"What were the societal impacts of the spread of COVID-19 notably the rise of misinformation, and the effects on mental health, school performance, and employment?"

Introduction:

Aims and objectives of the project:

The COVID-19 pandemic significantly impacted the worldwide population. We aimed to explore the pandemic's effects on: the rise of misinformation (the 'infodemic'), mental health challenges, school performance disruptions, employment shifts and the rise of vaccine hesitancy. Our research aimed to investigate these aspects of the pandemic, identify trends and provide recommendations for future policymakers in the government, healthcare, education and employment sectors.

Roadmap of the report:

The **Background** section outlines the context of our selected research area. The **Specifications and Design** section addresses technical and non-technical requirements, including architecture and tools used. **Implementation and Execution** discusses the development approach, team roles, tools, achievements, challenges, and the use of agile methodology. The **Data Collection** section explains the information required, available sources, and data collection processes, including the use of APIs. The **Result Reporting** section outlines our key results and visualisations. Finally, the **Conclusion** summarises our key findings and recommendations. Supporting tables and figures to this report can be found in the **Appendix**.

Background:

History of COVID-19 and its Societal Effects

COVID-19 was declared a global pandemic by WHO in March 2020[1]. The aftermath of the global pandemic was extensive, disrupting many facets of everyday life. Public health measures, such as lockdowns and social distancing, triggered significant societal impacts, relating to mental health, education, employment, and the spread of misinformation. The introduction of a new vaccine also created mixed feelings amongst the public, ranging from hope and relief to scepticism and fear, often influenced by misinformation and varying levels of trust in healthcare systems and authorities. The significance of this

Specifications and Design:

Requirements technical and non-technical:

Prior to undertaking the project we identified the key technical and non-technical requirements for the project as follows:

Technical requirements:

Employment of key python libraries taught over the duration of the course - **NumPy, Pandas** and **Matplotlib**. **Integration with at least one API**. Undertaking the **key stages in the data lifecycle: sourcing & pre-processing**, **cleaning**, **data analysis & visualisation**. The employment of the **versioning** tool of **Git**, with a **remote repository** on **GitHub** would allow us to collaborate on the same project via different branches for our different features. Finally the use of a **project management tool** to track progress.

Non Technical requirements:

The non-technical requirements mainly focused on **team collaboration** with the development of specific roles and how we **communicated** and **organised the workflow**. The development of a **research question** and **hypotheses** were central to **project planning**. **Ethical considerations** were also key in data processing, ensuring that the data was anonymous and compliant with guidelines. The output of our research would then be summarised in a **report** and a **group project presentation**. We would need to **identify and work through any challenges** through **collaborative problem-solving** and **adapt the project scope** based on data availability and the time constraints of the project.

This identification of technical and non-technical approaches ensured a comprehensive and well-coordinated project execution.

Key deliverables were identified:

1. **Cleaned and Merged Dataset:** Integrated data ready for analysis.
2. **Visualisations:** Key graphs answering the research questions.
3. **Project Report:** Detailed document covering objectives, methods, results, and conclusions.

[1] Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed. 2020 Mar 19;91(1):157-160. doi: 10.23750/abm.v91i1.9397. PMID: 32191675; PMCID: PMC7569573.

4. **Presentation:** Concise and engaging delivery of findings.

Design:

1. **Research Objectives** were set for each research area (Appendix 1)
2. **Scope of the Study**:
    a. **Geographical Boundaries**: Focused primarily on the United Kingdom, with some global comparisons for context (ESOC dataset).
    b. **Timeframe**: Examined data from 2020 (pandemic onset) through mid-2023 to assess short- and long-term impacts.This constraint was dependent on the available data sources timeframe so not filled for each area of research.
    c. **Population Sample**: Included a diverse demographic range, covering different age groups, genders, and socio-economic statuses.
3. **Ethics:** All data used in the study is publicly available and anonymised, ensuring that ethical guidelines regarding privacy and consent are adhered to. The study will not involve any direct data collection from individuals, nor will it require any interaction with participants.
4. **Methodological Framework**: The study followed a modular design to systematically address each focus area. Proposed methodology for each

Architecture:

We separated our design into five focus areas: misinformation, mental health, education, employment, and vaccine hesitancy. This modular design allowed for agile working with different team members able to work on specific parts of the project in parallel to others who were retrieving or cleaning other datasets.

The codebase was structured to allow for easy flow of data through a clear path. We used the ETL (Extract, Transform, Load) framework for the integration of data in this project. A file tree seen in appendix 2 allows for visualisation of this structure.

Extraction:
At the root of the file structure lies the directory: `1_raw_files`. This folder contains subdirectories for storing raw data (whether in Excel, CSV, or other formats) from the above sources. The raw datasets are then integrated into the data pipeline.

Jupyter notebook files (.ipynb) for processing the raw files (and selecting which ones are appropriate or conglomerating files which span several years are then saved to the `3_raw_database_to_csv_parsing_notebooks` and those which deal with APIS are saved to `2_api_data_pipeline_notebooks.`

After extracting data from APIs or processing raw CSV/excel files, the processed databases are saved in csv format to: `4_integrated_csv_files.`

Transformation:
Once the data is extracted, it undergoes transformation and is saved to the directory: `6_cleaned_files`. The notebooks used to clean the integrated data are saved in `5_cleaning_notebooks`. Files in this directory can access processed databases via a relative pathname: `r'../integrated_csv_files/name_of_folder/file.csv' '`. In this directory, the cleaning protocol and Python scripts used for data cleaning are stored.

Load:
For visualisation and analysis, the cleaned files are used by the jupyter notebooks found in: `7_analysis_and_visualisations.` The outputs of these visualisations are saved as PNG images in: `8_final_visua`.

Implementation and Execution:

Development approach and team member roles:

We first assigned roles for team members based on their strengths and interests as seen in appendix 3. Several key roles were identified: A **project manager** to help organise the workflow, **API specialists** for retrieval of data from key sources. A **data engineer** to help integrate the data into the workflow and several **analysts** responsible for the final cleaning, processing and visualisation of their expert field. Although these roles were employed, there was inevitable overlap between roles due to time constraints and the collaborative nature of the project.

Tools and libraries:

The **python** programming language was used to write in **jupyter notebook** files in order to clean, process, analyse and visualise the data. **GitHub** was employed as our **remote repository** and we utilised **Jira** for project management.

Scientific libraries: Pandas, NumPy, SciPi, nltk
Visualisation libraries: MatPlotLib, Seaborn, Plotly, Geopandas, wordcloud, squarify
Other tools: SQL alchemy, Slack, Google Meets, Jira

The project task distribution was agreed at standup meetings and tasks were created on Jira.

Implementation process:

Achievements

- A file structure was implemented allowing for smooth flow of data through our pipeline
- Data was successfully integrated from multiple sources including APIs, excel spreadsheets and CSVs.
- Many visualisations were produced in answer to our proposed research questions.
- Every member contributed code to the project.

Challenges

- Limitations to certain API rate limits (notably the COVID-19 API) made it challenging to extract data in a timely manner with requests taking several minutes at a time for selected countries and selected timeframes.
- Formatting of certain spreadsheets (OFCOM dataset) gave rise to many nan columns and tables had to be extracted before cleaning.
- The Labour Market Statistics Dataset, whilst comprehensive, was a particularly challenging shape to analyse and extract meaningful data from with 1831 columns and 1500 rows.
- ONS Vaccine Hesitancy Dataset saw some category adjustment between quarters with different ranges being collected. This had to be rationalised in order to effectively analyse.
- Storm Darragh cut the internet connection on the day prior to submission for a member of the team causing delays in integration.

