

Expanding Our Text Mining Toolkit

Sentiment Analysis and Term Selection in JMP Pro 16

Ross Metusalem

JMP Academic Ambassador

ross.metusalem@jmp.com

Text Explorer

Latent Class Analysis for 5 Clusters

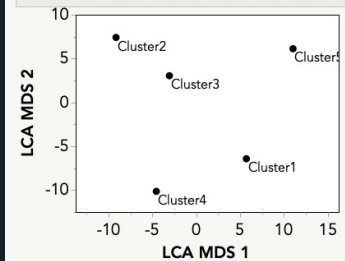
Using 2296 terms across 436 documents. Specified Maximum Number of Terms = 1000. Minimum Term Frequency=10

BIC 486939.04 [Show Text](#)

Top Terms by Cluster

Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Term	Score	Term	Score	Term
services	13.762	continues	7.4816	strength
tourism	12.442	prospects	6.8227	continues
costs	12.292	investment	6.7868	defense
wage	12.006	picture	6.4765	weakness
hiring	11.79	may	6.2545	reduced
workers	11.602	major	6.2061	apparel
quality	11.537	plant	5.9234	nonresidential
contacts	11.422	large	5.9123	yields
indicated	11.343	businessmen	5.9003	farm
labor	11.272	funds	5.8745	agriculture

MDS Plot



Text Explorer for text

Number of Terms	Number of Cases	Total Tokens	Tokens per Case	Number of Non-Empty Cases	Portion of Non-Empty Cases
7264	436	580828	1332.17	436	1.0000

Term and Phrase Lists

Term	Count	Phrase	Count	N
activity	4947	kansas city	3724	2
sales	4867	san francisco	3582	2
demand	3810	new york	2995	2
prices	3517	st louis	2774	2
construction	2520	real estate	1217	2
strong	2427	loan demand	769	2
increased	2227	retail sales	723	2
growth	2139	economic activity	562	2
manufacturing	1844	manufacturing activity	540	2
increases	1813	commercial real estate	508	3

Word Cloud



Context

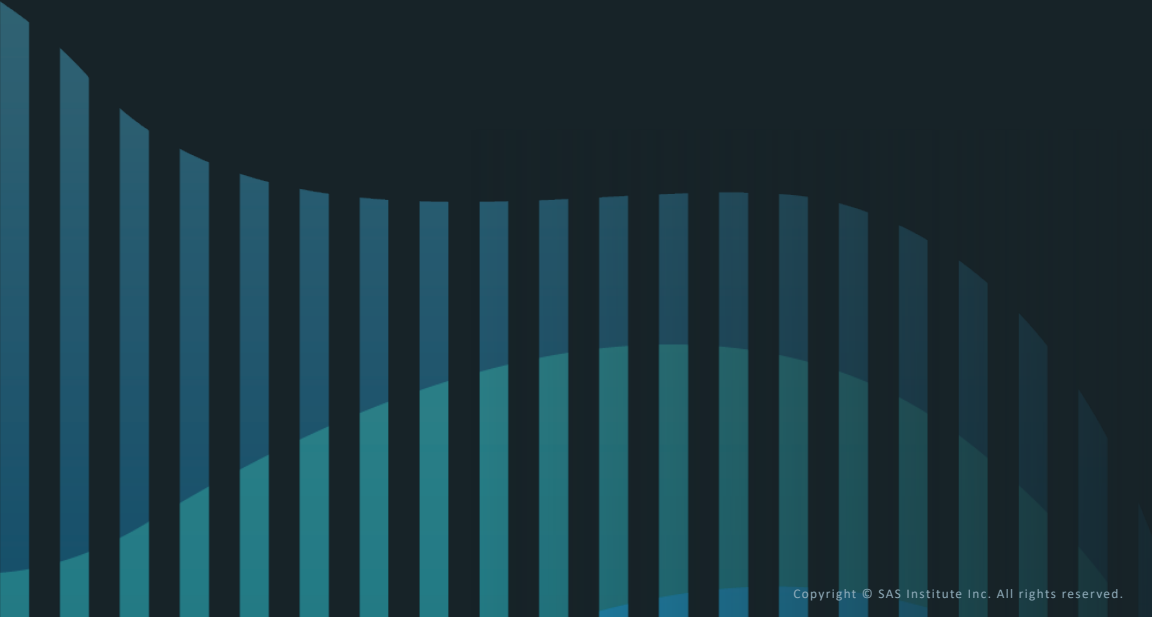
Run Script Debug Script Reformat Script Balance

