

IBM

# Data Science Experience



Manual para Workshop  
Febrero 2018

# Índice

## Introducción

### Workshop 1. Comenzar con DSX

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

### 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 Data Science Experience en Local

# Introducción

IBM Data Science Experience 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 Data Science Experience en Local y en Cloud. En este workshop nos centraremos en IBM Data Science Experience en Cloud, que forma parte de Watson Data Platform, que 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.

Data Science Experience 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:

DSX 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:**

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

Gracias a DSX 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 Data Science Experience puedes administrar los recursos del proyecto y la colaboración de los usuarios además de poder compartir, bifurcar y reutilizar assets con Github.

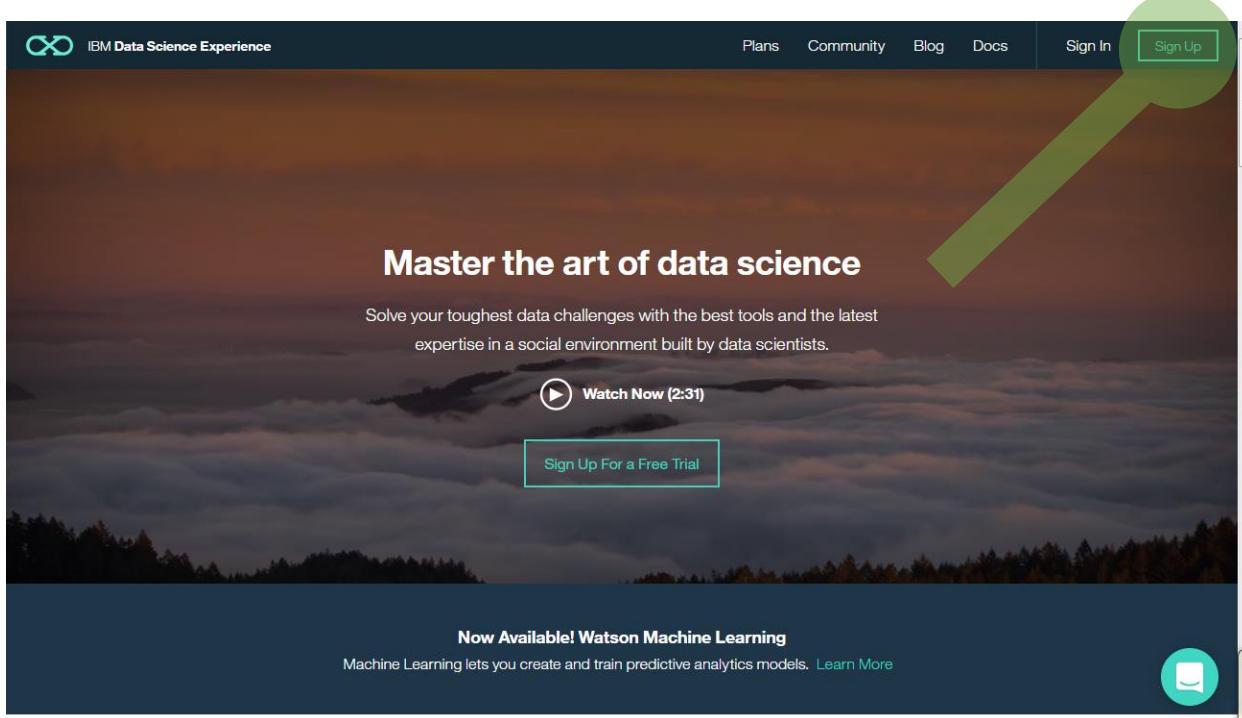
# Workshop 1.

## 1. Registrarse

IBM Data Science Experience forma parte de Watson Data Platform. Watson Data Platform es una plataforma que permite a los equipos colaborar, compartir datos y modelos y poner en producción dichos modelos.

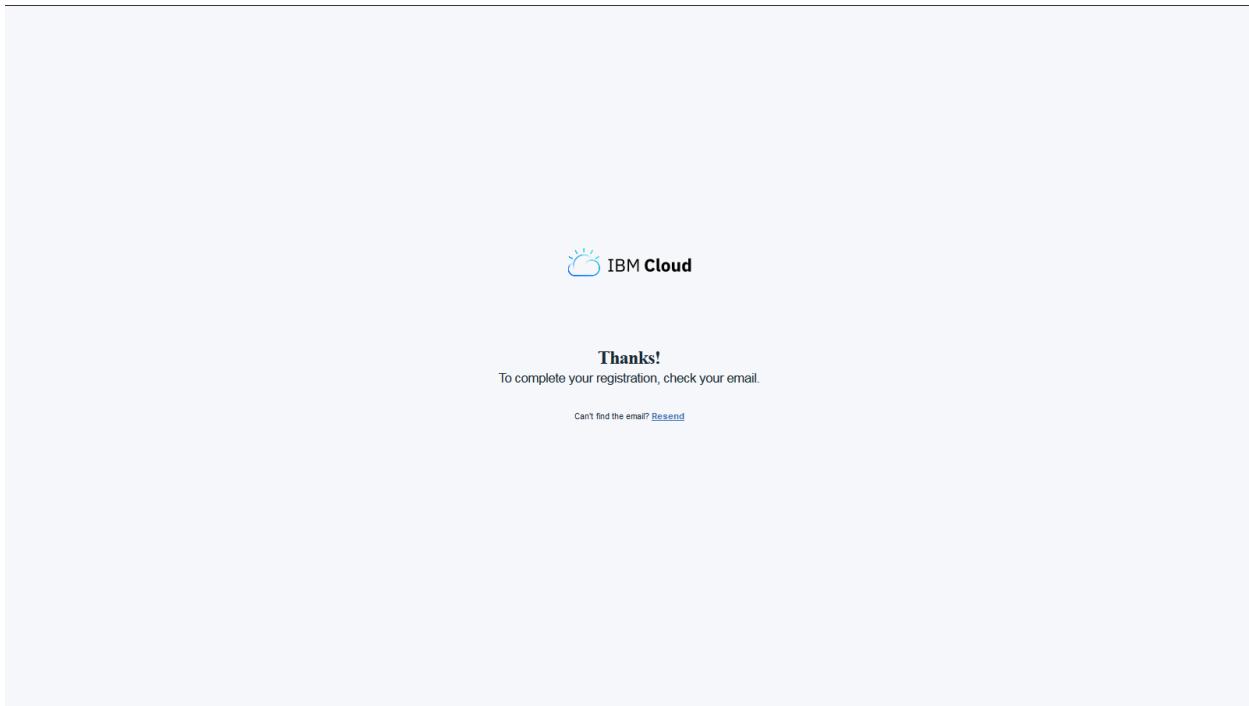
Para crearse una cuenta nueva, comienza por registrarte en Data Science Experience. Después de registrarte, puedes agregar otras aplicaciones de Watson Data Platform en cualquier momento desde la propia herramienta.

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

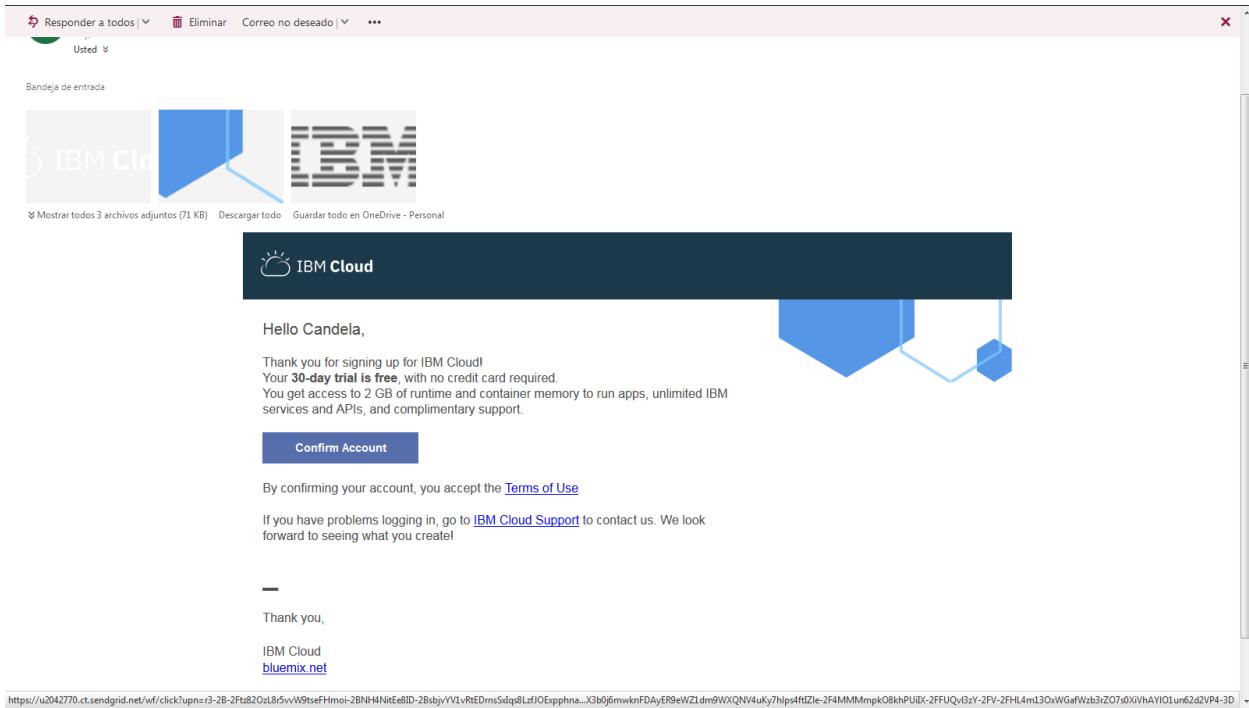


Si aún no tienes una cuenta IBMid e IBM Cloud, se crearán durante el proceso de suscripción.





Entramos al correo para confirmar la cuenta:



Y nos envía de vuelta a Data Science Experience, nos logeamos, nos pide que creamos un espacio y nombre para nuestra organización (podemos dejar lo que nos propone) y está listo para comenzar.

---

### Sign in to IBM

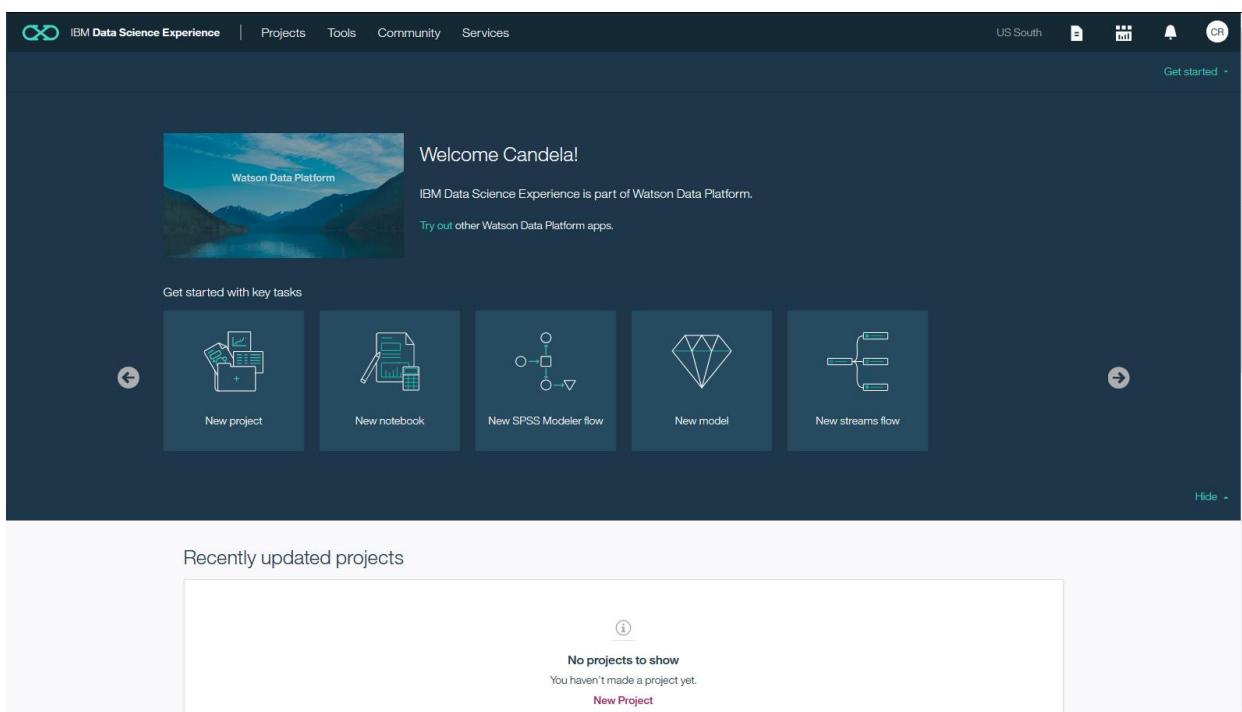
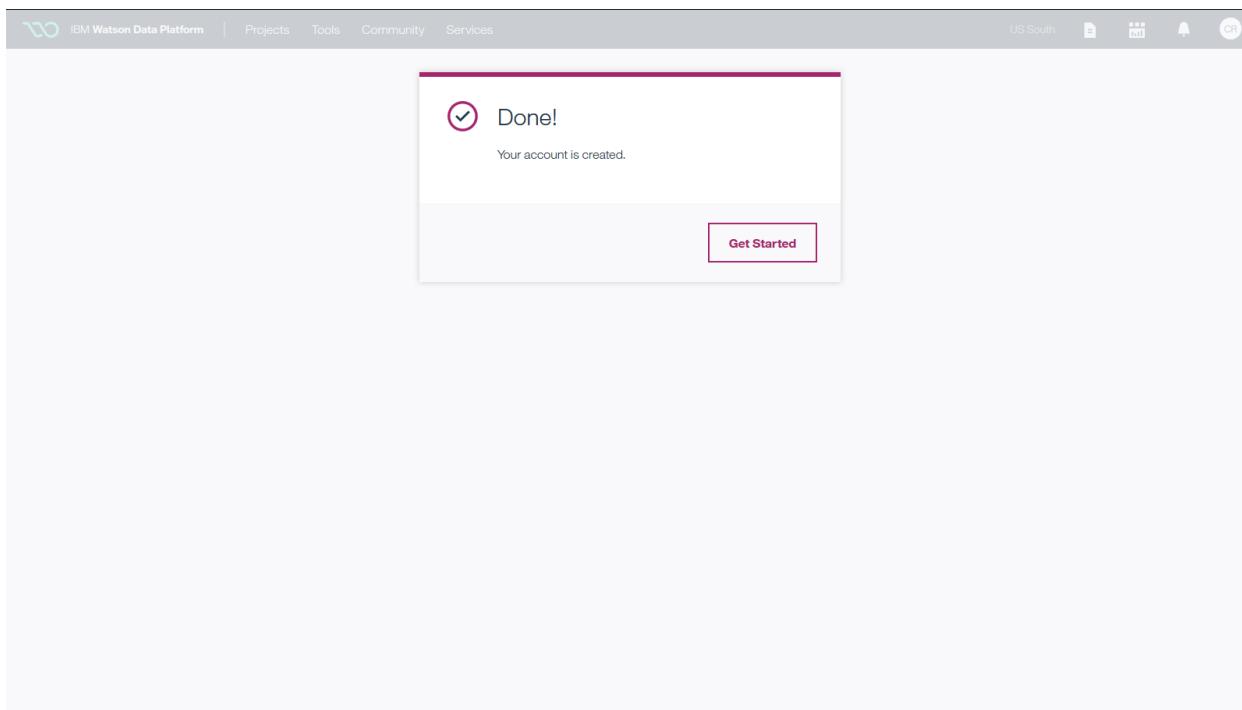
Enter IBMid or email [Forgot your IBMid?](#)

[Continue](#)

[New? Create an IBMid.](#)

<https://idaa.iam.ibm.com/idaa/mfim/sp/authsvc?PolicyId=urn:ibm:security:authentication:asf:basicldapuser>

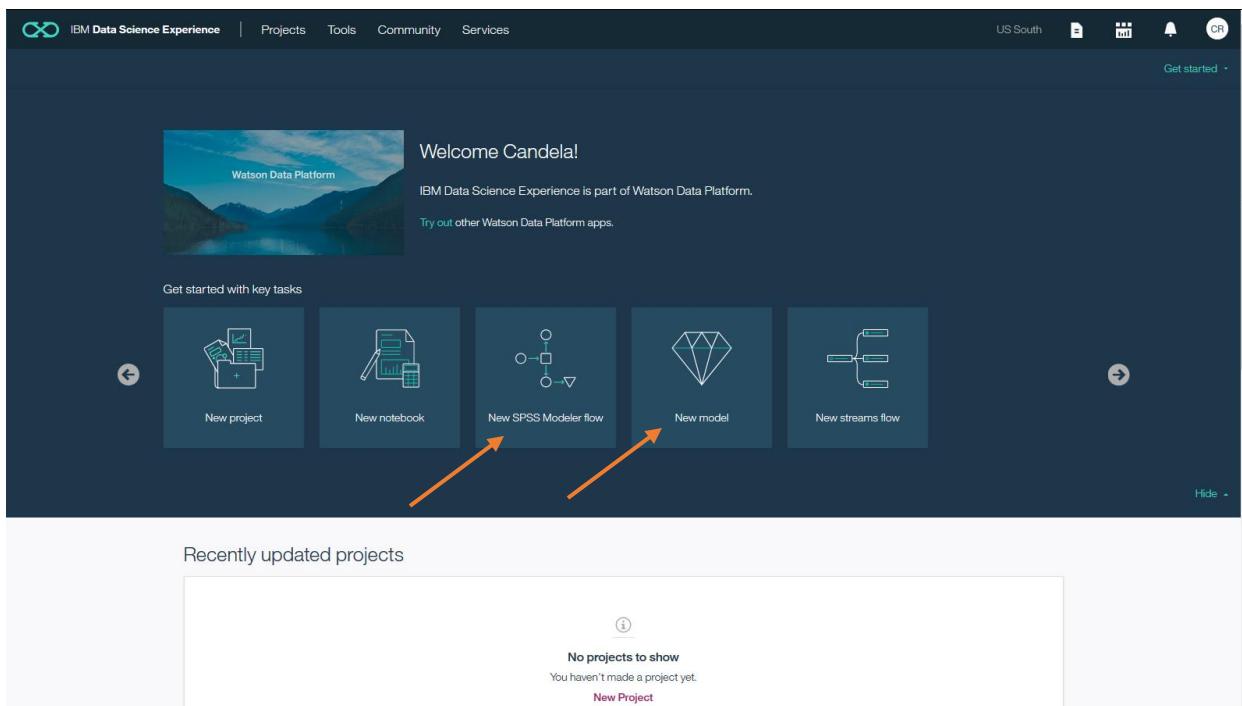
The screenshot shows the 'Create Organization and Space' page of the IBM Watson Data Platform. The URL in the address bar is <https://idaa.iam.ibm.com/idaa/mfim/sp/authsvc?PolicyId=urn:ibm:security:authentication:asf:basicldapuser>. The page header includes the IBM Watson Data Platform logo, navigation links for Projects, Tools, Community, and Services, and a user menu with 'US South', a file icon, a help icon, a notifications icon, and a profile icon. The main form is titled 'Create Organization and Space' and asks 'Name your IBM Cloud organization and space.' It contains two input fields: 'IBM Cloud Organization' with the value 'candela\_retolaza@hotmail.com\_us-south' and 'IBM Cloud Space' with the value 'WatsonDataPlatform'. A 'Continue' button is located at the bottom right of the form.



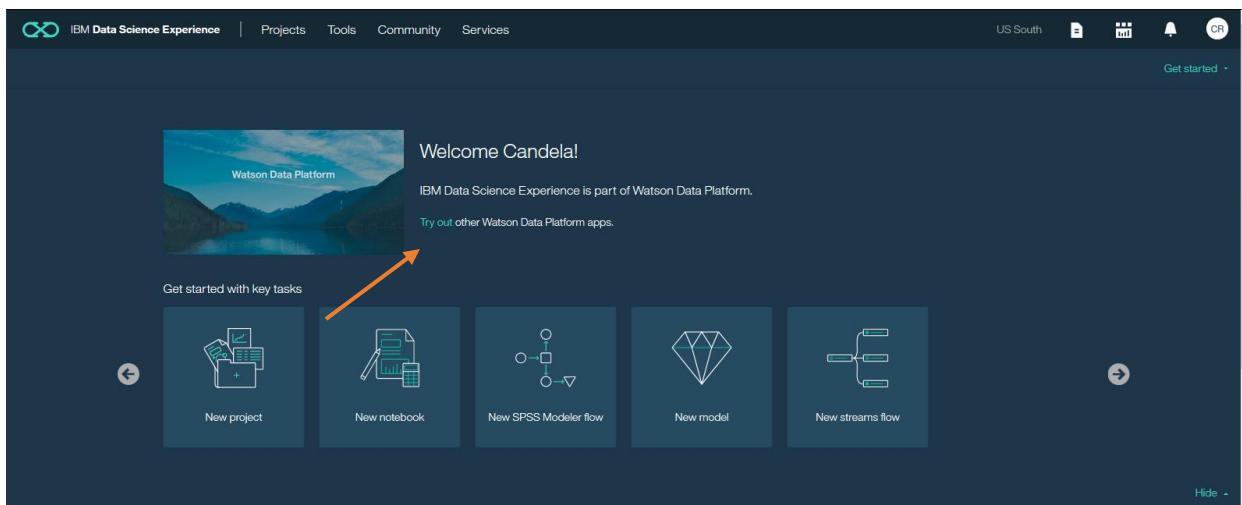
Nota: asegúrate de permanecer con tu navegador predeterminado durante el proceso de inicio de sesión. Si te encuentras en IBM Cloud Dashboard, simplemente regresa a la página de registro y sigue el enlace que indica que ya tienes una cuenta.

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

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.



Antes de pasar al siguiente apartado, asegúrate de que tienes todas las herramientas. Puedes confirmarlo clickando en **Try out**. Debe aparecer únicamente Data Refinery y Data Catalog. Si aparece Data Science Experience, simplemente pincha en **try for free** y créate una versión lite, y estará listo para comenzar.



Welcome Candela!

IBM Data Science Experience is part of Watson Data Platform.

