Mylan's EpiPen controversy: Leveraging Text Analytics during PR crisis

Nikhila Kambalapalli, Dr.Josh Wiener, Dr.Miriam McGaugh, Oklahoma State University, OK, USA

ABSTRACT:

The impact of social media on businesses is huge as the number of people sharing their unbiased opinions on platforms like Twitter has increased drastically over the past few years. Both good and bad publicity can spread to the masses in a matter of few seconds. Mylan N.V., an American global pharmaceutical company that acquired the right to market the EpiPen is one such company which has been appearing on the Social Media from the past one year. People tweeted their concerns when the price of EpiPens has been increased drastically in 2016, when more than 80,000 EpiPens have been recalled in March 2017 across multiple countries and in April 2017 when Sanofi filed antitrust lawsuit against Mylan over EpiPen[5].

The primary objective of this research is to know about the attitude of people towards the company and the industry and if prior reputation of the company (post price hike) influenced how people reacted to EpiPen product recall and lawsuit against EpiPen. To assess the opinion of people towards the company over time, I have collected the data from Twitter using Google Twitter API.I have used SAS Enterprise Miner and SAS Sentiment Analysis Studio to know the overall sentiment and the topics mentioned in the tweets. This analysis helps the company to stay ahead of the competition and retain its customer base especially when generic versions of EpiPen competitor are going to be available at cheaper prices.

INTRODUCTION:

EpiPen is an injection containing epinephrine[4] and is used to treat severe allergic reactions to insect bites, food, drugs and other allergens. The price of EpiPen had drastically increased from about $103.50 in 2009 to more than $608.61 in 2016.The price hike of EpiPen had created outrage among allergy sufferers and their families. People expressed their views and concerns about EpiPen pricing on Twitter and other Social Media platforms. Also more than 80,000 EpiPens have been recalled in March 2017 across multiple countries because they might fail to work in an emergency and this again led to people tweeting their opinions about the company. People used hashtags and tweeted about EpiPen and Mylan when Sanofi (a French multinational pharmaceutical company) filed antitrust lawsuit against Mylan over EpiPen in April 2017.Analysing the tweets about EpiPen after each incident might give insights about the opinion of people towards the company and helps to understand if people are reacting with prejudice.

DATA COLLECTION AND PREPARATION:

To collect the data from Twitter, I have used Twitter Archiving Add on in Google sheets. Twitter archiver is available in free and premium versions. I have used the free version to create a search rule which allows you to specify the combination of hashtags so that matching tweets would get saved in a spreadsheet. After you give the search rule and establish a connection to Twitter, the Google spreadsheet talks to Twitter through a Google Script and imports all the results obtained by the search rule into a Google spreadsheet. The connection is established every few minutes and the tweets are fetched. The data was collected for tweets related to EpiPen price hike, EpiPen product recall and lawsuit against Mylan over EpiPen and divided into three groups-one for each incident. EpiPen price hike data (Dataset1) was collected between August 2016 and November 2016.

The EpiPen product recall data was collected during March 2017(Dataset2) and tweets about EpiPen related to Sanofi lawsuit (Dataset3) were collected in the last week of April 2017.The challenge in data collection was to decide upon the search rule to be given to Twitter Archiver. After thorough research on twitter and knowing what the hashtags used are for the particular incident I have decided upon the search rules. The spreadsheet files were converted into three different SAS datasets one for each group. Data cleaning in the initial stage was done manually using Excel. Text mining process flow was performed for each dataset separately.

METHODOLOGY:

In Enterprise Miner, the first step was to create a library and add the SAS datasets as data sources. Text parsing was used to convert unstructured text to structured format which makes it easy for analysis. There are also Parts of Speech tags which help to extract more information from text by assigning each token the relevant part of speech. There is also a customized stop list which identifies the terms which occur rarely. Then the text filter node removes all the unnecessary terms by filtering the text data. This is the most important step and it allows the user to handle the filtering using the interactive filter viewer. The user can select which terms to keep and which terms to drop and can group similar terms together. We can see the concept links in the interactive filter, which help in understanding the association between words based on the co-occurrence of words in the documents. The thickness of the line joining the terms indicates the strength of association. The next step in the process flow is Text Cluster Node. Generally, term by document matrix is large and so it is rarely used directly for clustering. Singular Value Decomposition (SVD) technique is used to break down the data into clusters such that documents in a cluster are more similar to each other and documents that do not belong to the same cluster are not similar[1]. We can use either Hierarchical or Expectation Maximization (EM) algorithms for clustering in Enterprise Miner. The Text Topic Node is connected to the Text Cluster Node and it gives the topics or terms identified in each cluster.

RESULTS:

Analysis of EpiPen Price Hike Tweets:

The data was analyzed over the time to see the topic trend and later the tweets collected every few weeks are all combined into EpiPen_Price_Hike dataset to form clusters.
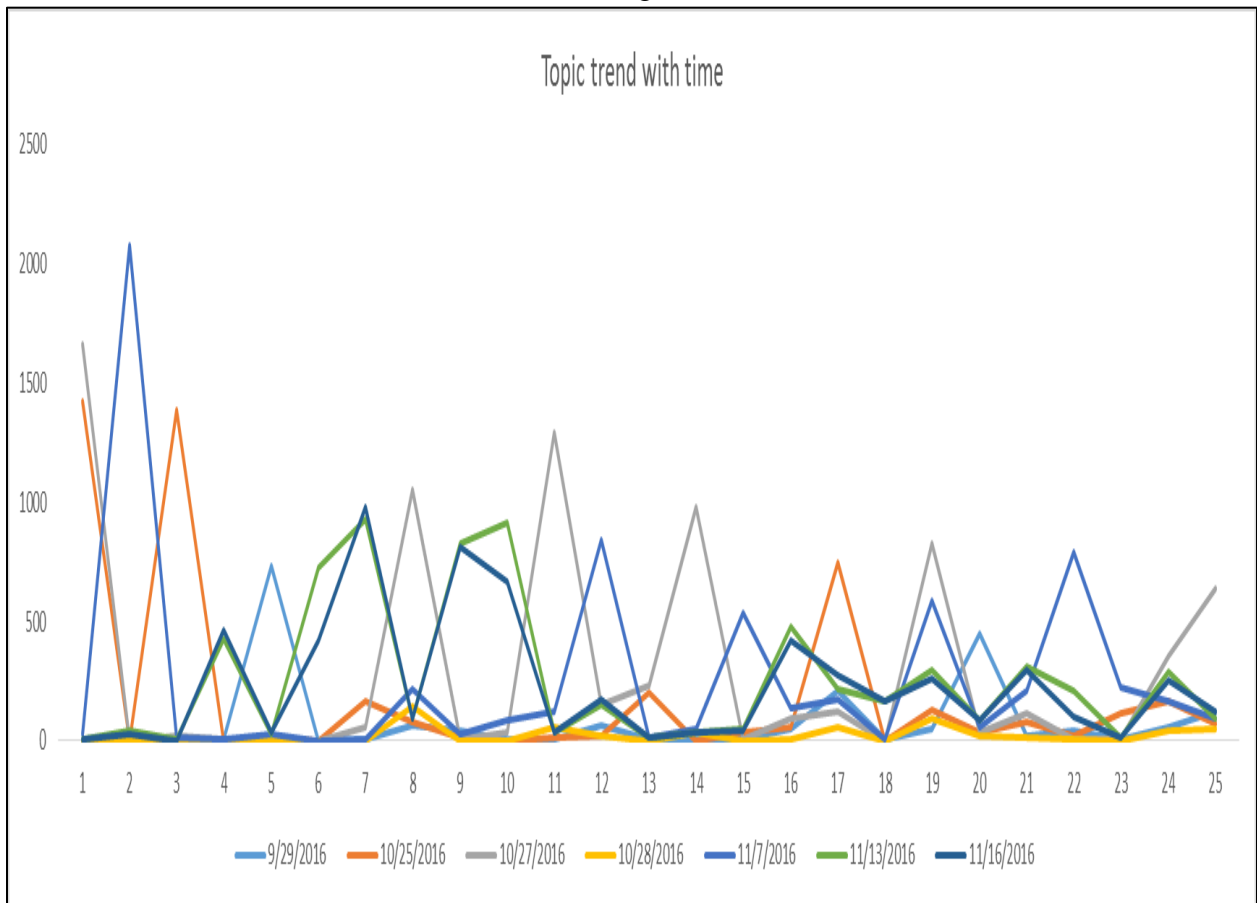
Figure 1
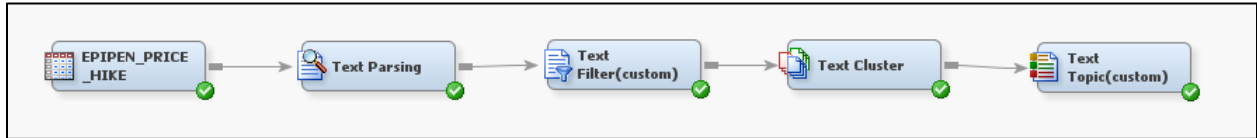
The process flow for the Epipen Price Hike data analysis is shown in figure 1.

Topic Trend with Time:

Figure 2

Topic trend analysis was done on the EpiPen price hike data by breaking down the data into seven different sets. The 25 topics shown in figure 2 are the same topics mentioned in the table below.

| Topic	Topic ID
nationwide,major health crisis,major,+crisis,medica	1
entire pharmaceutical industry,entire,out of control,+industry,+pharmaceutical	2
doubling,+triple,+drug company,+prevent,+law	3
el,de,jugar en el,xadamos,dos	4
dollar,+cost,+few,+family hundred,reason	5
reimbursement,military,+epipen military reimbursement,three,+senator	6
manufacturer,+price-gouge,mn,+mn loss,+post	7
millions,+add,+epipen price hike,+hike,exclusive	8
watchdog,+misclassify,medicare,medicaid,+senator	9
antitrust,senate,antitrust probe,ftc,+probe	10
return,kaleo,+pharmaceutical company,pharmaceutical,wednesday	11
grassley,defense,+explain,spending,department	12
failure,success,+view,accessible,sarahksilverman	13
return,rival,+shelf,ardent,+draw	14
virginia,justice,reject,west,+urge	15
pentagon,million,repay,+demand,+senator	16
doj,+maker,yes,+answer,+question	17
third-quarter,+third-quarter loss,+propose,+cost,+post	18
alternative,+competitor,back,coming,drugstores	19
tamaraleighllc,crookedhillary,hackers,+crony,+crookedhillary crony	20
lower,+charge,+hurt,loss,settlement	21
insulin,sticker,sticker shock,shock,+insulin price	22
close,reaches,window,also-ran,opportunity	23
overcharge,+pay,+lose,unlikely,wsj	24
buy,+market,vicelike,grip,allergy shot market	25

Concept link for EpiPen:

Figure 3
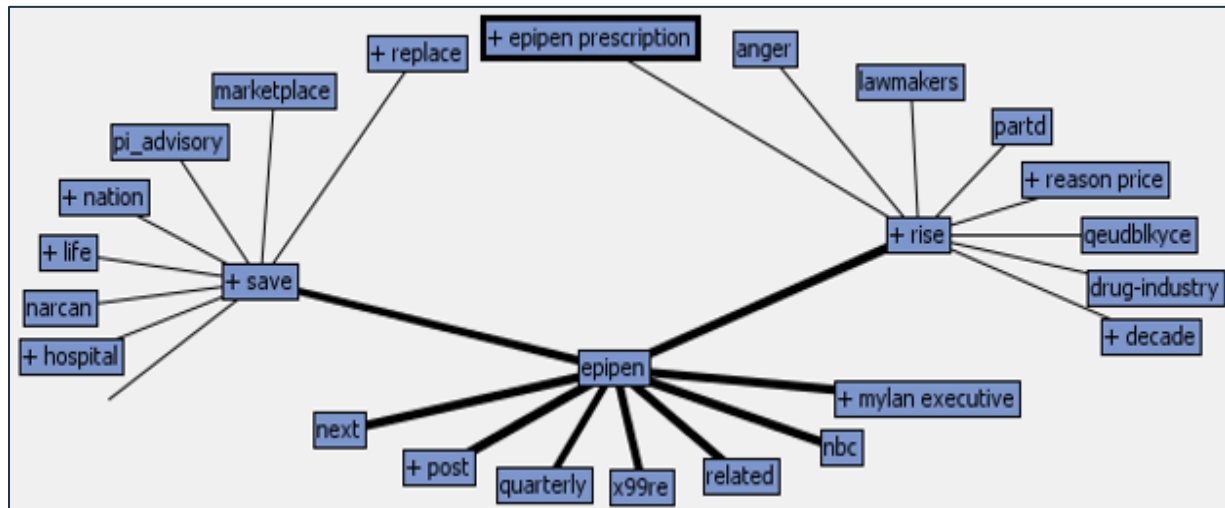
To know the answer for 'what are the people talking about EpiPen on Twitter?' the concept link shown in figure 3 would help us. The tweets show that people are angry and they want somebody to save the nation from the insane price of EpiPen.They also want the lawmakers to address this issue. They are also questioning the drug industry about this.
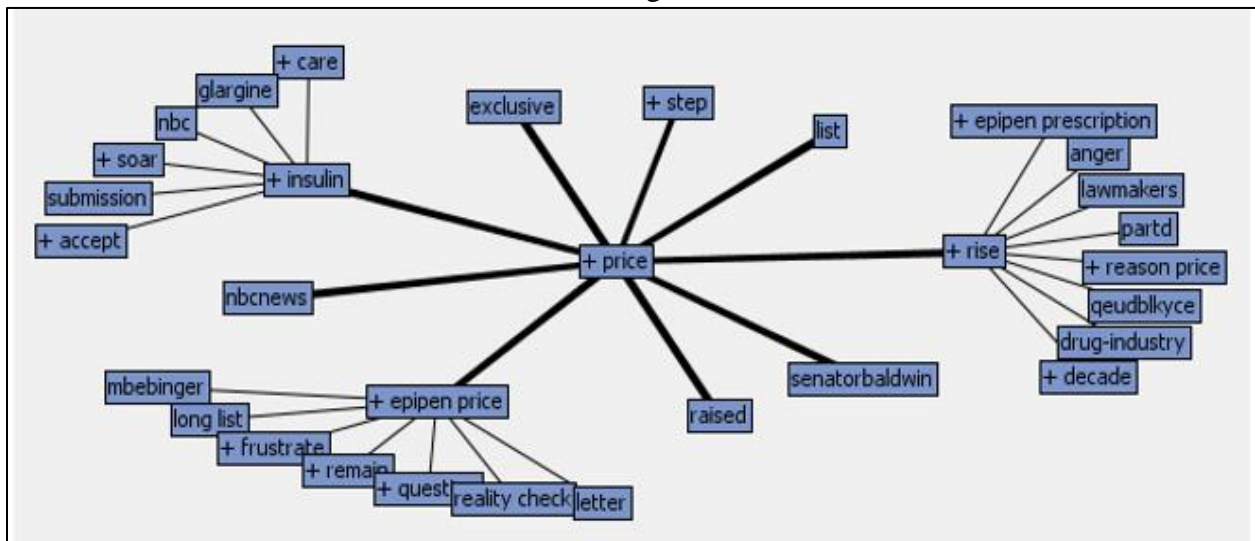
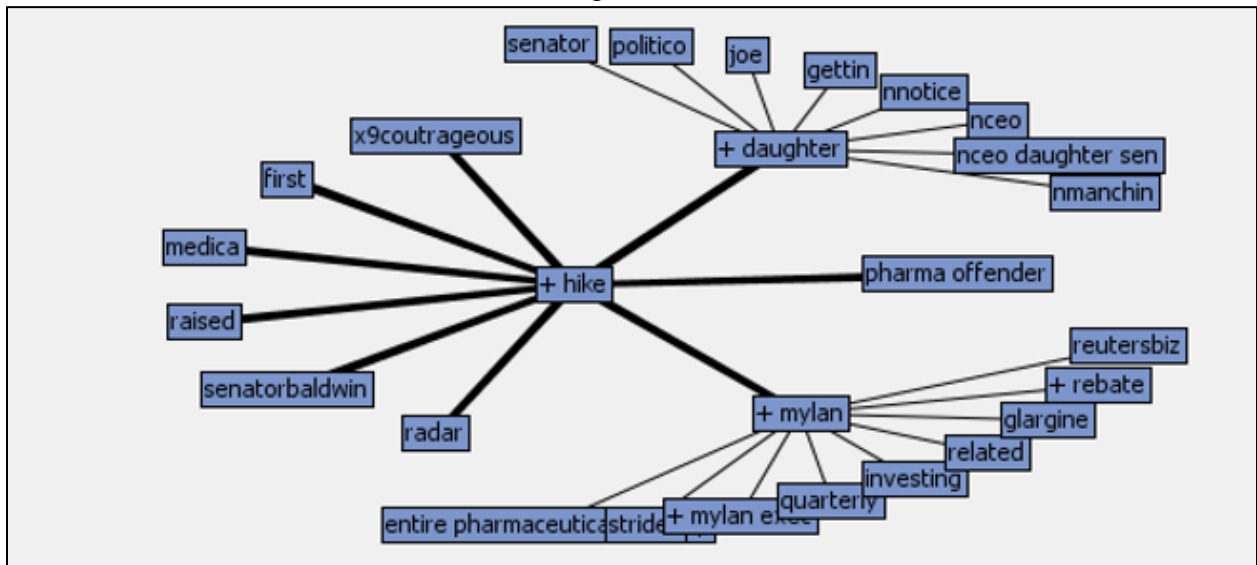**Concept link for Mylan:**

Figure 4
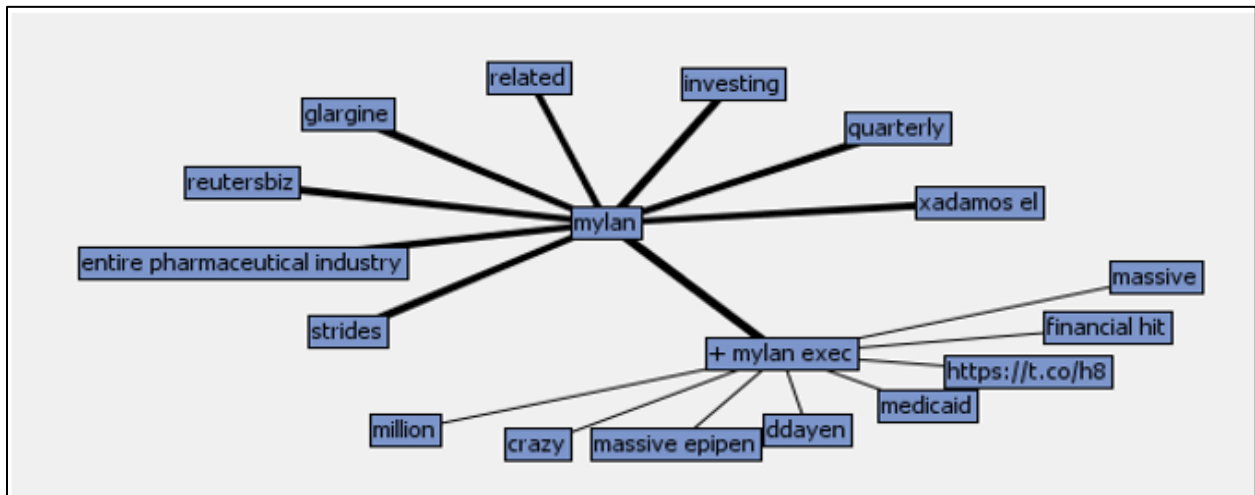
Concept link for price:

The concept link for price is shown in figure 5.When people are tweeting about EpiPen's price, price has strong association with rise, senator and EpiPen price has strong association with frustrated, reality check. In addition, when insulin prices went up people commented that 'Is Insulin the new EpiPen?' and some people related the Insulin price hike with that of EpiPen price hike and they questioned if anybody is even regulating the price hikes at all.

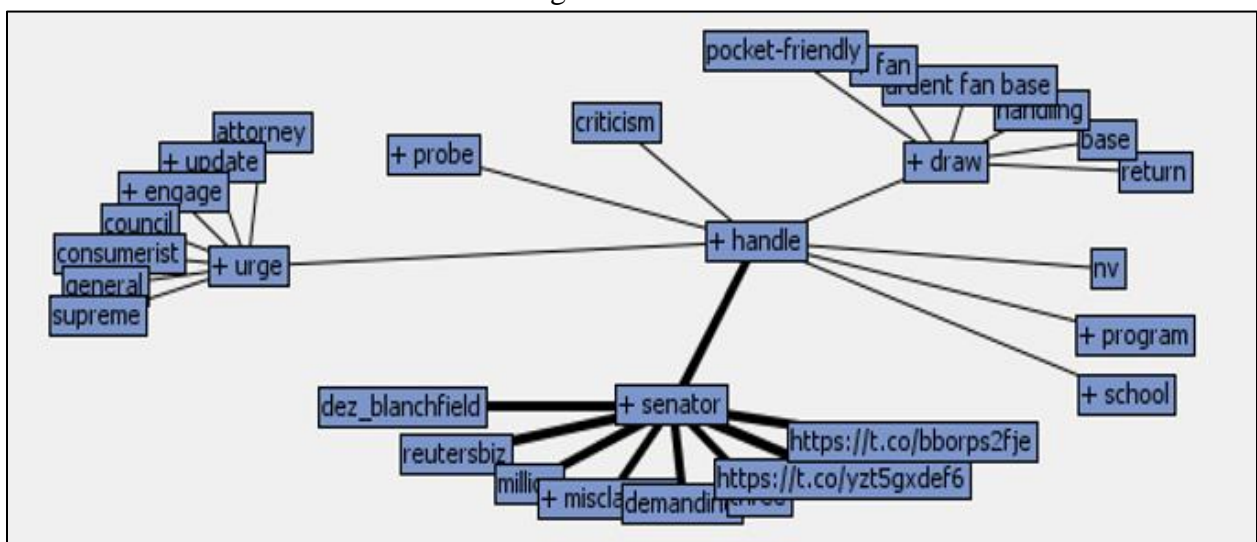Figure 5

Concept link for hike:

Figure 6

The concept link for hike shows the different terms that are strongly related to hike. The interesting thing is the term 'daughter' and they are referring to the CEO of the company who is the daughter of a politician. That is the reason the term 'daughter' is associated with 'senator' when you expand the term daughter. This can be understood from a tweet like 'Bresch CEO of #Mylan #EpiPen is the daughter of a Sentor as well, what was mom's salary?\n\nhttps://t.co/JvPK5PJHs7 https://t.co/AWjpYXUrTv"'

Concept link for Senator:

Figure 7

The concept link for senator is shown in figure 7. Some senators criticized the price hike and they urge Mylan probe over EpiPen school programs.

Text Clustering:

Figure 8

Cluster Id	Descriptive terms	Percentage	Explanation
1	+shelf +'epipen rival' +cheap return +alternative +pharma +window opportunity +stock +rival	7%	Cluster with terms showing optimism about EpiPen extending the shelf life and looking for a cheaper alternative.
2	+settlement +loss +post +urge +probe 'epipen settlement' epipen +senate +'west virginia' +senator +launch +answer +big	22%	Cluster with terms related to D9senators Mylan Probe and West Virginia urging Justice Department to reject settlement with Mylan
3	+epipen +price ceo epipen +insulin +family +competition mylan +read +generic +shock +gouge +sticker +'sticker shock' +'epipen price'	32%	Cluster with tweets about price hike of EpiPen and Insulin
4	+mylan +cost +entire +pharmaceutical +greed +industry 'out of control' +hike +pentagon +price +millions +add +'epipen price hike'	9%	Cluster with tweets about the greed of entire pharmaceutical industry
5	+mylan +price +epipens +berniesanders +crisis +soar +health +hike +major nationwide 'major health crisis' +cost +medicine 'year-old medica'	11%	Cluster with terms that express concern of the soaring cost of medicine
6	+epipen +alternative +competitor +market +return +senator medicaid +buy $30 +'epipen alternative' +medicare +allergy watchdog +misclassify +diy	18%	Cluster with terms about DIY alternative to EpiPen which costs $30 and EpiPen misclassification in 2009

Based on the clusters formed, conclusions made by looking at the descriptive terms the explanation is provided in figure 8.

The important topics are identified after the clustering was performed and shown in figure 9.

Figure 9

Topic ID	Topic
1	major health crisis,nationwide,+crisis,+major,year-old medica
2	doubling,+triple,+prevent,+law,+drug company
3	out of control,+entire,+industry,+pharmaceutical
4	+dollar,hundreds,+family hundred,+few,+reason
5	+millions,+add,+epipen price hike,+pentagon,+exclusive
6	alternative,+return,+pharmaceutical company

Analysis of EpiPen product recall tweets :
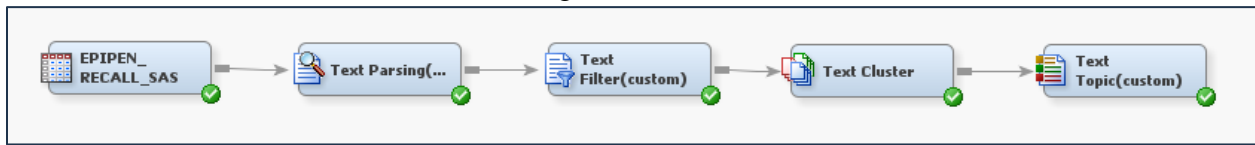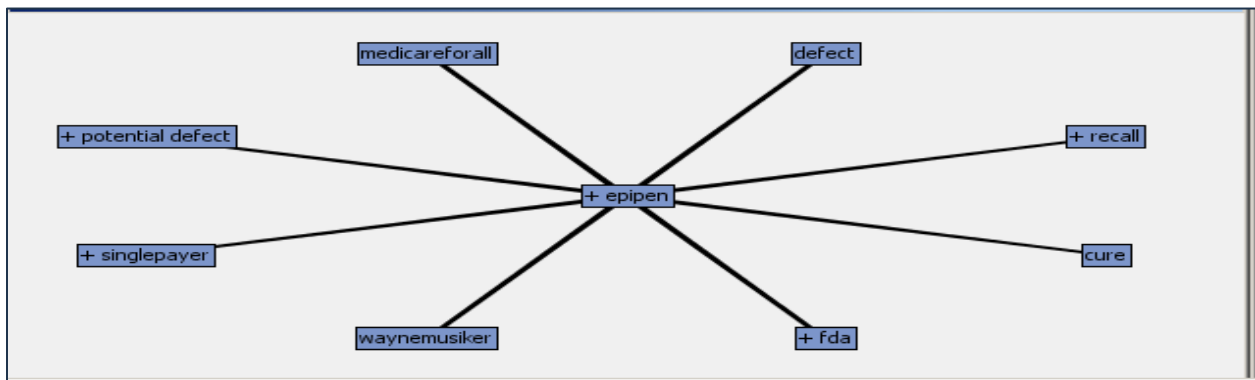
Figure 10

Figure 10 shows the process flow diagram for EpiPen product recall tweets.

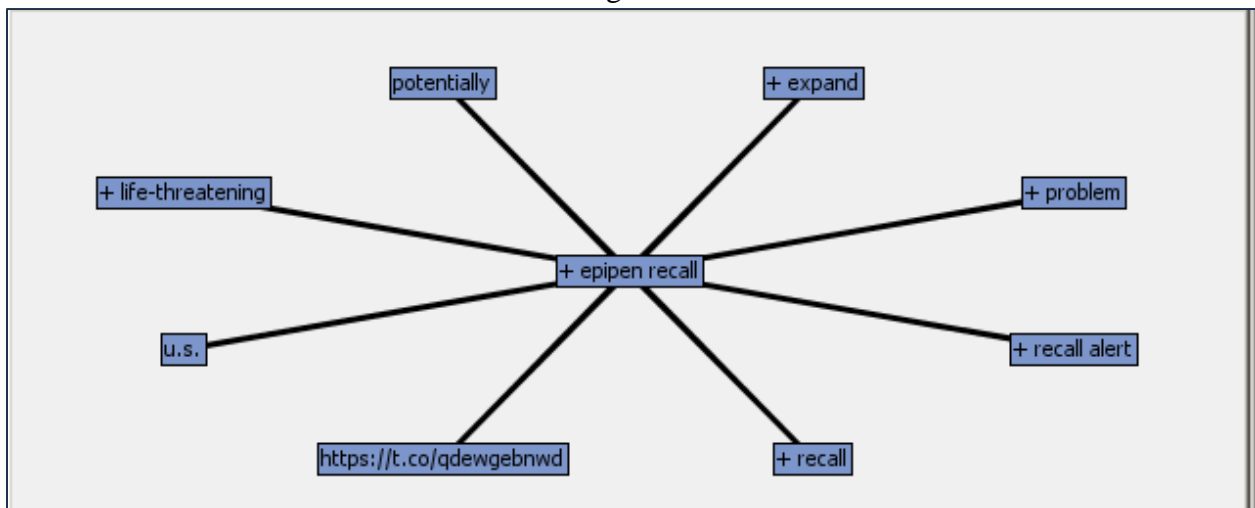**Concept link for EpiPen:**

The concept link for EpiPen is shown in figure 11.Epipen is strongly associated with the word defect, recall, single player, FDA.

