



Creación de un índice de movilidad para rankings de hoteles

XIMÉNEZ DE SANDOVAL, JOSÉ LUIS

Departamento de Economía Aplicada (Estructura Económica)
Facultad de Turismo. Universidad de Málaga
Correo electrónico: joseluis.xs@uma.es

FERNÁNDEZ-MORALES, ANTONIO

Instituto Universitario de Investigación de Inteligencia e Innovación Turística (i3t)
Universidad de Málaga
Correo electrónico: afdez@uma.es

GUEVARA-PLAZA, ANTONIO

Instituto Universitario de Investigación de Inteligencia e Innovación Turística (i3t)
Universidad de Málaga
Correo electrónico: guevara@uma.es

RESUMEN

Los rankings de hoteles ejercen gran influencia sobre los turistas a la hora de elegir un hotel. Los hoteleros también se esfuerzan por aparecer en lo más alto de la clasificación. Sin embargo, la evolución de estos rankings a lo largo del tiempo ha sido poco estudiada, por lo que es necesario profundizar en su análisis debido a las importantes implicaciones que tiene para los empresarios. ¿Permanecen los hoteles en las mismas posiciones o por el contrario experimentan grandes cambios en la clasificación? Este trabajo trata de aportar nuevas herramientas para el análisis de estos rankings con la creación de un índice que mide la movilidad de un hotel en el ranking de su ciudad. Se aplicará el modelo a una muestra de 727 hoteles con datos extraídos de Tripadvisor en cinco ciudades españolas (Barcelona, Madrid, Sevilla, Valencia y Zaragoza).

Palabras clave: TripAdvisor, ranking, hoteles, movilidad, minería de datos.

Clasificación JEL: L83; Z31; O33.

MSC2010: 91B82; 62P20; 62G30.

Designing a hotel ranking mobility index

ABSTRACT

Hotel rankings exert great influence on tourists when choosing a hotel. The hoteliers also strive to appear at the top of the ranking. However, the evolution over time of these rankings has been little studied, so it is necessary due to the important implications for entrepreneurs to deepen the analysis to answer questions like this: Do the hotels stay in the same (or quite similar) positions or do they experience big changes in the classification? This work tries to provide new tools for the analysis of these rankings with the creation of an index that measures the mobility of a hotel in the ranking. The model will be applied to a sample of 727 hotels with data extracted from Tripadvisor in five Spanish cities (Barcelona, Madrid, Seville, Valencia and Zaragoza).

Keywords: TripAdvisor, ranking, hotels, mobility, data mining.

JEL classification: L83; Z31; O33.

MSC2010: 62P20; 62G30.



1. Introducción.

Por varias razones los seres humanos nos sentimos atraídos por las listas. Nos ahorran esfuerzos para clasificar y elegir. Nos ayudan a ordenar, a tomar decisiones. Son básicamente útiles. Si bien las listas son casi tan antiguas como el ser humano, y así lo indica Umberto Eco (2011) cuando nos recuerda que el primer ejemplo de lista aparece en la *Ilíada* de Homero; este autor italiano también reconoce que la gran madre de todas las listas, infinita por definición porque está en continuo desarrollo es la World Wide Web.

Es en esta inmensa red donde surge en el año 2000 una empresa que ha revolucionado el sector turístico precisamente por sus listas: Tripadvisor. Ordena y clasifica hoteles, restaurantes y destinos de los cinco continentes a partir de los comentarios y valoraciones de los usuarios para facilitarnos la toma de decisiones.

La gran influencia que los rankings de hoteles ejercen sobre los clientes cuando estos van a elegir el establecimiento donde alojarse es reconocida por expertos en la materia, académicos y profesionales (Ghose, Ipeirotis, & Li, 2012). También los empresarios se ven afectados por estas clasificaciones en la medida en que una buena o mala posición puede influir no sólo en su reputación sino en el número de reservas y, por tanto, en la cifra de negocio (Callarisa-Fiol, Sánchez-García, Moliner, & Forgas-Coll, 2012).

Pese a la gran cantidad de trabajos académicos que en los últimos años han analizado esta aplicación (Miguéns, Baggio, & Costa, 2008; O'Connor, 2008; Melián, Bulchand, & González, 2010; Smyth, Wu, & Greene, 2010; O'Connor, 2010; González, Gidumal, & López-Valcárcel, 2010; Jeacle & Carter, 2011; Lee, Law, & Murphy, 2011; Vásquez, 2011; Tuominen, 2011; Callarisa-Fiol et al., 2012; Ayeh, Au, & Law, 2013; Filieri, Alguezaui, & McLeay, 2015; Kladou & Mavragani, 2015; Molinillo, Ximénez de Sandoval, Fernández-Morales, & Coca-Stefaniak, 2016) la movilidad de estos rankings, esto es, cómo evolucionan las posiciones de los hoteles a lo largo del tiempo, ha sido muy poco estudiada en la literatura científica.

Las revisiones bibliográficas más recientes (Chen & Law, 2016; Kwok, Xie, & Richards, 2017; Sotiriadis, 2017) dan buena muestra de ello al no mencionar ningún trabajo que se haya ocupado en estos últimos años de analizar las variaciones en la clasificación.

Es necesario, por tanto, que investigadores y expertos estudien el comportamiento de los algoritmos encargados de clasificar y ordenar los establecimientos hoteleros. Así podremos entender por qué hay ciudades como Madrid o Valencia donde los hoteles que eran número 1 del ranking TripAdvisor en junio de 2015, siguen en esa misma posición en febrero de 2018 (Hotel Orfila y Caro Hotel respectivamente), mientras que en ciudades como Barcelona o Sevilla, dos hoteles logran conquistar el puesto número uno en sus respectivas ciudades tan sólo unos meses después de su inauguración. El Hotel Mercer Sevilla, abierto en noviembre de 2016, alcanzó el primer puesto del ranking TripAdvisor un año después, en una ciudad con más de 170 hoteles, mientras que el hotel H10 Casa Mimosa, inaugurado en septiembre de 2016, llegó al número 1 del ranking de Barcelona en abril de 2017, apenas siete meses después de su apertura, en una ciudad con más de 500 hoteles.

Los ejemplos anteriores nos invitan a profundizar en el estudio de estos influyentes rankings para tratar de averiguar cómo funcionan y detectar y alertar sobre posibles ineficiencias o anomalías en su algoritmo.

Por todo ello, el primer objetivo de nuestro trabajo es el de contribuir al desarrollo de un marco conceptual que defina una nueva línea de investigación en turismo centrada en el análisis de los rankings de hoteles y restaurantes. Para ello presentaremos los principales conceptos que ayuden a definir este nuevo campo de estudio.

A continuación, expondremos detalladamente una novedosa metodología para la creación de un índice que nos permita cuantificar la movilidad de los establecimientos (ya sean hoteles o restaurantes) en el ranking de TripAdvisor. Además, se presentará una clasificación de hoteles en función del grado de movilidad.

Como tercer objetivo de nuestro estudio, queremos llamar la atención sobre las importantes implicaciones que la movilidad de los rankings puede llegar a tener para los agentes implicados en la industria hotelera, especialmente empresarios, inversores y profesionales del sector.

La estructura de nuestro trabajo es la siguiente. En primer lugar, realizamos una revisión bibliográfica sobre Tripadvisor y los rankings de hoteles. Posteriormente, en la sección 3 definimos los conceptos de ranking y presentamos las principales implicaciones de este concepto para los empresarios e inversores. En la sección 4, definimos la movilidad desde dos perspectivas: macro y micro. A continuación, detallamos los principales elementos que componen la metodología aplicada, a saber, el Vector Trayectoria y el Vector Movilidad. Además se analizan los dos componentes fundamentales de la movilidad (Tendencia y Volatilidad). Por último, formulamos los índices de Tendencia y Volatilidad, así como el índice de Movilidad. En la sección 6 presentamos una clasificación de los hoteles en función de la movilidad, mientras que las secciones 7 y 8 corresponden a la descripción de la muestra y a la presentación de resultados. Finalmente, recogemos las conclusiones y sugerencias para futuras investigaciones en esta materia.

2. Revisión bibliográfica.

Smyth et al. (2010) publican uno de los primeros trabajos que analizan y cuantifican la evolución en Tripadvisor de los hoteles, aunque en este caso tan sólo se centran en la variación de las puntuaciones a lo largo del tiempo. Estos autores estudian más de 30.000 comentarios sobre hoteles irlandeses y los comparan con 50.000 comentarios de hoteles de Las Vegas durante dos años.

Comprueban que el 64% de los hoteles irlandeses con una puntuación previa entre 2 y 3 puntos, tenían al final del periodo analizado una puntuación superior a 3. Además, descubren que mientras la puntuación media de los hoteles de Las Vegas se había mantenido constante en torno a 3,8 puntos en los últimos dos años, los hoteles de Irlanda pasaron en el mismo periodo de una puntuación media de 3,6 a 3,8.

Para los autores, estos cambios se deben a lo que ellos han llamado “El Efecto TripAdvisor”. Mientras en los hoteles de Las Vegas hace ya tiempo que se alcanzó un nivel aceptable de calidad, los hoteleros en Irlanda, conscientes de que unas valoraciones negativas en TripAdvisor podían dañar la imagen de sus establecimientos, se esforzaron por mejorar sus niveles de calidad y evitar así los comentarios negativos.

También Jurca, Garcin, Talwar y Faltings (2010) analizan la evolución a lo largo del tiempo de la posición de un hotel en el ranking, en este caso, a medida que van apareciendo nuevos comentarios. Estudian un hotel concreto de Nueva York y los resultados muestran que utilizando la mediana en lugar de la media aritmética se obtienen rankings más estables. Para estos autores, cuando la clasificación se basa en la media aritmética de las valoraciones, el usuario tiende a exagerar su puntuación para influir en la calificación global del hotel. Así, por ejemplo, un usuario cuya valoración personal (un 3) esté por debajo de la calificación global media del hotel (un 4), tiene un incentivo para dar una puntuación exageradamente negativa (un 1) sobre ese establecimiento con la intención de que la media global baje y se acerque a la puntuación que él considera debe tener ese hotel.

