Spotify Trends: How The Number Of Playlists Effects A Song's Place On The Charts

These days Spotify has become a staple of music listeners across the globe. Spotify charts are used by many people to evaluate what music is trending and to curate their own listening, but what do people really know about this chart?

I'm examining one aspect of the Spotify charts by looking into a possible positive correlation between the number of playlists a song is in and its place on the Spotify chart. I will also be comparing this relationship to the relationship of a song's place in the chart and several other aspects of a song. These include objective aspects like mode, bpm, number of artists, streams, and release year, and subjective aspects like danceability, valence, energy, and liveness.

For my methods, I'm using linear regression and Elastic Regression. I will be testing these models and comparing them. I will also be producing coefficients and creating graphs including a bubble graph and a figure containing multiple graphs. I will also clean the original dataset which includes using one hot encoding on the category "mode" and dropping unnecessary columns.

After running my models, I found no positive correlation between the number of playlists a song was in and its place on the Spotify chart. I did see a slight negative correlation but nothing linear or consistant. I also noticed an expected positive correlation between streams and placement in the chart, but it was not as correlated as expected.

A possible reason for the results I got is the endogeneity problem. Basically there is a potential problem of causality meaning it is unclear which direction causation flows. The number of playlists a song is in might influence its place on the chart, but its place on the chart might also influence the number of playlists a song is in.

The age of a song also could also play a factor specifically in why some songs had high numbers of streams and were in many playlists, but placed low on the charts. Older songs might be streamed as classics, but are not going to be tracked and charted the same way as more modern songs.

Ultimately my hypothesis was incorrect, but I gained valuable insight into the way the Spotify chart is created. If I were to repeat this project I would make sure to take into account the endogeneity problem and to factor in a song's release date.