



# Clustering 2020

## Songtexte

Lassen sich Musikgenres anhand  
ihrer Texte klassifizieren?

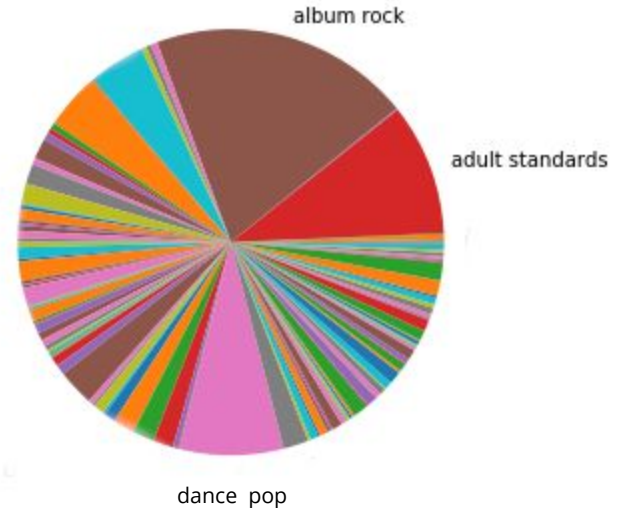
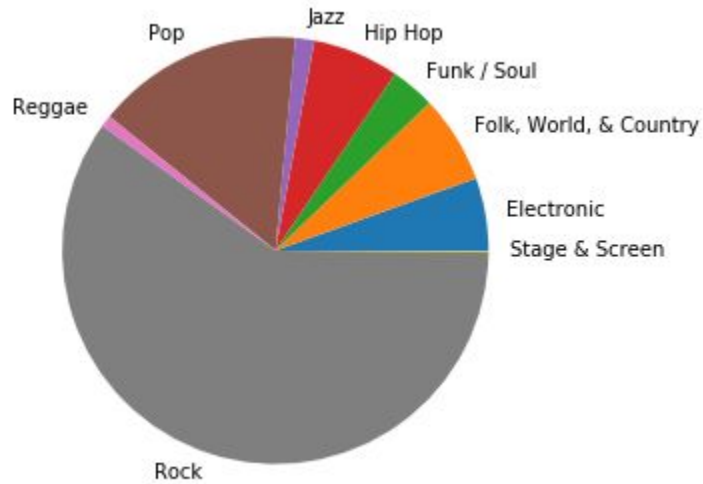


# Veränderung des Datensatzes

	artist	song	link	text
0	ABBA	Ahe's My Kind Of Girl	/a/abba/ahes+my+kind+of+girl_20598417.html	Look at her face, it's a wonderful face \nAnd...
1	ABBA	Andante, Andante	/a/abba/andante+andante_20002708.html	Take it easy with me, please \nTouch me gentl...
2	ABBA	As Good As New	/a/abba/as+good+as+new_20003033.html	I'll never know why I had to go \nWhy I had t...
3	ABBA	Bang	/a/abba/bang_20598415.html	Making somebody happy is a question of give an...
4	ABBA	Bang-A-Boomerang	/a/abba/bang+a+boomerang_20002668.html	Making somebody happy is a question of give an...

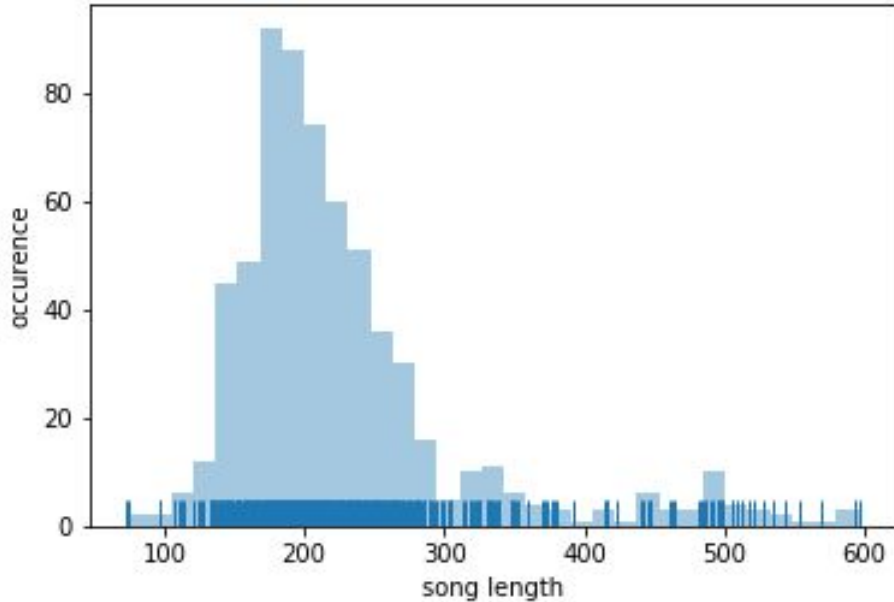
	artist	decades	text	stemmed_text	POS	Genre1	genre1	number_of_tokens	number_of_types
0	"Weird Al" Yankovic	1980s	potato skins potato cakes hash browns and i...	potato skin potato cake hash brown and instan...	french plain enough greasy sour lumpy tough en...	Rock	antiviral pop	12745.0	2264.0
1	"Weird Al" Yankovic	1990s	as i walk through the valley where i harvest m...	as i walk through the valley where i harvest ...	plain perfect fancy amish amish hard amish ami...	Rock	antiviral pop	10803.0	2317.0
2	"Weird Al" Yankovic	2000s	jillian was her name she was sweeter than asp...	jillian wa her name she wa sweeter than aspar...	same gorgeous perfect close sweet young ripe p...	Rock	antiviral pop	4352.0	1159.0
3	"Weird Al" Yankovic	2010s	my mama told me when i was hatched act like a...	my mama told me when i wa hatch act like a su...	red hard sure swiss crazy little okay crazy ev...	Rock	antiviral pop	2077.0	705.0
4	*NSYNC	1990s	here we go here we go one more time eve...	here we go here we go one more time everybodi...	fine fine free right fine right fine free time...	Electronic	boy band	1786.0	345.0

