

Evaluating events data for cultural analytics: a case study on the economic and social effects of Covid-19 on the Edinburgh Festivals

Suzanne R. Black, Rosa Filgueira, Lesley McAra, Brendan Miles, Mark Parsons & Melissa Terras

To cite this article: Suzanne R. Black, Rosa Filgueira, Lesley McAra, Brendan Miles, Mark Parsons & Melissa Terras (07 Feb 2024): Evaluating events data for cultural analytics: a case study on the economic and social effects of Covid-19 on the Edinburgh Festivals, Creative Industries Journal, DOI: [10.1080/17510694.2024.2311883](https://doi.org/10.1080/17510694.2024.2311883)

To link to this article: <https://doi.org/10.1080/17510694.2024.2311883>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 07 Feb 2024.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)

Evaluating events data for cultural analytics: a case study on the economic and social effects of Covid-19 on the Edinburgh Festivals

Suzanne R. Black^a , Rosa Filgueira^b , Lesley McAra^a, Brendan Miles^c, Mark Parsons^a  and Melissa Terras^a 

^aUniversity of Edinburgh, Edinburgh, UK; ^bUniversity of St Andrews, St Andrews, UK; ^cData Thistle, Edinburgh, UK

ABSTRACT

The effects of the Covid-19 pandemic on the Creative and Cultural Industries can be difficult to quantify. Metadata about events (theatre productions, music and comedy gigs, sporting fixtures, days out, and more) are an untapped resource for cultural analytics that can be used as a proxy metric for financial and social impact. This article uses a sample of large-scale cultural events data from UK industry providers Data Thistle to ask: how can events data at scale be used to quantify the financial and social effects of the Covid-19 pandemic on the cultural events sector in a particular region? We analysed the changes in event provision in Edinburgh in August 2018, 2019, 2020 and 2021, revealing an estimated 97.3% fall in ticketing revenue between 2019 and 2020. Additionally, the effects that pandemic restrictions had on different categories of event reveal a disparity in how different audience sectors were affected, with 'Visual Art' and 'Days Out' showing most resilience and 'Theatre', 'Comedy' and 'LGBT' events being most reduced. Our findings indicate that events data are a rich but heterogenous source of information regarding the cultural and creative economy, which is not yet routinely used by researchers.

ARTICLE HISTORY

Received 17 November 2022

Accepted 24 January 2024


KEYWORDS

Cultural value; data science; creative industries; cultural analytics; events data; lgbtqia+

1. Introduction

The UK has world-leading Creative and Cultural Industries (CCI), defined by UNESCO as conducting 'activities whose principal purpose is production or reproduction, promotion, distribution or commercialization of goods, services and activities of a cultural, artistic or heritage-related nature' across the domains of advertising, architecture, books, gaming, music, movies, newspapers and magazines, performing arts, radio, tv, and the visual arts (Lhermitte, Perrin, and Blanc 2015, 11). As well as contributing to a multi-billion-pound economy, access to arts and culture is related to issues of social

CONTACT Melissa Terras  m.terras@ed.ac.uk  University of Edinburgh, Edinburgh, UK

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/17510694.2024.2311883>.

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

inequality, health and wellbeing, and public infrastructure (Crossick and Kaszynska 2016, 7–8).

This heterogenous value of culture, both financial and social, has understandably proven difficult to quantify. A major report commissioned by the Arts and Humanities Research Council published in 2016 concludes that ‘[a]ccounting for human experiences of art and culture calls for multi-criteria analyses and a range of approaches, in order to span the depth and the breadth of research’ (Crossick and Kaszynska 2016, 10). A range of approaches has been implemented, ranging from financial analysis of individual sectors to quantify the economic value of CCI, to work that falls under the umbrella of cultural sociology, which attends to ‘politics, religion, social movements, race, civic engagement, and a variety of other social processes – from a cultural perspective’ (Alexander, Jacobs, and Smith 2017). This research by cultural sociologists works to tease out other forms of societal impact beyond the strictly financial.

The global Covid-19 pandemic has brought the importance of understanding the financial and social values of CCI into focus, for example, in determining the effectiveness of pandemic restrictions on the cultural workforce and how to better plan for future crises (Walmsley et al. 2022). The pandemic also caused changes to the cultural landscape that will have to be accounted for in future research, such as the reduction in prominence of revenue from live music (BBC News 2022) and the increased role of online activities and streamed performances (Bradbury et al. 2021, 4). These changes have only made it more important and more difficult to understand and measure the financial and social impacts of the pandemic on the CCI.

Before the pandemic, Hanquinet et al. espoused the value of ‘the information held by commercial organisations about consumer behaviours, particularly around consumption practices’ as ‘offer[ing] much more detailed, fine-grained, information on the social world compared with surveys and face-to-face interviews’ (Hanquinet et al. 2019, 199). Every year there are hundreds of thousands of events, festivals, concerts, plays and gigs, varying in scale from the very small and informal to the large and coordinated. The details of these events constitute a huge amount of data about events that have taken place (historical events) and events that are planned for the future (live events). Metadata about these events are an untapped resource that can be used as a proxy metric for financial and social impact.

This article addresses the usefulness of cultural events data in understanding the cultural and social value of CCI. To do this, we evaluate both an available source of events data (from the commercial providers Data Thistle) and large-scale data analysis and visualisation, using Python and pandas DataFrames, to ask: how can events data at scale be used to quantify the effect of the Covid-19 pandemic on the cultural events sector in a particular region? As we are assessing the suitability of a data source and a method – with the aim of providing guidance to other researchers working in this area – we take care to explain the decisions made when working with these data and their repercussions. After a detailed explanation of these data and how we approached them, there follow two examples of how cultural events data could be used. We calculate the estimated revenue difference between Edinburgh in August before and during the Covid-19 pandemic by creating an algorithm for estimating the revenue generated by ticket sales. We then compare the differing effects that pandemic restrictions had on different categories of event to go beyond

financial impact and draw inferences about social consequences that would otherwise be imperceptible. We demonstrate that events data are a rich but heterogeneous source of information regarding the cultural and creative economy and, although not yet routinely used by researchers, hold much future potential. Our findings augment the wealth of work that has been done to ascertain how the cultural sector in the UK was impacted by the pandemic, financially and otherwise. The key intervention here is to add another facet to the myriad approaches to quantifying and evaluating cultural value.

2. The role of cultural data in UK research

2.1. Determining cultural value

Determining cultural value, whether financial or otherwise, is not easy. Before the Covid-19 pandemic, estimates of the financial contribution that the Creative and Culture Industries make to the UK's economy painted a healthy picture, with official figures showing that the Creative Industries contributed £111.7 billion to the UK in 2018 (Department for Digital, Culture, Media & Sport, 2020) and £115.9 billion in 2019 (Department for Digital, Culture, and Media & Sport 2021). Beyond Gross Value Added (GVA), the government has been seeking ways to quantify cultural value more broadly. The Department for Digital, Culture, Media & Sport commissioned a report on how to quantify the value of the culture and heritage sectors with the aim of 'develop[ing] a formal approach to value the benefits of culture and heritage assets to society... to create publicly available statistics and guidance that will allow for more accurate articulation of the value of services provided by culture and heritage' (Sagger, Philips, and Haque 2021, 5). From academia there have also been calls to better assess cultural value. A project commissioned by the Arts and Humanities Research Council, 'The AHRC Cultural Value Project' (Crossick and Kaszynska 2016) sought to 'identify the various components that make up cultural value' and 'consider and develop the methodologies and the evidence that might be used to evaluate these components of cultural value' (6). This call for a better understanding of the multiple values of culture also comes from the public body Arts Council England, which describes the reach of arts and culture into society, wellbeing, education and the economy, and calls for better methods to 'establish causality between arts and culture and the wider societal impacts' (Arts Council England 2014, 8). There is work underway focusing on specific methods for quantifying the social impact of arts and culture, such as using wellbeing data (Oman 2021), and the pandemic has made this work all the more prescient.

2.2. Consequences of the Covid-19 pandemic

We know that the pandemic had a huge effect on business revenue, especially on the arts, culture and heritage sectors, after the UK government announced a stay-at-home order and the closing of all non-essential businesses on 24 March 2020 (BBC News 2020a). There were various timetables for arts venues to reopen across the UK in 2020 and 2021, although these often involved Covid-19 mitigation measures like

reduced capacities, which meant that venues were not necessarily able to generate enough revenue to be financially viable (BBC News 2020b; News Editor 2020; Telegraph Reporters 2021).

Globally, these measures have had a significant financial impact. The World Economic Forum has found that '[t]en million creative jobs have been lost because of the pandemic' (Bateman 2022). UNESCO found that most countries reported a fall in Gross Value Added over 2020 in the Cultural and Creative Industries, and that this fall was larger than the fall in their national economies overall, with the loss of revenue from CCI amounting to approximately 30–40% across different countries (Naylor et al. 2021, 12). In the UK, official GVA figures from DCMS for 2020 and 2021 are not yet available (Department for Digital, Culture, and Media & Sport 2022), although multiple reports have collated figures for specific aspects of the cultural sector.

