Assignment 2

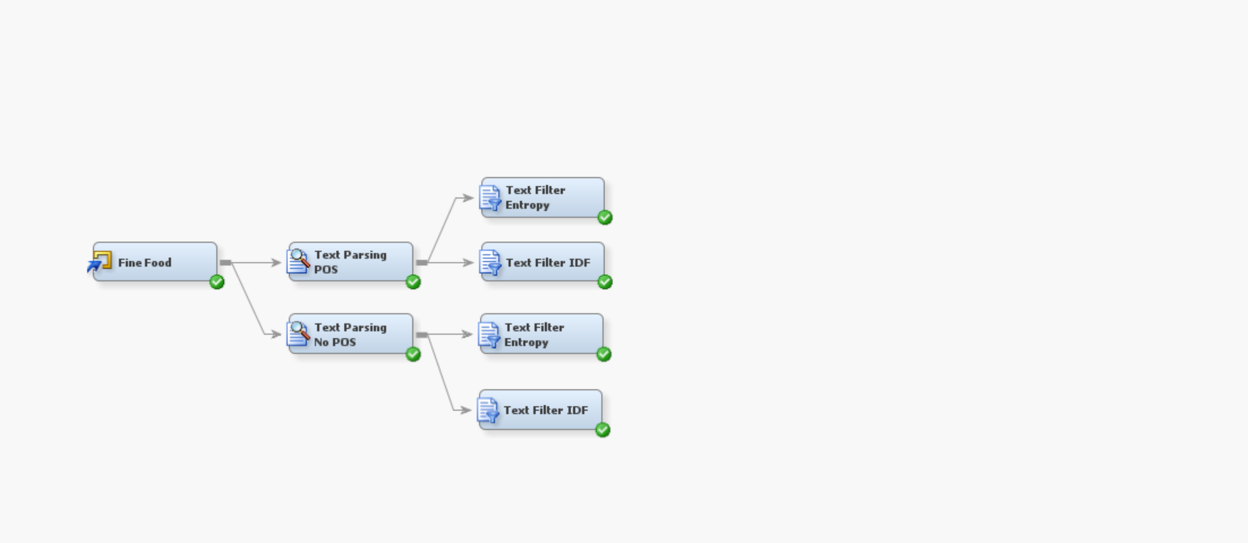
The goal of this assignment is to apply text parsing and filtering in real context. To do so do the following:

1- Go to the dataset: https://www.kaggle.com/snap/amazon-fine-food-reviews/downloads/amazon-fine-food-reviews.zip/2

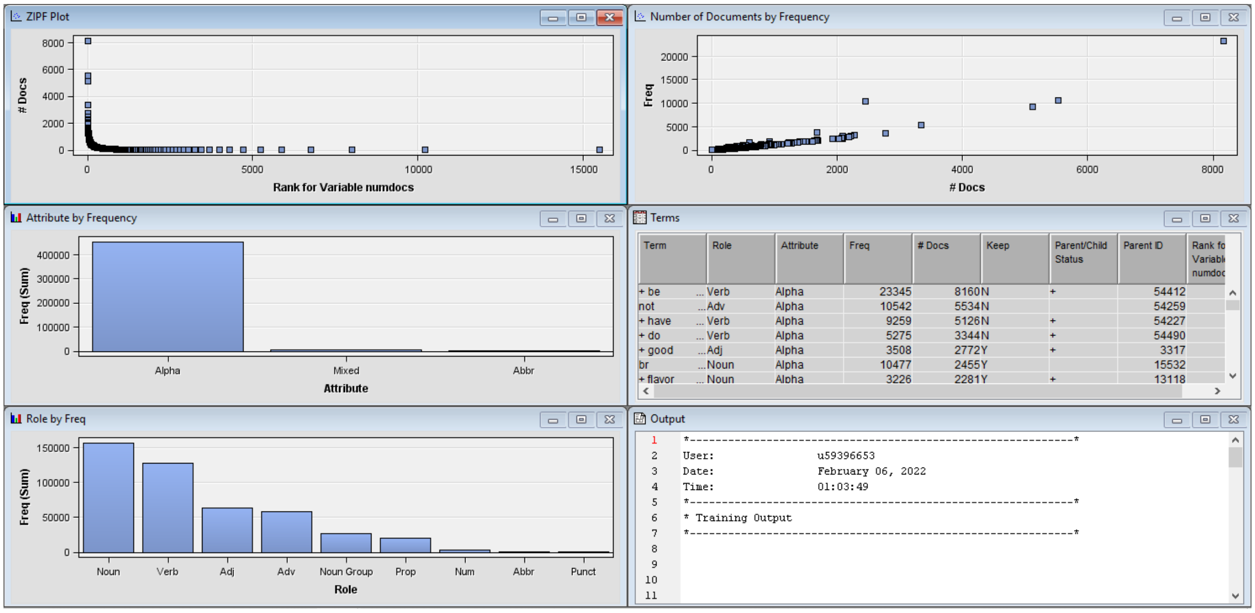
(You may need to register to Kaggle.com to download the dataset)

