A Novel Data Analytics Methodology for Discovering Behavioral Risk Profiles: The Case of Diners During a Pandemic

Thouraya Gherissi Labben [ID] and Gurdal Ertek *[ID]

College of Business and Economics, United Arab Emirates University,
Al Ain P.O. Box 15551, United Arab Emirates; thouraya.labben@uaeu.ac.ae
* Correspondence: gurdal@uaeu.ac.ae

Abstract: Understanding tourist profiles and behaviors during health pandemics is key to better preparedness for unforeseen future outbreaks, particularly for tourism and hospitality businesses. This study develops and applies a novel data analytics methodology to gain insights into the health risk reduction behavior of restaurant diners/patrons during their dining out experiences in a pandemic. The methodology builds on data relating to four constructs (question categories) and measurements (questions and attributes), with the constructs being worry, health risk prevention behavior, health risk reduction behavior, and demographic characteristics. As a unique contribution, the methodology generates a behavioral typology by identifying risk profiles, which are expressed as one- and two-level decision rules. For example, the results highlighted the significance of restaurants' adherence to cautionary measures and diners' perception of seclusion. These and other factors enable a multifaceted analysis, typology, and understanding of diners' risk profiles, offering valuable guidance for developing managerial strategies and skill development programs to promote safer dining experiences during pandemics. Besides yielding novel types of insights through rules, another practical contribution of the research is the development of a public web-based analytics dashboard for interactive insight discovery and decision support.

Keywords: data analytics; machine learning; risk profiling; behavioral typology; worry; risk preventive behavior; risk reduction behavior; restaurant; pandemic; healthcare

1. Introduction

Tourism is one of the most vulnerable economic sectors. It is directly affected by different types of risks, including political instability, natural disasters, urban risks, terrorism threats, and health emergencies [1–3]. In the past, the sector was impacted by many health outbreaks, e.g., Ebola, SARS, and Zika. However, the scope, duration, and consequences of the COVID-19 Coronavirus pandemic have been unprecedented. Consequently, research on the impact of the pandemic on the tourism business and tourists has saliently expanded, while only a limited number of studies have examined topics related to health risks and tourists' behaviors [4]. The continuous advancement of knowledge about tourist behavior in risky situations is of paramount importance because experts foresee multiple and repetitive health outbreaks to occur [5]. Ref. [6] called for more research in this area to help tourism sector professionals understand tourism stakeholders' behaviors and profiles to proactively adjust and prepare for future transformations.

In this context, data analytics methods could provide fresh perspectives and insights into understanding tourist behaviors in risky situations. Ref. [7] called for a more extensive use of data analytics methods in tourism. Ref. [8] expanded on the key role of the latter in tourist demand analysis and tourism emergency studies. Accordingly, the current study contributes to the application of data analytics in pandemic management.

The objective of this study is to better understand the typology of restaurant diner/patron behavior during pandemics. To this end, it was necessary to custom-develop a

formal data analytics methodology for behavioral risk profiling, as no such methodology was found in the literature. The methodology is constructed based on and illustrated through a case study where survey data were collected in the United Arab Emirates (UAE). The present study extends an earlier study [9] from the same research stream, which focused on overall patterns and significant factors characterizing restaurant diners/patrons during the pandemic. In contrast to [9], the present research focuses on the systematic methodological generation of an exhaustive list of notable behavioral risk profiles. These profiles are expressed in the form of one- and two-level "decision rules", where the rules are generated algorithmically using machine learning techniques and filtered and refined through visual analysis.

The two primary theoretical contributions of the paper in terms of analysis methodology are (1) the development and application of a novel integrated methodology and (2) the systematic profiling of restaurant customers during a pandemic using machine learning techniques and visualization. The latter were not encountered in the earlier literature. From a practical perspective, the first contribution of the paper is the novel types of insights obtained through the application of the developed methodology as well as the development of a public web-based analytics dashboard [10] for decision support. Furthermore, the survey data used in the study—except the demographic attributes, which contain sensitive data—have been posted online as public data in the Supplement document [11]. In terms of social sciences research, this paper incorporates the protection motivation theory (PMT) [12], rooted in the health belief model (HBM) [13], while the majority of tourism research is not based on solid theoretical frameworks.

The remainder of this paper is organized as follows. Section 2 offers an overview of data analytics methodologies in tourism and defines the different constructs used in this research. The following Section 3 is dedicated to a detailed description of the methodology adopted to employ data analytics techniques, including one-level and two-level rule discovery. Section 4 presents the main results based on the selected data analytics methods to understand the profiles and the behaviors of restaurant diners/patrons during the pandemic. Section 5 discusses the obtained results and threats to the validity of the study. Section 6 provides a conclusive summary and remarks.

2. Literature

In this section, first the field of data analytics is introduced. Next, a review and discussion of the literature on the application of data analytics in tourism and hospitality is provided. Then, the literature on the factors for tourist profiling in a health risk context is reviewed.

2.1. Data Analytics

Data analytics is the application of various methods of statistics, business intelligence, data visualization, and machine learning (ML) to analyze a particular dataset [14,15]. According to the *"Analytic Ascendancy Model"* by Gartner [16–18], there are four levels of data analytics maturity, namely descriptive, diagnostic, predictive, and prescriptive [18,19]. The primary objectives of the analytics techniques corresponding to these four maturity levels are as follows:

- Gaining insights into data by discovering hidden patterns (*descriptive*);
- Diagnosing root causes (*diagnostic*);
- Predicting future outcomes (*predictive*);
- Prescribing optimal decisions (*prescriptive*).

The primary questions answered at each level of maturity are "What happened?" (*descriptive*), "Why did it happen?" (*diagnostic*), "What will happen?" (*predictive*), and "How can we make it happen?" (*prescriptive*).

Data analytics is typically assumed to include the complete field of applied statistics, along with the machine learning techniques [20] from artificial intelligence (AI). Data analytics encompasses a multitude of computational techniques and algorithms that involve

extracting and analyzing data from database systems and structured/unstructured data sources. Overall, analytics, as a combination of analysis and mathematics, is a multidisciplinary field with applications in practically every field of science [21–23], technology [24–26], and business [27–29]. However, it is only recently that data analytics has become a key tool in tourism to enhance decision-making outcomes and augment the customer experience.

2.2. Data Analytics in Tourism and Hospitality

The interconnection between the three disciplines of marketing, statistics, and information technology has significantly contributed to the effectiveness of businesses through accurate segmentation and predictive modeling, anticipating customer behaviors and decisions [30]. For example, the ubiquitous use of social media in tourism generates an unlimited volume of valuable data about tourists and their behaviors [31]. Accordingly, studies using big data have recently proliferated, as they represent a new opportunity to gain valuable insights into understanding and predicting tourist behaviors [31,32].

Two aspects must be highlighted in this context. First, big data can create value only through data analytics, which extracts reliable information and insights that can be interpreted for decision-making [33]. Second, applying data analytics to big data from social media, for example, and to data from surveys is not exclusive. Different data sources and types can be integrated to uncover information with a wider scope and depth. When studying images of destinations using data from social media and a survey, Ref. [34] asserted that social media data should be considered as a supplementary source of information to the commonly used survey data. For the same authors, the application of data analytics to the two types of data identified similar key image phrases, but the survey contributed insights into selected local landmarks, and the social media data identified broader and diverse characteristics of the destination. This complementarity between small data (active data collected through surveys) and big data (passive data generated through technology) was earlier suggested by Ref. [35].

Most tourism studies have focused on using analytics for tourist data generated through their activities on social media for the development of travel recommender systems. Several studies have been conducted in this direction, such as Refs. [36–38], but subsequent research highlighted the underutilization of the potential of social media data to sustain the decisions of destination management organization [39].

Despite many recent developments in big data and analytics, ref. [31] highlighted the further need to develop specific methods to track individuals' movements and behavioral patterns. Refs. [40,41] added that most research on big data and data analytics in tourism is fragmented and restricted to a limited set of questions. According to [41], regression is the main analytical technique used in tourism research. In addition, although text analytics and machine learning are gaining popularity in the field, artificial intelligence and Bayesian classification approaches are applied seldomly [41].

Geoanalytics and web and social media analytics are popular research topics in tourism research. Geoanalytics examines tourist flows and localization together to optimize activities and enhance tourism offerings [42–44]. Tourism studies related to social media analytics generally use methods such as social network analysis, comparative analysis, and trend analysis in various sectors, such as travel, destination management, hospitality, food service, theme parks, and events [45].

Sentiment analytics is another approach conducted through natural language processing (NLP) and artificial intelligence (AI) to detect emotions and opinion polarity [46]. However, in tourism, there are only a limited number of scholars who have attempted to adequately deploy this kind of analysis. For example, Ref. [47] used sentiment analytics to analyze tourist preferences and match them with the features of relevant attractions. Based on a longitudinal approach, Ref. [48] analyzed the evolution of citizens' attitudes toward tourism. Ref. [49] used sentiment analytics in food services to improve the performance of five restaurants.

In general, data analytics has been used to explore many behavioral patterns of tourists. Ref. [50] analyzed the browsing behavior of hotel websites of visitors and their temporal activity on websites to optimize hotel marketing strategies. Ref. [51] predicted the online complaining behavior of hotel guests by identifying recurrent issues by hotel category, enabling the performance improvement of the businesses. Finally, sustainability topics in tourism are addressed by data analytics studies. For example, Ref. [52] analyzed online reviews and demonstrated that online "word of mouth" about sustainability is enhanced by those sharing online reviews encompassing environmental sustainability aspects. The researchers found that such reviews are more likely to be voted as useful by other customers.

Ref. [53] applied various analytics approaches (e.g., NLP and interactive heatmap) to identify challenges and reasons that prevent individuals from exercising at home using wearable fitness technology. The findings of this research helped US government authorities develop strategies and communication approaches to encourage the practice of physical activity.

In a comparative study of travel intention and actual travel behavior, Ref. [54] analyzed a large-scale survey to investigate prospective domestic travel and used data analytics to examine the actual behavior based on a mobile data set. The study provided empirical evidence of the influence of health risk perception, mitigated by age and gender, on travel intention. This finding was validated by data analytics when analyzing actual domestic travel behavior. Furthermore, the study also demonstrated that travel behavior depends on whether the destination is indoors or outdoors.

Based on the above, tourists' behavior is a key element in understanding their needs and expectations and enhancing managerial decisions. In the present study, a novel data analytics methodology is developed to analyze structured survey data (in tabular format) containing information about restaurant diners'/patrons' behavior in the context of health outbreaks.

2.3. Analysis of Risk Behavior in Tourism Research

Understanding tourist reactions in risky or even hostile settings could help organizations adjust to the tourists' changing behaviors in such settings. Therefore, tourism scholars investigated risk perception by considering consumer behavior theories [55]. In tourism, the nature of risk is categorized broadly as follows: political instability, natural disasters, urban risks, terrorism threats, and health emergency risks.

Refs. [56,57] have identified a variety of influencing factors, including political tensions, stereotyping, and psychosocial factors that hinder tourists from visiting destinations that are presently or formerly subject to conflicts. Political conflicts have been identified as being important, in particular, for learning about the long-term impact of political instability on the perception, attitude, and behavior of tourists and for developing an understanding of travel beyond the scope of perceived risk [58].

Natural disasters were considered by Ref. [59] as a type of risk that populations have learned how to cope with. Strategies to lessen the impact of natural disasters are instigated by governments, researchers, and international organizations [60]. The impact of natural hazards on tourists is intensified due to the lack of knowledge about the destination, the language barriers, and the difficulty of accessing key information for decision-making during a disaster [61,62].

Concerning urban violence, Ref. [63] has shown that tourists' perceptions differ based on the type of communication channel chosen. Ref. [64] observed that certain proactive suggestions such as changing the accommodation type, combined with an upgrade, information updates, and provision of security devices, are proven to be highly effective in preventing cancellations. Other studies have identified that instead of canceling their planned bookings, tourists opt to replace stays in cities with vacations in rural areas [65].

Regarding terrorism risks, researchers are particularly challenged in precisely identifying the mechanisms through which terrorism fears and threats shape tourist behavior [66].

The impact of health risk perception on tourist behavior is still controversial [67,68]. Refs. [68–70] demonstrated that perceived health risks associated with HIV, drugs, and alcohol do not necessarily imply a change in tourist precautionary behavior. Ref. [54] used data analytics to analyze a large-scale survey to investigate prospective domestic travel based on the examination of the actual behavior extracted from a mobile data set. Their study provided empirical evidence of the influence of health risk perception on travel intention, which is mitigated by age and gender.

Finally, with the COVID-19 pandemic, tourism studies on health risk and its impact on behaviors have proliferated. The limited amount of previous research before COVID-19 examined topics related to health outbreaks such as SARS, Zika, bird flu, and H1N1 [71].

In a post-COVID-19 study, Ref. [72] highlighted the importance of gaining a deeper understanding of how individual behavior changed during the COVID-19 pandemic. Data analytics methods and techniques were used to provide additional insights into the strategies adopted by tourists to protect themselves from infectious hazards while consuming tourism services. Confinement during the pandemic encouraged sedentary and unhealthy lifestyles that could cause mental and other health challenges for individuals. Ref. [73] developed novel approaches to extract information from posts on COVID-19 published on Reddit. This study is relevant to our work, as we also analyze data relevant to COVID-19 and use machine learning techniques. While the cited work used a supervised machine learning technique of classification to classify posts, we use the supervised machine learning technique of decision trees for descriptive and diagnostic analytics to create risk behavior profiles. In another study using deep learning models to analyze restaurant reviews during the pandemic, Ref. [74] concluded that restaurant diners/patrons were concerned with the following aspects, listed as per their importance: "Service", "Food", "Place", and "Experience". From a methodological perspective, the authors demonstrated that, when compared to machine learning algorithms, deep learning algorithms provide more reliable results in review score prediction and sentiment classification. Again, in relation to COVID-19, Ref. [75] established a strong relationship between restaurant complaints related to safety measures. The same study also reported cases using multiple methods, including neural-network-based deep learning algorithms and spatial modeling.

In this paper, a novel data analytics methodology is developed with the objective of analyzing structured survey data, which are in tabular format. The lasting presence of the COVID-19 health risk allowed the authors of this article to contribute to the existing knowledge by investigating the actual risk reduction behavior of restaurant diners/patrons, an unexplored area of study [76]. Additionally, this paper is based on the protection motivation theory [12], rooted in the health belief model [13]. Ref. [77] affirmed that only a handful of tourism articles embedded these frameworks when examining health behavior in risk circumstances. Hence, incorporating protection motivation theory within the context of a novel data analytics methodology is the theoretical contribution of the present research.

2.4. Factors for Tourist Profiling in Health Risk Context

As opposed to "segmentation", which refers to common patterns of a large group of people, "typology" designates the distinct characteristics of a small group of individuals [78]. Although segmentation and typology may use different methods, they both aim to categorize and profile groups based on certain characteristics and traits (as factors) to better understand and manage these groups. Several factors can be considered to profile customers, including geodemographic and socioeconomic factors, product-specific factors (frequency of purchase, loyalty, usage situation), and psychographic variables such as personality, behavior, motivations, and lifestyle [79]. In this study, restaurant diners'/patrons' behavior patterns, worry (psychological factor), and sociodemographic variables are used to explain a risk-coping typology of customers.

Tourism and, most importantly, hospitality businesses are among the most vulnerable fields in an unstable or risky economic or social environment [1–3]. Studying tourists' risk perceptions and, most importantly, their behaviors to cope with risk, is valuable and

important. Several scholars [80,81] have identified risk factors as the foremost category of factor variables affecting tourist decisions and intentions toward consuming tourism services and products. Identifying the variables explaining the risky (and not risky) behaviors of tourists during a health outbreak can help guide policymakers in managing health emergency crises and hastening recovery by designing targeted and customized awareness and communication strategies [82,83].

The unprecedented COVID-19 crisis triggered changes and transformations in the tourism sector [84]. According to Ref. [6], the pandemic dismantled ("re-mantled") the sectors at three levels: tourism demand, tourism supply, and destination management. From the perspective of tourism demand, the perspective adopted by this study is the research stream focused on exploring the formation of risk perception and its association with behavioral intention and patterns [85,86], or better preparedness for future outbreaks. Ref. [6] (p. 313) iterated on the need to better understand tourists' "behavioral, cognitive, emotional, psychological, and even ideological drivers, actions, and reactions" to health outbreaks.

Psychological antecedents [87], sociodemographic factors [88–90], culture [83,91], and previous experiences [91–93] have been identified as the main variables that shape tourist behavior. It is established that tourists' reactions and behaviors, as well as their aversion to risk, depend on sociodemographic factors and their personality traits [84,94,95]. At the same time, it is admitted that health risk directly affects people's well-being and psychology. Therefore, Ref. [25] stressed the importance of investigating psychological factors impacting tourist behavior. In tourism research, risk perception is commonly associated with worry (a psychological factor) [96,97]. For this reason, adding to the sociodemographic factors, this research considers the factor of worry as well as the risk preventive behavior and the risk reduction behavior. These variables were also selected for their practicality as they are observable and easy to measure for timely decision-making.

In this study, we consider four constructs (question categories) and their measurements (questions/attributes) with primary data from a survey:

A.   *Worry*, representing emotional and affective reactions;
B.   *Risk prevention behavior*, representing cognitive factors;
C.   *Risk reduction behavior*;
D.   *Demographics,* representing individual characteristics.

The following subsections expand on each of the listed constructs.

It is important to note that the attributes for *risk reduction behavior* (construct C) are used to create two derived attributes as targets/responses: a numerical attribute (BHV_SCORE) as the behavior score and a categorical attribute (BHV_CLASS) to represent the behavior class. Subsequently, behavioral risk profiling is conducted to reveal the typologies of respondents belonging to the low- vs. high-risk behavior classes.

2.4.1. Emotional and Affective Reactions

In the case of a health pandemic event, people's reactions can be expressed through fear, anxiety, and worry [87,98,99]. However, in tourism, perceived risk, uncertainty, anxiety, fear, and worry concepts are generally ambiguous. These concepts are often considered to define the same phenomenon, which causes inconsistencies in many studies in the field [100,101]. During COVID-19, these concepts were used in research studies to investigate their relationship with risk perception and behavior. Therefore, it is best to clarify the differences between these concepts, especially in the context of tourism studies, before advancing further.

Fear reflects the consciousness of danger and is generally associated with uncertainty, making an individual frightened and precautious when making decisions [102]. Fear is not the same as anxiety. Considered a mental disorder in psychology, anxiety is a new concept in tourism [103]. Anxiety describes mental tension and reaction to stress induced by unknown consequences [104]. Worry differs from anxiety, even if it is considered an interrelated concept [105]. According to Ref. [105], anxiety is linked to self-belief in problem-

solving incapability, whereas worry is about the negative impact of an unmanageable and chaotic series of thoughts about an unknown future.

Worry is an emotional reaction that has been widely researched in tourism [106] and is defined by Ref. [107] (p. 261) as "an individual's attempt to engage in mental problem-solving regarding tourist trip-related issues where outcomes are thought to be uncertain and contain possibilities for negative results". It is generally admitted that worry is an important determinant of tourist behavior [108]. Tourists with high levels of worry carefully plan their trips, conform to safety measures, and contract travel insurance [109,110]. Similarly, Ref. [111] demonstrated that worry reduces travel intentions.

