SPOTIFY

DATABASE SYSTEMS

TOP 50 SPOTIFY SONGS 2019

INTRODUCTION

Spotify is a digital music streaming services provider, that contains millions of songs, playlists and podcasts. The sample of the database below looks at the top 50 songs of 2019 on spotify. With 271 Million active users as of 31st December 2019, more users are beginning to use this streaming service to listen to their music and as new music is released, with a greater cultural diversity, I have chosen this dataset to evaluate the most popular music on this service as of 2019. I predict that for this dataset the top 50 spotify songs will include a greater diversity of songs, with music pop with a faster tempo being some of the popular songs of 2019.

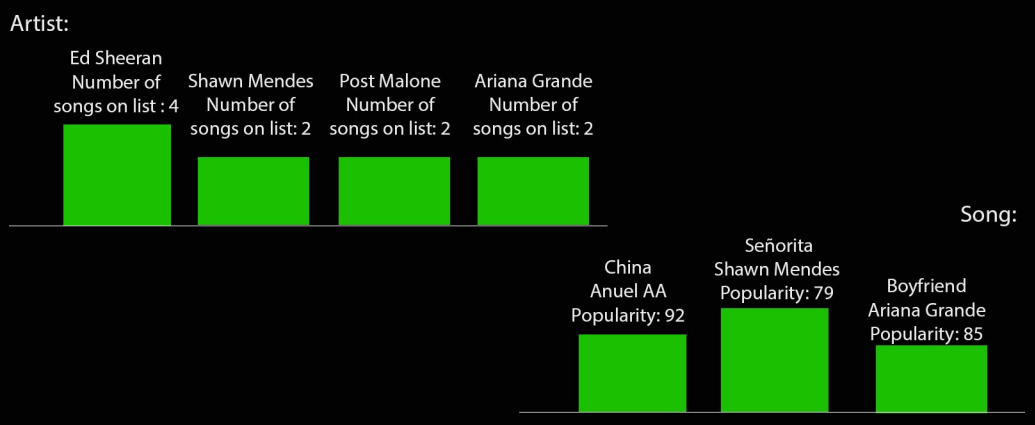
	Track.Name	Artist. Name	Genre	Beats.Per.Min	Energy
1	Senorita	Shawn Mendes	Canadian Pop	117	55
2	China	Anuel AA	Reggaeton flow	105	81
3	Boyfriend	Ariana Grande	Dance Pop	190	80
4	Beautiful People	Ed Sheeran	Pop	93	65
5	Goodbyes	Post Malone	Dfw Pop	150	65

ATTRIBUTES

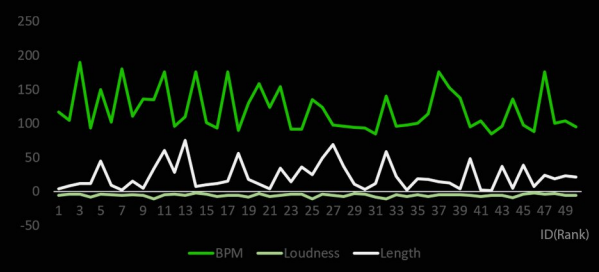
Records : 50
Primary Key : ID

- Track.Name
- Genre
- Liveness - higher value the more likely the song is a live recording
- Valence - higer value denotes a more posotive mood
- Energy - higher value are more energetic
- Loudness (dB)
- Artist.Name
- Beats.Per.Min- tempo of the song
- Length - duration of the song
- Danceability - higher values show that they are easier to dance to
- Acousticness - higher the value the more acoustic the song is
- Speechiness - higher the value the more words the song has
- Popularity