# Verteilung der Genres

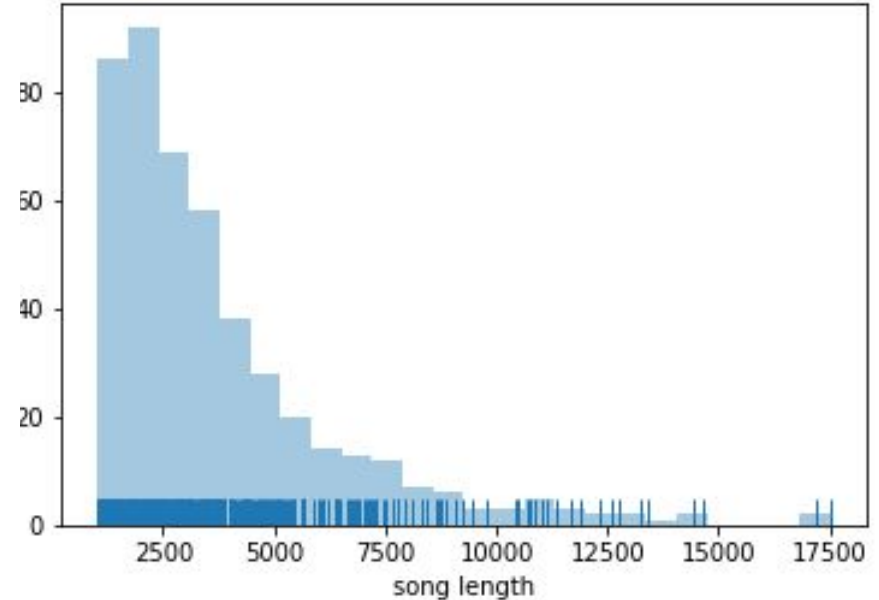


# Songlängen Verteilung

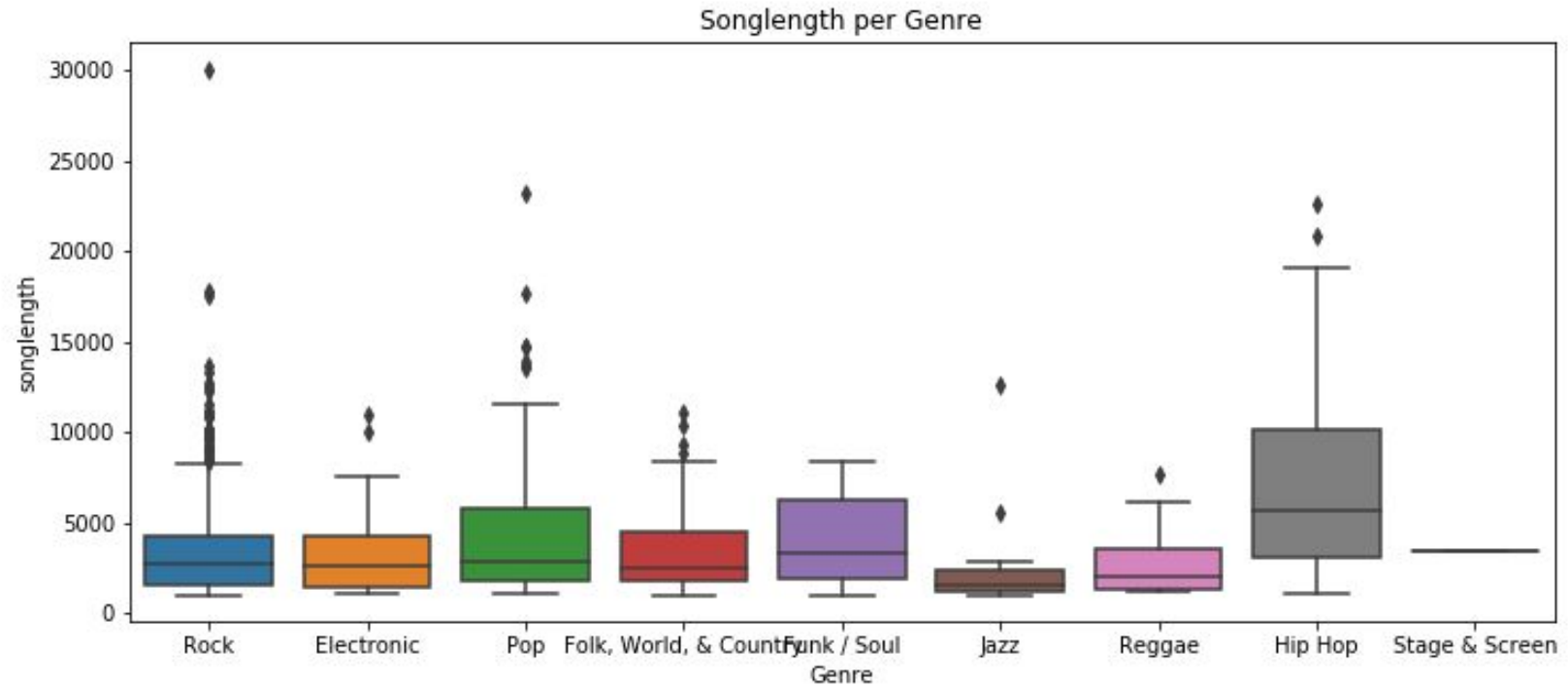
Distribution of Songlength



Distribution of Songlength

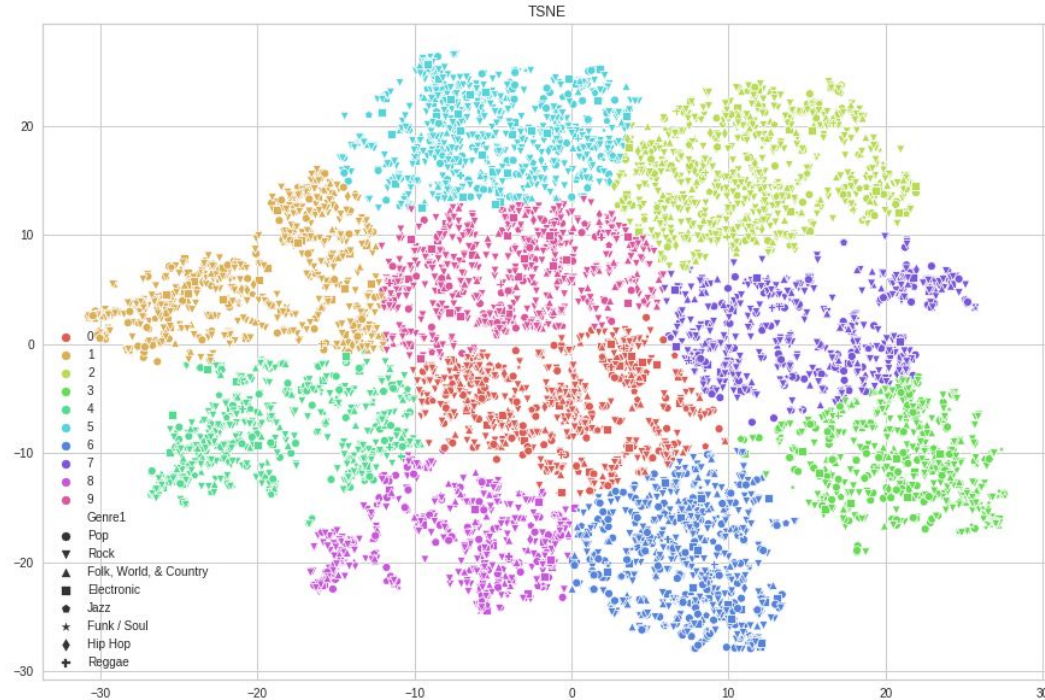


# Verteilung nach Genre



# Clustering Verfahren

# K-Means mit allen Genres

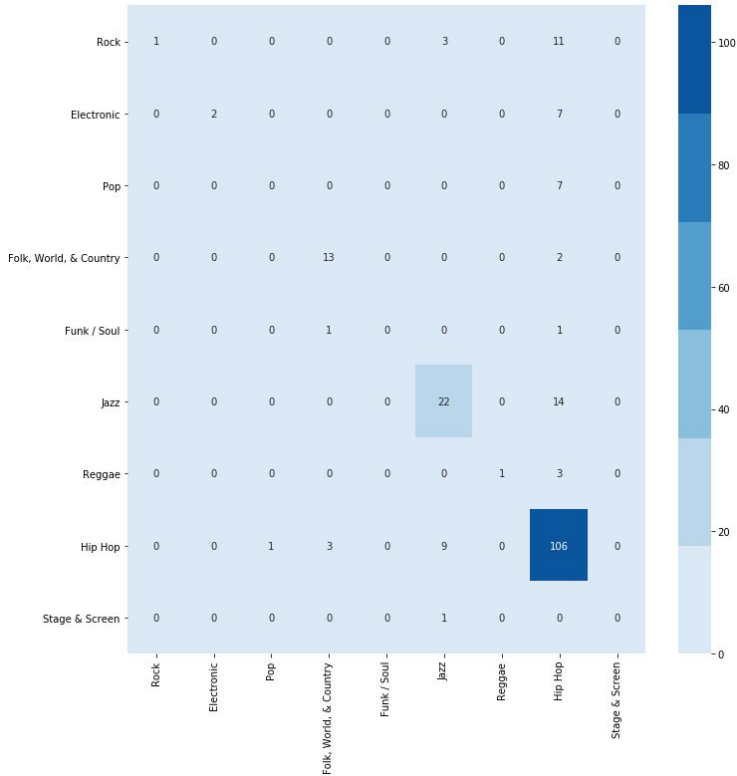






# SVM

# SVM zur Auswahl der besten Genres



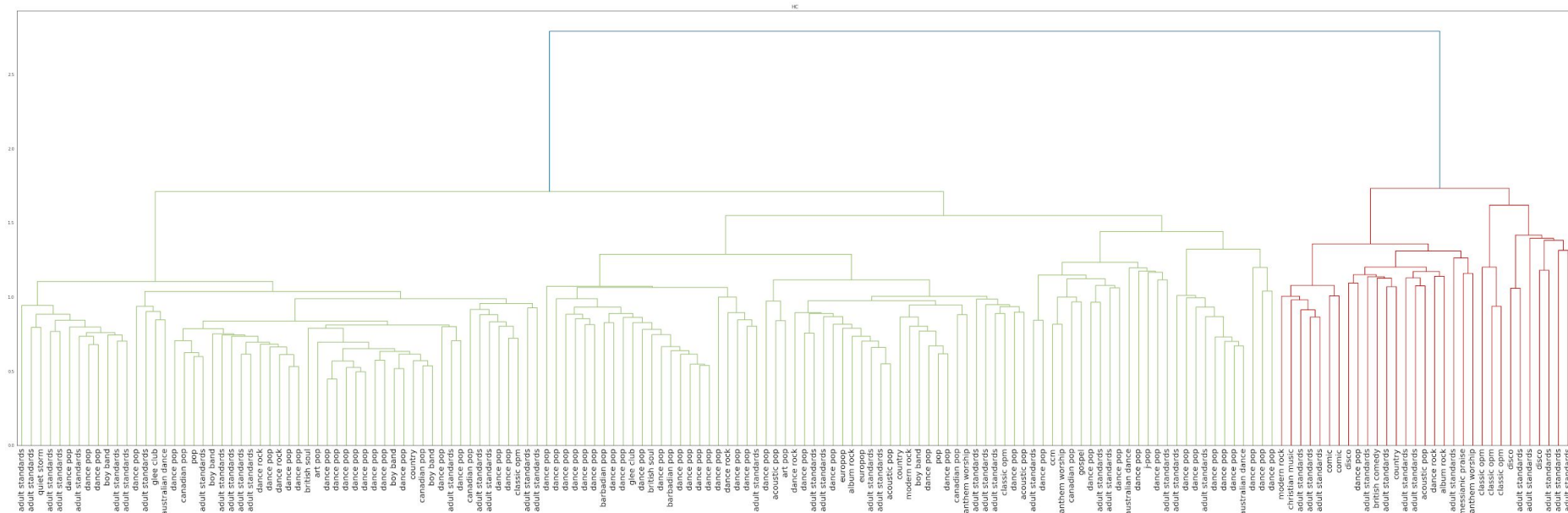
	precision	recall	f1-score	support
Rock	1.00	0.07	0.12	15
Electronic	1.00	0.22	0.36	9
Pop	0.00	0.00	0.00	7
Folk, World, & Country	0.76	0.87	0.81	15
Funk / Soul	0.00	0.00	0.00	2
Jazz	0.63	0.61	0.62	36
Reggae	1.00	0.25	0.40	4
Hip Hop	0.70	0.89	0.79	119
Stage & Screen	0.00	0.00	0.00	1
accuracy			0.70	208
macro avg	0.57	0.32	0.35	208
weighted avg	0.70	0.70	0.65	208



