



BrainStation

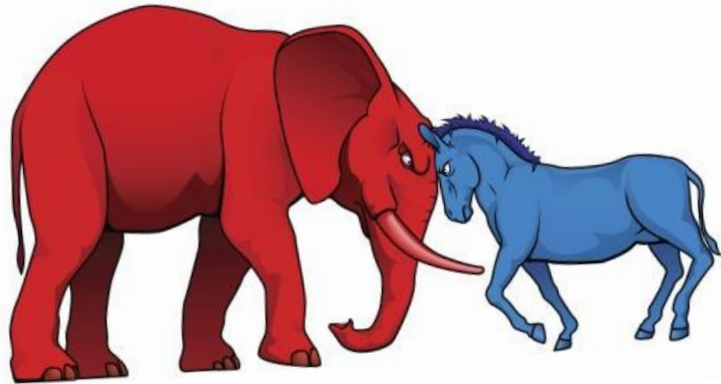
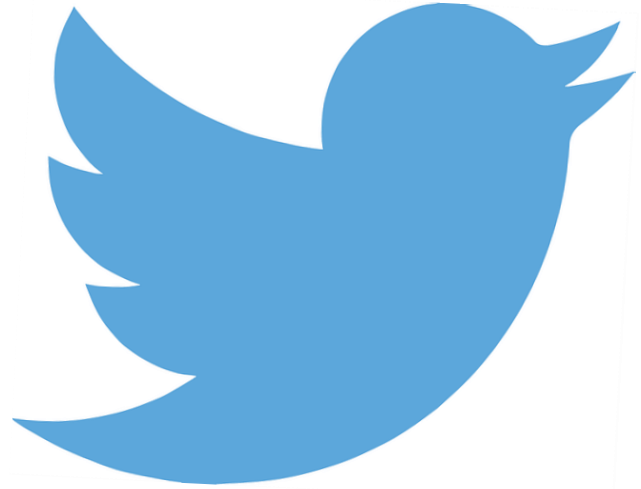
Twitter Sentiment Analysis

Sentiment and Clustering Analysis on Political Tweets

Democrat vs Republican Party

Erick Rudelman | BrainStation

Sprint 1 – Jan 3, 2024





BrainStation

Subject Area & Problem Statement

Twitter reaches 12.6% of internet users worldwide.

Twitter has over 500M users monthly.

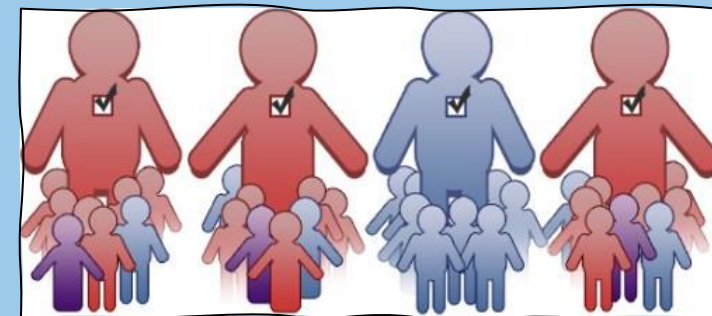
78% of Twitter users use it for news.

Map what people are talking about and understand the stories behind it on a political context.

Understand public sentiment.

Identify relevant cluster for these authors and conversations.

These numbers show just how big of an impact Twitter has on our conversations and perspectives.





BrainStation

Data Science Solution

Web Scraping

Exploratory Data Analysis

Natural Language Processing

- Basic text data processing
- Advanced text data processing

Clustering Algorithms

Logistic Regression

Data Visualization



BrainStation

Impact of Solution

Offer a nuanced understanding of public sentiment and discourse on Twitter.

Uncover patterns, trends, and sentiment dynamics within the digital community.

Informed decision-making.

- Businesses
- Politicians refining communication strategies
- Individuals or groups aiming for enhanced community engagement.

