

## Algorithm Regression

# ALGORITHM REGRESSION



# Algorithm Regression

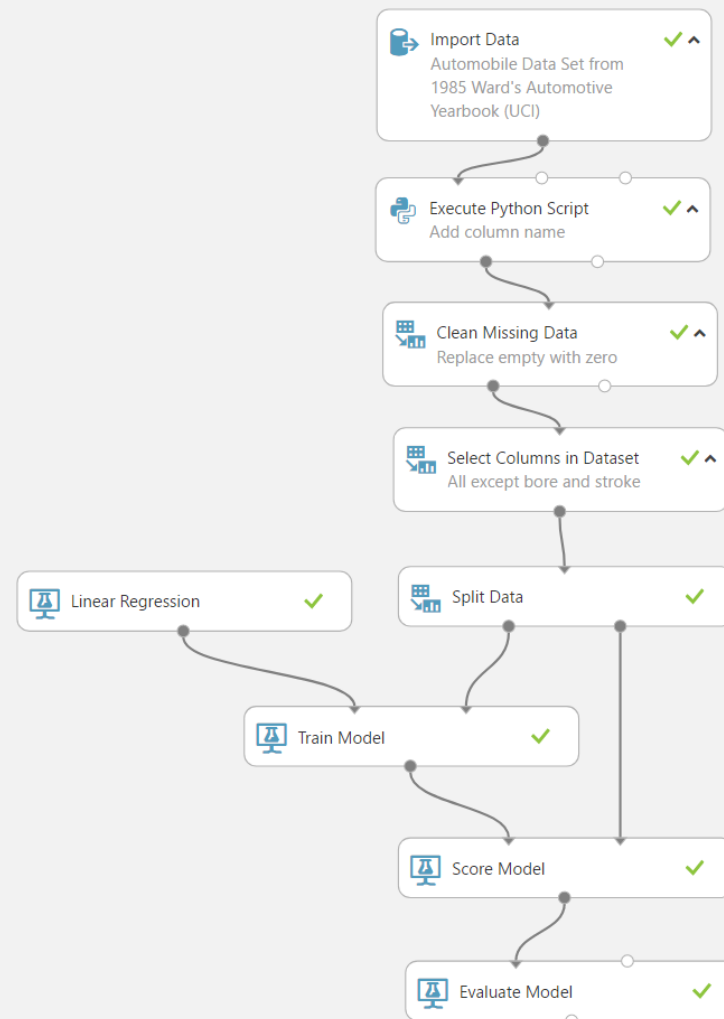
## In this session

- Regression Algorithms in Azure ML
- Create New ML Experiment
- Import auto data from UCI
- Add column name using python
- Clean missing data
- Select column (exclude column)
- Split data
- Add Linear Regression module
- Add Train Model
- Add Score Model
- Add Evaluate Model



# Algorithm Regression

## Over view



# Algorithm Regression

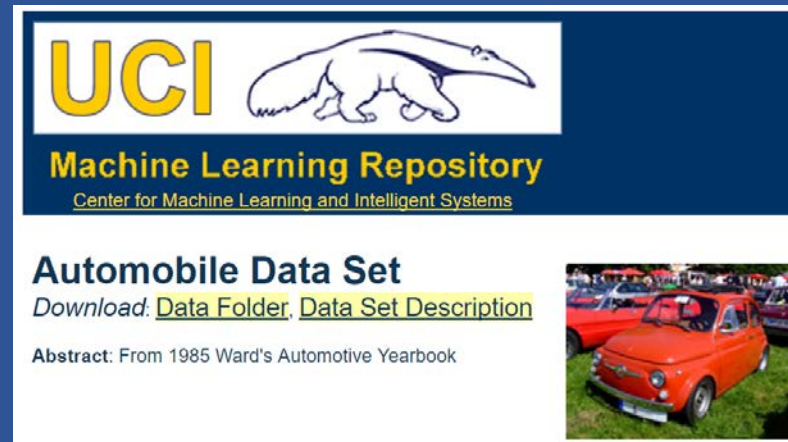
## Working steps

### Working Steps

1. Create New ML Experiment
2. Import auto data from UCI
3. Add column name using python
4. Clean missing data
5. Select column (exclude column)
6. Split data
7. Add Linear Regression module
8. Add Train Model
9. Add Score Model
10. Add Evaluate Model