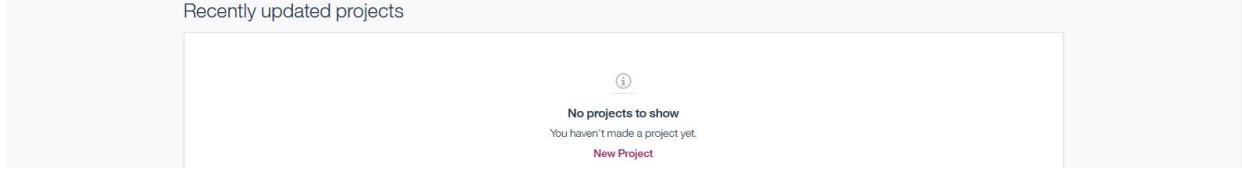
Try out other Watson Data Platform apps.

Get started with key tasks

- New project
- New notebook
- New SPSS Modeler flow
- New model
- New streams flow

Hide ▾

A red arrow points from the text "Try out other Watson Data Platform apps." to the "New notebook" button.

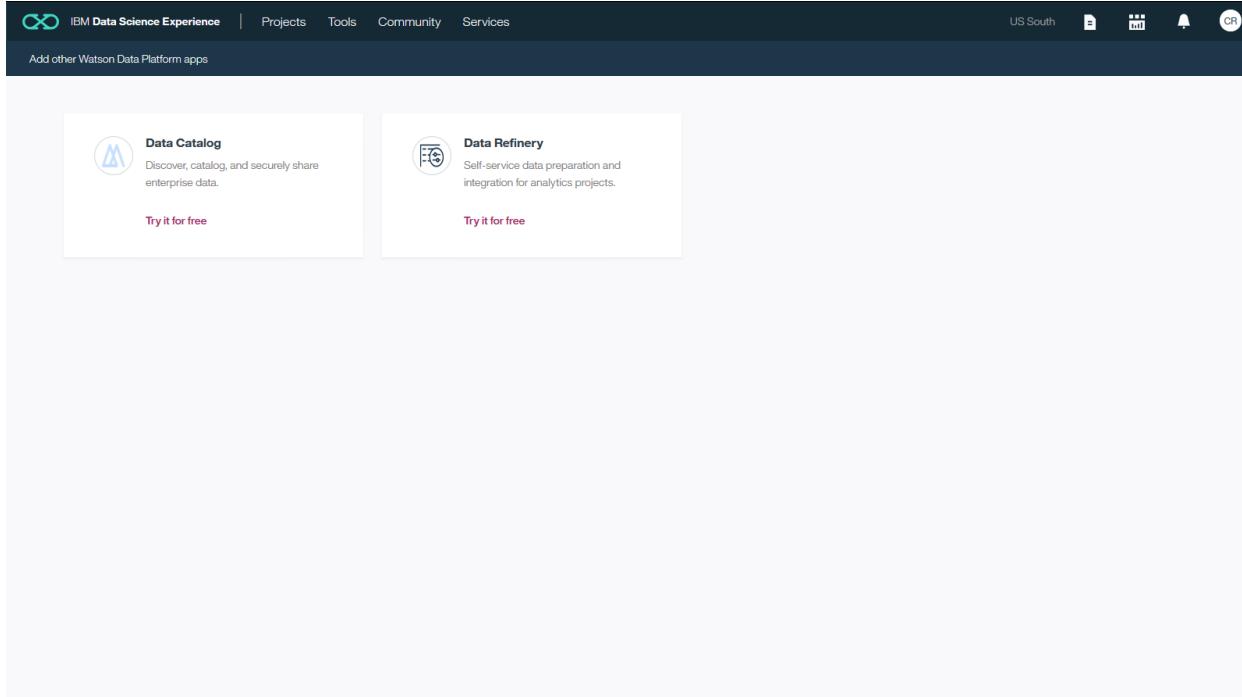


Recently updated projects

No projects to show

You haven't made a project yet.

New Project



IBM Data Science Experience | Projects Tools Community Services

Add other Watson Data Platform apps

US South

**Data Catalog**  
Discover, catalog, and securely share enterprise data.  
[Try it for free](#)

**Data Refinery**  
Self-service data preparation and integration for analytics projects.  
[Try it for free](#)

## 2. Añadir usuarios adicionales (opcional)

Las aplicaciones de Watson Data Platform, IBM Data Science Experience, Data Catalog 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.

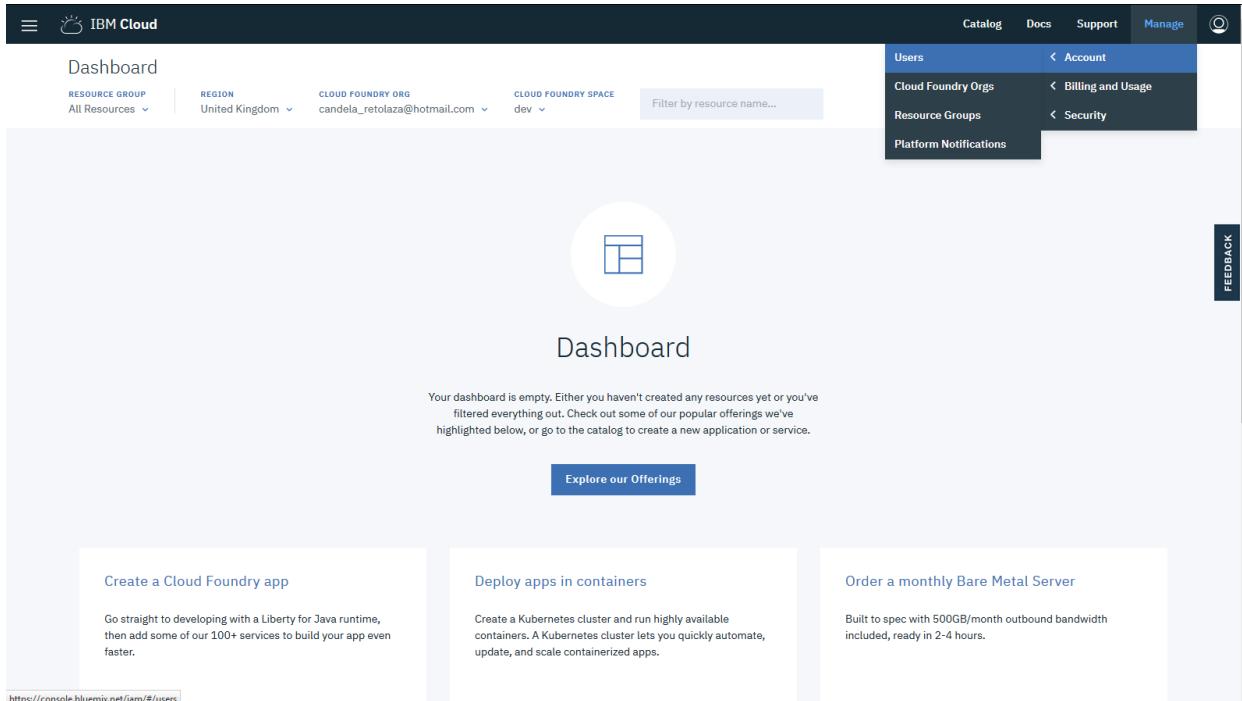
Lo vemos detalladamente:

### 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.

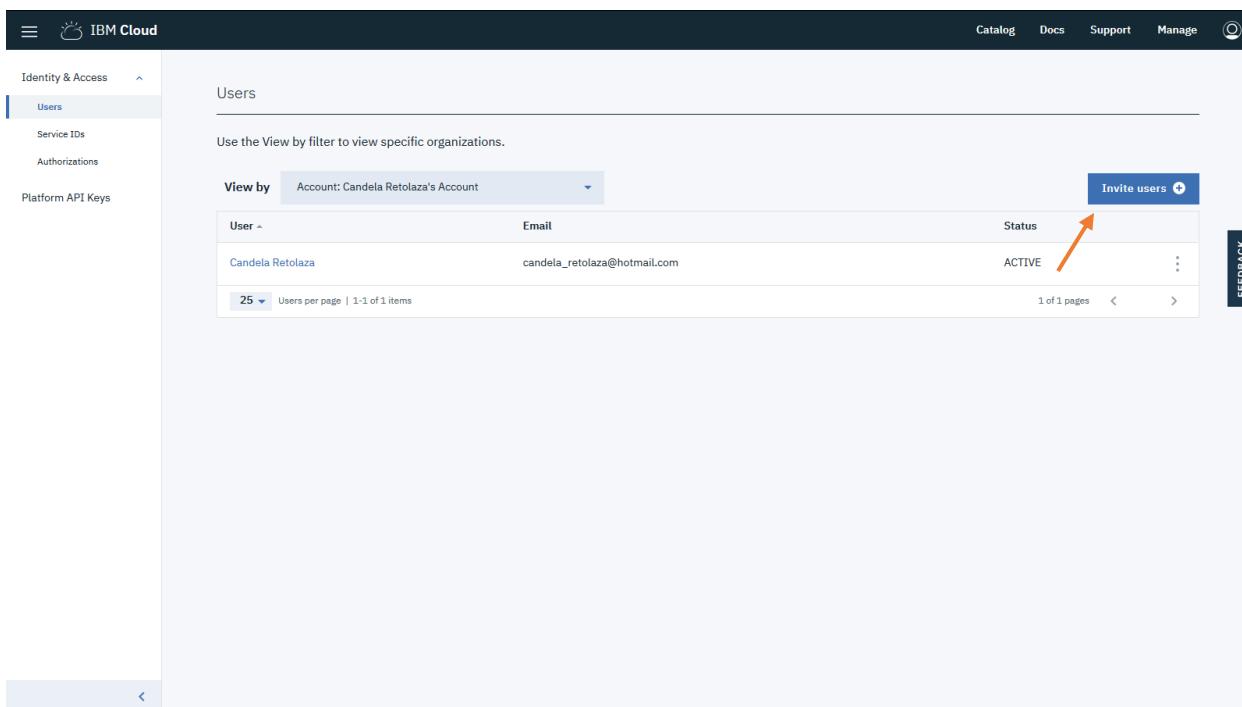
Entra en <https://console.bluemix.net/>

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 in blue, indicating it is selected. Other options in the 'Manage' dropdown include 'Cloud Foundry Orgs', 'Resource Groups', and 'Platform Notifications'. The main dashboard area is empty, showing a placeholder for creating a Cloud Foundry app.

En la página **Users**, haz clic en **Invitar usuarios**.



The screenshot shows the 'Users' page within the 'Identity & Access' section of the IBM Cloud console. A single user, 'Candela Retolaza', is listed in the table. The 'Status' column shows 'ACTIVE'. An orange arrow points to the 'Invite users' button in the top right corner of the table header. The table also includes columns for 'User' and 'Email'.

Identity & Access

Users

Service IDs

Authorizations

Platform API Keys

Invite users to Candela Retolaza's Account

Users

You can invite one or more users. All users invited together are granted the same access.

Email address

Enter up to 100 email addresses separated by commas, spaces, or line breaks.

Access

Assign access for all invited users. The default is no access.

Services

Cloud Foundry access

You can add users to one Cloud Foundry organization with this invite. You can edit this access or assign additional access after the user accepts the invite. All fields are required.

Organization

No Cloud Foundry access

Cancel

Invite users

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

Identity & Access

Users

Service IDs

Authorizations

Platform API Keys

Invite users to Candela Retolaza's Account

Users

You can invite one or more users. All users invited together are granted the same access.

Email address

candela016@hotmail.com  
candela.retolaza@es.ibm.com

Access

Assign access for all invited users. The default is no access.

Services

Assign access to

Resource

Services

Cloud Object Storage

Service instance

All service instances

Resource type

Cancel

Invite users

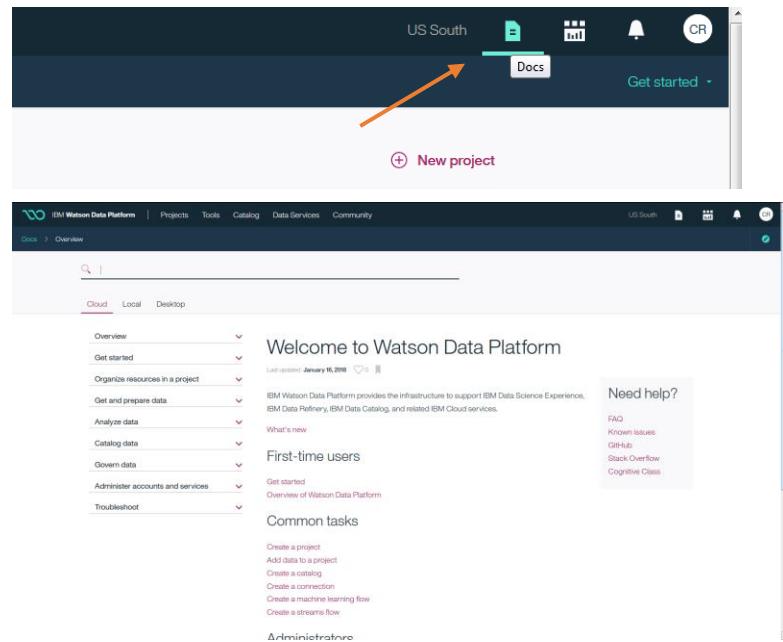
En la sección Acceso de la página Invitar usuarios, expanda el acceso a Cloud Foundry y seleccione la Organización a la que está agregando usuario.

- 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 DSX.

**NOTA: Donde encontrar la ayuda y documentación de Data Science Experience**

**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

Identity & Access

Users

Service IDs

Authorizations

Platform API Keys

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 DSX

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

IBM Data Science Experience

Watson Data Platform

Welcome Candela!

IBM Data Science Experience is part of Watson Data Platform.

Try out other Watson Data Platform apps.

Get started

Get started with key tasks

New project

New notebook

New SPSS Modeler flow

New model

New streamflow

Recently updated projects

No projects to show

You haven't made a project yet.

New Project

Para empezar a utilizar Data Science Experience 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.

¿Necesitas inspiración? Haz clic en el botón **Comunidad** en tu Data Science Experience 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.

NAME	ROLE	COLLABORATORS	DATE CREATED	LAST UPDATED
Demos DSX	Admin	2	Sep 25, 2017	Jan 09, 2018
Machine Learning, Data Science	Admin	3	Oct 06, 2017	Dec 22, 2017
Demo DSX	Admin	1	Nov 21, 2017	Dec 21, 2017

Creamos un proyecto nuevo, por ejemplo: Workshop, añadimos una descripción opcional: por ejemplo, Workshop DSX.

Añadimos un almacenamiento y un servicio de spark:

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

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

Plan	Features	Pricing
<input checked="" type="radio"/> <b>Lite</b>	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 <small>Applies to aggregate total across all storage bucket classes</small>	Free

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.

**Create**

Confirm Creation

Plan
<b>Lite</b>

Resource group  
**Default**

Service name  
**cloud-object-storage-ig**

**Confirm**

Confirmamos, y actualizamos pinchando en **Refresh**:

New project

Define project details

Name  
Workshop

Description  
Project description

Choose project options

Restrict who can be a collaborator  ⓘ

Define storage

① Select storage service  
**Add**  
Add an object storage instance and then return to this page and click Refresh.

② Refresh

Define compute engine

① Select Spark service  
Spark service  
**Add**  
Add IBM Analytics for Apache Spark, then return to this page and click Refresh.

② Refresh

**Create**

Ya hemos creado el almacenamiento. Ahora añadimos un servicio de Spark:

New project

Define project details

Name  
Workshop

Description

Choose project options

Restrict who can be a collaborator  ⓘ

Define storage

② Select storage service  
Target Cloud Object Storage Instance  
**cloud-object-storage-ig**

Define compute engine

① Select Spark service  
Spark service  
**Add**  
Add IBM Analytics for Apache Spark, then return to this page and click Refresh.

② Refresh

**Create**

De igual manera, creamos una versión lite 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
Lite	2 Spark Executors	Free

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

Terms Cancel **Create**

**Confirm Creation**

Organization: candela\_retolaza@hotmail.com\_us-south

Plan  
Lite

Space  
WatsonDataPlatform

Service name  
Spark-sc

Cancel **Confirm**

Terms Cancel **Create**

New project

Define project details

Name: Workshop

Description: Project description

Choose project options

Restrict who can be a collaborator  ⓘ

Define storage

✓ Select storage service

Target Cloud Object Storage Instance: **cloud-object-storage-ig**

Define compute engine

① Select Spark service

Spark service: **Add**

Add IBM Analytics for Apache Spark, then return to this page and click Refresh.

② Refresh

Cancel Create

Confirmamos, y actualizamos pinchando en **Refresh** y listo. Ya podemos crear un proyecto.

New project

Define project details

Name: Workshop

Description: Project description

Choose project options

Restrict who can be a collaborator  ⓘ

Define storage

✓ Select storage service

Target Cloud Object Storage Instance: **cloud-object-storage-ig**

Define compute engine

✓ Select Spark service

Spark service: **Spark-ke**

⚠ If you associate the same Spark service with multiple projects, the Spark history server will display job history information for all the projects.

Cancel Create

The screenshot shows the IBM Data Science Experience (DSX) interface. At the top, there is a navigation bar with the DSX logo, 'IBM Data Science Experience', and links for 'Projects', 'Tools', 'Community', and 'Services'. On the right side of the top bar are icons for 'US South', a file, a bar chart, a bell, and a user profile. Below the top bar, the main header 'My Projects' and 'Workshop' are displayed, along with a 'Last Updated: Feb 12 2018' timestamp. To the right of the header are buttons for 'Add to project', a user profile, and several small icons. The main content area is titled 'Workshop' and shows summary statistics: 0 Assets, 0 Bookmarks, and 1 Collaborators. Below these stats, there are sections for 'Date created' (Feb 12 2018), 'Description' (No description available), 'Storage' (0% of 5 GB used), 'Collaborators' (Candela Retolaza, Admin), and 'Bookmarks' (0). To the right, there is a 'Recent activity' section with a placeholder message: 'Alerts related to this project will show here when the project is active.' The interface has a clean, modern design with a light gray background and blue accents for links.

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.

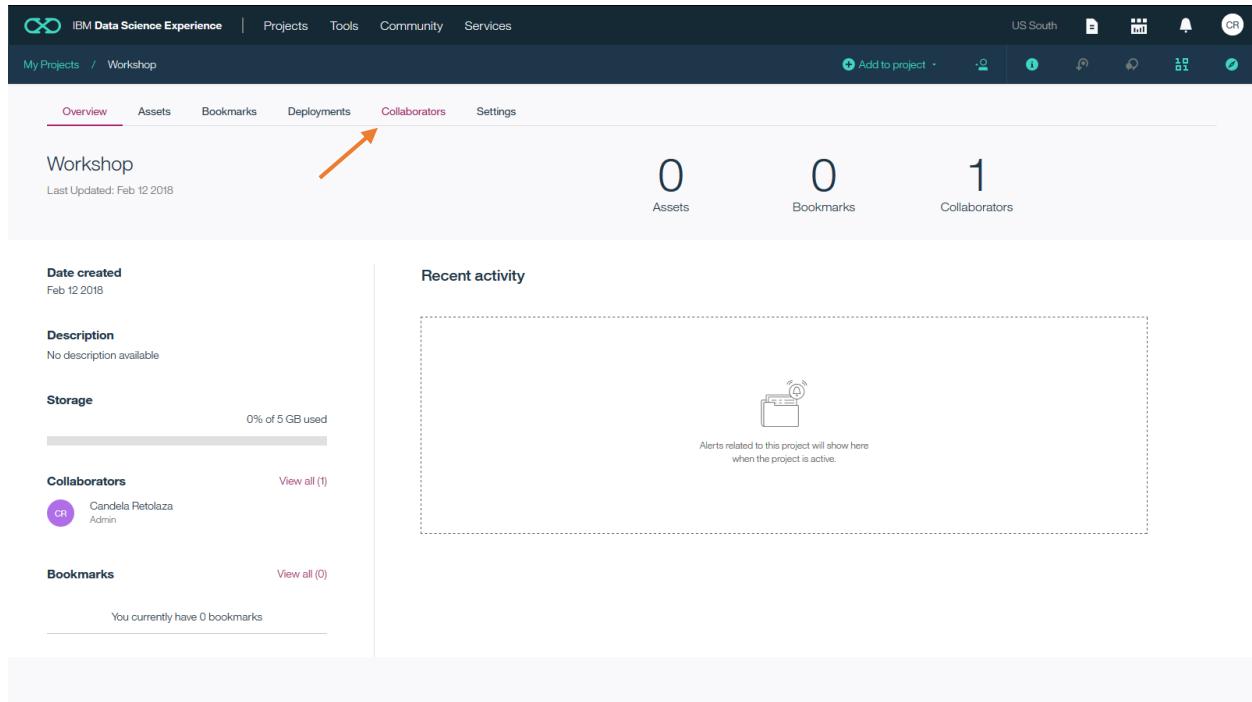
Un proyecto sirve para organizar tus recursos para trabajar y hacer minería de datos. Los recursos de su proyecto pueden incluir:

Notebooks, Modelos, Flows de SPSS, Streams flows, 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.

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.



The screenshot shows the IBM Data Science Experience (DSX) interface. At the top, there is a navigation bar with the DSX logo, 'IBM Data Science Experience', and links for 'Projects', 'Tools', 'Community', and 'Services'. On the far right of the top bar are icons for 'US South', a file, a bar chart, a bell, and a user profile. Below the top bar, the page title is 'My Projects / Workshop'. The main content area has a dark header with tabs: 'Overview' (highlighted in purple), 'Assets', 'Bookmarks', 'Deployments', 'Collaborators' (which has an orange arrow pointing to it), and 'Settings'. The 'Overview' section for the 'Workshop' project shows the following statistics: 0 Assets, 0 Bookmarks, and 1 Collaborator. Below these stats, there is a 'Recent activity' section with a placeholder message: 'Alerts related to this project will show here when the project is active.' On the left side of the main content area, there are sections for 'Date created' (Feb 12 2018), 'Description' (No description available), 'Storage' (0% of 5 GB used), 'Collaborators' (Candela Retolaza, Admin), and 'Bookmarks' (You currently have 0 bookmarks). At the bottom of the main content area, there is a large, empty light-gray box.

My Projects > Workshop

Overview Assets Bookmarks Deployments Collaborators Settings

Find collaborators

**Collaborators**

NAME	EMAIL	PERMISSION	STATUS	ACTIONS
cc 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.

My Projects > Workshop

Overview Assets Bookmarks Deployments Collaborators Settings

Find collaborators

**Collaborators**

NAME	EMAIL	PERMISSION	STATUS	ACTIONS
cc Candela Retolaza Conde	candela016@hotmail.com	Admin (1)	Active	⋮

**Add collaborators**

Workshop

Invite

candela016@hotmail.com Add more people...

