Big data and analytics in hospitality and tourism: a systematic literature review

Article

Supplemental Material

Mariani, M. M. and Baggio, R. (2022) Big data and analytics in hospitality and tourism: a systematic literature review. International Journal of Contemporary Hospitality Management, 34 (1). pp. 231-278. ISSN 0959-6119 doi: https://doi.org/10.1108/IJCHM-03-2021-0301 Available at https://centaur.reading.ac.uk/101148/

Table I The measures for all networks analyzed

		Scopus				WoS			
		Authors	Countries	Papers	Topics	Authors	Countries	Papers	Topics
No. nodes (order)		2243	75	883	30	4300	83	1419	30
No. links (size)		4879	215	19952	87	9085	369	66702	87
No. components		421	10	129	11	907	8	88	10
No. isolated nodes		68	8	125	10	91	7	84	9
Max component order		835	64	752	20	518	76	1329	21
Max component order (%)		37.2%	85.3%	85.2%	66.7%	12.0%	91.6%	93.7%	70.0%
Fragmentation index	(*)	0.860	0.272	0.275	0.563	0.984	0.163	0.123	0.517
Average degree	(*)	0.002	0.077	0.051	0.200	0.001	0.108	0.066	0.200
Average clustering coefficient	(*)	0.765	0.447	0.400	0.488	0.828	0.518	0.439	0.457
Average closeness	(*)	0.031	0.263	0.284	0.419	0.003	0.340	0.369	0.458
Average betweenness	(*)	0.000	0.014	0.001	0.010	0.000	0.012	0.001	0.011
Average eigenvalue centrality	(*)	0.002	0.080	0.019	0.134	0.001	0.080	0.016	0.137
Average local efficiency	(*)	0.774	0.526	0.557	0.537	0.836	0.624	0.632	0.523
Density	(*)	0.002	0.077	0.051	0.200	0.001	0.108	0.066	0.200
Diameter (LCC)	(*)	11	7	8	3	17	4	6	3
Average path length (LCC)	(*)	4.549	2.449	2.480	1.632	6.605	2.131	2.306	1.629
Assortativity	(*)	0.265	-0.249	0.194	-0.293	0.557	-0.325	0.160	-0.344
Global efficiency	(*)	0.035	0.346	0.328	0.312	0.004	0.437	0.419	0.338
Gini index degrees	(*)	0.416	0.556	0.600	0.540	0.366	0.554	0.530	0.523
Modularity (LCC)	(*)	0.781	0.300	0.326	0.089	0.866	0.221	0.368	0.140
No. communities (LCC)	(*)	23	6	6	3	22	6	4	3

LCC = largest connected compoment

Table II Most frequently used words and 2-grams

	Scopus					WoS			
	Words		2-grams			Words		2-grams	
Rank	Term	Occurr.	Term	Occurr.	Rank	Term	Occurr.	Term	Occurr.
1	data	44.4%	big_data	1.6%	1	data	47.9%	big_data	1.7%
2	tourism	16.8%	social_media	0.5%	2	tourism	20.8%	social_media	0.7%
3	travel	15.0%	data_analytics	0.3%	3	hotel	18.3%	online_reviews	0.5%
4	tourist	9.3%	online_reviews	0.2%	4	online	14.2%	data_analytics	0.3%
5	social	9.0%	machine_learning	0.2%	5	travel	13.5%	machine_learning	0.2%
6	time	8.6%	data_mining	0.2%	6	social	11.5%	user_generated	0.4%
7	analytics	8.3%	travel_behavior	0.2%	7	model	11.5%	customer_satisfaction	0.2%
8	information	8.1%	decision_making	0.2%	8	service	10.5%	sentiment_analysis	0.2%
9	urban	7.4%	real_time	0.1%	9	analytics	8.7%	methodology_approach	0.2%
10	online	6.6%	mobile_phone	0.1%	10	information	8.2%	design_methodology	0.2%
11	model	6.4%	travel_demand	0.1%	11	time	8.1%	decision_making	0.2%
12	management	6.0%	data_analysis	0.1%	12	city	7.6%	online_travel	0.1%
13	media	5.8%	media_analytics	0.1%	13	media	7.4%	data_analysis	0.1%
14	transportation	5.6%	hospitality_industry	0.1%	14	urban	6.9%	data_mining	0.1%
15	hotel	5.3%	smart_tourism	0.1%	15	destination	6.7%	mobile_phone	0.1%

