NLP Final Project_Joan Lee

March 17, 2018

0.0.1 Homework Prompt:

Illinois is famous for being one of the very few states in the country with negative population growth. The objective of your final project is to:

- Identify the key reasons for the declining population by extracting meaningful insights from unstructured text
- Provide actionable recommendations on what can be done to reverse this trend
- You have a collection of ~30 days' worth of news articles in RCC located at: /home/kadochnikov/news/news_chicago_il.pkl – related to either Chicago and / or Illinois.

To complete your assignment, I suggest considering the following steps:

- Clean-up the noise (eliminate articles irrelevant to the analysis)
- Detect major topics
- Identify top reasons for population decline (negative sentiment)
- Suggest corrective actions
- Demonstrate how the city / state can attract new businesses (positive sentiment)
- Leverage appropriate NLP techniques to identify organizations and people and apply targeted sentiment

Why businesses should stay in IL or move into IL? - Create appropriate visualization to summarize your recommendations (i.e. word cloud chart or bubble chart)
Why residents should stay in IL or move into IL?

- Create appropriate visualization to summarize your recommendations (i.e. word cloud chart or bubble chart)

Additional guidance:
Default sentiment will likely be wrong from any software package and will require significant tweaking Either keyword / dictionary approach or Labeling and classification You are encouraged to explore a combination several techniques to identify key topics: Topic modeling (i.e. LSA and TF-IDF) Classification (hand-label several topics on a sample and then train classifier) Clustering (cluster topics around pre-selected keywords or word vectors)
Please limit your work to 7 PowerPoint slides. On your slides you will want to provide: - Executive Summary - Methodology and source data overview - Actionable recommendations
Please submit your actual program codes (i.e. Python Notebook) along with your PowerPoint – as a separate attachment Your presentation should be targeted toward business audience and

must not contain any code snippets You are welcome to use any software packages of your choice to complete the assignment

```
In [2]: import sys
        print(sys.version)

3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32) [MSC v.1900 64 bit (AMD64)]


In [15]: import time
         import math
         import re
         from textblob import TextBlob
         import pandas as pd
         import numpy as np

         import nltk as nltk
         from nltk.corpus import stopwords
         from nltk.stem.wordnet import WordNetLemmatizer

         import string

         import warnings
         warnings.filterwarnings(action='ignore', category=UserWarning, module='gensim')

         import gensim
         from gensim import corpora, models

In [67]: import re
         import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         from statistics import mode
         import nltk as nltk
         import nltk.corpus
         from nltk.text import Text
         import sys
         from collections import Counter, defaultdict
         import sklearn
         from sklearn.model_selection import train_test_split
         from sklearn.feature_extraction.text import CountVectorizer, HashingVectorizer, Tfidf
         from sklearn.naive_bayes import MultinomialNB
         from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
         from sklearn.linear_model import LogisticRegression, SGDClassifier
         from sklearn import metrics

In [4]: import pickle
        directory ='C://Users//Joan//OneDrive//06.WI18_nlp//'
```

```
In [5]: text = pd.read_pickle(directory+'news_chicago_il.pkl')

In [6]: path = directory + text

In [7]: text.head(5)

Out[7]:                       crawled language  \
        0  2018-01-22T06:31:13.011+02:00  english
        1  2018-01-22T06:43:12.000+02:00  english
        2  2018-01-22T06:52:12.009+02:00  english
        3  2018-01-22T06:53:16.017+02:00  english
        4  2018-01-22T06:54:04.067+02:00  english


                                                      text  \
        0  Discussion in ' Midwest ' started by sportscra...
        1  CHICAGO  Hundreds of ballet dancers between t...
        2  Chicago rapper Fredo Santana has reportedly di...
        3  E-mail \nPolice are searching for a 27-year-ol...
        4  Documentary about Miss Continental pageant set...


                                                     title
        0               Best Stops Chicago Red Line?
        1  AP PHOTOS: Talented young dancers compete in C...
        2  Chicago Rapper Fredo Santana Reportedly Dead A...
        3  Man, 27, reported missing from Rogers Park | C...
        4  Documentary about Miss Continental pageant set...

In [8]: text.shape

Out[8]: (40585, 4)

In [9]: text['language'].value_counts()

Out[9]: english    40585
        Name: language, dtype: int64

In [10]: # Remove special characters to avoid problems with analysis
         text['text_clean'] = text['text'].map(lambda x: re.sub('[^a-zA-Z0-9 @ . , : - _]', ''

In [11]: pd.set_option('display.max_colwidth', 100)
         text[['text', 'text_clean']].head(5)

Out[11]:
         0  Discussion in ' Midwest ' started by sportscrazed2 , Jan 21, 2018 . sportscrazed2
         1  CHICAGO  Hundreds of ballet dancers between the ages of 9 and 19 have been audition
         2  Chicago rapper Fredo Santana has reportedly died, according to several of his frier
         3  E-mail \nPolice are searching for a 27-year-old man who went missing from the Roger
         4  Documentary about Miss Continental pageant set to debut in Chicago by Carrie Maxwel
```

3

```
          0  Discussion in  Midwest  started by sportscrazed2 , Jan 21, 2018 . sportscrazed2 Asp
          1  CHICAGO  Hundreds of ballet dancers between the ages of 9 and 19 have been audition
          2  Chicago rapper Fredo Santana has reportedly died, according to several of his frien
          3  Email Police are searching for a 27yearold man who went missing from the Rogers Pa
          4  Documentary about Miss Continental pageant set to debut in Chicago by Carrie Maxwel
```

In [12]: *#combining title with text-- could help train data to look for keywords and relevant*
         text['text_combi']= text['title'].astype(str) + ' '+text['text_clean']

In [13]: text.head(5)

Out[13]:   crawled language \
         0 2018-01-22T06:31:13.011+02:00   english
         1 2018-01-22T06:43:12.000+02:00   english
         2 2018-01-22T06:52:12.009+02:00   english
         3 2018-01-22T06:53:16.017+02:00   english
         4 2018-01-22T06:54:04.067+02:00   english

         0 Discussion in ' Midwest ' started by sportscrazed2 , Jan 21, 2018 . sportscrazed2 A
         1 CHICAGO Hundreds of ballet dancers between the ages of 9 and 19 have been audition
         2 Chicago rapper Fredo Santana has reportedly died, according to several of his frien
         3 E-mail \nPolice are searching for a 27-year-old man who went missing from the Roger
         4 Documentary about Miss Continental pageant set to debut in Chicago by Carrie Maxwel

                                                          title \
         0   Best Stops Chicago Red Line?
         1   AP PHOTOS: Talented young dancers compete in Chicago
         2   Chicago Rapper Fredo Santana Reportedly Dead At 27 - MTV
         3   Man, 27, reported missing from Rogers Park | Chicago Informer
         4 Documentary about Miss Continental pageant set to debut in Chicago

         0 Discussion in Midwest started by sportscrazed2 , Jan 21, 2018 . sportscrazed2 Asp
         1 CHICAGO Hundreds of ballet dancers between the ages of 9 and 19 have been audition
         2 Chicago rapper Fredo Santana has reportedly died, according to several of his frien
         3 Email Police are searching for a 27yearold man who went missing from the Rogers Pa
         4 Documentary about Miss Continental pageant set to debut in Chicago by Carrie Maxwel

         0 Best Stops Chicago Red Line? Discussion in Midwest started by sportscrazed2 , Jan
         1 AP PHOTOS: Talented young dancers compete in Chicago CHICAGO Hundreds of ballet da
         2 Chicago Rapper Fredo Santana Reportedly Dead At 27 - MTV Chicago rapper Fredo Santa
         3 Man, 27, reported missing from Rogers Park | Chicago Informer Email Police are sear
         4 Documentary about Miss Continental pageant set to debut in Chicago Documentary abou

In [41]: *#picking out a random sample to skim topics to determine whether to keep or reject--*
         # then train a model to do so on the entire data set

```python
from sklearn.model_selection import train_test_split
text_combi = text['text_combi']
#text_sample = text_combi.sample(n=300)

train, test = train_test_split(text_combi, test_size=0.992)
```

In [42]: `train.count()`

Out[42]: 324

In [47]:
```python
for index, i in enumerate(train):
    print(index, i)

train=train.reset_index(drop=True)
```

Out[47]:
```
0       Man dies hours after being shot leaving Southwest Side party - Chicago Tribune
1       The absurdity of Poland&apos;s Holocaust law - Chicago The Streisand effect is
2       My View: 2011 worker's comp changes beginning to benefit Illinois My View: 201
3       Instant analysis: Michigan State 87, Illinois 74 | MLive.com Registration on or
4       Illinois governor candidate stands by ad blasted as racist CHICAGO AP  A Repub
5       2012 Chicago White Sox Opening Day Ticket Vs Tigers Image not available Photos
6       BC-BKC--Missouri St.-Loyola of Chicago Box BCBKCMissouri St.Loyola of Chicago
7       51st Annual Elmhurst College Jazz Festival Begins Feb. 22 - Chicago Tribune The
8       Therapist / Respiratory Therapy Job in Chicago, Illinois / Hospital | USA Jobs
9       6 shot within 3 hours to start President's Day weekend in Chicago | 1 killed,
10      Toys R Us closing seven Chicago area stores Toys R Us closing seven Chicago are
11      Loyola-Chicago beats Drake 80-57 behind big 2nd-half run The Associated Press .
12      17 killed in mass shooting at high school in Parkland, Florida | Gary/Chicago (
13      Chicago ranks second worst of most populous U.S. cities for fiscal health Chica
14      Holocaust denier running for Congress has no opponents in Illinois Republican
15      Illinois mother pleads guilty in babys suffocation death ALORTON, Ill. AP  A s
16      Hoy 22.00 NBA Orlando Magic Vs Chicago Bulls EN VIVO  Esperar 10 Seg Publicida
17      Chicago Restaurant Becomes Go-To for Eagles Fans unable to retrieve fulltext c
18      Black NHL player Devante Smith-Pelly taunted by Chicago fans Black NHL player
19      Snow blankets Chicago as winter storm sweeps U.S. Midwest  12:40 PM EST Snow
20      Group aims to bridge gap between NU students, Chicago musicians Send Email Can
21      Loyola holds off Valparaiso for 22nd win, 11th in 12 games - Chicago Tribune D
22      Duerson Act would ban pre-teen tackle football in Illinois Duerson Act would ba
23      Oak Park Police Investigate Alleged Abduction ń CBS Chicago Sponsored By Oak P
24      College notes: Evanston&apos;s Will Jones paces Monmouth to MWC hoops title - (
25      Illinois lawmakers inch closer to new hospital tax despite worries Illinois la
26      Carrollton, Illinois - Insurance Agent - Carrollton, IL, , USA #jobs #Carrollt
27      What You Should Know About Sales Jobs In Chicago By Jeffrey Lee Thousands of pe
28      The Art Institute of Chicago Share Visit the worlds best museum, the Art Instit
29      Volvo and The Petfinder Foundation partner at 2018 Chicago Auto Show - Roadsho
                                    ...
294     Police: Chicago man cited for punching a co-worker in the face Police: Chicago
295     Catcher Kevan Smith prepared to battle for role with White Sox (Chris Kuc/Chic
296     Bearded Villain Chicago Respect Than Bearded Villain Chicago Respect Thank you
```

```
297    VITU Ushers in a New Era of Tax, Title and Registration Processing in Illinois
298    Illinois GOP Rep. Under Fire For Ad Mocking Transgender Community, Feminists Ne
299    Bulls lose 4th straight, fall 110-96 to Bucks - Chicago Tribune For three games
300    Credit Repair Dieterich, Illinois   Dieterich, Illinois   Dieterich, Illinois
301    Illinois schools say pension shift would undo funding revamp Illinois schools s
302    Woman killed in domestic shooting among 8 shot in Chicago - Chicago Tribune A 5
303    Oakton helps to spotlight Muslim-American culture by joining &apos;Coming Toget
304    To Buy Affordable Lawn Equipment Chicago Deserves To Be Prioritized By Robert I
305    Bradley downs Illinois St 70-58 in I-74 rivalry game Bradley downs Illinois St
306    Federal court in Chicago considers potentially landmark LGBT housing case Chica
307    Name: Courtney Davis, Michael Rohd, Nina Rohd, and LolaLocation: Evanston, Illi
308    Southern Illinois Soybean Wheat Meeting Did you head back to Ohio todayThere is
309    Fayne ties career-high with 24 pionts, Illinois State beats Valpo 76-65 Fayne t
310    LOG CABIN- Redwood, Live Edge, Mid Century Mountain Home, 6 piece set (Olde Chi
311    (USA-IL-CHICAGO) Copywriter Home View All Jobs  2,158,749  Robert Half The Crea
312    A Fond Farewell to Crate & Barrels Chicago Flagship Store After 27 years, Crate
313    (USA-IL-Chicago) Scala Architect Home View All Jobs  2,195,278  CyberCoders Sca
314    Chicago is bracing for its biggest snow since 2015 Chicago is bracing for its I
315    Wiretap records Illinois governor hopeful's comments on race Wiretap records II
316    Four Chicago teens charged with trying to break into Naperville house, police s
317    Share Buyback Plan Initiated by Owens-Illinois (OI) Board Tweet OwensIllinois I
318    This Is The Best Illinois City To Live In | Patch real estate Shared from the I
319    Did You Catch Chicago Wests Adorable Cameo in Kylie Jenners Birth Announcement'
320    Fast Hitch from Monroe Center, Illinois to Birmingham, Alabama I have a guy try
321    Illinois sets up new hotline for victims of sexual harassment Illinois sets up
322    Las Vegas man arrested in Chicago in death of infant U.S. Marshals have arreste
323    Victim: 'There was no way out' from clutches of 'violent pimp' from Naperville
Name: text_combi, Length: 324, dtype: object
```

```python
In [51]: for index, i in enumerate(train):
            print(index, i)
```

```
0 Man dies hours after being shot leaving Southwest Side party - Chicago Tribune   TribuneMan
1 The absurdity of Poland&apos;s Holocaust law - Chicago The Streisand effect is one of the cur
2 My View: 2011 worker's comp changes beginning to benefit Illinois My View: 2011 workers comp
3 Instant analysis: Michigan State 87, Illinois 74 | MLive.com Registration on or use of this s
4 Illinois governor candidate stands by ad blasted as racist CHICAGO AP  A Republican lawmaker
5 2012 Chicago White Sox Opening Day Ticket Vs Tigers Image not available Photos not available
6 BC-BKC--Missouri St.-Loyola of Chicago Box BCBKCMissouri St.Loyola of Chicago Box LOYOLA OF (
7 51st Annual Elmhurst College Jazz Festival Begins Feb. 22 - Chicago Tribune The Elmhurst Coll
8 Therapist / Respiratory Therapy Job in Chicago, Illinois / Hospital | USA Jobs Club Kindred I
9 6 shot within 3 hours to start President's Day weekend in Chicago | 1 killed, 12 wounded in s
10 Toys R Us closing seven Chicago area stores Toys R Us closing seven Chicago area stores Jan
11 Loyola-Chicago beats Drake 80-57 behind big 2nd-half run The Associated Press January 25, 20
12 17 killed in mass shooting at high school in Parkland, Florida | Gary/Chicago Crusader 17 ki
13 Chicago ranks second worst of most populous U.S. cities for fiscal health Chicago ranks seco
14 Holocaust denier running for Congress has no opponents in Illinois Republican primary Holoca
15 Illinois mother pleads guilty in babys suffocation death ALORTON, Ill. AP  A southern Illino
```

