PROJECT1

Gaurang

2024-03-15

##load libraries

```
## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'ggplot2' was built under R version 4.3.3

## Warning: package 'tidyr' was built under R version 4.3.3

## Warning: package 'readr' was built under R version 4.3.3

## Warning: package 'purrr' was built under R version 4.3.3

## Warning: package 'stringr' was built under R version 4.3.3

## Warning: package 'forcats' was built under R version 4.3.3

## Warning: package 'lubridate' was built under R version 4.3.3

## -- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

## Warning: package 'caret' was built under R version 4.3.3

## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift

## Warning: package 'randomForest' was built under R version 4.3.3

## randomForest 4.7-1.1
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```
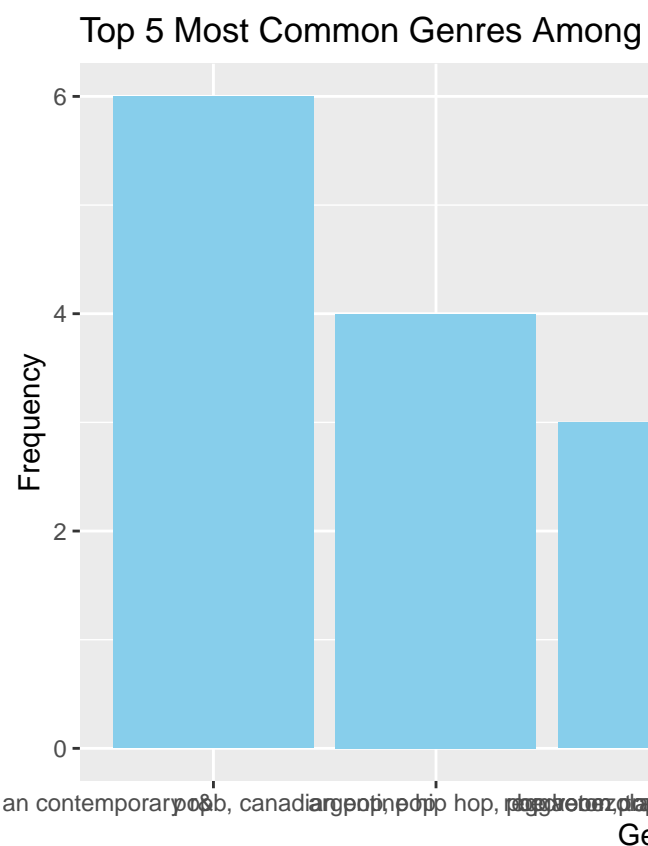
```
## 
## The following object is masked from 'package:ggplot2':
## 
##     margin

## Warning: package 'e1071' was built under R version 4.3.3

## Warning: package 'gbm' was built under R version 4.3.3

## Loaded gbm 2.1.9
## This version of gbm is no longer under development. Consider transitioning to gbm3, https://github.c
```

##read dataset

##cleaning

work

Top 5 Most Common Genres Among

#intro ##1.What are the most common genres among the top tracks?

```
##                                                                           genres
## 1                                                                            pop
## 2                                  canadian contemporary r&b, canadian pop, pop
## 3                                                                    bedroom pop
## 4                                      reggaeton, trap latino, urbano latino
## 5   argentine hip hop, pop venezolano, trap argentino, trap latino, urbano latino
## 6                                                                          k-pop
## 7                                                                  pop, r&b, rap
## 8                                                      afrobeats, nigerian pop
## 9                                      big room, dance pop, edm, pop, pop dance
## 10                                                                     chill pop
```
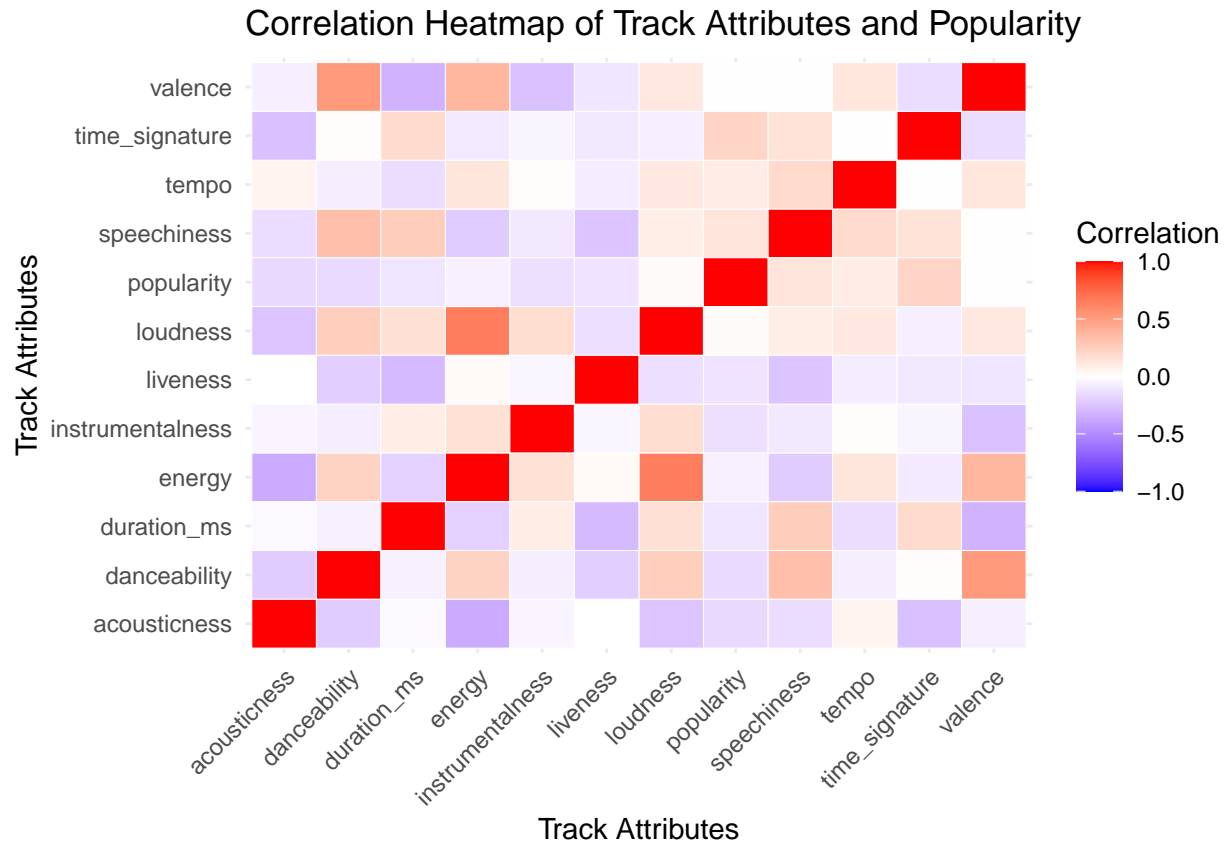
```
##    n
## 1  6
## 2  4
## 3  3
## 4  3
## 5  2
## 6  2
## 7  2
## 8  1
## 9  1
## 10 1
```

2nd

##Next, let's examine the correlation between track attributes and popularity.

