information

MDPI

Article

A Data-Driven Approach to Estimating the Impact of Audiovisual Art Events Through Web Presence

Andreas Giannakoulopoulos *, Minas Pergantis *, Aristeidis Lamprogeorgos and Stella Lampoura

Department of Audio and Visual Arts, Ionian University, 49100 Kerkira, Greece; a18labr@ionio.gr (A.L.); stella@ionio.gr (S.L.)

* Correspondence: agiannak@ionio.gr (A.G.); mperg@ionio.gr (M.P.)

Abstract: In the modern era, a major avenue of dissemination for cultural and artistic events is through the World-Wide Web, where every such event has a multifaceted distinct digital footprint. This digital footprint is an indicator of how strong the influence of each event is in the public's perception and to what extent it becomes part of the audiovisual art landscape. This study aims to present how the impact of an audiovisual event may be estimated using quantitative data collected through its online presence. This data-driven approach is made possible through web data extraction techniques and the use of generative artificial intelligence, which allows for structured information extraction from an endless variety of websites. Based on an event's innate characteristics, web outreach, estimated scope, and thematical popularity, an encompassing impact factor is calculated, which may be used to rank events on the basis of perceived influence. For the purposes of this study, a dataset consisting of thousands of events in Greece was collected over an extended period. These data were used for a computational statistical analysis. Through this process of data collection, impact calculation, and analysis, data-driven insights were derived concerning the landscape of audiovisual art events.

Keywords: audiovisual art; cultural impact; data analysis; world-wide web; generative artificial intelligence; web scraping

Academic Editor: María N. Moreno García

Received: 7 December 2024
Revised: 13 January 2025
Accepted: 20 January 2025
Published: 24 January 2025

Citation: Giannakoulopoulos, A.; Pergantis, M.; Lamprogeorgos, A.; Lampoura, S. A Data-Driven Approach to Estimating the Impact of Audiovisual Art Events Through Web Presence. *Information* **2025**, *16*, 88. https://doi.org/10.3390/info16020088

1. Introduction

Audiovisual art events are an integral part of modern human society and play an important role in human development. The contribution of art to social well-being is well documented in the international literature. There are studies that relate attendance at cultural events or participation in art projects with improvements in life quality and wellbeing [1–4]. There are also mentions about the important role of art in engaging people and helping them to develop social skills, improving their quality of life [5,6]. In addition to that, the impact of audio and visual art combined with digital technology is, in particular, an emerging research area, showing positive results both in individuals and communities [7] in the fields of social and physical health, creativity, knowledge, identity, and the economy.

Modern forms of expression through digital technology are the new avant-garde in the field of art, which seemingly stagnated at the end of the 20th century. These new forms must be part of a dialectic of modern digital society and man's innate need for expression and creativity [8]. At the same time, cutting-edge digital mediums involving extended realities can allow wider and more diverse audiences to engage with culture [9]. Audiovisual arts can help engage citizens and give meaning to a common future in the digital era. Furthermore, art in its modern version plays an important role and is a way of expressing human social intelligence, which receives new challenges from its coexistence with artificial intelligence [8].

The role of media coverage of art and art events has received a lot of research attention in the past. The work of Janssen [10] demonstrates that during the period of 1965 to 1990, important deviations appeared in the manner in which newspapers covered art-related matters. More specifically, the study showed that art, like pop music, literature, and cinema, increased the proportion of space they received and concluded that "These changes appear to be closely related to changes in the audience for the arts, developments in the arts supply, and pressures from advertisers and competitors" [11].

Media coverage provides the information products offered to society, using a standardized manner of formulating the media reality (wording, statements, images, expressions of a presenter, etc.). Especially in the case of an audience with low levels of media literacy, the media reality becomes social reality. We can observe the same tactics in digital media such as online media outlets and journalistic websites, which have evolved to imitate traditional media while appearing as refreshing ways of informing and commercializing [11].

In recent years, researchers have become increasingly interested in the field of big data and computational engineering. The term big data refers to a huge number of datasets that cannot be analyzed with conventional data processing systems but rather require innovative automated technologies to capture, store, distribute, manage, and analyze [12]. An example of such datasets is news items available online. Such diverse and large-scale information is impossible to process manually; this is achieved by breaking the data into sentences or words in order to analyze and extract the useful parts [13]. Furthermore, considerable research attention has been devoted to big data utilization, presenting the importance of the method as a key indicator of a firm's efficiency and productivity [14].

Big data research analysis in the past has been used through computational engineering for the study of cultural phenomena on a large scale. For example, Eurostat, the European Union's statistical office, has been using big data sources since 2013. In particular The European Statistical System (ESS) adopted the Scheveningen Memorandum in order to explore the potential of big data for producing official statistics [15] in 2013 and created the Big Data Action Plan and Roadmap [16] in 2014.

Consequently, distant reading and cultural analytics are two approaches that were generated by researchers at the beginning of the 21st century, i.e., by Moretti [17] and Manovich [18]. These approaches link social studies with big data and data mining, and both types of research focus on investigating the patterns and trends through wider perspectives [19]. This added distance from the subject matter allows for a focus on more macro concepts such as themes, systems, and tropes [17]. More specifically, cultural analytics is used to study cultural information on a massive scale through computational techniques [18]. Going one step further, Manovich [18] suggests a new field of research, called "computational media studies", which emphasizes the computational analysis of content derived from online activity.

Data-driven research is a scientific method of storing, analyzing, correlating, and querying data. Its main purpose is to gain insights through a large data volume by analyzing patterns and classification schemes that emerge in the dataset [20]. Ample evidence exists to suggest the importance of data-driven methods, especially in the field of industry, where a statistical positive correlation between the data-driven decisions and the productivity of a firm has been established [21]. This data-oriented approach is now being adopted by the creative industries in order to enable innovation and to drive creative decisions [22,23].

The impact of an event of cultural significance has been mostly studied in terms of economic results [24–26]. Quantitative metrics such as the number of visitors and locals that attended, as well as spending and local business income, provide useful subjective data [24]. This data collection process may be assisted by digital technologies such as action-tracking systems [26], but collecting this information is an intricate process, which makes it hard to gain insight from a large number and variety of events. Going beyond economic impact, methodologies have been proposed to assess social and cultural impact, as well as to analyze the cultural effects of various events [27,28]. These methodologies focus on defining the theoretical framework and proposing analysis tools [27] and also devising metrics and applying them on specific events [28].

This study presents a data-driven approach that implements cultural analytics and data-driven research methods in order to investigate the relationship between audiovisual art events and their potential impact, which is inferred through their perceived outreach. In this context, this study describes in detail its data collection methodology, which combines web data scraping with generative artificial intelligence (GenAI) and human oversight to produce structured information from an endless variety of websites. This collected information is used as the basis for a quantitative approach, which establishes an audiovisual event impact factor as a numerical representation of each event's footprint in the realm of contemporary events. Using descriptive and computational statistics, the impact factor alongside other quantitative data is used as a basis to answer the following research questions:

RQ1: Is it feasible to quantify the impact of an event of audiovisual art through its online presence using a data-driven approach?

RQ2: What insights can be derived regarding the Greek audiovisual event landscape through implementing this methodology?

Through answering these questions, the ultimate goal of this study is two-fold: firstly, it aims to present a valid methodology for the collection of objective data, the multi-layered classification of events, and the quantitative calculation of their impact, and secondly, it aims to investigate the contemporary landscape of audiovisual events in Greece and provide insights regarding the role of online media, the various prevalent event archetypes, and the concepts that emerge from the topical classification of the most popular events, while at the same time presenting an analysis paradigm.

2. Methodology

2.1. Research Design

The presented research was carried out in three stages. In the first stage, the data collection process took place using an advanced web data extraction algorithm in combination with GenAI technologies and human oversight. The process of web data extraction to collect factual information for the purpose of research in the fields of arts, the humanities, and medicine is becoming ever-more popular as more and more information becomes available online [29]. The value of using GenAI in data collection and analysis has been proven to bring forth novel insights and reduce the cost and time of analytical processes [30]. In the context of art- and culture-related texts, large language models have seen particular success, provided that the models are fine-tuned for their specific purpose [31].

In the second stage, as soon as a large amount of data was collected, the calculation of the impact factor was carried out based on an algorithm that utilized multiple quantitative metrics, including innate characteristics of the audiovisual events, their web outreach, their estimated scope, and the popularity of their various classifications. Both the transformation of unstructured data to texts and the use of texts and natural language processing to gain insights are very promising applications of this emerging technology [32]. Moreover, web presence is of vital importance in the field of art and culture content dissemination [33]. The collection of quantitative data from a large variety of events with different characteristics allows for an objective analysis of impact, even without access to detailed economic or activity data, as seen in previous methodologies [24,26].

Finally, in the last stage, the impact factor and the quantitative values from which it was derived were used as the basis for a thorough analysis of the dataset, leading to useful conclusions. Utilizing large datasets to enable quantitative research in the fields of art and culture allows for a new perspective that enriches research and discussion in these fields [34], and the combination of quantitative metrics with statistical analysis may lead to important insights, which would allow for better assessment of the landscape of art and culture [35].

2.2. Data Extraction

The main objective of the web data extraction algorithm was to extract information about actual contemporary audiovisual art events from the World-Wide Web. This information included textual descriptions of the event, its temporal and spatial characteristics, its classification with regard to the type of event, the methods and techniques used, and the subjects it covered. In order to achieve this, the algorithm was designed and implemented to be capable of taking HTML documents, converting their code into natural language, and subsequently extracting structured information from the produced text. In addition to that, the algorithm also collected information regarding the websites of media that reported on the audiovisual events and the people that contributed to those events.

The data extraction algorithm consisted of three main building blocks. The first part of the algorithm included the initial process of searching for publications on media websites. These publications were processed to create a collection of audiovisual events. The second and main part of the algorithm involved an iterative process through which the web was searched for additional publications about each of these events, and the content of these publications was converted into a graph of structured data that were merged with the previously recorded information about the event in the system's database. At the same time, the publications that were extracted but were not relevant to the investigated event led to the creation of new events to be investigated. Finally, the third part of the algorithm explored the websites that provided the various publications and extracted information about the media outlets themselves. The algorithm's output regarding the media outlets was curated and enriched through human oversight. The extraction algorithm's flowchart is presented in Figure 1.

As presented through the algorithm's flowchart, the process started with an initial small seed of websites and publications and through the iterative process of the web search and constantly expanded both to different events and to different media outlets. By configuring parameters such as how many initial publications and media were processed and how many non-relevant results were collected in the web search during the process of investigating each event, the algorithm could widen or narrow down the flow of new events. In order to create a graph of structured data from the free text of a publication, the algorithm made use of semantic web technologies as well as GenAI. The combination of these technologies enabled the processing of many different websites, which present information in vastly different manners. GenAI was also used in the process of comparing publications to existing events in order to determine relevancy.

When using GenAI, emphasis is placed by the algorithm on its capabilities to generate content summaries as well as extract information from an existing text or code sample. These two areas of use are widespread and provide satisfactory results [36]. The use of AI for tasks based on model knowledge (both current and historical) is deliberately avoided, since it often has a higher margin of error. During implementation, the OpenAI API was used, specifically the GPT-4o Mini model, which excels at generating natural text and code and combines satisfactory results with cost of use.

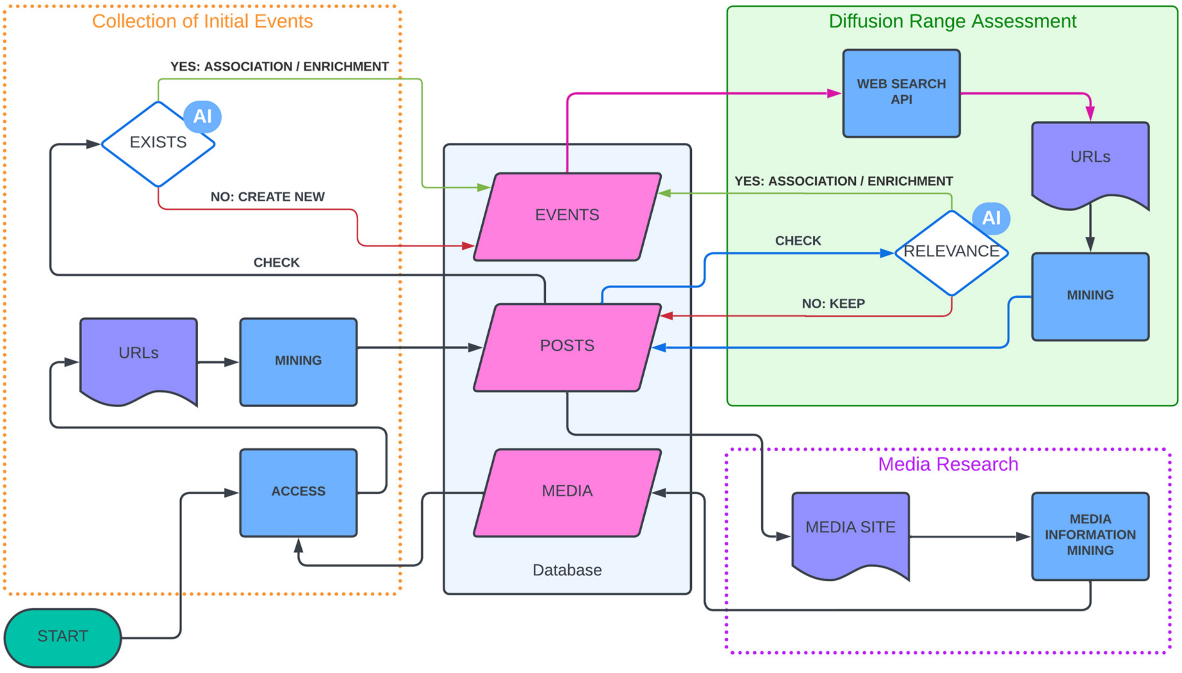

Figure 1. Data extraction algorithm flowchart.

Investigating the outreach of an audiovisual event demanded that the extraction algorithm was able to search the web for information without being limited to specific websites, no matter how many. In order to achieve this, the algorithm interacted with a programmatical search engine, through which it was able to search and find publications related to the event in question in real time. In the process of implementing the web search functionality programmatically, it was necessary to use an API that allowed for keyword-based search queries and provided relevant results including their URLs. The above process was achieved using Microsoft's Bing Web Search API. The algorithm asked the API for a list of relevant URLs and then extracted data from these URLs in order to enrich the information about the investigated event. The source media and webpages were also collected and stored in the system's database.

The data extraction algorithm was put into action during the spring and the summer of 2024 and provided information about thousands of events, publications, media, and contributors. The intention of the data gathering process was not of course to exhaustively discover every single audiovisual art event that took place in Greece but to create a representative sample of events that could provide interesting insights and be used for objective analysis. In the future, the data gathering process can be repeated to broaden the scope of the dataset as well as discover trends in the audiovisual art landscape.

During and after the data gathering process, human editors had access to the collected information through the graphical user interface (GUI) of an administration area. These editors included researchers and experts from the field of content management. Through the administration area, they could edit and update information collected by the algorithm, request the investigation of specific events or media outlets, and rectify (to an extent) the shortcomings of the results of the process. For instance, if during the data collection process an event reported by multiple publications was misjudged by GenAI to be two or more different events, a human editor could detect it by browsing events of similar titles and dates and, through an automated process, request the merging of these events into the main one. Human supervision was not aimed at altering the derived quantitative estimates but only at addressing simple consistency issues, leading to a more robust dataset.

2.3. Quantitative Metrics and Calculating the Impact Factor

For the study and analysis of the large volume of data collected and stored in the system's database, it was necessary to quantify the various characteristics that related to the impact of each event. An important role in this process was played by the intrinsic characteristics of the events that were directly measurable, such as the number of publications that referred to each event or the number of total contributors. Moreover, through analysis of the textual content of the publications, quantitative estimates concerning the size of each production and the range of the audience to which it was addressed were derived with the help of GenAI. At the same time, emphasis was also placed on the characteristics of the various media outlets that published information about each event and, in particular, on quantitative indicators of their popularity, such as their website traffic data and the size of their social media presence. Finally, the popularity of the tags that described and categorized each event was calculated based on the number of other collected events that shared common classification tags.

2.3.1. Media Outreach

An important part of the algorithm for estimating the impact of various audiovisual art events was the estimation of the corresponding influence of the online media outlets and websites that reported on these events.

This influence was based on the following two pillars:

- The first pillar concerned the online presence of websites, as described by their total monthly visits, backlinks, organic search results, popularity ranking in Greece, and their domain authority, which measures their relevance, as established by Moz [37].
- The second pillar concerned the presence of the media in the various social media platforms and the number of followers or subscribers on their official accounts.

The values obtained with the above measurements, as presented in detail in Table 1, were normalized with a logarithmic scale in base 10 to reduce the influence of outliers [38], and through their weighted average, two numerical indices were calculated, referring to each domain's web outreach and social media presence.

Table 1. Web and social outreach parameters and weights.

Indexes	Properties	Weight
Web Outreach	Total Visits	5
	Backlinks	4
	Organic Results	3.5
	Ranking in Greece	2
	Domain Authority	1.5
Social Outreach	YouTube	5
	Facebook	4.5
	X (Twitter)	3.5
	TikTok	3
	Instagram	1.5
	LinkedIn	1

With regard to web outreach, the most important factor was deemed to be the total number of visitors to the website and was thus given the biggest weight. The number of backlinks is a strong indication of how much a website is part of a larger informational ecosystem and, as such, was also given a large weight. This was followed by the number of organic results, which included only visits from a search engine. Finally, the Greek ranking and domain authority can also indicate the local or subject-related importance of a domain and, as such, were both included but with lower weights. All of the above metrics were collected using the third-party services of Similar Web (https://www.similarweb.com/, accessed on 5 December 2024) and Neil Patel (https://neilpatel.com/, accessed on 5 December 2024). Web analytics metrics such as the ones used here have been a source of objective information with regard to websites, visitors, and networks in many different applied research endeavors, and despite varied approaches, they continue to be a valuable tool of assessing influence [39]. In particular, in the field of news dissemination, where search engine optimization is paramount, web metrics become essential [40].

