

BUSINESS & TECH SCHOOL UNIVERSIDAD EUROPEA CON IBM

Where business
meets technology.

Laboratorio : Primeros pasos con Watson Studio

María Borbonés García

@mariaborbones

Tabla de contenidos

Introducción al laboratorio.....	3
1- Desplegamos el servicio en IBM Cloud	4
2- Creamos y configuramos un proyecto en Watson Studio.....	6
3- Cargamos datos en nuestro proyecto.....	11
4- Preparamos los datos.....	14
5- Creamos nuestro primer notebook.....	22

Introducción al laboratorio

En este laboratorio, adquirirás los conocimientos necesarios para utilizar Watson Studio. Entender qué es un proyecto, qué tipo de assets puedo importar y los diferentes elementos de la plataforma Watson Studio.

Requisitos:

- Tener cuenta de IBM Cloud
- Acceso a Internet

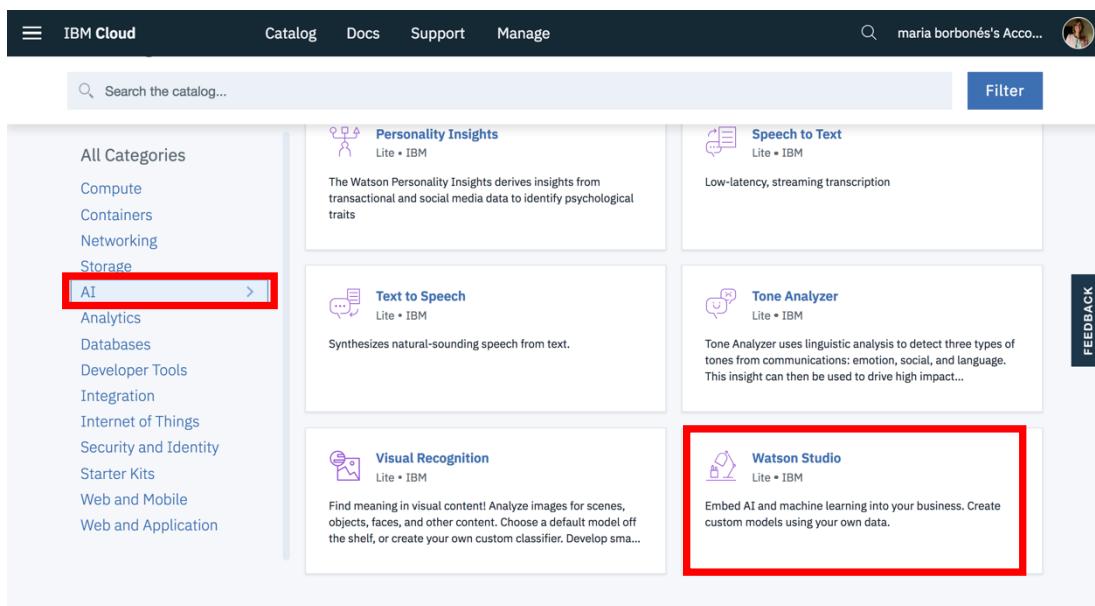
1- Desplegamos el servicio en IBM Cloud

Para crear el servicio de Watson Studio, necesitamos acceder al catálogo de **IBM Cloud** desde la siguiente URL: <https://console.cloud.ibm.com/>

Una vez hemos accedido a **IBM Cloud**, hacemos click en catálogo (barra superior) como se muestra en la imagen:



En el menú de la izquierda, donde se muestran todos los servicios de IBM Cloud, buscamos la categoría **Watson** debajo de **Todas las categorías** como se muestra en la imagen y hacemos click en ella:



Service	Description
Personality Insights	The Watson Personality Insights derives insights from transactional and social media data to identify psychological traits.
Speech to Text	Low-latency, streaming transcription.
Text to Speech	Synthesizes natural-sounding speech from text.
Tone Analyzer	Tone Analyzer uses linguistic analysis to detect three types of tones from communications: emotion, social, and language. This insight can then be used to drive high impact...
Visual Recognition	Find meaning in visual content! Analyze images for scenes, objects, faces, and other content. Choose a default model off the shelf, or create your own custom classifier. Develop sma...
Watson Studio	Embed AI and machine learning into your business. Create custom models using your own data.

En este caso, entre todo el conjunto de servicios, vamos a elegir desplegar **Watson Studio**, así que lo buscamos y hacemos click sobre el mismo.

En la siguiente pantalla, debemos asignarle un nombre al servicio (por ejemplo: labAI), una región (recomendable EEUU), y un grupo de recursos donde desplegarlo. Hacemos click en **Crear**.

IBM Cloud Catalog Docs Support Manage maria borbonés's Acco... 

[View all](#)

 **Watson Studio**
Lite • IBM

Watson Studio democratizes machine learning and deep learning to accelerate infusion of AI in your business to drive innovation. Watson Studio provides a suite of tools and a collaborative environment for data scientists, developers and domain experts.

Service name: CREATE

Choose a region/location to deploy in: Select a resource group: FEEDBACK

[View Docs](#) [Terms](#)

AUTHOR	IBM
PUBLISHED	09/18/2018
TYPE	Service

Features

- Use what you know, learn what you don't Start from a tutorial, start from a sample, or start from scratch. Tap into the power of the best of open source (RStudio, Jupyter Notebooks) and Watson services for flexible model creation. Use Python, R, or Scala. Stop downloading and configuring analysis environments and start creating insights.
- Power on demand Enterprise-scale features on demand. From data exploration and preparation, to enterprise-scale performance. Manage your data, your analytical assets, and your projects in a secured cloud environment.

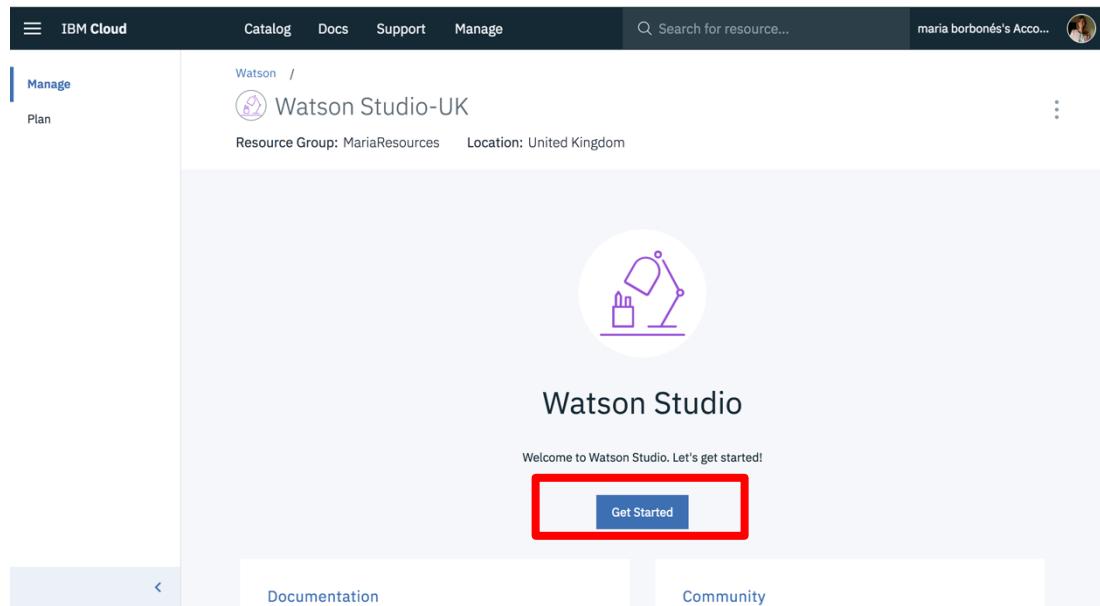
Need Help? [Contact IBM Cloud Support](#) [Estimate Monthly Cost](#) [Cost Calculator](#)

iEnhorabuena! Has completado la primera parte del laboratorio. Ya sabes cómo desplegar un servicio de Watson en IBM Cloud.

2- Creamos y configuramos un proyecto en Watson Studio

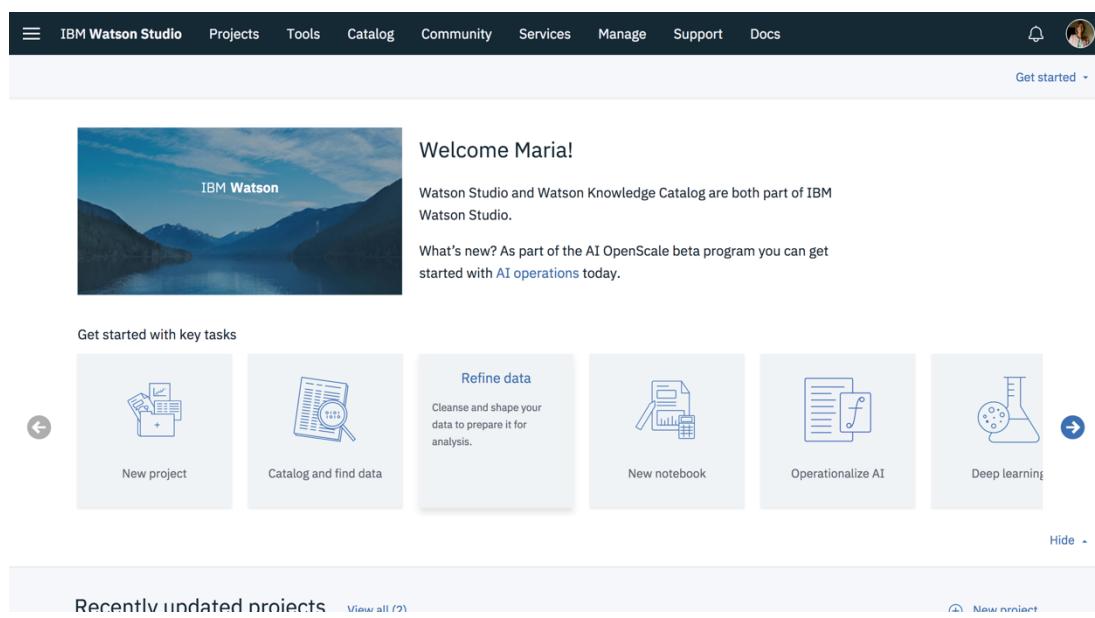
Watson Studio permite democratizar machine learning y deep learning para acelerar la introducción de la Inteligencia Artificial en tu negocio. Provee, además, un conjunto de herramientas que permite trabajar de forma colaborativa a los científicos de datos, desarrolladores y SMEs.

Para demostrar las capacidades de Watson Studio vamos a crear un proyecto. Hacemos click en **Get Started**



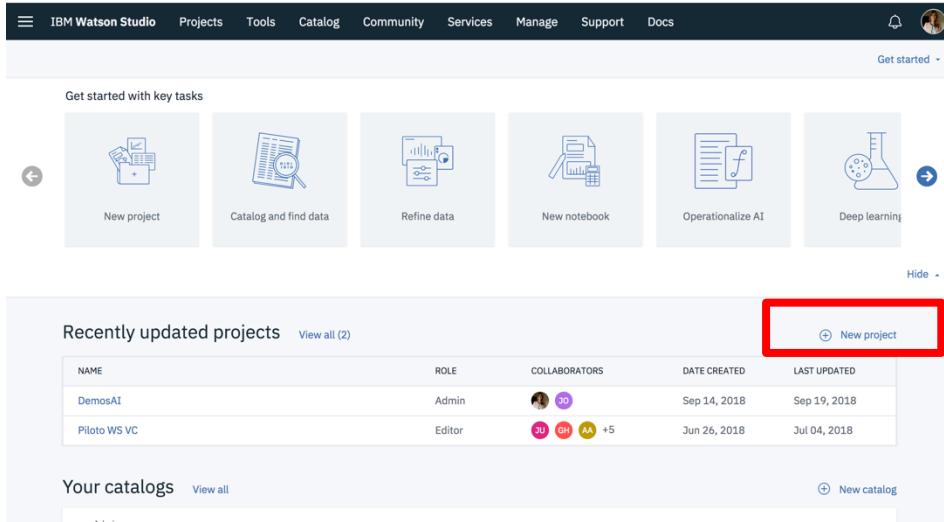
The screenshot shows the Watson Studio UK interface. At the top, there's a navigation bar with 'IBM Cloud', 'Catalog', 'Docs', 'Support', 'Manage', a search bar, and a user profile. Below the navigation is a sidebar with 'Manage' and 'Plan' options. The main content area displays 'Watson Studio-UK' under 'Watson /'. It shows 'Resource Group: MariaResources' and 'Location: United Kingdom'. A central feature is a circular icon with a person working at a desk. Below it, the text 'Watson Studio' and 'Welcome to Watson Studio. Let's get started!' is followed by a prominent blue 'Get Started' button, which is outlined in red. At the bottom of the main area are links for 'Documentation' and 'Community'.

Y accedemos a la consola de **Watson Studio**



The screenshot shows the IBM Watson Studio console. The top navigation bar includes 'IBM Watson Studio', 'Projects', 'Tools', 'Catalog', 'Community', 'Services', 'Manage', 'Support', 'Docs', and a user profile. A 'Get started' button is located in the top right corner. The main area features a large image of a lake with mountains, labeled 'IBM Watson'. To the right, a welcome message says 'Welcome Maria!' and notes that Watson Studio and Watson Knowledge Catalog are part of IBM Watson Studio. It also mentions the AI OpenScale beta program. Below this, a section titled 'Get started with key tasks' lists five items: 'New project', 'Catalog and find data', 'Refine data' (with a sub-note about preparing data for analysis), 'New notebook', 'Operationalize AI', and 'Deep learning'. At the bottom, there's a section for 'Recently updated projects' with a 'View all' link and a note about a new project.