# Hierarchisches Clustering und Gaussian Mixture Model



# HC mit Untergenres von Pop



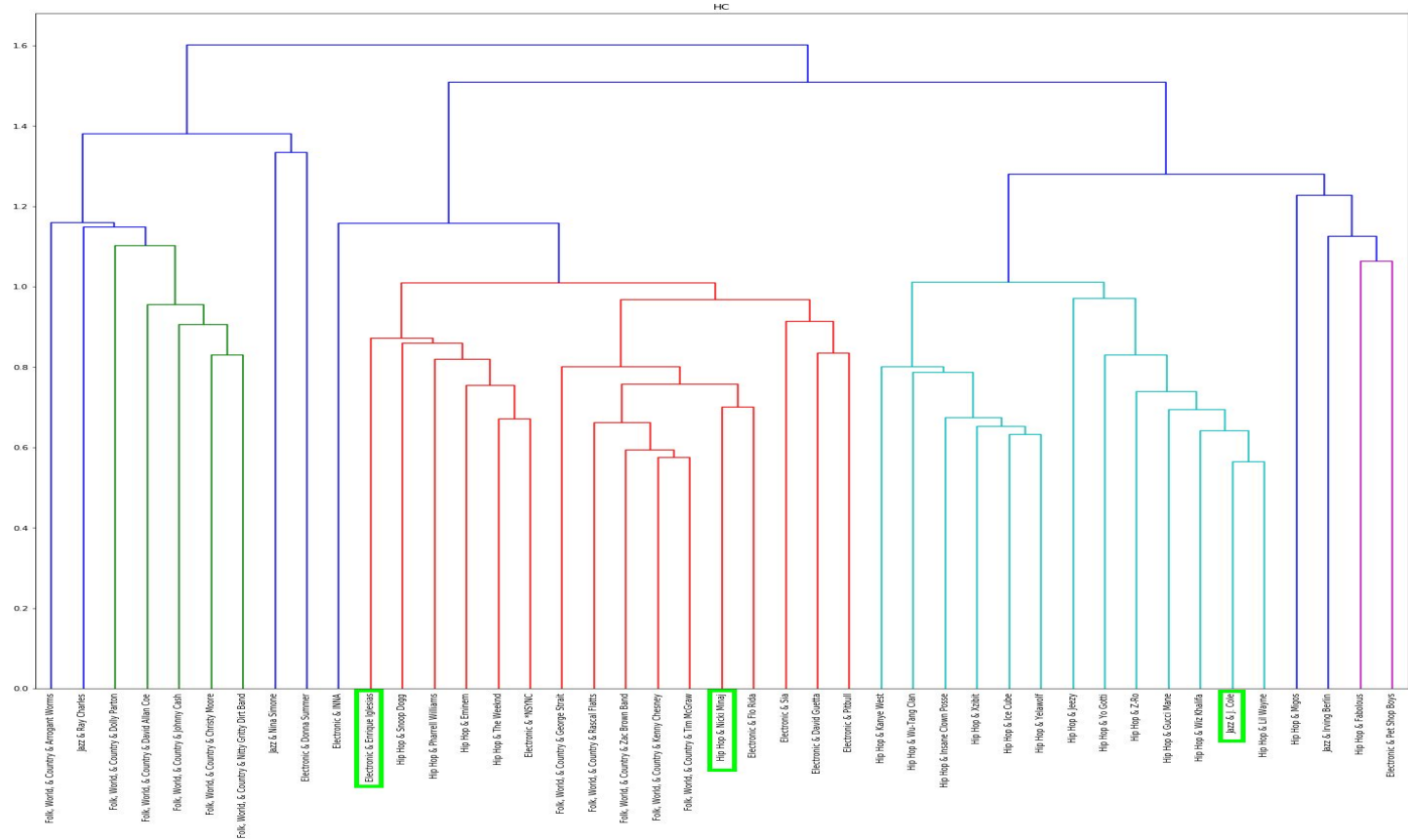
# Pop-Subgenres: adult standard vs. dance pop



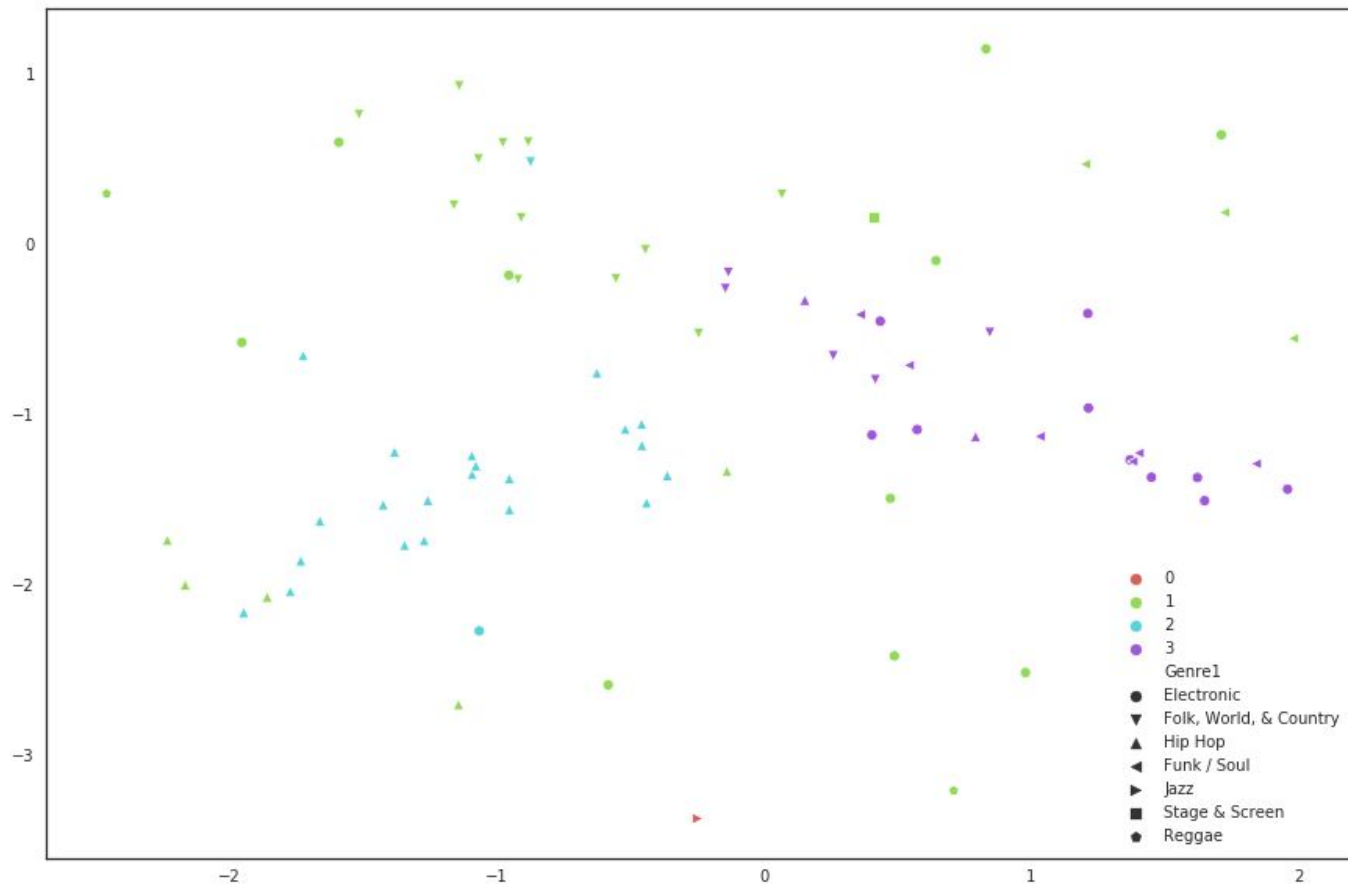
## HC



# HC mit 4 Genres in den 2010ern



# Gaussian Mixture Model





# Gaussian Mixture Model

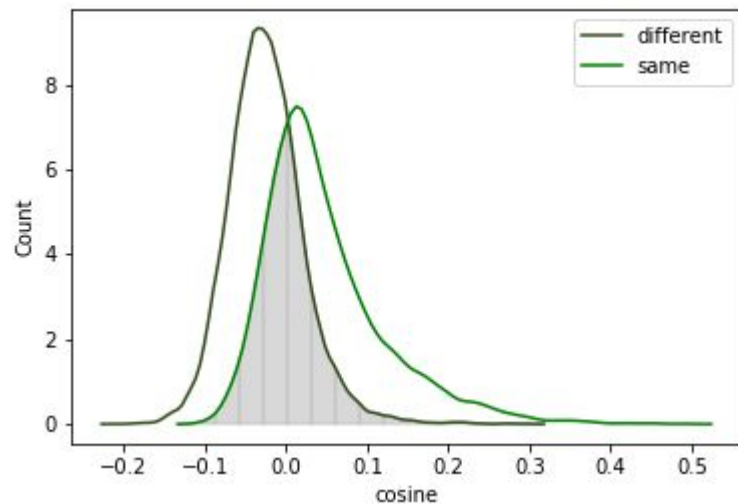
artist	genre	decade	0	1	2	3
Ms. Lauryn Hill	Hip Hop	2000s	0.0	3.265538e-302	1.000000e+00	0.000000e+00
Nine Inch Nails	Electronic	2000s	0.0	9.129822e-24	1.000000e+00	0.000000e+00
Reba McEntire	Folk, World, & Country	2000s	0.0	1.819207e-110	1.000000e+00	0.000000e+00
Culture Club	Electronic	2000s	0.0	0.000000e+00	3.973914e-230	1.000000e+00
Hank Williams	Folk, World, & Country	2000s	0.0	0.000000e+00	1.994558e-218	1.000000e+00
Insane Clown Posse	Hip Hop	2000s	0.0	0.000000e+00	7.066848e-280	1.000000e+00
Q-Tip	Hip Hop	2000s	0.0	0.000000e+00	1.780805e-214	1.000000e+00
Tim McGraw	Folk, World, & Country	2000s	0.0	1.000000e+00	4.653551e-288	0.000000e+00
David Guetta	Electronic	2000s	0.0	0.000000e+00	1.000000e+00	3.953938e-288
Dolly Parton	Folk, World, & Country	2000s	0.0	0.000000e+00	1.000000e+00	8.576813e-188
Gloria Gaynor	Funk / Soul	2000s	0.0	1.000000e+00	0.000000e+00	4.116459e-63
Gucci Mane	Hip Hop	2000s	1.0	0.000000e+00	0.000000e+00	7.261716e-167