Access level

Select an option

Viewer

Editor

Admin

Add

Cancel

Invite

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

Y damos a invitar:

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	candela_retolaza@es.ibm.com	Admin	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, flujos de aprendizaje automático, modelos, modelos de SPSS, Streams flows.

Para agregar flujos de datos a un proyecto, debe seleccionar la herramienta de Refinería de datos y comenzar a limpiar y dar forma a los datos en los activos de datos en su proyecto.

Si tiene permisos de administrador en un proyecto, puede 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.

**Project information**

Project name: Workshop

Description: Project description

Storage: 0 Byte Used (0% of 5 GB used)

**Save**

**Storage**

Type: Cloud Object Storage (Beta) Bucket Name: workshop984b65fe9c4b427392cac90bb0c561aa

0 Byte Used (0% of 5 GB used)

**Associated services**

NAME	SERVICE TYPE	PLAN	ACTIONS
Spark-ae	Spark	Personal	⋮

**Access tokens**

NAME	ROLE	CREATED	LAST USED	ACTIONS
you currently have no access tokens				

**Connect to a GitHub repository**

Repository URL: `https://github.com/owner/repository-name`

**Connect**

The screenshot shows the IBM Watson Data Platform Workshop interface. At the top, there is a navigation bar with links for Projects, Tools, Catalog, Data Services, and Community. The top right corner shows 'US South' and various system icons. The main content area is titled 'My Projects > Workshop'. It displays a table with one row: 'Spark-ae' (Name), 'Spark' (Role), 'Personal' (Scope), and a 'More' (Actions) button. Below this is a section titled 'Access tokens' with a 'New token' button. A message indicates 'you currently have no access tokens'. The next section is 'Connect to a GitHub repository', featuring a 'Repository URL' input field containing 'https://github.com/owner/repository-name' and a 'Connect' button. The final section is 'Project scope', which lists 'IBM Cloud account' as 'Candela Retolaza Conde's Account' and 'Restrict who can be a collaborator' as 'Yes'.

# Workshop 3.

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

## 1. Añadir datos

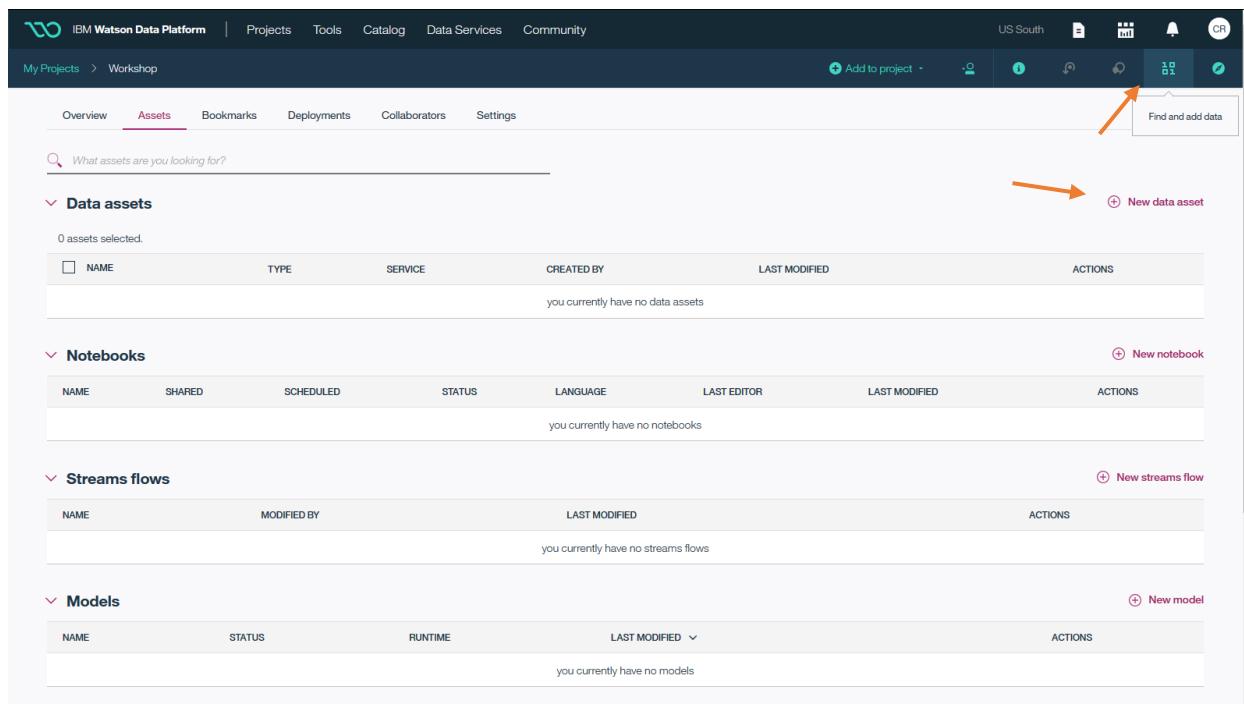
Después de crear un proyecto, tenemos que añadir datos para poder trabajar con los mismos. Todos los colaboradores en el proyecto están autorizados automáticamente para acceder a los datos en el proyecto.

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.

<input type="checkbox"/> NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
you currently have no data assets					

**▼ Notebooks**

0 assets selected.

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

**▼ Streams flows**

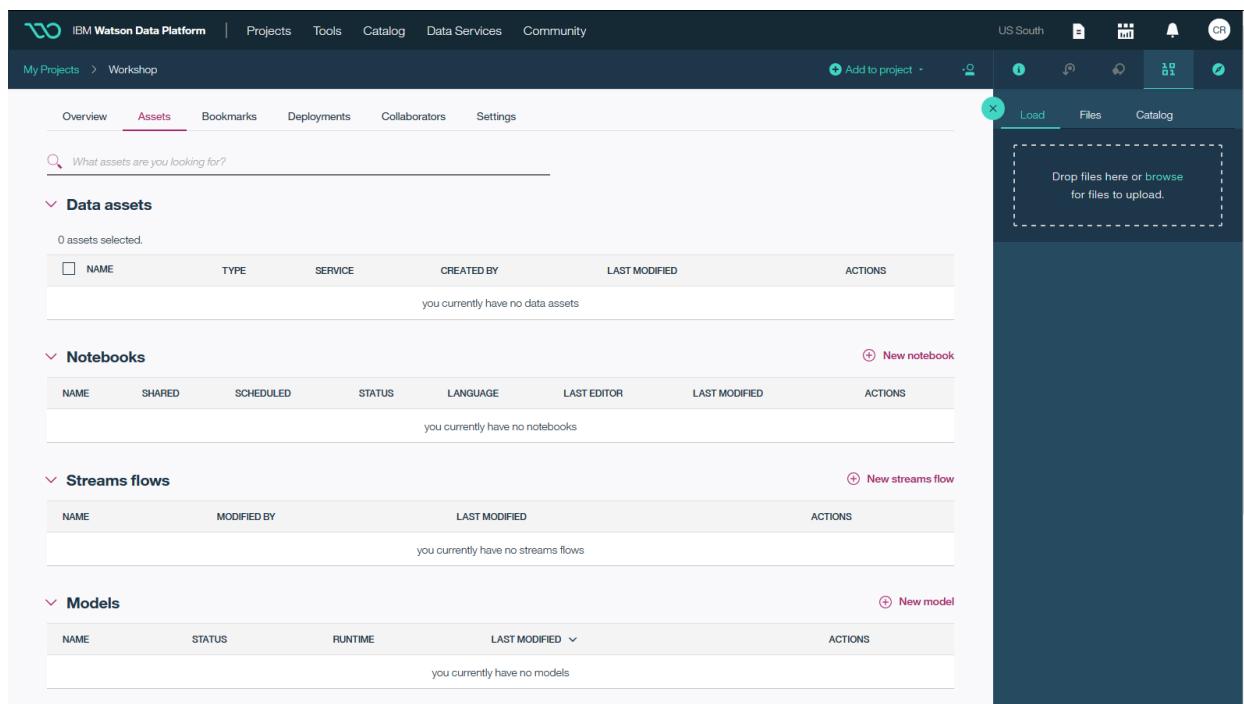
0 assets selected.

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

**▼ Models**

0 assets selected.

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.

<input type="checkbox"/> NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
you currently have no data assets					

**▼ Notebooks**

0 assets selected.

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

**▼ Streams flows**

0 assets selected.

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

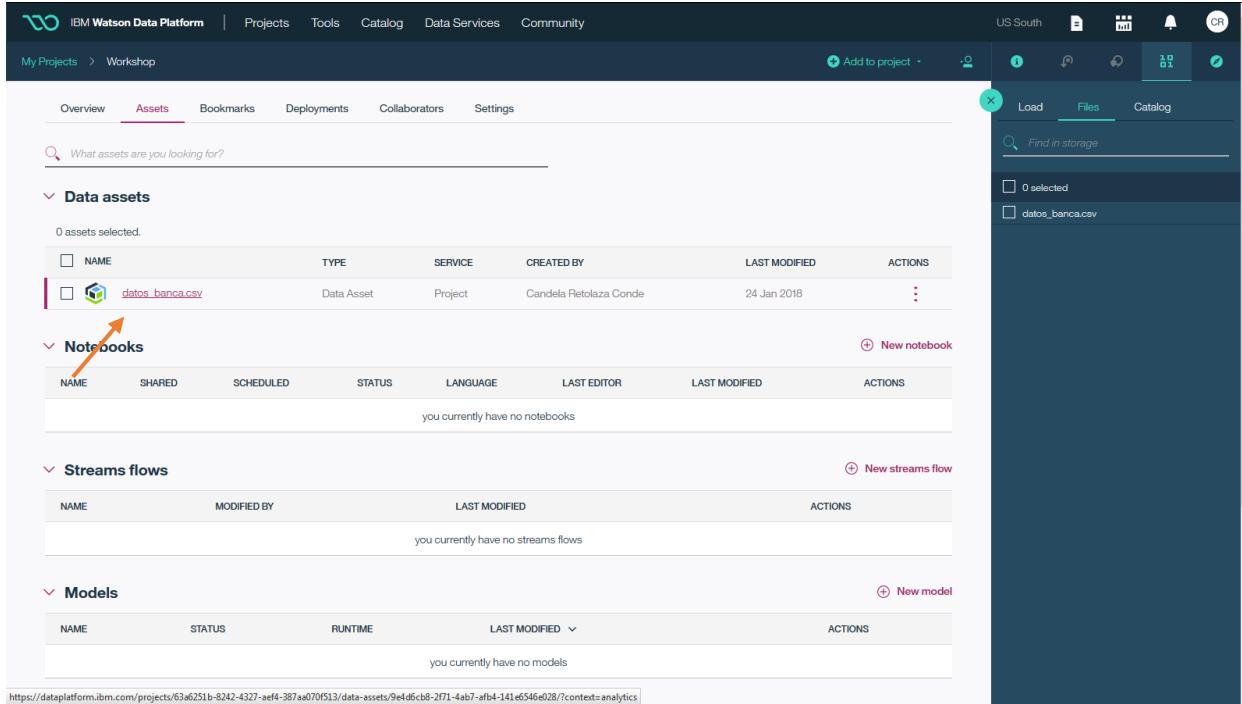
**▼ Models**

0 assets selected.

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.

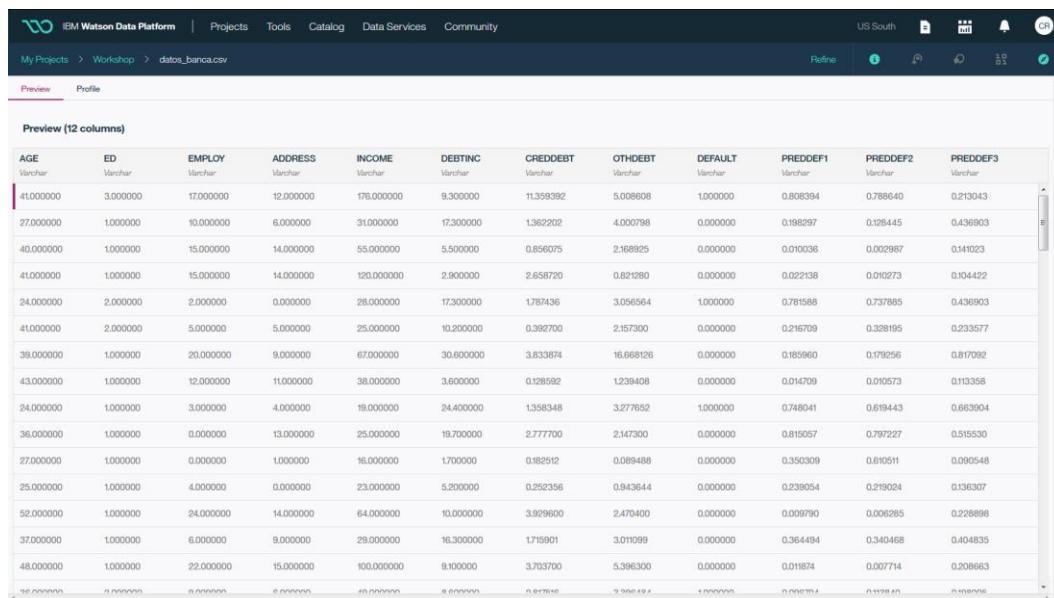
Subimos el fichero datos\_banca.csv. Debes permanecer en la página hasta que la carga esté completa. Puede cancelar un proceso de carga en curso si desea dejar de cargar un archivo.



The screenshot shows the 'Assets' section of the IBM Watson Data Platform. The 'Data assets' section lists a single item: 'datos\_banca.csv' (Data Asset, Project, Created by Candelaria Retolaza Conde, Last modified 24 Jan 2018). Below this is the 'Notebooks' section, which is currently empty. A red arrow points to the 'datos\_banca.csv' entry in the 'Data 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' section for the 'datos\_banca.csv' file. The preview table has 12 columns: AGE, ED, EMPLOY, ADDRESS, INCOME, DEBTINC, CREDDEBT, OTHDEBT, DEFAULT, PREDEF1, PREDEF2, and PREDEF3. The first row of data is highlighted with a red border. A red arrow points to the first row of the preview table.

AGE	ED	EMPLOY	ADDRESS	INCOME	DEBTINC	CREDDEBT	OTHDEBT	DEFAULT	PREDEF1	PREDEF2	PREDEF3
41.000000	3.000000	17.000000	12.000000	176.000000	9.300000	11.359392	5.006608	1.000000	0.808394	0.788640	0.213043
27.000000	1.000000	10.000000	6.000000	31.000000	17.300000	1.362202	4.006798	0.000000	0.198297	0.128445	0.436903
40.000000	1.000000	15.000000	14.000000	55.000000	5.500000	0.856075	2.168925	0.000000	0.010306	0.002987	0.141023
41.000000	1.000000	15.000000	14.000000	120.000000	2.900000	2.658720	0.821280	0.000000	0.022138	0.010273	0.104422
24.000000	2.000000	2.000000	0.000000	28.000000	17.300000	1.787436	3.050564	1.000000	0.781588	0.737885	0.436903
41.000000	2.000000	5.000000	5.000000	25.000000	10.200000	0.392700	2.157300	0.000000	0.216709	0.328195	0.233577
39.000000	1.000000	20.000000	9.000000	67.000000	30.600000	3.833874	16.668126	0.000000	0.185960	0.179256	0.817092
43.000000	1.000000	12.000000	11.000000	38.000000	3.600000	0.128592	12.39408	0.000000	0.014709	0.010573	0.113358
24.000000	1.000000	3.000000	4.000000	19.000000	24.400000	1.358348	3.277652	1.000000	0.748041	0.619443	0.663904
36.000000	1.000000	0.000000	13.000000	25.000000	19.700000	2.777700	2.147300	0.000000	0.815057	0.797227	0.315530
27.000000	1.000000	0.000000	1.000000	16.000000	1.700000	0.182512	0.0869488	0.000000	0.350309	0.610511	0.090548
25.000000	1.000000	4.000000	0.000000	23.000000	5.200000	0.252356	0.943644	0.000000	0.238054	0.219204	0.136307
52.000000	1.000000	24.000000	14.000000	64.000000	10.000000	3.929600	2.470400	0.000000	0.009790	0.006285	0.226898
37.000000	1.000000	6.000000	9.000000	29.000000	16.300000	1.715901	3.010999	0.000000	0.364494	0.340468	0.104833
48.000000	1.000000	22.000000	15.000000	100.000000	9.100000	3.703700	5.396300	0.000000	0.011874	0.007714	0.208663
36.000000	0.000000	6.000000	6.000000	49.000000	8.000000	0.187602	9.296192	4.000000	0.000000	0.000000	0.000000

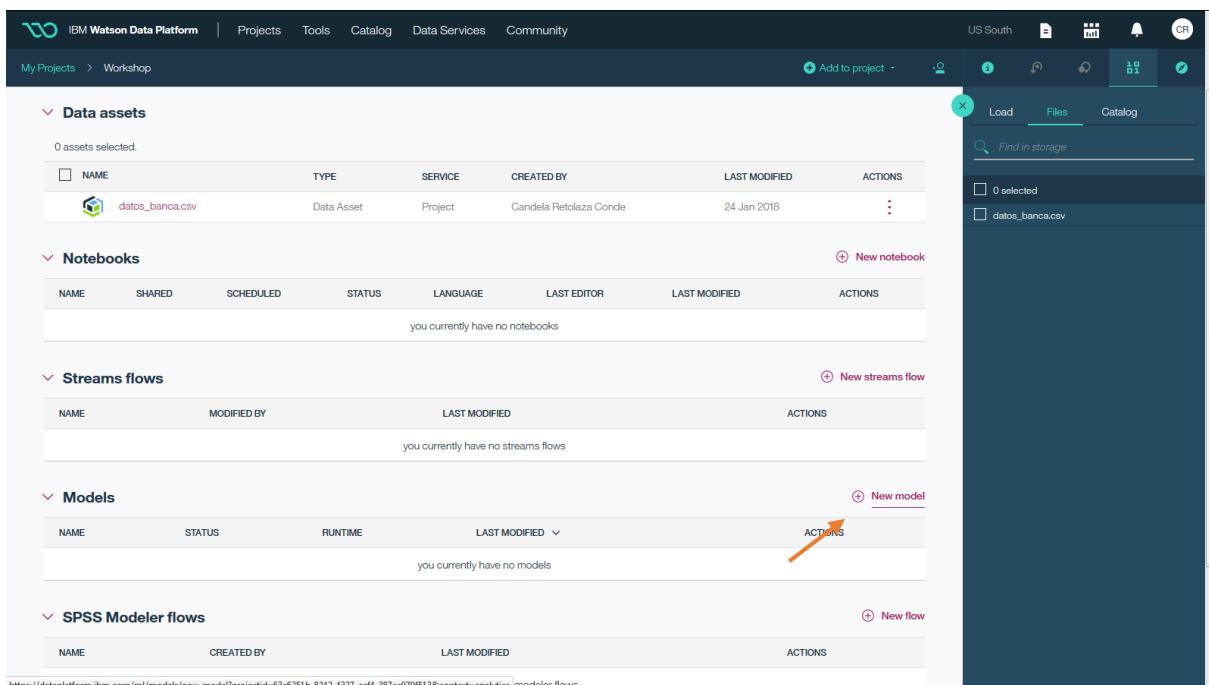
El fichero que acabamos de subir a nuestro proyecto contiene información de demográfica de clientes de un banco ficticio, con datos como: edad, nivel educativo, años en el trabajo actual, años en la misma vivienda, salario... en el que, además, tenemos información sobre los créditos que tiene cada cliente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

Podremos refinar los datos desde Data Renifery o haciendo click en **Refine** (aún en BETA).

## 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**.



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'. The 'Projects' tab is selected, showing 'My Projects' and 'Workshop'. The main content area is divided into sections: 'Data assets', 'Notebooks', 'Streams flows', 'Models', and 'SPSS Modeler flows'. The 'Models' section is currently active, displaying a table with columns: NAME, STATUS, RUNTIME, LAST MODIFIED, and ACTIONS. A red arrow points to the '+ New model' button in the ACTIONS column. The right side of the interface features a sidebar with 'Load', 'Files' (selected), and 'Catalog' tabs, and a search bar for 'Find in storage'.

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.