Decision to change something

- Our initial project was to look at silly newsets and unusual crimes (think Florida Man), but finding sources and useful data on this was a challenge! We were also concerned about how mean-spirited much of the social media coverage around these crimes was and we were not interested in engaging with this schadenfreude.
- We had hoped to use methods from the ML section of the course to build a predictive model using two datasets we had found with tweet text regarding COVID-19 news which were labelled as real/fake 0/1. Once we had constructed a model we had planned to integrate with the reddit API to extract posts regarding COVID-19 to use the model and see if we could identify real/fake ones with our model. We successfully integrated with the reddit API. We imported the tweet data and vectorized the text (i.e. converted it into numerical form for analysis using term frequency-inverse document frequency), split it into a train and test set when a logistic regression model was applied the resultant model was not successful and labelled everything as false (even though there was a reasonable split in the database!).

Data Collection:

Information we required

In order to answer the question we would need reliable information about the COVID-19 pandemic regarding cases, mortality and vaccination numbers for integration with the other datasets to gain valuable insights. We would need datasets which cover the above scope in order to draw comparisons between the different domains throughout the pandemic period.

Available information

A plethora of information on COVID was available. Individually the team identified several key resources for data which were pooled together and integrated into the data pipeline.

Data sources

The following data sources were included in our project

- COVID-19 Statistics API (COVID-19 Statistics API)
    - Information on cases and fatalities for different countries in the world
- UKHSA Datasets & Fingertips API (Overview - Access our data | UKHSA data dashboard)
    - Information on cases, fatalities and vaccinations for different regions in the UK
- OFCOM COVID-19 Survey Data (OFCOM COVID-19 Survey Data):
    - Insights into consumption and trust in the media during the pandemic
- ONS (Office for National Statistics) Vaccine Hesitancy Dataset:
    - For vaccine hesitancy statistics, and related indicators.

- ONS API (ONS developer Hub):
    - For wellbeing statistics, and related indicators.
- ESOC (Empirical Studies of Conflict) COVID-19 Misinformation Dataset (ESOC COVID-19 Misinformation Dataset | Empirical Studies of Conflict):
- AAAI (Association for the Advancement of Artificial Intelligence) COVID Misinformation Dataset
    - To analyze patterns and distribution channels of fake news.
- ONS Website: Labour Market Statistics Dataset (Labour Market Statistics Time Series) :
    - Information on employment rate, unemployment rate and weekly earnings
- ONS Website: Suicide Rate in England and Wales (Suicide Rate in England and Wales)
- ONS Website: Quarterly Wellbeing Estimates (Wellbeing Quarterly)
    - Seasonally and non seasonally-adjusted quarterly estimates of life satisfaction / feeling that the things done in life are worthwhile / happiness and anxiety in the UK.
- ONS Website: COVID-19 and vaccine hesitancy dataset Dataset: Coronavirus and vaccine hesitancy, Great Britain
- NHS Website: Mental Health Statistical Publications (NHS Mental Health Statistical Publications)
- NHS Website: Mental Health of Young People and Children in England (Follow Up to the 2017 Survey)

How the data was collected:

Where the dataset was available on a website these were downloaded and saved to the `raw files` directory.
API sources were integrated and the extracted data saved as a csv for processing:

1. COVID 19 API - This is a REST API which details
2. ONS API - This is a RESTful API
3. UKHSA API -

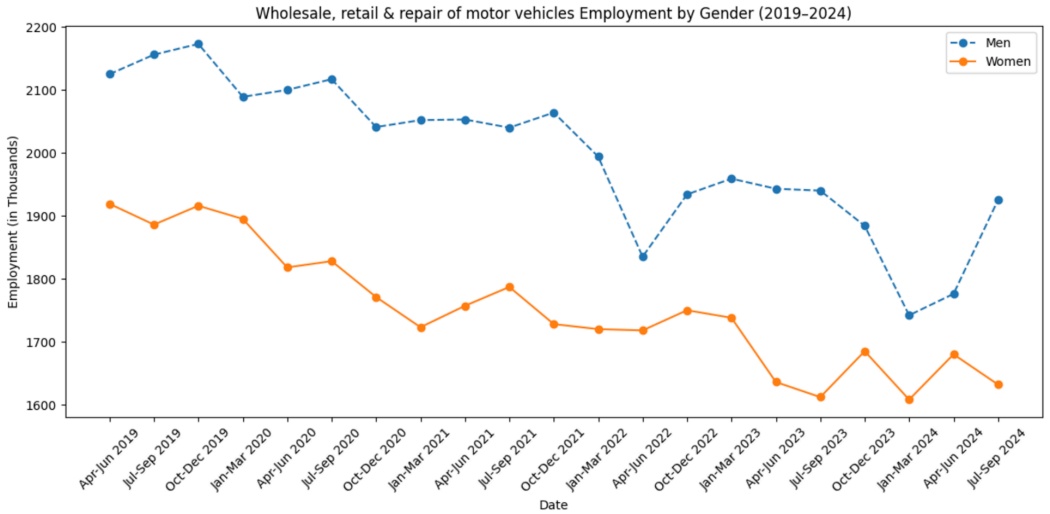
Result Reporting:

This section summarises the societal impacts of COVID-19 across education, employment, mental health, vaccine hesitancy, and misinformation. Findings are based on comprehensive data analysis from credible sources such as the ONS, OFCOM, and WHO. It summarises the societal impacts of COVID-19 across education, employment, mental health, vaccine hesitancy, and misinformation, emphasising the interconnected challenges faced globally. While this summary highlights key findings, the full results delve deeper into nuanced insights, such as the influence of regional disparities on digital access, the gendered impacts of job losses, and the role of fear-driven misinformation narratives in shaping public behaviour. These findings and visualisations collectively underscore the need for targeted, multi-sector responses to mitigate the pandemic's wide-ranging effect and are tailored to highlight key patterns and insights, supporting evidence-based decision-making for public health, education, and policy interventions.

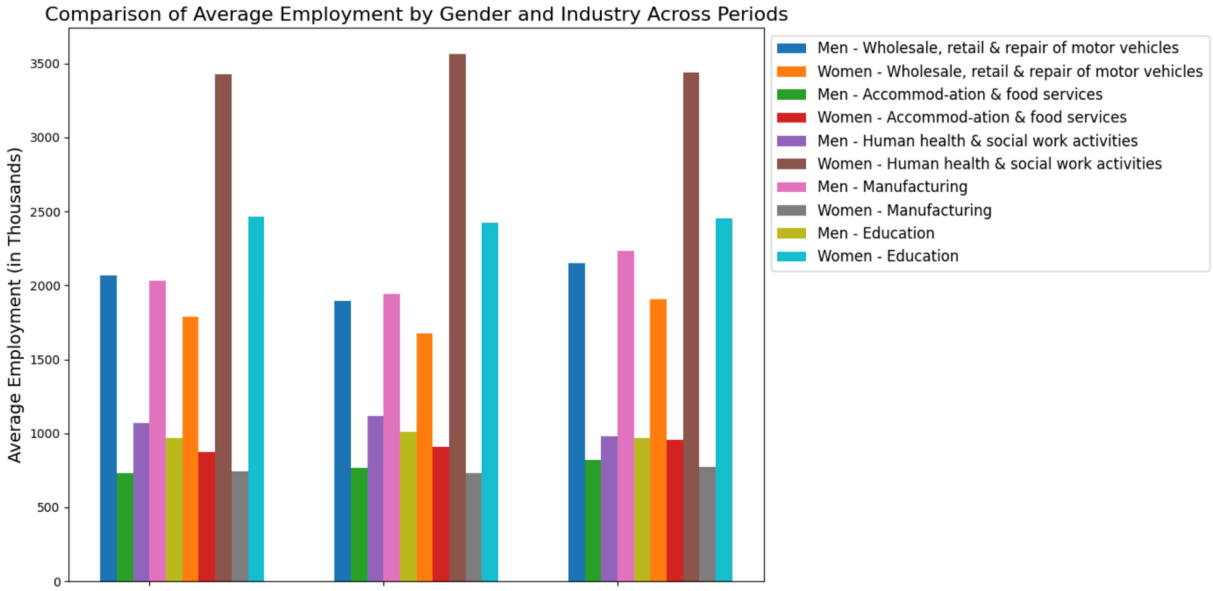
Employment and Mental Health

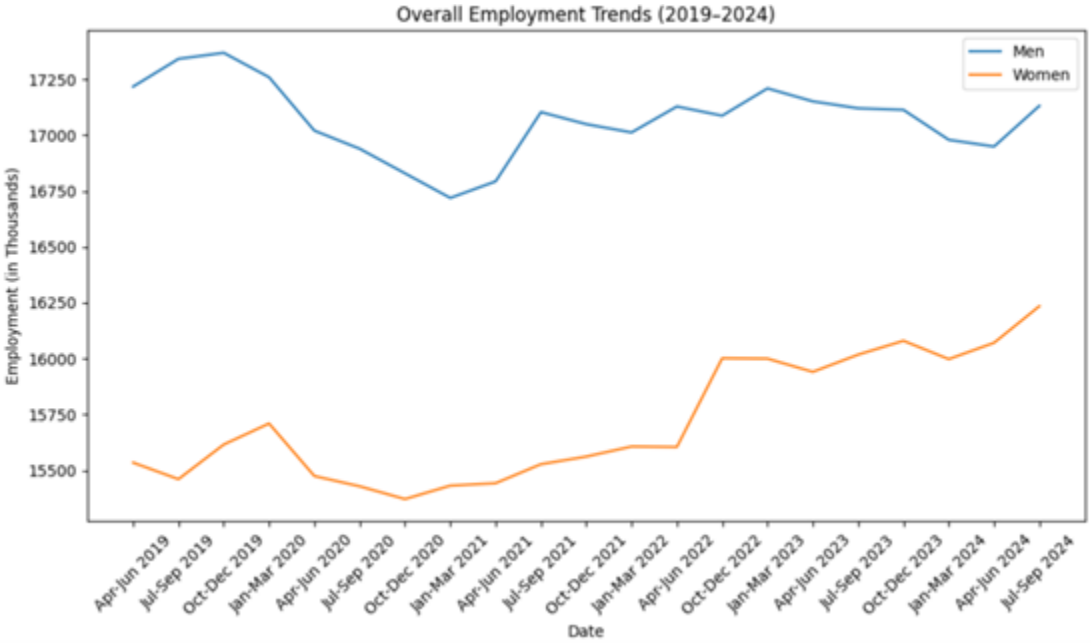
Labour market disruptions caused widespread job losses, with significant mental health implications:

- **Job Losses:** The pandemic caused the equivalent of 255 million full-time job losses globally, severely impacting sectors like hospitality and retail.



- **Mental Health Effects:** Anxiety and depression surged globally by 25%, disproportionately affecting unemployed individuals and key workers.
- **Women** experienced disproportionately higher job losses in service sectors, highlighting gendered vulnerabilities in the labour market.

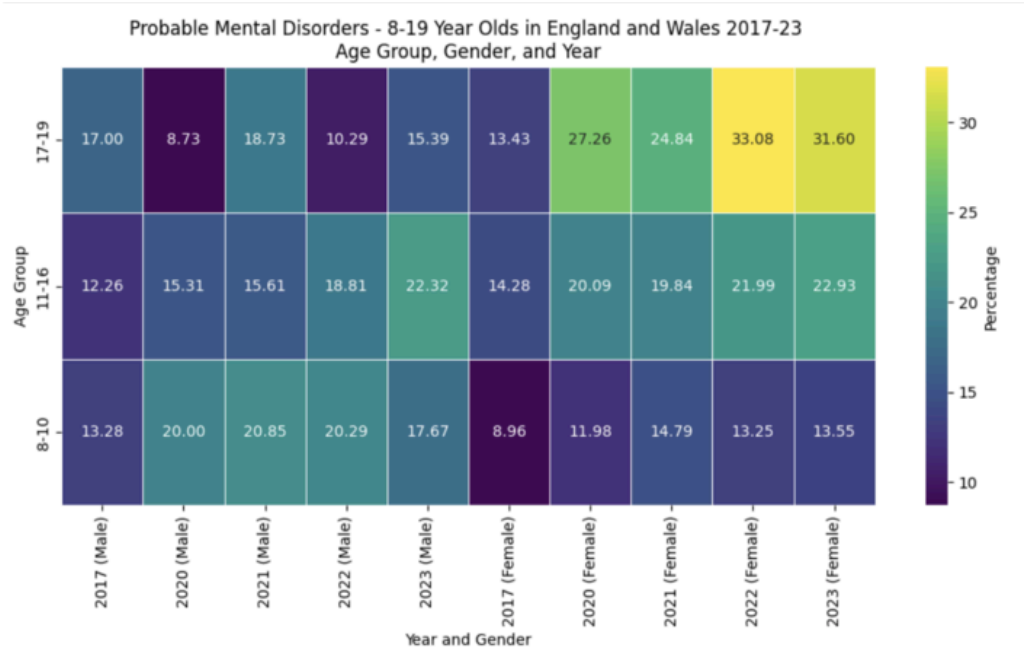




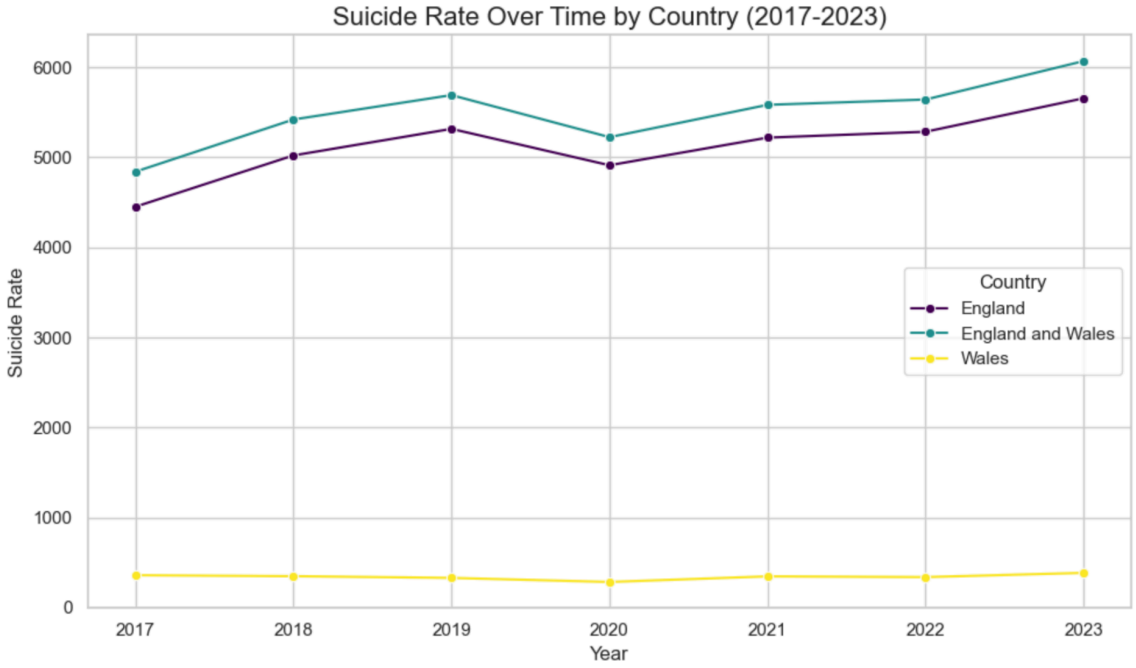
Mental Health

The pandemic exacerbated mental health challenges across demographics:

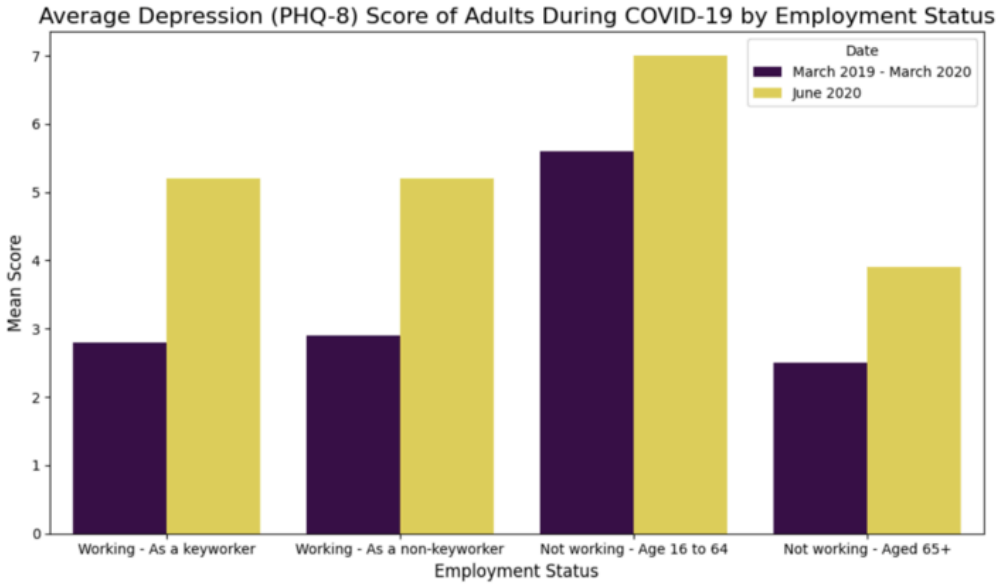
- **Anxiety and Depression:** These conditions rose globally by 25%, with children under 19 and key workers particularly affected.



- **Suicide Rates:** Regional variations in suicide rates underscore the need for targeted interventions.



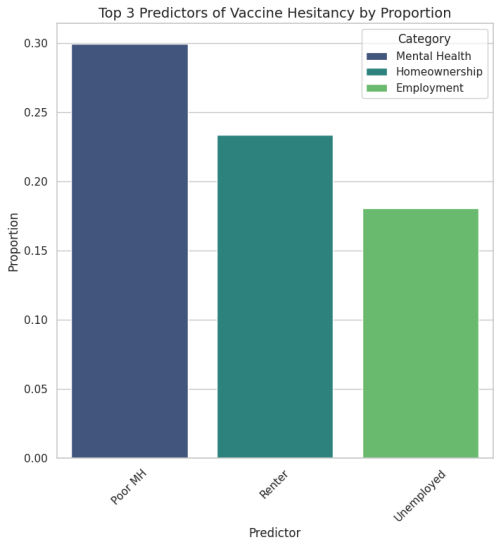
- **Depression** surged during the pandemic, with an 86% increase among key workers from 2019 to 2020, while weekly surveys in 2021 revealed the highest depressive symptoms among 16–24-year-olds, correlating with increased social media use.

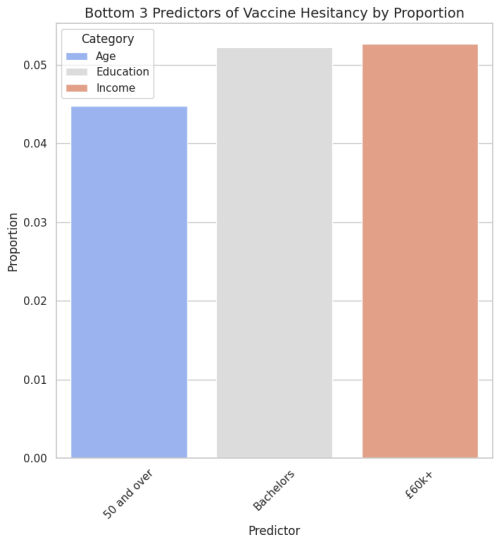


Vaccine Hesitancy & Vaccine Adoption

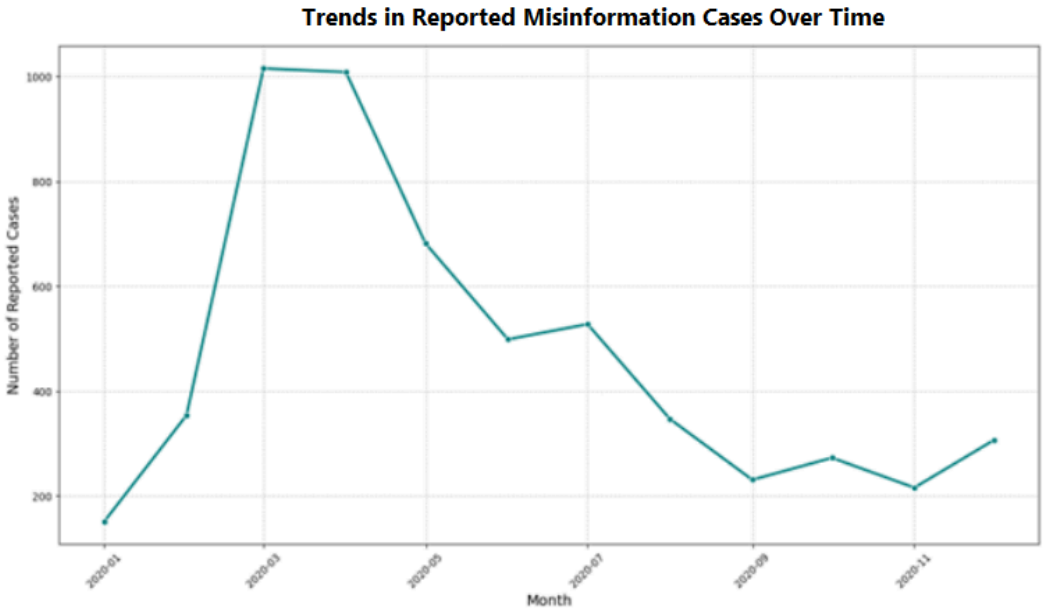
Vaccinations have proven to be one of the most effective tools in combating the spread of COVID-19; however, the pandemic also brought significant societal challenges. Vaccine hesitancy emerged as a critical challenge, linked to misinformation and socio-economic factors:

- **Predictors:** Hesitancy correlated strongly with low income, poor mental health, and younger age groups (16–29). Whereas Positive Sentiment was strongly correlated with Higher Income Groups(e.g., £60,000+ annually), Older Age (50+) and Good Mental Health.

