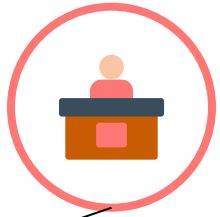
Fake News

- *And How Convincing Are They*

Lua Xuan Zi 3036181680

Fake news – And how convincing are they?

Overview



Problem Definition

- Data Science question
- Why is it important?
- How to answer?
- Difficulties in answering

1

Results

- Methods
- Results/Findings

2

Future works

- Future improvements
- Possible future studies

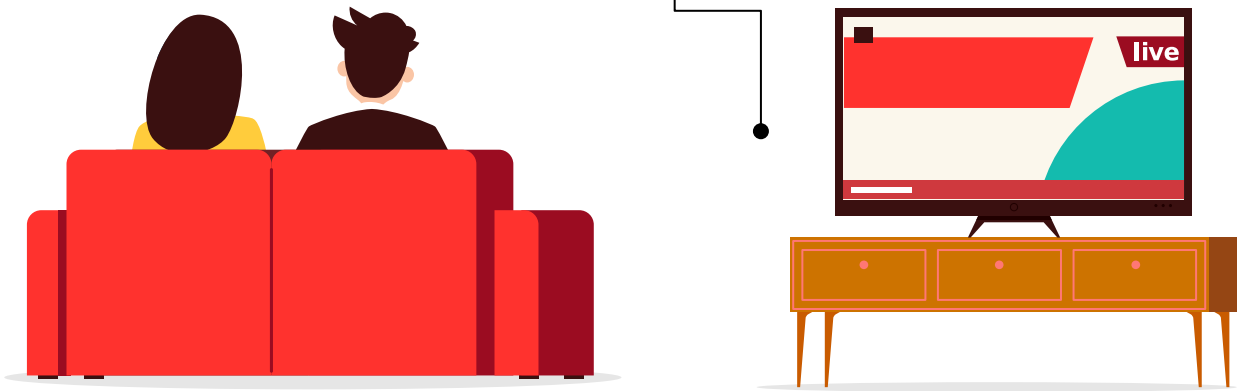
3

Conclusion & References

4

Data Science Question

How can we **quantify** the **belief level** of fake news on social media platforms?



How can we quantify belief levels on social media platforms:

Why is this question important?