Ahora vamos a crear un nuevo proyecto. Así que hacemos click en **New Project**



Recently updated projects [View all \(2\)](#)

NAME	ROLE	COLLABORATORS	DATE CREATED	LAST UPDATED
DemosAI	Admin		Sep 14, 2018	Sep 19, 2018
Piloto WS VC	Editor		Jun 26, 2018	Jul 04, 2018

Your catalogs [View all](#)

Elegimos el tipo de proyecto que queremos crear. Esta opción nos permite crear un tipo de proyecto que se adapte a nuestras necesidades, pero no te preocupes, que más tarde puedes configurar los elementos existentes en tu proyecto. Los tipos de proyectos posibles son:

Basic: te permite empezar subiendo assets y más tarde añade las herramientas que necesites

Data Science: Analiza tus datos para inferir conocimiento y compartirlo con tus compañeros

Visual Recognition: Etiqueta y clasifica contenido utilizando los servicios de Watson

Deep Learning: Etiqueta y clasifica contenido utilizando los servicios de Watson

Modeler: Construye flujos de modeler para entrenar modelos de SPSS y Spark

Business Analytics: Crea visualizaciones para ganar conocimiento de tus datos mucho más rápido

Data Engineering : Combina, limpia, analiza, y da forma a tus datos utilizando Data Refinery.

Streams Flow : Ingesta y analiza datos en tiempo real utilizando el servicio de Streaming Analytics.

Create a project

Choose the project starter for your work. Required services with Lite plans are provisioned automatically. You can add other assets and services later.

Standard Work with any type of asset. Add services for analytical assets as you need them. ASSETS All	Data Science Analyze data to discover insights and share your findings with others. ASSETS Data • Notebooks	Visual Recognition Tag and classify visual content using the Watson Visual Recognition service. ASSETS Data • Visual recognition model	Deep Learning Build neural networks and deploy deep learning models. ASSETS Data • Modeler flow • Model • Experiment	Modeler Build modeler flows to train SPSS models or design deep neural networks. ASSETS Data • Modeler Flow • Model • Experiment
Business Analytics Create visual dashboards from your data to gain insights faster. ASSETS Data • Dashboard	Data Engineering Combine, cleanse, analyze, and shape data using Data Refinery. ASSETS Data • Data Refinery flow	Streams Flow Ingest and analyze streaming data using the Streaming Analytics service. ASSETS Data • Streams flow		

En este caso como queremos probar toda la funcionalidad posible de Watson Studio, elegimos ‘Standard’.

En la siguiente pantalla se nos solicita un nombre y asociar un **Cloud Object Storage** al proyecto para poder almacenar todos los assets que necesitemos para trabajar.

Cloud Object Storage es un servicio altamente escalable de almacenamiento de objetos en la nube, diseñado para alta durabilidad, resiliencia y seguridad. Almacena, gestiona y accede a tus datos directamente desde un portal o a través de APIs REST.

COS permite almacenar grandes volúmenes de datos mientras aseguramos la seguridad de los mismos, disponibilidad y confiabilidad. Utiliza algoritmos IDAs para separar los datos en particiones que son distribuidas en diferentes data centers de forma encriptada.

Gracias a la tecnología de Aspera, se permite cargas de ficheros a velocidades mucho más rápidas.

Además nuestro servicio **IBM SQL Query** permite utilizar standard SQL para analizar datos almacenados en Object Storage.

Las comunicaciones entre **IBM Watson Studio** y **Cloud Object Storage** son realizadas mediante un endpoint privado y por tanto no se tarifica el ancho de banda de las operaciones realizadas entre ambos.

Puedes encontrar más información en <https://console.bluemix.net/catalog/services/cloud-object-storage>

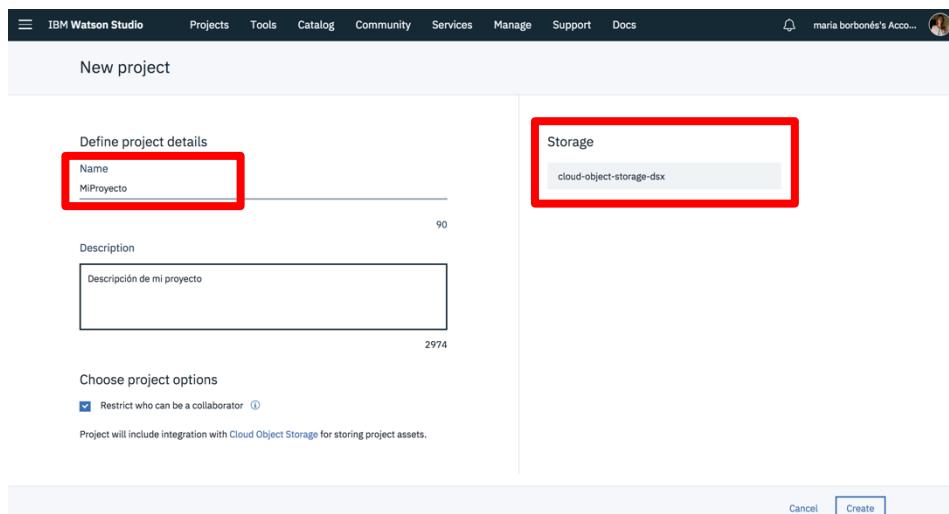
Se recomienda el uso de Project-lib para interactuar con COS desde notebooks

<https://dataplatform.cloud.ibm.com/docs/content/analyze-data/publish-library-documentation.html?context=analytics>

Para acceder a COS desde Python podéis ver este enlace

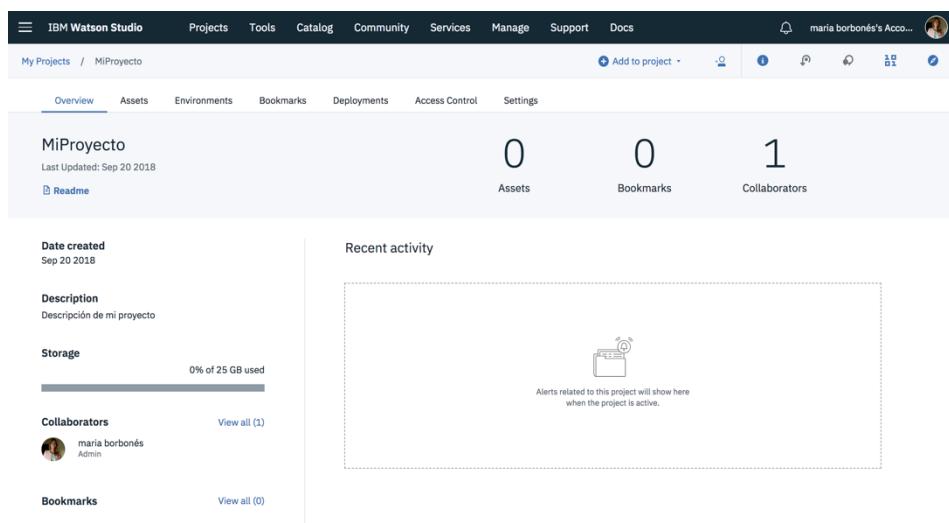
<https://medium.com/ibm-data-science-experience/working-with-ibm-cloud-object-storage-in-python-fe0ba8667d5f>

Elegimos por tanto el nombre y creamos un servicio de **Cloud Object Storage**, siguiendo los pasos que nos indica **IBM Watson Studio**.



The screenshot shows the 'Define project details' section. The 'Name' field is set to 'MiProyecto'. The 'Storage' field is set to 'cloud-object-storage-dsx'. Below these fields are sections for 'Description' (with placeholder 'Descripción de mi proyecto') and 'Choose project options' (with a checked checkbox for 'Restrict who can be a collaborator'). A note below says 'Project will include integration with Cloud Object Storage for storing project assets.' At the bottom right are 'Cancel' and 'Create' buttons.

Una vez finalizado el proceso de configuración, ya se nos da acceso de forma automática a la página de Overview de nuestro proyecto.



The screenshot shows the 'Overview' page for the 'MiProyecto' project. The project summary at the top shows 0 Assets, 0 Bookmarks, and 1 Collaborator (maria borbonés). The 'Recent activity' section is empty. On the left, there are sections for 'Date created' (Sep 20 2018), 'Description' (Descripción de mi proyecto), 'Storage' (0% of 25 GB used), 'Collaborators' (maria borbonés), and 'Bookmarks' (View all (0)).

El siguiente paso para terminar de configurar nuestro proyecto sería añadir al resto de colaboradores. Watson Studio se tarifica por número de usuarios autorizados que tengan el rol de Editor o Admin pero en cambio los usuarios tipo viewer son gratuitos e ilimitados. A continuación, se detallan cada uno de los roles.

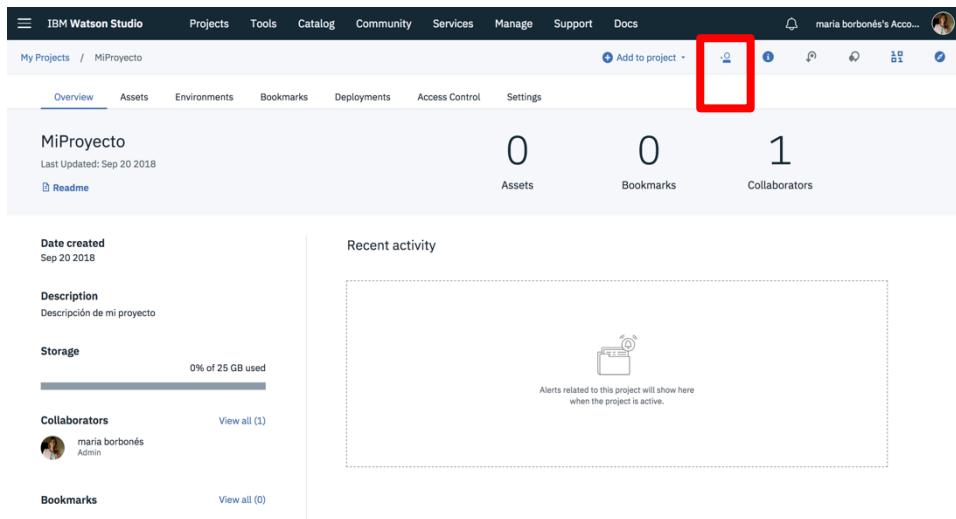
Admin: Tiene control total sobre el proyecto, assets y colaboradores

Editor: Puede añadir y utilizar los assets

Viewer: Puede visualizar los assets

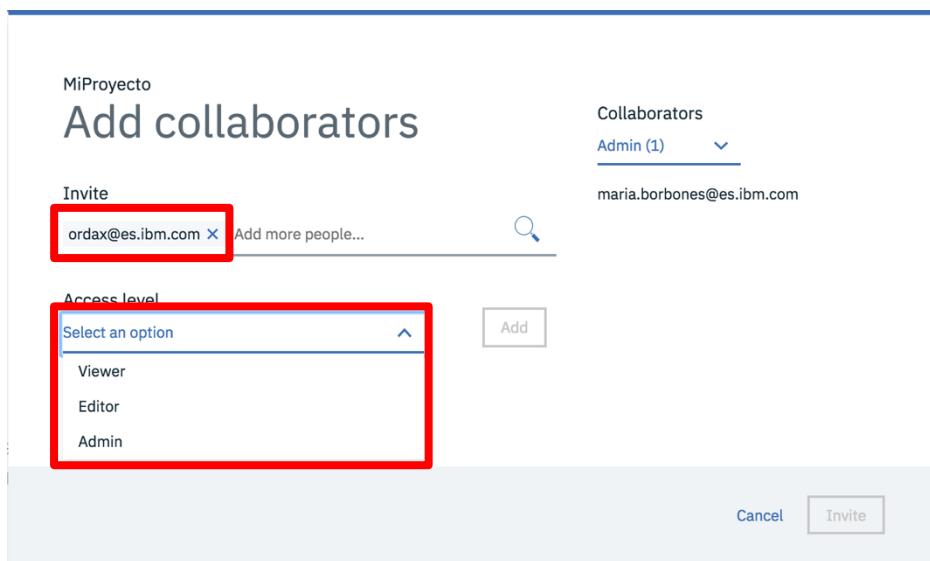
Más información en <https://dataplatform.cloud.ibm.com/docs/content/getting-started/roles.html>

Para ello, accedemos al ícono en la barra superior con forma de persona y hacemos click sobre él.



The screenshot shows the 'My Projects' section of the IBM Watson Studio interface. A specific project titled 'MiProyecto' is selected. At the top right of the project card, there is a user icon with a red box drawn around it. Below the project title, it says 'Last Updated: Sep 20 2018'. The project summary includes metrics: 0 Assets, 0 Bookmarks, and 1 Collaborators. On the left, there are sections for Date created (Sep 20 2018), Description (Descripción de mi proyecto), Storage (0% of 25 GB used), Collaborators (maria borbonés Admin), and Bookmarks (View all (0)). On the right, there is a 'Recent activity' section which is currently empty.

El siguiente paso es introducir los correos de aquellas personas que queramos tener como colaboradores y asignarles un rol a cada uno de ellos como se muestra en la imagen.