# Algorithm Regression

## Data set



Home

<https://archive.ics.uci.edu/ml/datasets/Automobile>

Data download

<https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.data>

Element column names/values

<https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.names>

# Algorithm Regression

## Data attribute

1. **symboling**: -3, -2, -1, 0, 1, 2, 3
2. **normalized-losses**: continuous from 65 to 256
3. **make**: alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo
4. **fuel-type**: diesel, gas
5. **aspiration**: std, turbo
6. **num-of-doors**: four, two
7. **body-style**: hardtop, wagon, sedan, hatchback, convertible
8. **drive-wheels**: 4wd, fwd, rwd
9. **engine-location**: front, rear
10. **wheel-base**: continuous from 86.6 to 120.9
11. **length**: continuous from 141.1 to 208.1
12. **width**: continuous from 60.3 to 72.3
13. **height**: continuous from 47.8 to 59.8

# Algorithm Regression

## Data attribute

14. **curb-weight**: continuous from 1488 to 4066
15. **engine-type**: dohc, dohcv, l, ohc, ohcf, ohcv, rotor
16. **num-of-cylinders**: eight, five, four, six, three, twelve, two
17. **engine-size**: continuous from 61 to 326
18. **fuel-system**: 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi
19. **bore**: continuous from 2.54 to 3.94
20. **stroke**: continuous from 2.07 to 4.17
21. **compression-ratio**: continuous from 7 to 23
22. **horsepower**: continuous from 48 to 288
23. **peak-rpm**: continuous from 4150 to 6600
24. **city-mpg**: continuous from 13 to 49
25. **highway-mpg**: continuous from 16 to 54
26. **price**: continuous from 5118 to 45400



# Algorithm Regression

## Add column name Python Script

```
1 #import pandas as pd
2 def azureml_main(dataframe1 = None, dataframe2 = None):
3     dataframe1.columns = [
4         'symboling',
5         'normalized-losses',
6         'make',
7         'fuel-type',
8         'aspiration',
9         'num-of-doors',
10        'body-style',
11        'drive-wheels',
12        'engine-location',
13        'wheel-base',
14        'length',
15        'width',
16        'height',
17        'curb-weight',
18        'engine-type',
19        'num-of-cylinders',
20        'engine-size',
21        'fuel-system',
22        'bore',
23        'stroke',
24        'compression-ratio',
25        'horsepower',
26        'peak-rpm',
27        'city-mpg',
28        'highway-mpg',
29        'price'
30    ]
31     return dataframe1,
```

# Algorithm Regression


Clean missing data

▲ Clean Missing Data


Columns to be cleaned

**Selected columns:**  
**All columns**

Launch column selector

Minimum missing value ra... 


0

Maximum missing value ra... 


1

Cleaning mode

Custom substitution value ▼

Replacement value 

0

☐ Generate missing valu... 

# Algorithm Regression

Select (exclude) column

## Select columns

BY NAME

WITH RULES

☐ Allow duplicates and preserve column order in selection

Begin With

ALL COLUMNS

NO COLUMNS

Exclude ▼

column names ▼

bore ✕

stroke ✕

+

-

✓

# Algorithm Regression

## Split data

Split Data

Splitting mode

Split Rows ▼

Fraction of rows in the first... ≡

0.8

☒ Randomized split ≡

Random seed ≡

0

Stratified split

False ▼

# Algorithm Regression

## Evaluation Metrics

- Add Linear Regression module
- Add Train Model
- Add Score Model
- Add Evaluate Model
- Run
- Inspect Score
- Inspect Evaluate metrics

Linear Regression

Solution method  
Ordinary Least Squares

L2 regularization weight  
0.001

☒ Include intercept term

Random number seed

☒ Allow unknown catego...

