

data_exploration

May 15, 2019

```
In [28]: import sys
         sys.path.append("../cs224u")
         import numpy as np
         import pandas as pd
         from matplotlib import pyplot as plt
         import math
         from nltk.corpus import stopwords
         from nltk.tokenize import word_tokenize
         from nltk.probability import FreqDist
         from collections import Counter
         import vsm
```

We'll begin by reading in all of the data and printing a few rows from each CSV file

```
In [9]: analysts = pd.read_csv('../data/analysts.csv')
        print(analysts.shape)
        analysts.head(10)
```

(187327, 12)

```
Out[9]:
```

	num	file-company-name	file-exchange	file-symbol	\
0	1	Ark Restaurants Corp.	NASDAQ	ARKR	
1	2	UniFirst Corp.	NYSE	UNF	
2	2	UniFirst Corp.	NYSE	UNF	
3	2	UniFirst Corp.	NYSE	UNF	
4	2	UniFirst Corp.	NYSE	UNF	
5	3	Resources Connection, Inc.	NASDAQ	RECN	
6	3	Resources Connection, Inc.	NASDAQ	RECN	
7	3	Resources Connection, Inc.	NASDAQ	RECN	
8	4	Sonic Corp.	NASDAQ	SONC	
9	4	Sonic Corp.	NASDAQ	SONC	

	file-published-on	file-title	\
0	2017-01-01 06:49:04	Ark Restaurants' (ARKR) CEO Michael Weinstein ...	
1	2017-01-04 13:33:06	UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...	
2	2017-01-04 13:33:06	UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...	
3	2017-01-04 13:33:06	UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...	

```

4 2017-01-04 13:33:06 UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...
5 2017-01-04 23:16:06 Resources Connection's (RECN) CEO Kate Duchene...
6 2017-01-04 23:16:06 Resources Connection's (RECN) CEO Kate Duchene...
7 2017-01-04 23:16:06 Resources Connection's (RECN) CEO Kate Duchene...
8 2017-01-04 21:41:07 Sonic's (SONC) CEO Cliff Hudson on Q1 2017 Res...
9 2017-01-04 21:41:07 Sonic's (SONC) CEO Cliff Hudson on Q1 2017 Res...

```

```

                                company-name                quarter-year-event \
0      Ark Restaurants Corp.                Q4 2016 Earnings Conference Call
1      UniFirst Corporation.                Q1 2017 Earnings Conference Call
2      UniFirst Corporation.                Q1 2017 Earnings Conference Call
3      UniFirst Corporation.                Q1 2017 Earnings Conference Call
4      UniFirst Corporation.                Q1 2017 Earnings Conference Call
5  Resources Connection, Inc.                Q2 2017 Earnings Conference Call
6  Resources Connection, Inc.                Q2 2017 Earnings Conference Call
7  Resources Connection, Inc.                Q2 2017 Earnings Conference Call
8                        Sonic Corp.  Q1 2017 Results Earnings Conference Call
9                        Sonic Corp.  Q1 2017 Results Earnings Conference Call

```

```

symbol-exchange symbol                datetime \
0      NaN      ARKR  December 30, 2016 10:00 A.M. ET
1      NaN      UNF   January 4, 2017 10:00 AM ET
2      NaN      UNF   January 4, 2017 10:00 AM ET
3      NaN      UNF   January 4, 2017 10:00 AM ET
4      NaN      UNF   January 4, 2017 10:00 AM ET
5      NaN      RECN   January 4, 2017 17:00 ET
6      NaN      RECN   January 4, 2017 17:00 ET
7      NaN      RECN   January 4, 2017 17:00 ET
8      NaN      SONC   January 04, 2017 05:00 PM ET
9      NaN      SONC   January 04, 2017 05:00 PM ET

```

```

                                analyst-company
0                        Bruce Geller - DGHM
1                        John Healy - Northcoast Research
2  Kevin Steinke - Barrington Research Associates...
3                        Joe Box - KeyBanc Capital Markets
4  Justin Hauke - Robert W. Baird & Company, Inc.
5                        Andrew Steinerman - JPMorgan
6                        Kevin McVeigh - Deutsche Bank
7                        Mark Marcon - Baird
8                        Greg Francfort - Bank of America
9                        Nicole Miller - Piper Jaffray

