Data Analytics

Professor Ernesto Lee

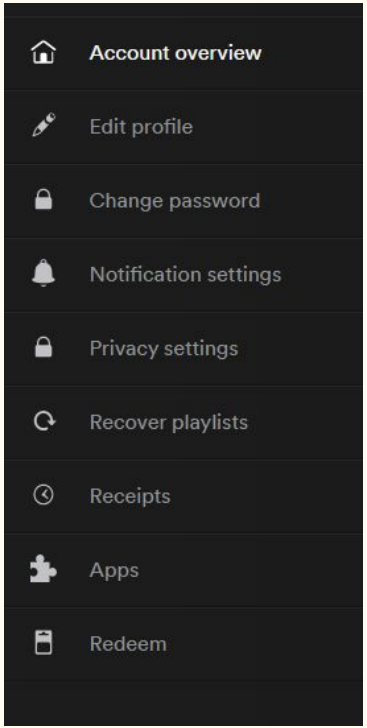
Data Analytics with Music (Preamble to Tableau)

What we will cover

1. Introduction to Spotify
2. Downloading the Data from Spotify Dashboard
3. Extracting Relevant Data
4. Conclusion

Downloading the Data from Spotify Dashboard

- Now to download your Spotify data the first and foremost step is to login into your Spotify account.
  - You can do this by visiting the official Spotify's website.
  - https://www.spotify.com/us/home/
- Once logged in, go to the top right corner of the screen.
- You will see your Profile section.
- Select your Account from the dropdown menu.
- Go to Privacy Settings and Download your data



Download your data

Most of the personal data that Spotify has about you is accessible through the Spotify app (e.g. playlists, search queries, followers, and streaming history). If you would like to get a consolidated copy of this data, you can download it by following the steps below.

The download will include a copy of your playlists, searches, streaming history for the past year, a list of items saved in your library, the number of followers you have, the number and names of the other users and artists you follow, and your payment and subscription data. For more information, see your data rights and privacy settings.

As the downloadable file you will receive will contain your profile information, you should keep it secure and be careful when storing, sending, or uploading it to any other services.

If you have any questions or concerns about the personal data contained in your downloadable file, please contact us.

STEP 1

Click the button to start the process of collecting your data.

Request

STEP 2

We are preparing your data file. This can take up to 30 days to complete.

You'll receive an email when it is ready to download.

STEP 3

We've sent you an email with a link to download your data file. If you can't find the email, check your junk/spam folders, or click below to resend.

Resend email

Extract Relevant Data

Follow.json — contains the following list as well as the current followers.

Identity.json — contains your information that is shown on the Spotify app such as name, photo, and verification etc.

Inferences.json — contains Spotify's understanding of you as a user i.e., what kind of content you consume on Spotify like education, business, dance, and so on.

Payments.json — contains your payment information.

Playlist1.json — This JSON has the playlists information that is created by you.

SearchQueries.json — It stores your search history for example: at what time and on which system you have queried for a song, artist or podcast.

StreamingHistory0.json — This has your streaming history i.e., when and which song you have heard and for how long.

UserData.json — This contains the personal information that you provide at the time of sign up, for instance, your username, date of birth, email, gender, etc.

YourLibrary.json —The content that you have saved or liked can be found in this JSON file.

Part 2: Data Cleaning

What we will cover

1. Importing Data
2. Dropping duplicate columns
3. Dropping duplicate rows
4. Null check
5. Data Format check
6. Value check
7. Exporting Data
8. Conclusion

Importing Data

Download and load:

final.csv (history of all songs played)

playlist.csv (your playlist)

- Save them as an excel workbook with two spreadsheets
- At the end of the session - load them into Tableau

Drop Duplicated Columns

Final: 'Unnamed: 0', 'name', 'endTime', 'type', 'uri', 'track_href', 'analysis_url', 'duration_ms'

Playlist: 'Unnamed: 0', 'id', 'spotify_id', 'list_id'

Deduplicate Rows

Find the excel function that:

- Counts the number of rows
- Counts the number of duplicate rows
- Removes duplicate rows

Null Check

Find the excel function that:

- Counts the number of empty rows
- Removes empty rows

Check all of the data types

- Are dates the date data type
- Are all numbers numeric data types
- Booleans? etc.

Value Check

- Find the max and min of each numeric data type
- Find the number of unique values for each categorical value
- Run the correlations on the numeric data columns

Part 3: Exploratory Data Analysis

Find INSIGHTS in your dataset

What we will cover

1. Music Attributes
2. My streaming History
3. How often did I listen to music?
4. Whom did I listen to most?
5. My mood throughout the year
6. Feature Analysis
7. Mean of audio features
8. Histogram of Features
9. Histogram for Tempo
10. Heatmap
11. Scatterplot
12. Pivot Table
13. Playlist Analysis
14. Songs by Year
15. Songs by Key
16. Conclusion

Music Attributes

- **Tempo**: The tempo of the song. The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, the tempo is the speed or pace of a given piece and derives directly from the average beat duration.
- **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. Higher the value more energetic the song.
- **Danceability**: 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. The value ranges from 0 to 1. Higher the value more suitable the song is for dancing.
- **Loudness**: Loudness values are averaged across the entire track. It is the quality of a song. It ranges from -60 to 0 DB. Higher the value, the louder the song.
- **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).
- **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 a strong likelihood that the track is live.
- **Acousticness**: A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
- **Speechiness**: Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audiobook, 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.
- **Mode**: Songs can be classified as major and minor. 1.0 represents major mode and 0 represents minor.
- **Key**: Key is the pitch, notes or scale of song that forms the basis of a song. 12 keys are ranging from 0 to 11.

Streaming History

Open song_data.csv

```
!pip install patool
import patoolib
patoolib.extract_archive("/content/train.csv.zip",outdir='/content/')
```

```
Collecting patool
  Downloading https://files.pythonhosted.org/packages/43/94/52243ddff508780dd2d8
     |████████████████████████████████| 81kB 2.3MB/s
Installing collected packages: patool
Successfully installed patool-1.12
patool: Extracting /content/train.csv.zip ...
patool: running /usr/bin/7z x -o/content/ -- /content/train.csv.zip
patool: ... /content/train.csv.zip extracted to `/content/'.
'/content/'
```

EDA

How often did I listen to music? (Convert ms to seconds)

EDA

Who did I listen to the most?

EDA

What was your mood throughout the year? Hint: check valence

Feature Analysis

Load distinct_song.csv

Calculate the mean value of the audio features

Histogram

Create a histogram of:

Danceability

Energy

Key

Loudness

*also create a count of songs by tempo

*create a heatmap (correlation of the songs)

Bivariate Analysis

Scatterplot:

energy/loudness

danceability/valence

energy/valence

valence/loudness

liveness/loudness

tempo/danceability

Extra Analysis (playlist_data.csv)

Songs by key

Songs by year

Conclusion

Did you learn anything new?