

Capstone Proposal

6/29/18

**Xiaoran Cai, Sarah Wu,
Chinmayee Mohapatra, Yerin Jung,
Vatsal Randhar, Aditi Wadhawan,
Kylene Ye, Thibault Twahirwa**

Agenda

01 Executive Summary

02 Framework

03 Market Research

04 Observations in Data

05 Research Questions

06 Proposed Approach

07 Clarifications and Queries

Executive Summary

Situation

- ▶ XXX faces **strong competition** from its peers, such as XXX Business News and XXX.



Challenge

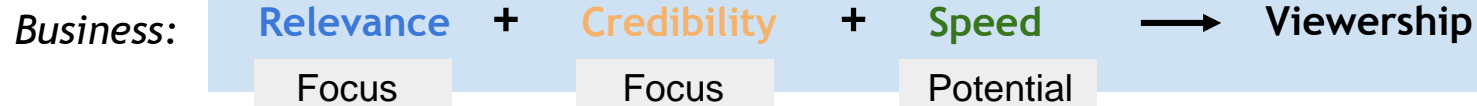
- ▶ To attract and retain viewership, explore new analytical strategies.



Approach

- ▶ We applied analytical methodologies such as word association and sentiment analysis to explore the dataset.

Proposed Criteria Framework



Viewership: Indicator of news story or program success

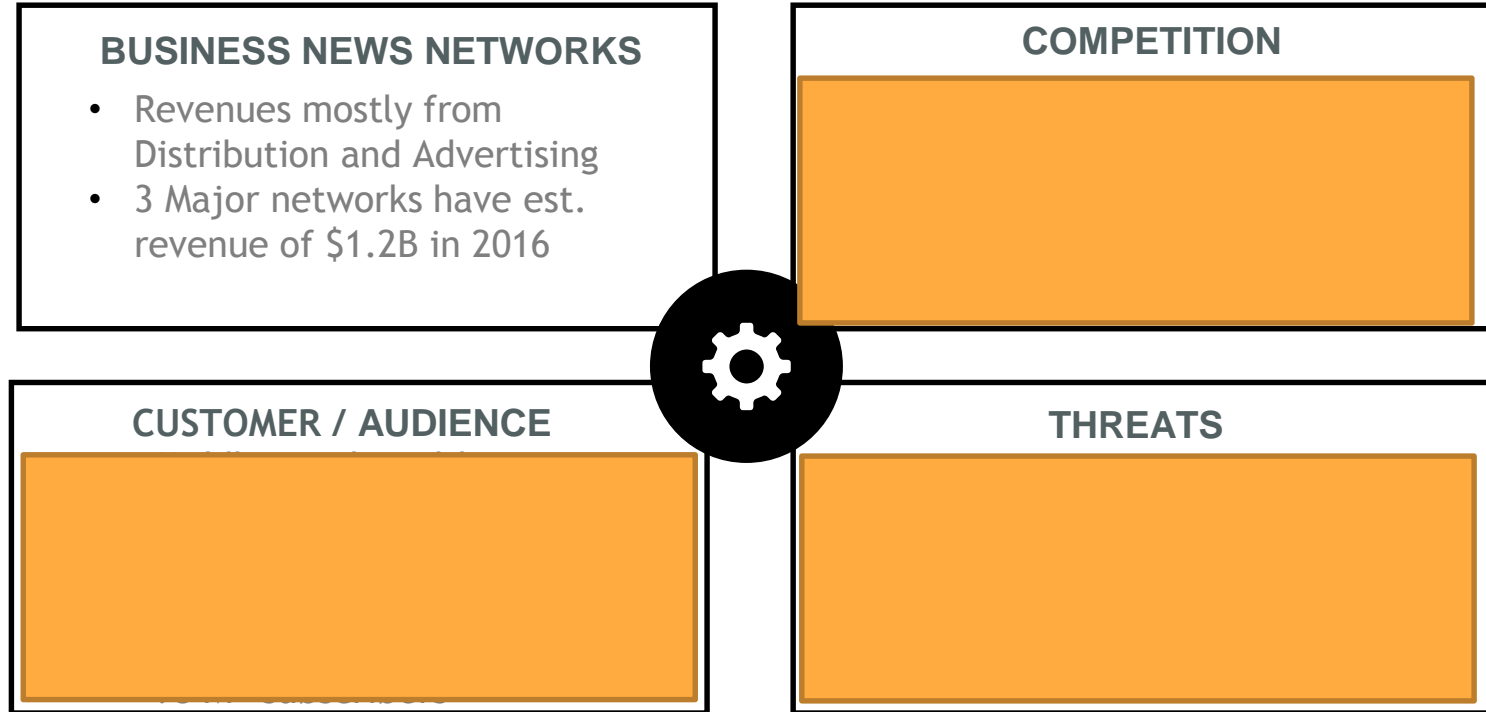
Relevance: How the words XXX uses are related to the public's interests

Credibility: Story-telling techniques; bias of content, trustworthiness, and fairness

Speed: Time gap between the occurrence of the events and reporting a news



Business Research



Source: Comcast Spotlight - Nielsen Scarborough USA Plus-MRI/Mosaic, De15-Ap17, Adults 18+, Cable Networks
 Watched Past 7 Days: XXX, Pew Research

