3주차 과제

```
# Week_3 Quiz
# Date : 2021-05-02
# Description
# 문제 바로 아래에 답안을 작성하여, <<r script 형태>>로 저장 후 제출해주시면 됩니다.
# 과제 제출 기한은 일요일(5/2) 저녁 10까지입니다.

library(dplyr)
library(caret)

## 3주차 과제는 kc_house와 titanic_train 데이터를 활용하여 과제를 진행해주세요.
## kc_house : 집과 관련된 정보를 바탕으로 가격을 예측하는 데이터 입니다. (수치형 데이터 예측)
## titanic_train : 타이타닉 호에 탑승한 승객 정보를 바탕으로 생존여부를 예측하는 데이터입니다. (범주형 데이터 예측)


## 데이터 불러오기
kc_house <- read.csv(file = "변경필요", header= TRUE)
titanic <-  read.csv(file = "변경필요", header= TRUE)



## 1. kc_house 데이터를 활용하여 다음 조건을 충족하는 훈련, 테스트 데이터를 생성해주세요. (10점)
## 조건1) Random seed number : 2021
## 조건2) 8:2의 비율의 Train과 Test 데이터셋 구분




## 2. price를 표준화(standardization)한 데이터를 kc_house에 추가해주세요. (변수명 무관) (10점)
## 힌트 : mean(평균), sd(표준편차) 함수 활용



## 3. sqft_living를 min-max 스케일링한 데이터를 kc_house에 추가해주세요. (변수명 무관) (10점)
## 힌트
## 1) sqft_living에 대한 min과 max값을 변수로 생성
## 2) 신규 생성 된 min, max 변수와 mutate 함수를 사용하여 신규 컬럼을 생성




## 4~7번 : titanic_train 데이터를 활용하여 문제를 풀어주세요.
## titanic은 train데이터이기에 추가적으로 train / test를 분리하지 않으셔도 됩니다.


## 4. Survived 변수를 범주형 데이터로 변경해주세요. (10점)



## 5. 다음 조건을 바탕으로 생존여부(Survived)를 측정하는 Rogistic Regression 모델을 생성해주세요. (10점)
### 1) Cross Validation : 5-fold
### 2) Cross Validation Repeat : 5
### 3) Boosting 기법이 적용된 Regression Method 사용

## 힌트
## 1) trainControl Help 문서
## 2) 정상 실행 되거나, "missing value 오류"가 출력되는 2가지 케이스 모두 정답입니다.
##    (missing value 오류 여부는 Console에 출력되는 Error 메세지를 통해 확인 가능)




## 6. 데이터 전처리
## 5번 문제에서 null인 데이터가 존재할 경우 모델링이 정상적으로 수행하지 않음을 확인하였습니다.
## 전체 컬럼 중 null인 데이터가 1개라도 있으면 모두 삭제하여 5번 모델링을 재실행해주세요. (10점)

## 힌트
## 1) View 함수를 통해 null로 표시된 데이터는 어떤식으로 보이는지 확인해보시기 바랍니다.
## 2) 1안 : 참고 링크(https://m.blog.naver.com/PostView.nhn?blogId=liberty264&logNo=220992831831&proxyReferer=https:%2F%2Fwww.google.com%
## 3) 2안 : train()함수에서 지원하는 파라미터 사용하기
```

```
## 7. 6번에서 생성된 모델에 대한 해석을 작성해주세요. (10점)

## 해석 : ___ 개의 Feature를 사용하여 생존 여부를 예측하였음
## 해석 : ___ 번의 반복이 가장 높은 정확도를 보이고 있음
```