

The

# Music

# Analysis

Project



# The Tasks

The main sprint project will be to analyse **Billboard data** from the last 25 years combined with **Spotify Track Audio data**.

There are two parts to this:

1. Build an **interactive dashboard** in smaller groups.
2. **Conduct an individual analysis** in a Jupyter Notebook

You will be graded on both these deliverables. Dashboards will be presented and notebooks handed in on **January 23th**.



# Course Goals

# Course Goals

K3

Describe key Python libraries  
and what they are used for.

# Course Goals

K4

Outline the role Python plays in  
**building data infrastructures**  
such as data structuring, data  
flows/pipelines as well as  
**exploratory data analysis** and  
**statistical analysis**

# Course Goals



**Access, transform, clean and explore data through Python based on previous learnings of data structuring.**

# Course Goals



Create **data visualisations**  
programmatically using  
Python libraries.

# Course Goals

**K3**

Describe key Python libraries and what they are used for.

**K4**

Outline the role Python plays in building data infrastructures such as data structuring, data flows/pipelines as well as exploratory data analysis and statistical analysis

**S3**

Access, transform, clean and explore data through Python based on previous learnings of data structuring.

**S4**

Create data visualisations programmatically using Python libraries.

# Individual

# Individual

## What?

You are assigned to conduct an **individual analysis of the datasets**. It should be understandable for any data-literate reader.

In your analysis you should also explain what you are doing, and why.

Find something that sparks your interest, and that you want to explore.

## How?

Make your analysis in a Jupyter Notebook. **Use the Pyramid-principle to structure your analysis.** Make sure that there is an Exploration as well as a Presentation part.

You are free to have classmates review your work - but please do not copy work from each other.

## Why?

By continuously analysing new datasets, you will grow your knowledge in Python. By working on your own, you will be challenged to really dig into the data yourself.

This is an interesting set of tables, and it will give you a lot of space to experiment and showcase your knowledge.

# Group

# Group

## What?

This is a group assignment where you will use Python and the library Streamlit to build an **interactive dashboard intended for a group of stakeholders**.

Who these stakeholders are will be up to you to define - and based on that definition you will build the dashboard.

## How?

This project will require you to write Python in an IDE like VS Code - as well as collaborating with your classmates. Depending on the complexity of your dashboard, you might want to share code using GitHub.

**Presentation will be 10 minutes**, showcasing your dashboard and some parts of the code

## Why?

Making data available and understandable to stakeholders is equally as important as being able to answer complex questions.

This assignment will challenge you to work as a group, learn new tools and **exploring the art of defining and building dashboards**.

# Datasets

Chart Performance

Tracks

Artists

Mapping

Audio Features (1)

Audio Features (2)

This dataset provides information about how tracks perform on Billboard Top 50 weekly charts from 2000 to 2024.

**File:**

chart\_positions

**Columns:**

chart\_week

The week of the chart in a string format (e.g., "2024-10-05").

list\_position

The position of the track on the chart during that week, with 1 being the top position.

**track\_id**

A unique identifier for each track.

**Legend**

Primary Key

Foreign Key

# Datasets

Chart Performance

Tracks

Artists

Mapping

Audio Features (1)

Audio Features (2)

This dataset contains general metadata about tracks, including their release information.

**Table:**

tracks

**Columns:**

**track\_id**: str

A unique identifier for each track.

name: str

The name of the track.

duration\_ms: int

The length of the track in milliseconds.

release\_date: str

The date when the track was released.

album\_type: str

The type of album the track is from (e.g., single, album).

explicit: bool

A boolean indicating whether the track contains explicit content.

**Legend**

Primary Key

Foreign Key

# Datasets

Chart Performance

Tracks

Artists

Mapping

Audio Features (1)

Audio Features (2)

This dataset contains information about artists and their current popularity.

## Table:

artists

## Columns:

**artist\_id**: str

A unique identifier for each artist.

name: str

The name of the artist.

popularity: int

A numeric score (0-100) representing how popular the artist is on Spotify. right now. A higher number means more popularity.

followers: int

The number of people following the artist on Spotify.

## Legend

Primary Key

Foreign Key

# Datasets

Chart Performance

Tracks

Artists

Mapping

Audio Features (1)

Audio Features (2)

This dataset links artists with the tracks they have created. Remember that one track may have multiple artists!

**Table:**

tracks\_artists\_mapping

**Columns:**

**artist\_id:** str

A unique identifier for each artist.

**track\_id:** str

A unique identifier for each track.

**Legend**

Primary Key

Foreign Key

# Datasets

Chart Performance

Tracks

Artists

Mapping

Audio Features (1)

Audio Features (2)

This dataset contains specific datapoints about the audio features of the tracks.

## Table:

audio\_features

## Columns:

**track\_id**: str

A unique identifier for each track.

danceability: float

Value (0-1) representing how suitable the track is for dancing based on tempo, rhythm stability, beat strength, and regularity.

energy: float

Value (0-1) that measures the intensity and activity of a track. Higher values represent more energetic songs.

key: int

Representing the musical key of the track.

loudness: float

Value (0-1) representing the overall loudness of a track in decibels (dB).

mode: int

Value representing the modality (major or minor) of the track, where 1 = major and 0 = minor.

## Legend

Primary Key

Foreign Key

# Datasets

Chart Performance

Tracks

Artists

Mapping

Audio Features (1)

Audio Features (2)

## Columns:

speechiness: float

Value (0-1) that detects the presence of spoken words in the track. Tracks closer to 1 are more speech-heavy.

acousticness: float

Value (0-1) that indicates the likelihood that the track is acoustic.

instrumentalness: float

Value (0-1) predicting whether a track contains no vocals.

liveness: float

Value (0-1) that detects the presence of a live audience in the recording.

valence: float

Value (0-1) describing the musical positiveness conveyed by a track, where higher values indicate more positive emotions.

tempo: float

Value representing the track's tempo in beats per minute (BPM).

time\_signature: int

Value representing the overall time signature of the track.

**Legend**  
Primary Key  
Foreign Key

# Questions