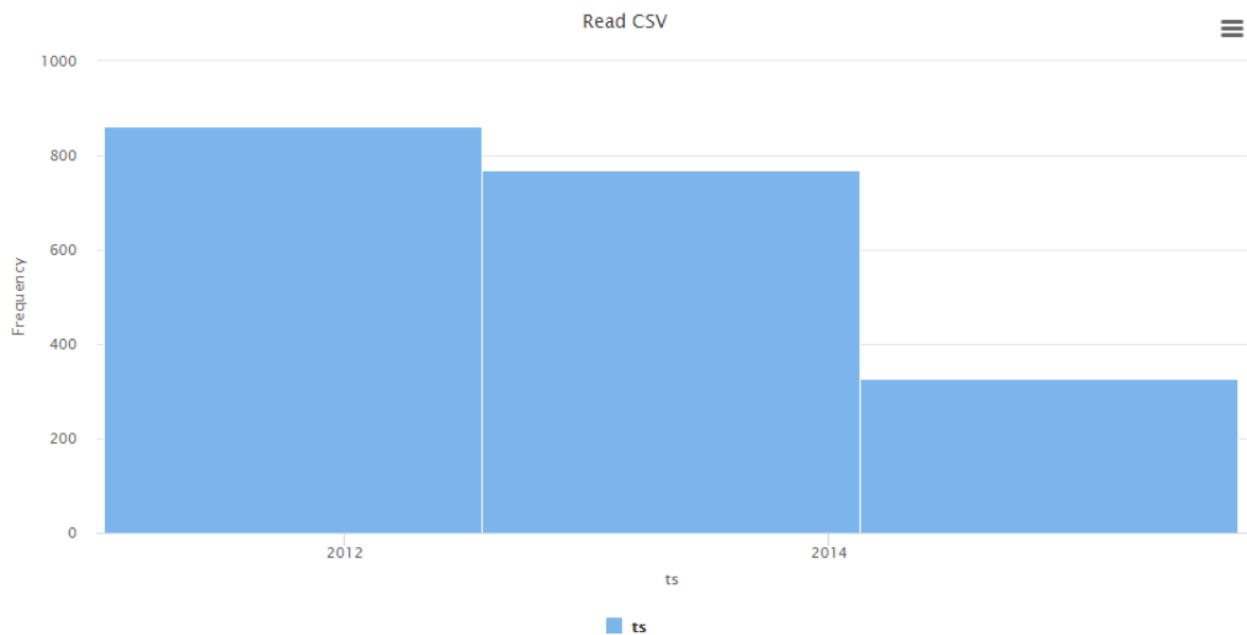
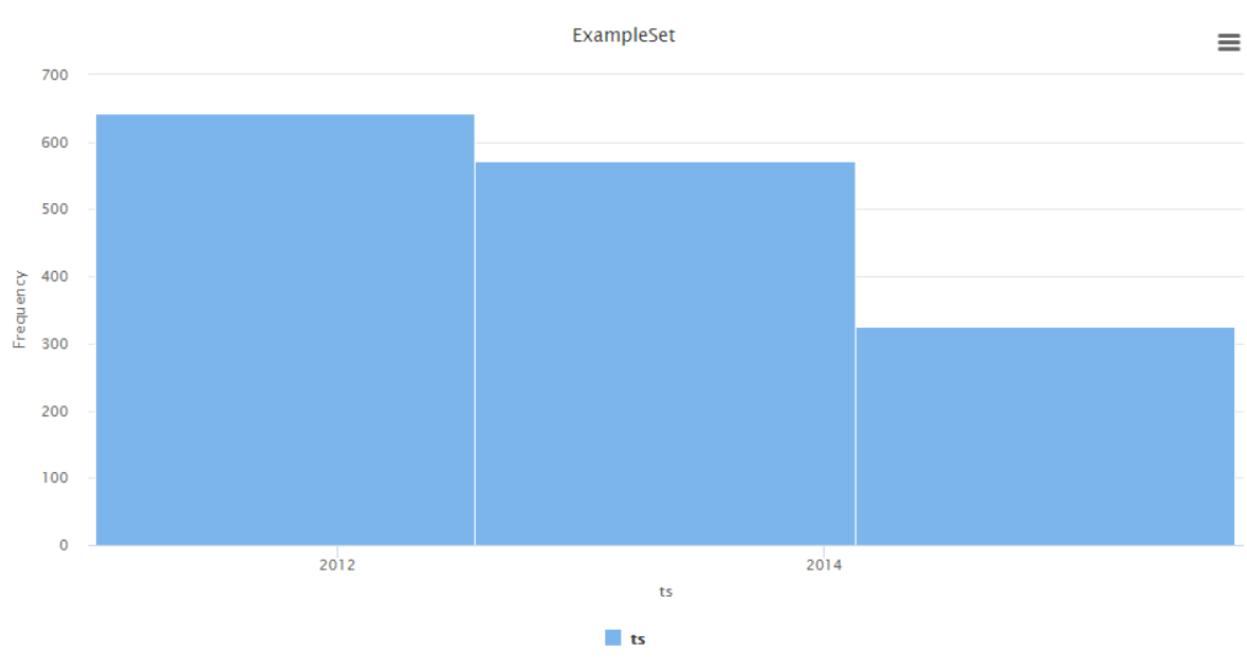


FASE 1: ESPOLORAZIONE DEL DATA SET

Encoding UTF-8 per eliminare simboli strani. Abbiamo 3 intervalli di tempo.



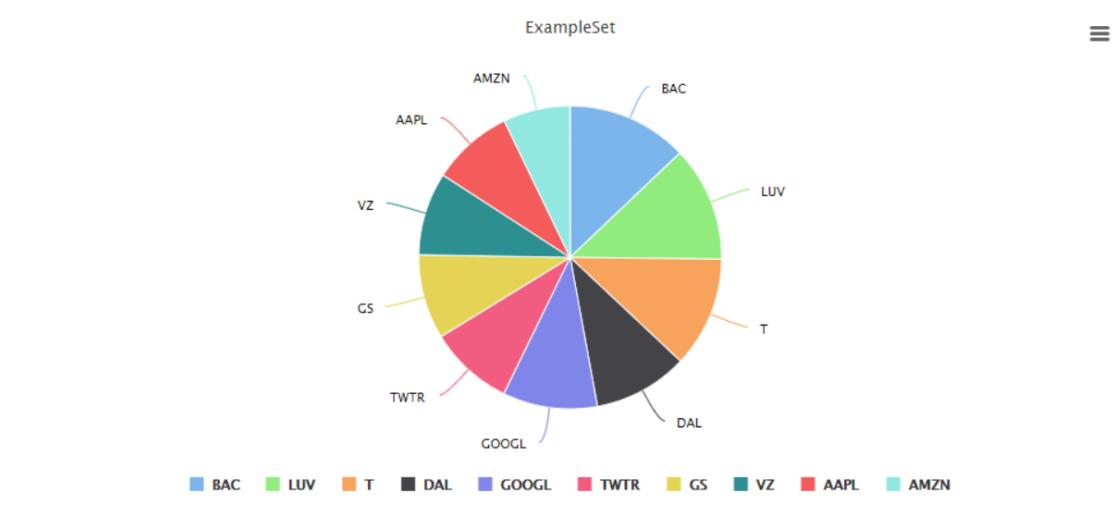
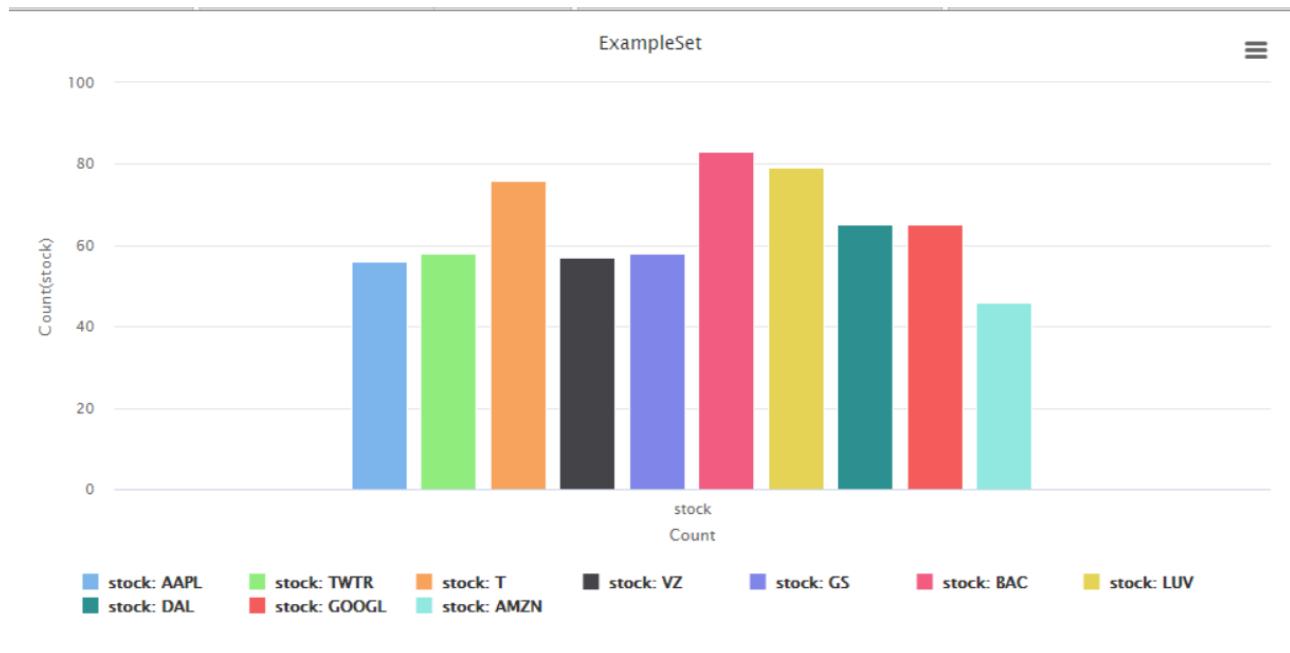
Così abbiamo tolto i dati relativi ai mesi di ottobre novembre e dicembre per gli anni 2011 e 2013 e abbiamo reso più omogeneo il nostro dataset, da un punto di vista quantitativo.



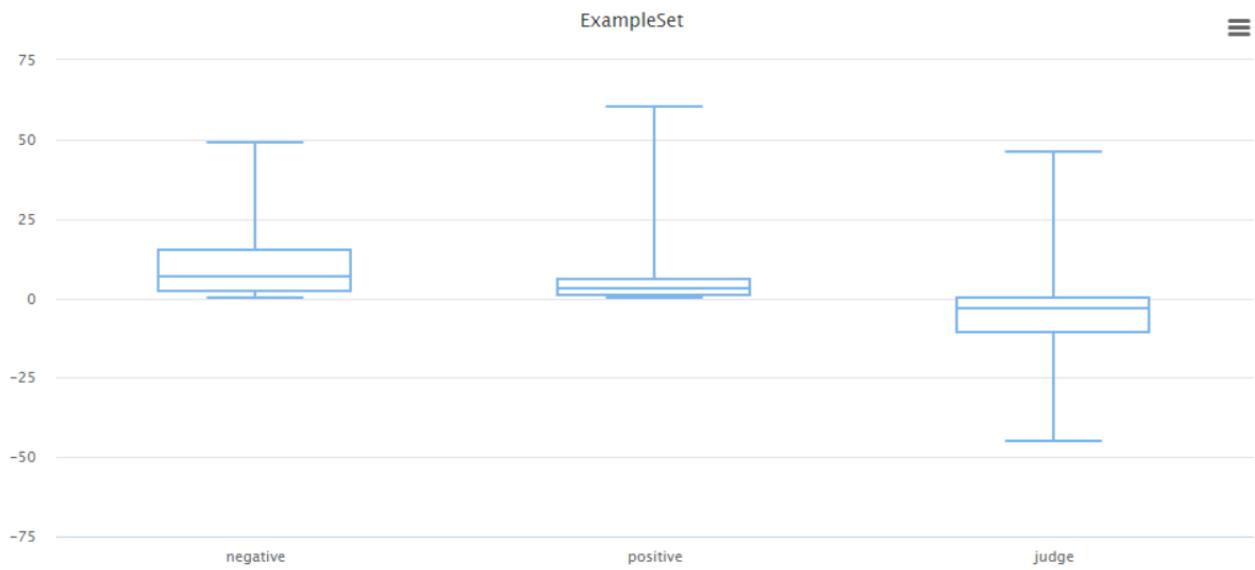
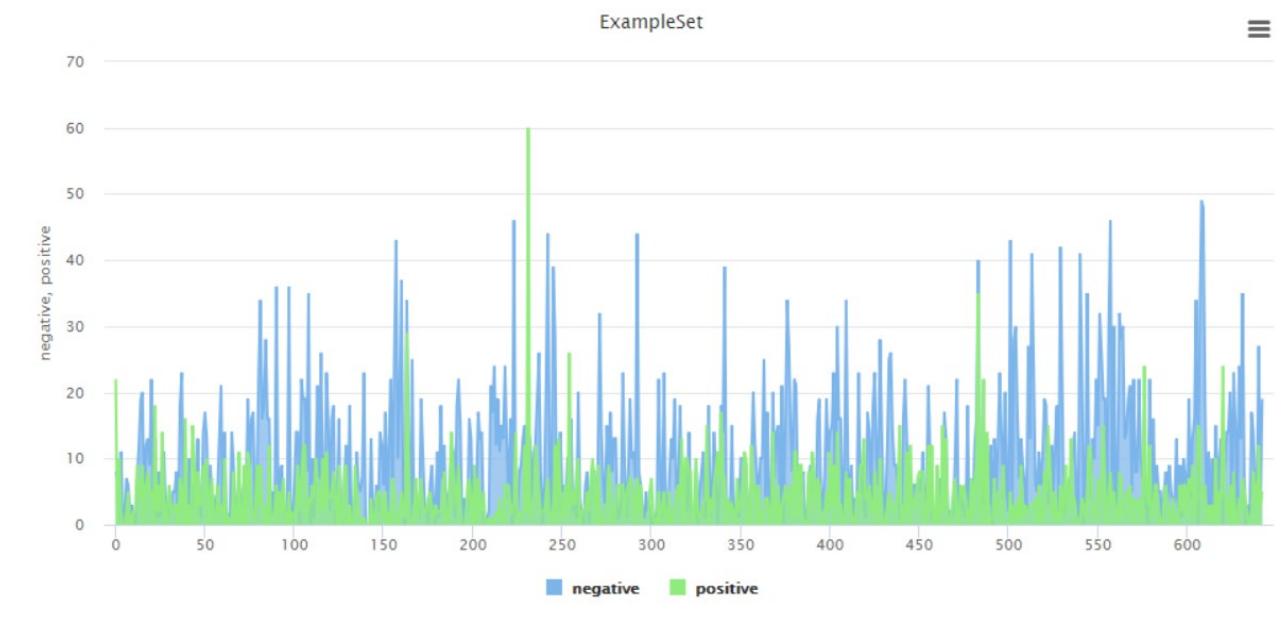
Diamo una breve occhiata a come sono disposte gli stock e i giudizi sulle varie frasi inerenti alle azioni che abbiamo a disposizione.