16	destination	5.1%	case_study	0.1%	16	tourist	6.6%	text_mining	0.1%
17	network	5.1%	tourism_industry	0.1%	17	hospitality	6.5%	real_time	0.1%
18	traffic	5.1%	smart_city	0.1%	18	satisfaction	6.2%	smart_tourism	0.1%
19	spatial	5.0%	sentiment_analysis	0.1%	19	customer	6.0%	deep_learning	0.1%
20	systems	4.6%	time_series	0.1%	20	management	6.0%	destination_image	0.1%

Table III The top ten topics along with their ten most relevant terms

Scopus

Topic	**Terms**								
Topic 1	data	travel	predict	model	time	trip	transport	city	system	learn
Topic 2	travel	data	use	transport	big	tourism	model	urban	servic	system
Topic 3	data	online	custom	review	travel	analysis	hotel	analytic	use	predict
Topic 4	data	media	social	analysi	onlin	hotel	model	custom	big	review
Topic 5	data	model	travel	analyt	big	hotel	media	social	transport	system
Topic 6	data	mobil	pattern	urban	tourism	spatial	travel	activ	analysi	social
Topic 7	data	big	analysi	tourism	custom	inform	use	manag	travel	mobil
Topic 8	data	model	review	social	hotel	onlin	analyt	big	use	media
Topic 9	data	model	urban	tourism	analysi	travel	use	tourist	destin	big
Topic 10	data	tourism	tourist	method	big	model	analysi	inform	travel	forecast

WoS

Topic	**Terms**									
Topic 1	review	onlin	data	model	analysi	use	manag	hotel	satisfact	custom
Topic 2	data	tourism	travel	use	big	tourist	inform	analysi	model	differ
Topic 3	data	mobil	health	time	travel	urban	servic	human	use	big
Topic 4	data	model	research	review	onlin	big	manag	tourism	travel	social
Topic 5	research	hospit	analysi	tourism	review	data	model	industri	use	big
Topic 6	hotel	data	manag	big	experi	onlin	review	analysi	tourism	citi
Topic 7	data	comput	tourism	inform	big	use	system	analysi	develop	fog
Topic 8	data	tourism	model	analysi	travel	big	network	use	servic	research
Topic 9	review	data	onlin	hotel	model	differ	custom	tourist	base	network
Topic 10	onlin	review	hotel	custom	travel	satisfact	servic	differ	attribut	analysi

Table IV Similarity of words and topics for the two time periods considered (up to 2017 included and after 2017)

	Words	**2-grams**	**Topics**
Scopus	0.77	0.61	0.92
WoS	0.76	0.60	0.86

Table V The Cosine similarity between all the networks' topological features

Network	**Similarity**
Authors	0.977
Countries	0.964
Papers	0.995
Topics	0.996

Table VI The top ten papers by citation count

Scopus

Rank	**Authors**	**Title**	**Journal**	**Year**
1	Gretzel, U. *et al*.	Smart tourism: foundations and developments	*Electronic Markets*	2015
2	Batty, M.	Big data, smart cities and city planning	*Dialogues in Human Geography*	2013
3	Sun, Y. *et al*.	Internet of Things and big data analytics for smart and connected communities	IEEE *Access*	2016
4	Xiang, Z. *et al*.	What can big data and text analytics tell us about hotel guest experience and satisfaction?	*International Journal of Hospitality Management*	2015
5	Wood, S.A. *et al*.	Using social media to quantify nature-based tourism and recreation	*Scientific Reports*	2013
6	Fan, W. and Gordon, M.D.	The power of social media analytics	*Communications of the ACM*	2014
7	Xiang, Z. *et al.*	A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism	*Tourism Management*	2017
8	Guo, Y. *et al*.	Mining meaning from online ratings and reviews: tourist satisfaction analysis using Latent Dirichlet Allocation	*Tourism Management*	2017
9	Zhong, C. *et al*.	Detecting the dynamics of urban structure through spatial network analysis	*International Journal of Geographical Information Science*	2014
10	Toole, J.L. *et al*.	The path most traveled: travel demand estimation using big data resources	*Transportation Research Part C: Emerging Technologies*	2015

WoS

Rank	**Authors**	**Title**	**Journal**	**Year**
1	Gretzel, U. *et al*.	Smart tourism: foundations and developments	*Electronic Markets*	2015
2	Xiang, Z. *et al*.	What can big data and text analytics tell us about hotel guest experience and satisfaction?	*International Journal of Hospitality Management*	2015

