# Scale 불량 예측 Project

임승호

# TIMELINE

## 과제정의

배경

분석하고자 하는 방향

예상 목표

## 분석계획

데이터 확인

데이터 전처리

가설 검정

모델링

## 데이터 현황

데이터 확인

기술 통계량 확인

이상치 확인

결측치 확인

## 탐색적 분석

데이터 시각화

가설 검정

## 모델링 및 평가

Regression

RandomForest

Gradient Boosting

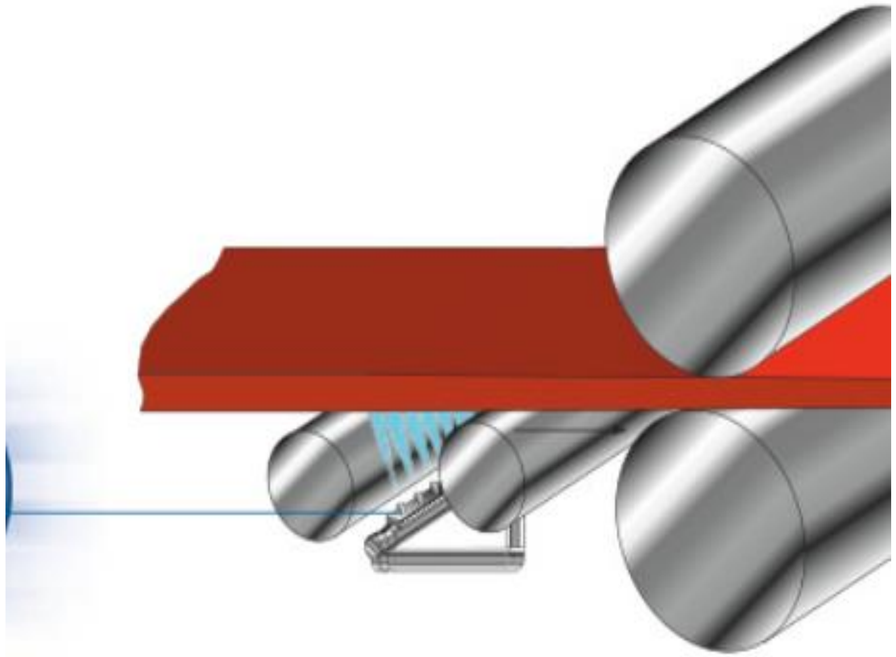
## 결론

결과 및 결론

## 종합결과

핵심인자 정리

# 과제정의



### Scale불량 급증

압연흠, Scratch 등 다양한 불량이 발생했으나, 특히 압연공정에서 scale불량이 급증함

### 불량의 근본 원인

수집된 데이터를 활용하여 다양한 분석을 통해 불량의 근본 원인을 찾고 불량예측 및 개선 기회를 도출

# 분석계획

## 데이터 확인

결측치 확인
기술통계량 확인
이상치 확인

## 데이터 전처리

이상치 처리
결측치 처리
파생변수 생성

## 가설 검정

범주형 설명변수와 목표변수간 분포확인
연속형 설명변수와 목표변수간 분포 확인
설명변수와 목표변수간 차이 검정

## 모델링

다중선형회귀분석
랜덤포레스트
그래디언트 부스팅

# 데이터 현황

| # | Column | Non-Null Count | Dtype |
|---|---|---|---|
| 0 | plate_no | 1000 non-null | object |
| 1 | rolling_date | 1000 non-null | object |
| 2 | scale | 1000 non-null | object |
| 3 | spec_long | 1000 non-null | object |
| 4 | spec_country | 1000 non-null | object |
| 5 | steel_kind | 1000 non-null | object |
| 6 | pt_thick | 1000 non-null | int64 |
| 7 | pt_width | 1000 non-null | int64 |
| 8 | pt_length | 1000 non-null | int64 |
| 9 | hsb | 1000 non-null | object |
| 10 | fur_no | 1000 non-null | object |
| 11 | fur_input_row | 1000 non-null | object |
| 12 | fur_heat_temp | 1000 non-null | int64 |
| 13 | fur_heat_time | 1000 non-null | int64 |
| 14 | fur_soak_temp | 1000 non-null | int64 |
| 15 | fur_soak_time | 1000 non-null | int64 |
| 16 | fur_total_time | 1000 non-null | int64 |
| 17 | fur_ex_temp | 1000 non-null | int64 |
| 18 | rolling_method | 1000 non-null | object |
| 19 | rolling_temp | 1000 non-null | int64 |
| 20 | descaling_count | 1000 non-null | int64 |
| 21 | work_group | 1000 non-null | object |

| | |
|---|---|
| plate_no | 0 |
| rolling_date | 0 |
| scale | 0 |
| spec_long | 0 |
| spec_country | 0 |
| steel_kind | 0 |
| pt_thick | 0 |
| pt_width | 0 |
| pt_length | 0 |
| hsb | 0 |
| fur_no | 0 |
| fur_input_row | 0 |
| fur_heat_temp | 0 |
| fur_heat_time | 0 |
| fur_soak_temp | 0 |
| fur_soak_time | 0 |
| fur_total_time | 0 |
| fur_ex_temp | 0 |
| rolling_method | 0 |
| rolling_temp | 0 |
| descaling_count | 0 |
| work_group | 0 |

현재 1000개의 데이터가 있으며 결측치가 없음을
알수 있습니다

⬇

결측치가 없기 때문에 이상치 및 파생변수의
중요성이 커지게 됩니다.
이상치를 제거하거나 파생변수를 생성하기 위해서는
많은 도메인지식이 필요합니다

# 이상치 확인

| | pt_thick | pt_width | pt_length | fur_heat_temp | fur_heat_time | fur_soak_temp | fur_soak_time | fur_total_time | fur_ex_temp | rolling_temp |
|---|---|---|---|---|---|---|---|---|---|---|
| count | 1000.00000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 26.78200 | 2831.900000 | 36788.200000 | 1157.245000 | 85.972000 | 1150.928000 | 71.720000 | 238.589000 | 1150.928000 | 934.637000 |
| std | 18.13757 | 494.081478 | 13912.387116 | 21.245007 | 26.346297 | 17.344384 | 20.602137 | 38.194828 | 17.344384 | 96.598015 |
| min | 12.00000 | 1800.000000 | 7900.000000 | 1103.000000 | 55.000000 | 1113.000000 | 35.000000 | 165.000000 | 1113.000000 | 0.000000 |
| 25% | 15.00000 | 2500.000000 | 26650.000000 | 1140.000000 | 66.000000 | 1135.750000 | 57.750000 | 210.000000 | 1135.750000 | 893.750000 |
| 50% | 19.00000 | 2800.000000 | 40400.000000 | 1159.000000 | 75.000000 | 1156.000000 | 66.000000 | 230.000000 | 1156.000000 | 948.000000 |
| 75% | 34.00000 | 3100.000000 | 49100.000000 | 1173.000000 | 102.250000 | 1164.000000 | 81.000000 | 263.000000 | 1164.000000 | 991.000000 |
| max | 100.00000 | 4600.000000 | 54900.000000 | 1206.000000 | 158.000000 | 1185.000000 | 145.000000 | 362.000000 | 1185.000000 | 1078.000000 |

Rolling_temp의 min이 0이다

⬇

가열된 슬라브의 온도가 0일수가 없다

# 파생변수 생성

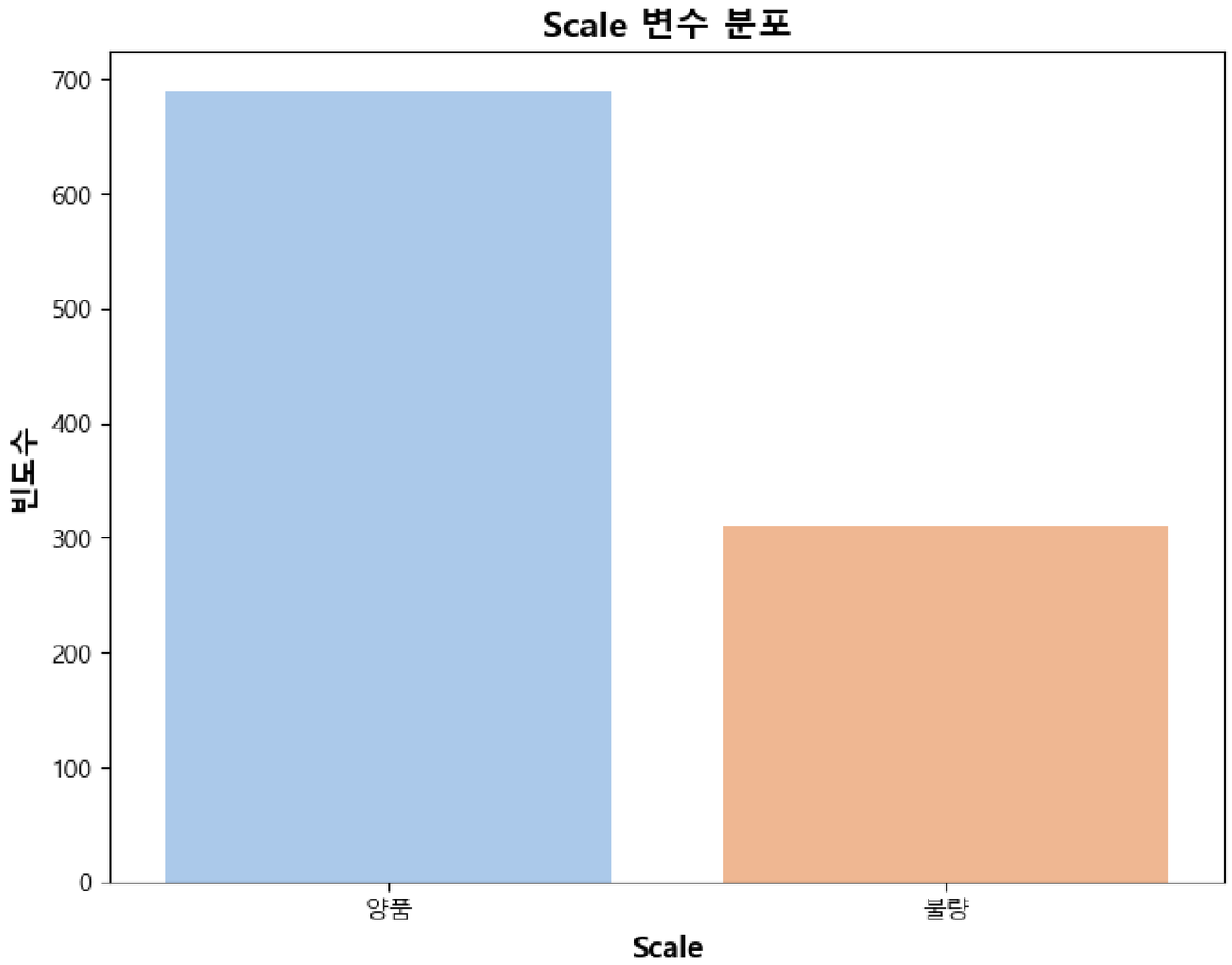
| | scale | spec_long | spec_country | steel_kind | pt_thick | pt_width | pt_length | hsb | fur_heat_temp | fur_heat_time | fur_soak_temp | fur_soak_time | fur_ex_temp | rolling_temp | descaling_count | work_group | pre_time |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | AB/EH32-TM | 미국 | T | 32 | 3700 | 15100 | 적용 | 1144 | 116 | 1133 | 59 | 1133 | 934.0 | 8 | 1조 | 84 |
| **1** | 0 | AB/EH32-TM | 미국 | T | 32 | 3700 | 15100 | 적용 | 1144 | 122 | 1135 | 53 | 1135 | 937.0 | 8 | 1조 | 63 |
| **2** | 0 | NV-E36-TM | 영국 | T | 33 | 3600 | 19200 | 적용 | 1129 | 116 | 1121 | 55 | 1121 | 889.0 | 8 | 1조 | 87 |
| **3** | 0 | NV-E36-TM | 영국 | T | 33 | 3600 | 19200 | 적용 | 1152 | 125 | 1127 | 68 | 1127 | 885.0 | 8 | 1조 | 73 |
| **4** | 0 | BV-EH36-TM | 프랑스 | T | 38 | 3100 | 13300 | 적용 | 1140 | 134 | 1128 | 48 | 1128 | 873.0 | 8 | 1조 | 64 |

```
#예열대 시간
df_raw["pre_time"] = df_raw["fur_total_time"] - df_raw["fur_heat_time"] - df_raw["fur_soak_time"]
```

