Digital Marketing Analytics: Social Care Analytics at Delta Airlines

Adil Wahab

Project Description

The data "Delta_social_media" includes a subsample of real Twitter data relating to Delta's social care activity. We see both twitter user mentions of "@Delta" as well as @Delta's replies to users. A brief description of the individual datasets follows.

- mentions table: Twitter user mentions of @Delta Each row is an individual tweet. The data excludes retweets and filters out some news mentions. Key variables (not exhaustive) include:

1. Original tweet variables: user_id, created_at, screen_name, text, favorite_count, retweet_count
2. delta_responded indicator: TRUE if @Delta responds to that tweet
3. Delta response variables: delta_reply_text, delta_reply_created_at, delta_reply_favorite_count Note: These variables relate to @Delta's initial response tweet and ignore any of @Delta's subsequent responses to that tweet

- replies table: @Delta's replies Each row is an individual tweet. The data only includes replies: it excludes @Delta's regular tweets and retweets. Key variables (not exhaustive) now include only the original tweet variables: user_id (all from @Delta), created_at, screen_name, text, favorite_count, retweet_count, etc.

EDA

```
dim(mentions)
```

```
## [1] 3202   96
```

```
dim(replies)
```

```
## [1] 2204   90
```

1. The scale of Delta's social care activities

a. Average number of daily replies

```
replies %>% mutate(day = day(created_at)) %>%
  group_by(day) %>% count() -> a1
a1[nrow(a1)+1,] = c("Average", round(colMeans(a1[,2])))
a1
```

	day <chr>	n <chr>
1	1	222
2	2	284
3	3	314
4	4	266
5	5	294
6	30	439
7	31	385
8	Average	315

8 rows

The average number of daily replies is about 315.

b. Average number of daily mentions

```
mentions %>% mutate(day = day(created_at)) %>%
  group_by(day) %>% count() -> b1
b1[nrow(b1)+1,] = c("Average", round(colMeans(b1[,2])))
b1
```

