

후판 공정 스케일 불량 개선 프로젝트

C1조
김채은
장수현
정옥용
정충원
조민정
최기영

2024.11.04

목차

1. 도메인 정보 파악
 2. 데이터 현황 확인 후 전처리
 3. 데이터 탐색적 분석
 4. 분석
 - 4-1. 핵심인자 도출
 - 4-2. 모델링
 5. 개선방안
 6. 배운 점 & 느낀 점
-

1.도메인 정보 파악

Data columns (total 21 columns):

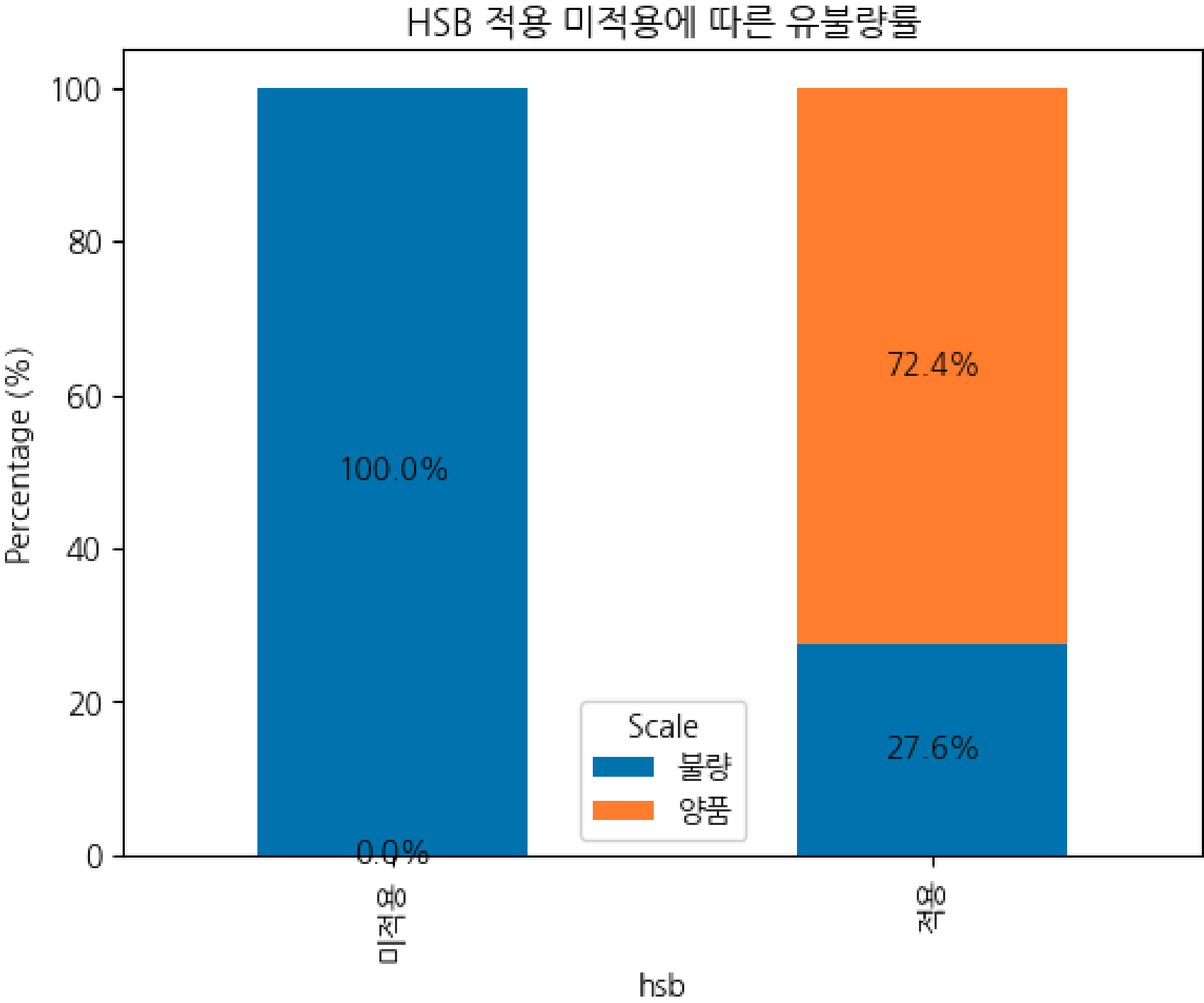
#	Column	Non-Null Count	Dtype
0	plate_no	1000 non-null	object
1	rolling_date	1000 non-null	object
2	scale	1000 non-null	object
3	spec_long	1000 non-null	object
4	spec_country	1000 non-null	object
5	steel_kind	1000 non-null	object
6	pt_thick	1000 non-null	int64
7	pt_width	1000 non-null	int64
8	pt_length	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	rolling_method	1000 non-null	object
18	rolling_temp	1000 non-null	int64
19	descaling_count	1000 non-null	int64
20	work_group	1000 non-null	object

삭제한 변수

- 1.plate_no : Plate 번호
- 2.rolling_date : 열연작업날짜
- 3.spec_long : 제품 규격
- 4.spec_count : 제품 규격 기준국가
- 5.pt_thick : 후판 지시 두께
- 6.pt_width : 후판 지시 폭
- 7.pt_length : 후판 지시 길이
- 8.fur_no : 가열로 호기
- 9.fur_input_row : 가열로 장입열
- 10.fur_total_time : 가열로 총 재로시간

1.도메인 정보 파악

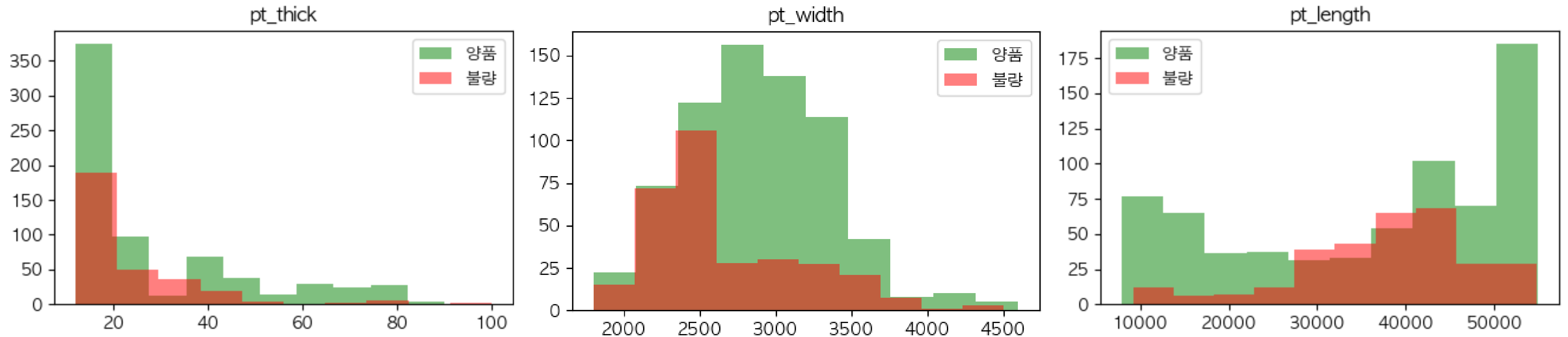
HSB가 적용되지 않은 것은 모두 불량이므로 HSB가 적용된 데이터만 사용



