"Hocance" is a combination of the words "hotel" and "vacation" referring to a staycation at a hotel. With the prolonged COVID-19 pandemic, the hocance trend is gaining huge popularity in Korea. For example, Hyundai Home Shopping Network earned 1 billion won (about US$855,431) in sales through the 'Alpensia Five Star Hotel Package' promotion. Moreover, according to Hotels Combine, a global hotel search engine, 8 out of 10 respondents in their 20s and 30s said they have experience or plan to go on hocance, and about 72.1% of respondents in their 40s and 50s said they have experience or are willing to experience hocance, indicating that they were interested in hocance.

This paper utilizes text mining techniques in R to find out whether hocance trend can be a recovery keyword to the hotel industry, which has been hit hardest by COVID-19, or whether it is meaningless because it is a temporary phenomenon. Since the word "hocance" is not English, this paper uses the words "staycation" and "hospitality" to replace the word "hocance".

When analyzed based on the frequency of tokens (Figure1), the token "trend" recorded the 17th highest frequency of 11. This value is not a high value when determined by numerical values, but it is a high frequency of about 2.4 percent of the total 1559 tokens. This is similar to the frequency of words that seem to need to be interested in, such as "domestic", "time" and "safety". This paper ran TF-IDF analysis as the next step, because simply checking the frequency of tokens was meaningless. The analysis (Figure2) shows that the token "trend" is significant from a business perspective. From a business perspective, the most significant word was "vacation". As a final step, this paper conducted an emotional analysis based on the NRC lexicon. Words can be analyzed into nine parts: anger, anticipation, Fear, joy, negative, positive, sadness, surprise, and trust. According to the results (Figure3), the words are generally positive. The most meaningful token in the previous step, "vacation", was identified as 'positive', 'joy', and 'anticipation' according to the emotional analysis. In other words, despite the pandemic, the customer has expectations. In addition, the token "economy" appears as 'trust' and the token "luxury" is also expressed as 'positive' and 'joy. ' Therefore, in the short term, it is believed that luxury packaging could help the hotel industry recover. In addition, 'surprise' and 'anticipation' can be found on the token "expect", and 'joy' and 'anticipation' can be found on the token "holiday". Taking this into consideration, this paper proposes an event packaging in addition to the luxury packaging, high-priced hocance. In the previous frequency analysis, "domestic" was one of the words that showed the same frequency as "trend". This paper presents an event packaging for domestic customers. By making changes to the interior props and diet provided, it will give customers a feeling of being abroad. This will satisfy customers' desire to travel abroad, which is being restricted by pandemic. Not only "domestic" but also "time" and "safe" are tokens that this paper had interest in the previous frequency analysis. Given that the token "time" is expressed as 'anticipation,' hocance in hotels in the city can be attractive to customers. The tokens "safe" are expressed as 'trust' and 'joy, ' and it would be nice to plan a hocance by appealing that the hotel is safe for customers. For example, after receiving a quarantine certification mark or sterilizing the room, this paper is suggested to create a step for customers to feel safe by attaching a seal that seals the door. The tokens highlighted in the sentiment analysis is "leisure" which appears as 'trust', 'surprise', 'positive', 'joy' and 'anticipation'. In addition, given that the token "beach" appears to be "joy", it is believed that a hocance, including marine leisure, will appeal to customers. On the other hand, low and middle priced hocance is likely to survive relatively late. The token "demand" appears as 'negative' and 'anger', and there are no factors that can neutralize it, such as the token "luxury" implied by the high priced hocance. Also, the expression of token "change" as 'fear' also seems to support the prediction of low and middle priced hocance.

In conclusion, hocance is a trend that can help the hotel industry recover. The word "staycation", which is currently used worldwide, is not suitable for this trend. It will be necessary to define new words for this trend, such as hocance. This paper proposes "Hocation" as a word for representing this trend. Hocation will not just be a short-lived phenomenon caused by pandemic, but it is expected to establish itself as a part of the hotel's profit structure.