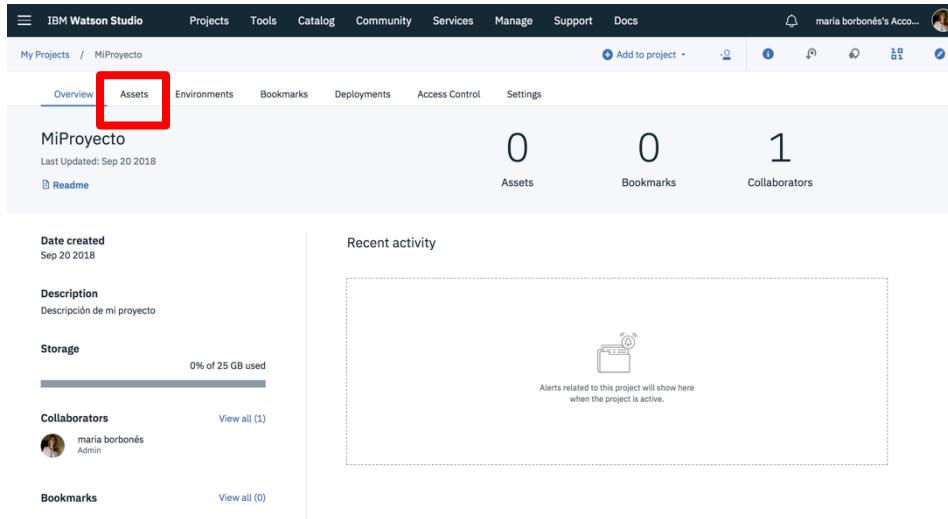


The screenshot shows the 'Add collaborators' dialog box. In the 'Invite' section, the email 'ordax@es.ibm.com' is entered into the input field, which is highlighted with a red box. Below this, there is a search icon. To the right, under 'Collaborators', it shows 'Admin (1)' and the email 'maria.borbones@es.ibm.com'. In the 'Access level' section, a dropdown menu is open, showing options: 'Select an option' (highlighted with a red box), 'Viewer', 'Editor', and 'Admin'. At the bottom right of the dialog are 'Cancel' and 'Invite' buttons.

¡Enhorabuena! Ya sabes como crear tu primer proyecto en Watson Studio y cómo configurarlo para empezar a trabajar.

3- Cargamos datos en nuestro proyecto

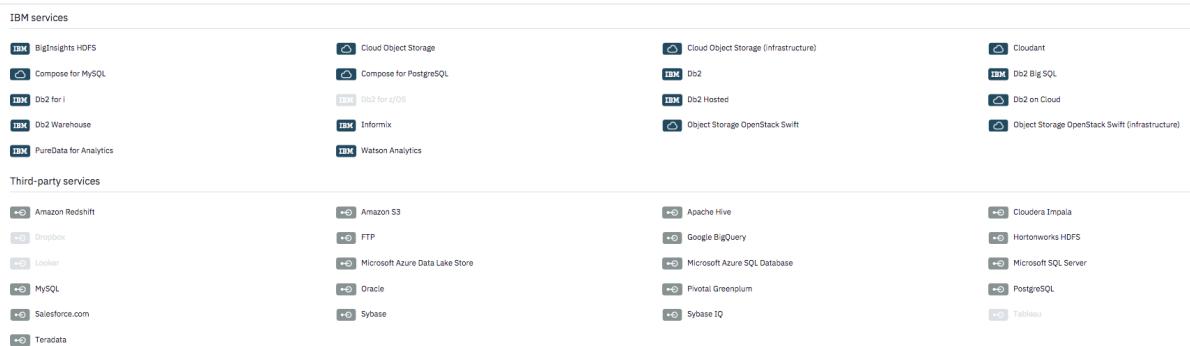
Vamos a empezar a trabajar con nuestro proyecto, y el primer paso es cargar datos para que podamos trabajar con ellos. Pero antes, necesitamos navegar a la pestaña de **Assets** de nuestro proyecto. Hacemos click en **Assets** en la misma página donde nos encontramos.



The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with links for Projects, Tools, Catalog, Community, Services, Manage, Support, and Docs. On the far right, it shows 'maria borbonés's Acco...' and a profile icon. Below the navigation is a breadcrumb trail: 'My Projects / MiProyecto'. The main content area has tabs for Overview, Assets (which is highlighted with a red box), Environments, Bookmarks, Deployments, Access Control, and Settings. Under the Overview tab, there's a summary for 'MiProyecto' with 'Last Updated: Sep 20 2018'. It shows 0 Assets, 0 Bookmarks, and 1 Collaborator. Below this, there are sections for Date created (Sep 20 2018), Description (Descripción de mi proyecto), Storage (0% of 25 GB used), Collaborators (maria borbonés, Admin), and Bookmarks (View all (0)). To the right, there's a 'Recent activity' section with a placeholder message: 'Alerts related to this project will show here when the project is active.' There's also a small icon of a document with an upward arrow.

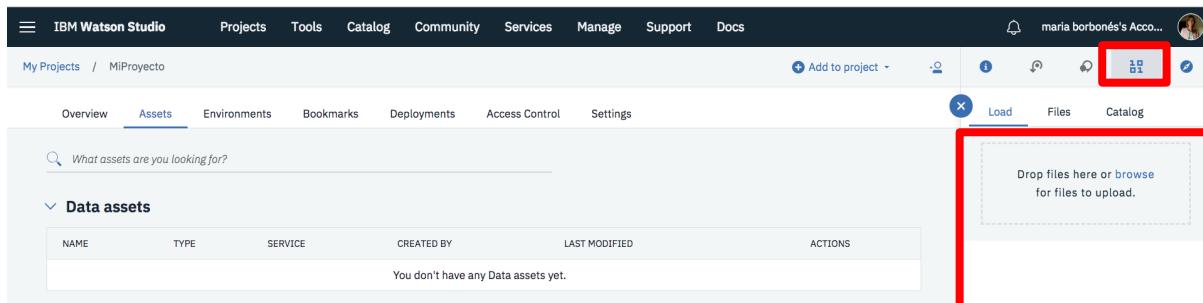
Veremos que aún no tenemos ningún Data Asset, pero lo vamos a solucionar muy rápido. IBM Watson Studio ya dispone de conectores para trabajar con algunas fuentes de datos, pero también nos permite cargar directamente nuestros propios ficheros.

Por simplicidad par este laboratorio, vamos a elegir esta última opción, pero aquí tenéis un listado de las diferentes opciones de las que disponéis, tanto de IBM como de terceros.



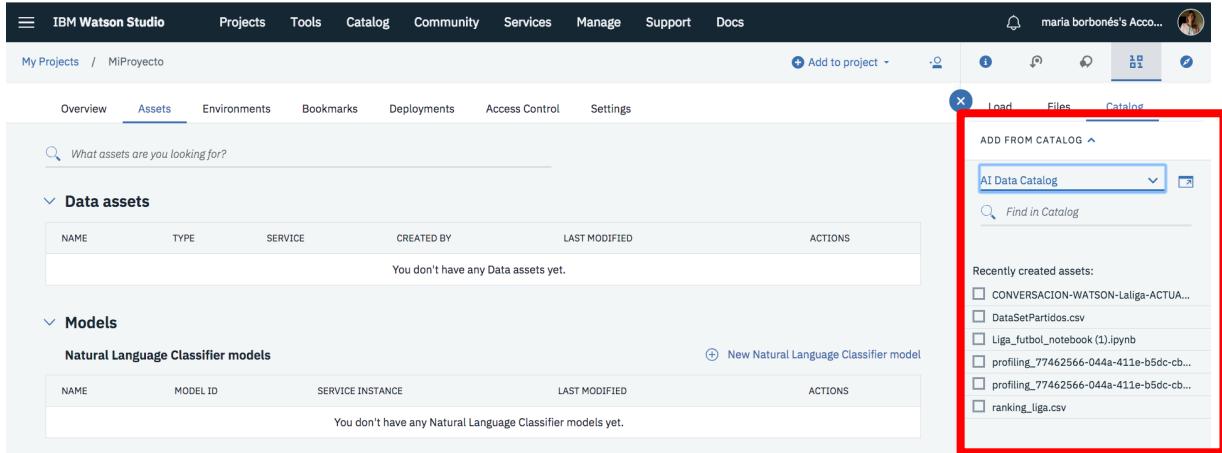
This screenshot shows the 'Catalog' section of the IBM Watson Studio interface. It's divided into two main sections: 'IBM services' and 'Third-party services'. In the 'IBM services' section, there are icons for BigInsights HDFS, Compose for MySQL, Db2 for i, Db2 Warehouse, and PureData for Analytics. In the 'Cloud Object Storage' section, there are icons for Cloud Object Storage and Cloud Object Storage (Infrastructure). In the 'Db2' section, there are icons for Db2 and Db2 Hosted. In the 'Object Storage OpenStack Swift' section, there is an icon for Object Storage OpenStack Swift (Infrastructure). In the 'Third-party services' section, there are icons for Amazon Redshift, Dropbox, Locker, MySQL, Salesforce.com, Teradata, Amazon S3, FTP, Microsoft Azure Data Lake Store, Oracle, Sybase, Apache Hive, Google BigQuery, Microsoft Azure SQL Database, Pivotal Greenplum, and Sybase IQ. There are also icons for Cloudera Impala, Hortonworks HDFS, Microsoft SQL Server, PostgreSQL, and Tableau.

Para cargar nuestro fichero, en este caso un **CSV** vamos a hacer click en el icono con unos y ceros de la parte superior, y a través del explorador de archivos vamos a seleccionar el fichero **(customer_churn)** que es el que vamos a utilizar en este ejercicio. Podéis encontrarlo en la carpeta **Data del lab.**



This screenshot shows the 'Assets' page of the IBM Watson Studio interface. At the top, there's a navigation bar with links for Projects, Tools, Catalog, Community, Services, Manage, Support, and Docs. On the far right, it shows 'maria borbonés's Acco...' and a profile icon. Below the navigation is a breadcrumb trail: 'My Projects / MiProyecto'. The main content area has tabs for Overview, Assets (selected and highlighted with a red box), Environments, Bookmarks, Deployments, Access Control, and Settings. Under the Assets tab, there's a search bar with the placeholder 'What assets are you looking for?'. Below it, there's a section titled 'Data assets' with a table header: NAME, TYPE, SERVICE, CREATED BY, LAST MODIFIED, and ACTIONS. The table body is empty with the message 'You don't have any Data assets yet.'. To the right of the table, there's a large red-bordered box containing the text 'Drop files here or browse for files to upload.'

Como ya os habrán explicado durante la sesión teórica, también disponemos de un catálogo de datos llamado **Watson Knowledge Catalog** y podríamos acceder a todos los assets a los que se nos hubiera concedido permiso para acceder de los diferentes catálogos.



The screenshot shows the IBM Watson Studio interface with the 'Catalog' tab selected. A modal window titled 'ADD FROM CATALOG' is open, containing a search bar with 'AI Data Catalog' and a list of recently created assets:

- CONVERSACION-WATSON-Laliga-ACTUA...
- DataSetPartidos.csv
- Liga_futbol_notebook (1).ipynb
- profiling_77462566-044a-411e-b5dc-cb...
- profiling_77462566-044a-411e-b5dc-cb...
- ranking_liga.csv

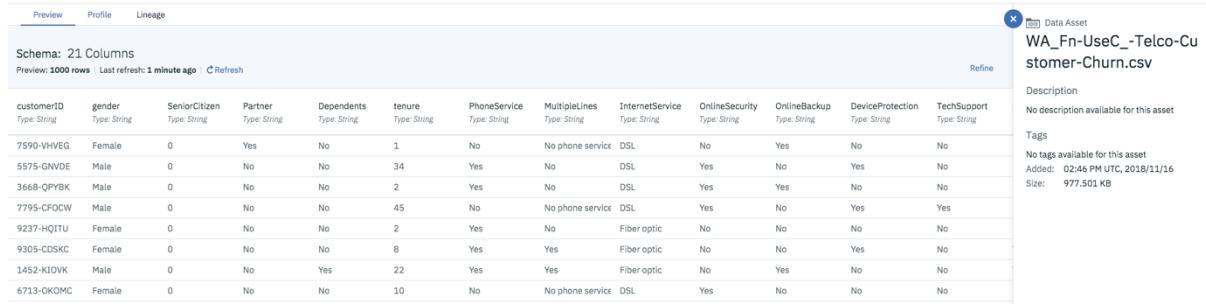
Una vez tenemos nuestro data asset listo en nuestro proyecto:



The screenshot shows the 'Assets' tab selected in the navigation bar. Under the 'Data assets' section, there is a single asset listed:

NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
WA_Fn-UseC_-Telco-Customer-Churn.csv	Data Asset	maria borbonés	16 Nov 2018, 3:46:15 pm	⋮

Podemos hacer click sobre él y ver más información sobre ese fichero, como una previsualización de las primeras 1000 filas.

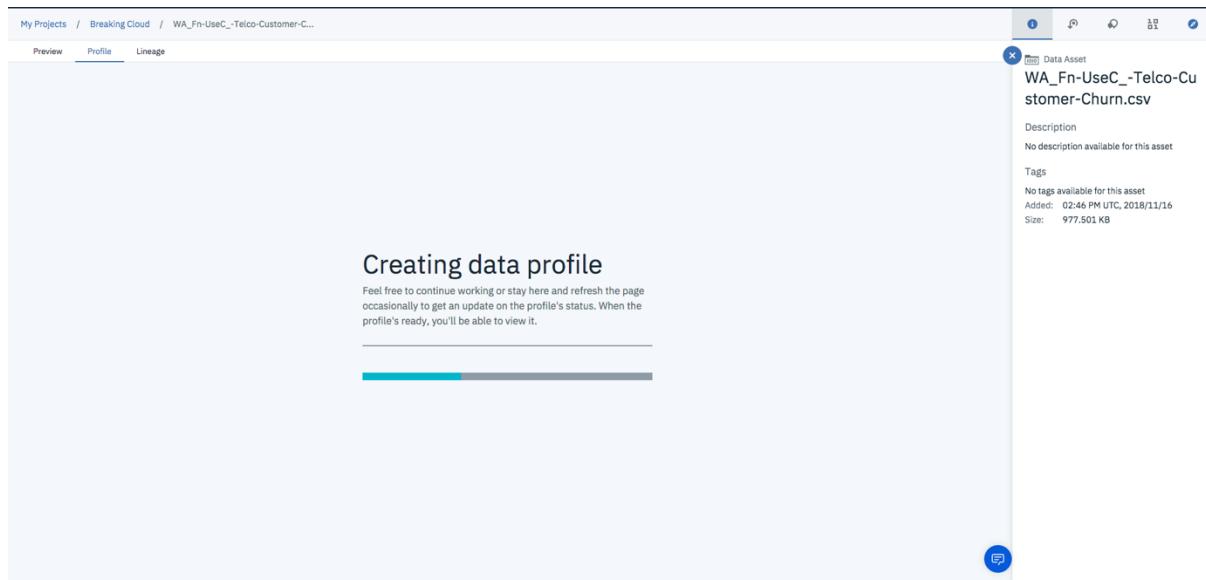


The screenshot shows the 'Profile' tab selected for the data asset. The preview table has 21 columns and 1000 rows. The first few rows of data are:

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport
7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No	No
5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No
7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes	Yes
9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No
9305-CDSKC	Female	0	No	No	8	Yes	Fiber optic	No	No	Yes	No	No
1452-KIOVK	Male	0	No	Yes	22	Yes	Fiber optic	No	Yes	No	No	No
6713-OKOMC	Female	0	No	No	10	No	No phone service	DSL	Yes	No	No	No

Si hacemos click en profile, podemos hacer click en **Create profile**, para que se inicie el proceso de perfilado de los datos y ayudarnos a entender mejor este data asset. En este caso es una tabla y nos mostrará medias, frecuencias... pero si trabajaramos con otros ficheros como PDFs, nos mostraría el resultado del procesamiento del lenguaje natural (NLP) tal como entidades, conceptos, palabras claves...

El proceso de creación del perfil puede tardar unos minutos... Podéis seguir trabajando y volver más tarde para ver el resultado



Creating data profile

Feel free to continue working or stay here and refresh the page occasionally to get an update on the profile's status. When the profile's ready, you'll be able to view it.

WA_Fn-UseC_-Telco-Customer-Churn.csv

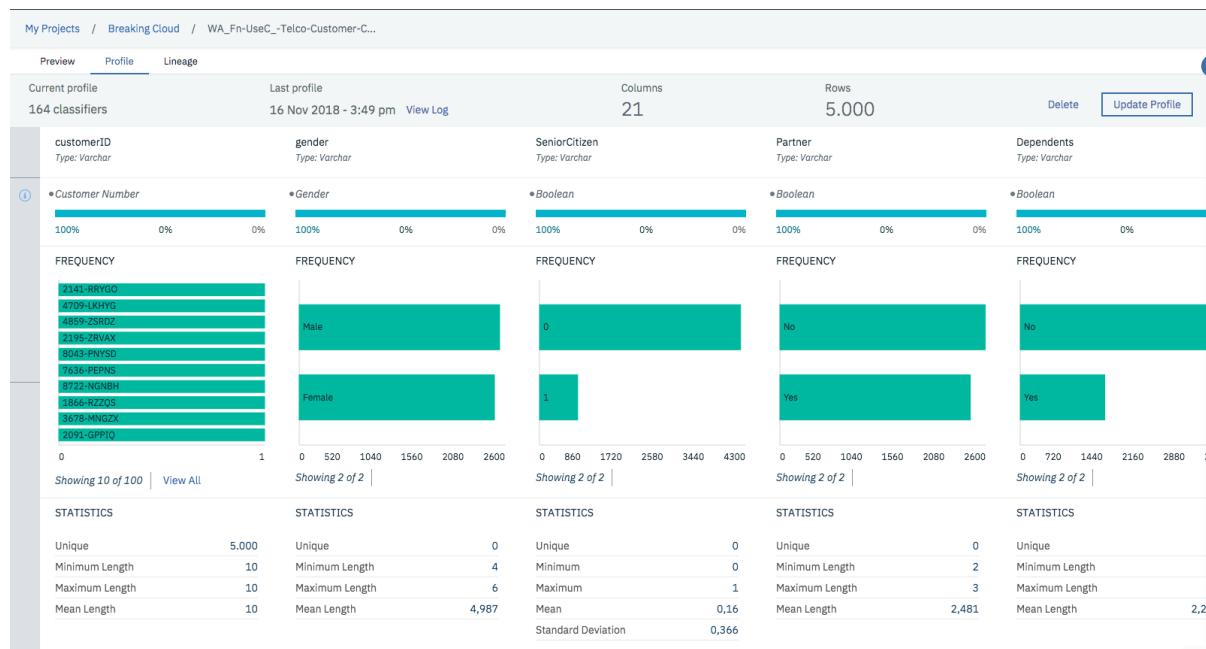
Description: No description available for this asset

Tags: No tags available for this asset

Added: 02:46 PM UTC, 2018/11/16

Size: 977.501 KB

El resultado obtenido



También hay una opción de linaje (no disponible para cuentas gratuitas) que permite ver quién, en qué momento y qué operaciones se han realizado sobre los diferentes datasets.

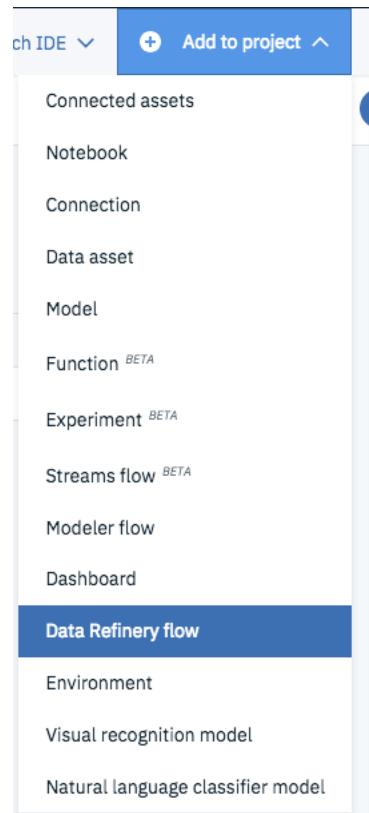
The screenshot shows the Watson Studio Lineage interface. At the top, there are tabs: Overview, Access, Review, Profile, and Lineage. The Lineage tab is selected. Below the tabs, there's a summary for a project named "Transportation Cat...". It shows 0 assets created from, 0 assets added to projects, and the last event was an "Asset downloaded" a few moments ago. To the right, a detailed view of the "Asset downloaded" event is shown. The event occurred on Aug 29, 2018 at 1:30 PM and was downloaded by SHARYN RICHARD. There are also icons for adding new assets and viewing the event.

Para volver a la vista de **Assets** de nuestro proyecto, simplemente hacemos click en la barra superior en el nombre de nuestro proyecto.

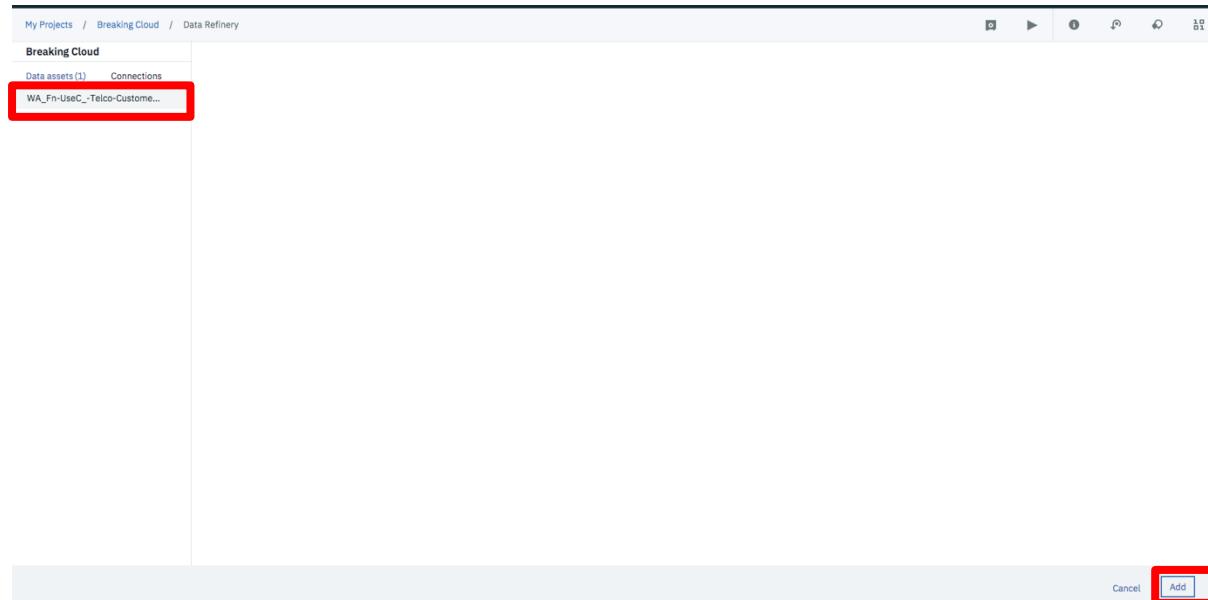
iEnhorabuena! Ya sabes como ingestar datos en Watson Studio

4- Preparamos los datos

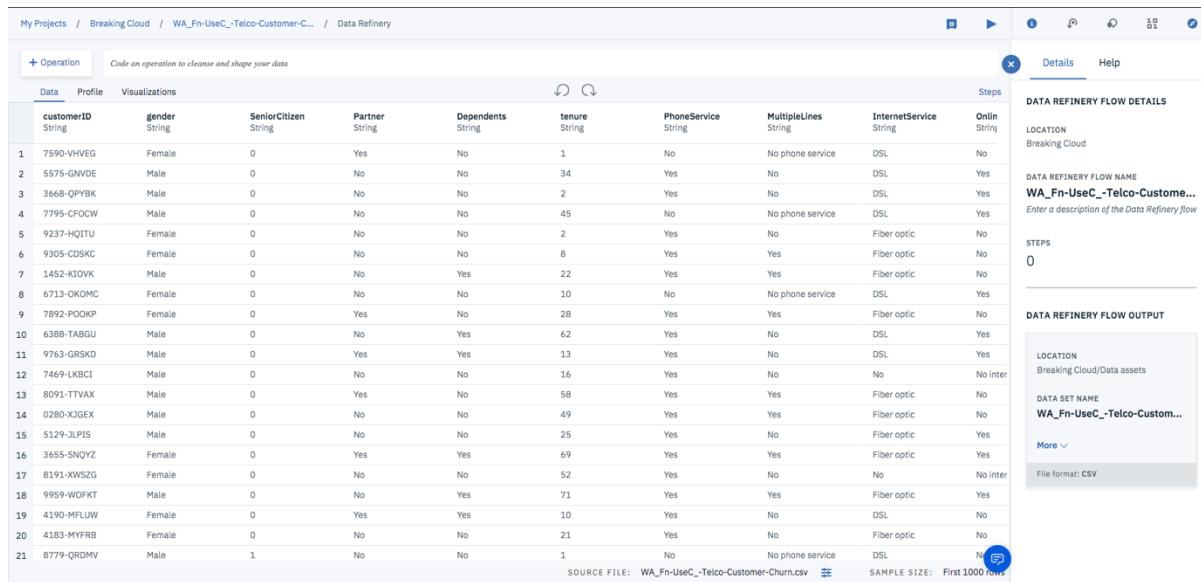
El siguiente paso es preparar los datos para que el científico de datos pueda trabajar con ellos. Hacemos click en **+Add to project** y hacemos click en **Data Refinery Flow**



Seleccionamos el asset que acabamos de cargar en nuestro proyecto (**customer-churn**) y hacemos click en **Add**

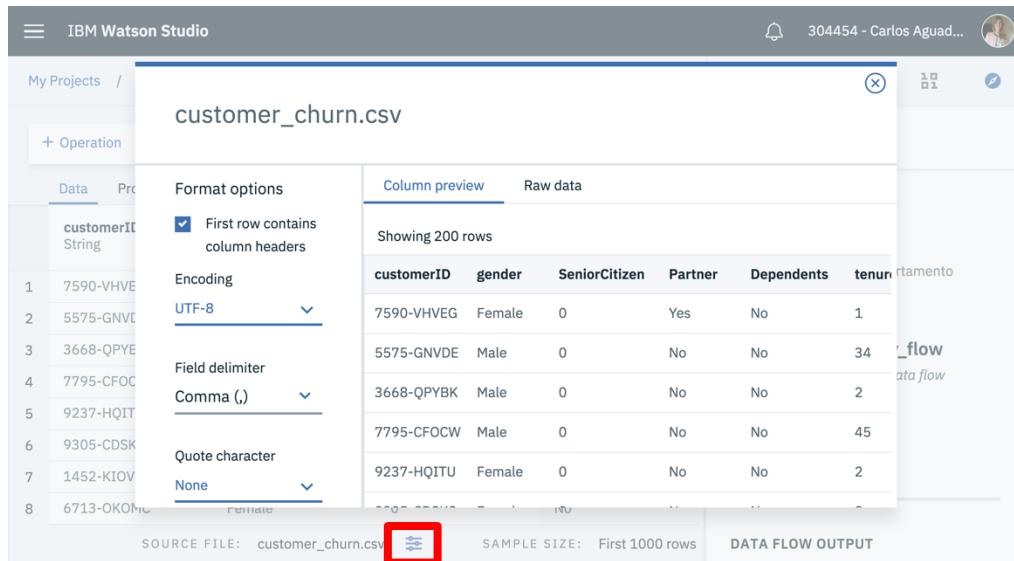


Veremos que se carga nuestra tabla en Data Refinery y ya estamos listos para empezar a realizar operaciones sobre nuestros datos.



The screenshot shows the IBM Watson Studio Data Refinery interface. At the top, there's a navigation bar with 'My Projects' (selected), 'Breaking Cloud', 'WA_Fn-UseC_-Telco-Customer-C...', and 'Data Refinery'. Below the navigation is a toolbar with icons for 'Operation', 'Code', 'Profile', 'Visualizations', and 'Details' (selected). The main area displays a data preview of 'customer_churn.csv' with 21 rows and various columns like 'customerID', 'gender', 'SeniorCitizen', etc. To the right, there are two panels: 'DATA REFINERY FLOW DETAILS' and 'DATA REFINERY FLOW OUTPUT'. The 'DATA REFINERY FLOW DETAILS' panel shows the flow name 'WA_Fn-UseC_-Telco-Customer-Churn' and step count '0'. The 'DATA REFINERY FLOW OUTPUT' panel shows the location 'Breaking Cloud/Data assets', dataset name 'WA_Fn-UseC_-Telco-Customer-Churn', and file format 'CSV'.

