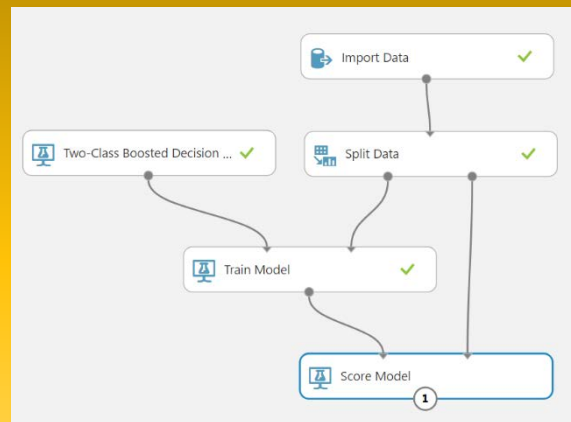


First Experimental

FIRST EXPERIMENTAL



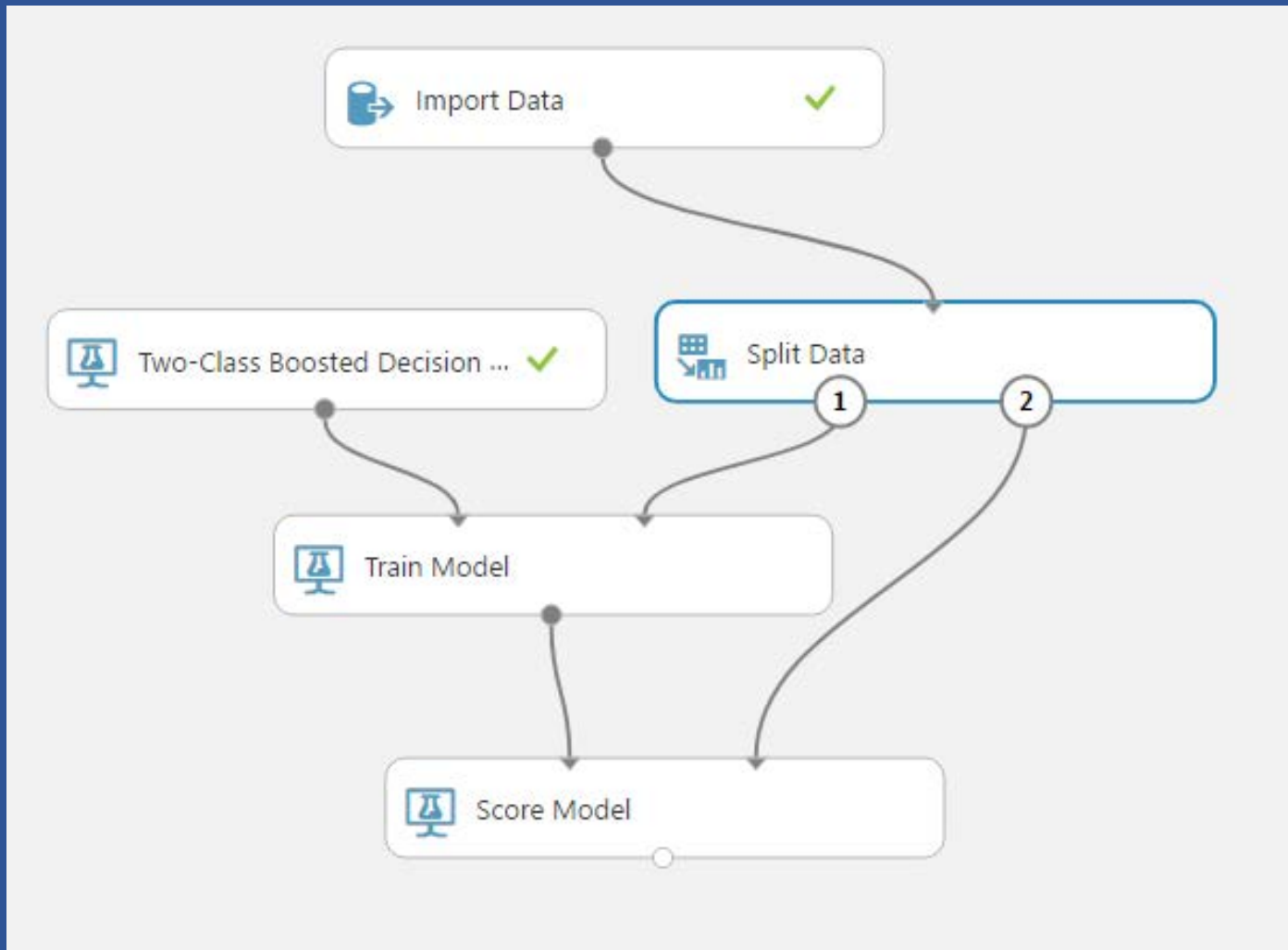
First Experimental

In this session

- Sign 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

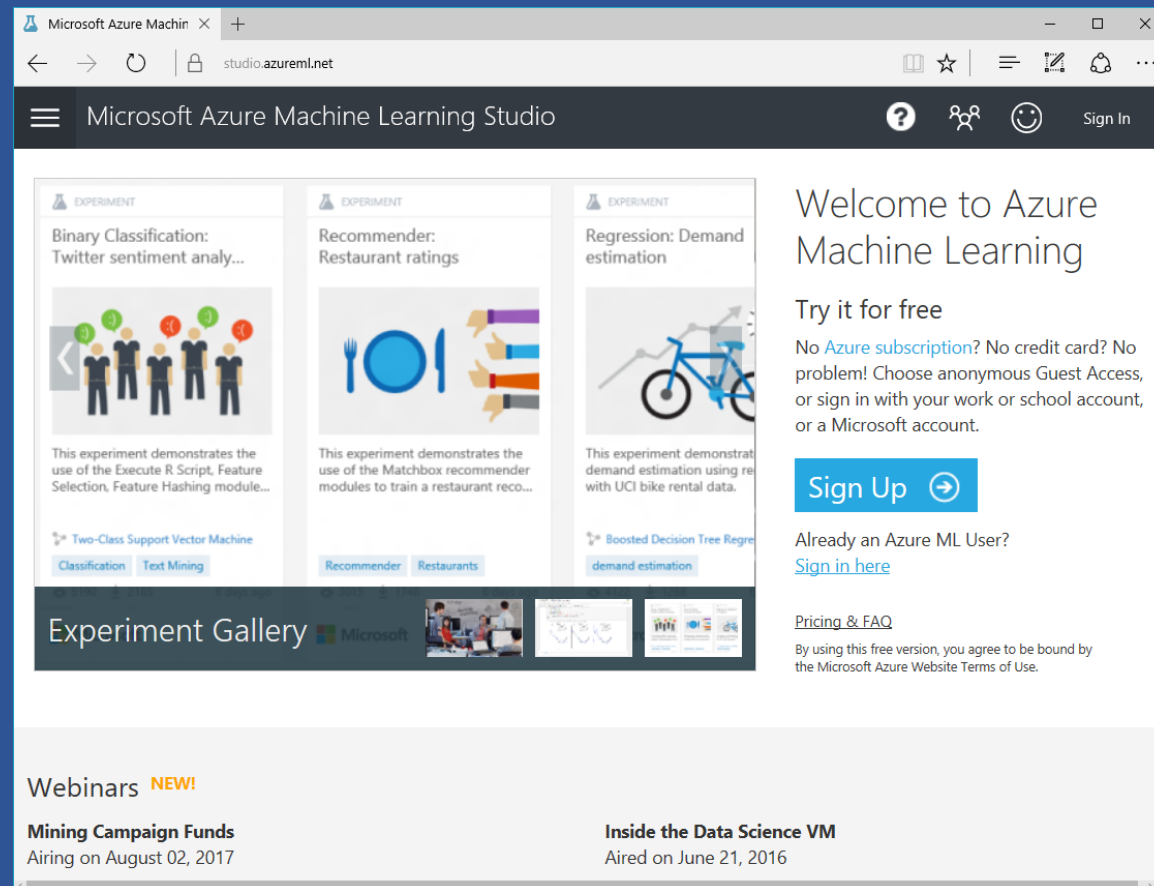
First experiment model



First Experimental

Sing Up FREE Azure ML Studio Subscription

<https://studio.azureml.net/>



First Experimental

Sing Up FREE Azure ML Studio Subscription

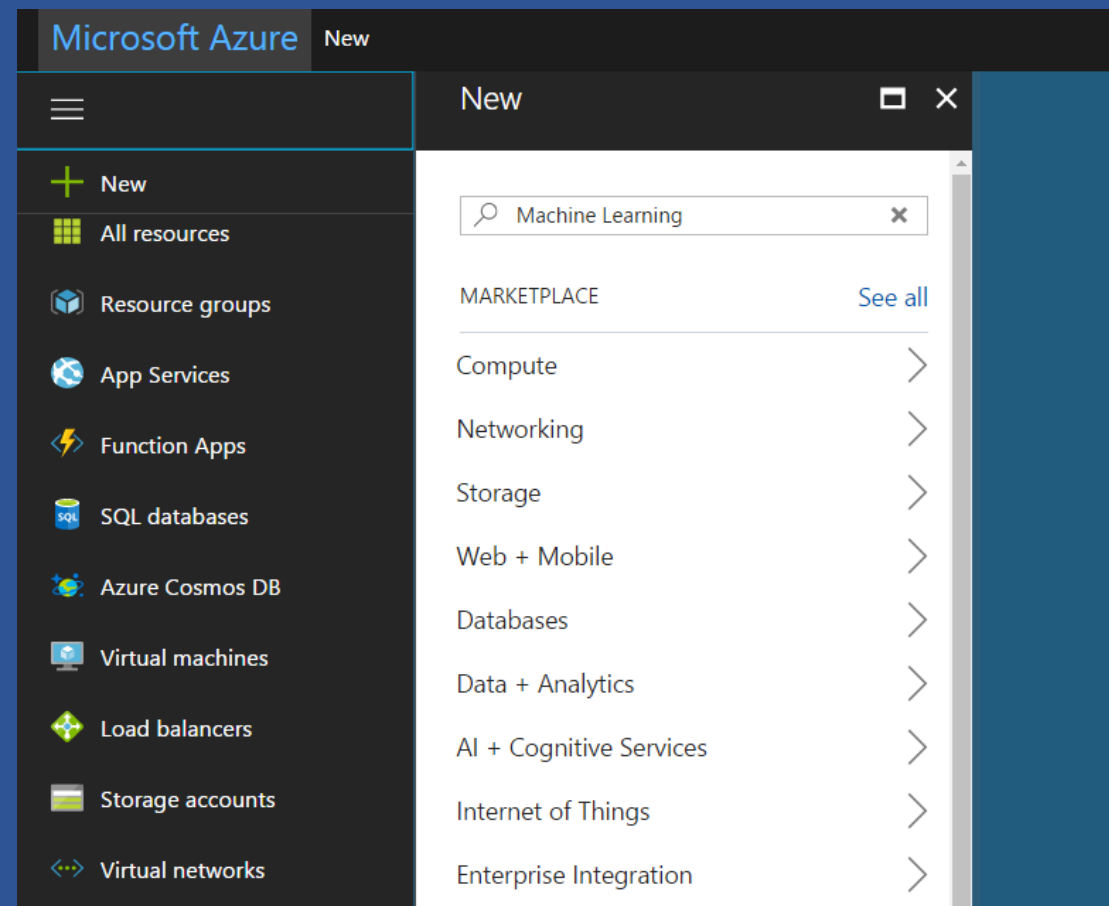
Free Workspace -> sign up here

Quick Evaluation	Most Popular	Enterprise Grade
Guest Workspace	Free Workspace	Standard Workspace
8-hour trial	\$0/month	\$9.99/month
No sign-in required.	Don't already have a Microsoft account? Simply sign up here .	Azure subscription required Other charges may apply. Read more .
Enter	Sign In	Create Workspace
<ul style="list-style-type: none">▪ No hassle instant access▪ Stock sample datasets▪ ML models built in minutes▪ Full range of ML algorithms	<ul style="list-style-type: none">▪ Free access that never expires▪ 10 GB storage on us▪ R and Python scripts support▪ Predictive web services	<ul style="list-style-type: none">▪ Full SLA Support▪ Bring your own Azure storage▪ Parallel graph execution▪ Elastic Web Service endpoints

