Text, Web and Social Media Analytics

Title

Sentiment analysis on billion song lyrics

# Description:

This analysis will identify the most common words used in the songs in the data set, as well as use sentiment analysis to identify the most positive and negative songs using the sentiment package. In addition, we can identify the artists with the highest and lowest average sentiment scores across their music.

# Context:

These are the lyrics for 57650 songs. They can be used for Natural Language Processing purposes, such as clustering of the words with similar meanings or predicting artist by the song.

# Content:

The dataset contains 4 columns:

1. Artist
2. Song Name
3. Link to a webpage with the song (for reference). This is to be concatenated with [*http://www.lyricsfreak.com*](http://www.lyricsfreak.com/) to form a real URL.
4. Lyrics of the song, unmodified.

# Dataset:

Songdata.csv

# R Code:

#load data

songdata <- read.csv("Enter the path /songdata.csv") #clean data

songdata$text <- tolower(songdata$text)

songdata$text <- gsub("[^0-9A-Za-z///' ]", "", songdata$text) #Build a corpus

library(tm)

Corpus <- Corpus(VectorSource(songdata$text)) #remove puctuation

Corpus <- tm\_map(Corpus, removePunctuation)

#remove numbers

Corpus <- tm\_map(Corpus, removeNumbers) #strip white space

Corpus <- tm\_map(Corpus, stripWhitespace) #remove stopwords

myStopwords <- c(stopwords('english'))

Corpus <- tm\_map(Corpus, removeWords, myStopwords) #wordcloud of top 100 words among all songs library(wordcloud)

wordcloud(Corpus, scale=c(3,.2), max.words=100, random.order=FALSE, rot.per=0.25, use.r.layout=FALSE, colors=brewer.pal(8, "Dark2"))



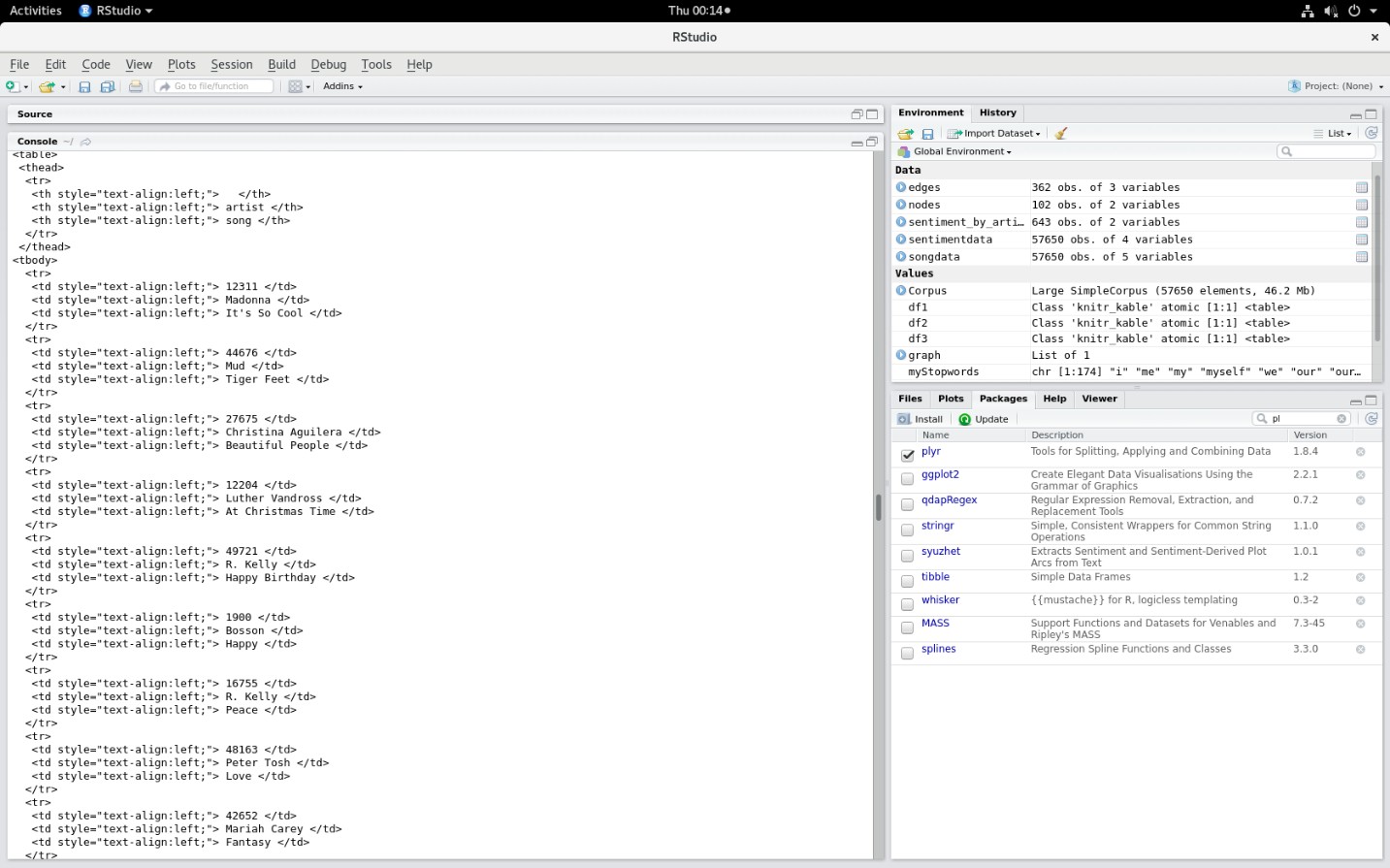
#sentiment analysis by polarity of song library(sentimentr)

