Microsoft

Two Class Decision Forest with Azure ML

Contents

Table of Contents

Goals and Requirements

Estimated time to complete lab is 20-30 minutes.

Goals

1. Develop a model to predict whether income exceeds base limit from censor data.

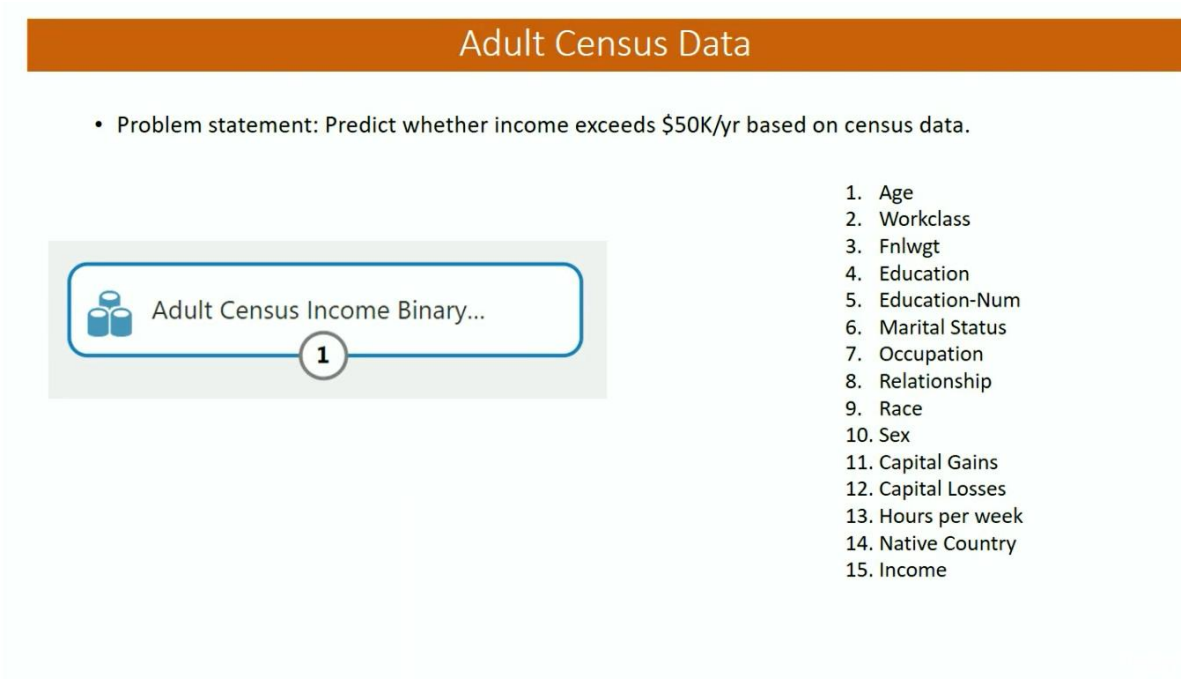
2. Develop the model using Decision Forest