```

```

In [10]: execs = pd.read_csv('./data/executives.csv')
          print(execs.shape)
          execs.head(10)

```

```
(111916, 12)
```

```

Out[10]:
num      file-company-name file-exchange file-symbol \
0      1      Ark Restaurants Corp.      NASDAQ      ARKR
1      1      Ark Restaurants Corp.      NASDAQ      ARKR
2      2      UniFirst Corp.      NYSE      UNF
3      2      UniFirst Corp.      NYSE      UNF
4      3      Resources Connection, Inc.      NASDAQ      RECN
5      3      Resources Connection, Inc.      NASDAQ      RECN
6      3      Resources Connection, Inc.      NASDAQ      RECN
7      4      Sonic Corp.      NASDAQ      SONC
8      4      Sonic Corp.      NASDAQ      SONC
9      5      Walgreens Boots Alliance, Inc.      NASDAQ      WBA

file-published-on      file-title \
0 2017-01-01 06:49:04 Ark Restaurants' (ARKR) CEO Michael Weinstein ...
1 2017-01-01 06:49:04 Ark Restaurants' (ARKR) CEO Michael Weinstein ...
2 2017-01-04 13:33:06 UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...
3 2017-01-04 13:33:06 UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...
4 2017-01-04 23:16:06 Resources Connection's (RECN) CEO Kate Duchene...
5 2017-01-04 23:16:06 Resources Connection's (RECN) CEO Kate Duchene...
6 2017-01-04 23:16:06 Resources Connection's (RECN) CEO Kate Duchene...
7 2017-01-04 21:41:07 Sonic's (SONC) CEO Cliff Hudson on Q1 2017 Res...
8 2017-01-04 21:41:07 Sonic's (SONC) CEO Cliff Hudson on Q1 2017 Res...
9 2017-01-05 16:33:05 Walgreens Boots Alliance's (WBA) CEO Stefano P...

company-name      quarter-year-event \
0      Ark Restaurants Corp.      Q4 2016 Earnings Conference Call
1      Ark Restaurants Corp.      Q4 2016 Earnings Conference Call
2      UniFirst Corporation.      Q1 2017 Earnings Conference Call
3      UniFirst Corporation.      Q1 2017 Earnings Conference Call
4      Resources Connection, Inc.      Q2 2017 Earnings Conference Call
5      Resources Connection, Inc.      Q2 2017 Earnings Conference Call
6      Resources Connection, Inc.      Q2 2017 Earnings Conference Call
7      Sonic Corp.      Q1 2017 Results Earnings Conference Call
8      Sonic Corp.      Q1 2017 Results Earnings Conference Call
9      Walgreens Boots Alliance, Inc.      Q1 2017 Results Earnings Conference Call

symbol-exchange symbol      datetime \
0      NaN      ARKR      December 30, 2016 10:00 A.M. ET
1      NaN      ARKR      December 30, 2016 10:00 A.M. ET
2      NaN      UNF      January 4, 2017 10:00 AM ET
3      NaN      UNF      January 4, 2017 10:00 AM ET
4      NaN      RECN      January 4, 2017 17:00 ET
5      NaN      RECN      January 4, 2017 17:00 ET
6      NaN      RECN      January 4, 2017 17:00 ET
7      NaN      SONC      January 04, 2017 05:00 PM ET
8      NaN      SONC      January 04, 2017 05:00 PM ET
9      NaN      WBA      January 05, 2017, 08:30 AM ET

```

```

                                executive-positions
0  Bob Stewart - President and Chief Financial Of...
1  Michael Weinstein - Chairman and Chief Executi...
2      Ronald Croatti - President and CEO
3      Steven Sintros - SVP and CFO
4      Alice Washington - Interim General Counsel
5      Kate Duchene - Chief Executive Officer
6      Herb Mueller - Chief Financial Officer
7      Cliff Hudson - CEO
8      Claudia San Pedro - CFO
9  Gerald Gradwell - Senior Vice President, Inves...

