CMP7202 — Web Social Media Analytics and Visualisation

Modern Social Media Intelligence

Web-Driven Analytics and Insights

Alexander Roberts

M.Sc. Big Data Analytics

Submitted: May 2025

Word count: 2,190

BIRMINGHAM CITY University

School of Computing and Digital Technology

Faculty of Computing, Engineering and the Built Environment

Birmingham City University

Contents

List of Tables

List of Figures

1 Introduction

Social media has become a cornerstone in society for communication and information sharing. According to Statista (2025), in February 2025, there were over 5.56 billion persons who use the Internet, of whom 5.24 billion used some form of social media. Hence, a tremendous amount of data and information is available through social media. This report will investigate how such data can be analysed and transformed to uncover new insights into trends and patterns.

1.1 Twitter Datasets Background

Two Twitter datasets have been identified for social network analysis, a tabular dataset by Naghshzan (2023), and a social network graph by Leskovec (2012), hosted on the Stanford Network Analysis Project (SNAP). The former contains a set of real-world tweets posted between 1st January 2022 and 22nd December 2022. It contains over 500,000 tweets. The data was gathered using web scrapping techniques, which was likely verified for its integrity, but this could imply that errors could exist (e.g., missing entries or invalid text) within the dataset. Additionally, the author could have purposefully or accidentally inputted bias into the dataset, but this would seem unlikely as there is no motive for influencing this data, nor is there any obvious evidence of this. For the latter dataset regarding the social network graph, the data was collected from various real-world Twitter sources in 2012, and then subsequently combined to form a large dataset. Its author, Jure Leskovec, is a professor at Stanford University. Given their academic background, it would seem likely they would understand the importance of maintaining the data's integrity by adding as little author bias as possible. Furthermore, Stanford has also sponsored the content of SNAP, helping authenticate the content's reliability.

2 Statistical Analysis of Social Media Data

All code will be available in a separate Jupyter Notebook that will accompany this report.

2.1 Tabular Twitter Dataset

2.1.1 Setup and initial data preprocessing

Once the data had been downloaded from Naghshzan (2023), any duplicate entries were removed. As indicated by the initial view of the dataset in Figure 2.1, there are missing values within the dataset. These should not be removed yet as it could mean valuable insights from other attributes are lost during the EDA. Instead, this should be done ad-hoc using entire row removal for complete case analysis. Further data preprocessing may be required for subsequent EDA steps.

date	content	likeCount	replyCount	retweetCount	viewCount	quoteCount	sourceLabel	links	...	vibe	username	UserDescripti
2022-01-01 23:59:59+00:00	I'm getting a sugar daddy this year!	0.0	10.0	0	210.006559	0	Twitter for iPhone	NaN	...	NaN	LaTera__	NCAT Alumr || For gigg
2022-01-01 23:59:59+00:00	might make em french toast after and spoon a l...	1.0	0.0	0	191.910011	0	Twitter for iPhone	NaN	...	NaN	allirica_rose	writer, read horror lov gene disaste
2022-01-01 23:59:59+00:00	Our platform can be matched to your company's ...	0.0	0.0	0	210.006559	0	HubSpot	[TextLink(text='hubs.ly/Q011g-C00', url='https...	...	NaN	matrixlms	An innovat way to train a learn. \nSign
2022-01-01 23:59:59+00:00	Oh my God this first song. I can't with how am...	5.0	2.0	0	542.144594	0	Twitter for Android	[TextLink(text='youtu.be/tA0-WXsm-T0', url='ht...	...	NaN	AnarkyIsMe	Vegan activ trying to mak change in the
2022-01-01 23:59:59+00:00	Happy new year ! 🥳🎉	1.0	0.0	0	191.910011	0	Twitter for iPhone	NaN	...	NaN	__amyya	5♥5

Figure 2.1: Initial view of the tabular Twitter dataset.

2.1.2 Hashtag trends

Analysing Twitter hashtag trends can provide insight into trending topics (during 2022). This is useful for user sentiment analysis and finding key topic trends. To visualise this, each hashtag in every tweet was counted, with the resulting counts being measured

against each tweet hashtag. Certain hashtags, such as 'nft', 'nfts', 'nftart', and 'nft-collector', were combined into one hashtag given they relate so closely to each other. Figure 2.2 illustrates the counts of the hashtags. As shown in the figure, non-fungible tokens (NFTs) were the most popular trend at the time, with other trends involving cryptocurrencies and Bitcoin also being popular at the time. This makes sense, given the significant rise in interest of NFTs in the past 5 years (Hammi, Zeadally and Perez, 2023, p. 46).

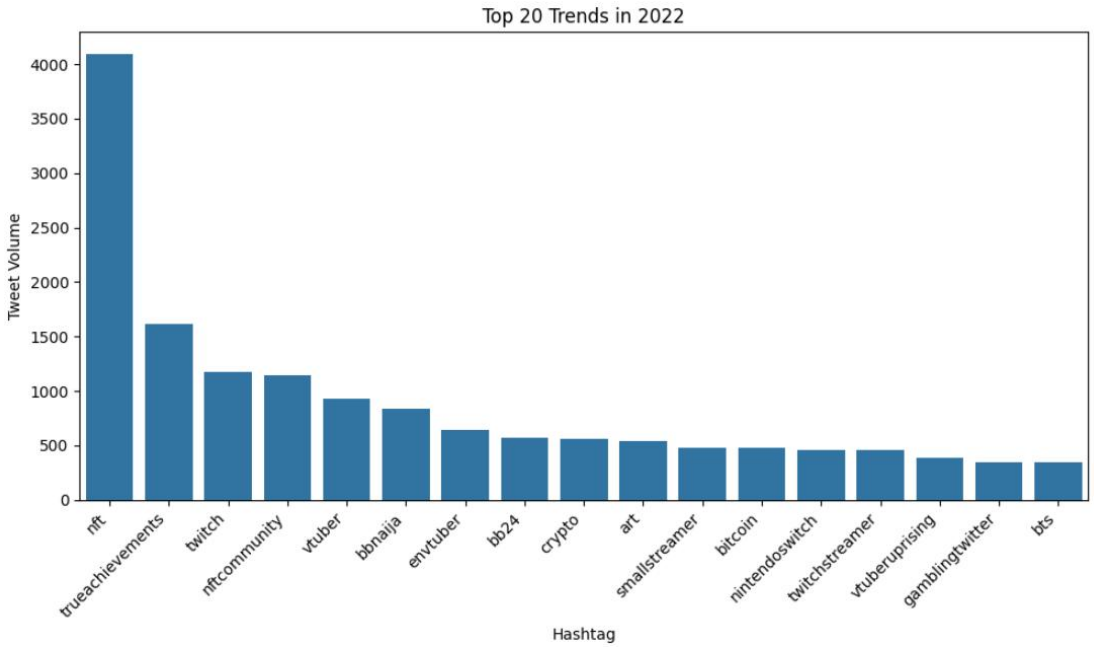

Figure 2.2: Top hashtag trends on Twitter in 2022.

2.1.3 Geographical analysis

The number of hashtags posted by a country can provide an idea of which country populations most frequently use Twitter. This is useful for identifying cultural trends, and targeted advertisements. To achieve this, each city / location was converted into their respective country by using the 'pycountry' Python module, along with some manual name adjustments. Then, each instance of each country was counted using aggregation, which formed a 'hashtag_count' by country.