3	Sun, Y.C. *et al*.	Internet of Things and big data analytics for smart and connected communities	IEEE *Access*	2016
4	Wood, S.A. *et al*.	Using social media to quantify nature-based tourism and recreation	*Scientific Reports*	2013
5	Xiang, Z. *et al*.	A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism	*Tourism Management*	2017
6	Guo, Y. *et al*.	Mining meaning from online ratings and reviews: tourist satisfaction analysis using Latent Dirichlet Allocation	*Tourism Management*	2017
7	Fan, W. and Gordon, M.D.	The power of social media analytics	*Communications of the ACM*	2014
8	Zhong, C. *et al*.	Detecting the dynamics of urban structure through spatial network analysis	*International Journal of Geographical Information Science*	2014
9	Toole, J.L. *et al*.	The path most traveled: travel demand estimation using big data resources	*Transportation Research Part C: Emerging Technologies*	2015
10	Chen, C. *et al*.	The promises of big data and small data for travel behavior (aka human mobility) analysis	*Transportation Research Part C: Emerging Technologies*	2016

Table VII Most represented journals for both databases

Scopus		WoS	
Rank	**Journal**	**Rank**	**Journal**
1	*Tourism Management*	1	*Sustainability*
2	*Sustainability*	2	*Tourism Management*
3	*Transportation Research Part C: Emerging Technologies*	3	*International Journal of Hospitality Management*
4	*International Journal of Hospitality Management*	4	*International Journal of Contemporary Hospitality Management*
5	*ISPRS International Journal of Geo-Information*	5	*ISPRS International Journal of Geo-Information*
6	IEEE *Access*	6	*Transportation Research Part C: Emerging Technologies*
7	*International Journal of Contemporary Hospitality Management*	7	IEEE *Access*
8	*IEEE Transactions on Intelligent Transportation Systems*	8	*Journal of Travel Research*
9	*Current Issues in Tourism*	9	*Journal of Hospitality and Tourism Technology*
10	*Transportation Research Record*	10	*Tourism Review*

Table VIII Big data and analytics works in hospitality and tourism (Top 40 most cited work at the time of retrieval in Scopus and WoS, without duplication; in "type of data and size" an asterisk indicates large quantities of data > 100,000 records)

Article (author and title)	Macro-topical area	Research topic	Type of paper (conceptual, review, empirical)	Sources of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Alaei *et al*. (2019), Sentiment Analysis in Tourism: capitalizing on big data	Methodological contributions shedding light on a specific technique or family of techniques	Sentiment analysis approaches applied in tourism	Review/methodology	N/A	N/A	N/A	N/A	N/A
Batista e Silva *et al*. (2018), Analyzing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Generates a dataset describing tourist density at high spatial resolution, with monthly breakdown for the European Union	Empirical	EUROSTAT, Booking.com, Tripadvisor	716,000 establishments in Booking.com and Tripadvisor combined	Secondary data from EUROSTAT, and collection from Booking.com and Tripadvisor	Statistical software and geographical information systems (GIS)	Tourist density grids, density maps, tables
Batty (2013), Big data, smart cities and city planning	Knowledge and value creation	Introduces urban big data, explains its importance for cities and short-term thinking and calls for new theory on smart travel card data in Greater London	Conceptual	N/A	N/A	N/A	N/A	N/A
Brandt *et al*. (2017), Social media analytics and value	Knowledge and value creation	Uses Twitter data for San Francisco and kernel density						

creation in urban smart tourism ecosystems		estimation and Latent Dirichlet Allocation to captures spatial patterns within the city	Empirical	600,000 geo-tagged Twitter posts	Structured and unstructured (*)	Data purchased from Twitter	Kernel density estimation and Latent Dirichlet Allocation	Heat maps, figures, tables
Buhalis and Foerste (2015), SoCoMo marketing for travel and tourism: empowering co-creation of value	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	Proposes social context mobile (SoCoMo) marketing as a new framework that enables marketers to increase value for all stakeholders at the destination	Conceptual	N/A	N/A	N/A	N/A	N/A
Buhalis and Leung (2018), Smart hospitality—interconnectivity and interoperability towards an ecosystem	Knowledge and value creation (in smart cities context)	Conceptualizes the smart and agile hospitality enterprises of the future and proposes a smart hospitality ecosystem. The model enables fully integrated applications, using big data to enhance hospitality decision-making	Conceptual	N/A	N/A	N/A	N/A	N/A
Buhalis and Sinarta (2019), Real-time co-creation and nowness service: lessons from tourism and hospitality	Knowledge and value creation (in smart cities context)	Conceptualizes the integration of real-time consumer intelligence, dynamic big data mining, artificial intelligence, and contextualization to illustrate service co-creation	Conceptual	N/A	N/A	N/A	N/A	N/A