New model BETA

**Define model details**

Name  
Modelo Predictivo Impago

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 sample

Spark Service  
Spark-ae

**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](#).

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**

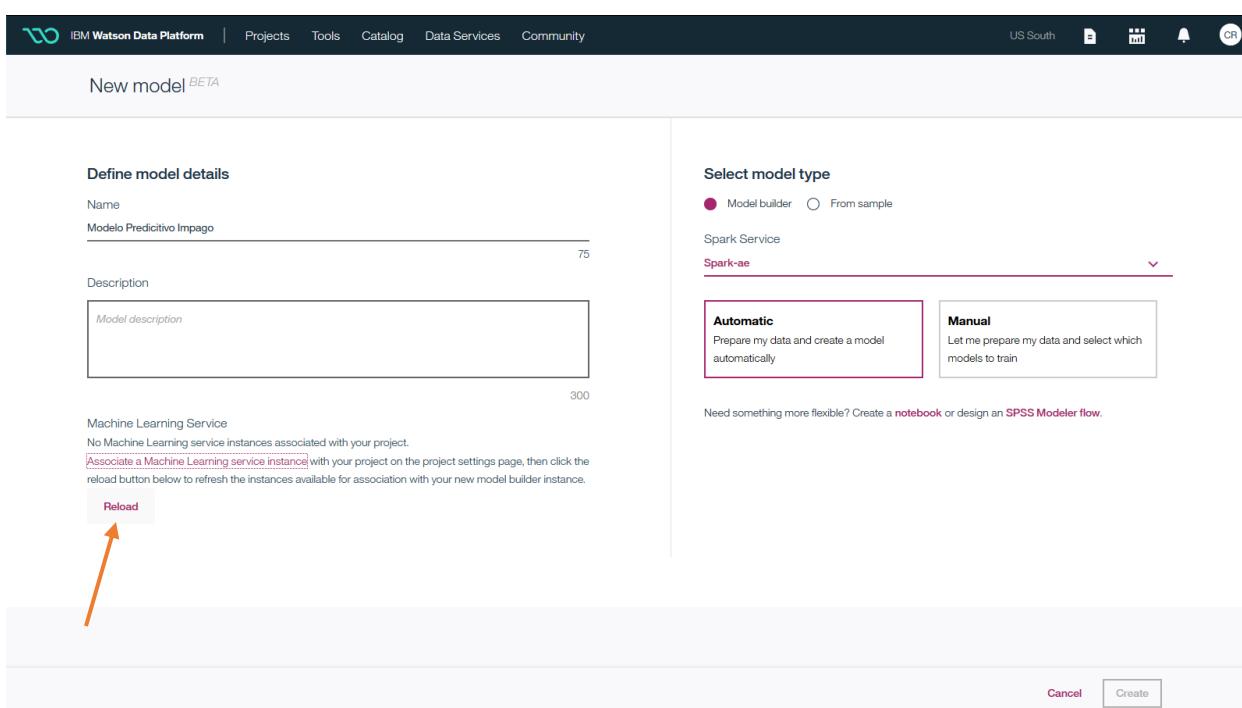
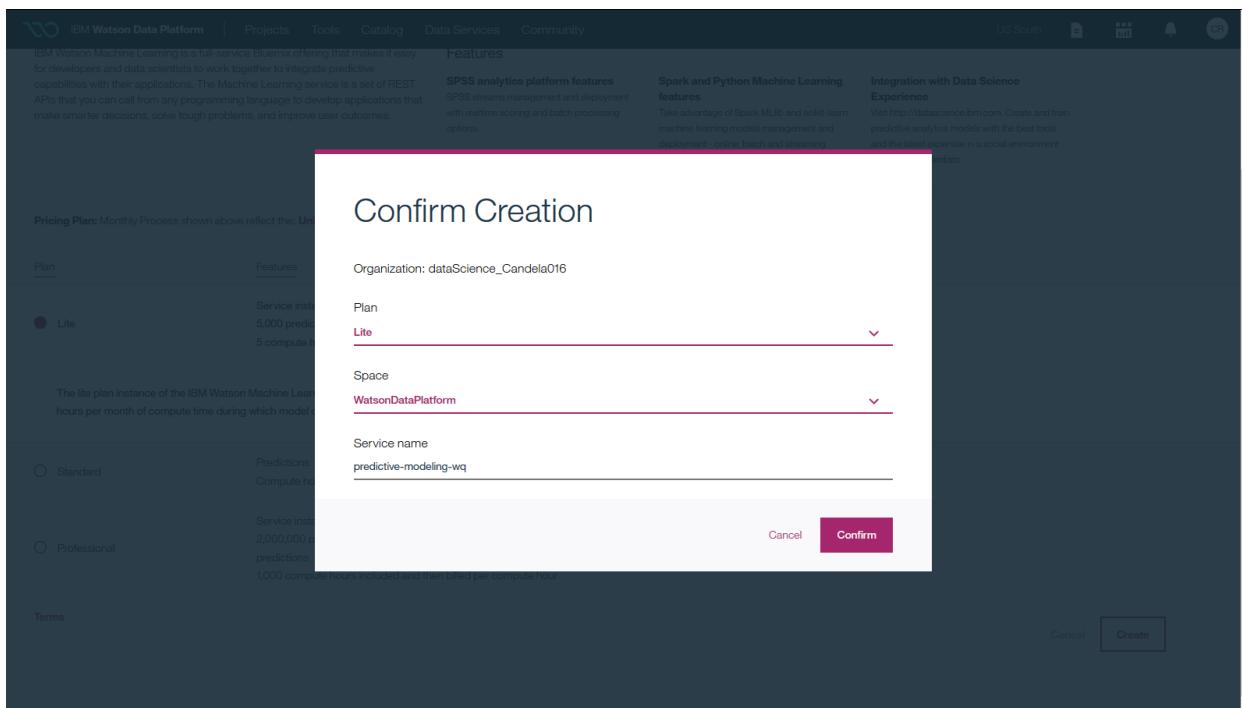
**SPSS analytics platform features**  
SPSS streams management and deployment with realtime scoring and batch processing options.

**Spark and Python Machine Learning features**  
Take advantage of Spark MLlib and scikit-learn machine learning models management and deployment - online, batch and streaming.

**Integration with Data Science Experience**  
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/instance



Si hacemos **Reload**, tendremos nuestro servicio de Machine Learning listo para usar:

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

IBM Watson Data Platform | Projects Tools Catalog Data Services Community

My Projects > Workshop > Modelo Predictivo Impago

US South

Select Data

Train Evaluate

What asset are you looking for?

NAME	TYPE	SERVICE
datos_banca.csv	Data Asset	Project

Click to preview data

Add Data Assets

Close Next

IBM Watson Data Platform | Projects Tools Catalog Data Services Community

My Projects > Workshop > Modelo Predictivo Impago

US South

Select Data

Train Evaluate

What asset are you looking for?

>Loading data

Status: loading data set...

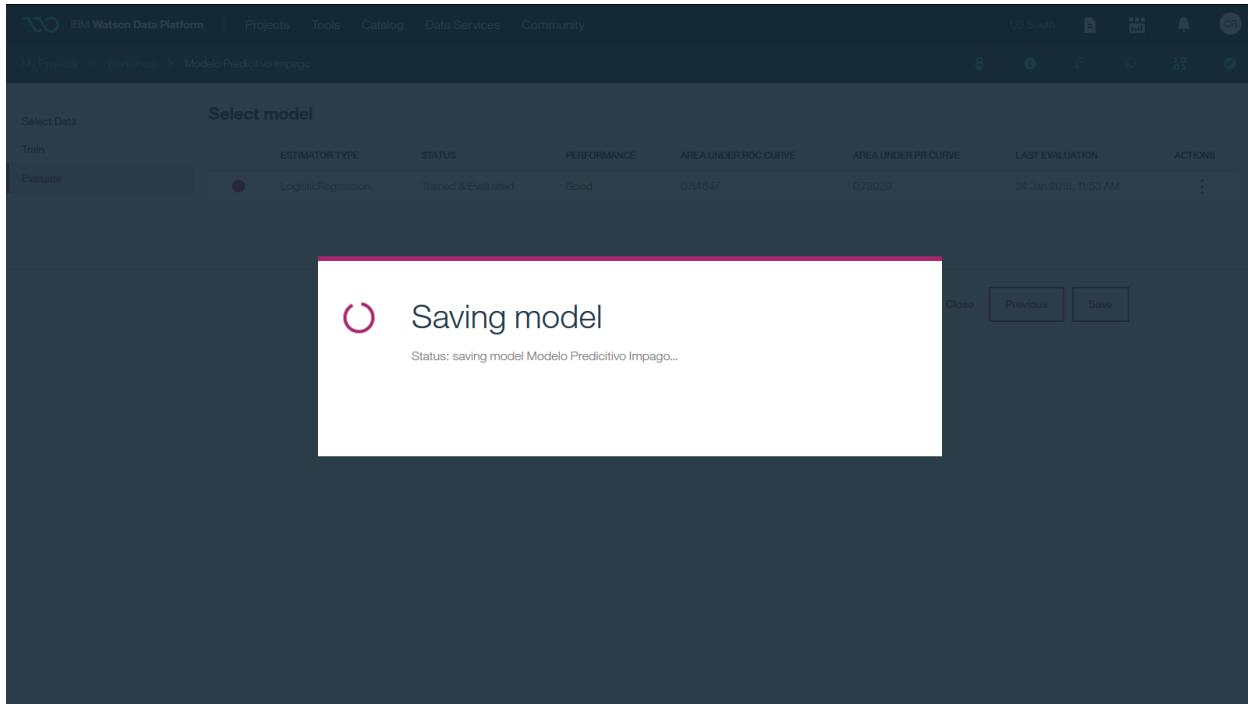
Add Data Assets

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.

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.

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 is a navigation bar with links for 'Projects', 'Tools', 'Catalog', 'Data Services', and 'Community'. Below the navigation bar, the path 'My Projects > Workshops > Modelo-Predictivo Impago' is visible. The main content area is titled 'Select model' and shows a table with two rows: 'Train' and 'Evaluate'. The 'Evaluate' row is selected, indicated by a blue dot. The table columns are 'ESTIMATOR TYPE', 'STATUS', 'PERFORMANCE', 'AREA UNDER ROC CURVE', 'AREA UNDER PR CURVE', 'LAST EVALUATION', and 'ACTIONS'. The 'Evaluate' row has values: 'LogisticRegression', 'Trained & Evaluated', 'Good', '0.64847', '0.72029', '24 Jan 2018, 11:53 AM', and a 'More' button. A modal dialog box is overlaid on the page, titled 'Saving model'. It contains the text 'Status: saving model Modelo Predictivo Impago...' and has three buttons: 'Close', 'Previous', 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-7ee99ebdfb33
Model builder details	<a href="#">View</a>

**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-7ee99ebdfb33
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-7ee99ebdfb33		⋮

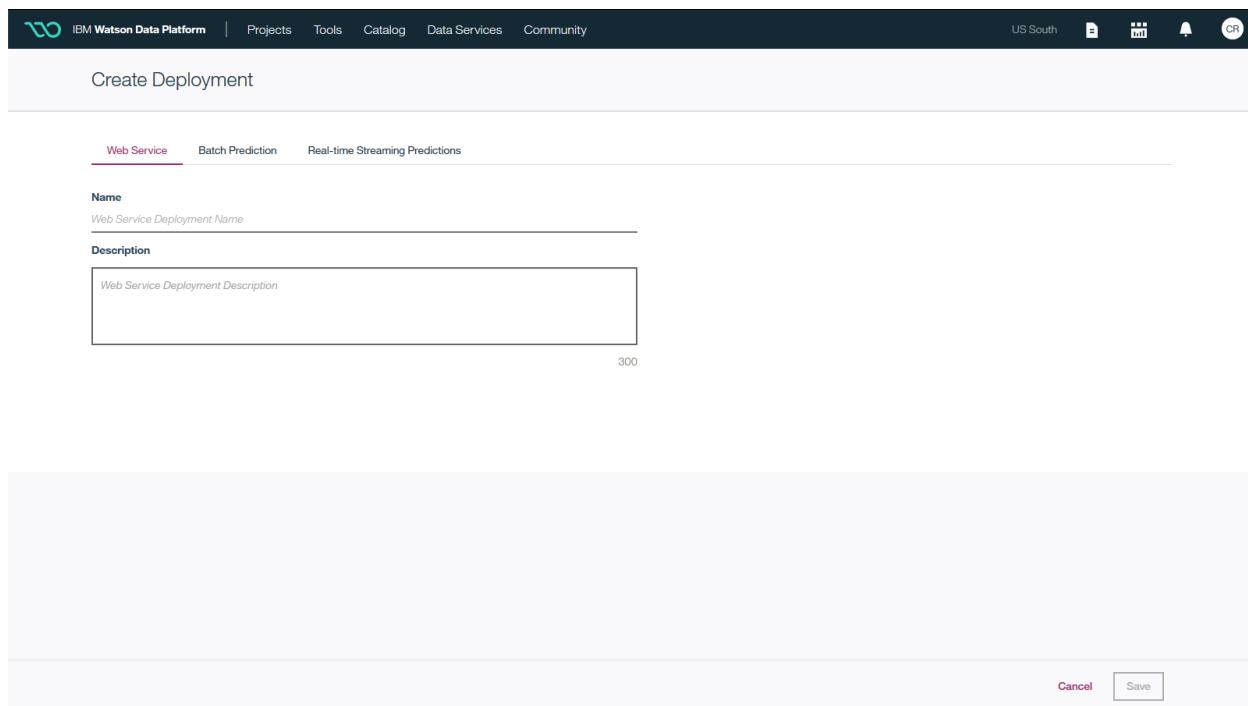
Ahora, podemos hacer un test y así, probar con otros datos nuestro modelo, y ver que output nos devuelve: por ejemplo 41 años, 1 educación que es nivel básico, 1 año en el mismo empleo y dejamos todo lo demás, le damos a probar y nos devuelve el resultado de la predicción.

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 is a navigation bar with the IBM Watson Data Platform logo, Projects, Tools, Catalog, Data Services, and Community. Below the navigation bar, the breadcrumb navigation shows 'My Projects > Workshop > Modelo Predictivo Impago'. The main content area is titled 'Modelo Predictivo Impago' with a subtitle '(1)'. Below the title, there are three tabs: Overview, Evaluation, and Deployments. The Deployments tab is selected and highlighted in red. To the right of the tabs, there is a red button labeled 'Add Deployment' with a plus sign. Below the tabs, there is a table with the following columns: NAME, STATUS, DEPLOYMENT TYPE, and ACTIONS. The table has a single row with the text 'Your model is not deployed.' The rest of the table is empty.

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



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

Create Deployment

**Web Service** Batch Prediction Real-time Streaming Predictions

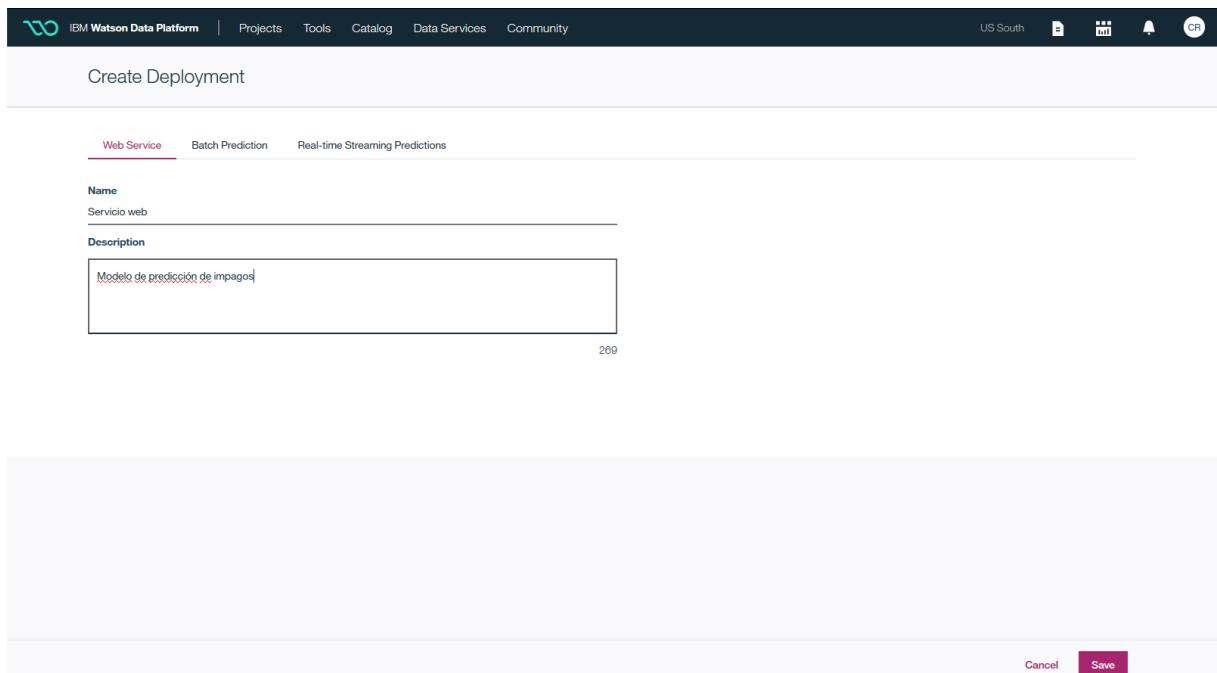
**Name**  
Web Service Deployment Name

**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.



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

Create Deployment

**Web Service** Batch Prediction Real-time Streaming Predictions

**Name**  
Servicio web

**Description**  
Modelo de predicción de imágenes

200

Cancel Save

### Modelo Predictivo Impago

Overview   Evaluation   **Deployments**

**Add Deployment**

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
Servicio web	ACTIVE	Web Service	⋮

### Modelo Predictivo de Impagos

Overview   Evaluation   **Deployments**

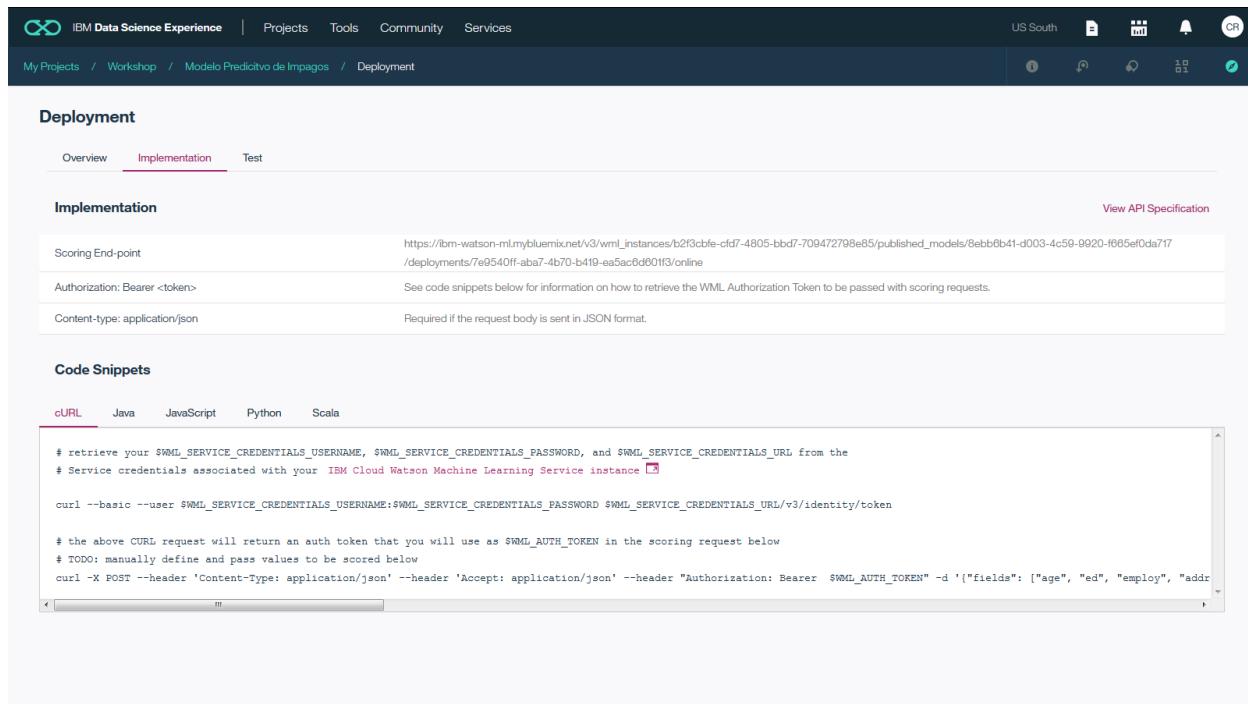
**Add Deployment**

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.



**Deployment**

Overview Implementation Test

**Implementation**

Scoring End-point: [https://ibm-watson-ml.mybluemix.net/v3/wml\\_instances/b2f3cbfe-cfd7-4805-bbd7-709472798e85/published\\_models/8ebb6b41-d003-4c59-9920-f865ef0da7?](https://ibm-watson-ml.mybluemix.net/v3/wml_instances/b2f3cbfe-cfd7-4805-bbd7-709472798e85/published_models/8ebb6b41-d003-4c59-9920-f865ef0da7?)

Authorization: Bearer <token>: See code snippets below for information on how to retrieve the WML Authorization Token to be passed with scoring requests.

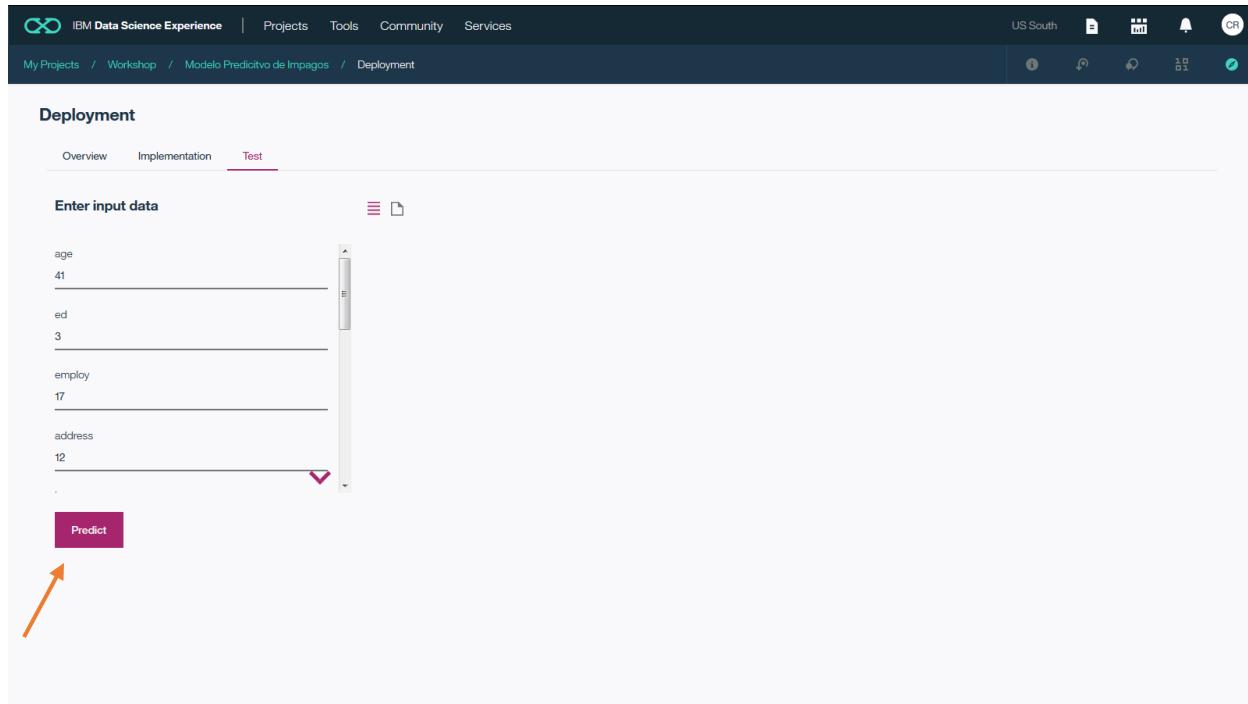
Content-type: application/json: Required if the request body is sent in JSON format.

