



IBM Watson Studio Hands-On

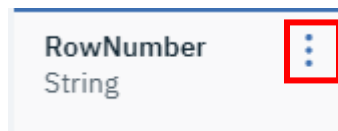
Samir Char Iglesias – IBM Global Entrepreneur Program

Preparar datos

1. Entre al su proyecto y de click sobre el archivo “Churn.csv” para tener una vista previa de los datos. Notará que el tipo de dato de todas las columnas es “String”.

ROWNUMBER	CUSTOMERID	SURNAME	CREDITSCORE	GEOGRAPHY	GENDER	AGE	TENURE	BALANCE	NUMOFPRODUC...	HASCRCARD
Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String
1	15634602	Hargrave	619	France	Female	42	2	0	1	1
2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0
3	15619304	Onio	502	France	Female	42	8	159660.8	3	1
4	15701354	Boni	699	France	Female	39	1	0	2	0
5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1
6	15574012	Chu	645	Spain	Male	44	8	113755.78	2	1
7	15592531	Bartlett	822	France	Male	50	7	0	2	1
8	15656148	Obinna	376	Germany	Female	29	4	115046.74	4	1
9	15792365	He	501	France	Male	44	4	142051.07	2	0

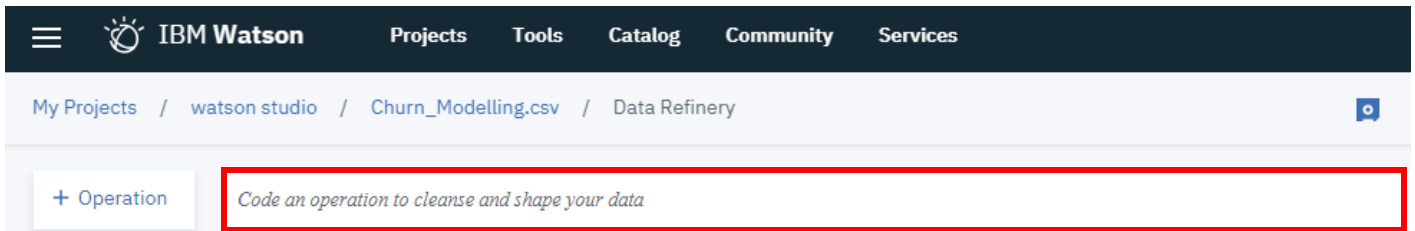
2. Para resolver esto se puede usar el servicio de “Data Refinery”. Diríjase a su proyecto y al final de la página encontrará la sección de “Data Flow” y al lado de click en “New data flow”. En el panel de la izquierda deberá ver la opción “Churn.csv”, seleccione esta opción y de click en “Add” en la parte baja de la página.
3. La columna “RowNumber” no se necesita ya que es un indicador del número de la fila. Para eliminarla posicione el mouse sobre el nombre “RowNumber” y de click sobre el ícono que se muestra a continuación:



En la ventana emergente seleccione “Remove”.

4. Repita el punto anterior para las columnas “CustomerId” y “Surname” ya que son columnas que proveen información muy específica y no aportan al modelo.
5. Finalmente, es necesario cambiar los tipos de las columnas de “String” a numérico es decir como número entero o decimal. Para hacer esto aprovecharemos la opción

de Data Refinery para usar la librería de Dplyr de R. En la parte superior de la pantalla encontrará el campo para escribir el código:



En este campo copie y pegue el siguiente código:

```
mutate_at(vars(`CreditScore`, `Age`, `Tenure`, `NumOfProducts`, `HasCrCard`, `IsActiveMember`, `Exited`), funs(as.integer))
```

“mutate_at” aplica una función a las columnas especificadas. “vars” indica las columnas que queremos modificar y “funs” la función que le queremos aplicar. “as.integer” significa que los valores de las columnas especificadas se tomarán como números enteros.

