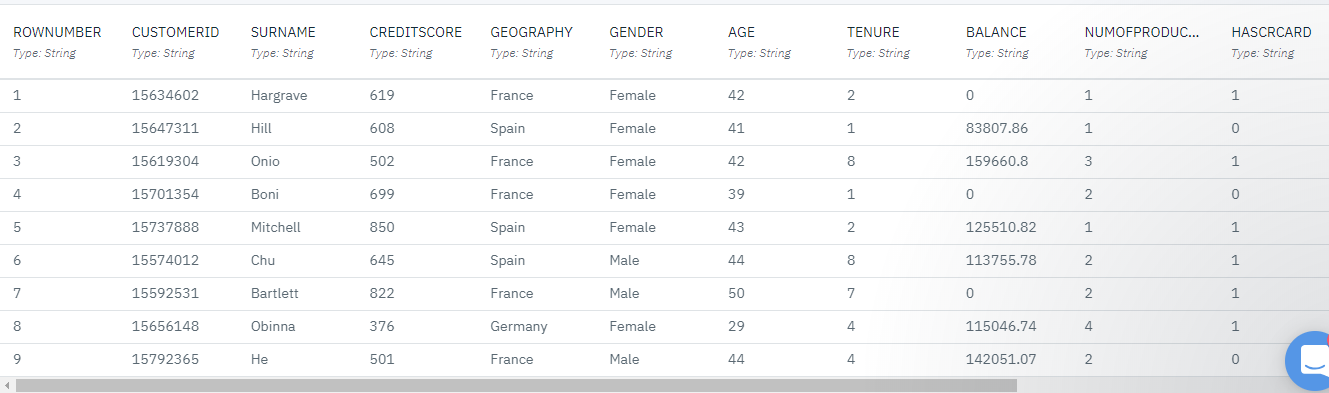


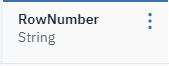
****

**IBM Watson Studio Hands-On**

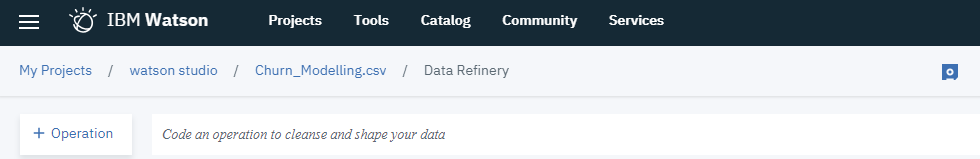
**Samir Char Iglesias – IBM Global Entrepreneur Program**

**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”.
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”.

1. Repita el punto anterior para las columnas “CustomerId” y “Surnmae” ya que son columnas que proveen información muy específica y no aportan al modelo.
2. 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” acplica 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.

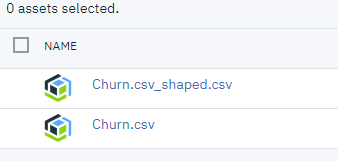
Este proceso lo repetimos para las columnas de “EstimatedSalary” y “Balance” pero transformando las columnas a números decimales. Se aplica la siguiente línea de código:

mutate\_at(vars(`Balance`,`EstimatedSalary`),funs(as.numeric))

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



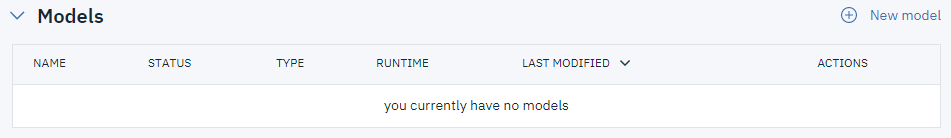
En la ventana que se abre, en la parte inferior derecha de click en la opción “Save and Run”. Regrese a su proyecto y deberá ver un nuevo data asset llamado Churn.csv\_shaped.csv



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

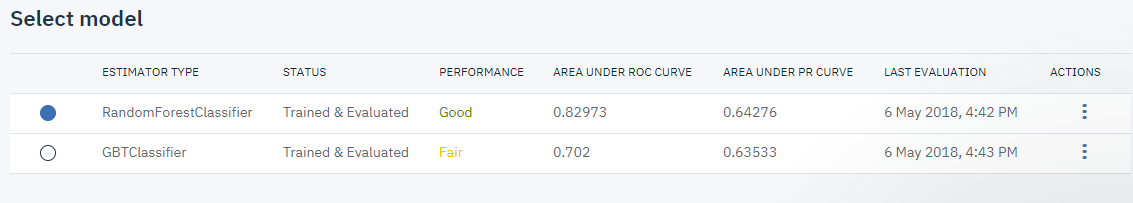
**Crear el modelo y desplegar modelo**

1. 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”.

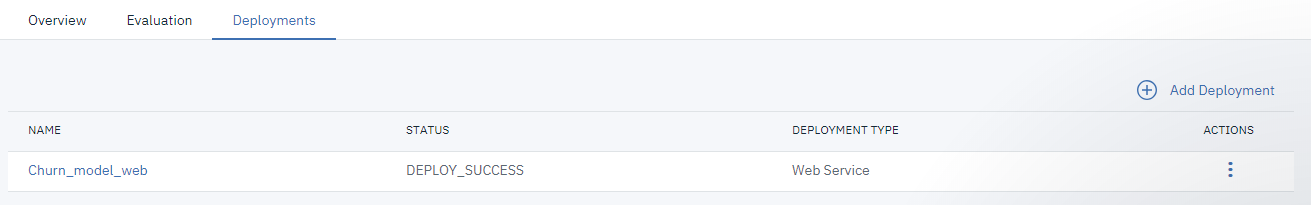
1. Seleccione la variable “Exited" para predecir y todas las demás columnas se usarán como predictores. Escoja la opción de “Clasificación” (Si no se escogió de manera automática). Arriba a la izquierda seleccione “Add estimators” y escoja los algoritmos “Random forest classifier” y “Gradient boosted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



1. Diríjase a la ventana “Deployments”, seleccione “Add deployment”, luego “Web Service” y nombre este “Churn\_model\_web” y finalmente guardar.

**Probar despliegue del modelo**

1. Después de guardar deberá ver el modelo. De click al nombre “Churn\_model\_web” para entrar al despliegue y poderlo probar.



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