Investigating Music through Data Science

A UVA Data Science Case Study by Mei Gilhousen, 2022

Prompt:

The music app developer Spotify has been struggling with business recently. They have had trouble with expanding their customer base and advertising popular music on the app. As a data analyst for the company, you have been tasked with investigating the available data from the Spotify algorithm to predict the top 5 songs from an available list of recently released songs on the app. The company will then use your predictions as a testing model to increase their advertising, and to target more customers who listen to those artists.

Investigating data:

The data given is a list of approximately 100 songs, including their name, artist, and other musical characteristics such as tempo and energy. Your task is to select 3 of the musical variables available and to create a multilinear regression model with them, with the dependent variable being the popularity ranking. Select the three explanatory variables that you think would have the greatest impact on the model. This model will then be used to predict which songs will be the most popular from the available list.

Deliverable:

Create a plan to investigate the chosen explanatory variables, and to create the final model. After creating the model, list the top 5 songs with the predicted highest ranking, as well as the multilinear regression model that can be used for further predictions. Also note any obstacles to the plan and analysis, such as if the variables did not meet the conditions to perform multilinear regression.