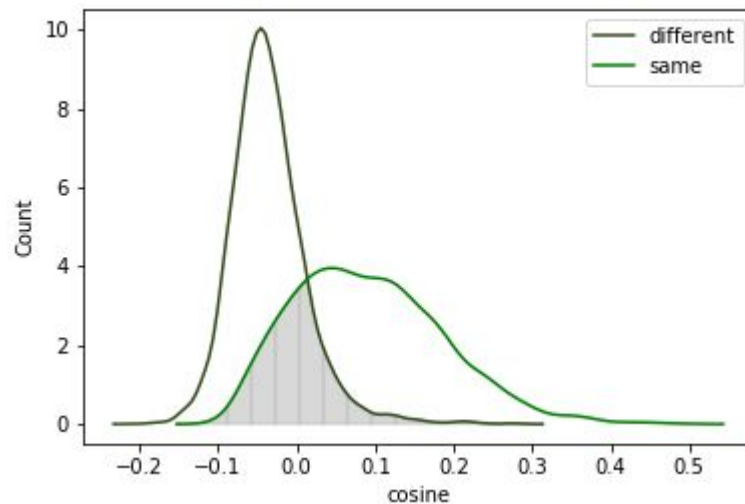
# Delta

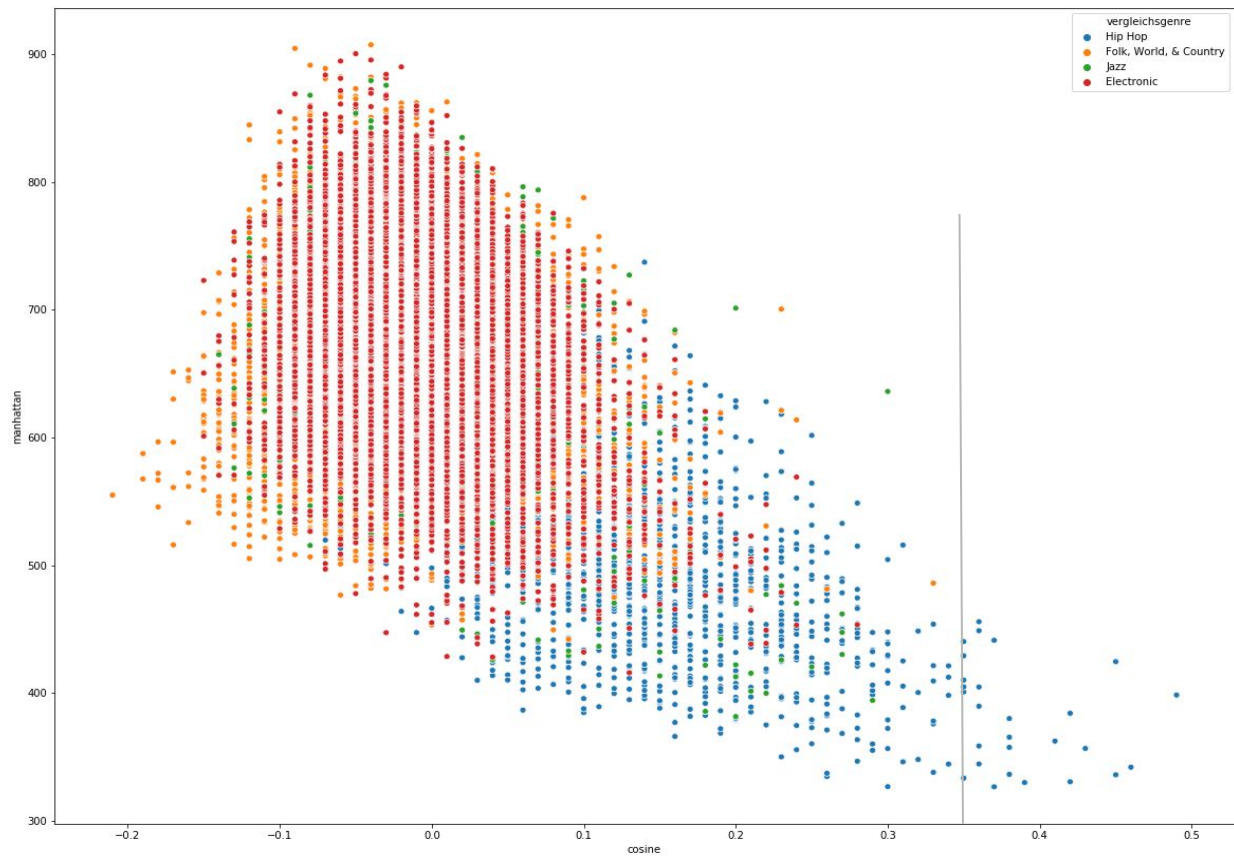
# Delta mit vier Genres

alle Songs



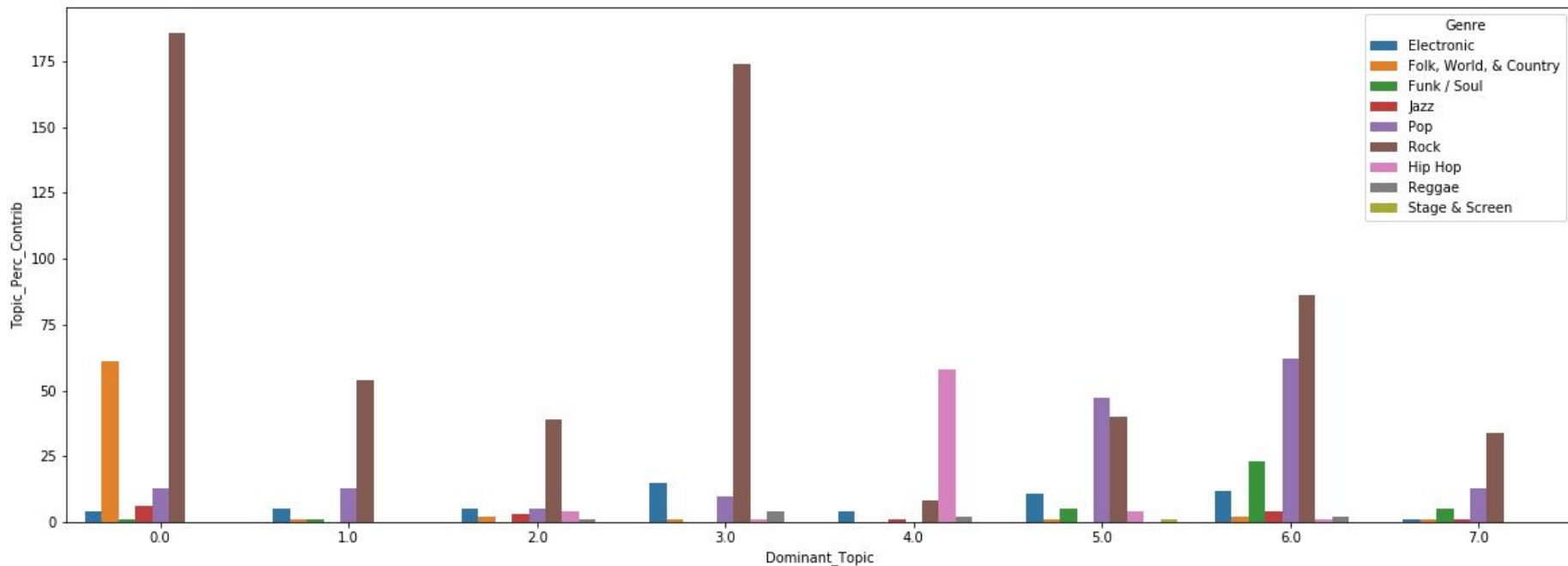
nur Hip Hop





# Topic Modelling

# Genre-Verteilung pro Topic



# Keywords pro Topic für Hip Hop und Country

shit ya ass  
bitch money  
fuck nigga yo em  
niggas

road went  
blue money  
woman  
rain saw  
sweet  
lord

# Topic Modelling mit Sklearn

```
Topic 01: know, time, want, way, life, need, heart, everything, tonight, nothing  
Topic 02: nigga, shit, bitch, ai, fuck, ass, money, niggas, wit, get  
Topic 03: baby, girl, want, ai, little, know, love, money, good, tonight  
Topic 04: love, heart, life, need, know, feel, time, lonely, night, say  
Topic 05: man, old, little, ai, good, woman, town, time, big, night  
Topic 06: life, world, blood, time, heart, death, soul, light, way, end
```



# Ergebnisse

# These 1: Lassen sich Musikgenres mittels Texte unterscheiden?

Antwort: Lyrics der Lieder ähneln sich sehr, zeigen keine klaren Cluster

- Welche Genres liegen enger zusammen?

Antwort: Pop und Rock → Pop ist aber intern gut differenzierbar, Rock nicht

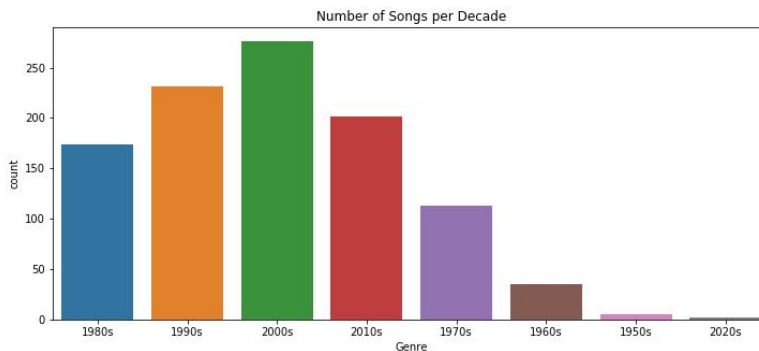
- Gibt es Genres, die trotzdem gute Ergebnisse zeigen?

Antwort: Ja, vor allem Hip Hop (und Country) lässt sich in allen Verfahren am besten identifizieren

