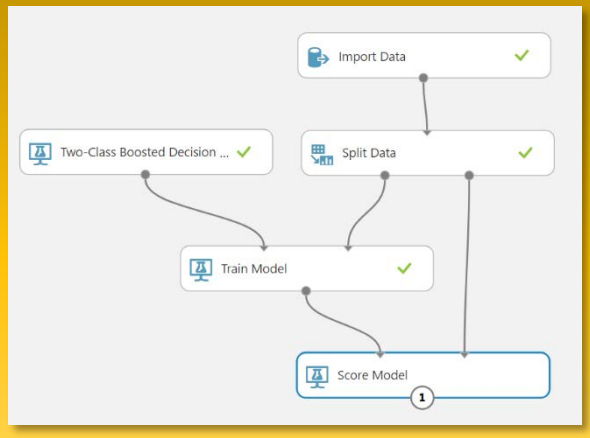
First Experimental

FIRST EXPERIMENTAL



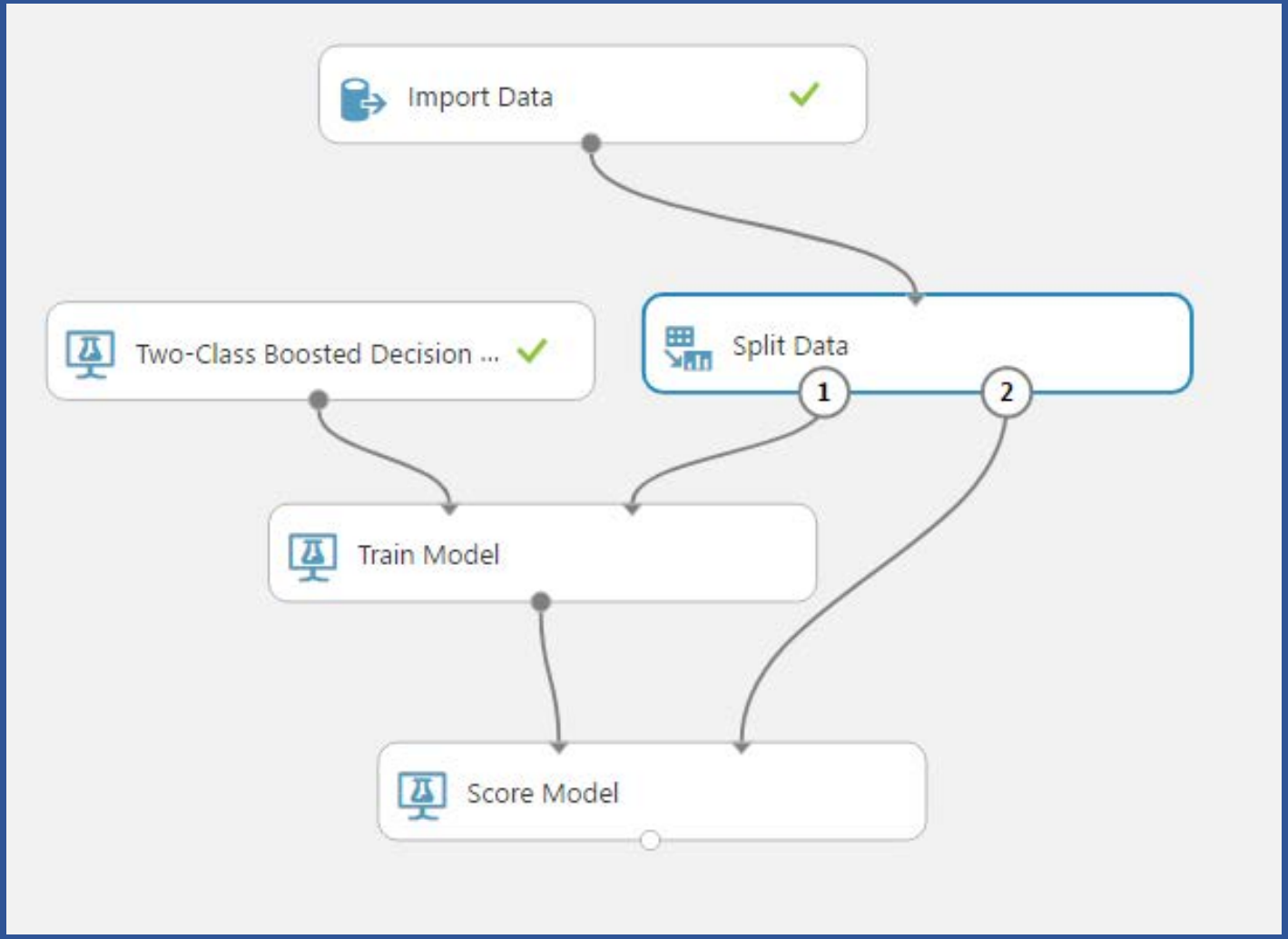
First Experimental

In this session

- Sing Up FREE Azure ML Studio Subscription
- Create Azure ML Studio workspace
- Train, Test, Evaluate for Two-Class Boosted Decision Tree
- Import census income dataset
- Create a new Azure Machine Learning experiment
- Train and evaluate a prediction model
- Type of datasets

