

IBM Watson Studio

(IBM Watson Studio)



Manual para Workshop

Abril 2018

Índice

Introducción

Workshop 1. Comenzar con WATSON STUDIO

- 1.** Registrarse
- 2.** Añadir usuarios adicionales
- 3.** Comenzar con WATSON STUDIO

Workshop 2. Crea y organiza los recursos en un proyecto

- 1.** Crear un proyecto
- 2.** Añade colaboradores
- 3.** Aprende a gestionarlo: borrar assets, añadir, crear conexiones, ...

Workshop 3. Consigue y prepara los datos y analiza los datos de manera sencilla

- 1.** Añadir datos
- 2.** Crear un modelo automático
- 3.** Despliega el modelo en Watson Machine Learning
- 4.** Crear un modelo semi-automático o manual

Workshop 4. Analiza en profundidad los datos

- 1.** Notebooks
- 2.** Visualizaciones
- 3.** Algoritmos de analítica predictiva de SPSS
- 4.** RStudio
- 5.** Librerías de Deep Learning

Workshop 5. Recursos para aprender DSX Local

Introducción

IBM Watson Studio es un entorno que reúne todo lo que necesita un Data Scientist. Incluye las herramientas de código abierto más populares, además los equipos de data scientist de IBM han unido al código abierto una serie de funcionalidades que aportan un gran valor añadido, todo integrado a la perfección en esta única herramienta para que tanto el análisis como los usuarios sean más efectivos y eficaces.

Actualmente existen dos: IBM DSX Local y en Cloud. En este workshop nos centraremos en IBM Watson Studio en Cloud, es una plataforma híbrida que interconecta los datos con servicios analíticos, dando solución a los problemas típicos a los que se enfrentan las empresas. Como pueden ser: mala colaboración entre los componentes de un equipo, falta de confianza en el dato o falta de confianza en el resultado, problemas de seguridad, etc.

Watson Studio se crea basándose en tres pilares fundamentales: aprender crear y colaborar.



Learn

Get started or get better with built-in learning.



Create

Use the best of open source tooling with IBM innovation.



Collaborate

Work smarter using community, work faster with your team.

Aprender:

WATSON STUDIO cuenta con herramientas de aprendizaje incluidas, con numerosos tutoriales de niveles que van desde niveles básicos a avanzados para que cualquiera pueda empezar a disfrutar de la herramienta. Además, puedes complementar el aprendizaje con los cursos y clases gratuitos sobre Data Mining y machine learning uniéndote a los más de 400,000 usuarios registrados en Cognitive Class.

Utiliza los conjuntos de datos, código de ejemplo, tutoriales y artículos técnicos que están a disposición de los usuarios.

Crear:

Watson Studio recomienda fusionar lo mejor del código abierto, con el valor añadido que aportan las herramientas de IBM para crear modelos de datos punteros. Además, WATSON STUDIO cuenta con una gran inversión en Spark, líder en la industria (posee más de 3500 desarrolladores e investigadores).

Gracias a WATSON STUDIO puedes usar el código abierto y las herramientas potentes de analítica avanzada de modo integrado, gobernado y seguro.

Colaborar:

Las características colaborativas proporcionan una ayuda importante para aumentar la productividad y el impacto en el negocio.

Con Watson Studio puedes administrar los recursos del proyecto y la colaboración de los usuarios además de poder compartir, bifurcar y reutilizar assets con Github.

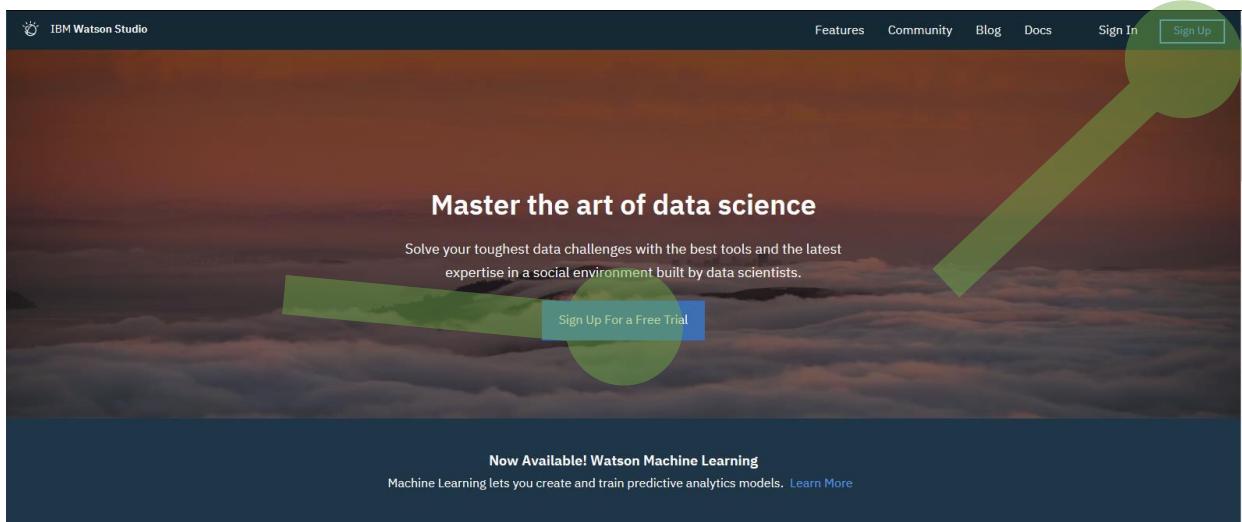
Workshop 1.

1. Registrarse

Watson Studio es una plataforma que permite a los equipos colaborar, compartir datos y modelos y poner en producción dichos modelos.

Para crearse una cuenta nueva, comienza por registrarte. Después de registrarte, puedes agregar otras aplicaciones de Watson Studio en cualquier momento desde la propia herramienta.

Para registrarse Entra en <https://datascience.ibm.com/> Haciendo Sing Up



Si aún no tienes una cuenta IBMid e IBM Cloud, se crearán durante el proceso de suscripción.

Try Watson

Powered by IBM Cloud

Start building smarter applications and quickly visualize, share and gain insights from your data with IBM Watson. No credit card required.

Watson Studio
Democratize ML/DL to accelerate infusion of AI in your business.
[Remove](#)

Try our available applications

Watson Knowledge Catalog
Securely discover, catalog, and govern enterprise data.
[Select](#)

Register for IBM Cloud

To sign up for IBM Watson, you must first register for IBM Cloud.

Create your IBM Cloud Account
candela.dsx@hotmail.com

Accept the terms
 I accept the [IBM Watson terms](#), the [IBM Cloud privacy policy](#) and [IBM Cloud terms](#).

Next

IBM Cloud

Sign up for an IBMid and create your IBM Cloud account
Try IBM Cloud free for 30 days

Start building immediately.
Your trial doesn't require a credit card. All you need to do is sign up and start building.

Production app? No problem.
We give you 2GB of runtime and container memory free for 30 days, plus access to provision up to 10 services.

We're here to help.
Your trial comes with free help desk support. Ask us anything along the way.

Email*
candela.dsx@hotmail.com

First Name*

*Enter your given name.

Last Name*

Company

Country or Region*
United Kingdom

Password*

Keep me informed of products, services, and offerings from IBM companies worldwide.
 By email

By clicking Create Account, I accept the [IBM Cloud privacy policy](#) and [IBM Cloud terms](#).

Create Account

Privacy Terms



Sign up for an IBMid and create your IBM Cloud account
Try IBM Cloud free for 30 days

Start building immediately.
Your trial doesn't require a credit card. All you need to do is sign up and start building.

Production app? No problem.
We give you 2GB of runtime and container memory free for 30 days, plus access to provision up to 10 services.

We're here to help.
Your trial comes with free help desk support. Ask us anything along the way.


Company

Last Name*

First Name*

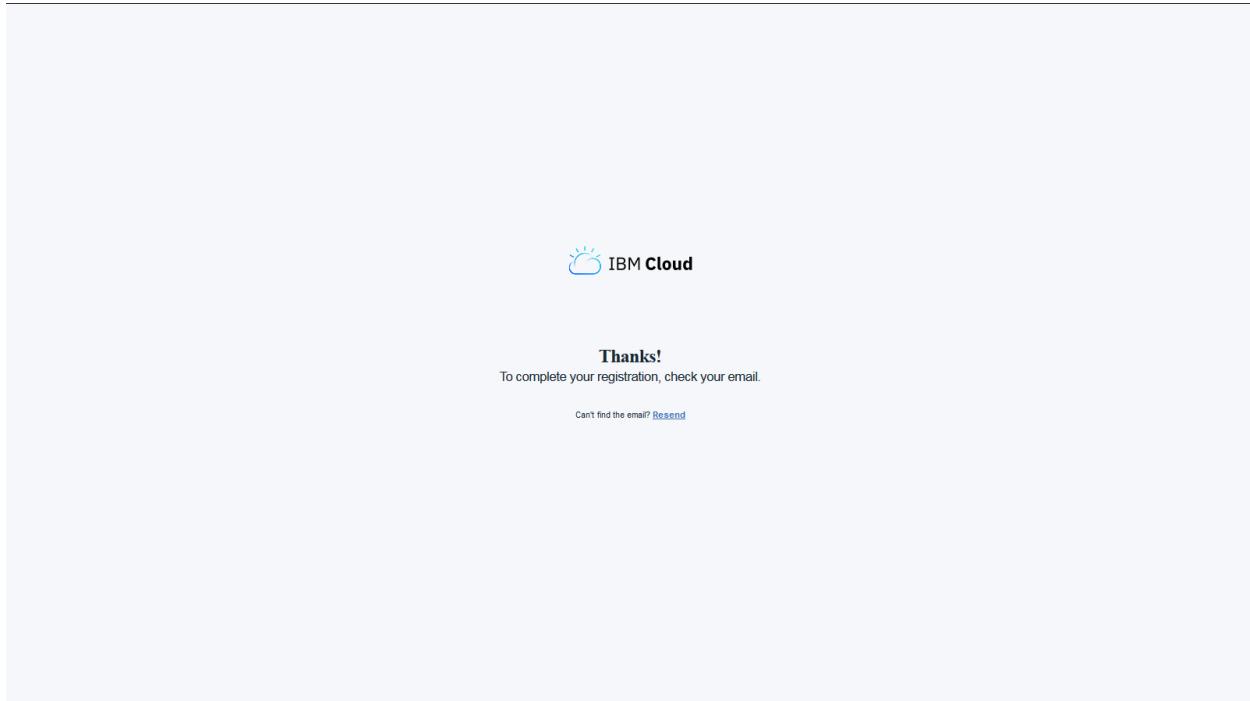
Country or Region*

Password*

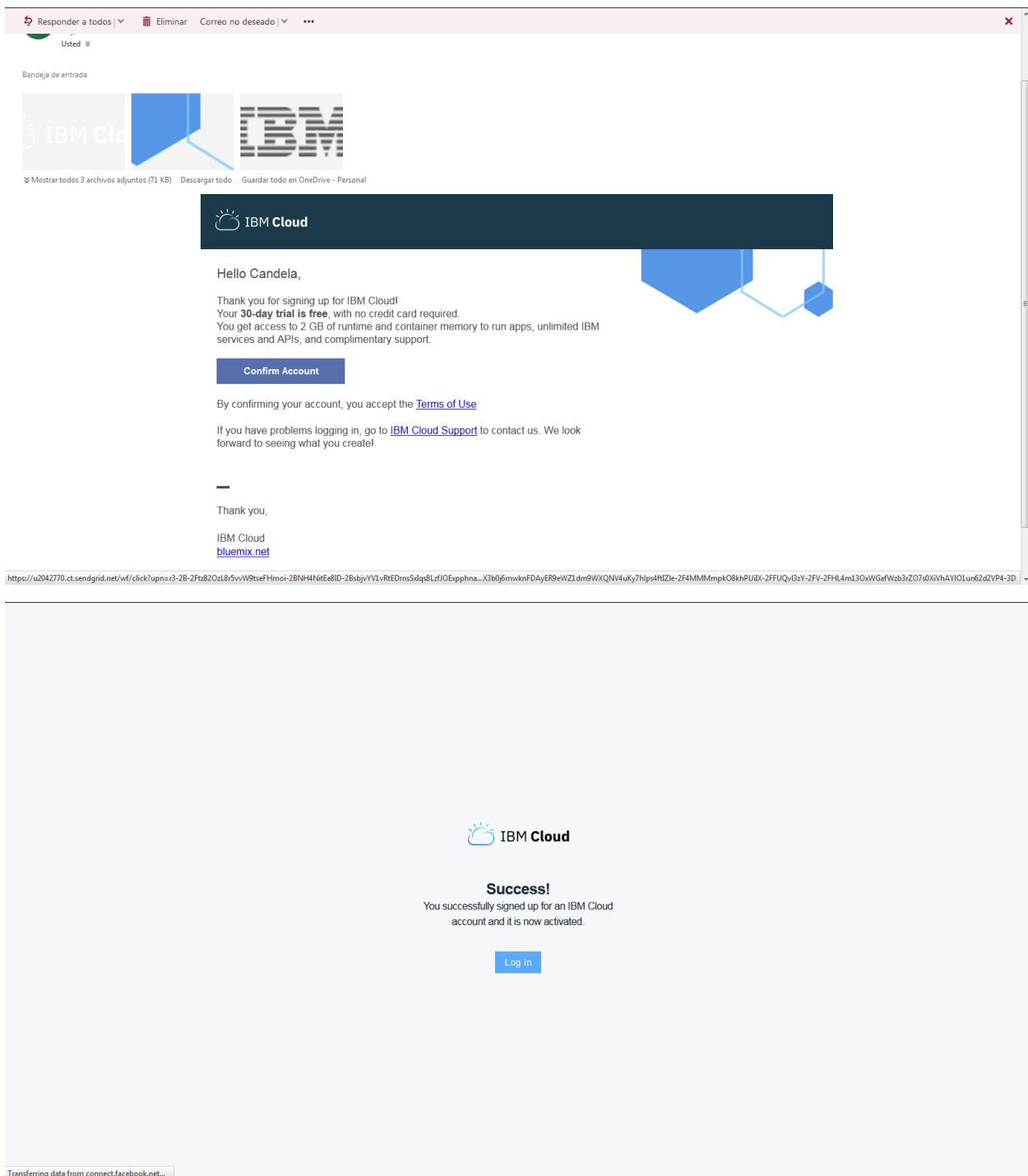
Keep me informed of products, services, and offerings from IBM companies worldwide.
 By email
 By clicking Create Account, I accept the [IBM Cloud privacy policy](#) and [IBM Cloud terms](#).

Create Account

Privacy Terms



Entramos al correo para confirmar la cuenta:



Y nos envía de vuelta a Watson Studio, nos logeamos.

Sign in to IBM

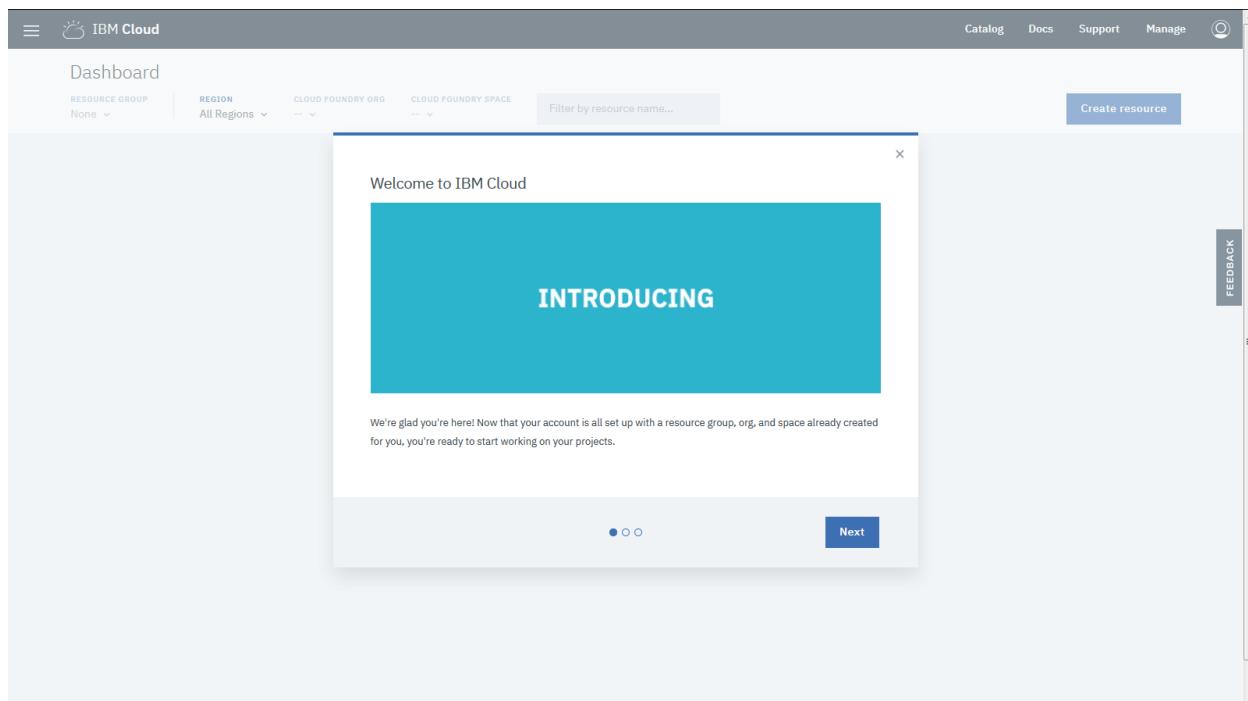
Enter IBMid or email
[Forgot your IBMid?](#)

[Continue](#)

[New? Create an IBMid.](#)

<https://idaas.iam.ibm.com/idaas/mtfim/sp/s/authsvc?PolicyId=urn:ibm:security:authentication:arf:basicDepUser>

Nos da la bienvenida a IBM Cloud.



En IBM Cloud, podemos crearnos las herramientas que necesitemos para construir nuestros modelos.

The screenshot shows the IBM Cloud dashboard with the 'Watson' service selected. A modal window titled 'What do you want to build?' is open, displaying a screenshot of the Watson service interface with various options like Functions, Integrate, Internet of Things, Mobile, Schematics, Security, Storage, and Watson.

Create a Cloud Foundry app
Go straight to developing with a Liberty for Java runtime, then add some of our 100+ services to build your app even faster.

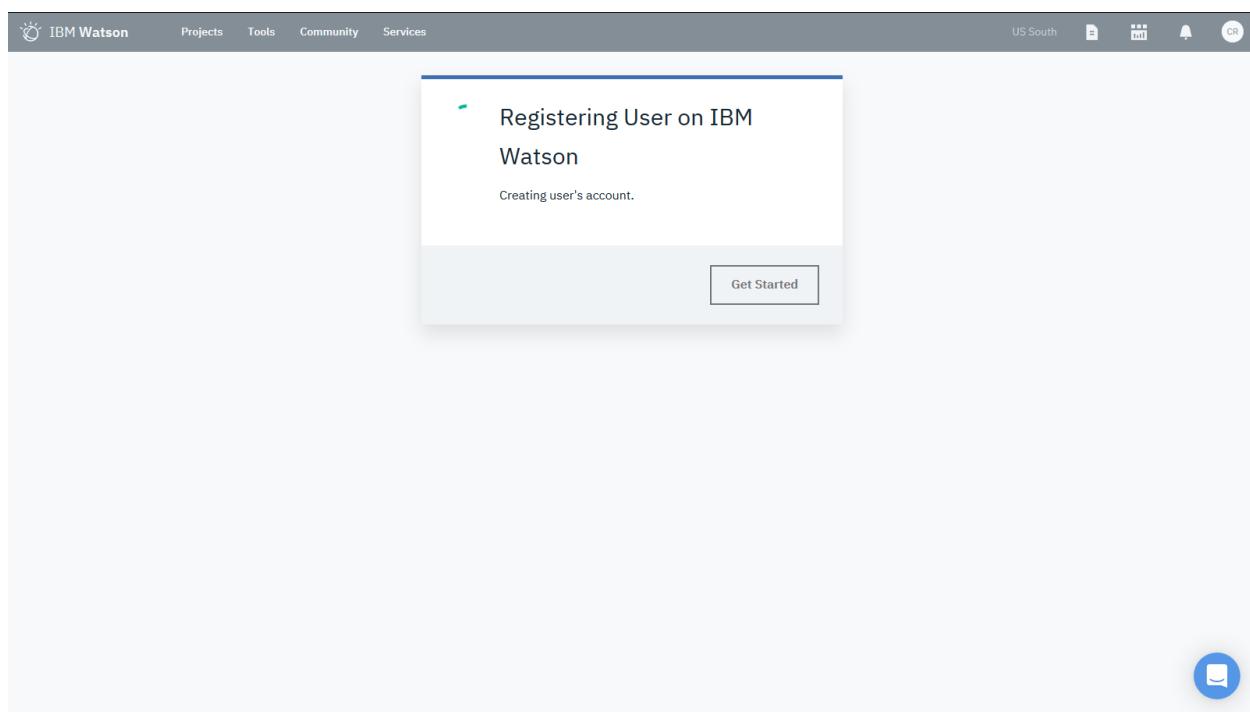
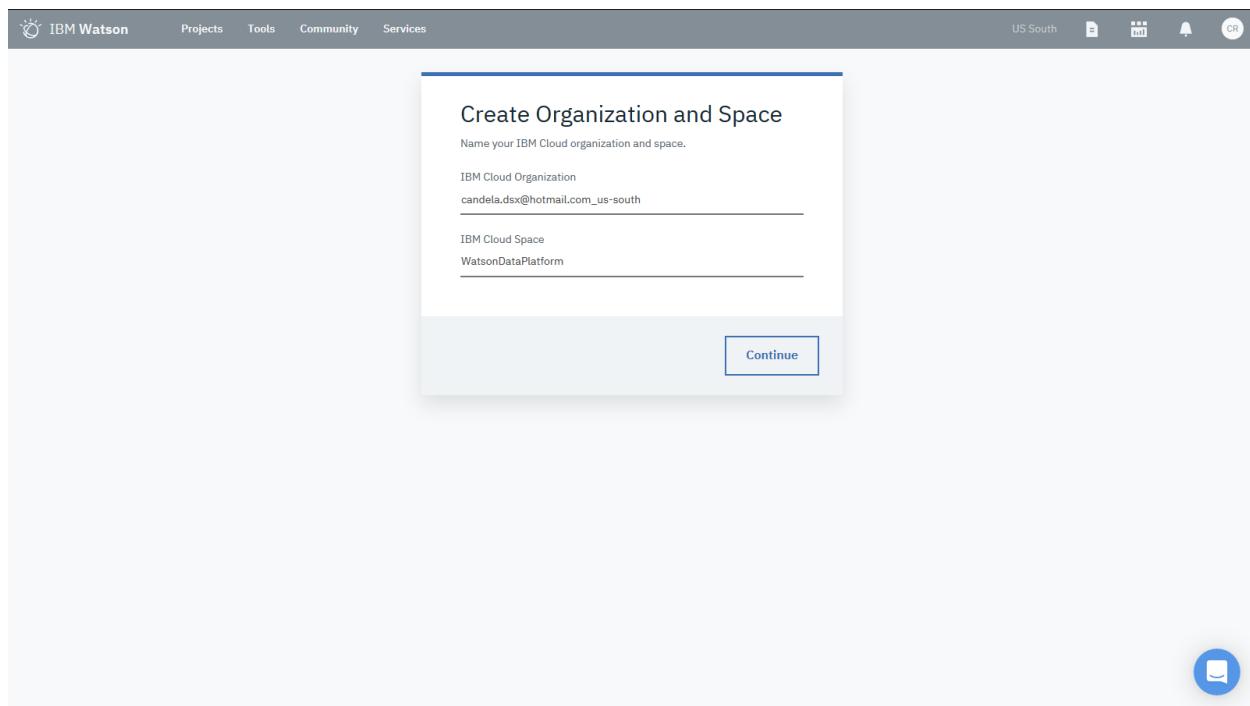
Deploy apps in containers
Create a Kubernetes cluster and run highly available containers. A Kubernetes cluster lets you quickly automate, update, and scale containerized apps.