Figure 2.3 shows the top 15 countries with the highest number of Twitter hashtags

posted. It is clear that the top three counties, the United States, Canada, and the United Kingdom, have the highest number of hashtags. This could be due to numerous factors, including how with larger populations, there is a higher chance of more persons using the service compared to smaller country populations. Another reason could also include how Twitter is an American company, so its only natural its sphere of influence would be the highest in the United States. Finally, their high hashtag count could be due to close shared cultural relationships, as all three Western countries have historically had close ties to one another (Hordiienko and Khamula, 2020, p. 10) (Norton and Horton, 2023, pp. 73-78). Figure 2.4 shows a heatmap of these hashtag counts, which was done by locating base coordinates for each country, which were then used to make the plot on a real-world map. It is important to note regarding this visualisation that it appears Europe and the Middle East have a higher concentration of hashtag counts compared to the United States. This is because when European and Middle Eastern countries hashtag counts are combined, they outnumber the United States.

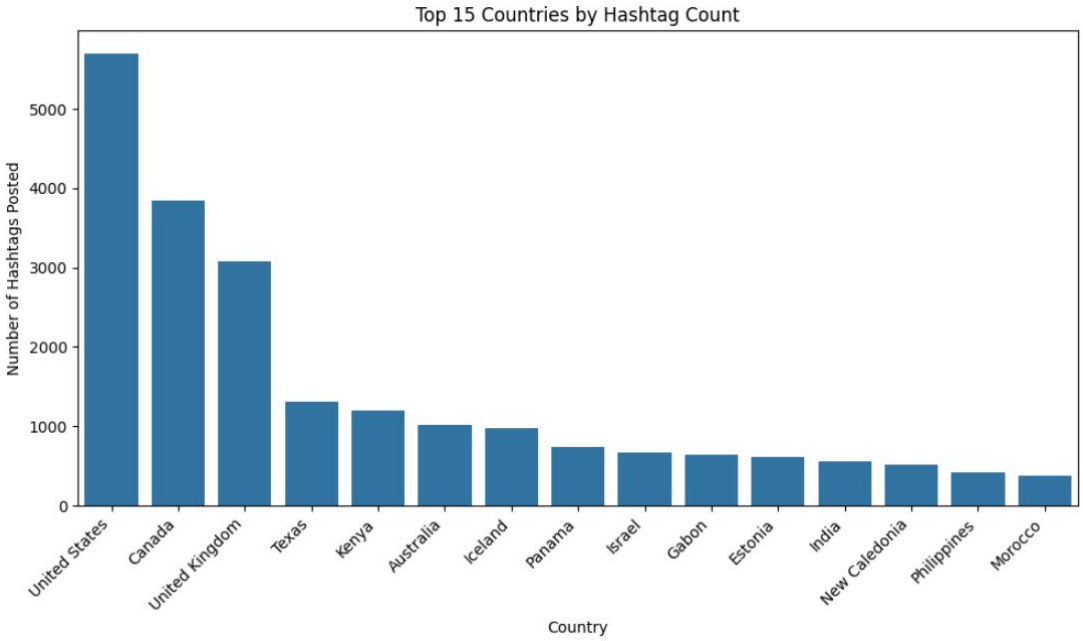

Figure 2.3: Top 15 countries with the highest number of Twitter hashtags between January 2022 to December 2022.

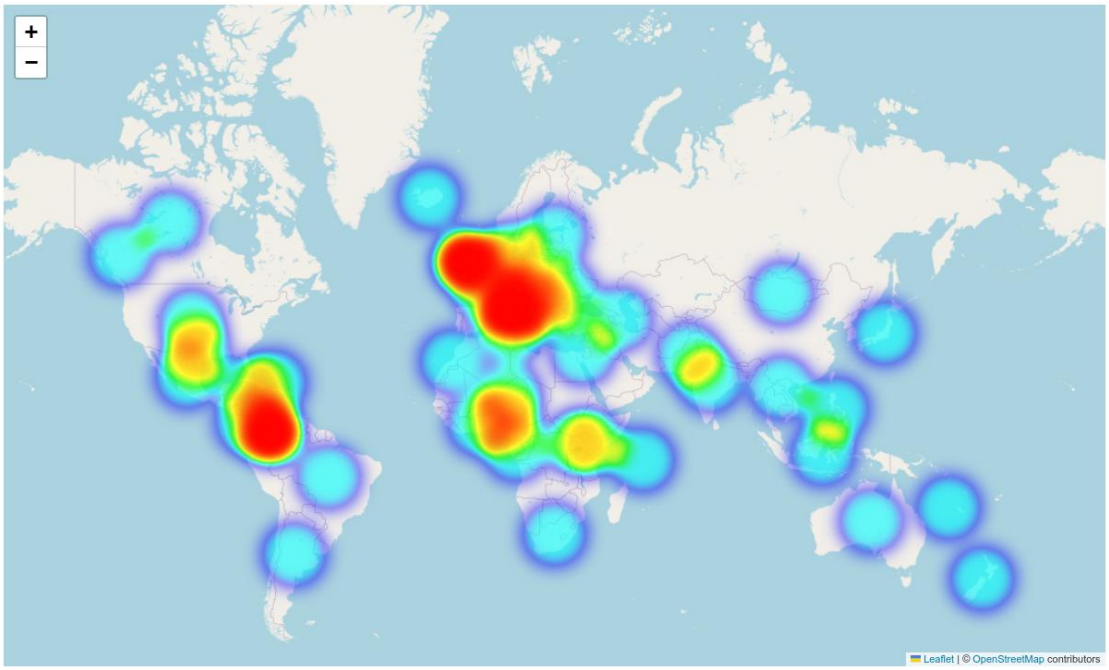

Figure 2.4: Heatmap of the hashtag counts by country, from January 2022 to December 2022.

2.1.4 Device mediums

Figure 2.5 shows the distribution of posts tweeted on several device mediums. Investigating this can indicate what devices Twitter, possibly along with other social media apps, should be optimised for. As indicated in the figure, tweets are mainly posted on iPhone, followed by Android, the online web app, and finally any other mediums (e.g., Google Pixel).

