

Projekt Opinion Mining

1. Wprowadzenie

Przedstawiona praca stanowi analizę opinii dotyczących najpopularniejszych modeli butów do biegania renomowanych marek: Nike, Adidas, New Balance, Puma oraz Under Armour. Głównym narzędziem wykorzystanym w badaniu jest metoda Data Miningu, a szczególnie techniki Text Miningu, które pozwalają na ekstrakcję i analizę istotnych informacji z pozyskanych opinii o wybranych produktach. W kolejnych etapach zostaną poruszone takie zagadnienia jak zebranie i przygotowanie danych, wizualizacja i analiza danych oraz podsumowanie i wnioski.

2. Cel i założenia

Głównym celem jest zidentyfikowanie oraz zobrazowanie tematów opinii oraz ogólne nastawienie konsumentów do danego produktu. Pod uwagę będą brane głównie to na jaki temat wypowiadają się opiniujący oraz jakie mają doświadczenia z zakupionym produktem. Będąc dobrze zaznajomionym z działaniem tych marek na rynku, pozwoliłem sobie wyprowadzić następujące hipotezy:

- We wszystkich markach, opinie będą głównie pozytywne
- Nike będzie miało największy odsetek opinii pozytywnych (głównie klienci cenią sobie aspekt wizualny i jakościowy, a Nike oferuje wysoki poziom obu)
- Under Armour będzie miał najwięcej opinii negatywnych

3. Przygotowanie danych

Pierwszym krokiem było zebranie danych o najpopularniejszych modelach butów do biegania pięciu renomowanych marek: Nike, Adidas, New Balance, Puma, Under Armour. W tym celu, skorzystałem z portalu Amazon oraz skupiłem się na produktach z największą liczbą opinii. W celu zwiększenia wiarygodności i rzetelności wyników, posortowałem opinie od najnowszych (stan dn. 11.06.2023r.) oraz wybrałem jedynie opinie w języku angielskim. Dla każdej marki zebrałem po 70 opinii, umieściłem w arkuszu Excel, w taki sposób aby w jednej kolumnie znajdował się tekst opinii, a w kolumnie obok nazwa marki. Zapisalem plik w formacie .csv, aby sprawniej wykonywać różne operacje na pliku w programie RStudio. Następnie rozpocząłem proces tworzenia kodu. Po zaimportowaniu pliku .csv do środowiska roboczego, przeszedłem do przygotowania samych danych. Poniżej, przedstawione są kolejno podjęte przeze mnie kroki:

1. Usunięcie znaków specjalnych, cyfr oraz znaków interpunkcyjnych.
2. Przekształcenie wielkich liter na małe.
3. Usunięcie "stopwords", czyli powszechnych i uniwersalnych słów, które nie odwołują się bezpośrednio do tematu badania (np. be, go, am, and).
4. Usunięcie pustych przestrzeni.
5. Stemming, czyli pozbawianie wyrazów ich końcówek oraz pozostawienie samego tematu wyrazu.

```
<<DocumentTermMatrix (documents: 70, terms: 336)>>  
Non-/sparse entries: 737/22783  
Sparsity           : 97%  
Maximal term length: 10  
weighting          : term frequency (tf)
```

Tak prezentuje się DTM po wstępnym przygotowaniu danych. **NATOMIAST**, są to jedynie wyniki oparte na marce Nike, gdyż na samym początku kodu zazaczyłem, że na razie chcę operować tylko na tej marce w celu tworzenia wszelkich wizualizacji i topic modellingu. Dzięki temu, zaoszczędziłem również wiele czasu w tworzeniu samego kodu, gdyż teraz, w celu zobaczenia wyników odpowiednich marek, należy jedynie zmienić 12. linijkę kodu.

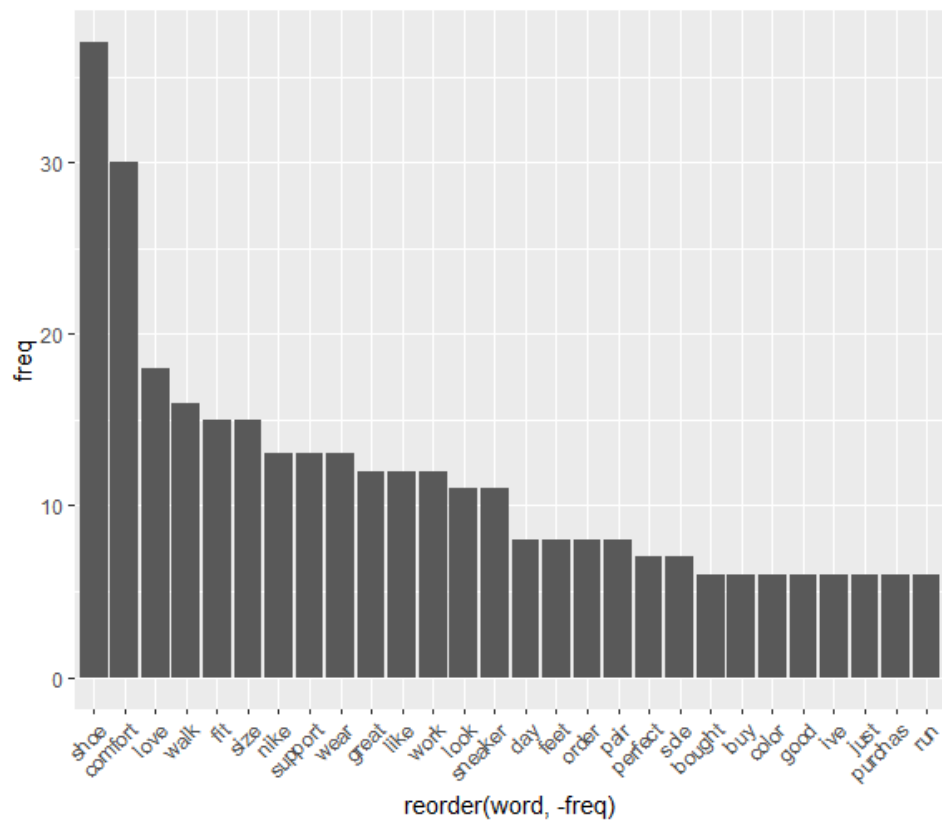
4. Wizualizacja i analiza danych

Poniżej, znajdują się wszelkie wizualizacje stanowiące część Topic Modellingu oraz Sentimental Analysis przeprowadzonych dla opinii każdej z wybranych marek.

4.1. Nike

Po przygotowaniu danych, dla marki Nike, pozostało 336 wyrazów.

Histogram i wordcloud najczęściej występujących słów w korpusie Nike.



Topic modelling:

| | Topic 1 | Topic 2 | Topic 3 | Topic 4 |
|-------|-----------|-----------|-----------|-----------|
| [1,] | "shoe" | "like" | "comfort" | "love" |
| [2,] | "fit" | "work" | "walk" | "nike" |
| [3,] | "look" | "size" | "wear" | "great" |
| [4,] | "support" | "order" | "just" | "sneaker" |
| [5,] | "pair" | "perfect" | "purchas" | "day" |
| [6,] | "sole" | "buy" | "realli" | "ive" |
| [7,] | "good" | "color" | "best" | "even" |
| [8,] | "size" | "run" | "bought" | "one" |
| [9,] | "feet" | "light" | "hurt" | "time" |
| [10,] | "wide" | "theyr" | "much" | "year" |

Powyżej ukazane jest 10 najczęściej występujących słów w danym temacie. Analizując podane słowa postanowiłem nazwać tematy następująco:

- Topic 1 – Rozmiar i dopasowanie
- Topic 2 – Stan wizualny i fizyczny
- Topic 3 - Doświadczenia podczas użytkowania
- Topic 4 – Doświadczenie z marką (postawa wobec marki, czas dostawy, częstotliwość użytkowania/kupowania produktów marki)

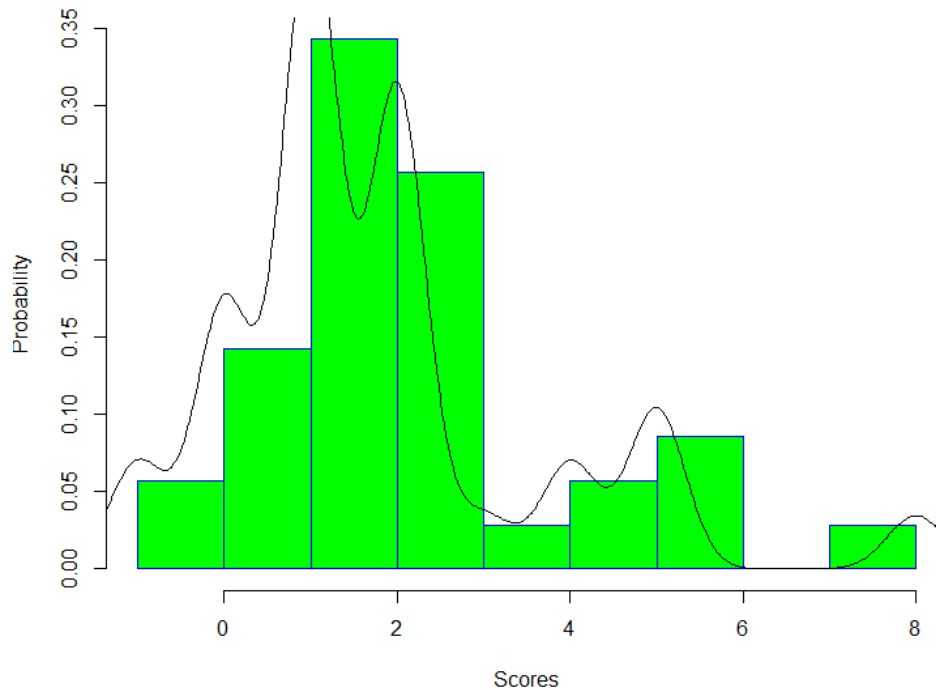
| | Topic 1 | Topic 2 | Topic 3 | Topic 4 |
|--------------|-----------------------|--------------------------|-----------------------------------|-----------------------|
| Nazwa | Rozmiar i dopasowanie | Stan wizualny i fizyczny | Doświadczenia podczas użytkowania | Doświadczenie z marką |
| Ilość opinii | 35 | 15 | 11 | 9 |

Ilość opinii oznacza ilość przynależących opinii do danego tematu

A. Topic 1: Rozmiar i dopasowanie

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--------|---------|--------|-------|---------|-------|
| -1.000 | 1.000 | 1.000 | 1.771 | 2.000 | 8.000 |

Histogram for the Sentiment by Topic Rozmiar i dopasowanie

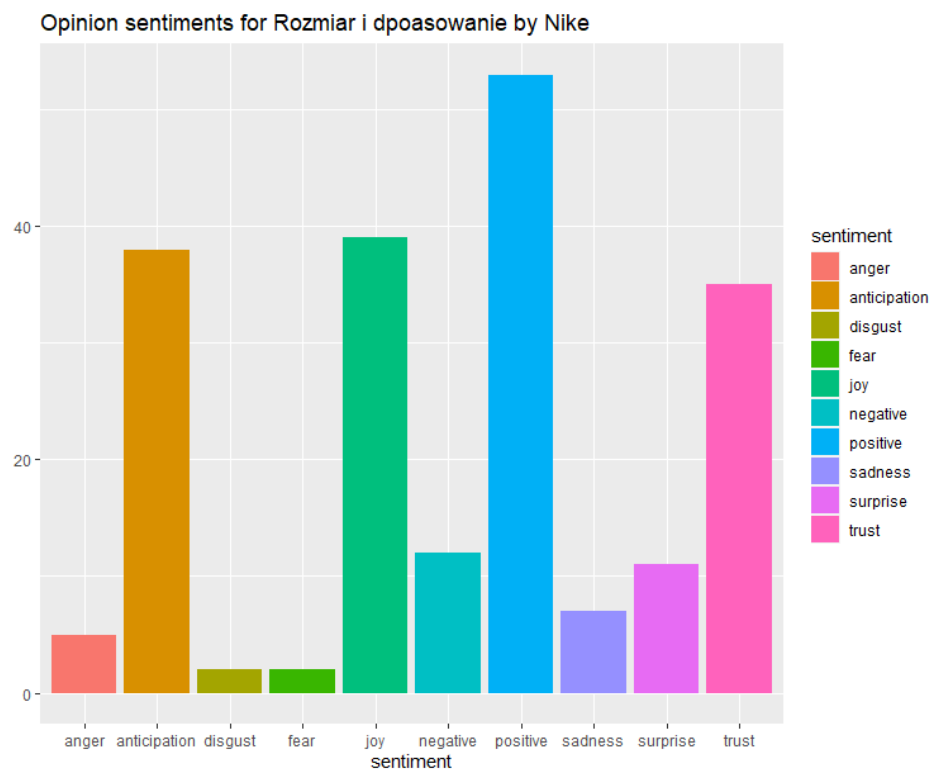


```
> pos1$Score
[1] 1 2 1 1 1 8 5 1 5 2 1 2 2 3 1 2 2 4 1 5 1 2 4 2 1 1 1 2
> length(pos1$Score)
[1] 28

> neu1$Score
[1] 0 0 0 0 0
> length(neu1$Score)
[1] 5

> neg1<-m1[m1$Score<0,]
> neg1$Score
[1] -1 -1
> length(neg1$Score)
[1] 2
```

Opinie dla Tematu Rozmiar i dopasowanie mają bez wątpienia pozytywne nacechowanie. Pojawiło się 28 opinii pozytywnych, 5 neutralnych i 2 negatywne.



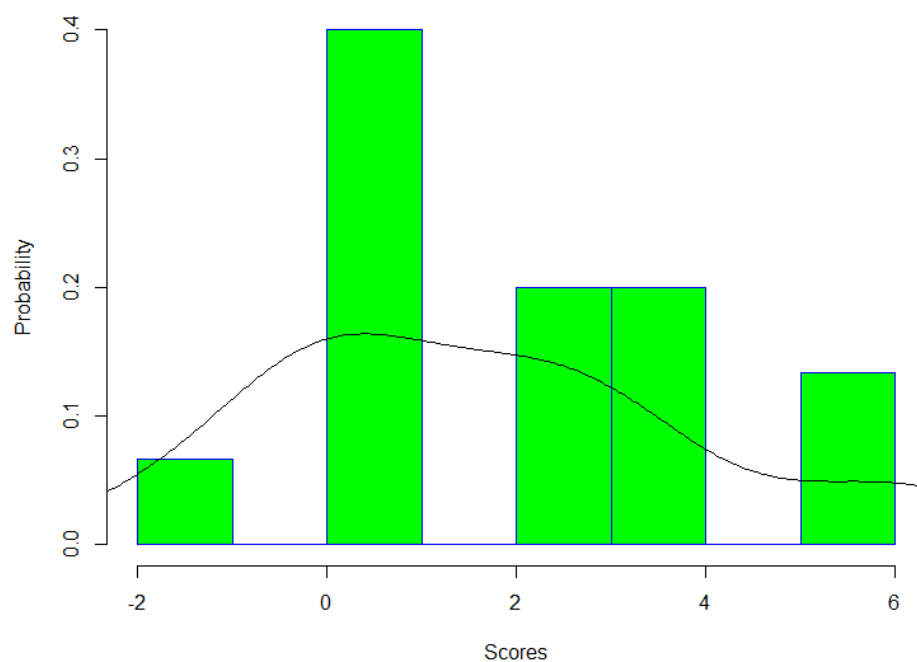
Powyżej, ukazane są emocje jakie wyrażali klienci poprzez słowa. Przede wszystkim byli pozytywni, czuli radość, zaufanie oraz ich oczekiwania zostały spełnione.