# These 1: Lassen sich Musikgenres mittels Texte unterscheiden?



# These 2: Werden die Songtexte im Laufe der Jahre “einfacher”?



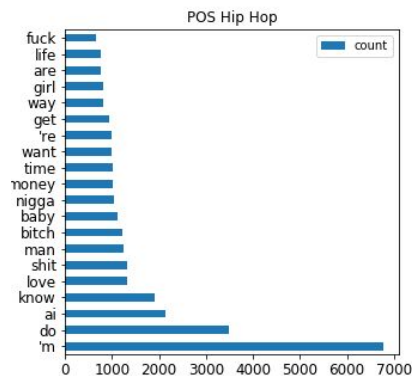
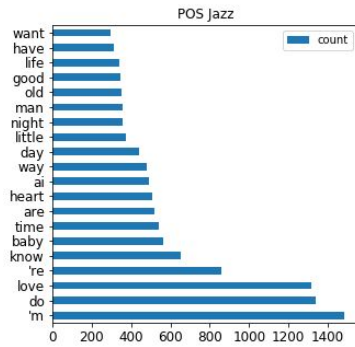
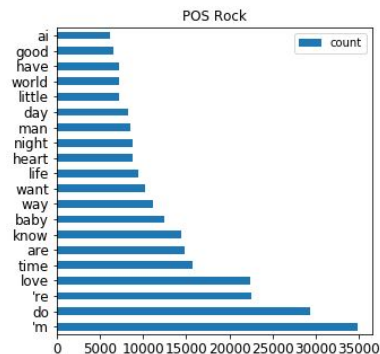
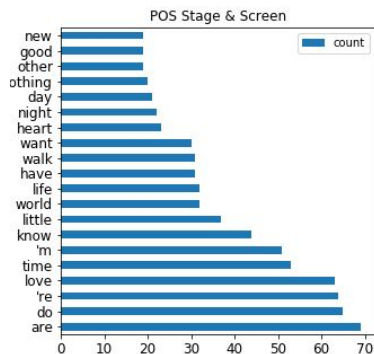
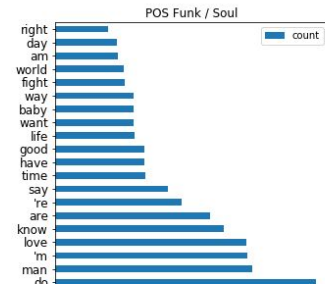
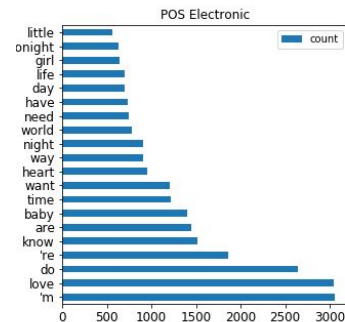
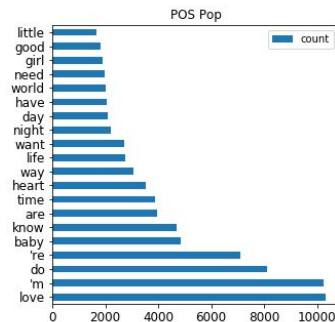
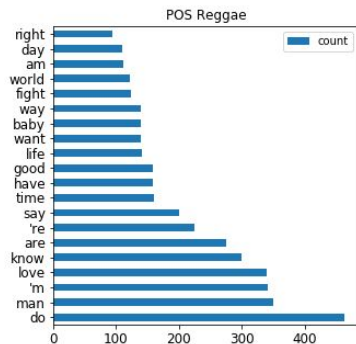
# Fazit

# Fazit:

- beste Genres: Hip Hop und Country
- beste Verfahren : HC + t-SNE + Text als Feature und Topic Modelling
- Zukunft:
  - Länge der Songtexte anreichern
  - außerdem evtl zusätzliche Features wie Melodie, Instrumente ...
  - ausgeglichenerer Datensatz