Anno 2011

Possiamo osservare come sono stati presi i vari stock. Innanzitutto notiamo che la distribuzione degli stock per l'anno 2011 è abbastanza uniforme, con prevalenza di osservazioni di BAC e LUV:

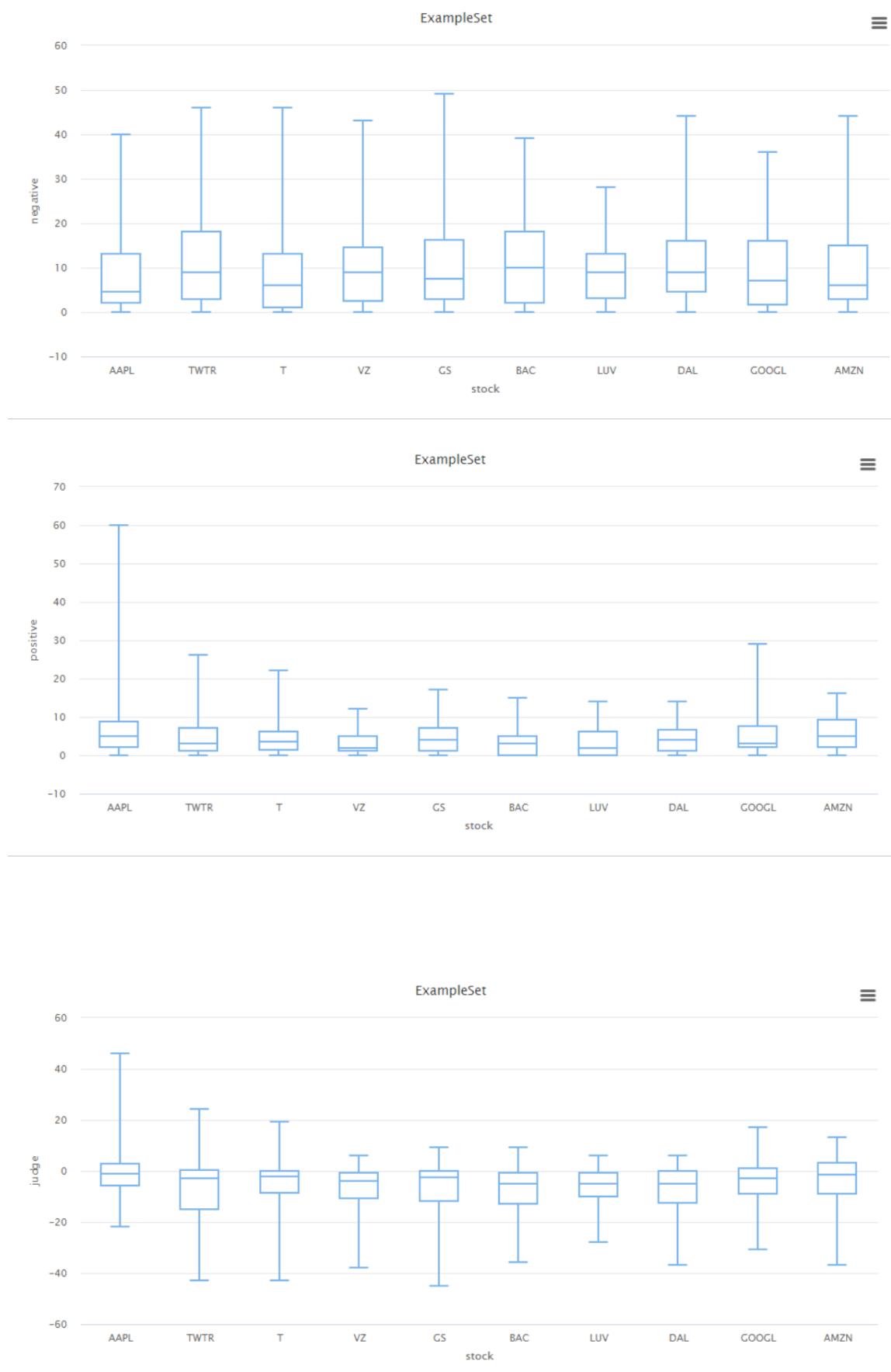


Guardiamo il valore dei giudizi negativi e positivi



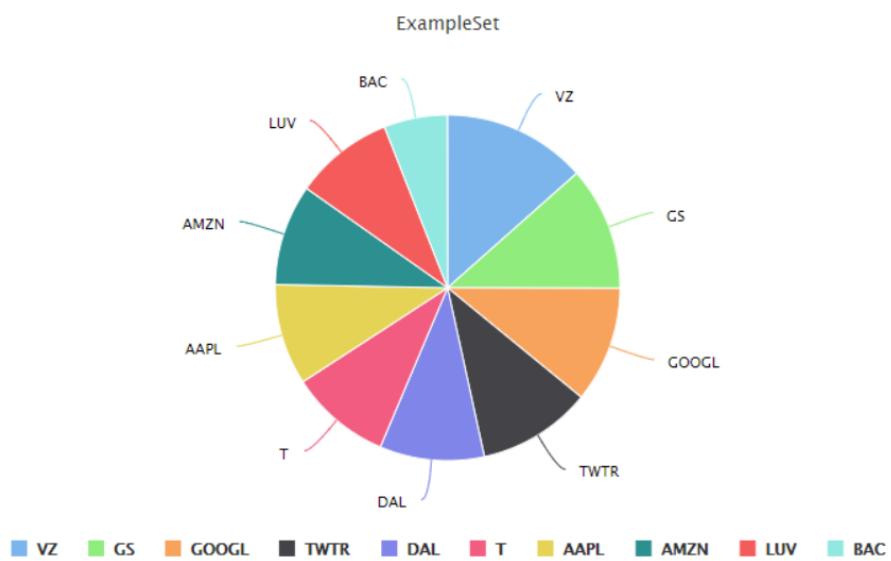
Row No.	average(judge)	sum(judge)	sum(positive)	sum(negative)
1	-5.719	-3677	2825	6502

Collegamento tra stock, positive, negative e judge

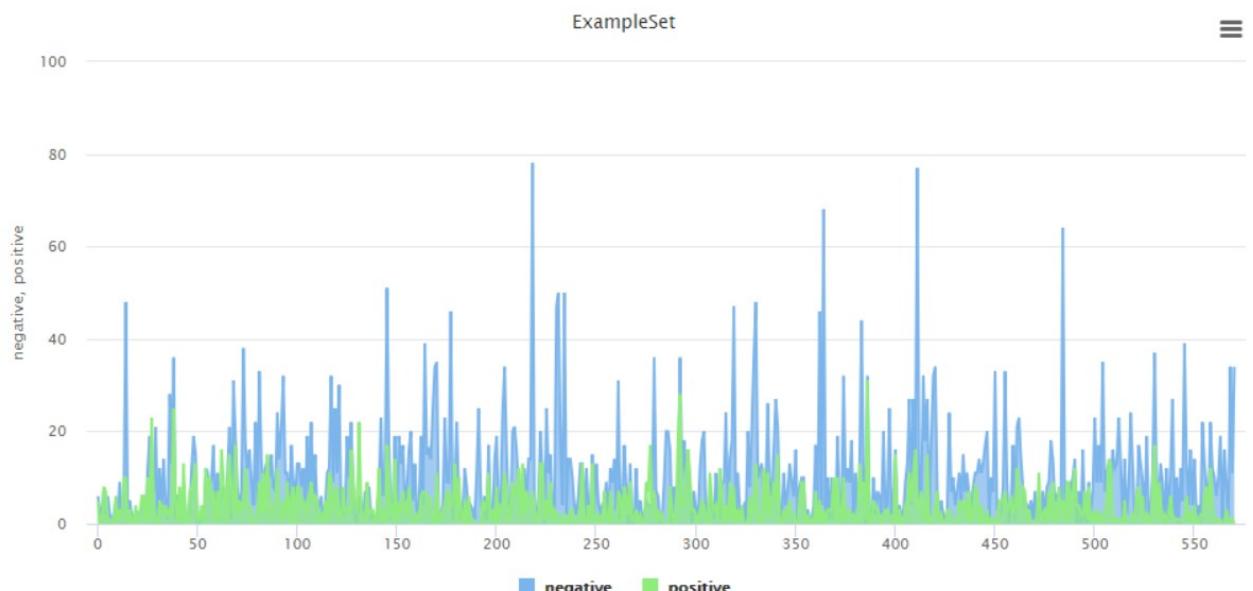


Anno 2013

Possiamo osservare come sono stati presi i vari stock. Innanzitutto notiamo che la distribuzione degli stock per l'anno 2011 è abbastanza uniforme, con prevalenza di osservazioni di VZ e GS e poche di BAC:

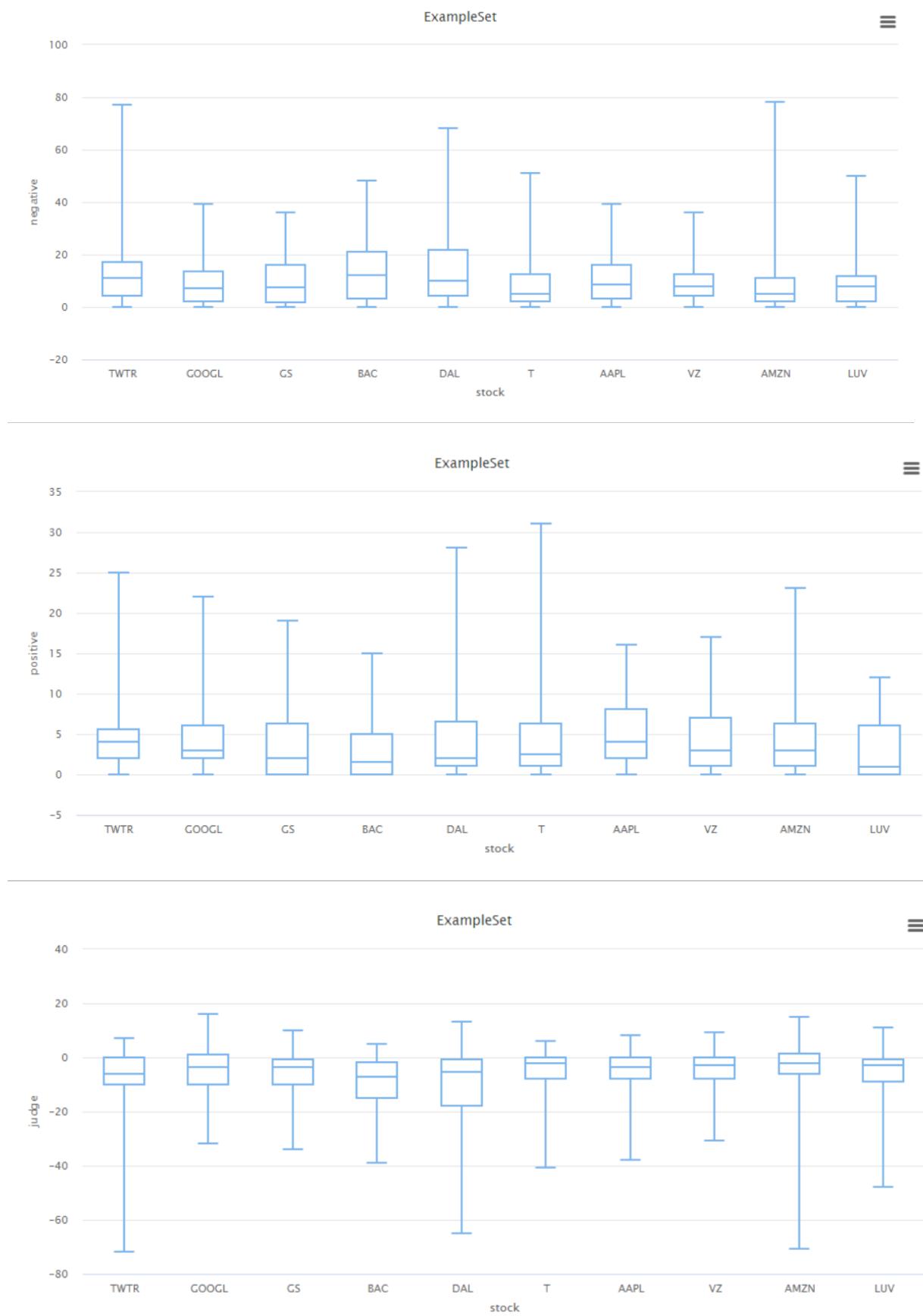


Guardiamo i valori dei giudizi positivi e negativi



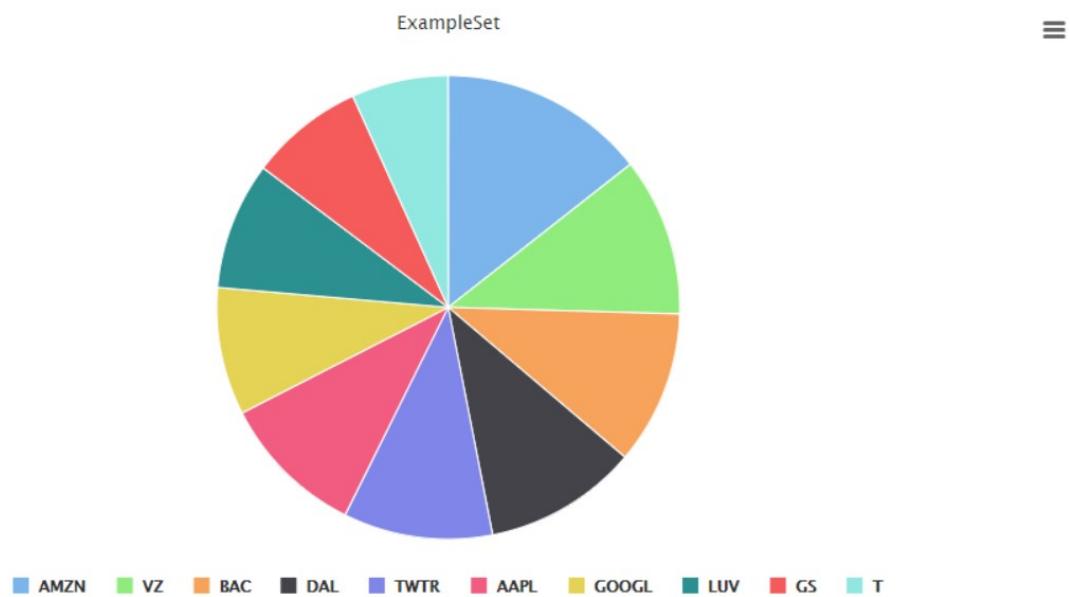
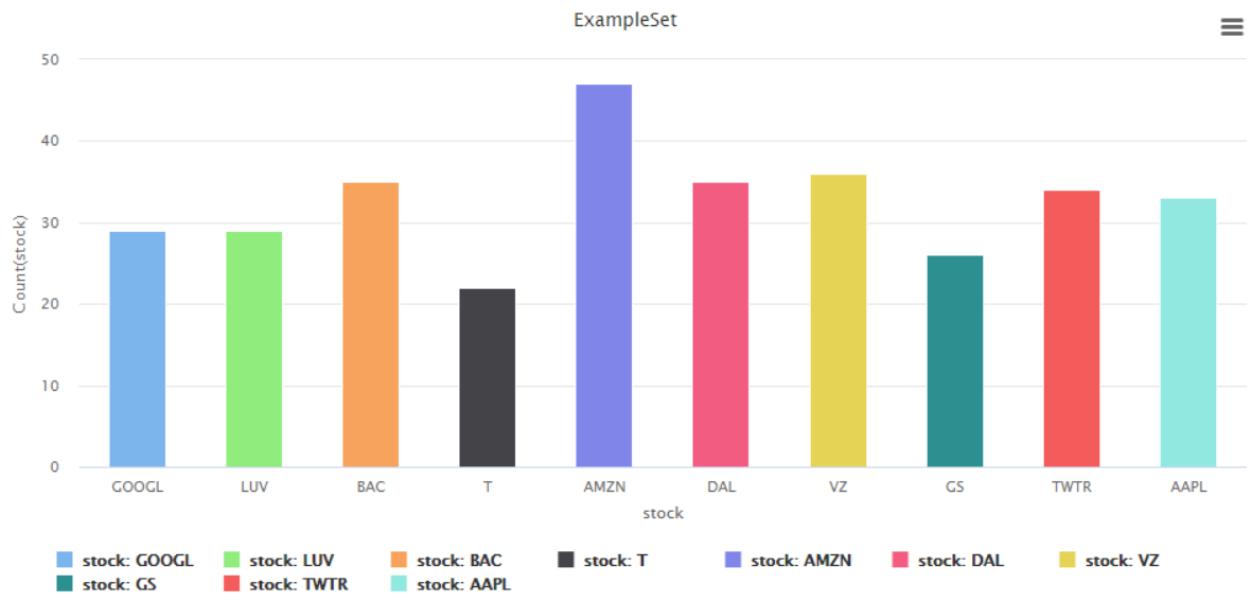
Row No.	average(judge)	sum(judge)	sum(positive)	sum(negative)
1	-6.580	-3757	2369	6126

