SPOTIFY DATA ANALYSIS

BUDT 704 | 0501

Pythonic Playlisteners

Aarti, Danny, Mehul, Prathamesh, Nishit, Saloni

TABLE OF CONTENTS

1

INTRODUCTION

INTRODUCTION

WHY?

Spotify®

The amount of data people consume every day makes it a necessity for digital companies to fine tune their selling for each user.

Spotify, an online music streaming platform, puts user personalization at the forefront through recommended playlists. What sets Spotify apart from other music streaming platforms is precisely the spot on recommendations that it offers users for their playlists.

Through this project, we aim to see how any platform can improve its user personalization and its implications for the future of targeted marketing.

ABOUT THE DATA

This dataset consists of **1 million Spotify playlists** and is sampled from over 4 billion public playlists on Spotify. However, **our analysis was done only on the first 1000 playlists** and their respective songs to serve the purpose of this project.

2

DATA PROCESSING

DATA PROCESSING

DATA EXTRACTION

Since the **tracks are included as lists within the playlist, we extracted them** in order to display the track information separately before we dug deeper

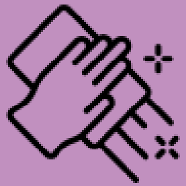
Additionally, since just the playlist data was insufficient, **we extracted additional song metrics from Spotify using API calls**

DATA WRANGLING

We also **needed the year that the playlist was modified** in order to analyse the trends over time

For that purpose, we **converted the timestamp of playlist modification from UTC format into datetime** format, extracting only the year

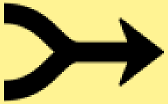
DATA PROCESSING



DATA CLEANING

We wanted to clean the data as it included data points which are out of the scope of our analysis.

The key tasks for this cleanup **included dropping redundant data points and retaining only those which would be useful** for our analysis.



DATA MERGING

To avoid doing API calls to Spotify everytime we wanted to use the data, **we consolidated the song metrics in a dataset**

Next, **we merged the 2 key data sets**: the first which had all the playlist data and the second one which had all the tracks data

3

DATA ANALYSIS

DATA ANALYSIS: FOCUS POINTS

TRACK ATTRIBUTES

Quantify Track Attributes and Correlations

ARTIST POPULARITY

Track Attributes which Contribute to Artist Popularity

RECOMMENDATION SYSTEM

Personalizing User Experience with Similarity Measures

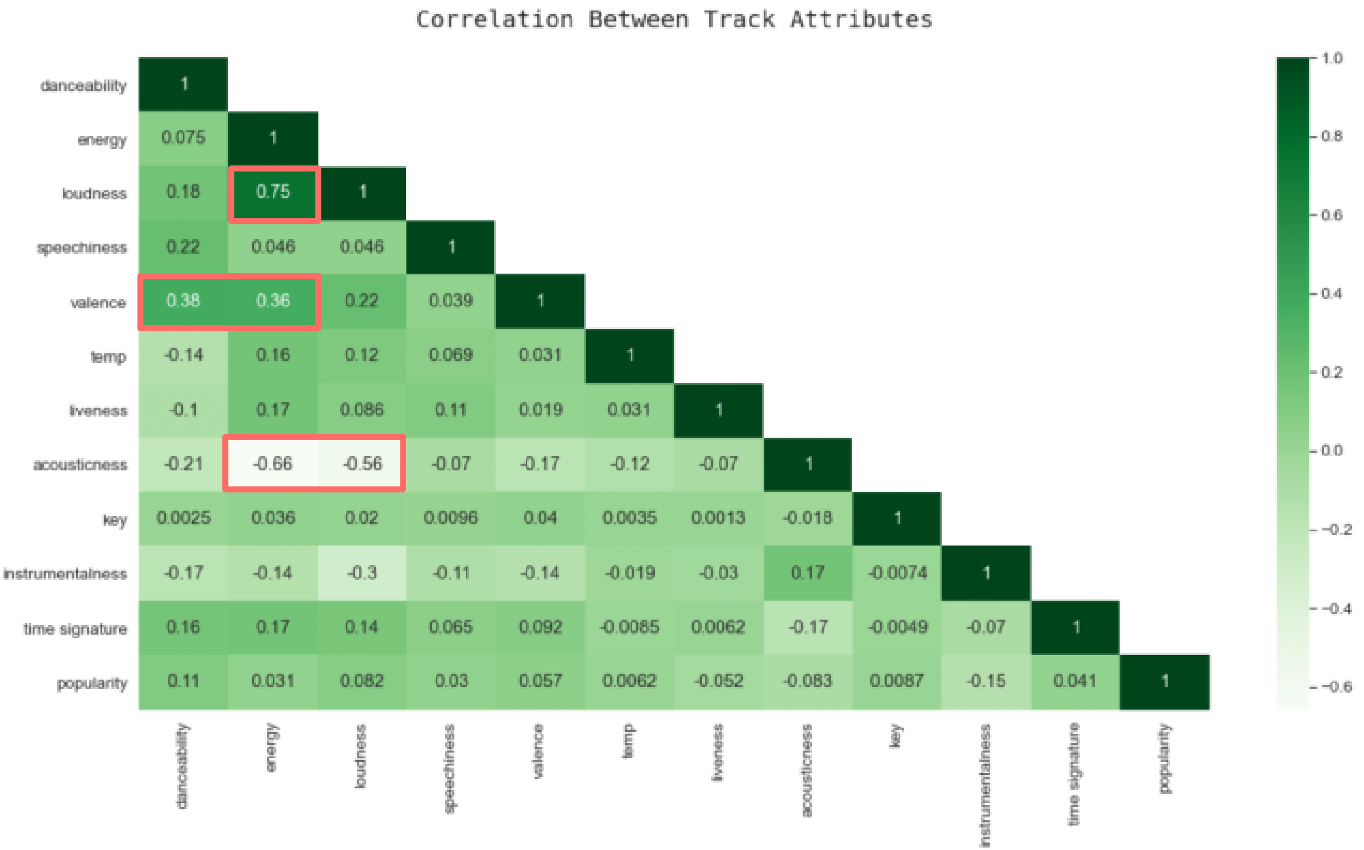
TRACK ATTRIBUTES

We aimed to understand the attributes of tracks with a heatmap correlation

- Danceability
- Energy
- Loudness
- Speechiness
- Valence
- Tempo
- Liveness
- Acousticness
- Key
- Instrumentalness
- Time Signature
- Popularity

All the track attributes were given as input and a correlation analysis was performed on them

RESULTS



Loudness v. Energy

Valence v. Danceability

Valence v. Energy

Energy v. Acousticness

Loudness v. Acousticness

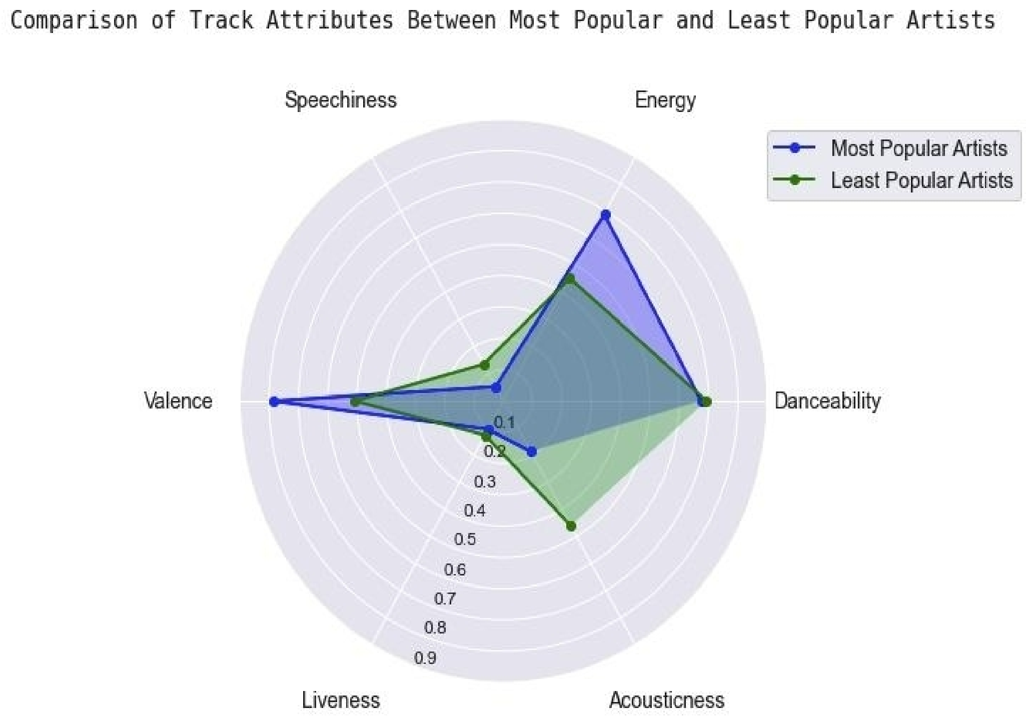
ARTIST POPULARITY

We aimed to understand the characteristics of the song which make an artist popular by analysing the attributes of all the songs released by them and the average popularity of the songs

Based on our previous research, to get the most accurate result, we narrowed down the list of attributes to:

- Danceability
- Energy
- Speechiness
- Valence
- Acousticness
- Liveness

RESULTS



Seeing that the most popular artists tend to have songs with a higher valence, energy and danceability, it can act as a formula for new musicians trying to build up the their name in the music industry.

They should **consider making songs that are positive and energetic** at the same time.

RECOMMENDATION SYSTEM

Now that we have sufficient knowledge about user's preference of songs and artists by analyzing song attributes and popularity, we have enough tools to build a recommendation system and personalize user experience.

To achieve this, we matched the genre of the song input by the user and its attributes to the rest of the songs in our dataset and generated the closest matching songs. The recommendation model is based on:

- Cosine similarity
- Hamilton similarity

RESULTS

Input:

```
Enter the Song Name    All Day
Enter the Artist Name    Kanye West
Your recommended playlist based on your song choice:
```

The recommendation system **suggests** songs that are similar in genre and most artists have similar level of popularity.

There's a **high likelihood these songs will be added by the user** to the playlist they want to build.

Output:

	track_name	artist_name
0	All Day	Kanye West
1	I Get The Bag (feat. Migos)	Gucci Mane
2	I Mean It	G-Eazy
3	Digits	Young Thug
4	No Problem (feat. Lil Wayne & 2 Chainz)	Chance The Rapper
5	i	Kendrick Lamar
6	Put On	Jeezy
7	Overnight Celebrity	Twista
8	We Dem Boyz	Wiz Khalifa
9	Day 'N' Nite (nightmare)	Kid Cudi
10	Real Big	Mannie Fresh
11	Luv U Better	LL Cool J
12	No Flex Zone	Rae Sremmurd
13	100 Shots	Young Dolph
14	Hustlin'	Rick Ross
15	CoCo	O.T. Genasis
16	Whatever You Like	T.I.
17	Litty (feat. Tory Lanez)	Meek Mill
18	Best Friend	Young Thug
19	Ridin'	Chamillionaire

4

FINAL THOUGHTS

FINAL THOUGHTS

Metrics are not individually related to popularity of song, but popular tracks are high in:

- Danceability
- Valence
- Energy

These attributes should be considered by upcoming artists and labels, but there may be external factors that can influence a track or artist's success. Examples include

- Chart performance
- Frequency of releases

The recommendation system highlights how quantitative metrics can influence recommendation systems in other media companies and media types.

THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution