

Welcome to my DS project



This app was developed by Juan Equihua on April 10th, 2022. The purpose of this app is analysing the Tweets dataset provided for the assignment.

Filter dataset by date:

Start date

2020/11/09

End date

2021/11/10

Do you want to filter by topics and analyse more specific data?

Poor Pay

Cost of Living

Wage Growth

Rich People

Low Income Families

Public Sector Pay

Government Support

Leaseholding

State Pension

Pay Rises

Income Tax

Council Tax

Small Businesses

Social Care

House Prices

Job

Minimum Wage Inc...

National Insurance

Gender Pay Gap

Mental Health

Long Hours

Poor People

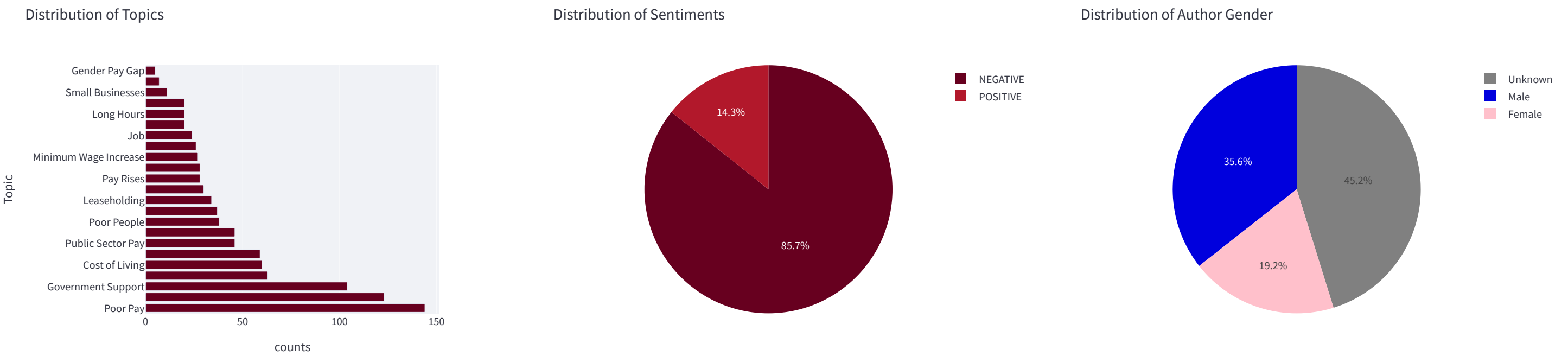
⌵

Statutory Sick Pay

Note: Filtered data contains 1000 records.

Analysis of the dataset:

This section intends to present initial visualisations of the data contained in tweeter dataset, such as the distributioon of topics, tweet sentiment obtained from the [BERT](#) transformers by [Hugging Face](#), and distribution of author gender.



Descriptive engagement stats of filtered data:

	total_engagement	comments	shares	likes
mean	9.3380	0.2910	2.4000	6.6470
max	4,349.0000	97.0000	1,126.0000	3,126.0000

Find the most relevants tweets in the dataset based on a string query using Gensim

This section intents to find the most relevant tweets in the dataset for a given query string provided by the user, query similarities are obtained with the use of the [gensim](#) library, a software framework for topic modelling with large corpora. For each relevant tweet retrieved, we extract the top 10 keywords in the text with the use of the [Yake](#) library.

Please type a string query

Cost of living in uk against Europe and US

How many similar tweetees would you like to see?

