Hello everyone, my name is Shan Jiang, and I will give you guys a simple introduction to my group project. Our project is about Spotify Music Analysis.

We start the project with a thought: we want to find general applications of the data which has possibilities that not only be useful to people, but also be interesting from a scientific perspective. So after discussion we choose Spotify Music Analysis.

The reason we choose music to analysis is that we know that music can bring beauty and express human emotions, but more important fact is that music is a symbol, a kind of sound symbol that expresses people think. Which means if you want to know somebody, you can learn it from the music he or she make or listen.

And why we choose Spotify music dataset to analyses? There are three reasons. The first one is that it is available in most parts of world. The second one is that it is available for most modern devices, including Windows, macOS, and Linux computers and so on. And the last one is that the music on Spotify can be searched for by artist, album, playlist, or record label. Users can create, edit, and share playlists and tracks on social media, and make playlists with other users. Nowadays Spotify had 191 million monthly active users, and it provides access to more than 40 million songs. Since it has lots of data and users, we think the analysis may be very valuable to lots of people.

In our project, we first will do some basic analysis of song metadata taken from Spotify. The dataset is from Kaggle, and the dataset contains 2017 songs with attributes from Spotify's API. There are 16 columns. 11 of which are song attributes, one column for song name, one for artist, and the column called "target" which is the label for the song. Each song labeled "1" means we like it and "0" for songs we don't like.

Since the music has many different features, we want to know that can we analyses these features to get that if there is a new song which someone likes it or not. We sought out to answer the following questions:

1. Are there specific relationships between those attributes?
2. Which methodology is the most accuracy one among those methodology?
3. Can we give someone recommendation of songs which he or she may like based on these patterns?