B. Topic 2: Stan wizualny i fizyczny

```
> summary(m1$Score)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--------|---------|--------|-------|---------|-------|
| -2.000 | 0.000 | 2.000 | 1.667 | 3.000 | 6.000 |

Histogram for the Sentiment by Topic Stan wizualny i fizyczny

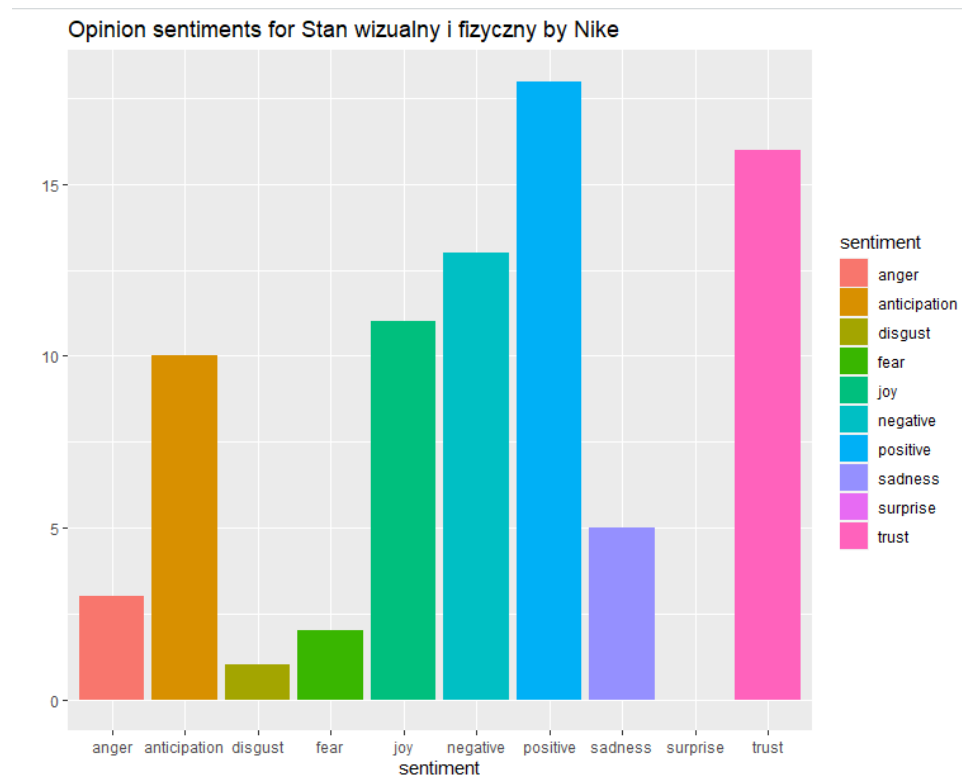


```

> pos1$score      > neu1$score      > neg1$score
[1] 6 3 3 6 2 2 2 3 [1] 0 0 0 0 0 0 [1] -2
> length(pos1$score) > length(neu1$score) > length(neg1$score)
[1] 8              [1] 6              [1] 1

```

Opinie dla Tematu Stan wizualny i fizyczny mają nieco zróżnicowane nacechowanie. Pojawiło się 8 opinii pozytywnych, 6 neutralnych i 1 negatywna. Zdecydowanie mniej osób pozostawiało opinię na ten temat, jednak te, które się pojawiły, były albo pozytywne albo neutralne.

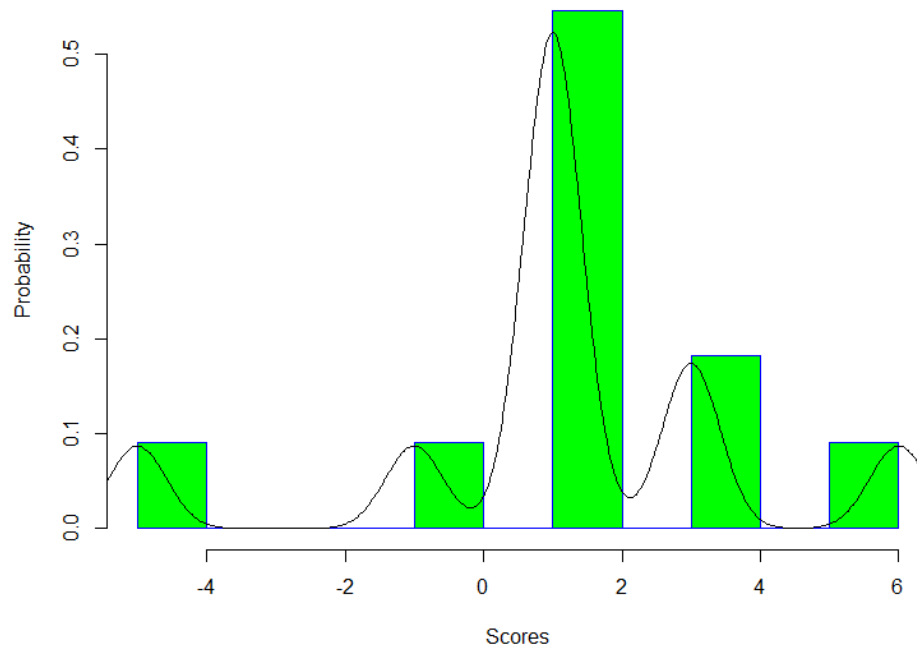


Klienci swoimi opiniami przede wszystkim wyrażali radość i zaufanie natomiast pojawiło się stosunkowo dużo emocji neutralnych takich jak antycypacja.

C. Topic 3: Doświadczenia podczas użytkowania

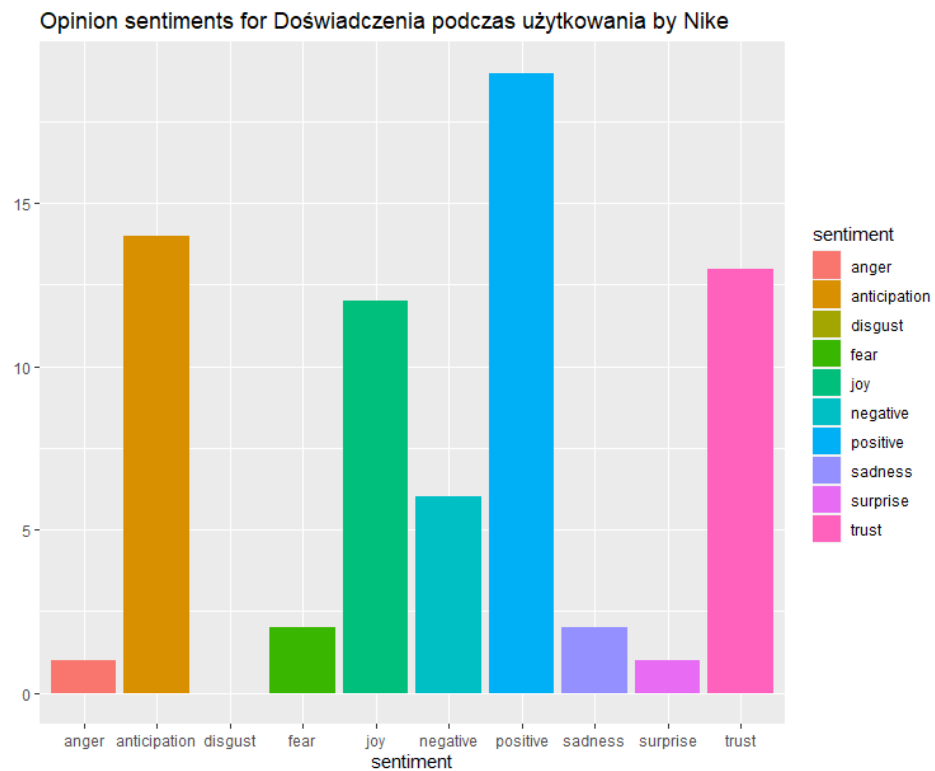
```
> summary(m1$score)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-5.000  1.000   1.000   1.091  2.000   6.000
```

Histogram for the Sentiment by Topic Doświadczenia podczas użytkowania



```
> pos1$score      > neu1$score      > neg1$score
[1] 6 1 3 1 1 1 1 1 3 integer(0)      [1] -1 -5
> length(pos1$score) > length(neu1$score) > length(neg1$score)
[1] 9              [1] 0              [1] 2
```

Opinie dla Tematu Doświadczenia podczas użytkowania mają bez wątpienia pozytywne nacechowanie. Pojawiło się 9 opinii pozytywnych, 0 neutralnych i 2 negatywne.

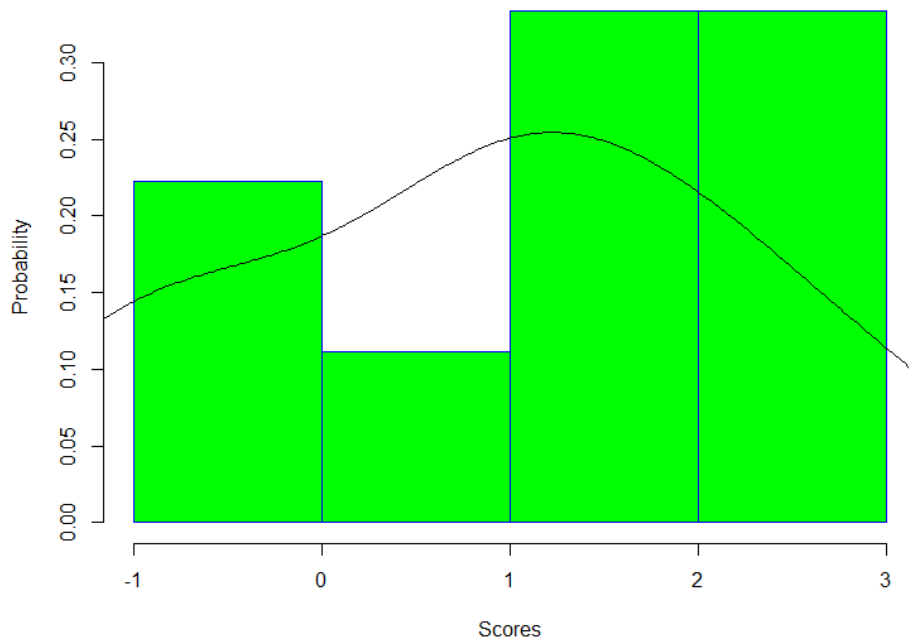


Klienci swoimi opiniami przede wszystkim wyrażali radość i zaufanie natomiast pojawiło się stosunkowo dużo emocji neutralnych takich jak antycypacja. Niektórym można przypisać uczucia negatywne jak smutek i złość, lecz odsetek jest bardzo niewielki.

D. Topic 4: Doświadczenie z marką

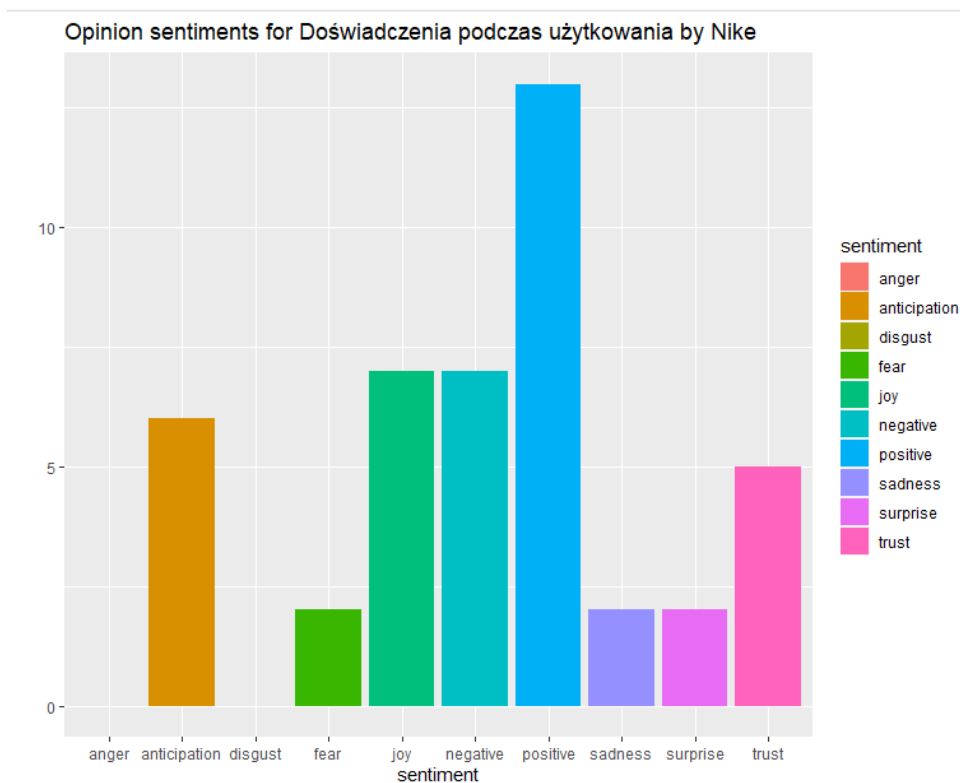
```
> summary(m1$Score)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.0000  0.0000  1.0000  0.8889  2.0000  3.0000
```

Histogram for the Sentiment by Topic Doświadczenie z marką



```
> pos1$Score      > neu1$Score      > neg1$Score
[1] 1 1 3 2 2 1     [1] 0                               [1] -1 -1
> length(pos1$Score) > length(neu1$Score) > length(neg1$Score)
[1] 6               [1] 1                               [1] 2
```

Opinie dla Tematu Doświadczenie z marką mają nieco bardziej zróżnicowane nacechowanie względem pozostałych tematów. Pojawiło się 6 opinii pozytywnych, 1 neutralna i 2 negatywne.

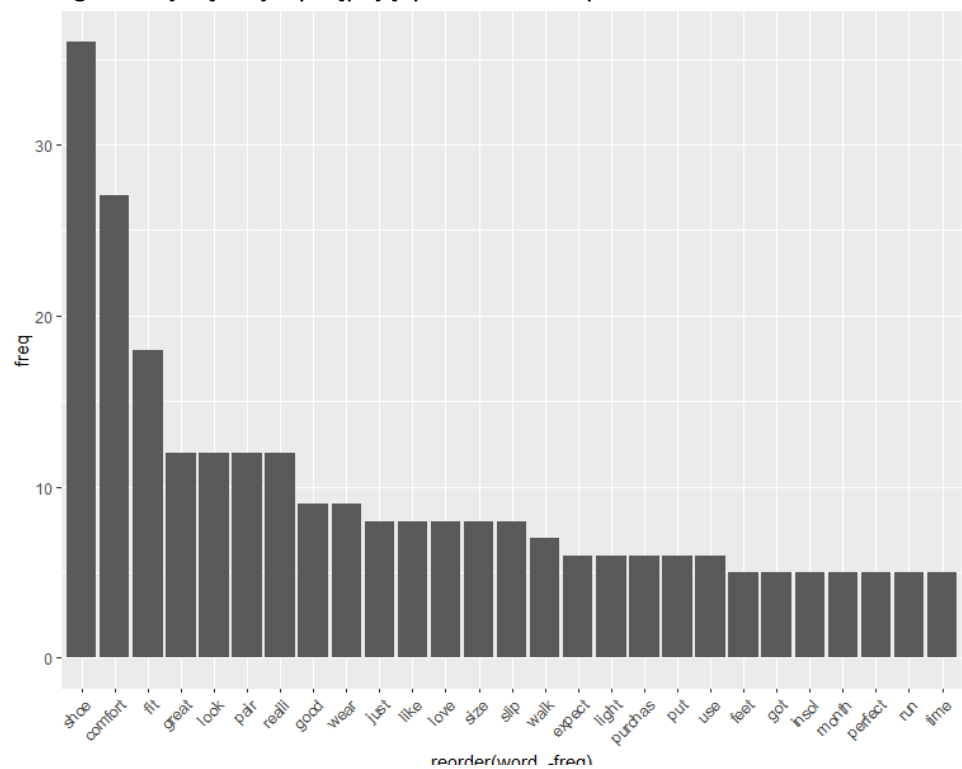


Klienci swoimi opiniami przede wszystkim wyrażali radość, zaufanie, ale też antycypację(neutralne), obawę i zaskoczenie.

4.2. Adidas

Po przygotowaniu danych, dla marki Adidas, pozostało 300 wyrazów.

Histogram najczęściej występujących słów w korpusie Adidas.



Topic modelling

```
> ldaOut.terms
      Topic 1   Topic 2 Topic 3   Topic 4
[1,] "look"    "fit"    "comfort" "shoe"
[2,] "slip"    "great" "just"    "realli"
[3,] "walk"    "pair"   "like"    "good"
[4,] "expect"  "got"    "size"    "wear"
[5,] "purchas" "feet"   "light"   "foot"
[6,] "use"     "broke"  "pair"    "ive"
[7,] "insol"   "can"    "month"   "put"
[8,] "love"    "eas"    "time"    "recommend"
[9,] "pad"     "high"   "dont"    "chang"
[10,] "perfect" "make"   "nice"    "comfi"
```

Powyżej ukazane jest 10 najczęściej występujących słów w danym temacie. Analizując podane słowa postanowiłem nazwać je następująco:

Topic 1 - Stan wizualny i fizyczny

Topic 2 - Doświadczenia podczas użytkowania

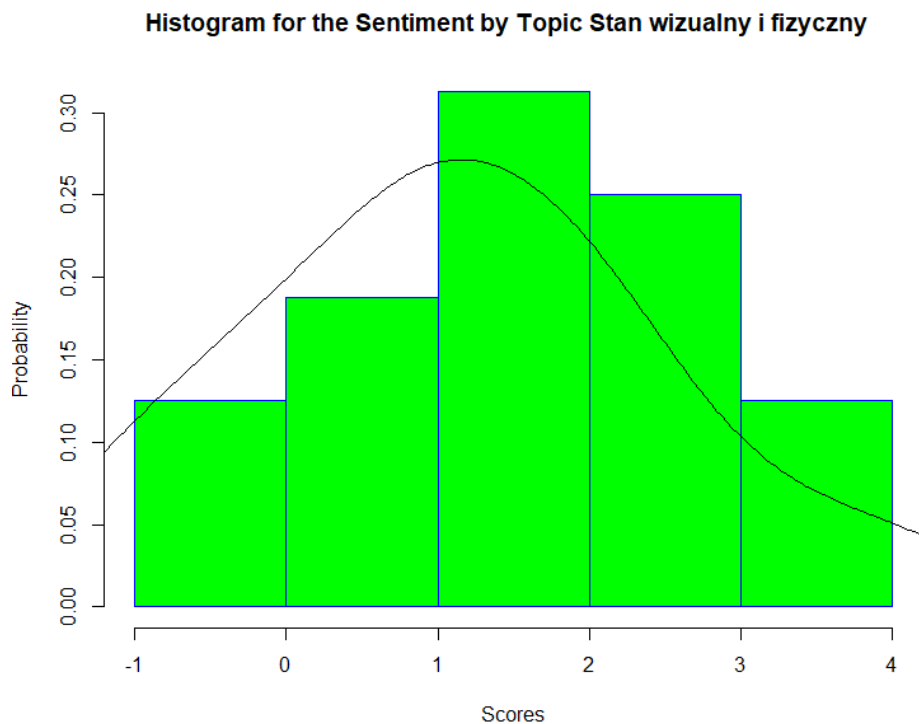
Topic 3 – Doświadczenie z marką (postawa wobec marki, czas dostawy, częstotliwość użytkowania/kupowania produktów marki)

Topic 4 – Rozmiar i dopasowanie

| | Topic 1 | Topic 2 | Topic 3 | Topic 4 |
|--------------|--------------------------|-----------------------------------|-----------------------|-----------------------|
| Nazwa | Stan wizualny i fizyczny | Doświadczenia podczas użytkowania | Doświadczenie z marką | Rozmiar i dopasowanie |
| Ilość opinii | 16 | 16 | 22 | 16 |

A. Topic 1: Stan wizualny i fizyczny

```
> summary(m1$score)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.000  0.000   1.000   1.125  2.000   4.000
```

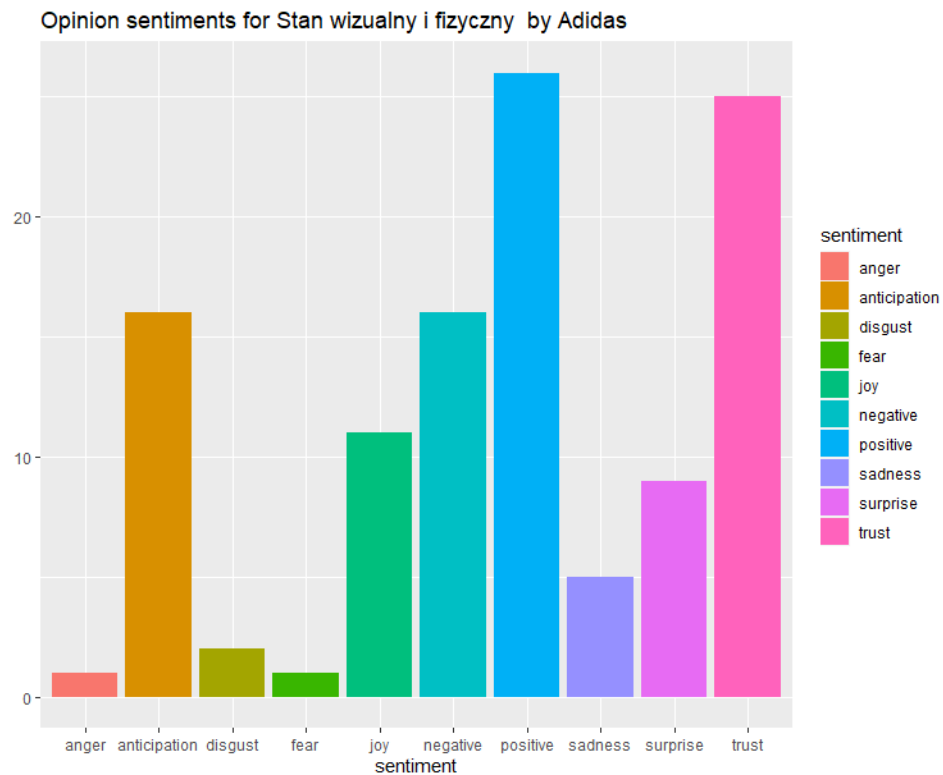


```
> pos1$score
[1] 1 2 1 2 1 3 2 1 2 1 4
> length(pos1$score)
[1] 11

> neu1$score
[1] 0 0 0
> length(neu1$score)
[1] 3

> neg1$score
[1] -1 -1
> length(neg1$score)
[1] 2
```

Opinie dla Tematu Stan wizualny i fizyczny prezentują się następująco: 11 pozytywnych, 3 neutralne i 2 negatywne.



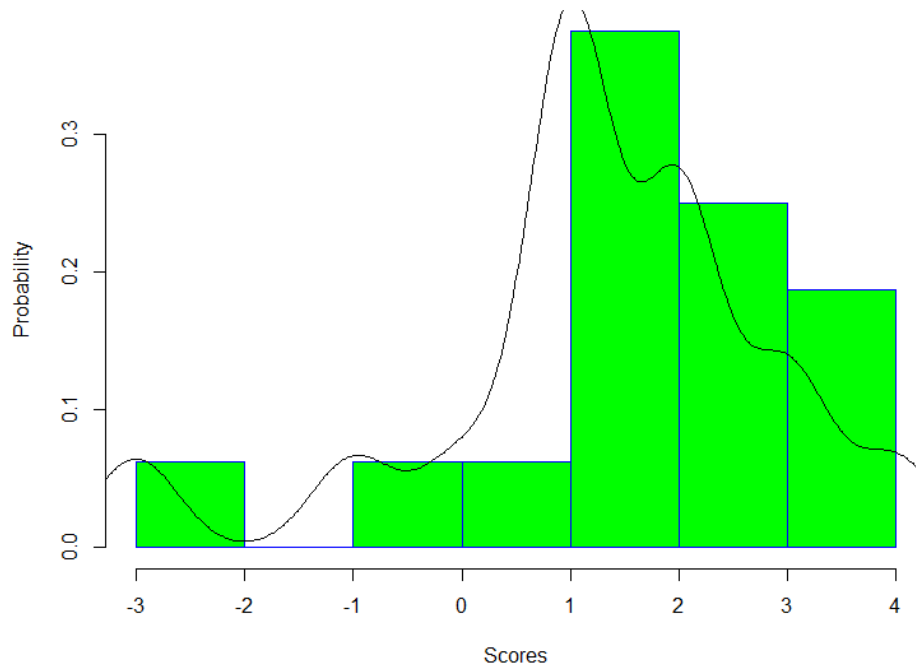
Powyżej, ukazane są emocje jakie wyrażali klienci poprzez słowa. Przede wszystkim w swoich opiniach wyrażali zaufanie i antycypację (neutralne).

B. Topic 2: Doświadczenia podczas użytkowania

```
> summary(m1$Score)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|------|---------|------|
| -3.00 | 1.00 | 1.00 | 1.25 | 2.00 | 4.00 |

Histogram for the Sentiment by Doświadczenia podczas użytkowania



```
> pos1$Score
```

```
[1] 1 1 4 2 1 1 2 1 2 3 1 3 2
```

```
> length(pos1$Score)
```

```
[1] 13
```

```
> neg1$Score
```

```
[1] -1 -3
```

```
> length(neg1$Score)
```

```
[1] 2
```

```
> neu1$Score
```

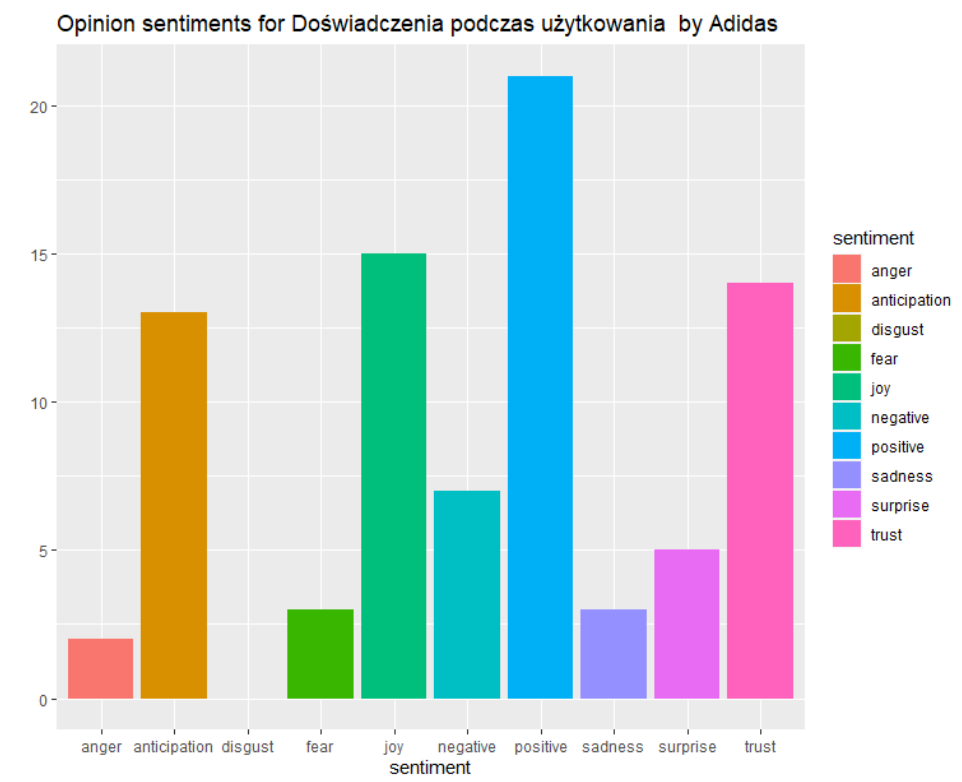
```
> neu1$Score
```

```
[1] 0
```

```
> length(neu1$Score)
```

```
[1] 1
```

Opinie dla Tematu Doświadczenia podczas użytkowania są głównie pozytywne (11). Występują 1 opinia neutralna i 2 negatywne.



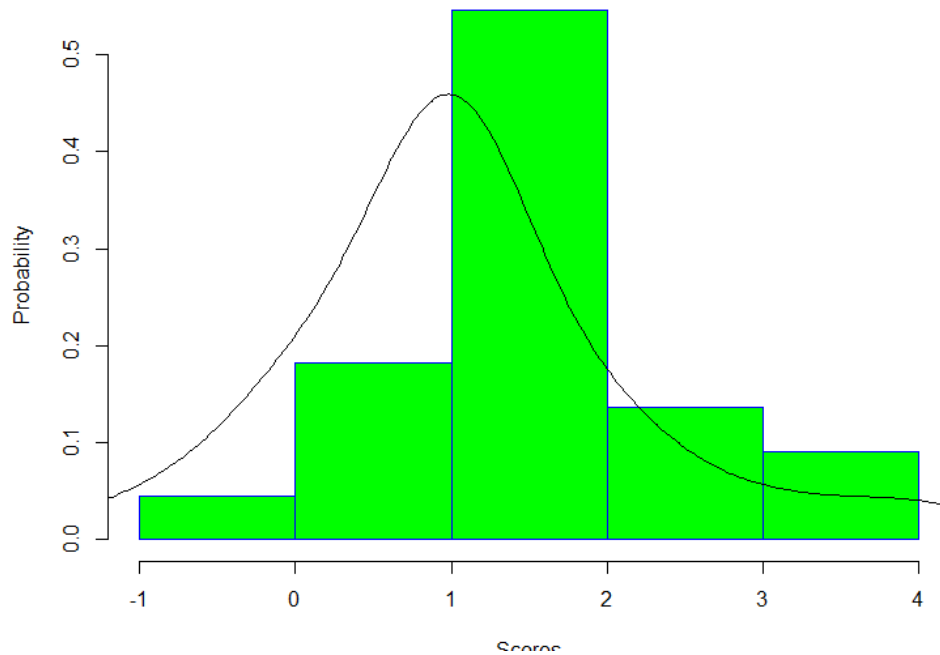
W swoich opiniach klienci głównie wyrażali radość i zaufanie.

C. Topic 3: Doświadczenie z marką

```
> summary(m1$Score)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--------|---------|--------|-------|---------|-------|
| -1.000 | 1.000 | 1.000 | 1.091 | 1.000 | 4.000 |

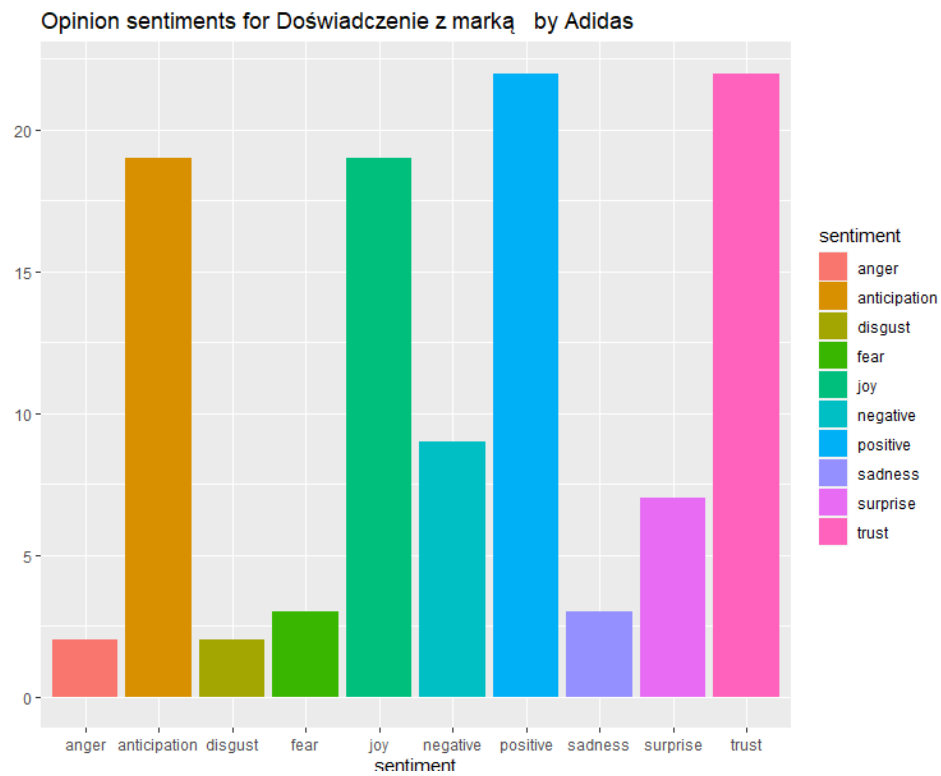
Histogram for the Sentiment by Doświadczenie z marką



```
> pos1$Score
[1] 1 1 1 1 1 1 2 2 2 1 3 1 1 1 1 4 1
> length(pos1$Score)
[1] 17
> neg1$Score
[1] -1
> length(neg1$Score)
[1] 1
```

```
> neu1$Score
[1] 0 0 0 0
> length(neu1$Score)
[1] 4
```

Opinie dla Tematu Doświadczenie z marką mają bardzo pozytywne nacechowanie. 17 opinii pozytywnych, 4 neutralne i tylko 1 negatywna.

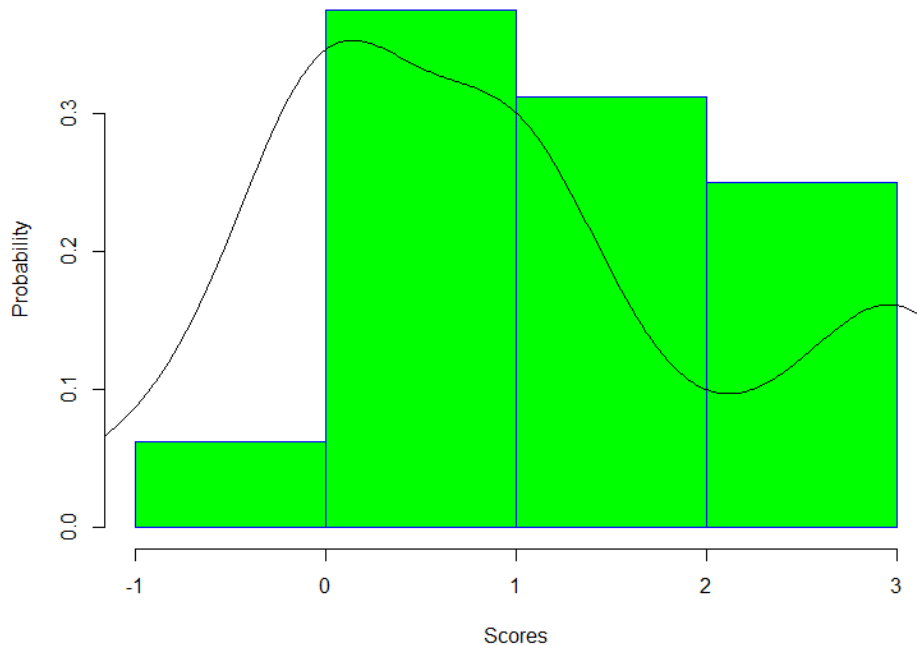


Klienci w swoich opiniach wyrażali głównie zaufanie, radość i antycypację.

D. Topic 4: Rozmiar i dopasowanie

```
> summary(m1$score)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.0000  0.0000  1.0000  0.9375  1.2500  3.0000
```

Histogram for the Sentiment by Rozmiar i dopasowanie

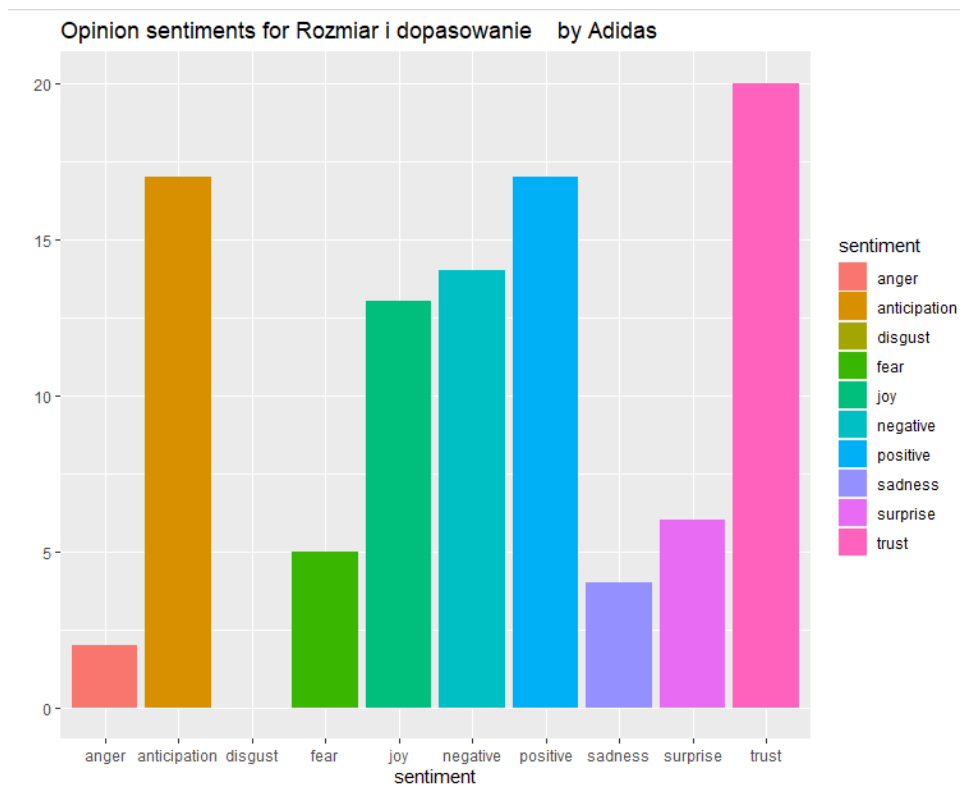


```
> pos1$score
[1] 1 1 1 2 3 1 3 3 1
> length(pos1$score)
[1] 9
```

```
> neu1$score
[1] 0 0 0 0 0 0
> length(neu1$score)
[1] 6
```

```
> neg1$score
[1] -1
> length(neg1$score)
[1] 1
```

Opinie dla Tematu Rozmiar i dopasowanie mają pozytywnie-neutralne opinie. 9 opini pozytywnych, 6 neutralnych i negatywną. Oceny pozytywne nie przekraczają jednak wartości 3.

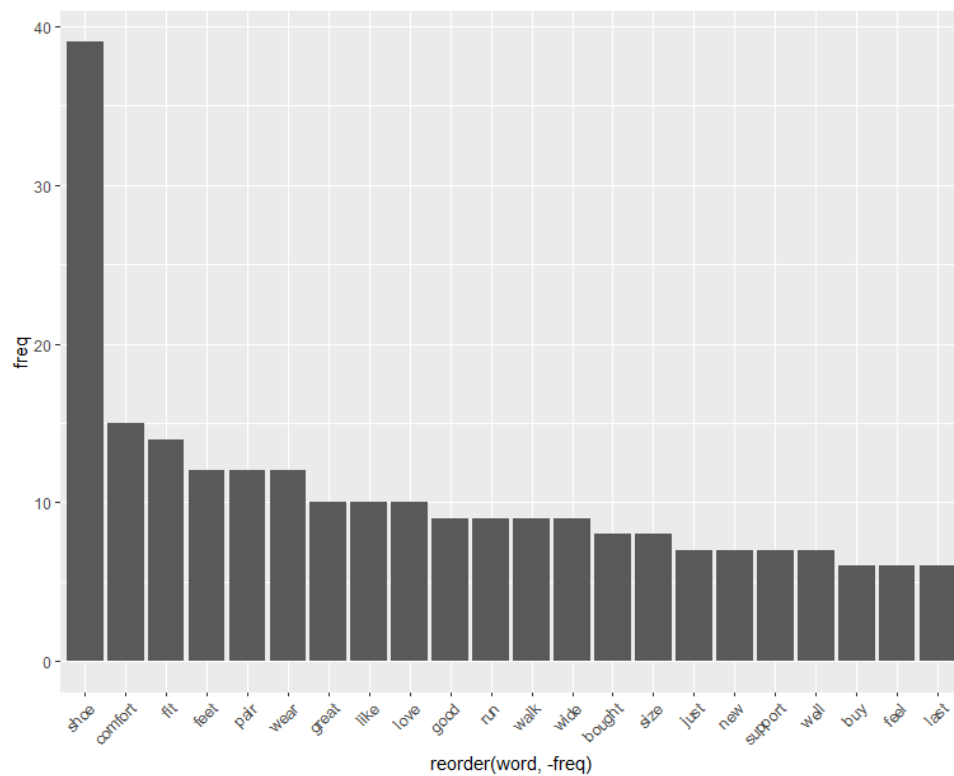


W swoich opiniach klienci wyrażali przeważnie zaufanie i antycypację.

4.3. New Balance

Po preporcessingu danych, w korpusie New Balance zostało 288 słów

Histogram i wordcloud najczęściej występujących słów w korpusie New Balance.



```
> ldaOut.terms
      Topic 1      Topic 2      Topic 3      Topic 4
[1,] "shoe"      "comfort"  "feet"      "great"
[2,] "like"      "fit"        "new"       "love"
[3,] "support"   "pair"       "wear"      "run"
[4,] "buy"       "bought"     "well"      "walk"
[5,] "need"      "just"       "last"      "size"
[6,] "nice"      "wide"       "bad"       "feel"
[7,] "alway"     "realli"     "everi"     "shoe"
[8,] "dont"      "wear"       "lightweight" "will"
[9,] "extra"     "comfi"      "balanc"    "good"
[10,] "good"     "excel"      "color"     "jog"
```

Powyżej ukazane jest 10 najczęściej występujących słów w danym temacie. Analizując podane słowa postanowiłem nazwać je następująco:

Topic 1 - Doświadczenie z marką

Topic 2 - Rozmiar i dopasowanie

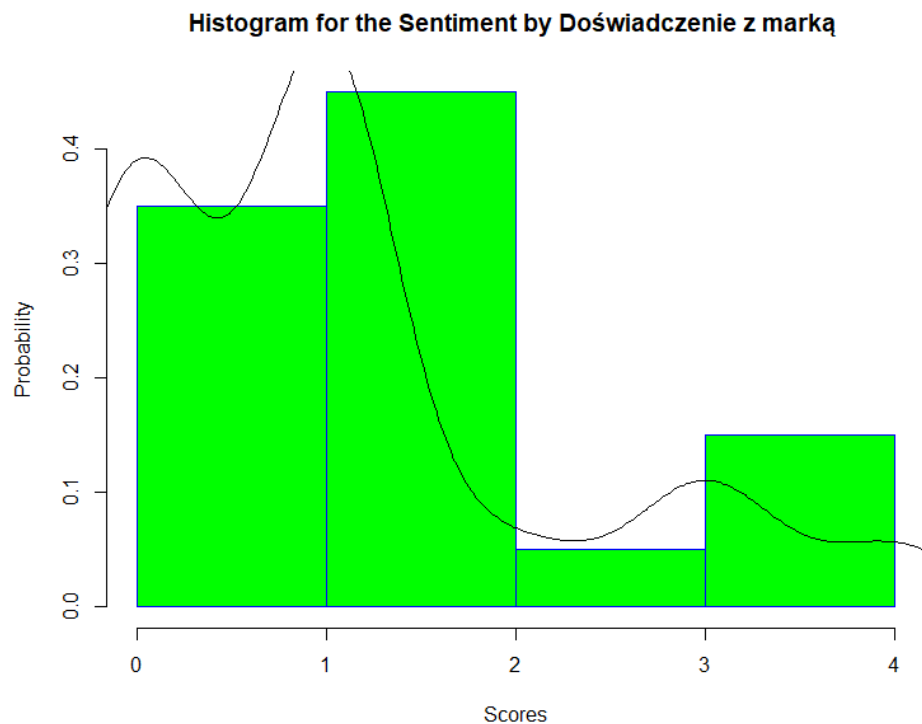
Topic 3 – Stan wizualny i fizyczny

Topic 4 – Doświadczenia podczas użytkowania

| | Topic 1 | Topic 2 | Topic 3 | Topic 4 |
|--------------|-----------------------|-----------------------|--------------------------|-----------------------------------|
| Nazwa | Doświadczenie z marką | Rozmiar i dopasowanie | Stan wizualny i fizyczny | Doświadczenia podczas użytkowania |
| Ilość opinii | 20 | 23 | 15 | 11 |

A. Topic 1: Doświadczenie z marką

```
> summary(m1$score)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00   0.00   1.00   1.05   1.00   4.00
```

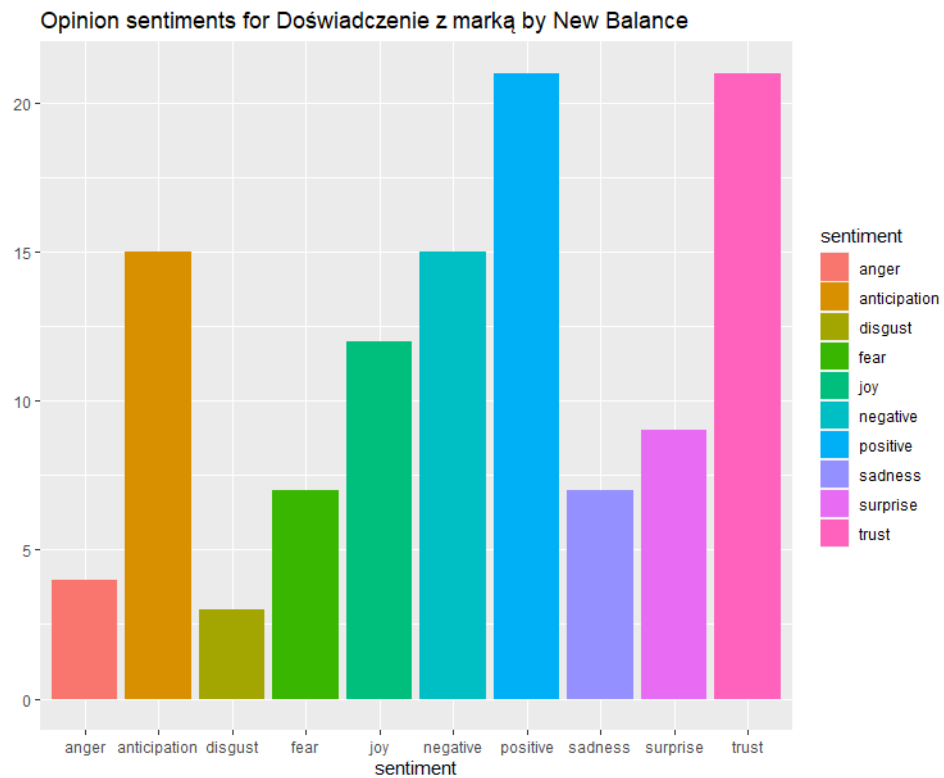
```

> pos1$score
[1] 3 4 1 1 1 1 1 2 1 3 1 1 1
> length(pos1$score)
[1] 13
> neg1$score
integer(0)
> length(neg1$score)
[1] 0

> neu1$score
[1] 0 0 0 0 0 0 0
> length(neu1$score)
[1] 7

```

Opinie dla Tematu Doświadczenie z marką są pozytywnie-neutralne. Nie ma żadnej opinii negatywnej, jest 13 opinii pozytywnych i 7 neutralnych.

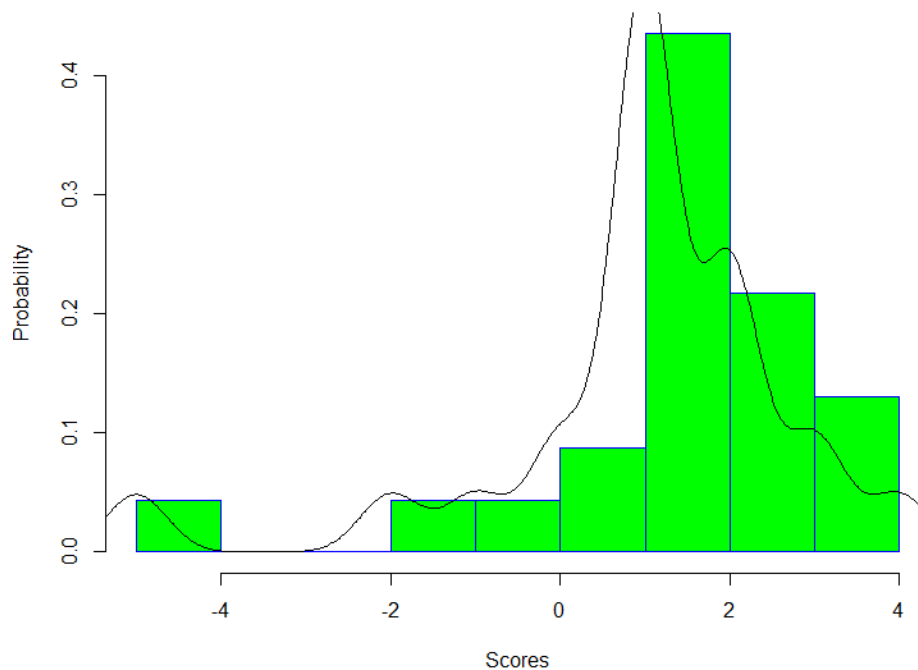


Powyżej, ukazane są emocje jakie wyrażali klienci w swoich opiniach. Głównie wyrażali zaufanie i radość.

B. Topic 2: Rozmiar i dopasowanie

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---------|---------|--------|--------|---------|--------|
| -5.0000 | 1.0000 | 1.0000 | 0.9565 | 2.0000 | 4.0000 |

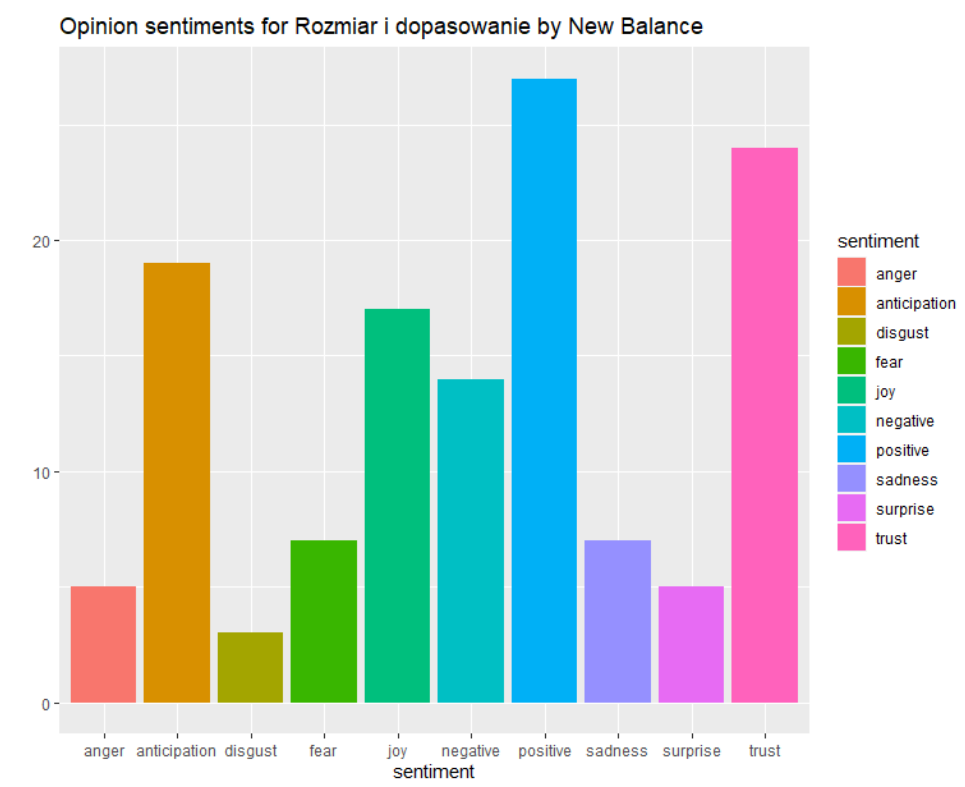
Histogram for the Sentiment by Rozmiar i dopasowanie



```
> pos1$score
[1] 1 1 1 1 1 1 2 4 1 2 2 3 2 1 3 2 1 1
> length(pos1$score)
[1] 18
> neg1$score
[1] -5 -1 -2
> length(neg1$score)
[1] 3

> neu1$score
[1] 0 0
> length(neu1$score)
[1] 2
```

Opinie dla Tematu Rozmiar i dopasowanie są głównie pozytywne. 18 opinii pozytywnych 2 neutralne i 3 negatywne.

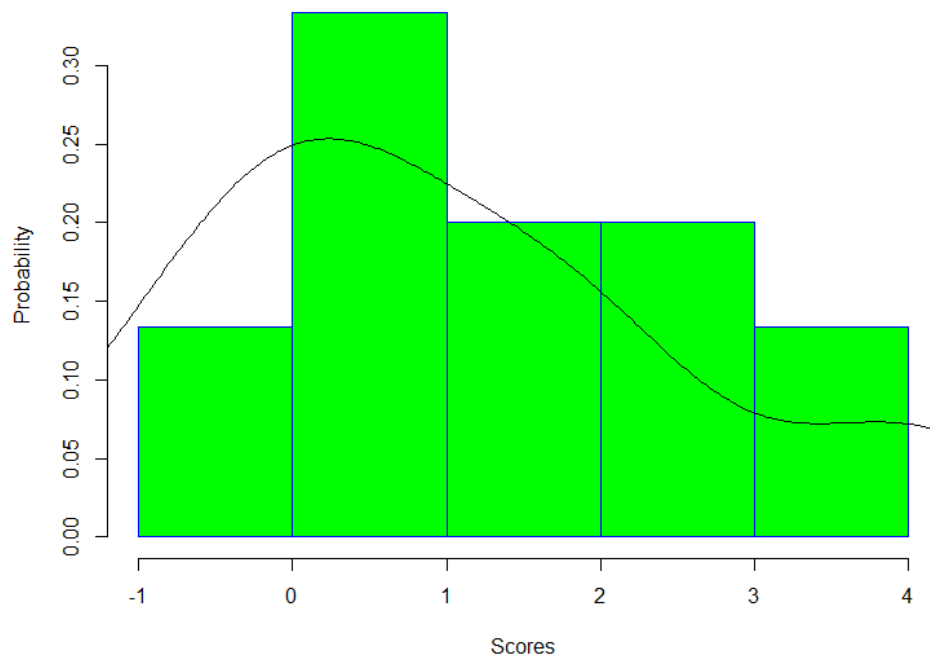


W swoich wypowiedziach, klienci głównie wyrażali radość zaufanie i antycypację.

C. Topic 3: Stan wizualny i fizyczny

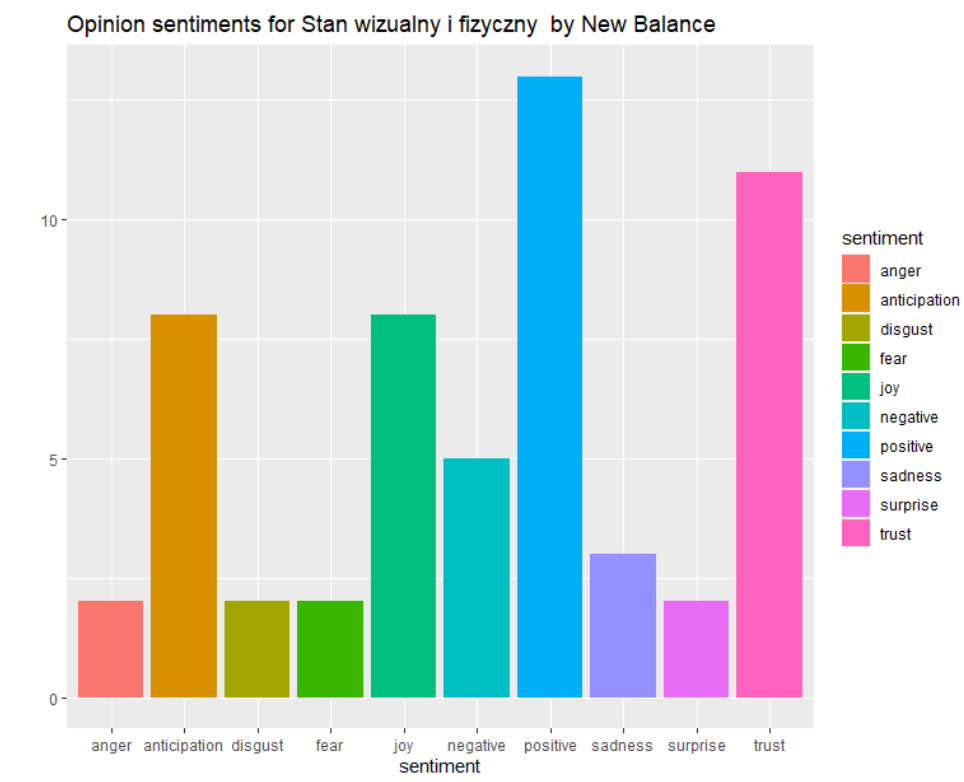
```
> summary(m1$score)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   -1.0   -0.0    0.5    1.0    1.5    4.0
```

Histogram for the Sentiment by Stan wizualny i fizyczny



```
> pos1$score
[1] 2 2 1 1 2 4 1 4
> length(pos1$score)
[1] 8
> neu1$score
[1] 0 0 0 0 0
> length(neu1$score)
[1] 5
> neg1$score
[1] -1 -1
> length(neg1$score)
[1] 2
```

Opinie dla Tematu Doświadczenie z marką mają nieco bardziej zróżnicowane nacechowanie względem pozostałych tematów. Pojawiło się 8 opinii pozytywnych, 5 neutralna i 2 negatywne.

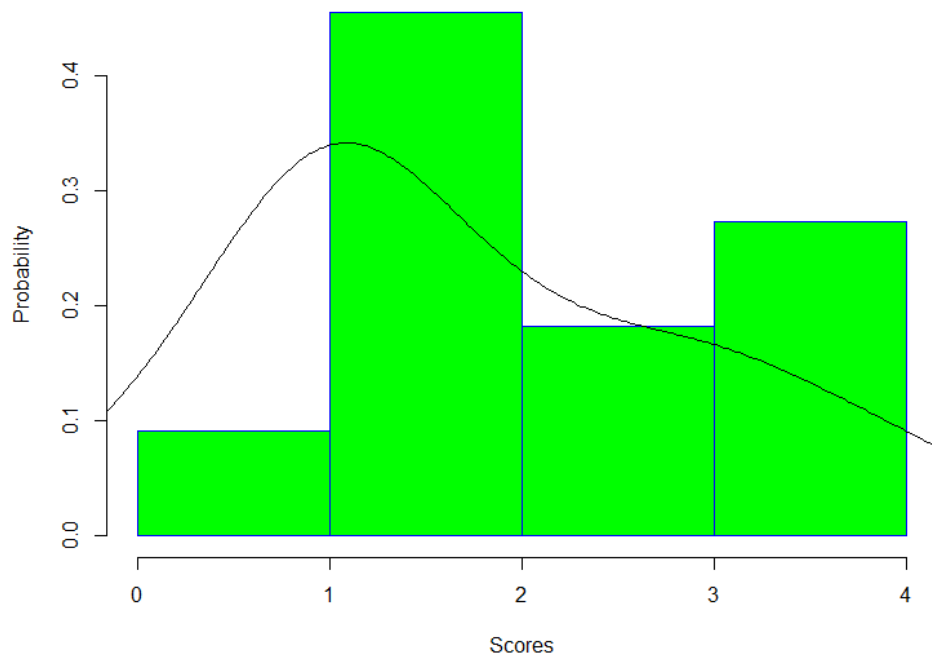


Powyżej, ukazane są emocje jakie wyrażali klienci poprzez słowa. Przede wszystkim byli pozytywni, wyrażali radość, zaufanie oraz ich oczekiwania zostały spełnione.

D. Topic 4: Doświadczenia podczas użytkowania

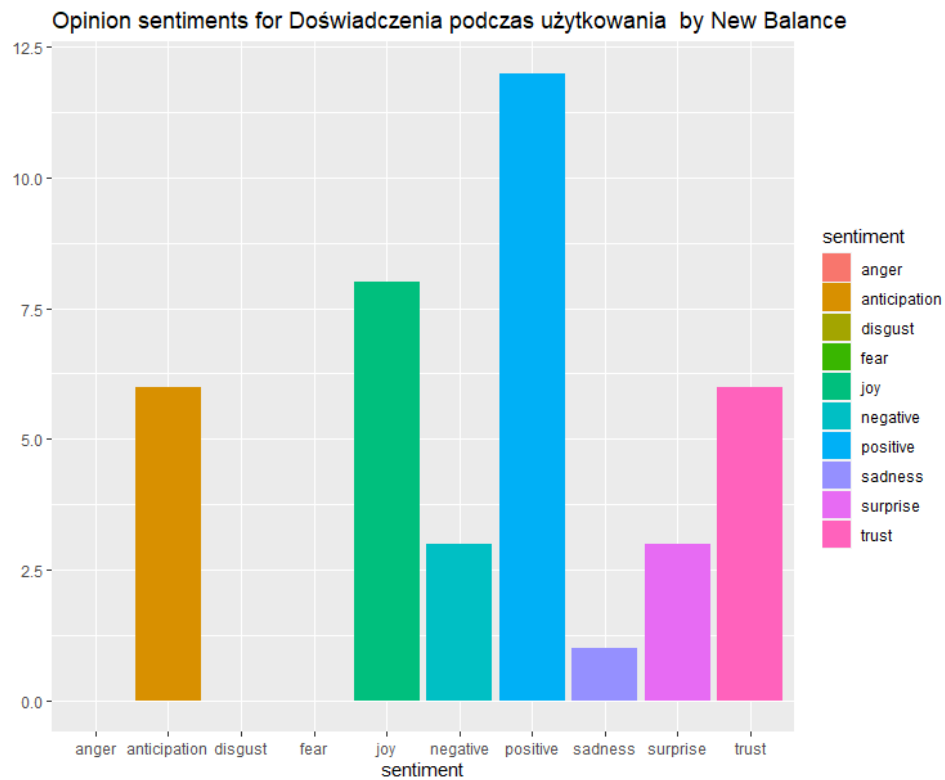
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|-------|
| 0.000 | 1.000 | 1.000 | 1.727 | 2.500 | 4.000 |

Histogram for the Sentiment by Doświadczenia podczas użytkowania



```
> pos1$score      > neu1$score      > length(neg1$score)
[1] 1 3 1 3 2 2 1 1 1 4 [1] 0 [1] 0
> length(pos1$score) > length(neu1$score) > neg1$Documents
[1] 10 [1] 1 character(0) Opinie
```

dla Tematu Doświadczenia podczas użytkowania są głównie pozytywne (10) z jednym komentarzem neutralnym.

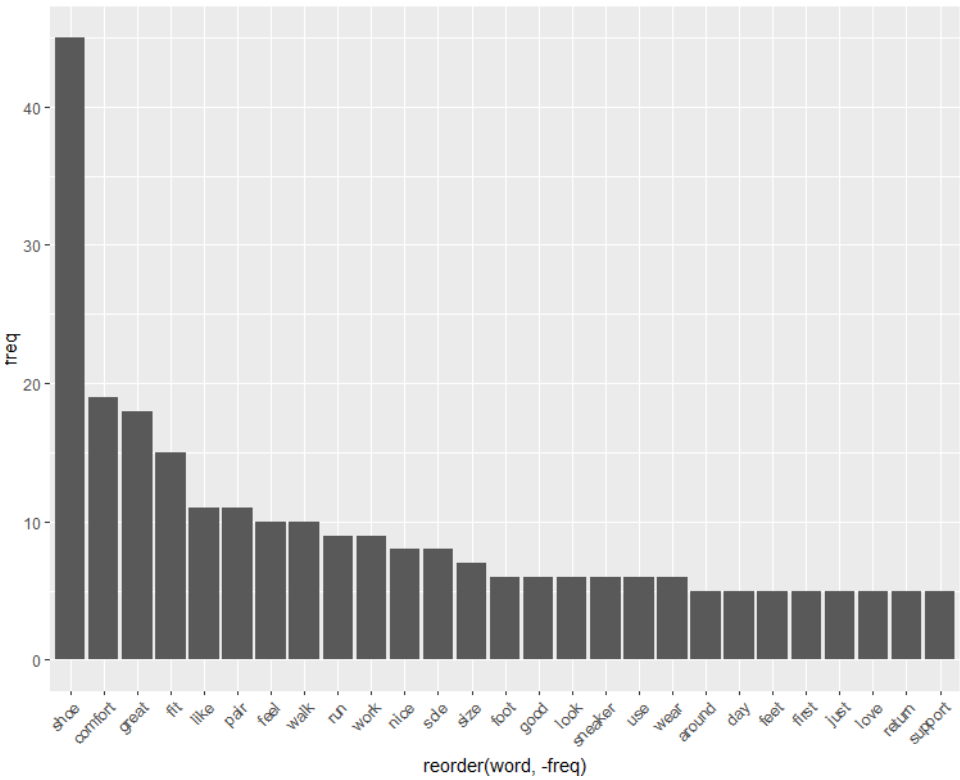


Klienci w swoich opiniach wyrażali przede wszystkim pozytywne emocje jak radość i zaufanie.

4.4. Puma

Po preprocessingu w korpusie Puma pozostało 307 wyrazów.

Histogram i wordcloud najczęściej występujących słów w korpusie Puma.



| | Topic 1 | Topic 2 | Topic 3 | Topic 4 |
|-------|--------------|-----------|-----------|---------|
| [1,] | "great" | "shoe" | "comfort" | "pair" |
| [2,] | "feel" | "run" | "fit" | "walk" |
| [3,] | "like" | "work" | "size" | "nice" |
| [4,] | "wear" | "sole" | "day" | "feet" |
| [5,] | "around" | "look" | "everyth" | "shoe" |
| [6,] | "just" | "first" | "love" | "good" |
| [7,] | "return" | "support" | "perfect" | "foot" |
| [8,] | "sneaker" | "anoth" | "toe" | "gym" |
| [9,] | "purchas" | "bought" | "foot" | "isnt" |
| [10,] | "disappoint" | "brand" | "got" | "ive" |

Powyżej ukazane jest 10 najczęściej występujących słów w danym temacie. Analizując podane słowa postanowiłem nazwać je następująco:

Topic 1 - Doświadczenie z marką

Topic 2 – Stan wizualny i fizyczny

Topic 3 – Rozmiar i dopasowanie

Topic 4 – Doświadczenia podczas użytkowania

| | Topic 1 | Topic 2 | Topic 3 | Topic 4 |
|--------------|-----------------------|--------------------------|-----------------------|-----------------------------------|
| Nazwa | Doświadczenie z marką | Stan wizualny i fizyczny | Rozmiar i dopasowanie | Doświadczenia podczas użytkowania |
| Ilość opinii | 27 | 19 | 12 | 12 |

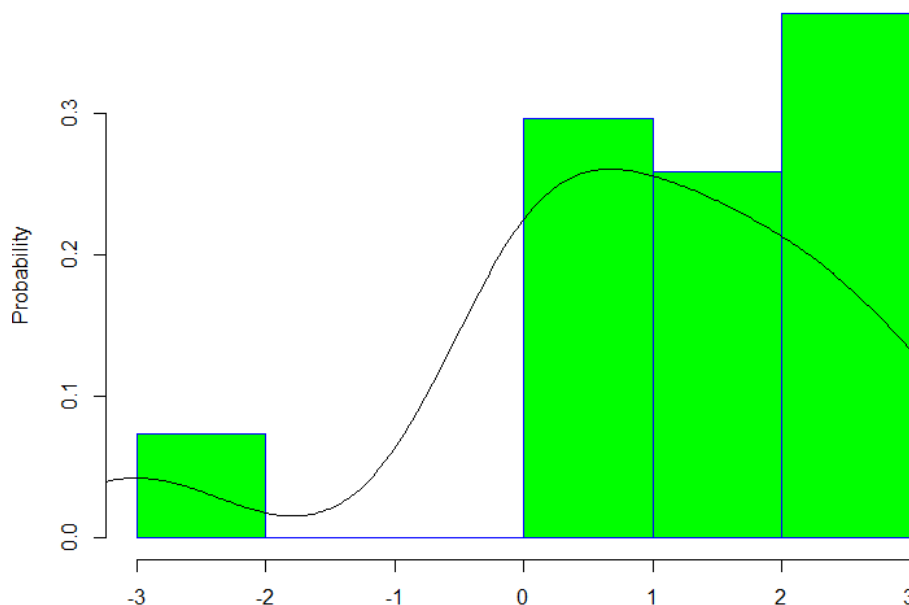
A. Topic 1: Doświadczenie z marką

```

Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-3.0000  0.0000   1.0000   0.9259  2.0000   3.0000

```

Histogram for the Sentiment by Doświadczenie z marką

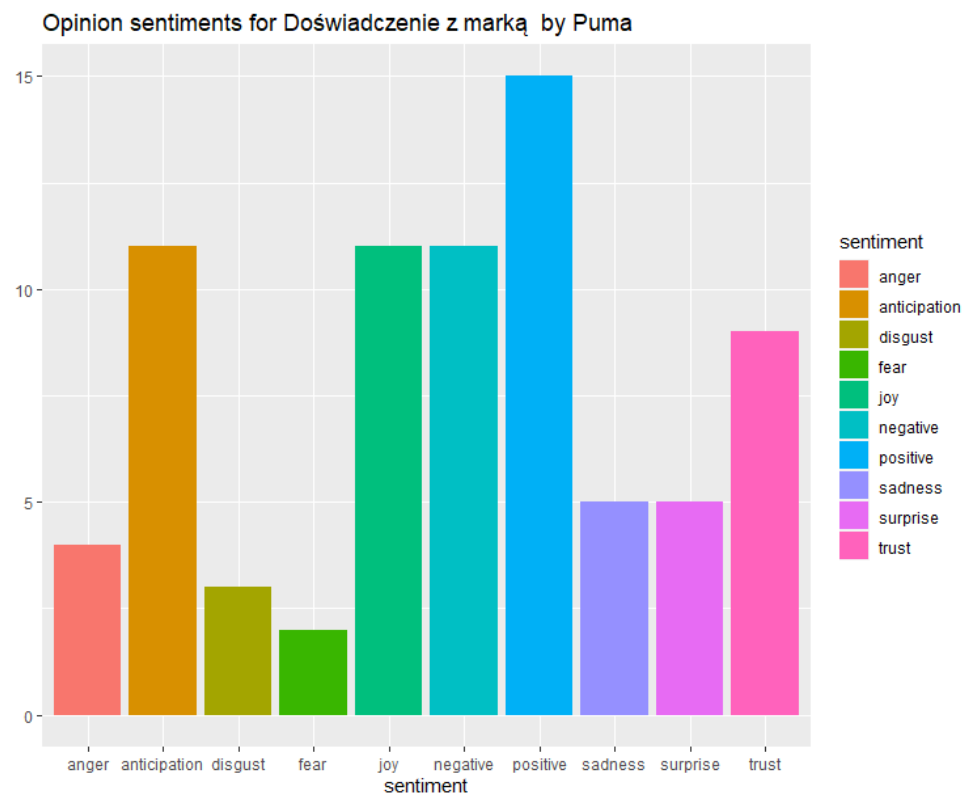


```

[1] 1 1 3 1 1 1 3 2 2 1 2 2 3 2 2 1 3
> length(pos1$score)
[1] 17
> neg1$score
[1] -3 -3
> length(neg1$score)
[1] 2
> neu1$score
[1] 0 0 0 0 0 0 0 0
> length(neu1$score)
[1] 8

```

Opinie dla Tematu Doświadczenie z marką mają pozytywne nacechowanie. Pojawiło się 17 opinii pozytywnych, 8 neutralnych i 2 negatywne.

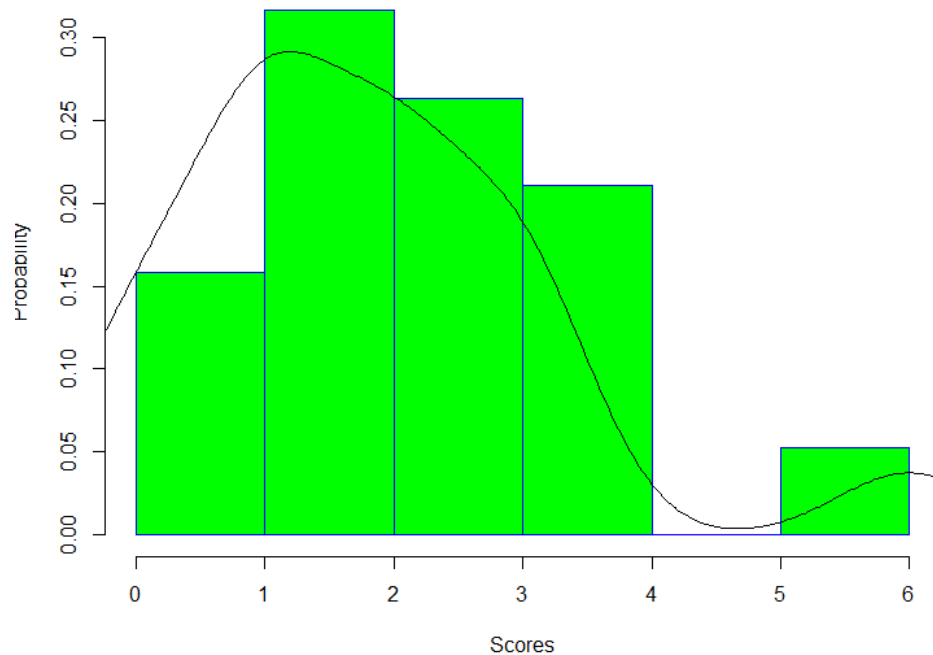


Klienci w swoich opiniach przede wszystkim wyrażali radość, zaufanie, natomiast pojawiły się też negatywne emocje jak złość, smutek a nawet obawy czy obrzydzenie.

B. Topic 2: Stan wizualny i fizyczny

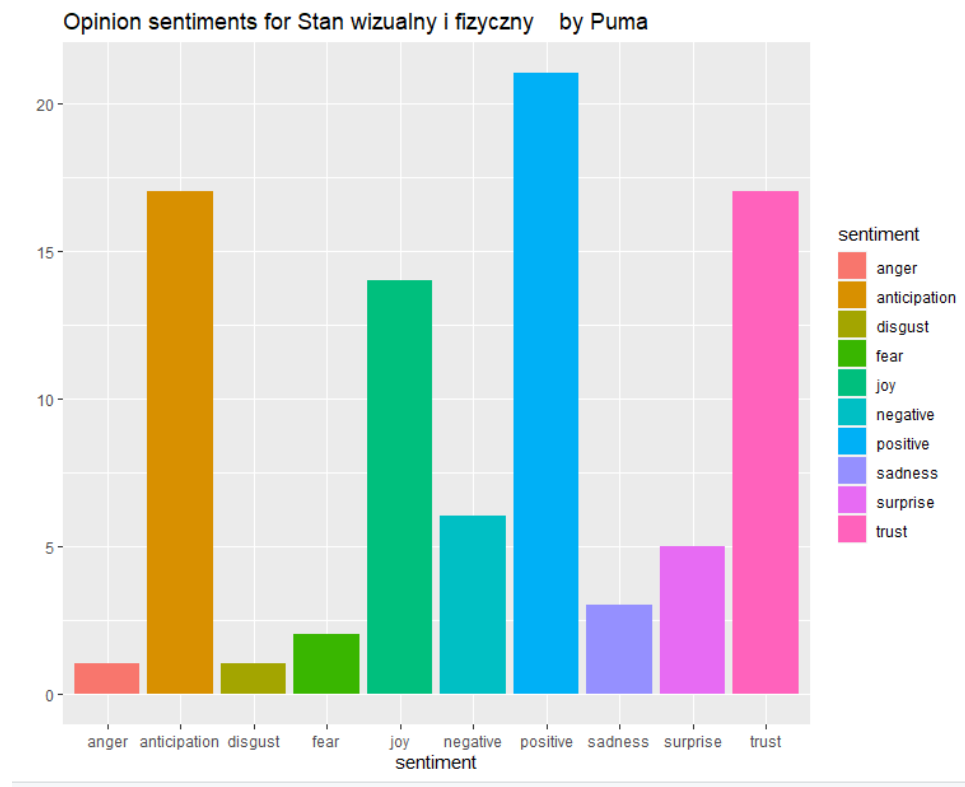
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|-------|
| 0.000 | 1.000 | 2.000 | 1.789 | 2.500 | 6.000 |

Histogram for the Sentiment by Stan wizualny i fizyczny



```
> pos1$score
[1] 2 2 3 6 2 2 2 1 1 1 3 1 3 1 3 1
> length(pos1$score)
[1] 16
> neu1$score
integer(0)
> length(neu1$score)
[1] 0
> neg1$score
[1] 0 0 0
> length(neg1$score)
[1] 3
```

Opinie dla Tematu Stan wizualny i fizyczny mają zdecydowanie nacechowanie pozytywne. 16 opinii pozytywnych i 3 neutralne. Brak odpowiedzi negatywnych.

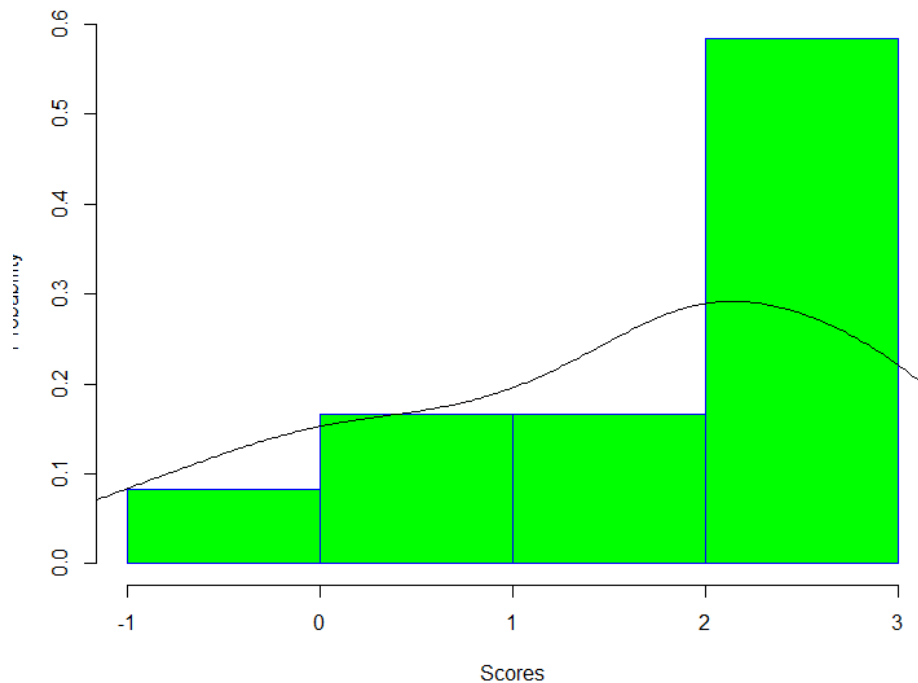


Powyżej, ukazane są emocje jakie wyrażali klienci poprzez słowa. Przede wszystkim byli pozytywni, czuli radość, zaufanie oraz antycypację (neutralne).

C. Topic 3: Rozmiar i dopasowanie

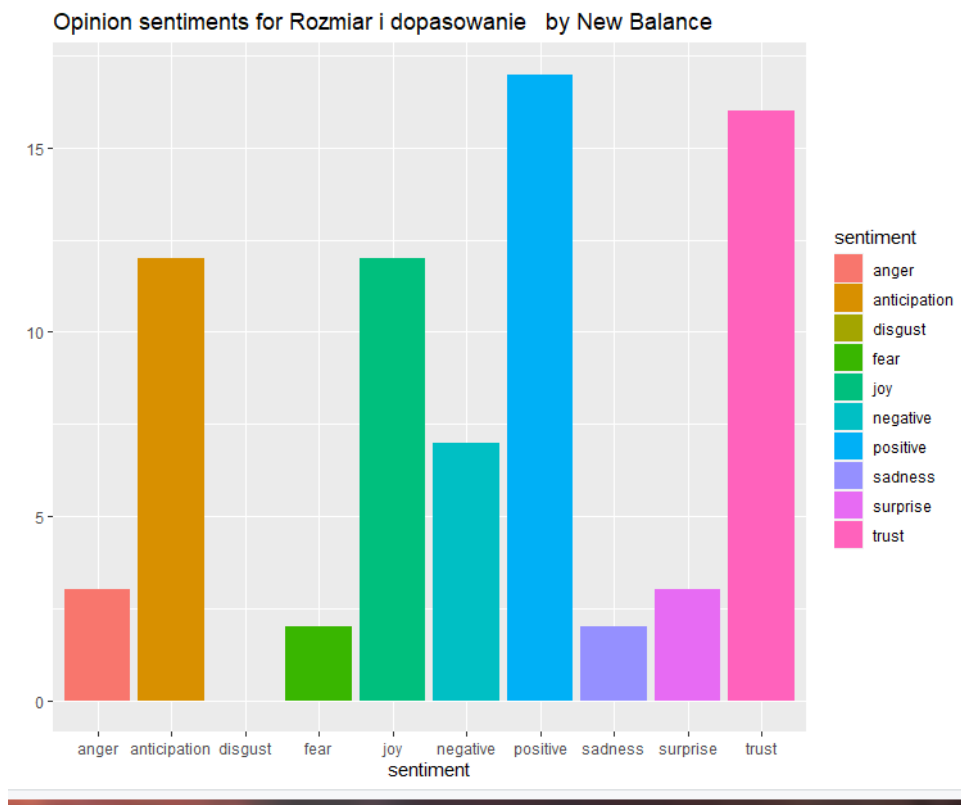
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|------|---------|------|
| -1.00 | 0.75 | 2.00 | 1.50 | 2.25 | 3.00 |

Histogram for the Sentiment by Rozmiar i dopasowanie



```
> pos1$score      > neu1$score      > neg1$score
[1] 3 2 2 2 3 1 3 2 1 [1] 0 0      [1] -1
> length(pos1$score) > length(neu1$score) > length(neg1$score)
[1] 9      [1] 2      [1] 1
```

Opinie dla Tematu Rozmiar i dopasowanie były głównie pozytywne. Pojawiło się 9 opinii pozytywnych, 2 neutralne i 1 negatywną.

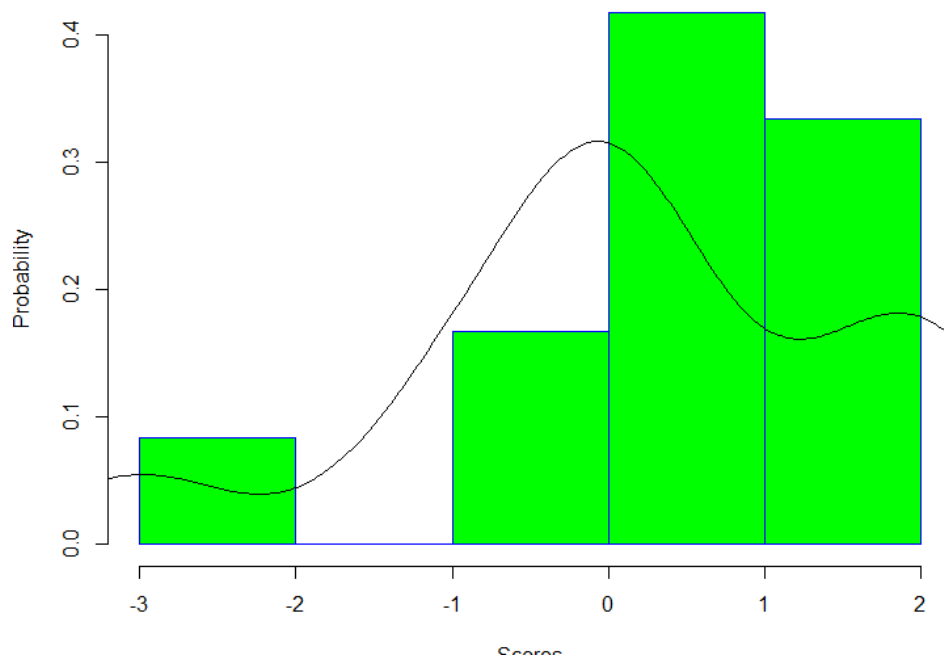


Klienci w swoich opiniach przede wszystkim byli pozytywni, czuli radość oraz zaufanie. Pojawilo się kilka negatywnych emocji jak złość, smutek i obawa.

D. Topic 4: Doświadczenia podczas użytkowania

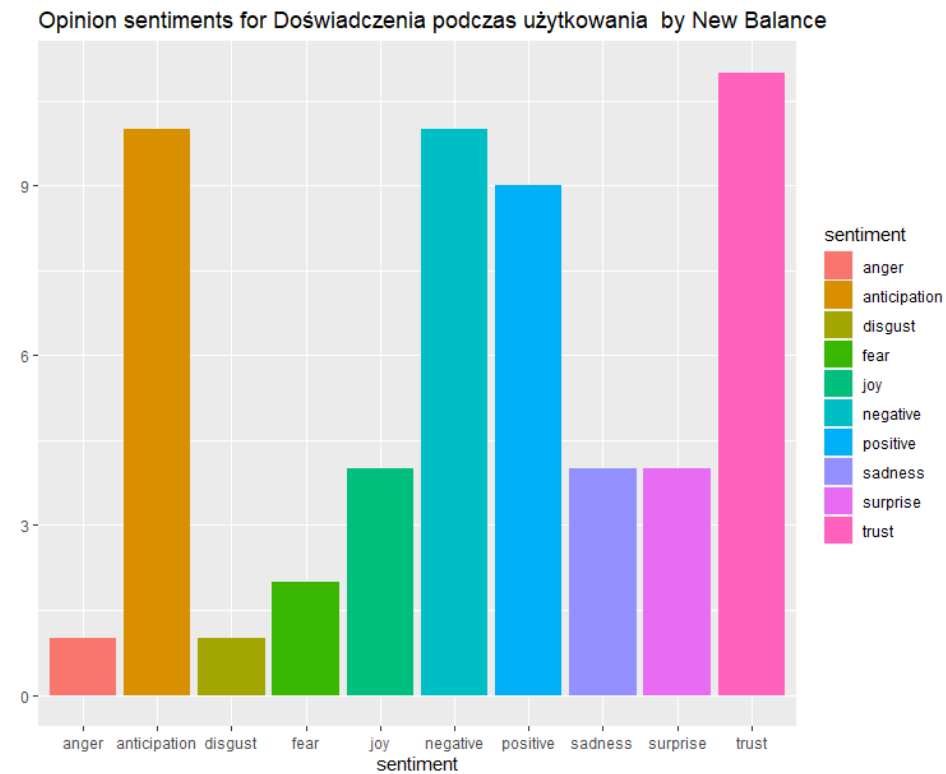
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---------|---------|--------|--------|---------|--------|
| -3.0000 | -0.2500 | 0.0000 | 0.1667 | 1.2500 | 2.0000 |

Histogram for the Sentiment by Doświadczenia podczas użytkowania



```
> pos1$score      > neu1$score      > neg1$score
[1] 2 1 2 2         [1] 0 0 0 0 0         [1] -1 -1 -3
> length(pos1$score) > length(neu1$score) > length(neg1$score)
[1] 4              [1] 5              [1] 3
```

Opinie dla Tematu Doświadczenia podczas użytkowania mają nieco bardziej zróżnicowane nacechowanie względem pozostałych tematów. Pojawiły się 4 opinie pozytywne, 5 neutralnych i 3 negatywne.

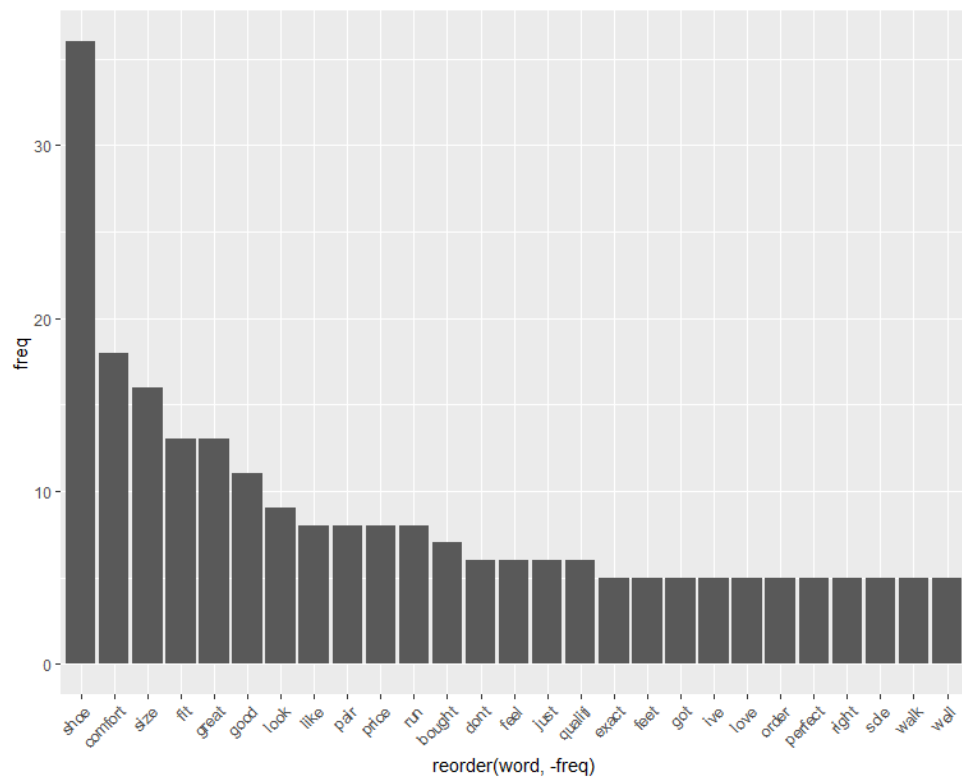


Emocje wyrażane w opiniach były w tym przypadku chyba najbardziej zróżnicowane. Większość z nich była stosunkowo często wyrażana.

4.5 Under Armour

Po preprocessingu, w korpusie Under Armour pozostały 304 wyrazy.

Histogram i wordcloud najczęściej występujących słów w korpusie Under Armour.



```

Topic 1 Topic 2 Topic 3 Topic 4
[1,] "great" "size" "shoe" "good"
[2,] "feel" "fit" "comfort" "look"
[3,] "just" "like" "qualiti" "price"
[4,] "bought" "pair" "sole" "dont"
[5,] "come" "run" "walk" "exact"
[6,] "perfect" "cheap" "purchas" "order"
[7,] "well" "will" "right" "feet"
[8,] "box" "bought" "alway" "got"
[9,] "brand" "can" "anoth" "wide"
[10,] "ive" "love" "armor" "arriv"

```

Powyżej ukazane jest 10 najczęściej występujących słów w danym temacie. Analizując podane słowa postanowiłem nazwać je następująco:

Topic 1 - Doświadczenie z marką

Topic 2 – Rozmiar i dopasowanie

Topic 3 – Doświadczenia podczas użytkowania

Topic 4 – Stan wizualny i fizyczny

| | Topic 1 | Topic 2 | Topic 3 | Topic 4 |
|--------------|-----------------------|-----------------------|-----------------------------------|--------------------------|
| Nazwa | Doświadczenie z marką | Rozmiar i dopasowanie | Doświadczenia podczas użytkowania | Stan wizualny i fizyczny |
| Ilość opinii | 20 | 19 | 19 | 11 |

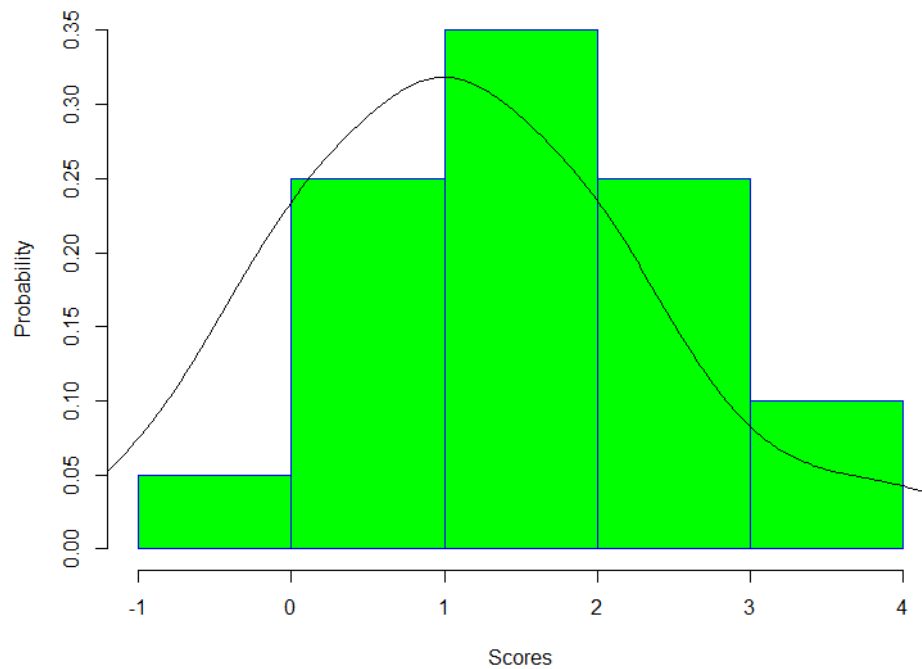
A. Topic 1: Doświadczenie z marką

```

Min. 1st Qu. Median Mean 3rd Qu. Max.
-1.00  0.00   1.00  1.15  2.00   4.00

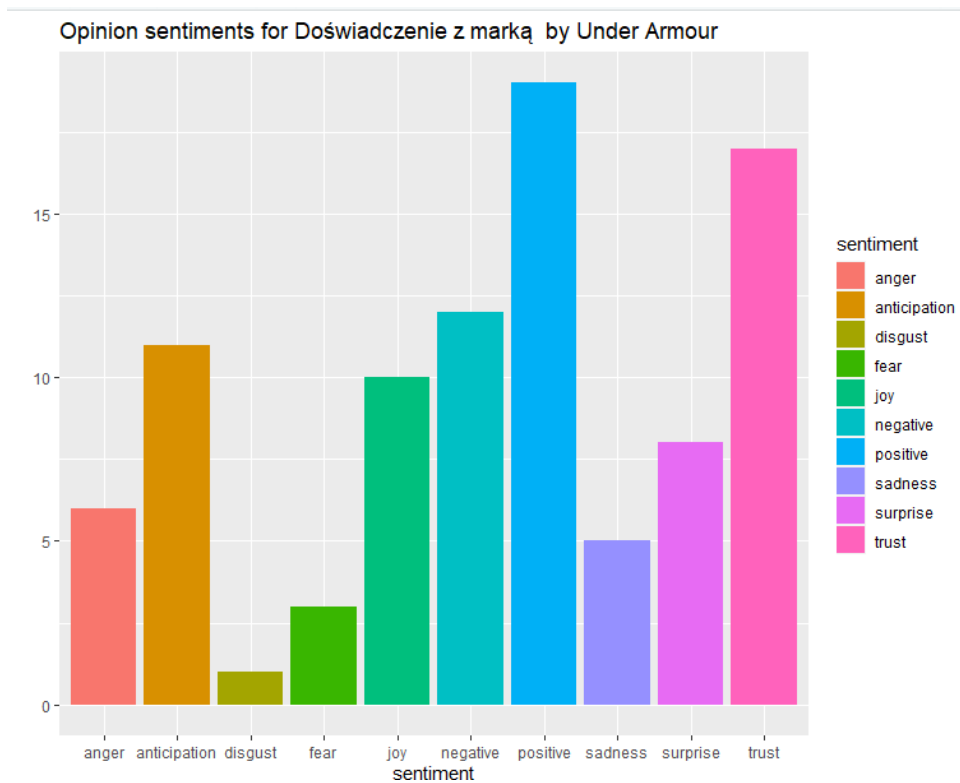
```

Histogram for the Sentiment by Doświadczenie z marką



```
> pos1$score
[1] 2 1 2 1 1 4 2 2 3 1 2 1 1 1
> length(pos1$score)
[1] 14
> neg1$score
[1] -1
> length(neg1$score)
[1] 1
> neu1$score
[1] 0 0 0 0 0
> length(neu1$score)
[1] 5
```

Opinie dla Tematu Doświadczenie z marką mają pozytywne nacechowanie. Pojawiło się 14 opinii pozytywnych, 5 neutralnych i 1 negatywna.

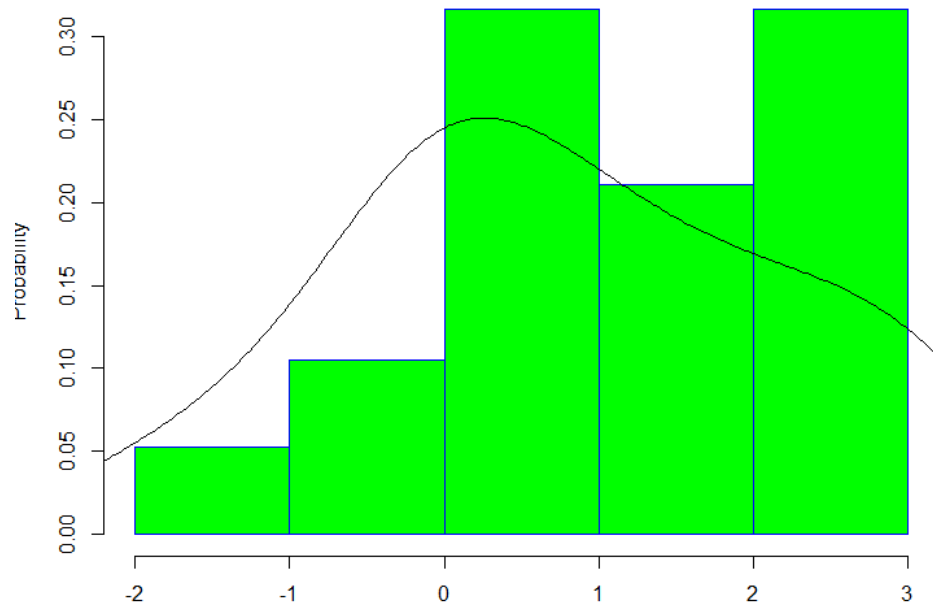


Klienci w swoich opiniach wyrażali głównie zaufanie, radość antycypację i złość.

B. Topic 2: Rozmiar i dopasowanie

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---------|---------|--------|--------|---------|--------|
| -2.0000 | 0.0000 | 1.0000 | 0.7895 | 2.0000 | 3.0000 |

Histogram for the Sentiment by Rozmiar i dopasowanie

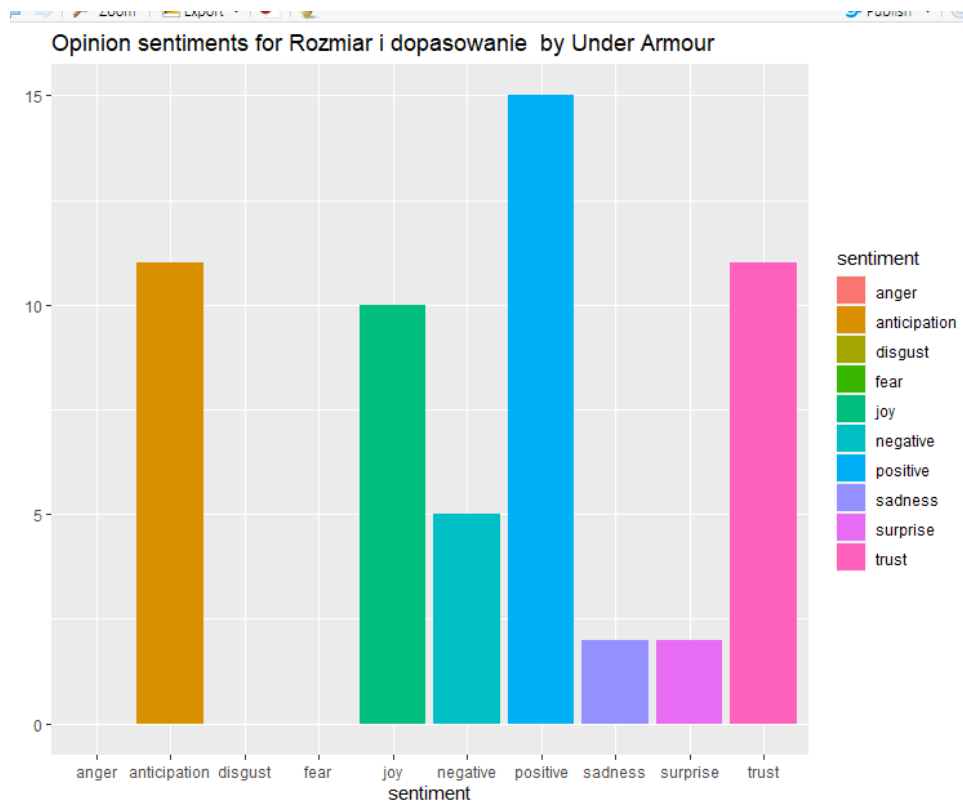


```
> pos1$Score
[1] 3 1 2 1 2 3 2 1 3 1
> length(pos1$Score)
[1] 10

> neu1$Score
[1] 0 0 0 0 0 0
> length(neu1$Score)
[1] 6

> neg1$Score
[1] -1 -1 -2
> length(neg1$Score)
[1] 3
```

Opinie dla Tematu Rozmiar i dopasowanie mają nieco bardziej zróżnicowane nacechowanie względem pozostałych tematów. Pojawiło się 10 opinii pozytywnych, 6 neutralnych i 3 negatywne.

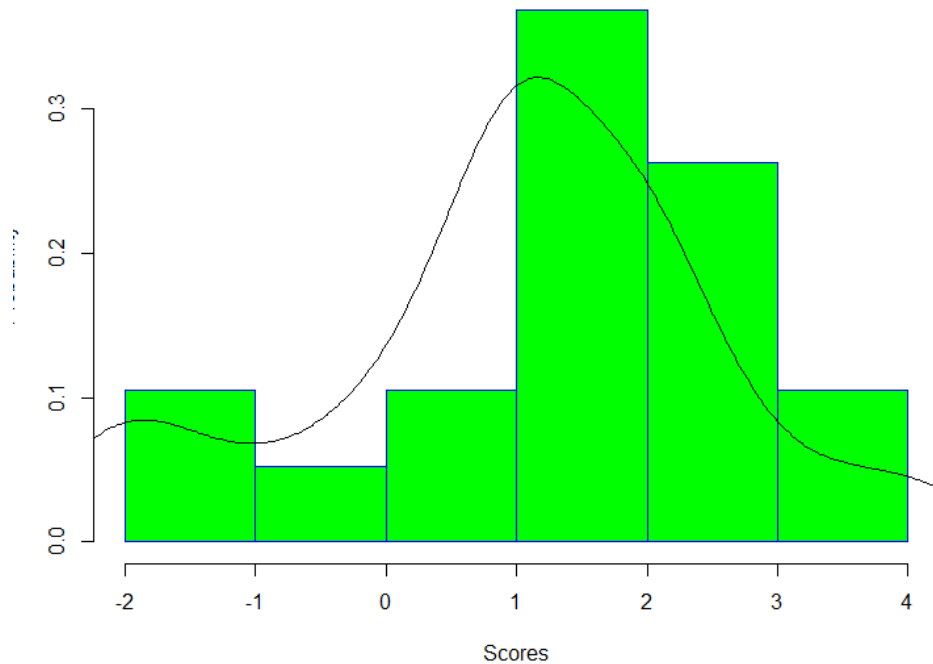


Powyżej, ukazane są emocje jakie wyrażali klienci poprzez słowa. Przede wszystkim wyrażali radość, zaufanie i antycypację.

C. Topic 3: Doświadczenia podczas użytkowania

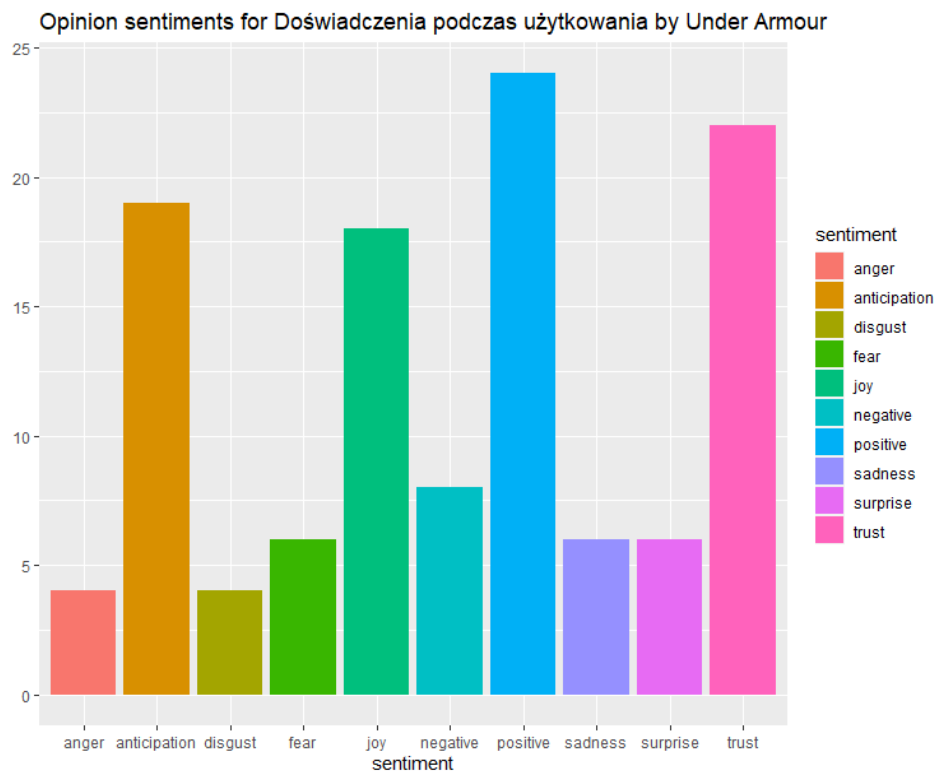
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|---------|--------|------|---------|------|
| -2.0 | 0.5 | 1.0 | 1.0 | 2.0 | 4.0 |

Histogram for the Sentiment by Doświadczenia podczas użytkowania



```
> pos1$Score
[1] 2 1 1 1 2 1 2 3 1 1 2 1 2 4
> length(pos1$Score)
[1] 14
> neg1$score
[1] -2 -2 -1
> length(neg1$score)
[1] 3
> neu1$score
[1] 0 0
> length(neu1$score)
[1] 2
```

Opinie dla Tematu Doświadczenia podczas użytkowania są głównie pozytywne. Pojawiło się 14 opinii pozytywnych, 2 neutralne i 3 negatywne.

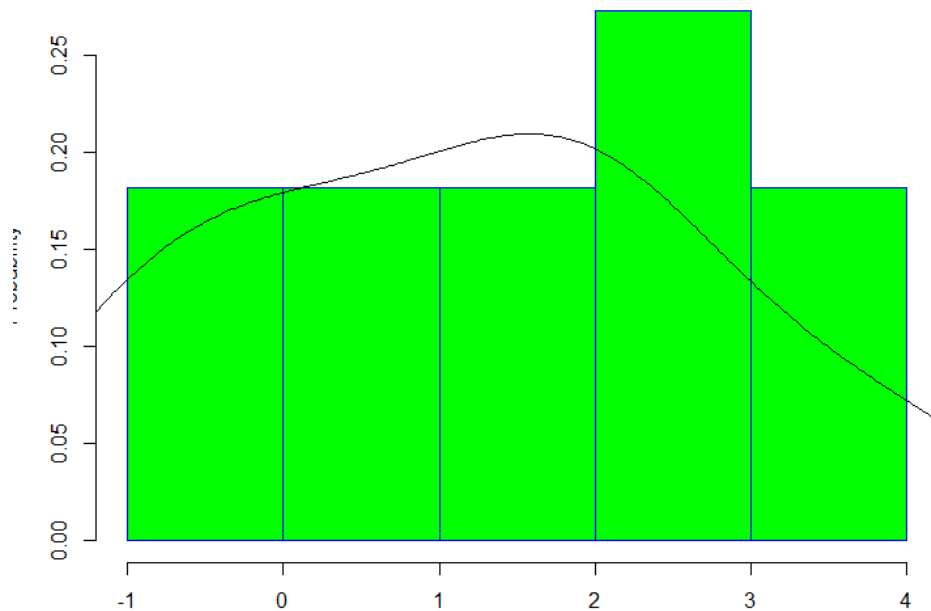


Klienci w swoich opiniach wyrażali przeważnie radość, zaufanie i antycypację. Niektórzy wyrażali złość, zaskoczenie, i obawę.

D. Topic 4: Stan wizualny i fizyczny

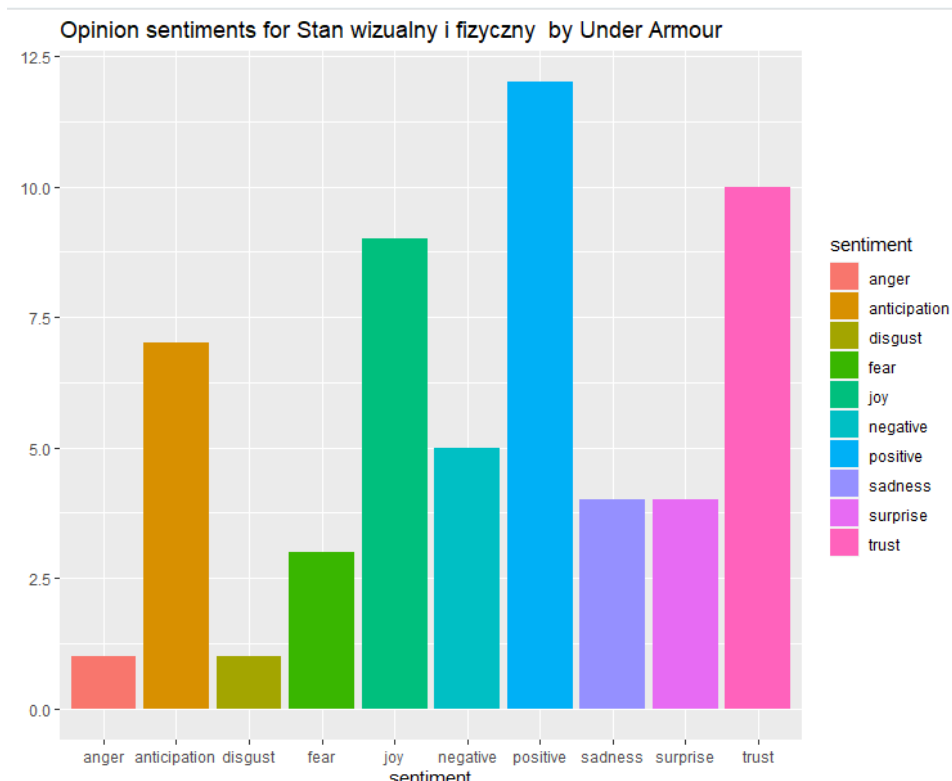
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--------|---------|--------|-------|---------|-------|
| -1.000 | 0.000 | 1.000 | 1.182 | 2.000 | 4.000 |

Histogram for the Sentiment by Stan wizualny i fizyczny



```
> pos1$score      > neu1$score      > neg1$score
[1] 2 3 2 4 2 1 1   [1] 0 0                        [1] -1 -1
> length(pos1$score) > length(neu1$score) > length(neg1$score)
[1] 7                [1] 2                        [1] 2
```

Opinie dla Tematu Stan wizualny i fizyczny mają nieco bardziej zróżnicowane nacechowanie względem pozostałych tematów. Pojawiło się 7 opinii pozytywnych, 2 neutralne i 2 negatywne.



W swoich opiniach klienci wyrażali głównie radość, zaufanie, antycypację, ale również smutek i zaskoczenie.

5. Podsumowanie i wnioski

Za obiekt mojej analizy obrałem sobie opinie klientów na temat najpopularniejszych modeli butów do biegania następujących marek: Nike, Adidas, New Balance, Puma, Under Armour. Z pozyskanych i przygotowanych danych udało mi się wyszczególnić cztery tematy, które pojawiały się wśród opinii konsumentów każdej z wyżej wymienionych firm. Poniżej przedstawiam jak często, poszczególne tematy znajdowały się w opiniach danych marek.

| | Stan wizualny i fizyczny | Doświadczenia podczas użytkowania | Rozmiar i dopasowanie | Doświadczenie z marką |
|--------------|--------------------------|-----------------------------------|-----------------------|-----------------------|
| Nike | 15 | 11 | 35 | 9 |
| Adidas | 16 | 16 | 16 | 22 |
| New Balance | 15 | 11 | 23 | 20 |
| Puma | 19 | 12 | 12 | 27 |
| Under Armour | 11 | 19 | 19 | 20 |
| Suma | 76 | 69 | 105 | 98 |

Z powyższej tabeli oraz dokładnej analizy poprzednich wizualizacji, można stwierdzić, że najczęściej poruszonymi tematami są: Rozmiar i dopasowanie oraz Doświadczenie z marką. To wszystko ma sens. Jako iż opinie były pozyskiwane z platformy Amazon.com, nic dziwnego, że klienci często w swoich opiniach poruszali kwestię rozmiaru buta. Brak możliwości przymierzenia obuwia przed zakupem on-line, zawsze budzi niepewność i skłania ludzi do podzielenia się doświadczeniami po zakupie, w celu pomocy innym konsumentom w odpowiednim doborze rozmiaru buta. Drugim z kolei najczęściej poruszonym tematem jest doświadczenie z marką. Jako iż mówimy o zakupach on-line, opinie te przede wszystkim zawierały informacje o stanie otrzymanego obuwia, czy jest dobrze zapakowane, czy jest uszkodzone oraz czy były opóźnienia w dostawie. Do tego tematu zaliczały się również opinie zawierające postawę konsumentów wobec danej marki. Przeważnie, pozytywnie wypowiadali się na temat swoich już kolejnych zakupów u tego producenta.

| | Pozytywne | Neutralne | Negatywne |
|--------------|-----------|-----------|-----------|
| Nike | 51 | 12 | 7 |
| Adidas | 50 | 14 | 6 |
| New Balance | 49 | 12 | 5 |
| Puma | 46 | 18 | 6 |
| Under Armour | 45 | 15 | 9 |

Z powyższej tabeli, można wyczytać liczbę pozytywnych, neutralnych i negatywnych opinii dla poszczególnych marek. Wyprowadzone przeze mnie hipotezy się potwierdziły. Nike ma najwięcej ocen pozytywnych, a Under Armour najwięcej negatywnych. Z tabeli wynika również to, że opinie pozostawiają głównie zadowoleni klienci. Rzadko ktoś pozostawia same negatywne komentarze. Dla każdej marki z wyłączeniem Under Armour'a, negatywne opinie to mniej niż 10%, co świadczy o wysokim poziomie produktów oraz o tym, że produkty się sprawdzają.