

Taller de Azure Machine Learning

Junio 2016 – Microsoft Argentina & Uruguay

Este documento presenta un recorrido paso por paso sobre cómo crear un modelo con [Azure Machine Learning](#) y está enfocado en desarrolladores de software, administradores de bases de datos, científicos de datos y arquitectos de software. Se espera que los lectores tengan conocimientos básicos de computación en la nube y de procesamiento y manejo de datos informáticos.

El objetivo de este documento es permitir a los usuarios familiarizarse con el ambiente de Azure Machine Learning a través de un ejemplo práctico.

El ejemplo elegido para este documento permitirá crear un modelo capaz de predecir, en base a diferentes datos sobre un individuo, si este tiene ingresos anuales mayores o menores a USD 50.000. Para dicho ejercicio, se utilizará un *dataset* de pruebas que contiene información anónima sobre aproximadamente 32.000 individuos.

El ejercicio está basado en el algoritmo *Two-Class Boosted Decision Tree*. Para más información sobre los algoritmos disponibles, visitar <http://aka.ms/algoritmosML>

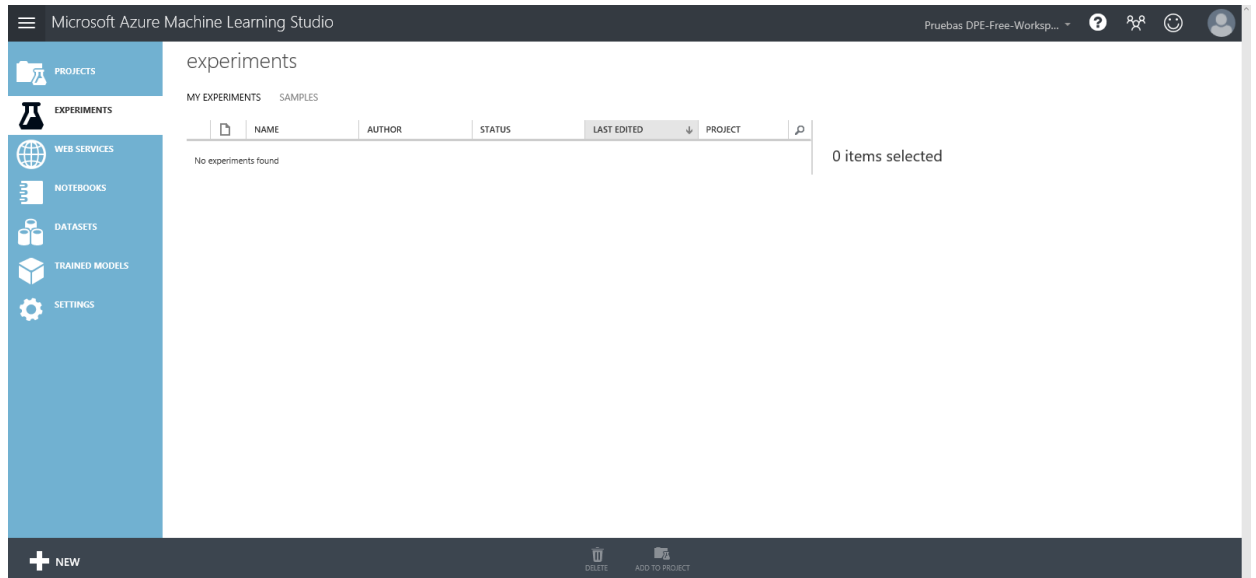
Tabla de contenidos

Sección #1: Creación de la suscripción para Azure Machine Learning	2
Sección #2: Creación del Experimento.....	2
Sección #3: Ingreso de la información	3
Sección #4: Proyección y transformación de los datos	4
Sección #5: Entrenar el modelo de Machine Learning	7
Sección #6: Evaluación y puntuación del modelo de Machine Learning	9
Sección #7: Probar el modelo de Machine Learning	10
Sección #8: Mejorar y calibrar el modelo de Machine Learning	11
Sección #9: Publicar como Web Service el modelo de Machine Learning	13
Sección #10: Probar un caso real del Web Service.....	14

Sección #1: Creación de la suscripción para Azure Machine Learning

Requisitos: contar con una Microsoft Account. De no contar con una, se puede crear en <http://signup.live.com/>.

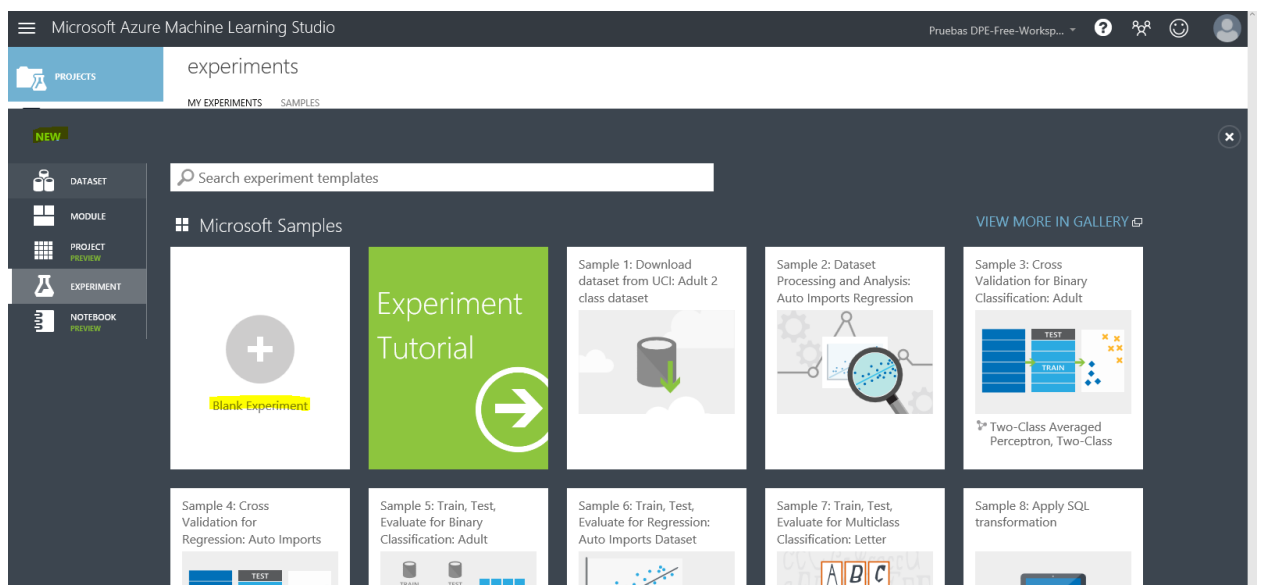
1.1 Ingresar en <http://studio.azureml.net/> e iniciar sesión (Sign in).

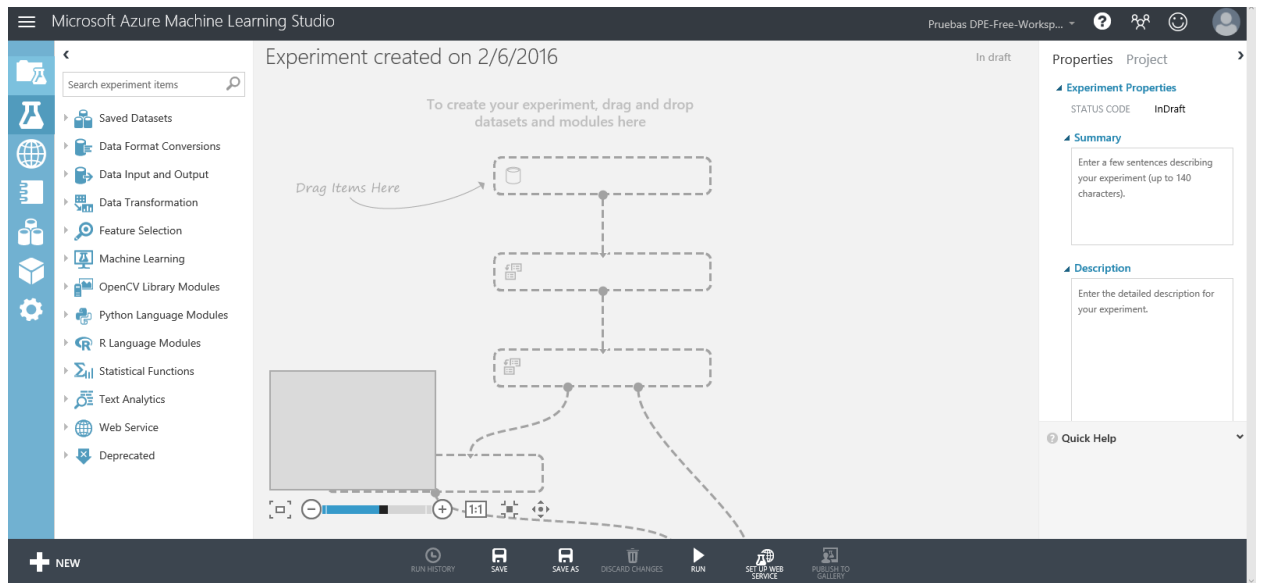


Se creará una nueva cuenta gratuita para pruebas de Azure Machine Learning.

