Music is an expression of art widely appreciated by millions of people all over the world. Every individual has a unique playlist and set of preferred genres catered to their own interest. Being that there are over 100 million recorded songs on the internet and musical streaming applications, it can be intimidating for someone to try and find new songs on their own. How could an individual expand their musical collection and add to their current libraries in a simple and non-overwhelming way?

Our group has unanimously decided on designing a predictive model that will give music listeners recommended playlist and songs based on their recent plays. Online music streaming services have been a dominating medium for people to use.  We want to take advantage of the information provided to us from user accounts to give listeners the best musical experience. We do this by providing a consistent, accurate, and satisfying playlist of songs that the user never listened to in hopes that they will enjoy the recommended selection.

We want to implement two different approaches, one being collaborative filtering and the other being content-based filtering.

Collaborative filtering uses algorithms that take in  large amounts of information on users’ behaviors, activities, or preferences and predicts what users will like based on their similarity to other users. The data being collected can be implicit (keeping records on how often and for how long a user views an item) or explicit (asking a user to rate an item). Results depend on the genre of music the user is listening to, to which this model will suggest genres related to that particular song.

Content-based filtering filters content based on an analysis of both the item being recommended and the user. Content-based filtering closely examines the actual item to determine which features are most important in making recommendations and how those features interact with the user’s preferences. Data collection for this process is considered more difficult since it is hard to select which features of an item will be important in creating some sort of predictive model. Machine learning techniques such as **Naive Bayesian classifiers** and **cluster analysis** are used to determine which features of an item can be used to classify it. Results from this stem from a collection of what the user already listened to. A method called “Pearson Correlation” is used to calculate the relationship between the collection of music listened to and potential suggestions.

Ideally it would be smart to incorporate unsupervised learning techniques to analyze the music played by the user and categorize similar songs into a playlist. We can then use supervised learning to determine whether it would accurately cater to the user’s interest. We have attached a link to a page of data sets from other machine learning projects related to the issue which we found useful in our research.

<http://millionsongdataset.com/>

<https://static.turi.com/datasets/millionsong/10000.txt>

<https://static.turi.com/datasets/millionsong/song_data.csv>

With this project, we hope to expose people to new and exciting musical selections and to deepen our understanding of machine learning principles.