

# Clean

Zonghong Yu

2022-12-04

```
df <- read.csv("genres_v2.csv",stringsAsFactors=FALSE)
head(df)
```

```

##  danceability energy key loudness mode speechiness acousticness
## 1      0.831  0.814  2  -7.364  1      0.4200      0.0598
## 2      0.719  0.493  8  -7.230  1      0.0794      0.4010
## 3      0.850  0.893  5  -4.783  1      0.0623      0.0138
## 4      0.476  0.781  0  -4.710  1      0.1030      0.0237
## 5      0.798  0.624  2  -7.668  1      0.2930      0.2170
## 6      0.721  0.568  0 -11.295  1      0.4140      0.0452
##  instrumentalness liveness valence tempo type
## 1      1.34e-02  0.0556  0.3890 156.985 audio_features
## 2      0.00e+00  0.1180  0.1240 115.080 audio_features
## 3      4.14e-06  0.3720  0.0391 218.050 audio_features
## 4      0.00e+00  0.1140  0.1750 186.948 audio_features
## 5      0.00e+00  0.1660  0.5910 147.988 audio_features
## 6      2.12e-01  0.1280  0.1090 144.915 audio_features
##      id uri
## 1 2Vc6NJ9PW9gD9q343XFRKx spotify:track:2Vc6NJ9PW9gD9q343XFRKx
## 2 7pgJBLVz5VmnL7uGHmRj6p spotify:track:7pgJBLVz5VmnL7uGHmRj6p
## 3 0vSWgAlfpye0WCGeNmuNhy spotify:track:0vSWgAlfpye0WCGeNmuNhy
## 4 0VSXnJqQkwuH2ei1nOQ1nu spotify:track:0VSXnJqQkwuH2ei1nOQ1nu
## 5 4jCeguq9rMTlbMmPHu07S3 spotify:track:4jCeguq9rMTlbMmPHu07S3
## 6 6fsypiJHyWmeINsOLC1cos spotify:track:6fsypiJHyWmeINsOLC1cos
##      track_href
## 1 https://api.spotify.com/v1/tracks/2Vc6NJ9PW9gD9q343XFRKx
## 2 https://api.spotify.com/v1/tracks/7pgJBLVz5VmnL7uGHmRj6p
## 3 https://api.spotify.com/v1/tracks/0vSWgAlfpye0WCGeNmuNhy
## 4 https://api.spotify.com/v1/tracks/0VSXnJqQkwuH2ei1nOQ1nu
## 5 https://api.spotify.com/v1/tracks/4jCeguq9rMTlbMmPHu07S3
## 6 https://api.spotify.com/v1/tracks/6fsypiJHyWmeINsOLC1cos
##      analysis_url duration_ms
## 1 https://api.spotify.com/v1/audio-analysis/2Vc6NJ9PW9gD9q343XFRKx 124539
## 2 https://api.spotify.com/v1/audio-analysis/7pgJBLVz5VmnL7uGHmRj6p 224427
## 3 https://api.spotify.com/v1/audio-analysis/0vSWgAlfpye0WCGeNmuNhy 98821
## 4 https://api.spotify.com/v1/audio-analysis/0VSXnJqQkwuH2ei1nOQ1nu 123661
## 5 https://api.spotify.com/v1/audio-analysis/4jCeguq9rMTlbMmPHu07S3 123298
## 6 https://api.spotify.com/v1/audio-analysis/6fsypiJHyWmeINsOLC1cos 112511
##  time_signature genre song_name
## 1      4 Dark Trap Mercury: Retrograde
## 2      4 Dark Trap Pathology
## 3      4 Dark Trap Symbiote
## 4      3 Dark Trap ProductOfDrugs (Prod. The Virus and Antidote)
## 5      4 Dark Trap Venom
## 6      4 Dark Trap Gatteka
##  Unnamed..0 title
## 1      NA
## 2      NA
## 3      NA
## 4      NA
## 5      NA
## 6      NA

