<R로 배우는 데이터 분석 입문 2차 과제 답안지>

가족자원경영학과 1916844 장은서

1.

```
frequency_three %>% filter(president == "황선혜"), frequency_three %>% filter(president == "강정애"), frequency_three %>% filter(president == "장윤금")
```

답 - 황선혜 : 우리대학,제도 / 강정애 : 헌신 / 장윤금 : 디지털

2.

```
raw_J <- raw_NYA %>% filter(president == "장윤금")
NYA_J_sentences <- raw_J %>% as_tibble() %>% unnest_tokens(input = value, output = sentence, token = "sentences")
NYA_J_sentences %>% filter(str_detect(sentence, "디지털") & str_detect(sentence, "데이터"))
```

답 - 이를, 인간의

3.

```
frequency_three %>% filter(president == "황선혜") %>%slice_max(n, n = 10)
frequency_three %>% filter(president == "황선혜") %>% slice_max(tf_idf, n = 10)
frequency_three %>% filter(president == "황선혜") %>% filter(n >= 5 & tf_idf >= 0.007042386)
```

답 - 우리대학, 제도

4.

```
frequency_three %>% filter(president == "강정애") %>%slice_max(n, n = 10)
frequency_three %>% filter(president == "강정애") %>% head(10)
frequency_three %>% filter(president == "강정애") %>% filter(n >= 4 & tf_idf >= 0.009117114)
```

답 - 헌신, 성과

5. 특정 텍스트에서의 빈도수는 높아 TF는 크지만, 세 총장님의 텍스트 모두에서 흔하게 사용되어 IDF가 작기 때문이다.

6.

```
frequency_two <- frequency_three %>% filter(president == "장윤금" | president == "강정애")
frequency_two$tf <- NULL
frequency_two$idf <- NULL
frequency_two$tf_idf <- NULL

frequency_wide <- frequency_two %>% pivot_wider(names_from = president, values_from = n, values_fill = list(n=0))

frequency_wide <- frequency_wide %>% mutate(sum = 강정애 + 장윤금)
frequency_wide %>% filter(장윤금 != 0 & 강정애 != 0) %>% arrange(-sum)
```

답 – 교육

7.

```
frequency_wide <- frequency_wide %>% mutate(ratio_jang = ((장윤금 + 1)/(sum(장윤금 + 1))), ratio_kang = ((강정애 + 1)/(sum(강정애 + 1))))
frequency_wide <- frequency_wide %>% mutate(odds_ratio = ratio_jang / ratio_kang)
frequency_wide %>% arrange(odds_ratio)
```

답 – 성과, 헌신, 0.130(성과 odds-ratio), 0.182(헌신 odds-ratio)

8.

```
frequency_wide %>% arrange(abs(1-odds_ratio))

frequency_wide <- frequency_wide %>% mutate(abs = abs(1-odds_ratio))

df <- frequency_wide %>% filter(odds_ratio == 0.9100346)
```

답 - 가족, 0.910

9.

```
dic <- read_csv("knu_sentiment_lexicon.csv")
```

```
word_NYA <- raw_NYA %>% unnest_tokens(input = value, output = word, token = "words", drop = F)

word_NYA <- left_join(word_NYA, dic, by = "word") %>% mutate(polarity = ifelse(is.na(polarity), 0, polarity))

df2 <- word_NYA %>% filter(polarity == 2)

n_distinct(df2$word)
```

답 – 15개

```
10. score_NYA <- word_NYA %>% group_by(president,value) %>% summarise(score = sum(polarity)) %>% ungroup()
```

답 – 강정애 : 23, 장윤금 : 25, 황선혜 : 24