

# Linear Regression 기반의 국내 주택 가격지수 예측 연구

최승은 2016162060

을지대학교 의료IT마케팅학과

# INDEX

01

서론

02

국내 집값 예측 Azure 실습

03

웹 서비스 배포 / R, Python 실습

04

결론

# 01

---

## 서론

## ▶ 대다수의 공통된 목표 : 내 집 마련, 부동산 투자

### 20대 94.8% '내 집 마련 필요하다'

※20대 청년층 2,889명 대상 조사, 자료 : 잡코리아X알바몬

Q. 내 집 마련이 필요한 이유 (\*복수응답)



JOBKOREA × albamon

〈머니투데이-KB국민은행 '주택구매의향 및 수익형 부동산 투자 의향 설문'〉

#### 수익형 부동산 투자 의향 설문 (단위: %)

■ 현재 수익형 부동산 투자 여부



■ 향후 수익형 부동산 투자 의향



■ 수익형 부동산 투자 의향이 없는 이유



■ 수익형 부동산에 투자 의향이 있다면 투자하고 싶은 상품



■ 수익형 부동산 투자 의사결정시 고려하는 최소 수익



■ 수익형 부동산 투자시 가장 걱정되는 부분



\*기간: 6월 중, 총 2795명 조사

\*자료: 머니투데이-KB국민은행 설문조사

그래픽: 이승현 디자인기자

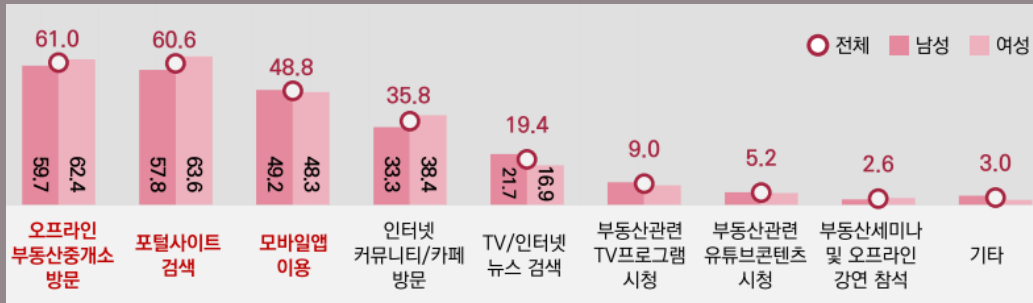


MT 머니투데이

## ▶ 부동산 투자 정보 부족

- 단순히 매매 정보만 탐색하는 모습

출처 : 한국방송광고진흥공사



&lt; 부동산 정보 습득 경로 &gt;



&lt; 알고 있는 부동산 정보제공서비스 &gt;

## ▶ 연구 내용

부동산 투자 관련 예측 지표 연구

# 02

---

## 국내 집값 예측 Azure 실습

2006년~2018년의  
전국 주택 가격동향  
= 10,144개의 데이터

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
tradeprice_sido_n1	region_cd	tradeprice_sido	building_type	construction_realized_amount	cd	spirit_deposit_rate	exchange_rate	composite_stock_price_index	economy_growth	exchequer_bond_three	household_loan_all	mortgage_all	numberofhouseholds	unsalenum_c
69.5	48000	69.4	0	4867793	4.15	4.06	971	1379.32	1.8	5.02				
64.6	48000	64.5	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
69.7	47000	69.3	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
94.2	45000	94.2	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
80	48000	80	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
82.9	42000	82.9	3	4867793	4.15	4.06	971	1379.32	1.8	5.02				
62.1	28000	62	0	4867793	4.15	4.06	971	1379.32	1.8	5.02				
94.6	46000	94.4	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
65.3	28000	65.2	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
67.6	26000	67.5	0	4867793	4.15	4.06	971	1379.32	1.8	5.02				
73.5	45000	73.5	3	4867793	4.15	4.06	971	1379.32	1.8	5.02				
73.2	46000	73.2	0	4867793	4.15	4.06	971	1379.32	1.8	5.02				
86.4	43000	86.4	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
77.2	41000	77.1	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
62.4	46000	62.4	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
66.9	11000	66.3	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
69.1	41000	68.7	0	4867793	4.15	4.06	971	1379.32	1.8	5.02				
90.5	42000	90.5	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
69.7	41000	69.1	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
77.6	50000	77.6	0	4867793	4.15	4.06	971	1379.32	1.8	5.02				
52.8	50000	52.8	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
72.3	29000	72	0	4867793	4.15	4.06	971	1379.32	1.8	5.02				
58.1	26000	58.1	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
92.5	47000	92.8	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
75.9	28000	75.9	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
97.6	29000	97.6	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
79.7	48000	79.9	3	4867793	4.15	4.06	971	1379.32	1.8	5.02				
82.1	50000	82.1	3	4867793	4.15	4.06	971	1379.32	1.8	5.02				
94.7	50000	94.7	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
72.2	42000	71.7	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
57	31000	56.9	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
74.2	43000	74.2	3	4867793	4.15	4.06	971	1379.32	1.8	5.02				
92	44000	92.1	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
77.4	27000	77.5	3	4867793	4.15	4.06	971	1379.32	1.8	5.02				
70.9	31000	70.5	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
68.5	43000	67.8	1	4867793	4.15	4.06	971	1379.32	1.8	5.02				
55.8	28000	55.9	3	4867793	4.15	4.06	971	1379.32	1.8	5.02				
84.6	26000	84.6	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				
85	30000	85.6	7	4867793	4.15	4.06	971	1379.32	1.8	5.02				