1.도메인 정보 파악

변수 삭제 이유

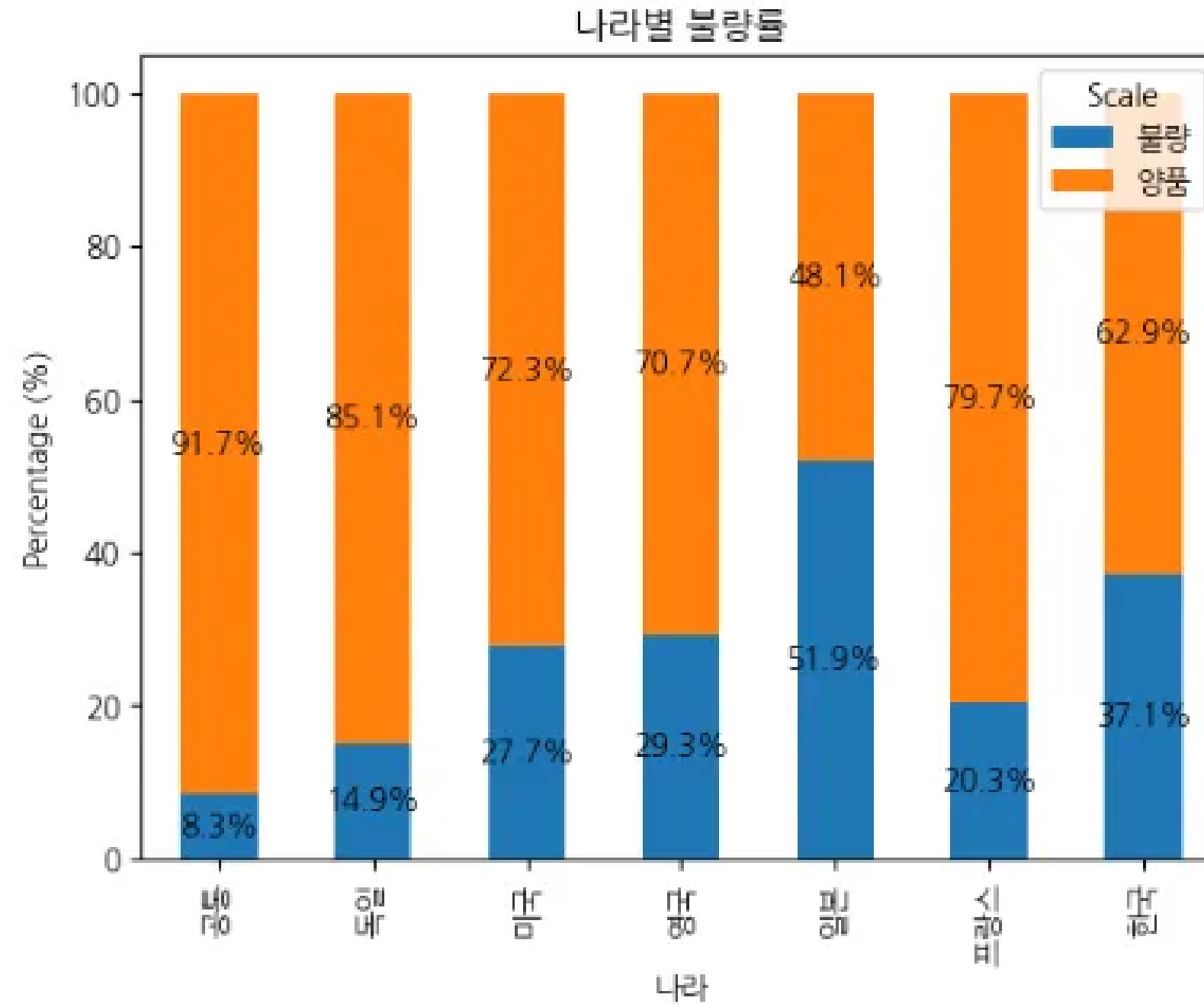
- pt_thick, pt_width, pt_length



주문을 의뢰한 회사는 MTO 방식 (선 주문 후 생산)이기 때문에
해당 변수들은 제어가 불가능, 따라서 무의미하다고 판단하여 삭제

1.도메인 정보 파악

- spec_count



제조 국가에서 요구하는 제조 규격에 따라 불량률이 상이
→ 이를 우리가 직접적으로 개선 불가능하다고 판단

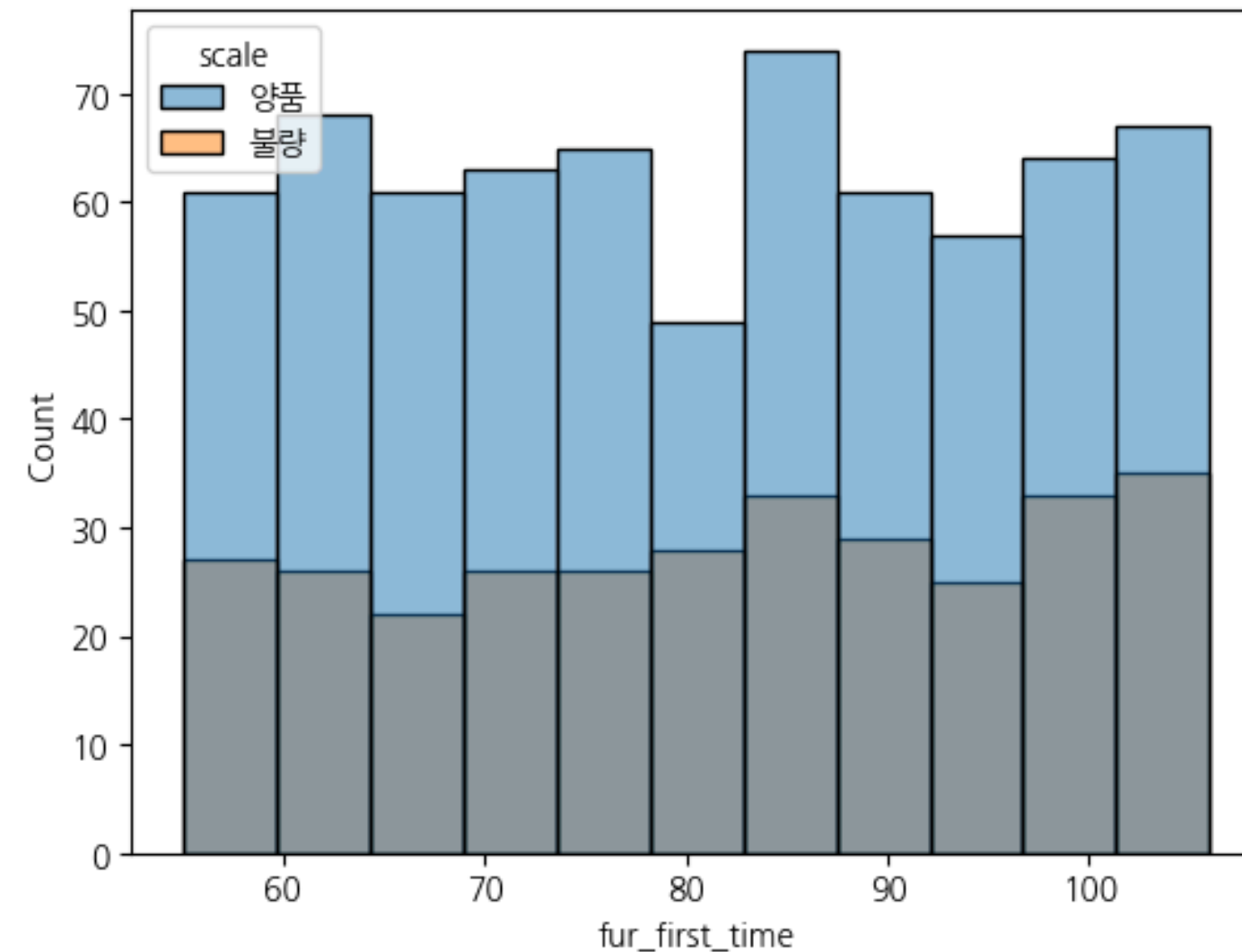
1.도메인 정보 파악

- fur_total_time

```
1 df['fur_first_time'] = df['fur_total_time'] - df['fur_soak_time'] - df['fur_heat_time']

1 sns.histplot(x='fur_first_time',hue = 'scale',data=df)

<Axes: xlabel='fur_first_time', ylabel='Count'>
```

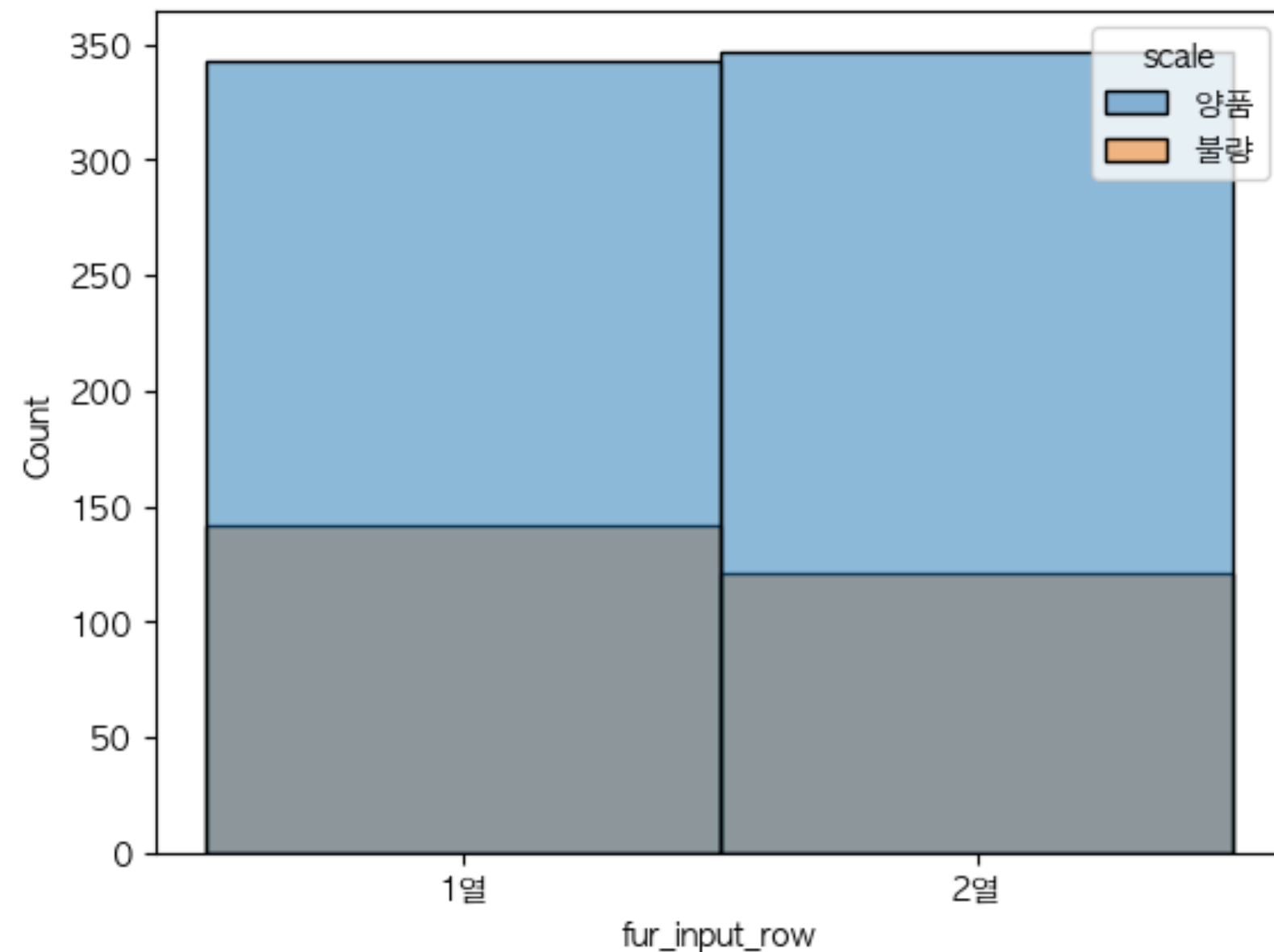


총 재로시간은 예열대, 가열대, 균열대의 시간을 합친 시간
현재 변수에 예열대의 시간만 존재하지 않음

→ 재로시간에서 가열대, 균열대의 시간을 제거한 뒤 예열대의 시간 확인 결과 의미가 없다고 판단
→ total_time 변수 삭제

1.도메인 정보 파악

- fur_input_row



```

1 sns.histplot(x='fur_input_row',hue='scale',data = df)
2
3 scale_input = pd.crosstab(df['scale'],df['fur_input_row'])
4
5 stats , p , dof , expected = chi2_contingency(scale_input)
6
7 if p > 0.05:
8     print('서로 독립임으로 scale에 영향을 주는 요인이 아니다.')
9 else:
10     print('서로 종속임으로 scale에 영향을 주는 요인이다.')
11
12 # 가열로 들어가는 열수가 scale에 영향을 주지 못함.

```

✓ 0.1s

서로 독립임으로 scale에 영향을 주는 요인이 아니다.

scale 변수와 가열로 장입열은 서로 독립이므로 scale에 영향을 주지 못하는 변수이기에 삭제

1.도메인 정보 파악

- fur_no

```
1 scale_fur_no = pd.crosstab(df['scale'],df['fur_no'])
2 scale_fur_no
3
4 stats , p , dof , expected = chi2_contingency(scale_fur_no)
5 print(stats,p)
6
7 # 고로의 호기가 불량에 영향을 주는지 보기위해 카이제곱 독립성 검정을 시행한 결과
8 # p-val 0.45로 독립이라는 귀무가설을 기각하지 못해 고로가 SCALE 불량에 영향을 준다고 보기 어렵다.
```

✓ 0.0s

1.5783415322367476 0.4542212948451264

scale과 호기수 간 관계가 있는지 확인하기 위해 카이제곱검정을 시행
→ p-value가 0.45로 귀무가설을 기각하지 못하여 서로 독립
→ 따라서 호기수가 scale 불량에 영향을 주지 않는다고 판단하여 삭제