Requirements

1. Access to an Azure Machine Learning and Provided dataset

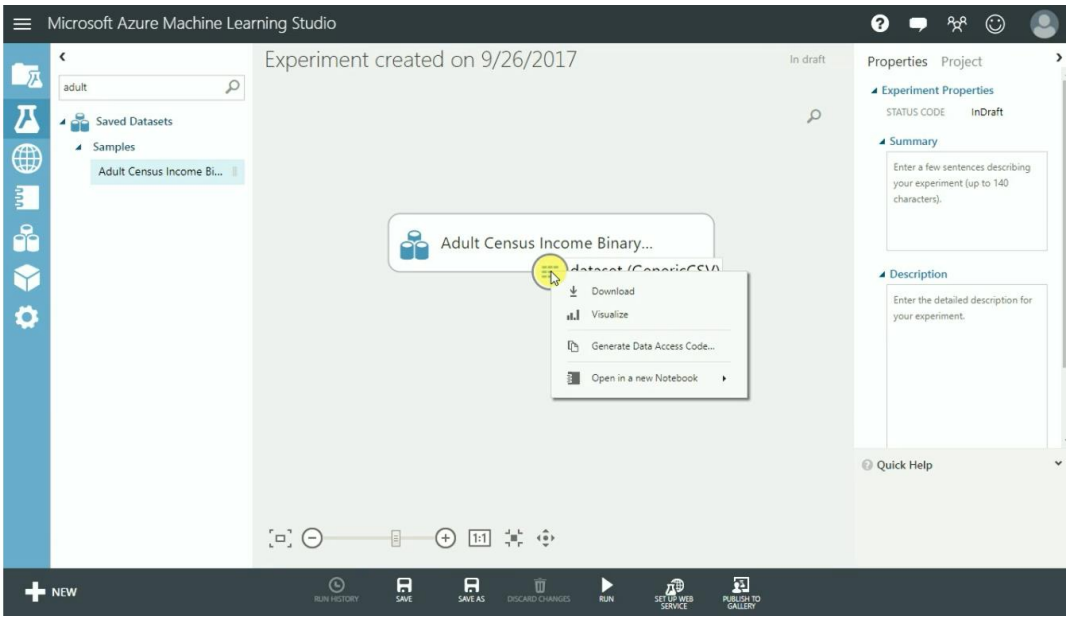
Two Class Decision Forest

Problem Statement: Predict whether income exceeds base amount on census data

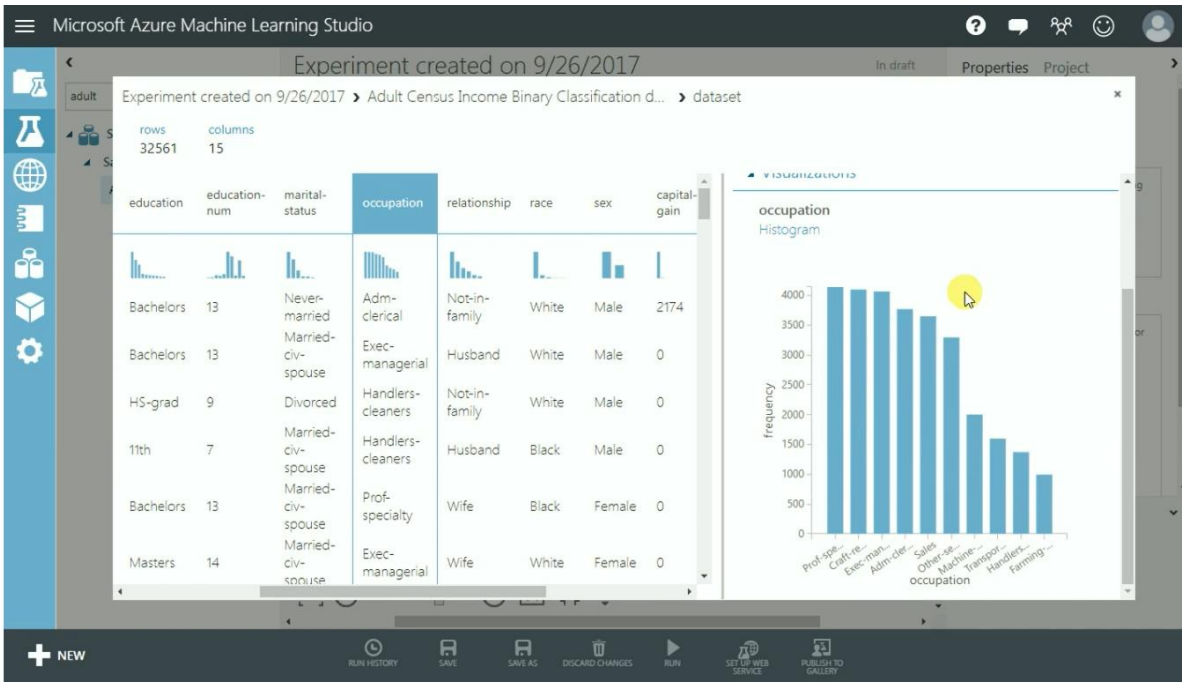


Dataset

Add the dataset and visualize

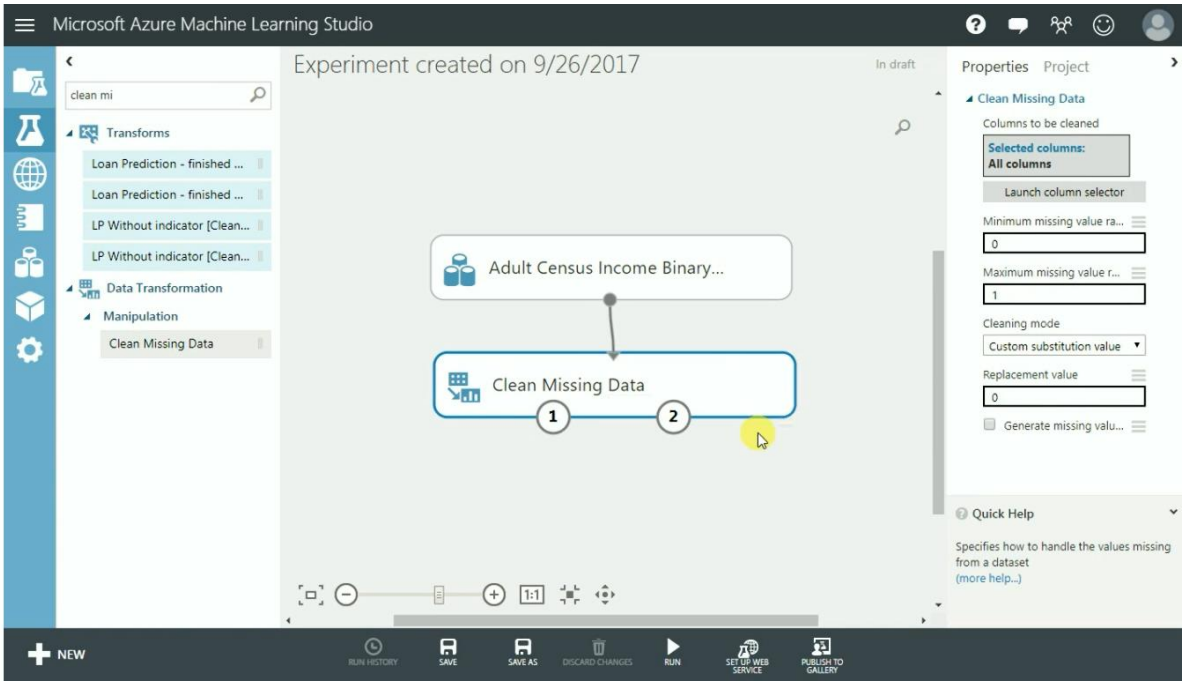


Go through the dataset columns one by one

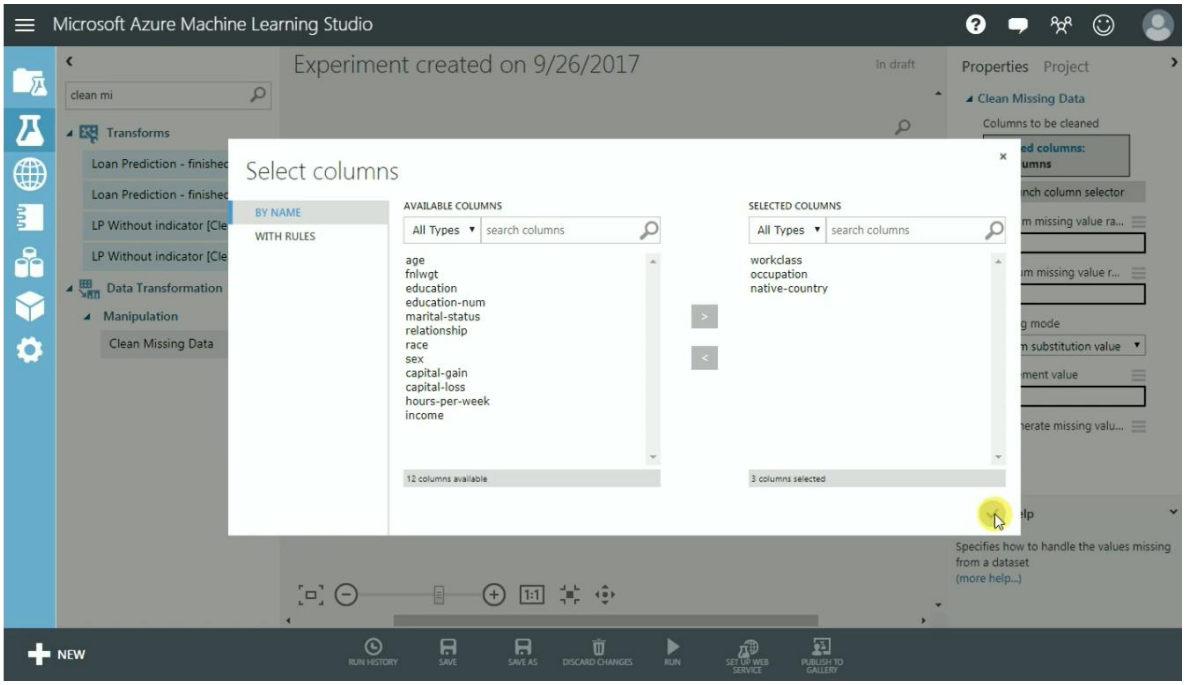


Clean Missing Data

Add clean missing data to canvas and connect the node

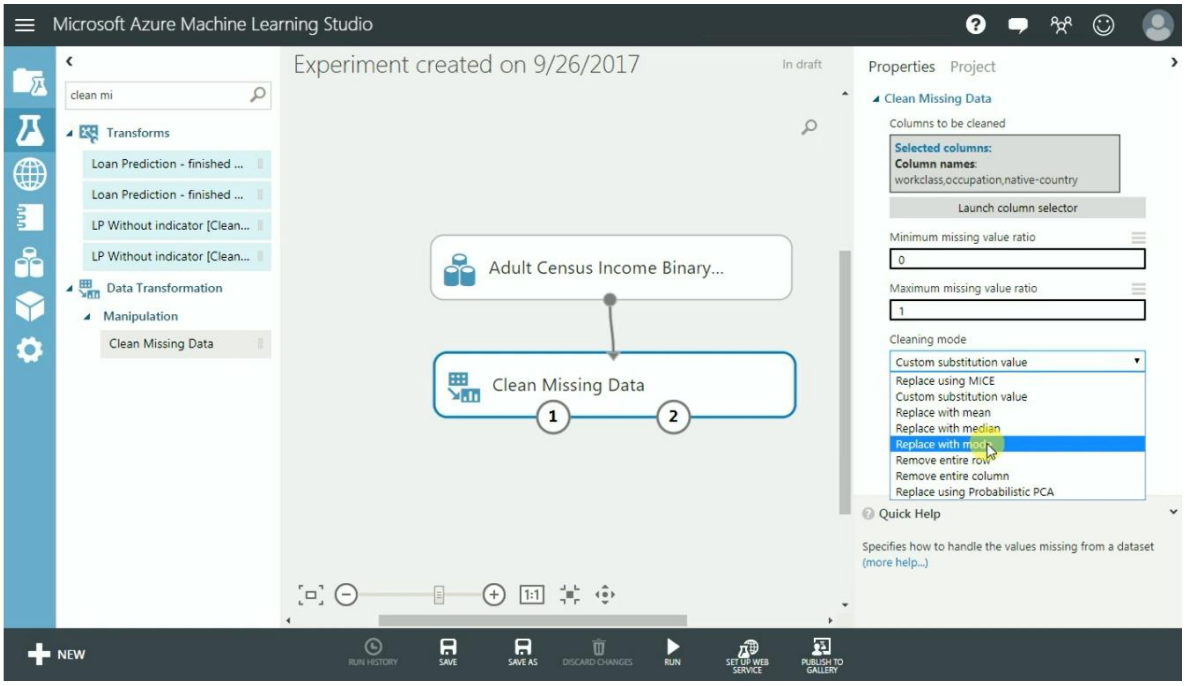


