouse, Path: AL_MORTGAGE / MORTGAGE_APPLICANTS /
 - Location:** No location added yet.

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.



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.

The screenshot shows the 'Assets' tab selected in the navigation bar. On the left, there's a sidebar with '11 assets' and a 'Data asset' section highlighted. The main area displays a table of data assets with columns for 'Name' and 'Last modified'. The asset 'MORTGAGE_APPLICANTS_TRUST_PROTECT' is selected, indicated by a blue border around its row.

Name	Last modified
MORTGAGE_APPLICANTS_TRUST_PROTECT	Now application/x-ibm-rel-table Maggie Assad (You)
MORTGAGE_APPLICANTS	4 hours ago application/x-ibm-rel-table Maggie Assad (You)
MORTGAGE_APPLICATION	4 hours ago application/x-ibm-rel-table Service
COMMERCIAL_CLIENT	4 hours ago application/x-ibm-rel-table

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:

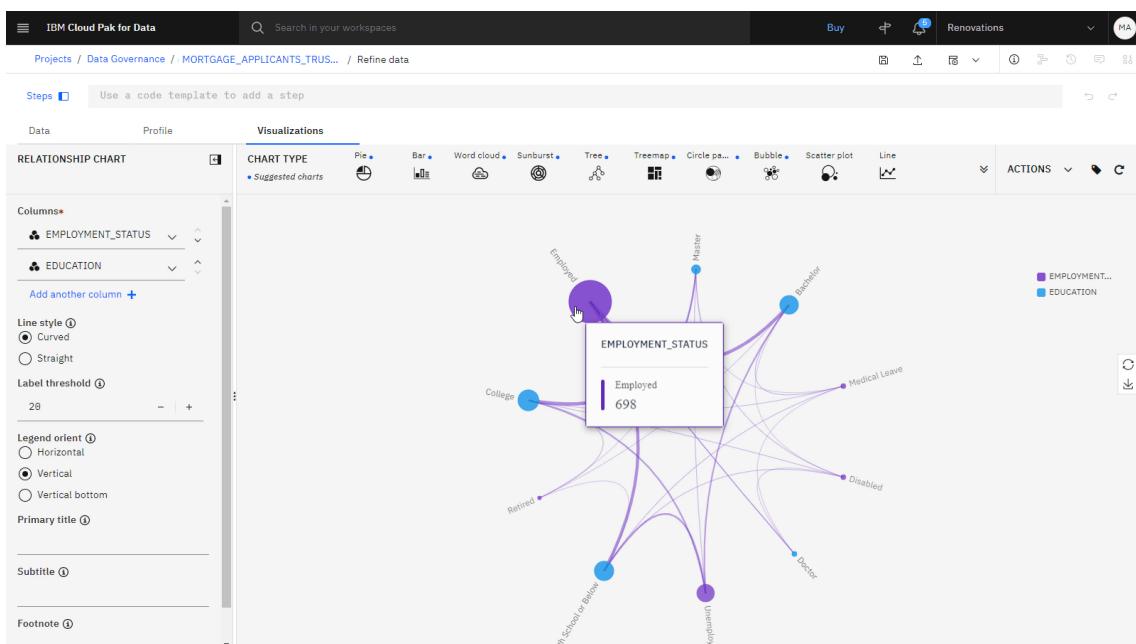
1. Pulse el activo de datos **HIPOTECAGE_APPLICANTS_TRUST_PROTECT** para obtener una vista previa de los datos.
2. Pulse **Preparar datos** para abrir el activo de datos en Data Refineryy espere a que los datos se lean y procesen.
3. En el panel *Acerca de este activo* , pulse la **X** para cerrar el 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 **EMPLOYEE_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:

- Para la primera columna, seleccione **IMPLEMENT_STATUS**.
- Pulse **Añadir otra columna**.
- 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.

The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with a menu icon, the text "IBM Cloud Pak for Data", and a search bar that says "Search in your workspaces". Below the navigation bar, the main title "Projects / Data Governance" is displayed in blue. The main content area shows a list of data entries, each with columns for ID, NAME, STREET ADDRESS, CITY, STATE, STATE ZIP, CODE, EMAIL, ADDRESS, PHONE NUMBER, GENDER, SOCIAL SECURITY, EDUCATION, and EMPLOYMENT/MARITAL STATUS. The data entries are listed from 1 to 36, corresponding to the rows in the CSV file.