First Experimental

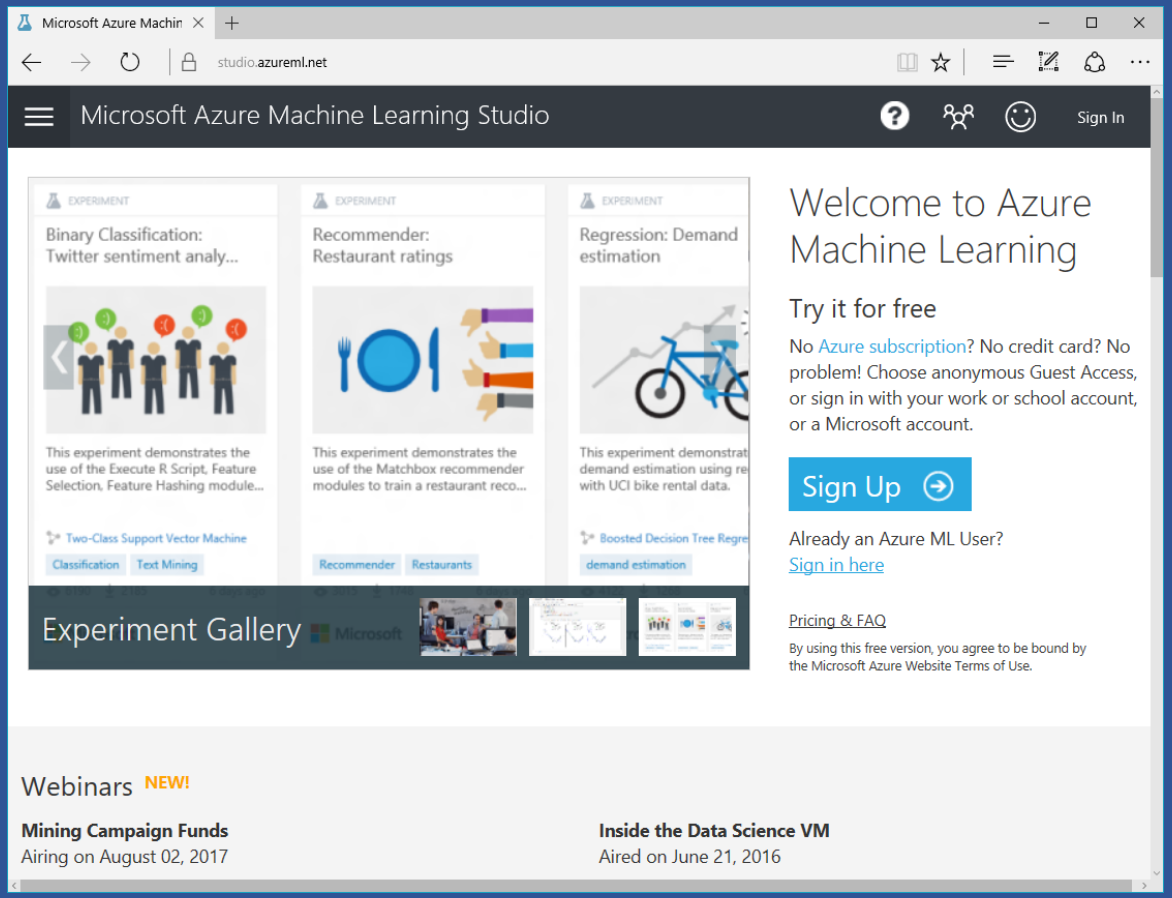
First experiment model



First Experimental

Sing Up FREE Azure ML Studio Subscription

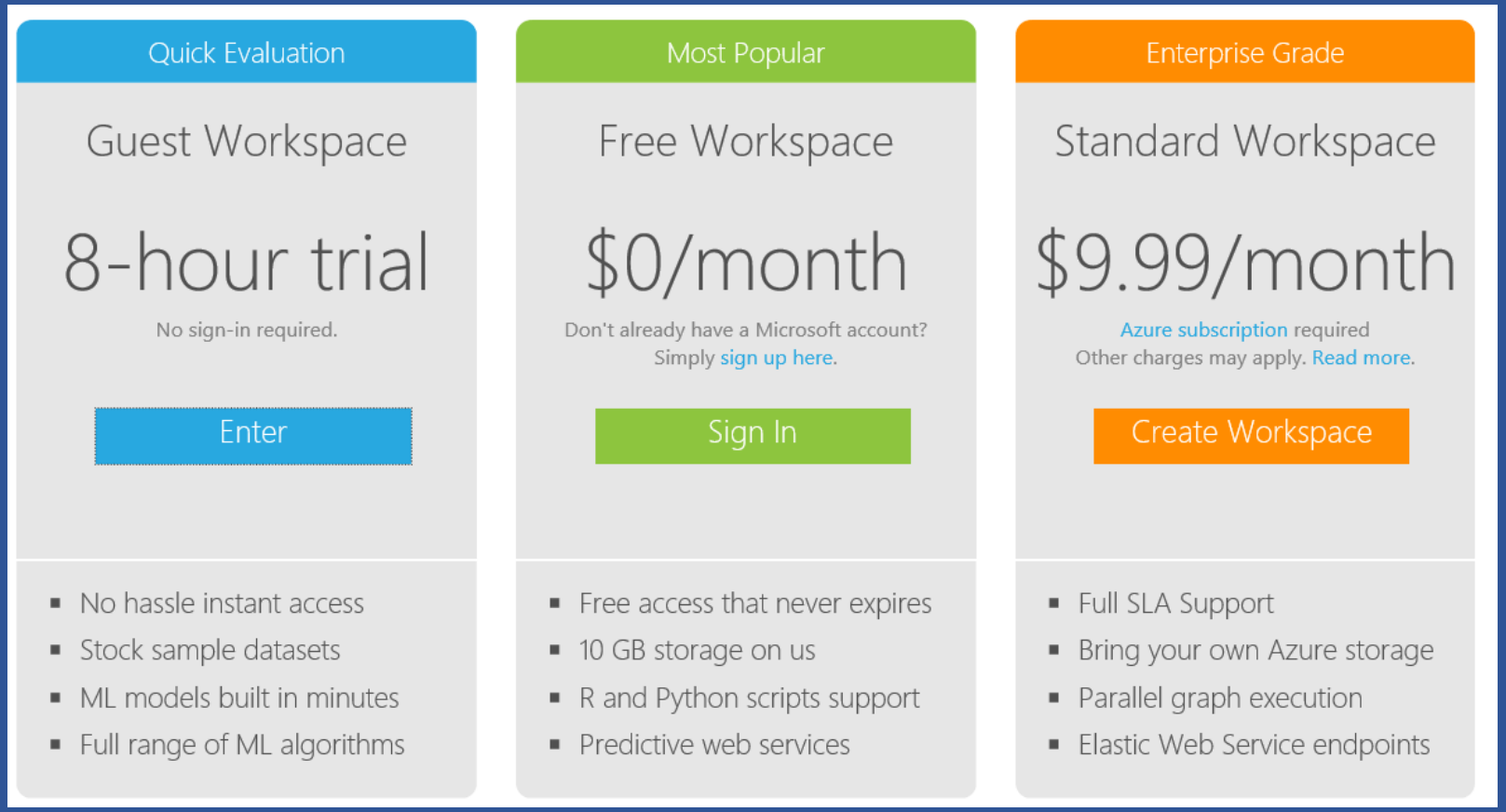
https://studio.azureml.net/



First Experimental

Sing Up FREE Azure ML Studio Subscription

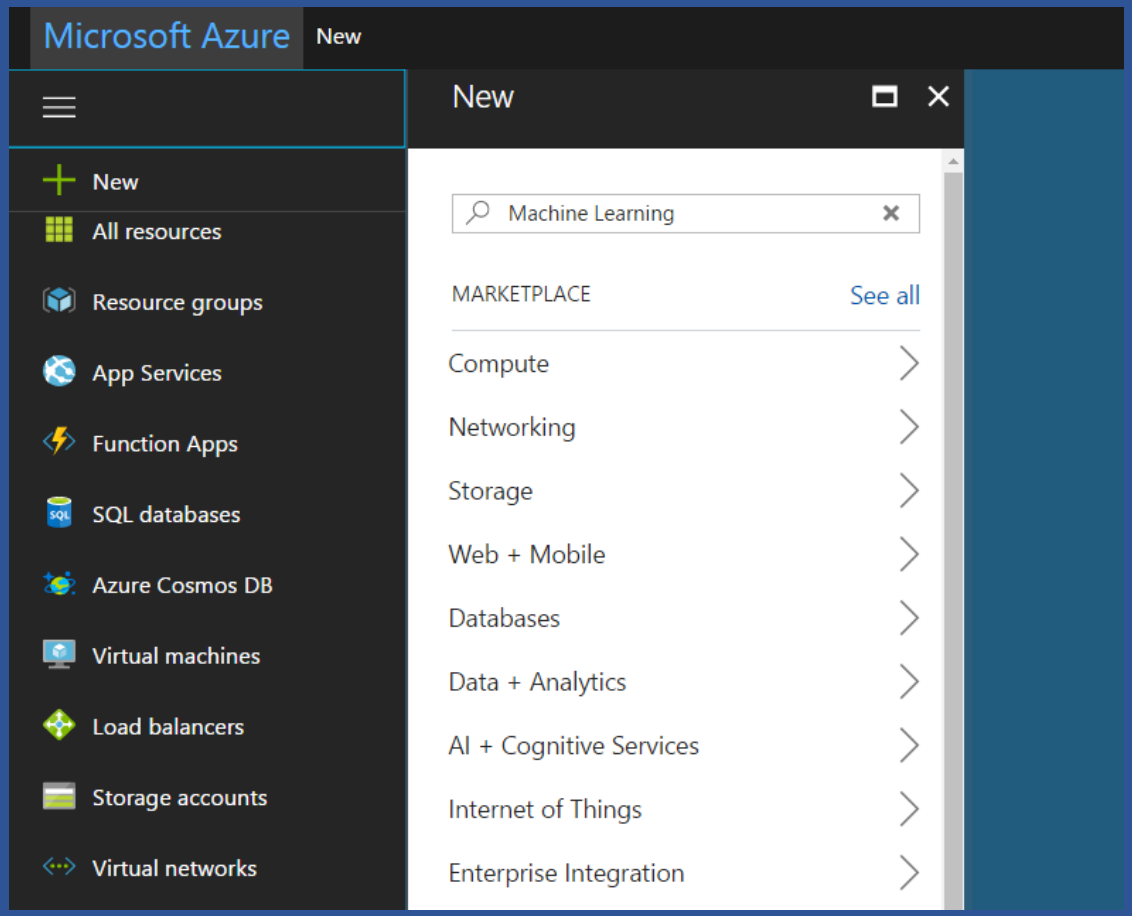
Free Workspace -> sign up here



First Experimental

Create Azure ML Studio workspace

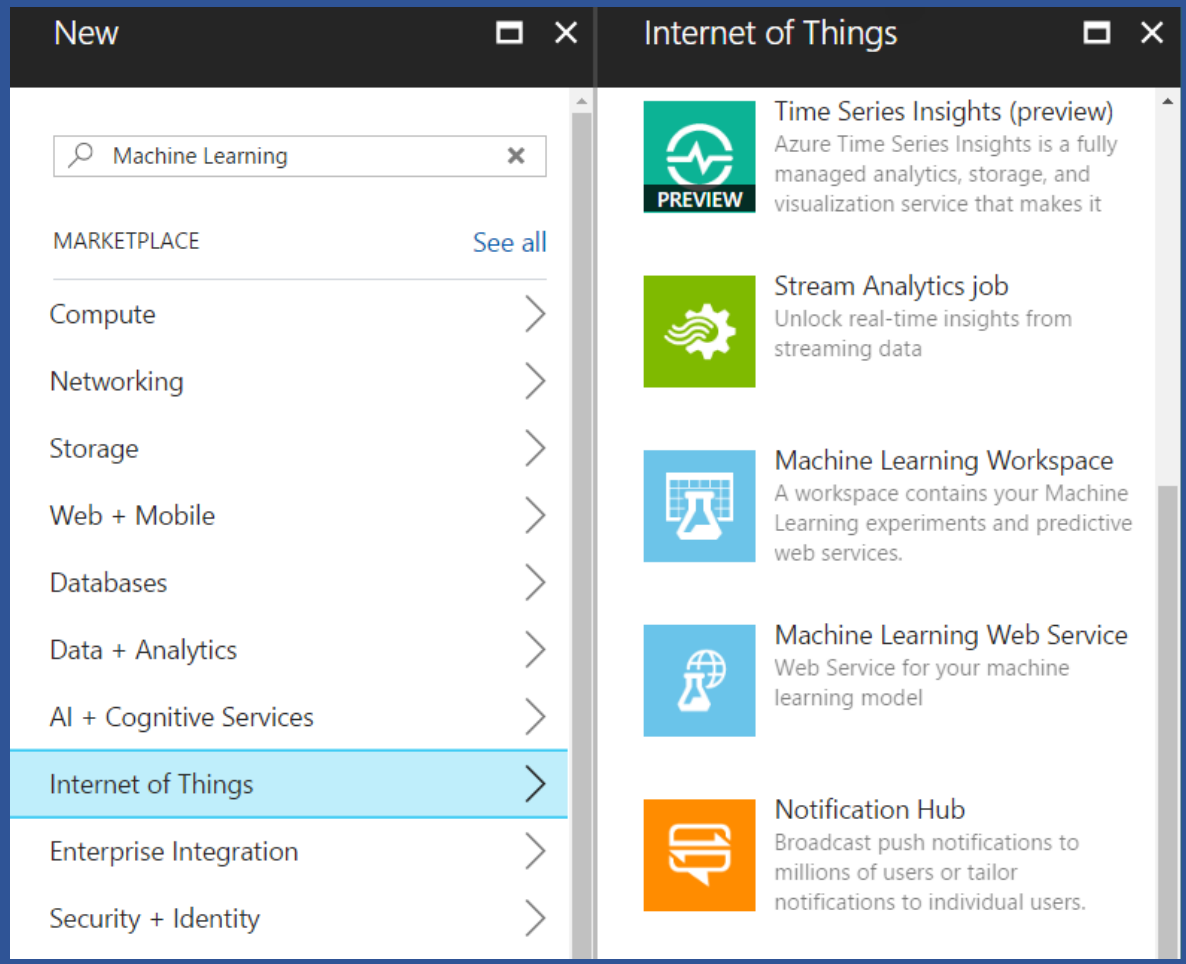
1. Go to the Azure portal https://portal.azure.com
2. Click +New



