

# Animal Crossing - A Sentiment Analysis on community Reviews

James Liang

August 5th 2023

## Introduction

A beloved exclusive of the Nintendo franchise, Animal Crossing: New Horizon, provides a unique escape from the hustle and bustle of modern day life, drawing gamers into the tranquil world of colorful characters and creative freedom to build their own civilisation.

In this report, I aim to gain insight into how the game was received by:

1. Professional critics
2. Casual gamers

during the initial few months, following Animal Crossings' launch in 2020.

Specifically, I cover the time period from 2020-03-20 to 2020-05-03, where the game was released, closely following the COVID-19 pandemic.

## Project Aim:

To determine the community sentiment of Animal Crossing: New Horizons, following its launch in 2020, as well as to ascertain the possible reasons behind the sentiments.

## Main Analysis Goals:

- Overall sentiments over time?
- Frequent discussion points amongst users?
- Differing sentiments between professional critics and casual gamers?
- What were the most prevalent sentiments and why did they exist?

## Loading in the Data

```
library(tidyverse)
library(stopwords)
library(tidyverse)
library(tidytext)
library(textdata)
library(wordcloud)
library(RColorBrewer)

library(broom)
library(gridExtra)
library(MASS)
library(kableExtra)
```

## Data

The gamers and critic reviews are scraped from Metacritic.

See: Animal Crossing: New Horizon

Sourced from: #TidyTuesday challenge

```
critics <- readr::read_tsv(critic.tsv)
user_reviews <- readr::read_tsv(user_reviews.tsv)
```

**Exploratory Data Analysis** In order to understand how we should process and clean the data, let us first conduct some basic exploratory data analysis to understand what we are working with.

```
dim_critics <- dim(critics)
dim_users <- dim(user_reviews)
```

When examining this dataset, one thing that should be prefaced is that there are a significantly greater number of gamer reviews (2999), as compared to professional critics (107). This discrepancy is unavoidable, given that there are only so many professional critics out there. However, these reviews usually represent the opinions of a collective group, and therefore, the validity of comparing gamer sentiments to critics still holds as an interesting analysis.

Note that the average grade given by users (converted to a percentage), is 42.17, whilst the average grade given by critics is 91. This difference in grades can be expected due to various reasons. For example, professional critics will often, holistically review all the components of a game, including design, gameplay and narrative, whilst casual gamers are far more subjective and tend to only review things that they have a strong opinion about.

In this sense, we would expect, in this analysis, that casual gamers **should** display a far greater, negative sentiment as opposed to critics. Let us examine whether this initial hypothesis regarding sentiments, is true.

## Methodology

To analyse the data, I employed a Bag of words-based approach.

- That is, I broke each review into individual words, and will remove the stop words.
- I would then determine the sentiment score of each word based on a lexicon and assign a sentiment score to each individual review. This is conducted through using left\_joins and inner\_joins to map particular words.

I used 2 lexicons for my analysis.

1. AFINN. This quantifies sentiments through a number, showing its intensity.
2. NRC, which categorises sentiments into moods like sad, joy, surprise ect.

## Preparation of Data

To conduct the analysis, we will first remove words that lack semantic meaning - or stop words.

Stopwords can be described as filler words. They include words like: “the”, “to”, “and”, ect.

```
# Break down critic review and user review into
# individual words
critic_words <- critics %>%
  unnest_tokens(output = word, input = text)
user_words <- user_reviews %>%
  unnest_tokens(output = word, input = text)
```

```

# Get the individual stop words
stopwords_smart <- get_stopwords(source = "smart")

# Remove stopwords
critics_words_no_stop <- critic_words %>%
  anti_join(stopwords_smart)
user_words_no_stop <- user_words %>%
  anti_join(stopwords_smart)

kable(head(user_words_no_stop))

```

grade	user_name	date	word
4	mds27272	2020-03-20	gf
4	mds27272	2020-03-20	started
4	mds27272	2020-03-20	playing
4	mds27272	2020-03-20	option
4	mds27272	2020-03-20	create
4	mds27272	2020-03-20	island

## Analysis

Are longer review more positive?

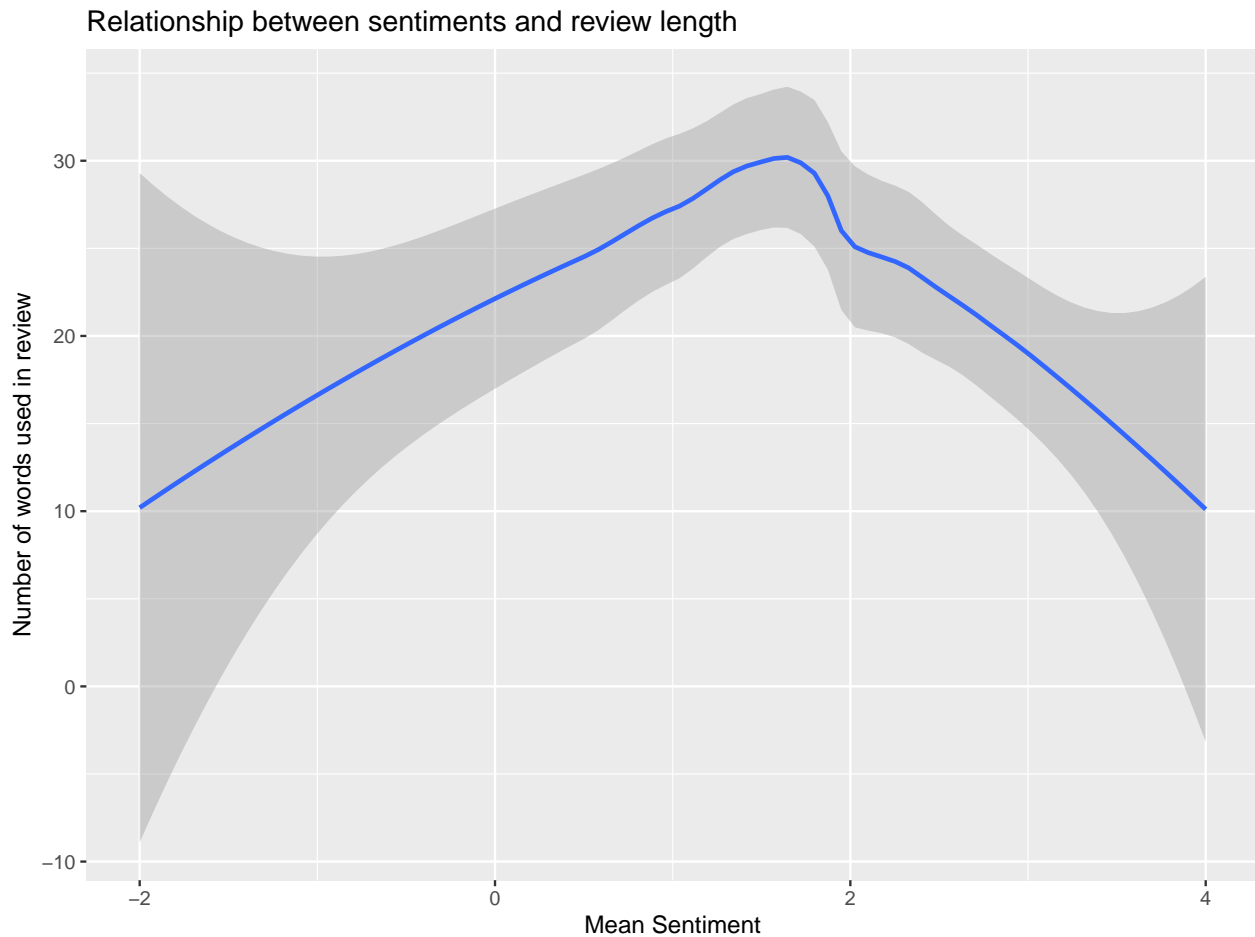
```

afinn <- get_sentiments("afinn")
nrc <- get_sentiments("nrc")
# Add the sentiment values to relevant words of
# critic_words
critic_words <- critics_words_no_stop %>%
  left_join(afinn)

# For each publication review, compute the
# average sentiment
sentiment_avg <- critic_words %>%
  group_by(publication) %>%
  summarise(mean_sentiment = mean(value, na.rm = TRUE),
    n_words = n())

sentiment_avg %>%
  ggplot(aes(x = mean_sentiment, y = n_words)) +
  geom_smooth() + labs(title = "Relationship between sentiments and review length",
    x = "Mean Sentiment", y = "Number of words used in review")

