1.0 INTRODUCTION

The international party Eurovision just took place. An event that I - before it got smudged into border politics - used to follow religiously. But except for both Australia's awkward and random space in the competition and Israel's more problematic one, what have we always known about the song contest? That you are doomed to the 12th place if you compete with a song in your local language aka with a song with non-English lyrics.

The aim of the project is to interrogate the importance of language in music recommendations on Spotify: Does Spotify have a bias towards non-English languages and how does that impact the visibility of artists.

2.0 EXPERIMENT

Made two accounts with user names Mette Hansen and Anvi Sen. I am gonna obtain their Daily Drive playlists for both users over almost a week. Based on this I am gonna look at the improvement of songs recommendation over the period of time I am collecting data. As Spotify has no available data on languages, I am gonna use the python library, langdetect in order to retrieve this information on the different recommended tracks.

Using matplotlib. I am gonna visualize how the distribution is for the different playlist for both users. For simplicity, I am only gonna do this for the songs recommended (after having used Spotify for a bit), the third and the fifth day. The analyses are based in a comparative analysis of the findings here.

2.1 TOP 5 LANGUAGES

According to music tomorrow the Spotify recommender algorithm is based on both content and collaborative filtering. The content-based filtering being based on the audio features that includes stuff like danceability, tempo and instrumentalness and are available through the Spotify API. Surely, tempo is an universal thing, so I am interested in seeing if it is true that Spotify recommends equally between different languages based on these features.

I used snippets from the class notebook 'Week 3.1 Spotify Playlists task' to look at 6 playlists with popular songs in the languages: Spanish, English, Hindi, Mandarin, Arabic and Korean. These languages were picked based on the amount of people in the world speaking those, and then I added Korean because K-pop is such a popular genre, and I wanted to incorporate that.

The playlists were found searching for 'most popular *language* songs by Spotify' in Google. I found playlists made officially by Spotify except for songs in Mandarin, already confirming a bias. I am therefore dropping using songs in Mandarin.

.

From the heatmap I observed that songs in different languages did indeed have similar scores.

This shows how 'Do I Wanna Know' by Arctic Monkeys has a score similarity score of 0.753 with اطيب بديت by Majid Almohandis from the Arabic playlist.

"Yo También" by Romeo Santos (feat. Marc Anthony) from the Spanish playlist has a score on 0.91 with "Cigara" by Bahaa Sultan from the Arabic, while this same track has almost noting in commen with العين يخزي by Bashaar Al Jawad - Yekhzi L 3ayn, even though they are from the same language playlist.

1) When asked to do recommendations based on Arabic 'Bahaa Sultan - Cigara' these are the recommended songs:

'Justin Bieber - Ghost', 'Manuel Turizo - Esperándote', **'Olivia Rodrigo - good 4 u'**, 'Tainy - Agua - Music From "Sponge On The Run" Movie', 'Mau y Ricky - Desconocidos', **'NewJeans - Zero'**, **'THE NEW SIX - Love or Die'**, 'Bad Bunny - Yonaguni', 'Aditya A - Chaand Baaliyan', 'YUJU - Without U']

The ones in bold are non-English songs. 4/10 are English songs. None of them are Arabic.

2) Recommendations based on Korean, BTS - Take Two.

'Arctic Monkeys - Do I Wanna Know?', 'GAYLE - abcdefu', **'Camila Cabello - Bam Bam (feat. Ed Sheeran)'**, 'Lunay - Soltera - Remix', 'YUJU - Without U', 'ENHYPEN - SHOUT OUT', 'Amr Diab - Habiby Ya Omry', **'Justin Bieber - Ghost'**, 'Maninder Buttar - Sakhiyan2.0 (From "BellBottom")', 'Majid Almohandis - اطيب بديت']

4/10 are English songs. 2/10 are Korean songs.

3) Recommendations based on the Spanish, King - Tu Aake Dekhle:

'Wael Kfoury - Set El Kel', 'Tamer Ashour - Hatwhashna', **'NewJeans - Zero'**, **'OneRepublic - I Ain't Worried'**, 'Alvaro Soler - Sofia', 'Ozuna - Dile Que Tu Me Quieres', 'Yasser Abd Alwahab - Allah Honej', 'Fadel Chaker - Ya Ghayeb', 'Aditya A - Chaand Baaliyan', 'A.R. Rahman - Param Sundari'

2/10 are English songs. 2/10 are Spanish songs.

4) Recommendations based on the Hindi, A.R. Rahman - Param Sundari

'aespa - Welcome To MY World (feat. nævis)', 'Jesse & Joy - ¡Corre!', 'Majid Almohandis - اطيب بديت', 'Pritam - Shayad (From "Love Aaj Kal")', 'Han Dong Geun - dawn call from you', 'ENHYPEN - SHOUT OUT', 'BTS - Take Two', **'Arctic Monkeys - Do I Wanna Know?'**, 'GAYLE - abcdefu', 'Amr Diab - Habiby Ya Omry']

1/10 is English. 1/10 is Hindi.

