Spotify Analysis on Turkish Music

Umre Metin

Mina Silahtaroğlu

Data Exploration

```
# The code below, with the use of API of Spotify, returns a dataframe that stores top tracks for an artist with the artist i
d

get_artist_top_tracks <- function(artist_id,essentials=TRUE){
  my_query <- paste0("https://api.spotify.com/v1/artists/",artist_id,"/top-tracks?country=TR&","access_token=",get_spotify_a
ccess_token())
  mydata <- jsonlite::fromJSON(my_query,flatten=TRUE)

  return_df <- mydata$tracks
  if(essentials){
    return_df <- return_df %>% select(track_id=id,track_name=name,track_popularity=popularity,album_id=album.id,album_name=a
lbum.name,duration_ms)
  }

  return(return_df)
}
```

```r
# Returns a list that contains artist id, artist name, artist popularity, followers and genres

get_artists_info <- function(artist_ids,essentials=TRUE){
  artist_ids_p <- paste0(artist_ids,collapse=",")
  my_query <- paste0("https://api.spotify.com/v1/artists?ids=",artist_ids_p,"&access_token=",get_spotify_access_token())
  mydata <- jsonlite::fromJSON(my_query,flatten=TRUE)
  mydata <- mydata$artists

  artist_info <- mydata %>% select(artist_id=id,artist_name=name,artist_popularity=popularity,artist_followers=followers.tot
al)
  # tibble(artist_id=mydata$id,artist_name=mydata$name,artist_popularity=mydata$popularity,artist_followers=mydata$followers
$total)

  artist_genres <-
    mydata %>% select(artist_id=id,genres) %>% unnest(genres)

  return_list <- list(artist_info=artist_info,artist_genres=artist_genres)

  return(return_list)
}
```

```r
# Returns all the attributes available for a track

get_track_audio_features <- function(track_ids,essentials=TRUE){
  track_ids_p <- paste0(track_ids,collapse=",")
  my_query <- paste0("https://api.spotify.com/v1/audio-features?ids=",track_ids_p,"&access_token=",get_spotify_access_token
())
  mydata <- jsonlite::fromJSON(my_query,flatten=TRUE)

  return_df <- mydata$audio_features
  if(essentials){
    return_df <- return_df %>% select(track_id=id,danceability:tempo)
  }

  return(return_df)
}


# For the input genre returns a list of top 20 artists, with their top 10 songs

top20_by_genre <- function(genre_name="turkish pop"){

  genre_data <- get_genre_artists(genre_name)
  artist_ids <- genre_data$id
  artist_list <- get_artists_info(artist_ids)
  top_tracks <- tibble()
  for(i in 1:length(artist_ids)){
    temp_top_tracks <- get_artist_top_tracks(artist_ids[i])
    track_af <- get_track_audio_features(temp_top_tracks$track_id)
    temp_df <- left_join(temp_top_tracks,track_af,by="track_id") %>% mutate(artist_id=artist_ids[i]) %>% select(artist_id,ev
erything())
    top_tracks <- bind_rows(top_tracks,temp_df)
  }

  return(list(artist_info=artist_list$artist_info,artist_genres=artist_list$artist_genres,top_tracks=top_tracks))
}
```

```
top20_info <- top20_by_genre()
```

```
## $q
## [1] "genre:\"turkish+pop\""
##
## $type
## [1] "artist"
##
## $market
## NULL
##
## $limit
## [1] 20
##
## $offset
## [1] 0
##
## $access_token
## [1] "BQCHLa2byqkeMsqghtyMTVLRHgVp8iNqCse8GPv2vt5Nl_cK8xQteWlgAhfaVeF1HMqojb42NI0e-lEI9IY"
```

```
glimpse(top20_info)
```

```
## List of 3
##  $ artist_info  :'data.frame':   20 obs. of  4 variables:
##   ..$ artist_id        : chr [1:20] "2yMN0IP20GOaN6q0p0zL5k" "64d1rUxfizSAOE9UbMnUZd" "2l0Ol9ASKE7E196nHpblB7" "1F2v33FQT
avJyaD7ZsyhdA" ...
##   ..$ artist_name      : chr [1:20] "Tarkan" "Sezen Aksu" "Toygar Işıklı" "Gülşen" ...
##   ..$ artist_popularity: int [1:20] 65 68 61 63 62 66 62 64 61 61 ...
##   ..$ artist_followers : int [1:20] 1421303 1871943 71313 592431 204422 672924 419011 397980 403279 372282 ...
##  $ artist_genres:'data.frame':   40 obs. of  2 variables:
##   ..$ artist_id: chr [1:40] "2yMN0IP20GOaN6q0p0zL5k" "2yMN0IP20GOaN6q0p0zL5k" "64d1rUxfizSAOE9UbMnUZd" "64d1rUxfizSAOE9Ub
MnUZd" ...
##   ..$ genres   : chr [1:40] "turkish pop" "turkish rock" "turkish pop" "turkish rock" ...
##  $ top_tracks   :Classes 'tbl_df', 'tbl' and 'data.frame':   200 obs. of  18 variables:
##   ..$ artist_id       : chr [1:200] "2yMN0IP20GOaN6q0p0zL5k" "2yMN0IP20GOaN6q0p0zL5k" "2yMN0IP20GOaN6q0p0zL5k" "2yMN0IP20
GOaN6q0p0zL5k" ...
##   ..$ track_id        : chr [1:200] "6vEL7E8TGzrxuxbrgdlkLL" "1BfpV5h16V1Gtr41GEF9cO" "1tiIKelgeYVi1KcylVnZdq" "0oVnsgyHc
lNnGeEb4hRoYO" ...
##   ..$ track_name      : chr [1:200] "Simarik" "Kedi Gibi" "Kış Güneşi" "Ay" ...
##   ..$ track_popularity: int [1:200] 63 57 57 57 56 55 52 52 51 51 ...
##   ..$ album_id        : chr [1:200] "6ebdbUXZn5AURKeRNXqiay" "4fJzakARJP2UfOjSj5Q9s1" "06EBiLh3V0q2dOUdXnic7e" "11KfFiHZx
XiBoEq52HAsJi" ...
##   ..$ album_name      : chr [1:200] "Ölürüm Sana" "10" "Aacayipsin" "Karma" ...
##   ..$ duration_ms     : int [1:200] 235218 249006 235171 260713 233639 246456 236054 248580 273307 256069 ...
##   ..$ danceability    : num [1:200] 0.775 0.699 0.652 0.685 0.705 0.753 0.649 0.806 0.743 0.729 ...
##   ..$ energy          : num [1:200] 0.818 0.897 0.706 0.763 0.883 0.926 0.807 0.865 0.871 0.788 ...
##   ..$ key             : int [1:200] 7 7 11 11 4 9 3 4 4 10 ...
##   ..$ loudness        : num [1:200] -8.69 -4.48 -13.41 -6.54 -8.35 ...
##   ..$ mode            : int [1:200] 1 1 0 0 0 1 0 0 0 0 ...
##   ..$ speechiness     : num [1:200] 0.164 0.0585 0.0436 0.0839 0.0508 0.103 0.135 0.0544 0.0618 0.0714 ...
##   ..$ acousticness    : num [1:200] 0.0518 0.0145 0.0158 0.00443 0.0223 0.00718 0.0613 0.0758 0.000997 0.0352 ...
##   ..$ instrumentalness: num [1:200] 0.00 4.05e-06 0.00 1.64e-04 8.34e-03 2.28e-06 5.05e-06 2.25e-05 1.18e-02 3.15e-05 ...
##   ..$ liveness        : num [1:200] 0.114 0.239 0.0843 0.217 0.0526 0.0407 0.0648 0.0696 0.177 0.0969 ...
##   ..$ valence         : num [1:200] 0.826 0.568 0.892 0.418 0.73 0.79 0.761 0.569 0.572 0.722 ...
##   ..$ tempo           : num [1:200] 97.1 98 91 90 93.9 ...
```