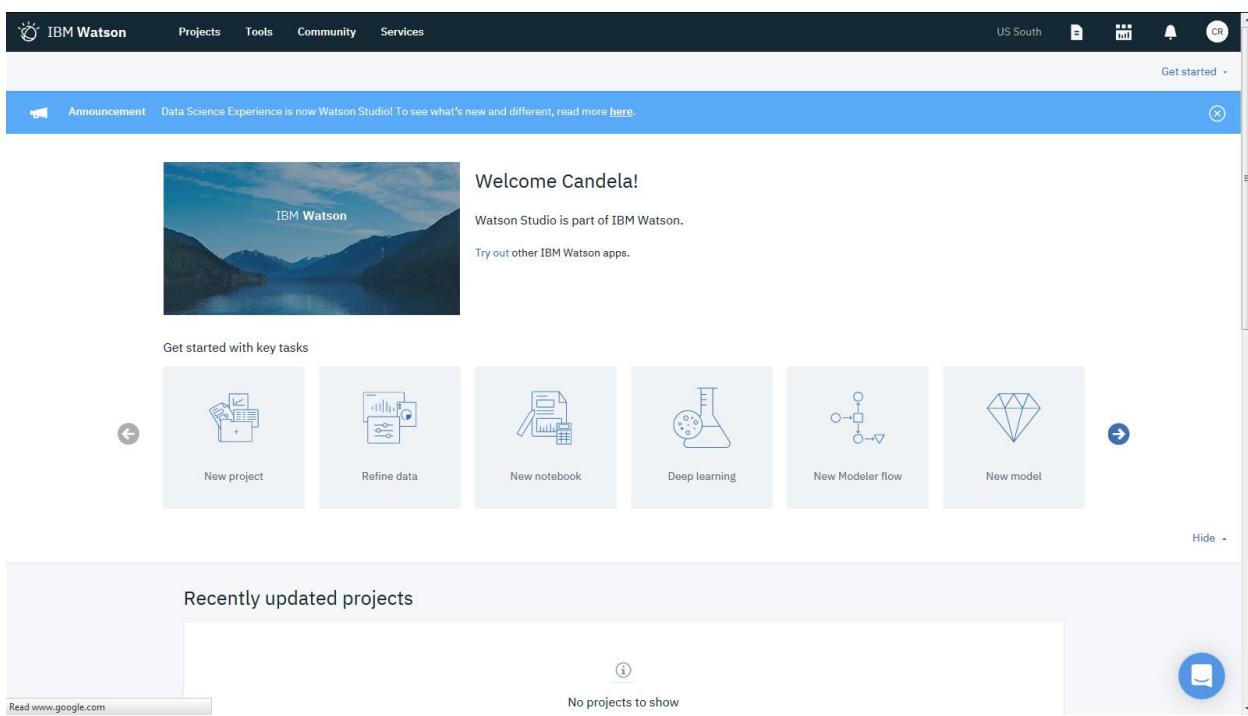
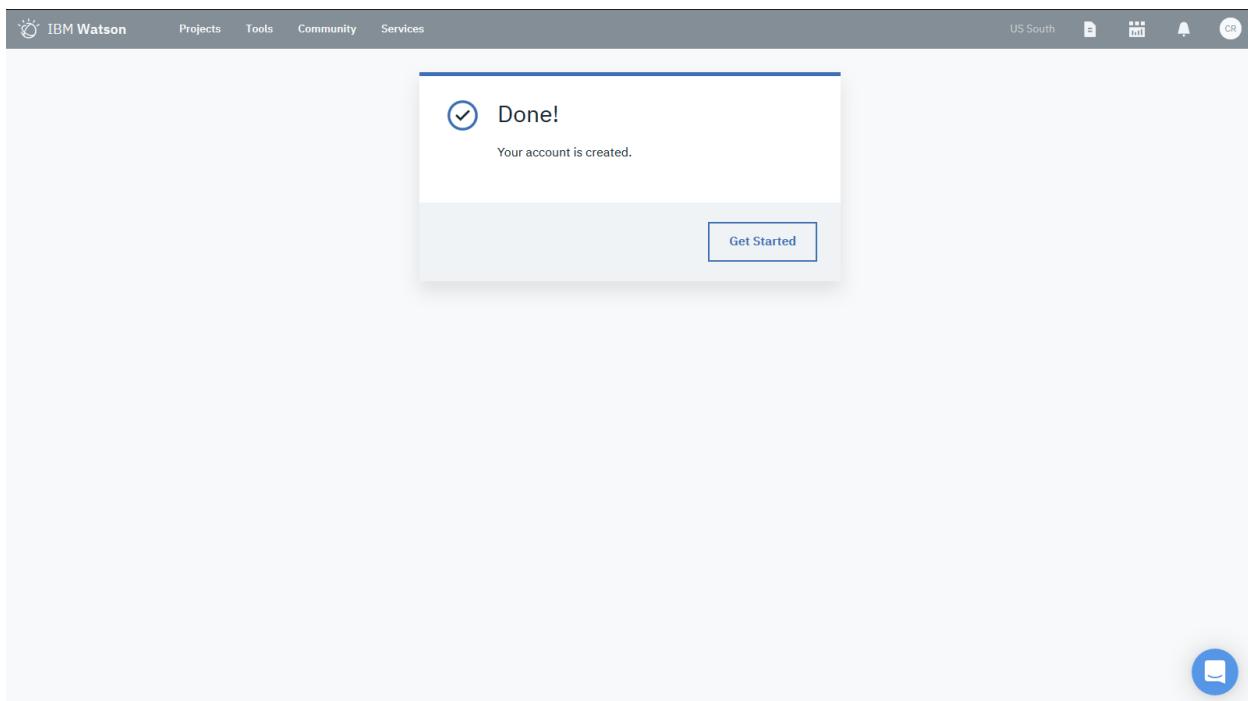
Order a monthly Bare Metal Server
Built to spec with 500GB/month outbound bandwidth included, ready in 2-4 hours.

The screenshot shows the IBM Cloud dashboard with an empty resource list. A large circular placeholder icon with a grid pattern is centered on the page. Below it, there are three service offerings:

- Create a Cloud Foundry app**: Go straight to developing with a Liberty for Java runtime, then add some of our 100+ services to build your app even faster.
- Deploy apps in containers**: Create a Kubernetes cluster and run highly available containers. A Kubernetes cluster lets you quickly automate, update, and scale containerized apps.
- Order a monthly Bare Metal Server**: Built to spec with 500GB/month outbound bandwidth included, ready in 2-4 hours.

Entramos de nuevo en: <https://datascience.ibm.com> y nos creamos una organización y un espacio de trabajo.



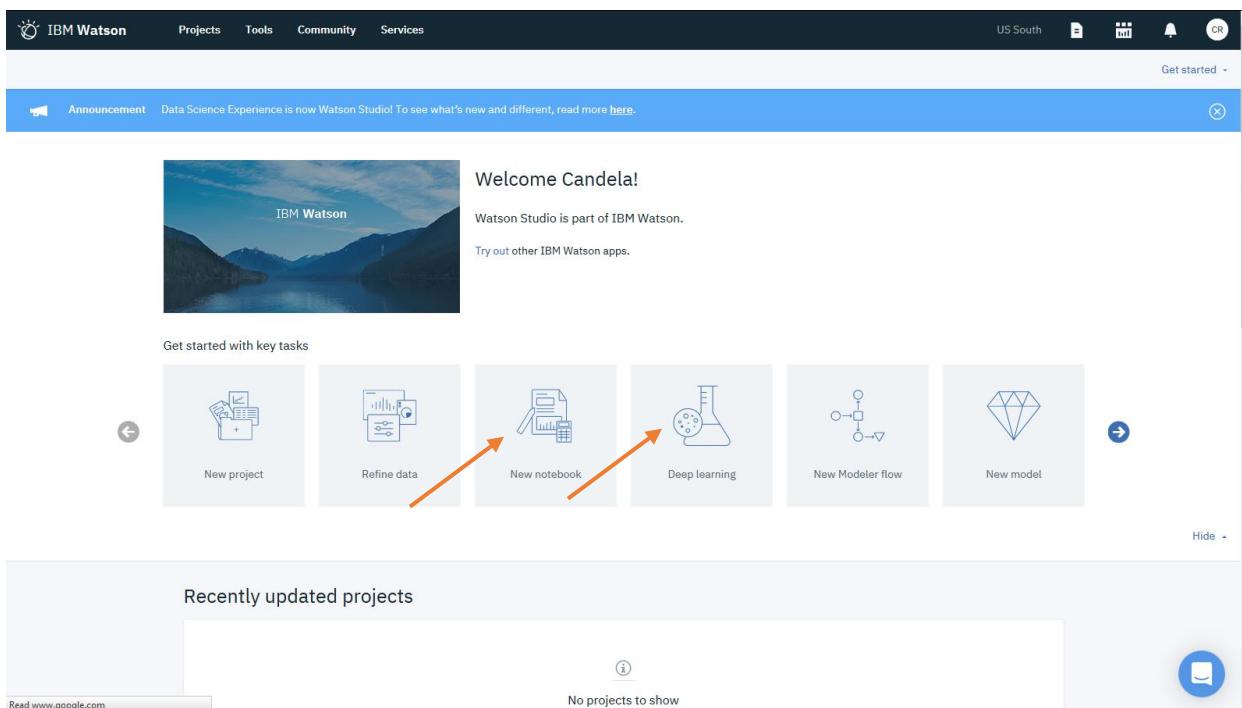


Ya nos hemos creado una cuenta.

Nota: asegúrate de permanecer con tu navegador predeterminado durante el proceso de inicio de sesión. Si te encuentras en IBM Cloud Dashboard, simplemente regresa a la página de registro y sigue el enlace que indica que ya tienes una cuenta.

Una vez que te registras, tu IBMID está vinculado a tu cuenta de IBM Cloud y a la cuenta de Watson. Usa tu IBMID (la dirección de correo electrónico que proporcionaste) para iniciar sesión en WATSON STUDIO.

Si eres el único usuario en la cuenta, ¡ya está todo listo! Como propietario de la cuenta IBM Cloud que se suscribió a una aplicación Watson Data Platform, tienes los permisos necesarios para agregar servicios, proyectos, catálogos, etc.



Desde IBM Watson, podemos hacer todo el ciclo de vida de un proyecto de machine y de Deep learning. Podemos preparar y refinar los datos, modelizarlos de diversas maneras y ponerlo en producción. Además podemos aprovechar todas las apis de Watson para enriquecer mis modelos, el lenguaje natural, reconocimiento de imágenes, etc.

2. Añadir usuarios adicionales (opcional)

Las aplicaciones de IBM Watson Studio, IBM Watson Knowledge e IBM Data Refinery están diseñadas para la colaboración entre muchos usuarios. Después de crear una cuenta, puedes agregar usuarios para que puedan compartir servicios y recursos que se aprovisionan para la cuenta.

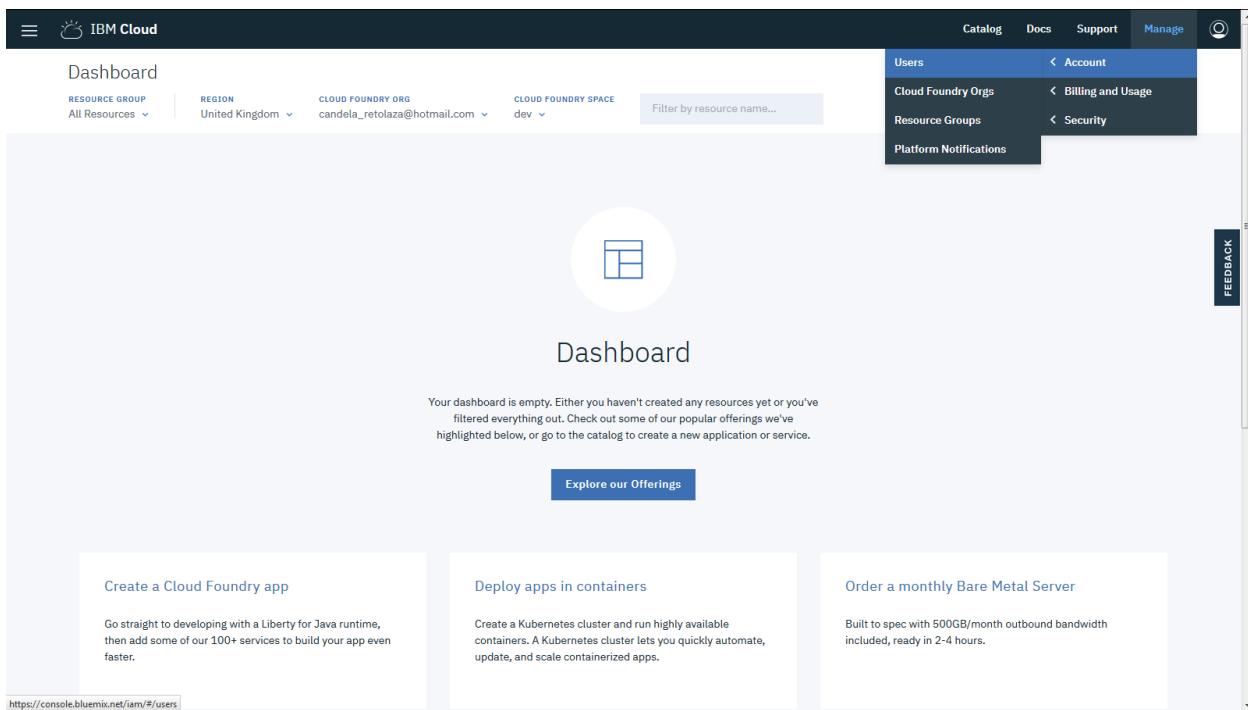
Lo vemos detalladamente:

Agrega usuarios y asigna roles de usuario

Los usuarios que invites a la cuenta pueden compartir los servicios y recursos en la cuenta. Por ejemplo, los usuarios pueden crear proyectos o catálogos utilizando una instancia existente de IBM Cloud Object Storage de la cuenta. Estos usuarios también pueden agregarse como colaboradores en catálogos y proyectos restringidos. Si el usuario invitado aún no tiene una cuenta IBM Cloud, el usuario recibirá un correo electrónico para completar el proceso de registro.

Entra en <https://console.bluemix.net/>

Haz clic en Manage> account> Users para agregar usuarios autorizados a tu organización.



The screenshot shows the IBM Cloud dashboard with the 'Manage' menu open. The 'Users' option is highlighted, indicating it is the active section. The main content area displays a message stating 'Your dashboard is empty. Either you haven't created any resources yet or you've filtered everything out.' Below this message is a blue button labeled 'Explore our Offerings'.

Dashboard

Your dashboard is empty. Either you haven't created any resources yet or you've filtered everything out. Check out some of our popular offerings we've highlighted below, or go to the catalog to create a new application or service.

[Explore our Offerings](#)

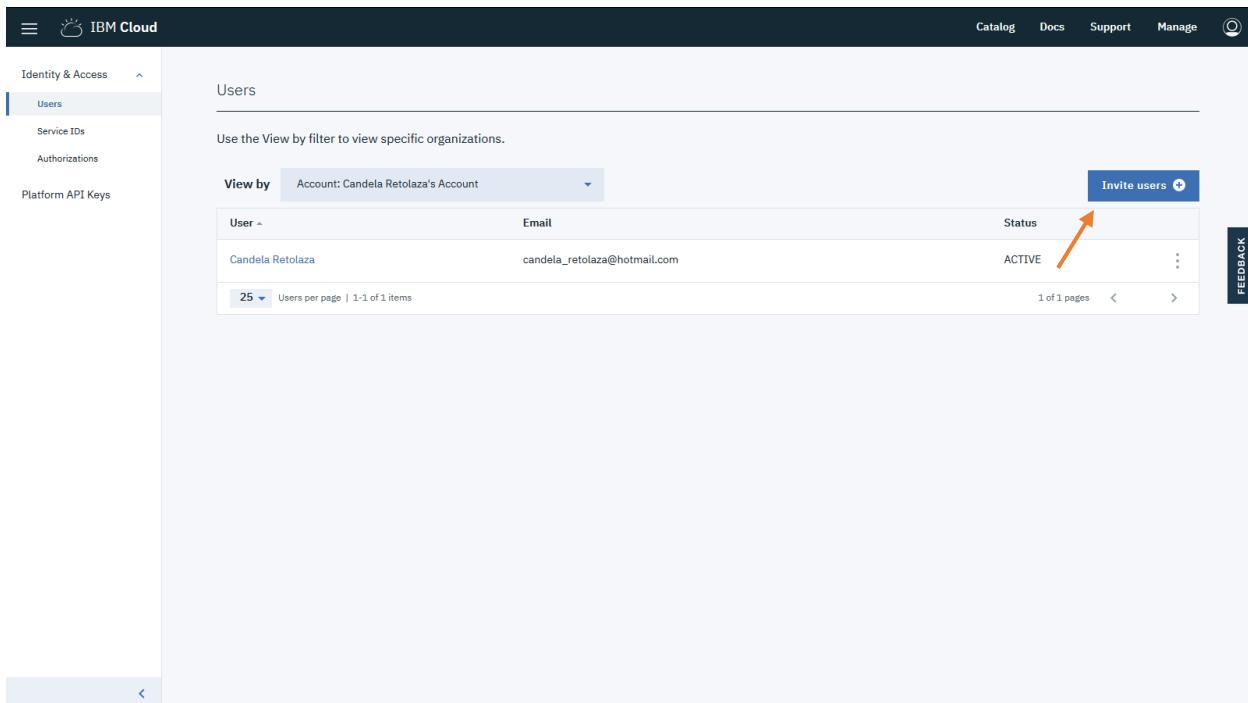
Create a Cloud Foundry app
Go straight to developing with a Liberty for Java runtime, then add some of our 100+ services to build your app even faster.

Deploy apps in containers
Create a Kubernetes cluster and run highly available containers. A Kubernetes cluster lets you quickly automate, update, and scale containerized apps.

Order a monthly Bare Metal Server
Built to spec with 500GB/month outbound bandwidth included, ready in 2-4 hours.

<https://console.bluemix.net/iam/#/users>

En la página Users, haz clic en Invitar usuarios.



The screenshot shows the 'Users' page within the 'Manage' section of the IBM Cloud interface. On the left sidebar, under 'Identity & Access', the 'Users' option is selected. In the main content area, there is a table listing one user: 'Candela Retolaza' with the email 'candela_retolaza@hotmail.com' and status 'ACTIVE'. At the top right of the table, there is a blue button labeled 'Invite users' with a small plus sign icon. An orange arrow points to this button, highlighting it.

User	Email	Status
Candela Retolaza	candela_retolaza@hotmail.com	ACTIVE

The screenshot shows the 'Identity & Access' section of the IBM Cloud interface. On the left, a sidebar lists 'Identity & Access' with sub-options: 'Users', 'Service IDs', 'Authorizations', and 'Platform API Keys'. The main content area is titled 'Invite users to Candela Retolaza's Account'. It has a 'Users' section with a note: 'You can invite one or more users. All users invited together are granted the same access.' Below it is an 'Email address' input field with placeholder text: 'Enter up to 100 email addresses separated by commas, spaces, or line breaks.' A 'Feedback' button is located in the top right corner of the main content area. At the bottom right are 'Cancel' and 'Invite users' buttons.

Selecciona un usuario existente de IBMid. Puedes agregar múltiples usuarios y la configuración posterior se aplicará a todos ellos.

This screenshot shows the same 'Identity & Access' interface as the previous one, but with two email addresses entered in the 'Email address' field: 'candela016@hotmail.com' and 'candela.retolaza@es.ibm.com'. The rest of the interface remains identical to the first screenshot, including the sidebar, main title, and bottom buttons.

The screenshot shows the 'Identity & Access' section of the IBM Watson Studio interface. It displays two main sections: 'Assign platform access roles' and 'Assign service access roles'. Under 'Assign platform access roles', there are four options: 'Administrator' (selected), 'Editor', 'Operator', and 'Viewer'. Each option has a brief description. Under 'Assign service access roles', there are three options: 'Manager' (selected), 'Writer', and 'Reader'. Each option also has a brief description. A 'Catalog', 'Docs', 'Support', and 'Manage' navigation bar is at the top right.

This screenshot shows the 'Cloud Foundry access' configuration screen. It includes fields for 'Organization' (set to 'candela_retolaza@hotmail.com_us-south'), 'Organization roles' (set to 'Manager'), 'Region' (set to 'All current regions'), and a note about specifying a region. A 'Feedback' button is visible on the right.

This screenshot shows a more detailed view of the 'Cloud Foundry access' configuration. It includes fields for 'Organization' (set to 'candela_retolaza@hotmail.com_us-south'), 'Organization roles' (set to 'Manager'), 'Region' (set to 'All current regions'), 'Space' (set to 'All current spaces'), and 'Space roles' (set to 'Developer'). A note indicates that the space role applies to selected spaces. At the bottom right are 'Cancel' and 'Invite users' buttons. A 'Feedback' button is also present.