Cai et al.(2014), Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet	Mapping, identification and representation of tourists, tourists' behaviors, tourist attractions, destinations and trips	Based on trajectory data of more than 11,800 taxis, evaluates how travel patterns mined from big data can inform public charging infrastructure development	Empirical	Trajectory data of 11,880 taxis	Unstructured (*)	Not specified	Probability density distributions	Charts, tables
Chen et al. (2016), The promises of big data and small data for travel behavior (aka human mobility) analysis	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Literature review trying to reconcile two separate fields, both involving understanding and modeling of how individuals move in time and space: "travel behavior analysis" and "human mobility analysis"	Conceptual	N/A	N/A	N/A	N/A	N/A
Cheng and Jin (2019), What do Airbnb users care about? An analysis of online review comments	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	By analyzing a "big data" set of Airbnb online reviews through text mining and sentiment analysis, the study finds that Airbnb users tend to evaluate their experience based on a frame of reference derived from past hotel stays. Key attributes identified in the data include	Empirical	181,263 Airbnb online reviews	Structured and unstructured (*)	Inside Airbnb website	Text mining through unsupervised learning, Leximancer	Charts, figures, tables

		“location”, “amenities” and “host”						
Chua *et al*. (2016), Mapping Cilento: using geotagged social media data to characterize tourist flows in southern Italy	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Analyzes geotagged social media data from Twitter to characterize spatial, temporal and demographic features of tourist flows in Cilento	Empirical	72,031geotagged Tweets	Structured and unstructured	Twitter's stream API, Twitter's REST API	Trajectory mining using graphs	Charts, figures, FlowSampler for flow visualization
Del Vecchio *et al*. (2018), Open innovation and social big data for sustainability: evidence from the tourism industry	Knowledge and value creation	Explores a set of regional tourist experiences related to a Southern European region and destination, to derive patterns and opportunities of value creation generated by big data in tourism	Empirical	7 social media accounts (including Instagram accounts, travel blogs, etc.). Posts, tweets/retweets up to 6,028	Structured and unstructured	Tracking accounts	Keyhole for clustering and Buzztrack for sentiment analysis	Charts, figures, tables
Fan and Gordon (2014), The power of social media analytics	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Discusses the benefits of social media analytics, the underlying stages of the analytics process, the most common social media analytic techniques, and the ways analytics creates business value	Conceptual	N/A	N/A	N/A	N/A	N/A

Fuchs *et al*. (2014), Big data analytics for knowledge generation in tourism destinations – a case from Sweden	Knowledge and value creation	Presents a knowledge infrastructure implemented at the Swedish mountain tourism destination, Åre and examples of use by tourism managers	Empirical	Web search, booking and feedback data (e.g., survey-based, user-generated content)	Structured and unstructured	Data warehouse (DW) including facts and dimensions tables	Online analytical processing (OLAP); support vector machines (SVM), Naïve Bayes (NB) and k-nearest neighbor (KNN)	Html-based web application
Gao *et al*. (2013), Discovering Spatial Interaction Communities from Mobile Phone Data	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	The study adopts an agglomerative clustering algorithm based on a Newman-Girvan modularity metric. By proposing an alternative modularity function incorporating a gravity model, this work helps discover the clustering structures of spatial-interaction communities using mobile phone datasets	Empirical	Phone records of 1,000,000 mobile subscribers in Harbin, China. A total of 74,000,000 phone call records	Structured (*) 74,000,000 phone call detail records	Not disclosed	Spatial and temporal analysis using networks	Tables, maps, figures matched with Google Earth
García-Palomares *et al*. (2015), Identification of tourist hot spots based on social networks: a comparative analysis of European metropolises using photo-sharing services and GIS	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Use of photo-sharing services for identifying and analyzing the main tourist attractions in eight major European cities	Empirical	Panoramio photos	Unstructured (*)	Panoramio website API + ArcGIS	Density graphs, spatial autocorrelation	Standard tables and Anselin Local Moran's I graph

