SUBMITTED TO:
MOHIT GUJAR SIR

DATE FINISHED: 30 SEPTEMBER 2021
DATE SUBMITTED: 4 OCTOBER 2021

A PROJECT REPORT ON

MUSIC DATA

ANALYSIS AND VISUALIZATION

DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION ENGINEERING

MENTOR :

Prof. Mohit Gujar

CONTRIBUTORS :

Sonali Bedade	19104A0016
Samruddhi Malawde	19104A0040
Anup Kanse	19104A0013

ABSTRACT:

Music is the art of organizing sound in accordance with suitable music at appropriate intervals of time. Music has hold its importance since a very long time and continues to grow and expand. From country to country it takes various forms involving people more into cultural values. Various types of music also have evolved over the time according to the changing generations and preferences. Considering the changing preferences of people the popularity of music can be contributed by various factors such as loudness, tempo, valence, etc and their relation with each other can be found out. In this project we try to find the correlation amongst various features of music like popularity, key, loudness, energy,danceability and many more.

INTRODUCTION:

Data Analysis and Data Visualization are now-a-days used in a variety of fields and it is an important tool that helps us to analyze data and draw useful insights. Here we are using Data Analysis and Data Visualization techniques to analyze the data from the dataset and plot the correlation between various features of the music genre and analyze the data for various correlations and use these analysis to visualize data.

The project is sub-divided following section. These are:

1.Loading necessary libraries.
2.Loading Dataset from a CSV file or from a Table.
3.Summarization of Data to understand Dataset (Descriptive Statistics)
4.Visualization of Data to understand Dataset (Plots, Graphs etc.)
5.Data pre-processing and Data transformation

DATA:

The music data has been taken from Kaggle website. The dataset contains various dependable and significant factors of music along with their respective values. It is in the format of a csv file which consists of 25709 rows and 16 columns. The columns includes features such as 'Artist Name', 'Track Name', 'Popularity', 'danceability', 'energy', 'key', 'loudness', 'mode', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo', 'duration_in min/ms', 'time_signature'. These features have different values for all the inputs and are correlated with each other. Of these 16 features, 'Artist Name' and 'Track Name' are of object type, 'mode' and 'time_signature' are of int type and all the remaining ones are of data type float. The data previously contained some null values which were replaced by 0 during the preprocessing.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Artist Nan	Track Nam	Popularity	danceabil	energy	key	loudness	mode	speechine	acousticn	instrumer	liveness	valence	tempo	duration_	time_signature	
2	Bruno Mai	That's Wh	60	0.854	0.564	1	-4.964	1	0.0485	0.0171		0.0849	0.899	134.071	234596	4	
3	Boston	Hitch a Ric	54	0.382	0.814	3	-7.23	1	0.0406	0.0011	0.00401	0.101	0.569	116.454	251733	4	
4	The Rainc	No Side to	35	0.434	0.614	6	-8.334	1	0.0525	0.486	0.000196	0.394	0.787	147.681	109667	4	
5	Deno	Lingo (fea	66	0.853	0.597	10	-6.528	0	0.0555	0.0212		0.122	0.569	107.033	173968	4	
6	Red Hot C	Nobody W	53	0.167	0.975	2	-4.279	1	0.216	0.000169	0.0161	0.172	0.0918	199.06	229960	4	
7	The Stoog	Search an	53	0.235	0.977	6	0.878	1	0.107	0.00353	0.00604	0.172	0.241	152.952	208133	4	
8	Solomon I	None Of U	48	0.674	0.658	5	-9.647	0	0.104	0.404	1.34E-06	0.0981	0.677	143.292	329387	4	
9	Randy Tra	On the Ot	55	0.657	0.415	5	-9.915	1	0.025	0.175	5.65E-06	0.132	0.347	96.03	3.105783	4	
10	Professior	Slow	29	0.431	0.776	10	-5.403	1	0.0527	2.21E-05	0.0013	0.179	0.318	120.857	237867	4	
11	Dudu Aha	â—ŠÃ¬â—	14	0.716	0.885	1	-4.348	0	0.0333	0.0614		0.253	0.833	128.043	164093	4	
12	Mohamm	Meri Dost	11	0.491	0.563	7	-8.588	0	0.0331	0.957	0.0023	0.263	0.843	95.558	4.33045	4	
13	Arctic Mor	The View	59	0.387	0.922	9	-5.192	0	0.0674	0.00487		0.187	0.417	146.478	222947	4	
14	Eyal Golar	â—ŠÃŸâ—	34	0.585	0.381	1	-7.622	0	0.0463	0.435		0.126	0.322	143.876	206875	4	
15	Harald Las	How it fee	34	0.462	0.374	11	-12.069	0	0.0331	0.949	0.929	0.131	0.174	122.939	4.44025	3	
16	Buffalo Sp	For What I	76	0.653	0.519	2	-10.164	1	0.0497	0.406	0.0209	0.101	0.822	98.883	153693	4	
17	Elmore Ja	Madison B	37	0.431	0.852	2	-6.522	1	0.0431	0.564	0.208	0.254	0.794	67.601	146800	4	
18	Dudu Aha	â—ŠÃºâ—	14	0.713	0.939	5	-3.609	0	0.0372	0.11	5.88E-06	0.331	0.88	108.034	179320	4	
19	Velvet Tw	Drinks	31	0.564	0.801	1	-4.73	0	0.0473	0.00233	0.0078	0.067	0.401	105.01	154293	4	
20	Eden Ben	â—ŠÃ»â—	29	0.57	0.458	9	-5.935	0	0.0236	0.502	2.35E-06	0.134	0.19	93.088	217529	4	
21	DaBaby	Baby Sitte	70	0.852	0.73	1	-5.803	1	0.0579	0.0219		0.336	0.394	92.506	157693	4	