Figure 11

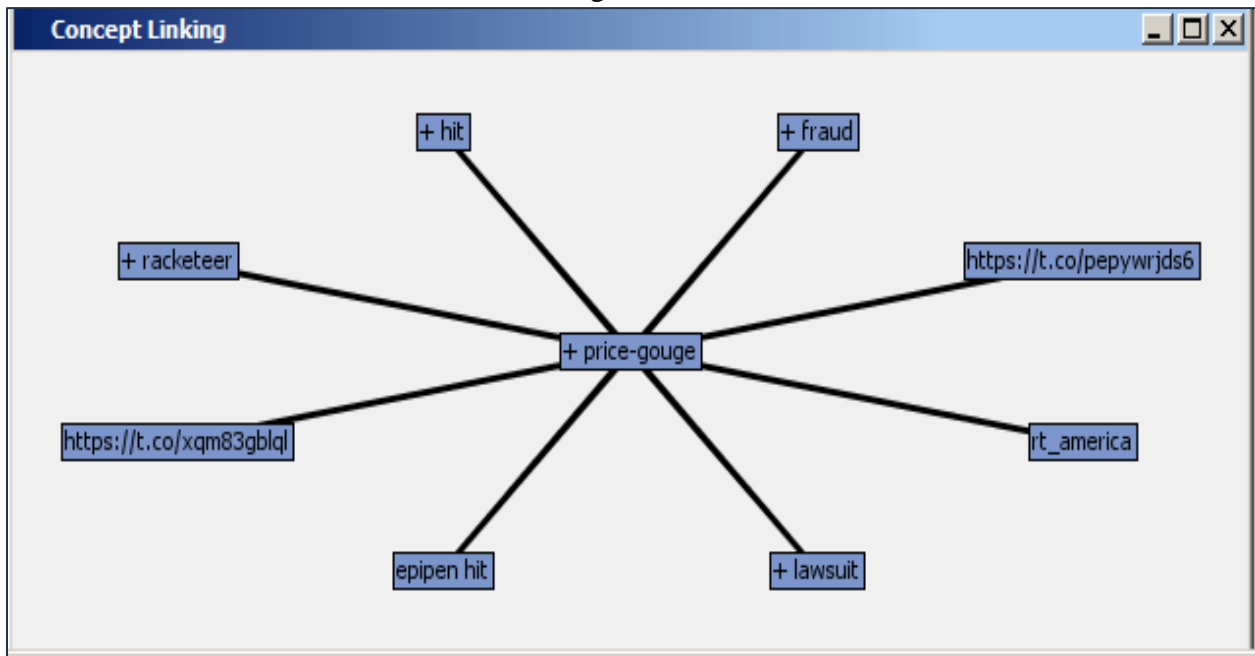**Concept link for EpiPen recall:**

Figure 12

Epipen recall is strongly associated with the terms recall alert,life threatening which shows that people are worried about the situation.

Concept link for price-gouge:

The concept link for price gouge is shown below in figure 13.It is strongly associated with fraud, lawsuit and some internet links

Figure 13

Text Clustering:

Figure 14 shows the different clusters formed from the EpiPen product recall data and shows the descriptive terms for each cluster.

The interpretation of the clusters along with the description is shown in figure 14.

Figure 14

Cluster ID	Descriptive Terms	Percentage	Explanation
1	epipen +foodallergy +lawsuit racketeering +maker +fraud price-gouging +hit defective 'epipen hit' +'defective part' 'mylan hit'	35%	Cluster with tweets showing that people are upset with expensive product like EpiPen being recalled.
2	distributed recalled +stock +return waynemusiker myles +'potential defect' +potential defect +mylan nationwide	26%	Cluster with tweets sharing information that the EpiPens have been recalled due to potential failure and
3	u.s. global +healthcare +bigpharma +mondaymotiviation +singlepayer cure disease medicareforall robinhoodtax extends 'news alert' +expand +medscape manufacturing	13%	Cluster with retweets of tweet '#BigPharma isn't healthcare, it's a disease. #MedicareForAll is a cure!
4	alert +fda nationwide +voluntary +'recall alert' 'nationwide voluntary recall' consumers alerts problems +life-threatening potentially +pfizer subject +info +canada	21%	Cluster with tweets about Pfizer in Canada recalling EpiPens and alerting other people about it and warning them to be careful
5	work +'need to' +life 'life saving' +report Cfocustohelp shots 'allergy shots' recalls +emergency lifesaving +want lives foodallergy	5%	Cluster with positive tweets

The results of text clustering are shown in figure 14.
The important topics identified by the Text Topic Node after Text clustering are shown in figure 15.

Figure 15

Topic ID	Topic
1	myles,waynemusiker,+distribute,+return,+stock
2	+fda,epipen jr,+voluntary,+alert,nationwide voluntary recall
3	+singlepayer,cure,medicareforall,disease,+mondaymotiviation
4	fraud,+lawsuit,+hit,epipen hit,+maker,+price hike
5	u.s.,epipen,+include,foodallergy,+mylannews
6	defect,+potential defect,+announce,+mylan
7	expand,+epipen recall,+alert,+problem

Analysis of EpiPen and Sanofi Lawsuit tweets:
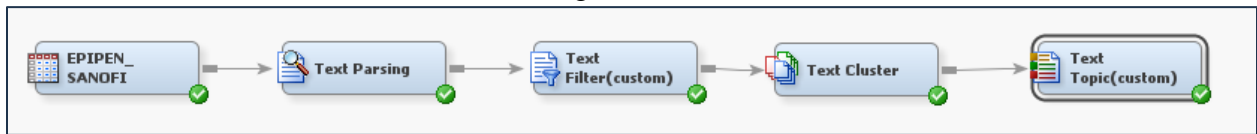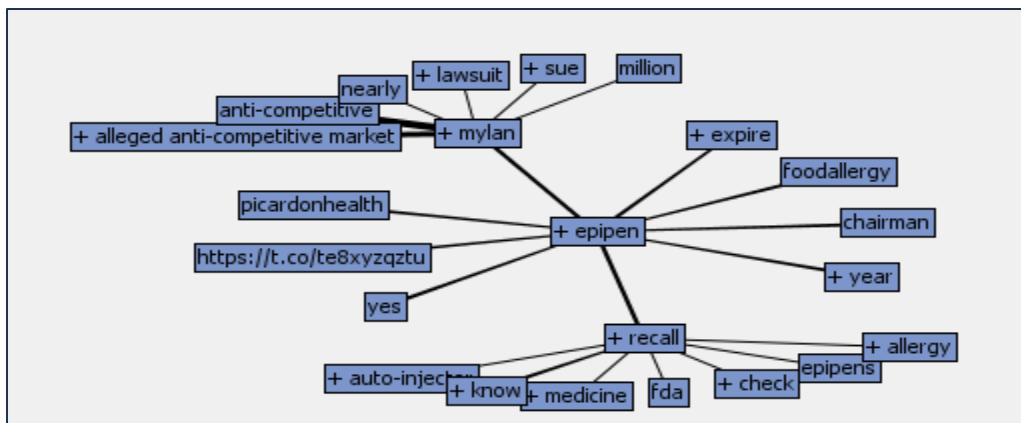
Figure 16

Figure 16 shows the process flow diagram for tweets when Sanofi filed a lawsuit against Mylan over EpiPen.
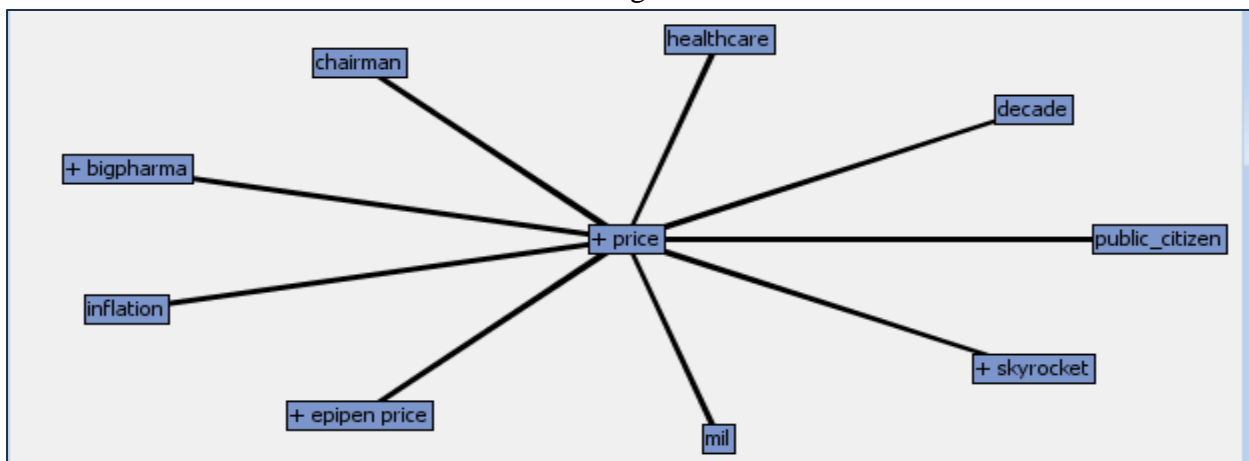
Concept link for EpiPen:

Figure 17

From figure 17, we can observe that EpiPen is strongly associated with the terms recall, mylan, and food allergy. Therefore, when Sanofi filed the lawsuit against Mylan over EpiPen people are talking about the past events too.
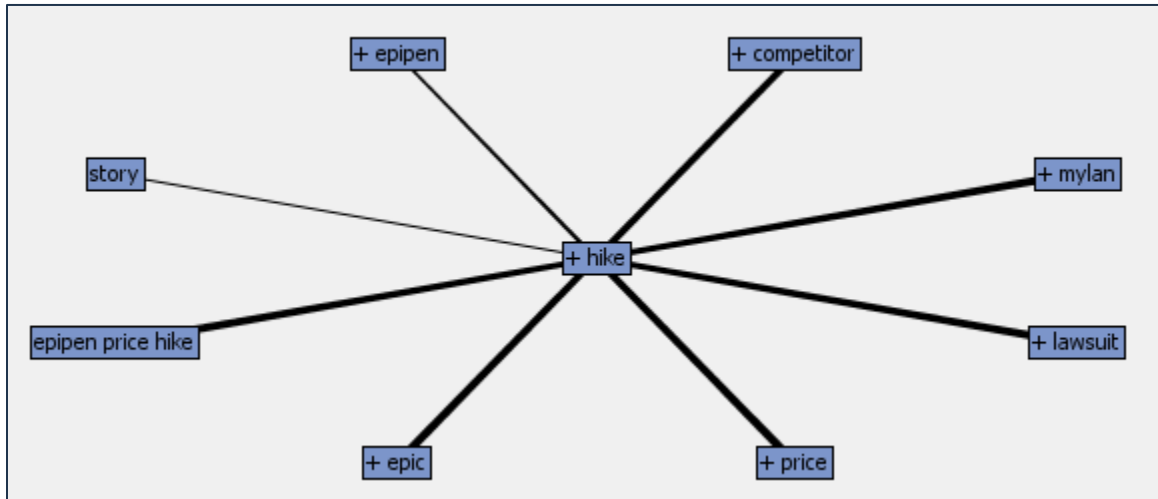
Concept link for price:

Figure 18

Price is strongly associated with the terms bigpharma, skyrocket, healthcare etc. In general the concluding statement from this could be EpiPen prices are skyrocketing.