Collegamento tra stock, positive, negative e judge

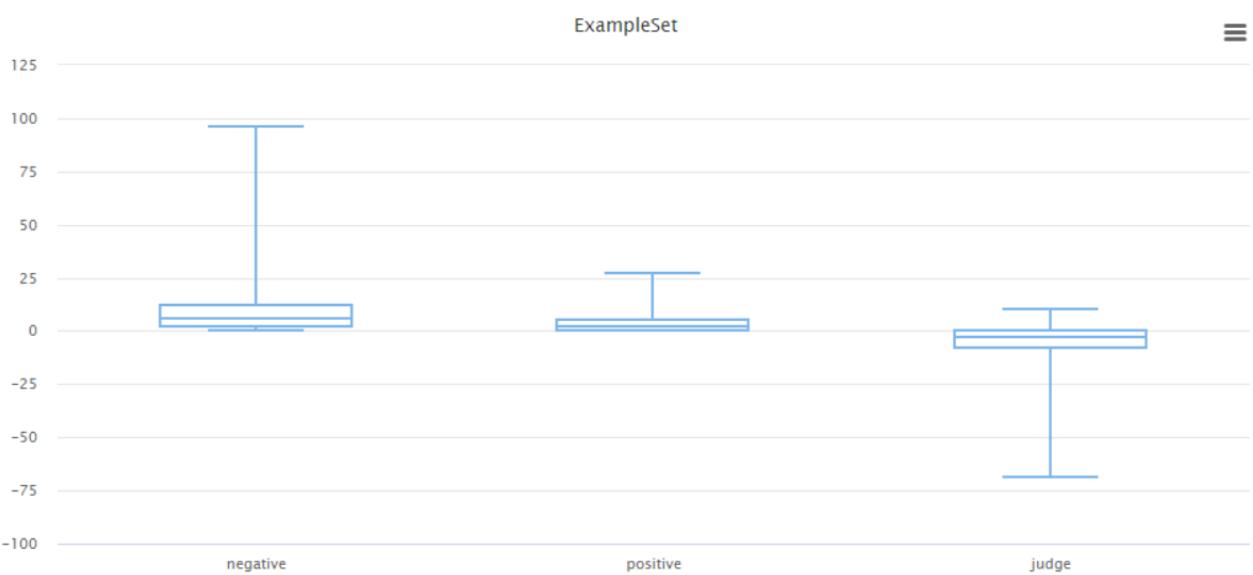
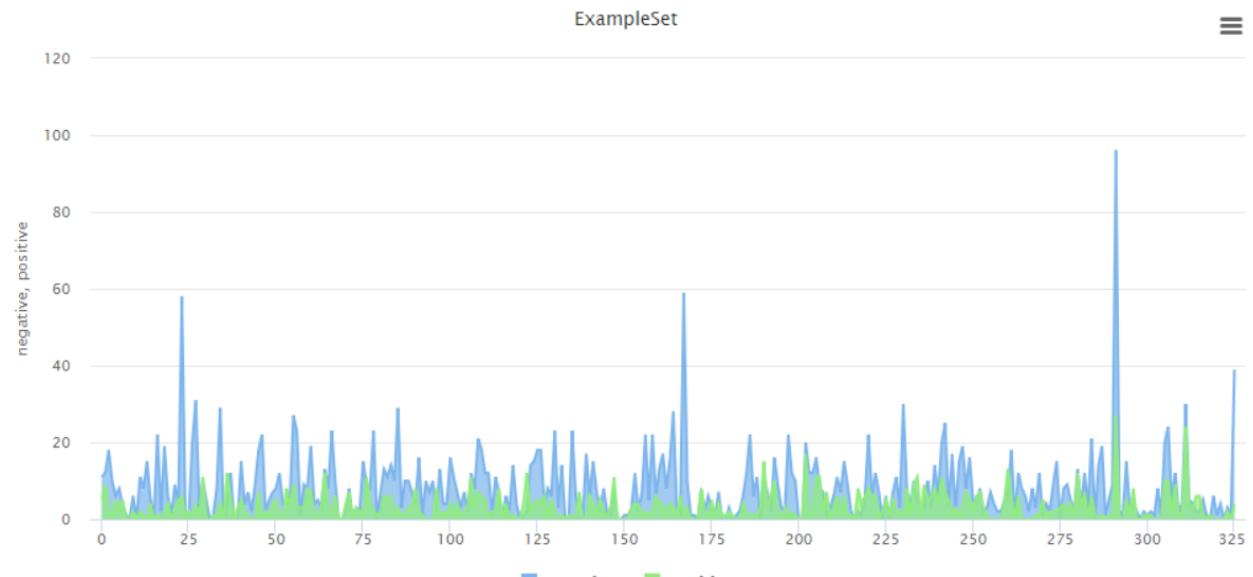


Anno 2015

Possiamo osservare come sono stati presi i vari stock. Innanzitutto notiamo che la distribuzione degli stock per l'anno 2015 è abbastanza uniforme, con prevalenza di osservazioni per AMZN e poche per T:

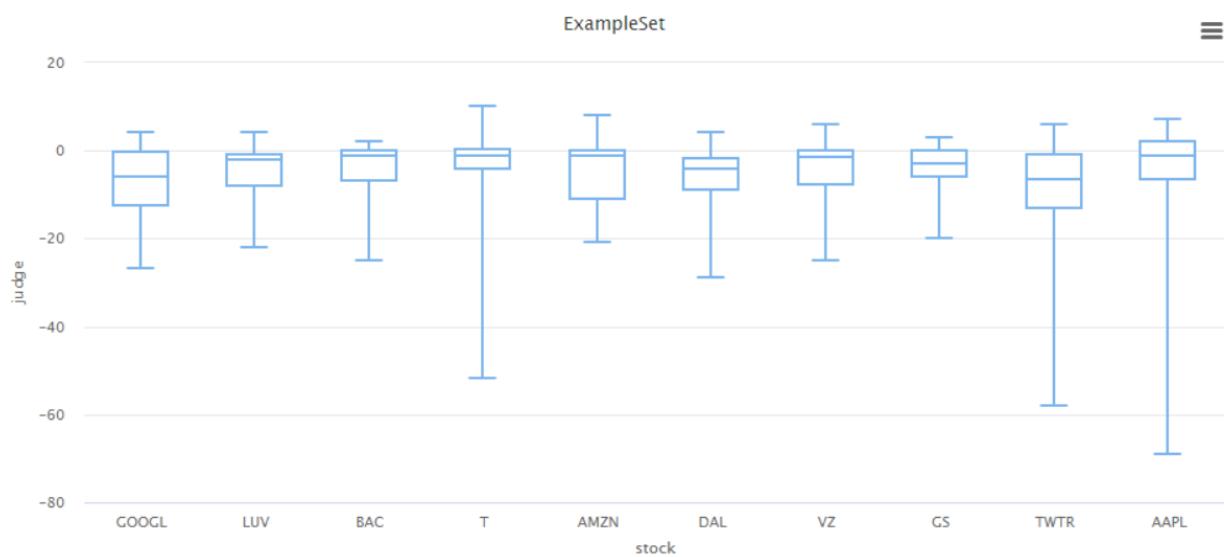
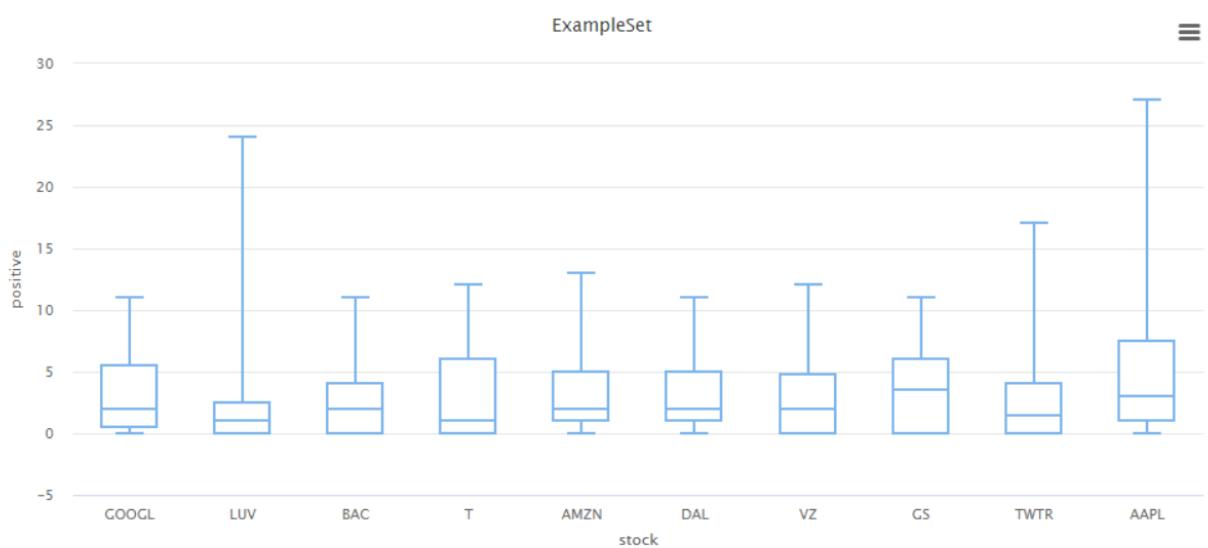
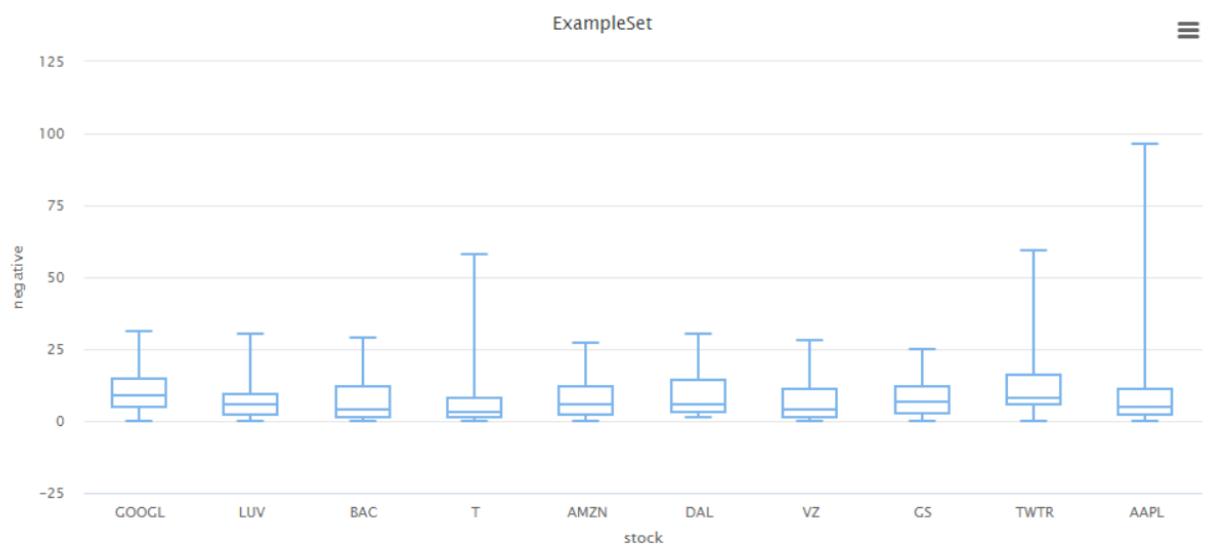


Guardiamo i valori dei giudizi positivi e negativi

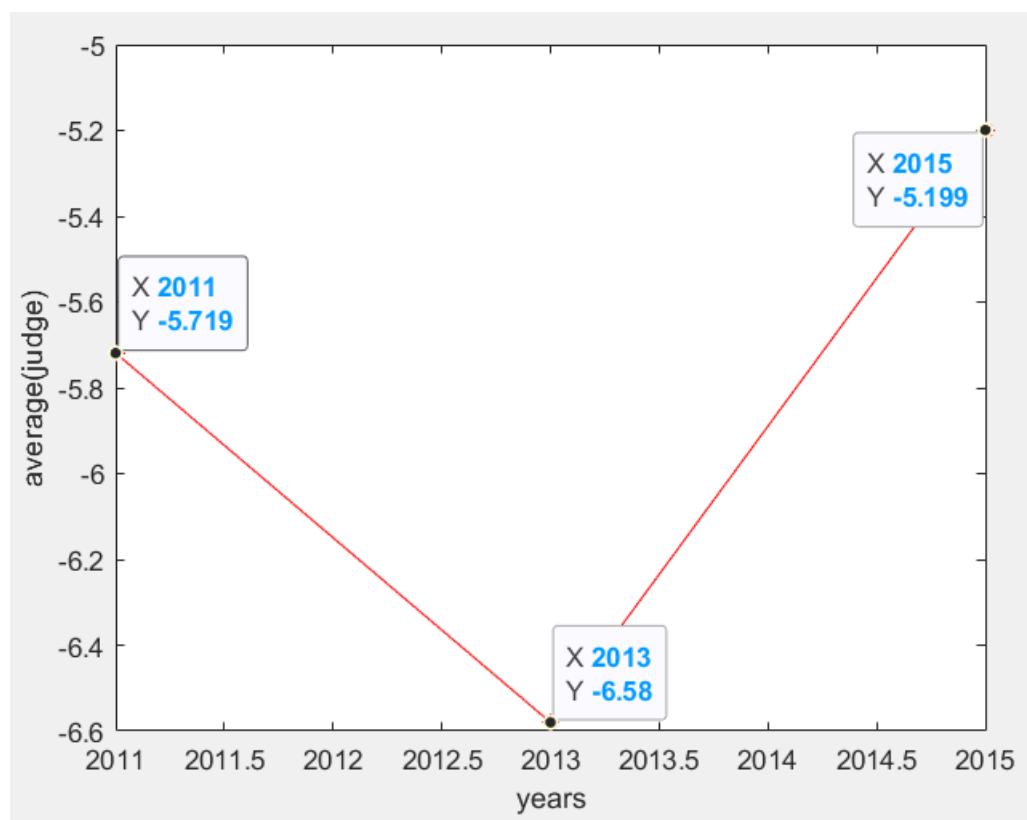


Row No.	average(judge)	sum(judge)	sum(positive)	sum(negative)
1	-5.199	-1695	1027	2722

Collegamento tra stock , positive, negative e judge

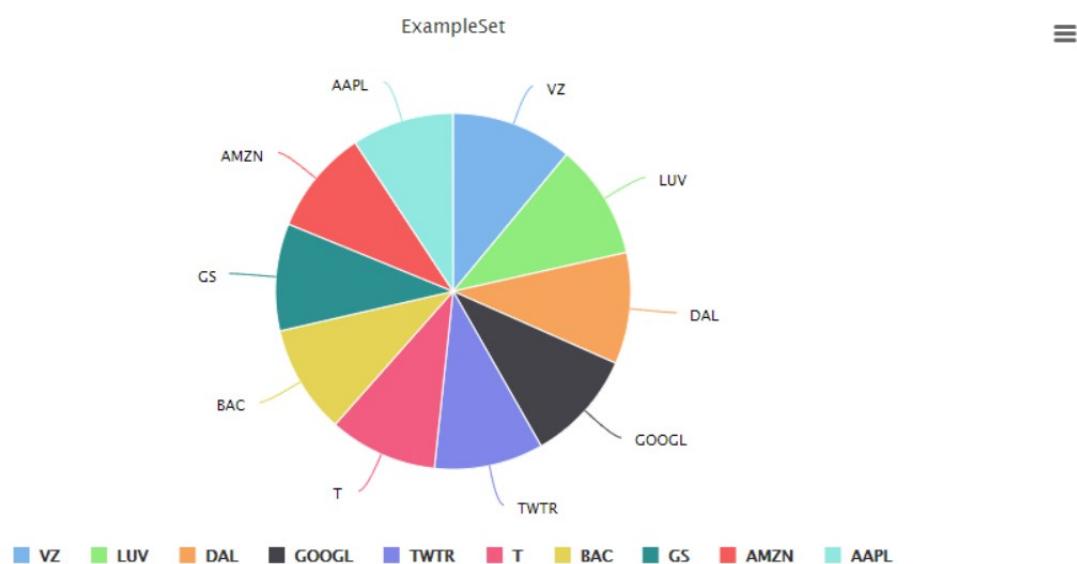
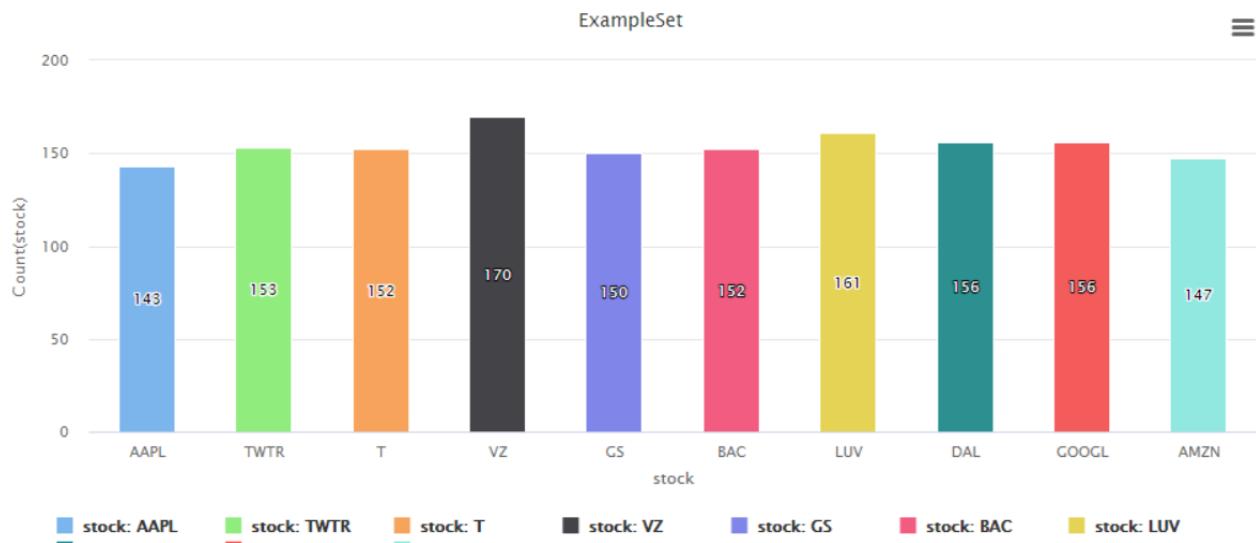


Possiamo riassumere quanto esplorato con questo grafico:

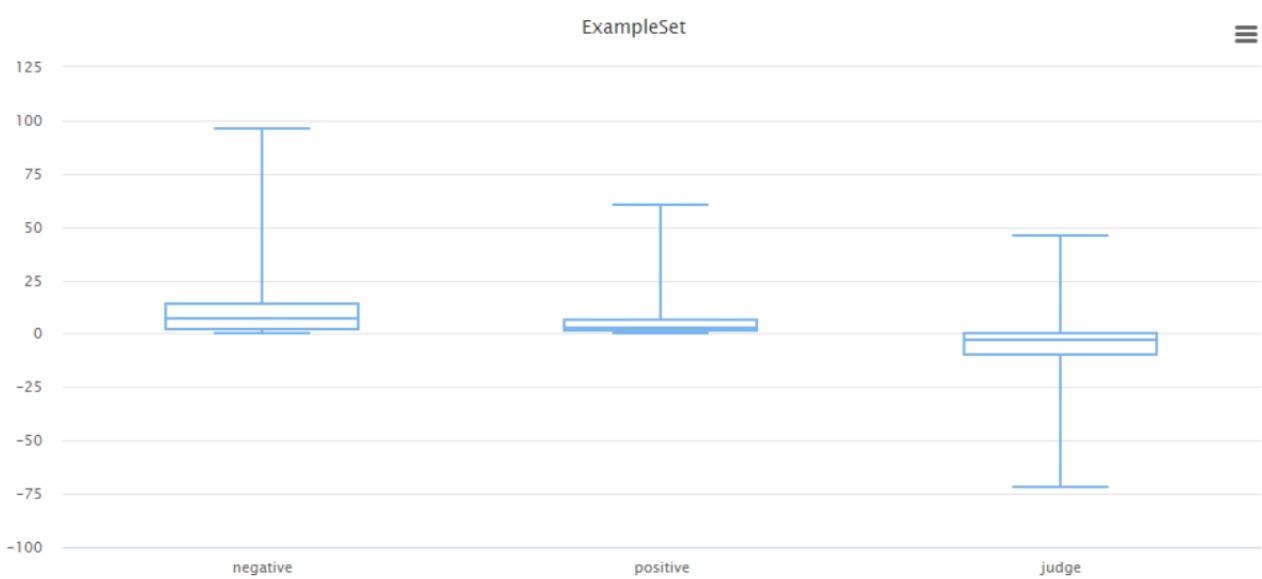


Anni 2011,2013,2015

Possiamo osservare come sono stati presi i vari stock. Innanzitutto notiamo che la distribuzione degli stock nei 3 anni è molto uniforme, con prevalenza di osservazioni per VZ:

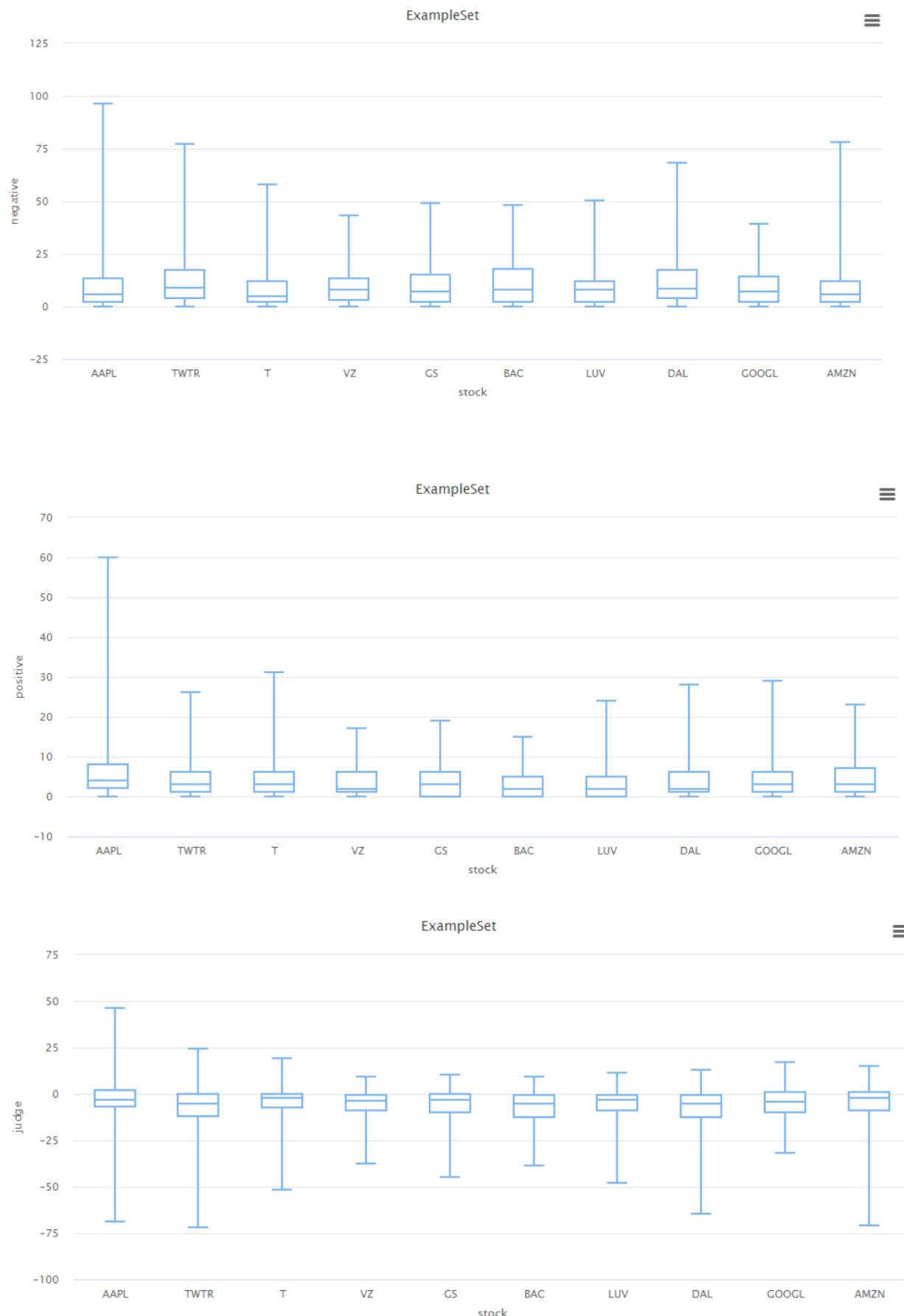


Guardiamo i valori dei giudizi positivi e negativi



Row No.	average(judge)	sum(judge)	sum(positive)	sum(negative)
1	-5.928	-9129	6221	15350

Collegamento tra stock, positive, negative e judge



FASE 2: CLUSTERING

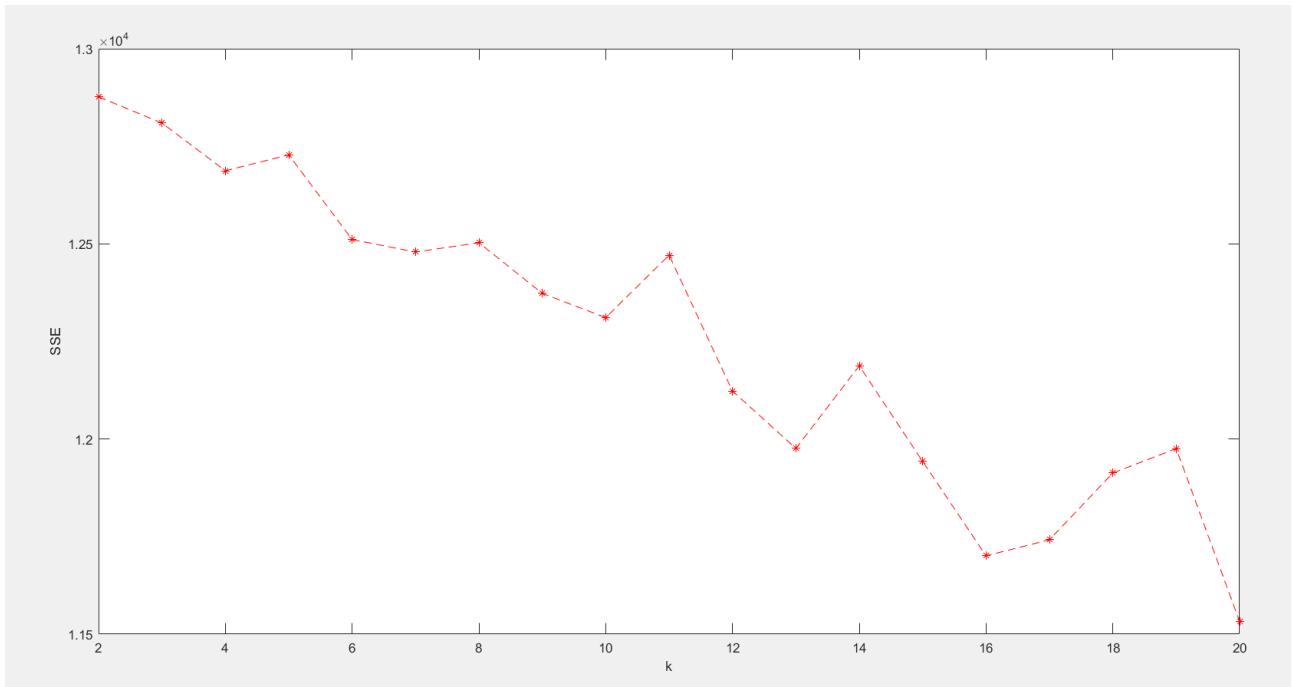
Prima di fare clustering, riduciamo il dataset eliminando i dati che contengono quelle news che non sono in inglese (text vectorization) e selezioniamo le variabili di interesse per fare clustering: “date” (“ts” rinominato), “positive”, “negative”, “body” e “stock”. Sull’attributo “body” applichiamo l’operatore *process document from data*, escludendo le parole che hanno frequenza inferiore al 5% e superiore al 99%.

2.1 K-means

1. JUDGE NORMALIZZATA, DATE NORMALIZZATA

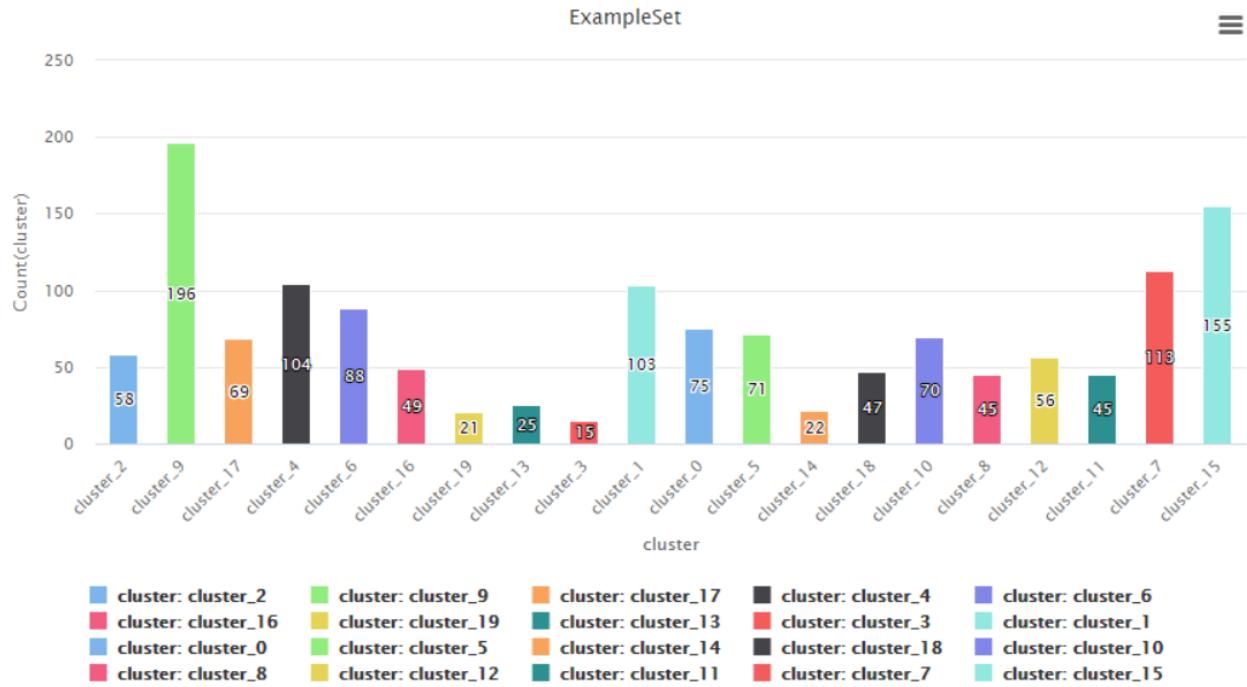
Introduciamo una nuova variabile, denominata “judge” (= “positive” - “negative”), che normalizziamo insieme all’attributo “date”, rimuovendo al contempo “positive” e “negative”.

Diamo uno sguardo al grafico dell’SSE:

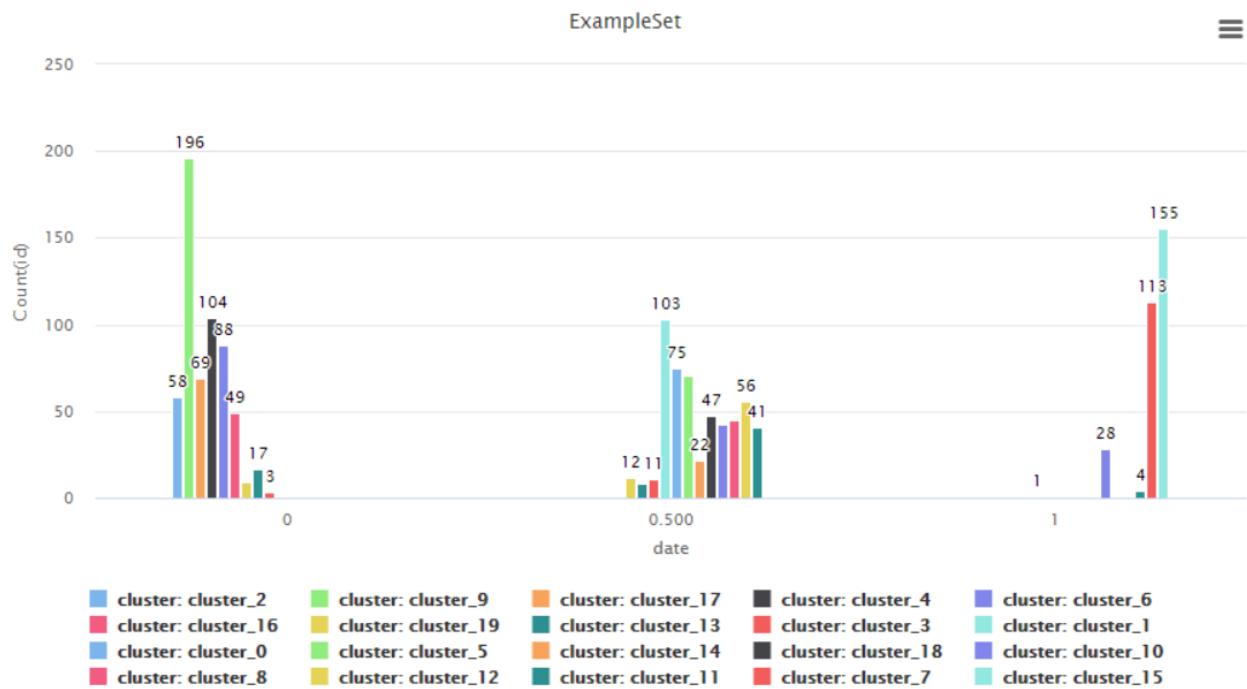


Non si intravede un gomito evidente.