```

```

In [11]: qna = pd.read_csv('./data/qna.csv')
         print(qna.shape)
         qna.head(10)

```

(1828257, 14)

```

Out[11]:  num      file-company-name  file-exchange  file-symbol  file-published-on \
0      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
1      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
2      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
3      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
4      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
5      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
6      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
7      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
8      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04
9      1  Ark Restaurants Corp.      NASDAQ      ARKR  2017-01-01 06:49:04

```

```

                                file-title      company-name \
0  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
1  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
2  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
3  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
4  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
5  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
6  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
7  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
8  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.
9  Ark Restaurants' (ARKR) CEO Michael Weinstein ...  Ark Restaurants Corp.

```

```

                                quarter-year-event  symbol-exchange  symbol \
0  Q4 2016 Earnings Conference Call      NaN      ARKR
1  Q4 2016 Earnings Conference Call      NaN      ARKR
2  Q4 2016 Earnings Conference Call      NaN      ARKR
3  Q4 2016 Earnings Conference Call      NaN      ARKR

```

4	Q4 2016 Earnings Conference Call	NaN	ARKR
5	Q4 2016 Earnings Conference Call	NaN	ARKR
6	Q4 2016 Earnings Conference Call	NaN	ARKR
7	Q4 2016 Earnings Conference Call	NaN	ARKR
8	Q4 2016 Earnings Conference Call	NaN	ARKR
9	Q4 2016 Earnings Conference Call	NaN	ARKR

	datetime	order	name \
0	December 30, 2016 10:00 A.M. ET	1	Operator
1	December 30, 2016 10:00 A.M. ET	2	Bruce Geller
2	December 30, 2016 10:00 A.M. ET	3	Bob Stewart
3	December 30, 2016 10:00 A.M. ET	4	Bruce Geller
4	December 30, 2016 10:00 A.M. ET	5	Bob Stewart
5	December 30, 2016 10:00 A.M. ET	6	Bruce Geller
6	December 30, 2016 10:00 A.M. ET	7	Bob Stewart
7	December 30, 2016 10:00 A.M. ET	8	Bruce Howard Geller
8	December 30, 2016 10:00 A.M. ET	9	Bob Stewart
9	December 30, 2016 10:00 A.M. ET	10	Bruce Geller

	content
0	[Operator Instructions] Our first question com...
1	Hi Good morning guys.
2	Hi Bruce, how are you?
3	Good thanks. Glad to hear things have turned a...
4	Yes, that includes an additional week last yea...
5	So it's not nearly as...
6	No, no.
7	as dismal as it might imply. What were the ap...
8	About 2%. Probably down, somewhere along that ...
9	Got it. Okay. Can you just kind of go through ...

```
In [12]: statements = pd.read_csv('./data/statements.csv')
print(statements.shape)
statements.head(10)
```

(204656, 14)

```
Out [12]:
```

	num	file-company-name	file-exchange	file-symbol \
0	1	Ark Restaurants Corp.	NASDAQ	ARKR
1	1	Ark Restaurants Corp.	NASDAQ	ARKR
2	1	Ark Restaurants Corp.	NASDAQ	ARKR
3	1	Ark Restaurants Corp.	NASDAQ	ARKR
4	2	UniFirst Corp.	NYSE	UNF
5	2	UniFirst Corp.	NYSE	UNF
6	2	UniFirst Corp.	NYSE	UNF
7	2	UniFirst Corp.	NYSE	UNF
8	2	UniFirst Corp.	NYSE	UNF

