ORIGINAL PAPER

A design concept of big data analytics model for managers in hospitality industries

Seyedmohammad Mousavian[1] · Shah J. Miah[1] · Yifan Zhong[2]

Abstract

The hospitality and tourism sector has long played a significant role in Australia's economy, especially in regional areas. Due to the onslaught of COVID-19, numerous businesses have experienced lockdowns, restrictions, and closures due to the fact that people's activity in restaurants, shopping centers, and recreational destinations was restricted, and many other places went into hibernation. After about 2 years since the outbreak, businesses in this sector are gradually starting to reopen and revitalize themselves, but in order to have better decision support about the future of this sector, thus being able to plan, businesses are suffering from an effective analytics solution due to the lack of broken data trends. Starting from fresh day-to-day real-time big data, the study aims to develop a new data analytics model, adopting the design science research methodology, which can provide invaluable options and techniques to make prediction easier from immediate past datasets. This study introduces an innovative design artifact as a big data solution for hospitality managers to utilize analytics for predictive strategic decision-making in post-COVID situation. The artifact can also be generalized for other sectors with tailoring aspects which are subject to further studies. The proposed artifact is then compared with other design artifacts related to big data solutions where it outperforms them in terms of comprehensiveness. The proposed artifact also shows promises for primarily available UGC in managers' decision support aids.

Keywords Big data · Data analytics · Decision support · Hospitality · Tourism · Machine learning

1 Introduction

The Australian and New Zealand Standard Industrial Classification (ANZSIC) categorizes hospitality businesses encompassing different trades such as accommodation, food, and beverages, which are served by cafes, restaurants, and takeaway food services; pubs, taverns, and bars; hotels, motels, and other accommodation; and hospitality clubs [6]. This sector plays a significant role in Australia's economy, particularly in regional and rural development [27]. These types of trading processes attract and enable people to travel and make growth in various occupations such as cafe and restaurant managers, retail managers, bar attendants, and receptionists. According to The Australian Tourism Industry [49], the total spending by domestic tourists and international visitors was $152 billion in 2019. However, the spread of COVID-19 in early 2020 impacted the hospitality and tourism sector seriously in Australia. According to the Australian Bureau of Statistics [5], largest falls in gross value added (GVA) among industries were seen in tourism- and hospitality-related industries, reflecting the restrictions imposed on movement. In addition, COVID-19 impacted regional employment seriously. For example, in food and beverage services, employment dropped from 818,900 in 2019 to 575,400 in 2020 [6]. The International Air Transport Association does not expect international travel to return to previous levels until 2024.

As comprehended, due to the nature of the hospitality sector, this industry has been at the forefront of the toughest restrictions and regulations during the outbreak. However, after more than 2 years since the outbreak, borders have

✉ Shah J. Miah
  shah.miah@newcastle.edu.au

  Seyedmohammad Mousavian
  seyed.mousavian@newcastle.edu.au

  Yifan Zhong
  yifan.zhong@curtin.edu.au

1 Newcastle Business School, University of Newcastle, Callaghan, NSW, Australia

2 School of Management and Marketing, Curtin University, Perth, WA, Australia

reopened, and restrictions were eased so that businesses started returning to their normal processes. While many studies have identified the great potential of big data analytics applications, such as Miah et al. [30] (impact of local tourism on business growth), Raun et al. [41] (impact of destination attraction), and Hu et al. [18] (people movement analysis), studies are yet to focus on immediate past data for generating quick indications for managers' decision support.

In this paper, we introduce an innovative concept of data analytics solution to assist hospitality managers with a certain hope of returning to normal and making up previous losses or relevant decision-making. We aim to achieve valuable insights by implementing "big data analytics," "machine learning algorithms," and other predictive models for anticipating what the hospitality sector is going to learn and observe in their regions. According to Lyu et al. [26], big data in tourism research can be divided into user-generated content (UGC) (72%), operations (web searching) (17%), and the Internet of Things (IoT) (10%). Additionally, UGC data which plays a key role in tourism and hospitality research has some benefits such as data ubiquity, speed, and simplicity of data extraction without necessarily paying cost, but relevant data transformation approaches are still limited. Design science research (DSR) offers options for problem-solving research with an interest in addressing information solution issues in organizations. The DSR helps to create computational artifact design to solve problems [16] by generating new models that may contribute to improving decision support practices or revitalizing design processes that are identified for designing and evaluating an artifact as a value-adding source of new understanding to the information systems (IS) research field. Since our research aims to develop a new data solution by analyzing UGC data, the DSR approach is found to be suitable to answer research questions. However, some methodological and ethical challenges might be faced considering the acquisition of a huge size of the dataset. Hence, our main research question relates to investigating the design activities of an innovative big data analytics solution for hospitality managers, as follows:

> RQ: How can we design a big data analytics solution for the hospitality industry?

In order to reach a preliminary response to our question, we use real hospitality data and try to implement a big data solution, adopting a supervised machine learning (ML) technique and some predictive models to see how impactful they all together would be. This paper can be seen as an important attempt to design a data analytics model for post-COVID time data transformation, which could assist the hospitality industry's growth in the future. To the best of our literature review, although data analytics using ML have been implemented in the hospitality sector previously, limited evidence is found regarding existing approaches that may offer a data-driven solution to deal with the post-pandemic situation (e.g., for the regrowing of businesses).

The rest of the paper is written as follows: Sect. 2 discusses the relevant literature on the application of big data analytics in the hospitality and tourism industries. Section 3 gives details of the design science research methodology, and subsequently, we specify details of the data solution framework, illustrating how to implement a related data-driven prototype. We then address the significance of visualization by providing relevant snapshots based on our sample dataset analysis. The last section discusses the contribution of the study, followed by the future perspective.