Reference:

Lim. S. (Aug. 10, 2020) *Why the staycation boom is exactly what hotels need*. retrieved from https://www.jll.co.kr/en/trends-and-insights/cities/why-the-staycation-boom-is-exactly-what-hotels-need

Krishnan. V and Mann R. (Jun. 10, 2020) *Hospitality and COVID-19: How long until 'no vacancy' for US hotels?* retrieved from https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/hospitality-and-covid-19-how-long-until-no-vacancy-for-us-hotels

Cohen A. (Jun. 12, 2020) *Hotels Are Betting on Staycations to Survive Covid-19*. retrieved from https://www.bloomberg.com/news/articles/2020-06-12/hotels-are-betting-on-staycations-to-survive-covid-19-recession

Economist (Aug. 31, 2020) *Staycationers are saving hotels and Airbnb from covid-19*. retrieved from https://www.economist.com/business/2020/08/31/staycationers-are-saving-hotels-and-airbnb-from-covid-19

Cohen A. (Jun. 14, 2020) *Survival during pandemic: Hotel industry is betting on staycations*. retrieved from https://www.business-standard.com/article/current-affairs/survival-during-pandemic-hotel-industry-is-betting-on-staycation-120061400025_1.html

Elliott C. (Sep. 4, 2020) *The Surprising Reason Staycations Are The New Vacations*. retrieved from https://www.forbes.com/sites/christopherelliott/2020/09/04/the-surprising-reason-staycations-are-the-new-vacations/?sh=6c23f38c140a

Song A. (Oct. 3, 2020) *Tickets for Hotels and Resorts Popular Thanks to Increasing Demand for 'Hocance'*. retrieved from http://koreabizwire.com/tickets-for-hotels-and-resorts-popular-thanks-to-increasing-demand-for-hocance/170919

Kim J. (Feb 3, 2021) *HOTELS COMBINE announces 2021 Hocance Trends*. retrieved from https://www.venturesquare.net/823081

Appendix

CODE

```
# Load packages
library(pdftools)
library(stringr)
library(dplyr)
library(tidyverse)
library(tidytext)
library(tidyr)
library(ggplot2)


# Import all PDF files from the same folder
setwd("C:/wd/HULT/Text Analytics and NLP/individual")
nm <- list.files(path="C:/wd/HULT/Text Analytics and NLP/individual")
my_pdf_text <- do.call(rbind, lapply(nm, function(x) paste(pdf_text(x),
collapse = " "))) #1 variable
colnames(my_pdf_text) <- c("text")
mydf <- data.frame(line=1:6, text = my_pdf_text[,1])


# Tokenize the mydf dataframe
token_list <- mydf %>% unnest_tokens(word, text)
# Check the result
#print(token_list)


# Check the token frequencies
frequencies_tokens <- mydf %>%
   unnest_tokens(word, text) %>%
   anti_join(stop_words) %>%
   count(word, sort=TRUE)
# Check the result
print(frequencies_tokens)


# Remove meaningless words: numbers, units, auxiliary verbs and etc.
mystopwords <- data_frame(word =
```

```
c("19","cent","months","percent","50","pre","2020","we’re","1",  "2",  "25", "60",

"70","a3","10","100","30","5","85","11","20","2019","3","36","40","90","a1",

"what's","1.1","180","35","4","400","66","750","three","month","was","during",
                              "you","now","those"))
clean_frequencies_tokens <- anti_join(frequencies_tokens, mystopwords, by = "word")


#############
# Look at the graphical approach:
freq_hist <- clean_frequencies_tokens %>%
   filter(n > 8) %>% # we need this to eliminate all the low count words
   mutate(word=reorder(word, n)) %>%
   ggplot(aes(word, n))+
   geom_col()+
   labs(x = NULL, y = "frequencies")+
   coord_flip()
print(freq_hist)



########################################
#####     TF-IDF framework      #####
########################################
each_words <- mydf %>%
   unnest_tokens(word, text) %>%
   count(line, word, sort = TRUE) %>%
   ungroup()

each_words <- each_words %>%
   bind_tf_idf(word, line, n)

# Remove meaningless words: numbers, units, auxiliary verbs and etc.
each_words <- anti_join(each_words, mystopwords, by = "word")
```

```
# Check the result
each_words %>%
   arrange(desc(tf_idf))

#############
# Look at the graphical approach:
each_words %>%
   arrange(desc(tf_idf)) %>%
   mutate(word = factor(word, levels = rev(unique(word)))) %>%
   group_by(line) %>%
   top_n(9) %>%
   ungroup %>%
   ggplot(aes(word, tf_idf, fill = line)) +
   geom_col(show.legend = FALSE) +
   labs(x = NULL, y = "tf-idf") +
   facet_wrap(~line, ncol = 2, scales = "free") +
   coord_flip()


#######################################
#####    Sentiment analysis     #####
#######################################
token_words <- mydf %>%
   unnest_tokens(word, text) %>%
   count(line, word, sort = TRUE) %>%
   ungroup()

# Remove meaningless words: numbers, units, auxiliary verbs and etc.
token_words <- anti_join(token_words, mystopwords, by = "word")

# Get sentiments based on NRC lexicon
token_words_sentiments <- token_words %>%
   inner_join(get_sentiments("nrc"), by = "word")
```

```
#############
# Look at the graphical approach:
token_words_sentiments %>%
    rename(
        document = line,
        term = word,
        count = n,
        sentiment = sentiment
    ) %>%
    count(sentiment, term, wt = count) %>%
    ungroup() %>%
    filter(n >= 5) %>%
    mutate(n = ifelse(sentiment == "negative", -n, n)) %>%
    mutate(term = reorder(term, n)) %>%
    ggplot(aes(n, term, fill = sentiment)) +
    geom_col() +
    labs(x = "Contribution to sentiment", y = NULL)
```

