

SAS University Edition Challenge

Day 1 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 out with some tasks.

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!

Day 1: **Based on a data set of the top 200 songs each day, for multiple regions, over a period of more than one year, analyse the trends in certain songs for certain regions.**

Click the link below to access the data.

<https://www.dropbox.com/sh/e9vgbslcou7ezel/AABfPkKn9BKhrHTUt2SK89ta?dl=0>

Summary of Skills Demonstrated

- Filtering data
- Summary Statistics

Task 1

You want analyse the summary statistics for the number of streams of the top song in New Zealand.

1. Filter your input data set for **region = 'nz'** and **position=1**.
2. Compute the mean of the **streams** variable for this summarized data set.

Tips:

1. Input data can be filtered with a WHERE statement or a WHERE data set option.
2. [PROC MEANS](#) can be used to produce summary statistics of numeric variables.

Question 1: *What is the **mean**, to the nearest integer, of the number of streams for the top song in New Zealand?*

Task 2

You now want to analyse data for a specific song, Mr. Brightside, to analyse the trends of this songs in within Australia.

1. Filter the songs_complete data set for **Track_Name = "Mr. Brightside"** and **region = "au"** (note the period after Mr)
2. Compute the median of the number of streams for the song Mr. Brightside over the period which we have data

Tips: The same techniques used for the previous question can be used here.

Question 2: *What is the **median** of the number of streams the song **Mr. Brightside** in **Australia**?*

Task 3

You want to further analyse the song Mr. Brightside and its number of streams in Australia. Specifically, you want to determine on what day in our data set did it have the most streams.

Tips: If you prefer to analyse data graphically, consider using [PROC SGPLOT](#) with a **series** statement. Combining this with the global statement **ods graphics / imagemap;** will allow you to hover your mouse of your chart produced and see the corresponding X and Y values for each point on the series plot.

If your prefer to use SQL for your analysis, consider using [PROC SQL](#).

Question 3: *On what day in our dataset did Mr. Brightside get the most streams in Australia and how many streams did it get? Write your answer for the date in format dd/mm/yyyy.*