



UNIVERSIDAD PONTIFICIA COMILLAS  
FACULTAD DE CIENCIAS ECONÓMICAS Y  
EMPRESARIALES

**TRABAJO DE FIN DE GRADO**  
**BIG DATA EN VALORACIÓN**

**Autor**

Jerónimo Ybarra Román

**Director**

Leandro Sergio Escobar Torres

## **AGRADECIMIENTOS**

*Quisiera expresar mi agradecimiento especial a mi supervisor, Leandro Sergio Escobar Torres, por su guía, absoluta disponibilidad y apoyo a lo largo de casi un año.*

*Gracias a su dirección he podido aventurarme a descubrir el mundo del Big Data y aprender nuevas herramientas y técnicas de valoración. Su supervisión me ha ayudado a superar muchos de los obstáculos que he encontrado por el camino.*

*Asimismo, quisiera agradecer a las instituciones de Lancaster University y Universidad Pontificia de Comillas, por haberme enseñado el valor del aprendizaje y del trabajo bien hecho. Cualidades, que me han permitido llevar a cabo este proyecto.*

*" Acknowledge all of your small victories.*

*They will eventually add up to something great."*

*Kara Goucher*

*10 de junio de 2022*

*Jerónimo Ybarra Román*

A handwritten signature in blue ink, appearing to read 'Jerónimo Ybarra Román', with a large, stylized flourish extending from the end of the name.

*Trabajo de Fin de Grado presentado a la facultad de Ciencias Económicas y Empresariales de la Universidad Pontificia Comillas en cumplimiento parcial de los requisitos para el Grado en Administración de Empresas con Mención Internacional (E4).*

## ***RESUMEN EJECUTIVO***

El presente trabajo de fin de grado trata de mostrar la posible aplicación del Big Data para conseguir valoraciones más precisas. A lo largo del mismo se indaga sobre el fenómeno del Big Data las diversas técnicas de valoración de empresas e inmuebles y el sector inmobiliario y sus fuentes de datos. Gracias a los conocimientos adquiridos durante este análisis de literatura se hace posible la realización de un caso que ponga en práctica la unión de los conceptos. De esta manera, el trabajo culmina con un ejercicio de valoración de los activos madrileños de la Inmobiliaria Colonial SOCIMI usando Big Data.

Los resultados del caso muestran ser prometedores y en base a ellos se teoriza que la aplicación del Big Data a mayor escala y con recursos más sofisticados podría llevar a valoraciones más precisas, tanto inmobiliarias como empresariales.

El fenómeno de los macrodatos, dada su popularidad, se ha tachado por muchos de sobrevalorado. Vistas las diversas aplicaciones productivas del mismo en combinación con ejercicios como el realizado en este trabajo, se aporta validez y se solidifica su posición como herramienta útil.

***PALABRAS CLAVE:*** Big Data, valoración, sector inmobiliario, raspado web, Colonial.

## ***EXECUTIVE SUMMARY***

This dissertation analyses the applications of Big Data, more specifically, focusing on how Big Data Analytics can improve valuation accuracy. Throughout the research paper, the phenomenon of Big Data is investigated, as well as the various techniques for valuing companies and real estate, and the real estate sector and its data sources. As a result of the knowledge acquired during this literature review, it is possible to combine these concepts in a case study. In this way, the paper culminates with a valuation exercise of the Madrid assets of the Inmobiliaria Colonial SOCIMI using Big Data.

The results presented, show to be promising and based on them it is theorized that the application of Big Data on a larger scale and with more sophisticated resources could indeed lead to more precise valuations, both in real estate and in business.

The phenomenon of big data, given its popularity, has been put into question by many and labeled as overrated. Given the various productive applications of Big Data shown, in addition to exercises such as the one carried out in this work, the validity of this field is proven and its position as a useful tool is solidified.

***KEYWORDS:*** Big Data, valuation, real estate, web scraping, Colonial.

## ÍNDICE DE CONTENIDOS

<b>AGRADECIMIENTOS</b> .....	<b>II</b>
<b>RESUMEN EJECUTIVO</b> .....	<b>III</b>
<b>ÍNDICE DE TABLAS</b> .....	<b>VI</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>VI</b>
<b>ÍNDICE DE ABREVIATURAS</b> .....	<b>VII</b>
<b>1. INTRODUCCIÓN</b> .....	<b>1</b>
1.1. PROPÓSITO GENERAL DE LA INVESTIGACIÓN.....	1
1.2. CONTEXTUALIZACIÓN DEL TEMA .....	1
1.3. JUSTIFICACIÓN DEL TEMA .....	2
1.4. OBJETIVOS .....	3
1.5. METODOLOGÍA.....	4
<b>2. BIG DATA</b> .....	<b>5</b>
2.1. HISTORIA .....	5
2.1.1. Agrupación y uso de datos .....	5
2.1.2. Primeros usos del término “Big Data” fuera del contexto actual.....	6
2.1.3. Primeros usos del término “Big Data” en el contexto actual .....	6
2.1.4. Transcurso y evolución .....	7
2.2. DEFINICIÓN .....	8
2.2.1. Dimensiones del Big Data.....	8
2.3. APLICACIONES DEL BIG DATA .....	10
2.4. CADENA DE VALOR Y TÉCNICAS.....	12
2.5. PRIVACIDAD Y REGULACIÓN.....	13
2.5.1. Pero ¿Cuánto se sabe sobre nosotros realmente? .....	13
2.5.2. ¿Como nos protegemos? .....	14
2.6. BIG DATA EN EL FUTURO.....	15
<b>3. VALORACIÓN</b> .....	<b>17</b>
3.1. LA IMPORTANCIA DEL VALOR .....	17
3.2. MÉTODOS DE VALORACIÓN .....	18
3.2.1. Métodos de valoración de empresas .....	18
3.2.2. Métodos de valoración de inmuebles .....	19
3.3. BIG DATA EN VALORACIÓN.....	20
<b>4. BIG DATA EN REAL ESTATE</b> .....	<b>21</b>
4.1. WEB SCRAPING .....	21
4.2. OBTENCIÓN DE DATOS INMOBILIARIOS .....	22
4.2.1. Big Data de Internet.....	23
4.2.2. Big Data de Sensores .....	24

4.2.3. <i>Big Data de IoT</i> .....	25
<b>5. CASO PRACTICO: VALORACIÓN INMOBILIARIA CON BIG DATA</b> .....	<b>27</b>
5.1. COLONIAL HISTORIA, DATOS Y ACTIVOS .....	27
5.2. OBTENCIÓN DE DATOS .....	29
5.2.1. <i>Datos sucios</i> .....	30
5.2.2. <i>Limpieza de Datos</i> .....	31
5.4. VALORACIÓN DE LOS ACTIVOS.....	33
5.4.1. <i>Valoración Intrínseca</i> .....	33
5.4.2. <i>Valoración Relativa</i> .....	36
5.5. CONCLUSIÓN.....	38
5.5.1. <i>Limitaciones y mejoras</i> .....	38
<b>ANEXO</b> .....	<b>39</b>
<b>BIBLIOGRAFÍA</b> .....	<b>47</b>

## ***ÍNDICE DE TABLAS***

Tabla 1. Las 10 Dimensiones de Big Data.....	9
Tabla 2. Principales metodos de valoración de empresas.....	18
Tabla 3. Tipos y fuentes de datos inmobiliarios .....	22
Tabla 4. Datos de Inmobiliaria Colonial.....	27
Tabla 5. Muestra de datos sucios .....	30
Tabla 6. Reducido datos sucios extracción precio alquiler.....	31
Tabla 7. Tabla reducida de precios de alquiler con datos limpios .....	32
Tabla 8. Tabla de alquiler medio para código postal 28004 .....	34
Tabla 9. Valor activo nº1 .....	35
Tabla 10. Valor total activos en € (método intrínseco).....	35
Tabla 11. Valor total activos en € (método relativo) .....	37

## ***ÍNDICE DE FIGURAS***

Figura 1. Cadena de Valor de Big Data.....	12
Figura 2. Porcentaje de acierto de análisis de “me gusta” (Facebook) por categorías .....	13
Figura 3. Visualización de Características de Activos de Colonial .....	27
Figura 4. Mapa del universo de 41 oficinas de Colonial en Madrid (escala 2km/cm) .....	28
Figura 5. Mapa del universo de 41 oficinas de Colonial en Madrid (escala 1km/cm) .....	28
Figura 6. Interfaz ParseHub (extracción precio alquileres de Belbex) .....	29
Figura 7. Proceso de Limpieza de Datos .....	31
Figura 8. Valor de mercado de activos de Madrid según Colonial.....	35
Figura 9. Método de selección de comparables .....	36

## *ÍNDICE DE ABREVIATURAS*

<b>AI</b>	Artificial Intelligence – Inteligencia Artificial
<b>CBD</b>	Central Business District
<b>DCF</b>	Discounted Cash Flow – Descuento de Flujos de Caja
<b>EBITDA</b>	Earnings Before Interest, Taxes, Depreciation, and Amortization.
<b>EV</b>	Enterprise Value
<b>FCF</b>	Free Cash Flow
<b>GAV</b>	Gross Asset Value
<b>IoT</b>	Internet of Things – Internet de las Cosas
<b>IT</b>	Information Technology – Tecnología de la Información
<b>LTV</b>	Loan To Value
<b>P/E</b>	Price/Earnings
<b>POI</b>	Points Of Interest – Puntos De Interés
<b>RCA</b>	Real Capital Analytics – Análisis de Capital Real
<b>RGPD</b>	General Data Protection Regulation – Reglamento General de Protección de Datos
<b>SABI</b>	Sistema de Análisis de Balances Ibéricos
<b>SOCIMI</b>	Sociedad Cotizada de Inversión Inmobiliaria
<b>WACC</b>	Weighted Average Cost of Capital
<b>WWW</b>	World Wide Web

# 1. INTRODUCCIÓN

## 1.1. Propósito general de la investigación

El propósito general del presente trabajo consiste en la investigación de las posibles aplicaciones del Big Data a la valoración, acompañado de una aplicación práctica de la teoría presentada a través de un caso de valoración usando los enfoques de valoración de mercado y de rentas de los activos de Madrid de la Inmobiliaria Colonial SOCIMI.

## 1.2. Contextualización del tema

La metodología que se debe usar para alcanzar una valoración lo más cerca del valor objetivo<sup>1</sup> posible, ha sido un tema debatido por académicos durante generaciones. A lo largo de los años se han desarrollado métodos (1) intrínsecos, como el DCF, que buscan alcanzar un valor no afectado por condiciones externas y; (2) extrínsecos, como el enfoque de mercado, que miran más allá del objeto valorado buscando incluir el valor percibido por el mercado en la valoración. Los métodos de valoración son muy diversos, esto los hace complementarios al usarse de manera conjunta, puesto que cada uno aporta ventajas y desventajas distintas que se cubren entre ellas.

Sin embargo, sigue existiendo una disparidad entre valoraciones y valores objetivos dada la dificultad de predecir condiciones futuras. Esto nos lleva a la cuestión de si el prometedor fenómeno del Big Data, notorio por utilizar datos del pasado para hacer predicciones del futuro, sería capaz aplicarse a la valoración. En finanzas, el pionero de las técnicas cuantitativas Jim Simons ya era capaz en 1990 de usar modelos informáticos que buscaran patrones históricos para identificar tendencias en el mercado (Zuckerman, 2019). De hecho, desde 1988 hasta 2018, el fondo estrella de Simons llamado “Medallion fund” generó un rendimiento anual promedio del 66 %, algo que ningún otro inversor, Warren Buffett, George Soros o Peter Lynch tiene en su historial (Zuckerman, 2019).

Siguiendo el ejemplo de Simons, este trabajo navega la literatura académica precedente sobre el Big Data y el análisis del Big Data en búsqueda de los avances que puedan aplicarse a la valoración. El Big Data, dados sus prometedores avances ha visto un crecimiento exponencial en popularidad que ha ido de la mano con el crecimiento de las tecnologías de red. Según una encuesta por QuinStreet, el 77 % de las empresas considera el Big Data como una prioridad (Arnaboldi, Busco, & Cuganesan, 2017). Esto ha actuado

---

<sup>1</sup> El valor objetivo es aquel que reside en la cosa o en la persona, y no en la percepción de valor individual, de modo que es absoluto, universal y cuantificable (Caballer, 1976).



como una espada de doble filo ya que, aunque por un lado ha permitido el avance de las tecnologías de análisis de Big Data, por otro lado; ha visto un creciente número de críticas poniendo en duda la validez y utilidad real de estos métodos.

A través del caso práctico de la valoración de los activos de Madrid de la Inmobiliaria Colonial SOCIMI, se trabaja de manera conjunta la validez del uso del Big Data y su aplicación en la búsqueda de valoraciones más precisas.

### 1.3. Justificación del tema

El fenómeno del Big Data no es solo relevante en el contexto empresarial, sino que nos afecta directamente en nuestro día a día. Resulta curioso, interesante y a la vez preocupante pensar en la magnitud de datos personales que forma nuestra huella digital. Esta se extiende desde nuestro historial de búsqueda hasta nuestros patrones de comportamiento registrados por redes sociales que muestran un conocimiento sobre nuestros intereses y maneras de actuar que ni nosotros entendemos en tanto detalle.

Cada vez se realiza un mayor porcentaje de nuestra vida en la red, factor el cual la pandemia mundial causada por el COVID 19 ha acelerado, y que muestra ir a más con el desarrollo de entornos web simulados en 3d como el metaverso. En consecuencia, el volumen de datos generado y el posible valor extraído de los mismo continuará aumentando exponencialmente. Es por lo tanto clave, tener un entendimiento de los usos y aplicaciones de los datos que generamos y generaremos.

De forma paralela, mi interés en finanzas me lleva a interesarme en el campo de la valoración y la complejidad de encontrar el valor objetivo de un activo. Pero, más allá de la curiosidad que me genera el hecho de que se pueda obtener el precio objetivo de un activo de manera certera; reconozco la importancia de la precisión necesaria en valoración dado el aspecto real de inversión que lo precede. Una valoración bien hecha es clave no solo para beneficiarse invirtiendo con mayor certidumbre, pero asegurando el funcionamiento de los mecanismos de mercado de manera justa e imparcial.

Por último, el sector inmobiliario al ser la categoría industrial más grande de mundo representa un mar de datos disponibles para la colección y uso. De esta manera, encaja a la perfección con el tema de Big Data en valoración, permitiendo realizar una colección de datos masivos para su posterior uso en una valoración que ponga a prueba la eficiencia de las técnicas de Big Data en este ambiente.

## 1.4. Objetivos

El presente trabajo se centrará en alcanzar los siguientes objetivos:

### **Big Data**

- Realizar un estudio literario del concepto del Big Data, indagando en su historia y origen con el fin de encontrar los aspectos comunes de las dispersas aportaciones académicas para alcanzar una definición más estandarizada.
- Explorar las aplicaciones, cadena de valor y técnicas del Big data en diferentes ámbitos y disciplinas para confirmar su validez como herramienta.
- Explorar el aspecto relacionado a la privacidad y como mantenerla a través de la regulación, frente a un mundo dominado por empresas que tienen como objetivo extraer el mayor número de datos sobre usuarios posibles.
- Proyectar a futuro el camino que le queda por recorrer a los macrodatos y analizar sus posibles aplicaciones en el tiempo.

### **Valoración**

- Profundizar en la importancia de la valoración y en los diversos métodos de valoración de empresas e inmuebles.
- Indagar en los aspectos en los que el uso del Big Data se podría aplicar para perfeccionar los modelos de valoración.

### **Big Data en Sector Inmobiliario**

- Realizar un análisis de la literatura existente en búsqueda de aplicaciones de Big Data a la valoración inmobiliaria con el fin de enfocar la metodología hacia el caso práctico.
- Examinar las técnicas de extracción web mostrando su funcionamiento, utilidad y aplicación.

### **Caso Práctico**

- Aplicar la teoría desarrollada a lo largo del trabajo.
- Creación de una herramienta de web scraping para la colección de datos.
- Limpieza de extracciones de datos y preparación de muestras de Big Data para análisis.
- Valoración inmobiliaria de los activos de Madrid de la Inmobiliaria Colonial usando la muestra de macrodatos obtenida y análisis de resultados.
- Conclusión validando o invalidando el uso de Big Data en valoración.

## 1.5. Metodología

La metodología para la realización del trabajo sigue el siguiente proceso:

Primero, se utilizará mayoritariamente el buscador académico *Google Scholar* en combinación con la herramienta web de información general y de cuentas anuales SABI además de recursos aportados por el tutor académico, para la obtención de materiales literarios en los que basar las investigaciones sobre:

1. El fenómeno del Big Data
2. La importancia del valor y las metodologías de valoración
3. El sector inmobiliario y sus tipos y fuentes de Big Data

En segundo lugar, se encuentra la realización de la valoración de activos de Madrid de Colonial. Para su realización se utilizará:

1. La aplicación ParseHub para la creación de una herramienta de raspado de datos web.
2. Las plataformas intermedias de venta y alquiler de inmuebles Belbex e Idealista como fuentes de datos del raspado.
3. El programa Microsoft Excel como herramienta de análisis de datos en la que se realizará la valoración. En este se realizarán dos modelos diferentes:
  - a. Modelo de valoración intrínseca que utilizará datos de alquileres y un valor de rentabilidad o 'yield' como factor de descuento.
  - b. Modelo de valoración relativa que utilizará datos de inmuebles en venta comparables para suponer cuanto debería valer un m<sup>2</sup> del activo valorado.