Sección #2: Creación del Experimento

2.1 Dentro de Machine Learning Studio, ingresar en **New > Blank Experiment**

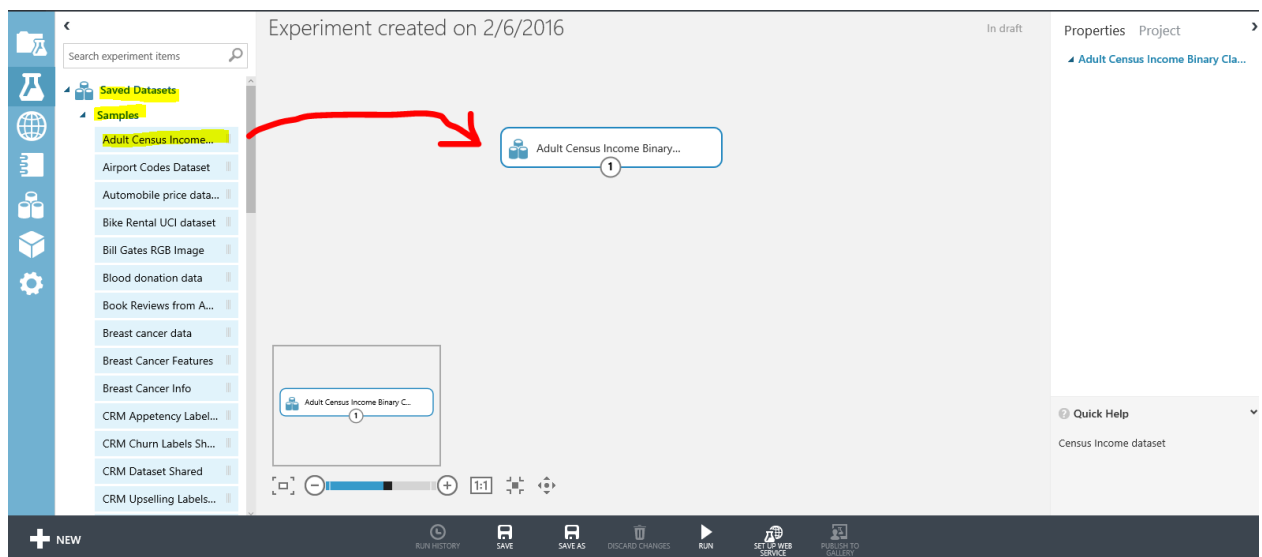




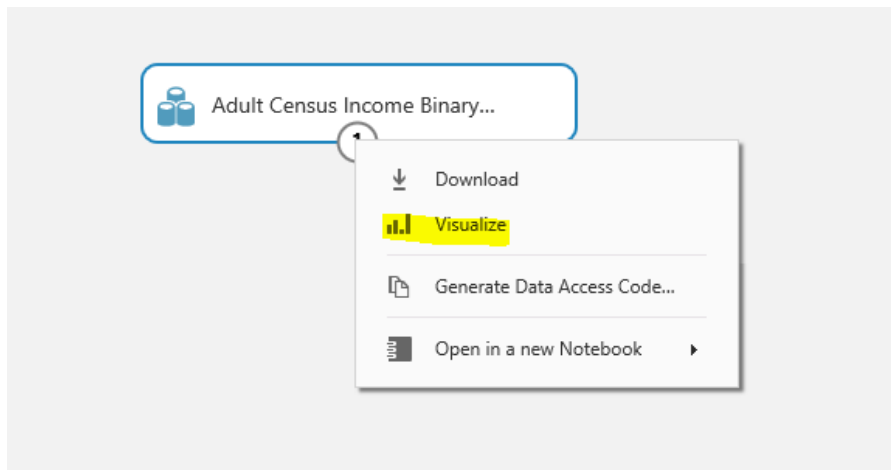
Sección #3: Ingreso de la información

3.1 Seleccionar en la solapa izquierda la opción **Saved Datasets > Samples**

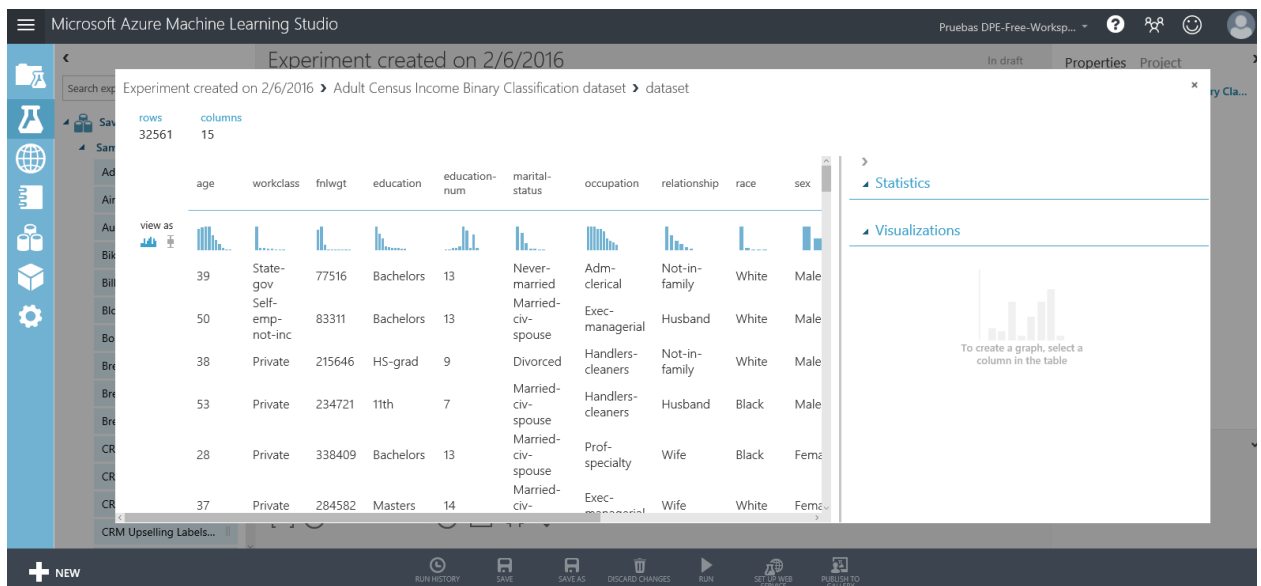
3.2 Seleccionar *Adult Census Income Binary Classification* y arrastrarlo hacia la pantalla principal (*drag & drop*).



3.3 Dar click en el número 1 bajo la caja, y luego **Visualize**

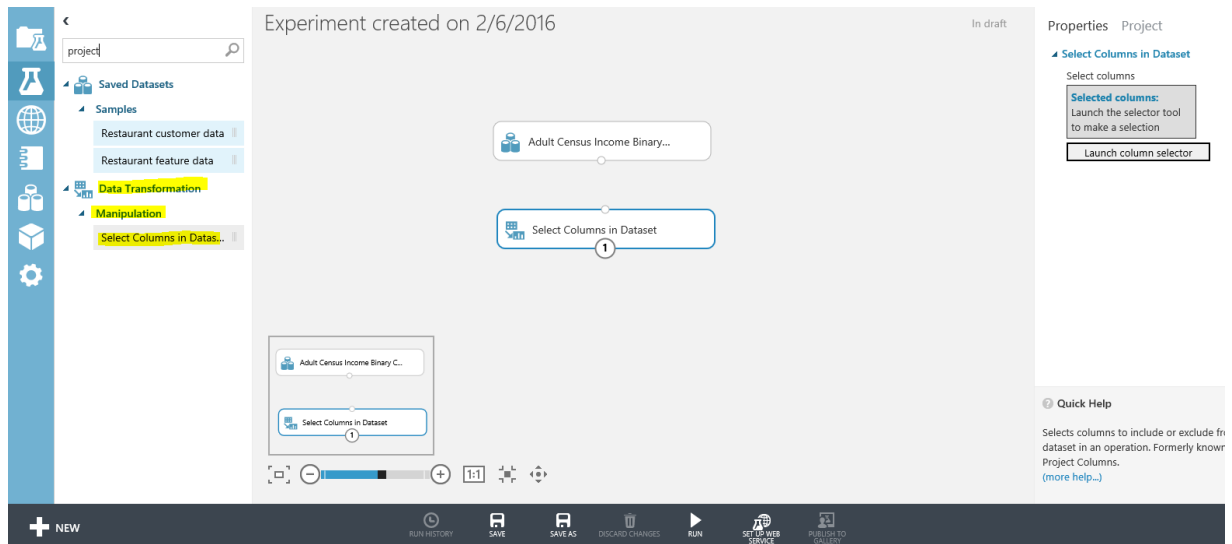


3.4 Esta opción permite visualizar el *dataset*. Notar la cantidad de columnas y registros, vista previa de la distribución para cada columna.



