UNIVERSIDAD DE MÁLAGA

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

PROGRAMA DE DOCTORADO EN ECONOMÍA Y EMPRESA

Big Data and Information Theory for Decision-Making: An Application to the Tourism Demand

TESIS DOCTORAL

(PhD Dissertation)

“Por compendio de publicaciones”

PhD Student (Author): Miguel Ángel Ruiz Reina

Supervisor and Tutor: Antonio Caparrós Ruiz

January 2022

UNIVERSIDAD DE MÁLAGA

AUTOR: Miguel Ángel Ruiz Reina

https://orcid.org/0000-0001-6055-7810

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña MIGUEL ÁNGEL RUIZ REINA
Estudiante del programa de doctorado EN ECONOMÍA Y EMPRESA de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: BIG DATA AND INFORMATION THEORY FOR DECISION-MAKING: AN APPLICATION TO THE TOURISM DEMAND

Realizada bajo la tutorización de ANTONIO CAPARRÓS RUIZ y dirección de ANTONIO CAPARRÓS RUIZ (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 24 de ENERO de 2022

Miguel Ángel Ruiz Reina Fdo.: Doctorando/a	 Fdo.: ANTONIO CAPARRÓS RUIZ Tutor/a
  Fdo.: ANTONIO CAPARRÓS RUIZ Director/es de tesis	

Edificio Pabellón de Gobierno. Campus El Ejido. 29071
Tel.: 952 13 10 28 / 952 13 14 61 / 952 13 71 10
E-mail: doctorado@uma.es

INFORME DEL DIRECTOR Y TUTOR PARA LA AUTORIZACIÓN DE DEFENSA DE TESIS DOCTORAL

Antonio Caparrós Ruiz, Profesor Titular del Departamento de Economía Aplicada (Estadística y Econometría) de la Universidad de Málaga, como director de tesis y tutor de Miguel Ángel Ruiz Reina emite un informe favorable sobre la idoneidad de la presentación de la tesis doctoral titulada "Big Data and Information Theory for Decision-Making: An Application to the Tourism Demand" por compendio de publicaciones.

El doctorando reúne el requisito mínimo fijado por el Programa de Doctorado en Economía y Empresa de la UMA para la lectura de la tesis doctoral; es decir, un artículo publicado en una revista JCR. En particular, el compendio de publicaciones que conforman la tesis está formado por tres contribuciones científicas "peer-reviewed" que siguen una línea de investigación en coherencia con la temática y los objetivos de la tesis, y que no han sido utilizadas para tesis anteriores. Este conjunto de estudios cumple los criterios de calidad exigidos por el Programa de Doctorado en Economía y Empresa, y está compuesto por un capítulo de libro y dos artículos JCR:

1. Capítulo de libro.

Reina, M.Á.R. (2020) Big Data: Forecasting and Control for Tourism Demand. In: Valenzuela O., Rojas F., Herrera L.J., Pomares H., Rojas I. (eds) Theory and Applications of Time Series Analysis. ITISE 2019. Contributions to Statistics. Springer, Cham. https://doi.org/10.1007/978-3-030-56219-9_18.

2. Artículos JCR.

2.1. Ruiz-Reina, M.Á. Entropy Method for Decision-Making: Uncertainty Cycles in Tourism Demand. Entropy 2021, 23, 1370. https://doi.org/10.3390/e23111370.

2.2. Ruiz-Reina, M. Á. (2021). Spatio-temporal clustering: Neighbourhoods based on median seasonal Entropy. Spatial Statistics, 45, 100535. https://doi.org/https://doi.org/10.1016/j.spasta.2021.100535.

En definitiva, la tesis doctoral aporta novedosos procedimientos metodológicos en los campos de investigación dedicados al Big Data y a la Teoría de la Información, los cuales son aplicados de forma idónea al análisis de la demanda turística desde una perspectiva econométrica. Los resultados obtenidos son relevantes para la toma de decisiones de los agentes económicos, ofreciendo nuevas herramientas de predicción a través del uso de Google Trends, la modelización de la incertidumbre en la demanda turística, y la aplicación del análisis cluster basado en la Entropía a las Ciencias Sociales.

Fdo: Antonio Caparrós Ruiz
Málaga, 24 de enero del 2022

Big Data and Information Theory for Decision-Making: An Application to the Tourism Demand

PhD Student (Author): Miguel Ángel Ruiz Reina

PhD Program in Economics and Business [Programa de Doctorado en Economía y Empresa de la Universidad de Málaga]. Faculty of Economics and Business. University of Malaga. e–mail: ruizreina@uma.es

Supervisor and Tutor: Antonio Caparrós Ruiz

University of Malaga. Faculty of Economics and Business. Department of Applied Economics (Statistics and Econometrics). Plaza de El Ejido s/n, 29013, Málaga (España); e–mail: antonio@uma.es. Tel: +34 952 13 11 63, Fax: +34 952137262.

Acknowledgements (Spanish)

Con la elaboración de esta tesis doctoral he contraído deudas de gratitud con personas e instituciones que han formado parte de mi educación en todos estos años. La elaboración de esta tesis es fruto de un largo periodo de estudio; me gustaría agradecer a las instituciones de la Universidad de Málaga (UMA), Universidad del País Vasco/ Euskal Herriko Unibertsitatea (UPV/EHU), Universidad Nacional de Educación a Distancia (UNED) y Escuela de Organización Industrial (EOI) por su labor de transformación de la sociedad a través de la cultura e investigación. Todas estas instituciones han impactado positivamente en mi desarrollo personal y profesional en un mundo cambiante donde su oferta formativa significó un "ascensor" competencial en la investigación realizada.

En mi caso personal, he sido fuertemente influenciado por las personas que han trabajado en estas instituciones, debido a esto muestro mi total agradecimiento y respeto por todas ellas. Seguramente seré injusto en los agradecimientos, puesto que la cantidad de personas que me ayudaron es innumerable y pido disculpas por adelantado.

En cuanto a la UMA, siempre estaré agradecido por ser mi alma máter y convertirme en un joven economista al inicio de mi carrera profesional. En mis estudios de Licenciatura en Economía tuve la suerte de encontrar a dos personas que me inspiraron y fueron el combustible inicial de toda mi carrera a partir de su innegable influencia intelectual. Por un lado, el profesor Bernardo Moreno Jiménez me mostró el camino del apasionante mundo de la Teoría Económica, su trabajo y ejemplo son inspiraciones constantes en este complicado mundo de la investigación. Por otro lado, el profesor Francisco Trujillo Aranda con su impecable actividad docente consiguió (conscientemente o no) que me "enamorara" de la Econometría. Es muy probable que, sin la aparición de estas dos personas en mi vida, no se hubiera dado nada de lo que hoy estamos agradeciendo.

Siendo ya un joven economista, decidí iniciar la experiencia de estudiar un máster en análisis económico especializado en técnicas cuantitativas en la UPV/EHU. El agradecimiento a la ciudad de Bilbao y a la facultad de Sarriko no se puede cuantificar. De esta experiencia y de esta prestigiosa institución hay una cantidad de personas de las que dedicaría decenas de páginas con agradecimientos. No puedo obviar a dos figuras que me influyeron en el análisis de series temporales, los catedráticos Francisco Javier Fernández Macho y Josu Arteche tienen su parte proporcional en el entendimiento de ciclos descubiertos en esta investigación. Además, estoy especialmente agradecido a dos personas sin minusvalorar al resto de personas que me ayudaron e inspiraron. Una de ellas es el catedrático en Estadística Vicente Nuñez Antón, gracias a él entendí el mundo estadístico como no lo había hecho antes y me proporcionó de herramientas de análisis poderosas. La otra persona es la profesora Ilaski Barañano Mendatxa su labor docente en los campos de Macroeconomía me han inspirado en años posteriores de mi docencia, pero me quedo con su inmensa generosidad, paciencia y excelentes consejos en conversaciones informales. De todas estas conversaciones siempre tengo en mi cabeza unas palabras que ella me transmitió sobre lo que significaba realizar una tesis doctoral: "realizar una contribución a la ciencia". En los momentos duros del camino en la realización de esta tesis, siempre he recordado esas palabras como inspiración y luz guía de este trabajo.

La UNED fue mi confirmación de que todo lo que había estudiado con anterioridad era lo adecuado y las puertas que me abrieron en la investigación fueron numerosas. En la EOI me formé en análisis de Big Data e Inteligencia de Negocio, lo cual significaba "cerrar el círculo" en la transición intelectual de un economista cuantitativo a una persona de negocio basada en datos. Gracias a la UNED y EOI encontré la motivación final antes

de decidirme a comenzar los estudios de doctorado, con todo esto entendí que podía "realizar una contribución a la ciencia".

Con la vuelta a la Universidad de Málaga, agradecer al Programa de Doctorado la admisión y soporte prestado en estos años. Agradecer a Antonio Caparrós Ruiz, tutor y director de esta tesis doctoral su excelente trabajo. Simplemente es la persona más importante en la elaboración de esta tesis, su generosidad, amabilidad, altruismo, buen hacer, paciencia e inteligencia han contribuido de un modo sobresaliente en la investigación asociada en cada trabajo que componen la tesis. Su excelsa dirección y gestión en este periodo de formación, anteponiendo el interés del doctorando generosamente con sus consejos e inteligentes comentarios son el mejor aval de este trabajo. No puedo obviar el papel de los componentes de la comisión anual de evaluación del Programa de Doctorado, sus comentarios y recomendaciones han encaminado este trabajo hacia un mejor entendimiento, mis mayores respetos hacia Ana Lozano, Germán Gemar y Mª Luz González. Agradecer también a Alfonso Delgado Bonal de la NASA por su contribución y su influencia intelectual. Agradecer también a todos los evaluadores de trabajos, tribunales, organizadores de congresos y demás personas externas a la tesis que han contribuido con sus inteligentes comentarios. Por último, agradecer a mi compañera de despacho en la UPV/EHU sus interminables horas de trabajo en la que pudimos avanzar en este complicado mundo de la Econometría. María siempre te recordaré y te llevaré en mi corazón allá donde estés.

Especial agradecimiento personal a mi familia y amigos, a mi madre por inculcarme la pasión por trabajar y ser siempre la persona que confiaba en mí. A mi pareja Mª José, por su generosidad y ayudarme a conseguir este sueño durante algo más de tres años de investigación. A mi padre, hermano, mi primo Isma por estar siempre ahí. No puedo obviar y dejar de recordar a todas las personas que me han dado trabajo para poder financiar mis estudios desde que tenía 18 años. Ser estudiante toda la vida a "tiempo parcial" es complicado, pero sin estos trabajos hubiera sido materialmente imposible poder ahorrar para invertir en mi sueño.

A todas las personas que entendieron este camino, simplemente gracias, este trabajo sin vuestra ayuda no sería posible.

Funding and additional acknowledgements (Spanish)

Agradecer la financiación parcial de la tesis y de los estudios relacionados por fondos públicos: préstamo renta ICO para realizar estudios de posgrado de Máster Universitario o de Doctorado (Orden EDU/3108/2009); financiación de grupo de investigación de la Universidad de Málaga "SEJ 157-INDICADORES SOCIALES"; asignación de gastos para PDI del Departamento de Economía Aplicada (Estadística y Econometría) y del Departamento de Teoría e Historia Económica de la Universidad de Málaga; fondos de la Universidad de Málaga "Funding for open access charge: Universidad de Málaga/CBUA".

"(…) El economista ha de poseer una rara combinación de dones. (…) Tiene que ser matemático, historiador, hombre de Estado, filósofo – en un cierto grado –. Tiene que entender símbolos y hablar con palabras. Tiene que contemplar lo que es particular en términos de lo que es general y tocar lo abstracto y lo concreto en el mismo vuelo de pensamiento. Tiene que estudiar el presente a la luz del pasado para los propósitos del futuro. Ningún rincón de la naturaleza del hombre o de sus instituciones tiene que escaparse del todo a su mirada. (…)."

JOHN MAYNARD KEYNES, "Alfred Marshall, 1842-1924" y "Mary Paley Marshall, 1850-1944" en Ensayos biográficos (1951), Crítica, 1992, p. 185)

Contents

Research: scientific production that supports the thesis and other activities

In this first section, the central research studies in the academic period 2018-2021 are cited. In particular, the activities and scientific production that support this thesis are outlined, in addition to other milestones of scientific dissemination.

Table 1 Activities in the training period and scientific production that endorse this thesis.

	Scientific production activities	
0.1.	*Papers included and book chapter that support the thesis*	3
0.2.	***Other contributions in the research period***	
0.2.1.	*Conferences Contributions*	12
0.2.2.	*Seminars*	3
0.2.3.	*Other non-indexed publications (Peer Reviewed)*	3
	Total activities	21

0.1. Papers included and a book chapter that support the thesis

1) *Reina, M.Á.R. (2020) Big Data: Forecasting and Control for Tourism Demand. In: Valenzuela O., Rojas F., Herrera L.J., Pomares H., Rojas I. (eds) Theory and Applications of Time Series Analysis. ITISE 2019. Contributions to Statistics. Springer, Cham. https://doi.org/10.1007/978-3-030-56219-9_18*

2) *Ruiz-Reina, M.Á. Entropy Method for Decision-Making: Uncertainty Cycles in Tourism Demand. Entropy 2021, 23, 1370. https://doi.org/10.3390/e23111370*

3) *Ruiz-Reina, M. Á. (2021). Spatio-temporal clustering: Neighbourhoods based on median seasonal Entropy. Spatial Statistics, 45, 100535. https://doi.org/https://doi.org/10.1016/j.spasta.2021.100535*

0.2. Other contributions in the research period

0.2.1. Conferences Contributions

1) *Ruiz-Reina, M.A., "Forecasting Methodology and Comparison for Tourism Demand in Spain", I Jornadas doctorales del programa en Economía y Empresa Málaga, Universidad de Málaga (Spain), 6 y 7 de junio 2019*

2) *Ruiz-Reina, M.A., "Google queries for Spanish Tourism Demand: A dynamic explanation in the digital market", I International workshop in Statistics and Econometrics methods applied to Tourism, Universidad Complutense de Madrid (Spain), September 2nd and 3rd 2019*

3) *Ruiz-Reina, M.A. (2019). "Big Data: Does it really improve Forecasting techniques for Tourism Demand in Spain?", International Conference on Time Series and Forecasting. ITISE 2019. 1, pp. 694-706. Granada (Spain): Godel Impresiones Digitales S.L. http://itise.ugr.es/ITISE2019_vol1.pdf*

4) *Ruiz-Reina, M. (2019). "Forecasting using Big Data: The case of Spanish Tourism Demand". International Conference on Time Series and Forecasting. ITISE 2019. 2, pp. 782-789. Granada (Spain): Godel Impresiones Digitales S.L. http://itise.ugr.es/ITISE2019_vol2.pdf*

5) *Ruiz-Reina, M. (2019). "Entropy of Tourism: the unseen side of tourism accommodation". ISBN: 978-609-485-004-2 pub. Proceedings of the International Conference on Applied Research in Business, Management and Economics 12 – 14 December, 2019. Universidad Pompeu Fabra (UPF), Barcelona, Spain*

6) *Ruiz-Reina, M. (2020). "Automatic forecasting: a selection criterion applied to the social sciences". VI Encuentro internacional de especialización para la investigación en Economía y Empresa, jaén 2020*

7) *Ruiz-Reina, M. (2020). "Google Trends and Tourism: Regression Cluster Analysis". 11th International Conference on Modern Research in Management, Economics and Accounting, ISBN: 978-609-485-093-6 pub. University of Oxford, UK. 2020*

8) *Ruiz-Reina, M.A. (2021), "Clustering Spatio-Temporal for Tourism Demand in Spain", II Jornadas doctorales del programa en Economía y Empresa Málaga\ II Doctoral Conference in Economics and Business, Universidad de Malaga (Spain), 3 y 4 de junio 2021*

9) *Ruiz-Reina, M. (2021). "Statistical learning: Bernoulli time series modelling for discrete decision choice". 4th International Conference on Econometrics and Statistics (EcoSta 2021), 24-26 June 2021, Virtual Conference, HKUST, Hong Kong. 2021. ISBN: 978-9925-7812-0-1*

10) *Ruiz-Reina, Miguel Á. 2021. "Tourism and Big Data: Forecasting with Hierarchical and Sequential Cluster Analysis", International Conference on Time Series and Forecasting. ITISE 2021. 19th - 21th July 2021, Las Palmas de Gran Canarias (Spain)*

11) *Ruiz-Reina, Miguel Á. 2021. "Cycles and Uncertainty: Applications in the Tourist Accommodation Market", International Conference on Time Series and Forecasting. ITISE 2021. 19th - 21th July 2021, Las Palmas de Gran Canarias (Spain)*

12) *Ruiz-Reina, Miguel Á. 2021. "Bernoulli Time Series Modelling with Application to Accommodation Tourism Demand", International Conference on Time Series and Forecasting. ITISE 2021. 19th - 21th July 2021, Las Palmas de Gran Canarias (Spain)*

0.2.2. Seminars

1) *Ruiz-Reina, M.A.; "Big Data and Machine Learning: some applications in forecasting for Tourism", Master en Análisis Económico y Empresarial, Universidad de Malaga (Spain), February 26th 2020*

2) *Ruiz-Reina, M.A.; University of Malaga (Spain), 2020. "Randomness for accommodation in Tourism Market"*

3) *Ruiz-Reina, M.A.; University of Malaga (Spain), 2021. "Big Data and R Programming: Automatic Forecasting and Evaluation Criteria"*

0.2.3. Other non-indexed publications (Peer Reviewed)

1) *Ruiz-Reina, M. Á. (2021). Cycles and Uncertainty: Applications in the Tourist Accommodation Market. Engineering Proceedings, 5(1), 3. doi:10.3390/engproc2021005003*

2) *Ruiz-Reina, M. Á. (2021). Tourism and Big Data: Forecasting with Hierarchical and Sequential Cluster Analysis. Engineering Proceedings, 5(1), 14. doi:10.3390/engproc2021005014*

3) *Ruiz-Reina, M. Á. (2021). Bernoulli Time Series Modelling with Application to Accommodation Tourism Demand. Engineering Proceedings, 5(1), 17. doi:10.3390/engproc2021005017*

1. Introduction

1.1. The current Research

This thesis belongs to the methodological field of Data Science, involving scientific methods, processes and systems to extract knowledge in data sets in business (Coussement & Benoit, 2021). Research improves learning and pattern representation by structuring Data Science and vice versa; the characterisations identified and modelled by Data Science allow decisions on large scales. For this, novel analysis methodologies are developed in fields such as Big Data, Information Theory, Spatial-Statistics, Time Series, Econometrics and Forecasting analytics to empower the decision-making of economic agents. Forecasting techniques have always been in the foreground in the context of planning and decisions. The uncertainty of individuals and organisations are about minimising together with maximising utility (Petropoulos et al., 2020). This PhD dissertation contributes to the methodological and empirical fields of science, thus filling gaps in knowledge and creating new lines of research (Martínez et al., 2021). The empirical application of this extensive work is analysed with monthly data on tourist accommodation for international visitors with data from the Spanish National Institute of Statistics (INE). In particular, this thesis develops innovative methodologies in the description of consumer decision-making. It creates tools for the intervention of future actions of firms with the analysis of behaviour patterns in the decision of tourist accommodation in Spain according to country of origin, analysing the last 15 years. This thesis is methodological and provides tools for public or private organisations. This analysis pursues the following objectives (Reina, 2020; Ruiz Reina, 2021; Ruiz-Reina, 2021): (i) theoretically demonstrate that internet searches with Google Trends search engines are produced prior to tourist accommodation using keywords; (ii) demonstrate temporary causal relationships in the decision-making of tourist accommodation between hotels and tourist apartments. This study will model demand in a secondary market (apartments demand) through a primary accommodation market (hotel demand). The linear and non-linear temporal causality tests will determine the direction of the demand for accommodation. Studying cycles of uncertainty in the domain in time and frequency will provide cyclical information; (iii) finally, we will develop unsupervised spatio-temporal clustering methods. These unsupervised methods will allow spatial and temporal ordering of demands to carry out an intervention based on the knowledge obtained by the techniques developed. Combining this unsupervised analysis and in a context without limiting assumptions will allow firms and organisations to make efficient decisions in contexts of uncertainty, complexity and spatial information. (Batty et al., 2014). These spatial results could complement previous studies on Spatial-Econometrics for tourist demand flows (Alvarez-Diaz et al., 2020).