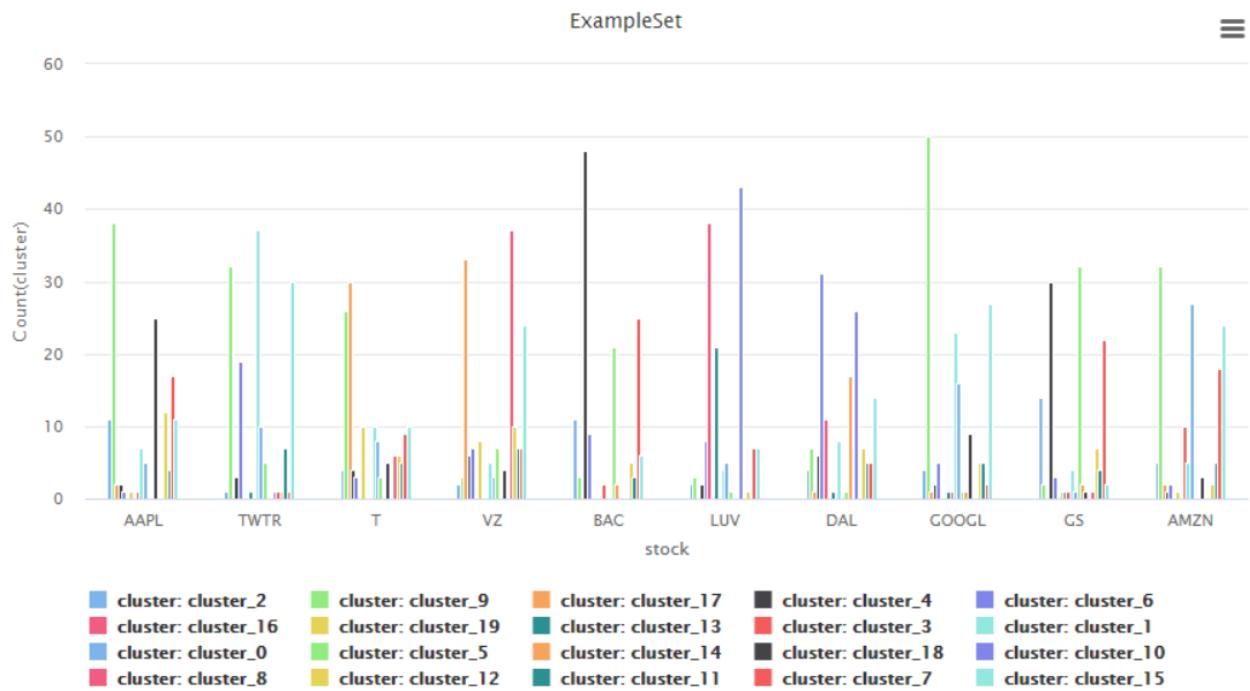
Se non altro per abbassare il valore dell'SSE, vediamo come si dispongono i cluster ottenuti con un valore di k pari a 20 (SSE=11532.133):



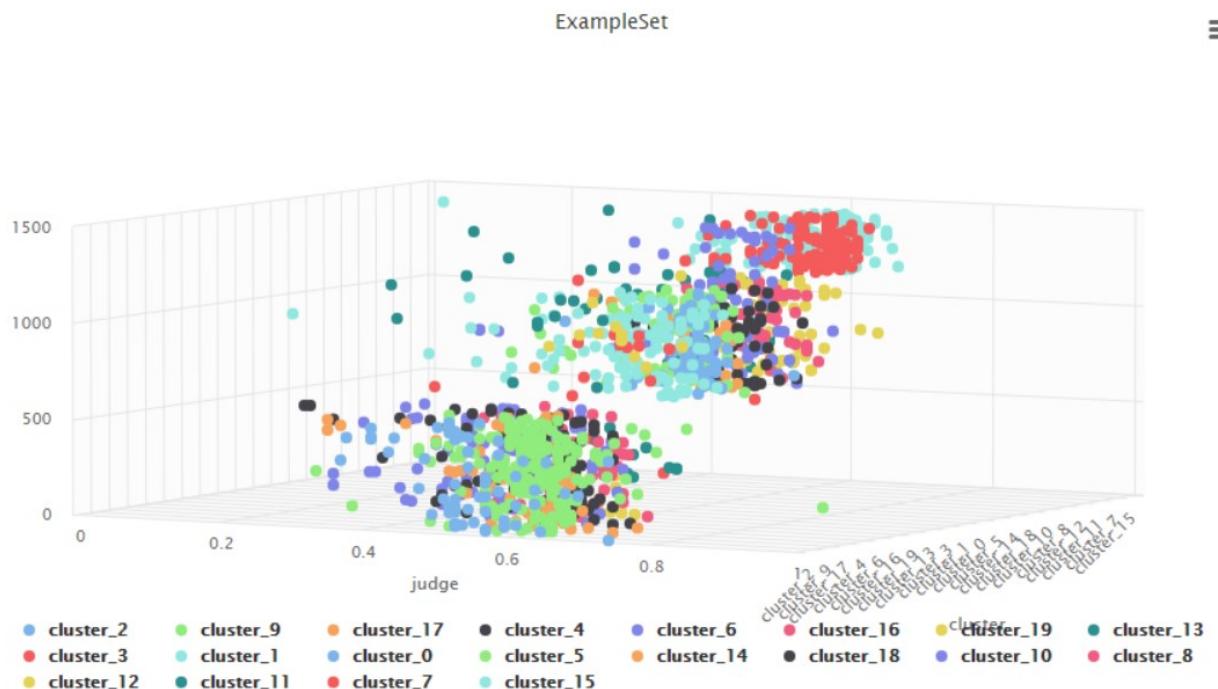
Rispetto a date:



Rispetto agli stock:



Rispetto a judge:

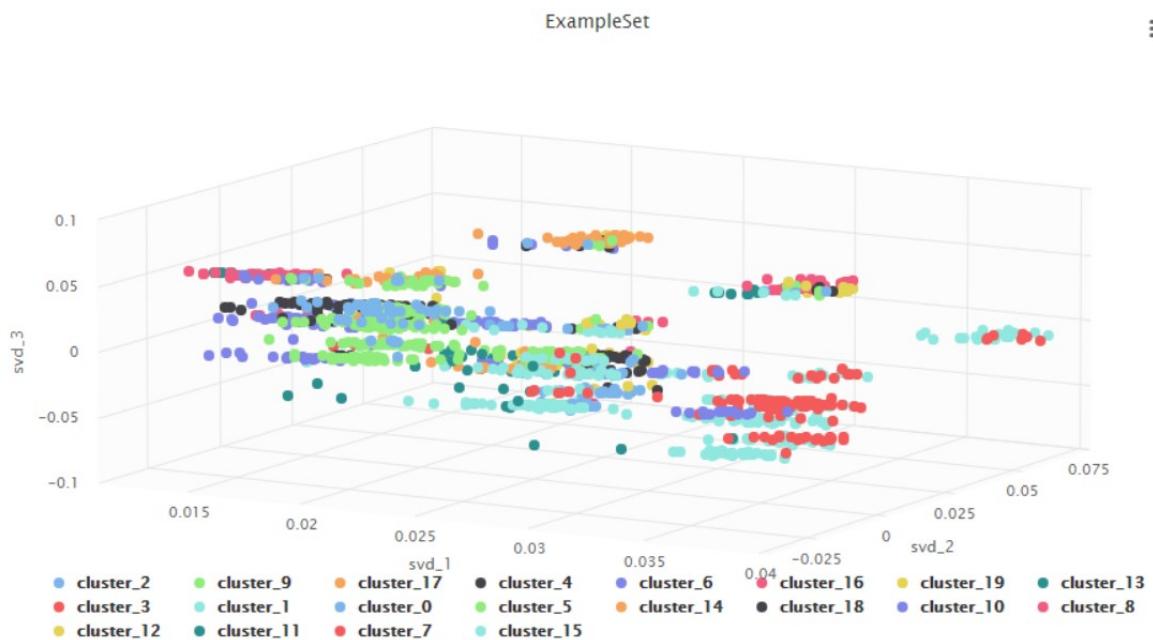


Osserviamo la misura di coesione dei cluster:

PerformanceVector

```
PerformanceVector:  
Avg. within cluster distance: -154.893  
Avg. within cluster distance for cluster 0: -117.738  
Avg. within cluster distance for cluster 1: -164.048  
Avg. within cluster distance for cluster 2: -84.736  
Avg. within cluster distance for cluster 3: -17.117  
Avg. within cluster distance for cluster 4: -159.620  
Avg. within cluster distance for cluster 5: -107.466  
Avg. within cluster distance for cluster 6: -139.297  
Avg. within cluster distance for cluster 7: -179.283  
Avg. within cluster distance for cluster 8: -58.672  
Avg. within cluster distance for cluster 9: -314.488  
Avg. within cluster distance for cluster 10: -98.532  
Avg. within cluster distance for cluster 11: -70.117  
Avg. within cluster distance for cluster 12: -84.441  
Avg. within cluster distance for cluster 13: -28.730  
Avg. within cluster distance for cluster 14: -30.189  
Avg. within cluster distance for cluster 15: -252.951  
Avg. within cluster distance for cluster 16: -62.131  
Avg. within cluster distance for cluster 17: -101.285  
Avg. within cluster distance for cluster 18: -67.632  
Avg. within cluster distance for cluster 19: -24.921
```

Guardiamo il grafico della svd su 3 dimensioni:



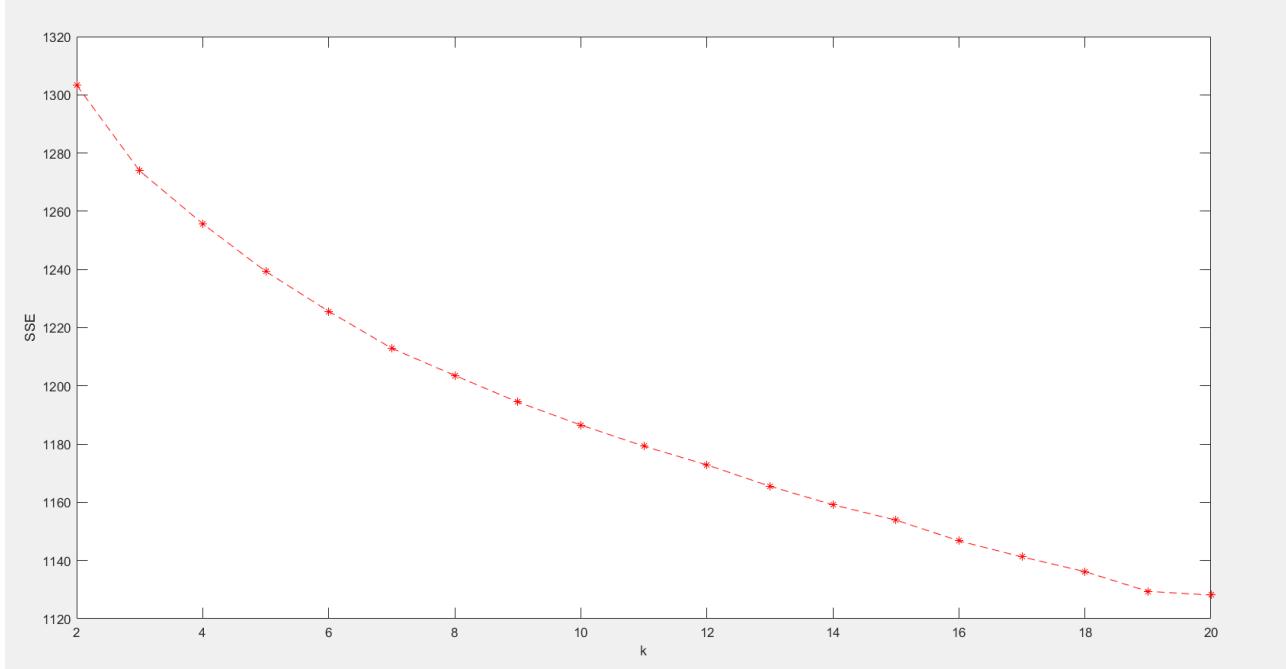
Varianza spiegata: 5.7%

Osserviamo che con 500 attributi della svd spiegheremmo l'82.6% della varianza.

2. POSITIVE E NEGATIVE NORMALIZZATI, NO DATE, NO STOCK

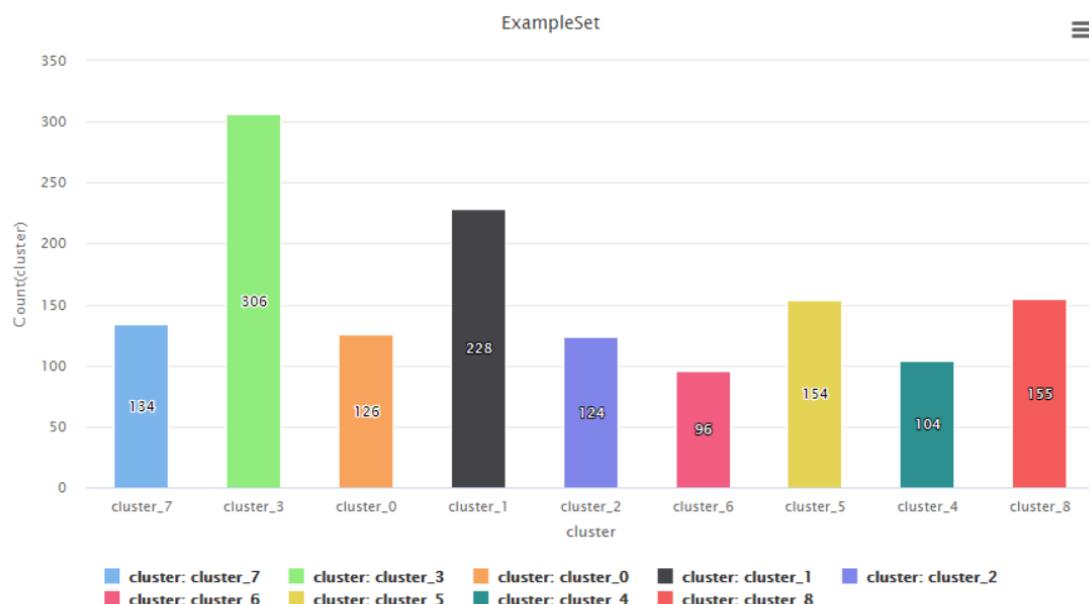
Al fine ridurre il valore di SSE e determinare cluster più omogenei in vista di una proficua ricerca di regole di associazione, manteniamo gli attributi “positive” e “negative”, normalizzandoli, mentre rimuoviamo “date” e “stock”.

Diamo uno sguardo al grafico dell'sse:

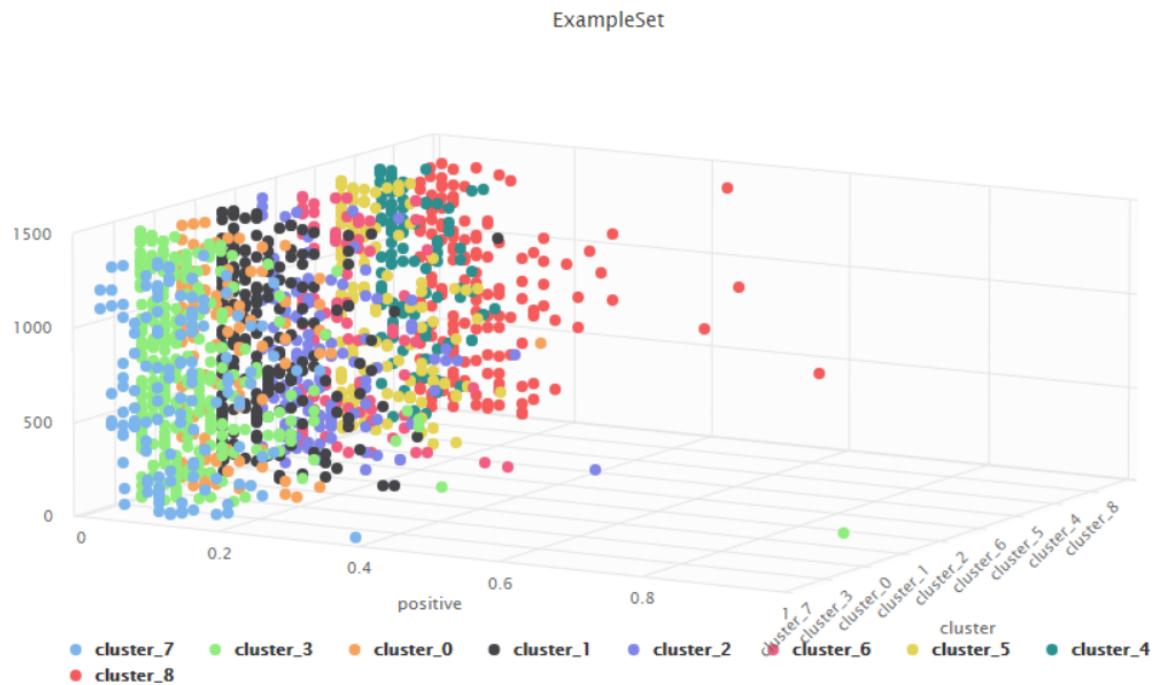


Per $k=9$ vengono creati dei cluster abbastanza bilanciati e, inoltre, il valore dell'SSE è decisamente più basso di prima (1194.458). Anche se il gomito del grafico potrebbe trovarsi in corrispondenza di valori di k maggiori di 9, riteniamo che questa misura sia adatta ad esplorare regole di associazione all'interno dei cluster, evitando problemi dovuti alla riduzione eccessiva delle loro dimensioni (come il disallineamento dei dati e la conseguente generazione di un numero enorme di itemset frequenti).

