This document is written to describe the data wrangling steps I took to clean the dataset for my first capstone project. The end goal of this data wrangling step would be to create a pandas dataframe that includes basic information and features of the tracks of a given playlist. The various steps included are below:

1. **Familiarization with Spotify API:** The first step consisted of understanding how to utilize the Spotify API to get the track information and features among other commands. Luckily, the Spotify API and documentation is comprehensive. In order to get information from Spotify, be it playlist data or track data, one would have to submit a request from a url issues commands from Spotify. For example to get track info, you would use the url, https://api.spotify.com/v1/tracks/{track_id}. On top of the url, an authorization token is also needed. This can be requested from Spotify. However, each token expires, so a new request would be needed every time the token expires. Commands that extract information are in json format.
2. **Extraction of track ids from playlist:** After familiarization with the API, a list of the track ids of a given playlist needs to be extracted. However, there was a max limit of 100 tracks issued for each command. Which means, if a playlist has more than 100 tracks, you would need to offset the command by a 100 or whatever set of hundred to get the rest of the tracks.
3. **Get track information:** Using the track ids, a command was issued to get the a specific set of (i.e. name, artists, album, etc.) information of each of the track ids. Once all the track info is combined into a list, that list is converted to a dataframe.
4. **Get track features:** Like the track info, a command was issued to get the features of each of the track ids. Also like the track info, this information was combined and then converted into a dataframe.
5. **Merge to form final dataframe:** Finally the end goal is reached by merging the two dataframes from above by their 'id'.

Fortunately, it seems to be that the information from Spotify is pretty complete so there were no need to deal with missing values. As for outliers, this would have to depend on the given playlist. However, each playlist is unique and each feature value may be important and cannot be excluded.

Using the steps above, I was able to successfully get information from the Spotify catalog using its API and converted it into a dataframe that would be the basis of further exploration and analysis.