



FACULTAD DE CIENCIAS ECONÓMICAS Y
EMPRESARIALES

**Oportunidades del uso de *Big Data* en
empresas del sector de alojamiento.
Implicaciones para la política de
competitividad del sector turístico a partir
del caso de la Unión Europea**

Autor: Gonzalo Sanz-Magallón Delhaize

Director: Dr. Carlos Martínez de Ibarreta Zorita

Madrid

Abril, 2018

Resumen

Este estudio aborda las oportunidades que el uso de *Big Data* ofrece a las empresas del sector de alojamiento. En primer lugar, se realiza un análisis crítico de la literatura reciente publicada al respecto. A través de la revisión de los estudios más recientes, se establece una definición de este concepto, se identifican las fuentes de datos y herramientas de las que disponen las empresas de este sector para su explotación, así como el coste que supone su implementación y las posibles aplicaciones a incorporar a su negocio.

Por otra parte, mediante el análisis estadístico de una base de datos de elaboración propia a partir de estadísticas de Eurostat, caracterizaremos el uso de *Big Data* en el sector de alojamiento en los países de la Unión Europea, estudiando la posible relación entre el uso de esta tecnología y el nivel de desarrollo de los países, así como entre el uso de *Big Data* y el nivel de rentabilidad de las empresas del sector.

Creemos que las conclusiones de esta investigación podrán ser de interés para el diseño de la política de promoción de competitividad turística en España y en el conjunto de la Unión Europea, economías en las que la aportación del sector turístico es muy relevante. Los resultados obtenidos a su vez sugieren que la adopción de *Big Data* mejora la rentabilidad del sector, lo que podría contribuir a aumentar las diferencias de renta entre países de la Unión Europea en el futuro. Cabe por ello plantear nuevas políticas públicas para promocionar la competitividad empresarial en las economías nacionales y regionales menos avanzadas. Por otra parte, los resultados muestran que el tamaño empresarial es un factor asociado a la adopción de esta tecnología, lo que sitúa a las empresas de menor dimensión en posición de desventaja.

Palabras clave: *Big Data*, innovación, sector de alojamiento, Unión Europea, niveles de renta, competitividad, tamaño empresarial.

Abstract

This study addresses the opportunities that the use of Big Data offers to companies in the accommodation sector. It starts with a critical analysis of recent literature published in this topic. Thus, through the analysis of the most recent studies, we will establish a definition of this concept, the data sources and tools available to companies, as well as the costs that would imply, and the possible applications that could be implemented for their business.

Also, by means of an statistical analysis of an own-elaboration database using Eurostat statistics, we will characterize the use of Big Data in the accommodation sector in the countries of the European Union, studying the possible relationship between the use of this technology and the development level of the countries, as well as between the use of Big Data and the level of profitability of companies in the sector.

The conclusions from this research may be of interest for the design of the public policy that try to promote tourism competitiveness in Spain and in rest of the European Union, economies in which the tourism industry plays a key role. The results obtained suggest that the adoption of Big Data could contribute to increase income differences between European Union countries in the future, so public policies should be proposed to promote competitiveness in less advanced economies. Also, the results show that business size is a factor associated with the adoption of this technology, which places smaller companies at a disadvantage competitive position.

Key words: *Big Data*, innovation, accommodation sector, European Union, income levels, competitiveness, company size.

Índice de contenido

1. Introducción	6
1.1. El uso de <i>Big Data</i> en el sector empresarial.....	6
1.2. Objetivos de la investigación	7
1.3. Metodología	8
1.4. Estructura	9
2. Aplicación de <i>Big Data</i> en el sector de alojamiento	11
2.1. Innovación como base de la competitividad empresarial en el siglo XXI	11
2.2. Etapas de la actividad hotelera: el ciclo del huésped	12
2.3. Fuentes de <i>Big Data</i> en el sector de alojamiento	13
2.3.1. Fuentes en la etapa del Pre-Arribo.....	14
2.3.2. Fuentes en la etapa del Arribo	15
2.3.3. Fuentes en la etapa de la Estadía.....	15
2.3.4. Fuentes de datos en la etapa de la Partida.....	16
2.3.5. Usos de las fuentes de datos.....	16
2.4. Técnicas de explotación de <i>Big Data</i> en el sector empresarial	16
2.4.1. Presentación de las técnicas de explotación de <i>Big Data</i> más comunes: <i>Hadoop MapReduce</i> y sistemas <i>DBMS</i>	16
2.4.2. Análisis de la “necesidad de negocio” como paso fundamental para decidir acerca de la implementación de <i>Hadoop</i>	18
2.4.3. Conclusiones en cuanto a la necesidad de implementación de herramientas de explotación de <i>Big Data</i>	19
2.5. Aplicaciones de <i>Big Data</i> a nivel empresarial en el sector de alojamiento.....	20
2.5.1. Ventajas del uso de <i>Big Data</i> en el sector del turismo.....	20
2.5.2. Casos de aplicación exitosa de <i>Big Data</i> en el sector empresarial.....	21
2.5.3. Posibles aplicaciones de <i>Big Data</i> en el sector de alojamiento.....	22
3. Análisis empírico: aplicación de <i>Big Data</i> en el sector de alojamiento en la Unión Europea	28
3.1. Principales características del uso de <i>Big Data</i> en el sector de alojamiento en la Unión Europea..	28
3.2. Planteamiento de las hipótesis de investigación.....	30
3.3. Fuentes, datos y variables empleadas.....	30
3.4. Análisis descriptivo de las variables	31
3.5. Desarrollo econométrico y resultados	33
3.5.1. Relación entre uso de <i>Big Data</i> y PIB per cápita.....	35
3.5.2. Relación entre uso de <i>Big Data</i> y rentabilidad empresarial del sector de alojamiento	36
4. Conclusiones y discusión a partir de los resultados obtenidos	38
5. Bibliografía	39
6. Anexos	42

Índice de gráficos

Gráfico I: Las 3Vs del Big Data	7
Gráfico II: Relación entre PIB pc de los países y su capacidad innovadora	12
Gráfico III: El ciclo del huésped	13
Gráfico IV: Porcentaje de empresas en la Unión Europea que usan <i>Big Data</i> por sectores de actividad..	28
Gráfico V: Uso de <i>Big Data</i> en la Unión Europea según tamaño de empresa.....	29
Gráfico VI: Fuentes de <i>Big Data</i> en el sector de alojamiento en la Unión Europea	29
Gráfico VII: Relación entre uso de <i>Big Data</i> y PIB pc en la UE	35
Gráfico VIII: Relación entre la rentabilidad y el uso de <i>Big Data</i> en el sector de alojamiento en la UE .	37
Gráfico IX: Relación entre la rentabilidad y los costes de personal por empleado en el sector de alojamiento en la UE	37

Índice de tablas

Tabla I: Fuentes de recolección de datos según etapas del ciclo del huésped	14
Tabla II: Recomendaciones de elección entre <i>Hadoop</i> y DBMS según las necesidades de la empresa....	20
Tabla III: Aplicaciones de <i>Big Data</i> en el sector de alojamiento.....	23
Tabla IV: Tabla de valores relativos de las variables definidas en países de la Unión Europea	32
Tabla V: Coeficientes de correlación lineal de Pearson entre las variables empleadas	34
Tabla VI: Estadísticos ecuación uso de <i>Big Data</i> (+250 e.)	35
Tabla VII: Estadísticos ecuación rentabilidad sector de alojamiento	36

1. Introducción

1.1. El uso de *Big Data* en el sector empresarial

Desde hace numerosos años, los datos se han convertido en uno de los activos más importantes a explotar por las empresas (Fosso Wamba, Akter, Edwards, Chopin y Gnanzou, 2015). En este sentido, el artículo más popular del año 2012 publicado por *Advertising Age*, líder global de noticias relativas al *marketing*, se tituló simplemente *Data Dominates*, a través del cual la agencia expone que nunca dos palabras han revolucionado tanto la industria desde el *social media*. El uso de la información difiere según su ámbito de aplicación, ya que en él se incluye desde la predicción de infecciones mortales y de incendios, la posibilidad de anticipar el mejor momento para comprar un billete de avión, el acceso al último dato de inflación en tiempo real, hasta el control de las redes sociales para identificar tendencias actuales (Mayer-Schönberger y Cukier, 2013).

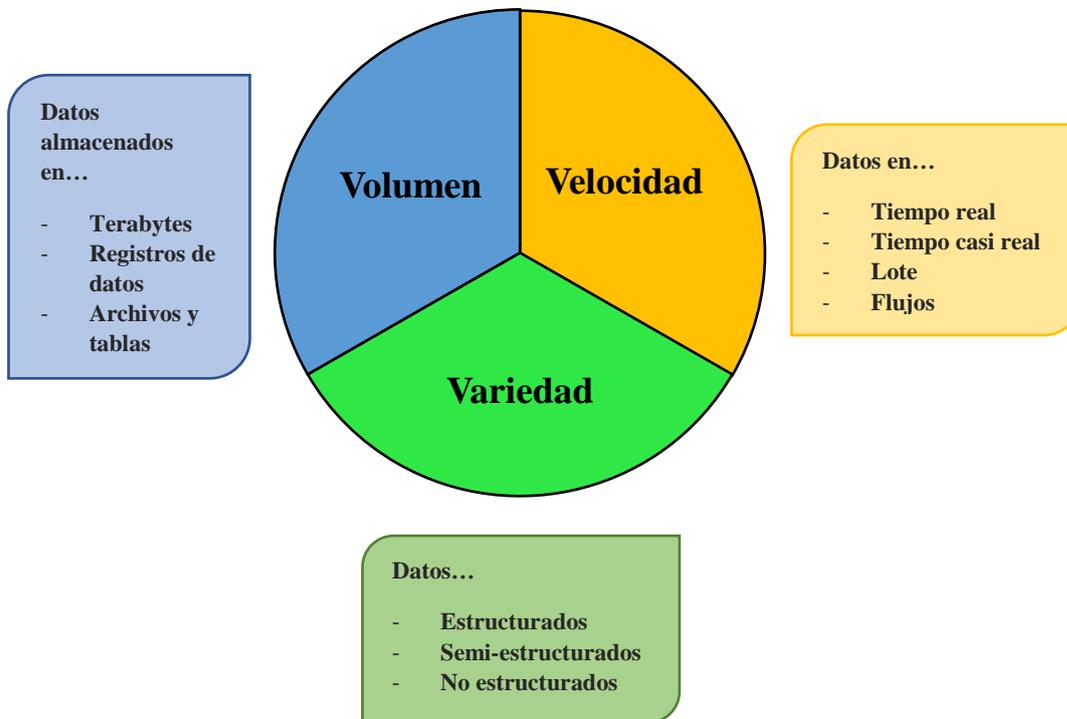
En el ámbito empresarial – área en el que se centrará este estudio – el objetivo es aprovechar el océano de información que crean los usuarios cada día para beneficiarse de una mejor toma de decisiones gracias al surgimiento de un nuevo fenómeno tecnológico: el *Big Data* (Camargo-Vega, Camargo-Ortega y Joyanes-Aguilar, 2015). Para ello, las empresas deben realizar grandes cambios en su infraestructura de almacenamiento (Redmond y Wilson, 2012), y en la arquitectura de administración de la información (Beyer y Laney, 2012), con el objeto de crear una ventaja competitiva (Lynn, 2013) mediante el análisis de estas enormes cantidades de datos para conseguir información y conocimiento que conlleve una mejor toma de decisiones estratégicas (López, 2013).

Históricamente, el término *Big Data* se ha venido utilizando en la esfera científica para hacer referencia a enormes conjuntos de datos cuya correcta interpretación requería el uso de superordenadores, pero a día de hoy no existe una definición universal, lo que explica que la Organización Mundial de Normalización (ISO) haya creado un grupo de trabajo para desarrollar la nueva norma de vocabulario – ISO 3435-5 – dedicada al área de *Big Data* y de la analítica predictiva (Maté Jiménez, 2014). A continuación, se presentan algunas de las definiciones más relevantes, y seleccionaremos la que mejor se adecúe a los objetivos de este trabajo.

La definición básica que podríamos establecer con respecto al *Big Data* hace referencia a “*bases de datos que no pueden ser almacenados, gestionados y procesados por ordenadores básicos dentro de un margen de error aceptable*” (Chen, Mao y Liu, 2014). Para el analista Kusnetzky (2010), el término hace referencia a “*las herramientas, procesos y procedimientos que permitan a una organización crear, manipular y administrar grandes conjuntos de datos e instalaciones de almacenamiento*”. Por su parte, Laney (2001) incorporó a la definición de *Big Data* el modelo de las “3Vs” que caracterizan este fenómeno: Volumen, Velocidad y Variedad. Según este autor, la primera “V” definidora de este fenómeno es la velocidad, y hace referencia a que la información que incluye se genera prácticamente en tiempo real, lo cual representa una enorme ventaja para la toma de decisiones de las empresas. El volumen es su segunda característica distintiva, en tanto en cuanto la cantidad de datos que engloba este fenómeno se mide en Terabytes y en Petabytes, es decir, el equivalente a 10^{12} y 10^{15} bytes respectivamente, dejando muy atrás las medidas tradicionales en Gigabytes (10^9 bytes). Finalmente, la tercera “V” representa la variedad, la cual hace alusión a la inmensa variedad de fuentes de las que proceden los datos englobados bajo esta definición, incluyendo desde tablas e informes, hasta datos generados a través de clics de ratón, redes sociales o páginas webs.

Exponemos estas características en el **Gráfico I**.

Gráfico I: Las 3Vs del Big Data



Fuente: elaboración propia.

Este modelo es recogido por Beyer y Laney (2012), que definen *Big Data* como “*activos de información caracterizados por su volumen elevado, velocidad elevada y alta variedad, que demandan soluciones innovadoras y eficientes de procesado para la mejora del conocimiento y la toma de decisiones en las organizaciones*”, añadiendo estos autores una cuarta “V”: la veracidad. Chen, Mao, Zhang y Leung (2014) añadieron una “V” más: el valor, para hacer esta definición perfecta. Por tanto, podemos afirmar que se ha llegado a producir una definición altamente consensuada de este fenómeno: “*Big Data representa aquellos activos de información caracterizados por un alto volumen, velocidad y variedad que requieren tecnología específica y métodos analíticos para su transformación en valor*” (De Mauro, Greco y Grimaldi, 2015).

