

SAS University Edition Challenge

Day 4 of 5

Challenge Overview

You've just concluded a series of successful interviews for a data scientist role at your dream job: A tech-startup that designs bespoke music playlists for businesses and individuals, using freely available music data from Spotify.

As one final test of your suitability for the role, the company has asked you to spend one week helping them analyse the songs that featured on triple J's hottest 100 list between 2014 and 2018.

You read an article last week about SAS University Edition and believe it would be the perfect tool to get the job done. You'd be correct in that belief.

What are you waiting for? Let's get to work!

Today you are working with 3 SAS data sets: **Items_Track_Cleaned**, **Artist_Info_Deduped**, and **Master_FT_Audio_Features**. Click the link below to access the data.

<https://www.dropbox.com/sh/0l24b40n631gstz/AABFUchQkumQecM7mfVEomula?dl=0>

You need to create a new data set from the columns in these 3 tables, and then perform some analysis on it.

Summary of Skills Demonstrated

- Importing data
- Joining tables
- Summarising data

Task 1

Create a new table by adding columns to **Items_Track_Cleaned** from the other 2 tables:

1. From the **Artist_Info_Deduped** table, add the **Artist_id**, **Genres1**, and **Genres2** columns. Rename **Genres1** as **Primary_Genre**, and **Genres2** as **Secondary_Genre**. Join on the **list_id** column.
2. From the **Master_FT_Audio_Features** table, add the **Danceability**, **Energy**, **Speechiness**, **Acousticness**, **Instrumentalness**, **Liveness**, **Valence**, and **Tempo** columns. Join on the **Track_id** column.

Give the resulting table the name **H_100_ENRICHED**. Merge the columns in such a way that only matching rows for each table are included in **H_100_ENRICHED**. The **H_100_ENRICHED** table should have 484 rows and 19 columns.

Tips: Try using the *PROC SQL* method in the *Combine Tables* Data Task. You may need to edit the code, to rename the columns.

Task 2

Use the `H_100_ENRICHED` table to answer the following questions:

Question 1: Which music genre appeared most frequently as the primary genre in the number 1 spot over the past 5 years (2014 – 2018)?

Question 2: Which year's top 100 list has the highest average song liveness?

Question 3: Which primary genre of music has the lowest average song acousticness?

Question 4: Which song in the table has a high rating (> 0.75) for both instrumentality and danceability?

Question 5: On average, which feature of songs on the triple J's Hottest 100 list has increased every year from 2014 to 2018?

Tips: Try using the *Query* Utility in SAS University Edition. You will need to decide how to filter and group the data, and which summary functions to use.