가열로 총 재로시간에서 가열대 재로시간과 균열대 재로시간을 뺀

'예열대 시간'이라는 파생변수를 만들었습니다

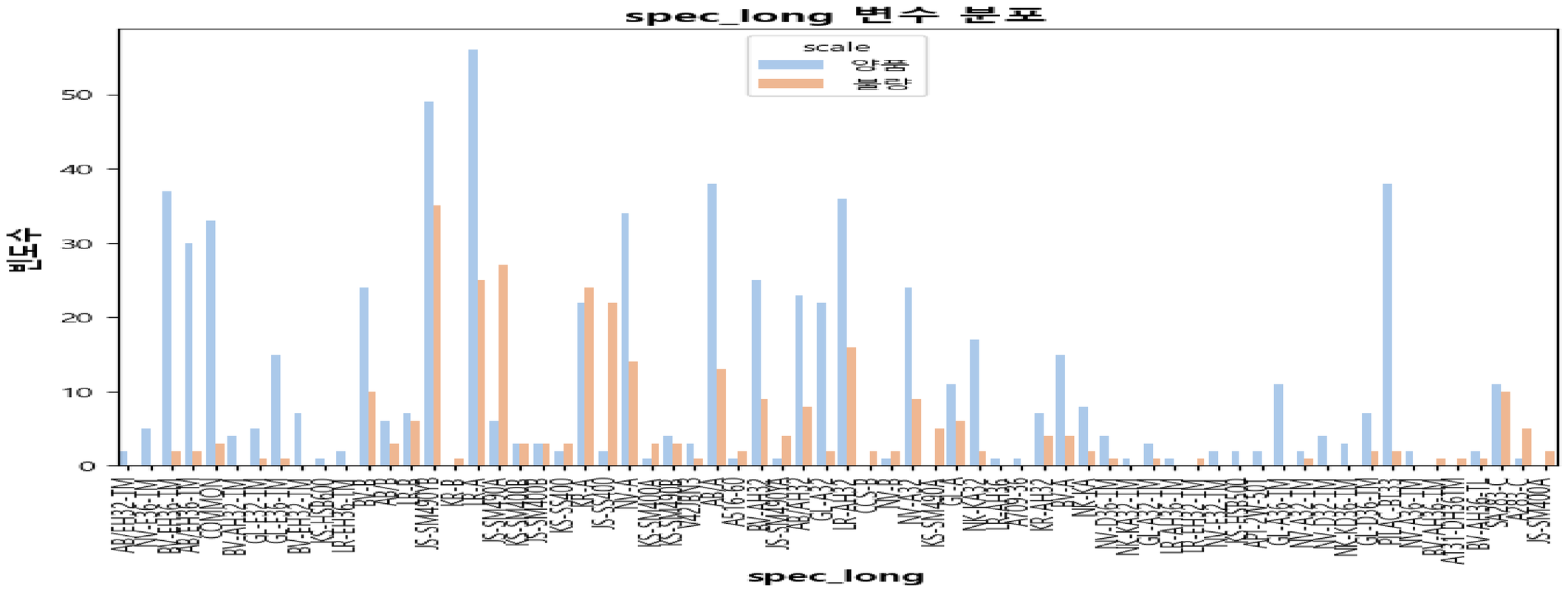
# 탐색적 분석




목표변수(scale)의 분포확인

Scale 불량율 자체가 높음을
알 수 있다

# 탐색적 분석

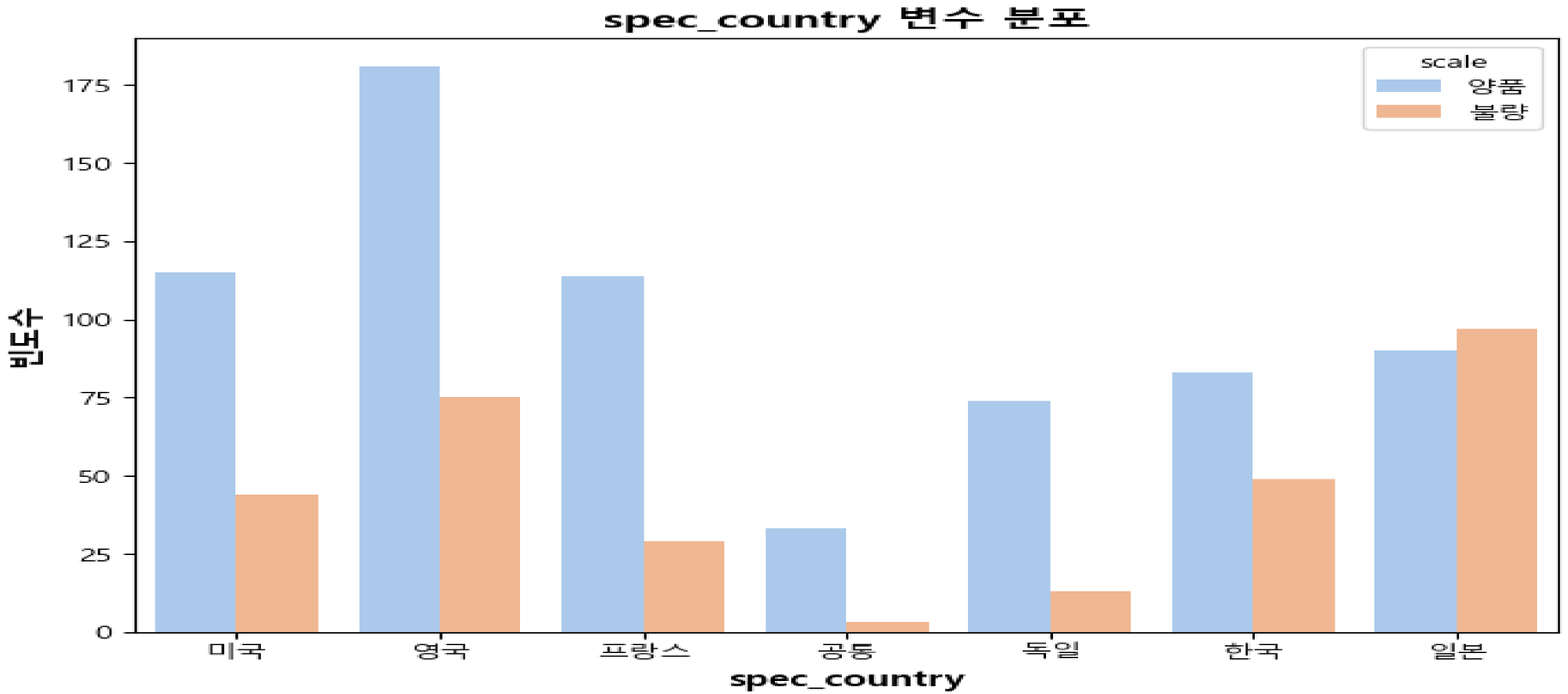



목표변수(scale)와 범주형 설명변수(spec_long)의 분포확인

규격별로 불량율이 다름을 알 수 있습니다.

