**Streaming to the Top:**

**Have The Grammys become the next thing to be Made By Spotify?**

*An ETL Exercise By: Jon + Yeni*

**DATA SETS:**

* We’re using this Spotify CSV data from Kaggle:

<https://www.kaggle.com/leonardopena/top-spotify-songs-from-20102019-by-year>

* We’re using this Grammy nominee data from Kaggle:

<https://www.kaggle.com/unanimad/grammy-awards>

**EXTRACT:**

*One CSV will be imported to pandas using from\_csv, the other (after converting it to sql in Postgres) will be loaded using from\_sql with pandas*

* Download Kaggle CSVs [5 min] -Jon + Yeni
* Import Grammy CSV and create a new column name joining the artist and song in Pandas. [15 min]-Yeni
* Import Spotify CSV into PGAdmin. [25 min]- Jon + Yeni
* Import Spotify SQL into Pandas and create new column in dataframe merging song title and artist name - [40 min] - Jon
* Create repo to house notebook and sql codes. [10 min]-Yeni

**TRANSFORM:**

* Once all data imported into Pandas, we will merge the dataframes on the newly created column (artist name & song title) [5 min]-Jon
* Rename all columns in Pandas and drop undesired columns to match desired PostGres table [10 minutes] -Jon + Yeni

**LOAD:**

* Insert merged dataframe to PGAdmin [10 min]- Jon
* Run queries to view final tables on PGadmin-[2 min]- Jon

Notes:

This was slated for approximately 2 hours. If we finished early, we intended to explore and uncover different queries in data analysis of hit songs over the past decade.We did! It took us approximately 2.5 hours to complete this task.