**Project: Capstone Project 1: Inferential Statistics and Insights**

- Are there variables that are particularly significant in terms of explaining the answer to the question?

  Following are the columns in the dataset :-

- genre
- artist_name
- Track_name
- Track_id
- popularity
- Acousticness
- Danceability
- Duration_ms
- energy
- Instrumentalness
- Key
- Liveness
- loudness
- mode
- Speechiness
- Tempo
- Time_signature
- Valence

In order to find the columns more significant to our results we can make use of a correlation matrix which gives a correlation among all the variables with each other .

On doing that we found out that among non categorical variables :-

- danceability
- energy
- loudness
- tempo

seem to have a positive correlation with songs popularity which is understandable since songs like despacito and shape of you which were extremely popular had a perfect balance of these features.

Apart from this, there are also features which have negative correlation with the popularity which include :-

- acousticness
- instrumentalnnss
- liveness
- Speechiness
-

Which means songs with a whole lot of words and instruments don’t tend to get that popular which again correlates to what the music culture s going on nowadays where instruments used are minimalistic and that too programmed .

One hypothesis testing performed was including the popularity of major key(happy) vs the minor key(sad) songs . Initially it appeared that on an average roughly the popularity of the sad and happy songs was the same but on conducting the hypothesis testing ( T-Test in our case) we found out that there was indeed a difference between the two . Statistically speaking people tend to like sad songs more than happy songs on an average .

Another test was conducted to test whether there was a significant difference among popularity when it came to rock and pop songs . Using the same T -Test it was again found out that there is a statistically significant difference popularity of rock and pop songs .