Por el contrario, si se utiliza la mediana, al ser ésta menos sensible a los valores extremos, el propio usuario no tendría incentivos para exagerar su puntuación (tanto positiva como negativa) y sería el primer interesado en proporcionar la valoración más exacta posible.

O'Mahony y Smith (2009) plantean el problema del exceso de información a disposición del usuario y ponen de manifiesto el gran volumen de producción de nuevas opiniones por parte de los usuarios y el hecho de que un gran número de ellas no son de interés.

Esta gran cantidad de comentarios publicados por los turistas ha generado también gran preocupación en el sector ante la posibilidad de fraude. Ong (2012) se preguntó si la cantidad de comentarios afecta a la fiabilidad percibida de los mismos concluyendo que efectivamente un alto número de opiniones puede proporcionar credibilidad. Los usuarios son conscientes de la posibilidad de que algunos comentarios sean falsos, pero cuando el número de estos es elevado (cientos) entonces el impacto de la manipulación de un comentario falso disminuye drásticamente.

Para Ott, Cardie y Hancock (2012), las webs con una baja dificultad para escribir comentarios y ponerlos al alcance de una gran audiencia, tienen un mayor nivel de falsos comentarios que aquellas páginas con mayores dificultades para exponer un comentario. Sus resultados señalan que sitios como Hotels.com tienen una tasa aproximada del 2% y TripAdvisor está en el 4% de comentarios fraudulentos.

Fruto de esta preocupación por el fraude, en 2014 la Comisión Europea (European Commission, 2014) publicó un informe específico sobre los comentarios online de los clientes en el sector hotelero donde, entre otras recomendaciones, indicaba que las webs de opiniones deben tomar medidas para verificar que sus usuarios se alojaron realmente en el establecimiento sobre el que están opinando.

En este sentido, como señala Biffaro (2015), en diciembre de 2014 la ICA (Italian Competition Authority) sancionó con una multa de 500.000 euros a Tripadvisor por infringir varios artículos del código de comercio italiano. El ICA encontró desleal la falta de medidas por parte de Tripadvisor para atajar las opiniones engañosas. Pero, como indica Hensens (2010) la mayor medida que tiene Tripadvisor para luchar contra los comentarios fraudulentos es el elevado número de opiniones por hotel. Su investigación sobre el fraude en los comentarios refleja que los hoteles con mayor número de opiniones son más insensibles a los comentarios falsos, considerando que a partir de 100 opiniones un hotel es prácticamente inmune a los comentarios fraudulentos.

Hemos querido dejar para el final un trabajo difícil de catalogar que, aunque se aleja en cierto modo del núcleo de nuestra investigación merece ser tenido en consideración por la originalidad con la que analizan la temática de los comentarios online. Para estos autores (Scott & Orlikowski, 2012) aunque los mecanismos de rankings no son nuevos, han adquirido gran capacidad de influencia y poder cuando se ejecutan a través de las tecnologías de la web 2.0, redistribuyendo la responsabilidad.

El estudio se centra en pequeños negocios de una remota zona geográfica no identificada para la que TripAdvisor ha cambiado las reglas del juego. Los autores analizan la implicación moral y estratégica de esta transformación. Estos autores han estado particularmente interesados en examinar el concepto de responsabilidad aplicado a las redes sociales, basado en la idea de la sabiduría de la multitud y la inteligencia colectiva. El poder de los ranking está en su capacidad de presentarse como una herramienta objetiva sin sesgo, mostrando la verdad, imponiendo su manera de ver las cosas pero sin ser visto.

TripAdvisor tiene un efecto material en los negocios y en la gestión. De hecho, para ciertas empresas del sector turístico una mejor o peor calificación en el ranking puede suponer la diferencia entre las ganancias o las pérdidas, sobreviviendo una temporada o teniendo que cerrar.

Por tanto, es necesario tomar en serio las relaciones de responsabilidad que se producen en TripAdvisor, ya que el resultado es en la práctica una nueva especie de autoridad que determina nuestras decisiones.

3. La movilidad de los rankings.

Como ya se indicó anteriormente, el concepto de movilidad de los rankings ha sido poco estudiado por investigadores y profesionales. Por tanto, creemos necesario presentar algunos conceptos.

Definimos en primer lugar un ranking como un sistema dinámico donde los distintos elementos que lo componen pueden cambiar de posición a lo largo del tiempo. Definimos a continuación la movilidad de un ranking como la medida de los cambios que se producen durante un periodo determinado de tiempo en las posiciones que ocupan los elementos que aparecen en la clasificación.

En función del grado de movilidad, nos encontraremos, por un lado, con ciudades donde la clasificación de los hoteles presenta poca movilidad. En este caso, los hoteles permanecen en las mismas (o similares) posiciones a lo largo del tiempo, y en consecuencia, los establecimientos situados en la zona alta del ranking seguirán muy probablemente en esa misma zona en el futuro. Igualmente, aquellos hoteles situados en las zonas media y baja también permanecerán en esas zonas en los próximos meses o años. Hablaremos en este caso de “Rankings Sólidos”.

En el extremo opuesto, nos encontraremos con rankings de máxima movilidad, donde los hoteles circulan con facilidad entre las zonas alta, media y baja. Los hoteles que hace unos meses estaban situados en las mejores posiciones, estarán ahora en la zona baja, y aquellos que ocupaban los peores puestos, serán ahora los primeros de la clasificación. Hablaremos entonces de “Rankings Líquidos”.

Cuando estudiamos la movilidad de un ranking en una ciudad concreta, estamos tratando de conocer cómo evolucionan las posiciones de los distintos hoteles en la clasificación a lo largo del tiempo. Esto tiene importantes implicaciones para los empresarios, trabajadores y profesionales del sector.

3.1. Implicaciones para los empresarios.

Podemos preguntarnos por ejemplo si los hoteleros estarían tan preocupados en mejorar su posición en la clasificación si supieran que el puesto que ocupan actualmente apenas va a variar en el futuro, porque el ranking de hoteles de su ciudad presenta un nivel de movilidad muy bajo.

En circunstancias así, los hoteles situados en la zona alta continuarán, en su mayoría, disfrutando de una posición privilegiada, en términos de visibilidad, reputación, etc., mientras que la mayor parte de los hoteles situados en la zona media y baja seguirán por largo tiempo en esas mismas (o similares) posiciones con unas probabilidades mínimas de alcanzar los mejores puestos.

Entonces, ante un ranking sólido, no tendrá sentido relacionar objetivos, políticas retributivas, incentivos al personal o medidas del desempeño a corto o medio plazo con una más que improbable mejora en la posición en el ranking.

Así mismo, las compañías hoteleras con establecimientos en varias ciudades deberían tener en cuenta el índice de movilidad del ranking de cada localidad. Sería un grave error comparar la evolución de las posiciones de sus hoteles en ciudades con índices de movilidad sensiblemente distintos. Una gestión similar puede tener efectos muy diferentes en la clasificación debido al nivel de movilidad del ranking de cada ciudad.

Las implicaciones para empresarios también afectan a los procesos de adquisición o apertura de nuevos establecimientos. Así, los inversores y las propias compañías hoteleras, deben plantearse la cuestión de si preferirán invertir en ciudades donde los rankings sean sólidos o, por el contrario, les resultarán más atractivas las ciudades con rankings líquidos, donde hay una mayor facilidad para alcanzar los mejores puestos en la clasificación, aunque también para caer a las posiciones más bajas. Para abordar esta cuestión, debemos distinguir entre la adquisición de un establecimiento ya existente o bien la apertura de uno nuevo.

Cuando la opción es comprar un hotel en funcionamiento, los inversores deben saber que, si adquieren un establecimiento situado en la zona media o baja del ranking en una ciudad con escasa movilidad, las posibilidades de situar su hotel a medio plazo en lo más alto de la clasificación son mínimas. En consecuencia, si el objetivo es llegar a estar entre los mejores en ciudades con rankings sólidos, deberían limitar sus compras a hoteles ya situados en la zona alta de la clasificación, al margen de otras consideraciones como la estrategia competitiva o la rentabilidad de la operación.

Si lo que se plantea es la apertura de un nuevo establecimiento, puede ser una gran oportunidad invertir en ciudades con poca movilidad. Un nuevo hotel tiene una gran ventaja con respecto a los ya existentes, pues tiene la posibilidad de alcanzar en muy poco tiempo las mejores posiciones, precisamente por ser nuevo. Si además hay poca movilidad, cuando logre una buena posición podrá permanecer en ella por mucho tiempo. El hotel H10 Casa Mimosa (Barcelona, España) es un buen ejemplo de esto. Se inauguró en septiembre de 2016 y llegó al número 1 del ranking de la ciudad condal en abril de 2017, apenas siete meses después de su inauguración, en una ciudad con más de 500 hoteles. En febrero de 2018 continuaba en esa posición.

Por consiguiente, en el proceso de compra de un hotel en ciudades con poca movilidad, habría que añadir una prima al precio de compraventa si el establecimiento se encuentra situado entre los mejores, ya que es muy probable que permanezca en esa misma posición a medio y largo plazo, por el hecho de estar en una ciudad con poca movilidad.

En sentido contrario, se debería aplicar un descuento al precio de compraventa en caso de que el establecimiento estuviera situado en la zona media o baja del ranking, ya que muy probablemente también seguirá en esos puestos a medio y largo plazo.

No sería extraño entonces que, ante esta falta de movilidad en las posiciones de la mayoría de los hoteles en el ranking, los hoteleros dejaran de insistir a sus clientes para que comenten sus experiencias en estas páginas si saben que no servirán para mejorar su posición. Al margen de que ciertamente estos comentarios pueden proporcionar información muy útil para corregir errores o realizar mejoras en las instalaciones y los servicios prestados, no es menos cierto que los empresarios, pueden sentir cierta sensación de estar haciéndole el trabajo gratis a TripAdvisor. Mientras el esfuerzo del empresario en captar nuevos comentarios de sus clientes no se va a traducir en una mejora de su posición en el ranking, TripAdvisor estará viendo crecer su base de datos de fotografías, comentarios y valoraciones de usuarios a coste cero.