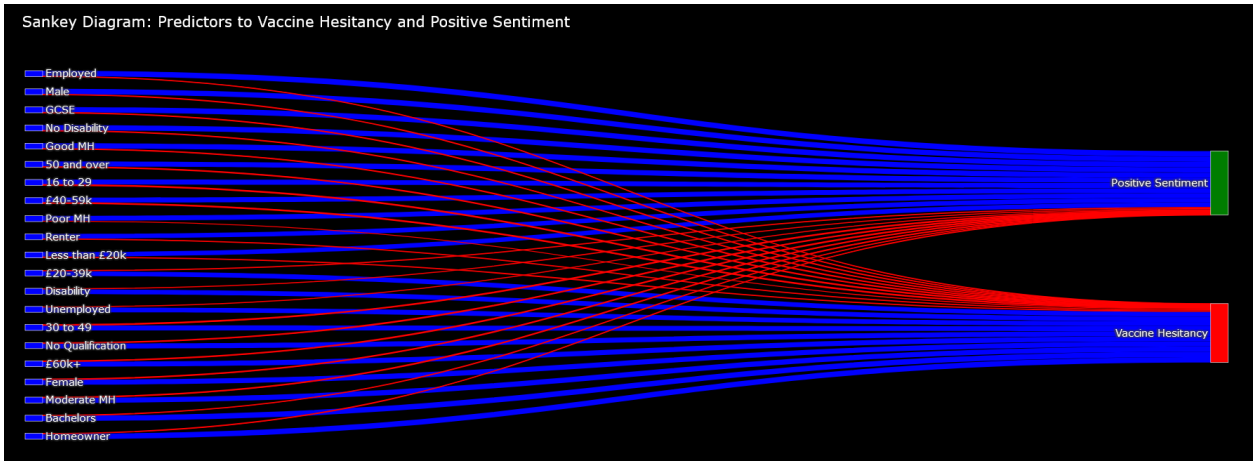




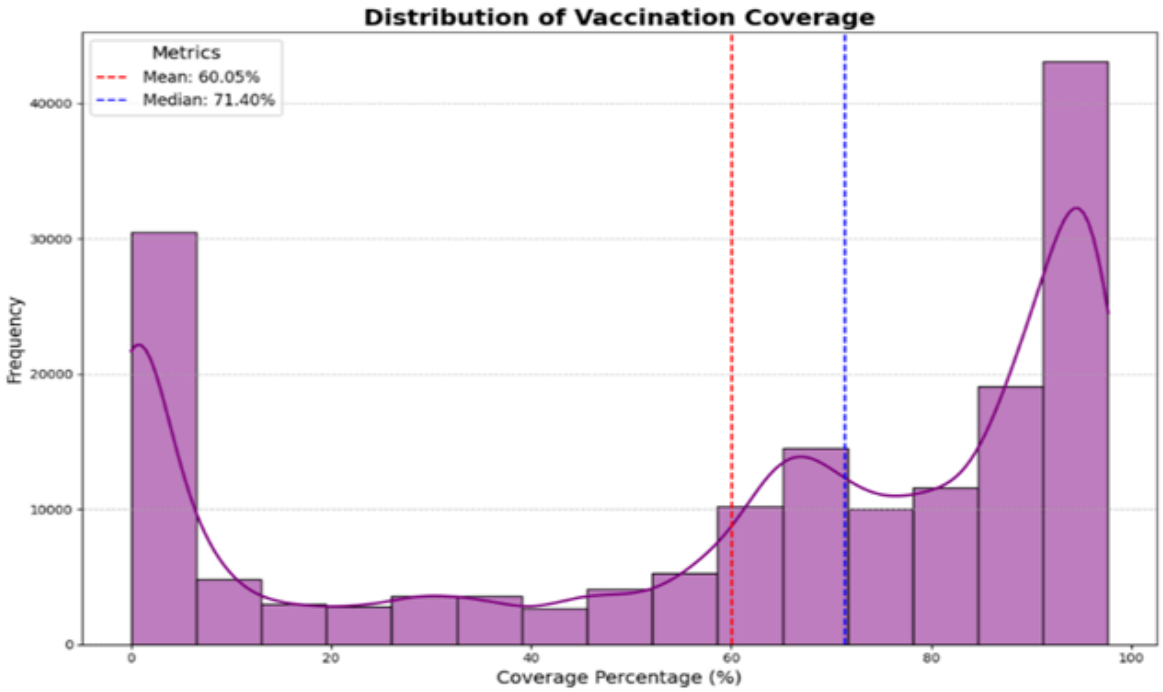
- **Role of Misinformation:** Social media platforms like Facebook and WhatsApp amplified hesitancy, with spikes in misinformation during major pandemic milestones.



- **Mid range age populations** (30–49) showed the highest levels of vaccine hesitancy, driven by both fear of side effects and susceptibility to misinformation



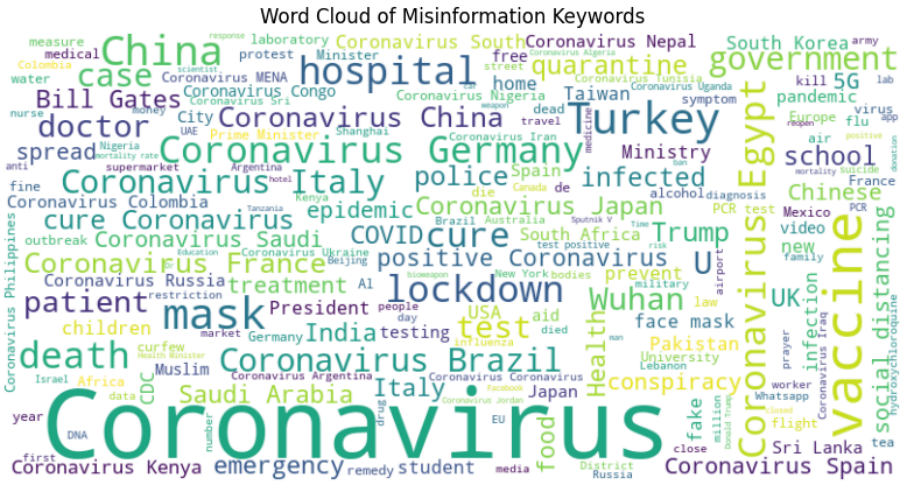
- **Regional variation in vaccine take-up** with regions like London and younger age groups, where vaccine hesitancy, misinformation, and logistical challenges have hindered uptake. Conversely, regions such as the Southeast and older age groups, which typically exhibit greater trust in vaccines and stronger compliance with vaccination campaigns.