9 3 Resources Connection, Inc. NASDAQ RECN

	file-published-on	file-title \
0	2017-01-01 06:49:04	Ark Restaurants' (ARKR) CEO Michael Weinstein ...
1	2017-01-01 06:49:04	Ark Restaurants' (ARKR) CEO Michael Weinstein ...
2	2017-01-01 06:49:04	Ark Restaurants' (ARKR) CEO Michael Weinstein ...
3	2017-01-01 06:49:04	Ark Restaurants' (ARKR) CEO Michael Weinstein ...
4	2017-01-04 13:33:06	UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...
5	2017-01-04 13:33:06	UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...
6	2017-01-04 13:33:06	UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...
7	2017-01-04 13:33:06	UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...
8	2017-01-04 13:33:06	UniFirst's (UNF) CEO Ronald Croatti on Q1 2017...
9	2017-01-04 23:16:06	Resources Connection's (RECN) CEO Kate Duchene...

	company-name	quarter-year-event \
0	Ark Restaurants Corp.	Q4 2016 Earnings Conference Call
1	Ark Restaurants Corp.	Q4 2016 Earnings Conference Call
2	Ark Restaurants Corp.	Q4 2016 Earnings Conference Call
3	Ark Restaurants Corp.	Q4 2016 Earnings Conference Call
4	UniFirst Corporation.	Q1 2017 Earnings Conference Call
5	UniFirst Corporation.	Q1 2017 Earnings Conference Call
6	UniFirst Corporation.	Q1 2017 Earnings Conference Call
7	UniFirst Corporation.	Q1 2017 Earnings Conference Call
8	UniFirst Corporation.	Q1 2017 Earnings Conference Call
9	Resources Connection, Inc.	Q2 2017 Earnings Conference Call

	symbol-exchange	symbol	datetime	order \
0	NaN	ARKR	December 30, 2016 10:00 A.M. ET	1
1	NaN	ARKR	December 30, 2016 10:00 A.M. ET	2
2	NaN	ARKR	December 30, 2016 10:00 A.M. ET	3
3	NaN	ARKR	December 30, 2016 10:00 A.M. ET	4
4	NaN	UNF	January 4, 2017 10:00 AM ET	1
5	NaN	UNF	January 4, 2017 10:00 AM ET	2
6	NaN	UNF	January 4, 2017 10:00 AM ET	3
7	NaN	UNF	January 4, 2017 10:00 AM ET	4
8	NaN	UNF	January 4, 2017 10:00 AM ET	5
9	NaN	RECN	January 4, 2017 17:00 ET	1

	name	content
0	Presentation	NaN
1	Operator	Greetings, and welcome to the Ark Restaurants ...
2	Bob Stewart	Thank you, operator. Good morning, and thank y...
3	Michael Weinstein	Hi, everybody. If you read the release, we're ...
4	Presentation	NaN
5	Operator	Ladies and gentlemen, thank you for standing b...
6	Steven Sintros	Thank you and welcome to the UniFirst Corporat...
7	Ronald Croatti	Thank you, Steve. I'd like to welcome everyone...
8	Steven Sintros	Thanks, Ron. As Ron mentioned, revenues for th...

Let's begin by seeing how many distinct statements and Q&A sessions we have (these numbers should be appx. the same, but some statements may not have a corresponding Q&A session).

```
In [13]: num_statements = 0
         for i in range(statements.shape[0]):
             if statements.iloc[i]['order'] == 1: num_statements += 1

         num_qna = 0
         for i in range(qna.shape[0]):
             if qna.iloc[i]['order'] == 1: num_qna += 1

         print("There are %s distinct statements" % str(num_statements))
         print("There are %s distinct Q&A sessions" % str(num_qna))
```

```
There are 33227 distinct statements
There are 33019 distinct Q&A sessions
```