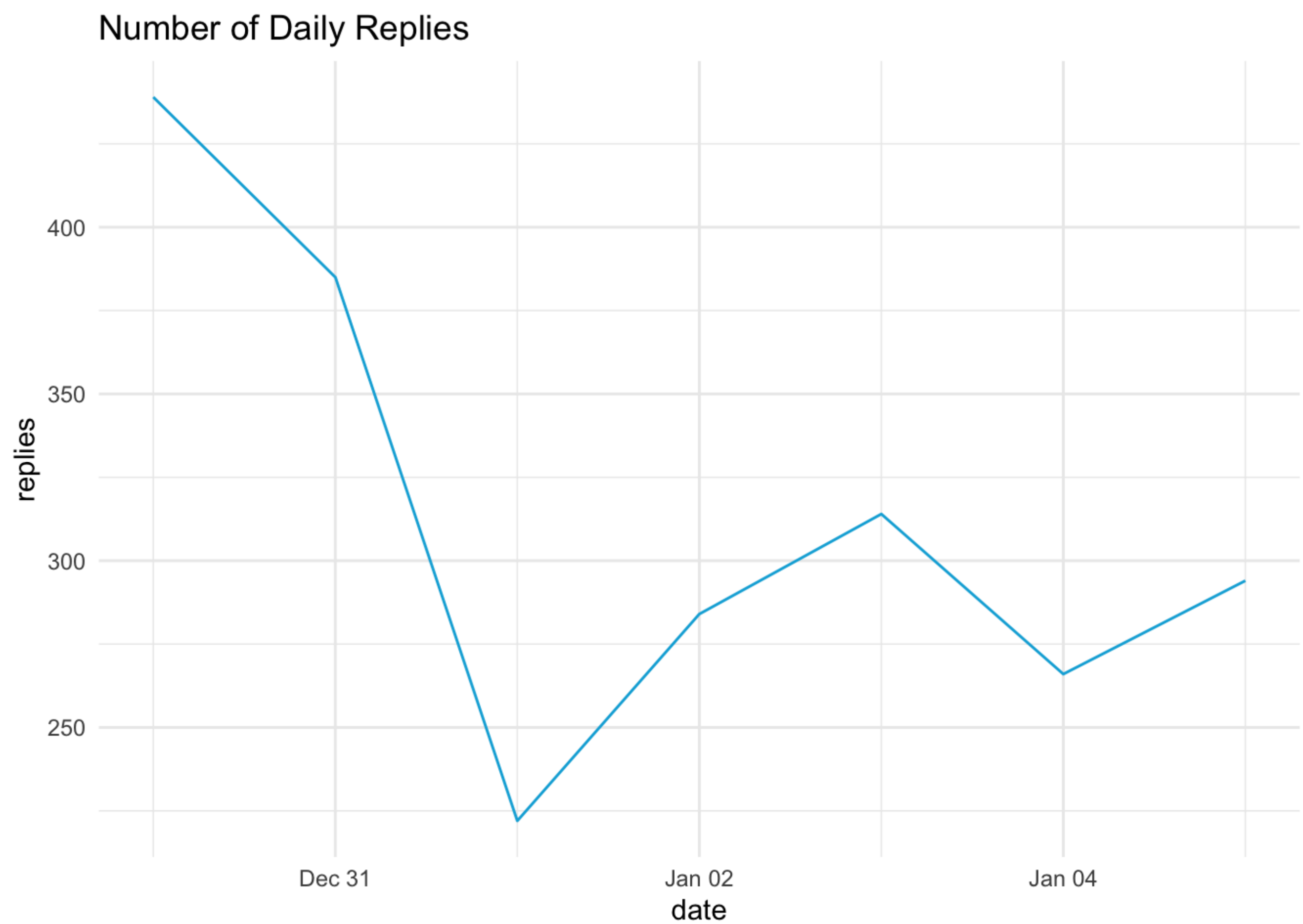
	day <chr>	n <chr>
1	1	343
2	2	441
3	3	478
4	4	353
5	5	503
6	30	608
7	31	476
8	Average	457

8 rows

The average number of daily mentions is about 457.