16 Hoy 22.00 NBA Orlando Magic Vs Chicago Bulls EN VIVO   Esperar 10 Seg Publicidad Wait 10 Se
17 Chicago Restaurant Becomes Go-To for Eagles Fans unable to retrieve fulltext contentIf youre
18 Black NHL player Devante Smith-Pelly taunted by Chicago fans Black NHL player Devante Smith
19 Snow blankets Chicago as winter storm sweeps U.S. Midwest   12:40 PM EST Snow blankets Chica
20 Group aims to bridge gap between NU students, Chicago musicians Send Email Cancel A new grou
21 Loyola holds off Valparaiso for 22nd win, 11th in 12 games - Chicago Tribune Donte Ingram ha
22 Duerson Act would ban pre-teen tackle football in Illinois Duerson Act would ban preteen ta
23 Oak Park Police Investigate Alleged Abduction ń CBS Chicago Sponsored By Oak Park Police Inv
24 College notes: Evanston's Will Jones paces Monmouth to MWC hoops title - Chicago Tribu
25 Illinois lawmakers inch closer to new hospital tax despite worries Illinois lawmakers inch
26 Carrollton, Illinois - Insurance Agent - Carrollton, IL, , USA #jobs #Carrollton pls RT   Ca
27 What You Should Know About Sales Jobs In Chicago By Jeffrey Lee Thousands of people depend
28 The Art Institute of Chicago Share Visit the worlds best museum, the Art Institute of Chicag
29 Volvo and The Petfinder Foundation partner at 2018 Chicago Auto Show - Roadshow To learn mo
30 Bernice Engelmann, 90, Nashville, Illinois Interment will follow in the church cemetery. Cre
31 Chicago Yacht Club Race to Mackinac update With a supporting team of many Old Goats, multihu
32 \n CHICAGO \n ,\n Feb. 13, 2018 \n \/PRNewswire\/ n CHICAGO n ,n Feb. 13, 2018 n PRNewswire
33 Church becomes 1st in southern Illinois with solar panels Its the first church in southern
34 Central Illinois Flying Aces vs Green Bay Gamblers, Feb 17, 2018 Preview, Watch and Bet, S
35 (USA-IL-Chicago) Global Business Services Shared Services Senior Consultant - Strategy & Op
36 Gymnastics notes: Nicole Kane makes state 2 months after breaking leg - Chicago Tribune The
37 Quality Inn & Suites Lincoln - Lincoln, Illinois, United States No news or articles are ava
38 Owens-Illinois Inc (OI) Files 10-K for the Fiscal Year Ended on December 31, 2017 Peter Lyn
39 2 of the top 3 U.S. hotels are in Chicago, report says Chicago is no longer home to the top
40 Bridges, No. 6 Michigan State hold off Illinois 87-74 Bridges, No. 6 Michigan State hold of
41 Two weeks in chicago - Illinois Hi... my husband and I are travelling to Chicago for two wee
42 North Central Illinois development group touts region North Central Illinois development gro
43 Chicago Bulls' Kris Dunn returns from concussion, starts against Toronto Raptors | NBA.com C
44 The Latest: Slain Chicago officer remembered as 'model cop' The Latest: Slain Chicago office
45 15-year-old Kentucky shooting suspect may be tried as an adult - Chicago The 15yearold boy a
46 Miami - OH vs. Northern Illinois, 2-3-2018 - Prediction & Preview
47 Miami Heat havent been the same since Tyler Johnsons injury in Chicago Jordan Mickey out o
48 CSO's Chinese New Year concert and more of the best things to do in Chicago this weekend - 2
49 Can middle-class candidate defeat millionaires in Illinois? Can middleclass candidate defea
50 All-Star Game over, Patrick Kane, Blackhawks have uphill climb ahead - Chicago Residual feel
51 (USA-IL-Chicago) Events Specialist Marriott Events Specialist in Chicago , Illinois Posting
52 Popular Chicago beauty blogger Jenna Colgrove revi Popular Chicago beauty blogger Jenna Colg
53 Illinois Supreme Court Knock Down Park Gun Ban by John Crump Illinois Supreme Court Knock D
54 Chicago police to meet with federal, state officials to combat rise in carjackings News Loca
55 U.S. Olympic figure skating team revels in Hollywood love after bronze-medal win - Chicago
56 Credit Repair Virden, Illinois   Virden, Illinois   Virden, Illinois   Virden, Illinois,   c
57 Illinois State Universities moving $35 million to Brookfield real estate fund Illinois State
58 Illinois Tool Works Inc. (ITW) Shares Sold by Sentry Investment Management LLC Tweet Sentry
59 FBI clashes with Trump, has 'grave concerns' on Russia memo | Chicago Sun-Times WASHINGTON
60 The Cadillac Escala Concept Sedan Wins the Beauty Pageant at the 2018 Chicago Auto Show The
61 Duclair speaks out on incident with fans and Smith-Pelly | NBC Sports Chicago FOLLOW US Face
62 Illinois Veterinarian Has More To Say About This Dog Flu Dog Insurance News Illinois Veteri
63 Ingredion for Emerging Business partners with food business incubator, The Hatchery Chicago

64 2019 Drive Chip & Putt Contest Illinois Section 24000 N Old McHenry Rd. Kildeer, IL 60047
65 Principal Consultant, Project Leadership - Chicago - Chicago, IL, , USA #jobs #Chicago pls
66 Chicago Sun-Times Axes Long Time Theatre Critic Hedy Weiss Weiss did not respond to a reques
67 Joe Maddon's 2018 theme: 'Putting art back in the game' - Chicago Tribune Jo
68 Chicago- "Colour My World" with Lyrics (3:00)
69 5-Year-Old Boy Playing In Snow Impaled By Curb Marker ń CBS Chicago Sponsored By 5YearOld B
70 (USA-IL-Chicago) Summer Associate (Cons. Level)- Technology, MBA, Consulting - Carnegie Mel
71 5 to Watch: Vonn, Chen, Rippon Highlight Star-Studded Competition - NBC Chicago Clive RoseGe
72 Chicago police commander killed, 5 others wounded in city shootings - Chicago Tribune A Chi
73 Automotive Appearance Chemicals Market Analysis, Market Size, Regional Outlook, Competitive
74 Transport a 3 month old Great Pyranese Pup from Orlando area to Illinois heathr.jpg 35.01 K
75 Chicago Cillian Shadow Day - lovemy3dogs We cast a shadow on something wherever we stand.
76 Upcoming Classes in Portland and Chicago: Writing Contemporary Fairy Tales 28 Jan Cannon Be
77 Basmati Rice Global Industry 2018 Sales, Supply and Consumption Forecasts to 2022...- Chica
78 Chicago Bulls vs Orlando Magic, Feb 12, 2018 Preview, Watch and Bet, Score NBA , Chicago
79 Selling at the Illinois Angus Futurity | THIS FRIDAY ADVERTISEMENT COMPLIMENTS OF SULLIVAN S
80 Chicago Fire S06E11 WEB x264 TBS ettv
81 2-Minute Drill: Rizzo returns, don't count out White Sox, Rihanna's birthday | Chicago Sun-
82 Luzviminda Benaires Lewis, 61, Mount Vernon, Illinois Next SEARCH OTHER SOURCES A greeting
83 Washington woes dissolve at your favorite Chicago restaurant   Food is comfort. And m
84 Toyota unveils new off-road trucks at the Chicago Auto Show Toyota unveils new offroad truc
85 Snow Totals Across the Chicago Area Parts of the Chicago area saw about 8 to 13 inches of sn
86 Sarah Leonard and Christie Nyquist, Suburban Chicago Real Estate Professionals Discuss The
87 What Makes This Death Penalty Case in Chicago on Murder of Yingying Zhang Unusual Newser U
88 Tivo Series 4 Dolby HDMI (Sherman, Illinois) $20 model name number: TCD 746320 QR Code Lin
89 Illinois gun rights group has other gun free zones in sights to be shot down More gun free z
90 Illinois Gov. Bruce Rauner to deliver State of the State Email Want to read through the spe
91 Rejuvenated Tim Anderson ready for big year after reflective offseason - Chicago Tribune Al
92 Illinois to get $1.7M from federal for anti-poverty programs | State and Regional | herald-
93 During Burn It Down, part of the floor was stolen. Peoria, Illinois : avengedsevenfold subm
94 The Cubs reportedly agree to 6-year deal with Yu Darvish | NBC Sports Chicago FOLLOW US Face
95 Linda Harvey, 72, Coulterville, Illinois We use to talk and talk at the post office I know s
96 Chicago fans ejected after heckling Capitals' Devante Smith-Pelly with racist taunt Black N
97 Keith B. Coley, 59, Caseyville, Illinois Next SEARCH OTHER SOURCES Funeral: Memorial service
98 Westwood Management Corp IL Has $4.04 Million Holdings in Illinois Tool Works Inc. (ITW) Tw
99 Consumer Credit Counseling Loop, Chicago, IL Loop, Chicago, IL Loop, Chicago, IL Loop
100 Watch Live Video Gov. Bruce Rauner delivers Illinois State of the State address | FOX2now
101 CBS 2 Weather Watch (11AM, Jan. 29, 2018) ń CBS Chicago 2018 CBS Broadcasting Inc. All Rig
102 Car Dolly New Tires w/ Straps (Rock Falls,Illinois) $600 condition: good QR Code Link to T
103 Photos : Manasuku Nacchindi Movie Meet and Greet in Chicago Manasuku Nacchindi Movie Meet
104 Brian Urlacher: 'I changed the way middle linebacker was played' - Chicago Tribu
105 African-American vegans in redefining soulful food - Chicago Tribune When the world first t
106 IP MKE/Chicago: FT: CWBG 17, 4pk NG SR; ISO: CW16,17
107 4 TICKETS CLEVELAND INDIANS @ CHICAGO WHITE SOX 6/12 *Sec 518 FRONT ROW AISLE* Image not av
108 Kurt Warner New England Patriots Quarterback Tom Bradys NFL dominance is more impressive t
109 Mold Removal and the Moldy Debris That Comes Along With It - Mold Solutions | Inspection |
110 Image from page 287 of The Chicago amateur base ball annual and inter-city base ball assoc
111 Man arrested after breaching security at Illinois airport 0 68 PEORIA, Ill. Several weeke

8

112 James Hudson delivers 'closeout' special as Andrean stops East Chicago Central James Hudson
113 MARCH 9 and 10, 2018 ~ Chicago Heights, Illinois Dancers Boot Camp 2018 Marked By Sonship
114 Mourners: Slain Chicago officer was 'one of the good guys' Mourners: Slain Chicago officer
115 Bored*** (Chicago) age: 25   Hmu if you down to cruise tonight or hang out. To weed out spa
116 Credit rating agency upgrades Chicago debt, says investors face 'small risk of loss' One Wa
117 The Chicago Women's March WOMEN , BEST OF This Saturday I attended the second inaugural Wom
118 Illinois Smoking Ban Has Not Damaged Casino Revenue, Study Claims News Live Casinos Illi
119 Chevrolet Traverse RS: Chicago 2018
120 Official Rules: Like to win Chicago Tickets | FOX2now.com Official Rules: Like to win Chica
121 GOP opponents Durkin and Straub say Illinois pension reform is priority, but have different
122 (USA-IL-Chicago) Staff Accountant- Entry Level Home View All Jobs 46,596 Publicis Media S
123 See first photos of Kim K takes her newborn baby girl, Chicago. Kim Kardashian West has bee
124 Another Chicago-area school closes due to flu Chicago Tribune Another Chicagoarea school c
125 No. 17 Ohio State holds off pesky Illinois to win 75-67 - StamfordAdvocate Close Image 1 of
126 Chicago to Cleveland Hyperloop project takes key first step Hyperloop Transportation Techno
127 Porter Airlines Canada 2-Days Flights/Tickets Seat Sale + Save up to Extra 25% Off with Pro
128 Chicago 1 May 7, 2009, 10:39:47 PM Software Adobe Photoshop Lightroom 5.0 Windows To
129 Wildwood and Chicago Music Exchange Experience?? I saw in AGF a few references to Wildwood
130 Milla Nova 2018 Wedding Dresses Collection | Chicago Campaign   dresses Milla Nova 2018 We
131 Search for Buried Treasure Linked to Illinois Man's Death At Yellowstone By Laurel Wamsley
132 Heavy Snow, White-Out Conditions Cause 11-Car Crash in Indiana - NBC Chicago The latest loc
133 YRS8-RMYF: STATEMENT OF ILLINOIS LAW ON ADVANCE DIRECTIVES YRS8RMYF: STATEMENT OF ILLINOIS
134 China Is Trying to Takeover the Chicago Stock Exchange, SEC Blocks Attempt DAHBOO77 Publish
135 Chicago Capital Management LLC Buys Shares of 204,602 Almost Family Inc (NASDAQ:AFAM) Poste
136 (USA-IL-Chicago) MRT-Cancer Registry You must be a U.S. Citizen to apply for this job Desig
137 (USA-IL-CHICAGO) Sr. Accountant Home View All Jobs 44,280 Robert Half Finance Accounting
138 // Yes Gun Control Works // Just look at Chicago and Baltimore // ... Idiots // Over 600 Mu
139 Anthony Rizzo addresses Florida school shooting vigil: 'I grew up in Stoneman Douglas&
140 Cynthia Ann Metz, 59, Cahokia, Illinois Next SEARCH OTHER SOURCES Funeral services will be
141 RIP John Mahoney. he was so, so nice to me once at the Chicago premiere of... RIP John Maho
142 Chicago Rivet & Machine Co. Announces Increase In Quarterly Cash Dividend   11:50 AM EST Ch
143 BC-BKC--W. Illinois-Nebraska-Omaha Box BCBKCW. IllinoisNebraskaOmaha Box NEBRASKAOMAHA 71,
144 Slain off-duty Chicago police officer was department veteran Slain offduty Chicago police o
145 Unruly Nigerian man, David Okeone restrained on Emirates flight from Dubai to Chicago An el
146 The Ski Jumpers Of A Small Illinois Town The Ski Jumpers Of A Small Illinois Town By Quinn
147 Chicago residents wonder if a root canal dental treatment is worth it   Chicago residents w
148 (USA-IL-Chicago) Engineer -On Call Wyndham Worldwide Engineer On Call in Chicago , Illinois
149 2018 Chicago Auto Show: Vehicle Coverage 2018 Chicago Auto Show: Vehicle Coverage Tuesday,
150 Video: 'Old Friends' from Porchlight's 'Merrily' - Chicago Tribune
151 Illinois Tool Works Inc. (ITW) Shares Sold by Loring Wolcott & Coolidge Fiduciary Advisors
152 Is Bode Miller sorry for blaming a female skier's decline on marriage or just inept?
153 Lady Doritos? Its Not Happening, PepsiCo Says ń CBS Chicago Lady Doritos Its Not Happening
154 Commercial HVAC Southern Illinois: How To Maintain Your HVAC Unit By Larry Cook Most busine
155 Alerta de Google: "hotel in chicago" Alerta de Google: hotel in chicago hotel in chicago
156 Man sentenced in 1985 rape, murder of suburban Illinois teen Topix Illinois Man sentenced
157 Biggest snow in over a year bearing down on large swath of the Chicago area beginning Thurs
158 (USA-IL-Chicago) UX/UI Designer Home View All Jobs 2,204,949 Robert Half The Creative Gro
159 Chicago Fire S06E10 iNTERNAL 720p WEB x264-BAMBOOZLE Torrent File Content 1 file Chicago.Fi