2- Import the data file to SAS enterprise miner

3- Apply text parsing and filtering on the dataset.



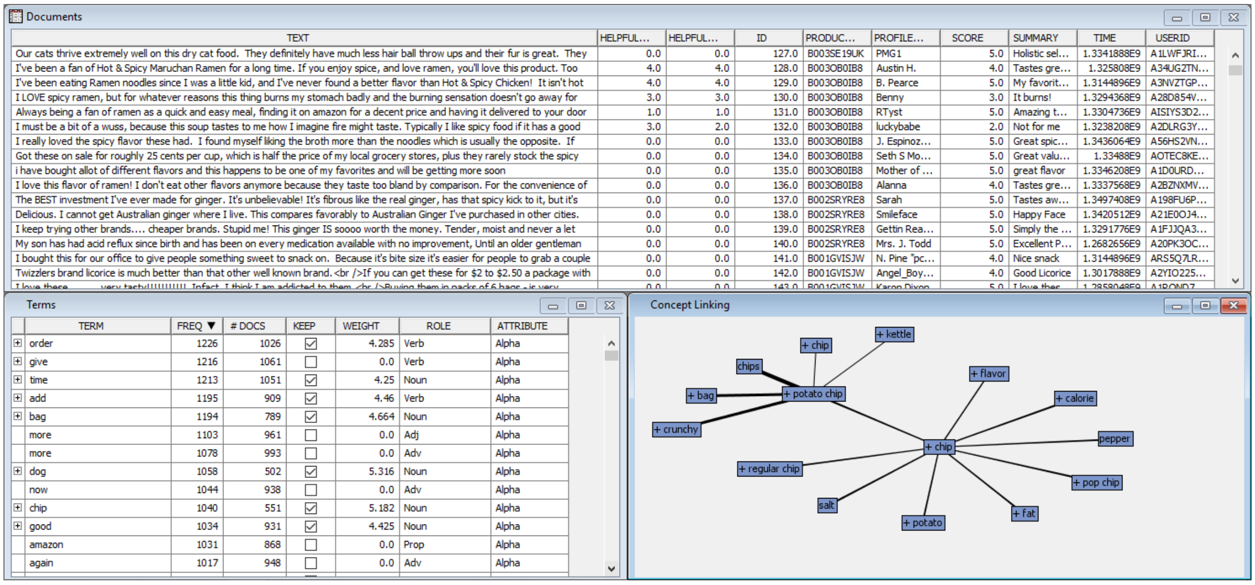
4- Take screen captures of the outcome of your graphs and findings



Term	Role	Attribute	Freq	# Docs	Keep	Parent/Child Status	Parent ID	Rank fo Variabl numdoc
+ be ...	Verb	Alpha	23345	8160	N	+	54412	
not ...	Adv	Alpha	10542	5534	N		54259	
+ have ...	Verb	Alpha	9259	5126	N	+	54227	
+ do ...	Verb	Alpha	5275	3344	N	+	54490	
+ good ...	Adj	Alpha	3508	2772	Y	+	3317	
br ...	Noun	Alpha	10477	2455	Y		15532	
+ flavor ...	Noun	Alpha	3226	2281	Y	+	13118	

```
*------------------------------------------------------------*
User:                   u59396653
Date:                   February 06, 2022
Time:                   01:03:49
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*
```