5) Recommendation based on English, Frank Ocean - Lost:

'Ahmed Saad - Mansethosh Men Mosalsal El Aghar', 'Ramy Gamal - بعادك لحظة', 'Nouamane Belaiachi - Wala Bghiti', 'Tamer Hosny - Nasseeny Leih', 'Yasser Abd Alwahab - Allah Honej', 'Rochak Kohli - Dil De Diya (From "Thank God")', 'Mohamed el Majzoub - Eftahli Albak', **'Tom Odell - Another Love'**, 'Angham - ارضيك عشان', 'Wael Kfoury - Set El Kel'],

1/10 is an English song.

More than half of the songs recommended are non-English, so looks like Spotify actually do quite varied recommendations.

2.2. USER ACCOUNTS

If Spotify is so good at recommendations, I wanna try it myself. I opened up two new accounts with the same characteristics: Both cis-women, both born in 1983 and both based in the UK. One, Mette Hansen will only listen to songs with English language and one, Anvi Sen to songs with only non-English language.

They got recommended the exact same daily list. This makes a good starting point. I will see how this playlist changes from the starting point to the end of this project.
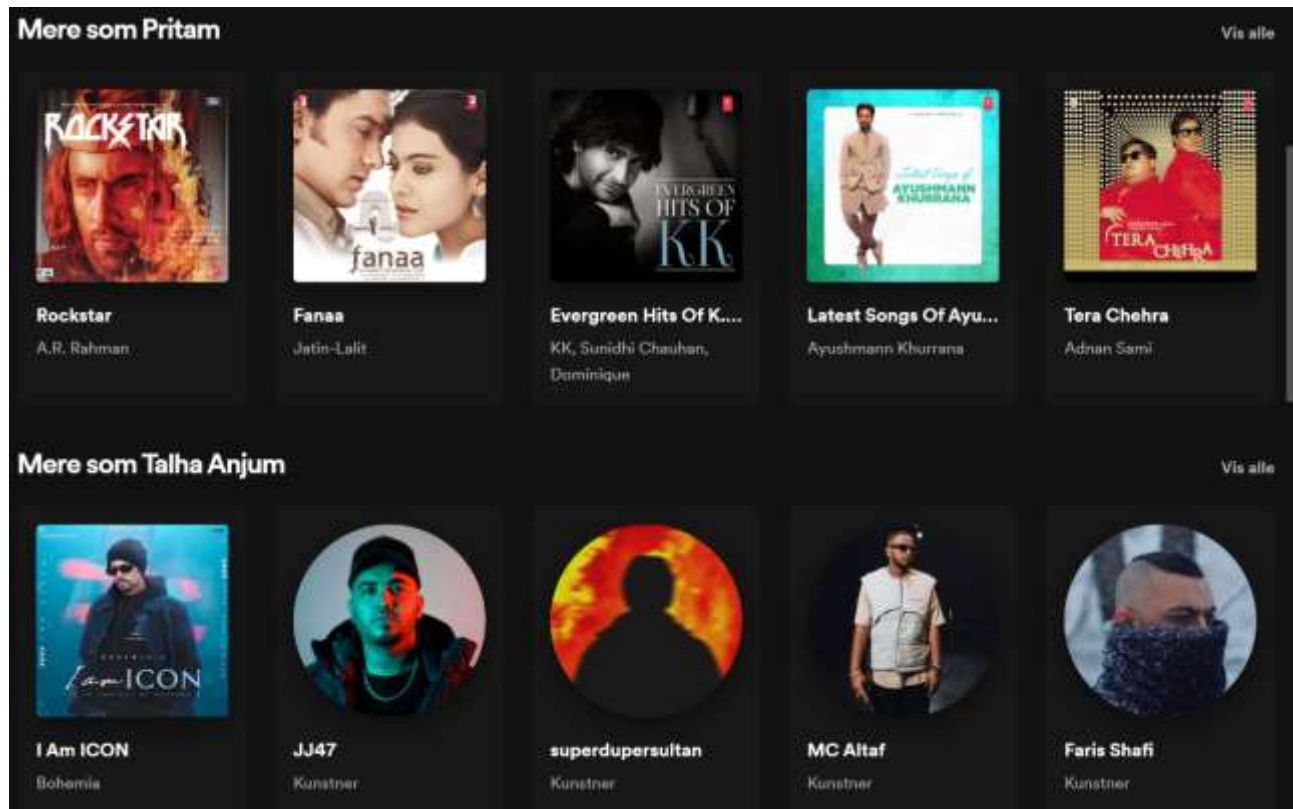
The list included songs like:

- Libianca - People | Kali Uchis - Moonlight | Tom Odell - Another Love | Lil Durk - All My Life (feat. J. Cole) | Peso Pluma - PRC | Dua Lipa - Dance The Night (From Barbie The Album) | PinkPantheress - Boy's a Liar Pt. 2 | The Weeknd - Popular (with Playboi Carti & Mado... | Taylor Swift - Blank Space | Manuel Turizo - La Bachata

9 out of 10 songs from this list is English.