Podemos especificar el formato de nuestros datos si hacemos click como se muestra en la figura:



The screenshot shows the 'Format options' section for 'customer_churn.csv'. It includes settings for 'First row contains column headers' (checked), 'Encoding' (UTF-8), 'Field delimiter' (Comma (,), highlighted with a red box), 'Quote character' (None), and 'Separator' (Tab). To the right, there are tabs for 'Column preview' (selected) and 'Raw data'. The 'Column preview' tab shows a preview of the first 200 rows of the data. At the bottom, there are buttons for 'SOURCE FILE: customer_churn.csv' and 'SAMPLE SIZE: First 1000 rows', followed by a 'DATA FLOW OUTPUT' section.

Empezamos a realizar operaciones sobre nuestros datos. Data Refinery tiene soporte para operaciones de la librería dplyr R a nivel scripting

https://dataplatform.cloud.ibm.com/docs/content/refinery/code_operations.html?audience=wdp

y a nivel interfaz de usuario se soportan las siguientes operaciones:

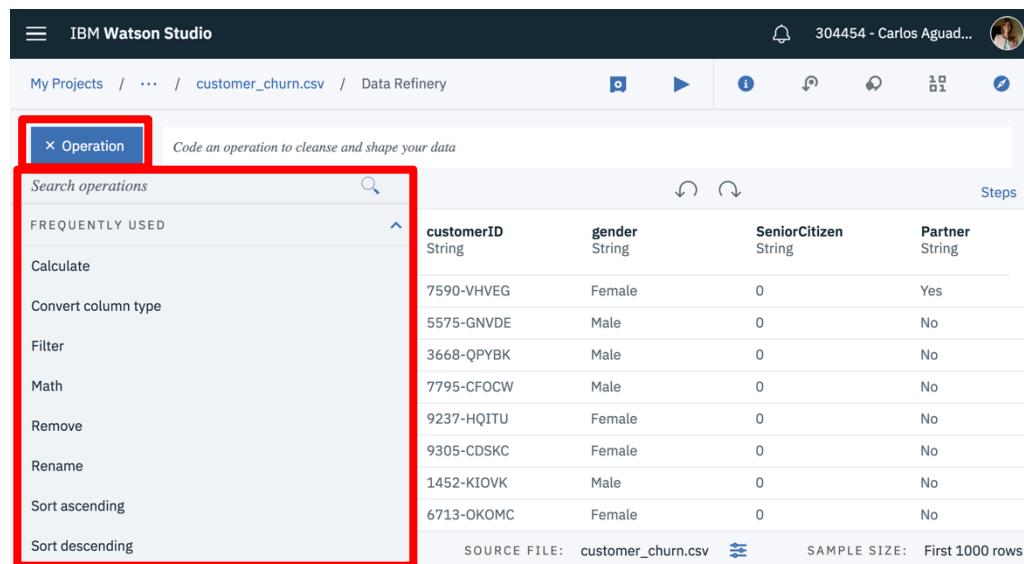
- [Frequently used](#)
- [Cleanse](#)
- [Math](#)
- [Natural language](#)

- [Organize](#)
- [Text](#)
- [Column menu](#)

Más detalle en:

https://dataplatform.cloud.ibm.com/docs/content/refinery/gui_operations.html?audience=wdp

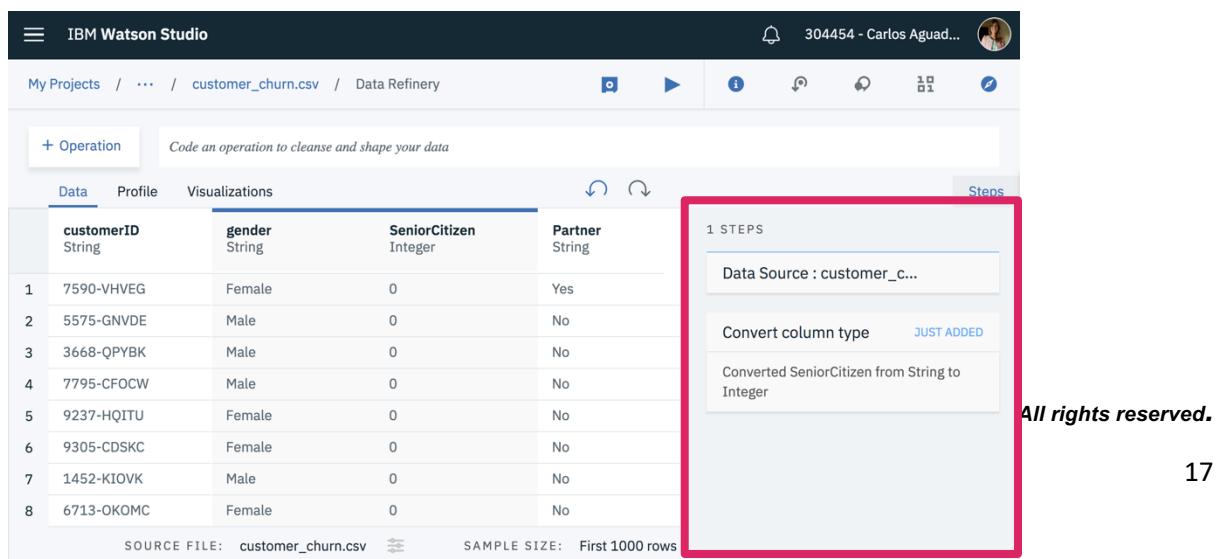
Hacemos click en + Operation para empezar a crear operaciones sobre nuestros datos como se muestra en la figura.



The screenshot shows the IBM Watson Studio interface with the 'Data Refinery' tab selected. The main area displays a dataset from 'customer_churn.csv' with columns: customerID, gender, SeniorCitizen, and Partner. Below the table, it says 'SOURCE FILE: customer_churn.csv' and 'SAMPLE SIZE: First 1000 rows'. To the left, a sidebar titled 'FREQUENTLY USED' lists various operations. A red box highlights the sidebar and the '+ Operation' button at the top left.

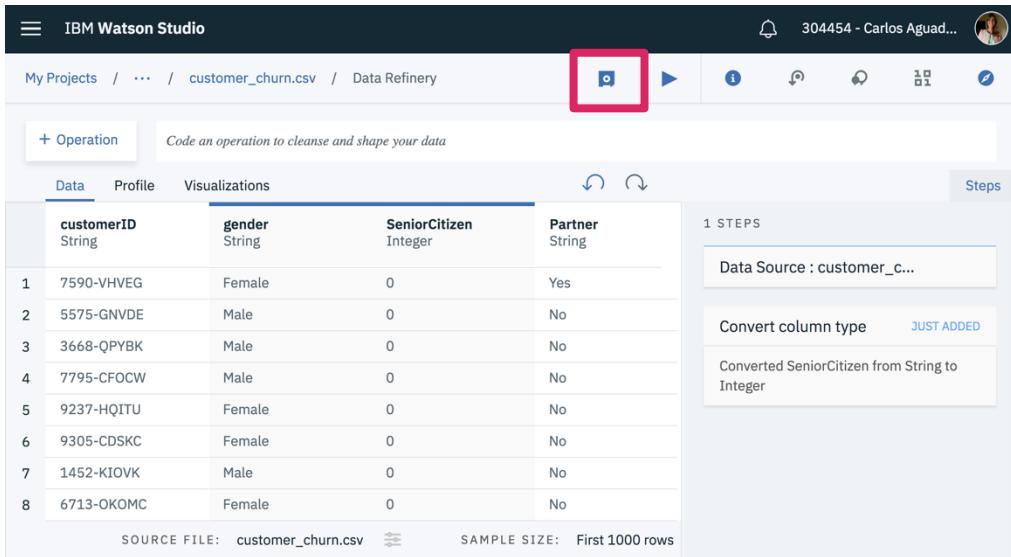
Por ejemplo, elegimos la operación **Convert Column Type** y seleccionamos la columna **LocalBillType** para convertirla de **String** a **Integer**.

Como resultado, cada una de las operaciones aplicadas sobre el dataset se guardan como pasos. Para más tarde poderlos ejecutarlos o incluso planificar la ejecución de un conjunto de pasos o data flow.



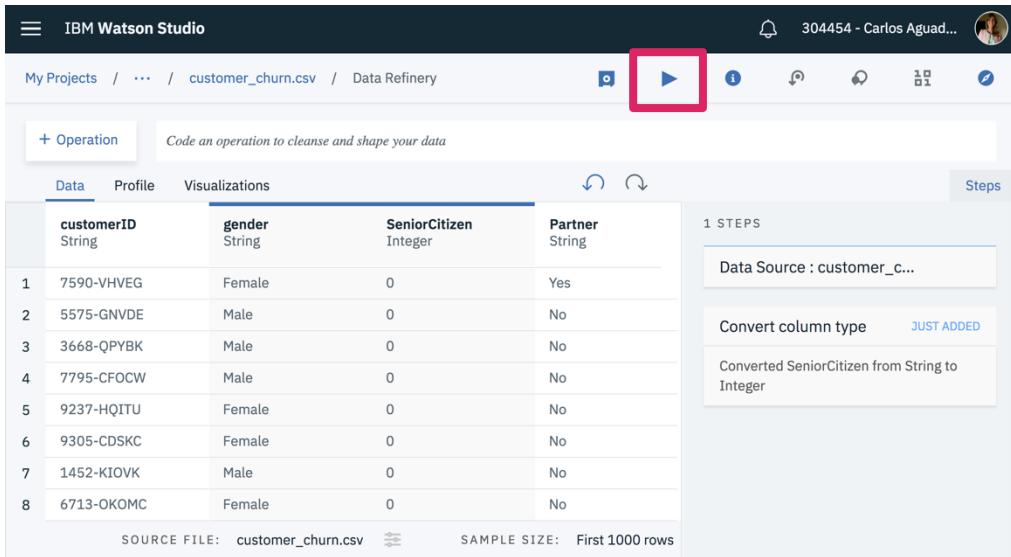
The screenshot shows the 'Data' tab selected in the IBM Watson Studio interface. The main area displays the same dataset with the 'Steps' panel on the right showing a single step: 'Convert column type' (JUST ADDED). The note indicates that 'Converted SeniorCitizen from String to Integer'. The bottom right corner of the screen contains the text 'All rights reserved.' and the number '17'.

Revisa todas las columnas, establece los tipos correctos de datos, todos ellos se guardarán como pasos y a continuación, guardamos nuestro flujo. Para ello hacemos click en el botón guardar como



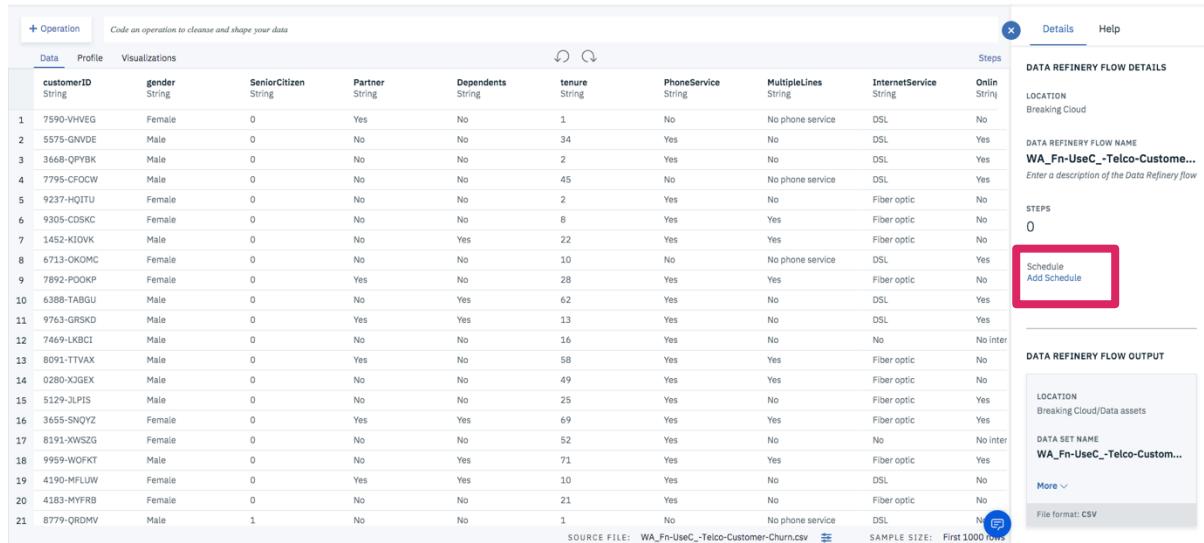
The screenshot shows the IBM Watson Studio Data Refinery interface. At the top, there's a navigation bar with 'My Projects / customer_churn.csv / Data Refinery'. Below it is a toolbar with several icons, one of which is highlighted with a red box. The main area contains a table with columns: customerID, gender, SeniorCitizen, and Partner. The 'Data' tab is selected. To the right, a sidebar titled 'Steps' shows a single step: 'Data Source : customer_c...'. Below the table, it says 'Convert column type JUST ADDED' and 'Converted SeniorCitizen from String to Integer'. At the bottom, it shows 'SOURCE FILE: customer_churn.csv' and 'SAMPLE SIZE: First 1000 rows'.

Tenemos un botón de ejecutar, que nos permite lanzar nuestro flujo o conjunto de pasos en cualquier momento.