sentimentdata <- sentimentr::sentiment(songdata$text) #25 most positive songs

writeLines("td, th { padding : 6px } th { background-color : brown ; color : white; border : 1px solid white; } td { color : brown ; border : 1px solid brown }", con = "mystyle.css")

df1 <- knitr::kable(head(songdata[order(sentimentdata$sentiment, decreasing = T),c(1,2)], 25), format = "html")

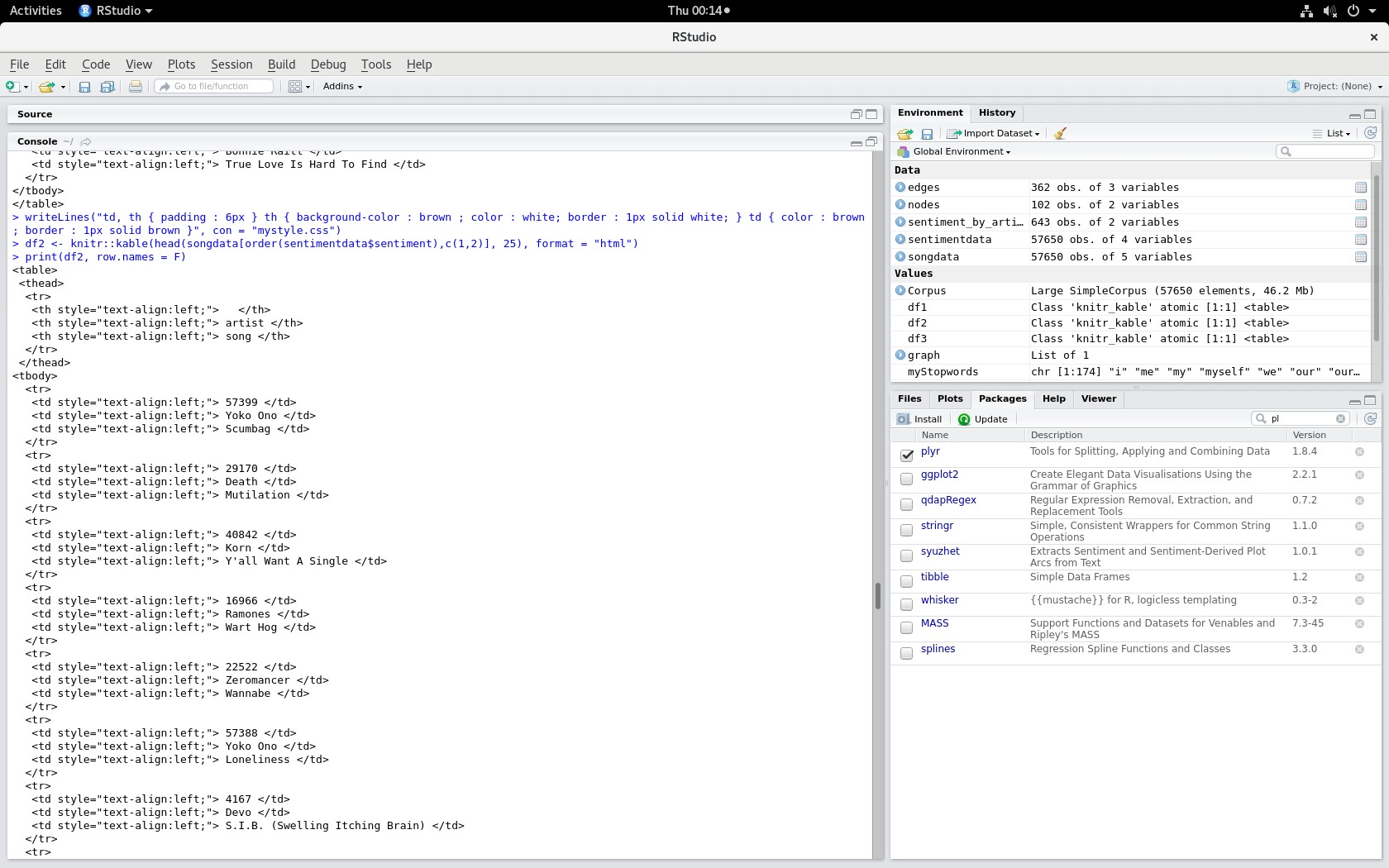
row.names(df1) <- NULL print(df1, row.names = F)