Words: **activity** demand strong growth conditions

Book: National Summary }} This initial economic **conditions** in the 12 Federal Reserve is based on information gathered from of the Reserve Banks, conversations with local businessmen and economists, regular monthly of manufacturing and trade industries conducted of the Reserve Banks, and selected statistical of regional economic **activity**. Reports Reserve Banks clearly indicate that the overriding domestic concern is inflation. hen generally expect that prices will continue to at a rapid rate during the remainder ear. There appears to be considerable sm regarding the ability of economic ation to achieve a significant reduction in the rate tion without generating an intolerable unemployment or a full-scale recession. y, there is evidence of extensive concern about istence of strong upward wage pressures,

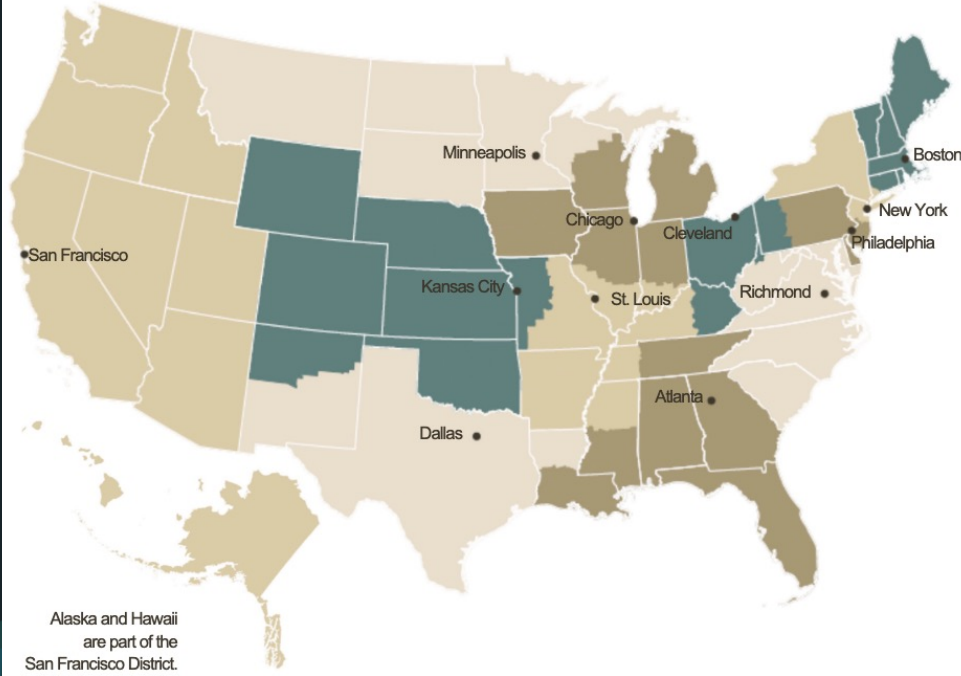
Sentiment Analysis

Quantifying positive-negative emotion in unstructured text



The Beige Book

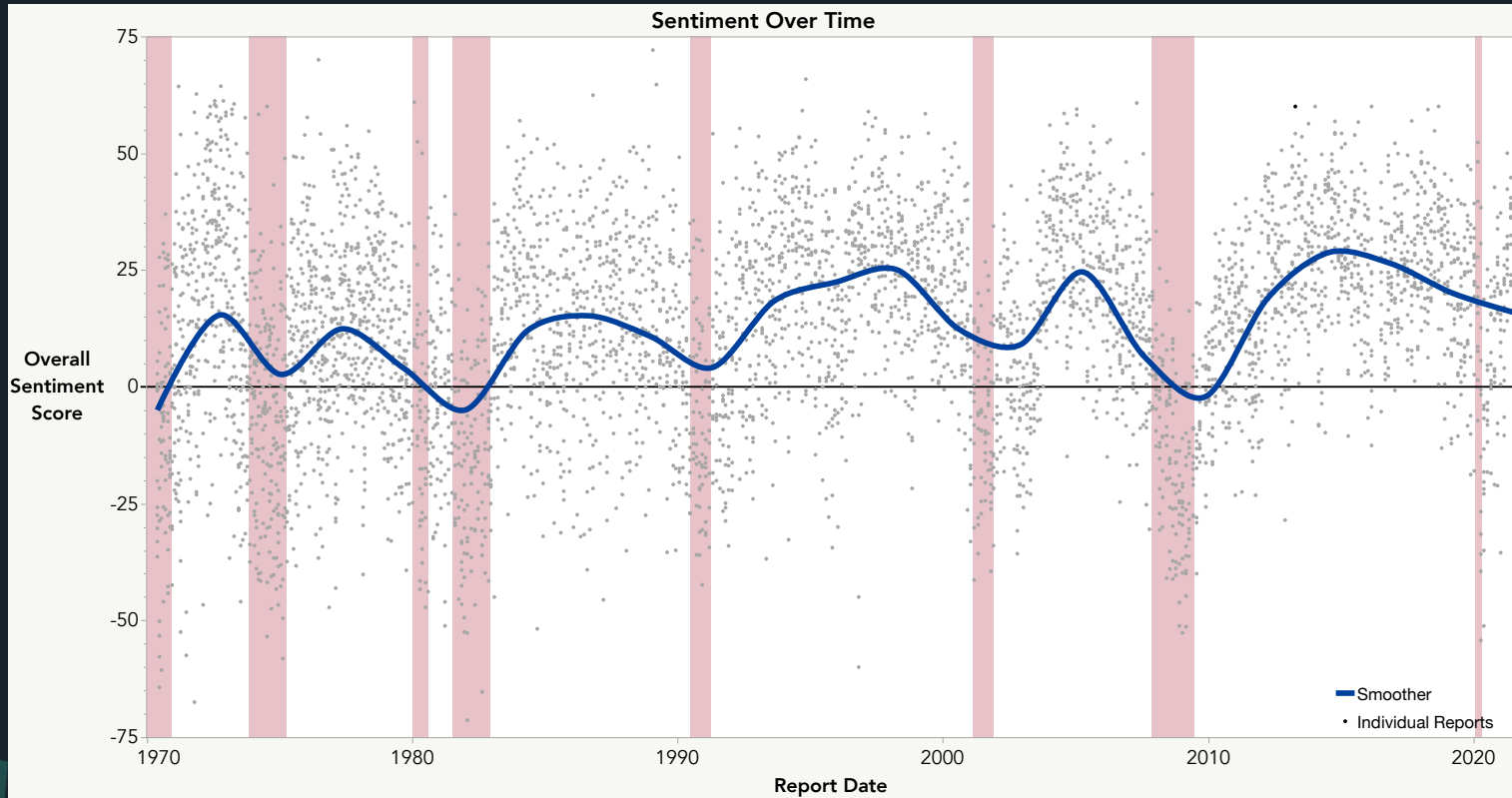
Summary of Commentary on Current Economic Conditions
By Federal Reserve District