With regard to social outreach, the platforms Facebook, YouTube, Instagram, TikTok, X (Twitter), and LinkedIn were selected. The first four are amongst the most popular in terms of traffic, as reported by SimilarWeb, while X (Twitter) was included for its importance in news dissemination and LinkedIn for its importance in business and academic circles. Social media with a main focus on chatting (like WhatsApp, WeChat, Messenger, etc.) or with minimal presence in Greece, like Eastern platforms (Douyin, Baidu, etc.), were omitted. YouTube was given the highest rank due to its expertise in providing audiovisual content, followed by the long-established and popular Facebook. X (Twitter) followed in significance due to its nature as a news disseminator. TikTok's audiovisual nature is important, but its limited audience scope led to an average weight. Finally, Instagram, which limits URL linking, and LinkedIn, which is geared towards commerce, were also important but to a lesser extent. The evaluation of social media in terms of audiovisual event impact is quintessential, since consistent social media presence and an active and growing number of followers are important qualities that lead to extended information diffusion [41] and influence.

2.3.2. Event Impact Factor

In order to now investigate the impact of each individual audiovisual art event, the impact factor was established as a combined index that takes into account the events' intrinsic characteristics, the web and social outreach of the media that published information about the event, the estimations about the audience and production of the events based on the published texts, and the popularity of the classification tags of each event. Table 2 presents the various derived quantitative metrics with a short explanation.

The various metrics of Table 2 were used to create a numerical representation of each event, with each measurement normalized and weighted to contribute to the overall impact factor. The means of the web and social media outreach of the various websites were calculated using harmonic means. The harmonic means were calculated by using a series of weights following the harmonic progression and ensured that the most popular media had more influence and that multiple publications in various lower-outreach media or posts in social media channels with less of a following did not subtract too much from the overall result [42]. In the case of classification tag popularity, a harmonically weighted sum was used. This sum is an aggregate in which every element is weighted by a factor that follows the harmonic progression, and it ensured that both the number of tags and the number of events in each tag were factored in, with more importance given to the most popular tags. The audience size and production cost metrics followed a scale from 1 to 10 and were estimated through the use of GenAI based on the textual content of the publications that reported on each event. The OpenAI GPT-4o model used excels in both reasoning and multimodal reasoning tasks, making these estimates a valuable metric.

Table 2. Quantitative metrics comprising the impact factor.

Metrics	Description	Weight
Web Outreach	Weighted harmonic mean of the web outreach of all media that reported on the event	5
Social Outreach	Weighted harmonic mean of the social outreach of all media that reported on the event	4.5
Number of Publications	The number of media that reported on the event	3.5
Audience Size	Audience size as estimated from 1 to 10 through textual analysis of the event descriptions by GenAI	2
Production Cost	Production cost as estimated from 1 to 10 through textual analysis of the event descriptions by GenAI	2
Number of Contributors	The total number of contributors related to the event	1
Type Popularity	Harmonically weighted sum of events with the same type tags with diminishing weights	1
Method Popularity	Harmonically weighted sum of events with the same method or technique tags with diminishing weights	1
Subject Popularity	Harmonically weighted sum of events with the same subject tags with diminishing weights	1

In order to combine the above metrics into the impact factor, all values were normalized using the z-score and were shifted and scaled to be between 0 and 1000. Normalization was necessary so that the measuring units of each different metric did not have an impact when they were considered in order to form the impact factor. Z-score normalization ensured that the distances between different values were retained [43], while the shifting and scaling to this specific value range allowed for better readability of the values by the general public. A four-digit scale was selected to allow for a variety of different values and distances between instances without the use of decimal numbers.

The final calculation of the impact factor used a weighted average approach incorporating the difference in importance for each metric. The weights for each metric are presented in the third column of Table 2. These weights were based on the importance of the various different metrics on the basis of the current bibliography and observation of the contemporary landscape. Web and social media outreach, encompassing multiple objective properties of the various media, were deemed the most important, with the number of publications following closely, since it is a good intrinsic indicator of outreach. The popularity of each classification category was weighted in a manner such that their total classification popularity was averagely important. The values estimated from publication texts were both weighted just below that, since they included an added insight but could be subjective or unreliable due to the nature of GenAI. Finally, the lowest-weighted metric was the number of contributors, which does have some intrinsic value in conveying the scale of an event but is not as reliable due to how media often fail to report on every contributor.

The metrics' transformation and impact factor calculation process were conducted for every investigated event in the dataset. The sum of the collected information and their outreach-related metrics are also openly available online through a digital repository at the address https://repo.artdata.gr (accessed on 5 December 2024) as well as a fully functional REST API. The repository includes many different ways of browsing events based on their classification and their spatial and temporal characteristics through an interactive map and a calendar system. Both the sum of the collected information and the repository itself are available in Greek, since this was the main language of the collected publications and derived events.

3. Results

Based on the processes and algorithms detailed in Section 2, a large number of audiovisual art events were discovered on the web and a series of numerical metrics were used for the data-driven quantification of their impact, culminating in the encompassing impact factor. During the discovery process, a large variety of data were collected, including audiovisual art events, online media outlets, news publications regarding the events, classification tags, and event-related people (or groups). Table 3 presents the number of instances collected per data entity.

Table 3. Overview of data collected.

Informational Entity	Number of Instances
Events	16,501
Media	2456
Publications	30,803
People	40,459
Classification Tags	7622

Out of a total of over 16,000 discovered events, 2679 were fully investigated with regard to their web presence, and their impact factor was accurately calculated using all available quantifiable metrics.

3.1. Impact Factor Statistics

The impact factor value, as derived from the various quantitative metrics, can act as a practical measurement of each event's performance, as demonstrated by its statistical values that are presented in Table 4.

Table 4. Overview of the impact factor's descriptive statistic values.

Informational Entity	Number of Instances
Mean	445.39
Standard Deviation	95.12
Q1	381
Q2—Median	441
Q3	512.5
Interquartile Range	131.5
Minimum	125
Maximum	835
Range	710

The impact factor statistics provide useful insights into the metric itself. The fact that the mean and median were relatively close to each other indicates a distribution that was mostly symmetrical with a very mild skew. The standard deviation and interquartile range values displayed reasonable variability without an excessively spread-out dataset. While a central concentration of values was demonstrated by the quartile values, the existence of some outliers was made apparent by the large overall range compared to the interquartile range. The above qualities of the metric and dataset are presented graphically through a normal distribution graph in Figure 2.

The properties discussed above present a metric that allows for a straight-forward and practical evaluation of each event's impact. The majority of the events displayed an impact factor close to the median and average as expected, while some special instances of both very impactful and very unimpactful events were also observed. The robustness of the metric is reinforced by evaluating these outliers, which included multiple high-profile concerts of local and international stars, as well as cultural performances of major significance on one end and small-scale independent performances and presentations of limited or local interest on the opposite end. The existence of true outliers demonstrates the metric's flexibility in handling a variety of events. It is safe to say that the metric presents a useful tool for conveying the estimated influence it is intended to represent to all interested parties.

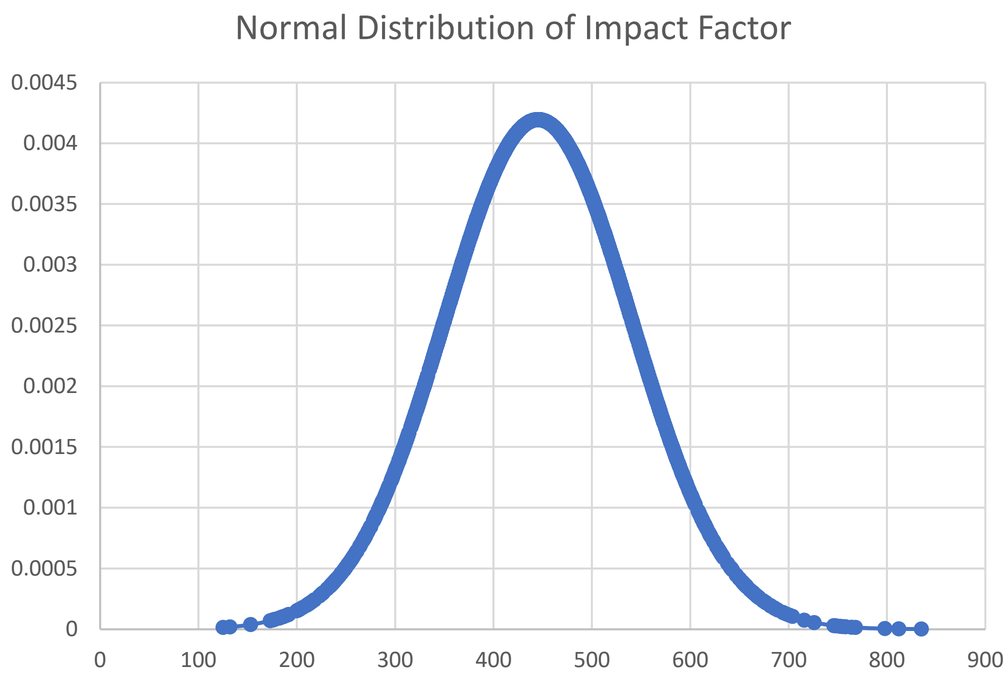

Figure 2. Graph of normal distribution of impact factor.

