

# INNOMATICS

## RESEARCH LABS



# SPOTIFY ANALYSIS

Team Member's  
N.Vijay

# About me

1. **N.Vijay**
  - **B.COM(General)**
  - **Want to make a career in Data science as i am interested in data and analysis**

# Table of Content

- ❑ Introduction
- ❑ Summary of the Data
- ❑ Steps Involved in Web Scrapping
- ❑ Challenges
- ❑ Problem Statement
- ❑ Songs Dataframe
- ❑ Spotify statistics
- ❑ Conclusion
- ❑

# Introduction

Spotify is a swedish audio streaming and media services provider founded in April 2006. It is the world's largest music streaming service provider and has over 381 million monthly active users , which also includes 172 million paid subscribers



# Summary of the Data

Spotify Statistics Data Frame will show the data of Top 50 Songs per year from (2010-2019) including details like song title, Name of the singer , genre of the song , Year of song released and some of the other details including Characteristics of songs.

# Steps Involved in Web Scraping

- Getting and Reading URL
- Extracting the HTML code
- Beautifying the HTML source code
- Extracting the Text from HTML
- Creating a Columns lists
- Creating a DataFrame from the Column lists
- Converting the Data Frame into .csv or .xlsx for storing the Data

# Challenges

- In web scraping , i have extracted raw data from table for that i have to use for loop for getting rows and columns
- As i extracted data from the table , all the columns were objects so i converted all the numerical columns into integers by using astype
- For finding top 10 songs i had to sort all the songs by popularity
- For finding top 20 artists i had to sort all the artists by popularity
- For finding top 20 genres i had to sort all the genres by popularity



# Problem Statements

1. Finding the most number of genres
2. Finding the high value of speechiness
3. Box plot for liveness
4. Finding top 10 popular songs
5. Finding the relationship between all the variables
6. Finding type of correlation between danceability and valence
7. Finding the type of correlation between Energy and Acousticness
8. Finding the duration of songs of different artists
9. Finding top 20 songs by Artists
10. Finding Top 20 songs by Genres

# Spotify Statistics

1. Found which genre has highest number of songs
2. Found high speechiness value
3. Found count for Liveness
4. Found what are the top 10 songs
5. Found relation through all variables through heat map
6. Found relation between Danceability and Valence
7. Found relation between Accousticness and energy
8. Found top 10 songs by popularity
9. Found top 20 artists by popularity
10. Found top 20 genres by popularity

# Song Dataframe

	Title	Artist_Name	Genre	Year	Beats_Per_MIN	Energy	Danceability	Loudness	Liveness	Valence	Duration	Acousticness	Speechiness	Popularity
0	Hey, Soul Sister	Train	neo mellow	2010	97	89	67	-4	8	80	217	19	4	83
1	Love The Way You Lie	Eminem	detroit hip hop	2010	87	93	75	-5	52	64	263	24	23	82
2	TiK ToK	Kesha	dance pop	2010	120	84	76	-3	29	71	200	10	14	80
3	Bad Romance	Lady Gaga	dance pop	2010	119	92	70	-4	8	71	295	0	4	79
4	Just the Way You Are	Bruno Mars	pop	2010	109	84	64	-5	9	43	221	2	4	78
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
598	Find U Again (feat. Camila Cabello)	Mark Ronson	dance pop	2019	104	66	61	-7	20	16	176	1	3	75
599	Cross Me (feat. Chance the Rapper & PnB Rock)	Ed Sheeran	pop	2019	95	79	75	-6	7	61	206	21	12	75
600	No Brainer (feat. Justin Bieber, Chance the Ra...	DJ Khaled	dance pop	2019	136	76	53	-5	9	65	260	7	34	70
601	Nothing Breaks Like a Heart (feat. Miley Cyrus)	Mark Ronson	dance pop	2019	114	79	60	-6	42	24	217	1	7	69
602	Kills You Slowly	The Chainsmokers	electropop	2019	150	44	70	-9	13	23	213	6	6	67

# Checking for duplicate values

```
1 song_df.duplicated()
executed in 20ms, finished 11:11:39 2022-07-08

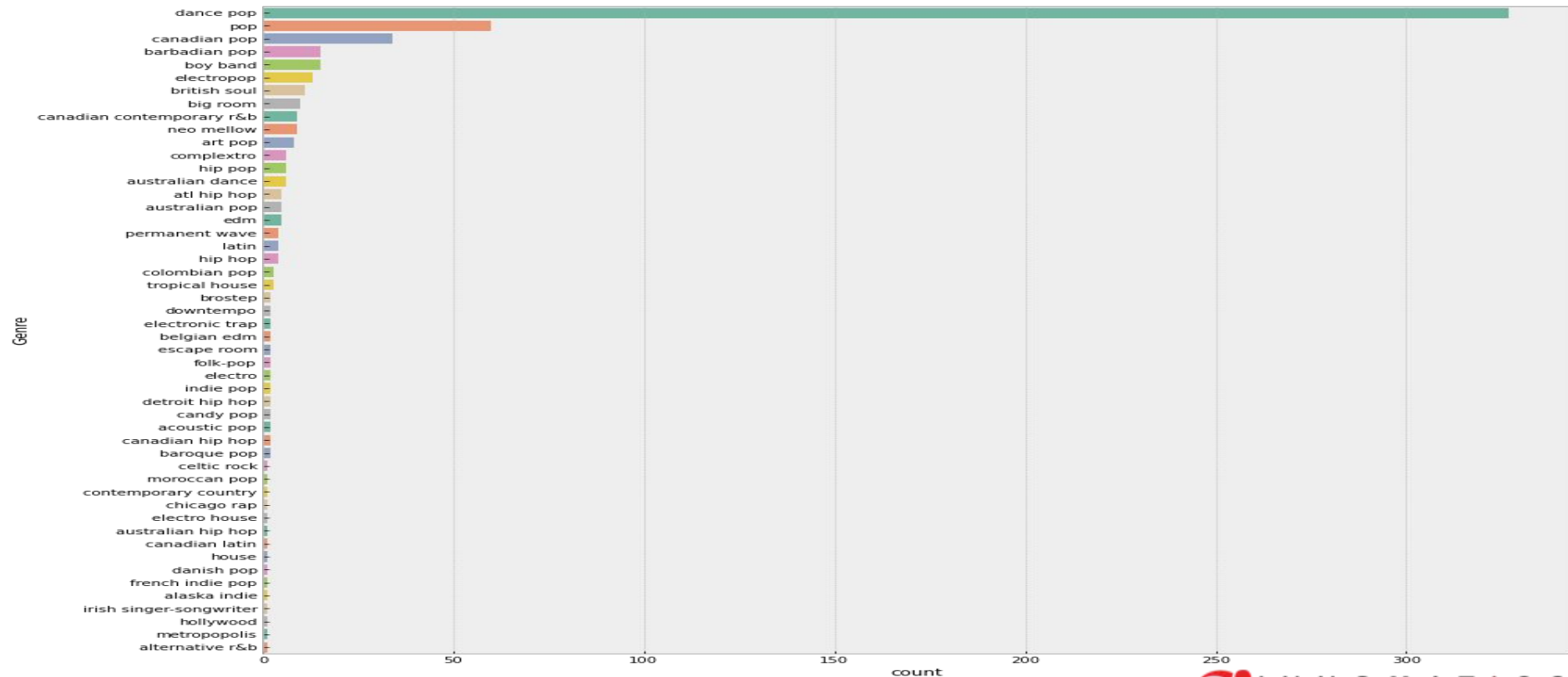
0      False
1      False
2      False
3      False
4      False
...
598    False
599    False
600    False
601    False
602    False
Length: 603, dtype: bool
```

# Checking for null values

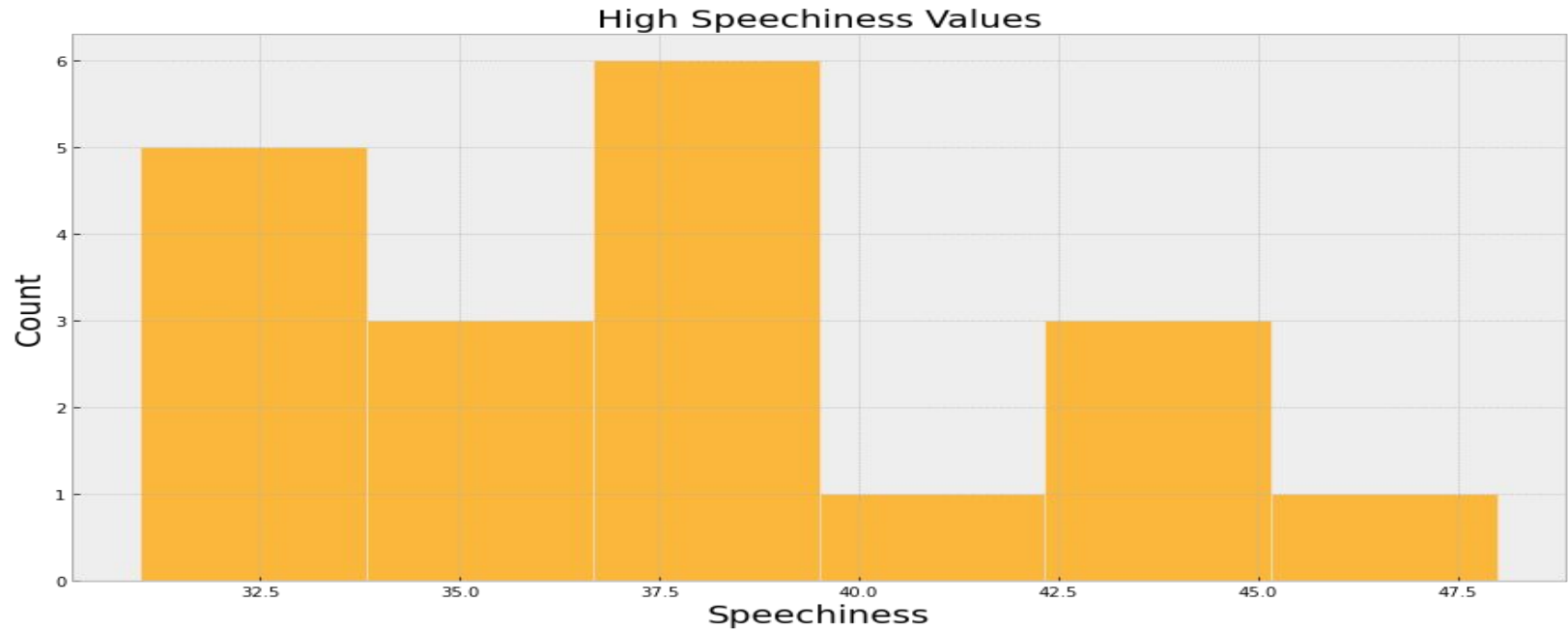
	Title	Artist_Name	Genre	Year	Beats_Per_MIN	Energy	Danceability	Loudness	Liveness	Valence	Duration	Acousticness	Speechiness	Popularity
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
598	False	False	False	False	False	False	False	False	False	False	False	False	False	False
599	False	False	False	False	False	False	False	False	False	False	False	False	False	False
600	False	False	False	False	False	False	False	False	False	False	False	False	False	False
601	False	False	False	False	False	False	False	False	False	False	False	False	False	False
602	False	False	False	False	False	False	False	False	False	False	False	False	False	False

603 rows × 14 columns

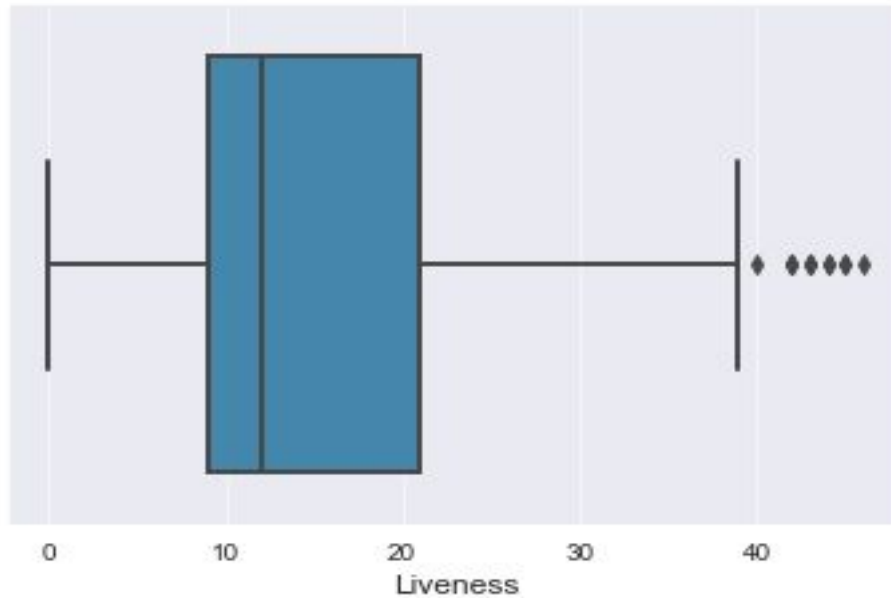
# Most number of genre of songs



# High Speechiness Values

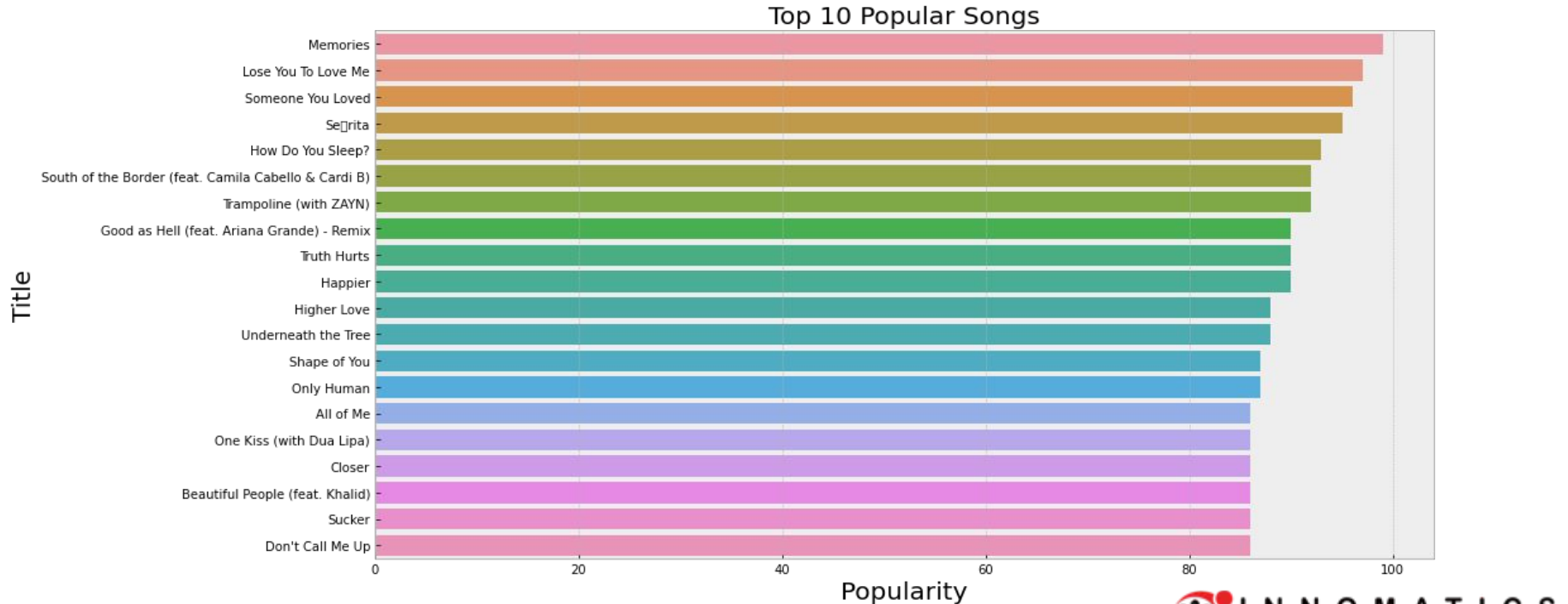


## Box plot for Liveness

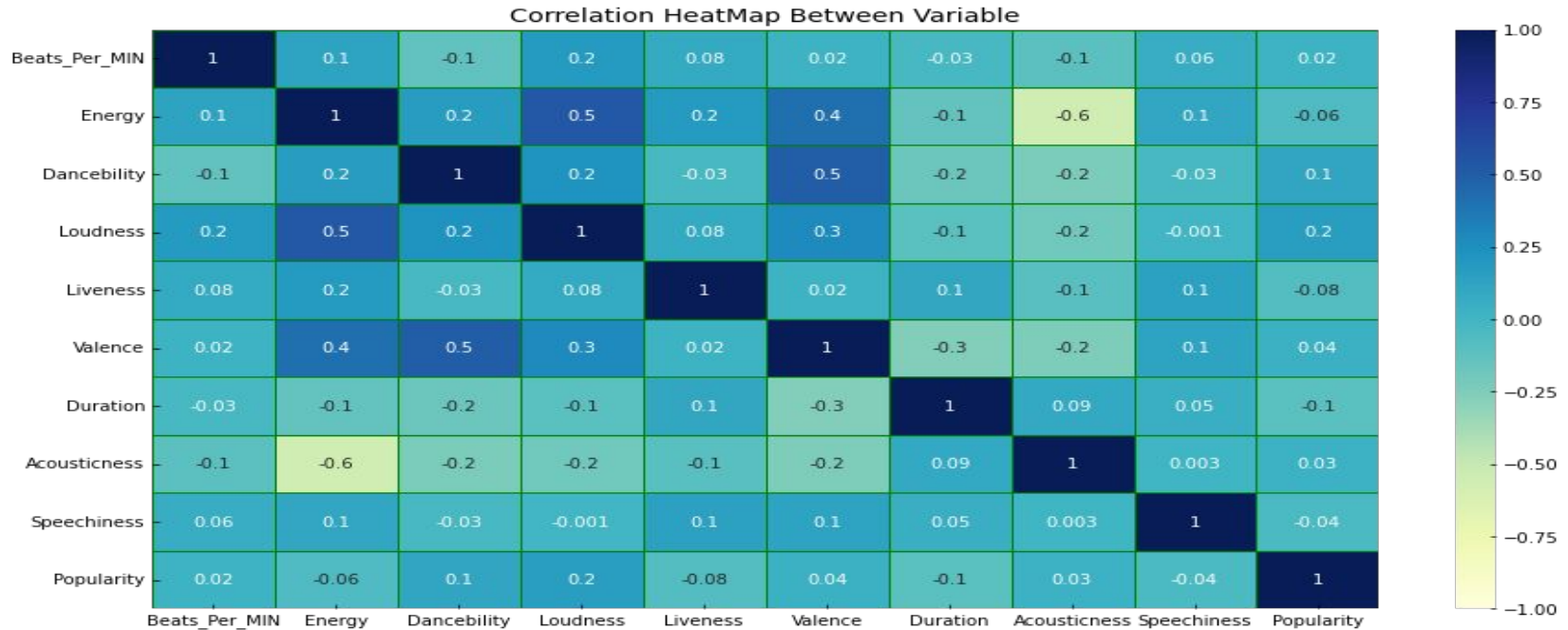




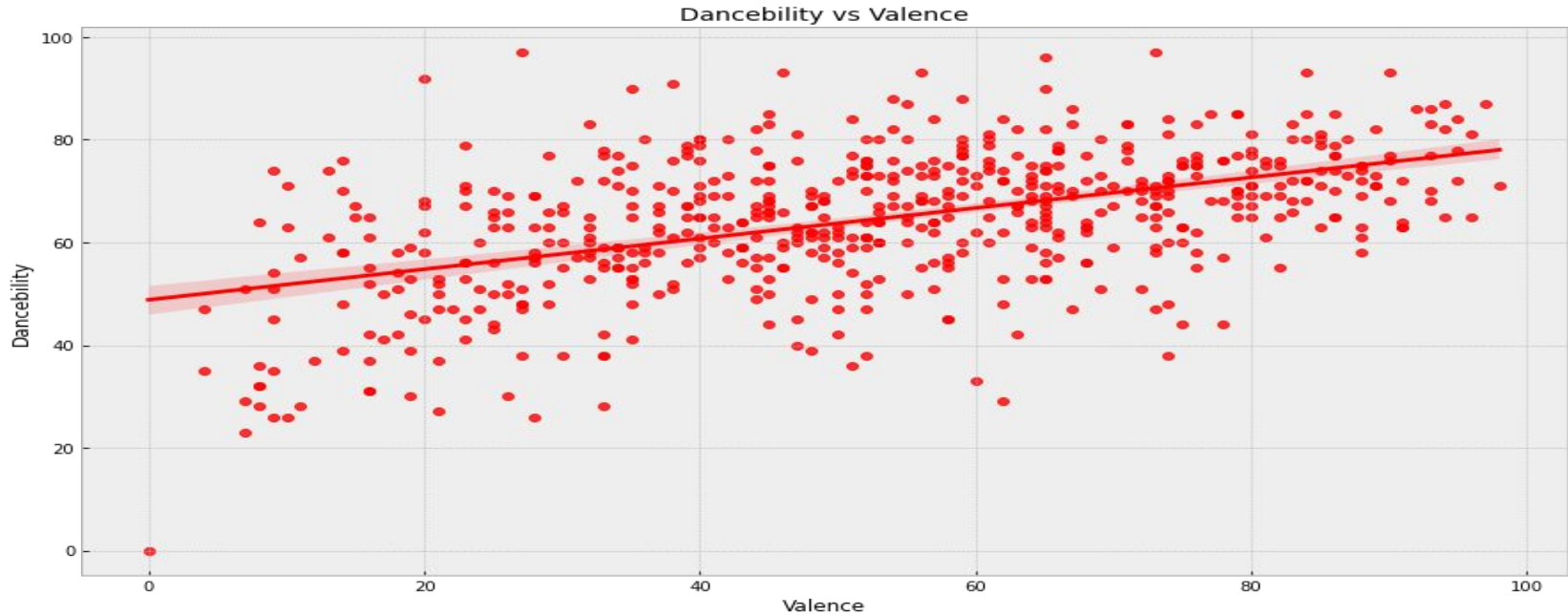
# Top 10 Popular Songs



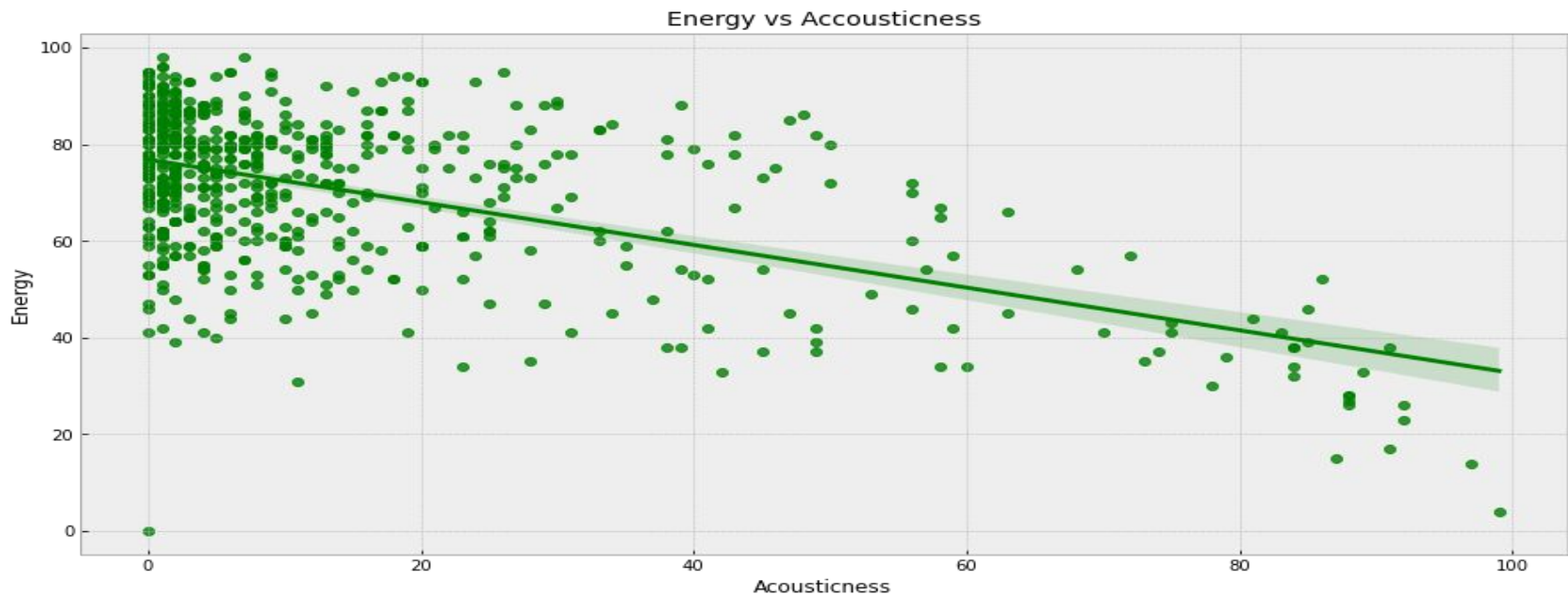
# Relationship between all the variables



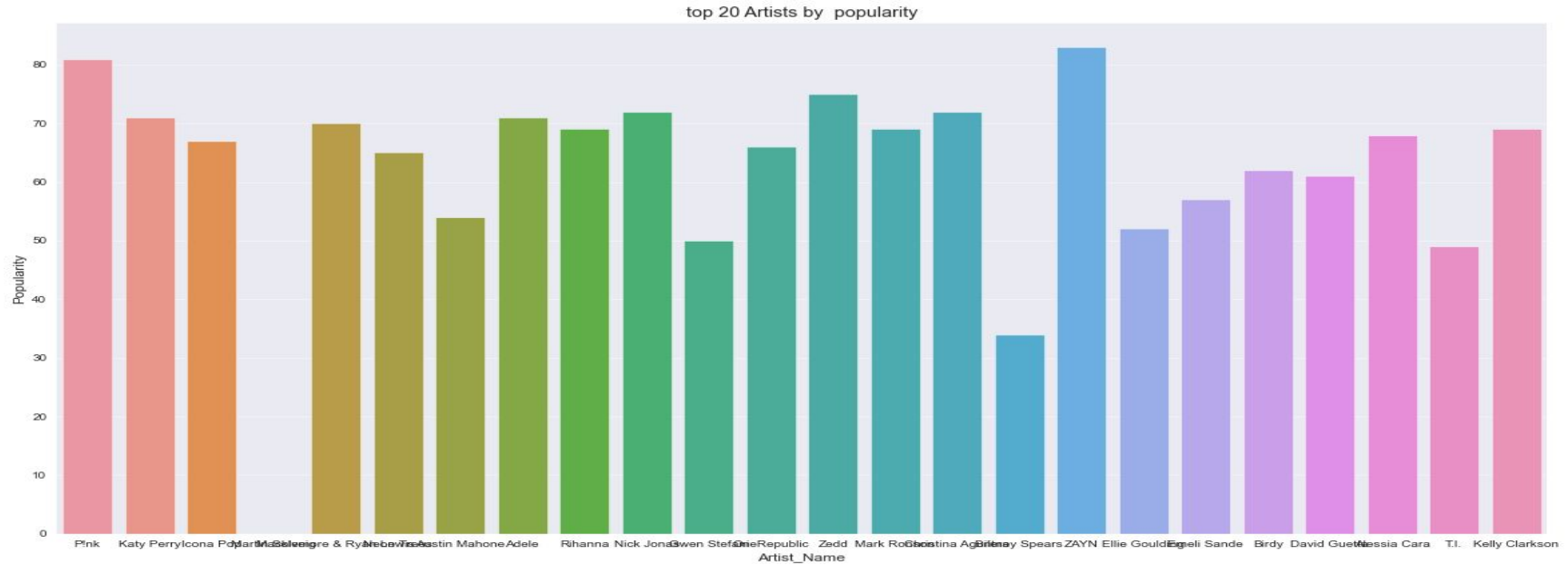
# Relation between Danceability and Valence