Los rankings líquidos, donde los hoteles fluyen entre las posiciones altas, medias y bajas con mucha facilidad, podrían identificarse con una planta hotelera muy competitiva con un alto grado de homogeneidad en cuanto a la calidad de los establecimientos y por tanto con valoraciones muy similares. Aquí, una pequeña variación en la puntuación puede provocar que

un hotel pase de estar en lo más alto a estar en la zona media o baja. Pero en estos casos, el hecho de estar en la zona alta, media, o baja ya no sería tan determinante para elegir un establecimiento, pues los clientes más que un ranking, estarían consultando un listado de hoteles donde todos tendrían un nivel de calidad similar y no habría diferencias significativas entre los de arriba y los de abajo.

En cualquier caso, TripAdvisor y Booking son quienes deciden mediante el diseño de sus algoritmos de clasificación, si el ranking debe tener más o menos movilidad.

En definitiva, la movilidad de los rankings existe, aunque hasta el momento nadie la haya cuantificado. Nosotros lo estamos haciendo por primera vez, pues resulta evidente que puede ejercer una influencia considerable sobre las empresas hoteleras. Lo que no vamos a hacer es juzgarla, es decir, considerarla como una característica positiva o negativa en sí misma. Siempre encontraremos grupos interesados en que el ranking de hoteles de una ciudad sea rígido, sólido. Pero también habrá en la misma ciudad otros colectivos interesados en que el ranking sea líquido.

En cualquier caso, es evidente la complejidad de este concepto y los innumerables pros y contras que ofrece a empresarios y profesionales. La movilidad será al mismo tiempo positiva para unos y negativa para otros.

4. Micro y Macro movilidad.

El análisis de la movilidad puede realizarse desde dos puntos de vista: Micro y Macro. La Micromovilidad estudia la trayectoria individual de cada elemento de la clasificación, mientras que la Macromovilidad, analiza el ranking en su conjunto.

Con la Micromovilidad analizamos la evolución individual en la clasificación de cada establecimiento, estudiando la tendencia (ascendente, descendente o neutra). Comparamos también el nivel de movilidad de distintos hoteles o bien la movilidad de un establecimiento en varios periodos de tiempo. Tratamos de responder en definitiva a preguntas del tipo: ¿Tiene más movilidad el hotel X o el Y? ¿Ha tenido más movilidad el Hotel Z en el año 1 que en el año 2?

Con la Macromovilidad, por el contrario, comparamos los rankings de distintas ciudades, identificamos si hay más hoteles que suben, bajan o permanecen en la misma posición, medimos el número de posiciones que por término medio suben o bajan los hoteles en la clasificación de una ciudad y obtenemos los índices de movilidad de los rankings en su conjunto. Tratamos de responder en este caso a preguntas del tipo ¿Hay más movilidad en el ranking de Barcelona o de Madrid? ¿Ha aumentado la movilidad en el ranking de hoteles de Valencia respecto al año anterior?

En este trabajo queremos aportar una nueva herramienta para el análisis de la movilidad de los hoteles en el ranking con la creación de un índice de Micromovilidad, que cuantifique el grado de movilidad de la trayectoria de un hotel a lo largo del tiempo.

5. Metodología.

Para llevar a cabo el análisis de la movilidad, es necesario en primer lugar recoger de forma periódica y prolongada en el tiempo la información sobre las posiciones que ocupan los hoteles en la clasificación en cada momento.

Una vez se han obtenido todos los datos, se obtiene la trayectoria de cada hotel. Definimos la trayectoria individual de un establecimiento como el conjunto ordenado de las distintas posiciones que ha ocupado en la clasificación a lo largo de un periodo de tiempo. El análisis de estas trayectorias constituye el objeto fundamental en el estudio de la movilidad de los rankings.

5.1. El Vector Trayectoria.

La trayectoria individual de cada hotel, quedará representada matemáticamente a través de un vector que llamaremos Vector Trayectoria, que recoge las diferentes posiciones que ha ocupado a lo largo del tiempo en el ranking,

Para iniciar el estudio de la movilidad es necesario, por tanto, configurar el vector trayectoria, lo que supone fijar los siguientes tres parámetros:

1. Dimensión (T): Se define como el número de observaciones disponibles para cada hotel o elemento.
2. Periodicidad (P): Las observaciones se realizarán con periodicidad diaria, semanal, quincenal, mensual, trimestral, etc.
3. Número de zonas (cuantiles) en que se divide el ranking (Q): Es necesario determinar si el ranking se divide en deciles (Q = 10), cuartiles (Q = 4), percentiles (Q = 100), etc. En los casos de rankings pequeños (N < 100, siendo N el número total de hoteles o elementos incluidos en el ranking) o cuando se requiera una precisión máxima, puede ser preferible indicar la posición exacta que ocupa cada elemento en la clasificación. En estos casos, Q = N.

El vector trayectoria \mathbf{x} , resumirá la información de la evolución temporal que ha experimentado un componente del ranking. Este vector tiene T elementos (x_t , $t = 1, 2, \dots, T$) que representan la zona o posición que ocupa en el ranking en el momento t.

$$\mathbf{x} = (x_1 \ x_2 \ x_3 \ \dots \ x_T)$$

En estudios estadísticos y económicos es frecuente identificar los cuantiles más elevados con los valores más “favorables” o “excelentes”. Por ejemplo, Jenkins, Brandolini, Micklewright, Nolan, & Basso (2012), al analizar la renta de los hogares en 21 países, sitúan en el decil 1 a los hogares más pobres y en el decil 10 a los hogares con más renta. Milanovic (2015) al estudiar la desigualdad sitúa a los individuos con menos rentas en el percentil 1. También en el estudio sobre discriminación salarial de Río, Gradín y Cantó (2011), el decil 1 corresponde a los salarios más bajos y el decil 10 a los más altos. Por último, Budria (2010) al comparar los salarios en el sector público y privado, clasifica a los individuos en la muestra por deciles, siendo el decil 1 el de menores salarios y el decil 10 el de mayores.

En nuestro caso, dado que en un ranking de hoteles el puesto número 1 se asocia al mejor establecimiento, cuando trabajamos con cuantiles, el decil 1 corresponde a la zona mejor y el decil 10 a la zona peor del ranking.

Así, por ejemplo, para un ranking dividido en deciles, Q = 10, el vector trayectoria $\mathbf{x} = (6 \ 7 \ 5 \ 5 \ 6)$ indica que en el primer periodo analizado el hotel estaba situado en el decil 6, bajó en el segundo periodo al decil 7, subió en el tercer periodo al decil 5, continuó en el cuarto periodo en el decil 5 y finalmente bajó al decil 6.

En definitiva, el vector trayectoria es el objeto fundamental del análisis de la movilidad, porque nos ofrece la información necesaria para realizar cualquier tipo de estudio posterior, ya que recoge las distintas posiciones por las que ha pasado un hotel a lo largo del tiempo.

5.2. El Vector Movilidad.

Utilizando la información recogida en el vector trayectoria, vamos a definir un conjunto de variables que nos ayudarán a conocer el comportamiento de un hotel a lo largo del tiempo, y a

medir su grado de movilidad. Todo esto nos permitirá posteriormente clasificarlo y compararlo con otros hoteles.

Aunque estas variables las utilizaremos para crear nuestros índices de movilidad, también nos ofrecen por sí mismas información suficiente para encontrar evidencias sobre la mayor o menor movilidad de un hotel (si realizamos el análisis de forma individual) o de un ranking (si se analizan de forma agregada).

Utilizaremos cuatro variables obtenidas a partir del vector trayectoria que aportan información diversa y complementaria acerca de la movilidad de cada elemento del ranking, y que están relacionadas con el número de saltos o cambios que ha experimentado en su trayectoria (M_1) y la distancia recorrida en dichos cambios (M_2, M_3, M_4).

M_1 : Esta variable se define como el número de cambios o saltos experimentados por un elemento del ranking en la trayectoria observada x .

Si definimos la variable c_t (que refleja si ha habido cambio entre el momento $t-1$ y t) como:

$$c_t = \begin{cases} 1 & x_t - x_{t-1} \neq 0 \\ 0 & x_t - x_{t-1} = 0 \end{cases}, \quad t = 2, \dots, T \quad (1)$$

tendremos que M_1 se obtiene mediante:

$$M_1 = \sum_{t=2}^T c_t \quad (2)$$

El valor mínimo de M_1 es 0 y su máximo es $T-1$.

M_2 : Esta variable se define como la distancia media recorrida por un elemento del ranking entre cada dos periodos consecutivos, ya sea positiva o negativa, en su trayectoria observada x .

$$M_2 = \frac{1}{T-1} \sum_{t=2}^T |x_t - x_{t-1}| \quad (3)$$

El rango de valores que puede tomar M_2 es $(0, Q-1)$.

Entre toda la información que nos ofrece el análisis de las trayectorias individuales de los hoteles, quizás la variable más asociada al concepto de movilidad sea la distancia recorrida por el hotel entre su posición inicial y final ya que, de forma intuitiva, cuando queremos analizar cómo ha evolucionado un hotel a lo largo del tiempo lo primero que hacemos es comparar su posición al inicio y al final del periodo. La variable M_3 nos permite cuantificar esta información.

M_3 : Esta variable se define como la distancia entre el primer y último valor observado del vector trayectoria x :

$$M_3 = x_1 - x_T \quad (4)$$

Así, diremos que un hotel con una trayectoria (6 6 5 5 4) ha mejorado su posición dos deciles a lo largo del periodo ($M_3 = 2$), mientras que el hotel con trayectoria (3 4 5 5 6) ha empeorado su situación en el ranking entre el momento inicial y final en tres deciles ($M_3 = -3$). Esta variable toma valores positivos y negativos, indicando los primeros una mejora en la clasificación y los segundos un empeoramiento en la posición en el ranking. La variable M_3 oscila

entre un valor mínimo de $M_3 = 1 - Q$ (en nuestro ejemplo $Q = 10$), que corresponde a una caída máxima en la clasificación pasando del primer al último cuantil; y el valor máximo $M_3 = Q - 1$, que indica una subida desde la zona más baja a la zona más alta del ranking.