9

160 Seeking Half-English Budgie Near Central Illinois My budgie, Ozzie, 8.5 yrs old died recent
161 4 injured in South Side fire, including 2 police officers - Chicago Tribune News Local news
162 Illinois early voting starts, but not available to all Illinois early voting starts, but no
163 Monaco scores 29 as Ball State outlasts Northern Illinois - Ball State Daily News (press re
164 Serial stowaway nabbed at Chicago airport days after release Serial stowaway nabbed at Chic
165 Indiana's defense delivers knockout punch against Illinois Juwan Morgan was one of thr
166 Chicago Police Commander Killed on Busy Downtown Block The commander, who was chasing a man
167 Grain volatility and La Niña; Chicago futures options fraud case; Cboe on crypto Giving New
168 Fremont Residence (Chicago) Fremont Residence Eclectic Laundry Room Chicago
169 Illinois Officials Propose Raising Smoking Age ń CBS St. Louis Illinois Officials Propose
170 Parents guide to the movies - Chicago Tribune PETER RABBITPGWhat its about: The beloved Bea
171 SEC Blocks Chinese Takeover Of Chicago Stock Exchange In a decision that many had anticipat
172 Food Stamps stopped in Chicago.. It's a TRAP
173 Long-dreaded Amazon threat to drug middlemen draws closer - Chicago A day that prescription
174 Mourners: Slain Chicago officer was one of the good guys Mourners: Slain Chicago officer wa
175 Finding The Best Expert For Marriage Counseling Chicago By Eric Robinson When you have a fa
176 Another Chicago-area school closes due to flu A school in North Riverside will be closed fo
177 White Sox to extend protective netting at Guaranteed Rate Field - Chicago The White Sox are
178 The best and worst numbers to have in your 2018 Super Bowl squares pool - Chicago Tribune T
179 (USA-IL-Chicago) SENIOR MANAGER ACCOUNTS PAYABLE Loyola University Chicago SENIOR MANAGER A
180 Southwest Air's Chicago De-Icing Stumble Was Third in Two Months Southwest Airlines Co.s ca
181 WH to Pull Nomination of Top Environmental Adviser, Climate Change Skeptic - NBC Chicago Au
182 AXA buys Chicago-based Maestro for $155 million - John Pletz Technology Blog - Crain's Chic
183 China Is Trying to Takeover the Chicago Stock Exchange, SEC Blocks Attempt   n a decision t
184 Chicago Restaurant Week Highlights 370 Spots to Dine Kids Chicago Restaurant Week Highlight
185 Shirley Bowman, 83, Medora, Illinois Next SEARCH OTHER SOURCES Visitation will be 58 pm Tue
186 University Of Chicago Law School - Law School University Of Chicago The University of Chica
187 CTA riders facing significant Monday morning delays | Chicago Sun-Times Nearly every CTA li
188 Kim and Kanye's baby daughter Chicago seen for first time Kim and Kanyes baby daughter Chic
189 Chicago West Birth Certificate OUT Kim Kardashian and Kanye West already revealed the name
190 Southern Illinois man killed when attempt to clear drain causes noxious gas, police say Sou
191 2018 Gentle Warrior Awards incorporates Pullman Porters celebrating their legacy and descen
192 NWT NBA Lauri Markkanen #24 Chicago Bulls Red Swingman Jersey Mens Allsorce.com Check your
193 Loyola-Chicago beats Drake 80-57 behind big 2nd-half run LoyolaChicago beats Drake 8057 beh
194 Loyola-Chicago clinches at least share of MVC title EVANSVILLE, Ind. Marques Townes scored
195 Holly Kramer, MD, MPH, Loyola Kidney Specialist and Researcher, Named President-elect of Na
196 Comment on Working fire in Chicago, 2-14-18 by Jason Maybe its good craftsmanship why the
197 Illinois State secures most wins in five seasons after sliding past the Sycamores of Indian
198 FPT5-UKYE: Press Release Chicago Police Department Wic... FPT5UKYE: Press Release Chicag
199 Style Fashion Week New York: Dertbag, Shay Kawaii, Chicago Playground & H33M Showcase FW18
200 Chicago Town have launched mac and cheese pizzas for just č1 but youll have to act fast Puz
201 FREEMASONS ARE GOING MIDWEST TO ILLINOIS ADDITIONAL 33 & BEYOND: THE ROYAL ART OF FREEMASO
202 The Latest: Slain Chicago officer remembered as 'model cop' The Latest: Slain Chicago offic
203 BRIEF-Owens-Illinois Reports Q4 Adjusted Earnings Per Share $0.55 Feb 6 Reuters OwensIllin
204 P.F. Changs offering Pyeongchang Lettuce Wraps following on-air flub by Chicago news statio
205 Chicago lawyer to co-lead case against Equifax Subscribe Chicago lawyer to colead case agai
206 S. Illinois fends off Indiana St. rally for 82-77 win S. Illinois fends off Indiana St. ral
207 Mossy Oakő Congratulates Chicago Team for Win at Bassmaster High School Series on Toledo Be

208 Bruce Rauner Brings Educational Freedom to Illinois Families January 30, 2018 4:00 AM A tax
209 (USA-IL-Chicago) IC - Insurance Loss Control Consultant Exl Service IC Insurance Loss Cont
210 Oboi Reed Blasts the Citys Failure to Address Its Biased Bike Ticketing Problem Streetsblo
211 Credit Repair Marissa, Illinois   Marissa, Illinois   Marissa, Illinois call 888 5021260 f
212 The Latest: Slain Chicago officer remembered as 'model cop' CHICAGO The Latest on the fune
213 Penn State at Illinois Preview: The Last Winnable Game Approaches Vegas Line: PSU 1 While t
214 (USA-IL-Chicago) Analytics + Information Management, Infor Cloud Financials Specialist Seni
215 Harden Furniture shuts down after sale to Illinois company   11:41 AM EST Harden Furniture
216 Lets Play Two Pearl Jam & Chicago Cubs Documentary Mike Pease Uncategorized Leave a comme
217 Winter deals at Echelon Chicago apartments, across from Jewel-Osco sponsored post 12218 Ecl
218 Coach Jerry Kill returns to Southern Illinois as ambassador Coach Jerry Kill returns to Sou
219 Chicago Equity Partners LLC Has $12.67 Million Stake in Morgan Stanley (NYSE:MS) Posted by
220 Watch Cleveland Monsters - Chicago Wolves Live Link2 Cleveland Monsters Chicago Wolves liv
221 "Imagine 2050 Angel of Hate Infiltrates Illinois Churches" is ILLEGAL SLANDER/DEFAMATION T
222 Instructional Designer - Chubb - Chicago, IL Print Job Job Description Combined Insurance,
223 NBCs Chicago Med Chicago Casting Call for Therapists NBCs Chicago Med is now casting for th
224 Cacklers, sheds, ice fishing, Botany of Desire, birding: WWW Chicago outdoors Chicago Sun-
225 Illinois nursing home cited in strangulation death A nursing home resident who was strangle
226 Chicago Auto Show attendees help the homeless Wednesday, February 14, 2018 06:04PM CHICAGO
227 York Township Honors United States Army Veteran - Chicago Tribune Richard Agemura is a comb
228 College basketball: No. 17 Ohio State holds off pesky Illinois to win 75-67 Save COLUMBUS,
229 Colts LB Edwin Jackson among 2 killed by suspected drunken driver - Chicago Tribune Indiana
230 Search for Buried Treasure Linked to Illinois Man's Death At Yellowstone By Laurel Wamsley
231 Credit Repair Byron, Illinois   Byron, Illinois   Byron, Illinois   Byron, Illinois,   are
232 Alzheimers research questioned as another drug fails - Chicago As more experimental drugs
233 New video by YoChicago: A 2-bedroom, 2-bath model in a keystone location at Echelon Chicago
234 Seattle hands Chicago State its 20th straight loss, 91-81 Seattle hands Chicago State its 2
235 US-Spec VW Arteon Lands In Chicago, Sales To Begin In Late 2018 USSpec VW Arteon Lands In C
236 Monochrome Minimalist Fashion in Harajuku w/ Dolls Kill, Kinji, Chicago & Adidas Monochrome
237 Men's Winter Coat (Chicago) $20 condition: like new QR Code Link to This Post Mens Wolverin
238 Do the Chicago Cubs feature a future Hall of Famer? Tweet By Jim Miloch , today at 9:07 am
239 (USA-IL-Chicago) Managing Director, Corporate Finance KPMG Managing Director, Corporate Fin
240 Sources: Several Chicago cops stripped of powers amid criminal probe Several members of a C
241 Considerations When Booking Boats For Sailing Chicago By Frank Robinson Depending on whethe
242 Serial stowaway nabbed at Chicago airport days after release Email CHICAGO Authorities say
243 Local Gymnastics Coach Faces Assault Charges ń CBS Chicago Sponsored By Local Gymnastics Co
244 Illinois hunters take more than 147,500 deer this season Illinois hunters take more than 14
245 Illinois lottery official quits after using Trump-like slur Illinois lottery official quits
246 Michael Madigan still reigns over Illinois You can interview the Democratic candidates for
247 Dishing Deep in Chicago on Salesforce Data and Metadata Backup and Restore Home Cybersecu
248 Rauner challenger Ives goes on attack during only forum of Republican governor race - Chica
249 Will assignment on third line be the charm for Blackhawks Brandon Saad? - Chicago Looks lik
250 1993 Chicago White Sox VS Toronto Blue Jays ALCS Game 1 Ticket Stub Image not available Pho
251 Chicago Manual of Style 17th Edition output style Hello, The new Chicago Manual of Style 17
252 Redesigned 2018 Hyundai Sonata Hybrid and PHEV revealed in Chicago Filed under: Chicago Aut
253 The Latest: Illinois GOP leader wants ad pulled, apology Save FILE In this Jan. 29, 2018
254 Some Illinois hospitals warn Medicaid change could mean cuts By Sarah Zimmerman Associated
255 BC-BKCS. Illinois-Missouri St. Box BCBKCS. IllinoisMissouri St. Box 17 pm 01272018 08:17pm

256 Nylander nets Leafs OT winner on penalty shot in Chicago CHICAGOAs the head coach of a team
257 Stocks in Review: Chicago Bridge & Iron Company N.V. (CBI) Privacy Policy Stocks in Review
258 #OpenHouse tomorrow from 12-2pm 4903 West Roberts Ridge Road #Monee Illinois https://t.co/
259 As Chicago Cubs unveil Yu Darvish, a winter of discontent bleeds into a spring of hope and
260 Illinois sees flood warnings as snow melts, heavy rains fall - KFVS12 News & Weather Cape
261 Chicago State sets D-I womens record with 59th loss in row Chicago State sets DI womens rec
262 Chicago Bears to play Ravens in Hall of Fame game The Chicago Bears and Baltimore Ravens w
263 Field Museum scientists in Chicago studying Michigan meteor Field Museum scientists in Chi
264 How to end gerrymandering in Illinois After months of bargaining, the Ohio legislature pass
265 (USA-IL-Chicago) Property Accounting Manager Home View All Jobs 2,156,577 Robert Half Fi
266 Meet the 3 Chicago-area dogs competing in the Puppy Bowl Miss Brie was adopted by a Palati
267 Woman found killed in West Side apartment Chicago Tribune Woman found killed in West Side
268 Illinois To Transform Assessment By Shane Vander Hart 2 Comments Add Illinois to the list
269 What is Loyola basketball's 'Wall of Culture'? It's a guide to team&ap
270 Chicago Cubs co-owner Todd Ricketts likely to be next Republican committee finance chair C
271 Offensive line decisions big dominoes at start of Bears' offseason - Chicago Tribune
272 Southern Illinois alone in 2nd in MVC after win over Bradley Southern Illinois alone in 2n
273 (USA-IL-Chicago) ARA Data Sciences C&IP Retail Senior Manager ARA Data Sciences CIP Retai
274 4 Chicago Cubs vs. St. Louis Cardinals Tickets 4/16 213   2 viewed per hour 4 Chicago Cubs
275 Chicago region digs out from snowstorm and it's not over Chicago region digs out from sno
276 Illinois Deputy Pushes Fellow Officers, Civilian Out of Way of Speeding Car | NT News View
277 Edison's Fowler Farm project wins Illinois Stewardship Alliance award Topix Illinois Edi
278 How the Illinois 16ths Congressional candidates felt about Pres. Trumps State of the Union
279 Benefit will help central Illinois 2-year-old with rare syndrome Benefit will help central
280 (9900748) Apparently not everyone hates Illinois Nazis ltnor : With the amount of old bigo
281 Kennedy Buys Second TV Ad in Race for Illinois Governor - NBC Chicago Receive the latest p
282 Max Strus stayed close to home to give DePaul a long-range shooting threat - Chicago Thing
283 Illinois Schools Say Pension Shift Would Be Financial Burden Illinois News Illinois School
284 Ameren Illinois : Decatur Civic Center Arena to be named for Ameren Illinois Feb. 01 DECA
285 Preparing For A New Paint Job From An Auto Body Repair Shop In Chicago By Donna Beley One
286 Legionnella testing underway at Illinois capitol Legionnella testing underway at Illinois
287 WOW! 1 Gig Internet Speeds Now Available in Chicago and Naperville, Ill. January 23, 2018 a
288 SystemDomain Chicago based IT firm Made to Finalists in Fastest Understanding and evaluati
289 People from Rhode Island, Illinois and Connecticut are ashamed to live there | Breaking UK
290 CHICAGO AUTO SHOW PARTNERS WITH COMCAST FOR THIRD CONSECUTIVE YEAR 0 Show Attendees to get
291 Hate the Grammys? The Oscars? So, it seems, does everybody - Chicago Tribune These are the
292 White Sox primed to change perceptions - Chicago Tribune The shocking news spread quickly
293 Prosecutors say Chicago gets 1st cryptocurrency case Bitcoin, Cryptocurrency and Fintech
294 Police: Chicago man cited for punching a co-worker in the face Police: Chicago man cited f
295 Catcher Kevan Smith prepared to battle for role with White Sox (Chris Kuc/Chicago Tribune)
296 Bearded Villain Chicago Respect Than Bearded Villain Chicago Respect Thank you to my BVC b
297 VITU Ushers in a New Era of Tax, Title and Registration Processing in Illinois Worldwide O
298 Illinois GOP Rep. Under Fire For Ad Mocking Transgender Community, Feminists Next Post Next
299 Bulls lose 4th straight, fall 110-96 to Bucks - Chicago Tribune For three games, Fred Hoib
300 Credit Repair Dieterich, Illinois Dieterich, Illinois Dieterich, Illinois Dieterich,
301 Illinois schools say pension shift would undo funding revamp Illinois schools say pension s
302 Woman killed in domestic shooting among 8 shot in Chicago - Chicago Tribune A 56yearold wo
303 Oakton helps to spotlight Muslim-American culture by joining 'Coming Together' -

```
304 To Buy Affordable Lawn Equipment Chicago Deserves To Be Prioritized By Robert Long In order
305 Bradley downs Illinois St 70-58 in I-74 rivalry game Bradley downs Illinois St 7058 in I74
306 Federal court in Chicago considers potentially landmark LGBT housing case Chicago News 0206
307 Name: Courtney Davis, Michael Rohd, Nina Rohd, and LolaLocation: Evanston, IllinoisSize: 1
308 Southern Illinois Soybean Wheat Meeting Did you head back to Ohio todayThere is a AgriMaxx
309 Fayne ties career-high with 24 pionts, Illinois State beats Valpo 76-65 Fayne ties careerhi
310 LOG CABIN- Redwood, Live Edge, Mid Century Mountain Home, 6 piece set (Olde Chicago Antique
311 (USA-IL-CHICAGO) Copywriter Home View All Jobs  2,158,749  Robert Half The Creative Group
312 A Fond Farewell to Crate & Barrels Chicago Flagship Store After 27 years, Crate  Barrel has
313 (USA-IL-Chicago) Scala Architect Home View All Jobs  2,195,278  CyberCoders Scala Architect
314 Chicago is bracing for its biggest snow since 2015 Chicago is bracing for its biggest snow
315 Wiretap records Illinois governor hopeful's comments on race Wiretap records Illinois gove
316 Four Chicago teens charged with trying to break into Naperville house, police say - Napervi
317 Share Buyback Plan Initiated by Owens-Illinois (OI) Board Tweet OwensIllinois NYSE:OI annou
318 This Is The Best Illinois City To Live In | Patch real estate Shared from the Naperville,
319 Did You Catch Chicago Wests Adorable Cameo in Kylie Jenners Birth Announcement? Did You Cat
320 Fast Hitch from Monroe Center, Illinois to Birmingham, Alabama I have a guy trying to get a
321 Illinois sets up new hotline for victims of sexual harassment Illinois sets up new hotline
322 Las Vegas man arrested in Chicago in death of infant U.S. Marshals have arrested a man in
323 Victim: 'There was no way out' from clutches of 'violent pimp' from Naperville | Chicago Su
```