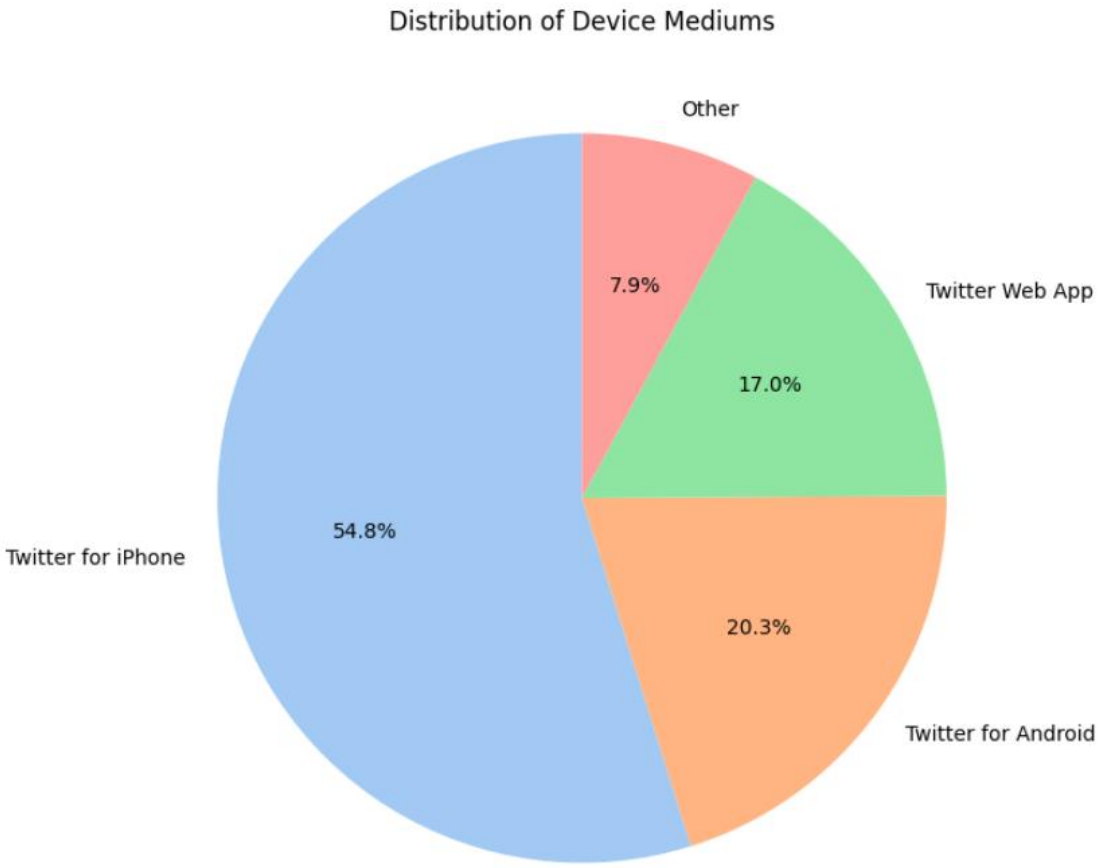

Figure 2.5: Distribution of Twitter posts by device mediums.

2.1.5 Tweet activity

Figure 2.6 shows the number of tweets posted per month in 2022. This is useful for analysing seasonal trends and determining user retention for the platform. As shown in the figure, the volume of tweets posted across the year is fairly consistent across each month. This signifies the continued popularity and usage of the platform. However, this trend could also be due to the data collector purposefully collecting roughly equal amounts of tweets for each month, to mitigate selection bias. For instance, the data collector could have performed random undersampling for selecting the tweets, but specified a minimum number of tweets that should be collected for each month (so that the distribution is not erratic).

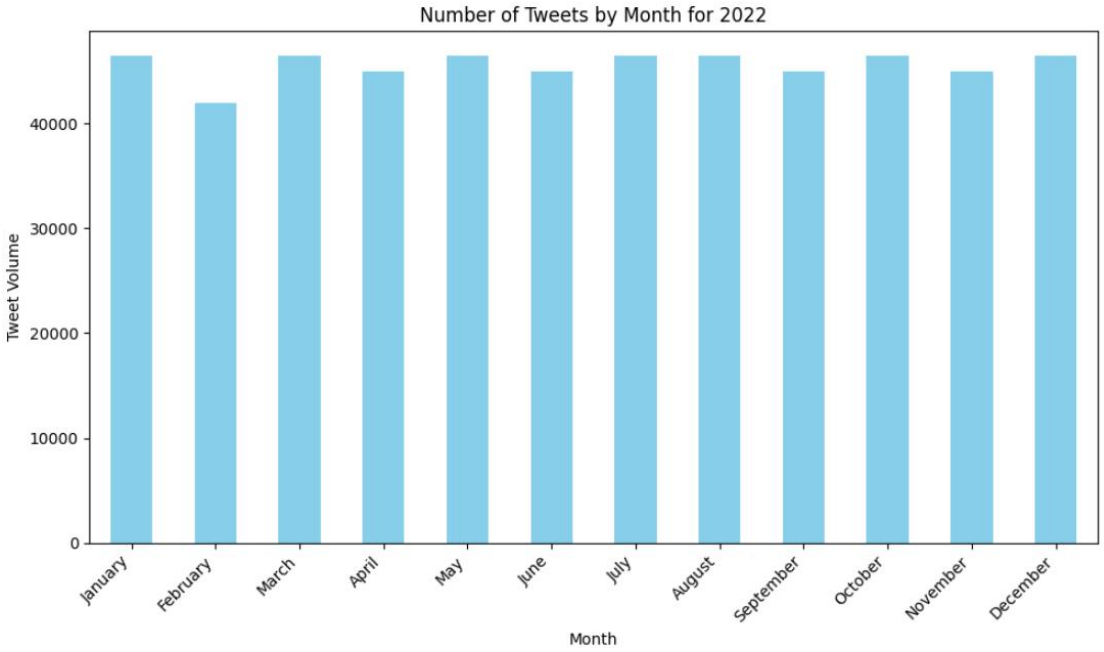

Figure 2.6: Tweet volume by month for 2022.

2.2 Social Graph Network Dataset

2.2.1 Setup

Once the data has been downloaded from Leskovec (2012), it can be loaded into a graph object using the Python module 'NetworkX'. Table 2.1 showing the total number of nodes, edges, and diameter. Unfortunately, for calculating certain metrics, such as the Betweenness Centrality, it can be extremely resource-heavy and time-consuming for larger graphs (due to its polynomial time complexity). To resolve this, a Python package called 'cuGraph' can be leveraged, which utilises graphics processing unit (GPU) parallel processing to accelerate calculations (as shown in Figure 2.7). It is also compatible with 'NetworkX'.

Table 2.1: Basic statistics of the Twitter SNAP graph dataset.

Metric	Value
No. of nodes	81,306
No. of edges	1,768,149
Diameter (longest shortest path)	7

Source: Rees (2020).

Node	Edges	Speedup	NetworkX Runtime (sec)	cuGraph Runtime (sec)
100	1,344	0.45	0.05	0.11
200	2,944	1.14	0.20	0.18
400	6,144	2.64	0.84	0.32
800	12,544	5.26	3.84	0.73
1,600	25,344	12.99	17.51	1.35
3,200	50,944	26.5	78.10	2.95
6,400	102,144	48.62	338.73	6.97
12,800	204,544	89.81	1,539.73	17.14
25,600	409,344	180.42	7,804.06	43.25
51,200	818,944	328.05	47,763.09	145.60

Figure 2.7: NetworkX vs. cuGraph performance comparison with graph scaling.

2.2.2 Investigating centrality

Centrality is a category of methods used to attempt to find the most central nodes within a graph. Some examples, according to Wan et al. (2021, pp. 104778-104786), include:

- degree centrality, which measures how many direct links to other nodes are present for a certain node;
- eigenvector centrality, which acts as a more advanced version of degree centrality, by assigning higher scores to highly connected nodes connected to other highly connected nodes;

- and betweenness centrality, which measures how well each node can act as a "bridge" for nodes attempting to find the shortest path to other nodes.

Figures 2.8, 2.9, and 2.10 show the centrality scores from the aforementioned metrics. Typically, these centrality measures can indicate different nodes for being the most important in the network, mainly due to their specific intended purpose. However, nodes which coincide in each of these metrics can be considered the most important in the network. Figure 2.11 shows the most important node(s) based the above metric. This was achieved by intersecting the top ten groups for each metric, with the resulting node(s) achieving good performance across all three metrics. As shown in Figure 2.11, the node '40981798' is the only node that has achieved well in all three metrics. Therefore, it can be considered the most important node in the graph. It is important to note, while the resulting node can be considered the most important node in the graph in general, for specific use cases, such as what node connects the most diverse groups across the graph, node '813286' would be a far better choice as its betweenness centrality is almost three times higher than node '40981798'.

degree_centrality	vertex
0.083218	115485051
0.079675	40981798
0.074067	813286
0.067843	43003845
0.061251	3359851
0.061103	22462180
0.060906	34428380
0.053010	7861312
0.052469	15913
0.044007	59804598

Figure 2.8: Top 10 nodes by degree centrality.

vertex	eigenvector_centrality
40981798	0.172362
43003845	0.165223
22462180	0.162535
34428380	0.162134
27633075	0.102091
31331740	0.101237
83943787	0.093309
18996905	0.092844
208132323	0.091336
117674417	0.088590

Figure 2.9: Top 10 nodes by eigenvector centrality.

vertex	betweenness_centrality
813286	0.062175
115485051	0.060170
3359851	0.039626
15846407	0.033075
40981798	0.027694
17093617	0.027193
7861312	0.022602
12925072	0.021829
14230524	0.021110
62581962	0.018021

Figure 2.10: Top 10 nodes by betweenness centrality

```
Common node with high scores across all aforementioned metrics,
with respect to each metric: {40981798}
```

Figure 2.11: Most important node(s) of the Twitter SNAP graph.

2.2.3 Visualising degree centrality

Figure 2.12 visualises the top 50 nodes by degree centrality, where the darker shades of blue for a given node represent a higher degree centrality for that node. This is useful for visualising the significance of these metrics, and to help identify some potential visual trends not apparent through the metrics.

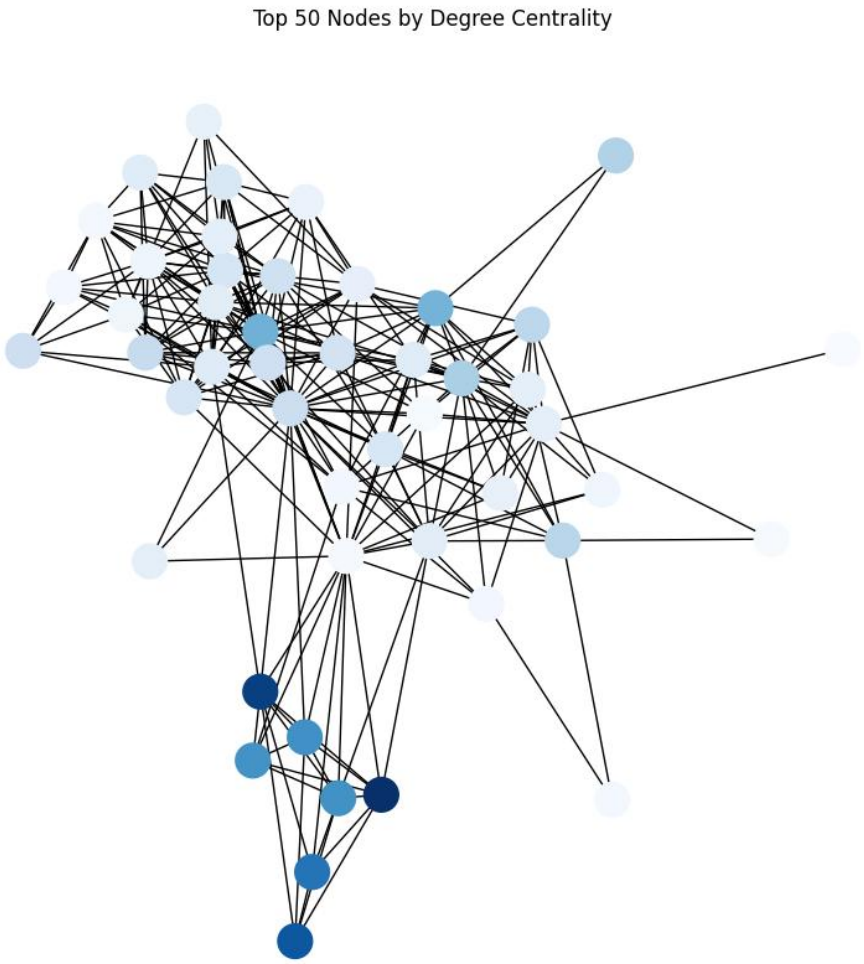

Figure 2.12: Top 100 nodes by degree centrality.

2.2.4 Community detection

Community detection is the process of attempting to identify groups of highly related individuals (that often have high cohesion), also known as communities, in graph networks. This relation(s) between individuals can represent factors such as interests, be-

haviours, and/or trends. This is useful for a range of purposes, particularly for pattern discovery and highlighting key areas for further analysis. Figure ... indicates the results of applying the Louvain community detection algorithm on the Twitter SNAP graph. This algorithm, a type of hierarchical clustering, was chosen namely due to its good performance for larger graph for detecting global communities (Zhang et al., 2021, pp. 1-3). Moreover, this is achieved by attempting to maximise the modularity of each community partition. Figure 2.13 illustrates the results of applying the Louvain detection algorithm on the graph. As indicated in the figure, the final modularity score is 0.8042 (4 s.f.), which suggests the communities detected have a strong and clear separation from other communities, while also having strong intra-community cohesion. Additionally, the difference between the number of unique communities and the maximum number of communities is that the unique communities are the actual meaningful communities identified by the algorithm, while the maximum number of communities is how many communities the algorithm trialled. As the former is significantly less than the latter, this suggests the algorithm has not largely over-segmented the communities. Figure 2.14 shows a sample of nodes from three different communities, and are colourised by the community they are a part of.

```
Modularity Score: 0.8042088747024536

          vertex  partition
0      115485051          0
1       40981798          3
2         813286          0
3       43003845          3
4        3359851          1
...          ...        ...
81301  567272538        142
81302  567685433         10
81303  567915211         60
81304  568655523          3
81305  568753458          4

[81306 rows x 2 columns]

Number of unique communities: 164
Maximum number of communities: 10003
```

Figure 2.13: Results of applying the Louvain community detection algorithm.