Everytime one of the things got recommended a song, I would listen to it and add it to the playlist. In that way everything got listened to. So I guess the only bias here is that the algorithm assumes you want western music.
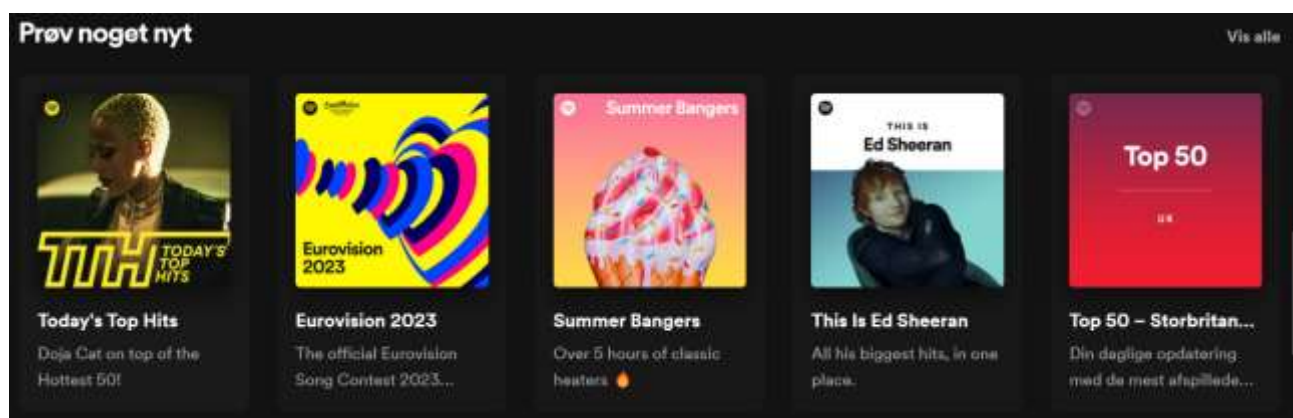
2.3 INITIAL OBSERVATIONS

Everything that gets recommended to both are quite western based, but as soon as Anvi started listening to artists like Jatin-Lalit and Adnan Sami it got more varied. Spotify has these content-based recommendations.
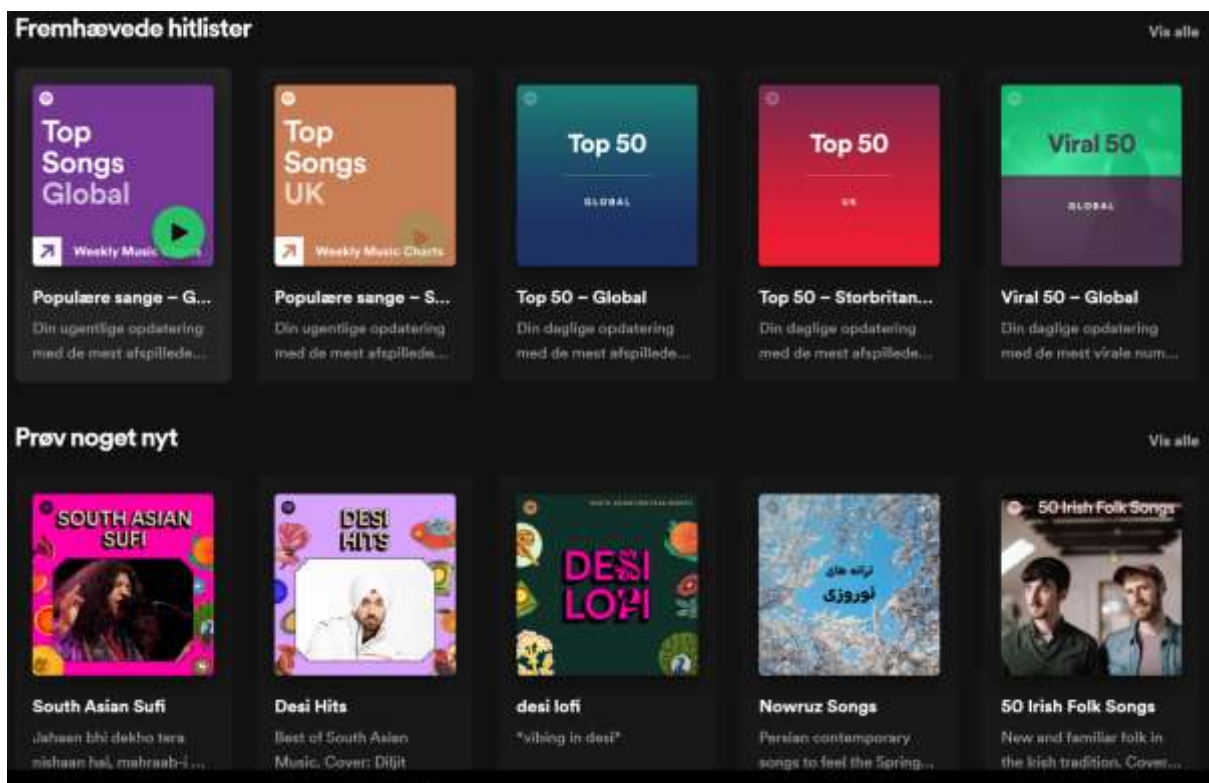

