IBM Cloud Pak Experiences

# 

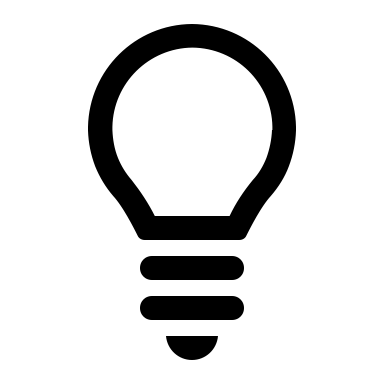
# Before you begin

Check out the Welcome Video: <http://ibm.biz/icpe-welcome>   
  
To complete the IBM Cloud Pak Experiences, you must have one of the following web browsers:

* Firefox Version 54 or later
* Chrome Version 59.03 or later

To get started:

1. Go to <https://www.ibm.com/cloud/garage/cloud-private-experiences/icpd> and sign up or log in.
2. Find the path you want to take and click **Explore**.
3. Locate your welcome email, which contains your user name and password.
4. Log in to the web client with the credentials from the email.



If you see this icon, we’re giving you insight into how the product works.

# The story

You work for a large multi-national bank and are responsible for processing mortgage applications.

Approving a mortgage application is manual and resource-intensive process. The process involves looking at a number of variables such as a client’s income, credit card debt, the sale price of the house, the loan amount, and so on, to determine whether the loan is a safe bet or a big gamble.

You’re responsible for automating the process by building a machine learning model that can accurately predict the likelihood of any given client defaulting on their mortgage.

The model will be invoked in real time by a web or mobile app available that customers use to apply for a mortgage. The model enables your company to analyze a mortgage application and decide whether to approve it in a few seconds with no manual intervention. This leads to significant cost savings and, more importantly, improved customer service because applications will be approved or rejected instantaneously.

# The journeys

## [Collect](#bookmark)

Connect to your existing databases so that you have quick and easy access to your data – no matter where it lives.

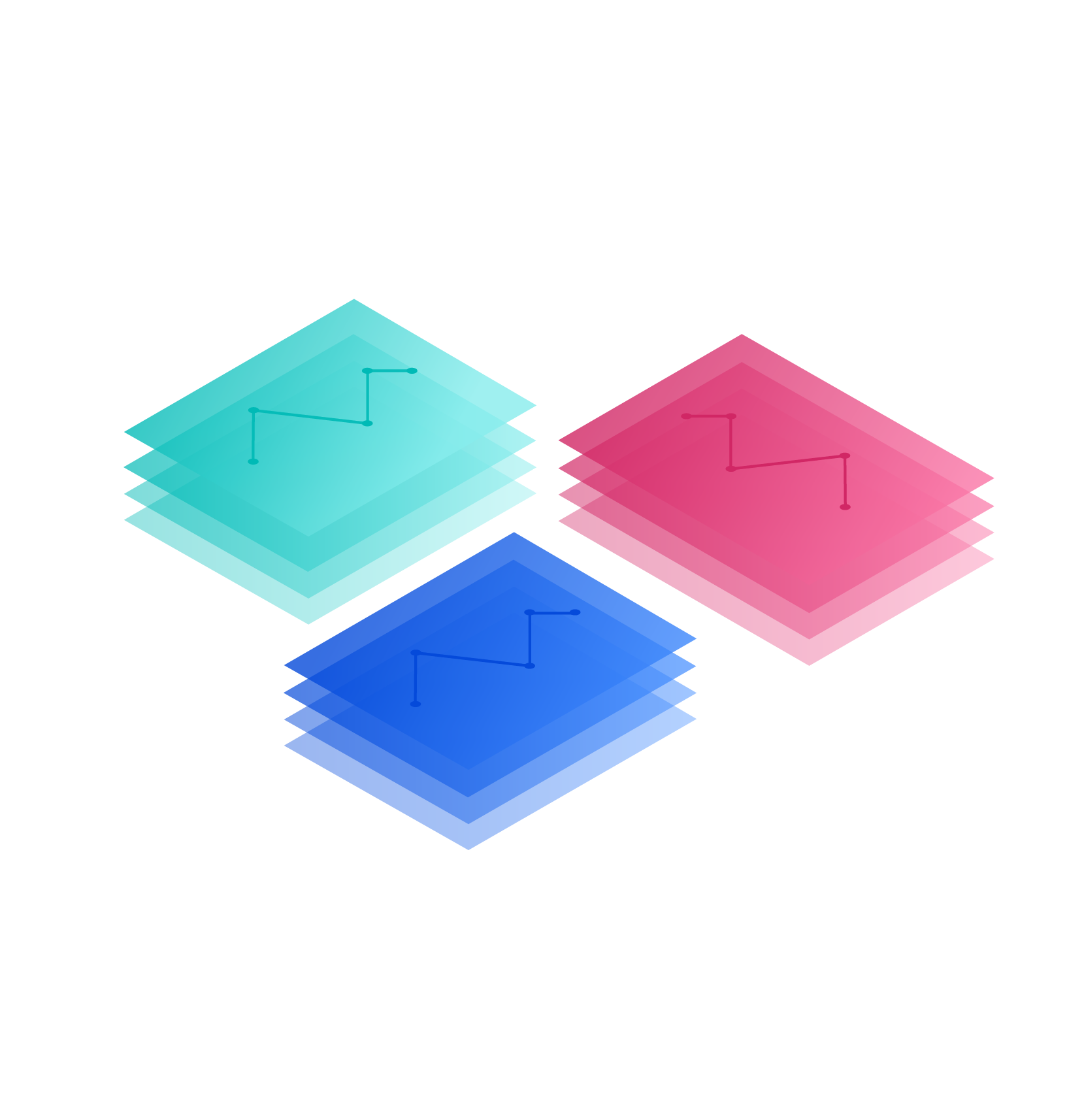
## [Organize](#bookmark1)

Turn your data into trusted data. Use automated discovery to import, analyze, and classify data. Define business terms so that your data is consistent. Apply rules and policies to make your data compliant with regulations. Transform the data to make it useful. But most importantly, make the data easy for users to find by publishing it to the enterprise data catalog.

## [Analyze](#bookmark2)

Build, deploy, and publish a machine learning model to predict whether clients will repay their mortgage or default on it.

# Organize

Learn to:

**Organize your data so it’s easy to find, trust and use**

5 – 7 minutes

**What will I learn?**  
In this stage of the journey, as a business user**,** you will utilize the self-service and AI assisted organizational capabilities to quickly and easily search for and find the right data, understand and trust it's content, infuse it into an analytics project and then refine and shape it in preparation for analytics and AI.

## 1. Find the Right Data

To get started:

A screenshot of a cell phone

Description automatically generated

We need to find the right data and business information related to the **Mortgage Default** analysis project. You can use the global search to search across catalogs, projects and the business glossary to find all assets that you may be interested in.

1. Enter the words **mortgage** in the global search area and press the enter key to start finding what you need.

A screenshot of a cell phone

Description automatically generated

The search returns all data and information assets related to the search criteria across all catalogs, projects and governance artifacts. Scroll down through the list to take a closer look at what was found.

A screenshot of a cell phone

Description automatically generated

You can further refine your search results by using the filters supplied by type, tag, catalog, project etc.

A screenshot of a cell phone

Description automatically generated

First and foremost, we need to verify that all the data needed for the project has been cataloged and available to the project team to use. We can do this very easily by refining the search to only display mortgage data that exists across all catalogs.

1. From the search drop down menu select **Catalogs > All catalogs.** Place your cursor inside the global search area next to the word **mortgage** and press the enter key.

A screenshot of a cell phone

Description automatically generated

All data assets across catalogs meeting the criteria are displayed. This is the data we are looking for.

The connection to the Analytics Data Warehouse and the 4 mortgage tables are what the project team requested; Mortgage Default, Applicant, Property and Customer are all in the **Enterprise** catalog.

However, before we proceed to the catalog we need to also find all the business information related to the project to review the terms and content of the data and identify if there are any policies and rules set by the business that the project team needs to be aware of and adhere to.

A screenshot of a cell phone

Description automatically generated

1. From the search drop down menu select **All.** Place your cursor inside the global search area next to the word **mortgage** and press the enter key.

