R만 하조

리뷰를 활용하여 나에게 맞는 호텔 찾기

학과	학번	이름
정보통계학과	20181067	이채연
정보통계학과	20181072	정현정
정보통계학과	20181076	주다희
정보통계학과	20181080	황세원

목차

서론

주제 선정 배경과
목적 및 기대효과

본론

자료 선정 및 분석
빅데이터 분석과 해석
- 워드크라우드
- 토픽모델링
- 감성분석

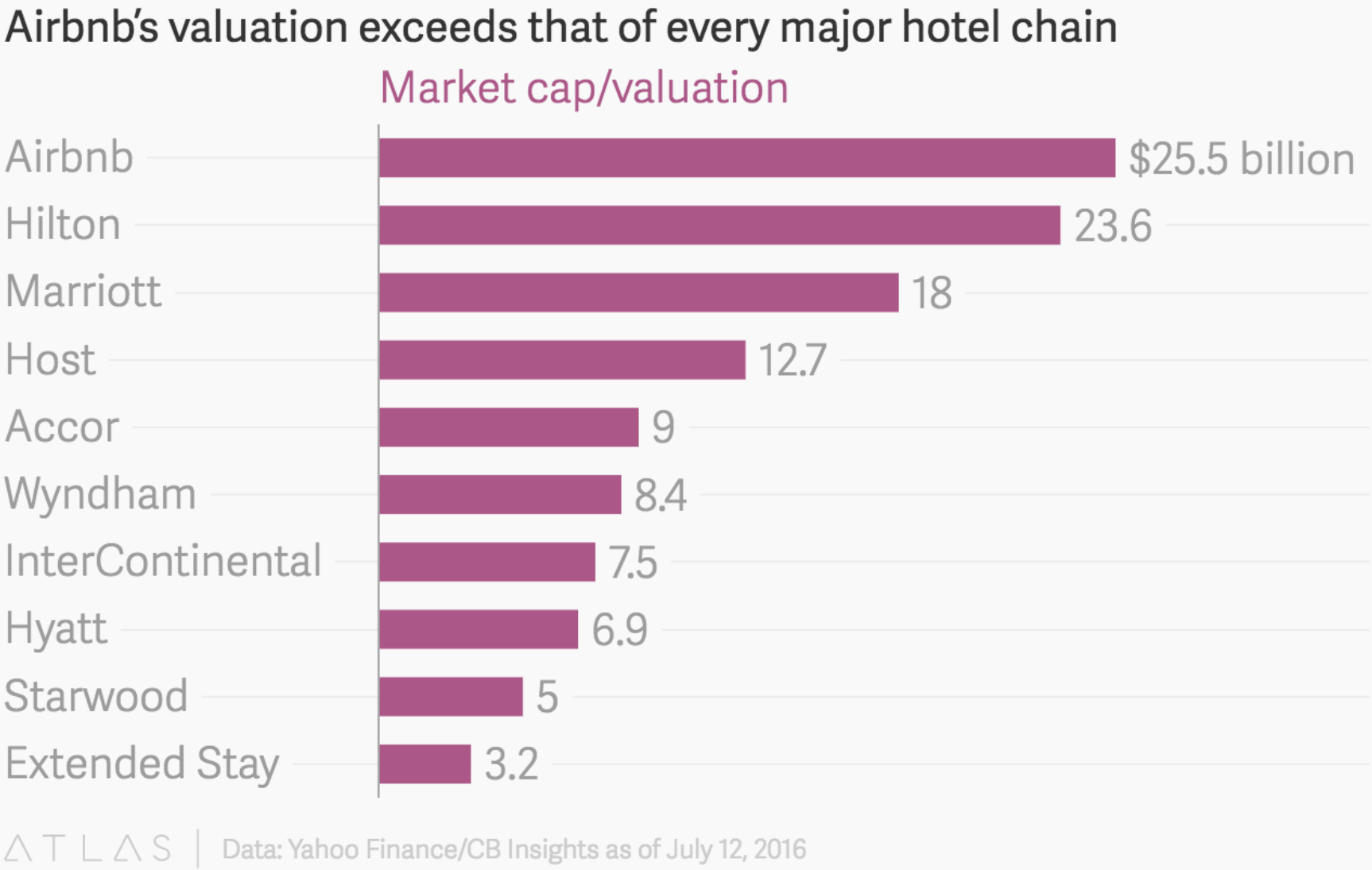
빅데이터 활용

결론

발표 요약 및 한계점,
향후 연구 방향

질의응답

주제 선정 배경


Airbnb's valuation exceeds that of every major hotel chain
Market cap/valuation
Airbnb
$25.5 billion
Hilton
23.6
Marriott
18
Host
12.7
Accor
9
Wyndham
8.4
InterContinental
7.5
Hyatt
6.9
Starwood
5
Extended Stay
3.2
ATLAS
Data: Yahoo Finance/CB Insights as of July 12, 2016