From the top20_info list a part of data is collected to make it easier to analyze.

```
top20Artist_songFeatures <-
  top20_info$top_tracks %>%
  select(artist_id, album_name, track_name, danceability:tempo)
```

For the top artists we have looked for the correlation between the top songs' attributes and danceability

With the correlation between the attributes below and danceability, we have seen that the most signicant value for relation is between "Valence", a plot for correlation will be added to this part.

```
correl_values = c(cor(top20Artist_songFeatures$danceability, top20Artist_songFeatures$energy),
                  cor(top20Artist_songFeatures$danceability, top20Artist_songFeatures$liveness),
                  cor(top20Artist_songFeatures$danceability, top20Artist_songFeatures$acousticness),
                  cor(top20Artist_songFeatures$danceability, top20Artist_songFeatures$instrumentalness),
                  cor(top20Artist_songFeatures$danceability, top20Artist_songFeatures$speechiness),
                  cor(top20Artist_songFeatures$danceability, top20Artist_songFeatures$valence),
                  cor(top20Artist_songFeatures$danceability, top20Artist_songFeatures$tempo))

correl_rnames= c('Danceability-Energy', 'Danceability-Liveness', 'Danceability-Acousticness', 'Danceability-Instrumentalnes
s', 'Danceability-Speechiness' , 'Danceability-Valence' , 'Danceability-Tempo')

correl_df_danceability = data.frame(Correlation = correl_values, row.names = correl_rnames)
correl_df_danceability %>%
  knitr::kable()
```