## 오픈API 상세

## XML 전국주택가격동향조사 통계 조회 서비스

한국감정원의 부동산(단독주택, 연립다세대, 아파트 등)의 전세가격 지수를 조사한 데이터

0

0

관심

## OpenAPI 정보

분류체계	지역개발 - 지역및도시	제공기관	한국감정원
관리부서명	ICT전략부	관리부서 전화번호	053-663-8474
API 유형	REST	데이터포맷	XML
활용신청	303	키워드	한국감정원,부동산통계,전세가격
등록	2016-08-09	수정	2018-05-09
심의유형	개발단계 : 허용 / 운영단계 : 허용		
비용부과유무	무료		
이용허락범위	<a href="#">이용허락범위 제한 없음</a>		
참고문서	<a href="#">기술문서 전국주택가격동향조사+통계+조회+서비스.docx</a>		

Dataset – 전국주택가격동향조사.csv  
(출처 : 공공데이터 포털 – 한국감정원 주택가격지수)

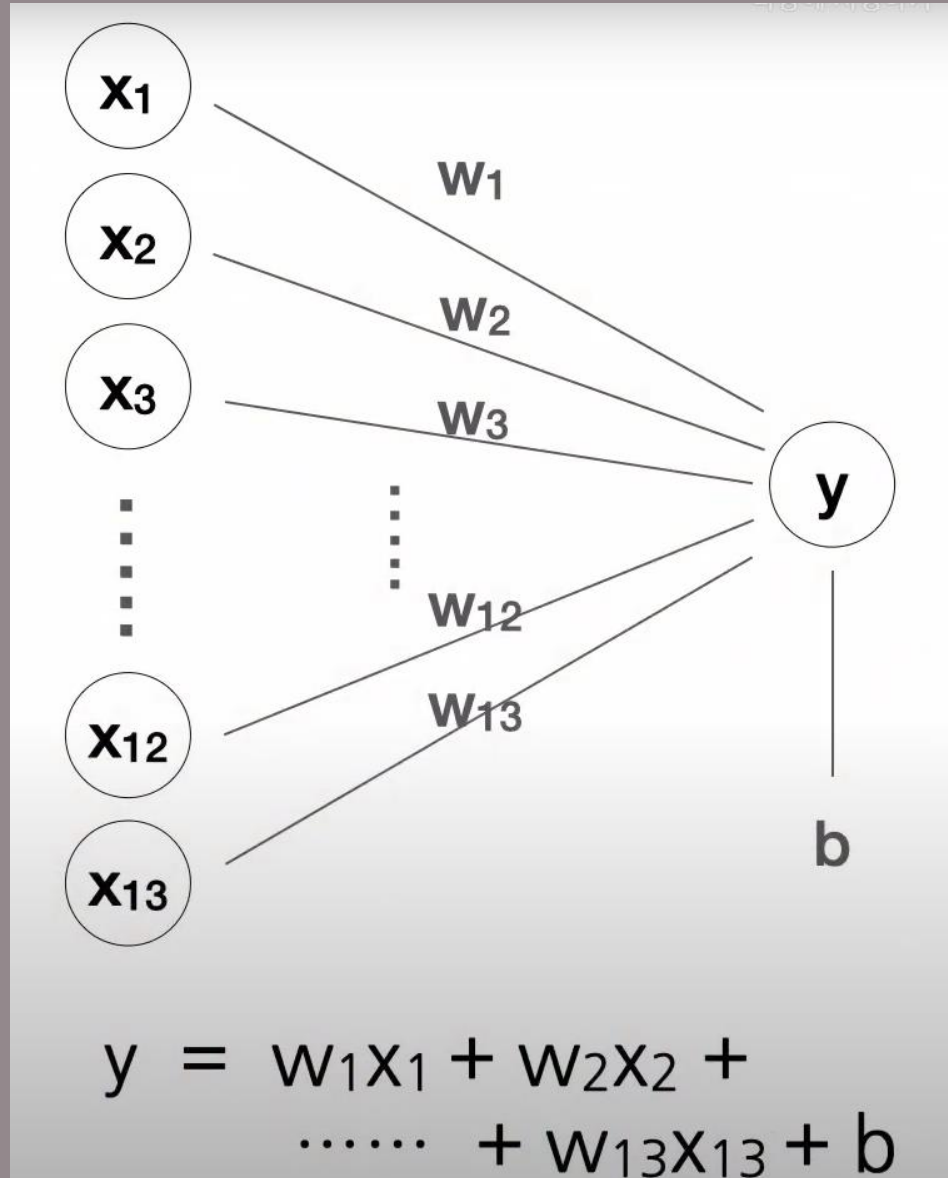
14개의  
독립변수

종속변수

변수 이름	변수 내용
region_cd	지역코드(시도)
tradeprice_sido	매매가격지수(시도)
building_type	부동산타입
construction_realized_amount	건설기성액(백만원)
cd	cd(91일물)
spirit_deposit_rate	정기예금금리
exchange_rate	환율
composite_stock_price_index	종합주가지수
economy	이달의 경제지표
exchange_rate_three	국고채 3년
household_loan_all	가계대출액(전국)
mortgage_all	주택대출액(전국)
numberofnosells	미분양 가구수(시도)
unsalenum_c	공사완료후 미분양(민간,시도)
tradeprice_sido_n1	한달 후의 매매가격지수(시도)