```



Firstly, a question that I was curious about, regards whether longer reviews, generally correlated to more positive sentiments? That is - are people who are writing a lot, happy with the game? To answer this question, I use `geom_smooth()` to show the overall pattern of the data. In this case, it appears that the length of a review, did not necessarily help with predicting the sentiment of a review. That is, it appears that reviews with highly negative sentiments use a similar number of words, as reviews that are highly positive.

Therefore, people who are writing a lot of text in their reviews, may either be displaying positive or negative emotions. The length of the review itself, is not indicative of a persons sentiment for the game.

## Average Sentiment over Time

```
# Replace NA values with a sentiment value of 0.
critic_avg <- critic_words %>%
  na.omit()

# Draw a line graph, that takes the average
# sentiment value for each date.
critic_df_plot <- critic_avg %>%
  group_by(date) %>%
  summarise(avg_critic_sentiment = mean(value))

# Add the sentiment values to relevant words of
# critic_words
user_words <- user_words_no_stop %>%
  left_join(afinn)
```

```

# Replace NA values with a sentiment value of 0.
user_avg <- user_words %>%
  na.omit()

# Draw a line graph, that takes the average
# sentiment value for each date.
user_df_plot <- user_avg %>%
  filter(date >= min(critic_avg$date) & date <= max(critic_avg$date)) %>%
  group_by(date) %>%
  summarise(avg_user_sentiment = mean(value))

inner_join(user_df_plot, critic_df_plot) %>%
  reshape2::melt(id.var = "date") %>%
  ggplot(aes(date, value, col = variable)) + geom_line() +
  labs(title = "Average sentiment in the first few months",
       x = "Date", y = "Average Sentiment") #+

