

What is IBM Studio?

Every business wants to work smarter, and to do that you need to tap into your company's greatest resource, your data. But extracting the full value out of your data isn't always an easy process. First, you end up juggling an incredibly large and complex collection of tools that are used for finding and cleaning data, analyzing and generating visualizations of that data, and using the data to build and deploy machine learning models. And to make matters worse these tools are often a time drain to individually manage, and can be difficult to integrate into your system, which can really slow down the workflow.

But not anymore. Using [Watson Studio](#) you

- can simplify your data projects with a streamlined process, that allows you to extract value and insights from your data to help your business get smarter, faster.
- It delivers an easy-to-use collaborative data science and machine learning environment for building and training models, preparing and analyzing data, and sharing insights, all in one place.
- Watson Studios easy to create visualizations and drag-and-drop code put the power of database decision-making into the hands of any member of your organization with no need for IT assistance.
- And if you need access to open source tools, the environment offers some of the most popular and powerful ones available.
- Watson Studio single environment also creates a workflow that's incredibly efficient so data scientists can share assets and work to solve problems within the system rather than starting from scratch every time a new issue arises. And developers can use that efficiency to quickly dive into building machine learning and deep learning algorithms.
- In fact, in the area of deep learning, Watson Studio supports some of the most popular frameworks and can deploy that deep learning on to the latest GPUs to

help accelerate modeling by making it easier to use. The environments built-in neural network modeler also helps you build models with a simplified graphical interface even if you don't have the dedicated resources to build a model from scratch,

- Watson's Studio can help you get started with modeling templates for areas such as visual recognition, language classification, and other tools from IBM Watson services.

Because Watson Studio is seamlessly integrated with the IBM Watson Knowledge Catalog, an intelligent asset discovery tool,

- you can transform data and models into trusted enterprise resources and collaborate with confidence, without compromising compliance, security or access control.
- Watson Studio provides many benefits for organizations helping to infuse AI into the business and drive innovation. You can train Watson Studio with embedded AI services including watson visual recognition.
- You can customize your models and deploy them as APIs or Core ML by using open source tools like Jupyter, Notebook, Anaconda and RStudio. Watson Studio supports most popular code libraries as well as no code visual modeling with neural network modeler for designing neural architectures using the most popular deep learning frameworks.
- In Watson Studio you can interactively discover, cleanse, and transform your data using data refinery. It helps you understand the quality and distribution of your data with built-in charts and statistics, and provides visualized results through interactive dashboards.
- Watson Studio includes an intuitive drag-and-drop interface that enables a non programmer to speed up the bottle building process by visually selecting, configuring, designing and auto coding neural networks. From development and training to production and evaluation,
- Watson Studio tracks your models over time to ensure you have the best performance for any given task using the best solutions across the entire lifecycle of your machine learning models.

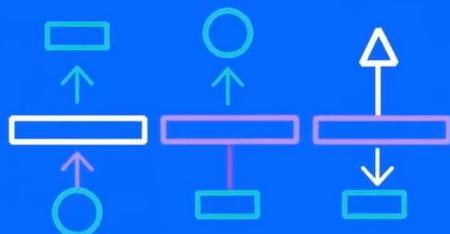
Data: your greatest resource

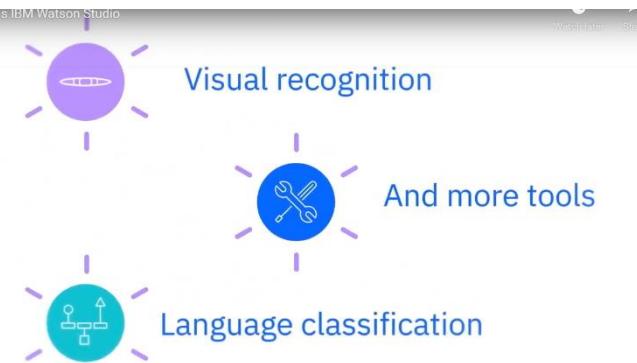
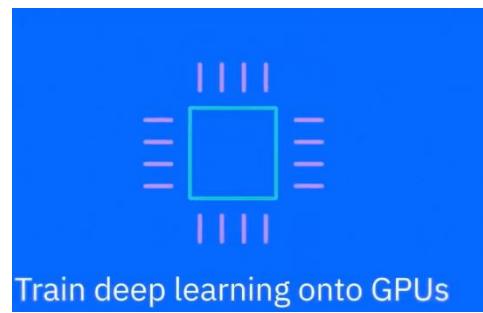
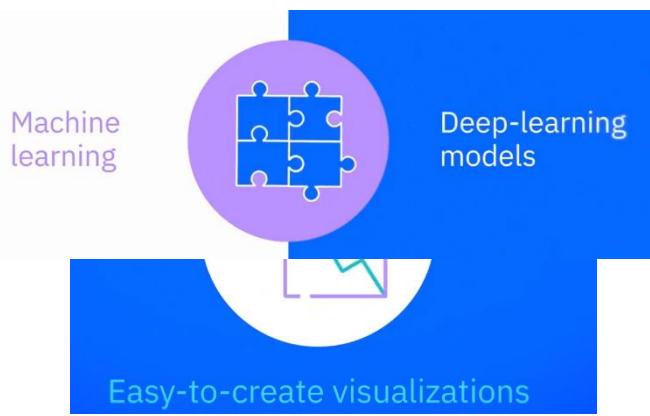
Find & clean the data

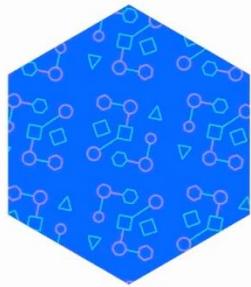


Visualize the data

Build machine-learning models



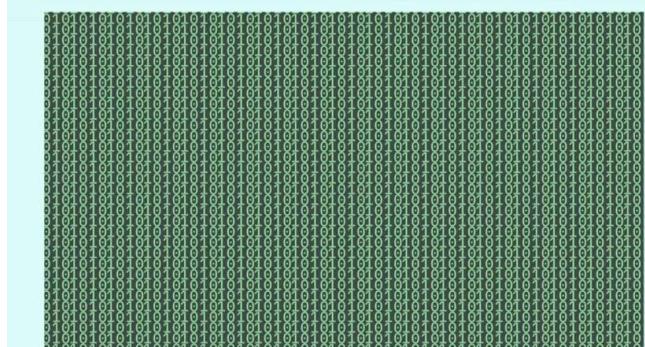
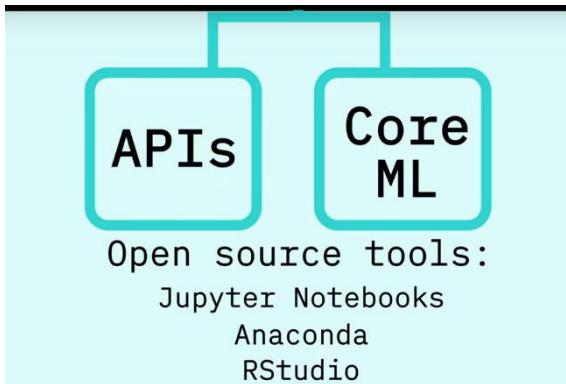




IBM Watson
Knowledge Catalog



Collaborate with
confidence



Most popular code libraries

No code visual modeling

with Neural Network Modeler. For designing neural architectures using the most popular deep learning frameworks.

Data Refinery



- Discover
- Cleanse
- Transform

Selecting

Configuring

Designing

Auto coding

Development and training

Production and evaluation



Watson Studio – An Introduction

Watson Studio is an integrated platform of tools, services, and data that helps companies accelerate their shift to become data-driven organizations.

You can start with a free account to explore its capabilities. Data science is a team sport; we have different types of people interested in the insights that data science can provide. This includes business analysts, data engineers, data stewards, data scientists, and developers. Data needs to be located and cleansed, models have to be created, tested, monitored, and updated. All this requires teamwork. For this reason

- Watson Studio was built as a collaborative platform a community of like-minded people.

There is a lot to cover in this introduction and will only scratch the surface. You can find more information on the digital technical engagement site at ibm.com/demos.

Once you are logged in you may see the Get Started Welcome screen. You can minimize the screen by clicking on the get started button in the upper right.

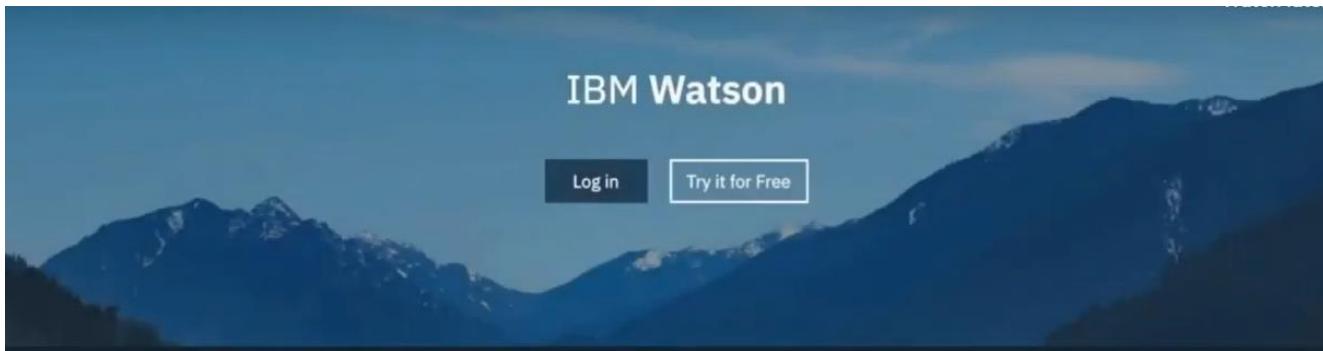
One important item that is easy to miss is the hamburger button in the upper left it gives you direct access to projects, catalogs, and services among other things. The gallery is particularly interesting. It is a collection of assets including tutorials, notebooks, data sets, articles, and papers from multiple sources. New assets are constantly added. Assets can be searched using filters for type, language, technology, topics and so on. The results can be sorted by features or by date. Manage gives you quick access to specific areas to manage-- finally we have integrated support and documentation in the Watson environment.

As mentioned earlier, the project is the center of the collaboration. It is very simple to create a project. You click "create a project" in the welcome screen or "new project" and either create an empty project or one from an existing one. Then you give it a name, possibly add a description, and you're ready to go. At the project level we also have a menu of options. It starts with the overview where you can see basic information on the project. This tab also includes a README section where you can get more details on what the

project is about. The next one is assets where you can see the data assets, models, notebooks, and other assets that are part of the project. You can go to add specific assets using the "Add to project" drop down menu at the top of the screen. We won't go into all of those menu items but one important one to know is "connection." This allows you to access data that comes from outside Watson Studio as you can see it includes a lot of data services from IBM but also quite a few from third parties such as Amazon and Microsoft. Going back to our project I'd like to point out the environment section. One important tool for that exploration, data manipulation, and model creation is the notebook. Depending on the amount of work that needs to be done we have a choice of resource allocation. We can also tailor the environment to include additional libraries so we have a complete environment from the start. I want to point out two more selections from the top menu: "Access Control" and "Settings." The access control allows you to control collaborators and their permissions and more. In the settings section you can among other things, add services. For example you click on the "Add service" drop-down menu, select "Watson" and add a "Machine-learning" service. You have the choice to add an existing service you may have created earlier in another project or create a new one. Note that most services include a light free version. This means that you can experiment with all sorts of capabilities for free.

Watson Studio

Introduction



IBM Watson

Log in Try it for Free

Explore our apps

Use IBM Watson to collaborate and build smarter applications. Quickly visualize and discover insights from your data and collaborate across teams.



IBM Watson Studio

Democratize ML/DL to accelerate infusion of AI in your business.



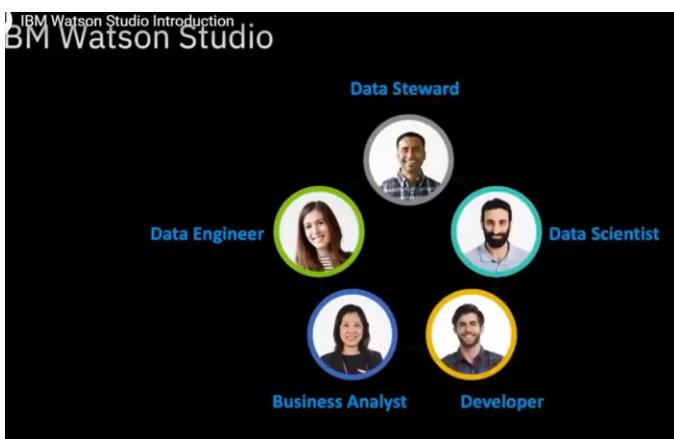
IBM Watson Machine Learning

Make smarter decisions, solve tough problems, and improve user outcomes.



IBM Watson Knowledge Catalog

Securely discover, catalog, and govern enterprise data.



IBM Demos Search My Demos

IBM Watson Studio Introduction

IBM Demos

ibm.com/demos

Watch later Share

Quickly discover the demos you want

Explore by Segment

Related products have been grouped together. Select from the list below to view all demos by segment.

Filter by Segment

Explore all demos →

Discover by Product

Have a specific product in mind? Select from the list below to view all demos by product.

Watson Studio |

IBM Watson Knowledge Studio

IBM Watson Studio

Browse by Expert

Know the author of the demo? Select from the list below to view all demos by expert.

Filter by Expert

Featured Collections

A collection is a distinct and organized group of demos that provide an in-depth and guided educational experience.

IDEOS Developer

Cloud Pak for Applications

Cloud Pak for Applications is an integrated solution that helps to modernize your applications to cloud native deployments under an offering with flexible entitlement.

Cloud Pak for Integration

Cloud Pak for Integration helps enable a multi-cloud approach to integrating data and applications.

Welcome Jacques!

Watson Studio • Watson Knowledge Catalog

Start by creating a project

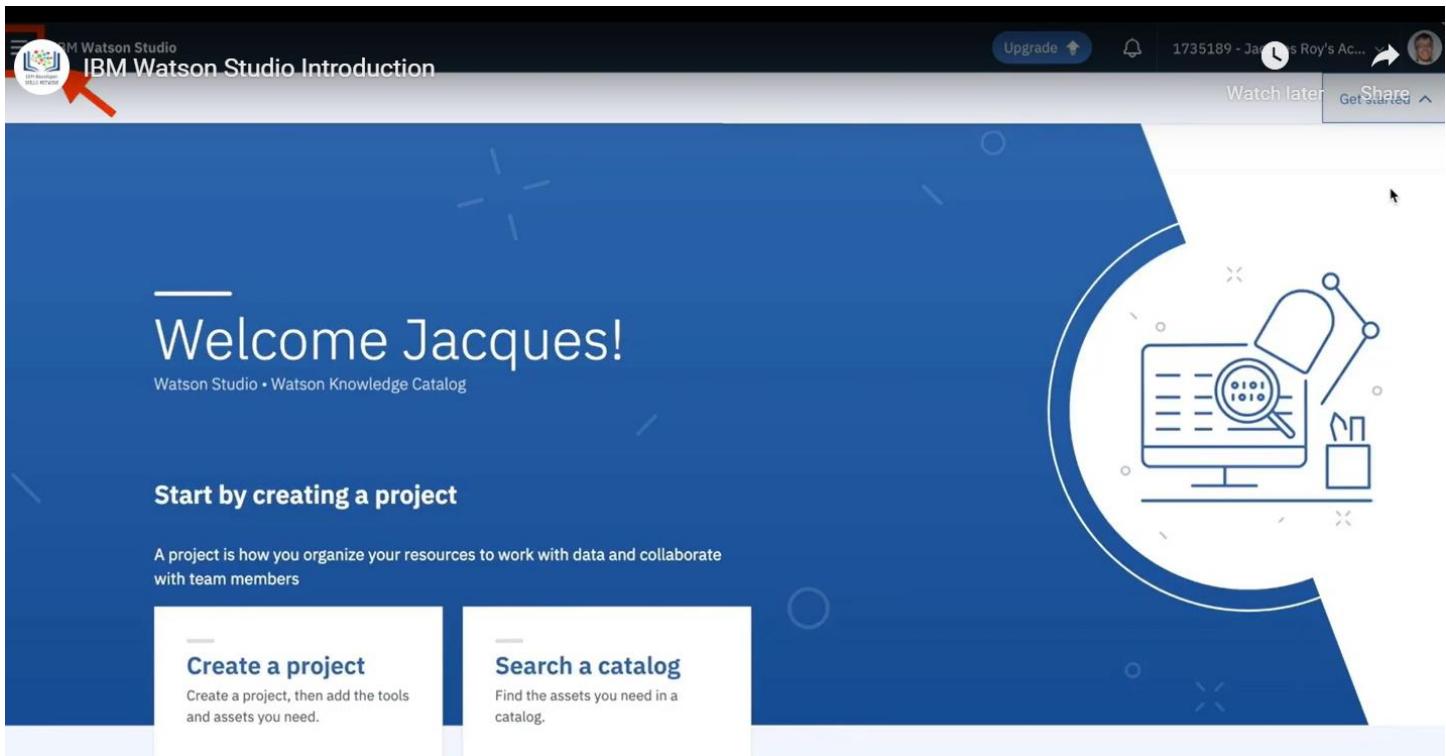
A project is how you organize your resources to work with data and collaborate with team members

Create a project

Create a project, then add the tools and assets you need.

Search a catalog

Find the assets you need in a catalog.



IBM Watson Studio
IBM Watson Studio Introduction

Welcome Jacques!

Watson Studio • Watson Knowledge Catalog

Start by creating a project

A project is how you organize your resources to work with data and collaborate with team members

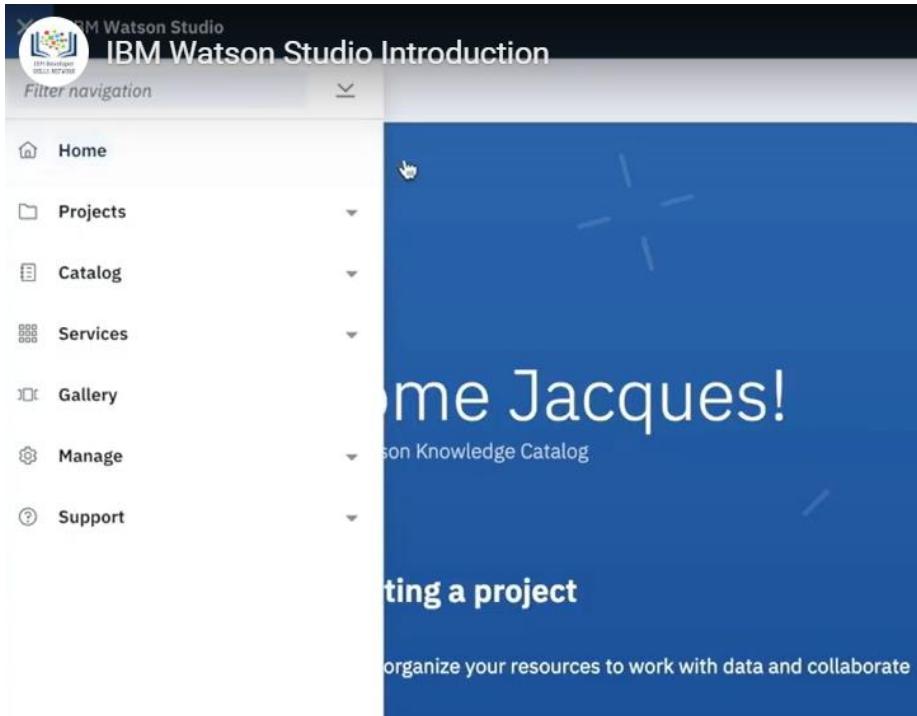
Create a project

Create a project, then add the tools and assets you need.

Search a catalog

Find the assets you need in a catalog.

Watch later Get Started



IBM Watson Studio
IBM Watson Studio Introduction

Filter navigation

- Home
- Projects
- Catalog
- Services
- Gallery
- Manage
- Support

Welcome Jacques!

Start by creating a project

A project is how you organize your resources to work with data and collaborate with team members

Gallery ⓘ

All filters Search by tags or asset titles

Featured

NOTEBOOK Deploying a Decision Optimization Model with... AUTHOR IBM MODIFIED Sep 12, 2019 <small>Science & Technology</small>	NOTEBOOK Use Spark and Python to Predict Equipment... AUTHOR IBM MODIFIED Jul 19, 2019 <small>Economy & Business</small>	DATA SET MNIST handwritten digits - PKL format AUTHOR IBM MODIFIED Jul 11, 2019 <small>Science & Technology</small>	DATA SET Greenhouse Gas Emissions worldwide AUTHOR IBM MODIFIED May 22, 2016 <small>Environment</small>
--	---	--	--

All content

DATA SET US County Boundary - geojson format AUTHOR IBM MODIFIED Oct 02, 2019 <small>Data Science</small>	NOTEBOOK Use PMML to predict iris species AUTHOR IBM MODIFIED Sep 17, 2019 <small>Economy & Business</small>	NOTEBOOK Deploying a Decision Optimization Model with... AUTHOR IBM MODIFIED Sep 12, 2019 <small>Science & Technology</small>	NOTEBOOK Measure Watson Assistant Performance AUTHOR IBM MODIFIED Aug 26, 2019 <small>Watson Assistant</small>
--	---	--	---

Sort by **Featured** ▾

All filters

Reset filters

Format

Data Set

Notebook

Topic

Communications

Economy & Business

Environment

Geography

Health

Law & Government

Leisure

Science & Technology

Society

Transportation

Gallery ⓘ

Search by tags or asset titles

Featured

NOTEBOOK Deploying a Decision Optimization Model with... AUTHOR IBM MODIFIED Sep 12, 2019 <small>Science & Technology</small>	NOTEBOOK Use Spark and Python to Predict Equipment... AUTHOR IBM MODIFIED Jul 19, 2019 <small>Economy & Business</small>	DATA SET MNIST handwritten digits - PKL format AUTHOR IBM MODIFIED Jul 11, 2019 <small>Science & Technology</small>
DATA SET Greenhouse Gas Emissions worldwide AUTHOR IBM MODIFIED May 22, 2016 <small>Environment</small>		

Filter navigation

- Home
- Projects
- Catalog
- Services
- Gallery
- Manage

Environment Runtimes

Catalogs

Storage Delegation

Authorized Users

Account

Billing and usage

Access (IAM)

Support

Welcome Jacques!

Person Knowledge Catalog

ting a project

organize your resources to work with data and collaborate

project

then add the tools

Search a catalog

Find the assets you need in a catalog.

The screenshot shows a software interface with a navigation bar on the left and a main content area on the right. The navigation bar includes links for Home, Projects, Catalog, Services, Gallery, Manage, Environment Runtimes, Catalogs, Storage Delegation, Authorized Users, Account, Billing and usage, Access (IAM), and Support. The 'Manage' link is currently selected. The main content area features a blue header with the text 'Welcome Jacques!' and 'Person Knowledge Catalog'. Below this, there are two cards: 'Create a project' (described as 'organize your resources to work with data and collaborate') and 'Search a catalog' (described as 'Find the assets you need in a catalog').

IBM Watson Studio
IBM Watson Studio Introduction

Filter navigation

- Home
- Projects
- Catalog
- Services
- Gallery
- Manage
- Support

Welcome Jacques!

Getting started with Watson Knowledge Catalog

Organize your resources to work with data and collaborate

Create a project

When you add the tools you need, you can add them here.

Search a catalog

Find the assets you need in a catalog.

← Back Watch!

Create a project

Choose whether to create an empty project or to preload your project with data and analytical assets. Add collaborators and data, and then choose the right tools to accomplish your goals. Add services as necessary.

Create an empty project



Add the data you want to prepare, analyze, or model. Choose tools based on how you want to work: write code, create a flow on a graphical canvas, or automatically build models.

NEW AutoAI experiment tool: Fully automated approach to building a classification or re...

USE TO

- Prepare and visualize data
- Analyze data in notebooks
- Train models

Create a project from a sample or file



Get started fast by loading existing assets. Choose a project file from your system, or choose a curated sample project.

USE TO

- Learn by example
- Build on existing work
- Run tutorials

New project

Define project details

Name

My first project

Description

Project description

Choose project options

Restrict who can be a collaborator [i](#)

Project will include integration with Cloud Object Storage for storing project assets.

Storage

cloud-object-storage-up

Watson Studio introduction

My Projects / Watson Studio introduction

Overview Assets Environments Jobs Deployments Access Control Settings

Watson Studio introduction

Last Updated: 14 Oct, 2019

[Readme](#)

1 Assets 1 Collaborators

Date created 14 Oct, 2019

Description No description available

Storage Cloud Object Storage 599.95 KB used

Collaborators View all (1)

Jacques Roy Admin

Recent activity

Alerts related to this project will show here when the project is active.

The screenshot shows the Watson Studio interface. At the top, there's a header bar with 'My Projects' and 'Watson Studio introduction'. On the left, there's a user profile icon for 'Admin'. To the right are icons for 'Launch IDE' and a '+' button. Below the header is a large, empty workspace area.

This screenshot shows a modal window titled 'Readme'. It contains a text area with the placeholder text: 'Document your project using standard Markdown syntax. See the [Markdown cheatsheet](#).'. There's a small pencil icon in the top right corner of the text area.

This screenshot shows the 'Assets' tab in the project dashboard. The top navigation bar includes 'Overview', 'Assets' (which is selected), 'Environments', 'Jobs', 'Deployments', 'Access Control', and 'Settings'. A search bar says 'What assets are you looking for?'. Below it, under 'Data assets', there's a table with columns: NAME, TYPE, CREATED BY, LAST MODIFIED, and ACTIONS. The message 'You don't have any Data assets yet.' is displayed. To the right, there's a file upload section with the instruction 'Drop files here or [browse](#) for files to upload.' and a 'Share' button. Under 'Notebooks', there's a table with columns: NAME, SHARED, SCHEDULED, STATUS, LANGUAGE, LAST EDITOR, LAST MODIFIED, and ACTIONS. One notebook named 'Data Exploration' is listed, created by Jacques Roy, last modified on 14 Oct 2019, and written in Python 3.6.

IBM Watson Studio Introduction

Projects / Watson Studio introduction

Add to project Watch later Share

Overview Assets Environments Jobs Deployments Access Control Settings

What assets are you looking for?

Data assets

NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
You don't have any Data assets yet.				

Notebooks

NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
Data Exploration				Python 3.6	Jacques Roy	14 Oct 2019	

Drop files here or [browse](#) for files to upload.

IBM Watson Studio Introduction

Projects / Watson Studio introduction

Add to project Watch later Share

Overview Assets Environments Jobs Deployments Access Control Settings

What assets are you looking for?

Data assets

NAME	TYPE

Notebooks

NAME	SHARE
Data Exploration	

Choose asset

Connection
The information necessary to create a connection to a data source.

AVAILABLE ASSET TYPES

Data	Connection NEW	Connected data	AutoAI experiment
Notebook	Dashboard	Visual Recognition ... NEW	Natural Language Cl...
Watson Machine Le...	Deep learning experi...	Modeler flow	Data Refinery flow
Streams flow	Synthesized neural n...	Function NEW	Decision Optimizatio...

Drop files here or [browse](#) for files to upload.

IBM Watson Studio Introduction

New connection Watch later Share

IBM services

Analytics Engine HDFS	Cloud Object Storage	Cloud Object Storage (infrastructure)	Cloudant
Cognos Analytics	Compose for MySQL	Compose for PostgreSQL	Db2
Db2 Big SQL	Db2 for i	Db2 for z/OS	Db2 Hosted
Db2 on Cloud	Db2 Warehouse	Informix	Object Storage OpenStack Swift (infrastructure)
PureData for Analytics	IBM Db2 warehouse database on Cloud		

Third-party services

Amazon Redshift	Amazon S3	Apache Hive	Cloudera Impala
Dropbox	FTP	Google BigQuery	Google Cloud Storage
Hortonworks HDFS	Looker	Microsoft Azure Data Lake Store	Microsoft Azure SQL Database
Microsoft SQL Server	MySQL	Oracle	Pivotal Greenplum
PostgreSQL	Salesforce.com	Snowflake	Sybase
Sybase IQ	Tableau	Teradata	

IBM Watson Studio Introduction

My Projects / Watson Studio introduction

Watch later Share

Environments

Define the runtime configuration for tools like the notebook editor, the model builder, or the flow editor and when you run Data Refinery flows.

You can use the default environment definitions or create custom environment definitions with different hardware and software configurations. Learn more.

Capacity Unit Hours (CUH) usage this month

- CUH

- CUH

used in this project

remaining

Which environment are you looking for?

Active environment runtimes

NAME	HARDWARE CONFIGURATION	TOOL	STARTED AT CAPACITY UNIT HOURS (CUH)	OWNER	ACTIONS

Environment definitions

New environment definition +

NAME	TOOL	HARDWARE CONFIGURATION	LANGUAGE	LAST MODIFIED	ACTIONS
Default R 3.6 S	Notebook	4 vCPU and 16 GB RAM	R 3.6	20 Aug 2019	

Projects / Watson Studio introduction

Launch IDE Add to project Watch later Share

Active environment runtimes

NAME	HARDWARE CONFIGURATION	TOOL	STARTED AT CAPACITY UNIT HOURS (CUH)	OWNER	ACTIONS
Default Python 3.6 XS - -GQTZYvcIS	2 vCPU and 8 GB RAM	Notebook	14 Oct 2019, 9:47:18 am 2.9 CUH	Jacques Roy	

Environment definitions

NAME	TOOL	HARDWARE CONFIGURATION	LANGUAGE	LAST MODIFIED	ACTIONS
Default R 3.6 S	Notebook	4 vCPU and 16 GB RAM	R 3.6	20 Aug 2019	
Default RStudio L	RStudio	16 vCPU and 64 GB RAM	R 3.6	24 May 2019	
Default RStudio M	RStudio	8 vCPU and 32 GB RAM	R 3.6	24 May 2019	
Default RStudio XS	RStudio	2 vCPU and 8 GB RAM	R 3.6	24 May 2019	
Default Python 3.6 Free	Notebook	1 vCPU and 4 GB RAM	Python 3.6	9 May 2019	
Default Python 3.6 S	Notebook	4 vCPU and 16 GB RAM	Python 3.6	9 May 2019	
Default Python 3.6 XS + DO	Notebook	2 vCPU and 8 GB RAM	Python 3.6	9 May 2019	
Default Python 3.6 XS	Notebook	2 vCPU and 8 GB RAM	Python 3.6	9 May 2019	

My Projects / Watson Studio introduction

Launch IDE Add to project Watch later Share

Overview Assets Environments Jobs Deployments Access Control Settings

Environments

Define the runtime configuration for tools like the notebook editor, the model builder, or the flow editor and when you run Data Refinery flows.

You can use the default environment definitions or create custom environment definitions with different hardware and software configurations. Learn more.

Capacity Unit Hours (CUH) usage this month
2.9 CUH 0 CUH
used in this project remaining ⓘ

Which environment are you looking for?

Active environment runtimes

NAME	HARDWARE CONFIGURATION	TOOL	STARTED AT CAPACITY UNIT HOURS (CUH)	OWNER	ACTIONS
Default Python 3.6 XS - -GQTZYvcIS	2 vCPU and 8 GB RAM	Notebook	14 Oct 2019, 9:47:18 am 2.9 CUH	Jacques Roy	

Environment definitions

NAME	TOOL	HARDWARE CONFIGURATION	LANGUAGE	LAST MODIFIED	ACTIONS
Default R 3.6 S	Notebook	4 vCPU and 16 GB RAM	R 3.6	20 Aug 2019	

The screenshot shows the 'Settings' tab selected in the top navigation bar of the Watson Studio interface. Below it, the 'Project information' section is displayed. It includes fields for 'Project name' (set to 'Watson Studio introduction') and 'Description' (with placeholder text 'Project description').

This screenshot shows the 'Storage' section with a Cloud Object Storage entry (599.95 KB used) and the 'Associated services' section, which currently displays a message: 'You don't have any Associated services yet.' A 'New token' button is visible in the top right of the associated services table.

This screenshot is similar to the previous one but includes a service catalog overlay on the right side. The catalog lists several services: Amazon EMR Spark, IBM Analytics Engine, Streaming Analytics, Dashboard, and Watson. The 'Add service' button is also visible in the top right of the catalog area.

Discovery Add a cognitive search and content analytics engine to applications. Add	Language Translator Translate text, documents, and websites from one language to another. Create industry or region-specific Add	Machine Learning IBM Watson Machine Learning - make smarter decisions, solve tough problems, and improve user outcome Add
Natural Language Classifier Natural Language Classifier uses advanced natural language processing and machine learning techniques to create Add	Natural Language Understanding Analyze text to extract meta-data from content such as concepts, entities, emotion, relations, sentiment Add	Personality Insights The Watson Personality Insights derives insights from transactional and social media data to identify personality Add
Speech to Text Low-latency, streaming transcription Add	Text to Speech Synthesizes natural-sounding speech from text. Add	Tone Analyzer Tone Analyzer uses linguistic analysis to detect three types of tones from communications: emotion. Add

Machine Learning

[Existing](#)

[New](#)

RESOURCE GROUP
All Resources ▾

LOCATION
All Locations ▾

CLOUD FOUNDRY ORG
jacquesr@us.ibm.com ▾

Existing Service Instance

Select service from the list

[Select](#)

[Cancel](#)

Machine Learning

Watch later Share

[Existing](#) [New](#)

Machine Learning

IBM Watson Machine Learning is a full-service IBM Cloud offering that makes it easy for developers and data scientists to work together to integrate predictive capabilities with their applications. The Machine Learning service is a set of REST APIs that you can call from any programming language to develop applications that make smarter decisions, solve tough problems, and improve user outcomes.

Features

Machine Learning features
 Take advantage of machine learning models management (continuous learning system) and deployment (online, batch, streaming). Select any of widely supported machine learning frameworks: TensorFlow, Keras, Caffe, PyTorch, Spark MLlib, scikit learn, xgboost and SPSS.

Wide choice of interfaces
 Use the command line interface and Python client to manage your artifacts. Extend your application with artificial intelligence through the Watson Machine Learning REST API.

Integration with Watson Studio
 Create and train machine learning models with the best tools and the latest expertise in a social environment built by and for data scientists.

