

## **Topic: Sentimental Analysis for Amazon Reviews**

### **What is Sentiment Analysis?**

- Sentiment analysis is a natural language processing technique to determine whether the data is positive, negative or neutral.
- It is performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand the customer needs.
- Sentiment analysis models focus on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, etc.), urgency (urgent, not urgent) and even intentions (interested v. not interested).

### **Dataset Description**

The dataset has been created using Amazon Reviews Exporter which is a google chrome extension. This extension helps to export Amazon Product Reviews to CSV format. Our Dataset contains 6 Attributes and 1735 Rows. We are analysing the reviews of a book based on a Japanese concept named - IKIGAI.

- id (type: chr): "R353MEFFMHYAY8" "R57Y9694P00OQ" "R1Y6LYA3EVVHII" "R2UBD10GP97TLL" ...
- profileName (type: chr): chr "Short of a Century" "Seethalakshmikshetty" "LibroReview" "Radhika Saimbi"
- text (type: chr): "\n The book does a decent job of relating the concept of Ikigai to modern day psychology (with Frankl's Logoth"| \_\_truncated\_\_ "\n Just read a back of book and it's enough dont waste like međŸ\230¢\n" "\n I personally believe that having a purpose on in life and then giving it your all is the most important to "| \_\_truncated\_\_ "\n IkigaiThe Japanese Secret to a Long and Happy LifeBy Hector Garcia and Francesc MirallesI will be confessin"| \_\_truncated\_\_ ...
- date (type: chr): "Reviewed in India on 12 October 2018" "Reviewed in India on 14 October 2019" "Reviewed in India on 8 April 2019" "Reviewed in India on 22 July 2019"...
- title (type: chr): "Simple & Light Reading. May Disappoint - Dependent on Reader's Expectation." "Ikigai" "The best book to read during your break time." "Little hyped!!"
- rating (type: int): 3 2 4 3 1 4 1 5 1 4

### **Aims and Objectives:**

- Product reviews are becoming more important with the evolution of online shopping. Customers are posting reviews directly on product pages in real time.

- Customers are more inclined towards the reviews to buy a particular product. So, analysing the data from those customer reviews to make the data more dynamic is essential nowadays.
- Hence, our project aims at performing Sentiment Analysis for Amazon Product Reviews.
- We are analysing the reviews of a book based on a Japanese concept named - IKIGAI.

## Analysis:

### *Import Libraries:*

```
> library(tm) # Text analytics - text mining
> library(wordcloud) #Create word cloud
> library(syuzhet) #for sentiment scores and emotion classification
> library(ggplot2) #for plotting graphs
```

### *Import and Explore dataset*

```
> # Import data into R
> reviews <- read.csv(file.choose(), header = T)
> # Check the structure of the file
> str(reviews)
```

```
> str(reviews)
'data.frame': 1735 obs. of 6 variables:
 $ id      : chr  "R353MEFFMHYAY8" "R57Y9694P000Q" "R1Y6LYA3EYVHII" "R2UBD10GP97TLL"
 ...
 $ profileName: chr  "Short of a Century" "Seethalakshmikshetty" "LibroReview" "Radhika Sa
imbi" ...
 $ text       : chr  "\n The book does a decent job of relating the concept of Ikigai to
modern day psychology (with Frankl's Logoth"| __truncated__ "\n Just read a back of book
and its enough dont waste like meðŸ\230c\n" "\n I personally believe that having a purpo
se on in life and then giving it your all is the most important to "| __truncated__ "\n I
kigaiThe Japanese Secret to a Long and Happy LifeBy Hector Garcia and Francesc MirallesI w
ill be confessin"| __truncated__ ...
 $ date      : chr  "Reviewed in India on 12 October 2018" "Reviewed in India on 14 Octob
er 2019" "Reviewed in India on 8 April 2019" "Reviewed in India on 22 July 2019" ...
 $ title     : chr  "Simple & Light Reading. May Disappoint - Dependent on Reader's Expec
tation." "Ikigai" "The best book to read during your break time." "Little hyped!!" ...
 $ rating    : int  3 2 4 3 1 4 1 5 1 4 ...
> |
```

### *Create Corpus*

```
> #Create a corpus of text column in the dataset.
> corpus <- iconv(reviews$text)
> corpus <- Corpus(VectorSource(corpus))
```