그런데 1000개의 데이터에 이렇게 많은 규격이 의미가 없다고 생각이 듭니다

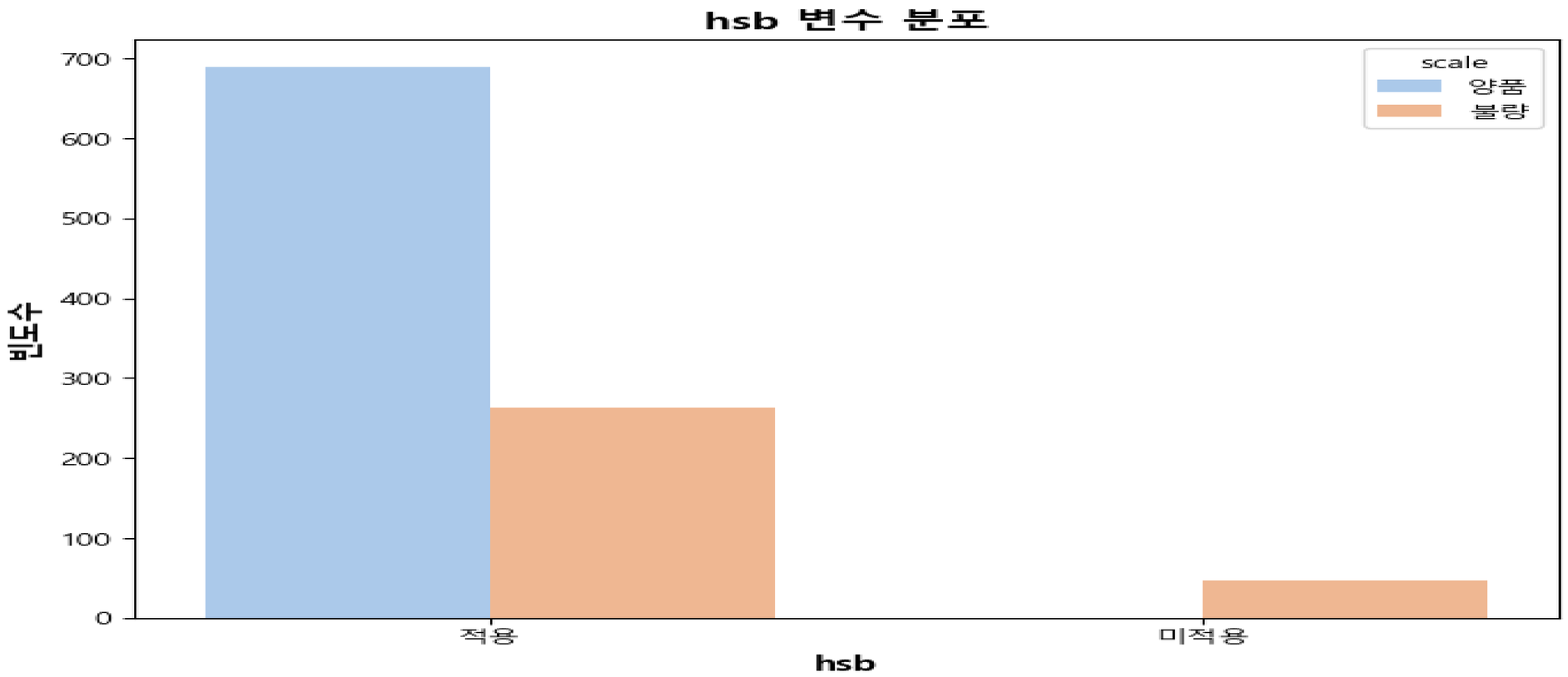
# 탐색적 분석




목표변수(scale)와 범주형 설명변수(spec_country)의 분포확인

**한국도 불량율이 굉장히 높으편이나, 일본은 불량율이 양품율보다 높음을 볼 수 있다**

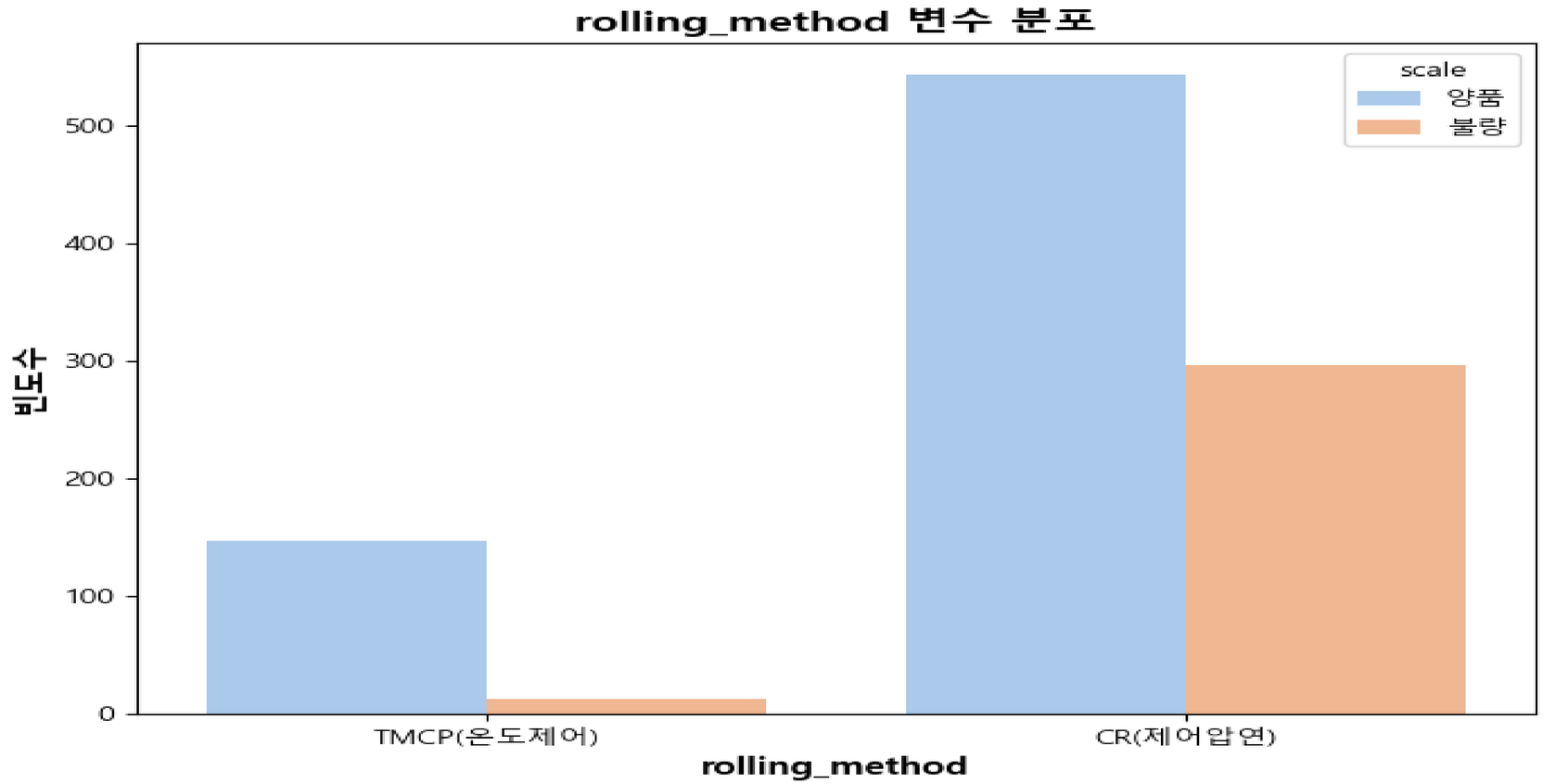
# 탐색적 분석




목표변수(scale)와 범주형 설명변수(hsb)의 분포확인

hsb를 미적용할 시, 모두 불량이 나왔음을 알 수 있습니다

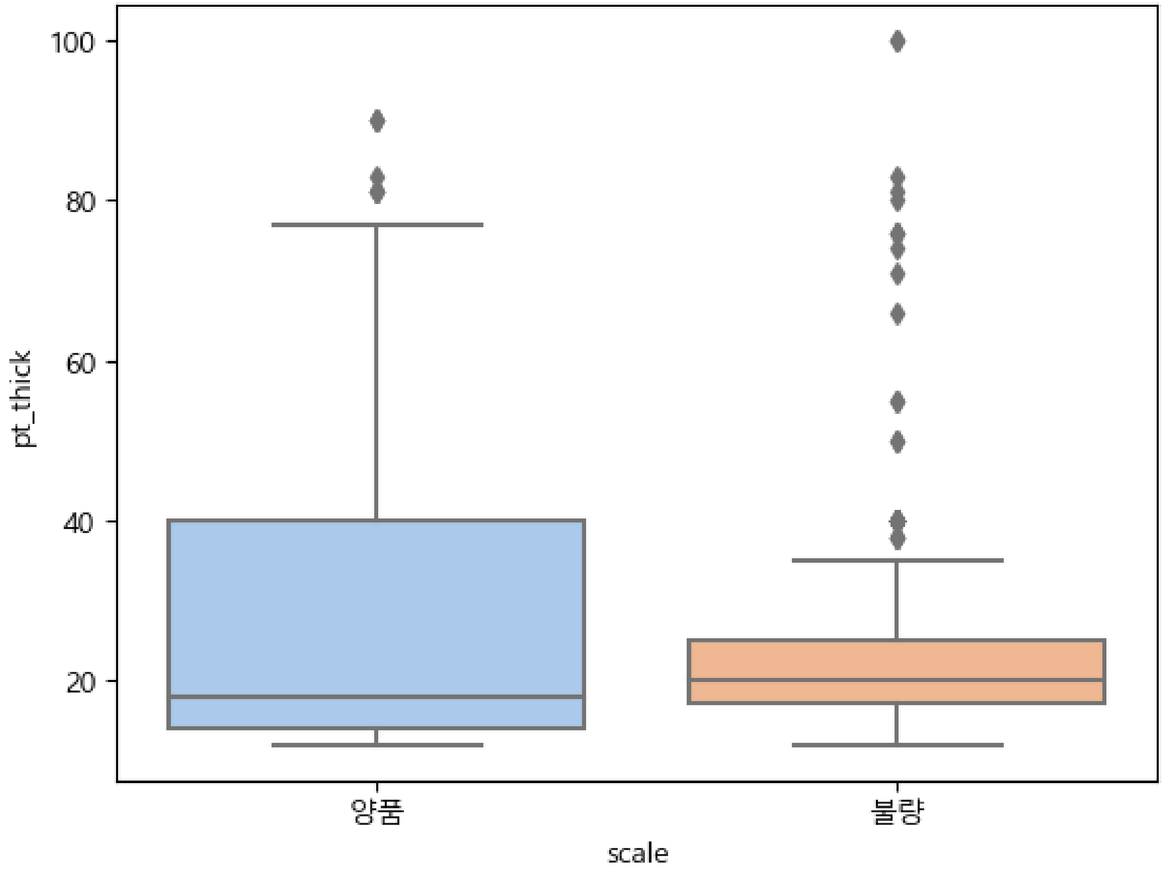
# 탐색적 분석




목표변수(scale)와 범주형 설명변수(rolling_method)의 분포확인

TMCP(온도제어)방식이 CR(제어압연)방식보다 불량율이 현저히 낮음을 알 수 있습니다

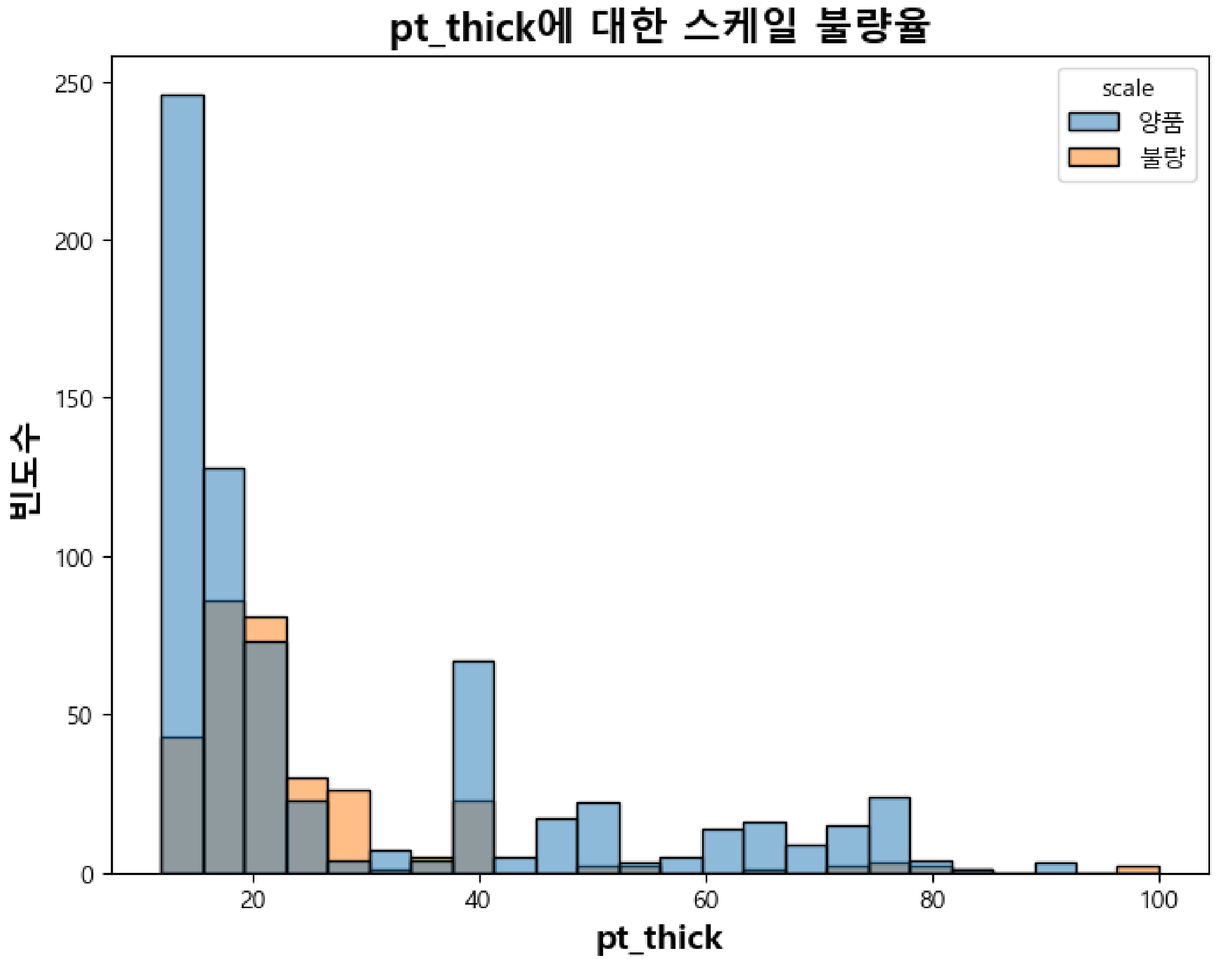
# 탐색적 분석




목표변수(scale)와 연속형 설명변수(pt_thick)의 분포확인

Box plot을 확인했을 때, 불량율에
이상치가 있어보이지만,
후판공정에 대해 찾아보았을 때
후판의 두께는 평균적으로
10~30까지 많이 사용하지만
100도 있다 했습니다 따라서
100은 이상치가 아니라고
판단하였습니다

