Spotify: Operation Popularity

# Team Members

May Chang, Christina Terranova, and Angely Lee

**Link to our presentation slides:**

<https://docs.google.com/presentation/d/10b5Qagsq26kUFR-ZKXfuXaeUyDI5_20SGkEMimnmCJI/edit?usp=sharing>

# Executive Summary of the Final Project

Our project revolves around leveraging a comprehensive Spotify dataset to provide valuable insights for artists and listeners. A tool will be created to help artists search what kind of songs are trending and which factors play a role in their popularity or success. By analyzing trends and factors contributing to the success of songs, we hope to empower artists to create meaningful music that resonates with their audiences. In addition, we hope listeners will also be able to discover a variety of songs based on their interests.

# Target Users or Analysis Consumers

This project targets artists by analyzing data to determine what factors influence a song's success, this will then allow these artists to understand what they could focus on in their music to create songs well-received by the general public and build in popularity.

# Technical Summary

Functions of Excel, such as sorting, filtering, and pivot charts were utilized for this project. Sorting and filtering aided in deriving the top three songs by number of streams as well as the #1 ranked songs on both Spotify and Apple Music charts. Pivot charts were then used to compare the top 3 songs and #1 ranked songs with factors, like valence, energy, bpm, etc to determine which factors had a greater influence on the song’s popularity.

In VBA, modeling was based on creating macros that would be able to effectively sort rows by hiding irrelevant ones. In such a big dataset, it can be difficult to visualize individual parts in a dataset and identify patterns. By using a loop, we were able to have the macro search for the name of the artist throughout the whole spreadsheet, and then have the macro hide any rows that did not contain the name. The other macros centered around sorting rows, and the “unhide” macro made sure that the user could easily see the rows that were previously hidden. We explored sites such as [Stack Overflow](http://stackoverflow.com) and the Microsoft help pages to brainstorm how to build the macros past what was learned in class. Given more time, we would have liked to implement a macro that would further filter the data when inputting a specific factor, such as a song with danceability greater than or equal to 80%, since Tableau was limited in that it could only produce songs with a factor of just 80% danceability.

For Tableau, the main modeling techniques we used revolved around the two dashboards created. Extensive preparation had to be done before putting them together. This included creating multiple worksheets that consisted of side-by-side bar graphs, tables, and highlighted charts. Tableau makes it easy to convert our data into line charts, bubble graphs, and treemaps. In addition, we were able to effortlessly add color, highlight certain points, sort, and filter our data. Certain parameters and calculated fields needed to be done to sort out the ranges to filter the top 10 similar tracks for the Music Recommendation Dashboard. This proved to be of great difficulty, but with the help of outside [resources](https://help.tableau.com/current/pro/desktop/en-us/changing-views-using-parameters.htm) and walkthroughs on [YouTube](https://www.youtube.com/watch?v=UnVpOIGry3o), we were able to overcome these obstacles. The use of maps would have been an intriguing addition to our dataset. Our dataset does not provide geographic points; however, if provided with more time to search for the data it would have been nice to explore other questions such as what tracks are streamed the most in different countries/regions, as well as which artists are the most popular in those areas.

# Data Needs and Sources

Extensive data is needed to answer the topic of our project which is to determine the factors that influence a song's success. Kaggle is a site that provides various types of datasets and after searching the dataset, [Top Spotify Songs](https://www.kaggle.com/datasets/arnavvvvv/spotify-music) is an excellent source for answering our project’s questions.

Parts of data were missing and had to be inputted manually, such as the title of some songs were incomplete, notated with ¿½, as well as rows being left blank in the in\_shazam\_charts column which were replaced by 0’s. Additionally, a row was deleted from the dataset due to a non-numerical number in the stream column. The dataset was downloaded, cleaned, and stored in Excel sheets as well as Tableau to test and analyze a plethora of factors for our project.

