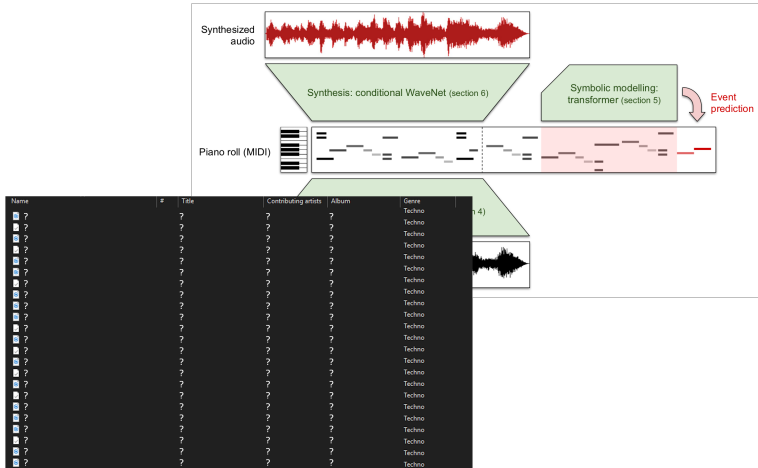


## Background and Implementation





## Execution and Relevance

```
total_elements = yearly_releases['pagination']['items']
if total_elements > 100000:
    max_range = 100000
else:
    max_range = (total_elements / 100) * 100

# New array to transform retrieved data in
data_sort = np.empty((max_range, 11), dtype=np.dtype('U500'))
```

```
print 'Currently processing Year', year
```

```
for j in range(0, max_range, 100):
```

```
    for release in yearly_releases['results']:
        # Data Calculations
        title_release = release.get('title', 'unknown')
        title_release_split = title_release.split('-', 2)
        title = title_release_split[0]
        delimited_title = title.replace(' ', '_')
        release_name = title_release_split[1]
        delimited_release = release_name.replace(' ', '_')
        community_input = release.get('community', 'unknown')
        have = community_input['have']
        want = community_input['want']
        have_want_mean = (have + want) / 2
```

```
    # Fill in array
    data_sort[i + j, 0] = release.get('id')
    data_sort[i + j, 1] = delimited_title
    data_sort[i + j, 2] = delimited_release
    data_sort[i + j, 3] = '%.1f' % (release.get('style'))
    data_sort[i + j, 4] = release.get('year')
    data_sort[i + j, 5] = (release.get('country')).replace(' ', '_')
    data_sort[i + j, 6] = ('%.1f' % (release.get('label'))).replace(' ', '_')
    data_sort[i + j, 7] = ('%.1f' % (release.get('format'))).replace(' ', '_')
    data_sort[i + j, 8] = have
    data_sort[i + j, 9] = want
    data_sort[i + j, 10] = format(have_want_mean, "09")
    i = i + 1
```

```
print 'Year:', year, ' | Number of entries processed:', j, ' | Number
```

```
of entries processed:', j, ' | Number
```

```
of entries processed:', j, ' | Number
```

```
of entries processed:', j, ' | Number
```

```
of entries processed:', j, ' | Number
```

```
of entries processed:', j, ' | Number
```

```
of entries processed:', j, ' | Number
```

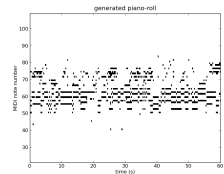
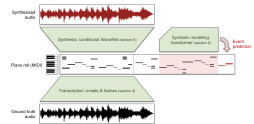
```
of entries processed:', j, ' | Number
```