```
##                  danceability      valence      energy    loudness acousticness
## danceability       1.00000000  0.525479001  0.23252789  0.25529997  -0.21553662
## valence            0.52547900  1.000000000  0.37666681  0.12073312  -0.06602598
## energy             0.23252789  0.376666812  1.00000000  0.65634275  -0.36113127
## loudness           0.25529997  0.120733122  0.65634275  1.00000000  -0.25261293
## acousticness      -0.21553662 -0.066025976 -0.36113127 -0.25261293   1.00000000
## instrumentalness  -0.07283305 -0.263328832  0.15954540  0.17861957  -0.04581003
## liveness          -0.20249659 -0.108042532  0.02637286 -0.13431443   0.00177927
## speechiness        0.33239089 -0.005241012 -0.21800696  0.09134069  -0.14651755
## tempo             -0.07283901  0.125452388  0.13242725  0.11727058   0.05959618
## duration_ms       -0.06482851 -0.330536497 -0.19769930  0.16479419  -0.02683308
## time_signature     0.01252882 -0.144077055 -0.08735310 -0.06601346  -0.27121863
## popularity        -0.15737185 -0.008203636 -0.05739763  0.01996697  -0.16201631
##                  instrumentalness     liveness   speechiness        tempo
## danceability          -0.07283305  -0.20249659   0.332390886 -0.072839009
## valence               -0.26332883  -0.10804253  -0.005241012  0.125452388
## energy                 0.15954540   0.02637286  -0.218006961  0.132427245
## loudness               0.17861957  -0.13431443   0.091340690  0.117270584
## acousticness          -0.04581003   0.00177927  -0.146517546  0.059596176
## instrumentalness       1.00000000  -0.03769426  -0.092468233  0.012058686
## liveness              -0.03769426   1.00000000  -0.253529920 -0.075386957
## speechiness           -0.09246823  -0.25352992   1.000000000  0.192251977
## tempo                  0.01205869  -0.07538696   0.192251977  1.000000000
## duration_ms            0.09538389  -0.29689832   0.263154582 -0.143338323
## time_signature        -0.04277509  -0.09246260   0.149505536  0.005251868
## popularity            -0.13125730  -0.11673491   0.140163766  0.097114931
##                  duration_ms time_signature   popularity
## danceability     -0.06482851    0.012528821 -0.157371855
## valence          -0.33053650   -0.144077055 -0.008203636
## energy           -0.19769930   -0.087353102 -0.057397629
## loudness          0.16479419   -0.066013456  0.019966968
## acousticness     -0.02683308   -0.271218629 -0.162016311
## instrumentalness  0.09538389   -0.042775094 -0.131257299
## liveness         -0.29689832   -0.092462598 -0.116734911
## speechiness       0.26315458    0.149505536  0.140163766
## tempo            -0.14333832    0.005251868  0.097114931
## duration_ms       1.00000000    0.189976090 -0.112414445
## time_signature    0.18997609    1.000000000  0.218950244
## popularity       -0.11241445    0.218950244  1.000000000
```

Correlation Heatmap of Track Attributes and Popularity

#3 ##Predicting Track Popularity Lastly, we'll attempt to predict track popularity based on its features.

```
## [1] "Random Forest RMSE: 9.46513510791638"
## [1] "Feature Importance:"
##                    %IncMSE IncNodePurity
## danceability      0.7285468     132.15223
## valence          -0.7147758      82.81624
## energy           -1.2894757      64.19214
## loudness          1.5135163      77.07723
## acousticness     -0.9358643      63.70563
## instrumentalness  0.6624328      48.40721
## liveness         -0.4491487      79.79069
## speechiness      -1.0495332      77.08184
## tempo            -1.8434936      82.78818
## duration_ms       1.1279341     119.66727
## time_signature    1.8943356      14.80190
```