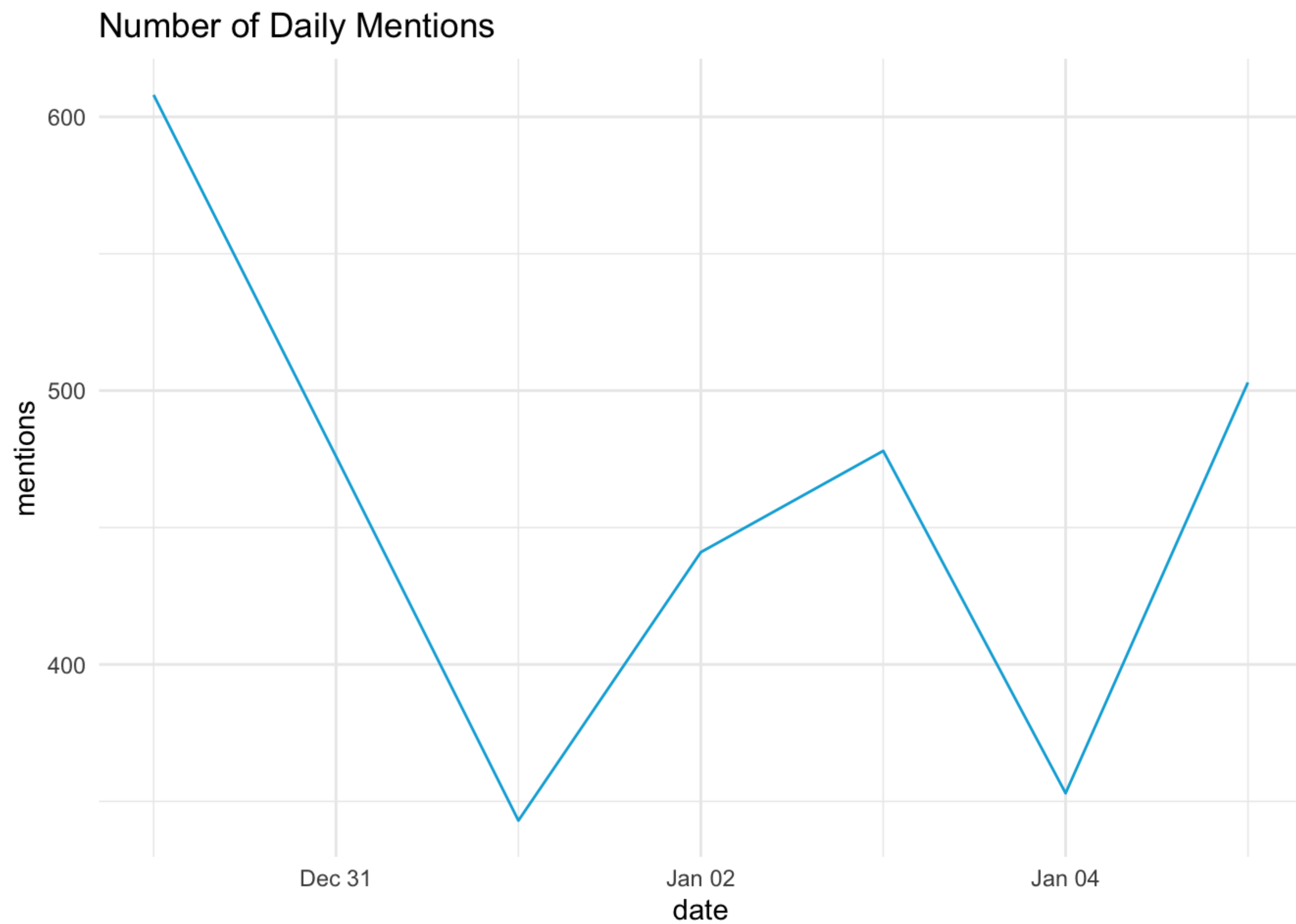
c. Plot of the number of replies by day

```
replies %>%
  mutate(date = as.Date(replies$created_at)) %>%
  group_by(date) %>%
  summarize(obs = n()) -> c1

ggplot(data=c1) + geom_line(aes(x=date,y=obs),color="#00aedb") +
  labs(title="Number of Daily Replies",y="replies") + theme_minimal()
```



c. Plot of the number of mentions by day

```
mentions %>%
  mutate(date = as.Date(mentions$created_at)) %>%
  group_by(date) %>%
  summarize(obs = n()) -> d1

ggplot(data=d1) + geom_line(aes(x=date,y=obs),color="#00aedb") +
  labs(title="Number of Daily Mentions",y="mentions") + theme_minimal()
```



2. Exploration of the mentions data

a. User's number of followers

i. Median number of followers by unique user

```
mentions %>% distinct(user_id, followers_count) %>%
  arrange(desc(followers_count)) -> median_followers

median(median_followers$followers_count)
```

```
## [1] 325
```

The median number of followers in Delta's mentions by unique user is 325.

ii. Screen name of the user with the #3 most followers

```
mentions %>% group_by(user_id, screen_name) %>%
  summarise(followers = sum(followers_count)) %>%
  arrange(desc(followers)) -> a2ii
head(a2ii,3)
```

user_id <chr>	screen_name <chr>	followers <int>
20264905	NYRangers	4368072
4119741	jdickerson	4014244
4170491	ajc	1045514

3 rows

Screen name “ajc” has the third most followers with 1,045,514 followers.

b. Engagement of the mention tweets

i. Average & maximum number of favorites

```
mentions %>% group_by(status_id) %>%
  summarise(favorites = sum(favorite_count))%>%
  arrange(desc(favorites)) -> b2i
print(paste0("Max: ",max(b2i$favorites)))
```

```
## [1] "Max: 994"
```

```
print(paste0("Mean: ",round(mean(b2i$favorites),2)))
```

```
## [1] "Mean: 3.27"
```