### *To see the corpus*

```
> inspect(corpus[1:5])
```

```
> # Create Corpus
> corpus <- iconv(reviews$text)
> corpus <- Corpus(VectorSource(corpus))
> # To see the corpus
> inspect(corpus[1:5])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 5
```

```
[1] \n The book does a decent job of relating the concept of Ikigai to modern day psychology (with Frankl's Logotherapy from Man's Search for Meaning among others) and a few scientific references in a simple manner. It talks about how purpose plays an important role in a man's life and the different ways in which it manifests itself. It also tackles some ways to 'find your flow' and ensure that what you do receives 100% of your attention and that you enjoy whatever you are creating. The book also discusses certain other Japanese concepts like takumi (specialized workers) and moai (connections with community or friend-circle). The brief discussions have the benefit of being to the point and simple but also pose the risk of trivializing them into regular self-help advice. The book also delves into Japanese perspectives on living life and persevering without getting caught up in artificially-created urgency. But again, maybe the authors wished for the readers to research more or meditate more on the content given the concise treatment of the same. The chapters on diet and exercises have more details and thus, may be more useful. Certain foods are dealt with in greater detail as is the concept of 'hara hachi bu' wherein one eats only 80% of what would actually assuage his hunger. The chapter on exercises includes illustrations and steps. While they may suffice for some of the purposes mentioned in the book - the philosophy behind them, progressive increments and other essential details are missing or insufficient. I was interested in the concept of Ikigai and wanted to read more about it. Despite being well-written and presenting modern applications, the book did not fulfill my requirements at all. Some of the condensed content made me think that the extended research, including on-site interviews, done by the authors for writing this book was clearly lost in translation or presentation in certain parts. The hardcover is pretty and soothing with its matte texture. Inner pages are smooth and heavy with a cream tinge. The font size is good. The spine as well as pages hold up well. Overall, the book is quite light and sturdy. Bought it for INR 460 against MRP of INR 499.\n
[2] \n Just read a back of book and its enough dont waste like meðŸˆˆ\n
```

### *Cleaning the Corpus*

Clean the corpus:

1. Make all text lowercase
2. Remove punctuation
3. Remove numbers
4. Remove common repeated words
5. Remove white spaces

```
> # Cleaning Corpus
> corpus <- tm_map(corpus, tolower)
> corpus <- tm_map(corpus, removePunctuation)
> corpus <- tm_map(corpus, removeNumbers)
> corpus <- tm_map(corpus, removeWords, stopwords('english'))
> corpus <- tm_map(corpus, removeWords, c("book", "read", "life", "will", "the", "and", "this", "for", "you", "that"))
> corpus <- tm_map(corpus, stripWhitespace)
> inspect(corpus[1:5])
> reviews_final <- corpus
```

```
> inspect(corpus[1:5])
```

```
<<SimpleCorpus>>
```

```
Metadata: corpus specific: 1, document level (indexed): 0
```

```
Content: documents: 5
```

```
[1] \n the book does a decent job of relating the concept of ikigai to modern day psychology with frankls logotherapy from mans search for meaning among others and a few scientific references in a simple manner it talks about how purpose plays an important role in a mans life and the different ways in which it manifests itself it also tackles some ways to find your flow and ensure that what you do receives of your attention and that you enjoy whatever you are creatingthe book also discusses certain other japanese concepts like takumi specialized workers and moai connections with community or friendcircle the brief discussions have the benefit of being to the point and simple but also pose the risk of trivializing them into regular selfhelp advice the book also delves into japanese perspectives on living life and persevering without getting caught up in artificiallycreated urgency but again maybe the authors wished for the readers to research more or meditate more on the content given the concise treatment of the samethe chapters on diet and exercises have more details and thus may be more useful certain foods are dealt with in greater detail as is the concept of hara hachi bu wherein one eats only of what would actually assuage his hunger the chapter on exercises includes illustrations and steps while they may suffice for some of the purposes mentioned in the book the philosophy behind them progressive increments and other essential details are missing or insufficienti was interested in the concept of ikigai and wanted to read more about it despite being wellwritten and presenting modern applications the book did not fulfill my requirements at all some of the condensed content made me think that the extended research including on-site interviews done by the authors for writing this book was clearly lost in translation or presentation in certain partsthe hardcover is pretty and soothing with its matte texture inner pages are smooth and heavy with a cream tinge the font size is good the spine as well as pages hold up well overall the book is quite light and sturdy bought it for inr against mrp of inr \n [2] \n just read a back of book and its enough dont waste like meðy~c\n
```

```
[3] \n personally believe purpose life giving important lead happy life book validates soikigai short short book based japanese concept according concept find deeply sown purpose lives within defining passion mission vocation profession gives tips simplest things life like sometimes taking pause will give us long happy life fei bought hardcover itâ€™s extremely pretty cover cover will soothe mind whenever look title go book personally thought wouldif expecting help find ikigai give step step solution can find reward millions dollars youâ€™ll disappointed book full facts real life experience japanâ€™s okinawa compels focus health choices focus inner happiness will probably book youâ€™ll come across tells take relax also burn passionthe language simplistic beautiful cover itâ€™s content many tough words beginnerfriendly reading world structure great prologue chapters epilogue list suggestions explore endcoming overall feel book itâ€™s cozy itâ€™s way also help slightly find ikigaii thought full hustle kind book felt relaxed read ready take tasks stressfree though something extremely crazy different definitely onetime read hustlers nonhustlers think punish get want\n
```

```
[4] \n ikigai the japanese secret long happy lifeby hector garcia francesc miralles i will confessing saying bought book just attractive cover soothing color combination calmed methis book finding passion life â€œikigaiâ€\u009d will help live happy content active long life talks art staying young growing old relaxing running behind materialistic things improve performance daily routine stay focusedone best lesson liked anti fragile talked creating options devoting time passion taking small risks increasing experience get rid things makes us fragilebefore reading book felt ma oeuvre follow finding ikigai just overview many therapies left little unsatisfied disappointed things like proverbs quotes traditions centritains found really required rocket science written book facts eye opening s seriously wish learn sun salutation exercise readingoverall will say itâ€™s easy simple read beginners setting small goalsonly staying active will make want live hundred yearsâ€”â€”japanese proverbbook starts proverb literally summarises book\n
```

```
[5] \n book good read seller definitely duping readers book neither hardcover authentic print cheap copy can bought roadside\n
```