A report collating figures from multiple government and industry sources estimates that in the creative, arts and entertainment activities sector, '[b]etween Q4 2019 (before the pandemic in the UK) and Q2 2021 (the most recent data to date), output declined by 37% in real terms in the sector in the UK', and that the music industry suffered a 46% fall in GVA contribution, and theatres a £630 million loss of income (Waitzman 2021). Further work on the economic effects of the Covid-19 pandemic has quantified this revenue loss for specific geographic regions (Chamberlain and Morris 2021), specific industries (Anayi et al. 2021; Clark 2022) and globally (Betzler et al. 2021). Focusing specifically on Edinburgh, The Edinburgh Festival Fringe Society estimate that the Edinburgh Festival Fringe brings between £200m and £1bn to Scotland and the UK each year, and that cancelling in-person shows at the Fringe in 2020 resulted in a revenue gap for the Fringe Society of £1.5m and for Fringe venues of £21m (Edinburgh Festival Fringe Society 2020).

The impact of the Covid-19 pandemic on the CCI extends beyond cancelled performances, refunded tickets and economic loss. As anthropologist Juliet Bedford writes:

No epidemic is ever just a health issue in isolation, and Covid-19 has emphasised this on the global stage... We need to be looking at it in terms of an economic issue, a livelihood issue, a social issue and a political issue too. (Wellcome 2020)

Reflections on the pandemic have revealed the effects it has had on the Cultural and Creative Industries themselves, like inequalities and precarities in the sector's working conditions (Comunian and England 2020; Walmsley et al. 2022) and how audiences' consumption of culture has changed, for example with the move to digital modes of engagement (Bradbury et al. 2021; Denk et al. 2022; Leung and Davies 2021; BBC News 2022).

2.3. Inequalities

Covid-related changes have not been experienced equally among all sectors of society. UK Research and Innovation (UKRI) has made efforts to focus research on the effects of the pandemic on sectors of society that may be impacted disproportionately, such as children, older people, people with disabilities, and LGBTQIA+ communities (Arts and Humanities Research Council 2021; Economic and Social Research Council 2021a; Economic and Social Research Council 2021b; Economic and Social Research Council

& Innovate UK, 2021). Other research finds that '[w]omen are affected more than men by the social and economic effects of infectious-disease outbreaks' because of greater caring responsibilities, less secure employment contracts and risk of domestic violence (Wenham et al. 2020, 194). Existing work on the effect of the pandemic on children and young people focuses on direct consequences like bereavement (Hillis et al. 2021) but also on the repercussions of mitigation efforts like home schooling (Deeker 2022). Studying the consequences of children's reduced access to arts and culture during this time could help us to better understand the ways in which children and young people have specific vulnerabilities during times of international crises. It has also been recognised that LGBTQIA+ communities, as seen in this early call by the charity Stonewall for renewed attention to data collection, have specific vulnerabilities (Munir 2020). In 2018, a survey of LGBTQIA+ individuals in the UK found that the respondents were 'less satisfied with their life than the general UK population (rating satisfaction 6.5 on average out of 10 compared with 7.7)' and that '[t]rans respondents had particularly low scores (around 5.4 out of 10)' (Government Equalities Office, 2018). Inequalities in participation in arts and culture are not new. Orian Brook et al. discern 'deeply unequal patterns of consumption and production' across multiple axes of oppression, including gender, race and class (Brook et al. 2020, 53). Better data about event provision where it intersects with minoritised sectors of society would enable a better understanding of how inequalities in access to arts and culture are experienced, their effects, and suggestions for policy and action to remedy this disparity.

2.4. Cultural analytics with events data

The interdisciplinary approach of cultural analytics, defined by Lev Manovich as 'the use of computational and design methods for exploration and analysis of contemporary global culture at scale' (Manovich 2020, 35) positions the metadata generated from the Creative and Cultural Industries as a valuable data stream. In his endeavour to 'work toward a more inclusive and democratic understanding of the cultural present and also of cultural histories' (9), Manovich hones in on events as 'cultural happenings that have duration in time and involve multiple people' (75). Existing work in UK universities that aligns with Manovich's description of cultural analytics often falls into the category of cultural sociology or the sociology of culture, which attend to 'culture industries and cultural consumption' and 'the entirety of social life... from a cultural perspective' (Alexander, Jacobs, and Smith 2017), respectively (see Brook, O'Brien, and Taylor 2020; Gross and Pitts 2016; Hadley et al. 2019; Hanquinet, O'Brien, and Taylor 2019; McAndrew, O'Brien, and Taylor 2020). In their comparison of the uses of survey data and social transactional data like ticketing data, Hanquinet et al. promote the utility of 'transactional data being particularly useful for shedding light on activities that are hard to gain insight into from surveys' like the Taking Part survey of cultural participation in England or Active Lives (Hanquinet et al. 2019, 201, 214). One such source of transactional data is events listings.

The data for the analyses in this article have been provided by Data Thistle (Data Thistle, n.d.), an Edinburgh-based listings technology business that provides live events data and has been in operation since 1985 (previously known as The List). They collate

'What's On' data about cultural events (theatre shows, music performances, art exhibitions, sporting fixtures, club nights, author events, nature walks and more) from over 10,000 venues across the UK. They make the event name, category, venue, date, time, event description, website links, accompanying images, prices, addresses and geo-location data available to be queried on their public 'What's On' website as well as providing data to users such as destination marketing organisations, transport companies and hotels. Data Thistle currently provide access to live events data (that is, events scheduled in the future) via widgets, spreadsheets, calendars, bespoke JSON feeds and an API.

Data Thistle are also constantly accruing a long tail of historical events data, as they operate on the understanding that they are creating a cultural record of the UK for posterity and so retain data pertaining to past events. This rich picture of cultural provision for the UK remains mostly untapped and under-researched. A few projects have begun to utilise these data. One such project is the Festivals & Communities Map (Currie and Correa 2021; Currie and Correa 2022), which combines data on: Edinburgh Festivals' community outreach efforts; Edinburgh Festival Fringe venues; ticket sales across the Fringe, Edinburgh International Festival and Edinburgh Book Festival; bike routes; the Scottish Index of Multiple Deprivation; and year-round venues. The year-round venues data were supplied by Data Thistle. Plotting all of these geospatially allows questions about event provision and equality to be interrogated. Another project that has found success by combining data from different sources is the Festival Mobility project (Ryan-Saha 2020a; Ryan-Saha 2020b). The project was designed with the aim of investigating how to combine events and transport data to better understand festival-related traffic and congestion in Edinburgh during August. When the Covid-19 pandemic happened the project pivoted to investigating festival infrastructure in Edinburgh from the perspective of public health (Ali-Knight 2021). Each of these pieces of work builds on Data Thistle's data and enriches it by adding data from other sources to extend the questions that can be asked of the data.

Combining events data with data from other sources, such as which Covid-19 restrictions were in place, the provision of public transport, or the number of tickets sold to each of these performances, enables many more questions to be asked. This would require access to further open and commercially-licensed data, which is beyond the scope of this case study, but we flag up here that the richness of the What's On data may be best explored in collaboration with other datasets.

3. Cultural events DATA

3.1. *Materials: a cultural events dataset*

We sourced live cultural events data under licensed agreement from Data Thistle, a UK based commercial company (Data Thistle, n.d.). Data Thistle maintain comprehensive listings of cultural events in the UK – what's on, where and when – with each event containing, at minimum, a title and a location including latitude/longitude and postcode. Data Thistle provided us with a sample of events data: the 'Data Thistle Dataset' (DTD) delimited by postcodes within the Edinburgh and South-East Scotland City Region Deal¹ and taking place between November 2017

and May 2022. The DTD is approximately 10% of the whole UK dataset for that time period.

The DTD is comprised of nine JSON files totalling 153MB and contains machine processable text information about 38,700 events, 48,790 schedules and 349,855 performances at 2474 places. The sample does not contain any online events. Each event in the dataset is allocated to 1 of 15 discrete genre categories ('Film', 'Music', etc) and events can also have tags conveying further information. The dataset contains 2175 unique tags used a total of 97,532 times. The distinctions between events, schedules and performances, and categories and tags are described below. The ontology of the data – how categories, tags and other values are defined and implemented, and whether they are used consistently – affects how we can query the data and the related findings.

3.2. Data complexities

Before asking specific questions of the Data Thistle Dataset, our first step was to establish how many events took place where, and what type of event they were. This brought with it complexities relating to two characteristics of the data taxonomy, which needed to be understood before conducting any analysis: 1) the difference between *events*, *schedules* and *performances*, and 2) the difference between *categories* and *tags*.

3.2.1. Events, schedules and performances

In the DTD:

- An *event* refers to the characteristics of a potential event. An event also needs a schedule and performance for it to be realised.

For example, a concert by the band Biffy Clyro is an event.

- A *schedule* is the range of instances of an event at a particular place.

For example, Biffy Clyro at Murrayfield Stadium on 3, 4 and 5 March 2023, and at Hampden Park on 6 and 7 March 2023 are two schedules.

- A *performance* is one instance of an event at a particular time and place.

For example, Biffy Clyro at Murrayfield Stadium on 5 March 2023 at 8pm is one performance.

Some implications of events versus schedules versus performances are:

- Working at the *event* level allows an overview of events without the specifics of time, place, ticket prices or performance-specific properties, such as whether a performance provides British Sign Language interpreting or if a film is screened in 3D.
- Working at the *schedule* level allows for time spans and places, but not ticket prices or performance-specific properties.

- Working at the *performance* level includes details of time, place, ticket prices and performance-specific properties. Information that may be pertinent to a particular research question is located in performance-specific properties, for example, whether an event is cancelled or sold out.