The average number of favorites by mention is about 3 mentions and the maximum number of favorites is 994.

ii. Text of the mention that receives the highest number of favorites

```
mentions %>%
  select(status_id,text,favorite_count) %>%
  arrange(desc(favorite_count)) -> b2ii
kable(b2ii[1,2])
```

x
Game winning moment from 30,000 feet. Thanks @Delta & congrats @Vikings. #SKOL https://t.co/mjlHb5pAez (https://t.co/mjlHb5pAez)

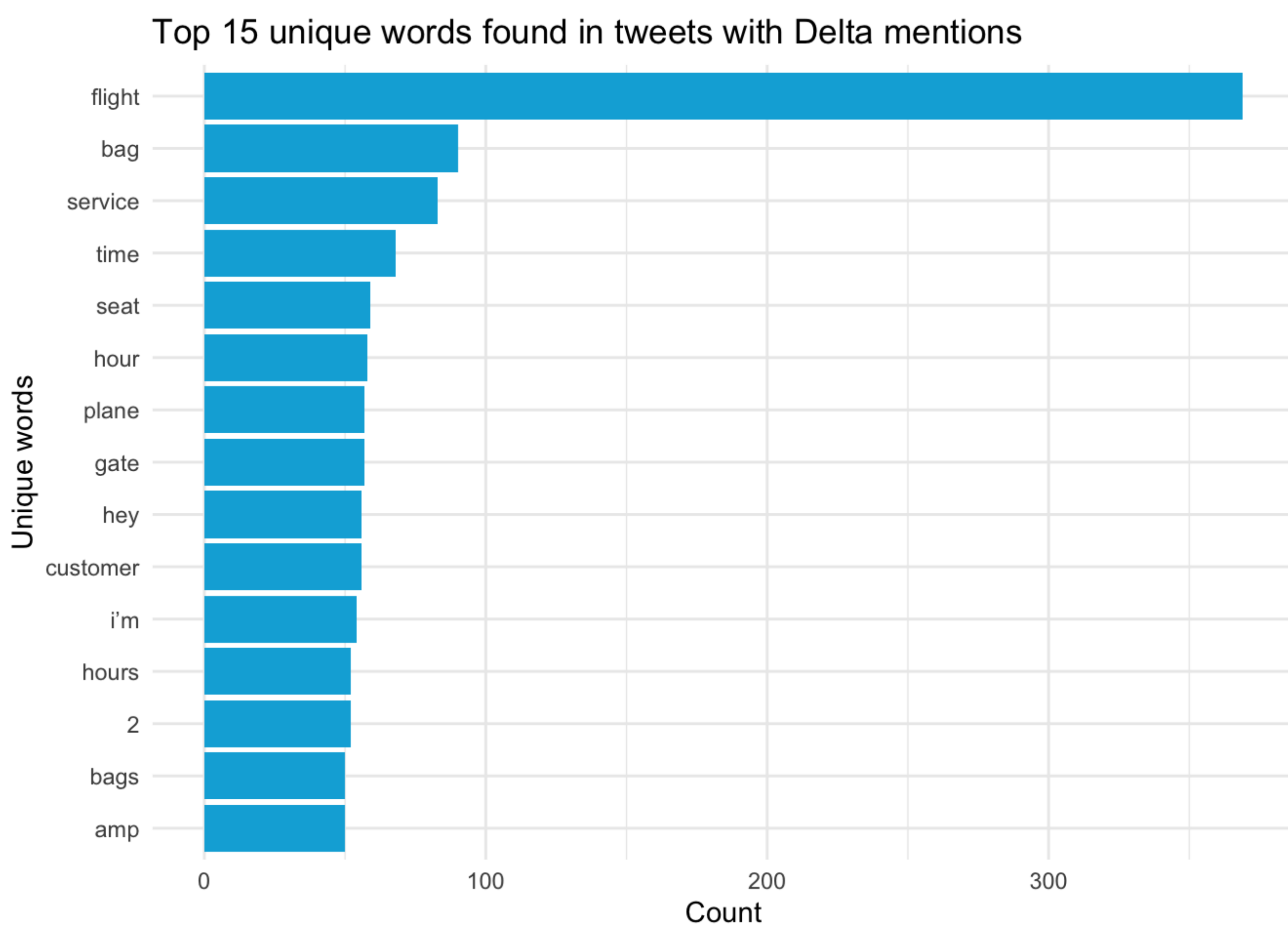
The following text is the mention with the most favorites: “Game winning moment from 30,000 feet. Thanks @Delta & congrats @Vikings. #SKOL https://t.co/mjlHb5pAez (https://t.co/mjlHb5pAez)”