First Experimental

Create Azure ML Studio workspace

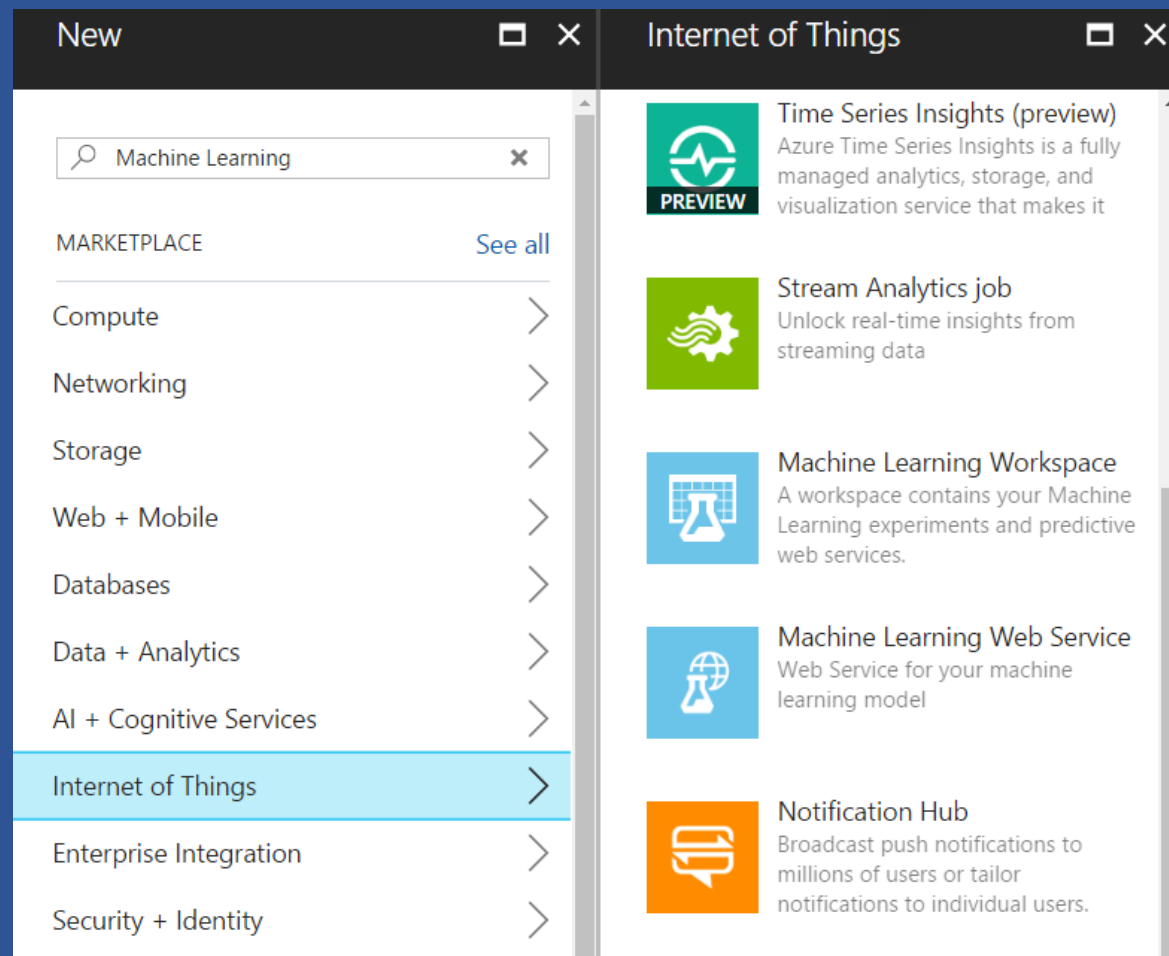
1. Go to the Azure portal <https://portal.azure.com>
2. Click +New



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 = **default**
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**

The screenshot shows the 'Machine Learning workspace' creation form in the Azure portal. The form is titled 'Machine Learning workspace' with a close button. It contains several fields and options:

- Workspace name:** A text input field containing 'ws1' with a green checkmark.
- Subscription:** A dropdown menu showing 'Loy2017a'.
- Resource group:** A section with two radio buttons: 'Create new' (selected) and 'Use existing'. Below it is a text input field containing 'r1' with a green checkmark.
- Location:** A dropdown menu showing 'Southeast Asia'.
- Storage account:** A section with two radio buttons: 'Create new' (selected) and 'Use existing'. Below it is a text input field containing 'ws123' with a green checkmark.
- Workspace pricing tier:** A dropdown menu showing 'Standard'.
- Web service plan:** A section with two radio buttons: 'Create new' (selected) and 'Use existing'. Below it is a text input field containing 'ws1Plan' with a green checkmark.
- Web service plan pricing tier:** A section with a blue background and a right arrow. It contains the text 'No pricing tier selected'.
- Pin to dashboard:** A checkbox that is currently unchecked.
- Create:** A blue button at the bottom left.
- Automation options:** A link at the bottom right.