2.4.2. Cognitive Factors

Cognitive factors include aspects related to health hazard severity level, risk event information and communication management, access to information, stereotype salience, perceived control over risk, and risk preventive measures or behavior [87,99,112].

Ref. [113] investigated dine-in habits during COVID-19 and found that the perceived severity of the risk, among other factors, strongly explains the diners'/patrons' co-creation behavior. Ref. [114] suggests that optimistic and pessimistic information impacts individuals' consumption and that their risk perception is asymmetric in the context of dairy product contamination. According to Ref. [115], exposure to coronavirus information on social media increases females' health risk perception and risk reduction behavior. The stereotype salience factor is the stereotype according to which a certain category of people is subject to or attracts adverse events [116,117]. This factor has not been sufficiently investigated in the specific case of health risks in the tourism and hospitality sectors. Amongst the very few studies, Ref. [118] investigated the infectability stereotype of tourists and found that "perceived COVID-19 infectability relates positively with tourist negative stereotype, which then relates negatively with resident hospitality." [118] (p. 1). When studying perceived control over risk, Ref. [87] confirmed that perceived control has a positive link with protective behavior. Ref. [119] stated that tourists prefer staycation when they perceive their control over risk as low. Ref [120] considered self-protective behavior consisting of adopting the coronavirus preventive measures as set by the World Health Organization [121]. Ref. [120] found that health risk perception was consistently and positively related to the adoption of regulatory precautionary measures. Refs. [113,122] have found that clients exhibiting behavior of compliance with preventive measures tend to actually visit restaurants.

2.4.3. Risk Reduction Behavior

Ref. [123] reports that, across tourism studies, risk has been investigated mostly in the form of risk perception (61.6%) and risk-taking behavior (29.1%). Other streams of research "focused on determining and predicting factors (27.9%) and on consequential impacts (5.8%) of risk perception and behavior" [123]. Select studies have applied the "Cusp Catastrophe Model" to explain tourist behavior during health emergencies. The mentioned model relies on the "catastrophe theory", which is a mathematical model that describes system behavior such that a progressively changing force can generate a sudden effect [124]. In relation to the SARS (severe acute respiratory syndrome) outbreak, Ref. [125] found that Hong Kong travelers exhibited a behavior that empirically fitted the Cusp Catastrophe Model.

With the COVID-19 outbreak, research into health risks and impact of risk on behaviors in hospitality and tourism have proliferated. A review of multiple recent publications shows that most of health-risk-related COVID-19 research has focused on investigating *intended* behavior and *planned behavior* instead of exploring actual behaviors during the outbreaks [126–129]. Most scholars surveyed tourists before their trips [130–132], while Ref. [87] examined tourist preventive health behavior (e.g., vaccination, health insurance, etc.) preceding an *actual* trip by collecting data from *actual travelers* at airports before their departure. Ref. [87] is among the few publications focusing on actual travelers instead of potential travelers.

When the risk is a direct stimulus, and while tourists are traveling, two concepts are commonly used to characterize behavior: risk *avoidance* behavior and risk *reduction* behavior. Adopting risk avoidance behavior refers to avoiding behaviors that may engage individuals in risky situations. Risk reduction behavior supposes that individuals engage in situations and activities that may carry risk, but at the same time, they adopt mitigating behavior to diminish the possible adverse consequences. Risk reduction behavior is frequently examined in the healthcare field [133] but has rarely been investigated in tourism studies [76]. Two examples of this stream of research are Refs. [134,135]. Ref. [134] found that backpackers exhibit risk reduction behavior by visiting attractions only when accompanied by local inhabitants of the destination. Ref. [135] found that the most significant risk reduction behavior in visiting highly rated restaurants is to mitigate food contamination risk.

This study included risk reduction behavior in the survey as a construct (question category) with 27 measurements (questions/attributes). Furthermore, two response variables (target attributes), namely behavior score and behavior class, were derived from the measurements of this construct.

2.4.4. Demographics

Demographic (individual characteristics) dimensions included all aspects related to sociodemographic differences, past experiences, cultural backgrounds, and personality traits. A study in China found that aging and uneducated individuals have inadequate knowledge about the coronavirus and tend to care less about adopting preventive measures [136]. Ref. [137] claimed that sociodemographic variables such as age and gender, as well as Hofstede's uncertainty index, influence risk perception and behavior. The latter index also significantly influences destination perception. Ref. [138] concluded that when the cultural distance between tourists and local communities is significant, tourists tend to adopt risk reduction behaviors (e.g., opt for organized trips and use tour guides).

Multiple studies [136,139] have shown that previous visits to destinations reduce the risk perception of destinations. Despite this commonly accepted relation, in the case of the COVID-19 pandemic [140], past travel experiences did not imply a lower risk perception.

The findings of a research study undertaken in Qatar, a country geographically close and culturally similar to the UAE, revealed that "conscientiousness, neuroticism, risk perception, and personal hygiene practices predicted social distancing" [141] (p. 237). The results of Ref. [142] suggest that neuroticism and conscientiousness decrease the intention to travel, whereas extroversion and openness predict a higher willingness to travel.

According to Ref. [143], tourists are willing to pay premiums to reduce risk-taking. Yet, this behavior depends on their age and level of revenue [144]. Ref. [145] confirmed that sociodemographic factors during COVID-19 significantly impacted trip frequency, experience with the pandemic, and risk perception.

3. Methods

This section discusses the study's location, the collected data's scope and constructs, the data collection and validation process, the steps of data preparation, the developed methodology and its steps, and the data analytics techniques used.

3.1. Location

This study was conducted in the United Arab Emirates (UAE), a Gulf country with an estimated population of >9 million residents [146,147]. The UAE is a union of seven "Emirates" (states/regions), with the Abu Dhabi Emirate being home to the capital city. The Emirates of Abu Dhabi and Dubai are especially well-known around the world for their inspiring architectural and sustainability projects and touristic sites [148]. While the UAE's primary economic sector is still oil and gas, with a share of 30% in the gross domestic product (GDP) [149], the country, over the years, has become a success story in transforming itself into a hub of commerce, finance, tourism, hospitality, and a multitude of other sectors.

The UAE is the top country in the world for the percentage of resident expatriates, representing 88% of the population [150]. The UAE ranks as the 32nd best country in the world to do business [151], resulting in a further flow of people, supporting the tourism and hospitality sectors. Furthermore, the UAE is a top tourist destination, with an estimated 14.4 million overnight visitors in 2022. Due to these facts, and in part owing to the systematic handling and management of the COVID pandemic [152,153], soon after the pandemic ended, the UAE was able to return to 86% of its tourism volume of the pre-COVID year 2019 [154]. Building on this achieved success, the UAE has higher goals in tourism. For example, the Dubai Emirate aims to double foreign trade by magnitude and make the city one of the top three cities in the world for tourism and business by 2033 [155].

3.2. Data

The data used in this research study were collected in the United Arab Emirates (UAE) over one month (December 2021) from a sample of 301 respondents. The survey included a section with sociodemographic questions and another to measure the restaurant patron's worry based on a recently validated scale (see point A below). The third section encompassed preventive behaviors per the World Health Organization (see point B below). The last section was dedicated to risk reduction behaviors, covering all dining-out journey stages, from information search to payment. The latter included 27 risk reduction behaviors identified through semi-structured interviews with restaurant diners/patrons (see point C below). There were 16 participants consisting of residents and tourists reflecting different nationalities and backgrounds. The saturation level was attained after the 12th interview. No new patterns were identified when interviewing subsequent participants. The survey was first tested on a small sample consisting of seven restaurant managers and ten restaurant customers across five restaurants, mainly to adjust measurements/categories A, C, and D. The restaurants were also varied, consisting of one fine dining, one fast food, one local food, and two ethnic restaurants. Once the survey questions were ensured for clarity, survey data were collected through an online survey following a snowball approach. The link was sent to the authors' families and friends who were hosting visitors and acquaintances on holidays in the UAE. In their turn, they shared the link with their network. The survey included two screening questions preventing individuals from responding when they were located outside the UAE and when they did not visit a restaurant within the last 2–3 weeks. The survey also prevented respondents from answering more than one time from the same IP.

In total, 67% of the sample consists of official residents of the UAE. The composition of the sample with respect to the emirate (state/region) of residence is similar to the population of each emirate within the UAE. The sample consisted of approximate participant percentages of 35% from Dubai, 28% from Abu Dhabi, 14% from Sharjah, 8% for Ajman, 7% for Al Ain, 3% for Fujairah and Umm Al Quwain, and 2% for Ras al Khaimah. The ages of 83% of the respondents ranged from 20 to 50 years. Other demographic attributes included gender (60% female and 40% male), marital status (46% married, 45% single, 12% divorced, and 7% widowed/widower), and latest educational diploma (54% bachelor, 27% master, 12% high school, 5% doctorate, and 2% primary school).

The constructs, measurements, targets/responses, derived datasets, and their relations are illustrated in Figure 1. Each construct is shown with a different fill color. Numerical attributes are shown with a green color and categorical attributes are shown with a purple color.

The constructs (question categories) in the survey were mainly the following (Figure 1), with the number of questions (measurements) in each construct (category) mentioned in parentheses:

A. *Worry* (7 questions/items)
B. *Risk Prevention Behavior* (5 questions/items)
C. *Risk Reduction Behavior* (27 questions/items)
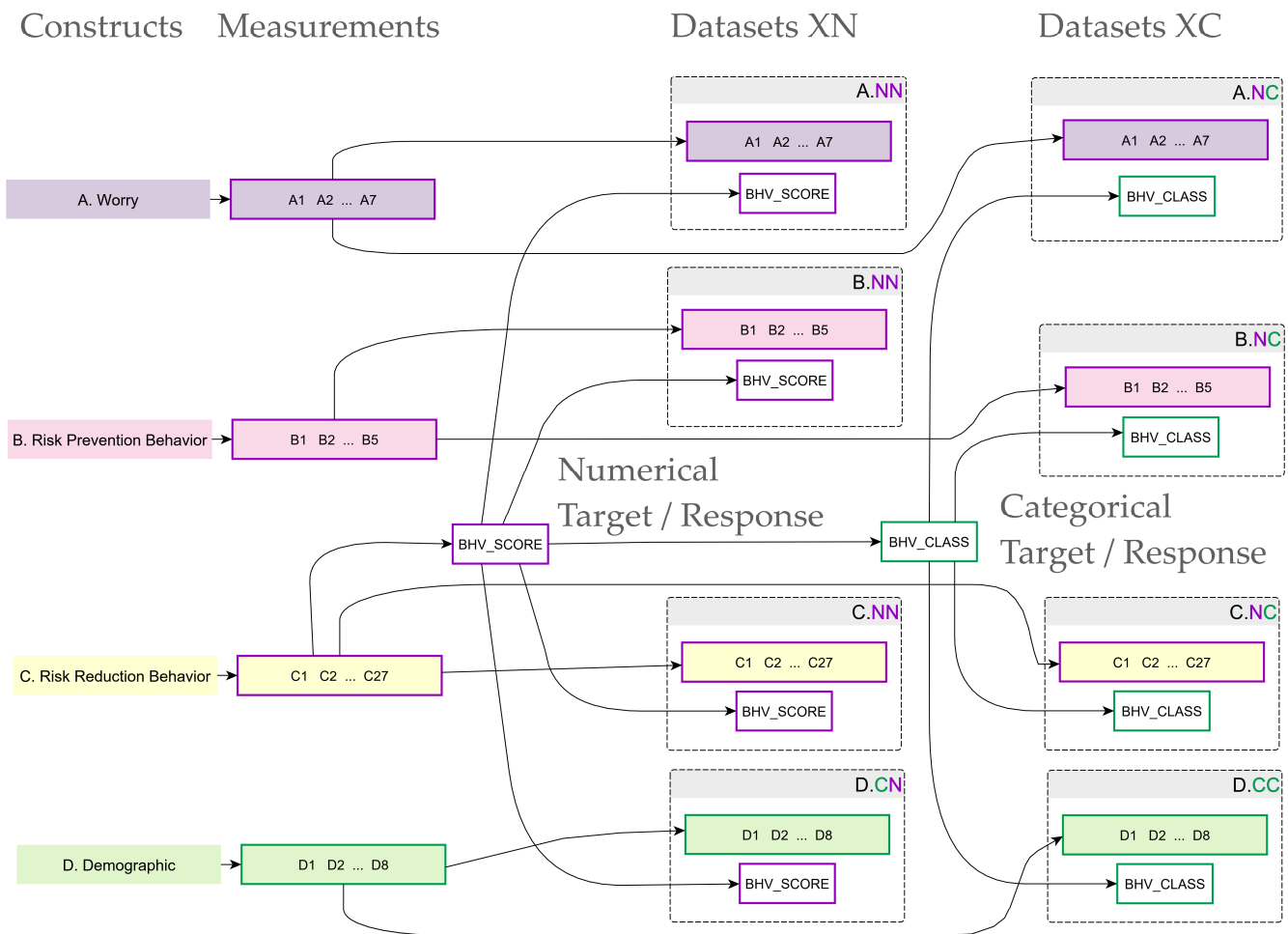D. *Demographics* (8 questions)

Figure 1. The constructs, measurements, targets/responses, datasets, and their relations.

The constructs, measurement items, and scales were adopted and/or developed as follows:

A. The construct and its measurement items and scales were adopted from Refs. [156,157].

B. The construct and its measurement items and scales were adopted from Refs. [158,159].

C. The construct was adopted from health research. Measurement items were developed by the authors through interviews; the scale was adopted from Ref. [160], with the addition of a measure from Ref. [46].

D. Factors/attributes were compiled from various research studies and adopted by the authors in the UAE context (e.g., different "Emirates," instead of different "states").

Additional justification for constructs A, B, and C (Figure 1) are as follows:

A. Ref. [156] focused on developing a specific scale to measure worry about the COVID-19 virus. This same scale was re-validated by Ref. [157]. This scale encompasses seven items measured using a Likert scale of four levels (1 = "Not at all"; 4 = "Very much").

B. Precautionary measures announced by the World Health Organization [158], including wearing masks, keeping social distancing, washing, or sanitizing hands amongst others, are preventive individual gestures that have shown efficiency in self-protection from the virus and in limiting the spread of the virus. In the present study, the preventive behavior was measured through five questions related to social distancing, with the questions being on touching the face, washing or sani-

tizing hands, wearing facemasks, and wearing gloves. The questions asked about the frequency and the observance of these precautionary behaviors. A five-point Likert scale was used ranging from 0 to 4 (0 = Never; 1 = Rarely; 2 = Sometimes; 3 = Frequently; 4 = Always).

C.   This aspect represents a significant contribution as it aims to measure the actual behavior of restaurant customers. The 27 patterns of "Risk Reduction Behavior" at restaurants were identified through semi-structured interviews. The interviews explored the adopted strategies by diners/patrons to reduce risk before and during their restaurant visits. As recommended by Ref. [161], the interviews also included questions dealing with information search regarding recommended restaurants. A total of 16 restaurant customers who had recently visited a restaurant were conveniently selected. The interviewees were equally composed of UAE residents (including two Emirati nationals) and tourists. The saturation level of answers was reached after the 12th interview. Analyzing the frequency of repetitive answers related to risk reduction behaviors allowed for the identification of 27 patterns. The question related to each behavioral pattern asked about diners'/patrons' adoption frequency of the same behaviors on a scale ranging from 0 = "None" to 4 = "Always", as suggested by similar research Ref [160].

Further motivation and justification for the questions are detailed in the "Data Collection" section of Ref. [9].

The survey was reviewed by the Social Sciences Research Ethics Committee (SS-REC) of the United Arab Emirates University (ERS 2021 8402), and ethical approval was issued on 23 November 2021. The interviewees and survey participants provided informed consent, and the questionnaires were anonymized. Both interviewees and questionnaire participants were free to withdraw from participation at any time.

3.3. Data Validity

A valid question that needs to be asked is whether the sample is representative of the targeted population. The targeted profiles were tourists and residents who had visited restaurants during the observed period. Thus, the sample data were collected to represent the tourists and residents in the UAE who visited restaurants rather than the broader population in the UAE. Having a representative sample of tourists is a serious challenge, mainly due to the highly changing profiles of tourists. This dynamism is because of multiple factors, including seasonality and ticket prices, which in turn depend on oil prices and other macroeconomic factors (e.g., local demand/supply balance). In terms of sampling, the research aimed to collect data from residents and tourists who had visited restaurants at least twice during the three weeks before the survey was conducted. Ref. [97] firmly asserted the infeasibility of perfectly identifying a random and representative sample of tourists and entertainment diners/patrons, given that this population is not a well-defined group and has a changing profile over time. For representation purposes, the initial approach was to directly collect data from restaurant diners/patrons as they left the dining outlets. However, the research team was not able to obtain the required authorization given the pandemic circumstances. Accordingly, the team used an online questionnaire following a non-probabilistic approach: the snowball technique. This technique is recommended when the population is unknown or rare. Accordingly, the snowball technique allowed the research team to collect data by sharing a link with a network of UAE residents and tourists visiting families and friends. It is challenging to assess the representation of the 301 observations and the external validity of the findings. However, this could be mitigated by the fact that during the pandemic, the population of diners/patrons (consisting of both residents and tourists) who visited restaurants in the UAE is unknown.

Another valid question that can be asked is whether the sample size is sufficient. A related valid question is what methodology was used to determine the sample size. Firstly, it is important to note that our study differs from most survey-based research, which uses structural equation modeling (SEM) and its variants [162]. Therefore, the

sample size calculations for SEM do not necessarily translate into direct values for sample selection for our research. Yet, as a benchmark, the sample size calculations for SEM were carried out [163] with Gpower software version 3.1.9.7 [164] and following the guidelines in Ref. [165]. Selecting the F test as the test family, linear regression (fixed model, R^2 deviation from zero) as the statistical test, a priori as the type of power analysis, and with the parameter values of f^2 = 0.15 (medium), α error prob = 0.05, power $(1 - \beta$ error prob) = 0.75, and the number of predictors as 27 (the number of questions for Category C, which is the largest question set), the required minimum sample size computed by Gpower software was 249. Even if an SEM model were to be constructed using all 47 questions, the required sample size, as computed by Gpower, was 314. The obtained required sample sizes of 249 and 314, if SEM were to be conducted, are, respectively, below or only slightly above the current sample size of 301 in our study. Thus, even though SEM was not the methodology used in our research, our sample size met the benchmark values of a possible SEM study.

Upon the completion of data collection, there were no missing data points in the different question categories. Since the different question categories were taken from different inventories in earlier research, their scales were not the same and were eventually scaled to a 1–5 Likert scale. The full list of questions in the survey can be found in Appendix A of the Supplement document [11].

Confirmatory analysis of the dataset produced Cronbach's alpha [166] values of 0.875, 0.62, and 0.865 for constructs (categories) A, B, and C, respectively. While Cronbach's alpha values of constructs A and C were above the recommended threshold of 0.7, that of construct B was not. Removing one of the five measurements (attributes/factors), namely B_05 "How often are you wearing gloves?", from the scale resulted in an increase in Cronbach's alpha from 0.62 to 0.74, which suggests that this item can be removed if structural equation modeling is applied. Because the main technique in our analytical methodology was decision trees, with the presented output being rules that profile diners/patrons, B_05 was retained in the sample during the analysis to be able to obtain richer insights.