c. Analyzing the text content of Delta’s mentions

```
mentions %>% filter(delta_responded==T) -> c2
# remove web links
c2$text <- gsub("http.*","",  c2$text)
c2$text <- gsub("https.*","", c2$text)
# extract words and remove stopwords
c2 %>% select(text) %>% unnest_tokens(word, text) -> c2_clean
stop_words_custom <- add_row(stop_words, word = "delta", lexicon = "SMART")
c2_clean %>% anti_join(stop_words_custom) -> c2_words
```

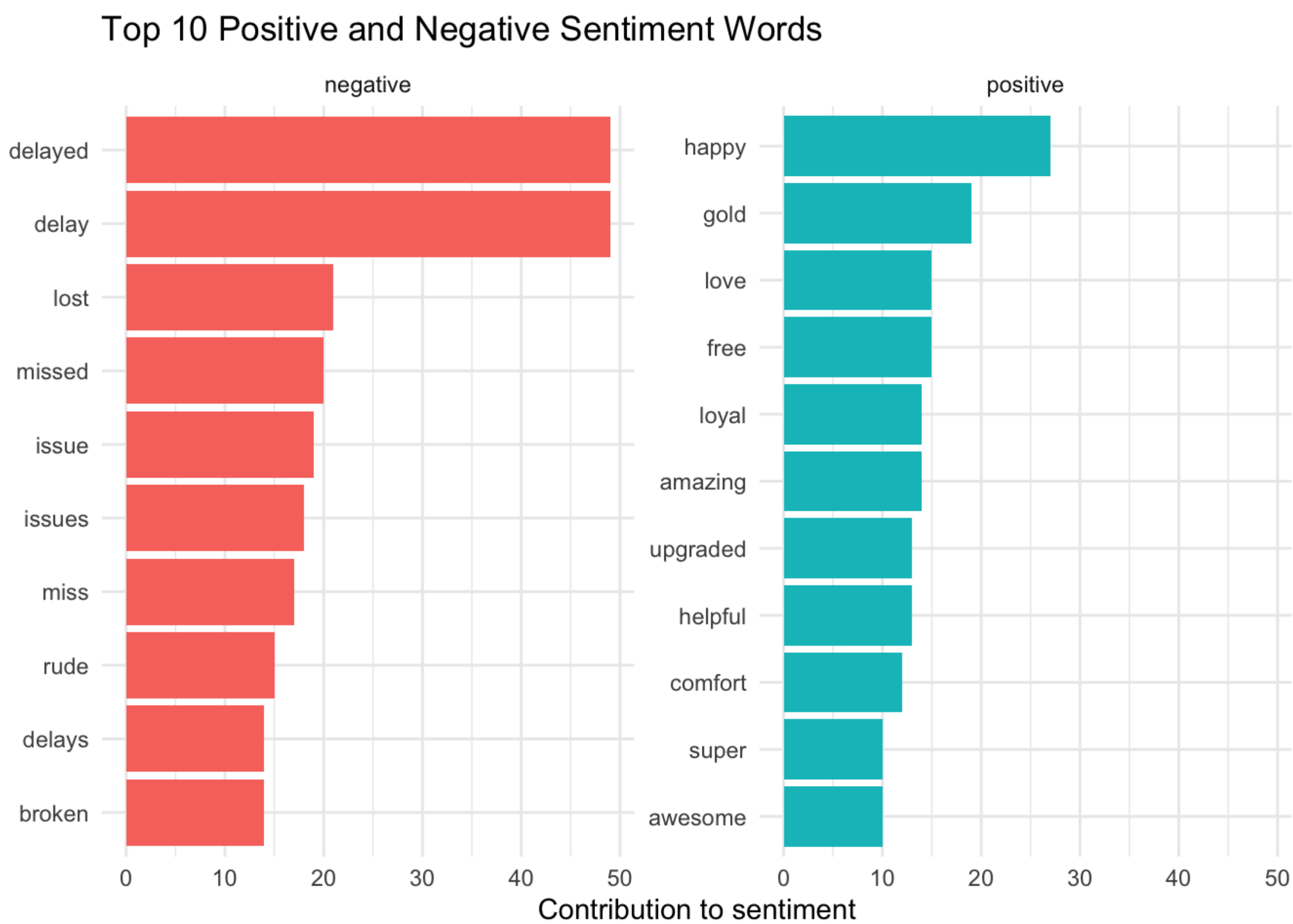
i. Bar plot of 15 unique words

```
c2_words %>% count(word, sort = TRUE) %>% top_n(15) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(x = word, y = n)) +
  geom_col(fill="#00aedb") +
  xlab(NULL) +
  coord_flip() +
  labs(y = "Count",
       x = "Unique words",
       title = "Top 15 unique words found in tweets with Delta mentions") +
  theme_minimal()
```



ii. Top 10 negative versus positive sentiment words

```
# join sentiment classification to the tweet words
bing_word_counts <- c2_words %>%
  inner_join(get_sentiments("bing")) %>%
  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(title = "Top 10 Positive and Negative Sentiment Words",
       y = "Contribution to sentiment",
       x = NULL) +
  coord_flip() + theme_minimal()
```

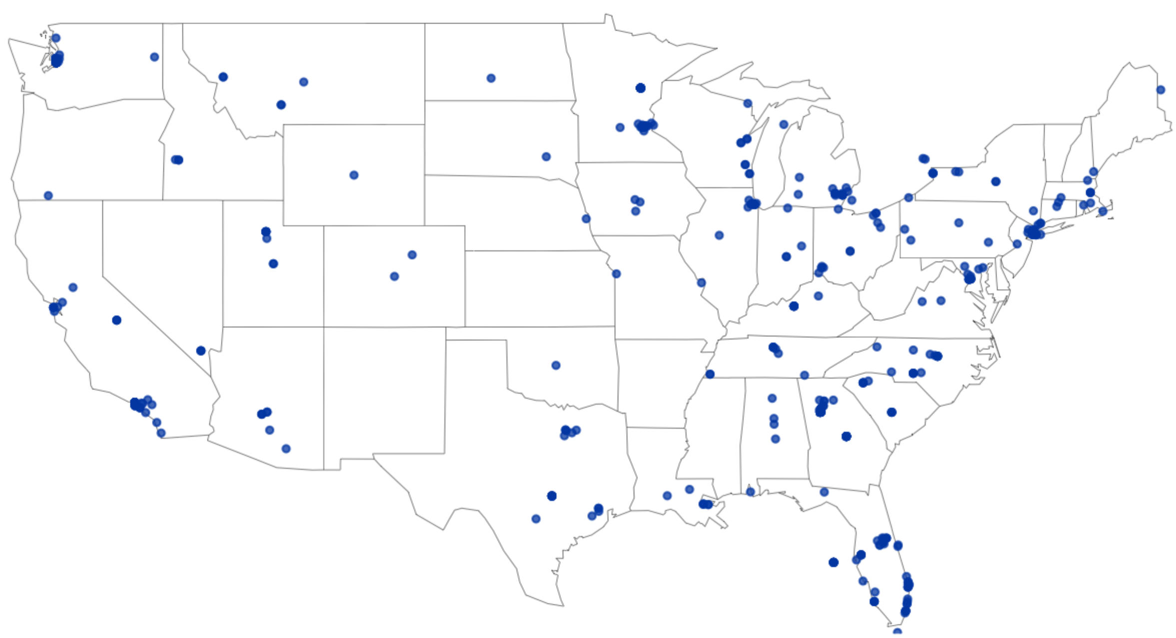


iii. Main recurring customer issues

According to the sentiment analysis, some of the main recurring customer issues are reated to flight delays, lost items, and missed flights. Naturally customers are most likely to complain about flight delays via Twitter and that often causes missed flights and lost baggage, which makes flight delays the main culprit for reasons behind a lot of the other complaints.

d. Location of Delta’s mentions in the US

```
rt <- lat_lng(mentions)
par(mar = c(0, 0, 0, 0))
maps::map("state", lwd = .25)

## lat and lng points for map
with(rt, points(lng, lat, pch = 20, cex = .75, col = rgb(0, .3, .7, .75)))
```



3. Quantitative metrics of customer success

a. Average, median, and maximum number of favorites for replies

```
print(paste0("average favorites for Delta's replies: ",round(mean(mentions$delta_reply_f
avorite_count, na.rm=T),2)))
```

```
## [1] "average favorites for Delta's replies: 0.28"
```

```
print(paste0("median favorites for Delta's replies: ",round(median(mentions$delta_reply_
favorite_count, na.rm=T),2)))
```

```
## [1] "median favorites for Delta's replies: 0"
```

```
print(paste0("maximum favorites for Delta's replies: ",round(max(mentions$delta_reply_fa
vorite_count, na.rm=T),2)))
```

```
## [1] "maximum favorites for Delta's replies: 26"
```

b. Delta’s response rates to its mentions (%)

```
mentions %>% mutate(responded = ifelse(delta_responded=="TRUE",1,0)) %>%
  group_by(responded) %>% add_count() %>% select(responded) -> responded