1.도메인 정보 파악

최종 사용 변수

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 953 entries, 0 to 999
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   scale                  953 non-null   object
1   steel_kind             953 non-null   object
2   hsb                    953 non-null   object
3   fur_heat_temp          953 non-null   int64
4   fur_heat_time          953 non-null   int64
5   fur_soak_temp          953 non-null   int64
6   fur_soak_time          953 non-null   int64
7   rolling_method         953 non-null   object
8   rolling_temp           953 non-null   int64
9   descailing_count       953 non-null   int64
10  work_group             953 non-null   object
dtypes: int64(6), object(5)
memory usage: 89.3+ KB
```

목표변수

: scale : 산화철 불량 유무

설명변수

1. steel_kind : 강종 (C:탄소 , T:티타늄)
2. hsb : 스케일 처리 적용 유무
3. fur_heat_temp : 가열대 온도
4. fur_heat_time : 가열 시간
5. fur_soak_temp : 균열대 온도
6. fur_soak_time : 균열대 시간
7. rolling_method : 압연 방식
8. rolling_temp : 압연 온도
9. descailing_count : 압연 descailing 횟수
10. work_group : 작업조

2. 데이터 현황 확인 후 전처리

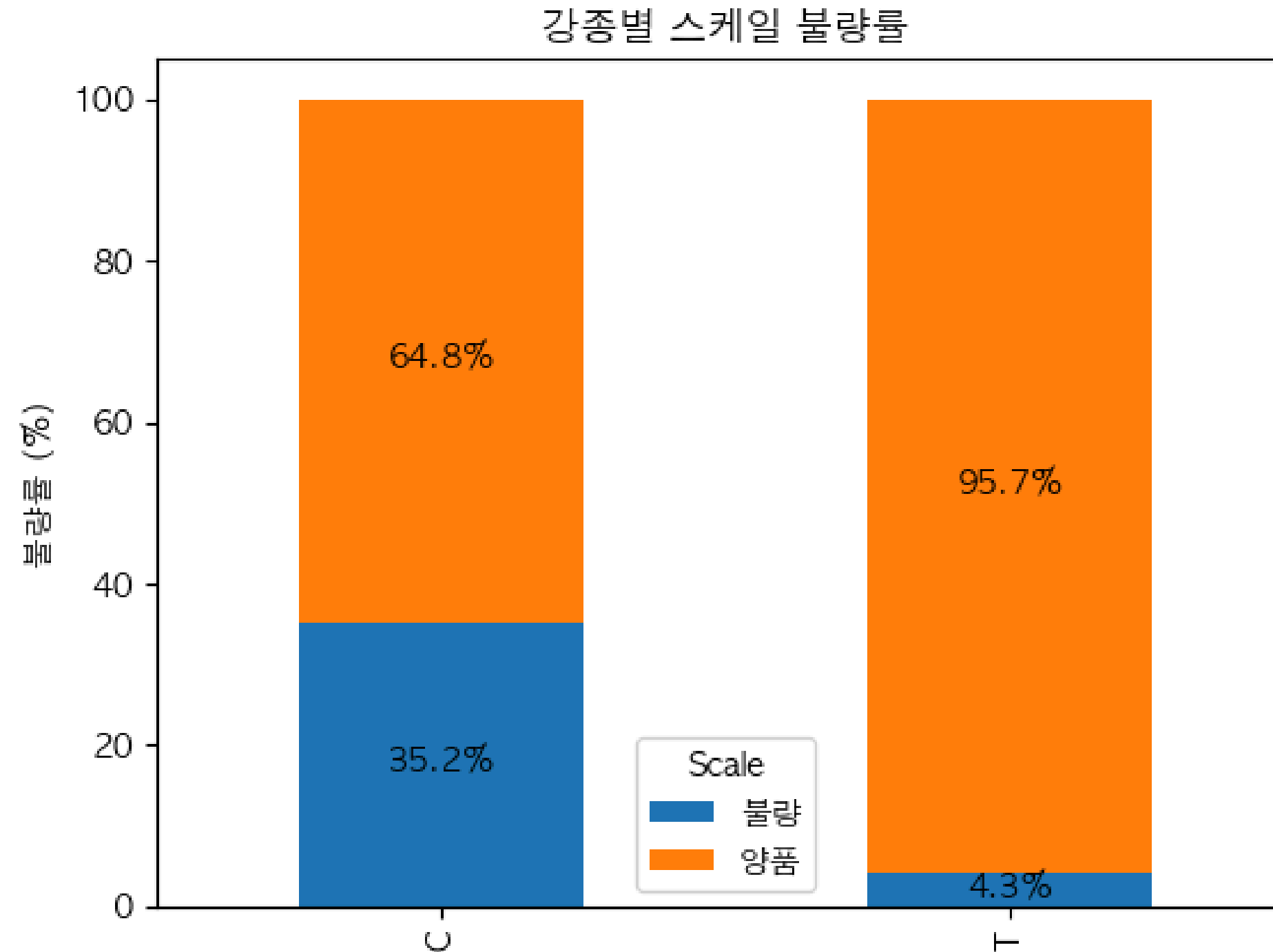
- 이상치 처리

rolling_method	rolling_temp	descaling_count	work_group	scale	rolling_temp
TMCP(온도제어)	0	6	1조	양품	0
CR(제어압연)	0	10	2조	양품	0
CR(제어압연)	0	10	2조	양품	0
CR(제어압연)	0	6	3조	양품	0
CR(제어압연)	0	6	3조	양품	0
TMCP(온도제어)	0	6	3조	양품	0

이상치가 양품의 데이터라
불량 판정에 큰 영향을 미치지 않는다고 생각하여 삭제

3. 데이터 탐색적 분석

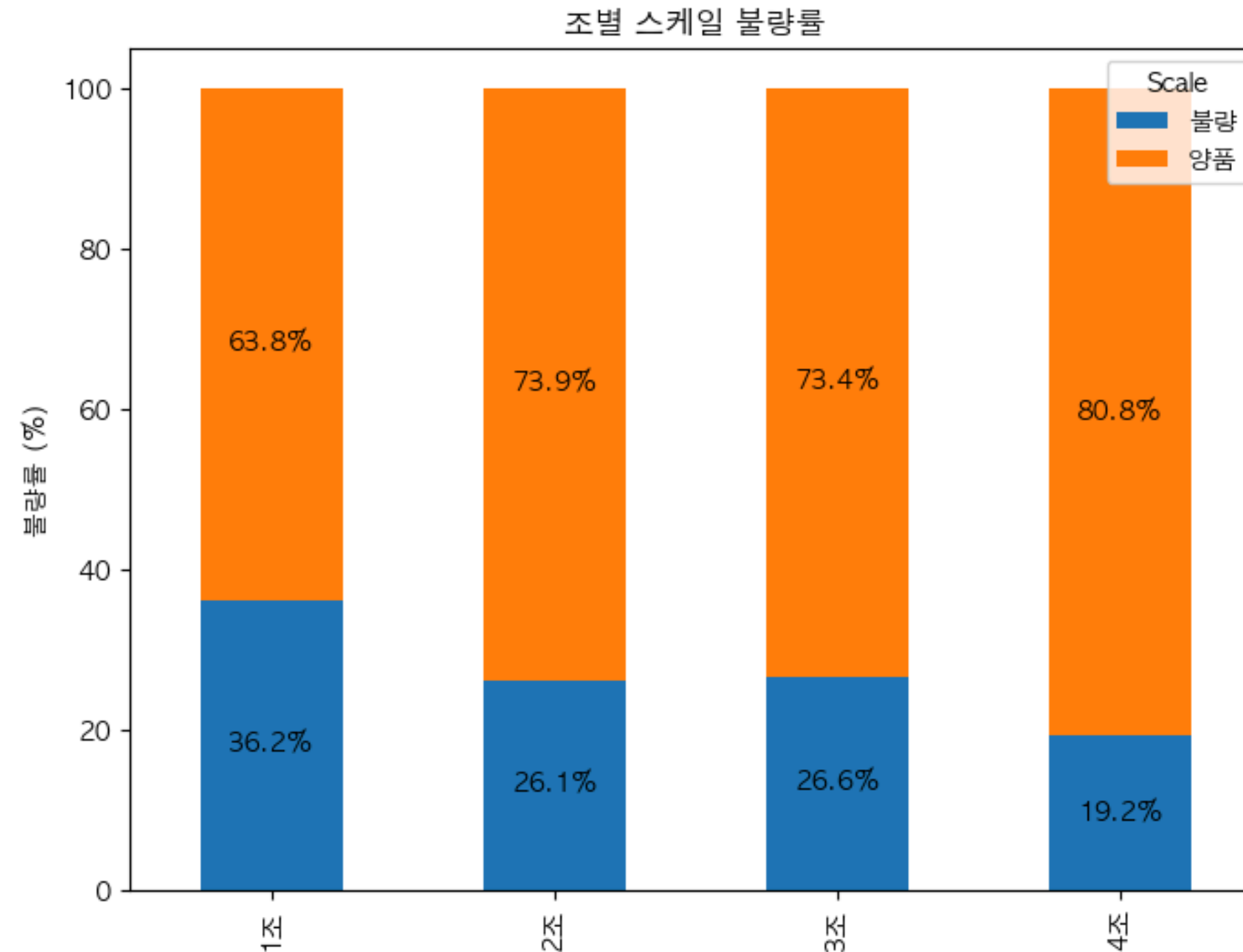
- 강종별 스케일 불량률



강종이 티타늄이 아닌 탄소일 때 불량률이 더 높음

3. 데이터 탐색적 분석

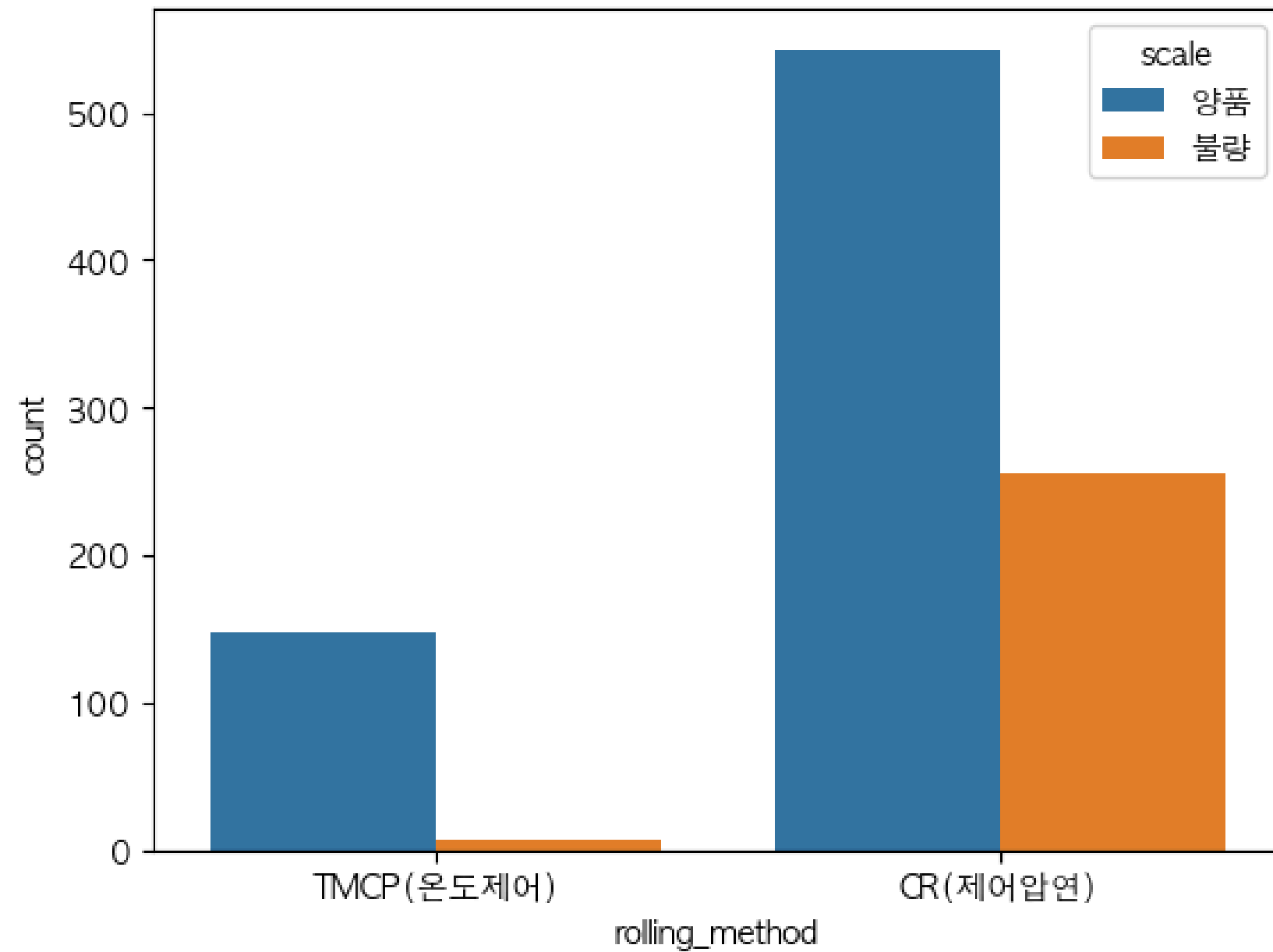
- 조별 스케일 불량률



1조와 4조의 불량률의 차이가 많이 나므로 (15%p) 개선교육이 필요