이번달의 매매가격지수와  
다른 경제 지표들을 알고 있을 때  
→ 다음달의 매매가격지수를 예측





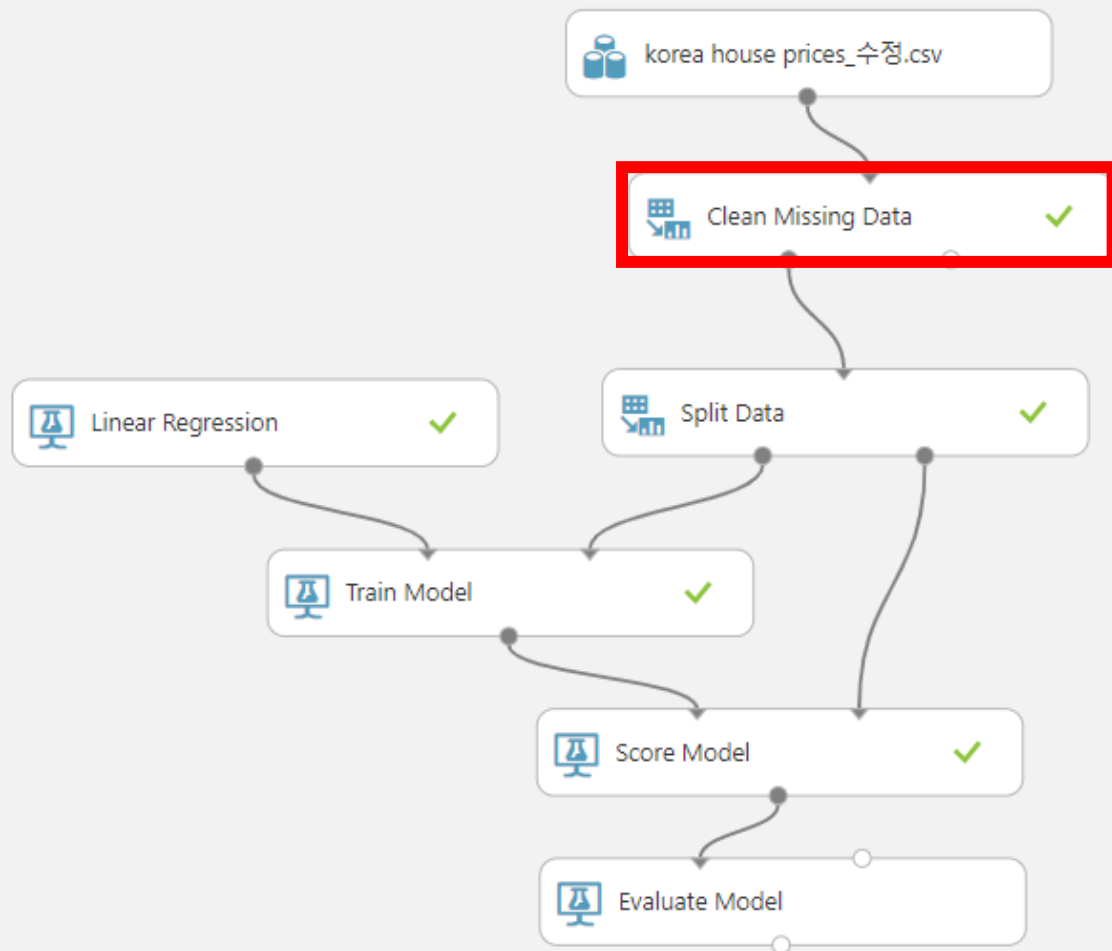
# Linear Regression

## 선형회귀

## 02

## 국내 집값 예측 Azure 실습

korea house prices\_



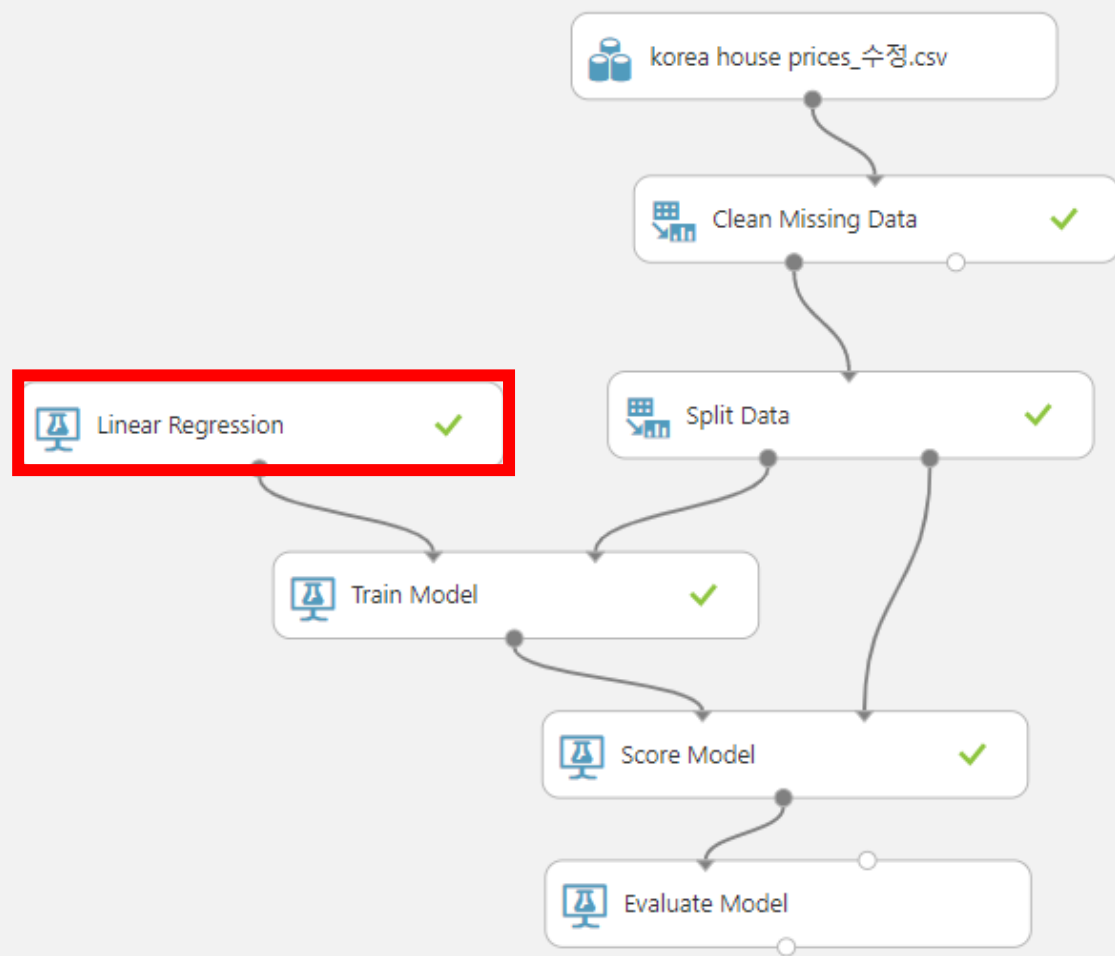
korea house prices\_ > Clean Missing Data > Cleaned dataset

rows	columns
6552	15

tradeprice_sido_n1	region_cd	tradeprice_sido	building_type	construction_realized_amount	cd	spirit_deposit_rate	exchange_rate	composite_stock
93.4	45000	93.3	7	6500576	2.88	3.93	1156.5	1682.16
83.9	30000	83.1	0	6500576	2.88	3.93	1156.5	1682.16
92.7	30000	92.3	7	6500576	2.88	3.93	1156.5	1682.16
80.8	30000	79.7	1	6500576	2.88	3.93	1156.5	1682.16
81.8	30000	81.8	3	6500576	2.88	3.93	1156.5	1682.16
75.7	50000	75.5	0	6500576	2.88	3.93	1156.5	1682.16
84.4	50000	84.4	7	6500576	2.88	3.93	1156.5	1682.16
61	50000	60.2	1	6500576	2.88	3.93	1156.5	1682.16
77.9	50000	78.2	3	6500576	2.88	3.93	1156.5	1682.16
72.1	48000	71.6	1	6500576	2.88	3.93	1156.5	1682.16
74.7	48000	74.3	0	6500576	2.88	3.93	1156.5	1682.16
79.6	48000	79.5	7	6500576	2.88	3.93	1156.5	1682.16
86.4	46000	86	0	6500576	2.88	3.93	1156.5	1682.16
94.5	46000	94.4	7	6500576	2.88	3.93	1156.5	1682.16
83.6	46000	83.6	3	6500576	2.88	3.93	1156.5	1682.16
80.5	46000	79.9	1	6500576	2.88	3.93	1156.5	1682.16
81.6	27000	81.5	3	6500576	2.88	3.93	1156.5	1682.16

&lt;결측값 제거 → 6,552개의 데이터&gt;

korea house prices\_



korea house prices\_ &gt; Train Model &gt; Trained model

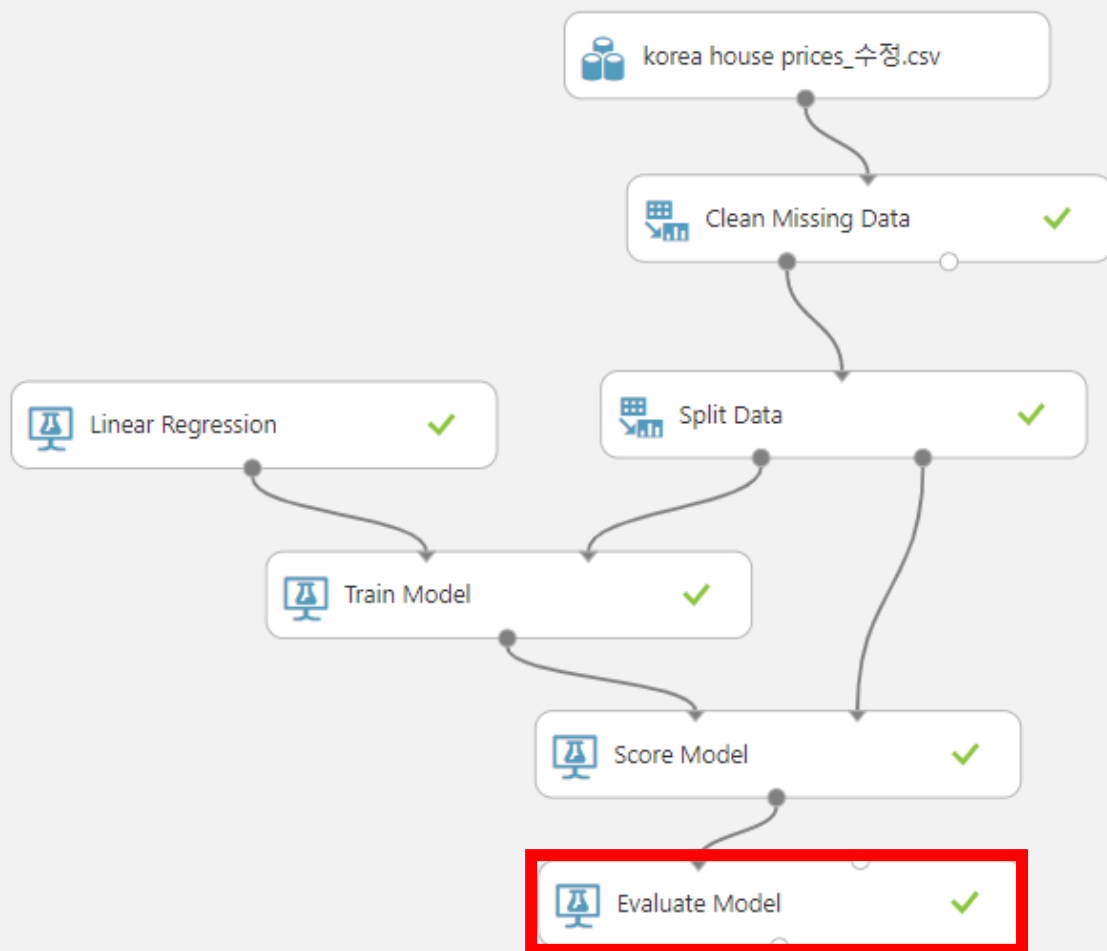
## Feature Weights

Feature	Weight
tradeprice_sido	0.978281
Bias	-0.181076
spirit_deposit_rate	0.0942917
economy_growth	0.0741057
exchequer_bond_three	-0.03904
cd	0.0160269

Tradeprice\_sido(해당 월의 매매가격지수)가 0.98로  
매우 강한 양의 상관관계를 보이고,  
spirit\_deposit\_rate(정기예금금리)가 0.09,  
economy\_growth(경제성장률)가 0.07로  
약한 양의 상관관계를 보여줌

&lt;선형 회귀 Visualization&gt;

korea house prices\_



korea house prices\_ &gt; Evaluate Model &gt; Evaluation results

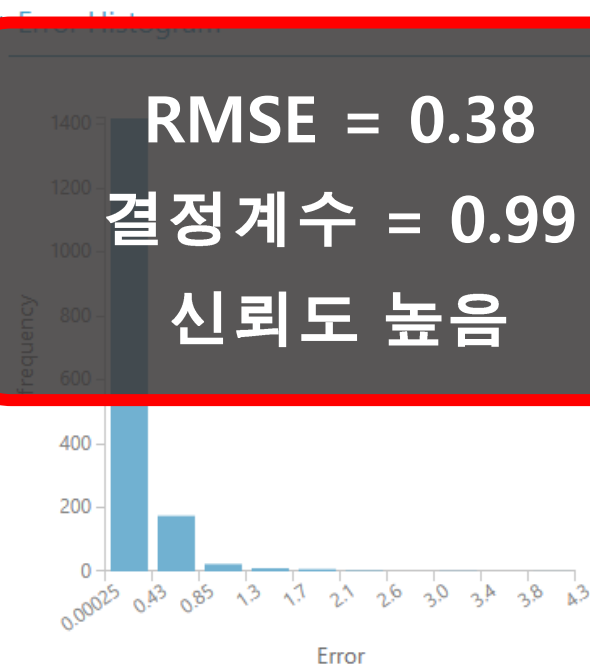
## Metrics

Mean Absolute Error	0.232313
Root Mean Squared Error	0.381929
Relative Absolute Error	0.046631
Relative Squared Error	0.003168
Coefficient of Determination	0.996832

RMSE = 0.38

결정계수 = 0.99

신뢰도 높음

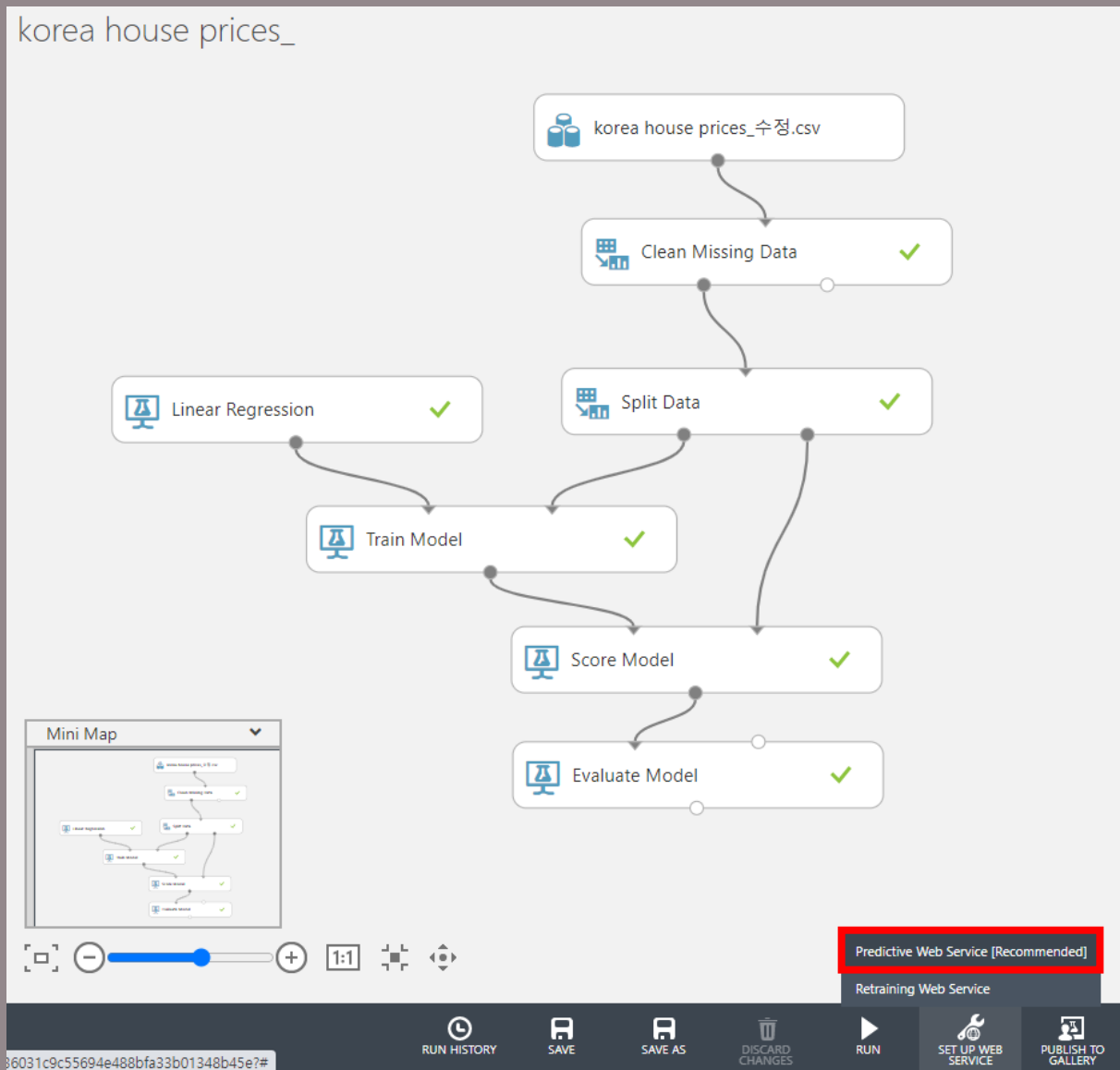


&lt;모델 평가 Visualization&gt;

# 03

---

웹 서비스 배포 / R, Python 실습



```

> library("curl")
library("httr")
library("rjson")

requestFailed = function(response) {
  return (response$status_code >= 400)
}

printHttpRequest = function(response, result) {
  if (requestFailed(response)) {
    print(paste("The request failed with status code:", response$status_code, sep=" "))
  }
  # Print the headers - they include the request ID and the timestamp, which are useful for debugging
  the failure
  print(response$headers)
}

print("Result:")
print(fromJSON(result))
}

req = list(
  inputs = list(
    "input1" = list(
      list(
        'column 0' = "1",
        'crim' = "1",
        'zn' = "1",
        'indus' = "1",
        'chas' = "1",
        'nox' = "1",
        'rm' = "1",
        'age' = "1",
        'dis' = "1",
        'rad' = "1",
        'tax' = "1",
        'ptratio' = "1",
        'black' = "1",
        'lstat' = "1"
      )
    )
  ),
  globalParameters = setNames(fromJSON('{}'), character(0))
)

body = enc2utf8(toJSON(req))
api_key = "cr97vkvTzPFuCAK/cNHTuF+iwyvzn0Ia1K9q4hsvDgSsBc1hhzVln1ks1xzQ0PqFJwhrvT1Zclt7+R3F2QbnxQ=="
authz_hdr = paste("Bearer ", api_key, sep=" ")

response=POST(url = "https://ussouthcentral.services.azureml.net/workspaces/36031c9c55694e488bfa33b01348b45e
/services/4065b07d5f934dc88304d880153b5339/execute?api-version=2.0&format=swagger",
  auth_headers( Content-type = "application/json", Authorization = authz_hdr),
  body=body)

result = content(response, type="text", encoding="UTF-8")
printHttpRequest(response, result)

```

```

[1] "Result:"
$Results
$Results$output1
$Results$output1[[1]]
$Results$output1[[1]]$`scored Labels`
[1] "41.315544128418"

```

< 결과값 출력 >

< R 실습 >

Primary Key와 Request-Respond 값을 입력

```

In [3]: import urllib.request
import json

data = {
    "Inputs": {
        "input1":
            {
                'Column 0': "1",
                'crim': "1",
                'zn': "1",
                'indus': "1",
                'chas': "1",
                'nox': "1",
                'ra': "1",
                'age': "1",
                'dis': "1",
                'rad': "1",
                'tax': "1",
                'ptratio': "1",
                'black': "1",
                'lstat': "1",
            },
        "GlobalParameters": {
        }
    }

    body = str.encode(json.dumps(data))

url = "https://ussouthcentral.services.azureml.net/workspaces/36031c9c-55694e488bfa33b01348b45e/services/40656b07d51934dc88304d1"
api_key = "cr97vkYtzPFuCAK/cNHUf+Iwyvzn01aIK9q4hSYDgS8BChhzVLnkSlxzQP9fJWhrv11zCLt7+R3F2QbnxQ=="
headers = {'Content-Type': 'application/json', 'Authorization': ('Bearer ' + api_key)}

req = urllib.request.Request(url, body, headers)

try:
    response = urllib.request.urlopen(req)
    result = response.read()
    print(result)
except urllib.error.HTTPError as error:
    print("The request failed with status code: " + str(error.code))

    # Print the headers - they include the request ID and the timestamp, which are useful for debugging the failure
    print(error.info())
    print(json.loads(error.read().decode('utf8', 'ignore')))
```

```
b'{"Results":{"output1":[{"Scored Labels":"41.315544128416"}]}}'
```

< 결과값 출력 >

< Python - Jupyter Notebook 실습 >  
Request-Response와 Primary Key 값 입력



# 04

---

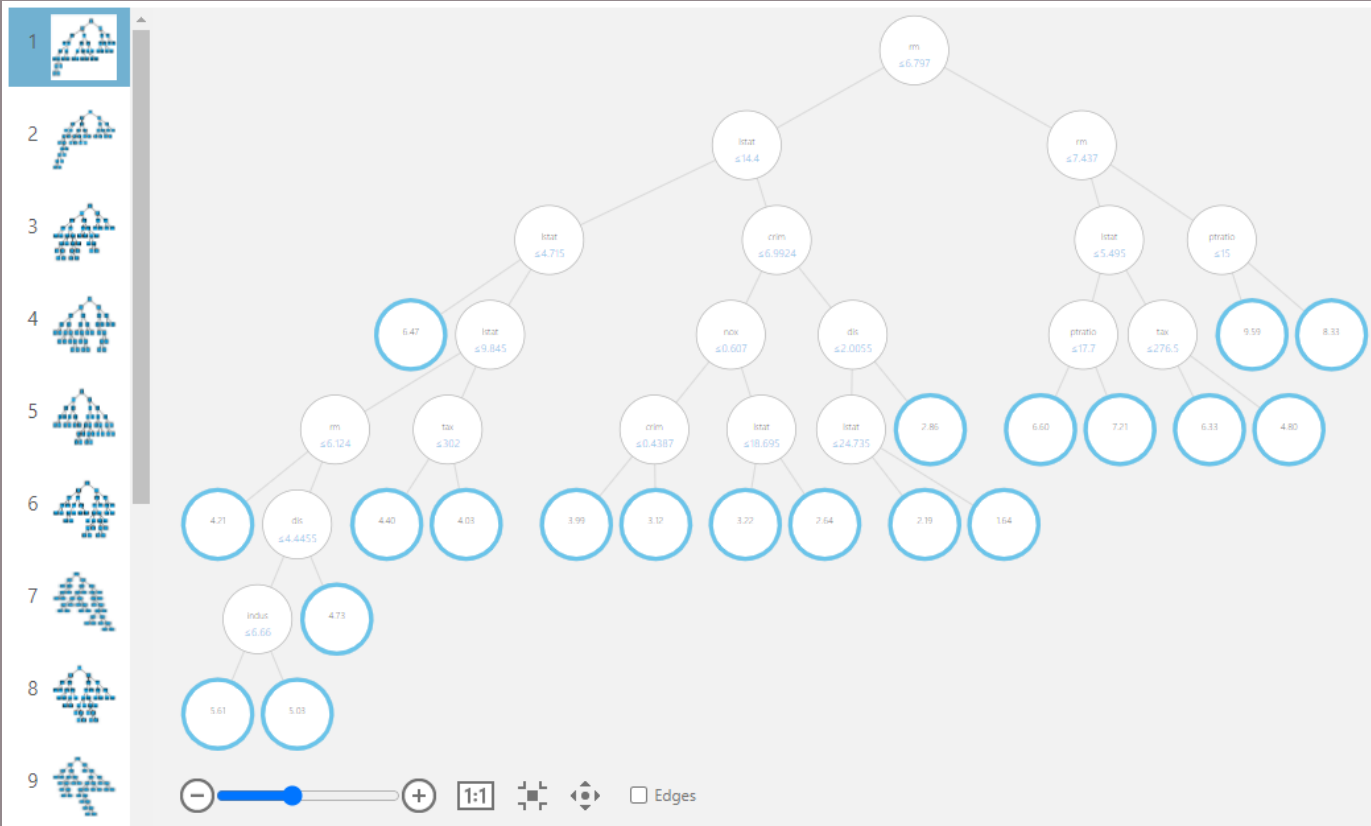
## 결론

## ▶ 결과

- 이번달 매매가격지수와 다른 경제 지표들을 알고 있을 때, 다음달 매매가격지수 예측 가능
- Tradeprice\_sido(해당 월의 매매가격지수)가 0.98로 매우 강한 양의 상관관계를 보이고, spirit\_deposit\_rate(정기예금금리)가 0.09, economy\_growth(경제성장률)가 0.07로 약한 양의 상관관계를 보여줌

## ▶ 개선방안

- 한달 후의 매매가격지수만이 아닌 6개월, 1년, 3년 이상의 매매가격지수 예측이 필요
- 웹 서비스를 만들어서 많은 사람들이 예측 서비스를 이용할 수 있도록 함
- 더욱 정확한 데이터 분석 방법 탐색