### *Data Visualisation*

```
> # Create Term Document  
> TextDoc_dtm <- TermDocumentMatrix(reviews_final)  
> dtm_m <- as.matrix(TextDoc_dtm)  
> dtm_m[1:10, 1:5]
```

```
> dtm_m[1:10, 1:5]  
      Docs  
Terms 1 2 3 4 5  
actually 1 0 0 0 0  
advice 1 0 0 0 0  
also 4 0 2 0 0  
among 1 0 0 0 0  
applications 1 0 0 0 0  
artificiallycreated 1 0 0 0 0  
assuage 1 0 0 0 0  
attention 1 0 0 0 0  
authors 2 0 0 0 0  
behind 1 0 0 1 0
```

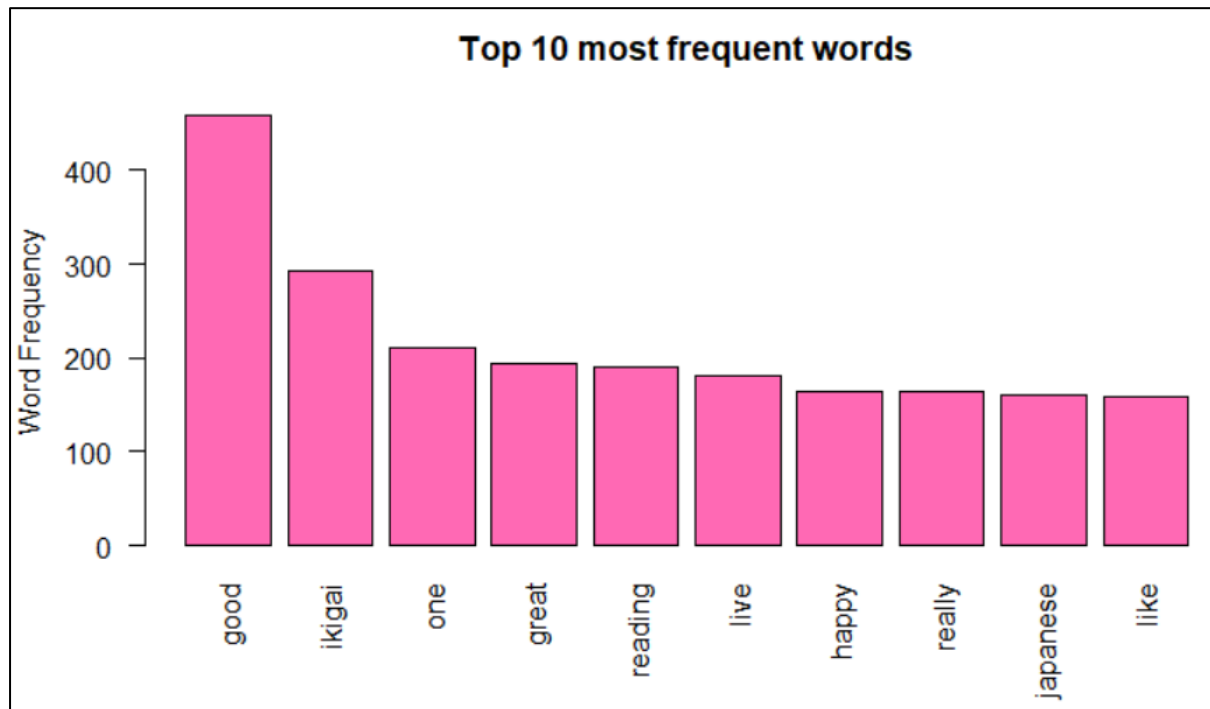
After cleaning the text data, the next step is to count the occurrence of each word, to identify popular or trending words. Using the function `TermDocumentMatrix()` from the text mining package, we have built a Document Matrix – a table containing the frequency of words. As you can see here are the top 10 most frequently found words in the review. The following table of word frequency is the expected output of the head command.

```
> # Sort by descending value of frequency  
> dtm_v <- sort(rowSums(dtm_m),decreasing=TRUE)  
> dtm_d <- data.frame(word = names(dtm_v),freq=dtm_v)  
> # Display the top 10 most frequent words  
> head(dtm_d, 10)
```