As working with the data at event, schedule or performance level will generate different results, this choice will be determined by the research question. For example, for an analysis of event descriptions using Natural Language Processing, the event level may be sufficient. Due to the complexities of this dataset, researchers working with it must have clear understanding of the decisions that are required to generate robust findings.

3.2.2. Categories and tags

Another aspect of the data required for analysis is the difference between *categories* and *tags*. Each event is assigned to one of 15 *categories* within the Data Thistle ontology: 'Film', 'Music', 'Days Out', 'Theatre', 'Kids', 'Comedy', 'LGBT' (Lesbian, Gay, Bisexual and Transgender²), 'Clubs', 'Visual Art', 'Dance', 'Books', 'Sport', 'Food & Drink', 'Talks & Lectures' and 'Workshops'. There is no data definition of these categories, but their scope can be inferred from the events in each category as described in Data Thistle's API documentation.³ Events are assigned only one category from a controlled vocabulary of the 15 categories listed above. These are chosen and inputted by event creators (event organisers, venues or promoters) and, when the data is imported, the given categories either match one of Data Thistle's categories or are mapped onto Data Thistle's ontology algorithmically using the closest match.

Events can also be assigned multiple *tags* and, rather than using a controlled vocabulary, these constitute a crowd-sourced list of descriptors generated by the event creators. There is no real limit on the number of tags an event can have as Data Thistle do not place any restrictions on this free text field. There are 2175 unique tags in the dataset and they function to add extra or more nuanced information about events, so an event categorised as 'Music' may also be tagged as 'Folk' to provide more detail about the genre of music. Tags become important when searching for all events relevant to a specific query. For example, an event that borders two categories, such as a musical theatre performance, may be categorised as 'Music' and tagged as 'Theatre' or vice versa. Tags are especially important for the categories of 'Kids' and 'LGBT' since these are not strictly genres of event but pertain to the audience for an event: an event aimed at children or LGBTQIA+ communities will also fall into another category, such as a kids' nature walk in 'Days Out', or a drag show in 'Theatre'. Formulating search queries to extract a sample of the DTD in order to answer specific research questions therefore requires an understanding of the difference between, and interplay of, categories and tags.

3.3. Method: notebooks

With our initial explorations of the DTD structure in mind, we ingested the nine JSON files that comprise the DTD using pandas DataFrames into a Python 3.9 environment,

created entity diagrams to understand the structure of the dataset and its main features (see Appendices A & B), and created two Jupyter Notebooks: one to generate DataFrames from the JSON files and one to enable a variety of analyses to be conducted on the DTD or subsets of it delimited by city, month and category of event.

3.3.1. Notebook one: *generate_list_dataframes*

The *Generate_List_Dataframes*⁴ notebook automatically ingests JSON files and processes them so the information can be analysed more easily by merging events that appear in more than one file, removing duplication, and ‘exploding’ nested lists of dictionaries to create several DataFrames with the information ‘separated’ at different levels (events, schedules, performances, tickets, places, etc.) and saving them to disk for use with the *Case_Study_Maker* notebook. This notebook also includes preparatory work to enable the estimation of revenue at the schedule level, which is discussed in more detail in [section 3.4](#).

3.3.2. Notebook two: *case_study_maker*

In the *Case_Study_Maker*⁵ notebook we developed a new method to perform custom analyses at run-time. It automatically loads and ingests the DataFrames generated in the previous notebook and generates different types of analyses (frequency, histograms, Gantt charts, maps, etc.) using the Plotly open-source interactive graphing library for Python at different levels (events, schedules, performances, tickets). This notebook requires three parameters from users: 1) *city* (e.g. ‘Edinburgh’); 2) list of *categories* to analyse (e.g. ‘Music’, ‘Visual Art’, ‘Film’, ‘Books’); 3) *month* to create a more focused analysis over the years (e.g. ‘August’). It enables the generation of consistent findings from the DTD and comparisons between subsets of the DTD.

These notebooks allowed us to generate the findings presented here, but could also be used in future research projects on any datasets supplied by Data Thistle.

3.4. Method: *revenue calculation*

For calculating the estimated revenue at the schedule level we had to establish a method. After workshoping an approach with Data Thistle, we applied the following steps (these are calculated automatically by running the *Generate_List_Dataframes* notebook). For all schedules, we retained the performances that we know have not been cancelled (this an attribute of performances properties, and some performances have that information available). For each of these performances, we obtained their ticket types (e.g. Standard, Concession, etc.). Each of those ticket types can have several prices (*max_price* and *min_price*). So, first we calculate the **ticket price per performance (steps 1a and 1b)** to obtain the revenue of each performance ticket. This involves:

- **Step 1a:** obtaining the ‘**list of performances ticket type price**’. For each ticket type price, we use the *max_price* value (which is a ticket’s attribute, as we can

see in the Events E-R Diagram shown in [Appendix A](#)) if that value is greater than or equal to zero. If *max_price* is equal to zero, we use *min_price* value (which is another attribute of tickets). This list has as many elements as ticket types per performance.

- **Step 1b:** obtaining the '**ticket price per performance**' by calculating the trimmed mean of the **list of performances ticket type prices** (from step 1a). We trim the 10% extreme scores (lowest and highest).

Once we have the **ticket price per performance** (with the revenue of each performance ticket), we calculate the '**performance revenue**' (**step 2**), which is all the revenue obtained by each performance.

- **Step 2:** If a performance is **sold out** then, the **performance revenue** is equal to **place capacity** (where the performance is being scheduled) multiplied by the **ticket price per performance** (calculated in step 1b). The place capacity is also an attribute of a place's properties, as we can see in the Places E-R Diagram (see [Appendix B](#)).
- If the performance is not sold out, the **performance revenue** is equal to **place capacity** multiplied by the **performance ticket price** (calculated in step 1b) and multiplied by a **capacity factor**.

For the purposes of these estimations we have set the capacity factor to **0.7**, that is, we have made the assumption that if a performance is not sold out, it will sell 70% of the maximum capacity of the venue. However, we have made adjustments for discrepancies potentially arising from high-capacity venues like sports stadia. If performances are not categorised as 'Music' but they are in places with high capacity (>20,000) then the capacity factor is set to **0.2**. Also, for performances categorised as 'Sport' with high capacity (>20,000), their capacity factor is set to **0.4**. We highlight below that these assumptions will be variable depending on the dataset, and other applications of this method should make decisions regarding these variables accordingly.

The final step is to calculate the **schedule revenue** (**step 3**). For each schedule we added the performance revenues (calculated in step 2). In this step we removed the performances designated as cancelled. See [Figure 1](#) for an algorithm combining the steps described above for estimating the revenue from each schedule.

This method was used in the following section to estimate the revenue generated from cultural events in Edinburgh in August 2018, 2019, 2020 and 2021.

4. Estimating revenue lost due to covid-19 pandemic: findings and discussion

In this section we investigate an approach to quantify the financial impact of the Covid-19 pandemic on cultural revenue using the method described in the previous section, which uses ticket prices, venue capacities and an assumption about the percentage of available tickets that are sold. This method is tested on data for Edinburgh during August 2018, 2019, 2020 and 2021, the city's busiest month for

Algorithm 1: Estimation of Each Schedule Revenue

```

1 for  $\forall$  (schedule) do
2   for  $\forall$  (performance)  $\neq$  cancelled do
3     for  $\forall$  (ticket) do
4       list_performances_ticket_type_price=[]
5       for  $\forall$  (ticket_type) do
6         if max_price  $\neq$  0.0 then
7           performance_ticket_type_price = max_price
8         else
9           performance_ticket_type_price = min_price
10        end
11        (step 1.a)  $\rightarrow$ 
12        list_performances_ticket_type_price.add(performance_ticket_type_price)
13      end
14      (step 1.b)  $\rightarrow$ 
15      ticket_price_per_performance = trim_mean
16      (list_performances_ticket_type_price)
17    end
18    (step 2)  $\rightarrow$ 
19    if performance == sold_out then
20      performance_revenue = place_capacity *
21      ticket_price_per_performance
22    else
23      capacity_factor = 0.7
24      if performance_category  $\neq$  music & (place_capacity > 20.000)
25        then
26          capacity_factor = 0.2
27        end
28      if performance_category == sport & (place_capacity > 20.000)
29        then
30          capacity_factor = 0.4
31        end
32      performance_revenue = place_capacity * capacity_factor *
33      ticket_price_per_performance
34    end
35  end
36 end
37 (step 3)  $\rightarrow$ 
38 schedule_revenue =  $\Sigma$ (performance_revenue)
39 end

```

Figure 1. Algorithm for calculating schedule revenue.

events. We arrive at some estimated figures for the revenue generated from tickets sales in those months. As this is exploratory work, we extensively work through the steps involved in this method, the assumptions this method relies upon, any issues arising from the data, and the work that would need to be undertaken on the DTD to achieve more accurate results.

4.1. Revenue findings

Using the method described in section 3.4, we estimated and visualised the revenue generated from each category of event for Edinburgh in August 2018, 2019,

2020 and 2021 in the Data Thistle Dataset. By just looking at these schedules, we can estimate the financial impact that the pandemic has had on Edinburgh’s event revenue. We chose August as a comparison point as the majority of Edinburgh Festivals happen that month (Edinburgh International Festival, Edinburgh Festival Fringe, Edinburgh Art Festival, Edinburgh International Book Festival, and the Royal Edinburgh Military Tattoo) as well as other events. As can be seen more clearly in Figure 2 and Table 1 below, the estimated revenues from ticket sales of in-person schedules in Edinburgh in just August 2018, 2019, 2020 and 2021 are as follows:

Figure 2 shows the estimated revenue generated from schedules in Edinburgh in August 2018, 2019, 2020 and 2021 per category, and the combined figures are given in Table 1.