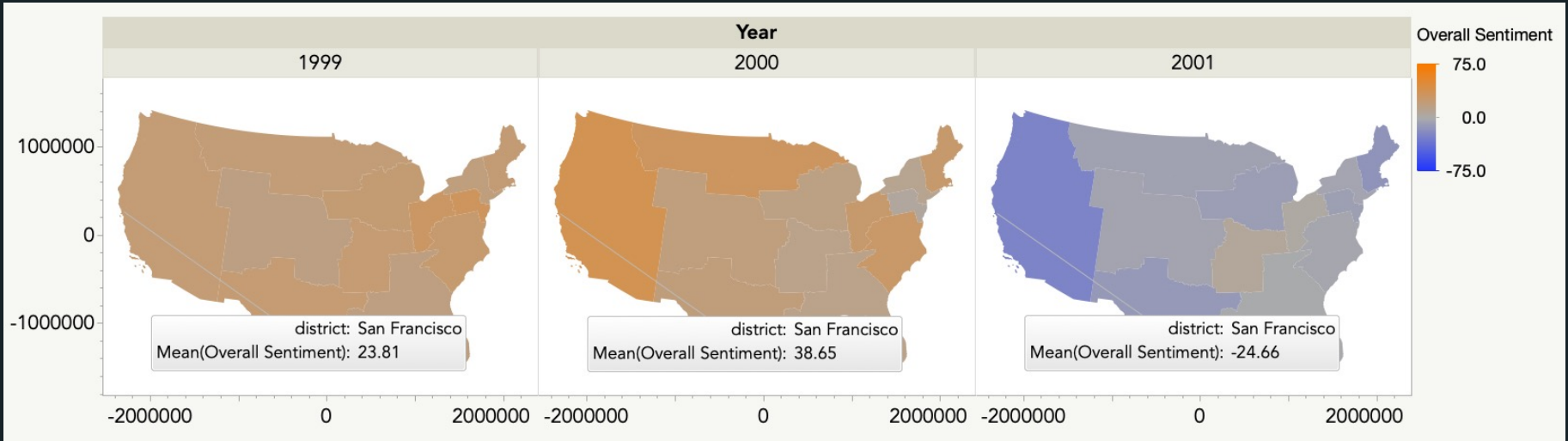
locations by major retail chains. Computer sales, on the other hand, have been severely depressed and are not expected to recover until yearend or early 1972. A second director, who is the chairman of a large diversified manufacturing firm (major divisions include auto components, defense and space products, and electronics), expressed the view that economic activity will probably improve during 1971. He also reported that there are a number of favorable "straws in the wind" from his firm's point of view:

there is considerable agreement that (1) consumers remain pessimistic, (2) outlays for plant and equipment are being cut back or deferred, (3) labor markets are easing but wage costs are not, and (4) growth in the money supply is too fast. Reports from around the District indicate widespread pessimism on the part of consumers. Large department stores in the region report poor sales for large luxury items all the way down to small inexpensive goods. One department store executive quipped that "... only

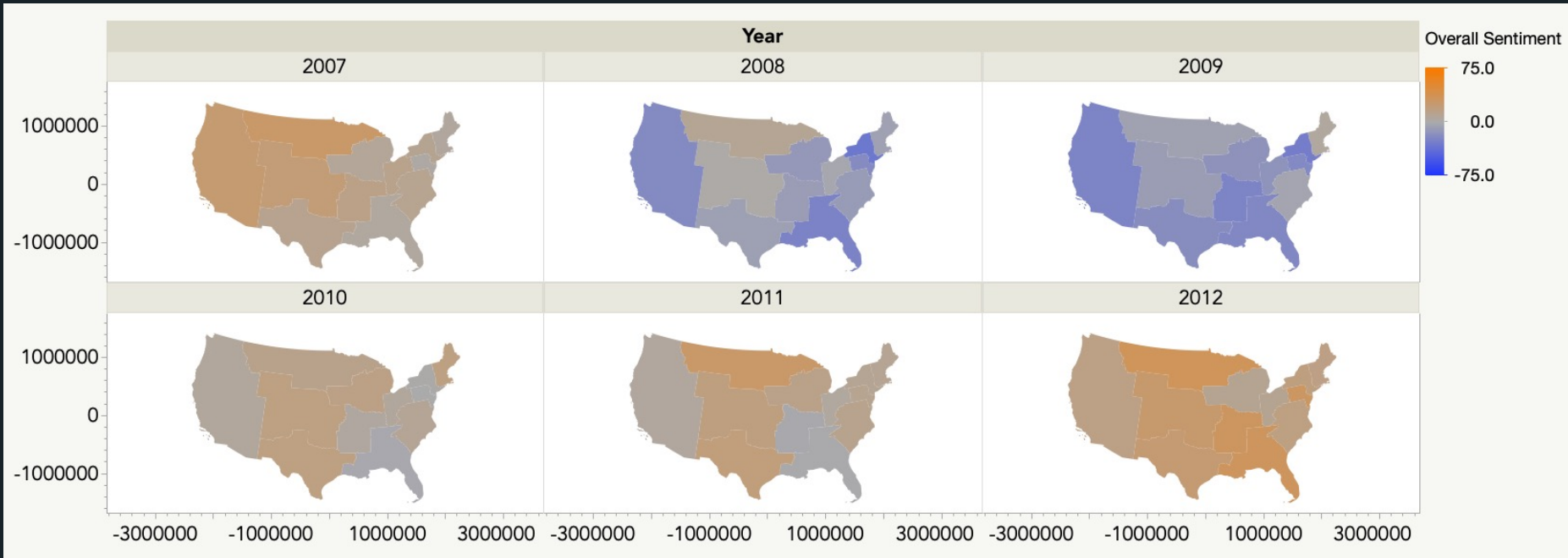
Sentiment Tracks with Recessions



Sentiment and the Dotcom Bust



Sentiment and the Great Recession



Lexical Sentiment Analysis

Sentiment Score



Intensifier



Negator (-1)

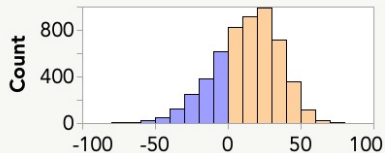
Term	Score
fantastic	90
favorable	40
feared	-60
fortunate	40
friendly	40
functional	20
garbage	-60
glad	60
good	60
gracious	70
great	80
greatest	90
grim	-75

Term	Multiplier
incredibly	1.40
insanely	1.90
little	0.30
mightily	1.60
more	1.20
most	1.50
much	1.20
not only	1.00
not the only	1.00
only	0.30
outrageously	1.70
over	1.00
overly	1.00

Term
no
non
none
nor
not
shouldn't
shouldn't
wasn't
wasn't
weren't
weren't
without
won't

▼ Sentiment Summary

	N	Mean Score
All Scored Documents	5339	4.1
Net Positive Documents	3902	22.9
Net Negative Documents	1415	-15.4
No Sentiment Documents	1	0.0



Scores of Documents with Net Sentiment

Document	Positive Sum	Positive Score Mean	Negative Sum	Negative Score Mean	Overall Score	Sentiment	Score	Count
