DATA SCIENCE Project : Spotify Data Analysis

Name - MUNISH SHARMA

Roll no. - 2018CSC1014

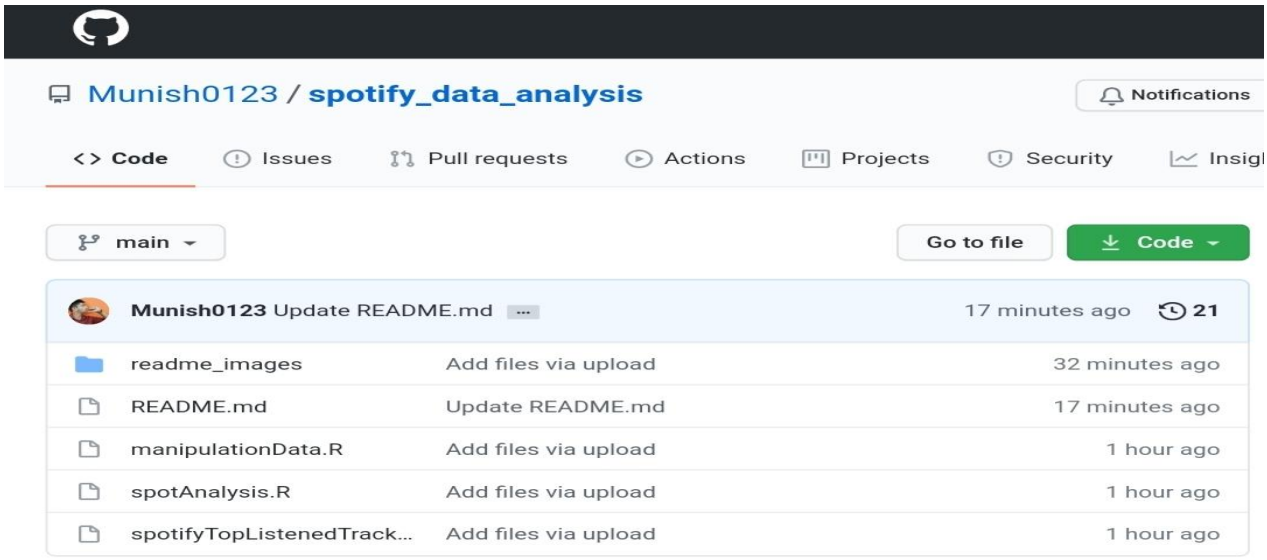
Subject - DATA SCIENCE

PROJECT NAME - SPOTIFY DATA ANALYSIS

Github Link:--

https://github.com/Munish0123/spotify_data_analysis.git

Screenshot:--



Data analysis…

	track.Name	track.Artist.Name	track.Album.Name
1	Bright Lights	Thirty Seconds To Mars	LOVE LUST FAITH + DREAMS
2	Cchaya	The Watson Brothers	Ohom
3	Some Say	Nea	Some Say
4	Everything I Own	Bread	Baby I'm-a Want You
5	If	Bread	Manna
6	Machinery	The Naked And Famous	Passive Me, Aggressive You (Remixes & B-Sides)
7	Canon in D Major, P.37	Johann Pachelbel	Chilled Piano Vol.3
9	Aubrey	Bread	Guitar Man
10	That's What I Like	Bruno Mars	24K Magic
11	Shanti	The Watson Brothers	Ohom
12	One Headlight	The Wallflowers	Bringing Down The Horse
13	Nidra	Cryptic Fate	Danob
14	Supermassive Black Hole	Muse	Black Holes and Revelations
15	Modern Loneliness	Lauv	~how i'm feeling~
16	...Shore Daraye, Shesh Barer Moto	The Watson Brothers	Ohom
17	Shattered	Trading Yesterday	More Than This
18	The Birthday Party	The 1975	The Birthday Party
19	In My Life	Imaginary Future	In My Life
20	Laid Low	The Naked And Famous	Simple Forms
21	Falling	First to Eleven	Falling
22	Now or Never Now	Metric	Art of Doubt
23	Adored	Collective Soul	Afterwords
24	Ahobaan	Cryptic Fate	Danob
25	Rain	Creed	Full Circle

track.Popularity	energy	acousticness	key	loudness
46	0.508	0.0466	0	-8.352
9	0.585	0.0903	1	-7.951
79	0.342	0.701	3	-6.316
69	0.338	0.735	2	-13.406
60	0.176	0.912	9	-16.952
18	0.816	0.00394	9	-3.991
58	0.122	0.988	2	-17.925
57	0.0902	0.647	7	-20.588
81	0.56	0.013	1	-4.961
4	0.542	0.0753	1	-9.669
71	0.539	0.000655	2	-8.968
2	0.41	0.935	9	-9.521
72	0.921	0.0492	7	-3.727
76	0.508	0.144	8	-7.105
6	0.604	0.0633	3	-5.974
51	0.263	0.355	5	-10.679
60	0.468	0.075	2	-9.736
55	0.0854	0.861	4	-15.775
37	0.888	0.03	6	-4.575
38	0.678	0.0291	1	-7.808
47	0.863	0.00731	1	-7.489
12	0.322	0.895	9	-9.195
2	0.346	0.739	2	-13.799
51	0.925	0.00209	11	-3.178

README.md

Spotify-data-Analysis

DATA Science project

Munish Sharma

Roll no. 2018CSC1014

Project name-- Spotify data analysis

Analysing top tracks on Spotify using spotifyr and ggplot2 packages (answering valuable questions with data)

Description:

This is a data visualisation project in R. I collected this dataset using spotify API with the help of spotifyr package.Web API provides access to user related data, like playlists and music that the user saves in the Your Music library. Such access is enabled through selective authorization, by the user. This is a datset of my top 100 most listened (or played) tracks on spotify.

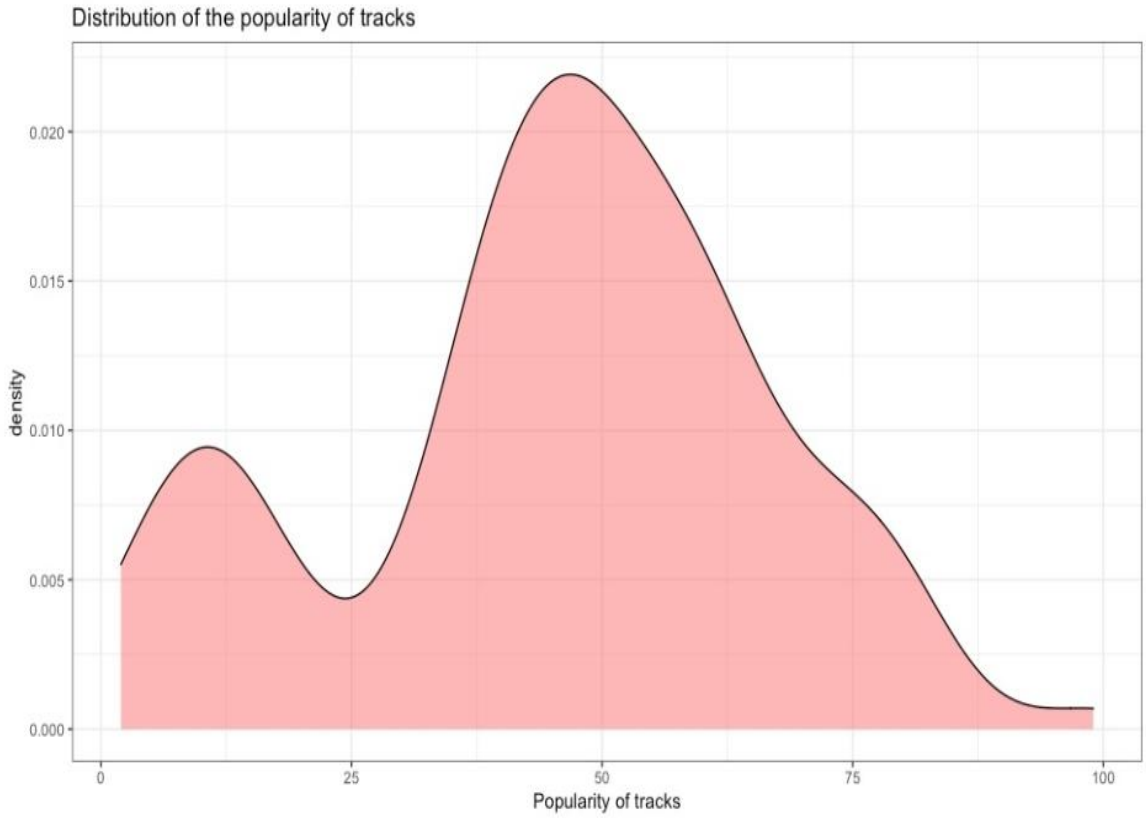
As data science is mostly about answering question with the help of data, visualising the data is a big part of it. So, In this project I tried to visualise with different plots(graphs) and tried to find some answer with the visualisation. which makes the things easy to understand

R packages used:

1. spotifyr
2. dplyr
3. httpuv
4. ggpubr
5. ggplot2
6. reshape2

Visualisation in detail:

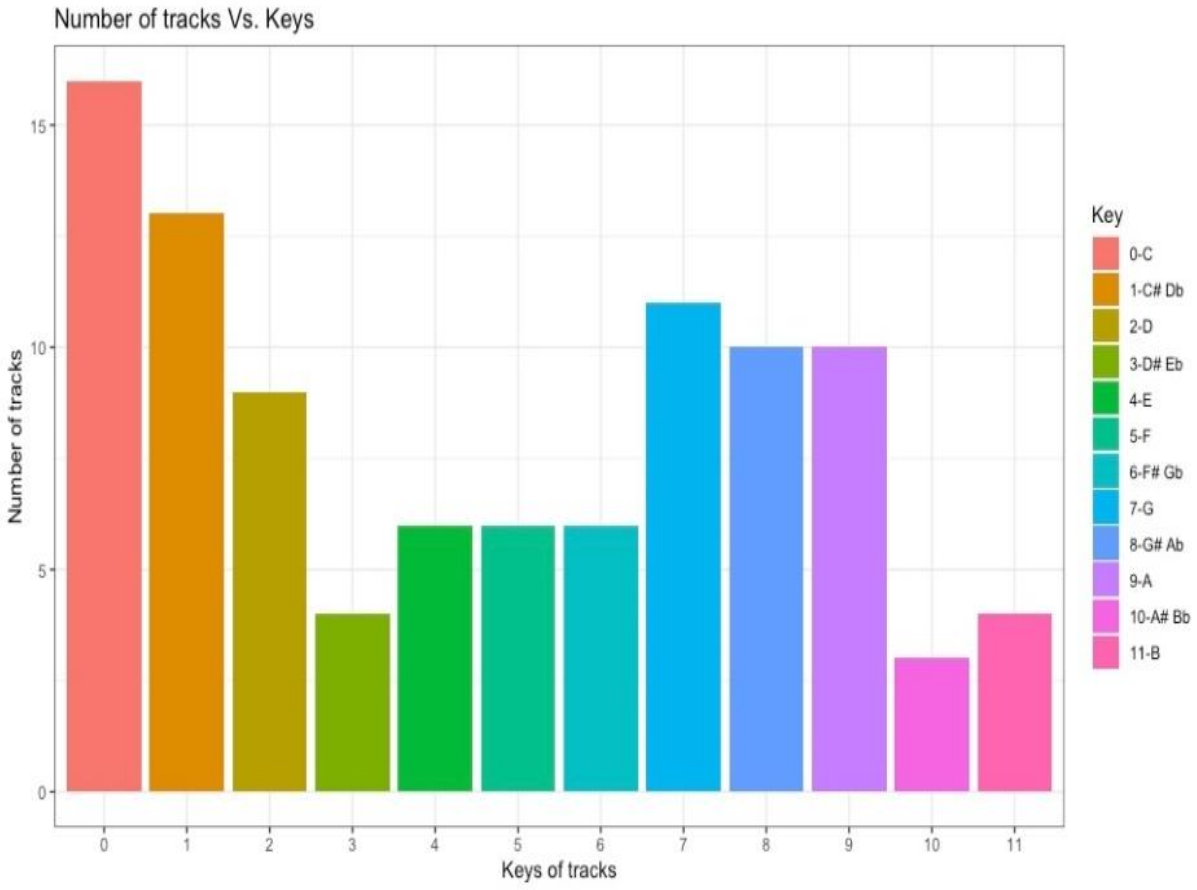
(Question-01): As these are the songs I have listened more. What is most of the track's popularity lies on the graph? Can I draw a statement that, my favourites have have specific range of popularity?



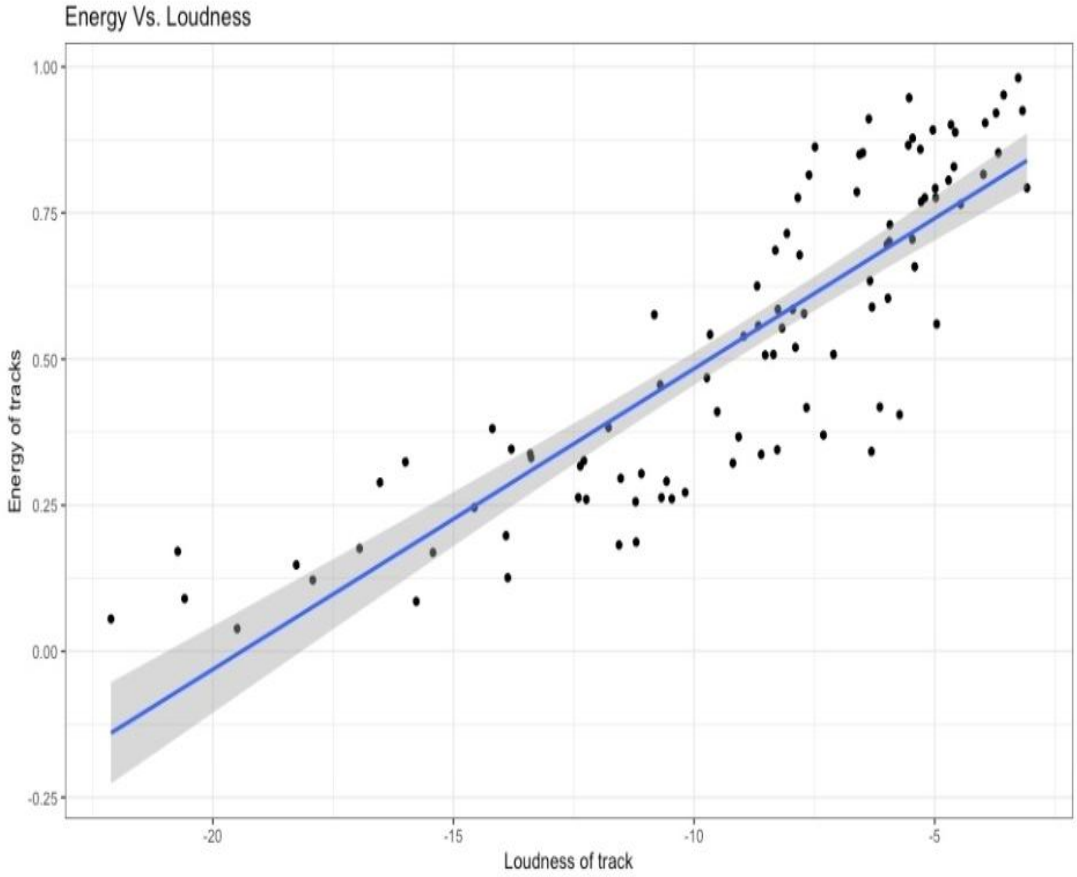
Here's the distribution of popularity of the tracks, by watching this plot I think I might have an idea that most of my listened track's have a popularity in between 35-70 (Guess am not a lover of highly popular songs)

(Question-02): If I group the tracks according to keys. What is the key that has most of the song, technically speaking, What is my most favourite Key?

Guess I am a fan of songs that is in scale C (noice..)

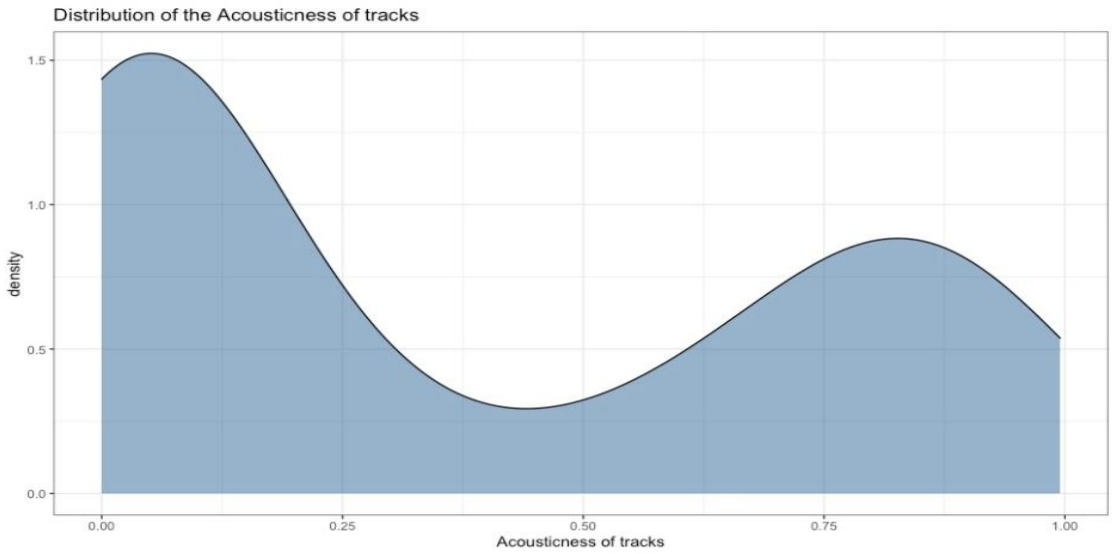


(Question-03): Does loud songs mean more energetic songs? most of the energy measured by loudness?

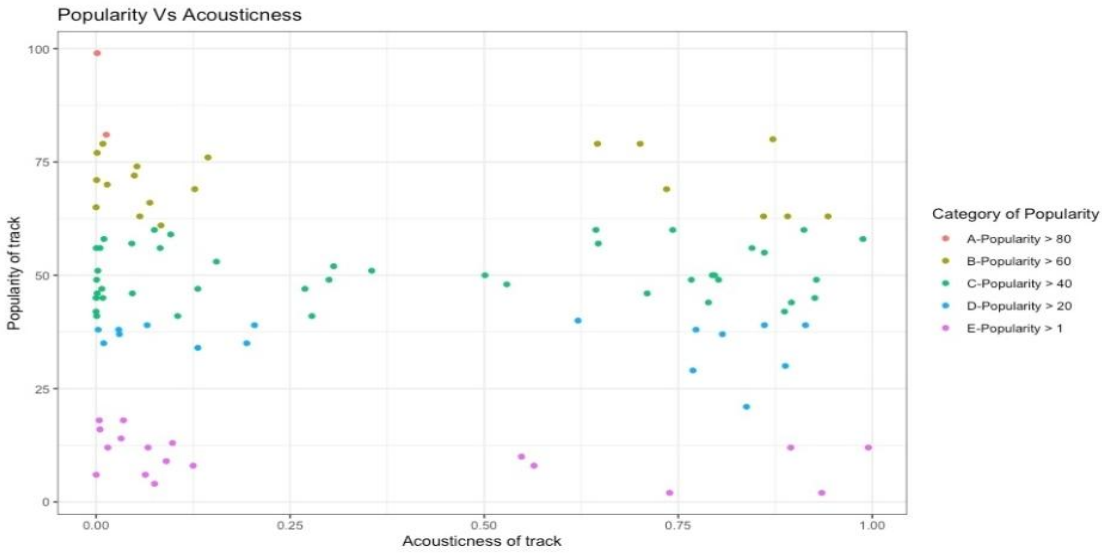


It looks like the relation between loudness and energy of a track is linear.

(Question-04): In my most played tracks what is the distribution of acoustic-ness? Do I prefer listening tracks those have specific acoustic range?

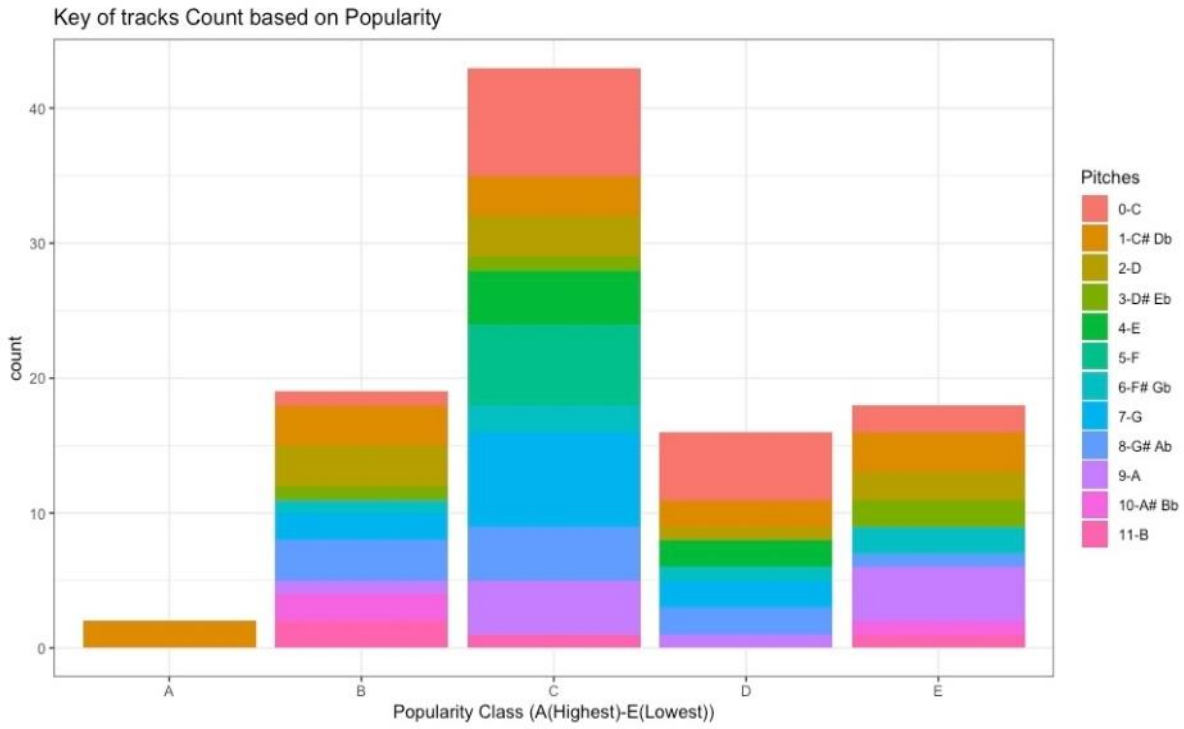


This one is tricky, maybe I like songs that have lower or higher acoustic-ness, maybe middle value are not that interesting to me. It's hard to tell, let's try another way:



This is 3-dimensional visualisation, the x and y axes are continuous variable and each of the track is coloured by it's popularity class. Popularity class was not on default, I created the column with conditions like if a track has a popularity in between 80-100 then it belongs to class-A and so on. It definitely gives us an interesting visualisation.

(Question-05): Based on popularity class, which key of tracks has been listened most(how much)?



clearly we can see most of my listened tracks are of class C (popularity between 40 and 60) which we thought 35-70 in our first plot, and inside of this class I think most of the songs are of scale C (what we found in Question-02). Impressive power of visualisation, right?

Now I am making a column on basis of tracks index. so, I am categorising all tracks in 5 groups where group A have first 20 tracks and Group B have next 20 songs and so on. Let's do it if we can find any interesting visuals.