Visualizing Keywords Over Time

Akanksha Shaktia ass546@nyu.edu
Zhongheng Li zl1761@nyu.edu
Priyanshi Shah ps3296@nyu.edu
Riya Patni rsp378@nyu.edu

Github Page : Github . Working Demo : Working Demo

What is the problem you want to solve and who has this problem?

The problem at hand is to develop a generic technique to visualize the change in keywords extracted from any given dataset over a period of time. The given dataset is Yelp's (reviews) dataset and the aim is to find the frequently occurring keywords, their changing trends over time, along with their appearance and disappearance over a period of time; augmenting better business decision based on the frequency. Analysis of these trends would help the business owners understand what people talk about their businesses and how these topics change over time.
To gain better insights, our team decided to focus on a specific class of business, which is restaurants, in a particular state of the United States, which is Pennsylvania. The idea is to realize that if the trends can be seen for a particular business category, then it can be scaled to accommodate other categories too, since the system has been strategically kept generic.
By visualizing such keywords, restaurant owners can gain insights into the factors that were most talked about in a particular month about restaurants in general by the people of Pennsylvania.
The resulting visualization essentially is aimed at serving other business models that would require making use of these trends.

What are the driving analytical questions you want to be able to answer with your visualization?

Questions:

1. What are the top 15 keywords (scalable upto 35) for each month extracted from the dataset over the year?
2. Which of these monthly keywords made it to the top keywords of the entire year – thereby implying they were most frequent throughout the year?
3. How does the ranking of the keywords change over time?
4. Which keywords appear or disappear over a period of time?
5. What are the top 5 reviews/comments that contain a particular word?

Description:

1. **What are the top 15 keywords (scalable upto 35) for each month extracted from the dataset over the years?**

   Given a dataset, which are the top 15-35 keywords that constitute the trending words or most frequently used words for restaurants reviews in Pennsylvania. By listing such words and sorting them by a factor, say, frequency, we can understand the driving factors for a business and what are the most commonly associated words with it. For example, keywords like 'BBQ, 'service', 'pizza' and frequency of each of them.

2. **Which of these monthly keywords made it to the top keywords of the entire year – thereby implying they were most frequent throughout the year?**

   Now that we can see all the frequent words for each month, question is, are these significantly important from the entire year's point of view? If yes, they would appear in the year's column that shows the top words for the entire year. If a word has made it to that list, we know its trend has been more or less frequent over the entire year.

3. **How does the ranking of the keywords change over time?**

   After finding all the top keywords, it is important to understand the life cycle of individual keyword over a period of time. That is, we need to visualize the ranking of this word at different points in time. So if a word appears as important in the month of January 2014, we look at its ranking over all the months of 2014 to see if this keyword has recently gained importance or has been important for many months now.

4. **Which keywords appear or disappear over a period of time?**

   This is a crucial question for the problem. Every month, a set of top 15-25 keywords are found. For example "thanksgiving" appears in the month of November as a frequent word, while "Valentines" appears as a frequent word in February. Since these words don't appear in any other months, we identify them as outliers. Certain words keep appearing and disappearing every month. Visualizing the set of newly appeared/disappeared keywords will thus give us an idea about which keywords started/stopped gaining importance over a period of time.

5. **What are the top 5 reviews that contain a particular word?**

   For a particular frequently occurring keyword, what are the top 5 reviews that contain the word most number of times? This will find interesting and important reviews and the user might want to dig in the details of that review.

What does your data look like? Where does it come from? What real-world phenomena does it capture?

The main dataset, Yelp Dataset Challenge, contains details and reviews of businesses over a period of 20 years across different locations in the United States. For the sake of simplicity, we have processed this data so that it shows the reviews and details for all the restaurants in Pennsylvania for the year 2014 and 2015. This processing was achieved using the unique business IDs, categories and addresses mentioned for each business. Following are the attributes and their descriptions used for the visualization.

Attribute Name	Attribute Type	Meaning	Values	Derived? (if yes explain how)
Business Id	Categorical	Unique ID assigned to each business unit.	Alphanumeric characters	No
Business Category	Categorical	Identify which category this business ID belongs to, whether it is a restaurant or a garage and so on.	Text	No
Text	Unstructured Text	The user comments posted on Yelp for each business.	Text	No
Keywords	Text	The frequent keywords derived from the text (comments) using Natural Language Processing	string	Yes. Keywords are generated from the existing dataset and stored in a new JSON file.
Frequency	Quantitative	For each derived keyword, this shows the number of occurrences of it in the entire dataset.	Number	Yes. By counting the number of occurrences of the derived keyword, we store the frequency of keywords in each month, which can be sorted and aggregated for better visualizations.
Mentioned	Unstructured Text	For each keyword, list the top reviews that contain the most occurrences of this keyword	Text	No. Just filtered out based on the top keywords.

What have others done to solve this or related problems?

For the visualization to impart correct and meaningful information, extracting the right keyword from the whole content is crucial. As a part of our initial research, we looked for appropriate algorithms to find such important words from the text. For this purpose, we referred to the following page published on ResearchGate which is a hub of such independent researches. This paper describes the efficiency of a keyword extraction technique known as RAKE (Rapid Automatic Keyword Extraction.)
Link to the paper: https://www.researchgate.net/publication/227988510_Automatic_Keyword_Extraction_from_Individual_Documents

Another beautiful visualization project that we came across is the Google Books N-gram Viewer. This is a simple and elegant visualization technique to show the occurrence of keywords over a time span. The ticks on x-axis depict a span of twenty years however, as you hover over the line, it shows information of the intermediate years. The user can also search for a keyword of his own choice to see its frequency. We plan to include a similar chart in our project to show the use of keywords over time.
Link to the project:
https://books.google.com/ngrams

Real-time project:
As the importance of data analytics is growing, companies are finding more and more applications for Data Mining. Often while browsing through web pages, we see dynamic advertisements on the margins or bottom of the page. However annoying they may seem, but if we notice them carefully, they are related to our previous web searches. This is a real- time application of frequent keyword extraction from the user's browsing history to extract what the user might be interested in and providing exactly those advertisements to him.
The project that we have shared here is a similar project created for a startup based in San Francisco to match advertisers to content providers in a context-specific way. In the post, the developer has shared an algorithm that helped in picking the right keywords from the entire dataset comprising of user's browsing history.
Link to the paper:
https://people.csail.mit.edu/lavanya/keywordfinder

How it is related: The project describes a step by step approach on how to extract keywords and rank them based on different attributes or features such as how frequent they are, or how recently they were used. This is similar to what we are planning to incorporate in our project.

Design Iterations

Initial Mockup:

How to read the visualization:

The main chart shows the the top 50 keywords for the year 2010, which automatically shows the words for two of its preceding and succeeding years, i.e., 2008, 2009, 2011, 2012. This is shown in the form of a list view. The blue bars in the background of each word indicates its frequency. Thus, we can compare that, for the year 2010, “wxv” has the least frequency and “kev” has the highest frequency.

This is the default visualization. User can choose a year of his choice by using the slider at the bottom of the list view. For the chosen year, we see the line chart and a word-matrix getting updated. The line chart shows the life-cycle of the words extracted for that year, over a period of 5 years. Note that, all these lines are initially greyed out. When the user chooses a word from the list or types a word in the search box, the line graph for that keyword is highlighted, and at the same time, its Calendar View is generated, as shown below. This calendar view shows the frequency ranking of this word throughout the year 14 2010, for each month of the year and each day in the month. This helps us compare the trends for a particular season or special trends on the weekends etc.

The word – matrix focuses on co-occurrence of words for the year 2010 and shows us which two words occurred together most frequently. When we click on a grid for say keyword Ki and Kj, we see that the line chart highlights the graph for those keywords and we can compare their trends over a period of 5 years, as shown below. These words will also be highlighted in the list view.

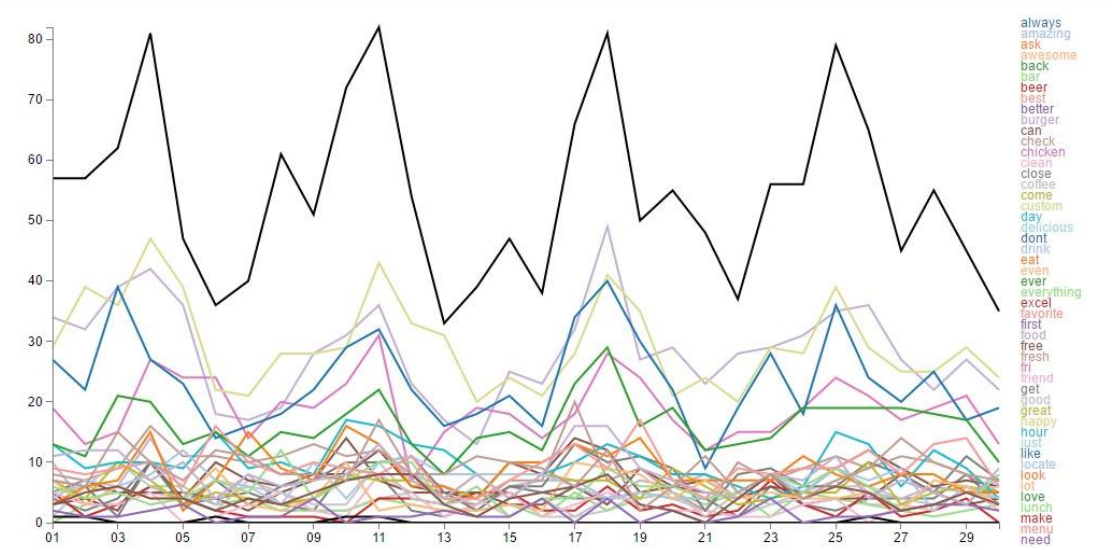
Project Intermediate Update

At this stage, we decided to show the visualization month-wise, instead of year-wise, since yearly generated keywords did not give an interesting trend. We processed the dataset such that the keywords and their frequency are stored month-wise. To add more granularity, we implemented a daily-data based line chart that shows how a word has performed on each day of the month. We planned to show this daily data based line chart such that it helps to compare two or more words, by selecting the words from the column on the right.

Main chart: List View chart

The main chart shows the top keywords of all the months of the selected year in a list-view where each month is represented as a column. The list chart is sorted by frequency in descending order. User can use the slider and change the year to view top keywords of all the months of the corresponding selected year.
The blue bars in the background of each word is proportionate to the frequency of each word.
Upon hovering on each word, that particular word is highlighted in all the columns allowing user to easily understand the change in ranking of that word over a period of 12 months.

The line chart shown below allows user to choose words from the side panel, and visualize their daily performance. It can either be used to analyze a particular word's performance, or comparison of multiple words. When a user selects a word, its line is highlighted, when a user selects another word, we see one more line with a new color getting highlighted. De-selecting these words will gray out existing lines.

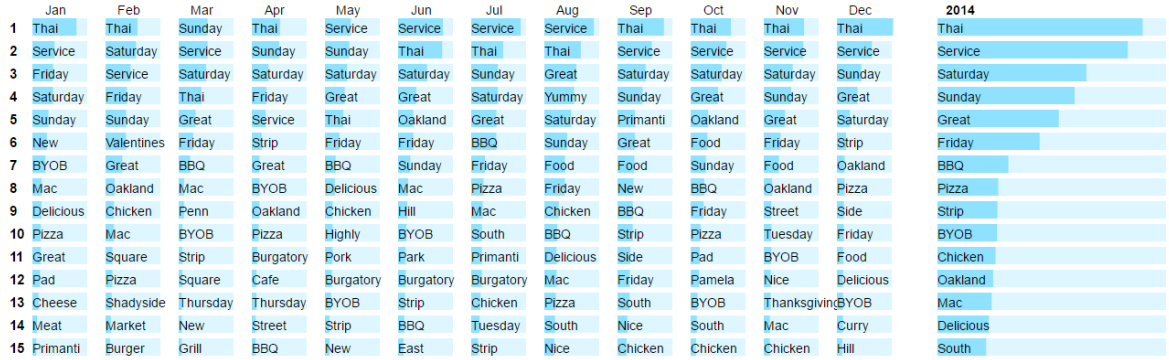


After discussion with the mentor we realized that the information that was being conveyed by the line chart could also be realized by joining word occurrences in adjacent columns of the List View chart using lines thus creating a word flow. Therefore, we dropped the idea of using the line chart.

Final Visualization

The final visualization was aimed at answering all the questions proposed in the proposal.

To answer the question about the top 15-35 words over a period of time, the list chart is designed as below that shows the words rank wise for each month, with their frequencies being visualized as horizontal bar charts in the background. By looking at each month's list, we can see the top words for that period. The last column shows the top words for the entire year. These are the words that were appeared most frequently in the comments over the entire year.



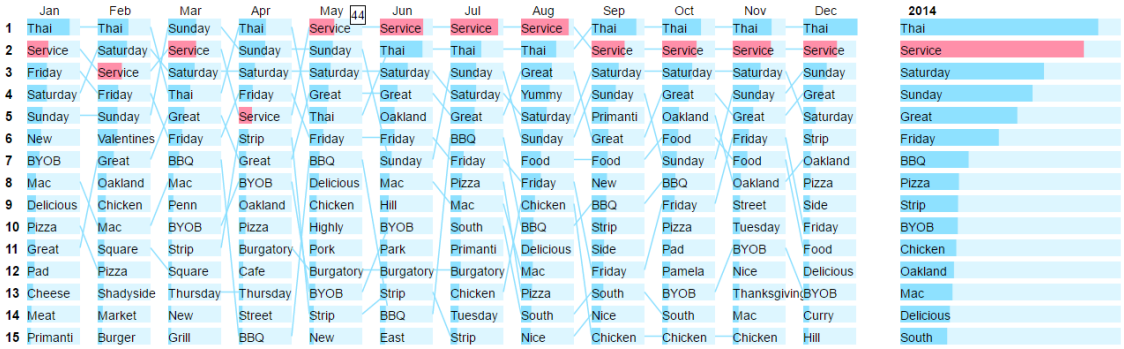
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	2014
1	Thai	Thai	Sunday	Thai	Service	Service	Service	Service	Thai	Thai	Thai	Thai	Thai
2	Service	Saturday	Service	Sunday	Sunday	Thai	Thai	Thai	Service	Service	Service	Service	Service
3	Friday	Service	Saturday	Saturday	Saturday	Saturday	Sunday	Great	Saturday	Saturday	Saturday	Sunday	Saturday
4	Saturday	Friday	Thai	Friday	Great	Great	Saturday	Yummy	Sunday	Great	Sunday	Great	Sunday
5	Sunday	Sunday	Great	Service	Thai	Oakland	Great	Saturday	Primanti	Oakland	Great	Saturday	Great
6	New	Valentines	Friday	Strip	Friday	Friday	BBQ	Sunday	Great	Food	Friday	Strip	Friday
7	BYOB	Great	BBQ	Great	BBQ	Sunday	Friday	Food	Food	Sunday	Food	Oakland	BBQ
8	Mac	Oakland	Mac	BYOB	Delicious	Mac	Pizza	Friday	New	BBQ	Oakland	Pizza	Pizza
9	Delicious	Chicken	Penn	Oakland	Chicken	Hill	Mac	Chicken	BBQ	Friday	Street	Side	Strip
10	Pizza	Mac	BYOB	Pizza	Highly	BYOB	South	BBQ	Strip	Pizza	Tuesday	Friday	BYOB
11	Great	Square	Strip	Burgatory	Pork	Park	Primanti	Delicious	Side	Pad	BYOB	Food	Chicken
12	Pad	Pizza	Square	Cafe	Burgatory	Burgatory	Burgatory	Mac	Friday	Pamela	Nice	Delicious	Oakland
13	Cheese	Shadyside	Thursday	Thursday	BYOB	Strip	Chicken	Pizza	South	BYOB	Thanksgiving	BYOB	Mac
14	Meat	Market	New	Street	Strip	BBQ	Tuesday	South	Nice	South	Mac	Curry	Delicious
15	Primanti	Burger	Grill	BBQ	New	East	Strip	Nice	Chicken	Chicken	Chicken	Hill	South

To answer the question about which words made it to the most frequent words of the year, we allow the user to hover on a word, upon which its occurrences get highlighted in all the columns wherever it appears, including the last column. So when a user hovers on a word 'Thai', it gets highlighted in all the columns and we can see that it seen in the 2014 column. However, hovering on the word 'food' shows that it does not appear in the year's list, meaning it wasn't that popular for the year.

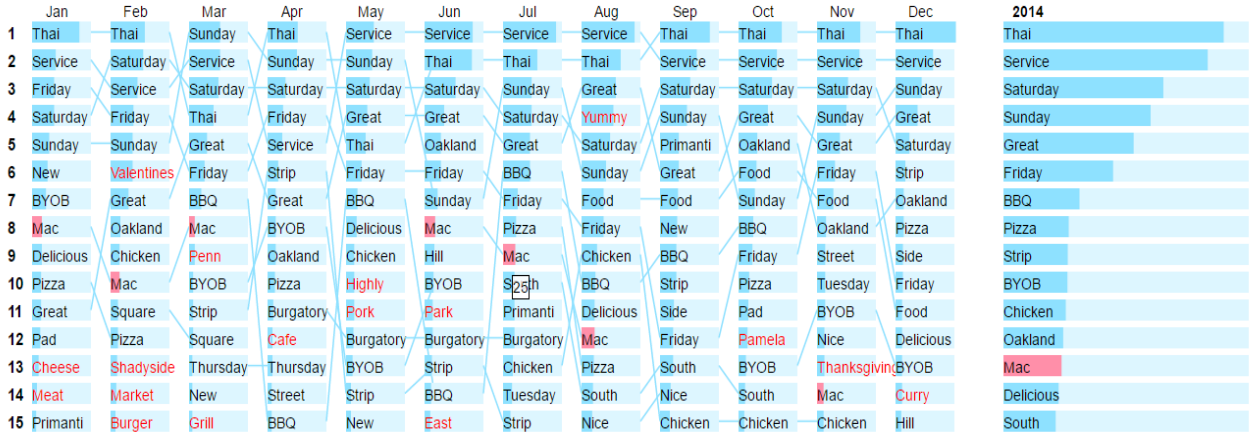
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	2014
1	Thai	Thai	Sunday	Thai	Service	Service	Service	Service	Thai	Thai	Thai	Thai	Thai
2	Service	Saturday	Service	Sunday	Sunday	Thai	Thai	Thai	Service	Service	Service	Service	Service
3	Friday	Service	Saturday	Saturday	Saturday	Saturday	Sunday	Great	Saturday	Saturday	Saturday	Sunday	Saturday
4	Saturday	Friday	Thai	Friday	Great	Great	Saturday	Yummy	Sunday	Great	Sunday	Great	Sunday
5	Sunday	Sunday	Great	Service	Thai	Oakland	Great	Saturday	Primanti	Oakland	Great	Saturday	Great
6	New	Valentines	Friday	Strip	Friday	Friday	BBQ	Sunday	Great	Food	Friday	Strip	Friday
7	BYOB	Great	BBQ	Great	BBQ	Sunday	Friday	Food	Food	Sunday	Food	Oakland	BBQ
8	Mac	Oakland	Mac	BYOB	Delicious	Mac	Pizza	Friday	New	BBQ	Oakland	Pizza	Pizza
9	Delicious	Chicken	Penn	Oakland	Chicken	Hill	Mac	Chicken	BBQ	Friday	Street	Side	Strip
10	Pizza	Mac	BYOB	Pizza	Highly	BYOB	South	BBQ	Strip	Pizza	Tuesday	Friday	BYOB
11	Great	Square	Strip	Burgatory	Pork	Park	Primanti	Delicious	Side	Pad	BYOB	Food	Chicken
12	Pad	Pizza	Square	Cafe	Burgatory	Burgatory	Burgatory	Mac	Friday	Pamela	Nice	Delicious	Oakland
13	Cheese	Shadyside	Thursday	Thursday	BYOB	Strip	Chicken	Pizza	South	BYOB	Thanksgiving	BYOB	Mac
14	Meat	Market	New	Street	Strip	BBQ	Tuesday	South	Nice	South	Mac	Curry	Delicious
15	Primanti	Burger	Grill	BBQ	New	East	Strip	Nice	Chicken	Chicken	Chicken	Hill	South

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	2014
1	Thai	Thai	Sunday	Thai	Service	Service	Service	Service	Thai	Thai	Thai	Thai	Thai
2	Service	Saturday	Service	Sunday	Sunday	Thai	Thai	Thai	Service	Service	Service	Service	Service
3	Friday	Service	Saturday	Saturday	Saturday	Saturday	Sunday	Great	Saturday	Saturday	Saturday	Sunday	Saturday
4	Saturday	Friday	Thai	Friday	Great	Great	Saturday	Yummy	Sunday	Great	Sunday	Great	Sunday
5	Sunday	Sunday	Great	Service	Thai	Oakland	Great	Saturday	Primanti	Oakland	Great	Saturday	Great
6	New	Valentines	Friday	Strip	Friday	Friday	BBQ	Sunday	Great	Food	Friday	Strip	Friday
7	BYOB	Great	BBQ	Great	BBQ	Sunday	Friday	Food	Food	Sunday	Food	Oakland	BBQ
8	Mac	Oakland	Mac	BYOB	Delicious	Mac	Pizza	Friday	New	BBQ	Oakland	Pizza	Pizza
9	Delicious	Chicken	Penn	Oakland	Chicken	Hill	Mac	Chicken	BBQ	Friday	Street	Side	Strip
10	Pizza	Mac	BYOB	Pizza	Highly	BYOB	South	BBQ	Strip	Pizza	Tuesday	Friday	BYOB
11	Great	Square	Strip	Burgatory	Pork	Park	Primanti	Delicious	Side	Pad	BYOB	Food	Chicken
12	Pad	Pizza	Square	Cafe	Burgatory	Burgatory	Burgatory	Mac	Friday	Pamela	Nice	Delicious	Oakland
13	Cheese	Shadyside	Thursday	Thursday	BYOB	Strip	Chicken	Pizza	South	BYOB	Thanksgiving	BYOB	Mac
14	Meat	Market	New	Street	Strip	BBQ	Tuesday	South	Nice	South	Mac	Curry	Delicious
15	Primanti	Burger	Grill	BBQ	New	East	Strip	Nice	Chicken	Chicken	Chicken	Hill	South

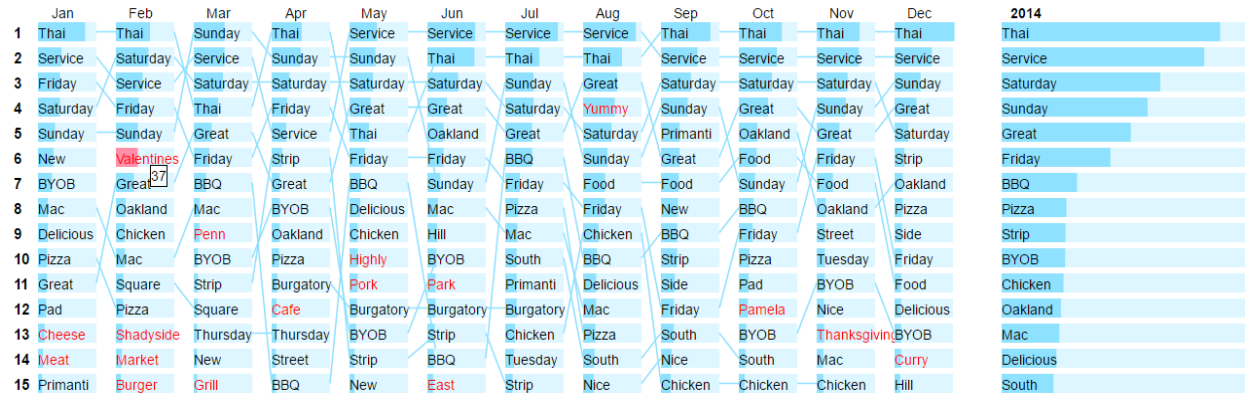
The next question is about how the rankings of the word change over time. To answer this, we use the lines combined with the bar highlights. When a user hovers on word, its occurrences in all the columns get highlighted, and interested user can follow the lines from one month to another to analyze its trend.



This implementation answers another question, which words disappear/re-appear over a period of time. As seen in the example of the keyword “Mac” in the chart below, we can see its appearance and disappearance trend. The highlighted words form a broken line which indicates that the word disappeared for some period only to re-appear again.

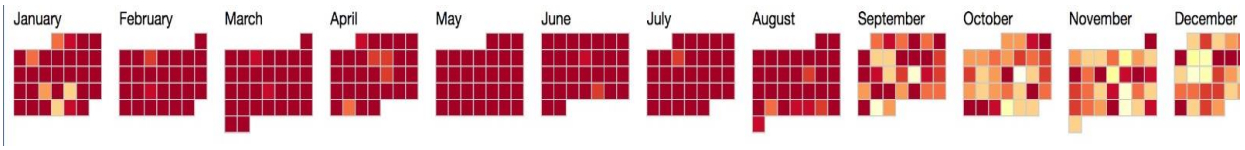


Considering another example of the word “Valentines” which was popular in the month of February or “Thanksgiving” whose frequency ranked 13 in the month of November. We see that such words are mostly seasonal and disappear for the rest of the months. To identify such outliers, we have highlighted them in red.



Upon clicking on a word, the bottom section of the dashboard populates the heatmap and comments for that word. The heat-map is a calendar-view representation of the word that shows how a word performed in that year, in each month, on each day. This chart is useful to find out whether a particular word performed better in a particular season – like fall or spring, or on special days like weekends or weekdays. It is interesting to find out trends using such a granular representation.

This calendar-view chart is populated for the clicked word which shows that this word was popular among the reviews during the months February-August, but its ranking reduced significantly in the subsequent months. This could be due to gaining importance of some other word over this.



Another tab in the bottom section populates 5 comments to understand the contexts in which the word was used.

