Executive Summary

Key reasons for declining population:

- The crime rate in Chicago is very high and it seems like the police were not able to fight it until recently
- Illinois is economically relatively weak and sees declining revenues for its businesses
- Wages are comparably low and unemployment is above national average
- Real estate prices in Chicago are surging
- Taxes in Chicago are above national average
- The educational system is under-funded and ranks badly
- Bad weather

Recommendations to reverse the trend:

- Hire more officers, build trust in the police, introduce harder laws, and use technology to fight crime
- Introduce intervention programs to prevent young people from joining gangs
- Companies in Chicago need to adapt new technologies
- Reduces taxes and regulations
- Ensure affordable housing
- Invest in the education system and support students from poor neighborhoods to get into college
- Invest in infrastructure, such as roads, schools etc.

Recommendations (positive sentiment)

Word cloud of recommendations: A healthy economic environment seems to be key

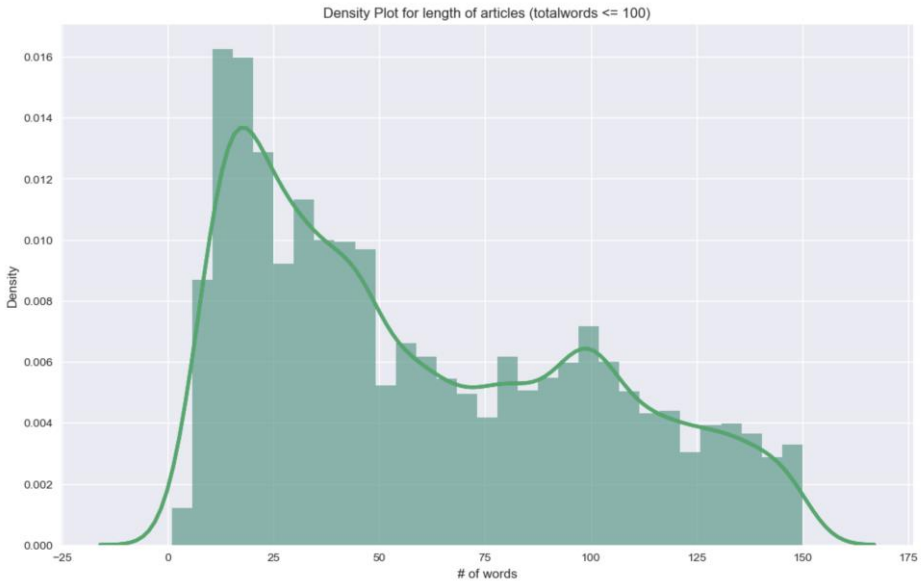
Article Clean-up and Filtering

1 Basic Cleaning:

column	# observations	dtypes
crawled_date	49,122	datetime
language	49,122	object
text	49,122	object
title	49,122	object

- Test whether all articles are in English language
- Combine titles and text columns to capture all words at once
- Adding stopwords often used in news articles to the dictionary and remove all stopwords from combined text base on that
- Remove punctuation, special characters, and tags
- Convert text to lowercase
- Lemmatize text

2 Filtering irrelevant articles:



- Looking at the distribution of word counts for articles with 100 or less words
- Assumption: Articles with **20 words or less** cannot convey meaningful insights for this task
- Most likely, those articles are advertisements etc.
- By removing those articles, removed ~10% of all articles in the corpus
- Further filtering applied based on topic modeling (see following steps)

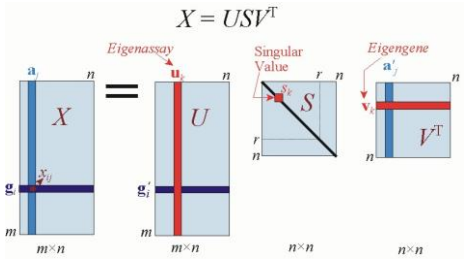
Topic Detection

1 Latent Semantic Analysis (LSA):

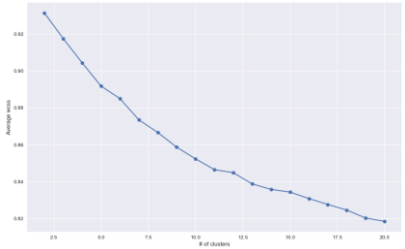
$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \times \text{idf}(t, D)$$

	Doc 1	Doc 2	...	Doc *n*
Term(s) 1	12	2	...	1
Term(s) 2	0	1	...	0
...	...	...	...	
Term(s) *n*	0	6	...	3