**Code Snippets**

cURL Java JavaScript Python Scala

```
# retrieve your $WML_SERVICE_CREDENTIALS_USERNAME, $WML_SERVICE_CREDENTIALS_PASSWORD, and $WML_SERVICE_CREDENTIALS_URL from the
# Service credentials associated with your IBM Cloud Watson Machine Learning Service instance
curl --basic --user $WML_SERVICE_CREDENTIALS_USERNAME:$WML_SERVICE_CREDENTIALS_PASSWORD $WML_SERVICE_CREDENTIALS_URL/v3/identity/token

# the above CURL request will return an auth token that you will use as $WML_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 $WML_AUTH_TOKEN" -d '{"fields": ["age", "ed", "employ", "addr"]}' $WML_SERVICE_CREDENTIALS_URL/v3/scoring
```



**Deployment**

Overview Implementation Test

**Enter input data**

age: 41

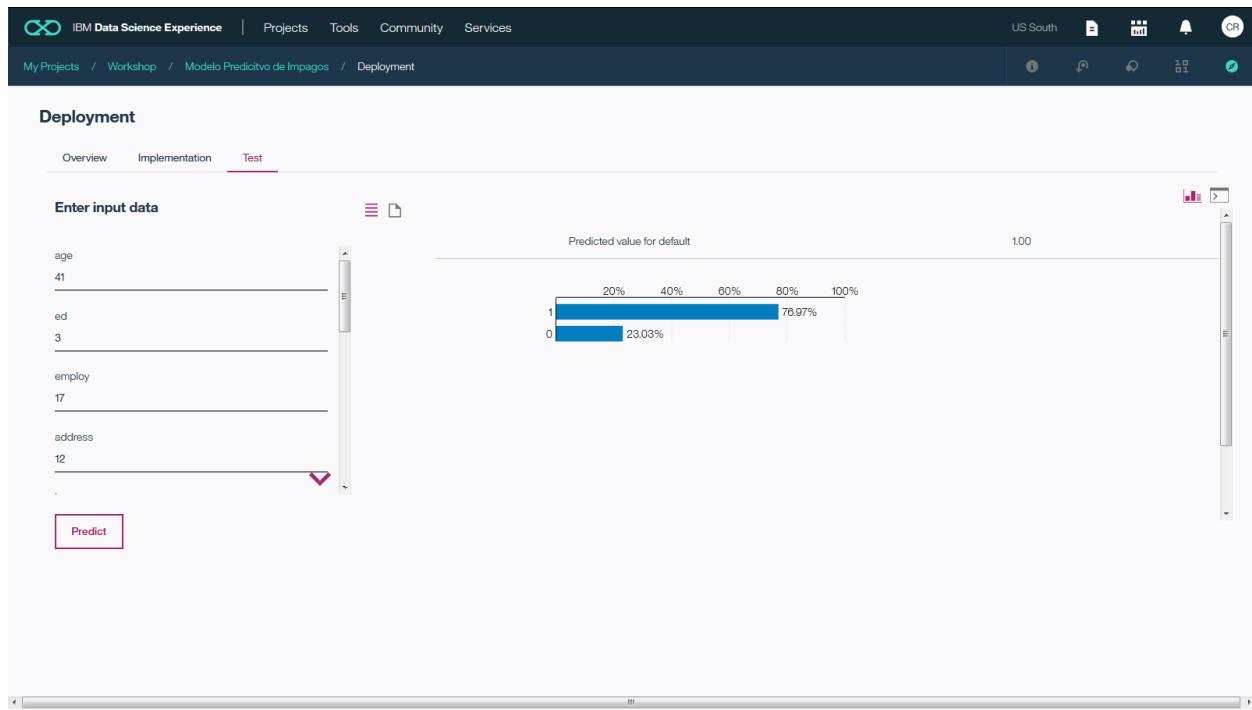
ed: 3

employ: 17

address: 12

**Predict**

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



IBM Data Science Experience | Projects Tools Community Services

My Projects / Workshop / Modelo Predictivo de Impagos / Deployment

US South

Deployment

Overview Implementation Test

Enter input data

age  
41

ed  
3

employ  
17

address  
12

Predicted value for default

1.00

20% 40% 60% 80% 100%

1 76.97%

0 23.03%

Predict

#### 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**.

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

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).

Podemos seleccionar uno o varios estimadores. Añadimos y ejecutamos.

IBM Watson Data Platform | Projects Tools Catalog Data Services Community

My Projects > Workshop > Modelo predictivo automático

US South

Select model

	ESTIMATOR TYPE	STATUS	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST EVALUATION	ACTIONS
<input checked="" type="radio"/>	RandomForestClassifier	Trained & Evaluated	Good	0.84825	0.634	24 Jan 2018, 12:39 PM	⋮
<input type="radio"/>	LogisticRegression	Trained & Evaluated	Good	0.83114	0.61084	24 Jan 2018, 12:39 PM	⋮
<input type="radio"/>	DecisionTreeClassifier	Trained & Evaluated	Poor	0.62931	0.43677	24 Jan 2018, 12:39 PM	⋮

Close Previous Save

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.

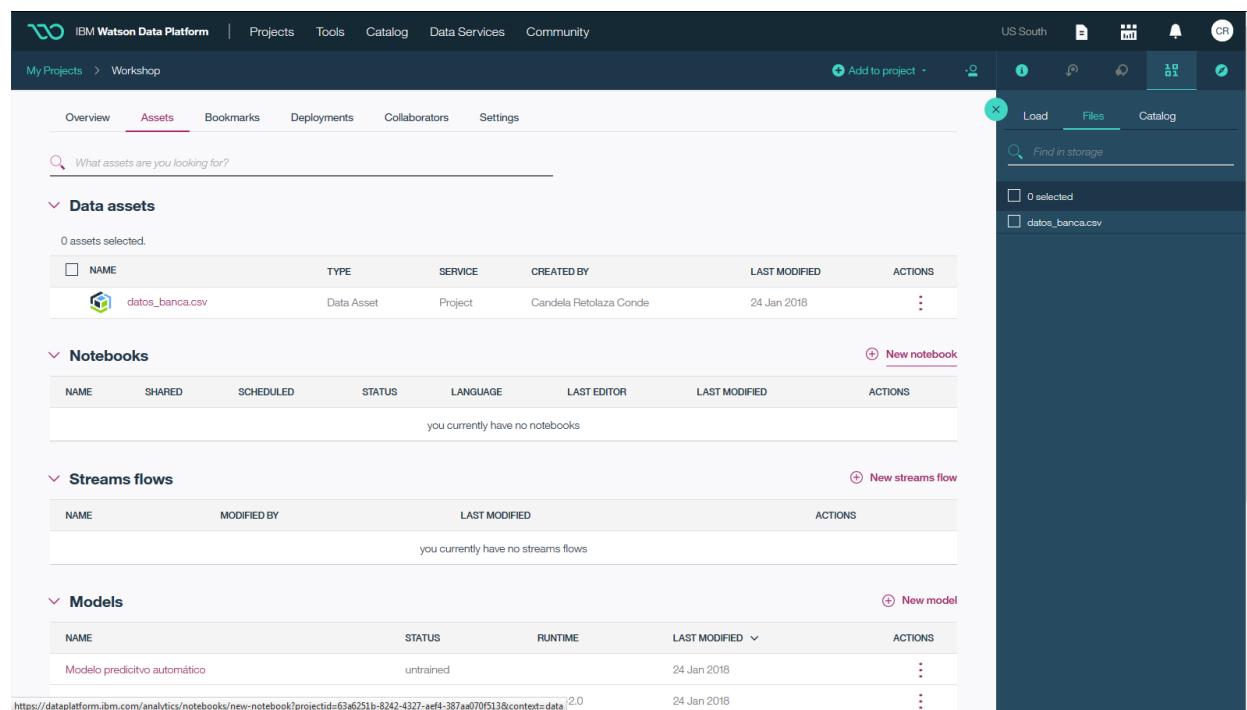
## Parte predictiva

### 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 Data Science Experience (DSX):

**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 'Assets' tab selected in the IBM Watson Data Platform interface. On the right, a sidebar titled 'Catalog' is open, showing a list of files including 'datos\_banca.csv'. At the top right of the main content area, there is a prominent blue button labeled '+ New notebook'.

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

**Description**  
Type your Description here

**Language\***  
Python 2    R    Scala    Python 3.5 Experimental

**Spark version\***  
2.1    2.0

**Spark service\***  
Spark-ae

Cancel    Create Notebook

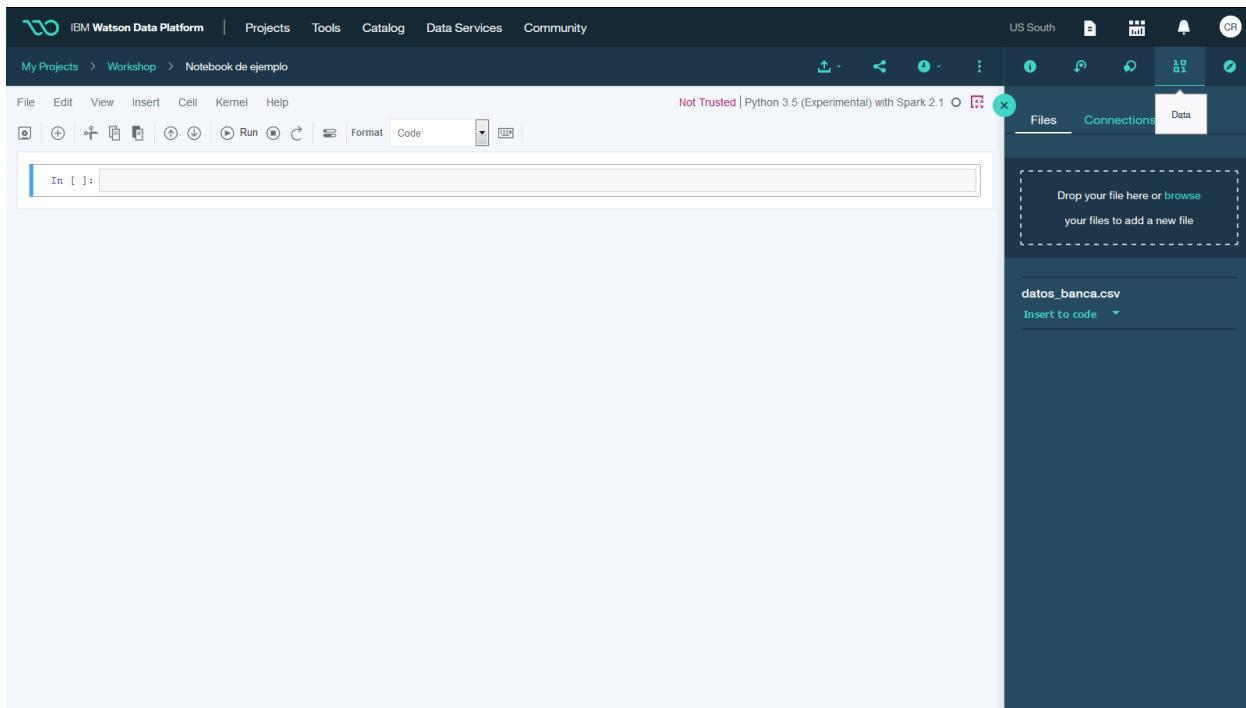
Puedes crear un notebook en blanco, cargar un archivo de notebook desde su sistema de archivos o cargar un archivo de notebook desde una URL. El notebook que crea o selecciona debe ser un archivo.ipynb.

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.

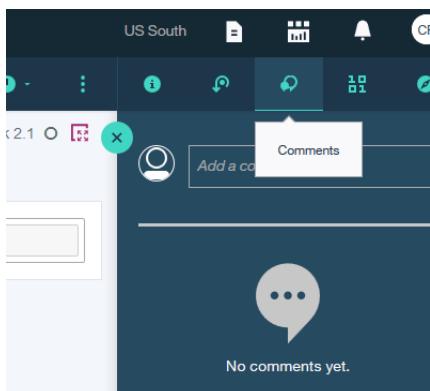
Para desarrollar aplicaciones analíticas en un Notebook, siga estos pasos generales:

- 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.

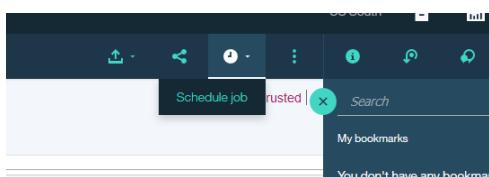
- iii. Cargar y acceder a los datos. Vemos que podemos añadir a nuestro notebook un dataset o un fichero de datos, además de conexiones de datos.

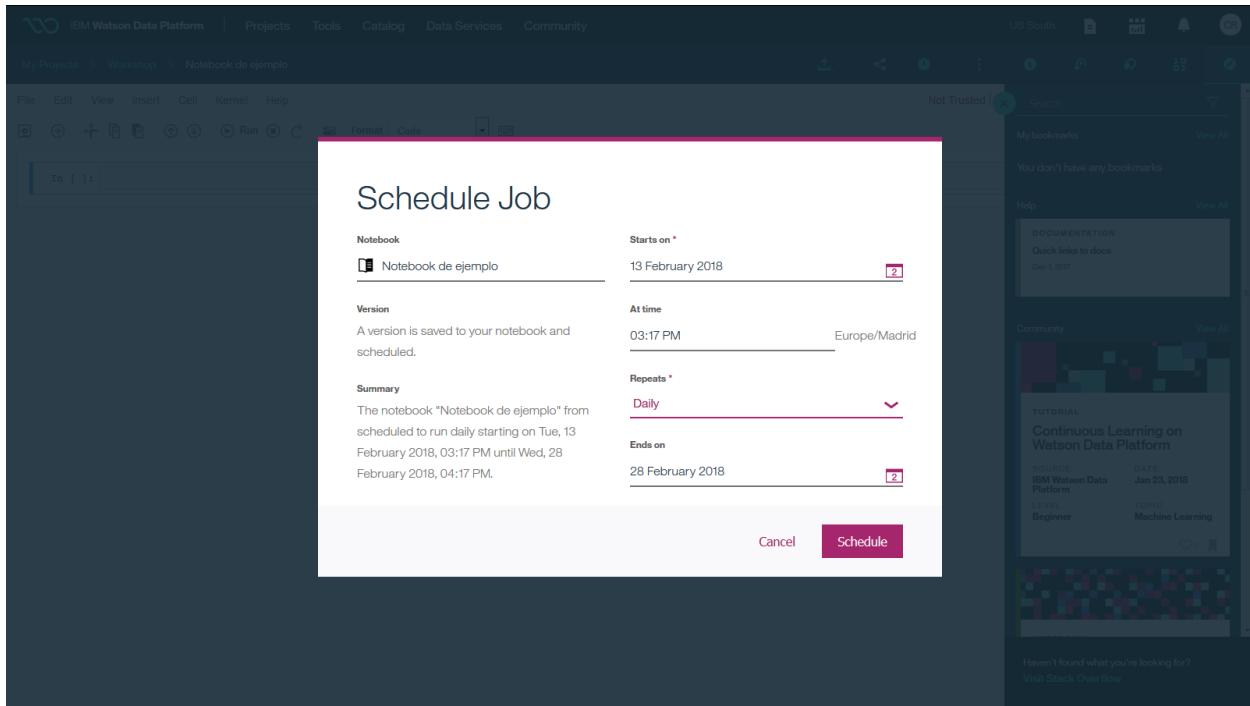


Colabora con otros miembros del proyecto. Puede agregar comentarios a los cuadernos haciendo clic en el ícono de comentario (ícono **Comentario**).

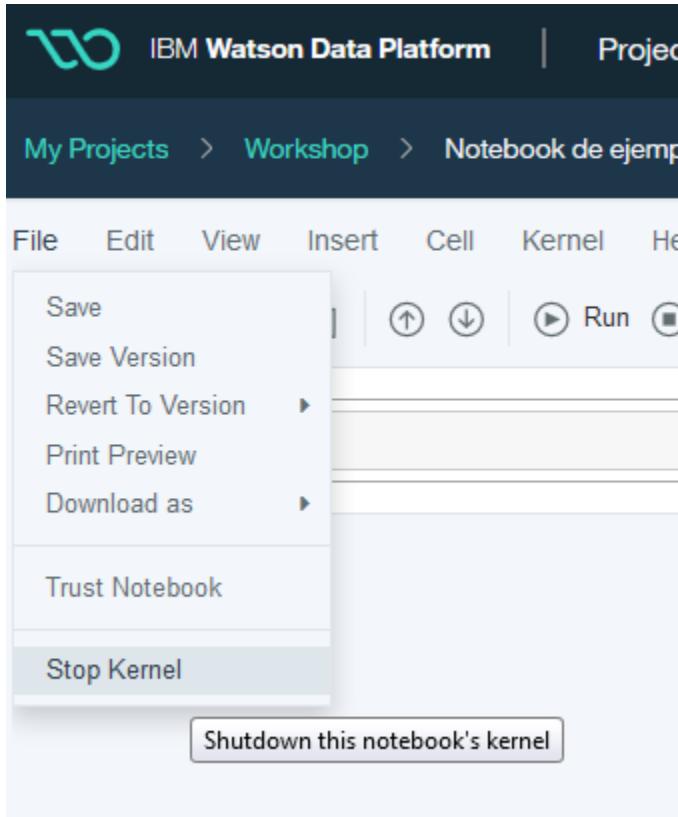


Si es necesario, programa el notebook para que se ejecute en otro momento.

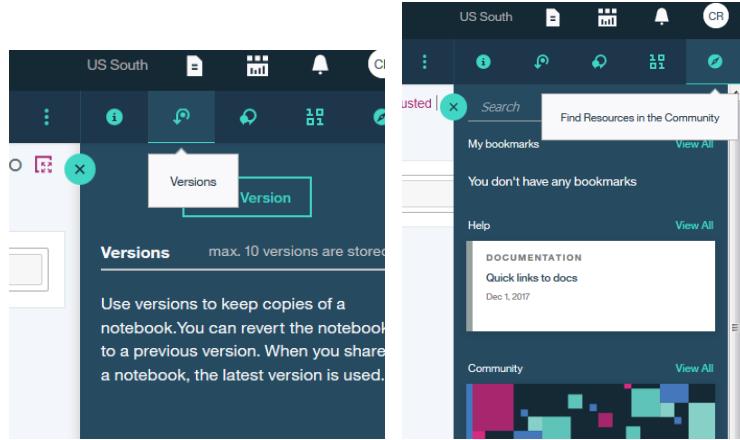




Cuando no estés trabajando activamente en el notebook, haz clic en **File> Stop kernel** para detener el kernel del notebook.

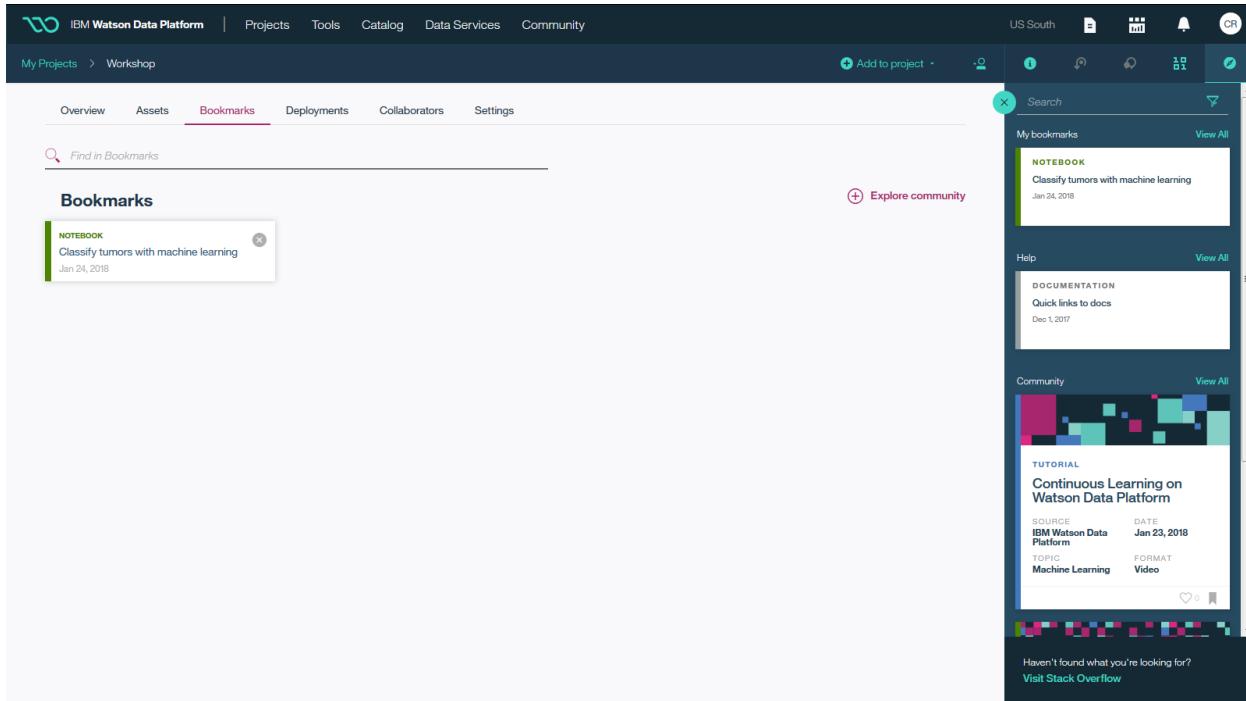


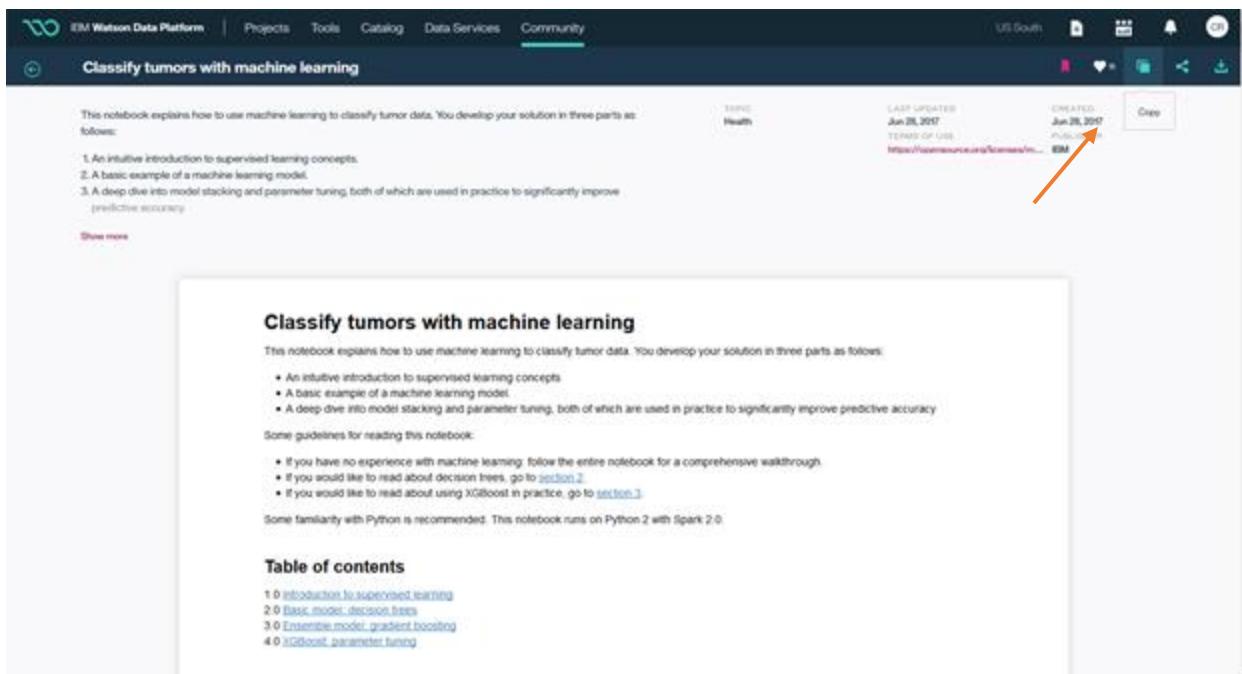
Además, podemos buscar recursos en la comunidad, y aprovechar esos recursos para enriquecer o comenzar un proyecto.