# Outputs

The top 3 ranking songs were found utilizing the sort feature of sorting songs by number of streams from greatest to least. The results were, “Blinding Lights” by The Weeknd with 3,703,895,074 streams, “Shape of You” by Ed Sheeran with 3,562,543,890 streams, and “Someone You Loved” by Lewis Capaldi with 2,887,241,814. A pivot chart was created to then analyze these top 3 songs according to various factors such as beats-per-minute (bpm), valence (%), danceability (%), acousticness (%), etc. Observations from the top 3 songs are that each song has a high bpm, fast tempo, and the range spans from 96-171 which is large but proves how consumers prefer faster-paced songs. Additionally, the danceability for each song is at least 50%; therefore, danceability may be an influential factor that makes a song popular. For stronger evidence, we filtered for songs that ranked #1 in Spotify and Apple Music charts. Between the two music platforms, “Daydreaming” and “Keep Driving” were ranked #1 in the charts and both these songs were by Harry Styles; therefore, a deep analysis of factors for Harry Styles may provide deeper insight into what makes a song popular. A second pivot chart was then made listing all the songs in the dataset by Harry Styles. Like the top 3 songs, the bpm for Daydreaming and Keep Driving are high at 114 and 126 respectively. Another observation is that the bpm for all of Harry Styles’ songs does not go below 95 bpm. The songs “Daydreaming” and “Keep Driving” also correlate with high valence (%) and danceability (%). According to the results, the following factors have more influence on popular songs: bpm, valence, and danceability. Artists can give a stronger focus to these elements when creating a song to propagate higher streams and listeners’ attention. Consumers may also use these factors to find other artists or songs with similar features they enjoy, such as high energy.

The key output of the VBA results is the artist search feature. This feature could be applied to many different concepts when working with a large amount of data as this project did. Target users can quickly and easily use this system without having prior knowledge of how to use VBA or Excel by pressing the buttons and following the prompts. This gives a quick overview of what is covered in the pivot tables and the Tableau dashboards. Something to note from the results that we would not have otherwise been aware of happened when sorting by ascending was that all of the songs before 1971 were Christmas songs. Sorting by Apple Music vs. Spotify charts using the macros also showed that different songs had more streams on certain streaming platforms than others, so not every artist had the same most streamed song on both platforms.

In addition to the findings from above, the Tableau dashboards make it easy for our targeted users to explore the Spotify dataset. With the Spotify Top Tracks Dashboard, users can look up certain songs or artists, find key attributes of a selected track, the key it’s in, and which music platforms the track appears in all within one window. Moreover, the Music Recommendation Dashboard has similar features to the first one but now has a field to input key attributes of a selected song to find similar tracks based on those attributes. Both dashboards have a field of all the songs in the dataset sorted in descending order to see which ones are the most streamed first.

One data visualization principle we paid attention to the most for the VBA and Tableau presentation was providing interactivity for our users to explore the data dynamically. It was also important for us to simplify the complexity of our project so that our users could use these applications with ease and be able to see the outputs straightforwardly. Since this project was based on Spotify, we also wanted to correlate with its theme colors throughout the VBA and Tableau applications. In addition, we tried to balance aesthetics and functionality by creating a visually appealing dashboard without overcrowding fields and overwhelming users with flashy visuals.

# Benefits to Target Audience

Our targeted audience includes artists in the music industry and listeners. We foresee several benefits from the successful completion of our project. From an artist’s perspective, one could use our Tableau dashboard to observe what factors may influence a song to become popular. The Spotify Top Tracks could also show useful information on music platforms which could be useful to marketing professionals. Since marketing does play a role in the popularity of a song, knowing which platforms are used the most is important, which was determined to be Spotify and Apple Music as Deezer and Shazam are not as recognizable music streaming platforms.

Additionally, listeners on Spotify can discover other artists and tracks by exploring the Music Recommendation Dashboard. Listeners can select a track within the dataset and see its key parameters which they can then use to look up and see a list of other similar tracks.