This doctoral thesis answers scientific questions related to the tourism industry: Big Data (chapter 1) — Is there a causal relationship between tourism demand and Big Data (Google Trends)? Seasonal analysis of searches in Google, modelling and selection criteria for predictive models; Information Theory and decision-making (chapter 2) — Is there a relationship between the demand for hotels and tourist apartments by country of origin? Is there a linear coincidence or not? Are there cycles of uncertainty in the behaviour of demand? Do all the nationalities of tourists present the same seasonality or cyclical behaviour? Spatio-Temporal Clustering (chapter 3) — How do we group tourist demands according to seasonal behaviour for accommodation in hotels and apartments? Are seasonality cycles and behaviours identified in decision-making?

1.2. Big Data and the data life cycle in tourism demand modelling

The focus of this thesis is methodological, developing novel data analysis techniques, and empirical with the application of the methodology to the Spanish tourist accommodation market. The benefits of understanding data applied to the tourism sector are high for the scientific community. The role of Big Data analysis and the understanding of the data life cycle is a central axis in the development of this thesis. The structuring of the data supports the effective communication of the modelling, improves the quality of decision-making, decreases approval time in the presentation of results to decision-makers (McKendry et al., 2021).

Applying a generic concept such as Big Data in a superconnected world responds to a prior question based on decision-making. Since NASA researchers Michael Cox and David Ellsworth used the term "Big Data" for the first time in 1997, the scientific literature has exceeded unsuspected limits. The researchers themselves highlighted applying this concept to numerous areas as a basis for understanding the visualisation of complex problems (Cox & Ellsworth, 1997). Understanding complex problems allows knowing and understanding the challenges in order to make efficient decisions.

The application of Big Data in a digital environment represents a competitive advantage over those who do not use advanced data analytics. This use plays a significant role for people and companies daily, revealing competitive advantages. Combining the five v's of Big Data (volume, velocity, value, variety and veracity) has led to a rapid expansion of data storage, analysis and visualisation techniques (Mikalef et al., 2018) The mechanisms for understanding and processing data through Big Data is based on added value for economic agents. The scientific literature has been based on all the aspects

collected under the concept of Big Data, assuming investments in infrastructure, business intelligence or analysis tools in areas not yet developed. (Gupta & George, 2016).

This rapid development of Big Data in the tourism industry has standardised and improved Data Science techniques for decision-making (J. Li et al., 2018). The information generated on the internet and the structuring of large volumes of data represent an unsuspected tool at the beginning of the international tourism market. Economic agents, particularly companies and agencies, have discovered methods of interacting with potential consumers. This exchange of information through data allows efficient decisions to be made at any given moment. Figure 1 shows the so-called life cycle of tourist demand data, which is the basis for understanding the Big Data environment for academics and professionals (Ruiz-Reina, 2019d).

From this Figure 1, all the knowledge developed in this doctoral thesis is extracted, it is convenient to describe the scheme for a better understanding of the later chapters in this work. Initially, we will begin to analyse the scheme from the ad-hoc point of view of the data set available in the "Data Warehouse" (Timakum et al., 2021). In this sense, we will work with two types of data: structured data (INE) and initially unstructured data (Google Trends). This unstructured data is initially structured by Google's engineering architectures, assuming a free and easily accessible ad-hoc resource in this research. The next point of the scheme is called "Analytics: Modelling and Forecasting", this is the central working scheme of this thesis. In particular, it is subdivided into three chapters: “Descriptive and Predictive Analytics (Chapter 1): Big Data — Forecasting and Control for Tourism Demand”; "Uncertainty Modelling (Chapter 2): Entropy Method for Decision-Making: Uncertainty Cycles in Tourism Demand"; "Prescriptive Analytics (Chapter 3): Spatio-temporal clustering: Neighbourhoods based on median seasonal Entropy."

The three chapters of the thesis will be developed extensively in the two articles and book chapter that make up this compendium of publications. This modelling will allow stakeholders to make data-driven decisions in efficiency in terms of knowledge, process improvements, cost reduction, time savings or management optimization. With this information, firms will structure their tourist offer on the web, including accommodation, transportation and additional services for potential consumers. Once consumers have this information on the Internet, they search Google, where potential consumers will find the offer. In this way, the tourist demand finally emerges, generating new data on the network that would be analysed again based on the behaviour of the life cycle of the data described above in the tourist demand.

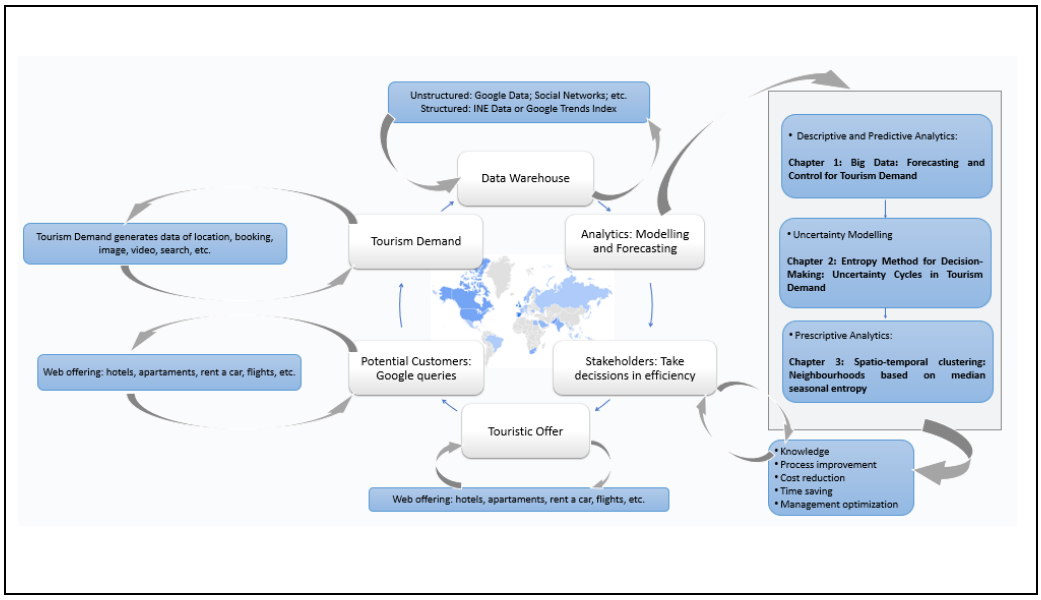

Fig. 1 The life cycle of the data in the Tourism Demand. Own elaboration.

Big Data and data analytics have grown with tourism demand's descriptive and predictive purpose in the last decade. The Data Science applied in this work summarizes and visualises the information obtained from the data to obtain conclusions in decision-making. The data produces knowledge, which supposes an added value in interpreting mathematical patterns to the organisations. However, despite the extensive written literature, unaddressed issues challenge developing this work (Wu et al., 2021). This work is supported by a book chapter publication and two scientific articles in which innovations in the methodology are developed in topics such as Big Data, Google Trends, Spatial-Statistics, Econometrics, Information Theory or unsupervised seasonal clustering.

Big Data analysis applied to the tourism industry involves knowing patterns of behaviour of individuals. Sometimes generalist and objectivist treatments of people are applied (Weaver, 2021). With the methodologies developed in this work, it is intended to obtain the most significant knowledge to adapt to the maximum patterns of tourists. Once the reader has understood what the general work scheme for modelling tourism demand is, we will focus on the three chapters that make up the preparation of this thesis by a compendium of publications. In particular, we will work with spatial data from the time series of demand for tourist accommodation in Spain. We will formally demonstrate the causality between Google searches and hotel demand. This analysis allows us to model the relationship between the demand for hotel accommodation and tourist apartments and

finally model the spatio-temporal classification algorithms. However, at this point, it is considered relevant to include a specific section on how important the tourism industry is in the Spanish economy.

1.3. *The tourism industry and relevance in modern economies: the case of Spain*

In an internationalised world economy where mobility flows have meant a sustained growth of the world tourism industry. Tourism has become one of the main drivers of growth in many regions, with a strong interaction between nature, industry prone to recession, terrorism or wars, natural disasters and infectious diseases (Hall, 2010; Cardenete et al., 2021; Pham et al., 2021). The contribution of this industry in the global Gross Domestic Product (GDP) has presented a sustained growth until the end of 2019, driven by the growth of the middle class in emerging economies, technological advances and bureaucratic simplification, among other factors. The tourism industry is one of the largest employers worldwide and the energy industry (ILO, 2020).

Despite the constant growth of the industry recently, it is fragile to exogenous factors that can mean negative perceptions about a balanced security environment such as strikes, wars, negative news about the destination area or citizen security. Studies have revealed the high exposure and vulnerability of the sector (Aliperti et al., 2019). The Severe Acute Respiratory Syndrome (SARS-CoV-2) virus that causes the COVID-19 disease is highly infectious and contagious. In addition to the damage to the tourism industry from SARS-CoV-2 (Zeng et al., 2005), there have been other epidemics such as Ebola (Novelli et al., 2018) or Foot and Mouth Disease (Frisby, 2003) The long-term consequences are unknown, and the dramatic effects on restrictions in social relationships due to social distance, mobility restrictions, protective equipment, transportation systems, hotel accommodation or events are yet to be determined in the tourism industry due to the crisis COVID-19. The studies indicate that tourism businesses have to pay attention to the change of personalised offers and towards digital technology (Abbas et al., 2021)

Since the end of 2019 and the beginning of 2020, the pandemic caused by COVID-19 has represented a negative exogenous shock to any previous modelling, which, although it may be unlikely, has been possible. Despite the gradual arrival of the vaccine to the countries, the dynamic adjustment of the sector is a progressive lifting of travel and social distance restrictions. Tourism supply chains may take years to adjust to the new circumstances (Gössling et al., 2020) The impact on the global economy due to the health emergency has been negative, especially in tourism (Zhang et al., 2021). The global tragedy with a very high number of infections and deaths has led to restrictions on the

international mobility of people. Exogenous conditions are unprecedented in recent industry analysis. According to official studies, the decline in the sector has meant different temporary recovery responses. We can say that tourism is a resilient economic sector and has shown rapid recoveries from exogenous shocks. Three international tourism crises have been the attacks of 11th September in NYC 2001 with an intermittent growth recovery in six months (Bonham et al., 2006), the SARS health crisis with a five-month recovery (Kuo et al., 2008) and the global economic-financial crisis of 2009 with the first month showing signs of recovery in the tenth month (UNWTO, 2013). Due to this, the analysis of this work is carried out in the context of international mobility without restrictions.

Spain, a country in southwestern Europe, finds itself in a political, social, economic, cultural and demographic environment conducive to developing the world's fastest-growing industry in the service sector. The entry of Spain into the European Union (EU) in 1986 represents a paradigm shift in the internal economy and the EU cluster due to the unrestricted mobility of people and capital in the environment. As a result, Spain in 2019 represents the second largest number of people in the world (82.7 million) of the arrival of people only behind France (89.4 million) and ahead of the world giant USA (79.6 million), in fourth position China with 62.9 million, in fifth Italy 62.1 and sixth position Turkey with 45.7 million people received (UNWTO, 2021). Figures from the World Economic Forum revealed that Spain is the most prepared country for the tourism industry with an index of 5.4 in 2019 ahead of economically more developed countries in other industries such as France, Germany, Japan, United States, United Kingdom, Australia, Italy, Canada or Switzerland (Uppink-Calderwood, 2019).

Official statistical sources indicate an approximate contribution of 12.4% to Spanish GDP (including 6.4% directly and 6% indirectly) and 12.8% of the workforce in the labour market for the year 2019. Since 2015, the tourism sector's contribution to the Spanish economy has grown by 1.3% and attained a figure of 2.72 million jobs in 2019. This figure is slightly lower (-0.1%) than in 2018. Similarly, workers in the tourism sector have grown by 0.8% since 2015 (INE, 2019). All this information analysed from Official Spanish data sources of the National Statistics Institute (INE) reveals significant growth in the sector, paralysed in 2020 by the global pandemic and with expectations of recovery in 2022 with the expectations of the frameworks designed in the literature (Rastegar et al., 2021).

The domestic tourism industry plays a pivotal role in the Spanish market (Arbulú et al., 2021). However, a broad vision of the international market allows the diversification of

potential consumers for a more significant contribution of the sector to GDP. Regarding the nationalities that visit Spain, the three largest emitting countries are the United Kingdom, Germany and France since 1999 according to the Big Data analysis, to a lesser extent under significantly Italy, the Netherlands or the USA (Ruiz Reina, 2021). The decision-making process is an added value in the tourist accommodation sector and the sectors involved. The exogenous shock experienced in 2020 in the Spanish economy and the consequent limitation of international mobility of potential consumers is not an obstacle to achieving a recovery in a similar way to the data prior to the crisis. The industry structure (accommodations, transport and service infrastructures) indicates the sector's recovery as soon as international mobility recovers to previous levels.

Macroeconomic conditions related to tourism and exogenous factors determine the contribution to GDP (Santamaria & Filis, 2019). The monthly study of the behaviour of international tourists with a tourist destination in Spain is essential to understand the market and make efficient decisions. This knowledge of the market and its granularity provide tools to the industry's economic agents of primary or secondary markets. In this way, the objective is to analyse from a "Big Data" approach to a "Small Data" approach for its understanding, overcoming the limitations of generality identified in the literature. (Weaver, 2021).

This doctoral thesis focuses on three chapters that make up the compendium to answer the scientific questions indicated at the beginning of this introductory section. In order to answer these questions applied to the field of the tourism industry, this thesis develops and innovates methodologies in the field of Data Science. The following subsections introduce the techniques developed to respond to all of the above.

1.4. Structure of the thesis

Once the importance of the data analysis and the tourism sector in the Spanish economy has been understood, it is necessary to underline that the tourism market is broad and made up of many economic agents that interact with each other. The complexity of the unsupervised analysis in a context of uncertainty gives value to this research to analyse the demand for tourist accommodation (Powell, 2019). The methodology of this thesis is described sequentially in the decision-making process contextualised in the life cycle of the data. In each chapter, it is described with an objective serving as the basis for further study development. In the following subsections, the three publications that compose the compendium of this thesis are introduced.

Three structural elements compose this research work; the first chapter demonstrates the causal relationships between internet searches and hotel demand. Once the relationship with Big Data has been demonstrated with primary data from official sources of the INE and secondary data from Google Trends, forecasting models are compared. A relative method of prediction accuracy is developed; right after the Information Theory chapter is introduced, this innovative line of research allows us to measure and quantify consumer decision-making. In addition, it allows us to know the existence of uncertainty cycles according to the country of origin by analysing the domain of time and frequency. Finally, in the third and final chapter, with the demonstrated knowledge and the methodological analysis developed, we highlight a new methodology of unsupervised clustering with seasonal series data to measure uncertainty in clients' decision-making process according to country of origin.

1.4.1.Forecasting with Big Data using Google Trends: keyword "visit Spain"

In this first introductory subsection, we try to present a methodology for analysing tourist data combined with keywords from previous searches on the internet, causality with the demand for accommodation for the first chapter. The amount of data generated in the network and communicated over the network exceeds the initially suspected limits. Initially, Big Data is linkable with large volumes of data and complex structures (Khoury & Ioannidis, 2014). Understanding the large volumes of data from social network data, commercial web pages, geographic data, among others, are the basis for decision-making by companies and consumers. In this way, it is possible to understand social changes and make predictions with non-primary quantitative data. In addition to the causal analysis explained with the Autorregressive Distributed Lag model extended to Seasonality (ARDL + seasonality), this chapter includes two novel contributions: Granger test-causality extended to seasonality, and accuracy matrix for forecasting called matrix U1 Theil that allows making comparisons between models in order to quantify and relativise models (Ruiz-Reina, 2019b). The decision matrix will allow us to reliably quantify the techniques used, being able to measure the quality of the proposed modelling of each model for the same comparative time horizon of tourism demand (G. Li et al., 2005; Song & Li, 2008; Peng et al., 2014; Jiao & Chen, 2019).

In the research carried out, we have analysed the correlation and causality of hotel demand with previous searches made by consumers. This study contributes to understanding the behaviour of demand through Google search engines. Our main goal was to understand the directions of causality between Big Data keywords to develop predictive models. Our study is based on developing and expanding previous models from the scientific literature

since 2008 using Google Trends (Jun et al., 2018). This research aims to provide modern information and communication generated by large volumes of tourism and tourist activities. The methodology developed allows interacting with potential consumers in the previous stages of tourist accommodation with the previously generated statistics for decision-making and forecast quantification.

Since 2006 Google has offered open data detailed analytics of its users' search terms. They found correlations in different fields, such as the unemployment rate and correlations between car sales and home purchases, among other fields (Jun et al., 2018). Researchers have used Google Trends as the leading search engine in the scientific literature applied to tourism and Baidu index as the central information server. Search engines show similarities and differences, highlighting the following as geographical area of use (Baidu only for China), data collection start (Google since 2004, Baidu since 2006) or indexing frequency (Google Trends: monthly and weekly; Baidu Weekly and daily) (J. Li et al., 2018). Considering the language / geographical limitation of Baidu and that according to official data from Spain, Chinese tourism does not represent a significant percentage of hotel overnight stays, we declined the use of the Chinese tool, focusing only on Big Data analysis from Google Trends. The conceptual map proposed for the analysis identifies essential tools for hotel demand in the Spanish market, covering needs in business intelligence with interdisciplinary analysis found in the scientific literature (Mariani et al., 2018).

The research works with data that allows obtaining knowledge of the behaviour of international tourism demand in Spain. In particular, forecasting methodologies are developed for the leading countries of origin that stay in Spanish hotels: Germany, France, Italy, the Netherlands, United Kingdom (UK), United States of America (USA) and a temporary variable added with the rest of countries called “residents abroad”. The analysis period is between January 2010 and June 2019, using a training period until December 2017. The out-sample training period is between January 2018 and June 2019, establishing evaluation windows with a time horizon of 3, 6, 12 and 18 months. The model developed is the ARDL + seasonality that allows us to study the dynamic elasticities (Peng et al., 2015), with a previous Granger-causality analysis extended to seasonality. The development of this seasonal methodology allows analysing the forecasting period in the short and long-term. We find a high predictive capacity in this sense which can be quantified with matrix U1 Theil developed in this work (Reina, 2020).

