Top Hits Spotify

Yixin Wang

Contents

An Overview of Dataset

The dataset is downloaded as a csv file and is obtained from (https://www.kaggle.com/datasets/paradisejoy/top-hits-spotify-from-20002019?resource=download&select=songs_normalize.csv)

```
top_hits <- read.csv("songs_normalize.csv")
summary(top_hits)
```

```
##     artist              song            duration_ms       explicit
##  Length:2000        Length:2000        Min.   :113000   Length:2000
##  Class :character   Class :character   1st Qu.:203580   Class :character
##  Mode  :character   Mode  :character   Median :223280   Mode  :character
##                                        Mean   :228748
##                                        3rd Qu.:248133
##                                        Max.   :484146
##       year        popularity     danceability        energy
##  Min.   :1998   Min.   : 0.00   Min.   :0.1290   Min.   :0.0549
##  1st Qu.:2004   1st Qu.:56.00   1st Qu.:0.5810   1st Qu.:0.6220
##  Median :2010   Median :65.50   Median :0.6760   Median :0.7360
##  Mean   :2009   Mean   :59.87   Mean   :0.6674   Mean   :0.7204
##  3rd Qu.:2015   3rd Qu.:73.00   3rd Qu.:0.7640   3rd Qu.:0.8390
##  Max.   :2020   Max.   :89.00   Max.   :0.9750   Max.   :0.9990
##       key            loudness            mode         speechiness
##  Min.   : 0.000   Min.   :-20.514   Min.   :0.0000   Min.   :0.02320
##  1st Qu.: 2.000   1st Qu.: -6.490   1st Qu.:0.0000   1st Qu.:0.03960
##  Median : 6.000   Median : -5.285   Median :1.0000   Median :0.05985
##  Mean   : 5.378   Mean   : -5.512   Mean   :0.5535   Mean   :0.10357
##  3rd Qu.: 8.000   3rd Qu.: -4.168   3rd Qu.:1.0000   3rd Qu.:0.12900
##  Max.   :11.000   Max.   : -0.276   Max.   :1.0000   Max.   :0.57600
##   acousticness       instrumentalness       liveness         valence
##  Min.   :0.0000192   Min.   :0.0000000   Min.   :0.0215   Min.   :0.0381
##  1st Qu.:0.0140000   1st Qu.:0.0000000   1st Qu.:0.0881   1st Qu.:0.3867
##  Median :0.0557000   Median :0.0000000   Median :0.1240   Median :0.5575
```

```
##  Mean   :0.1289549   Mean   :0.0152260   Mean   :0.1812   Mean   :0.5517
##  3rd Qu.:0.1762500   3rd Qu.:0.0000683   3rd Qu.:0.2410   3rd Qu.:0.7300
##  Max.   :0.9760000   Max.   :0.9850000   Max.   :0.8530   Max.   :0.9730
##      tempo           genre
##  Min.   : 60.02   Length:2000
##  1st Qu.: 98.99   Class :character
##  Median :120.02   Mode  :character
##  Mean   :120.12
##  3rd Qu.:134.27
##  Max.   :210.85
```

In summary, there are 2000 observations and 18 predictors. I think I will work with numbers mostly and there is no missing data.

An Overview of Research Questions

In this dataset, I am interested in predicting the popularity of these songs. The popularity is between 0 and 89. I think the energy, danceability, loudness, speechiness, liveness, acoustcness would all be useful in predicting popularity and they should be positively related.

Proposed Project Timeline

Since my data memo finished late, I am going to start the exploratory data analysis after the data memo passed immediately.

Questions or Concerns

I think everything is fine for now.