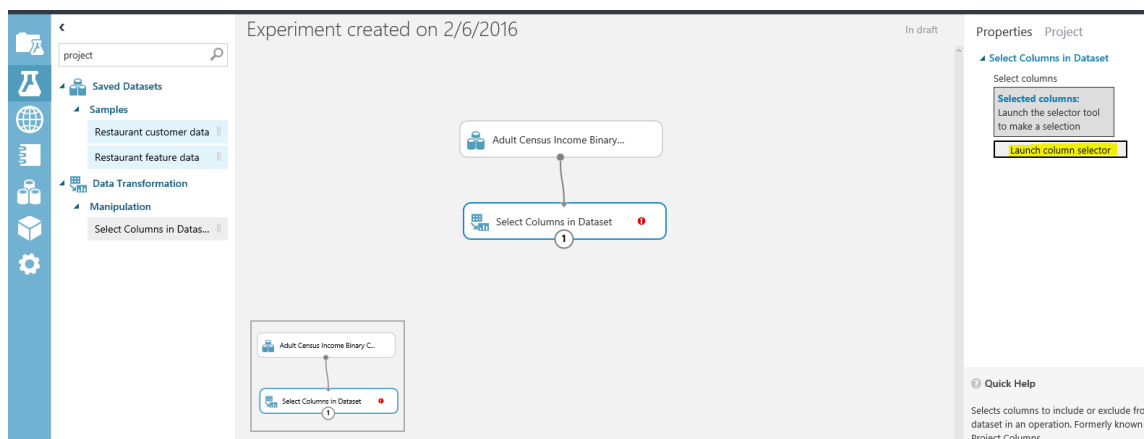
Sección #4: Proyección y transformación de los datos

4.1 Seleccionar el componente **Data Transformation** > **Manipulation** > **Select Columns in Dataset** y arrastrarlo debajo de la caja anterior



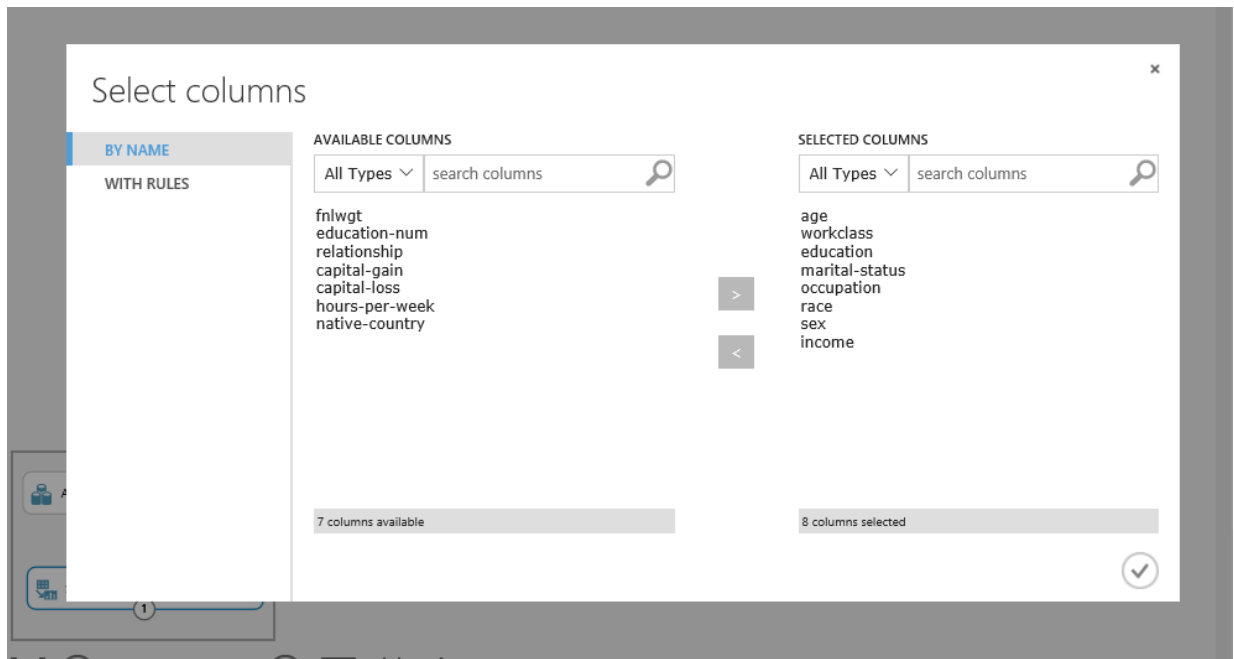
4.2 Unir la salida del *dataset* con la entrada del selector de columnas.

4.3 Notar el cartel de error. Hacer click en **Launch Column Selector**.

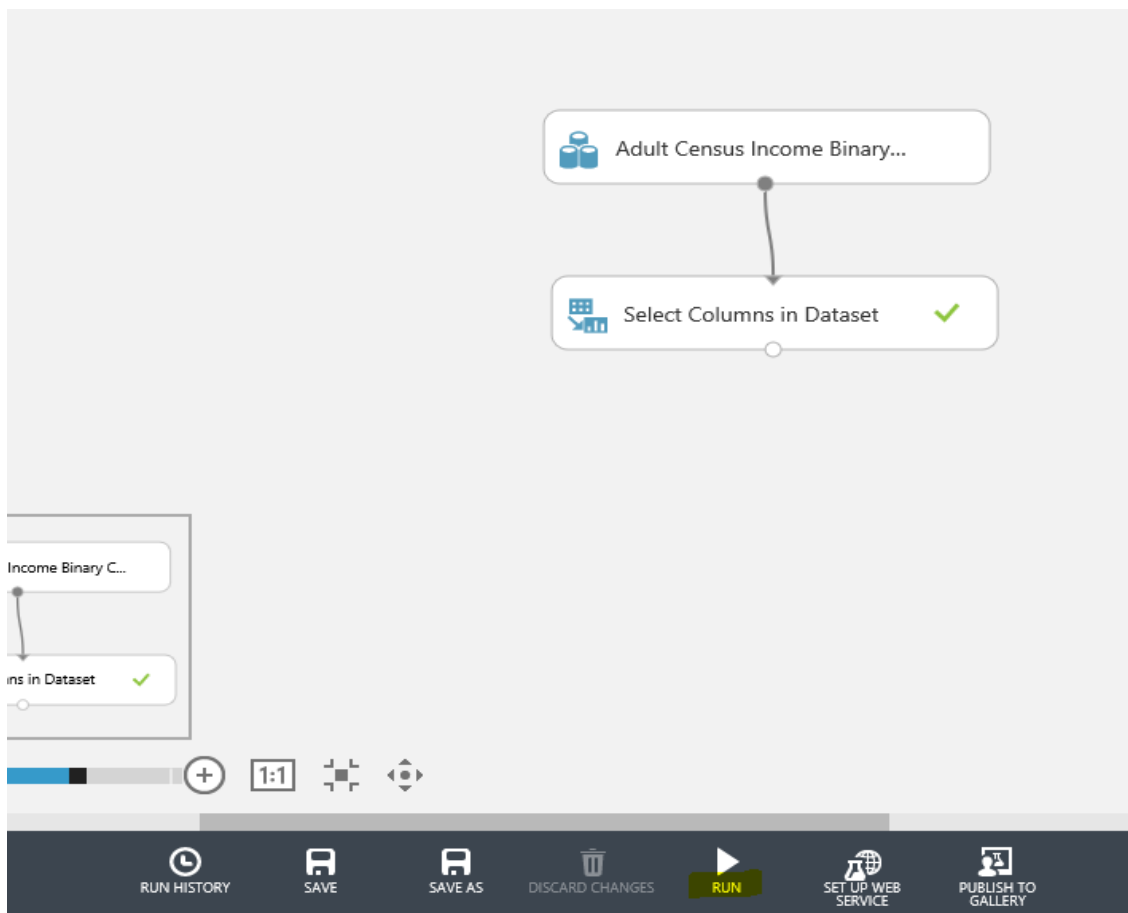


4.4 En primera instancia seleccionar los campos *age*, *workclass*, *education*, *marital-status*, *occupation*, *race*, *sex*, *income* y utilizando las flechas, moverlos al lado derecho del cuadro.