```

```
names(df)
```

```
## [1] "danceability" "energy" "key" "loudness"
## [5] "mode" "speechiness" "acousticness" "instrumentalness"
## [9] "liveness" "valence" "tempo" "type"
## [13] "id" "uri" "track_href" "analysis_url"
## [17] "duration_ms" "time_signature" "genre" "song_name"
## [21] "Unnamed..0" "title"
```

```
df = subset(df, select = -c(uri, id, track_href, Unnamed..0, duration_ms, analysis_url, title, type) )
```

```
head(df)
```

```
## danceability energy key loudness mode speechiness acousticness
## 1 0.831 0.814 2 -7.364 1 0.4200 0.0598
## 2 0.719 0.493 8 -7.230 1 0.0794 0.4010
## 3 0.850 0.893 5 -4.783 1 0.0623 0.0138
## 4 0.476 0.781 0 -4.710 1 0.1030 0.0237
## 5 0.798 0.624 2 -7.668 1 0.2930 0.2170
## 6 0.721 0.568 0 -11.295 1 0.4140 0.0452
## instrumentalness liveness valence tempo time_signature genre
## 1 1.34e-02 0.0556 0.3890 156.985 4 Dark Trap
## 2 0.00e+00 0.1180 0.1240 115.080 4 Dark Trap
## 3 4.14e-06 0.3720 0.0391 218.050 4 Dark Trap
## 4 0.00e+00 0.1140 0.1750 186.948 3 Dark Trap
## 5 0.00e+00 0.1660 0.5910 147.988 4 Dark Trap
## 6 2.12e-01 0.1280 0.1090 144.915 4 Dark Trap
## song_name
## 1 Mercury: Retrograde
## 2 Pathology
## 3 Symbiote
## 4 ProductOfDrugs (Prod. The Virus and Antidote)
## 5 Venom
## 6 Gatteka
```

```
df_dup = df[duplicated(df$song_name), ]
head(df_dup)
```

```
##      danceability energy key loudness mode speechiness acousticness
## 81      0.909  0.573   5   -6.856   1      0.1810      0.03320
## 294     0.756  0.746   1   -6.397   1      0.1420      0.23600
## 426     0.705  0.648   4  -10.467   0      0.1410      0.00476
## 489     0.675  0.628   4   -6.165   0      0.0838      0.00997
## 580     0.945  0.587   2   -5.104   1      0.2710      0.25300
## 669     0.733  0.513   6   -5.068   0      0.0389      0.55900
##      instrumentalness liveness valence  tempo time_signature  genre
## 81      1.63e-02    0.6570  0.3920 144.946              4 Dark Trap
## 294     0.00e+00    0.0999  0.0793 207.956              4 Dark Trap
## 426     5.95e-04    0.3730  0.3980 130.037              4 Dark Trap
## 489     0.00e+00    0.0868  0.5110 179.992              4 Dark Trap
## 580     1.01e-06    0.2640  0.5090 131.969              4 Dark Trap
## 669     1.48e-05    0.1740  0.5640 159.944              4 Dark Trap
##      song_name
## 81      Venom
## 294    Story: No Title
## 426      No Teeth
## 489 Make It Through Fall
## 580      Killer
## 669    Nightmare
```

```
dup <- df[df$song_name == "Venom",]
```

```
df = df[!duplicated(df$song_name), ]
```

```
head(df)
```

```
##  danceability energy key loudness mode speechiness acousticness
## 1      0.831  0.814  2  -7.364  1      0.4200      0.0598
## 2      0.719  0.493  8  -7.230  1      0.0794      0.4010
## 3      0.850  0.893  5  -4.783  1      0.0623      0.0138
## 4      0.476  0.781  0  -4.710  1      0.1030      0.0237
## 5      0.798  0.624  2  -7.668  1      0.2930      0.2170
## 6      0.721  0.568  0 -11.295  1      0.4140      0.0452
##  instrumentalness liveness valence  tempo time_signature  genre
## 1      1.34e-02  0.0556  0.3890 156.985      4 Dark Trap
## 2      0.00e+00  0.1180  0.1240 115.080      4 Dark Trap
## 3      4.14e-06  0.3720  0.0391 218.050      4 Dark Trap
## 4      0.00e+00  0.1140  0.1750 186.948      3 Dark Trap
## 5      0.00e+00  0.1660  0.5910 147.988      4 Dark Trap
## 6      2.12e-01  0.1280  0.1090 144.915      4 Dark Trap
##
##              song_name
## 1      Mercury: Retrograde
## 2              Pathology
## 3              Symbiote
## 4 ProductOfDrugs (Prod. The Virus and Antidote)
## 5              Venom
## 6              Gatteka
```

```
sum(duplicated(df$song_name))
```

```
## [1] 0
```