Now let's look at a complete statement and its corresponding Q&A session.

```
In [14]: def get_first_k_examples(k, verbose=False):
         results = []
         num_processed = 0
         row_statement, row_qna = 0, 0
         while num_processed < k:
             curr = {}
             curr['statement'] = []
             curr['qna'] = []
             while(True):
                 if len(curr['statement']) > 0 and statements.iloc[row_statement]['order'] == 1:
                     curr_speaker = str(statements.iloc[row_statement]['name'])
                     if verbose: print(curr_speaker + '\n')
                     curr_content = str(statements.iloc[row_statement]['content'])
                     if verbose: print(curr_content + '\n')
                     curr['statement'].append((row_statement, curr_speaker, curr_content))
                     row_statement += 1

             if verbose:
                 print('=' * 75 + '\n')
                 print("Moving on to Q&A\n")
                 print('=' * 75 + '\n')

             while(True):
                 if len(curr['qna']) > 0 and qna.iloc[row_qna]['order'] == 1: break
                 curr_speaker = str(qna.iloc[row_qna]['name'])
                 if verbose: print(curr_speaker + '\n')
                 curr_content = str(qna.iloc[row_qna]['content'])
```

```

        if verbose: print(curr_content + '\n')
        curr['qna'].append((row_qna, curr_speaker, curr_content))
        row_qna += 1
        results.append(curr)
        num_processed += 1
    return results

```

```
In [18]: get_first_k_examples(k=1, verbose=True)
```

Presentation

nan

Operator

Greetings, and welcome to the Ark Restaurants Fourth Quarter and Full Year 2016 Results Confer

Bob Stewart

Thank you, operator. Good morning, and thank you for joining us on our conference call for the

Michael Weinstein

Hi, everybody. If you read the release, we're comparing a 52-week year against a 53-week year.

=====

Moving on to Q&A

=====

Operator

[Operator Instructions] Our first question comes from the line of Bruce Geller with DGHM. Plea

Bruce Geller

Hi Good morning guys.

Bob Stewart

Hi Bruce, how are you?

Bruce Geller

Good thanks. Glad to hear things have turned around so far in the current quarter. Was that 7%

Bob Stewart

Yes, that includes an additional week last year. So when you're looking at it, you're comparing

Bruce Geller

So it's not nearly as...

Bob Stewart

No, no.

Bruce Howard Geller

as dismal as it might imply. What were the apples-to-apples comp be, if you could come up with

Bob Stewart

About 2%. Probably down, somewhere along that line, 2% to 3%.

Bruce Geller

Got it. Okay. Can you just kind of go through some of the trends by market? And then also, to t

Michael Weinstein

It's Michael, Bruce. There are - certainly, New York had a very strong 11 weeks. And we think v

Bruce Geller

Thank you.

Michael Weinstein

My pleasure Bruce.

Operator

Our next question comes from the line of Alan Goldberg, Private Investor. Please proceed with y

Unidentified Analyst

Good morning gentlemen. I am a private investor, but I'm not a novice investor. I have been in

Michael Weinstein

Please.

Unidentified Analyst

I live in Palm Beach and I am the one that travels 70 miles to get down there, more than once,

Michael Weinstein

Well, thank you very much. It's truly appreciated. I can address those issues with you. But st

Unidentified Analyst

Well, I don't know if you can still hear me, but very honestly in...

Michael Weinstein

I can.

Unidentified Analyst

In my previous life, I would have said, my God, this is the opportunity I've been waiting for.

Michael Weinstein

Yes. We have a problem here in terms of taking the company private. And number one is, I'm the

Unidentified Analyst

But I thought you've just agreed with me that we're both young men, sort of.

Michael Weinstein

We're very rigorous, but I don't want to sneeze one day and have a subordinated lender come to

Unidentified Analyst

I was just being a flippant. I understand completely. Go on please. I'm sorry.

Michael Weinstein

So those are the things, those are the parameters that are sort of controlling the fact that w

Unidentified Analyst

Would you like to sell it?

Michael Weinstein

Right. So when we bought it, we said, hey, this is earning \$1 million to \$1.5 million, probably,

Unidentified Analyst

You know anybody in the world would grab that.

Michael Weinstein

Yes. So the same situation is true in Shuckers. The same situation is true in Alabama. These g

Unidentified Analyst

What about the situation of your - let's call it, your buyback program?

Michael Weinstein

Buyback program, quite honestly, was instituted for 2 reasons. Number one, we're looking at th

Unidentified Analyst

You have answered my questions, I thank you so much for your time. I would like to at sometime

Michael Weinstein

Thank you.