	word	freq
good	good	458
ikigai	ikigai	292
one	one	211
great	great	194
reading	reading	191
live	live	180
happy	happy	164
really	really	164
japanese	japanese	161
like	like	158

```
> # Plot the most frequent words  
> barplot(dtm_d[1:10,]$freq,
```

```
+ las = 2,  
+ names.arg = dtm_d[1:10,]$word,  
+ col = "hotpink",  
+ main = "Top 10 most frequent words",  
+ ylab = "Word Frequency")
```

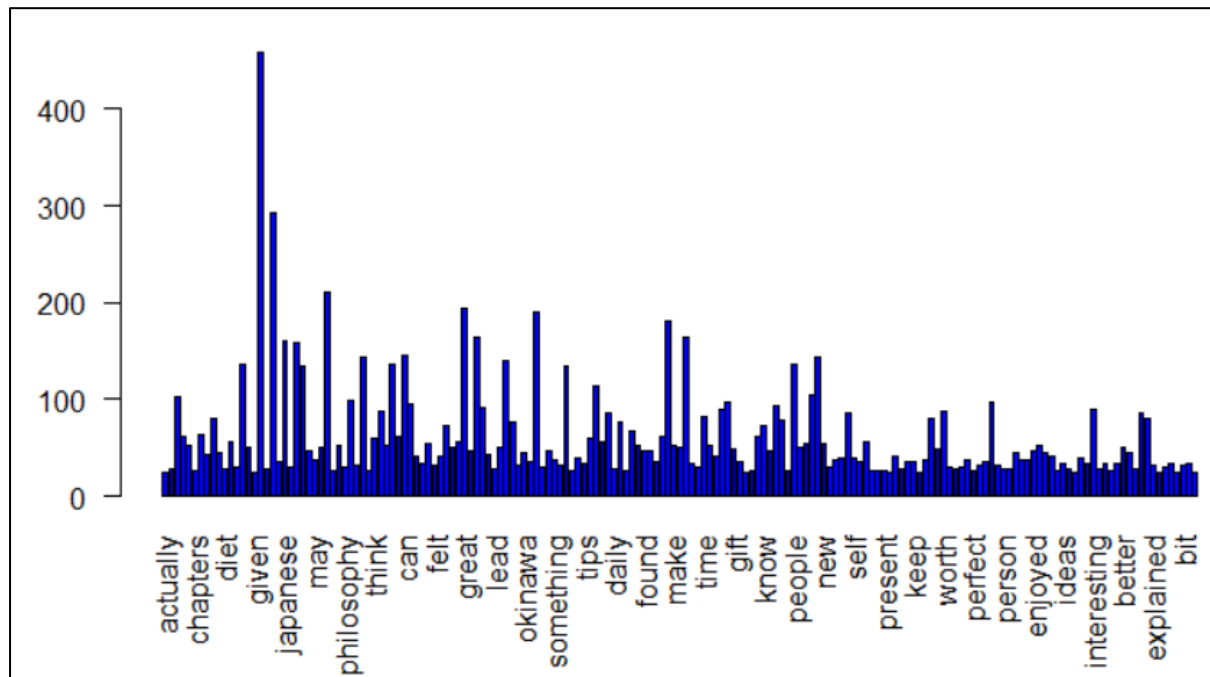


Plotting the top 10 most frequent words using a bar chart is a good basic way to visualize this word frequency data. So, I could interpret the following from this bar chart:

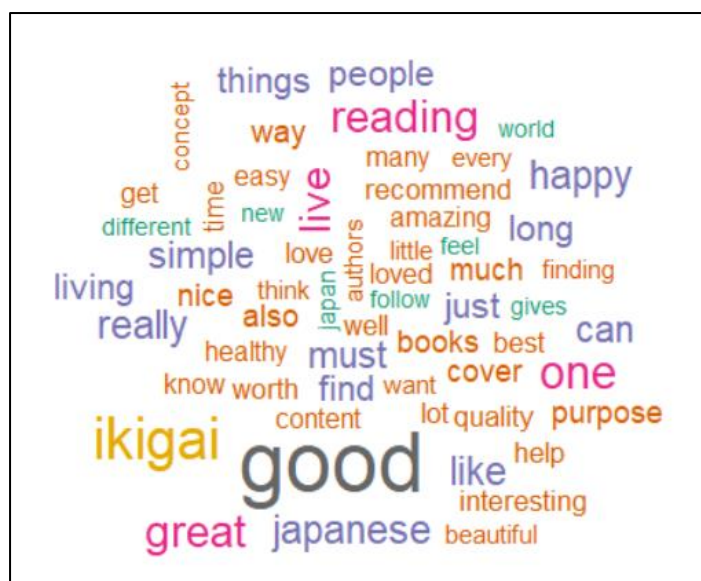
- The most frequently occurring word is “good”. Also notice that negative words like “not” don’t feature in the bar chart, which indicates there are no negative prefixes to change the context or meaning of the word “good” (In short, this indicates most responses don’t mention negative phrases like “not good”).
- Then the 2nd most frequent word used is IKIGAI which itself is the name of the book and the book is great.
- “Reading”, “happy”, “really” and “japanese” are the next four most frequently occurring words, which indicate that most of the people are really feeling happy by reading about the japanese book.
- Finally, the word like is also on the chart, which says that the readers liked the book in the positive context.

```
> # Bar plot  
> w <- rowSums(dtm_m)  
> w <- subset(w, w>=25)  
> barplot(w, las = 2, col = "blue")
```





```
> # Create Word cloud
> w <- sort(rowSums(dtm_m), decreasing = T)
> set.seed(2000)
> wordcloud(words = names(w),
+           freq = w,
+           max.words = 60,
+           random.order = T,
+           min.freq = 5,
+           colors = brewer.pal(25, 'Dark2'),
+           scale = c(3, 0.6))
```



A word cloud is one of the most popular ways to visualize and analyse qualitative data. It's an image composed of keywords found within a body of text, where the size of each word indicates its frequency in that body of text. Use the word frequency data frame (that table) created previously to generate the word cloud. The word cloud shows additional words that occur frequently and could be of interest for further analysis. Words like “content”, “quality”, “worth” (root for “worth-it” or “worthless”, etc. could provide more context around the most frequently occurring words and help to gain a better understanding of the main insight.

> # Find associations

> findAssocs(TextDoc\_dtm, terms = c("recommend", "ikigai", "happy", "people"), corlimit = 0.25)

\$recommend									
highly									
0.32									
\$ikigai									
find		concept		finding		activities			
0.39		0.33		0.29		0.29			
includes		different		inside		pulled			
0.28		0.27		0.27		0.27			
highlevel		long		still		bowl			
0.27		0.26		0.26		0.26			
decode		grab		gushing		limits			
0.26		0.26		0.26		0.26			
optimal		outdoor		selfdiscovery		showcasing			
0.26		0.26		0.26		0.26			
supercentenariansabove		worldevery		discovers		itso			
0.26		0.26		0.26		0.26			
meaning”-		notifies		puigcerverthe		refine			
0.26		0.26		0.26		0.26			
andres		components		concerning		diagramm			
0.26		0.26		0.26		0.26			
diagrammits		generalto		gorgeousnice		itnow			
0.26		0.26		0.26		0.26			
pictured		taint		therefore		youbut			
0.26		0.26		0.26		0.26			
zuzunaga		also		particular					
0.26		0.25		0.25					

\$happy											
long		secret		rocket		amazonians		balk		bulking	
0.45		0.31		0.28		0.27		0.27		0.27	
endless		etcif		foolish		futurefor		gleaning		hobbies	
0.27		0.27		0.27		0.27		0.27		0.27	
impart		infinitesimally		marriage		minded		noticeably		passes	
0.27		0.27		0.27		0.27		0.27		0.27	
payment		reinforce		sciencegood		scottish		sifting		views	
0.27		0.27		0.27		0.27		0.27		0.27	
waffle		web		wifei		things		active		smile	
0.27		0.27		0.27		0.26		0.26		0.26	

\$people																	
island		live		okinawa		cultivated		figure		involves		somebodys		japan		enjoy	
0.34		0.32		0.31		0.30		0.30		0.30		0.30		0.28		0.26	

Correlation is a statistical technique that can demonstrate whether, and how strongly, pairs of variables are related. This technique can be used effectively to analyze which words occur most often in association with the most frequently occurring words in the review of the book, which helps to see the context around these words.