First Experimental

Create Azure ML Studio workspace

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

Choose your pricing tier
Browse the available plans

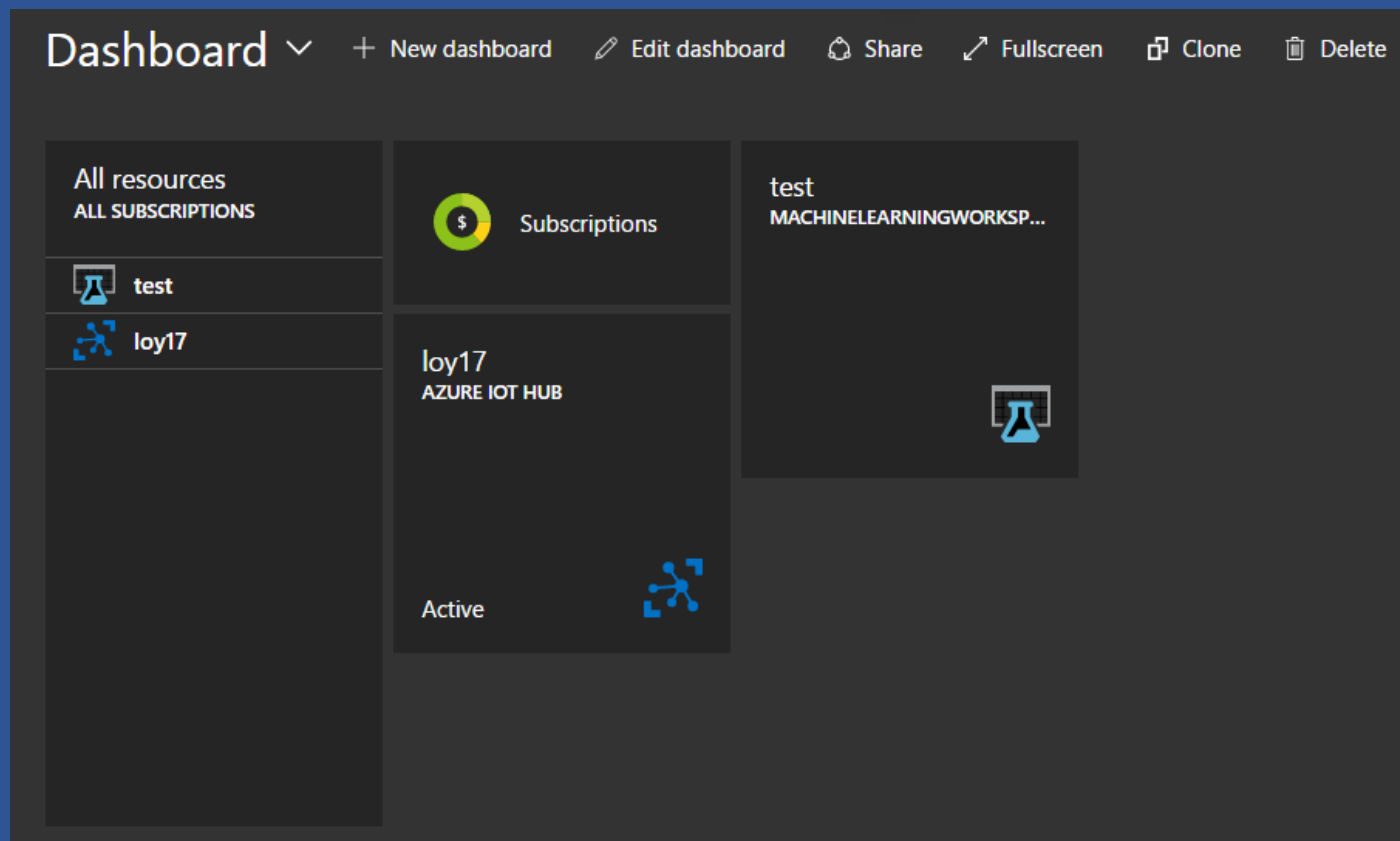
Choose the pricing tier for Machine Learning web services. [Learn more](#)

DEVTEST Standard	S1 Standard	S2 Standard
 2 Compute Hours	 25 Compute Hours	 500 Compute Hours
 1,000 Transactions	 100,000 Transactions	 2,000,000 Transactions
 Manual Scaling	 Manual Scaling	 Manual Scaling
0.00 USD/DAY (ESTIMATED)	3.23 USD/DAY (ESTIMATED)	32.26 USD/DAY (ESTIMATED)

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**

 Settings  Delete

Essentials ^

Resource group (change)	Type
res17	PaidStandard
Status	Storage
Enabled	ws123
Location	
Southeast Asia	
Subscription name (change)	
Loy2017a	
Subscription ID	
3964f3e9-c1fd-4fa5-b2c6-2a77b38eb7d6	