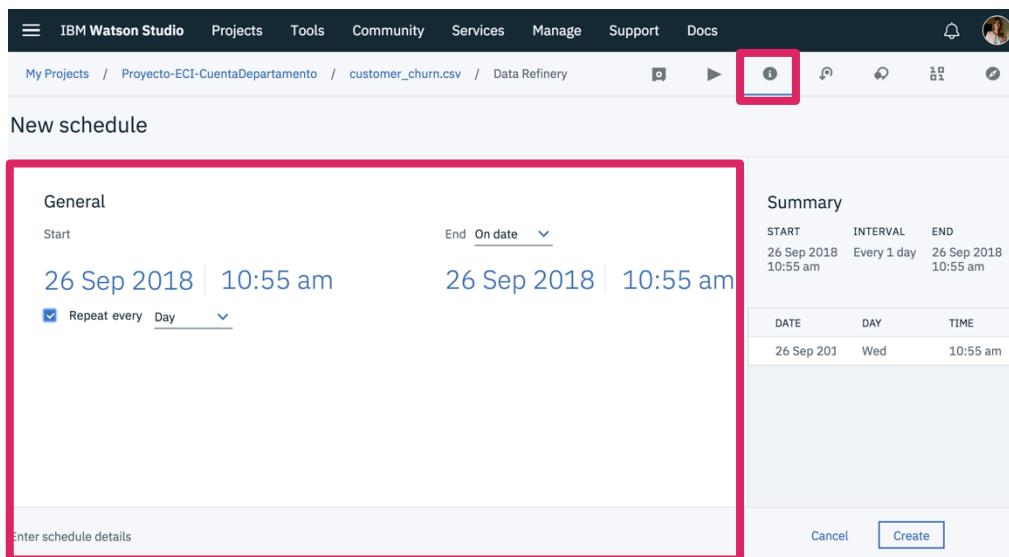
This screenshot is identical to the previous one, showing the IBM Watson Studio Data Refinery interface. The 'Run' button in the top toolbar is now highlighted with a red box. The rest of the interface, including the table data, steps sidebar, and footer information, remains the same.

Y también podemos planificarlo. Para ver las opciones, hacemos click en **Add Schedule**



The screenshot shows the IBM Watson Studio Data Refinery interface. On the left, there's a data preview table with columns: customerID, gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, MultipleLines, InternetService, and Online. The table contains 21 rows of customer data. Above the table, tabs for '+ Operation', 'Data', 'Profile', and 'Visualizations' are visible, with 'Data' being the active tab. To the right of the table, there's a 'Steps' section and a 'Details' tab. The 'DATA REFINERY FLOW DETAILS' panel shows the flow name 'WA_Fn-UseC_-Telco-Customer-Churn...' and a 'Schedule' button highlighted with a red box. Below it, the 'DATA REFINERY FLOW OUTPUT' panel shows the output location 'Breaking Cloud/Data assets' and the data set name 'WA_Fn-UseC_-Telco-Customer-Churn...'. A 'SAMPLE SIZE: First 1000 rows' link is also present.

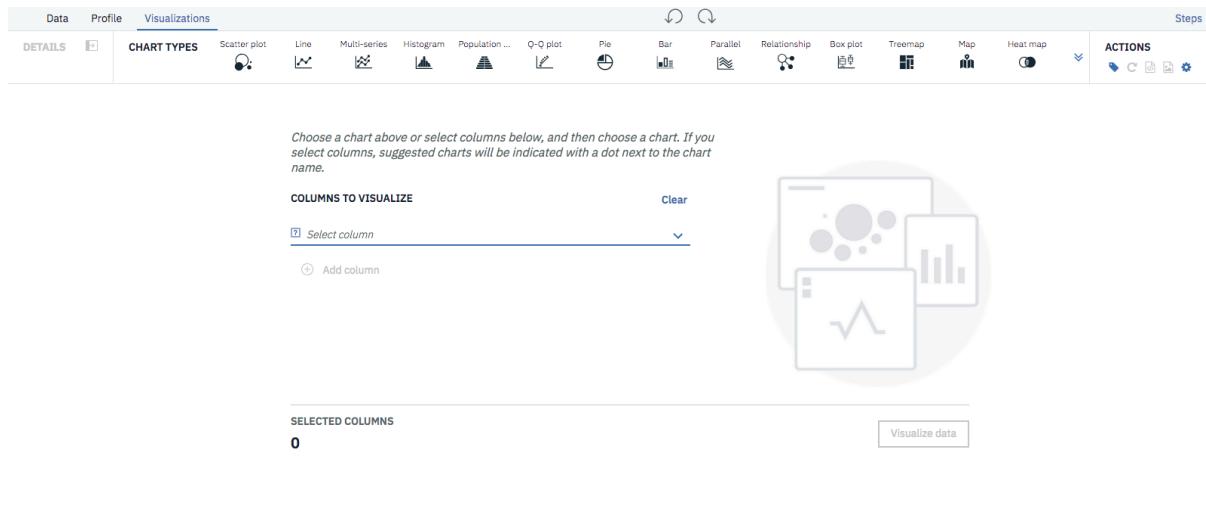
Y seleccionamos para cuando queremos planificar la ejecución de nuestro flujo



The screenshot shows the 'New schedule' dialog in IBM Watson Studio. It has two main sections: 'General' and 'Summary'. In the 'General' section, the start date is set to '26 Sep 2018 | 10:55 am' and the end date is 'On date'. A checkbox 'Repeat every Day' is checked. In the 'Summary' section, the start date is '26 Sep 2018 10:55 am', the interval is 'Every 1 day', and the end date is '26 Sep 2018 10:55 am'. A table below shows the specific date, day, and time for the scheduled run. Buttons for 'Cancel' and 'Create' are at the bottom. A red box highlights the 'i' icon in the top navigation bar of the dialog.

Adicionalmente, **Data Refinery** permite crear visualizaciones rápidas de los datos para entenderlos y obtener más conocimiento sobre ellos.

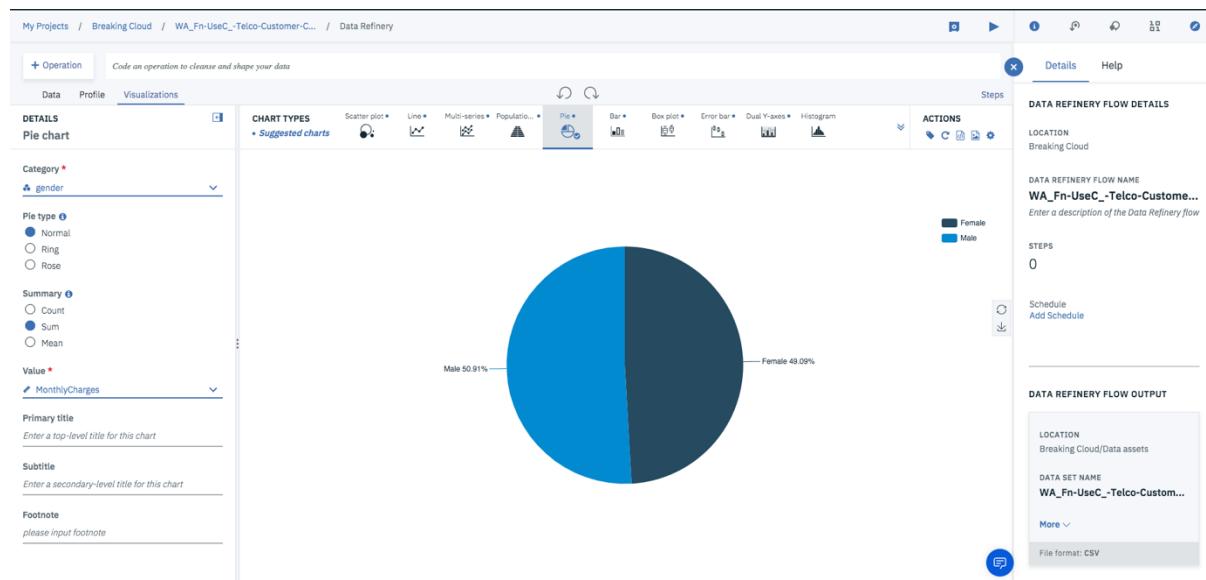
Hacemos click en **Visualizations** para crear una primera visualización.



The screenshot shows the Watson Studio interface with the 'Visualizations' tab selected. At the top, there are tabs for 'Data', 'Profile', and 'Visualizations'. Below the tabs is a toolbar with various chart types: Scatter plot, Line, Multi-series, Histogram, Population ..., Q-Q plot, Pie, Bar, Parallel, Relationship, Box plot, Treemap, Map, Heat map, and Actions. A large central area displays a placeholder message: 'Choose a chart above or select columns below, and then choose a chart. If you select columns, suggested charts will be indicated with a dot next to the chart name.' To the right of this message is a decorative graphic of overlapping charts. Below the message, there's a section titled 'COLUMNS TO VISUALIZE' with a dropdown menu labeled 'Select column' and a button to 'Add column'. Further down is a section titled 'SELECTED COLUMNS' with a count of '0' and a 'Visualize data' button.

Seleccionamos aquellas columnas que queramos visualizar. Por ejemplo: Usage y Gender

Y elegimos como tipo de gráfico **Pie**. De esta forma podemos tener de una forma rápida, una visualización donde se muestra el gasto mensual es muy equitativo entre hombres y mujeres.



Prueba a cambiar las visualizaciones, los datos... para probar mejor la experiencia de las visualizaciones.

Tened en cuenta que para visualizaciones más complejas disponemos de el componente **Dashboards** en Watson Studio, que se basa en Cognos Analytics.

Otra alternativa para crear subsets de datos es **IBM SQL Query**. Es un servicio gestionado de IBM Cloud que permite ejecutar consultas SQL (SELECT) para analizar, transformar o limpiar datos almacenados en COS.

Las fuentes de entradas pueden ser CSV, JSON, ORC, o ficheros localizados en IBM Cloud Object Storage.

Cada una de las consultas escribe como resultado un fichero CSV que se almacena en la instancia de Cloud Object Storage de tu elección.

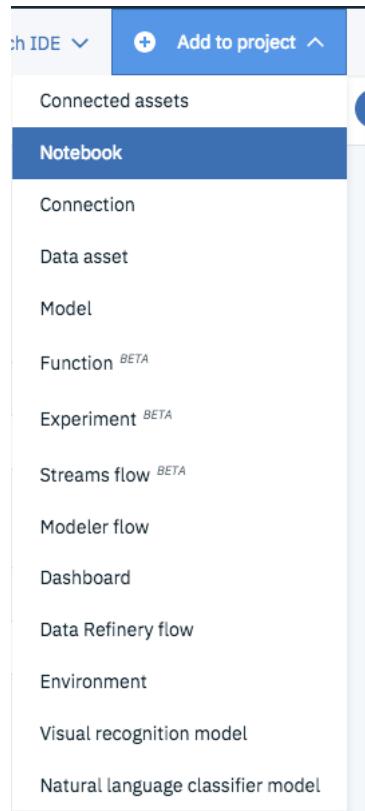
Además, incluye una interfaz web para realizar las consultas o vía API

iEnhорабуена! Ya sabes como utilizar Data Refinery para preparar los datos para tus modelos.

5- Creamos nuestro primer notebook

Vamos a ponernos manos en la masa. Para ello, vamos a volver a la pestaña de assets donde vamos a crear nuestro primer **notebook**. Los notebooks de Watson Studio son Jupyter, pero añadimos una capa adicional de colaboración, gestión de versiones e integración con fuentes de datos.

Para añadir nuestro primer notebook, hacemos click de nuevo en **+Add to project**. Elegimos **notebook**



Veréis que se pueden elegir tipos de environments. Esto quiere decir que la configuración de hardware y software para los notebooks es configurable y se puede cambiar en cualquier momento, lo que nos proporciona una gran agilidad en nuestros proyectos.

Los Notebooks Jupyter se ejecutan sobre Kernels Jupyter en entornos **Anaconda** o si usan Spark, se ejecutan sobre entornos **Spark**.

Con los entornos Anaconda, puedes definir la configuración de hardware y software del runtime que quieras utilizar en tus notebooks

Cada entorno tiene aprox. 2 GB de espacio para instalar paquetes conda o pip y ficheros temporales.

Las definiciones de 'environments' están disponibles tanto para Notebooks, como para Model Builder y Flow Editor.

Tipos de Environments disponibles:*Default Python 5.3 Free*

Gratis

Configurable la versión de R y Python*

Anaconda 5.0

1 Core / 4 GB RAM

Default Python 3.5 XS

2 Cores / 8GB RAM

Default Python 3.5 S

4 Cores / 16 GB RAM

Default R 3.4 XS

R-3.4 con r-essentials

2 Cores / 8 GB RAM

Default R 3.4 S

R-3.4 con r-essentials

4 Cores / 16 GB RAM

Si tus notebooks incluyen APIs de Spark o quieres crear modelos de machine learning o flujos con runtimes de Spark, puedes configurar el tamaño del driver de Spark y el número de ejecutores.

Se puede utilizar los entornos Spark de Watson Studio, que son Spark kernels como servicio (SparkR, PySpark y Scala). Cada kernel tiene un cluster y ejecutores de Spark dedicados.

También se pueden utilizar a través de IBM Cloud con IBM Analytics Engine. Proporciona una VM por nodo de cluster y tu propio HDFS. Spark + Hadoop.