Pricing Plan: Monthly Process shown above reflect the: United States

PLAN	FEATURES	PRICING
	Service instance 5 model deployments 5,000 predictions 50 capacity unit-hours (CUH) included: Capacity Type: <ul style="list-style-type: none"> • 1 (one) NVIDIA K80 GPU = 2 capacity units required per hour 	

PLAN	FEATURES	PRICING
 Lite	Service instance 5 model deployments 5,000 predictions 50 capacity unit-hours (CUH) included: Capacity Type: <ul style="list-style-type: none"> • 1 (one) NVIDIA K80 GPU = 2 capacity units required per hour • 1 (one) NVIDIA V100 GPU = 8 capacity units required per hour • 1 vCPU and 4 GB RAM = 0.5 capacity units required per hour • 2 vCPU and 8 GB RAM = 1 capacity units required per hour • 4 vCPU and 16 GB RAM = 2 capacity units required per hour • 8 vCPU and 32 GB RAM = 4 capacity units required per hour • 16 vCPU and 64 GB RAM = 8 capacity units required per hour Auto AI <ul style="list-style-type: none"> • 8 vCPU and 32 GB RAM = 20 capacity units required per hour Decision Optimization <ul style="list-style-type: none"> • 2 vCPU and 8 GB RAM = 30 capacity units required per hour 	Free

Creating an Account on IBM Watson Studio

IBM Watson gives you access to IBM Watson Studio, IBM Watson Knowledge Catalog, the data refinery, machine and deep learning, visual recognition models, dashboards and streams flows. At <https://dataplatform.cloud.ibm.com/> you can sign up for a free trial. When you sign up for an IBM Watson account you are automatically signed up for a free IBM Cloud account. Here you see the Watson applications that will be provisioned. IBM Watson Studio and IBM Watson Knowledge Catalog. If you already have an IBM cloud account then use your IBM ID to sign up for IBM Watson.

Otherwise, type your email address which will be used to create an IBM cloud account for you. On the next screen you are redirected to the IBM cloud registration page where you need to provide the typical basic information for an account then click create account.

Now check your email and confirm your account.

Now that you've registered for IBM cloud you can use those same credentials to sign-in. Next you'll see the process create the IBM Watson user account using your IBM cloud credentials and finally you'll see that your account was successfully created.

This IBM account has only one Associated IBM cloud account and one resource group.

If you have more than one associated account or one account with multiple resource groups then during the Watson Studio account setup you'll see this screen giving you the option to select an account and resource group to use.

IBM Cloud uses resource groups as a way for you to organize your account resources in customizable groupings so that you can quickly assign users access to more than one resource at a time. View the settings to verify the applications and services that are provisioned.

Now you're ready to start working in IBM Watson.

Jupyter Notebooks in Watson Studio Part 1

This video covers the basics for working with Jupyter Notebooks in Watson Studio.

Start in a Watson Studio project and add to the project a Notebook. Just provide a name in a description and create the Notebook.

Let's first load a file so you have some data to work with. From the filesy slide-out panel browse to select the file. After the file is added to the project its available to work with in this Notebook, just click insert to code and insert a panda's data frame.

Before running the Notebook it's a best practice to insert a cell at the top to describe what the Notebook does. Change the cell type to markdown so this cell will not be treated as code and then add the description.

Now you're ready to run the Notebook. The inserted code loads the data set into a data frame using your credentials for your cloud object storage instance and then displays the first five rows of the data set. Before returning to the project save the notebook.

On the assets tab you'll find the Notebook. If you open the Notebook it will be in read-only mode, but you can edit the Notebook and make changes.

For example, you can access the info panel and change the name of the Notebook and on the environment tab you could change the environment used to run the Notebook as well as stop or restart the runtime environment. If you'd like to share a read-only

version of the Notebook you can do that from here. You can select how much of the content you'd like to share and how you want to share the Notebook either through a link or social media. If you'd like to schedule the Notebook to run at a different time you can create a job.

Just provide a name for the job and select the scheduling options like specifying a date for the job to run and whether you'd like the job run to repeat. After you create and run the job you can see the status on the jobs tab in the project.

Jupyter Notebooks in Watson Studio Part 2

This video shows you how to create a Jupyter Notebook. Let's start by adding a data asset to the project. You can either browse to select files or drag files into the panel. Great, now the data file is uploaded to object storage and available as a data asset in this project.

Next, create a Notebook.

Provide a name and a description and then select the runtime to use when running this Notebook. Here you see the environments you could use. You'll learn more about environments later so for now just select the default Spark Python environment and verify the language and spark version. When you are ready create the Notebook. Now wait while the runtime environment is instantiated. Once the environment is ready, in the Notebook, access the data sources and locate the file. Click insert to code and choose how you want to insert the data. The choices in this drop-down box are dependent upon the language used in this Notebook and the file type. Notice that the inserted code includes the credentials you'll need to read the data file from the object storage instance. When you run the code the first five rows display.

Now let's take a closer look at environments on the environments tab you can define the hardware size and software configuration for the runtime associated with Watson Studio tools such as notebooks. You can see that there is one active environment runtime namely the runtime being used by the Notebook you just created and here are the other default environments. You can view any of the default environments to see a summary of the configuration and also create a new environment definition.

First provide a name in a description. If you select Spark for the type you'll see some additional configuration options. In this case just accept the defaults and choose Scala for the software version. When you are ready create the new environment. The environment is ready for you to use with a Notebook. To switch a notebook to use a different environment, you need to first stop the kernel.

Then you can change the environment and select the custom environment you just created and associate that with the Notebook.

Now open the Notebook in edit mode and wait for the new environment to be instantiated. Since this notebook was last saved using a different kernel you need to set the new kernel.

Let's delete the existing cell, locate the source data file, and insert a Spark session data frame. When you run the code the first five rows display.

Now you're ready to explore the community and find sample Notebooks and data sets to get started analyzing data.

Lab: Creating a Watson Studio Project with Jupyter Notebooks

Objective(s):

After completing this lab, you will be able to:

- Use Watson Studio service
- Create project in Watson Studio
- Add an interactive python notebook to a project in Watson Studio

Pre-requisite

You need an IBM Cloud account to create a project in Watson Studio. If you don't have an account created already, click and open this [link](#) and follow the instructions, to create an IBM Cloud account.

Exercise - Create a project on Watson Studio

If you have not created a Watson service before proceed with Task 1, otherwise go to Task 2

Task 1: Create Watson Studio Service:

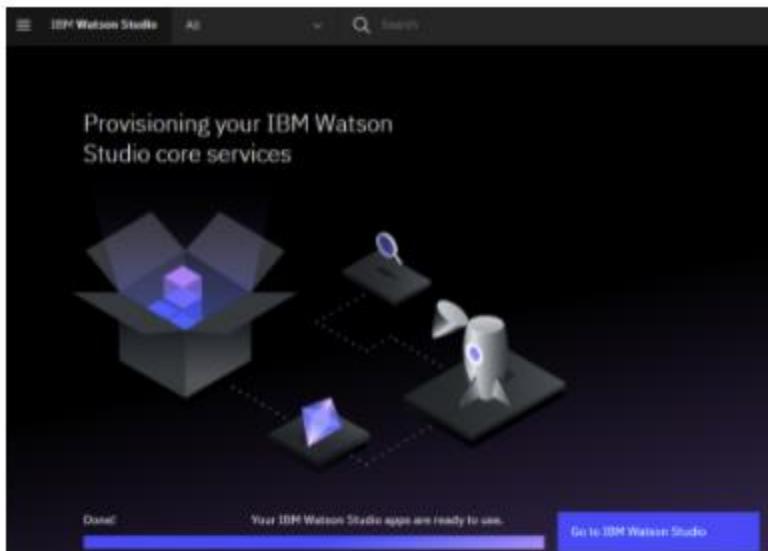
1. Click [here](#) to go to the IBM Cloud Watson Studio page. You will see the screen in the figure below. Click on the **Create** button.

The screenshot shows the IBM Cloud Catalog interface. On the left, there's a sidebar with 'Catalog / Services / Watson Studio'. Below it, a 'Create' button is highlighted in blue. The main area has tabs for 'Summary' and 'Create'. Under 'Summary', details are shown: Watson Studio, Region: London, Plan: Lite, Service name: Watson Studio-51, Resource group: Default. A 'Create' button is also present here. On the right, there's a table for selecting a pricing plan. The 'Lite' plan is selected, showing 1 authorized user, 10 capacity unit hours monthly limit, and environment = # of capacity units required per hour. It is labeled 'Free'. Other columns in the table include 'Plan', 'Features', and 'Pricing'. At the bottom of the table, there are buttons for 'Add to estimate' and 'View terms'.

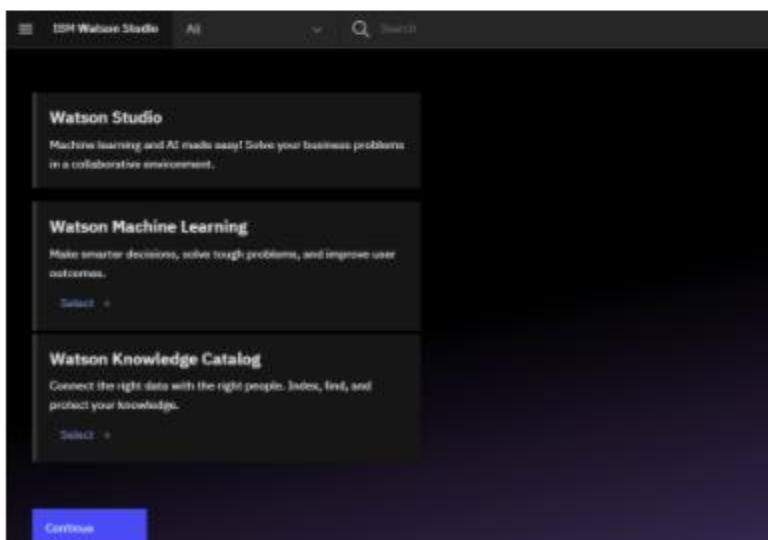
2. Now click **Get Started**.

The screenshot shows the details page for the Watson Studio service named 'Watson Studio-51'. The status is 'Active'. On the left, there's a 'Manage' tab and a 'Plan' section. The main area features a large circular icon with a purple silhouette of a person working at a desk. Below the icon, the text 'Watson Studio' is displayed. A message says 'Welcome to Watson Studio. Let's get started!' with a 'Get Started' button. On the right, there are 'Details' and 'Actions...' buttons.

3. Then click **Go to IBM Watson Studio**.

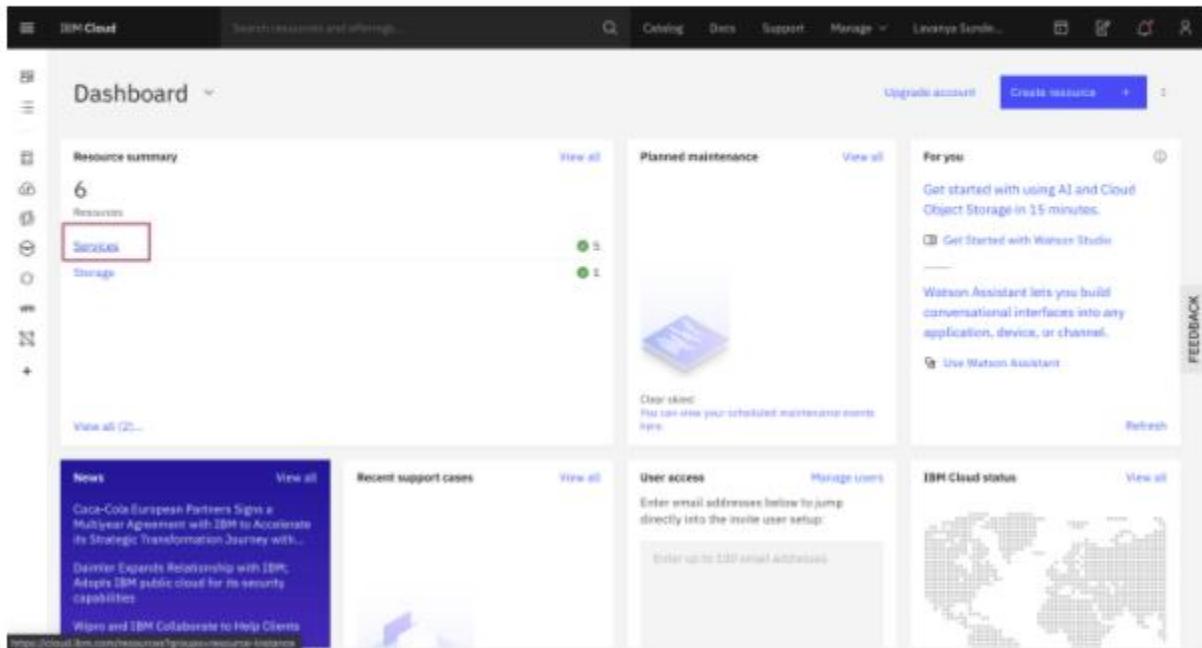


4. Then click **Continue**.



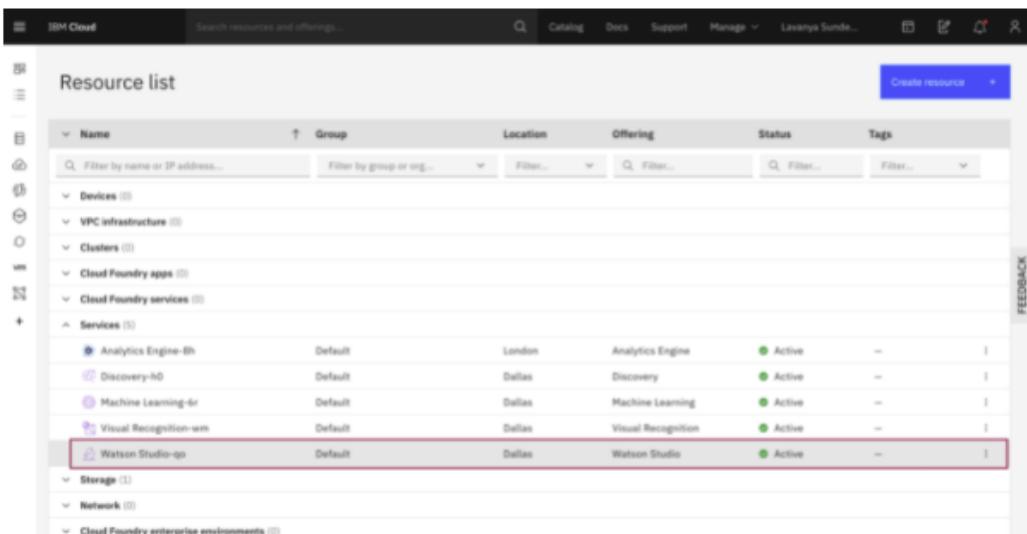
Task 2: Open Watson Studio

1. Go to the IBM Cloud Dashboard and click **Services**.



The screenshot shows the IBM Cloud Dashboard. On the left sidebar, under the 'Resource summary' section, there is a 'Services' button which is highlighted with a red box. The main content area displays various dashboard cards: 'Planned maintenance', 'For you' (with sections for AI and Cloud Object Storage, and Watson Studio), 'News' (with articles about Coca-Cola, Danone, and Wipro), 'Recent support cases', 'User access' (with a text input field for email addresses), and 'IBM Cloud status' (a world map). At the top right, there are buttons for 'Upgrade account', 'Create resource', and a feedback icon.

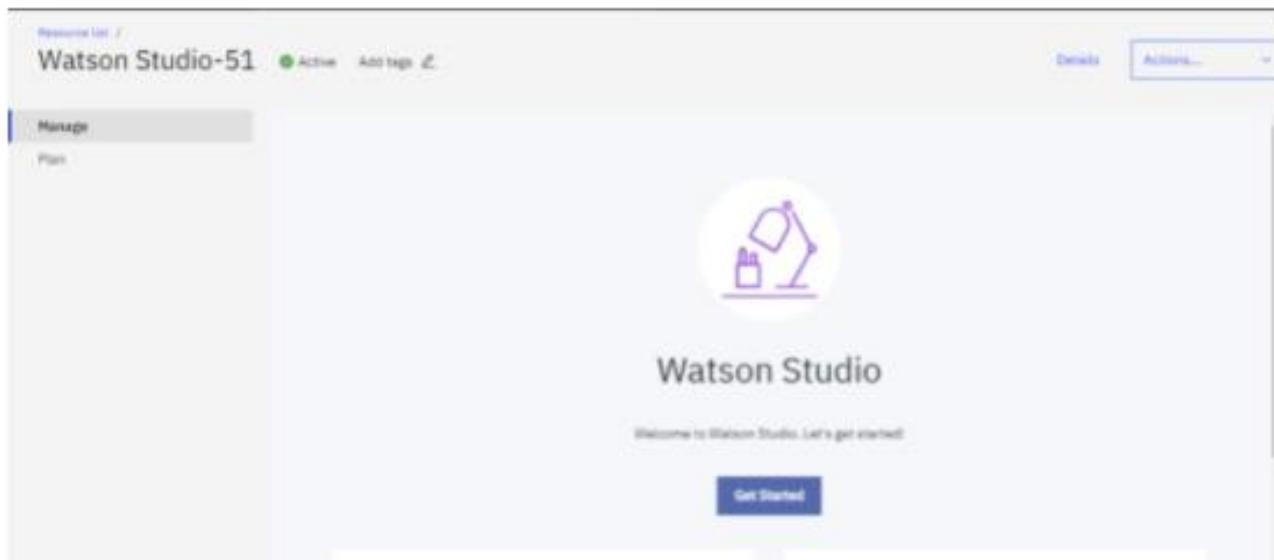
2. When you click on Services, all your existing services will be shown in the list. Click the Watson Studio service you created:



The screenshot shows the 'Resource list' page in the IBM Cloud interface. The left sidebar shows a tree view of resources: Devices, VPC infrastructure, Clusters, Cloud Foundry apps, Cloud Foundry services, and Services. Under 'Services', several services are listed: Analytics Engine-Bh, Discovery-h0, Machine Learning-6r, Visual Recognition-wm, and Watson Studio-go. The 'Watson Studio-go' service is highlighted with a red box. The main table lists these services with columns for Name, Group, Location, Offering, Status, and Tags. Each row has a 'View' and 'Edit' button at the end.

Name	Group	Location	Offering	Status	Tags
Analytics Engine-Bh	Default	London	Analytics Engine	Active	-
Discovery-h0	Default	Dallas	Discovery	Active	-
Machine Learning-6r	Default	Dallas	Machine Learning	Active	-
Visual Recognition-wm	Default	Dallas	Visual Recognition	Active	-
Watson Studio-go	Default	Dallas	Watson Studio	Active	-

3. Then click **Get Started**.



Task 3: Create a Project

1. Click on **Create a project**.

A screenshot of the IBM Watson Studio dashboard. The top navigation bar includes "IBM Watson Studio", a search bar, and an "Upgrade" button. The main area has a "Welcome" message and three sections: "Learn by example", "Work with data", and "Extend your capabilities". The "Work with data" section contains a "Create a project" button, which is highlighted with a red rectangle. On the left, there's a sidebar with "Quick navigation" sections for "Projects", "Deployments", and "Support". The central "Overview" section shows "Recent projects" and "Notifications" with a message "No recent projects".

2. On the **Create a project page**, click **Create an empty project**.

The screenshot shows the 'Create a project' page. At the top left is a 'Back' button. The main title is 'Create a project'. Below it is a sub-instruction: 'Choose whether to create an empty project or to preload your project with data and analytical assets. Add collaborators and data, and then choose the right tools to accomplish your goals. Add services as necessary.' There are two main options: 'Create an empty project' (selected) and 'Create a preloaded project'. The 'Create an empty project' section includes a circular icon with a blue gradient, a title 'Create an empty project', a description about adding data for preparation, analysis, or modeling, and a 'New' button for the AutoAI experiment tool. To the right is a 'USE TO' column with three items: 'Prepare and visualize data', 'Analyze data in notebooks', and 'Train models'.

3. Provide a **Project Name** and **Description**.

The screenshot shows the 'New project' creation form. It has three main sections: 'Define project details', 'Define storage', and 'Choose project options'. In the 'Define project details' section, there are fields for 'Name' (containing 'Project name') and 'Description' (containing 'Project description'). Both fields have red borders, indicating they are required. In the 'Define storage' section, there is a step-by-step guide: ① 'Select storage service' with a 'Add' button, and ② 'Refresh'. Below this is a note: 'Add an object storage instance, and then return to this page and click Refresh.' In the 'Choose project options' section, there is a checkbox 'Restrict who can be a collaborator' with a help icon, and a note below stating 'Project includes integration with Cloud Object Storage for storing project assets.' At the bottom right are 'Cancel' and 'Create' buttons.

4. You must also create storage for the project. Click **Add**

New project

The screenshot shows the 'New project' interface. On the left, there's a 'Define project details' section with fields for 'Project name' (containing 'Project A') and 'Project description'. On the right, under 'Define storage', there's a heading 'Select storage service' with a red box around the 'Add' button. Below it is a note: 'Add an object storage instance, and then return to this page and click Next step.' There's also a 'Refresh' button. At the bottom, there are 'Cancel' and 'Create' buttons.

5. On the Cloud Object Storage page, click **Create**.

The screenshot shows the 'Cloud Object Storage' creation page. On the left, there's a 'Pricing plan' section with a table comparing 'Lite' and '1 COG Service Instance' plans. The 'Lite' plan is free and includes 1 COG Service Instance, 25 GB/month storage, 2,000 Class A requests/month, 20,000 Class B requests/month, 10 GB/month data retrieval, and 50B of egress. It also applies to aggregate totals across all storage bucket classes. The '1 COG Service Instance' plan costs \$10/month and includes 1 COG Service Instance, 100 GB/month storage, 10,000 Class A requests/month, 100,000 Class B requests/month, 100 GB/month data retrieval, and 1TB of egress. On the right, there's a 'Summary' section with details: Region: Global, Plan: Lite, Service name: Cloud Object Storage-a, Resource group: Default. Below the summary is a 'Create' button, which is highlighted with a blue box, and a 'View terms' link.

6. On the New project page, note that the storage has been added, click **Refresh** and then click **Create**.

New project

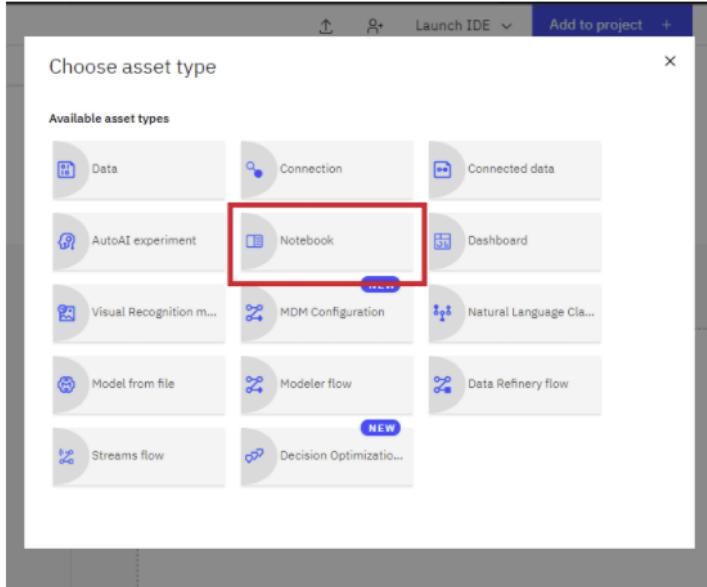
The screenshot shows the 'New project' interface. On the left, the 'Define project details' section includes fields for 'Name' (Project name) and 'Description' (Project description). On the right, the 'Define storage' section has a 'Select storage service' step with an 'Add' button and a note about adding an object storage instance. Below it is a 'Refresh' button. At the bottom, the 'Choose project options' section contains a checkbox for 'Restrict who can be a collaborator'. A note indicates project integration with Cloud Object Storage. At the very bottom are 'Cancel' and 'Create' buttons.

Task 4: Adding a Notebook to the Project:

1. Click **Add to project**.

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'IBM Watson Studio' and a search bar. Below it, a toolbar includes 'Upgrade', 'Launch IDE', and a settings icon. The main area shows a project overview with a title 'a' and a date 'Last Updated: Nov 25, 2020'. It also shows 'Assets' (0) and 'Collaborators' (1). A 'Readme' link is present. The 'Overview' tab is selected. On the right, there's a 'Recent activity' section. A blue rectangular box highlights the 'Add to project' button in the toolbar.

2. In the list of asset types, click **Notebook**.



3. On the New notebook page, click **Blank** and then add a name and optional description for the notebook. Specify the language as Python and runtime environment. Click **Create**.

New notebook

[Blank](#) [From file](#) [From URL](#)

Name
Type notebook name here

Description (optional)
Type your description here

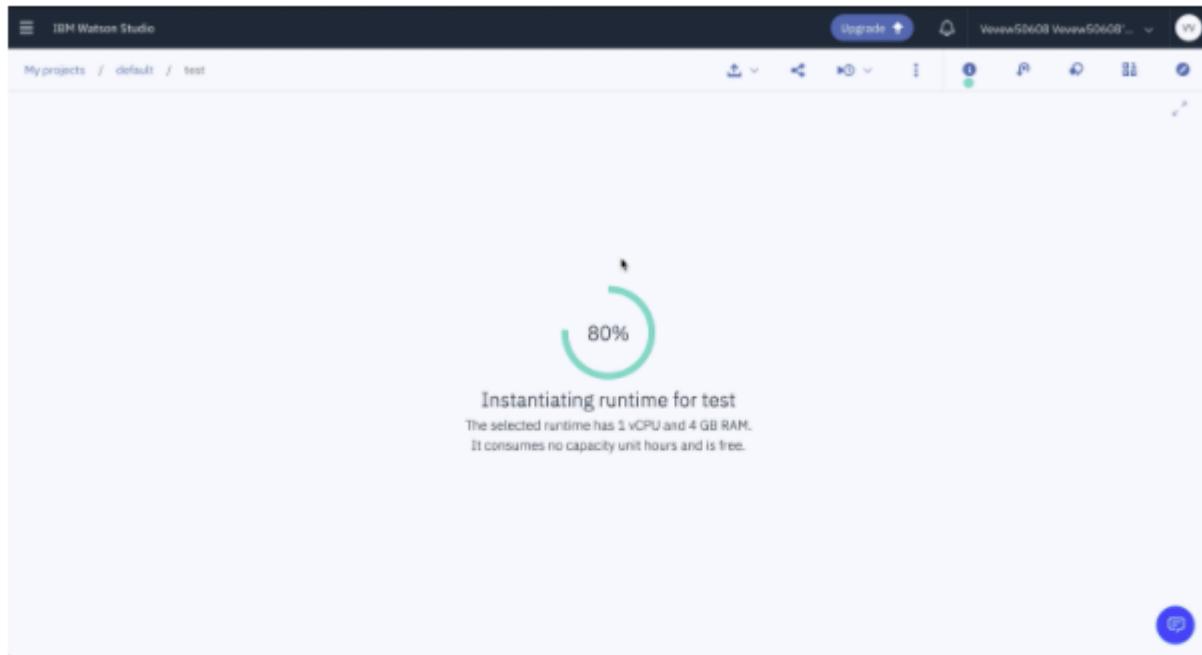
Select runtime
Default Python 3.7 XS (2 vCPU 8 GB RAM)

The selected runtime has 2 vCPU and 8 GB RAM.
It consumes 1 capacity unit per hour.
[Learn more](#) about capacity unit hours and Watson Studio pricing plans.

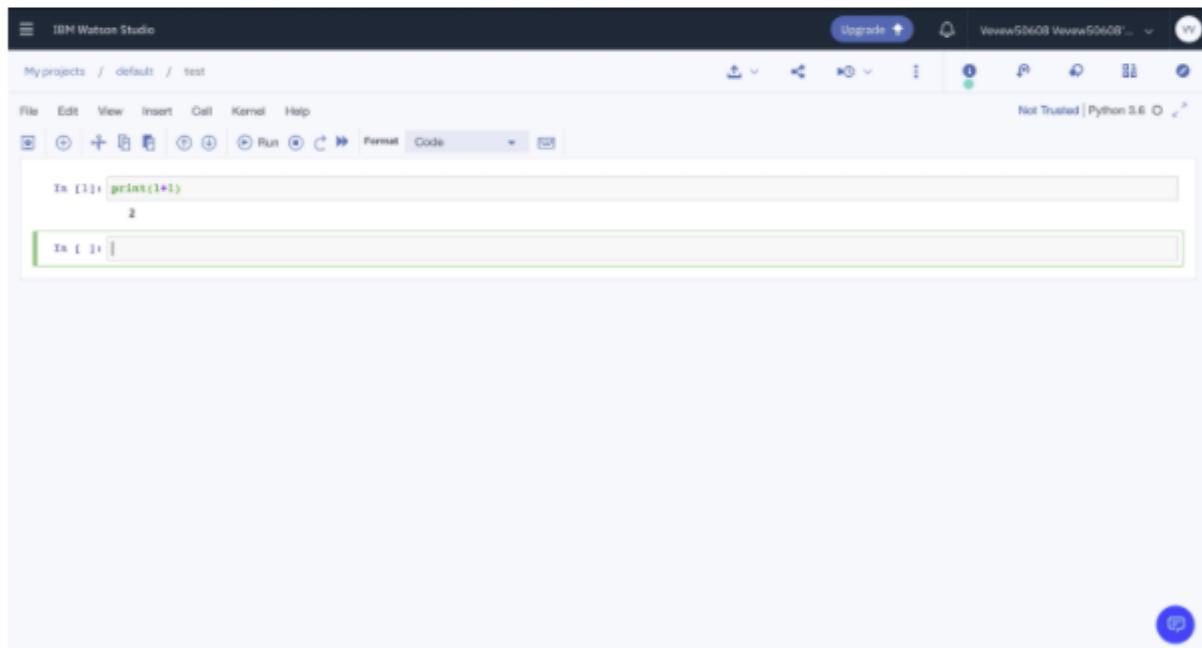
Language
 Python 3.7

[Cancel](#) [Create](#)

Just wait until the notebook appears. In case you are interested. The jupyter enterprise gateway has requested resources on the Kubernetes cluster IBM hosts for serving the jupyter kernel backing your notebook.



Now you're ready to code!



This concludes this tutorial.

Linking Github to Watson Studio

This video shows you how to connect your IBM Watson Studio account with your GitHub account. In Watson Studio navigate to your profile settings. On the integrations tab visit the link to generate a GitHub personal access token. Provide a descriptive name for the token and select the repo scope. Then generate the token. Copy the token, return to the GitHub integration settings, and paste the token. The token is validated when you save it to your profile settings. Now navigate to your projects. You enable GitHub integration at the project level. On the settings tab, simply scroll to the bottom and paste the existing GitHub repository URL. Once the URL is validated click connect. Go to the assets tab and open the Notebook you want to publish. Notice that this Notebook has the credentials replaced with X's. It's a best practice to remove or replace credentials before publishing to GitHub so this notebook is ready for publishing. You can provide the target path along with a commit message. You also have the option to publish content without hidden code which means that any cells in the notebook that began with the hidden cell comment will not be published. When you're ready click publish. The message tells you that the Notebook was published successfully and provides links to the Notebook, the repository, and the commit. Let's take a look at the commit so there's the commit and you can navigate to the repository to see the published Notebook. Lastly you can publish as a gist. Gists are another way to share your work on GitHub. Every gist is a git repository so it can be forked and cloned. There are two types of gists, public and secret. If you start out with a secret gist you can convert it to a public gist later and again you have the option to remove hidden cells. So that's the basics of Watson Studios GitHub integration.

Practice Quiz - Watson Studio

Bookmarked

Question 1

1/1 point (ungraded)

Fill in the blank: In Watson Studio, a _____ is how you organize your resources to achieve a particular goal. Resources can include data, collaborators, and analytic assets like notebooks and models.

Asset

Job

Notebook

Project



Submit

You have used 1 of 2 attempts

Reset

Question 2

1/1 point (ungraded)

Fill in the blank: It's a best practice to remove or replace _____ before publishing to GitHub.

Code cells

Charts

Markdown text

Credentials



Submit

You have used 1 of 2 attempts

Reset

✓ Correct (1/1 point)

Question 3

1/1 point (ungraded)

Which of the following do you need to create in order to publish a notebook to your GitHub repository?

Login credential

Apps

Access token

Profile



Submit

You have used 2 of 2 attempts

✓ Correct (1/1 point)

Question 4

1/1 point (ungraded)

Fill in the blank: If you'd like to schedule a notebook in Watson Studio to run at a different time you can create a(n) _____.

Job

Asset

API

Markdown cell



Submit

You have used 1 of 2 attempts

Reset

✓ Correct (1/1 point)

Question 5

1/1 point (ungraded)

Fill in the blank: On the environments tab you can define the _____.

- Software configuration
- Runtime configuration for flow editor
- Runtime configuration for notebook editor
- Hardware size
- All of the above



Submit

You have used 1 of 2 attempts

Reset

✓ Correct (1/1 point)

Question 6

1/1 point (ungraded)

Fill in the blank: When sharing a read only version of a notebook, you can choose to share _____.

- All content including code
- A permalink
- Only text and output
- All content, excluding sensitive code cells
- All of the above



Submit

You have used 1 of 2 attempts

Reset

Question 7

1/1 point (ungraded)

Fill in the blank: When working in a Jupyter Notebook, before returning to a project, it's important to _____.

Save your notebook

Insert cells

Insert to code

Run cells



Submit

You have used 1 of 2 attempts

Reset

✓ Correct (1/1 point)

Question 8

1/1 point (ungraded)

Fill in the blank: Before running a notebook, it's a best practice to _____ to describe what the notebook does.

Delete notebook cells

Refresh your page

Insert a cell at the bottom of the notebook

Insert a cell at the top of the notebook



Submit

You have used 1 of 2 attempts

Reset

✓ Correct (1/1 point)

Question 9

1/1 point (ungraded)

Fill in the blank: In the _____ tab you can define the hardware size and software configuration for the runtime associated with Watson Studio tools such as notebooks.

Environments

Settings

Overview

Assets



Submit

You have used 1 of 2 attempts

Reset

✓ Correct (1/1 point)

Question 10

1/1 point (ungraded)

Fill in the blank: IBM Cloud uses _____ as a way for you to organize your account resources in customizable groupings so that you can quickly assign users access to more than one resource at a time.

Resource groups

Catalogs

Projects

Services



Submit

You have used 1 of 2 attempts

Reset

✓ Correct (1/1 point)

IBM Watson Knowledge Catalog

Most organizations have huge amounts of data stored in many forms in various locations. Finding relevant data quickly and connecting disparate data sources can be challenging and time-consuming. [Watson Knowledge Catalog](#)

- [unites all information assets into a single metadata-rich catalog, based on Watson's understanding of relationships between assets and how they're being used and socialized among users in existing projects.](#)

Let's have a look at the overview of different tool categories that we've previously discussed.

- [Watson Knowledge Catalog corresponds to the Data Asset Management, Code Asset Management, Data Management, and Data Integration and Transformation.](#)
- [Watson Knowledge Catalog is a data catalog that is integrated with an enterprise data governance platform.](#)
- [It also merges the analytics capabilities of Watson Studio. The data catalog assists data scientists to easily find, prepare, understand, and use the data as needed.](#)
- [Watson Knowledge Catalog protects data from misuse and enables the sharing of assets with automated, dynamic masking of sensitive data elements. Data-profile visualizations, built-in charts and statistics help users to understand data assets.](#)
- [Seamless integration with Watson Studio helps data citizens to drive production of their data in a suite of powerful data science, AI, machine-learning and deep-learning tools.](#)
- [Joining with Watson Studio directs the building, training, and deploying of models. Users can interactively discover, cleanse, and prepare data with a built-in data refinery. Possible connections to more than 30 IBM and third-party data sources help to catalog and use your data in the original locations.](#)
- [IBM Watson Knowledge Catalog has various deployment choices on IBM Cloud™ and can be run anywhere with IBM Cloud Pak™ for Data. The latter is a fully-integrated data and AI platform built on Red Hat® OpenShift®](#)

Container base. It can be deployed easily into any public or private cloud or other enterprise platforms.