Research - Competitive Landscape

Strong TV presence, but competitors have deep strengths in their offerings

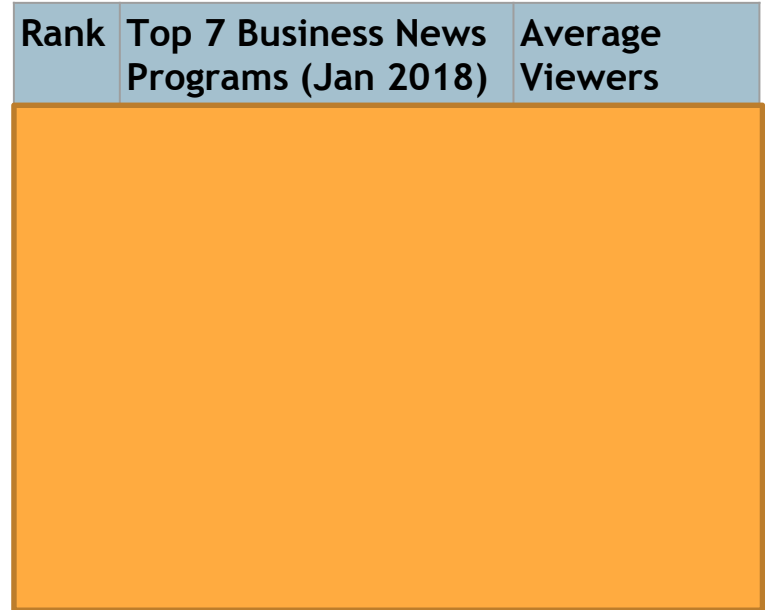
	Strong
	Medium
	Limited



Business News				
Entertainment/ Reality				
Political News				
Live Financial Coverage				
Lifestyle				



Business Day Viewership
(in 000s)



7

Data Analysis - Exploration and Collection

Tapping into the Competitors data

XXX

Captions	ID	Time_Stamp	Month	Hour	Weekday	Day
i have been working for this moment	1819_20170102100000A	1/2/2017 11:56	January	11	Mon	2
tilman tonight on billion dollar	1819_20170102100000B	1/2/2017 12:56	January	12	Mon	2
tilman tonight on the name of	1819_20170102100000C	1/2/2017 13:56	January	13	Mon	2
tilman tonight on you are the mad scientist	1819_20170102130000A	1/2/2017 14:56	January	14	Mon	2

2971 Transcripts

2017 weekdays data

Competitors

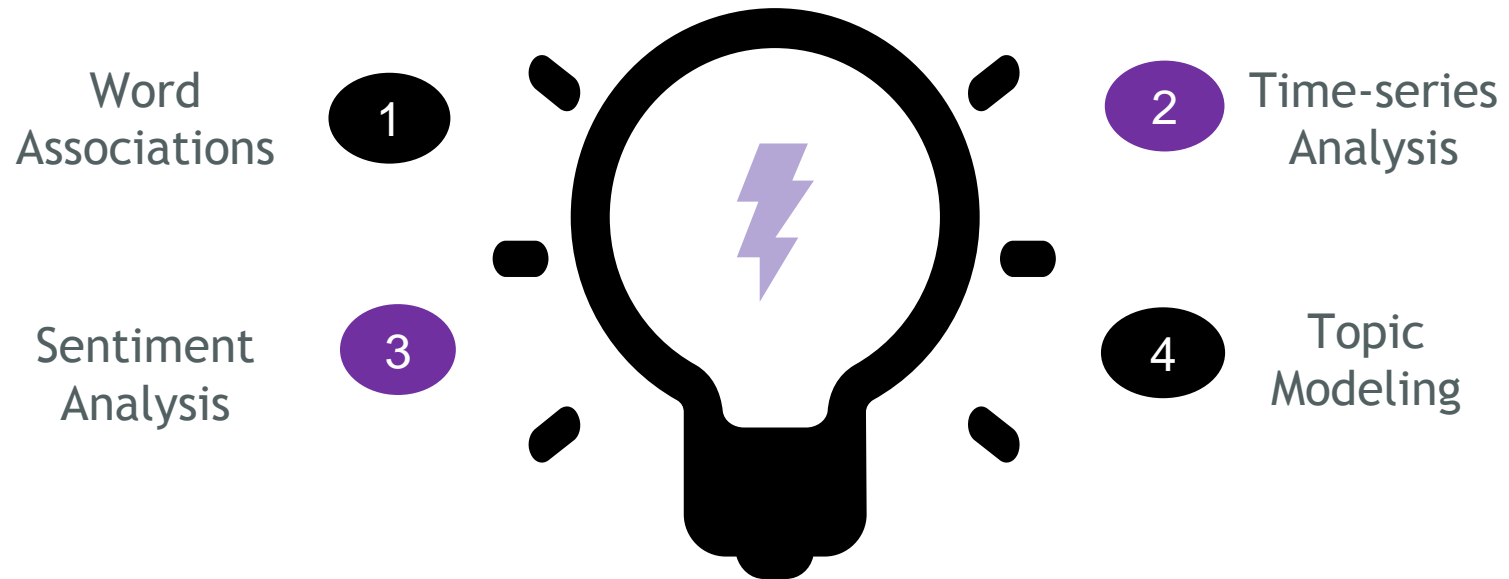
Archive.org
(Transcripts Collection)

import.io (Website)

BeautifulSoup (Python)

Data Analysis - Techniques

Potential Techniques to Use



Data Analysis - XXX

Changes in Top 10 Word Association with the key word 'Bitcoin'

 Bitcoin>

Month
October November December

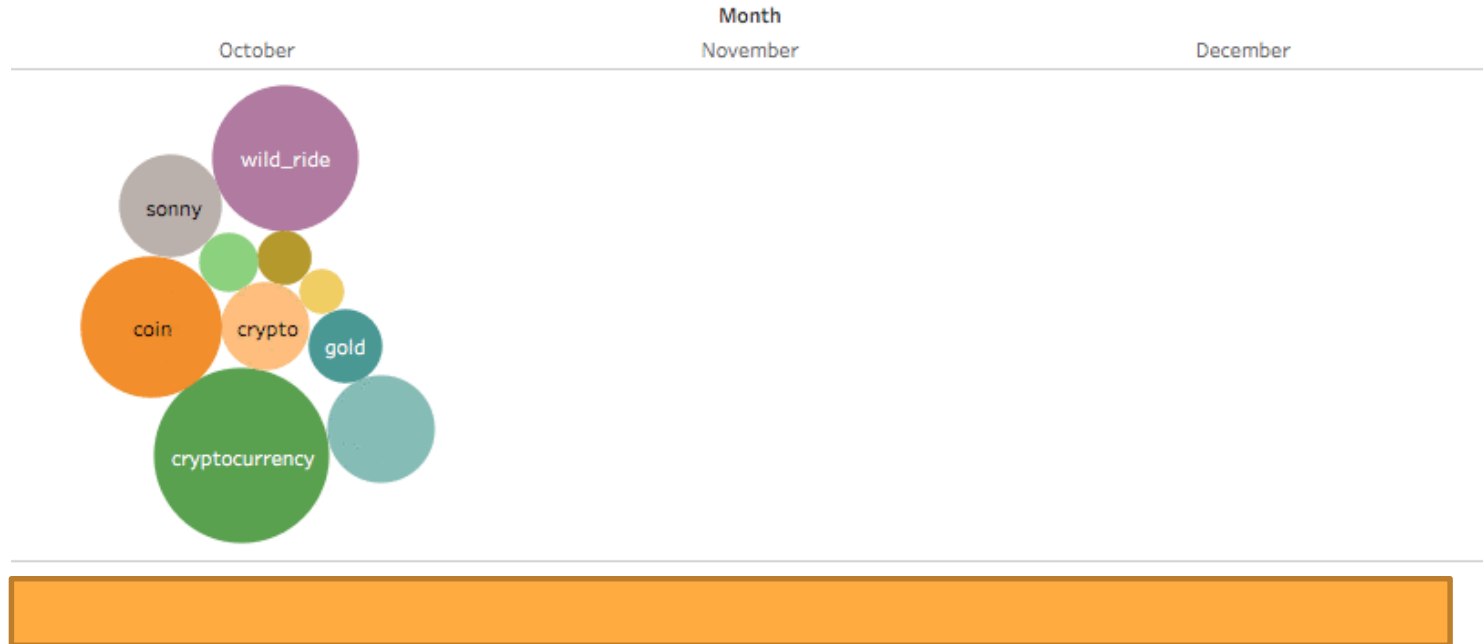


GIF showcases
differences in
word associations
over time.

Data Analysis - XXX

Changes in Top 10 Word Association with the key word 'Bitcoin'

< Bitcoin> - October



Data Analysis - XXX Business News

Changes in Top 10 Word Association with the key word 'Bitcoin'



Month
November

October

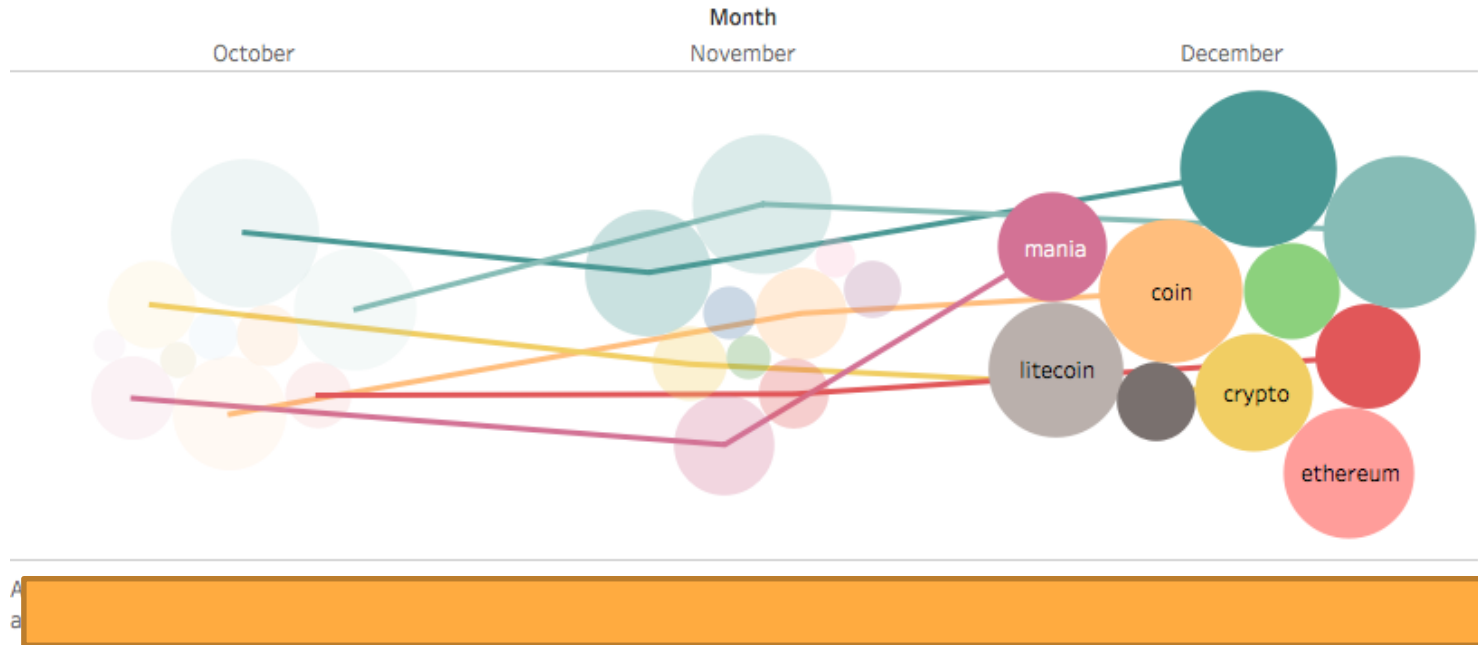
December



Data Analysis - XXX

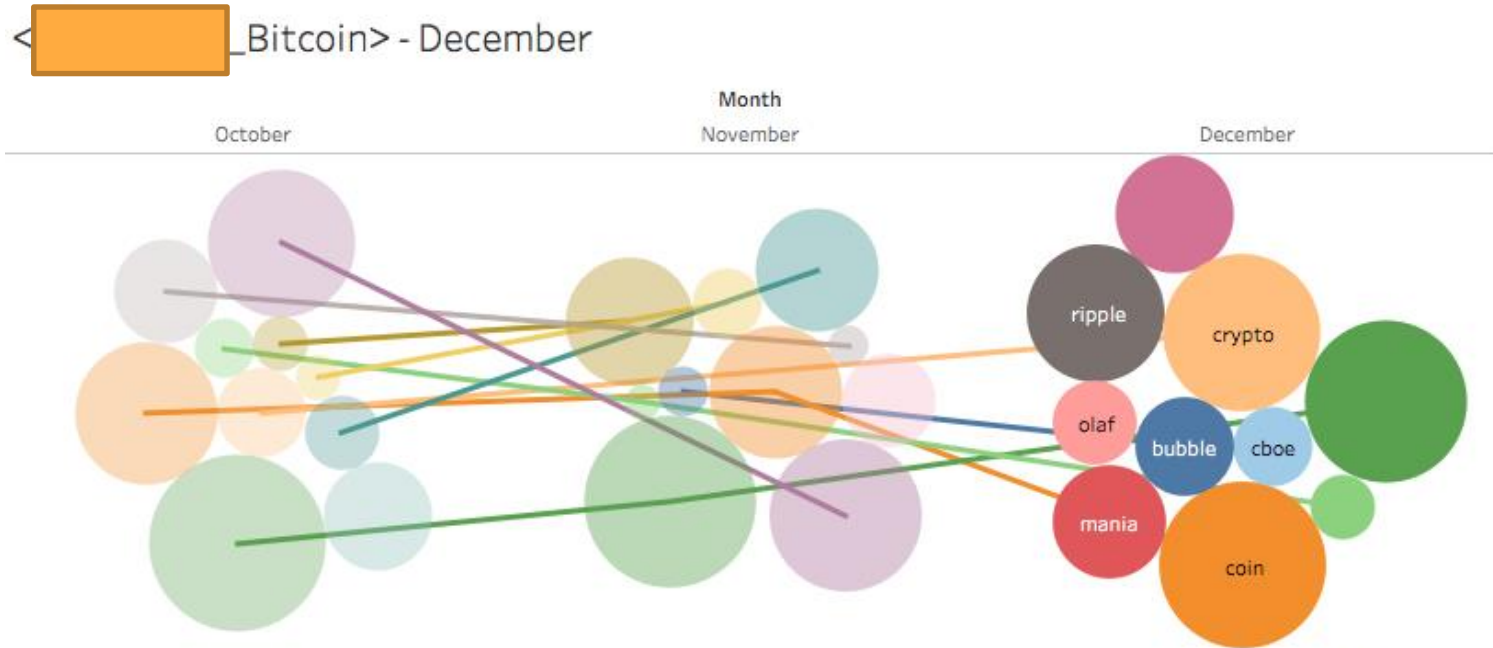
Top 10 Word Associations with the key word 'Bitcoin' in December

 Bitcoin>



Data Analysis - XXX

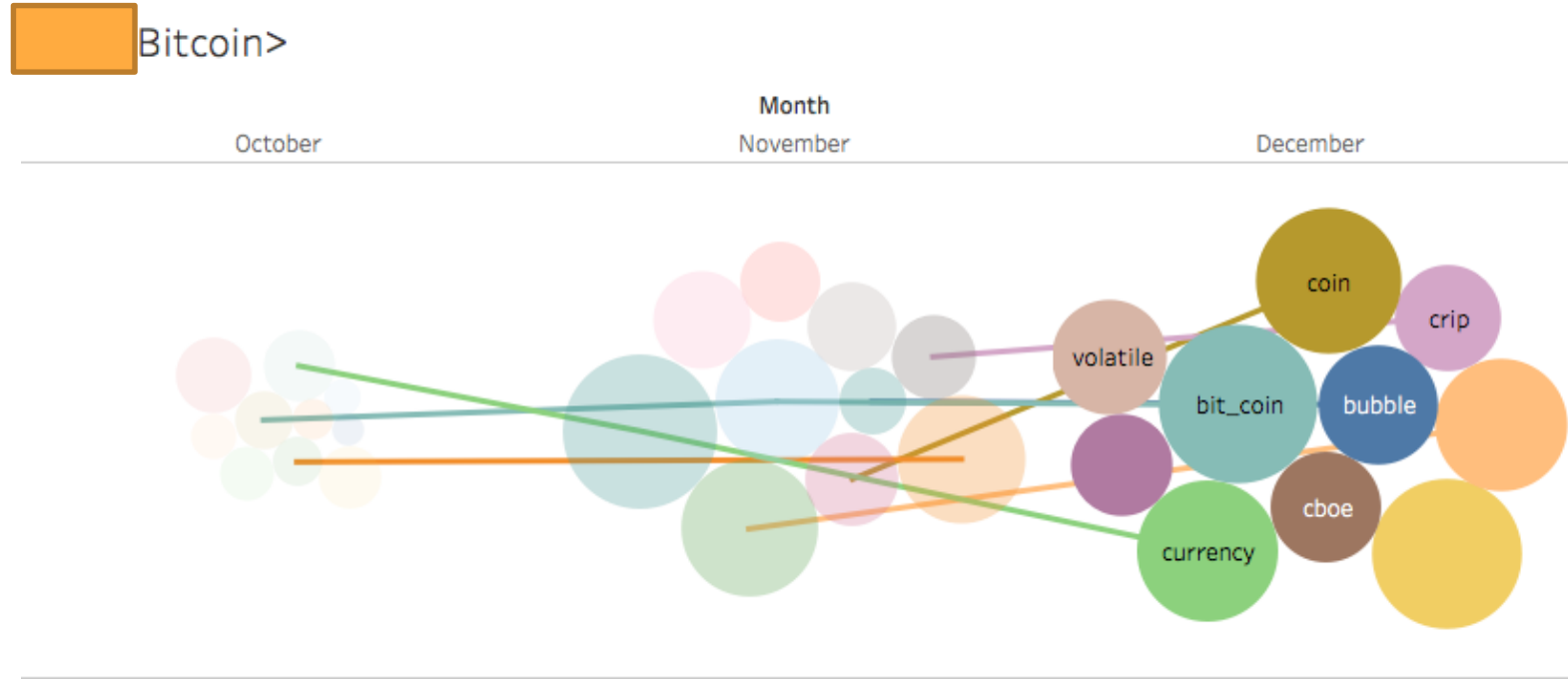
Top 10 Word Associations with the key word 'Bitcoin' in December



Asso-word broken down by Month Month on page December. Color shows details about Asso-word. Size shows details about Values as an attribute. The marks are labeled by Asso-word.

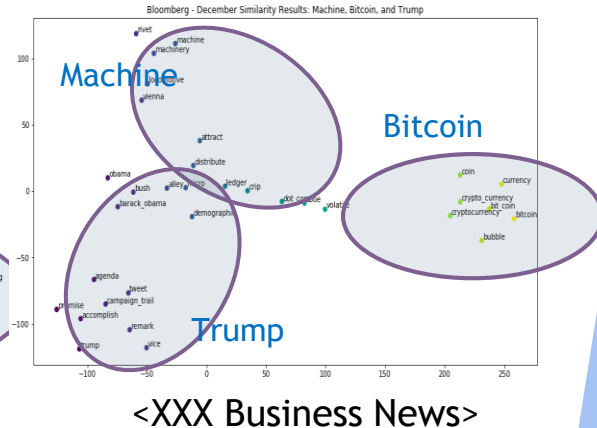
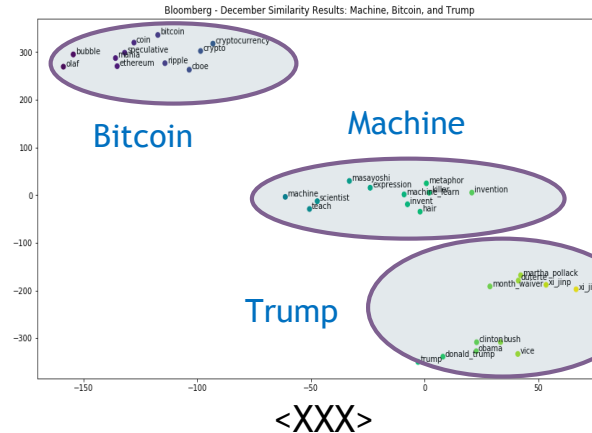
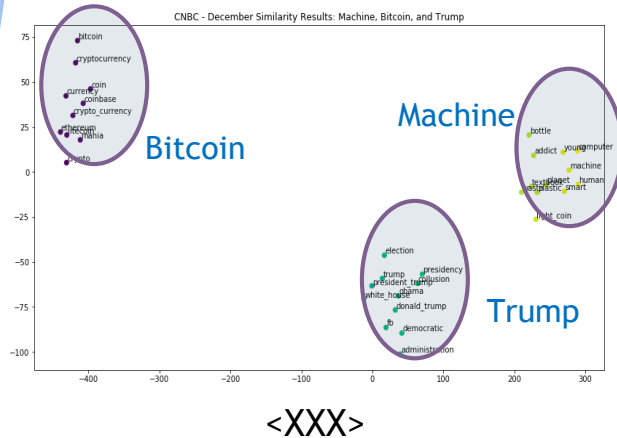
Data Analysis - XXX Business News

Top 10 Word Associations with the key word 'Bitcoin' in December



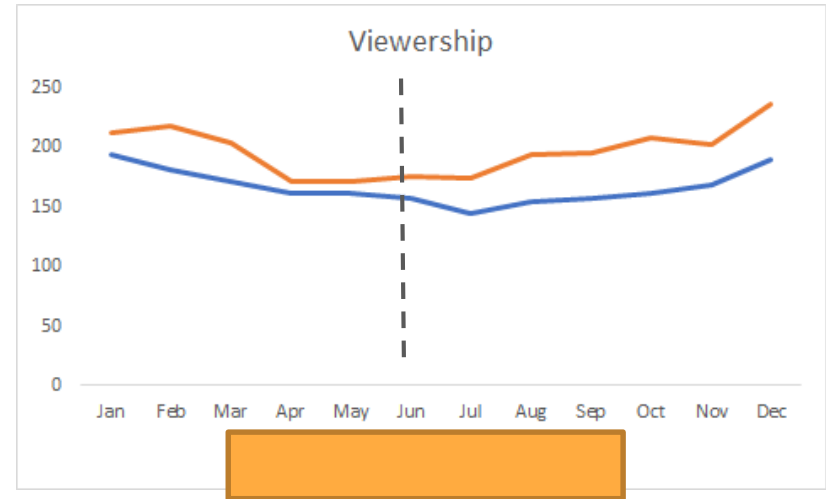
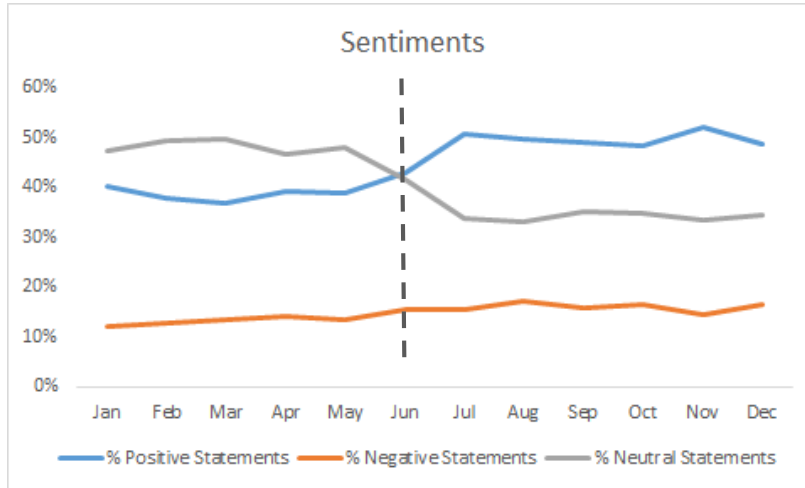
Data Analysis - Three Companies

Word Association Map with the three words: 'Bitcoin, Trump, and Machine'



Data Analysis - XXX

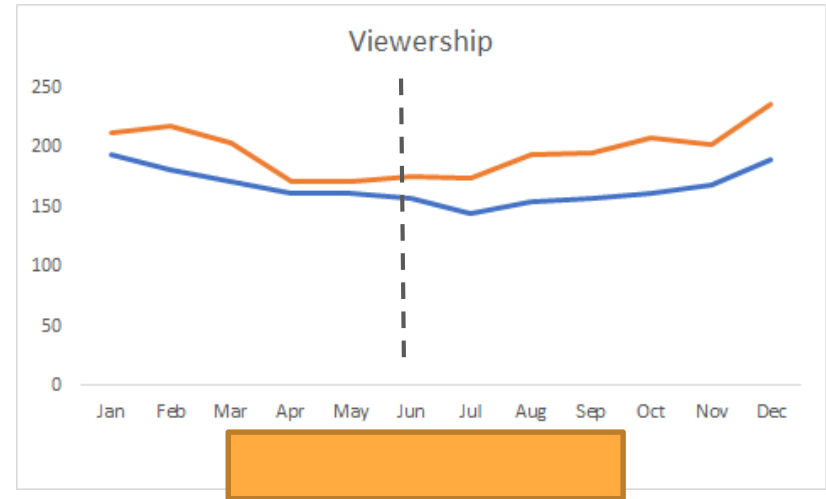
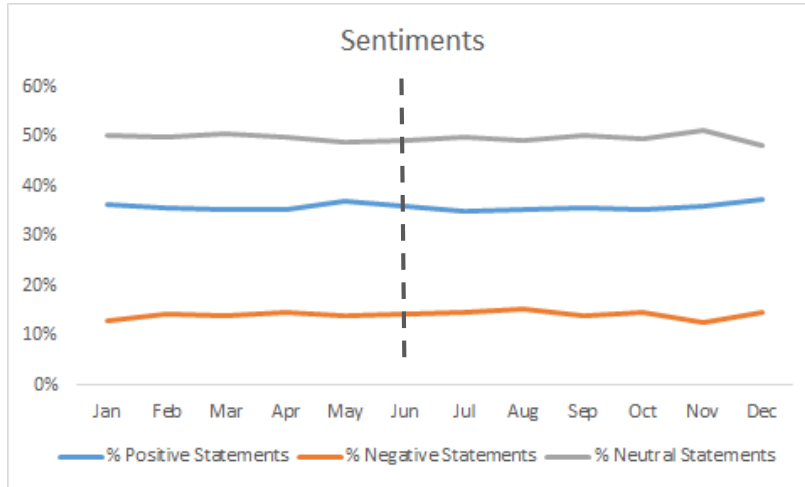
Sentiment Analysis of 'Trump' in 2017 with business day viewership



- During the mid year, XXX might have shifted its portrayal of Trump to a more positive light: observable decrease in neutrality.
- Hypothesis: Portrayal of content related to decrease in viewership