3.2. Metrics Relationships

Each individual quantitative metric presented in Table 2 contributed to the calculation of the impact factor, as detailed in Section 2.3.2. These metrics, while largely independent from each other, acted as separate properties, adding up to a conclusive result. In this context, it is important to observe the correlations between themselves as well as between them and the impact factor. Figure 3 presents a correlation matrix, which provides more insight into these relationships.

As presented in Figure 3, the majority of metrics demonstrated weak-to-moderate correlations with each other, despite representing independent measurements. Production size and audience size, which are both metrics inferred through the textual content, were very closely related and so were web outreach and social media outreach, which are derived through the presence of individual events on various online media. On the other hand, the lowest correlations were observed between these media related metrics and the metrics that derive from the popularity of various classification tags. In no case did we see negative correlations, which would indicate a metric working against its intended purpose. The correlation between each metric and the impact factor was of course expected, since the impact factor is the direct result of the combination of those metrics. The graph shown in Figure 3 reaffirms the robustness of the metric by hinting at the positive relationship between its various aspects, which, in turn, ensure its consistency and objectivity.

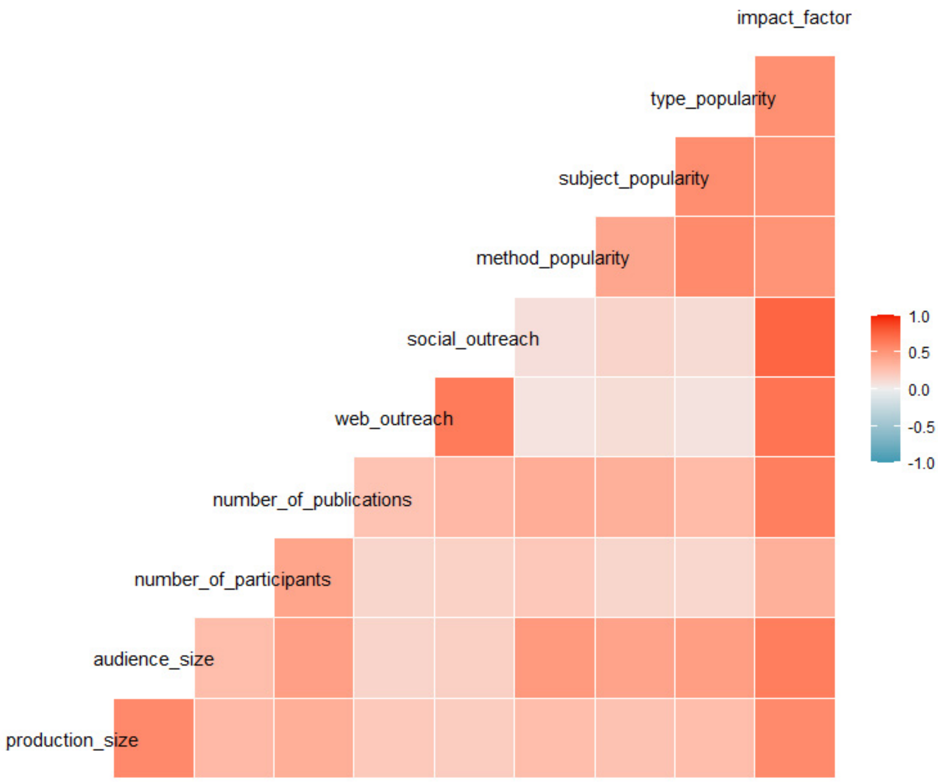

Figure 3. Correlation matrix between the various quantitative metrics.

3.3. Media Popularity and Media Relevance

Using the calculated impact factor as a means to identify impactful audiovisual art events in Greece provided insight into the difference between the general popularity of a website and its usefulness in providing information about popular events. Table 5 presents the top 20 most popular websites discovered in the first column, as indicated by their web presence and social presence metrics, while in the second column, the websites that provided the highest number of popular events are displayed.

Table 5. Comparison between popular media and media providing the most events.

Top 20 on the Web	Top 20 for Popular Events
nytimes.com	tickets.public.gr
imdb.com	musicity.gr
vice.com	typosthes.gr
leparisien.fr	thestival.gr
en.wikipedia.org	stagenews.gr
soundcloud.com	makthes.gr
linkedin.com	ogdoo.gr
songkick.com	zougla.gr
skai.gr	thessnews.gr
news247.gr	rockrooster.gr
gazzetta.gr	thessalonikicityguide.gr
antenna.gr	voria.gr
megatv.com	ticketservices.gr
newsbomb.gr	neolaia.gr
sport-fm.gr	thes.gr
lifo.gr	tetragwno.gr
zougla.gr	in.gr
athensvoice.gr	rockoverdose.gr
unboxholics.com	karfitsa.gr

The left column is dominated by popular Greek and global media outlets as well as repositories of audiovisual content. This is an indicator that the media-related metrics worked as intended, bringing the online media with the highest traffic and social media presence to the top. But these websites are different from the ones that appear in the second column of Table 5. This column is dominated by outlets focusing specifically on events and entertainment, as well as ticket providers and cultural guides. Some general-purpose Greek media outlets with strong culture and entertainment coverage also appear there, especially as we move further down the full lists of domains.

This difference is a good indicator that websites geared towards providing specialized information are especially relevant for audiovisual events, where practical details like ticket prices, venues, and reviews matter more. Event organizers, as they should, focus on promotion through local platforms, which are better aligned with the needs of the intended audiences. Both locality to the area of interest and specificity of the content are important factors in event dissemination, and this is made apparent in the second column. Using the impact factor as an analysis tool, we were able to generate this list, which may be used by organizers to target promotion and maximize the visibility of their event to their target groups as well as allow for the study of audience overlap and the possibilities of cross-promotion between the various domains.

3.4. Location Importance

The importance of the location where an audiovisual art event takes place may also be further investigated through the impact factor. In general, it was observed that events tended to take place near major cities, especially the larger ones. Figure 4 presents a visual representation of the locations of the audiovisual events with the above average impact factor.

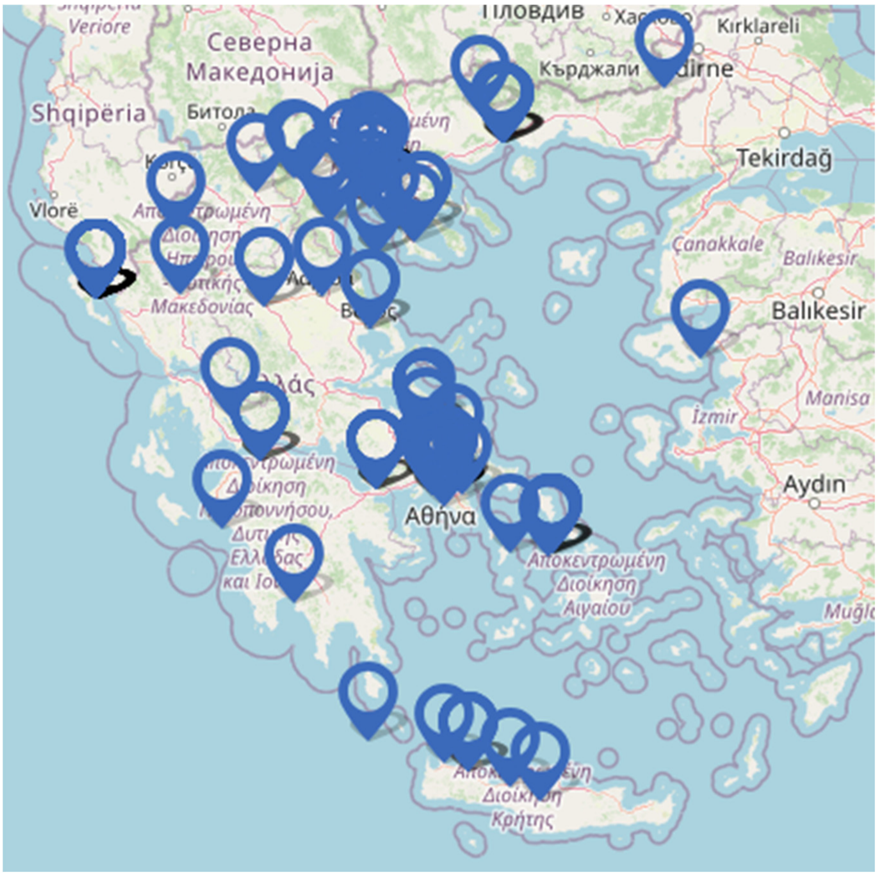

Figure 4. Map of events with above average impact factor.

As made clear by the map, events were much more prevalent in or near large metropolitan areas such as Athens and Thessaloniki. The events taking place in the rest of Greece were often located in rural cities and, in rarer cases, in open spaces of natural or cultural importance. As seen on the map, destinations with high popularity in terms of tourism also presented a significant number of events (Crete, Syros, Corfu, Lesbos, etc.). Table 6 presents a comparison between the percentage of events that took place in large cities compared to rural events, both for all the discovered events and for only the events with above average impact factor. Additionally, it presents the impact factor average in these two different groups of events.

Table 6. Large city versus rural events.

	Large City	**Rural**
Percentage of Total Events	72.17%	27.88%
Percentage of Popular Events	78.92%	21.08%
Average Impact Factor	456	417

It is notable that large city events were not only more numerous than rural events but also showed a tendency to further increase their advantage as the impact factor increased. Moreover, the average impact factor for events in large cities was observed to be higher than that of rural events.