Click Launch column selector and select missing value columns and click ok

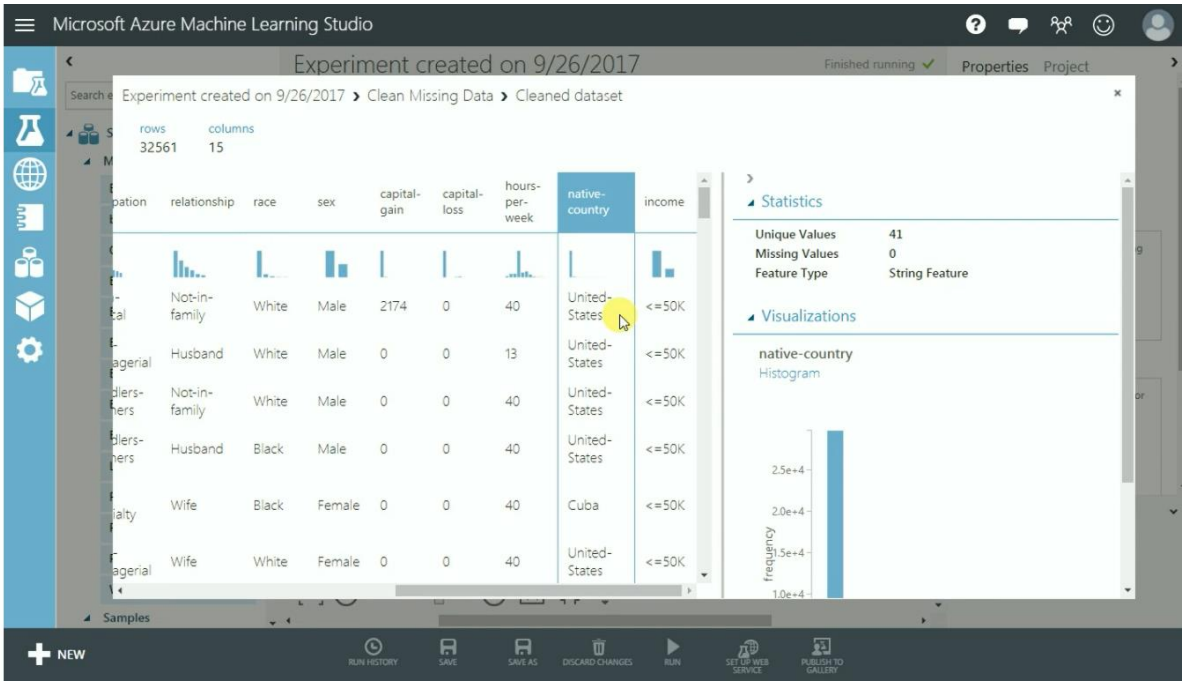


Replace Missing Data with Mode

Select parameters cleaning mode as Replace with mode

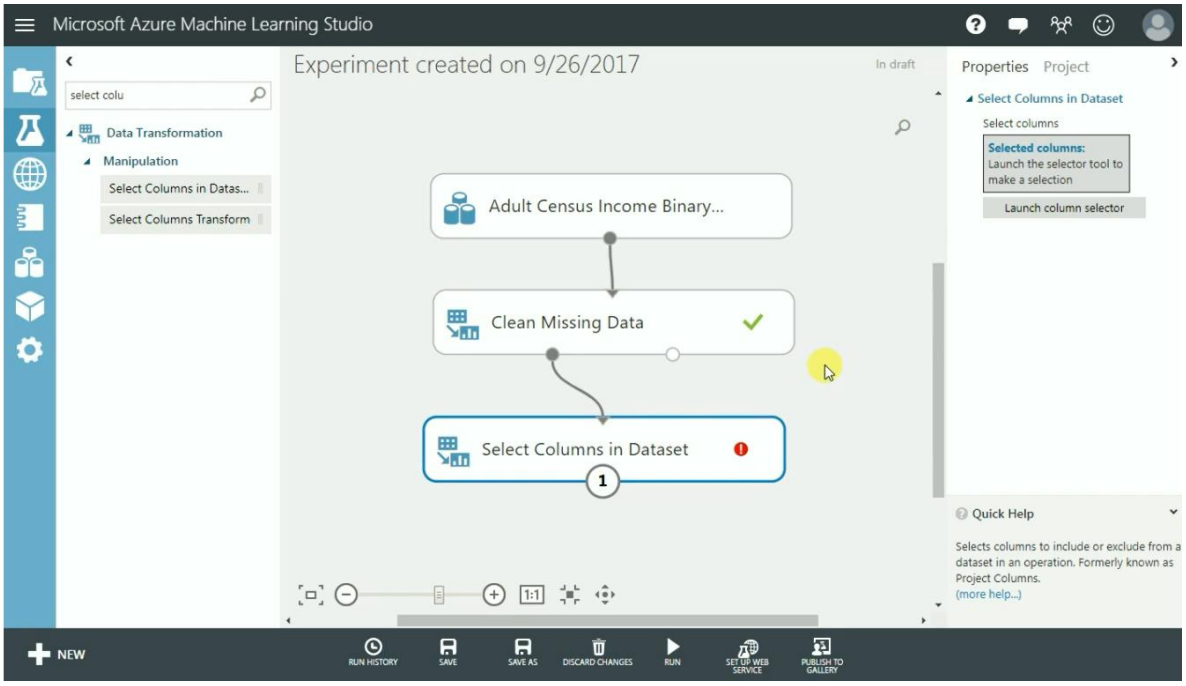


Run and visualize for result



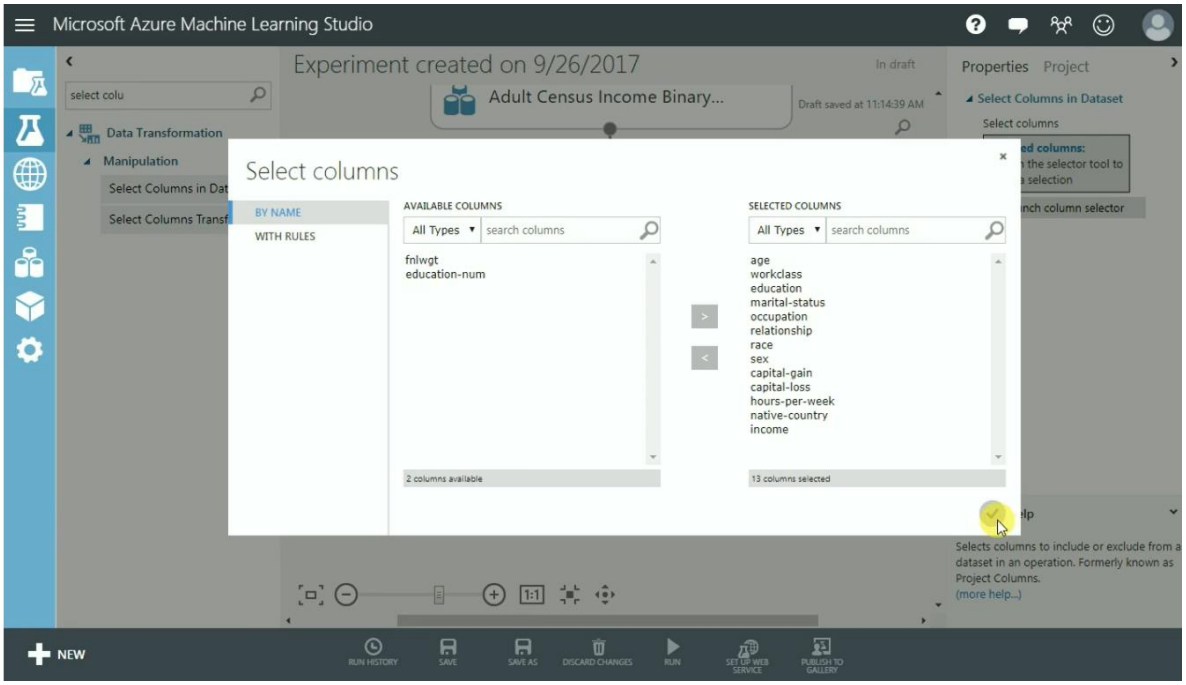
Select Columns in Dataset

Add Select columns in dataset and connect nodes



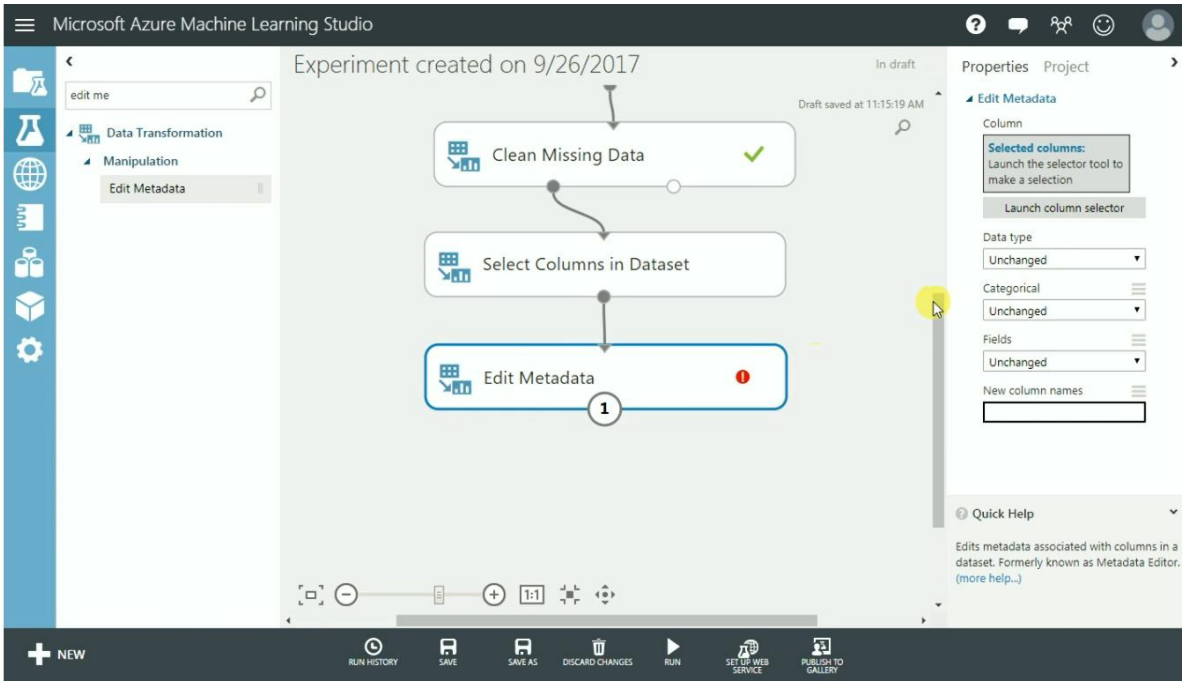
Launch columns selector and select all columns except fnlwgt and education-num

