

Alfred Custodio
Diana Melendez
Jon Arnold
Jorge Serrano

September 8th, 2021

ETL PROJECT 2 – Proposal GROUP 1

Objective: The objective is to analyze the different top music charts from different sites with an analysis of genres, artists, and streaming services.

Background and Inspiration: Viewing top 100 music charts is a great way to keep up with current trends and genres, but with music being accessible in digital form via streaming, more information can be gathered than just record sales and radio station requests. Since websites have different acceptance criteria for their top 100 lists, we want to observe if there are any differences in these lists regarding the general top 100.

Proposal: We will be reviewing the top 100 songs from Billboard.com, Rollingstone.com, and Music.Apple.com. These charts are updated on a weekly basis, so we will have a snapshot of the most current data. A PostgreSQL database will be created on GCP to handle the load process when ETL is performed on the websites.

Data Overview: Team will Web Scrape from Billboard Top 100 site, Rollingstone.com and Music.Apple.com using 'BeautifulSoup;. The extracted information will be stored in dataframes for cleaning and normalization before they are loaded in the database. Queries using joins and various 'where' and 'group by' clauses will allow us to analyze the data after ETL is complete.

Tables: Artist ID lookup table, Website ID lookup table, and Master Table with Rank, Song, Artist ID and Source ID.

Roles:

Alfred Custodio - Database

Diana Melendez - Database

Jon Arnold - Web Scrape 1

Jorge Serrano - Web Scrape 2

Report – All

SOURCES

- [Rolling Stone](#)
- [Billboard](#)
- [Apple Music](#)