**Spotipy 1.0 – Documentation**

by David Moody

Diagram

Description automatically generated

**Goal**

Create a song recommendation system based on an analysis of my playlist listening habits. This recommendation system will look at the playlist metrics and compare the subjectivity of a song to songs of similar metrics and make a prediction based on the base condition.

**Metrics**

***Danceability*** quantifies how suitable a track is for dancing based on a combination of musical elements, like tempo, rhythm, and beat. Songs with higher danceability have stronger and more regular beats. Like acousticness, danceability is measured on a scale of 0 (low danceability) to 100 (high danceability).

***Liveness*** detects the presence of an audience in a song. Liveness is measured on a scale of 0 (no audience) to 100 (audible audience). Songs with higher liveness are more likely to have been performed live.

***Loudness*** measures the decibel level of a song. Decibels are relative to a reference value, so songs with lower loudness values are quieter relative to the reference value of 0.

***Speechiness*** measures the presence of spoken words in a song. Speechiness is measured on a scale of 0 (low speechiness) to 100 (high speechiness). Songs with higher speechiness are mostly composed of spoken words, like poetry or a talk show.

***Tempo*** measures the beats per minute (bpm) of a song. Many popular songs range from 50 bpm to 200 bpm. Songs with higher tempo have a faster pace.

***Valence*** measures the positivity of a song. Valence is measured on a scale from 0 (low valence) to 100 (high valence). Songs with higher valence sound happier and more cheerful.

***Instrumentalness*** predicts whether a track contains vocals. Instrumentalness is measured on a scale of 0 (likely contains vocal content) to 100 (likely contains no vocal content). Songs with higher instrumentalness are less likely to have vocals

**Step 1:** Begin with the end in mind

**Questions:**

What is the best song for running?

What is the best song for energetic dancing?

References:

*Happy birthday*: **125bpm**

*Blurred Lines* – Robin Thicke: **120bpm**

Crazy – Gnarls Barkley : **100bpm**

*Hotel California –* Eagles:**80bpm**

What percentage of the songs have a valence above 50?

Is there a correlation between the valence and tempo of a song?

What artists show up more than once on the top 50 and in what years?

What artists show up with more than one song on the top 50?

What artist has the most songs in the playlist?

Percentage of songs with features

What words show up in song titles the most?

How what is the avg popularity of an artist’s songs?

What top 10 songs have the highest temp and danceability?

**Step 2:** Data pull *– extract.py*

1. obtain authorization token
2. make request to Spotify API
3. select playlist
4. pull raw data

**Step3:** Data cleaning *– clean.py*

1. prep data columns to extract by creating a dictionary

*{*

*“artist”: ””*

*“album”: ””,*

*“track\_name”: ””,*

*“track\_id”: ””,*

*“popularity”: ””,*

*“danceability”: ””,*

*“energy”: ””*

*“key”: ””,*

*“loudness”: ””,*

*“mode”: “”,*

*“speechiness”: ””,*

*“instrumentalness”: “”,*

*“liveness”: “”,*

*“valence”: “”,*

*“tempo”: “”,*

*“duration\_ms”: “”,*

*“time\_signature”: “”*

*}*

1. extract each value for every song and populate the dictionary
2. convert dictionary to a DataFrame
3. Clean Spotify top 50 (2018 - 2021) datasets

**Step 4:** Data transformation – *transform.py*

1. remove duplicates
2. combine Spotify top 50 data sets

**Step 5:** Load data – *load.py*

data = create\_df()

transform(clean(data))

**Step 6:** Data Analysis – *analyze.py*

1. perform manipulations and slicing based on initial inquiries
2. create views of the data that answer questions

**Step 7:** Data Visualization – *analyze.py*

1. create data visualizations with clean data
2. using matplotlib and seaborn

**Stats**

My playlist stats

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | danceability | energy | loudness | speechiness | instrumentalness | liveness | valence | tempo | duration\_m |
| max | 0.962 | 0.956 | -0.945 | 0.745 | 0.973 | 0.973 | 0.979 | 205.846 | 10.603 |
| min | 0.144 | 0.012 | -30.016 | 0.022 | 0 | 0.024 | 0.034 | 39.497 | 0.566 |
| mean | 0.620 | 0.588 | -7.555 | 0.108 | 0.083 | 0.183 | 0.464 | 116.46 | 3.797 |
| median | 0.635 | 0.59 | -7.167 | 0.0546 | 0.000 | 0.123 | 0.455 | 111.932 | 3.697 |
| std | 0.141 | 0.182 | 2.807 | 0.112 | 0.213 | 0.144 | 0.224 | 30.203 | 0.937 |

Ex. My playlist song popularity**Chart, pie chart

Description automatically generated**

**Insights**

Number of songs: **1569**

**42.766%** of songs are more positive than not (valence)

**81.070%** of songs are more danceable than not (danceability)

Average Popularity among top 5 artists with the most songs:

Kanye West (36): **48.944**

Drake (28): **47.75**

Tom Misch (26): **46.307**

Mac Miller (24): **60.541**

Kendrick Lamar (20): **51.2**

Top 5 most popular songs:

Yellow - Coldplay (**88**)

Starboy - The Weeknd (**88**)

lovely (with Khalid) - Billie Eilish (**87**)

Watermelon Sugar - Harry Styles (**89**)

THATS WHAT I WANT - Lil Nas X (**88**)

**Recommendations**

**Looking ahead**

**Spotipy 2.0**

1. I would pull the genre for each artist and acousticness
2. Delve deeper into the correlation between metrics
3. Create better visualization with a nicer GUI