1.2. Objetivos de la investigación

En primer lugar, el presente estudio revisa y organiza la literatura científica acerca del fenómeno del *Big Data* en el ámbito empresarial, prestando especial atención a su uso en el sector de alojamiento. Se abordarán las fuentes a disposición de las cadenas hoteleras y otras plataformas de servicios de alojamiento para explotar esta tecnología, así como las técnicas o herramientas más comunes para su implementación y el coste potencial que les podría suponer a las empresas de este sector incorporarlas a su actividad empresarial, así como las posibles aplicaciones de esta

tecnología que pueden darse en esta industria, todo ello con vistas a convertir esta práctica en una ventaja competitiva duradera frente a la competencia.

En segundo lugar, a través de la explotación mediante métodos estadísticos y econométricos de una base de datos de creación propia a partir de Eurostat (Oficina Europea de Estadística), estableceremos si el marco teórico se ajusta a la realidad de este sector a día de hoy. Para ello, analizaremos posibles asociaciones entre el uso de *Big Data* por parte de las empresas del sector de alojamiento y distintas variables que se definen en el siguiente apartado de Metodología. Se pretende caracterizar la situación del uso de esta tecnología en esta industria en los países europeos, analizando las posibles implicaciones que el uso de *Big Data* pueda tener sobre la competitividad de este sector.

De forma detallada, trataremos de dar respuesta a las siguientes preguntas de investigación:

- 1.1. ¿Qué fuentes y herramientas de *Big Data* están a disposición de las empresas pertenecientes al sector de servicios de alojamiento? ¿Qué coste les supone implementar las herramientas de explotación del *Big Data*?
- 1.2. ¿Qué posibles aplicaciones pueden aplicar en su negocio las empresas del sector de alojamiento gracias al uso de *Big Data*?
- 2.1. ¿Cómo puede caracterizarse el uso de *Big Data* en este sector a nivel europeo?
- 2.2. ¿Existe un déficit de adopción en los países con menor nivel de desarrollo económico?
- 2.3. ¿Puede concluirse que el uso de *Big Data* está íntimamente relacionado con el nivel de rentabilidad de las empresas y de los sectores?

Esta investigación pretende proporcionar conclusiones de utilidad tanto para las Administraciones públicas como para las empresas del sector privado. En cuanto a las primeras, los programas nacionales o europeos que pretenden promocionar la innovación en el sistema turístico, como la iniciativa *Smart Destination* lanzada por SEGITTUR, deberán ser capaces de ayudar a superar las posibles barreras a la innovación y a potenciar la competitividad de este sector. Si se confirma la existencia de una menor adopción de estas técnicas en países o regiones con menor nivel de desarrollo económico, estas actuaciones serán especialmente necesarias para reducir las desigualdades de renta entre países y regiones de la Unión Europea. Por otro lado, la investigación propuesta permitirá a los actores del sector privado enriquecer sus conocimientos sobre esta materia, obteniendo una visión global y actual del uso de *Big Data* a nivel empresarial, junto a los efectos beneficiosos sobre su negocio y los factores que los propician, así como la posibilidad de plantearse los riesgos y limitaciones que podría suponer el no adoptar el uso de esta tecnología con respecto a sus competidores en términos de competitividad.

1.3. Metodología

En primer lugar, revisaremos críticamente la literatura que hemos destacado en el primer apartado, ya que los estudios que resaltan los efectos beneficiosos del uso de *Big Data* pueden sacar conclusiones precipitadas o sesgadas. Hemos usado para ello las bases de datos EBSCO Business Source Premier y Google Scholar, utilizando palabras clave como “*Innovation*”, “*Big Data*”, “*Marketing*”, “*Hotel*”, “*Management*”, “*Accommodation*”, “*Applications*”, “*Airbnb*”, “*Analytics*” y “*Cost-Benefit Analysis*”, entre otras. En cuanto al análisis de la validez de la literatura, destacamos que se han encontrado estudios recientes y relevantes en la materia concreta que nos concierne, publicados en fuentes de reconocido prestigio como *Harvard Business Review*, *McKinsey Global Institute*, *Journal of Hospitality and Tourism Technology*, *International Journal*

of *Production Economics*, *Academy of Management Journal*, o *Cornell Hospitality Quarterly*, entre otras.

En un segundo lugar, emplearemos información estadística de Eurostat, que incluye un apartado dedicado a la implementación de *Big Data* en el ámbito empresarial y en el sector de alojamiento, con datos referidos a 2016. A través del uso de métodos estadísticos y econométricos, trataremos de revelar información que permita extraer conclusiones que den respuesta a las preguntas de investigación 2.1, 2.2 y 2.3.

Las estadísticas de Eurostat utilizadas son *Annual detailed enterprise statistics for services* (NACE Rev. 2 H-N and S95); *National Accounts 8ESA* (2010) y *Digital Economy and Society: Big Data Analysis*, disponibles en <http://ec.europa.eu/eurostat/data/database>, las cuales han permitido elaborar una base de datos de corte transversal homogénea que contiene para cada país de la Unión Europea las siguientes variables:

- **Información sobre *Big Data***
 - Porcentaje de empresas que usan *Big Data* (total nacional, según sectores de actividad y según tamaño de empresa);
 - Porcentaje de empresas del sector de alojamiento que usan *Big Data* con datos procedentes de: a) Dispositivos inteligentes o sensores de la empresa; b) Geolocalización de dispositivos portátiles; c) Generados por redes sociales; d) Otras fuentes;
 - Porcentaje de empresas del sector de alojamiento en que el análisis de *Big Data* es realizado por: a) empleados propios; b) un proveedor de servicios externo.
- **Información sobre características y resultados de las empresas del sector de alojamiento**
 - Número de empresas, número de empleados y tamaño medio empresarial;
 - Principales magnitudes económicas del sector de alojamiento: facturación, valor añadido, productividad, excedente bruto de explotación, costes de personal, inversiones en activos materiales, etc.
- **Características económicas de los Estados miembros de la Unión Europea**
 - PIB per cápita;
 - Número de visitantes al año a establecimientos de alojamiento;
 - Población.

1.4. Estructura

El presente estudio se dividirá en tres apartados principales. El primer capítulo, Aplicación de *Big Data* en el sector de alojamiento, partirá del concepto de innovación y de su importancia como base de la competitividad empresarial, para a continuación determinar el concepto de *Big Data*, las fuentes de datos a disposición de las empresas de esta industria y las técnicas más comunes para su explotación, así como los posibles usos que el *Big Data* ofrece a empresas de este sector.

En el segundo apartado, Análisis empírico: implementación de *Big Data* en el sector de alojamiento en la Unión Europea, se presentará la investigación realizada a partir de la base de datos con estadísticas de Eurostat. A partir de ella, contrastaremos nuestras hipótesis iniciales y daremos respuesta a los objetivos de investigación marcados.

Finalmente, el apartado de Conclusiones y discusión a partir de los resultados obtenidos recogerá y analizará conjuntamente los resultados tanto de la parte de revisión de literatura como del análisis empírico realizado.

Incluiremos por último un apartado de Bibliografía, que incluirá todas las fuentes bibliográficas empleadas para la realización del presente estudio y otro de Anexos, que presentará información adjuntada como parte separada del trabajo.

2. Aplicación de *Big Data* en el sector de alojamiento

2.1. Innovación como base de la competitividad empresarial en el siglo XXI

La actividad innovadora es en la actualidad una fuente esencial para la generación de ventajas competitivas de las empresas en la mayoría de los sectores productivos (OECD, 2005; Schwab y Sala-i-Martin, 2018). El Manual de Oslo publicado por la OCDE identifica cuatro tipos de innovaciones: de producto, proceso, organización y marketing, según las siguientes definiciones (OECD, 2005):

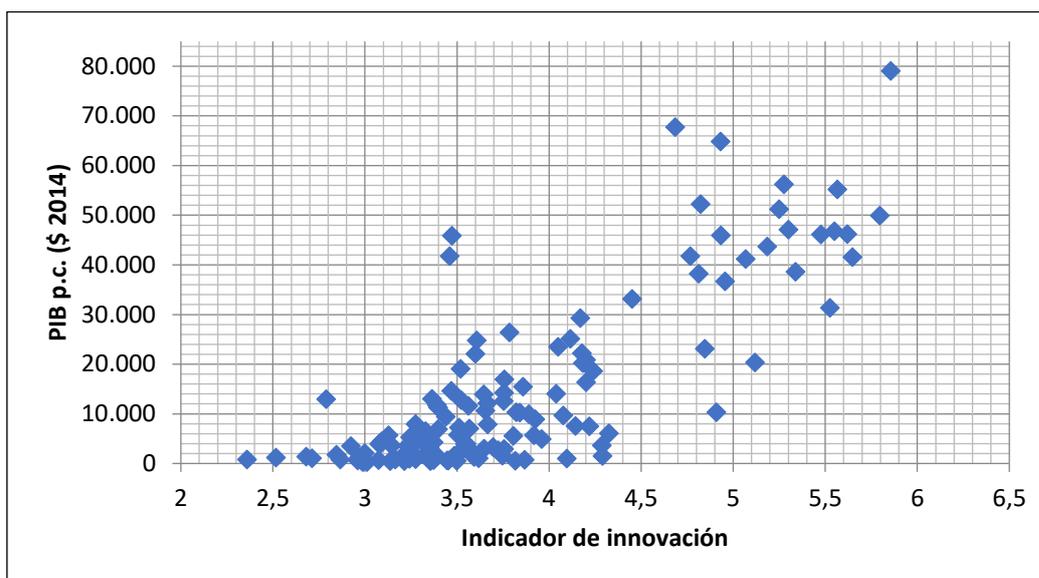
- a) Innovación de producto: una innovación de producto es un bien o un servicio nuevo o sensiblemente mejorado con respecto a sus características básicas, especificaciones técnicas, software incorporado u otros componentes intangibles, finalidades deseadas o prestaciones;
- b) Innovación de proceso: una innovación de proceso consiste en una tecnología de producción nueva o sensiblemente mejorada, así como en métodos nuevos y sensiblemente mejorados de suministro de servicios y de entrega de productos. El resultado debe ser significativo con respecto al nivel del volumen de producción, la calidad de los productos (bienes o servicios) o los costes de producción y distribución;
- c) Innovación de gestión: una innovación de gestión consiste en un método de organización nuevo o sensiblemente mejorado, así como en métodos nuevos y sensiblemente mejorados de suministro de servicios y de entrega de productos; e
- d) Innovación de marketing: una innovación de marketing es la introducción de un nuevo método de comercialización que implica cambios significativos del diseño o del envasado, de la colocación, la promoción o de la tarificación de un producto o servicio.

En la mayoría de las actividades de servicios, las actividades innovadoras incluyen varios tipos de innovación (OECD, 2005). Así, en el caso del uso de *Big Data* en establecimientos de alojamiento, como se verá más adelante, se corresponde tanto con una innovación de proceso, como de gestión y de marketing.

En general, la actividad innovadora de las empresas requiere disponer de recursos y capacidades internas a la empresa, y se ve a su vez influida por el entorno económico-cultural-tecnológico en el que opera la empresa, pudiendo identificarse distintas estrategias innovadoras entre los países desarrollados (Kao, 2009). Por otra parte, los indicadores que tratan de comparar la capacidad innovadora de las economías nacionales muestran una superior capacidad de las empresas ubicadas en países con un mayor nivel de desarrollo. El Ranking de Competitividad Mundial, publicado anualmente por el Foro Económico Mundial, calcula para las principales economías del mundo un indicador representativo de la capacidad innovadora¹, el cual, como muestra el **Gráfico II**, está claramente relacionado con el nivel de renta.

¹ El indicador se corresponde con el promedio de seis variables, todas ellas con un valor entre 0 y 7: Capacidad para la innovación, Calidad de las instituciones de investigación científica, Gasto de las empresas en I + D, Colaboración universidad-industria en I + D, Adquisición gubernamental de productos tecnológicos avanzados, Disponibilidad de científicos e ingenieros y patentes por cada millón de habitantes.

Gráfico II: Relación entre PIB pc de los países y su capacidad innovadora



Fuente: elaboración propia a partir de World Competitiveness Report 2017-2018, disponible en http://www3.weforum.org/docs/GCR2017-2018/GCI_Dataset_2007-2017.xlsx

2.2. Etapas de la actividad hotelera: el ciclo del huésped

Antes de proceder a detallar las distintas fuentes de datos de las que se nutre un hotel u otra empresa de servicios de alojamiento, procedemos a detallar las etapas del denominado “ciclo del huésped” (Kasavana y Brooks, 2001). Este proceso consiste en dividir el servicio que ofrece una empresa hotelera a un huésped en cuatro etapas perfectamente delimitadas e identificadas, cada una de las cuales posee funciones propias y consideraciones particulares, como podemos ver a través del **Gráfico III**. Usaremos esta fragmentación del proceso de alojamiento del huésped para clasificar las fuentes, métodos y aplicaciones del *Big Data* en apartados subsiguientes.

La primera etapa de este proceso es el “Pre-arribo”, etapa que comprende todas las actuaciones del potencial huésped hasta la creación de la reserva por el agente de reservas, acto que constituye el punto de partida del “ciclo del huésped”.

La siguiente etapa es el “Arribo”, que incluye el registro del huésped en el hotel, así como el proceso de asignación del cuarto. En esta etapa, el huésped entabla una relación comercial con el establecimiento a través de la recepción del hotel. El proceso de registro finaliza una vez que el huésped confirma un método de pago y su fecha de salida, otorgándosele la llave de su cuarto. Por tanto, esta etapa finaliza en el momento en que el huésped hace *check-in* en su cuarto.

La tercera etapa es la de la “Estadía”, etapa en la que el personal del hotel desempeña un papel fundamental, ya que se ocupa de coordinar las peticiones de todos los huéspedes, además de proporcionarles cualquier tipo de información. La seguridad del huésped también es un tema nuclear en esta etapa, y se incluyen aquí los cargos de distinta naturaleza (restaurante, internet, gimnasio, spa...) que se le imputarán por los servicios consumidos. Por último, la etapa “Partida” incluye el abandono del cuarto, satisfacción de las cuentas, devolución de la llave del cuarto y salida del hotel. También puede incluirse el relleno de un formulario de satisfacción acerca de la estancia del huésped en el hotel.

Gráfico III: El ciclo del huésped



Fuente: elaboración propia.

2.3. Fuentes de *Big Data* en el sector de alojamiento

El *Big Data* se genera diariamente a través de numerosas fuentes, que incluyen Internet, transacciones de teléfonos móvil, contenido generado por los usuarios, redes sociales, así como contenido “capturado” deliberadamente a través de redes de sensores, transacciones comerciales, y muchos otros ámbitos operativos como la bioinformática, los cuidados médicos, y las finanzas (George, Haas y Pentland, 2014). En este sentido, el análisis de *Big Data* busca generar una nueva percepción que permita completar sustancialmente los usos estadísticos tradicionales (como encuestas y otros métodos generalmente estáticos) con un método de explotación de datos mucho más dinámico, e incluso en tiempo real en determinadas ocasiones.

