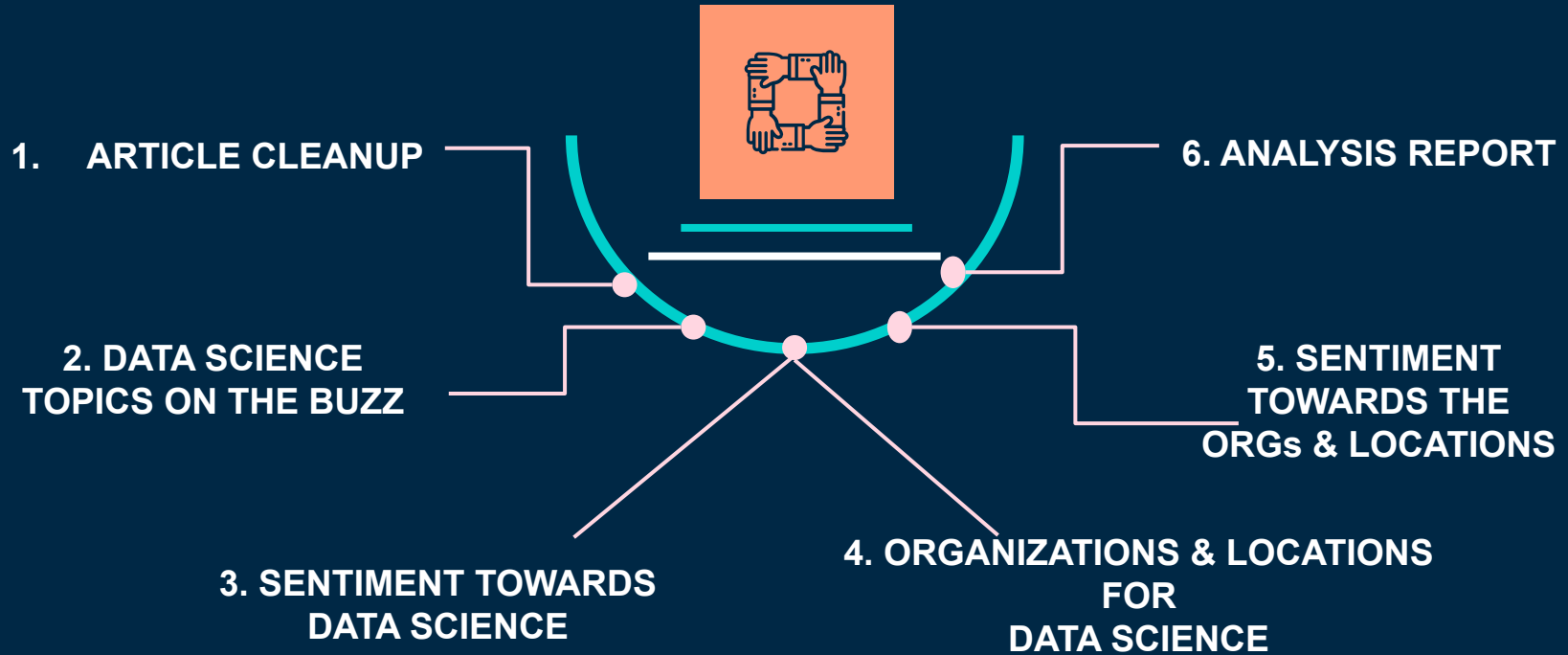




# DATA SCIENCE CAPABILITY ANALYSIS

Snigda Gedela

# CHAPTERS



# ARTICLE CLEANUP

## Feature Used

- ❑ Text
- ❑ The text also has **title** included in it.
- ❑ It will give **more context** and information about the news.