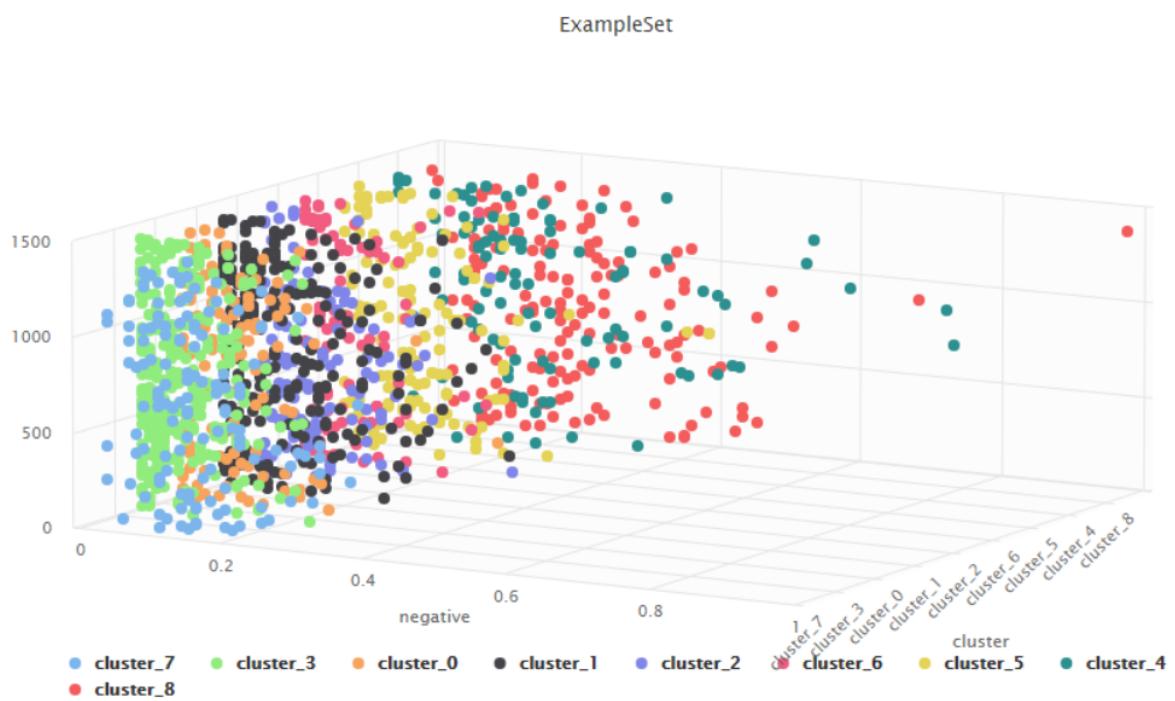
Diamo uno sguardo ai cluster che si creano per $k=9$ (sse=1194.458):



Rispetto a positive:



Rispetto a negative:



Osserviamo la misura di coesione dei cluster:

PerformanceVector

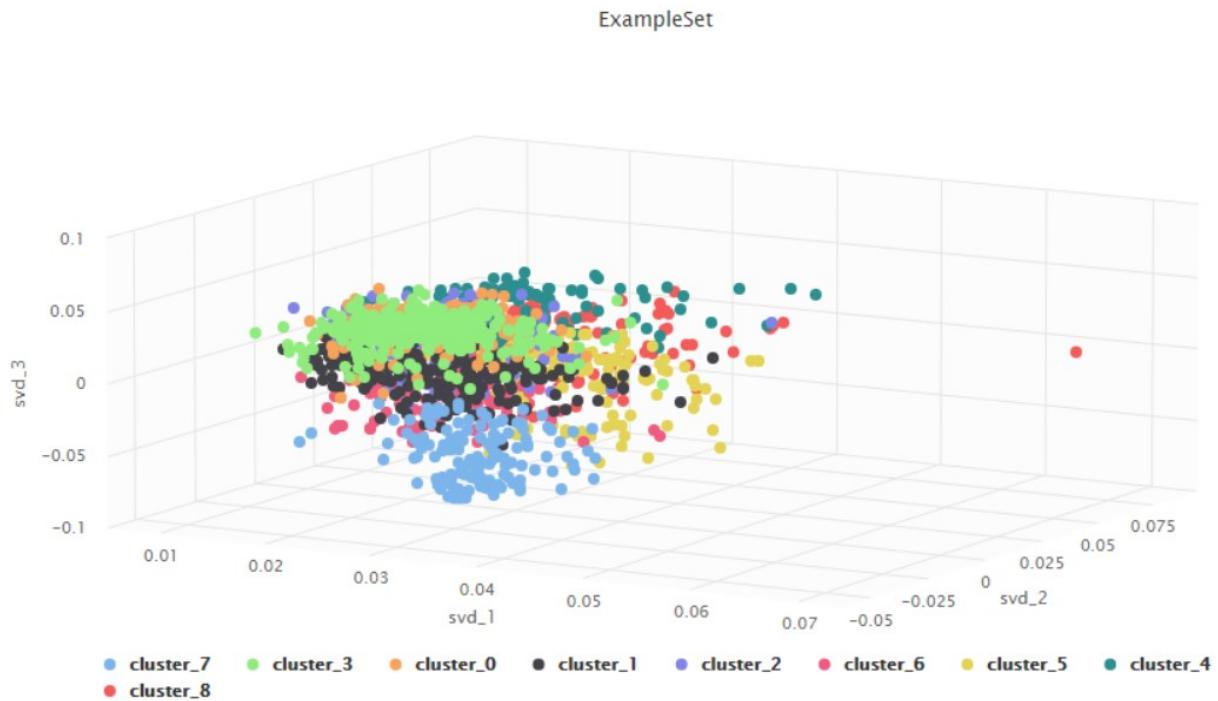
```
PerformanceVector:  
Avg. within cluster distance: -239.722  
Avg. within cluster distance for cluster 0: -157.318  
Avg. within cluster distance for cluster 1: -298.986  
Avg. within cluster distance for cluster 2: -155.601  
Avg. within cluster distance for cluster 3: -415.654  
Avg. within cluster distance for cluster 4: -133.693  
Avg. within cluster distance for cluster 5: -187.656  
Avg. within cluster distance for cluster 6: -120.555  
Avg. within cluster distance for cluster 7: -159.532  
Avg. within cluster distance for cluster 8: -205.508
```

Osserviamo la distanza tra i cluster:

PerformanceVector

```
PerformanceVector:  
Avg. within centroid distance: 0.837  
Avg. within centroid distance_cluster_0: 0.792  
Avg. within centroid distance_cluster_1: 0.866  
Avg. within centroid distance_cluster_2: 0.798  
Avg. within centroid distance_cluster_3: 0.927  
Avg. within centroid distance_cluster_4: 0.838  
Avg. within centroid distance_cluster_5: 0.754  
Avg. within centroid distance_cluster_6: 0.807  
Avg. within centroid distance_cluster_7: 0.718  
Avg. within centroid distance_cluster_8: 0.888  
Davies Bouldin: 5.032
```

Guardiamo il grafico della svd su 3 dimensioni:



Varianza spiegata: 2.8%

Notiamo che con 500 attributi della svd spiegheremmo l'80.7% della varianza.

2.2 Dbscan

1.JUDGE NORMALIZZATO, DATE NORMALIZZATO

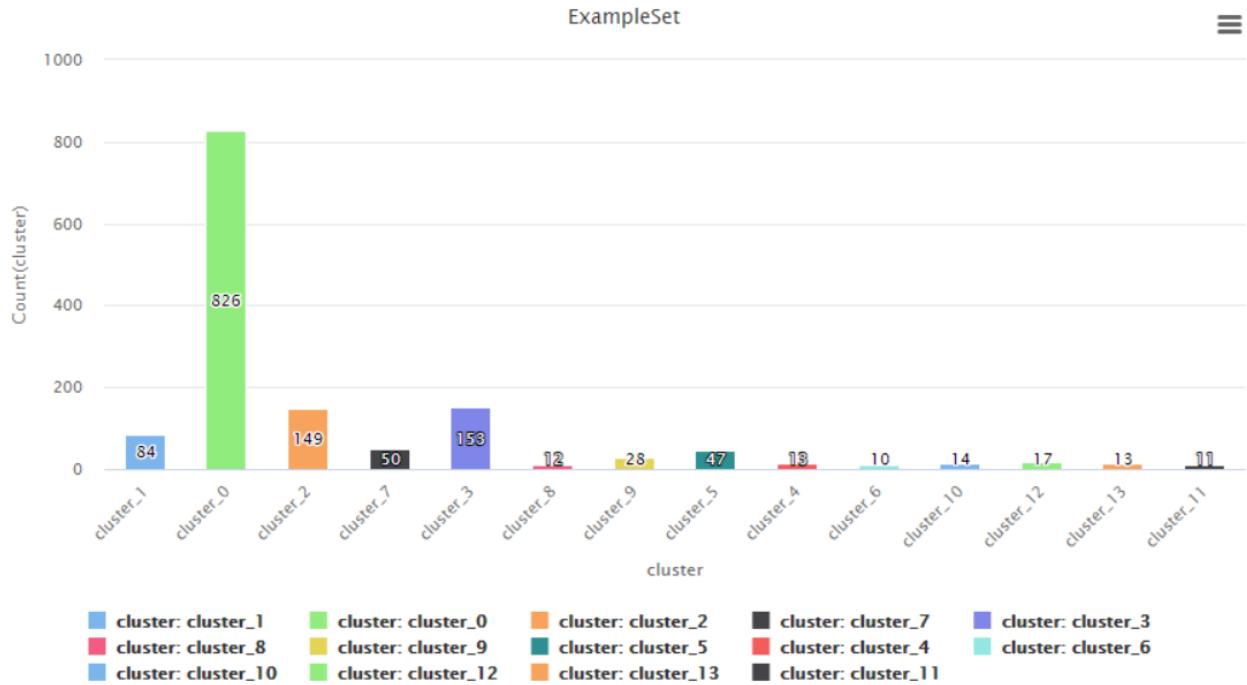
Anche in tal caso, come abbiamo fatto con il k-means, introduciamo una nuova variabile, denominata “judge” (= “positive” - “negative”), che normalizziamo insieme all’attributo “date”; quindi, rimuoviamo “positive” e “negative”.

Osserviamo il grafico del k dist per farci un’idea di come scegliere eps:



Da questo grafico si evince che un valore di eps accettabile potrebbe aggirarsi intorno ad 1.2.

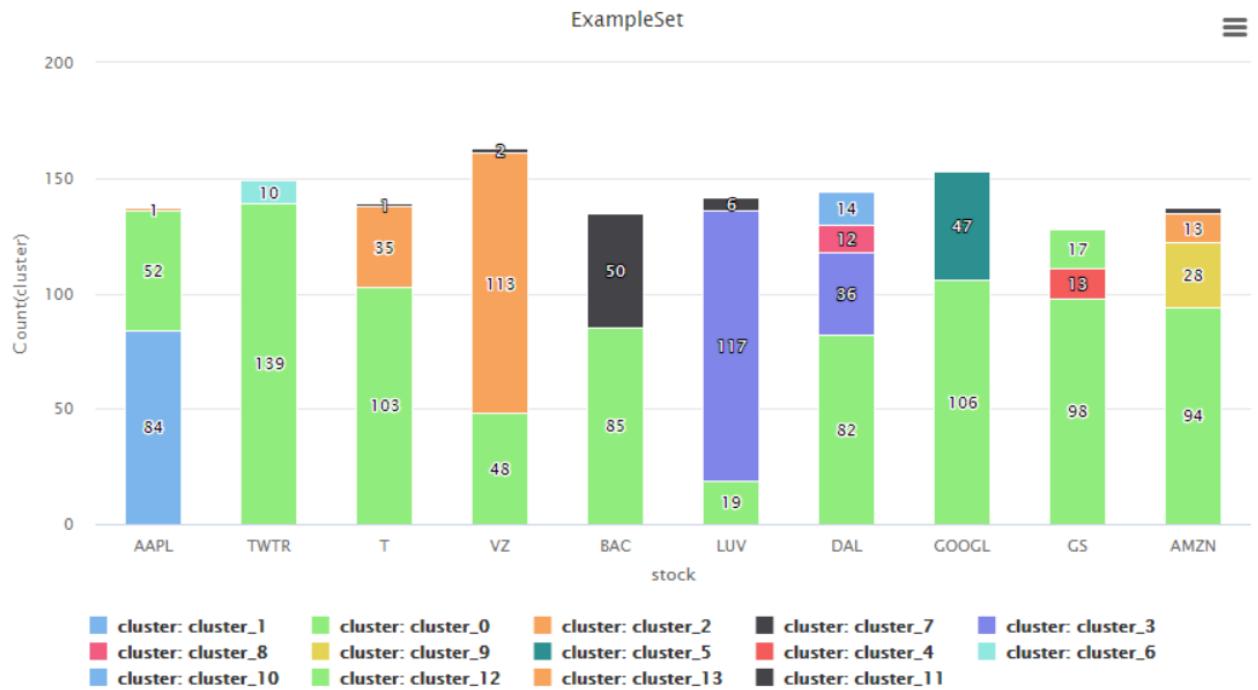
I cluster che si creano con minpoint=10 ed eps=1.2 (SSE=633.751) sono:



Rispetto a date:



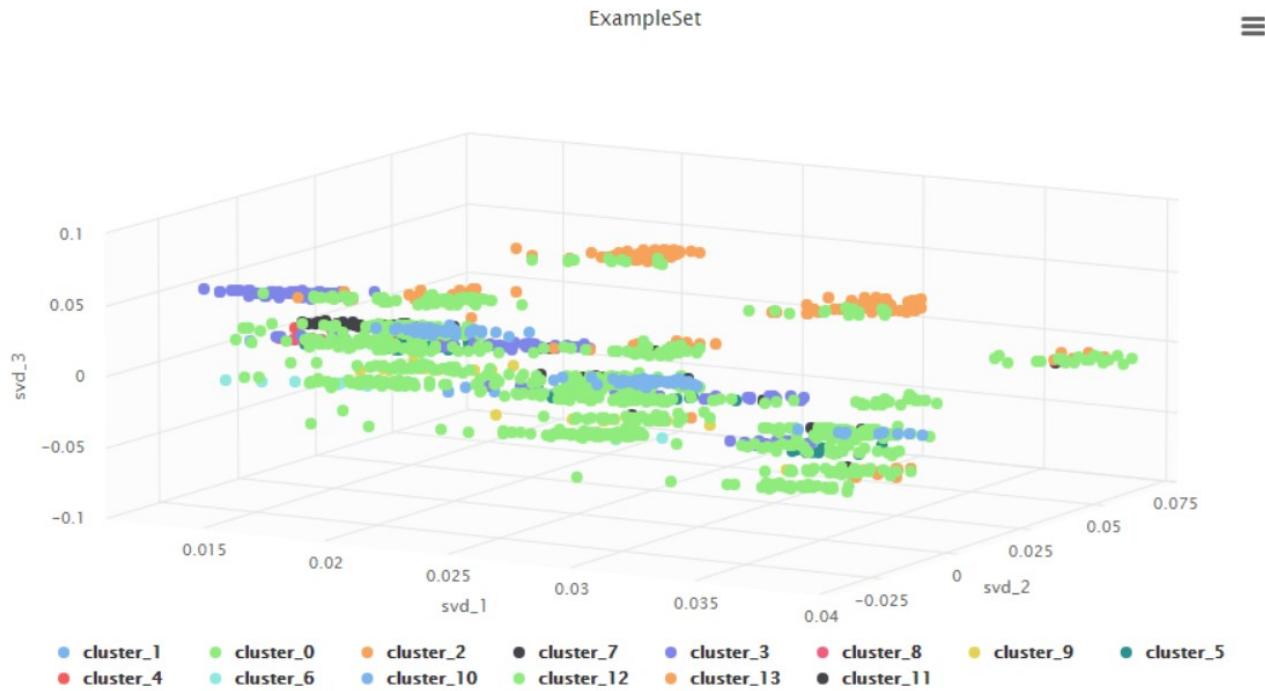
Rispetto a stock:



Rispetto a judge:



Guardiamo il grafico della svd su tre dimensioni:



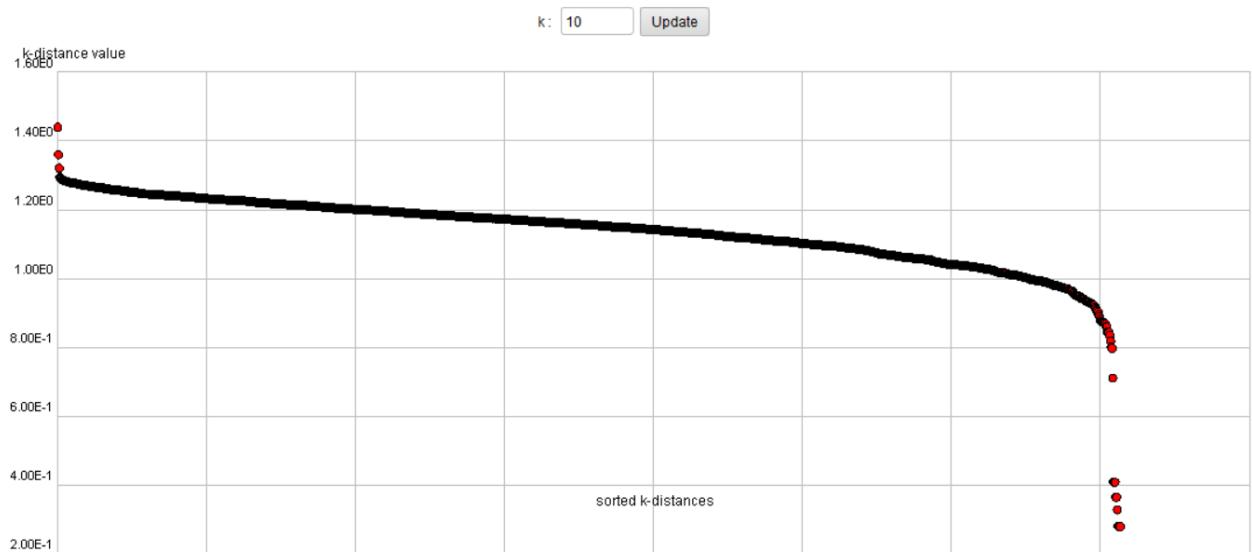
Varianza spiegata: 5.7%.