Estos autores establecen que el *Big Data* es un conjunto muy amplio de diferentes tipos de datos “granulares”, es decir, que se encuentran fraccionados en tantas pequeñas partes como sea posible para ser mejor definidos y tener una descripción lo más pormenorizada posible. Estos provienen de cinco fuentes claves: (i) datos públicos, que poseen gobiernos, organizaciones, e incluyen datos sobre transporte, uso energético, y atención médica, entre otros; (ii) datos privados, almacenados por empresas y organizaciones sin ánimo de lucro, de uso privativo ya que no se encuentran públicamente, incluyendo datos acerca de transacciones de los clientes, navegación por Internet, uso de teléfono móvil, entre muchas otras; (iii) “*data exhaust*” incluye aquellos datos resultantes de cualquier actividad digital u online, que se recogen pasivamente, tienen un valor limitado o nulo, y son por ello re combinados con otras fuentes de datos para crear nuevas fuentes de valor (interacción entre personas, compras en pequeños negocios o búsquedas de Internet, por ejemplo); (iv) “*community data*”, que se refiere a la conversión de datos no estructurados en auténticas redes dinámicas que recogen tendencias sociales (revisiones de clientes o contenidos de redes sociales, entre otros); y (v) “*self-quantification data*”, refiriéndose a tipos de datos que son revelados por el individuo a través de la cuantificación de sus acciones y comportamientos personales, como por ejemplo, los datos obtenidos por pulseras o relojes que monitorean el ejercicio y movimiento de un individuo que son posteriormente cargados y almacenados en aplicaciones móviles, lo que permite que puedan ser rastreados y agregados.

Por su parte, Purushothaman (2017) establece que existe un sinnúmero de posibilidades de uso de *Big Data*, especialmente para el sector hotelero, ya que en las empresas de esta industria se genera y almacena información de cada cliente desde que realiza una reserva online hasta que realiza su *check out*. Por tanto, podemos clasificar las posibles numerosas y variadas fuentes de recolección de datos de los huéspedes del hotel en cada etapa del “ciclo del huésped” que hemos definido en el apartado anterior, recogiéndolas en la **Tabla I**.

Tabla I: Fuentes de recolección de datos según etapas del ciclo del huésped

Etapa	Fuentes de recolección de datos en hoteles
Pre-Arribo	Página web, redes sociales (<i>data exhaust</i>), reservaciones, proveedores externos, acuerdos con socios / empresas de la industria
Arribo	<i>Check-in</i>
Estadía	Wifi, spa, <i>stands</i> y eventos, restaurantes, servicios de venta (<i>self-quantification data</i>)
Partida	<i>Check-out</i> , revisiones de clientes (<i>community data</i>)

Fuente: elaboración propia.

Podemos apreciar cómo estas fuentes pueden enmarcarse en todas las categorías definidas por George et al. (2014) que acabamos de ver – a excepción de los datos de naturaleza pública, lo cual es lógico debido a la naturaleza privada de este sector –, demostrándose así la extraordinaria riqueza que poseen los hoteles en cuanto a las fuentes que tienen a su disposición en su día a día para la recolección de datos de uso privativo de sus huéspedes.

2.3.1. Fuentes en la etapa del Pre-Arribo

En primer lugar, en la etapa del “Pre-Arribo”, podríamos distinguir como fuentes de recolección de datos, entre otras, la página web de la empresa, las redes sociales, las reservaciones de los huéspedes del hotel, los canales de distribución y los acuerdos de colaboración con proveedores de servicios externos (Purushothaman, 2017).

La página web de la empresa desde los primeros años del siglo XXI se ha convertido en una plataforma de interacción directa entre empresa y consumidor (Chen, Chiang y Storey, 2012). De esta forma, además de ofrecer la posibilidad de informar a sus consumidores acerca de sus productos y demás contenidos comerciales (promociones, horarios de apertura, formas de contacto, localización, entre muchas otras), las empresas han podido ir acumulando información detallada de sus clientes a través del uso de *cookies* y de los registros del servidor, constituyendo verdaderas minas de oro para los empresarios de forma a explotar información que les permita comprender las necesidades de sus clientes y detectar nuevas oportunidades de negocio (Chen, Chiang et al., 2012).

Las redes sociales constituyen un océano de información de utilidad para las empresas. Así, la mayoría de la información contenida en el universo digital es creada y consumida por los propios usuarios, el 68% en 2012 (Gantz y Reinsel, 2012), especialmente a través de sus interacciones a través de redes sociales. Sin embargo, debemos destacar que las empresas tienen dificultades de acceso y posible responsabilidad por la explotación de la mayor parte de esa información (cerca del 80%, en 2012, según estos autores), ya que deben enfrentarse a cuestiones de *copyright*, privacidad y cumplimiento de las regulaciones actuales en estas materias, a pesar de que dicha información sea generada directamente por los propios usuarios. Las empresas deben

utilizar dicha información para identificar tendencias actuales entre los usuarios, y entender con mayor precisión sus hábitos para, en última instancia, proporcionar una mayor satisfacción a las necesidades de sus clientes (Gantz y Reinsel, 2012).

Por su parte, el equipo de las reservaciones de hotel tiene como responsabilidad la captación de datos de los clientes a través de las reservas – tanto individuales como colectivas – que éstos realizan (generalmente de forme online), constituyendo registros de llamadas y de reservas online para poder definir un perfil del huésped que acude al hotel. De esta forma, incluso antes de su llegada al hotel, si se transmiten estos datos al equipo de la recepción del hotel adecuadamente, este podrá tener un trato más personalizado y ajustado a las necesidades de cada huésped.

Por último, además de las fuentes internas que acabamos de destacar, podemos incluir la adquisición de datos provenientes de proveedores de servicios especializados en la captación de datos de usuarios que, a través de acuerdos de colaboración, comparten con empresas que no disponen de los recursos necesarios para realizar esa tarea de captación de información. Por último, los acuerdos que las empresas de servicios de alojamiento realizan con socios de su grupo de empresas u otras empresas de su industria, de forma que puedan acceder a una mayor cantidad de datos de clientes representan una fuente adicional.

2.3.2. Fuentes en la etapa del Arribo

Pasando a analizar la etapa del “Arribo”, debemos destacar toda la información que se crea a través del *check-in* (método de pago y confirmación de las fechas de entrada y salida, entre otras), y que se deberá poner en común junto a los datos almacenados a partir de la creación y confirmación de la reserva del huésped. Esto le permitirá disponer al hotel de una mejor percepción del perfil del huésped, de tal forma que estará en mejor disposición de hacerle recomendaciones – turísticas, de ocio o de cualquier otra naturaleza – y de anticiparse a las demandas que le puedan surgir. Un ejemplo de ello sería si el huésped presenta alguna discapacidad o alergia alimenticia, ya que si el hotel cuenta con esa información antes del *check-in* del huésped, estará preparado para ajustar el trato a sus necesidades particulares, provocando una buena impresión, de trato personalizado al huésped.

2.3.3. Fuentes en la etapa de la Estadía

En cuanto a la etapa de la “Estadía”, podemos distinguir una gran variedad de fuentes de recolección de información. En primer lugar, el uso de red Wifi gratuito en los hoteles se encuentra, en numerosas ocasiones, condicionado a la previa inscripción del huésped a través de la inserción de unos determinados datos (dirección de correo electrónico, registro mediante perfil de Facebook, Twitter, Google...), lo cual puede constituir otra base de datos de utilidad para la empresa de alojamiento. En segundo lugar, destacamos todo lo relativo a la prestación de servicios – muchas veces, totalmente facultativos para el huésped –, como el servicio de spa, puestos y *stands* promocionales y eventos, de restaurantes o de servicios de venta. Todos estos servicios, además de potenciar la experiencia del huésped, pueden contribuir también a recoger información adicional del cliente, en cuanto a sus gustos y preferencias, así como su capacidad económica, entre otros.

2.3.4. Fuentes de datos en la etapa de la Partida

En lo relativo a esta última etapa, podemos establecer que pueden generarse datos adicionales del cliente, como pueden ser los datos creados en concepto de la satisfacción de las deudas generadas por el consumo de algún servicio de pago del hotel. Además, puede usarse el momento del *check-out* para tener un acercamiento individualizado al huésped, en base a la información que se haya podido recolectar en las fases anteriores, para ofrecerle una oferta personalizada, o simplemente para que la última experiencia que viva en el hotel sea lo más ajustada a su perfil como cliente. Por último, debemos destacar la relevancia que poseen las evaluaciones que realizan los huéspedes al haber finalizado su estancia (en páginas como *TripAdvisor* o *reviews* de la estancia en un alojamiento de Airbnb), así como a través directamente por el hotel.

2.3.5. Usos de las fuentes de datos

Toda esta información recogida por el hotel tiene varios usos potenciales: por un lado, debe servirle como guía individualizada del huésped, que estará a disposición de la empresa para (i) atraer el cliente mediante campañas de publicidad u ofertas individualizadas o agrupadas por segmentos de huéspedes, de cara a tratar de fidelizarlos y que regresen en un futuro; (ii) anticiparse a sus necesidades y ofrecerle una mejor experiencia desde la etapa del “Pre-Arribo”; y (iii) debe servirle para identificar oportunidades de negocio, adecuando las recomendaciones y servicios propuestos a las necesidades del huésped, lo que llevaría en última instancia a optimizar la experiencia del huésped.

Por tanto, podemos concluir que tal y como afirma Purushothaman (2017), las fuentes de datos que pueden estar a la disposición de las empresas de prestación de servicios de alojamiento son muy numerosas y diversas, las cuales representan un activo de gran importancia para éstas.

Establecida esta conclusión, debemos analizar a continuación las posibles técnicas de explotación de dichas bases de datos que ha constituido la empresa, con el propósito de determinar si el acceso a la explotación de grandes cantidades de información es accesible a todas las empresas del sector, o si, por el contrario, la adquisición y uso de las herramientas adecuadas para la correcta interpretación de las mismas requiere un gran esfuerzo tanto económico como tecnológico o de personal para las mismas.

2.4. Técnicas de explotación de *Big Data* en el sector empresarial

2.4.1. Presentación de las técnicas de explotación de *Big Data* más comunes: *Hadoop MapReduce* y sistemas *DBMS*

Según McAfee y Brynjolfsson (2012), las herramientas disponibles para lidiar con el volumen, velocidad y variedad del *Big Data* han aumentado considerablemente en los últimos años. Establecen los autores que, por lo general, estas tecnologías no son prohibitivamente costosas, y que la mayoría de los programas de *software* necesarios para explotar el *Big Data* son de fuente abierta. *Hadoop* es la herramienta más frecuentemente empleada, ya que combina un *hardware* básico con un *software* de fuente abierta, acogiendo flujos de datos recibidos y

distribuyéndolos en discos baratos, ofreciendo además herramientas para el análisis de los datos. Sin embargo, a pesar de la relativa facilidad que presenta esta herramienta frente a otras, estas tecnologías requieren un conjunto de habilidades que resultan novedosas para la mayoría de los departamentos de *IT*, por lo que éstos deberán esforzarse en integrar cuanto antes todas las fuentes de datos relevantes, tanto internas como externas.

Pavlo et al. (2009) establecieron en su estudio una comparativa entre la herramienta *Hadoop*, que ha implementado el modelo *MapReduce* (en adelante, “MR”), el cual es un modelo de programación para dar soporte a la computación paralela sobre cantidades voluminosas de datos, y los sistemas de gestión de bases de datos paralelas SQL (*parallel SQL database management systems*, en adelante, “DBMS”), ya que, si bien los DBMS han existido desde hace 20 años, ciertos autores han calificado MR como un drástico y nuevo modelo de computación. Pavlo et al. (2009) empiezan estableciendo que la característica más interesante de MR es que permite proporcionar un modelo sencillo a través del cual los usuarios pueden implementar programas de computación distribuida relativamente sofisticados, lo cual despertó un gran interés en la comunidad educativa. Por ejemplo, IBM y Google pusieron a disposición clústeres de procesadores *MapReduce* 1000 para enseñar programación distribuida.

En esta misma línea, establecen Dean y Ghemawat (2010) que MR ha sido ampliamente utilizado por las empresas, ya que ofrece la posibilidad de utilizar fácilmente los recursos presentes en un amplio sistema de distribución a programadores sin experiencia en el uso de sistemas paralelos o de distribución. Por tanto, a los programadores les resulta fácil el uso e implantación de este sistema, lo que ha provocado que, más de 100.000 trabajos de MR fueran ejecutados diariamente en los clústeres de Google en el año 2010.

Además, el modelo proporcionado por MR permite analizar datos almacenados en sistemas de almacenamiento heterogéneos – como bases de datos relacionales, que se suelen emplear para almacenar datos de clientes, así como sistemas de archivos, que suelen recoger las solicitudes realizadas por los usuarios –, aunque dichos datos migren a nuevos sistemas de almacenamiento. Sin embargo, los sistemas DBMS no presentan semejante facilidad para analizar los datos almacenados en bases de distinta naturaleza según Dean y Ghemawat, ya que los datos tienen que ser previamente copiados en el DBMS, y esta fase de aportación es inconveniente y costosa, siendo por un lado muy lenta, especialmente en el caso en que sólo vayan a ser analizados una o dos veces tras su aportación, y por el otro, la empresa debe poseer la herramienta de carga de datos apropiada.

Así, tal y como establecen Pavlo et al. (2009), existe un alto coste en lo relativo a la carga de datos si se emplean sistemas DBMS, ya que, para la mayoría de los indicadores de rendimiento que definen en su estudio – función básica y funciones analíticas de agregación, selección, relación y agregación UDF –, el tiempo requerido para cargar los datos en una base de datos paralela es de 5 a 50 veces mayor que el requerido para analizar los datos a través de *Hadoop*, lo cual representa una enorme diferencia en términos de coste de tiempo y de oportunidad. Así, a modo de ilustración, podríamos afirmar que pueden realizarse 50 análisis independientes sobre datos almacenados a través de MR antes de que sea posible cargarlos en una base de datos paralela y realizar un único análisis (Pavlo et al., 2009). Esta diferencia en el tiempo de carga de los datos puede no representar una carga muy gravosa para una empresa si tiene pensado realizar una gran cantidad de análisis distintos sobre los datos ya cargados, pero según Dean y Ghemawat, este supuesto no se da muy a menudo en las empresas, ya que lo más habitual es que busquen generar conjuntos de datos para ser procesados una o dos veces y ser descartados posteriormente.