res_rate = sum(responded) / nrow(responded)
print(paste0("Delta's response rates to its mentions is: ",(round(res_rate, 4) * 100),
"%"))
```

```
## [1] "Delta's response rates to its mentions is: 28.95%"
```

c. Average, median, and maximum response time (minutes) for Delta’s replies

```
mentions %>% mutate(responseTime = as.numeric(delta_reply_created_at - created_at)) %>%
  select(responseTime) -> secs

min <- secs/60

print(paste0("average response time for Delta's replies: ",round(mean(min$responseTime,
 na.rm=T),1)," minutes"))
```

```
## [1] "average response time for Delta's replies: 7.5 minutes"
```

```
print(paste0("median response time for Delta's replies: ",round(median(min$responseTime,
na.rm=T),1)," minutes"))
```

```
## [1] "median response time for Delta's replies: 4.6 minutes"
```

```
print(paste0("maximum response time for Delta's replies: ",round(max(min$responseTime, n
a.rm=T),1)," minutes"))
```

```
## [1] "maximum response time for Delta's replies: 76.1 minutes"
```

d. Strength and limitation for using these customer success metrics

Engagement (number of favorites for Delta's replies):

Pro: A cost-efficient way to get a quantitative overview of the level of insightfulness and usefulness of interactions with external stakeholders.

Con: Often times Delta's replies may not be anything that requires any sort of reaction from the customers, which means a low chance of favoriting the tweet. Therefore, a skewed view of customer success based on number of favorites.

Response rates:

Pro: Easily interpretable metric to understand responsiveness to customers.

Con: Many tweets Delta are mentioned in may not need any sort of response from Delta as the tweets can be a part of a thread only indirectly related to Delta where they don't require any input. These sort of situations can result in a misrepresented measure of customer success.