These figures show an expected drop-off from approx. £18.3m in 2019, before the Covid-19 pandemic, to approx. £0.5m in 2020, when pandemic restrictions were in full effect, and a slight resurgence to approx. £6.2m in 2021, when some of the restrictions were lifted. However, there is an unexpected difference between the 2018 and 2019 revenues of approx. £2.9m. Rather than reflecting any real-world circumstance, several key issues with the data are affecting these results. In the following paragraphs, we enumerate these issues and suggest how to mitigate them in order to generate more reliable figures.

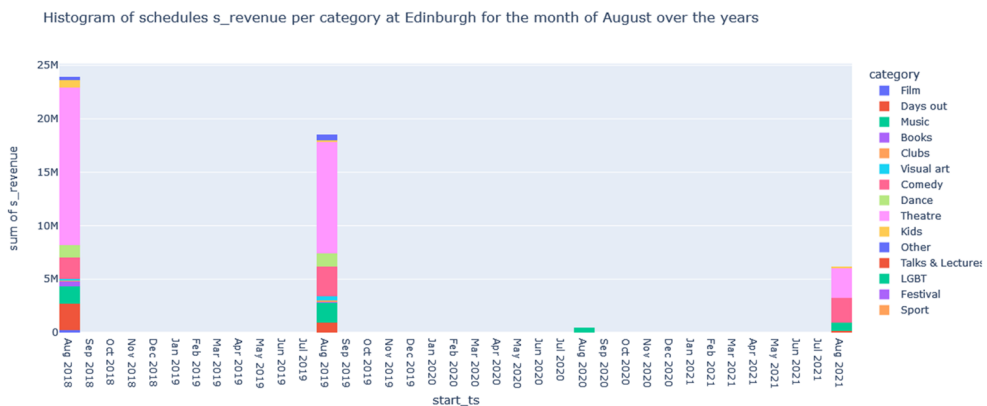


Figure 2. Revenue for all schedules in Edinburgh for Jul for August 2018, 2019, 2020 and 2021 by category.

Table 1. Estimated revenue for schedules in Edinburgh in August 2018, 2019, 2020 and 2021, and percentage change from 2019 to 2020, and 2019–2021.

	2018	2019	2020	2021	Revenue change 2019–2020	% change 2019–2020	Revenue change 2019–2021	% change 2019–2021
Estimated revenue (in £s)	24,200,040	18,263,400	492,512	6,216,536	−17,770,888	−97.3	−12,046,864	−66.0%

4.2. Revenue discussion

4.2.1. Assumptions

Beyond assuming that ticket availability can act as a proxy for cultural presence and consumption, the largest assumption in our method is the capacity factor. We have chosen a capacity factor of 0.7 as theatre industry figures suggest that ticket sales of around 50–80% of occupancy are required to break even. This is based on claims that '[m]ost major [West End] shows need around 70 to 80 per cent of seats filled merely to break even' (Telegraph Reporters 2021), '50% to 70% occupancy is typically needed to break even' (UK Theatre & Society of London Theatre 2020), and '[t]he ability of venues to absorb the constraints varies widely, but the Society of London Theatre (Solt), an industry body, has said social distancing measures mean auditoriums will have only between 15 and 30 per cent of seats available to sell, far below the 65 per cent or more capacity required by most productions to break even' (Pickford 2020). We acknowledge that not every performance will break even and that there are other factors in play, with various events subsidised with Arts Council funding, for example. In addition, not every artform will resemble the financial model of theatres. The Independent Cinema Office estimates that 'average rates for a 2, 3 or more screen venue are more likely to be in the 15–20% range' (Independent Cinema Office, n.d.). Furthermore, it is likely that smaller venues with very small operating margins will need to sell a much higher percentage of tickets to cover costs, even as much as 100%.

We also acknowledge that the capacity factor will likely be affected by many other circumstances, and a more nuanced estimation may be achieved by choosing different capacity factors for, say, a football cup final and a smaller local event, or distinguishing between a nationally recognised performer at the Edinburgh Festival Fringe and a new comedian in their debut year, or between a film performance on a Tuesday morning on a hot summer's day and a film performance of the newest film in the James Bond franchise on opening weekend. In all of these instances, combining the Data Thistle Dataset with further data would enable more accurate revenue modelling. For example, the event categories could be used to infer the distribution of child, family, senior concession or other types of tickets likely to be sold, or data about the types of events that often sell out could be used to nuance the expected percentage of total tickets sold.

4.2.2. Limitations

4.2.2.1. Capacity values. At present, the DTD only has capacity values for 120 performance places (out of 2474). Of the 1346 places in Edinburgh, only 82 have capacity values and therefore only a small subset of events at venues in the DTD appear in the above revenue calculation. This omission is made clear in 'Film' schedules, which we would expect to contribute a significant portion of the total revenue. However, the dataset does not include capacity values for the large cinema chains, either for the entire venue or for each individual screen, and therefore most of the revenue generated by 'Film' schedules does not appear in these figures. Obtaining the venue capacities for more venues would increase the relevant data in the DTD.

4.2.2.2. Capacity values and performance spaces. The capacity values present in the DTD relate to the capacity of the whole venue, not to each performance space, of which there may be several. The DTD does contain some data about which performance space a performance is located in. For example, the performance place Summerhall has a capacity value of 500 in the DTD but the venue is comprised of multiple smaller spaces, such as 'The Dissection Room', 'Red Lecture Theatre' and 'Basement', some of which are temporary structures erected during August for the Edinburgh festivals. Of the 38,704 events in the DTD, 923 take place at Summerhall and 80 have an associated performance space. Moreover, the maximum capacities of venues changed along with changing Covid-19 pandemic guidelines and this is not accounted for in the DTD. Obtaining the venue capacities for each performance space at each of the venues in the dataset – as well as tracking through time the changes due to Covid-19 distancing restrictions – would enable us to make a more accurate calculation.

4.2.2.3. Online performances. There are also online events, which are not represented in the DTD as they have no physical location associated with them and this dataset was generated by selecting events in a specific geographical area. The Edinburgh Festival Fringe reports that 300 online events took place in 2020 accruing £250,000 in ticket sales (Edinburgh Festival Fringe 2020), which are not represented in our calculations. A more accurate estimation of revenue would include online events, which raises the question of how to associate online events with a geographical location.

4.2.2.4. 'Free' events. There are a significant number of 'Free' or 'Pay What You Want' events in the dataset and, as these do not have fixed ticket prices, are not included in the revenue calculation. The following histogram (Figure 3) shows the number of free tickets available for performances in Edinburgh across 2018–2021, and Table 2 gives the number of free tickets in August 2018, 2019, 2020 and 2021.

As some performances have multiple ticket types (for example adult = £0, child = £0, family = £0 and student = £0), which are all represented in these figures, the number of free ticket types is larger than number of performances with free tickets. Free tickets have to be taken into consideration when estimating event revenue, especially during August.

4.2.3. Financial effect of pandemic on Edinburgh festivals

Again, it must be stated that these figures should be treated with an abundance of caution as our calculation rests on multiple assumptions and we are aware that we do not have all of the requisite venue capacities. Our estimation shows a loss of over £17 million for the sector in Edinburgh from 2019 to 2020, a figure that is not too far away from the revenue gap of £21 million reported by the Edinburgh Festival Fringe Society (Edinburgh Festival Fringe Society 2020). We should also stress that many of the venues and employees would have been supported by UK Government Covid-19 support grants and furlough schemes (Waitzman 2021), but that there were also many individuals, small traders, and freelancers that would not have been eligible

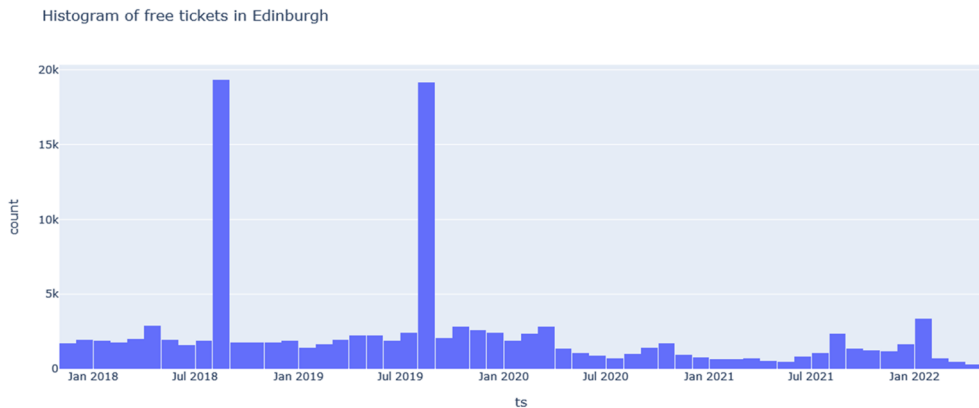


Figure 3. Histogram of free tickets frequency per category in Edinburgh in August 2018–2022.

Table 2. Number of free tickets in Edinburgh in August 2018–2021.