2 Related works

In this section, the paper includes a short literature review to understand and interpret existing works and data-driven methodologies in the field of tourism and hospitality.

2.1 Big data analytics research

Big data has opened new avenues for conducting data analytics research to improve the supporting mechanism of decision-making [40]. Large datasets are becoming increasingly available as technology advances in operating businesses. Various sources of data such as CCTV cameras, uploaded photos and videos on social media, digital texts, and UGC may provide input or insights that are to be significant elements for analysis. Xiang et al. [52] suggested that because of the nature of big data, it provides large details of unstructured information about experiences, sentiments, interests, and ideas. In general, the strategic value of big data in the field of tourism has been addressed by several scholars [20]; [25], [28] in areas relevant to destination preference and management. Although research is at an early stage (e.g., mainly for information systems design), the tourism industry, as a data-intensive industry, has the potential to use these possibilities, particularly for destination managers in the context of tourism planning and operations [18].

Previous literature has introduced various analytics methods using ML techniques. For instance, Miah et al. [30] used geo-tagged photos shared by tourists on social media to analyze and predict tourist behavioral patterns in Australia. They used several commonly accepted big data analytics, such as *textual metadata analytics*, to capture tourists' priorities over destinations. Additionally, they applied *clustering techniques* to the geographical data to identify interesting locations and the flow of people in those places. Chang et al. [9] applied *deep learning* solutions for sentiment analysis in the airline industry to demonstrate the impact of COVID-19 on the feelings of passengers in several aspects. The work

by Chang et al. used Tripadvisor to extract flight information, and they used average ratings for services such as customer service, cleanliness, food and beverage, and value for money. For data analysis, they utilized aspect-based sentiment analysis, deep learning–based natural language processing (NLP), and visualization. These studies are limited to focus on people activity data insights to assist hospitality managers and hosts to plan for their future business growth.

Biswas et al. [8] performed Poisson, quasi-Poisson, and negative binomial regressions with Airbnb dataset from ten cities around the world to find the determinants of the number of customer reviews received by these shared homes. They used clustering techniques to make clusters of cities, and for each cluster, they applied regressions separately. Biswas et al. [8] concluded that hosts could receive more reviews by improving six areas, namely, "website, host, property, historical reviews, rental policies, and availability" (p. 5). Giglio et al. [15] applied a supervised ML approach to capture images posted on Flickr to extract tourists' behavior and find out which areas are more attractive to them in Italy. Martinez-Torres and Toral [29] utilized an ML approach to distinguish between positive, negative, deceptive, and non-deceptive online reviews in the hospitality sector. They performed classification methods such as SVM, *K*-nearest neighbor, and gradient boosting random forest. Findings suggest that their model offers options to separate between deceptive and non-deceptive reviews by their polarity orientation. These studies show promises in conducting a new experimental design study in the hospitality industry.

Analytical approaches can provide managers with insights into which model(s) might be suitable to apply in their specific context. Whether they seek to predict a continuous value or discrete value, managers can resort to various machine learning approaches and visualizations.

Table 1 shows the top 10 relevant studies of analytics design in hospitality management that utilize various ML approaches.

2.2 Critical analysis

Tourism literature has several attempts to capture big data analytics as a tool for prediction. For instance, Miah et al. [30] designed an artifact for analyzing geo-tagged photos of users shared on Flickr social media. Their approach follows "big data analytics," which intends to assist hospitality managers in strategic decision-making. Their artifact contains four stages: textual meta-data processing, geographical data clustering, representative photo identification, and time-series modeling. They adopted the principles of design science research, and they suggested an IT artifact that is able to capture tourists' destination patterns and behaviors. Raun et al. [41] developed a methodology to capture visitor flows using mobile positioning data (big data). They measured

three out of five measurable dimensions of a tourist's destination as temporal, geographical, and compositional and concluded which destinations are more popular among various nationalities. Similar to this, Giglio et al. [15] introduced a big data model to identify the tourism's location of interest. They applied an artificial neural network model to train millions of images shared on Flickr platform and then used clustering techniques to homogenous group observations. Most literature in this area suggested possible techniques, theories, and approaches,however, the lack of suggesting a design artifact is obvious. One positive point about having an artifact is that although some sort of abstraction is used in its design, it can be implemented and thus be created. On the other hand, theory goes beyond the existed artifact and entails additional knowledge [16]. The theory that is considered in design science research model is called *design theory* which contains more prescriptive knowledge rather than descriptive knowledge [16], which is more useful for our prediction purposes as the future embraces the use of predictive and prescriptive analysis for big data analytics.

Geerdink [14] designed a specific kind of reference architecture developed by Angelov et al. [4] for predictive analytics using big data and open data sources. The big data solution reference architecture model contains three ArchiMate layers as business layer, application layer, and technology layer. Application and technology layers mostly deal with the required computation systems, storage, and technical infrastructures which are beyond our research purpose; however, business layer that intends to use predictive big data analytics is our research interest. Despite analyzing a good reference architecture, the main focus of Geerdink [14] is to suggest a reference architecture based on big data for predictive analytics, not necessarily suggesting big data analytics models or algorithms. Our study is contributing to the literature in terms of suggesting an artifact, possible big data models, and some real applications in the hospitality sector.

Previous literature extensively discussed big data analytics and its related artifact solutions separately. For example, L'Heureux et al. [22] discussed machine learning approaches and potential issues for big data analytics. On the other hand, [45] analyzed the fundamental role that big data can play in bringing insightful information to managers regarding their customer's demands. They designed an analytical framework to assist managers in achieving their goals and, at the same time, meet their customer's needs. The paper develops big data tourism analytical (BDTA) framework but without any empirical testing of the model.

Our study integrates big data analytics with a novel artifact managerial solution that, to the best of our knowledge, has not previously been discussed. Another aspect of this study that makes it unique is the implementation of design science research. Design science is critical for building a successful artifact in a special field.

