

IBM

Data Science

Experience Local



Manual para Workshop
Febrero 2018

Índice

Introducción

Workshop 1. Comenzar con DSX

- 1.** 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. Analítica avanzada

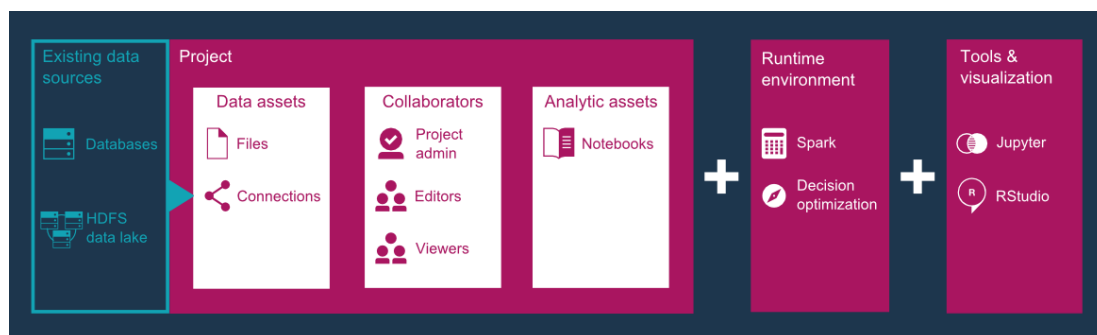
- 1.** Algoritmos de analítica predictiva de SPSS
- 2.** DSX Local Workshop
- 3.** Decision Optimization en DSX Local
- 4.** Deep Learning

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, pero incluye 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 variantes: IBM Data Science Experience en Local y en Cloud. En este workshop nos centraremos en IBM Data Science Experience en Local.

IBM Data Science Experience (DSX) Local es una solución empresarial lista para usar en tus instalaciones dirigida no solo a científicos de datos sino que también está dirigida a ingenieros de datos. Ofrece un conjunto de herramientas de data science, como los notebooks RStudio, Spark, Jupyter y Zeppelin, que están integradas con tecnologías propietarias de IBM.



La interfaz intuitiva de DSX proporciona un espacio de proyectos de colaboración para equipos e individuos para optimizar su tiempo y rendimiento. Los proyectos pueden contener notebooks, assets de datos y colaboradores.

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**.

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.

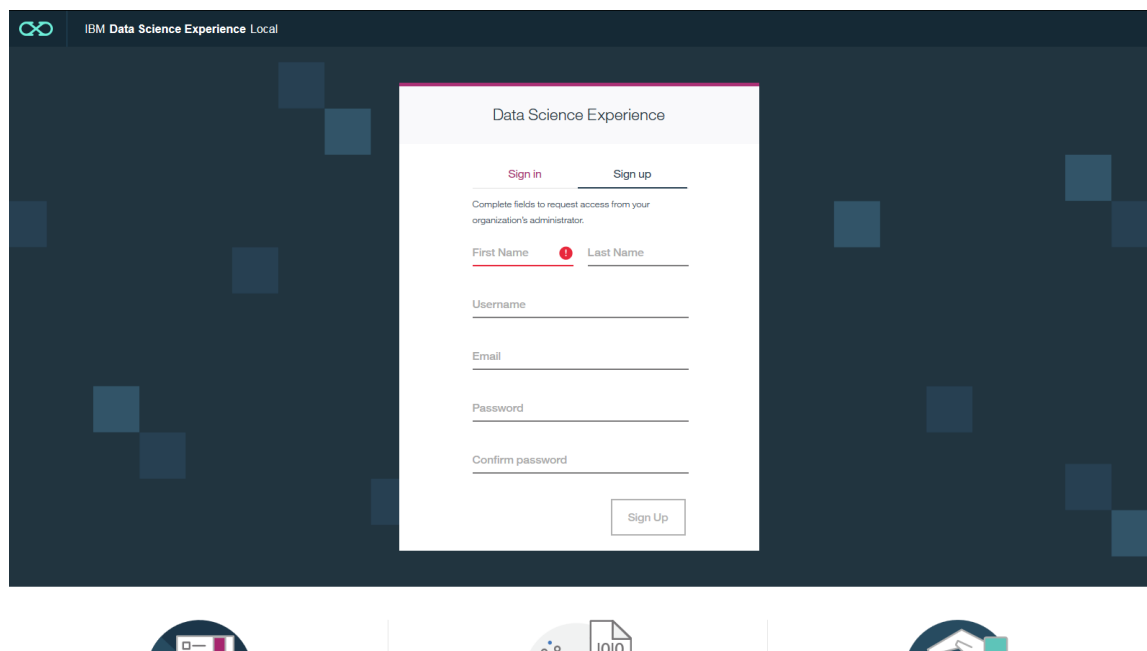
Con Data Science Experience puedes administrar los recursos del proyecto y la colaboración de los usuarios además de poder compartir, bifurcar y reutilizar assets con Github.

Workshop 1.

1. Comenzar con DSX

Entra en: <https://9.172.229.220/>

Registrate:



IBM Data Science Experience Local

Data Science Experience

[Sign in](#) [Sign up](#)

Complete fields to request access from your organization's administrator.

First Name ! Last Name

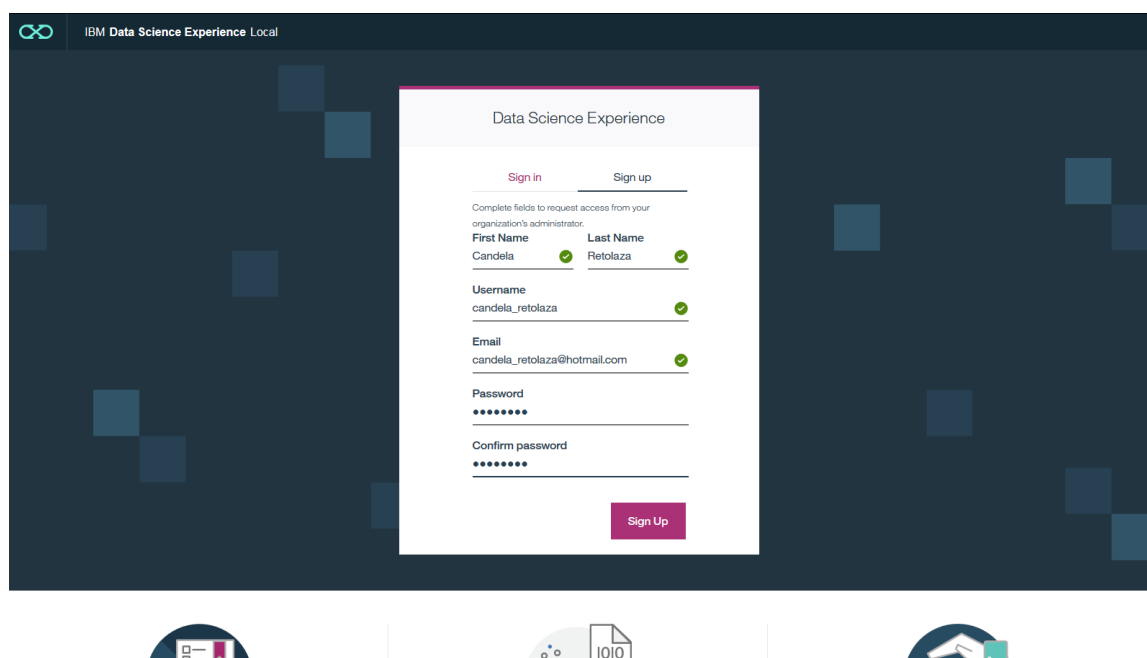
Username

Email

Password

Confirm password

Sign Up



IBM Data Science Experience Local

Data Science Experience

[Sign in](#) [Sign up](#)

Complete fields to request access from your organization's administrator.

First Name ✓ Last Name ✓

Candela Retolaza

Username ✓

candela_retolaza

Email ✓

candela_retolaza@hotmail.com

Password

Confirm password

Sign Up

El administrador/administradores tiene que aceptar la petición y podrás comenzar a utilizar DSX.

Sample notebooks

NOTEBOOK - JUPYTER	NOTEBOOK - JUPYTER	NOTEBOOK - JUPYTER	NOTEBOOK - JUPYTER
Analyze open data sets with...	Train and predict with Scala machine...	Using the Progress Callbacks with...	How to make targeted offers to customers?
AUTHOR: IBM	AUTHOR: IBM	AUTHOR: IBM	AUTHOR: IBM-DD
DATE: Aug 29, 2016	DATE: Mar 07, 2017	DATE: Mar 07, 2017	DATE: Mar 07, 2017
TOPIC: Society	TOPIC: Transportation	TOPIC: Science & Technology	TOPIC: Economy & Business
SOURCE: External	SOURCE: External	SOURCE: External	SOURCE: Self-Contained

Recently updated projects [View all](#) My Projects [New Project](#)

NAME	PROJECT TYPE	ROLE	COLLABORATORS	LAST UPDATED
DSX_Workshop	Standard	Admin		2018-02-08
FastStartTELCO	Standard	Admin		2018-02-02
DD Workshop	Standard	Admin		2018-01-31
DSX_Local_WorkshopJacques	Standard	Viewer	..	2018-01-30
DSX_Local_WorkshopCandela_Carmen	Standard	Admin		2018-01-30

Helpful links

- [Docs](#) Find the information you need
- [Discussion forum](#) Stack Overflow is a
- [Blog](#) Read and follow our blogs
- [Got ideas?](#) Have feedback on Data Science Experience?

Para empezar a utilizar Data Science Experience puedes seguir los siguientes pasos:

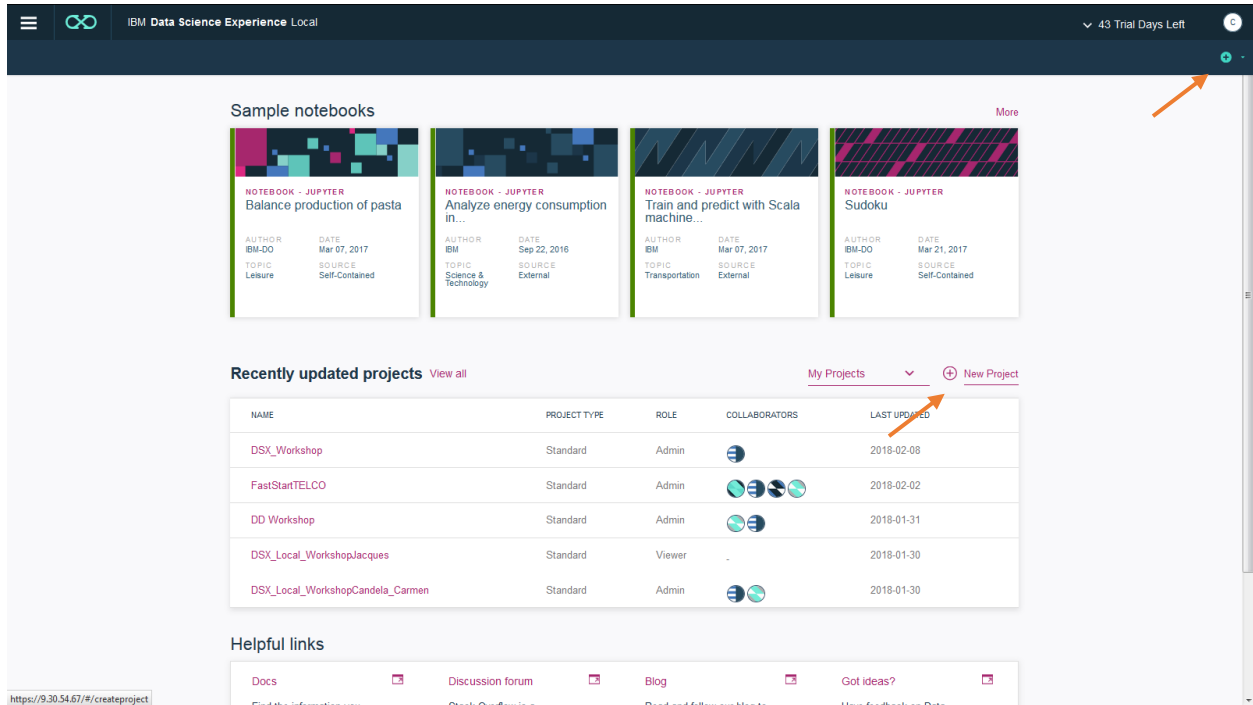
1. Configura un proyecto para organizar sus recursos.
2. Agrega colaboradores a tu proyecto.
3. Agrega datos a tu proyecto.
4. Comience a analizar datos. Por ejemplo, puede crear notebooks, y hacer un deployment para crear aplicaciones o servicios web.

¿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



Sample notebooks

NOTEBOOK - JUPYTER
Balance production of pasta
AUTHOR IBM-DO DATE Mar 07, 2017
TOPIC Leisure SOURCE Self-Contained

NOTEBOOK - JUPYTER
Analyze energy consumption in...
AUTHOR IBM DATE Sep 22, 2016
TOPIC Science & Technology SOURCE External

NOTEBOOK - JUPYTER
Train and predict with Scala machine...
AUTHOR IBM DATE Mar 07, 2017
TOPIC Transportation SOURCE External

NOTEBOOK - JUPYTER
Sudoku
AUTHOR IBM-DO DATE Mar 21, 2017
TOPIC Leisure SOURCE Self-Contained

Recently updated projects View all

My Projects New Project

NAME	PROJECT TYPE	ROLE	COLLABORATORS	LAST UPDATED
DSX_Workshop	Standard	Admin		2018-02-08
FastStartTELCO	Standard	Admin		2018-02-02
DD Workshop	Standard	Admin		2018-01-31
DSX_Local_WorkshopJacques	Standard	Viewer		2018-01-30
DSX_Local_WorkshopCandela_Carmen	Standard	Admin		2018-01-30

Helpful links

Docs Find the information you need

Discussion forum Start a discussion

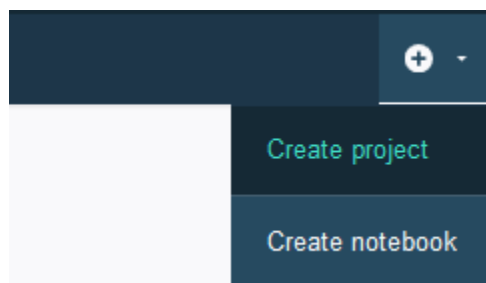
Blog Don't read, follow our blog

Got ideas? Have feedback on Data Science?

Podemos hacerlo de varias maneras, por ejemplo, arriba a la derecha, pinchando en



el icono, creamos un nuevo proyecto



Sample notebooks

NOTEBOOK - JUPYTER
Balance production of pasta

AUTHOR: IBM-DO, DATE: Mar 07, 2017, TOPIC: Leisure, SOURCE: Self-Contained

NOTEBOOK - JUPYTER
Analyze energy consumption in...

AUTHOR: IBM, DATE: Sep 22, 2016, TOPIC: Science & Technology, SOURCE: External

NOTEBOOK - JUPYTER
Train and predict with Scala machine...

AUTHOR: IBM, DATE: Mar 07, 2017, TOPIC: Transportation, SOURCE: External

NOTEBOOK - JUPYTER
Sudoku

AUTHOR: IBM-DO, DATE: Mar 21, 2017, TOPIC: Leisure, SOURCE: Self-Contained

Recently updated projects [View all](#)

NAME	PROJECT TYPE	ROLE	COLLABORATORS	LAST UPDATED
DSX_Workshop	Standard	Admin		2018-02-08
FastStartTELCO	Standard	Admin		2018-02-02
DD Workshop	Standard	Admin		2018-01-31
DSX_Local_WorkshopJacques	Standard	Viewer		2018-01-30
DSX_Local_WorkshopCandela_Carmen	Standard	Admin		2018-01-30

Helpful links

[Docs](#) [Discussion forum](#) [Blog](#) [Got ideas?](#)

<https://9.30.54.67/#/createProject>

Por ejemplo, llamamos al nuevo proyecto: Workshop_CR con tus iniciales, pues estando en el mismo clúster el nombre de los proyectos deber ser único. Añadimos una descripción opcional: por ejemplo, Workshop DSX. Vemos que existen tres maneras de crear un proyecto: En blanco, Desde un fichero o desde GitHub. Vamos a comenzar por crear uno en blanco.

Create Project

Blank From File From Github

Name*

Workshop

This name is valid

Description

Type your description here

☐ Library Project

Cancel Create

Gestión de proyectos:

Los proyectos son la base de operaciones para la colaboración. Puede usar proyectos para recopilar activos, como cuadernos, modelos, fuentes de datos y conjuntos de datos remotos y locales, en un solo lugar. Los proyectos son una excelente manera de trabajar con un equipo: compartir un conjunto de activos y luego construir nuevos modelos y análisis juntos. Puede colaborar en DSX Local o en un repositorio externo de GitHub.

DSX Local administra sus recursos y proporciona opciones automáticas de escalado y recuperación. La consola de administración (IBM Data Platform Manager) lo ayuda a administrar y supervisar su hardware, usuarios y servicios. La seguridad está integrada en DSX desde el cifrado de datos en reposo y en movimiento hasta la administración de usuarios.