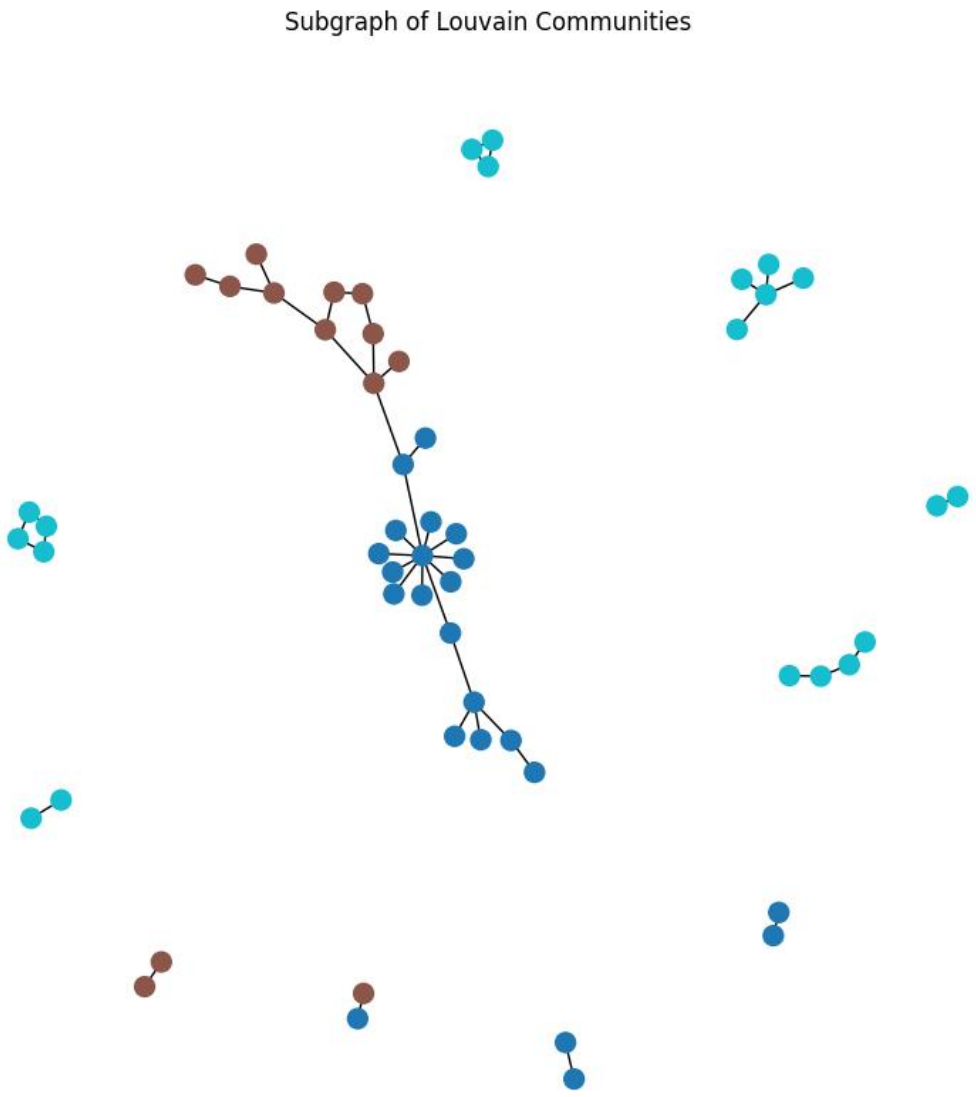

Figure 2.14: Subgraph of three of the detection communities identified by the Louvain algorithm.

3 Text Mining and the Cost of Living Crisis

During the years 2022-2023, the UK and numerous other nations worldwide experienced unprecedented increases in consumer price inflation (CPI) (Office for National Statistics, 2025), resulting in astronomical price hikes for households across industries such as food, energy, services, etc. (Moosa, 2024, pp. 1-4) This period become notoriously known as the Cost of Living Crisis.

3.1 Sentiment Analysis

3.1.1 Dataset background

A different Twitter dataset, by Tleonel (2022), was identified to be useful as it contains 144,000+ real-world tweets specifically about the Cost of Living Crisis, for 20th August-9th September 2022. The data was retrieved using web scrapping. Given there is little information regarding the author, it is difficult to determine how much bias they have applied to the dataset. However, given that there is no obvious motive for purposefully influencing the data negatively, it can be assumed they have input little bias. Any duplicates were removed.

3.1.2 Experimentation

Traditional text sentiment analysis tools leveraged straightforward techniques such as removing stop words, removing punctuation, performing lemmatisation / stemming and tokenisation to attempt to transform variable text into a more useful format. But within recent years, transformers, a special type of deep neural network architecture, has come about. Transformers can perform advanced text analysis, generally much more effectively compared to traditional means — making it useful for more complex text analysis tasks. Both methods have been used on the dataset, as described in Section 3.1.1, for sentiment analysis. 'TextBlob' was used as a more traditional approach, while a transformer model ('cardiffnlp/twitter-roberta-base-sentiment-latest') specifically trained on Twitter general text was also used. Both techniques automatically handle tasks such as removing stop words, adjusting for punctuation, etc. Other data preprocessing steps

included taking a sample of size 30,000 tweets for preprocessing (to help reduce resource demand). Figure 3.1 and 3.2 show the distributions of sentiment labels for 'TextBlob' and the transformer, respectively. Both approaches have classified most of the 30,000 tweets as negative, but contrastingly, 'TextBlob' has a significantly higher number of positively-classified tweets (33%) compared to the transformer approach (7%). The latter also has highlighted more of the tweets as neural and negative. Given the context that the Cost of Living Crisis was a period of challenging hardship for a wide range of people, it would seem more likely there would be few positive tweets regarding this. This implies the transformer model outperformed 'TextBlob' for sentiment analysis. Figures 3.3, 3.4, and 3.5 shows the word clouds for the most frequent used words in positively-classified, negatively-classified, and neutrally-classified tweets.

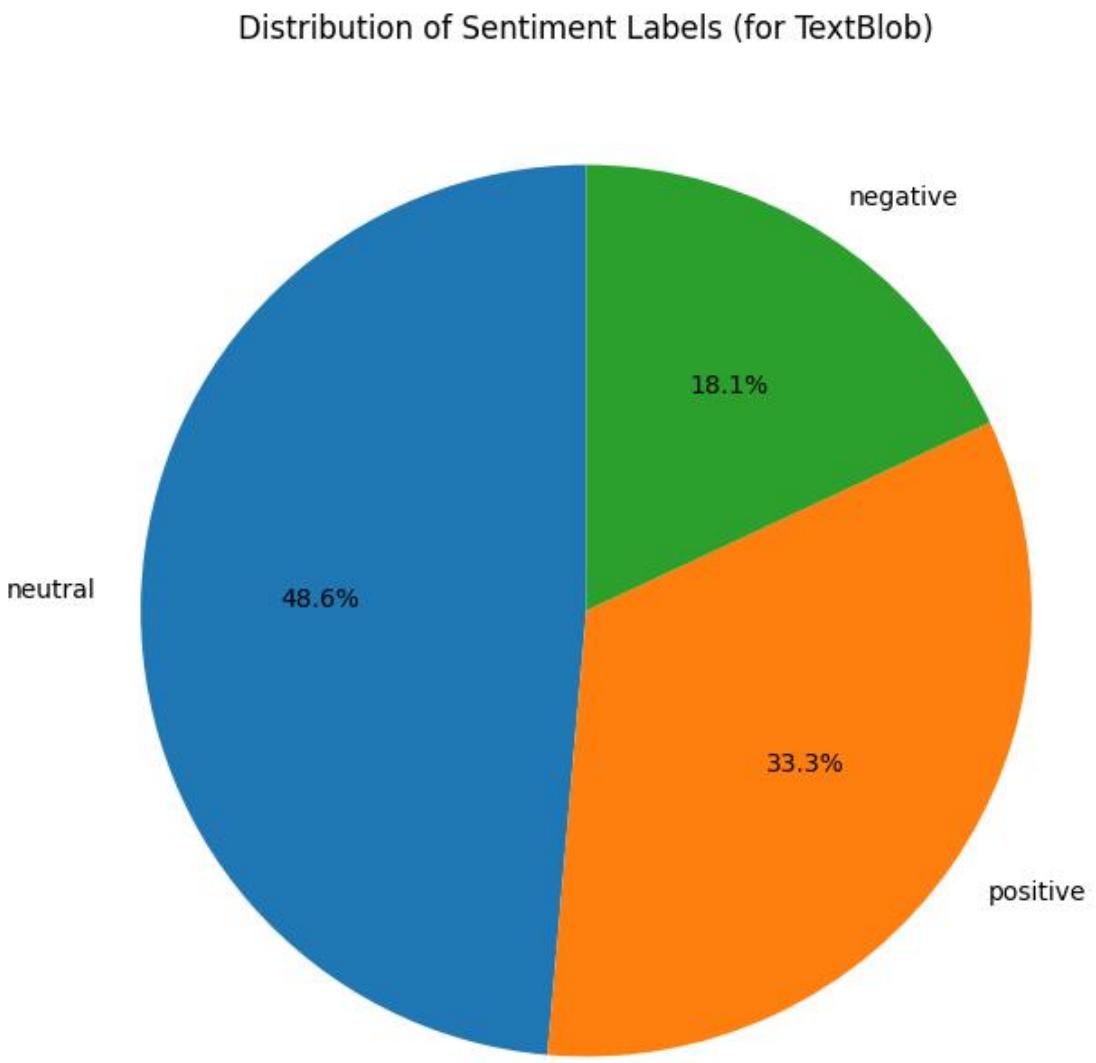

Figure 3.1: Distribution of sentiment labels for 'TextBlob'.

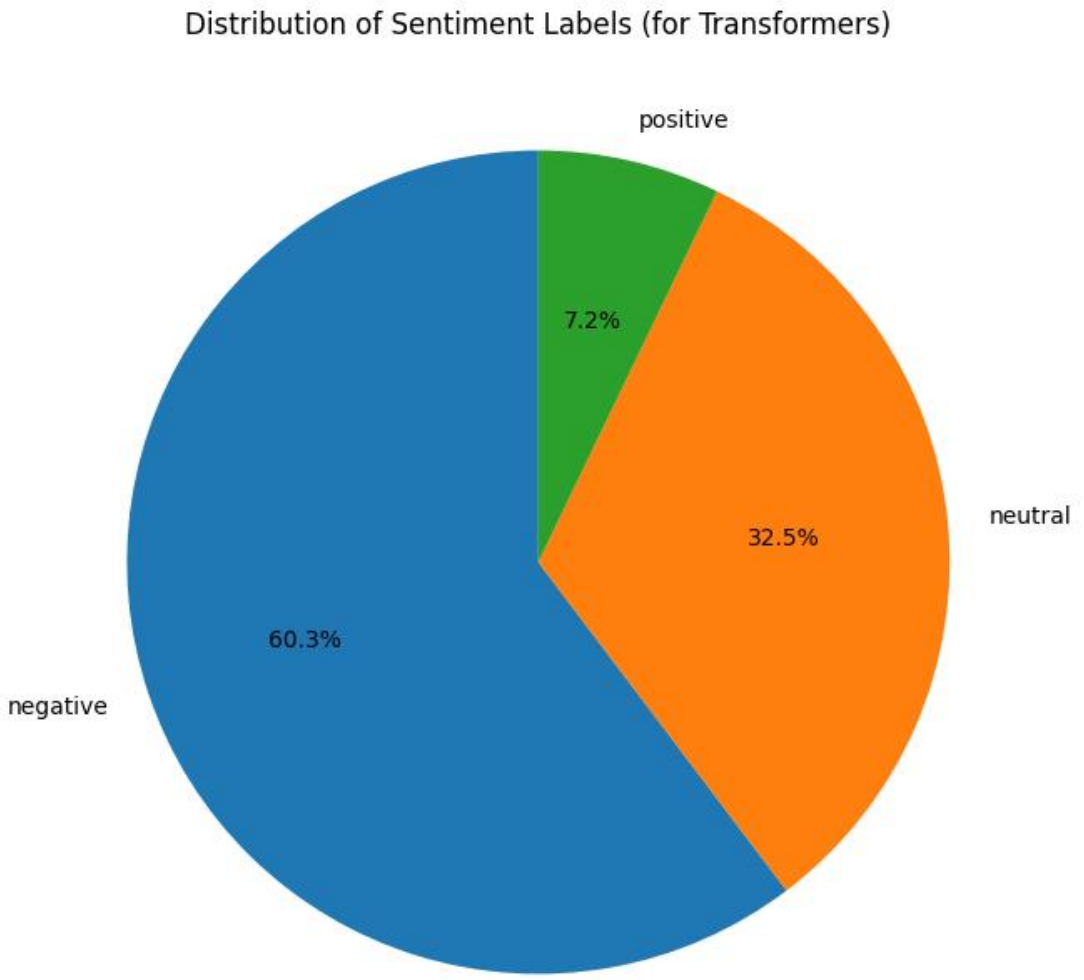

Figure 3.2: Distribution of sentiment labels for the transformer.

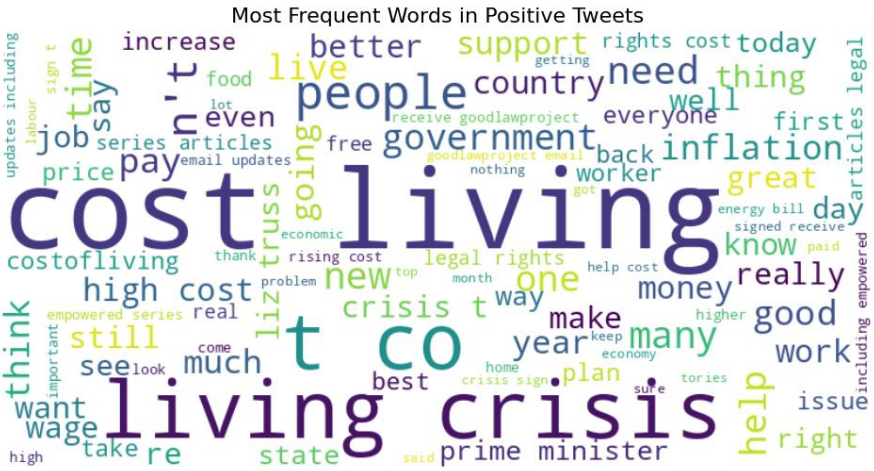

Figure 3.3: Word cloud of the most frequently used words in positively-classified tweets.

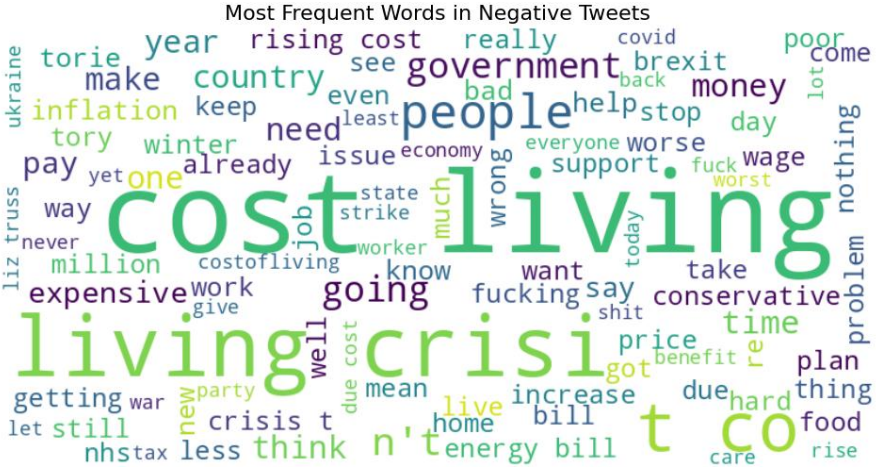

Figure 3.4: Word cloud of the most frequently used words in negatively-classified tweets.

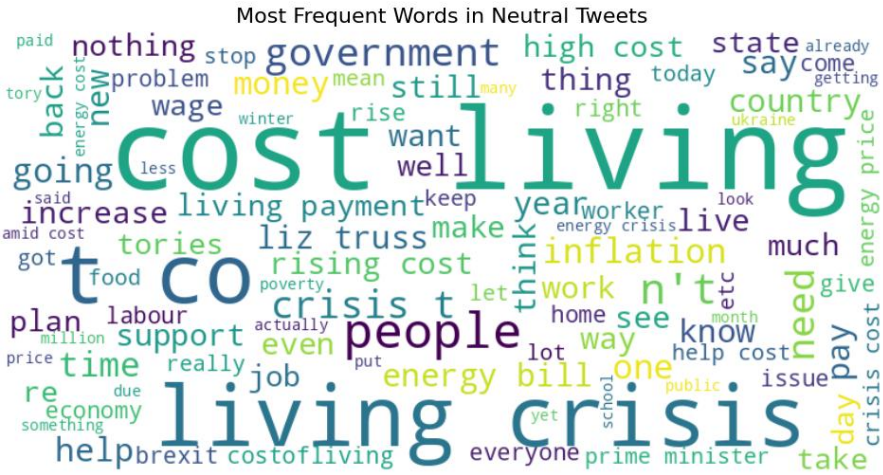

Figure 3.5: Word cloud of the most frequently used words in neutrally-classified tweets.

3.2 News Analysis and Topic Modelling

Recent news articles from late April and early May 2025 were retrieved using the Python 'newsapi' package.

3.2.1 Summary statistics

Figure 3.6 shows the process used to get the summary statistics shown in Tables 3.1 and 3.2. The number of words and sentences was calculated using the word and sentence tokenisers from the 'NLTK' Python package.

```
TITLE: Shoe thrown at President Ruto: Kenya government condemns 'shameful' incident
APPROX. NUM OF SENTENCES: 12
APPROX. NUM OF WORDS: 1052
PUBLISHED; 2025-05-05T09:16:44Z
----------------------------
TITLE: Canada election results: Liberals projected to win, Mark Carney on course to secure full term
APPROX. NUM OF SENTENCES: 62
APPROX. NUM OF WORDS: 2567
PUBLISHED; 2025-04-29T02:21:53Z
----------------------------
TITLE: Trump is acting like it's his choice whether he obeys the Constitution | CNN Politics
APPROX. NUM OF SENTENCES: 83
APPROX. NUM OF WORDS: 2628
PUBLISHED; 2025-05-05T04:00:50Z
----------------------------
TITLE: The White House billionaires have no idea how Americans live (and don't seem to care) | CNN Business
APPROX. NUM OF SENTENCES: 55
APPROX. NUM OF WORDS: 2011
PUBLISHED; 2025-05-01T18:46:52Z
----------------------------
TITLE: It took 250 years to build what Trump is trying to undo | CNN Politics
APPROX. NUM OF SENTENCES: 151
APPROX. NUM OF WORDS: 4429
PUBLISHED; 2025-05-09T08:00:50Z
----------------------------
Total number of articles: 13
Total number of words: 29313
Total number of sentences: 813
Mean number of words per article: 2254
Mean number of sentences per article: 62
```

Figure 3.6: Getting the summary statistics of the articles.

Table 3.1: Summary statistics of API-retrieved news articles.

Summary metric	**Value**
Number of articles	13
Total number of words	29313
Mean number of words per article	2254
Mean number of sentences per article	62

Table 3.2: Retrieved article titles with their publication date and word count.

Article title	**Publication date**	**Approximate number of words**
UK inflation rate jumps to highest in more than a year	21-05-2025	1170
Inflation news: Prices are falling, but it might be for bad reasons	14-05-2025	2181
How to trash an economic superpower in 100 days	28-05-2025	2224
CNN Poll: A growing majority says Trump has made the economy worse, with most sceptical of his tariff plans	28-05-2025	2384
Inflation surprise suggests outlook could be gloomier than we thought	21-05-2025	978
Memorial Day gas prices: Set for their cheapest Memorial Day since 2021	20-05-2025	2010
Australia's center-left Labour Party looks set to retain power, according to media projections	03-05-2025	2686
It's a dangerous time for a big, beautiful and expensive bill	22-05-2025	2993
Shoe thrown at President Ruto: Kenya government condemns 'shameful' incident	05-05-2025	1052
Canada election results: Liberals projected to win, Mark Carney on course to secure full term	29-04-2025	2567
Trump is acting like it's his choice whether he obeys the Constitution	05-05-2025	2628
The White House billionaires have no idea how Americans live (and don't seem to care)	01-05-2025	2011
It took 250 years to build what Trump is trying to undo	09-05-2025	4429

3.2.2 Keyword extraction

Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical measure used to evaluate the importance of words compared to a pre-defined collection of documents (also known as a corpus). This was used to extract the supposedly most important words from the article descriptions, as shown in Figure 3.7. These results are visualised in Figure 3.8. As highlighted by the visualisation, concepts such as government figures, inflation, and prices are seemingly common across the range of articles, suggesting these concepts are heavily focused on during financial hardships. This seems logical, as living costs are largely influenced by inflation and current prices indices, along with what actions governments take to either exacerbate or support the populations during these times.

```
Top Keywords: ['according' 'americans' 'cnn' 'country' 'donald' 'economic' 'election'
 'federal' 'government' 'inflation' 'liberal' 'march' 'minister' 'party'
 'possible' 'president' 'prices' 'prime' 'remake' 'trump']
```

Figure 3.7: TF-IDF designated most important words from article descriptions.