A screenshot of a cell phone

Description automatically generated

1. Click on the **Any type** filer and select the **Business term** type from the list to refine the results.

A screenshot of a cell phone

Description automatically generated

Notice that the governance team has been hard at work and has defined 25 business terms related to the mortgage data the project will be using. Also notice that they are tagged with the key word **Mortgage** to easily find them

1. Click on the **Sort Arrow** to the right of the **Name** column to sort ascending by Name.

**Note** that the Email Address has also been tagged as **Sensitive information**.

1. Scroll down the list of business terms to view them all.

**Note** that the Phone Number and Social Security Number are tagged as **Sensitive information**.

1. Click on the **Clear all** button to get ready for the next search task.

A screenshot of a cell phone

Description automatically generated

1. Click on the **Any type** filer and select the **Category** type from the list to refine the results.
2. Click on the **Any type** filer and select the **Policy** type from the list to refine the results.
3. Click on the **Any type** filer and select the **Rule** type from the list to refine the results.

A screenshot of a cell phone

Description automatically generated

Notice that there is a **Mortgage Default Analysis** category defined that contains all the business information related to the project and a category named **Sensitive Information** that is a sub-category along with a business policy related to the **Protection of Sensitive Information** and three rules outlining the protection of **Email Addresses, Phone Numbers and Social Security Numbers**. This is a good indication that the mortgage data being used by the project contains sensitive information that needs to be protected.

## 2. Understand the Data

The best data is data that is fully understood and trusted.

You can be confident in your data when you know where it comes from, that it complies to a set of policies and rules that address data privacy regulations and that it is clean and conforms to data quality policies, rules and standards, and that others have used it and trust it and are willing to share that information, to ensure you can produce meaningful and accurate analytical and AI results that will benefit better business outcomes.

### Understand Data Policies and Rules

In this section you will use the **Business glossary** to gain a deeper understanding of the business terminology defined by the data steward and the governance team responsible for establishing policies and rules to govern and protect the data. Since we see that there is an indication of sensitive information Let’s take a closer look at the **Sensitive Information** category content.

A screenshot of a cell phone

Description automatically generated

1. Click on the **Sensitive Information** category from the list.

A screenshot of a cell phone

Description automatically generated

The description clearly states that thissubcategory contains references to business terms that relate to data that will be used by the project team that need to be governed by data protection rules. Let’s get more information on the **Social Security Number.**

1. In the global search area, enter the words **social security number** and press the enter key.

A screenshot of a cell phone

Description automatically generated

A Social Security Number term appears in the list with tags of Mortgage applicant and Sensitive information. This is an indication that the Mortgage Applicant table has a social security number.

1. Click on the **Social Security Number** business term with the two tags applied from the list.

A screenshot of a cell phone

Description automatically generated

1. Click on the **Related Content** tab.
2. Click on the arrow next to the **MORTGAGE\_APPLICANT** table in the Catalog assets section. There is definitely a SOCIAL\_SECURITY\_NUMBER column that resides in this table.
3. Click on the **MORTGAGE\_APPLICANT** table in the Additional information assets section.

A screenshot of a cell phone

Description automatically generated

It has been assigned the three business terms that relate to sensitive information. It has 14 columns and is from the **MORTGAGE** schema from the **BLUDB** database. Let’s investigate further and Explore and visualize these relationships.

1. Click on the **ellipses…** in the top right corner and select the **Explore relationships** menu item.

A new tab will open in your browser and display the relationship graph.

A screenshot of a map

Description automatically generated

1. Click the **magnifying glass with a + sign** in the lower left corner to increase the graph size.
2. Click the plus sign next to **Database Columns** to expand it and view all the columns.

A close up of a map

Description automatically generated

1. Click on the **Center graph** button in the lower left corner to center the graph for easier viewing.

The table also contains EMAIL\_ADDRESS and PHONE\_NUMBER data that we know is sensitive.

1. Close this tab in your browser and go back to the source tab that got you here.

A screenshot of a cell phone

Description automatically generated

1. Click on the back **IBM Cloud Pak for Data** title to go back to the welcome page.

### Trust the Data Quality

The data quality analyst has informed the project team that the mortgage data quality meets the company’s standards and the data content was thoroughly analyzed and assessed before it was published to the catalog after it was discovered. You can see the results of that discovery and analysis in the Data Curation section of the Organize component of Cloud Pak for Data.

To review the data quality results:

A screenshot of a cell phone

Description automatically generated

* + - 1. Click on the Navigation men in the top left corner.

A screenshot of a social media post

Description automatically generated

* + - 1. Click on the **Organize > Curation > Data discovery** menu.

A screenshot of a cell phone

Description automatically generated

* + - 1. Click on the **View automated discovery results** link in the top right corner.

A screenshot of a cell phone

Description automatically generated

* + - 1. Click on the Discovery results **ID** for view the Mortgage Default Analysis quality results.