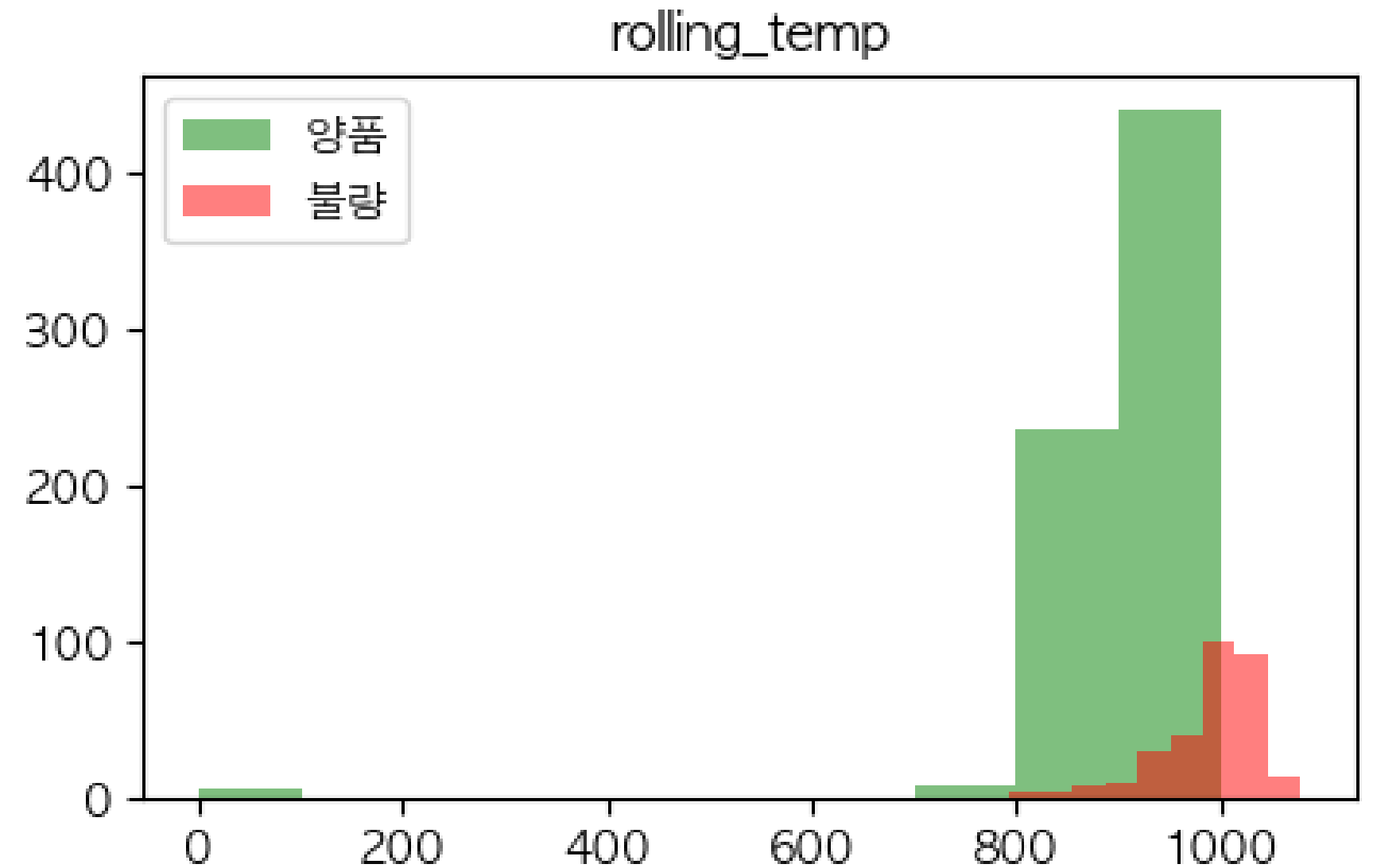
3. 데이터 탐색적 분석

- 압연 방식별 스케일 불량률



연마 방식에 따라 불량률에 차이가 존재

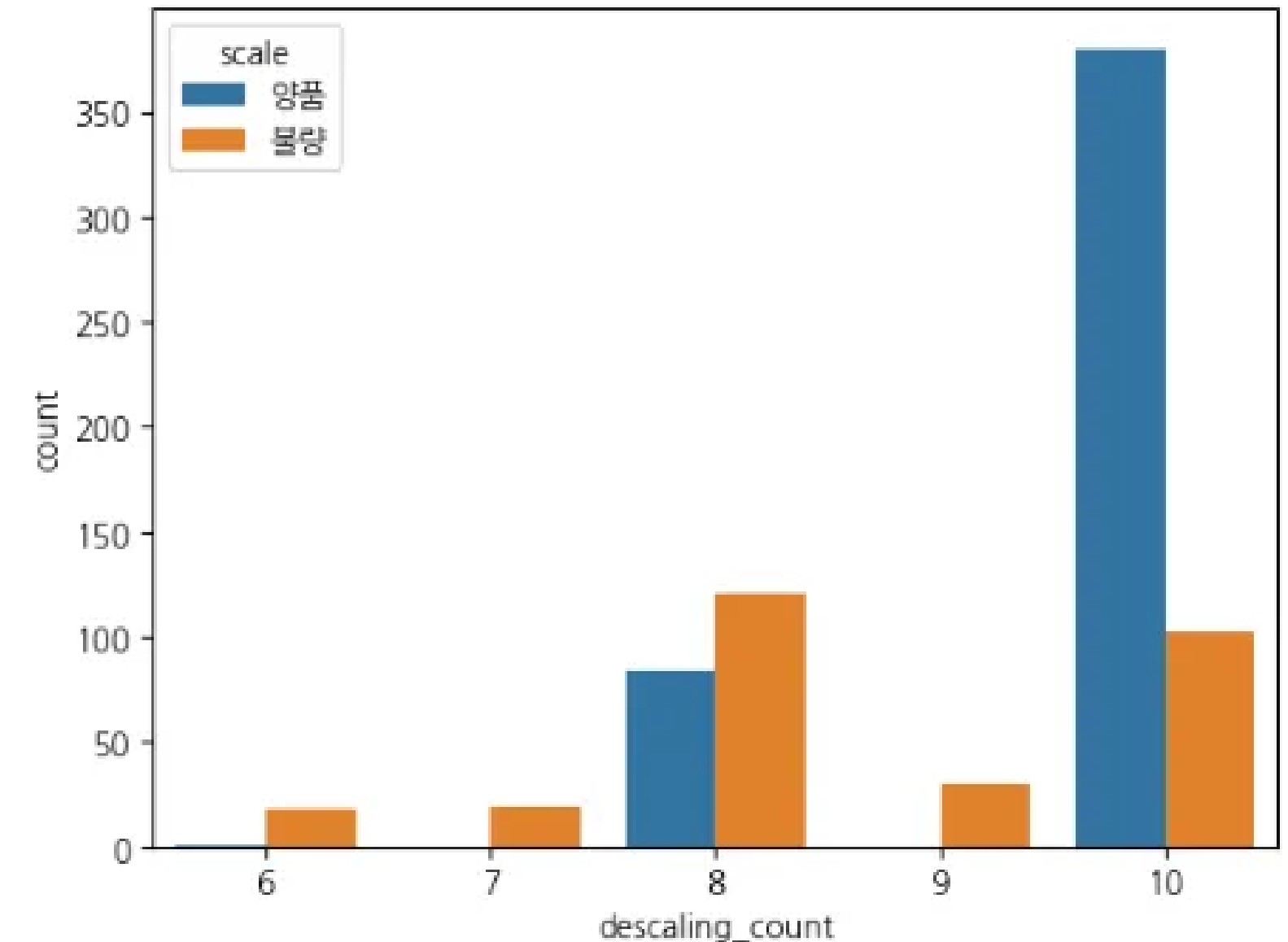
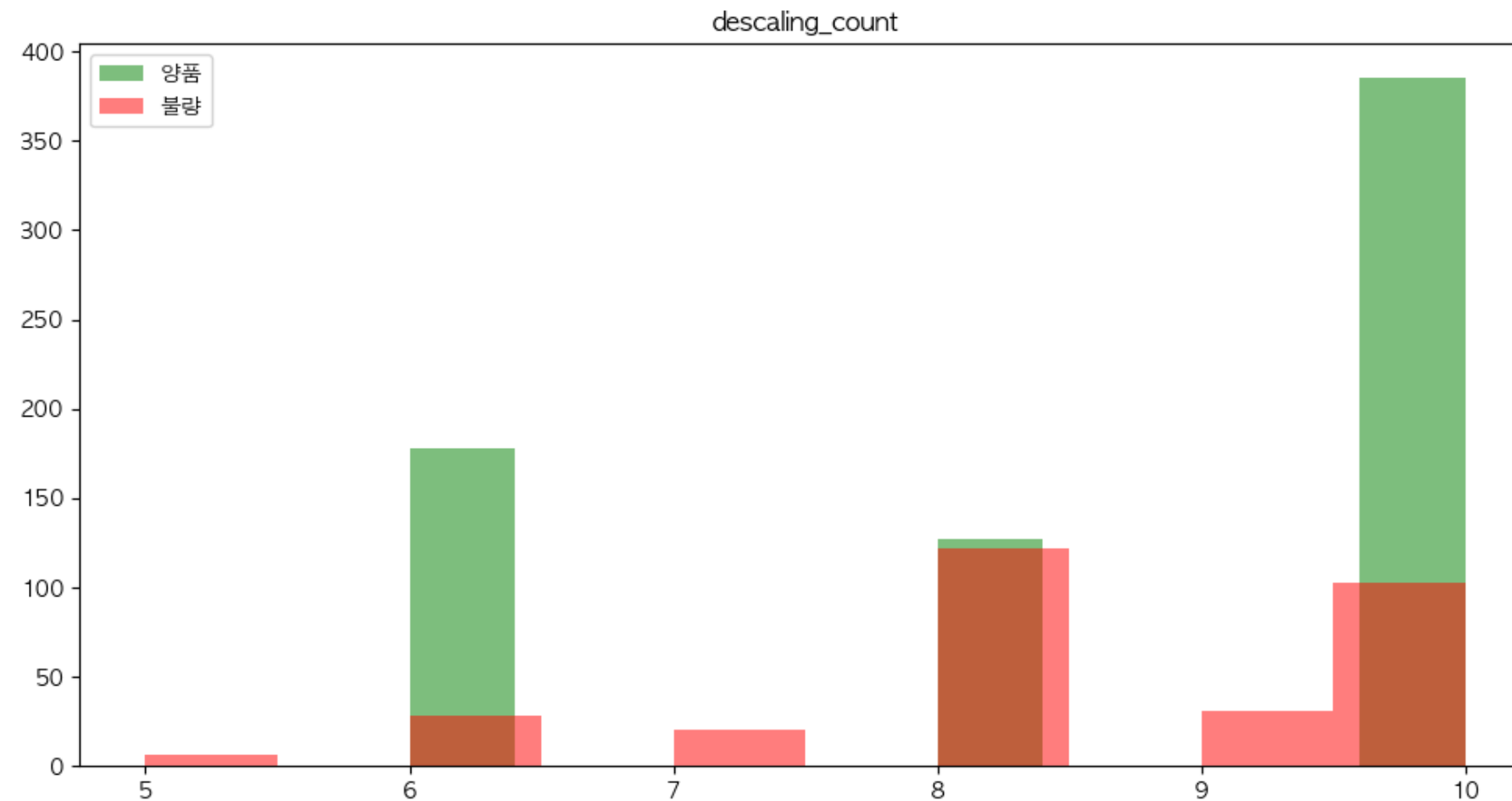
- 압연 온도별 스케일 불량률



1000도 이상일 때 100% 불량 발생

3. 데이터 탐색적 분석

- 압연 descaling별 스케일 불량률

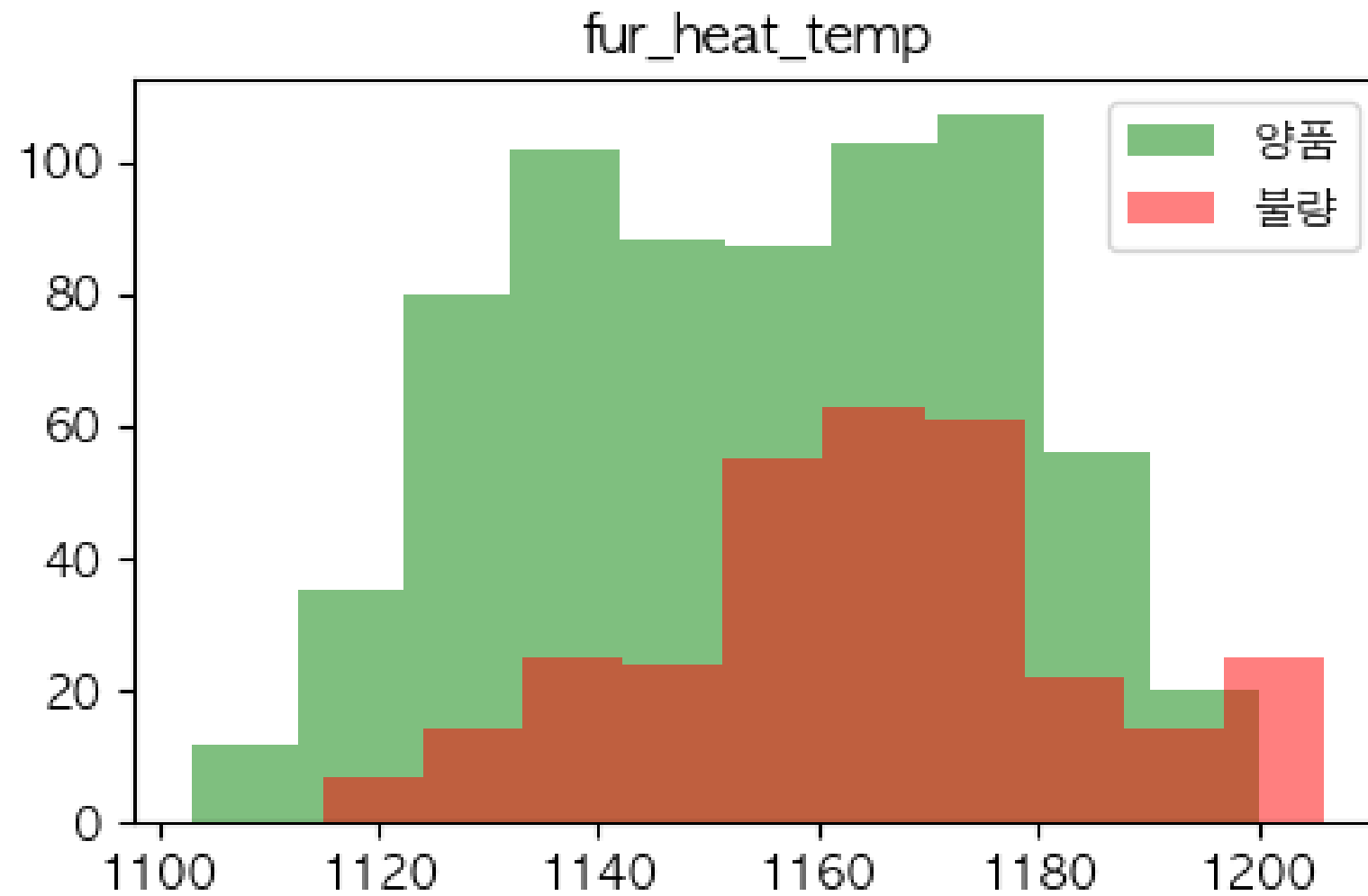


특히, 홀수 횟수는 100% 불량률을 기록하는 것을 볼 수 있음

탄소가 descaling 횟수에 따른 불량에 영향을 준다고 판단
 → 탄소의 데이터만 추출해 descaling과 scale의 카이제곱 검정 진행
 → descaling 횟수가 scale에 영향을 주는 것을 확인

3. 데이터 탐색적 분석

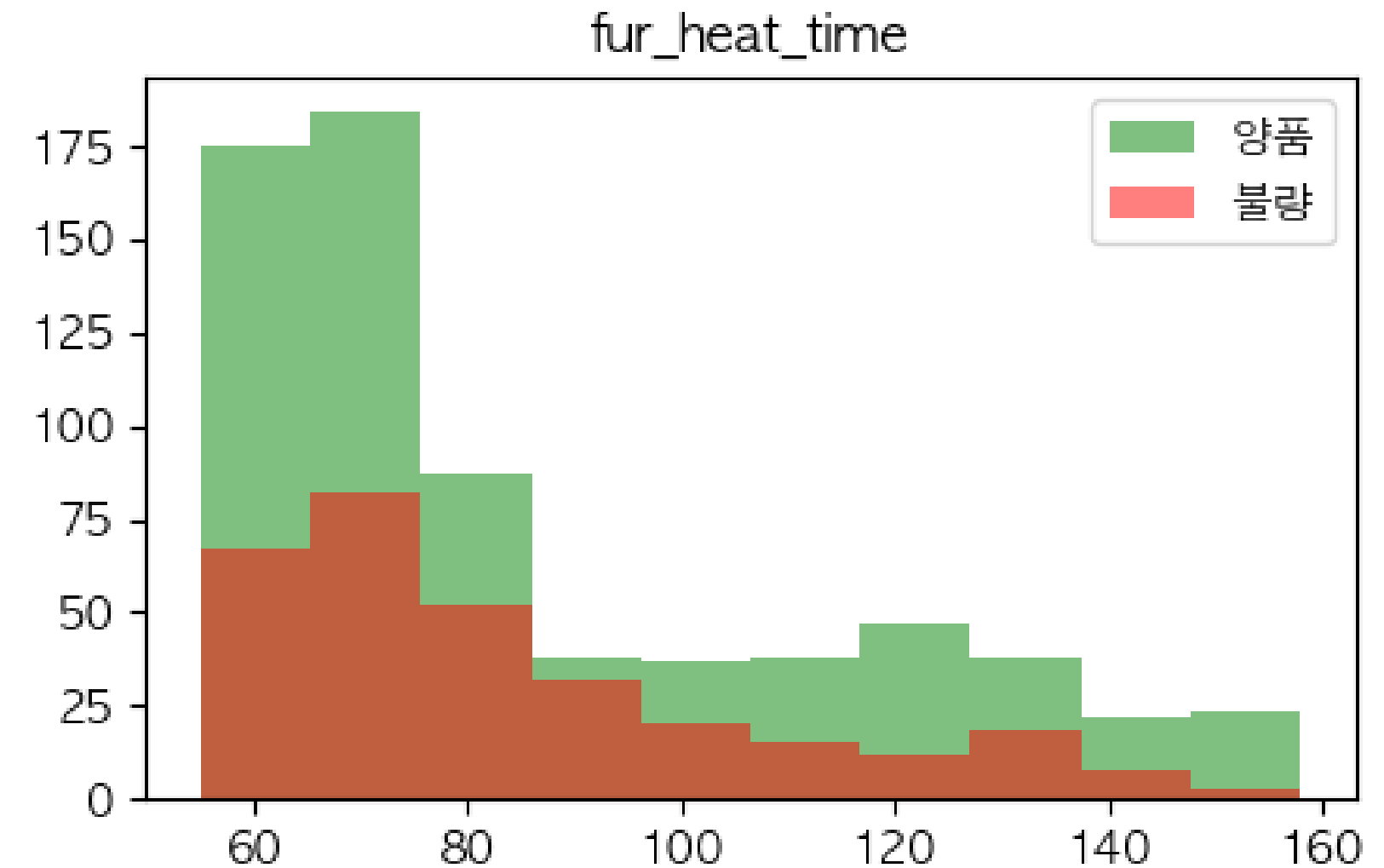
- 가열대 온도별 스케일 불량률



가열대 온도가 1200도를 넘어서면
100% 불량 발생

불량은 1120~1200도까지 나타나지만
1150도에서 1180도까지 가장 많이 분포

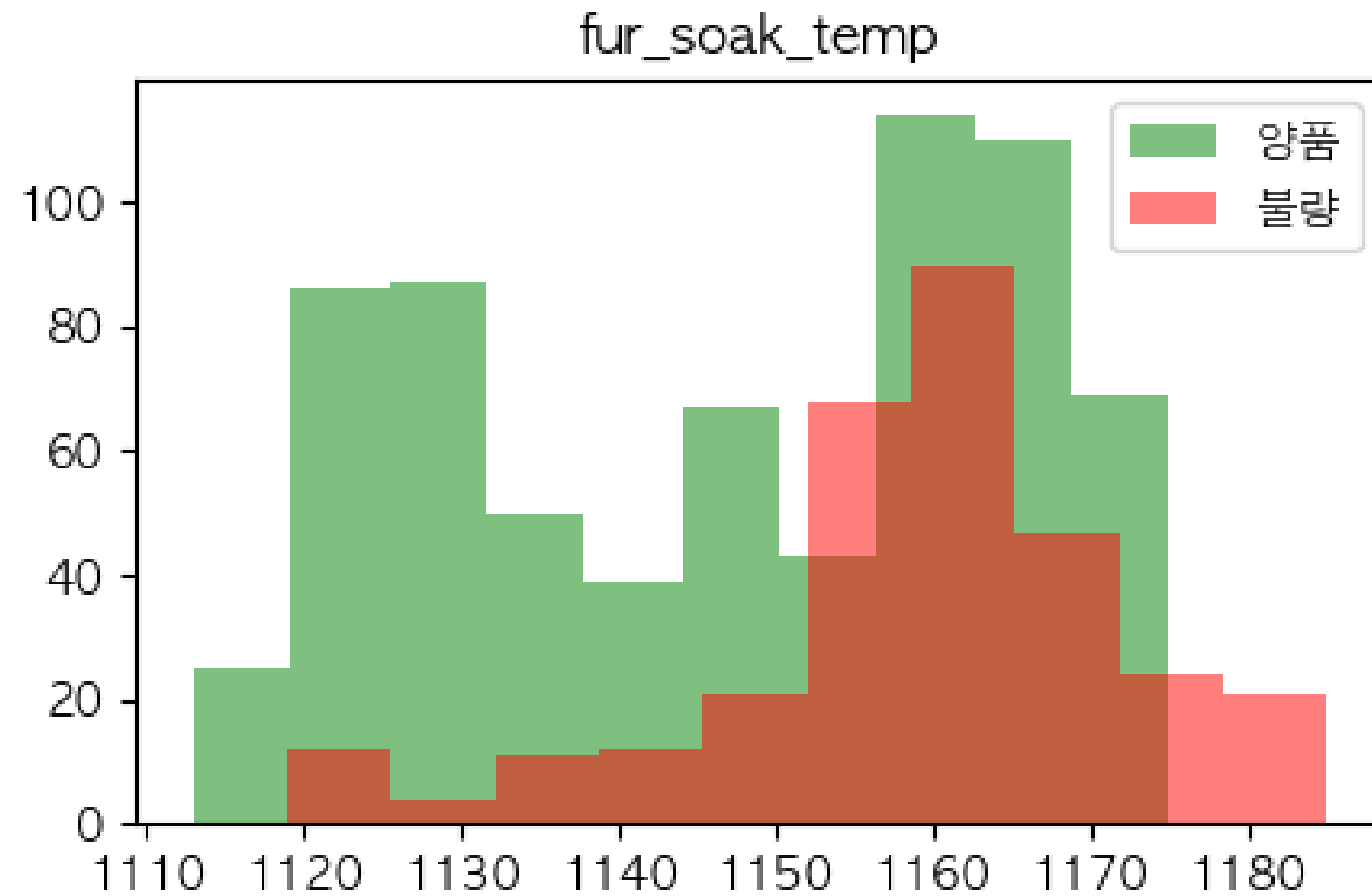
- 가열대 시간별 스케일 불량률



가열 시간이 증가함에 따라
불량횟수가 낮아지는 경향을 보임

3. 데이터 탐색적 분석

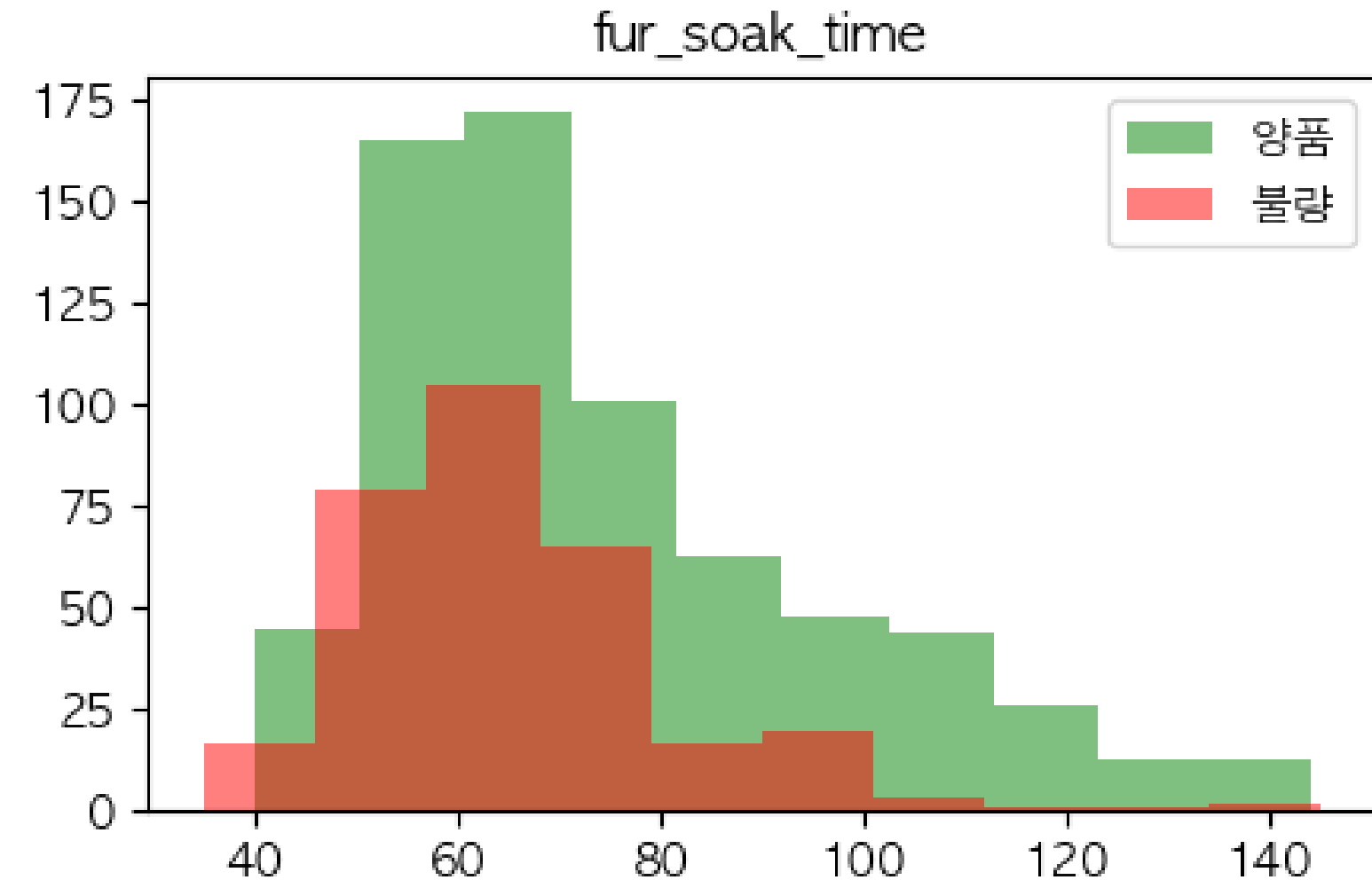