Regarding the measurement of prediction errors, the time series analysis is found in phenomena such as Statistics, Econometrics, Communications, Climate, Economy,

Finance, Machine Learning and other sciences with measurement purposes for the decision-making process. The first goal is to understand the phenomena in training periods and then evaluate their predictive periods in time horizons (Tsay, 2000). Predictions in social sciences are subject to many exogenous changes after the training period, such as outliers, shocks or structural changes (Inoue et al., 2017). It means that understanding prior data is essential to address specific problems. Thus, comparing methods with multiple measures can be cumbersome because of a single error measure (Armstrong & Collopy, 1992). The debate on error measurements has been extensive in the scientific literature, and no scientific consensus has been found at present (Makridakis et al., 2020).

The main error measurement measures can be differentiated into absolute or relative error criteria (Hyndman & Koehler, 2006). In the published work of the first chapter, predictive models are compared to quantify errors in different time horizons in hotel demand. To do this, we compare our ARDL + seasonality model with models such as Singular Spectrum Analysis (SSA) or Seasonal Autorregressive Integrated Moving Average (SARIMA). The training period for the predictions is from January 2018 to June 2019 (Reina, 2020).

Concluding with this initial work, it is shown that consumers carry out searches prior to their hotel accommodation decision and that with the use of this tool, they find a high explanatory capacity, giving rise to investigations that involve a more significant number of keywords in their hierarchical analysis (Ruiz-Reina, 2020; Reina, 2021c).

1.4.2.Measuring Uncertainty in Decision-Making

Once the causality among Google searches and hotel tourism demand has been demonstrated (Ruiz-Reina, 2020, 2019d; Reina, 2021c), the next step is to methodologically describe the agents' decision-making process with space-time data in a “big” to “small” data approach. Science reports have previously shown that non-trivial decision-making carries uncertainty (Stanton & Roelich, 2021). This uncertainty usually implies information gaps in decision-making that must be minimised. Economic agents have to maximise their benefits by minimising uncertainty. The lack of prior knowledge, the absence of prerequisites, and the seasonality analysis add difficulty to the analysis. Then, the methodological development of this second chapter models the spatio-temporal uncertainty, measuring and quantifying the decision-making of consumers of tourist accommodation according to country of origin for data from the INE. For this, the methodology developed is based on a static concept of Information Theory, particularly the Shannon Entropy (Shannon, 1948) under the hypothesis of adaptive expectations of

the markets. For this, understanding the concept of Entropy and maximum Entropy is the key (Delgado-Bonal & Marshak, 2019). This initial static analysis has allowed the construction of the Entropy time series and the verification of cyclical behaviours compatible with seasonality in tourism research (Chatfield & Baron, 1976; Smith, 2005; Vatsa, 2020). The pseudo-cycles identified with the methodology accurately determine the seasonal turning points in the decision-making of individuals in uncertain environments, finding a good performance in terms of precision (Camacho et al., 2020).

This novel modelling plays a successful role in the temporal analysis of the accommodation decision process between hotels and tourist apartments with seasonal cycles for data from the INE (Ruiz-Reina, 2021). They are expanding the spectrum of Information Theory analysis to fields unsuspected until Thermoeconomics (Saslow, 1999). In this sense, the Information Theory, the methodology, concept and solutions are widely recognised, understood and currently used (Golan & Maasoumi, 2008). The concepts of Entropy and their innovative applications to the field of decision-making suppose a new discipline of knowledge without limiting assumptions of laws and regularities of analysed data. The work scheme developed in this chapter allows the identification of information patterns unsuspected until now based on the concept of Maximum Entropy (Ruiz Reina, 2021). The study of the decision-making cycles between two possible temporal events, through the time domain using the correlogram and spectral analysis with the periodogram, contributes to the knowledge of amplitudes and frequencies of the identified cycles of consumer behaviour. The introduction of the heterodox concept supposes a contribution to Thermoeconomics by applying concepts of statistical mechanics to economic theory (Saslow, 1999).

Studying the theoretical aspects surrounding the Information Theory and Chaos Theory provides resources for analysing data from many disciplines (Delgado-Bonal & Marshak, 2019). To do this, in our work, we analyse time series without prior assumptions and establish causal time relationships. The causal relationships demonstrated in this work are not necessarily defined by an economic theory behind the events described. Instead, it is a causality of information transmission among systems (time series analysed) in the broad sense of the Shannon Entropy concept (Shannon, 1948). In particular, we propose a theoretical and empirical work scheme on uncertainty modelling for a temporal decision process. Causality in this work connected with Entropy is possible by defining time as a metric of causality in discrete time (Riek, 2020). In addition, this analysis must consider an assumption about the analysed sample, and in each period, the independent temporary random sample assumption must be fulfilled (Deaton, 1985). In this way, each period analysed individually would be independent of previous periods. For example, by the

definition of tourist accommodation, we assume that accommodation demands by nationalities of origin are made by temporarily independent individuals and share a cross-sectional characteristic such as nationality. Without performing an in-depth analysis, there is a simple analysis of pseudo-panel data with dependent analyses in the literature (Tovar et al., 2012) This type of data study could be framed within the pseudo-panel data with independent samples and lays the foundations for future research.

To understand the procedure and use of Entropy, we describe a mathematical and statistical scheme based on Information Theory. We describe a binary decision process between two mutually exclusive events and define the Entropy time series as a measure of uncertainty. In our work, the two processes are the choice of tourist accommodation among hotels and tourist apartments, the uncertainty measure being a factor that determines the transmission of information from one time series to another. Once the Entropy time series have been obtained, we continue to analyse the causal relationship between both variables to demonstrate linear (Granger-causality) or non-linear (Transfer Entropy) relationships in the context of decision-making uncertainty (Granger, 1969; Gençağa, 2018; Schreiber, 2000). In this sense, we seek to empirically demonstrate temporal causality between tourist accommodation in hotels and apartments. Once causality has been demonstrated, we analyse the time series of Shannon Entropy to determine the cycles of seasonal behaviour in the time domain (Box et al., 2013) and in the frequency domain (Harvey, 2006). Having demonstrated the existence of seasonal cycles and causal relationships, we model decision-making for accommodation using the Entropy uncertainty factor, obtaining a high explanatory capacity of the models for visiting tourists from Spain (Ruiz-Reina, 2021).

The modelling of this chapter shows relevant results in the cycles of seasonal behaviour in demand for tourist accommodation by country of origin. Among the relevant discoveries, the following stand out in summary: unit elasticities in decision exchanges between accommodation in tourist apartments and hotels for the period analysed; Entropy-based uncertainty factors indicate that when chaos increases, tourists prefer to stay in tourist apartments; finally, the study of cycles in the domain of time and frequency give rise to repetitive cycles according to the nationality of origin visiting tourist in Spain—having demonstrated these aspects of temporal causality and cycles of uncertainty. The study of this doctoral thesis continues with the innovative modelling of unsupervised spatio-temporal clustering based on seasonal Entropy cycles. This research represents a contribution in Spatial Statistics with theoretical and empirical implications in analysing tourism demand in contexts of uncertainty without prior assumptions for unsupervised clustering (Ruiz Reina, 2021). Finally, we can conclude that the

measurement of cyclical uncertainty makes it possible to quantify alternative concentrations to decomposition the Gini index in national and international markets (Fernández-Morales, 2021).

1.4.3.Clustering Spatio-temporal

The rise of Big Data technologies and their application in scientific business fields mean added value in decision-making processes. In the first two chapters, the theoretical methodology has initially focused on demonstrating temporal causality and measurement of uncertainty, applying it empirically to the Spanish tourist accommodation market according to the nationalities of origin of consumers in the last 15 years. Once the Entropy has been modelled, this chapter develops an innovative Spatio-temporal clustering methodology with seasonal patterns to group the homogeneous demands into conglomerates of unsupervised Entropy neighbours. This classification method based on a reference series allows the organisation and classification of information from the time series into flexible unsupervised groups.

This novel analysis deals with an unsupervised neighbourhood-based clustering modelling based on the median seasonal Entropy and contributes to Statistics and Spatial Economics. Since 1960, cluster analyses have tried to solve three questions: How many clusters are there? What is the best clustering algorithm? What should we do with the outlier values? The main goal of cluster analysis is to find groups with similar characteristics and more remarkable dissimilarities between the elements of each group (Luna-Romera et al., 2018). Time series clustering analyses became more popular since the 1990s, prompting the recent study of dynamic phenomena (Ruiz Reina, 2021). Emerging technologies and computing in a data-based knowledge environment justify clustering techniques to solve efficiency, quality, and complexity problems data analysis (Aghabozorgi et al., 2015).

The application and development of clustering methods have analysed the time domain, frequency domain, wavelet decomposition, or other transformations. In this third chapter, an unexplored framework for time series data with seasonality is developed whose clustering criterion is based on the median Entropy for seasonal cycle data. This unsupervised clustering analysis measures the median seasonal Entropy distances based on a reference series, this reference series assuming a centroid for the remainder of the analysis. This analysis implies developing a technique for non-Gaussian series based on Neighborhoods. Once the Neighborhoods are found, a similarity criterion called ‘‘coefficient of internal verification of the neighbourhood’’ is developed, this metric

chosen to evaluate the difference between data objects plays a fundamental role in clustering time series (Alonso et al., 2006). This foundation of space-time analysis is a consequence of the analysis of the two previous chapters that make up this thesis.

The seasonal space-time measurement based on Entropy develops a work scheme based on three steps: (1) data pre-processing, (2) uncertainty modelling, and (3) finally the clustering process. The first two steps are defined in the second chapter (Ruiz-Reina, 2021) and the third step is the one that makes up this third chapter (Ruiz Reina, 2021). This innovative methodology in Spatial Statistics, based on Information Theory, presents its empirical application in Spatial Economy applied to Spanish tourist accommodation time series. We have known the importance of the tourism sector and its growth in the Spanish and world economy. Its empirical application represents a decision-making knowledge tool for organisations based on the work scheme of Figure 1. The contribution of this clustering analysis can be exploited or compared with methodologies of main component analysis and on distances to a reference point (Blancas et al., 2010).

The decision-making grouping of tourist accommodation consumers among tourist apartments and hotels is based on the INE data's seasonal (temporal) and country of origin (spatial) criteria. The growing economic importance of the sector justifies the development of these techniques to improve efficiency in decision-making. It is a consequence of the analysis carried out in this thesis sequentially, assuming a comparative advantage of analysis and adaptation of commercial offers to potential consumers. After analysing more than 20 countries of origin, this study allows adjusting offers, adapting the labour market, identifying consumer behaviours in primary and secondary markets of the tourism industry based on seasonal Entropy measurement criteria.

Consequently, the sequential development of this third chapter presents contributions to theoretical statistics and economic implications in the Spanish tourism market. They are helping Big Data analysis tools such as Next Best Activity in the marketing process. This approach allows the prioritisation and subsequent selection of the best activity for a potential client at each moment of their demand process. Knowing the type of client in each season based on the clustering processes, the firms' campaign to be carried out is prioritised. Finally, depending on the sector involved in each stage, the activities will be carried out based on the results obtained by the classification methods for mutual benefit in decision-making.

1.5. The Added Value of this Thesis to the Scientific Field

The multidisciplinary analysis of this thesis attains a high level in several directions, combining methodology and empirical analysis. The combination of numerous tools adds value to Data Science. The methodological development allows understanding decision-making processes of tourist accommodation demand, translating into knowledge through their analysis. The study of the past and present provide tools for future decision-making.

The contribution of this thesis has involved adding theoretical and empirical analysis tools to the scientific field since the late 1990s. The process of digitisation of society and the impact is enormous for economic agents. The data revolution and its vast amounts are issues to be addressed. The spatial-temporal human behaviour data analysed in this thesis lack connection in the scientific literature. This thesis concatenates three chapters for the improvement of decision-making by economic agents. The data trail produced in the future by human behaviour and the accessibility of these data will be sources of analysis for this type of methodology developed in this research.

Big Data technologies provide researchers with valuable knowledge to establish profiles of consumer and business behaviour. The application schemes developed in the papers of this work direct strategies for fields such as Tourism, Finance, Sports, Health, among others. This information is widely regulated by Spanish and European legislation for any text, image, biometric or sound data. The European Union has focused its efforts on protecting this fundamental right, The General Data Protection Regulation (GDPR), the Data Protection Law Enforcement Directive, and other rules concerning personal data protection applied since May 23, 2018 (EU, 2016).

The techniques of causality, measurement of uncertainty and clustering developed in this thesis lay the groundwork for significant projects in the knowledge society. In this way, we can infer intelligence on the databases with the five v's mentioned at the beginning of this introductory section. This thesis has worked with Big Data from Google Trend technologies. However, applying these techniques could be inferred to the massive study of emails or social networks for researchers who have access to databases. Social interaction generates relevant information to define the behaviour patterns of people. Unstructured data such as videos, images, texts or locations can be classified to improve analysis. Structuring the data is not the objective of this thesis, but it is considered a previous step to the methodological framework developed to speed up and improve decision-making in the context of business intelligence. In this way, this thesis is framed within the analysis of the life cycle of the data in Figure 1. In the following sections of

this thesis, the three chapters are cited sequentially in the following sections, giving content to this thesis.

2. Chapter 1: Big Data: Forecasting and Control for Tourism Demand

2.1. Big Data: Forecasting and Control for Tourism Demand

Book Chapter (Peer Reviewed)

Book series: Contributions to Statistics

Book tittle: Theory and Applications of Time Series Analysis

Publisher: Springer Nature Switzerland AG

Year of publication: 2020

Indexed in Scholarly Publishers Indicators (SPI). The ICEE of this Editorial is 88 (position 2 of 26) in Economics 2018, (obtained from SPI Books in humanities and Social Sciences reviewed 19th January 2022)

Citation:

Reina, M.Á.R. (2020) Big Data: Forecasting and Control for Tourism Demand. In: Valenzuela O., Rojas F., Herrera L.J., Pomares H., Rojas I. (eds) Theory and Applications of Time Series Analysis. ITISE 2019. Contributions to Statistics. Springer, Cham. https://doi.org/10.1007/978-3-030-56219-9_18

Abstract

In this study, innovative forecasting techniques and data sources from Big Data are used for the study of Hotel Overnight Stays for Spain, from January 2018 to June 2019. The unstoppable development of the Tourism sector with the application of Big Data technologies, allow to make efficient decisions by economic agents. In this work, the use of the data collected from the Google Data Mining tools allows to obtain knowledge about Hotel Tourism Demand in Spain. The analysis carried out meets the four basic principles of Big Data analysis: volume, velocity, variety and veracity. In this setting, the methodology used corresponds to ARDL models, and ECM models being developed Granger-Causality extended to seasonality. The first one explains easily when economic agents will make their decisions; while the second one allows forecasting for short-term and long-term. This fact means that tourist offers and demands can be perfectly adjusted at every moment of the year. As a criterion for the selection of models, the innovative matrix U1 Theil is proposed, this allows to quantify how much a model is better than another in terms of forecasting.

Keywords: Big Data, Forecasting, Google trends

2.1.1.The BOOK CHAPTER is organised as follows:

2.1.1.1. Introduction

2.1.1.2. Literature Review

2.1.1.2.1. Forecasting Methods Using Search Engines (Google Trends)

2.1.1.3. Methodology

2.1.1.3.1. Modelling and Forecasting Evaluation

2.1.1.4. Data

2.1.1.5. Empirical Results

2.1.1.6. Conclusions

2.1.1.7. References

3. Chapter 2: Entropy Method for Decision-Making: Uncertainty Cycles in Tourism Demand

3.1. Entropy Method for Decision-Making: Uncertainty Cycles in Tourism Demand

Research paper (Peer Reviewed)

Scientific Journal: Entropy

Indexed in Journal Citation Reports (JCR). The Impact Factor of this journal is 2.587 in 2020, (obtained from Journal Impact Factor List- 2021)

Year of Publication: 2021

Citation:

Ruiz Reina, M.Á. Entropy Method for Decision-Making: Uncertainty Cycles in Tourism Demand. Entropy 2021, 23, 1370. https://doi.org/10.3390/e23111370

Abstract

A new methodology is presented for measuring, classifying and predicting the cycles of uncertainty that occur in temporary decision-making in the tourist accommodation market (apartments and hotels). Special attention is paid to the role of Entropy and cycles in the process under the Adaptive Markets Hypothesis. The work scheme analyses random cycles from time to time, and in the frequency domain, the linear and nonlinear causality relationships between variables are studied. The period analysed is from January 2005 to December 2018; the following empirical results stand out: (1) On longer scales, the periodicity of the uncertainty of decision-making is between 6 and 12 months, respectively, for all the nationalities described. (2) The elasticity of demand for tourist apartments is approximately 1% due to changes in demand for tourist hotels. (3) The elasticity of the uncertainty factor is highly correlated with the country of origin of tourists visiting Spain. For example, it has been empirically shown that increases of 1% in uncertainty cause increases in the demand for apartments of 2.12% (worldwide), 3.05% (UK), 1.91% (Germany), 1.78% (France), 7.21% (Ireland), 3.61% (The Netherlands) respectively. This modelling has an explanatory capacity of 99% in all the models analysed.

Keywords: Information Theory, Shannon Entropy, forecasting, decision-making, randomness, cycle, tourism

3.1.1.The RESEARCH PAPER is organised as follows:

3.1.1.1. Introduction

3.1.1.1.1. Tourism Sector, Gross Domestic Product and Randomness in Decision-Making

3.1.1.1.2. Literature Review

3.1.1.2. Material and Methods

3.1.1.2.1. Causality Testing: Linear and Nonlinear Relationships Data

3.1.1.2.1.1. Granger-Causality

3.1.1.2.1.2. Transfer Entropy

3.1.1.2.2. Information Theory: Shannon Entropy

3.1.1.2.3. Correlogram in the Time Domain and Cycles in the Frequency Domain

3.1.1.2.4. Causality Modelling

3.1.1.3. Results

3.1.1.3.1. Causality Testing

3.1.1.3.2. Randomness Measurement

3.1.1.3.3. Random Cycles

3.1.1.3.4. Causality Model and Forecasting

3.1.1.4. Theoretical Implications

3.1.1.5. Conclusions

3.1.1.6. Appendix A

3.1.1.7. Appendix B

4. Chapter 3: Spatio-temporal clustering: Neighbourhoods based on median seasonal Entropy

4.1. Spatio-temporal clustering: Neighbourhoods based on median seasonal Entropy

Research paper (Peer Reviewed)

Scientific Journal: Spatial Statistics

Indexed in Journal Citation Reports (JCR). The Impact Factor of this journal is 2.060 in 2020, (obtained from Journal Impact Factor List- 2021)

Year of publication: 2021

Citation:

Ruiz Reina, M. Á. (2021). Spatio-temporal clustering: Neighbourhoods based on median seasonal Entropy. Spatial Statistics, 45, 100535.
https://doi.org/https://doi.org/10.1016/j.spasta.2021.100535

Abstract

In this research, a new uncertainty clustering method has been developed and applied to the spatial time series with seasonality. The new unsupervised grouping method is based on Neighbourhoods and Median Seasonal Entropy. This classification method aims to discover similar behaviours for a time series group and find a dissimilarity measure concerning a reference series r. The Neighbourhood's Internal Verification Coefficient criterion makes it possible to measure intra-group similarity. This clustering criterion is flexible for spatial information. Our empirical approach allows us to measure accommodation decisions for tourists who visit Spain and decide to stay either in hotels or in tourist apartments. The results show the existence of dynamic seasonal patterns of behaviour. These insights support the decisions of economic agents.