Response time:

Pro: It gives an idea of how quickly are customers attended to, and an indicator of reply speed overall.

Con: Might not always paint a clear picture of time-efficiency as certain tweets may take longer to reply to due to event-based topics that will occur after a specific timeframe.

4. Deliver effective social care

a. What dictates which tweets get a response?

i. Linear probability model

```
ia4 <- lm(delta_responded ~ followers_count+favorite_count+retweet_count+verified,
   data = mentions)
summary(ia4)
```

```
##
## Call:
## lm(formula = delta_responded ~ followers_count + favorite_count +
##     retweet_count + verified, data = mentions)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3141 -0.2978 -0.2973  0.7022  0.8863
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.977e-01  8.239e-03  36.136  < 2e-16 ***
## followers_count  3.950e-08  1.188e-07   0.333 0.739517
## favorite_count  -4.233e-04  4.605e-04  -0.919 0.358084
## retweet_count    1.496e-03  4.395e-03   0.340 0.733548
## verifiedTRUE    -1.404e-01  3.781e-02  -3.714 0.000208 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4526 on 3197 degrees of freedom
## Multiple R-squared:  0.005524,   Adjusted R-squared:  0.00428
## F-statistic:  4.44 on 4 and 3197 DF,  p-value: 0.001402
```

ii. Summary and limitations

Based on the linear probability model, follower count and retweet count appear to have a positive relationship with Delta's response to a tweet. Surprisingly, if an account is verified and the favorite count have a negative relationship with whether Delta responds or not. However, based on the p-value the only significant variable in this model that dictates which tweets get a response is whether an account is verified or not. Therefore, if an account is NOT verified there is a higher chance of Delta responding to the account's tweets. It is important to note that the adjusted R-squared is quite low, which means the proportion of variation in the dependent variable explained by the independent variables is low.

b. Engage the customer publicly or privately via direct message?

i. What percent of delta's replies direct the customer to a private conversation?

```
replies$tactic_dm <- ifelse(grepl('DM|private message', replies$text, ignore.case=T), 1,
0)

print(paste0(round(sum(replies$tactic_dm) / nrow(replies),3)*100,"% of Delta's replies d
irect the customer to a private conversation"))
```

```
## [1] "32.9% of Delta's replies direct the customer to a private conversation"
```

ii. Why would Delta wish to direct customers to a private conversation?

1. Directing customers to DMs can help Delta reduce the damage done to brand reputation for negative reviews or experiences.
2. Delta can improve customer service satisfaction by adding a more personal touch to the communication between Delta representatives and customers. This can make customers feel valuable, which can gradually increase customer loyalty and brand image.
3. DMs are a good option for interactions between customer and Delta when it invovlves sharing sensitive information to resolve an issue or give further guidance.

5. Twitter responses based on Delta employees

a. How many different employees appear in the data?

```
length(unique(employee$`word(wo_url, -1)`))
```

```
## [1] 76
```

b. What percentage of Delta's replies are written by the top five employees collectively?

```
employee %>% group_by(`word(wo_url, -1)`) %>% count() %>%
  arrange(desc(n)) %>%
  select(employee=`word(wo_url, -1)`, replies_written = n) -> b5

prop_top5 <- colSums(b5[1:5,2])/colSums(b5[,2]) * 100

print(paste0(round(prop_top5,1),"% of Delta's replies are written by the top 5 employees collectively."))
```

```
## [1] "25.5% of Delta's replies are written by the top 5 employees collectively."
```

c. Why would Delta want its employees to sign each tweet?

The policy of having employees sign each tweet creates a greater sense of accountability for the employees that motivates them to keep up with the issue. It may often motivate employees to be more active, take on more cases, and feel a sense of accomplishment. Lastly, it makes it easier to keep track of the the replies. From a customer's standpoint seeing the initials of a an employee makes the customers feel important and diminishes any chance of thinking their concerns are just going into the "blackhole" or just handled by bots.