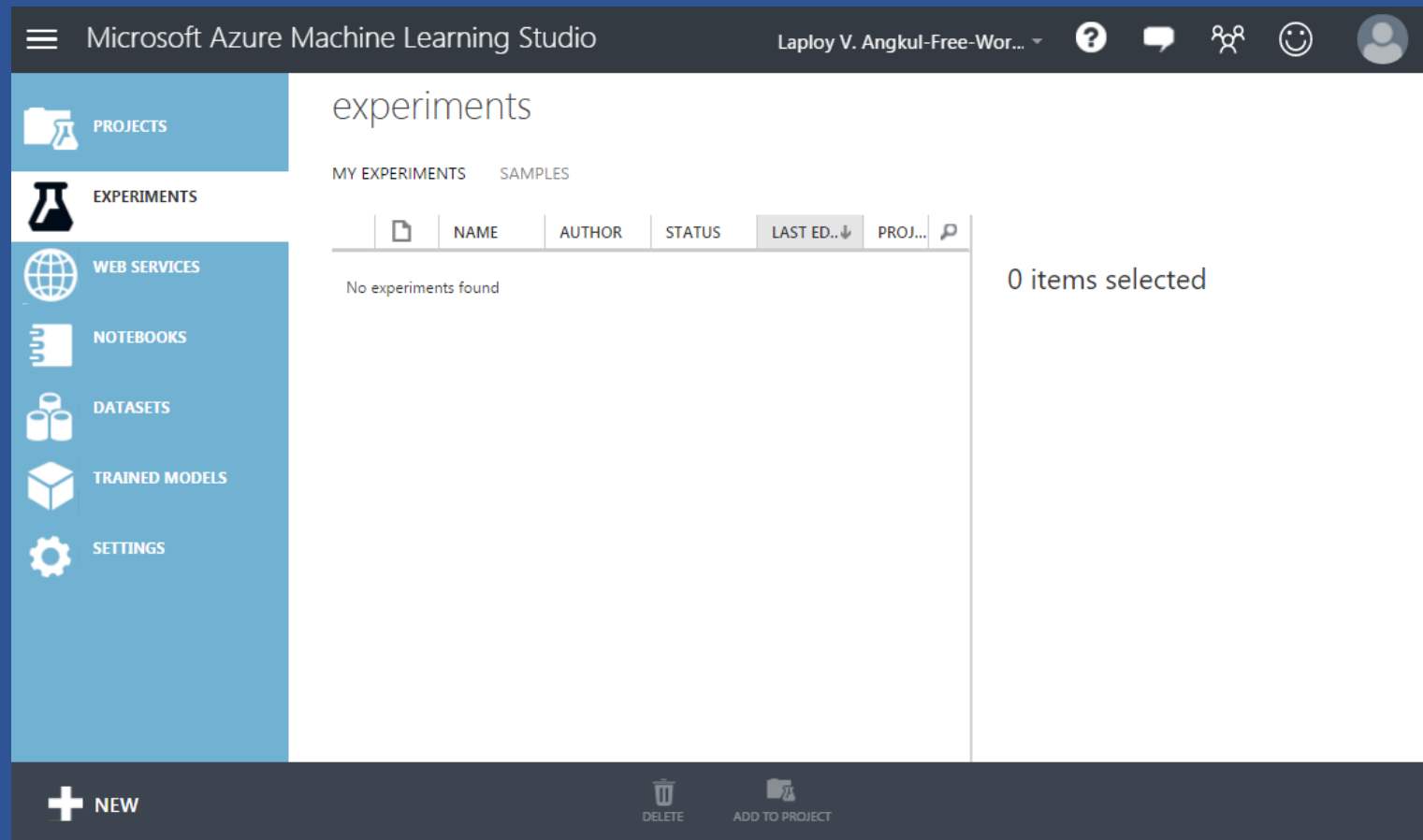
Additional Links

 Launch Machine Learning Studio 

First Experimental

Create Azure ML Studio workspace

Blank, new ML Studio workspace



First Experimental

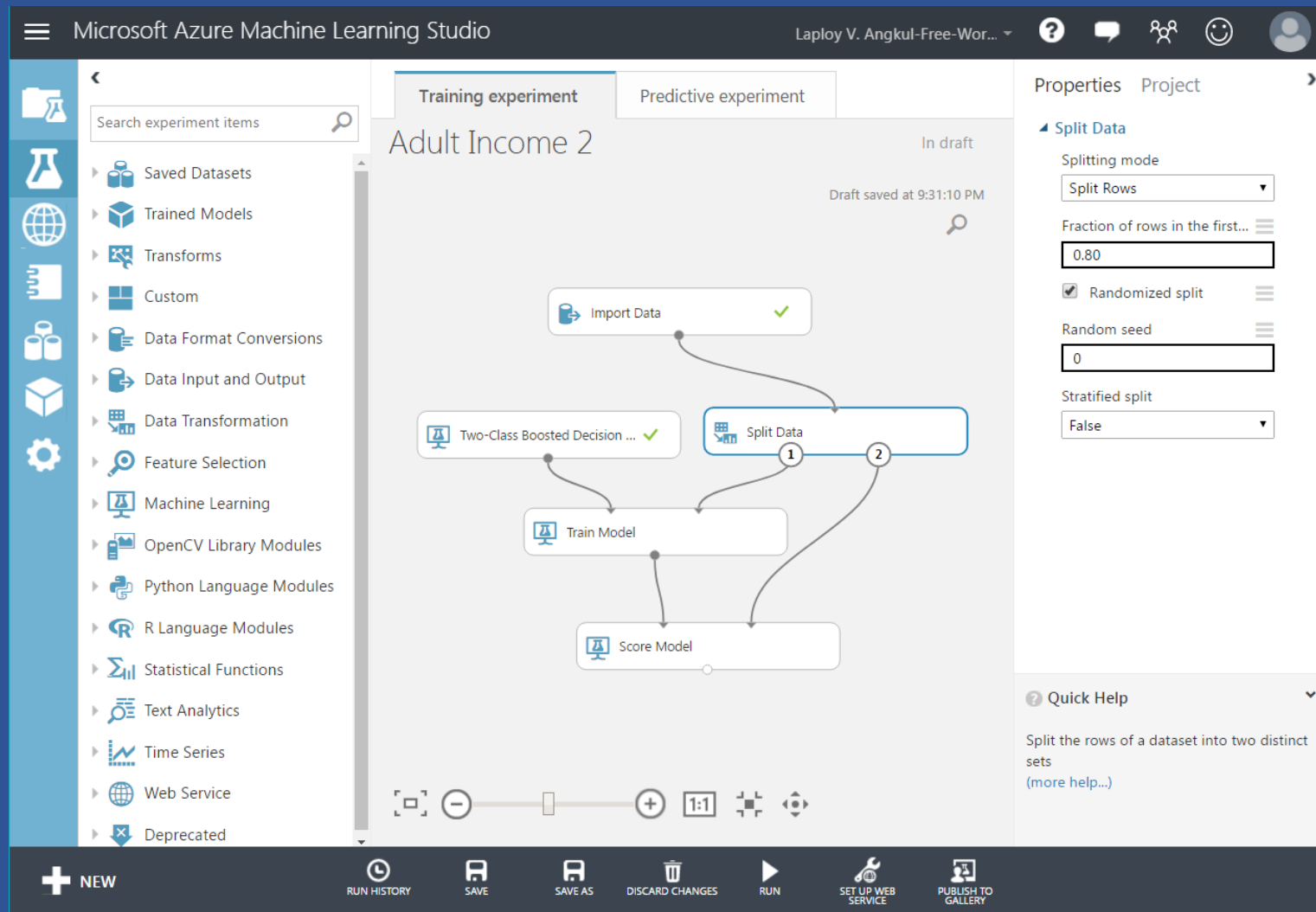
Train, Test, Evaluate for Binary Classification

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

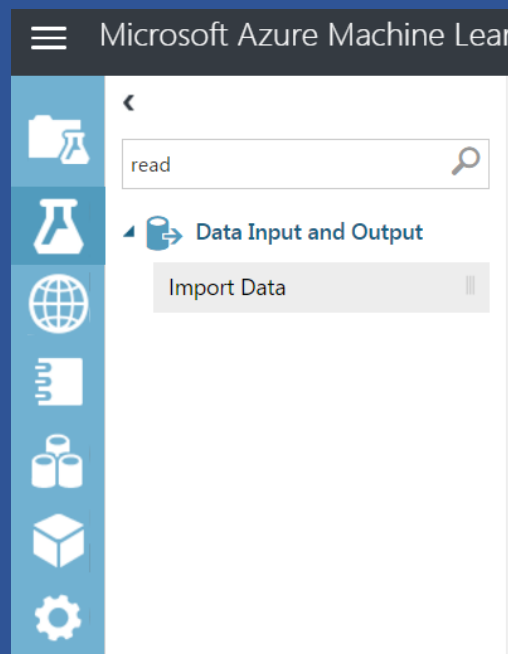
The overall workflow of the experiment



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>



[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

☒ Repository ☐ Web



[View ALL Data Sets](#)


Machine Learning Repository

[Center for Machine Learning and Intelligent Systems](#)


Welcome to the UC Irvine Machine Learning Repository!





We currently maintain 416 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable interface. Our [old web site](#) is still available, for those who prefer the old format. For a general overview of the Repository, please visit our [About page](#). For information about citing data sets in publications, please read our [citation policy](#). If you wish to donate a data set, please consult our [donation policy](#). For any other questions, feel free to [contact the Repository librarians](#). We have also set up a [mirror site](#) for the Repository.