And click ok

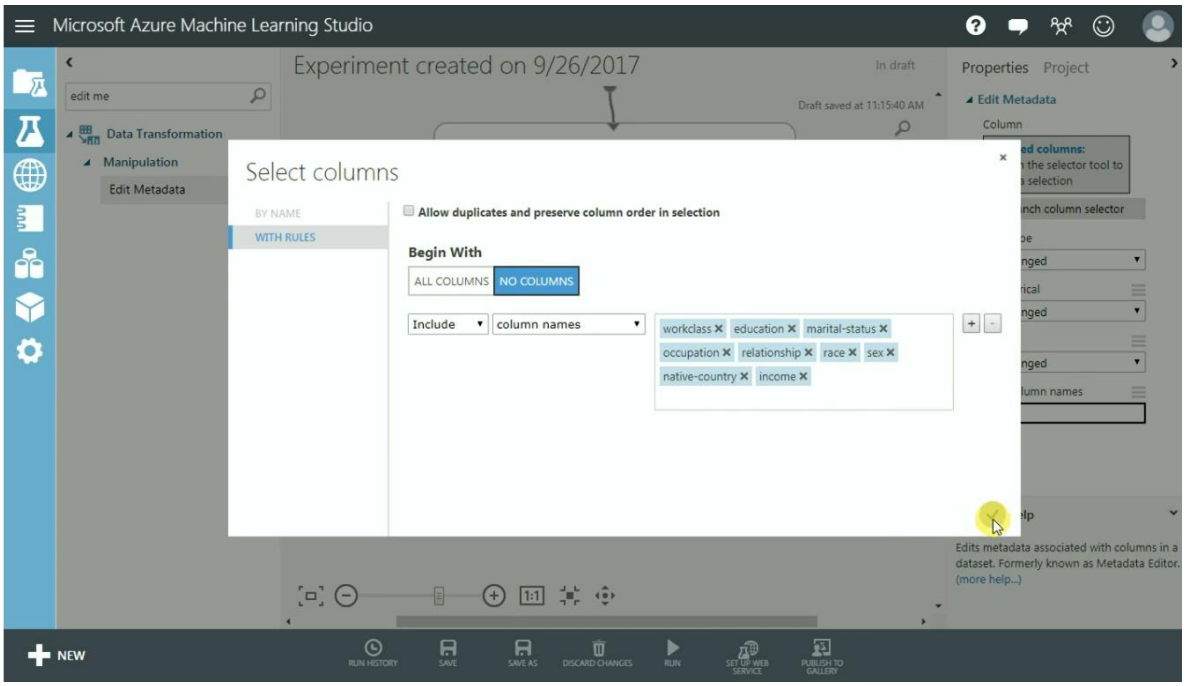


Edit Meta Data

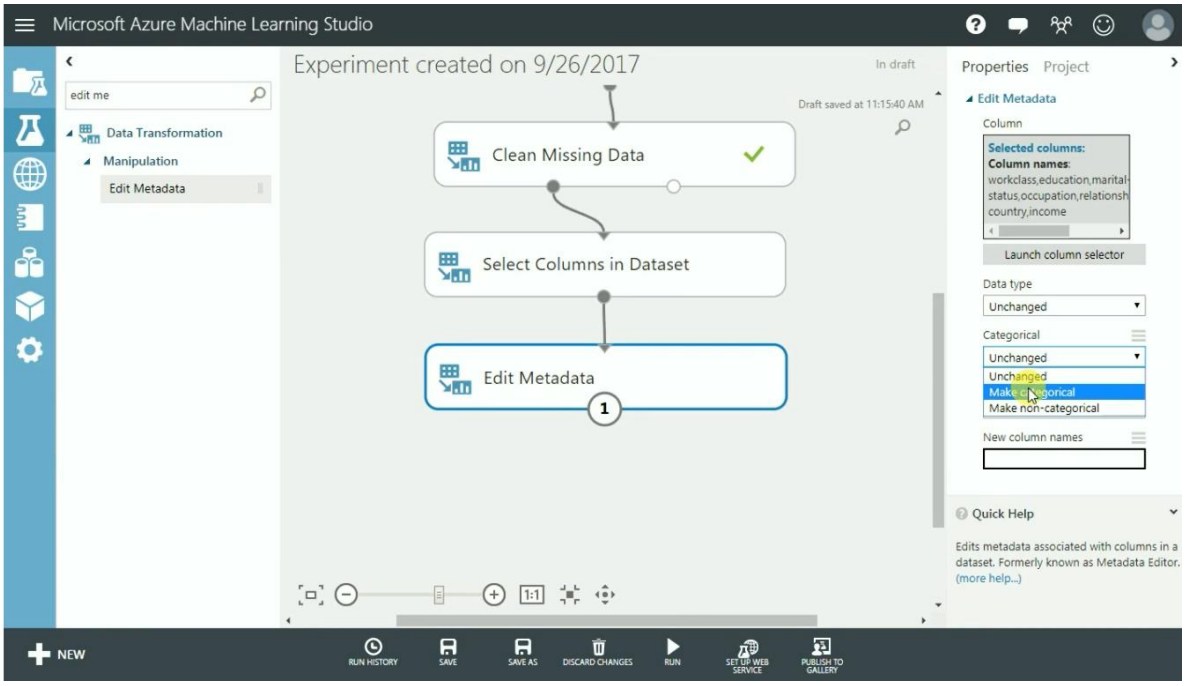
Add Edit metadata in canvas and make connections



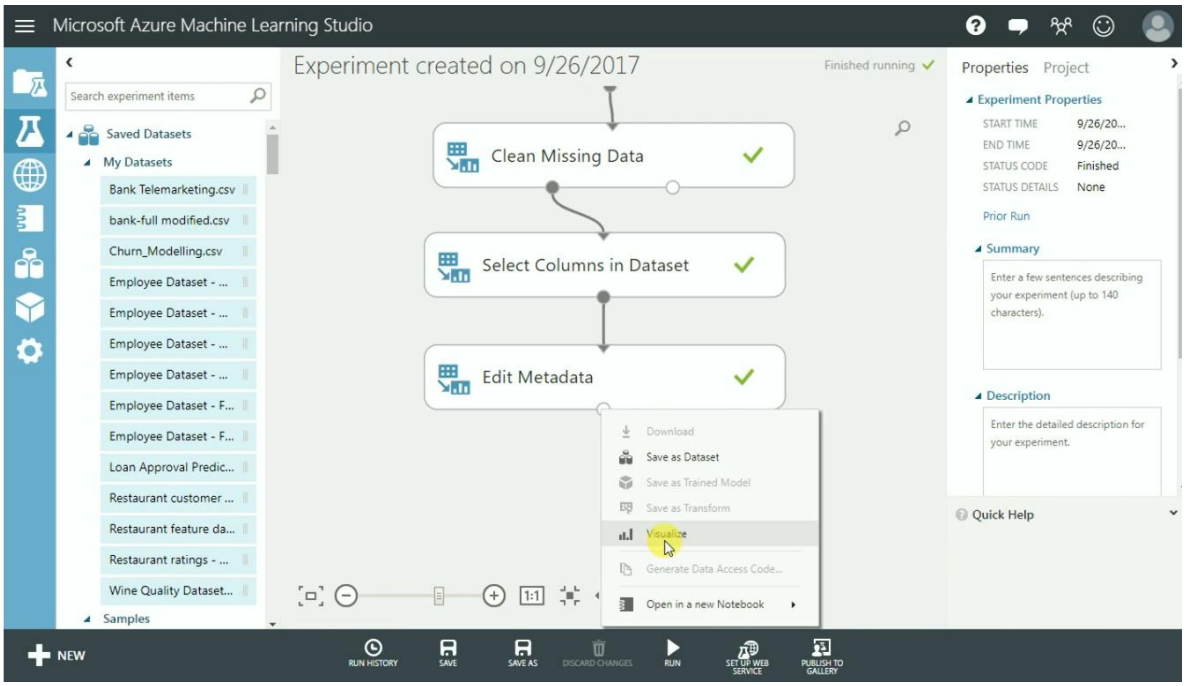
Launch column selector and add columns as shown