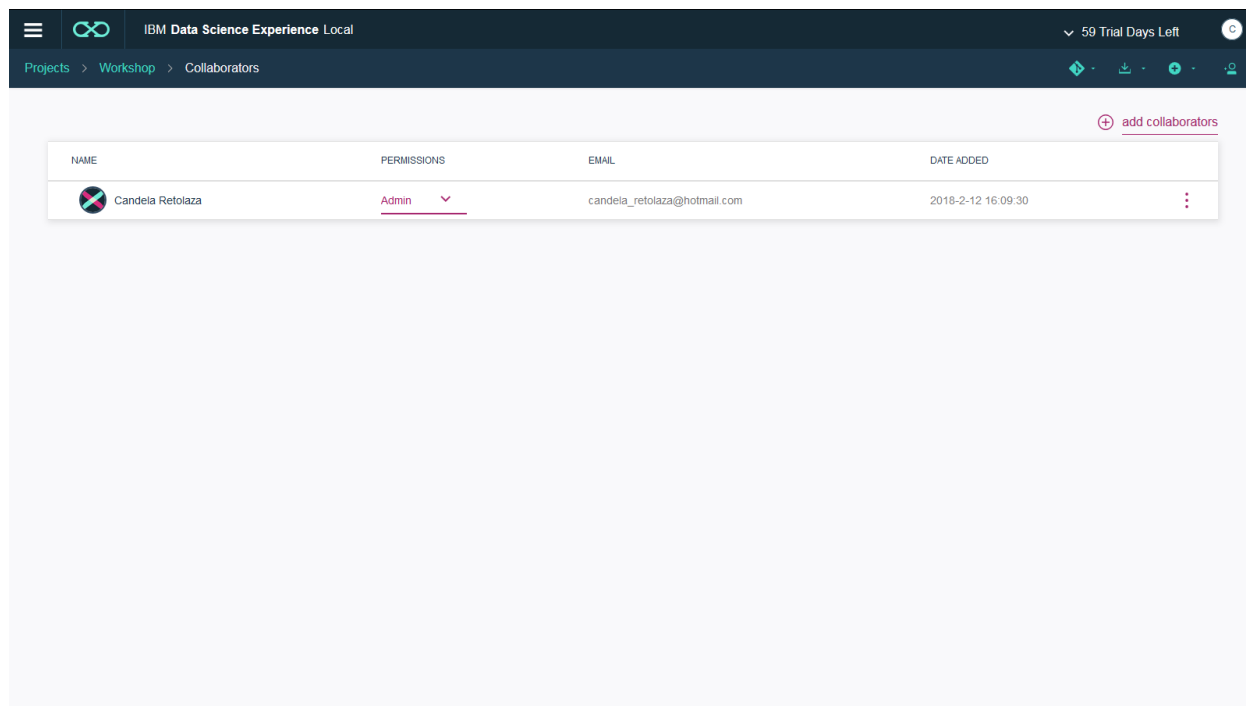
Si tienes permisos de administrador en un proyecto, tienes control total sobre él. Si tiene permisos de Editor, puedes agregar activos y colaboradores a un proyecto.

The screenshot shows the IBM Data Science Experience Local interface. At the top, there's a header with the IBM logo and 'IBM Data Science Experience Local'. Below the header, there's a navigation bar with 'Projects' and 'Worshop'. The main content area displays the 'Worshop' project details. It includes a description 'Description not available.', the date created 'Fri Feb 09 2018', and a summary of resources: 0 Assets, 5 Environments, 0 Jobs, 0 Data Sources, and 1 Collaborator. Below this, there's a section for 'Collaborators' showing 'Candela Retolaza admin' with a 'View all (1)' link. To the right, there's a 'Recent Assets' section with a table header: NAME, ASSET TYPE, and LAST MODIFIED. The table is currently empty, with a message 'you have no recently modified assets'.

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.

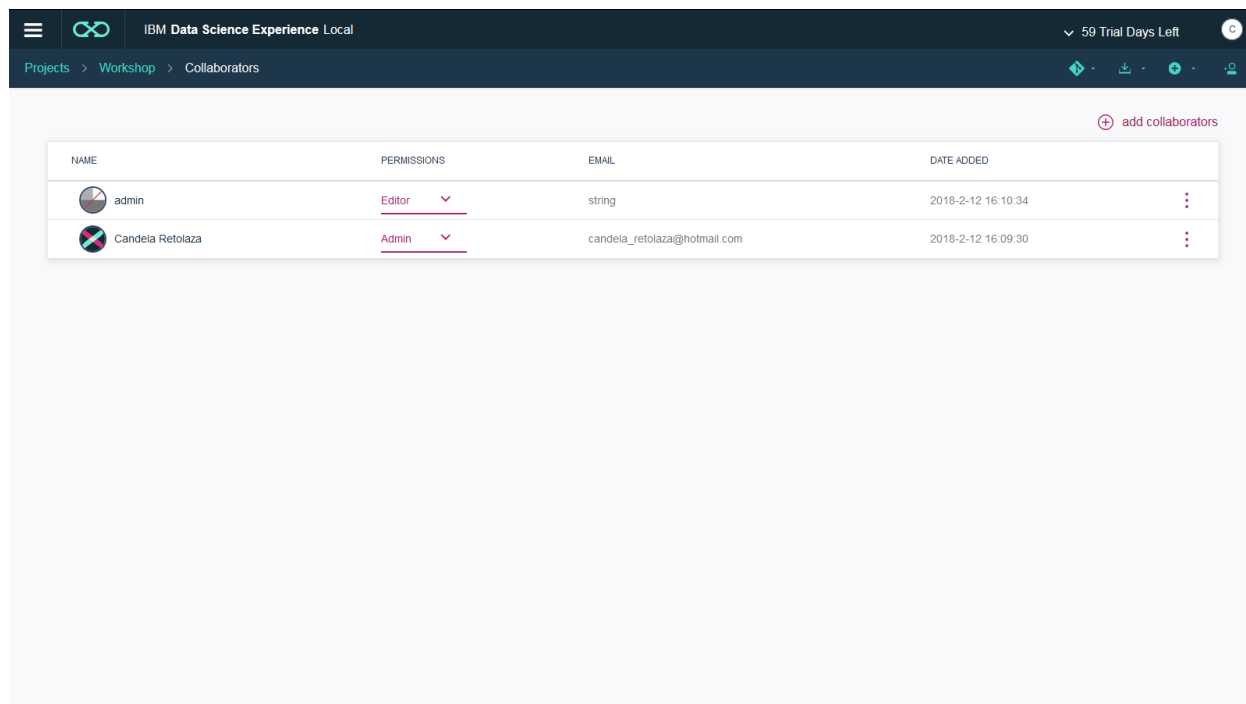
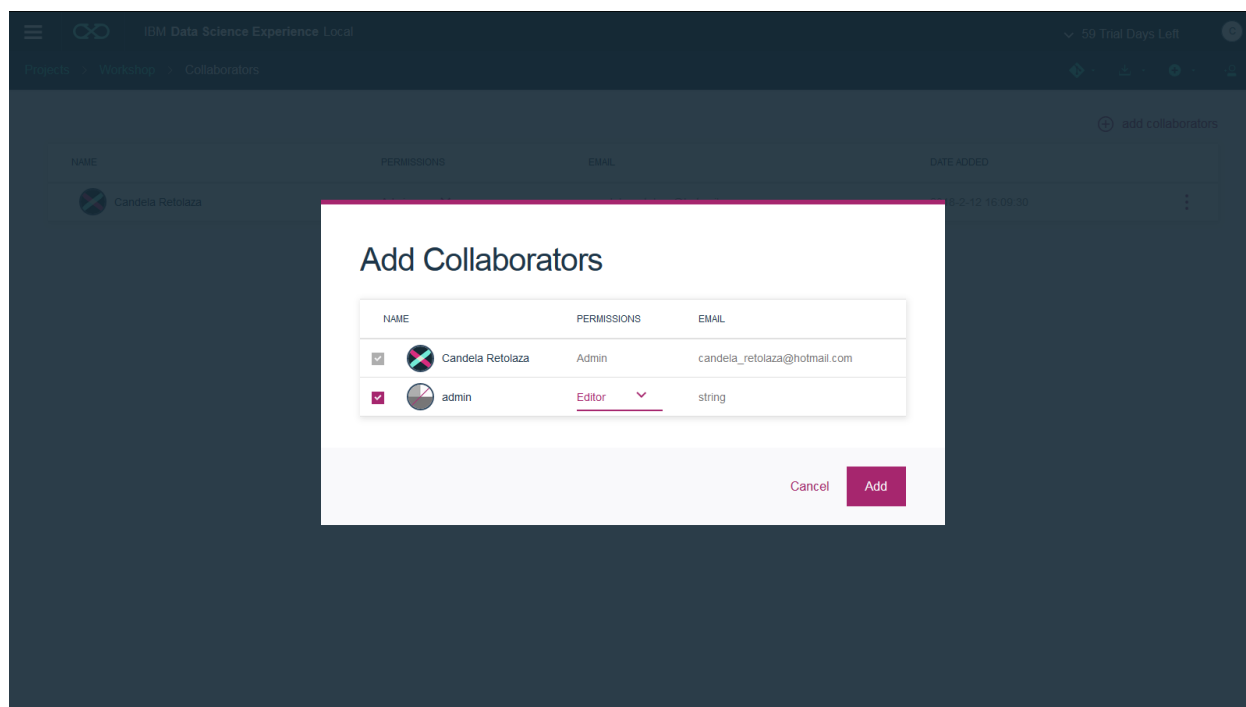
2. Añade colaboradores

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.

