

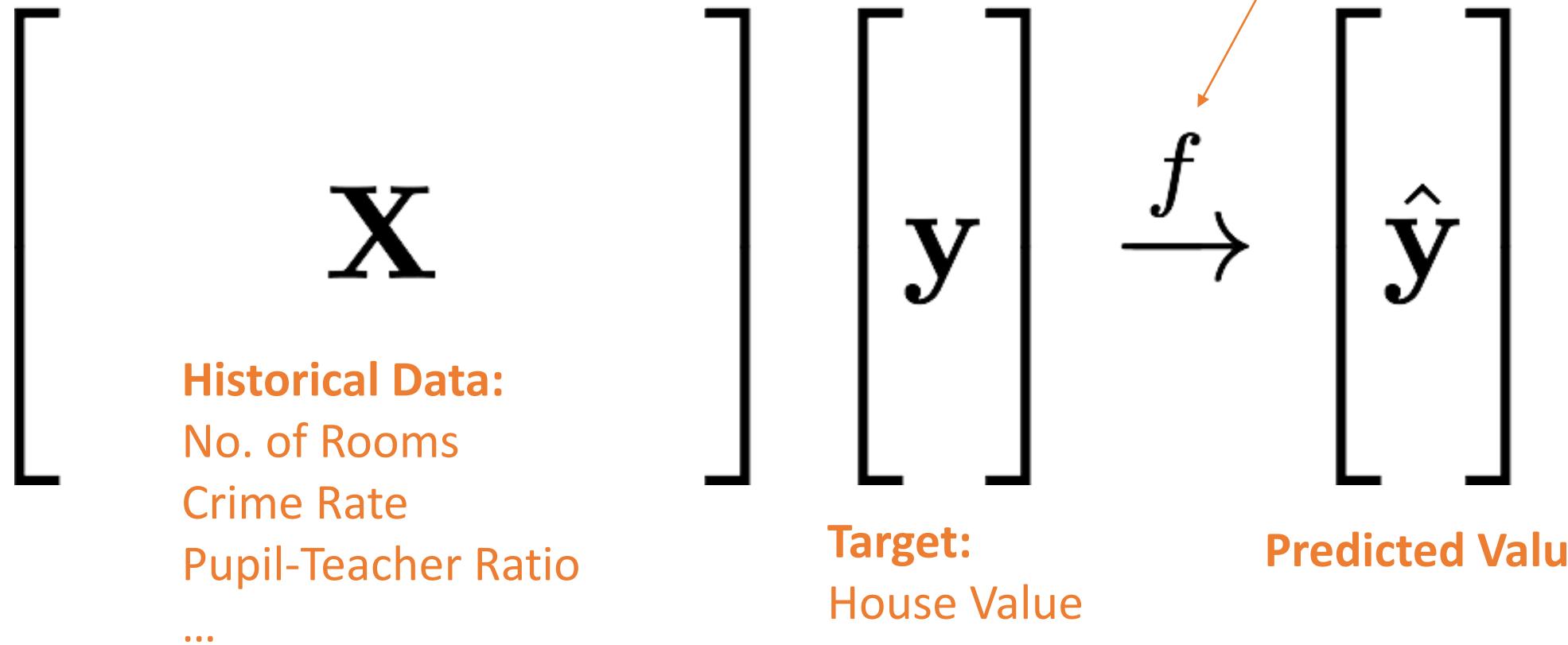
# Regression Example

- **Data:** Boston Housing (1978)
- **Source:** <http://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html>



# Supervised Learning Example

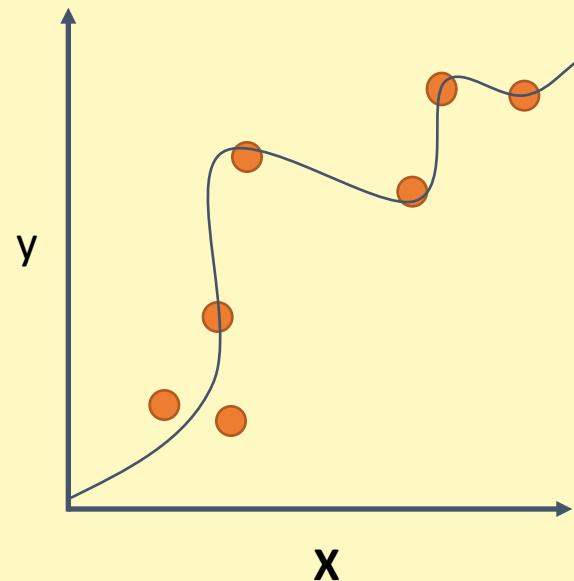
Machine Learning:  
Learn Patterns  
from Data



# Supervised Learning – You Already Have Target Data

**Regression:**

**How much will a customers spend?**

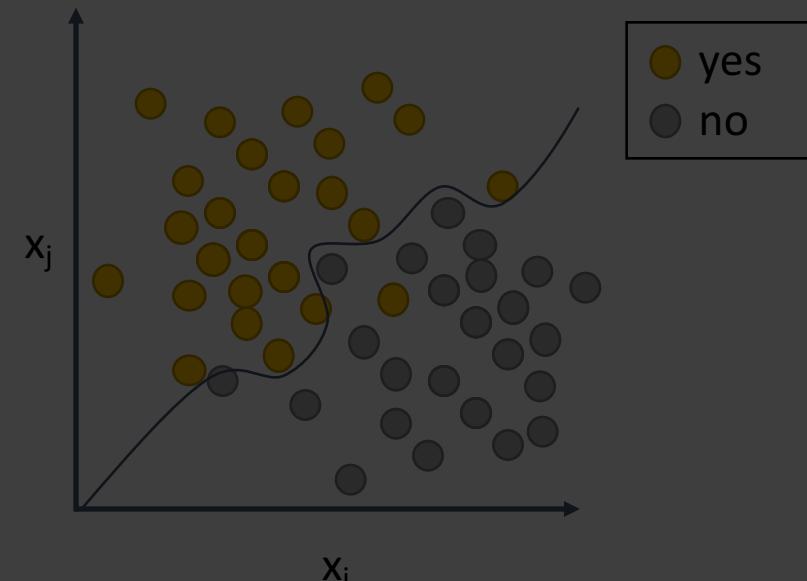


**H<sub>2</sub>O algos:**

**Penalized Linear Models**  
**Random Forest**  
**Gradient Boosting**  
**Neural Networks**  
**Stacked Ensembles**

**Classification:**

**Will a customer make a purchase? Yes or No**



**H<sub>2</sub>O algos:**

**Penalized Linear Models**  
**Naïve Bayes**  
**Random Forest**  
**Gradient Boosting**  
**Neural Networks**  
**Stacked Ensembles**

# Regression – Key Steps

- Import Data (CSV Files)
  - ./data/regression/ ...
  - Have a quick look
- Train a Random Forest Model
  - Look at variable importance
- Make Predictions
  - Compare with ground truth
- Build Partial Dependence Plots
  - Explain model behaviour



1. CRIM - per capita crime rate by town
2. ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
3. INDUS - proportion of non-retail business acres per town.
4. CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
5. NOX - nitric oxides concentration (parts per 10 million)
6. RM - average number of rooms per dwelling
7. AGE - proportion of owner-occupied units built prior to 1940
8. DIS - weighted distances to five Boston employment centres
9. RAD - index of accessibility to radial highways
10. TAX - full-value property-tax rate per \$10,000
11. PTRATIO - pupil-teacher ratio by town
12. B -  $1000(Bk - 0.63)^2$  where Bk is the proportion of blacks by town
13. LSTAT - % lower status of the population
14. MEDV - Median value of owner-occupied homes in \$1000's