Using term frequency-inverse document frequency (TFIDF) representation to quantify word importance in the corpus

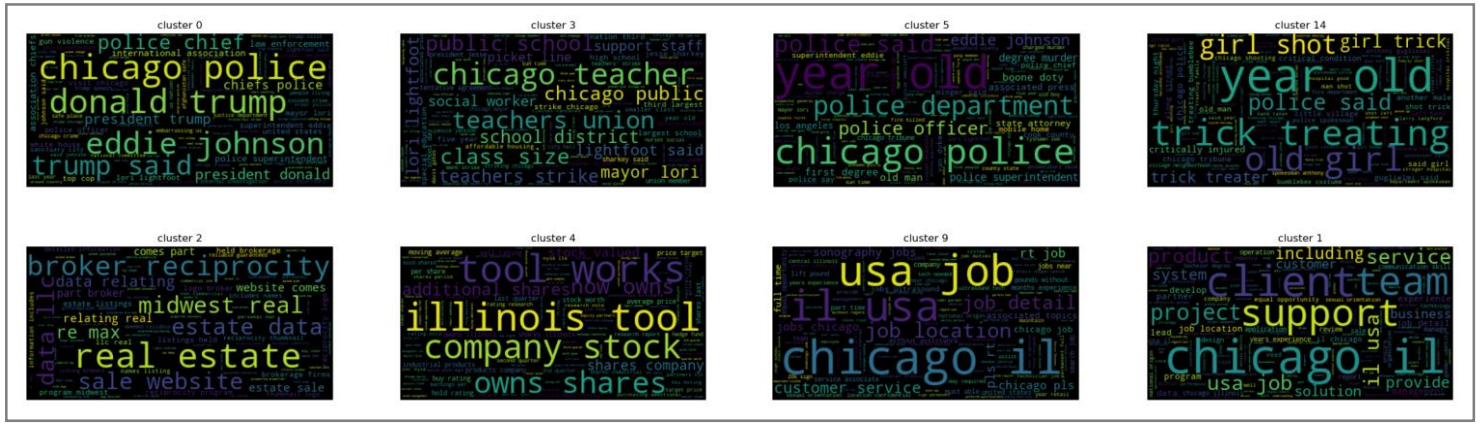


Reduced dimensions of TFIDF matrix to n = 1,500, while capturing ~60% variance using singular value decomposition (SVD) for better processing



Using within cluster sum of squares (wcss) and silouhette scores determined optimum number of clusters as **k = 15**

2 Topic labeling:



Relevant Groups

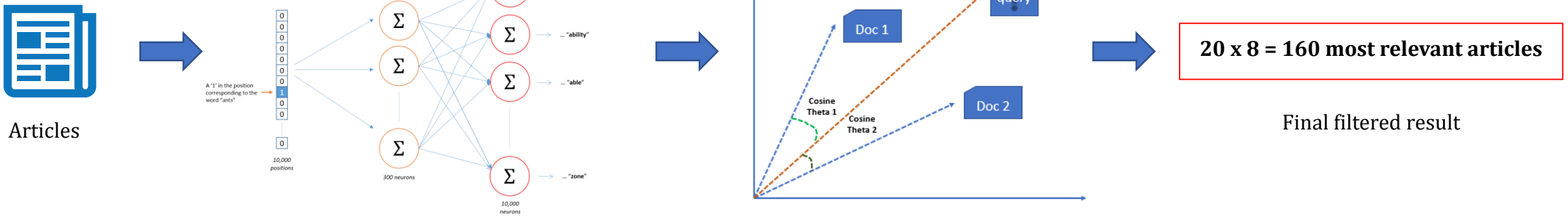
- Labeled articles based on LSA analysis
- Identified 7 irrelevant and 8 relevant groups
- Extracted top 15 keywords per cluster and created word clouds

- **Relevant groups:** Politics (0), jobs/support (1), real estate (2), education (3), financials/business (4), police (5), jobs (9), crime (14) (*see word clouds*)

- **Irrelevant groups (dropped):** URLs/advertisement (6), football (7), baseball (8), ice hockey/basketball (10), Chicago marathon (11), college sport (12), other sport (13)

Sentiment Analysis

1 Filtering similar articles by keyword query:



Google pre-trained word2vec

Cosine similarity between query and document (average sentence embeddings)

- Using Googles pre-trained word2vec model ('GoogleNews-vectors-negative300.bin.gz') to build a search engine that finds articles relevant to the topic
- Partition articles within each cluster into sentences, define similarity based on query 'population decline', get average similarity for every article based on sentence similarity, rank articles within each cluster based on average sentence similarity
- Selected 20 most relevant articles from each cluster, which leaves me with 160 articles in total

Sentiment Analysis (cont.)