Month	Total number of free tickets
August 2018	19,317
August 2019	19,179
August 2020	1025
August 2021	2382

for these schemes. The cultural sector across the UK did receive financial support in the form of a £157 billion recovery fund (Department for Digital, Culture, Media & Sport, & HM Treasury, 2020), which included the Coronavirus Job Retention Scheme and the Self-Employment Income Support Scheme (Waitzman 2021). The furlough scheme in particular has been interrogated elsewhere regarding its success and failures (Brook, O'Brien, and Taylor 2022).

The pandemic has altered how many industries expect to operate, making planning for events and audience attendance more difficult. Will Page, previously the Chief Economist of Spotify and PRS for Music, compiled statistics from PRS for Music, Office of National Statistics, HSBC Economics, Entertainment Retailers Association and the British Phonographic Industry to chart live and recorded music sales between 2019 and 2021. Page's overview of the music industry shows a drastically changed – and evolving – picture. He emphasises the importance of robust data analysis for 'evidence-based policy making, not policy-based evidence making' as the basis for 'policymakers and industry... to figure out what assistance and actions are required to get us back to where we once belonged' (Page 2022). Data work like Page's and like that made possible by Data Thistle's data have far-reaching possibilities for industry and beyond.

The pandemic did not just affect ticket sales:

Cultural and creative sectors are important in their own right in terms of their economic footprint and employment. They also spur innovation across the economy, as well as contribute to numerous other channels for positive social impact (well-being and health, education, inclusion, urban regeneration, etc.). (Organisation for Economic Co-operation and Development 2020)

The economic impact of cultural events is not limited to ticket buying: it extends to transport, food and drink, accommodation and spending on other local amenities. Future work could see events data being used as the source for economists to calculate the wider financial benefits that events have to a city such as Edinburgh.

5. Changes in event provision due to covid-19 pandemic: findings and discussion

5.1. Provision findings

We know that the Covid-19 pandemic drastically affected the landscape of cultural events. After lockdowns were announced on 24 March 2020 (BBC News 2020a), events were cancelled or postponed in their thousands. After the easing of restrictions, we would expect that some industries would be better placed to return to existing levels of event provision more easily, such as outdoor activities. Using the Data Thistle Dataset, we investigated which types of events recovered more quickly and, from this, inferred which audiences have been impacted to a greater or lesser extent by the change in event provision.

This graph (Figure 4) shows the total number of performances (rather than events or schedules) in Edinburgh in August for the years 2018, 2019, 2020 and 2021. August was chosen as it is the city's busiest month for events. The bars are coloured according to category. For example, in August 2018 and 2019, before the Covid-19 pandemic, the dominant categories are 'Comedy' and 'Theatre'. In 2020 and 2021 both the number and categories of performances that took place changes drastically. As expected, the total number of in-person performances is vastly reduced from approx. 70,000 in 2019 to approx. 1500 performances in 2020 and increases again to approx. 10,000 performances in 2021, which is far below the pre-pandemic level of event provision (see Table 3 for exact figures).

Looking at the change in performance provision in Edinburgh from pre-pandemic (August 2019) to the easing of restrictions (August 2021) for the different categories, not all categories recovered to the same extent (see Table 4).

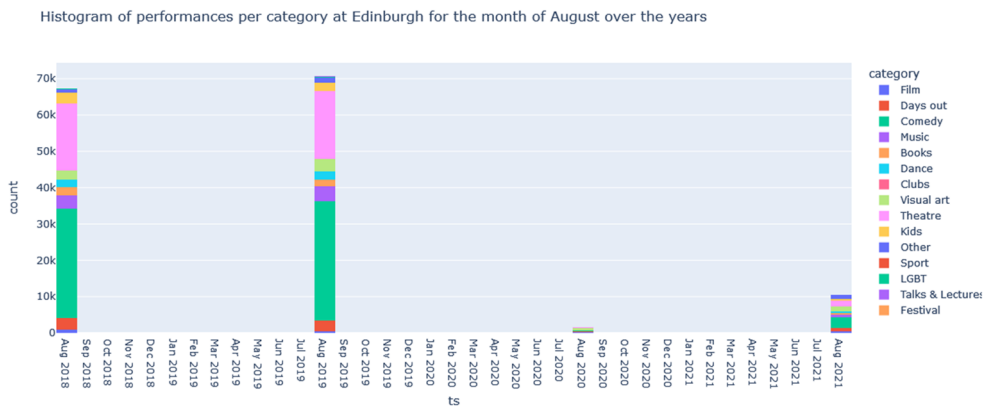


Figure 4. Performances in Edinburgh in August 2018, 2019, 2020 and 2021 per category.

Table 3. Total number of performances for all categories in Edinburgh for the years 2018, 2019, 2020 and 2021.

Year	Total number of performances
2018	67,384
2019	70,660
2020	1503
2021	10,405

Table 4. Number of performances in Edinburgh in 2019 and 2021, and the change between 2019 and 2021, by category.

Event category	No. of 2019 events	No. of 2021 events	% of events retained 2019–2021	% change 2019–2021
Visual Art	3427	1362	39.74	-60.26
Days Out	2954	875	29.62	-70.38
Books	1873	382	20.40	-79.60
Music	4082	783	19.18	-80.82
All Categories	70,660	10,405	14.73	-85.27
Kids	2416	323	13.37	-86.63
Theatre	18,605	1759	9.45	-90.55
Comedy	32,865	3006	9.15	-90.85
LGBT	261	2	0.77	-99.23

This table shows the number of performances in each of eight selected categories as well as a total for all categories. ‘Visual Art’ was the most resilient category of event, retaining 39.74% of pre-pandemic performances in 2021. ‘Days Out’ also managed to retain 29.62% of its pre-pandemic provision. Two categories that can be directly linked to specific demographics, and can therefore be used to intimate their relative impact on those demographics, are ‘Kids’ and ‘LGBT’. In 2021, event provision for children was only 13.37% of that in 2019. But the category with the biggest difference was ‘LGBT’, with 2021 event provision for LGBTQIA+ communities at only 0.77% of that in 2019.

5.2. Provision discussion

For industry, the wider implications of the disparity in event provision during (and potentially after) the pandemic are an understanding of which types of events are more robustly able to weather public health crises. There are also sociocultural implications about how sectors of society, such as LGBTQIA+ communities and children, have been affected disparately.

5.2.1. Effect of Covid-19 on LGBTQIA+ communities

Research on the impact of the pandemic on LGBTQIA+ communities indicates that they have been negatively affected by the pandemic in specific ways (Haworth 2021), especially when taking into account intersections with other minoritised identities (LGBT Foundation 2020). This disparity has been directly linked to the inability of LGBTQIA+ individuals to access safe spaces and events, as described in the NatCen report ‘The experiences of UK LGBT+ communities during the COVID-19 pandemic’:

The evidence identified by this review suggests that the COVID-19 pandemic has had a negative impact on the mental health of LGBT+ people living in the UK. This includes evidence of increased anxiety and depression, attributed to feelings of isolation and loneliness through the loss of safe, supportive, and identity-affirming peer-groups, communities and spaces. (Hudson et al. 2021, 2)

Billy Haworth of the Humanitarian and Conflict Response Institute notes in particular the importance of cultural events to mental health and wellbeing where the inaccessibility of cultural events and spaces during the pandemic has led to a situation where ‘disruptions to LGBTIQ+ spaces have been acutely felt, including nightlife, but also community spaces, support groups and activities like Pride festivals. These spaces usually provide many LGBTIQ+ people with essential opportunities to freely and safely express themselves and their identities’ (2021).

It should be noted that worries over declining numbers of LGBTQIA+ event and community spaces, and a commensurate effect on the mental health of LGBTQIA+ youths, especially, have been prevalent since before the Covid-19 pandemic (Marshall 2021; Walters 2015). The literature on this subject cites a lack of robust empirical evidence (Hudson et al. 2021, 2), such as McGowan et al.’s overview of research on the impact of the pandemic on the health and wellbeing of LGBTQIA+ communities, which finds a lack of high-quality research that attends to social and structural factors as well as a lack of sexual orientation and gender identity data (McGowan et al. 2021, 1). Given the importance of cultural events and spaces to the health and wellbeing of LGBTQIA+ communities, and the unique circumstances of the pandemic making those events and spaces inaccessible, the use of cultural events data, like those available through Data Thistle, provides an additional angle of insight into the effects of the pandemic. Cultural events metadata provides one more facet of evidence to combine with existing sociological methods of evidence gathering.

5.2.2. Limitations

5.2.2.1. Online events. However, there are limitations to what the DTD can tell us. The DTD does not contain data about online events, which have in some ways ameliorated the lack of access to in-person events when ‘moving activities online presented opportunities to connect with new and diverse audiences, in some ways improving accessibility’ (Haworth 2021). Lo Marshall describes how ‘[b]y removing geographical barriers...online platforms have shown how LGBTQ+ spaces can reach new audiences and connect with communities in difficult circumstances’ (2021). It should be noted that data about online events are available – Data Thistle separately collected a comprehensive set of online events during the Covid-19 pandemic – and future work focused on quantifying event provision for LGBTQIA+ or other specific communities would benefit from including them.

5.2.2.2. Intersectionality. Marshall goes on to highlight the ways in which online spaces can in some cases better cater to intersectional identities with ‘intimate digital gatherings for people who live under multiple identities’ (2021). Attending to intersections of identity is important in all social research, and particular care must be taken not to treat LGBTQIA+ individuals as a homogenous community, as ‘there

is no single LGBTIQ+ experience' but '[d]iversity of COVID-19 experiences reflects diversity within LGBTIQ+ populations, including diversity between different subcategories, but also intersections with other factors, like age, class, disability or race' (Haworth 2021) with transgender people facing disproportionate challenges (Haworth 2021; LGBT Foundation 2020).