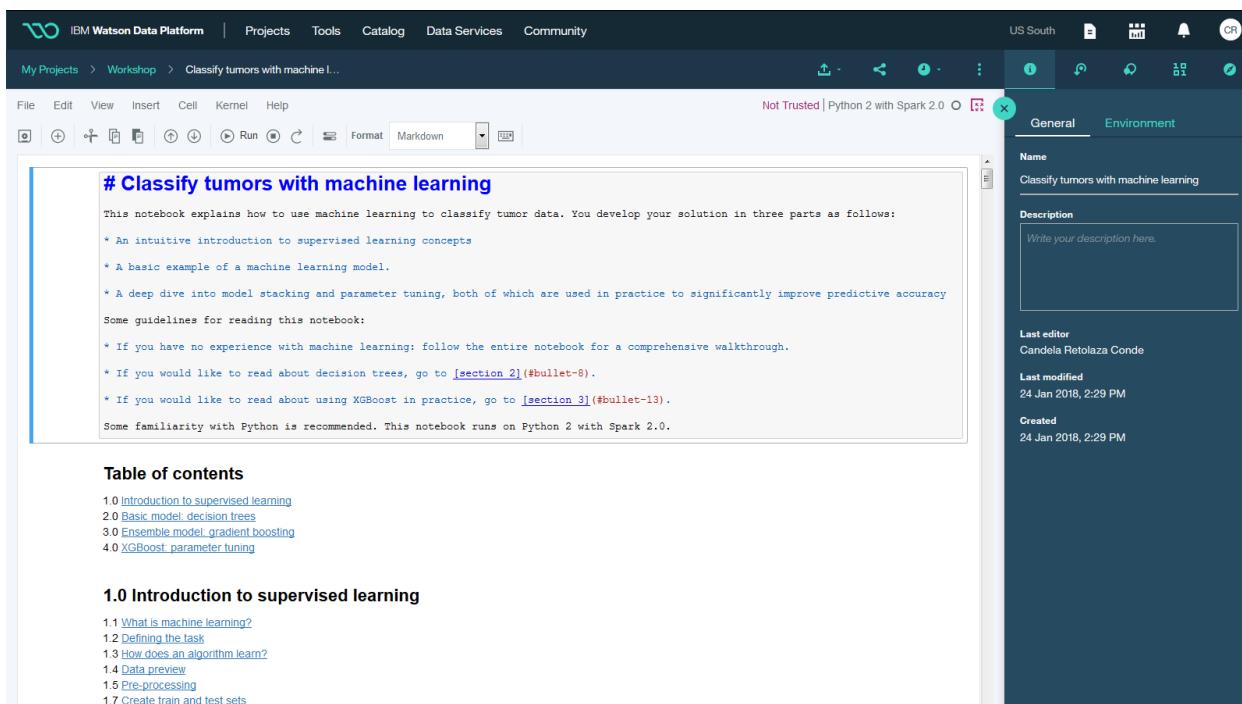
Para ello, debemos buscar en la comunidad algún notebook que nos pueda interesar, guardarlo en el proyecto, y copiarlo para utilizarlo o reutilizar ciertas partes de él.

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.





This screenshot shows the IBM Watson Data Platform interface. At the top, there are navigation links: Projects, Tools, Catalog, Data Services, and Community. On the right side, there are icons for Health, US South, and a bell. Below the navigation, the title 'Classify tumors with machine learning' is displayed. The notebook content includes a brief introduction, a table of contents, and some guidelines. In the top right corner of the notebook area, there is a 'Copilot' button, which is highlighted with an orange arrow.



This screenshot shows the same notebook in a different view within the IBM Watson Data Platform. The interface includes a top navigation bar with Project, Tools, Catalog, Data Services, and Community. Below that is a toolbar with File, Edit, View, Insert, Cell, Kernel, and Help. The notebook content is visible in the main area, and on the right, there is a sidebar with tabs for 'General' and 'Environment'. The 'General' tab shows the notebook's name, description, and last editor information. The 'Copilot' section is also present in this view.

Podemos copiar celdas o trozos para reutilizar, etc. Además, podemos compartir o notebooks para que las personas que no tienen cuentas DSX 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 DSX. 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 DSX: *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 DSX:

New notebook

Blank From file From URL

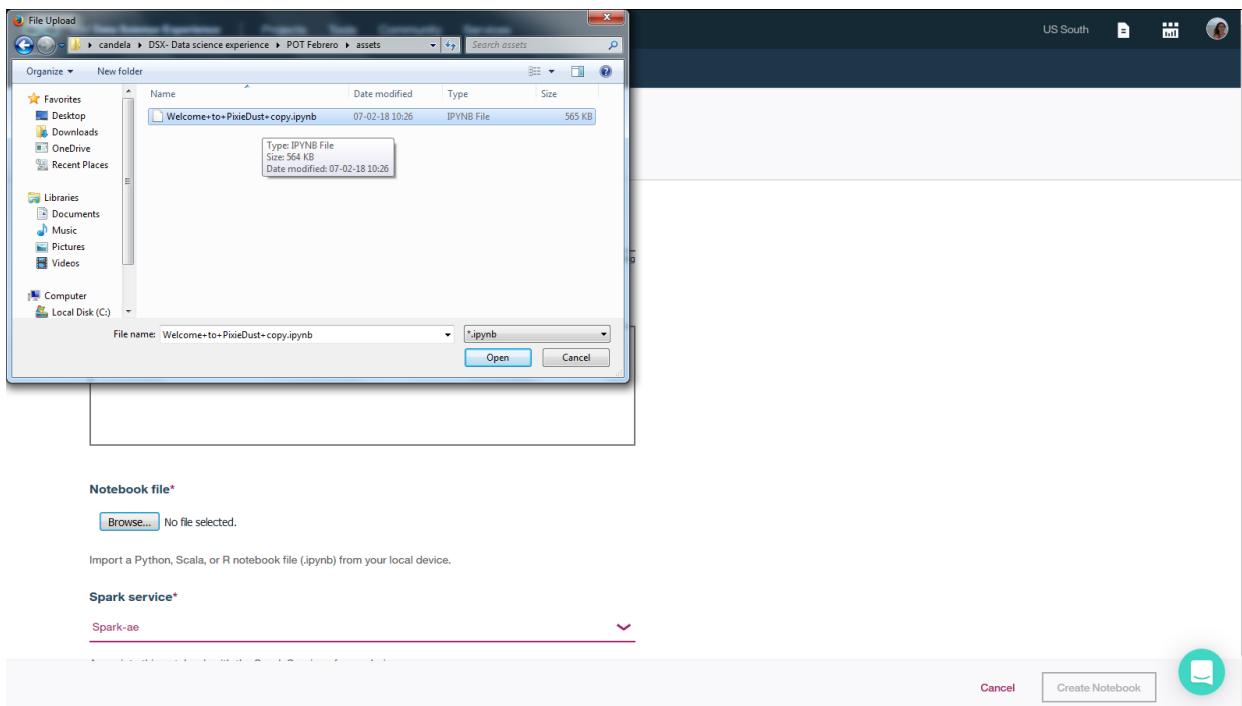
Name\*  
Type Notebook Name here

Description  
Type your Description here

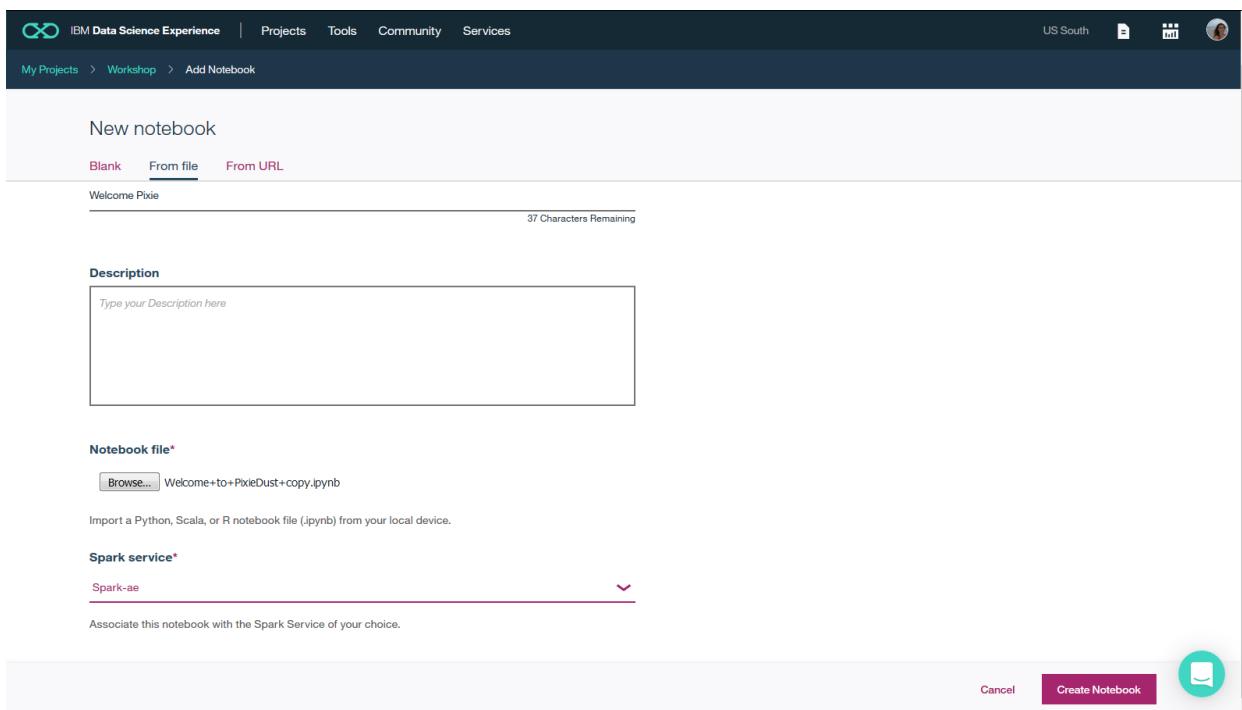
Notebook file\*  
Browse... No file selected.  
Import a Python, Scala, or R notebook file (.ipynb) from your local device.

Spark service\*  
Spark-ae

Create Notebook



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 PixelDust.

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.

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.

Cancel 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 (78kB)

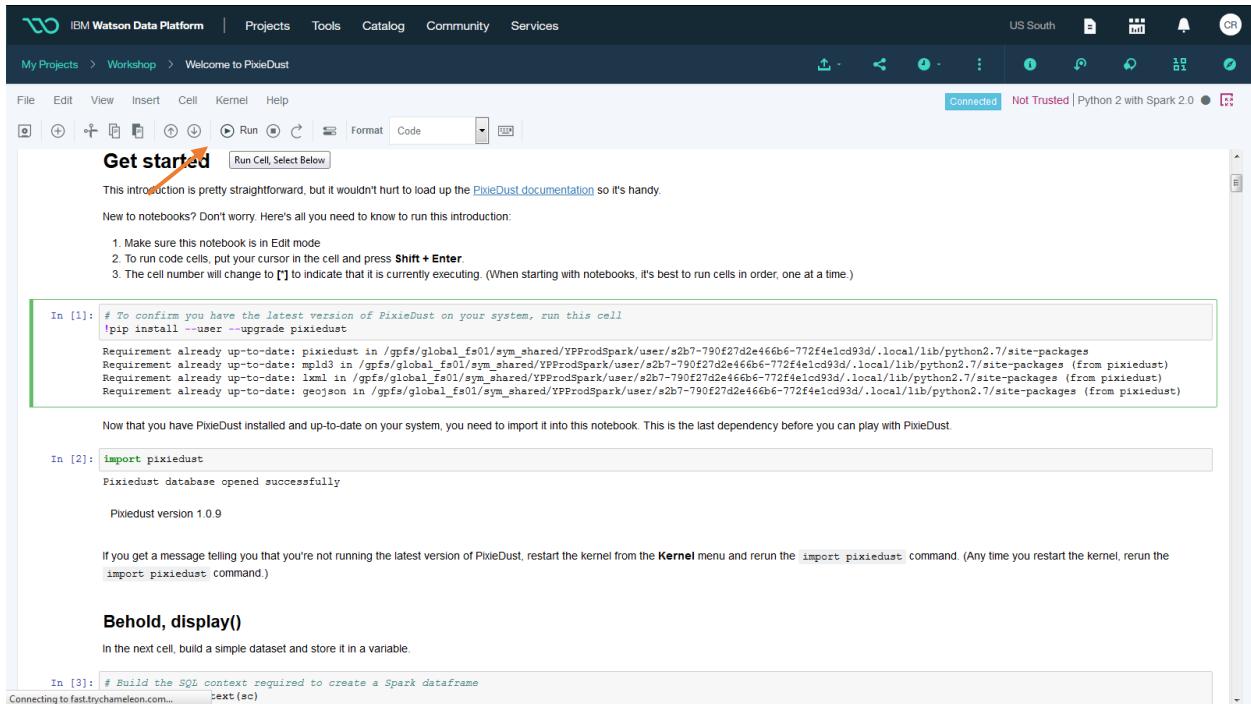
    100% |████████████████████████████████| 81kB 1.2MB/s eta 0:00:01

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 (from pixiedust)



The screenshot shows the IBM Watson Data Platform Workshop interface. At the top, there are navigation links for 'Projects', 'Tools', 'Catalog', 'Community', and 'Services'. The top right corner shows 'US South' and various status icons. The main area is a Jupyter notebook titled 'Welcome to PixieDust'. A red arrow points to the 'Get started' button. The notebook contains the following content:

```

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 /gpfs/global_fso1/sym_shared/YPProdSpark/user/s2b7-790f27d2e46b6-772f4e1cd93d/.local/lib/python2.7/site-packages
Requirement already up-to-date: mpfd3 in /gpfs/global_fso1/sym_shared/YPProdSpark/user/s2b7-790f27d2e46b6-772f4e1cd93d/.local/lib/python2.7/site-packages (from pixiedust)
Requirement already up-to-date: lxml in /gpfs/global_fso1/sym_shared/YPProdSpark/user/s2b7-790f27d2e46b6-772f4e1cd93d/.local/lib/python2.7/site-packages (from pixiedust)
Requirement already up-to-date: georjson in /gpfs/global_fso1/sym_shared/YPProdSpark/user/s2b7-790f27d2e46b6-772f4e1cd93d/.local/lib/python2.7/site-packages (from pixiedust)

Now that you have PixieDust installed and up-to-date on your system, you need to import it into this notebook. This is the last dependency before you can play with PixieDust.

In [2]: import pixiedust
Pixiedust database opened successfully

Pixiedust version 1.0.9

If you get a message telling you that you're not running the latest version of PixieDust, restart the kernel from the Kernel menu and rerun the import pixiedust command. (Any time you restart the kernel, rerun the import pixiedust command.)
```

**Behold, display()**

In the next cell, build a simple dataset and store it in a variable.

```

In [3]: # Build the SQL context required to create a Spark dataframe
Connecting to fast.trychameleon.com... text(s)

```

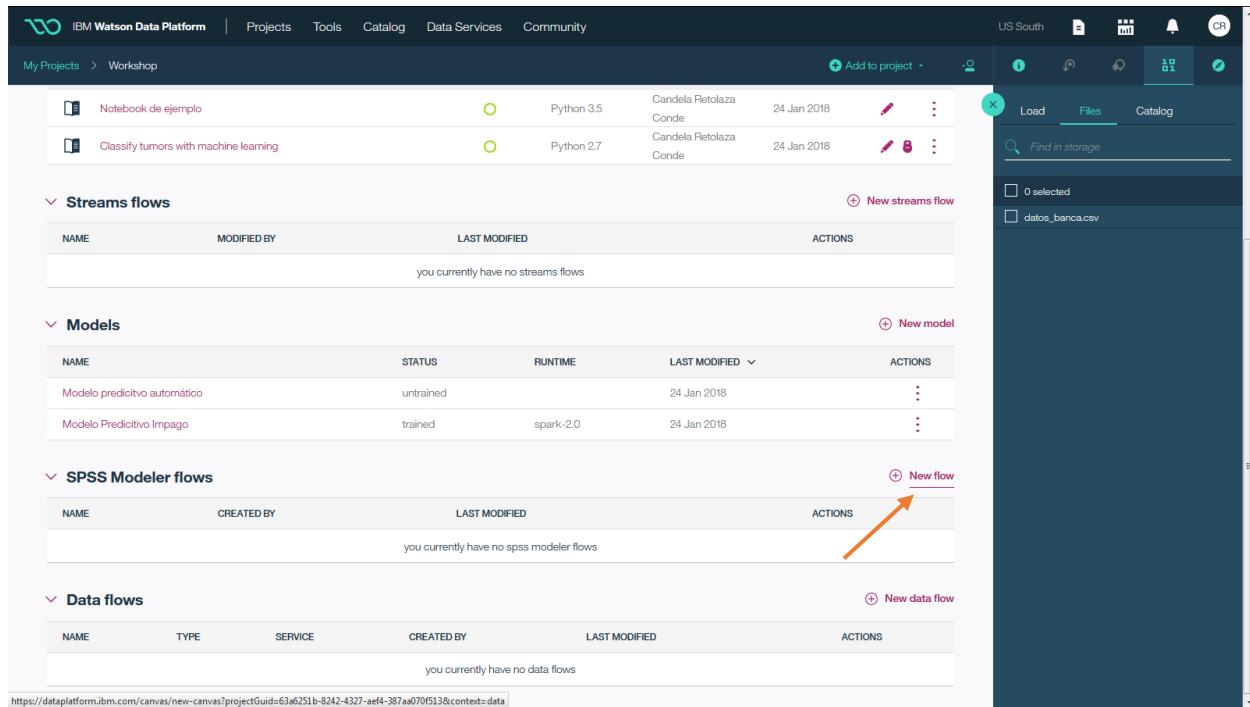
### 3. Algoritmos de analítica predictiva de SPSS

Una de las funcionalidades más interesantes que proporciona IBM como valor añadido a DSX 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.

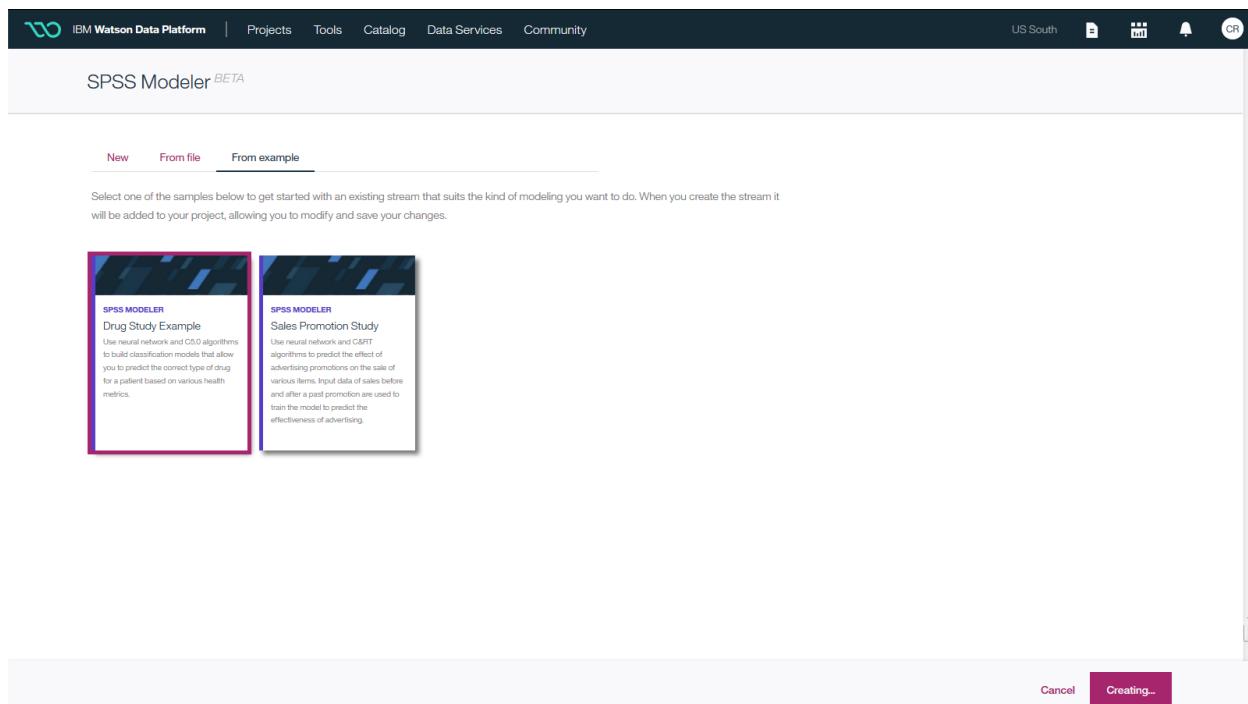
DSX está adquiriendo cada vez más funcionalidades de SPSS Modeler para añadirla a sus funcionalidades propias. Gracias a esto, el usuario es capaz de hacer minería de datos de una manera más sencilla, y así enriquecer los proyectos.

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 a list of projects and their details. On the right, there is a sidebar with tabs for 'Load', 'Files', and 'Catalog'. The 'Files' tab is active, showing a list of files including 'datos\_banca.csv'. The 'SPSS Modeler flows' section is expanded, showing a table with columns: NAME, CREATED BY, LAST MODIFIED, and ACTIONS. A red arrow points to the 'New flow' button in this section.

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.



