빅데이터 분석시스템 개발

김은표

Exam.csv 파일의 데이터프레임 출력

- df_exam <- read.csv('C:/Users/bigdata/Desktop/workspace/R/file/exam.csv')
- View(df_exam)

	id	class	math	english	science
1	1	1	50	98	50
2	2	1	60	97	60
3	3	1	45	86	78
4	4	1	30	98	58
5	5	2	25	80	65
6	6	2	50	89	98
7	7	2	80	90	45
8	8	2	90	78	25
9	9	3	20	98	15
10	10	3	50	98	45
11	11	3	65	65	65
12	12	3	45	85	32
13	13	4	46	98	65
14	14	4	48	87	12
15	15	4	75	56	78
16	16	4	58	98	65
17	17	5	65	68	98
18	18	5	80	78	90
19	19	5	89	68	87
20	20	5	78	83	58

Math, English, Science 출력

- df_math <- df_exam %>% select(math)
- df_english <- df_exam %>% select(english
- df_result1 <- df_exam %>% select(science)

```
> df_math
   math
1    50
2    60
3    45
4    30
5    25
6    50
7    80
8    90
9    20
10   50
11   65
12   45
13   46
14   48
15   75
16   58
17   65
18   80
19   89
20   78
```

```
> df_english
   english
1       98
2       97
3       86
4       98
5       80
6       89
7       90
8       78
9       98
10      98
11      65
12      85
13      98
14      87
15      56
16      98
17      68
18      78
19      68
20      83
```

```
> df_result1
   science
1       50
2       60
3       78
4       58
5       65
6       98
7       45
8       25
9       15
10      45
11      65
12      32
13      65
14      12
15      78
16      65
17      98
18      90
19      87
20      58
```

Class가 1인 모든 변수 출력

- df_filter1 <- df_exam %>% select(everything()) %>% filter(class == 1)
- df_filter1

	id	class	math	english	science
1	1	1	50	98	50
2	2	1	60	97	60
3	3	1	45	86	78
4	4	1	30	98	58

Math 데이터 출력

- df_filter2 <- df_exam %>% select(math) %>% filter(math >= 60 & math < 80)
- df_filter2

```
> df_filter2 <- df_exam %>% select(math) %>% filter(math >= 60 & math < 80)
> df_filter2
```

	math
1	60
2	65
3	75
4	65
5	78

English 데이터 출력

- df_filter2 <- df_exam %>% select(english) %>%
  filter(english >= 60 & english < 80)
- df_filter2

```
> df_filter2 <- df_exam %>% select(english) %>% filter(english >= 60 & english < 80)
> df_filter2
```

	english
1	78
2	65
3	68
4	78
5	68

Science 데이터 출력

- df_filter2 <- df_exam %>% select(science) %>%
  filter(science >= 60 & science < 80)
- df_filter2

```
> df_filter2 <- df_exam %>% select(science) %>% filter(science >= 60 & science < 80)
> df_filter2
```

	science
1	60
2	78
3	65
4	65
5	65
6	78
7	65

Science 데이터 출력

- df_filter6 <- df_exam %>% select(class,id,math) %>% filter(math >= 60)
- df_filter6

```
> df_filter6 <- df_exam %>% select(class,id,math) %>% filter(math >= 60)
> df_filter6
```

	class	id	math
1	1	2	60
2	2	7	80
3	2	8	90
4	3	11	65
5	4	15	75
6	5	17	65
7	5	18	80
8	5	19	89
9	5	20	78

Total 파생변수 출력

- df_total <- df_exam %>% mutate(total = math + english + science)
- df_total

```
df_total <- df_exam %>% mutate(total = math + english + science)
df_total
```

```
> df_total
   id class math english science total
1   1     1   50      98      50   198
2   2     1   60      97      60   217
3   3     1   45      86      78   209
4   4     1   30      98      58   186
5   5     2   25      80      65   170
6   6     2   50      89      98   237
7   7     2   80      90      45   215
8   8     2   90      78      25   193
9   9     3   20      98      15   133
10 10     3   50      98      45   193
11 11     3   65      65      65   195
12 12     3   45      85      32   162
13 13     4   46      98      65   209
14 14     4   48      87      12   147
15 15     4   75      56      78   209
16 16     4   58      98      65   221
17 17     5   65      68      98   231
18 18     5   80      78      90   248
19 19     5   89      68      87   244
20 20     5   78      83      58   219
```

Mean 파생변수 출력

- df_mean <- df_exam %>% mutate(total = math + english + science) %>% mutate(mean = total/3)
- df_mean

```
df_mean <- df_exam %>% mutate(total = math + english + science) %>% mutate(mean = total/3)
df_mean
```

```
> df_mean
   id class math english science total     mean
1   1     1   50      98      50   198 66.00000
2   2     1   60      97      60   217 72.33333
3   3     1   45      86      78   209 69.66667
4   4     1   30      98      58   186 62.00000
5   5     2   25      80      65   170 56.66667
6   6     2   50      89      98   237 79.00000
7   7     2   80      90      45   215 71.66667
8   8     2   90      78      25   193 64.33333
9   9     3   20      98      15   133 44.33333
10 10     3   50      98      45   193 64.33333
11 11     3   65      65      65   195 65.00000
12 12     3   45      85      32   162 54.00000
13 13     4   46      98      65   209 69.66667
14 14     4   48      87      12   147 49.00000
15 15     4   75      56      78   209 69.66667
16 16     4   58      98      65   221 73.66667
17 17     5   65      68      98   231 77.00000
18 18     5   80      78      90   248 82.66667
19 19     5   89      68      87   244 81.33333
20 20     5   78      83      58   219 73.00000
```

Grade 파생변수 출력

- df_grade <- df_exam %>% mutate(total = math + english + science) %>%
- mutate(mean = total/3) %>%
- mutate(grade = ifelse(mean >= 90, 'A', ifelse(mean >= 80, 'B',
- ifelse(mean >= 70, 'C', ifelse(mean >= 60, 'D', 'F')))))
- df_grade

```
> df_grade
   id class math english science total     mean grade
1   1     1   50      98      50   198 66.00000     D
2   2     1   60      97      60   217 72.33333     C
3   3     1   45      86      78   209 69.66667     D
4   4     1   30      98      58   186 62.00000     D
5   5     2   25      80      65   170 56.66667     F
6   6     2   50      89      98   237 79.00000     C
7   7     2   80      90      45   215 71.66667     C
8   8     2   90      78      25   193 64.33333     D
9   9     3   20      98      15   133 44.33333     F
10 10     3   50      98      45   193 64.33333     D
11 11     3   65      65      65   195 65.00000     D
12 12     3   45      85      32   162 54.00000     F
13 13     4   46      98      65   209 69.66667     D
14 14     4   48      87      12   147 49.00000     F
15 15     4   75      56      78   209 69.66667     D
16 16     4   58      98      65   221 73.66667     C
17 17     5   65      68      98   231 77.00000     C
18 18     5   80      78      90   248 82.66667     B
19 19     5   89      68      87   244 81.33333     B
20 20     5   78      83      58   219 73.00000     C
```