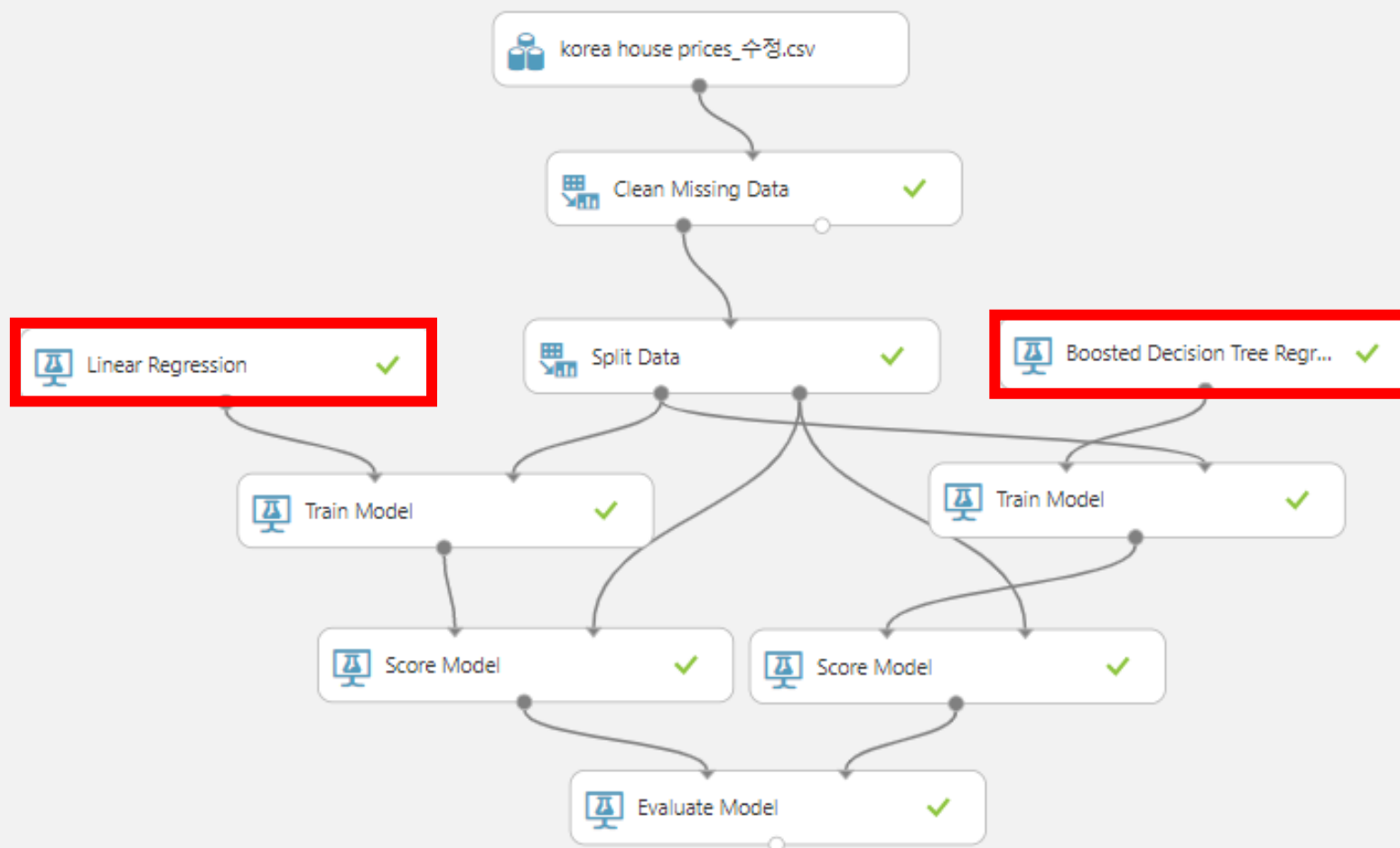


# Boosted Decision Tree Regression

향상된 의사결정트리 회귀분석

→ 정확도 개선

korea house prices\_



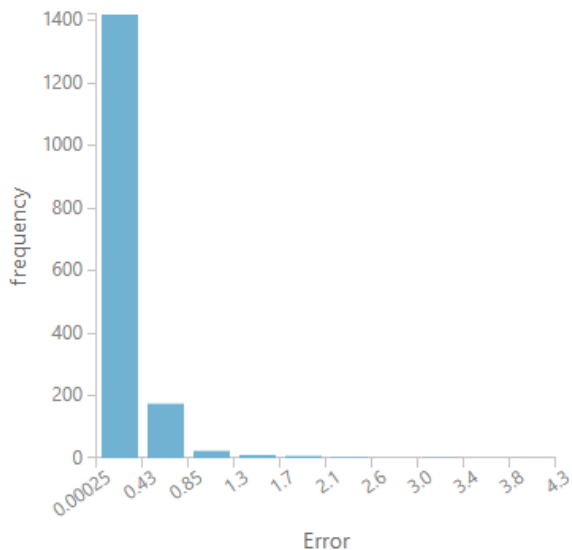
Linear Regression과 Boosted Decision Tree Regression을 비교

korea house prices\_ &gt; Evaluate Model &gt; Evaluation results

## Metrics

Mean Absolute Error	0.232313
Root Mean Squared Error	0.381929
Relative Absolute Error	0.046631
Relative Squared Error	0.003168
Coefficient of Determination	0.996832

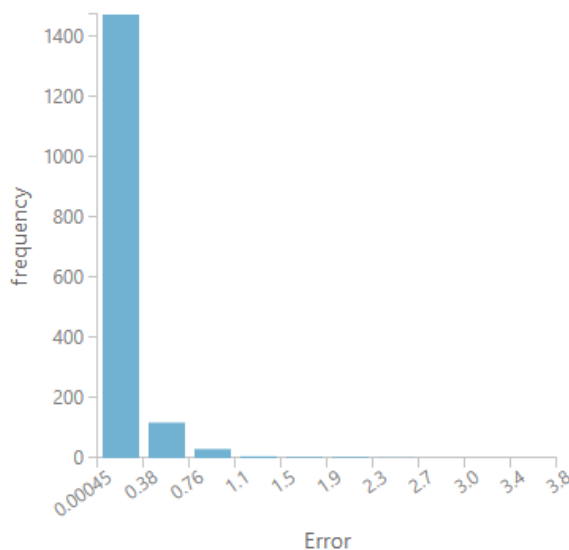
## Error Histogram



## Metrics

Mean Absolute Error	0.181425
Root Mean Squared Error	0.314613
Relative Absolute Error	0.036416
Relative Squared Error	0.00215
Coefficient of Determination	0.99785

## Error Histogram



## 선형회귀

RMSE : 0.38  
결정계수 : 0.9968

## 앙상블 트리

RMSE : 0.31  
결정계수 : 0.9979

## 앙상블 트리 모델의

RMSE(0.38 > 0.31)가 더 낮고,  
결정계수(0.9968 < 0.9979)는 더 높다

---

## Reference

잡코리아, 2020.10.05, 20대 94.8%, '내 집 마련 필요하다'

: [https://www.jobkorea.co.kr/goodjob/tip/view?News\\_No=18293](https://www.jobkorea.co.kr/goodjob/tip/view?News_No=18293)

머니투데이, 2018.07.04, 부동산 불패? 85% '투자계획 있거나 고민 중'

: <https://news.mt.co.kr/mtview.php?no=2018062911194265224>

한국방송광고진흥공사, 공개정보리스트

: [https://www.kobaco.co.kr/site/main/content/public\\_info\\_list](https://www.kobaco.co.kr/site/main/content/public_info_list)

공공데이터포털, 한국감정원\_전국주택가격동향조사 통계 조회 서비스

: <https://www.data.go.kr/data/15044014/openapi.do>

**THANK**

---

**YOU**