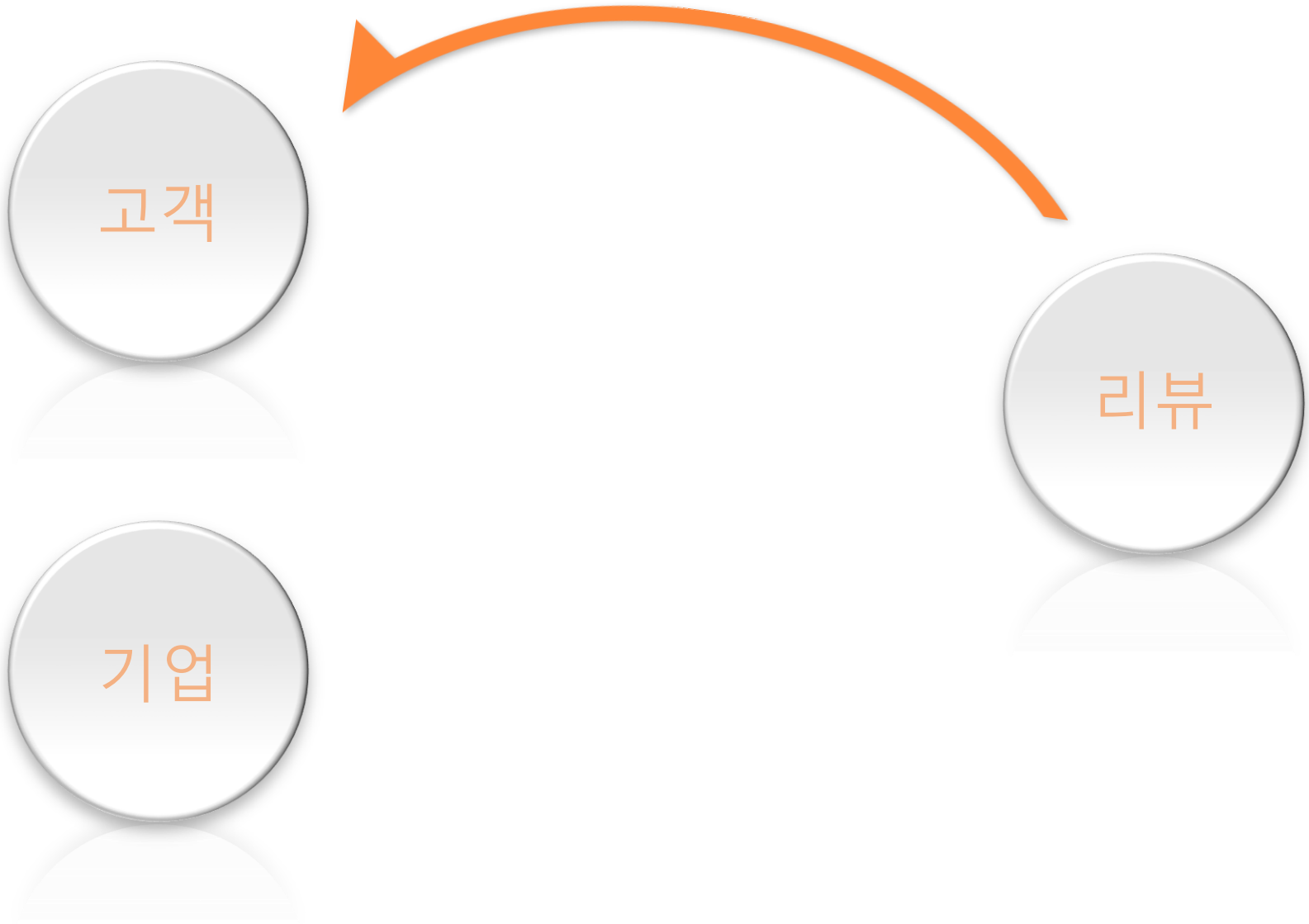
목적 및 기대효과

고객

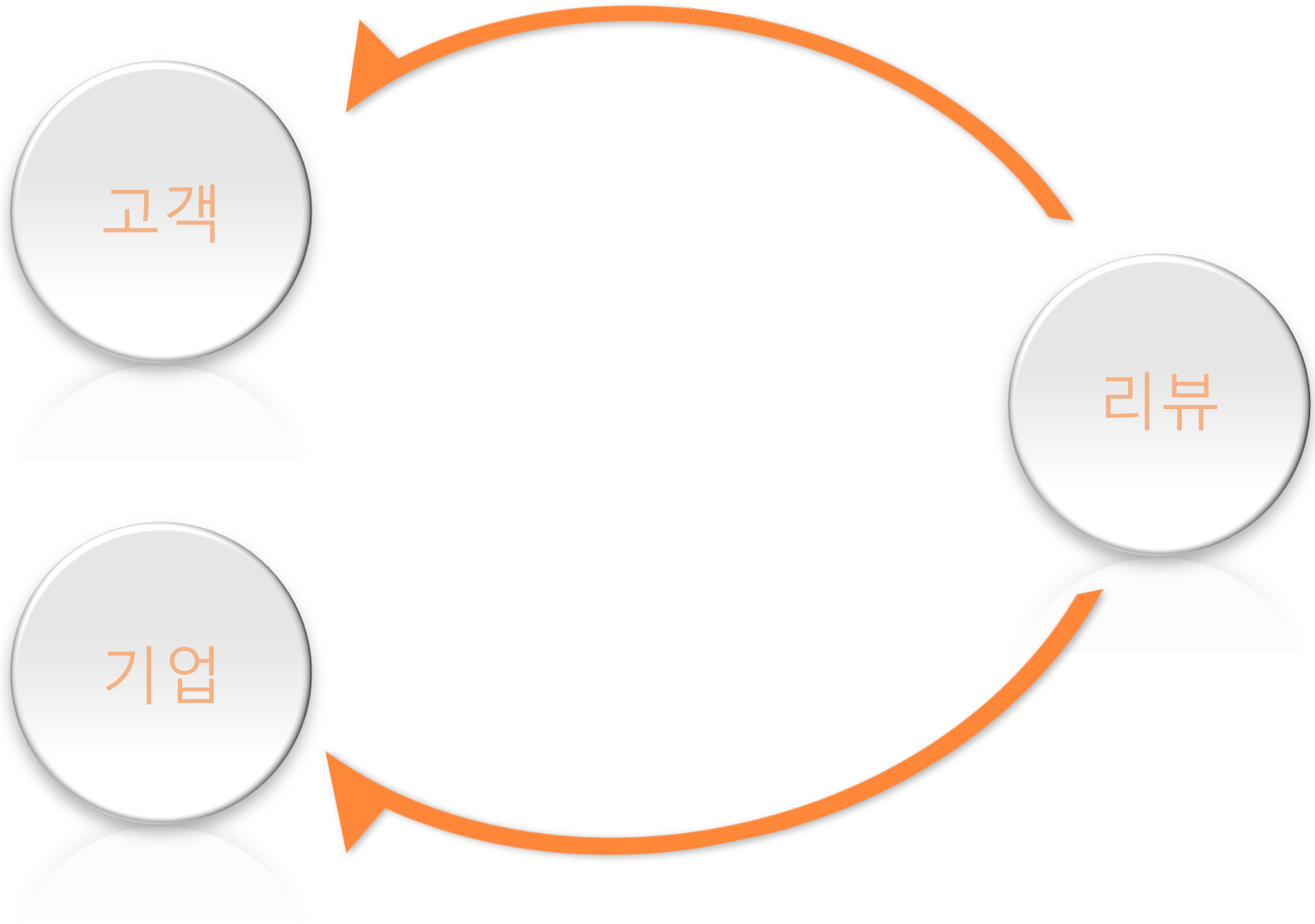
기업

리뷰

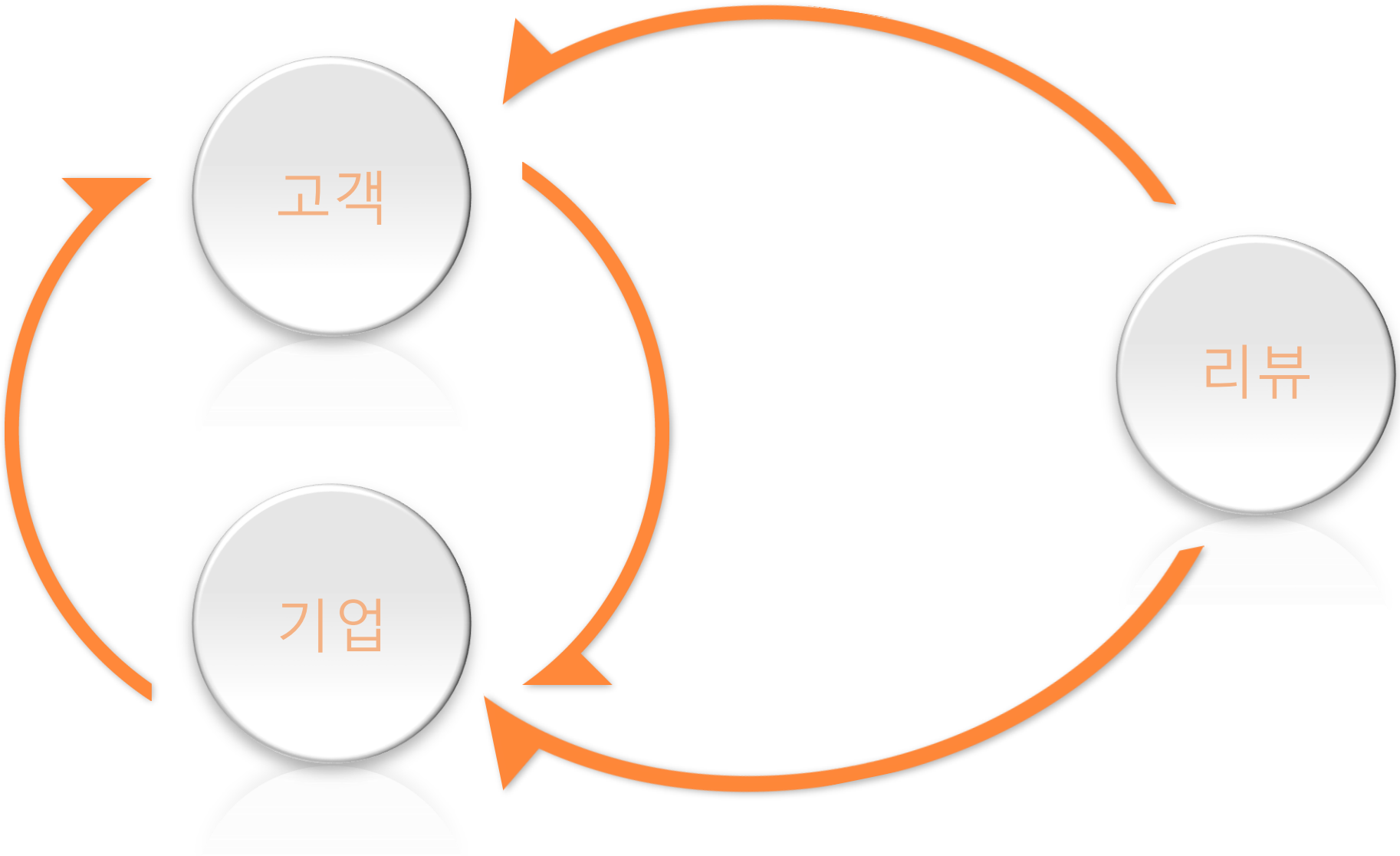
목적 및 기대효과



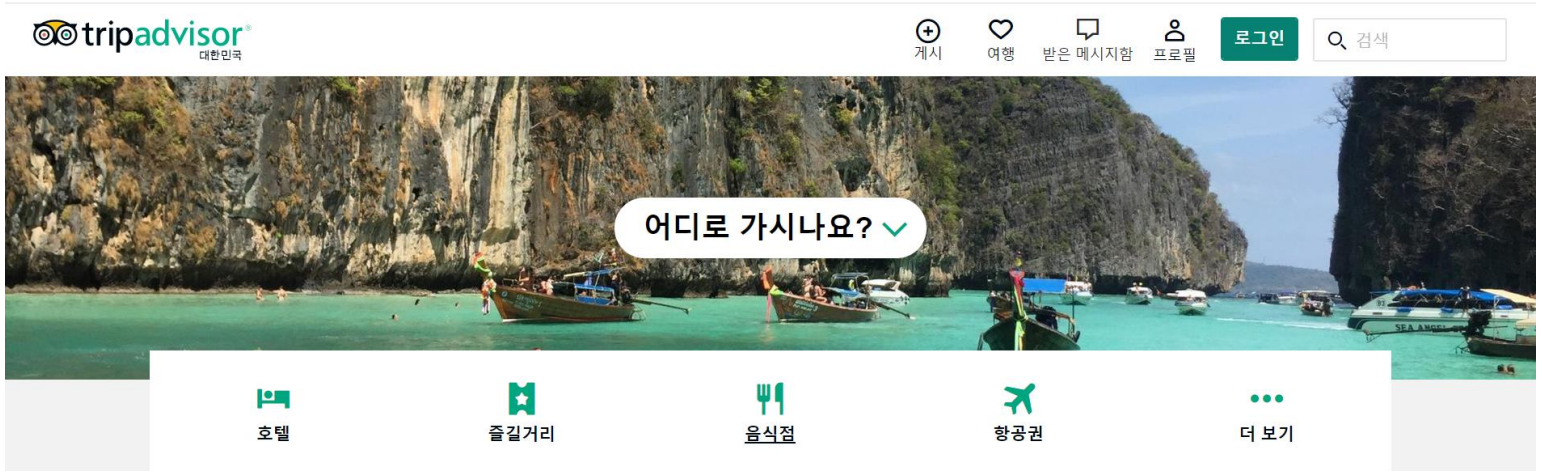
목적 및 기대효과



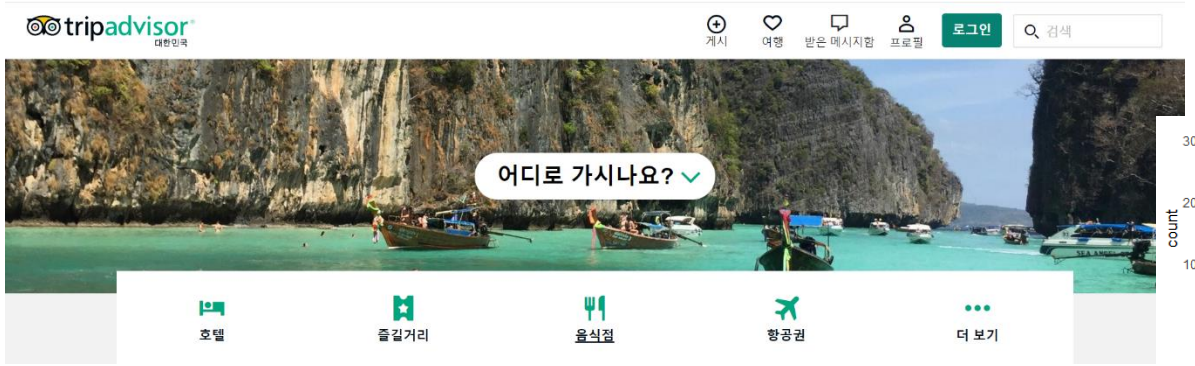
목적 및 기대효과

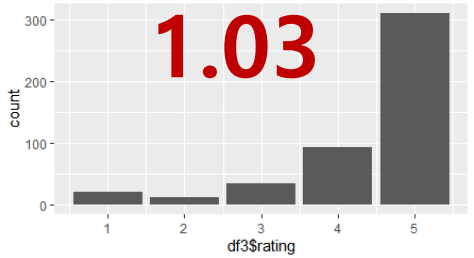


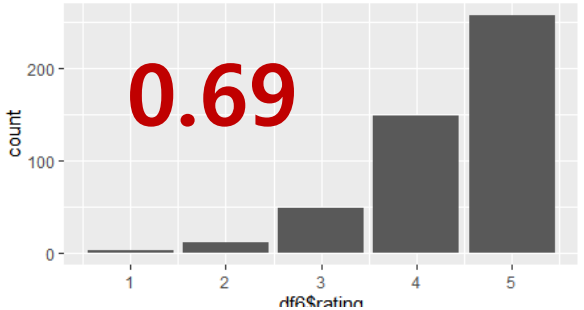
자료선정

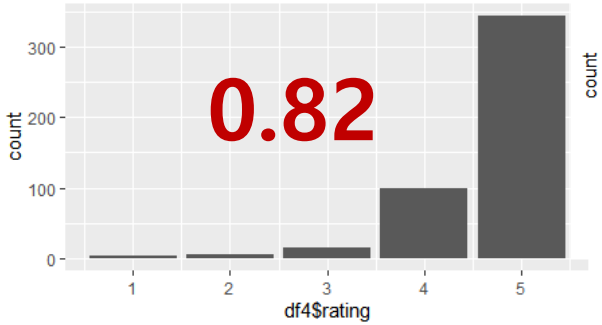

tripadvisor
대한민국
게시
여행
받은 메시지함
프로필
로그인
검색
어디로 가시나요?
호텔
즐길거리
음식점
항공권
더 보기

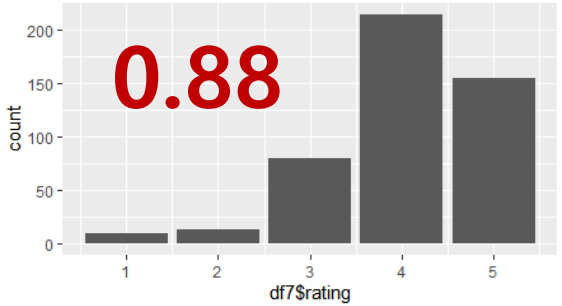
자료선정


tripadvisor
대한민국
게시
여행
받은 메시지함
프로필
로그인
검색
어디로 가시나요?
호텔
즐길거리
음식점
항공권
더 보기

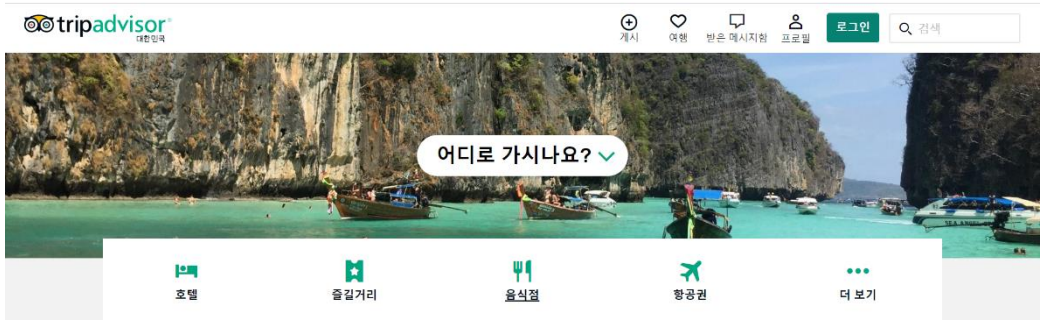

1.03
count
df3$rating

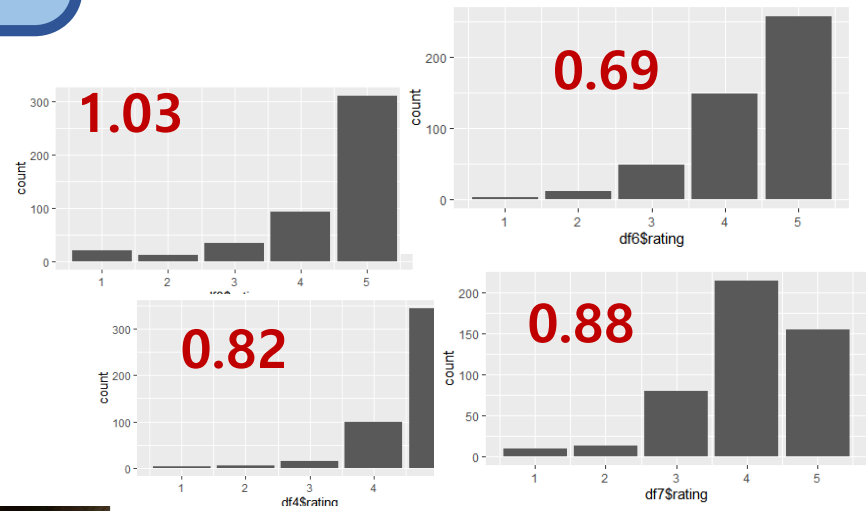

0.69
count

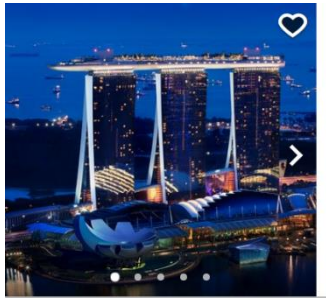

0.82