A catalog contains metadata about the contents of assets and how to access them. And a set of collaborators who need to use the assets for data analysis.

The metadata is stored in an encrypted IBM Cloud object storage instance.

- Any data that you want to store in the Cloud, you can upload to the cloud object storage of your choice, and then specify that object storage when you create the catalog. This split between where the data's metadata is stored and the actual location of the data is important. It means that you can keep your data where ever it is. You don't need to move it into the catalog because the catalog only contains metadata. You can have the data in unpremises data repositories in other IBM cloud services like Cloudant or Db2 on Cloud and in non-IBM cloud services like Amazon or Azure, in streaming data services or even dark data sources like PDFs.
- Included in the metadata is how to access the data asset. In other words, the location and credentials. That means that anyone who is a member of the catalog and has sufficient permissions can get to the data without knowing the credentials

or having to create their own connection to the data. Since the new catalog is empty, let's take a look at an existing catalog. On the Browse Assets tab you can see "recommendations", "highly rated assets", and "recently created assets", as well as a list of all the assets.

You can type a search term to find assets, and you can filter by asset type, such as Data Asset or Notebook. Or filter by tags that were assigned to the asset when it was added to the catalog. When you view an asset, you get a preview of the data and other information

like a description, ratings, tags, where the source is located, and any classifications. On the Access tab, those with permission can add members to view this particular asset. And the Review tab shows reviews and lets you contribute a review. When assets are added

to a catalog with Data Policies enabled, Watson Knowledge Catalog automatically profiles and classifies the content of the asset based on the values in those columns. The Profile

tab contains more detailed information about the inferred classifications. You can see the other possibilities for classifying each column and the confidence scores for those other possibilities. On the Lineage tab, you'll see the various events that Watson Knowledge

Catalog has captured that occurred in the lifecycle of this data asset, allowing you to trace what's happened to the asset since it was created. On the Access Control tab, you can see the current list of catalog members. You can also add members which is pretty similar to adding collaborators in a project. Most catalog members will likely have the editor role. The viewer role is intentionally restricted and only a select few users will have the admin role. Watson Knowledge Catalog includes capabilities to automatically mask sensitive data according to your organization's governance policies. For example, you can see in the diagram that the first name, last name, and gender data in the data set have been masked. You've learned how IBM Watson Knowledge Catalog can help organizations deal with their numerous data and other assets. In the next video we'll look at Data Refinery, a powerful tool for analyzing and preparing data.

Execution Environments

Data Asset Management

Data Management

Data Integration and Transformation

Data Visualization

Model Building

Model Deployment

Model Monitoring and Assessment

Code Asset Management

Development Environments

Watson Knowledge Catalog is a data catalog that is integrated with an enterprise data governance platform. It also merges the analytics capabilities of Watson Studio

Main features

- Find data
- Catalog data
- Govern data
- Understand data
- Power data science
- Prepare data
- Connect data

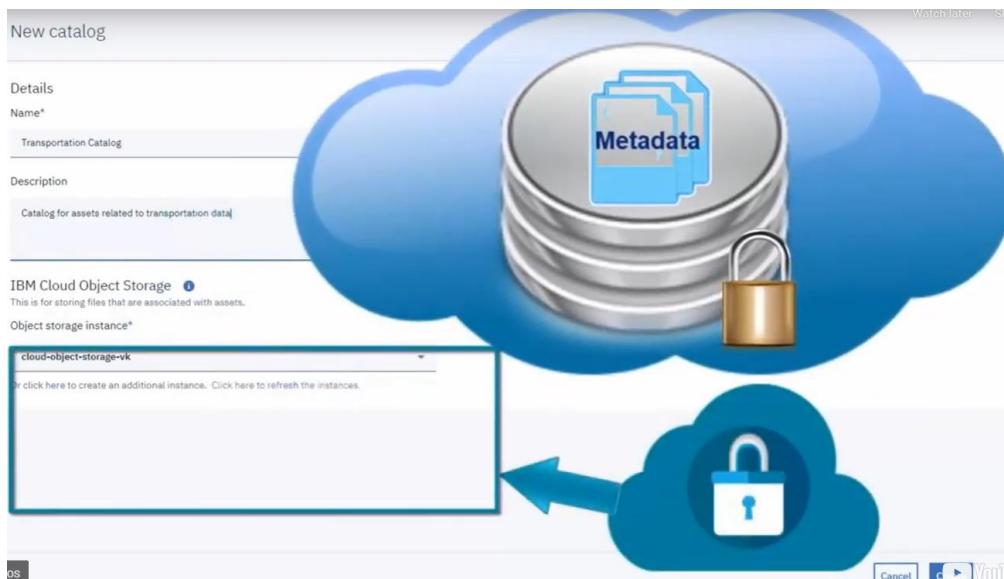


Watson Knowledge Catalog protects data from misuse and enables the sharing of assets with automated, dynamic masking of sensitive data elements

Main features

- Find data
- Catalog data
- Govern data
- Understand data
- Power data science
- Prepare data
- Connect data
- Deploy anywhere

Users can interactively discover, cleanse, and prepare data with a built-in data refinery



IBM Watson Knowledge Catalog

New catalog

Details

Name* Transportation Catalog

Description Catalog for assets related to transportation data

IBM Cloud Object Storage ⓘ
This is for storing files that are associated with assets.

Object storage instance*

cloud-object-storage-vk

Or click here to create an additional instance. Click here to refresh the instances.

VIDEOS

Watch later

IBM Watson Knowledge Catalog

New catalog

Details

Name* Transportation Catalog

Description

IBM Cloudant
IBM Db2 on Cloud
IBM Db2 Warehouse on Cloud
Amazon Web Services
Microsoft Azure

Twitter

PDF

Watch later



Catalogs / MyCo Data Catalog Watch later Add to Catalog

Browse Assets Access Control Settings

MyCo Data Catalog

What assets are you looking for?

Filter Any type Any business term Any tag Clear all

Watson Recommends	Highly Rated	Recently Added	Collapse	
Data Asset Housing Data Set.csv Owner: Madison Gooch Added: Feb 13, 2018 2:04 PM Tags: housing, people ★★★★★ 0 reviews	Data Asset PERSON Owner: Jay Limburn Added: Dec 13, 2017 5:20 AM Tags: disco..., SAMP... ★★★★★ 0 reviews	Data Asset ALL_2012.csv Owner: Tina Zeng Added: May 31, 2018 12:14 PM Tags: 2012, insur... ★★★★★ 0 reviews	Data Asset All US Banking Branches Owner: Jay Limburn Added: Feb 19, 2018 8:33 AM Tags: fss, bankin..., bran... ★★★★★ 1 review	Data Asset 2017 Loan F Owner: Un... Added: Apr... ★★★★★ 1

Showing 25 of 103 items

Name	Owner	Tags	Business Terms	Type	Date Added
2017 Loan Risk Data.csv	Unavailable			Data Asset	Apr 9, 2018

MORE VIDEOS 2017 Small Business Banking Loans Jay Limburn banking loan + 1 more Data Asset Feb 19, 2018 You!

Catalogs / MyCo Data Catalog Add to Catalog

Browse Assets Access Control Settings

MyCo Data Catalog

Q What assets are you looking for?

Filter Any type Any business term Any tag Clear all

<input type="checkbox"/> Connection	W <input type="checkbox"/> Dashboard	Highly Rated	Recently Added	Expand	
Showing 25 <input type="checkbox"/> Data Asset		Owner	Tags	Business Terms	Type Date Added
<input type="checkbox"/> N <input type="checkbox"/> Model		Unavailable			Data Asset Apr 9, 2018
<input type="checkbox"/> 2017 Small Business Banking Loans	 Jay Limburn	banking loan + 1 more			Data Asset Feb 19, 2018
<input type="checkbox"/> 2017 U.S. Auto Claims Satisfaction Study J.D...	 Ricardo Buglio	insurance auto			Data Asset Mar 15, 2018
<input type="checkbox"/> 2017_sales_data.csv	 Tina Zeng				Data Asset Apr 18, 2018
<input type="checkbox"/> ALL_2012.csv	 Tina Zeng	2012 insurance			Data Asset May 31, 2018
<input type="checkbox"/> ANCESTRY	 Jay Limburn	discovered SAMPLES	Account		Data Asset Dec 13, 2017
<input type="checkbox"/> AWS Data Warehouse	 Jay Limburn				Connection Dec 13, 2017
MORE VIDEOS  All US Banking Branches  Jay Limburn	fss banking + 1 more				Data Asset Feb 19, 2018 

Catalogs / MyCo Data Catalog Add to Catalog

Browse Assets Access Control Settings

MyCo Data Catalog

Q What assets are you looking for?

Filter Any type Any business term Any tag Clear all

<input type="checkbox"/> Watson Recommends	Highly Rated	Recently Added	Expand	
Showing 25 of 103 items		Owner	Tags	Type Date Added
<input type="checkbox"/> Name		Unavailable		Data Asset Apr 9, 2018
<input type="checkbox"/> 2017 Loan Risk Data.csv				Data Asset Feb 19, 2018
<input type="checkbox"/> 2017 Small Business Banking Loans	 Jay Limburn	banking loan + 1 more		Data Asset Mar 15, 2018
<input type="checkbox"/> 2017 U.S. Auto Claims Satisfaction Study J.D...	 Ricardo Buglio	insurance auto		Data Asset Apr 18, 2018
<input type="checkbox"/> 2017_sales_data.csv	 Tina Zeng			Data Asset May 31, 2018
<input type="checkbox"/> ALL_2012.csv	 Tina Zeng	2012 insurance		Data Asset Dec 13, 2017
<input type="checkbox"/> ANCESTRY	 Jay Limburn	discovered SAMPLES	Account	
<input type="checkbox"/> AWS Data Warehouse	 Jay Limburn			Connection Dec 13, 2017
MORE VIDEOS  All US Banking Branches  Jay Limburn	fss banking + 1 more			Data Asset Feb 19, 2018 

Catalogs / MyCo Data Catalog Watch later Add to Catalog

Browse Assets Access Control Settings

MyCo Data Catalog

What assets are you looking for?

Filter Type Any business term Any tag Clear all

Showing 25 Data Asset

	Owner	Tags	Business Terms	Type	Date Added
<input type="checkbox"/> N	Unavailable			Data Asset	Apr 9, 2018
<input type="checkbox"/> 21	Jay Limburn	banking loan + 1 more		Data Asset	Feb 19, 2018
<input type="checkbox"/> 21	Ricardo Buglio	insurance auto		Data Asset	Mar 15, 2018
<input type="checkbox"/> 2017_sales_data.csv	Tina Zeng			Data Asset	Apr 18, 2018
<input type="checkbox"/> ALL_2012.csv	Tina Zeng	2012 insurance		Data Asset	May 31, 2018
<input type="checkbox"/> ANCESTRY	Jay Limburn	discovered SAMPLES	Account	Data Asset	Dec 13, 2017
<input type="checkbox"/> All US Banking Branches	Jay Limburn	fss banking + 1 more		Data Asset	Feb 19, 2018
<input type="checkbox"/> AnalyticsReports.pdf	Carmen Ruppach	Reports		Data Asset	Mar 16, 2018
<input type="checkbox"/> Arron Countries	Arron La	sensitive confident...		Data Asset	Sep 27, 2018

MORE VIDEOS [play icon]

Catalogs / MyCo Data Catalog Watch later Add to Catalog

Browse Assets Access Control Settings

MyCo Data Catalog

What assets are you looking for?

Filter Type Any business term Tag Clear all

Showing 25 of 25 items

	Owner	Business Terms	Type	Date Added
<input type="checkbox"/> Name				
<input type="checkbox"/> ANCESTRY	Jay Limburn	Account	Data Asset	Dec 13, 2017
<input type="checkbox"/> CONVERSION_RATE	Jay Limburn	Data Asset	Dec 13, 2017	
<input type="checkbox"/> COUNTRY	Jay Limburn	Data Asset	Dec 13, 2017	
<input type="checkbox"/> CUST_WIN_BACK	Jay Limburn	Data Asset	Dec 13, 2017	
<input type="checkbox"/> EDUCATION	Jay Limburn	Data Asset	Dec 13, 2017	
<input type="checkbox"/> EMPLOYEE_SURVEY_TARGETS	Jay Limburn	Data Asset	Dec 13, 2017	
<input type="checkbox"/> EMP_POSITION_LOOKUP	Jay Limburn	Data Asset	Dec 13, 2017	
<input type="checkbox"/> GENDER_LOOKUP	Jay Limburn	Data Asset	Dec 13, 2017	
<input type="checkbox"/> MARITAL_STATUS	Jay Limburn	Data Asset	Dec 13, 2017	

MORE VIDEOS [play icon] [video icon]

Catalogs / MyCo Data Catalog / CONVERSION_RATE

[Watch later](#) [Add to Catalog](#) [Share](#)

Overview Access Review Profile Lineage

DATA ASSET
CONVERSION_RATE

Description
CONVERSION_RATE contains currency exchange values.

Added: Dec 13, 2017 5:31 AM Format: CSV

Business Terms
There are no terms available for this asset.

Tags
banking discovered GOSALES

Reviews
★★★★★ 1 review

Connection
Source: Db2 Warehouse on Cloud-z9
Source type: Db2 Warehouse

Schema: 4 Columns 624 Rows
Preview: 624 rows Last refresh: 12 minutes ago [Refresh](#)

COUNTRY_CODE	Type: Integer	CONVERSION_YEAR	Type: Smallint	CONVERSION_MONTH	Type: Smallint	CONVERSION_TO_LOCAL	Type: Double
1003		2010		1		1	
1003		2010		2		1	
1003		2010		3		1	
1003		2010		4		1	
1003		2010		5		1	
1003		2010		6		1	
1003		2010		7		1	
1003		2010		8		1	

[Remove](#) [Download](#) [Add to Project](#)

Overview [Access](#) Review Profile Lineage

DATA ASSET
CONVERSION_RATE

Asset owner Jay Limburn

Privacy Public
All catalog members can find and use the asset

Members

Q Who are you looking for?

Showing 1 member

<input type="checkbox"/> Name	Email	Edit
 SHARYN RICHARD	sharyn.richard@us.ibm.com	Edit

[Add member](#)

[Remove](#) [Download](#) [Add to Project](#)

IBM Watson Knowledge Catalog

Catalogs / MyCo Data Catalog / CONVERSION_RATE

[Watch later](#) [Add to Catalog](#) [Share](#)

Overview Access [Review](#) Profile Lineage

DATA ASSET
CONVERSION_RATE

Overall Rating
5.0
★★★★★ 1 review

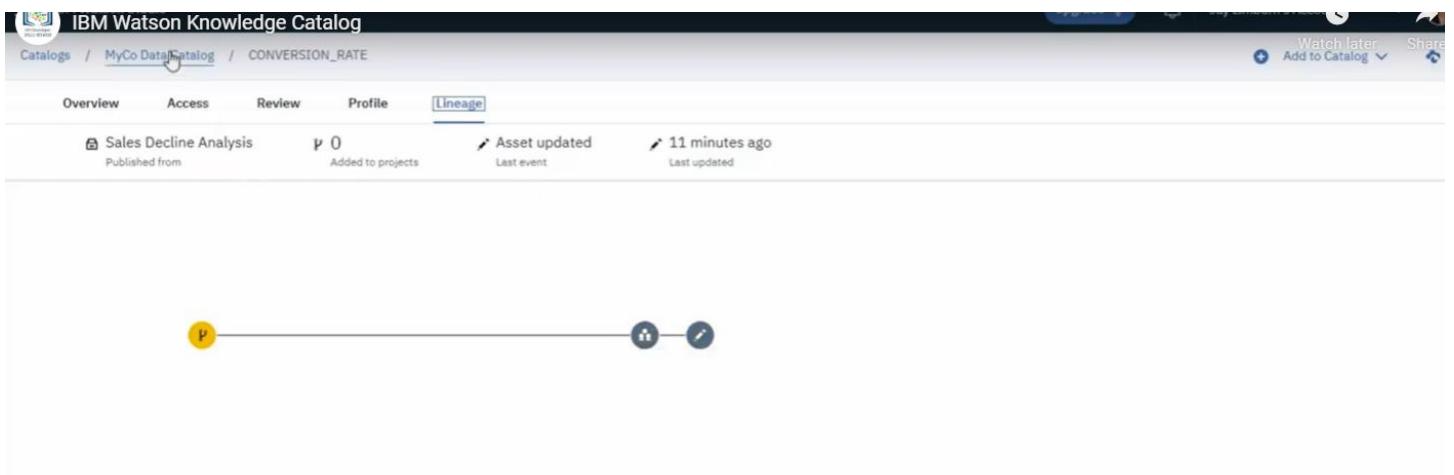
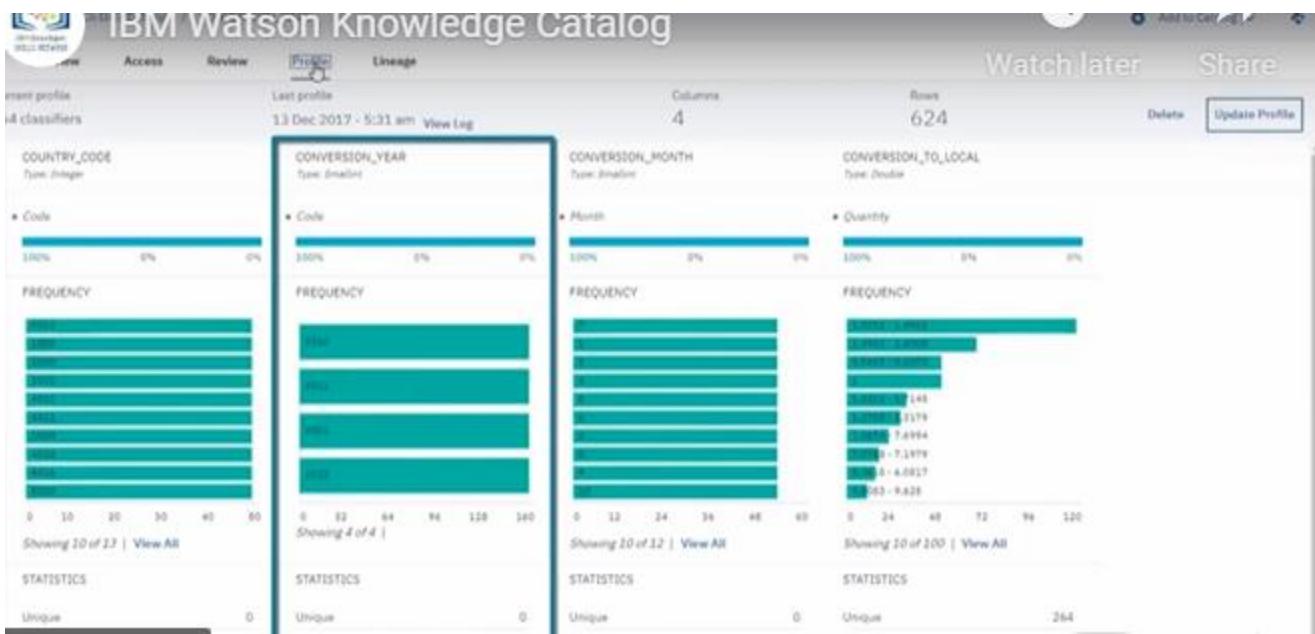
Review Summary
 (1)
 5 (0)
 4 (0)
 3 (0)
 2 (0)
 1 (0)

[My Review](#)

 SHARYN RICHARD | Mar 16, 2018
★★★★★
This asset contains some great data on customer conversion rates.

[Edit](#) [Delete](#)

[Remove](#) [Download](#) [Add to Project](#)



IBM Watson Knowledge Catalog

Catalogs / MyCo Data Catalog

Watch later Add to Catalog Share

Browse Assets Access Control **Settings**

Who are you looking for? Show: All

Collaborators: showing 56 collaborators

Name	Email	Role	Date Added
Deirdre Clyne	clyned@us.ibm.com	Admin	Mar 29, 2018
William Mathews	bmathews@us.ibm.com	Admin	Feb 28, 2018
Sonali Surange	ssurange@us.ibm.com	Admin	Mar 6, 2018
paul taylor	paultay@us.ibm.com	Admin	Jan 17, 2018

Add Collaborator

Access Groups: showing 0 groups

Name	Members	Role	Date Added
------	---------	------	------------

Add Access Group

No access groups available

Catalogs / MyCo Data Catalog

Watch later Add to Catalog Share

Browse Assets Access Control **Settings**

Who are you looking for? Show: All

Collaborators: showing 56 collaborators

Name
Deirdre Clyne
William Mathews
Sonali Surange
paul taylor

Add Collaborators as Viewers

Collaborators (0)

Role: Viewer

Having trouble adding collaborators? [Cancel](#) [Add](#)

No access groups available

The screenshot shows the IBM Watson Knowledge Catalog interface. At the top, there are navigation links: 'Browse Assets', 'Access Control', 'Settings', 'Watch later', and 'Share'. A search bar at the top left contains the placeholder 'Who are you looking for?'. Below the search bar, there's a sidebar titled 'Collaborators' showing 66 collaborators, with a list of names and their profile icons. Another sidebar titled 'Access Groups' shows 0 groups. The main content area has a title 'Add Collaborators as Viewers' and a sub-section 'Collaborators (3)' with a list item 'ARHAN RIZZI GARNETTE'. Below this, there are role selection fields: 'Role: Viewer' (selected), 'Admin', and 'Editor'. There are also 'Having' dropdowns for 'Editor' and 'Owner'. At the bottom right of the modal are 'Cancel' and 'Add' buttons.

The screenshot shows the IBM Watson Catalog interface. The top navigation bar includes 'IBM Watson', 'Projects', 'Tools', 'Catalog', 'Community', and 'Services'. Below the navigation, the path is 'logs / Transform Catalog / Customer Orders.csv'. The main content area has tabs: 'Overview' (selected), 'Access', and 'Profile'. Under 'Overview', there's a section for 'Data Asset' with the name 'Customer Orders.csv'. To the left, there are sections for 'Description' (no description available) and 'Tags' (subPersonalInfo). In the center, there's a table titled 'Schema: 15 Columns | 5000+ Rows | 4 Columns anonymized'. A tooltip for the 'Anonymized' column explains: 'The values and format of data in this column are substituted by this policy: Substitute Personal Information only'. The table lists columns: CITY, FIRST_NAME, LAST_NAME, GENDER, AGE, and MARITAL_STATUS. The first few rows of data are visible, showing various city names like Plymouth, Leipzig, and Jaipur, and gender codes like M and F.

Data Refinery

Data scientists often end up spending a lot of time doing mundane tasks like cleansing, shaping and preparing data.

Typically these tasks are roadblocks for starting the more enjoyable part of analyzing the data sets or building and training machine learning models.

This is because data sets typically are not in a format that can be readily used.

They first need to be cleansed, refined before they are useable by a data scientist.

IBM Data Refinery

- addresses this issue and simplifies the task of refining data and

its workflows. It provides a self-service data preparation environment where you can quickly analyze, cleanse and prepare data sets.

- Data refinery is available with Watson Studio on public cloud, private cloud and desktop.

In the rest of the video we will walk through a scenario and see Data Refinery in action.

In this scenario we will use Data Refinery to find the best deals using data about discounts offered over time.

We will then automate the analysis to run on a regular schedule.

Before the Data Scientist starts, she looks at the data distribution and notices that the in Sale column is missing data.

She visualizes the offer column and notices that it contains valuable information about discounts. Many fields contain the percent of information, some contain references to previous price indicating a new reduced price being available. She decides to derive sale from offer.

She uses a conditional decrease operation to derive if the product is on sale.

Next she uses a filter operation to eliminate deals that are not on sale

She then wants to pick up the bargains. She uses the replace substring operation and provides a pattern that extracts the discounts from the offer.

After converting the discount values to a decimal she can visually see the discounts that were available.

She needs to find the months that offered the best deals.

She visualizes the dateUpdated and notices that the date field has a variety of formats, some with dashes some with slashes and some with months as text. She hopes that data Refinery can normalize the data and extract a month.

She uses the convert column operation to convert to date and selects ymd.

Next she extracts month and creates a derived column called discountMonth.

The data now represents all brands and products providing sales and the month the offer was available. The data scientist is only interested in her preferred brands.

Over time she has built a list of preferred brands and has imported the data in her project. Data Refinery provides relational transformations such as left, inner right, full, semi and anti-join.

To ensure that the data only contains her preferred brand she uses a semi-join operation which narrows the brands to match her preferences. She then selects the keys for the join and the resulting fields. The visual results now confirms that the brands match the preferences.

To find the best possible deals she needs to perform some aggregations. Several features determine a good deal. She is interested in the best offer and duration when the discounts are active. Aggregating the sale data will help understand the deals. She groups the columns by brand and discountMonth and calculates the maximum discount. Finally she sorts the result in descending order

Data refinery is now displaying the best deals by brand preferences and the duration which the offer is available. The last step is to execute the analysis on the full dataset. She starts the full analysis, which she can monitor for the completion status.

It's time to automate the analysis which runs on a regular basis.

The data in the database can grow over time.

She uses a personalized runtime to match the larger data volumes and sets a schedule for automation.

The hourly schedule reads

from updated data from the database and writes to the target table.

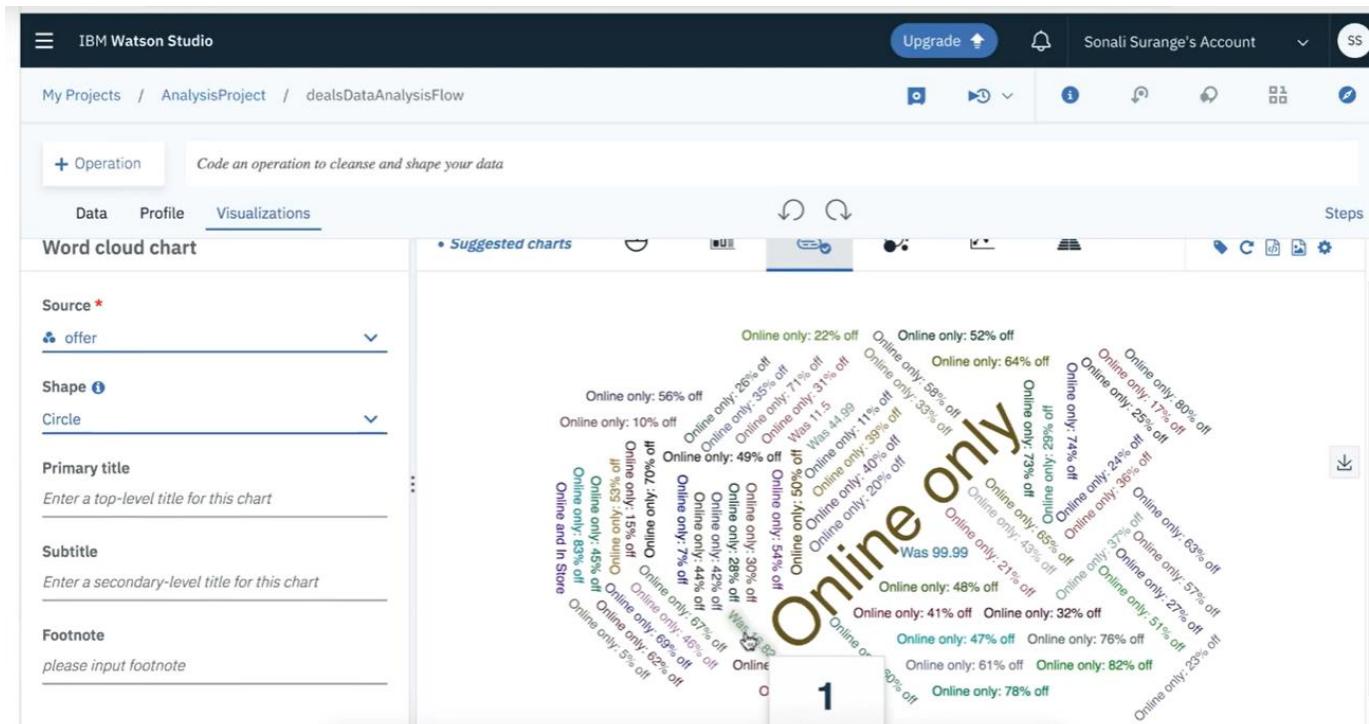
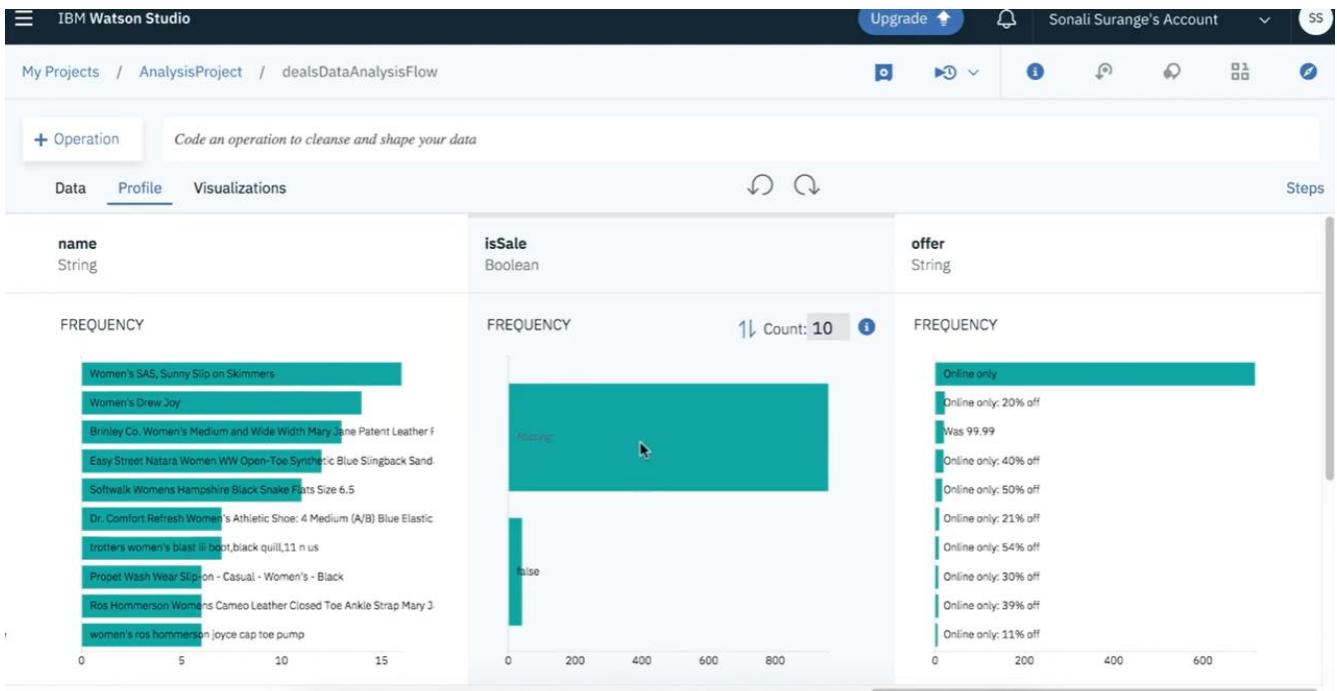
Data Refinery has helped her uncover deals in the raw data through a small set of operations and transformations with the bulk of the work done for her.

Simplifying Data Preparation

- Cleansing, Shaping, and Preparing data take up a lot of Data Scientist's time
- These tasks come in the way of the more enjoyable parts of Data Science: analyzing data and building ML models
- Data sets are typically not readily consumable. They need to be refined and cleansed
- IBM Data Refinery simplifies these tasks with an interactive visual interface that enables self-service data preparation
- Data Refinery comes with Watson Studio – on Public/Private Cloud and Desktop

The screenshot shows the IBM Watson Studio Data Refinery interface. At the top, there is a navigation bar with links for 'Data Refinery', 'IBM Watson Studio', 'Upgrade', 'Sonali Surange's Account', and a search bar. Below the navigation bar, the URL is displayed as `dataplatform.cloud.ibm.com/shaper?project_id=bac6667d-6a90-4c3b-8bce-7298b1fadb6d&flow_id=0612e9a2-ff97-45cb-8227-b06...`. The main area features a data preview table with the following columns:

COLUMN1	dateUpdated	timeUpdated	brand	categories	col
String	String	String	String	String	Str
1 AWpykE4QJbEilcB6Re54	2019 May, 1	2019-05-01T08:17:11Z	The Highest Heel	Clothing, Shoes, Womens Shoes, All Womens Shoes	Bla
2 AWpyg8JyAGTnQPR7wIev	2019 May, 1	2019-05-01T08:03:02Z	SoftWalk	Clothing, Shoes, Womens Shoes, All Womens Shoes	Bla
3 AWpx3JHC0U_gzG0hjpe	2019 May, 1	2019-05-01T05:20:01Z	City Classified	Clothing, Shoes, Womens Shoes, All Womens Shoes	Bla
4 AWpx9kE6AGTnQPR7v_oZ	2019 May, 1	2019-05-01T05:14:42Z	Madden Girl	Clothing, Shoes, Womens Shoes, All Womens Shoes	Bla
5 AWpx9zsdsM263mwCq-0pC	2019 May, 1	2019-05-01T05:14:31Z	City Classified	Clothing, Shoes, Womens Shoes, All Womens Shoes	Bla
6 AWpx7qRIM263mwCq-0j3	2019 May, 1	2019-05-01T05:13:42Z	SummitFashions	Clothing, Shoes, Womens Shoes, All Womens Shoes	Rec
7 AWpx8a1uAGTnQPR7v_ln	2019 May, 1	2019-05-01T05:11:19Z	SALE	Clothing, Shoes, Womens Shoes, All Womens Shoes	Bla
8 AWpx7wzuJbEilcB6RTr8	2019 May, 1	2019-05-01T05:09:16Z	FitFlop	Clothing, Shoes, Womens Shoes, All Womens Shoes	Urb
9 AWpx6oUKM263mwCq-0hj	2019 May, 1	2019-05-01T05:07:16Z	Aerosoles	Womens Shoes, Clothing, All Womens Shoes, Shoes	Wh
10 AWpx57bwJbEilcB6RTni	2019 May, 1	2019-05-01T05:06:01Z	New Balance	Clothing, Shoes, Womens Shoes, All Womens Shoes	Gg
11 AWpx9dYIM263mwCq-0oR	2019 May, 1	2019-05-01T05:02:38Z	Callisto	Clothing, Shoes, Womens Shoes, All Womens Shoes	Bla



Follow the video for step by step instructions on data refinery!

SPSS Modeler Flows in Watson Video

In this video, we will take a look at an easy-to-use, graphical way to build machine learning models

and pipelines.

SPSS Modeler Flows is a part of Watson Studio, which was inspired by another product, IBM

SPSS Modeler.

We'll discuss that product in a later unit.

Let's have a look again at the overview of different tool categories.

Modeler flows include some data management capabilities, as well as tools for data preparation,

visualization, and model building.