Related research [167] by one of the authors of this paper applied SEM to an extended version of this dataset and found the following:

- Worry is positively related to risk reduction behavior. In terms of the constructs in the present research, Construct A is positively related to Construct C, with p = 0.001 and coefficient = 0.19.
- Health risk perception (HRP) is a mediator construct that mediates the effects of worry on risk reduction behavior. HRP is not a construct that is included in the present research. In terms of the constructs in the present research, Construct HRP mediates the effects of Construct A on Construct C, with p = 0.0002 and coefficients = 0.43 and 0.22 for worry→HRP and HRP→C, respectively.
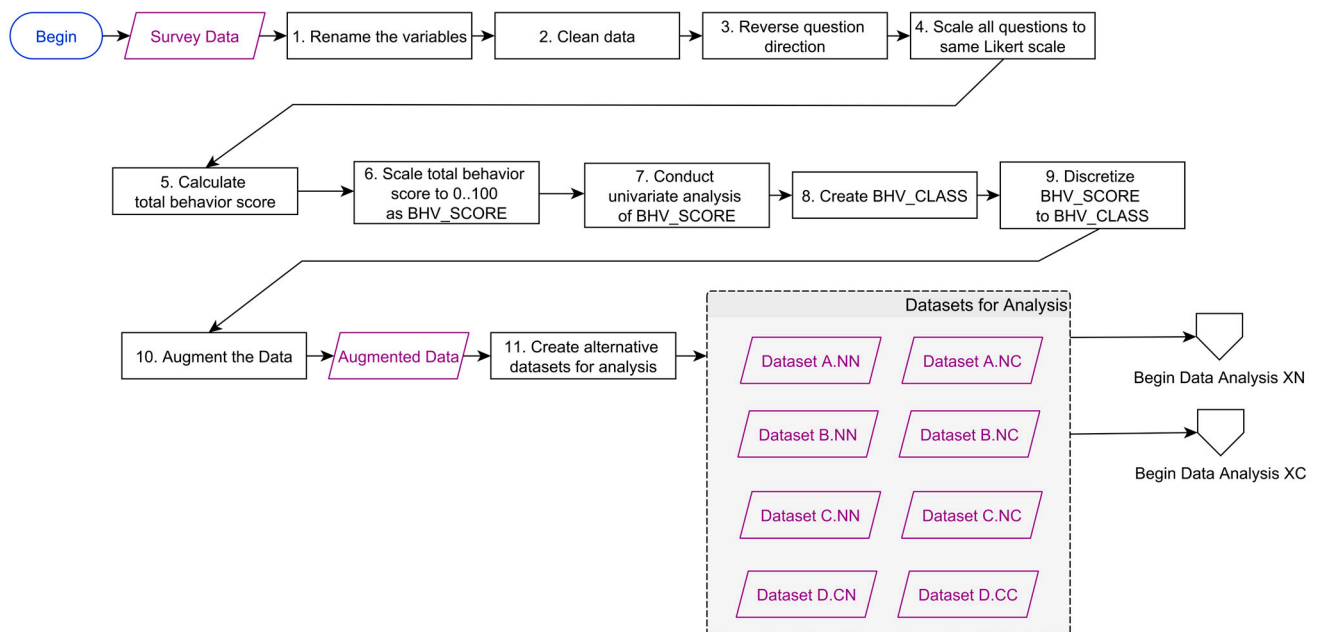
3.4. Steps of Data Preparation

After selecting the question categories to be included in the research, datasets for the analysis were created (Figure 1 and Table 1). For each of the above four question categories (A–D), two datasets were constructed, one with a numerical response and the other with a categorical response. Each dataset was then named based on the question category (A–D) and the data types of the factor and response variables (NN, NC, CN, or CC). The characteristics of the datasets and analyses corresponding to these categories are presented in Table 1. For example, the first row of Table 1 mentions Dataset A.**NN**, which is a dataset for question Category A (Worry) and has **n**umerical factors and a **n**umerical response.

The primary question category to characterize behavior was C, namely risk reduction behavior (BHV). The analysis focused mainly on the relationship between question Categories A–D as factors (independent variables) and Category C (risk reduction behavior) as the response (dependent variable). The synthesis of all the datasets in relation to the initially decided constructs and measurements is illustrated in Figure 1. Furthermore, Figure 2 shows the steps of the methodology until the end of the creation of these datasets. In Figure 2, Steps 5–11 illustrate how the datasets were generated.

Table 1. Datasets created based on the original data.

Dataset/Analysis	(Independent) Factor Category	Factors Data Type	(Dependent) Response	Response Data Type
A.**NN**	Worry (A)	**N**umerical	BHV_SCORE	**N**umerical
A.**NC**	Worry (A)	**N**umerical	BHV_CLASS	**C**ategorical
B.**NN**	Preventive Behavior (B)	**N**umerical	BHV_SCORE	**N**umerical
B.**NC**	Preventive Behavior (B)	**N**umerical	BHV_CLASS	**C**ategorical
C.**NN**	Risk Reduction Behavior (C)	**N**umerical	BHV_SCORE	**N**umerical
C.**NC**	Risk Reduction Behavior (C)	**N**umerical	BHV_CLASS	**C**ategorical
D.**CN**	Demographic (D)	**C**ategorical	BHV_SCORE	**N**umerical
D.**CC**	Demographic (D)	**C**ategorical	BHV_CLASS	**C**ategorical

Figure 2. Steps of the data analytics methodology, corresponding to data cleaning and preparation, resulting in the creation of the datasets in Table 1.

- To represent all answers to question Category C as a single number, the numerical answers to all questions in Category C were summed (Step 5 in Figure 1) and then scaled (Step 6), resulting in the calculation of a standardized BHV_SCORE (in the range of 0–100).
- Furthermore, BHV_SCORE was discretized to generate a categorical response (dependent) variable of BHV_CLASS (Steps 7–10). This variable takes the values of HighRisk, MediumRisk, and LowRisk, corresponding to low-, medium-, and high-risk reduction behaviors, respectively.

HighRisk and LowRisk were the main target values in the analysis of A.NC, B.NC, C.NC, and D.CC of Table 1.

- Respondents with the label HighRisk were the *least cautious* in avoiding any COVID-related risks, with the *lowest* BHV_SCORE *values*, and hence exhibited *high-risk behavior*.
- Conversely, respondents with the label LowRisk were the *most cautious* in avoiding any COVID-related risks, having the *highest* BHV_SCORE *values*, and hence exhibited *low-risk behavior*.

Once the data were prepared (Figure 1) and the different datasets were readied, two main types of analysis, as represented by the right-most symbols in Figure 2, were conducted:

- Analysis X**N**: Analysis of datasets with **n**umeric responses (left side of the gray panel in Figure 3).
- Analysis X**C**: Analysis of datasets with **c**ategorical responses (right side of the gray panel in Figure 3).
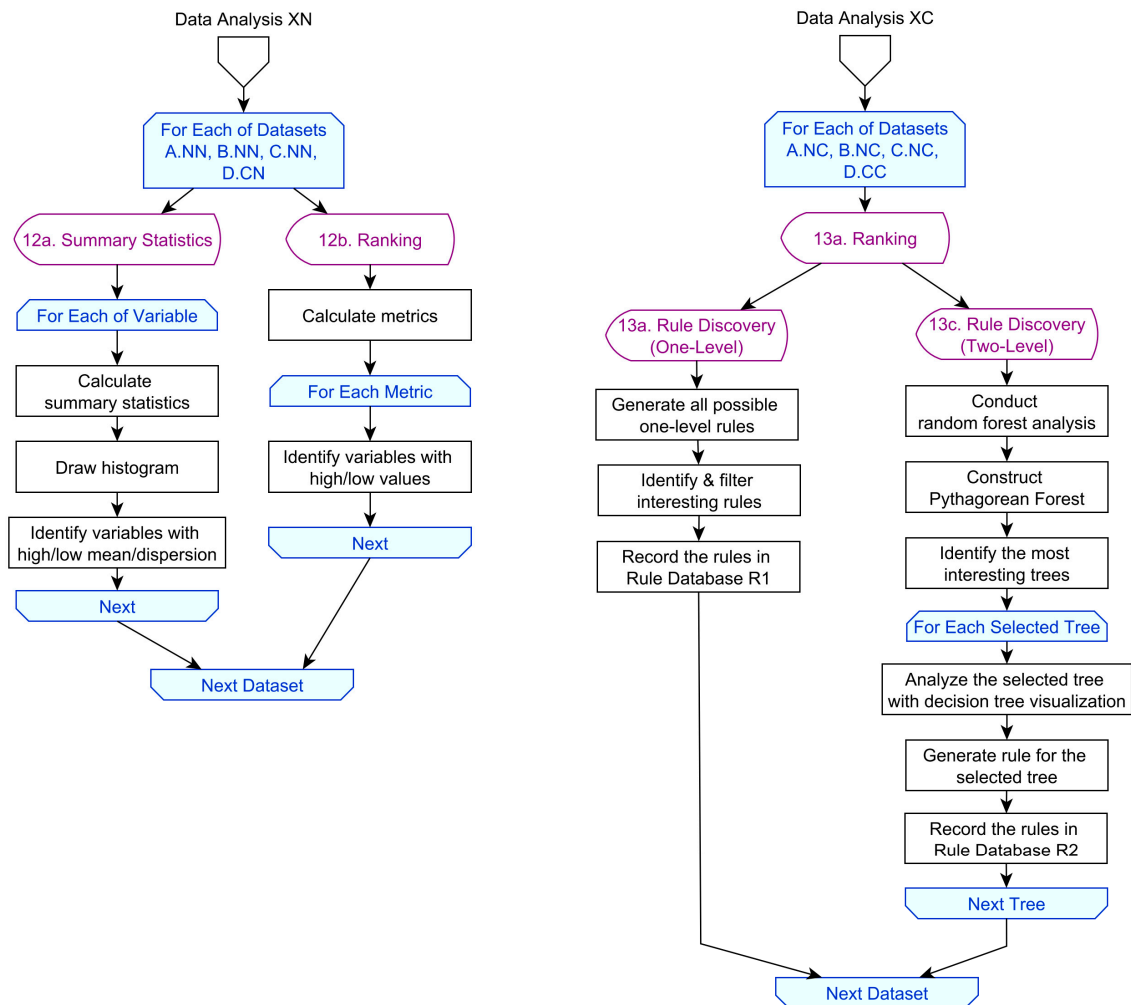

Figure 3. Steps of the data analytics methodology for the analysis of XN and XC datasets (XN: any type of factor and numerical response; XC: any type of factor and categorical response).

The details of these two analyses are presented in Figure 3 as workflows and explained in Section 3.6. The analysis of other questions has been left out of the scope of the current study and is reserved for future investigation.

3.5. A Novel Analytics Methodology

A novel integrated data analytics methodology was custom-developed and implemented to analyze the data collected in this study. The analytical methodology was designed specifically for this type of data to extract as many interpretable and actionable insights as possible. Given its novelty and applicability, the developed analytics methodology is the main methodological contribution of the present study. Such an integrated methodology, specifically using statistical summaries and machine learning techniques in tandem, was not encountered in the existing literature and is thus the major theoretical contribution of the present study.

The developed methodology is descriptive and diagnostic [18,19]. The methodology, analysis, and results are presented in the paper and the Supplement document [11] as follows:

- The characteristics of each dataset and the list of analyses corresponding to these datasets are coded and listed in Table 1;
- The constructs, measurements, targets/responses, and the datasets, and their relations are illustrated in Figure 1;
- The **data analytics methodology** developed for and applied in this research is presented as flow charts in Figures 2 and 3 and pseudo-code in Section 3.6;
- The detailed **steps of the data analytics** process applied in the methodology are provided in Section 3.6;
- The specific data analytics **techniques** integrated within the analytics methodology and applied in the study are **defined and described** with citations to sources in Section 3.7;
- The methodology (Figures 2 and 3) was implemented mostly within the Orange data-mining software, version 3.35 [168]. The **data analytics workflow (data science pipeline) in Orange** is shown in Figure 4. The software is a public domain open-source software whose source code can be accessed under Github [169];
- The sample **results** obtained by the applied techniques are presented in Section 4, Tables 2–10, and Figures 5 and 6;
- The results in this paper are only partial owing to space limitations. **Full analysis results** for all datasets in Table 1 are presented in Appendices B–E of the Supplement document [11].
- The **"one-level rule discovery"** in the methodology was carried out through a **custom-written** VBA (Visual Basic for Applications) code. The ChatGPT generative AI platform [170] was used to generate the VBA code, and then the VBA code was executed within MS Excel to generate the results. The ChatGPT prompt and VBA code are presented in Appendices F and G of the Supplement document [11], respectively.
- An interactive **web-based analytics dashboard** [10] has been developed that visualizes the one-level rules. The dashboard enables interactive visual exploration of the one-level rules, insight discovery, and decision support.
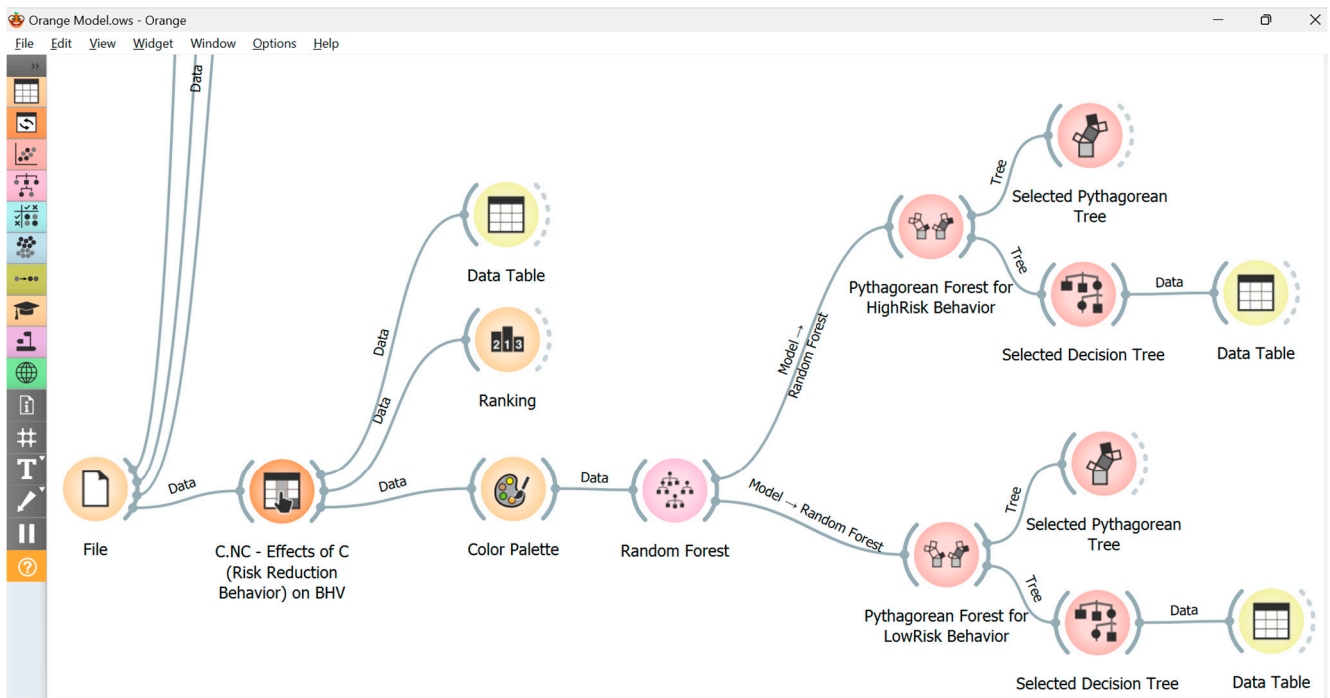

Figure 4. Analysis Steps 13a and 13c in the Orange data mining software, with a focus on C.NC.

Table 2. For Dataset C.NN, summary statistics for mean values for questions in Category C, displaying the questions in Category C with the highest and lowest mean values.

QuestionID	QuestionText	Mean
C_08	I dine out with people that I do not know necessarily well	3.78
C_21	I verify if the plate and the table cutlery are clean	3.66
C_19	I observe if the waiters are constantly wearing masks	3.38
C_27	I use WiFi payment means	3.35
C_11	I select dining outlets that are not crowded	3.34
. . .	. . .	. . .
C_10	I select dining outlets recommended by social media as COVID-19-safe	2.86
C_02	I do not dine out in fast-food	2.83
C_18	I ask the waiters to keep a reasonable social distance with me	2.74
C_24	I ask questions about how the dish was prepared	2.63
C_25	I ask the waiters to wear gloves when they are serving me	2.44

Table 3. For Dataset C.NN, summary statistics for dispersion values for questions in C, displaying the questions with the highest and lowest dispersion values.

QuestionID	QuestionText	Mean
C_25	I ask the waiters to wear gloves when they are serving me	0.51
C_18	I ask the waiters to keep a reasonable social distance with me	0.44
C_20	I observe if the waiters are constantly washing their hands with sanitizers	0.43
C_24	I ask questions about how the dish was prepared	0.42
C_26	I wear back my mask each time I finish eating	0.41
. . .	. . .	. . .
C_01	I select dining outlets offering healthier food	0.32
C_21	I verify if the plate and the table cutlery are clean	0.32
C_06	I dine out with my family members	0.32
C_08	I dine out with people that I do not know necessarily well	0.30
C_05	I dine out in seated dining outlets	0.29

Table 4. For Dataset C.NC, variables that rank highest and lowest with respect to their power in predicting BHV_CLASS.

| Rank	QuestionID	QuestionText	Gain Ratio	Gini
1	C_13	I eat in dining outlets clearly displaying the required precautionary measures	0.153	0.111
2	C_16	I complain if I observe that the dining outlet does not follow the precautionary measures	0.152	0.104
3	C_14	I leave the dining outlet if I do not get the first impression that it is COVID-19-safe	0.140	0.096
4	C_17	I request for a table that is located far from other clients	0.123	0.090
5	C_12	I book a table only when the dining outlet is not at the full authorized capacity	0.112	0.090
. . .	. . .	. . .	. . .	. . .
16	C_04	I order food instead of going out to dine	0.055	0.041
17	C_03	I dine out in high-end/high-category dining outlets	0.044	0.034
18	C_05	I dine out in seated dining outlets	0.039	0.032
19	C_02	I do not dine out in fast-food	0.041	0.032
20	C_08	I dine out with people that I do not know necessarily well	0.015	0.013
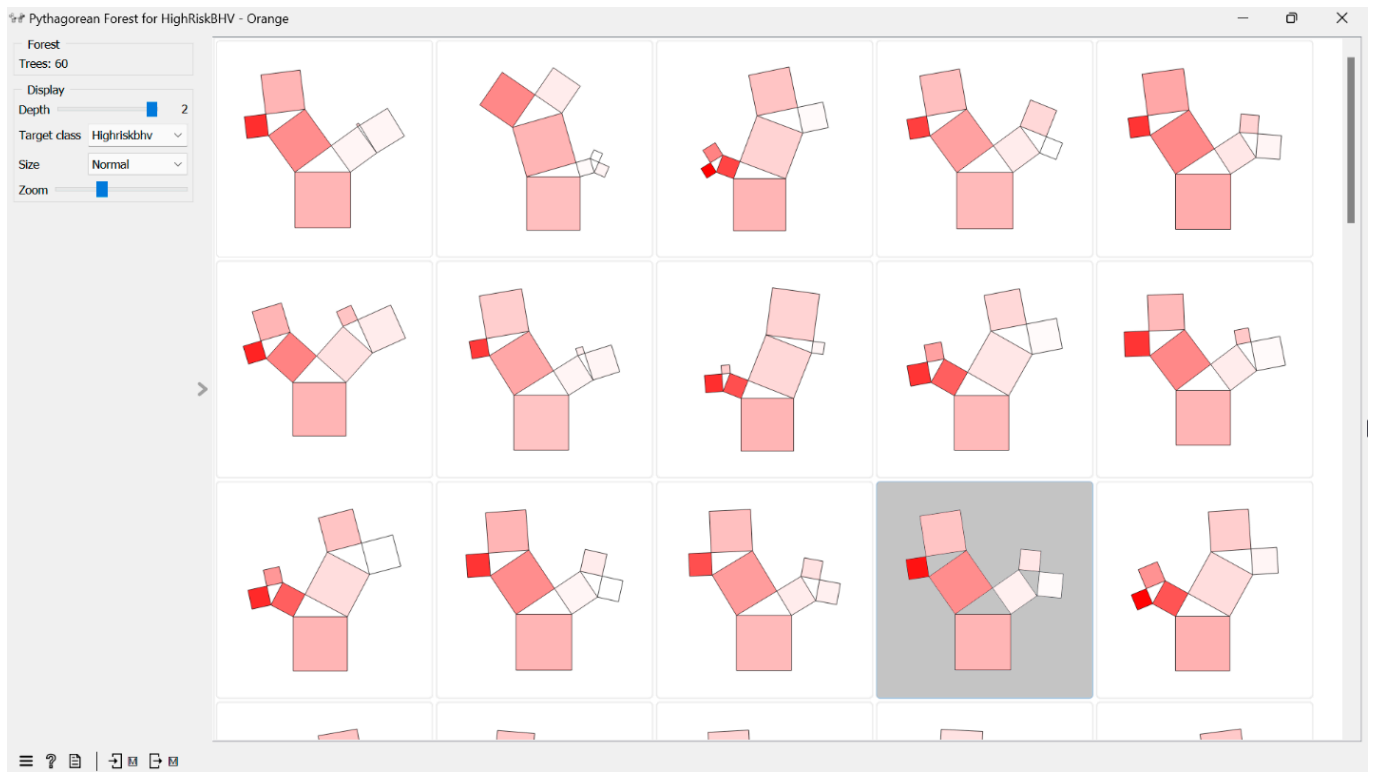
Figure 5. For Dataset C.NC, a partial view of Pythagorean trees for the generated random forest. Darker tones of red color denote higher probability of HighRisk behavior.

Table 5. Selected one-level rules for Category A (worry).

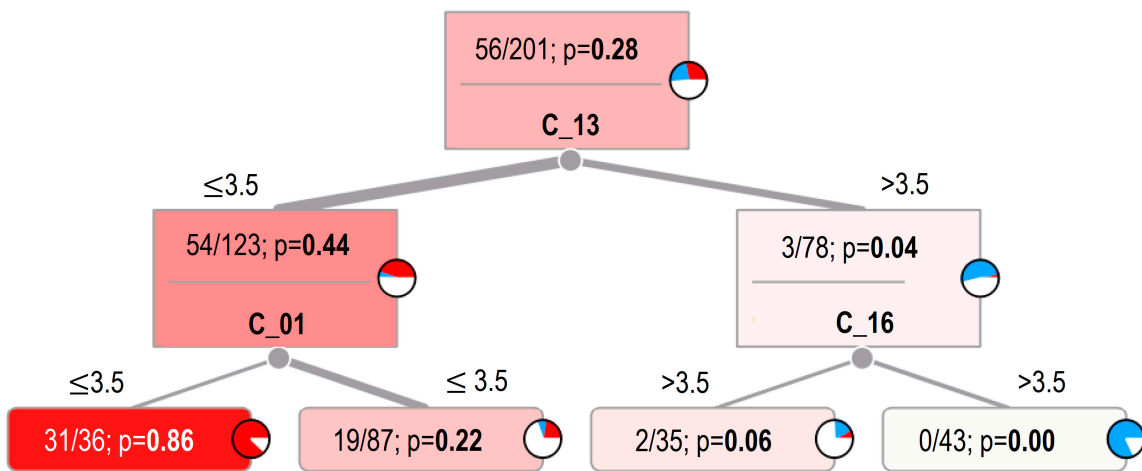| RuleID	QuestionID	Relation	Value	Rows	CountBHV	BHV	p	k
R01	**A_01**	$\geq$	**5**	**63**	**24**	**LowRisk**	**0.38**	**1.51**
R02	A_04	$\leq$	1	26	15	HighRisk	0.58	2.14
R03	A_02	$\leq$	1	25	14	HighRisk	0.56	2.08
R04	A_03	$\leq$	1	20	9	HighRisk	0.45	1.67

Figure 6. For Dataset C.NC, the decision tree visualization of the tree shown in Figure 5. Darker tones of red color and larger red slices denote higher probability of HighRisk behavior. *p*-values are rounded to two significant digits.

Table 6. Category A (worry) questions in Table 5.

QuestionID	QuestionText
A_01	**How concerned are you about yourself being affected by Coronavirus?**
A_02	How concerned are you about your family members being affected by Coronavirus?
A_03	How concerned are you about your close relatives being affected by Coronavirus?
A_04	How concerned are you about your friends being affected by Coronavirus?

Table 7. Selected one-level rules for Category D (demographic) questions.

RuleID	QuestionID	Relation	Value	Rows	CountBHV	BHV	p	k
R37	D_03	=	Nationality1	40	16	LowRisk	0.40	1.58
R38	**D_03**	**=**	**Nationality2**	**33**	**14**	**HighRisk**	**0.42**	**1.58**
R39	D_03	=	Nationality3	35	13	LowRisk	0.37	1.47
R40	D_04	=	Emirate1	106	38	LowRisk	0.36	1.42
R41	D_01	=	No	100	37	HighRisk	0.37	1.37
R42	D_02	=	DidNotTravel	117	39	LowRisk	0.33	1.32
R43	D_08	=	Master	81	27	LowRisk	0.33	1.32
R44	D_02	=	Nationality2	31	11	HighRisk	0.35	1.32

Table 8. Category D (demographic) questions in Table 7.

QuestionID	QuestionText
D_01	Are you resident in the UAE?
D_02	Did you travel outside UAE during the last 6 months?
D_03	**Your Nationality**
D_04	Your current location
D_08	Education

Table 9. Most significant two-level rules for each question and low-/high-risk behavior.

TreeID	Category	Node1	Relation	Value	Node2	Relation	Value	BHV_CLASS	NodeColor	p	k
T020	A	A_05	>	4.335	A_02	>	4.335	LowRisk	Blue	0.44	1.74
T018	**A**	**A_03**	**≤**	**3**	**A_04**	**≤**	**3**	**HighRisk**	**Red**	**0.47**	**1.75**
T101	**B**	**B_03**	**>**	**4.5**	**B_05**	**>**	**4.5**	**LowRisk**	**Blue**	**0.44**	**1.74**
T024	B	B_02	≤	2.5	B_03	≤	2.5	HighRisk	Red	0.45	1.67
T236	C	C_15	>	3.5	C_18	>	3.5	LowRisk	Blue	0.50	1.98
T184	C	C_13	≤	3.5	C_14	≤	3.5	HighRisk	Red	0.50	1.86

Table 10. Questions in Table 9.

QuestionID	Category	QuestionText
A_02	Worry	How concerned are you about your family members being affected by Coronavirus?
A_03	**Worry**	**How concerned are you about your close relatives being affected by Coronavirus?**
A_04	**Worry**	**How concerned are you about your friends being affected by Coronavirus?**
A_05	Worry	How concerned are you about getting hospitalized due to Coronavirus infection?
B_02	Risk Preventive Behavior	How often are you avoiding touching your face, eyes, mouth, and nose?
B_03	**Risk Preventive Behavior**	**How often are you washing your hands with water and soap or sanitizers?**
B_05	**Risk Preventive Behavior**	**How often are you wearing gloves?**
C_13	Risk Reduction Behavior	I eat in dining outlets clearly displaying the required precautionary measures.
C_14	Risk Reduction Behavior	I leave the dining outlet if I do not get the first impression that it is COVID-19-safe
C_15	Risk Reduction Behavior	I leave the dining outlet if I observe that it does not follow the precautionary measures.
C_18	Risk Reduction Behavior	I ask the waiters to keep a reasonable social distance with me

3.6. Steps of the Analytics Methodology

The steps of the methodology corresponding to *data cleaning and preparation*, as shown in Figure 2, are as follows:

1. **Rename the variables;**
2. **Clean the data;**
3. **Reverse the direction** of risk reduction behavior (BHV) questions posed in the opposite direction (C_08 only);
4. **Scale all Likert scale questions** to a 1–5 range (transform values of the questions that are in different ranges);
5. **Calculate total behavioral score** as SUM(C) = SUM(Values Given to All Questions in Category C). In the current research, we only used this simplest scoring method, and there are many possibilities for future research by defining a different scoring function;
6. **Scale the behavioral score** between 0 and 100 to obtain the scaled behavior score (BHV_SCORE). Higher BHV_SCORE values corresponded to lower risk behaviors;
7. **Conduct univariate analysis** of BHV_SCORE to help decide how to discretize BHV_SCORE;
8. **Create class label attributes for** BHV_CLASS, which takes the categorical values of HighRisk, MediumRisk, and LowRisk;
9. **Decide on the cut-off values** (based on approximate top and bottom 25% quartiles) to discretize BHV_SCORE as **class labels** under the categorical response attribute **BHV_CLASS:**
   a. Determine the ~25% quartile value to assign the class label "HighRisk" (Labeling Rule for the data used in this study: IF BHV_SCORE $\geq$ 68.89, THEN **HighRisk**);
   b. Determine the ~75% quartile value to assign the class label "LowRisk" (Labeling Rule for the data used in this study: IF BHV_SCORE $\leq$ 55.56, THEN **LowRisk**);
   c. Classify 25–75% with the class label "MediumRisk" (Labeling Rule for the data used in this study: IF 55.56 < BHV_SCORE < 68.89, THEN **MediumRisk**);
10. **Obtain the "Augmented Dataset",** which is the base for all the generated datasets and analysis;
11. **Create the datasets** in Table 1, where the dataset/analysis coding is based on the independent factors (A–D), and the letters that follow are based on the data type for the factors and response (NN, NC, CN, CC);
12. **Conduct Data Analysis XN:** For A.NN, B.NN, C.NN, and D.CN (Numerical/Categorical Factors, Numerical Response):
   a. Summary statistics;
   b. Ranking.
13. **Conduct Data Analysis XC:** For A.NC, B.NC, C.NC, D.CC, (Numerical/Categorical Factors, Categorical Response):
   a. Ranking;
   b. Rule discovery (one-level);
   c. Rule discovery (two-level).

Details of Steps 12 and 13 of the methodology, corresponding to *data analysis* (Figure 3), are as follows:

12a. Summary Statistics:

For each variable:

   Calculate summary statistics.
   Draw histogram.
   Identify the variables with the highest and lowest mean and dispersion values (standard deviation).

12b. Ranking:

Calculate ranking metrics (univariate regression coefficient and RreliefF).
For Each Metric:

Identify the variables that rank highest for each metric.

13a. Ranking:

Calculate ranking metrics (gain ratio and Gini index).
For Each Metric:

Identify the variables that rank highest for each metric.

13b. Rule Discovery (One-Level):

Generate all possible one-level rules.
Identify and filter interesting rules.
Record the rules in Rule Database R1.

13c. Rule Discovery (Two-Level):

Conduct random forest analysis.
Construct Pythagorean forest.
Identify the most interesting Pythagorean tree(s).
For Each Selected Tree:

Analyze the selected trees with Decision Tree Visualization.
Generate a rule based on the selected tree.
Record the rules in Rule Database R2.

3.7. Techniques Applied

In this section, the specific data analytics techniques selected for, implemented within, and applied as part of the methodology are defined and described. The applied techniques have inherent assumptions, which are discussed as they are each explained.

3.7.1. Summary Statistics

Summary statistics are the first analysis step in almost every data analytics project. The goal of summary statistics is to summarize the data at hand through various metrics, allowing for a glance into the overall characteristics, including the central tendency and dispersion, of data attributes [171] (pp. 37–66). The most popular metrics are the mean (estimated through sample average), median (sample middle value), mode (most common value in the sample), standard deviation (measure of dispersion/variability, estimated through sample standard deviation), and minimum, maximum, and range (calculated as the maximum minus the minimum). The benefit of summary statistics is that they can describe the overall characteristics of the data at hand through easy-to-compute metrics that can readily yield significant insights into the data. These quickly revealed insights can prove highly beneficial in the early stages of a data analytics project and can guide later stages [172].

3.7.2. Ranking

In machine learning, ranking refers to ordering a set of items, typically a set of predictor attributes, based on a score or other metric. Two metrics popular in data analytics/science practice for ranking are the Gini coefficient and the information gain ratio, which are used to accurately assess the relevance and significance of items selected for this research.

The Gini coefficient, originally developed by Corrado Gini [173,174], quantifies the level of dispersion, which is an indicator of inequality. In machine learning, it is typically used to measure the dispersion created by a split in a decision tree, thus measuring the discriminatory ability of a predictive attribute and its value ranges, as well as the predictive power of the attribute.

The information gain ratio, originally developed by Quinlan [175], uses Shannon's entropy, which quantifies the impurity or disorder within a set of data. Specifically, the information gain ratio is the ratio of the splitting information gain to the intrinsic information content/entropy of the split [176].

A viable question is which metrics should be chosen for ranking. In most studies, these two metrics were used together with others (e.g., Ref. [177]). Ref. [178], in a theoretical study that compared the Gini coefficient and information gain ratio, concluded that there was very little disagreement between the two metrics. In a more recent study, Ref. [179] reported that the information gain ratio, in comparison to the Gini coefficient, is "liable to unfairly favor attributes with large numbers of values or categories compared to those with few", thus favoring the Gini coefficient. We included both metrics in our study, yet eventually observed that the ranking outputs were quite similar, confirming the results of Ref. [179].

The assumptions of the two selected ranking metrics deserve discussion, as the metrics can be meaningfully interpreted only if their assumptions are satisfied. The chosen ranking metrics both assume that, at each split, the data can be divided into two subsets/subgroups that are as homogeneous as possible. They both also assume that lower impurity (higher homogeneity) is better for classification within the subsets/subgroups formed after the splits. The two selected ranking metrics, however, differ in the way they measure impurity. They do not require any scale of measurement for the factors and do not require any assumptions about the underlying distribution of the data. Since the Gini coefficient for a split is computed as the weighted sum of the Gini coefficients of the two child nodes, it is assumed in the Gini coefficient that the impurity of a split can be expressed as an additive function of the impurities of the resulting subsets. On the other hand, for gain ratio, the handling of the additive impurity calculation is more complex due to a normalization step.

3.7.3. Decision Tree Analysis

The main methodology used in the present study is the supervised machine learning technique of decision tree analysis [180]. Decision trees are hierarchical tree structures used for the classification of data instances. Each node within a decision tree corresponds to a set of conditions and a subset of the data that satisfies all the conditions reaching that node. Each branch in the tree represents a potential categorical or numerical value range that a child node takes. The classification process begins at the top root node and proceeds by arranging instances according to their respective feature values [181]. There are multiple decision tree algorithms, including CART, CHAID, C4.5, FACT, QUEST, and GUIDE [182].

In a survey of scholarly articles that used machine learning methods, decision tree analysis was the most cited, discussed, and implemented [20]. The most significant advantage of decision tree analysis is that the rules for classification can be stated in natural language and further visualized in a tree. This makes decision tree analysis one of the easiest techniques to comprehend and interpret among all supervised machine learning methods.

Our objective in this research is not to predict the risk behavior class but to create behavioral risk profiles expressed as IF–THEN decision rules. However, while we are not conducting predictive analytics, in numerous benchmark studies comparing classification algorithms, various decision tree algorithms have been reported to perform among the best with respect to classification accuracy and other performance metrics [183,184].

Our research focused on deriving and interpreting only one- and two-level decision trees for ease of interpretability and incorporation into decision-making. Yet, in a seminal paper, Ref. [185] suggests that such simple decision trees readily perform quite well for classification on the most commonly used datasets. Therefore, the rules derived for risk profiling have the potential to exhibit a high level of overall classification performance and, thus, characterization of intrinsic patterns.

3.7.4. Random Forest

Random forest was initially proposed by Tin Kam Ho [186] and Leo Breiman [186] and developed by subsequent researchers [187,188]. It is an algorithm used mainly for classification with high predictive power [189].

The algorithm creates a pool of decision trees for training and combines the predictions of the pool of trees to obtain a more robust prediction with a higher classification accuracy.

The main idea behind the random forest is to represent the collective wisdom of a diverse pool of decision trees in the form of predictive capability, making it an ensemble method. Because the random forest already has many decision trees at its core, it typically performs better than decision trees.

In this research, for each question category (construct), for the dataset with a categorical response (XC), a random forest is algorithmically constructed and is then visualized as a forest of Pythagorean trees, and then selected trees from the forest are visualized as decision trees.

3.7.5. Pythagorean Tree

The Pythagorean (Pythagoras) tree is a method for the hierarchical fractal visualization of a binary decision tree [190,191]. While variations of the basic Pythagorean tree exist, such as the overlap-free Pythagoras Tree [192], we used the Orange software [149], which implemented the original paper [193] on the method.

The Pythagorean tree is constructed by recursively placing squares with certain angles on top of existing trees. Each square denotes a node in the decision tree, which is a subset of data. The percentage of the target class is mapped to the color of the square [194]. For example, in our research, if the target class is HighRisk, then darker tones in red indicate a higher percentage of HighRisk in the data subset represented by that square. The visual objective in analyzing the Pythagorean tree is to search for low-hanging branches with large squares and large color tone deviations. Pythagorean trees may be more suitable than decision tree visualization (introduced next), especially for larger trees that go deeper. Some of the Pythagorean trees generated by the random forest algorithm in our study can be observed in Figure 5.

3.7.6. Decision Tree Visualization

Decision tree visualization visually represents decision trees [193,194]. The trees are generated through decision tree algorithms, including ensemble algorithms such as random forests. The decision tree algorithm partitions the data hierarchically into subsets based on the values or value ranges for selected attributes and one attribute at each level of the decision tree. In decision tree visualization, each node represents a partitioned subset. Each branch represents a split conditioned by the attribute's values at that split. The color tones of the nodes represent a numerical metric, typically the percentage of data points in that subset with the target class value. Furthermore, in some decision tree visualizations, a pie chart for each node represents the distribution of the class attribute values for that subset.