Supported By:



In Collaboration With:



Latest News:	Newest Data Sets:	Most Popular Data Sets (hits since 2007):
04-04-2013: Welcome to the new Repository admins Kevin Bache and Moshe Lichman!	01-04-2018:  Immunotherapy Dataset	1670845:  Iris
03-01-2010: Note from donor regarding Netflix data	01-04-2018:  Cryotherapy Dataset	1070219:  Adult
10-16-2009: Two new data sets have been		

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, Trinidad&T

First Experimental

adult.csv - Excel

File Home Insert Page Layout Formulas Data Review View Add-ins ACROBAT Team Tell me... Sign in Share

A1 : X ✓ f_x 39

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K	
2	50	Self-employed	83311	Bachelors	13	Married-civ	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K	
3	38	Private	215646	HS-grad	9	Divorced	Handlers-support	Not-in-family	White	Male	0	0	40	United-States	<=50K	
4	53	Private	234721	11th	7	Married-civ	Handlers-support	Husband	Black	Male	0	0	40	United-States	<=50K	
5	28	Private	338409	Bachelors	13	Married-civ	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K	
6	37	Private	284582	Masters	14	Married-civ	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K	
7	49	Private	160187	9th	5	Married-spouse	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica	<=50K	
8	52	Self-employed	209642	HS-grad	9	Married-civ	Exec-managerial	Husband	White	Male	0	0	45	United-States	>50K	
9	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-States	>50K	
10	42	Private	159449	Bachelors	13	Married-civ	Exec-managerial	Husband	White	Male	5178	0	40	United-States	>50K	
11	37	Private	280464	Some-college	10	Married-civ	Exec-managerial	Husband	Black	Male	0	0	80	United-States	>50K	
12	30	State-gov	141297	Bachelors	13	Married-civ	Prof-specialty	Husband	Asian-Pac-Islander	Male	0	0	40	India	>50K	
13	23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0	0	30	United-States	<=50K	
14	32	Private	205019	Assoc-acad	12	Never-married	Sales	Not-in-family	Black	Male	0	0	50	United-States	<=50K	
15	40	Private	121772	Assoc-voc	11	Married-civ	Craft-repair	Husband	Asian-Pac-Islander	Male	0	0	40	?	>50K	
16	34	Private	245487	7th-8th	4	Married-civ	Transportation	Husband	Amer-Indian-Alaska	Male	0	0	45	Mexico	<=50K	
17	25	Self-employed	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0	0	35	United-States	<=50K	
18	32	Private	186824	HS-grad	9	Never-married	Machine-oper	Unmarried	White	Male	0	0	40	United-States	<=50K	
19	38	Private	28887	11th	7	Married-civ	Sales	Husband	White	Male	0	0	50	United-States	<=50K	
20	43	Self-employed	292175	Masters	14	Divorced	Exec-managerial	Unmarried	White	Female	0	0	45	United-States	>50K	
21	40	Private	193524	Doctorate	16	Married-civ	Prof-specialty	Husband	White	Male	0	0	60	United-States	>50K	
22	54	Private	302146	HS-grad	9	Separated	Other-service	Unmarried	Black	Female	0	0	20	United-States	<=50K	
23	35	Federal-gov	76845	9th	5	Married-civ	Farming-fishing	Husband	Black	Male	0	0	40	United-States	<=50K	
24	43	Private	117037	11th	7	Married-civ	Transportation	Husband	White	Male	0	2042	40	United-States	<=50K	
25	59	Private	109015	HS-grad	9	Divorced	Tech-support	Unmarried	White	Female	0	0	40	United-States	<=50K	
26	56	Local-gov	216851	Bachelors	13	Married-civ	Tech-support	Husband	White	Male	0	0	40	United-States	>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

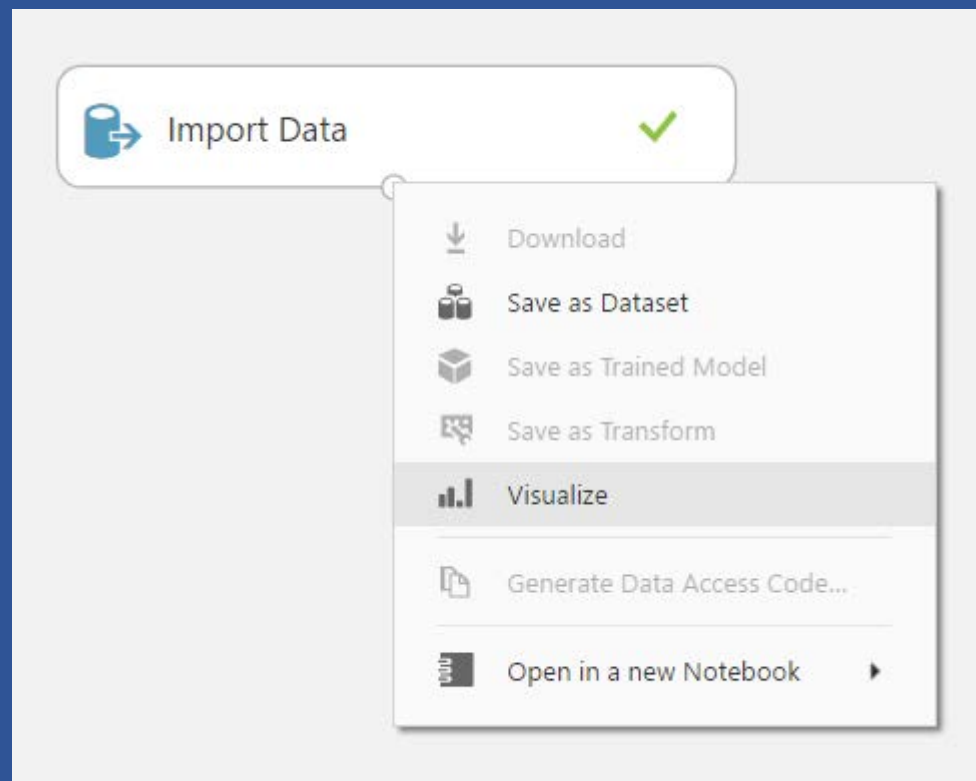
The screenshot shows the 'Import Data' module configuration in the Azure Machine Learning interface. The 'Data source' is set to 'Web URL via HTTP'. The 'Data source URL' field contains the URL 'http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data'. The 'Data format' is set to 'CSV'. There are checkboxes for 'CSV or TSV has header row' and 'Use cached results', both of which are currently unchecked. At the bottom, there is a table showing the execution status: START TIME (5/20/2017 9:13:05 PM), END TIME (5/20/2017 9:13:16 PM), ELAPSED TIME (0:00:11.502), STATUS CODE (Finished), and STATUS DETAILS (None).