This script shows which words are most frequently associated with the top 4 terms (corlimit = 0.25 is the lower limit/threshold I have set. We can set it lower to see more words, or higher to see less). The output indicates that “highly” occurs 32% of the time with the word “recommend”. We can interpret this as the context around the most frequently occurring word (“recommend”) is positive. Similarly, the name of the book “ikigai” is highly correlated with the “concept”, “finding” and “activities”. This indicates that most responses are saying that the book “includes optimal and conceptual findings” and can be interpreted in a positive context. Also, the output indicates that “happy” is strongly associated with the “long”, “secret” and “exciting”. This shows that most of the responses say that the book “has the secret of long and exciting life”. Also “island” occurs 34% and “japan” occurs 28% of the time with the word “people”. We can interpret this as the context as “most of the people described in the book lived on an island which is in japan”.

```
> # Obtain Sentiment Scores  
> sentiment_data <- iconv(reviews$text)  
> s <- get_nrc_sentiment(sentiment_data)  
> s[1:10,]
```

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	1	7	1	3	9	3	3	11	6	24
2	0	0	1	0	0	0	0	0	1	0
3	3	14	1	2	9	3	1	13	6	22
4	2	10	2	1	9	4	3	12	4	14
5	0	1	0	0	2	0	1	2	2	2
6	0	0	0	0	1	0	0	2	0	2
7	0	0	0	0	0	0	0	1	0	1
8	0	2	0	0	3	0	0	5	0	3
9	0	2	0	0	3	0	2	4	0	4
10	1	3	1	0	2	2	0	4	2	11

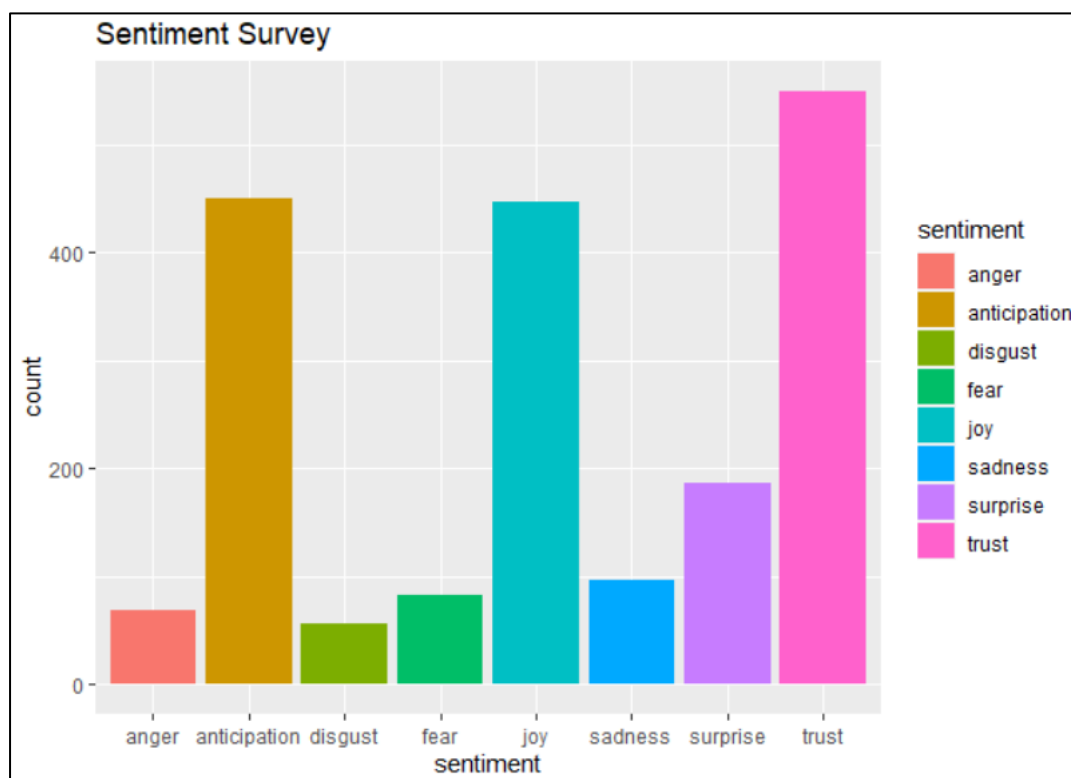
*Sentiments can be classified as positive, neutral or negative. They can also be represented on a numeric scale, to better express the degree of positive or negative strength of the sentiment contained in a body of text.*