1	338	68	-660	-60	-20	strong	60	14396
2	90	45	-490	-54	-36	improved	60	4633
3	438	63	-446	-64	-1	weak	-60	4627
4	350	70	-290	-48	5	good	60	4172
5	136	45	-530	-53	-30	stable	30	3702
6	325	54	-335	-56	-1	slow	-30	3154
7	80	40	-300	-50	-28	optimistic	50	2991
8	120	60	-352	-70	-33	concern	-60	2696
9	440	63	-432	-62	1	positive	60	2461
10	85	43	-240	-60	-26	weakness	-60	2072
						stronger	60	2048

[3] ...near at hand. Some bankers think loan demand remains as strong as ever; others believe demand has eased. The truckers' strikes...in some cases. Although output schedules for autos indicate a strong revival in May and June, pessimism in automotive centers has...

[4] ...in steel to deteriorate later in the year, with a strong increase in imports and an easing in exports. On the...

+SentimentxIntensifier

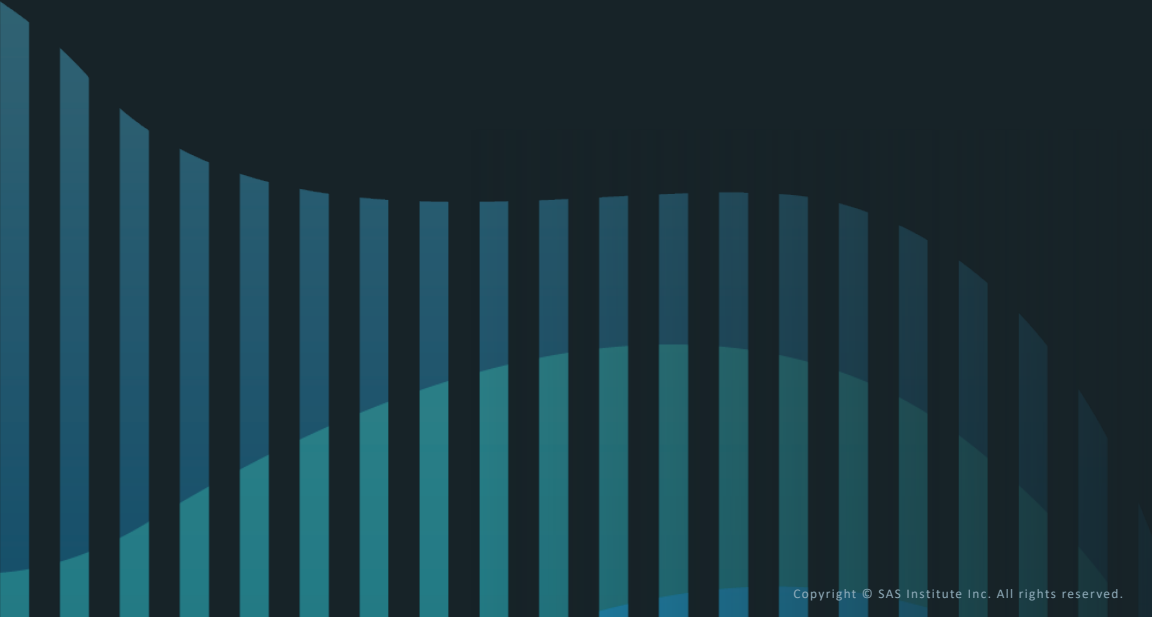
- +90 x1.6
- +60 x1.2
- +30 x0.8
- 30 x0.3
- 60
- 90