- 균일대 온도별 스케일 불량률



균일대 온도가 1175도를 넘으면
100% 불량 발생

균일대 온도가 1150도 이상이면
불량률이 높아짐

- 균일대 시간별 스케일 불량률



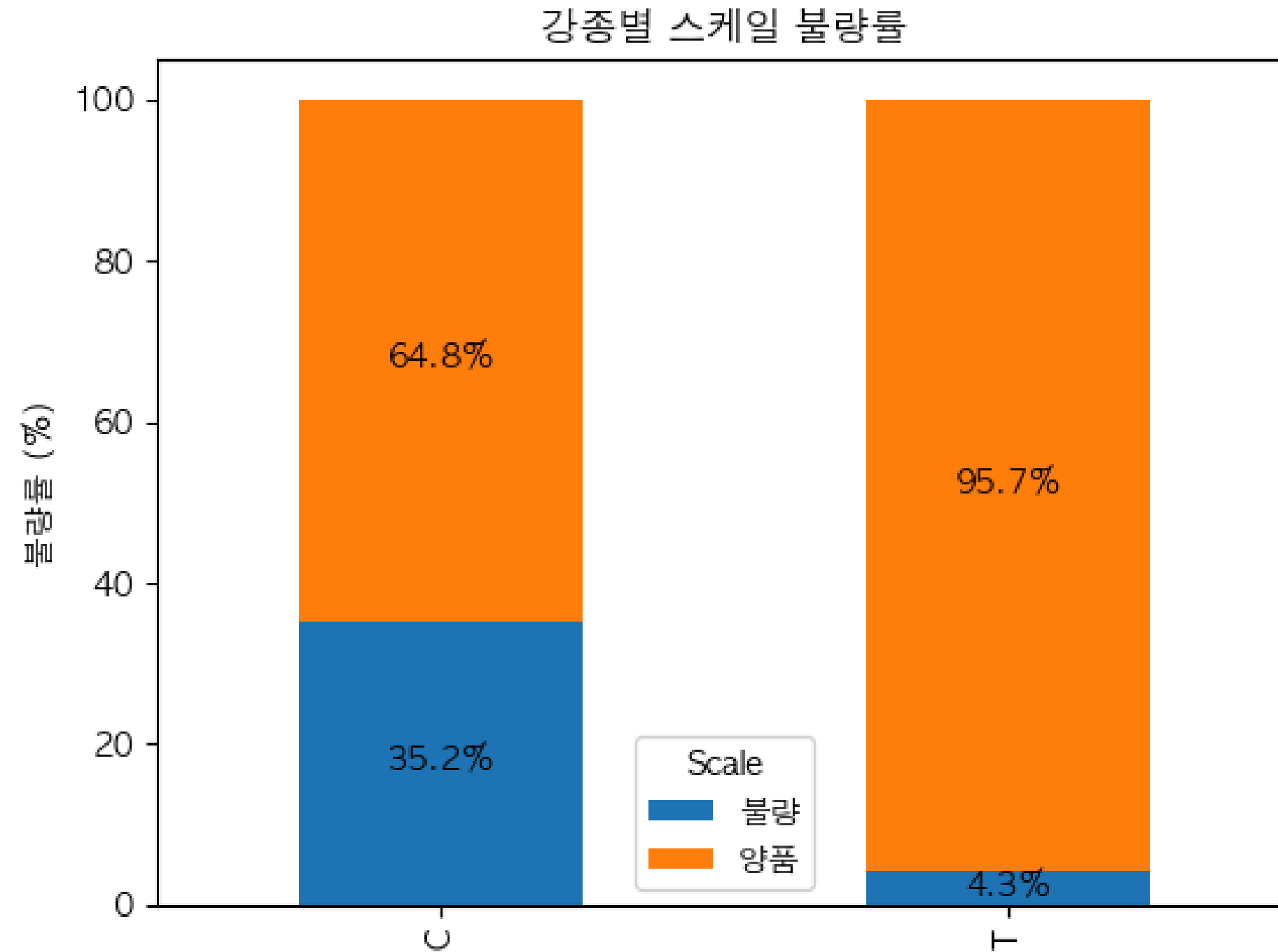
40분 이하는 불량 100% 발생

100분 이상부터 불량률이
거의 발생하지 않음

4. 분석

전체 데이터와 강종별로 나눠서 분석을 진행

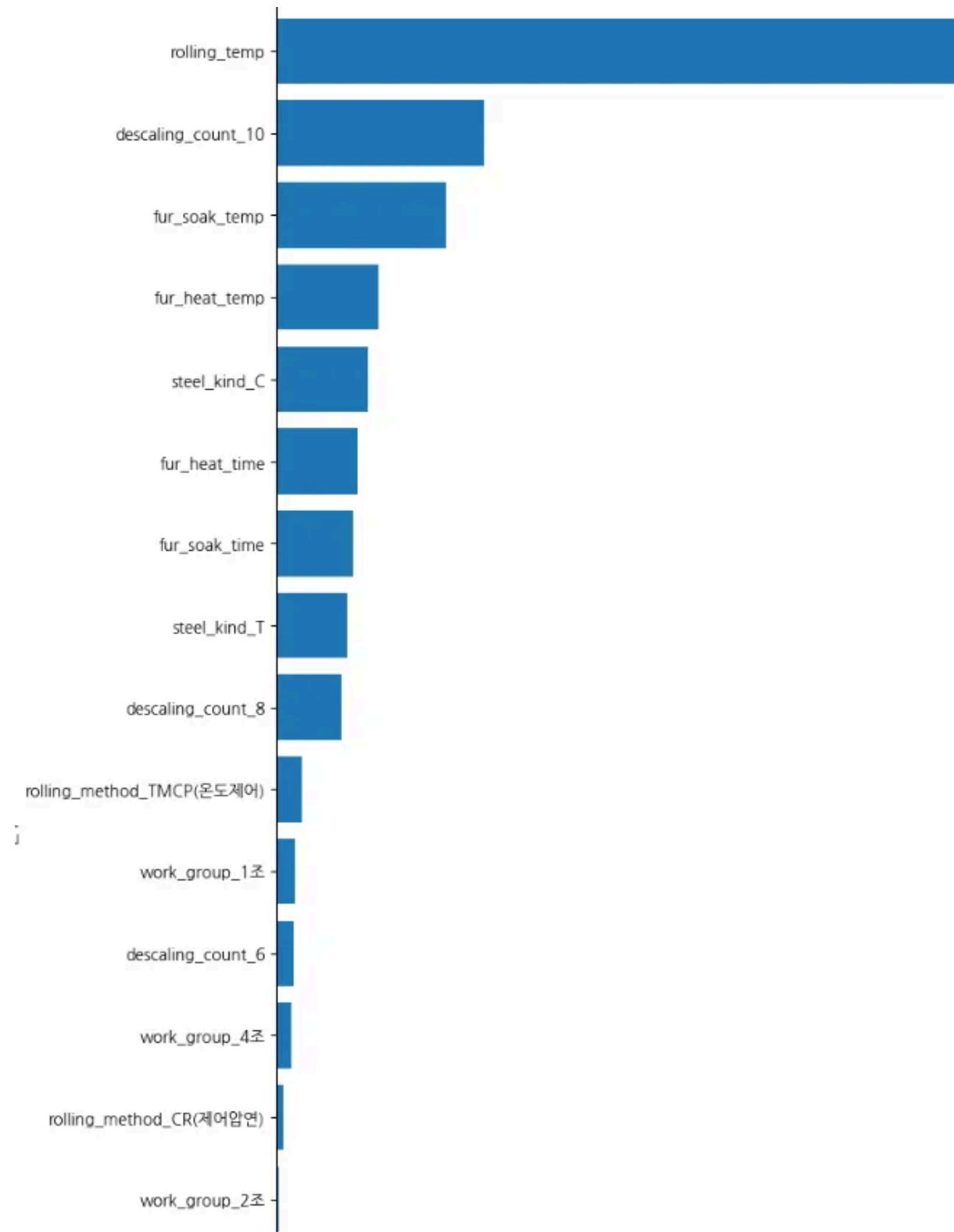
: 강종별로 불량률이 상이하므로 둘의 불량 원인 차이가 존재할 것이라 판단



4-1. 핵심인자 도출

: RF 모델을 활용하여 핵심인자를 구함

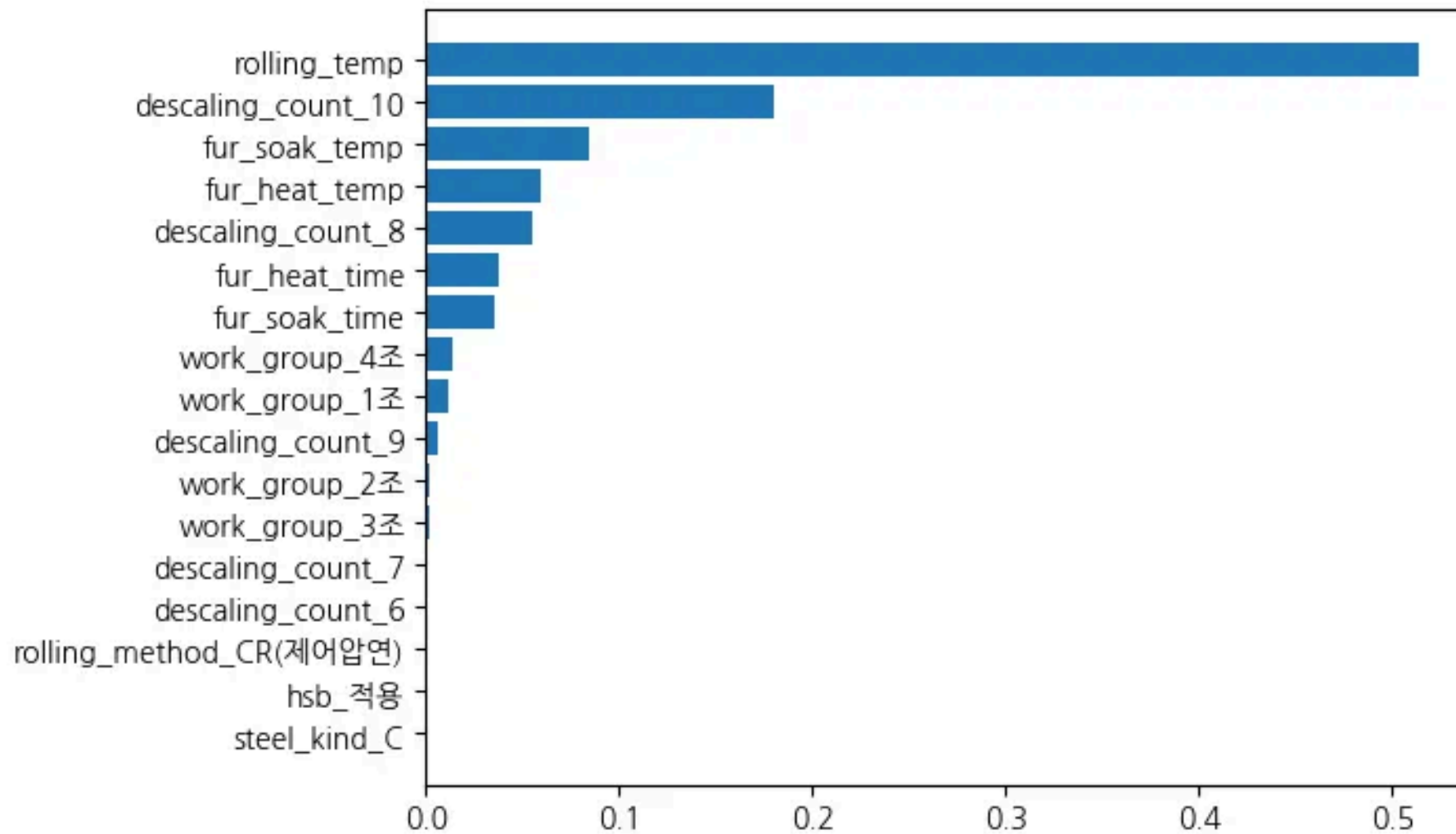
- 전체 데이터



- rolling_temp : 압연온도
- descaling_count : 압연횟수
- fur_soak_temp : 균열대 온도
- fur_heat_temp : 가열대 온도
- steel_kind_C : 강종이 탄소

4-1. 핵심인자 도출

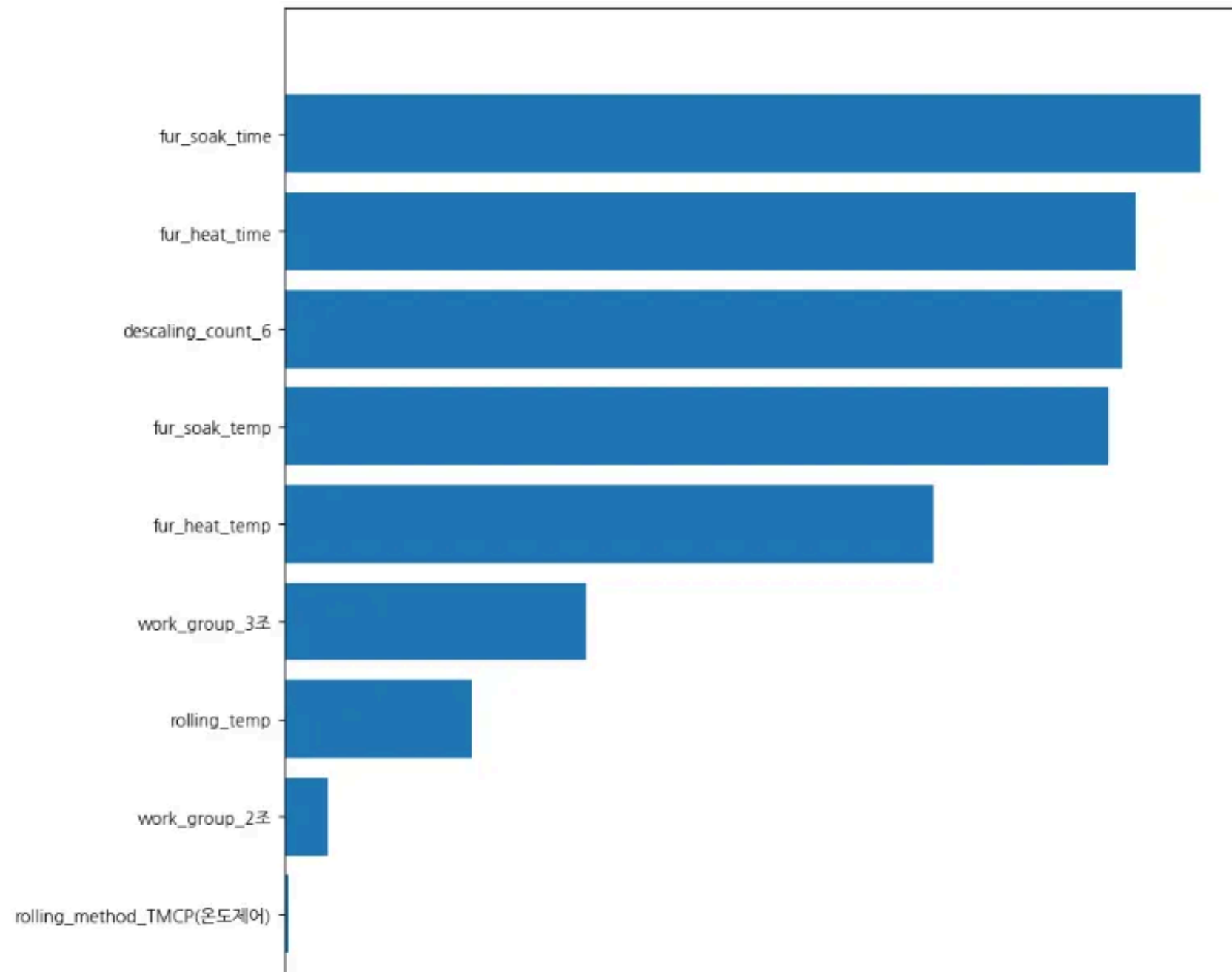
- 강종 = 탄소



- rolling_temp : 압연온도
- descaling_count_10 : 압연횟수가 10회
- fur_soak_temp : 균열대 온도
- fur_heat_temp : 가열대 온도
- descaling_count_8 : 압연횟수가 8회

4-1. 핵심인자 도출

- 강종 = 티타늄

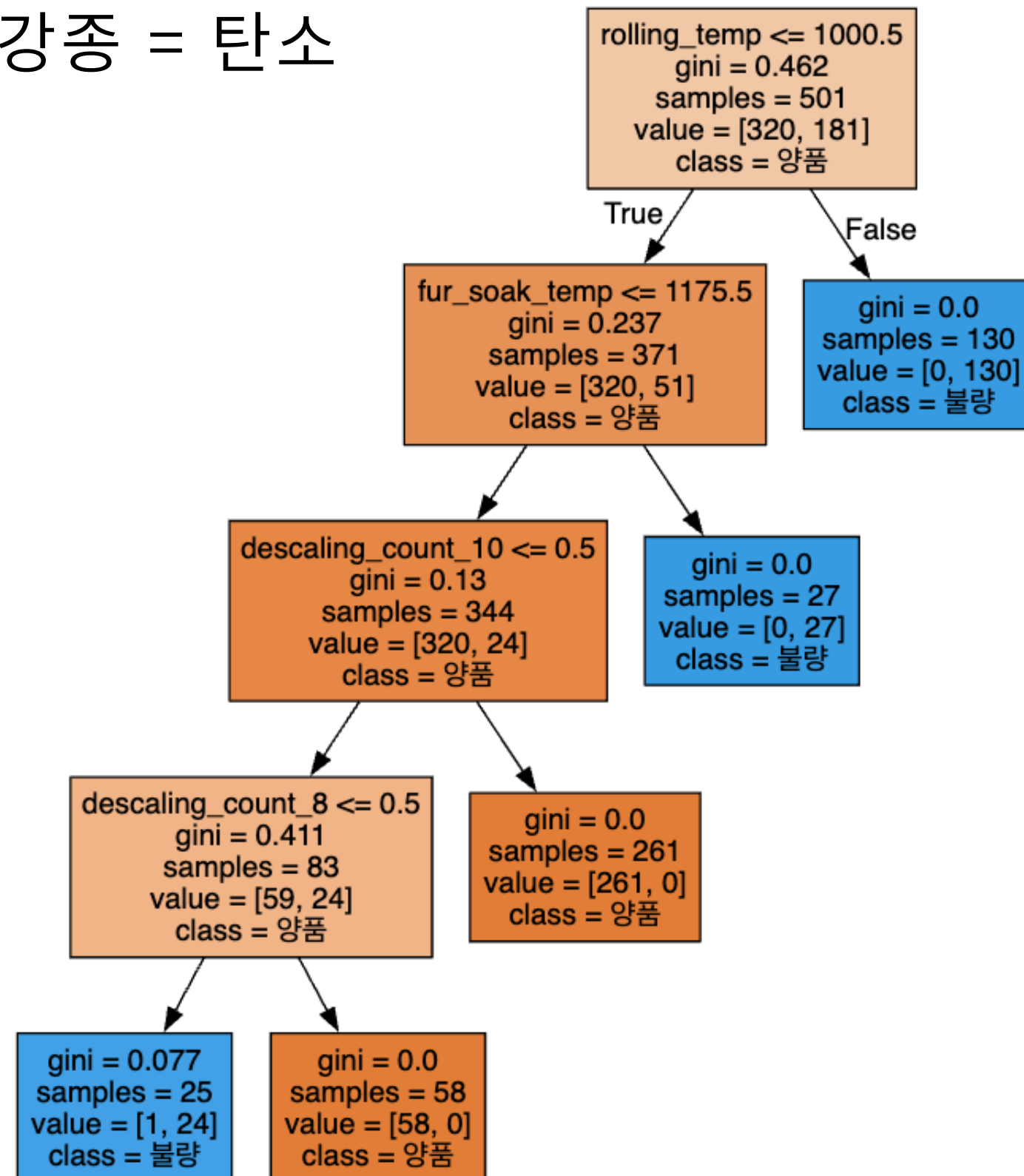


- fur_soak_time : 균열대 시간
- fur_heat_time : 가열대 시간
- descaling_count_6 : 압연횟수가 6회
- fur_soak_temp : 균열대 온도
- fur_heat_temp : 가열대 온도

4-2. 모델링

: DT 모델을 이용하여 분리기준을 시각화

- 강종 = 탄소



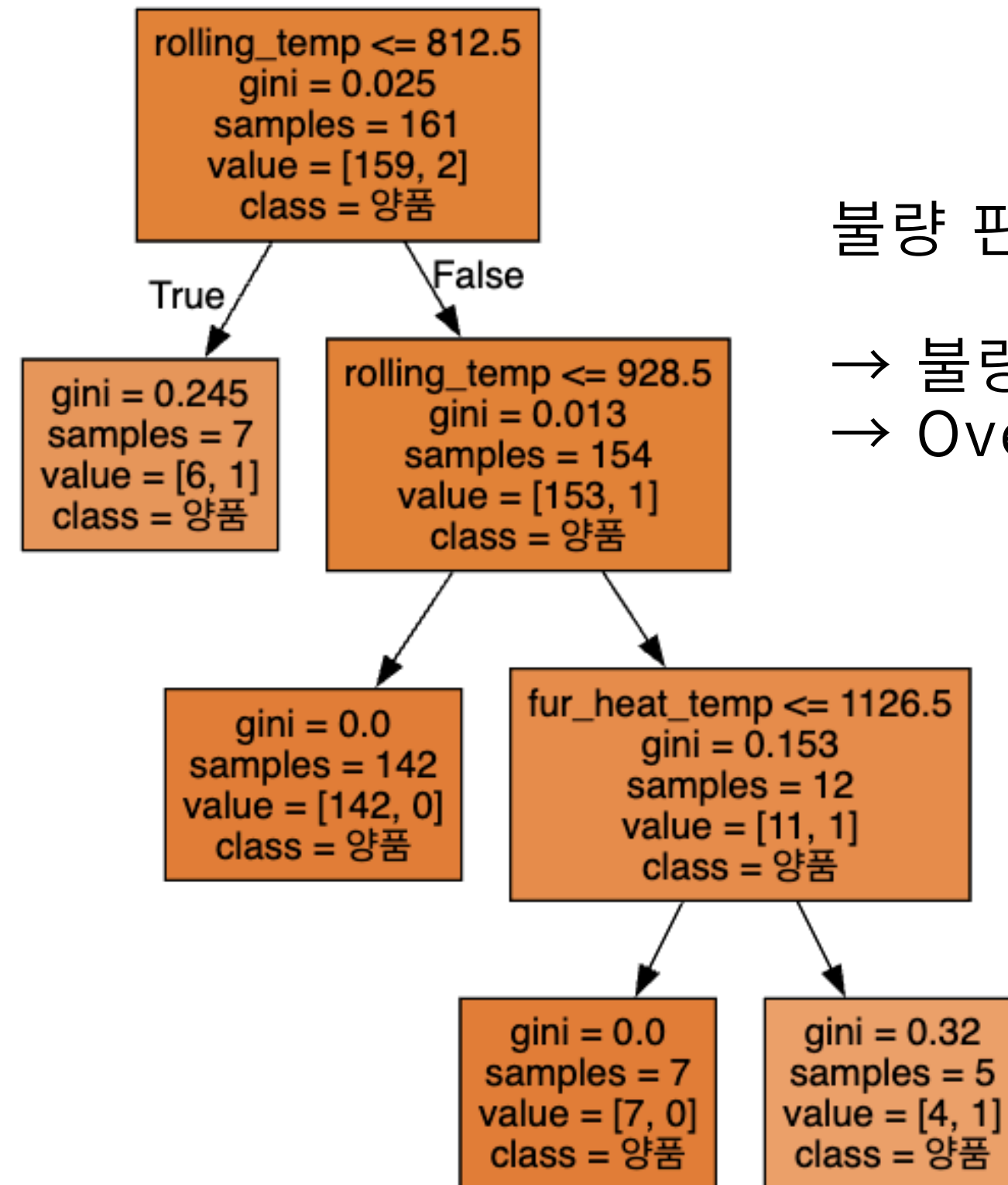
불량으로 판단되는 기준

- 압연온도가 1천도 이상일 때
- 균열대 온도가 1175도 이상일 때
- 압연횟수가 8회가 아닐 때

→ 탄소강의 후판공정을 할 때
압연온도와 균열대 온도의
세심한 조절이 필요할 것으로 판단

4-2. 모델링

- 강종 = 티타늄



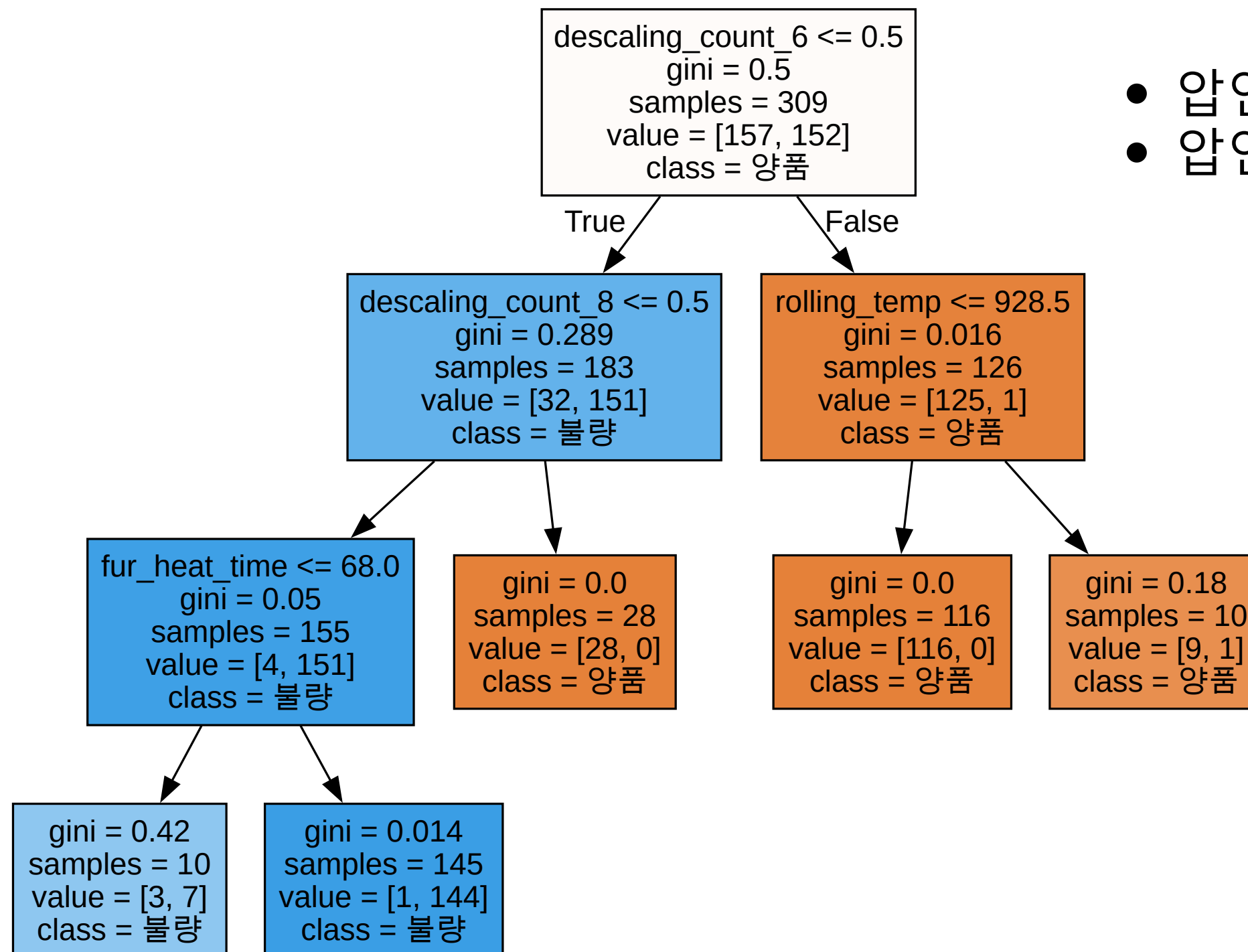
불량 판별이 되지 않음

→ 불량률의 빈도가 양품의 빈도의 10배에 가깝기 때문으로 판단
→ Oversampling 기법인 SMOTE 방법을 사용하여 모델 개선 시도

4-2. 모델링

- 강종 = 티타늄

: SMOTE 방식으로 Oversampling한 데이터

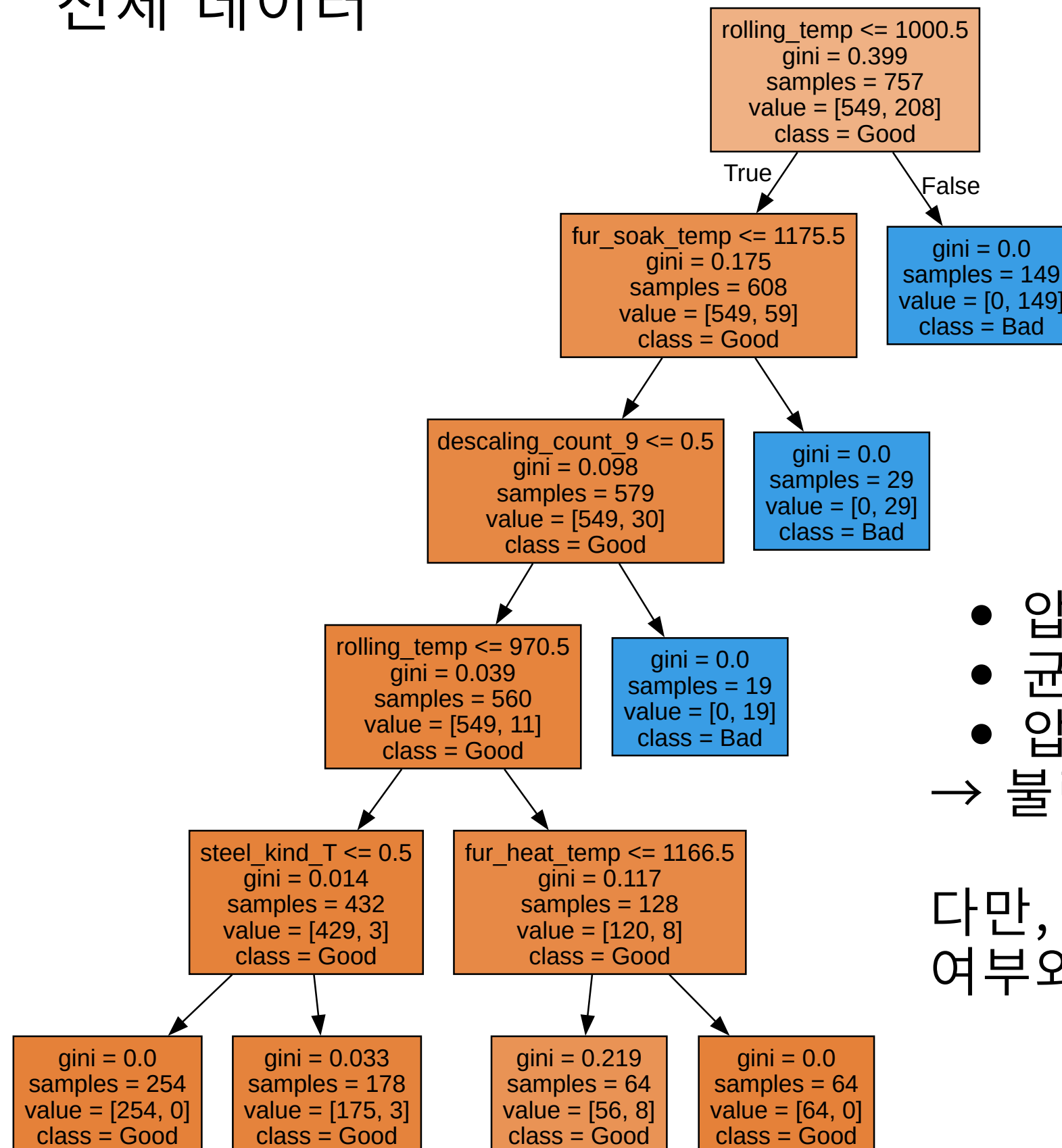


- 압연횟수가 6회이면 양품
- 압연횟수가 6회 또는 8회가 아니면 모두 불량 판정

→ 압연횟수를 6회나 8회로 고정시키는 개선 방안이 필요할 것으로 판단

4-2. 모델링

- 전체 데이터



불량으로 판단되는 기준

- 압연온도가 1000도 이상일 때