Si osservi che con 500 attributi della svd spiegheremmo l'82.6% della varianza.

2. POSITIVE E NEGATIVE NORMALIZZATI, NO DATE, NOSTOCK

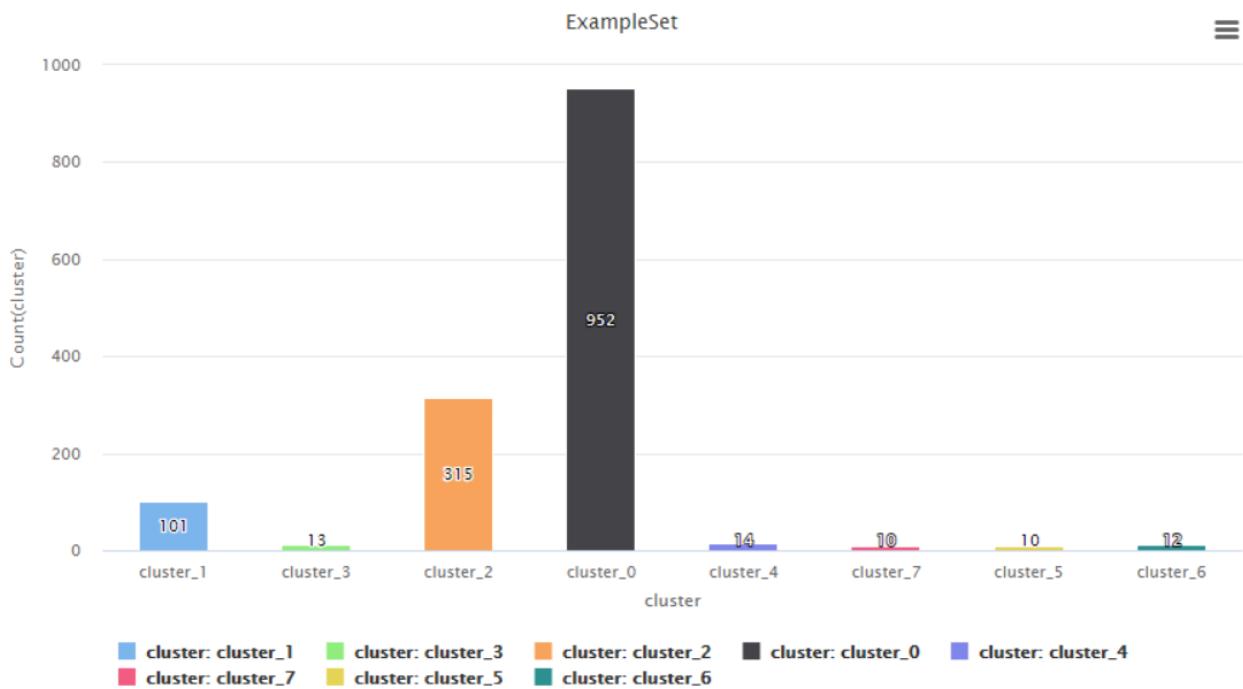
Rimuoviamo l'attributo "stock", mantenendo "positive" e "negative", con la speranza di ottenere risultati migliori di quelli prodotti dal k-means.

Prima di tutto osserviamo il grafico del k dist per farci un'idea di come scegliere eps:

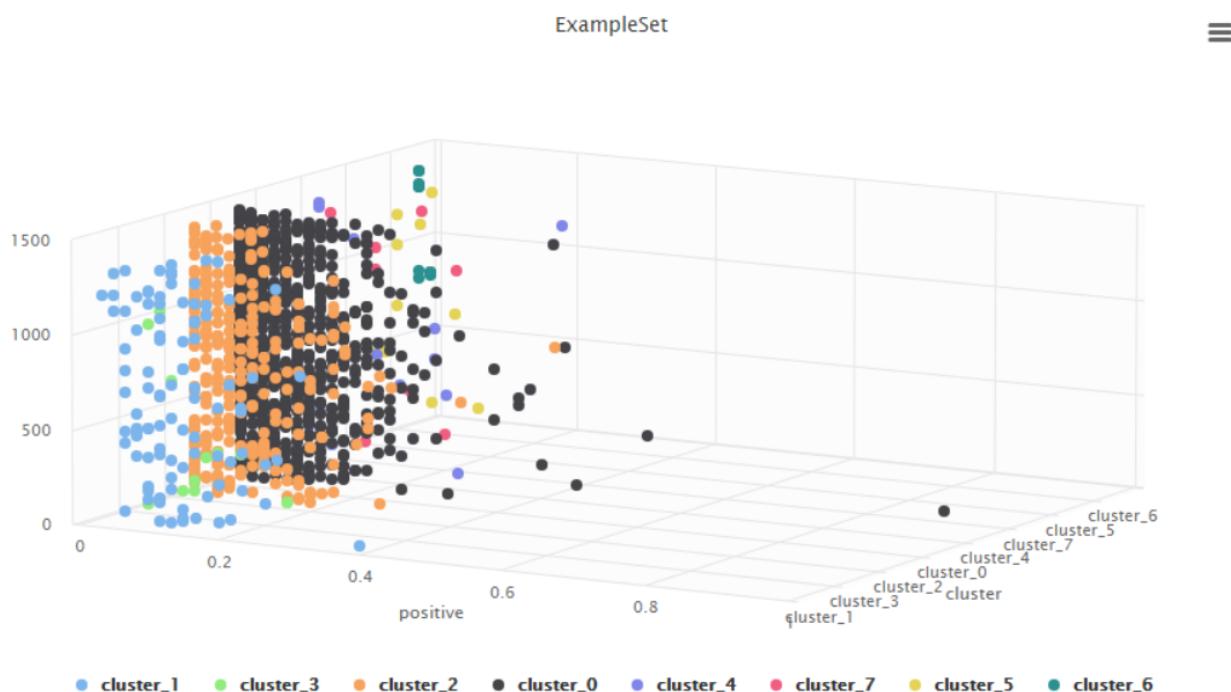


In tal caso, sembra che un valore di eps accettabile possa aggirarsi intorno a 1.1

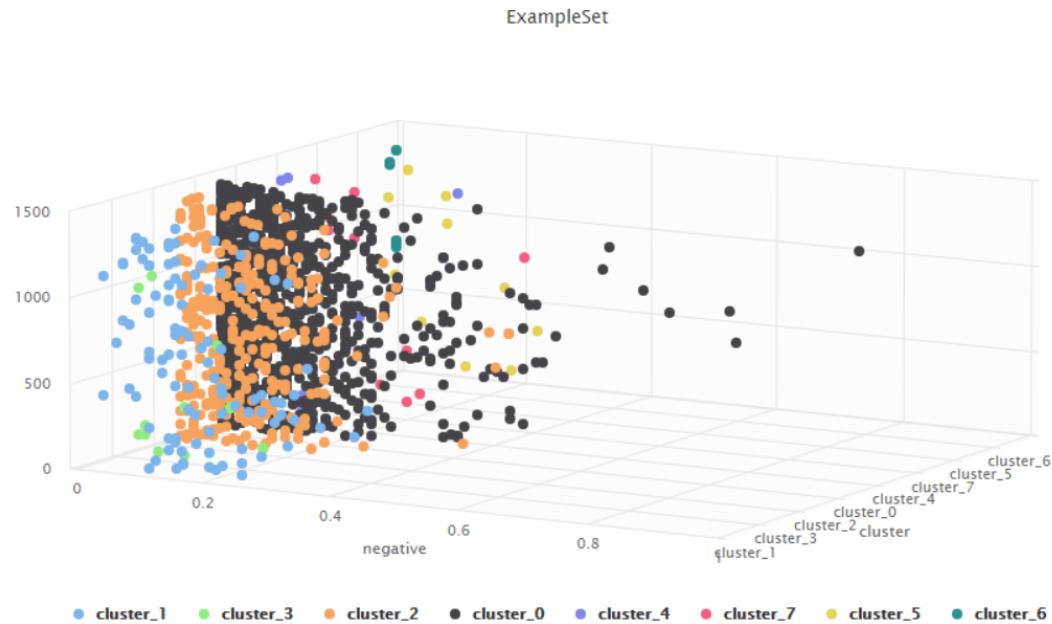
Con minpoints=10 ed eps=1.05, otteniamo la seguente configurazione di cluster (SSE=375.830):



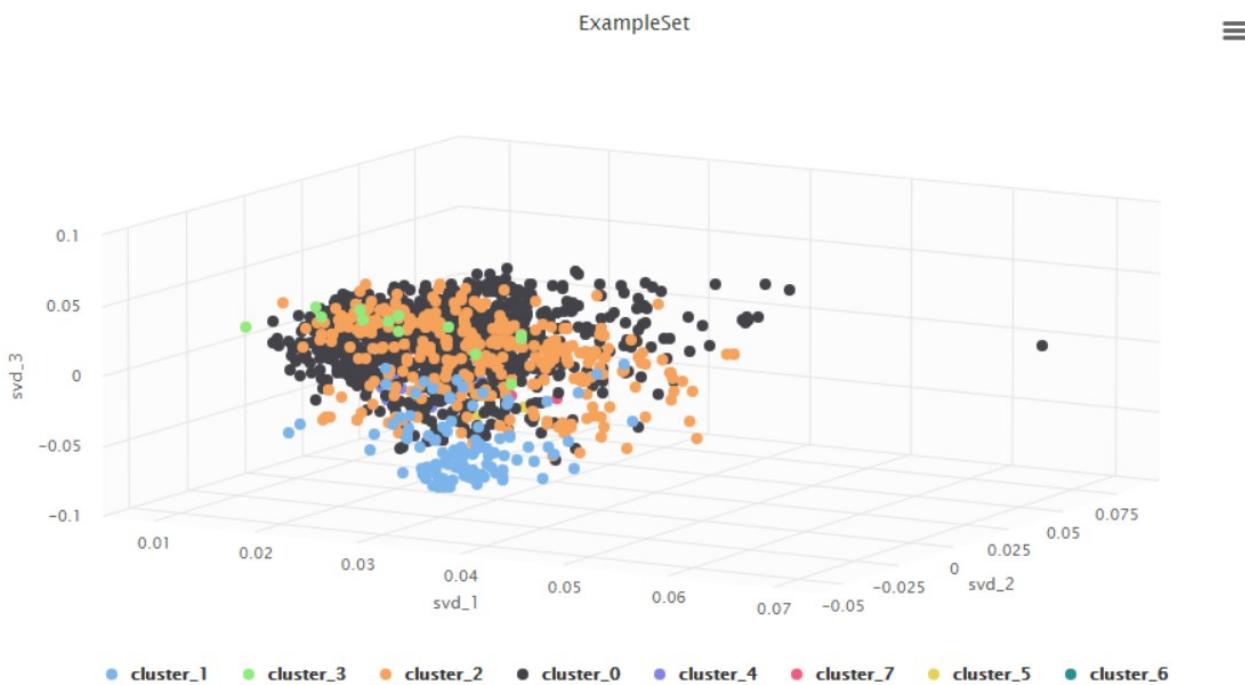
Rispetto a positive:



Rispetto a negative:



Guardiamo il grafico della svd su 3 dimensioni:



Varianza spiegata: 2.8%.

Si osservi che con 500 attributi della svd spiegheremmo l'80.7% della varianza.

Gli ultimi due modelli di clusters risultano più sbilanciati dei loro predecessori, a causa della disomogeneità presente nei dati a disposizione. Pertanto, sebbene i valori di SSE ricavati con l'algoritmo dbscan siano inferiori a quelli ottenuti mediante il clustering partizionale, riteniamo che quest'ultimo, in tal caso, fornisca i gruppi più adatti all'esplorazione di eventuali regole di associazione.

FASE 3: REGOLE DI ASSOCIAZIONE

Tenendo presente che l'algoritmo di estrazione di regole utilizzato, cioè l'FP-GROWTH, non è particolarmente performante in presenza di dati sparsi, analizziamo i risultati da esso ottenuti, relativamente ai 9 gruppi identificati con il k-means, nella seconda sessione di clustering.

- Cluster 0: minsup=0.5, minconf=0.5. Le dimensioni degli itemset frequenti variano da 1 a 7
Argomento: telefonia

Regole di associazione:

No.	Premises ↑	Conclusion	Support	Confidence	Lift	Conviction
775	positive = E	negative = E, group	0.556	0.565	1.016	1.021
776	positive = E	said, network	0.556	0.565	1.002	1.002
777	positive = E	compani, plan	0.556	0.565	1.002	1.002
778	positive = E	negative = E, said, year	0.556	0.565	1.016	1.021
779	positive = E	negative = E, said, servic	0.556	0.565	1.002	1.002
780	positive = E	negative = E, compani, year	0.556	0.565	1.016	1.021
781	positive = E	said, wireless, edit	0.556	0.565	1.016	1.021
782	positive = E	verizon, wireless, report	0.556	0.565	1.016	1.021
783	positive = E	communic, compani, edit	0.556	0.565	1.016	1.021
784	positive = E	communic, wireless, report	0.556	0.565	1.016	1.021
785	positive = E	communic, mobil, inc	0.556	0.565	0.988	0.984
786	positive = E	negative = E, said, inc, edit	0.556	0.565	1.016	1.021
787	positive = E	negative = E, compani, wireless, inc	0.556	0.565	1.016	1.021
789	positive = E	said, verizon, compani, wireless	0.556	0.565	1.016	1.021
790	positive = E	said, verizon, compani, mobil	0.556	0.565	1.016	1.021
791	positive = E	said, verizon, compani, inc	0.556	0.565	1.016	1.021
792	positive = E	negative = E, said, verizon, communic, edit	0.556	0.565	1.016	1.021
793	positive = E	negative = E, said, report, inc, edit	0.556	0.565	1.016	1.021
882	positive = E	group	0.563	0.573	1.016	1.021
883	positive = E	said, year	0.563	0.573	1.016	1.021
884	positive = E	compani, year	0.563	0.573	1.002	1.003
885	positive = E	negative = E, said, plan	0.563	0.573	1.016	1.021
886	positive = E	said, inc, edit	0.563	0.573	1.016	1.021

No.	Premises	Conclusion ↑	Support	Confidence	Lift	Conviction
24907	oper	positive = E	0.579	0.986	1.002	1.175
24917	negative = E, said, wireless, inc	positive = E	0.579	0.986	1.002	1.175
24921	negative = E, communic, wireless, mobil	positive = E	0.579	0.986	1.002	1.175
24929	said, verizon, communic, mobil	positive = E	0.579	0.986	1.002	1.175
24930	verizon, communic, compani, wireless	positive = E	0.579	0.986	1.002	1.175
24938	negative = E, said, verizon, communic, inc	positive = E	0.579	0.986	1.002	1.175
24941	negative = E, network	positive = E	0.587	0.987	1.003	1.190
24942	negative = E, servic	positive = E	0.587	0.987	1.003	1.190
24943	said, wireless, inc	positive = E	0.587	0.987	1.003	1.190
24944	compani, wireless, mobil	positive = E	0.587	0.987	1.003	1.190
24951	negative = E, said, mobil, inc	positive = E	0.587	0.987	1.003	1.190
24955	said, verizon, communic, inc	positive = E	0.587	0.987	1.003	1.190
24960	said, mobil, inc	positive = E	0.595	0.987	1.003	1.206
24961	verizon, compani, mobil	positive = E	0.595	0.987	1.003	1.206

Con minsup=0.4

Premises	Conclusion	Support	Confidence	Lift	Conviction
buy	negative = E	0.484	0.984	1.008	1.476
vodafon	negative = E	0.484	0.984	1.008	1.476
corp	negative = E	0.452	0.983	1.007	1.381

Positive: operation, year, plan.

Negative: inc, report, group, wireless, buy-Vodafone, corp.