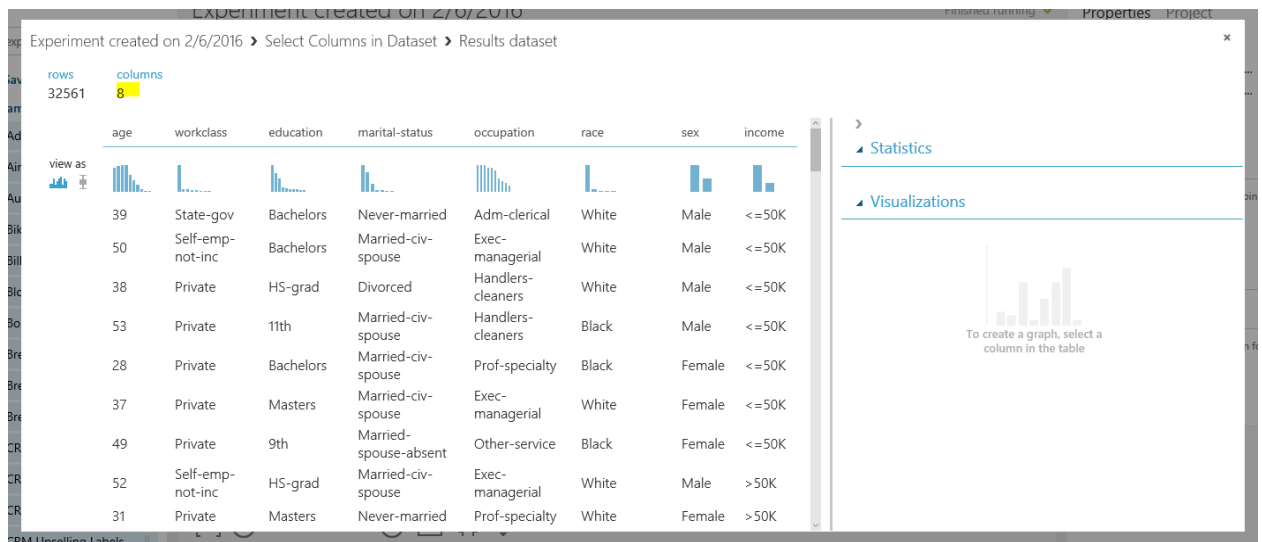
4.5 Dar click en OK.



4.6 Hacer click en **Run** en el menú inferior



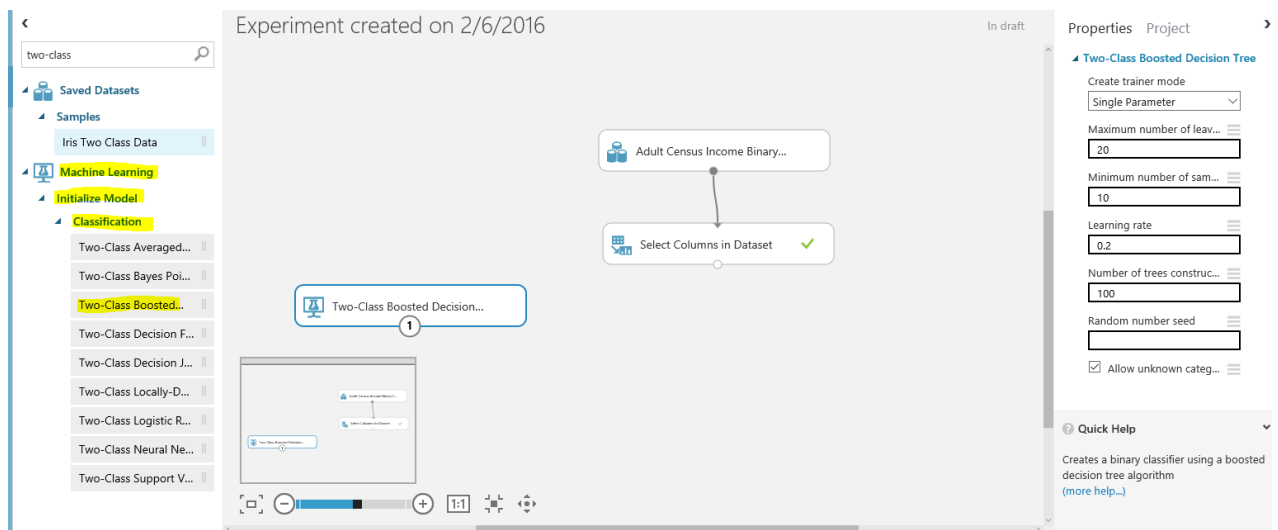
4.7 Visualizar la salida del selector de columnas.



Sección #5: Entrenar el modelo de Machine Learning

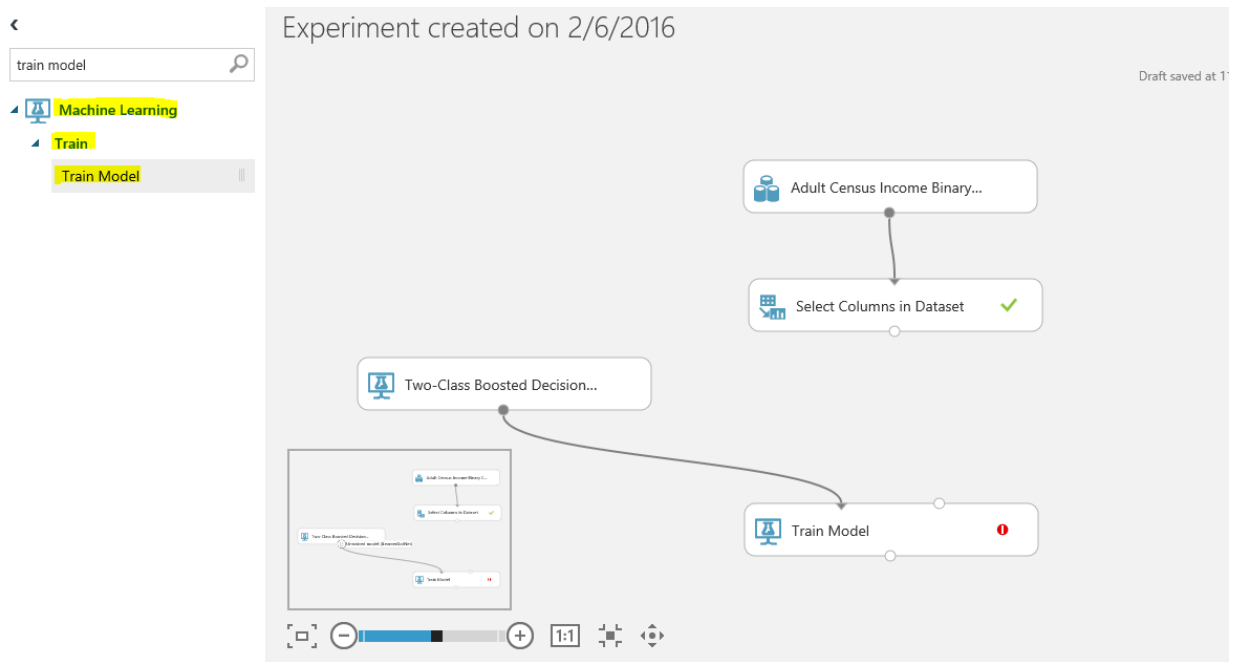
5.1 Para este ejemplo, utilizaremos el algoritmo **Two Class Boosted Decision Tree**.

5.2 Seleccionar **Machine Learning > Initialize Model > Classification > Two-Class Boosted Decision Tree** y arrastrarlo al menú principal.