SPSS Modeler BETA

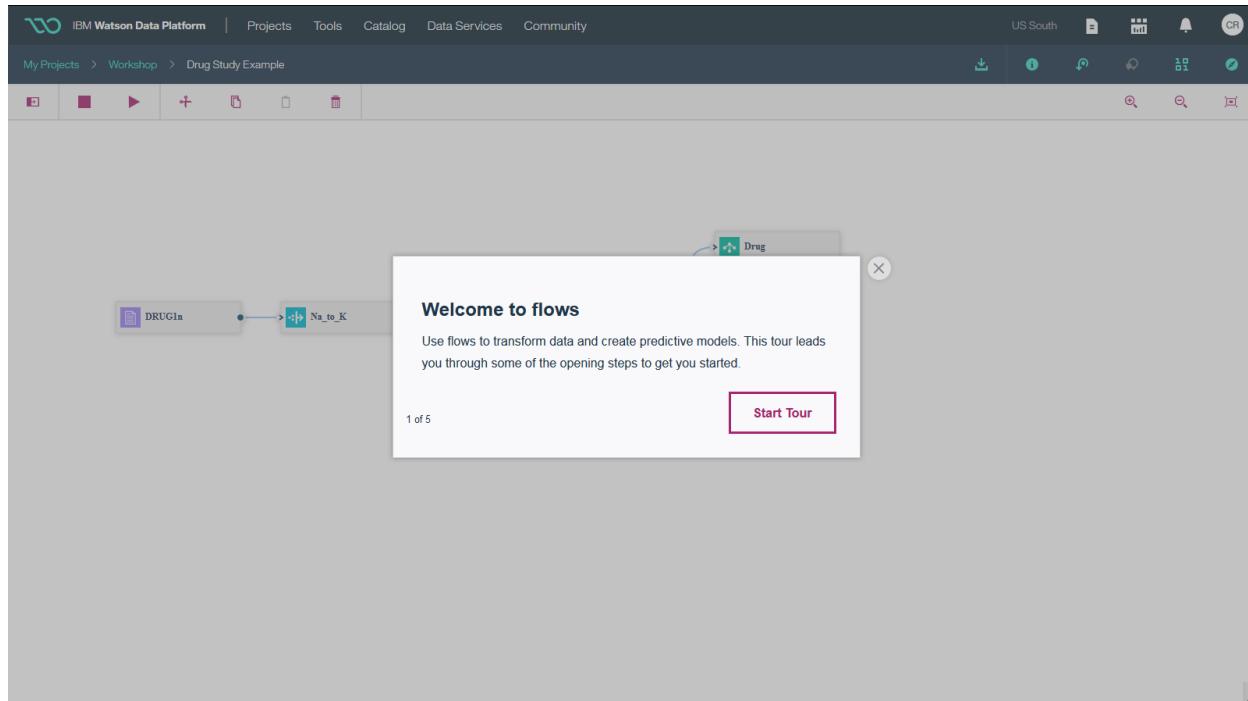
New    From file    From example

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

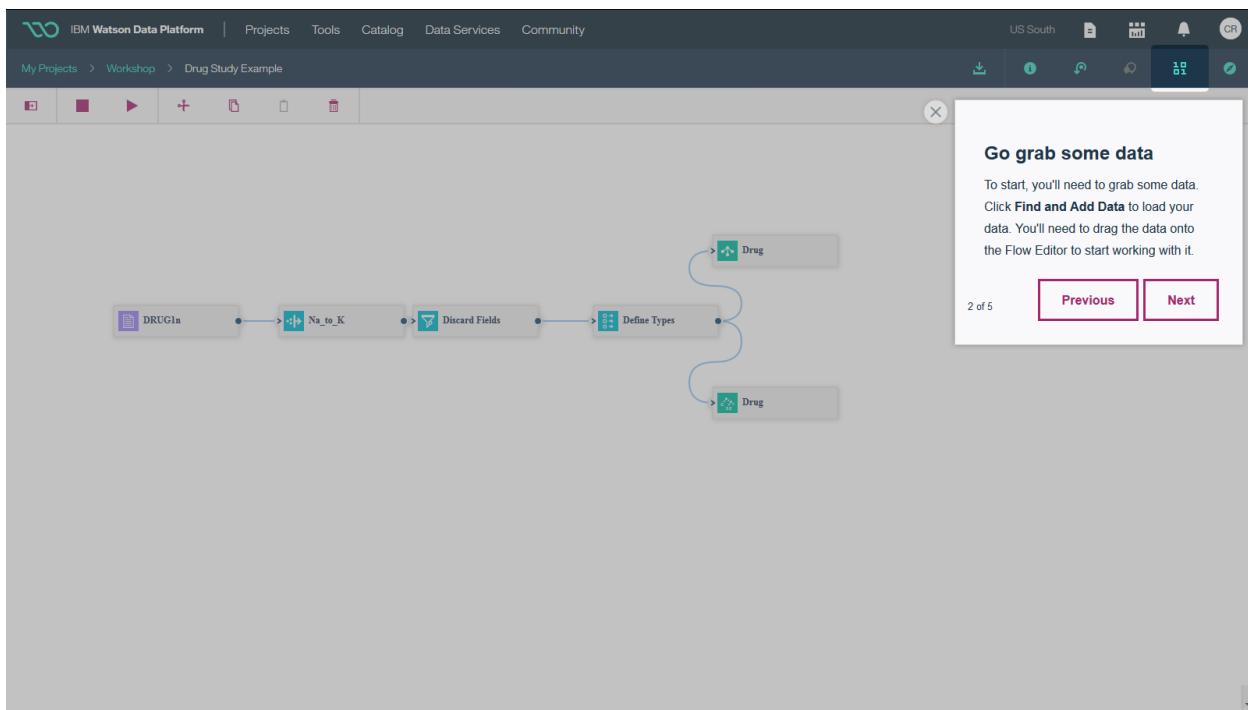
DRUG1a → Na\_to\_K

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

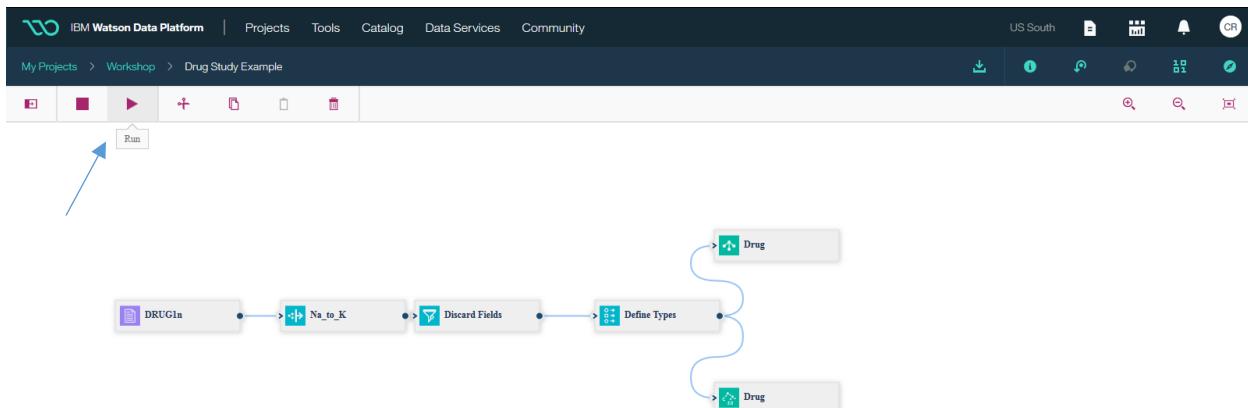
1 of 5

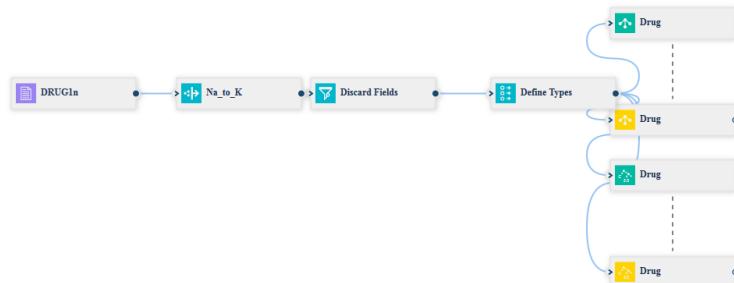


En este ejemplo, imagina que es un investigador médico que está recopilando datos para un estudio. Has recopilado información sobre un conjunto de pacientes, de los cuales todos sufrieron la misma enfermedad. Durante el curso del tratamiento, cada paciente respondió a un medicamento de un total de cinco. 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

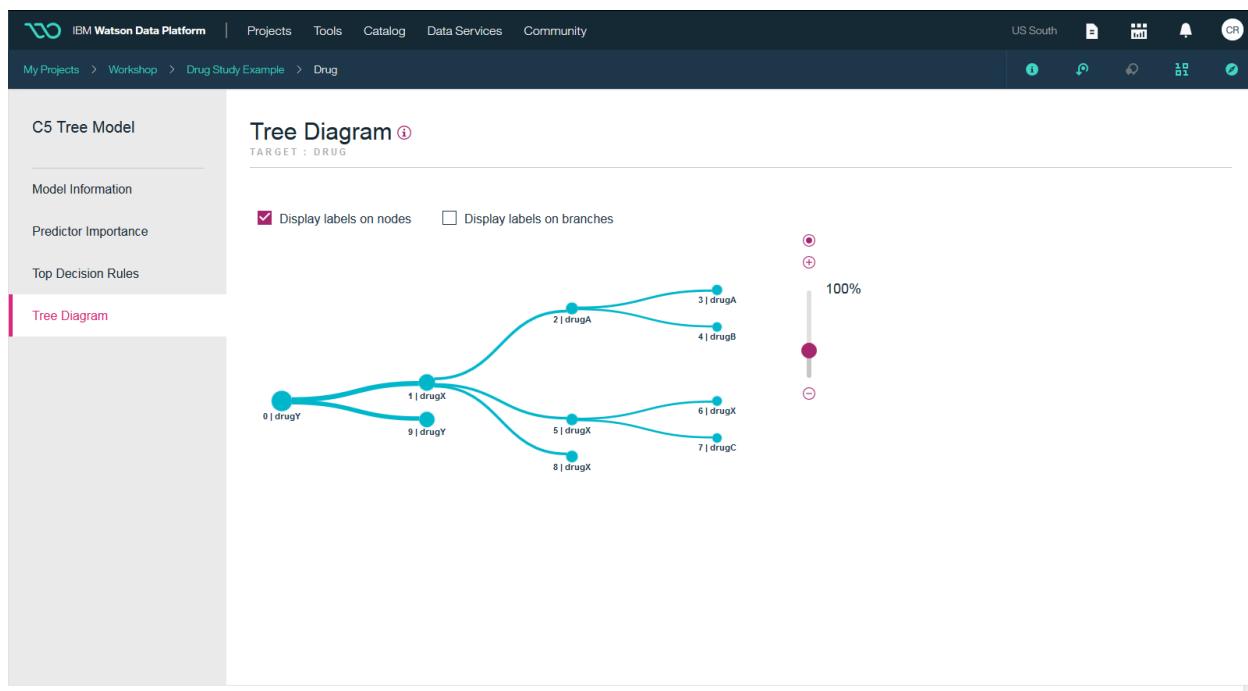
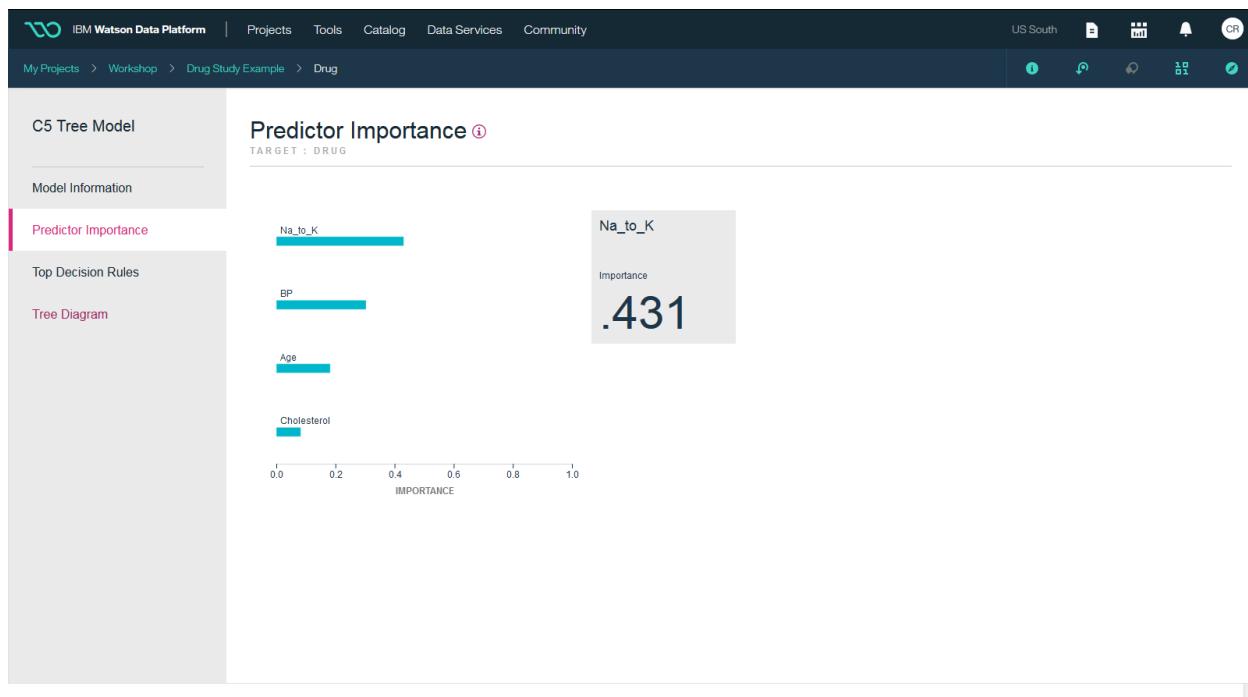
Debido a que el sodio y el potasio tienen una alta correlación, se puede observar que es así haciendo un gráfico, se crea una nueva variable para que pueda ser utilizada en el modelo.



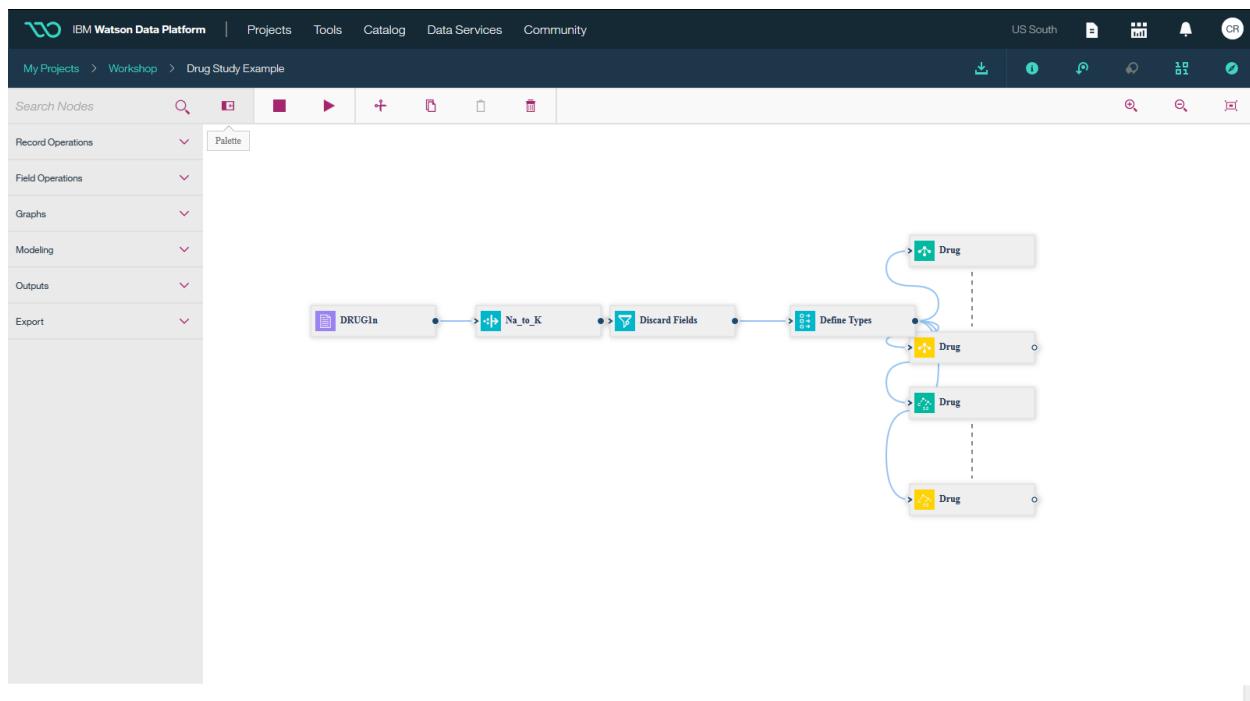


Para observar los resultados: Botón derecho: **view model**

Model Information	Value
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.

SPSS Modeler BETA

New    From file    From example

Name\*

Modelo de predicción de impago

Description

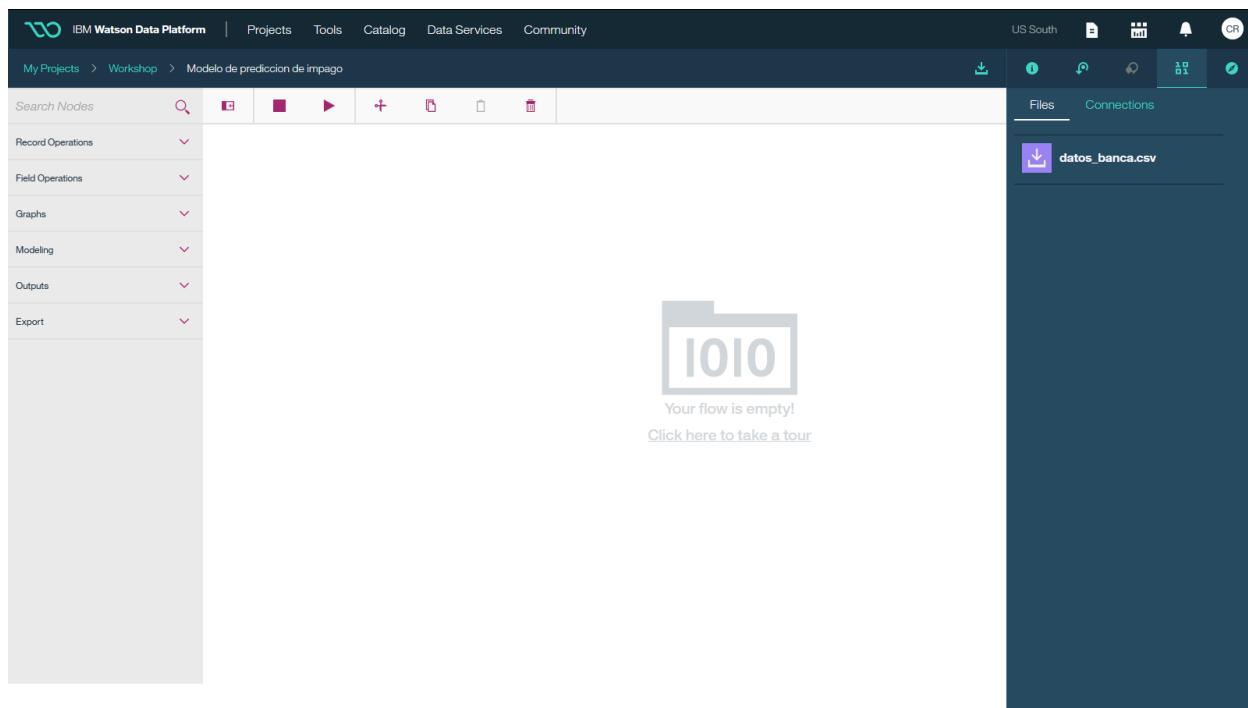
Type description here.

Runtime

IBM SPSS Modeler

Cancel    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.

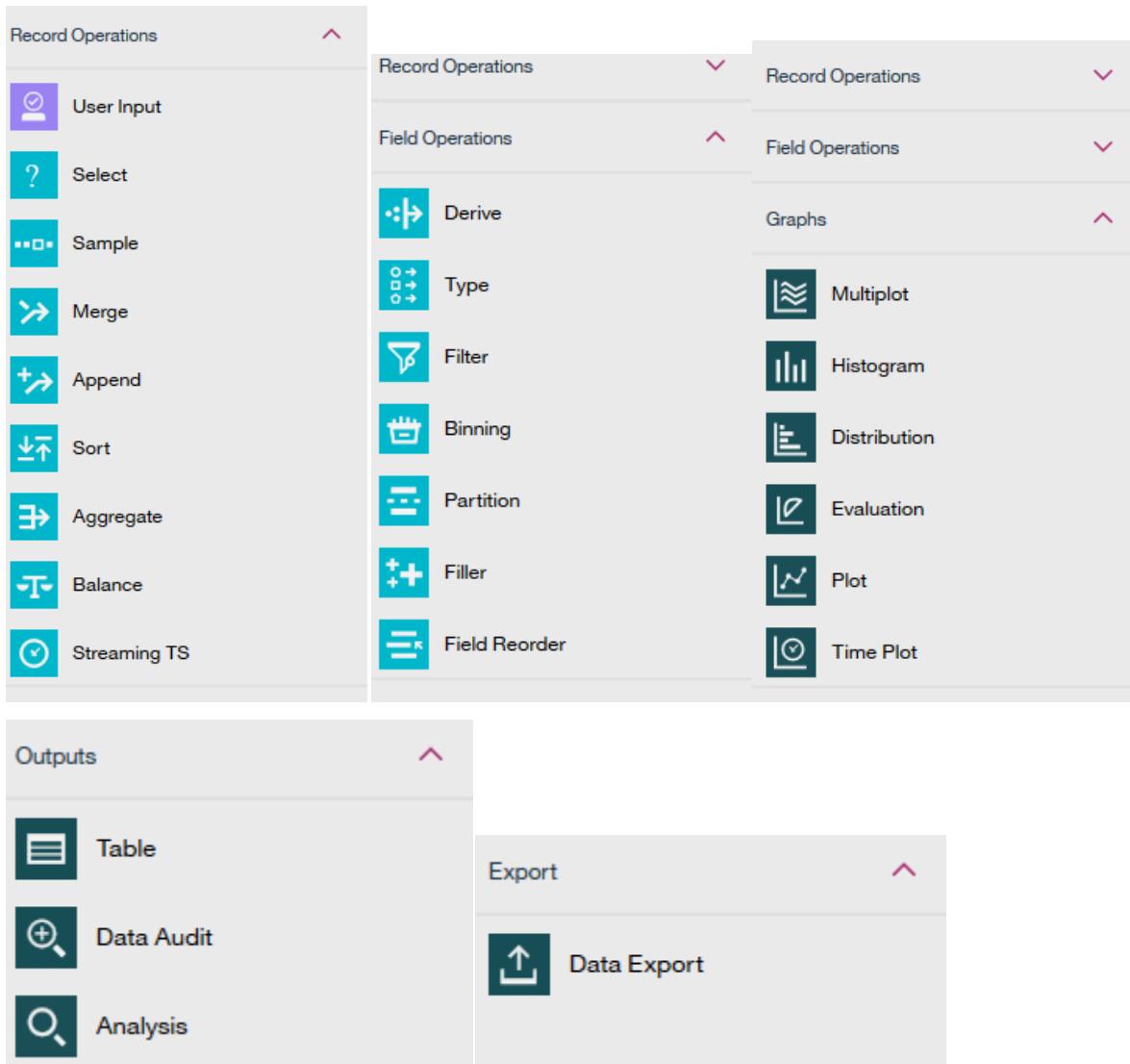
Los nodos Operaciones con campos realizan operaciones en campos de datos como, por ejemplo, filtrar, derivar campos nuevos y determinar el nivel de medición para campos dados.