En tercer y último lugar se expondrán los resultados y analizarán en base al juicio y saber hacer desarrollado a lo largo del trabajo para concluir si el Big Data tiene o no tiene futuro en el perfeccionamiento de las valoraciones.

## 2. BIG DATA

El Big Data o “macrodatos” en castellano, es un fenómeno del que muchos hablan y pocos entienden. Su definición ha sido fruto de discusiones contantes dada la naturaleza cambiante del asunto que se intenta definir. La evolución tecnológica; los nuevos usos, descubrimientos y aplicaciones de los datos; y la amplitud del tema, han generado una complicación a la hora de consolidarlo bajo una sola definición. Además, cabe resaltar la dificultad de alcanzar a una definición común dada la popularidad del tema.

### 2.1. Historia

Para captar la esencia del Big Data y llegar a una definición más precisa, es importante entender su origen e historia. Durante este apartado, se hará hincapié en las aportaciones de referentes en la materia como Laney (2001), Diebold (2012) o Visconti (2020) reflejadas en la literatura académica.

#### 2.1.1. Agrupación y uso de datos

El Big data tiene un énfasis en la consolidación de datos y su análisis con el fin de transformar datos en información. Por ello algunos académicos han relacionado el comienzo del Big Data con los primeros usos de la agrupación y análisis de datos.

En primer lugar, encontramos las ideas de Gil Press (2013) que proporcionó una breve historia del Big Data. Basándose en los estudios de Rider (1944), cubrió 68 años de historia en la evolución de Big Data entre 1944 y 2012, ilustrando 32 eventos con relación al Big Data en la ciencia de datos (Wu, Buyya, & Ramamohana, 2016). Press argumenta que hay una línea fina entre “crecimiento de datos” y “Big Data”, la cual se ha vuelto cada vez más borrosa en los últimos años. La tasa de crecimiento de datos se ha denominado “explosión de información”, pero los términos “información” y “datos”, no deberían de utilizarse de manera intercambiable (Press, 2013) (Wu, Buyya, & Ramamohana, 2016).

A diferencia de Press, Frank Ohlhorst (2012), atribuye el comienzo del Big Data a 1880 cuando se realizó el décimo censo de Estados Unidos. El problema en este momento fue uno de estadística, relacionado con como encuestar y documentar a los 50 millones de ciudadanos estadounidenses (Ohlhorst, 2012) (Wu, Buyya, & Ramamohana, 2016). Vemos que Ohlhorst enfoca el principio del Big Data hacia la primera concentración de datos que el considera de gran tamaño.

El que se aventura más atrás en el tiempo es Bernard Marr (2015) quien rastrea el origen del Big Data al año 18.000 A.C., remontándose a los primeros ejemplos que tenemos de humanos almacenando y analizando datos. El Ishango Bone fue descubierto en 1960 en la actual Uganda y se cree que es una de las primeras pruebas de almacenamiento de datos prehistóricos. Marr, argumenta que debemos prestar atención a los diferentes enfoques en los que los humanos capturan, almacenan, analizan y recuperan datos e información (Marr, 2017) (Wu, Buyya, & Ramamohana, 2016).

### *2.1.2. Primeros usos del término “Big Data” fuera del contexto actual*

Las primeras apariciones de las palabras “Big” y “Data” de manera juntas no significa que estas estuviesen siendo utilizadas para referirse a lo que ahora significan. A pesar de ello es interesante indagar en el pasado buscando los verdaderos primeros usos de “Big” y “Data” juntos. Francis X. Diebold (2012) quien le da importancia justamente a esto, hace un estudio en el que indaga tanto en el aspecto académico como en el no académico en busca de las primeras menciones del término.

En el aspecto académico, encuentra menciones por Tilly (1984) en un artículo que explora si el análisis estadístico genera valor para historiadores. En el aspecto no académico encuentra menciones en: (1) un programa de ordenador de 1987 llamado “small code, big data”, (2) un artículo del Washington Post de 1989 por Eric Larson titulado “They’re Making a List: Data Companies and the Pigeonholing of America,” y (3) un comunicado de PR Newswire de 1996 que menciona la tecnología de red "para aplicaciones de agrupamiento de CPU y Big Data..." (Diebold, 2012). Todas estas fuentes, a pesar de fascinantes, no usan el término “Big Data” en el contexto actual y demuestran no tener conciencia del mismo.

### *2.1.3. Primeros usos del término “Big Data” en el contexto actual*

Steve Lohr Periodista de tecnología, economía y fuerza laboral para el New York Times y galardonado con el Premio Pulitzer en 2013, explica que "El término Big Data es tan genérico que la búsqueda de su origen no es solo un esfuerzo por encontrar una referencia temprana a esas dos palabras que se usan juntas" (Wu, Buyya, & Ramamohana, 2016) y argumenta que simplemente adoptar el término por sí solo podría no tener la connotación de Big Data de hoy, es decir, no solo una gran cantidad de datos, sino diferentes tipos de datos manejados de nuevas maneras (Wu, Buyya, & Ramamohana, 2016). Siguiendo esta lógica, a continuación, se exponen los primeros usos de “Big Data” en el contexto actual.

El término Big Data, junto con la conciencia del fenómeno Big Data, se mostraba claramente en Silicon Graphics, Inc. (SGI) fabricante estadounidense de estaciones de trabajo informáticas, superordenadores y software de gráficos con sede en Mountain View, California (Hall, 2022) a mediados de la década de 1990 (Diebold, 2012). A pesar de haber sido de manera no académica, John Mashey, exjefe científico jubilado de SGI, produjo una serie de diapositivas de SGI de 1998 titulada "Big Data y la próxima ola de InfraStress", además, SGI publicó un anuncio que presentaba el término Big Data en Black Enterprise (Diebold, 2012).

Brevemente después de Mashey, en el campo académico de la comunidad informática, Weiss e Indurkha (1998) señalan que "... colecciones muy grandes de datos... ahora se compilan en almacenes de datos centralizados, lo que permite a los analistas hacer uso de métodos poderosos para examinar los datos de manera más completa. En teoría, el 'Big Data' pueden llevar a conclusiones mucho más sólidas para las aplicaciones de minería de datos, pero en la práctica surgen muchas dificultades" (Diebold, 2012).

El término "Big Data", que abarca la informática y la estadística/econometría, probablemente se originó en conversaciones en Silicon Graphics Inc. (SGI) en las que John Mashey ocupó un lugar destacado. Mientras que en el campo académico son Weiss e Indurkha (1998) en informática y Diebold (2000) en estadística/econometría quienes introducen el término los primeros (Diebold, 2012).

Cabe destacar que unos años más tarde en 2001 una nota de investigación inédita de Douglas Laney en Gartner enriqueció significativamente el concepto e introdujo las famosas 3V de big data (Diebold, 2012).

#### *2.1.4. Transcurso y evolución*

Desde las primeras agrupaciones de datos hasta hoy, la mayoría del desarrollo de macrodatos se ha realizado desde la invención del ordenador y más específicamente desde la creación de la World Wide Web (WWW). Avances tecnológicos en dispositivos móviles, sensores digitales, canales de comunicación telemáticos, poder informático y almacenamiento de datos han proporcionado los medios para la recopilación de datos a nivel masivo (Yaqoob, y otros, 2016) (Bryant, Katz, & Lazowska, 2008). Como producto de ello, compañías como Google, Yahoo o Facebook han podido analizar estos grandes volúmenes de datos para beneficiar sus negocios. Desde entonces, con el desarrollo y progreso de aplicaciones y tecnologías, se ha ido dando forma al término Big Data el cual se ha ido adaptando hasta lo que hoy en día conocemos (Yaqoob, y otros, 2016).

## 2.2. Definición

Como vemos reflejado en la historia, el origen y la definición de “Big Data” son un tanto dispersos. Sin embargo, vemos que por lo general todos los académicos, están de acuerdo con que “Big Data” significa una gran cantidad de información la cual requiere métodos especiales para ser procesada. De Mauro y otros (2016), matizan bien estos conceptos en su definición; según ellos, “Big Data es un activo de información tan extenso en volumen, velocidad y variedad que requiere tecnología específica y métodos analíticos para su transformación en valor” (De Mauro, Greco, & Grimaldi, 2016) (Narayan & Tan, 2019).

Asimismo, es importante entender que Big Data representa la interrelación de datos potencialmente provenientes de fuentes heterogéneas, por lo que no solo incluye **datos estructurados**, como bases de datos, sino también **datos no estructurados**, como imágenes, correos electrónicos, datos de GPS, información tomada de redes sociales, etc. (Snijders, Matzat, & Reips, 2012) (Moro Visconti, 2020).

### 2.2.1. Dimensiones del Big Data

Para llegar a una definición más precisa que englobe todos los aspectos del Big Data, es imprescindible indagar en las “dimensiones” del Big Data. Inicialmente presentadas por **Douglas Laney** en Gartner como **las 3V’s**: Volumen, Velocidad y Variedad. Laney (2001) argumenta que el comercio electrónico, en particular, había disparado los desafíos de gestión de datos a lo largo de estas tres dimensiones (Laney, 2001).

Sobre el modelo de Laney, se han hecho varias adaptaciones:

- **IBM** decide hacer una ampliación y añade una nueva V, la Veracidad, este nuevo modelo es conocido como **las 4V’s** de Big Data. IBM implementa esta dimensión para incluir la calidad y origen de los datos (Wu, Buyya, & Ramamohana, 2016).
- **Yuri Demchenko** (2013) hace una ampliación a las **5V’s**, añadiendo la Veracidad de IBM y una nueva V, el Valor (Wu, Buyya, & Ramamohana, 2016).
- **Microsoft**, en aras de maximizar el valor comercial convierte el modelo de 3V’s en uno de **6V’s**, incluyendo la Veracidad (al igual que IBM), la Variabilidad y la Visibilidad (Wu, Buyya, & Ramamohana, 2016).

Con el fin de abarcar el significado más preciso de “Big Data”, sin dejar dimensiones atrás, se utilizará como base el modelo de 10V’s propuesto por Moro Visconti (2020) en su informe “Big Data Valuation”. En la siguiente tabla se exponen las 10 dimensiones y sus significados.

**Tabla 1. Las 10 Dimensiones de Big Data**

<b>Volumen</b>	La cantidad de datos almacenada aumenta significativamente la profundidad, precisión y calidad de la información disponible (Moro Visconti, 2020). Hoy en día, se genera, cada dos días, la misma cantidad de datos que desde el principio de los tiempos hasta el 2003 (Marr, 2017). Con mayor cantidad de datos es posible reducir las diferencias entre predicciones y datos reales.
<b>Velocidad</b>	Relacionado con la rapidez a la que se genera y accede a los datos. Va de la mano de la tecnología puesto que la velocidad incrementa a medida que el sistema (e.g. bases de datos) mejora, gracias a inteligencia artificial o machine learning (Laney, 2001). De la misma manera que el volumen, puede reducir la brecha entre predicción y datos reales (Moro Visconti, 2020).
<b>Variiedad</b>	No existe mayor barrera para una gestión eficaz de datos que la variedad de formatos de datos (Laney, 2001). Se deben analizar gran variedad de datos estructurados, semiestructurados y no estructurados buscando (1) hacer coincidir predicciones con resultados reales, (2) predecir modelos de riesgo y (3) proporcionar un análisis más profundo y efectivo (Moro Visconti, 2020).
<b>Veracidad</b>	Como de fiables son los datos no solo en cuanto a su calidad, sino cuán confiable es la fuente, el tipo y el procesamiento de los mismos. Cuanto mayor velocidad y variedad de los datos, más difícil se hace comprobar la veracidad de estos (Manogaran, Thota, Lopez, & otros, 2017).
<b>Validez</b>	A diferencia de la veracidad, la validez se preocupa por la legalidad, integridad y transparencia de los datos. Es necesario tener en cuenta la base legal asociada con su procesamiento y si está dentro del período de retención apropiado de los datos (Moro Visconti, 2020).
<b>Variabilidad</b>	Entendido como la consistencia de los datos. Cuando hay alta variabilidad es importante correlacionarla con otras dimensiones como velocidad o variedad para extraer el máximo valor y reducir el riesgo sistemático (Manogaran, Thota, Lopez, & otros, 2017).
<b>Virilidad</b>	Representa el grado de difusión de los datos a través de la red, y la velocidad a la que se envía y recibe. Tiene en cuenta la participación de los diferentes stakeholders de la red (Moro Visconti, 2020) (Manogaran, Thota, Lopez, & otros, 2017).
<b>Visualización</b>	La conexión entre la presentación y análisis visual de los datos y la representación tecnológica de los datos, la cual ayuda al usuario a mejor entenderlos. La visualización de los datos revelada por el procesamiento del big data, es una parte integral de la toma de decisiones (Moro Visconti, 2020) (Manogaran, Thota, Lopez, & otros, 2017).
<b>Viscosidad</b>	La viscosidad es un elemento de la velocidad y representa la latencia o el tiempo de retraso en la transmisión de datos entre el origen y el destino. Caracteriza la resistencia a navegar un conjunto de datos o la complejidad de su procesamiento (Manogaran, Thota, Lopez, & otros, 2017).
<b>Valor</b>	El valor monetizable es la síntesis de las características del big data. Considera los datos como un recurso a ser explotado para impulsar la innovación (Walker, 2015) (Moro Visconti, 2020). Abre paso a la comprensión de que Big Data entendido como un conjunto grande de datos, por su cuenta no tiene valor sin un análisis enfocado en los beneficios y costos de analizar y recopilar datos.

*Fuente: Adaptación de (Moro Visconti, 2020)*



### 2.3. Aplicaciones del Big Data

Los macrodatos se han convertido en un fenómeno mundial dado su gran potencial para (1) **explicar** los patrones ocultos, (2) **predecir** futuras ocurrencias con cierta precisión y (3) **simular** posibles escenarios. Estas características dotan al Big Data con la capacidad de impactar multitud de disciplinas y es por lo que se ha convertido en una herramienta multiusos. Las aplicaciones de Big Data han mostrado alta adaptabilidad a los requisitos que surgen en los dominios científicos, tecnológicos y humanísticos y las organizaciones industriales. En ocasiones, problemas originados áreas distantes se han resuelto haciendo uso de las mismas técnicas y tipos de datos (De Mauro, Greco, & Grimaldi, 2016).

A continuación, se entrará en detalle en las aplicaciones reales en diferentes ámbitos:

**Medicina:** Facilita el acceso a datos médicos desde cualquier sitio a personal médico y a pacientes lo cual impulsa a los pacientes a adoptar un rol activo en el seguimiento de su salud (Narayan & Tan, 2019). Además, se usa para integrar datos médicos con indicadores sociales como hábitos diarios, renta, estatus social, etc. Ayudando a médicos a asimilar información sin la necesidad de contacto físico (Murdoch & Detsky, 2013).

**Trafico:** los datos generados por el “intelligent monitoring & recording system” (IMRS) en el monitoreo de los sistemas de tráfico benefician en la gestión eficiente del tráfico y en la prevención del delito (Xia et.al., 2016) (Narayan & Tan, 2019).

**Energía:** medidores inteligentes de energía han facilitado la colecta de grandes cantidades de información sobre cantidad y horarios de consumición de energía, permitiendo un manejo eficiente de los picos de carga eléctrica (Narayan & Tan, 2019). Por otro lado, macrodatos sobre clima y patrones meteorológicos en el tiempo, incrementan la eficiencia de la generación de energías renovables (Zhou & Yang, 2016) (Narayan & Tan, 2019)

**Retail:** la industria de la venta minorista aprovecha los datos generados por sus transacciones para analizar el comportamiento de sus clientes mientras que aprovecha datos de otras fuentes para enfocar su marketing (Lohr, 2012) (Narayan & Tan, 2019).

**Finanzas:** es una industria muy amplia con sectores como banca comercial, banca de inversión, seguros o fondos de inversión que ha visto mucho crecimiento en volumen en los últimos años. En el sector financiero el Big Data muestra utilidad en comercio electrónico, análisis de exposición al riesgo, gestión de carteras y análisis en tiempo real del mercado de valores (Fang & Zhang, 2016) (Narayan & Tan, 2019).

**Sector Inmobiliario:** Dado el énfasis del presente trabajo sobre el impacto del Big Data en el sector inmobiliario, se indagará en mayor detalle sobre sus aplicaciones en este ámbito. En el sector inmobiliarios, los macrodatos pueden ser de gran ayuda mejorando la toma de decisiones, reduciendo costes o incrementando la productividad. A continuación, se indaga en algunas de las aplicaciones:

*Gestión del inmueble:* Tener acceso a datos sobre el internet de las cosas en los diferentes inmuebles gestionados, permite identificar los diferentes problemas comunes que suelen ocurrir. Por consiguiente, dan la posibilidad de prever y contenerlos lo cual reduce costes.

*Marketing:* Según un estudio de 2021 de la Asociación Nacional de Agentes Inmobiliarios de EE. UU., el 97% de los compradores de viviendas utilizó Internet para buscar inmuebles en 2020, mientras que un 51% encontró la vivienda que compró en la web (Wei, y otros, 2022). El Big Data puede captar toda esta información y utilizarla para la reducción del gasto en marketing a través de publicidad dirigida.

*Tasación:* La precisión de las tasaciones ha sido un tema de debate popular y estudio académico durante varias décadas. Los hallazgos de Fisher, Miles y Webb (1999) comprueban que existe una desviación absoluta promedio de 9% a 12.5% entre tasaciones y precios de transacción. Asimismo, el informe MSCI de 2016 muestra una variación entre países y años, con una diferencia entre tasaciones de propiedades y transacciones reales de entre el 7,7 % (Italia) y el 13,9 % (Japón) (Kok, Koponen, & Adriana, 2017).