Figure 1 Text Parsing POS Results



TEXT	HELPFUL...	HELPFUL...	ID	PRODUC...	PROFILE...	SCORE	SUMMARY	TIME	USERID
Our cats thrive extremely well on this dry cat food. They definitely have much less hair ball throw ups and their fur is great. They	0.0	0.0	127.0	B003SE19UK	PMG1	5.0	Holistic sel...	1.3341888E9	A1LWFJRI...
I've been a fan of Hot & Spicy Maruchan Ramen for a long time. If you enjoy spice, and love ramen, you'll love this product. Too	4.0	4.0	128.0	B003OB0IB8	Austin H.	4.0	Tastes gre...	1.325808E9	A34UG2TN...
I've been eating Ramen noodles since I was a little kid, and I've never found a better flavor than Hot & Spicy Chicken! It isn't hot	4.0	4.0	129.0	B003OB0IB8	B. Pearce	5.0	My favorit...	1.3144896E9	A3NVZTGP...
I LOVE spicy ramen, but for whatever reasons this thing burns my stomach badly and the burning sensation doesn't go away for	3.0	3.0	130.0	B003OB0IB8	Benny	3.0	It burns!	1.3294368E9	A28D854V...
Always being a fan of ramen as a quick and easy meal, finding it on amazon for a decent price and having it delivered to your door	1.0	1.0	131.0	B003OB0IB8	RTyst	5.0	Amazing t...	1.3304736E9	AISIYS3D2...
I must be a bit of a wuss, because this soup tastes to me how I imagine fire might taste. Typically I like spicy food if it has a good	3.0	2.0	132.0	B003OB0IB8	luckybabe	2.0	Not for me	1.3238208E9	A2DLRG3Y...
I really loved the spicy flavor these had. I found myself liking the broth more than the noodles which is usually the opposite. If	0.0	0.0	133.0	B003OB0IB8	J. Espinoz...	5.0	Great spic...	1.3436064E9	A56HS2VN...
Got these on sale for roughly 25 cents per cup, which is half the price of my local grocery stores, plus they rarely stock the spicy	0.0	0.0	134.0	B003OB0IB8	Seth S Mo...	5.0	Great valu...	1.33488E9	AOTEC8KE...
i have bought allot of different flavors and this happens to be one of my favorites and will be getting more soon	0.0	0.0	135.0	B003OB0IB8	Mother of ...	5.0	great flavor	1.3346208E9	A1D0URD...
I love this flavor of ramen! I don't eat other flavors anymore because they taste too bland by comparison. For the convenience of	0.0	0.0	136.0	B003OB0IB8	Alanna	4.0	Tastes gre...	1.3337568E9	A2BZNXMV...
The BEST investment I've ever made for ginger. It's unbelievable! It's fibrous like the real ginger, has that spicy kick to it, but it's	0.0	0.0	137.0	B002SRYRE8	Sarah	5.0	Tastes aw...	1.3497408E9	A198FU6P...
Delicious. I cannot get Australian ginger where I live. This compares favorably to Australian Ginger I've purchased in other cities.	0.0	0.0	138.0	B002SRYRE8	Smileface	5.0	Happy Face	1.3420512E9	A21E0OJ4...
I keep trying other brands.... cheaper brands. Stupid me! This ginger IS soooo worth the money. Tender, moist and never a let	0.0	0.0	139.0	B002SRYRE8	Gettin Rea...	5.0	Simply the ...	1.3291776E9	A1FJJQA3...
My son has had acid reflux since birth and has been on every medication available with no improvement, Until an older gentleman	0.0	0.0	140.0	B002SRYRE8	Mrs. J. Todd	5.0	Excellent P...	1.2682656E9	A20PK3OC...
I bought this for our office to give people something sweet to snack on. Because it's bite size it's easier for people to grab a couple	0.0	0.0	141.0	B001GVISJW	N. Pine "pc...	4.0	Nice snack	1.3144896E9	ARS5Q7LR...
Twizzlers brand licorice is much better than that other well known brand.<br />If you can get these for $2 to $2.50 a package with	0.0	0.0	142.0	B001GVISJW	Angel_Boy...	4.0	Good Licorice	1.3017888E9	A2YIO225...