10

How many keywords would you like to see

5

	Top similar tweets	keywords
0	The issue with this is no one ever offers an alternative. Should women not have children? Should women not have jobs? How does that work if men want kids, which are basically impossible to afford on a single income? Does anyone have a suggestion, other than deriding "bad women"?	["offers an alternative","women","alternative","issue","offers"]
1	Staff Ts and Cs like Agenda for Change are national (just about stillwatch for ICS and the NHS Bill breaking up national pay bargaining). Poor pay is national. Nothing to do with the employer. The fight for NHSPay15 is national	["NHS Bill breaking","Agenda for Change","NHS Bill","stillwatch for ICS","Bill breaking"]
2	Red Herring 2: VC salaries probably should be high: if they had similar levels of responsibility in the forprofit sector, they'd be doubled, tripled or more. The problem is that their pay goes up, often with little oversight, while staff are being underpaid and overworked. 8/	["Red Herring","forprofit sector","similar levels","levels of responsibility","Herring"]
3	John McDonnell MP: "Sunak starts his budget speech boasting his economic policies are working. Tell that to the families struggling to make ends meet after the £20 cut to Universal Credit & the nurses who will be hit by an effective pay cut as inflation rises"	["budget speech boasting","Sunak starts","John McDonnell","policies are working","starts his budget"]
4	Indeed . The facts are that uk nursing pay has been cut by 8.3% since 2010; despite the conservatives trying to somehow convince the country that they are “ getting a pay rise “ . Even the 1% potential rise is below inflation.	["potential rise","convince the country","nursing pay","pay rise","pay"]
5	Politicians who tell you nurses are only worth 12.1% pay rises are enemies of the working class. Our nurses suffered a 7.4% pay cut over the last decade. In addition, they lost 1,199 health workers to CV19. Do not dare tell me they are not worth a proper pay rise. 15% Now!!!	["working class","pay","Politicians","class","nurses"]
6	How outrageous for Sunak to say business is "drunk on cheap labour", when for 11yr Tory govmnts have supported zero hrs contracts, gig economy, frozen public sec pay, and now refuse to increase minimum wage to a decent level.	["public sec pay","frozen public sec","increase minimum wage","Tory govmnts","gig economy"]
7	Our nurses have risked their lives to keep us safe throughout the pandemic. For the Gov’t to deny them the pay rise they deserve is an outrage, to deliberately mislead the public in order to get away with it is beyond despicable	["nurses have risked","risked their lives","pandemic","pay rise","rise they deserve"]
8	We are supporting NotTooMuchtoMask Older, mentally ill and disabled people, BAME and poorer communities, care staff largely low paid women have paid a heavy price in the pandemic. Pls respect our lives as well as your freedom.	["supporting NotTooMuchtoMask Older","care staff largely","staff largely low","NotTooMuchtoMask Older","BAME and poorer"]
9	3. What needs doing? Many of the bursary cuts need to be at least partially reversed. Biology, English, Geography are looking particularly in need of boosting It's also time to plan for teacher pay increases. Another pay freeze would likely lead to a teacher shortage disaster	["English","Geography","teacher","pay","Biology"]

Find top 3 topic predictions from Fine-tuned BERT model.

Predictions of Top 3 query classes (topics) are made via a a fine-tuned BERT model trained for 5 epochs with 70% of the data provided for this task. Fine-tuning code is contained in the notebook **Aux_notebook - Fine tuning BERT for multiclass.ipynb** file. The BERT model was deployed via a Flask API listening at localhost:80 and returns the top 3 predictions for the string query.

First class in your search: Government Support

Second class in your search: Poor Pay

Third class in your search: Low Income Families

Find the entities names in top similar Tweets using BERT.

Once we have identified the top 10 momst similar tweets in the data we are interested in identify people, locations, and organisations mentionend in the top most similar tweets, To achieve this we use the NER (Name Entity Recognition) system available from the [transformers Hugging Face](#) library with the use of [BERT](<https://huggingface.co/dslim/bert-base-NER>). This will identify on the fly, all entities for our top tweets and print them below:

(Please note that this process might take a while as streamlit cannot cache the NER system natively.)

- People mentioned in top similar queries: John McDonnell, Sunak
- Locations mentioned in top similar queries:
- Organisations mentioned in top queries: ##M, ICS, NHS, Not, Universal Credit, VC