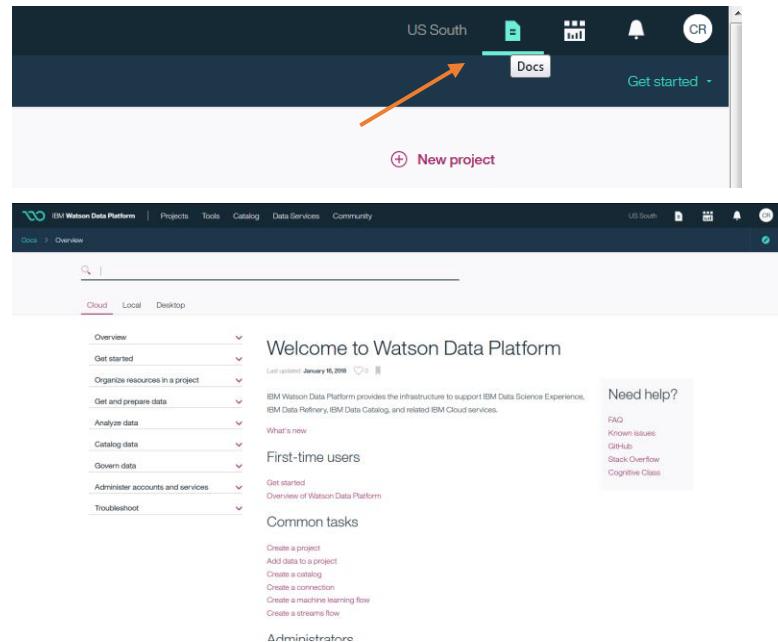
En la sección Acceso de la página Invitar usuarios, expanda el acceso a Cloud Foundry y seleccione la Organización a la que está agregando usuario.

- a. Asignar el rol del nuevo usuario en la organización.
- b. Secciona una Región y Espacio, o acepta los valores predeterminados.
- c. Para permitir que Watson Data Platform cree una instancia de Spark durante la creación del proyecto, asigne la función Desarrollador como la función de espacio.
- d. Para finalizar, haz clic en Invitar usuarios.

Para más información, visita la ayuda de WATSON STUDIO.

NOTA: Donde encontrar la ayuda y documentación de Watson Studio

Click en Docs.



Sus usuarios ahora pueden iniciar sesión y pueden cambiar su cuenta y organización en la Configuración del perfil. Los usuarios asociados a tu cuenta ahora pueden trabajar juntos y usar las aplicaciones y servicios disponibles de la cuenta.

IBM Cloud

Catalog Docs Support Manage

Identity & Access

- Users
- Service IDs
- Authorizations
- Platform API Keys

Users

Use the View by filter to view specific organizations.

View by Account: Candela Retolaza's Account

Invite users + FEEDBACK

User	Email	Status
Candela Retolaza	candela_retolaza@hotmail.com	ACTIVE
CANDELA RETOLAZA CONDE	candelaretolaza@es.ibm.com	ACTIVE
Candela Retolaza Conde	candela016@hotmail.com	ACTIVE

25 Users per page | 1-3 of 3 items 1 of 1 pages < >

3. Comenzar con WATSON STUDIO

Vuelve a IBM Watson Studio: Puedes abrir o cerrar el desplegable de inicio, para crear directamente proyectos, notebooks, modelos, etc.

IBM Watson

Projects Tools Community Services

US South

Get started +

Announcement Data Science Experience is now Watson Studio! To see what's new and different, read more [here](#).

Welcome Candela!

Watson Studio is part of IBM Watson.

Try out other IBM Watson apps.

Get started with key tasks

New project Refine data New notebook Deep learning New Modeler flow New model

Recently updated projects

No projects to show

Read www.google.com

The screenshot shows the IBM Data Science Experience interface. At the top, there's a navigation bar with links for Projects, Tools, Community, Services, and a location dropdown set to US South. On the right side of the header are icons for user profile, notifications, and help.

Recently updated projects: A section titled "Recently updated projects" displays a message: "No projects to show. You haven't made a project yet." with a "New Project" button.

New in the community: A section titled "New in the community" shows four cards for different types of content:

- ARTICLE:** Making Sense of the Bias / Variance... by ML Review, posted on Feb 09, 2018. Topic: Machine Learning.
- DATA SET:** Customer demographics and sales by IBM, posted on Nov 20, 2017. Topic: Economy & Business.
- NOTEBOOK:** Create and run a data flow with Watson Data... by IBM, posted on Feb 09, 2018. Topic: Science & Technology.
- TUTORIAL:** Working With IBM Cloud Object Storage In... by IBM Watson Data Platform, posted on Feb 02, 2018. Level: Beginner. Topic: DSX +1.

Helpful links: A section titled "Helpful links" with four categories:

- Docs:** Find the information you need.
- Discussion forum:** Stack Overflow is a community of
- Blog:** Read and follow our blog to keep up
- Got ideas?** Have feedback on Data Science

Para empezar a utilizar Watson Studio puedes seguir los siguientes pasos:

1. Configura un proyecto para organizar sus recursos.
2. Agrega colaboradores a tu proyecto.
3. Agrega datos a tu proyecto.
4. Opcional: agrega servicios analíticos como IBM Streaming Analytics o Watson Machine Learning.
5. Comience a analizar datos. Por ejemplo, puede crear notebooks, usar RStudio, crear flujos de SPSS o modelos de aprendizaje automático.
6. Despliega los modelos y entrénalos.

¿Necesitas inspiración? Haz clic en el botón Comunidad en tu Watson Studio para explorar los conjuntos de datos seleccionados, los Notebooks de ejemplo, los artículos y tutoriales, tanto para aprender de ellos como para utilizarlos como puntos de partida.

Workshop 2.

Crea y organiza los recursos en un proyecto

1. Crear un proyecto

Para crear un proyecto, haz click en Projects, y cree uno nuevo.

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'IBM Watson' and other options like 'Projects', 'Tools', 'Community', and 'Services'. On the right side, there are icons for 'US South', a file, a gear, a bell, and a 'CR' button. Below the navigation bar, a welcome message says 'Welcome Candela!' and 'Watson Studio is part of IBM Watson.' A link 'Try out other IBM Watson apps.' is also present. The main area has a section titled 'Get started with key tasks' with several tiles: 'New project' (selected), 'Refine data', 'New notebook', 'Deep learning', 'New Modeler flow', and 'New model'. An orange arrow points from the 'New project' tile to the 'Recently updated projects' section below. This section shows a message 'No projects to show' with a note 'You haven't made a project yet.' and a 'New Project' button. Another orange arrow points to this button. A large orange arrow points from the 'New Project' button down to the 'New project' dialog box. The dialog box is titled 'New project' and contains a sub-headline 'Select a project tile to get the right tools and services for your work.' It lists several project types: 'Basic', 'Jupyter Notebooks', 'Visual Recognition', 'Experiment Assistant', 'Modeler', 'Dashboards', 'Streams Designer', 'Data Preparation', and 'Complete'. The 'Complete' project type is highlighted with a checked checkbox. At the bottom of the dialog box are 'Cancel' and 'OK' buttons, with an orange arrow pointing to the 'OK' button.

Creamos un proyecto nuevo. Nos propone diferentes tipos de proyectos, dependiendo de lo que queramos hacer, pues existen proyectos con herramientas predefinidas, aunque posteriormente podríamos ampliarlo sin problema. En este caso, vamos a elegir uno completo en el que vamos a trabajar durante todo el Workshop.

Seleccionamos Complete y damos a OK.

por ejemplo: Workshop, añadimos una descripción opcional: por ejemplo, Workshop WATSON STUDIO.

New project

Define project details

Name
Project name

Description
Project description

Choose project options

Restrict who can be a collaborator (i)

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

Define storage

- ① Select storage service **Add**
- ② Refresh

Add an object storage instance and then return to this page and click Refresh.

Create

Añadimos un almacenamiento y un servicio de spark:

Primero añadimos un Object Storage: Creamos una versión lite gratuita:

Cloud Object Storage

Existing New

Cloud Object Storage

IBM Cloud Object Storage is a highly scalable cloud storage service, designed for high durability, resiliency and security. Store, manage and access your data via our self-service portal and RESTful APIs. Connect applications directly to Cloud Object Storage or use other IBM Cloud Services with your data.

Features

Storage for the IBM Cloud
IBM Cloud Object Storage provides unstructured data storage for cloud applications. Libraries and SDKs support a common set of S3 API functions for connecting new applications to scalable cloud storage and integrating your data into other services on the IBM Watson and Cloud Platform.

IAM Policies - Bucket level access management
IBM Identity and Access Management (IAM) integration allows for granular access control at the bucket level using role-based policies.

Encryption management
All data is encrypted at rest and in flight by default. Keys are automatically managed by default, but can optionally be self-managed or managed using IBM Key Protect. (*Key Protect is only available for buckets created in the US South (Dallas) and EU GB (London) regions.)

Regional and Cross Region resiliency options
Select the best resiliency option for your data. Choose "Cross Region" to store unstructured data across three regions, or choose "Regional" resiliency to store your data within a single region.

Data storage classes for Active, Less Active, Archive and Dynamic workloads
Choose storage classes for frequently accessed data, occasionally accessed data and long term data retention with Standard, Vault, and Cold Vault. Or, choose File class for dynamic data access needs that fluctuate month to month.

Lite and pay-as-you-go plans

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

Plan	Features	Pricing
<input checked="" type="radio"/> Lite	1 COS Service Instance Storage up to 25 GB/mo. Up to 20,000 GET requests/mo. Up to 2,000 PUT requests/mo. <small>Up to 100 Data Retrieval 10 GB/mo.</small> <small>Up to 5GB Public Outbound</small> <small>Applies to aggregate total across all storage bucket classes</small>	Free

IBM Watson Projects Tools Community Services US South CR

Integrating your data into other services on the IBM Watson and Cloud Platform.

PLAN **FEATURES** **PRICING**

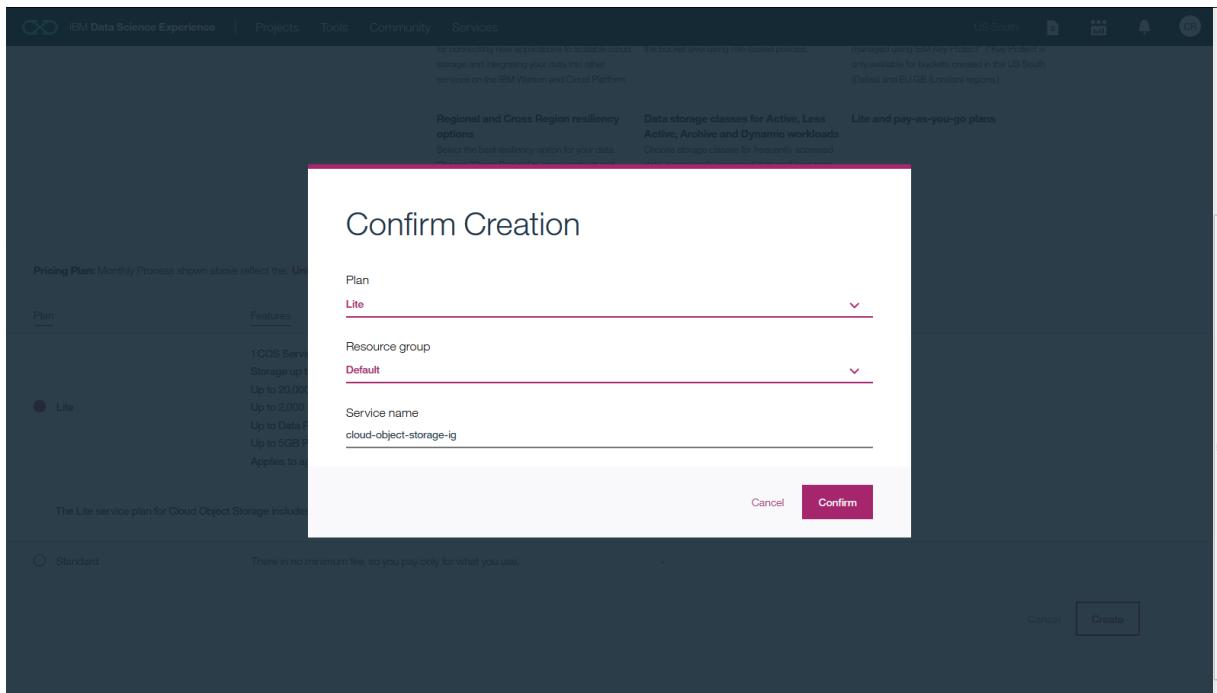
Lite

1 COS Service Instance
Storage up to 25 GB/mo.
Up to 20,000 GET requests/mo.
Up to 2,000 PUT requests/mo.
Up to Data Retrieval 10 GB/mo.
Up to 5GB Public Outbound
Applies to aggregate total across all storage bucket classes

The Lite service plan for Cloud Object Storage includes Regional and Cross Regional resiliency, flexible data classes, and built in security.

Standard **There is no minimum fee, so you pay only for what you use.**

Cancel **Create** 



Confirmamos, y actualizamos pinchando en Refresh:

Ya hemos creado el almacenamiento. Ahora ya podemos crear el proyecto.

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with links for 'Projects', 'Tools', 'Community', and 'Services'. On the right side of the top bar, there are icons for 'US South', 'Add to project', and user profile information. Below the top bar, the main header says 'My Projects / Workshop'. Underneath this, there's a sub-header 'Overview' followed by tabs for 'Assets', 'Environments', 'Bookmarks', 'Deployments', 'Collaborators', and 'Settings'. The main content area is titled 'Workshop' and shows the date 'Last Updated: Mar 22 2018'. It displays three summary statistics: 'Assets' (0), 'Bookmarks' (0), and 'Collaborators' (1). On the left, there are sections for 'Date created' (Mar 22 2018), 'Description' (No description available), 'Storage' (0% of 5 GB used), 'Collaborators' (Candelia R. Conde, Admin), and 'Bookmarks' (View all (0)). On the right, there's a 'Recent activity' section with a message from Armand from IBM: 'Hi Cande, Congratulations on getting your project started! The next step is to bring in some data to start analyzing. All your project collaborators are automatically authorized to access the data in the project.' Below this message, there are options for adding assets from 'Local files', 'Database connections', and 'Kafka topics'.

Ahora, ya tenemos un proyecto nuevo. Lo primero que nos muestra en el resumen es que no tenemos nada en el proyecto y que solo tiene un colaborador.

Un proyecto sirve para organizar tus recursos para trabajar y hacer machine learning, Deep learning, desplegar modelos, refinrar los datos, etc. Los recursos de su proyecto pueden incluir: Notebooks, Modelos, Visual recognition models, Flows de Modeler, Streams flows, Experimentos, Dashboard, deployments, archivos de assets de datos y conexiones, colaboradores, marcadores a los recursos de la comunidad, tokens de acceso, enlaces a repositorios de GitHub para publicar notebooks, servicios de spark u otros motores, y otros servicios asociados, como Watson Machine Learning o IBM Streaming o Servicios de Watson.

Si tienes permisos de administrador en un proyecto, tienes control total sobre él. Si tiene permisos de Editor, puedes agregar activos y colaboradores a un proyecto. La página Overview proporciona un resumen del estado actual del proyecto, incluida información sobre el uso del almacenamiento, la actividad reciente, los colaboradores, los marcadores y los activos.

2. Añade colaboradores

Lo primero que vamos a hacer en nuestro proyecto nuevo es añadir un nuevo colaborador al proyecto. Pincho en colaboradores y en añadir uno nuevo.

IBM Data Science Experience | Projects Tools Community Services

My Projects / Workshop

Overview Assets Bookmarks Deployments **Collaborators** Settings

Workshop
Last Updated: Feb 12 2018

Assets 0 **Bookmarks** 0 **Collaborators** 1

Date created
Feb 12 2018

Description
No description available

Storage
0% of 5 GB used

Collaborators View all (1)
CR Candela Retolaza Admin

Bookmarks View all (0)
You currently have 0 bookmarks

Recent activity

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

IBM Watson Data Platform | Projects Tools Catalog Data Services Community

My Projects / Workshop

Overview Assets Bookmarks Deployments **Collaborators** Settings

Find collaborators

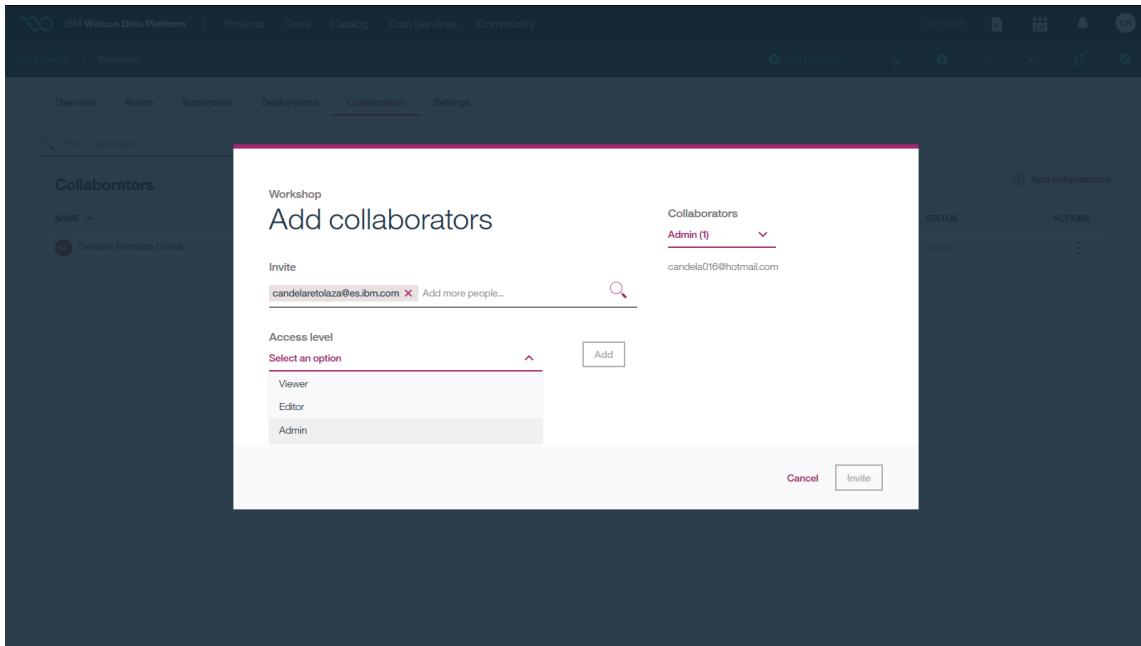
Collaborators

NAME	EMAIL	PERMISSION	STATUS	ACTIONS
CR Candela Retolaza Conde	candela016@hotmail.com	Admin	Active	⋮

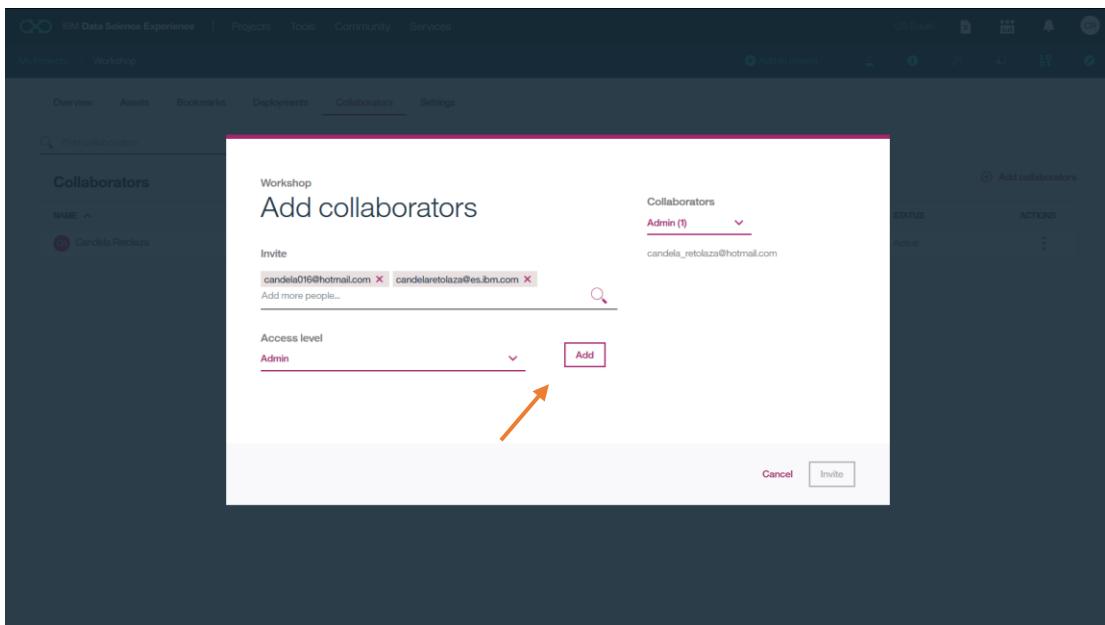
+ Add collaborators

Para poder hacer esto, solo necesito que la persona a la que quiero añadir tenga una cuenta en mi IBM Cloud. *Puedes crear cuentas asociadas a tu IBM Cloud tal y como hemos contado en el Workshop 1.2.*

Añadimos el correo, seleccionamos el tipo de acceso que queremos dar a esa persona y enviamos la invitación. Y ya estará en nuestro proyecto.



Añado el correo, doy a enter, y selecciono el tipo de acceso, pincho en Add:



Y damos a invitar:

Collaborators

Workshop Add collaborators

Invite

Search for name or enter email address

Access level

Admin

Add

Collaborators

Admin (2)

candela016@hotmail.com

candela016@es.ibm.com

STATUS

Active

ACTIONS

Cancel

Invite

Una vez añadido:

NAME	EMAIL	PERMISSION	STATUS	ACTIONS
Candela Retolaza	candela_retolaza@hotmail.com	Admin	Active	⋮
Candela Retolaza Conde	candela016@hotmail.com	Admin	Active	⋮
Candela Retolaza Conde	candela016@es.ibm.com	Editor	Active	⋮

Desde aquí puedo cambiar el tipo de permiso de cada colaborador.

3. Aprende a gestionarlo: borrar assets, añadir, crear conexiones...

Assets:

Si tiene permisos de administrador o editor en un proyecto, puedes agregar recursos.

Los tipos de activos enumerados están condicionados a las aplicaciones de Watson Data Platform que tiene. Para agregar assets a un proyecto, elija el tipo de asset en el menú Agregar al proyecto:

Conexiones, datos de fichero plano, datos de bases de datos, Notebooks, Modelos de reconocimiento visual, flujos de aprendizaje automático, dashboards, modelos, modelos de SPSS, experimentos, etc.

Si tienes permisos de administrador en un proyecto, puedes eliminar activos. Para eliminar un activo, elija Eliminar en el menú ACCIONES al lado del nombre del activo.

Gestión de proyectos:

Podremos gestionar el almacenamiento, los servicios asociados, los tokens, ver a quién pertenece la cuenta y conectar el proyecto en github. Se propone al lector que explore por la herramienta antes de comenzar el siguiente Workshop.

The screenshot shows the 'Project information' section of the 'Workshop' project settings. It includes fields for 'Project name' (Workshop) and 'Description' (Project description). A progress bar indicates 0% of 5 GB used. Buttons for 'Cancel' and 'Save' are at the bottom.