5.2.2.3. Data taxonomies. A limitation arising directly from the DTD is the way in which events are categorised and tagged. For example, at the event level, the DTD contains 75 events with category 'LGBT' and an additional 267 events tagged as 'LGBT' and assigned to other categories. This means that a search for events categorised as 'LGBT' will only return 21.93% (75 out of 342) of potentially relevant events. It should be noted that the 'LGBT' category is particularly susceptible to inconsistent use as event creators may opt to categorise their events as 'LGBT' and use the tags to give [supplementary information](#) about the event category or to categorise their events using one of the other categories and use the tags to indicate that the event may be relevant to audiences, or include work by people, from LGBTQIA+ communities. In addition to the differences between how tags and categories are attached to events, the textual description of each event may also contain terms indicating a specific audience, such as 'family-friendly' or 'queer'. Any future work using the DTD will require an awareness of these taxonomic discrepancies and how they factor into formulating search queries.

Future work on event provision for LGBTQIA+ communities could profitably examine the changing landscape of event provision, the link between access to arts and culture and mental health and wellbeing, and language use around the ways in which events are targeted towards specific demographics. This work could also be extended to focus on events targeted at children.

6. Conclusions

The question of how to determine cultural value – financially or otherwise – is a difficult one. There exist multiple approaches drawn from multiple disciplines and using heterogeneous data sources. In this article we aimed to evaluate large-scale metadata about cultural events as an additional data source. To this end, we sourced data from Data Thistle about events in Edinburgh from 2018 to 2022, evaluated the data's potential for analysis, and performed some exploratory analyses regarding changes in cultural revenue and cultural provision due to the Covid-19 pandemic.

We introduce a 153MB dataset comprised of events that took place in the Edinburgh and South-East Scotland City Region between 1 November 2017 and 1 May 2022 and describe one method of exploring and working with those data, using pandas DataFrames and Python. The dataset is rich and detailed, and so there is the potential for using it to ask nuanced and valuable questions about event provision and more in the UK. However, this richness comes with complexities. We highlighted some of these – the difference between events, schedules and performances, and between categories and tags – and discussed how they will impact findings.

After working through the characteristics of the data, we conducted two sets of analyses. The first investigated the possibility of estimating the difference in revenue generated during August in Edinburgh – the month when the largest arts festival in

the world is held – before and during the Covid-19 pandemic. We propose a method for estimating the revenue from cultural events by multiplying, for each event, the number of performances by the average ticket price, the capacity of the venue and a factor to reflect the average number of tickets sold. For the subset of venues with the relevant data in the DTD, we estimate that there was a 97.3% fall in ticketing revenue between 2019 and 2020. This work also surfaced issues around incomplete data regarding venue capacities and determining the most accurate capacity factor.

In an exploration of how event provision is affected by the Covid-19 pandemic, we looked at whether each category of event was affected equally. From the performances in each category in Edinburgh in August 2018, 2019, 2020 and 2021 we determined that ‘Visual Art’ and ‘Days Out’ were the most resilient categories of event retaining the most performances while the ‘Theatre’, ‘Comedy’ and ‘LGBT’ categories were the most reduced. We connected disproportionate event provision to inequalities in how sectors of society are affected by the pandemic in different ways.

Our key findings from these analyses are 1) large-scale cultural events data add an additional facet to existing approaches for assessing cultural value, both financially and socially, 2) that while the Data Thistle Dataset is rich enough to answer interesting questions it requires a substantial undertaking to understand and clean the data, and the research potential is vastly expanded if the dataset is enhanced or used alongside other datasets. A report on behalf of the Arts and Humanities Research Council in 2016 recommended the establishment of an Observatory for Cultural Value (Crossick and Kaszynska 2016, 10). We also suggest that infrastructural support for accessing and working with large-scale cultural data would allow novel research that can expand a pressing area of enquiry across multiple disciplines, encouraging research that intertwines with industry.

Notes

1. <https://esescityregiondeal.org.uk/>.
2. This category label has evolved over the years with Data Thistle using LGBT consistently in the dataset they provided to us. Throughout this article we use the label LGBTQIA+ as a non-exhaustive list indicating the many identities under this umbrella when not specifically referencing Data Thistle’s category label.
3. <https://api.datathistle.com/>.
4. The code for this notebook can be found here: https://github.com/rosafilgueira/Exploring_TheList_Data/blob/main/mining_LIST_2018_2021_data/Generate_List_Dataframes.ipynb.
5. The code for this notebook can be found here: https://github.com/rosafilgueira/Exploring_TheList_Data/blob/main/mining_LIST_2018_2021_data/Case_Study_Maker.ipynb.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Arts and Humanities Research Council under Grant AH/W007533/1.

Notes on contributors

Suzanne R Black is a postdoctoral researcher at the University of Edinburgh. With a background in English Literature, she combines humanities approaches with digital methods, and has worked across a range of projects involving data and the creative industries. Her research interests lie in digital literary culture and the development of data-led approaches to contemporary fiction and fanfiction.

Rosa Filgueira is a Lecturer at the School of Computer Science at the University of St Andrews. Her research is concerned with three closely-related topics. The first is to develop adaptive communication techniques which optimise the data movement for data-intensive applications at different HPC levels. The second is to facilitate the development of scientific workflows/application that can be run in many HPC environments while hiding the complexity to the users. The last is to develop intelligent frameworks that automatically understand domain-experts software and provide automated and tailored advice and suggestions.

Lesley McAra is Professor of Penology in the Law School and Director of the Institute for Advanced Studies in the Humanities, at Edinburgh University. For the past 24 years she has been Co-Director of the Edinburgh Study of Youth Transitions and Crime, a longitudinal research programme following a cohort of 4,300 young people. In 2019, she was joint winner of the ESRC prize for outstanding public policy impact. A past President of the European Society of Criminology, Lesley was awarded a CBE for services to Criminology in 2018, and elected as Fellow of the Royal Society of Edinburgh in 2021.

Brendan Miles is the Director of Data and Content for Data Thistle, the UK's leading live events data business. He has previously worked in arts, music, advertising and publishing and led commercial operations at Scotsman.com. Brendan is a specialist and occasional speaker on content monetisation, digital strategy and digital business models. At Data Thistle he leads their growing information services business working with the likes of Reach, trivago, LNER, Visit Scotland and the Department of Culture, Media and Sport.

Mark Parsons is the Associate Dean for e-Research within the College of Science & Engineering at the University of Edinburgh. In this role he provides advice and guidance to the College and University on all aspects of modern research infrastructures with a particular focus on high performance computing and data analytics.

Melissa Terras is the Professor of Digital Cultural Heritage at Edinburgh College of Art, which she joined in October 2017, previously directing UCL Centre for Digital Humanities. Her research focuses on the digitisation of cultural heritage, including its technologies, procedures, and impact, and how this intersects with internet technologies. She is Director of Creative Informatics, the AHRC funded Creative Cluster, and Turing Institute Fellow 2018–2022.

ORCID

Suzanne R. Black  <http://orcid.org/0000-0001-5479-0528>

Rosa Filgueira  <http://orcid.org/0000-0002-5715-3046>

Mark Parsons  <http://orcid.org/0000-0003-4097-7468>

Melissa Terras  <http://orcid.org/0000-0001-6496-3197>

Data availability statement

The data that support the findings of this study are available from Data Thistle. Restrictions apply to the availability of these data, which were used under license for this study.

The code that supports the findings of this study is openly available at https://github.com/rosafilgueira/Exploring_TheList_Data/blob/main/mining_LIST_2018_2021_data/Generate_List_Dataframes.ipynb.

https://github.com/rosafilgueira/Exploring_TheList_Data/blob/main/mining_LIST_2018_2021_data/Case_Study_Maker.ipynb.