3.5. Event Archetypes

Going beyond descriptive statistics, the use of computational methods may also lead to useful conclusions based on the various quantitative metrics of Table 2 and the impact factor. Using these metrics, each audiovisual art event may be described by numbers, which form the basis for applying a clustering methodology in order to classify each event instance into a different group. In order to achieve this, the K-Means clustering algorithm was used, since it ensures that all events are clustered and guarantees convergence. Using the metrics' mean values did not cause any issues, since the variability within the metrics was not extreme.

Before the clustering methodology was applied, all rows of the dataset with empty values were omitted, and the metrics were scaled with the z-score method to have a mean of 0 and a standard deviation of 1 [43]. This ensured that every metric was considered equally regardless of their units of measurement. Additionally, it was essential that the number of clusters was determined before the application of the algorithm. The "total within sum of square" method was utilized in tandem with the gap statistic method provided by R's factoextra [44] and cluster libraries [45], respectively. Figure 5 presents the results in the form of graphs. Specifically, for the calculations needed to plot the gap statistic chart, the parameters in Table 7 were employed.

Table 7. Parameters for the gap statistic calculation.

Name	**Value**
K.max	10
B	50
d.power	1
nstart	25
method	firstSEmax

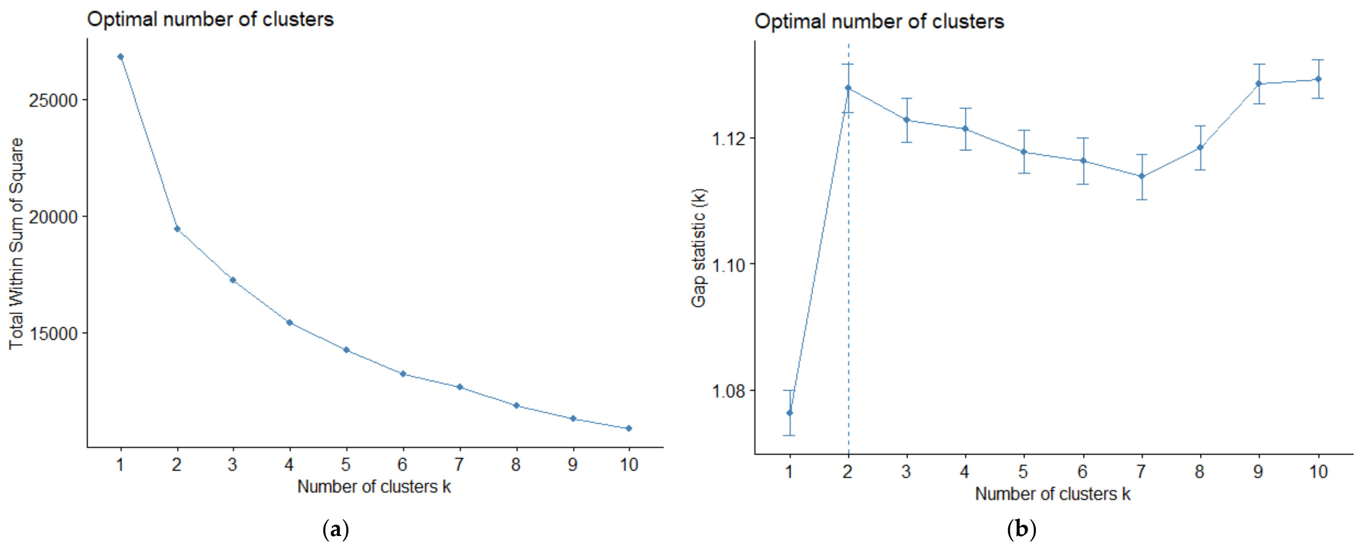

Figure 5. WSS and gap statistic graphs for determining the number of clusters. (**a**) WSS (**b**) Gap Statistic.

In order to strike a good balance between having a small but significant number of clusters and ensuring a low WSS value at the middle of the curve of the chart and a high gap statistic value, the optimal cluster value was decided to be 4. The decision process considered both the shape and values of the charts shown in Figure 5 and the purpose of the clustering analysis, which led to more robust conclusions with a limited number of clusters. The four clusters that were generated along with their number of events and the means for every metric used are presented in Table 8.

Table 8. Items and means for each different cluster.

	1	2	3	4
items	744	386	828	721
production_size	5.54	6.58	4.50	5.53
audience_size	6.69	7.35	5.20	5.91
number_of_participants	6.20	17.01	3.65	6.15
number_of_publications	3.58	12.99	1.53	3.02
web_outreach	551.66	640.72	516.42	631.84
social_outreach	350.61	491.92	289.24	480.24
method_popularity	3965.15	4290.64	861.03	1070.42
subject_popularity	1907.14	2418.11	543.01	767.85
type_popularity	10,155.69	10,228.12	4068.07	5357.25
impact_factor	463.72	588.45	341.78	468.90

Figure 6 presents the four clusters and their topography, including every point which corresponds to a single audiovisual art event, as described by its quantitative metrics.

Based on the calculated means and the topography of the chart, Cluster 3 (blue square) represents events that did not manage to score well in any of the metrics. The shape of the cluster indicates a variety of different combinations of properties, but none of these events gained enough traction to have a significant impact. On the other side of the spectrum, Cluster 2 (green triangle) represents events that had high metrics across the board. The elongated shape of the cluster indicates that some major outliers, including very influential events, displayed a big distance from the average event. Clusters 1 and 4 represent middle-of-the-pack events. The main differentiation between them is that Cluster 4 (purple cross) events showed high web and social outreach scores, which were accompanied with lower scores in other metrics, indicating events that gained some traction on the web but were limited by their scope, type, method, or subjects, while Cluster 1 (red circle) events had low outreach metrics despite belonging to popular classifications and aiming for bigger audiences. This paints the picture of some unique events aiming for a niche audience in Cluster 4 on one hand and events of popular types that did not manage to reach their intended audience in Cluster 1 on the other. Neither of these groups of events managed to achieve a high impact, but for apparently different reasons. Over time though, and by means of exposure through proper channels such as event aggregators or social media with more persistent attributes, events from Cluster 4 may gain traction in the future and move to Cluster 2.

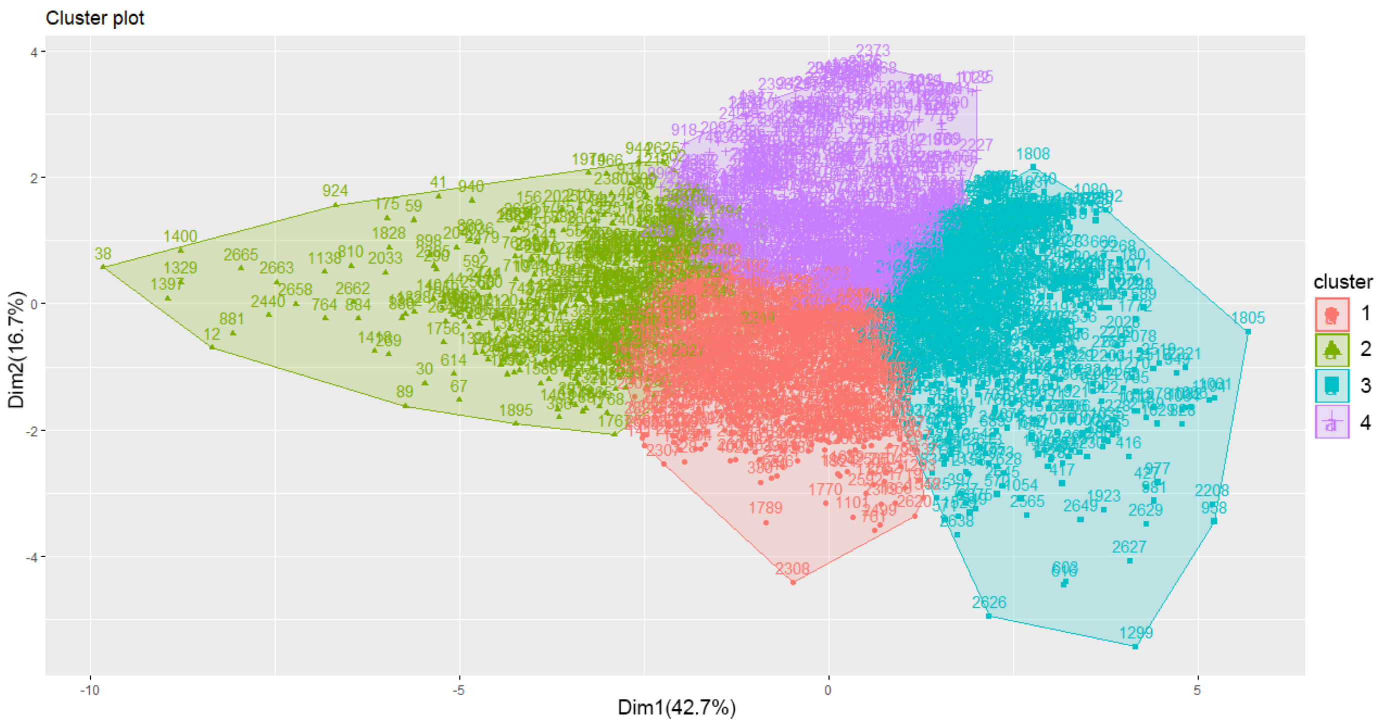

Figure 6. Clusters of audiovisual art events for four centers through K-Means.