OUTPUT

Figure1 : token frequencies

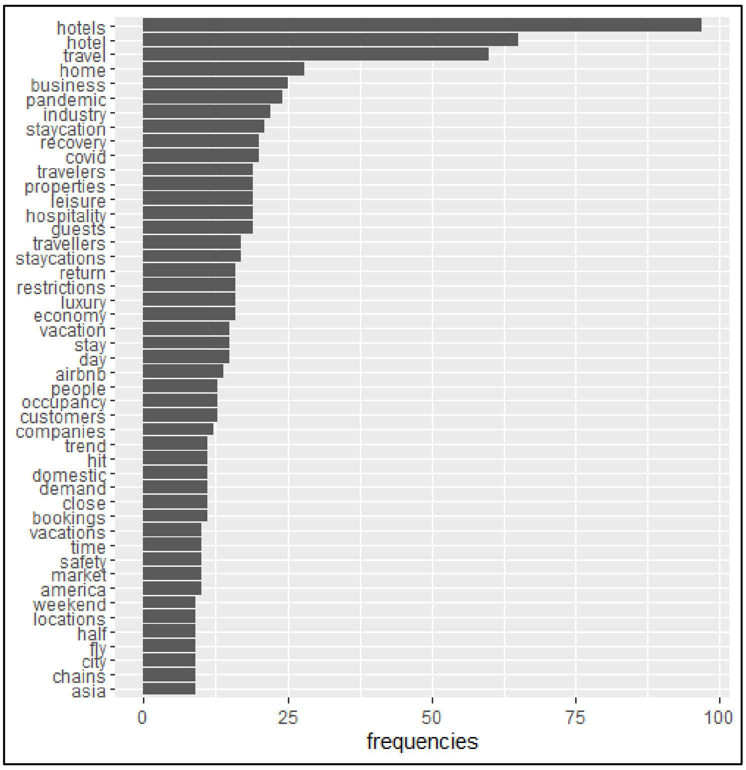

Figure2: TF-IDF framework

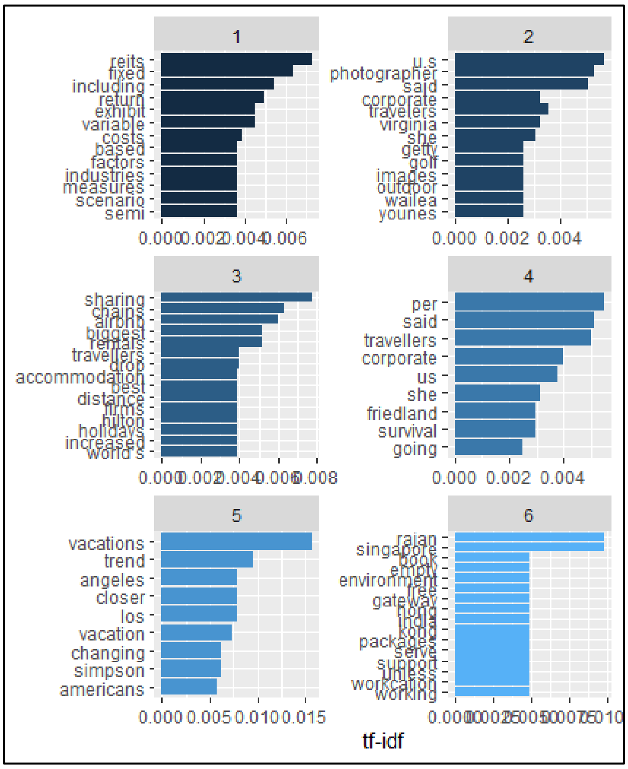

Figure3 : Sentiment analysis