Data Analysis - XXX Business Model

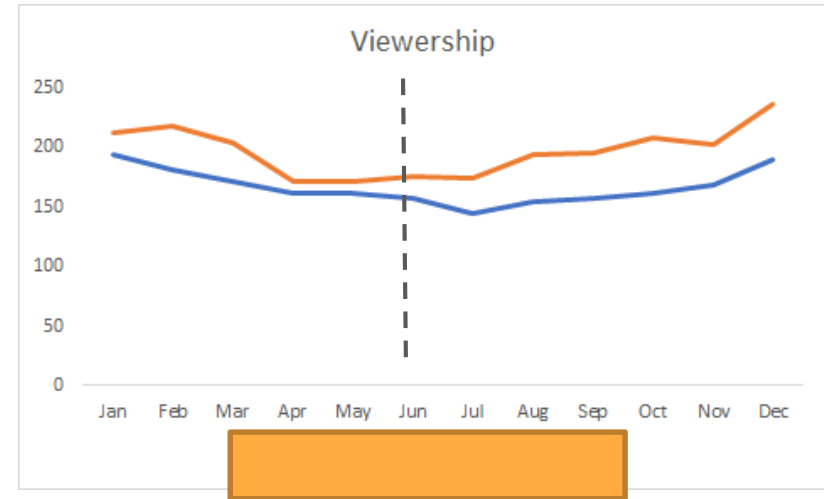
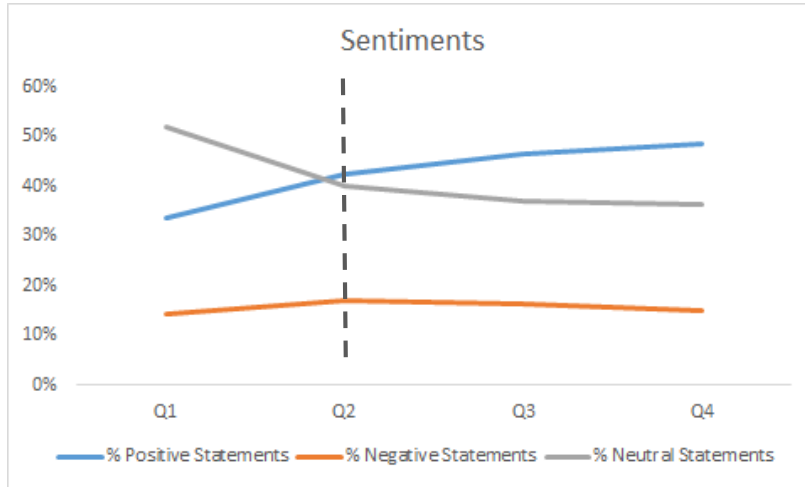
Sentiment Analysis of 'Trump' in 2017 with business day viewership



- XXX Business News has portrayed Trump consistently in the similar light.
- Hypothesis: Portrayal of content related to decrease in viewership

Data Analysis - XXX

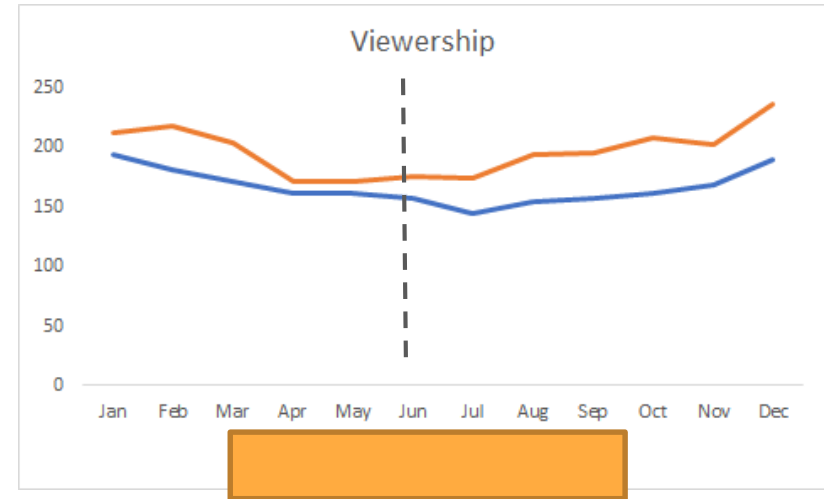
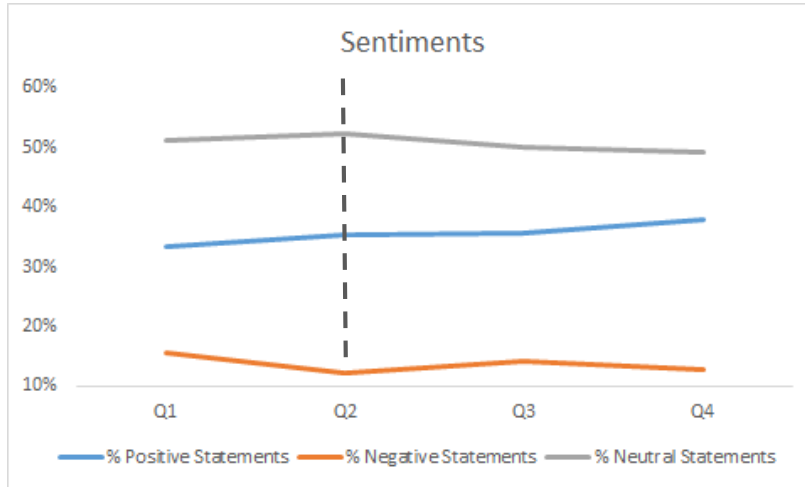
Sentiment Analysis of 'Bitcoin' in 2017 with business day viewership



- 20 percentage points increase in positive statements over 2017
- Hypothesis: Portrayal of content related to decrease in viewership

