



Search



Home



Library



Our Music Playlists

Analyzing Spotify:

# Project 4: Music Recommendations through Machine Learning!

Danny Flanagan, Albert Fuentes, Laura  
Gonzalez and Sarah Stegall-Rodriguez





# Table of Contents



01

Intro to  
Spotify & the  
Power of Data



3:15min

02

Challenges &  
Machine  
Learning



3:15min

03

Data  
Collection &  
Processing



3:15min

04

Filtering,  
Evaluation, &  
User  
Experience



3:15min



# About Spotify: **Timeline**



Founded in Sweden by Daniel Ek and Martin Lorentzon

## 2008 Launch

Total time: 10h34m

Genre: Pop  
Artist: Jane Doe  
Song: Mercury

## 2011 US Expansion

Total time: 12h45m

Genre: Reggaeton  
Artist: Sam Smith  
Song: Venus

## 2018 IPO: NYSE listing

Total time: 11h3m

Genre: Pop  
Artist: Nick Jones  
Song: Neptune

## 19-20: Diversify: content & podcasts

Total time: 10h5m

Genre: Rock  
Artist: Sam Smith  
Song: Mars

## 2020: Data-Driven Content

Total time: 10h59m

Genre: Latin  
Artist: Sam Smith  
Song: Mercury

## Present: High Profile Podcasts

Total time: 15h45m

Genre: Pop  
Artist: Sam Smith  
Song: Venus

## HOW SPOTIFY'S ALGORITHM WORKS



**WSJ | TECH BEHIND**



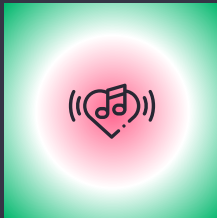


# Power of Data



## Importance of data in music industry:

- User preferences, content curation
- Artist insights, royalties and licensing
- Trend analysis and market research



## How Spotify collects data:

- Listening history and playlists
- User preferences: liking, disliking, skipping, and saving
- Location data: regional music
- App Usage
- Demographic info



**Marilyn  
Manson**

You listened 4h35min

Music is the  
strongest form of  
magic.

## Challenges in music recommendations

- Data quality & privacy concerns
- Scalability in music libraries
- Cold Start: New Users
- Diversity, Serendipity, & Precision: Taste Changes
- Contextual Recommendations: Mood
- Filter Bubbles & Feedback Loop
- Content Discovery Beyond Music



# Machine Learning & Data Collection



27.34% US

Follow

## Collaborative Filtering

Matrix factorization & deep learning models

4.65% Brazil

Follow

## Content-Based Filtering

Machine learning: analyzing song features, e.g. tempo, genre, rhythm, instrumentation

3.77% India

Follow

## Natural Language Processing (NLP)



# Data Collection: **Playlist**

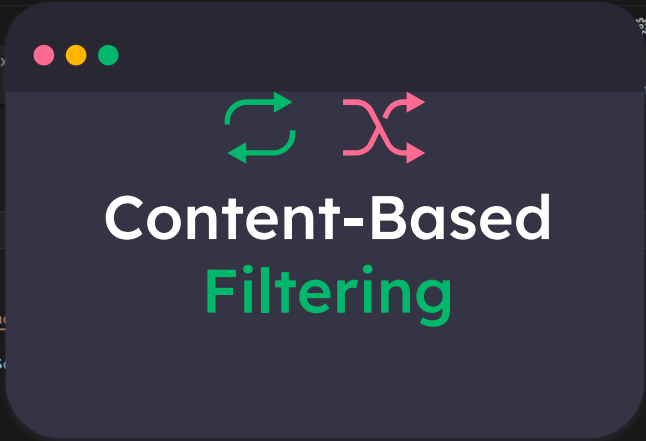


Order by ▾



#		Title	Album	Date	Duration
1		User Interaction Data	Album A	Jul 3, 2020	5:00
2		Playlist Data	Album B	Aug 4, 2019	3:45
3		Audio Features Extraction	Album C	Sep 7, 2019	2:50
4		Demographic & Location	Album D	Jul 2, 2019	4:00
5		Privacy Measures	Album E	Oct 1, 2020	3:50
6		Streaming Devices	Album F	Jan 2, 2019	4:00
7		Other App Data	Album G	Feb 9, 2020	3:01





File Edit Selection View Go Run Terminal Help

music-recommendations-machine-learning

EXPLORER

- MUSIC...
- Cleaning-Exploration
- Data
  - cleaned\_data.csv
  - data\_cluster\_pipelin...
  - data\_pd.csv
- .gitignore
- music\_recommendati...
- README.md
- Song\_Recommendatio...

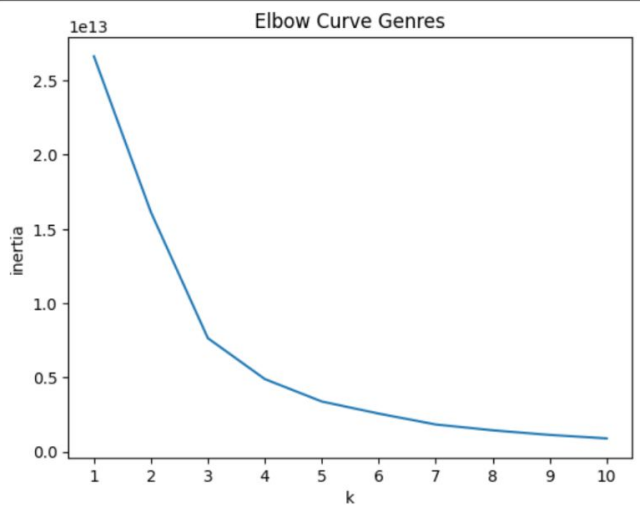
music\_recommendations (1).ipynb

C:\Users> laura > Downloads > music\_recommendations (1).ipynb > # Find the latest version of spark 3.x from <http://www.apache.org/dist/spark/> and enter as the spark version

+ Code + Markdown | Run All | Restart | Clear All Outputs | Variables | Outline

```
# Assuming genres_df_elbow is your DataFrame plot the Elbow Curve
sb.lineplot(x="k", y="inertia", data=genres_df_elbow)
plt.title("Elbow Curve Genres")
plt.xticks(k)
plt.show()
```

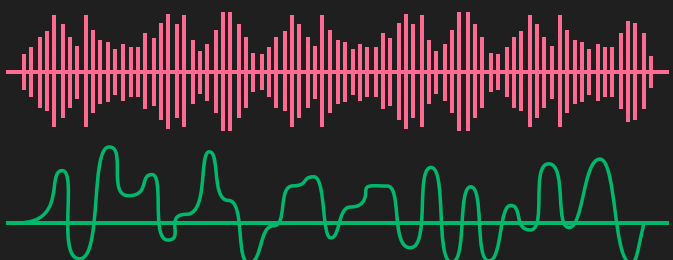
Elbow Curve Genres



k	inertia (approx. x 1e13)
1	2.6
2	1.8
3	0.8
4	0.5
5	0.35
6	0.28
7	0.22
8	0.18
9	0.15
10	0.12

```
# Create clusters for genres_pd using Pipeline
cluster_pipeline = Pipeline([('scaler', StandardScaler()), ('kmeans', KMeans(n_clusters=7))])
cluster_pipeline.fit(genres_pd_X)
genres_pd['cluster'] = cluster_pipeline.predict(genres_pd_X)
```

# Content Machine Learning



Cell 2 of 30 Go Live

File Edit Selection View Go Run Terminal Help

music\_recommendations (1).ipynb Song\_Recommendation.ipynb M X

import requests

+ Code + Markdown ▶ Run All ↺ Restart ⌵ Clear All Outputs 🔍 Go To ⌵ Variables ⌵ Outline

```
def recommend_songs( song_list, spotify_data, n_songs=10):

    metadata_cols = ['name', 'year', 'artists']
    song_dict = flatten_dict_list(song_list)

    song_center = get_mean_vector(song_list, spotify_data)
    scaler = data_cluster_pipeline.steps[0][1]
    scaled_data = scaler.transform(spotify_data[number_cols])
    scaled_song_center = scaler.transform(song_center.reshape(1, -1))
    distances = cdist(scaled_song_center, scaled_data, 'cosine')
    index = list(np.argsort(distances)[:,: ,n_songs][0])

    rec_songs = spotify_data.iloc[index]
    rec_songs = rec_songs[~rec_songs['name'].isin(song_dict['name'])]
    return rec_songs[metadata_cols].to_dict(orient='records')
```

[ ]

```
recommend_songs([{'name': 'Meltdown', 'year': 2023}], data_pd)
```

[ ]

Python

```
... [{"name": 'ROXANNE', 'year': 2019, 'artists': 'Arizona Zervas'},
    {'name': 'Hate The Way (feat. blackbear)',
     'year': 2020,
     'artists': "G-Eazy", 'blackbear'},
    {'name': 'All The Way Up (Remix)',
     'year': 2016,
     'artists': "Fat Joe", 'Remy Ma', 'JAY-Z', 'French Montana', 'InfaRed'},
    {'name': 'Show & Tell', 'year': 2019, 'artists': 'Melanie Martinez'},
    {'name': 'Lost In The World',
     'year': 2010,
     'artists': "Kanye West", 'Bon Iver'},
    {'name': 'What That Speed Bout!?',
     'year': 2020,
     'artists': "Mike Will Made-It", 'Nicki Minaj', 'YoungBoy Never Broke Again'},
    {'name': 'Da Rockwilder', 'year': 1999, 'artists': "Method Man", 'Redman'},
    {'name': 'Emotionally Scattered', 'year': 2020, 'artists': 'Lil Baby'},
    {'name': 'Lucid Dreams', 'year': 2018, 'artists': 'Juice WRLD'},
    {'name': 'Staring At The Sun (feat. SZA)',
     'year': 2019,
     'artists': "Post Malone", 'SZA"}]
```