In [50]: train.head(30)

Out[50]: 0   Man dies hours after being shot leaving Southwest Side party – Chicago Tribune
         1   The absurdity of Poland's Holocaust law - Chicago The Streisand effect is
         2   My View: 2011 worker's comp changes beginning to benefit Illinois My View: 201
         3   Instant analysis: Michigan State 87, Illinois 74 | MLive.com Registration on o
         4   Illinois governor candidate stands by ad blasted as racist CHICAGO AP A Repub
         5   2012 Chicago White Sox Opening Day Ticket Vs Tigers Image not available Photos
         6   BC-BKC--Missouri St.-Loyola of Chicago Box BCBKCMissouri St.Loyola of Chicago
         7   51st Annual Elmhurst College Jazz Festival Begins Feb. 22 - Chicago Tribune Th
         8   Therapist / Respiratory Therapy Job in Chicago, Illinois / Hospital | USA Jobs
         9   6 shot within 3 hours to start President's Day weekend in Chicago | 1 killed,
         10   Toys R Us closing seven Chicago area stores Toys R Us closing seven Chicago are
         11   Loyola-Chicago beats Drake 80-57 behind big 2nd-half run The Associated Press .
         12   17 killed in mass shooting at high school in Parkland, Florida | Gary/Chicago
         13   Chicago ranks second worst of most populous U.S. cities for fiscal health Chica
         14   Holocaust denier running for Congress has no opponents in Illinois Republican p
         15   Illinois mother pleads guilty in babys suffocation death ALORTON, Ill. AP A so
         16   Hoy 22.00 NBA Orlando Magic Vs Chicago Bulls EN VIVO   Esperar 10 Seg Publicida
         17   Chicago Restaurant Becomes Go-To for Eagles Fans unable to retrieve fulltext co
         18   Black NHL player Devante Smith-Pelly taunted by Chicago fans Black NHL player
         19   Snow blankets Chicago as winter storm sweeps U.S. Midwest   12:40 PM EST Snow
         20   Group aims to bridge gap between NU students, Chicago musicians Send Email Cano
         21   Loyola holds off Valparaiso for 22nd win, 11th in 12 games - Chicago Tribune Do
         22   Duerson Act would ban pre-teen tackle football in Illinois Duerson Act would ba
         23   Oak Park Police Investigate Alleged Abduction ń CBS Chicago Sponsored By Oak Pa

13

```
24      College notes: Evanston&apos;s Will Jones paces Monmouth to MWC hoops title -
25      Illinois lawmakers inch closer to new hospital tax despite worries Illinois la
26      Carrollton, Illinois - Insurance Agent - Carrollton, IL, , USA #jobs #Carrollt
27      What You Should Know About Sales Jobs In Chicago By Jeffrey Lee Thousands of p
28      The Art Institute of Chicago Share Visit the worlds best museum, the Art Insti
29      Volvo and The Petfinder Foundation partner at 2018 Chicago Auto Show - Roadsho
                                    ...
270     Chicago Cubs co-owner Todd Ricketts likely to be next Republican committee fin
271     Offensive line decisions big dominoes at start of Bears&apos; offseason - Chic
272     Southern Illinois alone in 2nd in MVC after win over Bradley Southern Illinois
273     (USA-IL-Chicago) ARA Data Sciences C&IP Retail Senior Manager   ARA Data Scien
274     4 Chicago Cubs vs. St. Louis Cardinals Tickets 4/16 213   2 viewed per hour 4
275     Chicago region digs out from snowstorm  and it's not over Chicago region digs
276     Illinois Deputy Pushes Fellow Officers, Civilian Out of Way of Speeding Car |
277     Edison's Fowler Farm project wins Illinois Stewardship Alliance award Topix  I
278     How the Illinois 16ths Congressional candidates felt about Pres. Trumps State
279     Benefit will help central Illinois 2-year-old with rare syndrome Benefit will
280     (9900748) Apparently not everyone hates Illinois Nazis ltnor : With the amount
281     Kennedy Buys Second TV Ad in Race for Illinois Governor - NBC Chicago Receive
282     Max Strus stayed close to home to give DePaul a long-range shooting threat - C
283     Illinois Schools Say Pension Shift Would Be Financial Burden Illinois News Ill
284     Ameren Illinois : Decatur Civic Center Arena to be named for Ameren Illinois F
285     Preparing For A New Paint Job From An Auto Body Repair Shop In Chicago By Donn
286     Legionnella testing underway at Illinois capitol Legionnella testing underway
287     WOW! 1 Gig Internet Speeds Now Available in Chicago and Naperville, Ill. Janua
288     SystemDomain Chicago based IT firm Made to Finalists in Fastest Understanding
289     People from Rhode Island, Illinois and Connecticut are ashamed to live there |
290     CHICAGO AUTO SHOW PARTNERS WITH COMCAST FOR THIRD CONSECUTIVE YEAR 0 Show Atte
291     Hate the Grammys? The Oscars? So, it seems, does everybody - Chicago Tribune T
292     White Sox primed to change perceptions - Chicago Tribune The shocking news spr
293     Prosecutors say Chicago gets 1st cryptocurrency case  Bitcoin, Cryptocurrency
294     Police: Chicago man cited for punching a co-worker in the face Police: Chicago
295     Catcher Kevan Smith prepared to battle for role with White Sox (Chris Kuc/Chic
296     Bearded Villain Chicago Respect Than Bearded Villain Chicago Respect Thank you
297     VITU Ushers in a New Era of Tax, Title and Registration Processing in Illinois
298     Illinois GOP Rep. Under Fire For Ad Mocking Transgender Community, Feminists N
299     Bulls lose 4th straight, fall 110-96 to Bucks - Chicago Tribune For three game
Name: text_combi, Length: 300, dtype: object

In [262]: y_train=[1,0,1,0,1,0,0,1,0,1,1,0,0,1,1,0,0,0,1,1,1,0,1,1,0,1,0,1,1,0,0,1,0,1,0,0,0,0

In [263]: training = pd.DataFrame({'text': train, 'keep': y_train})

In [264]: #looking at my combined data
          training.head(10)

Out[264]:    keep  \
          0     1
          1     0
```

```
2      1
3      0
4      1
5      0
6      0
7      1
8      0
9      1


0   Man dies hours after being shot leaving Southwest Side party - Chicago Tribune
1   The absurdity of Poland&apos;s Holocaust law - Chicago The Streisand effect is one
2   My View: 2011 worker's comp changes beginning to benefit Illinois My View: 2011 wc
3   Instant analysis: Michigan State 87, Illinois 74 | MLive.com Registration on or us
4   Illinois governor candidate stands by ad blasted as racist CHICAGO AP  A Republica
5   2012 Chicago White Sox Opening Day Ticket Vs Tigers Image not available Photos not
6   BC-BKC--Missouri St.-Loyola of Chicago Box BCBKCMissouri St.Loyola of Chicago Box
7   51st Annual Elmhurst College Jazz Festival Begins Feb. 22 - Chicago Tribune The El
8   Therapist / Respiratory Therapy Job in Chicago, Illinois / Hospital | USA Jobs Clu
9   6 shot within 3 hours to start President's Day weekend in Chicago | 1 killed, 12 v
```

In [265]: `#examining the shape of my dataframe`
`training.shape`

Out[265]: `(324, 2)`

In [266]: `#examining the class distribution`
`training.keep.value_counts().head(10)`

Out[266]:
```
0     252
1      72
Name: keep, dtype: int64
```

In [267]: `# how to define X and y (from the citation data) for use with COUNTVECTORIZER`
```
X = training.text
y = training.keep
print(X.shape)
print(y.shape)
```

```
(324,)
(324,)
```

In [268]: `# split X and y into training and testing sets`
```
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1)
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)# instantiate the vectorizer
vect = CountVectorizer()
```

```
(243,)
(81,)
(243,)
(81,)
```

0.1 Vectorizing dataset

```
In [269]: # instantiate the vectorizer
          vect = CountVectorizer()
```

```
In [270]: # learn training data vocabulary, then use it to create a document-term matrix
          vect.fit(X_train)
          X_train_dtm = vect.transform(X_train)
```

```
In [271]: # examine the document-term matrix
          X_train_dtm
```

```
Out[271]: <243x11724 sparse matrix of type '<class 'numpy.int64'>'
                  with 41327 stored elements in Compressed Sparse Row format>
```

```
In [272]: # transform testing data (using fitted vocabulary) into a document-term matrix
          X_test_dtm = vect.transform(X_test)
          X_test_dtm
```

```
Out[272]: <81x11724 sparse matrix of type '<class 'numpy.int64'>'
                  with 13650 stored elements in Compressed Sparse Row format>
```

0.2 Building and evaluating models

0.2.1 Naive Bayes Model

```
In [273]: # instantiate a Multinomial Naive Bayes model
          nb = MultinomialNB()
```

```
In [274]: # train and time the model using X_train_dtm
          %time nb.fit(X_train_dtm, y_train)
```

```
Wall time: 10.2 ms
```

```
Out[274]: MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)
```

```
In [275]: # make class predictions for X_test_dtm
          y_pred_class = nb.predict(X_test_dtm)
```

```
In [276]: # calculate accuracy of class predictions
          print(metrics.accuracy_score(y_test, y_pred_class))
```

```
0.8024691358024691
```

```
In [277]: # calculate precision and recall
          print(classification_report(y_test, y_pred_class))

            precision    recall  f1-score   support

         0       0.84      0.92      0.88        62
         1       0.62      0.42      0.50        19

avg / total       0.79      0.80      0.79        81
```

```
In [278]: # calculate the confusion matrix
          print(metrics.confusion_matrix(y_test, y_pred_class))

[[57  5]
 [11  8]]
```

0.2.2   Logisitc Regression Model

```
In [279]: # instantiate a logistic regression model
          logreg = LogisticRegression()
```

```
In [280]: # train the model using X_train_dtm
          %time logreg.fit(X_train_dtm, y_train)

Wall time: 36.2 ms
```

```
Out[280]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                    intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
                    penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
                    verbose=0, warm_start=False)
```

```
In [281]: # make class predictions for X_test_dtm
          y_pred_class = logreg.predict(X_test_dtm)
```

```
In [282]: # calculate accuracy of class predictions
          print(metrics.accuracy_score(y_test, y_pred_class))

0.7407407407407407
```

```
In [283]: # calculate precision and recall
          print(classification_report(y_test, y_pred_class))

            precision    recall  f1-score   support

         0       0.77      0.95      0.85        62
```

```
        1        0.25      0.05      0.09       19

avg / total       0.65      0.74      0.67       81
```

In [284]: `# calculate the confusion matrix`
          `print(metrics.confusion_matrix(y_test, y_pred_class))`

```
[[59  3]
 [18  1]]
```

0.3 Applying model to my remaining test data

From the above, it seems the naive bayes model has greater prediction accuracy, so I will use that on my remaining test set to determine which articles to keep for the remainder of my analysis.

0.3.1 vectoring my test data

In [285]: `# instantiate the vectorizer`
          `#vect = CountVectorizer()`

In [286]: `# transform testing data (using fitted vocabulary) into a document-term matrix`
          `X_test_dtm_2 = vect.transform(test)`
          `X_test_dtm_2`

Out[286]: `<40261x11724 sparse matrix of type '<class 'numpy.int64'>'`
          `        with 6559665 stored elements in Compressed Sparse Row format>`

In [287]: `# make class predictions for test`
          `y_pred_class = nb.predict(X_test_dtm_2)`

In [288]: `y_pred_class`

Out[288]: `array([0, 0, 0, ..., 1, 1, 0], dtype=int64)`

In [289]: `#now put all of the pieces back together`
          `test_2=pd.DataFrame({'text': test, 'keep': y_pred_class})`

In [290]: `test_2.head(10)`

Out[290]:   keep \
          38254   0
          19538   0
          29390   0
          1646   0
          35705   0
          35123   0
          36319   0

```
         30815      0
         24090      0
         23354      0


         38254  NBA commissioner Adam Silver lauds refs, players for airing differences - Chi
         19538  Statement By CUB Executive Director David Kolata Alerting Ameren Illinois Cust
         29390  This is Bill, a magnificent all natural mega bodybuilder in River North, Chica
         1646   Essentials When Looking For Custom Luxury Home Builders Illinois By Brian Ell:
         35705  A Chicago Winter Wedding Pt 1 This past weekend my family of four and I trave]
         35123  Steppenwolf announces spring 2018 LookOut lineup - Chicago Tribune Category :
         36319                      Hudson, Chicago, IL - Legal/Technical Writer Here are deta
         30815  BC-BKCIll.-Chicago-Green Bay Box Home  NCAA Basketball  BCBKCIll.ChicagoGreen
         24090  MLB players union claims owners threatening &apos;integrity&apos; of game - Cl
         23354  Forensic Accounting Chicago If want accurate and fair result in forensic accou
```

In [291]: `test_2.count()`

Out[291]:
```
keep    40261
text    40261
dtype: int64
```

In [292]: *#putting my training and test data back together for my final dataset*
```
text_final=training.append(test_2)
```

In [293]: `text_final.count()`

Out[293]:
```
keep    40585
text    40585
dtype: int64
```

In [294]: *# removing the "0" articles*
```
text_final_1 = text_final[text_final.keep==1]
```

In [295]: `text_final_1.count()`

Out[295]:
```
keep    7446
text    7446
dtype: int64
```

0.3.2 Topic Modeling

In [296]:
```python
def tf(word, blob):
    return blob.words.count(word) / len(blob.words)
    # tf(word, blob) computes "term frequency" which is the number of times a word appea
    # normalized by dividing by the total number of words in blob. We use TextBlob for b
    # and getting the word counts.
```

19

```python
def n_containing(word, bloblist):
    return sum(1 for blob in bloblist if word in blob.words)
# n_containing(word, bloblist) returns the number of documents containing word.
# A generator expression is passed to the sum() function.


def idf(word, bloblist):
    return math.log(len(bloblist) / (1 + n_containing(word, bloblist)))
# idf(word, bloblist) computes "inverse document frequency" which measures how commo
# among all documents in bloblist. The more common a word is, the lower its idf.
# We take the ratio of the total number of documents to the number of documents cont
# then take the log of that. Add 1 to the divisor to prevent division by zero


def tfidf(word, blob, bloblist):
    return tf(word, blob) * idf(word, bloblist)
# tfidf(word, blob, bloblist) computes the TF-IDF score. It is simply the product of
```