**Target:**  
**House Value**

..../regression/  
house\_price\_train.csv

The screenshot shows a Microsoft Excel spreadsheet titled "house\_price\_train". The table has 408 rows and 15 columns. The columns are labeled A through M and N. Column N is highlighted in green. The last column, O, is empty. The data includes various numerical values such as crime rates, property characteristics, and house prices.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	b	lstat	medv	
2	0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.9	9.14	21.6	
3	0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7	
4	0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4	
5	0.06905	0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.9	5.33	36.2	
6	0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605	5	311	15.2	395.6	12.43	22.9	
7	0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311	15.2	396.9	19.15	27.1	
8	0.21124	12.5	7.87	0	0.524	5.631	100	6.0821	5	311	15.2	386.63	29.93	16.5	
9	0.17004	12.5	7.87	0	0.524	6.004	85.9	6.5921	5	311	15.2	386.71	17.1	18.9	
10	0.22489	12.5	7.87	0	0.524	6.377	94.3	6.3467	5	311	15.2	392.52	20.45	15	
11	0.11747	12.5	7.87	0	0.524	6.009	82.9	6.2267	5	311	15.2	396.9	13.27	18.9	
12	0.09378	12.5	7.87	0	0.524	5.889	39	5.4509	5	311	15.2	390.5	15.71	21.7	
13	0.62976	0	8.14	0	0.538	5.949	61.8	4.7075	4	307	21	396.9	8.26	20.4	
14	0.63796	0	8.14	0	0.538	6.096	84.5	4.4619	4	307	21	380.02	10.26	18.2	
15	0.62739	0	8.14	0	0.538	5.834	56.5	4.4986	4	307	21	395.62	8.47	19.9	
16	1.05393	0	8.14	0	0.538	5.935	29.3	4.4986	4	307	21	386.85	6.58	23.1	
17	0.7842	0	8.14	0	0.538	5.99	81.7	4.2579	4	307	21	386.75	14.67	17.5	
18	0.80271	0	8.14	0	0.538	5.456	36.6	3.7965	4	307	21	288.99	11.69	20.2	
19	0.7258	0	8.14	0	0.538	5.727	69.5	3.7965	4	307	21	390.95	11.28	18.2	
20	1.25179	0	8.14	0	0.538	5.57	98.1	3.7979	4	307	21	376.57	21.02	13.6	
21	0.85204	0	8.14	0	0.538	5.965	89.2	4.0123	4	307	21	392.53	13.83	19.6	
22	0.75026	0	8.14	0	0.538	5.924	94.1	4.3996	4	307	21	394.33	16.3	15.6	
23	0.84054	0	8.14	0	0.538	5.599	85.7	4.4546	4	307	21	303.42	16.51	13.9	
24	0.67191	0	8.14	0	0.538	5.813	90.3	4.682	4	307	21	376.88	14.81	16.6	
25	0.95577	0	8.14	0	0.538	6.047	88.8	4.4534	4	307	21	306.38	17.28	14.8	
26	0.77299	0	8.14	0	0.538	6.495	94.4	4.4547	4	307	21	387.94	12.8	18.4	
27	1.00245	0	8.14	0	0.538	6.674	87.3	4.239	4	307	21	380.23	11.98	21	

The screenshot shows the H2O Flow web application interface. At the top, there is a navigation bar with tabs for "Data", "Model", "Score", "Admin", and "Help". The "Data" tab is currently active and highlighted in yellow. A context menu is open under the "Data" tab, listing options: "Import Files...", "Upload File...", "Split Frame...", "Merge Frames...", "List All Frames", and "Impute...". The "Upload File..." option is highlighted with a gray background. On the left side of the interface, there is a sidebar titled "Assistance" which lists various H2O routines with their descriptions. A yellow callout box with the text "Data → Upload File ..." is overlaid on the right side of the interface, pointing towards the "Upload File..." menu item. The bottom status bar shows the URL "localhost:54321/flow/index.html#" and connection information "Connections: 0".

Untitled Flow

H2O FLOW

Data ▾ Model ▾ Score ▾ Admin ▾ Help ▾

Import Files...

Upload File...

Split Frame...

Merge Frames...

List All Frames

Impute...

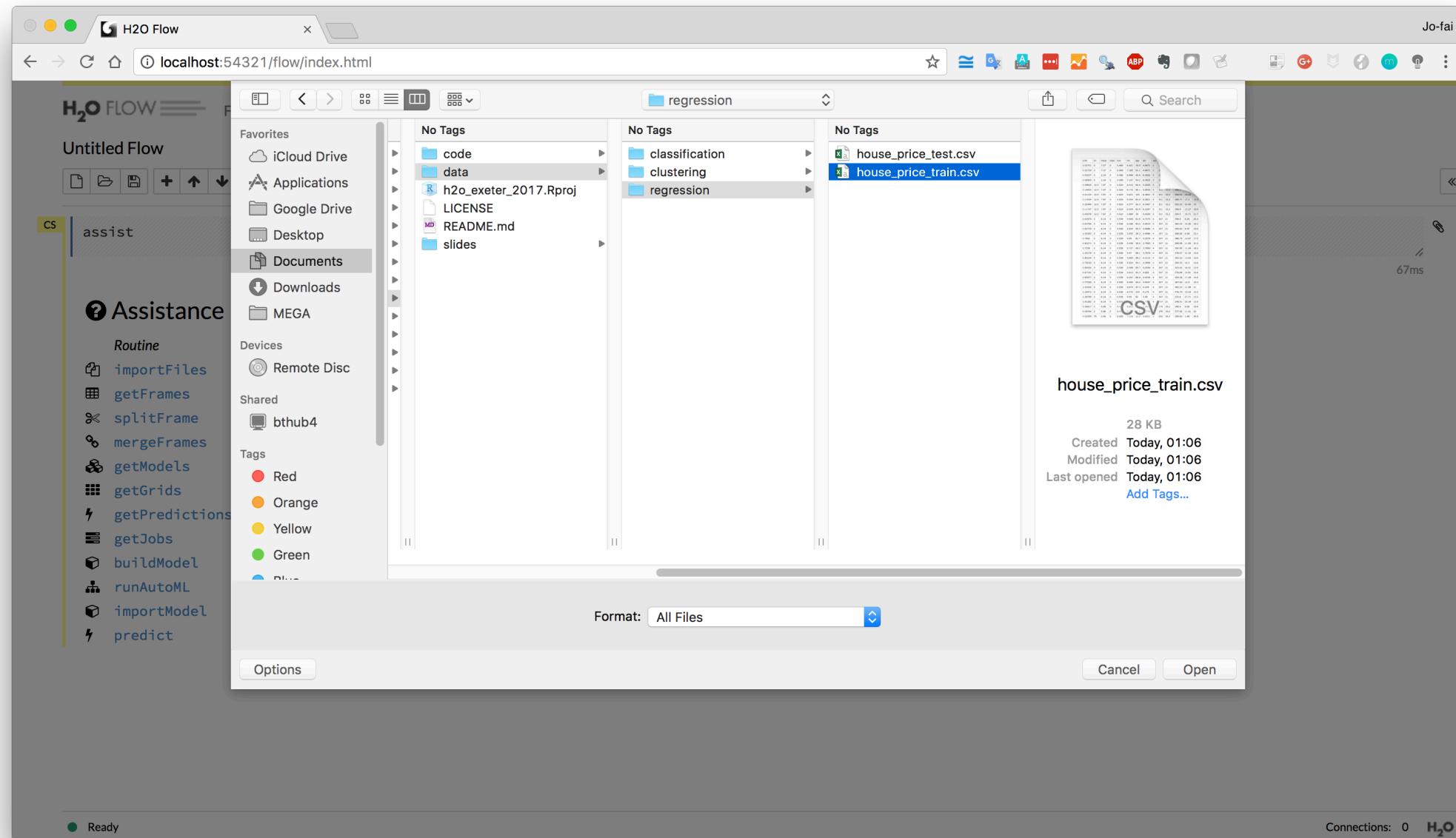
?

Assistance

Routine	Description
importFiles	Import file(s) into H <sub>2</sub> O
getFrames	Get a list of frames in H <sub>2</sub> O
splitFrame	Split a frame into two or more frames
mergeFrames	Merge two frames into one
getModels	Get a list of models in H <sub>2</sub> O
getGrids	Get a list of grid search results in H <sub>2</sub> O
getPredictions	Get a list of predictions in H <sub>2</sub> O
getJobs	Get a list of jobs running in H <sub>2</sub> O
buildModel	Build a model
runAutoML	Automatically train and tune many models
importModel	Import a saved model
predict	Make a prediction

localhost:54321/flow/index.html#

Connections: 0



**H<sub>2</sub>O FLOW**  Flow ▾ Cell ▾ Data ▾ Model ▾ Score ▾ Admin ▾ Help ▾

Untitled Flow

assit

?

**Assistance**

Routine	Description
 importFiles	Import file(s) into H <sub>2</sub> O
 getFrames	Get a list of frames in H <sub>2</sub> O
 splitFrame	Split a frame into two or more frames
 mergeFrames	Merge two frames into one
 getModels	Get a list of models in H <sub>2</sub> O
 getGrids	Get a list of grid search results in H <sub>2</sub> O
 getPredictions	Get a list of predictions in H <sub>2</sub> O
 getJobs	Get a list of jobs running in H <sub>2</sub> O
 buildModel	Build a model
 runAutoML	Automatically train and tune many models
 importModel	Import a saved model
 predict	Make a prediction

Upload Dataset... 

