

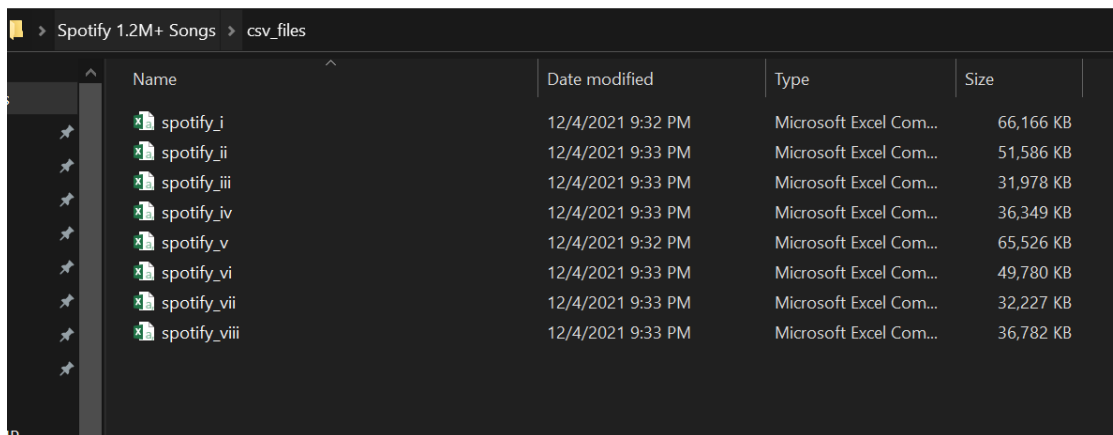
**Task: Download tracks\_features.csv. It contains audio features for over 1.2 million songs. This dataset is taken from Kaggle website. Reference for these audio features can be found in the Spotify website. Since this is a dataset with about 1.2 million of rows and 24 columns. You are required to write an R program to split the CSV file into eight CSV files. After tracks\_features.csv is split into eight CSV files, you need to import each of the CSV file into R and combine them as a data frame named complete.**

For this question, a dataset CSV file was split into eight files and then combined into a data frame. The tasks were broken down to answer this question by splitting the dataset into columns and uploading each one in its own data frame.

```
>
> # Setting a working directory and loading the dataset to put it into the dataframe
> setwd("/Users/Mohammed Tayfour/Desktop/Spotify 1.2M+ Songs")
> track_data <- read.csv("tracks_features.csv")
>
> #Splitting the entire dataset by columns first
> spotify <- track_data[,c(1,2, 3, 4)]
> spotify2 <- track_data[,c(5, 6, 7, 8)]
> spotify3 <- track_data[,c(9, 10, 11, 12, 13, 14, 15, 16)]
> spotify4 <- track_data[,c(17, 18, 19, 20, 21, 22, 23, 24)]
>
```

After that we split the “mini datasets” by rows as shown in the example below

```
> #Now lets wanna split the mini-datasets by rows to upload in csv files
> #spotify_i and spotify_v
>
> spotify1 <- spotify[1:602012,]
> write.csv(spotify1,"C:\\Users\\Mohammed Tayfour\\Desktop\\Spotify 1.2M+ Songs\\csv_files\\spotify_i.csv")
>
> spotify5 <- spotify[602013:1204025,]
> write.csv(spotify5,"C:\\Users\\Mohammed Tayfour\\Desktop\\Spotify 1.2M+ Songs\\csv_files\\spotify_v.csv")
>
```



| Name         | Date modified     | Type                   | Size      |
|--------------|-------------------|------------------------|-----------|
| spotify_i    | 12/4/2021 9:32 PM | Microsoft Excel Com... | 66,166 KB |
| spotify_ii   | 12/4/2021 9:33 PM | Microsoft Excel Com... | 51,586 KB |
| spotify_iii  | 12/4/2021 9:33 PM | Microsoft Excel Com... | 31,978 KB |
| spotify_iv   | 12/4/2021 9:33 PM | Microsoft Excel Com... | 36,349 KB |
| spotify_v    | 12/4/2021 9:32 PM | Microsoft Excel Com... | 65,526 KB |
| spotify_vi   | 12/4/2021 9:33 PM | Microsoft Excel Com... | 49,780 KB |
| spotify_vii  | 12/4/2021 9:33 PM | Microsoft Excel Com... | 32,227 KB |
| spotify_viii | 12/4/2021 9:33 PM | Microsoft Excel Com... | 36,782 KB |

After splitting the CSV files, they were uploaded into R to be combined into a complete data frame, as stated in the question. To solve this, each CSV file was placed into a data frame and then, using the merge function, each one was combined by rows first, then by columns, and finally by data frame, as shown below.

```
> i <- read.csv("spotify_i.csv")
> ii <- read.csv("spotify_ii.csv")
> iii <- read.csv("spotify_iii.csv")
> iv <- read.csv("spotify_iv.csv")
> v <- read.csv("spotify_v.csv")
> vi <- read.csv("spotify_vi.csv")
> vii <- read.csv("spotify_vii.csv")
> viii <- read.csv("spotify_viii.csv")
>
> #merging all imported csv into one dataframe called complete
>
> #we start by merging spotify_iv and spotify_viii
> d1 <- merge(x = iv, y = viii, by = c("X", "instrumentalness", "liveness", "valence", "tempo", "duration_ms", "time_signature", "year", "release_date"), all = TRUE)
>
> #spotify_vii and spotify_iii
>
> d2 <- merge(x = iii, y = vii, by = c("X", "explicit", "danceability", "energy", "key", "loudness", "mode", "speechiness", "acousticness"), all = TRUE)
>
> #merging the first half (spotify_iii, spotify_iv, spotify_vii, spotify_viii)
>
> half_1 <- merge(x = d1, y = d2, by = c("X"), all = TRUE)
>
>
> #lets merge both to make the second half
>
> half_2 <- merge(x = d4, y = d3, by= c("X"), all = TRUE)
>
> #now that we have both halves, we can finally create the dataframe "complete"
>
> complete <- merge(x = half_2, y = half_1, by = c("X"), all = TRUE)
> complete$X <-NULL
```