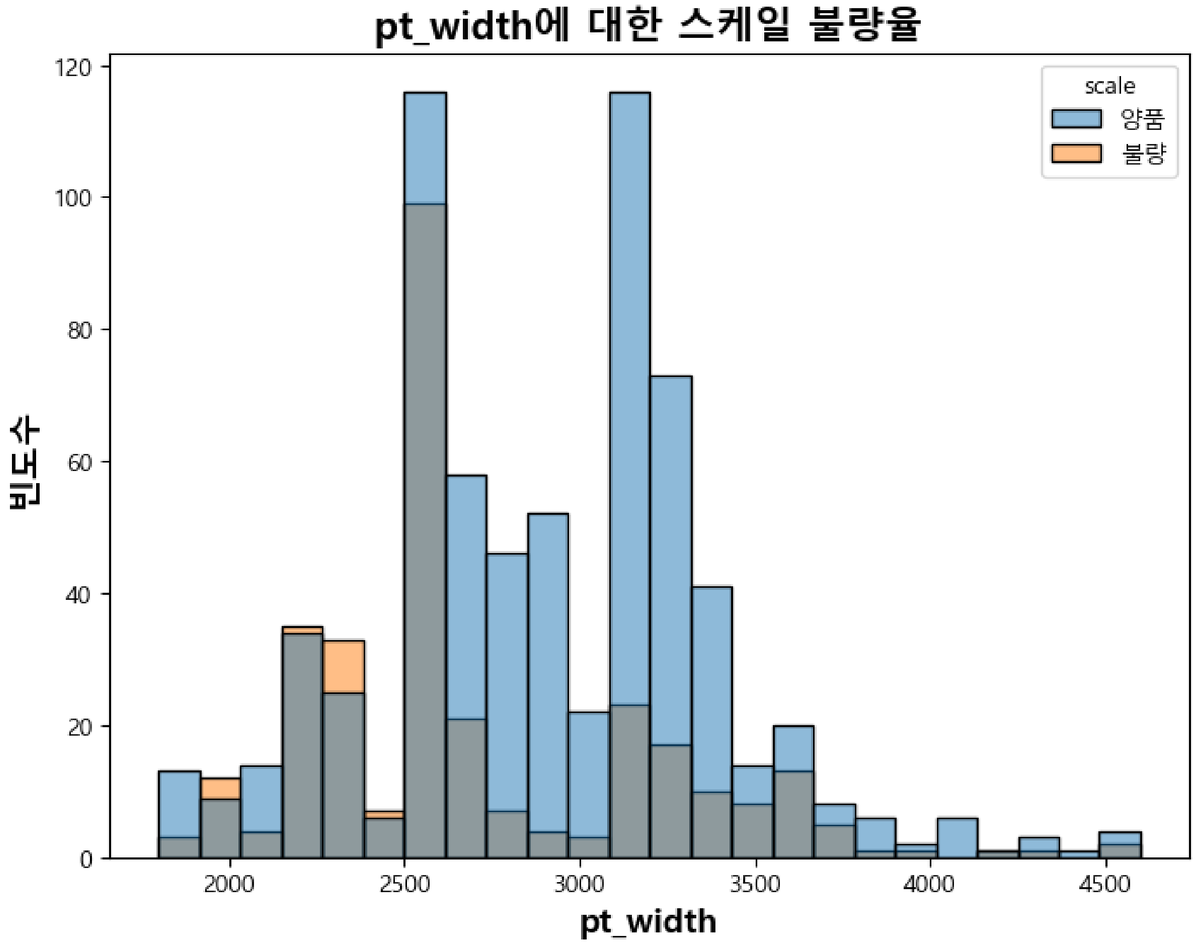
# 탐색적 분석




목표변수(scale)와 연속형 설명변수의 분포확인

후판의 두께가 15~25일때 불량율이 가장 높았으며, 그 이상일 경우 불량율이 줄어든다

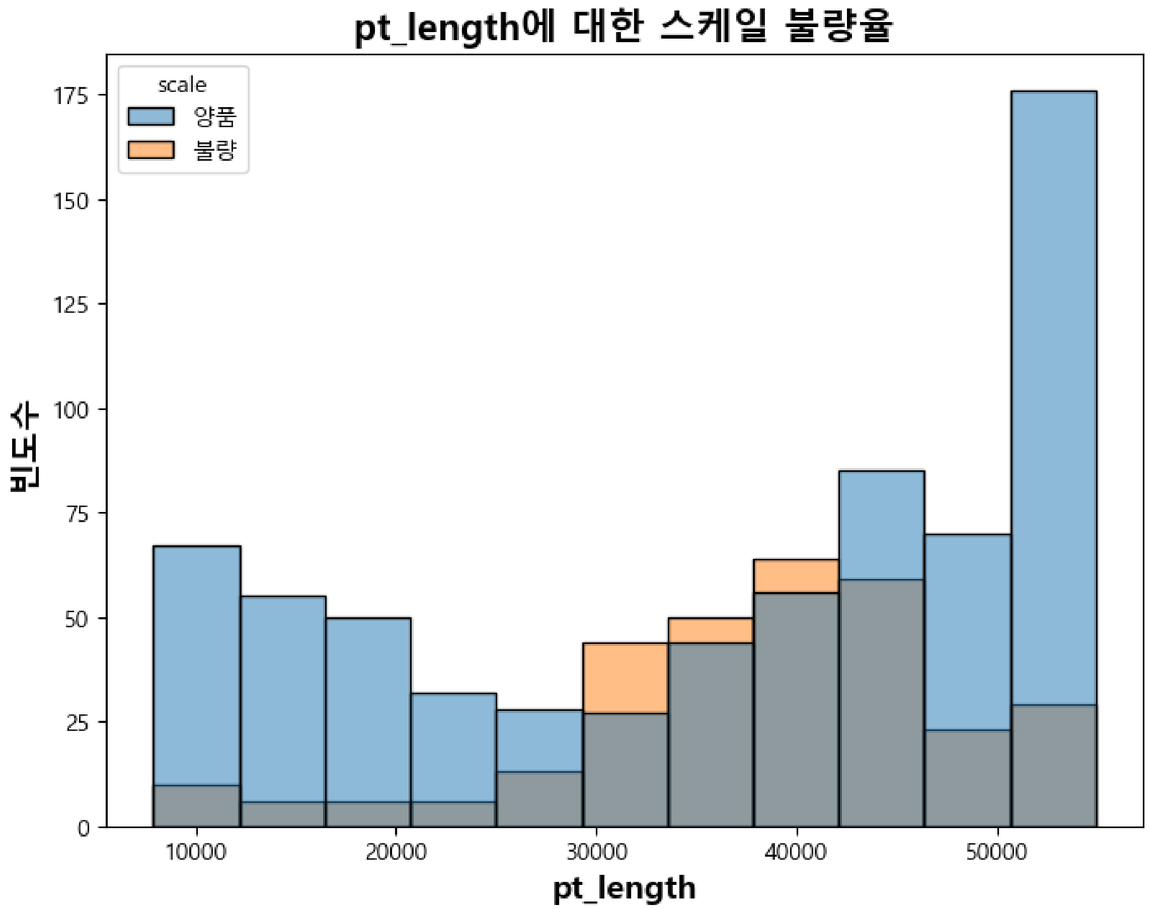
# 탐색적 분석




목표변수(scale)와 연속형 설명변수의 분포확인

후판의 폭은 작을 수록 불량율이 높음을 볼 수 있고, 양품데이터와 불량데이터 모두 정규분포를 띔을 볼 수 있다

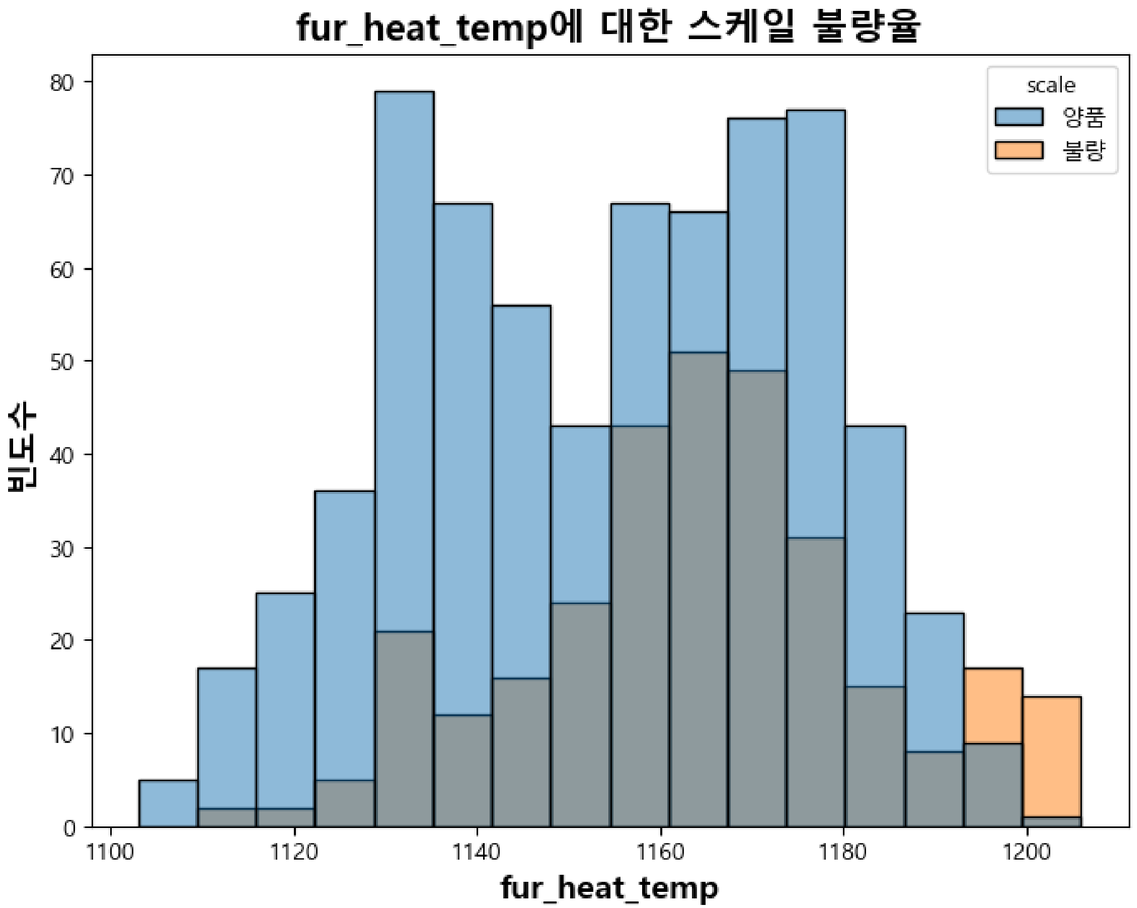
# 탐색적 분석




목표변수(scale)와 연속형 설명변수의 분포확인

후판의 길이가 30m기준으로 불량율이 높아짐을 알 수 있습니다

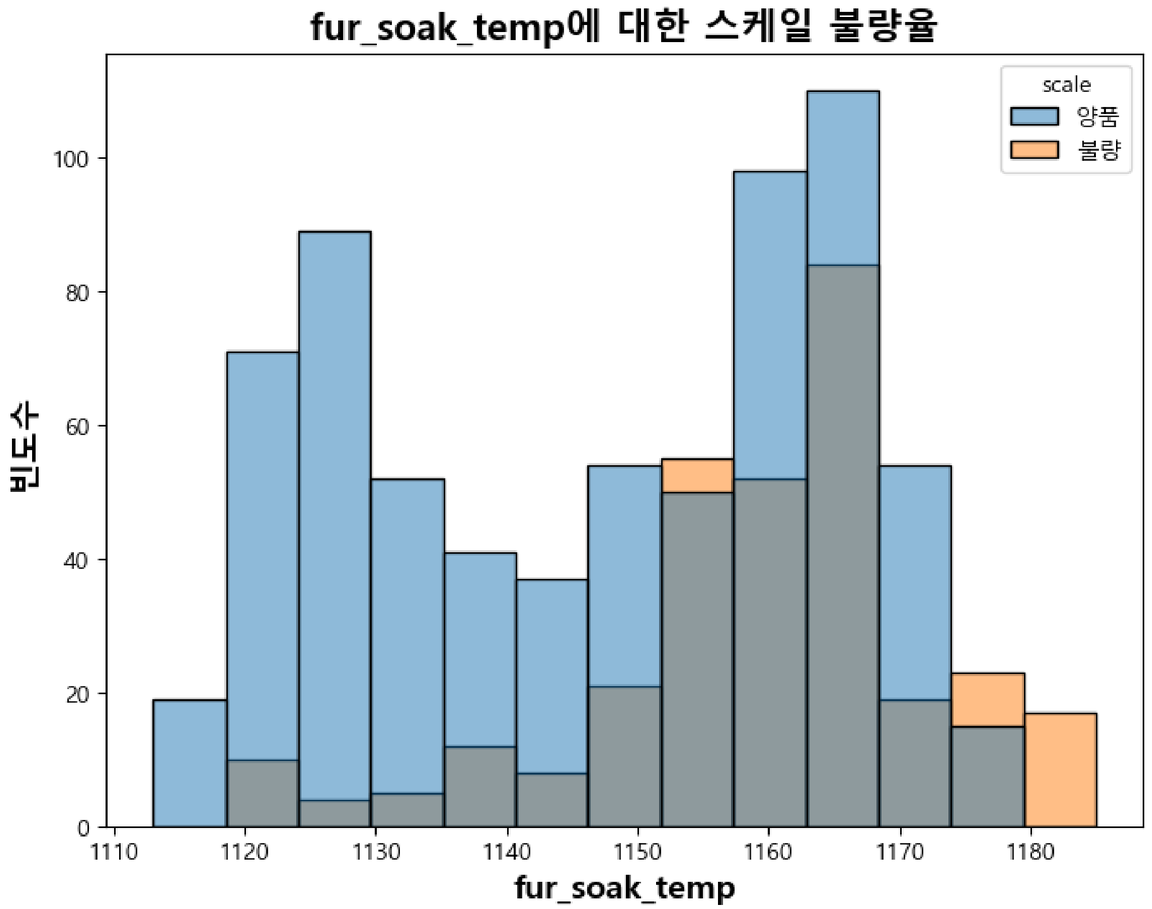
# 탐색적 분석




목표변수(scale)와 연속형 설명변수의 분포확인

가열로 가열대 소재온도가 높을 수록 불량율이 높음을 볼 수 있다. 또한 불량율이 정규분포를 띄는것을 볼 수 있다

# 탐색적 분석

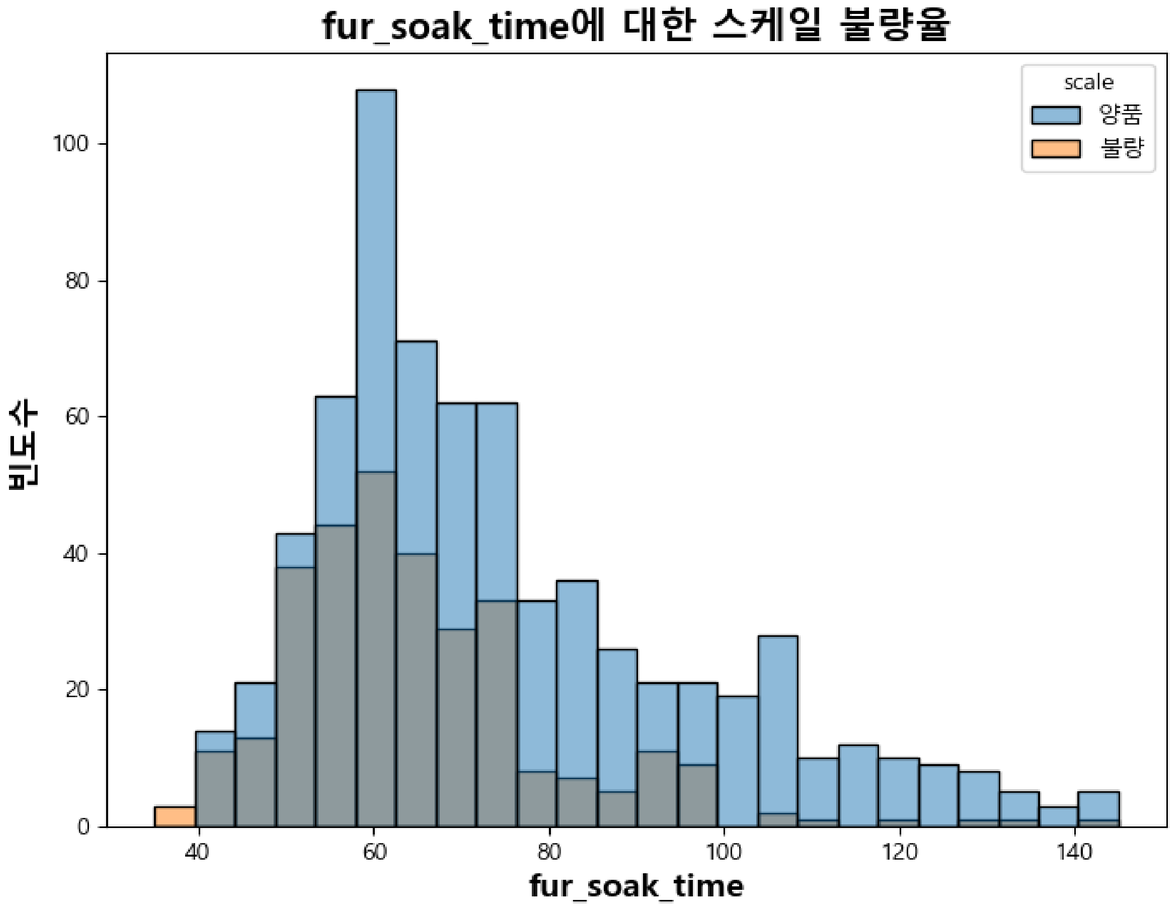



목표변수(scale)와 연속형 설명변수의 분포확인

균열대의 온도가 1155도를 기준으로

불량율이 높아짐을 볼 수 있습니다

# 탐색적 분석



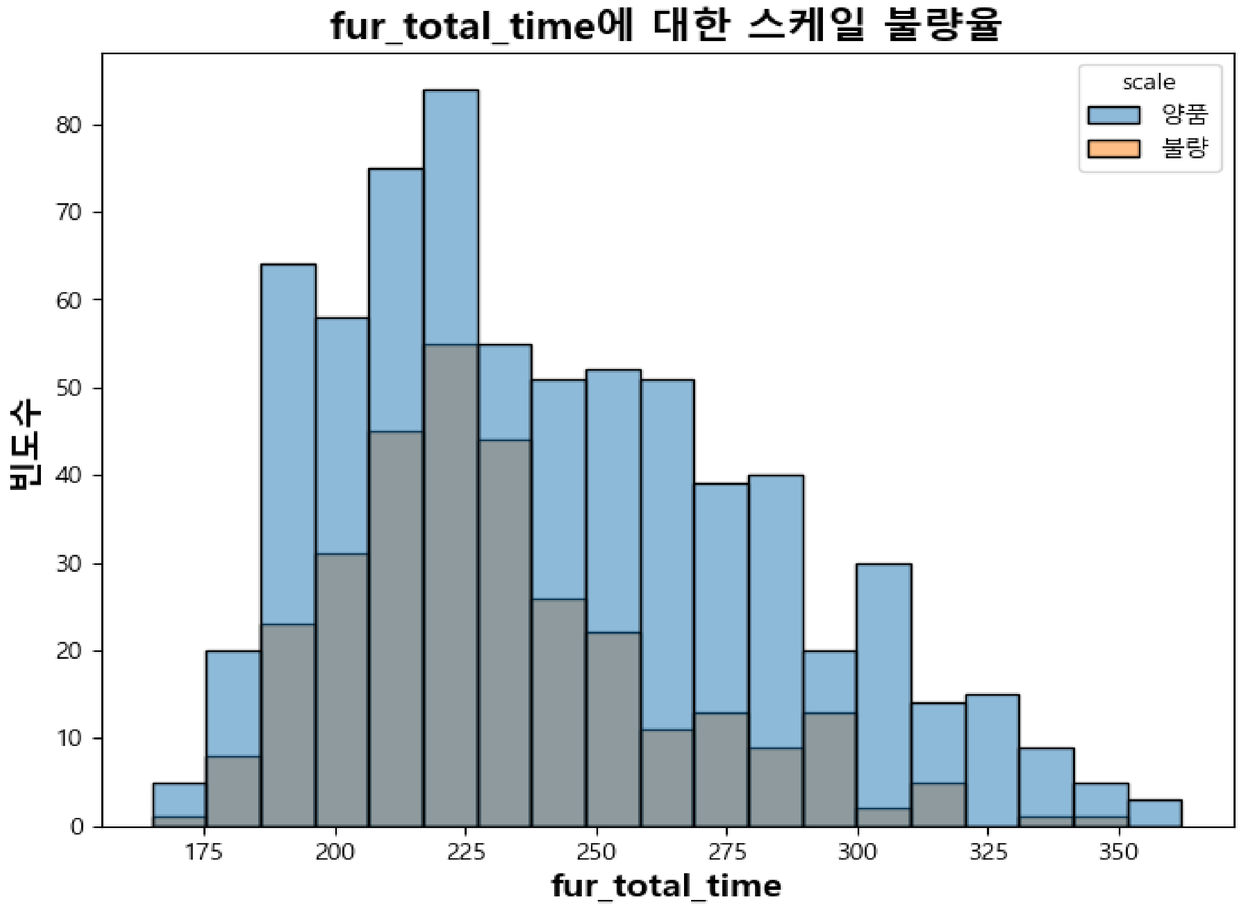

목표변수(scale)와 연속형 설명변수의 분포확인

균열대의 재로시간은 40분 이하는 불량이 많으며

이상인경우의 구간은 불량율이 비슷하며

정규분포를 띔을 볼 수 있다

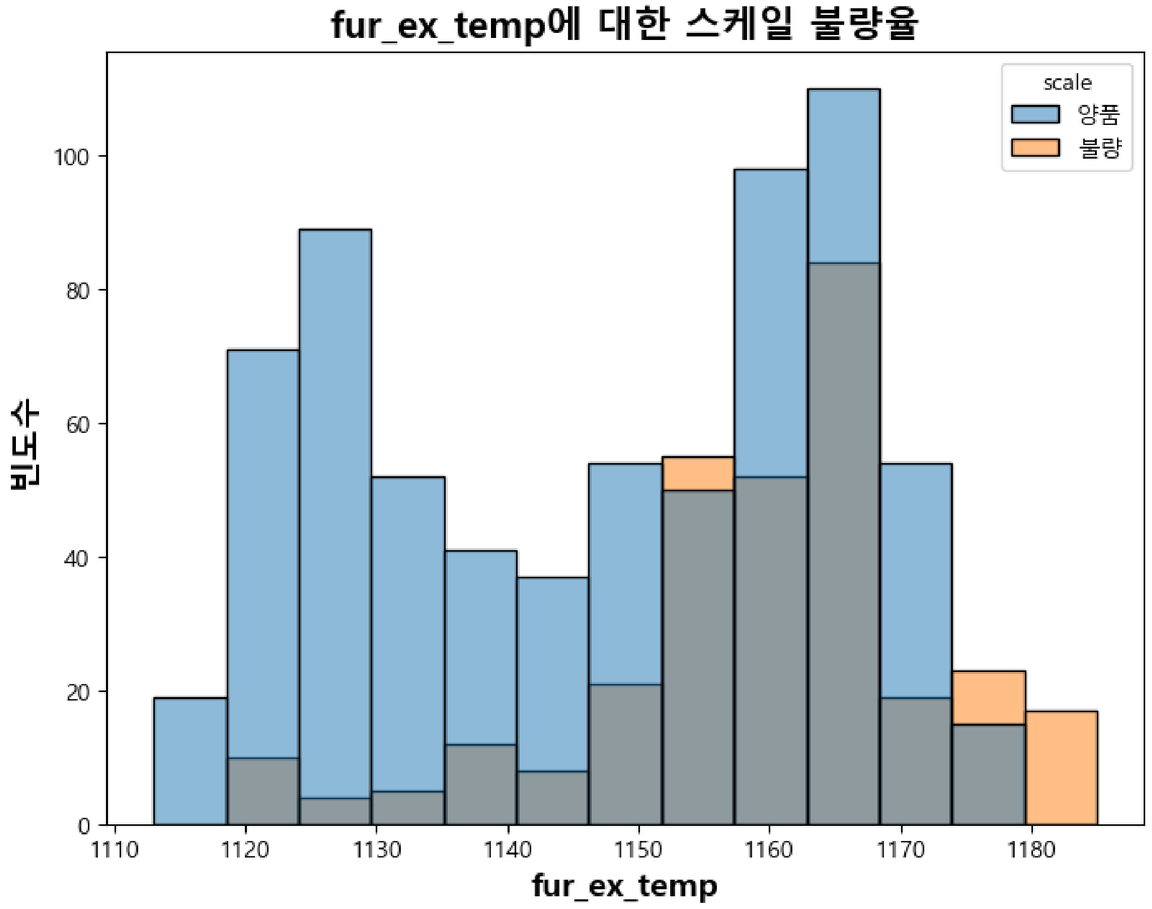
# 탐색적 분석




목표변수(scale)와 연속형 설명변수의 분포확인

가열로 총 재로시간은 구간마다의 불량율이 비슷하며, 정규분포를 띔을 볼 수 있다

# 탐색적 분석



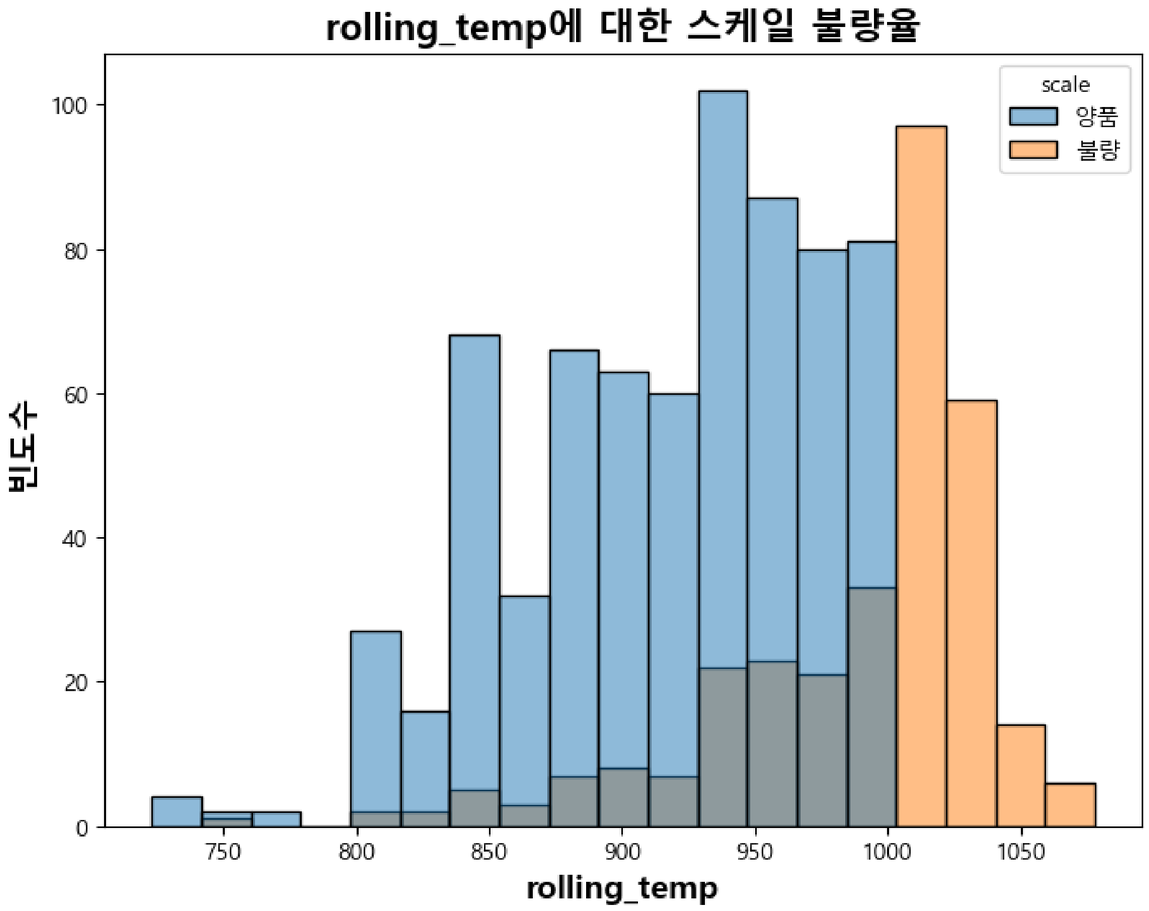

목표변수(scale)와 연속형 설명변수의 분포확인

균열대 소재온도와 비슷하게

가열로 추출온도는 1155도를 기점으로

불량율이 높아지는 것을 볼 수 있다

