

Spotify Artist Recommendation Engine

Group 4: Anu Thudianlackel, Dharmil Sanghavi, Dai Qing, Pooja Patil

Summary

At present, Music has become an important media for every individual. With the ease of internet, and applications like Spotify, Sound-cloud, Amazon-Music, Apple-Music, you can have access to your desired soundtracks at your fingertips. As a part of our project, we used the Spotify App's data in order to develop an artist recommendation model based on artist's genres, subgenres, followers and popularity which will help users to pick songs of their interest. We have used both Python and Tableau in order to achieve our project goals.

Project Motivation

We all in group are fan of few artists and kind of like only some music genres. This made us think that why not build a recommendation model which could suggest us artist of the same genre based on our favourite artist choice. We did find this for the song recommendation depending on our song choices but there was nothing available for artist recommendation. Hence, we thought to do this as our project.

Introduction

Our aim is to provide a recommendation system using content based filtering algorithm that recommends its users similar artist based on any particular genre, sub-genre, followers, popularity. I have chosen to build Content-based recommendation system to predict more accurate results by text based recommendation. Content-based recommendation mainly depends on item features for recommendations. In short, we will use Python to develop this model and use Tableau to help Spotify understand its user's choice and aid them to enhance their business & customer satisfaction.

Business Value

With this recommendation system, we would be able to easily predict insights for any particular artist ranging from its genres, sub-genres to its popularity, followers. Through Tableau Data visualization, business can scale up the views for other artists.

Technologies used

1. Python - For Data Analysis & Exploration
2. Tableau - For Data Visualization

A. Data Analysis (Python)

Implemented content-based filtering algorithm along with Natural Language Processing(NLP) to obtain the recommendation on similar artist that resulted in achieving 90% accuracy on search results.

- We have used python libraries like Pandas and NumPy for data gathering, data filtering, data processing and data cleaning steps.

	Artist	Genre	Sub-Genre	Followers	Popularity
687	Zarema	Pop	Deep Turkish Pop	29	16
688	ZAYN	Pop	Dance Pop	7,500,974	83
689	Zé Ramalho	Rock	Brazilian Rock	751,023	65
690	Zendaya	Pop	Dance Pop	2,243,059	76
691	Zezé Di Camargo & Luciano	Other	Axe	871,551	65

- We implement 'Bag of words(BOW)' technique on our new column 'recommend_solver'. BOW is a technique that describes the occurrence of a word within a document or an entry. So now recommend_solver merges the columns genres and sub-genres together which will help us find the artist.
- Artist is the index for which we will be using the 'recommend_solver' in order to predict similar artist based on genres and sub-genres.

```
In [8]: artist_data['recommend_solver'] = artist_data['Genre'] + ' ' + artist_data['Sub-Genre']
```

```
In [9]: artist_data.drop(['Genre', 'Sub-Genre'], axis=1, inplace=True)
```

```
In [10]: artist_data.set_index('Artist', inplace=True)
```

```
In [11]: artist_data.head()
```

```
Out[11]:
```

	Followers	Popularity	recommend_solver
Artist			
2NE1	1,118,731	63	Pop Dance Pop
14 Bis	129,166	46	Other Mpb
1976	8,553	31	Other Chinese Indie
"Weird Al" Yankovic	379,346	60	Pop Antiviral Pop
A Turma Do Balão Mágico	33,693	46	Other Musica Infantil

- Now, the data is ready to apply NLP. Here, we use Scikit-learn library so that we could find similar movies through functions CountVectorizer and Cosine_Similarity.

Now the Data is ready to apply NLP to it.

```
In [12]: from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

```
In [13]: count = CountVectorizer()
```

```
In [14]: count_matrix = count.fit_transform(artist_data['recommend_solver'])
```

```
In [15]: cosine_sim = cosine_similarity(count_matrix, count_matrix)
```

- Once we define the functions from Scikit-learn library, the next step is to build the recommendation model. The artist_recomender analyzes the artist name and then through Cosine_sim, it recommends top 10 artist by matching it with other artist's genres and sub-genres.
- Later, we call spotify_Artist_recommender() which creates the recommendation engine result.

Creating the Recommendation Model

```
In [16]: def artist_recomender(Artist, cosine_sim = cosine_sim):
    recommended_artist = []
    indices = pd.Series(artist_data.index)
    idx = indices[indices == Artist].index[0]
    recommend = pd.Series(cosine_sim[idx]).sort_values(ascending = False)
    recommend = recommend[1:11]
    top_10_index = recommend.index
    for i in top_10_index:
        recommended_artist.append(list(artist_data.index)[i])
    return recommended_artist
```

```
In [17]: # Artist similar to:
artist_recomender(Artist='Drake')
```

```
Out[17]: ['Preme',
'Metro Boomin',
'Wiz Khalifa',
'Tyler, The Creator',
'Ty Dolla $ign',
'SZA',
'Lil Jon & The East Side Boyz',
'Gorillaz',
'Aron Can',
'OJ Da Juiceman']
```

```
In [18]: def spotify_artist_recommender():
    print('Welcome to Spotify Artist Recommender')
    print('Enter any Artist from the Spotify 2018 List')
    artist_similar = input()
    result = artist_recomender(artist_similar)
    print('\nBecause you listen to:', artist_similar)
    print('\nYour artist recommendation should be:\n\n', result)
```

- The recommender for artist is ready. We need to input the name of any artist from the 691 distinct artist data set which the recommender will use to generate its top 10 artist recommendation based on artist's genre, sub-genre, popularity.

```
In [*]: ▶ spotify_artist_recommender()

Welcome to Spotify Artist Recommender
Enter any Artist from the Spotify 2018 List
```

- Let's consider 2 examples: When we input our artist as 'Taylor Swift', we get the output in the form of recommendations. The artist recommendations we get is Cyndi Lauper, Christina Aguilera, Ciara, Omarion, Olly Murs, Harry Styles, Nicki Minaj, Niall Horan, Natasha Bedingfield, David Guetta.
- Also, when we input 'Drake' as the artist name, the recommendations we get are Wiz Khalifa, Ty Dolla \$ign, Aron based on the artists genre and popularity.

```
In [19]: ▶ spotify_artist_recommender()

Welcome to Spotify Artist Recommender
Enter any Artist from the Spotify 2018 List
Taylor Swift

Because you listen to: Taylor Swift

Your artist recommendation should be:

['Cyndi Lauper', 'Christina Aguilera', 'Ciara', 'Omarion', 'Olly Murs', 'Harry Styles', 'Nicki Minaj', 'Niall Horan', 'Natasha Bedingfield', 'David Guetta']
```

```
In [20]: ▶ spotify_artist_recommender()

Welcome to Spotify Artist Recommender
Enter any Artist from the Spotify 2018 List
Drake

Because you listen to: Drake

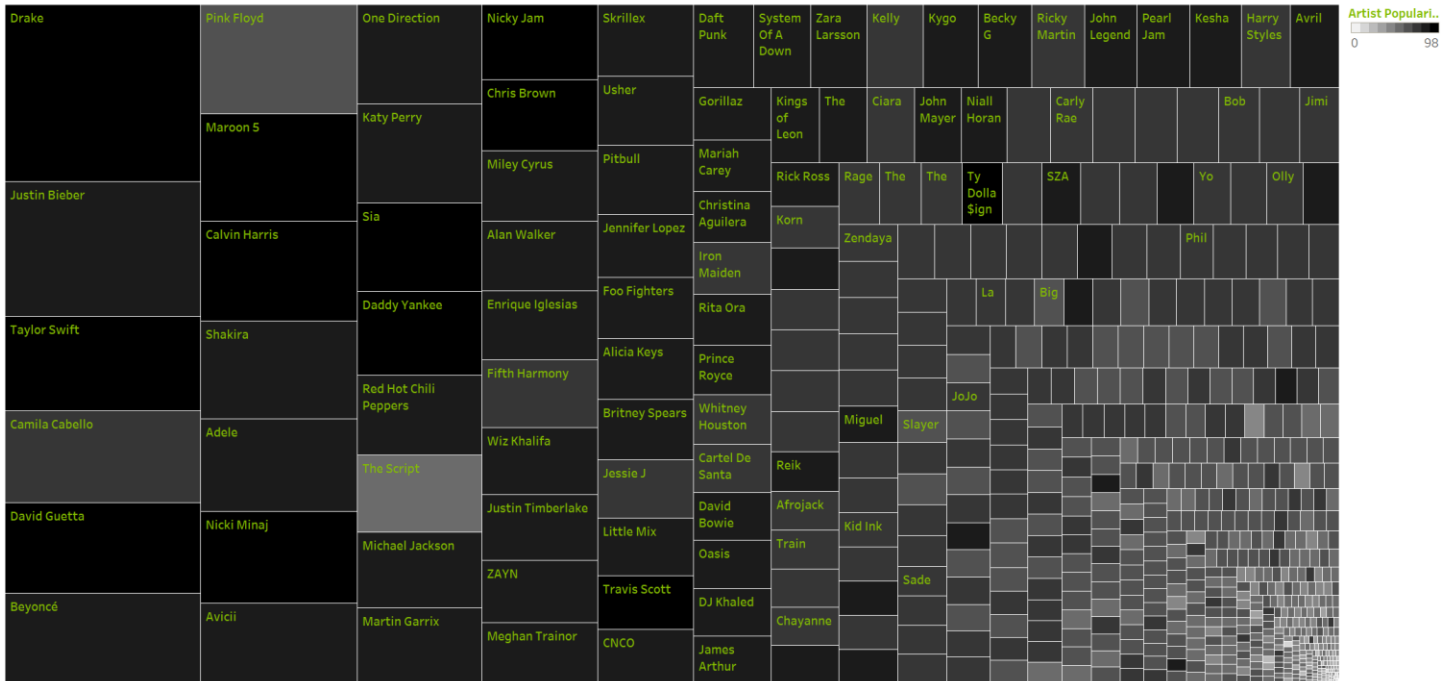
Your artist recommendation should be:

['Preme', 'Metro Boomin', 'Wiz Khalifa', 'Tyler, The Creator', 'Ty Dolla $ign', 'SZA', 'Lil Jon & The East Side Boyz', 'Gorillaz', 'Aron Can', 'OJ Da Juiceman']
```

B. Data Visualization (Tableau)

- The color represents the artist popularity. Darker the shade, higher is their popularity.
- The size represents the total number of followers of a particular artist. Bigger the size of the data mark, larger is their fan following.
- The marks are labeled by artist.

Artist Followers & Popularity



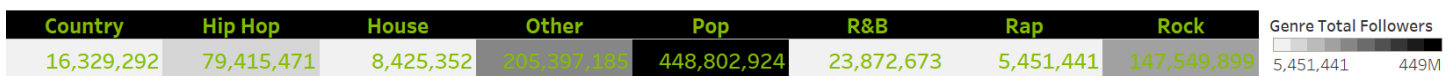
- The records show the sum of total artist is represented based on their music Genre. It represents the artist count for a particular genre. There are 8 main genres in total including Country, Hip Hop, House, Other, Pop, R&B, Rap and Rock.
- Darker the shade, higher is the count of artist for a particular genre.

Genre Count



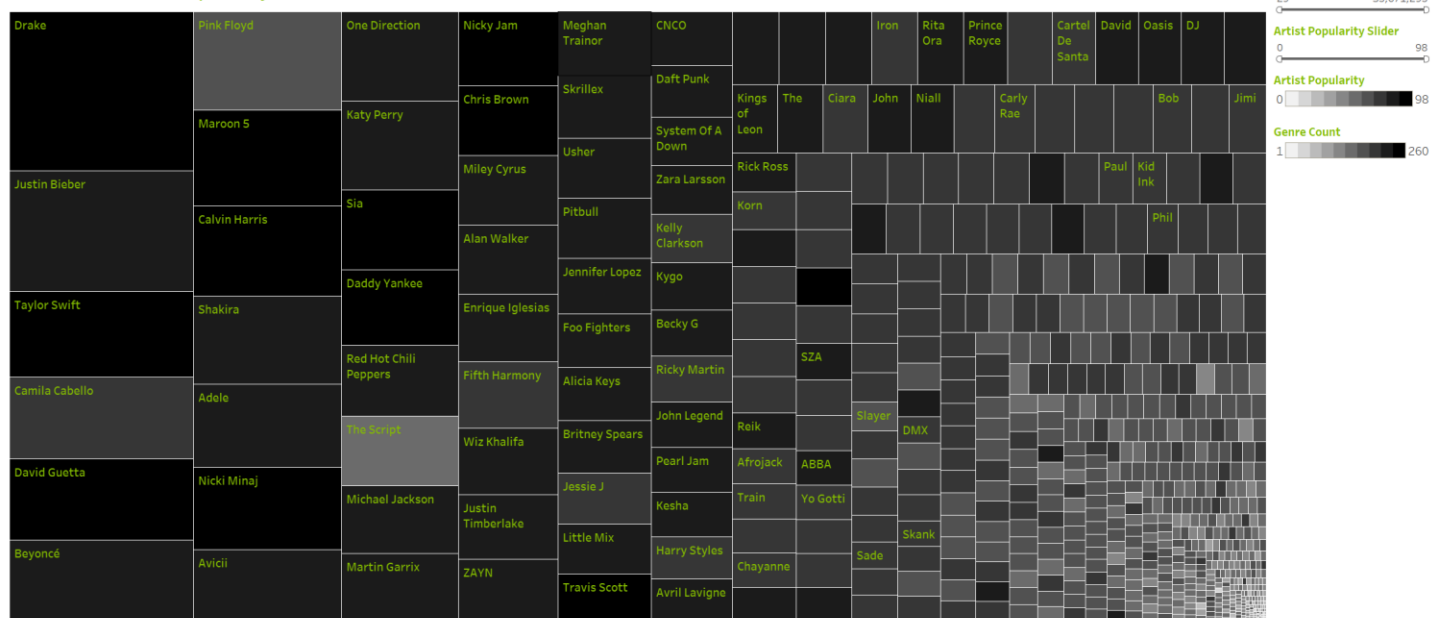
- The records show the sum of total artist is represented based on their music Genre. The marks are labeled by sum of followers. The view is filtered on sum of followers.
- Darker the shade, higher is the total followers for that particular genre.

Genre Total Followers



- This is our final dashboard. Our main focus is the 'Artist Followers & Popularity' represented using a tree-map. The other tables, Genre Count and Genre Total Followers are used in the form of filters for which the target is Artist Followers & Popularity. So if we select Hip-hop from Genre Count, we would be shown only the 31 artist that belongs to the Hip-Hop Category.
- We have two range sliders on top-right corner. The first one is Artist Followers which is for selecting the range for the total fan following the artist has on Spotify. The other is Artist Popularity Slider which is for selecting the range for artist's popularity. Once the range is applied, the changes gets reflected on the Artist Followers & Popularity tree-map.