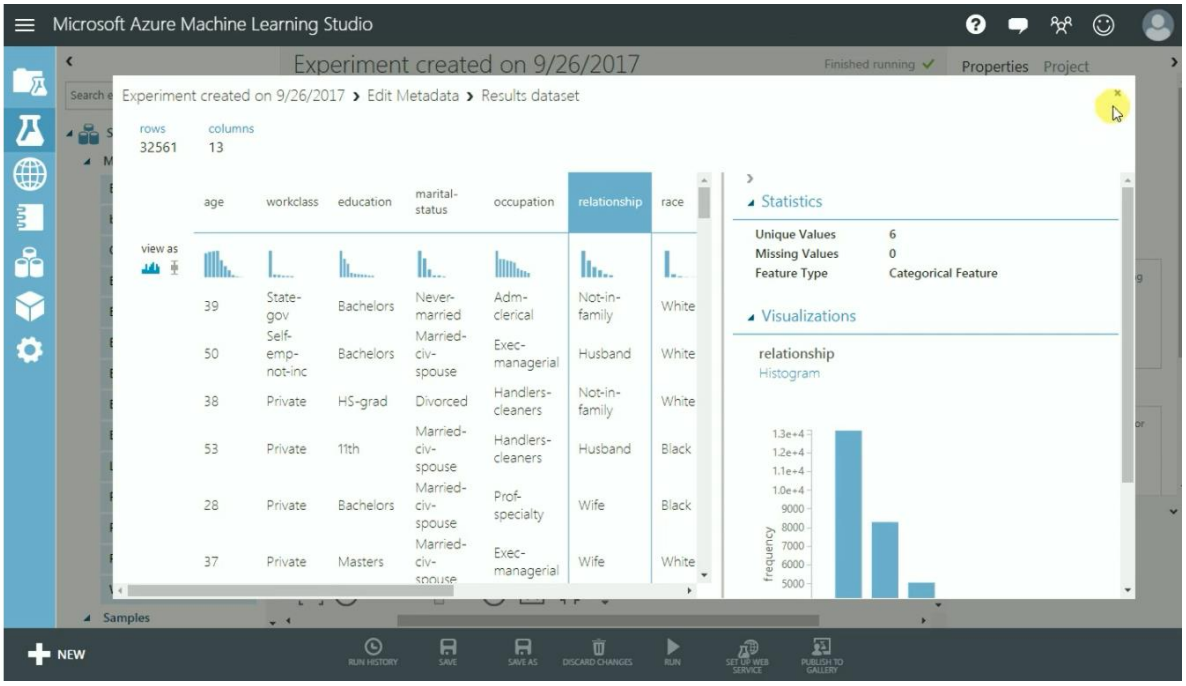
Change in parameter



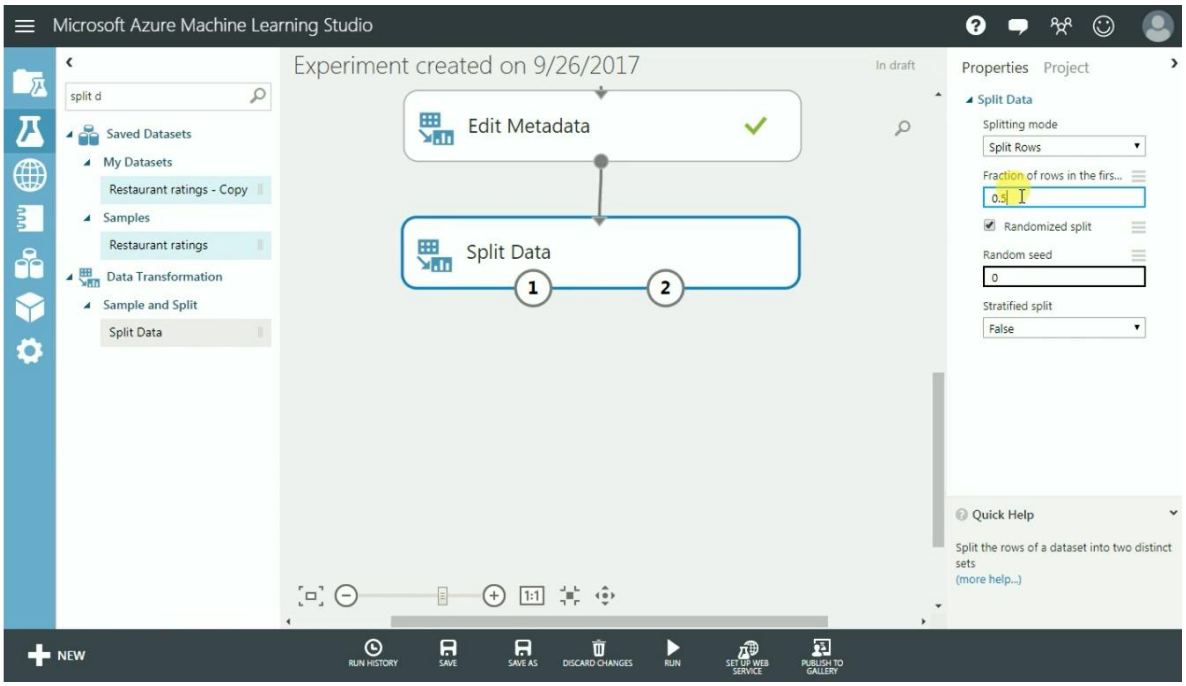
Run and visualize the output



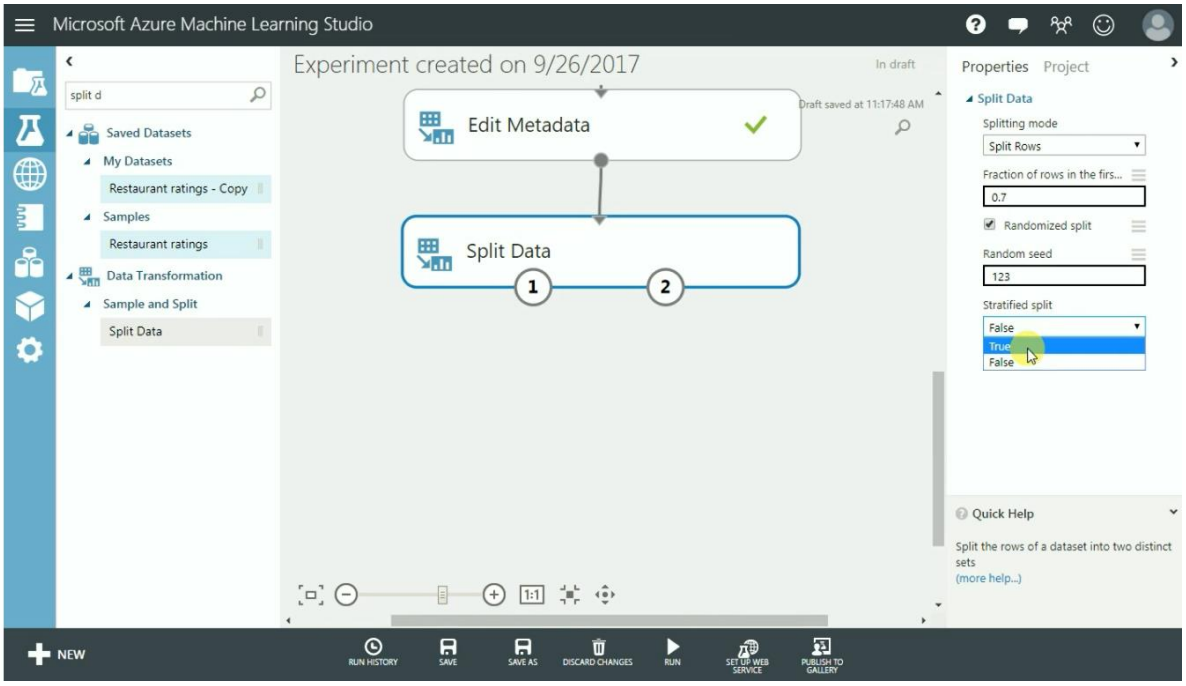
Output with categorical feature



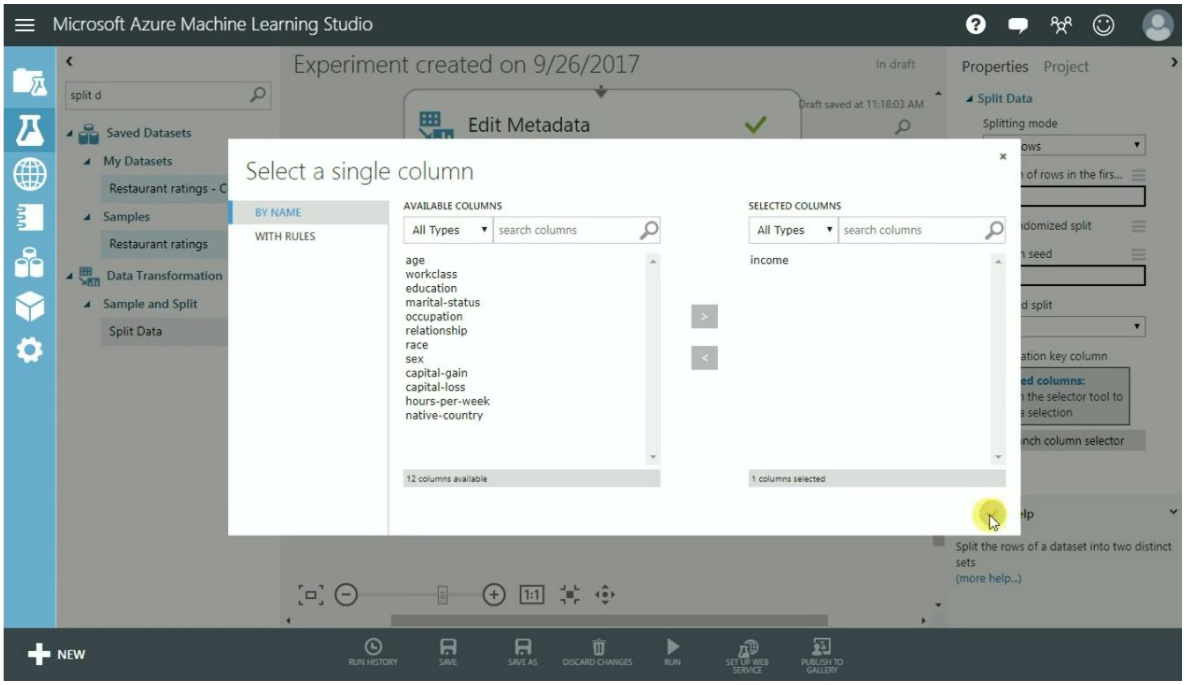
Add split data in canvas and make connections