Figur 1Anvi: Content-based recommendations

Even more interesting is the 'Try something new' recommendations, which are supposed to expose you music that - based on your listening patterns - you would not have been recommended. All of these are very based in English music.
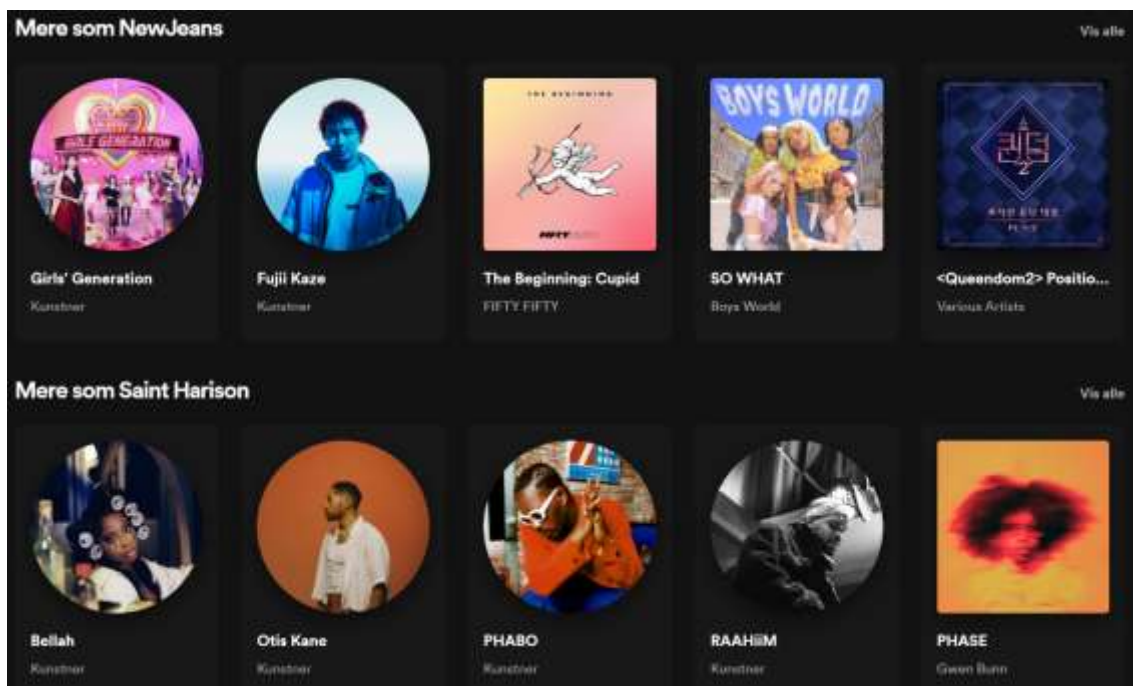

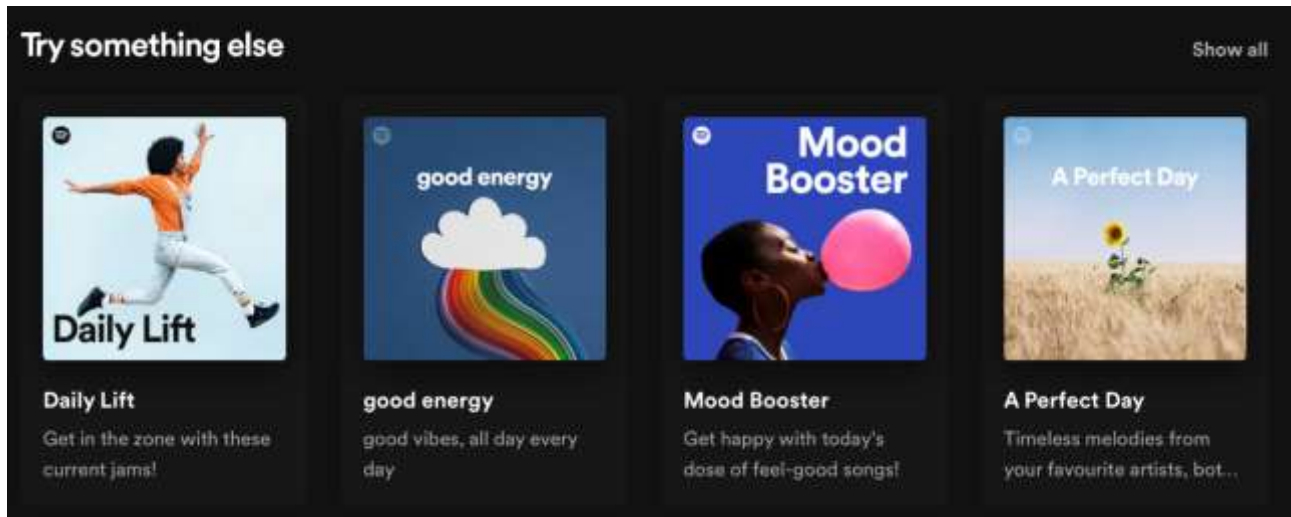*Figur 2Anvi: Try something new recommendations*

Figur 3 Anvi: recommendations after half a week

Mette got the same recommendations based on what she had listened to.

Figur 4 Mette: Content-based recommendations

She also got a 'Try something else' recommended playlist. But that didn't include music from non-english speaking artists.