Emotion classification is built on the NRC Word-Emotion Association Lexicon. The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). To understand this, the `get_nrc_sentiments` function, which returns a data frame with each row representing a sentence from the original file. The data frame has ten columns (one column for each of the eight emotions, one column for positive sentiment valence and one for negative sentiment valence). The data in the columns (anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative, positive) can be accessed individually. The output shows that the first line of text has;

- Least occurrences of words associated with emotions of anger, disgust, fear, sadness and negative sentiment
- Highest occurrences of words associated with emotions of joy, trust and positive sentiment
- Total of 6 occurrence of words associated with negative sentiment
- Total of 24 occurrences of words associated with positive sentiment

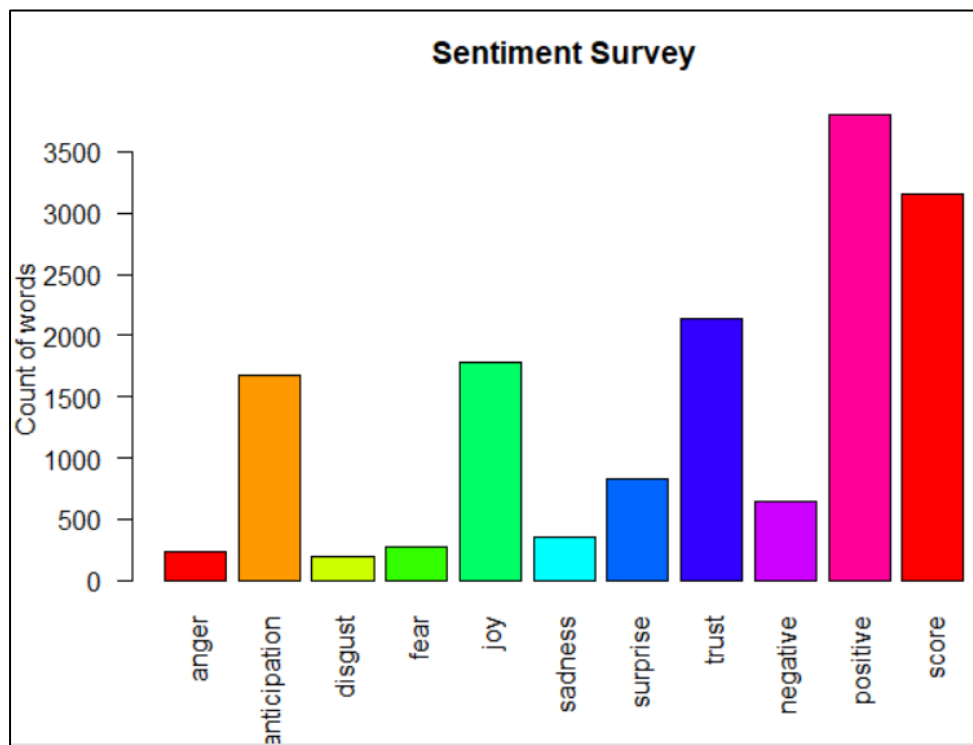
The next step is to create two plots charts to help visually analyze the emotions in this review text. First, perform some data transformation and clean-up steps before plotting charts. The first plot shows the total number of instances of words in the text, associated with each of the eight emotions.

```
> #transpose
> td<-data.frame(t(s))
> #The function rowSums computes column sums across rows for each level of a grouping variable.
> td_new <- data.frame(rowSums(td[2:253]))
> #Transformation and cleaning
> names(td_new)[1] <- "count"
> td_new <- cbind("sentiment" = rownames(td_new), td_new)
> rownames(td_new) <- NULL
> td_new2<-td_new[1:8,]
> #Plot One - count of words associated with each sentiment
> quickplot(sentiment, data=td_new2, weight=count, geom="bar", fill=sentiment, ylab="count") + ggtitle("Sentiment Survey")
```