```

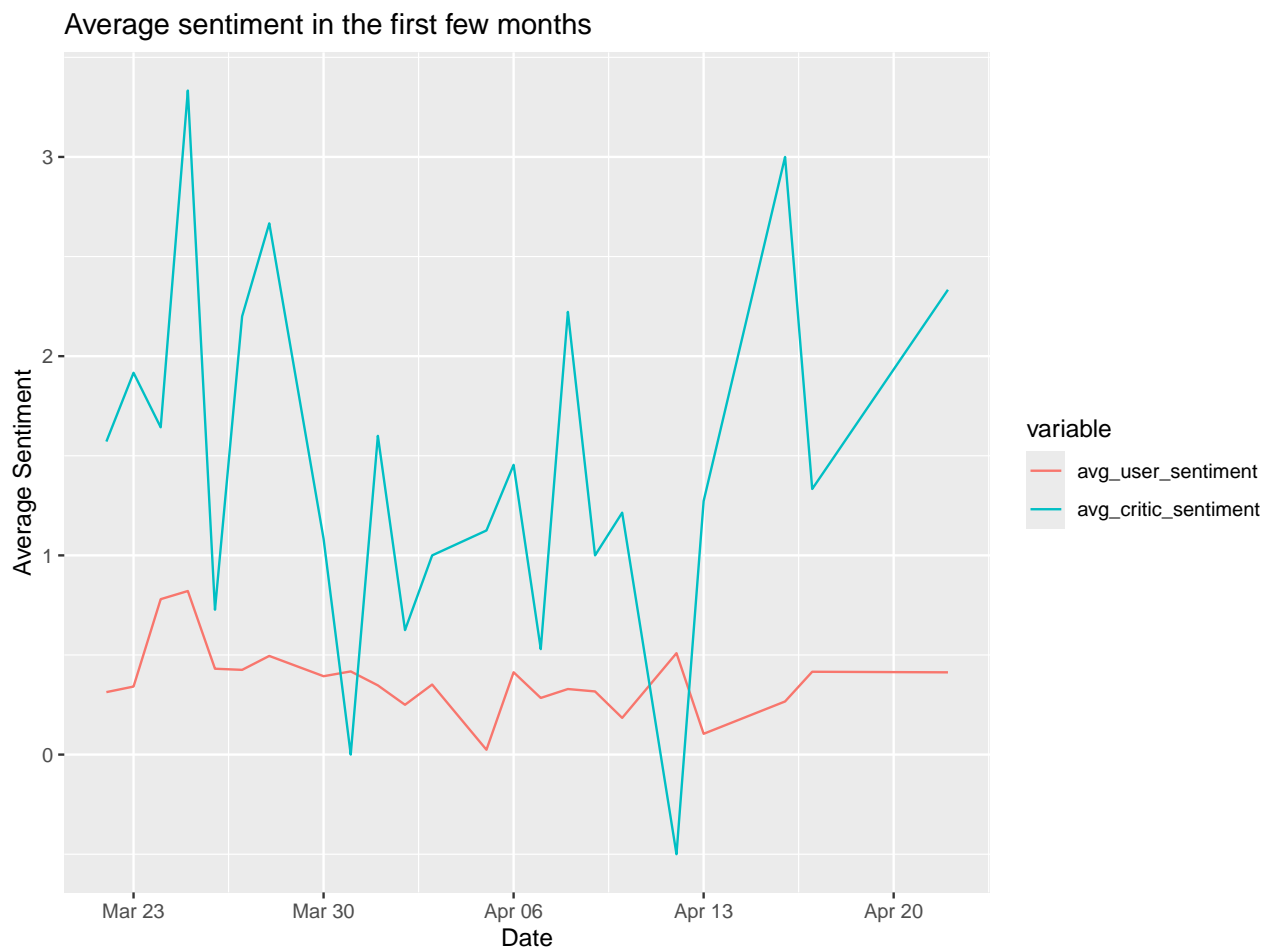


Figure 1: Average Sentiment over time - Critics vs Users

```

# geom_smooth(method = 'lm', se = FALSE)

```

Overall, sentiments of the reviews surrounding Animal Crossing have been positive following the games release. For both critics and casual gamers alike, the sentiment scores consistently stay above 0, which is an

indicator of a neutral sentiment.

In this analysis, we gain an interesting insight.

- (1) It appears that the high average scores ( $\sim 9/10$ ) that critics give, match the positive sentiments of their reviews. This was to be expected.
- (2) On the other hand, despite gamers' giving Animal Crossing such a harsh review ( $\sim 4/10$ ), **the sentiments of the words that they use are generally positive.**

This suggests that, despite the criticality of how gamers may rate a game, it appears that they do not actually “hate” the game as much as we may presume. This may be due to the lack of standardisation gamers use when rating a game, resulting a skewed, or biased score representations that do not exactly match the sentiments of their reviews.

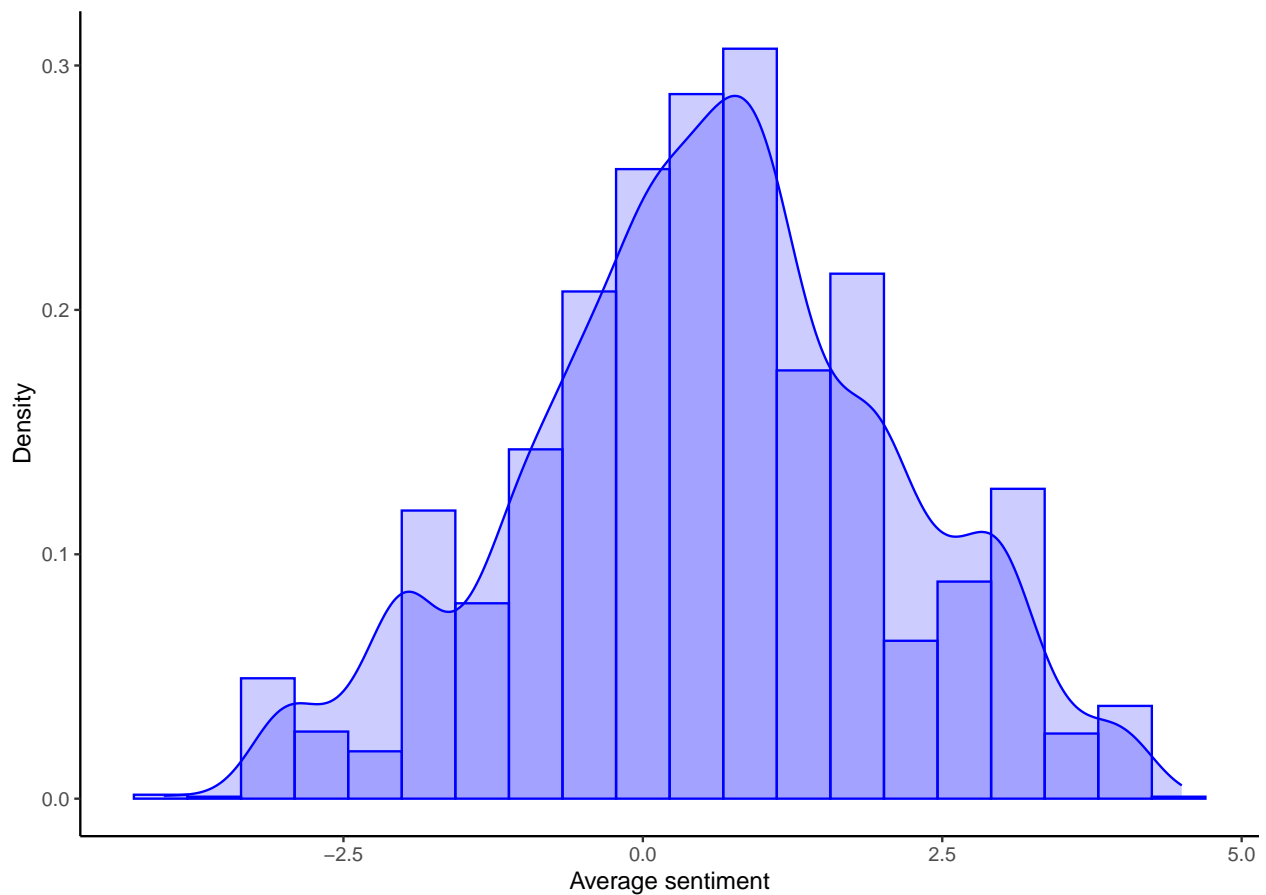
Would this still hold, if we assumed that we had a larger sample of gamer reviews to work with? Let us examine this idea through the idea of simulating resampling through bootstrapping.