Keywords: Spatial time series. Seasonal clustering, Entropy, Tourism economics, Neighbourhoods, Information Theory

4.1.1.The RESEARCH PAPER is organised as follows:

4.1.1.1. Introduction

4.1.1.1.1. The Motivation of the Technique and its Empirical Application

4.1.1.2. Methodology

4.1.1.2.1. Data Pre-processing

4.1.1.2.2. Uncertainty Modelling

4.1.1.2.3. Cluster Processing

4.1.1.3. A Case Study: Entropy in Decision-Making regarding the Tourist Demand for Spanish Accommodation

4.1.1.3.1. Theoretical Analysis Framework for Tourism Demand

4.1.1.3.2. Empirical Results

4.1.1.4. Discussions: Theoretical and Practical Implications in Tourism

4.1.1.4.1. Causality Modelling

4.1.1.5. Conclusions

4.1.1.6. Appendix A

4.1.1.7. Appendix B

5. Results, Discussion, Conclusions and Future Research Lines

5.1. Results and Discussion

Once all the research has been sequentially concatenated in the three chapters of this thesis, for the following, we will indicate the main results and the scientific discussion of the analysis under the criterion of parsimony. Starting with the first chapter, we can conclude a double aspect: the methodological component and the empirical results. These two aspects of the research have been supported by the well-known five v's of Big Data and give the following results: 1) A chance test extended to seasonality has been studied; 2) the relative measurement criterion based on the Theil ratio forming the matrix U1 Theil; 3) the economic modelling ARDL + seasonality with data from Big Data sources requires a high explanatory capacity of the model and improvements in forecasting; 4) the causal relationship among hotel demand and Google searches through a keyword with a seasonal component has been empirically demonstrated; 5) the cointegration relationships are statistically significant expressed in the Error Correction Model (ECM).

The methodology highlights the introduction of the ARDL + seasonality model; this model allows us to quantify the short-term relationships among endogenous and exogenous variables with a seasonal factor. In this way, it is possible to quantify the cyclical monthly effects. This causality relationship in the short and long-term (ECM) contrasts with the Granger-causality test extended to seasonality. It assumes an innovation compared to the traditional contrast proposed in the literature (Granger, 1969). To finalise the proposed methodology, it is worth highlighting the development of the modelling evaluation matrix called matrix U1 Theil. This matrix allows us to quantify the benefits of using our ARDL + seasonality model compared to other forecasting models in the literature such as SSA, SARIMA models, Hierarchical Neural Networks (HNN) or the multiplicative and additive versions of Holt-Winters (Ruiz-Reina, 2019b, 2019d, Reina, 2020). The development of this decision matrix for the evaluation of forecasting in the comparison of models under a relative criterion supposes to overcome certain limitations found in the literature and to avoid the appearance of "black swans" in the decision-making of the best model (Hyndman & Koehler, 2006; Makridakis et al., 2020).

From an empirical point of view, highlight primary data sources from the INE and secondary data from Google Trends. The Big Data tool from Google Trends favours the analysis of our modelling. The use of the keyword "visit Spain" as an exogenous variable shows a high explanatory capacity. With the application of the methodological approach, we were able to demonstrate that hotel accommodation consumers are relevant and that Big Data resources, in this case, demonstrate seasonal causality. The study applied to the main nationalities of origin that visit Spain reveals relationships in the short and long-

term among the hotel demand for visitors from Germany, France, Italy, the Netherlands, the UK and the USA. The explanatory capacity of the model in all cases is found with R-squared coefficients between 0.96 and 0.97. Matrix U1 Theil has allowed us to evaluate the predictive capacity of the modelling compared to other prediction models at time horizons of h = 3, 6, 12 and 18 months. The matrix results indicate that the best results are not always obtained by modelling the data from Big Data; in addition, the relationships in the long-term of cointegration expressed by the ECM explain future tourism demands. Finally, indicating that better models with less predictive capacity do not imply that they present less explanatory capacity with data from Big Data.

Once the existence of causal relationships between Big Data tools and Google trend data has been demonstrated, the work of Small Data in the decision-making process based on Entropy plays a significant role in the development of this thesis (Faraway & Augustin, 2018). Developing a working scheme is essential to understand the theoretical modelling of uncertainty measurement in the cycle theory. This work scheme allows solving uncertainty problems in the decision-making processes for two possible temporal events. The definition of the Entropy series and the chance contrast between two variables allows us to know two fundamental points: 1) the direction of causality between the variables to be studied; 2) the linear relationship or not for modelling the endogenous variable. The study of uncertainty cycles based on Entropy reveals the periodicity in which decisions are made. To finally model the causality following the work scheme for forecasting and control purposes.

The definition of the Entropy time series is based on a static context, and the proposal of this work allows one to know the possible unobserved cyclical behaviours of the series studied in the time domain using the correlogram and in the frequency domain using periodogram (Ruiz-Reina, 2019c). The causal model allows knowing the elasticities between exogenous and endogenous variables, and the uncertainty factor influences the decision-making of economic agents (Morikawa, 2020; Shoja & Soofi, 2017). The theoretical implications obtained with this modelling are relevant because it is possible to identify cycles in decision-making compatible with the seasonal patterns usually studied in the literature applied to tourism (Chatfield & Baron, 1976; Corluka, 2019).

The empirical results obtained for the main nationalities that visit Spain and spend the night (hotels and tourist apartments) are relevant for decision-making. We can indicate that for all the nationalities studied between January 2005 to December 2018, the causal relationships are unidirectional. In this way, first, it is theoretically demonstrated (Smallman & Moore, 2010) that hotel demand generates a secondary market for

accommodation in tourist apartments under the Adaptative Markets Hypothesis approach (Delgado-Bonal, 2019), in addition to being a linear relationship between both variables according to the main nationalities studied (UK, Germany, France, Ireland, The Netherlands and the rest of the world). Secondly, the models present tourist accommodation as a dependent variable and hotel demand as an explanatory variable; it should be noted that all have presented an expected positive sign among series with approximately unitary elasticity. The Entropy uncertainty factor shows a positive sign (Riek, 2020). From the latter, we can infer that the demand for tourist apartments is higher in the year in which uncertainty increases. The elasticities of the uncertainty factor the demand for tourist apartments are 2.12 rest of the world, 3.05 UK, 1.91 Germany, 1.78 France, 7.21 Ireland and 3.61 the Netherlands (Peng et al., 2015). Third, about the above, we can verify an empirical result with the theoretical demonstration; the Entropy grows with the increase in the probability of uncertainty, we find similarities with previous studies when the proportion is 0.5 in decision-making (Ruiz Reina, 2021). Fourth, we can quantify uncertainty cycles compatible with seasonality and previous literature results (Ruiz-Reina, 2019a). The most commonly encountered uncertainty cycles are between 6 and 12 months for all the nationalities studied except for Irish consumers, who present an additional 4-month cycle of uncertainty. In the short-term, this combination of intrinsic cycles coincides with long-period cycles is compatible with movements in the 174-month trend (Kaiser & Maravall, 1999). Finally, the empirical results have been obtained and contrasted with the theoretical work scheme on which the paper is based for the period analysed between January 2005 and December 2018 (Ruiz-Reina, 2021).

In the third chapter of the thesis, the development of the previous chapters is fundamental. Understanding causal relationships among Big Data, on the one hand, and on the other, the demonstration of uncertainty cycles is part of the data pre-processing with subsequent modelling, laying the foundations for understanding this thesis. Once the above is understood, the next step is developing the unsupervised seasonal clustering system based on medians of Entropy. This methodological contribution makes it possible to classify and organise individuals into groups based on their seasonal uncertainty. Mathematically initially, we must find a reference series that will be the dynamic centroid in a first step. The second part describes the dissimilarity measure called Median Seasonal Clustering Entropy against this reference series. Third, we classify clusters according to neighbourhood criteria whose intracluster similarity criterion is the Neighborhood's Internal Verification Coefficient representing robust data interpretation (Box, 1979).

As in previous chapters, we focus our empirical study on primary databases of the INE. It aims is to model the demand for tourist accommodation in hotels and apartments. In

this way, firms or organisations can intervene in each stage of the decision-making process for individualised tourism demand. This seasonal clustering allows knowing the behaviour of the accommodation decision for 20 nationalities that visited Spain between January 2005 and August 2019. The 20-uncertainty series analysed are divided into large geographical groupings (Other European Countries, Rest of the EU, Rest of the world, Africa) and the rest by countries such as Germany, Austria, Belgium, Denmark, USA, Finland, France, Greece, Ireland, Italy, Luxembourg, Norway, the Netherlands, Portugal, UK, and Sweden. According to official sources, this indicates that the highest number of monthly overnight stays occurred from the UK and Germany in the analysed period, the reference series finally being the global behaviour of all the countries. From the empirical results of this unsupervised clustering, the following stand out: 1) the decision processes under uncertainty vary seasonally, and the described technique is a robust tool, 2) the number of clusters is dynamic, varying the number of countries the number of clusters generated seasonally. We can conclude from this third chapter that the consequences of this analysis allow Stakeholders to make seasonal decisions based on Entropy as a measure of uncertainty and the individuality of the country of origin, with direct consequences on the primary market for hotel accommodation and secondary markets such as accommodation in tourist apartments, car rental, restaurants, business in the hotel accommodation environment. In addition, it has a direct implication in the labour market since the knowledge of the type of seasonal client will allow the hiring of qualified and adequate offer labour to the circumstances of the dynamic demand.

To conclude, the development of this thesis in the development of the chapters takes place from Big a to Small Data (Faraway & Augustin, 2018). As indicated, it started with a Big Data approach to finding causality between hotel demand and secondary data sources from Google Trends. Subsequently, the concept of uncertainty measurement has been worked on for its quantification, classification and empirically applied to the Spanish tourism market according to country of origin. The objective of this approach is the classification for the analysis of consumer behaviour in the tourism demand processes. From this, studies may arise related to the marginal propensity to consume, the calculation of the potential value of customers according to seasonal patterns or even regenerate predictive models with spatio-temporal behaviour patterns with the following purposes: pricing strategies, the market of work, or sociodemographic profiles.

This initial analysis, thought of Big Data and focused on Small Data, is carried out based on the principle of parsimony. The authors previously indicated that it is possible to represent a real-world with some simple model (Box, 1979). In particular, the criterion of parsimony used in this thesis is for several reasons: (i) simple reality illustrates,

complication obscures. In particular, we discovered causal relationships from Big Data keywords with a high explanatory capacity in modelling. With posterity, we developed an uncertainty methodology based on Entropy and seasonal clustering for the two main accommodation options highlighted by official Spanish statistics (hotels and tourist apartments); (ii) parsimony is typically rewarded by greater precision. Specifically, theoretical modelling describes seasonal behaviours with high precision and allows their classification to empower the decision-making process of organisations. (iii) In any case, indiscriminate modelling is not a practical option because this path is endless. In our case, the exogenous and uncontrollable factors mentioned in the introduction may be unlikely in a consolidated market such as tourism, where the legal environment, climate and industry structures will provide a speedy recovery after the COVID-19 crisis.

In practical terms, since this thesis is framed within a social sciences' scientific analysis, exogenous and uncontrollable factors can affect modelling. Practitioners and researchers can more easily access open database resources by facilitating modelling to fields of study. Theoretical modelling under the criterion of parsimony of this work allows obtaining robust results that contribute to science. The insensitivity can define the robustness of these results to deviations of the data in the analysis. Thus, the expansion of Information Theory overcomes the limitations of the usual ideal assumptions in econometric analyses. These ideal assumptions in scientific modelling can be widely simplified to understand the nature of the applied realities in this thesis. The need for robust, simple models for complex realities seems to be closely linked (Lisciandra, 2016).

Lastly, it should be noted that observing large volumes of data does not guarantee robust results for social realities. On certain occasions, the simplicity and parsimony criterion beats the analysis of large data sets that require a costly imbalance against large volumes of data. In terms of statistical inference, we can find fundamental multicollinearity problems or the inclusion of redundant variables, damaging the estimators' statistical analysis and properties (Faraway & Augustin, 2018). Due to the latter, an analysis at the microscopic level is justified in chapters 2 and 3 of this thesis.

5.2. Conclusions

This thesis has tried to analyse the reality of the Spanish tourist market and the decision-making between two possible options: accommodation in hotels or tourist apartments. The theoretical modelling of this thesis contributes to Spatial Statistics and the Information Theory of decision-making. On the other hand, empirical modelling applied to the tourism market provides robust results in decision-making modelling. It focuses

the analysis on the causality of keywords to model hotel demand, modelling uncertainty and clustering decision-making based on Information Theory as the central axis of this research. Data Science has been applied to gain knowledge by developing combined seasonal causality, Spatial Statistics, and Information Theory techniques. In this sense, the published works that make up this thesis are presented as novel contributions, filling gaps in the literature and opening new lines of future research.

This thesis has delved into the Spanish tourist accommodation market's theoretical and empirical research fields in the last 15 years. The combination of techniques allows understanding the decision-making procedure of consumers in order to achieve efficiency in decision-making by firms and organisations. Based on this, it has been demonstrated through a theoretical framework the previous search for information on the web before deciding on tourist accommodation for the main visiting nationalities in Spain. This analysis is completed with the final two chapters of this thesis; firstly, the modelling of uncertainty allows us to know how agents determine their decision to stay in tourist apartments or hotels. Finally, once we have modelled how agents seasonally decide their accommodation, we theoretically model the process of unsupervised clustering and similarity between groups to guarantee the homogeneity of foreign tourist groups in Spain. This analysis is valid regardless of structural or temporal changes in the series, mainly supported by the robust statistical procedures designed in this thesis.

It is necessary to emphasise that numerous theoretical and empirical recommendations are derived from this thesis. Its application or future development in the field of spatio-temporal Information Theory requires the involvement of investors (public and private institutions) and economic agents. The use of these empirical results represents a competitive advantage in the knowledge society. The use of empirical data in this thesis is open for the understanding or improvement of the theoretical and empirical analysis to obtain valid conclusions.

Another valid conclusion of this study is that individuals behave according to country of origin and that geographic proximity reveals seasonal demand behaviours. Assuming this, a powerful analysis tool in decision-making in the different stages of commercial strategies. Finally, the world reality of tourism is multivariate and immense in terms of data to be analysed. However, always under the criterion of parsimony, the methodological content of this thesis has been developed. Due to this, in the next section, we add future lines of research derived from this study.

5.3. Future Research Lines

This thesis supposes the first contribution of Information Theory in decision-making applied to the social sciences (tourism) with space-time data. The total absence of initial assumptions was the core of this thesis. In a context of unsupervised learning and without prior theoretical impositions, causality has been demonstrated among the existing information on the internet network collected by Google search engines and the actual demand for overnight hotel stays. This causality is the axis of work in which a linear temporal causal relationship has been demonstrated between the demand for hotels and tourist apartments in Spain for the different countries of origin. As a measure of uncertainty, entropy has allowed us to develop methodologies for measuring uncertainty cycles and spatio-temporal clustering with seasonality. Information Theory modelling has made a theoretical contribution in Data Science, Spatial Statistics and the theory of decision-making cycles in the domain of time and frequency. From the empirical point of view, the decision processes and behaviour of visiting tourists in Spain have been demonstrated in the period analysed. It means a contribution to a sector of constant growth in developed economies, particularly the Thermoeconomics. This contribution to Information Theory involves providing work structures among the agents participating in the decision-making process (Watson, 2019).

Although the theoretical and empirical results represent a contribution to science, the research branches derived from this thesis are numerous and by themselves would give rise to new doctoral theses, scientific papers or books. Taking into account that this thesis combines diverse knowledge of data analysis, business, Information Theory and Spatial Statistics. Any element outlined in figure 1 of the data life cycle represents a new line of research in its corresponding branch. For the recommendation of new lines of research, it is convenient to differentiate methodological aspects from an empirical point of view. We can highlight that the techniques can be applied to numerous lines of research.

The first chapter highlights that the techniques used can be a short or long-term analysis with relevant keywords. For this, it is recommended to use hierarchical decision algorithms based on keywords whose lines of research have already begun to be developed (Ruiz-Reina, 2020; Reina, 2021c). In addition, given the abundant and increasingly accessible information, it is recommended to add new models based on nowcasting and data flows in real-time with machine learning to forecasting techniques (Carrière-Swallow & Labbé, 2013; Richardson et al., 2021). Matrix U1 Theil assumes a contribution based on dimensionless selection criteria and allows quantifying the advantage of using some models for the model decision criteria. However, it is

recommended to deepen these techniques to overcome theoretical limitations as exact predictions (Hyndman & Koehler, 2005, 2006; Ruiz-Reina, 2019d; Makridakis et al., 2020).

On the other hand, from a statistical point of view, future researchers are recommended to delve into the idea of Small Data versus Big Data (Faraway & Augustin, 2018). In this way, it is possible to delve into classical problems of Econometrics such as collinearity or the inclusion of irrelevant/ redundant variables (Wooldridge, 2013). The empirical framework in the tourist accommodation market developed in this thesis using keywords from Google search engines can be extended to other fields of study such as Finance (Hu et al., 2018), Economy and growth (Niesert et al., 2020), insurance sector and unemployment (Aaronson et al., 2021), Mental health and emergency (Knipe et al., 2021), Psychiatry (Lopez-Agudo, 2020), pandemics (Simionescu & Raišienė, 2021), infections (Nishimura et al., 2021) among others. The research requirement, in this case, assumes that there are previous searches that generate Big Data content in Google search engines. In this way, in a digitalised and connected society, the lines of research are inexhaustible.

In the second chapter, Information Theory is the basis of the decision-making process. In particular, the core of work has been Shannon Entropy's classic concept of communication and perception (Tishby et al., 2011) this initial concept can be applied to other types of Entropy within Mechanical Statistics (Jaynes, 1957; Golan & Maasoumi, 2008). In particular, in this thesis, we have worked on a binary decision (two possible mutually exclusive events) of Shannon Entropy; this same analysis can be performed with the Bernoulli Entropy function. In the same way, the estimation methods could generate desirable properties based on the estimated parameters' consistency (Reina, 2021b). In addition, it is recommended to work with multivariate decision-making to broaden the spectrum of this primary work (Azami et al., 2019). The theoretical work scheme has been defined in time and frequency to analyse repetitive cycles; in this line, it is recommended to deepen the analysis of wavelets related to harmonic analysis (Fernández-Macho, 2018).

Regarding cycles, they have worked with short memory cycles; a possible line of research would be the study of seasonal long memory cycles (Arteche, 2007). These lines cited are the primary investigations that can expand this section. However, it is recommended to deepen the concept of Information Theory for subsequent contributions to the field of Econophysics applied to decision-making (Delgado-Bonal & Marshak, 2019). Regarding the empirical section, it should be noted that the decision of accommodation in hotels or tourist apartments has been worked on, but this can be extended to other options such as

rural accommodation or campsites, transport (car, train or aeroplane), destinations (beach, mountain or city), types of trips (business, leisure or cultural) and a wide range of possibilities. In addition, there are numerous empirical applications in fields such as finance (real estate investment in housing or rustic land), health (measurement of risk uncertainty), automobiles (selection of types of cars to buy), or any field of application that involves deciding in a context of uncertainty. All this serves as a working outline of what was developed in chapter 2 of this thesis.

The Entropy weight plays an essential role (third chapter) in decision-making, and the Entropy classification is a consequence of the uncertainty measurement (Yue, 2017). Unsupervised clustering criteria applied to time series are less developed than traditional analyses on cross-section data (Aghabozorgi et al., 2015). This type of analysis applied to Spatial Statistics is in itself new lines of research in multiple contexts. Entropy, complexity and spatial information based on the context of the Shannon Entropy are unsuspected fields in many scientific areas of study. In this way, the measurement of increasing information means adding complexity to the analysis in two-dimensional spatial systems (Batty et al., 2014). In this sense, the space-time study with seasonal data invites us to expand the fields of recent studies (Ruiz Reina, 2021). The technological improvement of data extraction, transformation and collection make Spatial Statistics applications a source of future research in any field (Gelfand, 2020): child mortality (Morales-Otero & Núñez-Antón, 2021), Big Data analysis (Banerjee, 2020), segmentation studies for international tourists for religious reasons according to spending levels (Mercadé-Melé & Barreal Pernas, 2021) or animal movements (Hooten et al., 2020). In short, Spatial Statistics has been placed among the first lines of research in the last 15 years; mainly, this thesis represents a beginning towards new frontiers of geospatial data in discrete time, for which additional lines would be the development of clustering models in continuous data. As in the previous chapters, the empirical applications are extensive and diverse to practitioners and researchers.

Finally, this thesis is an investigation for a better understanding of the decision-making processes unusual until now. This contribution has surpassed, contributed to and opened new frontiers of knowledge in the temporary causal relationships between information systems and clustering. Bibliographic references are relatively modern and involve up-to-date knowledge. This thesis as a whole can encourage future researchers to improve and understand the knowledge generated. The questions and answers of this thesis are the starting point for breaking the current frontiers of knowledge.

6. References

Aaronson, D., Brave, S. A., Butters, R. A., Fogarty, M., Sacks, D. W., & Seo, B. (2021). Forecasting unemployment insurance claims in real time with Google Trends. *International Journal of Forecasting*. https://doi.org/10.1016/J.IJFORECAST.2021.04.001

Abbas, J., Mubeen, R., Iorember, P. T., Raza, S., & Mamirkulova, G. (2021). Exploring the impact of COVID-19 on tourism: transformational potential and implications for a sustainable recovery of the travel and leisure industry. *Current Research in Behavioral Sciences*, *2*, 100033. https://doi.org/10.1016/j.crbeha.2021.100033

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering - A decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Aliperti, G., Sandholz, S., Hagenlocher, M., Rizzi, F., Frey, M., & Garschagen, M. (2019). Tourism, crisis, disaster: an interdisciplinary approach. *Annals of Tourism Research*, *79*. https://doi.org/10.1016/j.annals.2019.102808

Alonso, A. M., Berrendero, J. R., Hernández, A., & Justel, A. (2006). Time series clustering based on forecast densities. *Computational Statistics & Data Analysis*, *51*(2), 762–776. https://doi.org/10.1016/J.CSDA.2006.04.035

Alvarez-Diaz, M., D'Hombres, B., Ghisetti, C., & Pontarollo, N. (2020). Analysing domestic tourism flows at the provincial level in Spain by using spatial gravity models. *International Journal of Tourism Research*, *22*(4), 403–415. https://doi.org/10.1002/JTR.2344

Arbulú, I., Razumova, M., Rey-Maquieira, J., & Sastre, F. (2021). Can domestic tourism relieve the COVID-19 tourist industry crisis? The case of Spain. *Journal of Destination Marketing & Management*, *20*, 100568. https://doi.org/10.1016/J.JDMM.2021.100568

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, *08*, 69–80. https://doi.org/10.1016/0169-2070(92)90008-W

Arteche, J. (2007). The Analysis of Seasonal Long Memory: The Case of Spanish Inflation. *Oxford Bulletin of Economics and Statistics*, *69*(6), 749–772. https://doi.org/10.1111/J.1468-0084.2007.00478.X

Azami, H., Fernández, A., & Escudero, J. (2019). Multivariate Multiscale Dispersion Entropy of Biomedical Times Series. *Entropy 2019, Vol. 21, Page 913*, *21*(9), 913. https://doi.org/10.3390/E21090913

Banerjee, S. (2020). Modeling massive spatial datasets using a conjugate Bayesian linear modeling framework. *Spatial Statistics*, *37*, 100417. https://doi.org/10.1016/J.SPASTA.2020.100417

Batty, M., Morphet, R., Masucci, P., & Stanilov, K. (2014). Entropy, complexity, and spatial information. *Journal of Geographical Systems*, *16*(4), 363. https://doi.org/10.1007/S10109-014-0202-2

Blancas, F. J., González, M., Lozano-Oyola, M., & Pérez, F. (2010). The assessment of sustainable tourism: Application to Spanish coastal destinations. *Ecological Indicators*, *10*(2), 484–492. https://doi.org/10.1016/J.ECOLIND.2009.08.001

Bonham, C., Edmonds, C., & Mak, J. (2006). The Impact of 9/11 and Other Terrible Global Events on Tourism in the United States and Hawaii. *Journal of Travel Research*, *45*(1), 99–110. https://doi.org/10.1177/0047287506288812

Box, G. E. P. (1979). Robustness in the Strategy of Scientific Model Building. *Robustness in Statistics*, 201–236. https://doi.org/10.1016/B978-0-12-438150-6.50018-2

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2013). Time series analysis: Forecasting and control: Fourth edition. In *Time Series Analysis: Forecasting and Control: Fourth Edition*. https://doi.org/10.1002/9781118619193

Camacho, M., Gadea, M. D., & Loscos, A. G. (2020). A New Approach to Dating the Reference Cycle. *Journal of Business and Economic Statistics*. https://doi.org/10.1080/07350015.2020.1773834/SUPPL_FILE/UBES_A_1773834_SM1227.ZIP

Cardenete, M. A., Delgado, M. del C., & Villegas, P. (2021). Impact assessment of Covid-19 on the tourism sector in Andalusia: an economic approach. *Current Issues in Tourism*, 1–7. https://doi.org/10.1080/13683500.2021.1937073

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an Emerging Market. *Journal of Forecasting*, *32*(4), 289–298. https://doi.org/10.1002/FOR.1252

Chatfield, C., & Baron, R. R. v. (1976). Seasonality in Tourism. *Journal of the Royal Statistical Society. Series A (General)*. https://doi.org/10.2307/2344373

Corluka, G. (2019). Tourism Seasonality – An Overview. *Journal of Business Paradigms*.

Coussement, K., & Benoit, D. F. (2021). Interpretable data science for decision making. *Decision Support Systems*, *150*, 113664. https://doi.org/10.1016/J.DSS.2021.113664

Cox, M., & Ellsworth, david. (1997). Managing big data for scientific visualization. *ACM Siggraph*, *97*.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, *30*, 109–126. https://www.sciencedirect.com/science/article/pii/0304407685901344

Delgado-Bonal, A. (2019). Quantifying the randomness of the stock markets. *Scientific Reports 2019 9:1*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-49320-9

Delgado-Bonal, A., & Marshak, A. (2019). Approximate entropy and sample entropy: A comprehensive tutorial. In *Entropy* (pp. 1–37). https://doi.org/10.3390/e21060541

EU. (2016). *Data protection in the EU | European Commission*. https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

Faraway, J. J., & Augustin, N. H. (2018). When small data beats big data. *Statistics & Probability Letters*, *136*, 142–145. https://doi.org/10.1016/J.SPL.2018.02.031

Fernández-Macho, J. (2018). Time-localized wavelet multiple regression and correlation. *Physica A: Statistical Mechanics and Its Applications*, *492*, 1226–1238. https://doi.org/10.1016/J.PHYSA.2017.11.050

Fernández-Morales, A. (2021). Tourism Seasonality Across Markets. *Advances in Spatial Science*, 125–141. https://doi.org/10.1007/978-3-030-61274-0_7

Frisby, E. (2003). Communicating in a crisis: The British Tourist Authority's responses to the foot-and-mouth outbreak and 11th September, 2001. *Journal of Vacation Marketing*, *9*(1), 89–100. https://doi.org/10.1177/135676670200900107

Gelfand, A. E. (2020). Introduction to the special issue on frontiers in spatial research. *Spatial Statistics*, *37*, 100423. https://doi.org/10.1016/J.SPASTA.2020.100423

Gençağa, D. (2018). Transfer entropy. *Entropy*, *20*(288), 1–4. https://doi.org/10.3390/e20040288

Golan, A., & Maasoumi, E. (2008). Information Theoretic and Entropy Methods: an overview. *Econometric Reviews*, *27*(4–6), 317–328. https://doi.org/10.1080/07474930801959685

Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: a rapid assessment of COVID-19. *Journal of Sustainable Tourism*. https://doi.org/10.1080/09669582.2020.1758708

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049–1064. https://doi.org/10.1016/J.IM.2016.07.004

Hall, C. M. (2010). Crisis events in tourism: subjects of crisis in tourism. *Current Issues in Tourism*, *13*(5), 401–417. https://doi.org/10.1080/13683500.2010.491900

Harvey, A. (2006). Chapter 7 Forecasting with Unobserved Components Time Series Models. In *Handbook of Economic Forecasting*. https://doi.org/10.1016/S1574-0706(05)01007-4

Hooten, M. B., Lu, X., Garlick, M. J., & Powell, J. A. (2020). Animal movement models with mechanistic selection functions. *Spatial Statistics*, *37*, 100406. https://doi.org/10.1016/J.SPASTA.2019.100406

Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing*, *285*, 188–195. https://doi.org/10.1016/J.NEUCOM.2018.01.038

Hyndman, R. J., & Koehler, A. B. (2005). and Business Statistics Another Look at Measures of Forecast Accuracy Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(November), 679–688. http://www.sciencedirect.com/science/article/pii/S0169207006000239%5Cnhttp://core.ac.uk/download/pdf/6340761.pdf

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

INE. (2019). *INEbase / Services /Hotel Industry and Tourism /Spanish Tourism Satellite Account / Latest data*. Tourism Satellite Account of Spain. Year 2019. https://www.ine.es/dyngs/INEbase/en/operacion.htm?c=estadistica_C&cid=1254736169169&menu=ultiDatos&idp=1254735576863

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, *196*, 55–67. https://doi.org/10.1016/j.jeconom.2016.03.006

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, *106*(4), 620. https://doi.org/10.1103/PhysRev.106.620

Jiao, E. X., & Chen, J. L. (2019). Tourism forecasting: A review of methodological developments over the last decade. *Tourism Economics*. https://doi.org/10.1177/1354816618812588

Jun, S. P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, *130*, 69–87. https://doi.org/10.1016/J.TECHFORE.2017.11.009

Kaiser, R., & Maravall, A. (1999). *Short-term and long-term trends, seasonal adjustment, and the business cycle*. https://repositorio.bde.es/handle/123456789/6689

Khoury, M. J., & Ioannidis, J. P. A. (2014). Big data meets public health. *Science*, *346*(6213), 1054–1055. https://doi.org/10.1126/SCIENCE.AAA2709

Knipe, D., Gunnell, D., Evans, H., John, A., & Fancourt, D. (2021). Is Google Trends a useful tool for tracking mental and social distress during a public health emergency? A time–series analysis. *Journal of Affective Disorders*, *294*, 737–744. https://doi.org/10.1016/J.JAD.2021.06.086

Kuo, H.-I., Chen, C.-C., Tseng, W.-C., Ju, L.-F., & Huang, B.-W. (2008). Assessing impacts of SARS and Avian Flu on international tourism demand to Asia. *Tourism Management*, *29*(5), 917–928. https://doi.org/https://doi.org/10.1016/j.tourman.2007.10.006

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. In *Journal of Travel Research*. https://doi.org/10.1177/0047287505276594

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009

Lisciandra, C. (2016). Robustness analysis and tractability in modeling. *European Journal for Philosophy of Science 2016 7:1*, *7*(1), 79–95. https://doi.org/10.1007/S13194-016-0146-0

Lopez-Agudo, L. A. (2020). The association between internet searches and suicide in Spain. *Psychiatry Research*, *291*, 113215. https://doi.org/10.1016/J.PSYCHRES.2020.113215

Luna-Romera, J. M., García-Gutiérrez, J., Martínez-Ballesteros, M., & Riquelme Santos, J. C. (2018). An approach to validity indices for clustering techniques in Big Data. *Progress in Artificial Intelligence*, *7*, 81–94. https://doi.org/10.1007/s13748-017-0135-3

Makridakis, S., Hyndman, R. J., & Petropoulos, F. (2020). Forecasting in social settings: The state of the art. *International Journal of Forecasting*, *36*(1), 15–28. https://doi.org/10.1016/j.ijforecast.2019.05.011

Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. In *International Journal of Contemporary Hospitality Management*. https://doi.org/10.1108/IJCHM-07-2017-0461

Martínez, I., Viles, E., & G. Olaizola, I. (2021). Data Science Methodologies: Current Challenges and Future Approaches. *Big Data Research*, *24*, 100183. https://doi.org/10.1016/J.BDR.2020.100183

McKendry, D. A., Whitfield, R. I., & Duffy, A. H. B. (2021). Product Lifecycle Management implementation for high value Engineering to Order programmes: An informational perspective. *Journal of Industrial Information Integration*, 100264. https://doi.org/10.1016/J.JII.2021.100264

Mercadé-Melé, P., & Barreal Pernas, J. (2021). Study of expenditure and stay in the segmentation of the international tourist with religious motivation in Galicia. *Revista Galega de Economía*, *30*(3), 1–18. https://doi.org/10.15304/rge.30.3.7550

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and E-Business Management*, *16*(3), 547–578. https://doi.org/10.1007/s10257-017-0362-y

Morales-Otero, M., & Núñez-Antón, V. (2021). Comparing Bayesian Spatial Conditional Overdispersion and the Besag–York–Mollié Models: Application to Infant Mortality Rates. *Mathematics 2021, Vol. 9, Page 282*, *9*(3), 282. https://doi.org/10.3390/MATH9030282

Morikawa, M. (2020). Uncertainty in long-term macroeconomic forecasts: Ex post evaluation of forecasts by economics researchers. *Quarterly Review of Economics and Finance*. https://doi.org/10.1016/j.qref.2020.10.017

Niesert, R. F., Oorschot, J. A., Veldhuisen, C. P., Brons, K., & Lange, R. J. (2020). Can Google search data help predict macroeconomic series? *International Journal of Forecasting*, *36*(3), 1163–1172. https://doi.org/10.1016/J.IJFORECAST.2018.12.006

Nishimura, Y., Hagiya, H., Keitoku, K., Koyama, T., & Otsuka, F. (2021). Impact of the World Hand Hygiene and Global Handwashing Days on Public Awareness between 2016 and 2020: Google Trends Analysis. *American Journal of Infection Control*. https://doi.org/10.1016/J.AJIC.2021.08.033

Novelli, M., Gussing Burgess, L., Jones, A., & Ritchie, B. W. (2018). 'No Ebola…still doomed' – The Ebola-induced tourism crisis. *Annals of Tourism Research*, *70*, 76–87. https://doi.org/10.1016/j.annals.2018.03.006

Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*. https://doi.org/10.1016/j.tourman.2014.04.005

Peng, B., Song, H., Crouch, G. I., & Witt, S. F. (2015). A Meta-Analysis of International Tourism Demand Elasticities. *Journal of Travel Research*, *54*(5), 611–633. https://doi.org/10.1177/0047287514528283

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. ben, Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Oliveira, F. L. C., de Baets, S., Dokumentov, A., … Ziel, F. (2020). Forecasting: theory and practice. *Growth Dynamics*, *46*. https://arxiv.org/abs/2012.03854v3

Pham, T. D., Dwyer, L., Su, J. J., & Ngo, T. (2021). COVID-19 impacts of inbound tourism on Australian economy. *Annals of Tourism Research*, *88*, 103179. https://doi.org/10.1016/j.annals.2021.103179

Powell, W. B. (2019). A unified framework for stochastic optimization. *European Journal of Operational Research*, *275*(3), 795–821. https://doi.org/10.1016/J.EJOR.2018.07.014

Rastegar, R., Higgins-Desbiolles, F., & Ruhanen, L. (2021). COVID-19 and a justice framework to guide tourism recovery. *Annals of Tourism Research*, 103161. https://doi.org/10.1016/J.ANNALS.2021.103161

Reina, M. Á. R. (2020). Big Data: Forecasting and Control for Tourism Demand. In R. I. Valenzuela O., Rojas F., Herrera L.J., Pomares H. (Ed.), *Theory and Applications of Time Series Analysis. ITISE 2019* (pp. 273–286). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-030-56219-9_18

Reina, M. Á. R. (2021a). Cycles and Uncertainty: Applications in the Tourist Accommodation Market. *Engineering Proceedings 2021, Vol. 5, Page 3*, *5*(1), 3. https://doi.org/10.3390/ENGPROC2021005003

Reina, M. Á. R. (2021b). Bernoulli Time Series Modelling with Application to Accommodation Tourism Demand. *Engineering Proceedings 2021, Vol. 5, Page 17*, *5*(1), 17. https://doi.org/10.3390/ENGPROC2021005017

Reina, M. Á. R. (2021c). Tourism and Big Data: Forecasting with Hierarchical and Sequential Cluster Analysis. *Engineering Proceedings 2021, Vol. 5, Page 14*, *5*(1), 14. https://doi.org/10.3390/ENGPROC2021005014

Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, *37*(2), 941–948. https://doi.org/10.1016/J.IJFORECAST.2020.10.005

Riek, R. (2020). Entropy Derived from Causality. *Entropy 2020, Vol. 22, Page 647*, *22*(6), 647. https://doi.org/10.3390/E22060647

Ruiz Reina, M. Á. (2021). Spatio-temporal clustering: Neighbourhoods based on median seasonal entropy. *Spatial Statistics*, *45*, 100535. https://doi.org/https://doi.org/10.1016/j.spasta.2021.100535

Ruiz-Reina, M. Á. (2019a). Entropy of Tourism: the unseen side of tourism accommodation. *Proceedings of the International Conference on Applied Research in Business, Management and Economics*. https://www.dpublication.com/wp-content/uploads/2019/12/424.pdf

Ruiz-Reina, M. Á. (2020, December). Google Trends and Tourism: Regression Cluster Analysis. *Proceedings of The 11th International Conference on Modern Research in Management, Economics and Accounting*.

Ruiz-Reina, M. Á. (2021). Entropy Method for Decision-Making: Uncertainty Cycles in Tourism Demand. *Entropy 2021, Vol. 23, Page 1370*, *23*(11), 1370. https://doi.org/10.3390/E23111370

Ruiz-Reina, M. Á. (2019b). Big Data: Does it really improve Forecasting techniques for Tourism Demand in Spain? In I. Rojas (Ed.), *International Conference on Time Series and Forecasting* (pp. 694–706). Godel Impresiones Digitales S.L. https://itise.ugr.es/ITISE2019_Vol1.pdf

Ruiz-Reina, M. Á. (2019c). Entropy of Tourism: the unseen side of tourism accommodation. In Diamond Scientific Publications (Ed.), *Proceedings of the International Conference on Applied Research in Business, Management and Economics*. https://www.dpublication.com/wp-content/uploads/2019/12/424.pdf

Ruiz-Reina, M. Á. (2019d). Forecasting using Big Data: The case of Spanish Tourism Demand. *International Conference on Time Series and Forecasting*, 782–789. https://itise.ugr.es/ITISE2019_Vol2.pdf

Santamaria, D., & Filis, G. (2019). Tourism demand and economic growth in Spain: New insights based on the yield curve. *Tourism Management*, *75*, 447–459. https://doi.org/10.1016/J.TOURMAN.2019.06.008

Saslow, W. M. (1999). An economic analogy to thermodynamics. *American Journal of Physics*, *67*(12), 1239. https://doi.org/10.1119/1.19110

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, *85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shoja, M., & Soofi, E. S. (2017). Uncertainty, information, and disagreement of economic forecasters. *Econometric Reviews*, *36*(6–9), 796–817. https://doi.org/10.1080/07474938.2017.1307577

Simionescu, M., & Raišienė, A. G. (2021). A bridge between sentiment indicators: What does Google Trends tell us about COVID-19 pandemic and employment expectations in the EU new member states? *Technological Forecasting and Social Change*, *173*, 121170. https://doi.org/10.1016/J.TECHFORE.2021.121170

Smallman, C., & Moore, K. (2010). Process studies of tourists' decision-making. *Annals of Tourism Research*, *37*(2), 397–422. https://doi.org/10.1016/j.annals.2009.10.014

Smith, W. W. (2005). Seasonality in Tourism. *Annals of Tourism Research*. https://doi.org/10.1016/j.annals.2004.10.001

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting-A review of recent research. *Tourism Management*, *29*(2), 203–220. https://doi.org/10.1016/j.tourman.2007.07.016

Stanton, M. C. B., & Roelich, K. (2021). Decision making under deep uncertainties: A review of the applicability of methods in practice. *Technological Forecasting and Social Change*, *171*, 120939. https://doi.org/10.1016/J.TECHFORE.2021.120939

Timakum, T., Lee, S., & Song, M. (2021). Exploring the research landscape of data warehousing and mining based on DaWaK Conference full-text articles. *Data & Knowledge Engineering*, *135*, 101926. https://doi.org/10.1016/J.DATAK.2021.101926

Tishby, N., Polani, D., & Tishby, N. (2011). Information Theory of Decisions and Actions. *Perception-Action Cycle*, 601–636. https://doi.org/10.1007/978-1-4419-1452-1_19

Tovar, A. O., Zulaica, I. G., & Núñez-Antón, V. (2012). Analysis of pseudo-panel data with dependent samples. *Http://Dx.Doi.Org/10.1080/02664763.2012.696593*, *39*(9), 1921–1937. https://doi.org/10.1080/02664763.2012.696593

Tsay, R. S. (2000). Time series and forecasting: Brief history and future research. *Journal of the American Statistical Association*, *95*(450), 638–643. https://doi.org/10.1080/01621459.2000.10474241

UNWTO. (2013). *Economic Crisis, International Tourism Decline and its Impact on the Poor*. World Tourism Organization (UNWTO). https://doi.org/10.18111/9789284414444

UNWTO. (2021, November 10). *UNWTO statistics*. Https://Www.Unwto.Org/Statistics. https://www.unwto.org/statistics

Uppink-Calderwood, L. S. M. (2019). *The Travel & Tourism Competitiveness Report 2019 Travel and Tourism at a Tipping Point*. http://www3.weforum.org/docs/WEF_TTCR_2019.pdf

Vatsa, P. (2020). Seasonality and cycles in tourism demand—redux. *Annals of Tourism Research*, 103105. https://doi.org/10.1016/J.ANNALS.2020.103105

Watson, M. D. (2019). Information Theory Applied to Decision-Making Structures. In *Systems Engineering in Context*. Springer, Cham. https://doi.org/10.1007/978-3-030-00114-8_42

Weaver, A. (2021). Tourism, big data, and a crisis of analysis. *Annals of Tourism Research*, *88*, 103158. https://doi.org/10.1016/J.ANNALS.2021.103158

Wooldridge, J. M. (2013). Introductory econometrics: a modern approach / Jeffrey M. Wooldridge. In *Introductory econometrics: a modern approach*.

Wu, D. C., Wu, J., & Song, H. (2021). Special Issue: Big Data Analytics and Forecasting in Hospitality and Tourism. *International Journal of Contemporary Hospitality Management*, *33*(6), 1917–1921. https://doi.org/10.1108/ijchm-06-2021-035

Yue, C. (2017). Entropy-based weights on decision makers in group decision-making setting with hybrid preference representations. *Applied Soft Computing*, *60*, 737–749. https://doi.org/10.1016/J.ASOC.2017.07.033

Zeng, B., Carter, R. W., & de Lacy, T. (2005). Short-term perturbations and tourism effects: The case of SARS in China. In *Current Issues in Tourism* (Vol. 8, Issue 4, pp. 306–322). Multilingual Matters Ltd. https://doi.org/10.1080/13683500508668220

Zhang, H., Song, H., Wen, L., & Liu, C. (2021). Forecasting tourism recovery amid COVID-19. *Annals of Tourism Research*, *87*. https://doi.org/10.1016/j.annals.2021.103149

Appendix A: SUMMARY IN SPANISH (RESUMEN EN ESPAÑOL)

Esta tesis se encuadra en el marco metodológico de la Ciencia de Datos envolviendo métodos científicos, procesos y sistemas para extraer conocimiento con aplicaciones directas en la toma de decisiones (Coussement & Benoit, 2021). La investigación consiste en el aprendizaje y la representación de patrones de datos, y viceversa, las caracterizaciones identificadas y modeladas permiten tomar decisiones en grandes escalas. Para ello se desarrollan metodologías novedosas de análisis en campos como Big Data, Teoría de la Información, Estadística espacial, Series Temporales, Econometría y analítica predictiva para empoderar la toma de decisiones de los agentes económicos. Las técnicas de pronóstico siempre han estado en primer plano en el contexto de planificación de las organizaciones. Por esto, los individuos toman decisiones en un contexto de incertidumbre, dónde tratan de minimizar su riesgo maximizando beneficios particulares (Petropoulos et al., 2020). En este sentido, esta tesis doctoral se considera una contribución en los campos metodológicos y empíricos de la ciencia, cubriendo lagunas de información en la literatura científica (Martínez et al., 2021). La aplicación empírica de este extenso trabajo se analiza con datos mensuales de alojamiento turístico para visitantes internacionales con datos procedentes del Instituto Nacional de Estadística español (INE) en el periodo desde 2005 a 2019. Esta tesis es metodológica y provee de herramientas de análisis para organizaciones públicas o privadas. Este análisis persigue los siguientes objetivos (Reina, 2020; Ruiz Reina, 2021; Ruiz-Reina, 2021): (i) demostrar teóricamente que se producen previamente al alojamiento turístico búsquedas en internet con los motores de búsqueda de Google Trends con el uso de palabras clave; (ii) demostrar relaciones causales temporales en la toma de decisión del alojamiento turístico entre hoteles y apartamentos. A través de este estudio, podremos modelizar la demanda en un mercado secundario a través de un mercado primario de alojamiento. Los test de causalidad lineal y no lineal temporal nos determinarán el sentido de la demanda de alojamientos. El estudio de ciclos de incertidumbre en el dominio en el tiempo y la frecuencia proporcionarán información estacional sobre comportamientos de turistas según el país de origen; (iii) finalmente, desarrollaremos métodos de agrupamiento (clustering) espacio temporal no supervisados. Estos métodos no supervisados permitirán ordenar espacial y temporalmente las demandas turísticas con el objetivo de realizar una intervención basada en el conocimiento obtenido por las técnicas desarrolladas. La combinación de este análisis no supervisado y en un contexto sin suposiciones limitantes, permitirá a las empresas y organizaciones tomar decisiones en eficiencia en contextos de incertidumbre, complejidad e información espacial (Batty et al., 2014).

Esta tesis doctoral tratará de responder formalmente a preguntas científicas relacionadas con la industria turística: “Big Data (capítulo 1)” — ¿Existe relación causal entre la

demanda turística y Big Data (Google Trends) ?, Análisis estacional de búsquedas en Google, modelización y criterios de selección de modelos predictivos; "Information Theory and decision-making (capítulo 2)" — ¿Existe relación entre la demanda de hoteles y apartamentos turísticos?, ¿existe casualidad lineal o no? ¿Se producen ciclos de incertidumbre en el comportamiento en la demanda?, ¿Todas las nacionalidades de turistas presentan la misma estacionalidad?; "Clustering Spatio-Temporal (capítulo 3)" — ¿Cómo agrupamos a las demandas turísticas según comportamiento estacional para alojamientos en hoteles y apartamentos?, ¿Se identifican los ciclos de estacionalidad y los comportamientos en la toma de decisiones?

Para el desarrollo de esta tesis doctoral, es fundamental el entendimiento de grandes volúmenes y si la combinación de esta se realiza con un sector en auge como es el turismo, los beneficios de este análisis son aún mayores. El rol del análisis Big Data y el entendimiento del ciclo de vida del dato es un eje central en el desarrollo de esta tesis. La estructuración del dato respalda la comunicación efectiva de la modelización, mejora la calidad de toma de decisiones, disminuye tiempo de aprobación em la presentación de resultados a las personas decisoras (McKendry et al., 2021).

La aplicación de un concepto genérico como es el Big Data en un mundo digitalizado responde a una cuestión previa basada en la toma de decisiones. Desde que en 1997 los investigadores de la NASA Michael Cox y David Ellsworth usaron el término "Big Data" por primera vez, la literatura científica ha superado límites insospechados. Los propios investigadores resaltaban la aplicación de este concepto a numerosas áreas como base del entendimiento en la visualización de problemas complejos (Cox & Ellsworth, 1997). El entendimiento de problemas complejos permite conocer y entender los retos en aras tomar decisiones en eficiencia.

La aplicación de Big Data en un entorno digital supone una ventaja competitiva frente a los que no usan analítica de datos avanzada. Este uso juega un rol principal para personas y compañías diariamente revelando ventajas competitivas. La combinación de las cinco v's del Big Data (volumen, velocidad, valor, variedad y veracidad) ha supuesto una rápida expansión de las técnicas de almacenado de datos, análisis y visualización (Mikalef et al., 2018) Los mecanismos de entendimiento y procesado de datos a través de los cuales se basa el Big Data son la base del valor añadido para los agentes económicos. La literatura científica se ha basado en todos los aspectos recogidos bajo el concepto de Big Data suponiendo inversiones en infraestructura, inteligencia de negocio o herramientas de análisis en áreas no desarrolladas hasta el momento (Gupta & George, 2016).

Este rápido desarrollo del Big Data en la industria turística ha normalizado y mejorado las técnicas para la toma de decisiones (J. Li et al., 2018). La información generada en la red de internet y la estructuración de los grandes volúmenes de datos suponen una herramienta insospechada en los inicios del mercado turístico internacional. Los agentes económicos, en particular empresas y agencias, han descubierto métodos de interactuación con los consumidores potenciales. Este intercambio de información a través de los datos permite tomar decisiones en eficiencia en cada momento determinado. En la figura 2 se observa el denominado ciclo de vida del dato en la demanda turística, siendo la base del entendimiento del entorno Big Data para académicos y profesionales (Ruiz-Reina, 2019d).

De esta Figura 2 se extrae todo el conocimiento desarrollado en esta tesis doctoral, conviene describir el esquema para un mejor entendimiento de los capítulos que componen dicho trabajo. Inicialmente comenzaremos a analizar el esquema desde el punto de vista ad-hoc del conjunto de datos disponibles en el "Data Warehouse" (Timakum et al., 2021). En este sentido trabajaremos con dos tipos de datos: datos estructurados (INE) y datos inicialmente no estructurados (Google Trends). Estos datos no estructurados inicialmente son estructurados por las arquitecturas de ingeniería de Google suponiendo un recurso ad-hoc en esta investigación. El siguiente punto del esquema es el denominado "Analytics: Modelling and Forecasting", este es el esquema central de trabajo de la tesis. En particular se subdivide en tres capítulos: "Descriptive and Predictive Analytics (Chapter 1): Big Data — Forecasting and Control for Tourism Demand"; "Uncertainty Modelling (Chapter 2): Entropy Method for Decision-Making: Uncertainty Cycles in Tourism Demand"; "Prescriptive Analytics (Chapter 3): Spatio-temporal clustering: Neighbourhoods based on median seasonal entropy".

Los tres capítulos de la tesis, se desarrollan extensamente en los dos artículos y el capítulo de libro que componen este compendio de publicaciones. Esta modelización permitirá a los implicados tomar decisiones en eficiencia en términos de conocimiento, mejoras de procesos, reducción de costes, ahorro de tiempo u optimización de administración. Con esta información las firmas podrán estructurar su oferta turística en la web incluyendo alojamientos, transportes y servicios adicionales para los potenciales consumidores (Ruiz Reina, 2021). Una vez que los consumidores disponen de esa información en la red de internet, realizan búsquedas en Google dónde encontrarán la oferta los potenciales consumidores. De este modo emerge finalmente la demanda turística, generándose y retroalimentándose los nuevos datos en la red que volverían a ser analizados en base a los comportamientos del ciclo de vida del dato descrito anteriormente en la demanda turística.

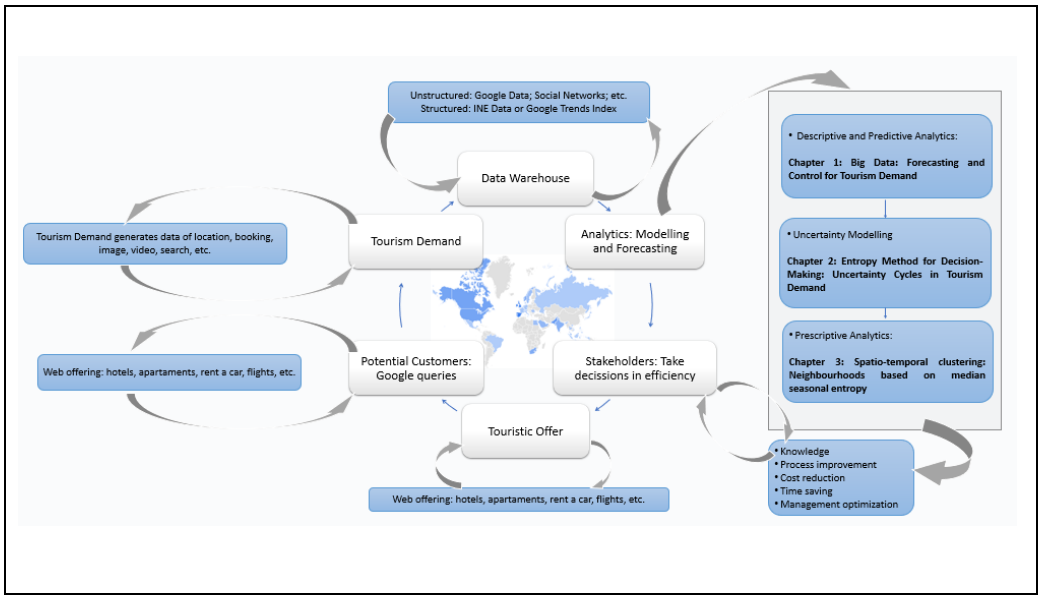

Fig. 2 El ciclo de vida del dato en la demanda turística. Elaboración propia (en inglés).

En la última década, el interés en Big Data y analítica de datos ha sido creciente con la finalidad descriptiva y predictiva de la demanda turística. La ciencia de datos aplicada en este trabajo resume y visualiza la información obtenida de los datos para obtener conclusiones en la toma de decisiones. Los datos producen conocimiento y esto supone un valor añadido en la interpretación de patrones matemáticos a las organizaciones. Sin embargo, a pesar de la amplia literatura escrita existen temas no abordados que suponen el reto de desarrollo de este trabajo (Wu et al., 2021).

El análisis Big Data aplicado a la industria turística supone conocer patrones de comportamiento de los individuos, en ocasiones se aplican tratamientos generalistas y objetivistas de personas perjudicando la sensación de bienestar del consumidor (Weaver, 2021). Con las metodologías desarrolladas en este trabajo, se pretende obtener el mayor conocimiento con la finalidad de adecuar al máximo los patrones de los turistas según país de origen. Una vez que el lector ha entendido cuál es el esquema general de trabajo para la modelización de la demanda turística, conviene resaltar la importancia del sector turístico en las economías y en particular en la española.

En una economía mundial internacionalizada donde los flujos de movilidad han supuesto un crecimiento sostenido de la industria turística mundial desde hace décadas. El turismo llega a ser uno de los principales conductores de crecimiento en muchas regiones con una fuerte interacción con la naturaleza, siendo una industria propensa a la recesión, problemas de terrorismo o guerras, desastres naturales y enfermedades infecciosas (Hall,

2010; Cardenete et al., 2021; Pham et al., 2021). La contribución de esta industria en el Producto Interior bruto (PIB) mundial ha presentado un crecimiento sostenido hasta finales de 2019 impulsado por el crecimiento de la clase media en economías emergentes, avances tecnológicos y simplificación burocrática entre otros factores. La industria turística es una de las mayores empleadoras a nivel mundial junto con la industria energética (ILO, 2020).

A pesar del crecimiento constante de la industria recientemente, es muy frágil a factores exógenos que puedan significar percepciones negativas sobre un ambiente de equilibrio de seguridad como huelgas, guerras, noticias negativas sobre el área de destino o de seguridad ciudadana. Los estudios han revelado la alta exposición y vulnerabilidad del sector (Aliperti et al., 2019). Del virus SARS-CoV-2 y COVID-19 debido a su alta infecciosidad e índice de contagio han limitado su crecimiento. Además de los daños en la industria turística del SARS-CoV-2 (Zeng et al., 2005), han existido otras epidemias tales como Ébola (Novelli et al., 2018) u otras enfermedades contagiosas (Frisby, 2003) Las consecuencias en el largo plazo son desconocidas y los efectos dramáticos sobre las restricciones en relaciones sociales por distancia social, restricciones de movilidad, el uso equipos de protección, efectos en sistemas de transportes, alojamientos hoteleros o eventos están aún por determinar en la industria turística debido a la crisis COVID-19. Los estudios indican que los negocios turísticos han de prestar atención en el cambio de oferta personalizada y hacia la tecnología digital (Abbas et al., 2021)

Desde finales de 2019 e inicios 2020 la pandemia causada por la COVID-19 ha supuesto un impacto exógeno negativo a cualquier modelado previo, que pudiendo ser poco probable ha sido posible. A pesar de la llegada paulatina de la vacuna a los países el ajuste dinámico del sector es progresivo con restricciones de viajes y distancia social, la oferta turística puede tardar en ajustarse a las nuevas circunstancias años (Gössling et al., 2020). El impacto en la economía global debido a la emergencia sanitaria ha sido negativo y especialmente en el turismo (Zhang et al., 2021). La tragedia mundial con altísimo número de infectados y muertes han supuesto restricciones de movilidad internacional de personas. Esto último supone impacto sobre las condiciones de estabilidad experimentadas previamente a la crisis y sin antecedentes por su volumen de implicación mundial. De acuerdo a cifras oficiales el decrecimiento del sector ha significado distintas respuestas temporales de recuperación. Podemos decir que el turismo es un sector económico resiliente y ha mostrado rápidas recuperaciones ante impactos exógenos. Tres crisis internacionales del turismo han sido relevantes: los ataques del 11 septiembre en la ciudad de Nueva York en 2001 con una recuperación intermitente de crecimiento en seis meses (Bonham et al., 2006); la crisis sanitaria del SARS-CoV-2 con una recuperación

de cinco meses (Kuo et al., 2008) y la crisis financiera económica global de 2009 con el primer mes con señales de recuperación en el décimo mes (UNWTO, 2013). Debido a estas recuperaciones, el análisis de este trabajo se realiza en un contexto de movilidad internacional sin restricciones.

España, país en el suroeste de Europa, se encuentra en un entorno político, social, económico, cultural y demográfico propicio para el desarrollo de la industria con mayor crecimiento mundial en el sector servicios. La entrada de España en la Unión Europea (UE) en el año 1986 supone un cambio de paradigma en la economía interna y del conjunto UE debido a la libre movilidad de personas y capitales en el entorno. A consecuencia de esto, España en el año 2019 representa la segunda mayor cifra mundial (82.7 millones) de llegada de personas únicamente por detrás de Francia (89,4 millones) y por delante del gigante mundial como los EEUU (79.6 millones), en cuarta posición se encuentra China con 62.9 millones, en quinta Italia 62.1 y sexta posición Turquía con 45.7 millones de personas recibidas (UNWTO, 2021). Las cifras del Foro Económico Mundial revelan que España es el país más preparado para la industria turística con un índice de 5.4 en 2019 por delante de países económicamente más desarrollados en otras industrias tales como Francia, Alemania, Japón, Estados Unidos, Reino Unido, Australia, Italia, Canadá o Suiza (Uppink-Calderwood, 2019).

Las fuentes estadísticas oficiales indican una contribución aproximada del 12.4% al GDP español (incluyendo 6.4% directamente y 6% indirectamente) y en el mercado laboral un 12.8% de la fuerza de trabajo para el año 2019. Desde el año 2015, la contribución del sector turístico a la economía española ha crecido por 1.3% y alcanzando una cifra de 2.72 millones de trabajos en 2019. Esta cifra ligeramente inferior (-0.1%) a la de 2018. De igual modo los trabajadores del sector turístico han crecido por 0.8% desde el año 2015 (INE, 2019). Toda esta información analizada de fuentes de datos oficiales españolas del INE revela un crecimiento significativo del sector, paralizado en el año 2020 por la pandemia mundial y con expectativas de recuperación en el año 2022 con las expectativas de los informes diseñados en la literatura (Rastegar et al., 2021).

La industria doméstica turística presenta un papel crucial en el mercado español (Arbulú et al., 2021), no obstante, una visión amplia del mercado internacional permite la diversificación de potenciales consumidores para una mayor aportación del sector al PIB. En cuanto a las nacionalidades que visitan España los tres mayores países emisores son en este orden Reino Unido, Alemania y Francia desde el año 1999 según fuentes oficiales, en menor medida pero muy significativa también Italia, Países Bajos o USA (Ruiz Reina, 2021). El análisis de un mercado global y su toma de decisiones según tipo de alojamiento

proveen de conocimiento a los interesados en este mercado. El impacto exógeno experimentado en 2020 en la economía española y la consecuente limitación de movilidad internacional de los potenciales consumidores no es óbice para conseguir una recuperación de modo similar a los datos previos a la crisis. La estructura de la industria, las infraestructuras de transportes y servicios, hacen indicar la recuperación de la movilidad internacional hasta niveles previos a la pandemia.

Debido a la importancia económica que representa la industria turística y a un peso mayor del sector servicios en las economías modernas (Santamaria & Filis, 2019), el estudio mensual del comportamiento de turistas internacionales con destino turístico en España es fundamental para entender el mercado y tomar decisiones en eficiencia. Este conocimiento del mercado y su desagregación, permiten proveer de herramientas a los agentes económicos de mercados primarios o secundarios en la Industria. De este modo el objetivo es analizar desde un enfoque "Big Data" a un enfoque "Small Data" para su entendimiento superando las limitaciones de generalidad identificadas en la literatura (Weaver, 2021).

Esta tesis doctoral se enfoca en tres capítulos que conforman el compendio para responder a preguntas científicas indicadas al inicio de este resumen. Una vez entendida la importancia del análisis de datos y del sector turístico en la economía española. Sin embargo, la realidad del mercado turístico es tan amplia y compuesto por muchos agentes económicos que interactúan entre sí. La complejidad del análisis no supervisado en un contexto de incertidumbre da valor a esta investigación para el análisis de la demanda de alojamientos turísticos (Powell, 2019). La metodología de esta tesis se describe secuencialmente en el proceso de toma de decisiones contextualizado en el ciclo de vida del dato. En cada capítulo se describe con un objetivo sirviendo de base para el estudio posterior desarrollo. En los siguientes párrafos se ofrece un resumen de las tres publicaciones que conforman el compendio de esta Tesis.

Para comenzar con los tres elementos estructurales, el primer capítulo está dedicado a demostrar las relaciones causales entre búsquedas en internet y demanda hotelera. Una vez demostrada la relación con Big Data con datos primarios de fuentes oficiales de estadística (INE) y secundarios de Google Trends se comparan modelos de forecasting. Para medir la bondad del forecasting se desarrolla un método relativo de acuracidad de las predicciones. Justo después, se introduce el capítulo de Teoría de la Información, esta línea de investigación innovadora en el campo de procesos de decisión contribuyendo a medir y cuantificar la toma de decisiones. Además, nos permite conocer la existencia de ciclos de incertidumbre según el país de origen analizando el dominio del tiempo y de la

frecuencia. Finalmente, en el tercer y último capítulo, con el conocimiento demostrado y el análisis metodológico desarrollado destacamos una nueva metodología de agrupación no supervisado con datos de series estacionales para la medición de incertidumbre en el proceso de decisión de los clientes turísticos según país de origen.

En este primer capítulo tratamos de exponer una metodología para el análisis de datos turísticos combinados con palabras clave procedentes de búsquedas previas en la red de internet, causalidad con la demanda de alojamientos. La cantidad de datos que son generados en la red y comunicados sobre la red superan los límites sospechados inicialmente. El entendimiento de los grandes volúmenes de datos procedentes de datos redes sociales, páginas web comerciales, datos geográficos entre otros suponen la base de la toma de decisiones de empresas y consumidores. De este modo, es posible entender los cambios sociales y realizar predicciones con datos cuantitativos no primarios. Además del análisis causal explicado con el modelo Autorregresivo de Retardos Distribuidos extendido a la estacionalidad (en inglés ARDL + seasonality), este capítulo incluye dos novedosas contribuciones: test de Granger- causalidad extendido a la estacionalidad una matriz de acuracidad para predicciones con denominada matriz U1 Theil que permite realizar comparaciones entre modelos de con la finalidad de cuantificar y relativizar modelos de pronóstico (Ruiz-Reina, 2019b). La matriz de decisión nos permitirá cuantificar de un modo fiable las técnicas utilizadas, pudiendo medirse la calidad del modelado propuesto de cada modelo para un mismo comparativo horizonte temporal de la demanda turística (G. Li et al., 2005; Song & Li, 2008; Peng et al., 2014; Jiao & Chen, 2019).

En la investigación llevada a cabo, hemos analizado la correlación y causalidad de la demanda hotelera con las búsquedas previas que han realizado los consumidores a través de una palabra clave por país de origen. Este estudio supone una contribución para el entendimiento del comportamiento de la demanda a través de los motores de búsqueda de datos abiertos de Google. Nuestra meta principal fue entender las direcciones de causalidad entre las palabras clave del Big Data para desarrollar modelos predictivos. Nuestro estudio está basado en desarrollar y ampliar modelados previos de la literatura científica desde 2008 usando Google Trends (Jun et al., 2018). La finalidad de esta investigación es proporcionar una moderna información y comunicación generada por grandes volúmenes de datos sobre el turismo y las actividades turísticas. La metodología desarrollada permite interactuar con potenciales consumidores en los estados previos de alojamiento turístico con las estadísticas generadas previamente con fines de tomas de decisiones y cuantificación del pronóstico.

Desde el año 2006 Google ofrece analíticas detalladas de los términos de búsquedas de sus usuarios de modo abierto. Encontrándose correlaciones en distintos campos como la ratio de desempleo, correlaciones entre ventas de coches y compras de casas, entre otros campos (Jun et al., 2018). En la literatura científica aplicada al turismo, los investigadores han usado como principales motores de búsqueda Google Trends e índices de Baidu como principales servidores de información. Los buscadores presentan similitudes entre sí y diferencias destacando las siguientes como área geográfica de uso (Baidu únicamente es para China), inicio de recogida de datos (Google desde 2004, Baidu desde 2006) o frecuencia de indexado (Google Trends: mensual y semanal; Baidu Semanal y diaria) (J. Li et al., 2018). Teniendo en cuenta la limitación idiomática/geográfica de Baidu y que según los datos oficiales de España el turismo chino no representa un porcentaje importante en las pernoctaciones hoteleras, declinamos el uso de la herramienta China, centrándonos únicamente en el análisis de Big Data procedente de Google Trends. El mapa conceptual propuesto para el análisis identifica herramientas claves para la demanda hotelera en el mercado español cubriendo necesidades en la inteligencia de negocio con análisis interdisciplinar encontrados en la literatura científica (Mariani et al., 2018).

En la investigación se trabaja con datos que permiten obtener conocimiento del comportamiento la demanda turística internacional en España. En particular se desarrollan metodologías de predicción para los principales países de origen que se alojan en hoteles españoles: Alemania, Francia, Italia, los Países Bajos, Reino Unido (RU), los Estados Unidos de América y una variable temporal agregada con el resto de países denominada “residents abroad”. El periodo de análisis es entre enero 2010 y junio 2019, utilizando un periodo de entrenamiento hasta diciembre 2017. El periodo de entrenamiento predictivo es entre enero 2018 a junio 2019 estableciendo ventanas de evaluación con horizonte temporal de 3, 6, 12 y 18 meses. El modelo desarrollado es el ARDL + seasonality nos permite estudiar las elasticidades dinámicas (Peng et al., 2015), con un previo análisis de Granger-causalidad extendido a la estacionalidad. El desarrollo de esta metodología estacional permite analizar el periodo de pronóstico en el corto y largo plazo. En este sentido encontramos una alta capacidad predictiva del modelo pudiéndose cuantificar con la matriz U1 de Theil desarrollada en este trabajo (Reina, 2020).

En cuanto a la medición de los errores de predicción, el análisis de las Series Temporales se encuentra en fenómenos tales como Estadística, Econometría, Comunicaciones, Clima, Economía, Finanzas, Machine Learning y otras ciencias con finalidades de medición para el proceso de toma de decisiones. La primera meta es entender los fenómenos en periodos de entrenamiento, para luego evaluar sus periodos predictivos en horizontes temporales

(Tsay, 2000). Las predicciones en ciencias sociales están sujetas a multitud de cambios exógenos tras el periodo de entrenamiento tales como outliers, shocks o cambios estructurales (Inoue et al., 2017). Esto significa que el entendimiento de datos previos es fundamental para abordar ciertos problemas. De este modo, la comparación entre métodos con múltiples medidas puede resultar engorrosa debido a la no existencia de una única medida de error (Armstrong & Collopy, 1992). En la literatura científica el debate sobre mediciones del error ha sido extenso, y no se ha encontrado consenso científico actualmente (Makridakis et al., 2020).

De la metodología desarrollada en este capítulo, destacar la introducción del modelo ARDL + seasonality, este modelo nos permite cuantificar las relaciones a corto plazo entre las variables endógenas y las exógenas con factor estacional mensual en el corto y largo plazo (error correction term). Cabe destacar el desarrollo de la matriz de evaluación de modelado denominada matriz U1 Theil. Esta matriz permite cuantificar los beneficios del utilizar nuestro modelo ARDL + seasonality frente a otros modelos de forecasting de la literatura como Singular Spectrum Analysis, modelos SARIMA, redes Neuronales Jerárquicas (HNN) o las versiones multiplicativas y aditivas de alisado Holt-Winters (Ruiz-Reina, 2019b, 2019d; Reina, 2020). El desarrollo de esta matriz de decisión para la evaluación de forecasting en la comparación de modelos bajo un criterio relativo, supone superar ciertas limitaciones encontradas en la literatura y evitar las apariciones de "black swans" en la toma de decisiones del mejor modelo (Hyndman & Koehler, 2006; Makridakis et al., 2020).

Desde un punto de vista empírico, destacar el uso de fuentes de datos primarias procedentes del INE y de datos secundarios de Google Trends. La herramienta Big Data de Google Trends favorece el análisis de nuestro modelado. El uso de la palabra clave "visit Spain" como variable exógena muestra una alta capacidad explicativa. El estudio aplicado a las principales nacionalidades de origen que visitan España revela relaciones en el corto plazo y largo plazo entre la demanda hotelera para los visitantes de Alemania, Francia, Italia, los Países bajos, Reino unido y los Estados Unidos. La capacidad explicativa del modelo en todos los casos se encuentra con coeficientes de determinación comprendidos entre 0.96 y 0.97. La matriz U1 de Theil nos ha permitido evaluar la capacidad predictiva del modelado frente a otros modelos de predicción en horizontes temporales de h= 3, 6, 12 y 18 meses. Los resultados de la matriz U1 Theil nos indica que no siempre se obtienen los mejores resultados modelizando los datos procedentes de Big Data, además las relaciones en el largo plazo de cointegración expresadas por el ECM explican las demandas turísticas futuras. Por último, indicar que existan mejores modelos

con menos capacidad predictiva, no implica que presenten menor capacidad explicativa con datos procedentes de Big Data.

Concluyendo con este trabajo inicial, queda demostrado que los consumidores realizan búsquedas previamente a su decisión de alojamiento hotelero, y que con el uso de esta herramienta encuentra una alta capacidad explicativa dando lugar a investigaciones que implican un mayor número de palabras clave en su análisis jerárquico (Ruiz-Reina, 2020; Reina, 2021c).

Una vez demostrada la causalidad entre las búsquedas de Google y la demanda turística hotelera (Ruiz-Reina, 2020, 2019d; Reina, 2021c). El siguiente paso es describir metodológicamente el proceso de toma de decisiones de los turistas con datos espacio temporales en enfoque "Big" a "small" de los datos. La ciencia ha demostrado previamente que la toma de decisiones no triviales conlleva incertidumbre (Stanton & Roelich, 2021). Esta incertidumbre habitualmente implica lagunas de información en la toma de decisiones y que han de ser minimizadas. La falta de conocimiento previo, la ausencia de prerrequisitos y el análisis de la estacionalidad añaden dificultad en el análisis. Entonces, el desarrollo metodológico de este segundo capítulo modeliza la incertidumbre espacio temporal midiendo y cuantificando la toma de decisiones de los consumidores de alojamientos turísticos según país de origen para datos del INE. Para ello la metodología desarrollada se basa en un concepto estático de Teoría de la Información, en particular la entropía de Shannon (Shannon, 1948) bajo la hipótesis de expectativas adaptativas de los mercados. Dando lugar al entendimiento del concepto de entropía y máxima entropía como eje vertebrador (Delgado-Bonal & Marshak, 2019). Este análisis estático inicial ha permitido construir las series temporales entropía y verificarse comportamientos cíclicos compatibles con la estacionalidad en el área de investigación del turismo (Chatfield & Baron, 1976; Smith, 2005; Vatsa, 2020).

Esta modelización novedosa juega un rol exitoso en el análisis temporal del proceso de decisión de alojamiento entre hoteles y apartamentos turísticos con ciclos estacionales para datos procedentes del INE (Ruiz-Reina, 2021). Ampliando el espectro del análisis de la Teoría de la Información a campos insospechados hasta el momento de la Termoeconomía (Saslow, 1999). En este sentido la Teoría de la Información, la metodología, concepto y soluciones son ampliamente reconocidos, entendidos y empleados actualmente (Golan & Maasoumi, 2008). Los conceptos de entropía sus innovadoras aplicaciones al campo de toma de decisiones suponen una nueva disciplina de conocimiento sin supuestos restrictivos y regularidades de datos analizados. El esquema de trabajo desarrollado en este capítulo permite identificar patrones de

información insospechados hasta la actualidad basados en el concepto de máxima entropía (Ruiz Reina, 2021). El estudio de los ciclos de toma de decisiones entre dos posibles eventos temporales, a través del dominio del tiempo (usando el correlograma) y a través del análisis espectral (con el periodograma) son una contribución en el conocimiento de amplitudes y frecuencias de los ciclos identificados. En relación a esto, la introducción del concepto heterodoxo supone una contribución a la Termoeconomía aplicando conceptos de mecánica estadística a la teoría económica (Saslow, 1999).

El estudio de los aspectos teóricos que envuelven la Teoría de la Información y Teoría del Caos, provee de recursos al análisis de datos de un gran número de disciplinas (Delgado-Bonal & Marshak, 2019). Para ello, en nuestro trabajo, analizamos series temporales sin suposiciones previas y estableciendo relaciones temporales causales. Las relaciones de causalidad demostradas en este trabajo no se definen necesariamente por la existencia de una teoría económica tras los hechos descritos, más bien es una causalidad de transmisión de información entre sistemas (entre series analizadas) en el sentido amplio del concepto de entropía Shannon (Shannon, 1948). En particular, proponemos un esquema de trabajo teórico y empírico sobre el modelado de incertidumbre para un proceso de decisión temporal. La causalidad en este trabajo conectada con entropía es posible al definir el tiempo como métrica de causalidad en tiempo discreto (Riek, 2020). Además, este análisis debe tener en cuenta una suposición sobre la muestra analizada, y en cada periodo temporal, el supuesto de muestra aleatoria temporal independiente debe de cumplirse (Deaton, 1985). De este modo, cada periodo analizado individualmente sería independientemente de periodos anteriores. Por ejemplo, por la propia definición de alojamiento turístico, suponemos que las demandas de alojamientos por nacionalidades de origen lo realizan individuos independientes temporalmente y compartiendo una característica de sección cruzada como la nacionalidad de origen.

Para el entendimiento del procedimiento y del uso de la Entropía, describimos un esquema matemático y estadístico basado en Teoría de la Información. Describimos un proceso de decisión binario entre dos eventos mutuamente excluyentes y definimos la serie temporal de entropía como medida de incertidumbre. En nuestro trabajo, los dos procesos son las elecciones de alojamiento turístico entre hoteles y apartamentos turísticos siendo la medida de incertidumbre un factor que determina transmisión de información de una serie temporal a otra. Una vez obtenidas las series temporales de entropía, continuamos analizando la relación causal entre ambas variables para demostrar relaciones lineales (Granger-causalidad) o no lineales (Transfer entropy) en el contexto de incertidumbre de la toma de decisión (Granger, 1969; Schreiber, 2000; Gençağa, 2018). En este sentido buscamos demostrar empíricamente en qué sentido existe la

causalidad temporal entre alojamientos turísticos en hoteles y apartamentos. Demostrada la causalidad, analizamos las series de entropía para determinar los ciclos de comportamiento estacional en el dominio del tiempo (Box et al., 2013) y en el dominio de la frecuencia (Harvey, 2006). Tras verificar la existencia de ciclos estacionales y las relaciones causales, modelizamos la toma de decisiones para alojamientos usando el factor de incertidumbre de entropía obteniendo una alta capacidad explicativa de los modelos para los turistas visitantes de España (Ruiz-Reina, 2021). Usando la entropía como medida de incertidumbre en la toma de decisiones de los consumidores, se observan claramente patrones de comportamientos cíclicos compatibles con la estacionalidad (Ruiz-Reina, 2019c; Reina, 2021a, 2021b).