In [297]:
```python
bloblist = []
del bloblist[:]

for i  in range(0,len(text_final_1)):
    bloblist.append(TextBlob(text_final_1['text'].iloc[i]))

len(bloblist)
```

Out[297]: 7446

In [298]:
```python
for i, blob in enumerate(bloblist):
    # Print top 5 values
        if i == 5:
            break
        print("Top words in news article {}".format(i + 1))
        scores = {word: tfidf(word, blob, bloblist) for word in blob.words}
        sorted_words = sorted(scores.items(), key=lambda x: x[1], reverse=True)
        for word, score in sorted_words[:10]:
            print("\tWord: {}, TF-IDF: {}".format(word, round(score, 5)))
```

```
Top words in news article 1
        Word: dies, TF-IDF: 0.16107
        Word: shot, TF-IDF: 0.13911
        Word: TribuneMan, TF-IDF: 0.12847
        Word: partyChicago, TF-IDF: 0.12847
        Word: 6300, TF-IDF: 0.12847
        Word: hours, TF-IDF: 0.12632
        Word: MenaCastro, TF-IDF: 0.12214
        Word: Northlake, TF-IDF: 0.12214
        Word: 6:16, TF-IDF: 0.12214
        Word: Fernando, TF-IDF: 0.11416
```

```
Top words in news article 2
        Word: Costs, TF-IDF: 0.09808
        Word: Insurance, TF-IDF: 0.08654
        Word: workers, TF-IDF: 0.0706
        Word: insurance, TF-IDF: 0.06014
        Word: compensation, TF-IDF: 0.05434
        Word: employers, TF-IDF: 0.0491
        Word: companies, TF-IDF: 0.04848
        Word: In, TF-IDF: 0.04718
        Word: costs, TF-IDF: 0.04291
        Word: NASI, TF-IDF: 0.04266
Top words in news article 3
        Word: blasted, TF-IDF: 0.2431
        Word: racist, TF-IDF: 0.20561
        Word: 2018.02.03, TF-IDF: 0.20054
        Word: 19:43, TF-IDF: 0.20054
        Word: dailymail.co.uk, TF-IDF: 0.17819
        Word: ad, TF-IDF: 0.16272
        Word: homophobic, TF-IDF: 0.14206
        Word: By, TF-IDF: 0.12653
        Word: stands, TF-IDF: 0.10825
        Word: unseat, TF-IDF: 0.10537
Top words in news article 4
        Word: Jazz, TF-IDF: 0.12729
        Word: jazz, TF-IDF: 0.12729
        Word: college, TF-IDF: 0.12649
        Word: Elmhurst, TF-IDF: 0.11056
        Word: College, TF-IDF: 0.10523
        Word: Festival, TF-IDF: 0.09007
        Word: Jiggs, TF-IDF: 0.04814
        Word: Whigham, TF-IDF: 0.04814
        Word: performances, TF-IDF: 0.04529
        Word: Begins, TF-IDF: 0.04472
Top words in news article 5
        Word: man, TF-IDF: 0.0464
        Word: condition, TF-IDF: 0.04351
        Word: shot, TF-IDF: 0.03885
        Word: Police, TF-IDF: 0.0323
        Word: A, TF-IDF: 0.03229
        Word: Stroger, TF-IDF: 0.0312
        Word: Hours, TF-IDF: 0.03096
        Word: block, TF-IDF: 0.03089
        Word: wounded, TF-IDF: 0.03047
        Word: Hospital, TF-IDF: 0.02992
```

Applying LDA to these news articles

```
In [299]: news_list = text_final_1['text'].tolist()
          news_list[:1]

Out[299]: ['Man dies hours after being shot leaving Southwest Side party - Chicago Tribune    T

In [300]: stop = set(stopwords.words('english'))
          exclude = set(string.punctuation)
          lemma = WordNetLemmatizer()
          def clean(doc):
              stop_free = " ".join([i for i in doc.lower().split() if i not in stop])
              punc_free = ''.join(ch for ch in stop_free if ch not in exclude)
              normalized = " ".join(lemma.lemmatize(word) for word in punc_free.split())
              return normalized

          doc_clean = [clean(doc).split() for doc in news_list]

In [301]: # Creating the term dictionary of our courpus, where every unique term is assigned a
          dictionary = corpora.Dictionary(doc_clean)

          # Converting list of documents (corpus) into Document Term Matrix using dictionary p
          doc_term_matrix = [dictionary.doc2bow(doc) for doc in doc_clean]

In [302]: # Creating the object for LDA model using gensim library
          Lda = gensim.models.ldamodel.LdaModel

In [303]: news_clean = [clean(doc).split() for doc in news_list]

In [304]: len(news_clean)

Out[304]: 7446

In [305]: print(*news_clean[:1], sep='\n\n')

['man', 'dy', 'hour', 'shot', 'leaving', 'southwest', 'side', 'party', 'chicago', 'tribune', 't

In [306]: # Creating the term dictionary of our courpus, where every unique term is assigned a

          dictionary = corpora.Dictionary(news_clean)

          # Converting list of documents (corpus) into Document Term Matrix using dictionary p

          %time doc_term_matrix = [dictionary.doc2bow(doc) for doc in news_clean]

Wall time: 2.68 s
```

3 topic model

```
In [307]: # Creating the object for LDA model using gensim library
          Lda = gensim.models.ldamodel.LdaModel
          numtopics = 3

          # Running and Training LDA model on the document term matrix.
          %time ldamodel = Lda(doc_term_matrix, num_topics=numtopics, id2word = dictionary, pas
```

Wall time: 24min 39s

```
In [308]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=3), sep='\n\n')
```

(0, '0.012*"chicago" + 0.008*"said" + 0.007*"illinois"')

(1, '0.016*"illinois" + 0.013*"state" + 0.010*"school"')

(2, '0.025*"chicago" + 0.011*"said" + 0.011*"police"')

```
In [309]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=5), sep='\n\n')
```

(0, '0.012*"chicago" + 0.008*"said" + 0.007*"illinois" + 0.005*"law" + 0.004*"court"')

(1, '0.016*"illinois" + 0.013*"state" + 0.010*"school" + 0.009*"said" + 0.008*"chicago"')

(2, '0.025*"chicago" + 0.011*"said" + 0.011*"police" + 0.010*"snow" + 0.006*"shooting"')

```
In [310]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=7), sep='\n\n')
```

(0, '0.012*"chicago" + 0.008*"said" + 0.007*"illinois" + 0.005*"law" + 0.004*"court" + 0.004*"

(1, '0.016*"illinois" + 0.013*"state" + 0.010*"school" + 0.009*"said" + 0.008*"chicago" + 0.00

(2, '0.025*"chicago" + 0.011*"said" + 0.011*"police" + 0.010*"snow" + 0.006*"shooting" + 0.005*

```
In [311]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=10), sep='\n\n')
```

(0, '0.012*"chicago" + 0.008*"said" + 0.007*"illinois" + 0.005*"law" + 0.004*"court" + 0.004*"

(1, '0.016*"illinois" + 0.013*"state" + 0.010*"school" + 0.009*"said" + 0.008*"chicago" + 0.006

(2, '0.025*"chicago" + 0.011*"said" + 0.011*"police" + 0.010*"snow" + 0.006*"shooting" + 0.005*

0.3.3 5 topic model

```
In [312]: # Creating the object for LDA model using gensim library
          Lda = gensim.models.ldamodel.LdaModel
          numtopics = 5

          # Running and Training LDA model on the document term matrix.
          %time ldamodel = Lda(doc_term_matrix, num_topics=numtopics, id2word = dictionary, pas
```

Wall time: 17min 38s

```
In [313]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=3), sep='\n\n')
```

(0, '0.015*"illinois" + 0.012*"state" + 0.011*"said"')

(1, '0.015*"illinois" + 0.012*"state" + 0.008*"chicago"')

(2, '0.013*"chicago" + 0.010*"said" + 0.008*"police"')

(3, '0.034*"chicago" + 0.015*"police" + 0.011*"said"')

(4, '0.021*"chicago" + 0.018*"snow" + 0.014*"school"')

```
In [314]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=5), sep='\n\n')
```

(0, '0.015*"illinois" + 0.012*"state" + 0.011*"said" + 0.009*"pritzker" + 0.009*"democratic"')

(1, '0.015*"illinois" + 0.012*"state" + 0.008*"chicago" + 0.007*"said" + 0.007*"school"')

(2, '0.013*"chicago" + 0.010*"said" + 0.008*"police" + 0.008*"court" + 0.007*"law"')

(3, '0.034*"chicago" + 0.015*"police" + 0.011*"said" + 0.009*"man" + 0.008*"shooting"')

(4, '0.021*"chicago" + 0.018*"snow" + 0.014*"school" + 0.010*"said" + 0.009*"football"')

```
In [315]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=7), sep='\n\n')
```

(0, '0.015*"illinois" + 0.012*"state" + 0.011*"said" + 0.009*"pritzker" + 0.009*"democratic" +

(1, '0.015*"illinois" + 0.012*"state" + 0.008*"chicago" + 0.007*"said" + 0.007*"school" + 0.007

(2, '0.013*"chicago" + 0.010*"said" + 0.008*"police" + 0.008*"court" + 0.007*"law" + 0.006*"of:

(3, '0.034*"chicago" + 0.015*"police" + 0.011*"said" + 0.009*"man" + 0.008*"shooting" + 0.008*"

(4, '0.021*"chicago" + 0.018*"snow" + 0.014*"school" + 0.010*"said" + 0.009*"football" + 0.008*

```
In [316]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=10), sep='\n\n')

(0, '0.015*"illinois" + 0.012*"state" + 0.011*"said" + 0.009*"pritzker" + 0.009*"democratic" +

(1, '0.015*"illinois" + 0.012*"state" + 0.008*"chicago" + 0.007*"said" + 0.007*"school" + 0.007

(2, '0.013*"chicago" + 0.010*"said" + 0.008*"police" + 0.008*"court" + 0.007*"law" + 0.006*"off

(3, '0.034*"chicago" + 0.015*"police" + 0.011*"said" + 0.009*"man" + 0.008*"shooting" + 0.008*"

(4, '0.021*"chicago" + 0.018*"snow" + 0.014*"school" + 0.010*"said" + 0.009*"football" + 0.008*
```

0.3.4   10 topic model

```
In [317]: # Creating the object for LDA model using gensim library
          Lda = gensim.models.ldamodel.LdaModel
          numtopics = 10

          # Running and Trainign LDA model on the document term matrix.
          %time ldamodel = Lda(doc_term_matrix, num_topics=numtopics, id2word = dictionary, pas

Wall time: 15min 38s


In [318]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=3), sep='\n\n')

(0, '0.011*"illinois" + 0.010*"chicago" + 0.009*"law"')

(1, '0.030*"snow" + 0.025*"chicago" + 0.014*"inch"')

(2, '0.068*"school" + 0.033*"student" + 0.021*"chicago"')

(3, '0.018*"pritzker" + 0.015*"said" + 0.011*"blagojevich"')

(4, '0.022*"illinois" + 0.021*"state" + 0.011*"tax"')

(5, '0.023*"chicago" + 0.009*"city" + 0.006*"park"')

(6, '0.024*"football" + 0.015*"illinois" + 0.015*"said"')

(7, '0.029*"chicago" + 0.023*"police" + 0.013*"said"')

(8, '0.016*"bi" + 0.013*"trump" + 0.012*"pritzker"')

(9, '0.016*"illinois" + 0.012*"democratic" + 0.012*"primary"')


In [319]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=5), sep='\n\n')
```

```
(0, '0.011*"illinois" + 0.010*"chicago" + 0.009*"law" + 0.008*"said" + 0.008*"county"')

(1, '0.030*"snow" + 0.025*"chicago" + 0.014*"inch" + 0.013*"weather" + 0.012*"friday"')

(2, '0.068*"school" + 0.033*"student" + 0.021*"chicago" + 0.012*"high" + 0.011*"district"')

(3, '0.018*"pritzker" + 0.015*"said" + 0.011*"blagojevich" + 0.010*"state" + 0.010*"white"')

(4, '0.022*"illinois" + 0.021*"state" + 0.011*"tax" + 0.009*"said" + 0.009*"would"')

(5, '0.023*"chicago" + 0.009*"city" + 0.006*"park" + 0.005*"said" + 0.005*"new"')

(6, '0.024*"football" + 0.015*"illinois" + 0.015*"said" + 0.014*"tackle" + 0.011*"duerson"')

(7, '0.029*"chicago" + 0.023*"police" + 0.013*"said" + 0.011*"officer" + 0.010*"shooting"')

(8, '0.016*"bi" + 0.013*"trump" + 0.012*"pritzker" + 0.012*"campaign" + 0.010*"illinois"')

(9, '0.016*"illinois" + 0.012*"democratic" + 0.012*"primary" + 0.012*"district" + 0.011*"party'
```

In [320]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=7), sep='\n\n')

```
(0, '0.011*"illinois" + 0.010*"chicago" + 0.009*"law" + 0.008*"said" + 0.008*"county" + 0.008*"

(1, '0.030*"snow" + 0.025*"chicago" + 0.014*"inch" + 0.013*"weather" + 0.012*"friday" + 0.010*"

(2, '0.068*"school" + 0.033*"student" + 0.021*"chicago" + 0.012*"high" + 0.011*"district" + 0.0

(3, '0.018*"pritzker" + 0.015*"said" + 0.011*"blagojevich" + 0.010*"state" + 0.010*"white" + 0

(4, '0.022*"illinois" + 0.021*"state" + 0.011*"tax" + 0.009*"said" + 0.009*"would" + 0.008*"yea

(5, '0.023*"chicago" + 0.009*"city" + 0.006*"park" + 0.005*"said" + 0.005*"new" + 0.004*"year"

(6, '0.024*"football" + 0.015*"illinois" + 0.015*"said" + 0.014*"tackle" + 0.011*"duerson" + 0

(7, '0.029*"chicago" + 0.023*"police" + 0.013*"said" + 0.011*"officer" + 0.010*"shooting" + 0.0

(8, '0.016*"bi" + 0.013*"trump" + 0.012*"pritzker" + 0.012*"campaign" + 0.010*"illinois" + 0.01

(9, '0.016*"illinois" + 0.012*"democratic" + 0.012*"primary" + 0.012*"district" + 0.011*"party'
```

In [321]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=10), sep='\n\n')

```
(0, '0.011*"illinois" + 0.010*"chicago" + 0.009*"law" + 0.008*"said" + 0.008*"county" + 0.008*"

(1, '0.030*"snow" + 0.025*"chicago" + 0.014*"inch" + 0.013*"weather" + 0.012*"friday" + 0.010*"
```

```
(2, '0.068*"school" + 0.033*"student" + 0.021*"chicago" + 0.012*"high" + 0.011*"district" + 0.0

(3, '0.018*"pritzker" + 0.015*"said" + 0.011*"blagojevich" + 0.010*"state" + 0.010*"white" + 0

(4, '0.022*"illinois" + 0.021*"state" + 0.011*"tax" + 0.009*"said" + 0.009*"would" + 0.008*"yea

(5, '0.023*"chicago" + 0.009*"city" + 0.006*"park" + 0.005*"said" + 0.005*"new" + 0.004*"year"

(6, '0.024*"football" + 0.015*"illinois" + 0.015*"said" + 0.014*"tackle" + 0.011*"duerson" + 0

(7, '0.029*"chicago" + 0.023*"police" + 0.013*"said" + 0.011*"officer" + 0.010*"shooting" + 0.0

(8, '0.016*"bi" + 0.013*"trump" + 0.012*"pritzker" + 0.012*"campaign" + 0.010*"illinois" + 0.01

(9, '0.016*"illinois" + 0.012*"democratic" + 0.012*"primary" + 0.012*"district" + 0.011*"party"
```