- Cluster 1: minsup=0.3, minconf=0.5. Le dimensioni degli itemset frequenti variano da 1 a 6.
Argomento: banche

Regole di associazione:

No.	Premis... ↑	Conclusion	Support	Confidence	LaPlace	Lift	Conviction
4103	positive = E	negative = E, bank, said	0.601	0.626	0.817	1.004	1.007
4252	positive = E	negative = E, year	0.605	0.630	0.819	0.991	0.984
4410	positive = E	bank, year	0.610	0.635	0.821	0.978	0.961
4697	positive = E	billion	0.618	0.644	0.826	0.985	0.973
5525	positive = E	market	0.645	0.671	0.839	0.981	0.961
5526	positive = E	said, year	0.645	0.671	0.839	0.994	0.987
6069	positive = E	report	0.662	0.689	0.848	0.995	0.989
7413	positive = E	negative = E, said	0.706	0.735	0.870	1.010	1.027
7539	positive = E	year	0.711	0.740	0.872	0.986	0.961
7661	positive = E	negative = E, bank	0.715	0.744	0.875	0.998	0.995
7662	positive = E	bank, said	0.715	0.744	0.875	0.998	0.995
10704	positive = E	said	0.825	0.858	0.931	1.004	1.022
10854	positive = E	bank	0.829	0.863	0.933	0.994	0.961
11235	positive = E	negative = E	0.842	0.877	0.940	1.004	1.032

No.	Premises	Conclu... ↑	Support	Confidence	Lift	Conviction
13212	offer	positive = E	0.395	0.968	1.008	1.224
13213	negative = E, capit	positive = E	0.395	0.968	1.008	1.224
13214	said, time	positive = E	0.395	0.968	1.008	1.224
13215	negative = E, bank, compani	positive = E	0.399	0.968	1.008	1.237
13218	negative = E, bank, said, america	positive = E	0.399	0.968	1.008	1.237
13221	negative = E, said, america	positive = E	0.412	0.969	1.009	1.276
13223	negative = E, said	positive = E	0.706	0.970	1.010	1.311
13224	negative = E, sell	positive = E	0.303	0.972	1.012	1.401
13225	compani, time	positive = E	0.303	0.972	1.012	1.401
13230	negative = E, said, time	positive = E	0.303	0.972	1.012	1.401
13232	said, billion, time	positive = E	0.303	0.972	1.012	1.401
13238	negative = E, bank, said, billion, america	positive = E	0.303	0.972	1.012	1.401
13240	negative = E, said, plan	positive = E	0.307	0.972	1.012	1.421
13249	said, compani, group	positive = E	0.311	0.973	1.013	1.441

Con minsup=0.2

Premises	Conclusion	Support	Confidence	Lift	Conviction
stake	negative = E	0.232	0.914	1.047	1.475
servic	negative = E	0.246	0.889	1.018	1.145
intern	positive = E	0.272	0.984	1.025	2.487
friday	positive = E	0.259	0.983	1.024	2.368
euro	positive = E	0.250	0.983	1.023	2.289
initi	positive = E	0.250	0.983	1.023	2.289
told	positive = E	0.250	0.983	1.023	2.289
corp	positive = E	0.246	0.982	1.023	2.250
work	positive = E	0.246	0.982	1.023	2.250
talk	positive = E	0.237	0.982	1.022	2.171
ad	positive = E	0.232	0.981	1.022	2.132
cost	positive = E	0.232	0.981	1.022	2.132
set	positive = E	0.228	0.981	1.021	2.092
head	positive = E	0.224	0.981	1.021	2.053
monday	positive = E	0.215	0.980	1.020	1.974
thursday	positive = E	0.215	0.980	1.020	1.974
demand	positive = E	0.211	0.980	1.020	1.934
potenti	positive = E	0.206	0.979	1.019	1.895
rais	positive = E	0.338	0.975	1.015	1.559
part	positive = E	0.294	0.971	1.011	1.362
govern	positive = E	0.285	0.970	1.010	1.322
offer	positive = E	0.395	0.968	1.008	1.224
london	positive = E	0.250	0.966	1.006	1.164
put	positive = E	0.246	0.966	1.005	1.145
corpor	positive = E	0.241	0.965	1.005	1.125
partner	negative = E	0.202	0.920	1.054	1.590

Positive: offer, raise, finance, intern, Friday, euro-initial-told, corp-work, talk, add-cost, set, head, Monday, Thursday, demand, potential, part, govern, offer, London, put, corporation.

Negative: base, follow, (in)equity, company, group, deal, inc, stake, service, partner.

- Cluster 2: minsup= 0.4, minconf=0.5. La dimensione massima degli itemset è 6.
Argomento: tecnologie smartphone

Regole di associazione:

No.	Premises ↑	Conclusion	Support	Confidence	LaPlace	Lift	Conviction
1606	positive = E	appl, said, maker	0.508	0.558	0.789	1.047	1.057
1607	positive = E	appl, report, edit	0.508	0.558	0.789	1.047	1.057
1608	positive = E	negative = E, said, maker	0.508	0.558	0.789	1.080	1.094
1609	positive = E	negative = E, compani, inc	0.508	0.558	0.789	1.032	1.039
1610	positive = E	appl, said, report, edit	0.508	0.558	0.789	1.064	1.075
1860	positive = E	appl, googl	0.516	0.566	0.793	0.989	0.986
1861	positive = E	appl, month	0.516	0.566	0.793	0.989	0.986
1862	positive = E	negative = E, googl	0.516	0.566	0.793	1.003	1.004
1863	positive = E	compani, report	0.516	0.566	0.793	1.018	1.023
1864	positive = E	appl, negative = E, percent	0.516	0.566	0.793	0.989	0.986
1865	positive = E	appl, negative = E, maker	0.516	0.566	0.793	1.080	1.097
1866	positive = E	said, inc, report	0.516	0.566	0.793	1.064	1.079
1867	positive = E	appl, negative = E, said, year	0.516	0.566	0.793	1.003	1.004
1868	positive = E	appl, negative = E, said, devic	0.516	0.566	0.793	1.003	1.004

Con minsup=0.3

Premises	Conclusion	Support	Confidence	Lift	Conviction
corp	positive = E	0.387	0.960	1.053	2.218
launch	positive = E	0.395	0.942	1.034	1.538
run	negative = E	0.379	0.959	1.017	1.383
phone	negative = E	0.484	0.952	1.009	1.185

Positive: maker, edit, launch, electron.

Negative: firm (apple operation), accord, operation, use, run, phone.

- Cluster 3: minsup:0.1, minconf:0.5. La dimensione massima degli itemset è 8.
Argomento: social network e amazon

Regole di associazione:

No.	Premises ↑	Conclusion	Support	Confidence	Lift	Conviction
98	positive = E	said, report	0.533	0.564	1.033	1.042
99	positive = E	negative = E, said, edit	0.533	0.564	1.052	1.064
131	positive = E	negative = E, year	0.552	0.585	0.973	0.960
140	positive = E	said, edit	0.556	0.588	1.053	1.071
171	positive = E	negative = E, compani	0.572	0.606	0.996	0.994
172	positive = E	negative = E, edit	0.572	0.606	1.053	1.077
179	positive = E	year	0.575	0.609	0.971	0.953
185	positive = E	negative = E, report	0.578	0.612	1.019	1.029
222	positive = E	compani	0.592	0.626	0.998	0.997
232	positive = E	edit	0.595	0.630	1.053	1.086
246	positive = E	report	0.601	0.637	1.020	1.034
434	positive = E	negative = E, said	0.758	0.803	1.011	1.044
465	positive = E	said	0.781	0.827	1.012	1.058
733	positive = E	negative = E	0.922	0.976	1.002	1.079

For the next following, confidence is > 0.8

No.	Premises	Conclusion ↑	Support	Lift	Conviction
91924	follow	negative = E	0.255	1.001	1.046
91925	earlier	negative = E	0.127	1.001	1.046
91926	photo	negative = E	0.127	1.001	1.046
91927	store	negative = E	0.127	1.001	1.046
92074	generat	negative = E	0.131	1.002	1.072
92075	rate	negative = E	0.131	1.002	1.072
92076	televis	negative = E	0.131	1.002	1.072
92209	critic	negative = E	0.134	1.002	1.098
92210	softwar	negative = E	0.134	1.002	1.098
92211	watch	negative = E	0.134	1.002	1.098
92231	san, francisco	negative = E	0.134	1.002	1.098
92298	state	negative = E	0.275	1.003	1.124
92299	hous	negative = E	0.137	1.003	1.124

For the next following confidence is =1

92981	meet	negative = E	0.137	1.027	∞
92982	known	negative = E	0.124	1.027	∞
92983	california	negative = E	0.121	1.027	∞
92984	confer	negative = E	0.121	1.027	∞
92985	area	negative = E	0.114	1.027	∞
92986	collect	negative = E	0.114	1.027	∞
92987	great	negative = E	0.114	1.027	∞
92988	job	negative = E	0.114	1.027	∞
92989	full	negative = E	0.111	1.027	∞
92990	washington	negative = E	0.111	1.027	∞
92991	corpor	negative = E	0.108	1.027	∞
92992	join	negative = E	0.108	1.027	∞
92993	valu	negative = E	0.108	1.027	∞
92994	further	negative = E	0.101	1.027	∞

Positive: april-governance-hold,official, pay, capital, statement, former, Tuesday, Thursday, edit, investor.

Negative: rate-television, critic-software –watch, feature-San Francisco-retail-team, state, house, Monday, platform-popular-trade, show, repression, ceo-file, global, hope, intern-raise, march, change, apple.

- **Cluster 4**: minsup=0.3, minconf=0.5. La dimensione massima degli itemset è 6.
Argomento: legislazione.

Regole di associazione:

No.	Premises ↑	Conclusion	Support	Confidence	Lift	Conviction
1112	negative = E	case, inc	0.308	0.552	1.125	1.137
1115	negative = E	positive = E, court, report	0.308	0.552	0.911	0.879
1125	negative = E	positive = E, case, compani	0.308	0.552	1.007	1.008
1129	negative = E	positive = E, case, inc	0.308	0.552	1.125	1.137
1146	negative = E	said, report, edit	0.308	0.552	0.989	0.987
1186	negative = E	positive = E, said, report, edit	0.308	0.552	0.989	0.987
1551	negative = E	said, compani	0.317	0.569	0.954	0.937
1555	negative = E	court, year	0.317	0.569	1.020	1.026
1558	negative = E	court, inc	0.317	0.569	1.138	1.160
1560	negative = E	case, year	0.317	0.569	1.020	1.026
1562	negative = E	report, edit	0.317	0.569	1.003	1.004
1563	negative = E	positive = E, said, compani	0.317	0.569	0.954	0.937
1571	negative = E	positive = E, court, year	0.317	0.569	1.020	1.026
1577	negative = E	positive = E, court, inc	0.317	0.569	1.138	1.160

Con minsup=0.2

Premises	Conclusion	Support	Confidence	Lift	Conviction
googl	negative = E	0.250	0.812	1.457	2.359

Con minsup=0.1

Premises	Conclusion	Support	Confiden...	Lift	Conviction
search	negative = E	0.135	0.875	1.569	3.538

Positive: -.

Negative: inc, edit, google, search.

- Cluster 5: minsup=0.4, minconf=0.5. Le dimensioni degli itemset variano da 1 a 8.
Argomento: Compagnie aeree.

Regole di associazione:

No.	Premises ↑	Conclusion	Support	Confidence	Lift	Conviction
3289	positive = E	airlin, report, delta	0.552	0.556	1.007	1.008
3290	positive = E	said, report, delta	0.552	0.556	1.007	1.008
3291	positive = E	negative = E, air, line	0.552	0.556	1.007	1.008
3303	positive = E	airlin, said, negative = E, flight	0.552	0.556	1.007	1.008
3304	positive = E	airlin, said, negative = E, delta	0.552	0.556	1.007	1.008
3305	positive = E	airlin, said, report, airport	0.552	0.556	1.007	1.008

Con minsup= 0.3:

Premises	Conclusion	Support	Confidence	Lift	Conviction
said, air, percent	positive = E, southwest, unit, delta	0.325	0.641	1.702	1.737
said, air, percent	positive = E, airlin, southwest, unit, delta	0.325	0.641	1.702	1.737
airlin, said, air, percent	positive = E, southwest, unit, delta	0.325	0.641	1.702	1.737
air, percent	positive = E, southwest, unit, delta	0.331	0.646	1.714	1.759
air, percent	positive = E, airlin, southwest, unit, delta	0.331	0.646	1.714	1.759
airlin, air, percent	positive = E, southwest, unit, delta	0.331	0.646	1.714	1.759
line, percent	positive = E, said, southwest, unit, delta	0.318	0.662	1.821	1.884
line, percent	positive = E, airlin, said, southwest, unit, delta	0.318	0.662	1.821	1.884
airlin, line, percent	positive = E, said, southwest, unit, delta	0.318	0.662	1.821	1.884
line, percent	positive = E, said, southwest, air, unit, delta	0.318	0.662	1.821	1.884
line, percent	positive = E, airlin, said, southwest, air, unit, delta	0.318	0.662	1.821	1.884
airlin, line, percent	positive = E, said, southwest, air, unit, delta	0.318	0.662	1.821	1.884
said, line, percent	positive = E, southwest, unit, delta	0.318	0.671	1.782	1.896
said, line, percent	positive = E, airlin, southwest, unit, delta	0.318	0.671	1.782	1.896

Positive: -.

Negative: edit (report flight), jet, inc, chief, price, market, executive, hold, group.

- Cluster 6: minsup=0.6, minconf=0.5. La dimensione massima degli itemset è 6.
Argomento: investimenti

Regole di associazione:

No.	Premis... ↑	Conclusion	Support	Confidence	Lift	Conviction
542	positive = E	compani, percent	0.688	0.710	1.002	1.005
543	positive = E	percent, edit	0.688	0.710	1.002	1.005
544	positive = E	percent, expect	0.688	0.710	1.002	1.005
545	positive = E	negative = E, year, percent	0.688	0.710	1.002	1.005

Con minsup= 0.4

Premises	Conclusion	Support	Confidence	Lift	Conviction
report	negative = E, year, percent, profit	0.448	0.551	1.103	1.114
report	negative = E, year, said, profit	0.448	0.551	1.151	1.161
report	negative = E, year, share, profit	0.448	0.551	1.151	1.161
report	negative = E, year, quarter, inc	0.448	0.551	1.203	1.207
report	negative = E, year, inc, million	0.448	0.551	1.203	1.207
report	negative = E, compani, percent, billion	0.448	0.551	1.151	1.161
report	negative = E, compani, edit, inc	0.448	0.551	1.176	1.184
report	negative = E, compani, expect, inc	0.448	0.551	1.203	1.207
report	negative = E, percent, share, profit	0.448	0.551	1.151	1.161
report	negative = E, percent, expect, billion	0.448	0.551	1.176	1.184
report	negative = E, percent, million, analyst	0.448	0.551	1.203	1.207
report	negative = E, said, share, profit	0.448	0.551	1.176	1.184
report	negative = E, said, edit, inc	0.448	0.551	1.176	1.184
report	negative = E, said, expect, market	0.448	0.551	1.203	1.207

Positive: percent, expect, invest, time, unit.

Negative: corporative-service, comparison, stock.

- Cluster 7: minsup=0.6, minconf=0.5. La dimensione massima degli itemset è 8.
Argomento: indici di mercato.

Regole di associazione:

No.	Premises ↑	Conclusion	Support	Confidence	Lift	Conviction
2689	positive = E	york	0.604	0.669	1.019	1.038
2690	positive = E	percent, york	0.604	0.669	1.019	1.038
2692	positive = E	said, jone	0.604	0.669	1.008	1.016
2733	positive = E	stock, index, week	0.604	0.669	1.043	1.084
2734	positive = E	stock, report, year	0.604	0.669	1.008	1.016
2735	positive = E	stock, industri, dow	0.604	0.669	1.019	1.038
2736	positive = E	stock, industri, jone	0.604	0.669	1.019	1.038
2737	positive = E	stock, dow, nasdaq	0.604	0.669	1.019	1.038
3321	positive = E	industri, dow	0.612	0.678	1.020	1.042
3322	positive = E	industri, jone	0.612	0.678	1.020	1.042
3323	positive = E	dow, nasdaq	0.612	0.678	1.020	1.042
3324	positive = E	percent, stock, nasdaq	0.612	0.678	1.020	1.042
3326	positive = E	percent, market, year	0.612	0.678	0.976	0.949
3328	positive = E	percent, market, averag	0.612	0.678	0.976	0.949
3330	positive = E	percent, report, industri	0.612	0.678	0.998	0.996
3332	positive = E	percent, industri, dow	0.612	0.678	1.020	1.042
3334	positive = E	percent, industri, jone	0.612	0.678	1.020	1.042

Positive: dow-trade-data, jone, nasdaq, investor, stock, expect, New York, fell (percent), week (percent).

Negative: rose (percent), point (percent), gain (percent).

- Cluster 8: minsup=0.2, minconf=0.5. La dimensione massima degli itemset è 7.
Argomento: governo.

Regole di associazione:

No.	Premises ↑	Conclusion	Support	Confidence	Lift	Conviction
23927	positive = E	govern, edit	0.542	0.583	1.039	1.053
25983	positive = E	said, report, edit	0.548	0.590	1.064	1.086
32178	positive = E	report, edit	0.574	0.618	1.064	1.098
44485	positive = E	said, state	0.626	0.674	0.985	0.969
46808	positive = E	said, year	0.639	0.688	0.969	0.929
49504	positive = E	said, report	0.652	0.701	1.016	1.037
52318	positive = E	said, edit	0.665	0.715	1.046	1.110
53620	positive = E	said, govern	0.671	0.722	1.009	1.022
54923	positive = E	state	0.677	0.729	0.991	0.977
57323	positive = E	edit	0.690	0.743	1.047	1.130
60112	positive = E	year	0.703	0.757	0.978	0.929
61110	positive = E	govern	0.710	0.764	1.012	1.038
61111	positive = E	report	0.710	0.764	1.021	1.066
82583	positive = E	said	0.858	0.924	0.994	0.929

Positive: provide, week, report, lead (year twitter), street, china, twitter, additional, appear, detail, Tuesday, edit, inform-system, effort.

Negative: follow.

Very negative: ani.