## NOISE CLEANING

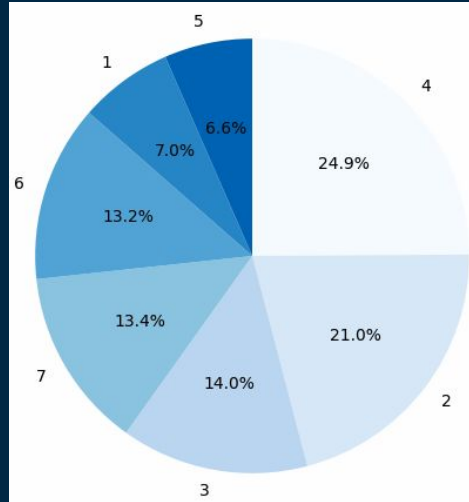
Lowercasing  
↓  
Removing HTML tags, URLs  
and web crawl remnants  
↓  
Removing punctuations, digits  
↓  
Removing symbols and  
non-printable characters  
↓  
Removing newlines, tabs and  
extra white spaces

## PRE-PROCESSING

Removing Stopwords  
↓  
Lemmatization  
using WordNetLemmatizer()

# Data Science Topics on the Buzz

- ❑ Algorithm Used: **LDA**
- ❑ Topic Frequency:

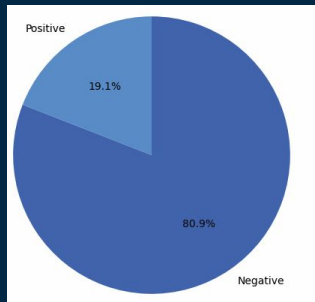


Topic	Description
Topic 1	New findings in Artificial Intelligence and Machine Learning
Topic 2	AI for issues worldwide
Topic 3	AI in product service companies
Topic 4	AI in Health Industry
Topic 5	AI in Video Games Industry
Topic 6	AI and technology used by customer-centric companies
Topic 7	AI and the stock market

# Sentiment Towards Data Science

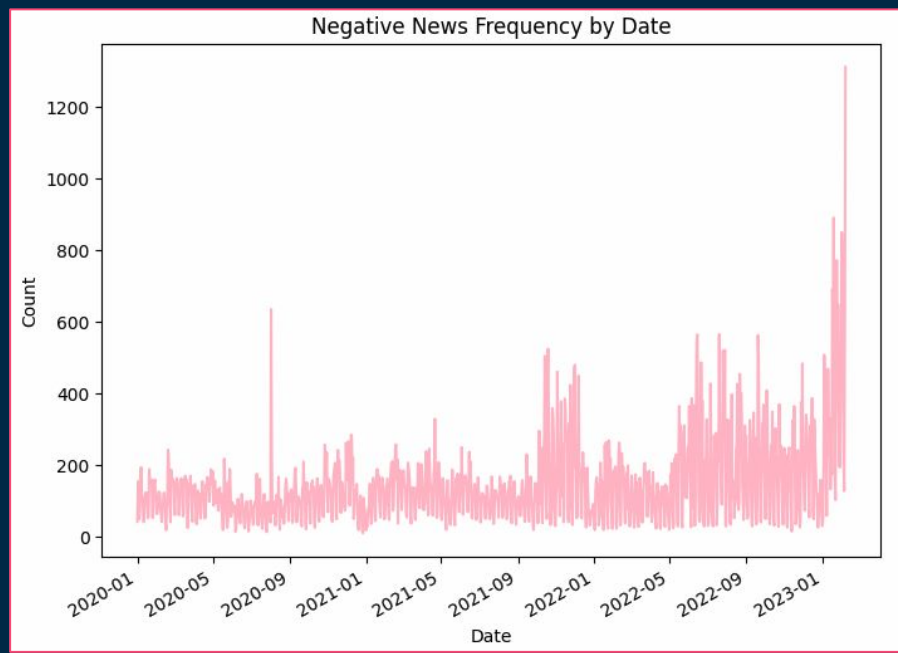
## Overview

- Algorithm Used: **biGRU**
- Sentiment Frequency:



- The negativity seem to have increased over time. But this is due to the fact that the usage of data science has increased too.