0.3.5 Sentiment analysis

```
In [323]: from urllib.request import urlretrieve

          import urllib
          outpath = 'http://www.unc.edu/~ncaren/haphazard/'
          directory ='C://Users//Joan//OneDrive//06.WI18_nlp//'
          article_directory ='C://Users//Joan//OneDrive//06.WI18_nlp//'
          file_pos = 'positive.txt'
          file_neg = 'negative.txt'

          corpus = [file_pos,file_neg]
          for file in corpus:
              urlretrieve(outpath+file,article_directory+file)

In [324]: article_list = text_final_1['text'].tolist()

In [325]: pos_sent = open(article_directory +file_pos).read()
          pos_words = pos_sent.split('\n')
          neg_sent = open(article_directory +file_neg).read()
          neg_words = neg_sent.split('\n')

In [326]: # customize the dictionaries by adding and removing your own positive and negative w

          pos_add = ['your_pos_term_1, your_pos_term_2']

          for term in pos_add:
              pos_words.append(term)

          neg_add = ['your_neg_term_1, your_neg_term_2']
```

```python
        for term in neg_add:
            neg_words.append(term)

        import re
        from string import punctuation
        from __future__ import division
        sentiment_scores=[]
        for article in article_list:
            sentiment_score=0
            for p in list(punctuation):
                article=article.replace(p,'')
                words=article.split(' ')
            for word in words:
                if word in pos_words:
                    sentiment_score=sentiment_score+1
                if word in neg_words:
                    sentiment_score=sentiment_score-1
            sentiment_scores.append(sentiment_score/len(words))

        article_sentiment=zip(article_list,sentiment_scores)
```

In [327]:
```python
# Create a dataframe from the results
column_names = ["article", "Sentiment_Score"]
sentiment_results = [article_list, sentiment_scores]
results_dict = dict(zip(column_names,sentiment_results))
all_articles_df = pd.DataFrame.from_dict(results_dict, orient='columns')
all_articles_df = all_articles_df[column_names]   # set specific column order
```

In [328]:
```python
# Create a list to store the sentiments
sent_list = []

# For each row in the column,
for row in all_articles_df['Sentiment_Score']:
    if row > 0:
        sent_list.append('Positive')
    elif row < 0:
        sent_list.append('Negative')
    else:
        sent_list.append('Neutral')

# Create a column from the list
all_articles_df['Sentiment_Label'] = sent_list
```

In [329]:
```python
#Make sure I didn't loose any records
len(article_list) - len(all_articles_df)
```

Out[329]: 0

In [331]:
```python
pd.set_option('display.max_colwidth', 150)
```

```
In [332]: all_articles_df.sample(frac=0.005, replace=True)

Out[332]:
    6273   NR Daily: Bruce Rauner Brings Educational Freedom to Illinois Families Somethin
    3348   Illinois proposing law to ban tackle football for ages 12 and under Illinois pi
    1340   Illinois schools say pension shift would be financial burden SPRINGFIELD, Ill.
    523    Weather Service forecasts heavy snow from Chicago to Detroit Weather Service fo
    5373   Ruttenberg Arts Foundation donates to Art Institute of Chicago – Consumer News
    1497   Proposed Legislation Aims to End Predatory Car Title Lending Rates in Illinois
    617    Bill would limit terms of Illinois legislative leaders Topix  Illinois  Bill we
    2628   African-Americans are leaving Chicago for a better quality of life in neighbori
    4997   Illinois Man, Mistaken for Robbery Suspect, Wrongfully Detained in Incident tha
    24     Chicago police to meet with federal, state officials to combat rise in carjacki
    3935   Duerson Act would ban pre-teen tackle football in Illinois Duerson Act would ba
    633    Alzheimers research questioned as another drug fails – Chicago Tribune As more
    6395   TRILLANES LIVE IN CHICAGO NAGPAKALAT NA NAMAN NG FAKE NEWS PARA SIRAAN SI PRESI
    5228   CPD Announces New Task Force Aimed At Curbing Carjackings ñ CBS Chicago Sponsor
    1923   Looking for P-ROC programmer in Chicago area Looking for a PROC programmer who
    2281   &apos;How dare you&apos;: Survivors of Florida school shooting lash out at Trum
    3565   Weather Service forecasts heavy snow from Chicago to Detroit Weather Service fo
    2853   Attack on DuPage couple was part of murder-for-hire plot by son, accomplice, o
    2031   W. Chicago High teachers ready to strike Friday if no agreement West Chicago Hi
    572    Chicago gets 9 consecutive days of snow By : Editor Comment: Off CHICAGO AP  A
    376    Chicago Sun-Times Editorial Board Meets With Marie Newman And Blue Dog Arch-Vil
    4707   College of Lake County laser/photonics/optics program unique in Illinois – Lake
    1130   Illinois man who killed wife's divorce lawyer dies in prison Illinois man who l
    4011   Judge OKs evaluation for Illinois teen accused in shooting CORRECTED: Local fan
    5806   Chicago, Detroit winter storm: Areas to see heavy snow, travel chaos CLOSE Auth
    5487   In wake of Illinois bail bond reform, advocates here push national legislation
    6216   Dramatic video of girl being kidnapped off Illinois street released by FBI Janu
    783                                                                            Red l
    7001   Former student opens fire at Florida high school, killing at least 17 – Chicago
    1328   Brazen child kidnapping on Chicago street Brazen child kidnapping on Chicago st
    7105   Chicago Public Schools Working to Close Achievement Gap Chicago Public Schools
    2191   Munster, Indiana Issues Flood Watch ñ CBS Chicago Sponsored By Munster, Indiana
    6717   &apos;Dreamers,&apos; Bill Nye among guests invited to Trump&apos;s State of th
    508    Chicago Gang or Graffiti Tag Hey, we live in the new building at 640 N Wells in
    4250   Illinois higher ed board asks state for $3.47B in funding Illinois higher ed bc
    1896   Chicago teen charged in carjacking of 84-year-old woman in Evanston – Evanston
    7142   Illinois consistent nationwide for cold-related deaths | WQAD.com Illinois cons

           Sentiment_Score Sentiment_Label
    6273          0.027723        Positive
    3348         -0.044118        Negative
    1340         -0.014085        Negative
    523           0.005731        Positive
    5373          0.018182        Positive
    1497          0.020440        Positive
```

```
617            -0.050633          Negative
2628            0.013550          Positive
4997           -0.030137          Negative
24              0.007519          Positive
3935           -0.008547          Negative
633             0.001708          Positive
6395           -0.013774          Negative
5228            0.055556          Positive
1923            0.039216          Positive
2281           -0.009224          Negative
3565            0.005731          Positive
2853           -0.018450          Negative
2031            0.020033          Positive
572            -0.009804          Negative
376             0.014795          Positive
4707            0.023810          Positive
1130           -0.055556          Negative
4011            0.010471          Positive
5806           -0.012077          Negative
5487            0.098039          Positive
6216           -0.016129          Negative
783             0.000000           Neutral
7001           -0.023511          Negative
1328           -0.013889          Negative
7105            0.037221          Positive
2191            0.000000           Neutral
6717            0.002439          Positive
508             0.000000           Neutral
4250            0.012270          Positive
1896           -0.001575          Negative
7142            0.000000           Neutral
```

In [330]: plt.figure().set_size_inches(10, 5)

CountSentiment = pd.value_counts(all_articles_df['Sentiment_Label'].values, sort=True

print (CountSentiment)

#CountStatus.plot.barh()
CountSentiment.plot.bar()
plt.show()

```
Negative    3367
Positive    2996
Neutral     1083
dtype: int64
```
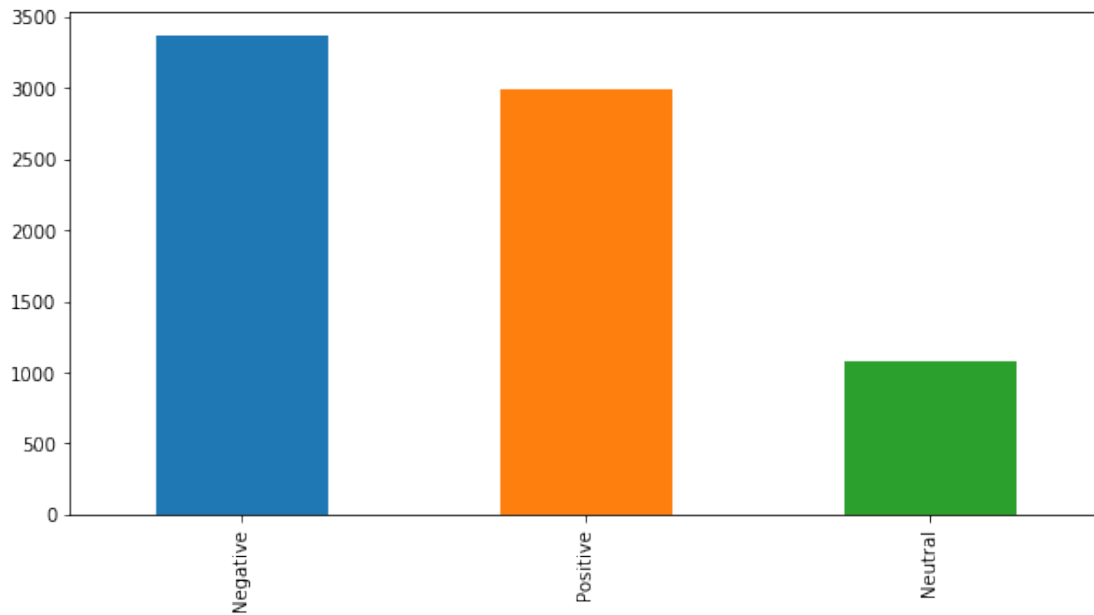
0.3.6 Looking at articles scored as having positive sentiment

```
In [333]: positive = all_articles_df[all_articles_df.Sentiment_Label=="Positive"]

In [348]: positive.count()

Out[348]: article            2996
          Sentiment_Score    2996
          Sentiment_Label    2996
          dtype: int64

In [337]: positive.head(10)

Out[337]:
          1   My View: 2011 worker's comp changes beginning to benefit Illinois My View: 2011 w
          3   51st Annual Elmhurst College Jazz Festival Begins Feb. 22 - Chicago Tribune The I
          10  Group aims to bridge gap between NU students, Chicago musicians Send Email Cancel
          13  Illinois lawmakers inch closer to new hospital tax despite worries Illinois lawma
          14  What You Should Know About Sales Jobs In Chicago By Jeffrey Lee Thousands of peop
          15  The Art Institute of Chicago Share Visit the worlds best museum, the Art Institut
          16  Chicago Yacht Club Race to Mackinac update With a supporting team of many Old Goa
          17  Church becomes 1st in southern Illinois with solar panels Its the first church in
          19  North Central Illinois development group touts region North Central Illinois deve
          22  Can middle-class candidate defeat millionaires in Illinois? Can middleclass candi

              Sentiment_Score Sentiment_Label
          1          0.036443         Positive
```

```
3         0.037209        Positive
10        0.043118        Positive
13        0.015837        Positive
14        0.057823        Positive
15        0.039514        Positive
16        0.023256        Positive
17        0.012821        Positive
19        0.035519        Positive
22        0.005780        Positive
```

In [341]: # going to check for top organizations in the positive articles
          import nltk as nltk
          import nltk.corpus
          from nltk.text import Text
          import pandas as pd
          import re
          import sys

          #turn positive text into string
          positive_text= str(positive)

In [386]: positive_text[:1000]

Out[386]: ''

In [483]: pos_result = positive.sort_values(by='Sentiment_Score', ascending=False)

In [485]: pos_result.head(50)

Out[485]:

```
496                                                              nbc, chee
2056                                                             nbc, chee
4569  Instead of property tax relief measure, Illinois lawmakers file tax reform sid
5209                                                         nbc, cheer, 
90                                                           nbc, cheer, 
3326                                                         Not sure 
3269  Awesome Day / Song: My God Is Awesome - Fellowship Church Chicago Hi This is a
5934  Classic  Chicago neighborhood Love the Willows perfect location for great food
5436  Good morning Chicago!! Vegas was great, Good morning Chicago Vegas was great, 
5197                            Know anything abt suburbs of chicago? moving from west
2723  Chicago skyline, really pretty in person  Chicago Picture Chicago skyline, rea
1632  How housing could help Chicago bag Amazon's HQ2 - Crain's Chicago Business By 
5984  chicago chic chicago chic by yugenstateofmind featuring levi jeans White tee 3
5463  Wilson Sporting Goods Office by Gensler, Chicago  Illinois Color, light and te
6259             Southern Illinois Soybean Wheat Meeting Having some really fun discu
3737  Winebid no longer allowing pickup in Illinois Im in Illinois... Even with the 
3667  Chicago Cop shot multiple times and died I will be there in April, luckily we 
5487  In wake of Illinois bail bond reform, advocates here push national legislation
645   Salem iuka even Carlyle asap - m4w (Illinois) favorite this post Salem iuka eve
```

```
5847  Chicago is a beautiful city. #lake #lake Chicago is a beautiful city. lake lak
5444          nbc, sure, chicago fire, why not, jesse spencer, matthew casey via g
1070
4911  Chicago To : Moving and Housing advice Hello Folks,We currently live in the Ch
5927  Southern Illinois University Edwardsville settles free speech lawsuit Southern
2451  Attorney / - Illinois-Chicago Illinois  Chicago Submit Resume Chicago office o
5884  Illinois: Red light camera to be installed in the Chicago Loop Illinois: Red l
3747  1910 Prairie  Chicago, IL  $525,000 Acres 0.21 Stunning woodwork and master cr
5501  ComEd's Chicago South Side microgrid hits a nerve Kolata said hes hopeful the
5461
771                                                          My View: 2011
414             State rep wants to mirror federal K-12 savings benefits in Illino
4053  Hollywood's 'go-to medium' brings psychic abilities to Hammond casino - Chicag
3243  Call an Amazon truce, top economists tell cities vying for HQ2 - Government Ne
5232  Amazing School | EveryBlock Chicago ExcellentSt Bruno Catholic School has exce
757   Historic Neighborhoods - Chicago Hi, I was curious what are some neighborhoods
827                    Feb 11: Illinois 8th Senate District Candidates Forum I hope you
4914  Mr. Illini maps Illinois: 2018 edition Also, well revisit a map that I shared
1460  Illinois State Rep. Breen files legislation to eliminate red light cameras in
6756  University of Chicago students, faculty protest potential campus visit from 'w
5179  North Shore Chicago Burbs PPI and Service Im about to get a 2011 Cayenne S. Lo
4812  Chicago NW Suburbs HP Tuner recommendation? Chicago NW Suburbs HP Tuner recomm
574   Chicago NW Suburbs HP Tuner recommendation? Chicago NW Suburbs HP Tuner recomm
7097
5951  is chicago a city that good to travel on budget? i is chicago a city that good
5640                                      What are the odds we will get a foot of s
2047  Illinois governor | Letters | pantagraph.com Save I viewed the You Tube video
634   Raoul, Biss, Garcia land significant endorsements - Blogs On Politics - Crain'
803   Happy to be here from Illinois Hi all, Im Stephen from Central Illinois. Im a
2766  Robert Rauschenberg's Rare Duchamp Readymade Goes to Art Institute of Chicago
5186  Chicago approves $700M mixed-use project, includes citys future 6th-tallest to
```

