

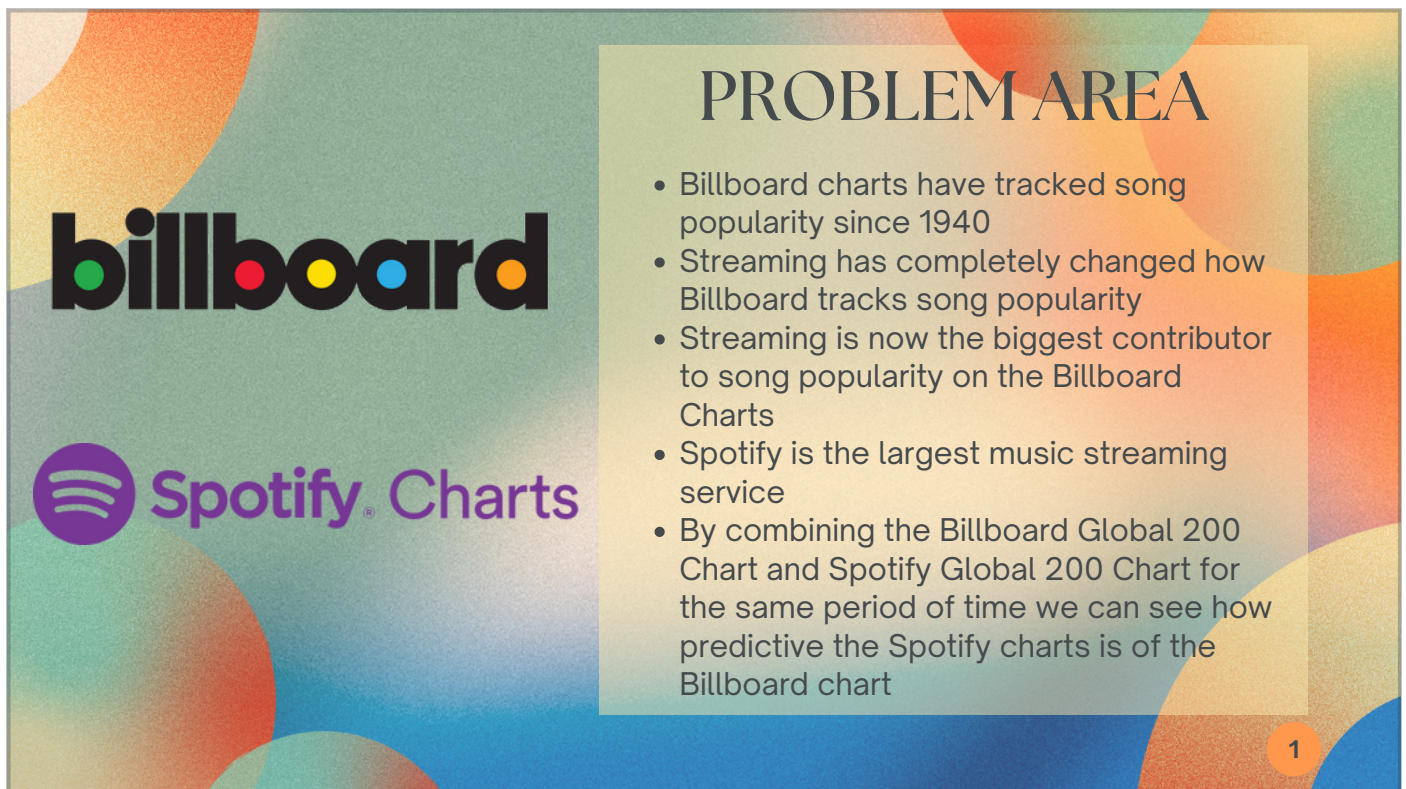


Fall 2023 MGMT 6600 Data Analytics

# PROJECT PROPOSAL

Billboard Global 200 Chart and Spotify Global 200 Chart Analysis

Presented by: Kaily Mejia



Since 1940 Billboard has tracked the most popular songs. However, since 1940 the way people have interacted and accessed music has changed drastically. Every time the way people access music has changed the Billboard charts have adjusted the way they calculate the popularity of a song.

Streaming has completely changed the way people listen to music. It has made music more accessible globally and is now the biggest income generator for the music industry. In the streaming business, Spotify has the largest market share with the largest number of streams when compared to any other streaming service.

As a music lover, I wanted to see just how affected the billboard chart is by streaming services. Billboard changed its rules in 2017 so that paid streams are worth the most in their point distribution of determining the rank of a song.