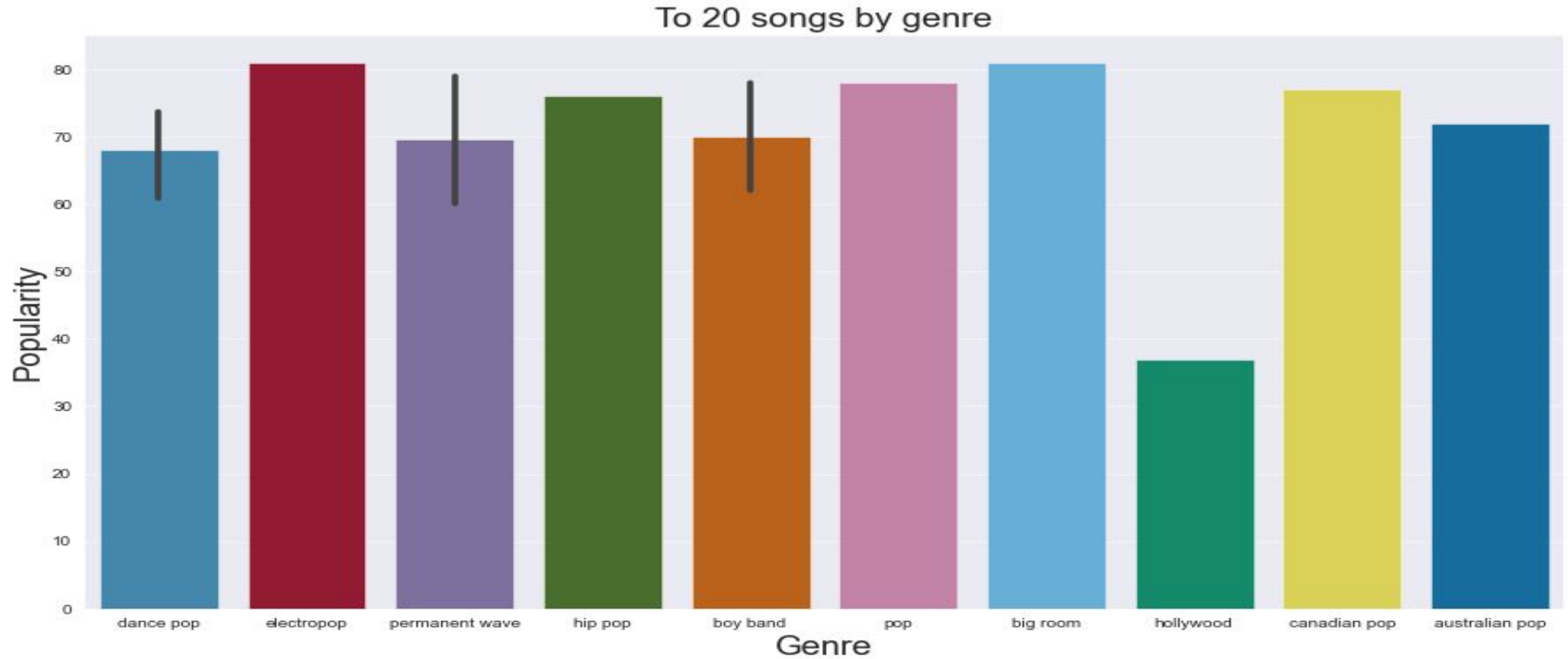
# Relation between Acoustictness and Energy



# Top 20 Artists by Popularity



# Top 20 songs by Genre



# Conclusions

- In overall analysis of Genre the most number of genre is Dance pop and when ,it comes to popularity of the genre dectro pop is the most popular one
- In overall analysis of speechiness 37.5 is the highest number of speechiness
- In overall analysis of top 10 songs analysis memories song is the top song
- In relation between Danceability and valence both danceability and valence has positive relation as danceability increases valence also increases and vice-a-versa
- In relation between Energy and Accousticness they have negative correlation as one increases the other decreases
- In analysis with top 20 artists by popularity Chainsmokers and Martin garrix are more popular

THANK  
YOU