Choose File **house\_price\_train.csv**

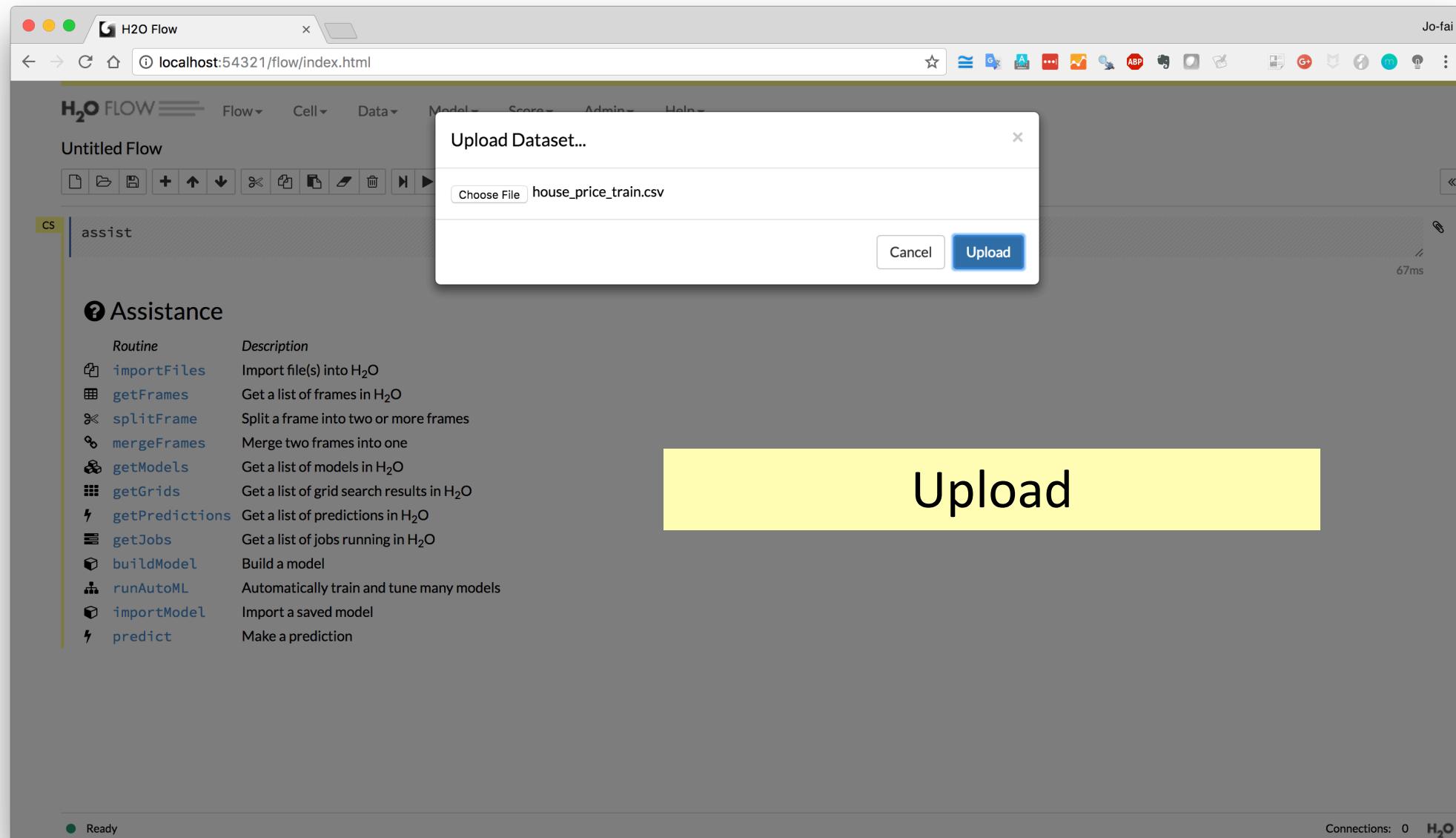
Cancel **Upload**

67ms

Connections: 0 

Ready

Upload



H2O Flow Jo-fai

localhost:54321/flow/index.html

Untitled Flow

Flow Cell Data Model Score Admin Help

Setup Parse

PARSE CONFIGURATION

Sources house\_price\_train.csv  
ID Key\_Frame\_house\_price\_train.hex  
Parser CSV  
Separator ;'44'  
Column Headers  Auto  First row contains column names  First row contains data  
Options  Enable single quotes as a field quotation character  Delete on done

EDIT COLUMN NAMES AND TYPES

Search by column name...

	crim	Numeric	0.02731	0.02729	0.03237	0.06905	0.08829	0.14455	0.21124	0.17004	0.22489
1	zn	Numeric	0	0	0	0	12.5	12.5	12.5	12.5	12.5
2	indus	Numeric	7.07	7.07	2.18	2.18	7.87	7.87	7.87	7.87	7.87
3	chas	Numeric	0	0	0	0	0	0	0	0	0
4	nox	Numeric	0.469	0.469	0.458	0.458	0.524	0.524	0.524	0.524	0.524
5	rm	Numeric	6.421	7.185	6.998	7.147	6.012	6.172	5.631	6.004	6.377
6	age	Numeric	78.9	61.1	45.8	54.2	66.6	96.1	100	85.9	94.3
7	dis	Numeric	4.9671	4.9671	6.0622	6.0622	5.5605	5.9505	6.0821	6.5921	6.3467

Ready Connections: 0 H2O

Scroll down

H<sub>2</sub>O Flow Jo-fai

localhost:54321/flow/index.html

H<sub>2</sub>O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

Delete on done

EDIT COLUMN NAMES AND TYPES

Search by column name...

	Column Name	Type	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7	Value 8	Value 9
1	crim	Numeric	0.02731	0.02729	0.03237	0.06905	0.08829	0.14455	0.21124	0.17004	0.22489
2	zn	Numeric	0	0	0	0	12.5	12.5	12.5	12.5	12.5
3	indus	Numeric	7.07	7.07	2.18	2.18	7.87	7.87	7.87	7.87	7.87
4	chas	Numeric	0	0	0	0	0	0	0	0	0
5	nox	Numeric	0.469	0.469	0.458	0.458	0.524	0.524	0.524	0.524	0.524
6	rm	Numeric	6.421	7.185	6.998	7.147	6.012	6.172	5.631	6.004	6.377
7	age	Numeric	78.9	61.1	45.8	54.2	66.6	96.1	100	85.9	94.3
8	dis	Numeric	4.9671	4.9671	6.0622	6.0622	5.5605	5.9505	6.0821	6.5921	6.3467
9	rad	Numeric	2	2	3	3	5	5	5	5	5
10	tax	Numeric	242	242	222	222	311	311	311	311	311
11	ptratio	Numeric	17.8	17.8	18.7	18.7	15.2	15.2	15.2	15.2	15.2
12	b	Numeric	396.9	392.83	394.63	396.9	395.6	396.9	386.63	386.71	392.52
13	lstat	Numeric	9.14	4.03	2.94	5.33	12.43	19.15	29.93	17.1	20.45
14	medv	Numeric	21.6	34.7	33.4	36.2	22.9	27.1	16.5	18.9	15

Click "Parse"

Ready Connections: 0 H<sub>2</sub>O

H2O Flow Jo-fai

localhost:54321/flow/index.html

## Untitled Flow

Flow Cell Data Model Score Admin Help

File Edit Run View

CS parseFiles

```
source_frames: ["house_price_train.csv"]
destination_frame: "Key_Frame__house_price_train.hex"
parse_type: "CSV"
separator: 44
number_columns: 14
single_quotes: false
column_names: ["crim","zn","indus","chas","nox","rm","age","dis","rad","tax","ptratio","b","lstat","medv"]
column_types:
["Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric"]
delete_on_done: true
check_header: 1
chunk_size: 4194304
```