The screenshot shows the 'Storage' section with Cloud Object Storage (Beta) and Bucket Name (workshop984b65fe9c4b427392cac90bb0c561aa) details. Below it is an 'Associated services' table with one entry: Spark-ae (Spark, Personal plan). The 'Access tokens' section shows no tokens. The 'Connect to a GitHub repository' section has a 'Repository URL' field with the placeholder <https://github.com/owner/repository-name> and a 'Connect' button.

The screenshot shows the IBM Watson Data Platform interface. At the top, there's a navigation bar with links for Projects, Tools, Catalog, Data Services, and Community. On the right side of the header, there are icons for US South, a file, a database, a bell, and a refresh button. Below the header, the main content area has a dark header bar with "My Projects > Workshop". The main body contains several sections:

- Access tokens**: A table with columns NAME, ROLE, CREATED, LAST USED, and ACTIONS. It displays one entry: "Spark-ae" with "Spark" under ROLE, "Personal" under CREATED, and "you currently have no access tokens" under LAST USED. There's a "New token" button at the top right.
- Connect to a GitHub repository**: A section with a "Repository URL" input field containing "https://github.com/owner/repository-name" and a "Connect" button.
- Project scope**: A section with two rows: "IBM Cloud account" (Candela Retolaza Conde's Account) and "Restrict who can be a collaborator" (Yes).

Workshop 3.

Consigue y prepara los datos y analiza los datos de manera sencilla

1. Añadir datos

Después de crear un proyecto, tenemos que añadir datos para poder trabajar con los mismos. Todos los colaboradores en el proyecto están autorizados automáticamente para acceder a los datos en el proyecto.

Puede añadir assets de datos de estas fuentes a un proyecto:

- Archivos locales
- Comunidad
- Conexiones de base

Vamos a comenzar añadiendo un fichero local a nuestro proyecto. Para agregar archivos de datos a un proyecto:

Desde la página Assets de su proyecto, puedes añadir datos de dos maneras, en el icono  arriba a la derecha, o clickando en New data asset.

My Projects > Workshop

Overview Assets Bookmarks Deployments Collaborators Settings

What assets are you looking for?

Data assets

0 assets selected.

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
you currently have no data assets					

Notebooks

NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
you currently have no notebooks							

Streams flows

NAME	MODIFIED BY	LAST MODIFIED	ACTIONS
you currently have no streams flows			

Models

NAME	STATUS	RUNTIME	LAST MODIFIED	ACTIONS
you currently have no models				

My Projects > Workshop

Overview Assets Bookmarks Deployments Collaborators Settings

What assets are you looking for?

Data assets

0 assets selected.

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
you currently have no data assets					

Notebooks

NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
you currently have no notebooks							

Streams flows

NAME	MODIFIED BY	LAST MODIFIED	ACTIONS
you currently have no streams flows			

Models

NAME	STATUS	RUNTIME	LAST MODIFIED	ACTIONS
you currently have no models				

Haga clic en Load y luego busque los archivos en el PC o arrástrelos.

Subimos el fichero datos_banca.csv. Debes permanecer en la página hasta que la carga esté completa. Puede cancelar un proceso de carga en curso si desea dejar de cargar un archivo.

The screenshot shows the IBM Watson Data Platform interface. In the top navigation bar, there are links for Projects, Tools, Catalog, Data Services, and Community. On the right side, there are icons for US South, Add to project, and notifications. The main content area is titled 'My Projects > Workshop'. Under the 'Assets' tab, there is a search bar with the placeholder 'What assets are you looking for?'. Below it, there's a section for 'Data assets' with a table showing one entry: 'datos_banca.csv' (Data Asset, Project, Created by Candelita Retolaza Conde, Last modified 24 Jan 2018). There are also sections for 'Notebooks', 'Streams flows', and 'Models', each with their own tables. On the right side, there is a sidebar with tabs for Load, Files, and Catalog. The 'Files' tab is selected, showing a list with '0 selected' and an item 'datos_banca.csv'. A red arrow points from the text 'Haciendo click en el fichero, podemos ver cómo son nuestros datos.' to the 'datos_banca.csv' entry in the Assets list.

Los archivos se guardan en el object storage que está asociado con su proyecto y se enumeran como assets de datos en la página de Assets de su proyecto.

Haciendo click en el fichero, podemos ver cómo son nuestros datos.

The screenshot shows the preview of the 'datos_banca.csv' file. At the top, there are links for Preview and Profile. Below that, it says 'Schema: 12 Columns' and 'Preview (100 rows)'. The preview table has 12 columns: AGE, ED, EMPLOY, ADDRESS, INCOME, DEBTINC, CREDDEBT, OTHDEBT, DEFAULT, PREDDEF1, PREC, and a Type column. The first few rows of data are listed. To the right of the preview, there is a sidebar with the title 'datos_banca.csv'. It includes sections for 'Description' (No description available for this asset), 'Tags' (No tags available for this asset), 'Creator' (candelita.dsx@hotmail.com), 'Added' (11:43 AM UTC, 2018/03/22), and 'Size' (94.34 KB). A blue arrow points from the text 'Podremos refinar los datos haciendo click en Refine.' to the 'Refine' button in the top right corner of the preview table.

Podremos refinar los datos haciendo click en Refine.

The screenshot shows the IBM Watson Studio interface with the 'Preview' tab selected for the 'datos_banca.csv' file. The data preview displays 100 rows of information across 12 columns. The columns are labeled: AGE, ED, EMPLOY, ADDRESS, INCOME, DEBTINC, CREDDEBT, OTHDEBT, DEFAULT, PREDDEF1, PREDDEF2, and PREDDEF3. Each row contains numerical values corresponding to these fields. The top navigation bar includes 'Projects', 'Tools', 'Community', 'Services', 'US South', and various icons for file operations.

El fichero que acabamos de subir a nuestro proyecto contiene información de demográfica de clientes de un banco ficticio, con datos como: edad, nivel educativo, años en el trabajo actual, años en la misma vivienda, salario... en el que, además, tenemos información sobre los créditos que tiene cada cliente y un histórico de datos de clientes sobre si han hecho impago.

Por tanto, vamos a suponer que un banco está preocupado por el posible impago de sus créditos. Vamos a utilizar datos de créditos anteriores para predecir los clientes potenciales que tendrán problemas para pagar sus créditos, a estos clientes de alto riesgo se les puede negar un crédito u ofrecer otros productos.

2. Crear un modelo automático

Vamos a crear un modelo, en este caso, creamos uno automático o semi-automático, con el fichero datos_banca.csv que acabamos de subir y entender, y queremos intentar predecir qué variables producen impago. Clicamos en New Model.

IBM Watson Data Platform | Projects Tools Catalog Data Services Community US South

Data assets
0 assets selected.

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
datos_banca.csv	Data Asset	Project	Candela Retolaza Conde	24 Jan 2018	⋮

Notebooks

NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
you currently have no notebooks							

Streams flows

NAME	MODIFIED BY	LAST MODIFIED	ACTIONS
you currently have no streams flows			

Models

NAME	STATUS	RUNTIME	LAST MODIFIED	ACTIONS
you currently have no models				

SPSS Modeler flows

NAME	CREATED BY	LAST MODIFIED	ACTIONS
modeler flows			

Definimos el nombre del modelo, y debemos de tener un servicio de machine learning asociado a nuestra cuenta. Nos creamos para comenzar una versión gratuita.

IBM Watson Projects Tools Community Services US South

New model BETA

Define model details

Name
Model name

Description

Machine Learning Service
No Machine Learning service instances associated with your project.
Associate a Machine Learning service instance with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

Reload

Select model type

Model builder From file From sample

Spark Service
No Spark instances associated with your project.
Associate an IBM Analytics for Apache Spark instance with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

Reload

Automatic
Prepare my data and create a model automatically

Manual
Let me prepare my data and select which models to train

Need something more flexible? Create a [notebook](#) or design an [SPSS Modeler flow](#).

Create

Machine Learning

Existing New

Machine Learning

IBM Watson Machine Learning is a full-service Bluemix 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	SPSS analytics platform features	Spark and Python Machine Learning features	Integration with Data Science Experience
SPSS streams management and deployment with realtime scoring and batch processing options.	Take advantage of Spark MLlib and scikit-learn machine learning models management and deployment - online, batch and streaming.	Visit http://datascience.ibm.com . Create and train predictive analytics models with the best tools and the latest expertise in a social environment built by data scientists.	

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

Plan	Features	Pricing
<input checked="" type="radio"/> Lite	Service instance (5 models per instance) 5,000 predictions 5 compute hours	Free
<input type="radio"/> Standard	Predictions Compute hours	\$0.5 USD/1,000 predictions \$0.45 USD/hour
	Service instance	\$1,000 USD/month

The lite plan instance of the IBM Watson Machine Learning service provides you with a maximum of 5 deployed models, 5,000 predictions per month, and 5 hours per month of compute time during which model can be trained, evaluated, and deployed to be available to accept prediction events.

IBM Watson Projects Tools Community Services US South

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

user outcomes.

deployment (online, batch, streaming). Select any of widely supported machine learning frameworks: Tensorflow, Keras, Caffe, Pytorch, Spark MLlib, scikit learn, xgboost and SPSS.

application with artificial intelligence through the Watson Machine Learning REST API.

social environment built by and for data scientists.

Confirm Creation

Organization: candela.dsx@hotmail.com_us-south

Plan: Lite

Space: WatsonDataPlatform

Service name: predictive-modeling-zr

Cancel Confirm

Create Cancel

2

New model BETA

Define model details

Name
Model name

Description
Model description

Machine Learning Service
No Machine Learning service instances associated with your project.
Associate a Machine Learning service instance with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

Reload

Select model type

Model builder From file From sample

Spark Service
No Spark instances associated with your project.
Associate an IBM Analytics for Apache Spark instance with your project on the project settings page, then click the reload button below to refresh the instances available for association with our new model builder instance.

Reload

Automatic
Prepare my data and create a model automatically

Manual
Let me prepare my data and select which models to train

Need something more flexible? Create a [notebook](#) or design an [SPSS Modeler flow](#).

Cancel Create

Si hacemos Reload, tendremos nuestro servicio de Machine Learning listo para usar. Falta seleccionar un servicio de Spark. De la misma manera creamos uno lite, con dos ejecutores de Spark.

Apache Spark

Existing New

Apache Spark

Apache Spark is an open source cluster computing framework optimized for extremely fast and large scale data processing, which you can access via the newly integrated notebook interface IBM Analytics for Apache Spark. You can connect to your existing data sources or take advantage of the on-demand big data optimization of Object Storage. Spark plans are based on the maximum number of executors available to process your analytic jobs. Executors exist only as long as they're needed for processing, so you're charged only for processing done.

Features

- Incredibly Fast**
Apache Spark delivers 100x the performance of Apache Hadoop for certain workloads because of its advanced in-memory computing engine.
- Easy to Use and Powerful**
Apache Spark's Streaming and SQL programming models backed by MLlib and GraphX make it incredibly easy for developers and data scientists to build apps that exploit machine learning and graph analytics. Because the service is 100% compatible with Apache Spark, developers can build their apps and run them against the IBM managed service to benefit from operational, maintenance, and hardware excellence.
- Convenient Data Storage**
Object Storage enables a convenient way to upload your data from a file for immediate use by your Spark instance. You can set up Object Storage directly from the Spark service interface.

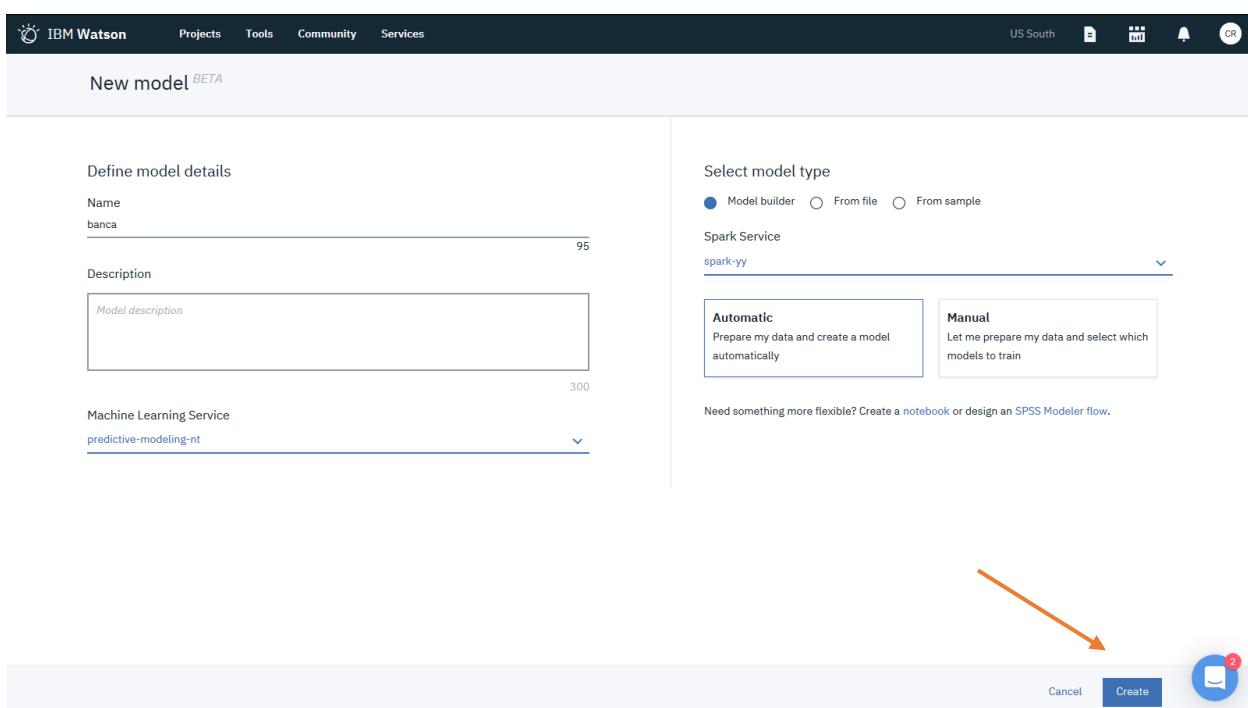
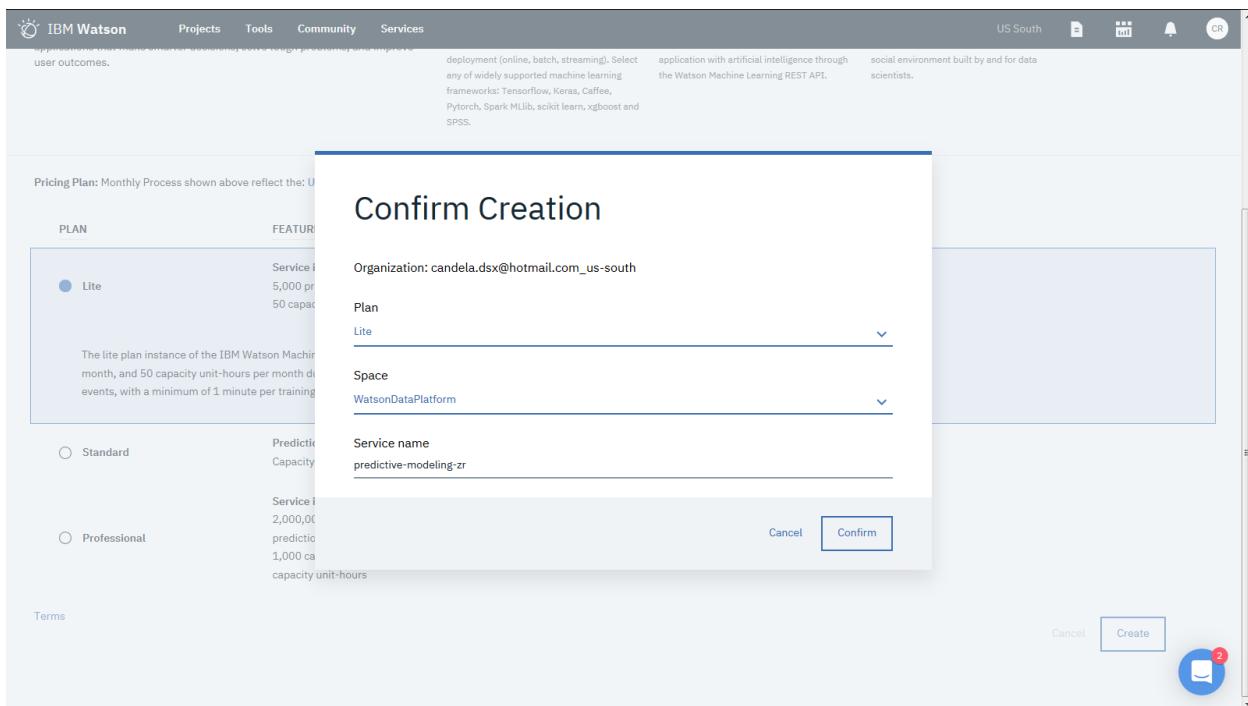
Pricing Plan: Monthly Process shown above reflect the: [United States](#)

PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	2 Spark Executors	Free

An entry level plan to run programs using up to 2 Spark executors

Terms

Create



Crearemos un modelo automático. Seleccionamos el fichero sobre el que queremos trabajar, en este caso solo tenemos uno:

The model builder currently supports CSV files and IBM Db2 Warehouse on Cloud data assets.

NAME	TYPE	SERVICE
datos_banca.csv	Data Asset	Project

Click to preview data

Add Data Assets

Close **Next**

Se tienen que inicializar los kernels y cargar los datos. Esperamos

The model builder currently supports CSV files and IBM Db2 Warehouse on Cloud data assets.

Kernel initializing

Your operation will begin immediately after kernel initialization completes.

Close **Next**

The model builder currently supports CSV files and IBM Db2 Warehouse on Cloud data assets.

What asset are you looking for?

NAME	TYPE	SERVICE
●	Project	

Status: loading data set...

Close Next

Para hacer un primer modelo, vamos a utilizar la variable de impago que se llama Default para analizar qué factores hacen que un cliente sea más propenso a hacer impago que otro. Además, le decimos que utilice todas las demás variables para predecir.

The screenshot shows the 'Select a technique' step in the IBM Watson Studio interface. The top navigation bar includes 'IBM Watson Data Platform', 'Projects', 'Tools', 'Catalog', 'Data Services', 'Community', 'US South', and various icons. The left sidebar shows 'My Projects > Workshop > Modelo Predictivo Impago'. The main area has a title 'Select a technique' and a sub-section 'Select Data'. Under 'Train', 'Column value to predict (Label Col)' is set to 'default (Decimal)'. Under 'Evaluate', 'Feature columns' is set to 'All (default)'. Below these are three options: 'Binary Classification' (selected), 'Multiclass Classification', and 'Regression'. A 'Validation Split' slider shows 'Train: 60', 'Test: 20', and 'Holdout: 20'. At the bottom right are 'Close', 'Previous', and 'Next' buttons.

Una vez hemos elegido que tipo de modelo de modelo queremos y seleccionadas quiénes son mis variables, vemos que también nos selecciona una parte de los datos para entrenamiento, y otra parte para testear. Ejecutamos.

The screenshot shows the 'Select model' step in the IBM Watson Studio interface. The top navigation bar and sidebar are identical to the previous screenshot. The main area has a title 'Select model' and a sub-section 'Select Data'. Under 'Evaluate', there is a table showing a single entry: 'ESTIMATOR TYPE' (LogisticRegression), 'STATUS' (Trained & Evaluated), 'PERFORMANCE' (Good), 'AREA UNDER ROC CURVE' (0.84647), 'AREA UNDER PR CURVE' (0.72029), 'LAST EVALUATION' (24 Jan 2018, 11:53 AM), and an 'ACTIONS' column with a three-dot menu icon. At the bottom right are 'Close', 'Previous', and 'Save' buttons.

Nos dice que tipo de estimador a utilizado (una regresión logística) y si es una buena predicción o no, y el área bajo la curva ROC y el área bajo la curva PR. Podemos guardar el modelo, o volver atrás y repetir con otro modelo.

The screenshot shows the IBM Watson Data Platform interface. At the top, there's a navigation bar with links for 'My Projects', 'Workshops' (selected), and 'Modelo Predictivo Impago'. Below the navigation is a 'Select model' section with two tabs: 'Train' and 'Evaluate'. Under 'Evaluate', a table lists a single model entry:

	ESTIMATOR TYPE	STATUS	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST EVALUATION	ACTIONS
●	LogisticRegression	Trained & Evaluated	Good	0.6847	0.7059	24 Jan 2018, 11:53 AM	⋮

In the center of the screen, a modal dialog box is open with the title 'Saving model'. It contains the message 'Status: saving model Modelo Predictivo Impago...' and two buttons: 'Close' and 'Save'.

3. Despliega el modelo en Watson Machine Learning

Cuando guardamos el modelo, nos muestra un resumen del modelo, nos deja evaluarlo y desplegarlo.

Modelo Predictivo Impago

Overview Evaluation Deployments

Summary

Machine learning service	predictive-modeling-wq
Runtime environment	spark-2.0
Training date	24 Jan 2018, 11:55 AM
Label column	default
Latest version	b13c03db-7754-41f9-88ed-7ee09ebdfb33
Model builder details	View

Input Schema

COLUMN	TYPE
age	decimal(31,6)
ed	decimal(31,6)
employ	decimal(31,6)
address	decimal(31,6)

Modelo Predictivo Impago

Overview **Evaluation** Deployments

Last Evaluation Result

Version	b13c03db-7754-41f9-88ed-7ee09ebdfb33
Phase	setup
AreaUnderPR	0.72
AreaUnderROC	0.846

Performance Monitoring

Configure performance monitoring to evaluate and retrain the model periodically to ensure the model performance is acceptable. You will need an existing IBM Db2 Warehouse on Cloud connection associated with your project to be used as your feedback data connection.

[Configure Performance Monitoring](#)

Versions

TIME	VERSION	DEPLOYED	ACTIONS
24 Jan 2018 11:58am	b13c03db-7754-41f9-88ed-7ee09ebdfb33		⋮

Ahora, podemos hacer un test y así, probar con otros datos nuestro modelo, y ver que output nos devuelve: por ejemplo 41 años, 1 educación que es nivel básico, 1 año en el mismo empleo y dejamos todo lo demás, le damos a probar y nos devuelve el resultado de la predicción.

Para poder hacer y configurar la supervisión del rendimiento para evaluar y volver a entrenar el modelo periódicamente para garantizar que el rendimiento del modelo sea aceptable, se necesita una conexión existente de IBM Db2 Warehouse en la nube asociada con su proyecto para utilizarla como su conexión de datos de retroalimentación.

También podemos ponerlo en producción, entrando a la pestaña de deployments. Añadimos un deployment nuevo.