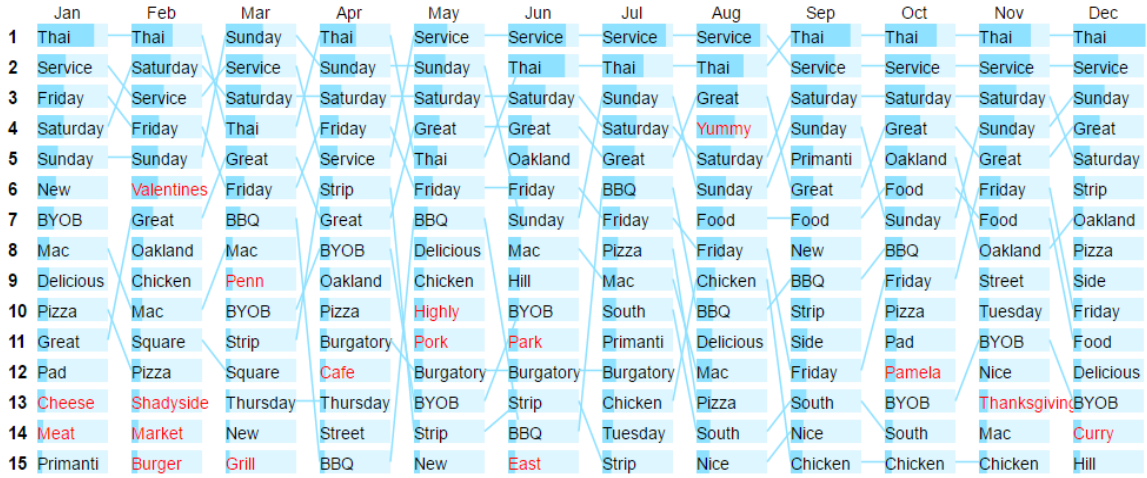
Top 5 Comments being Mentioned

- So after an evening of enjoying a comedy show at the Waterfront, we headed to Rock Bottom for some late night grub. The place was pretty busy, we had to walk around the bar area for several minutes until a nice couple told us they were about to leave and we could have their bar table. As nice as this was of them it was awkward standing over them as they finished their drinks paid their bill and left. So finally seated I look over the drinks and being the huge IPA lover that I am I went with that and it was indeed a good choice. The hubby had the Red Ale which was also rather good.The service was OK, not the best not the worst. But moving on to more important thing, the food. Talk about missing the mark hugely! I ordered the mini street fish tacos. You get four mini tacos with a marinated in lime piece of fish, and tomatillo salsa. They were just strokes away from inedible. I love fish tacos, I mean it is such a simple concept so why do people have such a hard time executing this dish? The shells were hard and literally split at the bottom when I picked them up. The sour cream sprayed all over the plate only made them messy and the poor fish was just so overcooked. If I wasn"t so hungry I wouldn"t have eaten them at all! Lesson learn on that.The husband had the BBQ Chicken Pizza. It had beans and corn with roasted red pepper on it along with some sour cream. My husband is usually the type that will order these lovely specialty pizza"s and ask for them to remove all the wonderful unique toppings! For instance, he would usually just say, Chicken BBQ and cheese, nothing else! I refused to let him do that this time, because I was really intrigued in the combination of ingredients on this pie. And boy was I jealous when it came out. It blew my crappy tacos out of the water! Should have stuck with something simple. Furthermore I could care less if I come back to this location. The beers were good but hell you can get a decent IPA just about anywhere anymore. And the underwhelming food sealed the deal. I know this is a chain, but it still doesn"t have to stink...
- Second time I try coming here for Korean BBQ. Apparently it"s only offered during weekends after 4:30.When we arrived it was pretty empty for wed night.Very overpriced. I got a bim bap that was very bland and has to add tons of soy sauce.
- After being told that Primanti Brothers was the best Pittsburgh had to offer, my faith in Pittsburgh food was gone. I thought, "If this is the best food in Pittsburgh, why would I possibly want to stay another day??"The answer: FAT HEADS SALOON.We went in expecting an okay lunch, and instead we both had multiple food-gasms. Fried perogie appetizer?? AWESOME. Perfect little pillows of potatoes and pleasure.I ordered the extreme pastrami and the boyfriend ordered the big beefy barbecue. My obsession with pastrami started about 3 months ago and this just made it more intense. The thick slices of meat and hot pepper mayo made for an amazing combination, but egg on a sandwich is always a selling point for me. Overall the best sandwich I have ever eaten. Ever. Based upon the sauce covering boyfriend"s face from nose-to-chin and ear-to-ear, I would say he was pretty pleased as well. North Carolina eastern BBQ is what I was raised on, so I"m not one to judge his sandwich. People of Pittsburgh, are you trying to get people out of your town by sending them to a hole in the wall with bland sandwiches? STOP IT! Send people to Fat Heads.
- We got PGH BBQ Company to cater a buffet dinner for ~100 people in our backyard. It was excellent! The food was all delicious. We had: Brisket, pulled pork, mac and cheese, cornbread, baked beans, portabella mushrooms, spinach salad and some trays of cookies/brownies. Guests raved about the brisket, but I was most impressed with the pulled pork - excellent eastern North Carolina style vinegar sauce! The bottles of hot sauce they provided were also very very good (on everything). Art took care of organizing everything for us and he was great to work with. He prefers phone to email but once I figured that out he was very responsive and he got all the details right. The women who came out do set up / clean up were also very good. Nice, prompt, efficient, hardworking - even helped us fit the leftovers in the freezer!
- Any place in Pittsburgh that serves Horchata automatically gets four stars. If you"ve never had it, you have to try it.Really good tacos. They are more gourmet than traditional, with interesting BBQ

Findings

Based on the derived data, we drilled through the patterns and found out the word rankings by sorting the words in a month based on frequency. This helped us display the data in a sorted order, which is a visually effective technique, since the most important ones appear at the top.

In order to find out the outliers in this data, we analyzed the rankings to find out those words which trended only in one particular month throughout the year. This technique brought out three interesting observations.



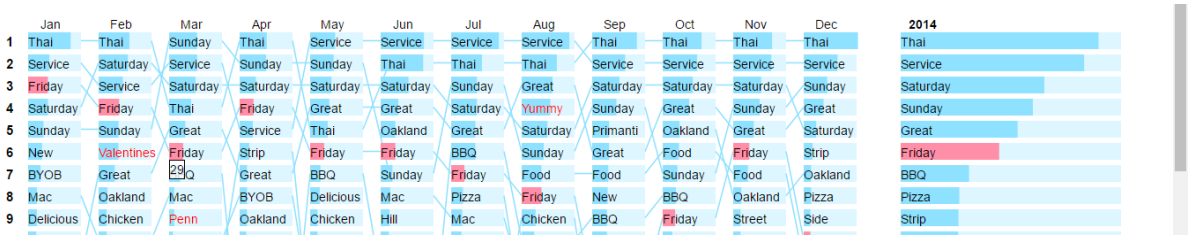
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	Thai	Thai	Sunday	Thai	Service	Service	Service	Service	Thai	Thai	Thai	Thai
2	Service	Saturday	Service	Sunday	Sunday	Thai	Thai	Thai	Service	Service	Service	Service
3	Friday	Service	Saturday	Saturday	Saturday	Saturday	Sunday	Great	Saturday	Saturday	Saturday	Sunday
4	Saturday	Friday	Thai	Friday	Great	Great	Saturday	Yummy	Sunday	Great	Sunday	Great
5	Sunday	Sunday	Great	Service	Thai	Oakland	Great	Saturday	Primanti	Oakland	Great	Saturday
6	New	Valentines	Friday	Strip	Friday	Friday	BBQ	Sunday	Great	Food	Friday	Strip
7	BYOB	Great	BBQ	Great	BBQ	Sunday	Friday	Food	Food	Sunday	Food	Oakland
8	Mac	Oakland	Mac	BYOB	Delicious	Mac	Pizza	Friday	New	BBQ	Oakland	Pizza
9	Delicious	Chicken	Penn	Oakland	Chicken	Hill	Mac	Chicken	BBQ	Friday	Street	Side
10	Pizza	Mac	BYOB	Pizza	Highly	BYOB	South	BBQ	Strip	Pizza	Tuesday	Friday
11	Great	Square	Strip	Burgatory	Pork	Park	Primanti	Delicious	Side	Pad	BYOB	Food
12	Pad	Pizza	Square	Cafe	Burgatory	Burgatory	Burgatory	Mac	Friday	Pamela	Nice	Delicious
13	Cheese	Shadyside	Thursday	Thursday	BYOB	Strip	Chicken	Pizza	South	BYOB	Thanksgiving	BYOB
14	Meat	Market	New	Street	Strip	BBQ	Tuesday	South	Nice	South	Mac	Curry
15	Primanti	Burger	Grill	BBQ	New	East	Strip	Nice	Chicken	Chicken	Chicken	Hill