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.

The screenshot shows a CSV file titled "MORTGAGE_APPLICANTS_TRUST_PROTECT_shaped.csv" with 36 rows of data. The columns include: ID, NAME, STREET ADDRESS, CITY, STATE, STATE ZIP, CODE, EMAIL, ADDRESS, PHONE NUMBER, GENDER, SOCIAL SECURITY, EDUCATION, and EMPLOYMENT/MARITAL STATUS. The data is presented in a grid format with horizontal and vertical scroll bars. To the right of the grid, there are several data visualization charts: a histogram for EDUCATION frequency, a scatter plot for FREQUENCY vs. EDUCATION, a pie chart for EDUCATION, and a bar chart for FREQUENCY. The bottom of the screen shows a status bar with "1000 Unique" and "1000 Unique" repeated across the bottom.

Step	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Data	1	110347 Hillary Murphy	4355 Bicknell Hills Dr	North Hills	California	CA	93485	Hillary.Murphy@idion.ca	937-255-8772	Female	457-12-7640	Doctor	Disabled	Divorced
Data	2	110350 Branson Crist	6095 Connecticut Avenue	Tahoe City	California	CA	43251	Branson.Crist@uiuanta.ma	845-512-1137	Male	591-81-1843	High School or Below	Unemployed	Married
Data	3	110353 Emma Friesen	6399 Torniford Street	Ashburn	Virginia	VA	43016	Emma.Friesen@elisha.net	630-893-4159	Male	479-51-0416	College	Retired	Single
Data	4	110363 Dennis Dooley	3161 Coonial Promenade Pkwy	Oseee	Florida	FL	42840	Dennis.Dooley@ace.biz	651-215-0230	Male	386-68-7811	Master	Employed	Married
Data	5	110369 Julia Gerhold	1474 West Bypass	Weberwood	New Jersey	NJ	42782	Reese.Wuensche@loriana.ca	201-920-0081	Male	220-76-0498	High School or Below	Employed	Married
Data	6	110374 Reese Wuensche	1474 West Bypass	Eastwood	Alaska	AK	42602	Reese.Wuensche@loriana.ca	540-771-7123	Male	409-02-1323	College	Retired	Single
Data	7	110381 Glen Marks	9881 Douglas Avenue	Burlington	Vermont	VT	42317	Glen.Marks@charonv.biz	249-89-0555	Male	409-02-1323	High School or Below	Unemployed	Divorced
Data	8	110385 Halle Cassin	10340 Spring Av SW	Madison	Wisconsin	WI	42317	Halle.Cassin@kevin.info	914-298-2798	Male	367-25-8238	Master	Employed	Married
Data	9	110391 Rene Leannon	65059 Aaron Arrow Drive	Oak Lawn	Illinois	IL	41961	Halle.Cassin@kevin.info	480-862-8327	Female	104-21-2766	Bachelor	Unemployed	Single
Data	10	110399 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
Data	11	110400 Tate Murzik	3525 Fort Morgan Road	Plainview	Texas	TX	41667	Tate.Murzik@idion.ca	918-292-9609	Male	374-87-7018	High School or Below	Employed	Married
Data	12	110401 Broderick Hauck	23071 Hwy 195	New Rochelle	New York	NY	41609	Broderick.Hauck@floyd.co.uk	714-620-3992	Female	314-45-7544	College	Unemployed	Single
Data	13	110402 Joanne Kline	6080 Lakewood Center Drive	Lakewood	Colorado	CO	40932	Joanne.Kline@idion.ca	361-533-1340	Male	367-25-8238	Bachelor	Unemployed	Single
Data	14	110403 Merrill Etchman	63421 Hwy 78 East	Reston	Virginia	VA	41194	Merrill.Etchman@ryca.us	970-288-2406	Male	394-17-0450	High School or Below	Unemployed	Single
Data	15	110404 Chance Adams	57385 Raintree Service Rd	Lakewood	California	CA	40964	Chance.Adams@fatima.me	928-505-4385	Female	690-41-9539	High School or Below	Disabled	Married
Data	16	110405 Drew Keebler	4095 Ann St	Denver	Colorado	CO	40780	Drew.Keebler@loyce.ca	516-668-0066	Female	899-32-0268	High School or Below	Unemployed	Married
