

# 2주차 과제

```
# Week_2 Quiz
# Description
# 문제 바로 아래에 답안을 작성하여, <<r script 형태>>로 저장 후 제출해주시면 됩니다. (.R 파일업로드 / Rproj파일은 업로드하지 않으셔도 됩니다.)
# 답안인 코드만 작성하시면 됩니다. 코드 실행 결과 포함 X
# [R Script Encoding Error] : File > Reopen with Encoding > 'UTF-8' > OK
# 과제 제출 기한은 일요일 저녁 10시까지입니다.

library(dplyr)
# 데이터는 배포해드린 kc_house_data를 사용하세요.

# (5점) 1-(1). 2주차 과제에 배포된 kc_house_data.csv를 다운받고 불러오세요.
# 데이터 출처 : https://www.kaggle.com/harlfoxem/housesalesprediction

# (5점) 1-(2). 데이터를 검토할 수 있는 함수를 3개 이상 작성하세요.
# Hint. Part3 > 07. Ch 08. 프로젝트를 통해 R 프로그래밍에 익숙해지기 - 01. 한국복지패널 데이터를 활용한 한국인의 삶 분석 참고

# (10점) 2. kc_house 데이터에는 층 수를 나타내는 floors 변수가 있습니다. floors 변수의 각 층별 빈도를 구하세요.

# (10점) 3. floors 변수가 1이면 'low', 1보다 크면 'high'로 구분하는 변수 'floor_level'을 만드세요.
# Hint : ifelse 함수와 $를 이용하여 해결해보세요.

# (10점) 4. 3 에서 생성한 floor_level에 따른 평균 집 값을 출력하는 함수를 작성하세요. 집 값에 관한 변수는 price입니다.
# %>%와 dplyr에서 제공하는 함수를 이용하여 해결하세요.

# (10점) 5. grade가 4인 집과 13인 집의 평균 집 값을 비교하고 싶습니다.
# grade가 4인 집은 group1, 13인 집은 group2라는 새로운 이름의 데이터 프레임 생성하세요. 이 때, 변수는 grade와 price만 추출하세요.

# (5점) 6-(1). 5에서 생성한 데이터프레임으로 grade가 4인 집과 13인 집에 대한 아래의 가설을 검정하려고 합니다.
# 귀무가설 : grade가 4인 집과 13인 집의 가격은 동일하다.
# 대립가설 : grade가 4인 집과 13인 집의 가격은 다르다.
# T검정을 이용하여 가설검정을 진행하세요. (단, 정규성과 등분산성을 만족한다고 가정합니다.***따로 검정 하지마세요.)

# (5점) 6-(2). 위 결과를 해석하세요.
# 해석 :

# r에서는 문자열의 위치에 따라 추출할 수 있는 substr함수가 있습니다. 아래 코드를 실행하여 month 라는 변수를 만들고 아래 7번 문제를 해결하세요.
#(참고: substr함수의 수행 결과는 character형태임)
# 주의 : kc_house는 1-(1)에서 지정한 데이터프레임 이름으로, 다를 경우 수정하여 코드를 실행하세요.
kc_house$month <- substr(kc_house$date, 5,6)

# (5점) 7-(1). 월 별로 평균 집값에 차이가 있는지 알고 싶습니다.
# 귀무가설 : 월 별로 평균 집 값에 차이가 없다.
# 대립가설 : 월 별로 평균 집 값에 차이가 있다.
# ANOVA 검정을 시행하세요. (단, 정규성과 등분산성을 만족한다고 가정합니다. ***따로 검정하지 마세요.)

# (5점) 7-(2). 7-(1) 결과를 해석하세요.
# 해석 :
```