

## General Instructions

- The deadline for this assignment is **07-03-22** at 11:55 pm. There will be no extensions.
- Submit a zip file which contains two files:
  1. *RollNo\_code.ipynb*: The ipython notebook **with outputs**.
  2. *RollNo\_Report.pdf*: The report **should be typed** and can include figures. It should contain your answers and analysis to the questions given in this document.
- Name the zip file as *RollNo.zip*
- Not following the above instructions will lead to a 20% penalty.
- Plagiarism is strictly prohibited.

## Pre-reading

This assignment deals with analysing the music put out by an artist over 30 of their songs. You should watch [Marcus Schedl's](#) talk on lyrical features of a song as well as go through [this link](#) about understanding and calculating the Gloom Index for an artist. You should go through the latter of the two links which performs an in-depth case study of the band 'Radiohead' to give you an idea about the motivation of this assignment.

## Q1 - Song Meta-Data extraction (2 marks)

- (a) Select an artist of your choice, the only constraint is that they should have 30 singles - easy enough, right?
- (b) Retrieve song meta-data and acoustic features using Spotipy and store them in separate dataframes. Print the head of each dataframe and paste this result in your report. **(1 mark)**
- (c) Retrieve popularity of each selected song using this [endpoint](#). **(1 mark)**

## Q2 - Feature extraction from Lyrics (5 marks)

- (a) Calculate the Gloom Index for the selected songs. **(1 mark)**

- (b) Calculate the **SMOG** (Simple Measure Of Gobbledygook) index for the selected songs. **(1 mark)**
- (c) Calculate the compressibility score (value should be between 0 and 1) of each song. **(1 mark)**
- (d) Create a table in which the first column indicates rank ordering of the songs based on your perception of how gloomy they are. The second column should contain the ranks based on the computational Gloom Index (GI). Compare the computational GI estimates with your perceptual estimates. To what extent do they agree? What are the typical discrepancies? Can you explain the reason for them? **(2 marks)**

### Q3 - Inference (6 marks)

- (a) Do you observe any correlation between compressibility and Gloom Index? Are the results inline with the hypothesis that: *Songs with a higher Gloom Index will have a lower compressibility score?* (Hint: You can read about the Spearman Correlation test of association which you can use to justify your answer.) **(1 mark)**
- (b) Do you observe any correlation between compressibility and SMOG Index. Are the results inline with they hypothesis that: *Songs with a higher SMOG Index will have a lower compressibility score?* **(1 mark)**
- (c) Which of the above features (extracted in Q1 and Q2) has high correlation with popularity of a song (Use Spearman correlation). Pick the feature that correlates the most with Song Popularity and provide a scatter plot accordingly. Provide a table where rows represent extracted features and entries are the Spearman correlation values. **(2 marks)**
- (d) Perform correlation between all the extracted features (11 acoustic features + GI + SMOG + Compressibility) pair-wise. Use a heat-map to demonstrate all these correlations. Report any inferences from this (For ex: Danceability might correlate positively with Energy and Compressibility). **(2 marks)**