A screenshot of a cell phone

Description automatically generated

* + - 1. Click on the **Review discovery results** button under **Actions** in the bottom right corner that is highlighted.

A screenshot of a cell phone

Description automatically generated

You immediately see the **Quality score** for each table analyzed with the lowest score being 98% and the highest being 100%. Also notice that business terms have been assigned at the table level.

Since we know the MORTGAGE\_APPLICANT table is a new addition to the project, that was just requested, and it contains sensitive information, we will drill down into the details of the analysis

* + - 1. Click on the **Arrow** next to the MORTGAGE\_APPLICANT table to expand it and review the column analysis.

A screenshot of a computer

Description automatically generated

You are presented with a **Quality score** for each column analyzed. The quality looks satisfactory and meets the quality standard thresholds and the data classes and business terms are accurately assigned and complete. This is important to ensure the data is autonomously protected and that the governance artifacts are related to data assets they apply to.

* + - 1. Scroll down to see the remaining columns.

A screen shot of a computer

Description automatically generated

* + - 1. Click on the **IBM Cloud Pak for Data** title to get ready for the next section.

### Understand the Data Content

You have gained an understanding of the policies and rules and information related to sensitive data and validated and trust the data quality.

In this section we will go to the **Enterprise** catalog, which is where we identified all the data we need resides, and use all of the features it provides to gain an even better understanding of the data content and have even more confidence in the data based on what others are saying and by utilizing the AI assisted recommendations, automatic profiling and additional data content statistics provided.

A screenshot of text

Description automatically generated

1. Click on the **Navigation** menu in the top left corner.

A screenshot of a cell phone

Description automatically generated

1. From the menu select **Organize > All catalogs**.

A screenshot of a cell phone

Description automatically generated

1. Click on the **Enterprise** catalog.

A screenshot of a social media post

Description automatically generated

Watson Knowledge Catalog provides suggested assets to you based on recommendations using AI, things you might be interested in based on your past viewing history. Notice that it is already recommending **Mortgage** data to you based on your past searches. It also keeps track of what’s hot and **Highly Rated** based on reviews and ranks them in order of their rating highest to lowest.

1. Click on the **High Rated** section to see what’s hot.

A screenshot of a social media post

Description automatically generated

Notice that the Analytics Data Warehouse, MORTGAGE\_APPLICANT and MORTGAGE\_CUSTOMER tables have been reviewed and are rated quite high. A good indication of their quality and usability.

Lastly, Watson Knowledge Catalog keeps track of what’s new that has been **Recently Added** since the last time you visited the catalog. This are all means to help you find and understand the data more quickly and easily.

1. Click on the **Recently Added** section to see what’s new.

A screenshot of a social media post

Description automatically generated

Notice that the MORTGAGE\_APPLICANT table was the most recent data asset added to the catalog.

1. Click on the **MORTGAGE\_APPLICANT** table to review its content and metadata.

A screenshot of a social media post

Description automatically generated

You are brought into the Overview section of the MORTGAGE\_APPLICANT table. You immediately see that Data masking is in progress and that 3 columns are being masked. The asset is being masked by the **Protect Sensitive Personal Information** data protection rule being enforced by the data governance team. Because you are not authorized to view the sensitive information, the data is being protected.

A screenshot of a social media post

Description automatically generated

The data that is masked is indicated with a **lock** icon next to their column names; EMAIL\_ADDRESS, PHONE\_NUMBER and SOCIAL\_SECURITY\_NUMBER.

1. Scroll to the right to view all the masked columns.

A screenshot of a social media post

Description automatically generated

1. Hover over the **lock** icon on the EMAIL\_ADDRESS column to view the data protection message. Do the same for the PHONE\_NUMBER and SOCIAL\_SECURITY\_NUMBER columns.
2. Click on the **Review** tab to read the review.

A screenshot of a social media post

Description automatically generated

The table has a 5-star rating with a very positive review. Reviews can be written by anyone who has access to the catalog and the asset to notify and inform others of the content and usability of the data.

1. Click on the **Profile** tab to view the data profile.

A screenshot of a computer

Description automatically generated

As data assets are discovered and added to the catalog they are automatically profiled and classified to give end users a more in-depth understanding of the data content, quality and usability. Data classifications are used to identify what type of data it is and to autonomously enforce data protect rules to mask sensitive data, like you just saw.