## Analysing the sentiments of gamers more closely

Bootstrapping is a resampling technique used in statistics for estimating the sampling distribution of a statistic by repeatedly sampling with replacement from the observed data. It is particularly useful when sample sizes are small, and can provide more accurate estimates and confidence intervals compared to traditional methods, which may rely on large sample approximations.

```
user_avg %>%
  group_by(user_name) %>%
  summarise(avg_user_sentiment = mean(value)) %>%
  ggplot(aes(x = avg_user_sentiment, y = after_stat(density))) +
  geom_histogram(color = "blue", fill = "blue", alpha = 0.2,
    bins = 20) + geom_density(color = "blue", fill = "blue",
    alpha = 0.2) + labs(x = "Average sentiment", y = "Density",
    title = "Distribution of Gamer sentiments") + theme_classic() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```

## Distribution of Gamer sentiments



Observe that the distribution of gamer sentiments are approximately bell shaped. Therefore, let us model the sentiments of gamers through a Normal Distribution. (normal distributions assume that data is evenly distributed about the mean).

### Use Maximum Likelihood to fit a Normal Distribution

```
##### the bootplot.f function #####

## This function 'bootplot.f' takes a vector of
## Bootstrap samples as the main argument
## ('stat_boot'), and produces a plot showing the
## histogram, with smooth density estimate
## overlay, and also provides a option (detail)
## for the number of *bins* used in the
## histogram.

bootplot.f <- function(stat.boot, bins = 50) {

  df <- tibble(stat = stat.boot)
  CI <- round(quantile(stat.boot, c(0.025, 0.975)),
    2)
  p <- df %>%
    ggplot(aes(x = stat, y = after_stat(density))) +
    geom_histogram(bins = bins, colour = "magenta",
```

```

        fill = "magenta", alpha = 0.2) + geom_density(fill = "magenta",
colour = "magenta", alpha = 0.2) + geom_vline(xintercept = CI,
colour = "magenta", linetype = 5) + theme_bw()

    p
}

##### end of bootplot.f function #####

set.seed(123) # Reproducibility

users <- user_avg %>%
  group_by(user_name) %>%
  summarise(avg_user_sentiment = mean(value))

x <- users$avg_user_sentiment
n <- nrow(users)
df <- tibble(id = 1:n, x = x)

# MLE Estimate
fit <- fitdistr(df$x, "normal")
params <- fit$estimate

# Creating 5000 bootstrap samples (with
# replacement), to calculate an MLE for each
# sample

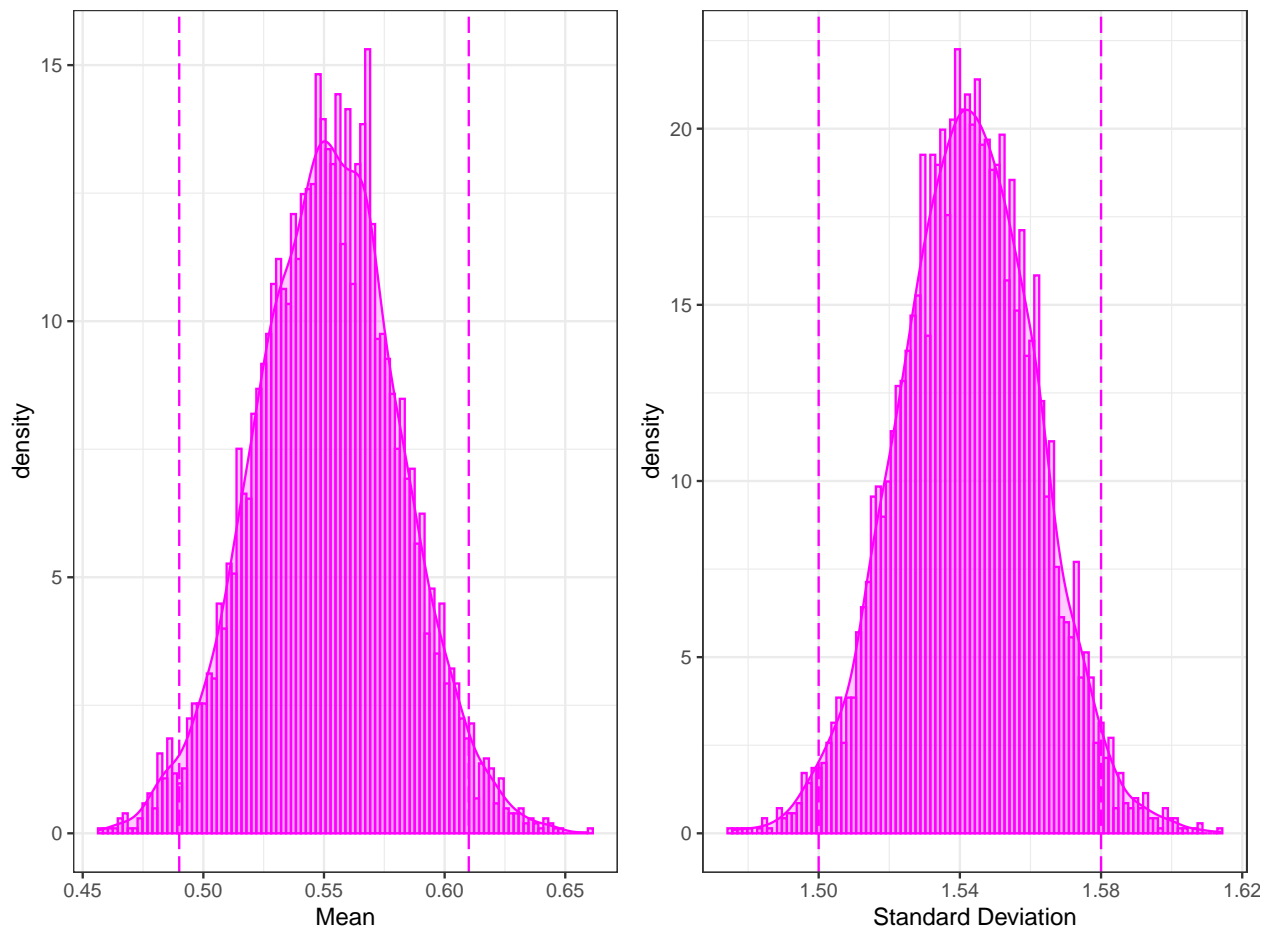
B <- 5000
param <- 2
MLE.x_boot <- matrix(rep(NA, param * B), nrow = B,
  ncol = param)
for (i in 1:B) {
  temp <- sample(df$x, size = n, replace = TRUE)
  MLE.x_boot[i, ] <- fitdistr(temp, "normal")$estimate
}

boot.LCI.mu <- quantile(MLE.x_boot[, 1], c(0.025, 0.975))

p_MLEboot.mu <- bootplot.f(MLE.x_boot[, 1], bins = 100) +
  xlab("Mean")
p_MLEboot.sig <- bootplot.f(MLE.x_boot[, 2], bins = 100) +
  xlab("Standard Deviation")
grid.arrange(p_MLEboot.mu, p_MLEboot.sig, ncol = 2)

```