First Experimental

Create Azure ML Studio workspace

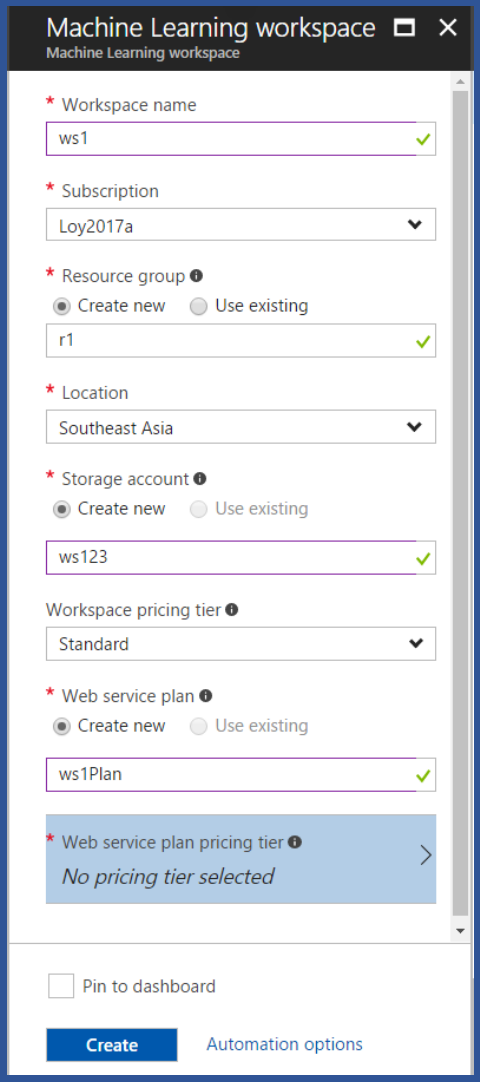
3. Select Internet of Things, click Machine Learning Workspace, then click Create



First Experimental

Create Azure ML Studio workspace

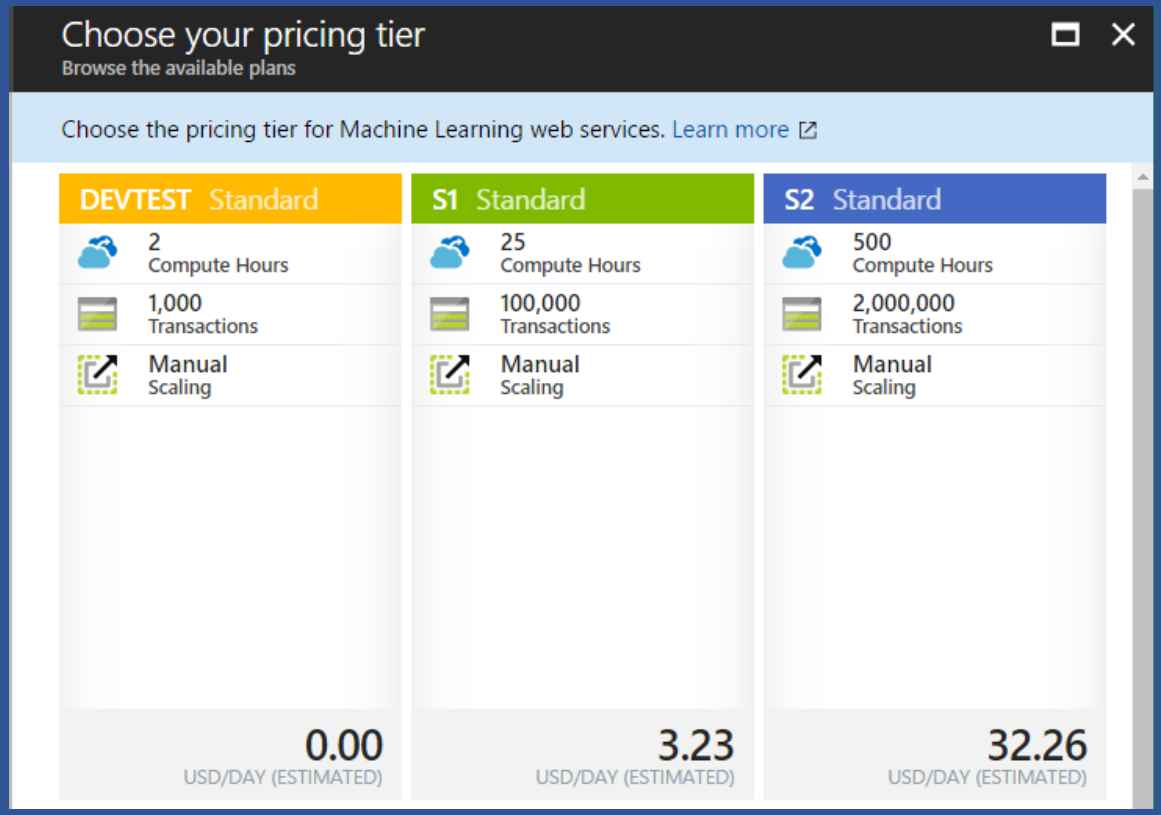
4. Workspace name = ws1
5. Subscription = defult
6. Resource group = Create new: rs1
7. Location = Southeast Asia
8. Storage account = Create new: names1
9. Workspace pricing tier = Standard
10. Web service plan = Create new: ws1Plan



First Experimental

Create Azure ML Studio workspace

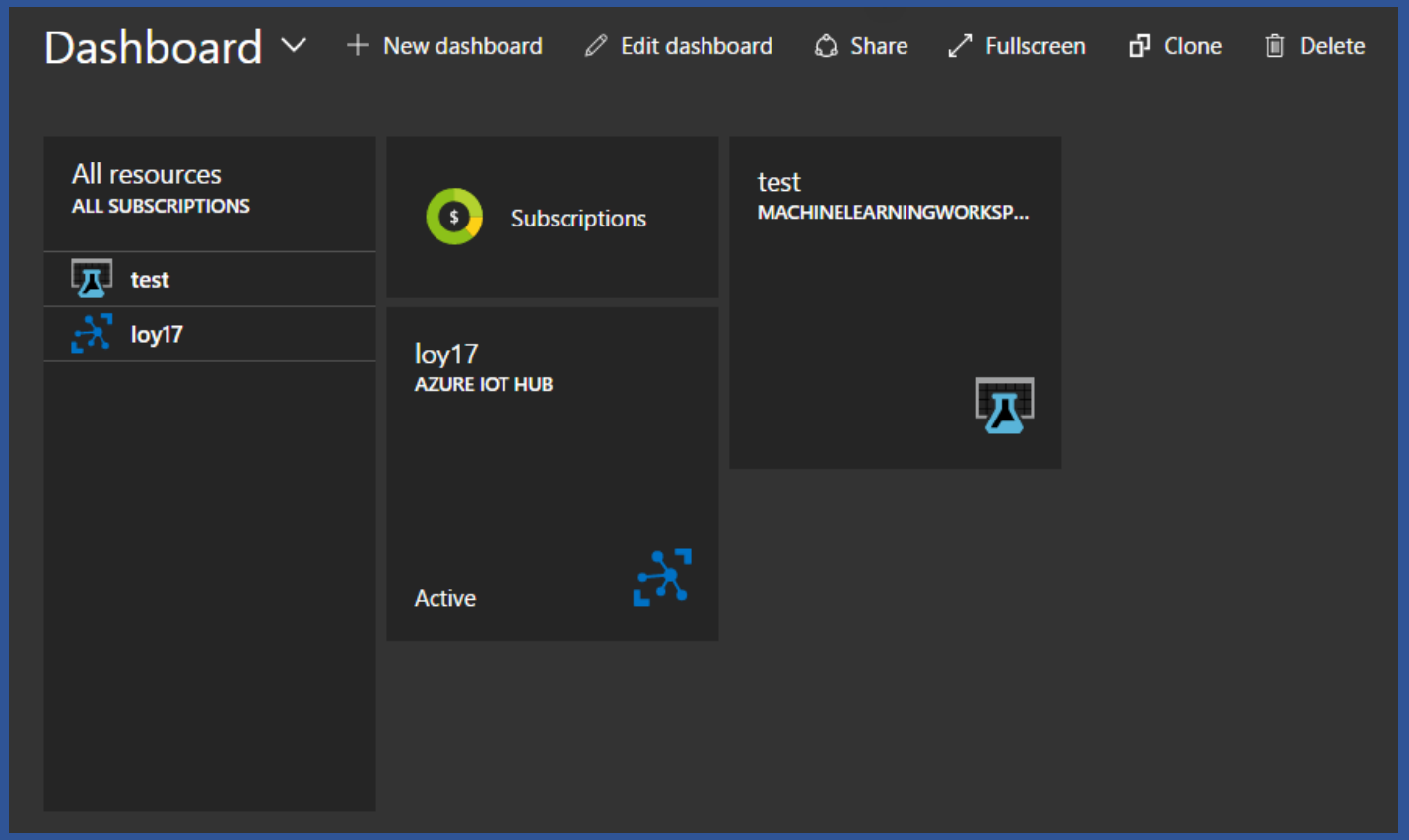
11. Click No pricing tier selected
12. Click DEVTEST
13. Click Pin to dashboard
14. Click Create