1. Scroll to the right to view the other columns.

A screenshot of a social media post

Description automatically generated

Notice that the protected data does not have any profile information displayed.

1. Click on the **Lineage** tab to view the data asset lineage.

A screenshot of a cell phone

Description automatically generated

Lineage is captured for every data asset in a catalog. It keeps track of where it came from, any updates or changes that have been made to its metadata and any movement of the data outside of the catalog.

1. Click on the yellow + sign to see the detail pane on the right. Do this for every node in the lineage graph to see what is tracked.

## 3. Prepare the Data for Analytics and AI

You have found the data you are looking for, have been assured that the sensitive data has been protected with no risk of being exposed, have a good understanding and trust in the data and the confidence that it will suit the needs of the project team . It is now time to get it into the project and begin preparing it and using it for the analytical and AI tasks.

**Note** – Since this a shared environment, and for the sake of time, the Mortgage Default Analysis project has already been created and populated with the data from the catalog and we have pre-built the data refinery data flow for you to review.

A screenshot of a social media post

Description automatically generated

1. Click the **Collapse** button to close the recommendation section and gain more real estate.

A screenshot of a cell phone

Description automatically generated

1. Enter the word **mortgage** in the search area to limit the data assets in the list.

A screenshot of a cell phone

Description automatically generated

1. Click the check box next to the **MORTGAGE\_APPLICANT, MORTGAGE\_CUSTOMER, MORTGAGE\_DEFAULT and MORTGAGE\_PROPERTY** assets in the list.
2. Click on the **Add to Project** button.

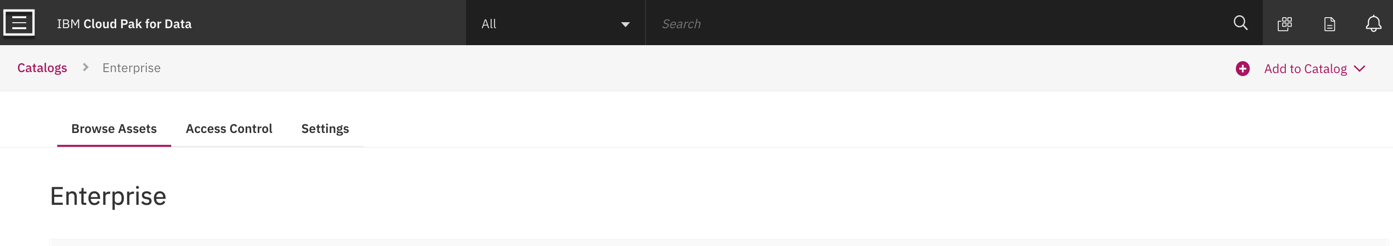
A screenshot of a cell phone

Description automatically generated

1. Select the **Mortgage Default Analysis** project from the Target list. **Notice** that even though you did not select the Analytics Data Warehouse connection, Knowledge Catalog knows it will be needed so it is being included so you can properly access the data one it is in the project.

**Note** – As stated earlier, the Mortgage Default Analysis project has already been created and populated with the data assets from the catalog so you **will not** add the data assets but experience how it was done and Cancel out.

1. Select the **Cancel** button.



1. Select the **Navigation** menu in the top left corner.

A screenshot of a cell phone

Description automatically generated

1. Select the **Projects** menu item.

A screenshot of a cell phone

Description automatically generated

1. Click on the **Mortgage Default Analysis** project.
2. When the project opens, close any Getting started dialogs that may appear.

A screenshot of a cell phone

Description automatically generated

Notice that the data assets are in the project.

1. Scroll down to the **Data Refinery flows** section.

A screenshot of a social media post

Description automatically generated

1. Click on the **Mortgage Records with Applicants** data flow.
2. When the data flow opens, close any Getting started dialogs that may appear.

A screenshot of a cell phone

Description automatically generated

1. Click on the **Steps** button to view all 12 data preparation recipe steps.

A screenshot of a cell phone

Description automatically generated

1. Scroll through the steps to see the data preparations operations performed to shape the data.
2. Click on the **IBM Cloud Pak for Data** title to go back to the welcome page.

## Summary

**You organized your data so it’s easy to find, understand and use**

By completing this journey, you:

* Quickly found the data that was needed for the project
* Easily found and understood the data governance terms, policies and rules
* Gained trust in the quality and content of the data
* Prepared the data for analytics and AI