```
boot.LCI.mu <- quantile(MLE.x_boot[, 1], c(0.025, 0.975))
boot.LCI.sig <- quantile(MLE.x_boot[, 2], c(0.025,
0.975))
```

Observe the above 95% confidence interval for the average sentiment. This tells us that we are 95% confident that the true average sentiment lies somewhere between the range of 0.4945212 and 0.6094607.

This process can also be repeated for the standard deviation of sentiments for gamers. From the results of the bootstrap, we are 95% confident that the true standard deviation of sentiments for gamers lies somewhere between the range of 1.5041186 and 1.5795391.

This suggests a large potential variability in the sentiments of gamers. Therefore, our conclusion based on this experiment can be summarised as such:

- Despite the low “game score” given to Animal Crossing by gamers, the sentiments of the reviews surrounding the game is generally positive, indicated by the positive mean sentiment. Having said that, it should also be taken into account, that there is a lot of variability in the sentiments.

**Potential Limitations** A potential limit of the bag of words approach is that when we tokenise the text, we end up losing a lot contextual meaning within the reviews. For example, see the result below, where people have given the game a review score of 10, but use scathing words in the review text itself. Therefore, when matching lexicons, we end up with a negative sentiment, despite an obvious positive review to the game.

This means that the analysis that has been conducted, is not perfect. Therefore, results taken from the analysis should be taken as a general indicator of the general sentiment surrounding the game, rather than an indicative sentiment of how all gamers felt.

```

mismatched_reviews <- user_avg %>%
  group_by(user_name) %>%
  summarise(avg_user_sentiment = mean(value), grade = mean(grade)) %>%
  filter(avg_user_sentiment < 0 & grade == 10)

temp <- user_reviews %>%
  filter(user_name == "Cephey")
kable(temp[, -2]) |>
  kable_styling(latex_options = c("scale_down", "small",
    "hold_position"))

```

grade	text	date
10	I hate HATERS, i set 0 to all heaters! There is no anymore point in user score when just some dump haters set 0, and that mean NOTHING.	2020-03-25

## So, what is everyone talking about?

Note that all words and their negativity/positivity has been determined by the afinn sentiment list of words.

```

# Get the negative sentiment words, alongside
# their occurrence
set.seed(10)
a <- user_avg %>%
  group_by(word, value) %>%
  filter(value < 0) %>%
  summarise(count = n()) %>%
  arrange(-count)

# Create the word cloud
user_neg_cloud <- wordcloud(a$word, a$count, min.freq = 0,
  max.words = 20, colors = brewer.pal(5, "Dark2"))

```

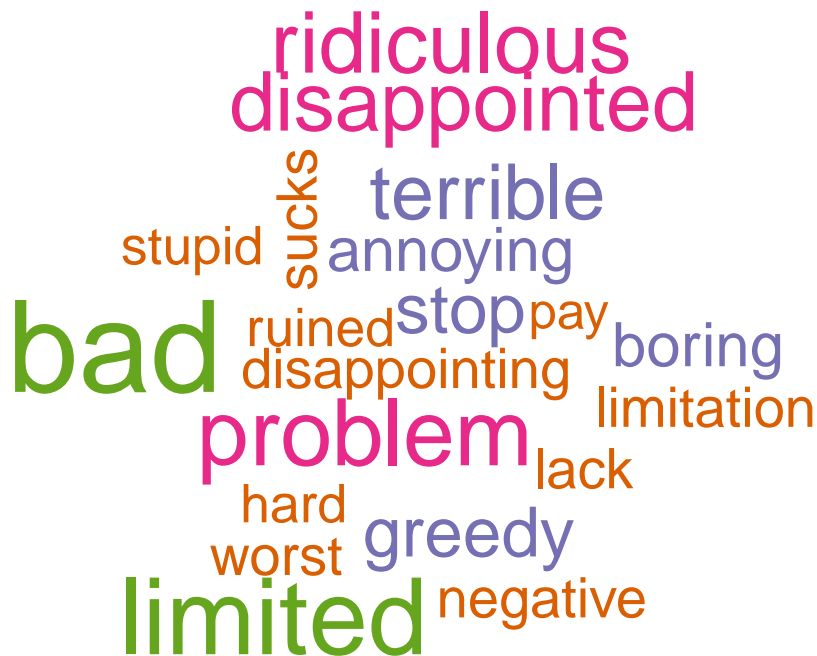


Figure 2: Gamers - Negative Word Cloud

The people that rate the game the most harshly in their reviews are casual gamers. Let us investigate what they are talking about:

It appears that the most frequently used negative terms, relate to simplistic and general terms that rate the game as 'bad', 'limited', or 'disappointed'. This can be from a variety of things. For example, Animal Crossing has limited multiplayer capabilities - supporting only local co-op play. This may have come as a disappointment to many avid fans of the game :(.

People have also described the game negatively, using terms such as 'pay', or 'greedy' in their reviews. The hidden meanings that can be logically deduced, is that people dislike the fact that Animal Crossing has elements where players can spend real-money to advance their progression in the game.

Therefore, negative reviews, at least extracted via `afinn`, relate to a dislike of microtransactions, and the limitations in gameplay potential of Animal Crossing.

```
# Get the positive sentiment words, alongside
# their occurrence
set.seed(20)
a <- user_avg %>%
  group_by(word, value) %>%
  filter(value > 0) %>%
  summarise(count = n()) %>%
  arrange(-count)

# Create the word cloud
user_pos_cloud <- wordcloud(a$word, a$count, min.freq = 0,
  max.words = 20, colors = brewer.pal(5, "Dark2"))
```



Figure 3: Users - Positive Word Cloud

```
# Get the positive sentiment words, alongside
# their occurrence
set.seed(10)
a <- critic_avg %>%
  group_by(word, value) %>%
  filter(value > 0) %>%
  summarise(count = n()) %>%
  arrange(-count)
```