The screenshot shows the IBM Watson Data Platform interface. At the top, there's a navigation bar with links for Projects, Tools, Catalog, Data Services, and Community. Below that, a secondary navigation bar shows 'My Projects' > 'Workshop' > 'Modelo Predictivo Impago'. The main content area is titled 'Modelo Predictivo Impago' with a small info icon. Below the title, there are three tabs: 'Overview', 'Evaluation', and 'Deployments', with 'Deployments' being the active one. A sub-header below the tabs says 'Your model is not deployed.' To the right of the tabs is a button labeled '+ Add Deployment'. A large, empty table follows, which typically lists deployment details like Name, Status, Deployment Type, and Actions. The entire interface has a clean, modern design with a dark header and light body.

Podemos desplegar de tres maneras diferentes nuestros modelos: Web service, Batch Prediction y Real-time Streaming Predictions.

Create Deployment

Web Service Batch Prediction Real-time Streaming Predictions

Name
Web Service Deployment Name

Description
Web Service Deployment Description 300

Cancel Save

Una vez creado el servicio, está listo para utilizarlo. Es decir, creamos por ejemplo un deployment ONLINE que nos servirá para crear una página web o una aplicación móvil, Podemos hacer un deployment en batch, para que se ejecute cada cierto tiempo, o utilizar el Streaming para hacer el deployment en tiempo real.

Create Deployment

Web Service Batch Prediction Real-time Streaming Predictions

Name
Servicio web

Description
Modelo de predicción de impago

269

My Projects > Workshop > Modelo Predictivo Impago

Modelo Predictivo Impago

Overview Evaluation Deployments

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
Servicio web	ACTIVE	Web Service	⋮

(+) Add Deployment

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
Deployment	ACTIVE	Web Service	View Delete

Podemos ver la información del despliegue, y el código para poder crear un api rest que llame a nuestro modelo, podemos testear y evaluar el modelo. Es importante que las variables con las que probemos el modelo sean las adecuadas, es decir, si nos pide un valor entero entre 1 y 5, cuidado no pongamos un valor decimal.

```

# retrieve your SWML_SERVICE_CREDENTIALS_USERNAME, SWML_SERVICE_CREDENTIALS_PASSWORD, and SWML_SERVICE_CREDENTIALS_URL from the
# Service credentials associated with your IBM Cloud Watson Machine Learning Service instance
curl --basic --user SWML_SERVICE_CREDENTIALS_USERNAME:SWML_SERVICE_CREDENTIALS_PASSWORD SWML_SERVICE_CREDENTIALS_URL/v3/identity/token

# the above CURL request will return an auth token that you will use as SWML_AUTH_TOKEN in the scoring request below
# TODO: manually define and pass values to be scored below
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/json' --header "Authorization: Bearer SWML_AUTH_TOKEN" -d '{"fields": ["age", "ed", "employ", "addr"]}'
  
```

The screenshot shows the 'Deployment' section of the IBM Data Science Experience interface. At the top, there are tabs for 'Overview', 'Implementation', and 'Test', with 'Test' being the active tab. Below the tabs is a form titled 'Enter input data' with the following fields:

- age: 41
- ed: 3
- employ: 17
- address: 12

Below the form is a large orange arrow pointing upwards towards the 'Predict' button. The 'Predict' button is located at the bottom left of the input area.

Se propone al lector que cree algún despliegue del modelo obtenido, lo pruebe y evalúe.

This screenshot shows the same 'Deployment' interface as the previous one, but with a visual representation of the predicted value. To the right of the input form, there is a chart titled 'Predicted value for default' with the following data:

Value	Probability (%)
1	76.97%
0	23.03%

The chart is a horizontal bar chart with two bars: a longer blue bar for value 1 and a shorter blue bar for value 0. The total probability is shown as 100% at the top of the chart.

4. Crear un modelo semi-automático o manual.

Para finalizar el Workshop 2, vamos a mostrar cómo hacer un modelo, en lugar de automático, manual. Seguimos los mismos pasos que en el apartado 2.2. Pinchamos en New Model.

The screenshot shows the IBM Watson Data Platform interface. The top navigation bar includes 'IBM Watson Data Platform', 'Projects', 'Tools', 'Catalog', 'Data Services', and 'Community'. A 'My Projects' dropdown shows 'Workshop' is selected. On the left, there are sections for 'Data assets', 'Notebooks', 'Streams flows', 'Models', and 'SPSS Modeler flows'. Under 'Data assets', there is a table with one row for 'datos_banca.csv'. The right sidebar has tabs for 'Load', 'Files' (which is selected), and 'Catalog', with a search bar and a list of selected files.

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
datos_banca.csv	Data Asset	Project	Candela Rietolaza Conde	24 Jan 2018	⋮

Igual que antes, definimos el modelo, y ahora seleccionamos modelo MANUAL

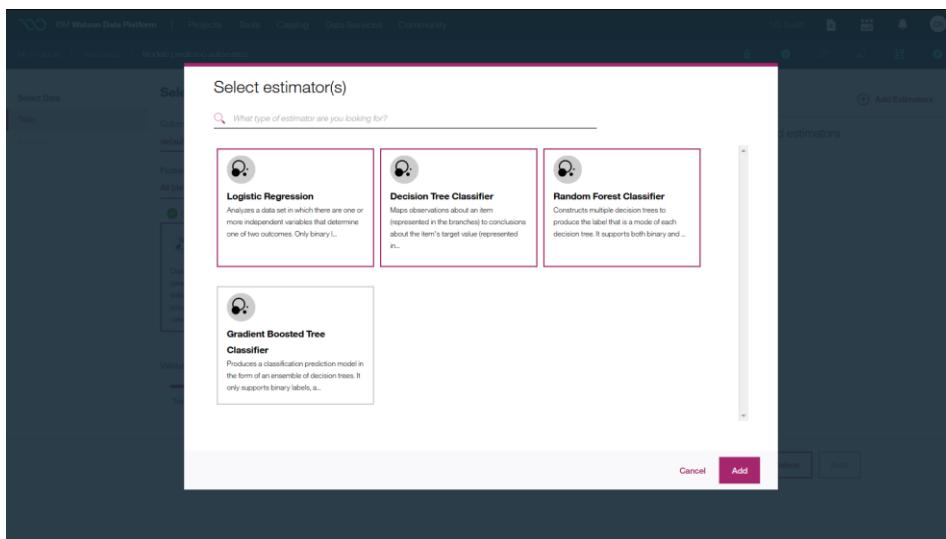
The screenshot shows the 'Define model details' section where the name is set to 'Modelo predictivo automático'. In the 'Select model type' section, 'Model builder' is selected, and 'Spark Service' is chosen. Under 'Automatic', it says 'Prepare my data and create a model automatically'. A note at the bottom suggests creating a notebook or SPSS Modeler flow. At the bottom right are 'Cancel' and 'Create' buttons.

Igual que antes, seleccionamos el fichero de datos de banca.

La diferencia es que ahora nos sugiere una de las técnicas y podemos añadir estimadores (que en la manera automática elegía por nosotros).

The 'Select a technique' dialog shows 'Binary Classification' as the suggested technique. It also lists 'Multiclass Classification' and 'Regression'. A validation split slider is set to Train: 80, Test: 20, Holdout: 20. On the right, there's a 'Configured estimators' section and a 'Add Estimators' button. At the bottom are 'Close', 'Previous', and 'Next' buttons.

Podemos seleccionar uno o varios estimadores. Añadimos y ejecutamos.



ESTIMATOR TYPE	STATUS	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST EVALUATION	ACTIONS
RandomForestClassifier	Trained & Evaluated	Good	0.84825	0.634	24 Jan 2018, 12:39 PM	⋮
LogisticRegression	Trained & Evaluated	Good	0.83114	0.61084	24 Jan 2018, 12:39 PM	⋮
DecisionTreeClassifier	Trained & Evaluated	Poor	0.62931	0.43677	24 Jan 2018, 12:39 PM	⋮

Ahora de los tres estimadores, dos son buenos, y podemos guardarlos y desplegarlos tal y como se explicó en el apartado 2.3.

Workshop 4.

Parte predictiva

1. Notebooks

Para aquellos que no están familiarizados con los Notebooks de Jupiter, los Notebooks permiten a los equipos combinar documentación y código, ejecutar programas línea por línea y combinar los resultados en convincentes visualizaciones. Se trata de un entorno unificado para la colaboración y totalmente accesible por profesionales no técnicos a través de un navegador web.

Para crear un Notebook en IBM Watson Studio(WATSON STUDIO):

Paso 1. Entrar en el proyecto, y desde la vista de Assets del proyecto, haga clic en el enlace Nuevo Notebook.

The screenshot shows the IBM Watson Data Platform interface. At the top, there's a navigation bar with links for Projects, Tools, Catalog, Data Services, and Community. Below that, the 'My Projects' section shows 'Workshop' is selected. The main content area has tabs for Overview, Assets (which is active), Bookmarks, Deployments, Collaborators, and Settings. Under the Assets tab, there's a search bar and a section titled 'Data assets' which lists a single item: 'datos_banca.csv' (Data Asset, Project, Created by Candelaria Retolaza Conde, Last modified 24 Jan 2018). Below that is a 'Notebooks' section with a message 'you currently have no notebooks'. Further down are sections for 'Streams flows' and 'Models'. On the right side, there's a sidebar with tabs for Load, Files (which is active), and Catalog. It shows a file browser with 'datos_banca.csv' listed under 'Find in storage'. The bottom of the screen shows a URL: <https://dataplatform.ibm.com/analytics/notebooks/new-notebook?projectId=63a6251b-8242-4327-aef4-387ad070f513&context=data>.

En la ventana Crear Notebook, especifique el método a usar para crear su notebook.

New notebook

Blank From file From URL

Name*

Notebook de ejemplo

31 Characters Remaining

Description

Type your Description here

Language*

Python 2 R Scala Python 3.5 Experimental

Spark version*

2.1 2.0

Spark service*

Spark-ae

Create Notebook

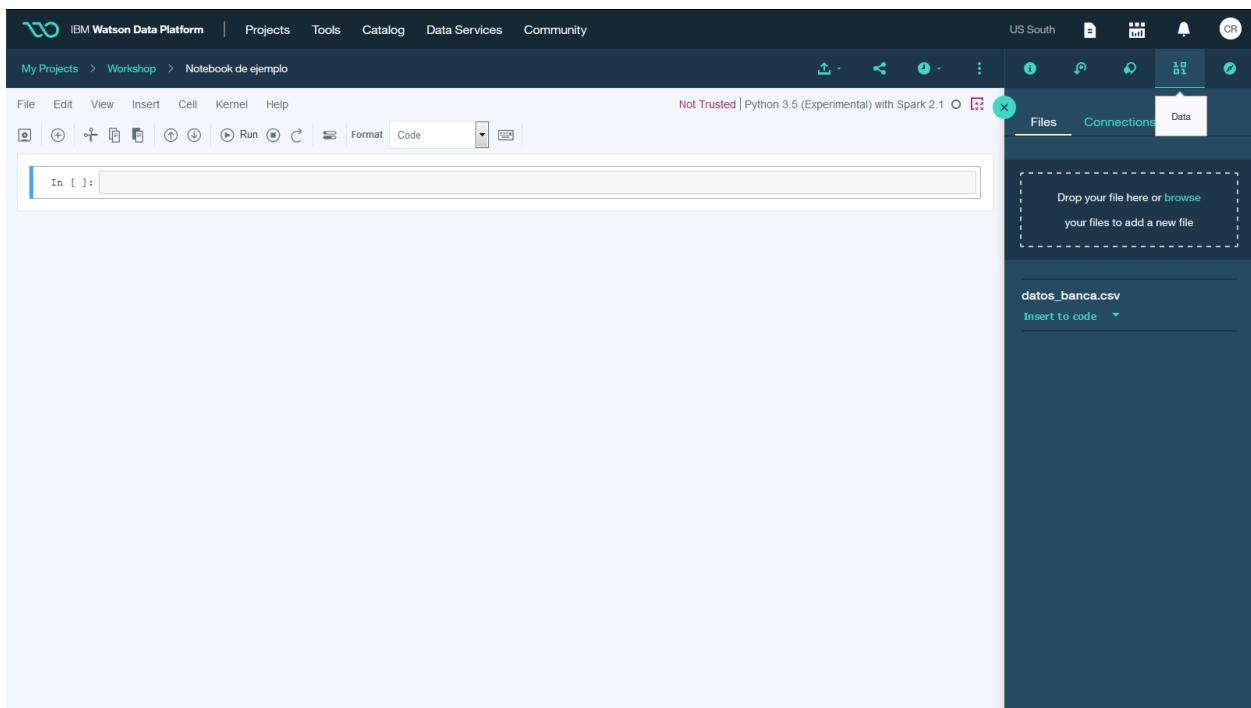
Puedes crear un notebook en blanco, cargar un archivo de notebook desde su sistema de archivos o cargar un archivo de notebook desde una URL. El notebook que crea o selecciona debe ser un archivo.ipynb.

Después de crear un Notebook, estás listo para comenzar a escribir y ejecutar código para analizar datos. Antes de comenzar a codificar, deberá familiarizarse con la interfaz del notebook y cómo codificar en Markdown para escribir el código. Los notebooks se ejecutan en un Kernel de Jupyter en un clúster Spark.

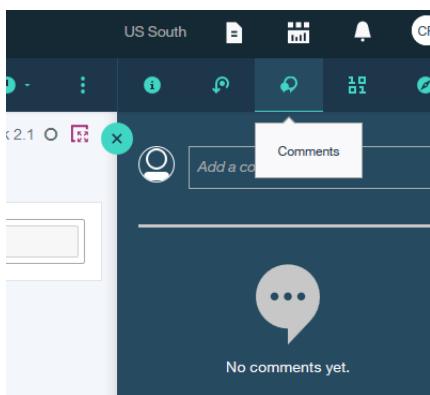
Para desarrollar aplicaciones analíticas en un Notebook, siga estos pasos generales:

- i. Importa bibliotecas preinstaladas para Python y R o instale sus propias bibliotecas.
- ii. Instale bibliotecas personalizadas o de terceros para cualquier idioma. Para Scala, no hay bibliotecas preinstaladas en el servicio Spark. Se almacenan en caché cuando los descarga y solo están disponibles durante el tiempo que se ejecuta el notebook.

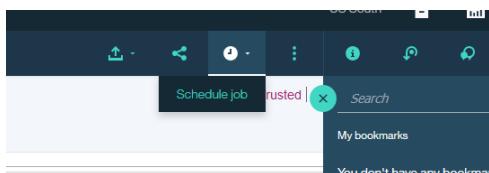
- iii. Cargar y acceder a los datos. Vemos que podemos añadir a nuestro notebook un dataset o un fichero de datos, además de conexiones de datos.

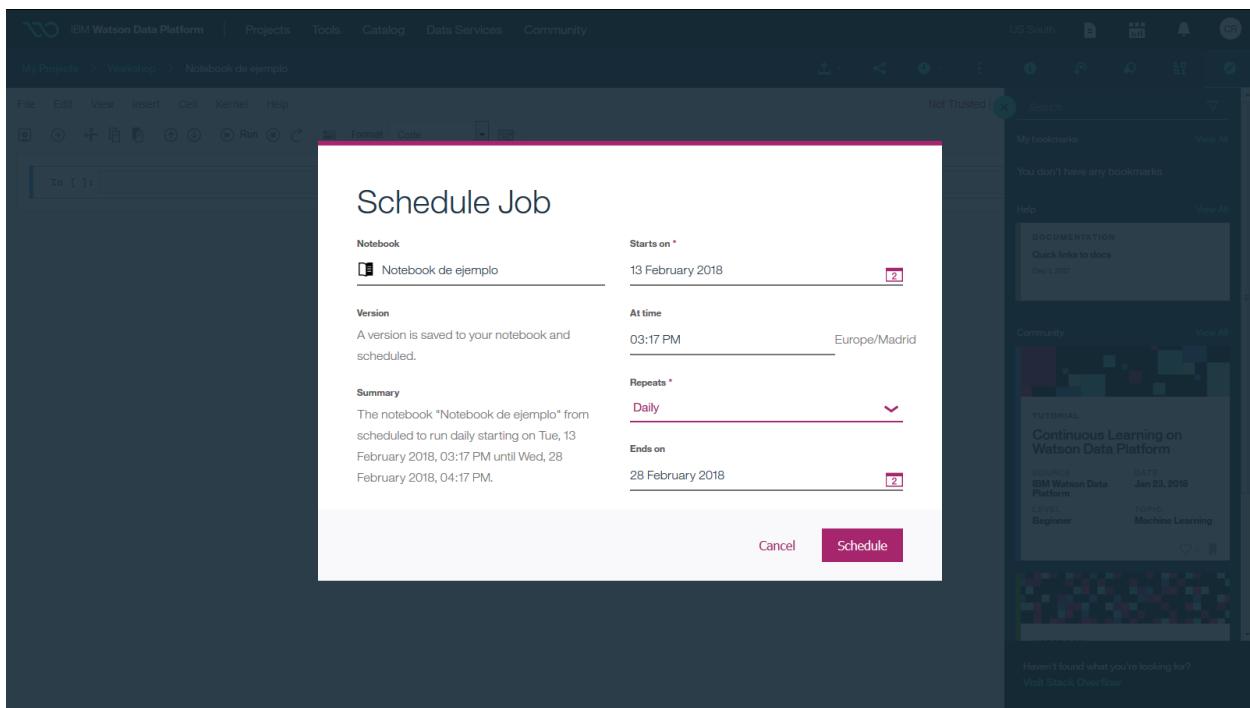


Colabora con otros miembros del proyecto. Puede agregar comentarios a los cuadernos haciendo clic en el icono de comentario (ícono Comentario).

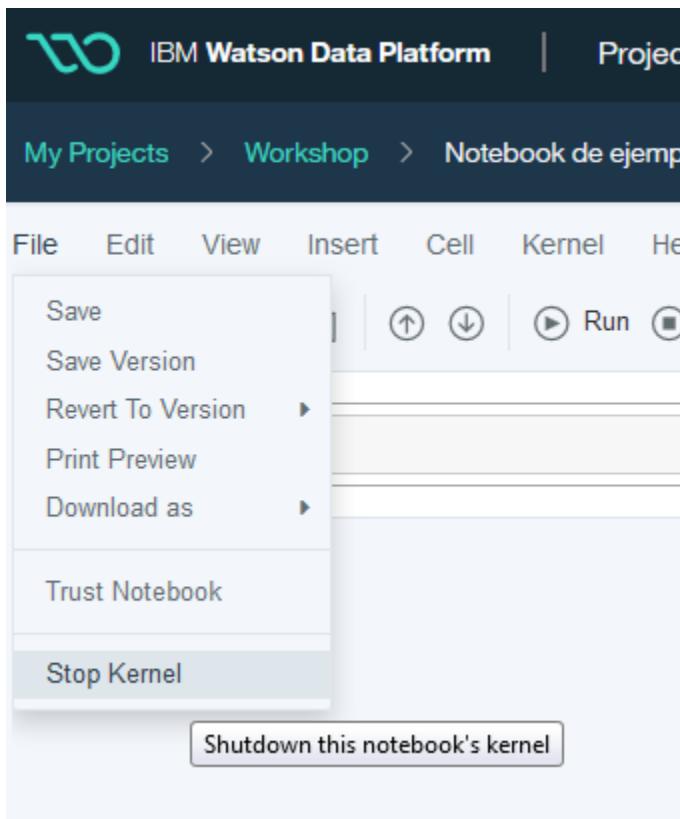


Si es necesario, programa el notebook para que se ejecute en otro momento.

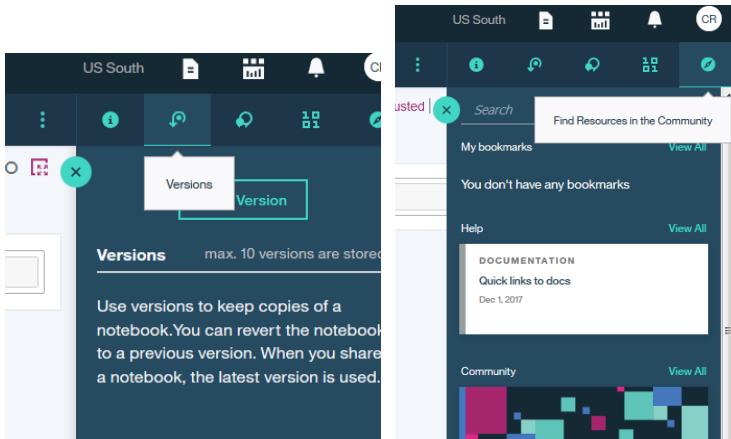




Cuando no estés trabajando activamente en el notebook, haz clic en File > Stop kernel para detener el kernel del notebook.



Además, podemos buscar recursos en la comunidad, y aprovechar esos recursos para enriquecer o comenzar un proyecto.



Para ello, debemos buscar en la comunidad algún notebook que nos pueda interesar, guardarla en el proyecto, y copiarlo para utilizarlo o reutilizar ciertas partes de él.

A continuación, vamos a probarlo. Buscamos un notebook que nos interese: por ejemplo, buscamos por machine learning y escogemos uno que nos resulte interesante. Guardamos, y podemos copiarlo a nuestro proyecto y lo tendremos listo para utilizar.

The screenshot shows the IBM Watson Data Platform interface. The top navigation bar includes 'Projects', 'Tools', 'Catalog', 'Data Services', and 'Community'. The main area shows the 'Workshop' project. The 'Bookmarks' tab is selected, displaying a bookmark for a 'NOTEBOOK' titled 'Classify tumors with machine learning' from Jan 24, 2018. To the right, a search bar is open, showing 'My bookmarks' and a result for the same notebook. Below the search results, there's a 'Help' section with 'DOCUMENTATION' and 'Community' links, and a 'TUTORIAL' section for 'Continuous Learning on Watson Data Platform'.

This screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with links for 'Projects', 'Tools', 'Catalog', 'Data Services', and 'Community'. Below the navigation bar, the title 'Classify tumors with machine learning' is displayed. The main content area contains the notebook's text, including its purpose, three parts, guidelines for reading, and some familiarity requirements. On the right side of the notebook view, there's a sidebar with tabs for 'General' and 'Environment'. The 'General' tab shows the notebook's name ('Classify tumors with machine learning'), a description field ('Write your description here.'), and details about the last editor ('Candela Retolaza Conde') and creation date ('24 Jan 2018, 2:29 PM').

This screenshot shows the same notebook in the IBM Watson Studio interface, but with the right sidebar expanded. The 'General' tab is active, displaying the notebook's name, description, and metadata like the last editor and creation date. The 'Environment' tab is also visible on the right.

Podemos copiar celdas o trozos para reutilizar, etc. Además, podemos compartir o notebooks para que las personas que no tienen cuentas WATSON STUDIO puedan verlos.

Si deseas enseñar a otras personas su notebook pero no quieres que puedan ejecutarlo, puedes darles una URL con una vista de solo lectura.

Si deseas publicar tu notebook para que otras personas puedan copiarlo y ejecutarlo, puedes publicarlo en Github o como gist.

¡asegúrate de ocultar cualquier código, como credenciales, que no quieras que otros vean!