Academic advancement in social media studies, offering a comprehensive exploration of online conversations.

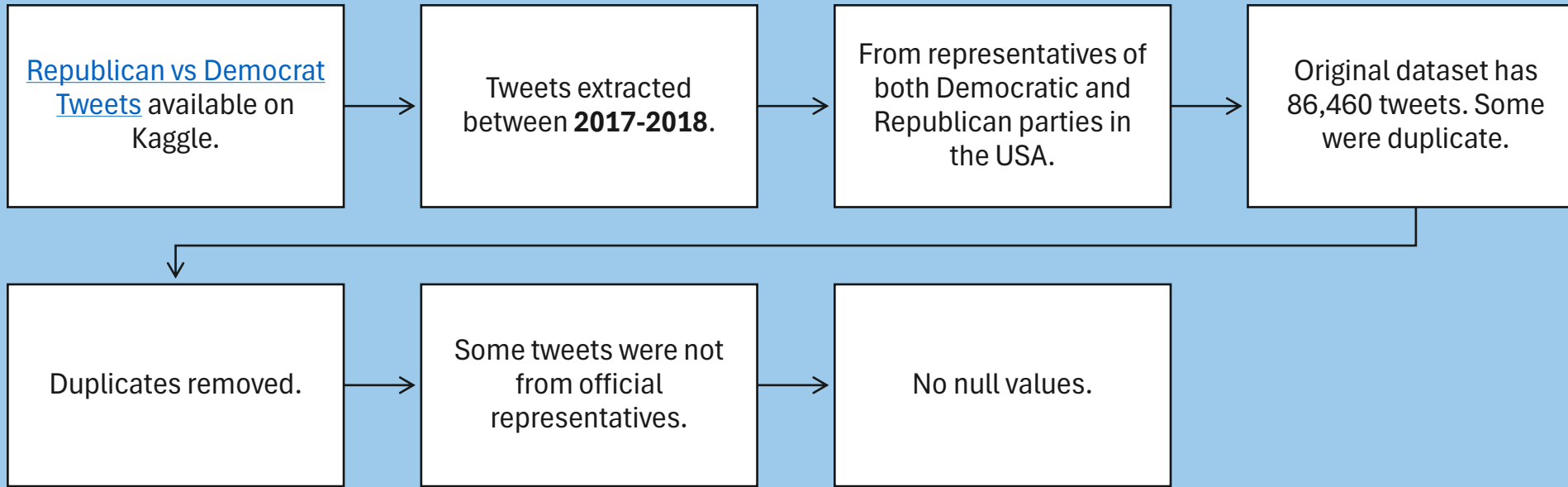
Empower stakeholders.



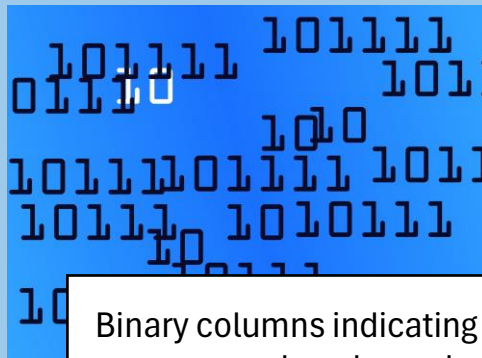


BrainStation

Dataset: EDA



Dataset: New Features



Binary columns indicating if it was a retweet or not.



Hashtag columns highlighting all hashtags for each tweet.



Created a list of all the hashtags for further analysis.





BrainStation

Data Processing: Next Steps

Basic features from text data

- ☐ Count of Words
- ☐ Removing Stopwords
- ☐ Extracting number of special characters
- ☐ Identifying uppercase words

Advanced features from text data

- ☐ CountVectorization
- ☐ Hashing Vectorization
- ☐ Stemming
- ☐ Lemmatization
- ☐ N-grams
- ☐ Term Frequency
- ☐ Sentiment Analysis