	Sentiment_Score	Sentiment_Label
496	0.166667	Positive
2056	0.166667	Positive
4569	0.153846	Positive
5209	0.153846	Positive
90	0.153846	Positive
3326	0.142857	Positive
3269	0.131868	Positive
5934	0.128205	Positive
5436	0.117647	Positive
5197	0.111111	Positive
2723	0.107143	Positive
1632	0.107143	Positive
5984	0.102564	Positive
5463	0.102041	Positive
6259	0.100000	Positive

```
3737          0.100000        Positive
3667          0.100000        Positive
5487          0.098039        Positive
645           0.095238        Positive
5847          0.093750        Positive
5444          0.090909        Positive
1070          0.090909        Positive
4911          0.088542        Positive
5927          0.088235        Positive
2451          0.085714        Positive
5884          0.085106        Positive
3747          0.084967        Positive
5501          0.084507        Positive
5461          0.083333        Positive
771           0.083333        Positive
414           0.083333        Positive
4053          0.083333        Positive
3243          0.081967        Positive
5232          0.081633        Positive
757           0.080000        Positive
827           0.080000        Positive
4914          0.079710        Positive
1460          0.079646        Positive
6756          0.079365        Positive
5179          0.078431        Positive
4812          0.078431        Positive
574           0.078431        Positive
7097          0.076923        Positive
5951          0.076923        Positive
5640          0.076923        Positive
2047          0.076923        Positive
634           0.075472        Positive
803           0.074627        Positive
2766          0.073529        Positive
5186          0.071429        Positive
```

In [490]: `positive_text= str(pos_result[:300])`

In [491]:
```python
import os
from graphviz import Source

java_path = "C:\ProgramData\Oracle\Java\javapath\java.exe"
os.environ['JAVAHOME'] = java_path

# Change the path according to your system
stanford_classifier = 'C:\Program Files\stanford-ner-2017-06-09\classifiers\english.a
stanford_ner_path = 'C:\Program Files\stanford-ner-2017-06-09\stanford-ner.jar'
```

```
            stanford_parser_path = 'C:\Program Files\stanford-parser-full-2017-06-09\stanford-pai
            stanford_parser_models_path = 'C:\Program Files\stanford-parser-full-2017-06-09\stani

            from nltk.parse.stanford import StanfordDependencyParser
            from nltk.tag.stanford import StanfordNERTagger

            # Creating Tagger Object
            st = StanfordNERTagger(stanford_classifier, stanford_ner_path, encoding='utf-8')
            sdp = StanfordDependencyParser(path_to_jar=stanford_parser_path, path_to_models_jar=s
```

C:\Users\Joan\Anaconda3\lib\site-packages\nltk\tag\stanford.py:183: DeprecationWarning:
The StanfordTokenizer will be deprecated in version 3.2.5.
Please use nltk.tag.corenlp.CoreNLPPOSTagger or nltk.tag.corenlp.CoreNLPNERTagger instead.
  super(StanfordNERTagger, self).__init__(*args, **kwargs)

In [492]: #tokenizing and tagging text
          tokenized_positive_text = nltk.word_tokenize(positive_text)
          classified_positive_text = st.tag(tokenized_positive_text)

          classified_p_text_df = pd.DataFrame(classified_positive_text)

          classified_p_text_df.drop_duplicates(keep='first', inplace=True)
          classified_p_text_df.reset_index(drop=True, inplace=True)
          classified_p_text_df.columns = ["Entities", "Labels"]

In [493]: #counts by labels
          classified_p_text_df.groupby("Labels").count()

Out[493]:   Entities
          Labels
          LOCATION   17
          O   703
          ORGANIZATION   71
          PERSON   22

In [494]: #list of organization
          entities_df = classified_p_text_df.loc[classified_p_text_df["Labels"].isin(['LOCATION
          entities_df.reset_index(drop=True, inplace=True)
          entities_df.head(20)

Out[494]:   Entities   Labels
          0   Illinois   LOCATION
          1   Chicago   LOCATION
          2   Vegas   PERSON
          3   Vegas   LOCATION
          4   Chicago ORGANIZATION
          5   Picture ORGANIZATION
          6   bag ORGANIZATION
```

```
7        Amazon   ORGANIZATION
8         Crain          PERSON
9        Dennis          PERSON
10       Rodkin          PERSON
11       Wilson   ORGANIZATION
12     Sporting   ORGANIZATION
13        Goods   ORGANIZATION
14       Office   ORGANIZATION
15      Gensler          PERSON
16     Southern   ORGANIZATION
17     Illinois   ORGANIZATION
18      Soybean   ORGANIZATION
19        Wheat   ORGANIZATION
```

In [495]: *#per class notes, the Stanford NER tagger does have capability to support multi-word*
```
          tokenized_text = nltk.word_tokenize(positive_text)
          classified_text = st.tag(tokenized_text)

          netagged_words = classified_text

          entities = []
          labels = []

          from itertools import groupby
          for tag, chunk in groupby(classified_text, lambda x:x[1]):
              if tag != "O":
                  #print("%-12s"%tag, " ".join(w for w, t in chunk))
                  entities.append(' '.join(w for w, t in chunk))
                  labels.append(tag)


          entities_all = list(zip(entities, labels))
          entities_unique = list(set(zip(entities, labels))) #unique entities
```

In [496]: entities_df = pd.DataFrame(entities_unique)
```
          entities_df.columns = ["Entities", "Labels"]
          entities_df.groupby('Labels').count()
```

Out[496]:   Entities
```
          Labels
          LOCATION          16
          ORGANIZATION      31
          PERSON            15
```

In [498]: entities_df = pd.DataFrame(entities_all)
```
          entities_df.columns = ["Entities", "Labels"]
          persons_df = entities_df.loc[entities_df["Labels"].isin(['LOCATION','ORGANIZATION','
          counts_df = persons_df.groupby('Entities').count()
```

36

```python
counts_df.rename(columns={"Labels": "Mentions"}, inplace=True)
counts_df.sort_values(by=['Mentions'], ascending=False).head(50)
```

Out[498]:

Entities	Mentions
Chicago	24
Illinois	20
Elmhurst College Jazz Festival Begins	2
Elmhurst College Jazz Festival	2
ComEd	2
Hyperloop Transportation Technologies	2
Salem	2
Southern Illinois University Edwardsville	2
Art Institute	2
West Chicago Street	2
Carlyle	2
Blockchain Illinois	2
MENU Split Illinois	1
Illinois Senate	1
Illinois Senate Committee	1
Wilson Sporting Goods Office	1
Illinois-Chicago Illinois Chicago Submit Resume Chicago	1
Illinois-Edwardsville	1
Kolata	1
Lincoln	1
Lisa Madigan	1
Tanya Manus	1
Split Illinois	1
MERRILLVILLE	1
Matt Finn	1
NBC Chicago	1
Obama Center	1
West Chicago High School	1
Reclaim Chicago Endorse	1
Vegas	1
Senate	1
Southern Illinois Soybean Wheat Meeting Having	1
RDA	1
Emerging Business Partners	1
Hatchery Chicago Breaking News Ingredion for Emerging Business Partners	1
Chicago Tribune Random News Chicago TribuneLocal	1
Art Institute of Chicago Share Visit	1
Boston	1
Bronzeville	1
CTU	1
Cassandra Smith	1
Chicago Loop Illinois	1
Chicago Picture Chicago	1

```
            Chicago Roof Top                                                    1
            Chicago South Side                                                  1
            Chicago Teachers Union Charter                                      1
            Chicago Tribune News                                                1
            Chicago Vegas                                                       1
            Hammond                                                             1
            Chicago bag Amazon                                                  1
```

In [499]: `from wordcloud import WordCloud`

In [505]: `entities= str(entities_df['Entities'])`

In [506]: `wordcloud = WordCloud().generate(entities)`

In [507]: `plt.imshow(wordcloud, interpolation='bilinear')`
          `plt.axis("off")`

Out[507]: `(-0.5, 399.5, 199.5, -0.5)`

Topic modeling among the postiive sentiment articles

In [423]: `pos_list = positive['article'].tolist()`
          `pos_list[:1]`

Out[423]: `["My View: 2011 worker's comp changes beginning to benefit Illinois My View: 2011 wor`

In [424]: `stop = set(stopwords.words('english'))`
          `exclude = set(string.punctuation)`
          `lemma = WordNetLemmatizer()`
          `def clean(doc):`

```python
        stop_free = " ".join([i for i in doc.lower().split() if i not in stop])
        punc_free = ''.join(ch for ch in stop_free if ch not in exclude)
        normalized = " ".join(lemma.lemmatize(word) for word in punc_free.split())
        return normalized

    doc_clean = [clean(doc).split() for doc in pos_list]
```

In [425]: # Creating the term dictionary of our courpus, where every unique term is assigned a
          dictionary = corpora.Dictionary(doc_clean)

          # Converting list of documents (corpus) into Document Term Matrix using dictionary p
          doc_term_matrix = [dictionary.doc2bow(doc) for doc in doc_clean]

In [426]: # Creating the object for LDA model using gensim library
          Lda = gensim.models.ldamodel.LdaModel

In [427]: pos_clean = [clean(doc).split() for doc in pos_list]

In [428]: len(pos_clean)

Out[428]: 2996

In [429]: print(*pos_clean[:1], sep='\n\n')

['view', '2011', 'worker', 'comp', 'change', 'beginning', 'benefit', 'illinois', 'view', '2011

In [430]: # Creating the term dictionary of our courpus, where every unique term is assigned a

          dictionary = corpora.Dictionary(pos_clean)

          # Converting list of documents (corpus) into Document Term Matrix using dictionary p

          %time doc_term_matrix = [dictionary.doc2bow(doc) for doc in pos_clean]

Wall time: 1.68 s

3 topic model

In [395]: # Creating the object for LDA model using gensim library
          Lda = gensim.models.ldamodel.LdaModel
          numtopics = 3

          # Running and Training LDA model on the document term matrix.
          %time ldamodel = Lda(doc_term_matrix, num_topics=numtopics, id2word = dictionary, pas

Wall time: 15min 50s

```
In [396]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=3), sep='\n\n')

(0, '0.014*"illinois" + 0.013*"state" + 0.008*"said"')

(1, '0.024*"chicago" + 0.009*"said" + 0.008*"snow"')

(2, '0.012*"democratic" + 0.012*"illinois" + 0.012*"pritzker"')


In [397]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=5), sep='\n\n')

(0, '0.014*"illinois" + 0.013*"state" + 0.008*"said" + 0.007*"chicago" + 0.006*"would"')

(1, '0.024*"chicago" + 0.009*"said" + 0.008*"snow" + 0.008*"city" + 0.005*"school"')

(2, '0.012*"democratic" + 0.012*"illinois" + 0.012*"pritzker" + 0.011*"candidate" + 0.011*"bi"

In [398]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=7), sep='\n\n')

(0, '0.014*"illinois" + 0.013*"state" + 0.008*"said" + 0.007*"chicago" + 0.006*"would" + 0.006*

(1, '0.024*"chicago" + 0.009*"said" + 0.008*"snow" + 0.008*"city" + 0.005*"school" + 0.005*"cou

(2, '0.012*"democratic" + 0.012*"illinois" + 0.012*"pritzker" + 0.011*"candidate" + 0.011*"bi"


In [399]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=10), sep='\n\n')

(0, '0.014*"illinois" + 0.013*"state" + 0.008*"said" + 0.007*"chicago" + 0.006*"would" + 0.006*

(1, '0.024*"chicago" + 0.009*"said" + 0.008*"snow" + 0.008*"city" + 0.005*"school" + 0.005*"cou

(2, '0.012*"democratic" + 0.012*"illinois" + 0.012*"pritzker" + 0.011*"candidate" + 0.011*"bi"
```

5 topic model

```
In [431]: # Creating the object for LDA model using gensim library
          Lda = gensim.models.ldamodel.LdaModel
          numtopics = 5

          # Running and Training LDA model on the document term matrix.
          %time ldamodel = Lda(doc_term_matrix, num_topics=numtopics, id2word = dictionary, pas

Wall time: 16min 53s


In [432]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=3), sep='\n\n')
```

```
(0, '0.019*"chicago" + 0.011*"school" + 0.010*"said"')

(1, '0.016*"pritzker" + 0.015*"bi" + 0.011*"illinois"')

(2, '0.012*"chicago" + 0.011*"snow" + 0.009*"illinois"')

(3, '0.023*"chicago" + 0.010*"city" + 0.007*"said"')

(4, '0.017*"illinois" + 0.016*"state" + 0.008*"said"')


In [433]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=5), sep='\n\n')

(0, '0.019*"chicago" + 0.011*"school" + 0.010*"said" + 0.008*"student" + 0.006*"illinois"')

(1, '0.016*"pritzker" + 0.015*"bi" + 0.011*"illinois" + 0.010*"campaign" + 0.010*"said"')

(2, '0.012*"chicago" + 0.011*"snow" + 0.009*"illinois" + 0.007*"said" + 0.006*"democratic"')

(3, '0.023*"chicago" + 0.010*"city" + 0.007*"said" + 0.005*"police" + 0.004*"park"')

(4, '0.017*"illinois" + 0.016*"state" + 0.008*"said" + 0.008*"tax" + 0.007*"would"')


In [434]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=7), sep='\n\n')

(0, '0.019*"chicago" + 0.011*"school" + 0.010*"said" + 0.008*"student" + 0.006*"illinois" + 0.0

(1, '0.016*"pritzker" + 0.015*"bi" + 0.011*"illinois" + 0.010*"campaign" + 0.010*"said" + 0.009

(2, '0.012*"chicago" + 0.011*"snow" + 0.009*"illinois" + 0.007*"said" + 0.006*"democratic" + 0

(3, '0.023*"chicago" + 0.010*"city" + 0.007*"said" + 0.005*"police" + 0.004*"park" + 0.004*"str

(4, '0.017*"illinois" + 0.016*"state" + 0.008*"said" + 0.008*"tax" + 0.007*"would" + 0.006*"yea


In [435]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=10), sep='\n\n')

(0, '0.019*"chicago" + 0.011*"school" + 0.010*"said" + 0.008*"student" + 0.006*"illinois" + 0.0

(1, '0.016*"pritzker" + 0.015*"bi" + 0.011*"illinois" + 0.010*"campaign" + 0.010*"said" + 0.009

(2, '0.012*"chicago" + 0.011*"snow" + 0.009*"illinois" + 0.007*"said" + 0.006*"democratic" + 0

(3, '0.023*"chicago" + 0.010*"city" + 0.007*"said" + 0.005*"police" + 0.004*"park" + 0.004*"str

(4, '0.017*"illinois" + 0.016*"state" + 0.008*"said" + 0.008*"tax" + 0.007*"would" + 0.006*"yea
```

0.3.7 Looking at the articles scored as having negative sentiment

In [334]: `negative = all_articles_df[all_articles_df.Sentiment_Label=="Negative"]`

In [436]: `negative.head(20)`

Out[436]:

```
0    Man dies hours after being shot leaving Southwest Side party - Chicago Tribune
2    Illinois governor candidate stands by ad blasted as racist CHICAGO AP  A Republi
4    6 shot within 3 hours to start President's Day weekend in Chicago | 1 killed, 12
5    Toys R Us closing seven Chicago area stores Toys R Us closing seven Chicago area
7    Holocaust denier running for Congress has no opponents in Illinois Republican pr
8    Black NHL player Devante Smith-Pelly taunted by Chicago fans Black NHL player Dev
11   Duerson Act would ban pre-teen tackle football in Illinois Duerson Act would ban
12   Oak Park Police Investigate Alleged Abduction ń CBS Chicago Sponsored By Oak Par
18   Owens-Illinois Inc (OI) Files 10-K for the Fiscal Year Ended on December 31, 201
20   The Latest: Slain Chicago officer remembered as 'model cop' The Latest: Slain Ch
26   Illinois Veterinarian Has More To Say About This  Dog Flu Dog Insurance News Ill
28   Chicago police commander killed, 5 others wounded in city shootings - Chicago Tr
34   Illinois Gov. Bruce Rauner to deliver State of the State Email Want to read throu
38   Illinois Smoking Ban Has Not Damaged Casino Revenue, Study Claims News  Live Cas
41   Heavy Snow, White-Out Conditions Cause 11-Car Crash in Indiana - NBC Chicago The
42   Biggest snow in over a year bearing down on large swath of the Chicago area begi
46   Another Chicago-area school closes due to flu A school in North Riverside will be
47   Southwest Air's Chicago De-Icing Stumble Was Third in Two Months Southwest Airli
50   CTA riders facing significant Monday morning delays | Chicago Sun-Times Nearly ev
52   Oboi Reed Blasts the Citys Failure to Address Its Biased Bike Ticketing Problem

     Sentiment_Score Sentiment_Label
0          -0.014706        Negative
2          -0.068182        Negative
4          -0.017391        Negative
5          -0.003597        Negative
7          -0.014409        Negative
8          -0.035599        Negative
11         -0.014056        Negative
12         -0.016949        Negative
18         -0.018779        Negative
20         -0.001845        Negative
26         -0.042969        Negative
28         -0.023973        Negative
34         -0.020270        Negative
38         -0.011050        Negative
41         -0.015810        Negative
42         -0.018692        Negative
46         -0.013115        Negative
47         -0.007519        Negative
50         -0.013514        Negative
52         -0.008403        Negative
```

```
In [441]: neg_result= negative.sort_values(by=['Sentiment_Score'])

In [478]: neg_result.head(20)

Out[478]:
          856                                                                   Chicago
          7289
          125    Illegals, Chicago and loathsome little demagogues Illegals, Chicago and loathso
          3287                                                        Ex-staffer: Top Il
          3315   Illegals, Chicago and 'loathsome little demagogues' | On Air Republican Chicago
          4444                                                             Shootings in
          4394                                         Government watchdog: States
          4297                                                             Shootings in
          7263                                    Calls for investigations, new leade
          5597                                                          Chicago ro
          5226   Illegals, Chicago and loathsome little demagogues Republican Chicago Alderman
          3895               Also seen from the Chicago Autoshow: 50th anniversary badging God Da
          7180   Chicago police teaming up with feds once more to fight shocking rise in violent
          3525   Lawyers argue for firing squad for Ohio death row inmate after unsuccessful let
          239    Slain Chicago cop advocated for stiffer sentences; suspect in his shooting is
          1816   Illinois cattle disease study in need of farm cooperators in Southern Illinois
          6936   Man convicted in deadly 2012 Indianapolis house explosion dies in prison  Chica
          292    Fellow detainees applauded suspected Chicago cop killer: report   Fellow detai
          669    Brian Mackey talks Illinois budget, sexual harassment claims against Madigan |
          3478   Northwestern professor takes leave of absence amid harassment allegations - Chi

                 Sentiment_Score Sentiment_Label
          856         -0.200000        Negative
          7289        -0.181818        Negative
          125         -0.176471        Negative
          3287        -0.166667        Negative
          3315        -0.161290        Negative
          4444        -0.153846        Negative
          4394        -0.153846        Negative
          4297        -0.153846        Negative
          7263        -0.142857        Negative
          5597        -0.133333        Negative
          5226        -0.133333        Negative
          3895        -0.130435        Negative
          7180        -0.125654        Negative
          3525        -0.120482        Negative
          239         -0.120000        Negative
          1816        -0.117647        Negative
          6936        -0.116279        Negative
          292         -0.115385        Negative
          669         -0.114754        Negative
          3478        -0.114754        Negative

In [455]: #turn NEGATIVE text into string
```

```
            negative_text= str(neg_result[:200])

In [456]:  negative_text

Out[456]:  "

In [457]:  #tokenizing and tagging text
           tokenized_text = nltk.word_tokenize(negative_text)
           classified_text = st.tag(tokenized_text)

           classified_text_df = pd.DataFrame(classified_text)

           classified_text_df.drop_duplicates(keep='first', inplace=True)
           classified_text_df.reset_index(drop=True, inplace=True)
           classified_text_df.columns = ["Entities", "Labels"]

In [458]:  #counts by labels
           classified_text_df.groupby("Labels").count()

Out[458]:                   Entities
           Labels
           LOCATION              19
           O                    621
           ORGANIZATION          42
           PERSON                20

In [459]:  #list of organization
           entities_df = classified_text_df.loc[classified_text_df["Labels"].isin(['LOCATION','O
           entities_df.reset_index(drop=True, inplace=True)
           entities_df.head(20)

Out[459]:            Entities        Labels
           0          Chicago      LOCATION
           1          Illinois     LOCATION
           2          Chicago        PERSON
           3          Alderman       PERSON
           4          Nicholas       PERSON
           5          Sposato        PERSON
           6          Tuckers   ORGANIZATION
           7        Democratic  ORGANIZATION
           8            Party   ORGANIZATION
           9               of   ORGANIZATION
           10         Illinois  ORGANIZATION
           11            Ohio      LOCATION
           12         Chicago   ORGANIZATION
           13         Tribune   ORGANIZATION
           14     Indianapolis    LOCATION
           15            Brian       PERSON
           16           Mackey       PERSON
```

```
          17      Madigan       PERSON
          18         WGN  ORGANIZATION
          19       Radio  ORGANIZATION
```

In [460]: *#per class notes, the Stanford NER tagger does have capability to support multi-word*
```
          tokenized_text = nltk.word_tokenize(negative_text)
          classified_text = st.tag(tokenized_text)

          netagged_words = classified_text

          entities = []
          labels = []

          from itertools import groupby
          for tag, chunk in groupby(classified_text, lambda x:x[1]):
              if tag != "O":
                  #print("%-12s"%tag, " ".join(w for w, t in chunk))
                  entities.append(' '.join(w for w, t in chunk))
                  labels.append(tag)


          entities_all = list(zip(entities, labels))
          entities_unique = list(set(zip(entities, labels))) #unique entities
```

In [461]: entities_df = pd.DataFrame(entities_unique)
```
          entities_df.columns = ["Entities", "Labels"]
          entities_df.groupby('Labels').count()
```

Out[461]:   Entities
```
          Labels
          LOCATION           13
          ORGANIZATION       21
          PERSON             10
```

In [462]: entities_df = pd.DataFrame(entities_all)
```
          entities_df.columns = ["Entities", "Labels"]
          persons_df = entities_df.loc[entities_df["Labels"].isin(['LOCATION','ORGANIZATION','
          counts_df = persons_df.groupby('Entities').count()
          counts_df.rename(columns={"Labels": "Mentions"}, inplace=True)
          counts_df.sort_values(by=['Mentions'], ascending=False).head(200)
```

Out[462]:   Mentions
```
          Entities
          Chicago                                                     48
          Illinois                                                    22
          Illinois Corrections Department                              4
          Chicago Tribune                                              4
          Chicago Alderman Nicholas Sposato                            3
          Brian Mackey                                                 2
```

```
       CHICAGO                                                          2
       Indianapolis                                                     2
       Tuckers                                                          2
       Illinois Arlington Heights Cook County                           2
       Home International News                                          2
       Steve Bannon                                                     2
       SOPHIA TAREEN                                                    1
       NPR                                                              1
       Ohio                                                             1
       PBS                                                              1
       Transgender Inmates Sue Illinois Corrections Department          1
       Thomas Lindstrom                                                 1
       Marquette Park                                                   1
       Transgender Inmates Sue Illinois Corrections Department Illinois News   1
       Tyler Durden                                                     1
       Ukrainian Village                                                1
       University of Chicago| Latest News Raw                           1
       WGN Radio                                                        1
       Washington Post                                                  1
       Mount Vernon Illinois Jefferson County                          1
       Illinois Illinois                                                1
       Madigan                                                          1
       Latoya Frost                                                     1
       La Salle                                                         1
       Bruce Rauner                                                     1
       Greece                                                           1
       Democratic Party of Illinois                                     1
       Democratic Party                                                 1
       Cicero                                                           1
       Chicago Tribune Storm                                            1
       Chicago Sun-Times Northwestern                                   1
       Chicago Loop                                                     1
       Chicago Daily Herald Elgin                                       1
       CTA                                                              1
       CHICAGO AP A Republican                                          1
       CHICAGO AP                                                       1
       West                                                             1
```

In [464]: `#topic modeling on negative sentiment articles`
```python
neg_list = neg_result['article'].tolist()
neg_list[:1]
```

Out[464]: `['Chicago public schools close tomorrow. Chicago public schools close tomorrow.']`

In [465]:
```python
stop = set(stopwords.words('english'))
exclude = set(string.punctuation)
lemma = WordNetLemmatizer()
def clean(doc):
```

```
            stop_free = " ".join([i for i in doc.lower().split() if i not in stop])
            punc_free = ''.join(ch for ch in stop_free if ch not in exclude)
            normalized = " ".join(lemma.lemmatize(word) for word in punc_free.split())
            return normalized

        neg_clean = [clean(doc).split() for doc in neg_list]
```

In [466]: # Creating the term dictionary of our courpus, where every unique term is assigned a
```
        dictionary = corpora.Dictionary(neg_clean)

        # Converting list of documents (corpus) into Document Term Matrix using dictionary p
        doc_term_matrix = [dictionary.doc2bow(doc) for doc in neg_clean]
```

In [467]: # Creating the object for LDA model using gensim library
```
        Lda = gensim.models.ldamodel.LdaModel
```

In [468]: neg_clean = [clean(doc).split() for doc in neg_list]

In [469]: len(neg_clean)

Out[469]: 3367

In [470]: print(*neg_clean[:1], sep='\n\n')

['chicago', 'public', 'school', 'close', 'tomorrow', 'chicago', 'public', 'school', 'close', 't

In [471]: # Creating the term dictionary of our courpus, where every unique term is assigned a

```
        dictionary = corpora.Dictionary(neg_clean)

        # Converting list of documents (corpus) into Document Term Matrix using dictionary p

        %time doc_term_matrix = [dictionary.doc2bow(doc) for doc in neg_clean]
```

Wall time: 1.21 s

3 topic model

In [472]: # Creating the object for LDA model |using gensim library
```
        Lda = gensim.models.ldamodel.LdaModel
        numtopics = 3

        # Running and Training LDA model on the document term matrix.
        %time ldamodel = Lda(doc_term_matrix, num_topics=numtopics, id2word = dictionary, pas
```

Wall time: 19min 44s

```
In [473]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=3), sep='\n\n')

(0, '0.016*"school" + 0.013*"chicago" + 0.010*"said"')

(1, '0.016*"illinois" + 0.013*"chicago" + 0.012*"state"')

(2, '0.021*"chicago" + 0.016*"police" + 0.013*"said"')


In [474]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=5), sep='\n\n')

(0, '0.016*"school" + 0.013*"chicago" + 0.010*"said" + 0.007*"student" + 0.006*"year"')

(1, '0.016*"illinois" + 0.013*"chicago" + 0.012*"state" + 0.010*"snow" + 0.008*"said"')

(2, '0.021*"chicago" + 0.016*"police" + 0.013*"said" + 0.010*"football" + 0.008*"shooting"')


In [475]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=7), sep='\n\n')

(0, '0.016*"school" + 0.013*"chicago" + 0.010*"said" + 0.007*"student" + 0.006*"year" + 0.005*"

(1, '0.016*"illinois" + 0.013*"chicago" + 0.012*"state" + 0.010*"snow" + 0.008*"said" + 0.005*"

(2, '0.021*"chicago" + 0.016*"police" + 0.013*"said" + 0.010*"football" + 0.008*"shooting" + 0


In [476]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=10 ), sep='\n\n')

(0, '0.016*"school" + 0.013*"chicago" + 0.010*"said" + 0.007*"student" + 0.006*"year" + 0.005*"

(1, '0.016*"illinois" + 0.013*"chicago" + 0.012*"state" + 0.010*"snow" + 0.008*"said" + 0.005*"

(2, '0.021*"chicago" + 0.016*"police" + 0.013*"said" + 0.010*"football" + 0.008*"shooting" + 0
```

0.3.8   5 topic modeling

```
In [418]: # Creating the object for LDA model using gensim library
          Lda = gensim.models.ldamodel.LdaModel
          numtopics = 5

          # Running and Training LDA model on the document term matrix.
          %time ldamodel = Lda(doc_term_matrix, num_topics=numtopics, id2word = dictionary, pa

Wall time: 16min 34s


In [419]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=3), sep='\n\n')
```

```
(0, '0.022*"football" + 0.016*"said" + 0.012*"tackle"')

(1, '0.012*"said" + 0.011*"chicago" + 0.007*"illinois"')

(2, '0.018*"illinois" + 0.017*"state" + 0.008*"said"')

(3, '0.022*"chicago" + 0.019*"snow" + 0.010*"inch"')

(4, '0.029*"chicago" + 0.020*"police" + 0.016*"school"')


In [420]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=5), sep='\n\n')
(0, '0.022*"football" + 0.016*"said" + 0.012*"tackle" + 0.012*"illinois" + 0.011*"duerson"')

(1, '0.012*"said" + 0.011*"chicago" + 0.007*"illinois" + 0.006*"year" + 0.006*"city"')

(2, '0.018*"illinois" + 0.017*"state" + 0.008*"said" + 0.006*"tax" + 0.006*"year"')

(3, '0.022*"chicago" + 0.019*"snow" + 0.010*"inch" + 0.009*"weather" + 0.008*"illinois"')

(4, '0.029*"chicago" + 0.020*"police" + 0.016*"school" + 0.012*"said" + 0.011*"man"')


In [421]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=7), sep='\n\n')
(0, '0.022*"football" + 0.016*"said" + 0.012*"tackle" + 0.012*"illinois" + 0.011*"duerson" + 0

(1, '0.012*"said" + 0.011*"chicago" + 0.007*"illinois" + 0.006*"year" + 0.006*"city" + 0.005*"

(2, '0.018*"illinois" + 0.017*"state" + 0.008*"said" + 0.006*"tax" + 0.006*"year" + 0.006*"wou

(3, '0.022*"chicago" + 0.019*"snow" + 0.010*"inch" + 0.009*"weather" + 0.008*"illinois" + 0.008

(4, '0.029*"chicago" + 0.020*"police" + 0.016*"school" + 0.012*"said" + 0.011*"man" + 0.010*"o


In [422]: print(*ldamodel.print_topics(num_topics=numtopics, num_words=10), sep='\n\n')
(0, '0.022*"football" + 0.016*"said" + 0.012*"tackle" + 0.012*"illinois" + 0.011*"duerson" + 0

(1, '0.012*"said" + 0.011*"chicago" + 0.007*"illinois" + 0.006*"year" + 0.006*"city" + 0.005*"

(2, '0.018*"illinois" + 0.017*"state" + 0.008*"said" + 0.006*"tax" + 0.006*"year" + 0.006*"wou

(3, '0.022*"chicago" + 0.019*"snow" + 0.010*"inch" + 0.009*"weather" + 0.008*"illinois" + 0.008

(4, '0.029*"chicago" + 0.020*"police" + 0.016*"school" + 0.012*"said" + 0.011*"man" + 0.010*"o
```