Let's see it in JMP

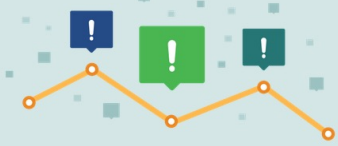
Term Selection

Identify words strongly associated with an outcome



Consumer Complaint Database

This database is a collection of complaints about consumer financial products and services that we sent to companies for response.



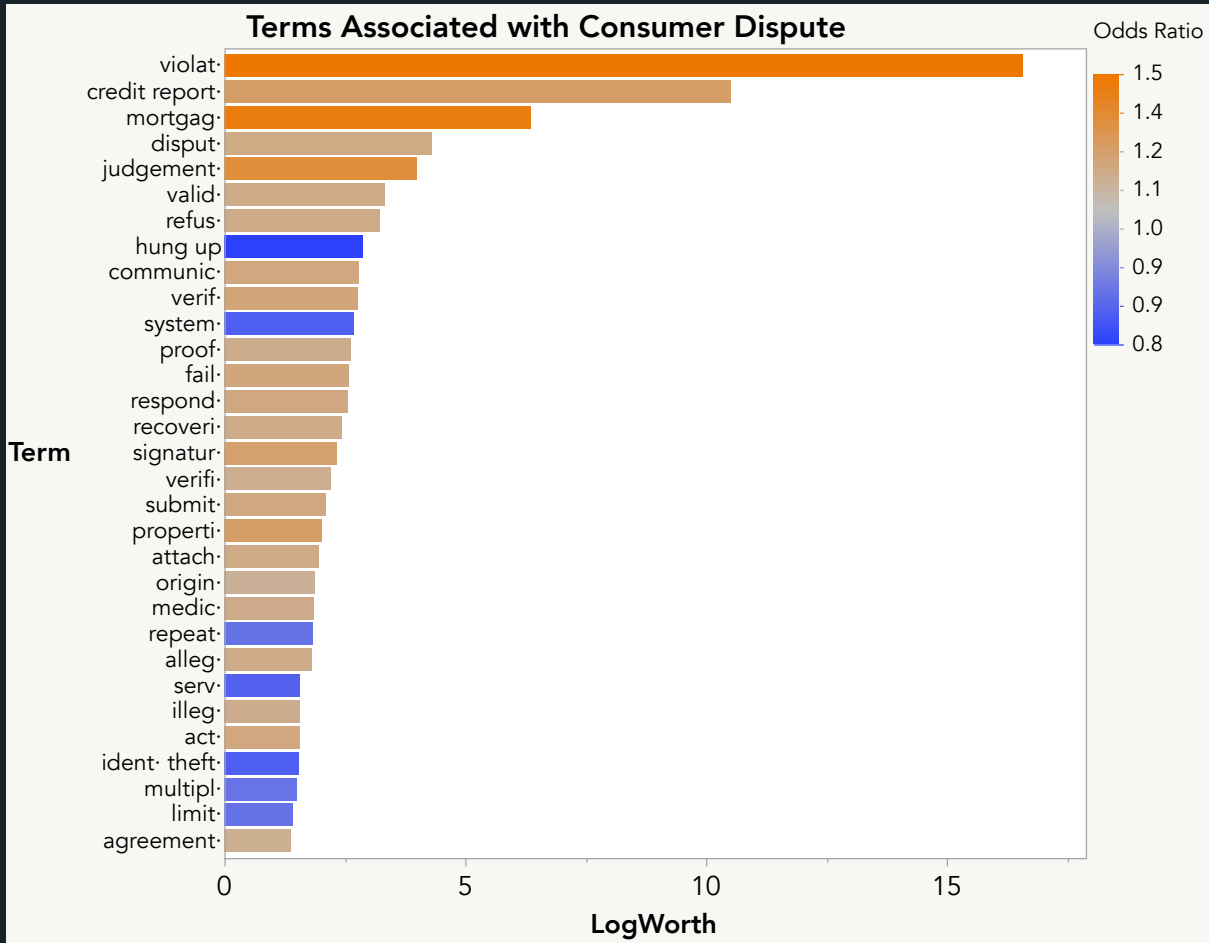
CFPB Debt Collection Complaints for Demoiing Term Selection JMP