First Experimental

Create Azure ML Studio workspace

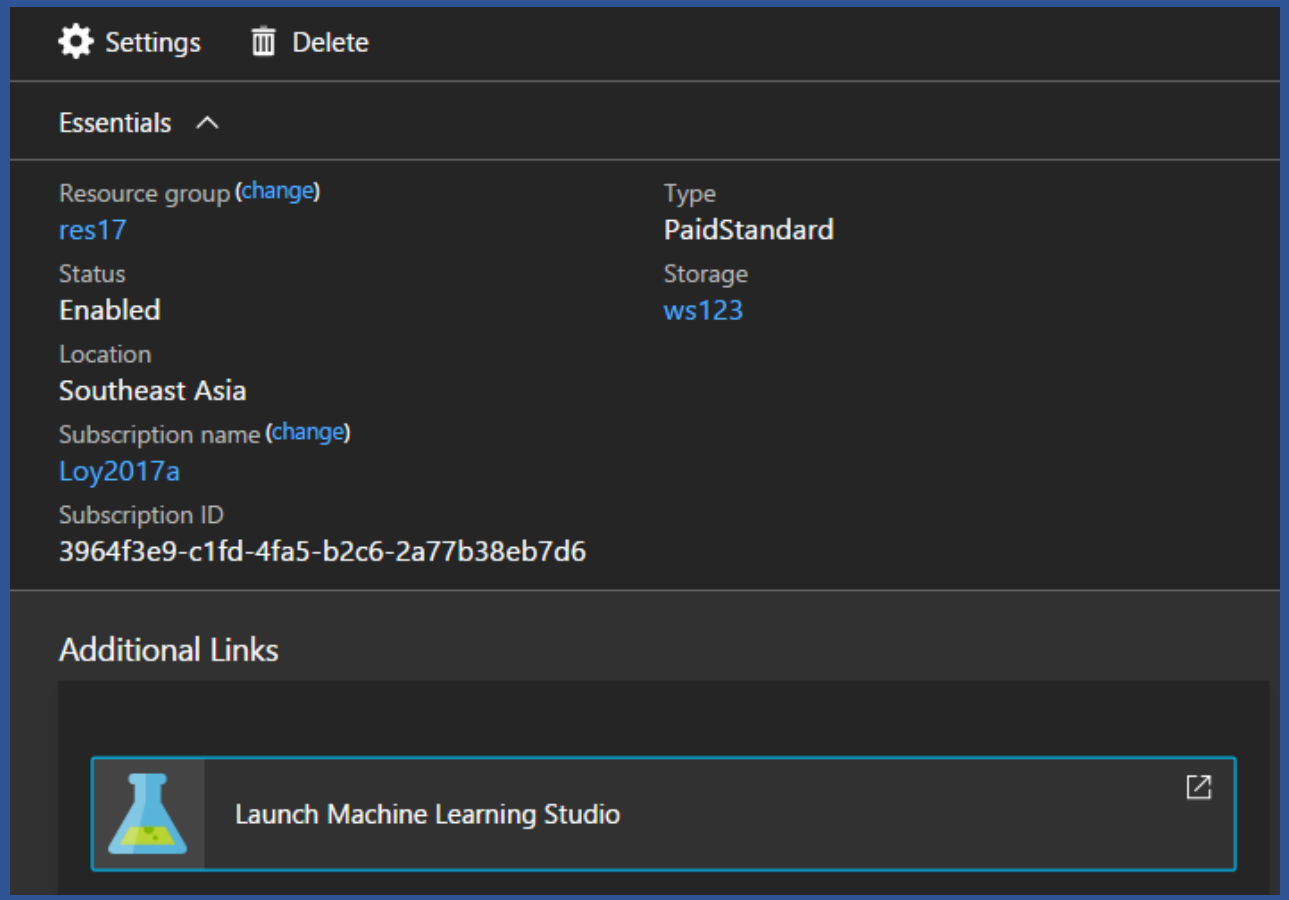
15. Click at Machine Learning workgroup on dashboard



First Experimental

Create Azure ML Studio workspace

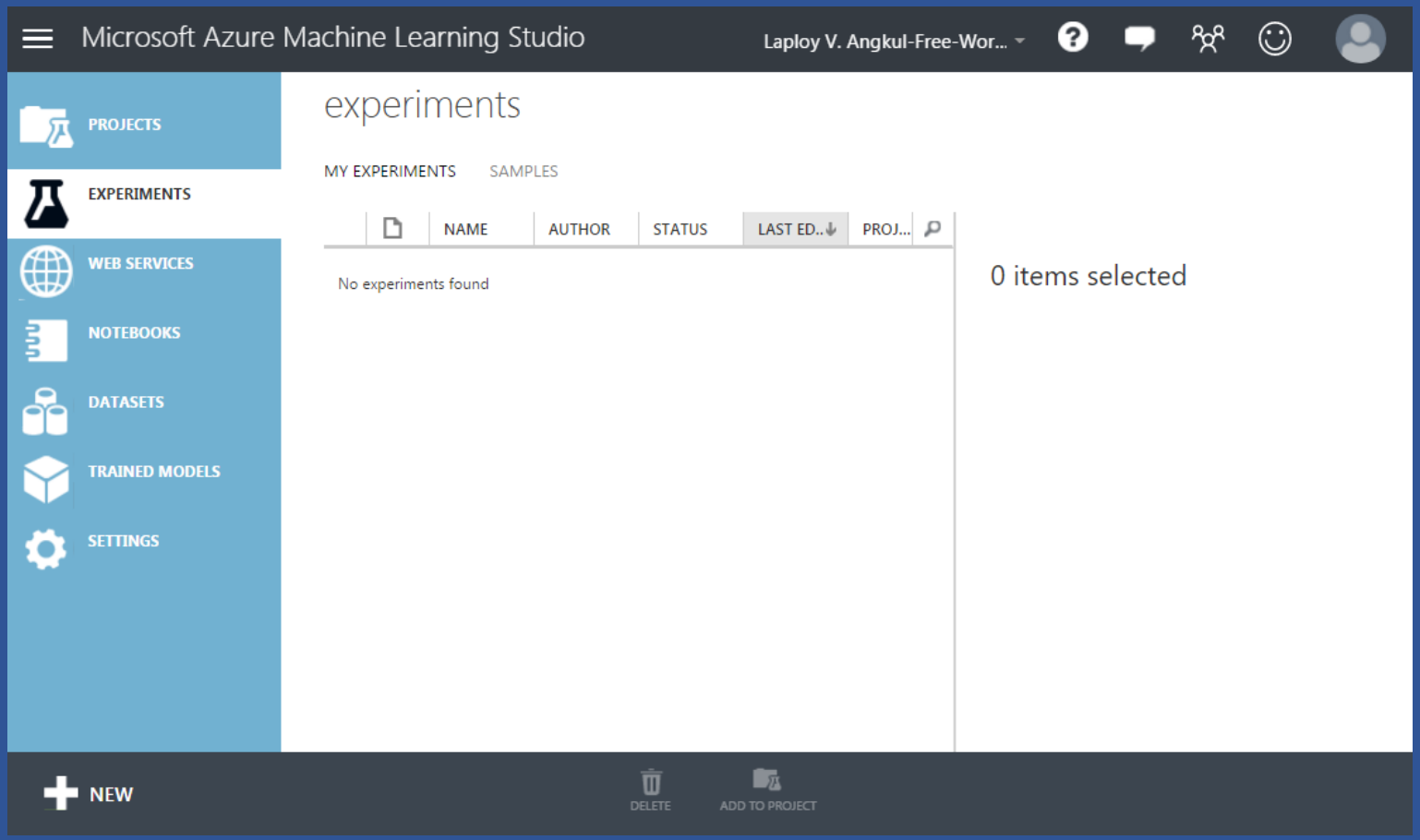
16. Click Launch Machine Learning Studio