Los resultados empíricos obtenidos para las principales nacionalidades que visitan España y que pernoctan (hoteles y apartamentos turísticos) son relevantes para el conocimiento de la toma de decisiones. Podemos indicar que para todas las nacionalidades estudiadas entre el periodo de enero 2005 a diciembre 2018 las relaciones causales son unidireccionales. De este modo, primero, se demuestra teóricamente (Smallman & Moore, 2010) que la demanda hotelera genera un mercado secundario de alojamiento en apartamentos turísticos bajo el enfoque de hipótesis de adaptativo del mercado (Delgado-Bonal, 2019), además de ser una relación lineal entre ambas variables según las principales nacionalidades estudiadas (Reno Unido, Alemania, Francia, Irlanda, los Países bajos y "resto del mundo"). Segundo, los modelos presentan como variable dependiente a los alojamientos turísticos y como variable explicativa a la demanda hotelera, de esto cabe destacar que todos han presentado un signo positivo esperado entre ambas series con elasticidad aproximadamente unitaria y que el factor de incertidumbre de entropía muestra un signo positivo (Riek, 2020). De esto último podemos inferir que en los periodos del año en los que la incertidumbre aumenta la demanda de apartamentos turísticos es mayor. Las elasticidades de los factores de incertidumbre frente a la demanda de apartamentos turísticos son: 2.12 resto del mundo, 3.05 Reino Unido, 1.91 Alemania, 1.78 Francia, 7,21 Irlanda y 3.61 los Países bajos (Peng et al., 2015). Tercero, en relación a lo anterior, podemos verificar un resultado empírico con la demostración teórica, la entropía crece con el incremento de probabilidad de incertidumbre, encontramos similitudes con estudios previos cuando la proporción es de 0.5 en la toma de decisiones (Ruiz Reina, 2021). Cuarto, a nivel microscópico podemos cuantificar los ciclos de incertidumbre compatibles con la estacionalidad y con resultados previos de la literatura (Ruiz-Reina, 2019a). Los ciclos de incertidumbre más comúnmente encontrados son de periodicidad 6 y 12 meses para todas las nacionalidades estudiadas con la excepción de los consumidores de Irlanda que presentan un ciclo de incertidumbre adicional de 4

meses. Esta combinación de ciclos intrínsecos en el corto plazo coincide con ciclos de periodo largo compatibles con los movimientos de la tendencia de 174 meses (Kaiser & Maravall, 1999).

La modelización de este capítulo muestra resultados relevantes para el periodo analizado entre enero 2005 y diciembre 2018 (Ruiz-Reina, 2021) en los ciclos de comportamiento estacional en la demanda de alojamiento turístico según país de origen. Entre los descubrimientos relevantes destacan resumidamente los siguientes: elasticidades unitarias en los intercambios de decisiones entre alojamiento de apartamentos turísticos y hoteles para el periodo analizado; los factores de incertidumbre basado en entropía indican que cuando el chaos aumenta, los turistas prefieren alojarse en apartamentos turísticos; finalmente, el estudio de los ciclos en el dominio del tiempo y la frecuencia dan lugar a ciclos repetitivos según la nacionalidad de origen del turista visitante en España.

Habiendo demostrado estos aspectos de causalidad temporal y de ciclos de incertidumbre, el estudio de esta tesis doctoral continúa con la modelización innovadora de agrupamientos no supervisados en el dominio espacio temporal basada en los ciclos de entropía estacionales. Dicha investigación supone una contribución en el área de Estadística espacial con implicaciones teórico y empíricas en el análisis de la demanda turística en contextos de incertidumbre sin suposiciones previas para clustering no supervisado (Ruiz Reina, 2021). La metodología desarrollada agrupa demandas y patrones estacionales homogéneos en conglomerados de vecinos de entropía no supervisados. Este método de clasificación basados en una serie de referencia, permite la organización y clasificación de información de las time series en grupos flexibles no supervisados.

Desde 1960, los análisis de agrupamiento tratan de resolver tres tipos de cuestiones: ¿Cuántos clusters hay? ¿cuál es el mejor algoritmo de clustering? ¿qué debemos de hacer con los valores outliers? La principal meta del análisis de agrupamiento es encontrar grupos con similares características entre grupos y mayores disimilitudes entre los elementos de cada grupo (Luna-Romera et al., 2018). Una aplicación desde los inicios del análisis es la aplicación a datos de sección cruzada, la aplicación de estas técnicas de clasificación desde inicio de los 90 para datos de series temporales (Ruiz Reina, 2021). Las tecnologías emergentes y la computación en un entorno de conocimiento basados en datos, justifican las técnicas de clustering para solucionar problemas de eficiencia, calidad y complejidad del análisis de datos (Aghabozorgi et al., 2015).

La aplicación y desarrollo de métodos de clustering se han basado en el análisis del dominio del tiempo, el dominio de la frecuencia, descomposición de ondas u otro tipo de transformaciones. En este tercer capítulo se desarrolla un marco de trabajo no explorado para datos de time series con estacionalidad cuyo criterio de agrupamiento está basado en la entropía mediana para datos de ciclos estacionales. Este análisis de clustering no supervisado mide las distancias medianas de entropía estacional basadas en una serie de referencia, suponiendo esta serie de referencia un centroide para el resto del análisis. Este análisis implica desarrollar una técnica para series no Gaussianas basadas en vecindarios y un criterio de similitud denominado ‘‘coefficient of internal verification of the neighbourhood’’.

La medición espacio temporal estacional basada en la entropía desarrolla un esquema de trabajo basado en tres pasos: (1) pre-procesado de datos, (2) modelización de la incertidumbre, y (3) finalmente el proceso de agrupamiento. Los dos primeros pasos se definen en el segundo capítulo (Ruiz-Reina, 2021) y el tercer paso es el que compone este tercer capítulo (Ruiz Reina, 2021). Esta metodología innovadora en el campo de la Estadística espacial basada en Teoría de la Información, presenta su aplicación empírica en el campo de la economía espacial aplicada a las series temporales del alojamiento turístico español. Conocida la importancia del sector turístico y el crecimiento de esta en la economía española y mundial. Su aplicación empírica representa una herramienta de conocimiento de decision-making para las organizaciones basado en el esquema de trabajo de la figura 2.

El agrupamiento de toma de decisiones de los consumidores de alojamiento turístico entre apartamentos turísticos y hoteles basados en criterios estacionales (temporales) y de país de origen (espaciales) para los datos del INE. La importancia económica creciente del sector justifica el desarrollo de estas técnicas en aras de mejorar la eficiencia en la toma de decisiones. Es consecuencia del análisis llevado en esta tesis secuencialmente, suponiendo una ventaja comparativa de análisis y de adecuación de ofertas comerciales a los consumidores potenciales. El entendimiento de las más de 20 países de origen nacionalidades permite ajustar ofertas, adecuar el mercado de trabajo, permite identificar comportamientos de los consumidores en mercados primarios y secundarios de la industria turística basados en criterios de medición de la entropía estacional.

Las 20 series de incertidumbre analizadas se dividen en grandes áreas geográficas (Otros países europeos, Resto de la UE, Resto del mundo, África) y el resto por países tales como Alemania, Austria, Bélgica, Dinamarca, Estados Unidos, Finlandia, Francia, Grecia, Irlanda, Italia, Luxemburgo, Noruega, Países Bajos, Portugal, Reino Unido y Suecia. De

los resultados empíricos de este clustering no supervisado destacan: 1) los procesos de decisión bajo incertidumbre varían estacionalmente y la técnica descrita es una herramienta robusta, 2) el número de clusters es dinámico, variando el número de países y el número de clúster generados estacionalmente. Podemos concluir de este tercer capítulo que las consecuencias de este análisis permiten a los agentes económicos tomar decisiones estacionales basadas en la entropía como medida de incertidumbre y en la individualidad del país de origen. Con consecuencias directas en el mercado primario de alojamiento hotelero y mercados secundarios tales como alojamiento en apartamentos turísticos, alquiler de coches, restauración, negocios en el entorno del alojamiento hotelero. Además, presenta una implicación directa en el mercado laboral, puesto que el conocimiento del tipo de cliente estacional permitirá la contratación de mano de obra facultada y adecuada a las circunstancias de la demanda dinámica.

Consecuentemente el desarrollo secuencial de este tercer capítulo, presenta contribuciones a la Estadística teórica e implicaciones económicas en el mercado turístico español. Ayudando a herramientas de análisis Big Data tales como Next Best Activity en el proceso de marketing, este acercamiento permite la priorización y subsecuente selección de la mejor actividad para un potencial cliente en cada momento de su proceso de demanda. Conociendo el tipo de cliente en cada época estacional basados en los procesos de agrupamiento, se prioriza la campaña a realizar por parte de las empresas. Finalmente, dependiendo del sector implicado en cada momento, las actividades se realizarán en función de los resultados obtenidos por los métodos de clasificación para beneficio mutuo en la toma de decisiones.

Para concluir, el desarrollo de esta tesis en el desarrollo de los capítulos transcurre en dirección del “Big Data” a “Small Data” (Faraway & Augustin, 2018). Como se ha indicado se ha iniciado con un enfoque de Big Data para encontrar causalidad entre demanda hotelera y fuentes secundarias de datos procedentes de Google Trends. Con posterioridad se ha trabajado con el concepto de medición de incertidumbre para su cuantificación, clasificación y aplicación empírica al mercado de alojamiento turístico español según país de origen. El objetivo de este enfoque es la clasificación para el análisis de los comportamientos de consumidores en aras de intervenir en los procesos de demanda turística. A partir de esto, pueden surgir estudios relacionados con la propensión marginal al consumo, el cálculo del valor potencial de los clientes según patrones estacionales o incluso para regenerar modelos predictivos con patrones de comportamiento espacio temporal con los siguientes fines: estrategias de precios, mercado de trabajo, perfiles sociodemográficos o industrias relacionadas al alojamiento turístico.

Este análisis inicial pensado en Big Data y enfocado en el Small Data, se realiza basado en el principio de parsimonia. Los autores de la literatura previamente indican que es posible representar un mundo real con algún modelo simple (Box, 1979). En particular, el criterio de parsimonia utilizado en esta tesis es debido a varias razones: (i) la realidad simple ilustra, la complicación oscurece. En particular descubrimos relaciones causales procedentes de palabras clave del Big Data con una alta capacidad explicativa en el modelado. Con posteridad, desarrollamos metodología de incertidumbre basadas en entropía y clustering estacional para las dos principales opciones de alojamiento que las estadísticas oficiales españolas destacan (hoteles y apartamentos turísticos); (ii) La parsimonia es típicamente recompensada por una mayor precisión. En concreto la modelización teórica describe con alta precisión comportamientos estacionales y permiten su clasificación para empoderar el procedimiento de toma de decisiones en las organizaciones. (iii) En cualquier caso, la elaboración indiscriminada de modelos no es una opción práctica porque este camino es interminable. En nuestro caso los factores exógenos e incontrolables citados en la introducción pueden ser poco probables en un mercado consolidado como es el mercado turístico dónde el ambiente legal, climático y estructuras de la industria proporcionarán una pronta recuperación tras la crisis de la COVID-19.

En términos prácticos, dado que esta tesis se encuadra dentro de un análisis científico de las ciencias sociales y los factores exógenos e incontrolables pueden afectar al modelado. Los investigadores pueden acceder más fácilmente a recursos de bases de datos abiertos facilitando la aplicación del modelado a los campos de estudio. La modelización teórica bajo el criterio de parsimonia de este trabajo permite obtener resultados robustos con una contribución a la ciencia claramente. La robustez de estos resultados puede definirse por la insensibilidad a las desviaciones de los datos en el análisis, de este modo la ampliación de Teoría de la Información supera las limitaciones de los supuestos ideales habituales en los análisis econométricos. Estos supuestos ideales en el modelado científico pueden simplificarse ampliamente para entender la naturaleza de las realidades aplicadas en esta tesis. La necesidad de modelos simples robustos para realidades complejas parece estar íntimamente ligada (Lisciandra, 2016).

Por último, indicar que el observar grandes volúmenes de datos no garantiza resultados robustos de realidades sociales. En determinadas ocasiones la simplicidad y el criterio de parsimonia bate al análisis de grandes conjuntos de datos que requieren un desequilibrio costoso en contra de los grandes volúmenes de datos. En términos de inferencia estadística podemos encontrarnos con problemas básicos de multicolinealidad o inclusión de variables redundantes perjudicando al análisis estadístico y las propiedades de los

estimadores (Faraway & Augustin, 2018). Por esto último, queda justificado un análisis a nivel microscópico en los capítulos 2 y 3 de esta tesis.

Una vez analizados los principales resultados de la investigación, los resultados teóricos y empíricos suponen una contribución a la ciencia. Debido a esto, las ramas de investigación derivadas de esta tesis son numerosas y por sí mismas darían lugar a nuevas tesis doctorales, artículos o libros científicos. Teniendo en cuenta que esta tesis aúna conocimiento diverso del análisis de datos, negocios, teoría de la información y estadística espacial. Cualquier elemento esquematizado en la figura 2 del ciclo de vida del dato supone una nueva línea de investigación en su rama correspondiente. Para la recomendación de nuevas líneas de investigación, es conveniente diferenciar aspectos metodológicos de aspectos empíricos. De los primeros podemos destacar que las técnicas pueden ser aplicadas a numerosas líneas de investigación.

Del primer capítulo, destacar que las técnicas utilizadas pueden ir encaminadas en un análisis a corto o largo plazo con palabras clave de relevancia. Para ello se recomiendan usar algoritmos de decisión jerárquica basadas en keywords cuyas líneas de investigación ya han comenzado a elaborarse (Ruiz-Reina, 2020; Reina, 2021c). Además, dada la información abundante y cada vez más accesible se recomiendan añadir a las técnicas de predicción nuevos modelos basados en nowcasting y flujos de datos en tiempo real con machine learning (Carrière-Swallow & Labbé, 2013; Richardson et al., 2021). Para los criterios de decisión de modelos, la matriz U1 de Theil supone una contribución basada en criterios de selección adimensionales y permite cuantificar la ventaja del uso de algunos modelos, no obstante, se recomienda la profundización en estas técnicas con el objetivo de superar las limitaciones teóricas como predicciones exactas (Hyndman & Koehler, 2005, 2006; Ruiz-Reina, 2019d; Makridakis et al., 2020). Por otro lado, desde el punto de vista estadístico, se recomienda a los futuros investigadores profundizar en la idea del Small Data frente al Big Data (Faraway & Augustin, 2018). De este modo se puede profundizar en problemas clásicos de la Econometría como la colinealidad o la inclusión de variables irrelevantes (Wooldridge, 2013). El marco empírico en el mercado de alojamiento turístico desarrollado en esta tesis usando palabras clave de los motores de búsqueda de Google puede ser ampliado a otros campos de estudio tales como Finanzas (Hu et al., 2018), Economía y crecimiento (Niesert et al., 2020), sector asegurador y desempleo (Aaronson et al., 2021), Salud mental y emergencia (Knipe et al., 2021), Psicología (Simionescu & Raišienė, 2021), pandemias (Simionescu & Raišienė, 2021), infecciones (Nishimura et al., 2021) entre otros. El requisito de investigación en este caso supone que existan búsquedas previas que generen contenido

Big Data en los motores de búsqueda de Google. De este modo, en una sociedad digitalizada y conectada hace que las líneas de investigación sean inagotables.

En el segundo capítulo, la Teoría de la Información es la base del proceso de decision-making. En particular el núcleo de trabajo ha sido el concepto clásico de la entropía de Shannon de comunicación y percepción (Tishby et al., 2011), este concepto inicial puede ser aplicado a otros tipos de entropía dentro de la Estadística Mecánica (Jaynes, 1957; Golan & Maasoumi, 2008). En particular, en esta tesis se ha trabajado sobre una decisión binaria (dos posibles eventos mutuamente exclusivos) para la entropía de Shannon, este mismo análisis se puede realizar con la función de entropía de Bernoulli. Del mismo modo los métodos de estimación podrían generar propiedades deseables sobre consistencia de los parámetros estimados (Reina, 2021b). Además, se recomienda trabajar con toma de decisiones multivariantes con el fin de ampliar el espectro de este trabajo primario (Azami et al., 2019). El esquema de trabajo teórico se ha definido en el dominio del tiempo y de la frecuencia para el análisis de ciclos repetitivos, en esta línea se recomienda profundizar en el análisis de ondas relacionadas con el análisis armónico (Fernández-Macho, 2018). En cuanto a los ciclos, se han trabajado con ciclos de memoria corta, una posible línea de investigación sería el estudio de ciclos de memoria larga (Arteche, 2007). Estas líneas citadas son las principales investigaciones que pueden ampliar este apartado, no obstante, se recomienda profundizar en el concepto de Teoría de la Información para aportaciones posteriores al campo de la Econofísica aplicadas al campo de toma de decisiones (Delgado-Bonal & Marshak, 2019). En cuanto al apartado empírico, destacar que se ha trabajado con la decisión de alojamiento en hoteles o apartamentos turísticos, pero este puede ser ampliado a otras opciones tales como alojamientos rurales o campings, transportes (coche, tren o avión), destinos (playa, montaña o ciudad), tipos de viajes (ocio o culturales) y un amplio conjunto de posibilidades. Además, existen numerosas aplicaciones empíricas en campos como las finanzas (inversión inmobiliaria en vivienda o terrenos rústicos), salud (medición de la incertidumbre del riesgo), automóviles (selección de tipos de coches para comprar), o cualquier campo de aplicación que suponga tomar una decisión en un contexto de incertidumbre. Todo esto sirviendo como esquema de trabajo lo desarrollado en el capítulo 2 de esta tesis.

Relacionado con el tercer capítulo, el peso de la entropía juega un rol importante en la toma de decisiones y la clasificación de entropía es consecuencia de la medición de incertidumbre (Yue, 2017). Los criterios de agrupamiento no supervisados aplicados a las series temporales se encuentran menos desarrollados que los análisis tradicionales en datos de sección cruzada (Aghabozorgi et al., 2015). La aplicación de este tipo de análisis aplicado a la Estadística Espacial supone de por sí novedosas líneas de investigación en

múltiples contextos. La entropía, complejidad y la información espacial basadas en el contexto de la entropía de Shannon son campos insospechados en numerosos campos científicos. De este modo la medición de información creciente supone añadir complejidad al análisis en los sistemas de doble dimensión espacial (Batty et al., 2014). En este sentido el estudio espacio temporal con datos estacionales invita a ampliar campos de estudios recientes (Ruiz Reina, 2021). La mejora tecnológica de la extracción, transformación y recogida de datos hace que las aplicaciones de Estadística Espacial sean fuente de futuras investigaciones en cualquier campo (Gelfand, 2020): mortalidad infantil (Morales-Otero & Núñez-Antón, 2021), análisis de datos masivos (Banerjee, 2020), estudios de segmentación para turistas internacionales por motivos religiosos según niveles de gastos (Mercadé-Melé & Barreal Pernas, 2021) o movimientos de animales (Hooten et al., 2020). En definitiva, la Estadística Espacial se ha colocado entre las primeras líneas de investigación en los últimos 15 años, particularmente esta tesis supone un inicio hacia nuevas fronteras de datos geoespaciales en tiempo discreto, por lo que líneas adicionales podrían ir enfocadas al desarrollo de modelos de clasificación en tiempo continuo. Al igual que en los capítulos anteriores, las aplicaciones empíricas son muy amplias y diversas para proveer de conocimiento a investigadores.

Finalmente, esta tesis es una investigación para el mejor entendimiento de los procesos de toma de decisiones inusitados hasta el momento. Esta contribución ha superado, contribuido y abierto nuevas fronteras del conocimiento en las relaciones de causalidad entre sistemas de información y agrupamiento. Las referencias bibliográficas son relativamente modernas e implican conocimiento actualizado. Esta tesis en su conjunto puede animar a investigadores futuros a mejorar y entender los conocimientos generados en este trabajo. Las preguntas y respuestas de esta tesis son el punto de partida de rupturas en las fronteras actuales del conocimiento.

“El tiempo es oro, no lo desperdicies porque no volverá” — Me dijeron

“Ojalá el oro fuera tiempo” — Yo pensé