POPULARITY



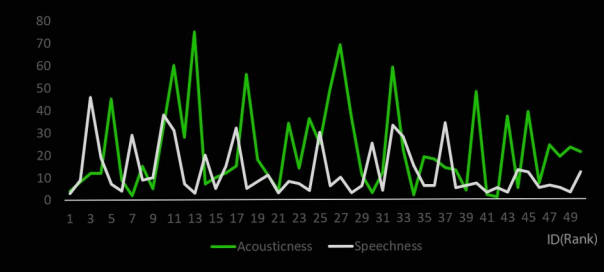
DATA



BPM:
Average - 120
Standard Deviation - 30.6
Min - 85
Max - 190

Loudness:
Average - -5.66
Standard Deviation - 2.04
Min - -11
Max - -2

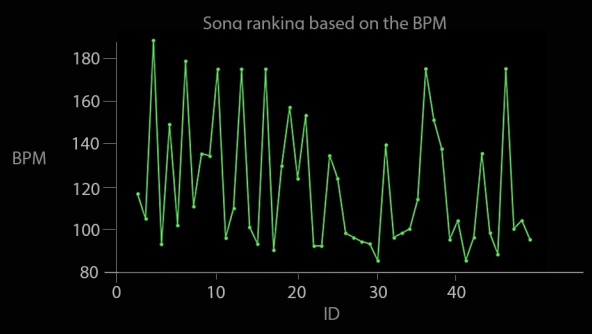
Length:
Average - 201
Standard Deviation - 38.8
Min - 115
Max - 309



Acoustic:
Average - 22.2
Standard Deviation - 18.8
Min - 1
Max - 75

Speech:
Average - 12.5
Standard Deviation - 11
Min - 3
Max - 46

```
# Libraries used: matplotlib
# Code adapted from:
# <http://www.learningaboutelectronics.com/
# Articles/How-to-plot-a-graph with-matplotlib
# -from-data-from-a-CSV-file-using-the-CSV-module-in
# Python.php>

import matplotlib.pyplot as plt
import csv


fig = plt.figure()
ax = fig.add_subplot(111)

x=[]
y=[]


with open('top50.csv', 'r') as csvfile:
    data = csv.reader(csvfile)
    for row in data:
        x.append(int(row[0]))
        y.append(int(row[4]))


plt.plot(x,y, marker='.',color = 'C2')

plt.title('Song ranking based on the BPM')

plt.xlabel('ID')
plt.ylabel('Bpm')


ax.patch.set_facecolor('black')
ax.patch.set_alpha(1)
plt.show()
```



^ Figure : Graph drawn using code written in python

DATA RELIABILITY

The data presented in the dataset may be seen as unreliable due to the lack of maintenance during that year, last being updated in July 2019 the data does not consider new song releases after this month therefore being unreliable to whether the given data remained within the top 50. Moreover, it is unclear whether the data given is considering songs released in other countries and while spotify does have a large quantity of music it still may be missing some data, consequently making the data less reliable.

TOP 3 GENRES

Dance Pop: 16%
Top 3 of this genre:
1. Ariana Grande - Boyfriend
2. DJ Snake - Loco Contigo
3. Chris Brown - No Guidance

Pop: 14%
Top 4 of this genre:
1. Ed Sheeran - Beautiful People
2. Ed Sheeran - I don't care
3. Lewis Capaldi - Someone you loved

Latin: 10%
Top 3 of this genre:
1. J Balvin - La Canción
2. Lunay - Soltera Remix
3. Daddy Yankee - Con Calma

CONCLUSION

To conclude, the data supports the majority of my hypothesis, Pop songs can be seen as one of the greater genres with 16% of the top 50 being dance pop and 14% being pop, and with latin being one of the top three genres. It can be said that songs on these charts are more diverse however, genres such as J - Pop, K - Pop, Bollywood genres are less well recognised on these charts, despite having more popularity on other streaming services such as YouTube. American pop music is evidently the most listened to music of 2019 on Spotify, while it is uncertain to defintively say why, one main reason could be influenced by more people using these streaming services in this area. In addition to this Señorita has a tempo of 117 beats per minute this is 3 below the mean, further seen through the graph, there is no correlation between the ranking on the track and their tempo. Personally this project has developed my skills in computing with data and using Python to create graphs using csv files.

DATASET REFERENCES

Organise your spotify music:
http://organizeyourmusic.playlistmachinery.com/

View Kaggle Dataset:
https://www.kaggle.com/leonardopena/top50spotify2019

Creating graphs using python:
http://www.learningaboutelectronics.com/Articles/How-to-plot-a-graph-with-matplotlib-from-data-from-a-CSV-file-using-the-CSV-module-in-Python.php

Matplotlib - Python Library Documentation:
https://matplotlib.org/