Tipos de Environments*Default Spark Scala 2.11*

2 Ejecutores: 1 vCPU y 4 GB RAM

Driver: 1 vCPU and 4 GB RAM

Default Spark Python 3.5 XS

2 Ejecutores: 1 vCPU and 4 GB RAM

Driver: 1 vCPU and 4 GB RAM

Opciones de Hardware:

1 vCPU y 4 GB RAM

2 vCPU y 8 GB RAM

Ejecutores:

1 vCPU y 4 GB RAM

2 vCPU y 8 GB RAM

Número de ejecutores:

De 1 a 10

Versión de Spark

Spark 2.3

Versiones de Software

Scala 2.11

Python 2.7

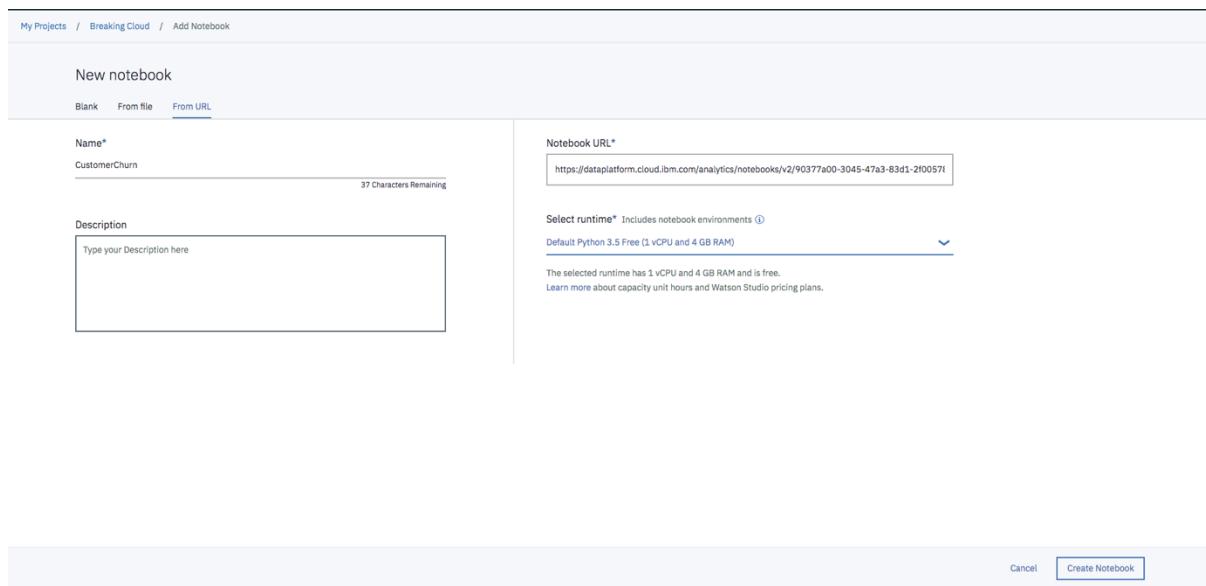
Python 3.5

R 3.4

* Cada notebook tiene su propio cluster de Spark dedicado, esto quiere decir que aunque dos notebooks comparten la misma definición de 'environment', se inician dos clusters con su propio driver de Spark y ejecutores

Podemos elegir empezar desde cero, subir un fichero que ya tenemos o desde una URL. Vamos a empezar con un ejemplo, así que elegimos la pestaña **From File** y subimos el fichero que podéis encontrar en la carpeta **Notebooks** que se llama Customer_churn_prediction

Elegimos el environment default Python 3.6 XS y hacemos click en **Create Notebook**



Y automáticamente se crea nuestro notebook que precide el customer churn.

This screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with links for IBM Watson Studio, Projects, Tools, Community, Services, Manage, Support, and Docs. Below the navigation bar is a toolbar with File, Edit, View, Insert, Cell, Kernel, and Help options. The main area displays a notebook titled "Predicting customer churn with IBM Watson Machine Learning". The notebook content includes a hexagonal grid graphic and some explanatory text. On the right side, there's a sidebar titled "Environment" which is currently active. It shows the following configuration details:

- LANGUAGE:** Python 3.5
- SPARK VERSION:** 2.3
- HARDWARE CONFIGURATION:**
 - 1 Driver: 1 vCPU and 4 GB RAM
 - 2 Executors: 1 vCPU and 4 GB RAM each
- SOFTWARE CONFIGURATION:** A "VIEW DETAILS" link.
- RUNTIME STATUS:** Running

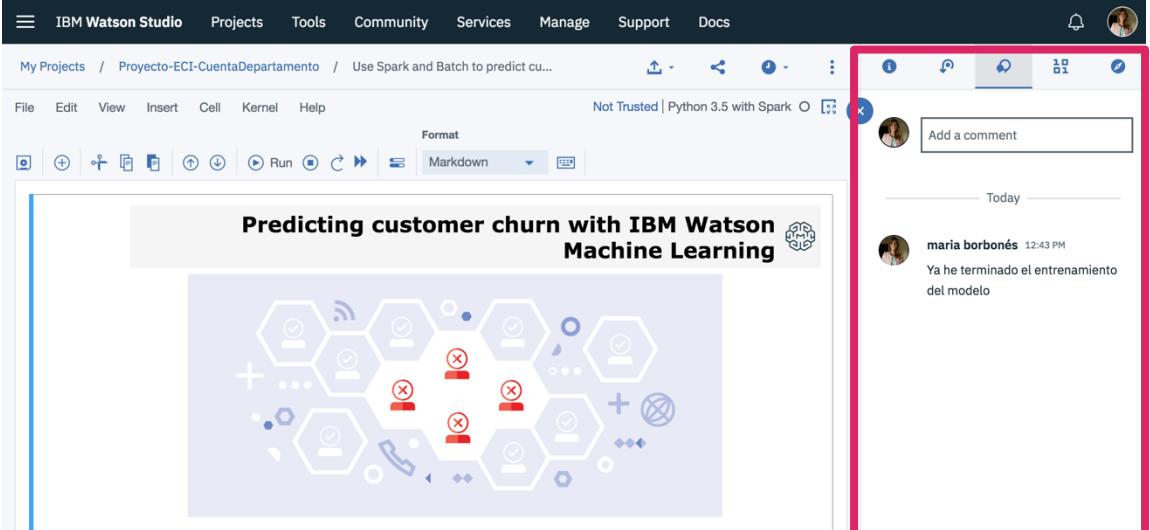
Si hacemos click en el ícono de información como se muestra en la figura anterior, vemos que se muestra información sobre el entorno elegido para ejecutar el notebook.

Si hacemos click en el siguiente ícono, vemos cómo podemos guardar diferentes versiones desde los notebooks.

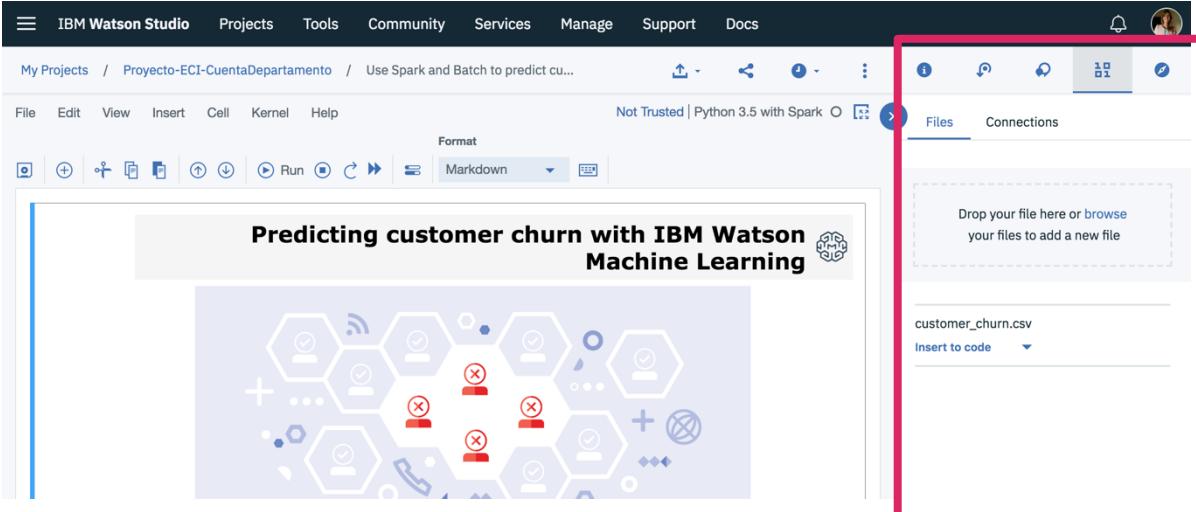
This screenshot shows the same IBM Watson Studio interface as the previous one, but with a red box highlighting the "Save Version" button in the top right corner of the sidebar. The sidebar also contains a "VERSIONS" section with the following text:

Use versions to keep copies of a notebook. You can revert the notebook to a previous version. When you share a notebook, the latest version is used.

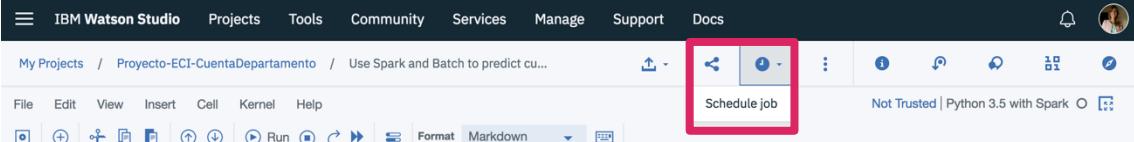
Permite dejar comentarios entre colaboradores...



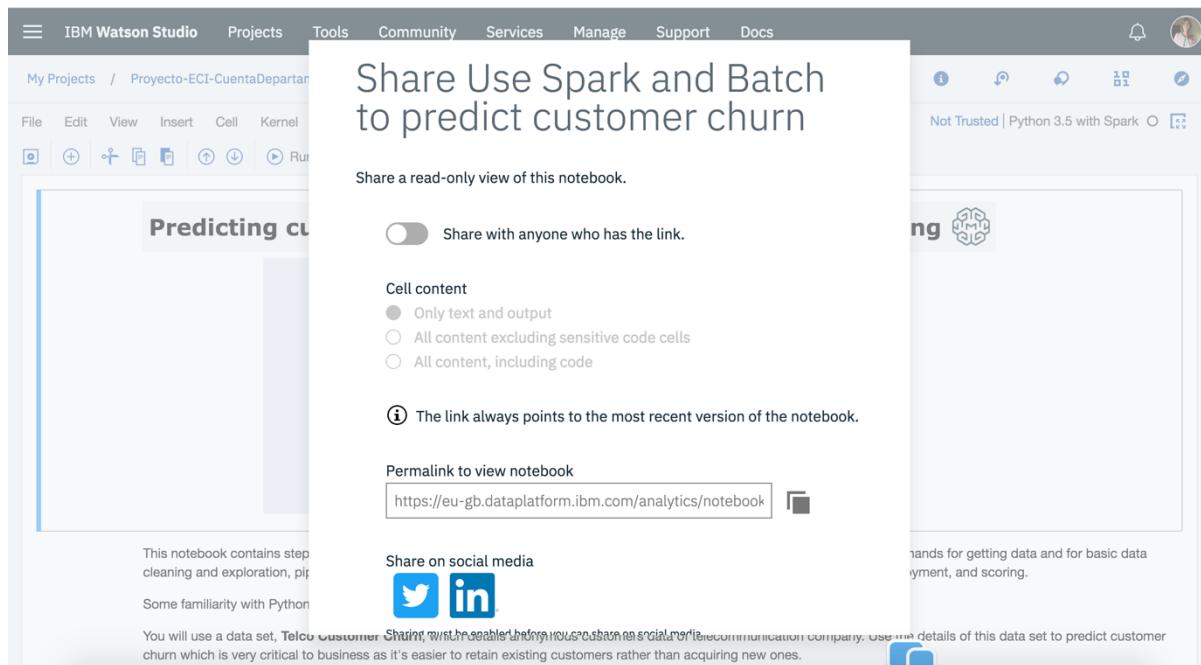
Y ver desde el notebook, los assets que tenemos en el proyecto disponibles para trabajar con ellos:



Los notebooks pueden ser planificados para ejecutarse en un momento determinado del tiempo



Pueden compartirse en redes sociales o a través de un link



Share Use Spark and Batch to predict customer churn

Share a read-only view of this notebook.

Share with anyone who has the link.

Cell content

- Only text and output
- All content excluding sensitive code cells
- All content, including code

(i) The link always points to the most recent version of the notebook.

Permalink to view notebook

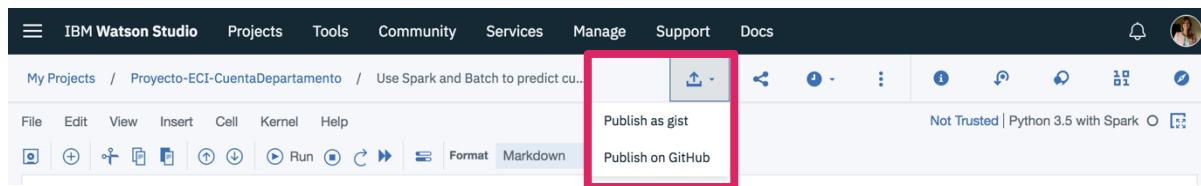
<https://eu-gb.dataplatform.ibm.com/analytics/notebook>

Share on social media

This notebook contains step cleaning and exploration, pip Some familiarity with Python You will use a data set, Telco Customer Churn, which is a data set from a telecommunications company. Use the details of this data set to predict customer churn which is very critical to business as it's easier to retain existing customers rather than acquiring new ones.

Y pueden publicarse como gist o en github directamente



IBM Watson Studio Projects Tools Community Services Manage Support Docs

My Projects / Proyecto-ECI-CuentaDepartamento / Use Spark and Batch to predict cu..

File Edit View Insert Cell Kernel Help

 Publish as gist

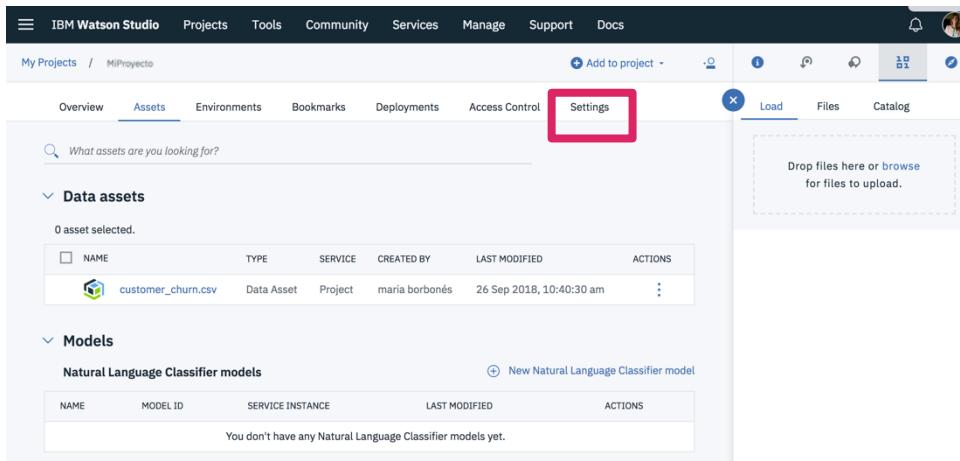
Publish on GitHub

Not Trusted | Python 3.5 with Spark O

El notebook que hemos importado contiene los pasos y el código para desarrollar un modelo predictivo, y empezar a evaluar datos. Este notebook incluye comandos para obtener datos y algunas exploraciones básicas, creaciones de pipelines, entrenamiento de un modelo, persistencia del modelo en Watson Machine Learning as a Service, despliegue del modelo y scoring.