# 탐색적 분석




목표변수(scale)와 연속형 설명변수의 분포확인

압연온도는 1000도 이상에서는

불량만을 보임을 알 수 있다

# 카이제곱 검정

카이제곱 검정 통계량: 68.38385440143148
p-value: 8.766804557877572e-13

***H0 : 강종 별 Scale 불량에 차이가 없다.***

***H1 : 강종 별 Scale 불량에 차이가 있다. = 유의미***

***p_value 값이 0.05미만이므로, 대립가설 채택 = 유의미한 설명변수***

카이제곱 검정 통계량: 77.7104243330205
p-value: 1.1931000837493562e-18

***H0 : HSB 적용 여부 별 Scale 불량에 차이가 없다.***

***H1 : HSB 적용 여부 별 Scale 불량에 차이가 있다. = 유의미***

***p_value 값이 0.05미만이므로, 대립가설 채택 = 유의미한 설명변수***

카이제곱 검정 통계량: 106.4133937362546
p-value: 5.985184936341528e-25

***H0 : 가열로 호기 별 Scale 불량에 차이가 없다.***

***H1 : 가열로 호기 별 Scale 불량에 차이가 있다. = 유의미***

***p_value 값이 0.05이상이므로, 귀무가설 채택 = 무의미한 설명변수***

***==> 가열로 호기는 공정 변수가 아니므로 공정에 영향을 미치지 않는다고 생각하였다. 따라서 drop 시킨다.***

**범주형 목표변수와 범주형 설명변수간의 검정**

**카이제곱검정으로 추출된 변수는**

**rolling_date, fur_input_row, fur_no**

# 로지스틱 회귀

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | scale | R-squared: | 0.294 |
| Model: | OLS | Adj. R-squared: | 0.286 |
| Method: | Least Squares | F-statistic: | 36.03 |
| Date: | Tue, 07 Mar 2023 | Prob (F-statistic): | 8.96e-48 |