	TERM	FREQ ▼	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
⊞	order	1226	1026	☑	4.285	Verb	Alpha
⊞	give	1216	1061	□	0.0	Verb	Alpha
⊞	time	1213	1051	☑	4.25	Noun	Alpha
⊞	add	1195	909	☑	4.46	Verb	Alpha
⊞	bag	1194	789	☑	4.664	Noun	Alpha
	more	1103	961	□	0.0	Adj	Alpha
	more	1078	993	□	0.0	Adv	Alpha
⊞	dog	1058	502	☑	5.316	Noun	Alpha
	now	1044	938	□	0.0	Adv	Alpha
⊞	chip	1040	551	☑	5.182	Noun	Alpha
⊞	good	1034	931	☑	4.425	Noun	Alpha
	amazon	1031	868	□	0.0	Prop	Alpha
	again	1017	948	□	0.0	Adv	Alpha

Figure 2 Text Filter Interactive View

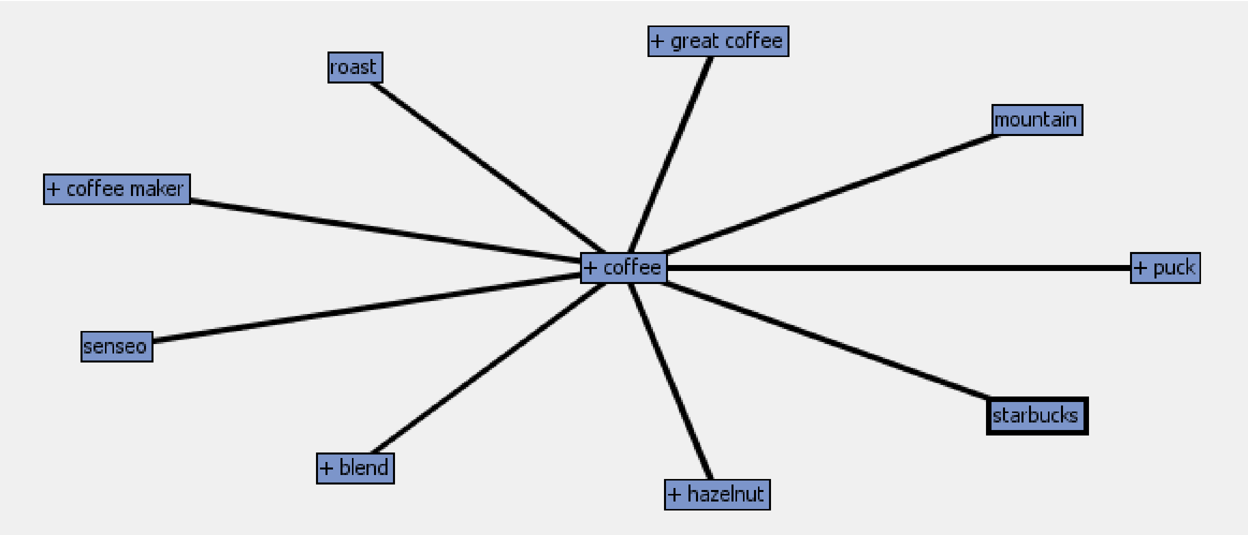


Figure 3 Identifying the word Coffee and words associated with it

5- Identify synonyms in your data and combine them together using interactive text explorer

6- What words/word-combinations have you combined?