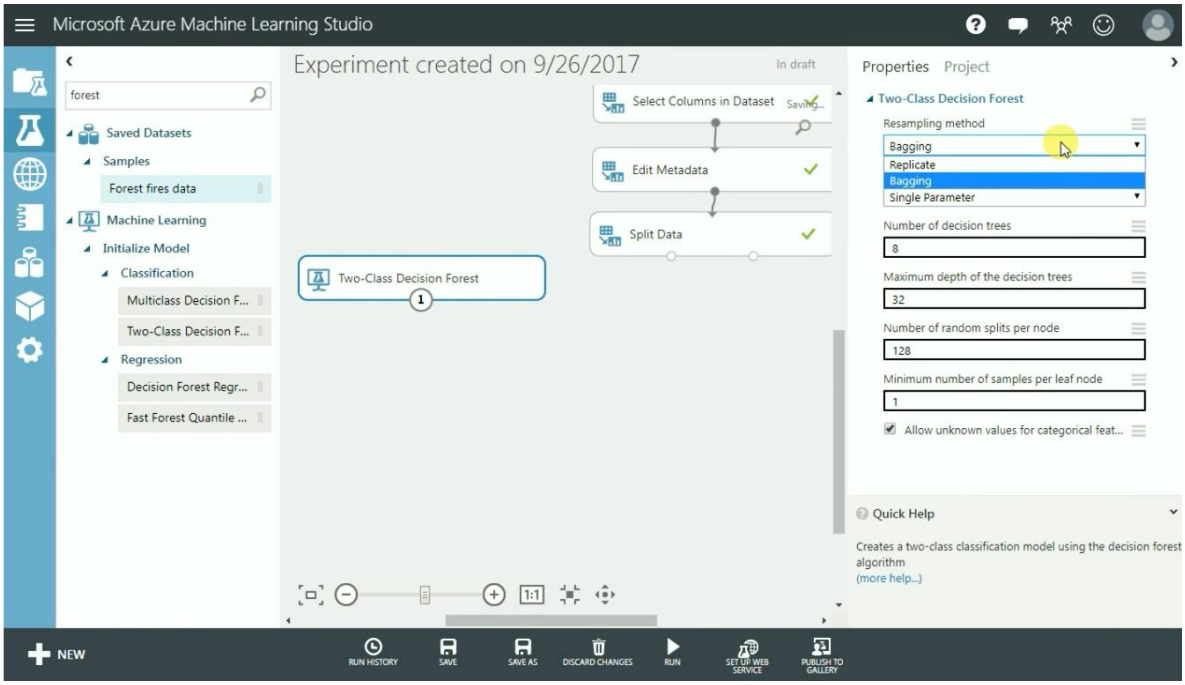
Change parameters fractions, random seed and stratified split



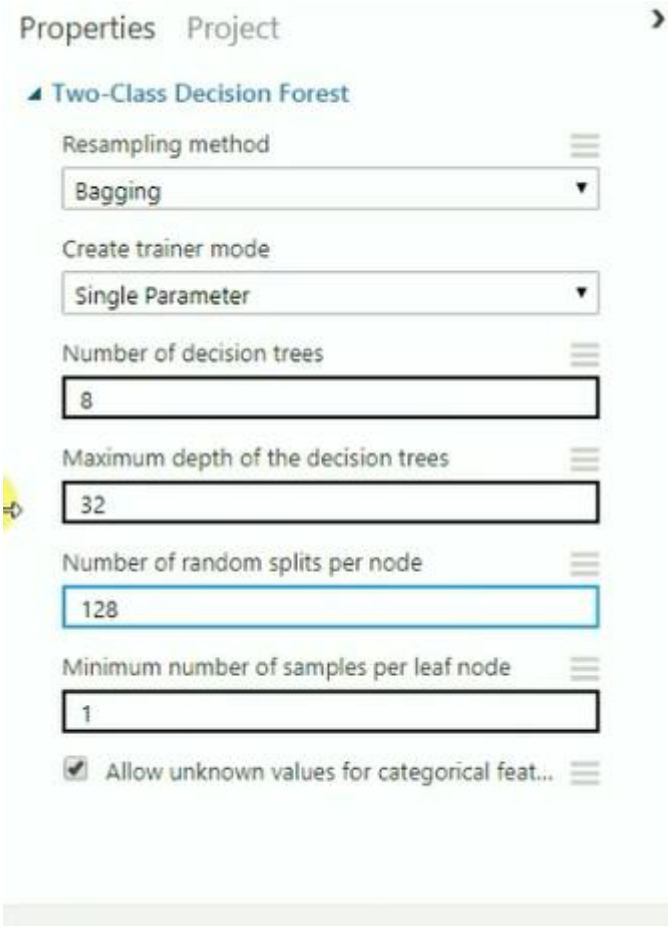
Launch column selector and select income column and click ok



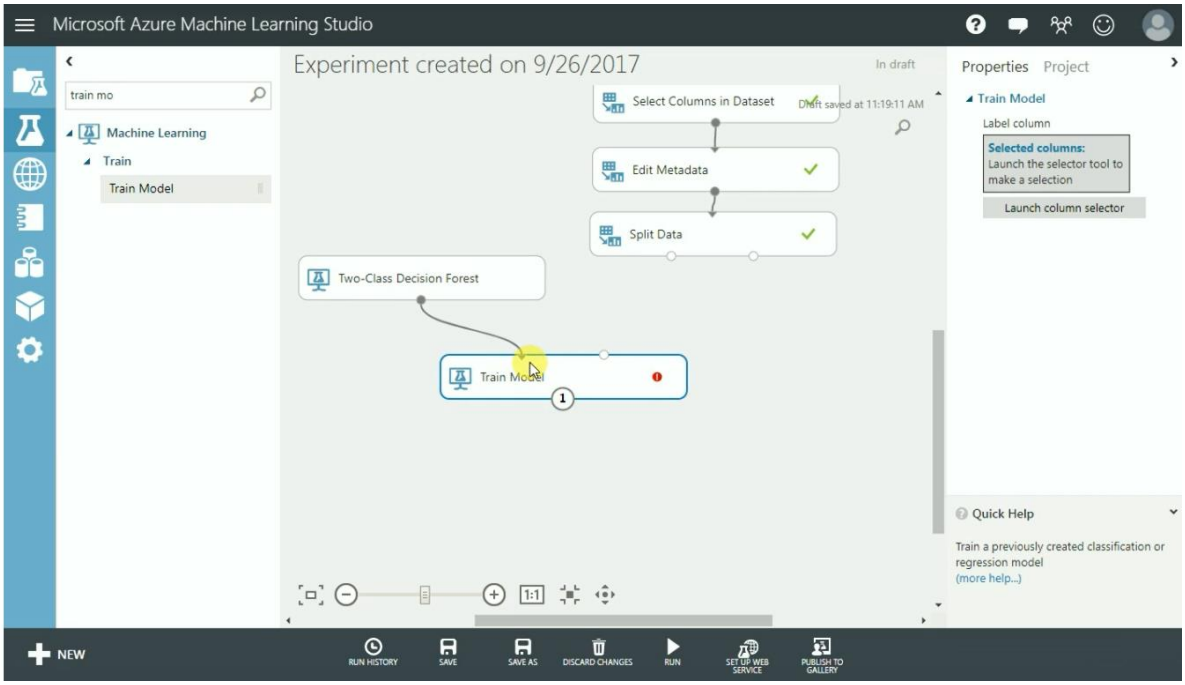
Add Two class decision forest in canvas and input parameters as required