| Time: | 15:05:08 | Log-Likelihood: | -330.57 |
| No. Observations: | 700 | AIC: | 679.1 |
| Df Residuals: | 691 | BIC: | 720.1 |
| Df Model: | 8 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 0.3034 | 0.015 | 20.529 | 0.000 | 0.274 | 0.332 |
| pt_thick | -0.1620 | 0.036 | -4.520 | 0.000 | -0.232 | -0.092 |
| pt_width | -0.1002 | 0.017 | -5.932 | 0.000 | -0.133 | -0.067 |
| pt_length | -0.0675 | 0.034 | -2.013 | 0.044 | -0.133 | -0.002 |
| fur_heat_temp | -0.0607 | 0.025 | -2.473 | 0.014 | -0.109 | -0.013 |
| fur_heat_time | 0.0691 | 0.016 | 4.228 | 0.000 | 0.037 | 0.101 |
| fur_soak_temp | 0.1618 | 0.017 | 9.642 | 0.000 | 0.129 | 0.195 |
| fur_soak_time | 0.0030 | 0.019 | 0.159 | 0.874 | -0.034 | 0.041 |
| fur_ex_temp | 0.1618 | 0.017 | 9.642 | 0.000 | 0.129 | 0.195 |
| descaling_count | -0.2309 | 0.031 | -7.535 | 0.000 | -0.291 | -0.171 |

| | | | |
|---|---|---|---|
| Omnibus: | 51.293 | Durbin-Watson: | 1.894 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 44.765 |
| Skew: | 0.545 | Prob(JB): | 1.90e-10 |
| Kurtosis: | 2.412 | Cond. No. | 1.99e+16 |

