Assignment 1

As part of the analysis, I performed an examination of song similarity and recommendations based on a Spotify list of my daily recommendations. This exploration primarily revolved around two tasks: assessing song similarity based on various feature subsets, and making song recommendations using cosine similarity measures.

Task 2: Song Similarity

At first, I did not change the subset features, and the results indicated that for each song, only the song itself had a cosine similarity of 1 compared to the other 49 songs. Afterwards，I subsequently listened to the songs that showed a high degree of similarity and those that showed a low degree of similarity. I discovered that from a subjective listening perspective, the songs with a high degree of similarity were closer in terms of features such as 'acousticness', 'valence', 'instrumentalness' and 'tempo', while the songs with a low degree of similarity were the exact opposite. For instance, 'Lauv - i'm so tired...' isthe most similar to 'Rita Ora - Let You Love Me' and the least similar to 'Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)' in the playlist. It was clear that RIP had a slower tempo and was more emotional.

In my findings, songs with a high cosine similarity (e.g., close to 1) indeed appeared to be similar in their respective features.

Out[47]:

	Lil Nas X - TALES OF DOMINICA	Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)	benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)	Khalid - Love Lies - Snakehips Remix	Lauv - i'm so tired...	Justin Bieber - Lonely (with benny blanco)	Aries - ONE PUNCH	Zara Larsson - Ruin My Life	Tones And I - Dance Monkey	Chelsea Cutler - The Reason	The Chainsmokers - Beach House	Jonas Blue - Rise	Lil Nas X INDUSTRY BABY (feat. Jack Harlow
Lil Nas X - TALES OF DOMINICA	1.000000	0.620052	0.048926	-0.285689	-0.051546	0.463777	-0.042751	0.380783	0.550754	-0.599315	-0.367114	-0.021350	0.244453
Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)	0.620052	1.000000	-0.122883	-0.177848	-0.506178	0.356124	-0.231877	0.122116	0.732593	-0.596931	-0.368033	0.089456	0.225185
benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)	0.048926	-0.122883	1.000000	-0.079466	0.492054	0.200714	0.189903	-0.061024	0.132625	0.320766	-0.001817	-0.180893	-0.131107
Khalid - Love Lies - Snakehips Remix	-0.285689	-0.177848	-0.079466	1.000000	0.240263	-0.390353	0.036516	0.358465	-0.615136	0.233499	0.668610	0.166595	-0.041168
Lauv - i'm so tired...	-0.051546	-0.506178	0.492054	0.240263	1.000000	0.019922	0.269017	0.330740	-0.324586	0.457619	-0.051369	-0.206704	-0.096643
Justin Bieber - Lonely (with benny blanco)	0.463777	0.356124	0.200714	-0.390353	0.019922	1.000000	0.101625	0.047134	0.479184	-0.038383	-0.530448	-0.493397	-0.344167
Aries - ONE PUNCH	-0.042751	-0.231877	0.189903	0.036516	0.269017	0.101625	1.000000	-0.142888	-0.289035	0.473715	0.104757	0.393946	-0.576002
Zara Larsson - Ruin My Life	0.380783	0.122116	-0.061024	0.358465	0.330740	0.047134	-0.142888	1.000000	0.032149	-0.406844	-0.000038	0.117397	0.364764
Tones And I - Dance Monkey	0.550754	0.732593	0.132625	-0.615136	-0.324586	0.479184	-0.289035	0.032149	1.000000	-0.435895	-0.474618	0.018984	0.132027
Chelsea Cutler - The Reason	-0.599315	-0.596931	0.320766	0.233499	0.457619	-0.038383	0.473715	-0.406844	-0.435895	1.000000	0.355730	0.037387	-0.617175
The Chainsmokers - Beach House	-0.367114	-0.368033	-0.001817	0.668610	-0.051369	-0.530448	0.104757	-0.000038	-0.474618	0.355730	1.000000	0.426517	-0.104307
Jonas Blue - Rise	-0.021350	0.089456	-0.180893	0.166595	-0.206704	-0.493397	0.393946	0.117397	0.018984	0.037387	0.426517	1.000000	-0.129049