All flows are created using a drag-and-drop editor and consist of "nodes" of various types, with data "flowing" from one node to the next according to their connections.

A sample Modeler flow shown here includes two data source nodes shown in purple on the

left; type, aggregate, filter, merge, filler, and partition nodes in the middle; 2 model building nodes shown in pentagons.

Once a flow is executed and the models are built, the upside-down pentagon "model nuggets"

are created.

They can be used to see information about the models and to get predictions for new data.

And the three green square nodes on the right provide model evaluation information in the

form of tables and charts.

You can build your SPSS Modeler flows by dragging different types of nodes from the left, the

part of the screen called the "palette," to the "canvas," the main part of the screen.

Each flow starts with one or more data sources located in the “Import” group, and can include some or all other types of nodes.

Watson Studio provides some sample flows to help new users.

In the Drug Study example shown here, we are using a small artificial data set.

The target variable is a categorical field, “Drug,” that has five categories, and there are several predictor variables.

This flow creates a new “derived” field by dividing the values of one of the predictors by values of another one, and at the end builds a small neural network model and a decision tree model.

When a user clicks the “Run” button on the top panel, denoted by a triangle, the flow is executed and the models build.

This is reflected in the new pentagon nodes, called “model nuggets,” that display under each model node.

If you click on the three dots in the upper right corner of one of those nodes and select “View Model”, you will see various types of model information.

By connecting new data sources to the model nugget, you can get predictions on new data.

The first window in the model viewer shows model accuracy and related measures, such as precision and recall.

This toy data example enabled us to get perfect accuracy, which is normally not the case with real life data.

The Confusion Matrix view shows how model predictions on the training data matched the observed target values.

Once again, in this toy example all cases were classified correctly.

We can also look at Model Information, which displays a table that tells us more about the details of the model.

Feature Importance displays a diagram that indicates the relative predictive strength of various model inputs.

Finally, the Network Diagram gives a visual representation of the neural network model we built.

On the left is the input layer, with units corresponding to each continuous predictor and each category of the categorical predictors, plus a bias unit that is usually present in each layer of a neural network.

In the middle, we see a “hidden layer” with 7 units, or neurons, and a bias unit.

On the right is the output layer with 5 units corresponding to the five target categories.

Controls on the right and bottom of the diagram enable some interactive exploration of the

model.

The colors of the connections between units indicate the values of the weights on those connections.

We can also look at the decision tree model built using the C5 algorithm.

A Model Information table and Feature Importance chart appear as before.

Additionally, a Top Decision Rules table is displayed.

Decision tree models are popular because they have a special structure that makes it easy

to explain predictions or extract decision rules.

The tree diagram is also displayed.

On the left side of the canvas, we see a part of the model palette that can be used in the flows.

At the top are “Auto Classifier” and “Auto Numeric” nodes that can be used for categorical and continuous targets, respectively.

Those nodes will build several kinds of models and pick the best one based on a certain criterion.

Later, we will talk about the AutoAI feature of Watson Studio; AutoAI takes this capability

to the next level by automatically finding not only the best model, but an entire data pipeline, which includes various data transformations.

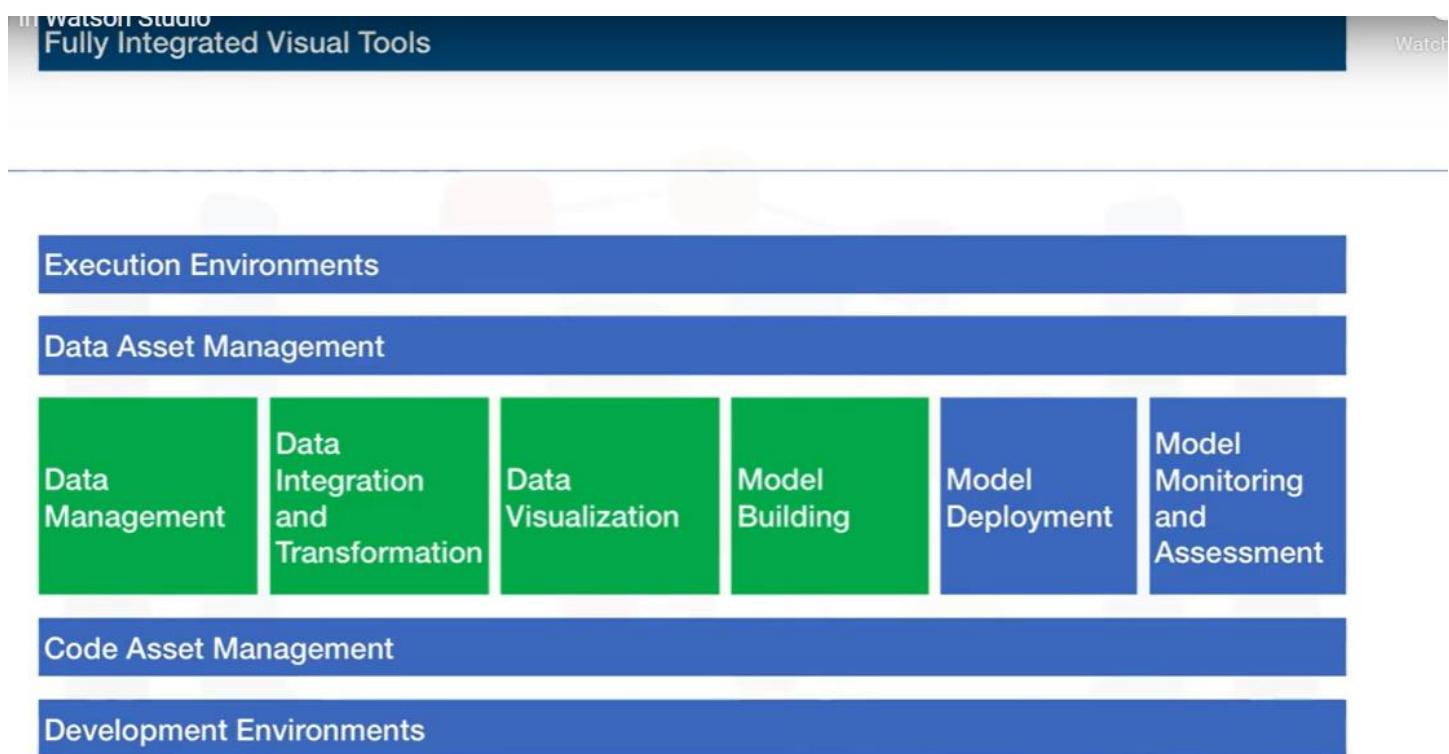
In this video, you've learned how Modeler Flows in Watson Studio can help analysts to create powerful machine learning pipelines using a graphical interface without the need to write any code.

This feature was based on IBM SPSS Modeler.

Next, after completing a lab to give you hands-on experience with this powerful technology,

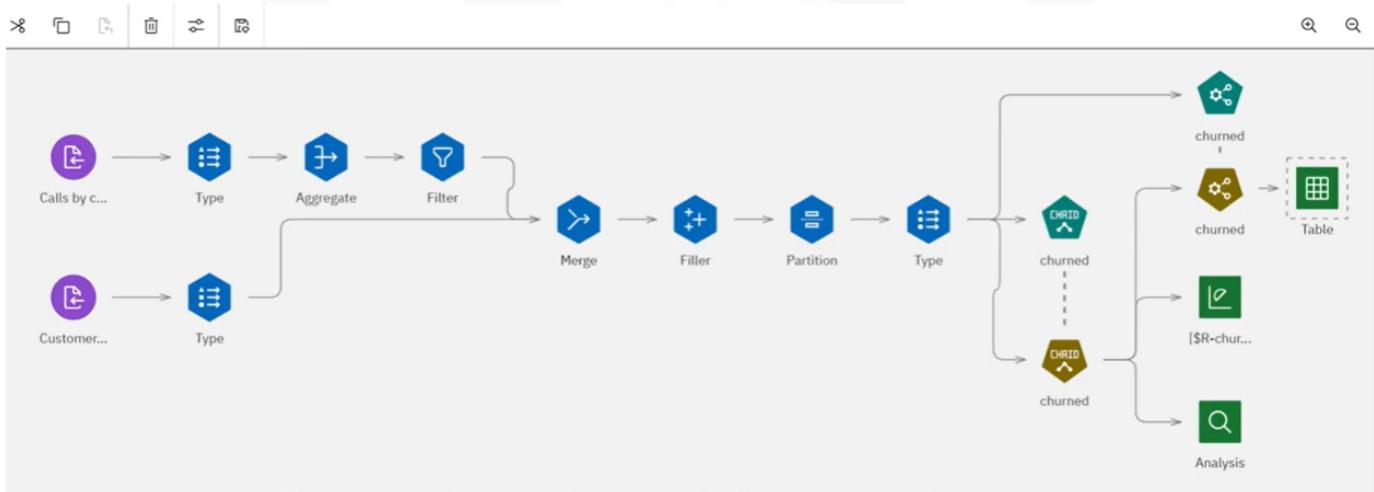
we will take a look at two other IBM products that can be used for Data Science: IBM SPSS

Modeler and IBM SPSS Statistics.



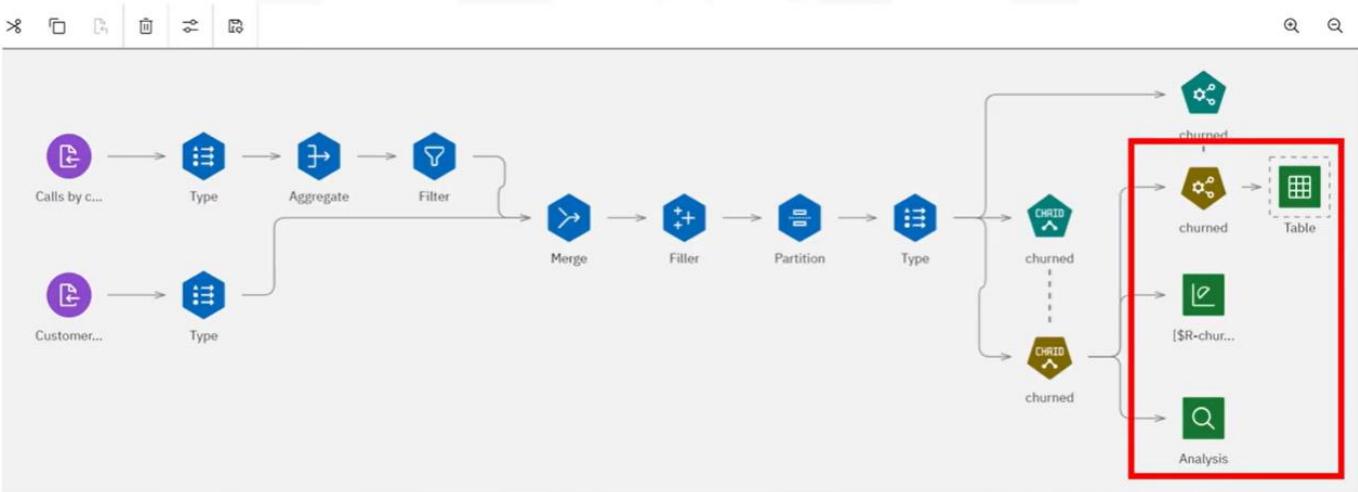
Modeler flows include some data management capabilities, as well as tools for data preparation, visualization, and model building.

An example flow

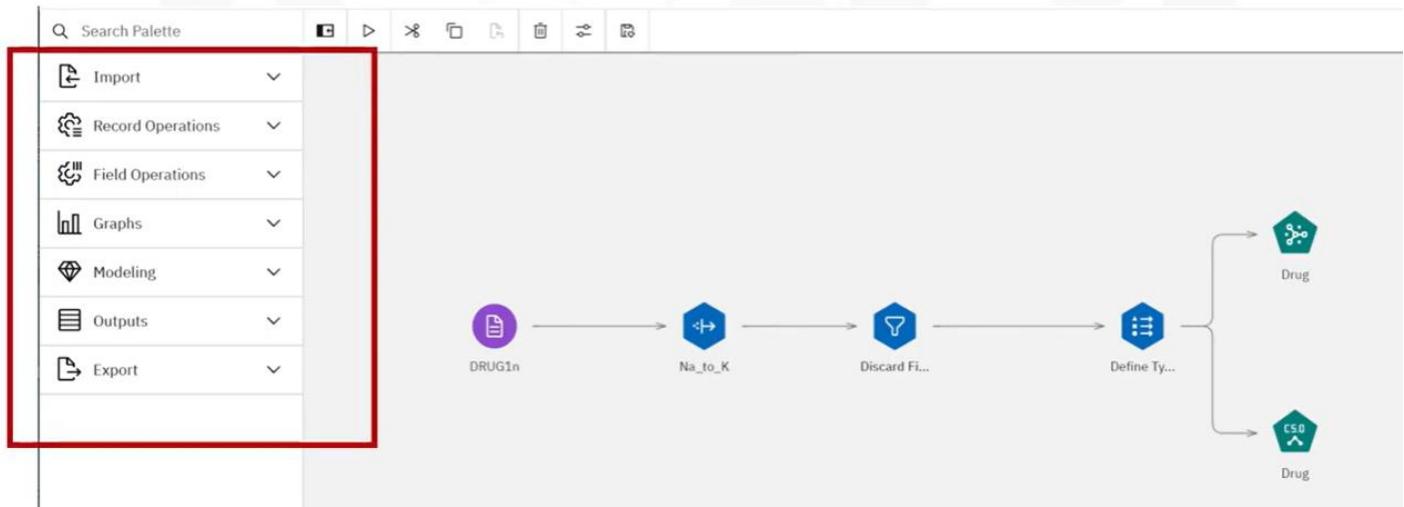


They can be used to see information about the models and to get predictions for new data

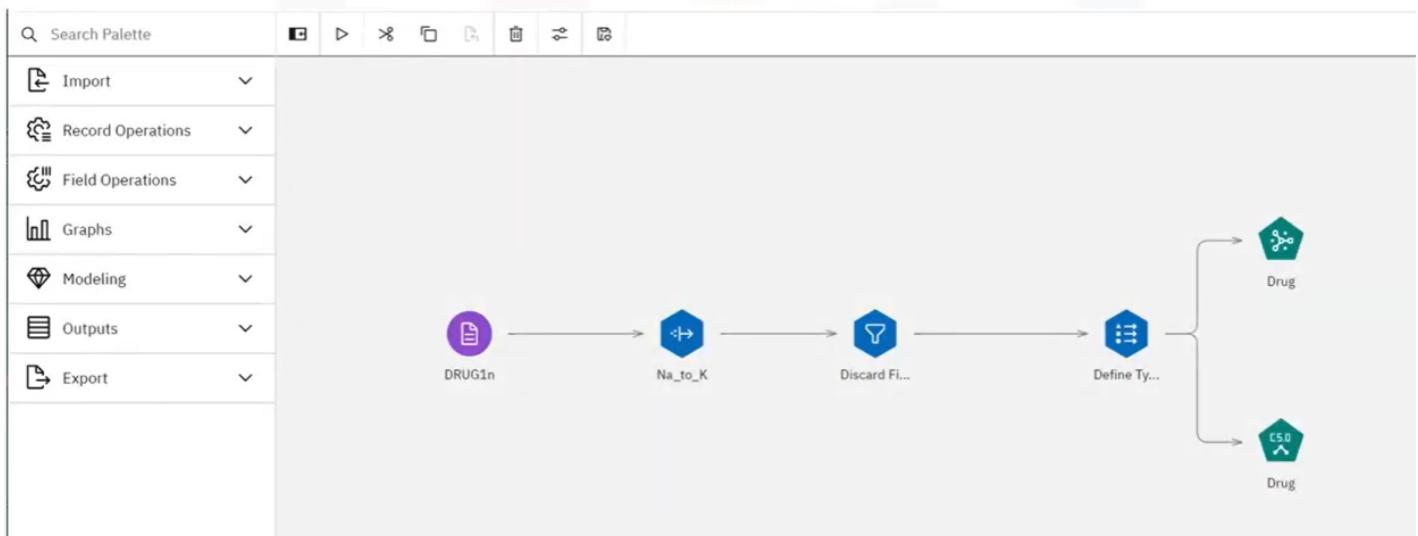
An example flow



Types of nodes

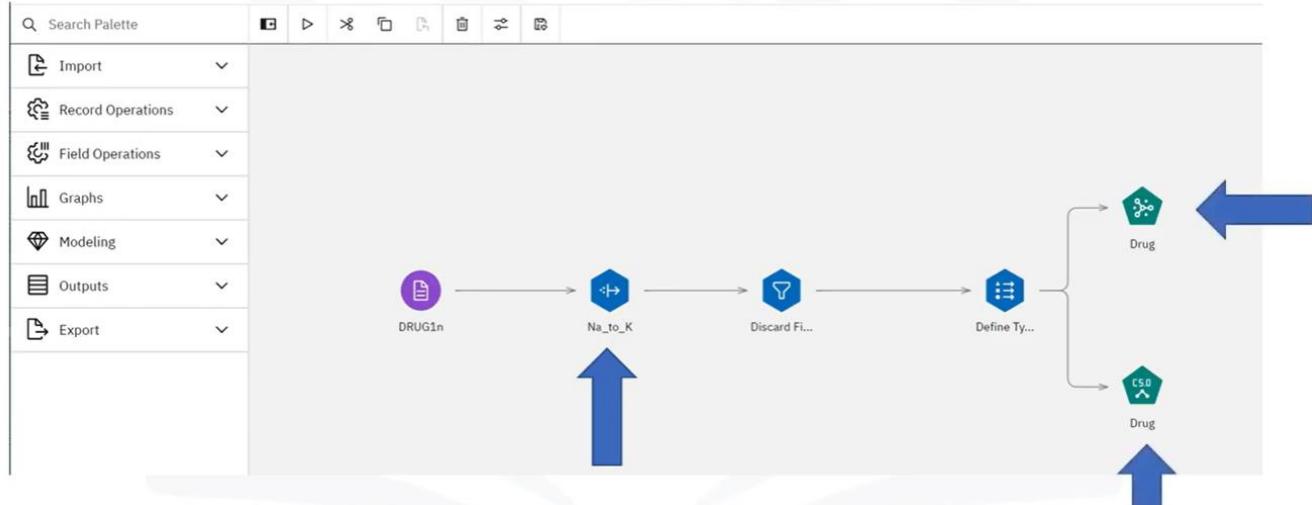


SPSS Modeler Flows

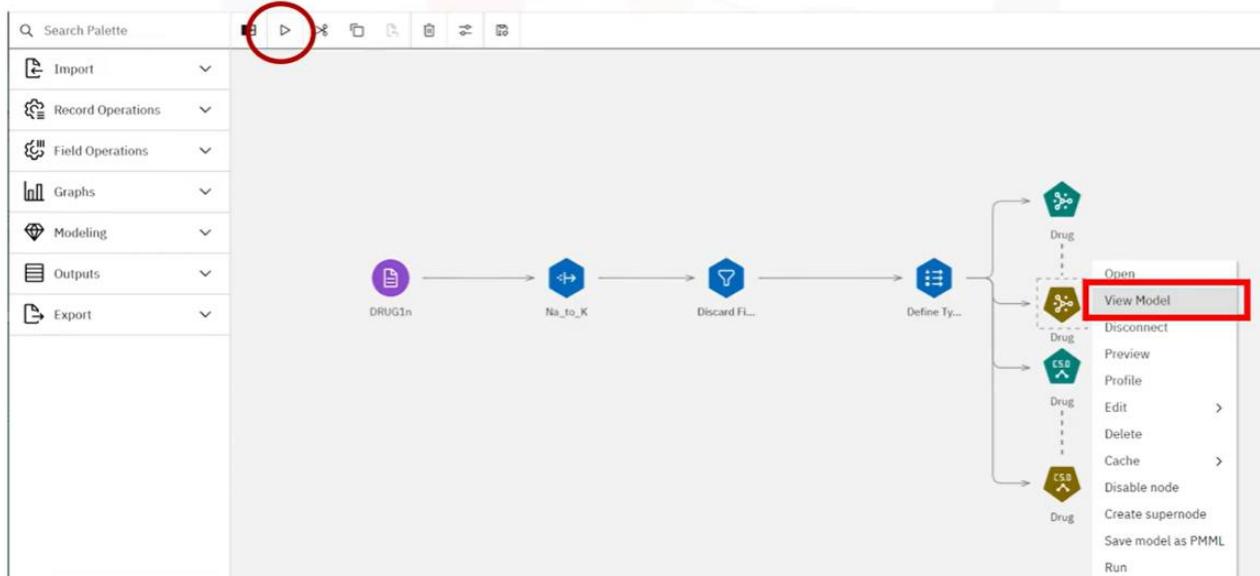


Watson Studio provides some sample flows to help new users

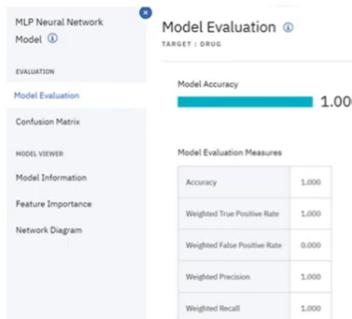
SPSS Modeler Flows



SPSS Modeler Flows

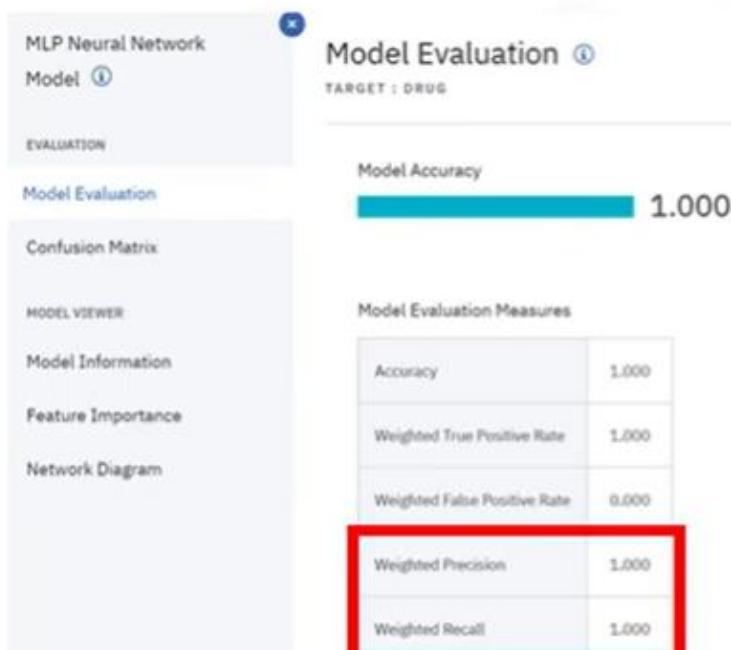


Examining models

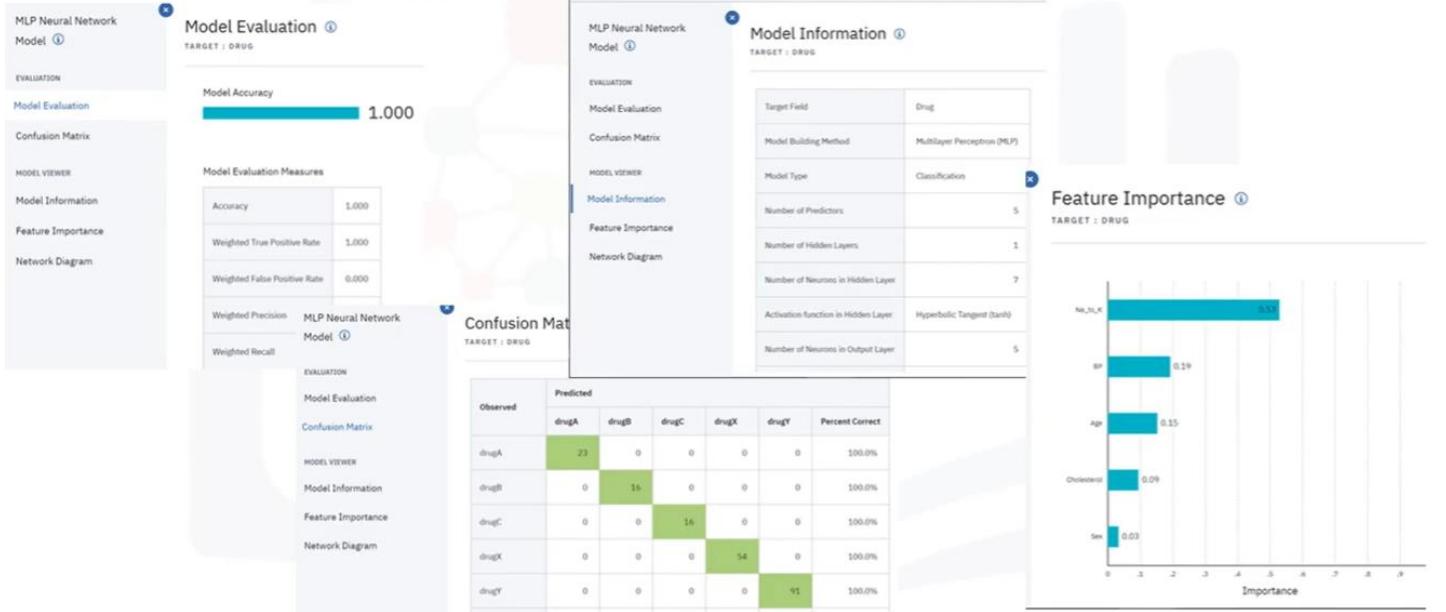


By connecting new data sources to the model nugget,
you can get predictions on new data