count
df4$rating


0.88
count
df7$rating

자료선정





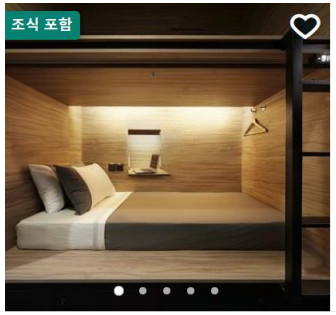


마리나 베이 샌즈

agoda
₩646,592
특가보기
✓ 무료 취소

MarinaBaySands ↗ ₩625,924
Expedia ↗ ₩887,938
모두투어 ↗ ₩833,400
8건의 특가 전체(최저 ₩625,924)

28,796건의 리뷰
싱가포르의 772 숙박 장소 중에서 1위로 가성비가 최고인 숙박 장소
무료 와이파이
수영장



더 포드

₩97,740
특가보기

Agoda.com ↗ ₩89,300
호텔스닷컴 ↗ ₩97,790
Trip.com ↗ ₩127,080
8건의 특가 전체(최저 ₩89,300)

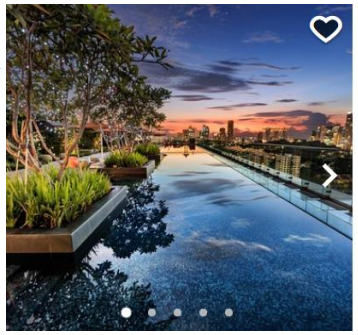
캡슐 호텔
1,390건의 리뷰
싱가포르에서 필터와 일치하는 가성비 최고 호텔 중 11위
무료 와이파이

래플스 호텔

₩965,384
특가보기

Raffles ↗ ₩969,729
Expedia ↗ ₩972,287
Booking.com ↗ ₩972,081
8건의 특가 전체(최저 ₩965,384)

4,660건의 리뷰
싱가포르의 772 숙박 장소 중에서 4위로 가성비가 최고인 숙박 장소
무료 와이파이
수영장
호텔 웹사이트 방문 ↗



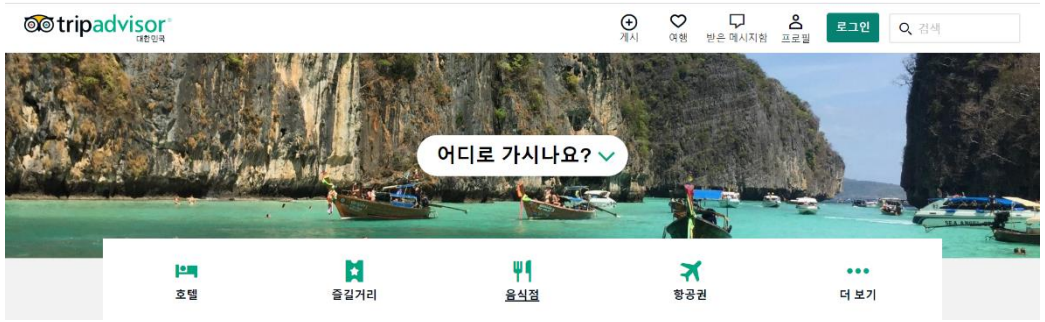
Hotel Jen Orchardgateway Singapore by Shangri-La