#	A	B	C	D	E	F	G	H	I	J	K
1	ID	Artist	Release Title	Style	Year	Country	Label	Format	Have	Want	Mean
2	13474305	Orbital	Buried Deeper	W Techno	2019	UK; Europe	ACP Recordings	AC Vinyl; 12"; 45	371	52	211
3	13803451	Barker*	BARKER 001	Techno	2019	Germany	Ostgut Ton; Scape	W Vinyl; 12"; Lim	182	219	200
4	13142871	Planetary Assault	Straight Shooting	Techno	2019	UK	Mote-Evolver	Vinyl; 12"	195	168	181
5	13178981	AUROD	Voltage EP	Techno	2019	Belgium	Lenske Records	Vinyl; 12"	209	121	165
6	13249290	AUROD	Universe Remix	E Techno	2019	France	Molek&W; Curve P	Vinyl; 12"; EP	145	177	161
7	13669663	Rikhter	RIK1	Techno	2019	Germany	R - Label Group; Hai	Vinyl; 12"; 33	143	176	159
8	13295785	Dark Comedy	Corbomite Manu	Techno	2019	UK	Mint Condition; Lar	Vinyl; 12"; EP	115	175	145
9	13212807	MDA Analog	Shine EP	Techno	2019	Unknown	MDA LABS	Vinyl; 12"; Rel	123	134	128
10	13045749	Shvachxsh	WOD'D EP	Techno	2019	Netherlands	Mord; Mord; Mord;	Vinyl; 12"; 33	162	80	125
11	13920892	Z@P	Brutalismo	Techno	2019	Germany	My Own Jupiter	Vinyl; 10"; Sin	106	145	125
12	13580912	Paula Temple	Edge Of Everything	Techno	2019	Netherlands	Noise Manifesto; W	Vinyl; 12"; Alp	137	106	121
13	13641415	Yak (19)	Termina EP	Techno	2019	UK	R & S Records	Vinyl; 12"	110	124	117
14	14238549	Blawan	Many Many Pings	Techno	2019	UK	Ternesc	Vinyl; 12"	131	102	116
15	13915901	Nina Kraviz	Stranno Neobatr	Techno	2019	Russia	Lenske Records	Vinyl; 12"; EP	164	65	114
16	13860569	Farrago (3)	Neontrance	Techno	2019	Belgium	Lenske Records	Vinyl; 12"	177	49	113
17	14121636	Amelie Lens	Little Robot	Techno	2019	Belgium	Lenske Records	Vinyl; 12"; EP	173	49	111
18	13430457	Corp	Time Out	Techno	2019	Italy	Spaziotempo	Vinyl; 12"; 33	123	100	111
19	14003054	Aurora Hala	Liquiditty	Techno	2019	US	Mutual Dreaming R	Vinyl; 12"	107	109	108
20	13354120	Koboski	RK4	Techno	2019	Germany	R - Label Group; Hai	Vinyl; 12"; 33	97	118	107
21	14000170	Ben Klock	Subzero / Comey	Techno	2019	Germany	Ostgut Ton; Copyrig	Vinyl; 12"; 45	164	49	106
22	13695405	Jeff Mills	If Remixes	Techno	2019	US	Purpose Maker; Coj	Vinyl; 12"; 45	149	64	106
23	13762703	Jeff Mills	The Director's Cu	Techno	2019	US	Axis; Axis Records	Vinyl; 12"; Cor	163	49	106
24	14014972	Anna* & Kittin'	Speicher 112	Techno	2019	Germany	Kompakt Extra; Spe	Vinyl; 12"; 45	141	67	104
25	14115405	Dominik Eulber	Mannigfaltig	Techno	2019	Germany	IK7 Records; K7 Mu	Vinyl; 12"; Alp	173	36	104
26	13951554	Sleeparchive	Revised Records	Techno	2019	Germany	Tresor; Tresor Reco	Vinyl; 12"; 33	145	60	102
27	1381432	Maceo Plex	Mutant Tour 2015	Techno	2019	UK	Ellum Audio	Vinyl; 12"; LP; Com	106	98	102
28	13780930	Die Orangen	Soft 2	Techno	2019	Israel	Malka Tut; Malka T	Vinyl; 12"	98	103	100
29	13956902	Milo Spikers	Observable 93	Techno	2019	Belgium	Lenske Records	Vinyl; 12"	144	57	100
30	13476873	TV-OUT	TV-OUT	Techno	2019	US	L.I.E.S. Records	Vinyl; 12"	105	93	99
31	13579861	Mike Parker	Early Works	Techno	2019	Germany	Overlife; The Excha	Vinyl; 12"; 33	116	90	98
32	13683853	Surgeon	Raw Trax 1	Techno	2019	UK	Dynamic Tension R	Vinyl; 12"; EP	136	57	96
33	13135986	Radical G & The RR2	13135986	Techno	2019	Germany	R - Label Group; Hai	Vinyl; 12"; 33	79	115	94
34	13111917	Glyn Hendry	Escape Club 99	Techno	2019	UK	Poly Kicks	Vinyl; 12"; 45	93	95	94
35	13651248	Various	Essential Memon	Techno	2019	Netherlands	Emerald; Hardgroo	Vinyl; 12"; 33	132	55	93
36	14012424	Stef Mendesi	Klockworks 26 & C	Techno	2019	Germany	Klockworks	Vinyl; 12"	133	54	93



My thesis will be looking into the development of a techno data set that can be used for the purposes of machine learning to produce a creative artificial intelligence program. As there is currently no techno-based data set available, I will need to construct one.

However before I create the data set, I need to identify what songs I want to use for the data set. Because techno music has evolved over the past 30 years, I want a data set which encompasses the most popular techno songs stretching over those 30 years.





The intention of this proof of concept is to be able to use Discog's API to extract the top 100 techno songs based on the mean of an entries 'have' and 'want' values for every year between 1989 - 2019, before exporting to a csv; 'have' referring to the number of uses who state they own the release compared to those who 'want' the release.

The script is intended to clean the data set before exporting, including removing duplicates, fixing nulls, and parsing unwanted commas.



The developed python script executes the following:

1. Authenticates with the API
2. Creates a dataset array with relevant headings
3. Starts a for loop from 2019 to 1989
4. Extracts the top 10000 entries by relevance and places them into an array
5. Removes entries which contain more than 1 style
6. Removes duplicates
7. Sorts by mean and concatenates the top 100 entries to a dataset array

Once this has completed, the dataset array is exported as a csv file

```
total_elements = yearly_releases['pagination']['items']
if total_elements > 10000:
    max_range = 10000
else:
    max_range = (total_elements / 100) * 100

# New array to transform retrieved data in
data_sort = np.empty((max_range, 11), dtype=np.dtype('U50'))

print 'Currently processing Year', year

for j in range(0, max_range, 100):
    for release in yearly_releases['results']:
        # Data Calculations
        title_release = release.get('title', 'unknown')
        title_release_split = title_release.split('-', 2)
        title = title_release_split[0]
        delimited_title = title.replace(' ', ' ')
        release_name = title_release_split[1]
        delimited_release = release_name.replace(' ', ' ')
        community_input = release.get('community', 'unknown')
        have = community_input['have']
        want = community_input['want']
        have_want_mean = (have + want) / 2

        # Fill in array
        data_sort[i + j, 0] = release.get('id')
        data_sort[i + j, 1] = delimited_title
        data_sort[i + j, 2] = delimited_release
        data_sort[i + j, 3] = ';' + '.join(release.get('style'))
        data_sort[i + j, 4] = release.get('year')
        data_sort[i + j, 5] = (release.get('country')).replace
        data_sort[i + j, 6] = (';' + '.join(release.get('label'))
        data_sort[i + j, 7] = (';' + '.join(release.get('format'))
        data_sort[i + j, 8] = have
        data_sort[i + j, 9] = want
        data_sort[i + j, 10] = format(have_want_mean, "05")
        i = i + 1

    print 'Year:', year, ' | Number of entries processed: ', j

# Safety net for request rate limit
end_timer = time.time()
elapsed_time = end_timer - start_timer
if elapsed_time < 1:
    sleep(1.1 - elapsed_time)
```



## Discogs' API's Limitations:

- ▶ Discogs catalogs releases, not individual tracks
- ▶ Discogs will only retrieve a maximum of 10,000 entries per search request
- ▶ Each individual request can only retrieve a page of 100 entries
- ▶ There is a rate limit of 60 requests per minute, resulting in a 50 minute run time with a timer running as a safeguard
- ▶ Discogs will produce numerous duplication of entries, which are removed prior to sorting
- ▶ When retrieving 'Techno' releases, it will include multi-style releases which contain 'techno' as one of their styles - these are removed before added to the dataset



