**File Name: spotify.xls**

**Data Source: Adopted from Kaggle**

**Data Set Information:**

This data set contains songs for the past two years from Spotify. The features and the associated description are described below:

**Attribute Information:**

|  |  |  |
| --- | --- | --- |
|  | acousticness | (Ranges from 0 to 1). A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic. |
|  | artists | Track Artists |
|  | danceability | (Ranges from 0 to 1). Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable. |
|  | duration\_ms | The duration of the track in milliseconds. (Integer typically ranging from 200k to 300k) |
|  | energy | Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. |
|  | explicit | The binary value whether the track contains explicit content or not |
|  | id | The Spotify ID for the track. |
|  | instrumentalness | Predicts whether a track contains no vocals. Ranges from 0.0 to 1.0. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. |
|  | key | The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation . E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1. |
|  | liveness | Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live. (Ranges from 0 to 1) |
|  | loudness | The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Values typical range between -60 and 0 db. |
|  | mode | Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0. |
|  | name | Track name |
|  | popularity | The popularity of the song lately, default country = US |
|  | release\_date | The date of release of the track in yyyy-mm-dd, yyyy-mm, or even yyyy format |
|  | speechiness | Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks. (Ranges from 0 to 1) |
|  | tempo | The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration. |
|  | valence | A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry). (Ranges from 0 to 1) |
|  | year | Year of release |

A start-up that plans to develop a product that generates music is interested in using the dataset. See (e.g. www. HookGen.com). Use clustering to profile the songs. Based on your results -what kind of features should the new product contain?

1. Use k-means (k=2 and k=3) to profile the songs. Include only numerical variables that characterise the songs (with the exception of year).
2. Check any artists’ popularity with the generated clusters.
3. What questions or concerns do you have about these clusters?
4. As a product development manager for the start-up- what might you do next if handed this report?