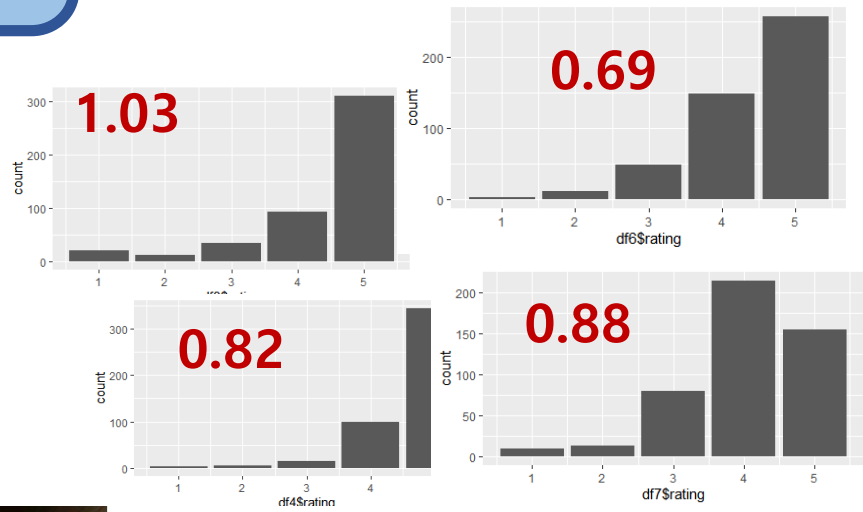
₩249,408
특가보기

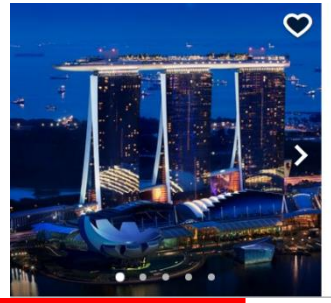
Agoda.com ↗ ₩240,827
Booking.com ↗ ₩241,246
ZenHotels.com ↗ ₩261,588
8건의 특가 전체(최저 ₩240,827)

2,867건의 리뷰
싱가포르의 772 숙박 장소 중에서 2위로 가성비가 최고인 숙박 장소
무료 와이파이
수영장
스페셜 프로모션
호텔 웹사이트 방문 ↗

자료선정





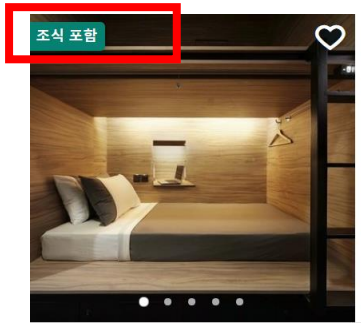


마리나 베이 샌즈

agoda
₩646,592
특가보기
✓ 무료 취소

MarinaBaySands ↗ ₩625,924
Expedia ↗ ₩887,938
모두투어 ↗ ₩833,400
8건의 특가 전체(최저 ₩625,924)

28,796건의 리뷰
싱가포르의 772 숙박 장소 중에서 1위로 가성비가 최고인 숙박 장소
무료 와이파이
수영장

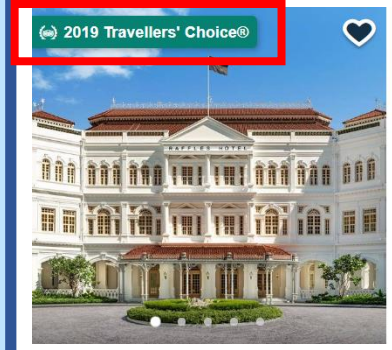


더 포드

₩97,740
특가보기

Agoda.com ↗ ₩89,300
호텔스닷컴 ↗ ₩97,790
Trip.com ↗ ₩127,080
8건의 특가 전체(최저 ₩89,300)

캡슐 호텔
1,390건의 리뷰
싱가포르에서 필터와 일치하는 가성비 최고 호텔 중 11위
무료 와이파이

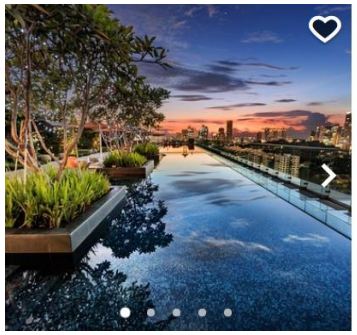


래플스 호텔

Trip.com
₩965,384
특가보기

Raffles ↗ ₩969,729
Expedia ↗ ₩972,287
Booking.com ↗ ₩972,081
8건의 특가 전체(최저 ₩965,384)

4,660건의 리뷰
싱가포르의 772 숙박 장소 중에서 4위로 가성비가 최고인 숙박 장소
무료 와이파이
수영장
호텔 웹사이트 방문 ↗



Hotel Jen Orchardgateway Singapore by Shangri-La

₩249,408
특가보기

Agoda.com ↗ ₩240,827
Booking.com ↗ ₩241,246
ZenHotels.com ↗ ₩261,588
8건의 특가 전체(최저 ₩240,827)

2,867건의 리뷰
싱가포르의 772 숙박 장소 중에서 2위로 가성비가 최고인 숙박 장소
무료 와이파이
수영장
스페셜 프로모션
호텔 웹사이트 방문 ↗

자료수집 - 크롤링

```r
tripadvisor_scraper <- function(pid, url){ # url주소에서 HTML소스 가져오기
  doc <- read_html(url)

  # 호텔명
  hotel <- doc %>%
    html_nodes("#HEADING") %>%
    html_text() %>%
    str_trim()

  # 작성자
  name <- doc %>%
    html_nodes("div.social-member-event-MemberEventOnObjectBlock__event_type--3njyv > span > a") %>%
    html_text() %>%
    str_trim()

  # 리뷰제목
  commenttitle <- doc %>%
    html_nodes("div.location-review-review-list-parts-ReviewTitle__reviewTitle--2GO9Z > a > span > span") %>%
    html_text() %>%
    str_trim()

  #리뷰내용
  comment <- doc %>%
    html_nodes("div.location-review-review-list-parts-ExpandableReview__containerStyles--1G0AE > div._2f_ruteS._1bona3Pu._2uD5bLZZ > div > q") %>%
    html_text() %>%
    str_trim()

  #평점
  rating <- doc %>%
    html_nodes("div.location-review-review-list-parts-SingleReview__mainCol--1hApa > div.location-review-review-list-parts-RatingLine__container--2bjtw > div > span") %>%
    html_attr("class")
  #평점그림 -> 숫자로 변형
  for (i in 1:5) {
    if (rating[i] == "ui_bubble_rating bubble_50") {
      rating[i] <- gsub("ui_bubble_rating bubble_50",5,rating[i])
    }else if (rating[i] == "ui_bubble_rating bubble_40") {
      rating[i] <- gsub("ui_bubble_rating bubble_40",4,rating[i])
    }else if (rating[i] == "ui_bubble_rating bubble_30") {
      rating[i] <- gsub("ui_bubble_rating bubble_30",3,rating[i])
    }else if (rating[i] == "ui_bubble_rating bubble_20") {
      rating[i] <- gsub("ui_bubble_rating bubble_20",2,rating[i])
    }else {
      rating[i] <- gsub("ui_bubble_rating bubble_10",1,rating[i])
    }
  }

  # 수집된 정보를 데이터로 결합
  df <- data.frame(
    pid = pid,
    hotel = hotel,
    name = name,
    commenttitle = commenttitle,
    comment = comment,
    rating=rating,
    stringsAsFactors = F
  )
  return(df)
}
```

자료수집 - 크롤링

```r
tripadvisor_scraper <- function(pid, url){ # url주소에서 HTML소스 가져오기
  doc <- read_html(url)

  # 호텔명
  hotel <- doc %>%
    html_nodes("#HEADING") %>%
    html_text() %>%
    str_trim()

  # 작성자
  name <- doc %>%
    html_nodes("div.social-member-event-MemberEventOnObjectBlock__event_type--3njyv > span > a") %>%
    html_text() %>%
    str_trim()

  # 리뷰제목
  commenttitle <- doc %>%
    html_nodes("div.location-review-review-list-parts-ReviewTitle__reviewTitle--2GO9Z > a > span > span") %>%
    html_text() %>%
    str_trim()

  #리뷰내용
  comment <- doc %>%
    html_nodes("div.location-review-review-list-parts-ExpandableReview__containerStyles--1G0AE > div._2f_ruteS._1bona3Pu._2uD5bLZZ > div > q") %>%
    html_text() %>%
    str_trim()

  #평점
  rating <- doc %>%
    html_nodes("div.location-review-review-list-parts-SingleReview__mainCol--1hApa > div.location-review-review-list-parts-RatingLine__container--2bjtw > div > span") %>%
    html_attr("class")
  #평점그림 -> 숫자로 변형
  for (i in 1:5) {
    if (rating[i] == "ui_bubble_rating bubble_50") {
      rating[i] <- gsub("ui_bubble_rating bubble_50",5,rating[i])
    }else if (rating[i] == "ui_bubble_rating bubble_40") {
      rating[i] <- gsub("ui_bubble_rating bubble_40",4,rating[i])
    }else if (rating[i] == "ui_bubble_rating bubble_30") {
      rating[i] <- gsub("ui_bubble_rating bubble_30",3,rating[i])
    }else if (rating[i] == "ui_bubble_rating bubble_20") {
      rating[i] <- gsub("ui_bubble_rating bubble_20",2,rating[i])
    }else {
      rating[i] <- gsub("ui_bubble_rating bubble_10",1,rating[i])
    }
  }

  # 수집된 정보를 데이터로 결합
  df <- data.frame(
    pid = pid,
    hotel = hotel,
    name = name,
    commenttitle = commenttitle,
    comment = comment,
    rating=rating,
    stringsAsFactors = F
  )
  return(df)
}
```