Para eludir el error inherente de las tasaciones, se han desarrollado técnicas que utilizan macrodatos como por ejemplo una serie de índices alternativos de propiedad, como los índices basados en transacciones a través de Real Capital Analytics (RCA). También existen técnicas sofisticadas de suavizado para lidiar con rezagos y puntos de referencia de rendimiento suavizados artificialmente (Kok, Koponen, & Adriana, 2017).

Sin embargo, estas medidas no resuelven el problema fundamental de encontrar el valor de un activo sin la necesidad de construir un índice. Por ejemplo, encontrar el valor de un activo con motivo de inversión o financiación (Kok, Koponen, & Adriana, 2017).

*Valoración:* Tal y como argumenta este trabajo, la precisión de una valoración depende muchísimo sobre los datos que se usen en modelo, es por ello que a través de un mayor volumen de datos se pueden hallar valores mas precisos que tengan en cuenta factores invisibles a simple vista. Se indagará en más detalle sobre esta aplicación en el caso práctico.

## 2.4. Cadena de Valor y Técnicas

Para extraer valor y aplicar Big Data a la toma de decisiones, es necesario procesar los datos, ya que, sin procesamiento, simplemente tienes mares de números, letras e imágenes sin sentido. La cadena de valor de Big Data propuesta por Visconti (2020) se basa en una serie de pasos los cuales tienen valor agregado incremental de importancia económica y legal. La cadena de valor se puede representar gráficamente de la siguiente manera:

**Figura 1.** Cadena de Valor de Big Data



**Fuente:** Adaptación de (Moro Visconti, 2020)

La recopilación de datos es el primer eslabón, fundamental, en la cadena de valor y se realiza extrayendo datos de diferentes fuentes que pueden ser:

- Plataformas digitales (para el intercambio intermedio de datos, transacciones, etc.)
- Sensores / Internet de las Cosas (IoT)<sup>2</sup>
- Redes sociales
- Nuevos medios de comunicación
- Protocolos inalámbricos
- Accediendo a Internet encriptado

Los siguientes pasos (organización, procesamiento, análisis, archivo) se realizan a través de algoritmos cada vez más automatizados desde una perspectiva secuencial. Cada una de estas funciones tiene un impacto significativo en la creación de valor económico, con consecuencias jurídicas que se evaluarán dependiendo del caso (Moro Visconti, 2020).

Entre las diversas técnicas de procesamiento de datos o Big Data Analytics (BDA), las más comunes según Manyika et al. (2011) y Chen (2012) son las siguientes (De Mauro, Greco, & Grimaldi, 2016): Análisis de Conglomerados, Algoritmos Genéticos, Procesamiento del Lenguaje Natural, Aprendizaje Automático, Redes Neuronales, Modelado Predictivo, Modelos de Regresión, Análisis de Sentimiento, Procesamiento de Señales, Visualización de Datos (De Mauro, Greco, & Grimaldi, 2016).

---

<sup>2</sup> El desarrollo de los datos y la tecnología ha hecho surgir el sector conocido como Internet de las Cosas, que indica una familia de tecnologías innovadoras, cuyo fin es hacer de cualquier tipo de objeto, incluso sin vocación digital, un dispositivo conectado a Internet, que es capaz de disfrutar de todas las funcionalidades que tienen los objetos nacidos para utilizar la web (Mignemi, 2014) (Moro Visconti, 2020).

## 2.5. Privacidad y Regulación

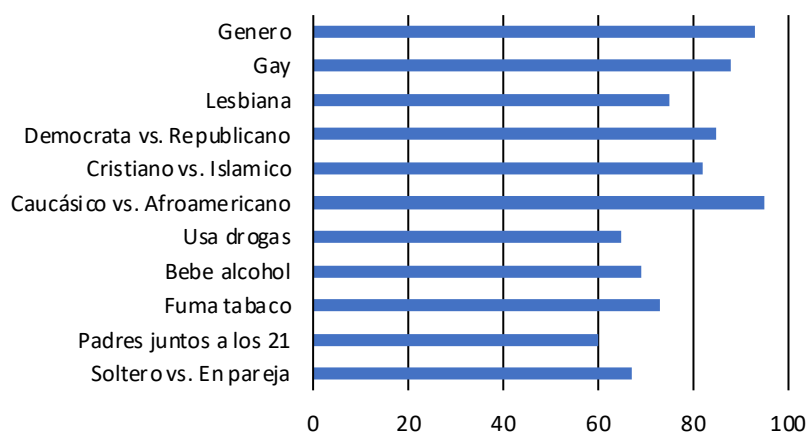
Para abastecer el abanico de aplicaciones del Big Data, organizaciones recogen cantidades masivas de todo tipo de datos, muchos sobre comportamiento e identidad. Esto significa que el Big Data sabe mucho sobre nosotros y por ello es necesario asegurar una gestión ética de nuestros datos que proteja nuestra privacidad y seguridad.

### 2.5.1. Pero ¿Cuánto se sabe sobre nosotros realmente?

Lo primero es entender que cuando surcamos la web, dejamos una huella digital. Es decir, nuestro proveedor de internet sabe todas las páginas web que hemos visitado en nuestra vida, incluso en navegación privada.

Google sabe nuestra edad y genero (incluso sin habérselo dado) y tienen un perfil extensivo de detalles sobre nosotros y nuestros intereses para personalizar sus anuncios. Del mismo modo, Meta (antes Facebook) a parte de saber quiénes son nuestros amigos y si estamos o no en una relación, es capaz de predecir si la relación va a durar o, si estas soltero, cuando estas cerca de empezar una relación y con quien (Marr, 2017). Asimismo, a través de un análisis de “me gusta”, son capaces de predecir una serie de características, entre ellas, tu inteligencia. En la *Figura 1* está representada la certidumbre de este análisis.

**Figura 2.** Porcentaje de acierto de análisis de “me gusta” (Facebook) por categorías



*Fuente:* Adaptación de (Fang & Zhang, 2016)

Como vemos en la *Figura 2* este análisis aporta un resultado bastante fiable de las características personales y hábitos de un individuo. Usando esta tecnología, compañías financieras son capaces de predecir información que sería ilegal pedir a los clientes (por leyes de discriminación) y usarla para puntuar sus calificaciones crediticias. En cambio, dado que las puntuaciones a través de Big Data usan algoritmos secretos, “es imposible analizar los algoritmos en busca de parcialidad discriminatoria”, escribió el National

Consumer Law Center en un artículo reciente sobre Big Data (Fang & Zhang, 2016). Si esta aplicación genera resultados negativos excesivos para un grupo en particular, puede convertirse en un problema de préstamo justo y muestra una violación de privacidad.

Por otro lado, a través del sistema de miles de cámaras de tráfico, que escanean y archivan fotos de coches y matrícula, la policía sabe cuándo y por donde conducimos. No solo puede saberse esto, sino que, a través de los datos de geolocalización de nuestros teléfonos móviles, puede saberse como de rápido conducimos. Por suerte, esta información no se comparte con la policía, pero cada vez más aseguradoras usan estos datos para deducir el riesgo de sus clientes y debido a ello sus cuotas (Marr, 2017).

### *2.5.2. ¿Como nos protegemos?*

En consecuencia, a las crecientes preocupaciones sobre la privacidad que planta el Big Data, se ha puesto en el punto de mira de grupos industriales reguladores y legisladores. En Estados Unidos, no existen (actualmente) leyes específicas que regulen los macrodatos. Más bien, las empresas que buscan participar en operaciones de Big Data deben asegurarse de que sus actividades cumplan con las leyes de privacidad que se aplican a los datos involucrados en sus operaciones, así como con las políticas de privacidad de empresas y los requisitos contractuales aplicables (Moro Visconti, 2020).

Por ejemplo, las instituciones financieras que deseen utilizar información personal no pública en relación con las operaciones de procesamiento de datos deberán asegurarse de cumplir con el Gramm Leach-Bliley Act (GLBA) (Moro Visconti, 2020).

Igualmente, con el fin de minimizar algunos riesgos asociados al uso de Big Data, las empresas a menudo optan por anonimizar o desidentificar los datos antes de realizar análisis. La anonimización de los datos puede ser un mecanismo eficaz para permitir que una empresa manipule, analice y estudie los datos sin tener que preocuparse por las consideraciones de privacidad (Moro Visconti, 2020).

A diferencia de Estados Unidos, la Unión Europea ha regulado el tratamiento del Big Data. De hecho, en mayo de 2018, un nuevo General Data Protection Regulation (RGPD) sustituyó a la Directiva 95/46, consolidando e innovando las normas de protección de datos. GDPR regirá la forma en que se almacenan y protegen los datos, y está destinado a devolver a los ciudadanos un mayor control de los datos que se tienen sobre ellos (Moro Visconti, 2020).

## 2.6. Big Data en el Futuro

Cierto es, que el avance y progreso que ha habido en el área de Big Data en los últimos años ha sido muy prometedor, sin embargo, todavía quedan muchos obstáculos que sobrepasar y metas que alcanzar. En este último apartado se expandirá sobre el camino que falta por recorrer viendo los retos y oportunidades futuras en el campo del Big Data.

Los macrodatos pueden encontrarse (1) estructurados o limpios, de forma que son legibles a software informáticos y están bien presentados; o (2) desestructurados o sucios, de forma que no son legibles y están mal formateados.

Estudios por la International Data Corporation (IDC) indican que el volumen de datos desestructurados está creciendo a un ritmo acelerado (Charkaborty & Patra, 2014). Además, es importante tener en cuenta que aproximadamente el 80 % de los datos generados son desestructurados y las soluciones disponibles no tienen la capacidad suficiente para analizar los datos no estructurados con precisión y presentar los conocimientos de manera comprensible (Yaqoob, y otros, 2016).

Muchos de los obstáculos en el análisis del Big Data originan de los problemas que plantea el análisis de datos no estructurados como dialectos, jerga, faltas de ortografía, formas abreviadas, acrónimos, coloquialismo, complejidades gramaticales y mezclar uno o más idiomas en el mismo texto, por nombrar algunos (Charkaborty & Patra, 2014).

Por lo tanto, actualmente, los investigadores se están enfocando en la optimización dentro de las técnicas existentes para manejar eficientemente los grandes problemas de análisis de datos no estructurados. En el futuro tendremos que desarrollar o mejores softwares capaces de analizar los datos no estructurados o mejores técnicas de limpieza de datos.

Además, el factor de complejidad en los macrodatos motiva a los investigadores a desarrollar nuevas técnicas y herramientas de análisis poderosas que pueden proporcionar información sobre datos de manera eficiente. Algunas de las áreas de investigación importantes que deben explorarse en el futuro se destacan a continuación:

**Distributed mining** (Minería distribuida): Permitiría usar diversas técnicas de análisis de forma paralela sobre el mismo set de macrodatos haciendo el proceso más eficiente. Hacer versiones distribuidas de los métodos de análisis existentes requiere mucha investigación y experiencia práctica (Yaqoob, y otros, 2016).

**Scalable Machine Learning** (Aprendizaje automático escalable): Los algoritmos de aprendizaje automático existentes no fueron diseñados para manejar la cantidad de

macrodatos actual. Por lo tanto, se requieren algoritmos de aprendizaje automático escalables para hacer frente al creciente volumen de datos (Yaqoob, y otros, 2016).

**Time Variable Data** (Datos variables de tiempo): Los datos cambian con el tiempo, por lo que es importante que las técnicas de análisis de Big Data, como la minería de datos y el aprendizaje automático, puedan adoptar y detectar estos cambios. El campo de minería de flujo es un ejemplo de minería de datos en tiempo real (Yaqoob, y otros, 2016).

**Mining from Sparse Data** (Minería de datos dispersos): Sparse data significa que hay muchas lagunas en el conjunto de datos, por lo que sacar conclusiones fiables se hace muy difícil. En algunos escenarios no muestran tendencias y distribución claras, lo que dificulta la aplicación de técnicas de minería (Yaqoob, y otros, 2016).

Explorado en las técnicas mencionadas se encuentra el campo de **Machine Learning** (ML) (aprendizaje automático), un subgrupo de la inteligencia artificial (AI) dedicado a enseñar a un programa a analizar datos por su cuenta. En conjunto, los elementos de ML permiten que las máquinas recopilen, escaneen y procesen datos, identifiquen patrones de comportamiento y luego ofrezcan predicciones para la conducta futura del grupo de personas o entidades que aportan los datos (Sivinski, Okuliar, & Kjolbye, 2017).

El análisis de datos a menudo implica una búsqueda de patrones o correlaciones en los datos. Si se encuentra un patrón que se mantenga estable en el tiempo se puede utilizar para hacer predicciones a futuro (Sivinski, Okuliar, & Kjolbye, 2017). Este campo tiene mucho camino por recorrer y es uno de los que muestra mayor potencial a futuro.

Por último, cabe destacar la aplicación del Big Data más compleja y voluminosa que podría desarrollarse en el futuro, las “**Smart cities**” o ciudades inteligentes. Estas combinarían los datos de varios aspectos de una ciudad, para posteriormente presentarlos en un formato para que todos puedan visualizar y analizar (Kitchin, 2014).

Hoy en día, algunas ciudades como Londres brindan datos en tiempo real sobre el clima, el nivel de contaminación del aire, la disponibilidad y los movimientos del transporte público, los niveles de las masas de agua, la demanda de electricidad, el estado del mercado de valores, las tendencias en las redes sociales, mira las imágenes de las cámaras de tráfico e incluso el nivel de felicidad para sus ciudadanos. Cada vez más ciudades se están convirtiendo en inteligentes y el papel de Big Data solo aumentará en el futuro (Narayan & Tan, 2019).

### 3. VALORACIÓN

Realizada la revisión de la literatura sobre el Big Data y con el mejor entendimiento obtenido sobre los conceptos del mismo, el trabajo se adentra en el tema de la valoración. Este, se analizará en búsqueda de los puntos flojos en los que se puede aplicar el Big Data.

#### 3.1. La importancia del valor

El valor es la dimensión determinante de medición en una economía de mercado. Individuos invierten con la expectativa de que cuando vendan, el valor de su inversión haya crecido lo suficiente como para compensarle por el riesgo asumido (Koller, Goedhart, & Wessels, 2010). Este factor se mantiene cierto para todo tipo de inversiones, ya sean bonos, derivados, cuentas bancarias, acciones o activos físicos como los inmobiliarios. De hecho, en una economía de mercado, la capacidad de una compañía para generar valor para los accionistas y la cantidad de valor generada son las principales medidas por las que es juzgada (Koller, Goedhart, & Wessels, 2010).

El valor es una medida de rendimiento especialmente útil porque tiene en cuenta los intereses a largo plazo de todas las partes interesadas de una empresa, no solo de los accionistas. Las medidas alternativas no son ni tan a largo plazo ni tan amplias. Por ejemplo, las ganancias contables evalúan solo el desempeño a corto plazo desde el punto de vista de los accionistas; y tomar decisiones anteponiéndolas a otros factores como el valor, puede tener consecuencias negativas para el bienestar de la compañía en el futuro. El valor, por el contrario, es relevante para todas las partes interesadas, tal y como se demuestra por la creciente vertiente del mismo en la literatura académica. En empresas, el objetivo de la maximización de valor para los accionistas a largo plazo, esta correlacionado con mayor creación de empleo, mayor satisfacción de empleados y clientes, mayor carga de responsabilidad corporativa, uso eficiente del capital humano y los recursos naturales, etc. (Koller, Goedhart, & Wessels, 2010).

La maximización de estos factores conduce a mejores niveles de vida para todos. Por ello es importante el conocimiento de cómo las empresas crean valor y cómo medir el valor.

La metodología de valoración se basa fundamentalmente en el funcionamiento de una economía de libre mercado. Por lo tanto, para una valoración efectiva es esencial comprender y posteriormente modelar la dinámica del mecanismo de precios de la oferta y la demanda que influye en los precios del mercado. Cualquier modelo normativo siempre necesita ser probado contra los fundamentos económicos del mundo real (Escobar Torres, 2021).



### 3.2. Métodos de valoración

Realizar una valoración y dar con el precio justo correcto es una ciencia compleja, especialmente por las suposiciones que se deben hacer para llevar a cabo los diferentes métodos de valoración. Algunos académicos, como el Prof. Dr. Hurhan Aydin y Aswath Damodaran, atribuyen esta dificultad a la falta de datos históricos estables, lo que puede llevar a la estimación de valores poco realistas e incorrectos (Damodaran, 2009).

El objetivo del presente trabajo es analizar si el Big Data puede tener un impacto positivo en la resolución de este problema, permitiendo realizar valoraciones más precisas. Para ver donde el Big Data puede aportar valor, es importante primero tener un conocimiento extenso sobre el funcionamiento de los métodos de valoración existentes.

#### 3.2.1. Métodos de valoración de empresas

En la valoración de empresas, según Fernández (2008), las diferentes técnicas de valoración se pueden clasificar en seis grupos, como se muestra en la *Tabla 2*: balance, cuenta de resultados, mixtos, descuento de flujos, creación de valor y opciones.

