Airline News Articles Sentiment Analysis

GENEVIEVE GONZALES

2025-03-10

```
library(tidytext)
library(readtext)
library(quanteda)
library(syuzhet)
library(dplyr)
library(ggplot2)
```

Asiana Airlines

```
# ASIANA

# file paths for the three articles
file_paths <- c("/Users/genevievegonzales/Desktop/DH/asiana/abc.txt",
                "/Users/genevievegonzales/Desktop/DH/asiana/bbc.txt",
                "/Users/genevievegonzales/Desktop/DH/asiana/cnn.txt")

# read all articles into a df
asiana_articles <- readtext(file_paths)

# clean text
asiana_articles$text <- gsub("[\"""]", "", asiana_articles$text)
asiana_articles$text <- gsub("[[:punct:]]", "", asiana_articles$text)
asiana_articles$text <- tolower(asiana_articles$text)

# covert text into tidy
tidy_asiana <- tibble(text = asiana_articles$text) %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words)

# syuzhet sentiment score for overall tone
syuzhet_scores <- get_sentiment(tidy_asiana, method = "syuzhet")

# AFINN sentiment score (ranges from -5 to +5 per word)
afinn_scores <- get_sentiment(tidy_asiana, method = "afinn")


sentiment_results <- tibble(
  Syuzhet_Score = syuzhet_scores,
  AFINN_Score = afinn_scores
```

```
)

write.csv(sentiment_results, "asiana_sentiment_results.csv", row.names = FALSE)


# ======================================================

bing_sentiments <- get_sentiments("bing")  # bing lexicon (Positive/Negative)
afinn_sentiments <- get_sentiments("afinn")  # AFINN lexicon (Numerical Score)

nrc_sentiments <- get_sentiments("nrc")  # NRC lexicon (Emotion Categories)
fear_words <- tidy_asiana %>%
  inner_join(filter(nrc_sentiments, sentiment == "fear")) %>%
  count(word, sort = TRUE)

# table of words and their sentiment values
word_sentiment_table <- tidy_asiana %>%
  inner_join(afinn_sentiments, by = "word") %>%
  count(word, wt = value, sort = TRUE) %>%
  rename(Total_Sentiment_Score = n)

write.csv(word_sentiment_table, "asiana_word_sentiment_table.csv", row.names = FALSE)


# ======================================================

# filter top 15 most frequent fear words
top_fear_words <- fear_words %>% slice_max(n, n = 9)

# remove "shanghai" and "descent" from fear words
top_fear_words_filtered <- top_fear_words %>%
  filter(!word %in% c("flying","shanghai"))

write.csv(top_fear_words_filtered, "asiana_filtered_fear_words.csv", row.names = FALSE)


# ======================================================

asiana_nrc <- tidy_asiana %>%
  inner_join(nrc_sentiments, by = "word") %>%
  count(sentiment, sort = TRUE)

write.csv(asiana_nrc, "asiana_nrc_sentiment.csv", row.names = FALSE)


# ======================================================

positive_words_count <- tidy_asiana %>%
  inner_join(afinn_sentiments, by = "word") %>%
  filter(value > 0) %>%
  nrow()

# count negative words
```

```r
negative_words_count <- tidy_asiana %>%
  inner_join(afinn_sentiments, by = "word") %>%
  filter(value < 0) %>%
  nrow()

# create df for the pie chart
pie_graph <- data.frame(
  Sentiment = c("Positive", "Negative"),
  Proportion = c(positive_words_count, negative_words_count)
)

write.csv(pie_graph, "asiana_sentiment_pie_data.csv", row.names = FALSE)
```