자료수집 - 크롤링

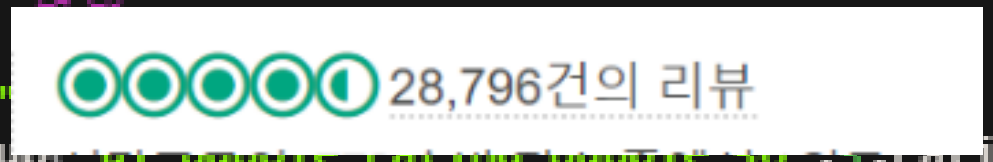
```r
#평점
rating <- doc %>%
  html_nodes("div.location-review-review-list-parts-SingleReview__mainCol--1hApa > div
  html_attr("class")
#평점그림 -> 숫자로 변환
for (i in 1:5) {
  if (rating[i] == "ui_bubble_rating bubble_50") {
    rating[i] <- gsub("ui_bubble_rating bubble_50",5,rating[i])
  }else if (rating[i] == "ui_bubble_rating bubble_40") {
    rating[i] <- gsub("ui_bubble_rating bubble_40",4,rating[i])
  }else if (rating[i] == "ui_bubble_rating bubble_30") {
    rating[i] <- gsub("ui_bubble_rating bubble_30",3,rating[i])
  }else if (rating[i] == "ui_bubble_rating bubble_20") {
    rating[i] <- gsub("ui_bubble_rating bubble_20",2,rating[i])
  }else {
    rating[i] <- gsub("ui_bubble_rating bubble_10",1,rating[i])
  }
}
```

자료수집 - 크롤링

```r
#평점
rating <- doc %>%
  html_nodes("div.location-review-review-list-parts-SingleReview__mainCol--1hApa > div
  html_attr("class")
#평점그림 -> 숫자로 변형
for (i in 1:5) {
  if (rating[i] == "ui_bubble_rating bubble_50") {
    rating[i] <- gsub("ui_bubble_rating bubble_50",5,rating[i])
  }else if (rating[i] == "ui_bubble_rating bubble_40") {
    rating[i] <- gsub("ui_bubble_rating bubble_40",4,rating[i])
  }else if (rating[i] == "
    rating[i] <- gsub("ui_
  }else if (rating[i] == "
    rating[i] <- gsub("ui_
  }else {
    rating[i] <- gsub("ui_
  }
}
```

자료수집 - 크롤링

```
#평점
rating <- doc %>%
  html_nodes("div.location-review-review-list-parts-SingleReview__mainCol--1hApa > div
  html_attr("class")
#평점그림 -> 숫자로 변환
for (i in 1:5) {
  if (rating[i] == "
    rating[i] <- gsub(
  }else if (rating[i] == "ui_bubble_rating bubble_40") {
    rating[i] <- gsub("ui_bubble_rating bubble_40",4,rating[i])
  }else if (rating[i] == "
    rating[i] <- gsub("ui_
  }else if (rating[i] == "
    rating[i] <- gsub("ui_
  }else {
    rating[i] <- gsub("ui_
  }
}
```



자료 분석 - 워드크라우드

```r
df <- read.csv("rdata_YMCA_One_Orchard_d310344.csv",stringsAsFactors = F)

library(tidytext)

ct_df <- data.frame(rid=1:nrow(df),text=df$commenttitle)
c_df <- data.frame(rid=1:nrow(df),text=df$comment)
ctc_df <- rbind(ct_df,c_df)

tidy_ctc <- ctc_df %>% unnest_tokens(word,text,token='ngrams',n=1)

#불용어 사전에 있는 단어 제거
data(stop_words)
tidy_ctc <- tidy_ctc %>% anti_join(stop_words)

#유사 용어 정제 및 삭제
#3.marinabay
tidy_ctc<- tidy_ctc %>% subset(subset=(word!='hotel'&word!='bay'&word!='marina'&word!='singapore'&word!='bay'&word!='sands'))
tidy_ctc$word <- gsub('stayed','stay',tidy_ctc$word)
tidy_ctc$word <- gsub('views','view',tidy_ctc$word)

#4.raffles
tidy_ctc<- tidy_ctc %>% subset(subset=(word!='hotel'&word!='bay'&word!='raffles'&word!='singapore'))
tidy_ctc$word <- gsub('stayed','stay',tidy_ctc$word)

#6.podcapsule
tidy_ctc<- tidy_ctc %>% subset(subset=(word!='hotel'&word!='pod'&word!='singapore'))
tidy_ctc$word <- gsub('stayed','stay',tidy_ctc$word)

#7.YMCAoneorchard
tidy_ctc<- tidy_ctc %>% subset(subset=(word!='hotel'&word!='ymca'&word!='singapore'&word!='orchard'))
tidy_ctc$word <- gsub('stayed','stay',tidy_ctc$word)
tidy_ctc$word <- gsub('mrt','station',tidy_ctc$word)
```

자료 분석 - 워드크라우드

전처리

```
df <- read.csv("rdata_YMCA_One_Orchard_d310344.csv",stringsAsFactors = F)

library(tidytext)

ct_df <- data.frame(rid=1:nrow(df),text=df$commenttitle)
c_df <- data.frame(rid=1:nrow(df),text=df$comment)
ctc_df <- rbind(ct_df,c_df)

tidy_ctc <- ctc_df %>% unnest_tokens(word,text,token='ngrams',n=1)

#불용어 사전에 있는 단어 제거
data(stop_words)
tidy_ctc <- tidy_ctc %>% anti_join(stop_words)

#유사 용어 정제 및 삭제
#3.marinabay
tidy_ctc<- tidy_ctc %>% subset(subset=(word!='hotel'&word!='bay'&word!='marina'&word!='singapore'&word!='bay'&word!='sands'))
tidy_ctc$word <- gsub('stayed','stay',tidy_ctc$word)
tidy_ctc$word <- gsub('views','view',tidy_ctc$word)

#4.raffles
tidy_ctc<- tidy_ctc %>% subset(subset=(word!='hotel'&word!='bay'&word!='raffles'&word!='singapore'))
tidy_ctc$word <- gsub('stayed','stay',tidy_ctc$word)

#6.podcapsule
tidy_ctc<- tidy_ctc %>% subset(subset=(word!='hotel'&word!='pod'&word!='singapore'))
tidy_ctc$word <- gsub('stayed','stay',tidy_ctc$word)

#7.YMCAoneorchard
tidy_ctc<- tidy_ctc %>% subset(subset=(word!='hotel'&word!='ymca'&word!='singapore'&word!='orchard'))
tidy_ctc$word <- gsub('stayed','stay',tidy_ctc$word)
tidy_ctc$word <- gsub('mrt','station',tidy_ctc$word)
```

자료 분석 - 워드크라우드

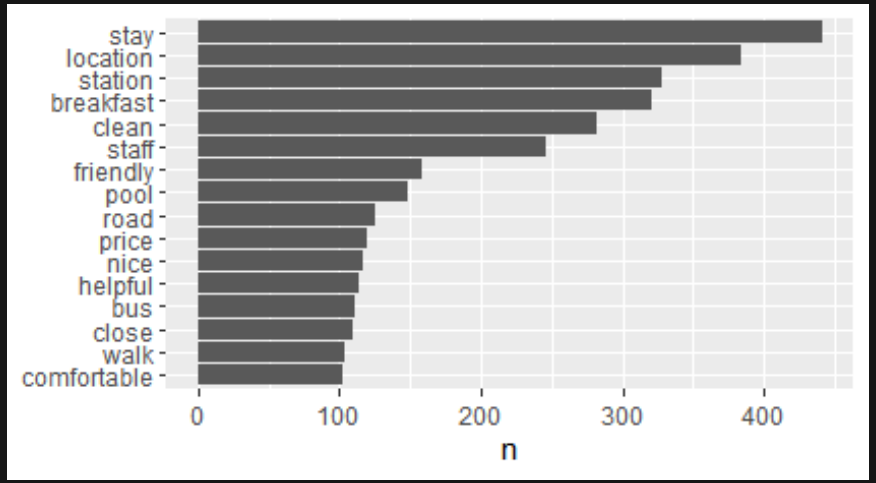
```
#단어 빈도 분석
library(dplyr)
library(ggplot2)
tidy_ctc %>%
  count(word,sort=TRUE) %>%
  filter(n>100) %>%
  mutate(word=reorder(word,n)) %>%
  ggplot(aes(word,n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

#워드클라우드 분석
library(wordcloud)
library(wordcloud2)

#version1
tidy_ctc %>%
  anti_join(stop_words) %>%
  count(word) %>%
  with(wordcloud(word,n,max.words=100))

#version2
tidy_ctc %>%
  anti_join(stop_words) %>%
  count(word,sort = TRUE)%>%
  head(100) %>%
  wordcloud2()

#version3
tidy_ctc %>%
  anti_join(stop_words) %>%
  count(word,sort = TRUE)%>%
  head(100) %>%
  wordcloud2(size = 1.2, color = "random-light", fontFamily = 'Tahoma', minRotation = -pi/2, maxRotation = -pi/2)

#긍정/부정 비교 워드클라우드
library(reshape2)

tidy_ctc %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment,sort=TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill=0) %>%
  comparison.cloud(colors = c("orange", "navy"), max.words = 100)
```

자료 분석 - 워드크라우드

```r
#단어 빈도 분석
library(dplyr)
library(ggplot2)
tidy_ctc %>%
  count(word,sort=TRUE) %>%
  filter(n>100) %>%
  mutate(word=reorder(word,n)) %>%
  ggplot(aes(word,n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

#워드클라우드 분석
library(wordcloud)
library(wordcloud2)

#version1
tidy_ctc %>%
  anti_join(stop_words) %>%
  count(word) %>%
  with(wordcloud(word,n,max.words=100))

#version2
tidy_ctc %>%
  anti_join(stop_words) %>%
  count(word,sort = TRUE)%>%
  head(100) %>%
  wordcloud2()

#version3
tidy_ctc %>%
  anti_join(stop_words) %>%
  count(word,sort = TRUE)%>%
  head(100) %>%
  wordcloud2(size = 1.2, color = "random-light", fontFamily = 'Tahoma', minRotation = -pi/2, maxRotation = -pi/2)

#긍정/부정 비교 워드클라우드
library(reshape2)

tidy_ctc %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment,sort=TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill=0) %>%
  comparison.cloud(colors = c("orange", "navy"), max.words = 100)
```

