

# 과제 2

## ■ 다음 조건에 맞는 프로그램을 작성하여 구동하시오.

- 어느 공장에서 제품을 생산하는 과정에서 3개의 변수를 정하여 측정하고, 만들어진 제품의 품질을 측정하여 우측의 데이터로 만들었고, 3개의 변수(재질, 온도, 색깔) 중 어떤 것이 가장 품질을 높이는데 영향을 주는지 알아내고자 한다.

*(matt : 재질, temp : 온도, color : 색깔, quality : 품질)*

- 우측의 데이터를 Excel 파일에 저장하고 csv 파일로 변환하시오.
- 다음의 내용을 포함하는 파이썬 프로그램을 만드시오
  - 1) matt의 값을 x축으로 하고 각 값 별로 quality=0과 quality=1의 개수를 계산하여 y축으로 하는 countplot을 설정하여 차트로 그리시오.
  - 2) 1)의 내용을 matt 대신 temp에 대하여, 그리고 color에 대하여 quality 개수를 y축으로 하는 차트를 그리시오.
  - 3) 피어슨 상관계수를 적용하여 matt, temp, color, quality 간의 상관계수를 구하여 화면에 프린트하시오.
  - 4) matt, temp, color 중 quality에 가장 영향을 많이 주는 것과 그 상관계수를 화면에 프린트하시오.
  - 5) quality, matt, temp, color 간의 상관계수를 이용하여 heatmap을 그리시오.

matt	temp	color	quality
1	3	1	0
2	2	1	1
3	1	3	1
2	2	2	1
3	2	3	1
2	3	2	0
3	1	1	1
1	3	3	0
2	3	1	0
3	2	1	1
2	3	3	0
3	1	2	1
2	2	3	0
1	3	2	0
3	2	2	1

# 과제 2

## ■ 프로그램 결과

- csv 파일 1개
- count plot 3개 (화면에 출력)
- 상관계수 matrix (4X4, 화면에 출력)
- quality 에 영향을 가장 많이 주는 변수 및 값 (화면에 출력)
- heatmap (화면에 출력)

# 과제 2

## ■ 제출 방법

- 제출할 파일
  - 프로그램(.py file)
  - Input 데이터 파일 (.csv file)
  - 수행 결과가 담긴 화면을 캡처한 그림 파일
- 메일의 제목을 다음과 같이 작성  
"BigData-01-2 " + **학생이름** + " " + **학번**  
예) BigData-01-2 홍길동 22100000
- 메일 보낼 곳 : wonchanjung@hanmail.net

*출력 결과가 한 화면에 보이지 않을 경우,  
스크롤하여 2개의 그림 파일로 만들어 제출*

## ■ 제출 일자

- 12/04(일) 23:59:59
- 제출 일에 제출하지 못하는 경우  
=> 추후에 제출하면 기간에 따라 얼마의 감점 후 점수가 부여됨