Aunque la distancia entre la posición inicial y final aporta gran información para determinar la movilidad de un hotel a lo largo del tiempo, es conveniente completar el análisis con la información que proporciona la distancia entre el cuantil máximo y mínimo visitado, con independencia del momento en el que se haya hecho. Así, por ejemplo, la trayectoria (3 4 5 4 3), aunque no presenta cambios entre la posición inicial y final ($M_3 = 0$), sí muestra por el contrario un cambio de dos deciles a lo largo del periodo entre el máximo y el mínimo, lo que nos indica cierto grado de movilidad, que la variable M_3 no refleja. Por este motivo, utilizamos también la variable M_4 .

M_4 : Esta variable se define como la distancia entre los valores mayor y menor del vector trayectoria observado x :

$$M_4 = \max_t(x_t) - \min_t(x_t), \quad t = 1, 2, \dots, T \quad (5)$$

Cuantifica, por tanto, la distancia entre el máximo y el mínimo cuantil por donde el establecimiento ha pasado, con independencia del momento en el que se hayan alcanzado dichas posiciones. Esta variable tiene un valor mínimo $M_4 = 0$ cuando no hay diferencias entre la posición máxima y mínima y un valor máximo $M_4 = Q - 1$.

Como vemos, todos estos indicadores ofrecen información diferente y útil para conocer la movilidad de un hotel a través de su vector trayectoria. Por ello, podemos construir el vector movilidad \mathbf{m} asociado a un vector trayectoria observado x con estas cuatro variables:

$$\mathbf{m} = (M_1 \ M_2 \ M_3 \ M_4) \quad (6)$$

A modo de ejemplo, podemos obtener el vector movilidad para un hotel cuya trayectoria en un ranking dividido en cuartiles durante cinco periodos consecutivos de tiempo ha sido $x = (1 \ 4 \ 4 \ 1 \ 3)$. Entonces, $M_1 = 3$, $M_2 = 2$, $M_3 = -2$ y $M_4 = 3$. La trayectoria x quedaría asociada al vector movilidad $\mathbf{m} = (3 \ 2 \ -2 \ 3)$.

A partir del vector movilidad de cada hotel podremos obtener el índice de movilidad, objetivo principal de nuestro estudio.

5.3. Componentes de la Movilidad.

Hemos definido la movilidad de un hotel en un ranking como la medida de los cambios que experimenta en la clasificación durante un periodo de tiempo. Esta movilidad, es el resultado del efecto conjunto de dos componentes: la tendencia y la volatilidad, para los que elaboramos sendos índices parciales de movilidad.

5.3.1. Tendencia

El concepto de tendencia ha sido ampliamente estudiado en los análisis de series temporales (Mills, 2011; Kirchgässner, Wolters, & Hassler, 2013; Box, Jenkins, Reinsel, & Ljung, 2015). En nuestro caso, la tendencia se limita a indicar la dirección de la trayectoria, y se mide por la diferencia entre la posición inicial y final del hotel en la clasificación. Así, una tendencia será ascendente si al final del periodo analizado el hotel se encuentra clasificado en mejor posición que al inicio, presentando por el contrario una tendencia descendente cuando la posición que ocupa al final del periodo es peor que la que ocupaba inicialmente. Hablaremos además de

tendencia fuerte o débil en función del valor de esas diferencias. Si no hay cambios en la posición inicial y final, diremos que el hotel presenta tendencia neutra.

Aunque ya quedó aclarado anteriormente, debemos recordar que cuando hablamos de posición, nos estaremos refiriendo bien al cuantil donde se encuentre situado el hotel (decil, cuartil, etc.) o bien a la posición exacta que ocupa en la clasificación, según se haya decidido o no dividir el ranking por zonas.

El índice de tendencia se obtiene utilizando el elemento M_3 del vector movilidad. Siendo Q el número de zonas en que se divide el ranking (deciles, quintiles, cuartiles...), el índice de tendencia (IT) se define mediante la siguiente expresión:

$$IT = \frac{M_3}{Q - 1} \quad (7)$$

Este índice nos muestra la relación entre el valor que toma la variable M_3 para un hotel y un periodo concreto (en el numerador), y el valor máximo que esa variable puede tomar ($Q - 1$).

El índice IT varía, por tanto, entre +1 y -1, donde $IT = +1$ (tendencia máxima ascendente) nos indicará que el hotel ha pasado de la peor a la mejor zona del ranking, mientras que $IT = -1$ corresponderá a un hotel que ha pasado de la mejor a la peor zona (tendencia máxima descendente). Por último, hablaremos de tendencia neutra, $IT = 0$, cuando la posición inicial y final del hotel en el ranking sea la misma ($M_3 = 0$). Además, hablaremos de tendencia fuerte, media o débil en función del valor que tome el índice de tendencia.

5.3.2. Volatilidad

La segunda componente de la movilidad es la volatilidad, que mide la variabilidad en el comportamiento de un hotel en el ranking y vendrá determinada en primer lugar por el número de cambios producidos en su trayectoria a lo largo del tiempo, y en segundo lugar por la dimensión de dichos cambios. Mediremos por tanto la volatilidad de un hotel por el número de saltos que se producen en su trayectoria y por la amplitud de esos saltos, cuantificados por las variables M_1 y M_2 del vector movilidad m .

Así, cuanto mayor sea el número de saltos que presenta el hotel a lo largo de su trayectoria, y mayor la amplitud de los mismos, mayor será la volatilidad de ese hotel. Si no hay saltos en toda la trayectoria, diremos que el hotel presenta volatilidad nula.

La componente derivada del número de saltos se define como la relación entre el número de saltos (M_1) en la trayectoria del hotel (numerador) y el número máximo de saltos posibles, ($T - 1$) donde T es el número total de periodos analizados.

Por tanto, el primer componente del Índice de volatilidad se obtiene:

$$IV_1 = \frac{M_1}{T - 1} \quad (8)$$

y toma valores entre 0 y 1. IV_1 será igual a 0 cuando no se produzca ningún salto en toda la trayectoria, lo que indicará que el hotel ha permanecido en la misma posición durante todo el periodo de tiempo analizado. En cambio, IV_1 será igual a 1 cuando se haya producido un salto en la clasificación en cada uno de los periodos analizados.

El segundo componente de la volatilidad cuantifica la amplitud de esos saltos. Para construir el segundo componente del índice de volatilidad, normalizaremos el elemento M_2 del

vector movilidad m , que es el promedio de las diferencias en valor absoluto entre las posiciones cada dos periodos consecutivos, dividiéndolo por el máximo que puede alcanzar, $Q-1$:

$$IV_2 = \frac{M_2}{Q-1} \quad (9)$$

De esta forma, IV_2 toma valores entre 0 y 1.

Finalmente, el índice de volatilidad, IV , estará compuesto por el promedio de los dos índices parciales IV_1 e IV_2 :

$$IV = \frac{IV_1 + IV_2}{2} \quad (10)$$

En general, la interpretación del índice de volatilidad depende de ambas componentes. Tendremos, en primer lugar, hoteles con mucha volatilidad, cuando tengan muchos saltos en su trayectoria, y estos sean de gran amplitud. Hablaremos de hoteles con poca volatilidad, cuando tengan pocos saltos y estos además sean de poca amplitud. Por último, nos encontraremos con trayectorias con un nivel medio de volatilidad cuando o bien tengan muchos saltos, pero de poca amplitud, o bien tengan pocos saltos, pero estos sean de gran amplitud.

Es importante señalar que los saltos vendrán condicionados por el modo en que definamos la posición de un hotel en el ranking. Si deseamos medir la posición con la máxima precisión posible consideraremos un salto cuando el hotel cambie su posición exacta en el ranking entre dos momentos consecutivos, al pasar por ejemplo del puesto 186 al 193.

Pero si por el contrario hemos considerado más conveniente definir la posición en el ranking por el cuantil en que se encuentre el hotel (decil, percentil, quintil, etc.) en lugar de su posición exacta, habrá un salto si el hotel está situado en cuantiles distintos en dos momentos consecutivos. Nótese entonces que, aunque la posición exacta haya variado, el hotel puede permanecer en el mismo cuantil durante los periodos analizados y por tanto no se habrá producido ningún salto.

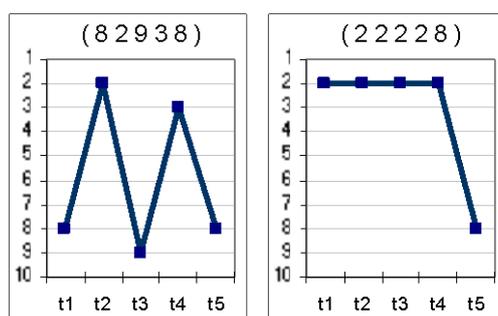
5.3.3. El Índice de Movilidad

Dado que nuestro objetivo en este trabajo es diseñar un índice que cuantifique la movilidad de la trayectoria de un hotel en un periodo de tiempo determinado y que como hemos definido anteriormente, la movilidad tiene dos componentes, el índice de movilidad de la trayectoria de un hotel en un ranking (IM) será un índice compuesto. Estará formado por la combinación del índice de volatilidad IV y del índice de tendencia IT , dando la misma ponderación a ambos. Teniendo en cuenta que el índice de tendencia puede tener signo positivo o negativo, tomaremos su valor absoluto para construir el índice de movilidad:

$$IM = \frac{1}{2}|IT| + \frac{1}{2}IV \quad (11)$$

El hecho de que la tendencia y la volatilidad tengan el mismo peso a la hora de medir la movilidad en nuestra propuesta, es una cuestión que plantea algunas controversias. Ilustraremos esta discusión con dos ejemplos extremos, los hoteles A y B, con trayectorias respectivas (8 2 9 3 8) y (2 2 2 2 8), representadas en el gráfico 1.

Gráfico 1. Ejemplo: trayectorias de A y B.



Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

El Hotel A presenta una trayectoria con muchos cambios en su posición en el ranking a lo largo del tiempo (mucha volatilidad) pero que al final del periodo se encuentra en la misma posición que al inicio (tendencia neutra). En cambio, el Hotel B permanece sin cambios durante la mayor parte del tiempo (mínima volatilidad) pero en el último periodo varía drásticamente su posición en el ranking, lo que provoca una gran diferencia entre la posición inicial y final (fuerte tendencia).

Es obligado entonces responder a la siguiente pregunta: ¿Qué hotel debemos considerar que tiene más movilidad? Según nuestro modelo los dos presentan un índice de movilidad similar en valores absolutos (0,417 para el hotel A y 0,438 para el hotel B) como se recoge en la tabla 1.

Tabla 1. Ejemplo: movilidad de los hoteles A y B.

Hotel	Vector trayectoria	Vector movilidad	Índices				
			IV ₁	IV ₂	IV	IT	IM
A	(8 2 9 3 8)	(4 6 0 7)	1,00	0,667	0,833	0,000	0,417
B	(2 2 2 2 8)	(1 1,5 -6 6)	0,25	0,167	0,208	-0,667	0,438

Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Decidir si el hotel A tiene un grado de movilidad mayor o menor que el hotel B siempre será un asunto problemático. En estos dos casos extremos elegidos por el investigador para ilustrar la cuestión, podría discutirse sin llegar a un consenso cuál de los dos hoteles presenta un mayor grado de movilidad.

Unos analistas defenderán que el hotel A es claramente más dinámico que el hotel B, si nos fijamos en los constantes cambios producidos en su posición. Pero también habrá quien defienda que el hotel B es más dinámico que el A, si lo que importa es la diferencia entre la posición inicial y la final.

Desde nuestro punto de vista, consideramos que la movilidad tiene dos componentes (la volatilidad y la tendencia) y que ambas tienen el mismo peso, y por tanto para tener un alto grado de movilidad es necesario tener un alto grado tanto de volatilidad como de tendencia.

En todo caso, creemos que el considerar la movilidad como combinación de estas dos variables nos permite caracterizar mejor la trayectoria y por tanto explicar mejor el comportamiento de un hotel.

También queremos destacar, que el grado de movilidad no permanece constante a lo largo del tiempo. De hecho, en el ejemplo anterior, si hubiéramos concluido nuestro análisis en el

periodo t_4 en lugar de t_5 , el hotel A hubiera presentado una movilidad extrema y el hotel B una movilidad nula.

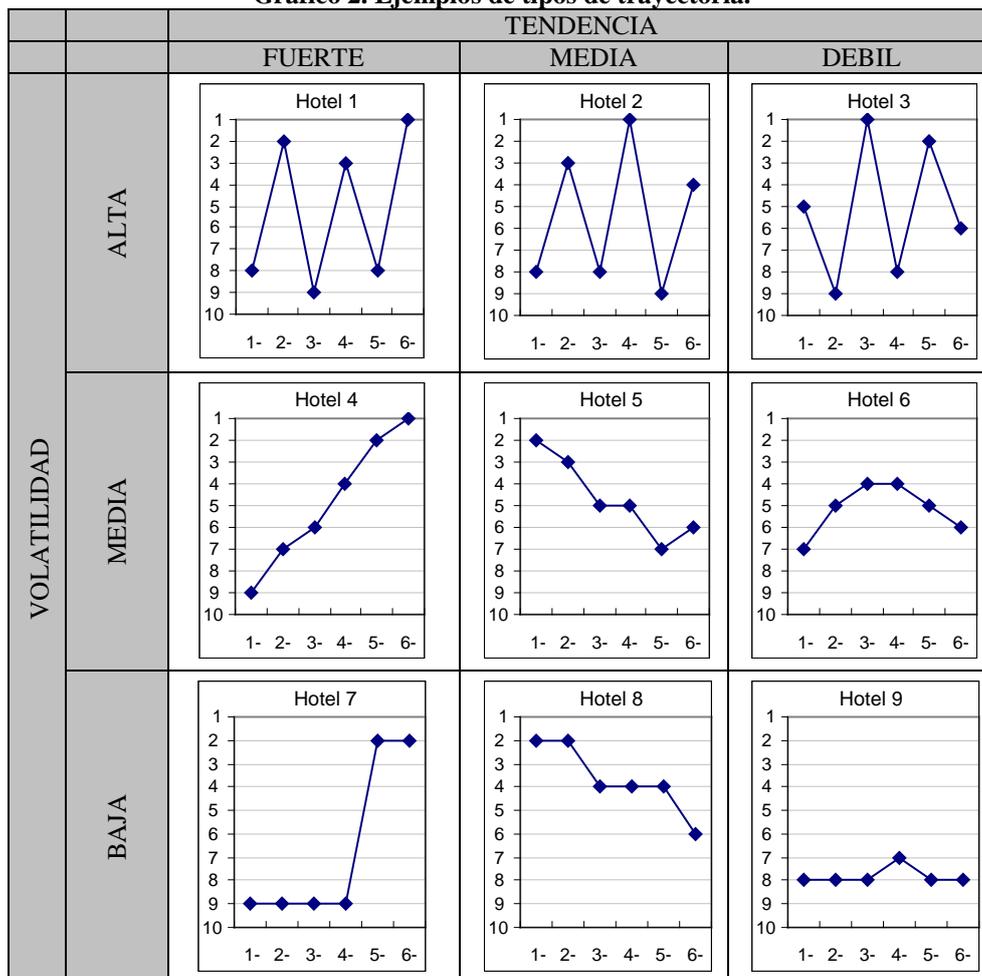
Todas estas circunstancias, lejos de ser complicaciones para el análisis, son interesantes retos para abrir nuevas líneas de investigación sobre el comportamiento de la movilidad. Habrá hoteles que presenten un grado de movilidad similar a lo largo del tiempo mientras que otros presentarán cambios bruscos en su movilidad como así ocurre en nuestros dos ejemplos anteriores.

6. Clasificación de los hoteles por su movilidad.

Si hemos definido la movilidad como el resultado de la combinación de volatilidad y tendencia, podríamos clasificar las trayectorias en función de los indicadores de volatilidad y de tendencia. Con fines descriptivos, y para facilitar el análisis consideraremos “alta/fuerte” una componente cuyo indicador (en valor absoluto en el caso del índice de tendencia) supere el valor 0,7. En el extremo opuesto, la consideraremos “baja/débil” si el indicador es inferior a 0,3. Con valores comprendidos entre 0,3 y 0,7 denominaremos “media” la componente de la movilidad.

En el gráfico 2 aparecen representadas diferentes trayectorias que nos podríamos encontrar al analizar la evolución en el ranking de un hotel a lo largo del tiempo en función de su grado de volatilidad (alta, media y baja) y de tendencia (fuerte, media débil), para un ranking clasificado en deciles y observado en seis momentos del tiempo.

Gráfico 2. Ejemplos de tipos de trayectoria.



Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Para las trayectorias de los ejemplos del gráfico 2, se han obtenido los vectores de movilidad y sus correspondientes índices que se muestran en la tabla 2.

Tabla 2. Vectores movilidad e índices de los ejemplos del gráfico 2.

Hotel	Vector movilidad	Índices				
		IV ₁	IV ₂	IV	IT	IM
1	(5 6,2 7 8)	1	0,689	0,844	0,778	0,811
2	(5 6 4 8)	1	0,667	0,833	0,444	0,639
3	(5 5,8 -1 8)	1	0,644	0,82	-0,11	0,47
4	(5 1,6 8 8)	1	0,178	0,589	0,889	0,739
5	(4 1,2 -4 5)	0,8	0,133	0,47	-0,44	0,46
6	(4 1 1 3)	0,8	0,111	0,456	0,111	0,283
7	(1 1,4 7 7)	0,2	0,156	0,178	0,778	0,478
8	(2 0,8 -4 4)	0,4	0,089	0,24	-0,44	0,34
9	(2 0,4 0 1)	0,4	0,044	0,222	0	0,111

Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Teniendo en cuenta la clasificación propuesta, se pueden catalogar los hoteles en función de su grado de movilidad, medido por el índice de movilidad IM del siguiente modo:

1. Hoteles con Movilidad Alta ($IM > 0,7$): Corresponde a hoteles con distintas combinaciones de tendencia fuerte y media con volatilidad alta y media.
2. Hoteles con Movilidad Media ($0,3 \leq IM \leq 0,7$): Corresponde a hoteles cuyas trayectorias presentan distintas combinaciones de tendencia fuerte, media y débil con volatilidad alta, media y baja.
3. Movilidad Baja ($IM < 0,3$): Corresponde a hoteles que han mostrado trayectorias con distintas combinaciones de tendencia media y débil con volatilidad media y baja. Aunque el valor 0 estará por tanto incluido en este apartado, a efectos informativos se indicará en una categoría especial (Movilidad Nula) dada su notoriedad.

7. Muestra de hoteles.

Aunque este trabajo forma parte de un proyecto de investigación más amplio que tiene por objetivo estudiar el funcionamiento y la evolución de los rankings de hoteles y restaurantes en las principales ciudades del mundo, en este artículo analizamos los rankings de hoteles de Tripadvisor de las cinco mayores ciudades de España por número de habitantes (Instituto Nacional de Estadística, 01/01/2015): Madrid (3.141.991), Barcelona (1.604.555), Valencia (786.189), Sevilla (693.878) y Zaragoza (664.953).

Desde junio de 2015 y de forma ininterrumpida se han ido extrayendo semanalmente las posiciones que los distintos hoteles ocupan en los rankings de estas ciudades. Debido a que la movilidad detectada tanto semanal como mensual era mínima, se decidió utilizar los datos trimestrales, donde ya se aprecian ciertos niveles de movilidad.

Por tanto, el presente trabajo analiza la evolución trimestral a lo largo de un año (de junio de 2015 a junio de 2016) de los rankings de las cinco primeras capitales de provincia españolas por número de habitantes (ver Tabla 3).

De este modo, para el periodo objeto de estudio, disponemos de cinco posiciones en el ranking para cada hotel (Jun-15, Sep-15, Dic-15, Mar-16 y Jun-16). Una vez procesada esta información, se han obtenido los vectores trayectoria para cada establecimiento, reflejando la evolución de sus posiciones a lo largo del periodo.

Tabla 3. Datos de la muestra.

Ciudad	(1) Hoteles en la muestra	(2) Hoteles en Tripadvisor	(3) = (1) / (2) %
Barcelona	271	524	51,72%
Madrid	250	447	55,93%
Sevilla	104	176	59,09%
Valencia	55	100	55,00%
Zaragoza	47	80	58,75%
Total	727	1.327	54,79%

Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Debemos señalar, por último, que los rankings han sido divididos en deciles ($Q = 10$), por lo que las posiciones que toman los hoteles variarán de 1 a 10 (donde el decil 1 corresponde a la zona más alta y el decil 10 a la zona más baja).

Del total de hoteles que aparecen en el ranking de Tripadvisor (1.327), se eliminaron de la muestra los establecimientos que no tenían comentarios, ni a la fecha de inicio de la investigación ni en ninguno momento posterior. Estos son pequeños establecimientos sobre los que nadie ha opinado. Como no tienen valoraciones de clientes tampoco tienen una puntuación global media. En estos casos TripAdvisor no asigna al hotel ninguna posición numérica y los sitúa al final del ranking.

Los autores, como es habitual en investigaciones de carácter longitudinal, también se han encontrado con elementos faltantes en algún momento. La pérdida o no disponibilidad de una proporción variable de los datos correspondientes a los sujetos seleccionados es siempre una realidad a considerar en todo estudio de investigación (Durán, 2005) y uno de los problemas metodológicos en estudios longitudinales (Twisk & De Vente, 2002).

En nuestro caso, se han suprimido aquellos hoteles con datos faltantes en al menos un periodo. En consecuencia, la muestra está compuesta por todos aquellos hoteles para los que se dispone de información en todos y cada uno de los momentos en que se tomaron datos durante el estudio.

No obstante, es importante identificar las causas de esta pérdida de información para realizar una adecuada interpretación de los resultados (Twisk & De Vente, 2002). Así, en nuestro caso, las razones por las que un hotel puede no aparecer en alguno de los momentos en que se tomaron los datos son básicamente cuatro:

1. Apertura de nuevos hoteles: Si un hotel no existía al inicio de nuestra investigación en junio de 2015, pero se inauguró en algún momento posterior, tendremos datos faltantes en los primeros periodos.
2. Cierres: Si un hotel estaba abierto al público en junio de 2015 pero ha cesado su actividad en algún momento posterior, nos encontraremos con datos faltantes en los últimos periodos de nuestro estudio.
3. Incidencias técnicas: El proceso de extracción de datos se realiza de forma automatizada por lo que cualquier incidencia técnica en servidores, ordenadores, conexión a Internet, o cambios en el diseño de la página web de TripAdvisor puede provocar la existencia de datos faltantes.

4. Por último, no podemos descartar errores humanos en la manipulación y el procesamiento de la información.

Los datos perdidos durante una investigación pueden causar principalmente dos tipos de problemas. En primer lugar, al reducirse el número de casos analizados se puede producir una pérdida de potencia o precisión en los resultados. En segundo lugar, los datos faltantes pueden afectar a la validez de los resultados debido a posibles sesgos (Deeg, 2002; Durán 2005).

En nuestro caso, y aunque se podrían aplicar diversas técnicas para reemplazar matemáticamente los datos perdidos (Engels & Diehr, 2003), no lo hemos considerado necesario ya que la muestra (727 hoteles con datos en todos los periodos) era suficientemente representativa, por lo que ni la precisión ni la validez de los resultados se han visto afectadas de forma significativa por los datos faltantes.

8. Resultados.

En este apartado aplicaremos a los hoteles seleccionados en la muestra la metodología de análisis de la movilidad del ranking de Tripadvisor expuesta en la sección 5. Primero analizaremos los datos en conjunto y posteriormente realizaremos un análisis comparado entre las ciudades.

8.1. Análisis de la muestra.

Para realizar el análisis de los índices hemos segmentado los valores obtenidos en cuatro categorías, de acuerdo con la metodología descrita en la sección anterior: Nivel Nulo, Bajo/Débil, Medio y Alto/Fuerte.

Además, para facilitar el análisis estudiaremos todos los índices con signo positivo (es decir, los valores del índice de tendencia se toman en valor absoluto), dado que para el presente estudio no es tan relevante el signo como la cuantía del índice.

La tabla 4 muestra la distribución de frecuencias para los tres índices: Volatilidad (IV), Tendencia (IT) y Movilidad (IM) del conjunto total de hoteles en la muestra. Se observa que el porcentaje de hoteles con índice nulo es muy significativo en los tres casos, especialmente en el caso del índice de tendencia donde el 76,62% de los hoteles presenta un valor igual a 0, lo que indica que para estos establecimientos no ha habido cambios entre su posición inicial y final. Agrupando los índices en los segmentos antes señalados, comprobamos que para los tres índices el porcentaje del segmento bajo/débil es el mayoritario. Sin embargo, la distribución de frecuencias de los tres índices muestra diferencias significativas.

Si clasificamos todas las trayectorias observadas en la muestra conjuntamente por su grado de volatilidad y su nivel de tendencia, de acuerdo con la metodología propuesta, (tabla 5) se observa que el 88,31% de los hoteles tienen Volatilidad Baja y Tendencia Débil, por lo que podemos concluir esta sección señalando que es muy elevado el porcentaje de hoteles con un nivel de movilidad bajo o nulo.

Tabla 4. Distribución de frecuencias de los índices de movilidad (%).

TIPO	NIVEL	IV	IT	IM
Alto/Fuerte	(0,9 - 1]	0%	0%	0%
	(0,8 - 0,9]	0%	0%	0%
	(0,7 - 0,8]	0%	0%	0%
Medio	(0,6 - 0,7]	0%	0%	0%
	(0,5 - 0,6]	2,61%	0%	0%
	(0,4 - 0,5]	7,98%	0,14%	0,14%
	(0,3 - 0,4]	1,10%	0,14%	3,16%
Bajo/Débil	(0,2 - 0,3]	21,46%	1,10%	9,22%
	(0,1 - 0,2]	33,29%	22,01%	26,82%
	(0 - 0,1]	0%	0%	27,10%
Nula	0	33,56%	76,62%	33,56%
TOTAL		100%	100%	100%

Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Tabla 5. Clasificación de las trayectorias observadas (%).

Volatilidad	Tendencia			Total
	Fuerte	Media	Débil	
Alta	0%	0%	0%	0%
Media	0%	0,28%	11,42%	11,69%
Baja	0%	0%	88,31%	88,31%
Total	0%	0,28%	99,72%	100%

Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

8.2. Análisis por ciudades

Analizamos en esta sección la movilidad de los hoteles de la muestra distinguiendo por la ciudad en la que están ubicados. La Tabla 6 recoge los porcentajes de hoteles para cada segmento de los tres índices.

Tabla 6. Distribución de los índices por ciudades (%).

	Barcelona	Madrid	Sevilla	Valencia	Zaragoza
IT					
Fuerte	0,0%	0,0%	0,0%	0,0%	0,0%
Media	0,0%	0,0%	1,0%	0,0%	2,1%
Débil	100,0%	100,0%	99,0%	100,0%	97,9%
IV					
Alta	0,0%	0,0%	0,0%	0,0%	0,0%
Media	8,5%	14,0%	13,5%	3,6%	23,4%
Baja	91,5%	86,0%	86,5%	96,4%	76,6%
IM					
Alta	0,0%	0,0%	0,0%	0,0%	0,0%
Media	2,6%	3,6%	3,8%	0,0%	8,5%
Baja	97,4%	96,4%	96,2%	100,0%	91,5%

Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Respecto al índice de tendencia (IT), se aprecia un comportamiento similar en todas las ciudades con unos niveles de tendencia débil superiores al 97% en todos los casos, llegando al 100% en tres de las cinco ciudades.

En cuanto al índice de volatilidad (IV), aunque no se encuentran casos de volatilidad alta en ninguna ciudad, al menos sí se aprecia un número considerable de hoteles con volatilidad media. No obstante, sigue siendo mayoritario el porcentaje de hoteles con volatilidad baja, con valores superiores al 86% en todas las ciudades con la excepción de Zaragoza (76,6%).

La distribución del índice de movilidad (IM) también se encuentra muy sesgada hacia los valores bajos, con cifras superiores al 91% en todas las ciudades para este nivel de movilidad, destacando especialmente Valencia donde el 100% de los hoteles muestran movilidad baja.

Por último, también se aprecia cierta correlación entre el tamaño de la ciudad y la movilidad del ranking. La ciudad más pequeña (Zaragoza) es la que presenta más movilidad (menor porcentaje de hoteles con movilidad Baja/Débil en los tres índices). Y la ciudad más grande en número de hoteles, Barcelona, es la que presenta una movilidad menor (porcentajes de movilidad Baja/Débil más altos).

En la tabla 7 presentamos los valores medios de las distribuciones de los índices de movilidad por ciudades.

Tabla 7. Índices de movilidad (medias) por ciudades.

Índice	Barcelona	Madrid	Sevilla	Valencia	Zaragoza
IV	0,1363	0,171	0,1769	0,1419	0,1973
IT	0,0241	0,024	0,0363	0,0262	0,0543
IM	0,0802	0,0975	0,1066	0,084	0,1258

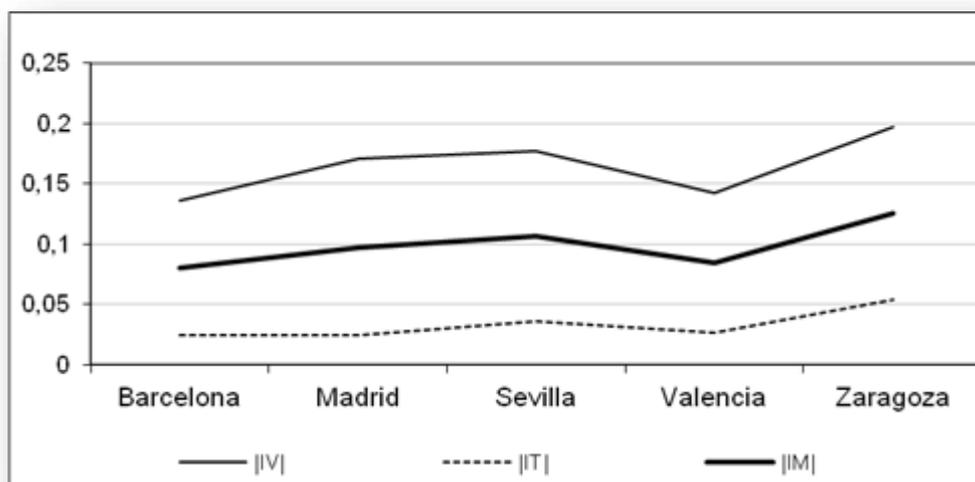
Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Los datos de la tabla 7 nos muestran unos valores muy bajos en todas las ciudades. No obstante, se podría apreciar en primer lugar que el índice de volatilidad presenta unas cifras superiores en todas las ciudades al índice de tendencia. Esto nos indicaría que hay cierto nivel de movilidad en términos de volatilidad, es decir, hay movimientos entre deciles pero no suponen una variación significativa en el ranking a medio plazo (índice de tendencia). Los hoteles estarían alternando subidas y bajadas sin efecto final en la clasificación.

Además, como se aprecia en el gráfico 3, que recoge los valores de los tres índices para las cinco ciudades, hay cierta relación entre el tamaño de la ciudad (medido por el número de hoteles) y la movilidad.

Se observa que, con la excepción de Valencia, hay cierta tendencia a aumentar el valor de los índices a medida que disminuye el tamaño de la ciudad. En cualquier caso, todos los índices nos señalan un grado de movilidad muy bajo en todas las ciudades.

Gráfico 3. Índices por ciudades.



Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Para confirmar si las ciudades se comportan de un modo similar se ha realizado el análisis de la varianza, tomando como factor la ciudad, cuyos resultados se presentan en la tabla 8, donde se aprecian diferencias significativas entre las ciudades.

Tabla 8. Análisis de la varianza por ciudades.

Índices	Suma de cuadrados	GL	F	Pr(>F)
IV	0,2939	4	3,4281	0,008677 **
IT	0,0480	4	4,2009	0,002276 **
IM	0,1236	4	3,7683	0,00483 **

**Significativo al 1%

Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Adicionalmente, también se ha realizado el test no paramétrico de Kruskal-Wallis, para reforzar las conclusiones acerca de las diferencias por ciudades. Las medianas por ciudad y los resultados de este test, para cada índice aparecen en la tabla 9.

Tabla 9. Medianas y test de Kruskal-Wallis por ciudades.

Índices	Medianas					Test de Kruskal-Wallis		
	Barcelona	Madrid	Sevilla	Valencia	Zaragoza	χ^2	g.l.	p-valor
IV	0,1388	0,1388	0,1388	0,1388	0,1388	15,1605	4	0,00438**
IT	0	0	0	0	0	11,123	4	0,02522*
IM	0,0694	0,0694	0,0694	0,0694	0,125	14,5893	4	0,005633*

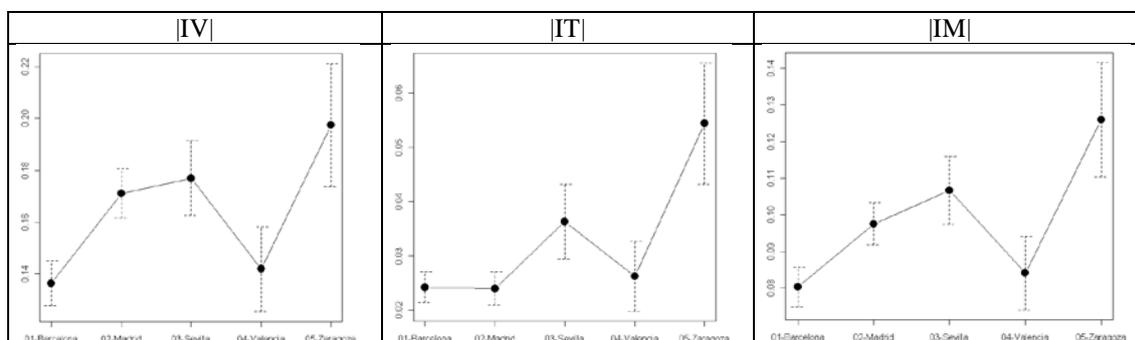
* Significativo al 5%, **Significativo al 1%

Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

Los resultados son significativos y confirmamos que existen diferencias en la movilidad de los rankings entre las distintas ciudades.

Como se aprecia nuevamente en los gráficos de medias (gráfico 4), hay cierta relación entre el tamaño de la ciudad y la movilidad del ranking. Más evidentes para los índices IV y IM (donde Valencia rompe una clara línea ascendente).

Gráfico 4. Índices (en valor absoluto). Gráfico de medias por ciudades.



Fuente: Elaboración propia a partir de los datos extraídos de TripAdvisor (2016).

9. Conclusiones.

Aunque la influencia que los rankings de hoteles ejercen en el cliente a la hora de tomar la decisión de donde alojarse es ampliamente reconocida tanto en el ámbito empresarial como académico, (Ghose et al., 2012; Callarisa-Fiol et al., 2012), pocos han sido los que han estudiado lo que hemos llamado la movilidad de los rankings, esto es, cómo varían las posiciones en la clasificación de los distintos establecimientos a lo largo del tiempo.

De hecho, en las últimas revisiones bibliográficas (Chen & Law, 2016; Kwok et al., 2017; Sotiriadis, 2017) no se menciona ningún trabajo que se haya ocupado de analizar la evolución de estos rankings.

Dada la revolución que han supuesto para el sector turístico compañías como TripAdvisor o Booking (Xiang & Gretzel, 2010; Ong, 2012), y sin poner en ningún momento en duda la honestidad y rigor de quienes diseñan y gestionan estos rankings, resulta sorprendente que hasta el momento no exista ningún tipo de supervisión o control por parte de instituciones públicas o privadas sobre los mismos, que auditen y verifiquen su correcto funcionamiento.

Nuestro trabajo ha querido aportar luz sobre unos rankings que en estos momentos ejercen una influencia excesiva sobre el sector (Mateos, Martín, & Martín-Velicia, 2014), y adolecen de falta de transparencia en su funcionamiento (Mellinas et al., 2015).

En este sentido, como señalan Scott y Orlikowski (2012), TripAdvisor está configurado de tal manera que produce una forma de transparencia para los comentarios de los turistas mientras que oscurece el mecanismo que ordena esos comentarios.

Así, por una parte, estas plataformas son consideradas transparentes al poner a disposición del usuario final una inabarcable cantidad de información en distintos formatos (puntuaciones, comentarios, fotografías), con valoraciones tanto positivas como negativas aportadas por otros usuarios que resultan de gran utilidad a la hora de seleccionar un establecimiento.

Pero al mismo tiempo que se pone a disposición del usuario toda esta información, se están ocultando los mecanismos utilizados para ordenar esos rankings. Más allá de unas leves nociones sobre el funcionamiento del algoritmo encargado de la configuración del ranking (TripAdvisor, 2016), no podemos conocer con claridad cómo se establece la valoración final del hotel y, por tanto, su posición en la clasificación.

El gran valor de TripAdvisor como creador de rankings no está en su algoritmo, sino en su capacidad de atraer a cientos de miles de usuarios de todo el mundo para que compartan sus opiniones, valoraciones y experiencias. Por eso entendemos que esta opacidad no está en ningún caso justificada, dando argumentos a quienes sospechan de comportamientos poco éticos en la elaboración de estos rankings.

En cuanto a las aportaciones teóricas de este trabajo, consideramos que un aspecto esencial de un ranking es su grado de movilidad, ya que mediante el estudio de esta variable podemos detectar posibles anomalías en su funcionamiento. Por ello hemos querido contribuir a la creación de un marco conceptual que permita desarrollar una línea de investigación apenas explorada en turismo hasta la fecha, centrada en el estudio de la movilidad en los rankings de hoteles más influyentes: TripAdvisor y Booking. Para ello se han definido los principales conceptos en los que se basa este nuevo campo de estudio.

Al mismo tiempo, presentamos una herramienta para medir la movilidad de los hoteles analizando las trayectorias individuales de los distintos establecimientos. Así, se ha formulado un índice de movilidad como combinación de otros dos índices (tendencia y volatilidad).

Respecto a las contribuciones prácticas, se han aplicado los tres índices anteriores a una muestra de hoteles formada por 727 establecimientos situados en las cinco primeras ciudades de España por número de habitantes (Madrid, Barcelona, Valencia, Sevilla y Zaragoza).

Los resultados indican unos niveles de movilidad mínimos. Así, a nivel global, el 96,7% de los hoteles presentan un índice de movilidad bajo o nulo. Y por ciudades, todas sin excepción muestran niveles de movilidad muy bajos (siempre inferiores a 0,13 para un índice que toma valores entre 0 y 1), si bien los test estadísticos nos indican que hay diferencias significativas entre las distintas capitales.

En este sentido se aprecia cierta relación entre el nivel de movilidad y el tamaño de la ciudad (en número de hoteles). Así, Barcelona, la ciudad con una mayor planta hotelera (más de 500 hoteles en TripAdvisor) es la que presenta un índice de movilidad más bajo ($IM = 0,0802$) mientras que la más pequeña, Zaragoza (con 80 hoteles en TripAdvisor), es la que presenta mayores índices de movilidad ($IM = 0,1258$).