	Correlation
Danceability-Energy	0.4199976
Danceability-Liveness	-0.0626638
Danceability-Acousticness	-0.4358045
Danceability-Instrumentalness	-0.0442729
Danceability-Speechiness	0.2235766
Danceability-Valence	0.4654431
Danceability-Tempo	-0.2297999

We had fun :)

We looked for the most positive songs in the top20 artists' best songs. And it was "Karabiberim" from Serdar Ortac

```
Musical_Positivity <-
  top20Artist_songFeatures %>%
  arrange(desc(valence)) %>%
  select(track_name, valence) %>%
  head(10)

Musical_Positivity %>%
  knitr::kable()
```

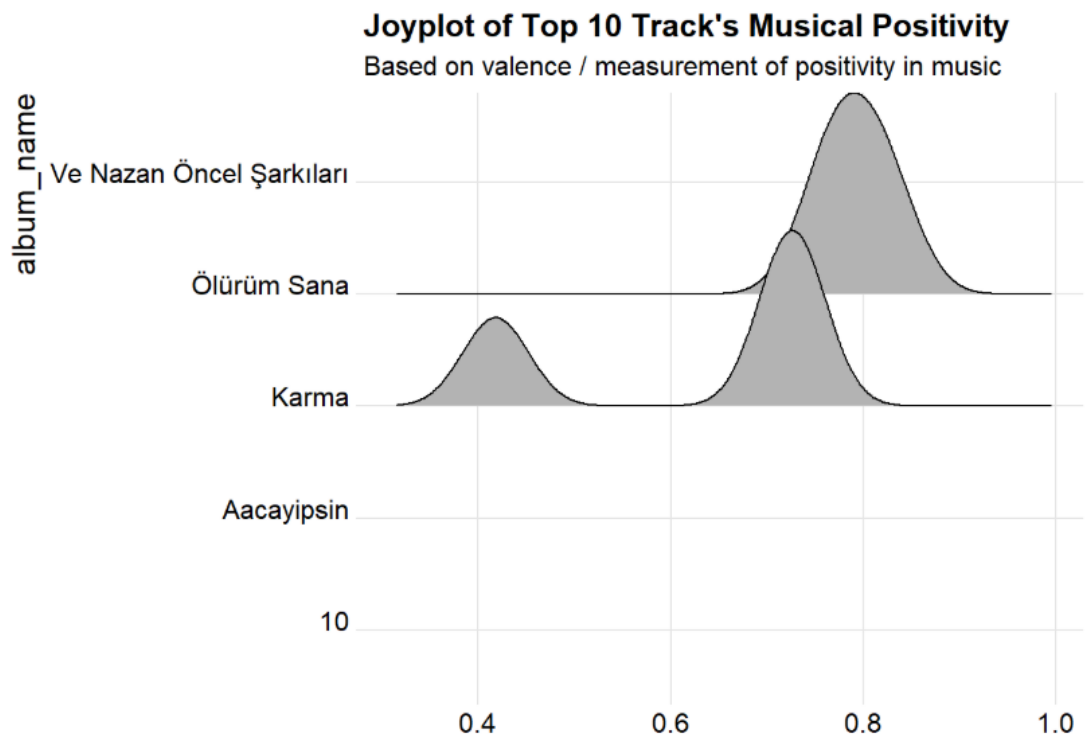
track_name	valence
Karabiberim	0.968
Mor	0.961
Aşkın Ateşi	0.961
Kafa	0.960
Tac Mahal	0.952
Kırmızı	0.945
Yukardan Ayarlı	0.943
Gamzelim	0.942
Jest Oldu	0.935
Çanta	0.927

Since the most significant relationship was maintained with valence; to demonstrate the distribution of valence (joy) for the top tracks' albums a joyplot is created

Since the data set was quite wide, we have sliced the data in 10 rows

```
ggplot(slice(top20Artist_songFeatures, 1:10), aes(x = valence, y = album_name)) +
  geom_joy() +
  theme_joy() +
  ggtitle("Joyplot of Top 10 Track's Musical Positivity", subtitle = "Based on valence / measurement of positivity in music"
)
```

```
## Picking joint bandwidth of 0.0343
```



Thank you for listening…