- Lastly, there are 2 cards, one being for Artist Popularity and other is Genre Count. Darker the shade, higher is the artist's popularity. Darker the shade, larger is the artist count for that genre.

Artist Followers & Popularity



Genre Count

Country	Hip Hop	House	Other	Pop	R&B	Rap	Rock
13	31	1	260	256	12	5	114

Genre Total Followers

Country	Hip Hop	House	Other	Pop	R&B	Rap	Rock
16,329,292	79,415,471	8,425,352	205,397,185	448,802,924	23,872,673	5,451,441	147,549,899



Conclusion

We collected data from Kaggle and then followed data cleaning techniques to remove null values and duplicates from dataset. That dataset is used to run the bag of words model to find the recommendations for given artist.

Also, we used content-based filtering algorithm along with NLP to obtain recommendation on similar artist that resulted in achieving 90% accuracy on search results. Lastly, through interactive Tableau dashboard we are able to differentiate the artist's followers and popularity based on the color shade and size of shape.

References

1. Dataset:

[https://public.tableau.com/vizql/w/SonyMusicArtistSpotifyData/v/SonyMusicArtistSpotify2/viewData/sessions/C82CF923CA844EE491831A8BC269A0E8-0:0/views/712297945292471785_7171632493191854206?maxrows=200&viz=%7B%22worksheet%22%3A%22Sony%20Music%20Artist%20List%22%2C%22dashboard%22%3A%22Sony%20Music%20Artist%20Spotify%20\(2\)%22%7D](https://public.tableau.com/vizql/w/SonyMusicArtistSpotifyData/v/SonyMusicArtistSpotify2/viewData/sessions/C82CF923CA844EE491831A8BC269A0E8-0:0/views/712297945292471785_7171632493191854206?maxrows=200&viz=%7B%22worksheet%22%3A%22Sony%20Music%20Artist%20List%22%2C%22dashboard%22%3A%22Sony%20Music%20Artist%20Spotify%20(2)%22%7D)

2. Project Idea:

https://github.com/SaikumarReddySandannagari/CMPE256_Spotify_Music_Recommendation_System