First Experimental

Create Azure ML Studio workspace

Blank, new ML Studio workspace



First Experimental

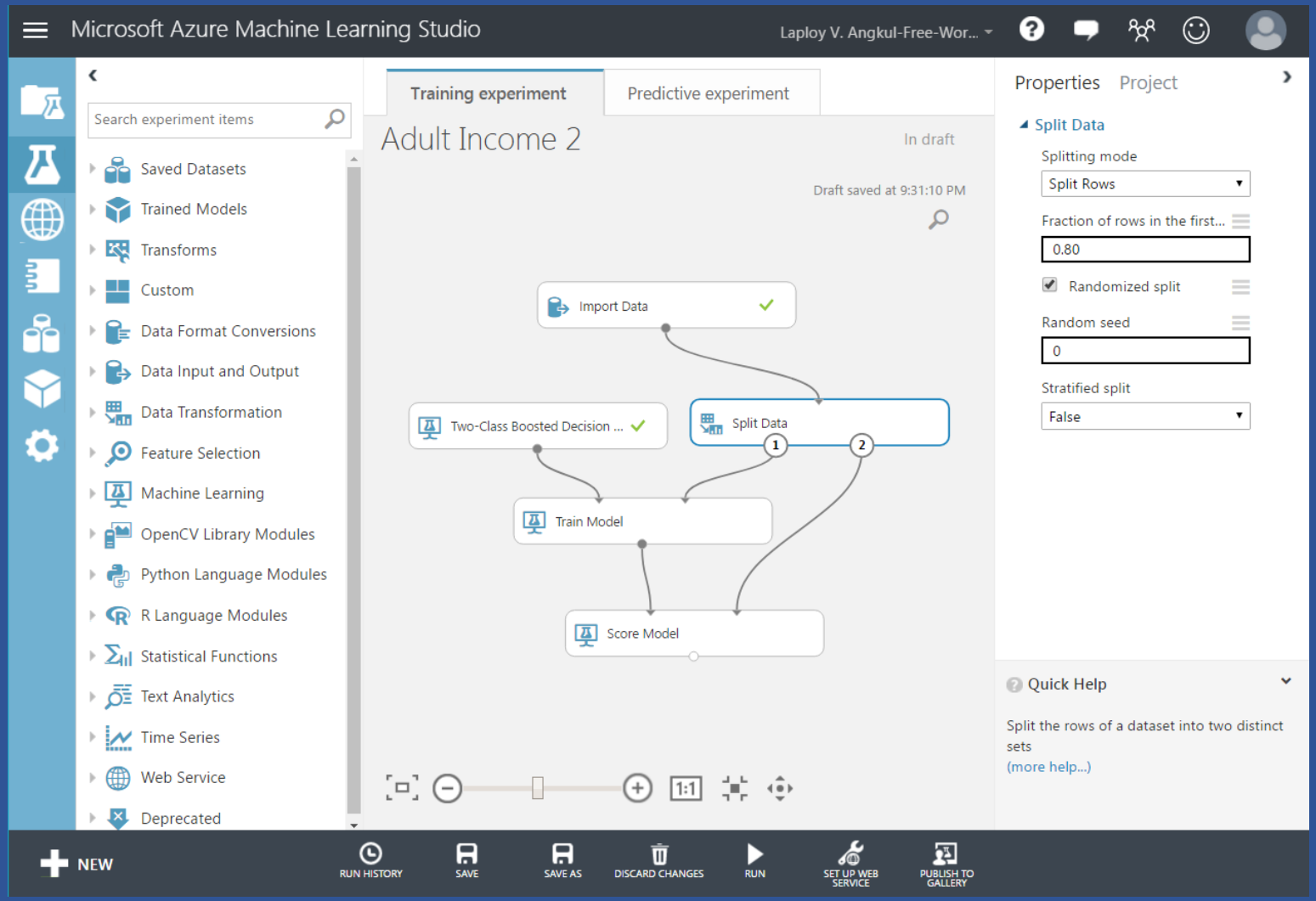
Train, Test, Evaluate for Binary Classification

Predicting whether a person's income exceeds $50,000 per year based on his demographics or census data

1. Download, prepare, and upload a census income dataset.
2. Create a new Azure Machine Learning experiment.
3. Train and evaluate a prediction model.

First Experimental

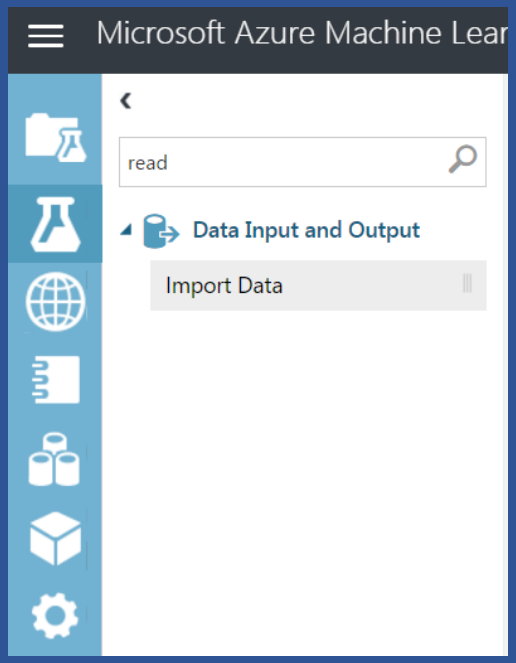
The overall workflow of the experiment



First Experimental

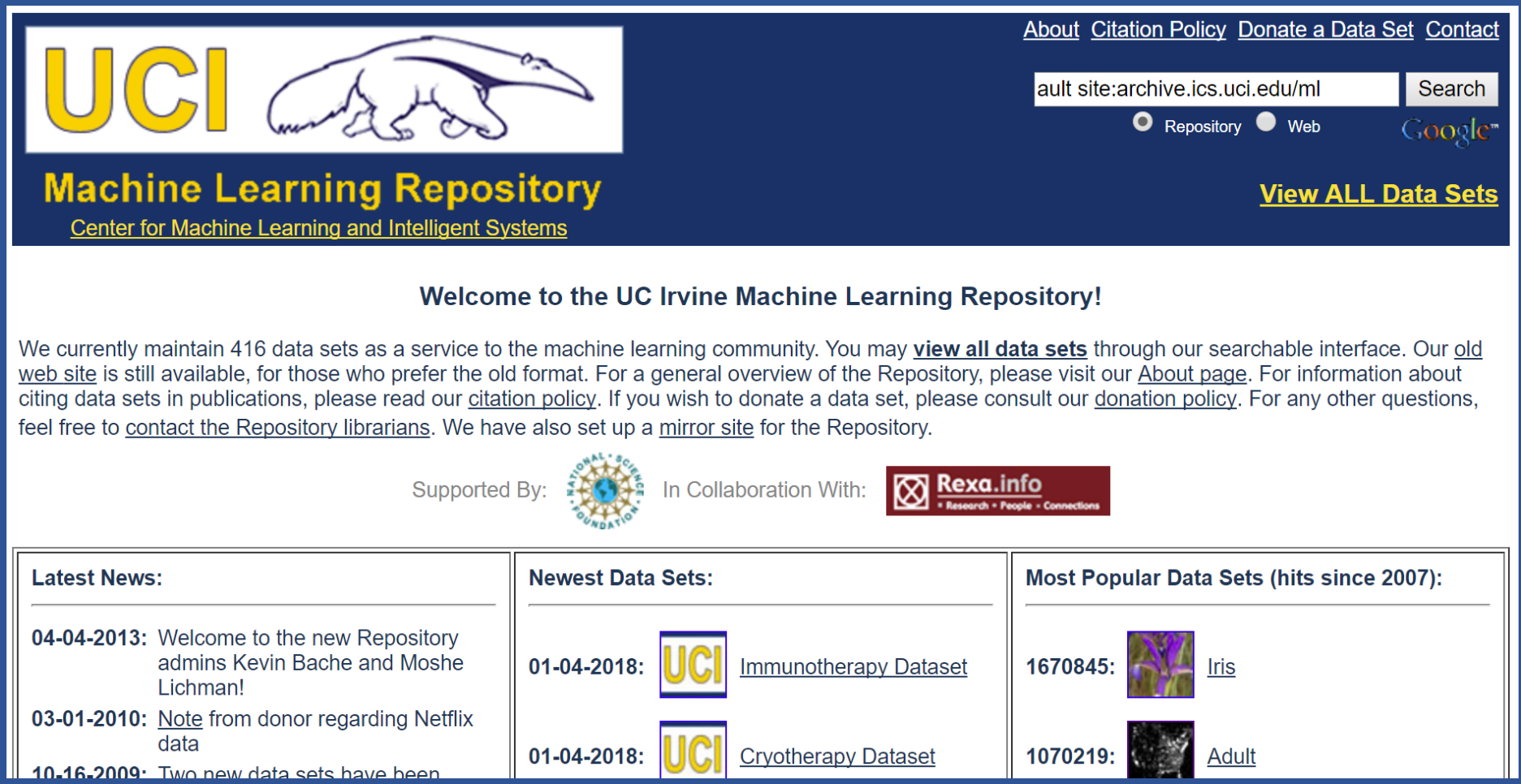
Train, Test, Evaluate for Binary Classification

- Create New blank experiment. Name = Adult Income 1
- Click Data Input and Output
- Drag & drop Import Data