Gretzel *et al*. (2015), Smart tourism: foundations and developments	Knowledge and value creation (in a smart cities context)	Defines smart tourism, sheds light on current smart tourism trends, and lays out its technological and business foundations	Conceptual	N/A	N/A	N/A	N/A	N/A
Guo *et al*. (2017), Mining meaning from online ratings and reviews: tourist satisfaction analysis using Latent Dirichlet Allocation	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	Uses Latent Dirichlet analysis (LDA) to identify key dimensions of customer service in online reviews. LDA uncovers 19 controllable dimensions that are key for hotels to manage their interactions with visitors	Empirical	Tripadvisor online reviews	Structured and unstructured (*)	Web crawler	LDA, correspondence analysis, ANOVA, regression analysis, perceptual mapping	Charts, tables, maps
Hasan and Ukkusuri (2014), Urban activity pattern classification using topic models from online geo-location data	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Develops a model to analyze large-scale geo-location data from social media to infer individual activity patterns	Empirical	Geo-located Twitter check-ins	Structured and unstructured (*)	Web crawler	Topic modeling (LDA)	Charts, tables
Hasan *et al*. (2013), Spatiotemporal patterns of urban human mobility	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Use of smart subway fare card transactions data to model urban mobility patterns. The popularity of places in the city is deployed as an interaction	Empirical	Smart card transactions over a three-month period of public transport users of London	Structured and unstructured	Data was collected by Transport for London (TfL) for operational purposes	Visit probability function	Charts, tables

		parameter between different individuals						
Khalilzadeh and Tasci (2018), Large sample size, significance level, and the effect size: solutions to perils of using big data for academic research	Methodological contributions shedding light on a specific technique or family of techniques (measurement problem)	Informs tourism and hospitality academia of the effect size, measures for the most commonly used statistical tests when dealing with big data	Methodological	Individuals	Structured	Survey	Tests	Tables
Kim *et al*. (2019), Quantifying nature-based tourism in protected areas in developing countries by using social big data	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Illustrates spatial patterns of visitation using 10 years of Flickr geo-tagged photographs	Empirical	Flickr geotagged photos	Structured and unstructured	OpenStreet Map, Python model, NatCap, InVEST.rec reation	Geographically weighted regression	Charts, tables
Kim *et al*. (2017), What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	Applies sentiment analysis to VirtualTourist online reviews of Paris	Empirical	19,835 online reviews from VirtualTourist	Structured and unstructured	Web crawling program developed in Python	Stanford sentiment analysis tool based on JAVA1.7.0_65, . Co-occurrence analysis	Tables, charts

Kirilenko *et al*. (2018), Automated sentiment analysis in tourism: comparison of approaches	Methodological contributions shedding light on a specific technique or family of techniques	Illustrates the suitability of different types of automated classifiers for applications typical in tourism, hospitality, and marketing studies by comparing their performance to that of human assessors/raters	Methodological and empirical	209 surveys of travelers, 332 Tripadvisor reviews, 200 English-language Twitter messages	Structured and unstructured	Online panel survey, Tripadvisor, Twitter	Sentiment analysis using four types of software (SentiStrength, Deeply Moving, two programs using RapidMiner) vs. human raters. Accuracy, precision and recall for software and Cohen's kappa, Kendall's τ, ratio of opposite classification for human raters	Tables, figures
Li *et al*. (2018), Big data in tourism research: a literature review	Knowledge and value creation	Reviews literature on different types of big data in tourism research, distinguishing UGC data, device data and transaction data. Research focuses, data characteristics, analytic techniques, major challenges and further directions are identified	Literature review	N/A	N/A	N/A	N/A	N/A

Li *et al*. (2017), Forecasting tourism demand with composite search index	Demand evaluation and forecast/prediction	Proposes and tests a framework and procedure to create a composite search index adopted in a generalized dynamic factor model (GDFM). This is used to predict tourist volumes to Beijing	Empirical	Baidu search engine trends, data series and tourist volumes	Structured (*)	Search using search engine	Econometric model	Charts, tables
Liu *et al*. (2017), Big data for big insights: investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	Examines the determinants of hotel customer satisfaction by discriminating among customers by language group. Study of hotel customer reviews written by guests speaking 10 different languages	Empirical	412,784 Tripadvisor online reviews of hotels located in five Chinese cities	Structured (*)	Crawler developed in PHP	Language-detection was through MySQL and the textcat package, regression analysis	Charts, tables
Ma *et al*. (2018), Effects of user-provided photos on hotel review helpfulness: an analytical approach with deep leaning	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	Compares deep learning models with other machine learning techniques to examine the effect of user-provided photos on review helpfulness across two social media	Empirical	24,960 Tripadvisor online reviews and 3,064 Yelp online reviews	Structured and unstructured	Web crawler in Python based on APIs	Deep learning model: neural network for sequence encoding, residual network for image representation, feature fusion	Tables and Figures
Mariani *et al*. (2018), Business intelligence and big data in hospitality and tourism: a systematic literature review	Knowledge and value creation	Performs a systematic literature review of business intelligence and big data	Literature Review	N/A	N/A	N/A	N/A	N/A