# Challenges

An unexpected challenge when deliberating on the factors that correlated between popular songs was defining the attributes/columns because some were not as straightforward, such as valence. However, once the attributes were clearly defined, the process of understanding the narrowed-down factors was much easier.

For building the VBA macros, first, we had to clean the data contained in the original Excel file since there were several errors such as symbols showing up in song names. After the data was cleaned, much research was needed to see how to build a macro that would run a search based on a name inputted into the search. We knew that we needed an Input Box to make it work, and had to work backward to think of different ways this could be achieved. After several different iterations of trial and error with different methods, such as AutoFilter, we were able to put the final product together. We realized when working with AutoFilters that while it was a good way to sort through the data with the Input Box command, there was no easy way to undo the filter and this caused Excel to have several problems. Once we got the final version of the macro functioning correctly, we reverse-engineered to create an “unhide” button that would show the rows again even after the user had entered the artist prompt. Another challenge was choosing which macros to make as far as filtering the data to best fit with the other parts of the project in the pivot tables and Tableau.

Moreover, other obstacles were met while building the dashboards on Tableau, including some programming from some of the worksheets which were resolved by more research online as well as trial and error. This included creating parameters and calculated fields to create ranges for finding similar tracks on the Music Recommendation Dashboard. Another opposition we faced was determining the best way to present the dashboards and some of the worksheets to align with the goal of the project. This matter was easily resolved through further research on other Tableau dashboard models and more experimentation. A third challenge we faced was ensuring the dataset was clean with no errors because data cannot be edited within Tableau, unlike with VBA on Excel. The database was meticulously combed-through with manual inputs and deletions where the final clean data was then uploaded to our shared drive for functional use.

# Personal Learning

Through this project we learned of the strengths and weaknesses of programs where the same information can be derived in both; however, the process differs. For example, pivot tables may be created to see information for a specific artist with specified factors which can be achieved as well on VBA by creating a button. Pivot tables are simpler to use and factors can be easily added or subtracted whereas VBA takes knowledge of the program to adequately create the button. Once the button is written, VBA can be much more efficient than pivot tables as results can be delivered in a click of a button.

Through this project, we continued by learning and application of VBA as a beginner. We feel we were able to understand VBA better after working on this project since we applied what we learned earlier in the semester and problem-solving skills. This project allowed great practice in programming and helped us get back into the mindset that you need to have when writing and editing code. As frustrating as it can be, perseverance will be rewarded with the macro performing the intended task. The project also will help prepare us for the future in methods we would like to store and analyze data, whether it is for work or personal use, with concepts like managing finances. We enjoyed applying what we learned, such as how Input Boxes can be used to search for terms that can change instead of a static concept.

With Tableau, we found benefits in exploring the many functions available on the application. Its strong data visualization tools aided in creating worksheets and dashboards to look at data from a business perspective rather than the technicalities for our targeted users. Tableau is a powerful tool because of how user-friendly the interface is and there are a lot of outside sources and videos we can look at for help. There were many new concepts we gained a deeper understanding in, one was creating parameters and utilizing calculated fields to facilitate our dashboards. Another was about “dashboard actions” which help with controlling what fields are affected by dynamic ranges. Lastly, we learned about the various styles to present our data and choose which was the most effective in answering the proposed business questions.

# Closing Thoughts

This project analyzed what factors make a song popular; therefore, equipping current and aspiring artists with knowledge on the foundation of building a good song while customizing it to their musical style. The foundation of creating music may be the same; however, the execution will depend on the artist and allow for the creation of an array of diverse music. Not only did our applications cater to artists, but they also targeted listeners who are interested in discovering more than what and who’s on their current playlist.

Something to make note of is that our dataset, and therefore research findings, are based solely on vocal pieces. As mentioned before, music is eclectic, and vocal pieces are only a slice of the whole where there are a plethora of other categories, instrumental music, jazz, and musical theater being a few examples, that could be analyzed and derive completely diverse results. Our recommendations are for artists of the vocal lyrics paired with the instruments category and should not be taken to other categories.