1st chart

Welcome everyone we are her to talk about music and if some attributes captured by the product spotify have any correlation to what makes a song popular. We are looking at from 3 different perspectives and each of us will go into more detail as we go through the presentation.

1) Is there a similar attribute (i.e acousticness/valence/energy/danceability) that a large portion of the songs have?

2) Is there a particular genre that dominates the top hits playlist?

3) Is there a connection between the hits globally and what countries it ‘blows up’ inData World 2 files (<https://data.world/kcmillersean/billboard-hot-100-1958-2017> and , the web page with the current information, and the API

2nd chart

Data world had 2 files

* The first file is just over 28,000 songs with 10 spotify identifiers which we will use to see if there is any correlation to the ranking. We used the columns highlighted
* Billboard-hot 100-1958-2019. It has roughly 320,000records. This represents the top 100 songs every week from 1957 to the end of 2019. This data has the information on rankings of the songs.
* The last data source that we used was a web interface to Spotify where we could download a csv file containing the top 200 songs based on time period and country.
* We also was able to interwork with the spotify API, and if funding is extended we can continue deeper analysis.

3rd Chart

Here is a cut paste of the songs with the spotify attributes. The fields of the file are shown on the left-hand side. The field highlighted in green is the key used to tie this file with the second file during the merge. The fields in yellow are the attributes that we spent time looking at. They are

Tempo — The tempo of the song. BPM

Energy — The energy of a song, the higher the value, the more energetic.

Danceability — The higher the value, the easier it is to dance to this song.

Loudness — The higher the value, the louder the song (in dB).

Valence — The higher the value, the more positive mood for the song.

Acousticness — The higher the value the more acoustic the song is.

Popularity — The higher the value the more popular the song is.

Instrumentalness - The higher the value the greater likelihood the track contains no vocal content.

Liveness – The higher the value the more likely the track was performed live.

Speechiness – The higher the value the more exclusively speech-like recording.

4th chart

The second file is the top 100 songs taken every week from 1958 to 2019 giving us 62 years of data. In this file there can be multiple entries since a song can be on the hit list multiple list. Part of what we needed to do is highest point that the song went to and used that for comparisons. While there is a peak position column, we needed to be careful as the last entry may not contain the appropriate information. We ended up using the min function to find the lowest ranking. The columns that have been highlighted in yellow are the ones that we have focused on

Once we pulled the data in and removed rows with no data. We pulled the peak position and used the SongID column to do an inner merge of the dataframes. We again cleaned up the rows if the merge brought in extra rows. Once that was done, we had a clean working dataframe.

To answer the first question, we took the clean merged dataframe and created scatter plots, box plots, calculate the mean and median for the different attributes of the songs, and finally separated the data by decades. We can’t show all the information, but in the next few charts we will show what results we saw and observations we made. In addition, after examining the information we took a look at the decade to see whether the songs in the decade had any similarity to them. To make life simple we used loops and functions to create the graphs and we are able to analyze the data quickly depending on the grouping of the songs.

5th chart

In general, the charts will have an x-axis of the based on the peak position of the song and the y-axis will be the attribute rating. There was only one solid correlation and that was track popularity.

6th chart

The next step of possible correlations are speechiness, danceability, and valence

7th chart

The final 6 attributes show no obvious correlation.

8th chart

This last chart is taking a look at the same attributes but looking at the mean and median of the attributes. The chart here shows the mean and it represents the correlations we saw earlier. The one standout is the loudness. It shows a slight trend that the more popular the song the lower the volume while we earlier scatter chart showed no correlation. The next step would be to eliminate the outliers and see if a stronger correlation exists.