*Tabla 2. Principales metodos de valoración de empresas*

Balance	Cuenta de resultados	Mixtos (goodwill)	Descuento de flujos	Creacion de valor	Opciones
Valor contable	Múltiplos de: Beneficio: PER Ventas Ebitda Otros múltiplos	Clásico	<i>Free Cash Flow</i>	EVA	Black y Scholes
Valor contable ajust.		Unión de expertos	Cash Flow Acciones	Beneficio econ.	Opción de invertir
Valor de liquidación		Contables europeos	<i>Dividendos</i>	Cash value Added	Ampliar el proyecto
Valor sustancial		Renta abreviada	<i>Capital Cash Flow</i>	CFROI	Aplazar la inversión
Activo neto real		Otros	APV		Usos alternativos

*Fuente: Adaptación de (Fernández, 2008)*

Se entrará en mayor detalle en los primeros cuatro métodos a los cuales Fernández (2008) les aporta mayor énfasis en su obra “Métodos de valoración de empresas”.

**Balance:** Estos se basan en estimaciones del valor del patrimonio para determinar el valor de la empresa. Consideran que el valor de una empresa arraiga esencialmente de su balance o de sus activos (Fernández, 2008). Este enfoque aporta una valoración desde una perspectiva estática, sin tener en cuenta el impacto que puedan tener el estado futuro de la empresa, el valor temporal del dinero o la situación de la economía (Fernández, 2008).

**Cuenta de resultados:** Estos métodos intentan determinar el valor de la empresa en base a medidas de magnitud de la cuenta de resultados como pueden ser los beneficios, las ventas o el EBITDA. Se buscan empresas comparables de las que se extraen las medidas llamadas “múltiplos” como EV/EBITDA o P/E, usadas para valorar la empresa objetivo.

**Mixtos:** Fernández (2008) clasifica estos métodos como mixtos puesto que por un lado realizan una valoración estática de los activos (valor conjunto de su patrimonio) y, por otro, añaden una valoración dinámica que intenta cuantificar el valor a futuro (plusvalía resultante del valor de sus beneficios futuros). El fondo de comercio o “goodwill”, procura representar el calor de los activos intangibles no reflejados en el balance como valor de marca, liderazgo sectorial o calidad de las carteras de clientes (Fernández, 2008).

Es importante matizar que los métodos mixtos cada vez se emplean menos, y en vez se recurre al método de descuento de flujos de fondos ya que este constituye el único método de valoración conceptualmente correct (Fernández, 2008).

**Descuento de flujos:** Este método basado en el valor temporal del dinero intenta delimitar el valor de una empresa estimando los flujos de dinero o “cash flows” que generará en el futuro, para posteriormente descontarlos a una tasa adecuada al riesgo de estos flujos (Fernández, 2008). La precisión de los modelos de Descuento de Flujos de Caja (DCF), independientemente del que se utilice, depende de la validez de sus inputs. Todos los modelos deben imitar el mundo real lo mejor posible. Entonces, si se asume que la tasa de capitalización es X y el mercado está proporcionando Y, el modelo no se estaría imitando a la realidad y la valoración será imprecisa (Rosenbaum & Pearl, 2009).

### *3.2.2. Métodos de valoración de inmuebles*

Dentro de la valoración de terrenos y edificios e inmuebles existen tres enfoques básicos: el **enfoque de mercado** (o comparativo), el **enfoque de rentas** y el **enfoque de costos**. Dentro de cada enfoque se pueden encontrar una serie de métodos diversos a menudo adaptados para ajustar el procedimiento a la situación de valoración, el tipo de propiedad, los datos disponibles, el propósito de la valoración, la naturaleza del cliente, el marco legal local, etc. (TEGoVA, T., 2016).

**Enfoque de mercado:** la valoración se produce comparando la propiedad objetivo con la evidencia obtenida de transacciones de mercado que cumplen con los criterios para la base de valor relevante (TEGoVA, T., 2016).

**Enfoque de rentas:** describe cualquier método que obtenga el valor del activo capitalizando o descontando los ingresos futuros estimados que generará el activo.

**Enfoque de costos:** proporciona una valoración basada en el principio económico de que un comprador no pagará más por una propiedad que el costo de obtener una propiedad de igual utilidad, ya sea por compra o por construcción (TEGoVA, T., 2016).

### 3.3. Big Data en valoración

Tras este breve análisis de las diversas metodologías de valoración y conociendo ya los métodos y aplicaciones de los macrodatos, se podría argumentar que implementar técnicas de Big Data y Big Data Analytics en valoración tendría un impacto positivo en la realización de valoraciones más precisas. La pregunta clave ahora es ¿Dónde podemos implementar las técnicas de Big Data para maximizar su efectividad?

Varias de las metodologías de valoración se basan en sistemas que a partir de un set de datos y o unas suposiciones son capaces de estimar de manera precisa el valor de un activo. Entre estos encontramos los métodos de cuenta de resultados y descuento de flujos en valoración de empresas; y los enfoques de mercado y de rentas en valoración de inmuebles. Estas metodologías al recaer fuertemente sobre sus inputs para llegar a la valoración final son susceptibles a errores si los datos son incorrectos o insuficientes.

Si se indaga en mayor detalle, se pueden agrupar los métodos de valoración de empresas e inmuebles por categorías similares.

Los métodos de cuenta de resultados buscan compañías comparables de la misma manera que el enfoque de mercado busca inmuebles comparables. Ambos métodos pueden mejorar la precisión de sus valoraciones con Big Data por las mismas razones. Como se mencionó en el apartado de Big Data, un mayor set de datos aporta resultados más correctos. Por lo tanto, se podrían usar técnicas de Big Data para tener acceso a un mayor set de datos el cual aporte una valoración más precisa. Además, sobre los datos y el activo a valorar, se podrían aplicar técnicas de machine learning para encontrar similitudes ocultas a simple vista y así obtener activos comparables más correctos.

Los métodos de descuento de flujos y el enfoque de renta comparten el principio de proyectar flujos a través de suposiciones y luego descontarlos a un tipo acorde con el riesgo asumido. Utilizando técnicas de Big Data para enriquecer el conjunto de datos y en consecuencia las suposiciones, se daría lugar a valoraciones más precisas en ambos métodos por igual.

Dada la mayor facilidad para encontrar y recopilar datos inmobiliarios, durante este trabajo se indagará en mayor detalle sobre la valoración inmobiliaria. Esto se hace teniendo en cuenta que los posibles descubrimientos y resultados se puedan aplicar de la misma manera a las metodologías de valoración de empresas similares expuestas.

## 4. BIG DATA EN REAL ESTATE

El sector inmobiliario es la clase de activos más grande del mundo. Recibió su propia clasificación de industria global en agosto de 2016 y representa, en promedio, el 5,1% de cualquier cartera institucional (Kok, Koponen, & Adriana, 2017).

Como se ha comentado, para hacer más exactas las valoraciones de activos inmobiliarios, es necesario una colección de datos lo más grande posible, por ello, a continuación, se expondrán primero la técnica del web scraping para la obtención de datos y segundo los diversos tipos y fuentes de los que se pueden obtener datos inmobiliarios.

### 4.1. Web scraping

En la era de los macrodatos, el valor de la información está clara, y por ello el acceso a el Big Data de los usuarios de internet está muy cotizado y generalmente pertenece a los gigantes de la WWW como Google o Facebook. El resultado de esto, un intercambio de datos pobre y pocos canales de acceso a los datos para el público (Wei, y otros, 2022).

Actualmente, el acceso del público a datos gratuitos y sencillos de encontrar proviene predominantemente de los motores de búsqueda de Internet. El problema, es que para la mayoría de páginas web, la recolección manual de información es inviable. Para superar este obstáculo, nace la tecnología de rastreadores web, que facilita la obtención de Big Data de la web de manera automatizada (Wei, y otros, 2022).

El **web scraping**, también conocido como “web harvesting”, extracción web o raspado web es una técnica para extraer datos de la World Wide Web (WWW) y guardarlos en un sistema de archivos o base de datos para su posterior recuperación o análisis (Zhao, 2017). Por lo general, los datos web se extraen a través de un navegador web o utilizando el protocolo de transferencia de hipertexto o Hyper-text Transfer Protocol (HTTP). Puesto que una enorme cantidad de datos heterogéneos se genera constantemente en la WWW, el web scraping es ampliamente reconocido como una técnica eficiente y potente para recolectar Big Data (Mooney, Westreich, & El-sayed, 2015) (Bar-Ilan, 2001).

El proceso de extracción de datos de Internet se puede dividir en dos pasos secuenciales:

1. Adquirir recursos web
2. Extraer la información deseada de los datos adquiridos

Específicamente, un programa de web scraping comienza con la redacción de una solicitud HTTP para adquirir los recursos del sitio web. Esta solicitud se puede formatear en una URL que contenga una consulta GET o en un fragmento de mensaje HTTP que

contenga una consulta POST (Zhao, 2017). Una vez que el sitio web de destino reciba y procese correctamente la solicitud, el recurso solicitado se recuperará del sitio web y luego se enviará de vuelta al programa de web scraping (Zhao, 2017).

Para entender cómo funciona a nivel práctico un web scraper pongamos el ejemplo de que nos interesa extraer el título de 400 páginas que tienen el mismo formato y se encuentran dentro del mismo sitio web. En cada una de las 400 páginas el título está dentro de un selector <h1> que a su vez está dentro de un <div> con la clase .header (Martí, 2022). Lo que hará nuestro web scraper es detectar ese selector h1 que está dentro de la clase header (.header h1) y extraerá esa información en cada una de estas 400 páginas. Finalmente, podemos obtener esta información exportando los datos en varios formatos como un listado en .json o un fichero .csv (Martí, 2022).

De los diversos tipos de programas de raspado web, algunos se crean para reconocer automáticamente la estructura de datos de una página, como Nutch o Scrapy, o para proporcionar una interfaz gráfica basada en la web que elimina la necesidad de escribir el código de raspado manualmente, como Import.io o ParseHub (Zhao, 2017).

#### 4.2. Obtención de datos inmobiliarios

Para la obtención de datos del sector inmobiliario, la tecnología de rastreo web, se ha convertido en el método de adquisición de datos más importante. Permite obtener rápidamente Big Data de transacciones, características de las instalaciones, información ambiental e información geográfica, entre otros (Wei, y otros, 2022).

A continuación, se detallan los datos más comunes buscados para la valoración de activos inmobiliarios. Se organizan en tres tipos de datos: Internet, sensores e IoT, de acuerdo con sus fuentes y métodos de adquisición.

**Tabla 3.** Tipos y fuentes de datos inmobiliarios

<b>Tipos</b>	<b>Big Data de Internet</b>	<b>Big data de Sensores</b>	<b>Big Data de IoT</b>
<b>Fuentes</b>	Plataformas intermedias Mapas de Internet Historial de búsqueda en Internet Imágenes de internet Información de texto	Imágenes multiespectrales  Luz nocturna  Radars láser	Tarjetas inteligentes  Posicionamiento GPS  Monitoreo ambiental

*Fuente:* Adaptación de (Wei, y otros, 2022)

#### *4.2.1. Big Data de Internet*

Como se ha expuesto, el desarrollo de Internet ha generado enormes recursos de datos de todo tipo accesibles a cualquier persona con una conexión a la red, convirtiéndolo en la fuente más importante de información de Big Data.

Desde la perspectiva de la valoración de activos inmobiliarios, los tipos de Big Data de Internet más utilizados incluyen:

- 1. Datos de plataformas intermedias:** Referidos a los datos relacionados con las transacciones o arrendamientos inmobiliarios proporcionados por plataformas intermedias de Internet como Idealista, Belbex o fotocasa. Estos datos incluyen: tiempo de transacción, precio de cotización<sup>3</sup>, datos de ubicación, datos de las características del edificio (estilo, edad, nº plantas, etc.), precio de alquileres, entre otros. Estos datos contienen muchos parámetros que hacen posible evaluar bienes inmuebles en lotes a través de información de Big Data (Wei, y otros, 2022).
- 2. Datos de Mapas de Internet:** Las plataformas de mapeo de Internet proporcionan dos tipos de datos importantes para la valoración de inmuebles. Las plataformas de mapas de Internet pueden proporcionar datos de puntos de interés o points of interest (POI) con disponibilidad (casi) en tiempo real. Estas incluyen coordenadas de latitud y longitud, nombres y direcciones, todo lo cual puede proporcionar información de ubicación y vecindario lo cual puede ser clave en la realización de la valoración (Wei, y otros, 2022).
- 3. Datos del Historial de búsqueda en Internet:** Como bien se ha expuesto anteriormente, un gran número de compradores utiliza internet para buscar inmuebles (97 % en 2020). Según los académicos, esto hace posible la utilización de los historiales de búsqueda como medida del comportamiento de los compradores lo cual refleja la intención de consumo y los cambios de demanda del mercado inmobiliario por adelantado (Wei, y otros, 2022).
- 4. Datos de Imágenes de Internet:** La estética de un inmueble juega un papel vital en su valor de mercado, sin embargo, la percepción visual es una experiencia subjetiva difícil de cuantificar. Gracias al importante avance de la tecnología de visión y reconocimiento espacial, los datos de imágenes han comenzado a utilizarse para la valoración de bienes inmuebles. Introduciendo imágenes

---

<sup>3</sup> Los datos de precios se refieren principalmente al precio de cotización de la propiedad inmobiliaria o arrendamiento. Dado que el precio real de la transacción es confidencial y difícil de obtener, generalmente se reemplaza por el precio de cotización.

exteriores e interiores del inmueble que reflejen apariencia, diseño interior, decoración y otra información, y utilizando redes neuronales convulsionales para cuantificar dichas características; investigadores han descubierto que pueden mejorar la precisión se sus valoraciones (Wei, y otros, 2022).

- 5. Datos de Información de texto:** Los datos de información de texto consisten en palabras escritas por personas las cuales contienen información sentimental valiosa para ser extraída. El tipo de datos que utilizan los académicos incluye anuncios de viviendas en línea y noticias de bienes raíces. Kang, Wu & Wang (2019), creen que los precios de la vivienda se ven afectados por las actitudes psicológicas humanas. En concreto, algunos investigadores han prestado atención a la información textual que refleja la mentalidad del vendedor, y han analizado cuantitativamente la influencia de la mentalidad de este en el precio de la vivienda (Kang, Wu, & Wang, 2019) (Wei, y otros, 2022).

#### *4.2.2. Big Data de Sensores*

La tecnología de sensores puede proporcionar datos de imágenes multirresolución, multiespectral y multitemporales, además de una vista aérea de arriba a abajo de la propiedad a evaluar y el entorno comunitario que la rodea. Gracias a esto, se ha convertido en una de las pocas tecnologías que pueden ayudar a percibir las condiciones ambientales externas de una casa (Wei, y otros, 2022). Se puede dividir en las siguientes fuentes:

- 1. Imágenes multiespectrales:** Aportan información de la superficie del entorno con mayor detalle que las imágenes normales. Esto permite una mejor evaluación de las condiciones ambientales circundantes de la propiedad, lo cual es muy importante para la valoración de bienes raíces (Wei, y otros, 2022).
- 2. Luz nocturna:** La valoración de inmuebles debe tener en cuenta atributos sociales y económicos. Se ha descubierto que la cantidad de luz nocturna esta correlacionada con el producto interno bruto (PIB) y la densidad de población. Además, mayor luz nocturna indica prosperidad de las ciudades y, por lo tanto, es útil en el modelaje del precio de los inmuebles (Wei, y otros, 2022).
- 3. Radares laser:** Los radares laser son especialmente útiles, ya que pueden proporcionar parámetros tridimensionales. Estudios relevantes muestran el uso de radares para medir el área de piso y el volumen de inmuebles además del, impacto del ángulo de visión del mar en el valor de las casas de playa a través del uso de la información geomórfica de la costa (Wei, y otros, 2022).



En general, el precio de los inmuebles está muy relacionado con el entorno en el que se ubica. El estudio comparativo muestra que la combinación de imágenes de sensores y otros datos geoespaciales puede mejorar significativamente la precisión de la valoración de bienes inmuebles. Sin embargo, uso de macrodatos de sensores es difícil en modelos de valoración, ya que requiere un proceso complejo de extracción (Wei, y otros, 2022).

#### *4.2.3. Big Data de IoT*

El IoT se utiliza a día de hoy en una diversidad de campos incluyendo: transporte inteligente, hogares inteligentes, posicionamiento y navegación, gestión logística, seguridad alimentaria, atención médica digital y movilidad de la población. Tiene la ventaja de ser un recuso interactivo y en tiempo real, lo cual permite el rastreo de ubicaciones o de datos ambientales (Wei, y otros, 2022). En consecuencia, se ha investigado experimentalmente sobre su uso para la valoración inmobiliaria, en la cual se han detectado tres fuentes de datos valiosas:

- 1. Datos de tarjetas inteligentes:** Datos de tarjetas como la Oyster card del metro de Londres se utilizan comúnmente para generar mapas de rutas y representar la accesibilidad a los diferentes recursos públicos presentes. Mediante el cálculo de indicadores de accesibilidad se puede explorar la relación entre estos y los precios de la vivienda. Gracias a estudios realizados por, Siripanich y otros (2019) o, Zhu y otros (2018), sabemos que los los macrodatos sobre viajes se pueden usar hasta cierto punto para representar el estatus social o el poder económico de los compradores de vivienda (Wei, y otros, 2022).
- 2. Datos de posicionamiento GPS:** su característica distintiva es la posibilidad de rastrear la posición de dispositivos equipados con GPS en tiempo real. Además, es posible representar la trayectoria de actividad de una o más personas simultáneamente. En relación con el sector inmobiliario, el GPS de vehículos y el GPS de teléfonos móviles se utilizan para proporcionar información sobre la actividad humana, que luego se puede analizar para determinar las preferencias de las personas por los inmuebles desde la perspectiva de la demanda. Podría argumentarse que la aplicación de datos de posicionamiento GPS en la valoración inmobiliaria, es similar al de tarjetas inteligentes (Wei, y otros, 2022).
- 3. Datos de monitoreo ambiental:** Estos datos incluyen intensidad de ruido, calidad del aire, contaminación del agua y otra información ambiental. Para su obencion se suele tener que recurrir a departamentos gubernamentales quienes extraen los



datos mediante el uso de sensores ambientales. Estos factores están directamente correlacionados con el valor de un inmueble. Un estudio realizado por Zambrano-Monserrate y Ruano (2019) en la ciudad de Machala, Ecuador, muestra que cada aumento de 1 decibelio en el ruido ambiental reduce los precios de la vivienda en un 1,97 % (Wei, y otros, 2022). A pesar de este estudio, suele ser difícil estimar con precisión los parámetros ambientales debido a la dificultad de obtener información ambiental y por lo tanto son raramente usados (Wei, y otros, 2022).

## 5. CASO PRACTICO: VALORACIÓN INMOBILIARIA CON BIG DATA

Con el fin de poner en práctica los conceptos expuestos se llevará a cabo una **valoración** de los **activos de Madrid** de la empresa **Inmobiliaria Colonial SOCIMI S.A.**

### 5.1. Colonial Historia, Datos y Activos

La Inmobiliaria Colonial fue constituida en el 1946 por el Banco Hispano Colonial, con el objetivo de gestionar el importante patrimonio de terrenos del banco y los activos inmobiliarios de otras entidades financieras y algunos particulares (Colonial, 2022). En los años 70-80 Inmobiliaria Colonial fue consolidando progresivamente su posición en el sector inmobiliario español. En la *Tabla 4* podemos ver el estado actual de Colonial.

*Tabla 4. Datos de Inmobiliaria Colonial*

GAV del Grupo (millones) a 12/2021	12.436 €
Nº Activos	64 España 18 Francia 1 Logístico
Superficie total (m2)	1.677.527
Capitalización Bursatil (millones)	4.452 €
Resultado Neto Recurrente 4T 2021 (millones)	128 €
LTV de Grupo	35,8%
Rating "Investment Grade"	BBB+ Stable por S&P

*Fuente: Adaptación de (Colonial, 2022)*

A lo largo de los años Colonial ha expandido su porfolio de activos a través de: la adquisición de la inmobiliaria francesa Soci t  Fonci re Lyonnaise (SFL) (2004), la aportaci n de activos por Mutua Madrile a (2005), la fusi n con el Grupo Inmocaral (2006), las adquisiciones de Riofisa y del 15% FCC (2007), la adquisici n del 30% de SIIC Paris (2010) y la adquisici n de inmuebles sueltos con  nfasis en zonas Prime CBD.

En la *Figura 3* podemos ver m s detalles sobre la naturaleza de los activos actuales.

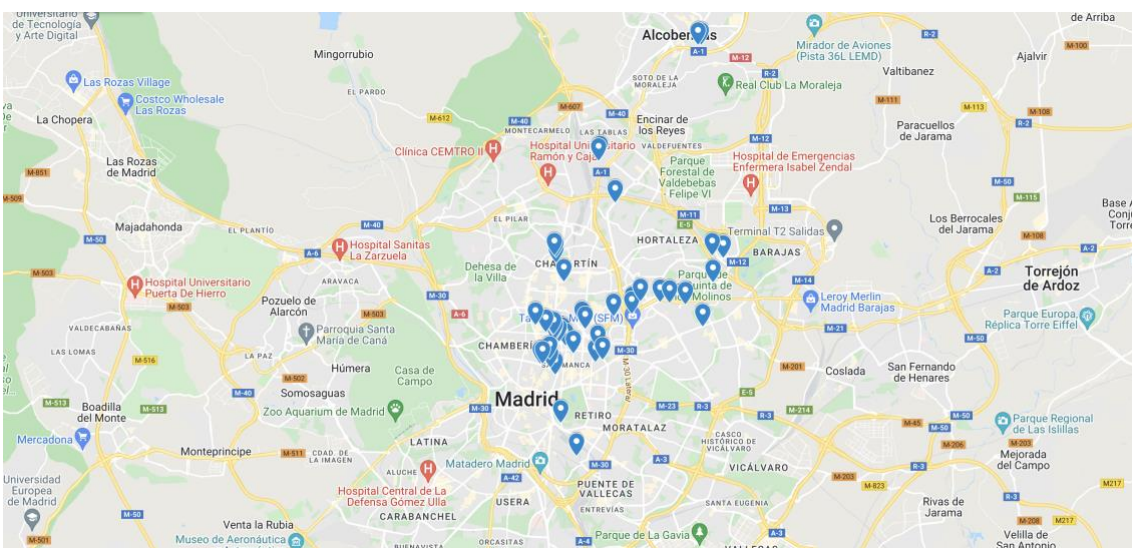
*Figura 3. Visualizaci n de Caracter sticas de Activos de Colonial*



*Fuente: Adaptaci n de (Colonial, 2022)*

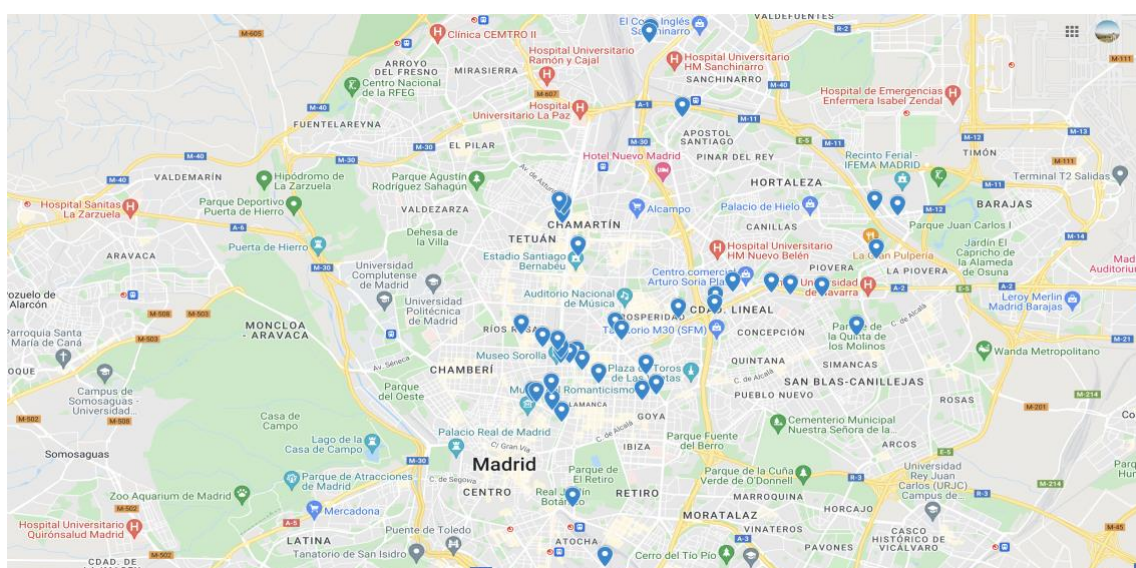
Como podemos ver en la *Figura 3* los activos de Colonial están muy concentrados a oficinas situadas en zonas Prime CBD. Es por ello que tuvo mucho sentido elegir Colonial como objeto para la valoración del caso práctico, dado que esta concentración de activos facilita significativamente la colecta y análisis de datos además del proceso de valoración. Para hacer más manejable el ejercicio de valoración, se ha decidido enfocar aún más el objeto valorado hacia los activos de Madrid. El portfolio de activos de Colonial en Madrid está compuesto por 41 inmuebles con una superficie total de 819.432m<sup>2</sup> los cuales tienen un valor de mercado de 3.265 €m, según ellos.

**Figura 4.** Mapa del universo de 41 oficinas de Colonial en Madrid (escala 2km/cm)



*Fuente: Elaboración propia*

**Figura 5.** Mapa del universo de 41 oficinas de Colonial en Madrid (escala 1km/cm)



*Fuente: Elaboración propia*

## 5.2. Obtención de datos

La información sobre precios de oficinas, ubicación de bienes raíces y características de edificios, se ha obtenido a través del web scraping, puesto que es el método más accesible. En particular, el scraping de datos para la valoración de los activos de Madrid de la inmobiliaria colonial se ha realizado utilizando el programa de raspado web ParseHub ya que esta aplicación proporciona una interfaz gráfica basada en la web que elimina la necesidad de escribir el código de raspado manualmente. Además, en ParseHub, el usuario puede crear su propia herramienta de raspado web, esto permite adaptar el scraper a las necesidades específicas de usuario.

Para realizar la valoración con las diferentes metodologías se han identificado tres sets de datos necesarios. El primero, los datos de los activos de la inmobiliaria colonial en Madrid; el segundo, un set de datos sobre precios de venta de oficinas en Madrid y el tercero, un set de datos sobre precios de alquileres de oficinas en Madrid.

A continuación, un ejemplo del proceso de extracción de datos de alquileres de oficinas.

**Figura 6.** Interfaz ParseHub (extracción precio alquileres de Belbex)

The screenshot displays the ParseHub web scraping tool interface. On the left, a sidebar shows the configuration steps for the scraper: 'Select Pagina (1)', 'Select Oficina', 'Extract Direccion', 'Relative Precio\_m2\_mes', 'Relative m2', 'Relative Planta', 'Select Next', 'Click each Next item', and 'and go to main template'. A 'Get Data' button is visible below these steps. The main area shows a browser window with the URL 'https://belbex.com/oficinas/madrid-provincia/alquiler/'. The browser displays the Belbex.com website for 'Madrid Provincia', showing a map of office locations and a table of office listings. The table below the browser window contains the following data:

Oficina_Direccion	Oficina_Precio_m2_mes	Oficina_m2	Oficina_Planta
C89, Paseo Castellana, 89 Tetuán, Madrid, 28046 Oficina en alquiler	29,50 €/m2/mes	1.130 - 3.482 m2	Plantas 4 - 8
Complejo Castellana Norte, Calle Isabel	13,50 €/m2/mes	285 - 7.775 m2	Plantas Baja - 6

**Fuente:** Programa ParseHub, elaboración propia

En la *Figura 6* vemos la interfaz que muestra la herramienta de scraping, la página web de Belbex.com y un resumen de los datos. Se ha utilizado esta página puesto que está enfocada en alquiler y venta de oficinas, lo cual encaja con el set de activos de Colonial.

En la parte izquierda de la *Figura 6* vemos la herramienta creada para extraer datos de la página de Belbex. La herramienta se puede dividir en dos códigos: El primero (Select **Oficina**) que coge los datos de oficinas extrayendo dirección, precio de alquiler (m<sup>2</sup> al



mes), superficie (m<sup>2</sup>) de la oficina y número de plantas; y el segundo (Select **Next**) que al llegar al final de la página encuentra el botón de siguiente o “next” y lleva a la herramienta la siguiente página para volver a comenzar la extracción con el primer código.

A la derecha de la *Figura 6* se pueden observar las diferentes oficinas en alquiler y sus características, las cuales la herramienta va a identificar para extraer sus datos. Finalmente, en la parte inferior de la interfaz, se encuentra una tabla que muestra un resumen de los datos que se están extrayendo y los diferentes formatos en los que se puede hacer la extracción .csv/Excel, .json o .csv/Excel Wide.

### 5.2.1. Datos sucios

Una vez realizada la extracción puedes descargar el archivo en el formato deseado. Puesto que la herramienta elegida para realizar el análisis y valoración de datos es Excel, lo más adecuado es realizar la extracción en formato csv. y posteriormente transformarlo a Excel.

No obstante, los datos todavía no están listos para ser analizados ya que están “sucios”. Datos sucios significa que hay datos faltantes, datos erróneos o representaciones no estándar de los mismos datos (Williams, 1997) (Cutter Information Corporation, 1998). Antes de que las técnicas de análisis de datos se apliquen contra cualquier dato, estos deben limpiarse para eliminar o reparar los datos sucios (Kim, 2003).

A continuación, se puede observar una muestra de los datos sucios:

**Tabla 5. Muestra de datos sucios**

Oficina_Direccion	Oficina_Precio_m2_mes	Oficina_m2	Oficina_Planta
C89, Paseo Castellana, 89 Tetuán, Madrid, 28046 Oficina en alquiler	29,50 €/m2/mes	1.130 - 3.482 m2	Plantas 4 - 8
Avenida Leonardo da Vinci, 15 Getafe, Madrid, 28906 Oficina en alquiler		554 - 3.867 m2	Plantas Baja - 2
Ronda Europa, 5 Tres Cantos, Madrid, 28760 Oficina en alquiler		234 - 1.982 m2	Plantas Baja - 1
Paseo Castellana, 94 Chamartín, Madrid, 28046 Oficina en alquiler	32 - 44 €/m2/mes	313 - 4.121 m2	Plantas Baja - Ático
Edificio Square Garden, Calle Marie Curie, 9 Rivas-Vaciamadrid, Madrid, 28529 Oficina en alquiler		440 - 41.213 m2	Plantas Baja - Ático
Edificio C9, Calle Vía de los Poblados, 3 Hortaleza, Madrid, 28033 Oficina en alquiler	13,50 €/m2/mes	858 - 5.148 m2	Plantas 1 - 6
Edificio Elipse, Avenida Manoteras, 18 Hortaleza, Madrid, 28050 Oficina en alquiler	9,50 €/m2/mes	294 - 3.554 m2	Plantas Baja - 6
Calle Santiago de Compostela, 94 Fuencarral - El Pardo, Madrid, 28029 Oficina en alquiler	16 €/m2/mes	676 - 3.382 m2	Plantas 2 - 6
Edificio Cadagua, Paseo Castellana, 93 Tetuán, Madrid, 28046 Oficina en alquiler	30 €/m2/mes	336 m2	Planta 7
Calle Valgrande, 8 Alcobendas, Madrid, 28100 Oficina en alquiler	10 €/m2/mes	700 - 6.445 m2	Plantas Baja - 2
Edificio Tanworth I, Calle María Tubau, 3 Fuencarral - El Pardo, Madrid, 28050 Oficina en alquiler	11,25 €/m2/mes	303 - 1.937 m2	Plantas Baja - 3
Oasis, Calle Basauri, 17 Moncloa - Aravaca, Madrid, 28023 Oficina en alquiler	10,50 €/m2/mes	510 m2	Planta Baja
A2 Plaza, Calle Nanclares de Oca, 17 San Blas - Canillejas, Madrid, 28022 Oficina en alquiler	13 €/m2/mes	1.068 - 24.285 m2	Plantas Baja - 4
Complejo Castellana Norte, Calle Isabel Colbrand, 22 Fuencarral - El Pardo, Madrid, 28050 Oficina en alquiler	13,50 €/m2/mes	285 - 7.775 m2	Plantas Baja - 6
Torre Rioja, Calle Rosario Pino, 14 - 16 Tetuán, Madrid, 28020 Oficina en alquiler	25 €/m2/mes	436 - 2.536 m2	Plantas 5 - 8
Paseo Castellana, 42 Salamanca, Madrid, 28046 Oficina en alquiler	32 €/m2/mes	453 - 3.624 m2	Plantas 1 - 9
Edificio OM8, Calle Julián Camarillo, 31 MADBIT San Blas - Canillejas, Madrid, 28037 Oficina en alquiler		857 - 31.494 m2	Plantas Baja - 4

**Fuente:** *Elaboración propia*

Como vemos en la *Tabla 5*, tras descargar los datos, estos contienen texto y números en las mismas casillas, texto excesivo como “Oficina en alquiler” o tienen falta de datos en algunas casillas.

En consecuencia, estos datos no son maleables ya que Excel no los reconocerá. Por ejemplo, bajo Oficina\_Precio\_m2\_mes vemos en la primera casilla el dato “29,50

€/m<sup>2</sup>/mes”, esto es ilegible para Excel, ya que no es capaz de quedarse con solo el “29,50”. Ocurre lo mismo en las diferentes columnas, por lo tanto, antes de realizar la valoración es necesario limpiar los datos.

### 5.2.2. Limpieza de Datos

El proceso de limpieza de datos o “data cleansing” es una operación que se realiza sobre los datos para eliminar anomalías y obtener un set de datos que represente de manera precisa y única el universo que intenta describir (Ridzuan & Zainon, 2019). Implica eliminar los errores, resolver las inconsistencias y transformar los datos a un formato uniforme. Con la gran cantidad de datos recopilados para el Big Data, la limpieza manual de los mismos es casi imposible, ya que requiere mucho tiempo y es propensa a errores (Ridzuan & Zainon, 2019). Es por ello, que el proceso de limpieza de datos es un proceso que debe automatizarse con herramientas de análisis de datos.

El proceso de limpieza de datos extraídos se ha estructurado de la siguiente manera:

**Figura 7. Proceso de Limpieza de Datos**



**Fuente:** Elaboración propia

Siguiendo el proceso de la *Figura 7*, lo primero es identificar los datos que vamos a querer extraer por ejemplo para el alquiler son: la dirección, el código postal y el precio del alquiler (€/m<sup>2</sup>/mes). Lo siguiente es identificar porque están sucios estos datos y cómo crear una fórmula en Excel para limpiar toda la columna por igual. Finalmente se cogen los sets de datos ya limpios y se presentan en una nueva tabla lista para ser analizada.

Para mejor entender el proceso de data cleansing, se utilizará de ejemple la extracción de datos limpios de la columna de precio de alquiler. Dentro de la columna encontramos 3 tipos diferentes de datos mostrados en la *Tabla 6*.