First Experimental

http://archive.ics.uci.edu/ml



First Experimental

http://archive.ics.uci.edu/ml/datasets/Adult

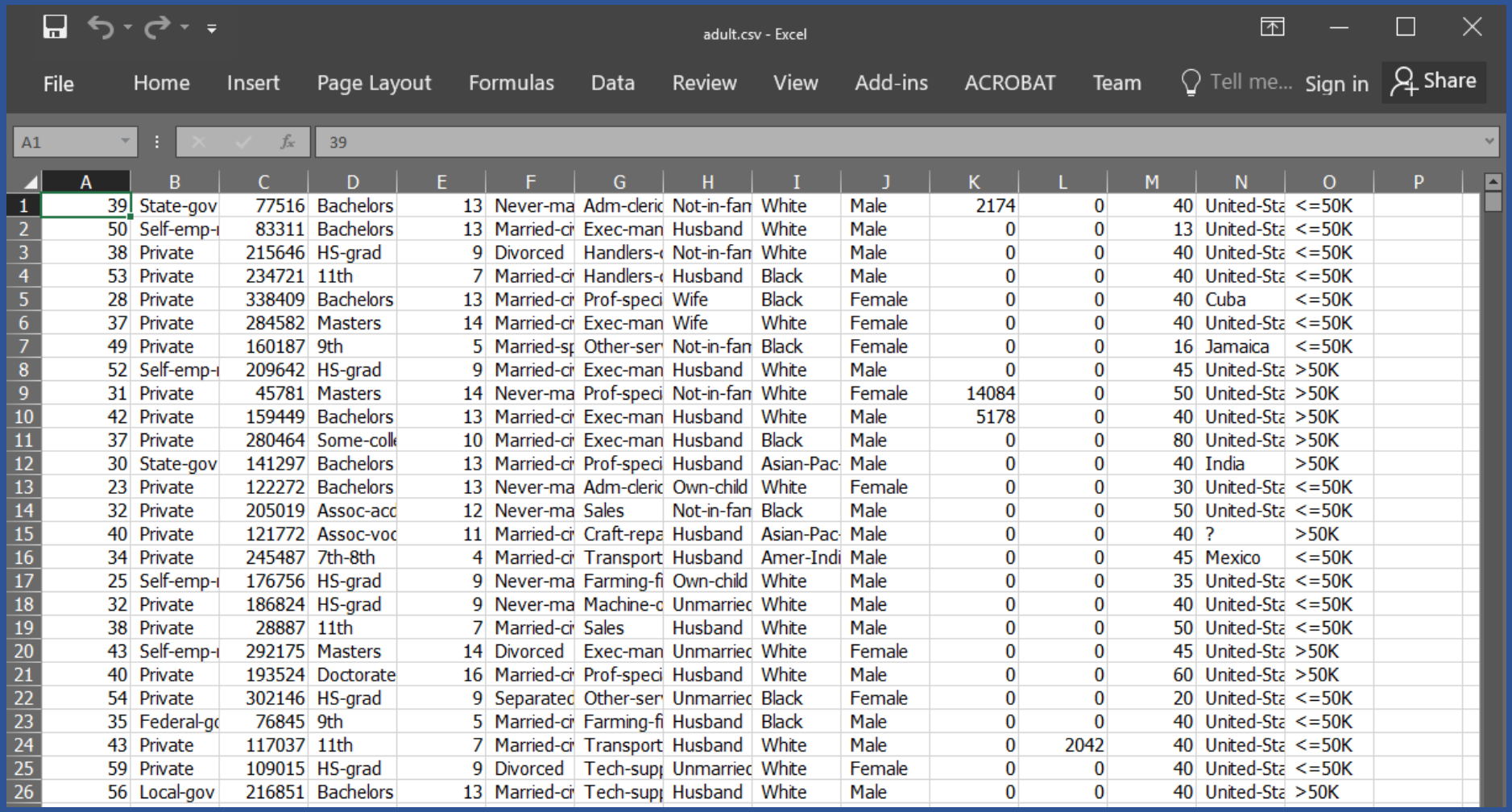
Attribute Information:

Listing of attributes:

>50K, <=50K.

age: continuous.
workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, N
fnlwgt: continuous.
education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8
Preschool.
education-num: continuous.
marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-a
occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handle
fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
sex: Female, Male.
capital-gain: continuous.
capital-loss: continuous.
hours-per-week: continuous.
native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Gu
Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France
Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&T

First Experimental



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	39	State-gov	77516	Bachelors	13	Never-ma	Adm-cleric	Not-in-fan	White	Male	2174	0	40	United-Sta	<=50K	
2	50	Self-emp-r	83311	Bachelors	13	Married-ci	Exec-man	Husband	White	Male	0	0	13	United-Sta	<=50K	
3	38	Private	215646	HS-grad	9	Divorced	Handlers-c	Not-in-fan	White	Male	0	0	40	United-Sta	<=50K	
4	53	Private	234721	11th	7	Married-ci	Handlers-c	Husband	Black	Male	0	0	40	United-Sta	<=50K	
5	28	Private	338409	Bachelors	13	Married-ci	Prof-speci	Wife	Black	Female	0	0	40	Cuba	<=50K	
6	37	Private	284582	Masters	14	Married-ci	Exec-man	Wife	White	Female	0	0	40	United-Sta	<=50K	
7	49	Private	160187	9th	5	Married-sp	Other-serv	Not-in-fan	Black	Female	0	0	16	Jamaica	<=50K	
8	52	Self-emp-r	209642	HS-grad	9	Married-ci	Exec-man	Husband	White	Male	0	0	45	United-Sta	>50K	
9	31	Private	45781	Masters	14	Never-ma	Prof-speci	Not-in-fan	White	Female	14084	0	50	United-Sta	>50K	
10	42	Private	159449	Bachelors	13	Married-ci	Exec-man	Husband	White	Male	5178	0	40	United-Sta	>50K	
11	37	Private	280464	Some-colle	10	Married-ci	Exec-man	Husband	Black	Male	0	0	80	United-Sta	>50K	
12	30	State-gov	141297	Bachelors	13	Married-ci	Prof-speci	Husband	Asian-Pac	Male	0	0	40	India	>50K	
13	23	Private	122272	Bachelors	13	Never-ma	Adm-cleric	Own-child	White	Female	0	0	30	United-Sta	<=50K	
14	32	Private	205019	Assoc-acd	12	Never-ma	Sales	Not-in-fan	Black	Male	0	0	50	United-Sta	<=50K	
15	40	Private	121772	Assoc-voc	11	Married-ci	Craft-repa	Husband	Asian-Pac	Male	0	0	40	?	>50K	
16	34	Private	245487	7th-8th	4	Married-ci	Transport	Husband	Amer-Indi	Male	0	0	45	Mexico	<=50K	
17	25	Self-emp-r	176756	HS-grad	9	Never-ma	Farming-fi	Own-child	White	Male	0	0	35	United-Sta	<=50K	
18	32	Private	186824	HS-grad	9	Never-ma	Machine-c	Unmarried	White	Male	0	0	40	United-Sta	<=50K	
19	38	Private	28887	11th	7	Married-ci	Sales	Husband	White	Male	0	0	50	United-Sta	<=50K	
20	43	Self-emp-r	292175	Masters	14	Divorced	Exec-man	Unmarried	White	Female	0	0	45	United-Sta	>50K	
21	40	Private	193524	Doctorate	16	Married-ci	Prof-speci	Husband	White	Male	0	0	60	United-Sta	>50K	
22	54	Private	302146	HS-grad	9	Separated	Other-serv	Unmarried	Black	Female	0	0	20	United-Sta	<=50K	
23	35	Federal-go	76845	9th	5	Married-ci	Farming-fi	Husband	Black	Male	0	0	40	United-Sta	<=50K	
24	43	Private	117037	11th	7	Married-ci	Transport	Husband	White	Male	0	2042	40	United-Sta	<=50K	
25	59	Private	109015	HS-grad	9	Divorced	Tech-supp	Unmarried	White	Female	0	0	40	United-Sta	<=50K	
26	56	Local-gov	216851	Bachelors	13	Married-ci	Tech-supp	Husband	White	Male	0	0	40	United-Sta	>50K	

First Experimental

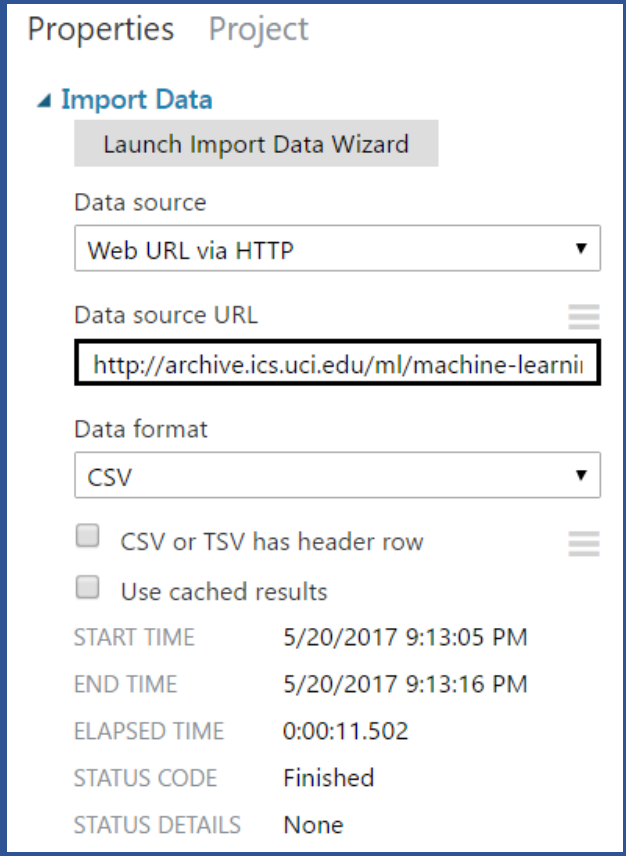
Train, Test, Evaluate for Binary Classification