Job

Run Time 00:00:00.107  
Remaining Time 00:00:00.0

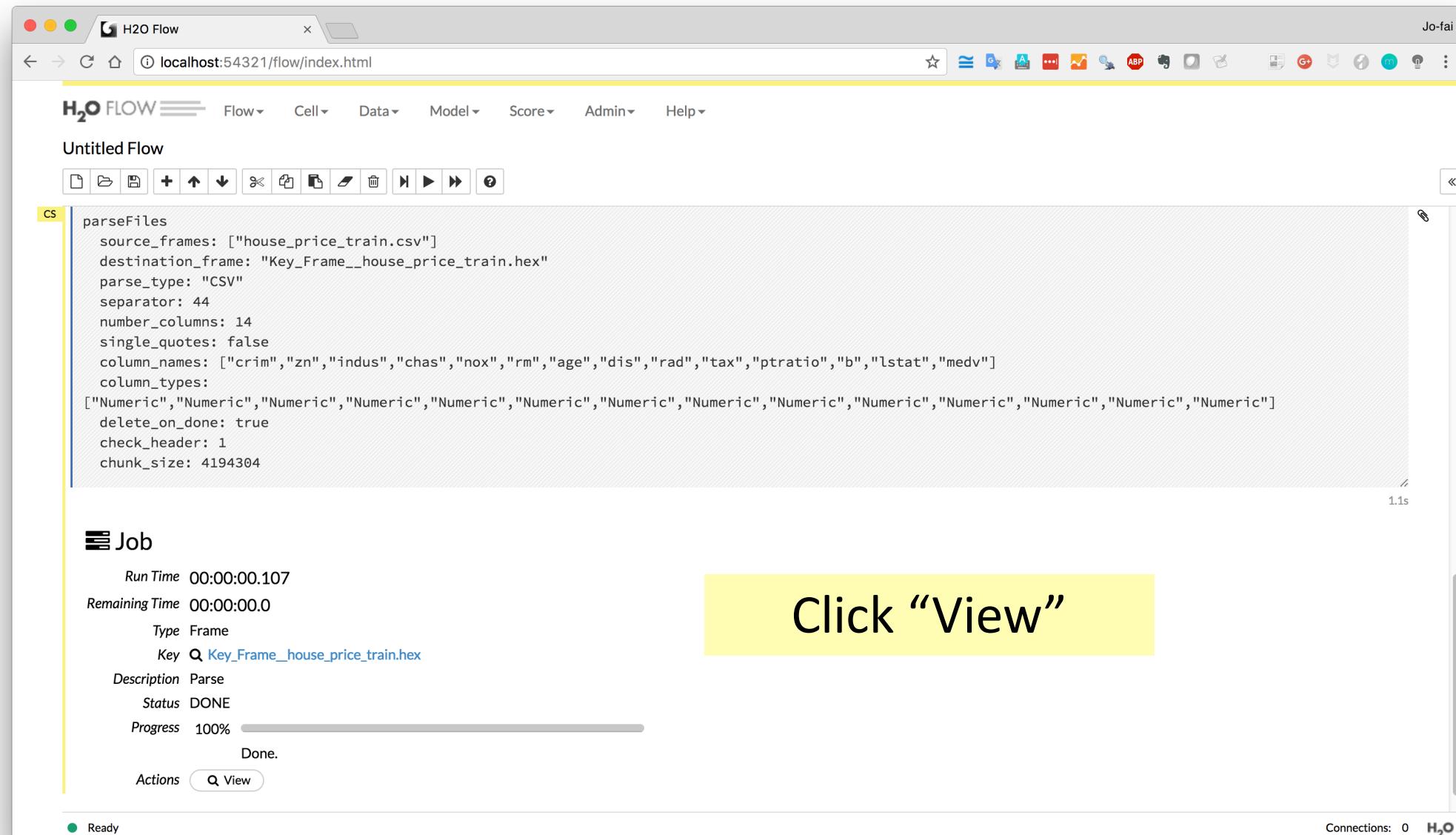
Type Frame  
Key [Q Key\\_Frame\\_house\\_price\\_train.hex](#)

Description Parse  
Status DONE  
Progress 100% Done.

Actions [View](#)

Click “View”

Ready Connections: 0 H2O



H2O Flow Jo-fai

localhost:54321/flow/index.html

Untitled Flow

getFrameSummary "Key\_Frame\_\_house\_price\_train.hex"

86ms

Key\_Frame\_\_house\_price\_train.hex

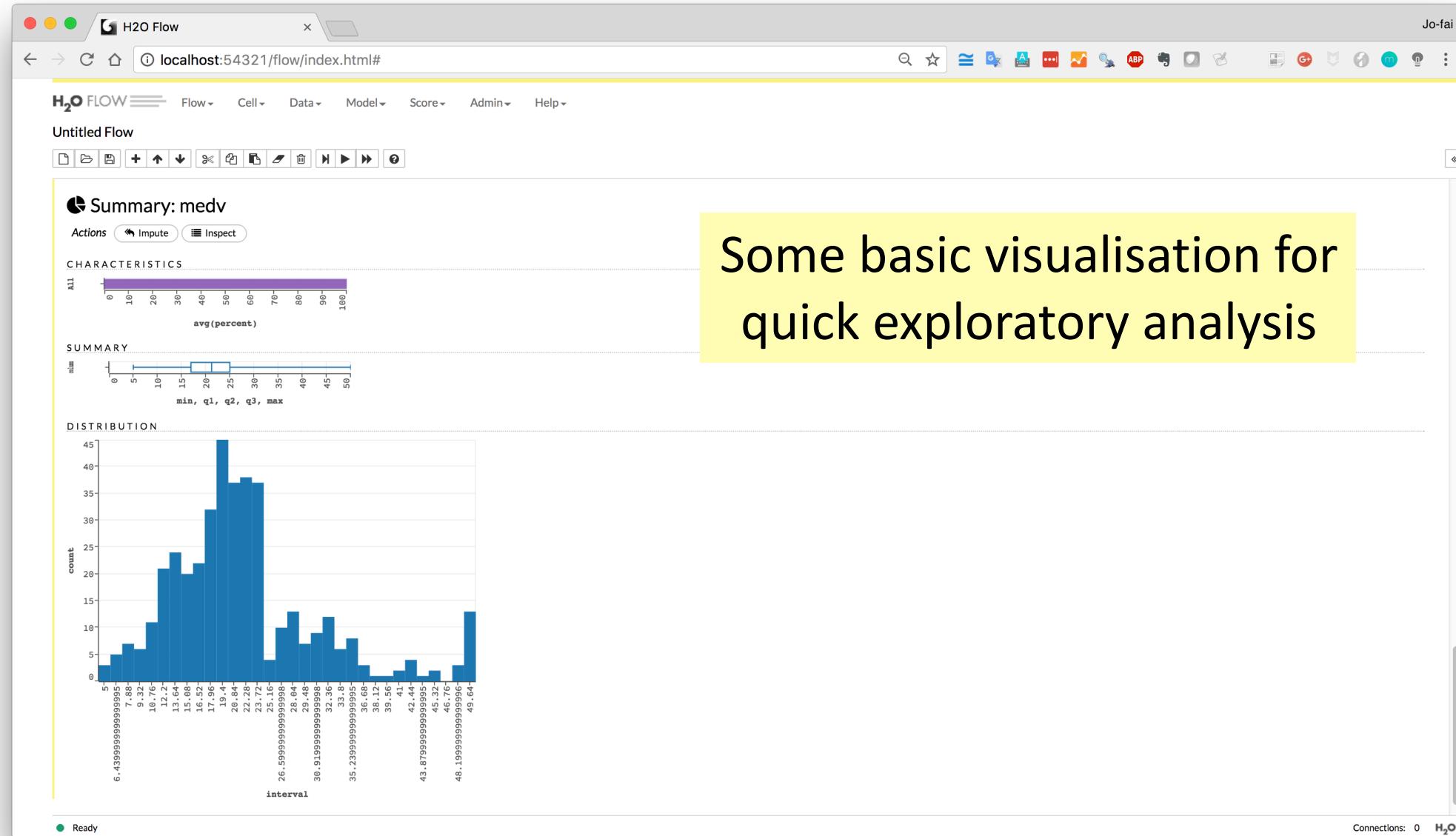
Actions: View Data Split... Build Model... Predict Download Export Delete

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
crim	real	0	0	0	0	0.0091	73.5341	3.5674	7.9480	...	
zn	real	0	301	0	0	0	100.0	10.5872	22.2598	...	
indus	real	0	0	0	0	0.4600	27.7400	11.4093	6.8145	...	
chas	int	0	379	0	0	0	1.0	0.0688	0.2534	...	Convert to enum
nox	real	0	0	0	0	0.3850	0.8710	0.5568	0.1156	...	
rm	real	0	0	0	0	3.8630	8.7250	6.2866	0.6909	...	
age	real	0	0	0	0	2.9000	100.0	69.3889	27.8179	...	
dis	real	0	0	0	0	1.1296	10.7103	3.7177	2.0152	...	
rad	int	0	0	0	0	1.0	24.0	9.8378	8.7844	...	Convert to enum
tax	int	0	0	0	0	188.0	711.0	412.3784	170.4474	...	Convert to enum
ptratio	real	0	0	0	0	12.6000	22.0	18.4474	2.1618	...	
b	real	0	0	0	0	0.3200	396.9000	354.4032	94.1752	...	
lstat	real	0	0	0	0	1.7300	37.9700	12.7920	7.0987	...	
medv	real	0	0	0	0	5.0	50.0	22.6248	9.1850	...	