**Tabla 6. Reducido datos sucios extracción precio alquiler**

32 - 44 €/m <sup>2</sup> /mes
13,50 €/m <sup>2</sup> /mes

**Fuente:** Elaboración propia

En las casillas encontramos: (1) rangos de precios + “€/m2/mes” o (2) casillas en blanco o (3) precios individuales con decimales + “€/m2/mes”.

**Paso 1:** Identificamos que lo que estamos buscando extraer es un solo precio individual con dos decimales.

**Paso 2:** Lo que hace que los datos estén sucios es el texto extra (“€/m2/mes”) después de los precios, las casillas en blanco y los rangos de precio.

**Paso 3:** Para eliminar las casillas en blanco aplicamos la opción de filtro y desmarcamos las casillas en blanco. Para eliminar el texto extra (“€/m2/mes”) aplicamos la fórmula `=LEFT([@[Precio m2/mes]]; LEN([@[Precio m2/mes]])-9)` la cual recorta los últimos 9 caracteres de cada casilla. Finalmente, para reducir el rango a un solo precio, utilizamos las fórmulas `=RIGHT([@[Precio m2/mes]]; 2)` y `=LEFT([@[Precio m2/mes]]; 2)` para separar los dos valores del rango y posteriormente aplicamos la fórmula `=AVERAGE([@[Precio m2/mes]];[@[Precio m2/mes]])` sacando la media de los valores.

**Paso 4:** Reformateamos y pasamos los datos limpios a una tabla nueva final.

Se puede ver un ejemplo del producto final una vez realizado este proceso en todas las columnas en la *Tabla 7* a continuación:

**Tabla 7.** *Tabla reducida de precios de alquiler con datos limpios*

Dirección	Codigo postal	Precio (€/m <sup>2</sup> /mes)
Paseo Castellana, 89 Tetuán, Madrid, 28046	28046	29,50
Paseo Castellana, 94 Chamartín, Madrid, 28046	28046	38,00
Edificio C9, Calle Vía de los Poblados, 3 Hortaleza, Madrid, 28033	28033	13,50
Edificio Elipse, Avenida Manoteras, 18 Hortaleza, Madrid, 28050	28050	9,50
Calle Santiago de Compostela, 94 Fuencarral - El Pardo, Madrid, 28029	28029	16,00

**Fuente:** *Elaboración propia*

En esta nueva tabla con datos limpios, Excel es capaz de reconocer con facilidad los datos, y por lo tanto se puede analizar el Big Data correctamente, permitiendo la valoración de los activos. Una vez terminados los procesos de extracción de datos y limpieza de datos, obtenemos una muestra de casi mil precios de alquileres de oficinas diferentes en Madrid.

Este proceso se realiza de la misma manera para obtener datos sobre precios de venta de oficinas en Madrid y sobre la cartera de activos de Colonial en Madrid.

## 5.4. Valoración de los activos

Para realizar una valuación fundada sobre la base de valor justo, se utilizarán tanto un método de valoración intrínseca a través del enfoque de renta, como un método de valoración relativa a través del enfoque de mercado. En valoración, es clave utilizar varios métodos para valorar el mismo activo, ya que de esta manera se pueden comparar los resultados a modo de comprobación.

### 5.4.1. Valoración Intrínseca

El primer método usado para valorar los activos es el **enfoque de renta**, el cual como hemos comentado proporciona una indicación de valor mediante la conversión de flujos futuros de efectivo en un único valor capital (Escobar Torres, 2021). Cabe destacar que este enfoque puede subdividirse en dos familias:

- **Métodos de Inflación implícita:** conocidos como métodos de capitalización en los que los flujos se mantienen iguales o similares en el tiempo.
- **Métodos de Inflación explícita:** Siendo el más conocido es el método de DCF, en estos, los flujos varían de año a año.

Dado que las rentas utilizadas para la valoración son los precios de los alquileres los cuales se mantienen iguales o similares en el tiempo, se utilizará el método de inflación implícita. En particular el método de capitalización directa que implica convertir las rentas esperadas en una indicación de valor mediante la aplicación de una rentabilidad apropiada o ‘yield’, para dichas rentas estimadas (frecuentemente, beneficios operativos netos). Cabe destacar que este método, al mantener las rentas iguales, no refleja el potencial futuro del incremento de rentas (Escobar Torres, 2021).

Para encontrar el valor para la rentabilidad o yield se ha utilizado en informe de Knight Frank de 2021 en el que hacen un estudio comparando las rentabilidades del bono español y alemán (activos sin riesgo) a 10 años contra las yields de oficinas europeas. En el muestran que Madrid y Barcelona ofrecen unas rentabilidades atractivas del 3,50% y 3,75% respectivamente, datos muy por encima del valor negativo de los bonos comparados (Knight Frank, 2021).

Puesto que como hemos visto, la mayoría de los activos de colonial se encuentran bajo la categoría de “Prime”, e incluso los que no siguen siendo de alto estatus, se utilizará la misma yield para valorar todos los activos. Para tener una visión más objetiva del valor se proyectarán 3 escenarios, un base case con yield 3,5% un bull a 4% y un bear a 3%.



Llegados a este punto ya se tienen todos los elementos necesarios para realizar la valoración ahora es cuestión de montar el modelo en Excel. Los datos de los que disponemos son los siguientes:

Macrodatos del precio de alquiler por metro cuadrado al mes (€/m<sup>2</sup>/mes), de 1.192 oficinas, filtrados por código postal.

Datos de la colección de los 41 inmuebles de Colonial en Madrid y su número de metros cuadrados (m<sup>2</sup>) totales, filtrados por código postal.

Valor de la Yield de inmuebles prime en Madrid que es un 3,5% adquirida del informe Knight Frank (2021).

Los códigos postales se utilizarán como medida para relacionar los precios de los alquileres con los activos valorados. tienen un papel clave en este método de valoración y se intuye que dentro de cada código postal los edificios serán comparables entre sí y por lo tanto aportan una medida correcta para la valoración.

Por lo tanto, la manera en la que se realizará la valoración es cojiendo todos los datos de precios de alquiler y filtrándolos por código postal para extraer el precio medio de alquiler por código postal. Se puede ver un reducido de esta tabla en el *Anexo 4.*

Una vez se ha generado esta tabla podemos se multiplica el alquiler medio mensual por 12 para llegar al alquiler medio anual. A continuación, un ejemplo de cómo quedaría para el código postal 28004:

**Tabla 8.** *Tabla de alquiler medio para código postal 28004*

Codigo Postal	Alquiler medio (€/m <sup>2</sup> /mes)	Alquiler medio (€/m <sup>2</sup> /año)
28004	20,96	251,5

*Fuente: Elaboración propia*

Al terminar este proceso, se obtiene una renta anual media para cada código postal. Siguiendo el método de capitalización directa del enfoque de rentas a continuación dividimos nuestra renta o sea el alquiler medio anual entre la yield que utilizaremos (3,5%) para obtener el valor objetivo de un m<sup>2</sup> por código postal. En nuestro ejemplo:

$$251,5 / 3,5\% = 7.185 \text{ €/m}^2$$

El último paso es por lo tanto multiplicar el número de metros cuadrados de cada activo por el valor objetivo de cada 1 de sus metros cuadrados según el código postal en el que se halle. A continuación, la realización para el inmueble nº 1 de la muestra de activos:

**Tabla 9. Valor activo nº1**

Nº	Codigo Postal	m <sup>2</sup>	Valor codigo postal 28004	Valor activo
1	28004	17.202	7.185,71	123.608.583

*Fuente: Elaboración propia*

Como vemos en la tabla 9 el valor del activo nº 1 es por lo tanto 123.608.583 €. Al realizar este proceso con la herramienta Excel, se utilizan fórmulas para automatizar el procedimiento. De esta manera se genera automáticamente el valor de los 41 activos disponible en el Anexo 5 y sumando los resultados llegamos al valor total. Este proceso se realiza con los 3 escenarios de yield para obtener los siguientes resultados:

**Tabla 10. Valor total activos en € (método intrínseco)**

Bull 4% yield	Base 3,5% yield	Bear 3% yield
2.869.615.477	3.279.560.545	3.826.153.969

*Fuente: Elaboración propia*

Como vemos en la Tabla 10 el resultado de este método de valoración usando Big Data nos aporta un rango entre 2,8B€ y 3,8B€. Para comprobar la precisión de esta metodología se encuentran dos opciones con las que se puede comparar (1) el valor de los activos que colonial mismo propone que valen y (2) el resultado del método de valoración relativa.

Por lo tanto, en primer lugar, comparamos nuestros resultados con el valor de mercado de los activos publicado en la página web corporativa de colonial qué es el siguiente:

**Figura 8. Valor de mercado de activos de Madrid según Colonial**



*Fuente: Adaptación de (Colonial, 2022)*

Comparando el caso base con valor de 3.279 €m, con el valor de mercado propuesto por colonial con valor de 3.265 €m, vemos un margen de diferencia del 0,45%, lo cual aporta mucha credibilidad a esta valoración.

Se encuentran dos explicaciones para este resultado, la primera, es que se podría argumentar que una combinación de factores no relacionados aleatorios que se contrarrestan entre sí ha culminado en una valoración que parece extremadamente precisa. La segunda, argumenta que esto es una clara demostración del potencial que tiene

la aplicación del Big Data a la valoración, ya que con recursos limitados y poca experiencia en este sector un estudiante de carrera es capaz de llegar a una valoración muy precisa. Con el objetivo de ir un paso más allá en la comparación del resultado, se llevará a cabo también una valoración relativa siguiendo el enfoque de mercado.

#### 5.4.2. Valoración Relativa

En el **enfoque de mercado**, la valoración se produce comparando la propiedad objetivo con la evidencia obtenida de transacciones de mercado que cumplen con los criterios para la base de valor relevante (TEGoVA, T., 2016).

Realizar la valoración a través del enfoque de mercado implica tener en cuenta la percepción del precio de un inmueble por parte del público. Se realiza una extracción de datos precios de inmuebles en venta. utilizando el precio total del inmueble y el número de metros del inmueble se extrae el precio por metro cuadrado y utilizando la dirección se posiciona en el mapa para utilizar como comparable.

El problema encontrado en este método es que no hay tantos inmuebles de oficinas en venta en Madrid, y por lo tanto la muestra alcanza solamente datos de 512 comparables. Esto se convierte en un problema mayor al utilizar el método anterior de contraposición por códigos postales ya que ahora al no ser tan amplia la muestra los valores están afectados mucho más por datos sin representativos.

Para evitar este obstáculo se ha realizado con uso de los mapas interactivos creados en Google Maps un estudio de comparables en detalle en el cual se han extraído cuatro comparables por cada 1 de los 41 inmuebles de colonial en Madrid. La selección de estas comparables se ha realizado superponiendo en el mapa, los activos de colonial con las oficinas en venta y eligiendo las cuatro oficinas en venta más cercanas a cada activo. Este ejercicio da como resultado la tabla vista en el *Anexo 8*.

**Figura 9.** Método de selección de comparables



**Fuente:** Elaboración propia

Una vez seleccionadas todas las comparables e introducidas sus direcciones en la tabla, se utiliza Excel para corresponder la dirección de cada comparable a la tabla en la que se encuentra su precio por metro cuadrado y devolverlo a la tabla final del estudio de comparables. A raíz de los precios de las comparables se extraen la media y la mediana del valor objetivo por metro cuadrado de cada activo de Madrid de Colonial. Multiplicando este valor por el número de metros cuadrados de cada activo, se llega al valor objetivo de cada uno.

Finalmente sumando el valor de los 41 activos llegamos al valor total el objetivo según este método relativo usando el enfoque de mercado:

**Tabla 11.** Valor total activos en € (método relativo)

Valor Total Media	Valor Total Mediana
2.063.613.578	2.060.086.999

*Fuente:* Elaboración propia

Como podemos observar esta valoración aporta un resultado más bajo que la anterior y que el valor según la página web corporativa de Colonial.

En este caso encuentran tres razones para justificar este resultado. En primer lugar, el resultado puede estar afectado por edificios comparables que no son de categoría prime y por lo tanto están disminuyendo el valor final total. En segundo lugar, se podría argumentar, que al no tener una muestra lo suficientemente extensa la valoración a través de Big Data pierde potencial y por lo tanto da un resultado sesgado, en este caso hacia la baja. En tercer lugar, se podría argumentar, que el resultado obtenido en este ejercicio se aproxima más al valor objetivo de los activos y que colonial podría haber inflado sus valoraciones en beneficio propio.

Sin embargo, al contrastar el resultado final de las dos metodologías se podría tachar este último argumento dictando que entonces la valoración a través del enfoque de rentas sería errónea.

## 5.5. Conclusión

Los métodos de valoración tradicionales se han perfeccionado gracias a las aportaciones de diversos académicos relevantes a lo largo de cientos de años. Gracias a ello, estos métodos se han consolidado en el panorama actual y su validez se ha comprobado a través de los resultados que han provisto a lo largo de los años. Sin embargo, su dependencia sobre las suposiciones necesarias y sobre los datos que se le introduzcan al modelo los hacen vulnerables a errores que se podrían solventar a través de la aplicación de técnicas de Big Data.

Como hemos observado en la literatura las aplicaciones del Big Data se pueden resumir en gran potencial para (1) **explicar** patrones ocultos, (2) **predecir** futuras ocurrencias con cierta precisión y (3) **simular** posibles escenarios. Estas 3 características en mi opinión lo hacen ideal para sobre saltar los obstáculos actuales en la valoración permitiendo alcanzar valores objetivos más precisos.

En lo que respecta a los resultados del caso práctico se ha puesto en evidencia la utilidad, valor y potencial del uso del Big Data en valoración, solidificando su posición como una herramienta útil y respondiendo a las críticas que dictan que el Big Data está sobrevalorado.

### 5.5.1. Limitaciones y mejoras

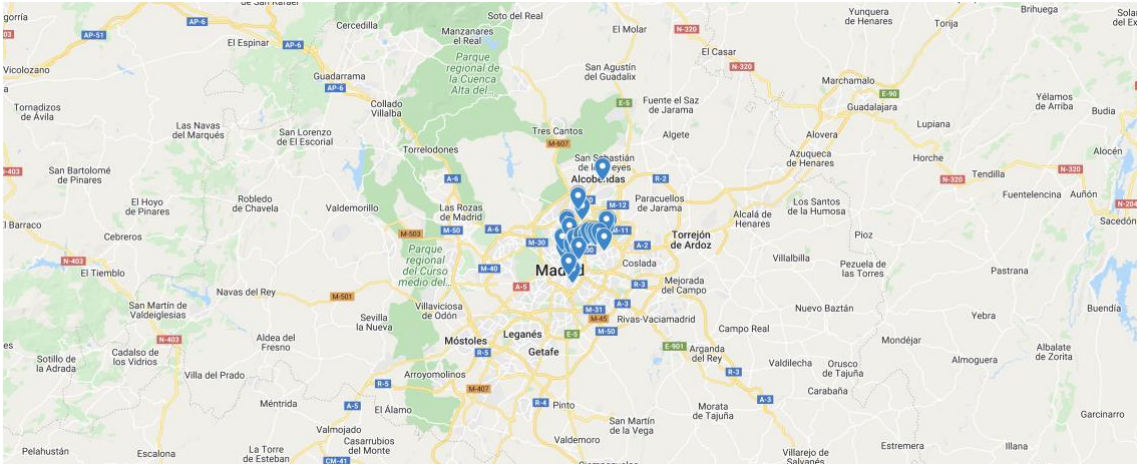
La mayor limitación con la que se ha tenido que lidiar en este caso práctico ha sido con los conocidos como “reCAPTCHA”. Estos, son los clásicos sistemas con los que todo internauta se ha topado a la hora de rellenar un formulario o crean una cuenta online y están diseñados para diferenciar entre humanos y ordenadores. Los reCAPTCHAS han impedido la utilización de la herramienta de scraping creada, para extraer datos de otras páginas inmobiliarias. De hecho, la razón por la que no se ha podido expandirla muestra de Big Data de oficinas en alquiler con un script de idealista ha sido a causa de un reCAPTCHA detectará a la herramienta de parche sáb en su primer la extracción impidiera una segunda. Otras limitaciones incluyen la inexperiencia en el sector de Big Data y la falta de capital humano para realizar un proyecto a mayor escala.

Existen además múltiples mejoras para indagar más allá sobre la aplicación del Big Data a la valoración, Incluyendo el uso de datos de las otras fuentes de datos expuestas cómo el Big Data de sensores o el Big Data de IoT.