Complaint ID	Date received	Consumer complaint narrative	Consumer disputed?
1	2265295 2016-12-28	About a year and a half ago I purchased a membe...	No
2	2436870 2017-04-15	I worked as a manager for a cell phone store. The ...	No
3	2035377 2016-07-28	This payday loan was opened back in 2012. I joine...	No
4	1991793 2016-06-29	XXXX/XXXX/XXXX I received a letter from ARA, In...	No
5	2284673 2017-01-11	To the Consumer Financial Protection Bureau : I X...	No
6	1301896 2015-03-26	This company continues to report on my credit re...	No
7	2200388 2016-11-08	Midwest Recovery Systems recently added them...	No
8	2426542 2017-04-10	I XXXX makes a claim with the Consumer Financia...	No
9	2401000 2017-03-23	I do not owe a balance with XXXX XXXX XXXX XX...	Yes
10	2195488 2016-11-07	Got an alert from credit monitoring for a new colle...	No
11	2194978 2016-11-06	I received a notice from XXXX on a new collection...	Yes
12	2009520 2016-07-12	I recently checked my experian report and noticed...	Yes
13	1999874 2016-07-06	ARA has called me numerous times harrassing me...	No
14	2197850 2016-11-07	I received a new account notification from a credit...	No
15	2010655 2016-07-13	I monitor my credit report, more frequently now a...	Yes
16	1737808 2016-01-12	I noticed on all XXXX of my credit reports that the...	No
17	2197689 2016-11-07	I recently went on my credit report and saw I have...	No
18	1925472 2016-05-13	On XXXX/XXXX/XXXX : Notified of unexpected di...	No
19	2237471 2016-12-06	XXXX XXXX called me regarding a debt that was ...	No
20	1445830 2015-06-30	My identity was stolen back in XX/XX/2012. I mad...	No
21	1602300 2015-10-10	There is a collection account on my credit for \$29...	Yes
22	1833657 2016-03-15	I have been paying a debt collector since XX/XX/...	No
23	2033485 2016-07-27	The item is FMS (Premiere Property), it should ne...	No

I monitor my credit report, more frequently now as we're attempting to buy our first house. A collection notice was filed for the amount of {\$3900.00}. We have received no notification of this debt, no verification of this debt, and consequently no notice of right to dispute this debt.

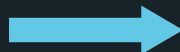
* Consumer disputed company's response

What types of complaints lead to consumer disputes?