Previous 20 Columns Next 20 Columns

Ready Connections: 0 H2O

Click on any variable  
e.g. crim, medv



Some basic visualisation for  
quick exploratory analysis

Now do the same for  
`house_price_test.csv`  
(e.g. new data for making  
predictions)

The screenshot shows the H2O Flow interface on a Mac OS X desktop. The main window displays a flow named "Untitled Flow" containing a step titled "Key\_Frame\_house\_price". The "COLUMN SUMMARIES" section shows the schema for the "house\_price" dataset, including columns like label, crim, zn, indus, chas, nox, rm, age, dis, rad, tax, ptratio, b, lstat, and medv. The "No Tags" section in the sidebar lists "classification", "clustering", and "regression", with "regression" currently selected. A file browser dialog is open, showing the file "house\_price\_test.csv" located in the "Documents" folder. The file details are: Size 7 KB, Created Today, 01:06, Modified Today, 01:06, Last opened Today, 01:06. Buttons for "Delete", "Cancel", and "Open" are visible at the bottom of the dialog.

Show all data  
Data → List All Frames

Untitled Flow

Type Frame  
Key **Q Key\_Frame\_house\_price\_hex**

Description Parse  
Status DONE  
Progress 100% Done.

Actions **Q View**

getFrames

29ms

Frames

	Rows	Columns	Size
<b>Key_Frame_house_price_hex</b>	99	14	4KB
<b>Key_Frame_house_price_train.hex</b>	407	14	12KB

**Q Predict on selected frames...** **Q Delete selected frames**

localhost:54321/flow/index.html#

Connections: 0 **H2O**

The screenshot shows the H2O Flow web application running on a Mac OS X system. The browser window title is "H2O Flow" and the address bar shows "localhost:54321/flow/index.html#". The main interface has a top navigation bar with "Model" (highlighted in yellow), "Score", "Admin", and "Help" dropdowns, and a toolbar with various icons.

The central area displays an "Untitled Flow" with the following details:

- Type: Frame
- Key: [Key\\_Frame\\_house\\_price\\_test.hex](#)
- Description: Parse
- Status: DONE
- Progress: 100%
- Actions: Done, View

A modal menu is open under the "Model" dropdown, listing various machine learning models:

- Aggregator...
- Deep Learning...
- Distributed Random Forest...
- Gradient Boosting Machine...
- Generalized Linear Modeling...
- Generalized Low Rank Modeling...
- K-means...
- Naive Bayes...
- Principal Components Analysis...
- Stacked Ensemble...
- Word2Vec...
- XGBoost...

The "Frames" section lists two frames:

- getFrames**:
  - Type: ID
  - Key: [Key\\_Frame\\_house\\_price\\_test.hex](#)
    - Build Model...
    - Predict...
    - Inspect
- Frames**:
  - Type: ID
  - Key: [Key\\_Frame\\_house\\_price\\_train.hex](#)
    - Build Model...
    - Predict...
    - Inspect

At the bottom, there are buttons for "Predict on selected frames..." and "Delete selected frames".

On the right side, there is a yellow callout box with the text:

Train a predictive model  
Start with Random Forest

Below the callout, there is a table showing frame statistics:

Rows	Columns	Size
99	14	4KB
407	14	12KB

The status bar at the bottom shows "Connections: 0" and the H2O logo.

H2O Flow Jo-fai

localhost:54321/flow/index.html#

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

Build a Model

Select an algorithm: Distributed Random Forest

PARAMETERS

model\_id my\_random\_forest  
training\_frame Key\_Frame\_house\_price\_train.hex  
validation\_frame (Choose...)  
nfolds 0  
response\_column medv  
ignored\_columns Search...

GRID ?

Destination id for this model; auto-generated if not specified.

Id of the training data frame (Not required, to allow initial validation of model parameters).

Id of the validation data frame.

Number of folds for N-fold cross-validation (0 to disable or >= 2).

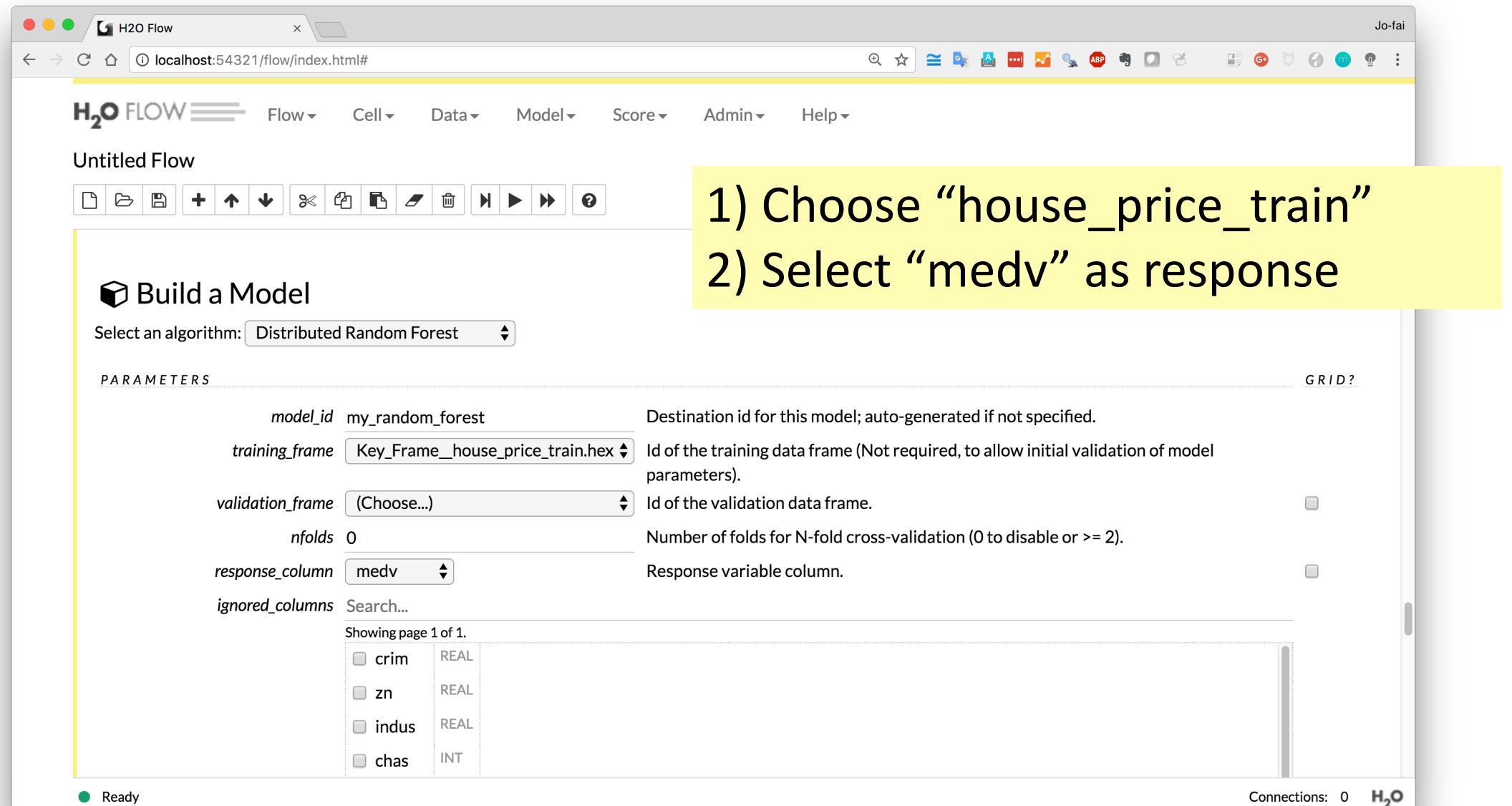
Response variable column.

Showing page 1 of 1.

crim	REAL
zn	REAL
indus	REAL
chas	INT

Connections: 0 H2O

1) Choose “house\_price\_train”  
2) Select “medv” as response



Untitled Flow

Flow ▾ Cell ▾ Data ▾ Model ▾ Score ▾ Admin ▾ Help ▾

Only show columns with more than  % missing values.