References

- Alexander, J. C., Jacobs, R. N., & Smith, P. (Eds.). 2017. *The Oxford Handbook of Cultural Sociology*. Oxford: Oxford University Press.
- Ali-Knight, J. 2021. *Modelling Mobility and Infection: The Reopening of Edinburgh's Festivals*. University of Edinburgh. <https://publichealthscotland.scot/repository/modelling-mobility-and-infection-the-reopening-of-edinburghs-festivals/>
- Anayi, L., N. Bloom, P. Bunn, S. Chen, P. Mizen, M. Oikonomou, G. Thwaites, and P. Smietanka. 2021. "Update: Which Firms and Industries Have Been Most Affected by Covid-19?" Economics Observatory, May 5. <https://www.economicsobservatory.com/update-which-firms-and-industries-have-been-most-affected-by-covid-19>
- Arts and Humanities Research Council. 2021. "Deafblind Community Affected by Lack of Touch." *UKRI*, March 17. <https://www.ukri.org/news-and-events/tackling-the-impact-of-covid-19/researching-the-impact-of-coronavirus/deafblind-community-affected-by-lack-of-touch/>
- Arts Council England. 2014. The value of arts and culture to people and society: An evidence review. https://www.artscouncil.org.uk/sites/default/files/download-file/The_value_of_arts_and_culture_to_people_and_society_an_evidence_review.pdf
- Bateman, K. 2022. "COVID-19 Hit the Creative Industries Particularly Hard. How Can They Be Supported in Future?" *World Economic Forum*, February 22. <https://www.weforum.org/agenda/2022/02/creatives-job-losses-covid-employment/>
- BBC News. 2020a. "Coronavirus: UK brings in strict curbs on life to fight virus." *BBC News*, March 24. <https://www.bbc.com/news/uk-52014472>
- BBC News. 2020b. "Coronavirus: Emergency money for culture 'won't save every job'" *BBC News*, July 7. <https://www.bbc.com/news/entertainment-arts-53302415>
- BBC News. 2022. "Live music revenue fell again in 2021, despite gigs returning." *BBC News*, April 25. <https://www.bbc.com/news/entertainment-arts-61187598>
- Betzler, D., E. Loots, M. Prokůpek, L. Marques, and P. Grafenauer. 2021. "COVID-19 and the Arts and Cultural Sectors: Investigating Countries' Contextual Factors and Early Policy Measures." *International Journal of Cultural Policy* 27 (6): 796–814. <https://doi.org/10.1080/10286632.2020.1842383>
- Bradbury, A., K. Warran, H. W. Mak, and D. Fancourt. 2021. The Role of the Arts during the COVID-19 Pandemic (p. 44). University College London. https://www.artscouncil.org.uk/sites/default/files/download-file/UCL_Role_of_the_Arts_during_COVID_13012022_0.pdf
- Brook, O., D. O'Brien, and M. Taylor. 2020. *Culture is Bad for You: Inequality in the Cultural and Creative Industries*. Manchester: Manchester University Press.
- Brook, O., D. O'Brien, and M. Taylor. 2022. "Covid, furlough and creative businesses." *CultureHive*. <https://www.culturehive.co.uk/CV/resources/covid-furlough-and-creative-businesses/>
- Chamberlain, P., and D. Morris. 2021. "The Economic Impact of Covid-19 on the Culture, Arts and Heritage (CAH) Sector in South Yorkshire and Comparator Regions." University of Sheffield. <https://www.sheffield.ac.uk/media/28397/download?attachment>
- Clark, H. 2022. "How Has the Pandemic Affected Industries and Labour in the UK? House of Commons Library. <https://commonslibrary.parliament.uk/how-has-the-pandemic-affected-industries-and-labour-in-the-uk/>
- Comunian, R., and L. England. 2020. "Creative and Cultural Work without Filters: Covid-19 and Exposed Precarity in the Creative Economy." *Cultural Trends* 29 (2): 112–128. <https://doi.org/10.1080/09548963.2020.1770577>
- Crossick, G., and P. Kaszynska. 2016. "Understanding the Value of Arts & Culture: The AHRC Cultural Value Project." Arts and Humanities Research Council. <https://www.ukri.org/publications/ahrc-cultural-value-project-report/>

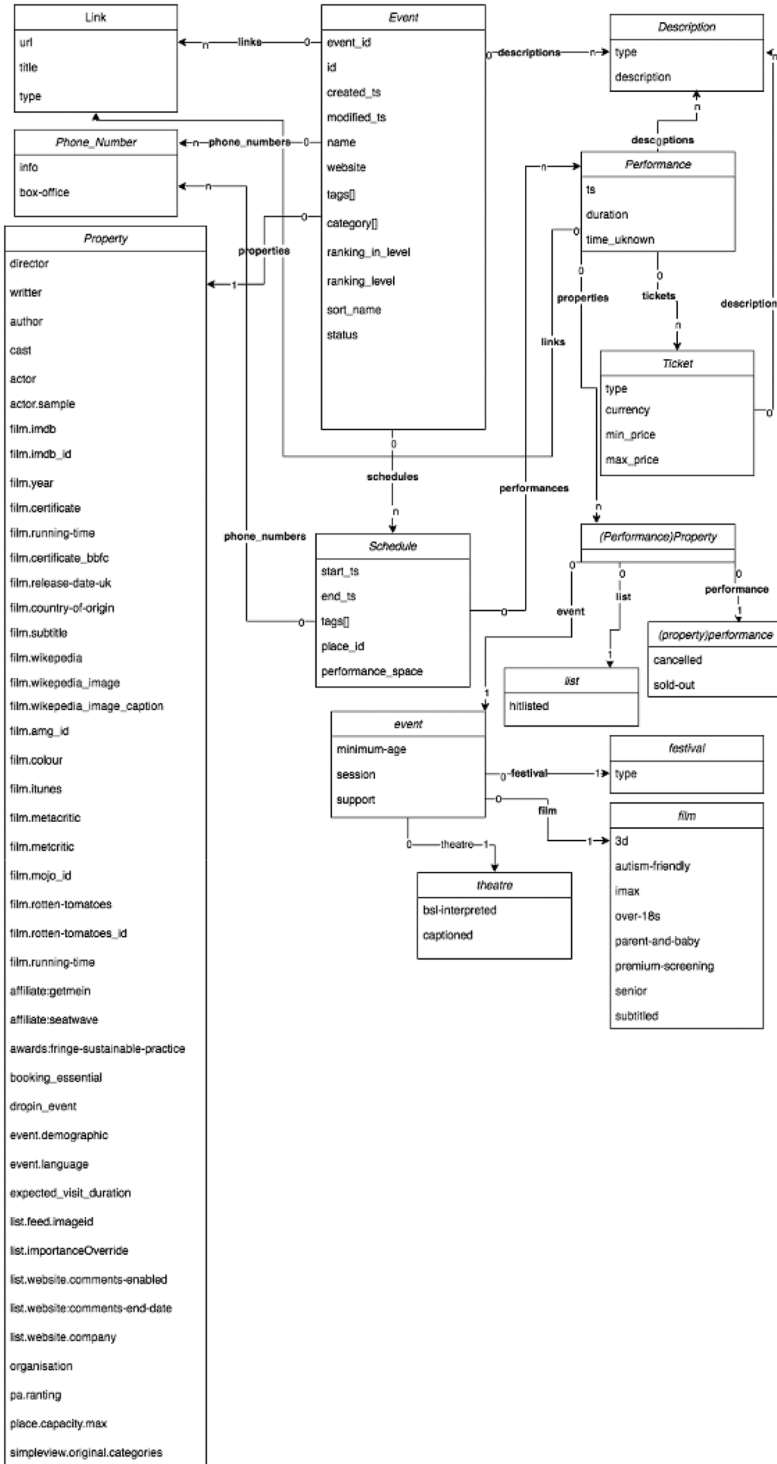
- Currie, M., and M. M. Correa. 2021. *The Culture and Communities Mapping Project*. London: Palgrave Macmillan. <https://link.springer.com/book/10.1007/978-3-030-88651-6>
- Currie, M., and M. M. Correa. 2022. "Tangibles, Intangibles and Other Tensions in the Culture and Communities Mapping Project." *Cultural Trends* 31 (1): 88–106. <https://doi.org/10.1080/09548963.2021.1910491>
- Data Thistle. n.d. Events Data. Retrieved 28 October 2022, from <https://www.datathistle.com/about/events-data/>
- Deeker, W. 2022. *The Covid Generation: The Effects of the Pandemic on Youth Mental Health*. Horizon: The EU Research & Innovation Magazine. <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/covid-generation-effects-pandemic-youth-mental-health>.
- Denk, J., A. Burmester, M. Kandziora, and M. Clement. 2022. "The Impact of COVID-19 on Music Consumption and Music Spending." *PLoS One* 17 (5): e0267640. <https://doi.org/10.1371/journal.pone.0267640>.
- Department for Digital, Culture, Media & Sport. 2020. "UK's Creative Industries Contributes Almost £13 Million to the UK Economy Every Hour." <https://www.gov.uk/government/news/uks-creative-industries-contributes-almost-13-million-to-the-uk-economy-every-hour>
- Department for Digital, Culture, Media & Sport. 2021. "DCMS Economic Estimates 2019 (Provisional): Gross Value Added." <https://www.gov.uk/government/statistics/dcms-economic-estimates-2019-gross-value-added/dcms-economic-estimates-2019-provisional-gross-value-added>
- Department for Digital, Culture, Media & Sport. 2022. "DCMS Sectors Economic Estimates 2020: GVA." <https://www.gov.uk/government/statistics/announcements/dcms-sectors-economic-estimates-2020-gva>
- Department for Digital, Culture, Media & Sport, & HM Treasury. 2020. "£1.57 Billion Investment to Protect Britain's World-Class Cultural, Arts and Heritage Institutions." <https://www.gov.uk/government/news/157-billion-investment-to-protect-britains-world-class-cultural-arts-and-heritage-institutions>
- Economic and Social Research Council. 2021a. *Older People's Lives and Coronavirus*. UKRI. <https://www.ukri.org/news-and-events/tackling-the-impact-of-covid-19/researching-the-impact-of-coronavirus/older-peoples-lives-and-coronavirus/>
- Economic and Social Research Council. 2021b. *The Pandemic's Impact on UK's LGBT+ Communities*. UKRI. <https://www.ukri.org/news-and-events/tackling-the-impact-of-covid-19/researching-the-impact-of-coronavirus/the-pandemics-impact-on-uks-lgbt-communities/>.
- Economic and Social Research Council & Innovate UK. 2021. *COVID-19 and Children*. UKRI. <https://www.ukri.org/news-and-events/tackling-the-impact-of-covid-19/researching-the-impact-of-coronavirus/covid-19-and-children/>.
- Edinburgh Festival Fringe. 2020. "Fringe 2020 draws to a close with almost 300 shows online and £250,000 raised for artists and venues." <https://www.edfringe.com/learn/news-and-events/fringe-2020-draws-to-a-close-with-almost-300-shows-online-and-ps250-000-raised-for-artists-and-venues>
- Edinburgh Festival Fringe Society. 2020. Impact of Covid-19 on DCMS sectors: Evidence to UK Parliament (Digital, Culture, Media and Sport Committee No. CVD0092). <https://committees.parliament.uk/writtenevidence/2942/pdf/>
- Government Equalities Office. 2018. "National LGBT Survey: Summary report." p. 28. <https://www.gov.uk/government/publications/national-lgbt-survey-summary-report>
- Gross, J., and S. Pitts. 2016. "Audiences for the Contemporary Arts: Exploring Varieties of Participation across Art Forms in Birmingham, UK." *Participations: Journal of Audience & Reception Studies* 13 (1): 1–23. <https://www.participations.org/13-01-02-gross.pdf>
- Hadley, S., K. Johanson, B. Walmsley, and A. Torreggiani. 2019. "Reflections on Audience Data and Research." *Cultural Trends* 28 (2-3): 81–85. <https://doi.org/10.1080/09548963.2019.1617486>
- Hanquinet, L., D. O'Brien, and M. Taylor. 2019. "The Coming Crisis of Cultural Engagement? Measurement, Methods, and the Nuances of Niche Activities." *Cultural Trends* 28 (2-3): 198–219. <https://doi.org/10.1080/09548963.2019.1617941>