#25 most negative songs

writeLines("td, th { padding : 6px } th { background-color : brown ; color : white; border : 1px solid white; } td { color : brown ; border : 1px solid brown }", con = "mystyle.css")

df2 <- knitr::kable(head(songdata[order(sentimentdata$sentiment),c(1,2)], 25), format = "html") print(df2, row.names = F)



#most positive artists

songdata$sentiment <- sentimentdata$sentiment

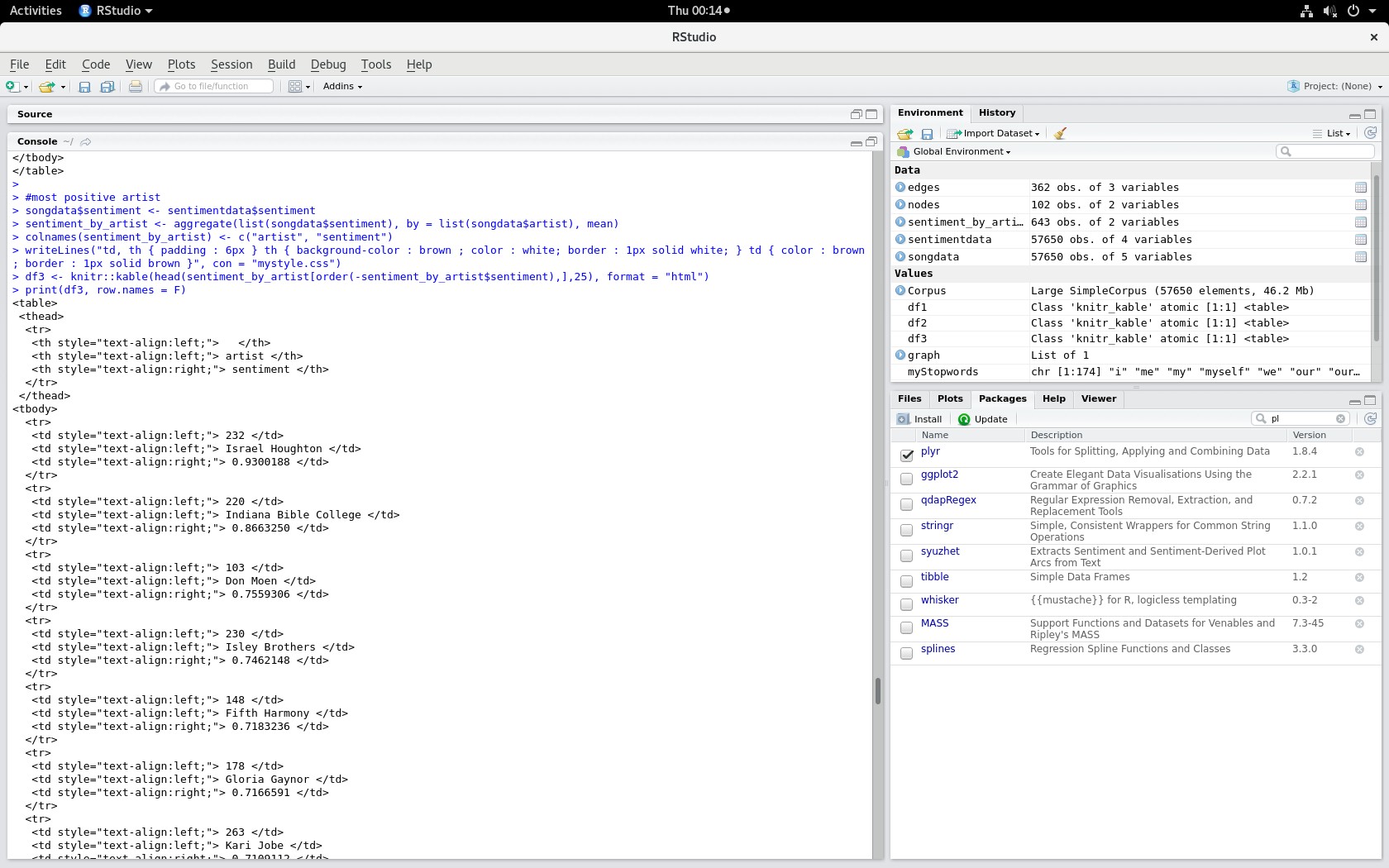
sentiment\_by\_artist <- aggregate(list(songdata$sentiment), by = list(songdata$artist), mean) colnames(sentiment\_by\_artist) <- c("artist", "sentiment")

writeLines("td, th { padding : 6px } th { background-color : brown ; color : white; border : 1px solid white; } td { color : brown ; border : 1px solid brown }", con = "mystyle.css")

df3 <- knitr::kable(head(sentiment\_by\_artist[order(-sentiment\_by\_artist$sentiment),],25), format

= "html")

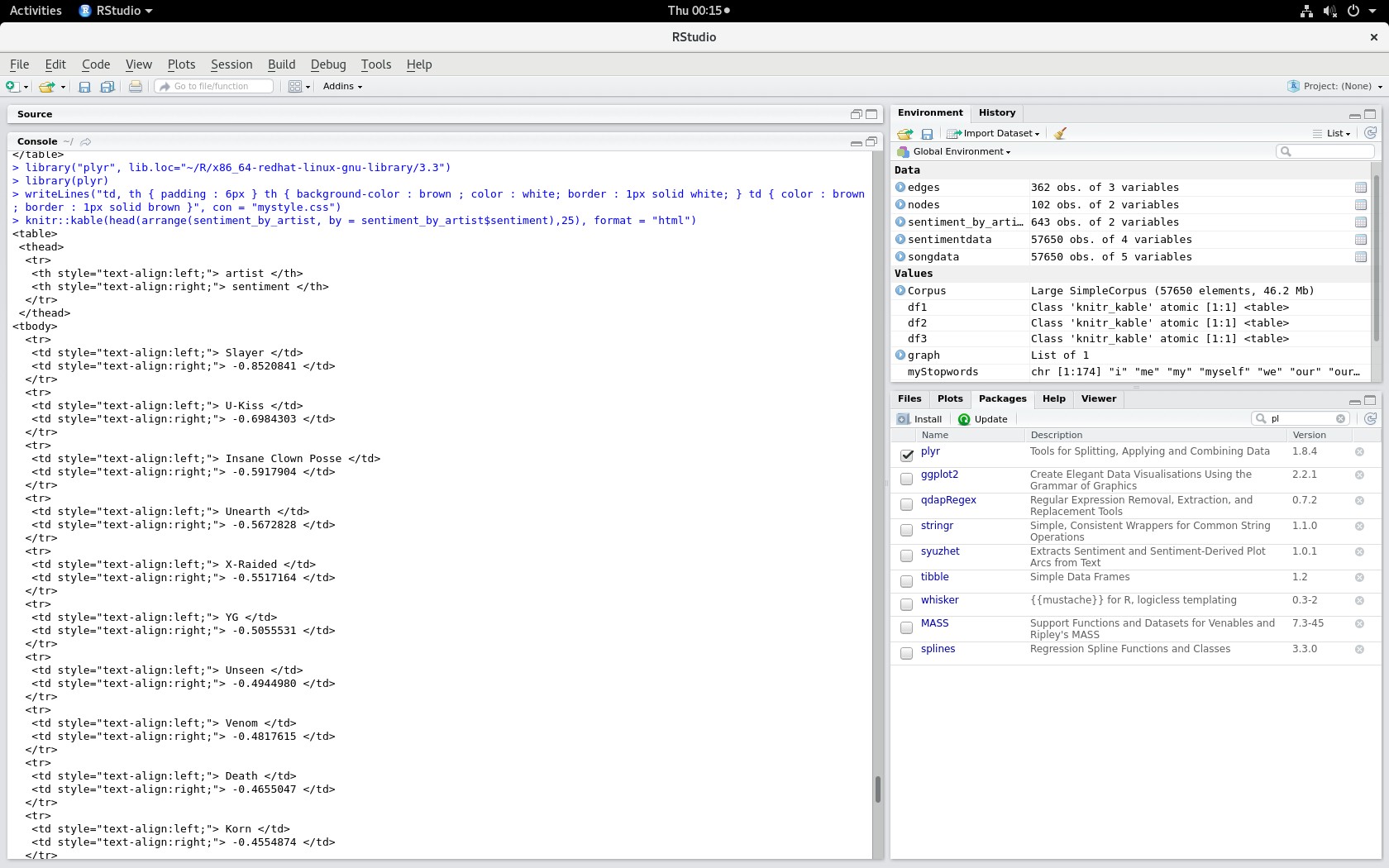
print(df3, row.names = F)



#most negative artists library(plyr)

writeLines("td, th { padding : 6px } th { background-color : brown ; color : white; border : 1px solid white; } td { color : brown ; border : 1px solid brown }", con = "mystyle.css")

knitr::kable(head(arrange(sentiment\_by\_artist, by = sentiment\_by\_artist$sentiment),25), format = "html")



# Output:

#25 most positive songs

<table>

<thead>

<tr>

<th style="text-align:left;"> </th>

<th style="text-align:left;"> artist </th>

<th style="text-align:left;"> song </th>

</tr>

</thead>

<tbody>

<tr>

<td style="text-align:left;"> 12311 </td>

<td style="text-align:left;"> Madonna </td>

<td style="text-align:left;"> It's So Cool </td>

</tr>

<tr>

<td style="text-align:left;"> 44676 </td>

<td style="text-align:left;"> Mud </td>

<td style="text-align:left;"> Tiger Feet </td>

</tr>

<tr>

<td style="text-align:left;"> 27675 </td>

<td style="text-align:left;"> Christina Aguilera </td>

<td style="text-align:left;"> Beautiful People </td>

</tr>

<tr>

<td style="text-align:left;"> 12204 </td>

<td style="text-align:left;"> Luther Vandross </td>

<td style="text-align:left;"> At Christmas Time </td>

</tr>

<tr>

<td style="text-align:left;"> 49721 </td>

<td style="text-align:left;"> R. Kelly </td>

<td style="text-align:left;"> Happy Birthday </td>

</tr>

<tr>

<td style="text-align:left;"> 1900 </td>

<td style="text-align:left;"> Bosson </td>

<td style="text-align:left;"> Happy </td>

</tr>

<tr>

<td style="text-align:left;"> 16755 </td>

<td style="text-align:left;"> R. Kelly </td>

<td style="text-align:left;"> Peace </td>

</tr>

<tr>

<td style="text-align:left;"> 48163 </td>

<td style="text-align:left;"> Peter Tosh </td>

<td style="text-align:left;"> Love </td>

</tr>

<tr>

<td style="text-align:left;"> 42652 </td>

<td style="text-align:left;"> Mariah Carey </td>

<td style="text-align:left;"> Fantasy </td>

</tr>

<tr>

<td style="text-align:left;"> 42653 </td>

<td style="text-align:left;"> Mariah Carey </td>

<td style="text-align:left;"> Fantasy (Bad Boy Fantasy) </td>

</tr>

<tr>

<td style="text-align:left;"> 8789 </td>

<td style="text-align:left;"> Isley Brothers </td>

<td style="text-align:left;"> Freedom </td>

</tr>

<tr>

<td style="text-align:left;"> 26999 </td>

<td style="text-align:left;"> Cat Stevens </td>

<td style="text-align:left;"> Ready </td>

</tr>

<tr>

<td style="text-align:left;"> 36767 </td>

<td style="text-align:left;"> Indiana Bible College </td>

<td style="text-align:left;"> Wonderful Savior </td>

</tr>

<tr>

<td style="text-align:left;"> 17864 </td>

<td style="text-align:left;"> Rod Stewart </td>

<td style="text-align:left;"> It Was Love That We Needed </td>

</tr>

<tr>

<td style="text-align:left;"> 24634 </td>

<td style="text-align:left;"> Beach Boys </td>

<td style="text-align:left;"> Dance, Dance, Dance </td>

</tr>

<tr>

<td style="text-align:left;"> 40631 </td>

<td style="text-align:left;"> Kiss </td>

<td style="text-align:left;"> Do You Love Me </td>

</tr>

<tr>

<td style="text-align:left;"> 56616 </td>

<td style="text-align:left;"> Wilson Phillips </td>

<td style="text-align:left;"> Dance Dance Dance </td>

</tr>

<tr>

<td style="text-align:left;"> 55463 </td>

<td style="text-align:left;"> Wang Chung </td>

<td style="text-align:left;"> True Love </td>

</tr>

<tr>

<td style="text-align:left;"> 44255 </td>

<td style="text-align:left;"> Modern Talking </td>

<td style="text-align:left;"> Nothing But The Truth </td>

</tr>

<tr>

<td style="text-align:left;"> 56269 </td>

<td style="text-align:left;"> Whitney Houston </td>

<td style="text-align:left;"> Joy To The World </td>

</tr>

<tr>

<td style="text-align:left;"> 4405 </td>

<td style="text-align:left;"> Donna Summer </td>

<td style="text-align:left;"> Love To Love You Baby </td>

</tr>

<tr>

<td style="text-align:left;"> 41481 </td>

<td style="text-align:left;"> Leo Sayer </td>

<td style="text-align:left;"> How Much Love </td>

</tr>

<tr>

<td style="text-align:left;"> 13504 </td>

<td style="text-align:left;"> Natalie Cole </td>

<td style="text-align:left;"> The Winner </td>

</tr>

<tr>

<td style="text-align:left;"> 24143 </td>

<td style="text-align:left;"> Andy Williams </td>

<td style="text-align:left;"> Promise Me, Love </td>

</tr>

<tr>

<td style="text-align:left;"> 26323 </td>

<td style="text-align:left;"> Bonnie Raitt </td>

<td style="text-align:left;"> True Love Is Hard To Find </td>

</tr>

</tbody>

</table>

#25 most negative songs

<table>

<thead>

<tr>

<th style="text-align:left;"> </th>

<th style="text-align:left;"> artist </th>

<th style="text-align:left;"> song </th>

</tr>

</thead>

<tbody>

<tr>

<td style="text-align:left;"> 57399 </td>

<td style="text-align:left;"> Yoko Ono </td>

<td style="text-align:left;"> Scumbag </td>

</tr>

<tr>

<td style="text-align:left;"> 29170 </td>

<td style="text-align:left;"> Death </td>

<td style="text-align:left;"> Mutilation </td>

</tr>

<tr>

<td style="text-align:left;"> 40842 </td>

<td style="text-align:left;"> Korn </td>

<td style="text-align:left;"> Y'all Want A Single </td>

</tr>

<tr>

<td style="text-align:left;"> 16966 </td>

<td style="text-align:left;"> Ramones </td>

<td style="text-align:left;"> Wart Hog </td>

</tr>

<tr>

<td style="text-align:left;"> 22522 </td>

<td style="text-align:left;"> Zeromancer </td>

<td style="text-align:left;"> Wannabe </td>

</tr>

<tr>

<td style="text-align:left;"> 57388 </td>

<td style="text-align:left;"> Yoko Ono </td>

<td style="text-align:left;"> Loneliness </td>

</tr>

<tr>

<td style="text-align:left;"> 4167 </td>

<td style="text-align:left;"> Devo </td>

<td style="text-align:left;"> S.I.B. (Swelling Itching Brain) </td>

</tr>

<tr>

<td style="text-align:left;"> 36991 </td>

<td style="text-align:left;"> Insane Clown Posse </td>

<td style="text-align:left;"> My Axe </td>

</tr>

<tr>

<td style="text-align:left;"> 51826 </td>

<td style="text-align:left;"> Snoop Dogg </td>

<td style="text-align:left;"> Gangsta Walk </td>

</tr>

<tr>

<td style="text-align:left;"> 20903 </td>

<td style="text-align:left;"> Violent Femmes </td>

<td style="text-align:left;"> I'm Bad </td>

</tr>

<tr>

<td style="text-align:left;"> 42045 </td>

<td style="text-align:left;"> Lou Reed </td>

<td style="text-align:left;"> Blind Rage </td>

</tr>

<tr>

<td style="text-align:left;"> 40670 </td>

<td style="text-align:left;"> Kiss </td>

<td style="text-align:left;"> Killer </td>

</tr>

<tr>

<td style="text-align:left;"> 25908 </td>

<td style="text-align:left;"> Bob Marley </td>

<td style="text-align:left;"> Jungle Fever </td>

</tr>

<tr>

<td style="text-align:left;"> 55897 </td>

<td style="text-align:left;"> Weird Al Yankovic </td>

<td style="text-align:left;"> Fat </td>

</tr>

<tr>

<td style="text-align:left;"> 29586 </td>

<td style="text-align:left;"> Depeche Mode </td>

<td style="text-align:left;"> Wrong </td>

</tr>

<tr>

<td style="text-align:left;"> 51734 </td>

<td style="text-align:left;"> Slayer </td>

<td style="text-align:left;"> World Painted Blood </td>

</tr>

<tr>

<td style="text-align:left;"> 19118 </td>

<td style="text-align:left;"> System Of A Down </td>

<td style="text-align:left;"> Psycho </td>

</tr>

<tr>

<td style="text-align:left;"> 43977 </td>

<td style="text-align:left;"> Migos </td>

<td style="text-align:left;"> Contraband </td>

</tr>

<tr>

<td style="text-align:left;"> 20774 </td>

<td style="text-align:left;"> Veruca Salt </td>

<td style="text-align:left;"> Centipede </td>

</tr>

<tr>

<td style="text-align:left;"> 45391 </td>

<td style="text-align:left;"> Neil Young </td>

<td style="text-align:left;"> Cry, Cry, Cry </td>

</tr>

<tr>

<td style="text-align:left;"> 12720 </td>

<td style="text-align:left;"> Megadeth </td>

<td style="text-align:left;"> Liar </td>

</tr>

<tr>

<td style="text-align:left;"> 200 </td>

<td style="text-align:left;"> Aerosmith </td>

<td style="text-align:left;"> Shame, Shame, Shame </td>

</tr>

<tr>

<td style="text-align:left;"> 22060 </td>

<td style="text-align:left;"> Yngwie Malmsteen </td>

<td style="text-align:left;"> Legion Of The Damned </td>

|  |
| --- |
| </tr>  <tr>  <td style="text-align:left;"> 47279 </td>  <td style="text-align:left;"> Ozzy Osbourne </td>  <td style="text-align:left;"> Psycho Man </td>  </tr>  <tr>  <td style="text-align:left;"> 2711 </td>  <td style="text-align:left;"> Chris Brown </td>  <td style="text-align:left;"> 100 Bottles </td>  </tr>  </tbody>  </table> |
| > |

#most positive artist

<table>

<thead>

<tr>

<th style="text-align:left;"> </th>

<th style="text-align:left;"> artist </th>

<th style="text-align:right;"> sentiment </th>

</tr>

</thead>

<tbody>

<tr>

<td style="text-align:left;"> 232 </td>

<td style="text-align:left;"> Israel Houghton </td>

<td style="text-align:right;"> 0.9300188 </td>

</tr>

<tr>

<td style="text-align:left;"> 220 </td>

<td style="text-align:left;"> Indiana Bible College </td>

<td style="text-align:right;"> 0.8663250 </td>

</tr>

<tr>

<td style="text-align:left;"> 103 </td>

<td style="text-align:left;"> Don Moen </td>

<td style="text-align:right;"> 0.7559306 </td>

</tr>

<tr>

<td style="text-align:left;"> 230 </td>

<td style="text-align:left;"> Isley Brothers </td>

<td style="text-align:right;"> 0.7462148 </td>

</tr>

<tr>

<td style="text-align:left;"> 148 </td>

<td style="text-align:left;"> Fifth Harmony </td>

<td style="text-align:right;"> 0.7183236 </td>

</tr>

<tr>

<td style="text-align:left;"> 178 </td>

<td style="text-align:left;"> Gloria Gaynor </td>

<td style="text-align:right;"> 0.7166591 </td>

</tr>

<tr>

<td style="text-align:left;"> 263 </td>

<td style="text-align:left;"> Kari Jobe </td>

<td style="text-align:right;"> 0.7109112 </td>

</tr>

<tr>

<td style="text-align:left;"> 447 </td>

<td style="text-align:left;"> Religious Music </td>

<td style="text-align:right;"> 0.7092089 </td>

</tr>

<tr>

<td style="text-align:left;"> 322 </td>

<td style="text-align:left;"> Matt Redman </td>

<td style="text-align:right;"> 0.7079324 </td>

</tr>

<tr>

<td style="text-align:left;"> 231 </td>

<td style="text-align:left;"> Israel </td>

<td style="text-align:right;"> 0.6992021 </td>

</tr>

<tr>

<td style="text-align:left;"> 200 </td>

<td style="text-align:left;"> Hillsong </td>

<td style="text-align:right;"> 0.6966906 </td>

</tr>

<tr>

<td style="text-align:left;"> 464 </td>

<td style="text-align:left;"> Selah </td>

<td style="text-align:right;"> 0.6481213 </td>

</tr>

<tr>

<td style="text-align:left;"> 433 </td>

<td style="text-align:left;"> R. Kelly </td>

<td style="text-align:right;"> 0.6379706 </td>

</tr>

<tr>

<td style="text-align:left;"> 614 </td>

<td style="text-align:left;"> Yolanda Adams </td>

<td style="text-align:right;"> 0.6081804 </td>

</tr>

<tr>

<td style="text-align:left;"> 310 </td>

<td style="text-align:left;"> Luther Vandross </td>

<td style="text-align:right;"> 0.5972277 </td>

</tr>

<tr>

<td style="text-align:left;"> 341 </td>

<td style="text-align:left;"> Natalie Cole </td>

<td style="text-align:right;"> 0.5841289 </td>

</tr>

<tr>

<td style="text-align:left;"> 69 </td>

<td style="text-align:left;"> Christmas Songs </td>

<td style="text-align:right;"> 0.5833079 </td>

</tr>

<tr>

<td style="text-align:left;"> 299 </td>

<td style="text-align:left;"> Lionel Richie </td>

<td style="text-align:right;"> 0.5755303 </td>

</tr>

<tr>

<td style="text-align:left;"> 158 </td>

<td style="text-align:left;"> Frankie Valli </td>

<td style="text-align:right;"> 0.5732300 </td>

</tr>

<tr>

<td style="text-align:left;"> 201 </td>

<td style="text-align:left;"> Hillsong United </td>

<td style="text-align:right;"> 0.5668084 </td>

</tr>

<tr>

<td style="text-align:left;"> 592 </td>

<td style="text-align:left;"> X-Treme </td>

<td style="text-align:right;"> 0.5613414 </td>

</tr>

<tr>

<td style="text-align:left;"> 219 </td>

