|  |
| --- |
| **데이터사이언스기초 3주차 실습과제** |
| **실습 #1 보고서** |

|  |  |
| --- | --- |
| **이름** | 김영식 |
| **학번** | 20175119 |
| **소속**  **학과/대학** | 빅데이터학과 |

## <주의사항>

* 개별 과제 입니다. (팀으로 진행하는 과제가 아니며, 모든 학생이 보고서를 제출해야 함)
* 각각의 문제 바로 아래에 답을 작성 후 제출해 주세요.
  + 소스코드/스크립트 등을 작성 한 경우, 해당 파일의 이름도 적어주세요.
* SMARTLEAD 제출 데드라인: 2022. 3. 14(월요일) ~ 3.20(일요일) 23:59 까지
  + 부정행위 적발 시, 원본(보여준 사람)과 복사본(베낀 사람) 모두 0점 처리함
  + 예외 없음
* SMARTLEAD에 아래의 파일을 제출 해 주세요
  + 보고서 형식으로 제출
  + 보고서 파일명에 이름과 학번을 입력 해 주세요.

## <실습 과제>

|  |
| --- |
| [Q 1] **서로다른 길이를 가지는 백터를 하나의 데이터프레임으로 저장하고 저장시 문제되는 해결방안에 대해서 자유롭게 서술하세요**     1. **데이터프레임 저장 코드** 2. **제시한 해결 방안에 대한 세부사항 작성**  * **작성 코드 설명** * **문제시 해결방법에 대한 설명** |

답변:

1. **데이터 프레임 저장코드**
2. **제시한 해결 방안에 대한 세부사항 작성**



그대로 실행하려 하자 오류 발생

**-오류 해결 부분**

1. **리스트로 변환 후 -> 데이터프레임으로 변환하여 처리하는 방법**

**# 백터 선언**

col1 <- c(1, 2, 3, 4)

col2 <- c("M", "F", "M", "M", "F")

col3 <- c(TRUE, FALSE, TRUE, FALSE)

**# 리스트 생성**

checkLi <- list(col1 = c(1, 2, 3, 4), col2 = c("M", "F", "M", "M", "F"), col3 = c(TRUE, FALSE, TRUE, FALSE))

**# 리스트를 데이터프레임으로 변환**

**# 세로 행을 결합해주는 rbind함수를 사용하였다.**

df1 <- data.frame(rbind(checkLi))

**#열과 행의 오류로 인하여 matrix를 transpose 해주는 t()함수를 사용하였다.**

df1 <- data.frame(t(rbind(checkLi)))

**#출력**

df1

1. **나머지 두 부분에 NA값을 넣어서 길이를 일치 시키는 방법**

**# 백터 선언**

col1 <- c(1,2,3,4)

col2 <- c("M","F","M","M","F")

col3 <- c(TRUE,FALSE,TRUE,FALSE)

**#최대 길이 구하는 부분**

maxLength <- max(length(col1), length(col2), length(col3))

**#최대길이 – 자신의 길이 만큼 NA값을 넣어준다**

col1 <- c(col1, rep(NA, maxLength-length(col1)))

col2 <- c(col2, rep(NA, maxLength-length(col2)))

col3 <- c(col3, rep(NA, maxLength-length(col3)))

**#데이터 프레임 생성 후 출력**

DF1 <- data.frame(col1,col2,col3)

DF1

## <실습 과제>

|  |
| --- |
| [Q 2] R의 내장 데이터셋인 ‘cars’를 이용하여 다음과 같은 결과를 추출하고자 한다.  **<결과 예시>**    **(1) 합계, 평균, 최대값 순으로 저장하여, speed와 dist 계산 결과를 하나의 데이터프레임으로 만들어 출력한다.**  **(2) 지금까지 실습했던 방법 아무거나 사용가능**  **(3) 자신이 작성했던 코드와 결과 첨부(작성한 코드에 설명 작성)** |

답변:

**# speed**

**# 합계**

sumSpeed <- sum(cars$speed)

**# 평균**

meanSpeed <- mean(cars$speed)

**# 최대값**

maxSpeed <- max(cars$speed)

**# speed 데이터프레임 생성**

speeddf <- data.frame(sumSpeed, meanSpeed, maxSpeed)

**# dist**

**# 합계**

sumdist <- sum(cars$dist)

**# 평균**

meandist <- mean(cars$dist)

**# 최대값**

maxdist <- max(cars$dist)

**# dist 데이터프레임 생성**

distdf <- data.frame(sumdist, meandist, maxdist)

speedDF <- t(speeddf)

distDF <- t(distdf)

**# speed와 dist 데이터프레임 결합**

DF <- cbind(speedDF, distDF)

rownames(DF) <- c("1","2","3")

colnames(DF) <- c("speed", "dist")

DF