3.6. *Topic Influence*

Computational statistics and machine learning methodologies may be employed not only by using the numerical representations of the discovered audiovisual art events but also on the basis of the textual content that describes them. Using the impact factor as an indicator, the more popular events of the dataset were collected, and the texts describing them formed a corpus of documents. A series of topics were produced on the basis of this corpus using the Latent Dirichlet Allocation (LDA) generative statistical model [46] through a fully unsupervised text-mining methodology [47]. The bag-of-words approach of the LDA algorithm combined with its unsupervised approach ensures that the topic results were representative of the concept and wording presented in the corpus and that they did not carry the prejudices of the researchers but only the perception conveyed by the media outlets that provided that information.

In order to proceed with the LDA process, it was necessary to devise the optimal number of topics in a manner similar to the numerical clustering technique. In order to achieve this, the R language's ldatuning library was employed [48], and the methodologies presented by Cao Juan et al. [49], Devaud et al. [50], Arun et al. [51], and Griffiths et al. [52] were used to create the four distinct plots that appear in Figure 7.

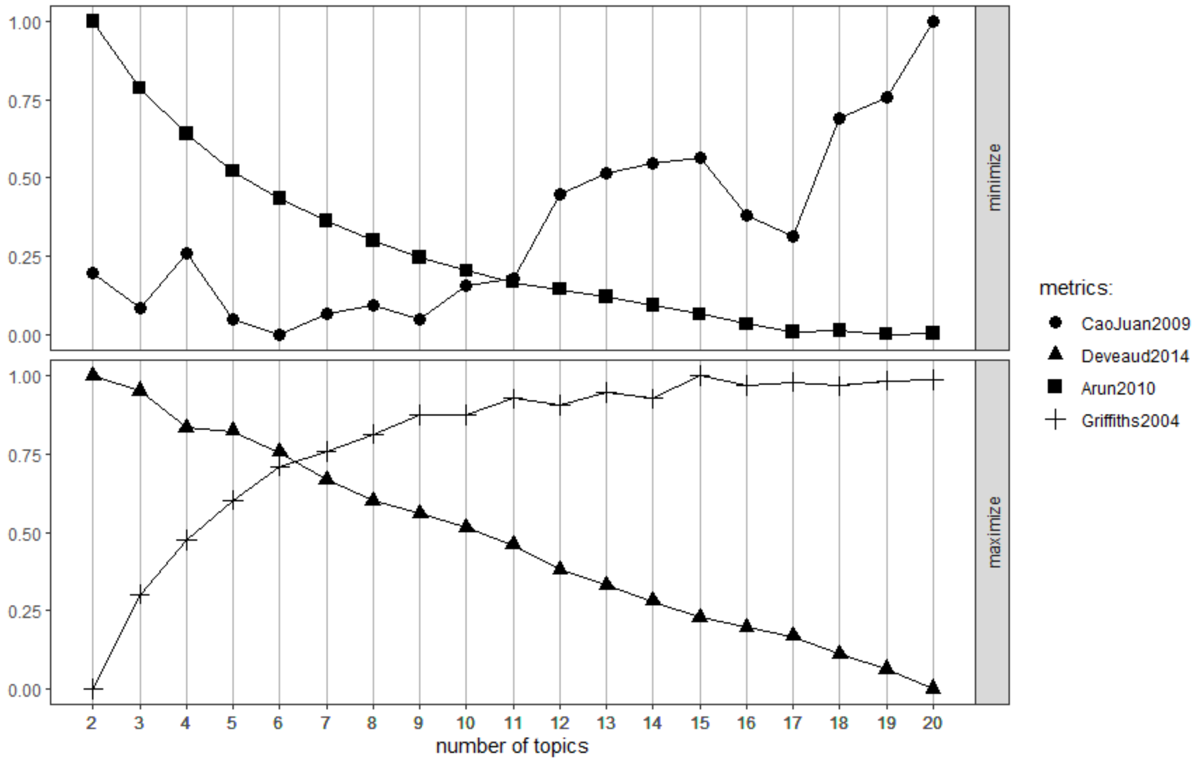

Figure 7. Chart used to determine the optimal number of topics [49–52].

Based on Figure 7, the optimal number of topics, which ensured the maximization of the Devaud and Griffiths plots and the minimization of the Cao Juan and Arun plots, was determined to be six. This number also fits our analysis in terms of context, since it provides a limited but not insignificant number of topics. In order to proceed with the LDA analysis, the model was fitted using the Gibbs method with a total of six topics through the R language's LDA library [53]. Figure 8 presents the intertopic distance map provided by the LDA methodology and the LDAVis library [54].

An overview of the various detected topics is presented in Table 9, which includes translations of high-occurrence tokens, relevant low-occurrence tokens, and an overview of the topic.

Table 9. Topics detected during LDA topic modeling.

Topic	High Occurrence	Low Occurrence	Overview
1	club, live, Saturday, night, band	show, experience, unique, audience, concert	Popular entertainment and music events within the context of weekend nightlife
2	festival, documentary, movie, feature, cinema	short, international, director, opening, screening	Movies and cinema, especially in the context of festivals and competitions
3	work, title, exhibition, art, creator	modern, museum, institution, production, collection	Artwork exhibitions and events in the context of culture and the visual arts
4	music, song, program, orchestra, hall	composer, leading, perform, symphony, conducting	Music events in the context of classical or contemporary artistic music
5	concert, year, event, Greek, celebrate	philharmonic, artist, plaza, traditional, summer	Musical concerts in the context of seasonal events and outdoor celebrations
6	performance, theater, stage, history, Sunday	direction, group, culture, actor, music	Performance art events such as theater or musical theater

Figure 8. Intertopic distance map.

As seen in the intertopic map, all topics were of a similar caliber with a less than 1% difference in token representation. Moreover, their spread throughout the map and in all four quadrants reflects their coherence. Topics 1 and 5, which deal with music and nightlife and musical concerts, respectively, showed a significant overlap, which was expected due to the similarities of the subject matter. Additionally, Topics 6 and 4, which deal with performance arts and musical events, were located closely together and presented a small overlap. Topics 2 and 3, which deal with movies and exhibitions, were isolated and unique.

Music, in general, appeared to have a dominant impact in the landscape of audiovisual art events, with the various musical aspects representing at least three different topics (1, 4, and 5) while also appearing in others. The topics of performance arts, cinema, and art exhibitions were also distinct and well defined. The vast array of different types of events is indicative of the multifaceted landscape of audiovisual art events which range from popular entertainment and local tradition to high art and culture. Detecting the various topical categories and their main concepts and terms through the process of topic modeling allows a broader perception of the field of contemporary audiovisual art.

4. Discussion

4.1. RQ1: Is It Feasible to Quantify the Impact of an Event of Audiovisual Art Through Its Online Presence Using a Data-Driven Approach?

As presented in the Methodology and Results sections, a series of sufficiently independent parameters that influence the impact of an audiovisual art event were established and used to create an encompassing impact factor. These quantitative metrics included inherent quantitative characteristics of the events, such as the number contributors and informational publications on the web. They were further complemented with objective web and social media outreach metrics, which have been established as important factors in dissemination [39,40,51], especially in the artistic field [55]. Moreover, the analysis of the textual content of each event's web publications allowed for the creation of numerical estimates of the audience size and the production cost through the analytic prowess of GenAI, which has been shown excel in similar tasks in the past [36]. Finally, the set of metrics was completed by measuring the popularity of the various classification tags describing each event within the collected event dataset. Leveraging popular themes and methodologies stimulates audience engagement and increases the chances of success [56]. These established quantitative metrics were combined to create the impact factor metric, which was calculated for each event of a real-world dataset extracted from the web.

As presented by the correlation matrix shown in Figure 3, the above parameters showed mostly weak positive correlations between them, which indicates that while mostly independent, they work well in tandem. Moreover, based on the collected event dataset, the calculated impact factor displays many statistical properties that allow it to be a good metric both in terms of consistency, flexibility, and representation and in terms of readability. The establishment of the various quantitative metrics and the positive attributes of the impact factor's statistical analysis demonstrate that it is in fact feasible to quantify the impact of an audiovisual art event through its online presence using data collected from the web. Additionally, the use of this dataset, the metrics, and the impact factor to produce useful insights, as presented in the Results section and discussed in RQ2 below, further reinforce this conclusion.

4.2. RQ2: What Insights Can Be Derived Regarding the Greek Audiovisual Event Landscape Through Implementing This Methodology?

Using the impact factor as a means to identify impactful audiovisual art events in the dataset and comparing the websites that provided the greatest number of such events with a general web and social media ranking of all collected websites provided a clear indication about the importance of specialized media outlets in the promotion of art events. Specialized websites allow for the cultivation of specific audience segments [57], while ticketing platforms promote content and assist both artists and audiences [58]. The second list of websites highlighted through the use of the impact factor demonstrates a substantial alternative to targeted promotion for artists and organizers.