Concept link for hike:

Figure 19

From the concept link for hike,we can say that when there is hike in the tweets it occurs either with epipen price hike or with the other terms lawsuit,mylan,epipen.

Text Clustering:

Figure 20

Cluster ID	Descriptive Terms	Percentage	Explanation
1	chairman +mylan million +received marketing 'alleged anti-competitive marketing' wsj +profit	22%	Cluster with tweets announcing the news that Mylan was sued for anti competitive marketing
2	ottawa_medic years picardonhealth +shot +year +overnight danidmedia hint	17%	Cluster with tweets sharing valuable information about not to use expired EpiPens
3	potent +epipen +mylan +recall expiration +auto-injector allergies +study +lawsuit recalls +'expiration date'	24%	Cluster with tweets about EpiPen's expiration and claims that the lawsuit was all about illegal business practices of Mylan.
4	sanofi results financial 'financial results' caianaphylaxis +learn +bigpharma decade +sanofi q1 results healthcare prices skyrocketed 'epipen price hike'	31%	Cluster with tweets about EpiPen's skyrocketed prices
5	foodallergy +help kids caianaphylaxis atericahealth +share +raise poster 4th anniversary food allergy awareness	6%	Cluster with tweets about 4th anniversary of EpiPen and also about how useful it is for kids and adults with allergies.

The results of text clustering are shown in figure 20 and the topics identified after clustering are shown in figure 21.

Figure 21

Topic ID	Topic
1	mylannews,+shoot,+bank,overnight
2	danidmedia,safe,picardonhealth,expired
3	sanofi,+result,today,+announce,infographic
4	foodallergy,+share,+kid,+epipen,+raise
5	healthcare,100%,bigpharma,decade,+skyrocket,+price,+recall
6	mylan,+million,nearly,$100 million,chairman

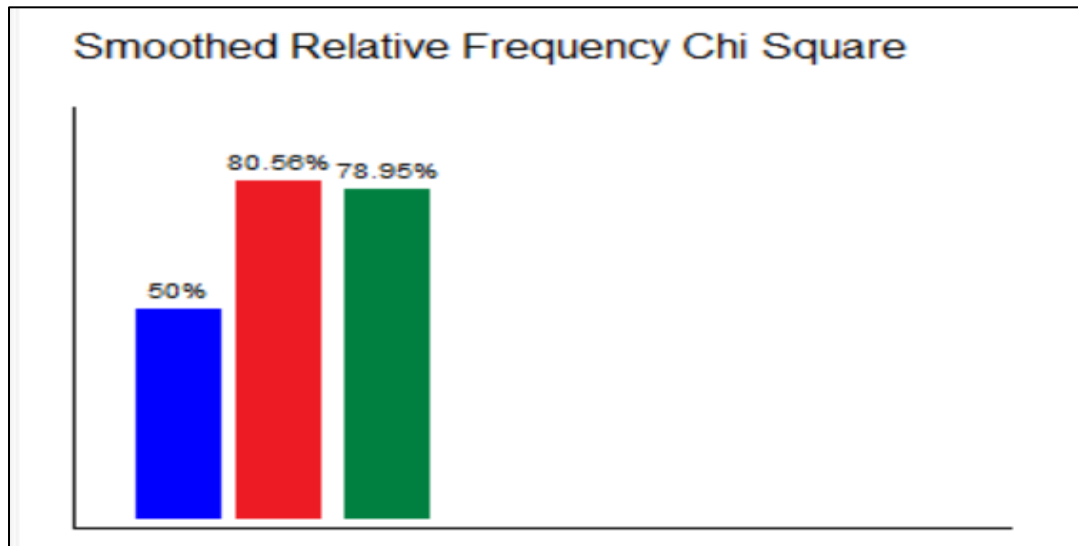## SENTIMENT ANALYSIS:

Figure 22

Figure 22 shows the results of Statistical Model from SAS Sentimental Analysis Studio.
- The overall sentiment is negative with very few positive tweets to work with.
- The other tweets are neutral through which people are sharing the information.
- The best model is smoothed relative frequency and feature ranking with overall precision of 78.95%,positive precision 50% and negative precision 80.56%

As a part of future scope, more detailed analysis can be carried out apart from statistical modelling.

CONCLUSION:

- From the results, it is clear that when a new incident like product recall happened some tweets referred to price hike and people reacted with some bias.
- The overall sentiment about the company is negative but not all the tweets were filled with anger or frustration, some of the tweets were neutral in context.
- The previous opinion about the company influenced how people have reacted about the current situation.
- Using this analysis, Mylan can gain more knowledge of its audience and measure how each incident affected the brand's overall reputation.

REFERENCES:

[1] *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS® by Goutam Chakraborty, Murali Pagolu,Satish Garla*

[2] http://www.scsug.org/wp-content/uploads/2016/11/Categorization-of-Fitbits-Customer-Complaints-on-Twitter_SapnaBhoir_JackyArora.pdf

[3] http://support.sas.com/resources/papers/proceedings12/306-2012.pdf

[4] https://en.wikipedia.org/wiki/Epinephrine_autoinjector

[5] http://www.dailymail.co.uk/wires/reuters/article-4440816/Sanofi-files-U-S-antitrust-lawsuit-against-Mylan-EpiPen.html

CONTACT:

Name: Nikhila Kambalapalli Enterprise: Oklahoma State University Address: Spears School of Business, Oklahoma State University City, State ZIP: Stillwater, OK - 74074 Email: nikhila.kambalapalli@okstate.edu

Name: Dr. Miriam McGaugh Enterprise: Oklahoma State University Address: Department of Marketing, Oklahoma State University City, State ZIP: Stillwater, OK - 74074 E-mail: Miriam.mcgaugh@okstate.edu