- 균열대온도가 1176도 이상일 때
- 압연횟수가 9회일 때

- 압연온도 1000도 미만
 - 균열대 온도 1175도 이하
 - 압연횟수 9회 제외
- 불량인 개수가 208개에서 11개로 확연히 감소

다만, 현장에서 압연온도의 조절과 균열대 온도 조절 가능 여부와 필요 압연횟수가 9회인지 파악 필요

5. 개선방안

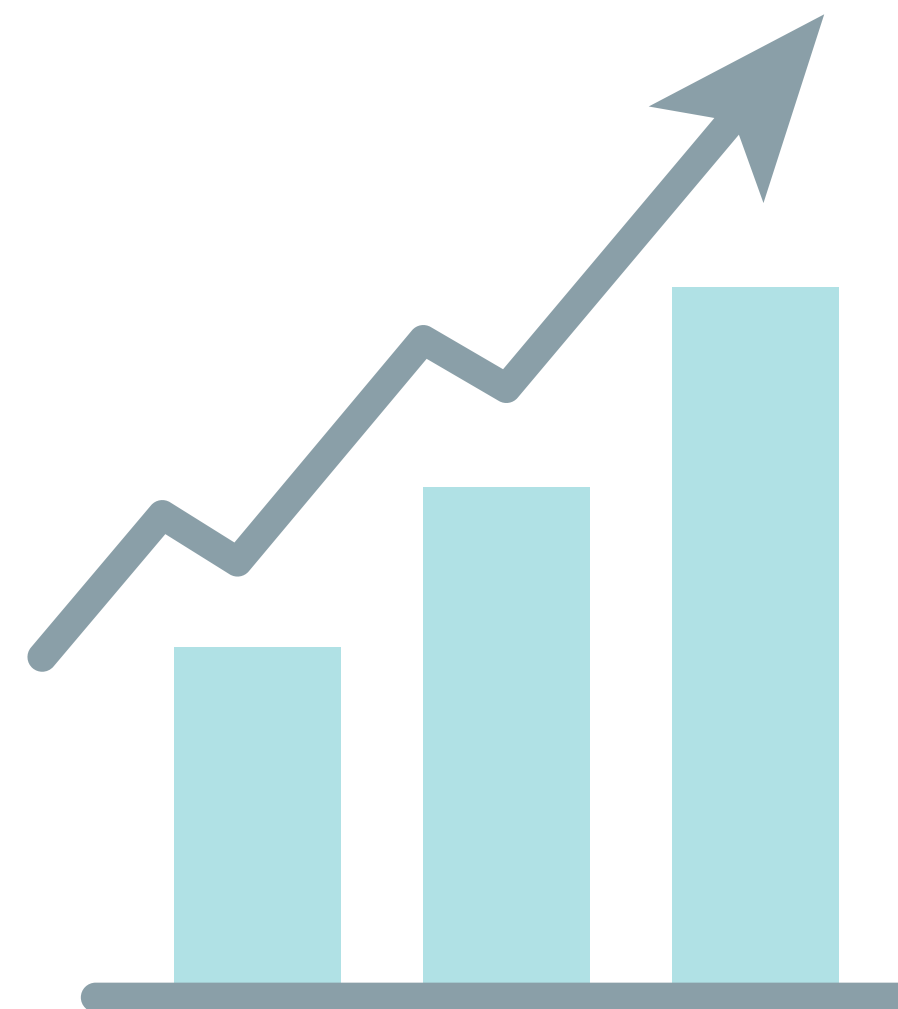
1. 압연 온도 관리 모니터링 시스템 (1000도 이하 유지)

2. 적정 descaling 실험

- 압연횟수 6회나 10회로 고정

3. 원자재 품질 관리(불순물 관리)

- 탄소강의 품질 상승
- 티타늄강의 생산-판매 강화



6. 배운 점 & 느낀 점

1. 전체 데이터만 분석을 하는 것이 아닌 범주형 데이터를 기준으로 층별화를 하여 보는 것도 분석을 하면 새롭게 보이는 정보가 있다는 점을 배웠습니다.
2. 공정에서 생성되는 다양한 데이터를 기반으로 문제의 원인을 분석하고 문제의 개선점을 도출하는 과정까지 신뢰할 수 있는 데이터의 중요성을 깨달았습니다. 또한, 이를 기반으로 한 논리적인 사고와 분석이 객관적인 인과관계를 파악하는 것에 필수적이며, 효율적인 문제 해결과 공정 최적화의 핵심임을 깨달았습니다.
3. 데이터의 전처리 과정과 변수 선택 과정에서 도메인 지식이 중요함을 배웠습니다.
4. 시각화와 통계적 검정, 일련의 데이터 분석 프로세스를 이해하였고 더불어 실제 분석결과를 해석하는 것이 어려움을 알게 되었습니다. 또한 이를 현장에 적용하는것에 많은 제약사항이 있다는 점을 깨달았습니다.