- Haworth, B. 2021. "LGBTIQ+ experiences of COVID-19 in the UK and the need for more inclusive crisis policies." *Policy@Manchester Blogs*, March 10. <https://blog.policy.manchester.ac.uk/posts/2021/03/lgbtiq-experiences-of-covid-19-in-the-uk-and-the-need-for-more-inclusive-crisis-policies/>
- Hillis, S. D., H. J. T. Unwin, Y. Chen, L. Cluver, L. Sherr, P. S. Goldman, O. Ratmann, et al. 2021. "Global Minimum Estimates of Children Affected by COVID-19-Associated Orphanhood and Deaths of Caregivers: A Modelling Study." *Lancet (London, England)* 398 (10298): 391–402. [https://doi.org/10.1016/S0140-6736\(21\)01253-8](https://doi.org/10.1016/S0140-6736(21)01253-8).
- Hudson, N., F. Kersting, S. Lynch-Huggins, L. MacNaboe, and S. Sharrock. 2021. *The Experiences of UK LGBT+ Communities during the COVID-19 Pandemic: A Review of Evidence* (p. 71). NatCen. <https://www.natcen.ac.uk/media/2120188/The-experiences-of-UK-LGBTplus-communities-during-the-COVID-19-pandemic.pdf>
- Independent Cinema Office. n.d. How to Start a Cinema: 10 The economics of the operation. Retrieved 15 June 2022, from <https://www.independentcinemaoffice.org.uk/advice-support/how-to-start-a-cinema/the-economics-of-the-operation/>
- Leung, R., and J. Davies. 2021. *On-Demand Culture: How the Lockdown is Changing Games and Streaming Services*. PEC/nesta. <https://cdn2.assets-servd.host/creative-pec/production/assets/publications/On-Demand-Culture-How-the-lockdown-is-changing-games-and-streaming-services-Insights-for-Policy-Makers-Jan-2021-3.pdf>
- LGBT Foundation. 2020. *Hidden Figures: The Impact of the Covid-19 Pandemic on LGBT Communities*. <https://lgbt.foundation/coronavirus/hiddenfigures>
- Lhermitte, M., B. Perrin, and S. Blanc. 2015. *Cultural Times: The First Global Map of Cultural and Creative Industries*. EY; Confederation of Authors and Composers Societies. https://en.unesco.org/creativity/sites/creativity/files/cultural_times_the_first_global_map_of_cultural_and_creative_industries.pdf
- Manovich, L. 2020. *Cultural Analytics*. Cambridge, MA: MIT Press.
- Marshall, L. 2021. "LGBTQ+ spaces face a new threat from the pandemic – here's how they can adapt." *The Conversation*, February 24. <http://theconversation.com/lgbtq-spaces-face-a-new-threat-from-the-pandemic-heres-how-they-can-adapt-153808>
- McAndrew, S., D. O'Brien, and M. Taylor. 2020. "The Values of Culture? Social Closure in the Political Identities, Policy Preferences, and Social Attitudes of Cultural and Creative Workers." *The Sociological Review* 68 (1): 33–54. <https://doi.org/10.1177/0038026119871401>
- McGowan, V. J., H. J. Lowther, and C. Meads. 2021. "Life under COVID-19 for LGBT+ People in the UK: Systematic Review of UK Research on the Impact of COVID-19 on Sexual and Gender Minority Populations." *BMJ Open* 11 (7): e050092. <https://doi.org/10.1136/bmjopen-2021-050092>
- Munir, M. 2020. *How COVID-19 is Affecting LGBT Communities*. Stonewall. <https://www.stonewall.org.uk/about-us/news/how-covid-19-affecting-lgbt-communities>
- Naylor, R., J. Todd, M. Moretto, and R. Traverso. 2021. *Cultural and Creative Industries in the Face of COVID-19: An Economic Impact Outlook* (p. 62). UNESCO Digital Library. <https://unesdoc.unesco.org/ark:/48223/pf0000377863>
- News Editor. 2020. "Theatres and Venues Claim August 1 Reopening Plan 'Not Viable.'" *The Ticketing Business News*, July 20. <https://www.theticketingbusiness.com/2020/07/20/theatres-venues-claim-august-1-reopening-plan-not-viable/>
- Oman, S. 2021. *Understanding Well-Being Data: Improving Social and Cultural Policy, Practice and Research*. Springer Nature. <https://doi.org/10.1007/978-3-030-72937-0>
- Organisation for Economic Co-operation and Development. 2020. Culture shock: COVID-19 and the cultural and creative sectors. <https://www.oecd.org/coronavirus/policy-responses/culture-shock-covid-19-and-the-cultural-and-creative-sectors-08da9e0e/>
- Page, W. 2022. "From Suffering to Recovering: The Stats That Explain Why the UK Music Industry is Set for a Dramatic Post-Lockdown Comeback." *Music Business Worldwide*, August 2. <https://www.musicbusinessworldwide.com/from-suffering-to-recovering-the-stats-that-explain-why-the-uk-music-industry-is-set-for-a-dramatic-post-lockdown/>
- Pickford, J. 2020. "Lights up again for England's struggling theatres." *Financial Times*, August 14. <https://www.ft.com/content/e7450de9-6e4f-47de-8a0d-8d163789d2ec>

- Ryan-Saha, J. 2020a. "City Mobility—Decisions based on events data." *Data Thistle*, February 17. <https://datathistle.medium.com/city-mobility-decisions-based-on-events-data-1a8d5fe51e09>
- Ryan-Saha, J. 2020b. "Festival Mobility: Decisions with Data." *The University of Edinburgh*, April 20. <https://www.ed.ac.uk/edinburgh-international-data-facility/updates-events/festival-mobility-decisions-with-data>
- Sagger, H., J. Philips, and M. Haque. 2021. "Valuing Culture and Heritage Capital: A Framework towards Informing Decision Making," p. 26. Department for Digital, Culture, Media & Sport. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/955203/GOV.UK_-_Framework_Accessible_v2.pdf
- Telegraph Reporters. 2021. "What the lockdown roadmap delay means for theatres, galleries and cinemas." *The Telegraph*, June 14. <https://www.telegraph.co.uk/theatre/what-to-see/theatres-cinemas-museums-lockdown-delay-june-21-rules-covid/>
- UK Theatre & Society of London Theatre. 2020. "Impact of Covid-19 on DCMS sectors (No. CVD0259)." Digital, Culture, Media and Sport Committee. <https://committees.parliament.uk/writtenevidence/6019/pdf/>
- Waitzman, E. 2021. Impact of government policy on the creative sector. <https://lordslibrary.parliament.uk/impact-of-government-policy-on-the-creative-sector/>
- Walmsley, B., A. Gilmore, D. O'Brien, and A. Torreggiani. 2022. Culture in Crisis: Impacts of Covid-19 on the UK cultural sector and where we go from here. Centre for Cultural Value. https://www.culturehive.co.uk/wp-content/uploads/2022/01/Exec_Summary_Culture_in_Crisis.pdf
- Walters, B. 2015. "Closing Time for Gay Pubs – a New Victim of London's Soaring Property Prices." *The Guardian*, February 4. <https://www.theguardian.com/society/2015/feb/04/closing-time-gay-pubs-lgbt-venues-property-prices>
- Wellcome. 2020. "More than a health crisis: Covid-19 impacts are far-reaching and long-term." <https://wellcome.org/news/more-health-crisis-covid-19-impacts-are-far-reaching-and-long-term>
- Wenham, C., J. Smith, S. E. Davies, H. Feng, K. A. Grépin, S. Harman, A. Herten-Crabb, and R. Morgan. 2020. "Women Are Most Affected by Pandemics—Lessons from past Outbreaks." *Nature* 583 (7815): 194–198. <https://doi.org/10.1038/d41586-020-02006-z>.

Appendix A

Events entity-relationship diagram.



Appendix B

Places entity-relationship diagram.