Example of similarity results

	Lil Nas X - TALES OF DOMINICA	Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)	benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)	Khalid - Love Lies - Snakehips Remix	Lauv - i'm so tired...	Justin Bieber - Lonely (with benny blanco)	Aries - ONE PUNCH	Zara Larsson - Ruin My Life	Tones And I - Dance Monkey	Chelsea Cutler - The Reason	The Chainsmokers - Beach House	Jonas Blue - Rise	Lil Nas X - INDUSTRY BABY (feat. Jack Harlow)
Lil Nas X - TALES OF DOMINICA	1.000000	0.620052	0.048926	-0.285689	-0.051546	0.463777	-0.042751	0.380783	0.550754	-0.599315	-0.367114	-0.021350	0.244453
Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)	0.620052	1.000000	-0.122883	-0.177848	-0.506178	0.356124	-0.231877	0.122116	0.732593	-0.596931	-0.368033	0.089456	0.225185
benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)	0.048926	-0.122883	1.000000	-0.079466	0.492054	0.200714	0.189903	-0.061024	0.132625	0.320766	-0.001817	-0.180893	-0.131106
Khalid - Love Lies - Snakehips Remix	-0.285689	-0.177848	-0.079466	1.000000	0.240263	-0.390353	0.036516	0.358465	-0.615136	0.233499	0.668610	0.166595	-0.041164
Lauv - i'm so tired...	-0.051546	-0.506178	0.492054	0.240263	1.000000	0.019922	0.269017	0.330740	-0.324586	0.457619	-0.051369	-0.206704	-0.096640

benny blanco - Unlearn (with Gracie Abrams)	Zedd - Happy Now	NOTD - I Wanna Know (feat. Bea Miller)	Marshmello - Happier	Khalid - OTW	Camila Cabello - Real Friends	Lil Nas X - THATS WHAT I WANT	benny blanco - Eastside (with Halsey & Khalid) - Acoustic	Zedd - The Middle	5 Seconds of Summer - Youngblood	Marshmello - Light It Up (with Tyga & Chris Brown)	Bebe Rexha - Meant to Be (feat. Florida Georgia Line)	Billie Eilish - bad guy (with Justin Bieber)	Lil Nas X - DEAD RIGHT NOW	Cha... - Tl...
0.116745	0.070419	-0.176398	-0.047216	-0.413940	-0.099485	0.328442	0.385013	0.174362	-0.124420	0.600468	0.188201	-0.297251	-0.201410	
-0.093564	0.194460	-0.044503	0.259144	0.255879	0.454503	0.310378	0.102476	0.332873	0.004124	0.407722	-0.260609	-0.406782	0.198463	
0.076477	-0.241081	-0.357321	-0.094317	-0.436501	-0.014636	0.081361	0.257645	-0.251340	0.340143	0.267291	-0.274551	-0.037638	-0.054721	
-0.269626	0.218360	0.538382	0.384142	0.145579	-0.446438	-0.045329	-0.389966	0.254889	0.326748	-0.178126	0.415186	-0.255960	0.041661	
0.010174	-0.041404	0.082320	-0.048078	-0.208971	-0.552605	0.063801	0.125174	-0.034166	0.404933	0.152574	0.360085	-0.027969	-0.475192	

The Chainsmokers - This Feeling	Lubalin - double helix	Ellie Goulding - Close To Me	Aries - BOUNTY HUNTER	Khalid - Better	Avicii - SOS (feat. Aloe Blacc)	Lil Nas X - Late To Da Party (F*CK BET) (feat. YoungBoy Never Broke Again)	Juice WRLD - Real Shit (with benny blanco)	benny blanco - You (with Marshmello & Vance Joy)	Y2K - Lalala	Kid Cudi - Sad People	Rita Ora - Let You Love Me	Halsey - Without Me	Lil Nas X - Old Town Road - Remix
-0.508159	0.146308	-0.127742	-0.166597	-0.224546	0.685949	-0.096275	0.220635	0.069010	-0.105120	0.407415	0.009616	-0.082758	0.169500
-0.229368	-0.118754	-0.080994	-0.352472	0.016400	0.620715	-0.100629	0.251406	-0.061716	-0.277905	0.264036	-0.026806	-0.059766	-0.132636
-0.635312	0.047712	0.285448	0.289241	-0.240730	-0.027554	0.251214	0.056230	-0.193405	-0.285422	-0.277359	0.301750	-0.493399	-0.191602
0.127353	-0.478351	0.291000	-0.053595	-0.123782	-0.478694	-0.402095	-0.219855	0.141492	-0.563612	0.120872	0.582512	-0.234713	-0.182080
-0.199452	-0.040818	-0.124296	0.449276	-0.228220	-0.182084	-0.142494	-0.047045	0.402924	-0.301418	-0.357517	0.657908	-0.302649	-0.002523

Khalid - Talk (feat. Disclosure)	benny blanco - Graduation (with Juice WRLD)	Jaden - BYE	Lauv - There's No Way	Lil Nas X - LIFE AFTER SALEM	Lil Nas X - SCOOP (feat. Doja Cat)	benny blanco - Bad Decisions (with BTS & Snoop Dogg)	Lil Nas X - AM I DREAMING (feat. Miley Cyrus)	benny blanco - Roses (with Juice WRLD feat. Brendon Urie)
-0.319260	-0.287540	-0.584923	0.101926	-0.242673	-0.127553	-0.079991	-0.060652	-0.444026
-0.131852	-0.388064	0.020171	-0.166314	0.158795	-0.419616	-0.145027	-0.076898	-0.358691
-0.538736	-0.046831	-0.349124	-0.402647	-0.273791	-0.230187	0.103824	0.179818	-0.152886
-0.393794	0.149229	0.325356	-0.137932	0.519897	-0.187540	0.139441	-0.215814	0.403432
-0.431253	-0.080438	-0.352943	-0.298322	-0.403094	-0.013696	-0.027174	-0.040543	0.210119

Comparison of 'Lauv - i'm so tired...' with other songs

I then replaced the subset features with 'danceability', 'energy', and 'tempo', followed by 'acousticness', 'instrumentalness', and 'valence', and the algorithm produced better results. Since the features 'duration_ms', 'time_signature', etc. are not really relevant to the content of the song, replacing the subset features gives more accurate results. For instance, when I used 'danceability', 'energy', and 'tempo', the similar songs seemed to share a similar vibe and pace, indicating a more energetic and danceable characteristic. On the other hand, when I used 'acousticness', 'instrumentalness', and 'valence', the similar songs tended to share a similar mood and instrumentation, which could indicate a more acoustic or melancholic feel.

	Lil Nas X - TALES OF DOMINICA	Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)	benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)	Khalid - Love Lies - Snakehips Remix	Lauv - i'm so tired...	Justin Bieber - Lonely (with benny blanco)	Aries - ONE PUNCH	Zara Larsson - Ruin My Life	Tones And I - Dance Monkey	Chelsea Cutler - The Reason	The Chainsmokers - Beach House	Jonas Blue - Rise	Lil Nas X - INDUSTRY BABY (feat. Jack Harlow)
Lil Nas X - TALES OF DOMINICA	1.000000	0.731289	-0.755572	-0.164682	0.064904	-0.624632	0.074973	0.245611	0.550913	-0.541982	-0.150476	0.846917	0.450234
Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)	0.731289	1.000000	-0.179813	-0.654955	0.032259	0.070902	0.613040	-0.369273	0.911941	-0.297115	-0.765116	0.637634	-0.148943
benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)	-0.755572	-0.179813	1.000000	-0.087481	0.299151	0.856367	0.596386	-0.430123	-0.143286	0.787994	-0.326479	-0.433424	-0.923050
Khalid - Love Lies - Snakehips Remix	-0.164682	-0.654955	-0.087481	1.000000	0.604302	-0.567059	-0.312502	0.913897	-0.899290	0.470978	0.917836	0.157572	0.144111
Lauv - i'm so tired...	0.064904	0.032259	0.299151	0.604302	1.000000	-0.163758	0.554996	0.573043	-0.353430	0.787330	0.245483	0.584167	-0.506797
Justin Bieber - Lonely (with benny blanco)	-0.624632	0.070902	0.856367	-0.567059	-0.163758	1.000000	0.525829	-0.833417	0.272973	0.363176	-0.673949	-0.564671	-0.764906
Aries - ONE PUNCH	0.074973	0.613040	0.596386	-0.312502	0.554996	0.525829	1.000000	-0.332299	0.439381	0.549989	-0.663769	0.387993	-0.855056
Zara Larsson - Ruin My Life	0.245611	-0.369273	-0.430123	0.913897	0.573043	-0.833417	-0.332299	1.000000	-0.667125	0.197127	0.862607	0.473427	0.375946
Tones And I - Dance Monkey	0.550913	0.911941	-0.143286	-0.899290	-0.353430	0.272973	0.439381	-0.667125	1.000000	-0.499090	-0.887246	0.288286	-0.068738
Chelsea Cutler - The Reason	-0.541982	-0.297115	0.787994	0.470978	0.787330	0.363176	0.549989	0.197127	-0.499090	1.000000	0.135836	-0.015562	-0.803695
The Chainsmokers - Beach House	-0.150476	-0.765116	-0.326479	0.917836	0.245483	-0.673949	-0.663769	0.862607	-0.887246	0.135836	1.000000	-0.028725	0.475993
Jonas Blue - Rise	0.846917	0.637634	-0.433424	0.157572	0.584167	-0.564671	0.387993	0.473427	0.288286	-0.015562	-0.028725	1.000000	0.068864
Lil Nas X - INDUSTRY BABY (feat. Jack Harlow)	0.450234	-0.148943	-0.923050	0.144111	-0.506797	-0.764906	-0.855056	0.375946	-0.068738	-0.803695	0.475993	0.068864	1.000000
benny blanco - Unlearn (with Gracie Abrams)	-0.939115	-0.702453	0.573161	-0.032653	-0.402463	0.605223	-0.284549	-0.397427	-0.406514	0.221874	0.083047	-0.978030	-0.21712

Result after subset features being replaced with 'danceability', 'energy', and 'tempo'

	Lil Nas X - TALES OF DOMINICA	Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)	benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)	Khalid - Love Lies - Snakehips Remix	Lauv - i'm so tired...	Justin Bieber - Lonely (with benny blanco)	Aries - ONE PUNCH	Zara Larsson - Ruin My Life	Tones And I - Dance Monkey	Chelsea Cutler - The Reason	The Chainsmokers - Beach House	Jonas Blue - Rise	Lil Nas X - INDUSTRY BABY (feat. Jack Harlow
Lil Nas X - TALES OF DOMINICA	1.000000	0.894968	0.633669	-0.880010	-0.765525	0.632315	-0.833263	0.836038	0.994370	-0.698969	-0.601136	0.472059	-0.231939
Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)	0.894968	1.000000	0.282784	-0.578304	-0.649832	0.873430	-0.517671	0.561731	0.934569	-0.307417	-0.669974	0.113494	-0.569439
benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)	0.633669	0.282784	1.000000	-0.823989	-0.255511	-0.193683	-0.761517	0.952024	0.551042	-0.921103	0.133145	0.979698	0.605415
Khalid - Love Lies - Snakehips Remix	-0.880010	-0.578304	-0.823989	1.000000	0.758265	-0.259382	0.988751	-0.905388	-0.830713	0.945593	0.445449	-0.699741	-0.136446
Lauv - i'm so tired...	-0.765525	-0.649832	-0.255511	0.758265	1.000000	-0.655290	0.812746	-0.446159	-0.775565	0.556963	0.910811	-0.067892	0.456244
Justin Bieber - Lonely (with benny blanco)	0.632315	0.873430	-0.193683	-0.259382	-0.655290	1.000000	-0.248919	0.114832	0.710057	0.068708	-0.846506	-0.371769	-0.895621
Aries - ONE PUNCH	-0.833263	-0.517671	-0.761517	0.988751	0.812746	-0.248919	1.000000	-0.835534	-0.784760	0.935602	0.505959	-0.632595	-0.108358
Zara Larsson - Ruin My Life	0.836038	0.561731	0.952024	-0.905388	-0.446159	0.114832	-0.835534	1.000000	0.777112	-0.901678	-0.115087	0.878011	0.335974
Tones And I - Dance Monkey	0.994370	0.934569	0.551042	-0.830713	-0.775565	0.710057	-0.784760	0.777112	1.000000	-0.621796	-0.653237	0.381085	-0.330514
Chelsea Cutler - The Reason	-0.698969	-0.307417	-0.921103	0.945593	0.556963	0.068708	0.935602	-0.901678	-0.621796	1.000000	0.169085	-0.853330	-0.445320
The Chainsmokers - Beach House	-0.601136	-0.669974	0.133145	0.445449	0.910811	-0.846506	0.505959	-0.115087	-0.653237	0.169085	1.000000	0.326020	0.777983
Jonas Blue - Rise	0.472059	0.113494	0.979698	-0.699741	-0.067892	-0.371769	-0.632595	0.878011	0.381085	-0.853330	0.326020	1.000000	0.745133
Lil Nas X - INDUSTRY BABY (feat. Jack Harlow)	-0.231939	-0.569439	0.605415	-0.136446	0.456244	-0.895621	-0.108358	0.335974	-0.330514	-0.445320	0.777983	0.745133	1.000000
benny blanco - Unlearn (with Gracie Abrams)	0.762521	0.938096	-0.013707	-0.424532	-0.730118	0.983345	-0.405622	0.291232	0.826038	-0.108085	-0.850808	-0.200711	-0.802900
Zedd - Happy Now	0.950530	0.984031	0.374721	-0.702572	-0.758067	0.836987	-0.657070	0.639657	0.978036	-0.446685	-0.722656	0.195539	-0.508335

Result after subset features being replaced with 'acousticness', 'instrumentalness', and 'valence'

Obviously, after the less correlated variables being excluded and similar variables for comparison being extracted, the similarity distinction between songs became more apparent.

Task 3 - Top N Recommendations

Following the exploration of song similarity, I used the cosine similarity to generate song recommendations based on a seed track. The recommendations were largely accurate with lower values of n, showing songs that had closely matching features with the seed track. However, as the value of n increased, the relevance of the recommendations started to fade. This was likely due to the fact that as we moved further from the seed track in the feature space, the selected songs started to diverge in terms of their characteristics.

```
In [35]:  track = "Khalid - Better"

In [52]:  #How many tracks to recommend
          n = 10

In [53]:  similarities.sort_values(by = track, ascending=False)[track].index[1:n+1]

Out[53]:  Index(['Billie Eilish - bad guy (with Justin Bieber)',
                 '5 Seconds of Summer - Youngblood', 'Lil Nas X - LIFE AFTER SALEM',
                 'benny blanco - You (with Marshmello & Vance Joy)',
                 'Aries - BOUNTY HUNTER', 'Khalid - OTW', 'Chelsea Cutler - The Reason',
                 'Khalid - Talk (feat. Disclosure)',
                 'Khalid - Love Lies - Snakehips Remix',
                 'Justin Bieber - Lonely (with benny blanco)'],
                dtype='object')

In [56]:  track = "Khalid - Better"

In [59]:  #How many tracks to recommend
          n = 20

In [60]:  similarities.sort_values(by = track, ascending=False)[track].index[1:n+1]

Out[60]:  Index(['benny blanco - I Can't Get Enough (benny blanco, Selena Gomez, J Balvin, Tainy)',
                 'Aries - BOUNTY HUNTER', 'The Chainsmokers - This Feeling',
                 'Lil Nas X - AM I DREAMING (feat. Miley Cyrus)', 'Aries - ONE PUNCH',
                 'Justin Bieber - Lonely (with benny blanco)',
                 'Chelsea Cutler - The Reason',
                 'benny blanco - You (with Marshmello & Vance Joy)',
                 'benny blanco - Eastside (with Halsey & Khalid) - Acoustic',
                 'Camila Cabello - Real Friends', 'Khalid - OTW', 'Zedd - Happy Now',
                 'Jaden - BYE', 'Juice WRLD - Real Shit (with benny blanco)',
                 'Lubalin - double helix', 'Lauv - i'm so tired...',
                 'Rita Ora - Let You Love Me',
                 'benny blanco - Unlearn (with Gracie Abrams)',
                 'Olivia O'Brien - RIP (feat. G-Eazy & Drew Love)',
                 'Lil Nas X - LIFE AFTER SALEM'],
                dtype='object')
```

Results when the subset features 'danceability', 'energy', and 'tempo' being used(track selected from 'Khalid - Better')

```
In [35]:  track = "Khalid - Better"

In [52]:  #How many tracks to recommend
          n = 10

In [53]:  similarities.sort_values(by = track, ascending=False)[track].index[1:n+1]

Out[53]:  Index(['Billie Eilish - bad guy (with Justin Bieber)',
                 '5 Seconds of Summer - Youngblood', 'Lil Nas X - LIFE AFTER SALEM',
                 'benny blanco - You (with Marshmello & Vance Joy)',
                 'Aries - BOUNTY HUNTER', 'Khalid - OTW', 'Chelsea Cutler - The Reason',
                 'Khalid - Talk (feat. Disclosure)',
                 'Khalid - Love Lies - Snakehips Remix',
                 'Justin Bieber - Lonely (with benny blanco)'],
                dtype='object')

In [35]:  track = "Khalid - Better"

In [44]:  #How many tracks to recommend
          n = 20

In [45]:  similarities.sort_values(by = track, ascending=False)[track].index[1:n+1]

Out[45]:  Index(['Billie Eilish - bad guy (with Justin Bieber)',
                 '5 Seconds of Summer - Youngblood', 'Lil Nas X - LIFE AFTER SALEM',
                 'benny blanco - You (with Marshmello & Vance Joy)',
                 'Aries - BOUNTY HUNTER', 'Khalid - OTW', 'Chelsea Cutler - The Reason',
                 'Khalid - Talk (feat. Disclosure)',
                 'Khalid - Love Lies - Snakehips Remix',
                 'Justin Bieber - Lonely (with benny blanco)',
                 'Avicii - SOS (feat. Aloe Blacc)',
                 'Lil Nas X - AM I DREAMING (feat. Miley Cyrus)',
                 'benny blanco - Unlearn (with Gracie Abrams)',
                 'The Chainsmokers - This Feeling', 'Aries - ONE PUNCH', 'Jaden - BYE',
                 'Kid Cudi - Sad People', 'Zedd - The Middle', 'Lauv - There's No Way',
                 'Ellie Goulding - Close To Me'],
                dtype='object')
```

Results when the subset features 'acousticness', 'instrumentalness' and 'valence' being used(track selected from 'Khalid - Better')

I also experimented with different subsets of features for generating recommendations. I found

that the choice of features significantly affected the quality of recommendations. For example, when I used more rhythm-related features like 'danceability', 'energy', and 'tempo', the recommended songs seemed to match the seed track's overall vibe and pace. Conversely, when I used more mood-related features like 'acousticness', 'instrumentalness', and 'valence', the recommended songs seemed to match the seed track's overall mood and feel.

When I used the subset features 'acousticness', 'instrumentalness' and 'validity', and n=10, the last track is 'Justin Bieber - Lonely (with benny blanco)'. By this point there was already a gap between the vibe and pace of 'Khalid - Better'. When n=20, the gap became even greater for the songs that followed.