5.3 Seleccionar **Machine Learning > Train > Train Model** y arrastrarlo hacia el menú principal.

5.4 Unir la salida del algoritmo con la primer entrada (izquierda) de **Train Model**

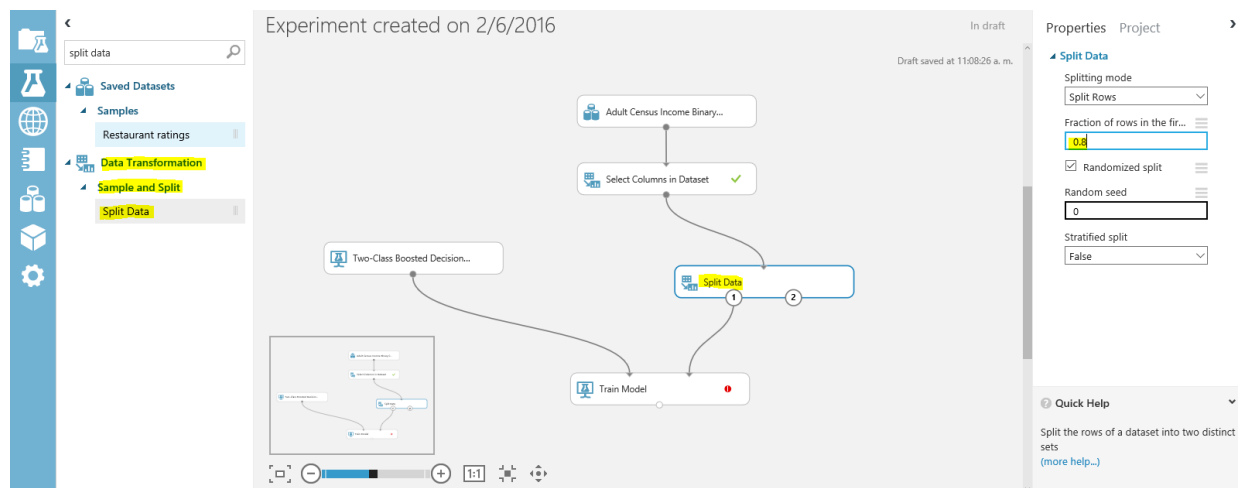


5.5 Seleccionar **Data Transformation** > **Sample and Edit** > **Split Data** y arrastrarlo al menu principal.

5.6 Unir la salida del selector con el componente **Split Data**.

5.7 Unir la primer salida de **Split Data** con la segunda entrada de **Train Model**.

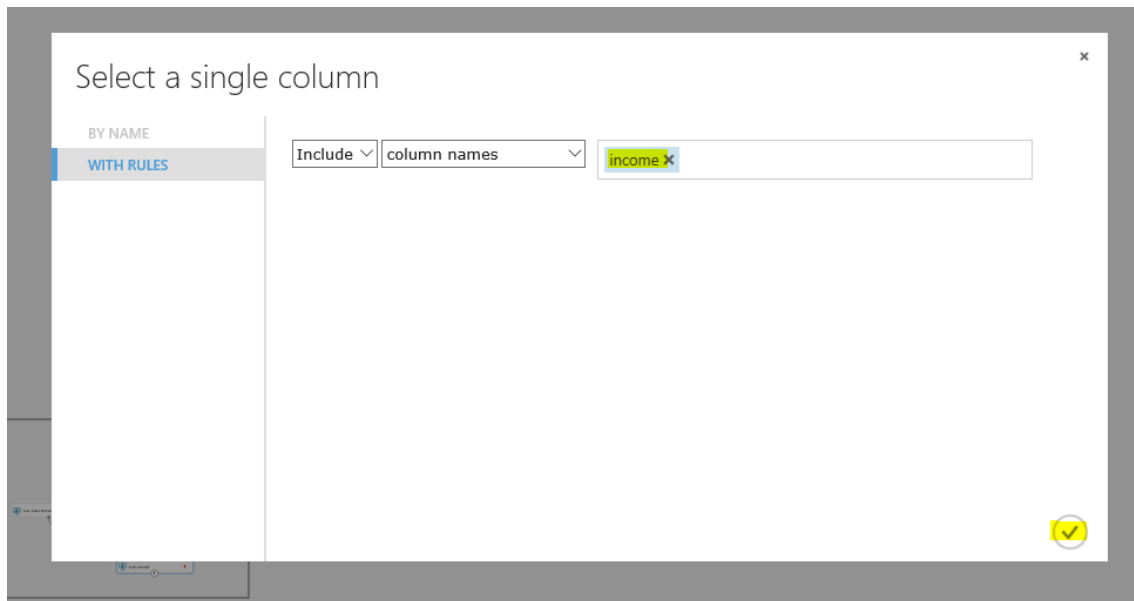
5.8 En la columna derecha, seleccionar 0.8 en el cuadro *Fractions of rows in the first output*.