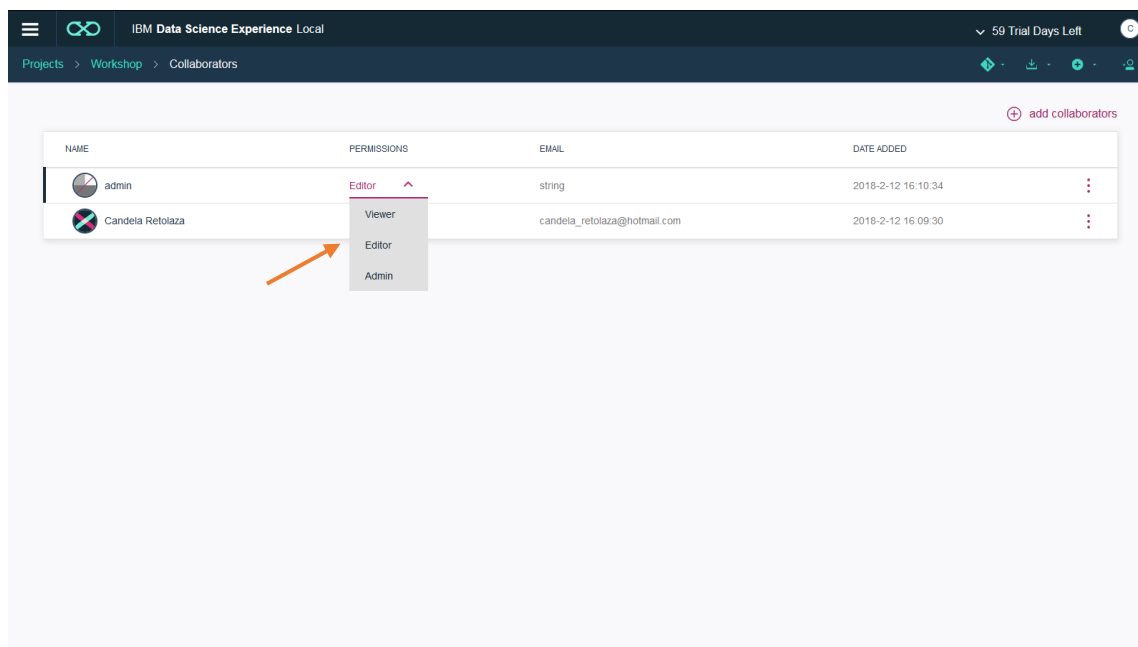


Para poder hacer esto, solo necesito que la persona a la que quiero añadir tenga acceso al mismo cluster.

Elegimos a quienes creemos que estén en nuestro proyecto, seleccionamos el tipo de acceso que queremos dar a esa persona y lo añadimos. Y ya estará en nuestro proyecto.



Una vez añadido, 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, puede agregar recursos como conexiones, datos de fichero plano, datos de bases de datos, Notebooks, flujos de aprendizaje automático, modelos.

Puede crear un cuaderno Jupyter en R, Python o Scala, o un cuaderno Zeppelin con una combinación de R, Python o Scala. También puede crear modelos automáticos o semiautomáticos y flows de SPSS Modeler. Puede comenzar desde cero, importar un bloc de notas existente o usar uno de los ejemplos de la comunidad.

Cuando abra un notebook nuevo o existente, la barra de acción está disponible. Desde la barra de acciones, puede ver y usar fuentes de datos conectadas, ver los comentarios y versiones, y compartir su proyecto o notebook con otras personas.

Si tiene permisos de administrador en un proyecto, puede eliminar assets. Para eliminar un asset, elija Eliminar en el menú ACCIONES al lado del nombre del asset.

Workshop 3.

Analítica avanzada

1. 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. Podemos entrar directamente desde la pantalla principal, o desde el proyecto, pinchando en crear un nuevo Flow de SPSS Modeler.

IBM Data Science Experience Local

Community

Published Assets

Projects

Model Management

Tools

Jupyter

RStudio

Zeppelin

H2O Flow

SPSS Modeler ^{RC}

Environments

Sample notebooks

More

NOTEBOOK - JUPYTER
Model a Golomb ruler using DO
AUTHOR: IBM-DO, DATE: Dec 12, 2016, TOPIC: Science & Technology, SOURCE: Self-Contained

NOTEBOOK - JUPYTER
Using the Progress Callbacks with...
AUTHOR: IBM, DATE: Mar 07, 2017, TOPIC: Science & Technology, SOURCE: External

NOTEBOOK - JUPYTER
Work with a remote Spark Sample One
AUTHOR: IBM, DATE: Jul 07, 2017, TOPIC: Science & Technology, SOURCE: Self-Contained

NOTEBOOK - JUPYTER
Work with a remote Spark Sample...
AUTHOR: IBM, DATE: Mar 07, 2017, TOPIC: Science & Technology, SOURCE: External

Recently updated projects

View all

My Projects New Project

NAME	PROJECT TYPE	ROLE	COLLABORATORS	LAST UPDATED
Intro_DeepLearning_Candela	Standard	Admin		2018-02-13
DSX_Local_Workshop_CR	Standard	Admin		2018-02-12
Workshop	Standard	Admin		2018-02-12
dsx-samples	-	Viewer	-	2018-02-12

Helpful links

Docs
Find the information you need. Watch videos of key tasks.

Discussion forum
Stack Overflow is a community of 6.9 million programmers just like you, helping each other. Join the conversation on Data Science Experience

Blog
Read and follow our blog to keep up with the latest updates about Data Science Experience.

Got ideas?
Have feedback on Data Science Experience? Submit your ideas in our product forum or vote on ideas submitted by others.

IBM Data Science Experience Local

Projects > Workshop > All

All Notebooks RStudio Models SPSS Modeler Streams Scripts Data Sets Other Files Published Assets

