

Task / Lead Data Scientist AI and Forecasting

Attached is a CSV file with weekly Top-200 Spotify chart data for Germany from January 2022 through August 2024.

We would like you to analyze the data to present a brief overview of the German Spotify Charts and create a model for predicting a track's chart performance.



1. Please conduct an initial EDA. What trends and developments can you identify in the charts? The following are some suggested areas of investigation, but we welcome any additional insights you may find:
 - What are the market shares of different genres, major labels, and track release ages (e.g., 80s, 90s, 2020s, frontline releases), and how have these shares changed over the past few years?
 - How long does it take a track to reach its peak position? How many streams does it take to reach the Top 10? How long does a track stay in the Top-200? Did this change over the last few years?
2. Develop a model to **predict** the **peak position** of a track as early as possible and **how many streams** it will generate during the first 6 months in the charts?

You are free to choose your own tools and presentation format. The following approach is only a suggestion:

- Load the CSV into a local database (PostgreSQL, MySQL, SQLite).
- Write a few SQL queries to analyze the data. You can support your EDA using a tool like Google Looker Studio, a free trial version of Tableau or other dashboarding tools.
- Conduct the forecasting task in a Jupyter Notebook.
- Prepare a 15-min presentation to showcase your approach and results.

Please keep in mind that your approach should be reproducible and that we are also interested in how you built your data process / pipeline. Please delete your copies of the data after the interview process.

The Spotify chart data has the following columns:

chart_market	Country of the Top-200 charts
week_start_date	Calendar date of first day of the week of the charts
chart_position	Which position did the track have in this week
spotify_artist_id, artist_name	The Spotify ID for the main artists of the track and the artist names for the track
isrc	ISRC stands for international standard recording code and is a music-industry-wide identifier for an individual track
spotify_track_id	The Spotify ID for a track
label	The name of the label that owns the track
major_label	The Major Label and other large labels that own the rights to this track
genre	Information about which genre the track might have
release_date	The date the track was released
streams	How many streams did the track have in this week on Spotify in Germany