```
# Create the word cloud
critic_pos_cloud <- wordcloud(a$word, a$count, min.freq = 0,
  max.words = 20, colors = brewer.pal(5, "Dark2"))
```



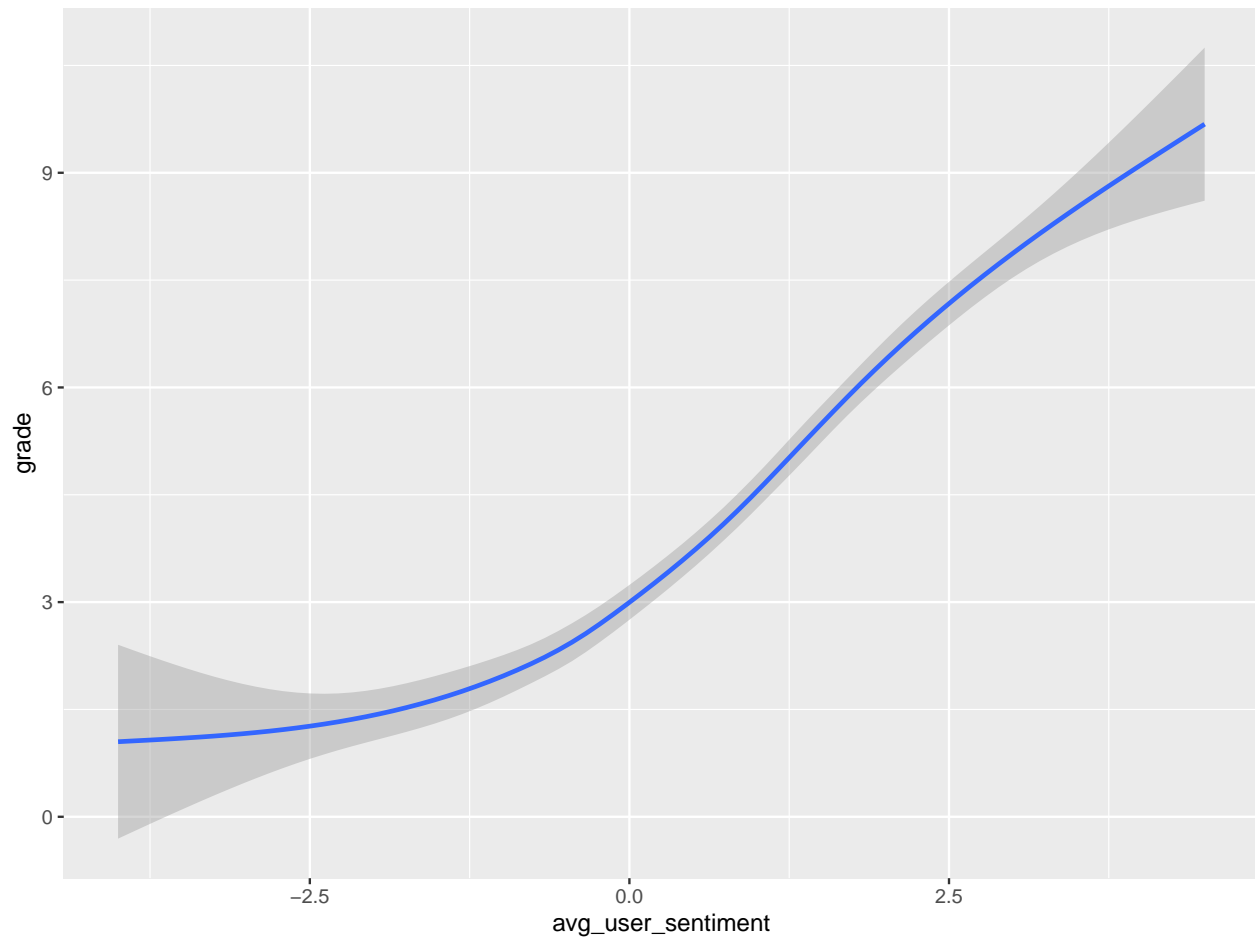
Figure 4: Critics - Positive Word Cloud

Now, let us compare the positive sentiments between gamers and critics. Whilst gamers appear to be more focused on aspects of the game relating to gameplay, critics, evidently seem more focused on the hollistic aspect of the game, opting to describe things like scenery, narrative, and its ambience.

Praise for the game frequently use words like 'fun', 'great', 'perfect', indicating that the gameplay of Animal Crossing is also received quite positively by the community. Words like 'freedom', 'worth' and 'rewarding', seem especially fitting for a game such as Animal Crossing, where a defining feature of the game, is its encouragement of creativity and peacefulness. If a life-simulating game such as Animal Crossing is praised with these words, then it is safe to say that the game is earning praise for its unique features.

## Graphs to summarise perceptions on Animal Crossing - New Horizons.

```
user_avg %>%
  group_by(user_name) %>%
  summarise(avg_user_sentiment = mean(value), grade = mean(grade)) %>%
  ggplot(aes(x = avg_user_sentiment, y = grade)) +
  geom_smooth()
```



A graph to show that the more positive the gamer sentiment for a game, the higher the review score they give to the game.

## Using NRC

```
# Add the sentiment values to relevant words of
# critic_words
critic_words_nrc <- critics_words_no_stop %>%
  left_join(nrc, by = "word")
critic_words_nrc %>%
  na.omit() %>%
  filter(sentiment %in% c("positive", "negative")) %>%
  ggplot(aes(sentiment, fill = sentiment, color = sentiment)) +
  geom_bar()
```