Properties	
Project	
▲ Import Data	
Launch Import Data Wizard	
Data source	Web URL via HTTP ▼
Data source URL	http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data
Data format	CSV ▼
<input type="checkbox"/> CSV or TSV has header row	
<input type="checkbox"/> Use cached results	
START TIME	5/20/2017 9:13:05 PM
END TIME	5/20/2017 9:13:16 PM
ELAPSED TIME	0:00:11.502
STATUS CODE	Finished
STATUS DETAILS	None

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

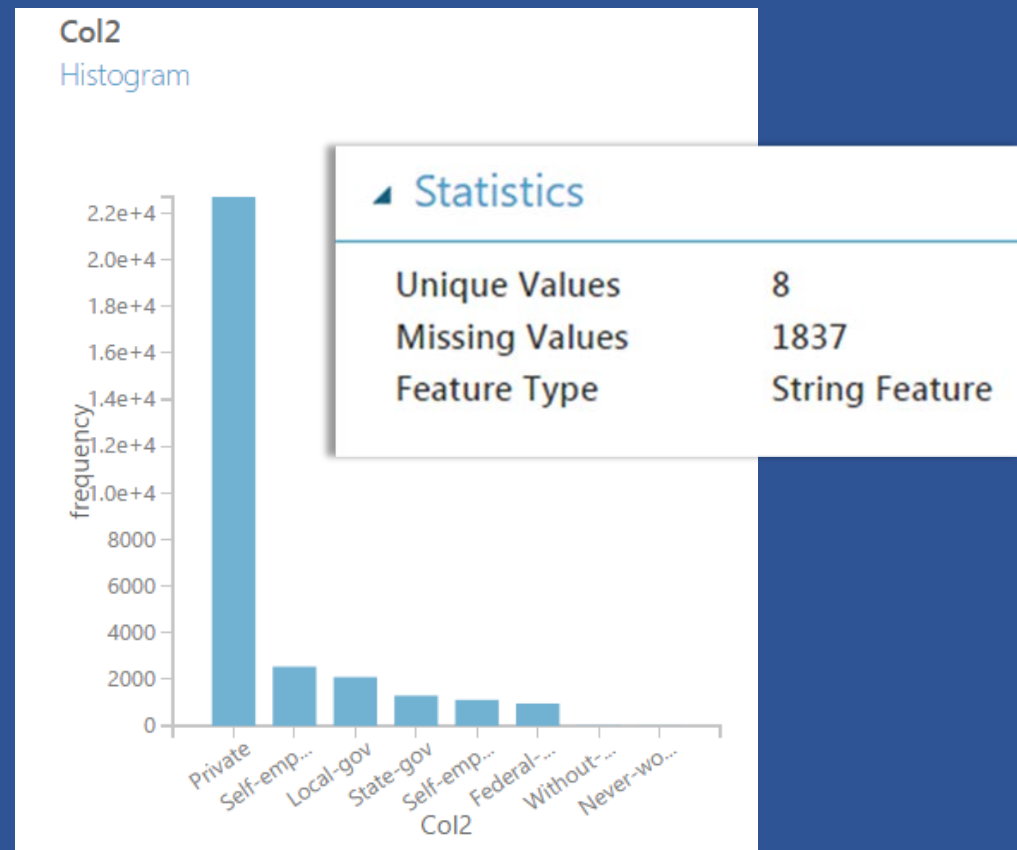
- Click on Col2
- Look at Statistics and Histogram

Adult Income ▶ Import Data ▶ Results dataset

rows 32562 columns 15

view as  

	Col1	Col2	Col3	Col4	Col5	Col6
39						
	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

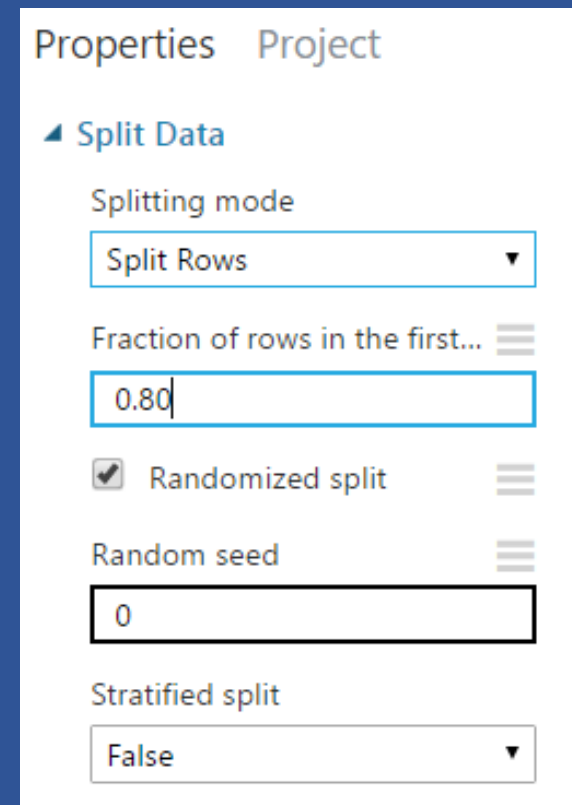
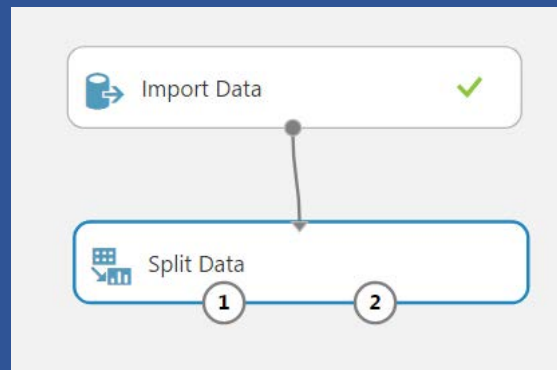
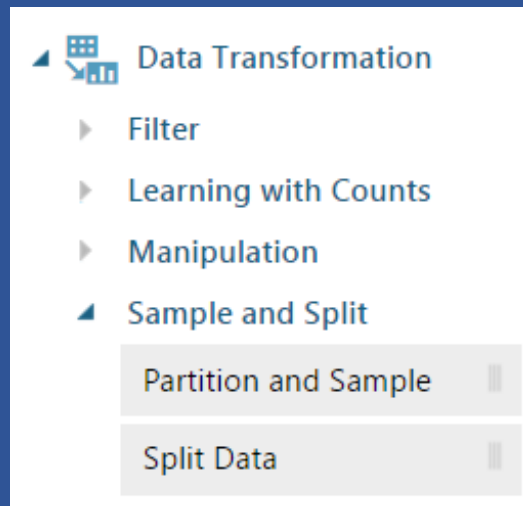
- **Training data** This grouping is used for creating our new predictive model based on the inherent patterns found in the historical data via the ML algorithm we use for the solution.
- **Validation data** This grouping is used for testing the new predictive model against known outcomes to determine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