En el apartado siguiente veremos cómo visualizar los resultados.

2. Visualizaciones

Usa visualizaciones en sus notebooks para presentar datos visualmente para ayudar a identificar patrones, obtener información y tomar decisiones.

Muchas de las bibliotecas de visualización de código abierto más comunes, como matplotlib, están preinstaladas en WATSON STUDIO. Todo lo que tienes que hacer es importarlos.

Para ver la lista de bibliotecas instaladas, ejecuta el comando apropiado desde una celda de notebook:

Python: !pip list --isolated

R: installed.packages()

Para importar una biblioteca instalada en tu notebook, ejecute el comando apropiado desde una celda de tu notebook con el nombre de la biblioteca:

Python: import library_name

R: library(library_name)

Puede instalar fácilmente otras bibliotecas y paquetes de visualización. Consulte en la ayuda de WATSON STUDIO: *Install custom or third-party libraries and packages.*

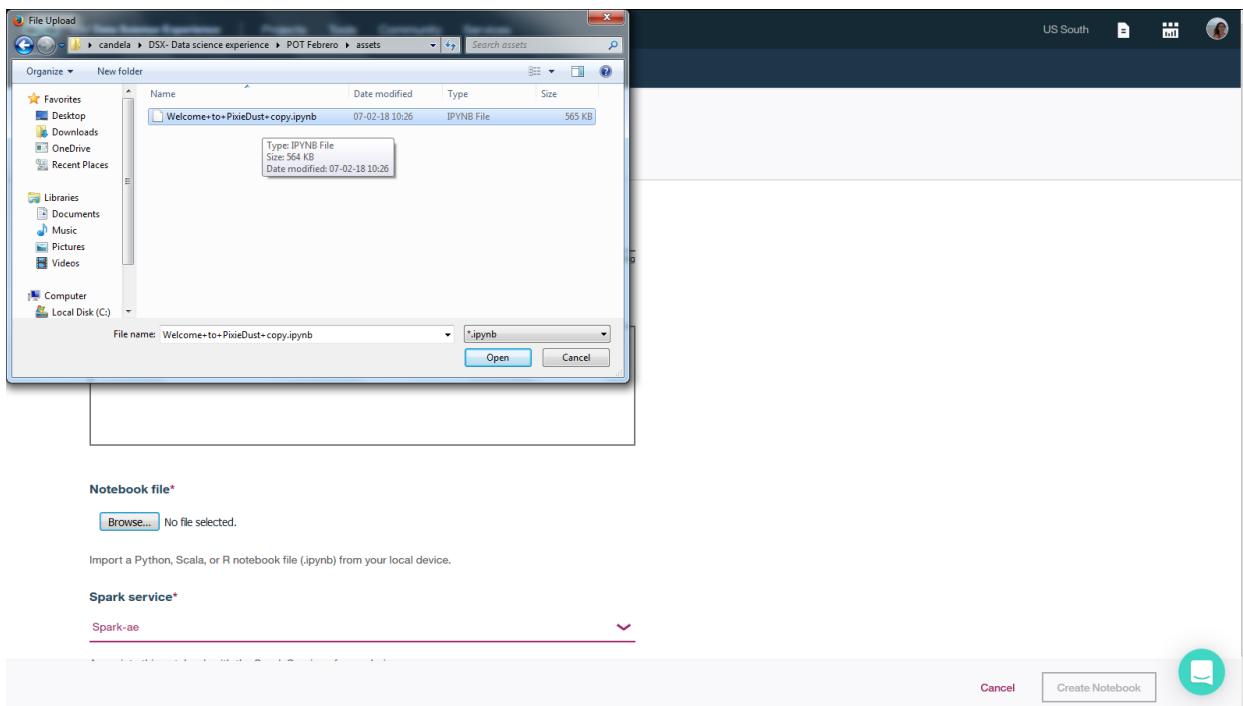
Además, puedes usar estas bibliotecas y herramientas de visualización de IBM:

- PixieDust: cree gráficos con un comando de una sola palabra y luego explore con una interfaz de usuario integrada en lugar de código. Ejecute el código de Scala dentro de los cuadernos de Python.
- Brunel: crea gráficos interactivos con código simple. Prueba en un cuaderno.
- Modelos SPSS: cree tablas y gráficos interactivos para ayudarlo a evaluar y mejorar un modelo de análisis predictivo creado con algoritmos de aprendizaje automático SPSS.

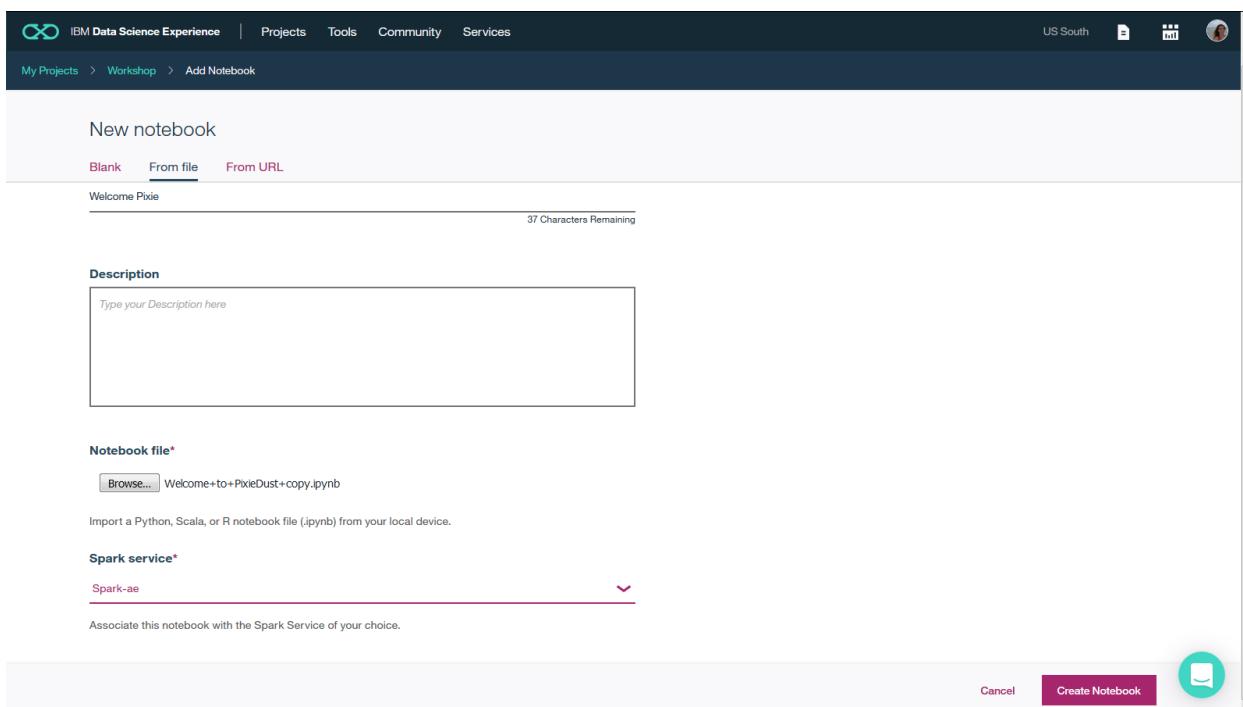
Puedes usar las siguientes bibliotecas de visualización en Notebooks de Scala: PixieDust, Brunel for Scala y Lightning for Scala.

Vamos a subir un Notebook llamado ‘Welcome Pixie Dust’ para comenzar a desenvolvernos. Creamos un nuevo notebook, pero vamos a subirlo desde un fichero. Tenemos el siguiente fichero en nuestro escritorio: *Welcome+to+PixieDust+copy.ipynb* ponemos un nombre y lo subimos a WATSON STUDIO:

The screenshot shows the 'Add Notebook' page in the IBM Data Science Experience. At the top, there's a navigation bar with links for 'Projects', 'Tools', 'Community', and 'Services'. Below that, a breadcrumb trail shows 'My Projects > Workshop > Add Notebook'. The main area is titled 'New notebook' and has three tabs: 'Blank' (selected), 'From file', and 'From URL'. A 'Name*' field is present with a placeholder 'Type Notebook Name here'. A 'Description' section contains a text input field with the placeholder 'Type your Description here'. A 'Notebook file*' section includes a 'Browse...' button and a message 'No file selected.'. Below it, a note says 'Import a Python, Scala, or R notebook file (.ipynb) from your local device.' A 'Spark service*' dropdown menu is set to 'Spark-ae'. At the bottom right are 'Cancel', 'Create Notebook', and a circular icon with a document symbol.



Solo podemos seleccionar un servicio de Spark y creamos el notebook:



Se propone al lector seguir los pasos del notebook para comenzar a familiarizarse con los notebooks y con las visualizaciones de PixieDust.

The screenshot shows the IBM Watson Data Platform interface. At the top, there's a navigation bar with links for Projects, Tools, Catalog, Data Services, and Community. The Community tab is highlighted. On the right side of the header, there are icons for US South, a file, a bell, and a CR button. Below the header, a URL https://opensource.org/licenses/m... and the IBM logo are visible.

The main content area displays a notebook titled "Welcome to PixieDust". It includes a brief introduction, a Table of Contents, and a "Get started" section. The "Get started" section contains instructions and a code cell:

```
In [1]: # To confirm you have the latest version of PixieDust on your system, run this cell
!pip install --user --upgrade pixiedust
```

Below the code cell, there's a detailed log of package requirements:

```
Requirement already up-to-date: pixiedust in /gpfs/global_fs01/sym_shared/YFProdSpark/user/s2b7-790f27d2e466b6-772f4elcd93d/.local/lib/python2.7/site-packages
Requirement already up-to-date: mld3 in /gpfs/global_fs01/sym_shared/YFProdSpark/user/s2b7-790f27d2e466b6-772f4elcd93d/.local/lib/python2.7/site-packages (from pixiedust)
Requirement already up-to-date: lmxl in /gpfs/global_fs01/sym_shared/YFProdSpark/user/s2b7-790f27d2e466b6-772f4elcd93d/.local/lib/python2.7/site-packages (from pixiedust)
Requirement already up-to-date: geojson in /gpfs/global_fs01/sym_shared/YFProdSpark/user/s2b7-790f27d2e466b6-772f4elcd93d/.local/lib/python2.7/site-packages (from pixiedust)
```

Otra forma de conseguir el mismo notebook:

Buscamos en la comunidad el notebook ‘Welcome to PixieDust’, lo abrimos, lo copiamos en un proyecto y lo editamos, para poder modificar o ejecutar celda a celda.

This screenshot shows the same notebook from the previous interface, but with a different perspective. A red arrow points to a "Copy" button located in the top right corner of the notebook's preview area. The rest of the interface is identical to the first one, showing the navigation bar, the notebook content, and the code cell at the bottom.

New notebook: Welcome to PixieDust

Project

Workshop

Add the notebook to an existing project.

Spark service*

Spark-ae

Associate this notebook with the IBM Analytics for Apache Spark Service of your choice.

Create Notebook

Una vez tenemos el notebook en nuestro proyecto, lo abrimos y le damos a editar. Podemos ejecutar celda a celda y seguir las instrucciones del notebook.

Welcome to PixieDust

This notebook features an introduction to [PixieDust](#), the Python library that makes data visualization easy.

This notebook runs on Python 2.7 and 3.5, with Spark 2.0.

Table of Contents

- Get started
- Load text data from remote sources
- Mix Scala and Python on the same notebook
- Add Spark packages and run inside your notebook
- Stash your data
- Contribute

Get started

This introduction is pretty straightforward, but it wouldn't hurt to load up the [PixieDust documentation](#) so it's handy.

New to notebooks? Don't worry. Here's all you need to know to run this introduction:

1. Make sure this notebook is in Edit mode
2. To run code cells, put your cursor in the cell and press **Shift + Enter**
3. The cell number will change to [1] to indicate that it is currently executing. (When starting with notebooks, it's best to run cells in order, one at a time.)

```
In [1]: # To confirm you have the latest version of PixieDust on your system, run this cell
!pip install --user --upgrade pixiedust
Requirement already up-to-date: pixiedust in /usr/local/src/bluemix_jupyter_bundle.v79/notebook/lib/extras
Collecting markdown (from pixiedust)
  Downloading Markdown-2.6.11-py2.py3-none-any.whl (79kB)
    100% |██████████| 81kB 1.2MB/s eta 0:00:01
Collecting lxml (from pixiedust)
  Downloading lxml-4.1.1-cp27mu-manylinux1_x86_64.whl (5.6MB)
    100% |██████████| 5.6MB 165kB/s eta 0:00:01
Requirement already up-to-date: astunparse in /usr/local/src/bluemix_jupyter_bundle.v79/notebook/lib/python2.7/site-packages
s (from pixiedust)
```

The screenshot shows the IBM Watson Studio interface with the 'My Projects' section selected. A Jupyter Notebook titled 'Welcome to PixieDust' is open. The notebook has three cells:

- Cell 1 (In [1]):** Contains the command `# To confirm you have the latest version of PixieDust on your system, run this cell
!pip install --user --upgrade pixiedust`. The output shows requirements for various Python packages from the pixiedust repository.
- Cell 2 (In [2]):** Contains the command `import pixiedust`. The output shows the message "Pixiedust database opened successfully" and "Pixiedust version 1.0.9".
- Cell 3 (In [3]):** Contains the command `# Build the SQL context required to create a Spark dataframe
text(sc)`. The output shows "Connecting to fast.trychameleon.com...".

3. Algoritmos de analítica predictiva de SPSS

Una de las funcionalidades más interesantes que proporciona IBM como valor añadido a WATSON STUDIO es ésta.

SPSS Modeler es una herramienta muy estable y muy potente para realizar minería de datos. Es una herramienta que permite al equipo de científicos de datos realizar todo el proceso de minería de datos siguiendo CRISP-DM, es decir, en SPSS podemos acceder a los datos (ya estén en cualquier base de datos, o ficheros planos) podemos limpiar y modificar los datos, después tenemos más de 50 modelos (árboles de regresión, clústeres, redes neuronales, regresiones, etc.) para analizarlos y posteriormente podemos exportar esos datos, hacer gráficos, o ponerlo en producción.

WATSON STUDIO está adquiriendo cada vez más funcionalidades de SPSS Modeler para añadirla a sus funcionalidades propias. Gracias a esto, el usuario es capaz de hacer minería de datos de una manera más sencilla, y así enriquecer los proyectos.

Ahora vamos a hacer analítica con los flujos de SPSS.

The screenshot shows the IBM Watson Data Platform interface with the 'Workshop' tab selected. The main area displays various projects and flows:

- Projects:** Notebook de ejemplo (Python 3.5), Classify tumors with machine learning (Python 2.7).
- Streams flows:** A table showing no streams flows.
- Models:** A table showing two models: 'Modelo predictivo automático' (untrained) and 'Modelo Predictivo Impago' (trained).
- SPSS Modeler flows:** A table showing no spss modeler flows. This section has a red arrow pointing to the '+ New flow' button.
- Data flows:** A table showing no data flows.

The right sidebar shows file management options like Load, Files, Catalog, and a search bar for 'Find in storage'.

Podemos crear un flujo nuevo o si somos usuarios de SPSS Modeler, podemos importar modelos que ya tengamos hechos o empezar con un ejemplo. Vamos a comenzar con un ejemplo para familiarizarnos y posteriormente crearemos una ruta con los datos que hemos cargado antes.

Select one of the samples below to get started with an existing stream that suits the kind of modeling you want to do. When you create the stream it will be added to your project, allowing you to modify and save your changes.

SPSS MODELER

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.

SPSS MODELER

Sales Promotion Study

Use neural network and C5.0 algorithms to predict the effect of advertising promotions on the sale of various items. Input data of sales before and after a post promotion are used to train the model to predict the effectiveness of advertising.

Cancel Creating...

My Projects > Workshop > Drug Study Example

Cancel Creating...

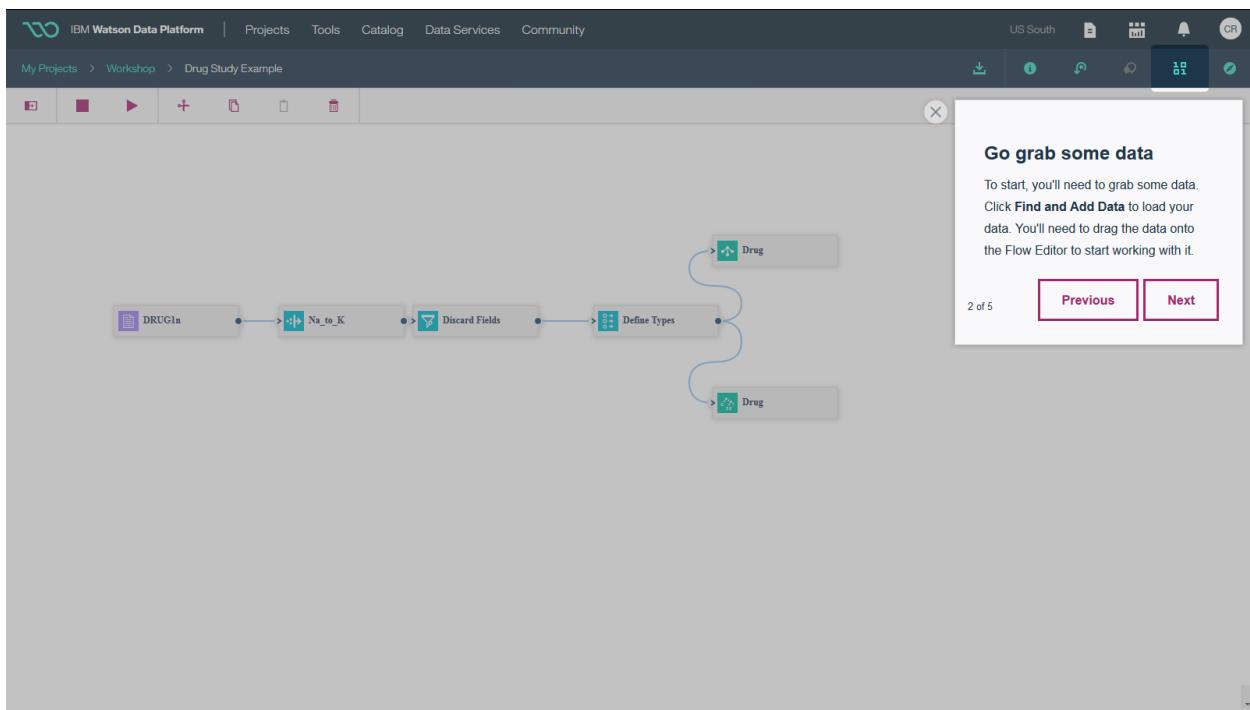
DRUG1a → Na_to_K

Welcome to flows

Use flows to transform data and create predictive models. This tour leads you through some of the opening steps to get you started.

Start Tour

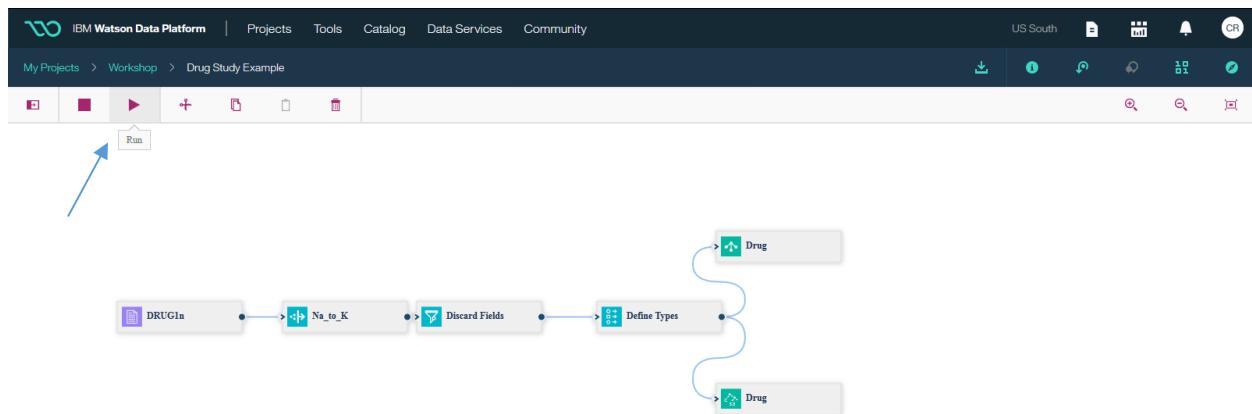
1 of 5



En este ejemplo, imagina que es un investigador médico que está recopilando datos para un estudio. Has recopilado información sobre un conjunto de pacientes, de los cuales todos sufrieron la misma enfermedad. Durante el curso del tratamiento, cada paciente respondió a un medicamento de un total de cinco. Parte de su trabajo consiste en utilizar minería de datos para averiguar qué medicamento es el adecuado para un futuro paciente con la misma enfermedad.

AGE	SEX	BP	CHOLESTEROL	NA	K	DRUG
23	F	HIGH	HIGH	0.792535	0.031258	drugY
47	M	LOW	HIGH	0.739309	0.056468	drugC
47	M	LOW	HIGH	0.697269	0.068944	drugC
28	F	NORMAL	HIGH	0.563682	0.072289	drugX
61	F	LOW	HIGH	0.559294	0.030998	drugY
22	F	NORMAL	HIGH	0.676901	0.078647	drugX
49	F	NORMAL	HIGH	0.789637	0.048518	drugY
41	M	LOW	HIGH	0.766635	0.069461	drugC
60	M	NORMAL	HIGH	0.777205	0.05123	drugY
43	M	LOW	NORMAL	0.526102	0.027164	drugY

Debido a que el sodio y el potasio tienen una alta correlación, se puede observar que es así haciendo un gráfico, se crea una nueva variable para que pueda ser utilizada en el modelo.