Las implicaciones de los resultados obtenidos afectan a los distintos agentes de la industria hotelera (principalmente empresarios, inversores y profesionales del sector). Así, los establecimientos situados en la zona alta de estos rankings tienen una elevada probabilidad de seguir en esa privilegiada zona durante mucho tiempo, dada su escasa movilidad. Dos ejemplos nos ilustran claramente esta realidad: los hoteles que eran número 1 del ranking TripAdvisor en Madrid (Hotel Orfila) y Valencia (Caro Hotel) en junio de 2015, seguían en esa misma posición en febrero de 2018, en ciudades con más de 400 hoteles la primera y 100 la segunda.

Del mismo modo, los hoteles situados en la zona media y baja permanecerán con una alta probabilidad en esas mismas zonas a medio plazo. Ante esta situación, puede resultar contraproducente fijar objetivos, establecer incentivos y medidas del desempeño basadas en una más que improbable mejora en la clasificación.

Los inversores, deberían negociar los precios de compraventa de los establecimientos hoteleros teniendo en cuenta no solo la posición que ocupan en el ranking en un momento dado, sino también el grado de movilidad de estos. En las ciudades analizadas (con rankings sólidos) habría que aplicar una prima sobre el precio inicial cuando el hotel esté situado en una zona de privilegio, pues muy probablemente seguirá ahí en los próximos años. Por el contrario, si el hotel ocupa una posición en la zona media o baja, se debería aplicar un descuento al precio de compra pues también seguirá (con una alta probabilidad) en esa misma zona del ranking a medio plazo.

Finalmente, entre las limitaciones de nuestro trabajo debemos reconocer que sería conveniente ampliar la investigación a ciudades de otros países y continentes, pues la muestra utilizada está formada únicamente por ciudades españolas. También sería conveniente aplicar el modelo a rankings desarrollados por otras webs (Booking.com), a ciudades con un estilo diferente de turismo (sol y playa) e incluso a otro tipo de establecimientos (restaurantes).

Para concluir, queremos invitar a investigadores y expertos en turismo y otras disciplinas a colaborar en este nuevo reto para aportar luz sobre unos rankings que han alcanzado hoy día una influencia inimaginable hace tan solo unos años.

Referencias.

- Ayeh, J. K., Au, N., & Law, R. (2013). "Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. *Journal of Travel Research*, 52(4), 437-452.
- Biffaro, L. (2015). Unfair Commercial Practices and Online Consumer Reviews: the Italian Tripadvisor Case. *Rivista Italiana di Antitrust/Italian Antitrust Review*, 2(1), 181-185.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*, 5th Ed. New York: Wiley.
- Budria, S. (2010). Schooling and the distribution of wages in the European private and public sectors. *Applied Economics*, 42(8), 1045-1054.
- Callarisa-Fiol, L.J., Sánchez-García, J., Moliner, M.Á., & Forgas-Coll, S. (2012). La importancia de las comunidades virtuales para el análisis del valor de marca. El caso de Tripadvisor en Hong Kong y París. *Papers de turisme*, 54, 89-115.
- Chen, Y.F., & Law, R. (2016). A Review of Research on Electronic Word-of-Mouth in Hospitality and Tourism Management. *International Journal of Hospitality & Tourism Administration*, 17(4), 347-372.
- Deeg, D.J. (2002). Attrition in longitudinal population studies: Does it affect the generalizability of the findings? *Journal of Clinical Epidemiology*, 55(3), 213-215.
- Durán, P. (2005). Los datos perdidos en estudios de investigación ¿son realmente datos perdidos? *Archivos argentinos de pediatría*, 103(6), 566-568.
- Eco, U. (2011). El vértigo de las listas. *IC Revista Científica de Información y Comunicación*, 8, 15-34.
- Engels, J. M., & Diehr, P. (2003). Imputation of missing longitudinal data: a comparison of methods. *Journal of Clinical Epidemiology*, 56(10), 968-976.
- EC-European Commission (2014). Study on Online Consumer Reviews in the Hotel Sector. Final report. Recuperado de: <http://bookshop.europa.eu/en/study-on-online-consumer-reviews-in-the-hotel-sector-pbND0414464/>
- Filieri, R., Algezau, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51, 174-185.

- Ghose, A., Ipeirotis, P.G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3), 493–520.
- González, S.M., Gidumal, J.B., & López-Valcárcel, B.G. (2010). La participación de los clientes en sitios web de valoración de servicios turísticos. El caso de Tripadvisor. *Revista de Análisis Turístico*, 10, 17-22.
- Hensens, W. (2010). *The reliability of data from guest reviews on Tripadvisor as a contemporary social media platform*. Paper submitted and presented at SAIMS 2010 in Mpekwini Resort.
- Instituto Nacional de Estadística (2015). Cifras oficiales de población de los municipios españoles: Revisión del Padrón Municipal. Madrid: Ministerio de Economía y Empresa. Recuperado de: https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica_P&cid=1254734710990
- Jeacle, I., & Carter, C. (2011). In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. *Accounting, Organizations and Society*, 36(4), 293-309.
- Jenkins, S.P., Brandolini, A., Micklewright, J. Nolan, B., & Basso, G. (2012). The Great Recession and its consequences for household incomes in 21 countries. En Jenkins, S.P., Brandolini, A., Micklewright, J., & Nolan, B. (eds.). *The Great Recession and the Distribution of Household Income*. Oxford: Oxford University Press, pp.33-89.
- Jurca, R., Garcin, F., Talwar, A., & Faltings, B. (2010). Reporting incentives and biases in online review forums. *ACM Transactions on the Web*, 4(2), 1-27.
- Kladou, S., & Mavragani, E. (2015). Assessing destination image: An online marketing approach and the case of TripAdvisor. *Journal of Destination Marketing & Management*, 4(3), 187-193.
- Kirchgässner, G., Wolters, J., & Hassler, U. (2013) *Introduction to Modern Time Series Analysis*. Berlin: Springer-Verlag.
- Kwok, L., Xie, K. L., & Richards, T. (2017). Thematic framework of online review research: A systematic analysis of contemporary literature on seven major hospitality and tourism journals. *International Journal of Contemporary Hospitality Management*, 29(1), 307-354.
- Lee, H. A., Law, R., & Murphy, J. (2011). Helpful reviewers in TripAdvisor, an online travel community. *Journal of Travel & Tourism Marketing*, 28(7), 675-688.
- Mateos, J. C., Martín, M.A.R., & Martín-Velicia, F. A. (2014). Disrupción en el uso de la web 2.0 en los hoteles independientes. *Cultura, desarrollo y nuevas tecnologías: VII Jornadas de investigación en turismo*, 169-183.
- Melián, S., Bulchand, J., & González, B. (2010). La participación de los clientes en sitios web de valoración de servicios turístico. El caso de TripAdvisor. *Análisis Turístico*, 10, 17-22.

- Mellinas, J. P., Martínez, S.M., & Bernal, J.J. (2015). Booking.com: The unexpected scoring system. *Tourism Management*, 49, 72-74.
- Miguéns, J., Baggio, R., & Costa, C. (2008). Social media and tourism destinations: TripAdvisor case study. *Advances in Tourism Research*, 26(28), 1-6.
- Milanovic, B. (2015). Global Inequality of Opportunity: How Much of Our Income Is Determined by Where We Live? *The Review of Economics and Statistics*, 97(2), 452-460.
- Mills, T.C. (2011). *The foundations of modern time series analysis*. Basingstoke, U.K.: Palgrave, McMillan.
- Molinillo, S., Ximénez-de-Sandoval, J. L., Fernández-Morales, A., & Coca-Stefaniak, A. (2016). Hotel assessment through social media: the case of TripAdvisor. *Tourism & Management Studies*, 12(1), 15-24.
- O'Connor, P. (2008). User-generated content and travel: A case study on Tripadvisor.com. In: O'Connor P., Höpken W., Gretzel U. (eds), *Information and Communication Technologies in Tourism 2008*. Springer, Vienna.
- O'Connor, P. (2010). Managing a hotel's image on TripAdvisor. *Journal of Hospitality Marketing & Management*, 19(7), 754-772.
- O'Mahony, M.P., & Smyth, B. (2009). *Learning to recommend helpful hotel reviews*. Proceedings of the third ACM conference on Recommender systems - RecSys '09, 305.
- Ong, B. S. (2012). The Perceived Influence of User Reviews in the Hospitality Industry. *Journal of Hospitality Marketing & Management*, 21(5), 463-485.
- Ott, M., Cardie, C., & Hancock, J. (2012). *Estimating the Prevalence of Deception in Online Review Communities*. Proceedings of the 21st international conference on World Wide Web WWW 12, 201-210.
- Río, C., Gradín, C., & Cantó, O. (2011). The Measurement of Gender Wage Discrimination: The Distributional Approach Revisited. *The Journal of Economic Inequality*, 9(1), 57-86.
- Scott, S.V, & Orlikowski, W. J. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. *Accounting, Organizations and Society*, 37(1), 26-40.
- Smyth, P.C.B., Wu, G., & Greene, D. (2010). *Does TripAdvisor makes hotels better?* Derek Greene School of Computer Science & Informatics, University College Dublin Belfield.
- Sotiriadis, M.D. (2017). Sharing tourism experiences in social media: A literature review and a set of suggested business strategies. *International Journal of Contemporary Hospitality Management*, 29(1), 179-225.
- TripAdvisor. (2016). Changes to the Tripadvisor Popularity Ranking Algorithm. Recuperado de: <https://www.tripadvisor.com/TripAdvisorInsights/n2701/changes-tripadvisor-popularity-ranking-algorithm>
- Tuominen, P. (2011). *The influence of TripAdvisor consumer-generated travel reviews on hotel performance*. University of Hertfordshire Business School Working Paper.

- Twisk, J., & De Vente, W. (2002). Attrition in longitudinal studies: how to deal with missing data. *Journal of Clinical Epidemiology*, 55(4), 329-337.
- Vásquez, C. (2011). Complaints online: The case of TripAdvisor. *Journal of Pragmatics*, 43(6), 1707-1717.
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179-188.