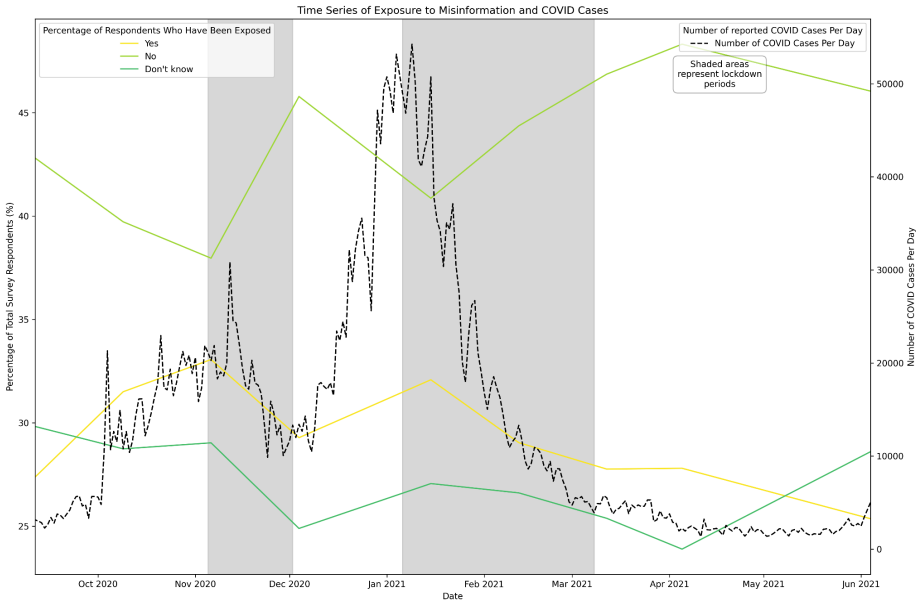
Misinformation

The pandemic saw a surge in misinformation, termed an "infodemic" by WHO:

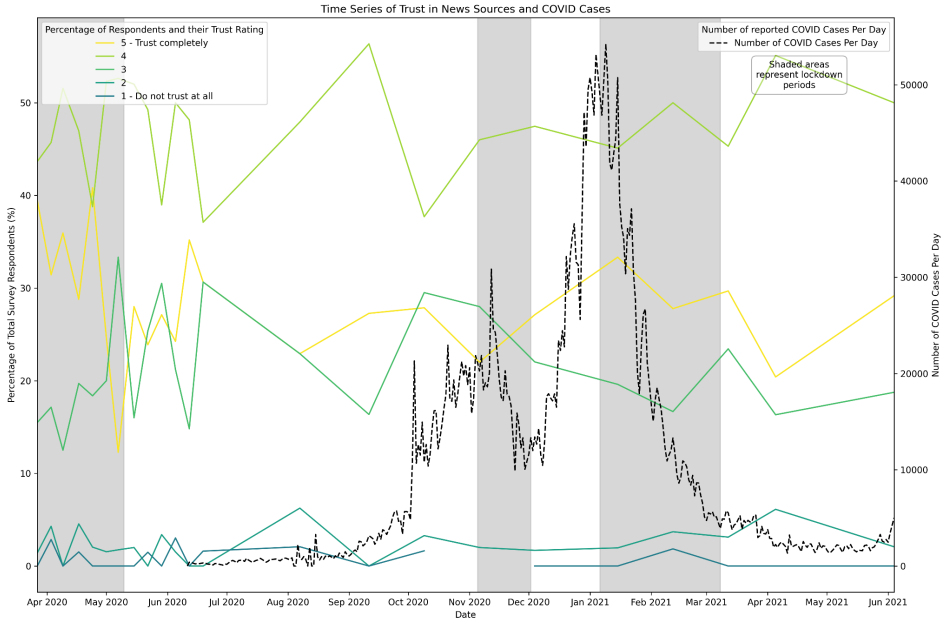
- **Key Narratives:** False cures, conspiracy theories, and downplaying COVID-19 were dominant themes. Exposure varied across demographic and regional lines.



- **Temporal Trends:** Peaks in misinformation aligned with major pandemic events, eroding trust in health organisations.



- **Trust** in traditional media sources like the BBC remained higher than social media, although traditional media trust also declined over time.

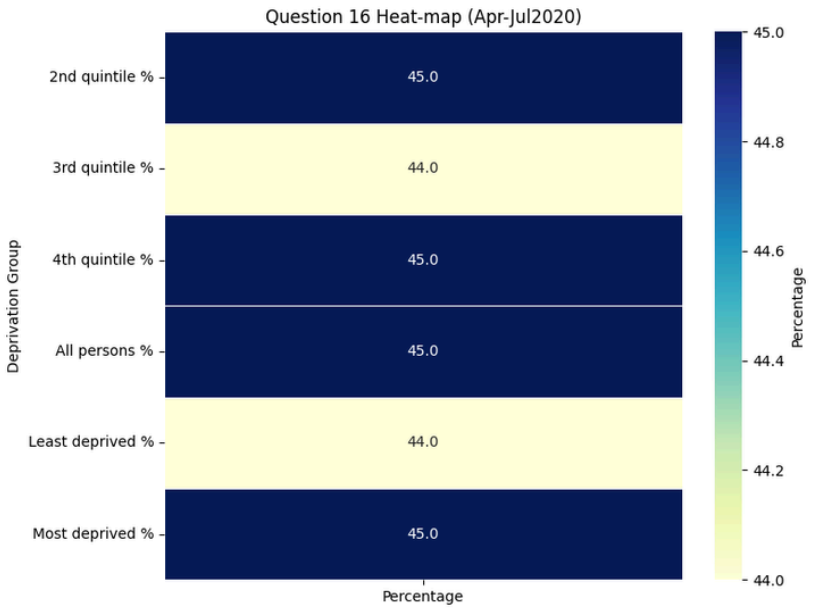


Education and Inequality

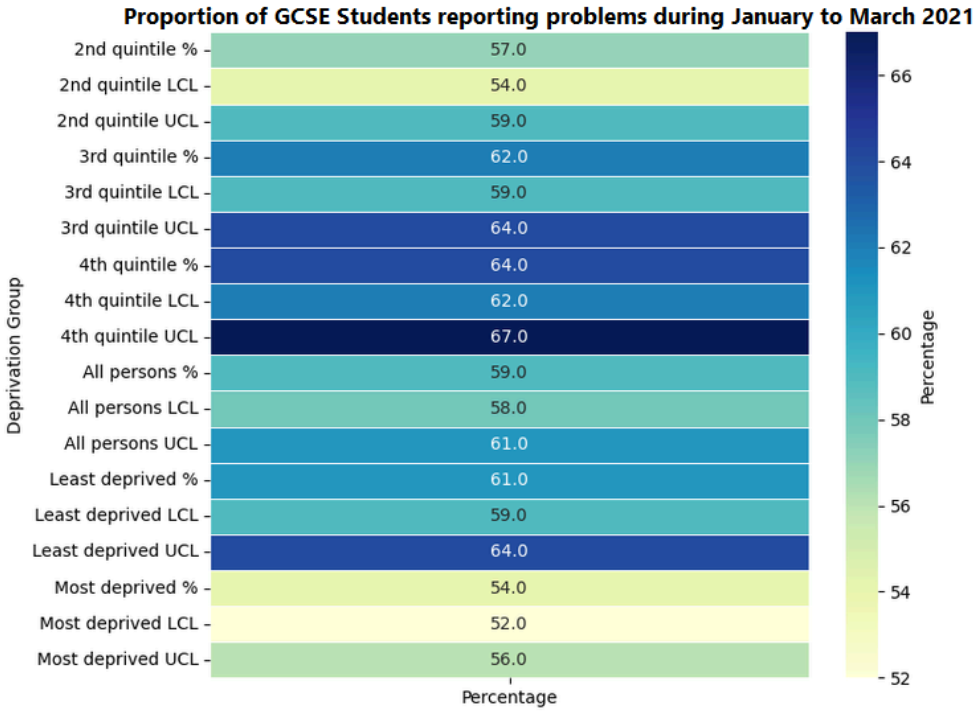
The pandemic significantly disrupted education systems, impacting over 1.6 billion students. Key findings include:

- **Digital Divide:** Over 40% of students globally lacked access to the internet or devices for remote learning, with low-income households most affected. This widened educational inequalities and posed risks for long-term academic outcomes.