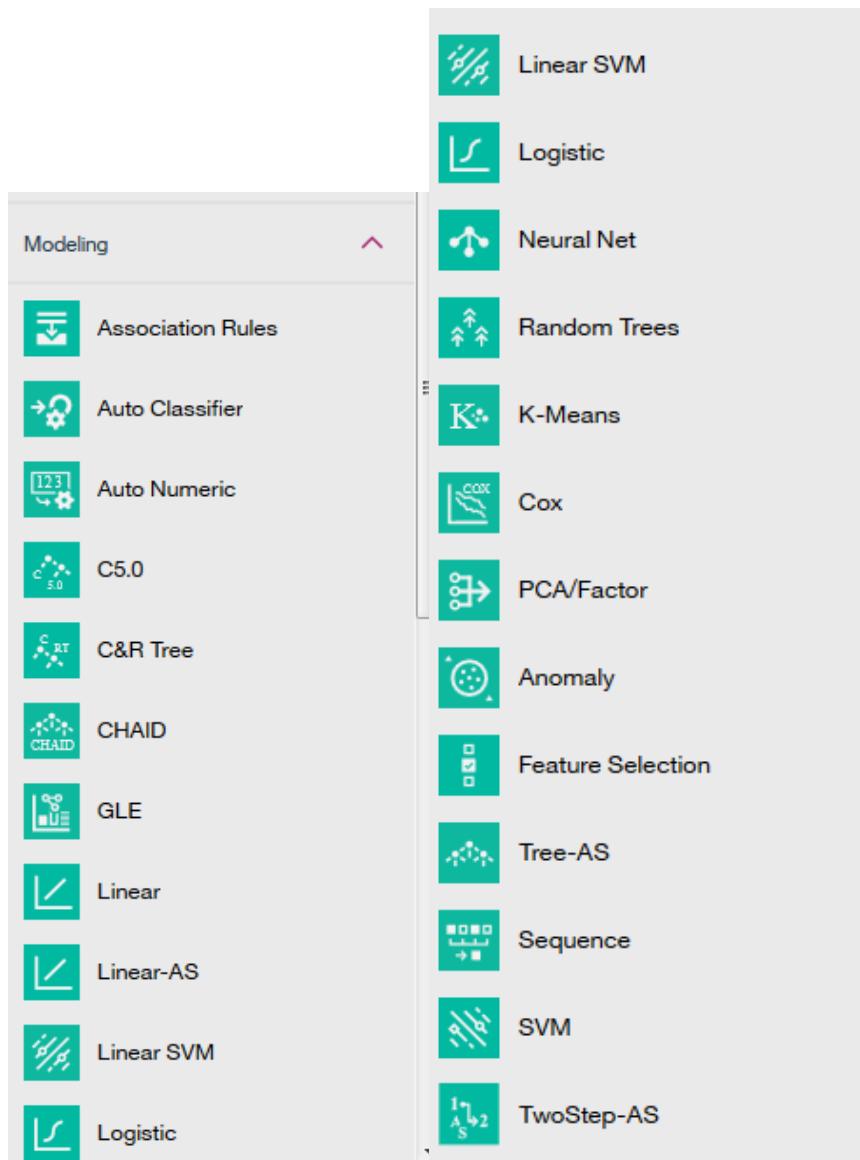
Los nodos Gráficos muestran gráficamente los datos antes y después del modelado. Entre ellos se incluyen gráficos, histogramas, nodos de malla y diagramas de evaluación. Los nodos Modelado utilizan los algoritmos de modelado disponibles en SPSS Modeler como, por ejemplo, redes neuronales, árboles de 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 DSX.

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

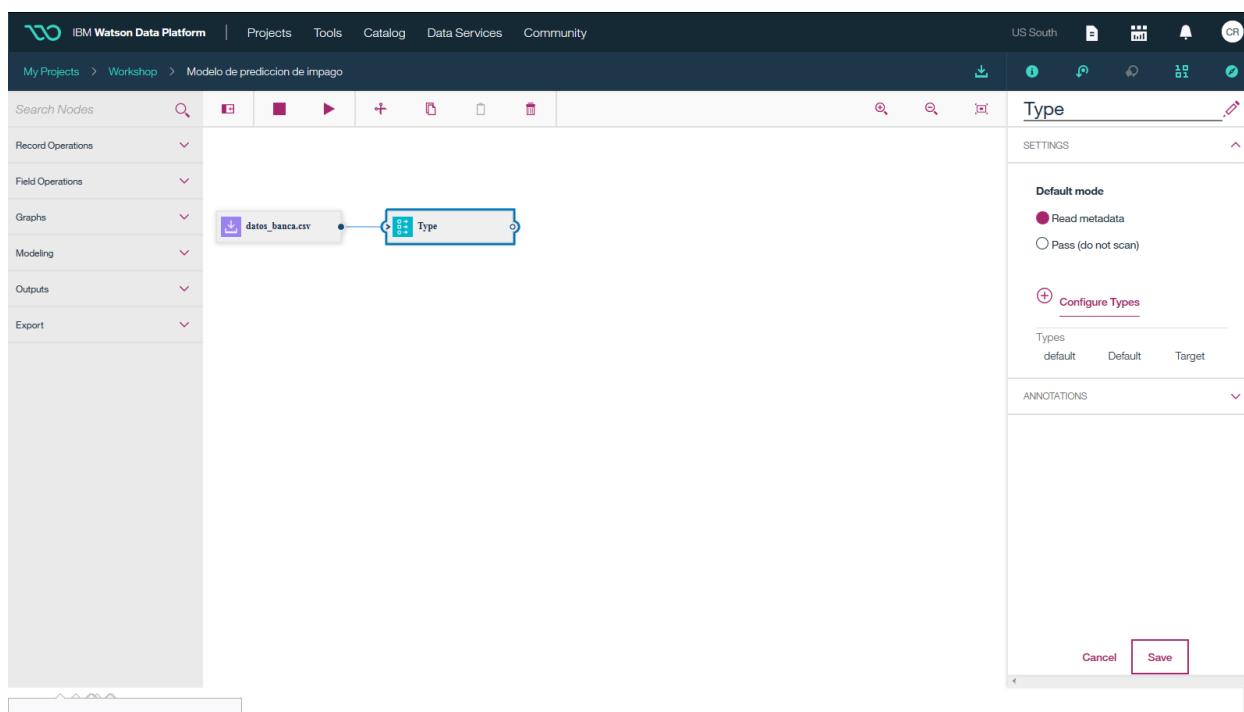
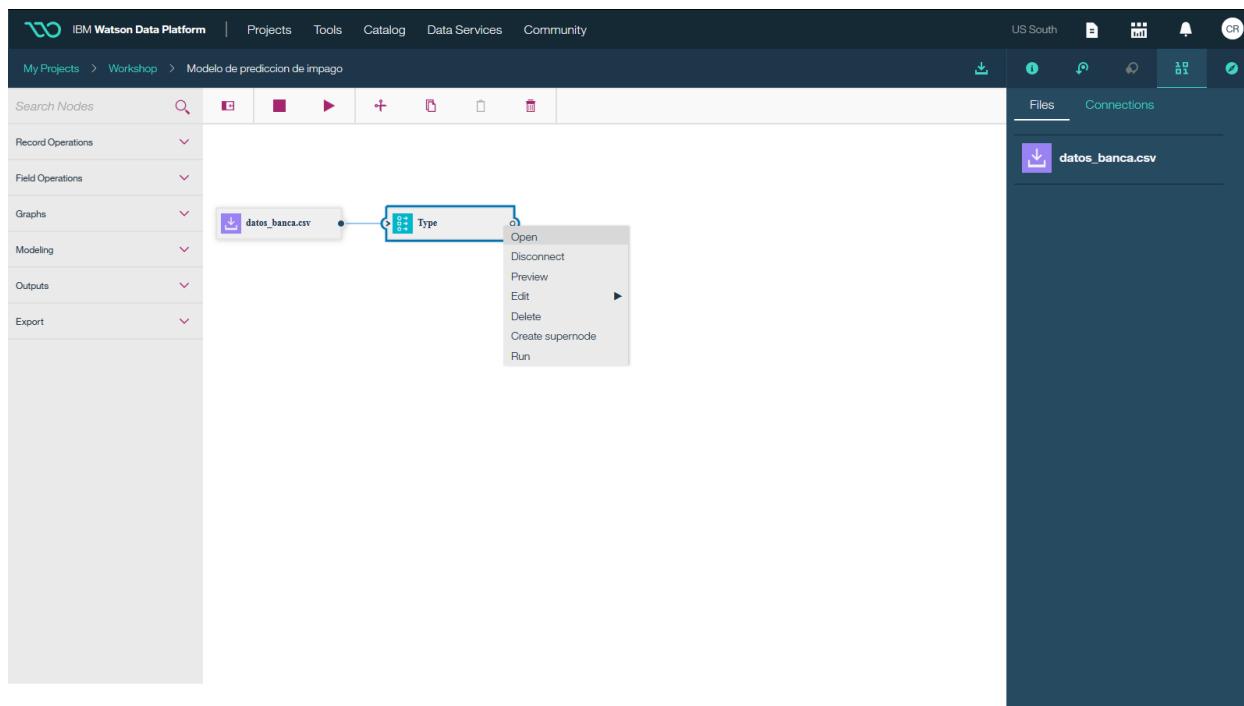
<https://dataplatform.ibm.com/docs/content/analyze-data/ml-canvas-spss.html?audience=wdp&context=analytics>

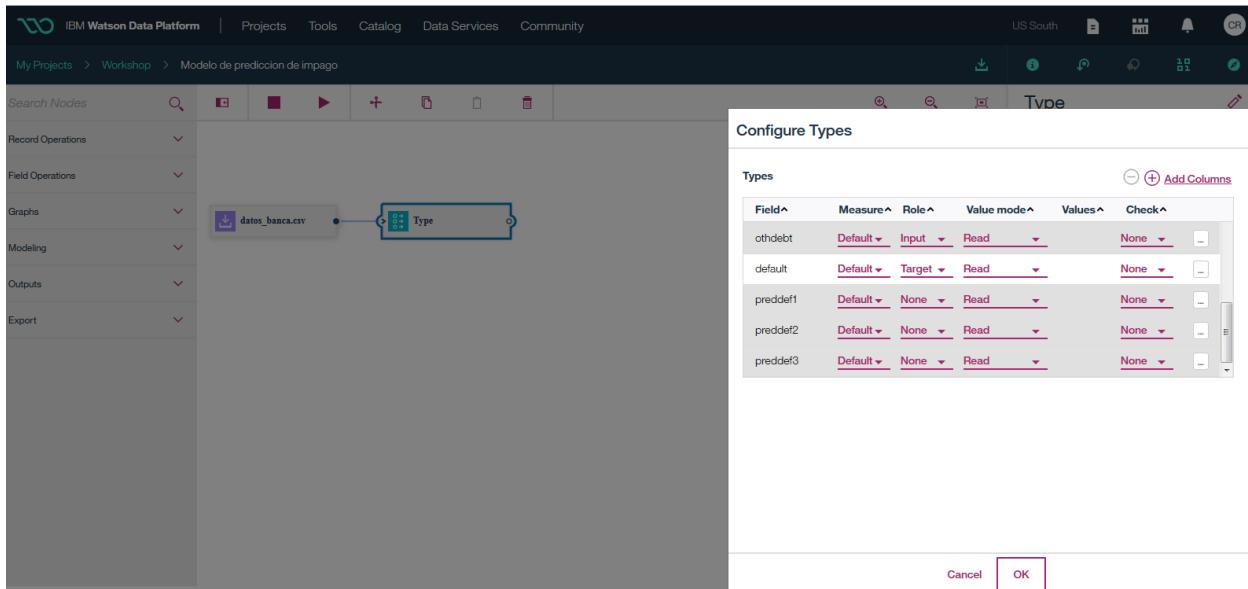
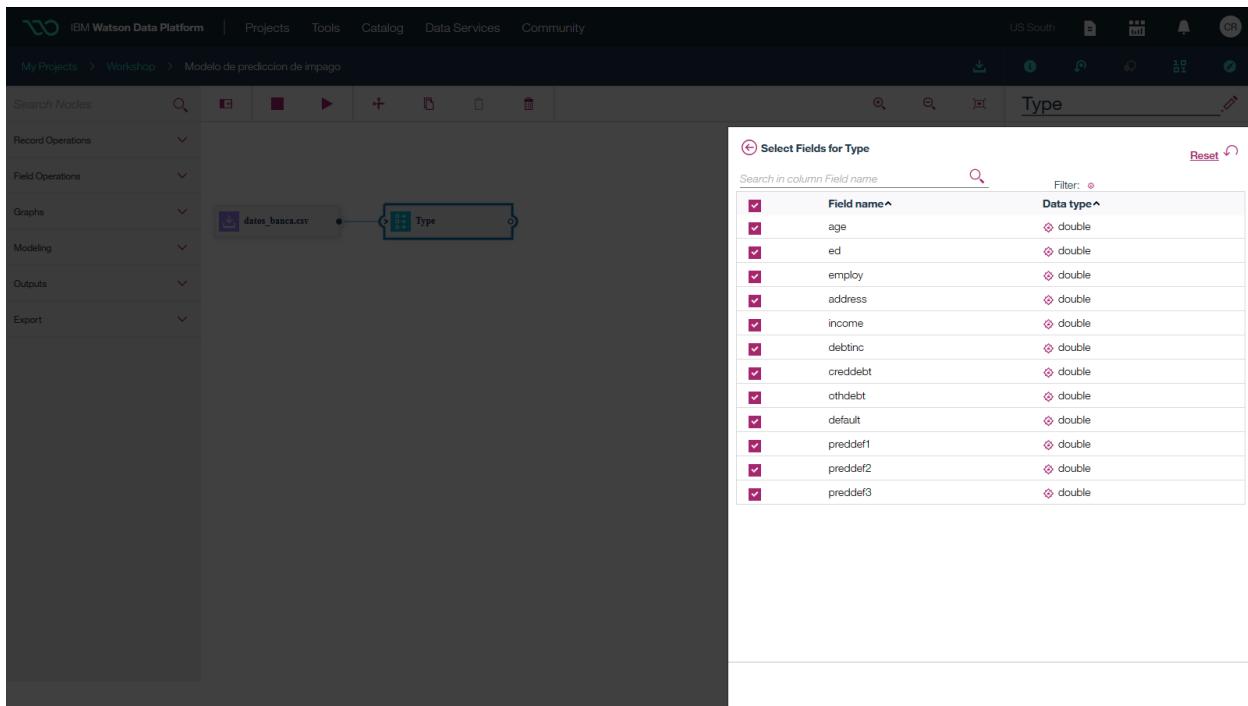




Para realizar el modelo, seleccionamos el fichero de entrada y vemos que incluso podemos añadir uno nuevo o conectarnos a una base de datos que tengamos. Arrastra el fichero al lienzo.

Y añade un nodo tipo, para instanciar los valores, y para seleccionar Default como **target**.





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.

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

My Projects > Workshop > Modelo de predicción de impago

Outputs Versions (12 fields, 10 records)

Auto Numeric  
C5.0  
C&R Tree  
CHAID  
GLE  
Linear  
Linear-AS  
Linear SVM  
Logistic  
Neural Net  
Random Trees  
K-Means  
Cox  
PCA/Factor  
Anomaly  
Feature Selection  
Tree-AS

Impago  
Arbol  
Impago  
default

C5 Tree Model

Model Information

Top Decision Rules

Tree Diagram

Tree Diagram ⓘ TARGET : default

Display labels on nodes  Display labels on branches

100%

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 Data Science Experience, proporciona un IDE para trabajar con R.

Una sesión de RStudio creada en Data Science Experience 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 'RStudio' tab is selected. On the left, there's a sidebar for 'Recently updated' projects: 'Workshop', 'SPSS Modeler', 'Data Refinery', and 'Streams Designer'. On the right, there's a table for 'COLLABORATORS' with columns for 'ROLE', 'COLLABORATORS', 'DATE CREATED', and 'LAST UPDATED'. A 'Get started' button is present, along with a 'New project' and 'New catalog' link. Below this, there's a 'Get started' section with a 'Create Catalog' button. The 'New in the community' section features four cards: 'Predicting Flight Cancellations Using...', 'Customer demographics and sales', 'Learn basics about notebooks and Apache Spark', and 'Continuous Learning on Watson Data Platform'.

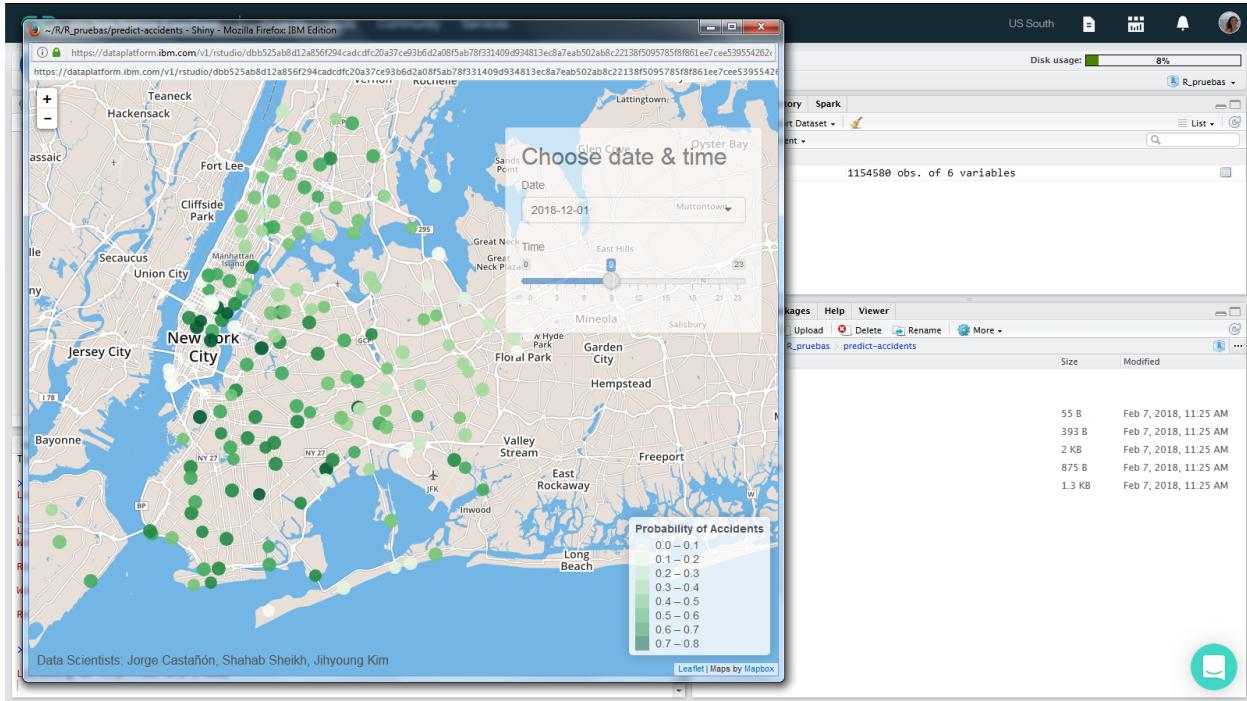
The bottom screenshot shows the RStudio interface within the IBM Watson Data Platform. The 'Console' tab is active, displaying the R version 3.3.2 startup message. The 'Files' tab is also visible, showing a directory structure for 'item-sparka-demos' with files like 'Rprofile', 'config.yml', 'item-sparka-demos', '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/dsx-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 Data Science Experience:

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 Data Science Experience 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](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](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](https://dataplatform.ibm.com/analytics/notebooks/d96fa67b-14f1-4db7-8b60-1af3c13699c3/view?access_token=c31fd96333af39811a78fe7773e421a50c7e20a450badb653bf4e0db39dc8f3f)

**Lasagne:**

→ Tutorial

[https://dataplatform.ibm.com/analytics/notebooks/c1bda39b-3fcf-4dae-a109-e71d11113633/view?access\\_token=18379e532a9953d4e97f2a75eee37a8ece9ee4745676e1a647493fbfd7b16fb](https://dataplatform.ibm.com/analytics/notebooks/c1bda39b-3fcf-4dae-a109-e71d11113633/view?access_token=18379e532a9953d4e97f2a75eee37a8ece9ee4745676e1a647493fbfd7b16fb)

**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](https://dataplatform.ibm.com/analytics/notebooks/39e93a50-cfc1-4097-b671-5261ba56e166/view?access_token=b7bd65f58805daf1f39465395dbb239c2f03d2cdeb611d8f413c81c7b1b06791)

Más información sobre Deep Learning en IBM Data Science Experience:

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

Una vez el lector tenga soltura con Data Science Experience y comprenda qué es el Deep learning, se recomienda este interesante proyecto: **Self Driving Car tutorials with Data Science Experience**. 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/aruizga7/Self-Driving-Car-in-DSX>

# Workshop 5.

## Recursos para trabajar en Data Science Experience en Local

Para todos aquellos que quieran aprender a utilizar Data Science Experience en Local:

[https://github.com/elenalowery/DSX\\_Local\\_Workshop](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](https://github.com/jc900/FastStart_DDLabs)