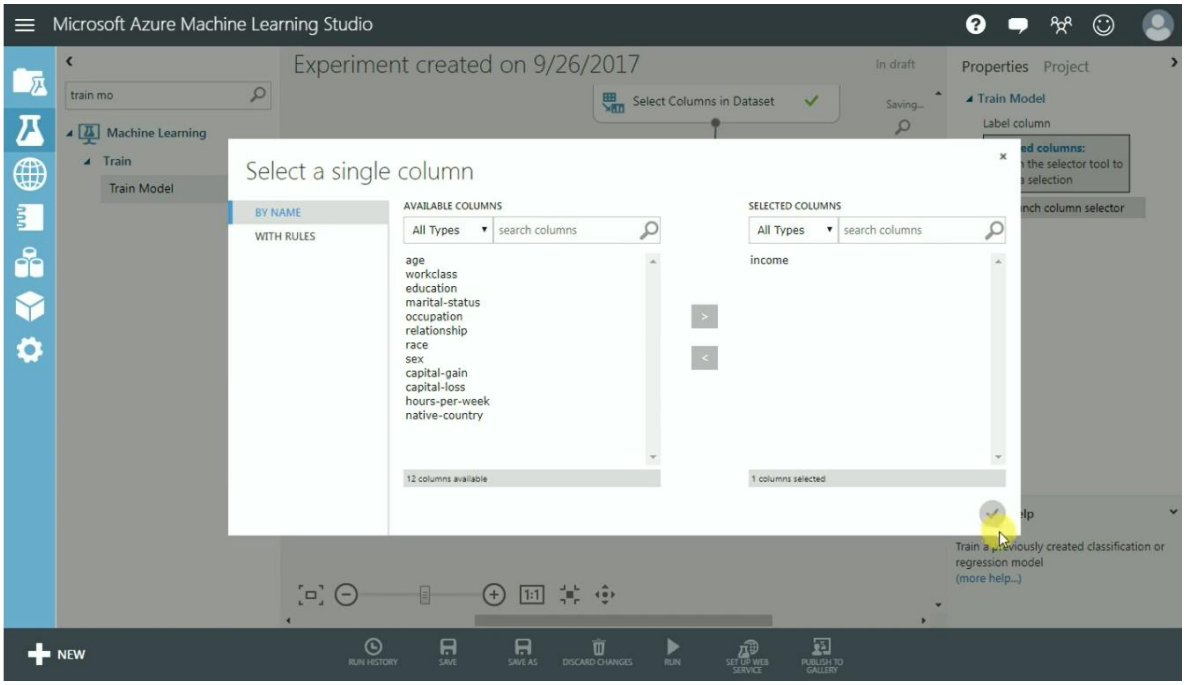
Parameters



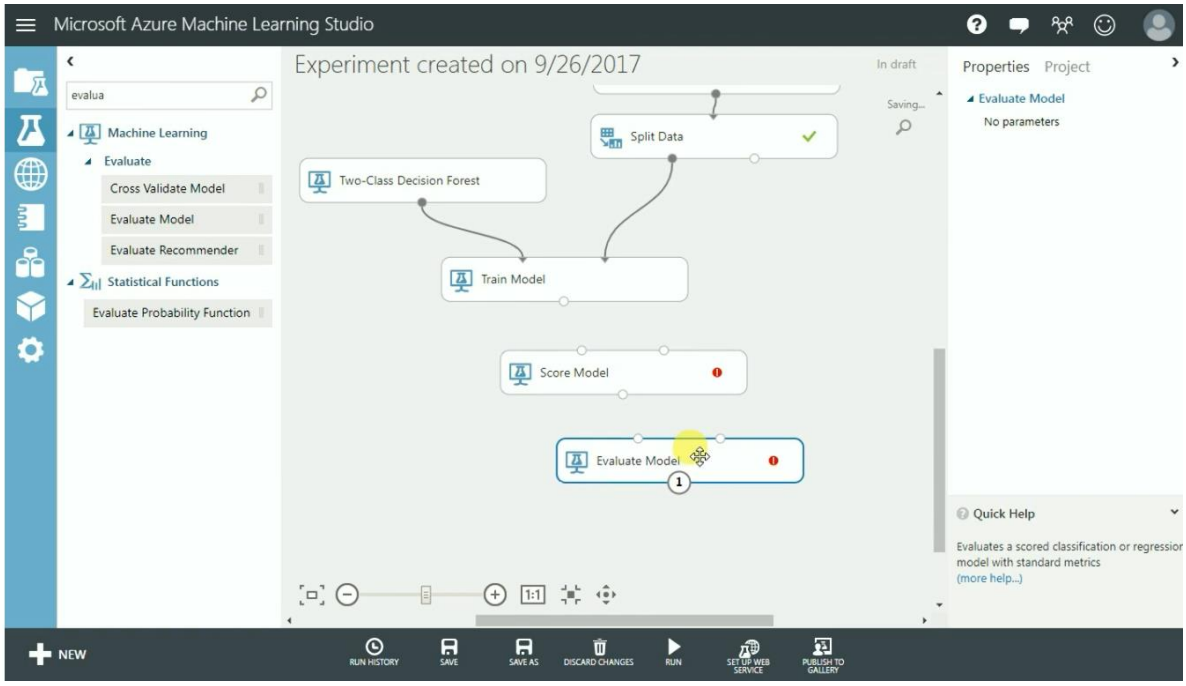
Add Train model and connect the nodes



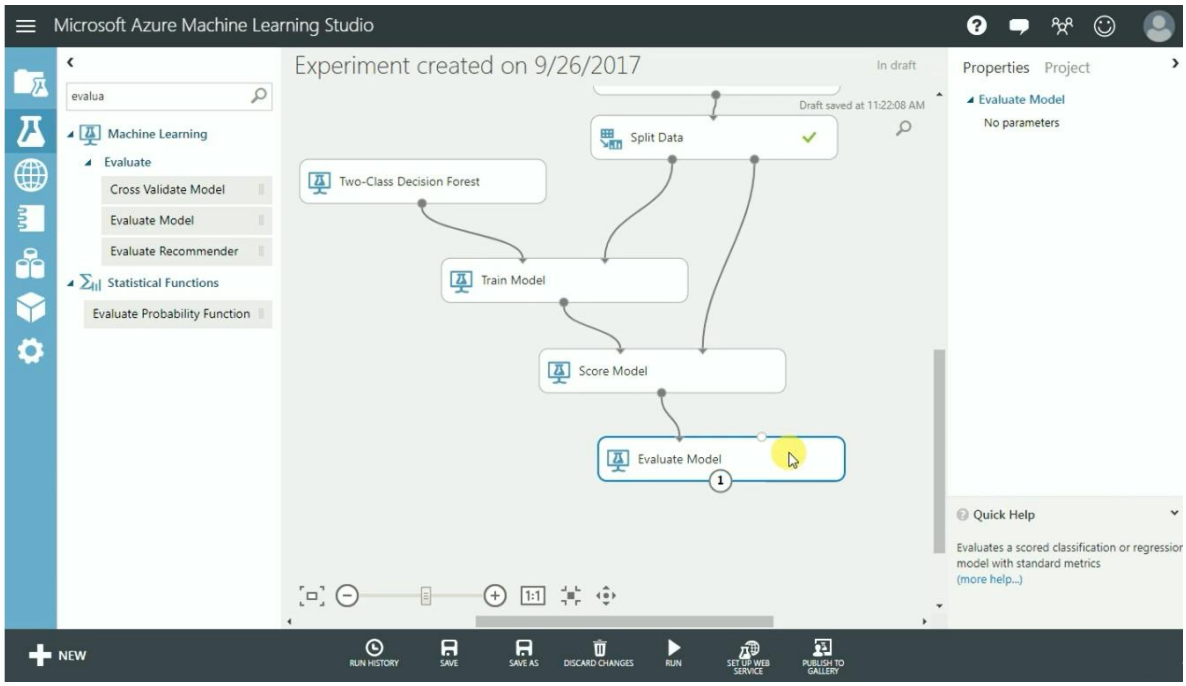
Launch column selector and select income then press ok