Examining models



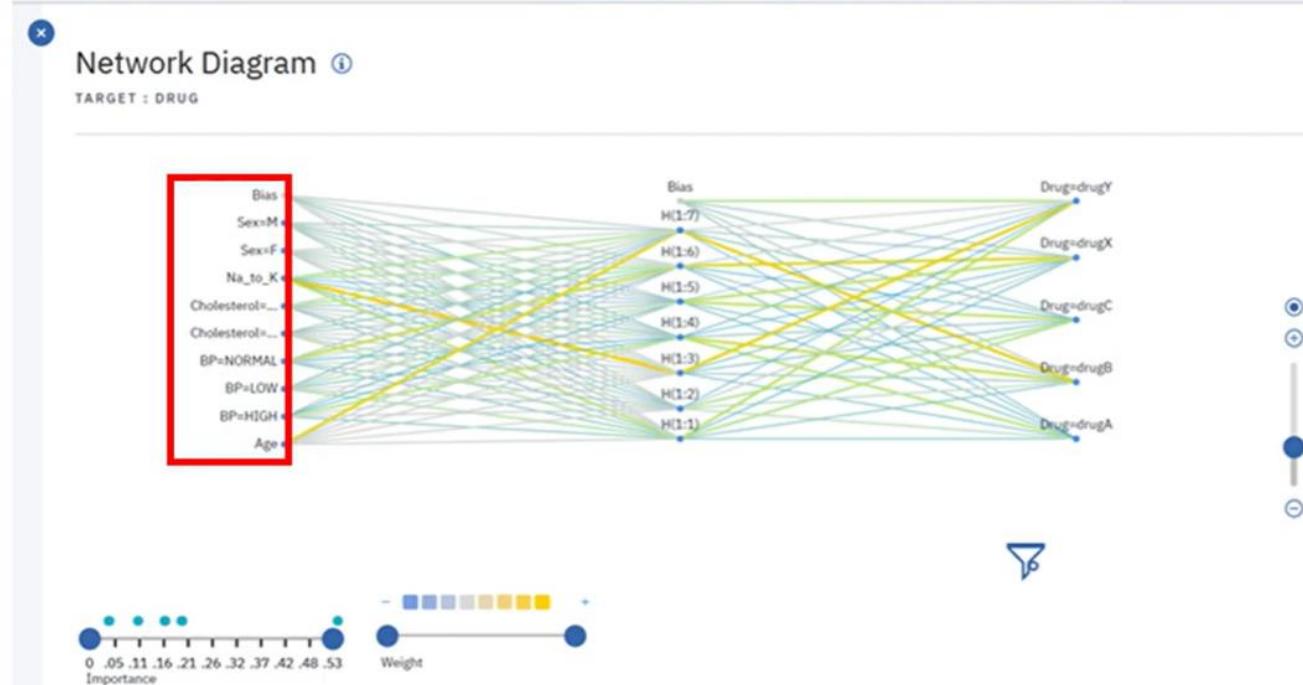
Examining models



IBM Developer

SKILLS NETWORK

Network diagram

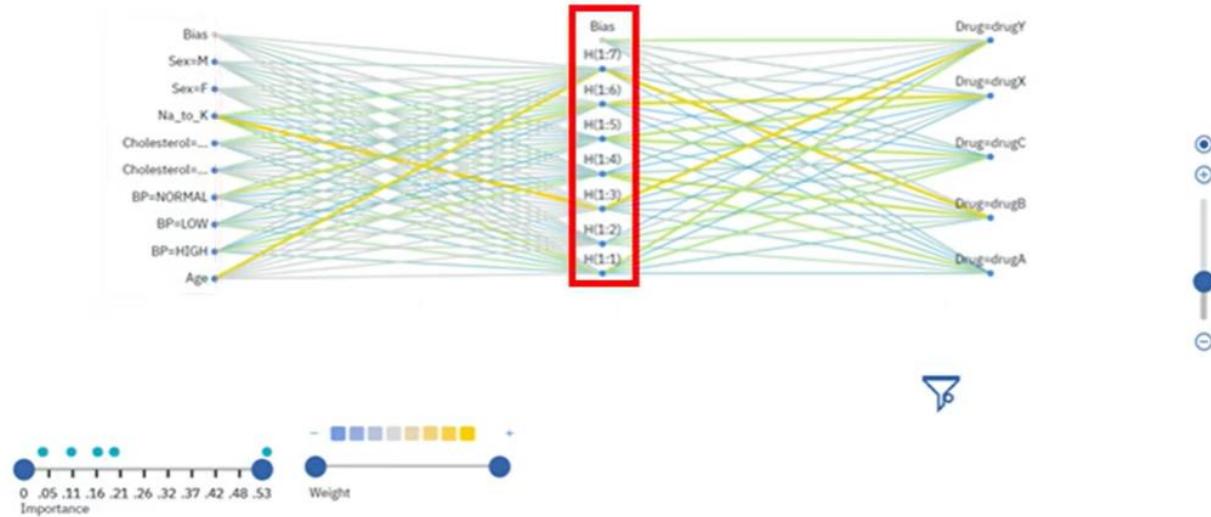


Network diagram



Network Diagram ⓘ

TARGET : DRUG

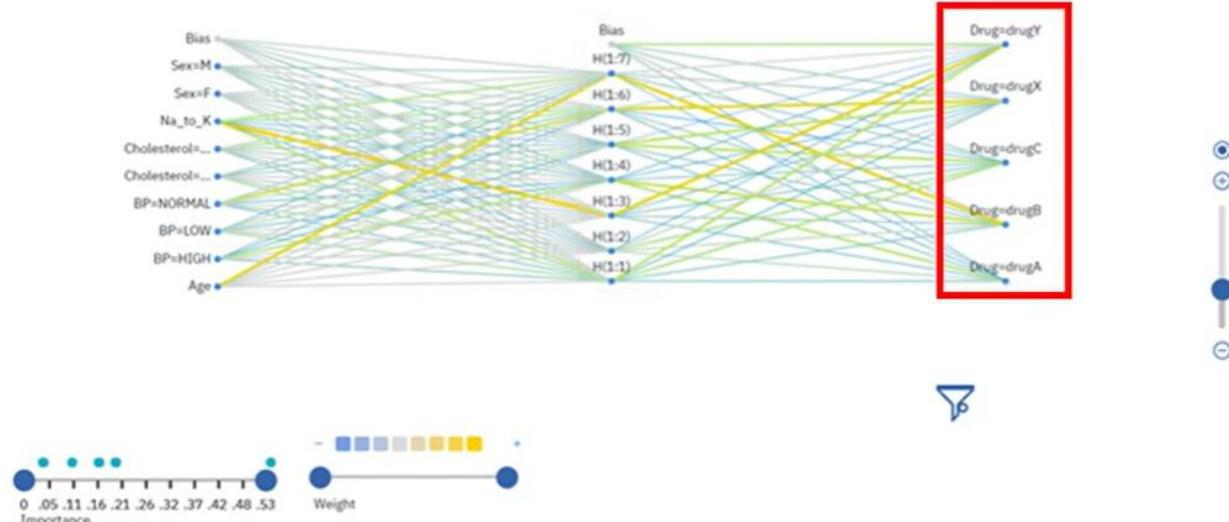


Network diagram

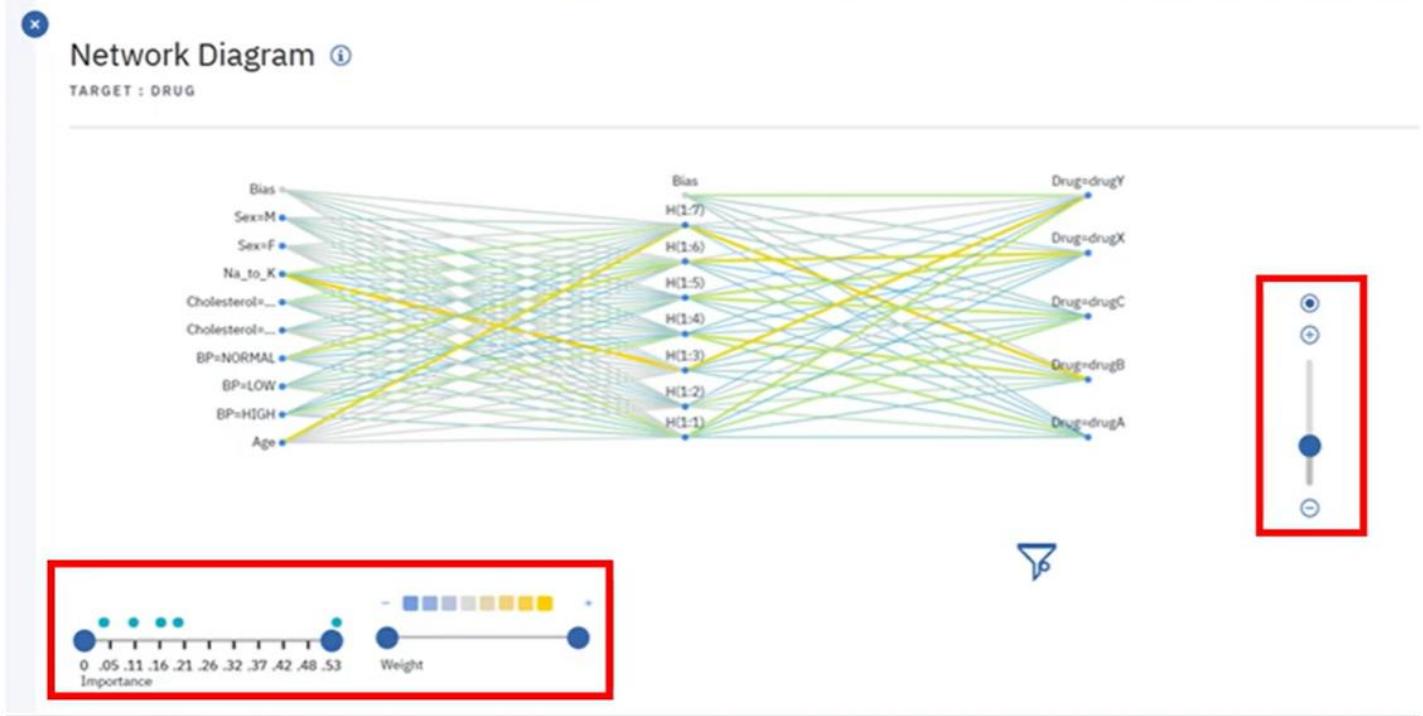


Network Diagram ⓘ

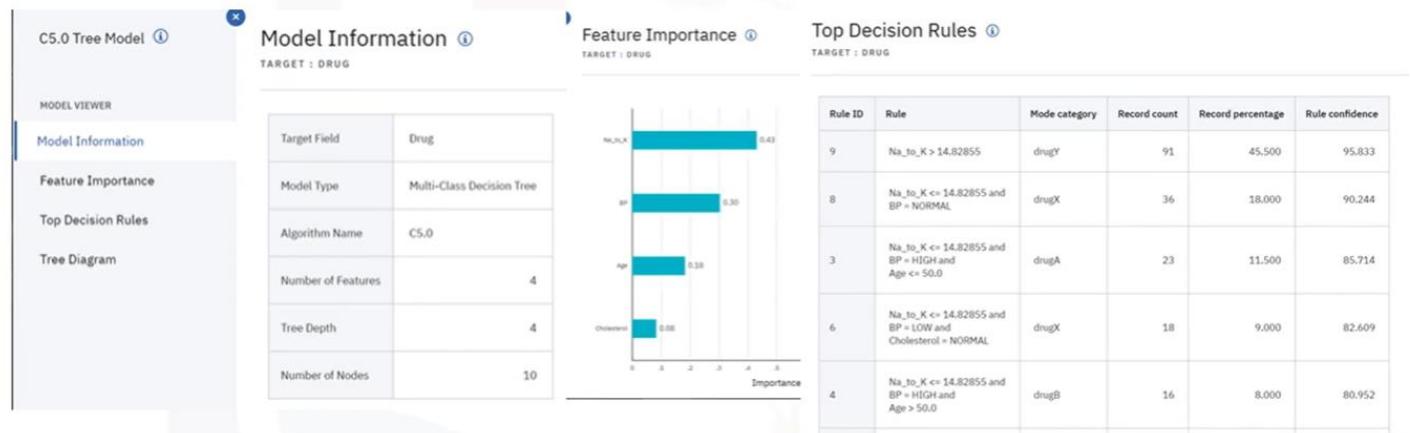
TARGET : DRUG



Network diagram

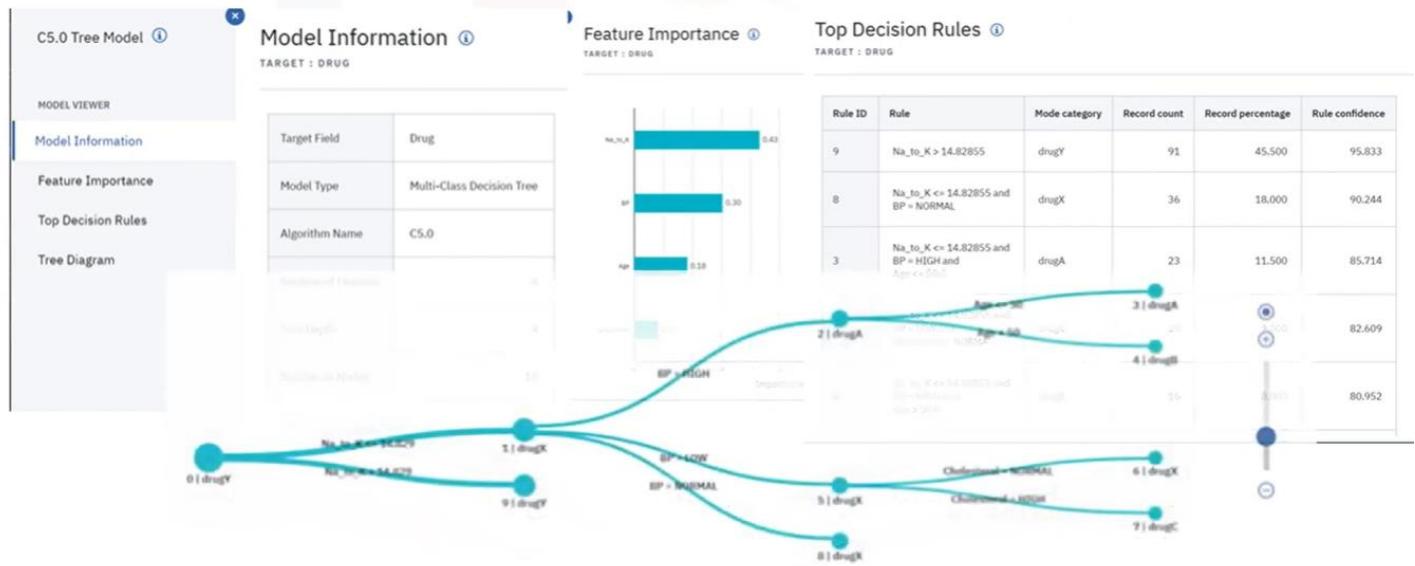


Examining decision tree model



Decision tree models are popular because they have a special structure that makes it easy to explain predictions or extract decision rules

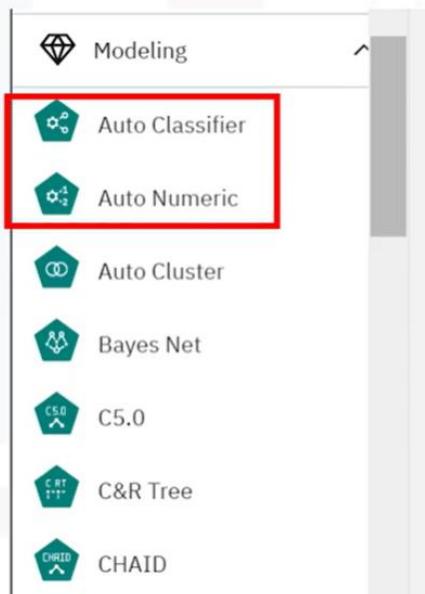
Examining decision tree model



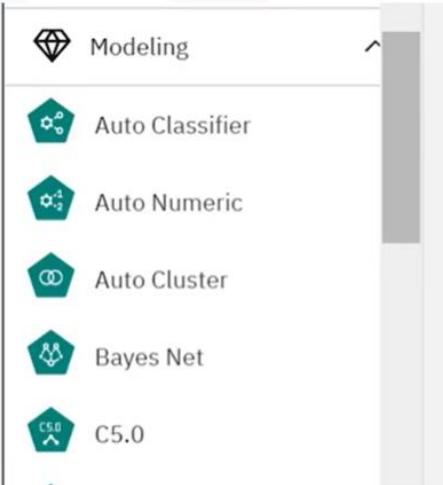
SPSS Modeler in Watson Studio

Watch later

Auto-Classifier and Auto-Numeric



Auto-Classifier and Auto-Numeric



Those nodes will build several kinds of models and pick the best one based on a certain criterion

hh

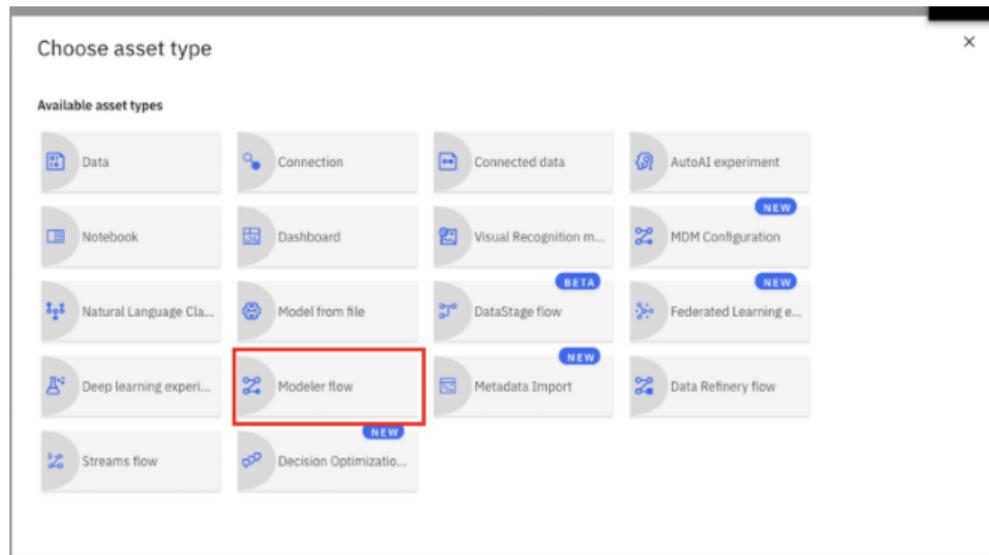
LAB: Modeler Flows in IBM Watson Studio (Updated)

For this lab, you will use the IBM Cloud Account and the Watson Studio project you created when you worked on the Lab in the previous lesson titled *Creating a Watson Studio Project with Jupyter Notebooks*. To create a free Watson Studio account, use the following link <https://cloud.ibm.com/catalog/services/watson-studio>

Objective for Exercise:

- Load an example flow, run it, and examine results.
- Add an Auto Numeric model, run it, and examine results.
- Get predictions for new cases using a model we built.

Step 1 - Open your project in Watson Studio, then click "Add Assets" by clicking the blue button on the top of the screen. In the panel that appears, select the "Modeler flow" option.



Step 2 - Next, select the tab "From example":

New modeler flow

New

From File

From Example

Select an example to build a model from sample data that you can download and add to your project.

SPSS

Use these samples to build predictive models without programming.

Drug Study Example

Use neural network and C5.0 algorithms to build classification models that allow you to predict the correct type of drug for a patient based on various health metrics.

Sales Promotion Study

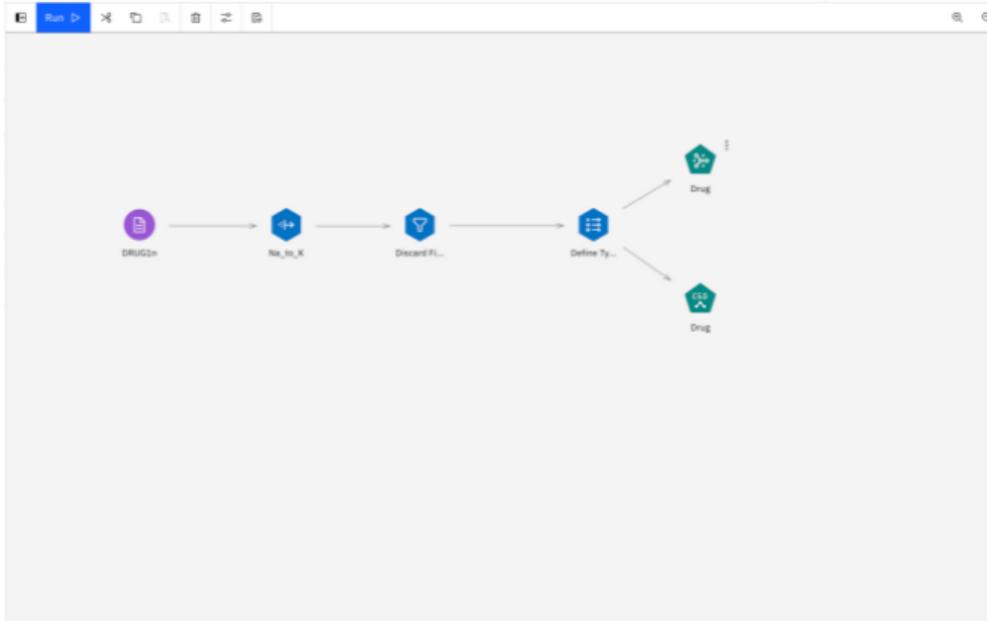
Use neural network and C&RT algorithms to predict the effect of advertising promotions on the sale of various items. Input data of sales before and after a past promotion are used to train the model to predict the effectiveness of advertising.

Hotel Satisfaction

A hotel manager is very interested in learning what customers think about the hotel. This example uses Text Analytics nodes to analyze fictional text data about hotel personnel, comfort, cleanliness, price, etc.

Step 3 - Selecting the Drug study example on the left would give us the flow we have already seen in the previous section, then click "Create" button in the lower right corner.

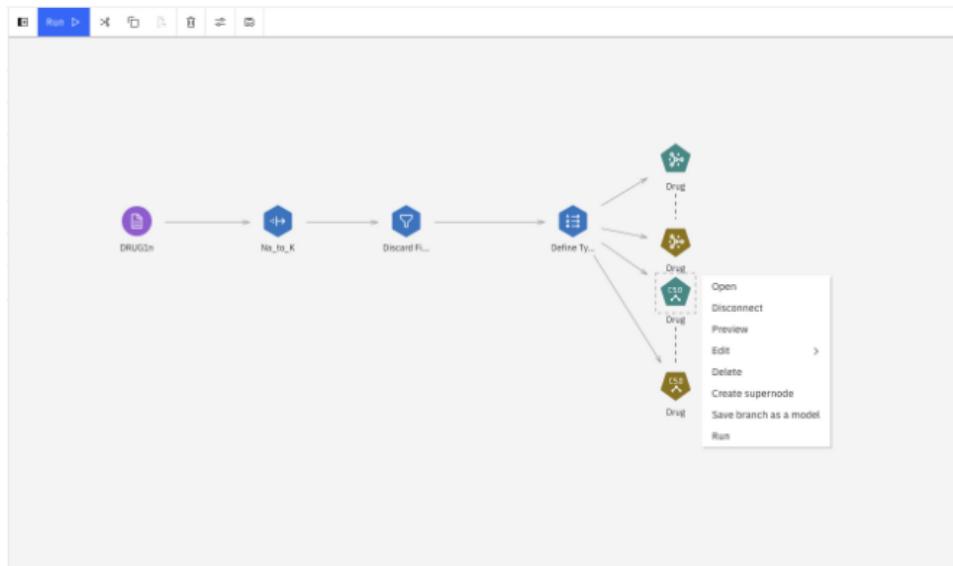
Once the flow loads, you will see an image similar to this:



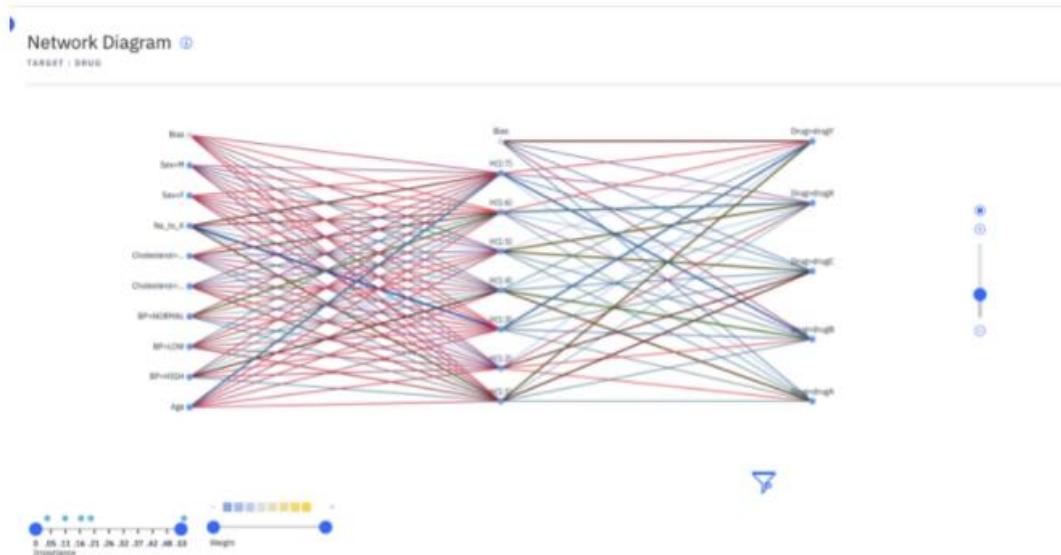
Step 4 - Reading from left to right, we can examine the flow. The purple circle corresponds to the data set of some fictional sales, the next node is deriving a field Increase that will be our target variable. It is based on the increase of sales. Then we see a type node which specifies our target variable and predictors. In this case our target variable is numeric and continuous. Finally, there are two models - a neural network and a C&RT decision tree. We can run the flow as is, or we can modify it to see other possibilities.

Step 5 - First, let's see what we can get with the existing models. Click **Run** button on the top of the canvas:

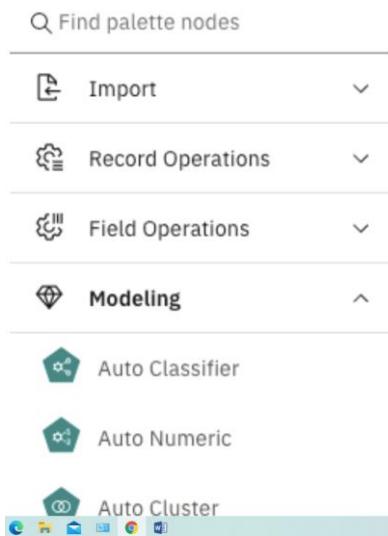
Step 6 - After some execution time, you'll get dark gold model nuggets for the two models. Clicking on three dots in the right side of one of them you can select "View model" and examine the information.



For example, the neural network diagram looks like this:



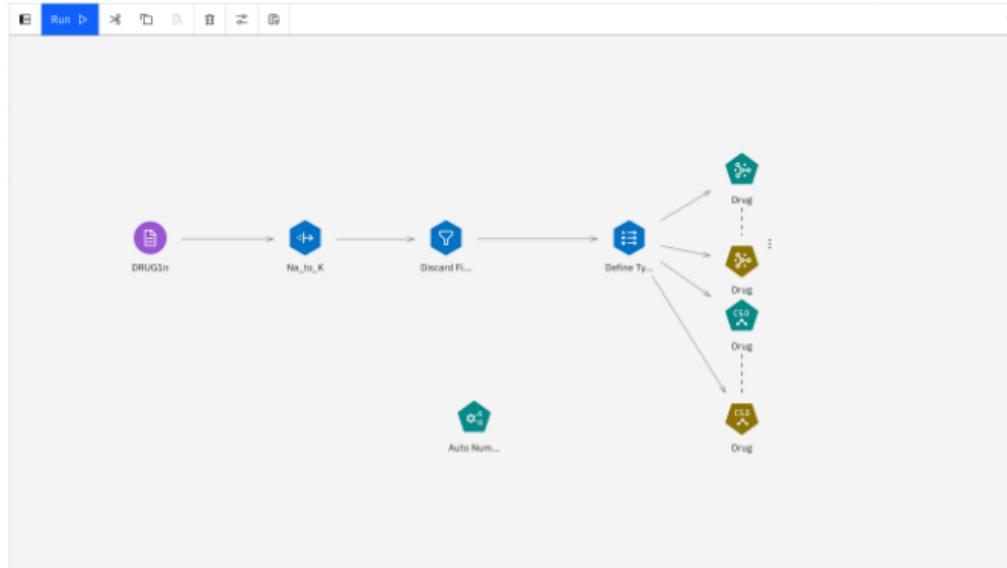
Step 7 - Now, let's add the "Auto-numeric" model to the flow. Open the modeling palette on the left side by clicking on Modeling:



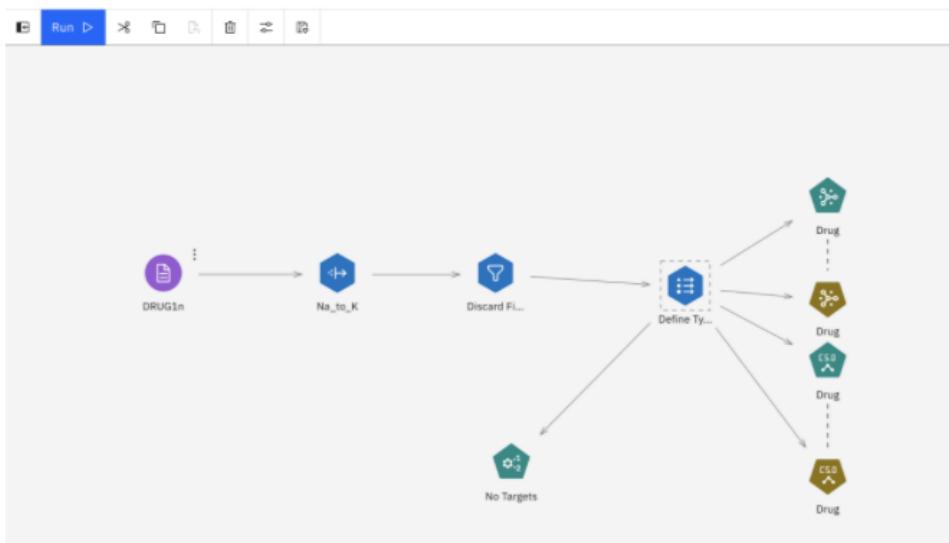
Step 8 - In the resulting modeling palette, pick the Auto-Numeric model:



Step 9 - Click on the Auto Numeric node and while keeping the left mouse button pressed, drag it onto the canvas.



Step 10 - Now, hover over the type node with your mouse and click on the circle with an arrow that appears on the right side. Make a connection to the dot on the left side of the Auto Numeric node.



Step 11 - Double-click the "Auto-numeric" and click "Use custom field roles", use the drop-down to chose a target, we will use "Na_to_K". Click "Add Columns" next to Inputs to add inputs.

The screenshot shows a software interface with a toolbar at the top featuring icons for sorting, search, refresh, and help. Below the toolbar, the text "No Targets" is displayed. Underneath, there's a section titled "Fields" with a collapse/expand arrow. A checkbox labeled "Use custom field roles" is checked. The "Target" section contains a dropdown menu with three options. The "Inputs" section has a header row with a "Remove" button and an "Add Columns" button followed by a plus sign. A single row is listed under "Inputs" with a checked checkbox and the label "Field Name".

Partition

...

▼

Splits

Remove Add Columns



Field Name

Cancel

Save

Step 12 - Check all. the fields we want to use for the input and click "OK" then "Save"

Select Fields for Na_to_K

Find in column Field name Filter: # abc Reset

<input checked="" type="checkbox"/>	Field Name	Data Type
<input checked="" type="checkbox"/>	Age	# integer
<input checked="" type="checkbox"/>	Sex	abc string
<input checked="" type="checkbox"/>	BP	abc string
<input checked="" type="checkbox"/>	Cholesterol	abc string
<input checked="" type="checkbox"/>	Drug	abc string

Cancel

OK

Step 13 - Finally, run the new branch of the flow by clicking on the three dots on the upper right part of Auto Numeric node and selecting the last option in the menu, "Run".



Step 14 - After the execution ends, we get a model nugget:



and viewing the model, we see the following table:

Auto Numeric - Models (i)

TARGET : NA_TO_K

USE	ESTIMATOR	ACCURACY ▾	RELATIVE ERROR	BUILD TIME (MINS)	NO. FIELDS USED	ACTIONS
<input checked="" type="checkbox"/>	XGBoost Tree 1	0.941	0.134	< 1	5	
<input checked="" type="checkbox"/>	Random Trees	0.859	0.290	< 1	5	
<input checked="" type="checkbox"/>	C&RT	0.838	0.300	< 1	5	
<input checked="" type="checkbox"/>	Generalized Linear 1	0.809	0.346	< 1	5	
<input checked="" type="checkbox"/>	LE	0.808	0.347	< 1	5	

This tells us that five different models have been built, and also some properties of the models.

XGBoost is a very popular model, representing gradient-boosted ensemble of decision trees. The algorithm was discovered relatively recently and has been used in many solutions and winning data science competitions. In this case, it created a model with the highest accuracy, which "won" as well. "C&RT" stands for Classification and Regression Tree", a decision tree algorithm that is widely used. This is the same decision tree we saw earlier when we built it separately. "LE" is "linear engine", an IBM implementation of linear regression model that includes automatic interaction detection. The model coefficients are shown in **Parameter Estimates** table. We can see that several coefficients correspond to a combination of one category of variable *class* with continuous variable *Promotion*. This is called "an interaction effect". You will not see such features in simple linear regression models.

← Auto Numeric -
Na_to_K

Linear Regression (i)

EVALUATION

Model Evaluation

Observed by Predicted

Residuals by Predicted

Normal P-P Plot of Residuals

Residual Histogram

MODEL VIEWER

Model Information

Records Summary

Feature Importance

Tests of Model Effects

Parameter Estimates

Model Evaluation (i)

TARGET : NA_TO_K

Model Accuracy 0.661

Model Evaluation Measures

R ²	0.661
Adjusted R ²	0.646

Step 15 - Now you can experiment with Modeler flows on your own. You can use various nodes, just don't forget to put a Type node before any modeling node and to specify your inputs and targets.

There are several other models to predict a continuous target, for categorical targets, you can try Logistic Regression, Support Vector Machines etc.

You will learn more about various models when you take a course on Machine Learning.

IBM SPSS Modeler

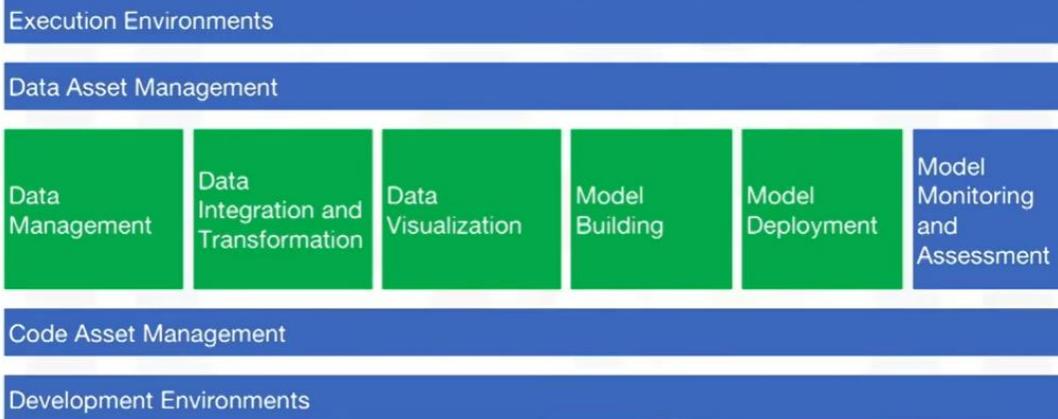
In this lesson we will discuss two products that are very helpful for data scientists. Both came to IBM with the SPSS acquisition in 2009. First is **IBM SPSS Modeler**. Let's review the different tool categories we discussed previously.

- IBM SPSS Modeler includes data management capabilities and tools for data preparation, visualization, model building and model deployment. The product was created by Integral Solutions Limited in the United Kingdom in 1994 and was originally called Clementine. It was acquired by a company called SPSS in 1998 and SPSS was in turn acquired by IBM in 2009.
- SPSS Modeler is a data mining and text analytics software application. It's used to build predictive models and conduct other analytics tasks.
- It has a visual interface that enables users to leverage statistical and data mining algorithms without programming. One of its main goals from the beginning was to create complex predictive modeling pipelines that are easily accessible. A sample modeler stream shown here includes one round data source node, three triangular graph nodes, one hexagonal node for computing, a new variable, and a square node for an output table.

Below the canvas, we can see the rich node palette with separate tabs for data sources, record in field operations, graphs, models, output and so on. Nodes and different tabs have different shapes with Pentagon's used for modeling nodes. Let's examine the sample stream that comes as an example with the product. It starts with a data set of telecommunications records and the goal is to build a model to predict which customers are about to leave the service otherwise known as churn. The data source is shown by the round node on the left side, a hexagon type node typically follows a data source node and it enables us to specify roles, target predictor or none. And measurement levels such as continuous nominal or flag for all variables. The term flag is used to denote a variable with two categories one of which can be considered positive and the other negative. In this example the measurement level for the churn

field is set to flag and the role is set to target. All others are set as predictors and inputs. The original data set has many fields and some of them are not relevant to the target variable, so we first need to decide which fields are more useful as predictors. There is a feature selection modeling node that helps to do this. After the stream with the feature selection node is executed a yellow model nugget gets created below it in the flow diagram. Using that nugget we can generate a filter node that filters out the variables that are not good predictors for the target. The data audit node located below the filtering node shows various properties of the data such as numbers of outliers in each variable and the percentage of valid values. It can also help to create a special node for missing value imputation that is replacing missing values of a variable with some valid values that can be selected based on domain knowledge. Here variable log toll has greater than 50% missing values and we will specify a value the mean to replace them. A super node in modeler is a special node that is not found in the palette but is created by the user with special functions included in it. The data audit node enables us to create a super node for imputing missing values. It is shaped as a star and shown on the right of the screen. Finally we attach the logistic regression model node to the stream and click run. Another model nugget appears and by clicking it we can see various model information and other output. In the output window that opens when we click on the model nugget the summary tab shows the target inputs and some model building settings. Based on certain advanced output settings that were specified before the model was built we can also see a classification table, accuracy, and some other generated outputs for the model. Note that these results are based on training data only. To assess how well the model generates two other real-world data you should always use a partition node to hold out a subset of records for the purposes of testing and validation. Then, in the model setup screen select the use partitioned

data check box. This will help detect and avoid model overfitting. Overfitting is defined as having significantly higher accuracy on the training data. Data used for training the model then on tests or unseen data. The yellow model nugget added earlier can also be used to compute predictions, also called scores on the original data or on a new data source. All we need to do is to connect the data source in question to the nugget, make sure it has the predictor variables used in the model, and create an output to a table or other structure for storing the scores. We can also specify settings for scoring inside the model nugget. Note that if the model was built on transformed predictor data, the same data transformation steps would be applied to the new data before it can be scored by the model. The analysis node is the final node in the stream. It attaches to a model nugget and when executed it will compute some model evaluation metrics, such as a confusion matrix and accuracy. In this example we've only looked at a logistic regression model. IBM SPSS Modeler offers a rich modeling palette that includes many classification, regression clustering, Association rules and other models. It also contains large selections of data source types, data transformations, graphs, and output notes. And we haven't even talked about text analytics, entity resolution and many other features of the product that can be extremely helpful to data scientists. We could create an entire course on IBM SPSS Modeler alone. You've learned how IBM SPSS Modeler helps analysts to create powerful machine learning pipelines using graphical interface. Next, we will talk about the original SPSS product now called IBM SPSS Statistics.



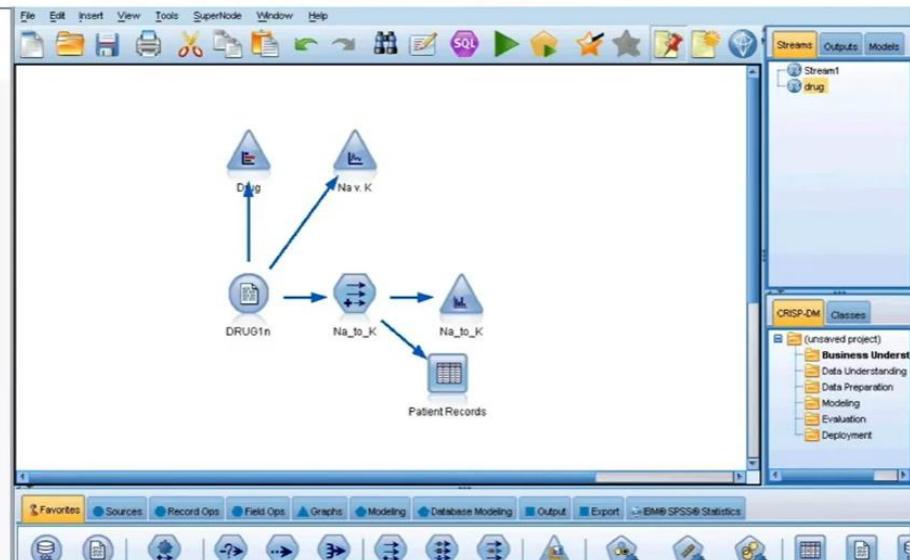
Includes data management capabilities and tools for data preparation, visualization, model building, and model deployment

IBM SPSS Modeler overview



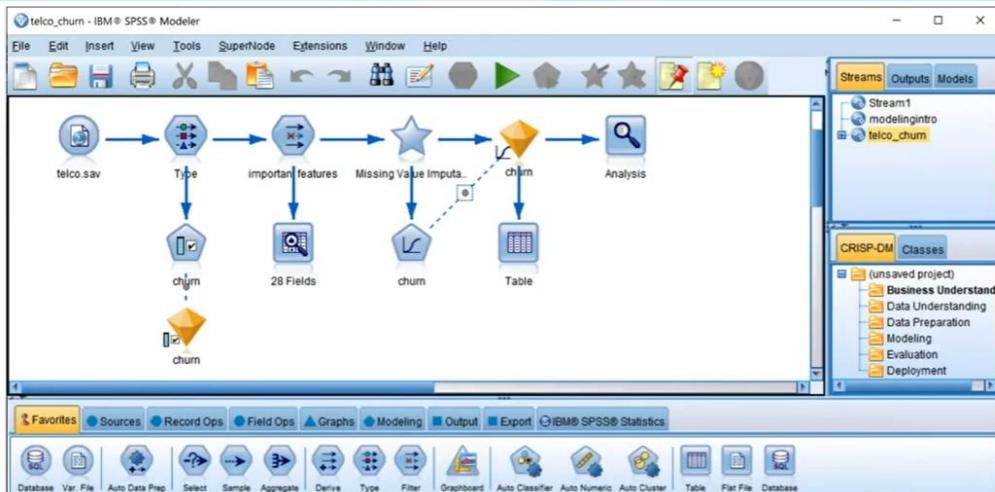
SPSS Modeler is a data mining and text analytics software application

IBM SPSS Modeler workspace



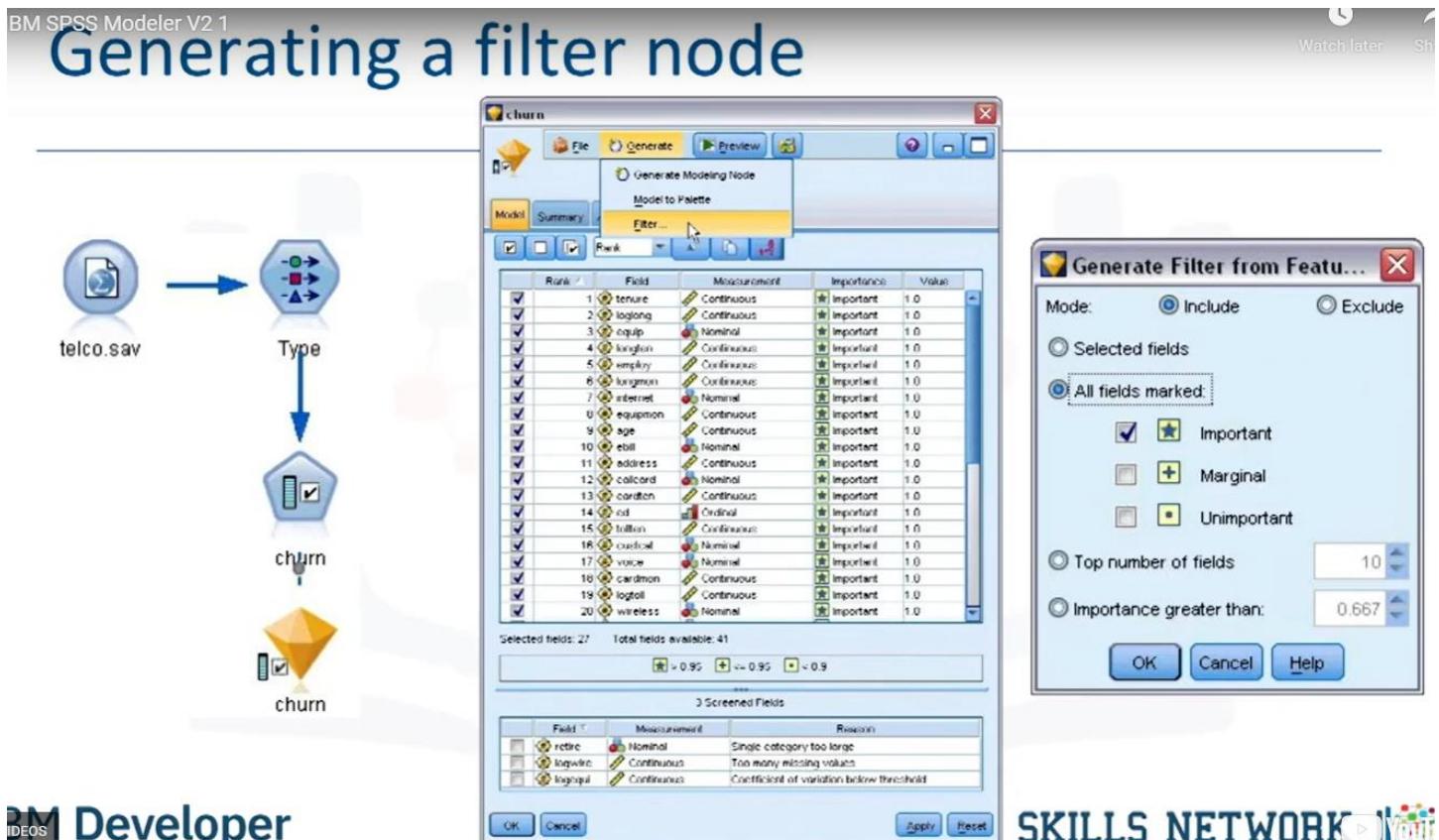
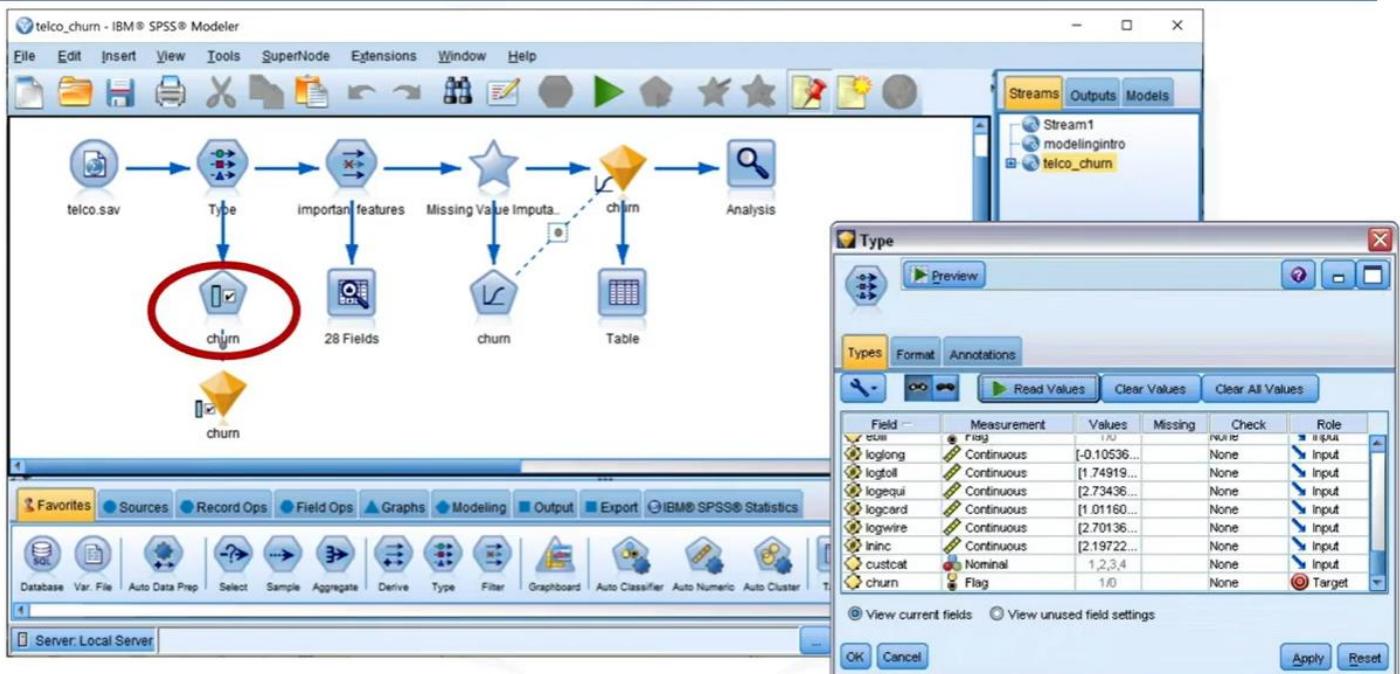
Nodes in different tabs have different shapes, with pentagons used for modeling nodes.

An example predicting churn

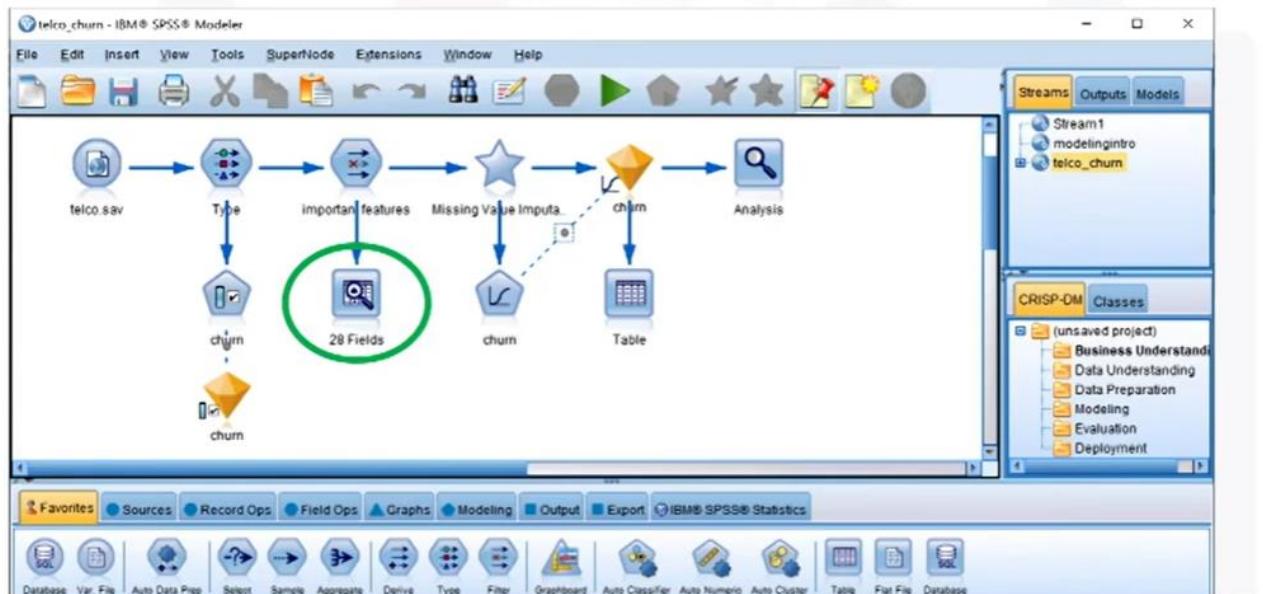


The goal is to build a model to predict which customer are about to leave the service, otherwise known as "churn."

An example predicting churn



Data audit node



Replacing missing values of a variable with some valid values that can be selected based on domain knowledge

Imputing missing values

The screenshot shows the "Data Audit of [28 fields] #2" window. It displays a table of fields with their measurements, outliers, extremes, and current actions. The "Audit" tab is selected. An "Annotations" tab is also visible. The "Impute Missing" column for the "logtoll" field is open, showing options like "Never", "Fixed", "Blank Values", "Null Values", "Blank & Null Values", "Condition...", and "Specify...". A cursor is hovering over the "Specify..." option.

On the right, an "Imputation Settings" dialog is open for the "logtoll" field. It shows the "Field" as "logtoll", "Storage" as "Real", and "Impute when:" set to "Blank & Null Values". The "Impute Method" is set to "Fixed". Under "Impute Fixed Values", the "Fixed as:" dropdown is set to "Mean", and the "Value:" dropdown has "Mean" selected. Buttons for "OK", "Cancel", and "Help" are at the bottom.

Generating a SuperNode

Field	Type	Action	Input Missing	Method	% Complete	Valid
tenure	Nominal	0 None	Blank If Null Val.	Fixed	100	100
age	Nominal	0 None	Never	Fixed	100	100
income	Nominal	0 None	Never	Fixed	100	100
ed	Nominal	0 None	Never	Fixed	100	100
employ	Nominal	0 None	Never	Fixed	100	100
equip	Nominal	0 None	Never	Fixed	100	100
month	Nominal	0 None	Never	Fixed	100	100
wireless	Nominal	0 None	Never	Fixed	100	100
longmon	Continuous	18	4 None	Never	Fixed	100
fullmon	Continuous	9	1 None	Never	Fixed	100
agephon	Continuous	2	0 None	Never	Fixed	100
cUSTOMER	Continuous	11	3 None	Never	Fixed	100
wiremon	Continuous	8	1 None	Never	Fixed	100
longline	Continuous	20	4 None	Never	Fixed	100
tollfree	Continuous	18	2 None	Never	Fixed	100
carter	Continuous	11	6 None	Never	Fixed	100
voice	Nominal	0 None	Never	Fixed	100	100
pauper	Nominal	0 None	Never	Fixed	100	100
internet	Nominal	0 None	Never	Fixed	100	100
cellwall	Nominal	0 None	Never	Fixed	100	100
confer	Nominal	0 None	Never	Fixed	100	100
still	Nominal	0 None	Never	Fixed	100	100



Missing Value Imputa...

The Data Audit node enables us to create a SuperNode for imputing missing values

Build and examine the model

telco_churn - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

telco_churn

CRISP-DM Classes

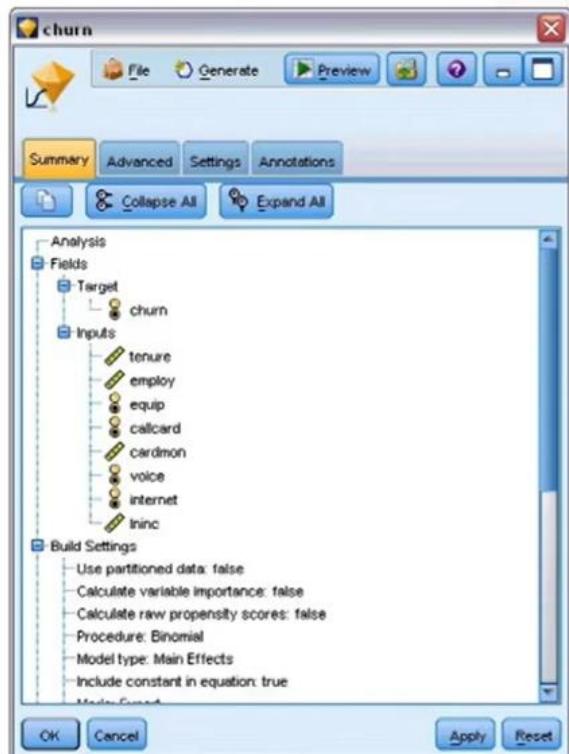
(unsaved project)

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics

Database Var File Auto Data Prep Select Sample Aggregate Drive Type Filter Grapboard Auto Classifier Auto Numeric Auto Cluster Table Flat File Database

Server: Local Server 129MB / 134MB

Examine output

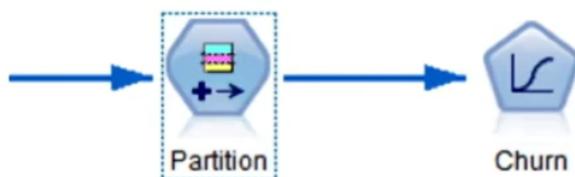


Observed	Predicted		Percentage Correct
	churn		
	No	Yes	
churn	726	0	100.0
	274	0	0
Overall Percentage			72.6

a. Constant is included in the model.
b. The cut value is .500
Variables in the Equation

OK Cancel Apply Reset

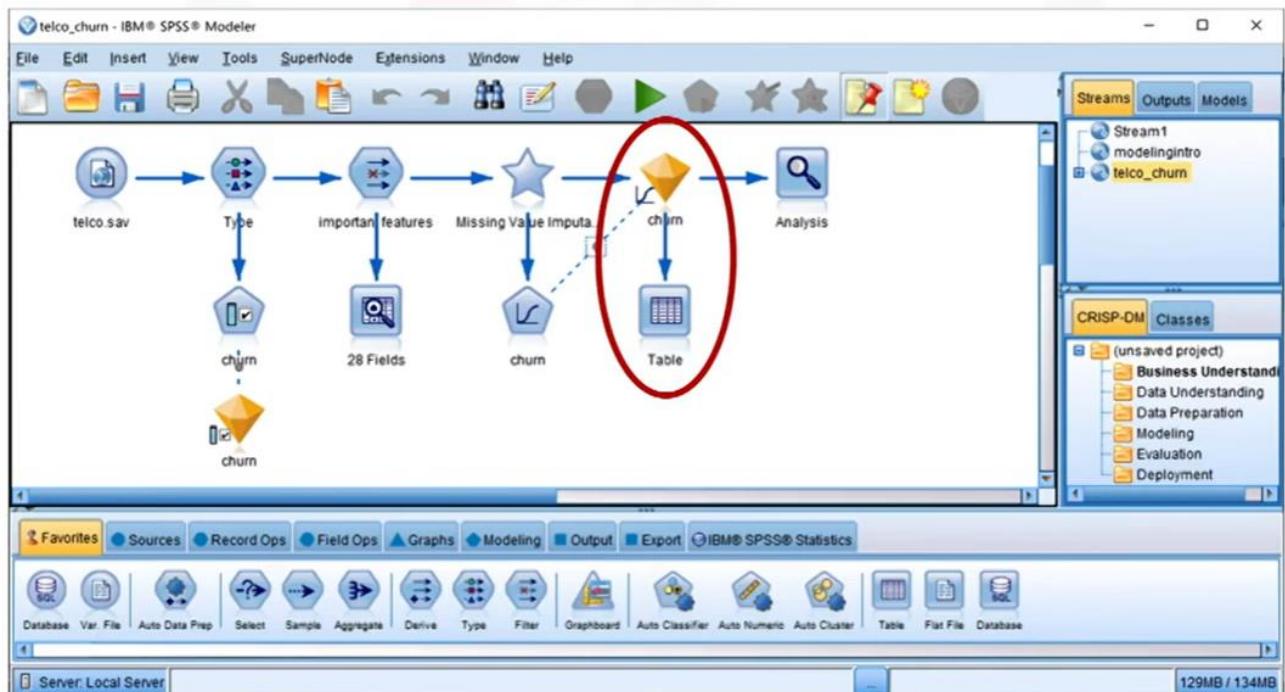
Always split data into training and test



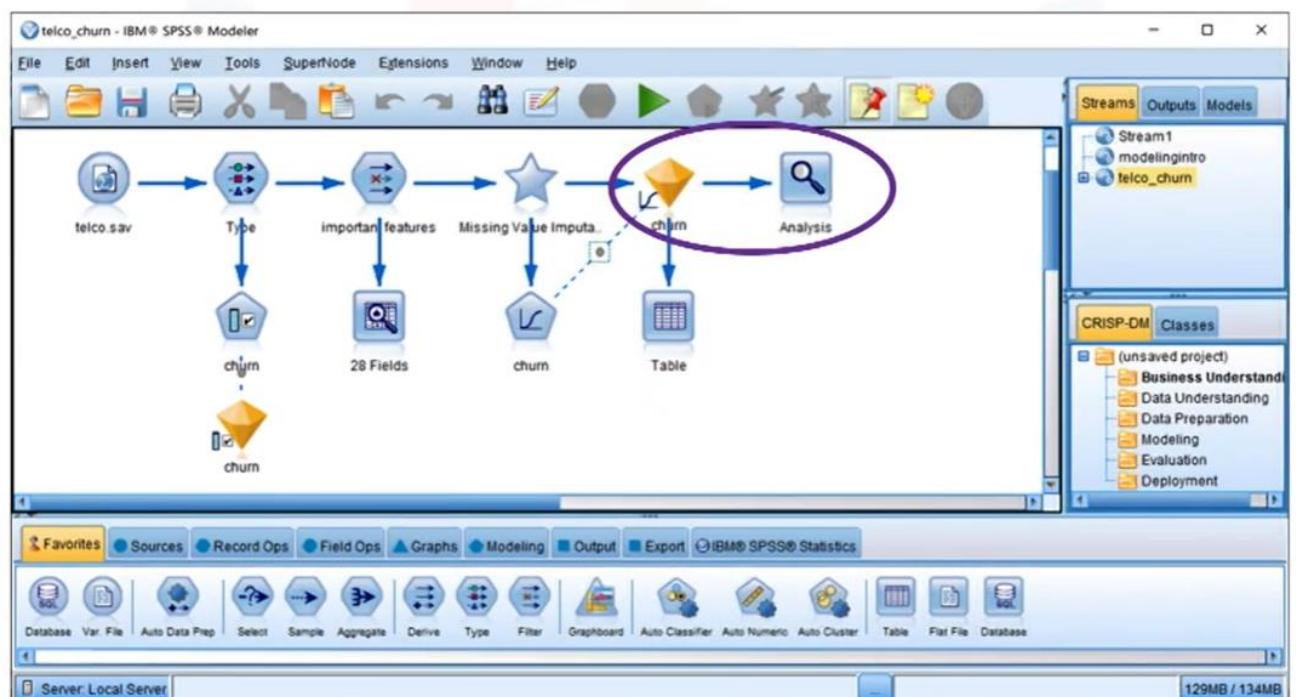
Use partitioned data

Always use a Partition node to hold out a subset of records for the purposes of testing and validation

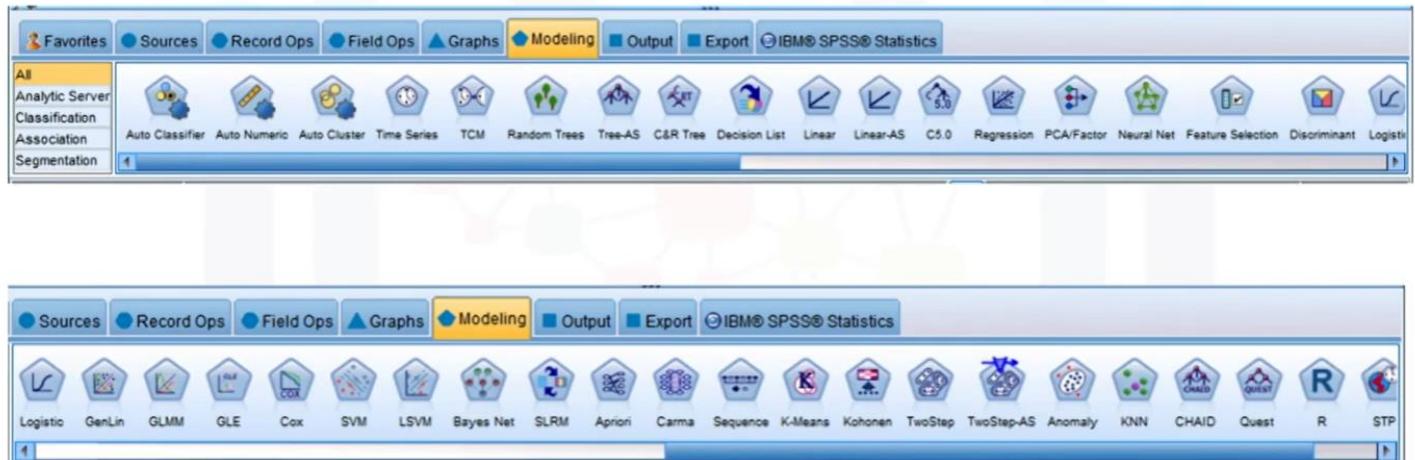
Compute model predictions (scores)



Model evaluation



Other models



IBM SPSS Modeler offers a rich modeling palette that includes many classification, regression, clustering, association rules, and other models.

SPSS Statistics

IBM SPSS Statistics evolved from an original product that was released in 1968. That product was called "Statistical Package for Social Sciences," or "SPSS."

- IBM SPSS Statistics is a statistical and machine learning software application and is widely used in academia, government agencies, and large enterprises.
- It's used to build predictive models, perform statistical analysis of data, and conduct other analytic tasks.
- It has a visual interface, which enables users to leverage statistical and data mining algorithms without programming, although the interface is very different from Modeler. As you can see, the main section of the screen looks very much like a spreadsheet; it displays data and allows manual editing.

This particular small data set, called "Employee Data", was created some time ago and does not represent real people. It is shipped with the product for use in demos and tutorials.

At the bottom of the screen, we can see two tabs: Data View and Variable View. In the Variable View, we can see and edit the information about all variables, including names, labels, data types, and measurement levels. We can also specify labels for values of categorical

variables, and missing values.

At the top of the data window is a menu. Under File, if you select "Import Data," you will see a list of a wide variety of data formats that you can import. The product uses its own data file format with the extension ".sav" that saves all the information about the variables we just saw in Variable view. The menu enables importing from and exporting to many other formats.

Under "Data," you'll find an extensive menu of possible data operations. Note that Data Validation can be performed using user-defined rules that specify the expected behavior of variable values. For example, if the date and month are kept in separate columns, the date cannot exceed "31," but for February, the date can't exceed "29." A special rule can therefore be created and applied during data validation. Additionally, you can enable some checks, such as percentage of missing values in a record or in the field. When you click the "Transform" menu item, you'll find a variety of available data transformations. Under "Compute Variable..." you can write a formula for a new variable based on existing variables. You can use any of the many mathematical and statistical functions available in the product.

You also have the option to use automatic data preparation, similar to Modeler. In the "Analyze" menu, you will see many types of statistical and machine learning analysis. Under "Regression," there are a variety of regression-related models. There are other kinds of regressions that appear separately on the Analyze menu, including General Linear Model, Generalized Linear Models, Mixed Models, and Loglinear. Now let's build a decision-tree model on the data. For this exercise we'll try to predict the "Employment category" field based on other fields. In the "Analyze" menu, select "Classify" and then "Tree". In the Decision Tree window, we can specify the dependent variable "Employment Category," and use most other fields -- except id and bdate -- as predictors, or independent variables. Usually the ID variable should not be used as a predictor, because it will not help with new cases, and the birthdate does not seem to be a useful predictor in this example either. We'll select "Exhaustive

CHAID" as our Growing Method, although there are also three other options available. Data scientists often try many different models to see which one works best for their data. Here we are just looking at one example model in order to illustrate how the product works. Click the "Validation" button to open the Decision Tree Validation window. Here, we select "Split-sample validation" to make sure we test the model on new data. Click "OK" in the Decision Tree window, to generate the output, including the tree diagram shown here. A Classification table is also displayed that shows how well the model works on training and test data. In this case, the accuracy is 91.2% on training data and only 85.6% on test data, which means the model does not generalize to new data very well. It's possible that by using different models, we can get better results.

Let's move to the next menu item. When you click "Graphs," you'll open a versatile Chart Builder, in addition to several other options.

The Chart Builder enables us to choose a style from the gallery and to drag required fields onto the canvas, select colors, and choose from other options.

Here's an example after we drag the "Previous Experience," "Current Salary," and Gender variables to the corresponding slots to define the axis and colors for the dots on the chart. The plot in the canvas is not based on real data, this example simply gives you an idea of what to expect.

Here is the real plot obtained from the data that we've been using. It shows different colored dots for gender, and regression lines that show the relationship of the current salary to previous experience for each gender.

Throughout IBM SPSS Statistics, you'll see a "Paste" button. When you click the "Paste" button, instead of executing the task right away the application will open another window, called the Syntax editor. Here, you can see the code called "syntax" pasted for you. SPSS syntax is a special programming language.

For example, here is the code for the decision tree we just built. Once we have the syntax, we can execute it, manually edit it, store it for later use, or send it to other users of IBM SPSS Statistics. Experienced SPSS users can write the code from scratch, while others

might prefer to have it generated by the graphical interface. Remember, the option to paste syntax

is available in throughout the program. If the syntax is generated by all the steps in a data analytics process -- opening the data set, applying any data transformations, building models -- and then saved as a syntax file with the extension ".sps", it's similar to saving a stream in IBM SPSS Modeler. However, one important difference is that it does not allow for an easy way of scoring new records with the model. We'll talk about different ways to deploy models in the next section.

You've learned how IBM SPSS Statistics helps data scientists to analyze their data using many statistical and machine learning techniques. Using a graphical user interface, we can create

complicated analysis that can be saved in the form of syntax and reused later.

Next, we will talk about predictive model deployment, an important part of the overall data science lifecycle.

IBM SPSS Statistics

Svetlana Levitan

© IBM Corporation. All rights reserved.

Evolved from an original product that was released in 1968

IBM SPSS Statistics overview

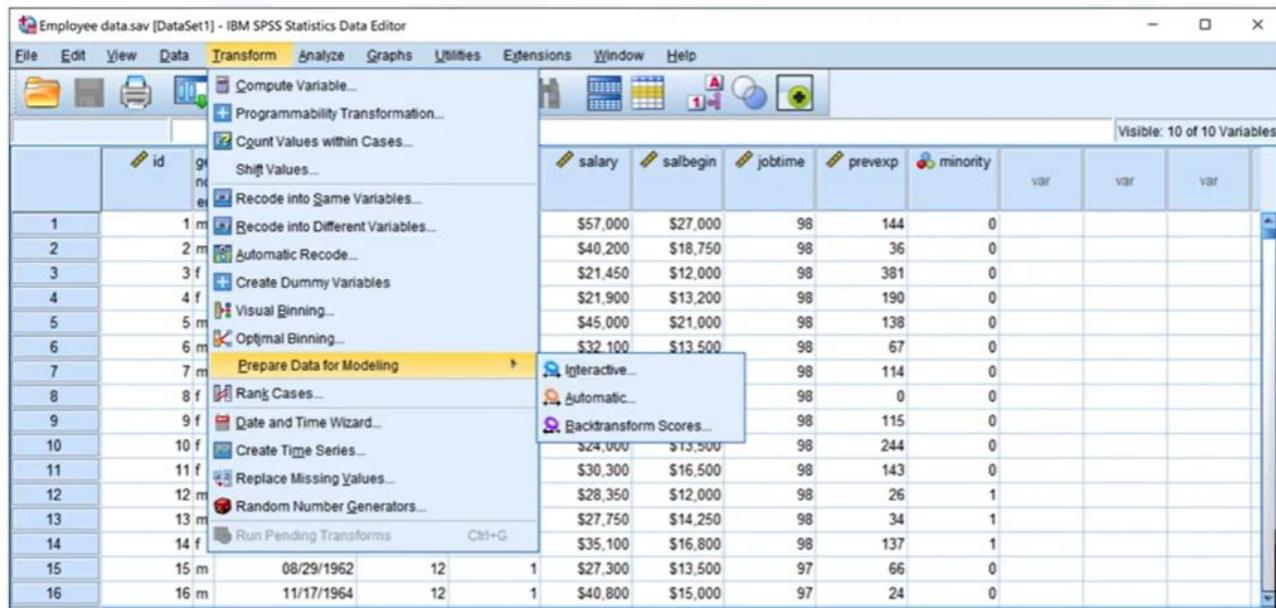
The screenshot shows the IBM SPSS Statistics Data Editor window. The title bar reads "Employee data.sav [DataSet1] - IBM SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Extensions, Window, and Help. Below the menu is a toolbar with various icons. The main area displays a table with 16 rows and 10 columns. The columns are labeled: id, gender, bdate, educ, jobcat, salary, salbegin, jobtime, prevexp, and minority. The table contains numerical and categorical data. A status bar at the bottom right indicates "Visible: 10 of 10 Variables".

Used to build predictive models, perform statistical analysis of data, and conduct other analytic tasks

The screenshot shows the IBM SPSS Statistics interface with the main menu open. The "File" menu is selected, showing options like New, Open, Import Data, Save As..., Save All Data, Export, and others. The "Import Data" option is highlighted. The "Validation" option under "File" is also highlighted. To the right, a detailed view of the "Validation" submenu is shown, listing various validation tools: Define Variable Properties..., Set Measurement Level for Unknown..., Copy Data Properties..., New Custom Attribute..., Define date and time..., Define Multiple Response Sets..., Validation, Identify Duplicate Cases..., Identify Unusual Cases..., Compare Datasets..., Sort Cases..., Sort Variables..., Transpose..., Adjust String Widths Across Files, Merge Files, Restructure..., Rake Weights..., Propensity Score Matching..., Case Control Matching..., Aggregate..., Orthogonal Design, and Split Into Files.

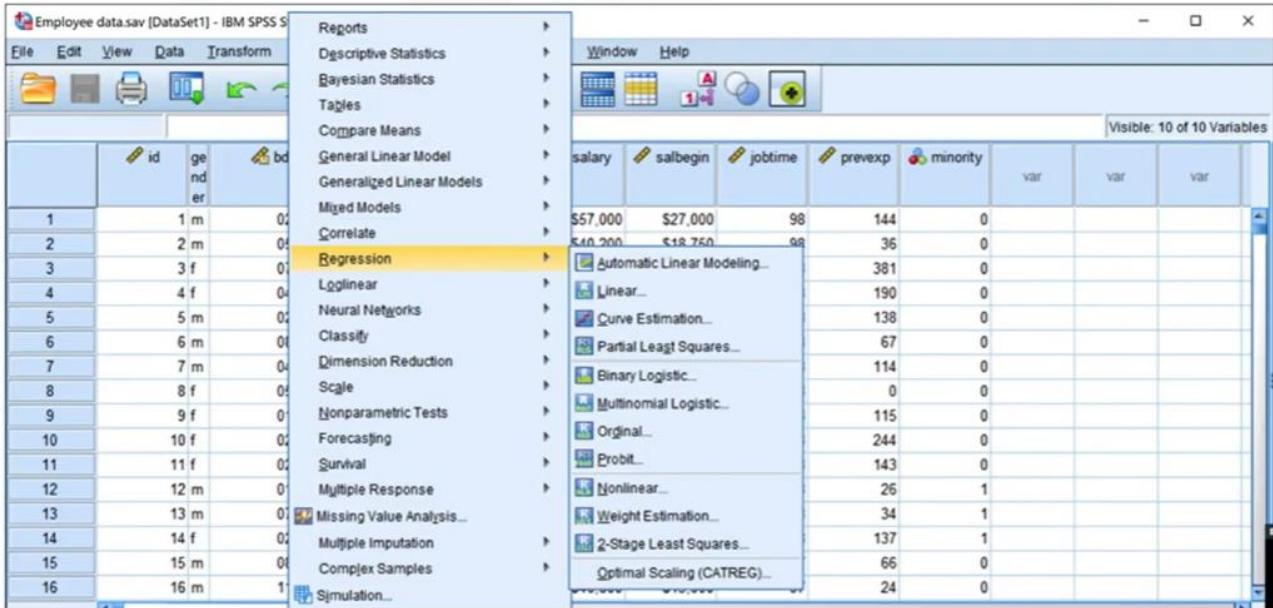
Data Validation can be performed using user-defined rules that specify the expected behavior of variable values

Data transformations



Write a formula for a new variable based on existing variables

Statistical analysis



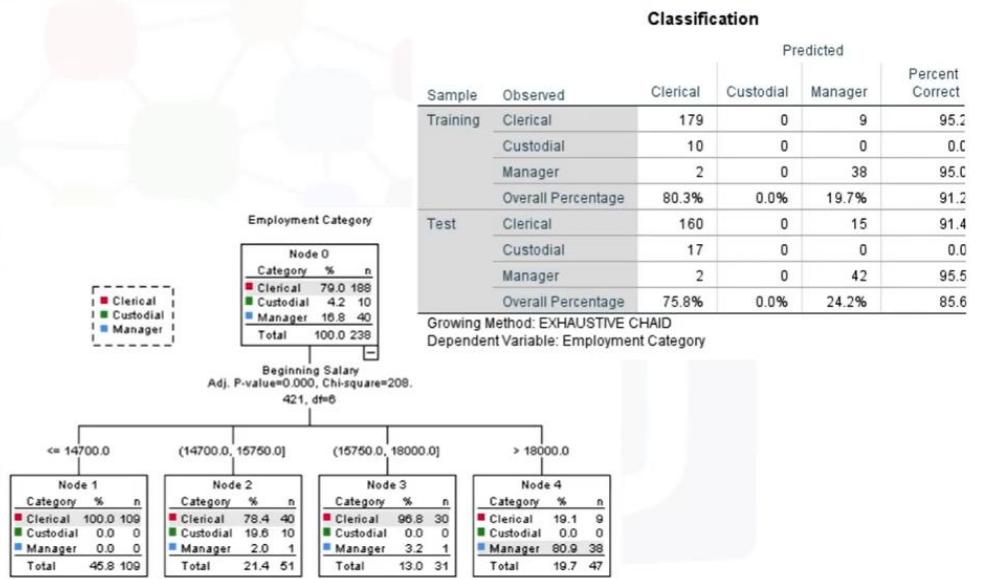
See many types of statistical and machine learning analysis

Building a tree model

The screenshot shows the SPSS Decision Tree dialog box. In the 'Variables' list, 'Employee Code [id]' and 'Date of Birth [bdate]' are selected. The 'Independent Variables' list includes 'Gender [gender]', 'Educational Level [level]', 'Current Salary [salary]', 'Beginning Salary [beginning]', 'Months since Hire [months]', and 'Previous Experience [experience]'. The 'Growing Method' is set to 'Exhaustive CHAID'. Other buttons include 'Output...', 'Validation...', 'Criteria...', 'Save...', and 'Options...'. A note says 'Right-click a variable to change its measurement level in the Variables list'.

Building a tree model

The 'Decision Tree: Validation' dialog box shows validation settings: 'Crossvalidation' (10 folds), 'Split-sample validation' (50% test sample), and 'Use variable' (Employee Code [id]). It also displays training and test sample statistics and a note about split-sample assignment.



Accuracy is 91.2% on training data and only 85.6% on test data

Graphs – chart builder

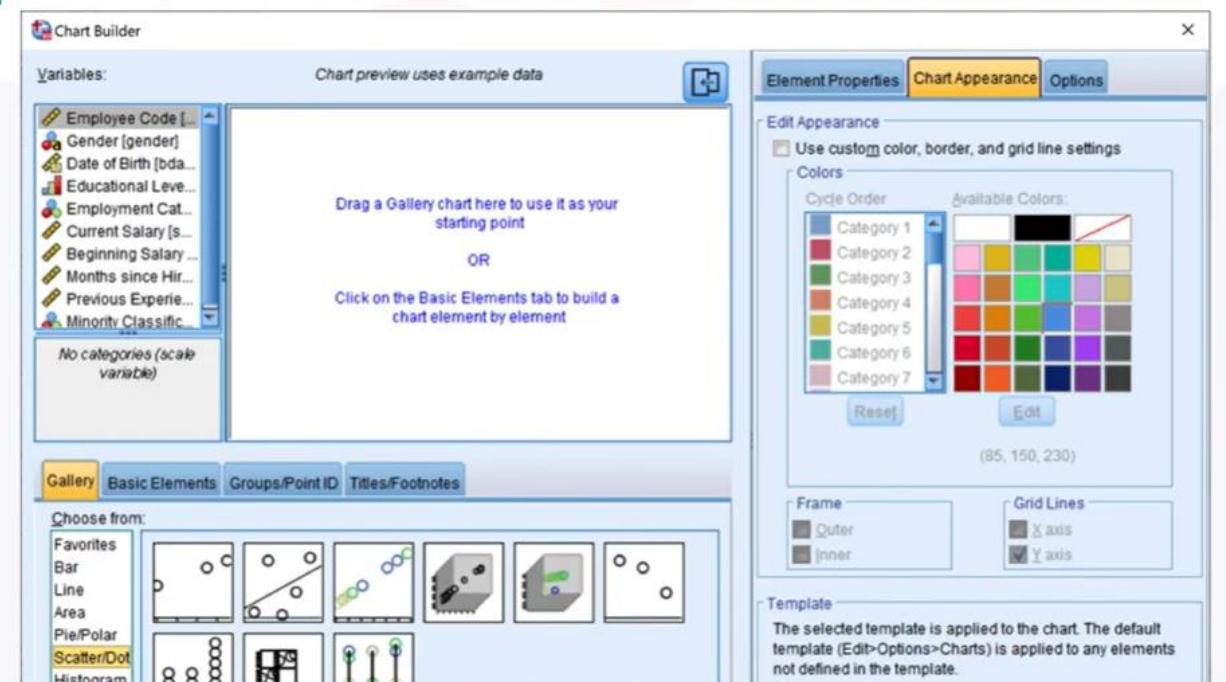


Chart Builder enables us to choose a style from the gallery and to drag required fields onto the canvas

Chart builder continued

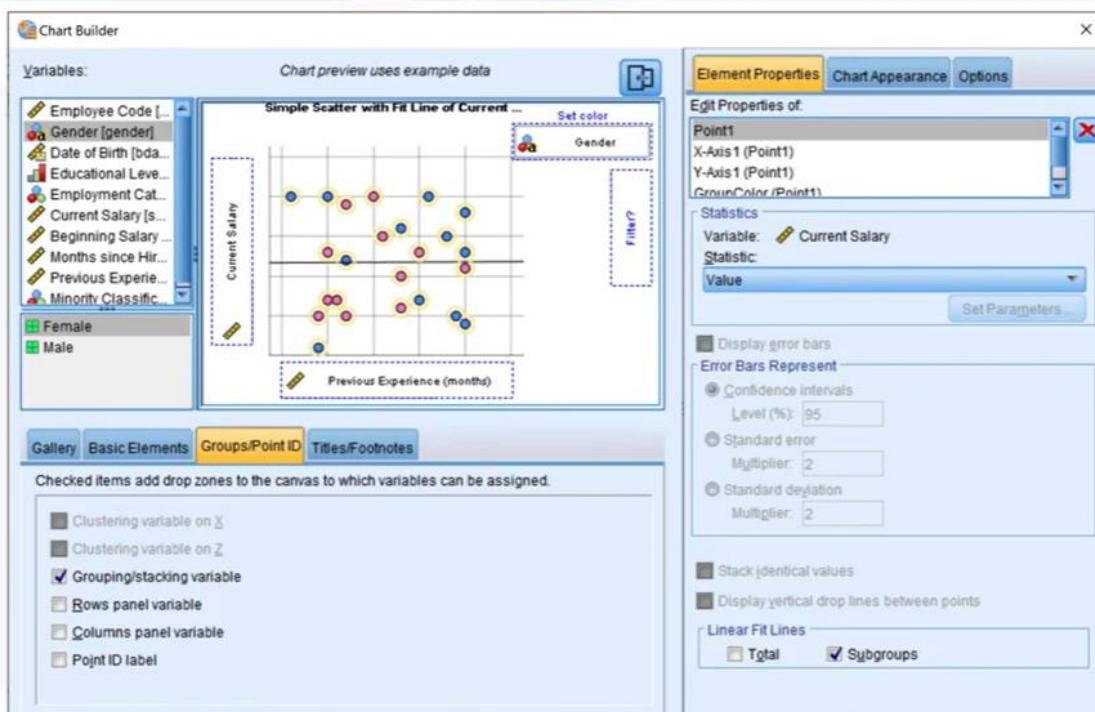
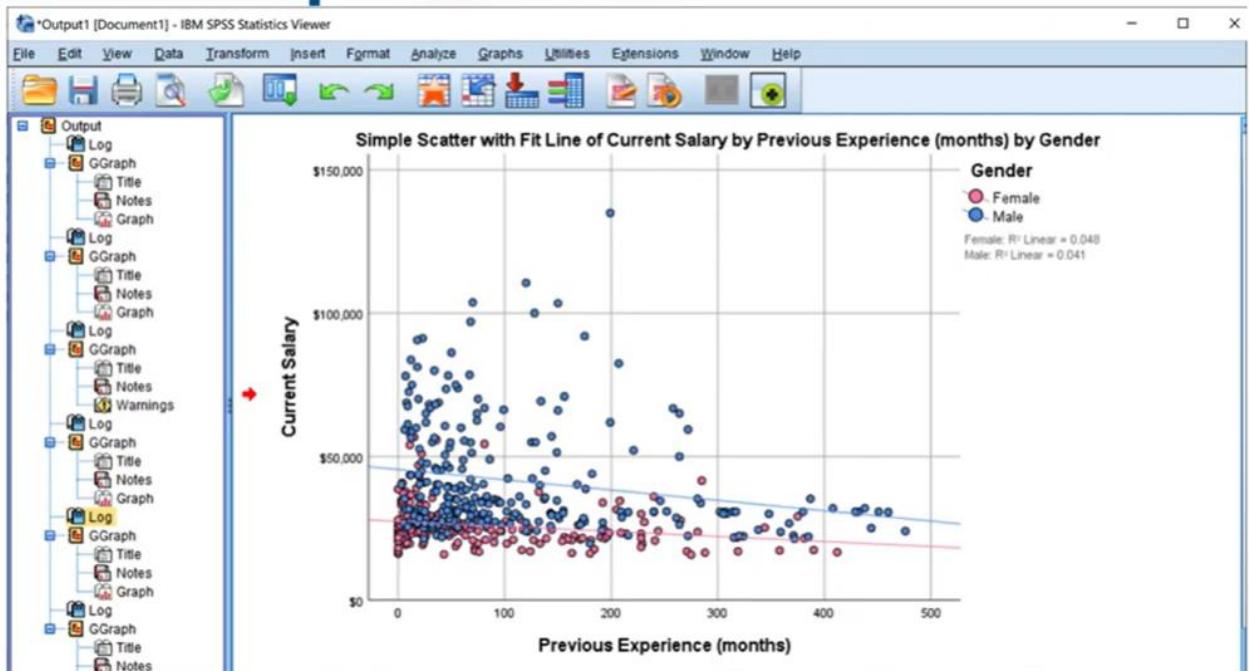
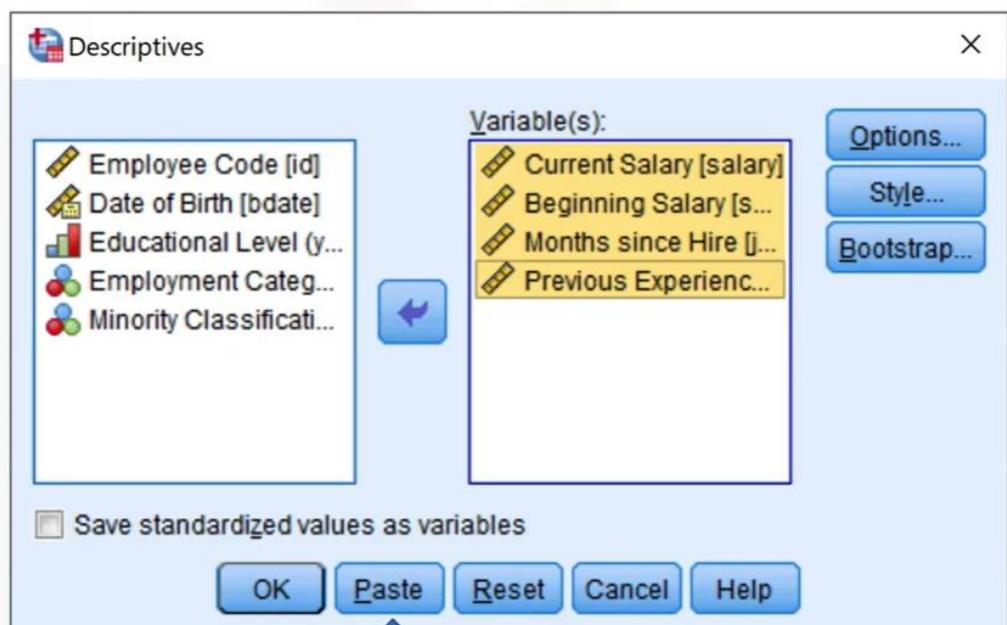


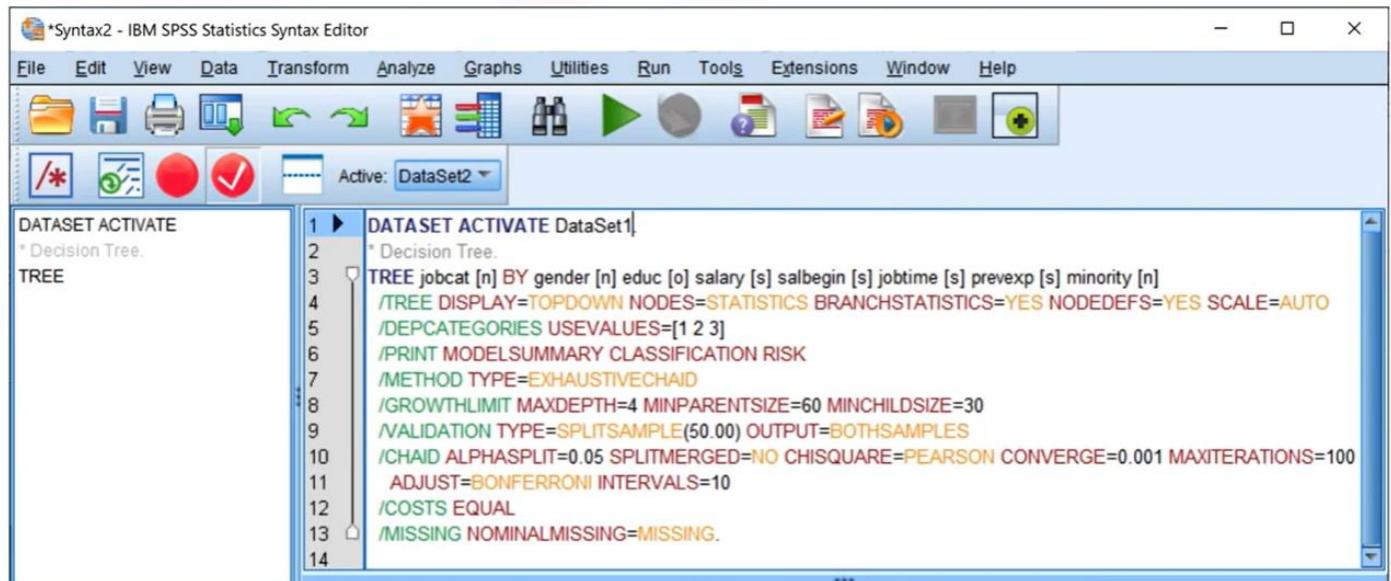
Chart output



Introducing SPSS Syntax



SPSS Syntax



The screenshot shows the SPSS Syntax Editor window. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Run, Tools, Extensions, Window, and Help. The toolbar contains various icons for file operations like Open, Save, Print, and Run. Below the toolbar, there are icons for Data Set, Output, and Syntax. The status bar indicates "Active: DataSet2". The main area displays SPSS syntax code for a decision tree:

```
1 DATASET ACTIVATE DataSet1.
2 * Decision Tree.
3 TREE jobcat [n] BY gender [n] educ [o] salary [s] salbegin [s] jobtime [s] prevexp [s] minority [n]
4 /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO
5 /DEPCATEGORIES USEVALUES=[1 2 3]
6 /PRINT MODELSUMMARY CLASSIFICATION RISK
7 /METHOD TYPE=EXHAUSTIVECHAID
8 /GROWTHLIMIT MAXDEPTH=4 MINPARENTSIZE=60 MINCHILDSIZE=30
9 /VALIDATION TYPE=SPLITSAMPLE(50.00) OUTPUT=BOTHSAMPLES
10 /CHAID ALPHASPLIT=0.05 SPLITMERGED=NO CHISQUARE=PEARSON CONVERGE=0.001 MAXITERATIONS=100
11 ADJUST=BONFERRONI INTERVALS=10
12 /COSTS EQUAL
13 /MISSING NOMINALMISSING=MISSING.
14
```

SPSS syntax is a special programming language

Model Deployment in Watson Machine Learning

So far, we've talked about building machine learning models and pipelines. In most practical applications, the return on investment is obtained when the model or pipeline is put into production, where it is used to get predictions, or scores, for the new cases.

Let's look back at our overview of different tool categories. In this unit, [Model Deployment](#) is our focus.

Suppose you worked hard to create the best possible machine learning model and the data preparation pipeline for it. How will you deploy your models?

In many practical scenarios, models are built and deployed by different teams, using different programming, and perhaps human languages. The teams will use different computing and data storage environments, and It might prove difficult to translate your program and the associated data preparation and post-processing steps from one environment to the other. Currently there are several approaches you can use to solve this problem, some commercial, some open source. Yet each one typically supports only a subset of all possible models, from building them to deploying, so a user gets locked into a specific framework.

Open standards for model deployment are designed to support model exchange between a wider variety of proprietary and open source models. Predictive Model Markup Language, or “PMML,” was the first such standard, based on XML.

It was created in the 1990s by the Data Mining Group, a group of companies working together on the open standards for predictive model deployment. IBM and SPSS were among the founding members of the Data Mining Group.

PMML 4.4 was recently released.

- It includes 17 statistical and machine learning models and many data transformations, built-in

functions, ways to combine multiple models together, and other features. This standard is widely known and used.

- The products we looked at earlier -- Watson Studio, IBM SPSS Statistics, IBM SPSS Modeler -- enable users to export most models in PMML.

In 2013, a demand for a new standard grew, one that did not describe models and their features, but rather the scoring procedure directly, and one that was based on JSON rather than XML. This led to the creation of Portable Format for Analytics, or PFA. PFA is now used by a number of companies and open source packages.

After 2012, deep learning models became widely popular. Yet PMML and PFA did not react quickly enough to their proliferation. The need for a standard intermediate representation was amplified by the wide variety of emerging deep learning frameworks and specialized hardware.

In 2017, Microsoft and Facebook created and open-sourced Open Neural Network Exchange, or “ONNX.” Originally created for neural networks, this format was later extended to support “traditional machine learning” as well.

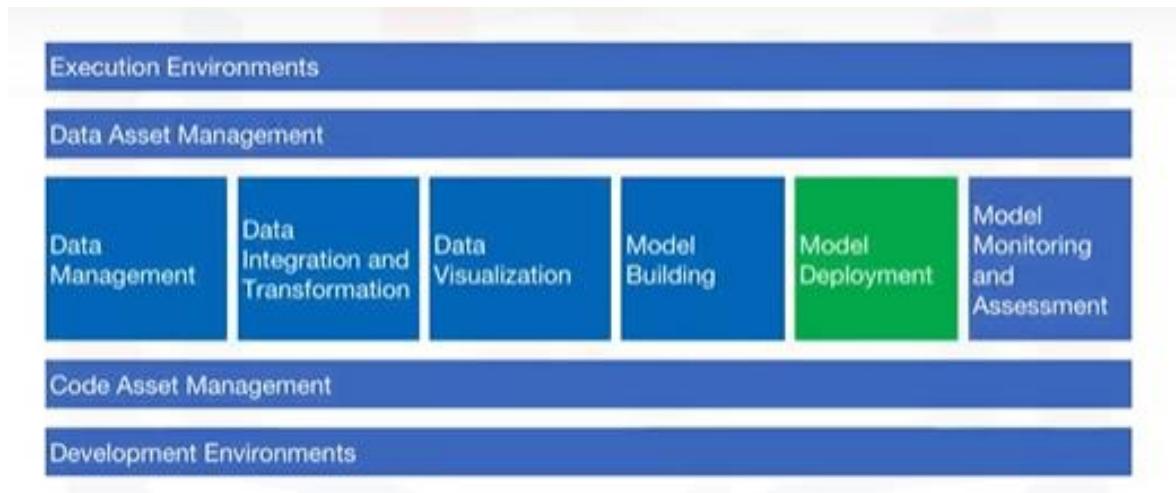
There are currently many companies working together to further develop and expand ONNX, and a wide range of products and open source packages are adding support for it.

Watson Machine Learning is IBM’s commercial offering designed for model deployment. It supports deployment of models built with most open source packages, as well as those expressed in PMML or ONNX. It also supports deployment of IBM SPSS Modeler streams and Modeler flows from Watson Studio. Deployment can be done using a graphical interface or Python code, and can be for online scoring through a REST API or batch scoring.

Watson Machine Learning helps integrate a deployed model into applications in the form of code snippets in several programming languages.

In this video, you've learned how open standards and Watson Machine Learning can help users to deploy their models into various application.

Next we'll talk about AutoAI and OpenScale, two advanced Watson Studio features that help to further simplify a data scientist's work.



How to deploy models?

- Workflow solutions (tend to support only models built with the same product)

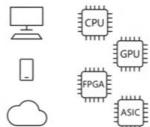
In many practical scenarios, models are built and deployed by different teams, using different programming, and perhaps human languages

How to deploy models?

- Workflow solutions (tend to support only models built with the same product)

Or

- Open standards:



Some workflow solutions

- Sage Maker by Amazon
- MLFlow by Databricks
- Airflow from Airbnb, now at Apache
- Kubeflow from Google

Open standards for model deployment



Open standards for model deployment are designed to support model exchange between a wider variety of proprietary and open source models

Open standards for model deployment

Predictive Model Markup Language



Created by Data Mining Group (DMG) in 1990s

XML based, version 4.4 has 17 models + many transformations and model combination methods

Supported by > 30 companies and Open source packages

Can be generated by Watson Studio and SPSS products

Open standards for model deployment

PFA: Portable Format for Analytics from DMG



JSON based small programming language

~5-6 years old

Supported by several companies and open source packages

ONNX: Open Neural Network eXchange



IBM Watson Machine Learning

- Deploy models built with Watson Studio, SPSS Modeler, many open source packages
- Support for PMML and ONNX
- Batch or online scoring

Deployment can be done using a graphical interface or Python code, and can be for online scoring through a REST API or batch scoring

Code snippets for integrating scoring

Code Snippets

CURL Java JavaScript **Python** Scala

```
import urllib3, requests, json

# retrieve your wml_service_credentials_username, wml_service_credentials_password, and wml_service_credentials_url from the
# Service credentials associated with your IBM Cloud Watson Machine Learning Service instance

wml_credentials={
    "url": wml_service_credentials_url,
    "username": wml_service_credentials_username,
    "password": wml_service_credentials_password
}

headers = urllib3.util.make_headers(basic_auth='{}:{}'.format(username=wml_credentials['username'], password=wml_credentials['password']))
url = '{}{}/v3/identity/token'.format(wml_credentials['url'])
response = requests.get(url, headers=headers)
mltoken = json.loads(response.text).get('token')
```

AutoAI in Watson Studio

In earlier sections we saw how IBM SPSS Modeler and Watson Studio Modeler flows allow you to graphically create a stream or flow that includes **data transformation steps and machine learning models.**

Such sequences of steps are called **data pipelines or ML pipelines.**

This section examines a feature of Watson Studio that

- **helps to automate the creation of machine learning pipelines.**
- This allows data scientists to produce results much faster and to focus on more creative work.

There is currently a shortage of qualified data scientists.

Many operations that a data scientist typically performs are repetitive and time-consuming. Therefore, automating some of that repetitive work will help free up both new and experienced data scientists to do the important work that they are trained to do.

The **AutoAI system**

- **was developed by IBM Research experts in collaboration with IBM**

Distinguished Engineer and two-time Kaggle Grandmaster Jean-Francois Puget.

- It provides a graphical interface to create and deploy machine learning models with real time visualizations.
- AutoAI automatically performs typical machine learning steps, such as:

Data preparation Model selection Feature engineering Hyper-parameter optimization. Users can view the progress on the graphical interface.

This example shows the training of a model to predict whether or not a customer is likely to buy a tent from an outdoor equipment store.

We start with structured data.

In this historical data, there are four feature, or “predictor,” columns:

GENDER: The customer’s gender AGE: The customer’s age

MARITAL_STATUS: “Married”, “Single”, or “Unspecified”

and PROFESSION: The general category of the customer’s profession, such “Hospitality” or “Sales”, or simply “Other.”

The model will learn to predict the value for the IS_TENT column; that is, whether or not the customer bought a tent.

After we choose IS_TENT as the column to predict, AutoAI analyzes the data and determines that

the IS_TENT column contains True/False information, making this data suitable for a binary classification

model.

The default metric for a binary classification is ROC/AUC.

After we click Run experiment, an infographic shows the process of building the pipelines as the model trains.

Once the pipeline creation is complete, we can view and compare the ranked pipelines in a leaderboard.

The pipelines for the sample binary classification model are quite uniform because of the underlying sample data.

To see pipelines in action, re-run the experiment as a regression experiment to predict purchase amount.

That experiment gives better variation in the resulting pipelines.

After clicking “Pipeline comparison,” we can see how the pipelines differ on various measures of model quality.

The pipelines can be saved as Machine Learning assets in the Watson Studio project. Then they can be deployed and tested.

Currently AutoAI is available only for classification and regression models; there is a plan to add time series model support in the future.

In this unit, you have learned how AutoAI automates typical data science tasks and helps get better performing data pipelines more quickly, while also simplifying pipeline deployment into production in Watson Machine Learning.

In the next section, we will discuss Watson OpenScale, which helps to ensure that your models are fair, explainable, and up.

AutoAI in Watson Studio

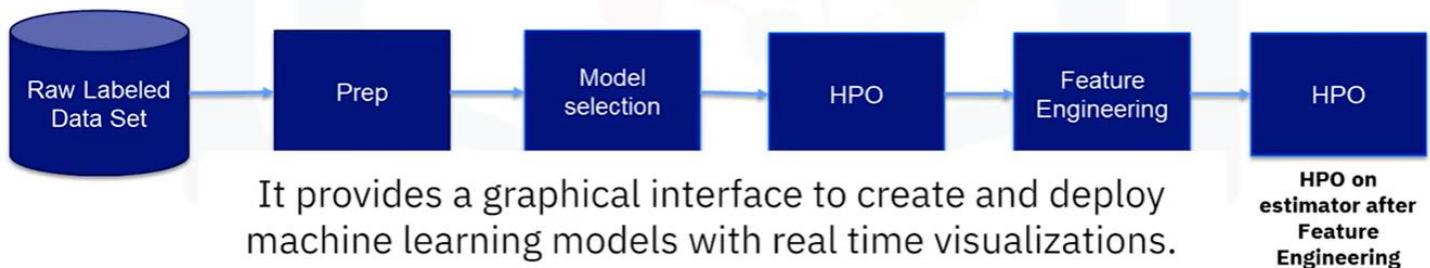
Svetlana Levitan

© IBM Corporation. All rights reserved.

Such sequences of steps are called data pipelines or ML pipelines.

AutoAI features

To help simplify an AI lifecycle management, AutoAI automates:



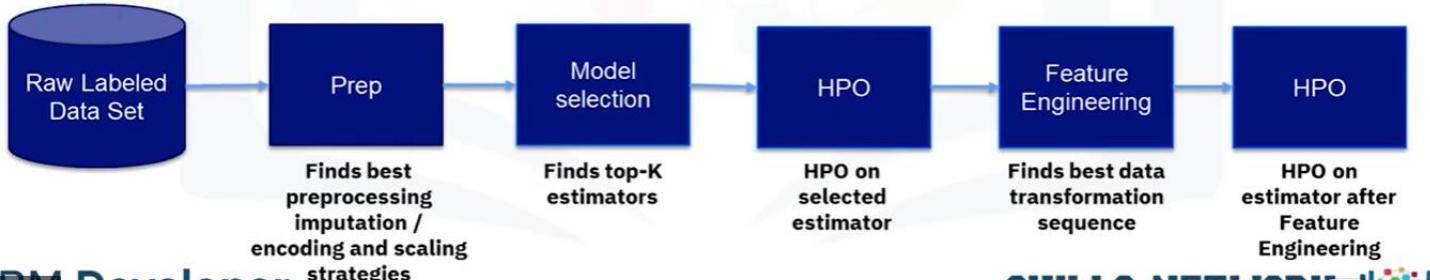
It provides a graphical interface to create and deploy machine learning models with real time visualizations.

Watch later Share

AutoAI features

To help simplify an AI lifecycle management, AutoAI automates:

- Data preparation
- Model development
- Feature engineering
- Hyper-parameter optimization



An example of using AutoAI - data

	A	B	C	D	E	F	G
1	GENDER	AGE	MARITAL_STATUS	PROFESSION	IS_TENT	PRODUCT_LINE	PURCHASE_AMOUNT
2	M	27	Single	Professional	TRUE	Camping Equipment	144.78
3	F	39	Married	Other	FALSE	Outdoor Protection	144.83
4	F	39	Married	Other	FALSE	Outdoor Protection	137.37
5	F	56	Unspecified	Hospitality	FALSE	Personal Accessories	92.61
6	M	45	Married	Retired	FALSE	Golf Equipment	119.04
7	M	45	Married	Retired	FALSE	Golf Equipment	123.76
8	F	39	Married	Other	FALSE	Outdoor Protection	142.23
9	F	49	Married	Other	FALSE	Golf Equipment	105.96
10	F	49	Married	Other	FALSE	Golf Equipment	109.21
11	M	47	Married	Retired	FALSE	Golf Equipment	117.58
12	M	47	Married	Retired	FALSE	Golf Equipment	115.03
13	M	21	Single	Retail	FALSE	Personal Accessories	112.03
14	F	66	Married	Other	FALSE	Golf Equipment	108.11
15	F	35	Married	Professional	FALSE	Golf Equipment	152.95
16	M	20	Single	Sales	TRUE	Mountaineering Equipment	124.66
17	M	20	Single	Sales	TRUE	Mountaineering Equipment	122.39

Set up the model

Select column to predict

DATA SOURCE: Outtaek.csv

Column name	Type
GENDER	String
AGE	Integer
MARITAL_STATUS	String
PROFESSION	String
IS_TENT	Boolean
PRODUCT_LINE	String
PURCHASE_AMOUNT	Decimal

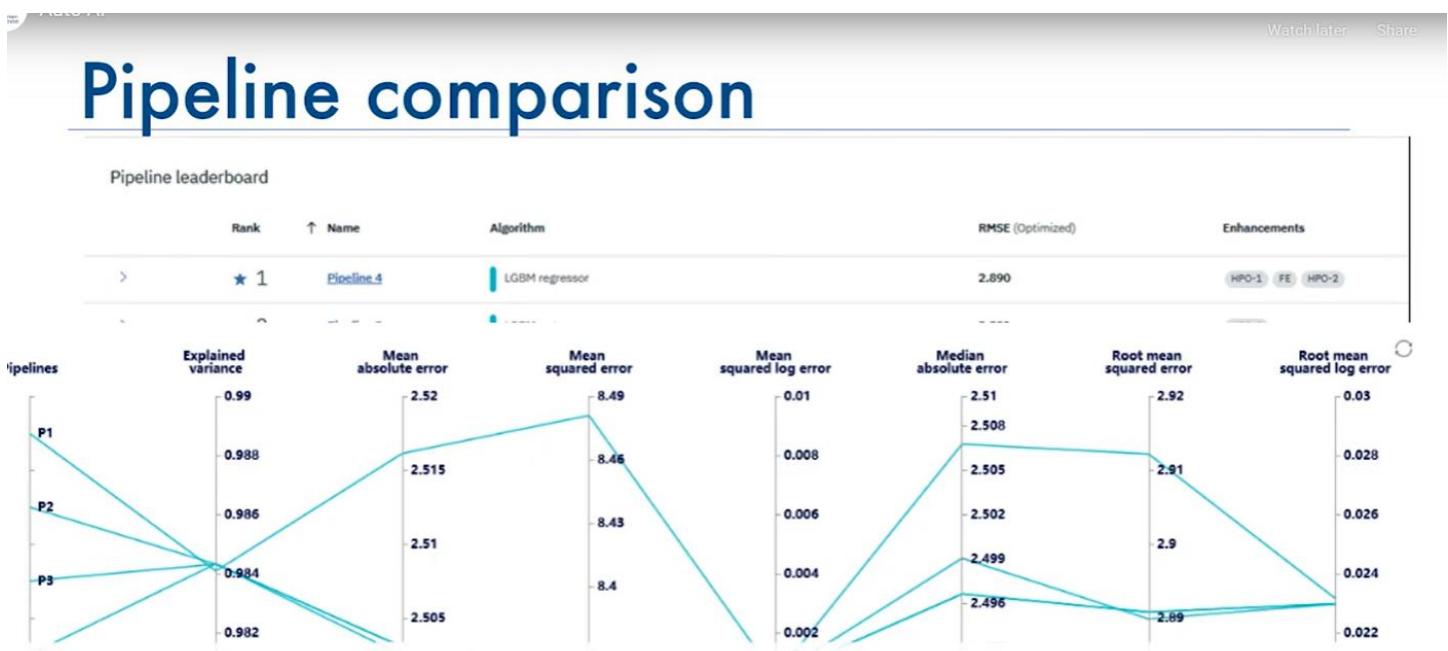
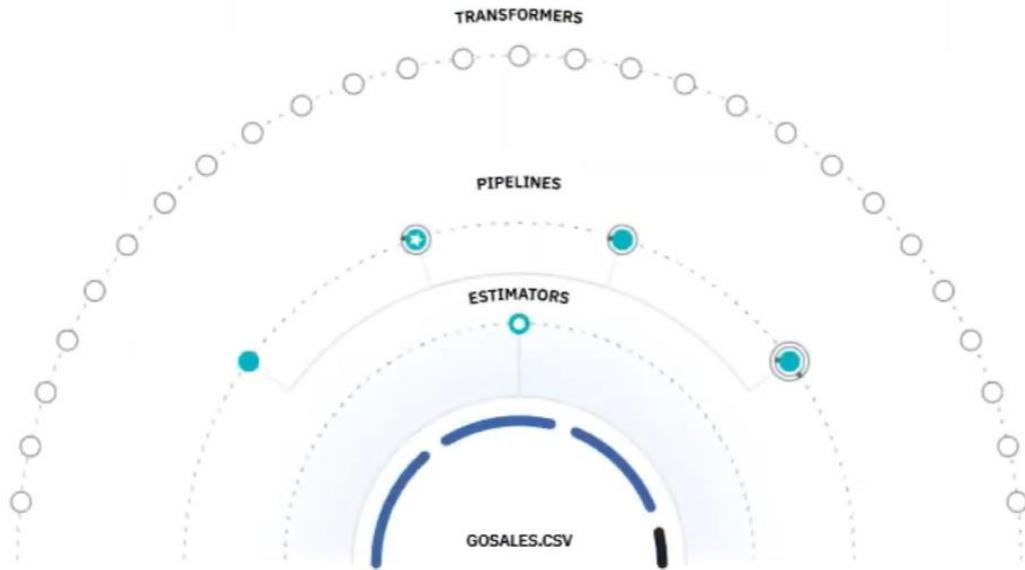
Selected prediction

PREDICTION TYPE: Binary Classification ?

OPTIMIZED METRIC: ROC AUC ?

Save and close Run experiment ○

Running AutoAI to build pipelines



The pipelines can be saved as Machine Learning assets in the Watson Studio project. Then they can be deployed and tested.

IBM Watson Open Scale

This video examines the features of Watson Studio that helps to ensure fairness and explainability of machine learning pipelines, as well as monitor their performance after deployment.

IBM Watson OpenScale is a product that includes several important features.

- It can test the model and its predictions for fairness and apply ways to overcome bias.
- It can also help to provide explanations for model predictions that are often hard to get but are necessary for compliance in some application areas.
- It monitors the model performance and can detect its deterioration, or "model drift," over time.
- It can alert the users when drift is detected and explain which predictors are causing it. We can specify criteria under which the model gets automatically retrained on fresh data.
- It also helps to measure how the model helps the business.

The attributes to monitor for bias are automatically recommended based on prior experience. They can be edited as needed.

- OpenScale then keeps track of model predictions for the specified groups and checks for bias in the predictions.

Users need to know that their AI models are fair, but the data their models were trained on can include unwanted biases which may unintentionally be included in the resulting models.

- IBM Watson OpenScale can detect bias when a model is in production and not just when it's being built.

In this demo of Watson OpenScale, we'll monitor the Credit Risk model which has been trained to determine whether or not someone is eligible for a loan based on a variety of different features such as their credit history, age, and their number of dependents.

After launching OpenScale we can see a metrics for the monitored model such as its quality and a fairness score.

What OpenScale does is measure a model's fairness by calculating the difference between the rates at which different groups, for example, women versus men, receive the same outcome. A fairness value below 100% means that the monitored group receives an unfavorable outcome more often than the reference group.

In this case, we see that women are receiving a "no risk" outcome or getting approved for loans at a lower rate than men.

OpenScale enables the inspection of each model's training data and this reveals that there was more training data for men than women.

This can give some insight as to why the model exhibits bias against women who apply for loans.

Data scientists can use this information to improve the model.

Now detecting bias is one thing, OpenScale can also mitigate it by creating a debiased model that runs alongside the monitored one.

In this case, the debiased model is 12% more fair than the production model.

The debiased model has been trained to detect when your production model will make a bias prediction.

So that you can isolate a specific transaction that results in a bias.

For each of these transactions Watson OpenScale will flip the monitored value in a record to the reference value, in this case from female to male, and leave all other data points in that record the same.

If this changes the prediction from risk to no-risk, then the debiased model will surface the no-risk outcome as the debiased result.

This is just one of the ways Watson OpenScale helps you ensure that your models are fair, explainable and compliant, wherever your model was built or is running. Insurance underwriters can use machine learning and OpenScale to more consistently and accurately assess claims risk, ensure fair outcomes for customers and explain AI recommendations for regulatory and business intelligence purposes.

Why does an AI model arrive at a given recommendation or prediction?

Users and customers want an explanation and with most models providing this information is not an easy task.

IBM Watson OpenScale explains predictions in business-friendly language.

This credit application for instance was predicted to be at risk.

OpenScale determines the features which contributed positively or negatively to that prediction and spells them out.

The explanation is presented visually, as well as in a sentence based text summary in

order to ensure maximum clarity.

Using proprietary IBM research technology, OpenScale also generates a contrast of explanation.

Here we see the minimum changes for this input record which would produce a different output, changing the prediction from risk to no-risk.

The explanations provided by Watson OpenScale can help organizations comply with regulations such as the Fair Credit Reporting Act and GDPR, which give customers the right to ask for reasons why their applications were denied.

Before an AI model is put into production, it must prove it can make accurate predictions on test data, a subset of its training data, however, over time production data can begin to look different than training data, causing the model to start making less accurate predictions.

This is called drift.

IBM Watson OpenScale monitors a model's accuracy on production data and compares it to accuracy on its training data.

When a difference in accuracy exceeds a chosen threshold, OpenScale generates an alert. Watson OpenScale reveals which transactions caused drift and identifies the top transaction. For instance, 25% of the transactions causing drift in this loan approval model were problematic because of these features, which contain data crucially different than the training data.

The transactions causing drift can be sent from manual labeling and used to retrain the model so that its predictive accuracy does not drop at run time.

Watson OpenScale not only helps identify drift but also highlights its root cause and provides transactions which can be turned into training data useful in fixing drift.

It gives you the insight you need to ensure that your models will consistently deliver the results you want over time.

For instance, the retrain version of the model based on the recommendations made by Watson OpenScale, started making accurate recommendations alleviating the drift.

This is just one of the ways that Watson OpenScale helps you ensure your models are fair, explainable

and compliant wherever your model was built or is running.

In this video, you have learned how OpenScale ensures fairness and explainability of models and monitors for model drift in production.

OpenScale features

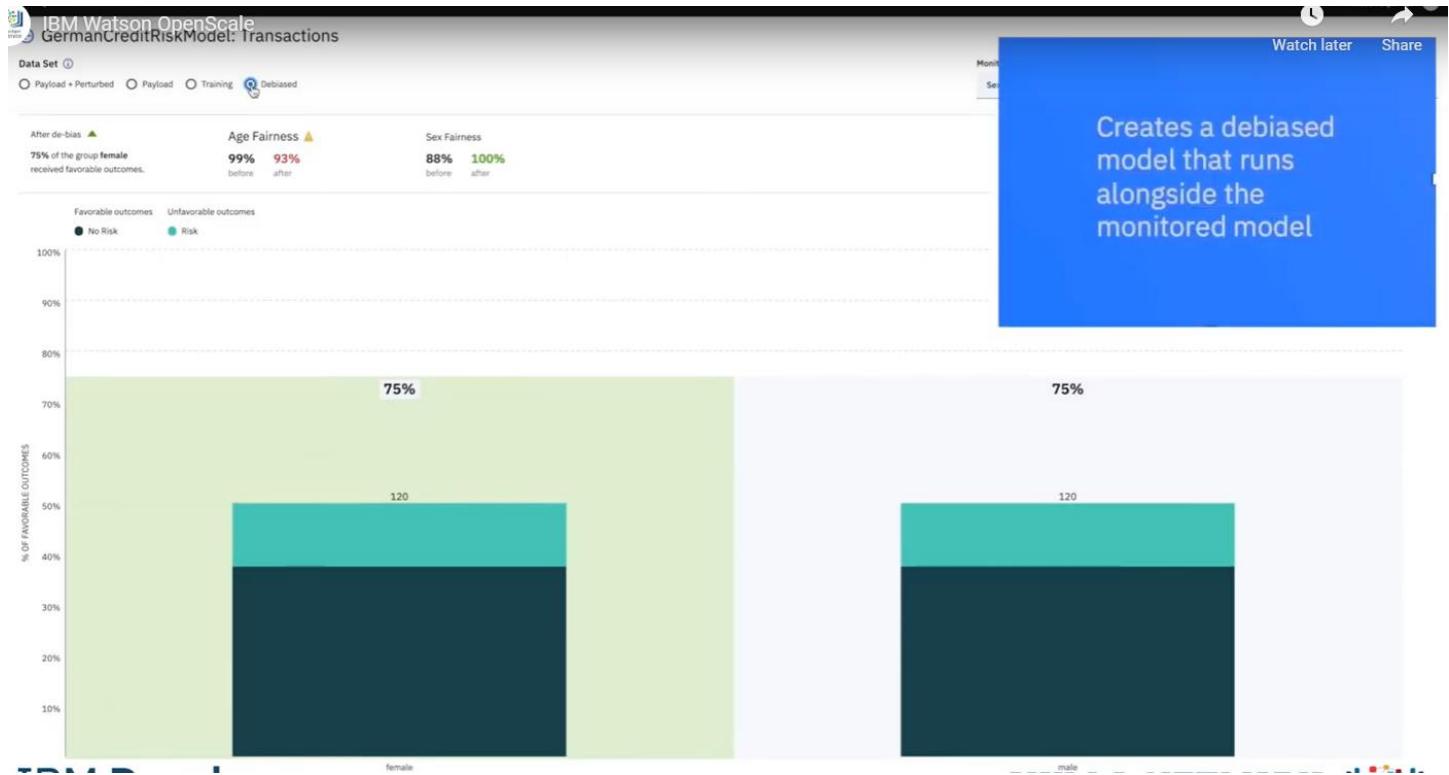
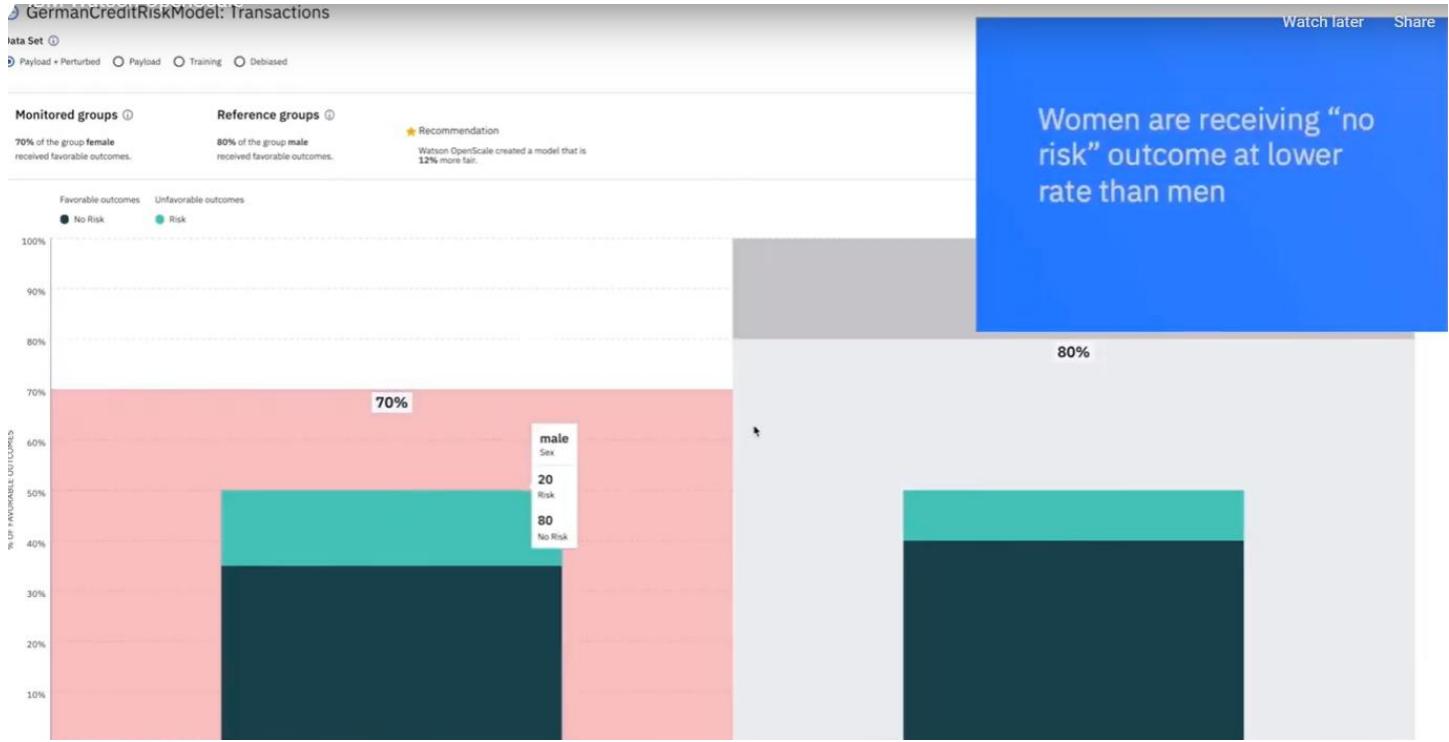
- **Fairness:** Detect and mitigate model bias; automatically recommend which attributes to monitor for bias.
- **Explainability:** audit and explain model decisions.
- **Model monitoring:** Monitor model performance, help to find causes when model drift is detected, possibly trigger retraining
- **Business impact:** Correlate model metrics and business KPIs to measure business impact; actionable metrics and alerts

The dashboard displays three deployed models:

- Watson Machine Learning GermanCreditRiskModel: 3 Deployments Monitored, 1 Quality Alert, 1 Fairness Alert. Issues: 2. Quality: 69%, Fairness: 90% (1 of 2 attributes reported). Status: Evaluation pending.
- Watson Machine Learning drug-test-deploy: 0 Issues. Quality: N/A, Fairness: N/A. Status: Evaluation pending.
- Watson Machine Learning Heart-Drug-Deployment: 0 Issues. Quality: N/A, Fairness: N/A. Status: Evaluation pending.

Metrics are updated hourly. A blue callout box on the right states: "IBM Watson OpenScale can detect bias when a model is in production".

This page shows the fairness monitoring for the GermanCreditRiskModel. It includes a sidebar with various metrics like Age, Sex, Area under ROC, and Precision, many of which have red warning icons. The main area displays a line chart titled "Fairness for Sex" showing the Fairness Score over time (Hourly, Daily, Weekly, Past 3 months, Past week, Yesterday, Today, Custom range from 8/23/2019 to 8/25/2019). The chart shows a score fluctuating around 100%. A blue callout box in the center states: "Fairness compares the difference between rates at which different groups receive the same outcome". On the right, there's a summary section with a Fairness Score of 97%, a note about being 1% below threshold, and details about monitored groups (Average female: 97%, Bias: 97%). Below that is a "Schedule" section with evaluation times and a link to check fairness now.



IBM Watson OpenScale
GermanCreditRiskModel: Transactions
August 24, 2019, 5:00 AM

Watch later Share
Sex

View All transactions Biased transactions

Review the transactions that were received for the selected time period. Click an Explain link to determine the features that contributed to each outcome.

Transaction ID	Sex	Bias Detected	Outcome	Action
0a5f44790c6d4c008979e776aaeb7f10-1	male	Risk	Explain	
10aca08977894ae09456e00a1f89941a-1	male	Risk	Explain	
125c52c7fb0d4465ea1c4e20a44e88e60-1	male	No Risk	Explain	
16cb52350cf543298ab2c1fed56236-1	male	No Risk	Explain	
19eb57b96b7041a7ba7e7ead9b17f3-1	male	No Risk	Explain	
1ab5b10b31e94085a6469c6dd09134be-1	male	No Risk	Explain	
1240c191123428d09ef1f785d4a521-1	male	No Risk	Explain	
2164ec326959487e0a0ff8a7ccad99d1d-1	male	No Risk	Explain	
288e7ceaf284542be90a904eda56db-1	male	Risk	Explain	
30349682679149f3913ee695f5b08915-1	male	No Risk	Explain	
337b3293019473fa27007dc134d3e95-1	male	No Risk	Explain	
36c42c2b0fd5401380904dbd84052a7-1	male	No Risk	Explain	
373e333960dd426a64891ad8777c07-1	male	No Risk	Explain	
378a2b1e8c342858df1491af8269a7a-1	male	No Risk	Explain	
3845782229d419294fe6b86a79d258-1	female	No Risk	Explain	
3944598f673a43029397c7a22b965bbf-1	female	Risk	Explain	
39e0c0fa5f994e45892e0482baea9fc0-1	male	Risk	Explain	
3c9212cd9fd54769a2c3207273b59781-1	male	No Risk	Explain	

Payload Table Payload_3999afe5-35bf-466b-a882-ceabd9961ddf
Corrected Records -

No Risk : Favorable Outcome
Current Model: 70.00%
De-biased Model: 75.00%

Risk : Unfavorable Outcome
Current Model: 30.00%
De-biased Model: 25.00%

Isolates transactions causing bias

IBM Watson OpenScale
GermanCreditRiskModel: Transactions
August 24, 2019, 5:00 AM

Watch later Share
Sex

All transactions Biased transactions

This subset of transactions received biased outcomes. Click the Explain link to determine how the monitored feature contributed to each unfavorable outcome. ⓘ

Transaction ID	Type	Outcome	Action
36c42cf0fd5401380904dbd84052a7-1	Altered	Risk	Explain
9c09c2d5540249e782da3f9c77a7c3-1	Altered	No Risk	Explain
3c9212cd89d54769a2c3207273b59781-1	Original	No Risk	Explain
e44281d106754393a5f40e21a634f9d9-1	Original	No Risk	Explain
3c9212cd89d54769a2c3207273b59781-1	Altered	Risk	Explain
e44281d106754393a5f40e21a634f9d9-1	Altered	Risk	Explain
6b0da069492e46a2b108fed0c362f79-1	Altered	Risk	Explain
36c42cf0fd5401380904dbd84052a7-1	Original	No Risk	Explain
9c09c2d5540249e782da3f9c77a7c3-1	Original	Risk	Explain
6b0da069492e46a2b108fed0c362f79-1	Original	No Risk	Explain

Fairness Correction Table Manual_Labeling_3999afe5-35bf-466b-a882-ceabd9961ddf

No Risk : Favorable Outcome
Current Model: 70.00%
De-biased Model: 75.00%

Risk : Unfavorable Outcome
Current Model: 30.00%
De-biased Model: 25.00%

Flips monitored value to reference value, in this case from female to male, and leaves all other data the same

IBM Watson OpenScale
GermanCreditRiskModel: Transactions
August 24, 2019, 5:00 AM

All transactions Biased transactions

This subset of transactions received biased outcomes. Click the Explain link to determine how the monitored feature contributed to each unfavorable outcome. ⓘ

Transaction ID	Type	Outcome	Action
36c42cf0fd5401380904dbd84052a47-1	Altered	Risk	Explain
9c09c2d5540249e782da23f69c77a7c3-1	Altered	No Risk	Explain
3c9212cd8fd54769a2c3207273b59781-1	Original	No Risk	Explain
e64281d106754393a5f40e21a634f9d9-1	Original	No Risk	Explain
3c9212cd8fd54769a2c3207273b59781-1	Altered	Risk	Explain
e64281d106754393a5f40e21a634f9d9-1	Altered	Risk	Explain
6b0da069492e46a2b108fedb0c362f79-1	Altered	Risk	Explain
36c42cf0fd5401380904dbd84052a47-1	Original	No Risk	Explain
9c09c2d5540249e782da23f69c77a7c3-1	Original	Risk	Explain
6b0da069492e46a2b108fedb0c362f79-1	Original	No Risk	Explain

[Load more](#)

Fairness Correction Table ⓘ Manual_Labeling_3999afe5-35bf-466b-a882-ceabdf961ddff

Category	Current Model	De-biased Model
No Risk : Favorable Outcome	70.1	75.1
Risk : Unfavorable Outcome	30.1	25.1

If this changes the prediction from risk to no-risk, then the debias model will surface the no-risk outcome as the debiased result

IBM Watson OpenScale
GermanCreditRiskModel: Transactions
August 24, 2019, 5:00 AM

All transactions Biased transactions

This subset of transactions received biased outcomes. Click the Explain link to determine how the monitored feature contributed to each unfavorable outcome. ⓘ

Transaction ID	Type	Outcome	Action
36c42cf0fd5401380904dbd84052a47-1	Altered	Risk	Explain
9c09c2d5540249e782da23f69c77a7c3-1	Altered	No Risk	Explain
3c9212cd8fd54769a2c3207273b59781-1	Original	No Risk	Explain
e64281d106754393a5f40e21a634f9d9-1	Original	No Risk	Explain
3c9212cd8fd54769a2c3207273b59781-1	Altered	Risk	Explain
e64281d106754393a5f40e21a634f9d9-1	Altered	Risk	Explain
6b0da069492e46a2b108fedb0c362f79-1	Altered	Risk	Explain
36c42cf0fd5401380904dbd84052a47-1	Original	No Risk	Explain
9c09c2d5540249e782da23f69c77a7c3-1	Original	Risk	Explain
6b0da069492e46a2b108fedb0c362f79-1	Original	No Risk	Explain

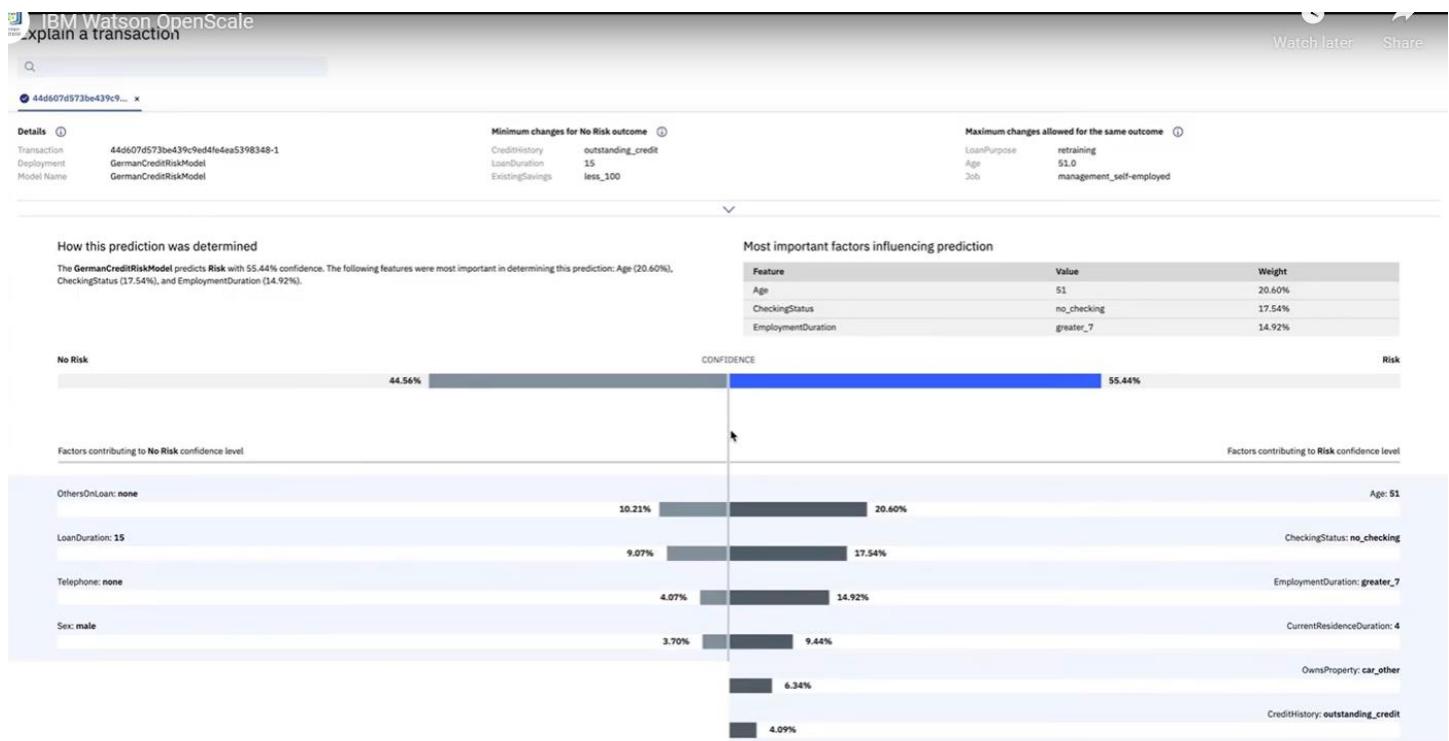
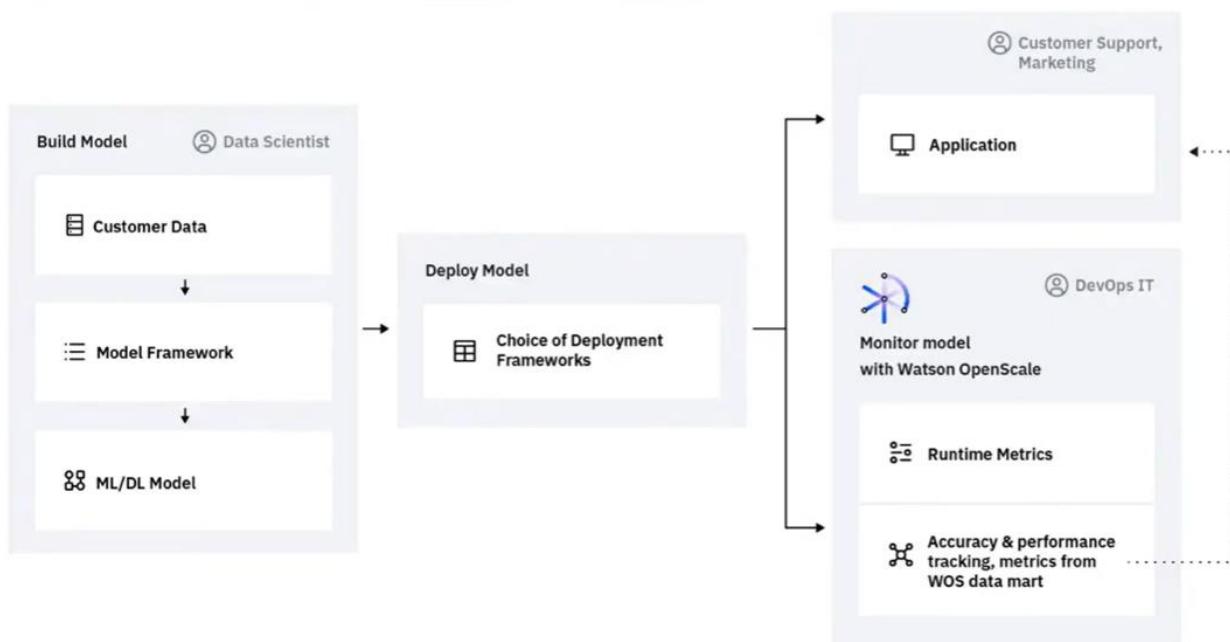
[Load more](#)

Fairness Correction Table ⓘ Manual_Labeling_3999afe5-35bf-466b-a882-ceabdf961ddff

Category	Current Model	De-biased Model
No Risk : Favorable Outcome	70.00%	75.00%
Risk : Unfavorable Outcome	30.00%	25.00%

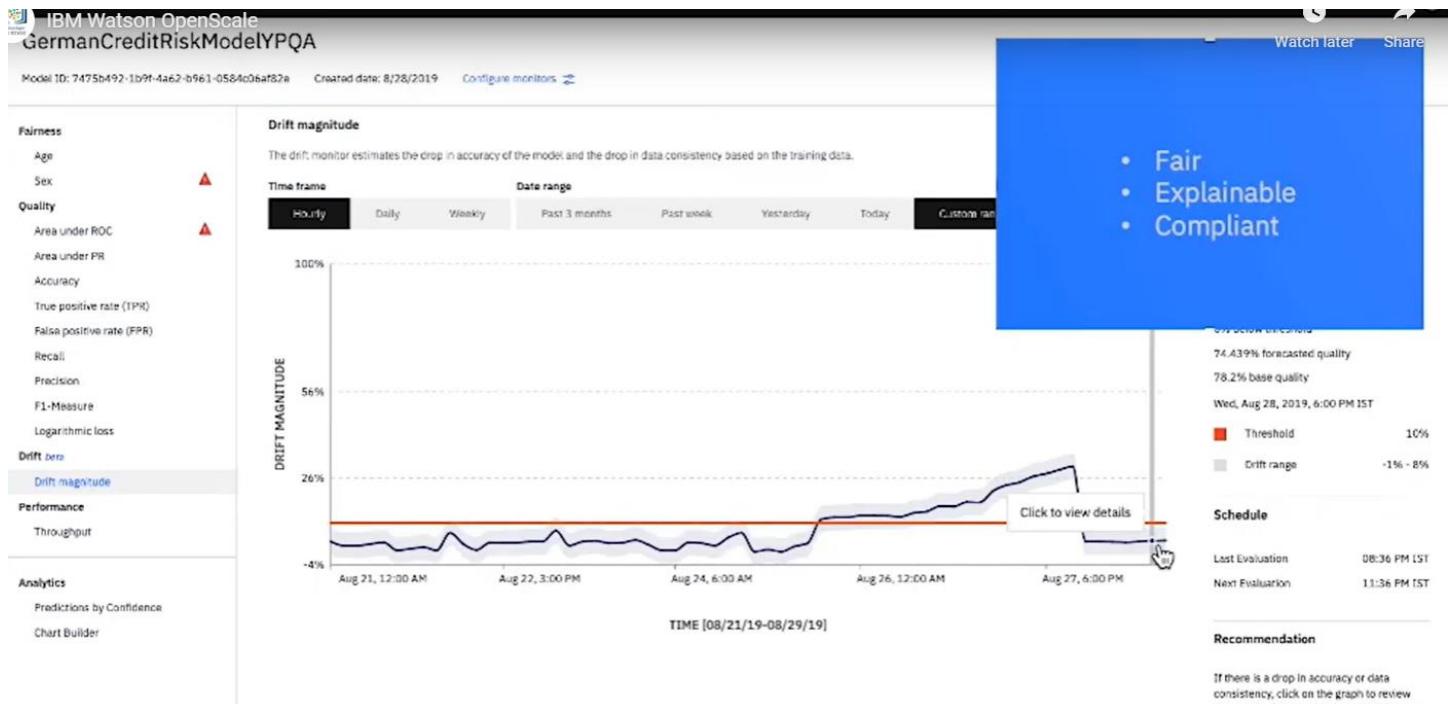
- Fair
- Explainable
- Compliant

Explaining model predictions



Transactions causing drift can become training data useful in fixing drift

		Model Output	ExistingSavings	OwnsProperty	EmploymentDuration	LoanAmount	Age	Confidence	Actions
AM	No Risk	100_to_500	savings_insurance	1_to_4	2165	37	84.0%	Explain prediction	
AM	No Risk	less_100	car_other	4_to_7	4892	31	75.1%	Explain prediction	
AM	No Risk	500_to_1000	car_other	4_to_7	2192	34	68.5%	Explain prediction	
AM	No Risk	500_to_1000	car_other	greater_7	250	36	57.0%	Explain prediction	
ff6b54adca435c87db1d16ecd8fc-1	Aug 26, 2019, 10:34:50 AM	No Risk	500_to_1000	unknown	4_to_7	4880	43	57.5%	Explain prediction
37b7f6e80dea41d58eccbdb7c849ba52-1	Aug 26, 2019, 09:34:50 AM	No Risk	500_to_1000	car_other	1_to_4	3355	48	60.1%	Explain prediction
46328aad12bb4bd2b926f5a260a3729e-1	Aug 26, 2019, 09:34:50 AM	No Risk	100_to_500	car_other	1_to_4	4153	41	68.0%	Explain prediction
b1ae76ae834842788c0159e38f0f06b-1	Aug 26, 2019, 09:34:50 AM	Risk	100_to_500	unknown	1_to_4	9016	46	74.6%	Explain prediction
f36dale2091e43bdf5deb617de7f448-1	Aug 26, 2019, 09:34:50 AM	No Risk	100_to_500	car_other	4_to_7	3988	44	85.3%	Explain prediction
25662b5e21514fe9bd47da10208f22b-1	Aug 26, 2019, 09:34:50 AM	No Risk	greater_1000	car_other	4_to_7	3540	46	70.3%	Explain prediction
4a48d90c63364478ae87ed06c9953a3a-1	Aug 26, 2019, 09:34:50 AM	Risk	500_to_1000	car_other	4_to_7	3324	52	78.4%	Explain prediction
87f1d19b8434406c95b51d2cda9cc9f9-1	Aug 26, 2019, 09:34:50 AM	Risk	500_to_1000	car_other	4_to_7	5520	43	78.2%	Explain prediction
bc313f5505f74f61924012df89b5442-1	Aug 26, 2019, 09:34:50 AM	No Risk	500_to_1000	savings_insurance	4_to_7	2074	42	57.8%	Explain prediction
13e93bc98af40dcbd1a5db105a1933a-1	Aug 26, 2019, 08:34:50 AM	No Risk	500_to_1000	car_other	1_to_4	5093	48	63.6%	Explain prediction
2e5eccb5102f412e821d3c74a36d0629-1	Aug 26, 2019, 08:34:50 AM	No Risk	less_100	car_other	1_to_4	1773	20	57.4%	Explain prediction
58a0abc3f6724ca8b422cfef555c90291-1	Aug 26, 2019, 08:34:50 AM	No Risk	100_to_500	car_other	1_to_4	1471	33	56.7%	Explain prediction



TRM Developer

CIVIC NETWORKS

Practice Quiz – Other IBM Tools

Q1. Which products (of those we covered) allow you to build data pipelines using a graphical user interface and no coding?

- **IBM SPSS Modeler and Modeler Flows in Watson Studio**

Q2. Which features of Data Refinery help save hours and days of data preparation?

- **All of the above**

Q3. Watson Knowledge Catalog provides what functionality?

- **Catalog data and ML assets, help to find relevant assets, keep track of asset lineage, enforce data governance**

Q4. Fill in the blank: PMML, PFA, and ONNX are _____.

- **Opens standards for predictive model serialization, exchange, and deployment**

Q5. Which node must be used in Modeler flows before any modeling node?

- **Type node**

Q6. Fill in the blank: Auto Classification node can be used for data with _____.

- **A categorical target variable**

Q7. Fill in the blank: Auto Numeric node can be used for data with _____.

- **A continuous target variable**

Q8. IBM SPSS Modeler evolved from which product?

- **Clementine**

Q9. Fill in the blank: IBM SPSS Statistics syntax can be created using _____.

- **Graphical user interface the IBM SPSS statistics product or syntax editor**

Q10. AutoAI provides which of the following services?

- **Automatic finding of optimal data preparation steps model selection and hyperparameter optimization**

Q11. OpenScale provides which of the following services?

- **Monitoring for fairness, bias, and model drift**

Q12. Predictive Model Markup Language (PMML) was created by which entity?

- **The Data Mining Group**

Graded Quiz >> Week 3 >> Tools for Data Science

1. Which feature in Watson Studio helps to keep track of and discover relevant Machine Learning assets?

- OpenScale
- AutoAI
- Modeler Flows
- **Watson Knowledge Catalog**
- All of the above

2. Data Refinery provides which of the following services?

- Catalog the data assets.
- Monitor for bias and model drift.
- **Visualize and prepare data.**
- Automatically build models.
- All of the above.

3. How does Data Refinery help build repeatable Data Pipelines for workloads of almost any size?

- **Create a scheduled Job and use a custom environment to run the data flow/pipeline on different workloads.**
- Not supported.
- Feature is available only in the UI, not API.
- Only a fixed workload size is supported.
- Manually write APIs to provide automation.

4. Modeler flows in Watson Studio always begin with which type of node?

- **A modeling node**
- A type node
- A data source node
- An output node
- All of the above

5. IBM SPSS Modeler includes what kind of models?

- Classification models (for data with a categorical target).
- Regression models (for data with a continuous target).
- Clustering models (for data with no target variables).
- Other kinds of models.
- **All of the above.**

8. Fill in the blank: If you'd like to schedule a notebook in Watson Studio to run at a different time, you can create a(n) _____.

- API
- markdown cell
- asset
- **job**

9. Fill in the blank: In the _____ tab you can define the hardware size and software configuration for the runtime associated with Watson Studio tools such as Notebook.

- settings
- assets
- overview
- **environments**

6. What feature of IBM SPSS Statistics allows easy saving and modifying of previous tasks?

- Charts
- Graphical user interface
- **SPSS syntax**
- SPSS Modeler streams
- All of the above

7. Open Neural Network eXchange (ONNX) was originally created for what models?

- Decision trees
- **Deep learning models.**
- Support Vector Machines (SVM).
- Clustering models
- Regression models

8. Fill in the blank: If you'd like to schedule a notebook in Watson Studio to run at a different time, you can create a(n) _____.

- API
- markdown cell
- asset
- **job**

9. Fill in the blank: In the _____ tab you can define the hardware size and software configuration for the runtime associated with Watson Studio tools such as Notebook.

- settings
- assets
- overview
- **environments**

10. Fill in the blank: It's a best practice to remove or replace _____ before publishing to GitHub.

- **credentials**
- markdown text
- code cells
- charts

Welcome to the final assignment!

Objective

- In this final assignment you will import a notebook into Watson Studio, modify it, run it, and share it with your peers to be graded.

Instructions:

For this assignment, you must import a Jupyter Notebook from GitHub into Watson Studio. You must then change the code in the notebook so that the chart displays "Corona" cases from a country other than the United States. Finally, you must save the notebook and share it with your peers for grading.

Step 1. Log in to your IBM Cloud account and launch Watson Studio.

- If you do not yet have an account, please follow this [link](#), by right-clicking and selecting copy link.
- If you're not sure how to import notebooks from GitHub into Watson Studio, read this [tutorial](#).

Step 2. In Watson Studio, create a new project, and import this [notebook](#).

Step 3. The notebook displays a chart of "Corona" cases from the "United States". Change the code so that the chart displays the "Corona" cases from a different country.

Step 4. Once you've changed the notebook, in the Cell menu, click Run All cells.

Step 5. You must save your notebook before you can share it. Click File, then click Save, as shown below:

My projects / project202002 / a7_final

File Edit View Insert Cell Kernel Help

Save

Save Version

Revert To Version

```
os
hashlib
os.path.abspath(os.getcwd())
coded = cwd.encode('utf-8')
b.md5(cwd_encoded).digest()
```

Print Preview

Download as

Trusted Notebook

Stop Kernel

Step 6. Now, click share  and choose to **Share with anyone who has the link**

Step 7. Activate sharing by clicking Share with anyone who has the link, and then select All content, including code.

Step 8. Click Copy link  to copy the URL to your notebook to the clipboard.

Step 9. Submit the link in the final assignment for your peers to review.