자료 분석 - 워드크라우드

```
#단어 빈도 분석
library(dplyr)
library(ggplot2)
tidy_ctc %>%
  count(word,sort=TRUE) %>%
  filter(n>100) %>%
  mutate(word=reorder(word,n)) %>%
  ggplot(aes(word,n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()
```



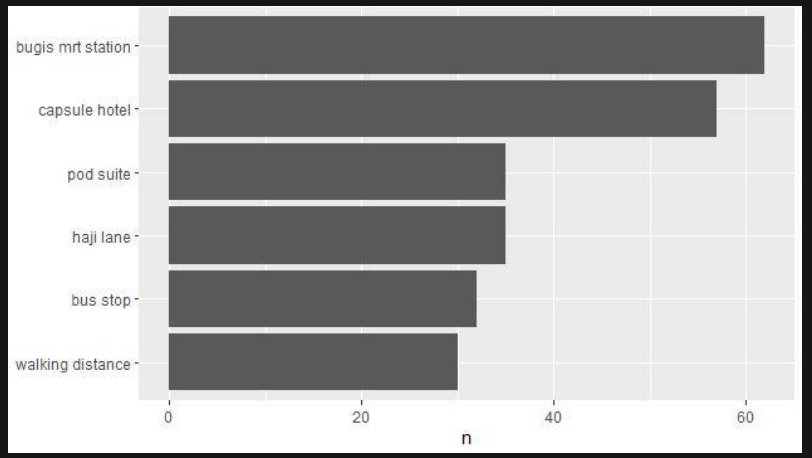

```
#version2
tidy_ctc %>%
  anti_join(stop_words) %>%
  count(word,sort = TRUE)%>%
  head(100) %>%
  wordcloud2()
```

데이터 전처리 – 토픽모델링

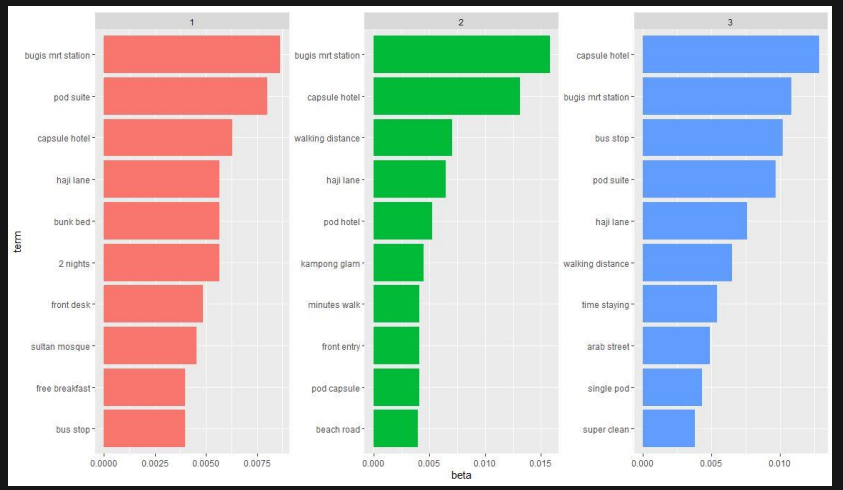
```
text_df <- tibble(rid=1:nrow(df),text=df$comment)
tidy_bigram <- text_df %>% unnest_tokens(bigram,text,token = "ngrams", n = 2)
text_df <- tibble(rid=1:nrow(df),text=df$comment)
tidy_bigram <- text_df %>% unnest_tokens(bigram,text,token = "ngrams", n = 2)

tidy_bigram %>%count(bigram,sort=TRUE)

library(tidyr)
bigrams_separated <- tidy_bigram %>%
  separate(bigram, c("word1", "word2"), sep = " ")

bigrams_filtered <- bigrams_separated %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)
car<-data.frame(c('marina bay','bay sands'))
names(car)<-c("stop")


bigram_counts <- bigrams_filtered %>%
  count(word1, word2, sort = TRUE)


bigrams_united <- bigrams_filtered %>%
  unite(bigram, word1, word2, sep = " ")
bigrams_united <-bigrams_united %>%
  filter(!bigram %in% car$stop)

bigrams_united %>%
  count(bigram,sort=TRUE) %>%
  filter(n > 20) %>%
  mutate(bigram=reorder(bigram,n)) %>%
  ggplot(aes(bigram,n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()



word_counts <- bigrams_united %>%
  count(rid,bigram, sort=TRUE)


library(topicmodels)

#DocumentTermMatrix
dtm <- word_counts %>%
  cast_dtm(rid,bigram, n)
dtm

lda <- LDA(dtm, k =3, control = list(seed = 1234))
lda

topics <- tidy(lda, matrix = "beta")
topics

top_terms <- topics %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

top_terms %>%
  mutate(term = reorder_within(term, beta, topic)) %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip() +
  scale_x_reordered()
```

데이터 전처리 – 토픽모델링

```r
# 기여도
contributions <- tidy_text %>%
  inner_join(afinn1, by = "word") %>%
  group_by(word) %>%
  summarize(빈도 = n(),기여도 = sum(value))

contributions %>%
  top_n(30, abs(기여도)) %>%
  mutate(word = reorder(word, 기여도)) %>%
  ggplot(aes(word, 기여도, fill = 기여도 > 0)) +
  geom_col(show.legend = FALSE) +
  coord_flip()
```




```r
# 빈도
bing_word_counts <- tidy_text %>%
  inner_join(bing1) %>% #bing 감성사전 활용
  count(word, sentiment, sort = TRUE) %>%
  ungroup()


bing_word_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Contribution to sentiment", x = NULL) +
  coord_flip()
```



자료 분석 – 감성 분석

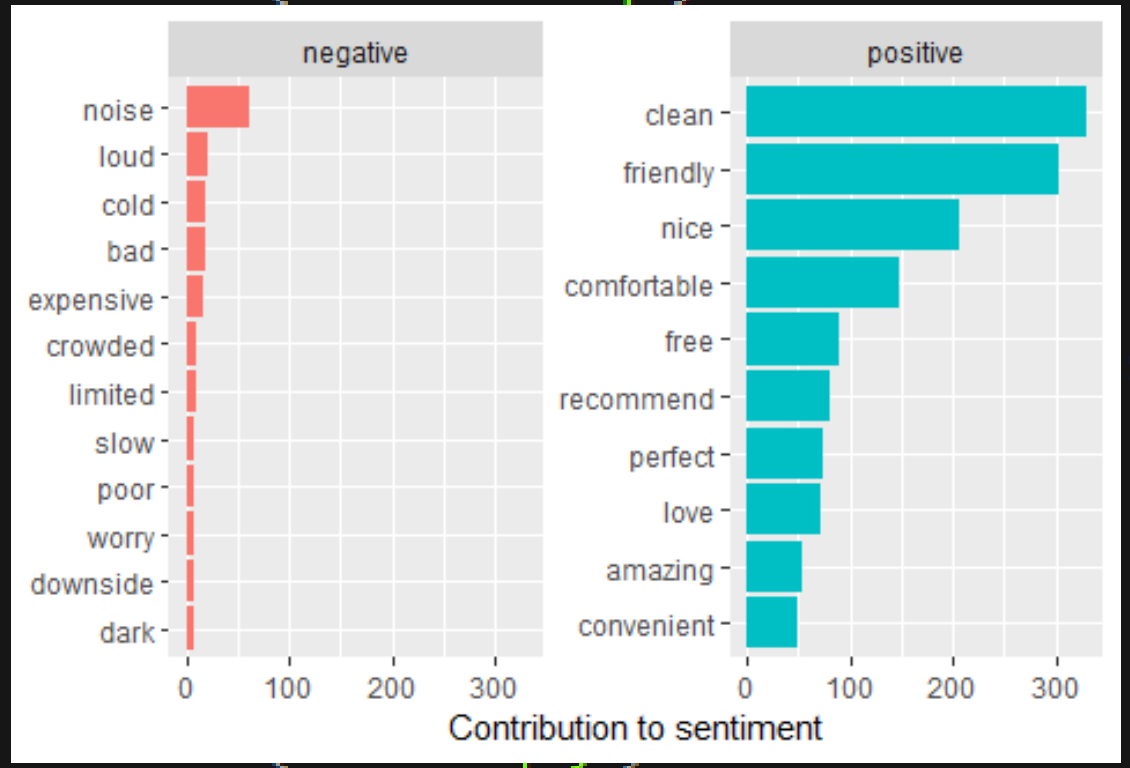
```
df_TF <- read.csv("rdata_YMCA_One_Orchard_d310344 _TF-IDF.csv", stringsAsFactors = FALSE)
df <- read.csv("rdata_YMCA_One_Orchard_d310344.csv", stringsAsFactors = FALSE)
nrow(df_TF)

bing <- get_sentiments("bing")
afinn <- get_sentiments("afinn")
nrc <- get_sentiments("nrc")

text_tdf <- data.frame(rid=1:nrow(df), text=df$commenttitle)
text_cdf <- data.frame(rid=1:nrow(df), text=df$comment)
text_df<-rbind(text_tdf, text_cdf)
text_df <- text_df %>% unnest_tokens(word,text,token = "ngrams", n = 1)
text_df <- text_df %>% anti_join(stop_words)

word_TF_IDF <- merge(x=text_df, y=df_TF, by = 'word')
word_TF_IDF <- word_TF_IDF[with(word_TF_IDF, order(desc(TF_IDF))), ]


data(stop_words) #불용어(stopword) 사전 가져오기
tidy_text <- word_TF_IDF %>% anti_join(stop_words)

bing1<-subset(bing, subset=(word=="lobby"&word=="huge"&word=="top"))
bing %>% subset(word != c('lobby','huge','top'))

bing1 <- bing %>% filter(word != "top") %>%
  filter(word != "hard") %>%
  filter(word != "cheap") %>%
  filter(word != "miss") %>%
  filter(word != "issues") %>%
  filter(word != "issue") %>%
  filter(word != "bunk") %>%
  filter(word != "sink") %>%
  filter(word != "cons") %>%
  filter(word != "blind")
afinn1 <- afinn %>% filter(word != "top") %>%
  filter(word != "lobby") %>%
  filter(word != "huge") %>%
  filter(word != "stop") %>%
  filter(word != "care") %>%
  filter(word != "chance") %>%
  filter(word != "greeted") %>%
  filter(word != "welcomed") %>%
  filter(word != "pay") %>%
  filter(word != "shared") %>%
  filter(word != "hard") %>%
  filter(word != "share")

tidy_text$word <- gsub("loves|loved|loving|lovely","love",tidy_text$word)
tidy_text$word <- gsub("amazing|awesome","amazing",tidy_text$word)
tidy_text$word <- gsub("enjoy|fun","enjoy",tidy_text$word)
tidy_text$word <- gsub("excellent|perfect","perfect",tidy_text$word)
tidy_text$word <- gsub("superb|super","superb",tidy_text$word)
tidy_text$word <- gsub("impressive|impressed","impressive",tidy_text$word)
tidy_text$word <- gsub("attraction|attractions","attraction",tidy_text$word)
tidy_text$word <- gsub("disappointing|disappointed","disappointing",tidy_text$word)
tidy_text$word <- gsub("crowded|complex","crowded",tidy_text$word)
tidy_text$word <- gsub("pricey|expensive","expensive",tidy_text$word)
tidy_text$word <- gsub("friendly|helpful|helping","friendly",tidy_text$word)
tidy_text$word <- gsub("recommend|recommended","recommend",tidy_text$word)
tidy_text$word <- gsub("happy|glad","happy",tidy_text$word)
tidy_text$word <- gsub("pleasant|pleasure","pleasure",tidy_text$word)
tidy_text$word <- gsub("lost|missed","lost",tidy_text$word)
tidy_text$word <- gsub("comfort|comfortable","comfortable",tidy_text$word)
tidy_text$word <- gsub("secure|safe","safe",tidy_text$word)
tidy_text$word <- gsub("secure|safe","safe",tidy_text$word)
tidy_text$word <- gsub("noise|noisy","noise",tidy_text$word)
tidy_text$word <- gsub("improve|improvement","improvement",tidy_text$word)
```