## Negative Sentiment over Time

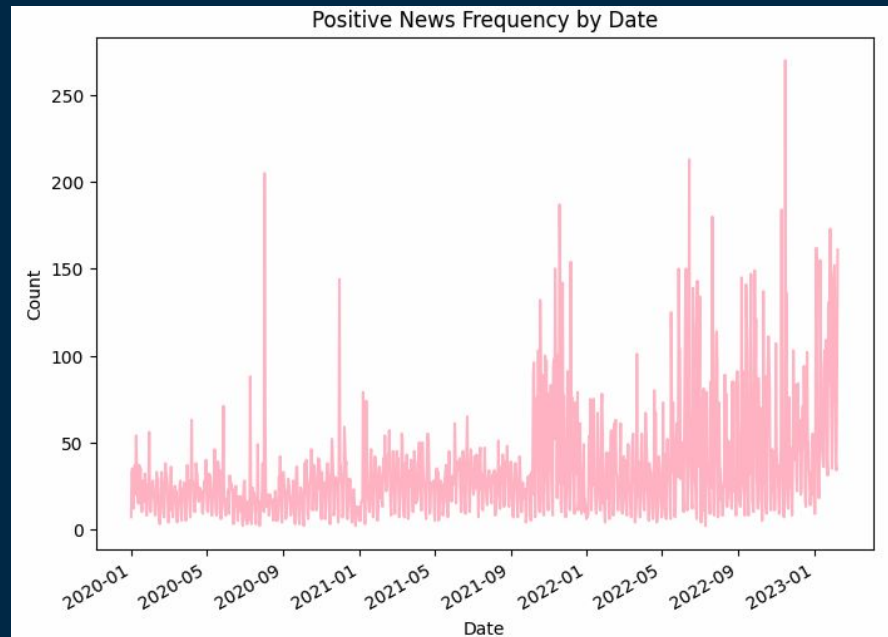


# Sentiment Towards Data Science

## Overview

- ❑ The positivity seem to have increased over time. But this is due to the fact that the usage of data science has increased too.
- ❑ Positivity is observed to be much stronger than negativity towards data science as time passes by; especially towards the last quarter of 2021.

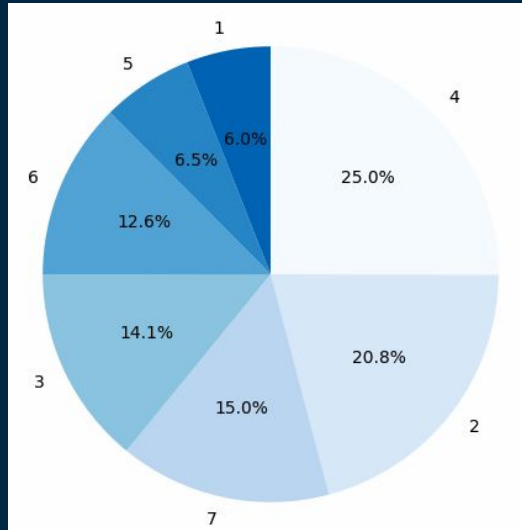
## Positive Sentiment over Time



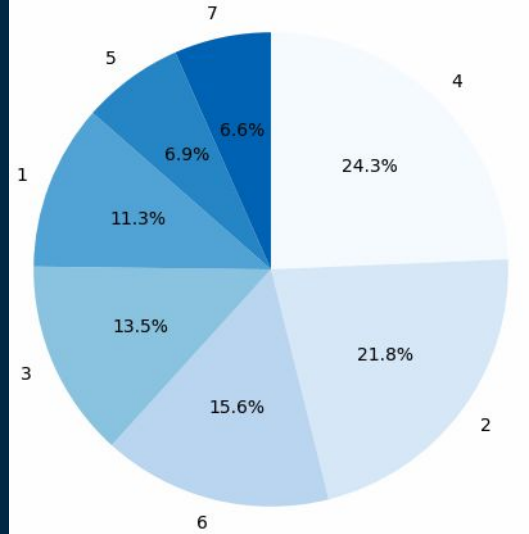
# Sentiment and Topic Analysis

## By Sentiment

Negative News



Positive News



## Observations

- ❑ Topic 4 ( Video Games) has the highest percentage of negativity as well as positivity.
- ❑ As data science is receiving more light day by day, it is not shocking to see that there is different perspectives on every time.
- ❑ Hence, similar levels of positivity and negativity is observed for all the topics.