All  None

ignore\_const\_cols  Ignore constant columns.

ntrees 50 Number of trees.

max\_depth 20 Maximum tree depth.

min\_rows 1 Fewest allowed (weighted) observations in a leaf.

nbins 20 For numerical columns (real/int), build a histogram of (at least) this many bins, then split at the best point

seed 54321 Seed for pseudo random number generator (if applicable)

mtries -1 Number of variables randomly sampled as candidates at each split. If set to -1, defaults to  $\sqrt{p}$  for classification and  $p/3$  for regression (where  $p$  is the # of predictors)

sample\_rate 0.6320000290870667 Row sample rate per tree (from 0.0 to 1.0)

ADVANCED GRID ?

score\_each\_iteration  Whether to score during each iteration of model training.

score\_tree\_interval 0 Score the model after every so many trees. Disabled if set to 0.

fold\_column (Choose...) Column with cross-validation fold index assignment per observation.

offset\_column (Choose...) Offset column. This will be added to the combination of columns before applying the link function.

weights\_column (Choose...) Column with observation weights. Giving some observation a weight of zero is equivalent to excluding it from the dataset; giving an observation a relative weight of 2 is equivalent to repeating that row twice. Negative weights are not allowed.

nbins\_top\_level 1024 For numerical columns (real/int), build a histogram of (at most) this many bins at the root level, then decrease by factor of two per level

nbins\_cats 1024 For categorical columns (factors), build a histogram of this many bins, then split at the best point. Higher values can lead to more overfitting.

r2\_stopping 1.7976931348623157e+308 r2\_stopping is no longer supported and will be ignored if set - please use stopping\_rounds, stopping\_metric and stopping\_tolerance

Ready Connections: 0 H<sub>2</sub>O

Enter a seed number  
(if you want to reproduce  
your results in future)

Untitled Flow

Flow ▾ Cell ▾ Data ▾ Model ▾ Score ▾ Admin ▾ Help ▾

histogram\_type: AUTO What type of histogram to use for finding optimal split points

categorical\_encoding: AUTO Encoding scheme for categorical features

**EXPERT**

build\_tree\_one\_node:  Run on one node only; no network overhead but fewer cpus used. Suitable for small datasets.

sample\_rate\_per\_class:  A list of row sample rates per class (relative fraction for each class, from 0.0 to 1.0), for each tree

binomial\_double\_trees:  For binary classification: Build 2x as many trees (one per class) - can lead to higher accuracy.

col\_sample\_rate\_change\_per\_level: 1 Relative change of the column sampling rate for every level (from 0.0 to 2.0)

calibrate\_model:  Use Platt Scaling to calculate calibrated class probabilities. Calibration can provide more accurate estimates of class probabilities.

calibration\_frame: (Choose...) Calibration frame for Platt Scaling

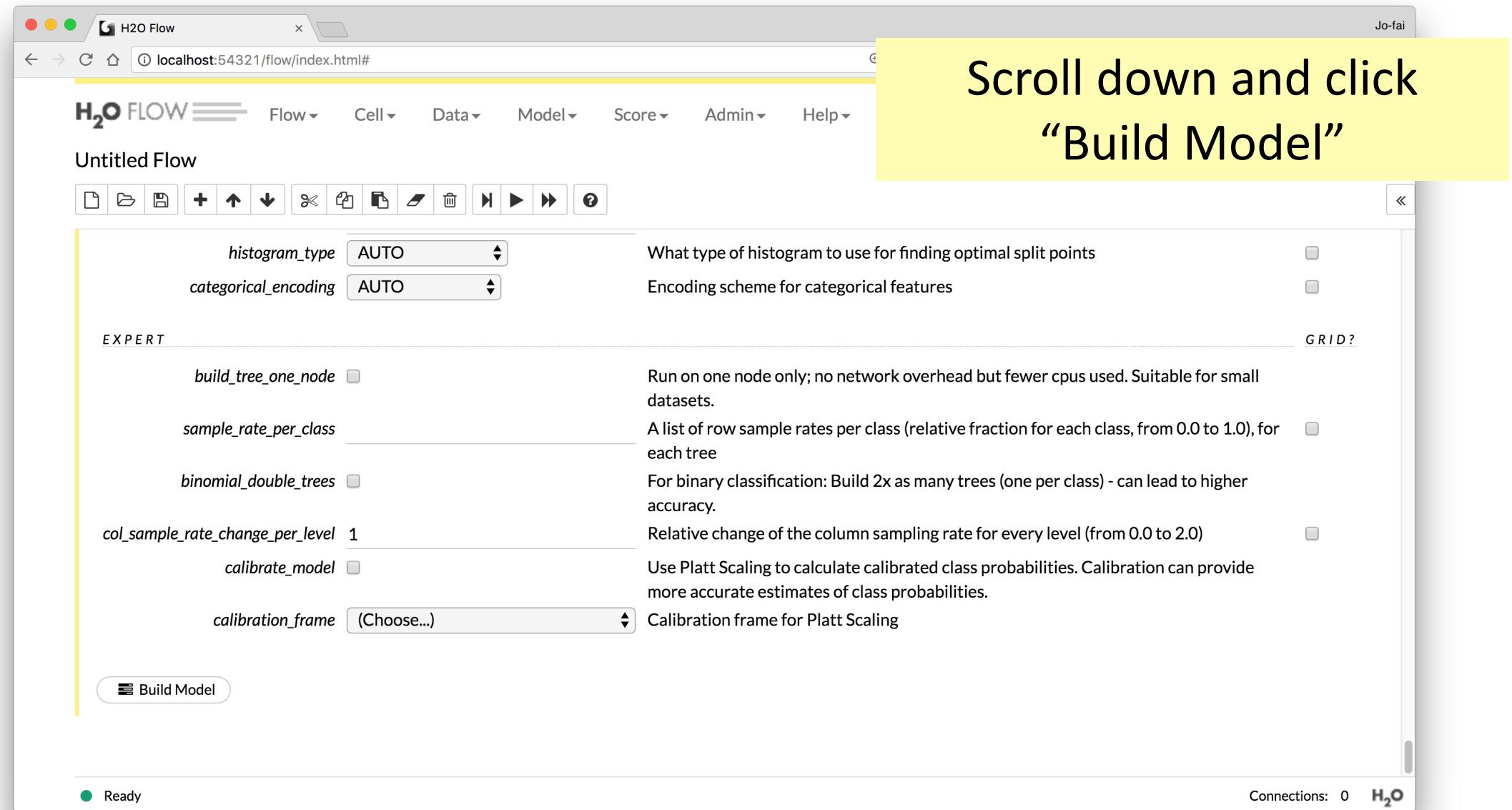
**Build Model**

Ready

Connections: 0 H<sub>2</sub>O

Jo-fai

Scroll down and click “Build Model”



H2O Flow Jo-fai

localhost:54321/flow/index.html#

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

Build Model

```
buildModel 'drf', {"model_id":"my_random_forest","training_frame":"Key_Frame__house_price_train.hex","nfolds":0,"response_column":"medv","ignored_columns":[],"ignore_const_cols":true,"ntrees":50,"max_depth":20,"min_rows":1,"nbins":20,"seed":54321,"mtries":-1,"sample_rate":0.6320000290870667,"score_each_iteration":false,"score_tree_interval":0,"nbins_top_level":1024,"nbins_cats":1024,"r2_stopping":1.7976931348623157e+308,"stopping_rounds":0,"stopping_metric":"AUTO","stopping_tolerance":0.001,"max_runtime_secs":0,"checkpoint":"","col_sample_rate_per_tree":1,"min_split_improvement":0.00001,"histogram_type":"AUTO","categorical_encoding":"AUTO","build_tree_one_node":false,"sample_rate_per_class":[],"binomial_double_trees":false,"col_sample_rate_change_per_level":1,"calibrate_model":false}
```

1.1s

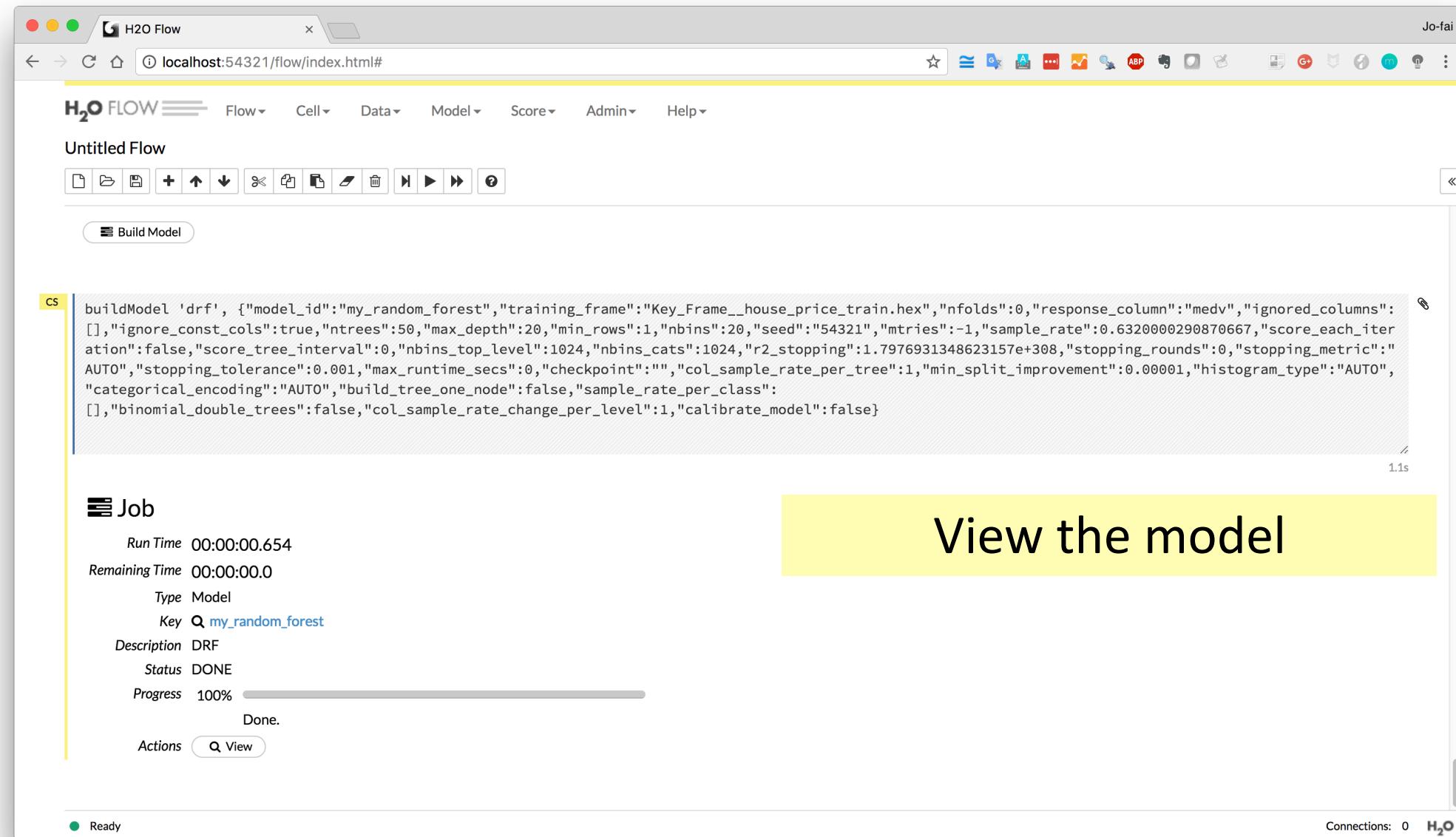
Job

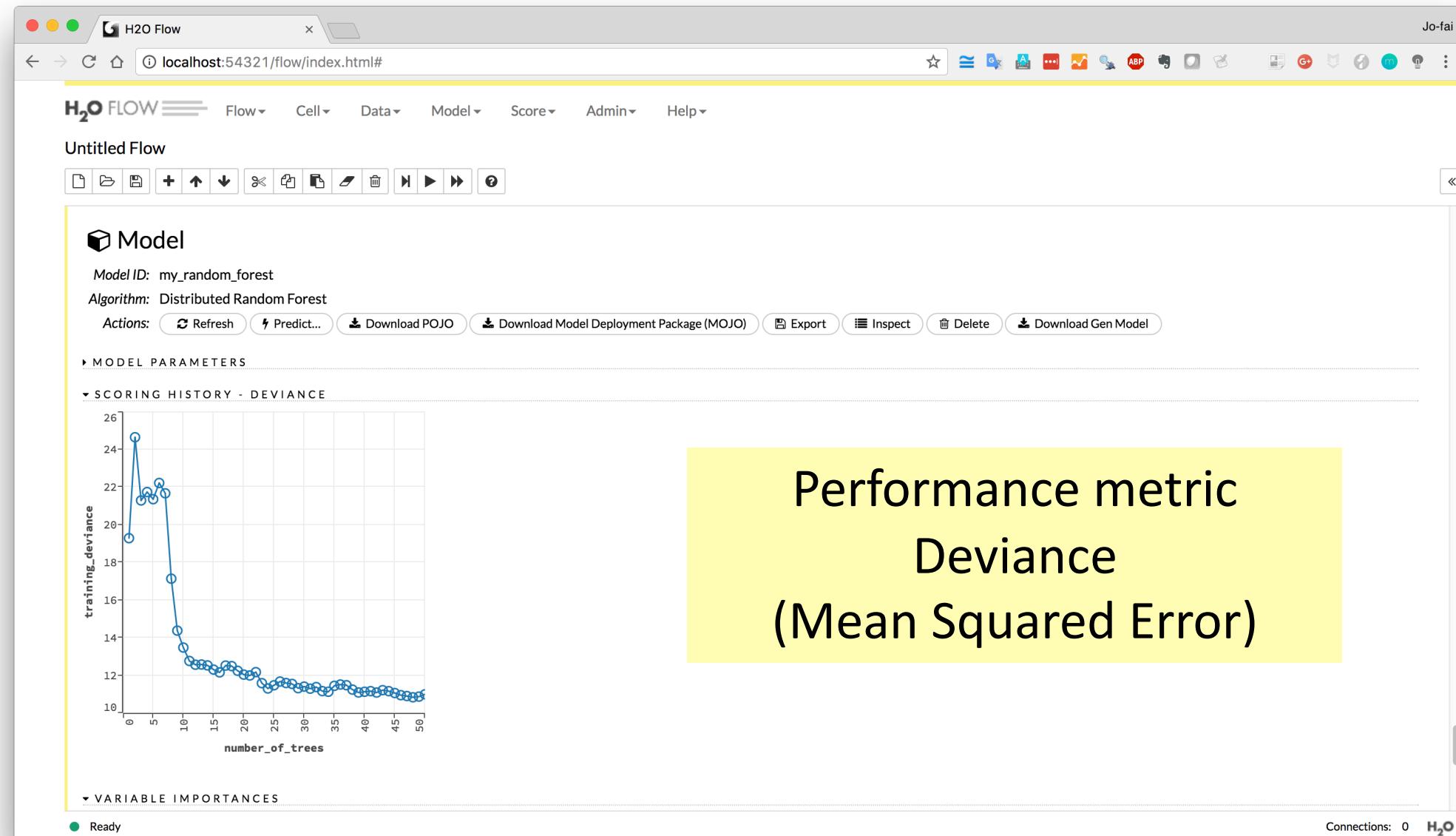
Run Time 00:00:00.654  
Remaining Time 00:00:00.0

Type Model  
Key Q my\_random\_forest  
Description DRF  
Status DONE  
Progress 100% Done.  
Actions View