Unidentified Analyst

Thank you so much for the wonderful job you are doing and I appreciate it. Thank you so very m

Michael Weinstein

Thank you. I guess we are ready for other questions.

Operator

Our next question comes from the line of Jeff Kay [ph] Private Investor. Please proceed with y

Unidentified Analyst

Hi, good morning gentlemen and congratulations on continued good performance. I have a couple o

Michael Weinstein

We have two. One is Canyon Road on First Avenue and 75th Street, which is a small restaurant t

Bob Stewart

16.

Michael Weinstein

We have 16 years left on that lease. That lease was bought by a New York real estate developer

Unidentified Analyst

Yes, I'm familiar with it.

Michael Weinstein

We think we have a great installation. Reviews for Clyde's are sensational if you go to OpenTa

Unidentified Analyst

Okay. Thank you. And just a follow-up to the previous caller's question regarding possibly goin

Michael Weinstein

Yes. We had a few discussions with Simpson Thacher on that. And quite frankly, this may be very

Unidentified Analyst

Okay. Happy New Year.

Michael Weinstein

And the second point, we don't like to have calls on December 30 at 10:00 in the morning, espe

Unidentified Analyst

Alright. Happy and healthy New Year.

Michael Weinstein

Same to you.

Operator

Mr. Weinstein, theres no further questions at this time. So, I will turn it back to you for cl

Michael Weinstein

Well, thank you very much. I'm pleased that at least we have two people that seem to be very, v

```
Out[18]: [{'statement': [(0, 'Presentation', 'nan'),  
                        (1,  
                        'Operator',
```


'Unidentified Analyst',
 "Well, I don't know if you can still hear me, but very honestly in..."),
 (19, 'Michael Weinstein', 'I can.'),
 (20,
 'Unidentified Analyst',
 "In my previous life, I would have said, my God, this is the opportunity I've been
 (21,
 'Michael Weinstein',
 "Yes. We have a problem here in terms of taking the company private. And number one
 (22,
 'Unidentified Analyst',
 "But I thought you've just agreed with me that we're both young men, sort of."),
 (23,
 'Michael Weinstein',
 "We're very rigorous, but I don't want to sneeze one day and have a subordinated
 (24,
 'Unidentified Analyst',
 "I was just being a flippant. I understand completely. Go on please. I'm sorry.")
 (25,
 'Michael Weinstein',
 "So those are the things, those are the parameters that are sort of controlling the
 (26, 'Unidentified Analyst', 'Would you like to sell it?'),
 (27,
 'Michael Weinstein',
 "Right. So when we bought it, we said, hey, this is earning \$1 million to \$1.5 million
 (28,
 'Unidentified Analyst',
 'You know anybody in the world would grab that.'),
 (29,
 'Michael Weinstein',
 "Yes. So the same situation is true in Shuckers. The same situation is true in Al
 (30,
 'Unidentified Analyst',
 "What about the situation of your - let's call it, your buyback program?"),
 (31,
 'Michael Weinstein',
 "Buyback program, quite honestly, was instituted for 2 reasons. Number one, we're
 (32,
 'Unidentified Analyst',
 'You have answered my questions, I thank you so much for your time. I would like to
 (33, 'Michael Weinstein', 'Thank you.'),
 (34,
 'Unidentified Analyst',
 "Thank you so much for the wonderful job you are doing and I appreciate it. Thank
 (35,
 'Michael Weinstein',
 "Thank you. I guess we are ready for other questions.'),
 (36,

```

'Operator',
'Our next question comes from the line of Jeff Kay [ph] Private Investor. Please p
(37,
'Unidentified Analyst',
"Hi, good morning gentlemen and congratulations on continued good performance. I l
(38,
'Michael Weinstein',
"We have two. One is Canyon Road on First Avenue and 75th Street, which is a smal
(39, 'Bob Stewart', '16.'),
(40,
'Michael Weinstein',
"We have 16 years left on that lease. That lease was bought by a New York real es
(41, 'Unidentified Analyst', "Yes, I'm familiar with it."),
(42,
'Michael Weinstein',
"We think we have a great installation. Reviews for Clyde's are sensational if yo
(43,
'Unidentified Analyst',
"Okay. Thank you. And just a follow-up to the previous caller's question regarding
(44,
'Michael Weinstein',
"Yes. We had a few discussions with Simpson Thacher on that. And quite frankly, th
(45, 'Unidentified Analyst', 'Okay. Happy New Year.'),
(46,
'Michael Weinstein',
"And the second point, we don't like to have calls on December 30 at 10:00 in the
(47, 'Unidentified Analyst', 'Alright. Happy and healthy New Year.'),
(48, 'Michael Weinstein', 'Same to you.'),
(49,
'Operator',
'Mr. Weinstein, theres no further questions at this time. So, I will turn it back
(50,
'Michael Weinstein',
"Well, thank you very much. I'm pleased that at least we have two people that seen

