



INSIGHT REPORT BUSINESS

Text Analytics

PROFESSOR: THOMAS KURNICKI

SUBMITTED BY: SKANDER DRISS

1. Overview:

Records labels are the most important player in the music industry. They are responsible for finding musical talent- Based on several characteristics for example some singer have a unique voice and other have high writing songs skills. Afterwards they record their works with a strict guidance and sell it to retail store or to music streaming stores (IBISWorld).

The music production industry has a total revenue of \$8bn in 2019 with an annual growth of 8%.

SWOT Analysis

Strength:

- High and steady barrier to enter
- Low customer concentration
- High labor efficiency

Weaknesses:

- High Competition
- Low profit
- High product concentration

Opportunities:

- Outlier growth
- High performance drivers

Threats:

- Very low long-term growth
- Number of broadband connections

In this report, I will be providing several insights and thoughts to a music records label based on an analysis that I will conduct.

To start my analysis I choose to extract data from the albums of the top three rappers in the USA

Kendrick Lamar, Eminem and Migos.

Using the library genius allowed me to dive into their albums and lyrics.

2. Word Cloud:

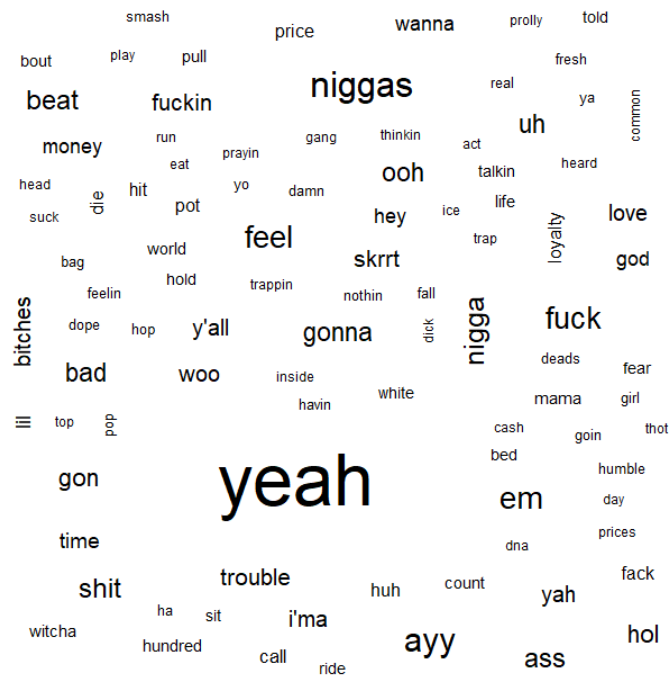


Figure 1: Word Cloud

When we have a look at the word cloud of these songs, we could notice that the biggest word here is yeah. All the rappers use those kind of words such as ‘yeah’, ‘Ayy’, ‘Yah’, ‘Woo’ as a link between the rhymes. In addition, these words help the rappers to think about the next rhyme while jamming and writing a song and when it goes with the flow of the song, they keep it in the lyrics. In addition, here we can see the word ‘skrrrt’ that was first said by Cardi B. Now it is a buzzword that all singers use.

Besides, we notice here that there are too many street and bad words such as ‘B****’, ‘N***’ ‘F***’ etc.

These words are used to show that they are powerful and to impose their authority as Bad Boys.

To sum up here we could encourage a new growing rapper to try to invent new words and use them as coordinators between lines of lyrics.

3. Sentiment Analysis:

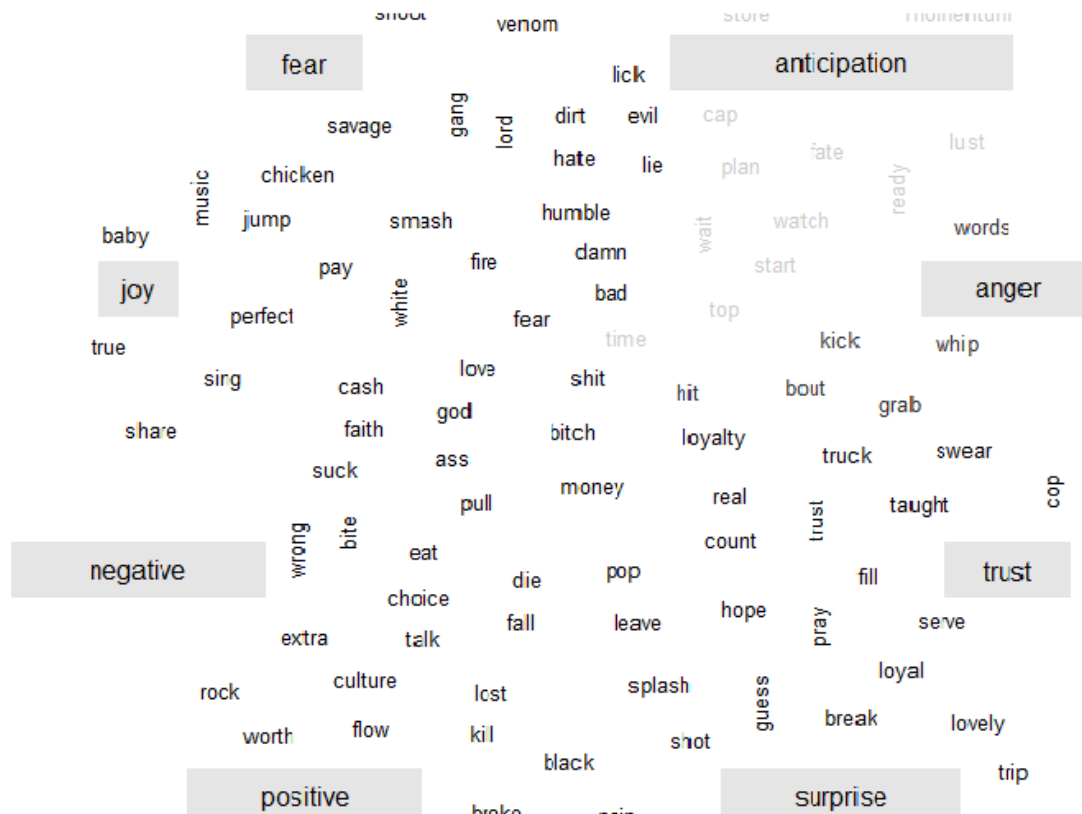


Figure 2: NRC Sentiments

As you can see in the figure above, rappers do not show any specific feeling while writing a song. Artists have minds that work very different from a regular person. That is the reason why we cannot understand their actions or follow their emotions. Joining the NRC sentiments with the tokens extracted from the albums shows how these words are spread from anger to joy, trust to fear and so on ...

A great business insight that we could come up with here is that an artist should never show a unique feeling in a song. In order to have a successful one, the rapper should differ his feelings from sadness to happiness. Artists do not expect to be understood in fact they prefer not to be. They are living in a world of their own where feelings are just word that they use to create a confusion among the listeners (15 Reasons Why Artistic People are Difficult to Understand).

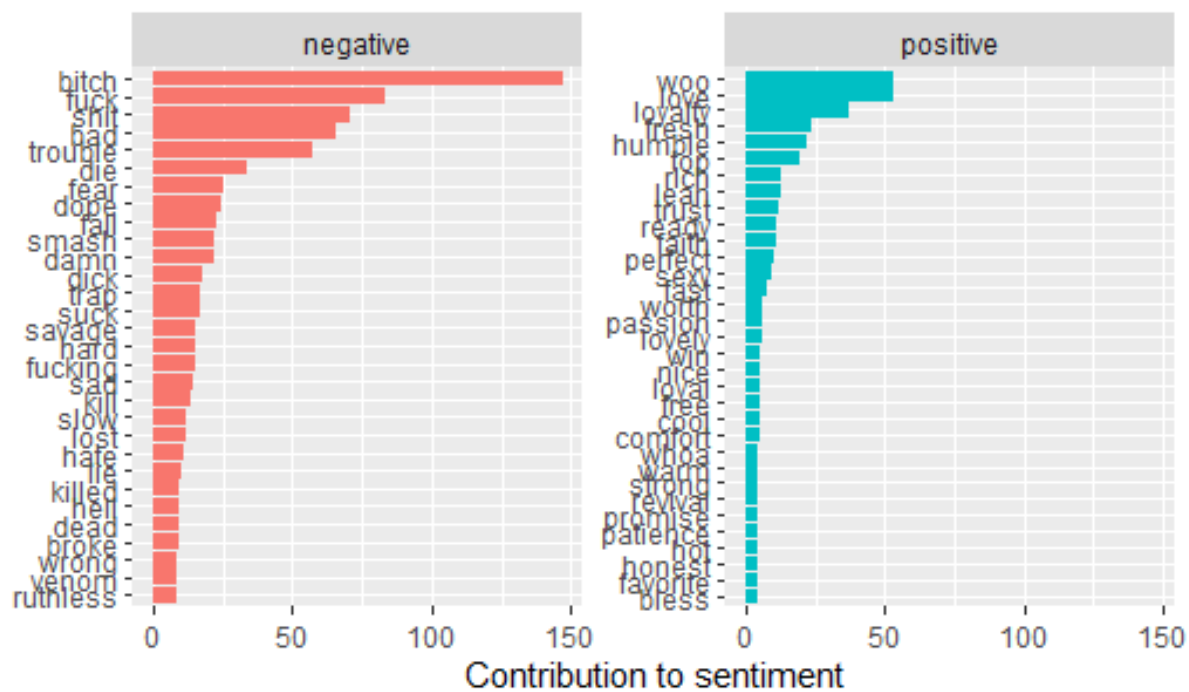


Figure 3: Bing Sentiments

When it comes the Bing sentiments, we notice that all the tokens extracted from the lyrics vary between positive and negative. As it was mentioned in the analysis above, usually artist do not have a stable feeling. In fact, if we dig a little bit deeper in this analysis we could admit that from the text that we have, rappers seem to be more negative than positive. Out of 513, 371 tokens are negative, which represent around 70% of the total words.

To sum up here, a successful rapper could have negative attitude and that is because due to the harsh condition of living or it may be because of the people that they are meeting with every day.

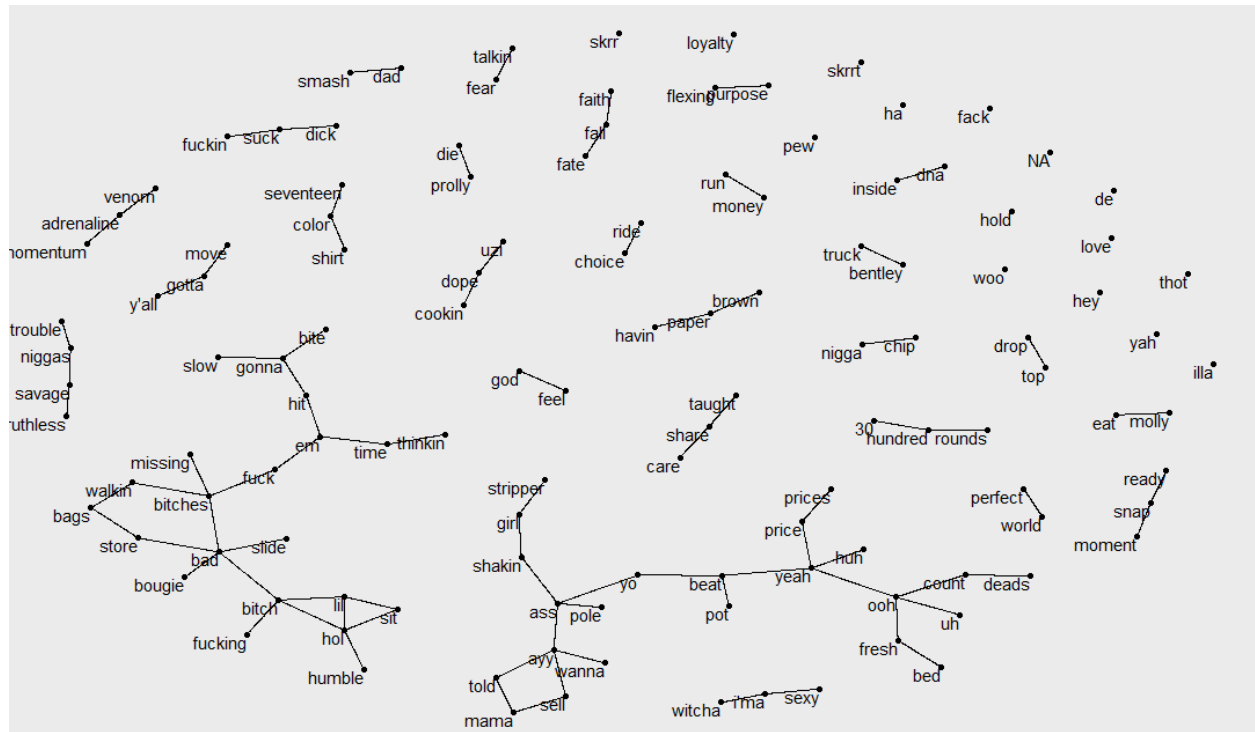


Figure 4: Bigram Graph

Last but not least, here is a bigram of the words that the rappers of these days are using together. I would like to suggest couple of recommendations extracted from the graph above:

- Run money
- Perfect world
- Feel god
- Ride choice
- Beat pot
- Share care
- Drop top

4. Conclusion:

In order to be able to conduct this analysis, I have used couple of frameworks that we covered in class. Tidy text, sentiment lexicon, Word cloud and bigram. I have also included tf-idf in my code However i was not able to generate the appropriate graph.

5. Appendix Code:

```
library(textreadr)
library(dplyr)
library(reshape2)
# text mining library
library(tidytext)
# plotting packages
library(igraph)
library(ggraph)
library(ggplot2)
#Corpus
library(tm)
#Word cloud
library(wordcloud)
#lyrics
library(genius)

#eminem Album
eminem <- genius_album(artist = "Eminem", album = "Kamikaze")
tidy_emi <- eminem %>%
  unnest_tokens(word, lyric) %>%
  anti_join(stop_words)

#Kendrick Album
Kendrick <- genius_album(artist = "Kendrick Lamar", album = "DAMN")
tidy_ken <- Kendrick %>%
  unnest_tokens(word, lyric) %>%
```

```

anti_join(stop_words)

#migos Album
migos <- genius_album(artist = "Migos", album = "Culture")
tidy_mig <- migos %>%
  unnest_tokens(word, lyric) %>%
  anti_join(stop_words)

### Combining all 3 tidy data frames and creating correlograms

df <- bind_rows(mutate(tidy_emi, author="Eminem"),
  mutate(tidy_ken, author= "Kendrick"),
  mutate(tidy_mig, author="Migos")) #closing bind_rows

#Frequency of the words
frequency <- df %>%
  unnest_tokens(word, word)%>%
  count(word, sort=T)

#World cloud
frequency %>%
  with(wordcloud(word, n, max.words = 100))

#sentiment NRC
df_nrc <- df %>%
  inner_join(get_sentiments("nrc")) %>%
  count(word, sentiment, sort=TRUE) %>%
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("grey20", "gray80"),
    max.words=150,

```



```

        scale=c(0.8,0.8),

        fixed.asp=TRUE,

        title.size = 1)

#sentiment bing
df_bing <- df %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=T) %>%
  ungroup() #%>%

#positive and negative graph
df_bing %>%
  group_by(sentiment) %>%
  top_n(30) %>%
  ungroup() %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="Contribution to sentiment", x=NULL)+
  coord_flip()

#number of negative words
df_bing_neg <- df %>%
  unnest_tokens(word, word) %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=T) %>%
  ungroup() %>%
  filter(sentiment == 'negative')

#number of positive words

```

```

df_bing_pos <- df %>%
  unnest_tokens(word, word) %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=T) %>%
  ungroup() %>%
  filter(sentiment == 'positive')

df %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=TRUE) %>%
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("grey20", "black"),
    max.words=100,
    scale=c(1,1),
    fixed.asp=TRUE,
    title.size =2
  )

##### TF_IDF #####

rap_tfidf<- inner_join(df, frequency,by= 'word')

rap_tfidf2 <- rap_tfidf %>%
  bind_tf_idf(word, author, n)%>%
  arrange(desc(tf_idf))%>%
  filter (tf_idf > 0.0000001)

##### Bigram #####

```

```

my_bigrams <- df %>%
  unnest_tokens(bigram, word, token = "ngrams", n=2)
bigrams_separated <- my_bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ")
bigram_counts <- bigrams_separated%>%
  count(word1, word2, sort = TRUE)

bigram_graph <- bigram_counts %>%
  filter(n>5) %>%
  graph_from_data_frame()

ggraph(bigram_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)

```

6. References:

IbisWorld- Retrieved from:

<https://my-ibisworld-com.hult.idm.oclc.org/us/en/industry/51222/industry-at-a-glance>

Alex Filipovic- 15 Reasons Why Artistic People are Difficult to Understand. Retrieved from

<https://www.lifehack.org/articles/communication/15-reasons-why-artistic-people-are-difficult-understand.html>