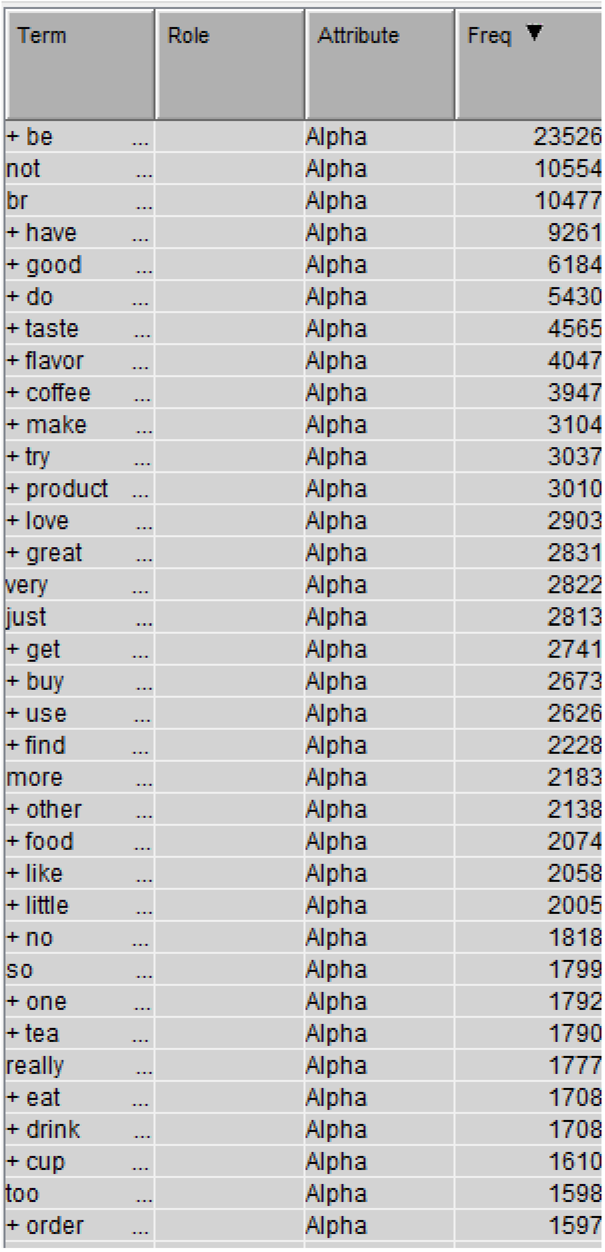
- br occurred 10477 and it is code for line break, Amazon (Prop & noun) Occurred 497 times and it adds no meaning therefore they were dropped.
- Purchase & buy gives the same meaning in the data set.
- Great & Excellent has the same level in sentiment and depending on the analysis other words can be added as synonyms like good.
- Cat & Dog can be added if we are referring to pets.
- Coffee, Tea, Cocoa (Cocoa & Hot Cocoa are synonyms) can be combined as hot beverages.

7- Give some examples of the entities identified by the text parser

Starbuck, coffee, tea, bag, sugar, chocolate, juice, soda, cookie, potato chip

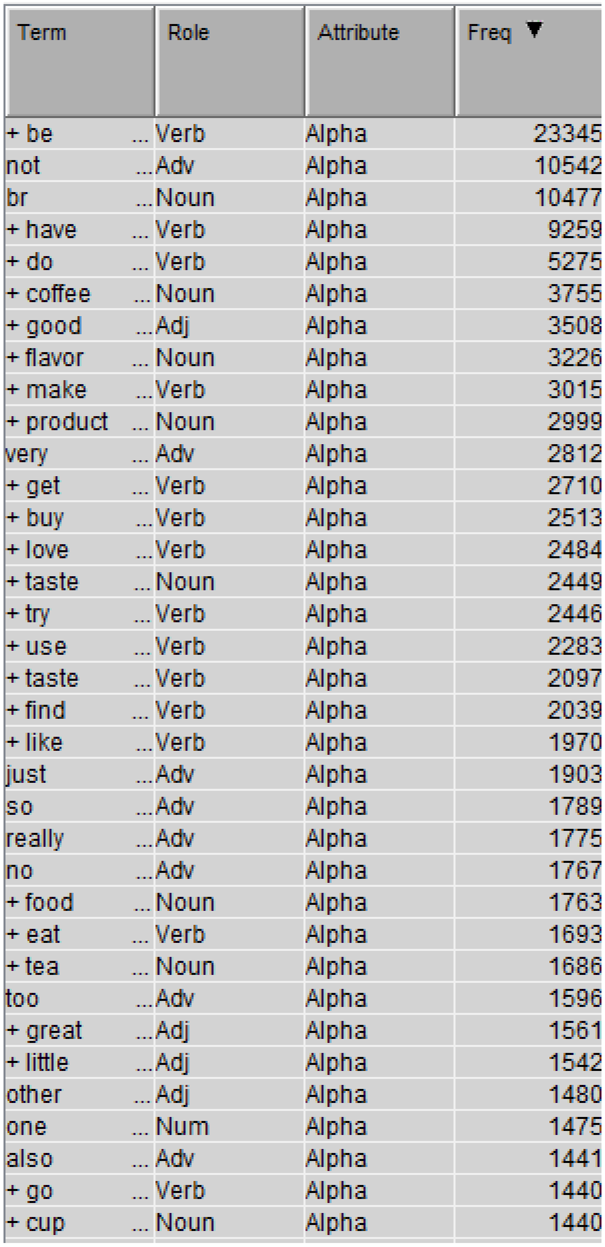
8- Repeat the previous steps while disabling POS. how are your results affected?