Since Spotify publishes its top 200 songs streamed globally every week just like Billboard does I thought to combine the two data sets to see how accurately the Spotify chart predicts the Billboard chart.



# DATA CONSTRUCTION

Billboard Global 200 (4/15/2023 - 10/14/23) & Spotify Global 200 (4/13/2023 - 10/12/2023)  
Merged on Rank

COMBINED GLOBAL  
200 CHART  
(4/15/2023 - 10/14/2023)

WEEKLY GLOBAL 200  
CHART  
(4/15/2023 - 10/14/2023)

BILLBOARD CHARTS

COMBINED GLOBAL  
200 CHART  
(4/13/2023 - 10/12/2023)

WEEKLY GLOBAL 200  
CHART  
(4/13/2023 - 10/12/2023)

SPOTIFY CHARTS

2

Both the data sources are publically available Spotify publishes its Global 200 weekly and Billboard also publishes its Global 200 weekly.

Since the data is published weekly it is only available by week. In order to have a dataset with a longer time frame the weeks would have to be combined.

I collected the last 27 weeks of both the Spotify and Billboard Global 200 charts. I wanted to focus on recent trends and I thought about 6 months of data would be enough to begin to identify trends in the dataset.

I knew I wanted to create a dataset where you could compare the rankings of the billboard chart and Spotify side by side.

In order to build that dataset I realized I had to combine each week of the Billboard with the corresponding Spotify chart for that week. After all 27 weeks were joined I combined them to create the desired dataset.

I also created a dataset with all the weeks of Billboard combined by itself and I did the same with the Spotify charts.

Through the individual datasets for both Billboard and Spotify and the larger merged one, I can perform data analysis on each chart individually and together.

# ANALYSIS TECHNIQUES

## CLUSTERING

### K-MEANS CLUSTERING

Could be used to measure the degree of correlation between the ranks of songs on the two charts within a group

## LINEAR REGRESSION

### MULTIPLE LINEAR REGRESSION

Could use the charts as variables to measure the strength and direction of the linear relationship between the ranks of songs on the two charts

## CORRELATION

### PEARSON CORRELATION ANALYSIS

Could use to measure the strength of the relationship between the ranks of songs on the two charts

## TIME SERIES

### CROSS-CORRELATION ANALYSIS

Could be used to measure the degree of correlation between the ranks of songs on the two charts over time

3

The following are the four analysis techniques I plan on using.

K-Means clustering can be used to see if the Spotify data could be used to predict the Billboard rank within a range. For example, if a song is ranked 7th on the Spotify chart the cluster could check if the song is ranked between 5-10 on the Billboard chart.

Multiple Linear Regression can measure the relationship between the Spotify chart rank as an independent variable and/or the streaming numbers and the Billboard chart rank as the dependent variable.

Pearson Correlation Analysis can be used to measure the strength and direction of the relationship between the two charts.

Cross-correlation analysis could measure the correlation in ranks over the time period of the dataset.

# DESIRED OUTCOMES

## PREDICTIONS

- Emerging trends in streaming and charting data
- Features contributing to songs charting including streaming numbers
- Predicting the billboard charts based on the strength of the relationship with the spotify chart

## UNCERTAINTIES

- Billboard has rules about ranking low for consecutive weeks that don't apply to spotify
- Spotify isn't a globally available app so there are regions of the world that aren't counted in spotify's streaming data

## FURTHER ANALYSIS

- Connecting social media use of songs to predict songs that are streaming
- Building a model that could model how many weeks a song is at a certain rank

4

Through the application of those analysis techniques on the data, the models made would hopefully predict any emerging trends in the relationship between streaming and music charts. In addition, it might provide some insights into what features of streaming data could provide the most accurate prediction for a song's chart ranking.

There are some uncertainties to take into consideration when looking at the model and its possible flaws. Billboard has specific rules about a song charting below a certain threshold for a specific chart that automatically removes them from charting. These rules don't apply to Spotify's charts because they simply chart based on streaming numbers on the platform. Additionally, while Spotify does have the largest market share in the music streaming business it is not available in every country, and in certain countries, other streaming services are more popular. So the quality of the model is going to be based on how effective the Spotify data is in representing global music trends.

However, this analysis could later be applied to how people are discovering and sharing new songs on social media platforms like TikTok. The use of music on many social media platforms has begun to show on charting and streaming data. In this analysis, the Spotify data will be treated as the independent variable for the charting data but in reality, streaming data and data consisting of social media use of music would probably provide a more holistic view as to what is affecting how songs chart and trends in music.





THANK YOU!

5