> OUTLINE

> TIMELINE

main\* 3 10 0

Cell 1 of 7 Go Live

# Content Machine Learning

# Music Recommendation



['Aerosmith']  
I Don't Want to Miss a Thing -  
From "Armageddon" Soundtrack

['Birdy']  
Skinny Love

['Breaking Benjamin']  
The Diary of Jane -

['AC/DC']  
Hell's Bells

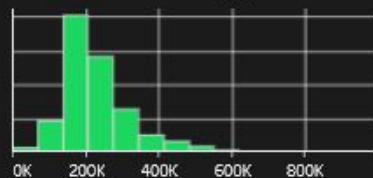
['Adele']  
When We Were  
Young

['Queen']  
Bohemian  
Rhapsody -

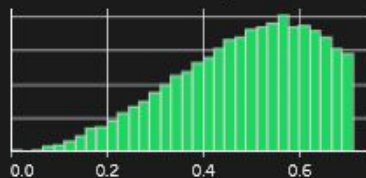
['Air Supply']  
Making Love Out  
of Nothing at All

['Bread']  
If

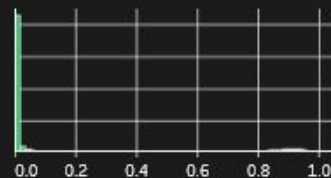
Duration (Ms)



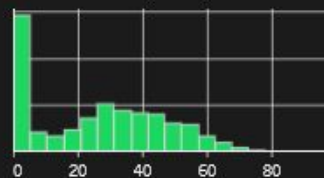
Danceability



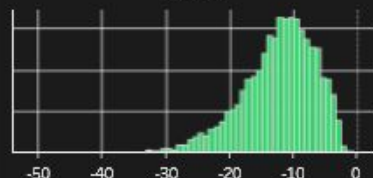
Instrumentalness



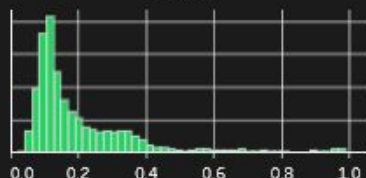
Popularity



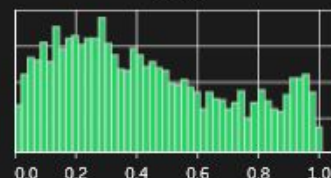
Loudness



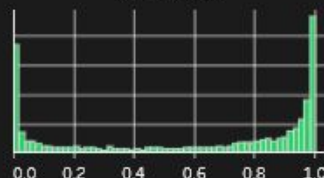
Liveness



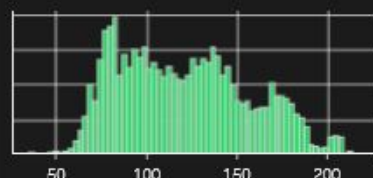
Energy



Acousticness



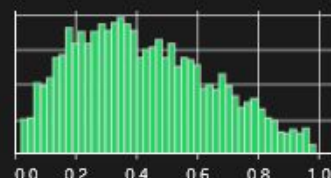
Tempo



Speechiness



Valence



Explicit

Clean

Mode

Major



## In Summary



**Goal:** Develop a machine learning algorithm that can provide music recommendations based on user input (song name and year).

To complete the project, we:

1. Cleaned the dataset by purging unnecessary variables.
2. Created a Tableau dashboard to explore the variables.
3. Trained our model using our data.
4. Connected to Spotify API to enable users to input a song of their choice.



# Lessons & Findings



## Future Directions & Additional Features

- Premium access
- Larger and more recent dataset

## Project Challenges

- The model demonstrates meaningful predictive power at least 75% classification accuracy or 0.80 R-squared.\*
- Overall model performance is printed or displayed at the end of the script\*



Search



Home



Library



Our Music Playlists

“Spotify’s mission is to unlock the potential of **human creativity** by giving a million creative artists the opportunity to **live off their art** and billions of fans the opportunity **to enjoy and be inspired** by it.”

—Spotify’s Mission Statement







Search



Home



Library



Our Music Playlists

# Thanks!

Do you have any questions?



Your very  
own DJ



Spotify Technology SA

NYSE: SPOT 1

Overview

Compare

Financials

Market Summary > Spotify Technology SA

159.24 USD

+0.71 (0.45%) ↑ today

Closed: Sep 18, 4:24 PM EDT • Disclaimer

After hours 159.11 -0.13 (0.082%)

1D 5D 1M 6M YTD 1Y 5Y Max



Open	157.36	Mkt cap	31.00B	52-wk high	182.00
High	160.20	P/E ratio	-	52-wk low	69.28
Low	155.54	Div yield	-		

More about Spotify Technology

Feedback

Explore more

Related

Following

Apple Inc	172.97 USD	↑ 1.65%
Netflix Inc	394.40 USD	↓ 0.64%
Amazon.com, Inc.	139.98 USD	↓ 0.29%
Microsoft Corp	329.06 USD	↓ 0.35%

Disclaimer

About

open.spotify.com

Spotify is a Swedish audio streaming and media services provider founded on 23 April 2006 by Daniel Ek and Martin Lorentzon. It is one of the largest music streaming service