

Exploratory Data Analysis and Visualization - MSc AIDA UoM

Professor George Evaggelidis

Academic Year: 2021-2022

You are provided with the file `spotify_data_by_year.csv`, which contains the average values of certain characteristics for songs each year from 1921 to 2020. You can see what each characteristic means on the this webpage

Additionally, the following information is provided regarding the 'popularity' characteristic: "A number between 0 and 100, with 100 being the most popular. The popularity is calculated by an algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played a lot now will have a higher popularity than songs that were played a lot in the past."

Use univariate and multivariate analysis techniques, as well as time series analysis, and provide interesting graphs for the above data along with appropriate commentary. It is clarified that if you want to group your data by decade, the decades are defined as follows: 1921-1930, 1931-1940, ..., 2011-2020.