The primary visual objective in analyzing decision tree visualization is to search for noticeable changes in class probabilities in earlier branches in the visualization, as indicated by notable changes in the node colors and pie chart slices. One of the decision tree visualizations constructed in our study can be observed in Figure 6.

4. Results

This section presents selected analysis results for summary statistics, rankings, one-level rules, and two-level rules. The objective is to illustrate the abundance of interpretable and actionable insights that can be obtained through the developed novel analytics methodology. Owing to space considerations, much more extended results are presented in Appendices B–E of the Supplement [11].

4.1. Summary Statistics (Step 12a)

Tables 2 and 3 display some of the summary statistics for Dataset C.NN of Table 1, which corresponds to numerical factors and numerical response, both regarding risk reduction behavior. Specifically, Tables 2 and 3 list the top and bottom five BHV (risk reduction behavior) factors with respect to the highest and lowest values for the two metrics of mean (Table 2) and standard deviation (Table 3). Both tables display the factors with respect to the decreasing values of the metrics.

According to Table 2, risk reduction behavior (BHV) factors 08, 21, 19, 27, and 11 of Category C have the highest mean values, meaning that these are the most frequently practiced behaviors across the sample. The most practiced risk reduction behavior is avoiding dining out with less well-known people.

Since the most frequently practiced risk reduction behavior is avoiding less well-known people and dining out with well-known people, restaurants during pandemics can focus on attracting people together with those they know well. For example, breakfast/lunch/dinner offers/bundles/deals for a la carte and buffet dining can be designed and promoted so as to attract families, co-workers, and close friends.

Other common behaviors include verifying the plate and cutlery, observing waiters wearing masks, using touchless payment means (in the daily language in the UAE, touchless payment methods are referred to as "WiFi"), and selecting outlets that are not crowded. These results are aligned with earlier research: Ref. [195] concludes, in a study conducted within the COVID-19 context, that the dining environment, as well as communication and hygiene, predict the customers' trust perception of the restaurant and their intentions to pay more.

The practical implication of these additional insights obtained from Table 2 is that restaurant staff should be trained to ensure the hygiene of plates and cutlery and adhere to mask-wearing rules during pandemics. Furthermore, reception staff can be trained to take restaurant reservations to balance client traffic throughout the day of the week and hours of the day to reduce overcrowding at the restaurant.

According to Table 3, risk reduction behavior (BHV) factors 25, 18, 20, 24, and 26 of Category C have the highest standard deviation values, and thus the highest level of dispersion. This means that these five behavior patterns are inconsistent across the sample and instead vary considerably. The highest behavioral dispersion was observed in the participants' interaction with waiters, which is the theme of the top four behaviors in Table 3. A practical implication is that the waiters and other staff at the restaurants should be trained very well in communication and problem-solving skills to be able to accommodate the different attitudes, expectations, and behaviors of diners.

A joint consideration of Tables 2 and 3 reveals that question/behavior C_21, which verifies the cleanliness of the plate and the cutlery, appears in both tables, with very high mean and very low standard deviation values. Thus, this is very frequent behavior consistently exhibited by restaurant diners/patrons. The immediate practical implication for restaurant managers is that they should make sure to always keep the plates and cutlery visibly clean, and the staff should be trained to pay the highest attention to achieving, observing, and maintaining this expectation.

4.2. Ranking (Step 12b)

Table 4 displays, this time for Dataset C.NC, the ranking of the risk-related factors (Category C) with respect to determining the categorical BHV_CLASS. Here, the response variable (target attribute) is categorical, as denoted by the second "C" in "C.NC".

According to Table 4, risk reduction behavior (BHV) factors 13, 16, 14, 17, and 12 of Category C have the highest RReliefF scores. This means these factors (attributes) have the highest predictive power for BHV_CLASS. In other words, when applying predictive supervised machine learning techniques for classification, an analyst can aim for higher predictive accuracies in predicting BHV_CLASS by using these top-ranking factors before others.

Table 4 reveals that the top three factors are related to restaurants following cautionary measures, and the latter two are related to the seclusion perception of the diner.

The ranking analysis does not specify how (positively or negatively, and to what magnitude) these factors affect BHV_CLASS. This insight can only be obtained by applying predictive analytics using machine learning techniques. Predictive analytics is not included within the scope of the current research because abundant results have already been obtained with the current descriptive/diagnostic state of the developed methodology. Still,

sample results for ranking are provided because these results suggest which questions should be included in surveys with higher priority than other questions.

4.3. One-Level Rule Discovery (Step 13b)

4.3.1. Definition of One-Level Rules

The following analysis involved the discovery of one-level rules in the form "**IF** Condition **THEN** BHV_CLASS". These rules are referred to as "one-level", because the behavior class outcome (the premise/consequent in the rule coming after "THEN") depends only on a single condition (the antecedent/expression between "IF" and "THEN").

While many possibilities can be generated, we focused on the rules that revealed the conditions underlying the most drastic changes in the behavior class. The metric used to characterize the degree of change in the behavior class was defined as "k", a multiplier representing the extent to which a behavior class is observed compared to the default frequency. The definition of this multiplier is similar to the definition of the lift metric in association mining [196]. More specifically, for the rule "**IF** X **THEN** Y":

- k is defined as the ratio $k = p/p0$;
- p is the probability of observing the behavior class BHV_CLASS value Y given that the condition value is X;
- $p0$ is the default probability of observing the same BHV_CLASS value in the complete sample (assumed to represent the population).

4.3.2. Sample One-Level Rules

Sample results for one-level rules are presented in Tables 5–8, from among the one-level rules that yield the largest k-values. Tables 5 and 7 present the top one-level rules for question categories A and D, respectively, and Tables 6 and 8 list the questions referred to in Tables 5 and 7, respectively.

For example, Table 5 displays the most significant rules for Category A, where questions with the highest k-values are shown in bold. Table 6 lists the questions referred to in Table 6.

Let us next illustrate how Tables 5 and 6 would be interpreted with an example.

4.3.3. Sample One-Level Rule R01 for Worry

The sample results for one-level rules are presented in Tables 5–8. The first rule in Table 5, namely Rule **R01**, is directly read as

"**IF** *A_01 ≥ 5* **THEN** *LowRisk; p = 0.38, k = 1.51*"

Thus, the one-level rule, which conditions on question A_01 within worry (Category A), can be expressed in natural language as follows:

"If the answer of a respondent to question A_01 (level of concern about a person's own self being affected by Coronavirus) is greater than or equal to 5, then, with a probability of p = 24/63 = 0.3809, that respondent exhibits low-risk behavior with respect to the COVID pandemic. Compared to the complete sample, where p0 = 76/301 = 0.2525 for low-risk, this respondent is k = 0.3809/0.2525 ≅ 1.51 times as likely to exhibit low-risk behavior. In other words, for a respondent who gives such an answer, there is a 51% higher chance for her/him (compared to the whole) to exhibit low-risk behavior."

4.3.4. Sample One-Level Rule R38 for Demographics

Table 7 displays the most significant rules for question/behavior Category D, where the questions with the highest k-values are shown in bold. Table 8 lists the questions listed in Table 7.

The information provided in Tables 7 and 8 makes it possible to relate behavior classes to demographic attributes. The true values of the demographic attributes have been masked to respect confidentiality concerns. Let us illustrate how Tables 7 and 8 can be interpreted using the following example:

The second rule in Table 7, Rule **R38,** is directly expressed as

"**IF** D_03 = *Nationality2* **THEN** *HighRisk; p = 0.42, k = 1.58*"

This one-level rule, which conditions demographics (Category D), can be expressed in natural language as follows:

"If the answer of a respondent to question D_03 (nationality) is equal to "Nationality2", then, with a probability of p = 14/33 = 0.4242, that respondent exhibits high-risk behavior with respect to the COVID pandemic. Compared to the complete sample, where p0 = 81/301 = 0.2691 for high-risk, this respondent was k = 0.4242/0.2691 = 1.58 times as likely to exhibit high-risk behavior. In other words, for a respondent who gives such an answer, there is a 58% higher chance for her/him (compared to the whole) to exhibit high-risk behavior."

As can be observed from the above natural language expression of the rule, this rule can be used to identify which nationalities may exhibit high-risk behaviors. Identifying such populations can help shape policies regarding flight rules and health screening tests at airports, reducing the risk of disease transmission before entering the UAE. Furthermore, training for precautionary measures can be developed in the languages of these populations to increase awareness and levels with other populations.

4.4. Two-Level Rule Discovery (Step 13c)

4.4.1. Definition of Two-Level Rules

The subsequent analysis was the discovery of two-level rules in the form "**IF** Condition1 **and** Condition2 **THEN** BHV_CLASS". These rules are referred to as "two-level" because the behavior class outcome (the premise/consequent in the rule coming after "THEN") depends on two conditions (the antecedent/expressions between "IF" and "THEN"), which would be visualized with a two-level deep decision tree.

While many possible rules can be generated, we focused on the rules that revealed the conditions underlying the most drastic changes in the behavior class and with at least eight observations in the selected leaf node (the node that does not branch further). The metric used to characterize the degree of change in the behavior class was again "k", the multiplier representing how much more a behavior class is observed compared to the default probability in the complete sample. However, there was a slight modification to this definition. More specifically:

- k is still defined as the ratio k = $p/p0$, but this time for a two-level rule in the form "**IF** X1 **and** X2 **THEN** Y";
- p is the probability of observing the behavior class BHV_CLASS value Y under conditions X1 **and** X2 being satisfied simultaneously;
- $p0$ is the default probability of observing the same BHV_CLASS value in the sample (which is assumed representative of the population).

The sample results for the two-level rules are presented in Tables 9 and 10 for the two-level rules that yield the largest k-values. Table 9 presents the top two-level rules for all question categories A–D. Table 10 lists the questions listed in Table 9.

4.4.2. Generation of Two-Level Rules Through Decision Tree Analysis

Two-level rules were generated using a method different from the generation of one-level rules. The process is based on decision tree analysis, which consists of successively applying the four steps of (1) random forest generation, (2) Pythagorean tree visualization, (3) decision tree visualization, and (4) rule identification, as follows:

- Decision tree analysis for generating the two-level rules for XC (data with categorical response) was initiated by first conducting random forest analysis, where BHV_CLASS was predicted using BHV factors.
- The trees in the forest were then visualized using Pythagorean trees (Figure 5).

- Next, the Pythagorean trees with the most significant change in the target color (red for identifying HighRisk; blue for LowRisk) in the branches were filtered, and decision trees (Figure 6) were drawn for each filtered tree.
- Finally, the two-level rules visualized in each decision tree were explicitly identified and recorded in Rule Database R2, which contained the most significant two-level rules.

4.4.3. Random Forest

A separate random forest was generated for each (BHV_CLASS) combination. For example, Figures 5 and 6 show the results from the random forest generated for Category D (BHV_CLASS = HighRisk). The source screenshots from the Orange software for Figures 5 and 6 have been edited for clear visual communication to obtain the shown final figures. The original Orange screenshots for the generated rules can be obtained from the authors upon request as proof of analysis.

Refs. [197–200] guided the choice of parameters for the random forests in our study. Each random forest included 60 trees, with six attributes considered in each split. Because the focus was only on two-level rules, the depth of the trees was set to 2. Only subsets with >20 observations were split at each level.

4.4.4. Pythagorean Tree

As shown in Figure 5, the results of the random forest analysis were visualized using Pythagorean forest. Figure 5 displays 15 of the 60 trees generated in our analysis using the random forest algorithm for only a single category of questions, namely those questions in Category C. In Figure 5, BHV_CLASS = HighRisk is the target behavior class. Among the different Pythagorean trees within the forest, the most interesting tree(s) were selected for successive analysis as decision tree(s).

The primary interest is towards Pythagorean trees that result in the largest changes in the distribution of classes, which are reflected in notable changes in the color of branches, especially for the target value of interest (HighRisk, in this case, shown with tones of red color). Color changes at the immediate lower branches of Pythagorean trees are preferred, rather than changes farther in higher branches.

In Figure 5, one of the Pythagorean trees is highlighted (third row and fourth column) with a shaded background. For illustration, this tree was selected for successive analysis using decision tree visualization (Figure 6). This tree was chosen for illustration because the Pythagorean tree visualization for this tree suggests a clear classification immediately at immediate lower branches for HighRisk (red color), without having to browse to farther higher branches in the tree. There was a significant change in color tone on the leftmost branch of the tree, and the tone at the end of the left branch was dark red.

4.4.5. Decision Tree Visualization and Sample Two-Level Rule for Risk Reduction Behavior

The Pythagorean tree selected earlier in Figure 5 is visualized as a decision tree in Figure 6. This decision tree was visually analyzed to obtain two-level classification rules that could be recorded in Rule Database R2 for two-level rules. The primary interest lies in rules that result in the largest changes in the distribution of classes, which can be detected through large changes in the colors of the boxes and the relative sizes of pie chart slices, especially for the target value of interest (HighRisk, in this case).

When constructing the decision visualizations, the following settings were chosen: depth: four levels; edge width: relative to root; target class = HighRisk or LowRisk; show details on non-leaves.

The root node of the tree (box on the top) indicates 201 observations in the analyzed tree. This is a subset of the sample with 301 observations. Of the 201 respondents, 56 were classified as HighRisk (red slice in the pie chart), with a ratio/probability of $p = 56/201 \cong 0.28$. This is very close to the ratio of HighRisk in the sample, which was $81/301 = 0.2691$. If the answer to C_13 is ≤ 3.5, then the branch to the left of the root node

is traversed, and the ratio/probability of HighRisk becomes $p = 54/123 \cong 0.44$. There is already a significant increase in the ratio/probability of HighRisk, which is an interesting pattern even at the first level of decision tree visualization (and the one-level rule that explains the first level).

There is a consideration regarding the results of Orange and how they are reflected in the paper: In Orange, rather than probability values, percentage values were displayed. Furthermore, the calculation of the percentages in Orange assumed a numerator value +0.5 higher than the actual value. In the paper, in Figure 6, instead of percentages, probabilities are shown. Furthermore, the p-values are based on directly dividing the original numerator value (number of observations in the subset with BHV_CLASS = High-Risk) by the value in the denominator (number of observations in the subset). These choices were made to achieve consistency with the earlier one-level rule results and clarity of communication.

Since our analysis aims to identify two-level rules, we dive one level deeper and branch again from the current node; if we branch out to the left from the last-mentioned node, then there is an even more significant pattern. From among the 123 respondents in the subset, if the answer to question C_01 is also ≤3.5, then the ratio of HighRisk (red slice in the pie chart) significantly increases, reaching a much higher ratio/probability of $p = 31/36 = 0.8611\%$. This probability is k = $0.8611/0.2691 \cong 3.20$ times higher than what is observed in the complete sample. This stark increase is even more interesting, yielding the following two-level rule:

"**IF** *C_13 ≤ 3.5 **and** C_01 ≤ 3.5 **THEN** HighRisk; p = 0.86, k = 3.20*"

which can be expressed in natural language as follows:

"If the answer of a respondent to question C_13 (eating in dining outlets that clearly display the required precautionary measures) is less than or equal to 3.5, and, furthermore, if the answer of the same respondent to question C_01 (selecting dining outlets that offer healthier food) is also less than or equal to 3.5, then the following can be stated: With a probability of 0.86, that respondent exhibits high-risk behavior with respect to the COVID pandemic. Compared to the complete sample, this respondent is 3.09 times more likely to exhibit high-risk behavior. In other words, for a respondent who gives such an answer, there is a 209% higher chance for her/him (compared to the whole population) to exhibit high-risk behavior."

Therefore, this decision tree (Figure 6) resulted in a two-level rule that suggests one reason for the significant increases in high-risk behavior. In other words, it gave us a risk behavior profile described by a two-level rule.

4.4.6. Sample Two-Level Rules

Table 9 displays the most significant two-level rules for all categories with the highest k-values, which were filtered for each category (BHV_CLASS). Table 10 lists the questions listed in Table 9.

Let us next illustrate how Tables 9 and 10 would be interpreted with two examples, namely rules T018 and T101.

4.4.7. Sample Interpretation of Two-Level Rules: T018

The second rule listed in Table 5, namely Rule **T018**, is directly read as

"**IF** *A_03 ≤ 3 **and** A_04 ≤ 3 **THEN** HighRisk; p = 0.47, k = 1.75*"

This two-level rule, which conditions on worry (Category A), can be expressed in natural language as follows:

"If the answer of a respondent to question A_03 (concern about close relatives being affected by Coronavirus) is less than or equal to 3, and, in addition, if the answer of the same respondent to question A_04 (concern about your friends being affected by Coronavirus) is also less than or equal to 3, then the following can be stated: With a

probability of 0.47, that respondent exhibits high risk behavior with respect to the COVID pandemic. Compared to the complete sample, this respondent is 1.75 times as likely to exhibit high-risk behavior. In other words, for a respondent who gives such an answer, there is a 75% higher chance for her/him (compared to the whole population) to exhibit high-risk behavior."

4.4.8. Sample Interpretation of Two-Level Rules: T101

The third rule in Table 5, namely Rule **T101**, is directly read as

"**IF** *B_03 > 4.5* **and** *B_05 > 4.5* **THEN** *LowRisk; p = 0.44, k = 1.74*"

This two-level rule, which conditions on risk preventive behavior (Category B), can be expressed in natural language as follows:

"If the answer of a respondent to question B_03 (frequency of washing hands with water and soap or sanitizers) is greater than 4.5, and, in addition, if the answer of the same respondent to question B_05 (frequency of wearing gloves) is also greater than 4.5, then the following can be stated: With a probability of 0.44, that respondent exhibits low risk behavior with respect to the COVID pandemic. Compared to the complete sample, this respondent is 1.74 times as likely to exhibit low-risk behavior. In other words, for a respondent who gives such an answer, there is a 74% higher chance for her/him (compared to the whole population) to exhibit low-risk behavior."

4.4.9. Discussion

Our study analyzed all two-level trees obtained from the random forest analysis using the set parameters. The analysis resulted in 166 notable two-level rules, fully presented in Appendix E of the Supplement document [11].

The different one- and two-level rules identified and described in this section provide different risk profiles, each of which can be a target for customized training programs, which can be rapidly generated with novel AI and other information technologies, such as generative AI [201–203].

4.5. Web-Based Analytics Dashboard

The risk profile rules obtained through one-level and two-level rule discovery can be used to construct interactive analytics dashboards, that can facilitate planning and decision-making in healthcare management [204]. To this end, a web-based visual analytics dashboard has been developed [10] using the Tableau Public platform [205]. Compared to tabular data representation, such dashboards are much easier to browse through data, and comprehend and derive insights, hence the motivation for developing the dashboard.

4.5.1. Dashboard Design

The developed analytics dashboard [10] displays each rule as a circular glyph, with the k-value on the x-axis and QuestionID on the y-axis. Furthermore, the color of each circle tells whether the rule is for profiling HighRisk or LowRisk people. The dashboard is interactive: when the user holds the mouse over a circle, information about that rule is displayed. The user would especially be interested in rules with highest or lowest k-values. Rules for a particular question are all horizontally aligned on a horizontal line corresponding to that question on the y-axis.