Configure Import data module:

- Data source = Web URL via HTTP
- Data source URL =

  http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data

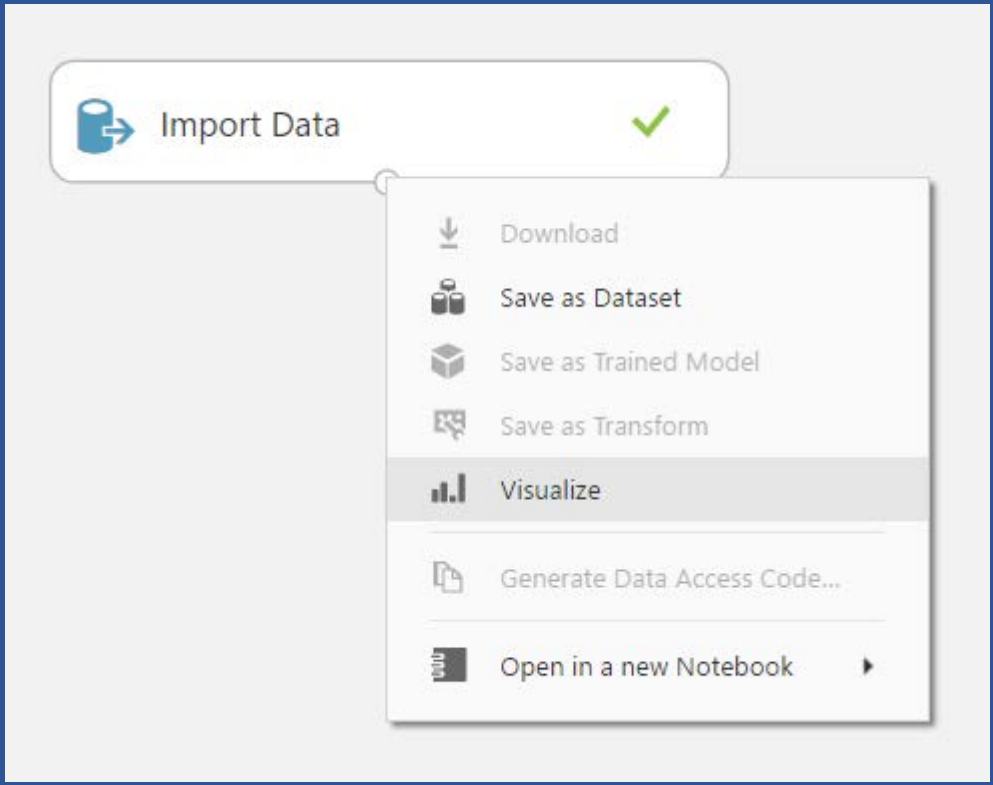
- Data format = CSV
- Run experiment



First Experimental

Train, Test, Evaluate for Binary Classification

- Right click at the output of Import Data
- Click Visualize



First Experimental

Train, Test, Evaluate for Binary Classification

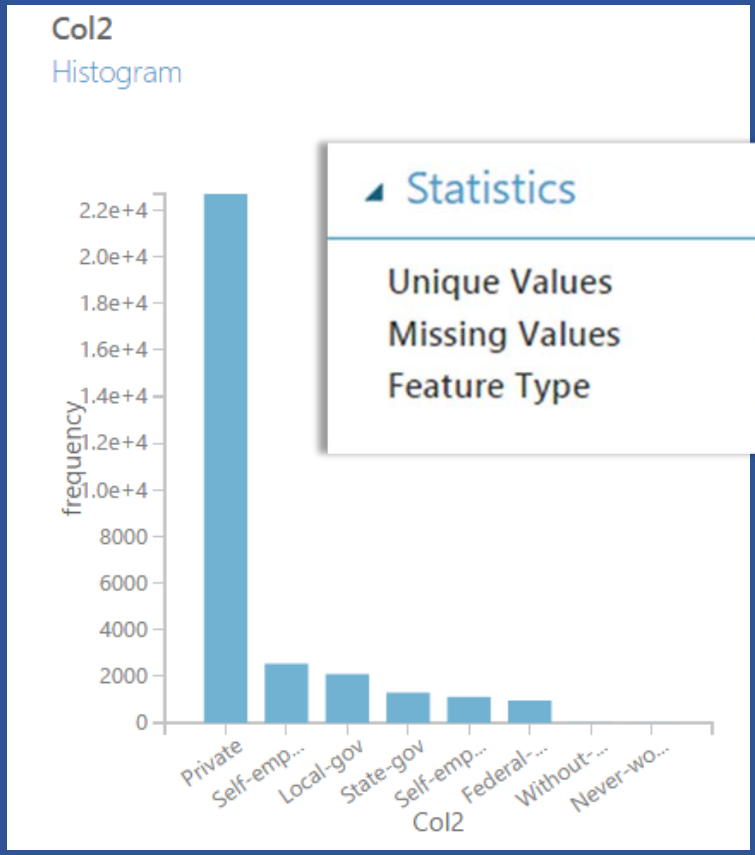
- Click on Col2
- Look at Statistics and Histogram

Adult Income › Import Data › Results dataset

rows 32562 columns 15

view as

Col1	Col2	Col3	Col4	Col5	Col6
39	State-gov	77516	Bachelors	13	Never-married
50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse
38	Private	215646	HS-grad	9	Divorced
53	Private	234721	11th	7	Married-civ-spouse
28	Private	338409	Bachelors	13	Married-civ-spouse
37	Private	284582	Masters	14	Married-civ-spouse



First Experimental

Train, Test, Evaluate for Binary Classification

Split up the dataset

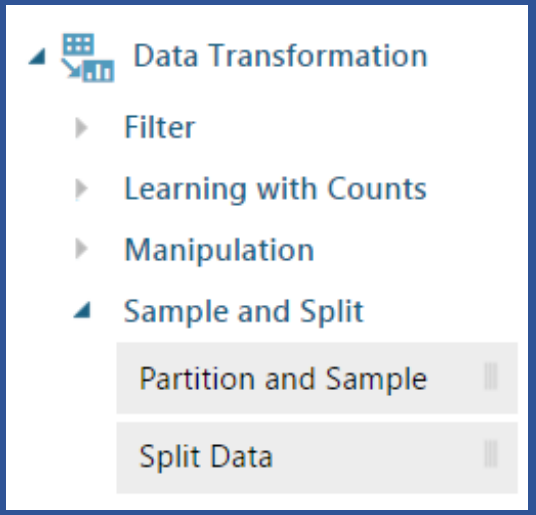
- Training data This grouping is used for creating our new predictive model based on the inherent patterns found in the historical data via the ML algorithm we use for the solution.
- Validation data This grouping is used for testing the new predictive model against known outcomes to determine accuracy and probabilities.

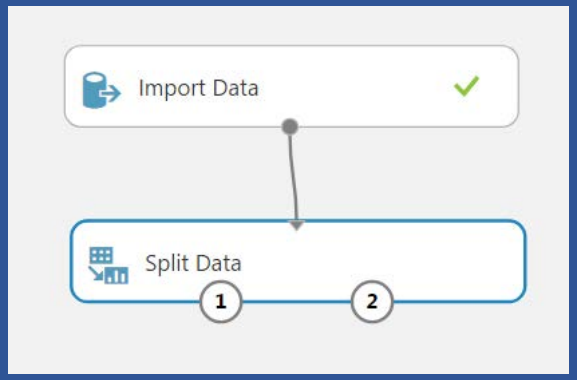
First Experimental

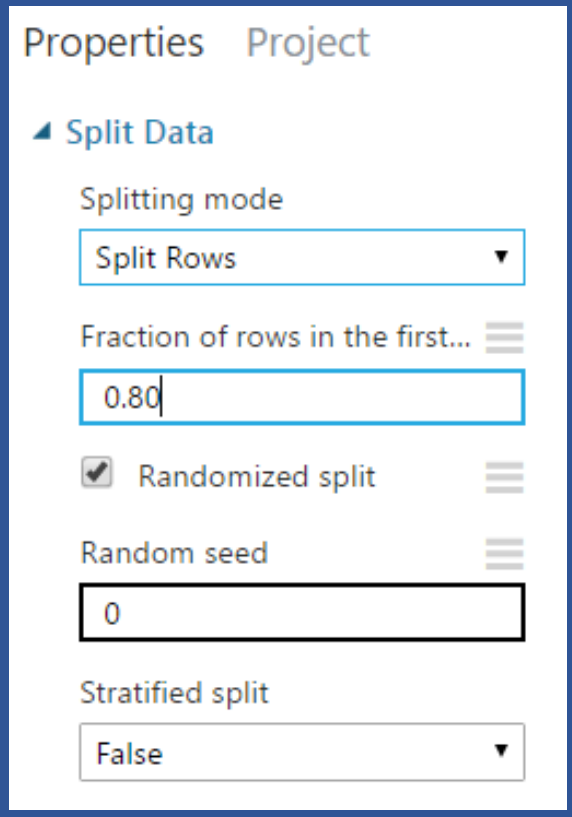
Train, Test, Evaluate for Binary Classification