5.9 Seleccionar **Train Model**, y hacer click en **Launch Column Selector** (solapa derecha).

5.10 Seleccionar únicamente la columna *income*, que será la columna que se busca predecir.

5.11 Click en **OK**.

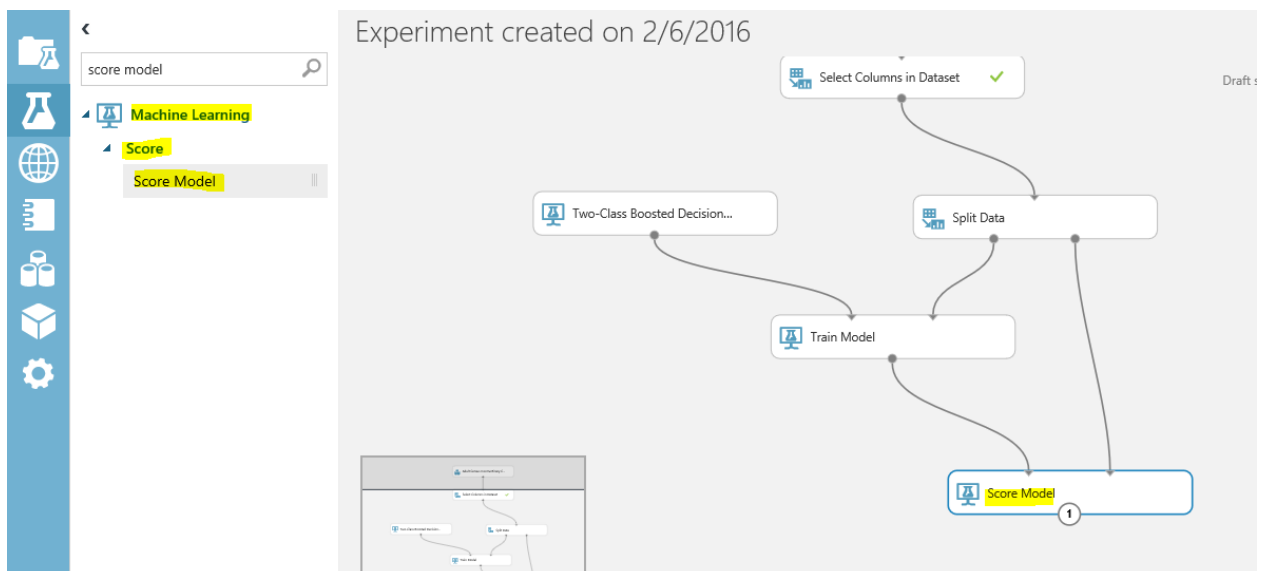


Sección #6: Evaluación y puntuación del modelo de Machine Learning

6.1 Seleccionar el componente **Machine Learning > Score > Score Model** y arrastrarlo hacia la ventana principal.

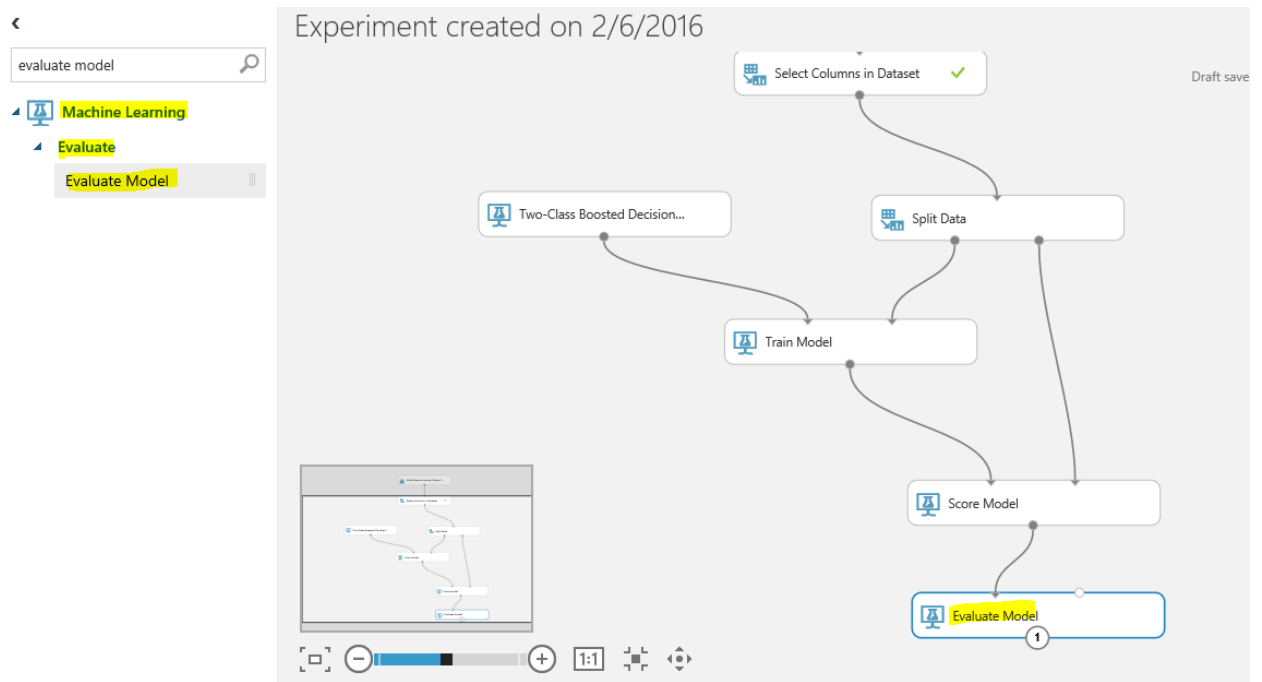
6.2 Conectar la salida de **Train Model** con la primer entrada (izquierda) de **Score Model**

6.3 Conectar la segunda salida (derecha) de **Split Data** con la segunda entrada (derecha) de **Score Model**.



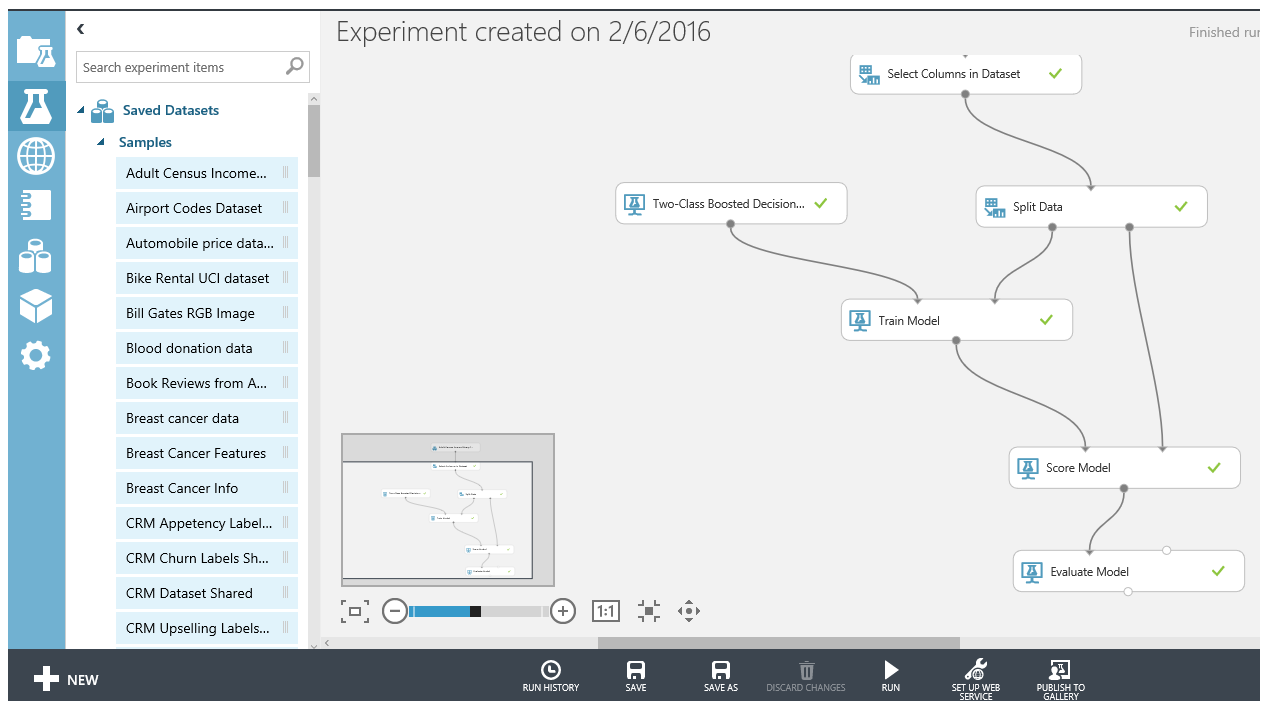
6.4 Seleccionar el componente **Machine Learning > Evaluate > Evaluate Model** y arrastrarlo a la ventana principal.