```
# Add the sentiment values to relevant words of
# gamers
user_words_nrc <- user_words_no_stop %>%
  left_join(nrc, by = "word")
user_words_nrc %>%
  na.omit() %>%
  filter(sentiment %in% c("positive", "negative")) %>%
  ggplot(aes(sentiment, fill = sentiment, color = sentiment)) +
  geom_bar()
```

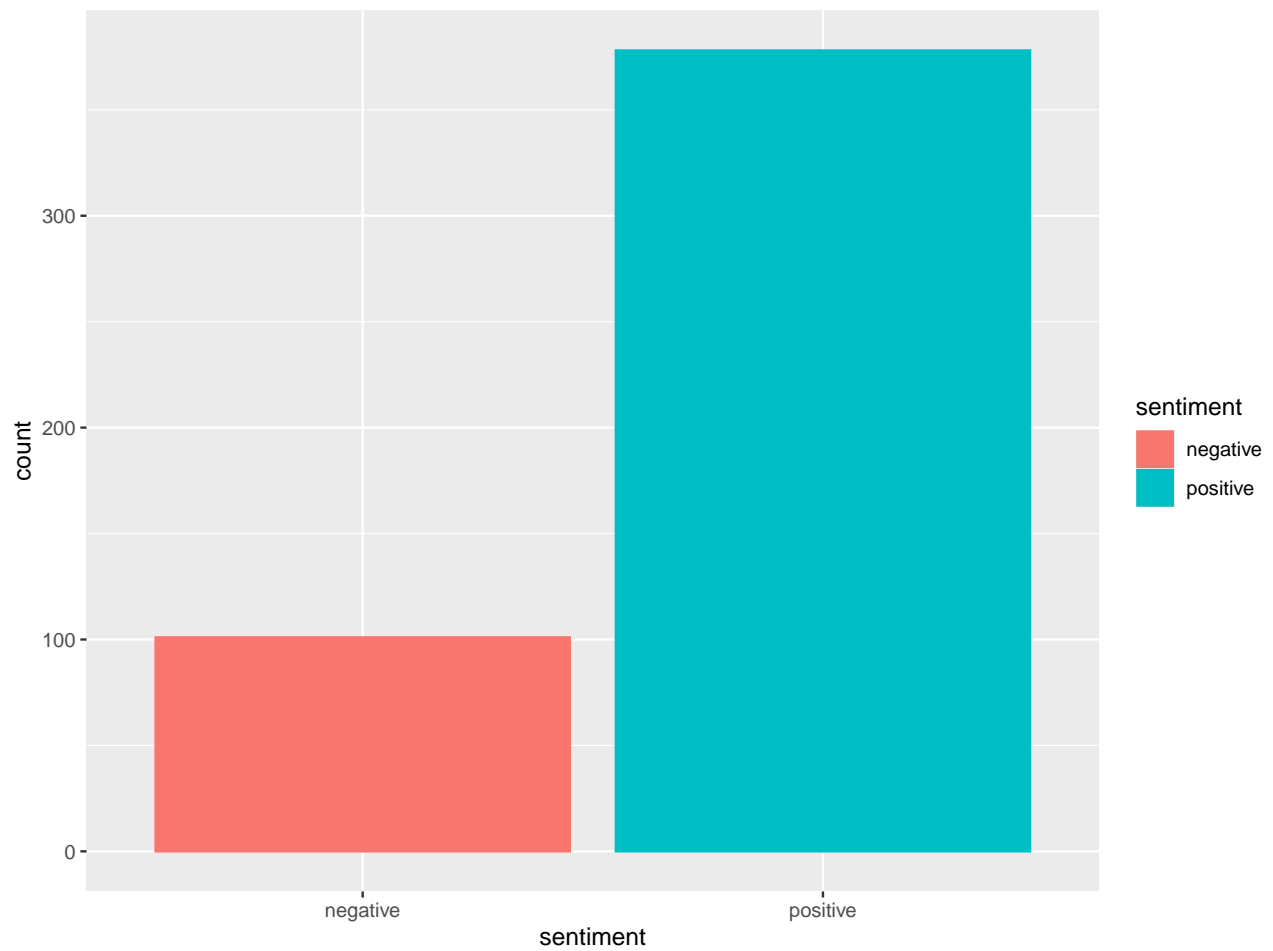
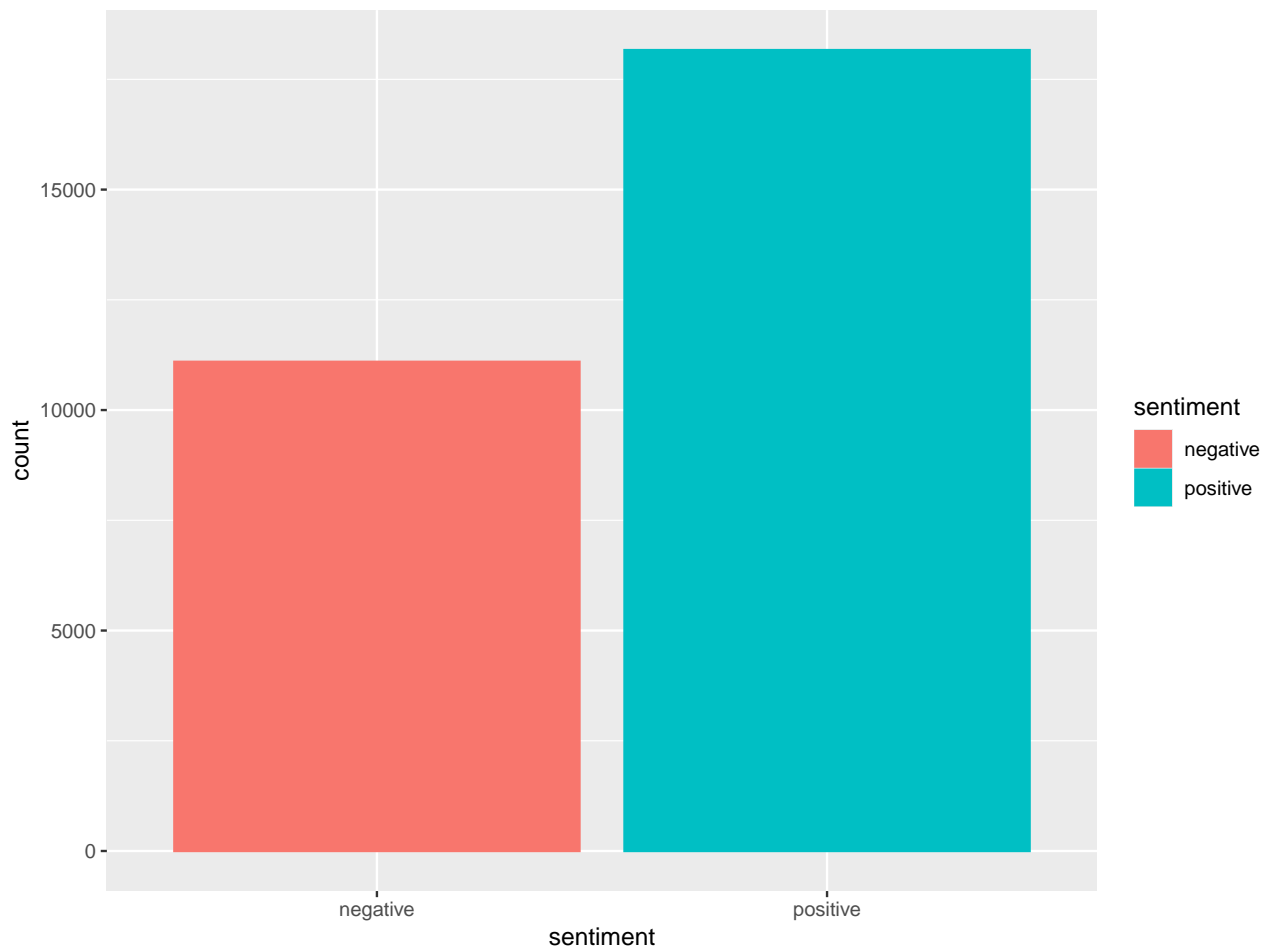


Figure 5: Positive vs Negative reviews - Critics

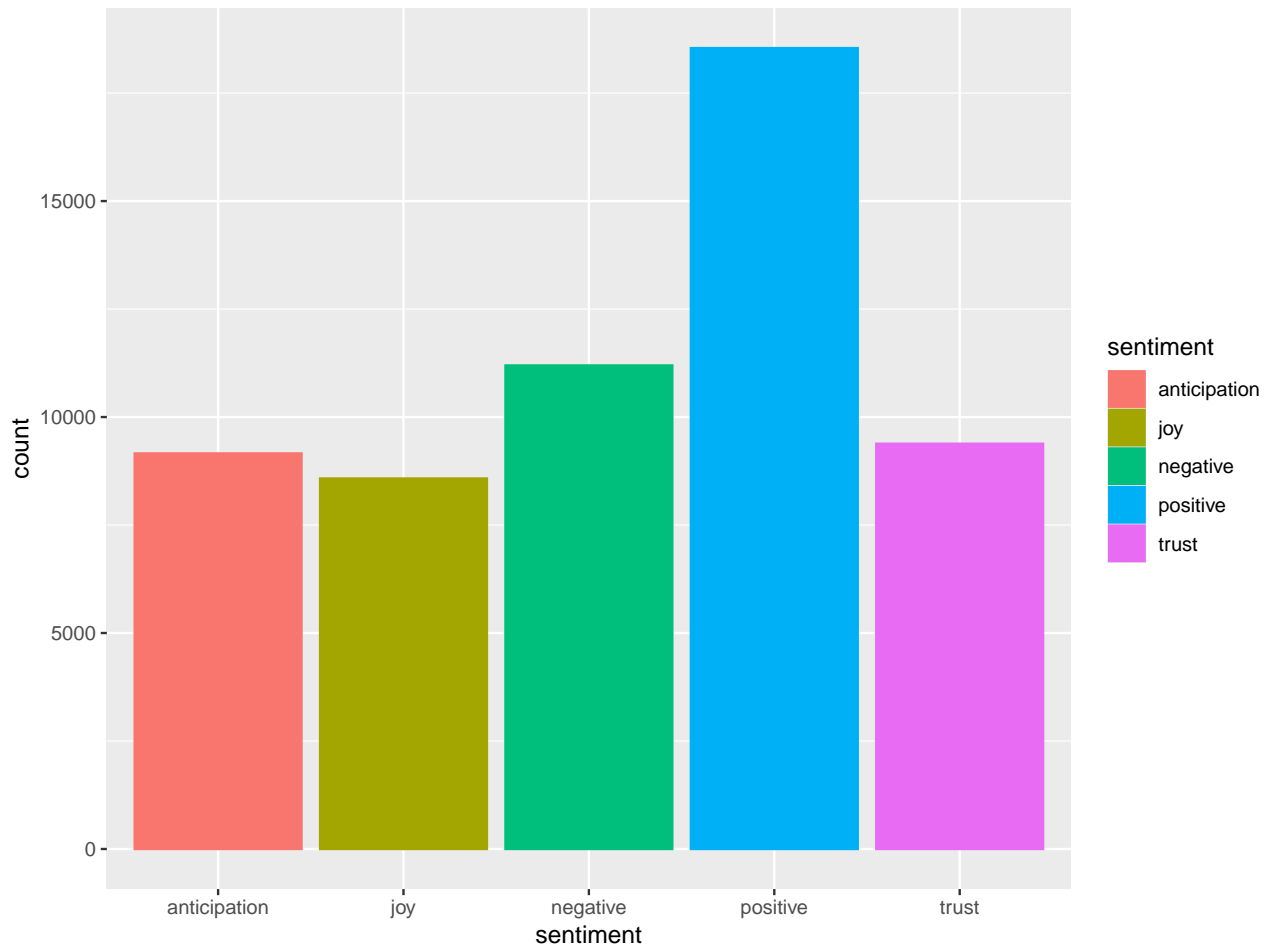


Expectedly, gamers give more negative reviews to Animal Crossing, as compared to critics.