Document-Term Matrix

Complaint ID	report-Binary	contact-Binary	letter-Binary	credit report-...
2.56M	1	1	1	1
1.29M	0	0	0	0
2265295	1	0	0	0
2436870	0	0	0	0
2035377	0	1	1	1
1991793	1	1	1	0
2284673	0	0	1	0
1301896	1	0	1	1
2200388	0	1	0	1
2426542	0	0	1	1
2401000	0	0	0	0
2195488	0	0	0	0
2194978	0	0	0	1
2009520	1	0	0	1
1999874	1	0	0	0
2197850	0	0	0	1
2010655	0	0	0	1



Generalized Regression

Model Specification

Select Columns

218 Columns

Q Enter column name

- Complaint ID
- Date received
- Product
- Sub-product
- Issue
- Sub-issue
- Consumer complaint narrative
- Company public response
- Company
- State
- ZIP code
- Tags
- Consumer consent provided?
- Submitted via
- Date sent to company
- Company response to consumer
- Timely response?
- Consumer disputed?
- Document Terms by Binary (200/0)

Pick Role Variables

Y Consumer disputed?

Weight optional numeric

Freq optional numeric

Validation optional numeric

By optional

Construct Model Effects

Add report: Binary

Cross contact: Binary

Nest letter: Binary

Macros credit report: Binary

Degree 2 bill: Binary

Attributes provid: Binary

Transform disput: Binary

No Intercept paid: Binary

servic: Binary

name: Binary

Personality: Generalized Regression

Distribution: Binomial

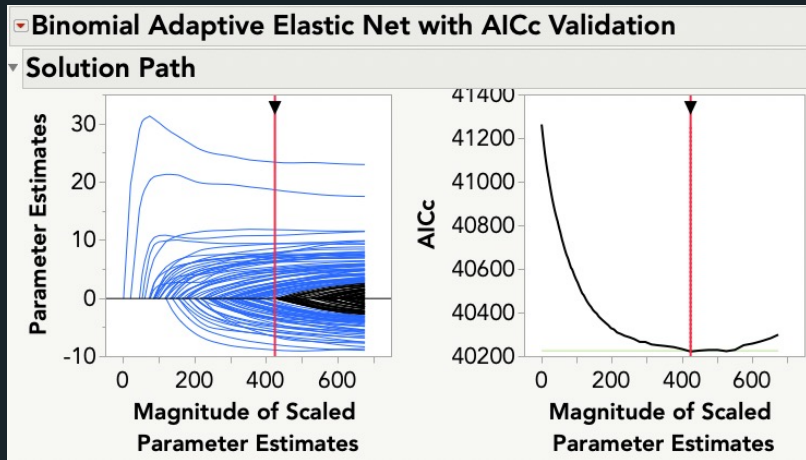
Target Level: Yes

Help Run

Recall Keep dialog open

Remove

$$\ln\left(\frac{P(\text{Dispute})}{P(\text{NOT Dispute})}\right) = -1.7 + 0.4 * \text{violate} - 0.2 * \text{hung up} \dots$$



Term	Coefficient	LogWorth	Count
violate	0.404	16.580	5328
credit report	0.209	10.501	17579
mortgag	0.382	6.363	2020
disput	0.142	4.302	11679
judgement	0.295	3.987	1690
valid	0.137	3.334	9697
refus	0.134	3.228	6340
hung up	-0.220	2.883	2075
communic	0.159	2.798	3515
verif	0.172	2.769	3103
system	-0.165	2.693	3781



Let's see it in JMP

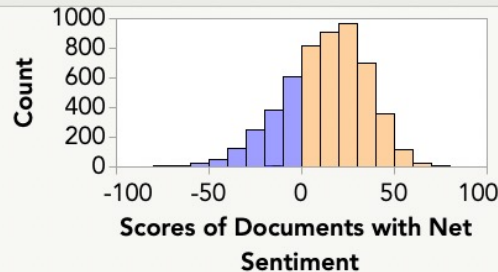
Term Selection

Identify words strongly associated with an outcome

Term	Coefficient	LogWorth	Count
violat·	0.404	16.580	5328
credit report·	0.209	10.501	17579
mortgag·	0.382	6.363	2020
disput·	0.142	4.302	11679
judgement·	0.295	3.987	1690
valid·	0.137	3.334	9697
refus·	0.134	3.228	6340
hung up	-0.220	2.883	2075
communic·	0.159	2.798	3515

Sentiment Summary

	N	Mean Score
All Scored Documents	5267	4.1
Net Positive Documents	3836	22.8
Net Negative Documents	1409	-15.4
No Sentiment Documents	1	0.0



Sentiment Analysis

Quantify positive-negative emotion in texts