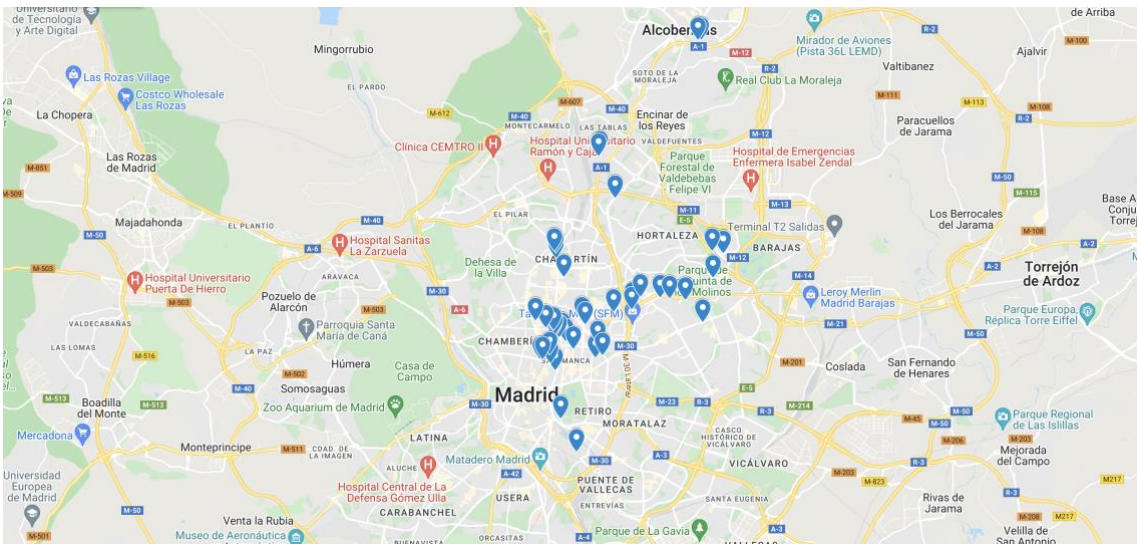


# ANEXO

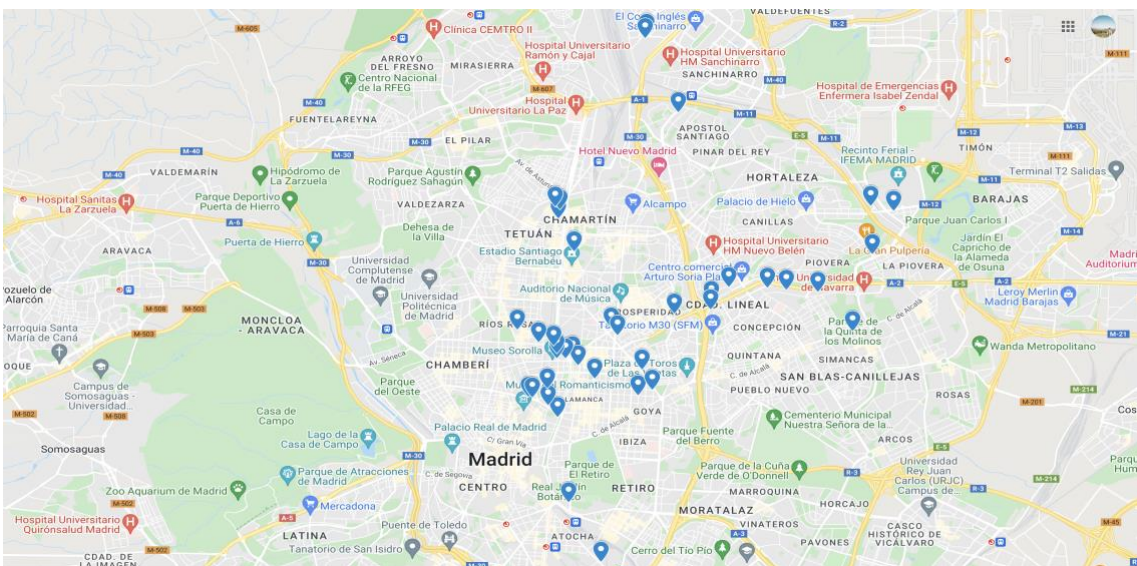
## Anexo 1. Mapas de Activos de Madrid de Inmobiliaria Colonial



Mapa del universo de 41 oficinas de Colonial en la Comunidad de Madrid (escala 10km/cm)



Mapa del universo de 41 oficinas de Colonial en la Comunidad de Madrid (escala 2km/cm)



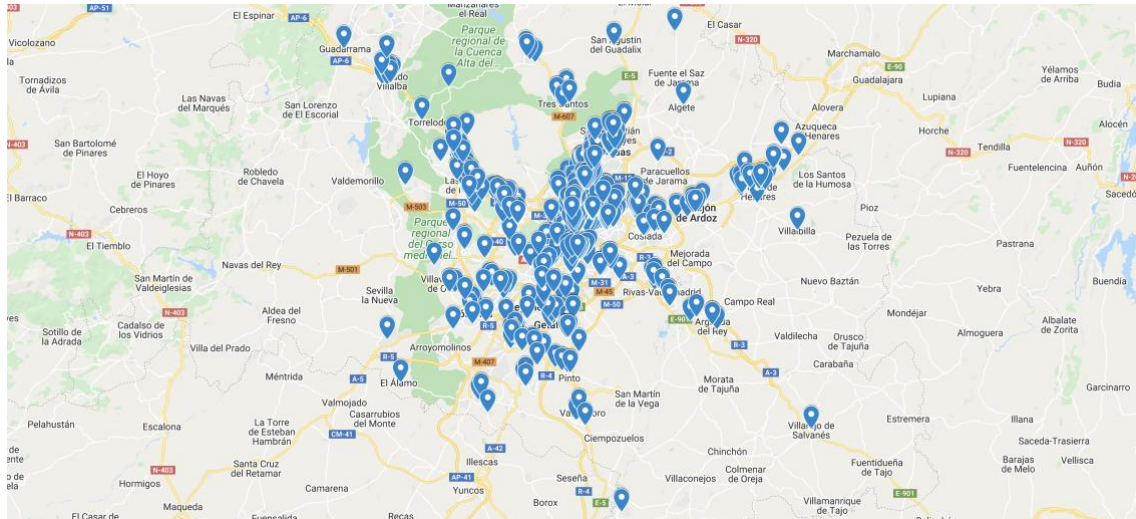
Mapa del universo de 41 oficinas de Colonial en la Comunidad de Madrid (escala 1km/cm)

*Anexo 2. Activos de Madrid de Inmobiliaria Colonial*

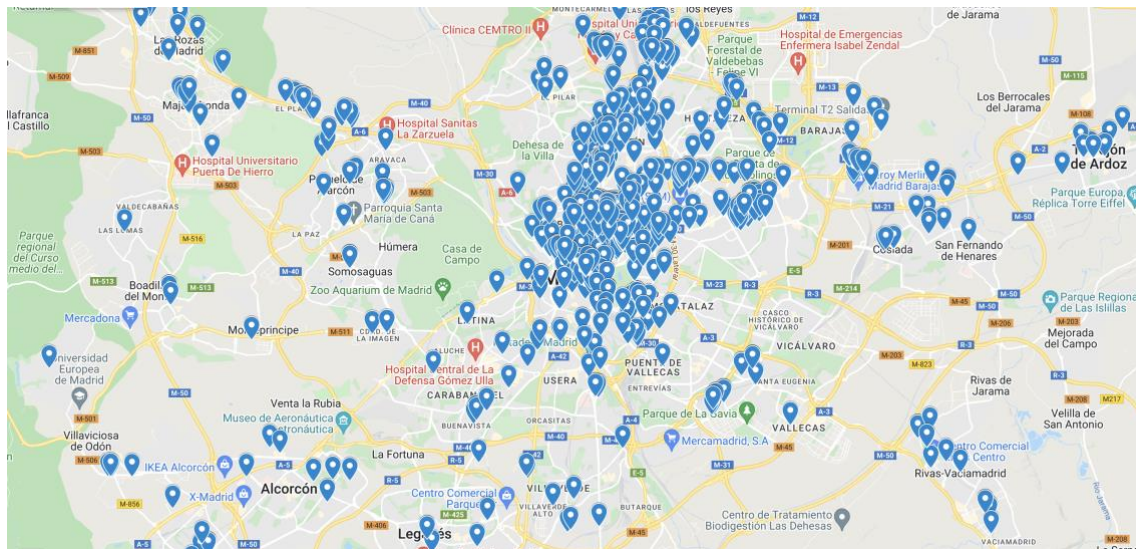
Nº	Dirección	Código Postal	m <sup>2</sup>	Plazas Parking	Nº Plantas
1	P.º de Recoletos, 37	28004	17.202	175	10
2	Calle de Génova, 17	28004	4.544	Si	9
3	P.º de la Castellana, 52	28046	7.523	Si	7
4	C. Serrano, 73	28006	4.242	100	8
5	P.º de la Castellana, 43	28046	5.998	81	9
6	C. de Miguel Ángel, 11	28010	6.300	Si	7
7	C. de Miguel Ángel, 23	28010	8.057	99	7
8	Calle Almagro, 9	28010	15.094	201	8
9	C. de Sagasta, 27	28004	4.481	0	3
10	C. de Sagasta, 31	28004	7.054	93	5
11	C. de Méndez Álvaro, 61	28045	90.000	-	-
12	Calle de José Abascal, 56	28003	12.338	219	8
13	Calle de José Abascal, 45	28003	5.300	54	8
14	Pl. del Descubridor Diego de Ordás, 3	28003	13.430	180	7
15	C. de Velázquez, 86D	28006	16.816	155	7
16	Calle de Don Ramón de la Cruz, 84	28006	9.339	91	4
17	P.º de la Castellana, 163	28046	11.000	57	12
18	C. Estébanez Calderón, 3	28020	10.000	101	9
19	C. del Poeta Joan Maragall, 53	28020	16.915	Si	9
20	C. de Manuel de Falla, 7	28036	6.252	41	4
21	C. de Arturo Soria, 336	28033	8.660	193	8
22	C. de López de Hoyos, 35	28002	7.140	Si	6
23	C. de Francisco Silvela, 42	28028	5.725	Si	7
24	C. del Príncipe de Vergara, 112	28002	11.300	115	9
25	Calle de Sta Hortensia, 26, 28	28002	46.928	946	10
26	C. Ramírez de Arellano, 37	28043	5.988	Si	8
27	C. de Martínez Villergas, 49	28027	24.135	Si	13
28	C. de José Ortega y Gasset, 100	28006	7.792	Si	7
29	Av. del Partenón, 4	28042	18.254	350	6
30	C. de Alfonso XII, 62	28014	13.374	70	8
31	C. de Puerto de Somport 8	28050	9.280	370	4
32	C. Puerto de Somport, 10, 18	28050	23.200	529	4
33	C. de Anabel Segura, 14	28108	17.138	381	5
34	C. de Francisca Delgado, 11	28108	17.033	420	5
35	Gta. Mar Caribe, 1	28043	6.327	170	5
36	C. de la Ribera del Loira, 28	28042	12.822	370	6
37	C. Ramírez de Arellano, 15	28043	6.832	112	7
38	C. de Josefa Valcárcel, 24	28027	5.652	90	7
39	C. de Josefa Valcárcel, 40	28027	8.825	Si	7
40	C. de Juan Ignacio Luca de Tena, 7	28027	10.147	260	4
41	C. de Alcalá, 506	28027	6.260	185	5



### Anexo 3. Mapas de oficinas en alquiler



Mapa de universo de oficinas en alquiler en la Comunidad de Madrid (escala 10km/cm)



Mapa de universo de oficinas en alquiler en la Comunidad de Madrid (escala 2km/cm)

Mapas accesibles en el siguiente enlace:

<https://www.google.com/maps/d/u/0/edit?mid=1bx-ySiB2x3JJW6O-Tmsiaz3oMPaEVrCg&usp=sharing>



*Anexo 4. Reducido de tabla de alquileres medios por código postal*

<b>Codigo Postal</b>	<b>Alquiler medio (€/m<sup>2</sup>/mes)</b>	<b>Alquiler medio (€/m<sup>2</sup>/año)</b>
28001	24,7	296,4
28002	16,1	192,7
28003	19,6	235,2
28004	21,0	251,5
28005	14,1	169,4
28006	22,5	269,8
28007	17,5	209,6
28008	25,9	311,2
28009	14,2	170,4
28010	18,7	224,0
28011	5,3	63,0
28223	16,4	196,8
28224	14,6	174,7
28229	12,8	153,7
28230	11,3	135,1
28231	14,7	176,0
28240	10,0	120,0
28250	18,0	216,0
28290	11,6	138,6
28400	7,1	85,0
28430	6,6	78,8
28440	6,8	81,8
28500	10,8	130,1
28529	10,3	123,4
28590	4,1	49,3
28600	8,4	100,8
28670	8,9	106,4
28700	9,7	115,8
28830	10,2	122,0
28850	12,4	148,7
28933	10,9	130,9
28934	13,7	163,9
28946	4,5	53,8
28947	7,4	88,9
28971	5,6	67,4
28978	7,8	93,4
28980	13,8	165,2
28981	15,0	180,0

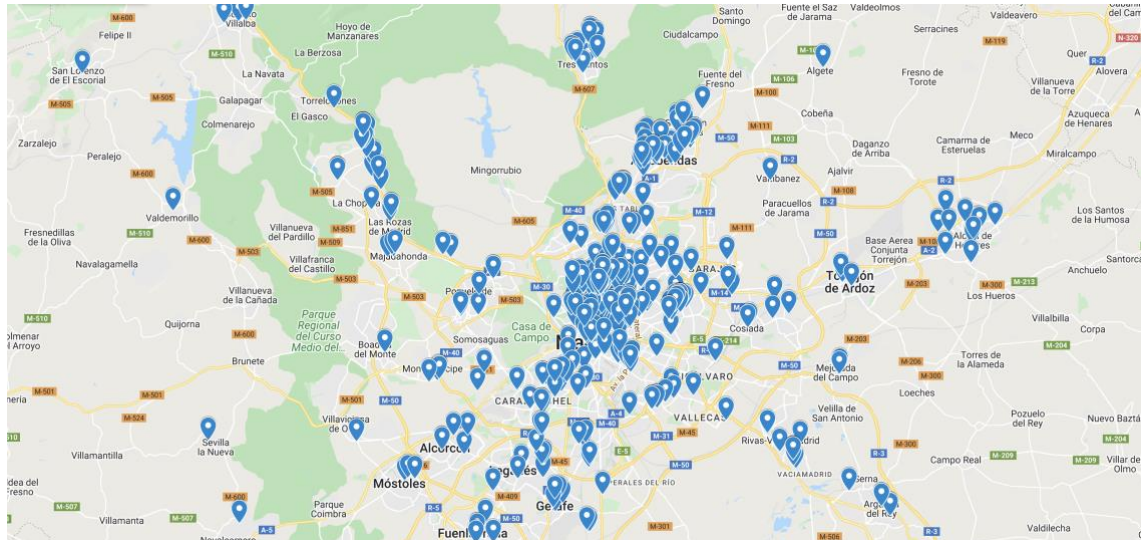
*Anexo 5. Resultado por inmueble valoración Intrínseca*

<b>Nº</b>	<b>Codigo Postal</b>	<b>m2</b>	<b>Valor activo</b>
1	28004	17.202	151.838.332
2	28004	4.544	40.108.905
3	28046	7.523	91.622.221
4	28006	4.242	40.157.507
5	28046	5.998	73.049.326
6	28010	6.300	49.531.905
7	28010	8.057	63.345.803
8	28010	15.094	118.672.154
9	28004	4.481	39.552.817
10	28004	7.054	62.264.132
11	28045	90.000	613.760.991
12	28003	12.338	101.797.728
13	28003	5.300	43.728.964
14	28003	13.430	110.807.545
15	28006	16.816	159.191.098
16	28006	9.339	88.408.995
17	28046	11.000	133.968.421
18	28020	10.000	92.943.158
19	28020	16.915	157.213.352
20	28036	6.252	49.261.373
21	28033	8.660	59.116.502
22	28002	7.140	48.286.442
23	28028	5.725	35.537.184
24	28002	11.300	76.419.719
25	28002	46.928	317.365.008
26	28043	5.988	40.639.400
27	28027	24.135	145.019.220
28	28006	7.792	73.764.096
29	28042	18.254	90.780.024
30	28014	13.374	165.627.839
31	28050	9.280	55.383.441
32	28050	23.200	138.458.602
33	28108	17.138	80.879.333
34	28108	17.033	80.383.807
35	28043	6.327	42.940.128
36	28042	12.822	63.765.831
37	28043	6.832	46.367.466
38	28027	5.652	33.960.996
39	28027	8.825	53.026.502
40	28027	10.147	60.969.962
41	28027	6.260	37.614.266

## Anexo 6. Mapas de oficinas en venta



Mapa de universo de oficinas en venta en la Comunidad de Madrid (escala 10km/cm)



Mapa de universo de oficinas en alquiler en la Comunidad de Madrid (escala 2km/cm)

Mapas accesibles en el siguiente enlace:

<https://www.google.com/maps/d/u/0/edit?mid=1bx-ySiB2x3JJW6O-Tmsiaz3oMPaEVrCg&usp=sharing>

*Anexo 7. Reducido de tabla de precios comparables*

<b>Dirección</b>	<b>Precio/m<sup>2</sup></b>
Calle Velázquez, 50 Salamanca, Madrid	8.000
Calle Villanueva Salamanca, Madrid	10.000
Calle Velázquez Salamanca, Madrid	8.889
Calle Goya, 25 Salamanca, Madrid	8.000
Calle Príncipe de Vergara, 39 Salamanca, Madrid	7.843
Calle Núñez de Balboa, 17 Salamanca, Madrid	6.667
Calle Ayala Salamanca, Madrid	6.587
Calle Príncipe de Vergara Salamanca, Madrid	5.852
Calle Espartinas Salamanca, Madrid	5.697
Calle Gil de Santivañes, 5 Salamanca, Madrid	4.767
Calle General Pardiñas, 3 Salamanca, Madrid	4.040
Calle Villalar Salamanca, Madrid	3.913
Calle Villanueva Salamanca, Madrid	3.315
Zona Goya Salamanca, Madrid	3.136
Calle Don Ramón de La Cruz Salamanca, Madrid	2.982
Calle Castelló, 23 Salamanca, Madrid	2.435
El Viso Chamartín, Madrid	9.886
Calle Serrano Chamartín, Madrid	6.145
Calle Velázquez Chamartín, Madrid	5.464
Prosperidad Chamartín, Madrid	3.917
Calle Luis Cabrera 61-63 Chamartín, Madrid	3.151
Zona Prosperidad-Chamartín Chamartín, Madrid	2.304
Calle Clara del Rey Chamartín, Madrid	581
Edificio Germania, Calle del Maestro Ángel Llorca, 6 Chamberí, Madrid	2.893
Calle Bravo Murillo Chamberí, Madrid	3.316
Calle Modesto Lafuente Chamberí, Madrid	4.286
Calle Bravo Murillo, 52 Chamberí, Madrid	3.016
Paseo San Francisco de Sales Chamberí, Madrid	2.870
Calle Modesto Lafuente Chamberí, Madrid	2.736
Avenida Reina Victoria Chamberí, Madrid	2.189
Calle Almirante Centro, Madrid	7.353
Zona Chueca - Justicia Centro, Madrid	7.266
Calle Fuencarral, 5 Centro, Madrid	6.931
Calle Argensola Centro, Madrid	5.617
Calle Argensola Centro, Madrid	5.542
Calle Sagasta Centro, Madrid	5.000
Calle Pelayo Centro, Madrid	3.614
Calle Barquillo, 38 Centro, Madrid	3.609