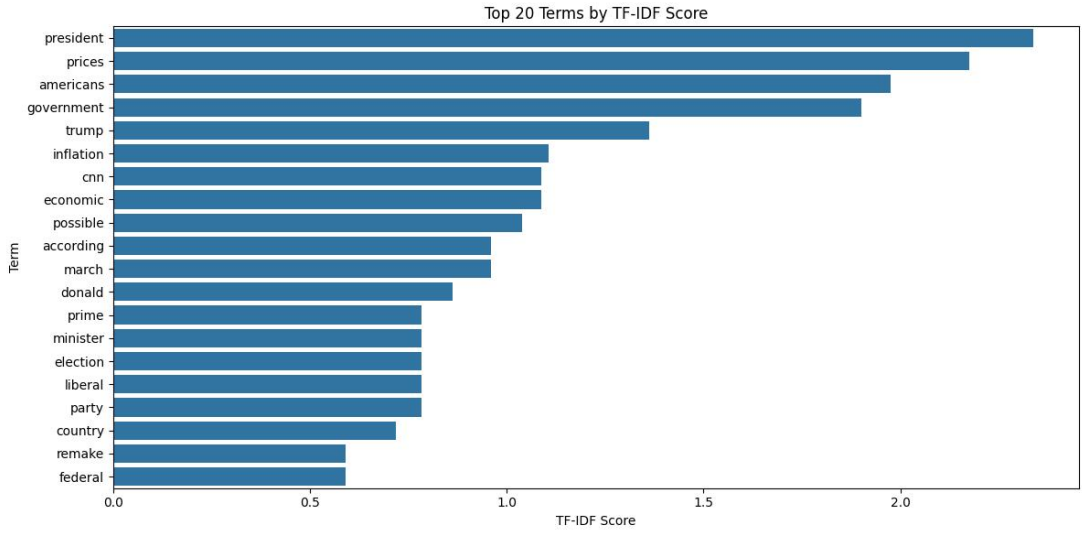

Figure 3.8: Bar chart of the most important words selected by TF-IDF.

3.2.3 Topic modelling

The TD-IDF results from Section 3.2.2 were used to influence the topic modelling. Non-negative matrix factorisation (NMF) was used to designate the appropriate words to the

appropriate topic category. NMF was selected due to its simple and explainable topic modelling results, mainly because it only requires TF-IDF. Three topic categories were set for this. Figure 3.9 presents the identified topic results. As shown in the figure, it appears it has been somewhat successful at picking out topics involving government reshuffling, the relation between governments, inflation and prices, and political parties, respectively. Figure 3.10 shows word clouds for each of these topics identified.

```
Topic 1: possible, remake, trump, president, government
Topic 2: president, possible, inflation, americans, prices
Topic 3: election, minister, party, liberal, prime
```

Figure 3.9: Identified topics.

Figure 3.10: Word cloud of the topic modelling results by NMF.

3.2.4 Article summarisation

A Large Language Model (LLM), 't5-small', was used to perform abstractive summarisation for a random BBC News article. This is so the model itself can "understand" the text itself, and provide a summary in its own words of its understanding. This is more ideal compared to extractive summarisation as it is more human-understandable and helps the analyst better identify key insights. Figure 3.11 shows the original selected article text, while Figure 3.12 shows the model-generated summary of this text. Overall, the quality of the summary is fairly good at picking out key statistics and points (regarding inflation metrics and analyst expectations) from the original text, while ignoring less significant information or invalid words. Although, possibly due to it being a light-weight LLM, its punctuation and proper formation of sentences could be improved further.

```
ARTICLE TITLE: UK inflation rate jumps to highest
in more than a year

ORIGINAL TEXT: UK inflation rate rises to highest
in more than a year - BBC NewsBBC HomepageSkip to
contentAccessibility HelpYour accountHomeNewsSport
EarthReelWorklifeTravelCultureFutureMusicTVWeather
SoundsMore menuMore menuSearch BBCHomeNewsSportEar
thReelWorklifeTravelCultureFutureMusicTVWeatherSou
ndsClose menuBBC NewsMenuHomeIsrael-Gaza warWar in
UkraineClimateVideoWorldAsiaUKBusinessTechMoreScie
nceEntertainment & ArtsHealthWorld News TVIn
PicturesBBC VerifyNewsbeatBusinessNew Tech
EconomyTechnology of BusinessArtificial
IntelligencePaths to SuccessBills push inflation
to highest in more than a yearImage source, Getty
ImagesCharlotte EdwardsBusiness reporter, BBC
NewsPublished21 May 20254322 CommentsA rise in the
cost of household bills has pushed UK inflation to
its highest rate in more than a year, according to
official data.Inflation was 3.5% in April, up from
2.6% in March and higher than economists had
expected.Water, gas and electricity prices all
went up on 1 April along with a host of other
```

Figure 3.11: Original article title and text before summarisation.

```
SUMMARY: inflation was 3.5% in April, up from 2.6%
in March and higher than economists had expected .
the bank of England has previously said it expects
inflation to peak at 3.7% between July and
September 2025 . a sharp jump in the cost of
airfares compared to April last year also helped
drive up inflation .
```

Figure 3.12: Model-generated summary.

4 Conclusions and Future Work

In summary, this report investigated natural language processing analytics and text manipulation techniques to illustrate the usefulness of these techniques, and how they could be applied to real-world data.

Future work includes investigating how Large Language Models (LLMs) can be optimised to better handle low amounts of textual context. Additionally, bias in LLMs could also be investigated to help develop their explainability and reliability.

Reference List

Hammi, B., Zeadally, S. and Perez, A. J. (2023). Non-Fungible Tokens: A Review. *IEEE Internet of Things Magazine*, 6 (1), pp. 46–50. Available at: 10.1109/IOTM.001.2200244.

Hordiienko, L. and Khamula, S. (2020). Special relationship between the United Kingdom and the USA: current state and future prospects. *Political Science and Security Studies Journal*, 1 (2), pp. 10–16. Available at: 10.5281/zenodo.4395111.

Leskovec, J. (2012). *Social circles: Twitter*. Available at: https://snap.stanford.edu/data/ego-Twitter.html [Accessed on 8th Apr. 2025].

Moosa, I. A. (2024). *The Cost of Living Crisis: Implications for Economic Theory and Public Policy*. Edward Elgar Publishing.

Naghshzan, A. H. (2023). *English Tweets of 2022*. [dataset]. Available at: https://www.kaggle.com/datasets/amirhosseinnaghshzan/twitter-2022 [Accessed on 8th Apr. 2025].

Norton, R. and Horton, D. (2023). 'Together and Apart: Canada and the United States and the World Beyond'. In: *Canada and the United States: Differences That Count*. Ed. by D. M. Thomas and C. Sands. 5th ed. Toronto, USA: University of Toronto Press, pp. 73–103.

Office for National Statistics (2025). *CPI ANNUAL RATE 00: ALL ITEMS 2015=100*. Available at: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7g7/mm23 [Accessed on 28th Feb. 2025].

Rees, B. (2020). RAPIDS cuGraph adds NetworkX and DiGraph Compatibility. *RAPIDS AI*. [blog] 2 October. Available at: https://medium.com/rapids-ai/rapids-cugraph-networkx-compatibility-d119e417557c [Accessed on 15th Apr. 2025].

Statista (2025). *Number of internet and social media users worldwide as of February 2025*. Available at: https://www.statista.com/statistics/617136/digital-population-worldwide/ [Accessed on 20th Apr. 2025].

Tleonel (2022). *Cost of Living — +144k Tweets - ENG — Aug/Sep 2022*. [dataset]. Available at: https://www.kaggle.com/datasets/tleonel/cost-of-living [Accessed on 24th Apr. 2025].

Wan, Z., Mahajan, Y., Kang, B. W. et al. (2021). A Survey on Centrality Metrics and Their Network Resilience Analysis. *IEEE Access*, 9 (1), pp. 104773–104819. Available at: 10.1109/ACCESS.2021.3094196.

Zhang, J., Fei, J., Song, X. et al. (2021). An Improved Louvain Algorithm for Community Detection. *Mathematical Problems in Engineering*, 2021 (1), pp. 1–14. Available at: 10.1155/2021/1485592.