자료 분석 – 감성 분석

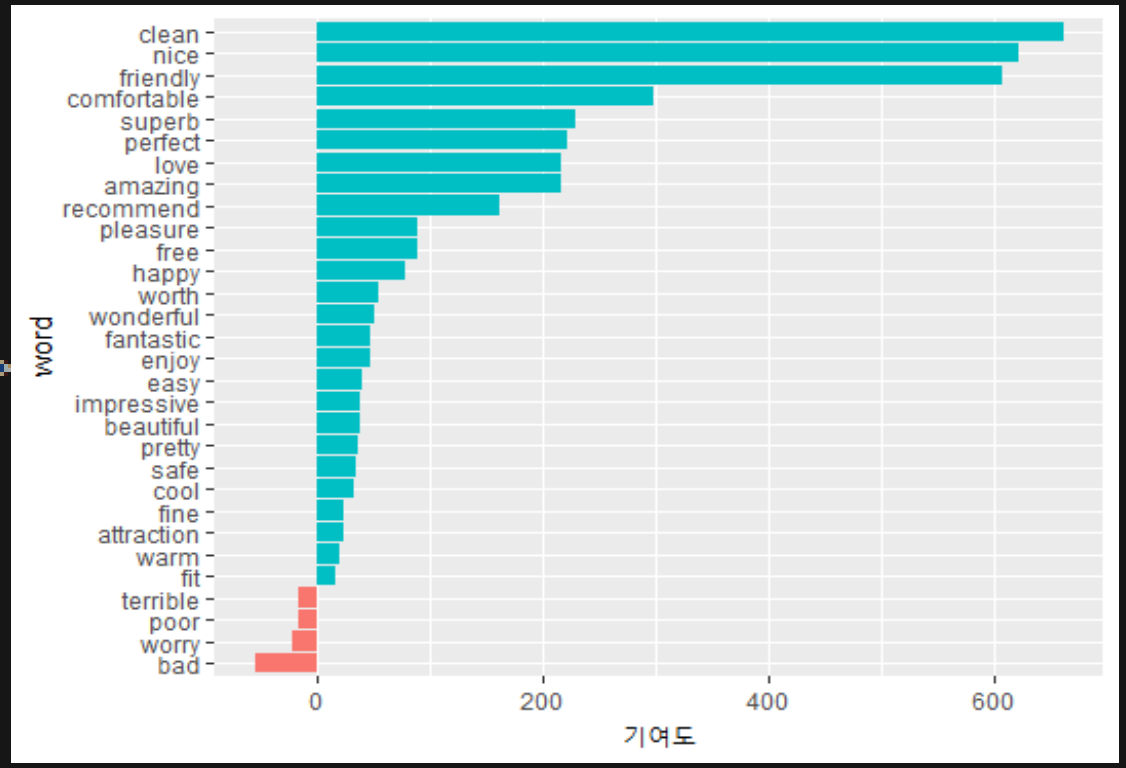
```
bing1 <- bing %>% filter(word != "top") %>%
  filter(word != "hard") %>%
  filter(word != "cheap") %>%
  filter(word != "miss") %>%
  filter(word != "issues") %>%
  filter(word != "issue") %>%
  filter(word != "bunk") %>%
  filter(word != "sink") %>%
  filter(word != "cons") %>%
  filter(word != "blind")
afinn1 <- afinn %>% filter(word != "top") %>%
  filter(word != "lobby") %>%
  filter(word != "huge") %>%
  filter(word != "stop") %>%
  filter(word != "care") %>%
  filter(word != "chance") %>%
  filter(word != "greeted") %>%
  filter(word != "welcomed") %>%
  filter(word != "pay") %>%
  filter(word != "shared") %>%
  filter(word != "hard") %>%
  filter(word != "share")
```

자료 분석 – 감성 분석

```
bing1 <- bing %>% filter(word != "top") %>%
  filter(word != "hard") %>%
  filter(word != "cheap") %>%
```