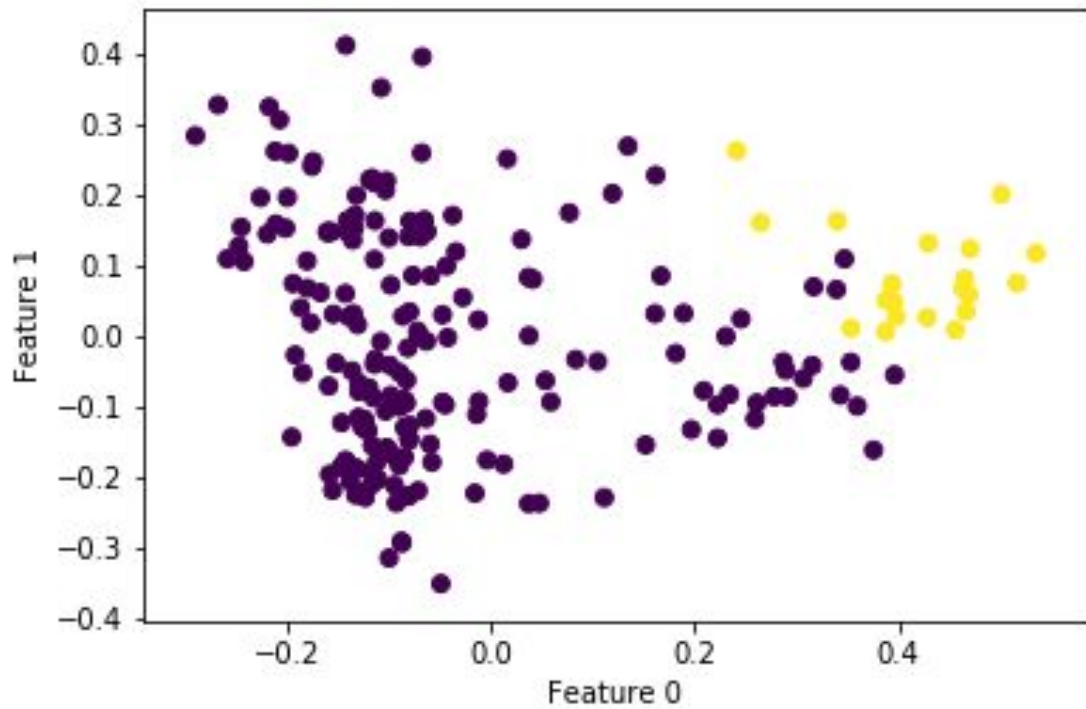
# Backup-Folien

# Genres - most common words





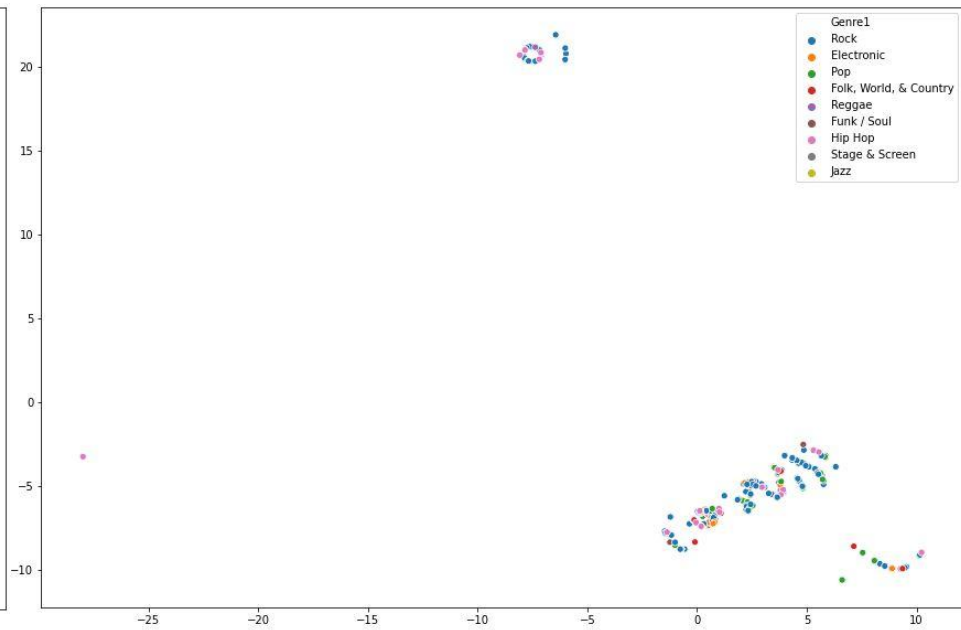
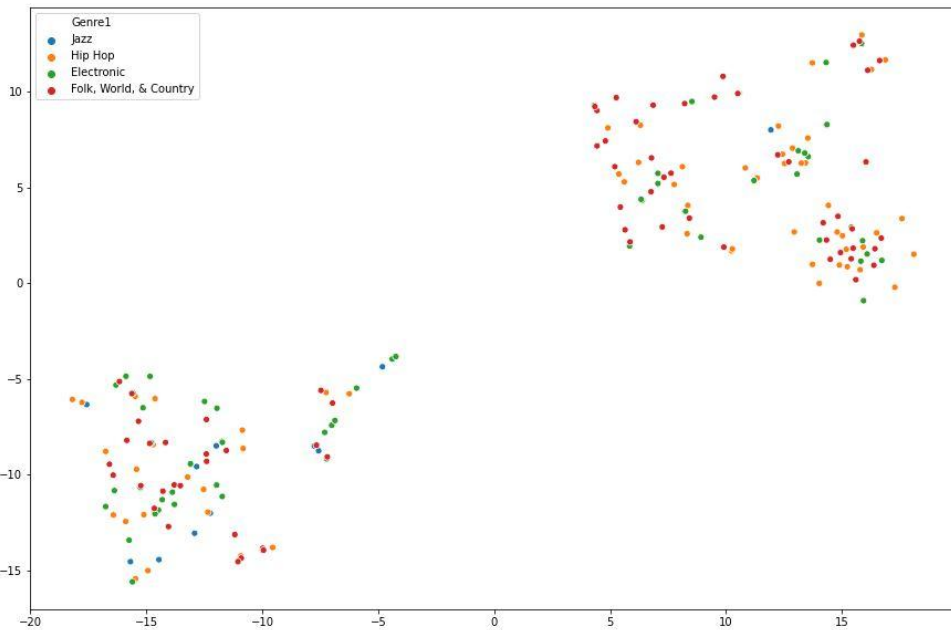
# DBscan



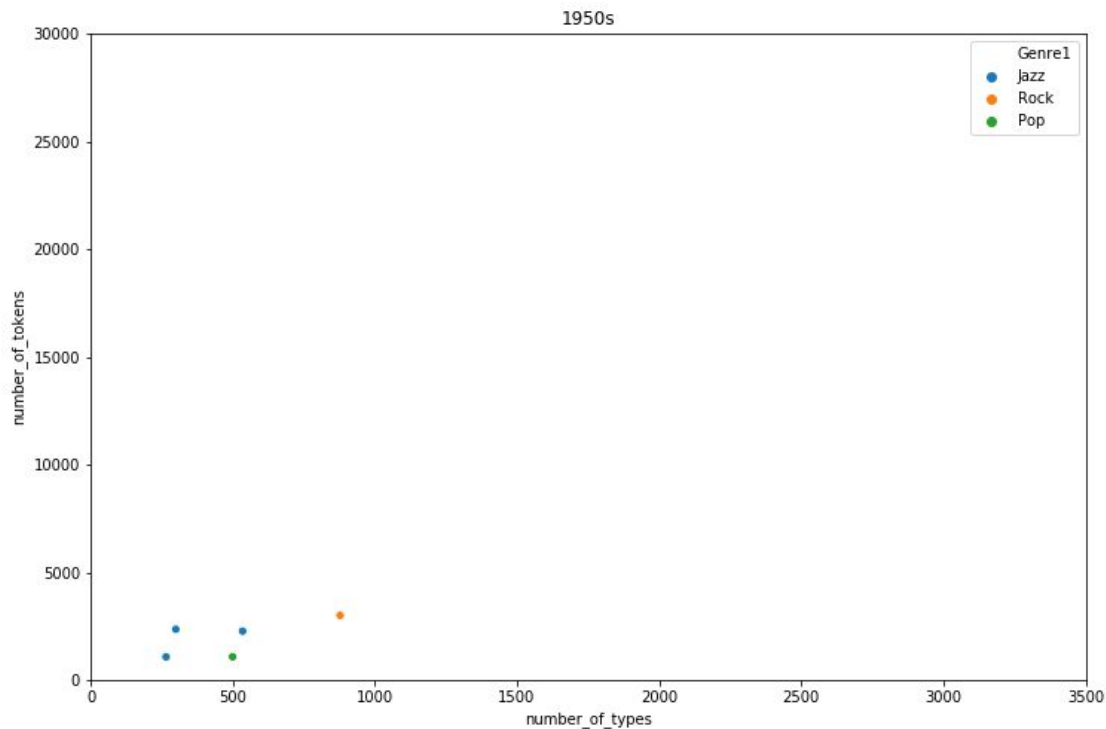
# Topic Modelling mit MALLET

```
0      0,04175 heavens rza champagne betcha fell clique cowards skinny mutha wang model quavo busting wig titty turnin biz slice hug skills
1      0,24597 love baby tonight feel girl give ooh i'm party heart don't night yeah you're stop heaven fucking wanna hey she's
2      0,03742 titties struggle threw chuck passin countin tony lemons nying hitting jokers dissed baking moms snake ova grade lovers scared tuye
3      0,31161 talkin bout nigga lay ain't shit niggas dream lord liza woman diddle bitch jack drop odl hoes break jane york
4      0,43531 i'm nigga fuck ain't don't shit bitch niggas money yeah make man ass that's it's back y'all wit real hit
5      0,54542 lookin south death kids queen kinda work full balls clean mom sold rocks underground sir control favorite mary dig guys
6      4,51323 i'm don't it's love time back you're can't baby i'll make gonna life yeah day that's good man night ain't
7      0,56041 i've home she's i'd there's lord man country christmas heaven road song big ain't lonely river town lovin long he'd
```

# Autoencoder



# Types-Tokens-Verteilung



Viktoria Ermisch:	1986364	Referatsnote
Timo Günther:	2033581	Hausarbeitsnote
Julia Jäger:	2124649	Referatsnote
Teresa Kaiser:	2353056	Hausarbeitsnote