Train Model

Label column

Selected columns:  
Column names: price

Launch column selector

### Metrics

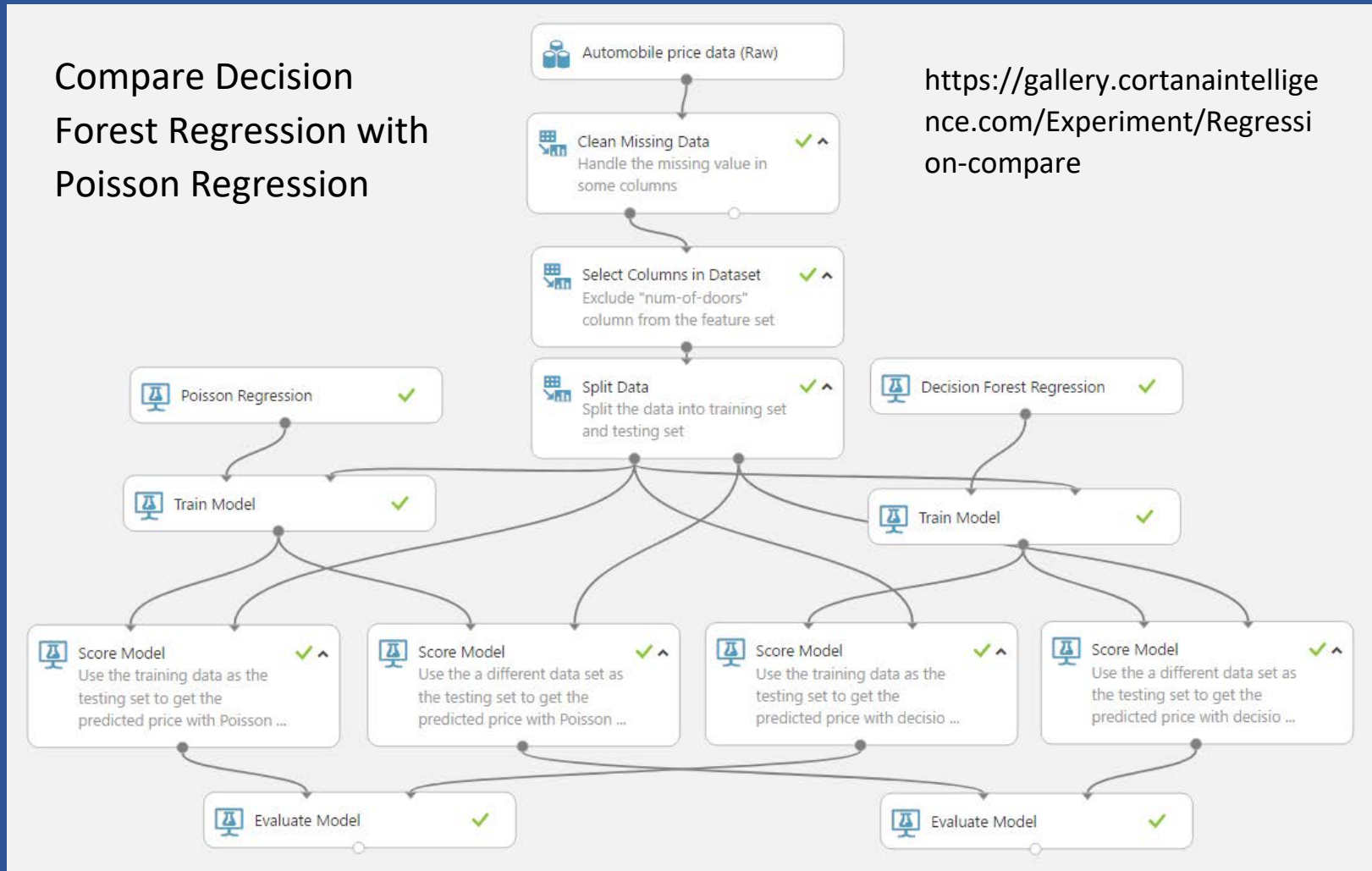
Mean Absolute Error	2728.117473
Root Mean Squared Error	4130.273478
Relative Absolute Error	0.388729
Relative Squared Error	0.200384
Coefficient of Determination	0.799616

# Algorithm Regression

## Regression compare

### Compare Decision Forest Regression with Poisson Regression

<https://gallery.cortanaintelligence.com/Experiment/Regression-compare>



# Algorithm Regression

More information

Using linear regression in Azure Machine Learning

<https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-linear-regression-in-azure>

This Experiment

<https://gallery.cortanaintelligence.com/Experiment/Regression-3>

Regression compare

<https://gallery.cortanaintelligence.com/Experiment/Regression-compare>