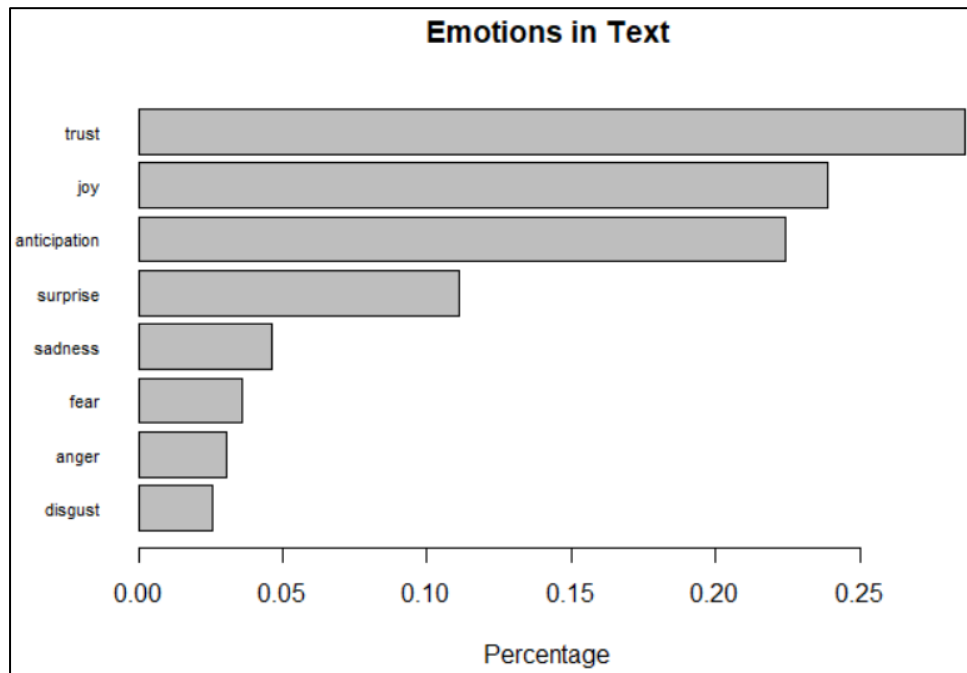
This bar chart demonstrates that words associated with the positive emotion of “trust” occurred about five hundred times in the text, whereas words associated with the negative emotion of “disgust” occurred less than 25 times. A deeper understanding of the overall emotions occurring in the review response can be gained by comparing these numbers as a percentage of the total number of meaningful words.

```
> # Simple Bar plot  
> barplot(colSums(s),  
+       las = 2,  
+       col = rainbow(10),  
+       ylab = 'Count of words',  
+       main = 'Sentiment Survey')
```



```
> #Plot two - count of words associated with each sentiment, expressed as a percentage  
> barplot(  
+   sort(colSums(prop.table(s[, 1:8]))),  
+   horiz = TRUE,  
+   cex.names = 0.7,  
+   las = 1,  
+   main = "Emotions in Text", xlab="Percentage"  
+ )
```

This bar plot allows for a quick and easy comparison of the proportion of words associated with each emotion in the text. The emotion “trust” has the longest bar and shows that words associated with this positive emotion constitute just over 35% of all the meaningful words in this text. On the other hand, the emotion of “disgust” has the shortest bar and shows that words associated with this negative emotion constitute less than 2% of all the meaningful words in this text. Overall, words associated with the positive emotions of “trust” and “joy” account for almost 60% of the meaningful words in the text, which can be interpreted as a good sign of positive review of the book.



```
> # Calculate Review wise Scores
> s$score <- s$positive - s$negative
> s[1:10,]
```

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive	score
1	1	7	1	3	9	3	3	11	6	24	18
2	0	0	1	0	0	0	0	0	1	0	-1
3	3	14	1	2	9	3	1	13	6	22	16
4	2	10	2	1	9	4	3	12	4	14	10
5	0	1	0	0	2	0	1	2	2	2	0
6	0	0	0	0	1	0	0	2	0	2	2
7	0	0	0	0	0	0	0	1	0	1	1
8	0	2	0	0	3	0	0	5	0	3	3
9	0	2	0	0	3	0	2	4	0	4	4
10	1	3	1	0	2	2	0	4	2	11	9

```
> # Check overall sentiment of the product
> review_score <- colSums(s[,])
> print(review_score)
```

anger	anticipation	disgust	fear	joy	sadness	surprise
229	1676	192	270	1786	347	829
trust	negative	positive	score			
2141	641	3801	3160			

This is the overall review score of the following sentiments which shows that the lowest sentiment score is 229 of anger and highest score is 3801 of positive. Which clearly states that the book ikigai has more positive reviews than the negative ones. And people have rated the book in a positive and effective context.

## **Conclusion:**

1. In our project, we have created a word frequency table and plotted a word cloud, to identify prominent insights in the review.
2. Word association analysis using correlation, explored four methods to generate sentiment scores, which proved useful in assigning a numeric value to strength (of positivity or negativity) of sentiments in the text and allowed interpreting that the average sentiment through the text is trending positive.
3. Lastly, our project demonstrates how to implement an emotion classification with NRC sentiment and creates plots to analyse and interpret emotions found in the text.
4. Now by performing the exploratory sentimental analysis on Amazon book review we have found out that the Japanese book ikigai has more positive and productive reviews
5. Overall, the book is rated in a good context.