- Around 45% of children could only access work via a mobile phone ,



- **Student Mental Health:** Isolation and resource limitations exacerbated mental health challenges among students, further hindering their learning experiences.



Results Conclusion

The findings highlight the multifaceted and interconnected societal impacts of COVID-19. Educational inequalities, labour market disruptions, and mental health challenges were compounded by misinformation and vaccine hesitancy. These issues necessitate targeted, multi-sector interventions, including enhancing digital equity, integrating mental health services into recovery programs, and combating misinformation through improved digital literacy.

Addressing vulnerability to misinformation requires not only improving access to accurate information but also fostering critical thinking and resilience against misinformation through education and media literacy programs. Proactively engaging with communities, particularly those vulnerable and most at risk, and engaging trusted local and public figures to disseminate factual information can rebuild trust in institutions and reduce susceptibility to false narratives. Such interventions will not only mitigate current challenges but also create a more informed, resilient society better equipped to navigate future crises.

Effective visualisation and communication of these insights will be instrumental in guiding policymakers and organisations toward impactful, long-term solutions.

Project Retrospective

An example of some projects previously successfully delivered would have been useful for managing, predicting and planning for expectations and refining our research question.

Although we achieved a lot as a group with our project there are many incomplete data pipelines. On reflection it would have been beneficial to limit our area of research to a more narrow field e.g. COVID and misinformation.

We aimed to integrate more data sources but achieved this for only a few pipelines, limiting our ability to create richer, multi-dimensional visuals by combining complementary datasets.

More time to rationalise findings would have been useful, in order to ensure consistency in reporting and in format both for narrative elements and for visualisations.

For full analyses and visualisations please refer to our extended project report.

Appendix

Appendix 1:
Research questions, hypotheses and proposed visualisations for each subdomain:

Question	Null Hypothesis	Suggested Graphs
MISINFORMATION UK – OFCOM DATASET		
DEVICE USAGE		
What devices were used to access information on the pandemic?	There is no significant distribution of devices used to access information about the pandemic.	Pie Chart: Show the percentage distribution of devices used to access information (smartphones, laptops, tablets, etc.).
NEWS CONSUMPTION FREQUENCY		
Did the frequency of obtaining news change over time during the pandemic?	The frequency of obtaining news did not change over time during the pandemic.	Time Series Plot: Display news engagement frequency (daily, weekly) over time, with key pandemic events and case numbers.
Did the frequency of obtaining news correlate with peaks in mortality or number of cases?	There is no correlation between the frequency of obtaining news and the number of COVID-19 cases.	Time Series Plot: Compare frequency of news engagement with COVID-19 case peaks over the pandemic.
MOST COMMONLY USED NEWS SOURCE		
Which sources were most frequently used to access pandemic news?	There is no significant difference in the frequency of use of different news sources.	Treemap: Display the relative frequency of news sources (e.g., BBC, ITV, Facebook, NHS) based on survey responses.
Did the use of different news sources change over time during the pandemic?	The use of different news sources did not change over time during the pandemic.	Time Series Plot: Show trends in the usage of grouped sources (social media, traditional media, official sources) over time with case peaks highlighted.
TRUST IN NEWS SOURCES		
Were certain news sources regarded as more important or trusted than others?	The perceived importance of different news sources does not vary across sources.	Pie Chart: Compare the percentages of respondents choosing specific sources as the most important or trusted during the pandemic.
Did public trust in the news change over time during the pandemic?	Public trust in the news did not change significantly over time during the pandemic.	Time Series Plot: Illustrate changes in public trust levels over time, alongside case surges and lockdown periods.
EXPOSURE TO MISINFORMATION		

Did the frequency of exposure to fake news change throughout the pandemic?	The frequency of exposure to fake news did not change throughout the pandemic.	Time Series Plot: Show fluctuations in the percentage of respondents exposed to misinformation over time, with major case spikes annotated.
Were there gender or social class-specific patterns in exposure to misinformation?	There are no significant differences in exposure to misinformation claims based on gender or social class.	Heatmap: Display differences in exposure to specific misinformation claims by gender and social class.
SOURCES OF MISINFORMATION		
What were the most commonly used sources for the spreading of misinformation during the COVID-19 Pandemic?	There is no significant difference in the frequency of use of various sources for spreading misinformation during the COVID-19 pandemic.	Wordcloud: Display the frequency of use of reported media source over the survey period
VACCINATION SENTIMENT AND UPTAKE		
Was there a regional difference in feelings towards vaccination, and how did this compare to vaccine uptake?	There is no regional difference in feelings towards vaccination, and willingness does not correlate with vaccine uptake.	Side-by-Side Maps: Compare regional vaccine sentiment with vac
MISINFORMATION WORLDWIDE - ESOC DATASET		
Distribution Channels: Most common platforms.	Distribution channels have equal representation in misinformation sharing.	Bar Chart: Show platform usage frequency (e.g., Facebook, Twitter, WhatsApp) by count.
Narratives and Types: Narrative distribution.	Misinformation narratives are evenly distributed across regions and types.	Stacked Bar Chart: Narrative frequency vs. misinformation type by region.
Temporal Trends: Patterns in misinformation spikes.	No significant temporal patterns exist in misinformation dissemination.	Line Plot: Trend in misinformation volume over time, annotated with key pandemic events (e.g., vaccine rollout).
Regional Influence: Country-specific narratives.	Misinformation narratives are independent of regional or geopolitical factors.	Heatmap: Regional narrative distribution frequency.
Motives and Patterns: Dominance of fear-based motives.	Motives do not vary by narrative or region.	Grouped Bar Chart: Motive frequency by country.
MISINFORMATION WORLDWIDE - AAAI DATASET		
Keyword Trends	No significant difference in keyword frequency between real and fake tweets.	Bar chart and word cloud visualisations to compare top keywords in real vs fake tweets.
Sentiment Analysis	Sentiment distribution (positive, neutral, negative) does not differ between real and fake tweets.	Grouped bar charts comparing sentiment distribution across real and fake tweets.