Table 1 Top 10 relevant studies in the hospitality management

The main point of the studies	Used analytical methods	Research methodologies	Problem gist	Authors
Big data analytics method for tourist behavior analysis	Image analytics	Hevner et al.'s (2004) seven guidelines	Addressing issues of destination management for their strategic decision-making	Miah et al. [30]
Predicting aspect-based sentiment using deep learning in the airline industry	Deep learning–based NLP, visualization	Used un-specified design research methodologies	Examining flight reviews to evaluate the impact of COVID-19 on passenger travel sentiment in several aspects	Chang et al. [9]
Examining the determinants of the count of customer reviews using clustering	Used ML for analyzing customer review	Used un-specified design research methodologies	Analyzing customer reviews received	Biswas et al. [8]
Using social media to identify tourism attractiveness	Used supervised ML approach plus clustering techniques	Used un-specified design research methodologies	Discovering which places are attractive for Italian tourists	Giglio et al. [15]
Social media analytics in hospitality and tourism	Used ML and text analytics	Designed a framework for information quality. A review has four features (linguistic, semantic, sentiment, and its source)	Examining platforms in terms of information quality related to online reviews about the entire hotel population in Manhattan	Xiang et al. [51]
A large-scale granular data investigation	Used text mining and econometrics model	Customer repeated booking data using a Python program were collected		Wu et al. [50]
Ranking of high-value social audiences on Twitter	Used LDA and support vector machine	Used un-specified design research methodologies	Using mechanism that is capable of identifying the top-k social audience members on Twitter	Lo et al. [24]
A machine learning approach for the identification of deceptive reviews in the hospitality sector	Used SVM, K-nearest neighbor, and gradient boosting random forest	Used un-specified design research methodologies	Focusing on the hospitality sector and follows a content analysis approach based on a set of unique attributes and the sentiment orientation of reviews	Martinez-Torres and Toral [29]
Measuring tourism destinations using mobile tracking data	Used mobile positioning and tracking	Used un-specified design research methodologies	Developing a database based on roaming data of the foreign mobile phone call in Estonia for destination management	Raun et al. [41]
Olympics on Twitter perspectives of hosts and guests	Used text mining approach	Sentiment analysis is performed to extract emotional expressions by classifying them as positive, negative, or neutral	Analyzing differences in pre- and post-Games sentiment toward issues surrounding the Olympics	Kirilenko and Stepchenkova [19]

3 Design science research methodology

Design science research has been a popular research methodology for IS researchers due to the fact that it enables the creation of new knowledge from the design and evaluation of new artifacts. In this research, we adopt a significant class of design science called the computational design science paradigm which is rapidly growing for its supportive guidelines to develop a novel ML-based data analytics framework. For the hospitality industry, the proposed innovation is to meet the standard of supporting managerial insights that are needed to better enable them for future planning. The computational design science research will be adopted in this research (extending the framework of DDSR given by). According to the computational DSR provides IS scholars with three guidelines to design an innovative artifact, which can be an algorithm, computational model, and prototype solutions for advancing the current data solution applications. The guidelines we applied in our design research are given in Table 2.

Previous literature used DSR for designing and evaluating solution artifacts for stakeholders such as Singh and Miah [48], who argue that big data management and data analytics are the critical aspects of a design artifact. Moreover, Miah et al. [32, 38] utilized DSR to design and develop their People Tracker artifact, which is based on other design studies [32, 34, 38], Miah et al. [33], Miah et al. [35]). They believe that DSR is not only a methodology for designing an artifact but also can be used for researchers to learn from artifact solutions. Also, compared with the design science methods by Peffers et al. [39] that provide various design steps such as identifying the problem, defining the solution objectives, designing and developing, demonstrating, evaluating, and communicating, it is important to outline a specific methodological approach that may better guide our design research.

3.1 Big data in hospitality

According to De Mauro et al. [10], big data is an information asset characterized by its high volume, velocity, and variety, which requires specific technologies and analytical methods to extract value from it. Due to the increased availability of the Internet around the world, social media activity has increased, and a large number of user-generated content (UGC) such as posts, images, videos, or reviews are being posted by people every day [47]. According to Lyu et al. [26], about 72% of the big data in tourism and hospitality research is derived from the UGC. UGC is defined as creative work that is published on publicly accessible websites and is created without a direct link to monetary profit or commercial interest. It can be as forms of consumer reviews (e.g., Yelp), personal blogs, microblogging platforms (e.g.,

Table 2 Three design science guidelines adopted in this study

Guidelines	Descriptions	In this research
Design inspiration: design meeting problem domain's rules	The artifact design should be inspired by key domain characteristics. For instance, developed a healthcare artifact in which key contextual cues from the electronic health records guided a novel Bayesian multitask learning approach for the prediction of relevant events	The artifact design in this research addresses the characteristics of the hospitality and tourism sector for data acquisition. For example, data capturing on people's activities and interactions led us to design an artifact with its appropriate contextual elements
Novelty demonstration: defining technical superiority	Design researchers must demonstrate the innovative nature of the artifact with its technical superiority by comparing baseline approaches via quantitative metrics (e.g., accuracy, precision, recall, F1, and others)	The design of the artifact has not been previously introduced. It is novel because we combined existing ideas to apply them in solving data solutions in the hospitality industry
Key contributions of implementations: leading new IS knowledge	Artifact's design should contribute situated implementations or design theory (e.g., design principles) and/or well-developed design theory to the IS knowledge base [16]	The artifact's design requires basic knowledge of machine learning and visualization to be implemented. With the base knowledge, it can bring value to the hospitality industries (cafes, restaurants, and pubs) and tourism industries (hotels and motels)

Tumblr), social networks (e.g., Facebook), and media-sharing tools (e.g., YouTube) [37]. As a "digital footprint," UGC has a number of advantages in the research context, including data availability, speed, and simplicity of data collection. However, a large amount of data imposes problems in collecting, organizing, and analyzing the bulk of this material. In addition, extracting the identity of UGC creators and their locations creates methodological and ethical challenges. With all limitations, UGC is still popular among researchers to extract people's mobility patterns, including visualization of digital footprints, sentiment analysis of consumer reviews, and tourist experiences [25].