2.4.2. Análisis de la “necesidad de negocio” como paso fundamental para decidir acerca de la implementación de *Hadoop*

Si bien es cierto que existen numerosas alternativas a *Hadoop* en el universo del *Big Data*, y que, de hecho, ya existían mucho antes de que esta herramienta revolucionara el panorama del análisis de cantidades extremadamente voluminosas de datos con su eficaz solución a bajo coste, aquéllas eran tremendamente costosas y por ello, solamente un grupo reducido de empresas las implementaba en su modelo de negocio. Por tanto, aunque aprender a usar *Hadoop* implica adentrarse de lleno en el mundo del *Big Data*, esto no quiere decir que sea la única herramienta que lo permita, o que represente la herramienta más adecuada para cada empresa. Por tanto, resulta difícil generalizar una respuesta que sirva para todas y cada una de las entidades que buscan iniciarse en el mundo del tratamiento de cantidades masivas de datos.

Es por ello por lo que, a la hora de analizar su mayor o menor adecuación para una empresa en concreto, además de contemplar las diversas virtudes que esta herramienta ofrece con respecto a sus alternativas (bajo coste, rico ecosistema, estructura de código abierto, tolerancia a fallos y fácil implementación, entre otras, como ya hemos visto), resulta determinante tener en cuenta otro factor: la necesidad de negocio (Power Data, 2015). Por esta noción entendemos la necesidad específica que posee una entidad y que desea solventar a través de la implementación de alguna herramienta de *Big Data*. Es decir, que la implementación de esta tecnología en la empresa debe responder a alguna funcionalidad previamente determinada por la propia empresa. Por tanto, en función de la naturaleza de dicha necesidad, será más adecuado el empleo de *Hadoop* o, por el contrario, el de algún otro sistema alternativo, en tanto en cuanto ya hemos visto que esta herramienta que, recordemos, explota el modelo de MapReduce, posee unas características concretas que la hacen diferenciarse de la competencia, características por las que podemos determinar frente a qué tipo de necesidades empresariales puede recomendarse su uso específico.

Por tanto, por su bajo coste y fácil implementación, esta herramienta resultará muy útil de implementar cuando la empresa busque iniciar un pequeño proyecto o primera experiencia de trato de *Big Data*, utilizando *Hadoop* como un repositorio centralizado de datos que vaya progresivamente aumentando su capacidad de procesamiento y reduzca el coste de almacenamiento con el tiempo (Power Data, 2015).

Por el contrario, si la intención de una empresa es la de llevar a cabo un proyecto específico, como puede ser una tarea que requiera controlar la seguridad, o alguna otra variable más técnica, esta herramienta no será la más apropiada pues, aunque ofrezca soluciones para ejecutar un gran número de tareas complejas, probablemente no resulte posible poder desempeñarlas correctamente para principiantes en esta materia, por lo que, al ser temas sensibles para la empresa, será más recomendable no arriesgarse a emplear esta tecnología precipitadamente. Otra debilidad que posee *Hadoop* reside en el trabajo en tiempo real, ya que, al partir de un ecosistema de datos tan heterogéneo, le resulta más difícil realizar un análisis en tiempo real que a otras herramientas – como *Apache Spark* – concebidas para resolver problemas de latencia, resolviendo de manera más exitosa esta necesidad (Power Data, 2015). Además, si bien es cierto que, en comparación con otras herramientas, *Hadoop* presenta un modo sencillo de utilización, debemos advertir de que, a pesar de esta condición, su implementación y posterior uso puede resultar complejo ya que exige poseer un mínimo de conocimientos matemáticos, estadísticos y técnicos. Por tanto, no puede presumirse que, desde su implementación, pueda explotarse todo el potencial de esta herramienta, sino que la mejor manera de ir familiarizándose con su funcionamiento es ponerla en acción e ir probando sus funcionalidades progresivamente (Power Data, 2015).

2.4.3. Conclusiones en cuanto a la necesidad de implementación de herramientas de explotación de *Big Data*

Antes de plantearse la adopción de *Hadoop*, toda empresa debe haberse planteado seriamente si realmente la aplicación de tecnologías de *Big Data* es una necesidad real y presente, ya que, de lo contrario, su implementación puede no ser recomendada. Además, aunque determine su necesidad de implementación de esta tecnología, debe estimar si (i) dicha inversión responde a una necesidad puntual y a corto plazo o (ii) responde a una necesidad que se proyecta en un amplio período temporal, estimando que será necesario ejecutar recurrentemente tareas a través del uso de la herramienta objeto de la inversión.

En el primer supuesto, la elección de *Hadoop* prácticamente se impone, debido a todas las características estudiadas. Sin embargo, en el segundo caso expuesto, nos encontraríamos ante una situación en la que podría ser recomendable acudir a otro servicio que, implicando un mayor coste empresarial en su implementación, permita acceder a funcionalidades exclusivas de dicha alternativa que estime imprescindibles para el análisis que quiera realizar a los grandes volúmenes de datos almacenados de sus clientes (Linke It, 2016). En este sentido, podemos destacar el servicio *SAP HANA*, cuyo rasgo diferenciador es el uso de su tecnología con rapidez e inmediatez, lo que se diferencia de la mayor dificultad de *Hadoop* para ejecutar tareas en tiempo real. Por tanto, una empresa que esté dispuesta a asumir un mayor coste preferirá acceder a servicios exclusivos de *SAP*, que no puedan ejecutarse con *Hadoop*, con el fin de realizar una inversión rentable a largo plazo.

En conclusión, hoy en día las empresas que apuestan por *Big Data* recurren mayoritariamente a *Hadoop* como marco de trabajo, debido a su reducido coste y a su facilidad de implementación, siendo *Big Data* prácticamente sinónimo de *Hadoop* (Power Data, 2015). Sin embargo, cada empresa deberá valorar qué sistema es el más adecuado para su caso específico según: (i) el tipo de necesidad que quiera satisfacer; (ii) el plazo de tiempo en el que se proyecte realizándola; (iii) las funcionalidades que aspira explotar por considerarlas imprescindibles para su negocio; y (iv) el coste en recursos económicos, temporales y humanos en los que desee incurrir.

La **Tabla II** que exponemos a continuación recoge las conclusiones a las que hemos llegado en cuanto a la elección de *Hadoop* frente a sistemas alternativos *DBMS* (como *Hana SAP*), en función de las variables que hemos definido:

Tabla II: Recomendaciones de elección entre *Hadoop* y DBMS según las necesidades de la empresa

	<i>Hadoop</i>	DBMS (SAP)
1. Necesidad...		
• De iniciación a BD	X	
• Específica/Técnica		X
2. Uso...		
• Inmediato o puntual	X	
• Previsión de uso reiterado a largo plazo		X
3. Funcionalidades...		
• Básicas	X	
• Avanzadas		X
4. Coste...		
• Gratis/Mínimo	X	
• Más elevado		X

Fuente: elaboración propia.

2.5. Aplicaciones de *Big Data* a nivel empresarial en el sector de alojamiento

Una vez estudiadas la diversidad de fuentes posibles de recolección de datos en una empresa perteneciente al sector de alojamiento, las distintas herramientas existentes para explotárlas, así como la viabilidad de su implementación, procede responder a una tercera pregunta: ¿para qué implementar *Big Data*? ¿Qué aplicaciones concretas puede tener en mi negocio?

2.5.1. Ventajas del uso de *Big Data* en el sector del turismo

Tal y como establecen Song y Liu (2017), existen varias ventajas que la implementación del *Big Data* aporta tanto a los consumidores como a los proveedores de productos turísticos. Así, esta tecnología ofrece la posibilidad de desarrollar un *marketing* personalizado y de proporcionar mejores servicios y productos destinados específicamente a determinados consumidores (Pries y

Dunnigan, 2015). Las ventajas que presenta el *Big Data* frente a metodologías tradicionales para la industria del turismo son las siguientes:

a) Fiabilidad: el *Big Data* se basa en acciones reales de los usuarios, no en encuestas. En otras palabras, acciones reales son analizadas en lugar de intenciones declaradas o respuestas a preguntas. Juntando todas las fuentes de información, puede afirmarse que el *Big Data* incrementa la base de la muestra en varias órdenes de magnitud con respecto a la que la investigación convencional tendía a basarse (Meeker y Hong, 2014). La fiabilidad del análisis de *Big Data* nos permite considerar todos los aspectos de la información con el fin de proporcionar resultados integrales en lugar de conclusiones sesgadas debido a la pérdida de información en la muestra de datos.

b) Nuevos flujos de información: el uso de *Big Data* en la industria del turismo incluye un tipo de información producida por los propios turistas, lo cual enriquece el conocimiento del mercado objetivo de las empresas turísticas y es muy útil para analizar la demanda de los consumidores de diferentes productos y servicios turísticos (Hendrik y Perdana, 2014). Dado que el *Big Data* incluye datos estructurados y reubicados, es posible establecer una referencia cruzada con otras fuentes de datos tales como redes sociales y datos de acceso público. Además, el análisis del *Big Data* puede ser contrastado con datos internos de cada empresa con el fin de determinar si el suministro de productos y/o servicios turísticos en cada área distinta de una ciudad están en sintonía con los turistas que los demandan.

c) Datos en tiempo real y a muy corto plazo: Uno de los usos innovadores del *Big Data* es el uso de datos para describir actividades en tiempo real antes de que las fuentes de datos oficiales estén abiertas al público (Bollier y Firestone, 2010). Varian (2014) argumentó que las consultas de búsquedas en tiempo real de Google son una buena forma de detectar en tiempo real actividades de los consumidores, ya que la correlación inmediata obtenida de los datos de correlación de Google posee una ventaja de seis semanas sobre los datos reportados.

2.5.2. Casos de aplicación exitosa de *Big Data* en el sector empresarial

Varios autores se han centrado en analizar los efectos que la incorporación del uso de esta tecnología está generando sobre el negocio de las empresas. En este sentido, podemos destacar el estudio llevado a cabo por el MIT Center for Digital Business y recogido por la *Harvard Business Review* (McAfee y Brynjolfsson, 2012), que, tras establecer que el fenómeno del *Big Data* – caracterizado por sus tres “V” – ha llegado para quedarse, aportó evidencias empíricas para convencer a los escépticos en cuanto a la importancia de la explotación de *Big Data* para casi cualquier negocio.

En primer lugar, interrogó a 330 empresas públicas de Norteamérica sobre sus prácticas gerenciales en materia organizacional y tecnológica. Tras analizar las respuestas en conjunto con otros datos de sus informes anuales y otras fuentes independientes, llegó a la conclusión de que, cuanto más puede caracterizarse una empresa en poseer un modelo de toma de decisiones basado en la explotación de datos – “*characterized as data-driven companies*” –, mejores resultados en indicadores objetivos financieros y organizativos obtienen. Concretamente, las empresas situadas en el tercio superior de cada sector en esta clasificación fueron de media un 5% más productivas y un 6% más rentables que sus competidores.

Para apoyar lo expuesto, aportan los autores dos casos en los que el uso de *Big Data* cambió drásticamente el rumbo de las empresas que decidieron ser pioneros en su implementación. El

primero es el caso de una importante aerolínea americana que contrató una empresa proveedora de servicios tecnológicos llamada PASSUR, que empleaba *Big Data* para realizar predicciones en cuanto a la hora de aterrizaje de los vuelos mucho más precisas que las previsiones antiguas, realizadas por los pilotos, de las que disponía la aerolínea, lo cual resultó en un ahorro en costes de millones de dólares. En segundo lugar, exponen el caso de Sears Holding, que adoptó prácticas de *Big Data* para sacarle mayor partido a los datos de los clientes de sus distintas marcas. Según el director tecnológico de la empresa, Phil Shelley, la adopción de esta tecnología permitió reducir el tiempo de generación de las promociones personalizadas de ocho semanas a solamente una, y afirmó que estaba muy sorprendido por la facilidad y el bajo coste por las que Sears realizó la transición entre los viejos sistemas de administración de datos de los que disponía y el implementado, coincidiendo estos comentarios con las características distintivas que hemos establecido de *Hadoop* a lo largo del apartado 2.4.

2.5.3. Posibles aplicaciones de *Big Data* en el sector de alojamiento

Pasando a analizar el sector en que se centra esta línea de investigación, debemos señalar que en la industria de prestación de servicios de alojamiento también se ha analizado recientemente el uso de *Big Data*. Marr (2016) ha apuntado que, en este sector, si bien el uso de *Big Data* es relativamente reciente, resulta determinante para identificar las características y necesidades heterogéneas de cada cliente, y poder cumplir con sus expectativas particulares para conseguir fidelizarlos, creando así una ventaja competitiva frente a la competencia. El autor también establece que el *Big Data* permite a los hoteles obtener información que les permite tomar decisiones en tiempo real, anticiparse a las necesidades de los clientes, y calcular el precio óptimo de las habitaciones, lo que les permite situarse en una posición privilegiada en el mercado.

En este sentido, ya desde hace más de una década, Magnini, Honeycutt y Hodge (2003) analizaron la importancia que tiene la extracción e interpretación de datos de clientes en el sector hotelero. Así, establecen los autores que, en esta industria, el conocimiento de los clientes – de dónde son, cuanto, cuando y en qué gastan su dinero – puede ayudar a los hoteles a implementar mejores estrategias de *marketing* y maximizar los beneficios. Por tanto, dicho proceso de extracción de datos permite identificar patrones de comportamiento significativos y construir modelos predictivos de comportamiento del cliente, clasificados por segmentos.

Por tanto, el uso de tecnologías de *Big Data* en empresas de este sector puede ofrecer a investigadores, gerentes y legisladores la evidencia basada en datos necesaria para tomar decisiones en base a números y análisis, más que en anécdotas, intuición o en su experiencia pasada (Frederiksen, 2012), lo que puede conducir a establecer análisis más precisos, una toma de decisiones más segura y mayores eficiencias operativas, y a una reducción de costes y de riesgos (De Mauro et al., 2015). En este sentido, las posibles aplicaciones de *Big Data* es una cuestión abundantemente tratada en la literatura científica, especialmente en el sector hotelero. Así, un gran número de autores (Waller y Fawcett, 2013; Bajari, Nekipelov, Ryan y Yang, 2015; Huet, 2015; Kahn y Liu, 2016; Marr, 2016; Purushothaman, 2017; Song y Liu, 2017; y Yang y Pan, 2017), cuyos estudios detallaremos a continuación, exponen posibles aplicaciones prácticas del uso de *Big Data* en cadenas hoteleras y otras plataformas de alojamiento como Airbnb, una vez establecidos los objetivos de la administración de datos, identificadas las fuentes de los datos y obtenido un conocimiento avanzado de los procesos y herramientas para su uso (Purushothaman, 2017).