Los datos que se van a utilizar son los que ya hemos importado en nuestro proyecto.

Lo primero que necesitamos para poder ejecutar nuestro notebook es crear el servicio de **Machine Learning as a Service** y asociarlo a nuestro proyecto. Hacemos click en la pestaña **Settings** de nuestro proyecto.



IBM Watson Studio Projects Tools Community Services Manage Support Docs

My Projects / MiProyecto

Overview Assets Environments Bookmarks Deployments Access Control **Settings**

What assets are you looking for?

Data assets

0 asset selected.

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
 customer_churn.csv	Data Asset	Project	maria borbonés	26 Sep 2018, 10:40:30 am	

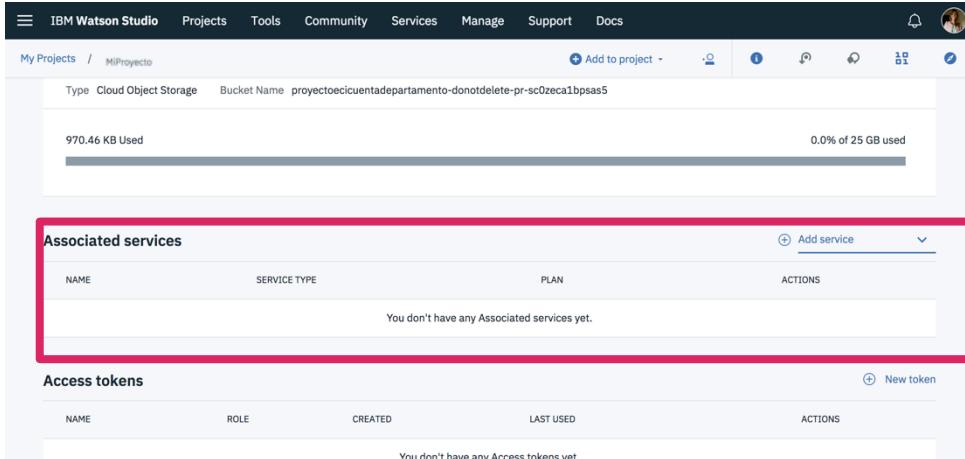
Models

Natural Language Classifier models [+ New Natural Language Classifier model](#)

NAME	MODEL ID	SERVICE INSTANCE	LAST MODIFIED	ACTIONS
You don't have any Natural Language Classifier models yet.				

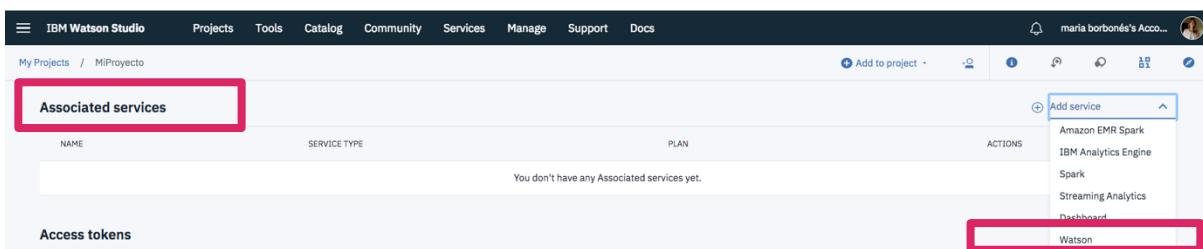
118. All rights reserved.

Buscamos la categoría **Associated services**



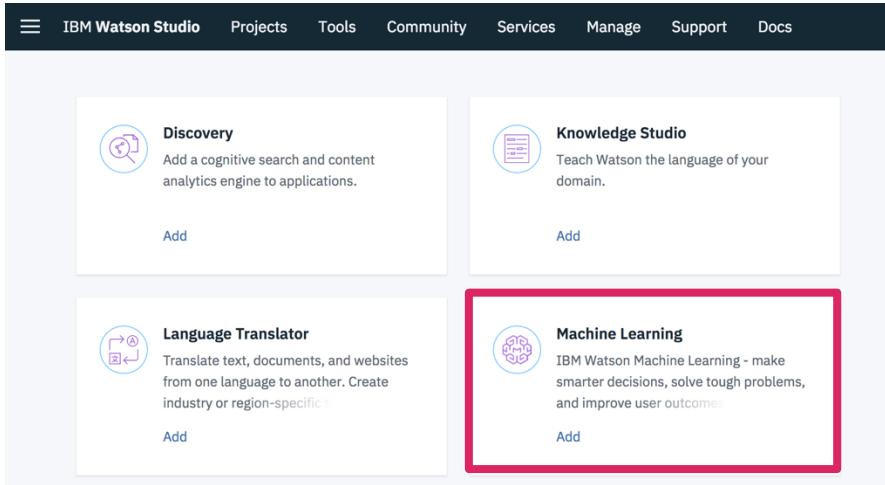
The screenshot shows the IBM Watson Studio interface with the 'Associated services' section highlighted by a red box. The section header is 'Associated services'. It includes columns for NAME, SERVICE TYPE, PLAN, and ACTIONS. A message at the bottom states: 'You don't have any Associated services yet.' Below this section is another one titled 'Access tokens'.

Elegimos la opción **Watson**



The screenshot shows the 'Associated services' section with the 'Add service' button highlighted by a red box. A dropdown menu is open, listing various service options: Amazon EMR Spark, IBM Analytics Engine, Spark, Streaming Analytics, Dashboard, and Watson. The 'Watson' option is also highlighted by a red box.

Elegimos el servicio de **Machine Learning as a service**



The screenshot shows the 'Machine Learning as a service' section with four options: Discovery, Knowledge Studio, Language Translator, and Machine Learning. The 'Machine Learning' option is highlighted by a red box. Each option has a description and an 'Add' button.

Seleccionamos crear nuevo

IBM Watson Studio Projects Tools Community Services Manage Support Docs

Machine Learning

Existing New

Machine Learning

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

Features

Machine Learning features	Wide choice of interfaces
Take advantage of machine learning models management (continuous learning system) and deployment (online, batch, streaming). Select any of widely supported machine learning frameworks: Tensorflow, Keras, Caffe, Pytorch, Spark MLlib, scikit learn, xgboost and SPSS.	Use the command line interface and Python client to manage your artifacts. Extend your application with artificial intelligence through the Watson Machine Learning REST API.

Integration with Watson Studio

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

Elegimos un plan de precios y hacemos click en **Create**

IBM Watson Studio Projects Tools Community Services Manage Support Docs

PLAN	FEATURES	PRICING
<input type="radio"/> Enterprise (beta)	Service instance Deploy models and functions in production Trace transactions, model and data lineage	Free
<input checked="" type="radio"/> Lite	Service instance (5 models per instance) 5,000 predictions 50 capacity unit-hours: Compute Tier: k80 = 2 capacity units for 1 training hour Compute Tier: k80x2 = 4 capacity units for 1 training hour Compute Tier: k80x4 = 8 capacity units for 1 training hour Otherwise 1 capacity unit for 1 computation hour Max 8 k80 GPUs (Deep Learning Training)	Free

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 50 capacity unit-hours per month during which model can be trained, evaluated, and deployed to be available to accept prediction events, with a minimum of 1 minute per training job.

Predictions
Capacity unit-hours:
Compute Tier: k80 = 2 capacity units for 1

Nos aparecerá nuestro servicio de MLaaS asociado a nuestro proyecto en la pestaña **Settings**

Associated services

Add service

NAME	SERVICE TYPE	PLAN	ACTIONS
pm-20-uq	Watson - Machine Learning	IBM Watson Studio	⋮

Ahora necesitamos obtener las credenciales del servicio, para ello, debemos ir a IBM Cloud en <https://console.ng.bluemix.net/> y desde la consola accedemos a nuestro servicio de **Machine Learning as a Service**

IBM Cloud Catalog Docs Support Manage Search for resource... maria.borbones's Account

Dashboard

Cloud Foundry Services

Name	Region	CF Org	CF Space	Plan	Service Offering
Machine Learning-TM	Dallas	maria.borbones@es.ibm.com	Cognitive	lite	pm-20
Machine Learning-deeplearning	Dallas	maria.borbones@es.ibm.com	Cognitive	lite	pm-20

Services

Name	Location	Resource Group	Plan	Details	Service Offering
Machine Learning-TM	Dallas	MariaResources	Lite	Provisioned	Machine Learning

Create resource

Y hacemos click sobre el servicio que acabamos de crear. Seleccionamos en el menú lateral **Service Credentials** y copiamos en algún editor de texto las credenciales para usarlas en nuestro notebook.

Manage

Service credentials

Plan

Connections

Dashboard / Machine Learning-TM

Resource Group: MariaResources Location: Dallas

Service credentials

Credentials are provided in JSON format. The JSON snippet lists credentials, such as the API key and secret, as well as

Service credentials

Items per page 10 | 1-1 of 1 items

KEY NAME	DATE CREATED
wdp-writer	NOV 16, 2018 - 05:20:38 PM

```
{
  "apikey": "sF5m_z07gmtqoeFsBu2IFSnNjyqF1-sK8rxN7UQps0VQ",
  "iam_apikey_description": "Auto generated apikey during resource-key operation for Instance - ci",
  "iam_apikey_name": "auto-generated-apikey-9c616ce2-6585-4fb0-887a-a6fc50e95ee",
  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",
  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a:290f0f438c15a26b2b129419c1c1b952::se",
  "instance_id": "74d2975c-382d-436c-877a-a640faca04d",
  "password": "b491c1f7-708f-4614-9f58-8e081e3dfeef",
  "url": "https://us-south.ml.cloud.ibm.com",
  "username": "9c616ce2-6585-4fb0-887a-a6fc50e95ee"
}
```

Volvemos a Watson Studio y a nuestro notebook desde la pestaña assets. Si has cerrado la pestaña de Watson Studio sin querer, aquí te dejo un recordatorio de la URL:
<https://dataplatform.cloud.ibm.com>

En la tercera celda de código, eliminamos el contenido, dejamos la celda seleccionada y desde el menú lateral, pinchamos en el botón de unos y ceros. Seleccionamos el fichero de customer churn y hacemos click en **Insert to code > Insert as DataFrame**

```
In [8]:
```

```

import types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# #hidden cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
client_f38eecd070d457997e73598bcfa557 = ibm_boto3.client(service_name='s3',
    ibm_api_endpoint='https://s3.us-gov.cloud-object-storage.appdomain.cloud',
    ibm_auth_endpoint='https://iam.ng.bluemix.net/oidc/token',
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3-api.us-geo.objectstorage.service.networklayer.com')
body = client_f38eecd070d457997e73598bcfa557.get_object(Bucket='uematoshesainpractica-donotdelete-pr-os2ckmhsvlligp',Key='customer_churn.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, '__iter__'): body.__iter__ = types.MethodType( __iter__, body )

df_data = pd.read_csv(body)
df_data.head()

```

```
Out[8]:
```

ID	CHURN	Gender	Status	Children	Est Income	Car Owner	Age	LongDistance	International	Local	Dropped	PaymentMethod	LocalBillType	LongDistanceBillType	Usage	RatePlan	
0	1	T	F	S	1.0	38000.00	N	24.393333	23.56	0.0	206.08	0.0	CC	Budget	Intnl_discount	229.64	3.0
1	6	F	M	M	2.0	29616.00	N	49.426667	29.78	0.0	45.50	0.0	CH	FreeLocal	Standard	75.29	2.0
2	8	F	M	M	0.0	19732.80	N	50.673333	24.81	0.0	22.44	0.0	CC	FreeLocal	Standard	47.25	3.0
3	11	F	M	S	2.0	96.33	N	56.473333	26.13	0.0	32.88	1.0	CC	Budget	Standard	59.01	1.0
4	14	F	F	M	2.0	52004.80	N	25.140000	5.03	0.0	23.11	0.0	CH	Budget	Intnl_discount	28.14	1.0

Drop files here or browse for files to upload.

WA_Fn-UseC_-Telco-Customer-Churn.csv

Insert to code

customer_churn.csv

Insert to code

pandas DataFrame

Credentials

Acabáis de ver, lo fácil que es insertar un objeto de cualquiera de nuestras fuentes de datos en un notebook para empezar a trabajar con él.

Para terminar de configurar nuestro notebook, buscamos la celda que contiene las credenciales de WML.

```
wml_credentials = {  
  
    "url": "",  
  
    "access_key": "",  
  
    "username": "",  
  
    "password": "",  
  
    "instance_id": ""  
  
}
```

E introducimos las credenciales que hemos almacenado anteriormente.

Una vez hecho esto ya podemos ejecutar nuestro notebook. Para ejecutar todas las celdas de una, podéis hacer click en **Cell > Run All** o ir ejecutando una por una con las teclas shift + enter.

Analiza el código una vez ejecutado y mira que hace cada una de las celdas, porque como te adelantaba al principio no sólo entrena un modelo, sino que además lo persiste en **Watson Machine Learning as a Service**, que vamos a aprender vamos sobre él, en el siguiente apartado de este laboratorio.

¡Enhorabuena! Ya sabes como trabajar con notebooks desde Watson Studio.