Notebooks

view all (0) add notebook

NAME	STATUS	ENVIRONMENT	TOOL	LAST MODIFIED
you have no notebooks				

RStudio

view all (0) open RStudio

NAME	TYPE	LAST MODIFIED
you have no rstudio files		

Models

view all (0) add model

NAME	TYPE	STATUS	LAST MODIFIED
you have no model files			

SPSS Modeler Streams

view all (1) add stream

NAME	TYPE	LAST MODIFIED	ACTIONS
Drug Study Example	SPSS	02-14-2018	

Scripts

view all (0) add script

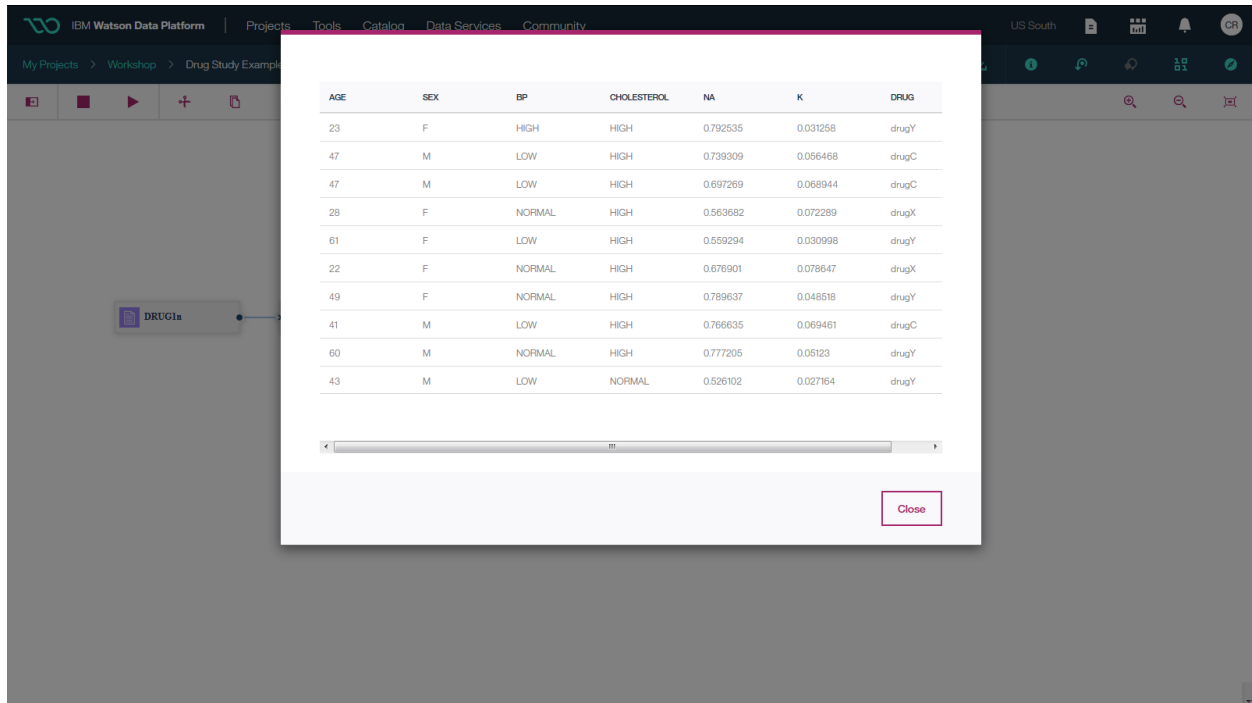
Podemos crear un flujo nuevo o si somos usuarios de SPSS Modeler, podemos importar modelos que ya tengamos hechos o empezar con un ejemplo.

Vamos a comenzar con un ejemplo:



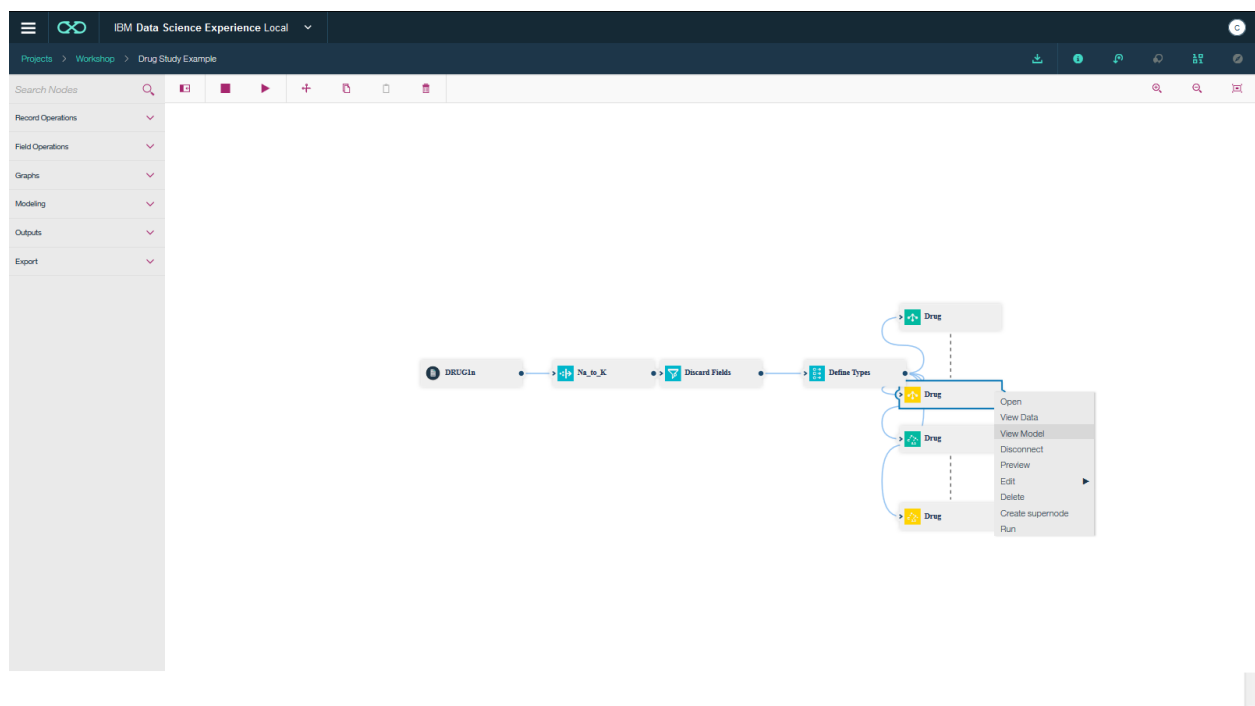
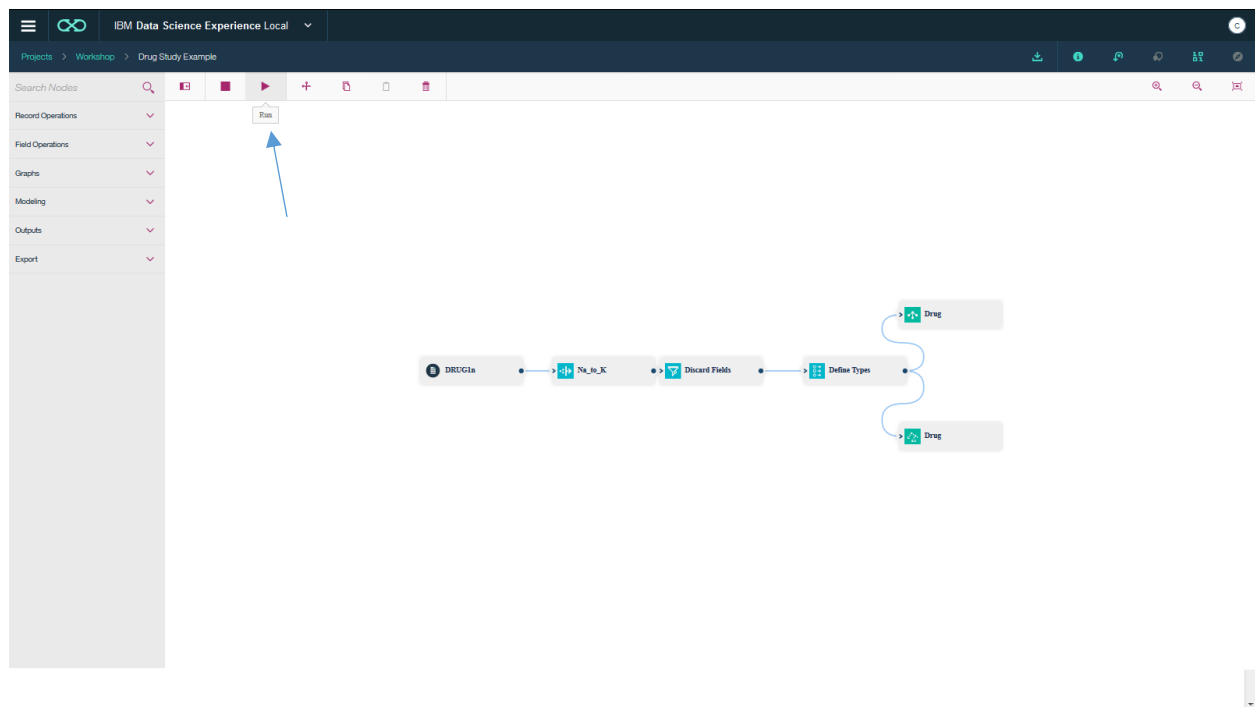
13

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.