While using parts of speech words like great and excellent had to be matched with the right POS before treating them as synonyms, if POS was not used then it’s not a problem.
By comparing the two methods some words changed in position in terms of frequency, but I don’t think it make much of a difference while looking at the top 100 words for example.



Term	Role	Attribute	Freq ▼
+ be ...		Alpha	23526
not ...		Alpha	10554
br ...		Alpha	10477
+ have ...		Alpha	9261
+ good ...		Alpha	6184
+ do ...		Alpha	5430
+ taste ...		Alpha	4565
+ flavor ...		Alpha	4047
+ coffee ...		Alpha	3947
+ make ...		Alpha	3104
+ try ...		Alpha	3037
+ product ...		Alpha	3010
+ love ...		Alpha	2903
+ great ...		Alpha	2831
very ...		Alpha	2822
just ...		Alpha	2813
+ get ...		Alpha	2741
+ buy ...		Alpha	2673
+ use ...		Alpha	2626
+ find ...		Alpha	2228
more ...		Alpha	2183
+ other ...		Alpha	2138
+ food ...		Alpha	2074
+ like ...		Alpha	2058
+ little ...		Alpha	2005
+ no ...		Alpha	1818
so ...		Alpha	1799
+ one ...		Alpha	1792
+ tea ...		Alpha	1790
really ...		Alpha	1777
+ eat ...		Alpha	1708
+ drink ...		Alpha	1708
+ cup ...		Alpha	1610
too ...		Alpha	1598
+ order ...		Alpha	1597

Figure 4 No POS



Term	Role	Attribute	Freq ▼
+ be ...	Verb	Alpha	23345
not ...	Adv	Alpha	10542
br ...	Noun	Alpha	10477
+ have ...	Verb	Alpha	9259
+ do ...	Verb	Alpha	5275
+ coffee ...	Noun	Alpha	3755
+ good ...	Adj	Alpha	3508
+ flavor ...	Noun	Alpha	3226
+ make ...	Verb	Alpha	3015
+ product ...	Noun	Alpha	2999
very ...	Adv	Alpha	2812
+ get ...	Verb	Alpha	2710
+ buy ...	Verb	Alpha	2513
+ love ...	Verb	Alpha	2484
+ taste ...	Noun	Alpha	2449
+ try ...	Verb	Alpha	2446
+ use ...	Verb	Alpha	2283
+ taste ...	Verb	Alpha	2097
+ find ...	Verb	Alpha	2039
+ like ...	Verb	Alpha	1970
just ...	Adv	Alpha	1903
so ...	Adv	Alpha	1789
really ...	Adv	Alpha	1775
no ...	Adv	Alpha	1767
+ food ...	Noun	Alpha	1763
+ eat ...	Verb	Alpha	1693
+ tea ...	Noun	Alpha	1686
too ...	Adv	Alpha	1596
+ great ...	Adj	Alpha	1561
+ little ...	Adj	Alpha	1542
other ...	Adj	Alpha	1480
one ...	Num	Alpha	1475
also ...	Adv	Alpha	1441
+ go ...	Verb	Alpha	1440
+ cup ...	Noun	Alpha	1440

Figure 5 POS