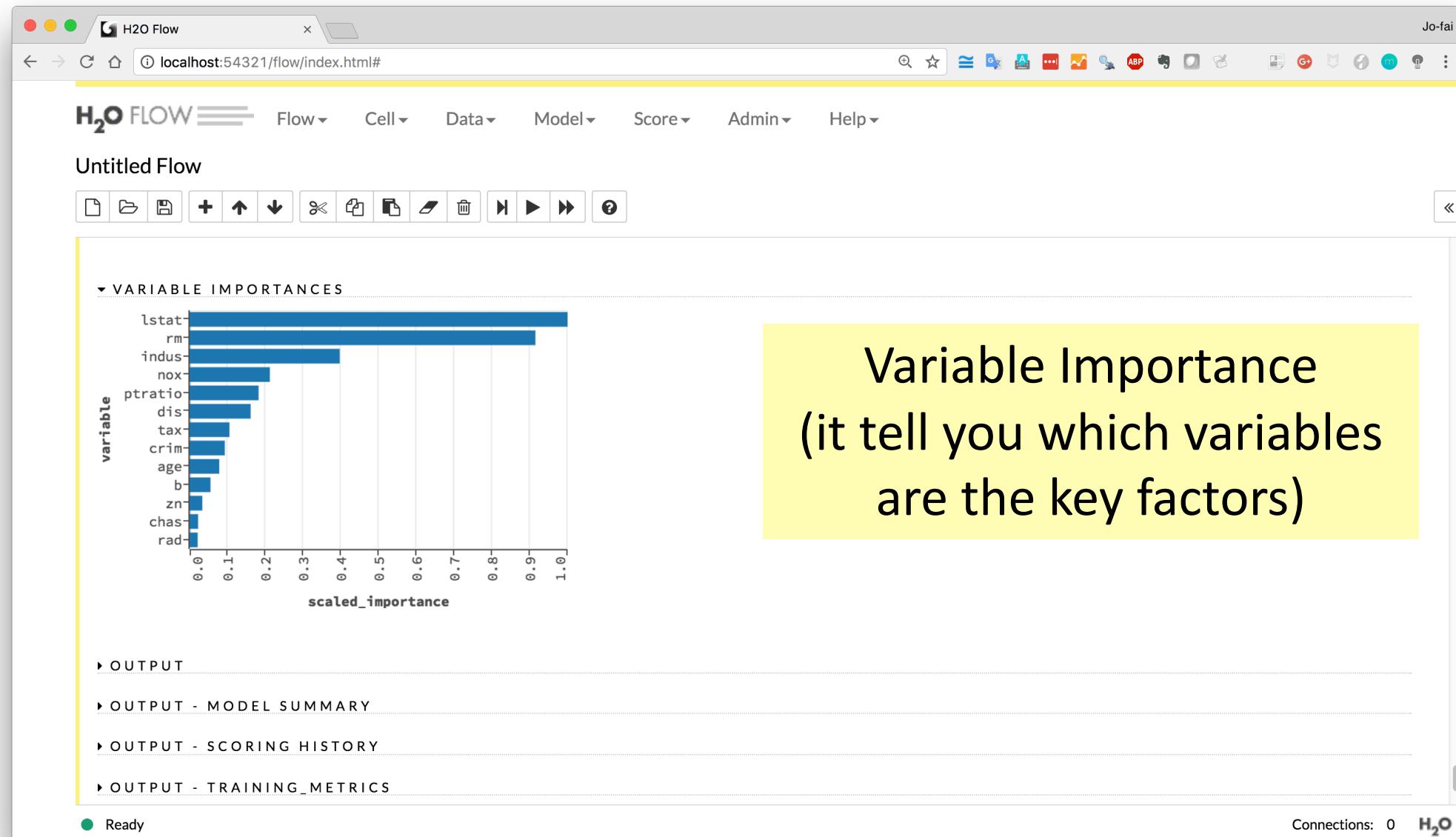
View the model

Ready Connections: 0 H2O





Performance metric  
Deviance  
(Mean Squared Error)



The screenshot shows the H2O Flow web application running in a browser. The title bar says "H2O Flow". The top navigation bar includes "Flow", "Cell", "Data", "Model", "Score", "Admin", and "Help". A dropdown menu under "Score" is open, showing "Predict...", "Partial Dependence Plots...", and "List All Predictions". The main workspace is titled "Untitled Flow" and contains a single step labeled "predict". Below this, there is a "Predict" section with the following fields:

- Name: my\_predictions
- Model: my\_random\_forest
- Frame: Key\_Frame\_house\_price\_test.hex
- Actions: Predict

A yellow callout box on the right side of the screen contains the following text:

We have a model.  
We can now make some predictions on the test dataset.

- 1) Score → Predict
- 2) Select the model
- 3) Select “test” data
- 4) Click “Predict”

The screenshot shows the H2O Flow web application running in a browser window. The title bar reads "H2O Flow" and the address bar shows "localhost:54321/flow/index.html#". The main interface has a top navigation bar with "Flow", "Cell", "Data", "Model", "Score", "Admin", and "Help" dropdowns, along with various toolbar icons. Below the navigation is a toolbar with icons for file operations like open, save, and copy. The main workspace contains a code editor with the following content:

```
predict model: "my_random_forest", frame: "Key_Frame__house_price_test.hex", predictions_frame: "my_predictions"
```

A yellow callout box with a black border and rounded corners is overlaid on the right side of the screen. It contains the text:

Combine predictions with original test data so we can compare the predictions with ground truth.

H2O Flow Jo-fai

localhost:54321/flow/index.html#

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

combined-my\_predictions

Prediction

Ground Truth

Previous 20 Columns Next 20 Columns

Row	predict	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	b	lstat	medv
1	27.0670	0.0063	18.0	2.3100	0	0.5380	6.5750	65.2000	4.0900	1.0	296.0	15.3000	396.9000	4.9800	24.0
2	25.9460	0.0299	0	2.1800	0	0.4580	6.4300	58.7000	6.0622	3.0	222.0	18.7000	394.1200	5.2100	28.7000
3	16.9850	1.2325	0	8.1400	0	0.5380	6.1420	91.7000	3.9769	4.0	307.0	21.0	396.9000	18.7200	15.2000
4	15.5270	0.9884	0	8.1400	0	0.5380	5.8130	100.0	4.0952	4.0	307.0	21.0	394.5400	19.8800	14.5000
5	15.0130	1.1308	0	8.1400	0	0.5380	5.7130	94.1000	4.2330	4.0	307.0	21.0	360.1700	22.6000	12.7000
6	15.6990	1.1517	0	8.1400	0	0.5380	5.7010	95.0	3.7872	4.0	307.0	21.0	358.7700	18.3500	13.1000
7	21.8920	0.0801	0	5.9600	0	0.4990	5.8500	41.5000	3.9342	5.0	279.0	19.2000	396.9000	8.7700	21.0
8	20.3540	0.1751	0	5.9600	0	0.4990	5.9660	30.2000	3.8473	5.0	279.0	19.2000	393.4300	10.1300	24.7000
9	29.0590	0.0276	75.0	2.9500	0	0.4280	6.5950	21.8000	5.4011	3.0	252.0	18.3000	395.6300	4.3200	30.8000
10	20.9370	0.0887	21.0	5.6400	0	0.4390	5.9630	45.7000	6.8147	4.0	243.0	16.8000	395.5600	13.4500	19.7000
11	23.3900	0.0205	85.0	0.7400	0	0.4100	6.3830	35.7000	9.1876	2.0	313.0	17.3000	396.9000	5.7700	24.7000
12	19.1160	0.1717	25.0	5.1300	0	0.4530	5.9660	93.4000	6.8185	8.0	284.0	19.7000	378.0800	14.4400	16.0
13	23.7010	0.1103	25.0	5.1300	0	0.4530	6.4560	67.8000	7.2255	8.0	284.0	19.7000	396.9000	6.7300	22.2000
14	21.9820	0.0839	0	12.8300	0	0.4370	5.8740	36.6000	4.5026	5.0	398.0	18.7000	396.0600	9.1000	20.3000
15	23.2093	0.0355	25.0	4.8600	0	0.4260	6.1670	46.7000	5.4007	4.0	281.0	19.0	390.6400	7.5100	22.9000
16	22.9100	0.0715	0	1.1900	0	0.1100	6.1210	56.8000	3.7476	3.0	217.0	18.5000	395.1500	8.1100	22.2000

Ready Connections: 0 H2O

**Download the results**

**Actions:** View Data Split... Build Model... Predict Download Export Delete

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
<code>predict</code>	real	0	0	0	0	8.1564	46.3040	22.7819	7.9191	.	.
<code>crim</code>	real	0	0	0	0	0.0063	88.9762	3.8033	10.9319	.	.
<code>zn</code>	real	0	71	0	0	0	95.0	14.5556	27.1545	.	.
<code>indus</code>	real	0	0	0	0	0.7400	27.7400	10.0166	6.9690	.	.
<code>chas</code>	int	0	92	0	0	0	1.0	0.0707	0.2576	.	Convert to enum
<code>nox</code>	real	0	0	0	0	0.3890	0.8710	0.5462	0.1174	.	.
<code>rm</code>	real	0	0	0	0	3.5610	8.7800	6.2766	0.7525	.	.
<code>age</code>	real	0	0	0	0	9.9000	100.0	65.2283	29.3788	.	.
<code>dis</code>	real	0	0	0	0	1.1781	12.1265	4.1131	2.4283	.	.
<code>rad</code>	int	0	0	0	0	1.0	24.0	8.3636	8.3207	.	Convert to enum
<code>tax</code>	int	0	0	0	0	187.0	711.0	391.2121	160.1494	.	Convert to enum
<code>ptratio</code>	real	0	0	0	0	12.6000	22.0	18.4889	2.1887	.	.
<code>b</code>	real	0	0	0	0	6.6800	396.9000	366.0096	78.0843	.	.
<code>lstat</code>	real	0	0	0	0	2.8800	34.4100	12.0817	7.3215	.	.
<code>medv</code>	real	0	0	0	0	5.6000	50.0	22.1545	9.2839	.	.

Previous 20 Columns Next 20 Columns

CHUNK COMPRESSION SUMMARY

Ready Connections: 0 H<sub>2</sub>O

H2O Flow Jo-fai

localhost:54321/flow/index.html#

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

buildPartialDependence

cs

Partial Dependence

Save Destination PDP  as:

Model: my\_random\_forest

Frame: Key\_Frame\_house\_price\_train.hex

nbins 20

Select columns?

Actions: Compute

How many levels should PDP compute. More levels will make it slower.

Checking this will allow you to select custom columns for PDP. By default, the top 10 features are used. Those features are sorted by variable importance.

Connections: 0 H2O

Using Partial Dependence  
Plots to further explain  
results

17ms