Comair

```r
# COMAIR

# file paths for the three articles
file_paths <- c("/Users/genevievegonzales/Desktop/DH/comair/abc.txt",
                "/Users/genevievegonzales/Desktop/DH/comair/nbc.txt",
                "/Users/genevievegonzales/Desktop/DH/comair/guardian.txt")

# read all articles into a df
comair_articles <- readtext(file_paths)

# clean text
comair_articles$text <- gsub("[\"""]", "", comair_articles$text)
comair_articles$text <- gsub("[[:punct:]]", "", comair_articles$text)
comair_articles$text <- tolower(comair_articles$text)

# convert text into tidy
tidy_comair <- tibble(text = comair_articles$text) %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words)

# syuzhet sentiment score for overall tone
syuzhet_scores <- get_sentiment(tidy_comair, method = "syuzhet")

# AFINN sentiment score (ranges from -5 to +5 per word)
afinn_scores <- get_sentiment(tidy_comair, method = "afinn")

sentiment_results <- tibble(
  Syuzhet_Score = syuzhet_scores,
  AFINN_Score = afinn_scores
)

write.csv(sentiment_results, "comair_sentiment_results.csv", row.names = FALSE)


# ====================================================
```

```r
bing_sentiments <- get_sentiments("bing")  # bing lexicon (Positive/Negative)
afinn_sentiments <- get_sentiments("afinn")  # AFINN lexicon (Numerical Score)

nrc_sentiments <- get_sentiments("nrc")  # NRC lexicon (Emotion Categories)
fear_words <- tidy_comair %>%
  inner_join(filter(nrc_sentiments, sentiment == "fear")) %>%
  count(word, sort = TRUE)


# Create a table of words and their sentiment values
word_sentiment_table <- tidy_comair %>%
  inner_join(afinn_sentiments, by = "word") %>%
  count(word, wt = value, sort = TRUE) %>%
  rename(Total_Sentiment_Score = n)

write.csv(word_sentiment_table, "comair_word_sentiment_table.csv", row.names = FALSE)


# ============================================================

# filter top 15 most frequent fear words
top_fear_words <- fear_words %>% slice_max(n, n = 10)

write.csv(top_fear_words, "comair_top_fear_words.csv", row.names = FALSE)


# ============================================================

comair_nrc <- tidy_comair %>%
  inner_join(nrc_sentiments, by = "word") %>%
  count(sentiment, sort = TRUE)

write.csv(comair_nrc, "comair_nrc_sentiment.csv", row.names = FALSE)


# ============================================================

positive_words_count <- tidy_comair %>%
  inner_join(afinn_sentiments, by = "word") %>%
  filter(value > 0) %>%
  nrow()

# count negative words
negative_words_count <- tidy_comair %>%
  inner_join(afinn_sentiments, by = "word") %>%
  filter(value < 0) %>%
  nrow()

# create df for the pie chart
pie_graph <- data.frame(
  Sentiment = c("Positive", "Negative"),
  Proportion = c(positive_words_count, negative_words_count)
)
```

```r
write.csv(pie_graph, "comair_sentiment_pie_data.csv", row.names = FALSE)
```

Alaska Airlines

```r
# ALASKA AIRLINES

# file paths for the three articles
file_paths <- c("/Users/genevievegonzales/Desktop/DH/alaska/abc.txt",
                "/Users/genevievegonzales/Desktop/DH/alaska/time.txt",
                "/Users/genevievegonzales/Desktop/DH/alaska/cnn.txt")

# read all articles into a df
alaska_articles <- readtext(file_paths)

# clean text
alaska_articles$text <- gsub("[\"""]", "", alaska_articles$text)
alaska_articles$text <- gsub("[[:punct:]]", "", alaska_articles$text)
alaska_articles$text <- tolower(alaska_articles$text)

# convert text into tidy
tidy_alaska <- tibble(text = alaska_articles$text) %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words)

# syuzhet sentiment score for overall tone
syuzhet_scores <- get_sentiment(tidy_alaska, method = "syuzhet")

# AFINN sentiment score (ranges from -5 to +5 per word)
afinn_scores <- get_sentiment(tidy_alaska, method = "afinn")

sentiment_results <- tibble(
  Syuzhet_Score = syuzhet_scores,
  AFINN_Score = afinn_scores
)

write.csv(sentiment_results, "alaska_sentiment_results.csv", row.names = FALSE)


# ========================================================

bing_sentiments <- get_sentiments("bing")  # bing lexicon (Positive/Negative)
afinn_sentiments <- get_sentiments("afinn")  # AFINN lexicon (Numerical Score)

nrc_sentiments <- get_sentiments("nrc")  # NRC lexicon (Emotion Categories)
fear_words <- tidy_alaska %>%
  inner_join(filter(nrc_sentiments, sentiment == "fear")) %>%
  count(word, sort = TRUE)


# table of words and their sentiment values
```

```r
word_sentiment_table <- tidy_alaska %>%
  inner_join(afinn_sentiments, by = "word") %>%
  count(word, wt = value, sort = TRUE) %>%
  rename(Total_Sentiment_Score = n)

write.csv(word_sentiment_table, "alaska_word_sentiment_table.csv", row.names = FALSE)


# ======================================================

# filter top 15 most frequent fear words
top_fear_words <- fear_words %>% slice_max(n, n = 10)

write.csv(top_fear_words, "alaska_top_fear_words.csv", row.names = FALSE)

# ======================================================
alaska_nrc <- tidy_alaska %>%
  inner_join(nrc_sentiments, by = "word") %>%
  count(sentiment, sort = TRUE)

write.csv(alaska_nrc, "alaska_nrc_sentiment.csv", row.names = FALSE)


# ======================================================

positive_words_count <- tidy_alaska %>%
  inner_join(afinn_sentiments, by = "word") %>%
  filter(value > 0) %>%
  nrow()

# count negative words
negative_words_count <- tidy_alaska %>%
  inner_join(afinn_sentiments, by = "word") %>%
  filter(value < 0) %>%
  nrow()

# create df for the pie chart
pie_graph <- data.frame(
  Sentiment = c("Positive", "Negative"),
  Proportion = c(positive_words_count, negative_words_count)
)

write.csv(pie_graph, "alaska_sentiment_pie_data.csv", row.names = FALSE)
```