Data	17	110406 Rickey Smitham	5188 West Avalon Ave	Myrtle Beach	South Carolina	SC	40720	Rickey.Smitham@yesenia.org	412-340-3623	Female	180-67-1839	Bachelor	Employed	Divorced
Data	18	110407 Matilde Hegmann	3051 Pelham Pkwy	Manchester	New Hampshire	NH	40551	Matilde.Hegmann@phylis.org	405-752-1411	Male	495-97-7784	Bachelor	Employed	Married
Data	19	110408 Sophie Boyer	17951 Hwy 14 East	Webster	Texas	TX	40318	Sophie.Boyer@illie.info	325-212-6229	Male	531-39-8976	College	Employed	Married
Data	20	110409 Esteban Stracke	5000 E Colorado	Centerville	Ohio	OH	40605	Esteban.Stracke@villia.org	409-44-1007	Male	367-25-8238	High School or Below	Employed	Married
Data	21	110410 Daniel Beier	6090 Brookton Avenue	San Diego	California	CA	39900	Daniel.Beier@villia.org	856-655-1224	Male	379-43-0399	College	Unemployed	Single
Data	22	110411 Michael Beier	6090 Brookton Avenue	San Diego	California	CA	39900	Michael.Beier@villia.org	856-655-1224	Female	379-43-0399	College	Unemployed	Single
Data	23	110471 Estevan Stracke	6432 Hartford Avenue	Cherry Hill	New Jersey	NJ	39840	Estevan.Stracke@victoria.us	501-734-9982	Male	404-74-8270	High School or Below	Unemployed	Married
Data	24	110472 Marlie Hirthe	1773 Fairhaven Commons Way	Cedar Park	Texas	TX	39673	Marlie.Hirthe@skrek.us	516-752-4601	Male	855-57-1844	High School or Below	Unemployed	Single
Data	25	110485 Savanna Hand	5335 Soojian Dr	Salmon Creek	Washington	WA	39430	Savanna.Hand@ellinor.info	573-592-3306	Male	317-42-4694	High School or Below	Employed	Divorced
Data	26	110493 Emelie Leuschke	6537 Main Street	Minneapolis	Minnesota	MN	39204	Emelie.Leuschke@mfafida.name	516-403-3230	Male	367-20-2500	Bachelor	Employed	Married
Data	27	110495 Alexane Weber	5071 Colony Plaza	Rogers	Arkansas	AR	39028	Alexane.Weber@stephania.me	734-235-1617	Female	879-11-8599	High School or Below	Unemployed	Married
Data	28	110496 Michaela Hickle	2040 Fall River Avenue	Gaffney	South Carolina	SC	39811	Michaela.Hickle@pferry.org	602-391-5353	Male	507-27-7752	Bachelor	Employed	Married
Data	29	110500 Dorothy Leannon	1009 Hwy 14 East	Wilmington	North Carolina	NC	39720	Dorothy.Leannon@villia.org	910-44-0406	Male	387-44-0406	High School or Below	Employed	Single
Data	30	110511 Jocelyn Reilly	4493 Ft 59	Phoenix	Arizona	AZ	38576	Jocelyn.Reilly@villia.org	573-220-9571	Male	389-67-8882	Master	Employed	Divorced
Data	31	110517 June Durgan	30393 Veterans Memorial Drive	Syracuse	New York	NY	38500	June.Durgan@frieda.biz	954-475-0429	Female	356-78-3824	High School or Below	Employed	Single
Data	32	110525 Stephanie Fadel	8504 Breverton Rd	Lentil	Colorado	CO	38266	Stephanie.Fadel@cherry.us	440-442-6907	Male	385-65-7822	Doctor	Employed	Married
Data	33	110530 Ansley Bogisch	27358 Basile Roive	Chino Hills	California	CA	38031	Ansley.Bogisch@grado.io	212-851-7240	Female	297-81-4190	Master	Employed	Divorced
Data	34	110539 Abel Fees	56109 Lakeville Road	Council Bluffs	Iowa	IA	37853	Abel.Fees@nyah.co.uk	434-527-8028	Female	865-56-1681	College	Employed	Single
Data	35	110541 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
Data	36	110547 Camille Renner	55560 State Route 36	St. Charles	Illinois	IL	37621	Camille.Renner@marijane.org	519-697-1003	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.