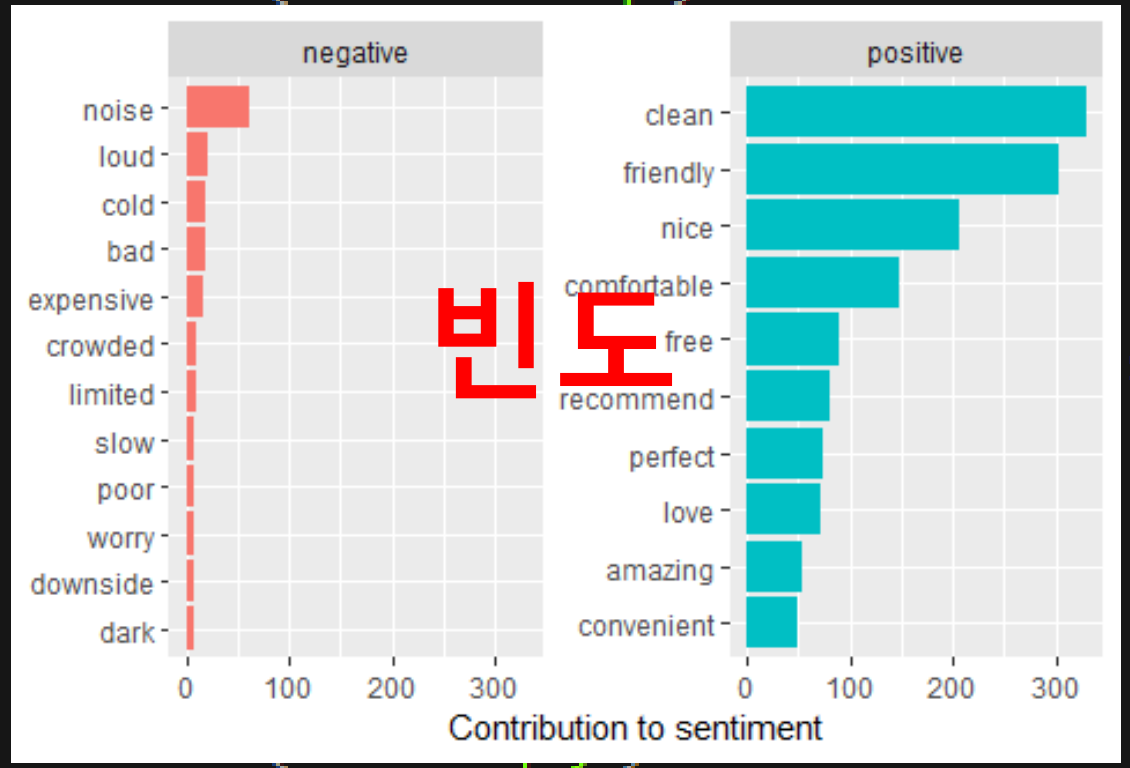
%>%



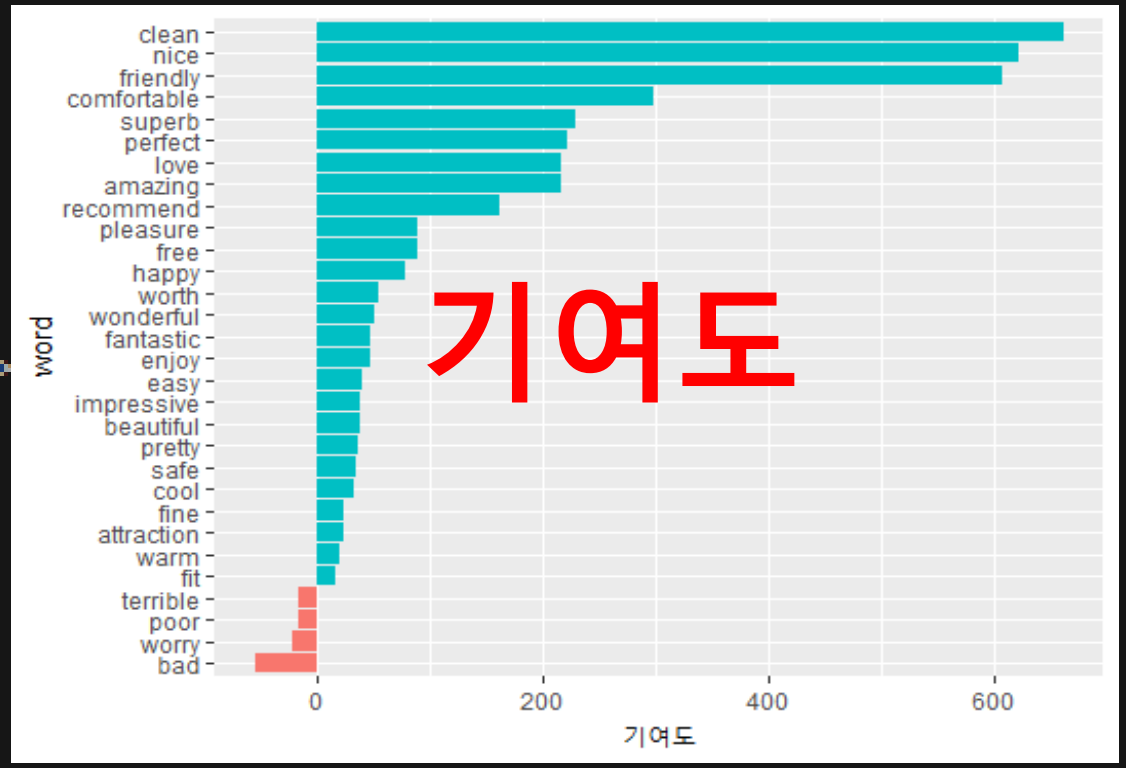

```
  filter(word != "shared") %>%
  filter(word != "hard") %>%
  filter(word != "share")
```

자료 분석 – 감성 분석

```
bing1 <- bing %>% filter(word != "top") %>%
  filter(word != "hard") %>%
  filter(word != "cheap") %>%
```

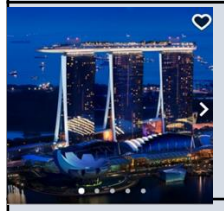
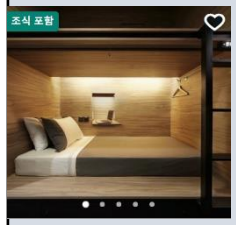
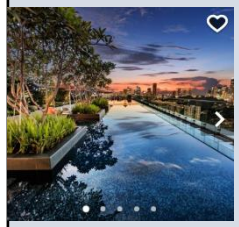


%>%



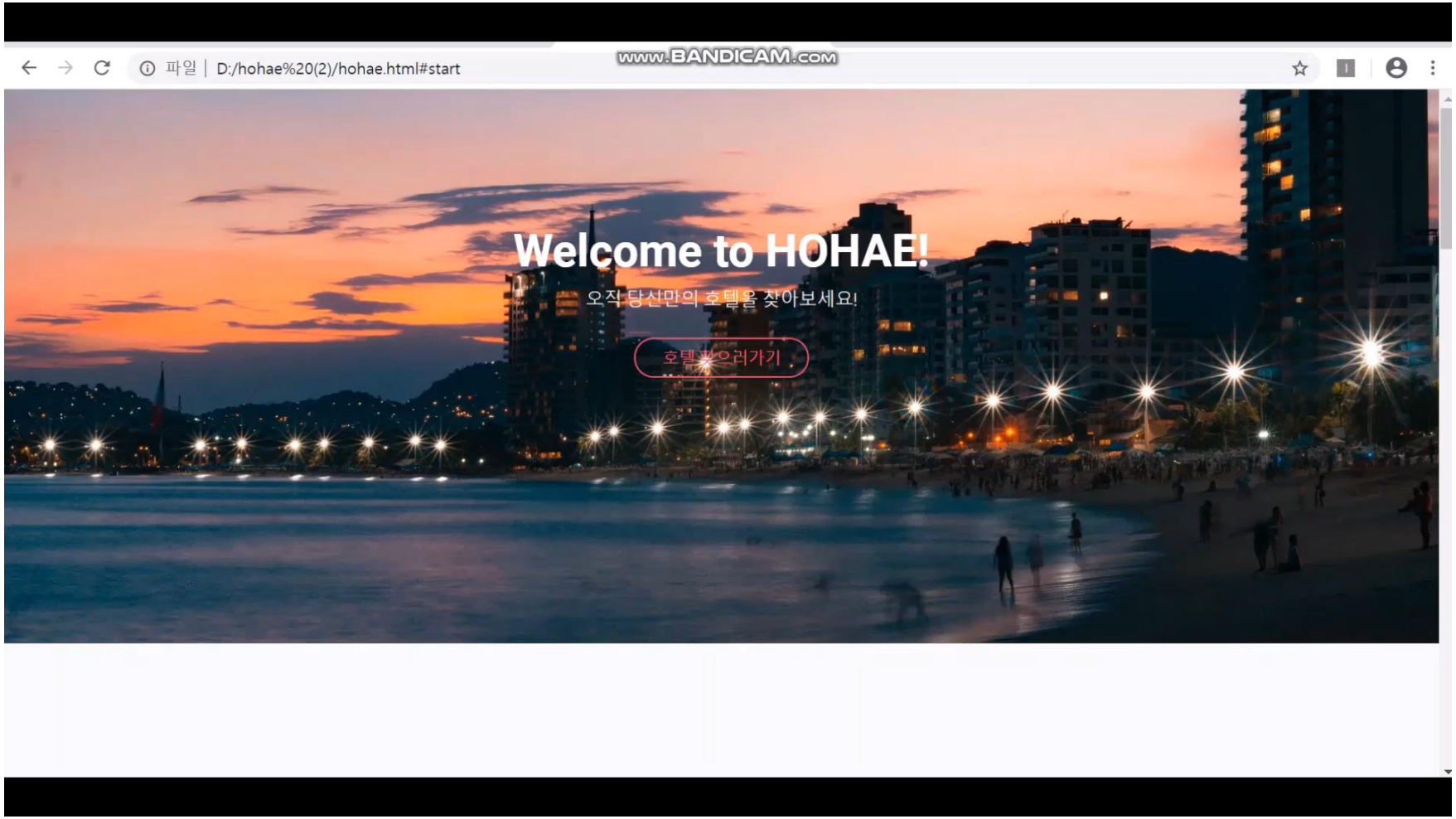

```
  filter(word != "shared") %>%
  filter(word != "hard") %>%
  filter(word != "share")
```

자료 분석 – 종합 분석

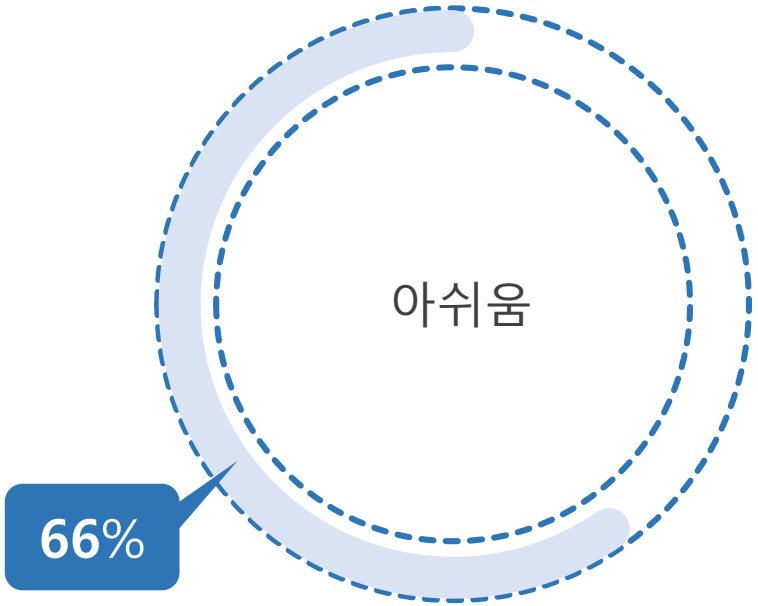
호텔 사진	키워드
 마리나 베이 샌즈	루프탑(정원, 수영장, 전망), 5성급, 전망, 시설, 기본 옵션 청결 낮음, 비협조적, 쓸데없는 비용지불
 레플스	유명, 특이, 우아하고 정직, 접근성 안 좋음, 고장 잘 남. 땅콩 껍질-> 땅콩 껍질을 바닥에 던져 놓고 걷는게 전통, 특이, 우아하고 정직
 더 포드	캡슐 호텔, 접근성(관광지(haji lane), 교통) 좋음, 무료조식, 청결도 좋음, 시설 편리, 일처리 느리고 한정적이고 어두움
 호텔 젠	접근성 좋음, 서비스 좋음, 시설 편리(수영장), 조식 품질 좋음, 가격 낮음, 시설이 낡은, 한정적, 더러움

분석 데이터 실전 활용

hohae.html



한계점 및 향후 연구 방향



키워드 세분화

- 땅콩 뿌리기, 인피니티 풀, 특색있는 술을 파는 bar 등등

세분화 능력 30%

데이터 수집 능력 50%

관점 45%

다양한 관점

– 여러 방면의 고차원적인 전공 지식 필요

데이터 표본 크기

- 보다 많은 데이터 필요

Thank U,
For watching
this presentation.