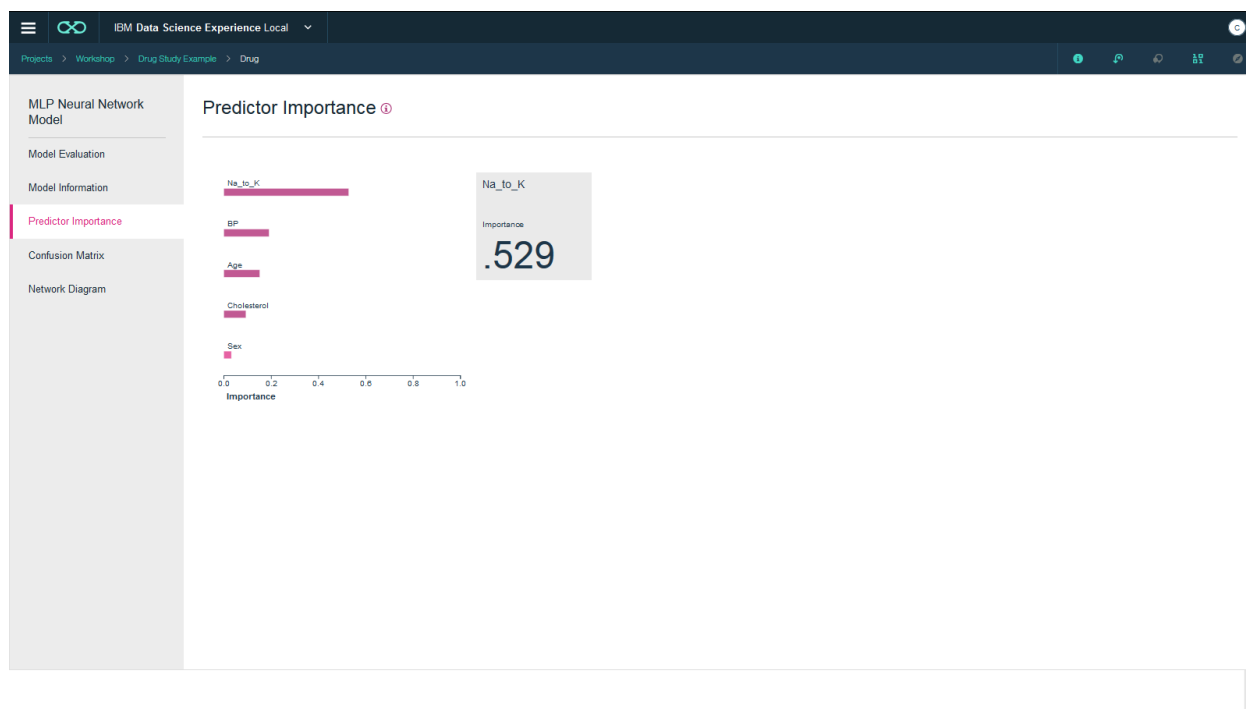


AGE	SEX	BP	CHOLESTEROL	NA	K	DRUG
23	F	HIGH	HIGH	0.792535	0.031258	drugY
47	M	LOW	HIGH	0.739309	0.056468	drugC
47	M	LOW	HIGH	0.697269	0.068944	drugC
28	F	NORMAL	HIGH	0.563682	0.072289	drugX
61	F	LOW	HIGH	0.559294	0.030998	drugY
22	F	NORMAL	HIGH	0.676901	0.078647	drugX
49	F	NORMAL	HIGH	0.789637	0.048518	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. Se propone que el lector entienda el modelo, y lo ejecute:



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



IBM Data Science Experience Local

Projects > Workshop > Drug Study Example > Drug

MLP Neural Network Model

Model Evaluation

Model Information

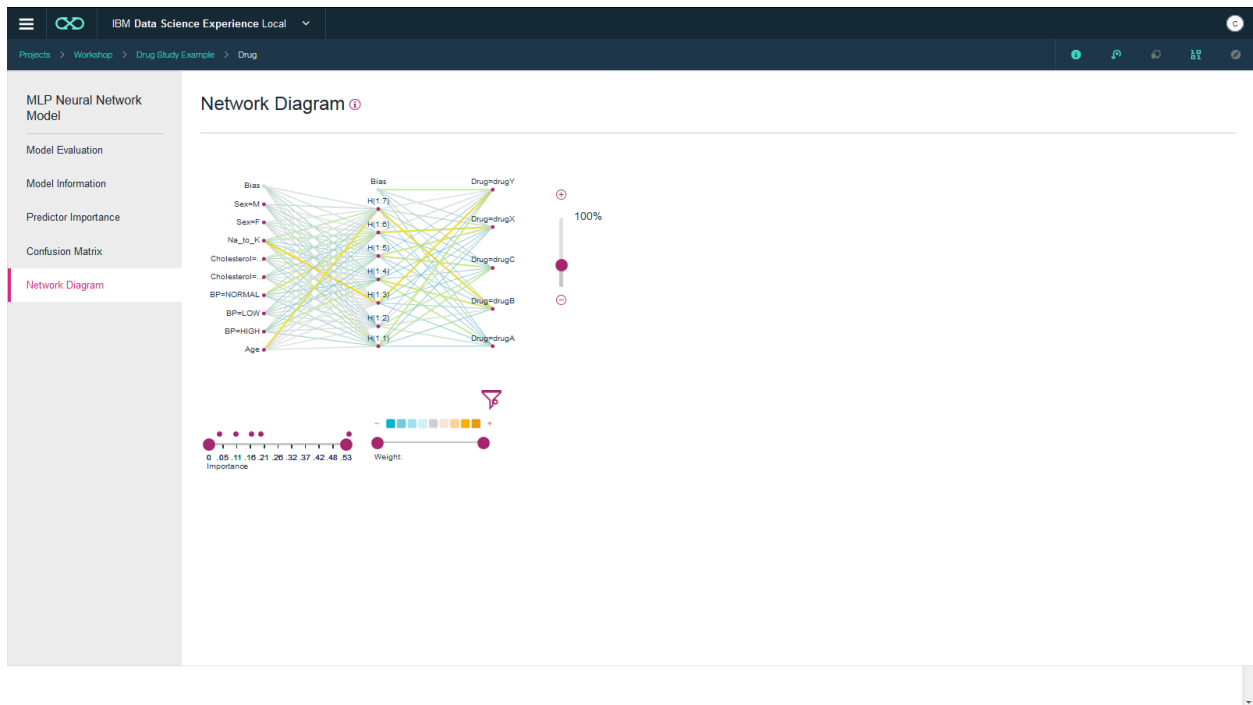
Predictor Importance

Confusion Matrix

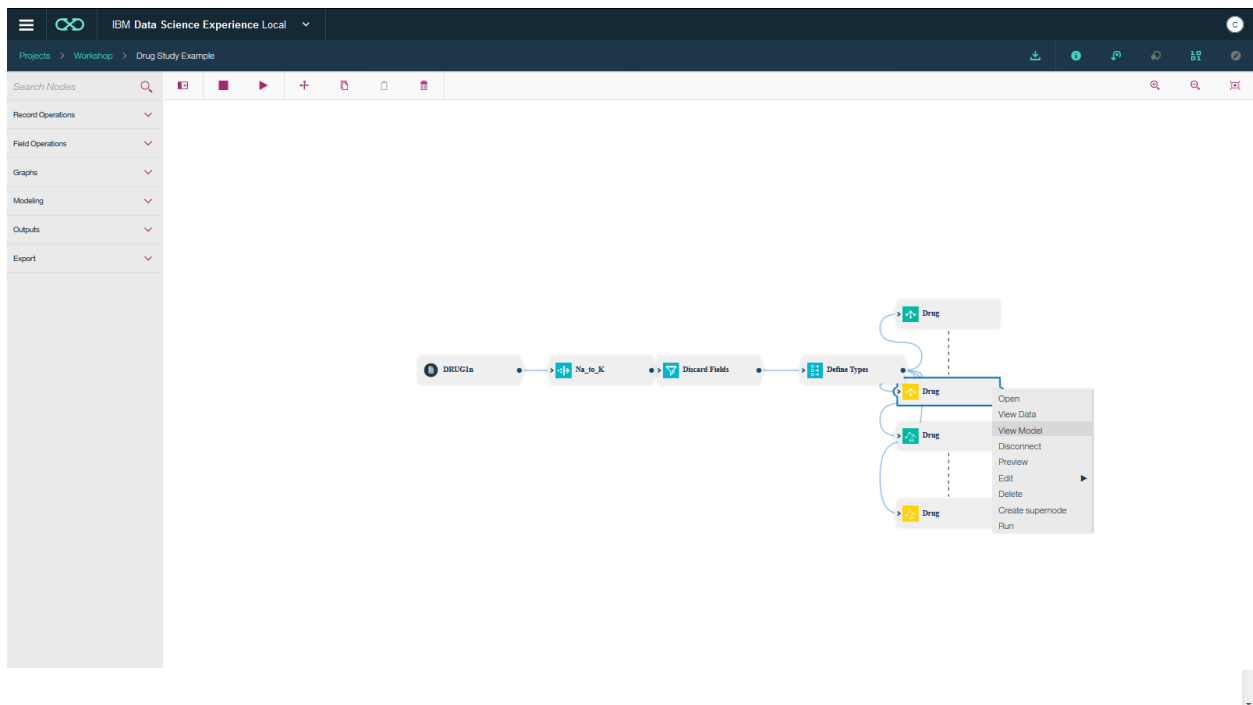
Network Diagram

Confusion Matrix ⓘ

Observed	Predicted					Percent Correct
	drugA	drugB	drugC	drugX	drugY	
drugA	23	0	0	0	0	100.0%
drugB	0	16	0	0	0	100.0%
drugC	0	0	16	0	0	100.0%
drugX	0	0	0	54	0	100.0%
drugY	0	0	0	0	91	100.0%
Percent Correct	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%



Si damos a **view data** podemos visualizar los datos y hacer auditoria de mis datos, ver la calidad, ver los estadísticos, ver gráficos, modificarlos, descargarlos, seleccionar datos desde el gráfico y crear nodos directamente desde el propio gráfico, etc.



IBM Data Science Experience Local

Projects

Workshop

Drug Study Example

Spreadsheet

Data Audit

Chart

Preference

	Age	Sex	BP	Cholesterol	Drug	Na_to_K	\$N-Drug	\$NC-Drug	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	23	F	HIGH	HIGH	drugI	25.5482629462098	drugI	0.0999906926913030																		
2	47	M	LOW	HIGH	drugC	13.092530280237954	drugC	0.0714542690229971																		
3	47	M	LOW	HIGH	drugG	10.13555926439989	drugG	0.0900318307707856																		
4	26	F	NORMAL	HIGH	drugK	7.797017894838772	drugK	0.00054937341953756																		
5	61	F	LOW	HIGH	drugI	18.042935993933009	drugY	0.0999966031181702																		
6	22	F	NORMAL	HIGH	drugK	8.60062543972352	drugK	0.0907488893285983																		
7	49	F	NORMAL	HIGH	drugI	16.275105001442763	drugY	0.0945744231486006																		
8	41	M	LOW	HIGH	drugC	11.036912799684884	drugY	0.0697773494947042																		
9	60	M	NORMAL	HIGH	drugY	15.170895959396791	drugY	0.06480333431903805																		
10	43	M	LOW	NORMAL	drugI	19.36761890737381	drugY	0.0999985649826816																		
11	47	F	LOW	HIGH	drugC	11.76144980704272	drugC	0.089186959945347																		
12	34	F	HIGH	NORMAL	drugY	19.1988972307953	drugY	0.0991655922920844																		
13	43	M	LOW	HIGH	drugY	15.376400456946078	drugY	0.0922630276784077																		
14	74	F	LOW	HIGH	drugI	20.849569162397972	drugI	0.0999981361902288																		
15	50	F	NORMAL	HIGH	drugI	12.72836341405821	drugK	0.0988974926939829																		
16	16	F	HIGH	NORMAL	drugI	15.055005903505468	drugY	0.02680454454064203																		
17	69	M	LOW	NORMAL	drugK	11.635087638912003	drugK	0.09068070444403																		
18	43	M	HIGH	HIGH	drugA	13.970583547734739	drugA	0.0691605870550982																		
19	23	M	LOW	HIGH	drugC	7.297576032819592	drugC	0.0900229542825006																		
20	52	F	HIGH	NORMAL	drugI	25.07404323447828	drugY	0.0999930734588394																		
21	57	M	LOW	NORMAL	drugY	19.12782886992907	drugY	0.0999995027848801																		
22	63	M	NORMAL	HIGH	drugY	25.9165707204392	drugY	0.09999984483135086																		
23	47	M	LOW	NORMAL	drugI	30.56810970819566	drugY	0.099998492781657																		
24	48	F	LOW	HIGH	drugI	15.036308320525077	drugY	0.0305703202740586																		
25	33	F	LOW	HIGH	drugI	33.4862682769707	drugY	0.099997947495023																		
26	36	F	HIGH	NORMAL	drugI	18.629033867544596	drugY	0.088476326502128																		
27	31	M	HIGH	HIGH	drugI	30.3662290039672	drugY	0.099999905675889																		
28	49	F	NORMAL	NORMAL	drugI	9.38070733603569	drugY	0.090585974281212																		
29	39	F	LOW	NORMAL	drugI	22.69725590520873	drugY	0.000099548768878																		
30	45	M	LOW	HIGH	drugI	17.00305904518868	drugY	0.099992487354931																		
31	16	F	NORMAL	NORMAL	drugK	8.78997233875832	drugK	0.099930386264515																		

SQL Editor

select * from ThisTable

IBM Data Science Experience Local

Projects > Workshop > Drug Study Example

Spreadsheet

Data Audit

Chart

Preference

Audit Quality Statistics Pearson Correlations

Age

Statistics

Count	200
Mean	44,315
Min	15
Max	74
Range	59
Variance	273,714
Standard Deviation	16,544
Standard Error of Mean	1.17
Median	45
Mode	47

Pearson Correlations

Na_to_K	-0.063	Weak
SNC-Drug	-0.095	Weak

SNC-Drug

Statistics

Count	200
Mean	0.986
Min	0.531
Max	1
Range	0.469
Variance	0.002
Standard Deviation	0.046
Standard Error of Mean	0.003
Median	0.998
Mode	1

*Multiple modes exist. The smallest value is shown.

Pearson Correlations

Age	-0.095	Weak
Na_to_K	0.134	Weak

Na_to_K

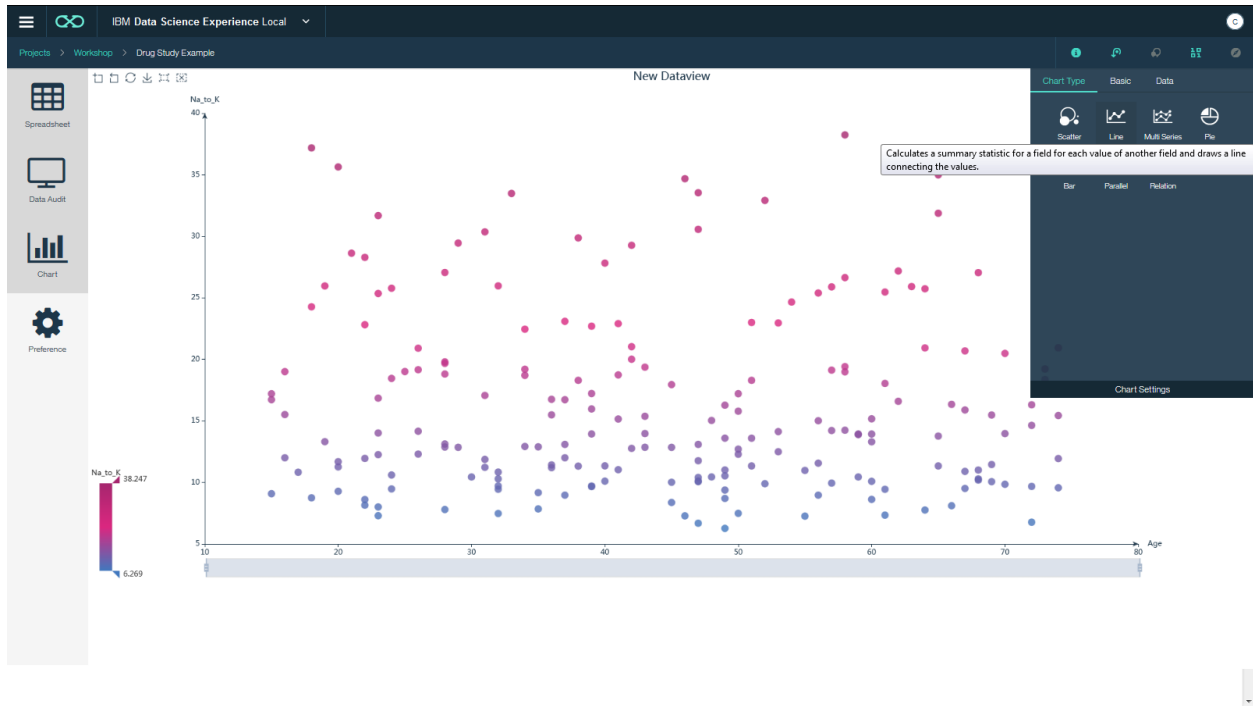
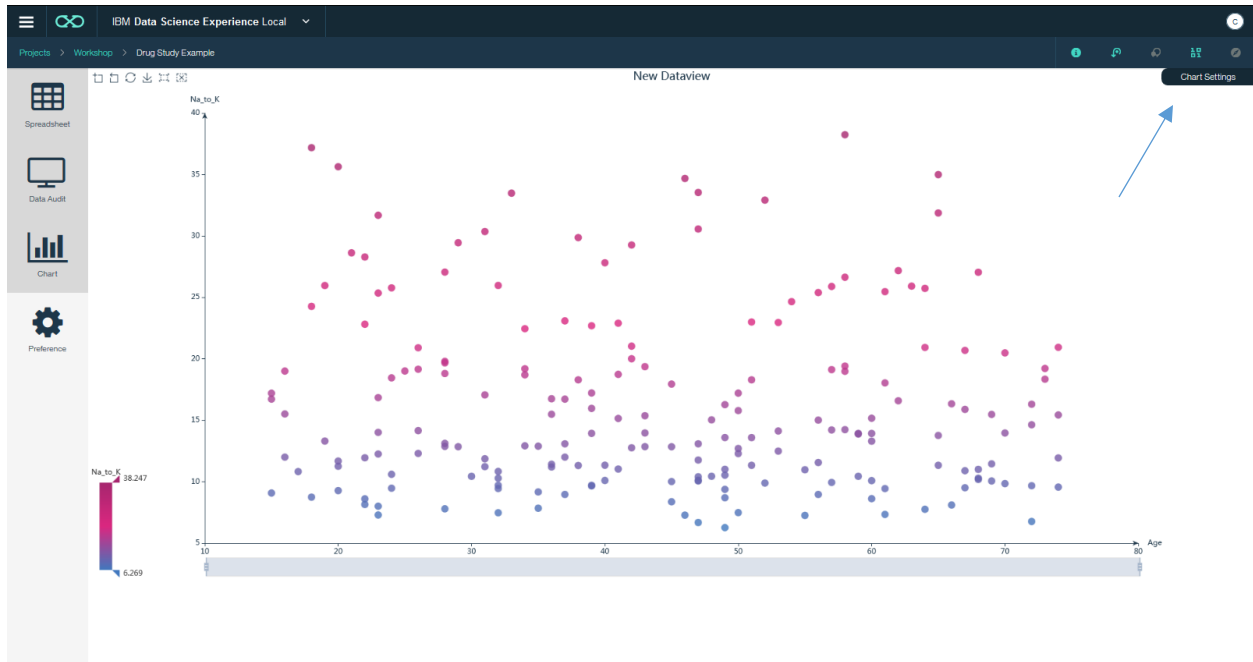
Statistics