Clasificaremos el listado de aplicaciones posibles de uso de *Big Data* en el sector de alojamiento según si se dan antes de la llegada del huésped al hotel (coincidiendo con la etapa del “Pre-arribo”, definida en el apartado 2.3.1), y aquellas cuando el huésped ya se encuentra disfrutando de su estadía en el hotel (etapas de “Arribo”, “Estadía”, y “Partida”), recogiénolas en la **Tabla III**.

Tabla III: Aplicaciones de *Big Data* en el sector de alojamiento

Etapa	Posibles aplicaciones de <i>Big Data</i> en el sector de alojamiento
Antes de iniciarse la estancia	<ul style="list-style-type: none"> - Políticas de <i>marketing</i> personalizadas (Marr, 2016); - Precios dinámicos y en tiempo real (Huet, 2015 y Marr, 2016); - Predicción ocupación hoteles (Yang y Pan, 2017); - Explotación de datos de redes sociales y revisión de comentarios de <i>feedback</i> de clientes (Song y Liu, 2017)
Una vez empezada la estancia	<ul style="list-style-type: none"> - Mejora satisfacción estancia huéspedes (Marr, 2016); - Mejor gestión del inventario (Waller y Fawcett, 2013); - Aumento eficiencia energética (Kahn y Liu, 2016)

Fuente: elaboración propia.

A) Aplicaciones antes del inicio de la estancia

- Establecer políticas de *marketing* más personalizadas: según Marr (2016), como en prácticamente todos los sectores, la mayoría del contenido analítico en la industria de prestación de servicios de alojamiento se centra en el *marketing*. El objetivo general es a menudo la realización de campañas personalizadas de *marketing* para los huéspedes basadas en datos y en forma de correo electrónico o de publicidad dirigida a través de las redes sociales. Esto implica tener que analizar toda la información disponible sobre los clientes que se están hospedando, mediante la recopilación de comentarios y evaluaciones de *feedback* de huéspedes que se alojaron previamente, de su actividad transaccional (como realizar el *check-in* o pedir comida en un restaurante del hotel), del uso de programas de fidelización de clientes, y de datos demográficos adquiridos a terceros. Este análisis se usa en última instancia para decidir si es probable de fidelizar un cliente mediante el ofrecimiento de una oferta que incluya una comida gratis en un restaurante, o una entrada para un espectáculo en un teatro cercano al hotel.

El autor expone esta aplicación a través del caso pionero de la cadena de hoteles americana Red Roof Inn, la cual, durante el invierno del año 2013-2014, en un periodo en el que la tasa de cancelación de vuelos era del 3% aproximadamente, lanzó una campaña de *marketing* dirigida específicamente a usuarios de teléfonos móviles en las áreas geográficas que probablemente se fueran a ver más afectadas durante aquel mes, imaginando que los pasajeros cuyo vuelo fuese cancelado buscarían alojamiento cercano a través de sus dispositivos móviles. El resultado fue un incremento de ingresos en un 10% en las áreas en las que se implementó la estrategia.

Sin embargo, tal y como establece el autor, si bien tanto la industria hotelera como la de prestación de servicios de alojamiento en general están empezando a utilizar *Big Data*, éstas tienen un gran volumen y variedad de datos con los que trabajar, resultando en un

gran número de aplicaciones más allá del *marketing* – como para los casos de las cadenas Marriott y Starwood, que veremos más adelante –, ya que los huéspedes dejan un rastro de datos desde el momento en que reservan su estancia hasta su *check-out*. Así, los analistas empiezan a familiarizarse con el análisis de esos datos para convertirlos en ideas procesables, resultando en estancias más satisfactorias para todos nosotros como clientes.

- Establecer precios dinámicos y en tiempo real: esta funcionalidad se ha convertido en una posibilidad real para las empresas del sector de alojamiento. Huet (2015) ilustra esta aplicación mediante el ejemplo de Airbnb. La compañía, que siempre ha dejado libertad a sus usuarios para determinar el precio al que ofrecen sus alojamientos, ha desarrollado un modelo llamado *Price Tips* que sugiere en tiempo real y para cada día del año la probabilidad de que un usuario que ofrece alojamiento obtenga una reserva según el precio que ha establecido, y además sugiere una franja de precios óptimos, combinando tecnologías de *Big Data* y *Machine Learning* – que podemos definir como “*un conjunto de métodos que pueden detectar automáticamente los patrones en los datos, y luego usar los patrones descubiertos para predecir los datos futuros, o realizar otros tipos de decisiones bajo incertidumbre*” (Robert, 2014) –. Algunas tendencias que emplea el modelo son intuitivas, como por ejemplo el hecho de que una ciudad hospede algún evento importante y otras no lo son tanto. Así, por un lado, el valor de los servicios que ofrece el alojamiento influye mucho en la demanda de este (Wi-Fi, aire acondicionado en climas cálidos), siendo el domingo el día menos probable de obtener una reserva, y el sábado el día más popular. Por el otro, el número de revisiones puede jugar un papel muy importante en la probabilidad de obtención de una reserva. Un alojamiento sin revisiones se considera incierto, pero tan solo una buena revisión hace que su reserva sea mucho más probable.

Además de estos factores, que Airbnb sabía cómo modelar, la compañía desarrolló *Aerosolve*, un paquete de *machine learning* de fuente abierta desde 2015, para buscar determinar más relaciones entre las características de los alojamientos listados y los precios que puedan exigir los usuarios. Con *Aerosolve*, Airbnb puede descubrir nuevos patrones que luego usa para comprender mejor qué hace que a un alojamiento se le atribuya un determinado precio. Con la implementación de *Aerosolve* y *Price Tips*, Airbnb pretende aumentar sus beneficios, buscando maximizar las reservas y evitar que el usuario – y, por ende, Airbnb – obtenga un precio inferior al que los clientes estarían dispuestos a pagar según este algoritmo.

Por su parte, Marr (2016) ilustra esta aplicación mediante los casos de Marriott y de Starwood Hotels and Resorts. El primero de estos ha usado la tecnología de *Big Data* en todas sus operaciones, mediante la explotación de conjuntos de datos no estructurados y semiestructurados, como informes meteorológicos y cronogramas de eventos locales, que se utilizan para pronosticar la demanda y determinar un valor para cada habitación durante todo el año, permitiendo a la cadena establecer precios con eficiencia óptima, lo cual es vital en una época en la que los clientes están acostumbrados a ahorrar céntimos escaneando los servicios de comparación de precios para obtener las mejores ofertas.

En cuanto a Starwood Hotels and Resorts, éste también posee un sistema que se basa en la optimización del precio de las habitaciones mediante el análisis de datos sobre factores económicos y eventos locales a escala mundial, así como informes meteorológicos. Saber cómo el clima hogareño de su sede central en América del Norte

afecta el precio que los clientes están dispuestos a pagar por una semana bajo el sol del Caribe les proporciona indicaciones sobre el mejor momento para reducir los precios o lanzar promociones de *marketing*. Esta estrategia le ha llevado a un aumento de casi el 5% en sus ingresos por habitación.

- Predecir los niveles de ocupación de los hoteles: este se ha convertido en es un aspecto clave en la gestión de los ingresos de los hoteles, ya que permite a los gerentes asignar los recursos más eficientemente, así como fijar precios estratégicamente. En su estudio, Yang y Pan (2017) analizan la cuestión de si el uso de una combinación de varias fuentes de *Big Data* – incluyendo consultas relacionadas con los motores de búsqueda, datos de tráfico del sitio web de la oficina local de turismo e información meteorológica detallada – aumenta considerablemente la precisión de las predicciones del nivel de ocupación semanal de huéspedes en la ciudad de Charleston, Carolina del Sur, ya que, si bien se había estudiado abundantemente la determinación de la predicción de la ocupación de huéspedes en el sector hotelero hasta la fecha, no se habían empleado para ello tecnologías de *Big Data*.

Para ello, los autores emplean el nivel medio de ocupación de hoteles en el área de Charleston usando cuatro variables externas: (i) el uso de *Google Correlate* para identificar las consultas más correlacionadas con la ocupación hotelera en el condado de Charleston; (ii) la recopilación de datos semanales acerca del tráfico del sitio web de la Oficina de Convenciones y Visitantes del Área de Charleston utilizando *Google Analytics*, con el fin de rastrear a los visitantes del sitio web; (iii) la creación de un *script* (documento que contiene un conjunto de códigos) personalizado en lenguaje R (un entorno y lenguaje de programación, con enfoque al análisis estadístico) para descargar información meteorológica de la página <http://www.forecast.io>; y (iv) la concurrencia de eventos o vacaciones que puedan impulsar o inhibir los volúmenes de visitantes. La conclusión a la que llegan los autores es que efectivamente, implementando fuentes de *Big Data*, se determinan predicciones más precisas, a través de la ejecución de varios test sobre dos modelos estadísticos – ARMAX y MSDR –. Sin embargo, también establecen limitaciones a la validez predictiva de este fenómeno, estableciendo que los futuros investigadores de esta línea incluyan fuentes más diversas para lograr predicciones aún mejores, como comentarios en redes sociales o datos de teléfonos móviles, que servirían como valiosos predictores.

- Explotación de datos de redes sociales y revisión de los comentarios de *feedback* de los clientes: Song y Liu (2017) establecen que la retroalimentación por parte de los huéspedes en la industria del turismo es clave en la búsqueda por identificar las preferencias del cliente y ofrecer experiencias de alojamiento positivas. Así, solicitar comentarios de los clientes es uno de los elementos más importantes para lograr un alto nivel de crecimiento de la compañía y construir una estrategia alrededor de una mejor satisfacción de las necesidades de los clientes. Un creciente conjunto de servicios en la nube nos da la capacidad presente y ubicua de interactuar usando teléfonos inteligentes, tabletas, o incluso relojes (Chen, Mao, Zhang et al., 2014).

El *machine learning*, definido anteriormente, es uno de los principales métodos técnicos utilizados en la industria del turismo para construir el mecanismo de retroalimentación entre los clientes y los operadores turísticos (Bajari et al., 2015). Por

ejemplo, a través de la cooperación entre los proveedores de servicios turísticos, las instituciones financieras y los operadores de telecomunicaciones, el *machine learning* puede identificar si una persona acaba de cambiar su residencia o si se ha desplazado al extranjero revisando sus cargos habituales.

Sobre la experiencia del cliente, el *machine learning*, junto con el *Big Data*, puede permitir que los negocios de viajes y operadores turísticos envíen de forma automática y proactiva mensajes de texto o establezcan llamadas con los clientes presentando nuevas ofertas tras adquirir algún producto o servicio. Sin embargo, el éxito del análisis predictivo depende tanto de la calidad del *Big Data* como de los mecanismos de retroalimentación del cliente, los cuales deben estar diseñados correctamente e integralmente para proporcionar datos procesables inmediatamente.

B) Aplicaciones una vez empezada la estancia del huésped

- Mejorar el reconocimiento y la experiencia de los huéspedes durante su estancia y ayudar a combinar mejor las habilidades de los empleados con las necesidades de éstos: Marr (2016), expone el caso de otra cadena americana de hoteles, Denihan Hospitality, reconocida por su uso innovador de la analítica. Denihan usó tecnología analítica de IBM para agrupar tanto datos transaccionales como de clientes de sus diferentes cadenas, y combinarlos con datos no estructurados como comentarios y revisiones de huéspedes proporcionando *feedback* de su estancia en el hotel en webs de calificación como *TripAdvisor*.

Tras evaluar los comentarios de *feedback* de los huéspedes y los datos transaccionales, la cadena adoptó decisiones estratégicas basadas en estos datos, para reorganizar muchas de sus habitaciones. Así, distinguió aquellos cuartos para viajeros de negocios de aquellos destinados a viajeros por ocio, proporcionó más espacio en el cuarto de baño de habitaciones reservadas para familias, y estableció una mayor gama de servicios, tales como cocinas equipadas, en las habitaciones donde sabían de antemano que los huéspedes apreciarían. Denihan llegó incluso a poner la analítica en manos del personal del hotel, que disponían de cuadros de mando en sus *smartphones*, permitiéndoles anticipar lo que un huésped en particular podría esperar de su estadía, como preferencias de comidas en restaurantes, servicios de conserjería o excursiones a lugares locales de interés. Además, el personal de limpieza recibe actualizaciones en tiempo real sobre necesidades de clientes, tales como si en una habitación en particular requieren una almohada adicional o si es probable que llamen al servicio de habitaciones para tomar un sándwich y un café a las 2 de la madrugada.

- Mejorar la gestión del inventario y de la logística: según Waller y Fawcett (2013), el *Big Data* tiene el potencial de revolucionar la dinámica de las cadenas de suministro. Establecen los autores que, debido al volumen, velocidad y variedad de este fenómeno, junto a la facilidad y bajo coste que posee, se pueden manejar mayores cantidades de datos a día de hoy. Existen más datos también debido a que los datos son captados más detalladamente. Por ejemplo, en lugar de registrar solamente la venta de una determinada unidad en una ubicación particular, también puede almacenarse el momento en que se vendió y la cantidad de inventario restante al tiempo de la transacción. Así, muchas empresas que no registraron sus ventas diarias por localización y por el nivel de existencias del inventario, ahora lo hacen gracias al *Big Data*. Además, como hemos visto, existe una proliferación de datos de opinión

del consumidor que pueden captarse a través de *Tweets*, *Likes*, y *reviews* de productos en páginas web. Establecen que por las tres grandes “Vs” del *Big Data*, los inventarios (i) se perpetúan en más localizaciones y en niveles más desagregados (volumen); (ii) pasan de actualizaciones mensuales a cada hora (velocidad); y (iii) se constituyen en almacenes, tiendas, así como tiendas y vendedores online (variedad).

Por otro lado, en cuanto a los usos potenciales de *Big Data* en el ámbito de la logística, Waller y Fawcett (2013) establecen posibles funcionalidades tales como la disponibilidad en tiempo real del nivel de capacidad logística del transportista, la mayor eficiencia y rapidez de respuesta de los fabricantes, y la mejora del nivel de precisión del sistema de inventarios, así como funcionalidades adicionales en el ámbito de la predicción – estimación de la hora de entrega, establecimiento respuestas en tiempo real con respecto a las opiniones de los clientes –, del transporte (trazar rutas óptimas de entrega según el clima, el nivel de tráfico y las características del conductor), y de los recursos humanos (seguimiento más efectivo de la productividad).

