# **결측치, 이상치 정제**

각 파일을 이용해서 작업 실시

## Oxidation

- 이상치 조정: (Oxid\_time)

- 같으면 해당 값으로 대체, 다르면 평균값  
 - 같은 날짜의 한 챔버 내에서 dry/wet 둘 중 하나로

## photo\_litho

- 삭제(1107, 1161)

## Ion\_Iplantation.csv

- 삭제(1107, 1161)

- mean 값 대체(1239, 1323, 1229)

-> 대호와 정용

## etching

- 삭제(1107, 1161)  
 - mean 값 대체(580, 462, 276, 483, 575)

- 결측치: Thin F1,2,3, 4 -> 음수 값 조정 실시 필요

-> 나희와 윤슬

# **파생변수 생성**

## Path(효림)

Merge\_Jihyun.csv 이용

* 6개 데이터 모두 합쳐 놓은 파일임. 이 파일 이용해서 path 칼럼 추가
* 1107, 1161 행은 삭제한 상태에서 작업해줄 것

## Etching Rate(나희 윤슬)

* Etching.csv 이용
* 1107, 1161 행은 삭제한 상태에서 작업해줄 것

## Edge Error(보류)

* QualityInspection.csv 이용
* 1107, 1161 행은 삭제한 상태에서 작업해줄 것

## Defect(효림)

* QualityInsepction.csv 이용
* Target 수가 200 이상이면 1, 아래면 0으로 분류하여 새로운 칼럼 생성
* 1107, 1161 행은 삭제한 상태에서 작업해줄 것

모든 정제 과정이 끝난 뒤 하나의 파일로 합쳐서 아래의 분석 실시

3:30까지 하나의 취합 파일 생성할 것



# 공정 과부하 및 이상 발생현황 확인

1. C 관리도 그리기
2. path 별로 웨이퍼가 UCL, LCL 넘어갔는지를 확인
3. path 별 빈도수를 체크하고, 가장 많이 운영된 path 5개, 가장 적게 운영된 path 5개  
   -> 이게 과부하/안정 여부에 차이가 있는지? (관리도를 통해서 비교)
4. 해당 공정들을 분류하여 카이스퀘어로 동일성 검정:   
   불량이 동일한 비율로 발생하는가?
5. Insight: 불량률에 차이가 있다면 공정이 과부하 걸릴 수록 불량 유발. 따라서 공정 안정화 방안을 마련할 필요가 있음

## 이상공정(1), 안정공정(0) 분류 후 이상(불량)을 초래하는 운전 조건(파라미터) 확인

공정 별 파라미터에 대해서 0, 1 집단 간 평균 차이가 있는지 확인

모든 파라미터에 대해서 평균 차이 확인할 것임

boxplot과 2 sample t test 이용

## Chamber path 별 부하 현황 확인

막대 그래프로 path 별 부하 현황 차이

카이스퀘어 검정:

불안정한 공정 5개, 안정 공정

## 각 변수 별 특성 파악

공정 별 목표변수를 두고 영향인자 파악

## Modeling

1차 모델링으로 주요 변수 추출

## Web

..