Various methods, analytical techniques, and machine learning algorithms can be applied to UGC big data and other sorts of datasets. Some top analytical approaches for the UGC analysis that may have potential for our artifact design innovations are sentiment analysis, clustering, classification, regression, ensemble methods, density-based models, ANN, and content analysis.

4 Design and development
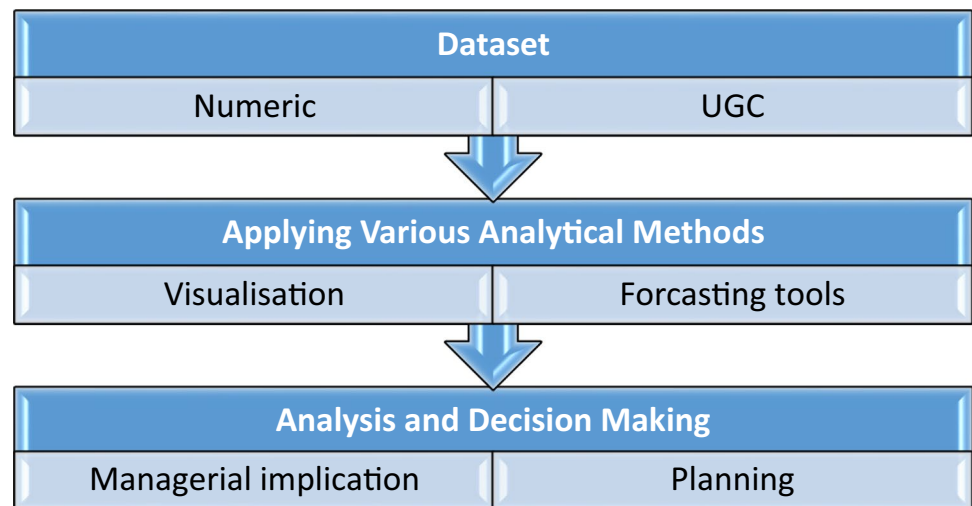
4.1 Big data analytics method

Applying forecasting models for transforming big data is new to tourism research, e.g., for capturing tourism demand by using various analytics techniques [17]. For example, some website traffic data, such as Google Analytics website traffic indicators, can be utilized to develop forecasts which improve managerial decisions. In this study, we propose a problem–solution framework based on data analytics to address the most current issues of exploring future directions for business growth. As a core part of the proposed data analytics solutions, the visualization, in our viewpoint, can significantly assist hospitality managers in planning

their customer's growth and retention for the future. We take either numeric datasets or UGC as input leading to generating managerial implications and planning solutions as output. Various tools can be used as forecasting tools, such as vector autoregression models [17], time series models [3], and seasonal autoregressive integrated moving average (SARIMA), which are often used in tourism forecasting [7]. Figure 1 shows the methodology with the sequence of activities in the proposed solution.

Figure 1 shows how a specific dataset can be utilized to generate insights. Both types of datasets in our model might be used for generating different analytical methods. For instance, one set of UGC retrieved from a social platform can be interpreted via sentiment analysis. The last part is the insight that is the result of our analytical implementation, which can be used for planning and managerial decision-making. Additionally, Fig. 2 is the big data analytics framework as an extended illustration of how our proposed analytics solution would operate.

As shown in Fig. 2, two main sections, big data and data analytics, have been integrated and make big data analytics as a comprehensive artifact. "Big data" section generally deals with data management which entails processes and required technologies to acquire, store, prepare, and retrieve data for further analysis. "Data analytics" section refers to the techniques and models applied to big data to gain intelligence [13]. In addition, data pre-processing is applied once we extract our data from external sources,this is to ensure that the dataset is consistent and is ready for analytics. The next step contains data sampling and feature extraction, which are necessary when we are dealing with high-dimensional big data and when we have imbalance class(es) [43]. Moreover, according to Seiffert et al [44] imbalance class issues regarding big data may make the machine learning model biased toward the dominant class(es). Data sampling will alter datasets in such a way as to treat this issue. To

Fig. 1 Big data methodology

Fig. 2 Big data analytics framework
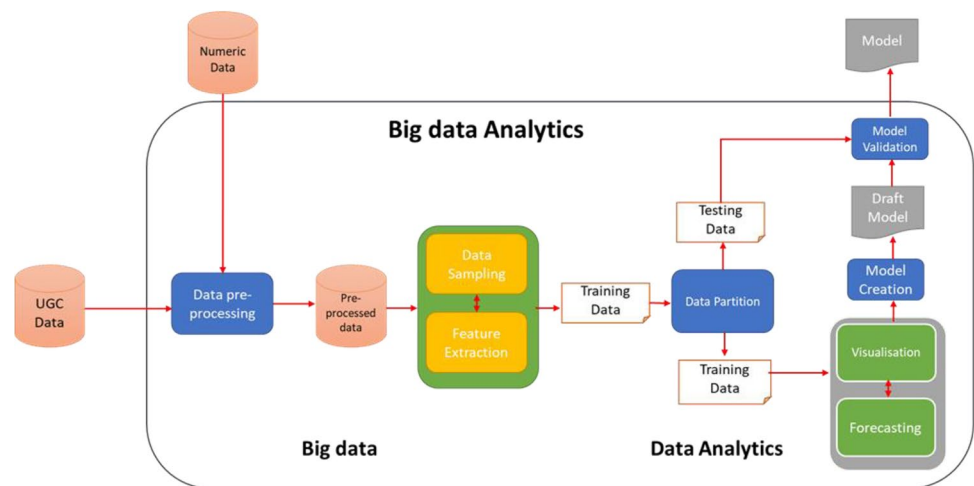
execute a supervised machine learning (e.g., regression, classification), for example, splitting data into training and testing is necessary for building an ML model. After training the machine, it is ready to be used for prediction, and also, its accuracy can be measured by using a confusion matrix.