Figur 5 Mette: Try something else

After that, I loaded in the data using the Spotify API for developers. Had to set up my own - think that the one used in class had reached it's limit.
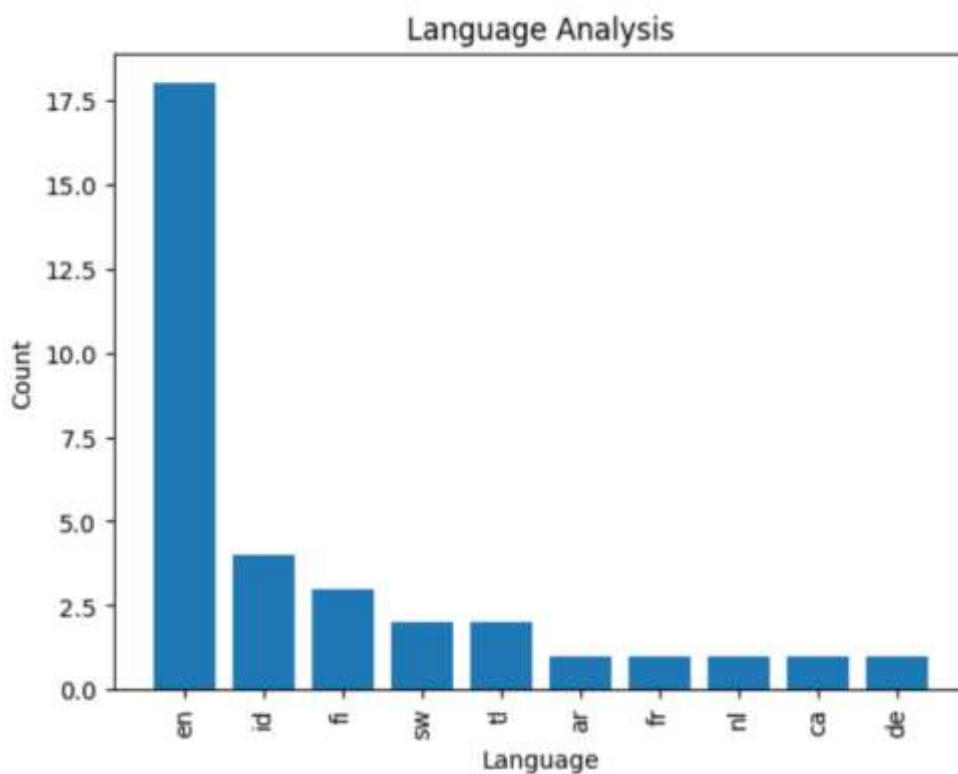
3.0 LOOKING AT THE DATA

3.1 ANVI SEN

Tabular data for the user Anvi Sen who only listened to non-English songs during the period of the experiment.

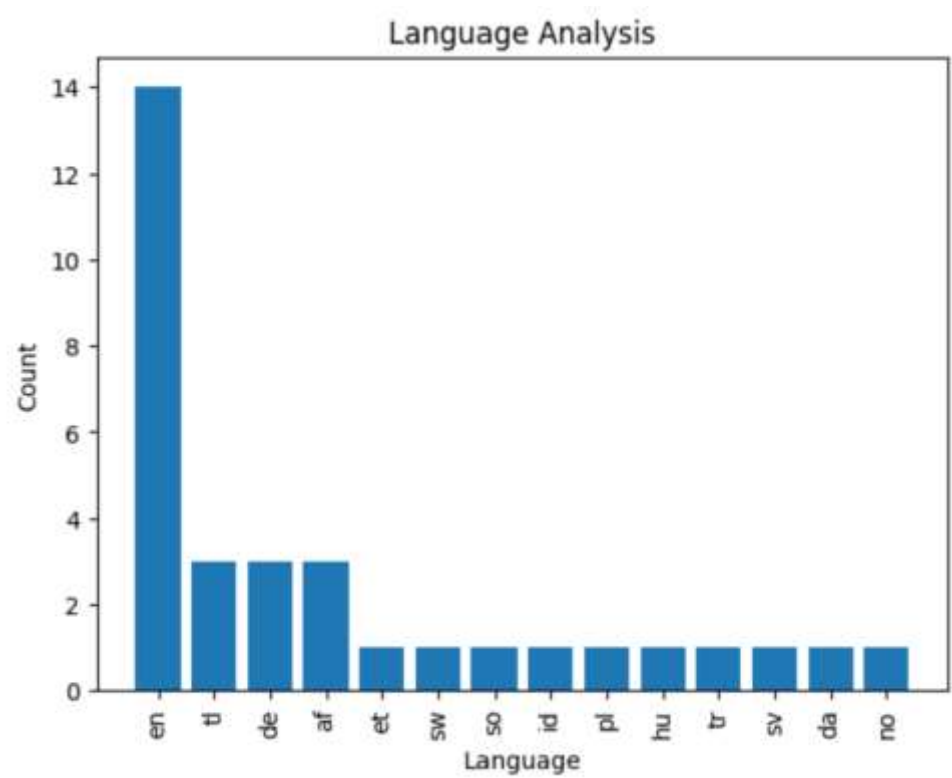|	Unnamed: 0	Track Name	Artist	Album	Duration (ms)	Popularity;
70	70	Entammede Jimikki Kammal	Vineeth Sreenivasan	Velipadinte Pusthakam (Original Motion Picture...	201600.0	58;
71	71	Pepeta	Nora Fatehi	Pepeta	213695.0	50;
72	72	One Night in Dubai (feat. Helena)	Arash	One Night in Dubai (feat. Helena)	158106.0	60;
73	73	Masalei Ni	TM Bax	Masalei Ni	255425.0	48;
74	74	ABUSADAMENTE - 3000beats Remix	3000BEATS	ABUSADAMENTE (3000beats Remix)	203763.0	51;
75	75	I Like You (A Happier Song) (with Doja Cat)	Post Malone	Twelve Carat Toothache	192840.0	87;
76	76	UNAVAILABLE (feat. Musa Keys)	Davido	Timeless	169911.0	81;
77	77	Dua Lipa	Jack Harlow	Come Home The Kids Miss You	135053.0	75;
78	78	Be Honest (feat. Burna Boy)	Jorja Smith	Be Honest (feat. Burna Boy)	207030.0	73;
79	79	Calm Down (with Selena Gomez)	Rema	Calm Down (with Selena Gomez)	239317.0	95;
80	80		Sand Lamigrad		240044.0	40;