1

News on Social Media

Most **popular** news platform in the U.S. for people aged **18-64.**

2

Evaluate Potential Impact

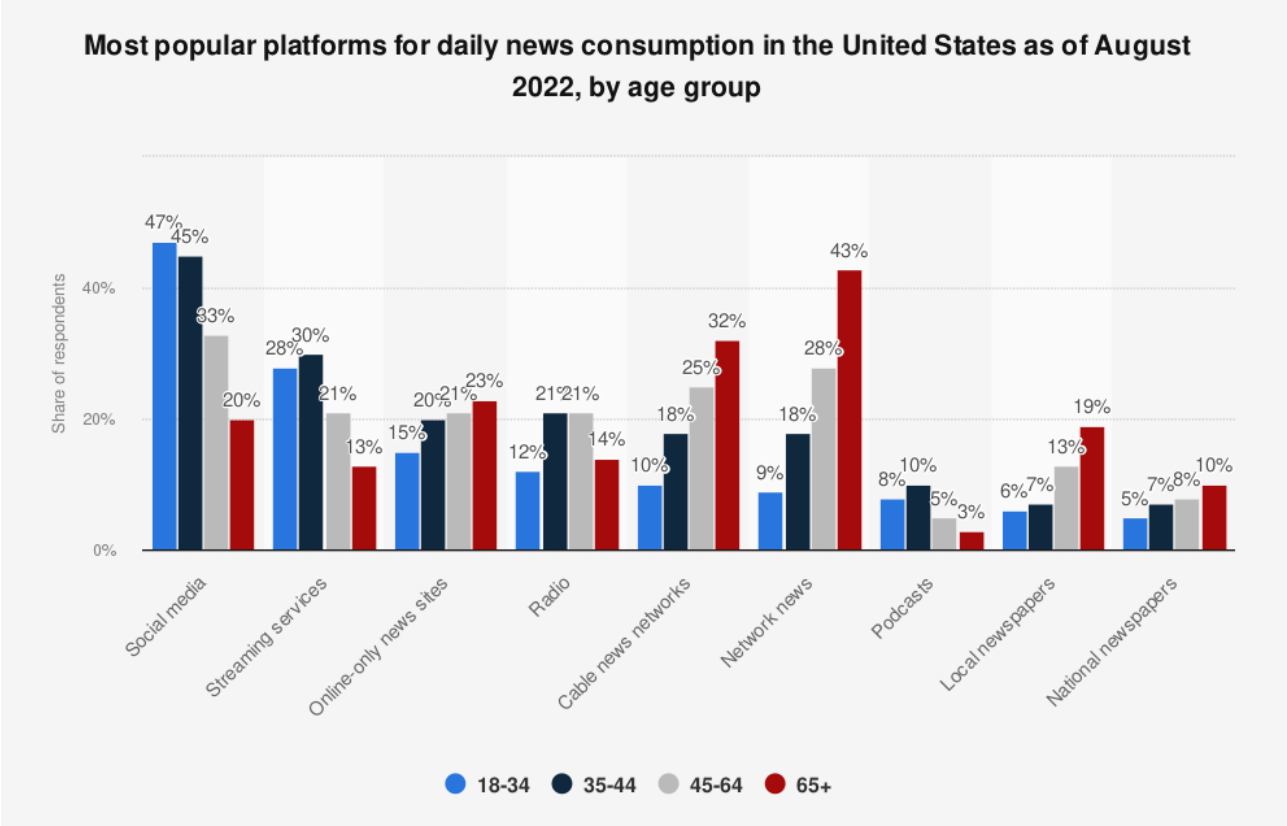
By understanding the spread on platforms, and public perception.

3

Evaluate Success of efforts

e.g. By comparing belief levels before and after.

Why is this question important?



How can we quantify belief levels on social media platforms:

How to answer this question?

Likes? Reposts? Comments?

Is the number of likes or shares an indicator of belief level? Which is more reliable?

No. of fake news?

Demand and supply. Is ‘higher belief level’ detected on these platforms a reason for the higher number of fake news?

Comments?

Positive and negative

Difficulties & Limitations

Limited access to data of social media platforms
- Second hand data used
- Manually collected some data

Shares & Likes

Belief? Exposure?
- Hard to distinguish between genuine belief and mere exposure

Shares, Likes, &Comments

Collecting data
- Manually collected data
- Small sample size

No. of fake news

Other factors
- Monetary incentives
- Ease of content creation

How can we quantify belief levels on social media platforms:

By the Number of Fake News

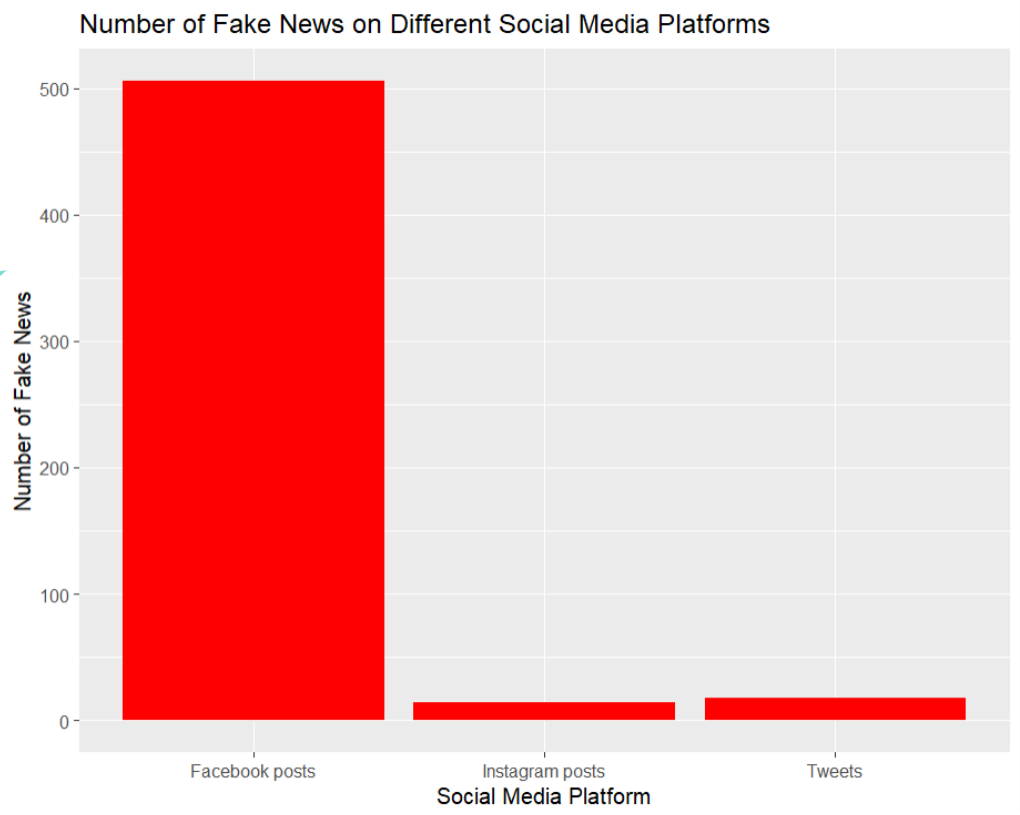
Method

```
library(tidyverse)
library(dplyr)
library(ggplot2)

dat <- read_csv("C:\\Users\\luaxu\\Desktop\\New Task.csv")
names(dat)
dat
unique(dat$Label)

filtered_dat <- dat |>
  filter(Label %in% c('pants-fire', 'FALSE') & Source %in% c("Facebook posts",
                                                             "Tweets",
                                                             "TikTok posts",
                                                             "Youtube videos",
                                                             "Instagram posts"))


platform_counts <- filtered_dat |>
  group_by(Source) |>
  summarize(Count = n())

ggplot(platform_counts, aes(x = Source, y = Count)) +
  geom_bar(stat = 'identity', fill = 'red') +
  labs(title = "Number of Fake News on Different Social Media Platforms",
       x = "Social Media Platform",
       y = "Number of Fake News")
```

Filter desired sources

How can we quantify belief levels on social media platforms:

By the Number of Fake News



Results

How can we quantify belief levels on social media platforms:

By the Number of Shares, Likes, & Comments

Indicators

Belief indicators

Quantifying number of shares and likes as indicators of belief

Assumption

Sharing/Liking == Belief

Individuals who engage with these posts are more likely to belief it

Limitations

Not always the case

- Engagement does not always equal to belief
- Might be sharing to criticize it

How can we quantify belief levels on social media platforms:

By the Number of Shares, Likes, & Comments

Method

POLITIFACT

Setting on Apple's new Journal app "lets anyone near you know your FULL NAME and EXACTLY where you're geo-located."

By Jeff Cercone • March 4, 2024

	A	B	C	D
1	Comments	Shares	Likes	Followers
2	560	2000	1500	19000
3	8	7	65	362
4	2900	1400	7900	132000
5	520	144	1800	1900000
6	1600	165	2200	1800000
7	10	17	29	472000
8	1400	6400	17000	1100000
9	5	20000	608	
10	547	301	2800	3600000
11	756	1200	5900	838000
12	37	1000	69	781
13	77	435	311	691
14	86	227	637	16000
15	680	2800	6300	965000
16	98	131	345	500000
17	180	6900	723	
18	41	160	170	16000
19	64	96	6200	128
20	23	59	302	

Social media only

Collecting data

How can we quantify belief levels on social media platforms:

By the Number of Shares, Likes, & Comments

Method

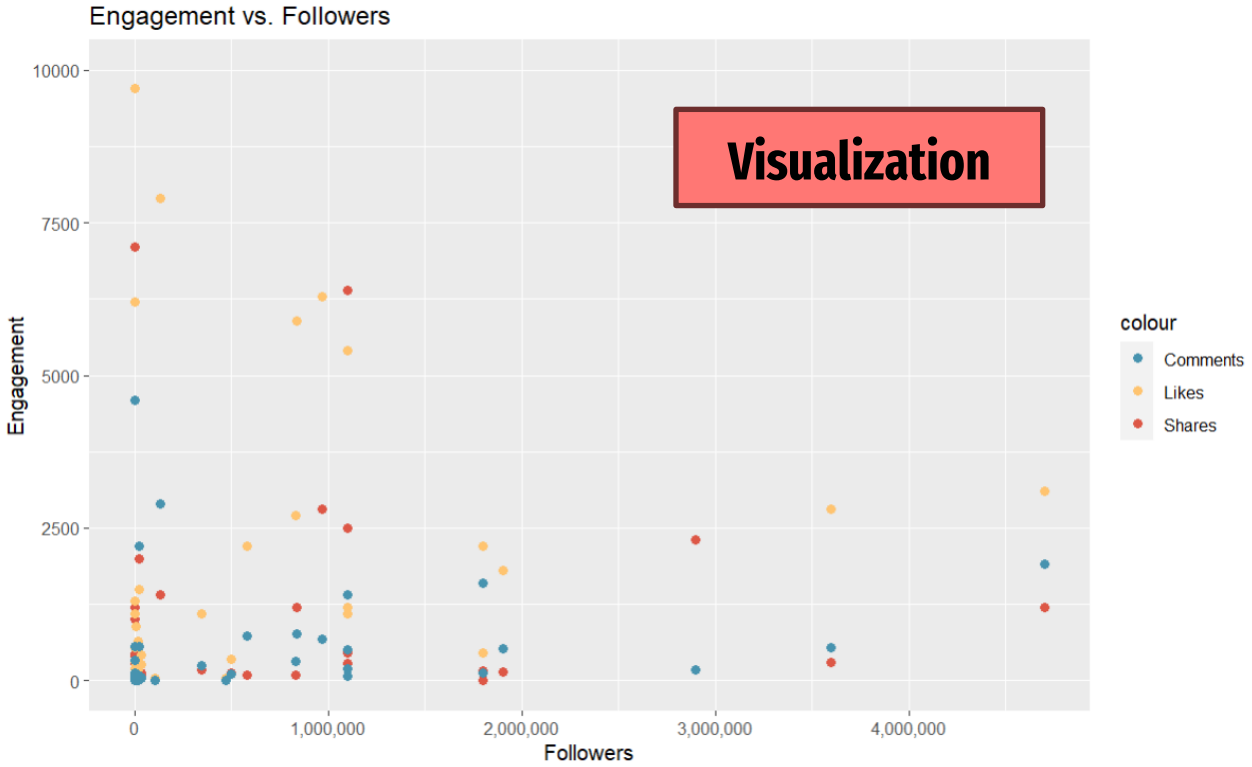
```
library(ggplot2)
library(scales)

data <- read_csv("C:\\Users\\luaxu\\Desktop\\Social media only.csv")

data <- na.omit(data[c("Shares", "Likes", "Comments", "Followers")])

x_limit <- c(0, 1e+06)
y_limit <- c(0, 10000)

ggplot(data, aes(x = Followers)) +
  geom_point(aes(y = Shares, color = "Shares"), size = 2) +
  geom_point(aes(y = Likes, color = "Likes"), size = 2) +
  geom_point(aes(y = Comments, color = "Comments"), size = 2) +
  labs(x = "Followers", y = "Engagement") +
  scale_color_manual(values = c("Shares" = "#DD5746", "Likes" = "#FFC470", "Comments" = "#4793AF")) +
  ggtitle("Engagement vs. Followers") +
  xlim(x_limit) +
  ylim(y_limit) +
  scale_x_continuous(labels = comma)
```

Removed NA values (posts from individual private accounts

How can we quantify belief levels on social media platforms:

By the Number of Shares, Likes, & Comments

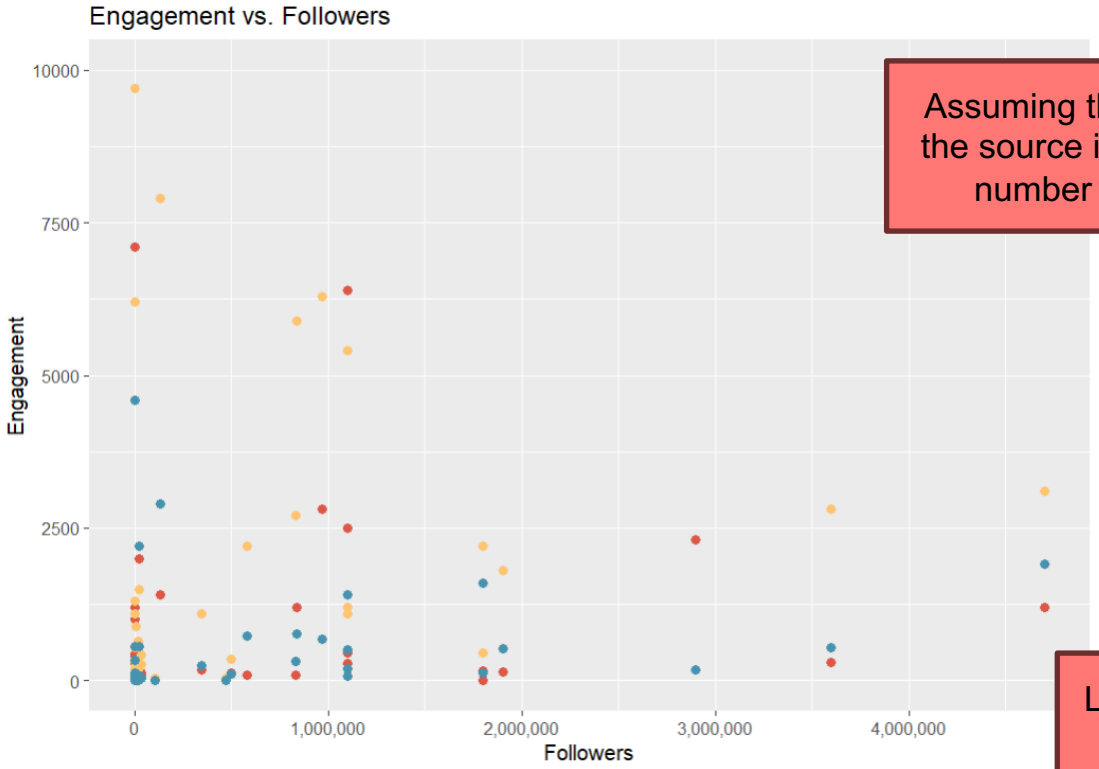
Results



How can we quantify belief levels on social media platforms:

By the Number of Shares, Likes, & Comments

Findings



Assuming the credibility of the source is related to the number of followers

Limitations for likes: Unable to consider different ‘like’ emojis in Facebook

Which is more reliable? Shares? Or Likes?

Endorsement from others

people who have liked the post has remained (Leskin, 2019). Research has found that, in online environments, even if the source of the content is unknown, endorsements from others can overcome people's initial skepticism about that source (Metzger & Flanagin, 2013). In fact, social endorsements may be an even more powerful determinant of news consumption than source cues (Messing & Westwood, 2014).

Reliability?

Proof

liked the post. After concerns over the impact of the "like" feature on users' mental health, Instagram has recently started to test hiding the number of likes on a post in certain countries. The ability to see the usernames of

Future works and directions

Surveys	Research	Data	Other factors	Interventions
To better understand belief levels	Sentiment analysis for comments	Better collection of data with access to APIs	Consider the factors influencing ‘belief’	To analyze the impact of future interventions

References

1. Mena, P., Barbe, D., & Chan-Olmsted, S. (2020). Misinformation on Instagram: The Impact of Trusted Endorsements on Message Credibility. *SageJournals.* https://doi.org/10.1177/205630512093
2. PolitiFact. (2024). https://www.politifact.com/
3. Watson, A. (2022). Statistica. https://www.statista.com/statistics/717651/most-popular-news-platforms/
4. Yadav, K. (2020). https://www.kaggle.com/datasets/techykajal/fakereal-news

THANK YOU