```

Now let's examine the value returned by `get_first_k_examples`

```

In [19]: samples = get_first_k_examples(100, verbose=False)
         print("%s samples were returned" % len(samples))

```

100 samples were returned

Now let's examine the speakers in both the statement and Q&A sections

```

In [20]: # Parse out the speakers
         statement_speakers = []
         qna_speakers = []
         for ex in samples:

```

```

    for elem in ex['statement']:
        curr_speaker = elem[1]
        statement_speakers.append(curr_speaker)
    for elem in ex['qna']:
        curr_speaker = elem[1]
        qna_speakers.append(curr_speaker)
print("There are %s distinct speakers in the statements section" % len(set(statement_speakers)))
print("There are %s distinct speakers in the Q&A section" % len(set(qna_speakers)))

```

There are 265 distinct speakers in the statements section
There are 858 distinct speakers in the Q&A section

```

In [21]: # Count the number of times each speaker spoke
statement_speakers_counts = Counter(statement_speakers)
qna_speakers_counts = Counter(qna_speakers)

# Remove 'Operator' and 'Presentation' speakers
del statement_speakers_counts['Presentation']
del statement_speakers_counts['Operator']
del qna_speakers_counts['Operator']

print("There are %s distinct speaker count values for the statements" % len(set(statement_speakers_counts.values())))
print(sorted(set(statement_speakers_counts.values())))
print('')

print("There are %s distinct speaker count values for the Q&A" % len(set(qna_speakers_counts.values())))
print(sorted(set(qna_speakers_counts.values())))

```

There are 5 distinct speaker count values for the statements
[1, 2, 3, 10, 12]

There are 44 distinct speaker count values for the Q&A
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44]

```

In [22]: # Let's get the key corresponding to the highest Q&A speaker count
for key in qna_speakers_counts.keys():
    if qna_speakers_counts[key] == 95: print(key)

```

Unidentified Analyst

Now let's examine the vocabularies in both the statement and Q&A sections. We will drop stop words and for each token we will ensure its length is greater than 1 to get rid of punctuation.

```

In [23]: # Define our set of stop words
stop_words = set(stopwords.words('english'))
print(stop_words)

```



```
{'which', "weren't", 'while', 'myself', 'yourselves', 'needn', 'y', 'didn', 'were', 'had', "don
```

```
In [24]: # Parse out content
statement_tokens = []
qna_tokens = []
for ex in samples:
    for elem in ex['statement']:
        curr_tokens = elem[2]
        curr_tokens = word_tokenize(curr_tokens)
        curr_tokens = [tok.lower() for tok in curr_tokens if (not tok.lower() in stopp
        statement_tokens.extend(curr_tokens)
    for elem in ex['qna']:
        curr_tokens = elem[2]
        curr_tokens = word_tokenize(curr_tokens)
        curr_tokens = [tok.lower() for tok in curr_tokens if (not tok.lower() in stopp
        qna_tokens.extend(curr_tokens)
print("There are %s distinct tokens in the statements section" % len(set(statement_tok
print("There are %s distinct tokens in the Q&A section" % len(set(qna_tokens)))
```

There are 10098 distinct tokens in the statements section

There are 11421 distinct tokens in the Q&A section

```
In [25]: # Count the number of times each token appeared
statement_token_counts = Counter(statement_tokens)
qna_token_counts = Counter(qna_tokens)

print("There are %s distinct token count values for the statements" % len(set(statement
print(sorted(set(statement_token_counts.values()))))
print('')

print("There are %s distinct token count values for the Q&A" % len(set(qna_token_count
print(sorted(set(qna_token_counts.values()))))
```

There are 272 distinct token count values for the statements

[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26

There are 336 distinct token count values for the Q&A

[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26

```
In [42]: # Let's look at the frequency distributions for both the statement and Q&A sections
statements_fdist = FreqDist(statement_tokens)
qna_fdist = FreqDist(qna_tokens)
```

```
In [43]: print("Here are the most common words in the statements:")
print(statements_fdist.most_common(100))
```

Here are the most common words in the statements:

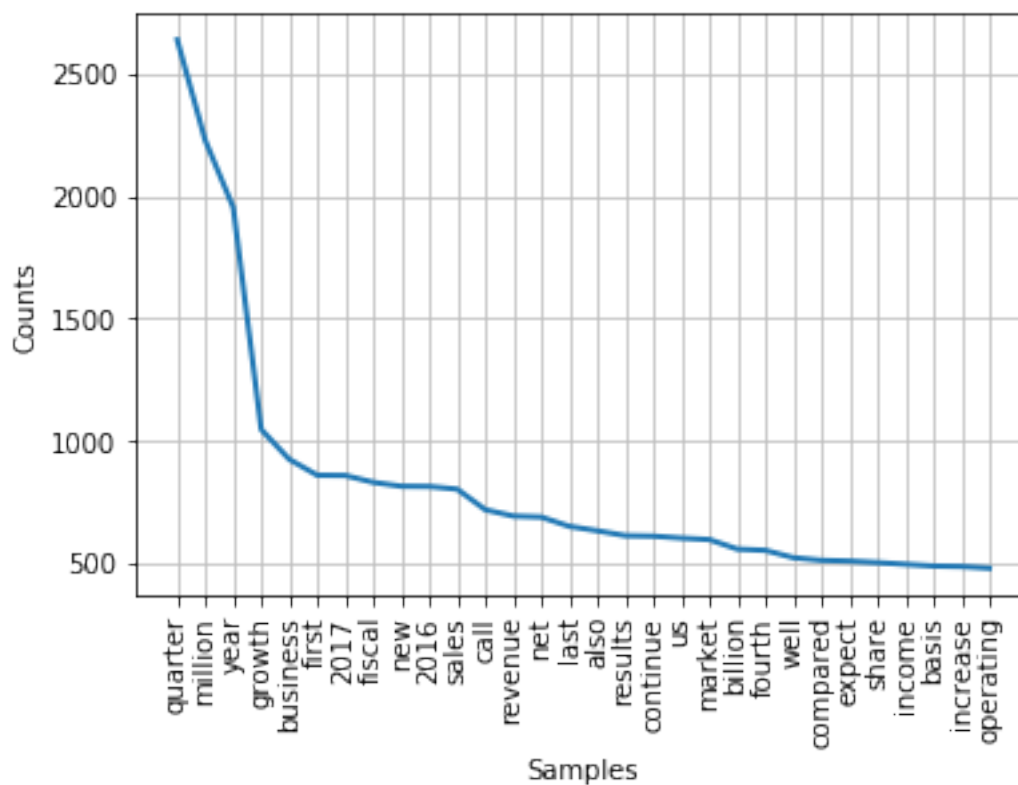
```
[('quarter', 2639), ('million', 2228), ('year', 1954), ('growth', 1046), ('business', 924), ('
```

```
In [44]: print("Here are the most common words in the statements:")  
         print(qna_fdist.most_common(100))
```

Here are the most common words in the statements:

```
[('think', 3023), ('would', 1858), ('question', 1809), ('year', 1801), ('going', 1652), ('quar
```

```
In [38]: statements_fdist.plot(30, cumulative=False)
```



```
In [37]: qna_fdist.plot(30, cumulative=False)
```