1 Methodology:

- Used corpus of negative and positive words compiled by the University of Pittsburgh and customized the dictionaries
- For each sentence within the respective clusters, computed the sentiment score between -1 and 1 (negative=reason; positive=recommendation)

2 Reasons (negative sentiment):

Police:
- Unsatisfactory police work (e.g., only 10% of murders charged)
- Harsh policing lead to distrust into the police in Chicago
- There are not enough police officers to make the needed amount of arrests

Crime:
- Increased crime (30% increase over past years)
- Marijuana use is associated with higher crime rates (and has been legalized)

Business:
- Revenue for companies in Illinois is declining by ~ 1-3% on average
- Downstate Illinois is especially impacted by labor shortage
- Relative economic weakness of Illinois compared to other states

Jobs:
- Wages do not grow fast enough (e.g., for technicians)
- Jobless rate in Illinois is higher than the national average
- Between 50-60% of the jobs in Illinois could be at risk of becoming automated (outdated industries)
- Largest sectors of job loss are manufacturing, education, and health services

Real estate:
- Rising real estate prices (rose by ~ 25% compared to 2018)
- Investors who wholesale real estate in Illinois are now required to run their deals through a broker/agent (this could further increases prices)

Weather:
- Bad weather in Chicago
- Cold weather is impacting businesses as well (e.g., hotels)

Education:
- Inequality in children's knowledge and skills at school entry
- Fewer than 23% of the high schools received the highest rating
- The Chicago public schools' pension fund is only 50% funded

Fiscal/governmental:
- Chicago has the nation's worst pension crisis under almost every measure
- Chicago is an extreme outlier fiscally (high taxes)
- Chicago has one of the nation's worst credit rating and ~ $ 30 billion in debt (no money to invest)
- Chicago has a $ 838 million budget deficit

Sentiment Analysis (cont.)

3 Recommendations (positive sentiment):

Police:
- Need for more officers
- Start building trust between police and the community
- improving policing and stopping the flow of illegal guns will require new and better gun laws, a big investment in police training and more support for the police to help them cope with the pressures of the job
- creating a unit devoted solely to investigating homicides and improving training

Crime:
- Chicago has seen nearly 30% decrease in murder rates over the past several years (down by 11% this year compared to previous year)
- Better use of technology drives the violent crime rate down
- Violent crime fell to the second lowest point in five years
- invest in intervention programs piloted by organizations like the youth peace center in Roseland, the institute for non-violence in Austin, and Enlace in little village
- Criminal justice reforms
- cut off the pipeline of young people into gangs by providing many more safe spaces for an estimated 50,000 kids in Chicago's 15 most violence-plagued neighborhoods

Business:
- Companies in Illinois need to adapt new technologies to reduce costs, improve accuracy, and for better compliance
- pockets of strength need to be leveraged into an economy that allows for broader-based growth

Jobs:
- rather than allowing this deterioration to continue, Illinois policy leaders should reverse another trend: that of passing into law more taxes and regulations that will stunt local hiring

Real estate:
- Ensure affordable housing

Education:
- Individuals with low income working parents and people looking to transition into a new career need to be enabled to do so
- The city is doling out $ 31 million in equity grants to 219 schools with low or declining enrollment
- Larger class sizes will not affect academic performance and will save taxpayers millions of dollars
- Check whether students from struggling schools are prepared for college
- Implement after-school, evening, and weekend programming courses for kids that are at severe risk of leaving school otherwise

Fiscal/governmental:
- Invest in infrastructure (e.g., roads)

Entity Identification and Targeted Sentiment Analysis

- Applied nltk tagger to all relevant articles
- Filtered all organizations and all persons, calculated the count, and ranked them by their count in descending order
- Calculated the average sentiment score for all sentences that the respective organization or person was mentioned in (within the 160 relevant articles)
- Extracted most important sentences based for organizations and persons respectively based on count

Organizations Recommendations

- Businesses should come to Chicago for the software and manufacturing market
- Chicago seems to be interesting for supply chain businesses
- Chicago is a global city with a considerable market
- Overall, there is still some growth in Chicago/Illinois

Persons Recommendations

- People can find software jobs as well as blue collar jobs
- Most reasons for people to come to Chicago seem to be job related