First Experimental

Train, Test, Evaluate for Binary Classification

Add Split Data:

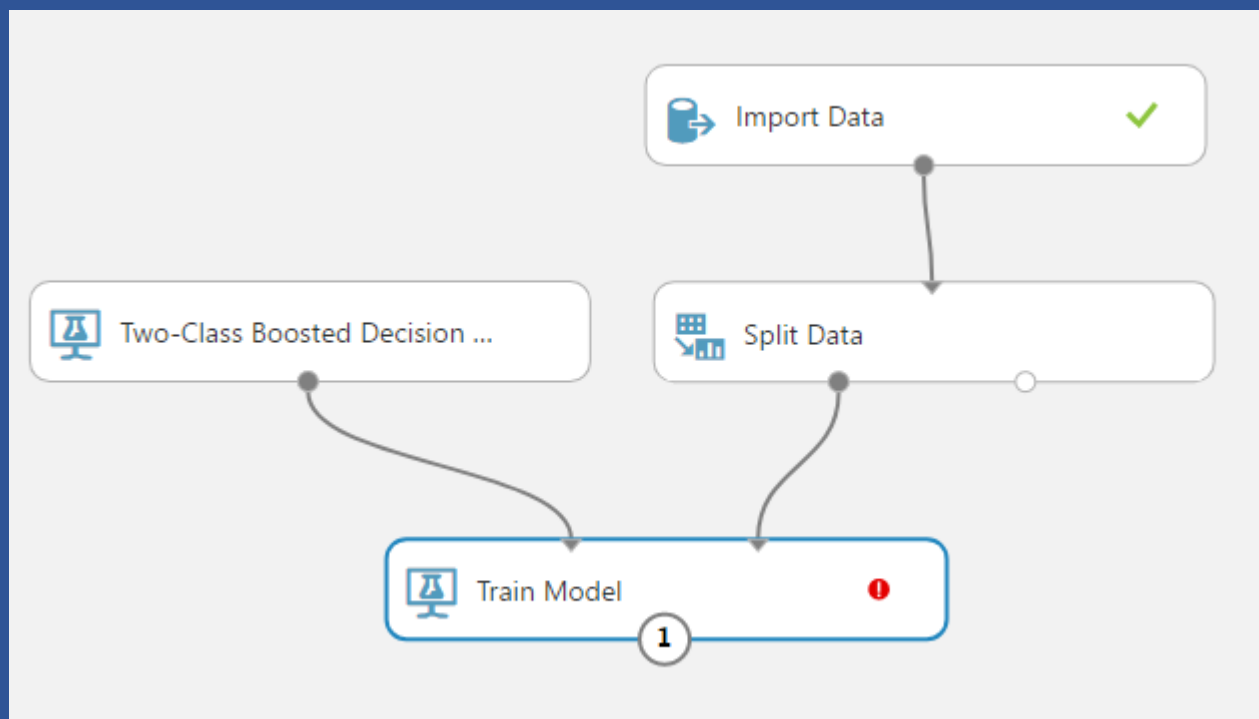
- Click **Data Transformation**
- Click **Sample and Split**
- Drag & drop **Split Data** module into canvas
- Connect **Import Data** to **Split Data**
- Set properties Fraction of row to 0.80



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**



Two-Class Boosted Decision T..

Create trainer mode
Single Parameter ▼

Maximum number of le...
20

Minimum number of sa...
10


Learning rate
0.2

Number of trees constru..
100

Random number seed

☒ Allow unknown cate...

First Experimental Label column specification

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

Select a single column

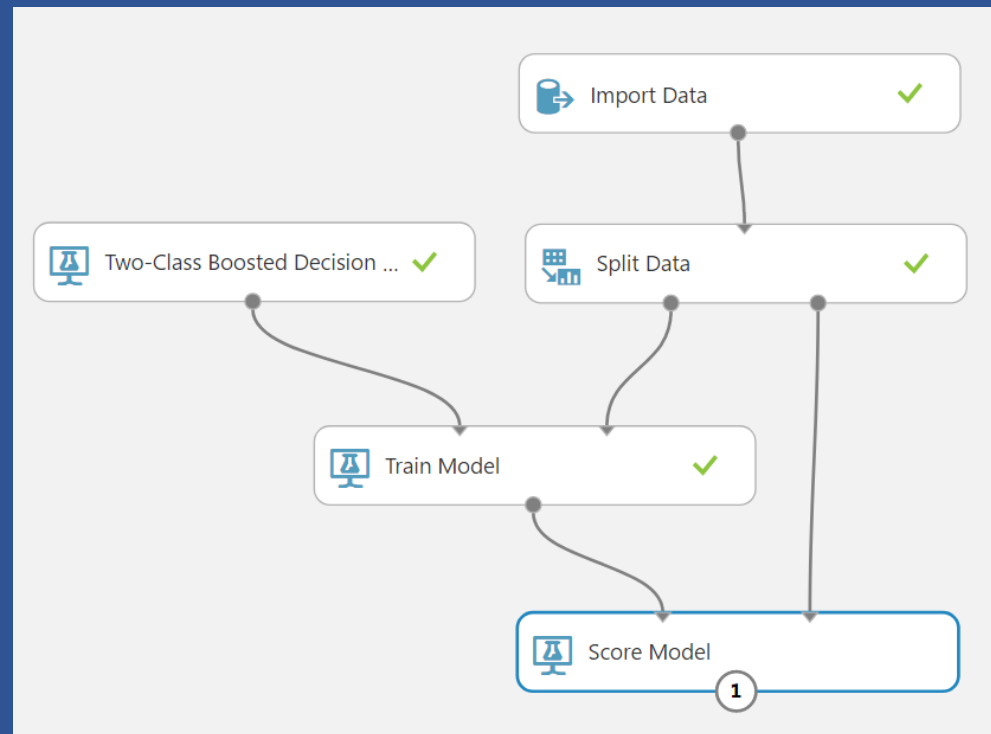
BY NAME	Include ▼	column names ▼	Col15 ✕
WITH RULES			

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>