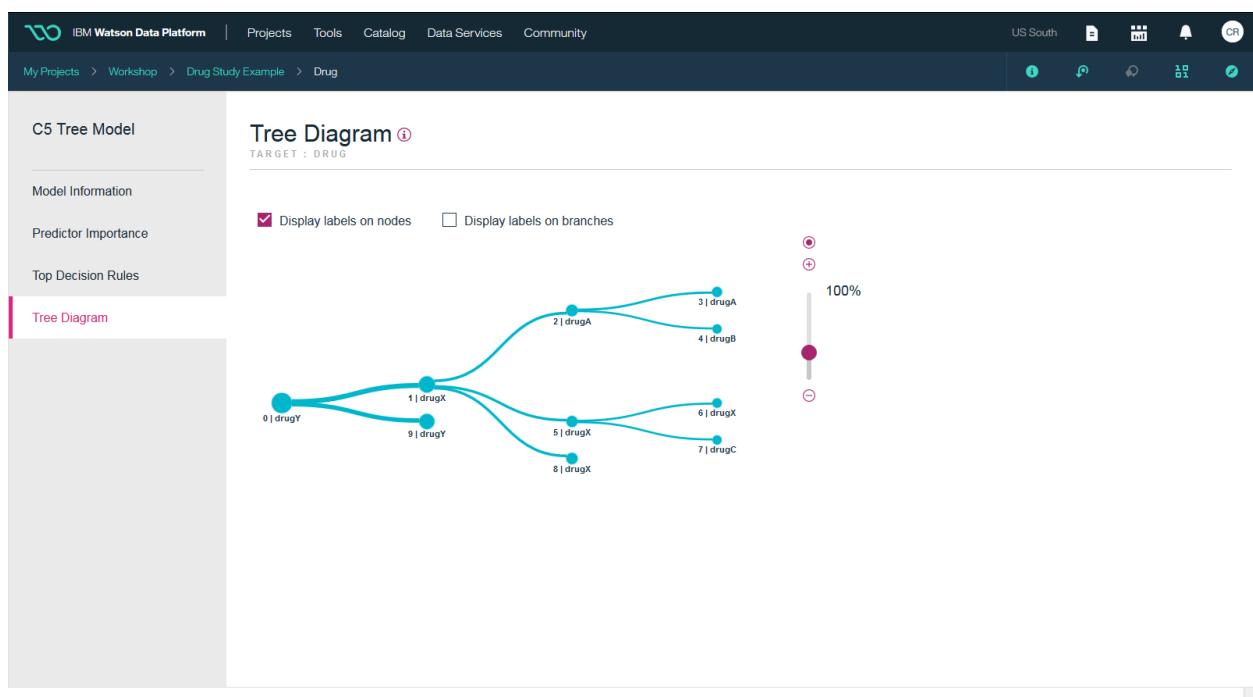
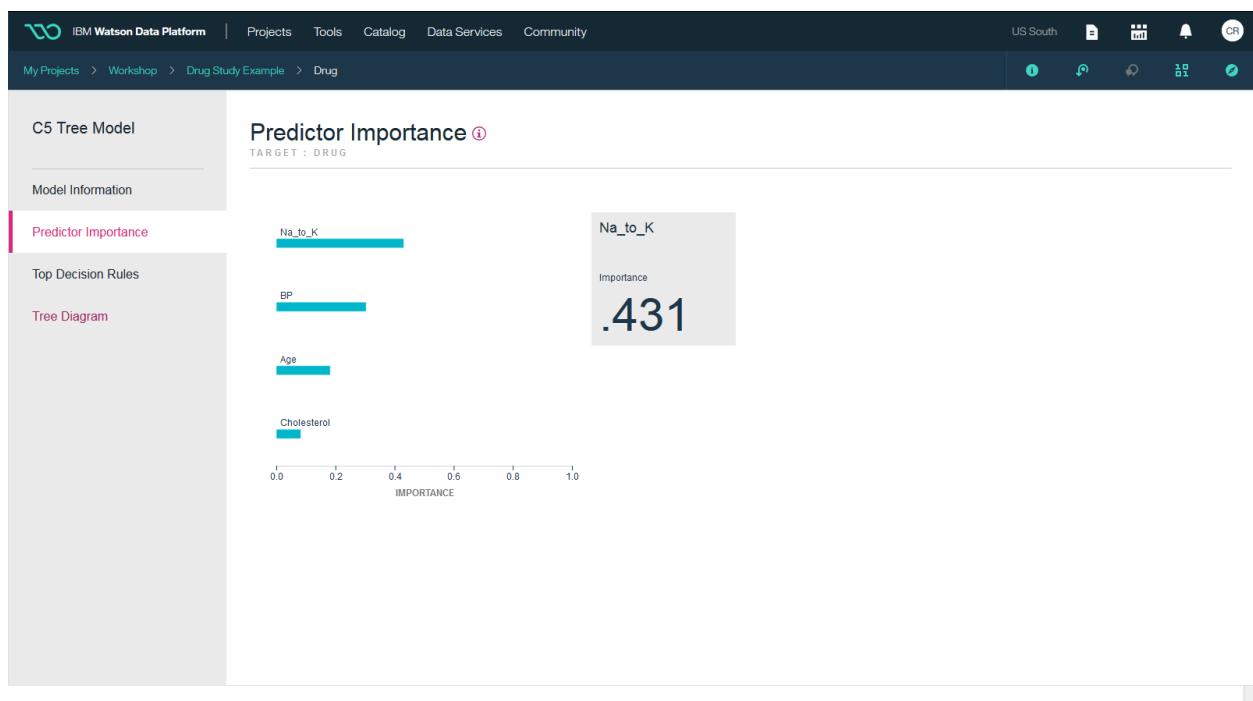


The screenshot shows the IBM Watson Data Platform Workshop interface. At the top, there's a navigation bar with links for Projects, Tools, Catalog, Data Services, and Community. On the right side of the header, there are icons for US South, download, search, and other workspace functions. Below the header, the main area displays a data pipeline flow. The flow starts with a 'DRUG1a' dataset, followed by a 'Na_to_K' transformation, a 'Discard Fields' step, and a 'Define Types' step. From the 'Define Types' step, four parallel paths lead to four separate 'Drug' nodes, each represented by a green icon with a cross.

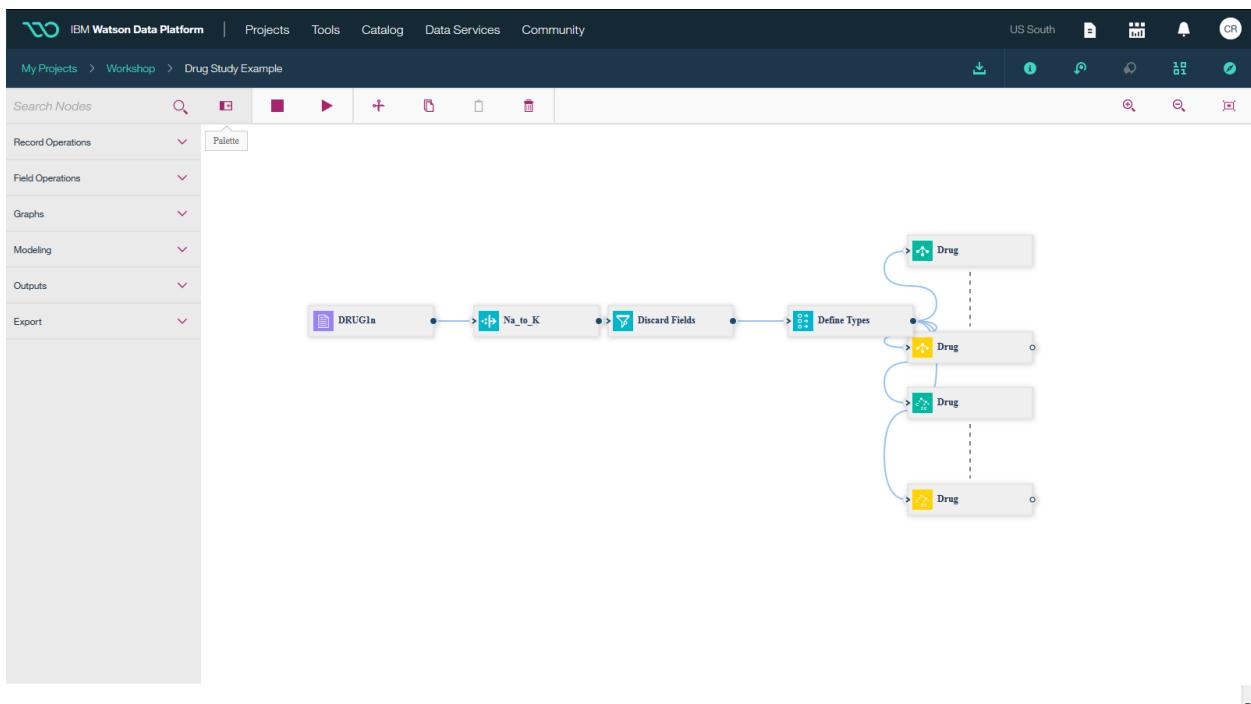
Para observar los resultados: Botón derecho: view model

The screenshot shows the 'Model Information' view for a C5 Tree Model. The title is 'C5 Tree Model' and the target field is 'Drug'. The 'Model Information' section contains the following details:

Predictor Importance	Drug
Top Decision Rules	
Tree Diagram	
Target Field	Drug
Model Type	Multi-Class Decision Tree
Algorithm Name	C5
Number of Features	4
Tree Depth	4
Number of Nodes	10



Se sugiere al lector que entienda el modelo y que lo modifique si es necesario. Podemos copiar el modelo en nuestro proyecto.



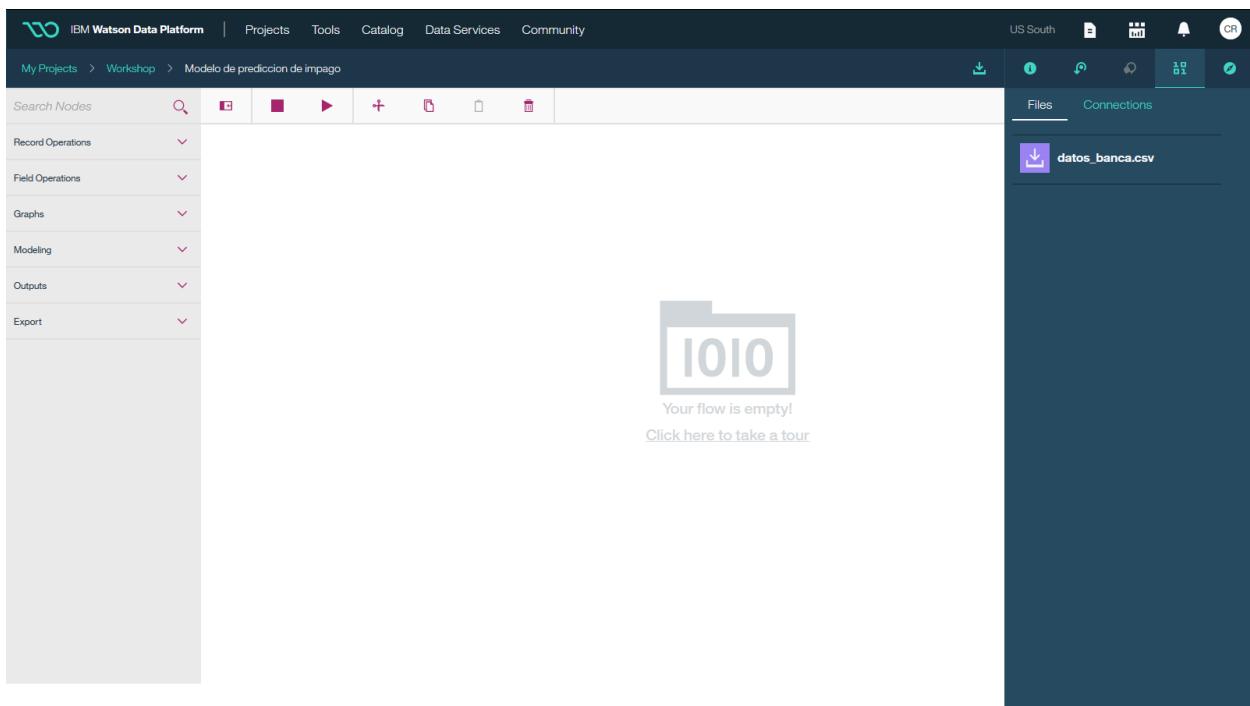
A continuación, vamos a hacer uno nuevo.

The screenshot shows the creation of a new model in SPSS Modeler BETA. The configuration screen includes the following fields:

- Name***: Modelo de predicción de impago
- Description**: Type description here.
- Runtime**: IBM SPSS Modeler

A progress bar at the bottom right shows the status as "Creating...".

seleccionamos el runtime de spss modeler y creamos:



Como veis tenemos una paleta a la izquierda en la que tenemos los nodos, y un lienzo en blanco en el que hacer nuestra ruta o Flow. Se recomienda al lector que explore los nodos.

Por ejemplo, la pestaña de la paleta Operaciones con registro contiene nodos que puede utilizar para realizar operaciones en los *registros* de datos como, por ejemplo, seleccionar, fusionar y añadir.

Los nodos Operaciones con campos realizan operaciones en *campos* de datos como, por ejemplo, filtrar, derivar campos nuevos y determinar el nivel de medición para campos dados.

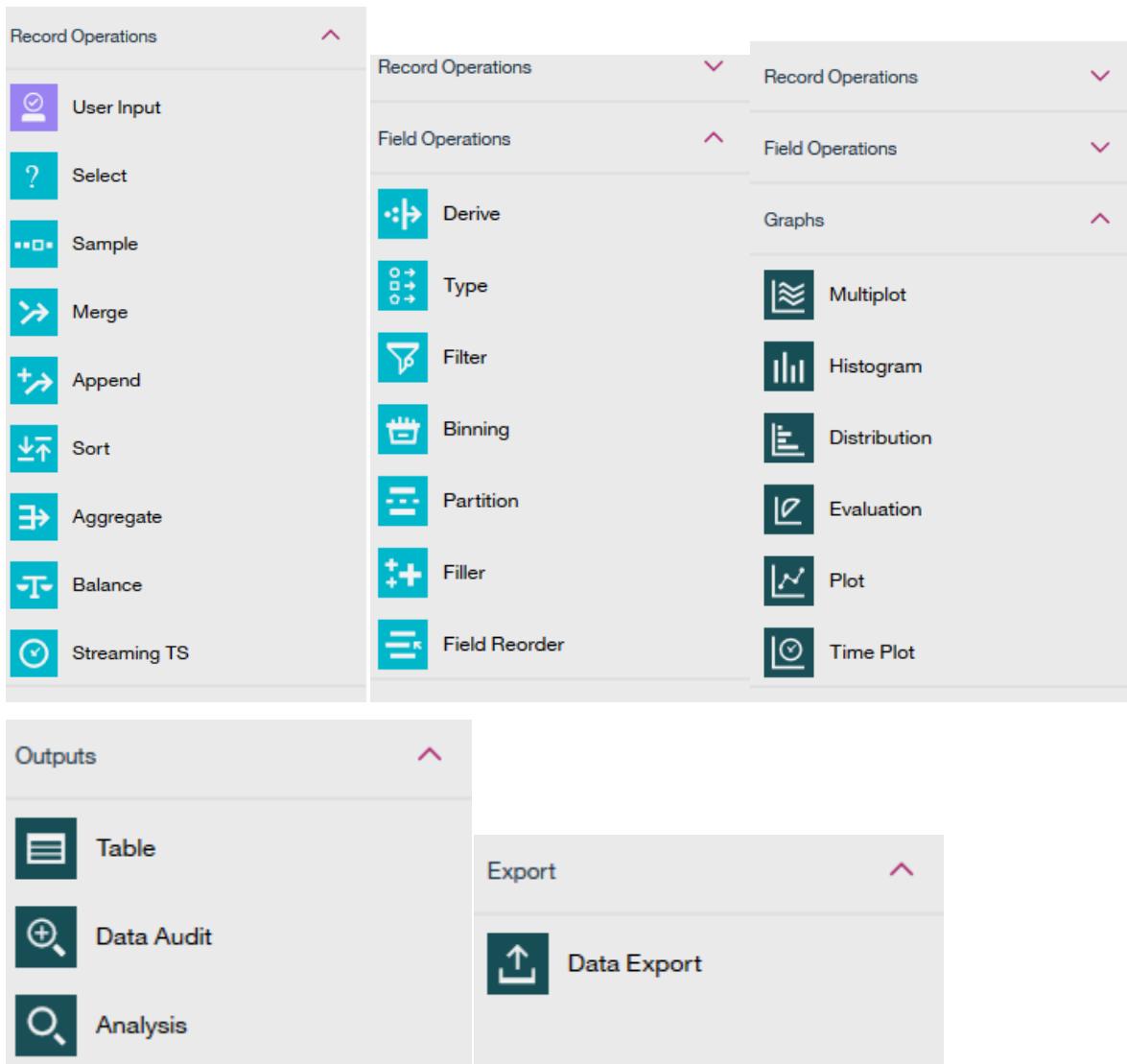
Los nodos Gráficos muestran gráficamente los datos antes y después del modelado. Entre ellos se incluyen gráficos, histogramas, nodos de malla y diagramas de evaluación.

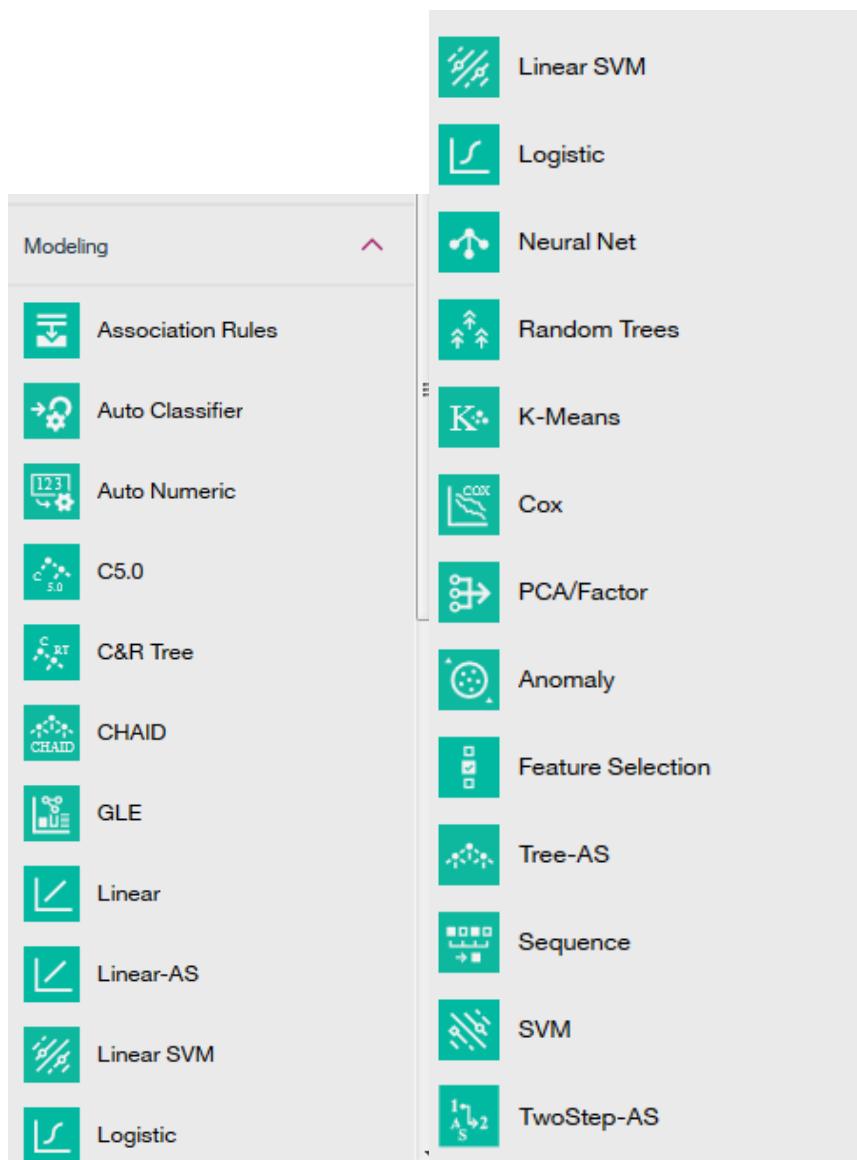
Los nodos Modelado utilizan los algoritmos de modelado disponibles en SPSS Modeler como, por ejemplo, redes neuronales, árboles de decisión, algoritmos de agrupación en clúster y secuenciación de datos.

Los nodos Salida generan diferentes salidas para resultados de datos, gráficos y modelos que se pueden visualizar en WATSON STUDIO.

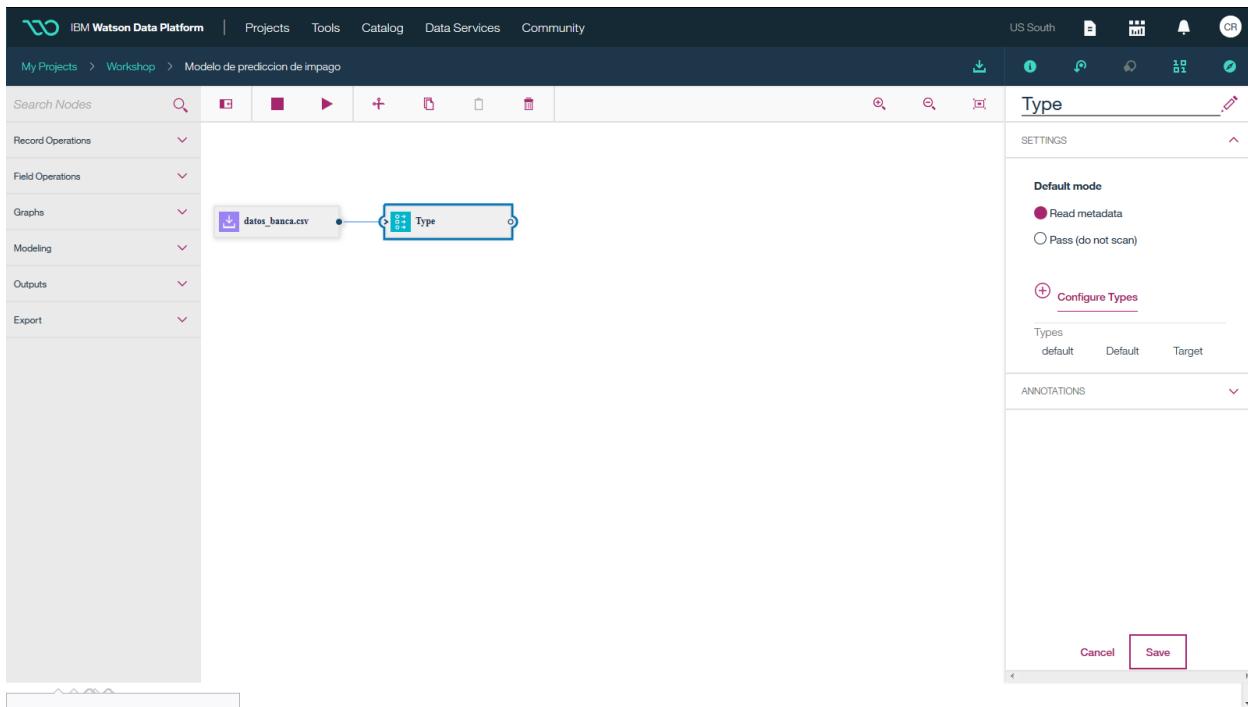
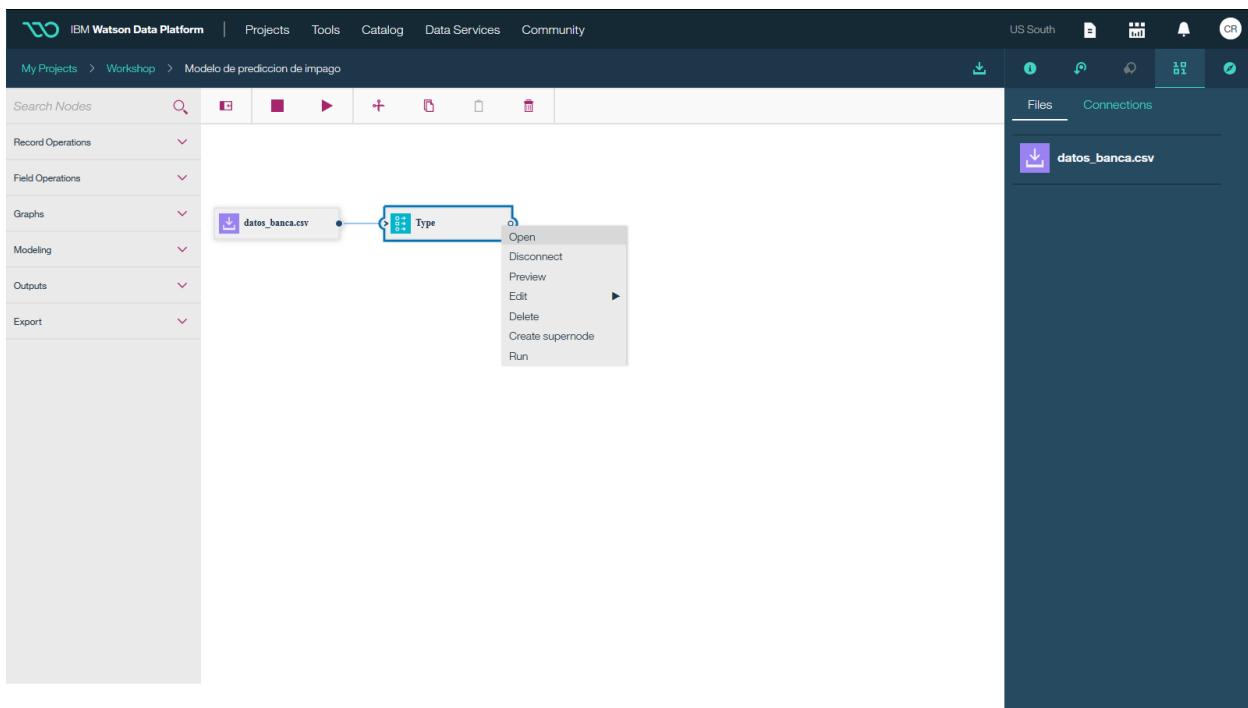
En la documentación completa de WATSON STUDIO se encuentra una descripción completa de cada nodo.