The proposed solution architecture in Fig. 2 illustrates how the input (numeric data, UGC) is going to convert into a model for managerial decision-making in the hospitality and tourism industry. In the framework's first step, the big data (e.g., customers' flight satisfaction retrieved from a social platform) is collected. In general, data can be structured like tabular reports and spreadsheets or unstructured like photos, texts, or videos [13]. Then, it is sent for pre-processing (e.g., data cleansing, data integration, stop word removal, or another anomaly removal). Data pre-processing is a vital step that usually contains identifying possible anomalies and cleaning them. Moreover, data cleansing deals with detecting and correcting errors in datasets [42]. Subsequently, it is ready for the next step, which is feature extraction and sampling processes. As mentioned earlier, feature extraction and data sampling are mainly used for treating high-dimensional and imbalanced class datasets, which is very common in large datasets. Following that, the sample data will be sent for separating into training and testing datasets (i.e., for supervised machine learning such as classification or regression, we need to split the data into training and testing). Based on what a decision maker seeks as an outcome (discrete or continuous value), predictive techniques can be categorized as regression (continuous value) or classification (discrete value) [13]. Data analytics models such as visualization and forecasting can be applied to generate a draft model. After evaluating its validation, it is going to be converted into a solid model that can be used for prediction, planning, and decision-making.

For an industrial application, for example, when customers' flight satisfaction is retrieved from a social platform, it probably requires some pre-processing as they usually come from various heterogenous sources in unstructured formats with inconsistencies and noises [1]. In order to increase the quality of a dataset and thus a predictive model, data pre-processing is necessary. Different data sampling and feature extraction methods can be applied after the pre-processing step. Subsequently, data is ready for applying one or multiple analytics models. Choosing supervised machine learning methods (e.g., regression) requires data split into testing and training, however, unsupervised methods (e.g., clustering) do not require training. Visualization can also be a useful tool for dealing with extracting intelligence from big unstructured data [23]. There are several visualization models such as histograms, heatmaps, bar charts, pie charts, and spider webs. For instance, Talón-Ballestero et al. [46] utilized two visualization models for their big dataset, namely chromosome proportions plot and the spider web representation.

Similar kinds of frameworks have been used by scholars in different research papers. For instance, Dittert et al. [11] depicted a big data analytics framework for small- and medium-sized companies (SMEs), which includes the steps as follows: define a task, collect and analyze the data, choose and set up a model, format data, evaluate results, and report to decision makers. Another model was the cross-industry standard process for data mining (CRISP-DM) which is the most commonly used framework for data mining solutions introduced by Kotu and Deshpande [21] and shows different steps of the data mining process in a business environment.

Although some big data analytics frameworks have been suggested previously to deal with business problems, our model specifically focuses on the hospitality and tourism industry. Due to the nature of this industry and the high volume of UGCs and unstructured data, pre-processing is an important step before entering the data analytics section. Also, visualization was considered in our framework, which makes it richer because, as illustrated in Figs. 3, 4, and 5, visualizations can provide highly valuable managerial insights.
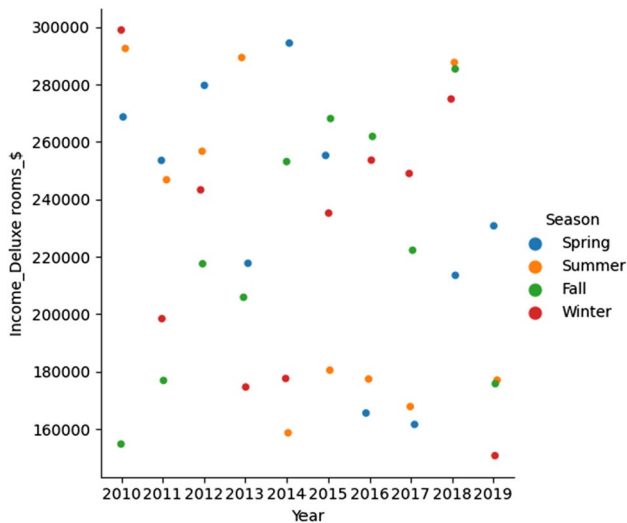
Fig. 3 Income of deluxe rooms in different seasons (2010–2019)

4.2 Visualization for the validity of the proposed framework

Visualization through plotting and scattering using Python language would provide support for transforming helpful data. As a case demonstration, we extracted "customer flight satisfaction" and "hospitality cash flow" from https://www.kaggle.com/. Customer flight satisfaction consists of the details of customers that have already flown with them. The feedback of the customers on various contexts and their flight data has been consolidated. The main purpose of this dataset is to predict whether a future customer would be

satisfied with their service, given the details of the other parameter values (Fig. 5). The hospitality cash flow dataset shows the inflow of income into this sector separated by different services (room types and restaurants) as it is shown in Fig. 3. Figure 3 shows the income generated from deluxe rooms in various seasons.