Impact of Language and Tone	No significant variation in emotional language between real and fake tweets.	Sentiment heatmap or textual analysis of language patterns in real vs fake tweets & Categorised Word Association Network of Real and Fake Keywords
LABOUR MARKET STATISTIC DATASET		
EMPLOYMENT		
Were employment rates affected by the pandemic?	The pandemic did not significantly affect employment rates	Time Series Plot: Plot the overall employment rate over time with number of cases of COVID-19
Was there a difference in employment rates in each sector/industry?	Employment rates were equally affected by the pandemic across all sectors and industries	Time Series Plots: Plot the overall employment rates by sector, allowing comparisons across sectors. Along with the number of COVID-19 cases.
		Stacked Bar Chart: Compare the employment distribution across sectors pre, during and post-pandemic - choose peak number of cases as height of the pandemic
EARNINGS		
Were weekly earnings affected by the pandemic?	Weekly earnings were not significantly affected by the pandemic	Time Series Plot: Compare the change in weekly earnings (£) with number of cases of COVID-19
Was there a difference in weekly earnings across sectors/industries?	The pandemic had the same effect on weekly earnings across all sectors	Time Series Plot: Plot each sectors weekly earnings rate (£) over time with number of cases of COVID-19
	There were no sector-specific differences in the change in weekly earnings during the pandemic	Bar Chart: Display the average weekly earnings (£) across different sectors pre, during and post-pandemic - choose peak number of cases as height of the pandemic
UNEMPLOYMENT		
Were unemployment rates affected by the pandemic?	The pandemic did not cause any significant change in unemployment rates	Time Series Plot: Plot the overall employment rate for the UK over time with number of cases of COVID-19 to visualize any trends or changes
Was there a difference in regional unemployment?	Unemployment rates across all regions were equally affected by the pandemic	Time Series Plots: Plot the overall unemployment rates by region. Along with the number of COVID-19 cases.
		Stacked Bar Chart: Compare the unemployment distribution across regions pre, during and post-pandemic - choose peak number of cases as height of the pandemic

Was there a difference in unemployment for each age?	The pandemic had no effect on unemployment rates for any age group	Time Series Plots: Plot the overall unemployment rates by age group. Along with the number of COVID-19 cases.
		Stacked Bar Chart: Compare the unemployment rate across age groups pre, during and post-pandemic - choose peak number of cases as height of the pandemic
Was there a difference in unemployment for each gender?	The pandemic had no effect on unemployment rates for any gender	Time Series Plots: Plot the overall unemployment rates by gender. Along with the number of COVID-19 cases.
		Stacked Bar Chart: Compare the unemployment rate across genders pre, during and post-pandemic - choose peak number of cases as height of the pandemic
MENTAL HEALTH – NHS & ONS Wellbeing Data		
How did the prevalence of anxiety and depression change during the COVID-19 pandemic compared to before and how did levels of wellbeing evolve?	The pandemic had no significant impact on the wellbeing of the country.	Line graph charting the changes in response to the ONS wellbeing survey. Pie charts to show a snapshot of the distributions of responses
How were key workers affected in comparison to non-key workers and the unemployed?	Key workers were not affected any differently to those in other states of employment.	Bar graph to display changes in reported feelings of depression in key workers compared to others.
How was the mental health of children affected?	The mental health in those under 19 was unaffected.	Line graphs and bar graph charting the changes in mental health before, during and after the pandemic.
What were the most common mental health issues reported during the pandemic?	There were no differences to the mental health issues reported during the pandemic.	Bar graph showing the percentage increase in different mental health issues related to psychosocial factors.
Were there notable changes in the suicide rate pre, during, and post pandemic?	There were no notable differences in suicide rate.	Line graph charting the changes in overall suicide rate.
Vaccine Hesitancy Analysis (ONS Dataset)		
Predictors of Hesitancy:	Predictors have no significant association with vaccine hesitancy.	Bar Chart: Hesitancy-to-Sentiment proportions.
Positive Sentiment:	Predictors are not correlated with positive vaccine sentiment.	Sankey Diagram: Predictors → Outcomes (Sentiment & Hesitancy).

Temporal Trends:	Vaccine hesitancy does not fluctuate over time.	Line Plot, Stacked Area Chart, Bar & Grouped Bar Chart: Trends in weighted totals by predictor over time.
Category Distribution (e.g., Income):	Vaccine hesitancy is not influenced by demographic categories.	Circular Barplot: Predictor categories and hesitancy/positive sentiment.
Cross-category Comparisons:	Mental health and income categories are equally associated with hesitancy.	Dumbbell Plot: Comparison of hesitancy vs. sentiment by category.
Education		
How did the Covid-19 lockdowns affect GCSE students in the first and last lockdowns in terms of academic performance and engagement across different levels of deprivation?	The covid-19 lockdown did not disproportionately impact GCSE students from deprived backgrounds which could lead to lower results and more mental health problems	Heatmaps to show relationships between variables

Appendix 2:
Proposed file structure:

```
.
├── 1_raw_files
├── 2_api_data_pipeline_notebooks
├── 3_raw_database_to_csv_parsing_notebooks
├── 4_integrated_csv_files
├── 5_cleaning_notebooks
├── 6_cleaned_files
├── 7_analysis _and_visualisation_notebooks
├── 8_final_visualisations
├── 9_reports
└── README.md
```

Appendix 3:
Table detailing the role assignment

Role	Responsibility	Assigned Team Member(s)
Team Leader	Coordination, Documentation, and Role assignment	Zenzi Mansaray
Data Engineer	Data Cleaning and Integration	ALL
API Specialists	Fetching Data via APIs/Data Sources	Alexandra Howland Natalie Ellis
Analyst(s) 1	Misinformation and Media Analysts - Worldwide	Heledd Davies
Analyst 2	Misinformation and Media Analysts - UK	Alexandra Howland

		Heledd Davies
Analyst 3	Mental Health Analyst	Elizabeth (Lisa) Franchetti
Analyst 4	Employment and Economic Impact Analyst	Hayley Lawrence
Analyst 5	Education Analyst	Zenzi Mansaray
Analyst(s) 6	Vaccination Hesitancy & Vaccine Adoption Analysts	Heledd Davies Hafsa Hussain
Report Manager	Drawing Conclusions from the Data & Writing a Narrative.	Alexandra Howland Heledd Davies

Appendix 4:
Table outlining proposed project timeline
Week 1: Project Selection

Day	Task	Team Member(s)	Output
1-6	Discuss Project Ideas	All	Individual proposals for
7	Decide Upon Research Question	All	Finalised Research Question

Week 2: Project Refinement and Data Collection

Day	Task	Team Member(s)	Output
1	Kickoff Meeting	All	Task breakdown, timeline agreement
2-7	Data Collection	API Specialist, Data Engineer	Access to the data in a useable format, initial clean of the dataset where necessary

Week 3: Data Cleaning & Visualisation

Day	Task	Team Member(s)	Output
1-2	Initial Data Cleaning	Data Engineer & Analysts	Pre-processed dataset
3-4	Exploratory Analysis	API Specialist	Initial visualizations to understand trends in cases, fatalities and vaccinations
5-7	Filtering of Datasets, Finalising the Data Selection and Final Clean	All	Data Ready for Analysis & Visualisation

Week 4: In Depth Analysis & Finalizing Visualisations

Day	Task	Team Member(s)	Output
1-4	Misinformation Analysis: Analyse misinformation types, motives, and distribution channels	Data Engineer & Analyst 1	Initial visualisations to understand trends in misinformation correlated with cases/fatalities and vaccination rate
1-4	Mental Health Analysis: Correlate mental health metrics with COVID-19 trends	Analyst 2	Initial visualisations to understand trends in mental health correlated with

			cases/fatalities and vaccination rate
1-4	Employment Analysis: Correlate Employment Data with COVID-19 trends	Analyst 3	Initial visualisations to understand trends in employment correlated with cases/fatalities and vaccination rate
1-4	Education Analysis: Correlate Educational Data with COVID-19 Trends	Analyst 4	Initial visualisations to understand relationships between educational parameters correlated with cases/fatalities and vaccination rate
5-7	Finalising visualisations	All	Final visualisations

Week 5: Finalization of Report and Presentation Development

Day	**Task**	**Team Member(s)**	**Output**
1-5	Documentation & Report	Analysts, Report Manager	Draft Report
1-5	Presentation Prep	All	Polished Presentation
6-7	Final Presentation	All	Completed Project