# Organisations for Data Science

## 1: New Findings

Youtube
OpenAi (ChatGPT)
Samsung
Microsoft
Google

Negative

Samsung
Microsoft
Sony
Google
Youtube

Positive

## 2: Issues Worldwide

Gray Television
-----------------

Gray Television
Gray medium group

## 3: Product Services

FCC
OpenAi (ChatGPT)
Google
Nexstar Media Group

OpenAi (ChatGPT)
Samsung

## 4: Health

Google
IBM
OpenAi (ChatGPT)
Microsoft
NASA

Samsung
Microsoft
Medium LLC
Google
OpenAi (ChatGPT)



# Organisations for Data Science

## 5: Video Gaming

FCC
Wsi

## 6: Customer-centric Companies

Gray Television
-----------------

## 7: Stock Market

IBM
Microsoft
Intel
Samsung
NASA

Negative

Positive

LLC
FCC
Wsi
Google
Youtube

Gray Television
Gray medium group
LinkedIn

iBM
Microsoft
NASA
Intel
Samsung

# Locations for Data Science

## 1: New Findings

New York
India
California
China
Washington

Negative

India
Australia
New York
Japan
Canada

Positive

## 2: Issues Worldwide

Israel
India
Canada
Germany
France

Israel
Germany
India
France
Texas

## 3: Product Services

California
China
New York
Washington
India

New York
China
India
California
London

## 4: Health

India
China
California
Japan
UK

India
China
Japan
Australia
Singapore

# Locations for Data Science

## 5: Video Gaming

Doylestown
Allenstown
China
New York
Pennsylvania

Doylestown
Allenstown
China
Japan
California

## 6: Customer-centric Companies

Japan
China
Israel
India
California

India
China
Japan
Singapore
London

## 7: Stock Market

Japan
China
India
Germany
Brazil

Japan
China
India
Germany
France

Negative

Positive

# Sentiment of Organisations/ Locations towards DS

## Organisations

- ❑ Australia
- ❑ California
- ❑ Canada
- ❑ China
- ❑ France
- ❑ Germany
- ❑ New York
- ❑ Israel

## Locations

- ❑ OpenAI (ChatGPT)
- ❑ Google
- ❑ Gray Television
- ❑ Intel
- ❑ Microsoft
- ❑ Samsung
- ❑ IBM
- ❑ NASA
- ❑ Gray Medium
- ❑ FCC

## Recommendation

- ❑ These organizations and the locations have been mostly taken positively about. Hence, working on a data science project in these organisations and at these locations would increase the probability of the success of your project.

# ANALYSIS REPORT

- ❑ The negativity seem to have increased over time. But this is due to the fact that the usage of data science has increased too.
- ❑ Topic 4 ( Video Games) has the highest percentage of negativity as well as positivity.
- ❑ As data science is receiving more light day by day, it is not shocking to see that there is different perspectives on every time.
- ❑ Hence, similar levels of positivity and negativity is observed for all the topics.
- ❑ It is recommended to focus on Artificial Intelligence in product and customer services, health industry, video game industry and stock market.
- ❑ Youtube, OpenAI, Google, Microsoft and Samsung are investing their resources on innovations in Artificial Intelligence.
- ❑ Overall, India, China, Japan, America and Canada are highly active in the field of Data Science.
- ❑ Israel is especially mostly working towards resolving worldwide issues using Artificial Intelligence.
- ❑ Singapore is highly focussed towards utilising Artificial Intelligence in improving healthcare whereas UK is highly against the concept of using Artificial Intelligence in healthcare.
- ❑ America, especially Pennsylvania ( Doylestown , Allentown ) is highly involved in using AI in the video gaming industry.

# Factors for Successful Data Science Projects

## Recommendations

- ❑ Industries and locations (respectively) to focus on :
  - ❖ Product Services : London, India, New York
  - ❖ Customer-centric services : Singapore, London, India
  - ❖ Health Industry : Australia, Singapore, India, China
  - ❖ Video Gaming : Pennsylvania, Japan, California
  - ❖ Stock Market : France, Germany India
  
- ❑ Companies you should approach for sponsorship with innovative ideas in Data Science:
  - ❖ OpenAI
  - ❖ Google
  - ❖ Youtube
  - ❖ Samsung
  - ❖ Sony
  - ❖ This is because these companies are active in new findings and also in most of the other buzzing topics in Data Science.

**Note:** The above recommendations have been made using the NER, sentiment analysis and Topic modelling. Selected locations and organisations that have no or less negativity/ more positivity.

THANK YOU!