As it is shown in Fig. 3, seasons are shown in different colors. The *X*-axis is the year, and the *Y*-axis is Income_Deluxe_rooms_$. For example, in 2010 during spring, summer, and winter, the hotel's deluxe rooms made a good income each season of more than $260,000. However, in fall, it was less than $160,000. In spring 2019, deluxe rooms made income just above $220,000, but in summer, winter, and fall, the income was less than $180,000. Figure 4 is very helpful; it shows the advertising/marketing expenses from 2010 to 2019 in various seasons.

Based on the chart (in Fig. 4) in 2010, 2011, and 2012, marketing expenses increased in winter, and 2013 dropped significantly. In 2019, spring, summer, and winter expenses were almost around $80,000; however, in fall, it was around $50,000. Figure 5 shows how males and females were satisfied with their flight experience.

As you can see, between the ages of 40 and 50, more men were dissatisfied compared to women. On the other hand, between the ages of 20 and 30, more women were dissatisfied compared to men. The flight company should pay more attention to men in 40–50 and women in 20 to 30.

In summary, Figs. 3, 4, and 5 are illustrative examples of how the big data analytics framework (Fig. 2) can be followed for generating information for managerial decision-making. For example, the flight customer satisfaction

Fig. 4 Advertising expenses in different seasons (2010–2019)
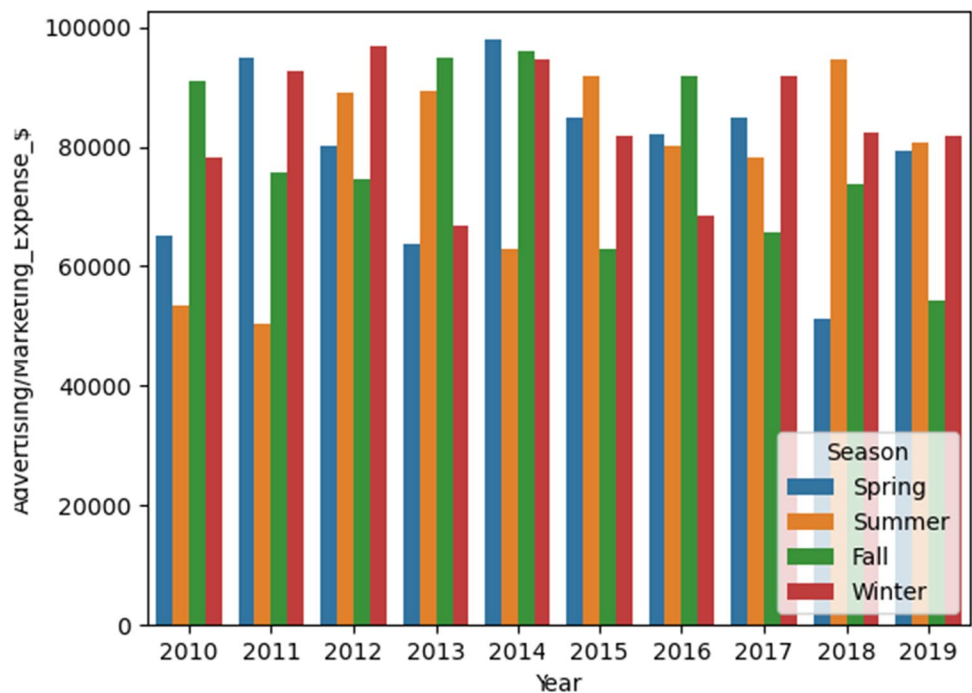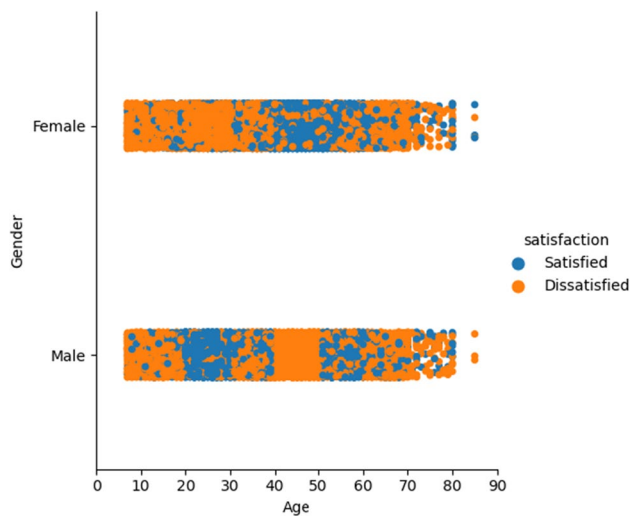
Fig. 5 Flight customer satisfaction

dataset was collected as a sample from the *Kaggle* website to do some visualization. After preparing the dataset such as removing some unwanted columns, it was then tested to make sure which visualization technique can be more informative (e.g., bar chart, pie chart, and plotting). Finally, after testing and comparing the visualized model with our real dataset, then, a brief report will be added to explain the visualized model.

5 Discussion and conclusion

This research intends to bring a new requirement for developing a big data analytics solution. As an initial component, we described insights that are valuable for managers in the hospitality and tourism industries and what can be captured from the UGC big data, but also, it can be applied to numeric data (as shown in the visualization as a proof of concept). Although previous literature has extensively worked on big data analytics in the tourism industry, our attempt was to enhance our design understanding of data analytics solutions for improving business practices in the post-pandemic situation, providing a solid data-driven solution for business revitalization. In this paper, we tried to introduce a novel framework that aims to capture UGC or numeric data applying them to précised analysis for providing managerial implications for future planning. In the future, we intend to develop a meta-artifact so that it would be able to work for other industries apart from hospitality and tourism. According to Miah et al. [36], previous successfully implemented empirical works in applying DSR methods and strategies can be generalized as new knowledge to design more general solution concepts as meta-artifact. However, this might be quite challenging because, as mentioned, designing an

artifact requires familiarity with the characteristics of the domain of interest. So, although designing a more general artifact that can bring analytical value to more industries might be challenging, it is not far from imagination. Motivating into data analytics design research for making more effective decisions and guidance to "navigate our digital future," we enhance the current knowledge of computational DSR for better research support and guidance for future AI research, in our case, leading to designing smart automation for better data-driven strategic decision-making in this business domain. For exploration of current idea in terms of its generalizability, further research may extend big data analytics research in other problem-solving domains (e.g. higher education [2, 12, 31] and healthcare information management [2]).