	A	B	C	D	E	F	G	H	I	J	K
1	ID	Artist	Release_Title	Style	Year	Country	Label	Format	Have	Want	Mean
2	13474305	Orbital	Buried Deeper Within / Impa	Techno	2019	UK; Europe	ACP Recordings; ACP Recordi	Vinyl; 12"; 45 F	371	52	211
3	13803451	Barker*	BARKER 001	Techno	2019	Germany	Ostgut Ton; Scape Mastering	Vinyl; 12"; Lim	182	219	200
4	13142871	Planetary Assault System	Straight Shooting	Techno	2019	UK	Mote-Evolver	Vinyl; 12"	195	168	181
5	13178981	AIROD	Voltage EP	Techno	2019	Belgium	Lenke Records	Vinyl; 12"	209	121	165
6	13249290	AIROD	Universe Remix EP	Techno	2019	France	Molek&ki; Curve Pusher	Vinyl; 12"; EP	145	177	161
7	13669663	Rikhter	RIK1	Techno	2019	Germany	R - Label Group; Hard Wax; Gl	Vinyl; 12"; 33	143	176	159
8	13295785	Dark Comedy	Corbomite Manuever EP	Techno	2019	UK	Mint Condition; Lark Music; A	Vinyl; 12"; EP;	115	175	145
9	13212807	MDA Analog	Shine EP	Techno	2019	Unknown	MDA LABS	Vinyl; 12"; Rel	123	134	128
10	13045749	Shxcxchxsh	WOD T EP	Techno	2019	Netherlands	Mord; Mord; Mord; The Exche	Vinyl; 12"; 33	162	88	125
11	13920892	Z@P	Brutalismo	Techno	2019	Germany	My Own Jupiter	Vinyl; 10"; Sing	106	145	125
12	13580912	Paula Temple	Edge Of Everything	Techno	2019	Netherlands	Noise Manifesto; Wordandsc	Vinyl; LP; Albu	137	106	121
13	13641415	Yak (19)	Termina EP	Techno	2019	UK	R & S Records	Vinyl; 12"	110	124	117
14	14238549	Blawan	Many Many Pings	Techno	2019	UK	Ternesc	Vinyl; 12"	131	102	116
15	13915901	Nina Kraviz	Stranno Neoblatno	Techno	2019	Russia	Lenke Records	Vinyl; 12"; EP	164	65	114
16	13860569	Farrago (3)	Neontrance	Techno	2019	Belgium	Lenke Records	Vinyl; 12"	177	49	113
17	14121636	Amelie Lens	Little Robot	Techno	2019	Belgium	Lenke Records	Vinyl; 12"; EP	173	49	111
18	13430457	Corp	Time Out	Techno	2019	Italy	Spaziotempo	Vinyl; 12"; 33	123	100	111
19	14003054	Aurora Halal	Liquiditty	Techno	2019	US	Mutual Dreaming Recordings	Vinyl; 12"	107	109	108
20	13354120	Kobosil	RK4	Techno	2019	Germany	R - Label Group; Hard Wax; Gl	Vinyl; 12"; 33	97	118	107
21	1400178	Ben Klock	Subzero / Coney Island	Techno	2019	Germany	Ostgut Ton; Copyright Contro	Vinyl; 12"; 45 F	164	49	106
22	13695405	Jeff Mills	If Remixes	Techno	2019	US	Purpose Maker; Copyright Co	Vinyl; 12"; 45 F	149	64	106
23	13762703	Jeff Mills	The Director's Cut Chapter 3	Techno	2019	US	Axis; Axis Records; Axis Reco	Vinyl; 12"; Cor	163	49	106
24	14014972	Anna* & Kittin*	Speicher 112	Techno	2019	Germany	Kompakt Extra; Speicher; Kor	Vinyl; 12"; 45 F	141	67	104
25	14115405	Dominik Eulberg	Mannigfaltig	Techno	2019	Germany	IK7 Records; K7 Music GmbH;	Vinyl; LP; Albu	173	36	104
26	13551554	Sleeparchive	Revised Recordings EP	Techno	2019	Germany	Tresor; Tresor Records; Treso	Vinyl; 12"; 33	145	60	102
27	13281432	Maceo Plex	Mutant Tour 2019	Techno	2019	UK	Ellum Audio	Vinyl; LP; Com	106	98	102

The results are cleaned as the program runs, and automatically exports the data set to a csv file. The resulting file has a column for the headings following by 3000 entries sorted by its mean value and by year.



## In Summary:

- ▶ This proof of concept produces a data set of meta data containing the top 100 techno releases from 1989 - 2019
- ▶ The relevance of this is it can be used to quantitatively choose the contents of a data set when developing a set for the purposes of machine learning
- ▶ The proof of concept is implemented in python using Discogs' API
- ▶ The script extracts and transforms data year by year, cleaning as it goes. The top 100 tracks are sliced from the rest of the sorted set and concatenated to the dataset array
- ▶ After execution, the dataset array is exported to a csv file
- ▶ The average run time is roughly 50 minutes due to the API's request limit.