Through the use of the impact factor, an objective and quantified measurement was carried out regarding event locations and their relevance to both event quantity and influence. Art events in major cities become part of an effort for urban entrepreneurialism and shape urban communities and local culture [59,60] and thus have a stronger influence in the cultural landscape. Rural areas, on the other hand, attempt to achieve a cultural revitalization, but cultural policy directed to them remains underdeveloped [61], as the findings of this study also suggest being the case in Greece.

A better understanding of the art event landscape may be achieved through the employment of computational machine learning methods, as suggested by Manovich [18] and also established through this research. The four clusters of art events presented in the results section are indicative of high- and low-impact events (Clusters 2 and 3) and also middling events that struggle to gain major influence due to various reasons that revolve around limited dissemination (Cluster 1) or a niche subject and audience (Cluster 4). Major art events that combine intrinsic value, subject popularity, production scale, and active dissemination (all metrics that the impact factor assesses) stand out from the pack and become culturally significant [59].

Going beyond numerical metrics and into the topical analysis of terms and concepts of impactful events, the emergence of music as a dominant force in the Greek event landscape was established. The popularity, the importance, and the multiple facets of music as, presented by Topics 1, 4, and 5, are a result of both an appreciation of musical esthetics and the communal nature of sharing musical experiences [62], which, in turn, further increases impact. This importance of music and audio is even more pronounced in the younger Gen-Z generation [63]. Through the process of topically modeling these musical topics alongside the other major topical groups, which involve cinema, the performing arts, and exhibitions, a complete abstract overview of the Greek audiovisual event landscape emerges.

5. Conclusions

Within the context of this research, a distinct methodology of collected factual contemporary information about audiovisual art events from the web through the use of data extraction and generative artificial intelligence was presented. This information was distilled in multiple quantitative metrics based on event characteristics, web and social media outreach, estimated scope and type, method, and subject popularity. These metrics were consolidated in an impact factor value, which may abstractly represent each event's perceived influence. Through the use of this impact factor and descriptive and computational statistical analysis of all metrics, the landscape of audiovisual art events in Greece was outlined.

Of course, any quantitative impact estimation is limited due to the multi-faceted nature of audiovisual art events. Some important events might have limited exposure on the web due to their subversive or underground nature, while the presence of others might be artificially maximized through traditional and viral marketing tactics. Moreover, annual or celebrational events might have added influence, which is hard to quantify without a trend analysis. The temporal nature of impact may be better addressed by repeating the collection process at various points in the future and assessing the evolution of each event's impact. More importantly, these metrics work best in tandem with expert knowledge and qualitative insight, which help paint a clearer picture of how contemporary art and culture events reach the audience. The proposed impact factor has the potential to become a useful tool to assist in decision-making regarding audiovisual art events but is not intended to replace expert knowledge and critical thinking.

Utilizing and expanding the existing information in the future may lead to new insights about trends and changes in the landscape.

Beyond collecting more events in the future, repeating the outreach measurement of the same events at a later date may also lead to interesting conclusions regarding the impact of more persistent forms of dissemination. At the same time, multiple other analytical approaches may be employed to reach conclusions, especially with regard to the effect of method and subject classification on the impact. Specific subsets of events may be analyzed based on the existing or an expanded dataset. These may include events taking place in specific places (like popular tourist destinations) or annual, seasonal, and celebrational events. These new research avenues signify that both the analysis methodology and the collected dataset itself are important results of this research. The dataset, which is available openly and through an API, may be used in future work by other researchers and interested parties to reach conclusions based on these subsets or regarding different matters.

During the course of this research, it became apparent that the web provides a plethora of information, which, with proper methodology, may be used to provide quantitative feedback in various aspects of art and culture that are usually discussed in qualitative terms. Modern digital technologies, including the semantic web, GenAI, and search technologies, allow us to harvest this information and use it in tandem with expertise and qualitative observation to achieve more informed decision-making. This approach may be useful to a vast array of fields, including art, culture, social sciences, and the humanities in general.

The collected dataset, which is also openly available, alongside the methodologies for extracting data and establishing the quantitative representation of each event's impact and the presented paradigm of analysis have the potential to become useful tools in the service of event organizers, artists, marketing professionals, and any other interested parties. Through these tools, a better understanding of both artistic intention and audience needs and motivations can be achieved.

Author Contributions: Conceptualization, A.G. and M.P.; data curation, A.L.; formal analysis, M.P. and A.L.; funding acquisition, A.G.; investigation, A.L. and S.L.; methodology, A.G., M.P. and S.L.; project administration, S.L.; resources, A.G. and S.L.; software, M.P. and A.L.; supervision, A.G.; validation, A.L. and S.L.; visualization, A.L.; writing—original draft, M.P.; writing—review and editing, A.G., M.P., A.L. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hellenic Foundation for Research & Innovation, within the framework of the "2nd Call for H.F.R.I.'s Research Projects to Support Faculty Members & Researchers" (project number 3607).

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: The study did not involve humans.

Data Availability Statement: The data presented in this study are openly available in Zenodo at [https://zenodo.org/records/14295447 (accessed on 7 December 2024)], reference number [10.5281/zenodo.14295447].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bygren, L.O.; Konlaan, B.B.; Johansson, S.E. Attendance at cultural events, reading books or periodicals, and making music or singing in a choir as determinants for survival: Swedish interview survey of living conditions. *BMJ* **1996**, *313*, 1577–1580. [CrossRef] [PubMed]
2. Soman, T.; English, A. The uses of art in the context of health and diversity. *Paediatr. Child Health* **2010**, *15*, 9–10. [CrossRef]
3. Wikström, B.M. The dynamics of visual art dialogues: Experiences to be used in hospital settings with visual art enrichment. *Nurs. Res. Pract.* **2011**, *2011*, 204594. [CrossRef] [PubMed]
4. Tepper, S.J. *Artful Living: Examining the Relationship between Artistic Practice and Subjective Wellbeing Across Three National Surveys*; Curb Center for Art, Enterprise, and Public Policy at Vanderbilt University: Nashville, TN, USA, 2014.
5. Allan, J. Arts and the inclusive imagination: Socially engaged arts practices and Sistema Scotland. *J. Soc. Incl.* **2010**, *1*, 111–122. [CrossRef]
6. Charlton, M.; Brandt, D.; Dennis, K.; Donegan, R. Transforming Communities Through the Arts: A Study of Three Toronto Neighbourhoods. *Tor. Arts Found.* **2013**. Available online: https://torontoartsfoundation.org/tac/media/taf/Research/Transforming%20Communities%20Through%20the%20Arts/Transforming-Communities-Through-the-Arts-Full-Report.pdf (accessed on 5 December 2024).
7. Davies, C.R.; Knuiman, M.; Wright, P.; Rosenberg, M. The art of being healthy: A qualitative study to develop a thematic framework for understanding the relationship between health and the arts. *BMJ Open* **2014**, *4*, e004790. [CrossRef]
8. Josephson, S.G. *From Idolatry to Advertising: Visual Art and Contemporary Culture*; Routledge: London, UK, 2016.
9. Komianos, V.; Tsipis, A.; Kontopanagou, K. Introducing Digitized Cultural Heritage to Wider Audiences by Employing Virtual and Augmented Reality Experiences: The Case of the v-Corfu Project. *Technologies* **2024**, *12*, 196. [CrossRef]
10. Janssen, S. Art journalism and cultural change: The coverage of the arts in Dutch newspapers 1965–1990. *Poet* **1999**, *26*, 329–348. [CrossRef]
11. Pavelka, J. The factors affecting the presentation of events and the media coverage of topics in the mass media. *Procedia-Soc. Behav. Sci.* **2014**, *140*, 623–629. [CrossRef]
12. Bhosale, H.S.; Gadekar, D.P. A review paper on big data and hadoop. *Int. J. Sci. Res. Publ.* **2014**, *4*, 1–7.
13. von Bloh, J.; Broekel, T.; Özgun, B.; Sternberg, R. New (s) data for entrepreneurship research? An innovative approach to use Big Data on media coverage. *Small Bus. Econ.* **2019**, *55*, 673–694. [CrossRef]
14. Rosenbush, S.; Totty, M. How big data is changing the whole equation for business. *Wall Str. J.* **2013**, *11*, 1–3.

15. European Statistical System Committee. Scheveningen Memorandum on 'Big Data and Official Statistics'. 2013. Available online: https://ec.europa.eu/eurostat/documents/13019146/13237859/Scheveningen-memorandum-27-09-13.pdf/2e730cdc-862f-4f27-bb43-2486c30298b6?t=1401195050000 (accessed on 5 December 2024).
16. European Statistical System Committee. Big Data Action Plan and Roadmap. 2014. Available online: https://www.cepal.org/sites/default/files/presentations/presentacion-eurostat-and-big-data-rosemary-montgomery-eurostat.pdf (accessed on 5 December 2024).
17. Moretti, F. Conjectures on world literature. *New Left Rev.* **2000**, *2*, 54–68.
18. Manovich, L. The science of culture? Social computing, digital humanities and cultural analytics. *J. Cult. Anal.* **2016**, *1*. [CrossRef]
19. Martinho, T.D. Researching Culture through Big Data: Computational Engineering and the Human and Social Sciences. *Soc. Sci.* **2018**, *7*, 264. [CrossRef]
20. Amer-Yahia, S.; Leroy, V.; Termier, A.; Kirchgessner, M.; Omidvar-Tehrani, B. *Interactive Data-Driven Research: The Place Where Databases and Data Mining Research Meet*; LIG: Tokyo, Japan, 2015.
21. Brynjolfsson, E.; Hitt, L.; Kim, H. How does data-driven decision-making affect firm performance. In Proceedings of the Workshop for Information Systems and Economics at the 9th Annual Industrial Organization Conference, Busan, Republic of Korea, 16–18 August 2011; pp. 8–10.
22. Tamm, T.; Hallikainen, P.; Tim, Y. Creative analytics: Towards data-inspired creative decisions. *Inf. Syst. J.* **2022**, *32*, 729–753. [CrossRef]
23. Panneels, I.; Lechelt, S.; Schmidt, A.; Coşkun, A. Data-driven innovation for sustainable practice in the creative economy. In *Professor Christopher Smith, Executive Chair of the Arts*; Routledge: London, UK, 2024; p. 243.
24. Saayman, M.; Saayman, A. Economic impact of cultural events. *S. Afr. J. Econ. Manag. Sci.* **2004**, *7*, 629–641. [CrossRef]
25. Herrero, L.C.; Sanz, J.Á.; Devesa, M.; Bedate, A.; Del Barrio, M.J. The economic impact of cultural events: A case-study of Salamanca 2002, European Capital of Culture. *Eur. Urban Reg. Stud.* **2006**, *13*, 41–57. [CrossRef]
26. Lucia, M.D.; Zeni, N.; Mich, L.; Franch, M. Assessing the economic impact of cultural events: A methodology based on applying action-tracking technologies. *Inf. Technol. Tour.* **2010**, *12*, 249–267. [CrossRef]
27. Colombo, A. How to evaluate cultural impacts of events? A model and methodology proposal. *Scand. J. Hosp. Tour.* **2016**, *16*, 500–511. [CrossRef]
28. Devesa, M.; Roitvan, A. Beyond economic impact: The cultural and social effects of arts festivals. In *Managing Cultural Festivals*; Routledge: London, UK, 2022; pp. 189–209.
29. Landers, R.N.; Brusso, R.C.; Cavanaugh, K.J.; Collmus, A.B. A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychol. Methods* **2016**, *21*, 475. [CrossRef] [PubMed]
30. Dhoni, P. Exploring the synergy between generative AI, data and analytics in the modern age. *TechRxiv* **2023**. [CrossRef]
31. Trichopoulos, G.; Konstantakis, M.; Alexandridis, G.; Caridakis, G. Large Language Models as Recommendation Systems in Museums. *Electronics* **2023**, *12*, 3829. [CrossRef]
32. Vigenesh, M.; RaviKumar, V.; Patil, V.D.; Kumar, S.S.; Geyi, N.; Chattopadhyay, S. Assessing the Ability of AI-Driven Natural Language Processing to Accurately Analyze Unstructured Text Data. In Proceedings of the 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Bangalore, India, 29–31 December 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
33. Giannakoulopoulos, A.; Pergantis, M.; Konstantinou, N.; Kouretsis, A.; Lamprogeorgos, A.; Varlamis, I. Estimation on the importance of semantic web integration for art and culture related online media outlets. *Future Internet* **2022**, *14*, 36. [CrossRef]
34. Christ, A.; Penthin, M.; Kröner, S. Big data and digital aesthetic, arts, and cultural education: Hot spots of current quantitative research. *Soc. Sci. Comput. Rev.* **2021**, *39*, 821–843. [CrossRef]
35. Pergantis, M.; Varlamis, I.; Kanellopoulos, N.G.; Giannakoulopoulos, A. Searching Online for Art and Culture: User Behavior Analysis. *Future Internet* **2023**, *15*, 211. [CrossRef]
36. Khashabi, D.; Min, S.; Khot, T.; Sabharwal, A.; Tafjord, O.; Clark, P.; Hajishirzi, H. Unifiedqa: Crossing format boundaries with a single qa system. *arXiv* **2020**, arXiv:2005.00700.
37. Cameron, J. Domain Authority: What Is It and How Is It Calculated. Available online: https://moz.com/learn/seo/domain-authority (accessed on 5 December 2024).
38. Durga, V.S.; Jeyaprakash, T. An effective data normalization strategy for academic datasets using log values. In Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 17–19 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 610–612.
39. Jansen, B.J.; Jung, S.G.; Salminen, J. Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *PLoS ONE* **2022**, *17*, e0268212. [CrossRef]
40. Giomelakis, D.; Veglis, A. Employing search engine optimization techniques in online news. *Stud. Media Commun.* **2015**, *3*, 22–33. [CrossRef]

41. Church, E.M.; Zhao, X.; Iyer, L. Media-generating activities and follower growth within social networks. *J. Comput. Inf. Syst.* **2021**, *61*, 551–560. [CrossRef]
42. Rozovsky, L.V. Comparison of arithmetic, geometric, and harmonic means. *Math. Notes* **2021**, *110*, 118–125. [CrossRef]
43. Adeyemo, A.; Wimmer, H.; Powell, L. Effects of normalization techniques on logistic regression in data science. In Proceedings of the Conference on Information Systems Applied Research, Norfolk, VA, USA, 31 October–3 November 2018; Volume 2167, p. 1508.
44. Kassambara, A.; Mundt, F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. 2020. R Package Version 1.0.7. Available online: https://cran.r-project.org/web/packages/factoextra/readme/README.html (accessed on 5 December 2024).
45. Maechler, M.; Rousseeuw, P.; Struyf, A.; Hubert, M.; Hornik, K. Cluster: Cluster Analysis Basics and Extensions. 2022. R Package Version 2.1.4. Available online: https://cran.r-project.org/web/packages/cluster/index.html (accessed on 5 December 2024).
46. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
47. Albuquerque, P.H.; Valle, D.R.D.; Li, D. Bayesian LDA for mixed-membership clustering analysis: The Rlda package. *Knowl.-Based Syst.* **2018**, *163*, 988–995. [CrossRef]
48. Nikita, M. ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. R Package Version 1.0.2. 2020. Available online: https://cran.r-project.org/web/packages/ldatuning/index.html (accessed on 5 December 2024).
49. Cao, J.; Xia, T.; Li, J.; Zhang, Y.; Tang, S. A density-based method for adaptive LDA model selection. *Neurocomputing* **2009**, *72*, 1775–1781. [CrossRef]
50. Deveaud, R.; Sanjuan, E.; Bellot, P. Accurate and effective latent concept modeling for ad hoc information retrieval. *Doc. Numérique* **2014**, *17*, 61–84. [CrossRef]
51. Arun, R.; Suresh, V.; Veni Madhavan, C.E.; Narasimha Murthy, M.N. On finding the natural number of topics with latent dirichlet allocation: Some observations. In Proceedings of the Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference, PAKDD, Hyderabad, India, 21–24 June 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 391–402.
52. Griffiths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (Suppl. S1), 5228–5235. [CrossRef]
53. Chang, J. lda: Collapsed Gibbs Sampling Methods for Topic Models. R Package Version 1.4.2. 2015. Available online: https://cran.r-project.org/web/packages/lda/index.html (accessed on 5 December 2024).
54. Sievert, C.; Shirley, K. LDAvis: Interactive Visualization of Topic Models. R Package Version 0.3.2. 2015. Available online: https://cran.r-project.org/web/packages/LDAvis/index.html (accessed on 5 December 2024).
55. Coleman, L.J.; Jain, A.; Bahnan, N.; Chene, D. Marketing the performing arts: Efficacy of web 2.0 social networks. *J. Mark. Dev. Compet.* **2019**, *13*, 23–28.
56. Berridge, G. Event Experiences: Design, Management and Impact. Doctoral Dissertation, University of West London, London, UK, 2015.
57. Turrini, A.; Soscia, I.; Maulini, A. Web communication can help theaters attract and keep younger audiences. *Int. J. Cult. Policy* **2012**, *18*, 474–485. [CrossRef]
58. Browning, C. Connecting Artists to Fans: Concert Ticketing in the Digital Age. 2019. Available online: https://soar.suny.edu/handle/20.500.12648/14108 (accessed on 5 December 2024).
59. Carmichael, B.A. Global competitiveness and special events in cultural tourism: The example of the Barnes Exhibit at the Art Gallery of Ontario, Toronto. *Can. Geogr./Le Géographe Can.* **2002**, *46*, 310–324. [CrossRef]
60. Quinn, B. Arts festivals and the city. In *Culture-Led Urban Regeneration*; Routledge: London, UK, 2020; pp. 85–101.
61. Duxbury, N. Cultural and creative work in rural and remote areas: An emerging international conversation. *Int. J. Cult. Policy* **2021**, *27*, 753–767. [CrossRef]
62. Dearn, L.K.; Price, S.M. Sharing music: Social and communal aspects of concert-going. *Netw. Knowl. J. MeCCSA Postgrad. Netw.* **2016**, *9*. [CrossRef]
63. Nicolaou, C.; Matsiola, M.; Dimoulas, C.A.; Kalliris, G. Discovering the Radio and Music Preferences of Generation Z: An Empirical Greek Case from and through the Internet. *J. Media* **2024**, *5*, 814–845. [CrossRef]