6.5 Conectar la salida de **Score Model** con la primer entrada (izquierda) de **Evaluate Model**

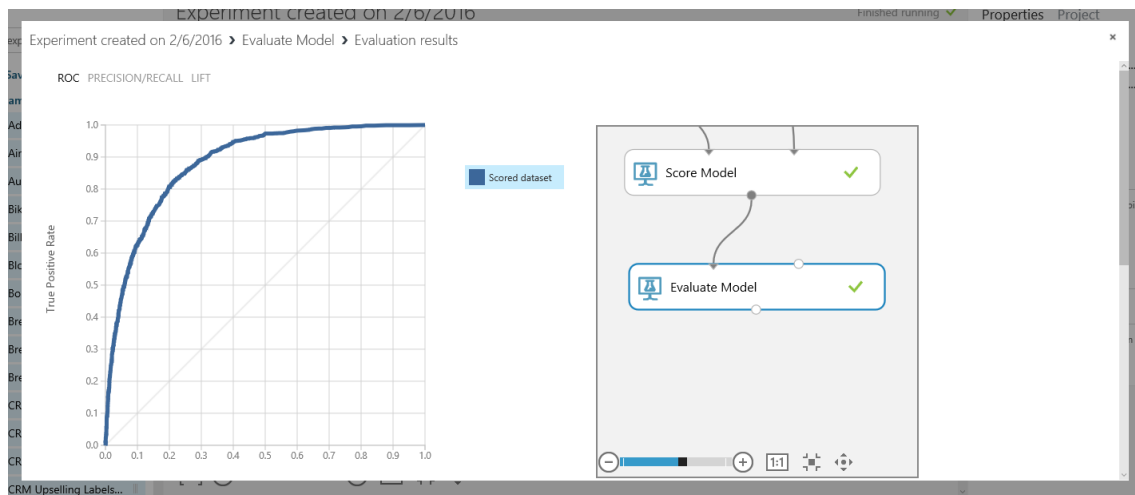


Sección #7: Probar el modelo de Machine Learning

7.1 Hacer click en **Run**. Esperar a que todas las cajas tengan íconos verdes (éxito).



7.2 Evaluar la salida del componente **Evaluate Model**, haciendo click en **Visualize**.



7.3 Notar bajo los gráficos, el desempeño del algoritmo a través de los indicadores **Accuracy** y **Precision**.

Experiment created on 2/6/2016 > Evaluate Model > Evaluation results

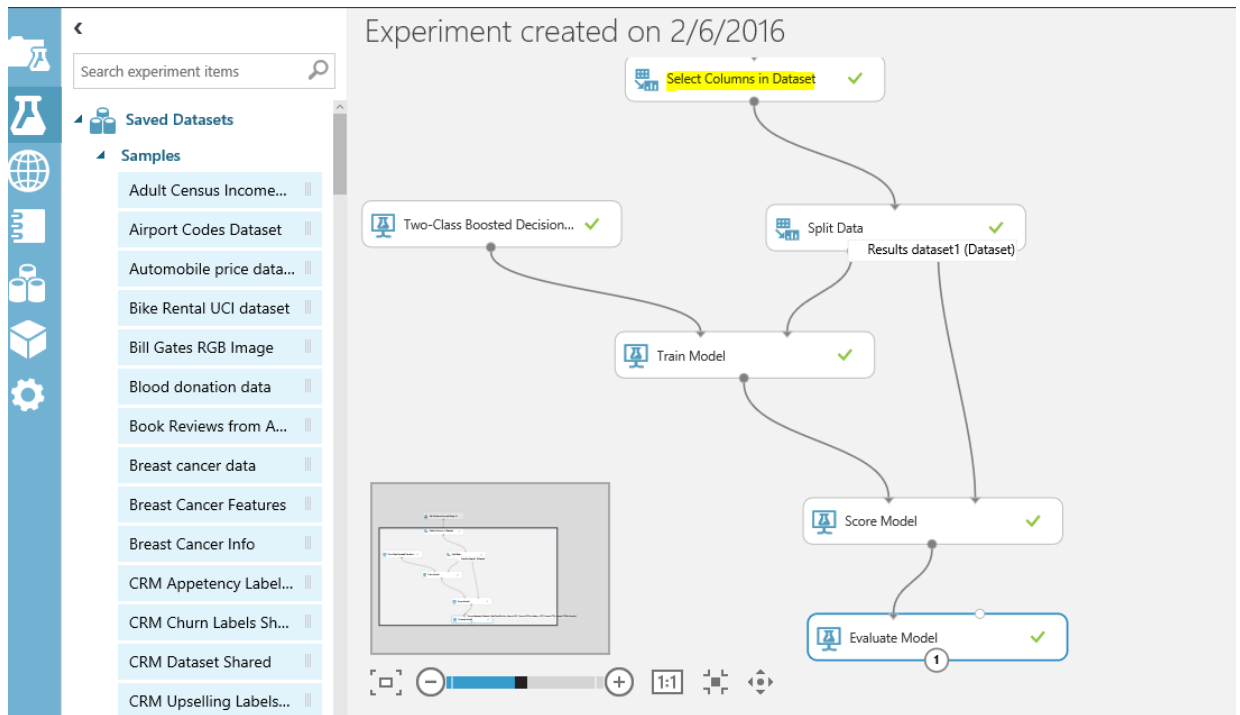
True Positive	False Negative	Accuracy	Precision	Threshold	AUC
958	579	0.833	0.681	0.5	0.883
False Positive	True Negative	Recall	F1 Score		
449	4154	0.623	0.651		
Positive Label	Negative Label				
>50K	<=50K				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	143	24	0.027	0.769	0.168	0.856	0.093	0.767	0.995	0.000
(0.800,0.900]	272	68	0.083	0.802	0.406	0.819	0.270	0.801	0.980	0.003
(0.700,0.800]	217	96	0.134	0.822	0.536	0.771	0.411	0.830	0.959	0.010
(0.600,0.700]	168	105	0.178	0.832	0.608	0.732	0.520	0.854	0.936	0.021
(0.500,0.600]	158	155	0.229	0.833	0.651	0.681	0.623	0.878	0.903	0.040
(0.400,0.500]	123	177	0.278	0.824	0.667	0.634	0.703	0.897	0.864	0.066
(0.300,0.400]	113	207	0.330	0.809	0.670	0.589	0.777	0.917	0.819	0.099
(0.200,0.300]	113	311	0.399	0.776	0.656	0.533	0.850	0.938	0.752	0.154
(0.100,0.200]	111	484	0.496	0.716	0.619	0.466	0.923	0.962	0.647	0.248
(0.000,0.100]	119	2976	1.000	0.250	0.400	0.250	1.000	1.000	0.000	0.883

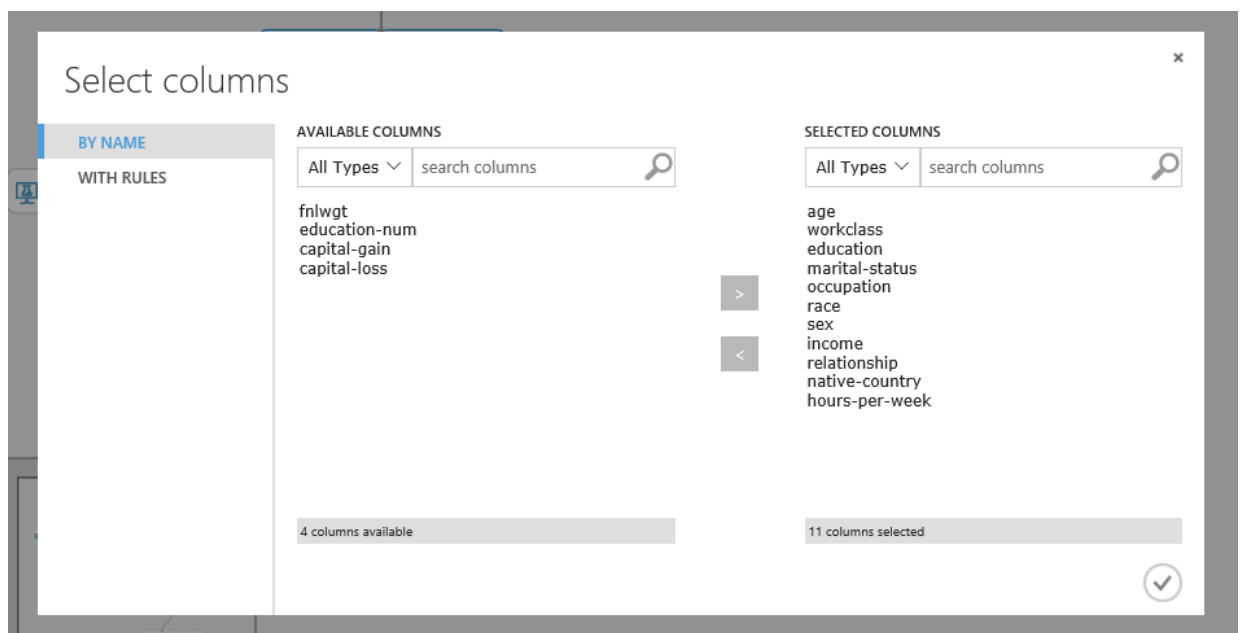
Sección #8: Mejorar y calibrar el modelo de Machine Learning

8.1 Regresar a la vista principal, seleccionar el componente **Select Columns in Dataset**.

8.2 Abrir el selector de columnas (solapa derecha).



8.3 Agregar en el cuadro derecho los campos *relationship*, *native-country*, *hours-per-week*.



8.4 Click en OK.

8.5 Click en Run.

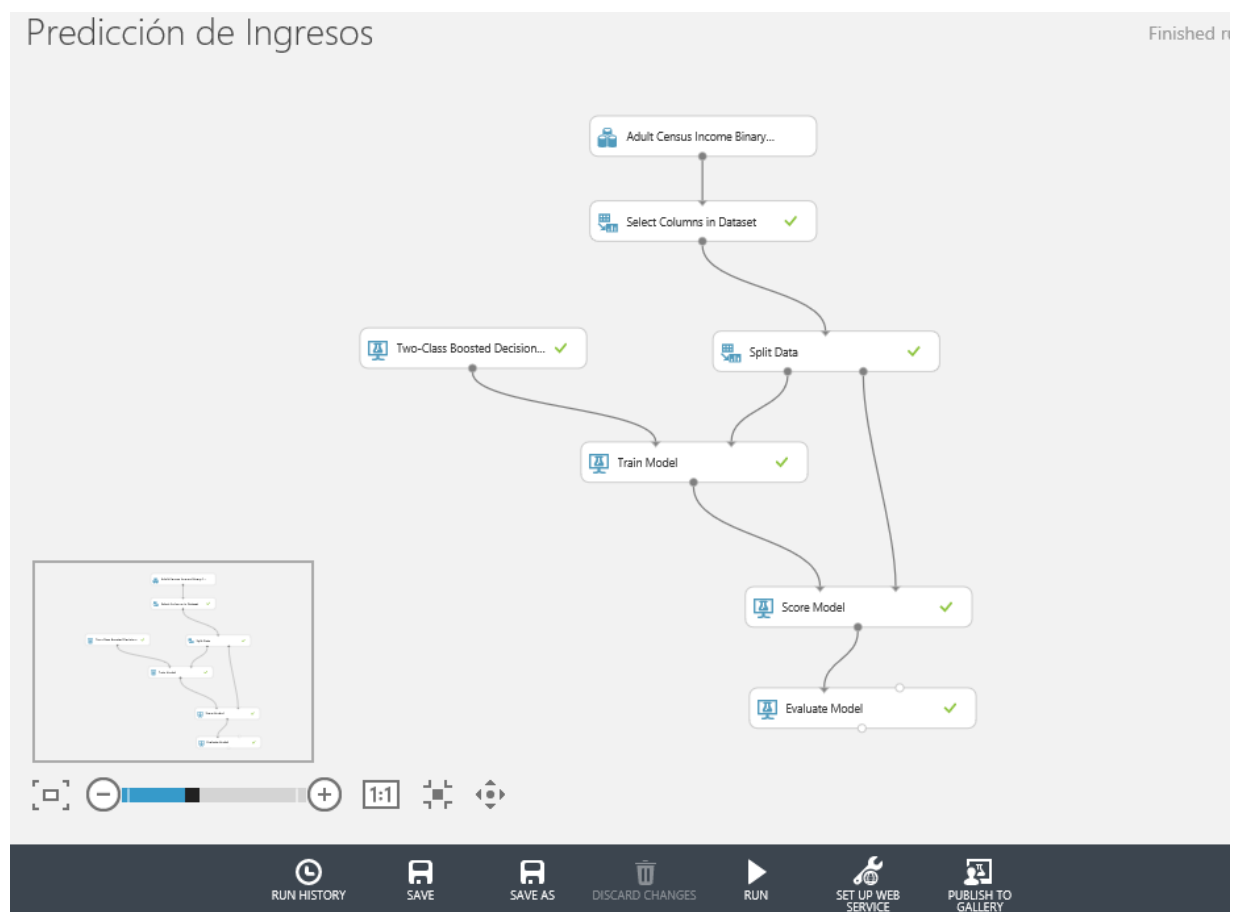
8.6 Luego de la ejecución, volver al menú **Visualize** de la salida de **Evaluate Model**

