Cat Couch Sentiment

Johanna Schmidle

This is part 4 (the bonus part!) of my Amazon Cat couch review project. I am just practicing my sentiment analysis skills in R. This will be a pretty simple notebook

Part 1: Data Prep I will be doing two main things: 1. General setup 2. Create Dataset for text analysis

Part 2: Analysis Here are my main plots and points of EDA: 1. WordCloud 2. Words in Sentiments 3. Distribution of Sentiments 4. Summary Table Sentiment vs Couch Colour 5. Review Sentiments vs Couch Colour

Data Prep

Set up

1. Load the libraries
2. Read in database
3. Clean the column names

```r
library(readr)
library(dplyr)
library(kableExtra)
library(tidyverse)
library(tidytext)
library(wordcloud)
library(RColorBrewer)
library(ggplot2)
library(textdata)
library(sentimentr)
```

```r
couch <- read_csv("/Users/johannaschmidle/Desktop/Data Analyst Projects/Cat Couch/CatCouchReviewsTable.c
                  col_names = TRUE,
                  col_types = cols())
```

```r
# Edit column names
couch<-couch %>%
  rename(Location=`Review Location`)  %>%
  rename(Reviewer=`Reviewer Name`) %>%
  rename(Date=`Review Date`) %>%
  rename(Text=`Review Text`) %>%
  rename(Colour=`Colour Name`)

couch %>% glimpse
```

```
## Rows: 100
## Columns: 7
## $ Reviewer <chr> "Shu", "Rebecca Baker", "CMac", "Cj", "MIKE", "Randi F.", "RF~
## $ Title    <chr> "True Value", "Elderly cat loves it!", "Better than I expecte~
```

```
## $ Rating   <dbl> 5, 5, 5, 4, 5, 1, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 4, 4, 5, 3~
## $ Location <chr> "Canada", "Canada", "Canada", "Canada", "Canada", "Canada", "~
## $ Date     <date> 2024-02-19, 2024-02-20, 2024-02-21, 2024-02-22, 2024-02-23, ~
## $ Text     <chr> "The mini sofa is great, big enough to fit my large cat. Ador~
## $ Colour   <chr> "Yellow", "Green", "Green", "Yellow", "Green", "Blue", "Pink"~
```

Text Dataset

This is for text analysis focused on the comments column, so I will create a special version of the dataset for this purpose.

1. Create ufo_txt dataset
2. Unnest tokens
3. Remove numbers from comments
4. Remove stop words
5. Add sentiment column

```r
v_stopwords <- get_stopwords() %>% .$word
numb <- "[[:digit:]]+"

couch_txt<-couch %>%
  unnest_tokens(word,Text)

couch_txt <- couch_txt %>%
  filter(!str_detect(word,numb))%>%
  filter(!(word %in% v_stopwords))%>%
  left_join(get_sentiments("nrc"), by="word")

dim_desc(couch_txt)
```

```
## [1] "[1,873 x 8]"
```

Text Analysis

WordCloud for Review Text

```r
pallet <- brewer.pal(15,"Set2")
couch_txt %>%
  count(word,sort=TRUE)%>%
  with(wordcloud(word,n, colors=pallet, random.order = FALSE, min.freq = 500))
```

Some of the top words are **cat**, **love**, **perfect**, **happy**, **good**. These are all postive words so this indicates that the sentiments of the Review Text should be mostly positive. But note that the word **disappointed** is also big, so this indicates that the people who had negative feelings towards the couch were mostly feeling disappointed. I will analyze the sentiments further and see what I can find.

Distribution of Sentiments

I want to see how most people when reviewing the Cat Couch. I will do this by plotting and comparing the amount of times each sentiment is seen in the *emotion* column. I will do this using a pie chart. Although pie charts are not always the most effective way to plot data, I think this plot is the easiest to visually analyze as you clearly see which sentiment is dominant.

```
#Choose sentiments I want to see
senti <- c("positive", "fear","surprise", "negative", "anger", "disgust")


couch_txt%>%
  count(sentiment) %>%
  filter(sentiment %in% senti) %>%
  mutate(sentiment=fct_reorder(sentiment,n)) %>%
  ggplot() +
  geom_bar(aes(x="", y=n,fill=sentiment), stat = "identity") +
  coord_polar("y",start=0) +
  theme(axis.text.x=element_blank()) +
  labs(x="",y="",title="Sentiments",fill="")
```

Sentiments
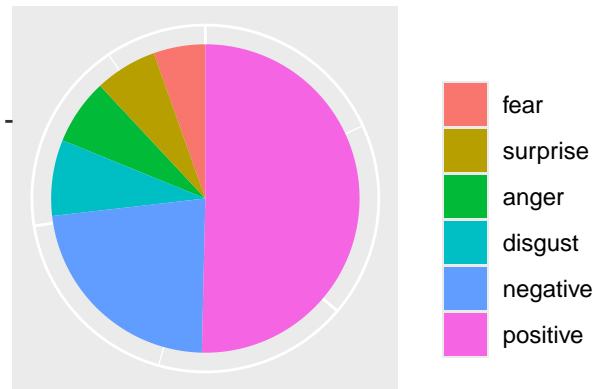
As we can see the sentiments are overwhelmingly positive. This makes sense with the words we saw in the WordCloud. The second most common emotion was negative, which could be where the word *disappointed* came from.