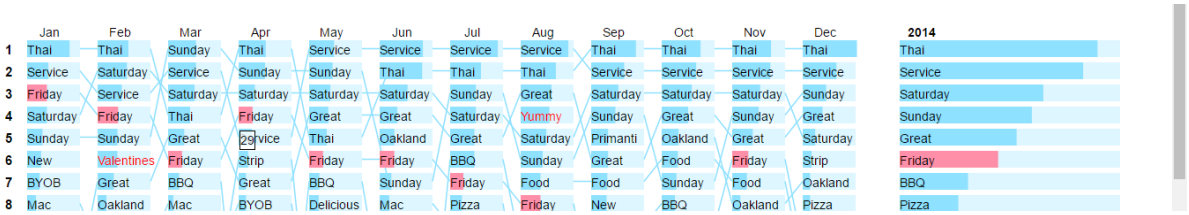
In the month of February, the word ‘Valentines’ trended as a popular keyword, since most comments mentioned eating at the restaurant on Valentine’s Eve. This shows that Valentines Eve is an important aspect to be considered for the month of February and restaurants can come up with special offers for Valentines. Similarly, in the month of November, the word thanksgiving trended.

The appearance of keyword Thai as the most popular implies that Thai is the most talked about cuisine among the people of Pennsylvania.

Another good observation is that, the trends are specific to the month, meaning, even if a word’s frequency has decreased or remains same in the subsequent month, it can still rise up in the ranking since it may be trending for that particular month compared to other words of that month. This can be seen from the following example.

The frequency of keyword ‘Friday’ in the month of March is 29, which is same in the month of April. However, it still rises up in the rank, since in the month of April, ‘Friday’ with the frequency of 29 is more popular than other words which have lower frequencies.




Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec 2014
1 Thai Thai Sunday Thai Service Service Service Service Thai Thai Thai Thai Thai
2 Service Saturday Service Sunday Sunday Thai Thai Thai Service Service Service Service Service
3 Friday Service Saturday Saturday Saturday Saturday Sunday Great Saturday Saturday Saturday Sunday Saturday
4 Saturday Friday Thai Friday Great Great Saturday Yummy Sunday Great Sunday Great Sunday
5 Sunday Sunday Great 29vice Thai Oakland Great Saturday Primanti Oakland Great Saturday Great
6 New Valentines Friday Strip Friday Friday BBQ Sunday Great Food Friday Strip Friday
7 BYOB Great BBQ Great BBQ Sunday Friday Food Food Sunday Food Oakland BBQ
8 Mac Oakland Mac BYOB Delicious Mac Pizza Friday New BBQ Oakland Pizza

Limitations and Future Works

The major limitation of this application is failure to understand the sentimental impact of a keyword with respect to the business in consideration. Meaning, given a word in the list, was it used in a positive way or a negative way? Are users happy about it or complaining about it? To answer these questions, given a chance to further develop this project, we would apply sentiment analysis and tag each word with a positive/negative/neutral attribute, and color code them such that the user of the application understands that this word is popular because people love something about it, or hate something about it. This will give deeper insights into the business trends.

To add to the existing functionality, we can process the data further to visualize the words for each business unit. For example, by choosing a restaurant name from a drop down, we can visualize all the information shown above, for just that one restaurant. This way, owners can track the trending words and find out what are people talking about "their" restaurant, and not restaurants in general.

Further, we would also like to understand co-occurrence of two words using a matrix, which will give an idea of the frequently co-occurring words, that can tell us which topics are closely related. Such as 'Cheese' and 'Pizza'. The idea of this visualization would be something like this.