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | scale | R-squared: | 0.294 |
| Model: | OLS | Adj. R-squared: | 0.287 |
| Method: | Least Squares | F-statistic: | 41.23 |
| Date: | Tue, 07 Mar 2023 | Prob (F-statistic): | 1.36e-48 |
| Time: | 15:05:08 | Log-Likelihood: | -330.58 |
| No. Observations: | 700 | AIC: | 677.2 |
| Df Residuals: | 692 | BIC: | 713.6 |
| Df Model: | 7 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 0.3034 | 0.015 | 20.544 | 0.000 | 0.274 | 0.332 |
| pt_thick | -0.1626 | 0.036 | -4.561 | 0.000 | -0.233 | -0.093 |
| pt_width | -0.1002 | 0.017 | -5.934 | 0.000 | -0.133 | -0.067 |
| pt_length | -0.0673 | 0.033 | -2.010 | 0.045 | -0.133 | -0.002 |
| fur_heat_temp | -0.0599 | 0.024 | -2.502 | 0.013 | -0.107 | -0.013 |
| fur_heat_time | 0.0687 | 0.016 | 4.265 | 0.000 | 0.037 | 0.100 |
| fur_soak_temp | 0.1604 | 0.014 | 11.070 | 0.000 | 0.132 | 0.189 |
| fur_ex_temp | 0.1604 | 0.014 | 11.070 | 0.000 | 0.132 | 0.189 |
| descaling_count | -0.2315 | 0.030 | -7.624 | 0.000 | -0.291 | -0.172 |

| | | | |
|---|---|---|---|
| Omnibus: | 51.327 | Durbin-Watson: | 1.894 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 44.801 |
| Skew: | 0.545 | Prob(JB): | 1.87e-10 |
| Kurtosis: | 2.412 | Cond. No. | 5.69e+16 |

**p-value가 0.05 이하가 되게끔 변수를 선택한다**

**제외된 변수**

**fur_soak_time**

# 모델링

```
Train 예측결과
 96     0
792     0
218     0
967     0
170     1
dtype: int32

Confusion Matrix:
[[440  44]
 [ 92 124]]

Test 예측결과
 681     0
990     0
155     1
768     0
438     0
dtype: int32

Confusion Matrix:
[[193  13]
 [ 54  40]]
```

```
Train 예측/분류 결과
Accuracy: 0.806

Confusion Matrix:
[[440  44]
 [ 92 124]]

              precision    recall  f1-score   support

           0      0.827     0.909     0.866       484
           1      0.738     0.574     0.646       216

    accuracy                          0.806       700
   macro avg      0.783     0.742     0.756       700
weighted avg      0.800     0.806     0.798       700

Test 예측/분류 결과
Accuracy: 0.777

Confusion Matrix:
[[193  13]
 [ 54  40]]

              precision    recall  f1-score   support

           0      0.781     0.937     0.852       206
           1      0.755     0.426     0.544        94

    accuracy                          0.777       300
   macro avg      0.768     0.681     0.698       300
weighted avg      0.773     0.777     0.756       300
```

## 로지스틱 모델링

**Train : 정확도 80.6%, 재현율 57.4%**

**Test : 정확도 77.7%, 재현율 42.6%**

**Train에 대한 학습은 적절히 이뤄졌지만,**

**Test에 대한 분류 결과를 보면,**

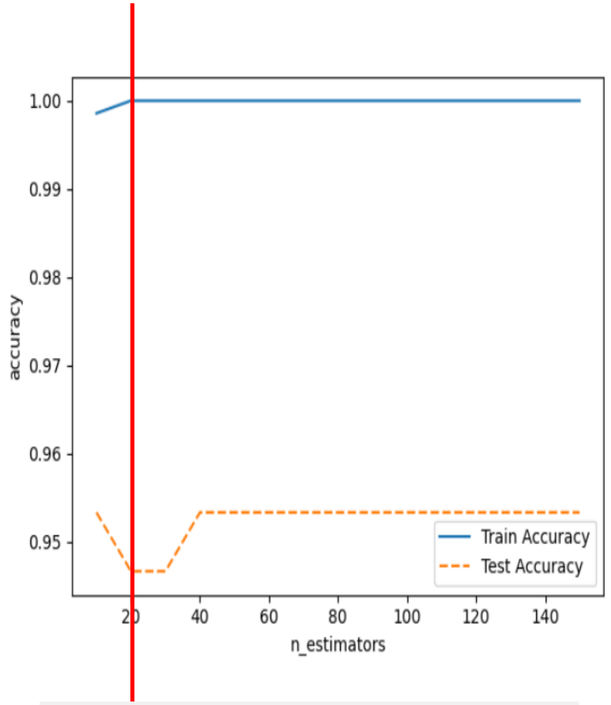
**"1"의 Recall값이 42.6%이다. 즉,**

**클래스를 제대로 분류하지 못했다.**

**이 모델은 예측/분류 모델로 사용하기에**

**부적합하다고 판단할 수 있다.**

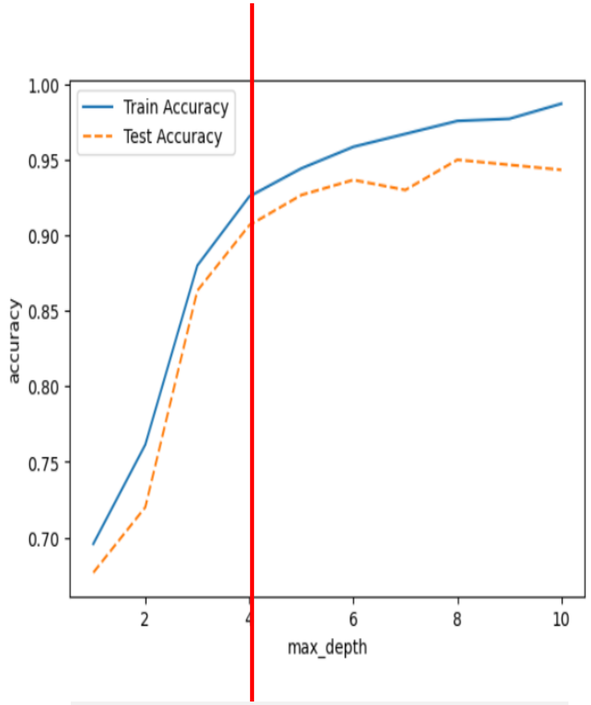
# 랜덤포레스트




## n_estimators

20선택

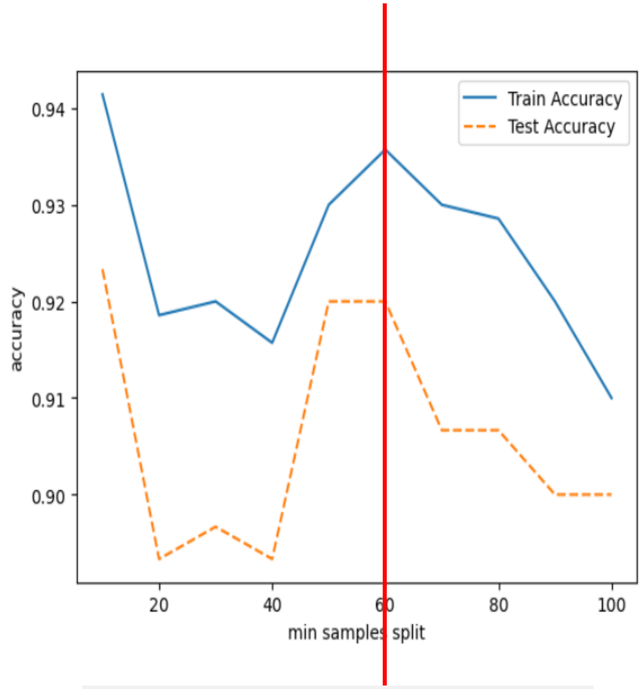
Train성능이 과적합 되지않으며
Test성능이 높은 경우 선택




## max_depth

4선택

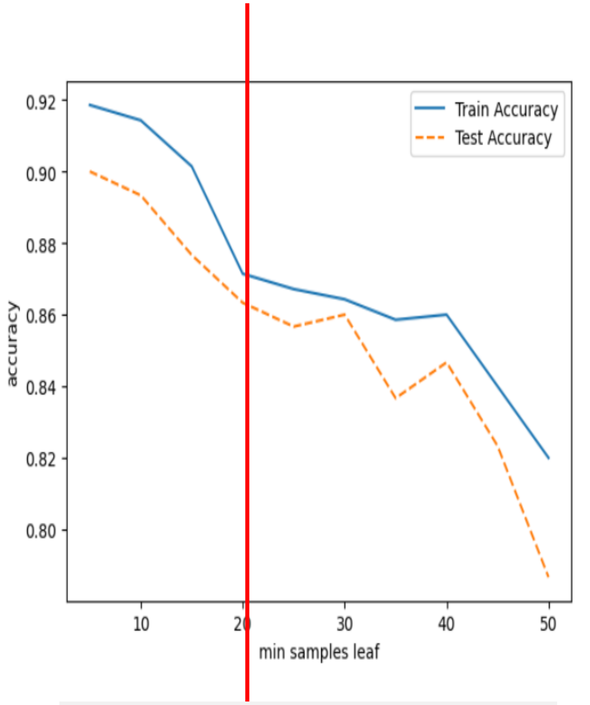
Test성능이 높은 경우 선택




## min_simple_split

60선택

Test성능이 높으며 train 데이터와
Gap차이가 크지 않은부분 선택




## min_samples_leaf

20선택

Test성능이 높으며 train 데이터와
Gap차이가 크지 않은부분 선택

# 랜덤포레스트

| | scale | Importance |
|---|---|---|
| 8 | rolling_temp | 0.260 |
| 7 | fur_ex_temp | 0.124 |
| 0 | pt_thick | 0.108 |
| 9 | descaling_count | 0.077 |
| 1 | pt_width | 0.059 |
| 6 | fur_soak_time | 0.053 |
| 81 | spec_country_일본 | 0.053 |
| 5 | fur_soak_temp | 0.050 |
| 3 | fur_heat_temp | 0.040 |
| 2 | pt_length | 0.028 |

```
Accuracy on training set:0.871
Accuracy on test set:0.863
Confusion matrix:
[[202   1]
 [ 23  74]]
              precision    recall  f1-score   support

           0       0.898     0.995     0.944       203
           1       0.987     0.763     0.860        97

    accuracy                         0.920       300
   macro avg       0.942     0.879     0.902       300
weighted avg       0.927     0.920     0.917       300
```

