ETL Challenge- Top Songs

Team Details:

Satish Nirankari

Thiripurasundari Kannappan

Overview

This project analyzed top songs data from 2010-2019 and find out if those songs have won Grammy Awards.

What we can analyse from the final data:

- If top songs are getting nominated for Grammy Awards.
- Number of top songs per artist.
- Which genre has the maximum number of top songs.
- Number of wins vs nominations.

Extract:

With the help of Pandas we read csv files and saved them in DataFrame.

Data Source:

We used "Kaggle" to extract csv data to do the analysis:

      A- [Top Songs](#)

      B- [Grammy Awards](#)

Transformation:

1. **Artist**: Used songs df to extract unique artist names and given a sequence generated value as artist id.
2. **Genre**: Used songs df to extract unique genre names and given a sequence generated value as genre id.
3. **Songs**: Used top_songs csv file to get the songs name and year and included the genre_id and artist_id as foreign key from their respective dataframes.
4. **Accolades**:Used grammy award csv file to get the year,category,worker , nominee, winner and included artist_id as foreign key from artist dataframes.

Load:

Using QuickDBD we created ERD and exported the [PostgresSQL](#) , this export was used to create table structure in musicdb database.

Used sqlalchemy for loading all the dataframe in the respective tables.