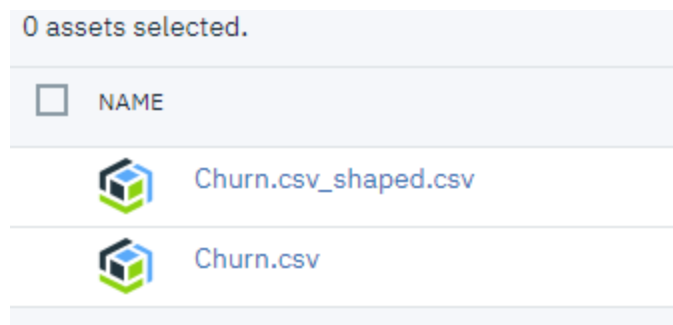
Este proceso lo repetimos para las columnas de “EstimatedSalary” y “Balance” pero transformando las columnas a números decimales. Se aplica la siguiente línea de código:

```
mutate_at(vars(`Balance`, `EstimatedSalary`), funs(as.numeric))
```

6. Ya que los datos están listos para usar, en la parte superior derecha de click el siguiente ícono para guardar y correr el flujo:



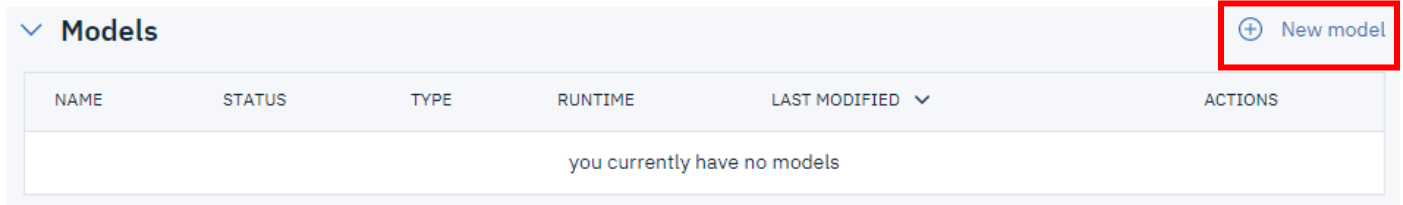
En la ventana que se abre, en la parte inferior derecha de click en la opción “Save and Run”. Regrese a su proyecto y deberá ver un nuevo data asset llamado Churn.csv_shaped.csv



Este ya es la data modificada y es la que se usará de ahora en adelante.

Crear el modelo y desplegar modelo

7. Baje hasta encontrar la opción de “New model”



Nombre el modelo como “Churn_model”. Deberá ver asociados los servicio de Machine Learning y Spark, seleccione la opción “Model builder” y “Manual”. De click en “Create”. Seleccione el data asset modificado en la sección anterior y presione “next”.

8. Seleccione la variable “Exited” para predecir y todas las demás columnas se usarán como predictores. Escoja la opción de “Clasificación” (Si no se escogió de manera automática). Arriba a la izquierda seleccione “Add estimators” y escoja los algoritmos “Random forest classifier” y “Gradient boosted tree classifier”. Seleccione “Next” y espere que se entrenen los modelos.
9. Deberá ver la siguiente ventana en donde encontrarán todas las métricas asociadas al desempeño de los algoritmos. Seleccione “RandomForestClassifier” y luego “save”.

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	0.82973	0.64276	6 May 2018, 4:42 PM	⋮
<input type="radio"/>	GBClassifier	Trained & Evaluated	Fair	0.702	0.63533	6 May 2018, 4:43 PM	⋮

10. Diríjase a la ventana “Deployments”, seleccione “Add deployment”, luego “Web Service” y nombre este “Churn_model_web” y finalmente guardar.

Probar despliegue del modelo

11. Después de guardar deberá ver el modelo. De click al nombre “Churn_model_web” para entrar al despliegue y poderlo probar.

OverviewEvaluationDeployments

Add Deployment

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
Churn_model_web	DEPLOY_SUCCESS	Web Service	

12. Entre a la ventana de “Test”, provea valores de prueba y prediga.