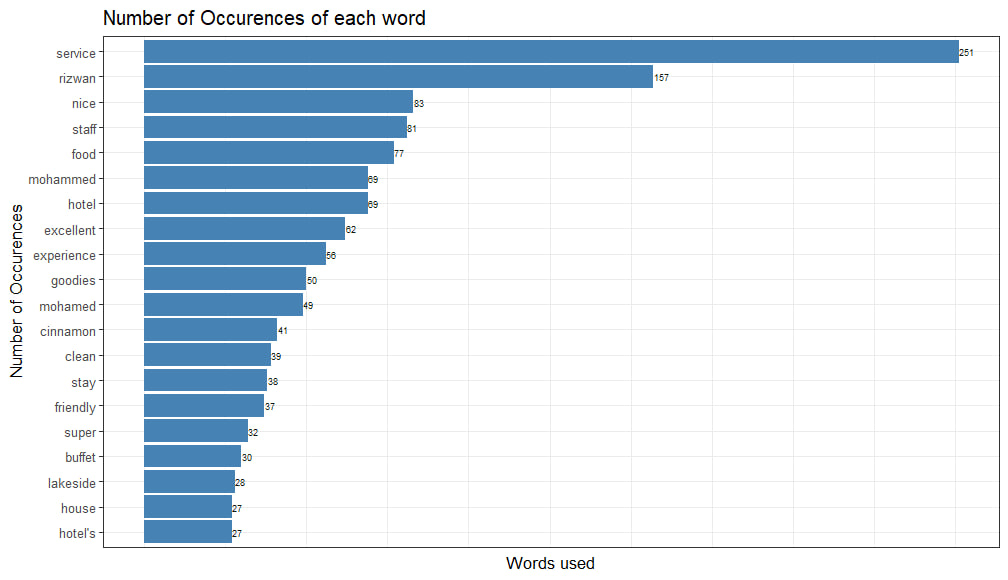
|  |
| --- |
| R version 4.3.2 (2023-10-31 ucrt) -- "Eye Holes"  Copyright (C) 2023 The R Foundation for Statistical Computing  Platform: x86\_64-w64-mingw32/x64 (64-bit)  R is free software and comes with ABSOLUTELY NO WARRANTY.  You are welcome to redistribute it under certain conditions.  Type 'license()' or 'licence()' for distribution details.  Natural language support but running in an English locale  R is a collaborative project with many contributors.  Type 'contributors()' for more information and  'citation()' on how to cite R or R packages in publications.  Type 'demo()' for some demos, 'help()' for on-line help, or  'help.start()' for an HTML browser interface to help.  Type 'q()' to quit R.  [Workspace loaded from ~/.RData]  > library(tidyverse)  ── **Attaching core tidyverse packages** ────────────────────────────────────────── tidyverse 2.0.0 ──  ✔ dplyr 1.1.4 ✔ readr 2.1.4  ✔ forcats 1.0.0 ✔ stringr 1.5.1  ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  ✔ purrr 1.0.2  ── **Conflicts** ──────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  ✖ dplyr::filter() masks stats::filter()  ✖ dplyr::lag() masks stats::lag()  ℹ Use the conflicted package to force all conflicts to become errors  > library(tidytext)  Warning message:  package ‘tidytext’ was built under R version 4.3.3  > reviews <- read\_csv("C:/Users/Prarthana/OneDrive - General Sir John Kotelawala Defence University/5th sem/Marketing/lexicons/Outscraper-cinnamon")  Error: 'C:/Users/Prarthana/OneDrive - General Sir John Kotelawala Defence University/5th sem/Marketing/lexicons/Outscraper-cinnamon' does not exist.  > reviews <- read\_csv("C:/Users/Prarthana/OneDrive - General Sir John Kotelawala Defence University/5th sem/Marketing/lexicons/Outscraper-cinnamon.csv")  **Rows:** 408 **Columns:** 1  ── **Column specification** ──────────────────────────────────────────────────────────────────────────  **Delimiter:** ","  chr (1): review\_text  ℹ Use `spec()` to retrieve the full column specification for this data.  ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  > names(reviews)  [1] "review\_text"  > review\_words<-reviews%>%  + unnest\_tokens(word,review\_text)  > head(stop\_words)  # A tibble: 6 × 2  word lexicon  *<chr>* *<chr>*  1 a SMART  2 a's SMART  3 able SMART  4 about SMART  5 above SMART  6 according SMART  > review\_sentiment<-review\_words%>%  + anti\_join(stop\_words,by="word")  > review\_sentiment%>%count(word,sort=T)%>%top\_n(20)  Selecting by n  # A tibble: 20 × 2  word n  *<chr>* *<int>*  1 service 251  2 rizwan 157  3 nice 83  4 staff 81  5 food 77  6 hotel 69  7 mohammed 69  8 excellent 62  9 experience 56  10 goodies 50  11 mohamed 49  12 cinnamon 41  13 clean 39  14 stay 38  15 friendly 37  16 super 32  17 buffet 30  18 lakeside 28  19 hotel's 27  20 house 27  > library(ggplot2)  > ggplot(review\_sentiment%>%count(word,sort=T)%>%top\_n(20),aes(reorder(word,n),n))+  + geom\_bar(stat = "identity", fill = "steelblue")+  + geom\_text(aes(label = n),color="#0f190f", hjust = -0.05, size = 2)+  + theme\_bw()+  + coord\_flip()+  + xlab("Number of Occurences")+  + ylab("Words used")+  + ggtitle("Number of Occurences of each word")+  + theme(axis.text.x = element\_blank(),  + axis.ticks.x = element\_blank())  Selecting by n  > review\_bing<-review\_sentiment%>%  + inner\_join(get\_sentiments("bing"),by="word")%>%  + ungroup()  > head(review\_bing, 10)  # A tibble: 10 × 2  word sentiment  *<chr>* *<chr>*  1 fantastic positive  2 variety positive  3 recommended positive  4 perfect positive  5 perfect positive  6 amazing positive  7 fantastic positive  8 excellent positive  9 clean positive  10 clean positive  > review\_bing\_sentiment\_freq <- review\_bing %>%  + count(sentiment)  > review\_bing\_sentiment\_freq  # A tibble: 2 × 2  sentiment n  *<chr>* *<int>*  1 negative 162  2 positive 742  > review\_bing %>%  + count(word, sentiment) %>%  + group\_by(sentiment) %>%  + top\_n(10, n) %>%  + ungroup() %>%  + ggplot(aes(x = reorder(word, n), y = n, fill = sentiment)) +  + geom\_col(show.legend = FALSE) +  + coord\_flip() +  + facet\_wrap(~ sentiment, scales = "free") +  + labs(x = "Number of Occurrences", y = "Words", title = "Top 10 Positive and Negative Sentiment Words Used in Headlines (using Bing lexicon)") +  + theme(plot.title = element\_text(size = 8, face = "bold")) +  + geom\_text(aes(label = n), hjust = --1.3, size = 3, color = "black")  > review\_afinn <- review\_sentiment %>%  + inner\_join(get\_sentiments("afinn"), by = "word") %>%  + mutate(sentiment = ifelse(value < 0, "Negative", "Positive"))  Error: The textdata package is required to download the AFINN lexicon.  Install the textdata package to access this dataset.  > install.packages("textdata")  WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:  https://cran.rstudio.com/bin/windows/Rtools/  Installing package into ‘C:/Users/Prarthana/AppData/Local/R/win-library/4.3’  (as ‘lib’ is unspecified)  trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/textdata\_0.4.4.zip'  Content type 'application/zip' length 502149 bytes (490 KB)  downloaded 490 KB  package ‘textdata’ successfully unpacked and MD5 sums checked  The downloaded binary packages are in  C:\Users\Prarthana\AppData\Local\Temp\Rtmpe2Ib6F\downloaded\_packages  > library(textdata)  Warning message:  package ‘textdata’ was built under R version 4.3.3  > review\_afinn <- review\_sentiment %>%  + inner\_join(get\_sentiments("afinn"), by = "word") %>%  + mutate(sentiment = ifelse(value < 0, "Negative", "Positive"))  Do you want to download:  Name: AFINN-111  URL: http://www2.imm.dtu.dk/pubdb/views/publication\_details.php?id=6010  License: Open Database License (ODbL) v1.0  Size: 78 KB (cleaned 59 KB)  Download mechanism: https  1: Yes  2: No  Selection: table(review\_afinn$value,review\_afinn$sentiment)  Enter an item from the menu, or 0 to exit  Selection: 1  trying URL 'http://www2.imm.dtu.dk/pubdb/views/edoc\_download.php/6010/zip/imm6010.zip'  Content type 'application/zip' length 16227 bytes (15 KB)  downloaded 15 KB    > review\_afinn%>%  + group\_by(sentiment)%>%  + count(word,sentiment)%>%  + top\_n(10,n)%>%  + ungroup()%>%  + ggplot(aes(x=reorder(word,n),y=n,fill=sentiment))+  + geom\_col(show.legend=F)+  + geom\_text(aes(label = n), hjust = --1.3, size = 3, color = "black") +  + facet\_wrap(~sentiment,scales="free")+  + coord\_flip()+  + labs(x="number of occurences",y="Words",title="Top 10 words for each sentiment used captions using afinn lexicon")+  + theme(plot.title = element\_text(size = 8, face = "bold")) |
|  |
| |  | | --- | | > | |



A graph with different colored bars

Description automatically generated with medium confidence

A graph of different colored squares

Description automatically generated with medium confidence