변수중요도는

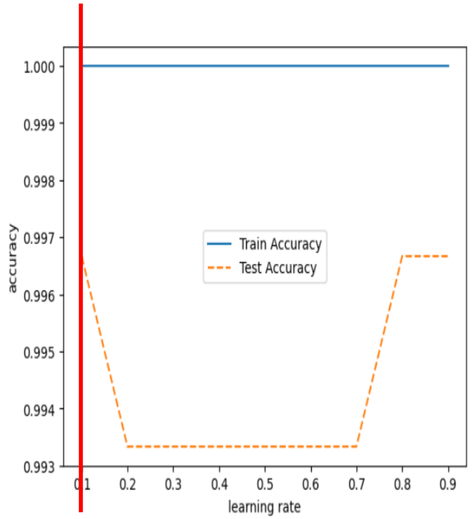
rolling_temp > fur_ex_temp > pt_thick > descaling_count

순서로 높았습니다

그리고 train성능은 87.1%이며, test성능은 86.3%로 과적합 되지는 않았지만

성능자체가 떨어지는 것을 볼 수 있습니다

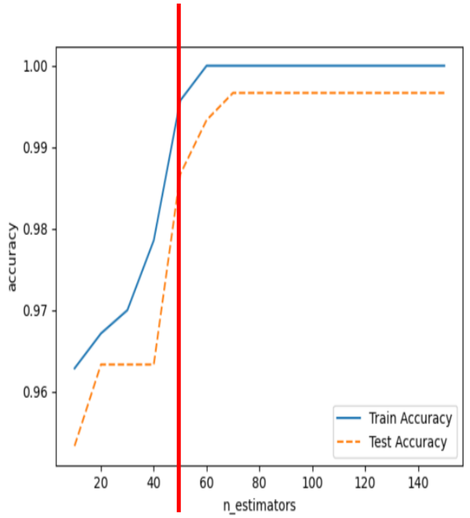
# 그래디언트 부스팅




## Learning_rate

0.1 선택

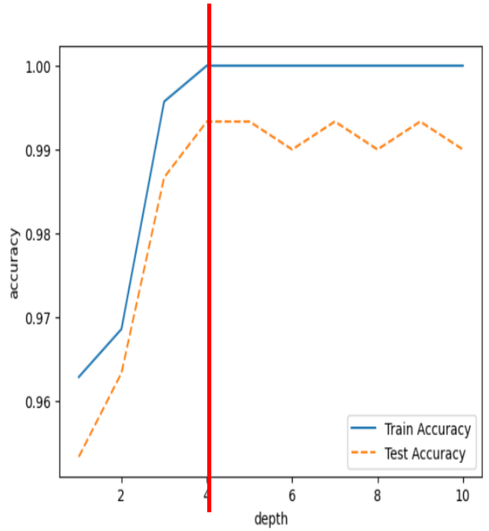
Test성능이 높은 경우
선택




## n_estimators

50 선택

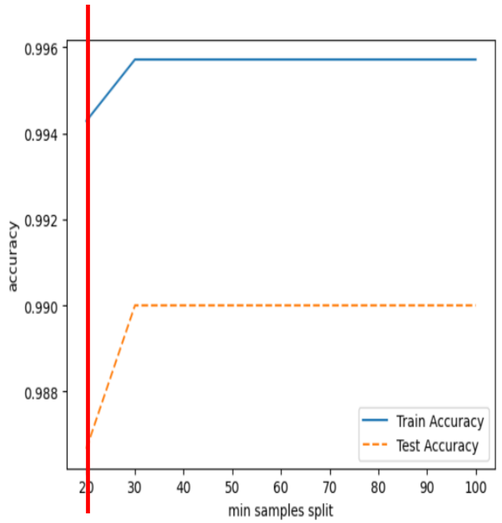
Train성능이 과적합 되지않으며
Test성능이 높은 경우 선택




## max_depth

4 선택

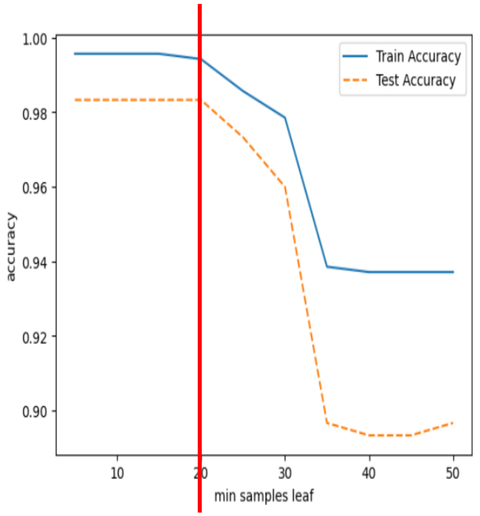
Train성능이 과적합 되지않으며
Test성능이 높은 경우 선택




## min_simple_split

20 선택

Train성능이 과적합 되지않으며
Test성능이 높은 경우 선택




## min_samples_leaf

20 선택

Test성능이 높으며
train 데이터와
Gap차이가 크지
않은부분 선택

# 그래디언트 부스팅

| | Feature | Importance |
|---|---|---|
| 8 | rolling_temp | 0.569 |
| 86 | hsb_미적용 | 0.150 |
| 5 | fur_soak_temp | 0.094 |
| 7 | fur_ex_temp | 0.066 |
| 9 | descaling_count | 0.054 |
| 87 | hsb_적용 | 0.029 |
| 0 | pt_thick | 0.020 |
| 2 | pt_length | 0.014 |
| 90 | work_group_1조 | 0.002 |
| 6 | fur_soak_time | 0.001 |

```
Accuracy on training set: 0.996

Accuracy on test set: 0.980

Confusion matrix:
[[203   0]
 [  6  91]]
              precision    recall  f1-score   support

           0       0.971     1.000     0.985       203
           1       1.000     0.938     0.968        97

    accuracy                          0.980       300
   macro avg       0.986     0.969     0.977       300
weighted avg       0.981     0.980     0.980       300
```

변수중요도는

rolling_temp > hsb_미적용 > fur_soak_templ > fur_ex_temp

순서로 높았습니다

그리고 train성능은 99.6%이며, test성능은98%로 과적합은 일어나지 않았고,

test성능 98%로 좋은 것을 알 수 있습니다

# 결론

GradientBoosting이 가장 높은 정확도와 F1 Score를 보였습니다.

따라서 주요 설명변수를 GradientBoosting을 통해 나온 변수로 판단할 것입니다.

가열대 온도, 가열로 균열대 온도, HSB 적용 여부, 압연 중 Descaling 횟수, Plate 두께입니다

그리고 앞서 탐색적 분석에서 가열로 균열대 온도와 압연온도(FUR_EXTEMP)이 완벽히 정비례해 압연온도를 제거했습니다.

따라서 압연온도 또한 주요 인자라고 볼 수 있습니다. 또한 Logistic으로 나온 중요도를 봤을 때, 각 변수의 온도를 높일지,

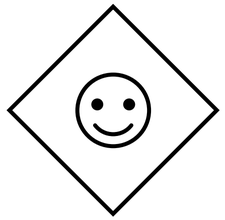
두께의 크기가 얼마나 되야할지, HSB를 적용할 지 말지를 결정할 수 있습니다.

즉, 압연공정에서 SCALE발생을 줄이기 위한 방법은 다음과 같습니다.

1. 가열대 온도를 낮춘다.
2. 가열로 균열대 온도를 낮춘다.
3. HSB를 적용한다.
4. 압연 중 Descailing 횟수를 늘린다.
5. Plate 두께를 얇게 한다.

# 종합 결과

| 변수 | 변수설명 | 변수역할 | 변수형태 | 분석 제외사유 | 탐색적 기법 | | 모델링 기법 | | | | 종점 | 선정 (사유) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 그래프 | 로지스틱 회귀 | x2분석 | 로지스틱 | RF | GB | | |
| plate_no | Plate번호 | ID | 범주형 | 생산된 슬라브의 인덱스이므로 제거한다. | X | | | | | | train set: 0.996<br>test set: 0.980 | 그래디언트 부스팅 (test set의 정확도가 매우 높고 적절하게 나타났기 때문) |
| rolling_date | 열연작업시각 | 날짜 | 연속형 | 작업을 시작한 시각과 공정과는 유의미한 차이가 없다. | X | | | | | | | |
| scale | Scale(산화철) 불량 | 목표변수 | 범주형 | | histogram | | | 선택 X | 선택 X | 선택 O | | |
| spec_long | 제품 규격 | 설명변수 | 범주형 | | histogram | | O | | O | O | | |
| spec_country | 제품 규격 기준국 | 설명변수 | 범주형 | | histogram | | O | | O | O | | |
| steel_kind | 강종 | 설명변수 | 범주형 | | histogram | | O | | O | O | | |
| pt_thick | Plate(후판) 지시두께(mm) | 설명변수 | 연속형 | | boxplot | O | | O | 3 | O | | |
| pt_width | Plate(후판) 지시폭(mm) | 설명변수 | 연속형 | | boxplot | O | | O | 5 | O | | |
| pt_length | Plate(후판) 지시길이(mm) | 설명변수 | 연속형 | | boxplot | O | | O | O | O | | |
| hsb | HSB(Hot Scale Braker)적용여부 | 설명변수 | 범주형 | | histogram | | O | | O | 2 | | |
| fur_no | 가열로 호기 | 설명변수 | 범주형 | 가열로 호기는 공정과정의 변수가 아니다. | histogram | | | | | | | |
| fur_input_row | 가열로 장입열 | 설명변수 | 범주형 | 카이제곱 검정으로 무의미한 변수의 근거가 있다. | histogram | | | | | | | |
| fur_heat_temp | 가열로 가열대 소재온도(℃) | 설명변수 | 연속형 | | boxplot | O | | O | O | O | | |
| fur_heat_time | 가열로 가열대 재로시간(분) | 설명변수 | 연속형 | | boxplot | O | | O | O | O | | |
| fur_soak_temp | 가열로 균열대 소재온도(℃) | 설명변수 | 연속형 | | boxplot | O | | O | O | 3 | | |
| fur_soak_time | 가열로 균열대 재로시간(분) | 설명변수 | 연속형 | 로지스틱 회귀분석으로 유의하지 않으므로 제거 | boxplot | | | | | | | |
| fur_total_time | 가열로 총 재로시간(분) | 설명변수 | 연속형 | 예열대 시간 파생변수생성을 통해 제거를 하였다. | boxplot | | | | | | | |
| fur_ex_temp | 가열로 추출온도((℃),계산치) | 설명변수 | 연속형 | | boxplot | O | | O | 2 | 4 | | |
| rolling_method | 압연방법 | 설명변수 | 범주형 | | histogram | | O | | O | O | | |
| rolling_temp | 압연온도(℃) | 설명변수 | 범주형 | | boxplot, histogram | | O | | 1 | 1 | | |
| descaling_count | 압연Descaling 횟수 | 설명변수 | 연속형 | | boxplot | O | | O | 4 | 5 | | |
| work_group | 작업조 | 설명변수 | 범주형 | | histogram | | O | | O | O | | |
| pre_time | 예열대 시간 | 설명변수 | 연속형 | | | | | | | | | |



# 끝

감사합니다