

Data Science Project

Week 2

Henri Heyden Mika Materzok Nike Pulow Senanur Şimşek
Group “Für Fortnite”

Christian-Albrechts-Universität zu Kiel

March 14, 2025

Overview

- 1 Main Question
- 2 Aspects to analyze
- 3 Datasets and Apis
- 4 Plans

“How do various musical elements vary in the most popular songs of the last 20 years?”

Overview

- 1 Main Question
- 2 Aspects to analyze
- 3 Datasets and Apis
- 4 Plans

- What chord-progressions repeat most frequently? Nike & Henri
- Do certain chord-progressions dominate specific genres? Mika & Henri
- What chords and chord types (minor, major, ...) dominate / how did the amount change? Nike & Henri
- What key signatures dominate? Sena

- How did song...
 - ▶ ...duration **Sena**
 - ▶ ...tempo **Sena**
 - ▶ ...polarity **Mika**
 - ▶ ...metrics on Spotify (e.g. "*danceability*") **Sena**

...vary and change over the years? Do they correlate?
- Which words and combinations are the most frequent? **Mika**

Overview

- 1 Main Question
- 2 Aspects to analyze
- 3 Datasets and Apis**
- 4 Plans

Datasets and Apis

What we have...

- **Main source:** “Hot 100 Songs” from the Billboard Year-End of 2005 up to 2024 **Nike**
- **Refined sources using songs from the main source:**
 - ▶ Ultimate-Guitar Webscraping links for tabs **Nike**
 - ▶ Ultimate-Guitar Webscraping tabs-links to chords **Henri**
 - ▶ Spotify-API for features **Sena**
 - ▶ Last.fm API for “topTags” (genres) **Mika**
 - ▶ azapi API & Genius API for lyrics **Mika**

Overview

- 1 Main Question
- 2 Aspects to analyze
- 3 Datasets and Apis
- 4 Plans**

Plans

What we do with it...

- Clean collected URLs for chords to filter out wrong results **Nike**
- Finalize and clean chords dataset from URLs **Henri**
- Identify main harmonies of songs by determining key signatures from chords **Henri**
- Extract relevant metrics and lyrics from Spotify, Last.fm, azapi/Genius **Sena & Mika**
- Implement algorithms for the concrete data analysis **Nike, Henri, Sena & Mika**

Questions?