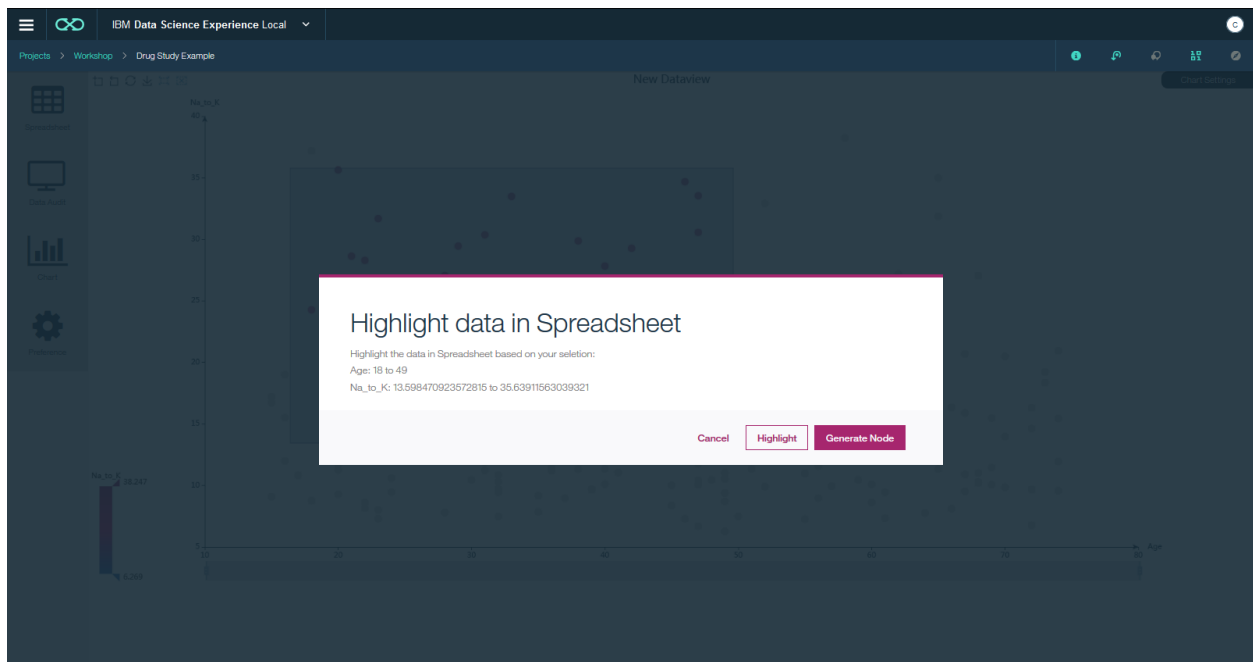
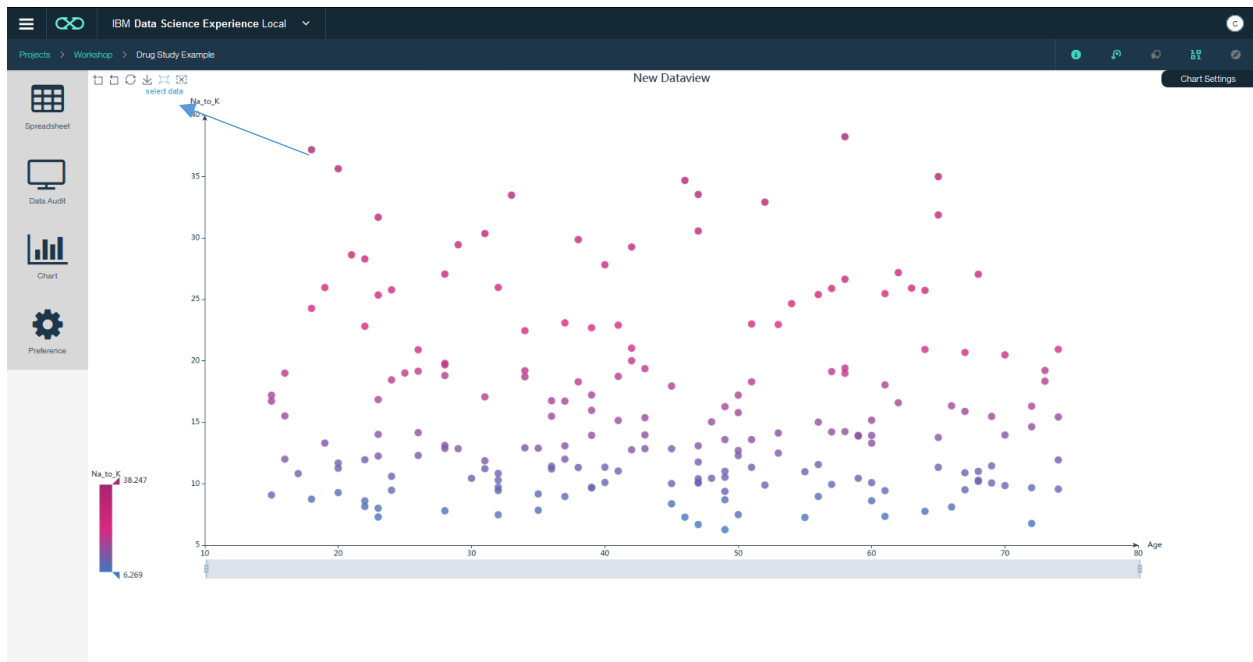
Count	200
Mean	16,084
Min	6,269
Max	38,247
Range	31,978
Variance	52,195
Standard Deviation	7,224
Standard Error of Mean	0.811
Median	13,938
Mode	10,977

*Multiple modes exist. The smallest value is shown.

Pearson Correlations

Age	-0.063	Weak
SNC-Drug	0.134	Weak





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

2. DSX Local Workshop

En esta parte del workshop, aprenderás a desarrollar e implementar aplicaciones en DSX Local. El workshop se ha dividido en varias partes independientes para aquellos que están interesados en una determinada herramienta de desarrollo o una determinada tarea de implementación.

Vamos a seguir el siguiente repositorio:

<https://github.com/CandelaRetolaza/Data-Science-Experience-en-LOCAL>

3. 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 se propone seguir el siguiente Workshop:

https://github.com/jc900/FastStart_DDLabs

4. 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 múltiples 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.

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

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-0bfb-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=c31fd96333af39811a78fe7773e421a50c7e20a450ba6b653bf4e0db39dc8f3f

Lasagne:

→ Tutorial https://dataplatform.ibm.com/analytics/notebooks/c1bda39b-3fcd-4dae-a109-e71d11113633/view?access_token=18379e532a9953d4e97f2a75eee37a8ece9ee4745676e1a647493fbfdd7b16fb

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

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

Más información sobre Deep Learning en IBM Data Science Experience:

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

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/aruzga7/Self-Driving-Car-in-DSX>