**Data Science with R- Project 1**

**Problem Statement**: To understand what exactly a listener prefers listening to on radio, every detail is recorded online. This recorded information is used for recommending music that the listener is likely to enjoy and to come up with a focused marketing strategy that sends out advertisements for music that a listener may wish to buy. However, this results in wasting money on scarce advertising.

Suppose that you are provided with data from a music community site, giving you details of each user. This will further help you get access on a log of every artist that the listed users have downloaded on their computer. With this data, you will also get information on the demographics of the listed users (such as age, sex, location, occupation, and interests). The objective of providing this data lies in building a system that recommends new music to the users in this listed community. From the available information, it is usually not difficult to determine the support for various individual artists (that is, the frequencies of a specific music genre/artist or song that a user is listening to) as well as the joint support for pairs (or larger groupings) of artists. Here, you need to count the number of incidences across all your network members. After this, you need to divide those frequencies with the number of members. Using the support value, you can calculate the values of confidence and lift.

In the mentioned data set data setnti,” a large chunk of information close to 300,000 records of song (or artistsr) selections is listed that is per the listening frequency given by 15,000 users. Each row of the data set contains the name of the artist that the user has been listening to. The first user is a German lady, who has listened to 16 artists. This has resulted in the first 16 rows of the data matrix.

First, what you need to accomplish is to transform the data given here into an incidence matrix, where each listener represents a row, with 0s and 1s across the columns. This indicates if he or she has played a certain artist or not. Then, the support for each of the listed 1004 artists needs to be calculated by displaying the support for all artists with support larger than 0.08.

Then, construct the association rules using the function Aprioriin the R package arulesand then look for artists (or groups of artists) who have support that is larger than 0.01 (1%). After the calculation is checked, another music collection of an artist turns out to be larger than 0.50 (50%).