Mariani and Borghi (2018), Effects of the Booking.com rating system: bringing hotel class into the picture	Methodological contributions shedding light on a specific technique or family of techniques (data quality)	Investigates the effects of the Booking.com rating system on the distribution of hotel ratings for the overall population of hotels located in London over two years	Empirical	1,228,089 Booking online reviews	Structured and unstructured (*)	Web crawler developed in Python	Nonparametric kernel density estimators, parametric and nonparametric tests	Tables and Figures
Mariani *et al*. (2016), Facebook as a destination marketing tool: evidence from Italian regional destination management organizations	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	Explores how Italian regional destination management organizations (DMOs) strategically employ Facebook to promote and market their destinations, and improves on the current metrics for capturing user engagement	Empirical	Overall number of Facebook posts posted on the official Italian regional DMOs' Facebook pages	Structured (*)	Data extractor based on Facebook APIs	Data parser and analyzer, calculating per post statistics	Tables created through data analyzer module. Graphs created through the data visualizer module
Marine-Roig and Anton Clavé (2015), Tourism analytics with massive user-generated content: A case study of Barcelona	Knowledge and value creation	Studying the online image of Barcelona as transmitted via social media through the analysis of more than 100,000 relevant travel blogs and online travel reviews (OTRs) written in English	Empirical	Heterogeneous, including the travel blogs, webpages, travelogues and travel reviews about Barcelona (250,000 pages)	Unstructured (*)	Data was extracted through Offline Explorer Enterprise (OEE)	Pre-processing: web content mining, language detection, user's hometown, cleaning, debugging. Processing: parser settings and categorizations through site content analyzer	Tables created through word count

Miah *et al*. (2017), a big data analytics method for tourist behaviour analysis	Knowledge and value creation	Using a design science research approach, designs and evaluates a big data analytics method to support strategic decision-making in tourism destination management	Empirical and methodological	Geotagged Flickr photos uploaded by tourists	Unstructured (*)	Flickr API	Textual metadata processing (GATE); geographical data clustering (P-DBSCAN); representative photo identification (visual content representation and Kernel density estimation); time series modeling (MAE)	Maps, charts and tables
Mocanu *et al*. (2013), The twitter of babel: mapping world languages through microblogging platforms	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Survey on worldwide linguistic indicators and trends	Empirical	Large-scale dataset of geotagged tweets	Structured and unstructured (*)	Twitter API	Language detection (Google Chromium Compact Language Detector) – geographical analyses	Maps, charts and tables
Paldino *et al*. (2015), Urban magnetism through the lens of geo-tagged photography	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Tastes of individuals, and what attracts them to live in a particular city or spend their vacation there	Empirical	Geo-tagged photos	Structured: metadata from photos (*)	Flickr API	Identification of resident, tourist and unknown, statistical analysis, network analysis (origin/destination)	Maps, charts, tables

Pan and Yang (2017), Forecasting destination weekly hotel occupancy with big data	Demand evaluation and forecast/prediction	Deploys time-series models incorporating tourism big data sources, including search engine queries, website traffic, and weekly weather information, to construct an accurate forecasting model of weekly hotel occupancy	Empirical	Google queries (Google Correlate), weekly session data on website traffic for the Charleston Area Convention and Visitors Bureau, weekly weather information from National Weather Service for North America and Europe. In addition, STR hotel occupancy data	Structured data	Google Correlate, Charleston Area Convention and Visitors Bureau, STR	Time series models (ARMAX, MSDR)	Charts, tables
Park *et al*. (2016), Using Twitter data for cruise tourism marketing and research	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Illustrates social media analytics using Twitter data referring to cruise travel	Empirical	50,414 cruise related Tweets	Structured and unstructured	ScraperWiki, Twitter API	Frequency analysis, network mapping	Tables, figures
Plaza (2011), Google Analytics for measuring website performance	Methodological contributions shedding light on a specific technique or family of techniques (platform issues and features)	Measures the effectiveness of website visit behavior and length of sessions, depending on the traffic source	Empirical	7,561 entries for 1,092 days drawn from Google Analytics	Structured	Excel	Regressions	Figures, tables

Puiu *et al*. (2016), CityPulse: large scale data analytics framework for smart cities	Methodological contributions shedding light on a specific technique or family of techniques (platform issues and features)	Presents a framework named CityPulse, describes its components, and demonstrates how they interact to support easy development of custom-made applications for citizens	Conceptual	Development of an architecture	N/A	N/A	N/A	N/A
Raun *et al*. (2016), Measuring tourism destinations using mobile tracking data	Demand evaluation and forecast/prediction	Measure space-time tracking data to analyze, monitor and compare destinations based on data describing actual visits	Empirical	Anonymized roaming data of foreign mobile phones (406,590 visits by 215,643 different phone IDs)	Structured (*)	From telcom operator	Statistical analyses, ArcGIS for spatial analyses, binary logistic regression	Charts, tables, maps
Rossetti *et al*. (2016), Analyzing user reviews in tourism with topic models	Methodological contributions shedding light on a specific technique or family of techniques	A description of the topic model method with application focus on the tourism domain	Empirical	Yelp Dataset Challenge; Tripadvisor dataset	Structured and unstructured	Yelp's existing dataset; Tripadvisor, automatically collected by crawler	K-nearest neighbor user based (KNN-UB), k-nearest neighbor item based (KNN-IB), probabilistic matrix factorization (PMF)	Illustrative examples for selected topics related to multi-criteria dimensions

Salas-Olmedo *et al*. (2018), Tourists' digital footprint in cities: comparing big data sources	Methodological contributions shedding light on a specific technique or family of techniques (platform issues and features)	Analyzes the digital footprint of urban tourists through big data from Panoramio (sightseeing), Foursquare (consumption) and Twitter (being connected – accommodation)	Empirical	307,062 geolocated Panoramio photographs, 234,159 Tweets, of which 20,076 Foursquare check-ins	Structured and unstructured (*)	Panoramio API	Statistical analyses, MongoDB, ArcGIS for spatial analyses, OLS regression, spatial autocorrelation	Charts, tables, maps
Schuckert *et al*. (2016), Insights into suspicious online ratings: direct evidence from Tripadvisor	Methodological contributions shedding light on a specific technique or family of techniques (data quality)	Examines gap between overall rating and individual ratings, as well as the proportion of suspicious ratings	Empirical	41,572 Tripadvisor reviews	Structured and unstructured	Crawler was developed to retrieve Tripadvisor online review data	Java-based program to parse HTML and XML web pages, regression analyses	Charts, tables
Sun *et al*. (2019), Forecasting tourist arrivals with machine learning and internet search index	Demand evaluation and forecast/prediction	Verifies the Granger causality and co-integration relationship between internet search index and tourist arrivals of Beijing	Empirical	Tourist arrivals, search query data from Baidu Index and Google Trends	Structured	Tourist arrivals from Wind Database	Kernel extreme learning machine, Granger causality test	Tables and charts

Sun *et al*. (2016), Internet of Things and big data analytics for smart and connected communities	Knowledge and value creation (in smart cities context)	Integration of Internet of Things (IoT) and big data analytics for smart connected communities	Conceptual and case study	Design of an IoT system personal sensors, open data and participatory sensing to enhance the services in the area of tourism and cultural heritage with a context-aware recommendation system	N/A	N/A	N/A	N/A
Tenkanen et al. (2017), Instagram, Flickr or Twitter: assessing the usability of social media data for visitor monitoring in protected areas	Methodological contributions shedding light on a specific technique or family of techniques (platform issues and features)	Estimation of visitation in protected areas through social media analytics	Empirical	Tourists' posts on Instagram, Flickr or Twitter	Structured and unstructured	APIs	Spearman correlation, Pearson correlation	Tables and charts
Toole *et al*. (2015), The path most traveled: travel demand estimation using big data resources	Demand evaluation and forecast/prediction	Develops a software system to estimate multiple aspects of travel demand using call detail records (CDRs) from mobile phones in conjunction with open- and crowdsourced geospatial data, census records, and surveys	Empirical	Call detail records, geospatial data, census records, surveys	Structured and unstructured (*)	Data from companies and cities, GIS, OpenStreet Map, surveys	Origin – destination matrices, network analysis, other analytics	Charts, tables, maps

Tu *et al*. (2016), Optimizing the locations of electric taxi charging stations: a spatial–temporal demand coverage approach	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Develops a spatial–temporal demand coverage approach for optimizing the placement of electric taxis charging stations in a space–time context. The study also evaluates the carbon emission generated by used electric taxis	Empirical	Taxi GPS data, transportation network, charging stations	Structured and unstructured	GPD, NAVInfo China	Map matching, segmentation, spatial–temporal demand coverage location model	Charts, tables, maps
Tussyadiah and Zach (2017), Identifying salient attributes of peer-to-peer accommodation experience	Methodological contributions shedding light on a specific technique or family of techniques	Explores themes from Airbnb online reviews to explain major service attributes sought by guests. Conducts a lexical analysis	Empirical	41,560 Airbnb reviews	Structured and unstructured	Inside Airbnb website	Lexical analysis, tokenization, word co-occurrence, regressions	Charts, tables, maps

Vilajosana *et al*. (2013), Bootstrapping smart cities through a self-sustainable model based on big data flows	Methodological contributions shedding light on a specific technique or family of techniques	Elaborates a procedure and a framework to make smart cities happen based on big data exploitation through the API stores concept	Conceptual	N/A	N/A	N/A	N/A	N/A
Wood *et al*. (2013), Using social media to quantify nature-based tourism and recreation	Methodological contributions shedding light on a specific technique or family of techniques measurement	Online posted photos are used to estimate visitation rates and travelers' origins, compared to empirical data showing that crowd-sourced information can serve as reliable proxy for empirical visitation rates	Empirical	Empirical datasets that quantified visitation to 836 sites in 31 countries around the world, plus Flickr metadata	Structured (*)	Dataset + Flickr API	Statistical and spatial analyses	Charts, tables, maps
Xiang *et al*. (2017), A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism	Methodological contributions shedding light on a specific technique or family of techniques (platform issues and features)	Three major online review platforms (Tripadvisor, Expedia and Yelp) are examined comparatively through text analytics in relation to hotel population in Manhattan (NYC). Discrepancies in the representation of the hotel industry on the platforms are identified	Empirical	Online reviews from three different platforms (Tripadvisor, Expedia, Yelp)	Structured and unstructured (*)	Web crawlers written in the Python and Java programmin g languages	Topic modeling (LDA), sentiment analysis, ML techniques, regression analyses	Charts, tables, maps

Xiang *et al*. (2015), What can big data and text analytics tell us about hotel guest experience and satisfaction?	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	Text analytics are produced for a large quantity of Expedia consumer reviews to deconstruct hotel guest experience and examine the association with satisfaction ratings. The association between guest experience and satisfaction is empirically proved	Empirical	Expedia online reviews	Structured and unstructured	Automated web crawler	Text analytics process involving data pre-processing, domain identification, statistical association analysis	Charts, tables
Xu *et al*. (2017), Business intelligence in online customer textual reviews: understanding consumer perceptions and influential factors	Tourists' (residents and service providers) perceptions, experiences, emotions, satisfaction and engagement with hospitality and tourism services	Examines customer satisfaction and dissatisfaction toward attributes of hotel products and services based on online customer textual reviews	Empirical	3,596 Tripadvisor online reviews	Structured and unstructured	Manual collection	Latent semantic analysis, regression analysis	Charts, tables
Yang *et al*. (2014), Predicting hotel demand using destination marketing organization's web traffic data	Demand evaluation and forecast/prediction	Deploys web traffic volume data of a destination marketing organization (DMO) to predict hotel demand for the destination	Empirical	Google Analytics account of the CACVB website, and (STR) hotel demand and occupancy data for the Charleston area	Structured	Google analytics plus standard data	Statistical and time series forecasts	Charts, tables
Zhao *et al.* (2019), Predicting overall customer satisfaction: big data evidence from hotel online textual reviews	Demand evaluation and forecast/prediction	Predicts overall customer satisfaction using the technical attributes of online textual reviews and customers'	Empirical	127,629 Tripadvisor online reviews	Structured and unstructured (*)	Customized scripts in Python	Regression analyses	Charts, tables

		involvement in the review community						
Zhong *et al*. (2014), Detecting the dynamics of urban structure through spatial network analysis	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Develops quantitatively a spatial structure of urban movements by constructing a weighted directed graph from smart card travel records in Singapore. Graph properties are used to obtain a view of travel demand, urban centers, neighborhoods	Empirical	Smart card data	Structured and unstructured (*)	Data provided by the Singapore Land Transport Authority	Network analysis package (R), community detection analysis (MapEquation), spatial analysis tool (ArcGIS)	Charts, tables, maps
Zhou *et al*. (2015), Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform	Mapping, identification and representation of tourists, tourist behaviors, tourist attractions, destinations and trips	Faces the challenges inherent in spatial analytics and automates the detection of places of interest in multiple cities based on spatial and temporal features of Flickr images from 2007. RHadoop platform is used	Empirical	Flickr Creative Dataset from Yahoo Lab, containing 99.3 million images, 49 million of which were geotagged	Structured and unstructured (*)	Hadoop cluster, Geospatial Data Abstraction Library (GDAL)	Tag calculation, tag classification	Charts, tables, maps