<https://dataplatform.ibm.com/docs/content/analyze-data/ml-canvas-spss.html?audience=wdp&context=analytics>





Para realizar el modelo, seleccionamos el fichero de entrada y vemos que incluso podemos añadir uno nuevo o conectarnos a una base de datos que tengamos. Arrastra el fichero al lienzo. Y añade un nodo tipo, para instanciar los valores, y para seleccionar Default como target.



The screenshot shows the IBM Watson Studio interface with a project titled "Modelo de predicción de impago". On the left, a sidebar lists various operations like Record Operations, Field Operations, Graphs, Modeling, Outputs, and Export. In the center, a flow diagram shows a CSV file node connected to a "Type" node. A modal window titled "Select Fields for Type" is open, displaying a table of fields and their data types. The table includes:

Field name	Data type
age	double
ed	double
employ	double
address	double
income	double
debtinc	double
creddebt	double
othdebt	double
default	double
preddef1	double
preddef2	double
preddef3	double

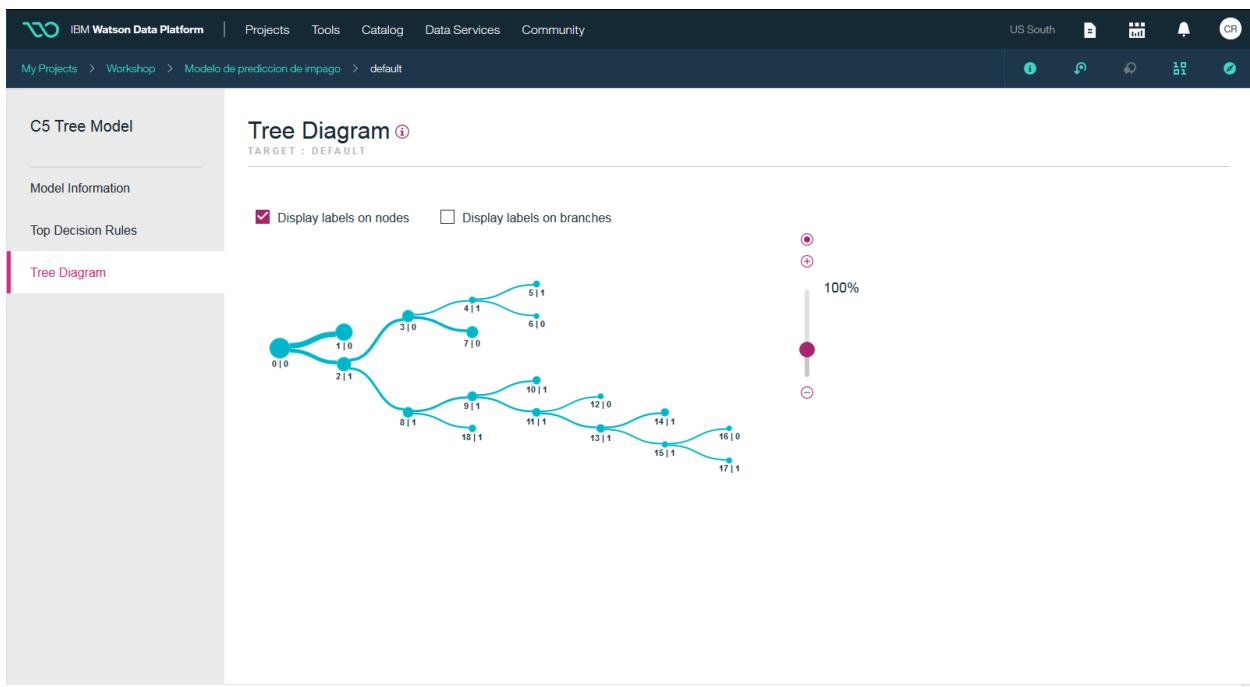
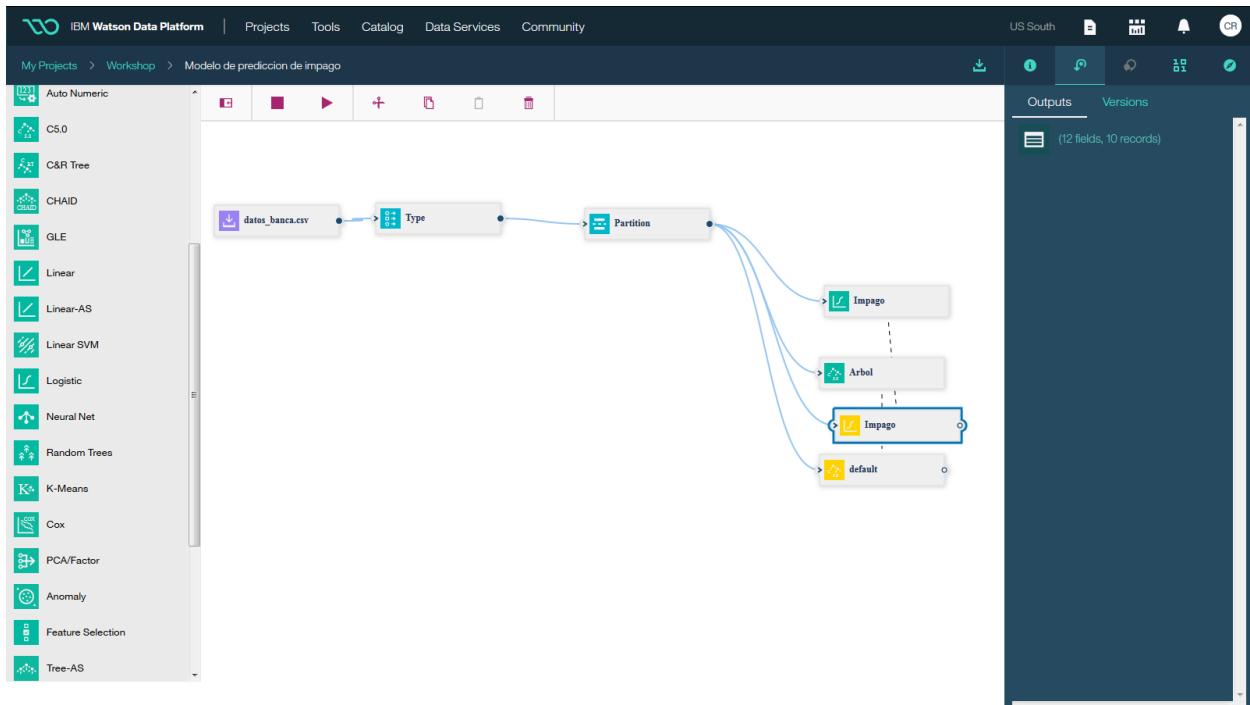
The screenshot shows the "Configure Types" dialog in IBM Watson Studio. It lists five entries: "othdebt", "default", "preddef1", "preddef2", and "preddef3". Each entry has dropdown menus for "Field", "Measure", "Role", "Value mode", "Values", and "Check". The "Role" column for "othdebt" is set to "Input". At the bottom right of the dialog are "Cancel" and "OK" buttons.

Field	Measure	Role	Value mode	Values	Check
othdebt	Default	Input	Read	None	None
default	Default	Target	Read	None	None
preddef1	Default	None	Read	None	None
preddef2	Default	None	Read	None	None
preddef3	Default	None	Read	None	None

Como es un fichero preparado no se necesita hacer una limpieza de los datos, pero todo esto es posible hacerlo también aquí, con los nodos de operaciones con los datos, si no lo hemos hecho en un paso previo con la parte de Data Refinery.

Además, podemos añadir varias fuentes de entrada y cruzar las tablas, elegir lo que queramos, etc.

Se invita al lector que acabe el modelo y consiga predecir las variables que generan impago.



Se deja al lector otro fichero, llamado baskets.csv para practicar ya sea con Modelos automáticos, con Notebooks, o con flows de SPSS. El objetivo será el siguiente:

Este ejemplo está relacionado con datos ficticios que describen el contenido de cestas de supermercado (es decir, una colección de artículos comprados a la vez) junto con los datos personales del comprador, que pueden obtenerse a través de las tarjetas de fidelidad. El objetivo es descubrir grupos de clientes que compren productos parecidos calificables desde el punto de vista demográfico, como por edad, ingresos, etc.

Este ejemplo muestra dos fases de la minería de datos:

- Modelado de reglas de asociación y una visualización de malla que muestra enlaces entre los artículos comprados.
- Perfilado de reglas de inducción C5.0 de los compradores de grupos identificados de productos.

Nota: Esta aplicación no utiliza directamente el modelado predictivo y, por tanto, no hay ninguna medición de la precisión de los modelos resultantes ni entrenamiento asociado/distinción de comprobaciones en el proceso de minería de datos.

4. RStudio

R es un popular paquete de análisis estadístico y aprendizaje automático que permite la gestión de datos e incluye pruebas, modelos, análisis y gráficos, y permite la gestión de datos. RStudio, incluido en IBM Watson Studio, proporciona un IDE para trabajar con R.

Una sesión de RStudio creada en Watson Studio incluye 2 GB de almacenamiento y 5 GB de memoria disponible para su uso.

También podemos hacer que nuestro análisis en R sea accesible para los no programadores a través de Shiny. Shiny es una excelente opción para implementar el análisis de minería de datos a los usuarios de negocio.

Shiny es un marco de aplicaciones web para R que permite convertir el análisis de R en aplicaciones web interactivas. No se requieren conocimientos de HTML, CSS o JavaScript.

The top screenshot shows the IBM Watson Data Platform interface. The navigation bar includes 'Projects', 'Tools', 'Catalog', 'Data Services', and 'Community'. The 'Catalog' tab is selected, showing a 'Get started' button and a 'Create Catalog' button. Below this, there's a section titled 'New in the community' featuring four items: an article about predicting flight cancellations, a dataset for customer demographics and sales, a notebook for learning basics about notebooks and Apache Spark, and a tutorial on continuous learning on Watson Data Platform.

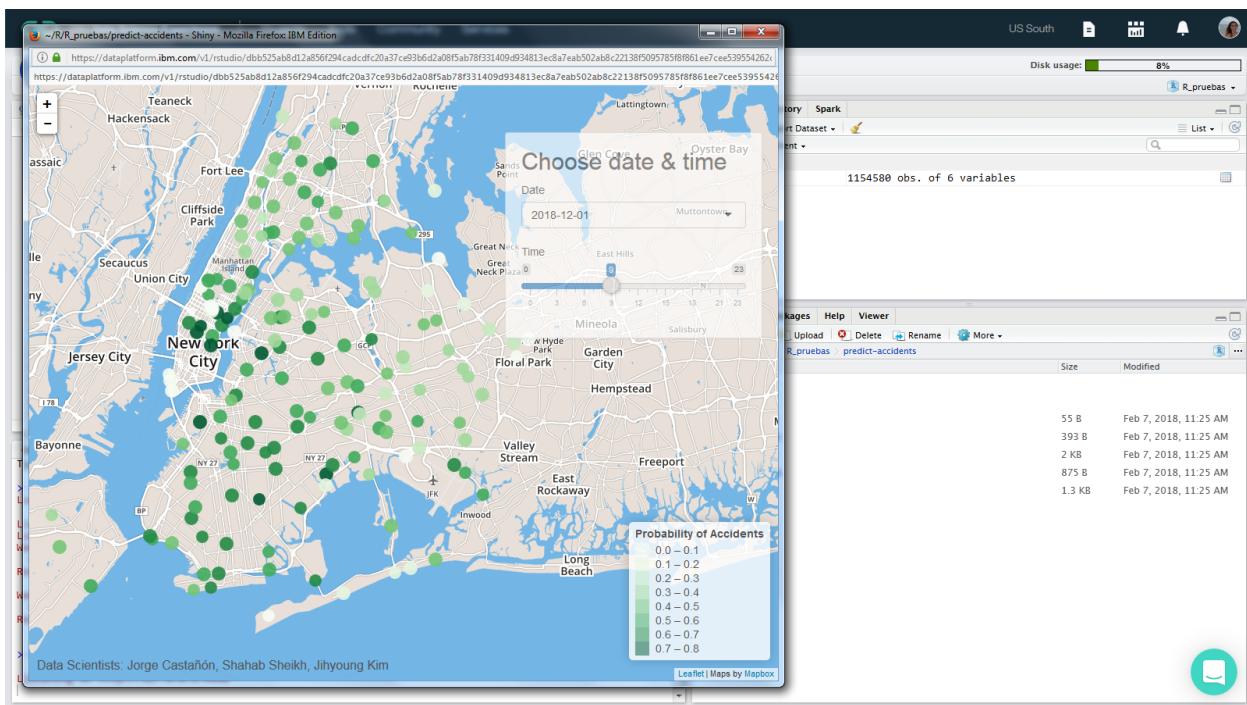
The bottom screenshot shows the RStudio interface within the Watson Data Platform. The R console displays the startup message for R version 3.3.2. The file browser on the right shows a directory structure with files like 'profile', 'config.yml', 'item-sparkaa-demos', 'lost+found', and 'R'.

Proponemos al lector seguir los pasos de los siguientes proyectos (Predictions of car accident in NYC based on weather data, analyzing flight delays, y Blocpower) que se encuentran en el siguiente repositorio:

<https://github.com/IBMDatascience/Watson Studio-shiny-apps>

Cuando accedas al repositorio de GitHub, primero lee el Readme.md y sigue los pasos.

El primero de los proyectos te muestra con la aplicación de Shiny los resultados de un modelo predictivo de accidentes de coche en New York. Es un modelo entrenado con datos históricos de accidentes e información meteorológica.



5. Librerías de Deep Learning

Deep learning es una rama de machine learning que utiliza grandes cantidades de datos para enseñar a las máquinas cómo hacer tareas o cosas que antes sólo eran capaces de hacer los seres humanos.

Buenos ejemplos de Deep learning son la percepción, el reconocimiento de lo que hay en una imagen, lo que las personas dicen cuando hablan, o ayudar a los robots a explorar el mundo e interactuar con él. El Deep learning está emergiendo como una herramienta central para

resolver problemas de percepción en los últimos años. Son los modelos que están detrás de la visión artificial y el reconocimiento de voz. Cada vez más personas descubren que el Deep learning es una herramienta muy potente para resolver multiples problemas.

Muchas empresas de hoy en día han convertido el Deep learning en una parte central de su conjunto de herramientas de aprendizaje automático. Por ejemplo, Facebook, Google y Uber están utilizando el Deep Learning en sus productos. En IBM estamos colaborando con los líderes en el mercado para impulsar la investigación y liderar en ese espacio.

Para comenzar con Deep Learning en Python con Watson Studio:

Existe una comunidad cada vez mayor de investigadores, ingenieros y científicos de datos que comparten un conjunto común y muy potente de herramientas, y la mayoría de ellas son de código abierto.

Una de las cosas buenas del Deep learning es que es realmente una familia de técnicas que se adapta a todo tipo de datos y todo tipo de problemas, todos utilizan una infraestructura común y un lenguaje común para describir items.

Lo que se aconseja al lector es comenzar con modelos muy simples y posteriormente comenzar con los que son más complejos y grandes. Es sencillo comenzar con tu propio ordenador ya que con IBM Watson Studio tienes todo lo que necesitas para comenzar a experimentar con las tecnologías de Deep Learning.

Las bibliotecas y tutoriales más populares de Deep Learning en Python son:

Theano: una de las bibliotecas de Deep Learning más conocidas.

→ TUTORIAL: https://dataplatform.ibm.com/analytics/notebooks/b4f6f269-6cd6-4adc-b63d-d19e5b0e90a0/view?access_token=647ed3ebaf725ffd9d4cf77fbc41066e093e15f764d5c810620a43044e362780

Tensorflow: es una biblioteca de bajo nivel que está menos madura que Theano. Sin embargo, es compatible con Google y ofrece computación distribuida lista para usar. →Tutorial:

https://dataplatform.ibm.com/analytics/notebooks/91440c8b-0fb-471e-b04e-235e4d9f510d/view?access_token=fb4380415a903111e26cec3bd95d8ba91a04746185c866fecde9d36643fa5585

Keras: Esta es nuestra biblioteca favorita de Python para Deep Learning y es el mejor lugar para comenzar si eres principiante.

→ Tutorial https://dataplatform.ibm.com/analytics/notebooks/d96fa67b-14f1-4db7-8b60-1af3c13699c3/view?access_token=c31fd96333af39811a78fe7773e421a50c7e20a450badb653bf4e0db39dc8f3f

Lasagne:

→ Tutorial

https://dataplatform.ibm.com/analytics/notebooks/c1bda39b-3fc-4dae-a109-e71d11113633/view?access_token=18379e532a9953d4e97f2a75eee37a8ece9ee4745676e1a647493fbdd7b16fb

MXNet- Es otra biblioteca de alto nivel similar a Keras. Ofrece enlaces para múltiples idiomas y soporte para computación distribuida.

→ Tutorial https://dataplatform.ibm.com/analytics/notebooks/39e93a50-cfc1-4097-b671-5261ba56e166/view?access_token=b7bd65f58805daf1f39465395dbb239c2f03d2cdeb611d8f413c81c7b1b06791

Más información sobre Deep Learning en IBM Watson Studio:

<https://medium.com/ibm-data-science-experience/deep-learning-with-data-science-experience-8478cc0f81ac>

Proponemos comenzar con el siguiente notebook, crea un proyecto o cópiale el notebook en algún proyecto que ya tengas creado, y sigue los pasos:

https://dataplatform.ibm.com/analytics/notebooks/d96fa67b-14f1-4db7-8b60-1af3c13699c3/view?access_token=c31fd96333af39811a78fe7773e421a50c7e20a450bad-b653bf4e0db39dc8f3f

Una vez el lector tenga soltura con Watson Studio y comprenda qué es el Deep learning, se recomienda este interesante proyecto: *Self Driving Car tutorials with Watson Studio*. Que se encuentra en el repositorio siguiente. Proponemos leer con atención el archivo README.md y seguir los pasos que se indican en él.

<https://github.com/arui7/Self-Driving-Car-in-WATSON STUDIO>

Workshop 5.

Recursos para trabajar en DSX en Local

Para todos aquellos que quieran aprender a utilizar DSX en Local:

https://github.com/elenalowery/DSX_Local_Workshop

ANEXO: Decision Optimization en DSX Local

Se usa con frecuencia el término optimización para referirse a hacer algo mejor. Aunque la optimización a menudo mejora las cosas, significa mucho más que eso: la optimización significa encontrar la solución más adecuada para una situación definida con precisión. Esta sofisticada tecnología, también llamada Analítica Prescriptiva, consiste en explorar una amplia gama de escenarios posibles antes de sugerir la mejor manera de responder a una situación presente o futura.

Generalmente se basa en problemas de negocios, como planificación compleja, programación, fijación de precios, inventario o administración de recursos. La analítica prescriptiva consiste en una multitud de problemas operacionales que están más allá de las capacidades del cerebro humano o del software de oficina estándar.

Para cualquier problema, se comienza a resolverlo con el modelo de optimización, que es la formulación matemática del problema que puede ser interpretada y resuelta por un motor de optimización. El modelo de optimización especifica las relaciones entre los objetivos, límites y elecciones que están involucradas en las decisiones. Pero son los datos de entrada los que hacen que estas relaciones sean concretas. Un modelo de optimización para la planificación de la producción, por ejemplo, puede tener la misma forma si está produciendo tres productos o

mil. El modelo de optimización más los datos de entrada crea una instancia de un problema de optimización.

Los motores de optimización (o solucionadores) aplican algoritmos matemáticos para encontrar una solución, un conjunto de decisiones que alcanza los mejores valores de los objetivos y respeta los límites impuestos. El motor de optimización implementa algoritmos especializados que se han desarrollado y ajustado para resolver de manera eficiente una gran variedad de problemas diferentes. Decision Optimization utiliza los motores de optimización IBM CPLEX que han demostrado ser especialmente útiles para las aplicaciones del mundo real.

Decision Optimization permite crear diferentes escenarios pues proporciona una plataforma configurable para dar soporte a los responsables de tomar las decisiones con analítica para resolver sus retos de planificación y programación. Reduce el esfuerzo, el tiempo y el riesgo asociado a la creación de soluciones personalizadas que mejoran los resultados de negocio.

Para comenzar y aprender a utilizarlo en DSX en Local:

https://github.com/jc900/FastStart_DDLabs