

# Brexit Twitter Data Analysis



# Data Preprocessing Steps

# 1. Data Transformation

- Reading Rows as string using line function
- Eliminating Comma delimiter
- Removing Unnecessary Characters from line
- Slicing the strings to store in attributes

```
date_time = line[line.find("{created_at") + 12 : line.find("id:")]
index = line.index(date_time)
line = line [index : :]
```

# 2. Data Cleaning

## RegEX for cleaning the tweet text

```
def remove_usernames(text):
    return re.sub('(?<=^|(?<=[^a-zA-Z0-9-_\.]))@([A-Za-z]+[A-Za-z0-9-_]+)', '', text)

def remove_hyperlinks(text):
    return re.sub("((http|https)\:\/\/)?[a-zA-Z0-9\.\\/\?\:@\-_#=]+\.([a-zA-Z]{2,6})[a-zA-Z0-9-_#=]+", "", text)

def remove_punctuation(text):
    return re.sub("[^a-zA-Z]+", " ", str(text))

def to_lower_case(text):
    return text.lower()
```

```
from nltk.stem import PorterStemmer  
porter_stemmer=PorterStemmer()  
def stemming(text):  
    words=re.split("\s+",text)  
    stemmed_words=[porter_stemmer.stem(word=word) for word in words]  
    return ' '.join(stemmed_words)
```

- **Stemming the words in tweet text**
- Convert "string null" to **numpy.NaN**

### 3. Data Reduction

- Dropped attributes with 80% NULL values
- Removed the duplicates of tweet\_id

# Eploratory Data Analysis

Counting the occurence of famous words in **TWITTR\_TEXT**

BREXIT_COUNT
31400

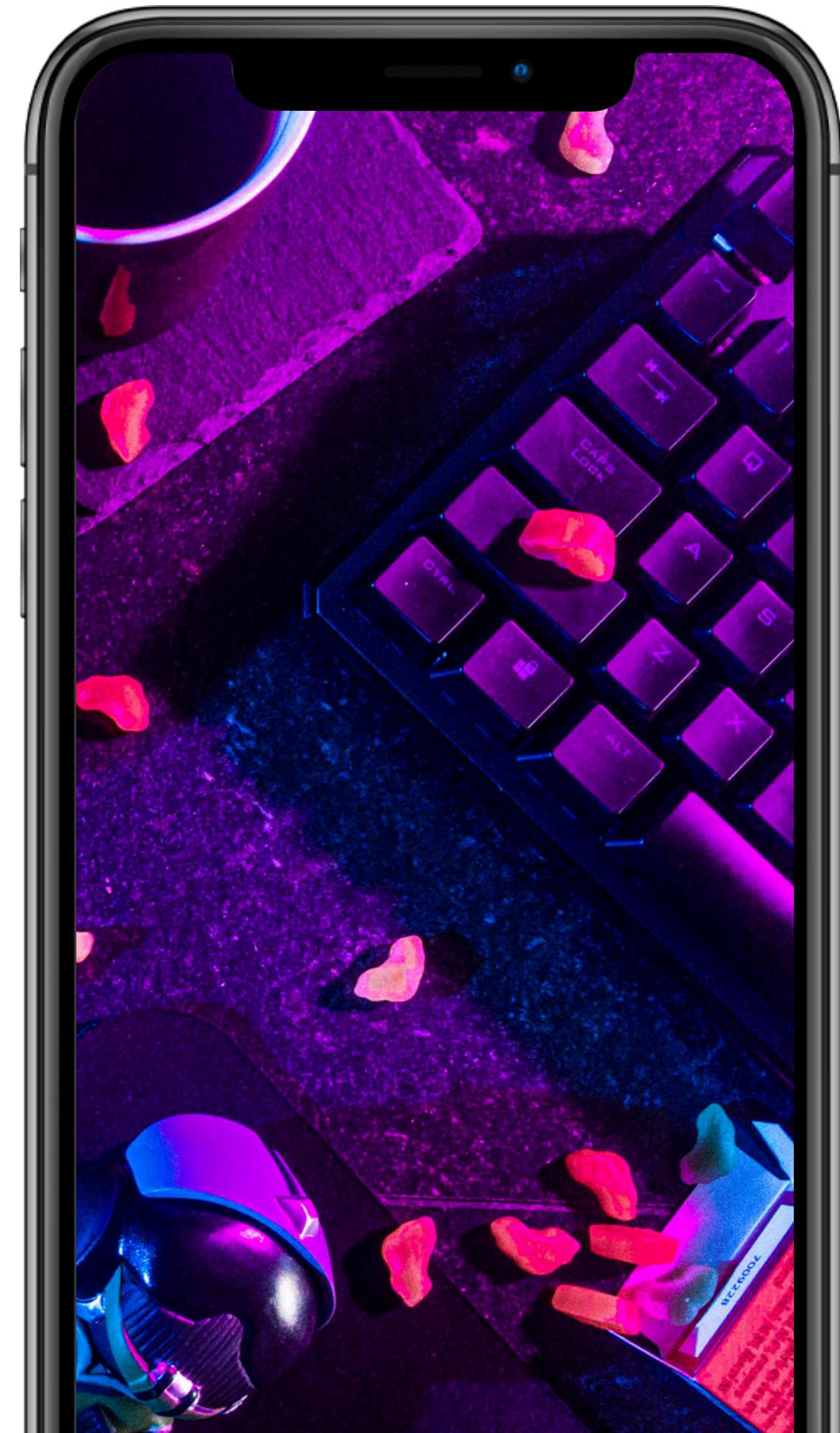
LABOUR_PARTY_COUNT
1752

UK_word_COUNT
11324

CORBYN_COUNT
1133

LeaveEU_COUNT
4602

VOTE_WORD_COUNT
9748



09

# IoT Frameworks

Presentations are communication tools that can be used as demonstrations, lectures, speeches, reports, and more. It is mostly presented before an audience. It serves a variety of purposes, making presentations powerful tools for convincing and teaching.

10

# The Challenges of IoT





# Privacy

Presentations are communication tools that can be used as demonstrations, lectures, speeches, reports, and more.

Presentations are communication tools that can be used as demonstrations, lectures, speeches, reports, and more.

Presentations are communication tools that can be used as demonstrations, lectures, speeches, reports, and more.

# Data storage and analysis

Presentations are communication tools that can be used as demonstrations, lectures, speeches, reports, and more. It is mostly presented before an audience. It serves a variety of purposes, making presentations powerful tools for convincing and teaching.





# The Future of IoT

Where to next?

Presentations are communication tools that can be used as lectures, speeches, reports, and more. It is mostly presented before an audience.

14

# \$1.1 trillion

Presentations are communication tools that can be used as lectures, speeches, reports, and more. It is mostly presented before an audience.





Presentations are communication tools that can be used as demonstrations, lectures, speeches, reports, and more. It is mostly presented before an audience. It serves a variety of purposes, making presentations powerful tools for convincing and teaching.