As there is no available information on the languages in the different tracks from Spotify, I am gonna use the python library, langdetect to look at the actual recommendations given.

The language codes given in the plots are for: English, German, French, Afrikaans, Estonian
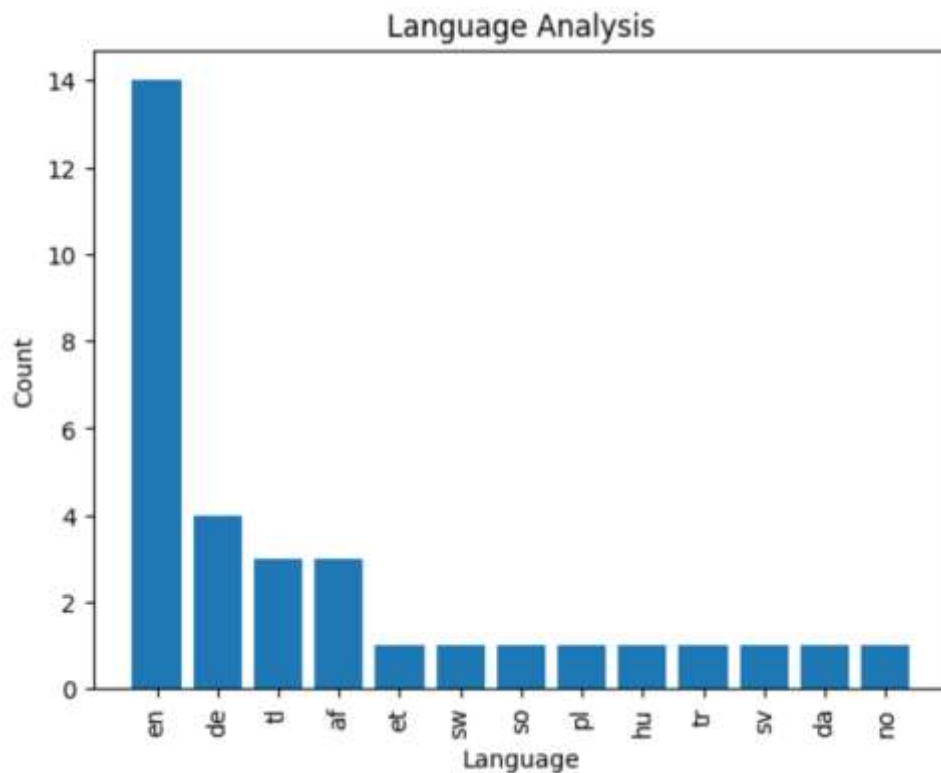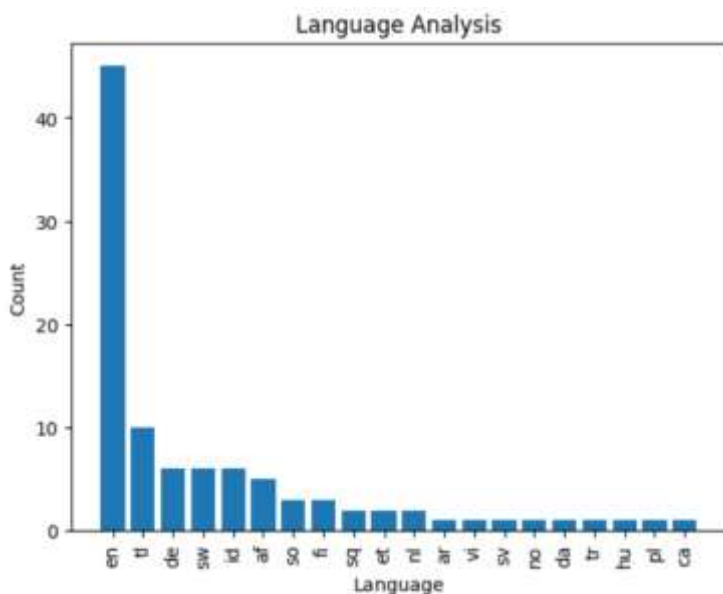
Recommendation from the first day:

Recommendations from the third day:

Language Analysis

Recommendation from the fifth day:

Language Analysis

These look very similar, but significant is the amount of songs in English that gets recommended, only falling from 17,5 to 14 between day 1 and 5 is not a lot out of 35. It is also worth noticing that the languages recommended are mostly European languages even though I went straight into the Top 30 best Hollywood songs playlist.
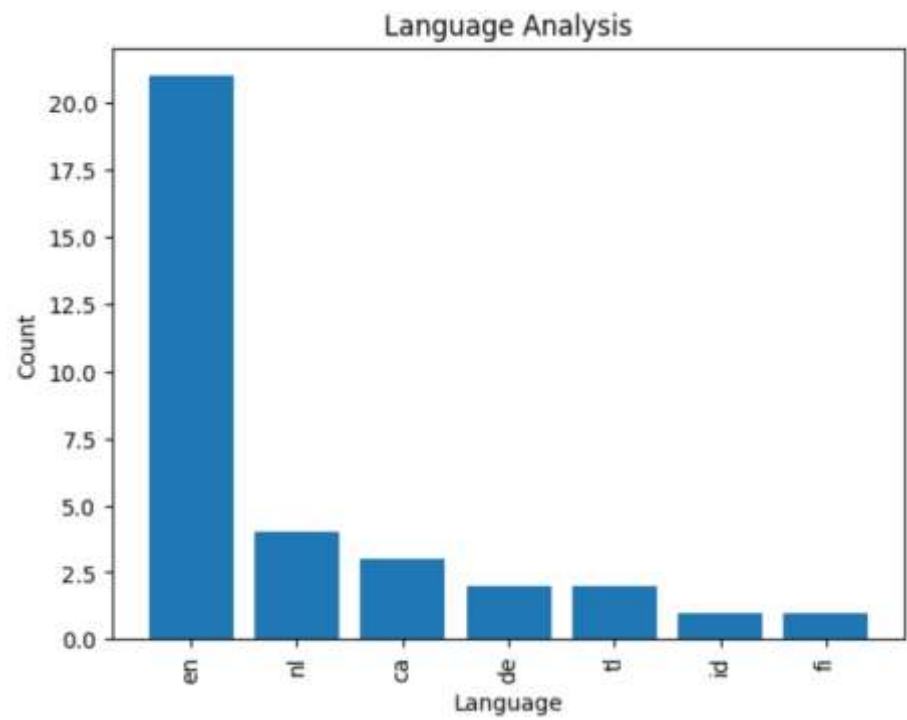

Language Analysis

Same pattern is seen in the total distribution of languages in recommended songs for Anvi Sen - the user who only listened to non-English songs.

This clearly shows that most of the songs were actually English even if the user only listened to non-English music, showing a clear bias.
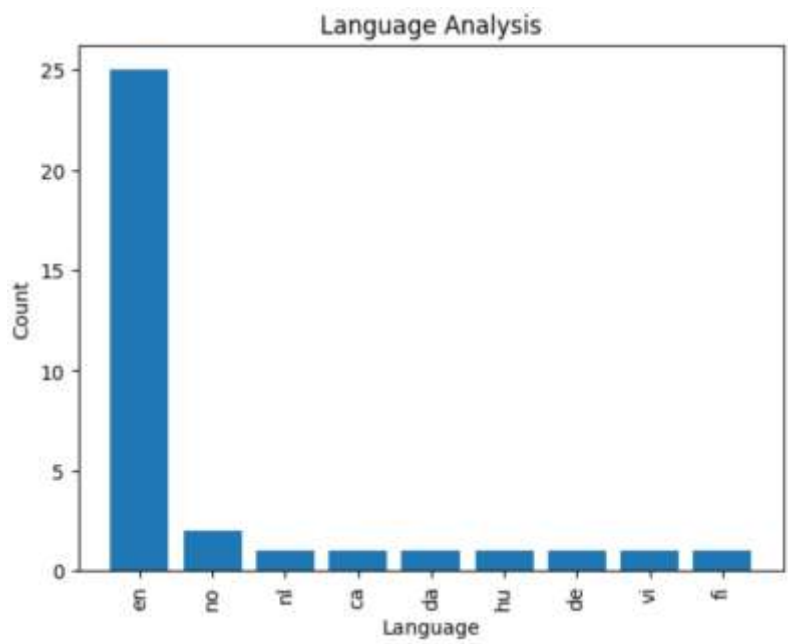
3.2 METTE HANSEN