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Al-Taie MZ, Kadry S, Lucas JP (2019) Online data preprocessing: a case study approach. Int J Electr Comput Eng 9(4):2620
2. Aghdam AR, Watson J, Cliff C, Miah SJ (2020) Improving theoretical understanding towards patientdriven healthcare innovation: online value cocreation perspective: a systematic review. J Med Internet Res 22(4):e16324. https://doi.org/10.2196/16324
3. Andrawis RR, Atiya AF, El-Shishiny H (2011) Combination of long term and short term forecasts, with application to tourism demand forecasting. Int J Forecast 27(3):870–886. https://doi.org/10.1016/j.ijforecast.2010.05.019
4. Angelov S, Grefen P, Greefhorst D (2012) A framework for analysis and design of software reference architectures. Inf Softw Technol 54(4):417–431. https://doi.org/10.1016/j.infsof.2011.11.009
5. Australian Bureau of Statistics (2022) Effects of COVID-19 strains on the Australian economy. [online] Available at: https://www.abs.gov.au/articles/effects-covid-19-strains-australian-economy [Accessed 20 July 2022]
6. Australian Industry and Skills Committee (2022) Hospitality. [online] Available at: https://nationalindustryinsights.aisc.net.au/industries/tourism-travel-and-hospitality/hospitality#:~:text=The%20Australian%20and%20New%20Zealand,and%20other%20accommodation%20and%20hospitality [Accessed 20 July 2022]
7. Bangwayo-Skeete PF, Skeete RW (2015) Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. Tour Manage 46:454–464. https://doi.org/10.1016/j.tourman.2014.07.014

8. Biswas B, Sengupta P, Chatterjee D (2020) Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques. Decis Support Syst 135:113324. https://doi.org/10.1016/j.dss.2020.113324

9. Chang Y-C, Ku C-H, Nguyen D-DL (2022) Predicting aspect-based sentiment using deep learning and information visualization: the impact of COVID-19 on the airline industry. Inf Manag 59(2):103587. https://doi.org/10.1016/j.im.2021.103587

10. De Mauro A, Greco M, Grimaldi M (2016) A formal definition of big data based on its essential features. Libr Rev 65(3):122–135. https://doi.org/10.1108/LR-06-2015-0061

11. Dittert M, Härting R-C, Reichstein C, and Bayer C (2018) A data analytics framework for business in small and medium-sized organizations. Paper presented at the Intelligent Decision Technologies 2017, Cham

12. Fahd K, Miah SJ, Ahmed K (2021) Predicting student performance in a blended learning environment using learning management system interaction data. App Comp Inform. https://doi.org/10.1108/ACI-06-2021-0150

13. Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manage 35(2):137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

14. Geerdink B (2015) A reference architecture for big data solutions - introducing a model to perform predictive analytics using big data technology. Int J Big Data Intell 2(4):236–249. https://doi.org/10.1504/IJBDI.2015.072160

15. Giglio S, Bertacchini F, Bilotta E, Pantano P (2019) Using social media to identify tourism attractiveness in six Italian cities. Tour Manage 72:306–312. https://doi.org/10.1016/j.tourman.2018.12.007

16. Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. MIS Q 37(2):337–355

17. Gunter U, Önder I (2016) Forecasting city arrivals with Google Analytics. Ann Tour Res 61:199–212. https://doi.org/10.1016/j.annals.2016.10.007

18. Hu F, Li Z, Yang C, Jiang Y (2019) A graph-based approach to detecting tourist movement patterns using social media data. Cartogr Geogr Inf Sci 46(4):368–382. https://doi.org/10.1080/15230406.2018.1496036

19. Kirilenko AP, Stepchenkova SO (2017) Sochi 2014 Olympics on Twitter: perspectives of hosts and guests. Tour Manage 63:54–65. https://doi.org/10.1016/j.tourman.2017.06.007

20. Költringer C, Dickinger A (2015) Analyzing destination branding and image from online sources: a web content mining approach. J Bus Res 68(9):1836–1843

21. Kotu V, Deshpande B (2014) Predictive analytics and data mining: concepts and practice with rapidminer: Morgan Kaufmann. Elsevier

22. L'Heureux A, Grolinger K, Elyamany HF, Capretz MAM (2017) Machine learning with big data: challenges and approaches. IEEE Access 5:7776–7797. https://doi.org/10.1109/ACCESS.2017.2696365

23. Liu Y-Y, Tseng F-M, Tseng Y-H (2018) Big data analytics for forecasting tourism destination arrivals with the applied vector autoregression model. Technol Forecast Soc Chang 130:123–134. https://doi.org/10.1016/j.techfore.2018.01.018

24. Lo SL, Chiong R, Cornforth D (2016) Ranking of high-value social audiences on Twitter. Decis Support Syst 85:34–48. https://doi.org/10.1016/j.dss.2016.02.010

25. Lu W, Stepchenkova S (2015) User-generated content as a research mode in tourism and hospitality applications: topics, methods, and software. J Hosp Market Manag 24(2):119–154. https://doi.org/10.1080/19368623.2014.907758

26. Lyu J, Khan A, Bibi S, Chan JH, Qi X (2022) Big data in action: an overview of big data studies in tourism and hospitality literature. J Hosp Tour Manag 51:346–360. https://doi.org/10.1016/j.jhtm.2022.03.014

27. Main J (2022) 6 reasons why tourism is so important to Australia's post COVID-19 economy. [online] Torrens Blog. Available at: https://www.torrens.edu.au/blog/6-reasons-why-tourism-is-so-important-to-australias-post-covid19-economy#:~:text=Perhaps%20most%20obviously%2C%20hospitality%20and,important%20mainstay%20for%20rural%20communities. [Accessed 20 July 2022]

28. Marine-Roig E, Clavé SA (2015) Tourism analytics with massive user-generated content: a case study of Barcelona. J Destin Mark Manag 4(3):162–172

29. Martinez-Torres MR, Toral SL (2019) A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. Tour Manage 75:393–403. https://doi.org/10.1016/j.tourman.2019.06.003

30. Miah SJ, Vu HQ, Gammack J, McGrath M (2017) A big data analytics method for tourist behaviour analysis. Inf Manag 54(6):771–785. https://doi.org/10.1016/j.im.2016.11.011

31. Miah SJ, Miah M, Shen J (2020) Learning management systems and big data technologies for higher education. Educ Inf Technol 25:725–730. https://doi.org/10.1007/s10639-020-10129-z

32. Miah SJ, Gammack J, McKay J (2019) A metadesign theory for tailorable decision support. J Assoc Inf Syst 20(5):570–603

33. Miah SJ (2010) A new semantic knowledge sharing approach for e-government systems, 4th IEEE International Conference on Digital Ecosystems and Technologies, 457–462, Dubai, United Arab Emirates

34. Miah SJ (2008) An ontology based design environment for rural decision support, Unpublished PhD Thesis, Griffith Business School, Griffith University, QLD, Australia

35. Miah SJ, Gammack J (2014) Ensemble artifact design for context sensitive decision support. Australas J Inf Syst 18(2):5–20

36. Miah SJ, McGrath M, and Kerr D (2016) Design science research for decision support systems development: recent publication trends in the premier IS journals. Australas J InfSyst 20. https://doi.org/10.3127/ajis.v20i0.1482

37. Miah SJ, Vu HQ, and Gammack J (2018) A location analytics method for the utilization of geo-tagged photos in travel marketing decision-making.J Inf Knowl Manag 18(1) https://doi.org/10.1142/S0219649219500047

38. Miah SJ, Vu H, Gammack J (2019) A big-data analytics method for capturing visitor activities and flows: the case of an island country. Inf Technol Manage 20(4):203–221. https://doi.org/10.1007/s10799-019-00303-2

39. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. J Manag Inf Syst 24(3):45–77. https://doi.org/10.2753/MIS0742-1222240302

40. Power DJ (2015) 'Big data'decision making use cases. Paper presented at the International Conference on Decision Support System Technology, p 1–9. https://link.springer.com/chapter/10.1007/978-3-319-18533-0_1. Accessed 25 June 2022

41. Raun J, Ahas R, Tiru M (2016) Measuring tourism destinations using mobile tracking data. Tour Manage 57:202–212. https://doi.org/10.1016/j.tourman.2016.06.006

42. Ridzuan F, Wan Zainon WMN (2019) A review on data cleansing methods for big data. Proc Comput Sci 161:731–738. https://doi.org/10.1016/j.procs.2019.11.177

43. Salekshahrezaee Z, Leevy JL, Khoshgoftaar TM (2021) Feature extraction for class imbalance using a convolutional autoencoder and data sampling. Paper presented at the 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI). https://doi.org/10.1109/ICTAI52525.2021.00037. Accessed 23 June 2022

44. Seiffert C, Khoshgoftaar TM, Hulse JV, and Napolitano A (2008) A comparative study of data sampling and cost sensitive learning. Paper presented at the 2008 IEEE International Conference on Data Mining Workshops, 15–19 December 2008, Pisa, Italy, https://doi.org/10.1109/ICDMW.2008.119

45. Stylos N, Zwiegelaar J (2019) Big data as a game changer: how does it shape business intelligence within a tourism and hospitality industry context? In: Sigala M, Rahimi R, Thelwall M (eds) Big data and innovation in tourism, travel, and hospitality: managerial approaches, techniques, and applications. Springer, Singapore, pp 163–181

46. Talón-Ballestero P, González-Serrano L, Soguero-Ruiz C, Muñoz-Romero S, Rojo-Álvarez JL (2018) Using big data from customer relationship management information systems to determine the client profile in the hotel sector. Tour Manag 68:187–197. https://doi.org/10.1016/j.tourman.2018.03.017

47. Shahbazi Z and Byun Y-C (2020) Analyzing the performance of user generated contents in B2B firms based on big data and machine learning. 86:30–39

48. Singh H, Miah SJ (2019) Design of a mobile-based learning management system for incorporating employment demands: case context of an Australian university. Educ Inf Technol 24(2):995–1014. https://doi.org/10.1007/s10639-018-9816-1

49. The Australian Tourism Industry (2022) Tourism recovery scenarios. [PDF] Available at: https://www.tra.gov.au/ArticleDocuments/185/Tourism-Recovery-Scenarios.pdf.aspx. Accessed 20 July 2022

50. Wu J, Cai J, Luo X, Benitez J (2021) How to increase customer repeated bookings in the short-term room rental market? A large-scale granular data investigation. Decis Support Syst 143:113495. https://doi.org/10.1016/j.dss.2021.113495

51. Xiang Z, Du Q, Ma Y, Fan W (2017) A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tour Manage 58:51–65. https://doi.org/10.1016/j.tourman.2016.10.001

52. Xiang Z, Schwartz Z, Gerdes JH, Uysal M (2015) What can big data and text analytics tell us about hotel guest experience and satisfaction? Int J Hosp Manag 44:120–130. https://doi.org/10.1016/j.ijhm.2014.10.013