Add Split Data:

- Click Data Transformation
- Click Sample and Split
- Drag & drop Split Data module into canvas
- Connect Import Data to Split Data
- Set properties Fraction of row to 0.80



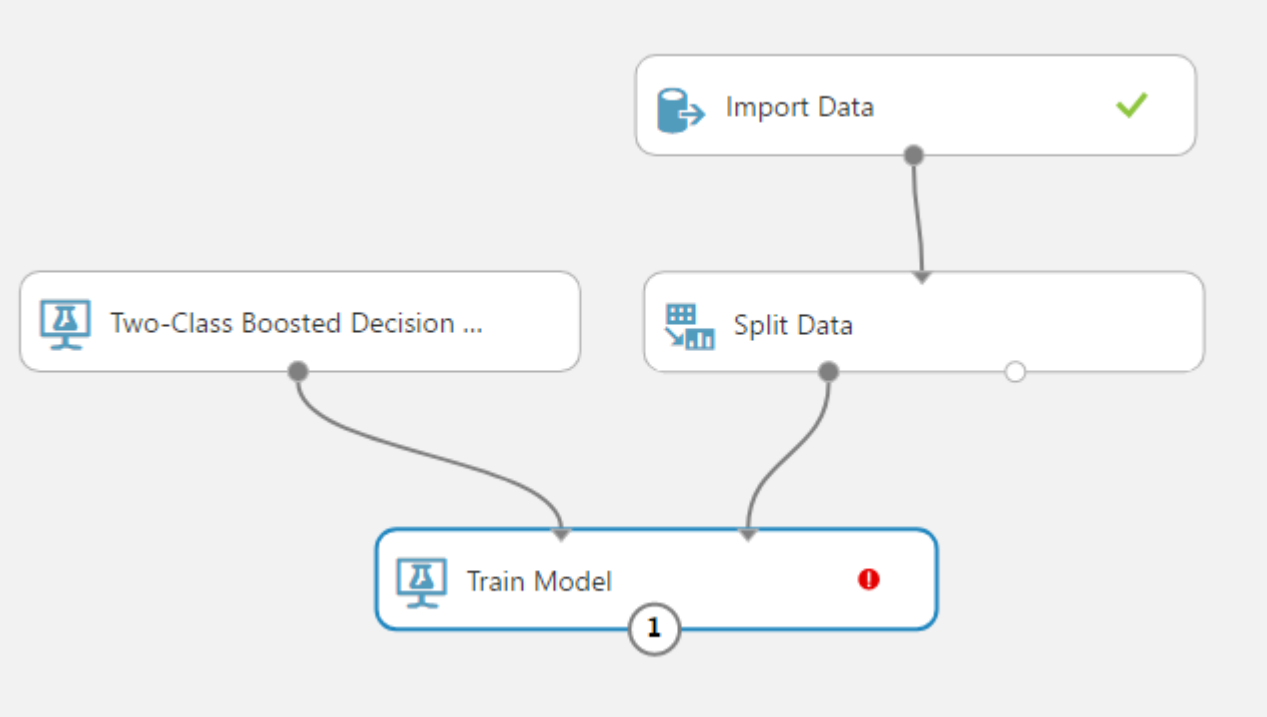


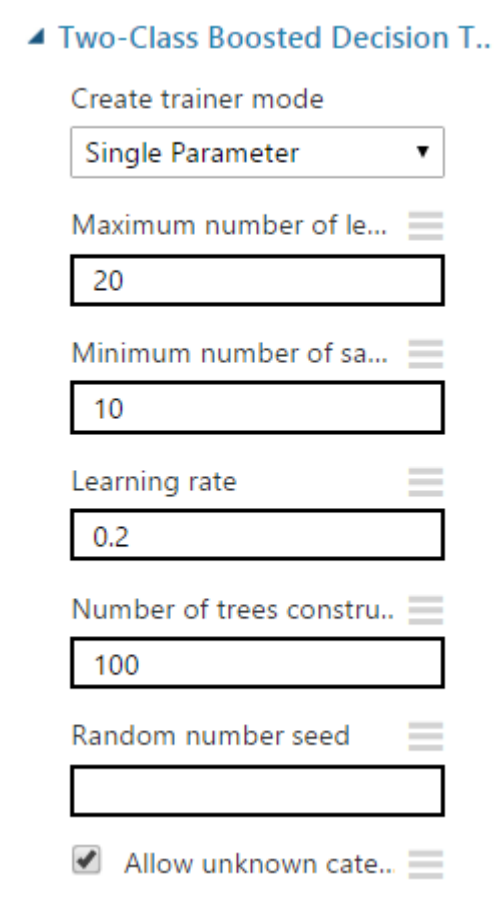


First Experimental

Train, Test, Evaluate for Binary Classification

- Add Two-Class Boosted Decision Tree and Train Model
- Connect Two-Class Boosted Decision Tree to Train Model
- Connect Split Data to Train Model





First Experimental

Label column specification

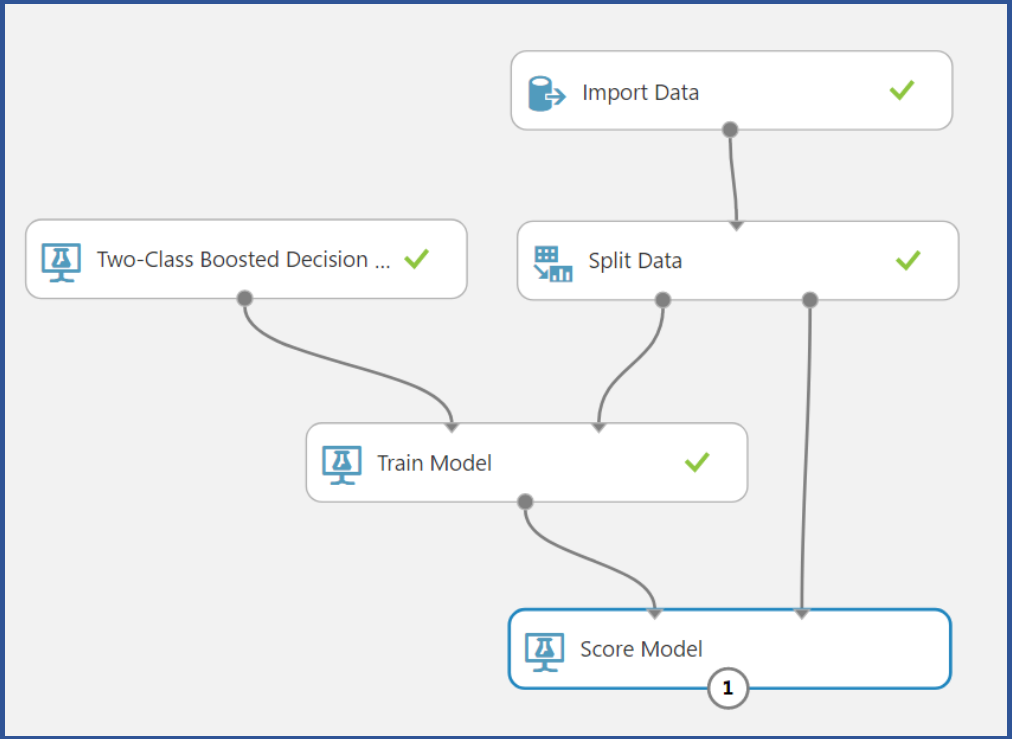
- Click Train Model
- Click Launch column selector
- Include col15
- Click 
- Save
- Run

First Experimental

Score the model

Score the model:

- Add Score Model to canvas
- Connect Score Model to Train and Split model
- Run



First Experimental

Visualize the model results

Col11	Col12	Col13	Col14	Col15	Scored Labels	Scored Probabilities
0	0	50	United-States	<=50K	<=50K	0.425173
0	0	40	Puerto-Rico	<=50K	<=50K	0.008254
0	0	35	United-States	<=50K	<=50K	0.002206

Visualize the model results:

- Visualize output of Score Model
- Scored Labels This column denotes the model's prediction for this row of the dataset.
- Scored Probabilities This column denotes the numerical probability (or the likelihood) of whether the income level for this row exceeds $50,000.

First Experimental

Type of datasets

Training set

- A set of examples used for learning
- Where the answer value is known.

Validation set

- A set of examples data
- Used to tune the architecture of a classifier
- And estimate the error

Test set

- Use to test the performances of a classifier
- Never used during the training process
- Give estimate of error

First Experimental

More Information

Two-Class Boosted Decision Tree

https://msdn.microsoft.com/en-us/library/azure/dn906025.aspx

Score Model

https://msdn.microsoft.com/en-us/library/azure/dn905995.aspx

Published Experiment

https://gallery.cortanaintelligence.com/Experiment/Adult-Income-1