METHOD:

Steps that we followed in this project:

1. Gathering the data
2. Understanding the data
3. Development of problem statement
4. Importing the required libraries
5. Data Analysis and Pre-processing
6. Removal of null values
7. Visualizing the data
8. Insights and Conclusion

APPROACH:

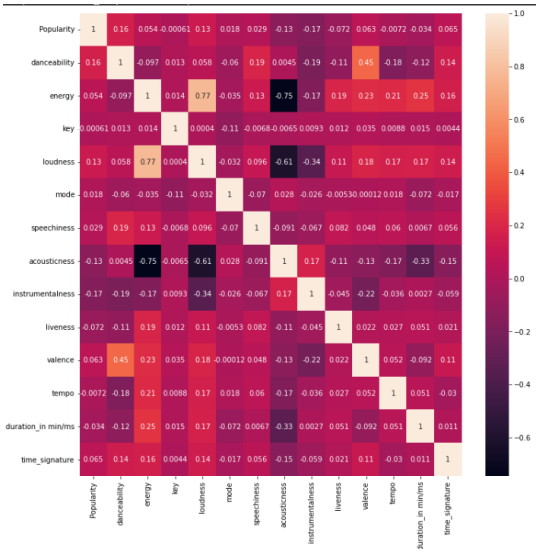
Here we are using Google Colaboratory for writing the code for our project.
We have imported necessary libraries like *Numpy* for numerical arrays, *Pandas* used for data analysis it allows
various file formats such as JSON, CSV, Microsoft excel.

Matplotlib is a plotting library for Python programming language and it is used for creating static, animated, and interactive visualization in python. *Seaborn* provides a variety of visualization patterns.

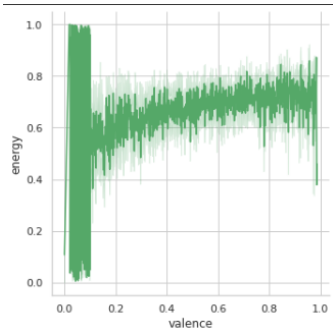
Following this, we have started with our data Pre-processing and checked all required quantities, then checked for any null values in the data set and removed null values or NaN values using command df.fillna(0) basically replaces null values into zeros. Then further plotted heat map of correlation with all the quantities. In this, we found out that light color shade show high correlation and the dark color show less correlation diagonal 1's are self-correlated.

After that, we have visualized the data as per the problem statements for this we used various types of plots like jointplot, relplot, scatterplot, and many more.

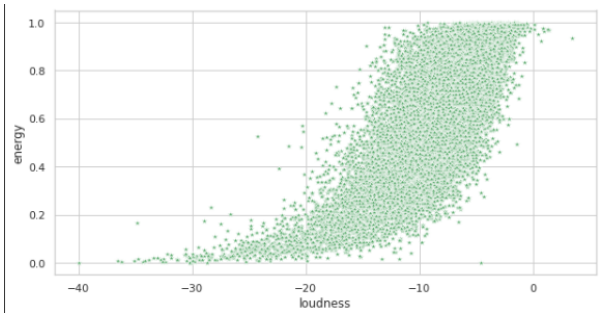
RESULTS AND CONCLUSION:



Here we have used heat Map to show correlation with all the quantities



Here we have plotted energy vs valence using relplot



Here we have plotted energy vs valence using scatterplot

CONCLUSION: The features energy and loudness are highly correlated with each other. The graph shows an exponential increase in loudness as energy goes on increasing. The energy of the music mostly depends on factors duration in mins, valence, tempo, and liveness. The feature loudness is dependable on valence.

REFERENCES:

Link to our project repository:
https://github.com/MalawdeSamruddhi/Music-Data-Analysis

Link to our YouTube video: **https://youtu.be/TDseI4IHNS4**

Dataset link: **https://www.kaggle.com/purumalgi/music-genre-classification**