Applied Data Science Capstone - Project approach and plan

1. Introduction:

On this 'Social Data Science' project the main purpose will be to answer the following question, with analysis of real data on R taken from different sources:

"How can music show the different lifestyles, based on a sentimental analysis of the most listened songs on different European countries?"

To answer this question data from the music streaming platform 'Spotify' will be taken. This one provides different attributes, which allows measuring different aspects for each song. This can be translated then into an emotional sense that triggers different behaviours on the users who listen to them. Based on this, and on the different playlists that compile the most listened songs of each country, an emotional analysis will be carried out with different methods. As last step, the results will be compared between the countries that are included on the study.

2. Business Problem:

On this task all the necessary steps to achieve the answer will be explained briefly:

- **Data obtainment and Processing:**

As row data, it was downloaded with 'SpotifyR' the name of playlists created by the user 'Spotify' (a lot, and plenty of them out of the scope for the purpose). The ones that are of big interest for this project are the ones that contain the 'Top 50' songs. As all of this lists names have the same format (*country_name* + ' Top 50') a join with a list of the European countries (obtained from the WDI) gives the names of the desired lists for the study. (*See below, the list of playlists that will be analyzed*).

1	Austria Top 50	12	Iceland Top 50
2	Belgium Top 50	13	Italy Top 50
3	Switzerland Top 50	14	Lithuania Top 50
4	Germany Top 50	15	Luxembourg Top 50
5	Denmark Top 50	16	Latvia Top 50
6	Estonia Top 50	17	Netherlands Top 50
7	Spain Top 50	18	Norway Top 50
8	Finland Top 50	19	Poland Top 50
9	France Top 50	20	Portugal Top 50
10	United Kingdom Top 50	21	Sweden Top 50
11	Ireland Top 50	22	Turkey Top 50

Illustration 1. European countries Top songs playlists

The result is now complemented with the songs of each playlist, adding as well the parameters that will be useful for later steps, such as, 'Danceability', 'Valence', 'Energy' and 'Track features'.

- **Sentimental analysis**

Once the data is ready to be analysed, a sentimental analysis will be carried out to identify the sentimental connotation of each song, this will be done by different methods:

- On one hand, a first approach will be done by using the parameters spotify provides on the songs (Danceability, Valence, Energy...).
- On the other hand, different tools on R (QDAP) or external tools ('SentiStrength', ANEW, VADER,...) for text processing will be used. This analysis will be also done in two parts.

Ion Garaicoechea Guruceta

First, focusing on the titles of the songs, and then for a more microscopic and accurate result, the same method will be applied to the whole lyric of a reduced group of relevant songs.

- **Comparison of results**

After the sentimental analysis is done and results are obtained for each of the countries, the comparison will be done between them.

As well, the results will be compared to other projects results with a similar scope but that used different methods.

3. <u>Conclusions and expected results</u>

The conclusions will be based on the comparison made in the task before.

- From the point of view of measuring and classifying the songs by parameters, this can be relevant to understand or visualize some different lifestyles across European countries populations. Defining the role the music plays into their daily lives.

- What wants to be proved with this is if the population, depending on the country they belong to, comes away from other countries which show other sentimental levels through the music they listen to. If so, maybe countries will show similar patterns of results, depending on the geographical location, language, culture or other facts that can influence this aspect.