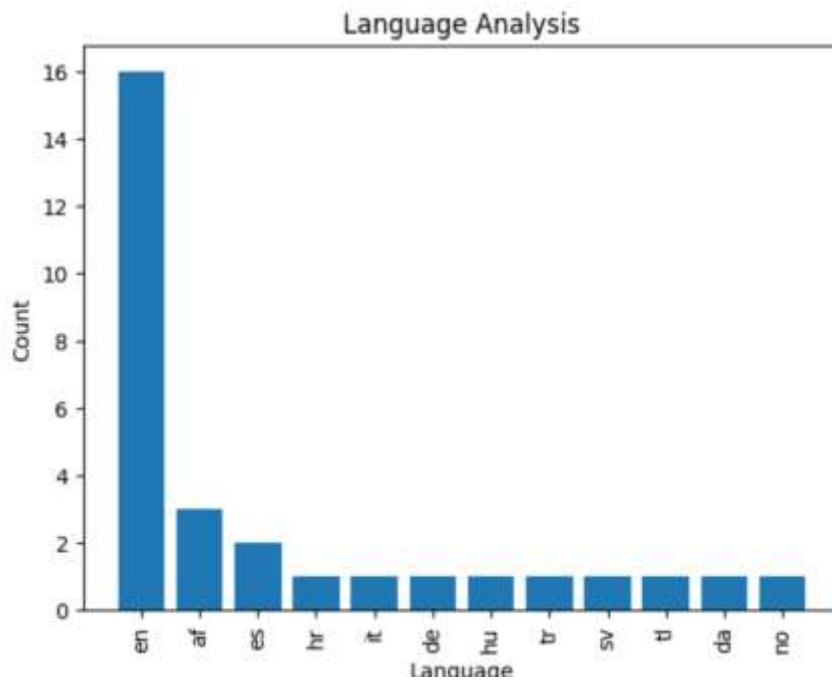I am gonna do the same for the user, Mette Hansen, who only listened to English songs.
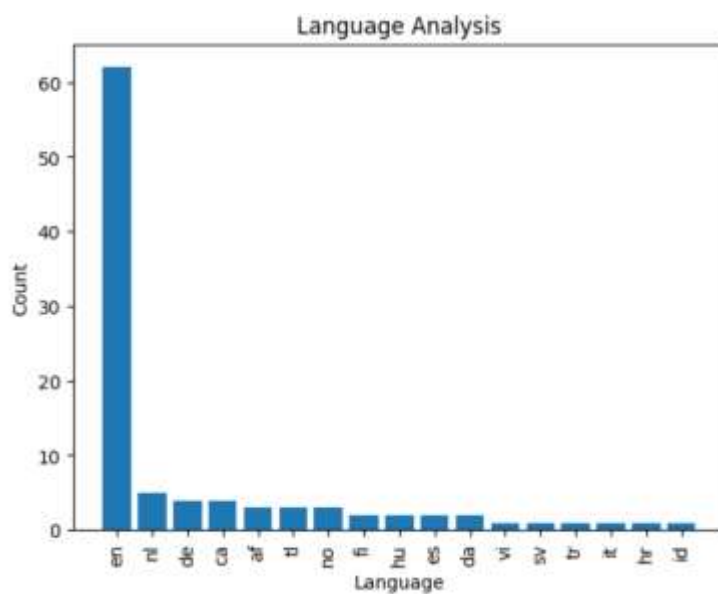
Recommendations from the first day:

Recommendations from the third day:

Recommendation based on fifth day:

Language Analysis

The results here are honestly not surprising. Mette, who only listened to English songs are also getting recommended English songs. In the beginning, the songs are only English, but they drop from 25 on the third day to 16 on the fifth, which is a bit surprising. Also, worth noticing though is it that there are a few songs detected as Afrikaans, so not a European language. Same pattern with the total distribution:

Language Analysis

4.0 CONCLUSION AND FINAL REFLECTIONS

While I do think that this small experiment did show some patterns in the recommender systems on Spotify, I do also think that there was a lot of pitfalls that might have influenced the results quite a bit.

One of them being that both users were based in the UK adding a bias that clearly shows by a lot of the recommended playlists being top ten something in the UK, even if they were not in an English language. To avoid this I could have used a VPN for one of the users.

Another thing is the python library, langdetect, that I used for plotting the languages. It is a library that I discovered quite recently and as with so many other biases in software, I fear that the reason that English and also European languages are detected more often than languages like Hindi, Arabic and Spanish are that the library software are just not as good at those languages. We need to wait for Spotify to release data on this before we can tell.

Last, but not least, the time period of the project most definitely has something to say here. While the data was gathered over a few days, and Spotify is surprisingly good at optimizing the algorithm quickly it still might not have been enough time. I have used my personal Spotify account for over 5 years now and I still think that the recommendations can be off, so this experiment could definitely benefit from longer time.

The last Daily Drive playlist that I looked at had a lot of English songs compared to the fact that the users had exclusively been listening to non-English songs. This shows directly opposite results from the ones retrieved when using Spotify's own API to get the audio features for different songs. This showed a much more varied distribution in languages.

In conclusion, the analysis of the distribution of languages in song recommendations show a clear bias towards English speaking languages.

5.0 REFERENCES

Pastrukhov, Dmitry, 2022. Inside Spotify's Recommender System: A Complete Guide to Spotify Recommendation Algorithms. Retrieved June 14 from (https://www.music-tomorrow.com/blog/how-spotify-recommendation-system-works-a-complete-guide-2022)