



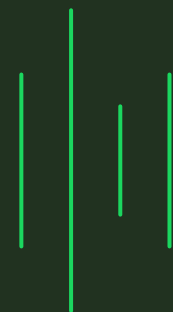
Big Data Management and
Processing

Music trend Analysis

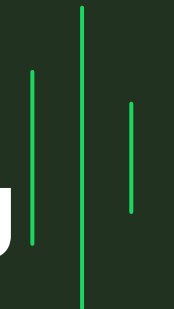
Martin Marinov and Nikola Mihaylov

Table of Contents

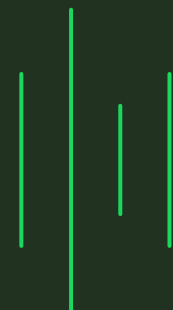
**01. Data
Ingestion**



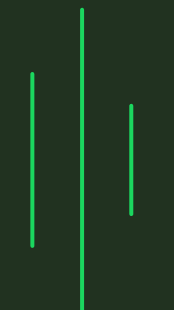
**02. Data
Preprocessing**



**03. EDA and
Insights**



**04. Pipeline
Automation**

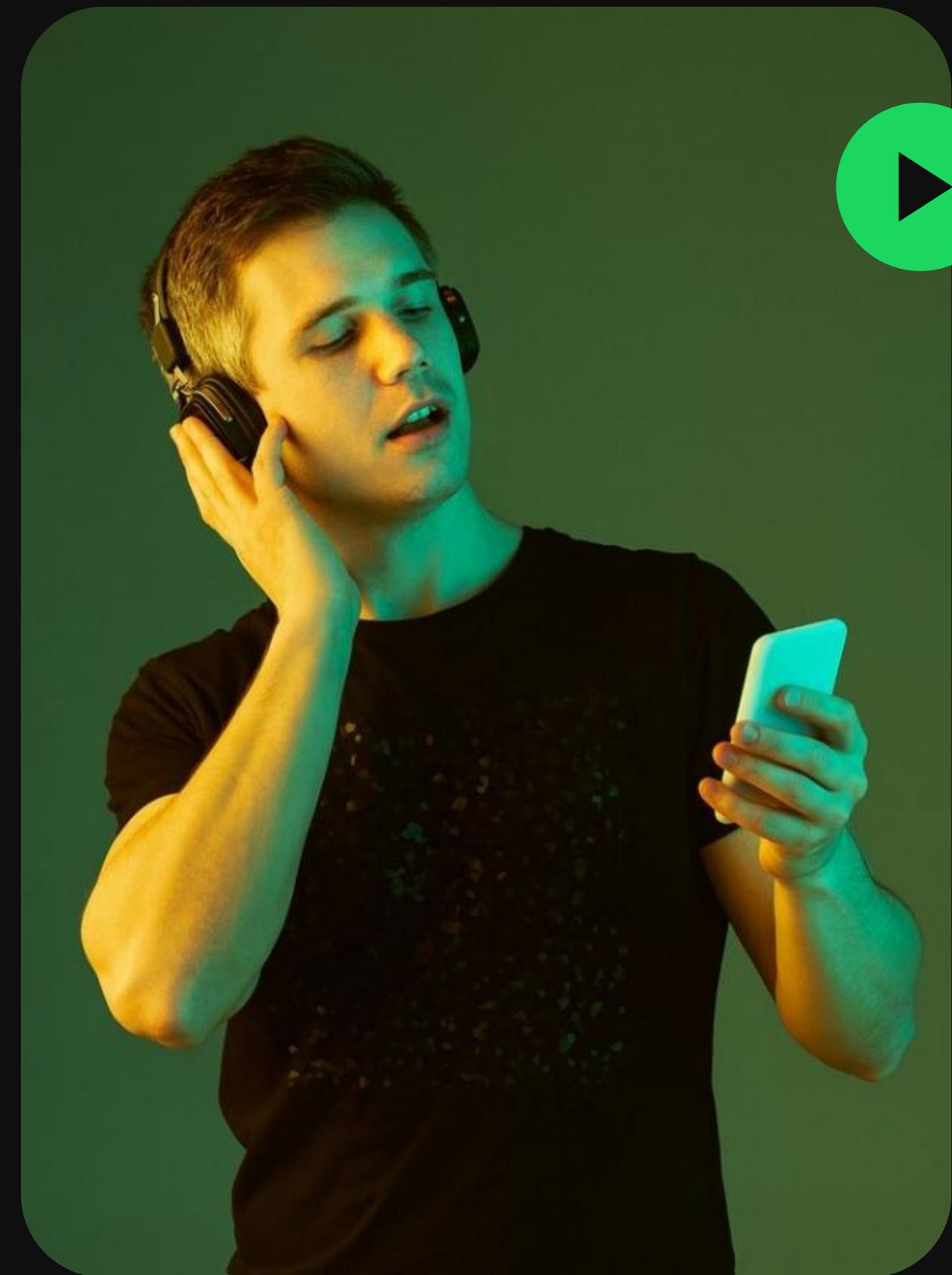




01. Data Ingestion

- ▶ Kaggle dataset called "Top Spotify Songs in 73 Countries (Daily Updated)"
 - As of January 9, 2025, the dataset comprises 1,584,257 rows and 25 columns, spanning the period from October 18, 2023, to January 8, 2025.;
- ▶ Acquired using the Spotify API

kaggle





01. Data Ingestion



#	Columns Overview				
1		name: The title of the song.		09:00 AM	
2		artists: The name(s) of the artist(s) associated with the song.		10:00 AM	
3		daily_rank: The daily rank of the song in the top 50 list.		11:00 AM	
4		country: The ISO code of the country of the Top 50 Playlist.		12:00 PM	
5		snapshot_date: The date on which the data was collected from the Spotify API.		01:00 PM	
6		popularity: A measure of the song's current popularity on Spotify.		02:00 PM	
7		danceability: A measure of how suitable the song is for dancing based on various musical elements.		03:00 PM	
+ 18 more columns					



02. Data Preprocessing



Data Cleaning

- Removing null values from the **name** and **artists** columns;
- Filling the null value in the **country** column with 'Global' value;
- Removing duplicates;
- Removing columns that don't give information we will use – **spotify_id**.





















Data Transformation

- Add new column - **continent**;
- Split the **artists** column and identify the **main_artists** and up to 3 feature artists;
- Convert the **snapshot_date** and **album_release_date** to pandas datetime.





03. EDA and Insights

#	Charts Overview				
1		Top 10 Artists by Appearance with and without features worldwide		09:00 AM	
2		Top 10 Songs and Albums by Appearance		10:00 AM	
3		Songs with the most #1 Rankings and artists with the most unique songs at #1		11:00 AM	
4		Top 10 most popular artists in every continent		12:00 PM	
5		Ranking each country based on musical characteristics such as danceability, energy, valence, tempo, and other key attributes		01:00 PM	
6		Album Distribution Across Different Time Intervals and the Top 5 Most Popular Songs in Each Interval		02:00 PM	
7		Most popular artists and songs in Bulgaria		03:00 PM	
8		Correlation between different musical characteristics		04:00 PM	



03. EDA and Insights



Most Popular Artists



#1 Feid



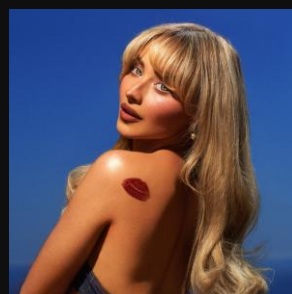
#2 Bad Bunny



#3 Karol G



#4 Billie Eilish



#5 Sabrina Carpenter



#6 Taylor Swift



#7 Peso Pluma



#8 The Weeknd



03. EDA and Insights



"Who" - Jimin

1337

#1 Rankings

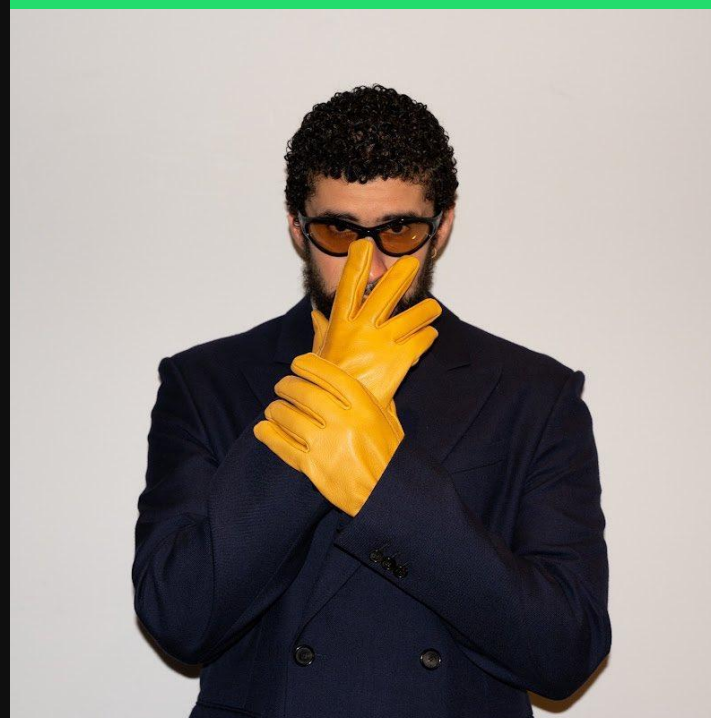


Bad Bunny

8

Unique Songs

#1 Ranking



Most Consecutive #1 in the
"Global" chart



Die With A Smile

Lady Gaga, Bruno Mars



26/08/2024

57 days

21/10/2024

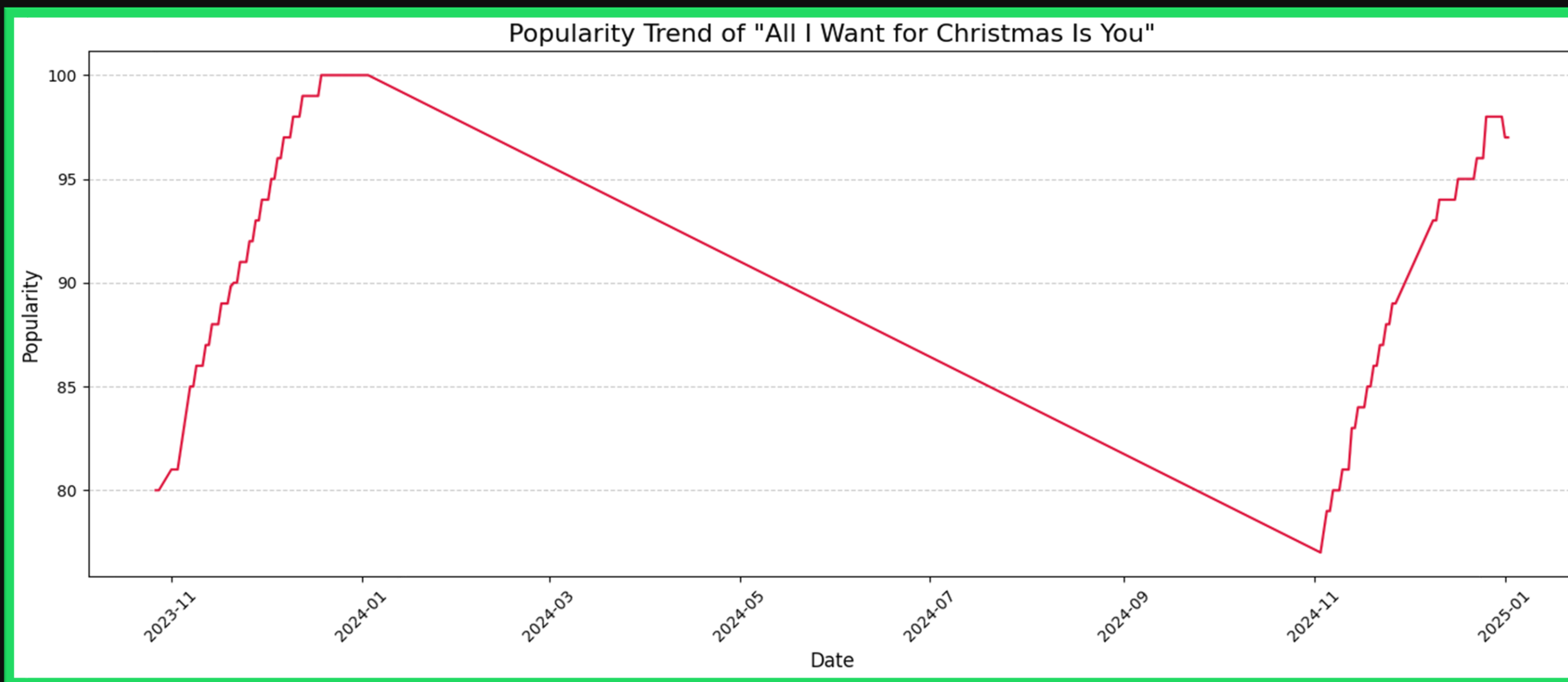




0.55



03. EDA and Insights





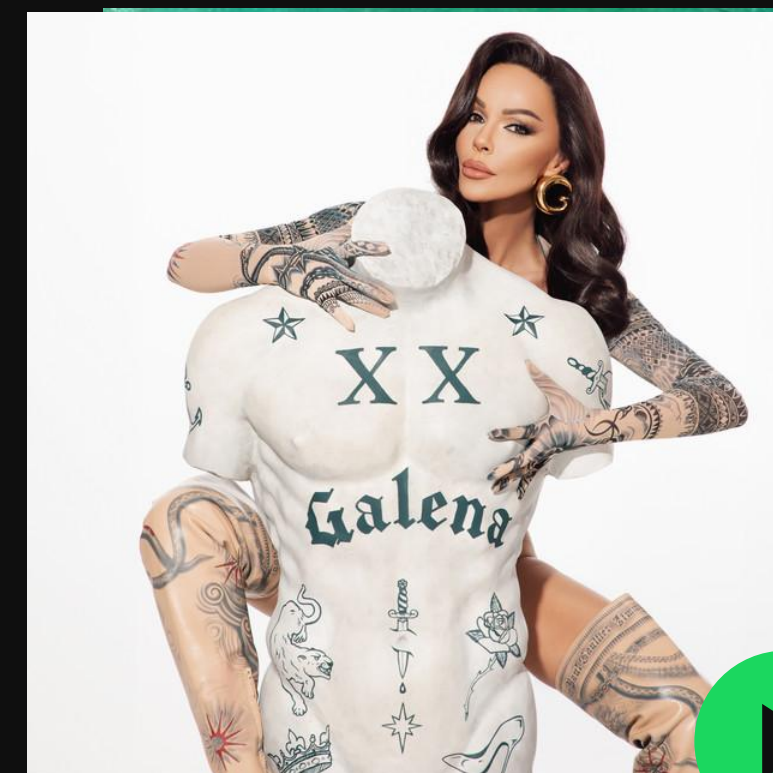
03. EDA and Insights



#1 Medi



#2 Djordan



#3 Galena





04. Pipeline Automation



01

Store the processed data in CSV files

02

Automate the pipeline with Apache Airflow

03

Visualize the created plots



PLAY

FOLLOWING



Thank you for your attention!

Questions?



In session

Jan
20

**Development of Big Data
Management and Processing**
Mon 12:00 PM | GATE Institute
Coursework

Jan
27

**Development of Big Data
Management and Processing**
Mon 3:00 PM | GATE Institute
Exam

Feb
03

Data Enrichment
Mon 3:00 PM | GATE Institute
Exam