Bar Charts for Top Words in Sentiments

I'm curious to explore the most frequently used words associated with each emotion. I plan to visualize this information using bar graphs for better readability.

```
#Sentiment Bar Charts
couch_txt %>%
  count(sentiment, word) %>%
  filter(sentiment %in% senti) %>%
  group_by(sentiment) %>%
  top_n(10,n) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(x=word, y=n, fill = sentiment)) +
  geom_bar(stat = "identity", show.legend = FALSE) +
  guides(x=guide_axis(angle = 45)) + coord_flip() +
  facet_wrap(~sentiment, scales = "free") +
  labs(y = "", x = "",
       title = "Sentiment analysis of Cat Couch Reviews",
       subtitle = "Using the NRC Lexicon")
```

Sentiment analysis of Cat Couch Reviews
Using the NRC Lexicon
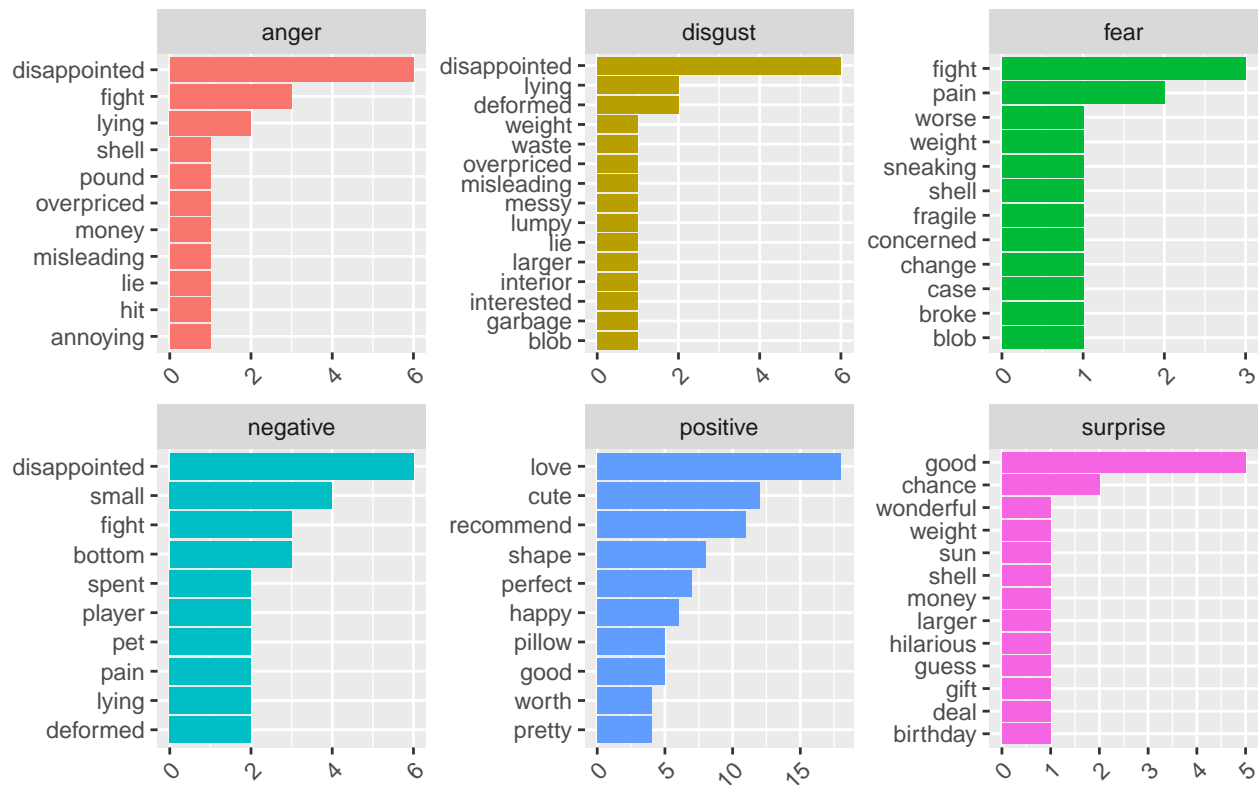
Notice that the word **disappointed** is the top word for the emotions *anger*, *disgust*, and *negative*. This could explain why the word was so large in our WordCloud.

Summary table

I am creating a table summary to analyze some statistics for the relationship between sentiment (positive or negative) and the color of the purchased couch.

```
couch_summary <- couch_txt %>%
  filter(!is.na(Colour) & !is.na(sentiment)) %>%
  group_by(Colour) %>%
  summarize(
    mean_sentiment = mean(ifelse(sentiment == "positive", 1, ifelse(sentiment == "negative", -1, 0))),
    num_positive = sum(sentiment == "positive"),
    num_negative = sum(sentiment == "negative")
  )

couch_summary
```

```
## # A tibble: 6 x 4
##   Colour mean_sentiment num_positive num_negative
##   <chr>           <dbl>        <int>        <int>
## 1 Black          0.35             9            2
## 2 Blue           0.0361          17           14
## 3 Green          0.133           32           14
## 4 Pink           0.235           12            4
## 5 White          0.163           22            7
```

5

```
## 6 Yellow          0.125              17              9
```

Overall Positivity: Colours like *Black*, *Pink*, *White*, and *Yellow* have mean sentiment scores above 0.1, indicating generally positive sentiment among reviews for these couch colors. But note *Black* and *Pink* have relatively low counts of negative reviews (2 and 4, respectively), suggesting they are perceived positively by most reviewers.

Varied Sentiment Levels: *Blue* and *Green* have mean sentiment scores closer to zero, suggesting mixed sentiments or a more neutral sentiment distribution compared to other colors.

Positive vs. Negative Reviews: While most colors have more positive reviews than negative ones, Blue stands out with a higher count of negative reviews (14) compared to its positive reviews (17).

Popular Colours: Green appears to have a higher number of reviews overall (46 in total)

Review Sentiment vs Couch Colour

I am going to display the information from above. I want to compare the positive and negative sentiments per couch colour, and I also want to compare all sentiments per couch colour. I will use grouped bar charts for this

```r
couch_txt %>%
  filter(!is.na(sentiment)) %>%
  filter(!is.na(Colour)) %>%
  filter(sentiment %in% c("positive", "negative")) %>%
  count(Colour, sentiment) %>%
  ggplot(aes(x = Colour, y = n, fill = sentiment)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Sentiment Distribution by Couch Color",
       x = "Couch Color",
       y = "Number of Reviews") +
  theme_minimal()
```
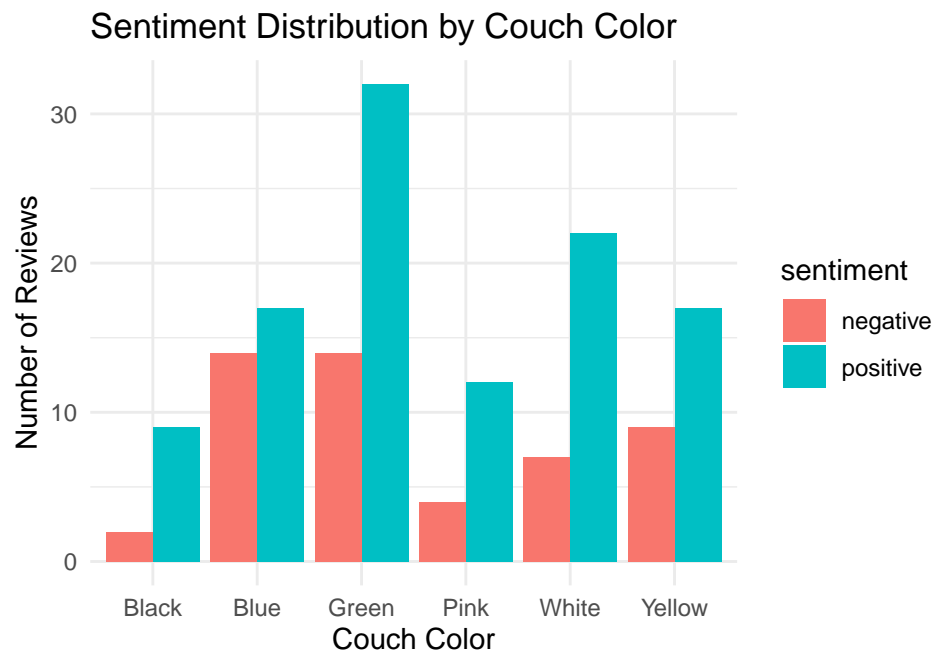


```r
couch_txt %>%
  filter(!is.na(sentiment)) %>%
  filter(!is.na(Colour)) %>%
```

```
count(Colour, sentiment) %>%
ggplot(aes(x = Colour, y = n, fill = sentiment)) +
geom_bar(stat = "identity", position = "dodge") +
labs(title = "Sentiment Distribution by Couch Color",
     x = "Couch Color",
     y = "Number of Reviews") +
theme_minimal()
```

Sentiment Distribution by Couch Color