4.5.2. Sample One-Level Rule R01 for Risk Reduction Behavior

As an example rule in the dashboard, the mouse held over the rightmost red circle for question C_01 would have a pop-up tooltip displays the following rule:

"**IF** *C_01 ≤ 1* **THEN** *HighRisk; p = 0.81, k = 3.02*"

where C_01 is "I select dining outlets offering healthier food."
This rule can be expressed in natural language as follows:

"If the answer of a respondent to question C_01 (selecting dining outlets offering healthier food) is less than or equal to 1, then, with a probability of p = 0.81, that respondent exhibits high-risk behavior with respect to the COVID pandemic. Compared to the complete sample, where p0 = 81/301 = 0.2691, this respondent is k = 0.81/0.2691 ≅ 3.02 times as likely to exhibit low-risk behavior. In other words, for a respondent who gives such an answer, there is a 202% higher chance for her/him (compared to the whole) to exhibit high-risk behavior."

5. Discussion

Overall, as illustrated by the sample analysis results, using data analytics contributed to the identification of the most/least adopted health risk reduction behaviors throughout the dining journey of restaurant diners/patrons. To our knowledge, this is the only exhaustive empirical study dedicated to customer behavioral strategies in the food service industry during COVID-19. Measuring the customer perception, Ref. [140] found that the restaurant dining environment, communication, cleanliness, and contactless devices are the main restaurant features contributing to diners' trust in restaurants. This supports our finding, as among the most adopted behaviors are "I verify if the plate and the table cutlery are clean", "I observe if the waiters are constantly wearing masks", and "I use WiFi payment means". Furthermore, the two-level rule discovery helped the authors to find that when restaurant diners/patrons rate low "eating in dining outlets that clearly display the required precautionary measures" and "selecting dining outlets that offer healthier food", they have a high probability of exhibiting high-risk behavior with respect to the COVID pandemic.

As explained in the literature, worry has a significant influence on tourist behaviors. Contrary to other studies considering the effect of worry as the combination of the different items measuring it, our study, based on the analytics adopted, identified the item among the seven of the scale explaining the high low-risk score. This article identified that if the item of "level of concern about a person's own self being affected by Coronavirus" is high, then respondents are inclined to exhibit low-risk behavior with respect to the COVID-19 pandemic. This result has practical added value compared to considering worry as an entire construct. It allows for professionals to design and plan actions to raise individual's concerns about their own safety and accordingly encourage preventive behaviors.

Considering different countries, Ref. [91] found that although the populations corresponding to each country perceived coronavirus as a high risk, UK nationals had the highest level of concern about COVID-19. This result sustains the role of nationality in predicting individuals' behavior. Accordingly, our study used the one-level rule discovery to conclude that restaurant diners/patrons of "Nationality2" would exhibit high-risk behavior. This is again a very practical result that would help governments to develop customized and, accordingly, more effective communication for specific communities in order to better manage health crises.

The results above confirm the usefulness of using data analytics in tourism to provide practical and actionable results, which are keys for better preparedness for future health crises and timely decisions in the management of future similar pandemics.

Although the presented analysis and results illustrate the applicability of the developed methodology and approach, there are multiple threats to the validity of the research, which are discussed next.

1. The first threat to validity is the sample coming from a single country, namely the United Arab Emirates (UAE), which is uniquely different from all other countries in the world in the sense of having the highest percentage of expatriates. Thus, the results may not generalize to other countries, including those in the Gulf region or the larger Middle East North Africa (MENA) region. However, it is also valuable and important to analyze this unique country because it has the highest percentage of expatriates. Furthermore, the UAE was one of the countries that managed the

COVID-19 pandemic in the most professional and coordinated way, making it a valuable and significant choice.

2. A second threat was mentioned in an earlier published work [9] in the same research stream: There may be other constructs and/or measurement items that may affect risk reduction behavior and risk profiles, which are much more effective and influential than the ones chosen. This will be a topic for future research.

3. A third item is the following: The sample size is 301 valid observations. This may be considered a small sample to capture the diners'/patrons' behaviors during COVID-19. In addition to what have been presented as justifications regarding measures taken by researchers to mitigate issues related to sample representativeness (Section 3.2), the authors would like to support the sufficiency of the sample size for the purpose of this research. First, as confirmed by [206–208] and Ref. [146], it is particularly challenging in tourism research to ensure the perfect population size, to identify a random and representative sample, and to compute beforehand an optimal sample size. Still, ref. [209] stated that there are some rules of thumb that have been used by researchers to determine a sample size. Based on the experience of the authors of [209], "a sample between 160 and 300 valid observations is well suited for multivariate statistical analysis techniques (e.g., CB-SEM, PLS-SEM) most of the time" (p. xiv). Due to the above listed reasons, the sample size of 301 of the present study could be considered sufficient for this research. At the same time, it is recommended to test the presented novel methodology on a bigger sample to re-validate its robustness.

4. A fourth threat is the following: As mentioned in Section 3.2, different question categories were taken from different inventories in earlier research, and their scales were not the same. They were eventually scaled to a 1–5 Likert scale, resulting in different numerical values in the rules, rather than only integer 1–5 Likert scale values. A solution to this problem could be to use inventories/constructs that are aligned/consistent in terms of having the same scale (e.g., 1–5 Likert scale). While Likert scaling did result in non-integer values in the rules, this does not affect the main theoretical contribution of the study, which is the development of a novel analytics methodology that yielded novel types of insights for the domain that were not previously given in the literature.

5. A fifth threat to validity concerns internal validity. As mentioned in Section 3.2, the Cronbach alpha value was low for only a single construct due to a single question, B_05, which is less than the popular threshold of 0.7 for the Cronbach alpha metric. As a counterargument, there are multiple reasons why this may not pose a serious threat to the validity or consistency of the study. First, the value of 0.62 is not too low, still close to the recommended value of 0.70. Second, our research is not applying SEM or its variants; hence, the Cronbach alpha value is not as important as would be if SEM was applied. Third, the constructed trees are for each category of questions; hence, the results for Category B do not affect the results for the other categories. Fourth, while the data analysis can be conducted without that question, the same consistency can be achieved by ignoring any rules that include that question.

6. A sixth item could be related to the use of the snowball sampling technique, a non-probabilistic method, which raises the external validity challenge. The discussions on the validity of the snowball technique, in the context of the selected domain and research question, can be carried out through the following steps:

   a. The snowball sampling technique uses respondents/participants to recruit new respondents from their network, such as friends, acquaintances, and workmates [210];

   b. This method can especially be used when the targeted population is unknown, inaccessible, or hard to reach. This was the case as it was not randomly possible to survey patrons directly when dining out during the COVID-19 pandemic (see Section 3.2: Data);

      c.   As shared earlier in Section 3.3: Data Validity, Ref. [97] asserted the infeasibility of identifying a random and representative sample of tourists and entertainment patrons, as this population is not a well-defined group and has a changing profile over time. This is even more true in the case of this study led during the unprecedented COVID-19 pandemic, as no information is available worldwide, not only in the UAE, about the characteristics of the population who frequently visited restaurants;

      d.   To hedge against the possible biases and risks of the snowball technique, and to increase the randomness of the sample, the authors involved different social networks in the survey. When selecting the first level of the networks, a relatively large group of 40 participants was formed. These initial participants were selected considering the composition of nationalities representing the UAE population as well as the nationality of tourists visiting the UAE. Therefore, the initial layer/level/cohort of participants can be argued to be a good representation of the different types of tourists and residents by nationality. This kind of multiple snowball sampling, also called the chain of referrals, is cautiously meshed, allowing for the formation of a sample that could be closely similar to a representative sample of the study group;

      e.   In addition to the careful selection of the initial layer of respondents, the survey included two screening questions that prevented ineligible individuals from participating. These ineligible individuals were those located outside the UAE at the time of survey and those who did not visit a restaurant in the UAE within the last 2–3 weeks.

      f.   Finally, the survey also prevented respondents from the same IP from submitting survey answers more than once.

      g.   For evaluating the validity of the snowball technique further, as it was applied in this research, readers can refer to discussions and guidelines for sampling in Refs. [210–212].

7.   Last, but not least, a seventh threat to the validity of the research is the simple scoring method that was used to compute BHV_SCORE. An unweighted summation is simplistic and lacks theoretical rigor. Furthermore, because it treats all measurement items of the same importance, the method is most likely not the best scoring method for a multitude, if not the majority, of scenarios or cases. However, this method is also the first scoring method that would be considered and implemented by practitioners, at least as a default benchmark method. Thus, it is important to investigate this scoring method. Future research can use other scoring methods and algorithms, including methods adopted from other domains, such as financial risk scoring [213].

6. Conclusions

This paper presented a novel data analytics methodology for discovering behavioral risk profiles in the context of diners during a pandemic. Furthermore, the applicability of the methodology was illustrated through survey data collected in the United Arab Emirates (UAE) during the COVID-19 pandemic. The scope of the methodology and presented case study is descriptive and diagnostic analytics, focusing on different risk profiles.

The developed methodology analyzes a survey dataset by transforming it into a collection of datasets consisting of both numeric and categorical responses. The objective of the methodology is to gain insights into the behavior of individuals about risk reduction and preventive measures through risk profiling. This is achieved through the exhaustive identification of behavioral risk profiles, expressed in terms of one-level and two-level rules. The methodology combines various statistical methods, business intelligence, data visualization, and machine learning to analyze the data at hand. Rather than having a single model, such as SEM, that investigates the relations between the constructs for the whole sample, the approach followed in our research aims to identify the multitude of different risk profiles.

The case study was conducted in the United Arab Emirates, a country known for its diverse population and thriving tourism and hospitality sectors. In addition to summary statistics and ranking, the most significant one- and two-level rules are obtained from the analysis, which creates a diverse collection of profiles for high-risk and low-risk behavior. Research findings revealed insights into the factors influencing risk reduction behavior, shedding light on the interplay between psychological factors and risk reduction behavior. Notably, the study highlighted the significance of restaurants' adherence to cautionary measures and diners' perception of seclusion. These factors emerged as key predictors of risk reduction behavior, offering valuable guidance for developing managerial strategies and skill development programs to promote safer dining experiences during the pandemic.

Overall, the present study contributes to the field of data analytics in hospitality by providing a practice-oriented integrated methodology for understanding the behavioral risk profiles of diners/patrons during a pandemic. The study reveals the skills required to equip human capital in the hospitality sector to accommodate possible future pandemics. The study also provides examples of the policies and practices that can be adopted and the competencies to be developed by businesses and government entities.

The main theoretical contribution of the present research study is the custom-developed data analytics methodology, which can be directly applied to any similar data, regardless of geographic region or demographic attributes. While the methodology carries out the analysis independent of any such attributes, it reveals the hidden implicit patterns explicit. The patterns relate to the mentioned attributes and others in the form of one-level and two-level rules. These rules unearth the hidden patterns and state them as risk profiles. Therefore, the methodology itself and the insight types (rules and rankings) are independent from geographic or demographic characteristics, while the specific results that the methodology generates yield insights that relate to these and other characteristics of the chosen domain and targeted population (restaurant diners/patrons in the UAE during the COVID-19 pandemic).

The methodology is designed for behavioral risk profiling, specifically in the context of the hospitality sector and during a pandemic. Yet, as an avenue for future research, it is possible to adapt the methodology to new domains and study fields.

References

1. Karl, M.; Muskat, B.; Ritchie, B.W. Which travel risks are more salient for destination choice? An examination of the tourist's decision-making process. *J. Destin. Mark. Manag.* **2020**, *18*, 100487. [CrossRef]
2. Tasci, A.D.; Sönmez, S. Lenient gun laws, perceived risk of gun violence, and attitude towards a destination. *J. Destin. Mark. Manag.* **2019**, *13*, 24–38. [CrossRef]
3. Lee, W.; Park, S.; Jeong, C. Repositioning risk perception as a necessary condition of travel decision: The case of North Korea tourism. *J. Hosp. Tour. Manag.* **2022**, *52*, 252–263. [CrossRef]
4. Donaire, J.A.; Galí, N.; Camprubi, R. Empty Summer: International Tourist Behavior in Spain during COVID-19. *Sustainability* **2021**, *13*, 4356. [CrossRef]
5. Cookson, C. Scientists in Race to Protect Humanity from Future Pandemics. Financial Times 2020. Available online: https://www.ft.com/content/8521d81e-1c0f-11ea-81f0-0c253907d3e0 (accessed on 24 November 2023).
6. Sigala, M. Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *J. Bus. Res.* **2020**, *117*, 312–321. [CrossRef]
7. Mariani, M.; Baggio, R.; Fuchs, M.; Höepken, W. Business intelligence and big data in hospitality and tourism: A systematic literature review. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 3514–3554. [CrossRef]
8. Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. *Tour. Manag.* **2018**, *68*, 301–323. [CrossRef]
9. Labben, T.G.; Chen, J.S.; Kim, H. Factors Shaping Diner's COVID-19 Preventive Behavior: A Case Study in the United Arab Emirates. In *Advances in Hospitality and Leisure*; Chen, J.S., Ed.; Emerald Publishing Limited: Bingley, UK, 2023; Volume 18, pp. 17–35. [CrossRef]
10. Tableau Public. Rule Database 1 by GE. Available online: https://public.tableau.com/app/profile/gurdal.ertek/viz/RuleDatabase1/Sheet1 (accessed on 14 October 2024).
11. Supplement. Supplement to "A Novel Data Analytics Methodology for Discovering Behavioral Risk Profiles: The Case of Diners During a Pandemic". 2023. Available online: https://ertekprojects.com/ftp/supp/17.zip (accessed on 25 November 2023).
12. Floyd, D.L.; Prentice-Dunn, S.; Rogers, R.W. A meta-analysis of research on protection motivation theory. *J. Appl. Soc. Psychol.* **2020**, *30*, 407–429. [CrossRef]
13. Janz, N.K.; Becker, M.H. The health belief model: A decade later. *Health Educ. Q.* **1984**, *11*, 1–47. [CrossRef]
14. Tsai, C.W.; Lai, C.F.; Chao, H.C.; Vasilikos, A.V. Big data analytics: A survey. *J. Big Data* **2015**, *2*, 21. [CrossRef]
15. Runkler, T.A. *Data Analytics*; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2020.
16. Gartner. *What Is Data and Analytics?* Available online: https://www.gartner.com/en/topics/data-and-analytics (accessed on 24 November 2023).
17. O'Leary, D.E. The impact of Gartner's maturity curve, adoption curve, strategic technologies on information systems research, with applications to artificial intelligence, ERP, BPM, and RFID. *J. Emerg. Technol. Account.* **2009**, *6*, 45–66. [CrossRef]
18. Eriksson, T.; Bigi, A.; Bonera, M. Think with me, or think for me? On the future role of artificial intelligence in marketing strategy formulation. *TQM J.* **2020**, *32*, 795–814. [CrossRef]
19. Lepenioti, K.; Bousdekis, A.; Apostolou, D.; Mentzas, G. Prescriptive analytics: Literature review and research challenges. *Int. J. Inf. Manag.* **2020**, *50*, 57–70. [CrossRef]
20. Alloghani, M.; Al-Jumeily, D.; Mustafina, J.; Hussain, A.; Aljaaf, A.J.A. systematic review on supervised and unsupervised machine learning algorithms for data science. In *Supervised and Unsupervised Learning for Data Science*; Springer: Cham, Switzerland, 2020; pp. 3–21. [CrossRef]
21. Romero, C.; Ventura, S. Educational data mining and learning analytics: An updated survey. *WIREs Data Min. Knowl. Discov.* **2020**, *10*, e1355. [CrossRef]
22. Jahanbakht, M.; Xiang, W.; Hanzo, L.; Azghadi, M.R. Internet of underwater things and big marine data analytics—A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 904–956. [CrossRef]
23. Maier-Hein, L.; Eisenmann, M.; Sarikaya, D.; März, K.; Collins, T.; Malpani, A.; Fallert, J.; Feussner, H.; Giannarou, S.; Mascagni, P.; et al. Surgical data science–from concepts toward clinical translation. *Med. Image Anal.* **2022**, *76*, 102306. [CrossRef]
24. Kamble, S.S.; Gunasekaran, A.; Gawankar, S.A. Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. *Int. J. Prod. Econ.* **2020**, *219*, 179–194. [CrossRef]
25. Ertek, G.; Kailas, L. Analyzing a decade of wind turbine accident news with topic modeling. *Sustainability* **2021**, *13*, 12757. [CrossRef]
26. Wang, J.; Xu, C.; Zhang, J.; Zhong, R. Big data analytics for intelligent manufacturing systems: A review. *J. Manuf. Syst.* **2022**, *62*, 738–752. [CrossRef]
27. Çinicioğlu, E.N.; Ertek, G.; Demirer, D.; Yörük, H.E. A framework for automated association mining over multiple databases. In Proceedings of the 2011 International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, Turkey, 15–18 June 2011; pp. 79–85. [CrossRef]
28. Akter, S.; Michael, K.; Uddin, M.R.; McCarthy, G.; Rahman, M. Transforming business using digital innovations: The application of AI, blockchain, cloud and data analytics. *Ann. Oper. Res.* **2022**, *308*, 7–39. [CrossRef]
29. Mariani, M.M.; Perez-Vega, R.; Wirtz, J. AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychol. Mark.* **2022**, *39*, 755–776. [CrossRef]

30. Verdenhofs, A.; Tambovceva, T. Evolution of customer segmentation in the era of Big Data. *Mark. Manag. Innov.* **2019**, *1*, 238–243. [CrossRef]

31. Miah, S.J.; Vu, H.Q.; Gammack, J.; McGrath, M. A big data analytics method for tourist behaviour analysis. *Inf. Manag.* **2017**, *54*, 771–785. [CrossRef]

32. Liu, Y.Y.; Tseng, F.M.; Tseng, Y.H. Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model. *Technol. Forecast. Soc. Change* **2018**, *130*, 123–134. [CrossRef]

33. Centobelli, P.; Ndou, V. Managing customer knowledge through the use of big data analytics in tourism research. *Curr. Issues Tour.* **2019**, *22*, 1862–1882. [CrossRef]

34. Lin, M.S.; Liang, Y.; Xue, J.X.; Pan, B.; Schroeder, A. Destination image through social media analytics and survey method. *Int. J. Contemp. Hosp. Manag.* **2021**, *33*, 2219–2238. [CrossRef]

35. Chen, C.; Ma, J.; Susilo, Y.; Liu, Y.; Wang, M. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp. Res. Part C Emerg. Technol.* **2016**, *68*, 285–299. [CrossRef]

36. Bao, J.; Zheng, Y.; Wilkie, D.; Mokbel, M. Recommendations in location-based social networks: A survey. *Geoinformatica* **2015**, *19*, 525–565. [CrossRef]

37. Cheng, M.; Edwards, D. Social media in tourism: A visual analytic approach. *Curr. Issues Tour.* **2015**, *18*, 1080–1087. [CrossRef]

38. Khotimah, H.; Djatna, T.; Nurhadryani, Y. Tourism recommendation based on vector space model using composite social media extraction. In Proceedings of the 2014 International Conference on Advanced Computer Science and Information System, Jakarta, Indonesia, 18–19 October 2014; pp. 303–308. [CrossRef]

39. Kurashima, T.; Iwata, T.; Irie, G.; Fujimura, K. Travel route recommendation using geotagged photos. *Knowl. Inf. Syst.* **2013**, *37*, 37–60. [CrossRef]

40. Xiang, Z.; Fesenmaier, D.R. *Analytics in Smart Tourism Design: Concepts and Methods*; Springer International Publishing: Cham, Switzerland, 2017.

41. Mariani, M. Big data and analytics in tourism and hospitality: A perspective article. *Tour. Rev.* **2020**, *75*, 299–303. [CrossRef]

42. Li, D.; Yang, Y. GIS monitoring of traveler flows based on big data. In *Analytics in Smart Tourism Design: Concepts and Methods*; Xiang, Z., Fesenmaier, D.R., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 111–126. [CrossRef]

43. Park, S.B.; Kim, J.; Lee, Y.K.; Ok, C.M. Visualizing theme park visitors' emotions using social media analytics and geospatial analytics. *Tour. Manag.* **2020**, *80*, 104127. [CrossRef]

44. Supak, S.; Brothers, G.; Ghahramani, L.; Van Berkel, D. Geospatial analytics for park & protected land visitor reservation data. In *Analytics in Smart Tourism Design: Concepts and Methods*; Springer: Cham, Switzerland, 2017; pp. 81–109. [CrossRef]

45. Mirzaalian, F.; Halpenny, E. Social media analytics in hospitality and tourism: A systematic literature review and future trends. *J. Hosp. Tour. Technol.* **2019**, *10*, 764–790. [CrossRef]

46. Fu, Y.; Hao, J.X.; Li, X.; Hsu, C.H. Predictive accuracy of sentiment analytics for tourism: A metalearning perspective on Chinese travel news. *J. Travel Res.* **2019**, *58*, 666–679. [CrossRef]

47. Abbasi-Moud, Z.; Vahdat-Nejad, H.; Sadri, J. Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Syst. Appl.* **2021**, *167*, 114324. [CrossRef]

48. Hao, J.X.; Fu, Y.; Hsu, C.; Li, X.; Chen, N. Introducing news media sentiment analytics to residents' attitudes research. *J. Travel Res.* **2020**, *59*, 1353–1369. [CrossRef]

49. Ching, M.R.D.; de Dios Bulos, R. Improving restaurants' business performance using Yelp data sets through sentiment analysis. In Proceedings of the 3rd International Conference on E-commerce, E-Business and E-Government, Lyon, France, 18–21 June 2019; pp. 62–67. [CrossRef]

50. Chan, I.C.C.; Ma, J.; Law, R.; Buhalis, D.; Hatter, R. Dynamics of hotel website browsing activity: The power of informatics and data analytics. *Ind. Manag. Data Syst.* **2021**, *121*, 1398–1416. [CrossRef]

51. Sann, R.; Lai, P.C.; Liaw, S.Y.; Chen, C.T. Predicting online complaining behavior in the hospitality industry: Application of Big Data Analytics to Online Reviews. *Sustainability* **2022**, *14*, 1800. [CrossRef]

52. Mariani, M.M.; Borghi, M. Effects of the Booking. com rating system: Bringing hotel class into the picture. *Tour. Manag.* **2018**, *66*, 47–52. [CrossRef]

53. Du, J.; Floyd, C.; Kim, A.C.; Baker, B.J.; Sato, M.; James, J.D.; Funk, D.C. To be or not to be: Negotiating leisure constraints with technology and data analytics amid the COVID-19 pandemic. *Leis. Stud.* **2021**, *40*, 561–574. [CrossRef]

54. Shin, H.; Ahn, J.; Kang, J.; Cho, J.; Yoon, D.; Lee, H. A comparative analysis of domestic travel intentions and actual travel behaviors in COVID-19: Focused on attitude-behavioral gap. *Asia Pac. J. Tour. Res.* **2022**, *27*, 1193–1206. [CrossRef]

55. Jonas, A.; Mansfeld, Y.; Paz, S.; Potasman, I. Determinants of health risk perception among low-risk-taking tourists traveling to developing countries. *J. Travel Res.* **2011**, *50*, 87–99. [CrossRef]

56. Kim, S.S.; Prideaux, B. Tourism, peace, politics and ideology: Impacts of the Mt. Gumgang tour project in the Korean Peninsula. *Tour. Manag.* **2003**, *24*, 675–685. [CrossRef]

57. Berno, T.; Ward, C. Innocence abroad: A pocket guide to psychological research on tourism. *Am. Psychol.* **2005**, *60*, 593. [CrossRef]

58. Farmaki, A.; Khalilzadeh, J.; and Altinay, L. Travel Motivation and Demotivation Withing Politically Unstable Nations. *Tour. Manag. Perspect.* **2019**, *29*, 118–130. [CrossRef]

59. Rossello, J.; Becken, S.; Santana-Gallego, M. The Effects of Natural Disasters on International Tourism: A Global Analysis. *Tour. Manag.* **2020**, *79*, 104080. [CrossRef]

60. AlQahtany, A.M.; Abubakar, I.R. Public perception and attitudes to disaster risks in a coastal metropolis of Saudi Arabia. *Int. J. Disaster Risk Reduct.* **2020**, *44*, 101422. [CrossRef]

61. Ritchie, B.W. Tourism Disaster Planning and Management: From Response and Recovery to Reduction and Readiness. *Curr. Issues Tour.* **2008**, *11*, 315–348. [CrossRef]

62. Cahyanto, I.; Pennington-Gray, L.; Thapa, B.; Srinivasan, S.; Villegas, J.; Matyas, C.; Kiousis, S. Predicting information seeking regarding hurricane evacuation in the destination. *Tour. Manag.* **2016**, *52*, 264–275. [CrossRef]

63. Giusti, G.; Raya, J.M. The effect of crime perception and information format on tourists' willingness/intention to travel. *J. Destin. Mark. Manag.* **2019**, *11*, 101–107. [CrossRef]

64. Hajibaba, H.; Boztuğ, Y.; Dolnicar, S. Preventing tourists from canceling in times of crises. *Ann. Tour. Res.* **2016**, *60*, 48–62. [CrossRef]

65. Song, H.; Livat, F.; Ye, S. Effects of Terrorist Attacks on Tourist Flows to France: Is Wine Tourism a Substitute to Urban Tourism. *J. Destin. Mark. Manag.* **2019**, *14*, 100385. [CrossRef]

66. Baumert, T.; de Obesso, M.M.; Valbuena, E. How does the terrorist experience alter consumer behaviour? An analysis of the Spanish case. *J. Bus. Res.* **2020**, *115*, 357–364. [CrossRef]

67. MacLaurin, T.L. The Importance of Food Safety in Travel Planning and Decision Selection. *J. Travel Tour. Mark.* **2003**, *15*, 233–257. [CrossRef]

68. Richens, J. Sexually Transmitted Infections and HIV among Travelers: A Review. *Travel Med. Infect. Dis.* **2006**, *4*, 184–195. [CrossRef]

69. Cossens, J.H.; Gin, S. Tourism and AIDS: The Perceived Risk of HIV Infection and Destination Choice. *J. Travel Tour. Mark.* **1994**, *3*, 1–20. [CrossRef]

70. Memish, Z.A.; Osoba, A.O. International Travel and Sexually Transmitted Diseases. *Travel Med. Infect. Dis.* **2006**, *4*, 86–93. [CrossRef]

71. Foroudi, P.; Tabaghdehi, S.A.H.; Marvi, R. The gloom of the COVID-19 shock in the hospitality industry: A study of consumer risk perception and adaptive belief in the dark cloud of a pandemic. *Int. J. Hosp. Manag.* **2021**, *92*, 102717. [CrossRef]

72. Zhong, Y.; Oh, S.; Moon, H.C. What can drive consumers' dining-out behavior in China and Korea during the COVID-19 pandemic? *Sustainability* **2021**, *13*, 1724. [CrossRef]

73. Bonifazi, G.; Corradini, E.; Ursino, D.; Virgili, L. New approaches to extract information from posts on COVID-19 published on Reddit. *Int. J. Inf. Technol. Decis. Mak.* **2022**, *21*, 1385–1431. [CrossRef]

74. Luo, Y.; Xu, X. Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic. *Int. J. Hosp. Manag.* **2021**, *94*, 102849. [CrossRef] [PubMed]

75. Huang, A.; de la Mora Velasco, E.; Farhangi, A.; Bilgihan, A.; Jahromi, M.F. Leveraging data analytics to understand the relationship between restaurants' safety violations and COVID-19 transmission. *Int. J. Hosp. Manag.* **2022**, *104*, 103241. [CrossRef] [PubMed]

76. Zhang, Y.W.; Choi, J.G.; Akhmedov, A.R. The impacts of perceived risks on information search and risk reduction strategies: A study of the hotel industry during the COVID-19 pandemic. *Sustainability* **2021**, *13*, 12221. [CrossRef]

77. Huang, X.; Dai, S.; Xu, H. Predicting tourists' health risk preventative behaviour and travelling satisfaction in Tibet: Combining the theory of planned behaviour and health belief model. *Tour. Manag. Perspect.* **2020**, *33*, 100589. [CrossRef]

78. Çakar, K.; Aykol, Ş. The past of tourist behaviour in hospitality and tourism in difficult times: A systematic review of literature (1978–2020). *Int. J. Contemp. Hosp. Manag.* **2023**, *35*, 630–656. [CrossRef]

79. Sun, S. An analysis on the conditions and methods of market segmentation. *Int. J. Bus. Manag.* **2009**, *4*, 63–70. [CrossRef]

80. Artuger, S. The effect of risk perceptions on tourists' revisit intentions. *Eur. J. Bus. Manag.* **2015**, *7*, 36–43. [CrossRef]

81. Wolff, K.; Larsen, S.; Øgaard, T. How to define and measure risk perceptions. *Ann. Tour. Res.* **2019**, *79*, 102759. [CrossRef]

82. Kim, D.K.D.; Kreps, G.L. An analysis of government communication in the United States during the COVID-19 pandemic: Recommendations for effective government health risk communication. *World Med. Health Policy* **2020**, *12*, 398–412. [CrossRef]

83. Han, S.; Yoon, A.; Kim, M.J.; Yoon, J.H. What influences tourist behaviors during and after the COVID-19 pandemic? Focusing on theories of risk, coping, and resilience. *J. Hosp. Tour. Manag.* **2022**, *50*, 355–365. [CrossRef]

84. Hall, C.M.; Scott, D.; Gössling, S. Pandemics, transformations and tourism: Be careful what you wish for. *Tour. Geogr.* **2020**, *22*, 577–598. [CrossRef]

85. Orden-Mejía, M.; Carvache-Franco, M.; Huertas, A.; Carvache-Franco, W.; Landeta-Bejarano, N.; Carvache-Franco, O. Post-COVID-19 tourists' preferences, attitudes and travel expectations: A Study in Guayaquil, Ecuador. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4822. [CrossRef] [PubMed]

86. Aliperti, G.; Cruz, A.M. Investigating tourists' risk information processing. *Ann. Tour. Res.* **2019**, *79*, 102803. [CrossRef]

87. Chien, P.M.; Sharifpour, M.; Ritchie, B.W.; Watson, B. Travelers' health risk perceptions and protective behavior: A psychological approach. *J. Travel Res.* **2017**, *56*, 744–759. [CrossRef]

88. Sascha, F.; Schiemenz, C.; Bartl, E.; Lindner, E.; Namberger, P.; Schmude, J. Travel participation of Germans before and during the COVID-19 pandemic–the effects of sociodemographic variables. *Curr. Issues Tour.* **2022**, *25*, 4031–4046. [CrossRef]

89. Chen, X.; Hao, Y.; Duan, Y.; Zhang, Q.; Hu, X. Gender and Culture Differences in Consumers' Travel Behavior during the COVID-19 Pandemic. *Sustainability* **2023**, *15*, 1186. [CrossRef]

36 of 40

90. Chunlan, G.; Lu, X.; Huang, S.; Zhao, Y.; Zhao, D. Understanding the Post-Pandemic Travel Intentions Among Chinese Residents: Impact of Sociodemographic Factors, COVID Experiences, Travel Planned Behaviours, Health Beliefs, and Resilience. *Int. J. Tour. Res.* **2024**, *26*, e2752. [CrossRef]

91. Carballo, R.R.; Carmelo, J.L.; Carballo, M.M. The influence of Muslim and Christian destinations on tourists' behavioural intentions and risk perceptions. *Behav. Sci.* **2024**, *14*, 347. [CrossRef]

92. Souissi, A.; Idi Cheffou, A.; Foued, B.S. Impact of anxiety and tourists' habits on their intention to vacation during and after the COVID-19 pandemic: Treatment effect analysis. *J. Tour. Manag. Res.* **2023**, *10*, 62–79. [CrossRef]

93. Kim, T.; Ha, J. Applying a goal-directed behavior model to determine risk perception of COVID-19 and war on potential travelers' behavioral intentions. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2562. [CrossRef]

94. Promsivapallop, P.; Kannaovakun, P. Travel risk dimensions, personal-related factors, and intention to visit a destination: A study of young educated German adults. *Asia Pac. J. Tour. Res.* **2018**, *23*, 639–655. [CrossRef]

95. Reisinger, Y.; Mavondo, F. Travel anxiety and intentions to travel internationally: Implications of travel risk perception. *J. Travel Res.* **2005**, *43*, 212–225. [CrossRef]

96. Godovykh, M.; Pizam, A.; Bahja, F. Antecedents and outcomes of health risk perceptions in tourism, following the COVID-19 pandemic. *Tour. Rev.* **2021**, *76*, 737–748. [CrossRef]

97. Wolff, K.; Larsen, S. Can terrorism make us feel safer? Risk perceptions and worries before and after the July 22nd attacks. *Ann. Tour. Res.* **2014**, *44*, 200–209. [CrossRef]

98. Block, L.G.; Keller, P.A. When to accentuate the negative: The effects of perceived efficacy and message framing on intentions to perform a health-related behavior. *J. Mark. Res.* **1995**, *32*, 192–203. [CrossRef]

99. Ritchie, B.W.; Chien, P.M.; Watson, B.M. It can't happen to me: Travel risk perceptions. In *Tourists' Behaviors and Evaluations*; Advances in Culture, Tourism and Hospitality Research; Emerald Group Publishing Limited: Leeds, UK, 2014; Volume 9. [CrossRef]

100. Quintal, V.A.; Lee, J.A.; Soutar, G.N. Risk, uncertainty and the theory of planned behavior: A tourism example. *Tour. Manag.* **2010**, *31*, 797–805. [CrossRef]

101. Yang, C.L.; Nair, V. Risk Perception Study in Tourism: Are we really measuring perceived risk? *Procedia—Soc. Behav. Sci.* **2014**, *144*, 322–327. [CrossRef]

102. Lerner, J.S.; Keltner, D. Fear, anger, and risk. *J. Personal. Soc. Psychol.* **2001**, *81*, 146–159. [CrossRef]

103. Zenker, S.; Braun, E.; Gyimothy, S. Too afraid to travel? Development of a pandemic (COVID-19) anxiety travel scale (PATS). *Tour. Manag.* **2021**, *84*, 104286. [CrossRef]

104. Luo, J.M.; Lam, C.F. Travel anxiety, risk attitude and travel intentions towards "travel bubble" destinations in Hong Kong: Effect of the fear of COVID-19. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7859. [CrossRef]

105. Davey, G.C.; Hampton, J.; Farrell, J.; Davidson, S. Some characteristics of worrying: Evidence for worrying and anxiety as separate constructs. *Personal. Individ. Differ.* **1992**, *13*, 133–147. [CrossRef]

106. Wolff, K.; Larsen, S. Tourist worries–Here and now vs. there and then: The effect of item wording in the Tourist Worry Scale. *Tour. Manag.* **2013**, *35*, 284–287. [CrossRef]

107. Larsen, S.; Brun, W.; Øgaard, T. What tourists worry about–Construction of a scale measuring tourist worries. *Tour. Manag.* **2009**, *30*, 260–265. [CrossRef]

108. Fennell, D.A. Towards a model of travel fear. *Ann. Tour. Res.* **2017**, *66*, 140–150. [CrossRef]

109. Wang, J.; Liu-Lastres, B.; Ritchie, B.W.; Mills, D.J. Travellers' self-protections against health risks: An application of the full Protection Motivation Theory. *Ann. Tour. Res.* **2019**, *78*, 102743. [CrossRef]

110. Shou, Y.; Olney, J. Attitudes Toward Risk and Uncertainty: The Role of Subjective Knowledge and Affect. *J. Behav. Decis. Mak.* **2020**, *34*, 393–404. [CrossRef]

111. Kovačić, S.; Jovanović, T.; Miljković, Đ.; Lukić, T.; Marković, S.B.; Vasiljević, Đ.A.; Ivkov, M. Are Serbian tourists worried? The effect of psychological factors on tourists' behavior based on the perceived risk. *Open Geosci.* **2019**, *11*, 273–287. [CrossRef]

112. Paek, H.J.; Hove, T. Risk perceptions and risk characteristics. In *Oxford Research Encyclopedia of Communication*; Oxford University Press: Oxford, UK, 2017. [CrossRef]

113. Min, J.; Kim, J.; Yang, K. How generations differ in coping with a pandemic: The case of restaurant industry. *J. Hosp. Tour. Manag.* **2021**, *48*, 280–288. [CrossRef]

114. Liu, S.; Huang, J.C.; Brown, G.L. Information and risk perception: A dynamic adjustment process. *Risk Anal.* **1998**, *18*, 689–699. [CrossRef]

115. Rivas, D.R.Z.; Jaldin, L.M.L.; Canaviri, N.B.; Escalante, P.L.F.; Fernández, A.A.M.; Ticona, A.J.P. Social media exposure, risk perception, preventive behaviors and attitudes during the COVID-19 epidemic in La Paz, Bolivia: A cross sectional study. *PLoS ONE* **2021**, *16*, e0245859. [CrossRef]

116. Heine, S.J.; Lehman, D.R. Cultural variation in unrealistic optimism: Does the West feel more vulnerable than the East? *J. Personal. Soc. Psychol.* **1995**, *68*, 595. [CrossRef]

117. Thornton, B.; Gibbons, F.X.; Gerrard, M. Risk perception and prototype perception: Independent processes predicting risk behavior. *Personal. Soc. Psychol. Bull.* **2002**, *28*, 986–999. [CrossRef]

118. Antwi, C.O.; Ntim, S.Y.; Boadi, E.A.; Asante, E.A.; Brobbey, P.; Ren, J. Sustainable cross-border tourism management: COVID-19 avoidance motive on resident hospitality. *J. Sustain. Tour.* **2022**, *31*, 1831–1851. [CrossRef]

119. Pichierri, M.; Petruzzellis, L.; Passaro, P. Investigating staycation intention: The influence of risk aversion, community attachment and perceived control during the pandemic. *Curr. Issues Tour.* **2023**, *26*, 511–517. [CrossRef]

120. Schneider, C.R.; Dryhurst, S.; Kerr, J.; Freeman, A.L.; Recchia, G.; Spiegelhalter, D.; van der Linden, S. COVID-19 risk perception: A longitudinal analysis of its predictors and associations with health protective behaviours in the United Kingdom. *J. Risk Res.* **2021**, *24*, 294–313. [CrossRef]

121. World Health Organization. Coronavirus Disease (COVID-19) Advice for the Public. 2020. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public (accessed on 24 November 2023).

122. Chou, P.F. An analysis of influenza prevention measures from air travellers' perspective. *Int. Nurs. Rev.* **2014**, *61*, 371–379. [CrossRef]

123. Yang, E.C.L.; Khoo-Lattimore, C.; Arcodia, C. A systematic literature review of risk and gender research in tourism. *Tour. Manag.* **2017**, *58*, 89–100. [CrossRef]

124. Chen, D.G.; Chen, X. Cusp catastrophe regression and its application in public health and behavioral research. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1220. [CrossRef]

125. Mao, C.K.; Ding, C.G.; Lee, H.Y. Post-SARS tourist arrival recovery patterns: An analysis based on a catastrophe theory. *Tour. Manag.* **2010**, *31*, 855–861. [CrossRef]

126. Hsieh, Y.J.; Chen, Y.L.; Wang, Y.C. Government and social trust vs. hotel response efficacy: A protection motivation perspective on hotel stay intention during the COVID-19 pandemic. *Int. J. Hosp. Manag.* **2021**, *97*, 102991. [CrossRef]

127. Wu, J.; Zhang, X.; Zhu, Y.; Yu-Buck, G.F. Get close to the robot: The effect of risk perception of COVID-19 pandemic on customer–robot engagement. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6314. [CrossRef] [PubMed]

128. Quan, L.; Al-Ansi, A.; Han, H. Assessing customer financial risk perception and attitude in the hotel industry: Exploring the role of protective measures against COVID-19. *Int. J. Hosp. Manag.* **2022**, *101*, 103123. [CrossRef] [PubMed]

129. Ryu, K.; Jarumaneerat, T.; Promsivapallop, P.; Kim, M. What influences restaurant dining and diners' self-protective intention during the COVID-19 pandemic: Applying the Protection Motivation Theory. *Int. J. Hosp. Manag.* **2023**, *109*, 103400. [CrossRef] [PubMed]

130. Lepp, A.; Gibson, H. Tourist roles, perceived risk and international tourism. *Ann. Tour. Res.* **2003**, *30*, 606–624. [CrossRef]

131. Simpson, P.M.; Siguaw, J.A. Perceived travel risks: The traveller perspective and manageability. *Int. J. Tour. Res.* **2008**, *10*, 315–327. [CrossRef]

132. Teng, W. Risks perceived by Mainland Chinese tourists towards Southeast Asia destinations: A fuzzy logic model. *Asia Pac. J. Tour. Res.* **2005**, *10*, 97–115. [CrossRef]

133. Skeen, S.; Laurenzi, C.A.; Gordon, S.L.; du Toit, S.; Tomlinson, M.; Dua, T.; Fleischmann, A.; Kohl, K.; Ross, D.; Servili, C.; et al. Adolescent mental health program components and behavior risk reduction: A meta-analysis. *Pediatrics* **2019**, *144*, 20183488. [CrossRef]

134. Adam, I. Backpackers' risk perceptions and risk reduction strategies in Ghana. *Tour. Manag.* **2015**, *49*, 99–108. [CrossRef]

135. Danelon, M.S.; Salay, E. Perceived physical risk and risk-reducing strategies in the consumption of raw vegetable salads in restaurants. *Food Control* **2012**, *28*, 412–419. [CrossRef]

136. Chan, E.Y.Y.; Huang, Z.; Lo, E.S.K.; Hung, K.K.C.; Wong, E.L.Y.; Wong, S.Y.S. Sociodemographic predictors of health risk perception, attitude and behavior practices associated with health-emergency disaster risk management for biological hazards: The case of COVID-19 pandemic in Hong Kong, SAR China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3869. [CrossRef]

137. Kim, H.; Schroeder, A.; Pennington-Gary, L. Does culture influence risk perception? *Tour. Rev. Int.* **2016**, *20*, 11–28. [CrossRef]

138. Crotts, J.C. The effect of cultural distance on overseas travel behaviors. *J. Travel Res.* **2004**, *43*, 83–88. [CrossRef]

139. Sönmez, S.F.; Graefe, A.R. Determining future travel behavior from past travel experience and perceptions of risk and safety. *J. Travel Res.* **1998**, *37*, 171–177. [CrossRef]

140. Ertaş, M.; Kırlar-Can, B. Tourists' risk perception, travel behaviour and behavioural intention during the COVID-19. *Eur. J. Tour. Res.* **2022**, *32*, 3205. [CrossRef]

141. Abdelrahman, M. Personality traits, risk perception, and protective behaviors of Arab residents of Qatar during the COVID-19 pandemic. *Int. J. Ment. Health Addict.* **2022**, *20*, 237–248. [CrossRef]

142. Tepavčević, J.; Blešić, I.; Petrović, M.D.; Vukosav, S.; Bradić, M.; Garača, V.; Gajić, T.; Lukić, D. Personality traits that affect travel intentions during pandemic COVID-19: The case study of Serbia. *Sustainability* **2021**, *13*, 12845. [CrossRef]

143. Qiu, R.T.; Park, J.; Li, S.; Song, H. Social costs of tourism during the COVID-19 pandemic. *Ann. Tour. Res.* **2020**, *84*, 102994. [CrossRef]

144. Sánchez-Cañizares, S.M.; Cabeza-Ramírez, L.J.; Muñoz-Ferñandez, G.; Fuentes-García, F.J. Impact of the perceived risk from COVID-19 on intention to travel. *Curr. Issues Tour.* **2021**, *24*, 970–984. [CrossRef]

145. Airak, S.; Sukor, N.S.A.; Abd Rahman, N. Travel behaviour changes and risk perception during COVID-19: A case study of Malaysia. *Transp. Res. Interdiscip. Perspect.* **2023**, *18*, 100784. [CrossRef]

146. Britannica. United Arab Emirates. Available online: https://www.britannica.com/place/United-Arab-Emirates (accessed on 14 October 2024).

147. World Bank Group Data. Population, Total, United Arab Emirates. Available online: https://data.worldbank.org/indicator/SP.POP.TOTL?end=2023&locations=AE (accessed on 14 October 2024).

148. Forbes Travel Guide. Dubai. Available online: https://www.forbestravelguide.com/destinations/dubai-united-arab-emirates/travel-guide (accessed on 14 October 2024).

149. International Trade Administration. United Arab Emirates—Country Commercial Guide. Available online: https://www.trade.gov/country-commercial-guides/united-arab-emirates-oil-and-gas (accessed on 14 October 2024).

150. Global Media Insight. United Arab Emirates Population Statistics 2024. Available online: https://www.globalmediainsight.com/blog/uae-population-statistics/ (accessed on 14 October 2024).

151. Forbes. Best Countries for Business 2018. United Arab Emirates. Available online: https://www.forbes.com/places/united-arab-emirates/ (accessed on 14 October 2024).

152. Sahu, M. Public policy measures for COVID-19 crisis management: Lessons from the UAE. *Fulbright Rev. Econ. Policy* **2021**, *1*, 246–265. [CrossRef]

153. Abbas Zaher, W.; Ahamed, F.; Ganesan, S.; Warren, K.; Koshy, A. COVID-19 crisis management: Lessons from the United Arab Emirates leaders. *Front. Public Health* **2021**, *9*, 724494. [CrossRef] [PubMed]

154. Hotel News Source. Hospitality Hotspots: The Latest Middle East & North Africa Tourism Statistics [2022–2023]—By Catalina Brinza. Available online: https://www.hotelnewsresource.com/article126095.html (accessed on 14 October 2024).

155. UAE. Dubai Economic Agenda D33. Available online: https://u.ae/en/about-the-uae/strategies-initiatives-and-awards/strategies-plans-and-visions/finance-and-economy/dubai-economic-agenda-d33 (accessed on 14 October 2024).

156. Ahmed, O.; Ahmed, M.Z.; Alim, S.M.A.H.M.; Khan, M.A.U.; Jobe, M.C. COVID-19 outbreak in Bangladesh and associated psychological problems: An online survey. *Death Stud.* **2022**, *46*, 1080–1089. [CrossRef] [PubMed]

157. Faisal, R.A.; Jobe, M.C.; Ahmed, O.; Sharker, T. Replication analysis of the COVID-19 Worry Scale. *Death Stud.* **2022**, *46*, 574–580. [CrossRef] [PubMed]

158. Imbriano, G.; Larsen, E.M.; Mackin, D.M.; An, A.K.; Luhmann, C.C.; Mohanty, A.; Jin, J. Online survey of the impact of COVID-19 risk and cost estimates on worry and health behavior compliance in young adults. *Front. Public Health* **2021**, *9*, 612725. [CrossRef]

159. Shahnazi, H.; Ahmadi-Livani, M.; Pahlavanzadeh, B.; Rajabi, A.; Hamrah, M.S.; Charkazi, A. Assessing preventive health behaviors from COVID-19: A cross sectional study with health belief model in Golestan Province, Northern of Iran. *Infect. Dis. Poverty* **2020**, *9*, 91–99. [CrossRef]

160. Kondo, A.; Abuliezi, R.; Naruse, K.; Oki, T.; Niitsu, K.; Ezeonwu, M.C. Perceived control, preventative health behaviors, and the mental health of nursing students during the COVID-19 pandemic: A cross-sectional study. *J. Health Care Organ. Provis. Financ.* **2021**, *58*, 1–11. [CrossRef]

161. Hales, C.; Shams, H. Cautious incremental consumption: A neglected consumer risk reduction strategy. *Eur. J. Mark.* **1991**, *25*, 7–21. [CrossRef]

162. Bowen, N.K.; Guo, S. *Structural Equation Modeling*; Oxford University Press: Oxford, UK, 2011. [CrossRef]

163. Portland State University. Summary of Minimum Sample Size Recommendations. Available online: https://web.pdx.edu/~newsomj/semclass/ho_sample%20size.pdf (accessed on 14 October 2024).

164. G*Power Statistical Power Analyses for Mac and Windows. Available online: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower (accessed on 14 October 2024).

165. Wolf, E.J.; Harrington, K.M.; Clark, S.L.; Miller, M.W. Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educ. Psychol. Meas.* **2013**, *73*, 913–934. [CrossRef]

166. UCLA Advanced Research Computing. What Does Cronbach's Alpha Mean? | SPSS FAQ. Available online: https://stats.oarc.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/ (accessed on 14 October 2024).

167. Labben, T.G.; Chen, J.S.; Ghoudi, K.; Elrazaz, T. Motivational, Cognitive, and Affective Antecedents' Impact on Health Risk Perception and Health Risk Reduction Behavior During the Dining Out Experiences. *under review, available from authors upon acceptance and request.*

168. Demsar, J.; Curk, T.; Erjavec, A.; Gorup, C.; Hocevar, T.; Milutinovic, M.; Mozina, M.; Polajnar, M.; Toplak, M.; Staric, A.; et al. Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14*, 2349–2353. Available online: http://jmlr.org/papers/volume14/demsar13a/demsar13a.pdf (accessed on 25 November 2023).

169. GitHub. Orange OpenSource. Available online: https://github.com/Orange-OpenSource (accessed on 14 October 2024).

170. ChatGPT. Available online: https://chatgpt.com/ (accessed on 14 October 2024).

171. Heumann, C.; Shalabh, M.S. *Introduction to Statistics and Data Analysis*; Springer International Publishing: Cham, Switzerland, 2016; pp. 36–44. [CrossRef]

172. Ertek, G.; Tokdemir, G.; Hammoudi, M.M. Graph-Based Visualization of Stochastic Dominance in Statistical Comparisons. In Proceedings of the 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, 3–7 November 2019; pp. 1–7. [CrossRef]

173. Gini, C. *Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche*; C. Cuppini: Bologna, Italy, 1912.

174. Ceriani, L.; Verme, P. The origins of the Gini index: Extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *J. Econ. Inequal.* **2012**, *10*, 421–443. [CrossRef]

175. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]

176. Yang, X.; Lo, D.; Li, L.; Xia, X.; Bissyandé, T.F.; Klein, J. Characterizing malicious Android apps by mining topic-specific data flow signatures. *Inf. Softw. Technol.* **2017**, *90*, 27–39. [CrossRef]

177. Iqbal, K.; Khan, M.S. Email classification analysis using machine learning techniques. *Appl. Comput. Inform.* 2022, *ahead-of-print*. [CrossRef]
178. Raileanu, L.E.; Stoffel, K. Theoretical comparison between the Gini Index and Information Gain criteria. *Ann. Math. Artif. Intell.* **2004**, *41*, 77–79. [CrossRef]
179. Zhao, X.; Nie, X. Splitting choice and computational complexity analysis of decision trees. *Entropy* **2021**, *23*, 1241. [CrossRef] [PubMed]
180. Rokach, L.; Maimon, O. Decision Trees. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: Boston, MA, USA, 2005; pp. 165–192. [CrossRef]
181. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. In *Emerging Artificial Intelligence Applications in Computer Engineering; Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*; Frontiers in Artificial Intelligence and Applications; Maglogiannis, I., Karpouzis, K., Wallace, M., Soldatos, J., Eds.; IOS Press: Amsterdam, The Netherlands, 2007; Volume 160, pp. 3–24, ISBN 978-1-58603-780-2/978-1-60750-270-8. Available online: https://tinyurl.com/bp6jm5zz (accessed on 14 October 2024).
182. Loh, W.Y. Fifty years of classification and regression trees. *Int. Stat. Rev.* **2014**, *82*, 329–348. [CrossRef]
183. Lim, T.S.; Loh, W.Y.; Shih, Y.S. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Mach. Learn.* **2000**, *40*, 203–228. [CrossRef]
184. Caruana, R.; Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 161–168. [CrossRef]
185. Holte, R.C. Very simple classification rules perform well on most commonly used datasets. *Mach. Learn.* **1993**, *11*, 63–90. [CrossRef]
186. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844. [CrossRef]
187. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
188. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random Forests. In *Ensemble Machine Learning*; Zhang, C., Ma, Y., Eds.; Springer: New York, NY, USA, 2012. [CrossRef]
189. Hamza, M.; Larocque, D. An empirical comparison of ensemble methods based on classification trees. *J. Stat. Comput. Simul.* **2005**, *75*, 629–643. [CrossRef]
190. Beck, F.; Burch, M.; Munz, T.; Di Silvestro, L.; Weiskopf, D. Generalized pythagoras trees for visualizing hierarchies. In Proceedings of the 2014 International Conference on Information Visualization Theory and Applications (IVAPP), Lisbon, Portugal, 5–8 January 2014; pp. 17–28. [CrossRef]
191. Scheibel, W.; Trapp, M.; Limberger, D.; Döllner, J. A taxonomy of treemap visualization techniques. In Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020)–IVAPP, Valletta, Malta, 27–29 February 2020; Volume 3, pp. 273–280. [CrossRef]
192. Munz, T.; Burch, M.; van Benthem, T.; Poels, Y.; Beck, F.; Weiskopf, D. Overlap-free drawing of generalized pythagoras trees for hierarchy visualization. In Proceedings of the 2019 IEEE Visualization Conference (VIS), Vancouver, BC, Canada, 20–25 October 2019; pp. 251–255. [CrossRef]
193. Phillips, N.D.; Neth, H.; Woike, J.K.; Gaissmaier, W. FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgm. Decis. Mak.* **2017**, *12*, 344–368. [CrossRef]
194. Dunn, B.K.A.; Worland, A.; Wagle, S. Interactive Decision Tree Creation and Enhancement with Complete Visualization for Explainable Modeling. *arXiv* **2023**, arXiv:2305.18432. Available online: https://arxiv.org/ftp/arxiv/papers/2305/2305.18432.pdf (accessed on 14 October 2024).
195. Jung, H.S.; Yoon, H.H.; Song, M.K. A study on dining-out trends using big data: Focusing on changes since COVID-19. *Sustainability* **2021**, *13*, 11480. [CrossRef]
196. Budaraju, R.R.; Jammalamadaka, S.K.R. Mining Negative Associations from Medical Databases Considering Frequent, Regular, Closed and Maximal Patterns. *Computers* **2024**, *13*, 18. [CrossRef]
197. Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [CrossRef]
198. Probst, P.; Boulesteix, A.L. To tune or not to tune the number of trees in random forest. *J. Mach. Learn. Res.* **2018**, *18*, 6673–6690. Available online: https://dl.acm.org/doi/pdf/10.5555/3122009.3242038 (accessed on 14 October 2024).
199. Probst, P.; Wright, M.N.; Boulesteix, A.L. Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]
200. Van Rijn, J.N.; Hutter, F. Hyperparameter importance across datasets. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2367–2376. [CrossRef]
201. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
202. Roumeliotis, K.I.; Tselikas, N.D. ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet* **2023**, *15*, 192. [CrossRef]
203. Ray, P.P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Phys. Syst.* **2023**, *3*, 121–154. [CrossRef]
204. Preim, B.; Lawonn, K. A survey of visual analytics for public health. *Comput. Graph. Forum* **2020**, *39*, 543–580. [CrossRef]

205. Tableau Public. Available online: https://public.tableau.com/ (accessed on 14 October 2024).

206. Luanglath, P.; Rewtrakunphaiboon, W. Determination of a minimum sample size for film-induced tourism research. In Proceedings of the Silpakorn 70th Anniversary International Conference, Bangkok, Thailand, 16–18 January 2013.

207. Memon, M.A.; Ting, H.; Ramayah, T.; Chuah, F.; Cheah, J.H. A review of the methodological misconceptions and guidelines related to the application of structural equation modeling: A Malaysian scenario. *J. Appl. Struct. Equ. Model.* **2017**, *1*, i–xiii. [CrossRef] [PubMed]

208. Polit, D.F.; Beck, C.T. Generalization in quantitative and qualitative research: Myths and strategies. *Int. J. Nurs. Stud.* **2010**, *47*, 1451–1458. [CrossRef] [PubMed]

209. Memon, M.A.; Ting, H.; Cheah, J.H.; Thurasamy, R.; Chuah, F.; Cham, T.H. Sample size for survey research: Review and recommendations. *J. Appl. Struct. Equ. Model.* **2020**, *4*, 1–20. [CrossRef] [PubMed]

210. Sharma, G. Pros and Cons of Different Sampling Techniques. *Int. J. Appl. Res.* **2017**, *3*, 749–752.

211. Kalton, G. Sampling considerations in research on HIV risk and ilness. In *Methodological Issues in AIDS Behavioral Research*; Ostrow, D.G., Kessler, R.C., Eds.; Plenum Press: New York, NY, USA, 1993; pp. 53–72. [CrossRef]

212. Penrod, J.; Preston, D.B.; Cain, R.E.; Starks, M.T. A discussion of chain referral as a method of sampling hard-to-reach populations. *J. Transcult. Nurs.* **2003**, *14*, 100–107. [CrossRef]

213. Ertek, G.; Kaya, M.; Kefeli, C.; Onur, Ö.; Uzer, K. Scoring and predicting risk preferences. In *Behavior Computing*; Cao, L., Yu, P.S., Eds.; Springer: London, UK, 2012; pp. 143–163. [CrossRef]