Anexo 8. Estudio de comparables para activos de Colonial

Nº	Comparable 1	Precio 1	Comparable 2	Precio 2	Comparable 3	Precio 3	Comparable 4	Precio 4	Media	Mediana
1	Calle Argensola Centro, N	5.617	Calle Barquillo, 38 Centr	3.609	Calle Almirante Centro, N	7.353	Calle Gil de Santivañes, 5	4.767	5.337	5.192
2	Calle Argensola Centro, N	5.617	Calle Sagasta Centro, M	5.000	Calle Zurbarán, 15 Chaml	7.813	Calle Almagro Chamberí,	7.388	6.455	6.503
3	Calle Pinar, 15 Chamartir	4.444	Calle Miguel Ángel Chan	4.286	Calle Eraso, 36 Salamanc	2.381	Paseo General Martínez	6.098	4.302	4.365
4	Calle Eraso, 36 Salamanc	2.381	Calle Miguel Ángel Chan	4.286	Calle Pinar, 15 Chamartir	4.444	Calle Velázquez Salaman	7.121	4.558	4.365
5	Calle Pinar, 15 Chamartir	4.444	Calle Miguel Ángel Chan	4.286	Calle Fernández de la Ho	4.778	Calle Eraso, 36 Salamanc	2.381	3.972	4.365
6	Calle Miguel Ángel Cham	4.286	Paseo General Martínez	6.098	Calle Fernández de la Ho	4.778	Calle Pinar, 15 Chamartir	4.444	4.902	4.611
7	Calle Miguel Ángel Cham	4.286	Paseo General Martínez	6.098	Calle Fernández de la Ho	4.778	Calle Pinar, 15 Chamartir	4.444	4.902	4.611
8	Calle Almagro Chamberí,	7.388	Calle Zurbarán, 15 Cham	7.813	Calle Sagasta Centro, Ma	5.000	Calle Fernandez de la Ho.	4.778	6.245	6.194
9	Calle Sagasta Centro, Ma	5.000	Calle Acuerdo, 20 Centrc	2.344	Calle Argensola Centro, N	6.931	Calle Argensola Centro, N	5.617	4.973	5.309
10	Calle Sagasta Centro, Ma	5.000	Calle Argensola Centro, N	5.617	Calle Zurbarán, 15 Chaml	7.813	Calle Acuerdo, 20 Centro,	2.344	5.194	5.309
11	Zona Adelfas-Retiro Reti	2.439	Calle Valderribas, 72 Re	2.151	Calle Luis Mitjans, 18 Rei	1.181	Calle Francisco Iglesias P	1.500	1.818	1.826
12	Calle Pinar, 15 Chamartir	4.444	Calle Miguel Ángel Chan	4.286	Paseo General Martínez	6.098	Calle Fernandez de la Ho.	4.778	4.902	4.611
13	Calle Santa Engracia Cha	2.794	Glorieta Pintor Sorolla C	2.771	Calle Miguel Ángel Cham	4.286	Paseo General Martínez	6.098	3.987	3.540
14	Calle Modesto Lafuente C	4.286	Calle Bravo Murillo, 52 C	3.016	Calle Santa Engracia Cha	2.794	Glorieta Pintor Sorolla Cf	2.771	3.217	2.905
15	Calle Velázquez Salaman	7.121	Calle Don Ramón de la C	4.375	Calle Eraso, 36 Salamanc	2.381	Calle Jorge Juan, 98 Salai	4.610	4.622	4.493
16	Plaza Manuel Becerra Sa	4.118	Calle Alcalá Salamanca,	3.977	Plaza Manuel Becerra Sa	4.118	Calle Antonia Mercé, 2 St	5.604	4.454	4.118
17	Calle Doctor Fleming, 3 C	5.396	Edificio Cuzco III, Paseo	4.898	Paseo Castellana, 123 Te	4.708	Avenida Brasil, 23 Tetuái	3.602	4.651	4.803
18	Calle Doctor Fleming, 53	3.726	Calle Doctor Fleming, 3	5.396	Edificio Cuzco III, Paseo C	4.898	Paseo Castellana, 123 Te	4.708	4.682	4.803
19	Calle Doctor Fleming, 53	3.726	Calle Doctor Fleming, 3	5.396	Edificio Cuzco III, Paseo C	4.898	Paseo Castellana, 123 Te	4.708	4.682	4.803
20	Feygon, Paseo Castellana	6.475	Paseo de la Habana Cha	6.051	Paseo Castellana, 123 Te	4.708	Avenida Concha Espina, 5	5.190	5.606	5.621
21	Calle de Manuel Tovar, 2	3.066	Avenida Manoteras, 30 I	2.738	Edificio B, Avenida Miano	3.836	Calle de Manuel Tovar, 2	3.066	3.177	3.066
22	El Viso Chamartín, Madri	9.886	Calle Príncipe de Vergar.	7.500	El Viso Chamartín, Madri	9.886	Calle Príncipe de Vergar.	7.500	8.693	8.693
23	Avenida Baviera, 16 Sala	1.122	Calle Doctor Gómez Ulla	4.538	Calle Francisco Navacerr:	4.444	Calle María de Molina, 3i	4.444	3.637	4.444
24	Calle Príncipe de Vergara	7.500	El Viso Chamartín, Madr	9.886	Zona Guindalera Salama	5.494	Calle María de Molina Sa	4.231	6.778	6.497
25	Zona Prosperidad-Chama	2.304	Calle Luis Cabrera 61-63	3.151	Calle López de Hoyos Sali	4.762	Calle Luis Cabrera 61-63	3.151	3.342	3.151
26	Zona Prosperidad-Chama	2.304	Calle Bueso Pineda, 12 C	3.034	Calle Fernández Caro, 7 C	1.369	Plaza Virgen del Romero	2.742	2.362	2.523
27	Zona Prosperidad-Chama	2.304	Calle Bueso Pineda, 12 C	3.034	Calle Fernández Caro, 7 C	1.369	Plaza Virgen del Romero	2.742	2.362	2.523
28	Calle Alcalá Salamanca, I	3.977	Calle Doctor Gómez Ulla	4.538	Plaza Manuel Becerra Sa	4.118	Paseo Marqués de Zafra	2.617	3.813	4.048
29	Avenida Partenón Baraja	4.439	Conde Orgaz Hortaleza,	4.825	Zona IFEMA Hortaleza, N	4.851	Avenida Partenón Baraja:	4.439	4.639	4.632
30	Calle Alcalá Retiro, Madr	7.981	Calle Alcalá Retiro, Madr	7.981	Calle Alcalá Retiro, Madr	7.981	Calle Alcalá Retiro, Madr	7.981	7.981	7.981
31	Edificio B, Avenida Mano	3.836	Avenida Manoteras, 30 I	2.738	Edificio Alfa, Avenida de	873	Campus Empresarial Arb	3.597	2.761	3.168
32	Calle Manuel Pombo Ang	3.595	Edificio B, Avenida Mani	3.836	Avenida Manoteras, 30 H	2.738	Calle Manuel Pombo Ang	3.595	3.441	3.595
33	Calle Anabel Segura Alco	3.340	Arroyo de La Vega - La N	2.292	Calle Anabel Segura Alco	3.340	Arroyo de La Vega - La M	2.292	2.816	2.816
34	Calle Anabel Segura Alco	3.340	Arroyo de La Vega - La N	2.292	Calle Anabel Segura Alco	3.340	Arroyo de La Vega - La M	2.292	2.816	2.816
35	Conde Orgaz Hortaleza, N	4.825	Avenida Partenón Baraja:	4.439	Conde Orgaz Hortaleza, N	4.825	Avenida Partenón Baraja:	4.439	4.632	4.632
36	Conde Orgaz Hortaleza, N	4.825	Avenida Partenón Baraja:	4.439	Conde Orgaz Hortaleza, N	4.825	Avenida Partenón Baraja:	4.439	4.632	4.632
37	Calle Bueso Pineda, 12 Ci	3.034	Calle Fernández Caro, 7	1.369	Calle López de Hoyos Sali:	4.762	Zona Prosperidad-Chama	2.304	2.867	2.669
38	Calle Bueso Pineda, 12 C	3.034	Calle Fernández Caro, 7 C	1.369	Plaza Virgen del Romero	2.742	Calle Bueso Pineda, 12 C	3.034	2.545	2.888
39	Calle Bueso Pineda, 12 C	3.034	Calle Fernández Caro, 7 C	1.369	Plaza Virgen del Romero	2.742	Calle Bueso Pineda, 12 C	3.034	2.545	2.888
40	Calle Doctor Zamenhof, 3	2.663	Calle Doctor Zamenhof, 3	2.663	Calle Doctor Zamenhof, 3	2.663	Calle Doctor Zamenhof, 3	2.663	2.663	2.663
41	Calle Ruffino González, 8	2.014	Calle Alfonso Gómez, 17	2.328	Calle Valentin Beato, 11	1.071	Calle Marathón, 6 San Bl:	3.243	2.164	2.171



## ***BIBLIOGRAFÍA***

- Arnaboldi, M., Busco, C., & Cuganesan, S. (2017). Accounting, accountability, social media and big data: revolution or hype? *Accounting, auditing & accountability journal*.
- Bar-Ilan, J. (2001). Data collection methods on the web for infometric purposes – A review and analysis. *Scientometrics*, 50(1), 7–32.
- Barea Mendoza, J., Rovira Fornes, J., Quecedo Gutiérrez, L., Gol Montserrat, J., & del Llano Señarís, J. (2017). Oportunidades y Retos de los Macrodatos (Big Data) En la toma de decisiones sanitarias. Fundación Gaspar Casal.
- Barnes, T. (2013). Big Data, Little History. *Dialogues in Human Geography* 3 (3) 297–302.
- Batty, M. (2013). Big data, big issues. *Geographical* 85(1): 75.
- Bryant, R., Katz, R., & Lazowska, E. (2008). Big-data computing: creating revolutionary breakthroughs in commerce, science and society.
- Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big data. *Business & Information Systems Engineering*, 5(2), 65-69.
- Caballer, V. (1976). El valor objetivo en los métodos estadísticos de valoración de inmovilizados. *Revista de Economía Política*, 74, pp. 57-75.
- Charkaborty, N., & Patra, G. (2014). Functional encryption for secured big data analytics. *International Journal of Computer Applications*, 107(16).
- Chen, H., Chiang, R. H., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188.
- Colonial. (2022, 05 15). Compañía: Colonial de un Vistazo. Retrieved from Inmobiliaria Colonial 

Colonial	Página	Web	Corporativa:
----------	--------	-----	--------------

  
<https://www.inmocolonial.com/compania/historia-colonial>
- Cutter Information Corporation. (1998). Data management strategies newsletter on the state of the data warehousing industry. *Management Science*, 31:150–162.
- Damodaran, A. (2009). Valuing Young, Start-up and Growth Companies: Estimation Issues and Valuation Challenges. Stern School of Business, New York University.

- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A Formal Definition of Big Data Based on its Essential Features. *Library Review*, Vol. 65 Iss: 3, pp.122 – 135.
- Diebold, F. (2012). On the Origin(s) and Development of the Term "Big Data".
- Escobar Torres, L. S. (2021). Métodos de valoración de inmuebles en el enfoque de renta.
- Fang, B., & Zhang, P. (2016). Big data in finance. *Big data concepts, theories, and applications*, 391-412.
- Fernández, P. (2008). *Metodos de valoración de empresas*. IESE Business School-Universidad de Navarra, 771.
- Garlasu, D., Sandulescu, V., Halcu, I., Neculoiu, G., Grigoriu, O., Marinescu, M., & Marinescu, V. (2013). A Big Data implementation based on Grid Computing. 11th RoEduNet International Conference (pp. 1-4). IEEE.
- Hall, M. (2022, Febrero 17). SGI. *Encyclopedia Britannica*.
- Kang, L., Wu, C., & Wang, B. (2019). Principles, approaches and challenges of applying big data in safety psychology research. *Frontiers in psychology*, 10, 1596.
- Kim, W. (2003). A taxonomy of dirty data. *Data mining and knowledge discovery*, 7(1), 81-99.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, 1(1), 2053951714528481.
- Knight Frank. (2021). *Oficinas: Mercado de oficinas nacional y europeo, Tendencias y oportunidades y El impulso de Madrid*. Knight Frank.
- Kok, N., Koponen, E.-I., & Adriana, C. (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal Of Portfolio Management*, 202 - 211.
- Koller, T., Goedhart, M., & Wessels, D. (2010). *Valuation: measuring and managing the value of companies* (Vol. 499). John Wiley and sons.
- Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety, Application Delivery Strategies*. Meta group.
- Lohr, S. (2012). How big data became so big. *New York Times*, 11, p. BU3.
- Manogaran, G., Thota, C., Lopez, D., & otros. (2017). Big Data Knowledge System in Healthcare. In C. Bhatt, N. Dey, & A. Ashour, *Internet of Things and Big Data*

- Technologies for Next Generation Healthcare. *Studies in Big Data* . (pp. 133-157). Springer International Publishing.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Marr, B. (2017). *Data strategy: How to profit from a world of big data, analytics and the internet of things*. Kognan Page.
- Martí, M. (2022, 05 18). *Qué es el Web scraping? Introducción y herramientas*. Retrieved from Sitalabs: <https://sitalabs.es/web-scraping-introduccion-y-herramientas/>
- Mignemi, M. (2014). *Business analytics contro la complessità di Big Data e Internet of Things*. *Harvard business review*, 11 - 52.
- Mooney, S., Westreich, D., & El-sayed, A. (2015). *Epidemiology in the era of big data*. *Epidemiology*, 26(3), 390.
- Moro Visconti, R. (2020). *Big Data Valuation*. In R. Moro Visconti, *The Valuation of Digital Intangibles*. (pp. 345-360). Palgrave Macmillan.
- Murdoch, T., & Detsky, A. (2013). *The Inevitable Application of Big Data to Health Care*. *Jama*, 309(13), pp. 1351-1352.
- Narayan, S., & Tan, H. (2019). *Adopting big data to forecast success of construction projects: A review*. *Malaysian Construction Research Journal*, 6(1), 132-143.
- Ohlhorst, F. (2012). *Big data analytics: turning big data into big money (Vol. 65)*. John Wiley & Sons.
- Press, G. (2013). *A Very Short History Of Big Data*. Forbes Tech Magazine.
- Rider, F. (1944). *The Scholar and the Future of the Research Library. A Problem and Its Solution*. Hadham Press.
- Ridzuan, F., & Zainon, W. (2019). *A review on data cleansing methods for big data*. *Procedia Computer Science*, 161, 731-738.
- Rosenbaum, J., & Pearl, J. (2009). *Investment Banking Valuation, Leveraged Buyouts, and Mergers & Acquisitions*. Wiley | Finance.



- Sivinski, G., Okuliar, A., & Kjolbye, L. (2017). Is big data a big deal? A competition law approach to big data. *European Competition Journal*, 13:2-3, 199-227, DOI: 10.1080/17441056.2017.1362866.
- Snijders, C., Matzat, U., & Reips, U.-D. (2012). "Big Data": Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science*.
- TEGoVA, T. (2016). *European Valuation Standards*.
- Walker, R. (2015). *From Big Data to Big Profits*. Oxford: Oxford University press.
- Wei, C., Fu, M., Wang, L., Yang, H., Tang, F., & Xiong, Y. (2022). The Research Development of Hedonic Price Model-Based Real Estate Appraisal in the Era of Big Data. *Land*.
- Williams, J. (1997). Tools for traveling data. *DBMS*, 10(7), 69-76.
- Wu, C., Buyya, R., & Ramamohana, K. (2016). Big Data Analytics = Machine Learning + Cloud Computing. arXiv preprint.
- Yaqoob, I., Targio Hashem, I. A., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N., & Vasilakos, A. V. (2016). Big data: From beginning to future, *International Journal of Information Management*, Volume 36, Issue 6, Part B,.
- Zhao, B. (2017). Web scraping. *Encyclopedia of big data*. Springer International Publishing AG.
- Zhou, K., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56, pp. 215 -225.
- Zillner, S., Rusitschka, S., & Skubacz, M. (2014). *Big Data Story: Demystifying Big Data with Special Focus on and Examples from Industrial Sectors*, Whitepaper. Siemens AG.
- Zuckerman, G. (2019). *The man who solved the market: How Jim Simmons launched the quant revolution*. Portfolio / Penguin.