<td style="text-align:left;"> Independence Day </td>

<td style="text-align:right;"> 0.5528929 </td>

</tr>

<tr>

<td style="text-align:left;"> 545 </td>

<td style="text-align:left;"> Vanessa Williams </td>

<td style="text-align:right;"> 0.5521797 </td>

</tr>

<tr>

<td style="text-align:left;"> 381 </td>

<td style="text-align:left;"> Otis Redding </td>

<td style="text-align:right;"> 0.5509968 </td>

</tr>

|  |
| --- |
| <tr>  <td style="text-align:left;"> 472 </td>  <td style="text-align:left;"> Soundtracks </td>  <td style="text-align:right;"> 0.5506100 </td>  </tr>  </tbody>  </table> |
| > |

#most negative artist

<table>

<thead>

<tr>

<th style="text-align:left;"> artist </th>

<th style="text-align:right;"> sentiment </th>

</tr>

</thead>

<tbody>

<tr>

<td style="text-align:left;"> Slayer </td>

<td style="text-align:right;"> -0.8520841 </td>

</tr>

<tr>

<td style="text-align:left;"> U-Kiss </td>

<td style="text-align:right;"> -0.6984303 </td>

</tr>

<tr>

<td style="text-align:left;"> Insane Clown Posse </td>

<td style="text-align:right;"> -0.5917904 </td>

</tr>

<tr>

<td style="text-align:left;"> Unearth </td>

<td style="text-align:right;"> -0.5672828 </td>

</tr>

<tr>

<td style="text-align:left;"> X-Raided </td>

<td style="text-align:right;"> -0.5517164 </td>

</tr>

<tr>

<td style="text-align:left;"> YG </td>

<td style="text-align:right;"> -0.5055531 </td>

</tr>

<tr>

<td style="text-align:left;"> Unseen </td>

<td style="text-align:right;"> -0.4944980 </td>

</tr>

<tr>

<td style="text-align:left;"> Venom </td>

<td style="text-align:right;"> -0.4817615 </td>

</tr>

<tr>

<td style="text-align:left;"> Death </td>

<td style="text-align:right;"> -0.4655047 </td>

</tr>

<tr>

<td style="text-align:left;"> Korn </td>

<td style="text-align:right;"> -0.4554874 </td>

</tr>

<tr>

<td style="text-align:left;"> Megadeth </td>

<td style="text-align:right;"> -0.4330749 </td>

</tr>

<tr>

<td style="text-align:left;"> Virgin Steele </td>

<td style="text-align:right;"> -0.4148325 </td>

</tr>

<tr>

<td style="text-align:left;"> Misfits </td>

<td style="text-align:right;"> -0.4000387 </td>

</tr>

<tr>

<td style="text-align:left;"> Quincy Punx </td>

<td style="text-align:right;"> -0.3864502 </td>

</tr>

<tr>

<td style="text-align:left;"> Eminem </td>

<td style="text-align:right;"> -0.3795108 </td>

</tr>

<tr>

<td style="text-align:left;"> Opeth </td>

<td style="text-align:right;"> -0.3790620 </td>

</tr>

<tr>

<td style="text-align:left;"> Exo </td>

<td style="text-align:right;"> -0.3753109 </td>

</tr>

<tr>

<td style="text-align:left;"> Ice Cube </td>

<td style="text-align:right;"> -0.3556014 </td>

</tr>

<tr>

<td style="text-align:left;"> Metallica </td>

<td style="text-align:right;"> -0.3509291 </td>

</tr>

<tr>

<td style="text-align:left;"> System Of A Down </td>

<td style="text-align:right;"> -0.3450699 </td>

</tr>

<tr>

<td style="text-align:left;"> Migos </td>

<td style="text-align:right;"> -0.3398299 </td>

</tr>

<tr>

<td style="text-align:left;"> Exo-K </td>

<td style="text-align:right;"> -0.3244116 </td>

</tr>

<tr>

<td style="text-align:left;"> Marilyn Manson </td>

<td style="text-align:right;"> -0.3142416 </td>

</tr>

<tr>

<td style="text-align:left;"> W.A.S.P. </td>

<td style="text-align:right;"> -0.3107639 </td>

</tr>

<tr>

<td style="text-align:left;"> Yo Gotti </td>

<td style="text-align:right;"> -0.2812385 </td>

</tr>

</tbody>

</table>