```
# Add the sentiment values to relevant words of
# critic_words
user_words_nrc <- user_words_no_stop %>%
  left_join(nrc, by = "word")

stack_critic <- critic_words_nrc %>%
  rename(user_name = publication)

user_words_nrc %>%
  bind_rows(stack_critic) %>%
  na.omit() %>%
  filter(sentiment %in% c("positive", "negative",
    "anticipation", "trust", "joy")) %>%
  ggplot(aes(sentiment, fill = sentiment, color = sentiment)) +
  geom_bar()
```



On account with the type of sentiment that is most prominent, it appears that most are positive. This is created through mapping tokens to particular words that belong to a sentiment category determined by NRC.

```
sentiment_words <- user_words_nrc %>%
  na.omit() %>%
  group_by(sentiment) %>%
  summarise(mean_score = mean(grade)) %>%
  arrange(-mean_score)

row_indices <- which(sentiment_words$sentiment %in%
  c("fear", "sadness", "negative", "anger", "disgust"))

sentiment_words %>%
  kable() %>%
  kable_styling() %>%
  row_spec(row_indices, background = "#ADD8E6") %>%
  row_spec(c(1:5), background = "#90EE90")
```

## Conclusion

Despite the critical scores that gamers may have given to the game, it should be noted that through a detailed sentiment analysis of Animal Crossing, a majority of gamers actually write with positivity in their reviews - much like critics have done. As they say: "Don't judge a book by its cover." In spite of how review scores of Animal Crossing may look (given with how ruthless players can be when rating this game), it is important to realise that these people are still using positive words in their reviews!



sentiment	mean_score
joy	4.149667
anticipation	4.067761
trust	4.047873
surprise	4.013894
positive	3.991576
fear	3.442298
sadness	3.297095
negative	3.162610
anger	3.069593
disgust	2.833437

Note, that “Animal Crossing: New Horizons” achieved exceptionally high sales figures, especially considering its release timing coincided with the onset of the COVID-19 pandemic. Many people turned to the game as a form of escapism during lockdowns and social distancing measures. Overall, “Animal Crossing: New Horizons” was considered a commercial and critical success, resonating strongly with players - in spite of the reviews on metacritic.

**Final Remarks** To improve the analysis:

- More sentiment word\_lists should be tested
- Exploring phrases, instead of breaking words into individual components may be beneficial for gaining the context behind words. (for example: bi-grams)
- Metacritic does not represent all the opinions of Animal Crossing. More sources of information, such as YouTube or Twitter could be scraped to have a more robust dataset.
- The timeframe of the dataset is quite narrow. It would be interesting to see the longer term sentiments surrounding the game.

Finally, it should be noted that the analysis conducted has its limitations. For example, with the afinn list of sentiment words cannot recognise context. Therefore, some sentiment values may be represented in the analysis.