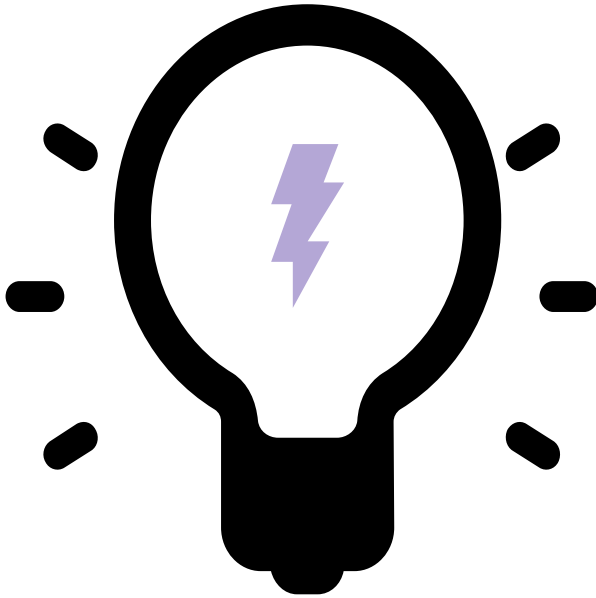
Data Analysis - XXX Business Model

Sentiment Analysis of 'Bitcoin' in 2017 with business day viewership



- 10 percentage points increase in positive statements over 2017
- Hypothesis: Portrayal of content related to decrease in viewership

Proposed Research Questions



How are the XXX's word associations different from the public's interests?

Identify how XXX's word association changes over time and how the changes differ from the public data's and competitors' word associations.

How does XXX portray topics?

Identify how certain topics are being portrayed differently from the competition in polarity and style.

Suggested Approach - Word Association

Technique:	Word Association Analysis
Data:	<ul style="list-style-type: none"> • XXX • Competitors (XXX and XXX) • Public data (Twitter or Google Trends)
Idea:	<ul style="list-style-type: none"> • Use selected business topic words from Fortune's Top Business Keywords • Assess <u>"How XXX's word associations differ from competitors' and the public's."</u>
Measurement:	<ul style="list-style-type: none"> • Measure how each company's word associations differ from public's word associations • Use two measures: <u>Information Entropy</u> and <u>Difference Counts</u>

Suggested Approach - Word Association

Information Entropy - For the same associated words

Words Associated with ' <i>Bitcoin</i> '	XXX	XXX Business News	Public (Google Trend / Twitter)	Entropy Change b/w XXX & Public	Entropy Change b/w XXX & Public	Which is more disparate ?
`Crypto`	0.58	0.34	0.67	$(0.58 - 0.67)^*$ $(\log(0.58/0.67))$	$(0.67 - 0.34)^*$ $(\log(0.67/0.34))$	XXX
`Legitimize`	0.63	0.84	0.76	$(0.63 - 0.76)^*$ $(\log(0.63/0.76))$	$(0.84 - 0.76)^*$ $(\log(0.84/0.76))$	XXX
`Litecoin`	0.43	0.56	0.57	$(0.43 - 0.57)^*$ $(\log(0.43/0.57))$	$(0.56 - 0.57)^*$ $(\log(0.56/0.57))$	XXX

Suggested Approach - Word Association

Penalized Difference Count : Count the number of different top N associated words and penalize them by values of those words

	Public(GoogleTrend/ Twitter)	XXX	XXX Business News
Top 7 words associated with 'Bitcoin'	A	B : 0.64	A: 0.54
	B	C : 0.54	B: 0.56
	C	D: 0.56	C: 0.34
	D	H: 0.34	F: 0.65
	E	J: 0.76	K: 0.89
	F	K: 0.23	D: 0.34
	G	E: 0.43	L: 0.56
Sum of Values	-	3.5	3.88
Difference Count (DC)	0/7	3/7	2/7
DC * Penalization		3/7* (0.34+0.76+0.23)/3.5	2/7 * (0.89+0.56) / 3.88

Suggested Approach - Sentiment Analysis

Technique:	Sentiment Analysis
Data:	<ul style="list-style-type: none"> • XXX • Competitors (XXX and XXX)
Idea:	<ul style="list-style-type: none"> • Use closed captioning data from XXX programs • Identify associated sentiments over multiple months in 2017, compare this with competitors' sentiment • Assess <u>"how certain topics are being portrayed differently in polarity and style."</u>
Measurement:	<ul style="list-style-type: none"> • Polarity scale from -1 to 1 • Equals negative and positive sentiments respectively.

Clarifications and Queries

Information Requested	Clarifying Questions	Success Criteria
<ul style="list-style-type: none"> Revenue breakdown per show <ul style="list-style-type: none"> Orders of magnitude Time-focused Targeted demographic for prime-time and business-time <ul style="list-style-type: none"> Focused bias analysis 	<ul style="list-style-type: none"> Are the 3 dimensions (Relevance, Credibility, Speed) a good framework to use? <ul style="list-style-type: none"> Which one should we prioritize? Is XXX seeing the same viewership decline as we hypothesized? 	<ul style="list-style-type: none"> Increase in relevance (<i>similarity in sentiments/ association</i>) by x % Identify probable drivers of shift in viewership

Thank you!

Any Questions?

Appendix

1. <https://www.adweek.com/tvnewser/2017-business-news-ratings-fbn-is-no-1-in-business-day/353311> : **2017 Business Day: FBN:** 197,000 total viewers, +28 percent / 25,000 A25-54, +14 percent, XXX: 167,000 total viewers, -11 percent / 30,000 A25-54, -19 percent
2. http://www.uwyo.edu/cojo/_files/docs/24600118gladney.pdf
3. https://www.researchgate.net/publication/237343900_THE_CREDIBILITY_OF_NEWSPAPERS_TELEVISION_NEWS_AND_ONLINE_NEWS
4. Allan Bell, The Language of News Media
5. https://link.springer.com/chapter/10.1007/978-3-642-37210-0_48