- Aumentar la eficiencia energética de la industria hotelera: tal y como establecen Kahn y Liu (2016), la implementación de *Big Data* puede ayudar a las cadenas hoteleras a establecer y analizar su propia eficiencia energética con el fin de encontrar un equilibrio entre la optimización de energía empleada y la calidad del servicio ofrecidos a los huéspedes. Por ejemplo, si los sistemas de ventilación y de aire acondicionado son reducidos en un día caluroso, el uso de electricidad será menor pero los huéspedes estarán menos cómodos. Por tanto, es importante que los hoteles equilibren sus beneficios privados con la eficiencia energética, hecho que podría ofrecer resultados tangibles si se implementan asociaciones fructíferas entre ingenieros y economistas, ya que, si bien los ingenieros son expertos en la eficiencia de los sistemas, los economistas producen una comprensión más precisa de cómo establecer sistemas de monitoreo para recompensar la eficiencia sin sacrificar la calidad de la estancia de los huéspedes.

Como podemos observar, se pueden distinguir una gran cantidad de aplicaciones que dan uso al *Big Data* y que afectan a todas las áreas de negocio de la industria de prestación de servicios de alojamiento. Es menester que las empresas de un sector adopten estas técnicas como parte del proceso de innovación del negocio (Manyika et al., 2011).

Sin embargo, la adopción de las mejores técnicas disponibles no siempre es generalizada en las empresas de un sector, siendo habitual que existan carencias en determinados ámbitos, como las limitaciones que sufren las empresas de menor tamaño o las economías con menor nivel de desarrollo económico, que pueden obedecer a una postura de oposición al cambio por parte del personal, falta de información, insuficiencia de los recursos financieros, humanos (Russom, 2011), o tecnológicos necesarios, entre otros (Bilgihan y Nejad, 2015).

3. Análisis empírico: aplicación de *Big Data* en el sector de alojamiento en la Unión Europea

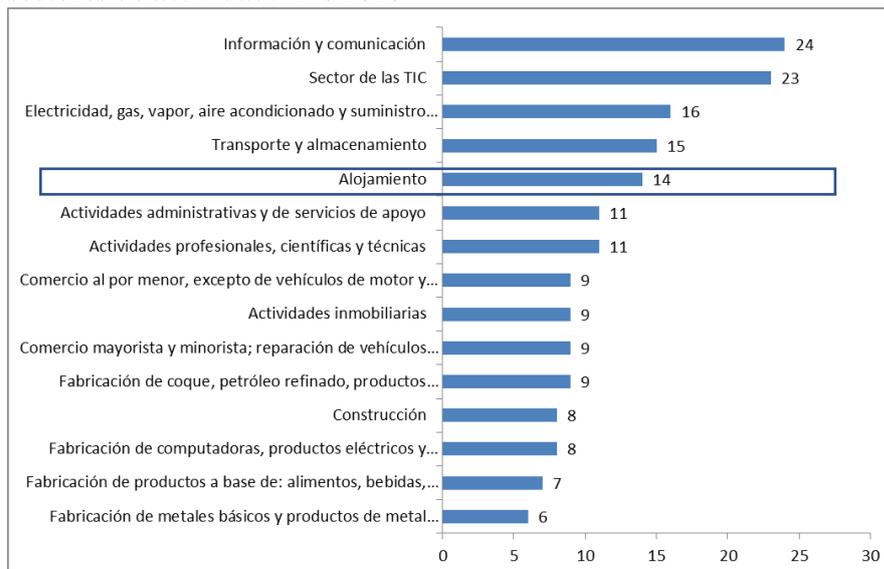
Una vez analizadas las posibilidades que ofrece la implementación de *Big Data* en las empresas del sector de alojamiento, procedemos a abordar el análisis empírico planteado. En este sentido, hemos visto al principio de este estudio (vid. Apartado 2.1) que el nivel de competitividad de las empresas se encuentra muy relacionado con su capacidad de innovación. Por tanto, habiendo establecido dicho marco teórico, procedemos a continuación a caracterizar el uso de *Big Data* en esta industria en la Unión Europea a partir de datos recogidos de Eurostat, y a analizar las implicaciones que puede tener el uso de esta tecnología en cuanto a la competitividad del sector de alojamiento.

3.1. Principales características del uso de *Big Data* en el sector de alojamiento en la Unión Europea

Como paso previo al análisis econométrico que permitirá contrastar las hipótesis planteadas, resulta relevante identificar las principales características del uso de *Big Data* en el sector de alojamiento en la Unión Europea.

En primer lugar, a través del **Gráfico IV** se puede observar que el promedio de empresas del sector de alojamiento que usan *Big Data* es del 14%, sobre el total, situándose entre los sectores que más han introducido esta técnica, sólo superado por el sector de Información y comunicación (24%), Tecnologías de la información (23%), Suministros de energía y agua (16%) y Transporte y almacenamiento (15%). Este hecho es acorde con lo analizado en epígrafes anteriores, que describían el potencial que tiene el uso de esta tecnología en este sector dada su fácil implementación y sus interesantes aplicaciones (vid. apartados 2.4 y 2.5, respectivamente).

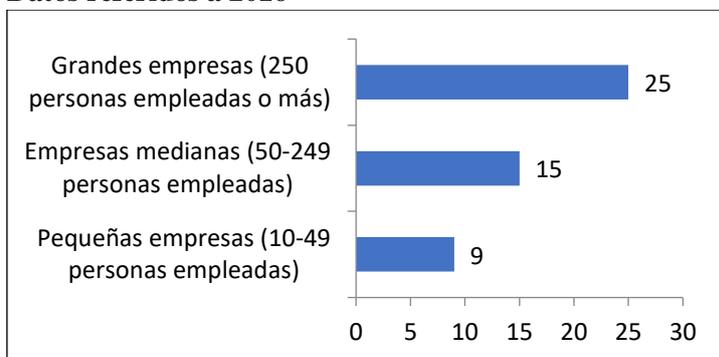
Gráfico IV: Porcentaje de empresas de la Unión Europea que usan *Big Data* según sectores de actividad. Año 2016



Fuente: *Digital Economy and Society: Big Data Analysis* (Eurostat).

En segundo lugar, el **Gráfico V** muestra que, en la Unión Europea, las grandes empresas (aquellas con 250 empleados o más), declaran usar técnicas de *Big Data* en una cantidad casi tres veces mayor que en las de menos de 50 empleados, pudiéndose deber este hecho al mayor número de barreras que poseen las empresas de menor tamaño frente a las de mayor capacidad económico-financiera y recursos tecnológicos.

Gráfico V: Uso de *Big Data* en la Unión Europea según tamaño de empresa (% empresas). Datos referidos a 2016



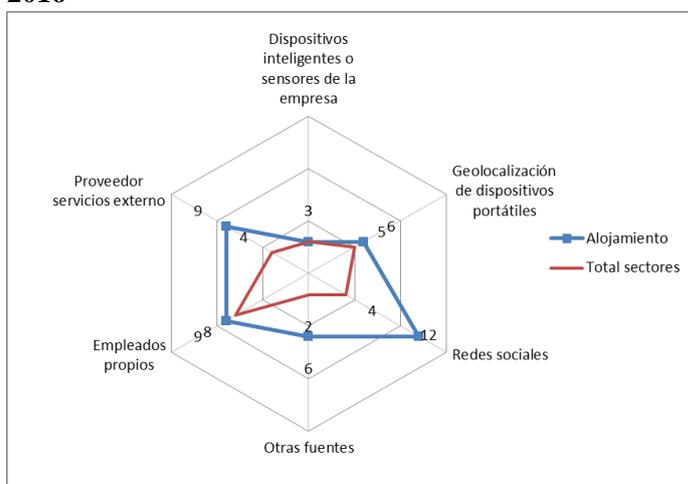
Fuente: *Digital Economy and Society: Big Data Analysis* (Eurostat).

Por último, en cuanto a las distintas fuentes de *Big Data* en el sector de alojamiento en la Unión Europea, el **Gráfico VI** muestra:

- que esta industria explota más intensivamente sobre todo las fuentes de proveedores externos (9% de las empresas del sector frente al 4% de promedio del conjunto de sectores); y
- que también hacer mayor uso de las técnicas de geolocalización de dispositivos portátiles, y triplica la explotación de redes sociales, en comparación con la media de los demás sectores.

Esta información confirma la extraordinaria diversidad y calidad de fuentes a disposición de este sector, como se analizó anteriormente (vid. apartado 2.3).

Gráfico VI: Fuentes de *Big Data* en el sector de alojamiento en la Unión Europea. Año 2016



Nota: se representa el porcentaje de empresas que utilizan cada tipo de fuente.

Fuente: *Digital Economy and Society: Big Data Analysis* (Eurostat).

3.2. Planteamiento de las hipótesis de investigación

Una vez establecidas las principales características del uso de *Big Data* en el sector de alojamiento en la Unión Europea, nos disponemos a contestar a las preguntas de investigación formuladas a partir del análisis econométrico de la base de datos creada.

Las hipótesis planteadas para su contrastación son:

1. “Existe un mayor uso de *Big Data* en las empresas pertenecientes a países con mayor PIB per cápita”. Esta relación procede de la menor actividad innovadora que en general presentan las economías con un menor nivel de desarrollo y de productividad, debido a la carencia ya sea de información, o de recursos humanos, tecnológicos o financieros necesarios. Para poder caracterizar mejor el uso de *Big Data* en el sector de alojamiento, realizaremos una clasificación agrupando países según sus características económicas y uso de *Big Data*.
2. “El uso de *Big Data* en empresas de alojamiento está asociado con mejores resultados económico-financieros, resultando en una mayor competitividad del sector de alojamiento”.

En el caso de confirmarse la primera hipótesis, cabe deducir que el crecimiento económico del sector de prestación de servicios de alojamiento de las economías menos desarrolladas se verá limitado en el futuro, lo que exigiría una actuación de las administraciones públicas europeas encaminada a promover su competitividad y capacidad innovadora, dentro del objetivo de alcanzar una mayor equidad en los niveles de renta y bienestar entre los Estados miembros, recogido en el Tratado de la Unión Europea y denominado el “principio de cohesión económica y social”.

Por su parte, la segunda hipótesis que planteamos, de confirmarse, deberá ser igualmente considerada dentro de las estrategias de promoción de la competitividad empresarial de las distintas Administraciones públicas implicadas en el desarrollo del sector turístico y especialmente en la industria de prestación de servicios de alojamiento. En este sentido, sería adecuado tratar de identificar las principales barreras existentes frente a una mayor adopción de las técnicas de *Big Data* en las distintas regiones y especialmente entre las empresas de menor tamaño, lo cual permitiría diseñar programas de información y formación de personal orientados a la innovación en el sector.

3.3. Fuentes, datos y variables empleadas

Debemos mencionar la escasez de estadísticas oficiales tanto a nivel nacional como europeo en materia de *Big Data*. Así, el Instituto Nacional de Estadística (INE) no incluye en su Encuesta sobre innovación en las empresas, de la que existen microdatos – y por tanto presenta un gran potencial para su uso en investigaciones –, un apartado sobre esta materia. Por su parte, en la Encuesta sobre el uso de TIC y comercio electrónico en las empresas, sí aparece información sobre el uso de *Big Data*, pero la carencia de microdatos impide su explotación para los fines de esta investigación. En cuanto a las estadísticas de Eurostat, sí que incluyen un apartado dedicado a esta materia concreta, pero cabe destacar que sólo se incluye en el año 2016. Es significativo y resulta hasta cierto punto extraño, que ni en años anteriores, ni en el último disponible, correspondiente de 2017 (publicado en enero de 2018), se contenga información sobre *Big Data*.

Por tanto, aunque no puede establecerse una serie histórica con datos de distintos años, sí existe información con capacidad para extraer conclusiones que den respuesta a las hipótesis de investigación. Así, a partir de tres estadísticas de Eurostat (Annual detailed enterprise statistics for services (NACE Rev. 2 H-N and S95), National Accounts 8ESA (2010) y Digital Economy and Society: *Big Data* Analysis, disponibles en <http://ec.europa.eu/eurostat/data/database>, se ha elaborado una base de datos de corte transversal homogénea, en la que se han incluido a las economías nacionales de más de 1 millón habitantes, lo que ha supuesto no incluir a Malta, Chipre y Luxemburgo, que presentan para algunas variables valores atípicos de difícil justificación. De esta forma, la base de datos contiene, para cada país de la Unión Europea, las siguientes variables, todas correspondientes al año 2016:

▪ **Como variables dependientes:**

- Porcentaje de empresas de más de 250 empleados que declaran utilizar técnicas de *Big Data* para cada país de la Unión Europea.
- Excedente bruto de explotación (EBE) respecto de la facturación en el sector de alojamiento, expresado en porcentaje.

▪ **Como variables independientes:**

- Porcentaje de empresas del sector de alojamiento que declaran utilizar técnicas de *Big Data*, para cada país de la Unión Europea.
- Producto Interior Bruto per cápita (PIB pc), expresado en euros.
- Tamaño empresarial, expresado como facturación por empresa y en miles de euros.
- Valor añadido bruto por empleado (VAB por empleado) en el sector de alojamiento, representativo del nivel de productividad, expresado en miles de euros.
- Costes de personal por empleado en el sector de alojamiento, expresado en miles de euros.
- Número total de visitantes, calculado como el número de llegadas de visitantes a establecimientos de alojamiento, para cada país de la Unión Europea.
- Visitantes per cápita, calculado como el número de llegadas de visitantes a establecimientos de alojamiento entre el número total de habitantes, para cada país de la Unión Europea.

3.4. Análisis descriptivo de las variables

Antes de proceder al desarrollo econométrico y a la interpretación de resultados, procederemos a exponer los valores, para la muestra definida, de todas las variables expuestas. La tabla que recoge los datos en términos absolutos se adjunta como **Anexo I**. En segundo lugar, hemos realizado la **Tabla IV**, que muestra los valores de las variables de cada país de la Unión Europea expresados, esta vez, como índice con respecto a la media europea, representando los valores un color (i) rojo, si su valor es inferior al 75% de la media europea; (ii) blanco, si se sitúa entre el 75% y el 125% de la media; o (iii) verde, si su valor es superior al 125% del promedio de la Unión Europea .

Tabla IV: Valores relativos de las variables definidas según países de la Unión Europea

Países UE	PIB pc	Big Data (+250 e.)	EBE sobre facturación	VAB por empleado	Facturación por empresa	Costes de personal por empleado	Uso Big Data sector alojamiento	Visitantes	Visitantes per cápita
Bélgica	128,9	168,0	98,0	161,6	125,6	155,5	n.d.	56,9	90,0
Bulgaria	23,4	92,0	107,1	25,1	31,0	21,8	28,6	25,7	64,4
Republica checa	57,4	88,0	79,1	56,5	33,2	55,5	28,6	70,8	120,1
Dinamarca	166,3	160,0	68,9	144,3	249,0	166,8	n.d.	21,0	66,0
Alemania	130,9	68,0	93,9	76,8	108,6	78,6	50,0	269,1	58,6
Estonia	55,0	132,0	68,9	45,1	66,7	51,4	85,7	15,6	212,5
Grecia	55,7	96,0	138,3	81,6	31,9	67,7	100,0	119,0	197,6
España	82,8	84,0	93,9	112,4	153,9	121,8	121,4	466,2	179,6
Francia	114,4	96,0	50,0	145,7	92,6	182,7	114,3	341,7	91,6
Croacia	38,1	96,0	143,9	80,8	115,9	64,5	171,4	104,2	444,9
Italia	95,2	120,0	98,0	137,3	84,7	131,4	92,9	431,3	127,2
Chipre	73,2	112,0	143,9	83,0	322,5	71,4	7,1	17,2	363,6
Lituania	46,4	80,0	115,3	41,6	15,5	36,4	128,6	11,5	71,5
Luxemburgo	311,7	136,0	97,4	142,2	198,4	153,6	92,9	8,0	248,7
Hungría	39,9	48,0	83,2	44,6	49,2	44,5	57,1	40,3	73,3
Malta	78,0	168,0	102,6	75,1	475,0	70,0	157,1	11,2	443,3
Países Bajos	141,9	172,0	102,0	104,1	136,4	96,4	157,1	120,3	126,8
Polonia	38,1	72,0	109,7	56,8	33,7	47,3	42,9	48,5	22,8
Portugal	61,5	96,0	103,1	73,0	35,9	70,5	150,0	93,8	162,3
Rumanía	29,6	72,0	108,7	27,8	36,8	23,6	114,3	18,8	17,0
Eslovenia	67,4	152,0	80,1	79,7	43,8	88,2	128,6	22,7	197,1
Eslovaquia	51,2	96,0	58,7	38,9	25,9	45,9	0,0	15,2	50,0
Finlandia	134,7	160,0	38,3	108,1	135,7	140,0	114,3	21,2	69,1
Suecia	161,2	120,0	51,0	132,2	144,1	170,5	92,9	49,8	90,4
Reino Unido	125,4	140,0	152,6	113,2	363,2	90,0	185,7	n.d.	n.d.
UE (28 países)	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Fuente: elaboración propia a partir de datos de Eurostat.

Como podemos ver a través de los índices expuestos, el primer rasgo significativo es las notables diferencias en renta per cápita entre los países de la Unión Europea. Esta misma característica se observa en buena parte de las variables analizadas relativas a la rentabilidad y productividad del sector de alojamiento, y el resto de indicadores.

Por otra parte, determinados países como Bélgica, Dinamarca, Luxemburgo, Finlandia o Suecia, a pesar de tener un índice de visitantes muy por debajo de la media, presentan en el resto de las variables índices por encima de ésta. En el caso de España, vemos que registra para todas las variables valores o bien en torno a la media, o bien en niveles superiores, destacando tanto el número de visitantes (466,2) y el de visitantes per cápita (179,6), como en la facturación por empresa (153,9). Otros países en los que todos los índices expuestos se sitúan en torno a la media europea o por encima de ésta son Italia, Países Bajos y Reino Unido. Finalmente, destacamos también que la gran mayoría de los países de Europa del Este (Bulgaria, República Checa, Hungría, Rumanía, Eslovaquia) registran valores generalmente inferiores a la media Europa.

3.5. Desarrollo econométrico y resultados

Una vez planteadas las hipótesis y definidas las variables objeto del presente análisis, procedemos a explotar la base de datos elaborada a partir de las estadísticas de Eurostat con el programa IBM SPSS Statistics (versión 19).

En primer lugar, como análisis preliminar, resulta interesante analizar las relaciones existentes entre el conjunto de variables definidas. Para ello, se calculará el coeficiente de correlación lineal de Pearson, que se muestra en la **Tabla V**.

Los resultados de mayor interés de este análisis, significativas para un nivel de significación del 5%, son:

- A) El nivel de PIB pc se encuentra correlacionado con el VAB por empleado en el sector de alojamiento, la facturación por empresa, los costes de personal por empleado y el uso total de *Big Data* en empresas de más de 250 empleados.
- B) A su vez, el uso de *Big Data* en empresas de más de 250 empleados está correlacionado con el PIB pc, el VAB por empleado, la facturación por empresa, los costes de personal por empleado y el uso de *Big Data* en el sector de alojamiento.
- C) El excedente bruto de explotación sobre facturación está correlacionado de forma negativa con los costes de personal por empleado.
- D) El VAB por empleado en el sector de alojamiento se relaciona positivamente con el PIB pc, el uso de *Big Data* en empresas de más de 250 empleados, la facturación por empresa, los costes de personal por empleado, el uso de *Big Data* en el sector de alojamiento y el número de visitantes en el país.
- E) Finalmente, la facturación por empresa en el sector de alojamiento, además de las anteriores relaciones expuestas, también está relacionada con el uso de *Big Data* en el sector de alojamiento y con los costes de personal por empleado.

Tabla V: Coeficientes de correlación lineal de Pearson entre las variables empleadas

		PIB pc	<i>Big Data</i> (+250 e.)	EBE sobre facturación	VAB por empleado	Facturación por empresa	Costes de personal por empleado	Uso <i>Big Data</i> sector alojamiento	Visitantes	Visitantes per cápita
PIB pc	C. Pearson	1	,645**	-,356	,822**	,704**	,843**	,304	,217	-,190
	Sig. (bilateral)		,001	,104	,000	,000	,000	,192	,345	,410
	N	22	22	22	22	22	22	20	21	21
<i>Big Data</i> (+250 e.)	C. Pearson	,645**	1	-,195	,600**	,512*	,560**	,460*	-,163	,112
	Sig. (bilateral)	,001		,385	,003	,015	,007	,041	,481	,629
	N	22	22	22	22	22	22	20	21	21
EBE sobre facturación	C. Pearson	-,356	-,195	1	-,191	,126	-,454*	,411	,046	,415
	Sig. (bilateral)	,104	,385		,394	,577	,034	,072	,841	,062
	N	22	22	22	22	22	22	20	21	21
VAB por empleado	C. Pearson	,822**	,600**	-,191	1	,621**	,952**	,460*	,459*	,073
	Sig. (bilateral)	,000	,003	,394		,002	,000	,041	,036	,752
	N	22	22	22	22	22	22	20	21	21
Facturación por empresa	C. Pearson	,704**	,512*	,126	,621**	1	,534*	,544*	,242	,085
	Sig. (bilateral)	,000	,015	,577	,002		,010	,013	,291	,715
	N	22	22	22	22	22	22	20	21	21
Costes de personal por empleado	C. Pearson	,843**	,560**	-,454*	,952**	,534*	1	,309	,399	-,028
	Sig. (bilateral)	,000	,007	,034	,000	,010		,185	,073	,905
	N	22	22	22	22	22	22	20	21	21
Uso <i>Big</i> <i>Data</i> sector alojamiento	C. Pearson	,304	,460*	,411	,460*	,544*	,309	1	,161	,552*
	Sig. (bilateral)	,192	,041	,072	,041	,013	,185		,512	,014
	N	20	20	20	20	20	20	20	19	19
Visitantes	C. Pearson	,217	-,163	,046	,459*	,242	,399	,161	1	,142
	Sig. (bilateral)	,345	,481	,841	,036	,291	,073	,512		,539
	N	21	21	21	21	21	21	19	21	21
Visitantes per cápita	C. Pearson	-,190	,112	,415	,073	,085	-,028	,552*	,142	1
	Sig. (bilateral)	,410	,629	,062	,752	,715	,905	,014	,539	
	N	21	21	21	21	21	21	19	21	21

** . La correlación es significativa al nivel 0,01 (bilateral).

* . La correlación es significativa al nivel 0,05 (bilateral).

Fuente: elaboración propia a partir de datos de Eurostat.

3.5.1. Relación entre uso de *Big Data* y PIB per cápita

La primera hipótesis de investigación es el mayor uso de *Big Data* en el ámbito empresarial en los países con mayor nivel de desarrollo económico. Para contrastarla, hemos estimado la siguiente ecuación (I) a través de un modelo de regresión lineal:

$$I) \quad \text{Big Data (+ 250 e.)} = B_0 + B_1(\text{PIB pc}) + B_2(\text{Facturación por empresa}) + u$$

Los estadísticos que se recogen en la **Tabla VI** confirman la relación positiva entre el uso de *Big Data* en la empresa y el PIB pc, con un R cuadrado del 0.42, siendo esta variable significativa para un nivel de significación del 5% ($t= 2,3$). El **Gráfico VII** ilustra la relación positiva entre las dos variables. Por el contrario, debemos rechazar la relación entre el tamaño medio de la empresa de cada país y el uso de esta técnica ($t=0,46$).

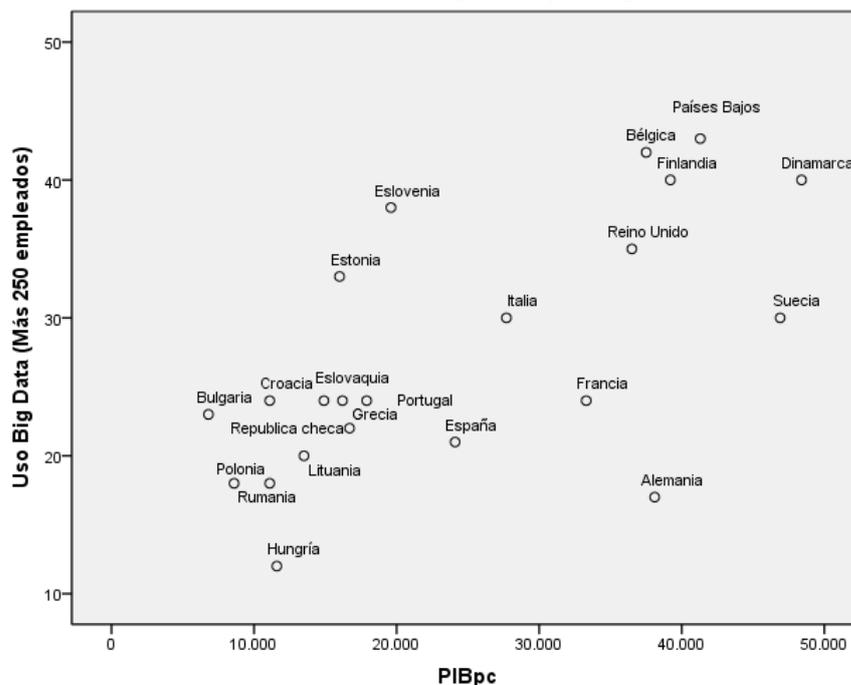
Tabla VI: Estadísticos ecuación uso de *Big Data* (+250 e.)

Parámetro	B	Error típ.	T	Sig.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Intersección	16,828	3,280	5,130	,000	9,962	23,694
PIB pc	,000	,000	2,304	,033	3,516E-5	,001
Facturación por empresa	,002	,005	,463	,649	-,008	,012

R cuadrado = ,423 (R cuadrado corregida = ,362)

Fuente: elaboración propia con datos de Eurostat.

Gráfico VII: Relación entre uso de *Big Data* y PIB pc en la UE



Fuente: elaboración propia con datos de Eurostat.

3.5.2. Relación entre uso de *Big Data* y rentabilidad empresarial del sector de alojamiento

La segunda hipótesis de contraste es que el nivel de rentabilidad en el sector de alojamiento, medido como el porcentaje de excedente bruto de explotación sobre la facturación, está relacionado con el uso de *Big Data*. Para contrastarla, y teniendo en cuenta las correlaciones bilaterales observadas, hemos estimado la siguiente ecuación (II) a través de un modelo de regresión lineal:

$$\text{II) EBE sobre facturación} = B_0 + B_1(\text{Big Data alojamiento}) + B_2(\text{Costes de personal por empleado}) + u$$

Los resultados que se muestran en la **Tabla VII** confirman la relación entre la rentabilidad y el uso de *Big Data* en el sector de alojamiento, controlando la influencia de los costes de personal por empleado. El R cuadrado de la ecuación es de 0.55, y ambas variables son significativas para un nivel de significación del 5%. En el **Gráfico VIII** y el **Gráfico IX** se ilustran las relaciones existentes entre las variables definidas en esta segunda ecuación.

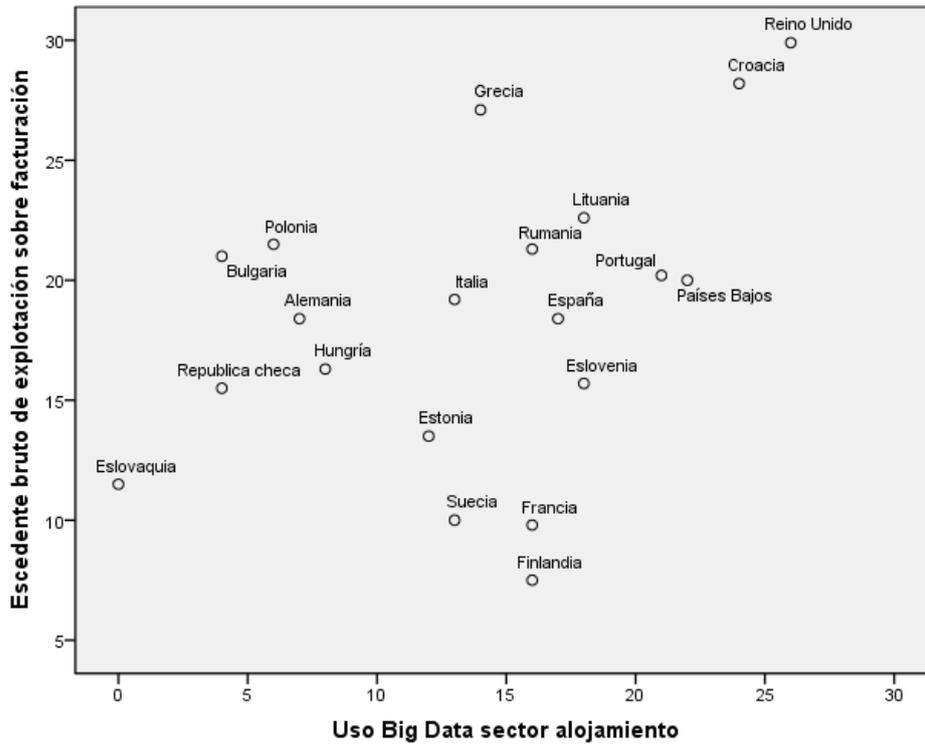
Tabla VII: Estadísticos ecuación rentabilidad sector de alojamiento

Parámetro	B	Error típ.	T	Sig.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Intersección	18,121	2,458	7,371	,000	12,935	23,308
Costes de personal por empleado	-,389	,102	-3,822	,001	-,604	-,174
Uso de <i>Big Data</i> en el sector de alojamiento	,526	,146	3,592	,002	,217	,835

R cuadrado = ,553 (R cuadrado corregida = ,500)

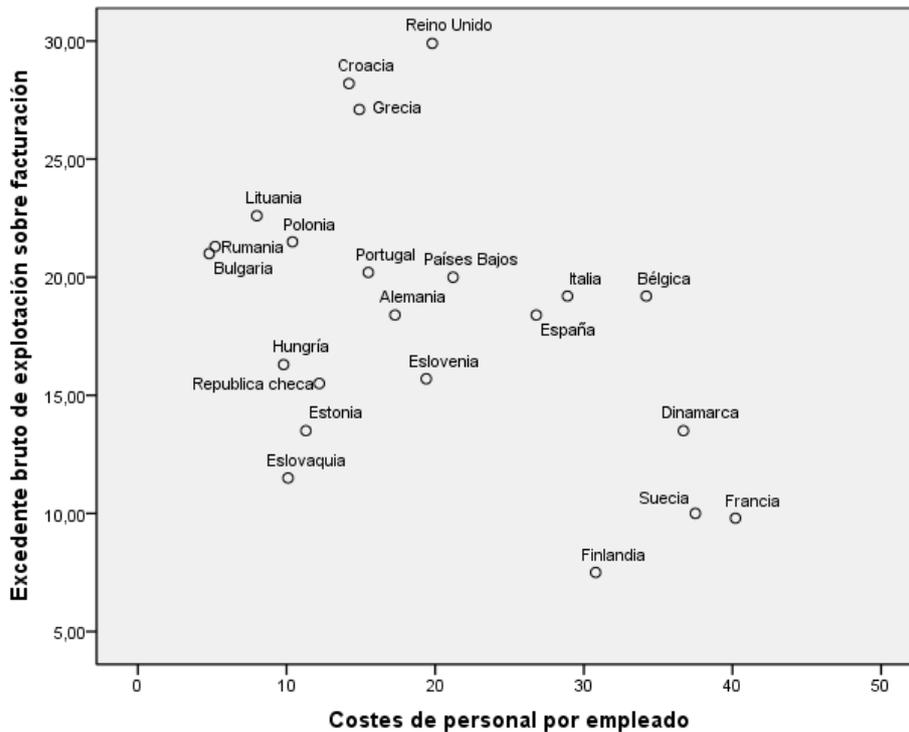
Fuente: elaboración propia con datos de Eurostat.

Gráfico VIII: Relación entre la rentabilidad y el uso de *Big Data* en el sector de alojamiento en la UE



Fuente: elaboración propia con datos de Eurostat.

Gráfico IX: Relación entre la rentabilidad y los costes de personal por empleado en el sector de alojamiento en la UE



Fuente: elaboración propia con datos de Eurostat.

4. Conclusiones y discusión a partir de los resultados obtenidos

El éxito económico y empresarial está cada vez más condicionado por la capacidad de innovar, quedando los factores productivos tradicionales (capital, climatología, recursos naturales, personal de baja cualificación, etc.) en posiciones de menor relevancia. En el actual contexto de avance tecnológico, la obtención y explotación de *Big Data* se vislumbra como un elemento de vital importancia para la obtención y preservación de ventajas competitivas de las empresas.

A pesar del gran potencial que presenta, en la actualidad la información estadística relativa al uso de *Big Data* es escasa, lo que limita su estudio en trabajos de investigación en ciencias sociales. A través de la información contenida en Eurostat, que incluye datos relativos a los países de la Unión Europea, se ha podido constatar de forma inequívoca que el uso de *Big Data* está más desarrollado entre las empresas de mayor tamaño, y los países con mayor nivel de renta per cápita. A su vez, su uso en el sector de alojamiento se asocia con una mayor rentabilidad empresarial. Esta conclusión, no obstante, debe tomarse con precaución, ya que como el uso de *Big Data* está asociado a las empresas de mayor tamaño, no es viable determinar con los datos disponibles la influencia individual que cada una ejerce sobre la rentabilidad.

Las implicaciones que tienen estos resultados para la política regional europea son relevantes, ya que la búsqueda de una reducción de las disparidades regionales entre los países miembros de la Unión Europea exigiría promocionar la actividad innovadora, incluyendo el uso de *Big Data* entre los países con menor nivel de renta per cápita, si se desean reducir las disparidades económicas actualmente existentes.

Por otra parte, se ha constatado que el uso de *Big Data* es deficiente en las empresas de menor tamaño, lo que las sitúa en una situación de desventaja con respecto a grupos con mayor capacidad para la implementación de técnicas innovadoras. La literatura revisada muestra la existencia de soluciones de *Big Data* interesantes de bajo coste, por lo que quizás existan problemas de falta de información o de medios humanos en las empresas de alojamiento de menor tamaño.

En este sentido, los programas de mejora de competitividad de Pequeñas y medianas empresas (Pymes) deberían evaluar las barreras que actualmente existen en este segmento empresarial ante la mayor utilización de este tipo de técnicas, como paso previo para diseñar programas de mejora de su competitividad. Dado que España presenta un tejido productivo caracterizado por la predominancia de pequeñas empresas, esta conclusión resulta especialmente relevante en nuestro país.

5. Bibliografía

- BAJARI, P., NEKIPELOV, D., RYAN, S. P., y YANG, M. (2015). Machine learning methods for demand estimation. *American Economic Review*, 105(5), 481–485.
- BEYER, M., y LANEY, D. (2012). The Importance of “Big Data”: A Definition. *Gartner Publications*, 1-9.
- BILGIHAN, A., y NEJAD, M. (2015). Innovation in hospitality and tourism industries. *Journal of Hospitality and Tourism Technology*, 6(3).
- BOLLIER, D., y FIRESTONE, C. M. (2010). *The promise and peril of big data*. Washington, DC: Aspen Institute.
- CAMARGO-VEGA, J., CAMARGO-ORTEGA, J., y JOYANES-AGUILAR, L. (2015). Conociendo Big Data. *Facultad de Ingeniería*, 24(38), 63-77.
- CHEN, H., CHIANG, R., y STOREY, V. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.
- CHEN, M., MAO, S., y LIU, Y. (2014). Big Data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.
- CHEN, M., MAO, S., ZHANG, Y., y LEUNG, V. (2014). *Big data: Related technologies, challenges and future prospects*. Heidelberg: Springer.
- DE MAURO, A., GRECO, M., y GRIMALDI, M. (2015). What is Big Data? A consensual definition and a review of key research topics. *AIP Conference Proceedings*, 1644(1), 97-104.
- DEAN, J., y GHEMAWAT, S. (2010). MapReduce: A Flexible Data Processing Tool. *Comm. ACM*, 53(1), 72-77.
- FOSSO WAMBA, S., AKTER, S., EDWARDS, A., CHOPIN, G., y GNANZOU, D. (2015). How ‘Big Data’ Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study. *International Journal of Production Economics*, 165, 234-246.
- FREDERIKSEN, L. (2012). Big Data. *Public Services Quarterly*, 8(4), 345-349.
- GANTZ, J., y REINSEL, D. (2012). The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. *IDC iView: IDC Anal. Future*, 2007, 1-16.
- GEORGE, G., HAAS, M., y PENTLAND, A. (2014). Big Data and Management: From the Editors. *Academy of Management Journal*, 57(2), 321-326.
- HENDRIK, H., y PERDANA, D. H. F. (2014). Trip guidance: A linked data based mobile tourists guide. *Advanced Science Letters*, 20(1), 75–79.
- HUET, E. (2015). How Airbnb Uses Big Data And Machine Learning To Guide Hosts To The Perfect Price. *Forbes*, 6/2015.
- KAHN, M., y LIU, P. (2016). Utilizing “Big Data” to Improve the Hotel Sector’s Energy Efficiency. *Cornell Hospitality Quarterly*, 57(2), 202-210.
- KAO, J. (2009). Tapping the World’s Innovation Hot Spots. *Harvard Business Review*, 109-14.
- KASAVANA, M., y BROOKS, R. (2001). *Managing Front Office Operations*. Michigan: Educational Institute of the American Hotel & Lodging Association.

- KUSNETZKY, D. (2010). What is "Big Data"? *ZDNet*. Extraído el 2 de enero de 2018 desde <http://www.zdnet.com/article/what-is-big-data/>
- LANEY, D. (2001). 3D data management: Controlling data volume, velocity and variety. *Meta Group*.
- LINKE IT (2016). SAP HANA vs Hadoop. Extraído el 23 de marzo de 2018 desde <http://blog.linkeit.com/sap-hana-vs-hadoop>
- LÓPEZ, D. (2013). Análisis de las posibilidades de uso de Big Data en las organizaciones. *Universidad de Cantabria*.
- LYNN, M. (2013). Big Data: management must interpret and apply data to gain the competitive advantage. *HVS Global Hospitality Report*, 1-4.
- MAGNINI, V., HONEYCUTT, E., y HODGE, S. (2003). Data Mining for Hotel Firms: Use and Limitations. *Cornell Hotel and Restaurant Administration Quarterly* 2003, 44, 94-105.
- MANYIKA, J., CHUI, M., BROWN, B., BUGHIN, J., DOBBS, R., ROXBURGH, C. y BYERS, A. (2011). Big data: The next frontier for innovation, competition, and productivity, *McKinsey Global Institute*.
- MARR, M. (2016). How Big Data And Analytics Are Changing Hotels And The Hospitality Industry. *Forbes*, 01/2016.
- MATÉ JIMÉNEZ, C. (2014). Big Data. Un nuevo paradigma de análisis de datos. *Anales de mecánica y electricidad*, 10-16.
- MAYER-SCHÖNBERGER, V., y CUKIER, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. USA: Houghton Mifflin Harcourt.
- MCAFEE, A., y BRYNJOLFSSON, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, 10/2012.
- MEEKER, W. Q., y HONG, Y. (2014). Reliability meets big data; Opportunities and challenges. *Quality Engineering*, 26(1), 102-116.
- OECD (2005). *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*. Organization for Economic Co-operation and Development.
- PAVLO, A., PAULSON, E., RASIN, A., ABADI, D., DEWITT, D., MADDEN, S., y STONEBRAKER, M. (2009). A comparison of Approaches to Large-Scale Data Analysis. *Proceedings of the 2009 ACM SIGMOD International Conference*.
- POWER DATA (2015). ¿Para iniciarme en Big Data necesito Hadoop? Extraído el 23 de marzo de 2018 en <https://blog.powerdata.es>
- PRIES, K., y DUNNIGAN, R. (2015). *Big Data analytics: A practical guide for managers*. Boca Raton, FL: CRC Press.
- PURUSHOTHAMAN, V. (2017). How to get started with Big Data in your hotel. *Hotelier*. Extraído el 2 de enero de 2018 desde <https://ehotelier.com/insights/2017/05/12/big-data-in-your-hotel/>
- REDMOND, E., y WILSON, J. (2012). *Seven Databases in Seven Weeks*. USA: O'Reilly Media, Inc., Pragmatic Programmers, LLC.
- ROBERT, C. (2014). Machine Learning, a Probabilistic Perspective. *CHANCE*, 27(2), 62-63.

- RUSSOM, P. (2011). Big Data Analytics. *TDWI Best Practices Report, Fourth Quarter 2011*.
- SCHWAB, K., y SALA-I-MARTÍN, X. (2018). *Global Competitiveness Report 2017-2018*. World Economic Forum.
- SONG, H. & LIU, H. (2017). Predicting Tourist Demand Using Big Data. *Analytics in Smart Tourism Design: Concepts and Methods. Switzerland: Springer*.
- VARIAN, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-27.
- WALLER, M., y FAWCETT, S. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77-84
- YANG, Y., y PAN, B. (2017). Forecasting Destination Weekly Hotel Occupancy with Big Data. *Journal of Travel Research*, 56(7), 957-970.

6. Anexos

Anexo I: Tabla de valores brutos de las variables definidas

País UE	PIB pc (€)	Big Data +250 e. (%)	EBE sobre facturación (%)	VAB por empleado (miles de €)	Facturación por empresa (miles de €)	Costes de personal por empleado (miles de €)	Uso Big Data sector alojamiento (%)	Visitantes	Visitantes p.c.
Bélgica	37.500	42	19,2	59,8	697,0	34,2	n.d.	7.481.422,0	0,66
Bulgaria	6.800	23	21	9,3	172,0	4,8	4,0	3.386.932,0	0,47
Republica checa	16.700	22	15,5	20,9	184,0	12,2	4,0	9.321.440,0	0,88
Dinamarca	48.400	40	13,5	53,4	1.382,0	36,7	n.d.	2.767.651,0	0,48
Alemania	38.100	17	18,4	28,4	603	17,3	7	35.404.924	0,43
Estonia	16.000	33	13,5	16,7	370	11,3	12	2.056.526	1,56
Grecia	16.200	24	27,1	30,2	177	14,9	14	15.664.091	1,45
España	24.100	21	18,4	41,6	854	26,8	17	61.341.839	1,32
Francia	33.300	24	9,8	53,9	514	40,2	16	44.958.968	0,67
Croacia	11.100	24	28,2	29,9	643	14,2	24	13.707.444	3,27
Italia	27.700	30	19,2	50,8	470	28,9	13	56.752.647	0,94
Chipre	21.300	28	28,2	30,7	1.790	15,7	1	2.268.176	2,67
Lituania	13.500	20	22,6	15,4	86	8	18	1.517.890	0,53
Luxemburgo	90.700	34	19,1	52,6	1.101	33,8	13	1.053.653	1,83
Hungría	11.600	12	16,3	16,5	273	9,8	8	5.301.843	0,54
Malta	22.700	42	20,1	27,8	2.636	15,4	22	1.468.143	3,26
Países Bajos	41.300	43	20	38,5	757	21,2	22	15.829.488	0,93
Polonia	11.100	18	21,5	21	187	10,4	6	6.378.793	0,17
Portugal	17.900	24	20,2	27	199	15,5	21	12.343.982	1,19
Rumania	8.600	18	21,3	10,3	204	5,2	16	2.471.411	0,13
Eslovenia	19.600	38	15,7	29,5	243	19,4	18	2.990.794	1,45
Eslovaquia	14.900	24	11,5	14,4	144	10,1	0	1.995.812	0,37
Finlandia	39.200	40	7,5	40	753	30,8	16	2.788.903	0,51
Suecia	46.900	30	10	48,9	800	37,5	13	6.550.337	0,66
Reino Unido	36.500	35	29,9	41,9	2.016	19,8	26	n.d.	n.d.
Unión Europea (28 países)	29.100	25	19,6	37	555,0	22	14,0	13.158.462,9	0,74

Fuente: elaboración propia a partir de datos de Eurostat