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
947	560	0.836	0.689	0.5	0.889
False Positive	True Negative	Recall	F1 Score		
427	4086	0.628	0.657		
Positive Label	Negative Label				
>50K	<=50K				

8.7 Nótese que, al agregar más campos, se mejora la precisión y exactitud a expensas de un mayor tiempo de ejecución.

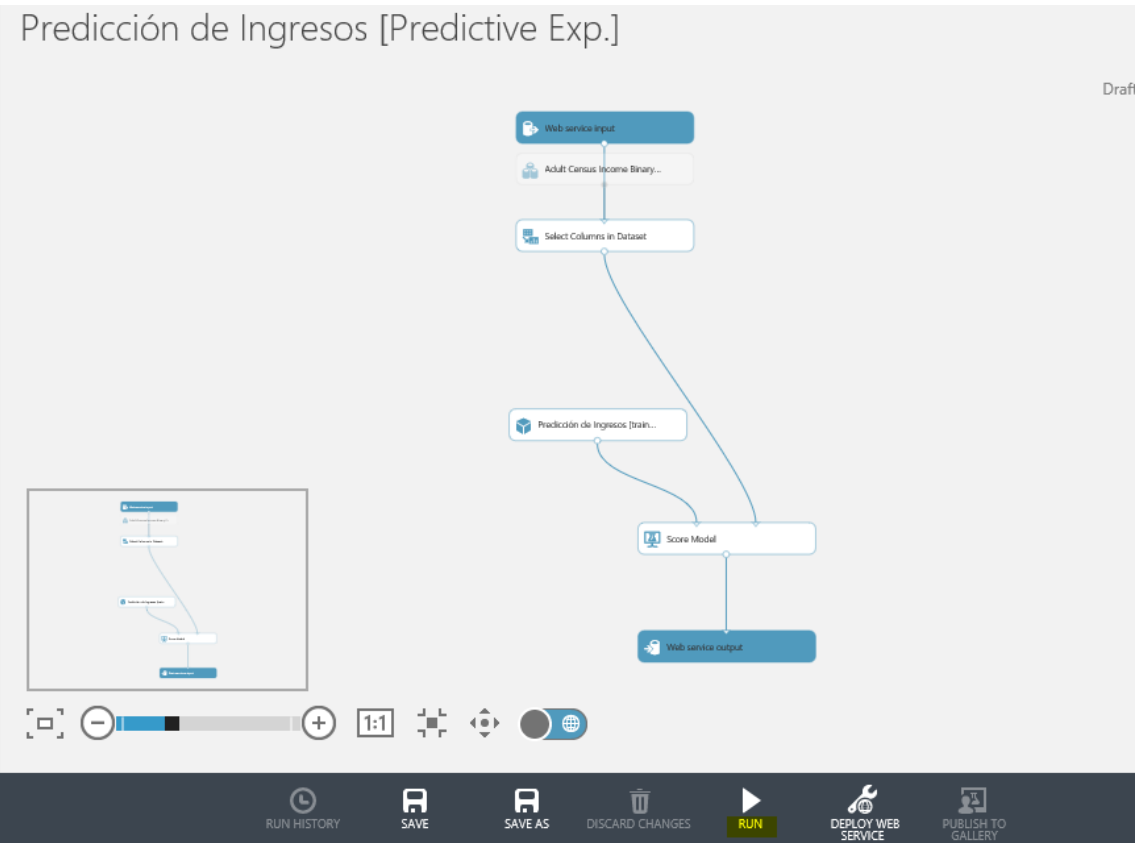
Sección #9: Publicar como Web Service el modelo de Machine Learning

9.1 Hacer Click sobre Set Up Web Service > Predictive Web Service



9.2 Esperar la publicación luego de la animación.

9.3 Ejecutar el **Web Service**, haciendo click en **Run**.



9.4 Hacer click en Deploy Web Service

predicción de ingresos [predictive exp.]

DASHBOARD CONFIGURATION

General

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

z7Iz7ZSy2ywFC3LFTlhk7QKUITJcw3D3GRKg510iv8Kxol6X/J/KR5XtNGx0WcxpmFJTjyJ6XlleXr8sTAZQ==

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	Test	Excel 2013 or later Excel 2010 or earlier workbook	6/2/2016 5:44:33 PM
BATCH EXECUTION		Excel 2013 or later workbook	6/2/2016 5:44:33 PM

Sección #10: Probar un caso real del Web Service

10.1 Hacer click en Test.

Test Predicción de Ingresos [Predictive Exp.] Service

Enter data to predict

AGE
p

WORKCLASS

FNLWGT
0

EDUCATION

EDUCATION-NUM
0

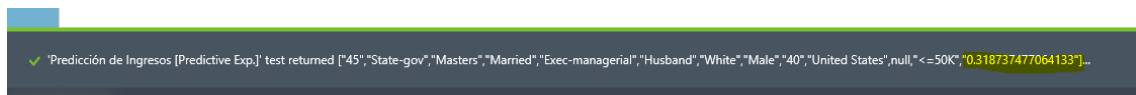
10.2 Ingresar datos de prueba para todos los campos.

10.3 Sugerencia de algunos datos de prueba:

- Age: 45
- Workclass: State-gov
- Education: Masters
- Marital Status: Married
- Relationship: Husband
- Race: White
- Sex: Male
- Hours per week: 40
- Native Country: United States

No todos los campos son obligatorios (incluyendo, por supuesto, *income*).

10.4 El último campo numérico indicará la probabilidad de ocurrencia para ese escenario.



10.5 La probabilidad de que un hombre blanco de 45 años con un Masters, casado, nacido en USA tenga ingresos menores a USD \$50.000 anuales es igual al 31%.

Evaluación

1. Un programa aprende si P función de T, mejora con E. ¿Cuál es la definición de P?
 - Probabilidad
 - Posibilidad
 - Performance
 - Programación
2. Un programa aprende si P función de T, mejora con E. ¿Cuál es la definición de E?
 - Energía
 - Evolución
 - Experiencia
 - Experimentos
3. ¿Cuál es otra forma de llamar a Machine Learning?
 - Análisis predictivo (*predictive analysis*)
 - *Data warehousing*
 - *Data mining*
 - *Random sampling*
4. Verdadero o falso: Entrenamiento Supervisado sirve para predecir respuestas de casos que ya conozco.
5. Elegir un algoritmo de Machine Learning para cada uno de estos escenarios:
 - *Customer Churn*
 - Recomendación de música
 - Predecir si un pasajero sobrevive al Titanic o no

Competencia

Desarrollar un modelo de Machine Learning que permita predecir la probabilidad de que un vuelo se demore (al menos un minuto) o no. Utilizar el *data set* disponible en AzureML Studio.

El ganador será el que mejor precisión tenga.