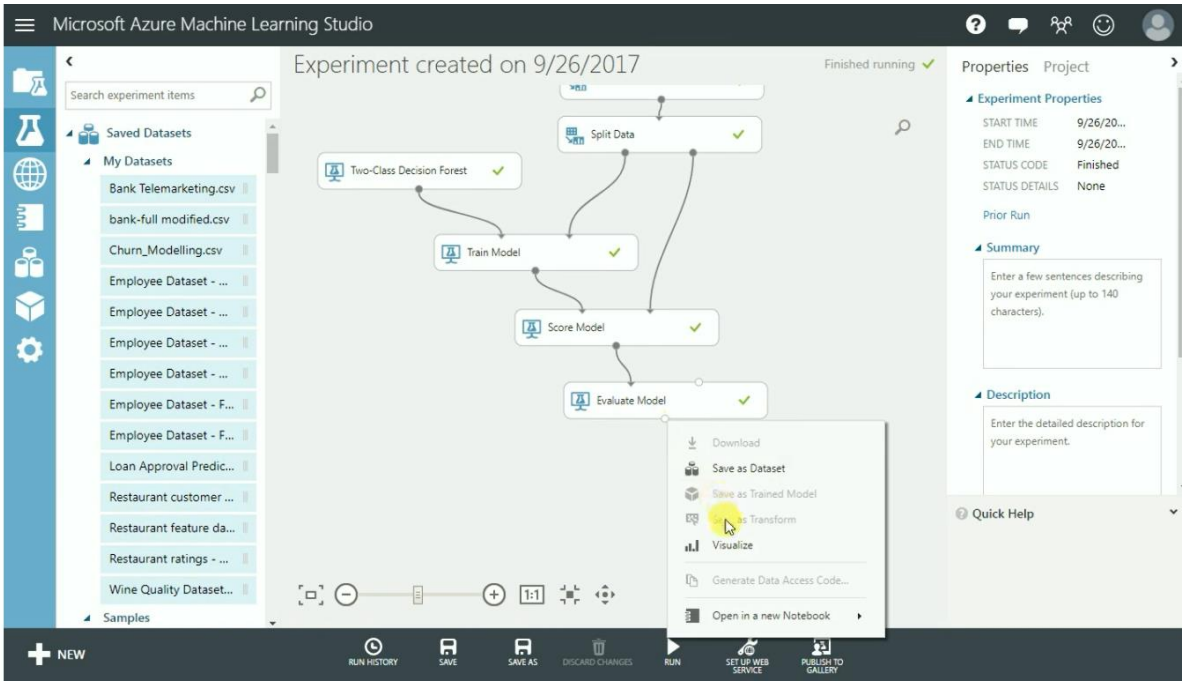
Add score model and evaluate model



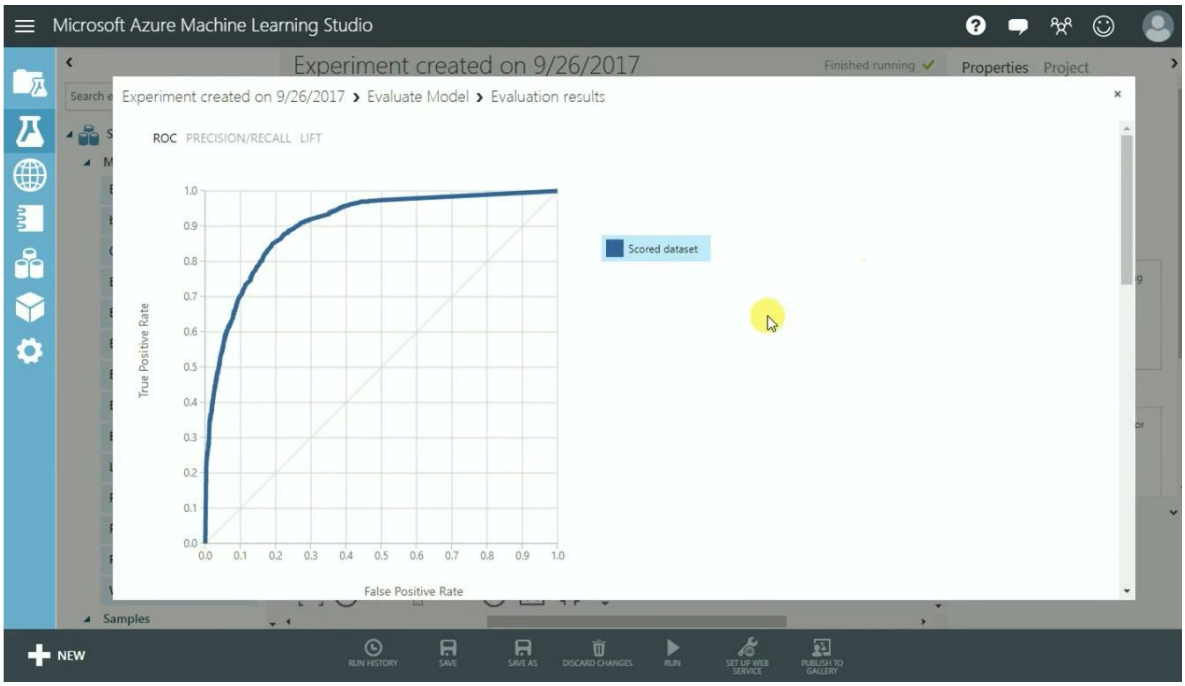
Connect all nodes as shown



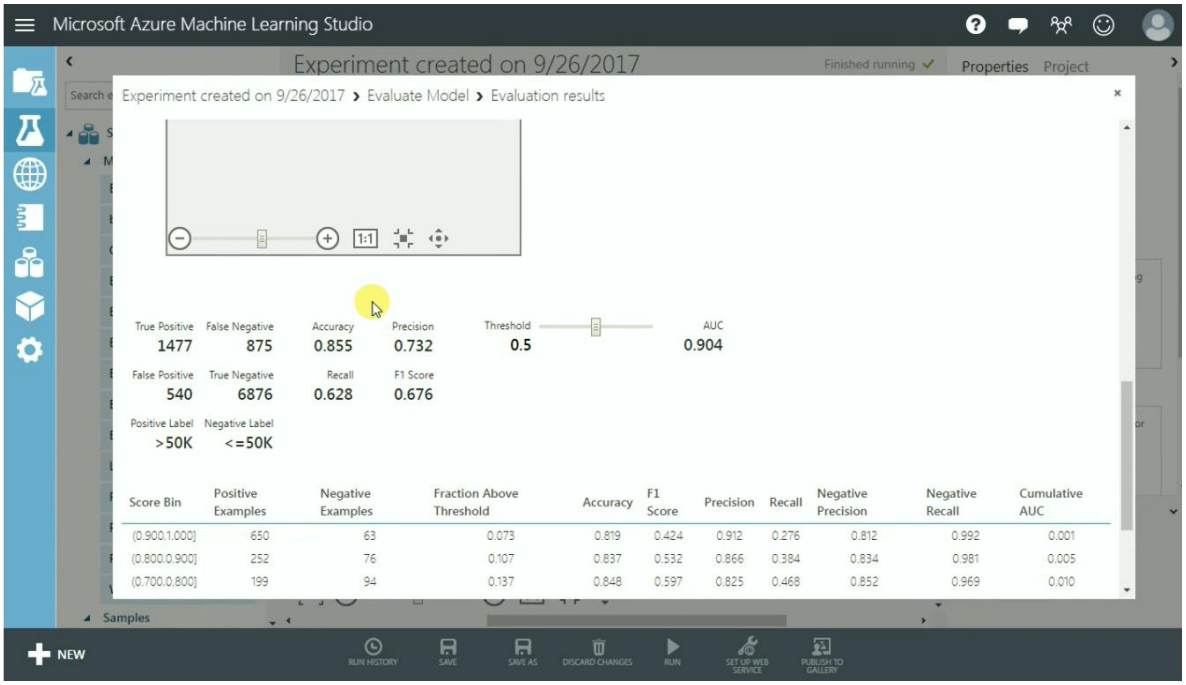
Run and visualize the module



Result obtained successful



Two class decision forest output with good accuracy


Microsoft Azure Machine Learning Studio
Experiment created on 9/26/2017
Finished running
Properties
Project
Experiment created on 9/26/2017 › Evaluate Model › Evaluation results
True Positive 1477
False Negative 875
Accuracy 0.855
Precision 0.732
Threshold 0.5
AUC 0.904
False Positive 540
True Negative 6876
Recall 0.628
F1 Score 0.676
Positive Label >50K
Negative Label <=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] | 650 | 63 | 0.073 | 0.819 | 0.424 | 0.912 | 0.276 | 0.812 | 0.992 | 0.001
(0.800,0.900] | 252 | 76 | 0.107 | 0.837 | 0.532 | 0.866 | 0.384 | 0.834 | 0.981 | 0.005
(0.700,0.800] | 199 | 94 | 0.137 | 0.848 | 0.597 | 0.825 | 0.468 | 0.852 | 0.969 | 0.010
Samples
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY