ISSN: 2806-5697
Vol. 7 – Núm. E2 / 2026

Volumen y composición de los datos abiertos demográficos en los catálogos autonómicos españoles: un análisis correlacional

Volume and composition of open demographic data in Spanish regional catalogues: a correlational analysis

Volume e composição dos dados demográficos abertos nos catálogos regionais espanhóis: uma análise correlacional

Benitez-Perez, Nathaly Annya
Universidad Técnica Estatal de Quevedo
nbenitezp@uteq.edu.ec
https://orcid.org/0009-0008-8357-7812

Sierra-Velez, Mauricio Alexander
Universidad Técnica Estatal de Quevedo
msierrav@uteq.edu.ec
https://orcid.org/0009-0004-9097-4913

Toledo-Espinosa, Anderson Jahir
Universidad Técnica Estatal de Quevedo
atoledoe@uteq.edu.ec
https://orcid.org/0009-0009-2964-4639

Vera-Moreno, Maria Jose
Universidad Técnica Estatal de Quevedo
mveram18@uteq.edu.ec
https://orcid.org/0009-0002-7674-5714

Concha-Ramírez, Jhonny Antonio
Universidad Técnica Estatal de Quevedo
jconchar@uteq.edu.ec
https://orcid.org/0000-0002-6031-5480

DOI / URL: https://doi.org/10.55813/gaea/ccri/v7/nE2/1694

Como citar:

Benitez-Perez, N. A., Sierra-Velez, M. A., Toledo-Espinosa, A. J., Vera-Moreno, M. J., & Concha-Ramírez, J. A. (2026). Volumen y composición de los datos abiertos demográficos en los catálogos autonómicos españoles: un análisis correlacional. *Código Científico Revista De Investigación*, *7*(E2), 458–478.

Recibido: 03/08/2026 **Aceptado**: 02/09/2026 **Publicado**: 30/09/2026

Resumen
La disponibilidad de datos abiertos demográficos en los portales autonómicos españoles presenta diferencias territoriales que dificultan identificar si los catálogos más amplios ofrecen también una mayor cobertura de esta categoría. El objetivo del estudio fue analizar cuantitativamente la relación entre el número total de data sets y la disponibilidad de data sets demográficos, diferenciando el volumen absoluto de su peso relativo dentro de cada portal. Se desarrolló una investigación cuantitativa, no experimental, descriptiva y correlacional, basada en el análisis secundario de datos correspondientes a las 17 comunidades autónomas españolas. Para el tratamiento estadístico se calcularon medidas descriptivas, se evaluó la normalidad mediante la prueba de Shapiro-Wilk y se aplicó el coeficiente rho de Spearman para contrastar dos hipótesis. Los resultados evidenciaron una asociación positiva, fuerte y estadísticamente significativa entre el número total de data sets y los data sets demográficos ($\rho = 0{,}778$; $p < 0{,}001$). En cambio, la relación entre el tamaño del catálogo y el peso relativo de la categoría Demografía no fue significativa ($\rho = 0{,}310$; $p = 0{,}226$). Se concluye que el volumen y la composición temática deben evaluarse como dimensiones diferenciadas.

Palabras clave: datos abiertos; datos demográficos; portales de datos; análisis de datos; segmentación de mercados.

Abstract
The availability of open demographic data in Spanish regional portals presents territorial differences that make it difficult to determine whether larger catalogues also imply broader coverage of this category. The study aimed to quantitatively analyze the relationship between the total number of datasets and the availability of demographic datasets, distinguishing absolute volume from relative weight. A quantitative, non-experimental, descriptive, and correlational study was conducted through secondary analysis of data published for the 17 Spanish autonomous communities. Descriptive statistics were calculated, normality was assessed using the Shapiro-Wilk test, and Spearman's rho was applied to test two hypotheses. The results showed a strong, positive, and statistically significant association between the total number of datasets and the number of demographic datasets ($\rho = 0.778$; $p < 0.001$). In contrast, the relationship between catalogue size and the relative weight of the Demography category was not statistically significant ($\rho = 0.310$; $p = 0.226$). The findings indicate that a larger volume of open data is associated with greater absolute availability of demographic information, but not with greater proportional representation. It is concluded that volume and thematic composition should be assessed as distinct dimensions of open-data availability.
Keywords: open data; demographic data; data portals; data analysis; market segmentation.

Resumo
A disponibilidade de dados demográficos abertos nos portais regionais espanhóis apresenta diferenças territoriais que dificultam determinar se catálogos mais vastos implicam também uma cobertura mais ampla desta categoria. O estudo teve como objetivo analisar quantitativamente a relação entre o número total de conjuntos de dados e a disponibilidade de conjuntos de dados demográficos, distinguindo o volume absoluto do peso relativo. Foi realizado um estudo quantitativo, não experimental, descritivo e correlacional, através da análise secundária de dados publicados relativos às 17 comunidades autónomas espanholas. Foram calculadas estatísticas descritivas, a normalidade foi avaliada através do teste de Shapiro-Wilk e o coeficiente rho de Spearman foi aplicado para testar duas hipóteses. Os resultados revelaram uma associação forte, positiva e estatisticamente significativa entre o

número total de conjuntos de dados e o número de conjuntos de dados demográficos (ρ = 0,778; p < 0,001). Em contrapartida, a relação entre a dimensão do catálogo e o peso relativo da categoria «Demografia» não foi estatisticamente significativa (ρ = 0,310; p = 0,226). Os resultados indicam que um maior volume de dados abertos está associado a uma maior disponibilidade absoluta de informação demográfica, mas não a uma maior representação proporcional. Conclui-se que o volume e a composição temática devem ser avaliados como dimensões distintas da disponibilidade de dados abertos.
Palavras-chave: dados abertos; dados demográficos; portais de dados; análise de dados; segmentação de mercados

Introducción

La apertura de datos gubernamentales ha transformado la manera en que las instituciones públicas organizan y comparten información con la sociedad. Más allá de fortalecer la transparencia, estos datos pueden convertirse en una base para generar conocimiento, innovación y valor económico o social, siempre que existan condiciones que faciliten su acceso, comprensión y reutilización por distintos actores. Por ello, analizar el uso y aprovechamiento de los datos abiertos permite comprender su aporte a la innovación y a la generación de valor público. (Janssen et al., 2012; Jetzek et al., 2019).

Este potencial también resulta relevante para el sector privado, ya que las empresas pueden utilizar la información pública como insumo para diseñar servicios, identificar oportunidades y apoyar decisiones basadas en datos. En este sentido, Ahmadi Zeleti et al. (2016) destacan modelos de negocio vinculados con datos abiertos, mientras que Magalhaes & Roseira (2020) muestran formas en que las empresas incorporan datos gubernamentales en sus actividades. Sin embargo, la creación de valor no depende solo de publicar más información, sino de contar con condiciones institucionales, tecnológicas y organizacionales que hagan posible su reutilización efectiva (Jetzek et al., 2014; Zuiderwijk et al., 2014).

Cuando los datos abiertos se aplican a procesos de inteligencia y toma de decisiones, su utilidad depende de que puedan ser localizados, interpretados y combinados con otras fuentes. Gottfried et al. (2021) evidencian su uso en inteligencia empresarial y visualización.

Magalhaes & Roseira (2020) resaltan el potencial los datos libres para generar aplicaciones con valor económico. En este marco, la información demográfica adquiere especial importancia porque permite caracterizar poblaciones y territorios, aspecto clave para la segmentación de mercados y el diseño de estrategias diferenciadas (Smith, 1956).

No obstante, el crecimiento de los portales de datos abiertos no significa que todas las categorías estén igualmente representadas. Un catálogo puede ser amplio y, al mismo tiempo, ofrecer poca información demográfica. Por ello, no basta con observar el número total de data sets: también es necesario considerar su composición temática, documentación y calidad. Peng et al. (2022) subrayan la importancia de documentar adecuadamente los data sets, mientras que Rangel-Carrillo et al. (2020) proponen evaluar la calidad de los datos abiertos mediante distintas dimensiones y atributos. Así, la disponibilidad cuantitativa no debe confundirse con utilidad real para el análisis.

En el contexto español, los portales autonómicos ofrecen un escenario pertinente para examinar estas diferencias. Curto Rodríguez & Marcos Sánchez (2024) analizaron los portales de datos abiertos de las comunidades autónomas y encontraron variaciones importantes tanto en el volumen total de data sets como en la presencia de categorías temáticas. Su estudio constituye la base empírica secundaria de esta investigación y permite analizar de manera específica la relación entre el tamaño general de los catálogos y la disponibilidad de información demográfica.

A partir de este contexto, el presente estudio busca analizar cuantitativamente la relación entre la disponibilidad general de datos abiertos y la disponibilidad de información demográfica en los portales de las 17 comunidades autónomas españolas, utilizando datos secundarios de abril de 2024. De forma específica, se examina si el número total de data sets se asocia con la cantidad absoluta de data sets demográficos y si el tamaño de los catálogos se relaciona con el peso relativo de esta categoría. Con ello, se pretende aportar evidencia sobre

la composición de los catálogos autonómicos y sus implicaciones para el análisis territorial y el marketing basado en datos.

Fundamentación teórica

Los datos abiertos gubernamentales son información pública disponible para ser usada y reutilizada por distintos actores. Su valor no se limita a la transparencia, ya que también pueden apoyar la innovación, la toma de decisiones y la creación de valor económico y social. Sin embargo, publicar datos no garantiza automáticamente su uso ni sus beneficios; para ello se requieren condiciones que faciliten el acceso, la comprensión y la reutilización efectiva (Janssen et al., 2012). En esta línea, los datos abiertos pueden integrarse en modelos de negocio y procesos organizacionales cuando existen mecanismos de intercambio, gobernanza e infraestructura digital adecuados (Ahmadi Zeleti et al., 2016; Jetzek et al., 2019; Van Loenen et al., 2021; Zuiderwijk et al., 2014).

La cantidad de data sets publicados permite aproximarse al tamaño de un portal, pero no refleja por sí sola la calidad ni la utilidad de la información. Para que los datos sean aprovechables, deben contar con documentación clara, metadatos suficientes, formatos adecuados e interoperabilidad (Peng et al., 2022; Rangel-Carrillo et al., 2020). Además, los ecosistemas de datos abiertos dependen de la participación de administraciones, empresas, investigadores y usuarios, así como de mecanismos de gobernanza que permitan mantener información actualizada, accesible y útil en el tiempo (Jetzek et al., 2019; Runeson et al., 2021; Van Loenen et al., 2021).

En el ámbito empresarial, los datos abiertos pueden servir como insumo para actividades de inteligencia, innovación y toma de decisiones. Diversos estudios muestran que las empresas pueden convertir estos datos en propuestas de valor y usarlos en procesos de visualización e inteligencia empresarial (Ahmadi Zeleti et al., 2016; Gottfried et al., 2021; Magalhaes & Roseira, 2020). En marketing, esta utilidad se vincula con la segmentación de

mercados, entendida como la identificación de grupos con características diferenciadas para orientar estrategias específicas (Smith, 1956). Desde esta perspectiva, los datos demográficos abiertos pueden apoyar el análisis territorial y la investigación de mercados, aunque su disponibilidad no implica necesariamente uso efectivo.

También es necesario distinguir entre el volumen total de data sets y la composición temática de los catálogos. Un portal puede ser amplio y, aun así, tener poca información demográfica; del mismo modo, un catálogo pequeño puede dedicar una mayor proporción a esta categoría. Por ello, evaluar la disponibilidad de datos abiertos exige observar tanto la cantidad absoluta como el peso relativo de la información, además de sus condiciones de calidad, documentación y reutilización (Janssen et al., 2012; Peng et al., 2022; Rangel-Carrillo et al., 2020).

En España, Curto Rodríguez & Marcos Sánchez (2024) analizaron los portales autonómicos de datos abiertos y encontraron diferencias importantes en el número de data sets y en la distribución de categorías temáticas. Este antecedente aporta la base secundaria para el presente estudio y permite profundizar en una brecha específica: la relación cuantitativa entre el tamaño general de los catálogos autonómicos y la disponibilidad de información demográfica, tanto en términos absolutos como relativos (Ahmadi Zeleti et al., 2016; Jetzek et al., 2019; Magalhaes & Roseira, 2020; Van Loenen et al., 2021).

Modelo de investigación e hipótesis

A partir de la revisión teórica, el estudio plantea un modelo cuantitativo para analizar la relación entre la disponibilidad general de datos abiertos y la disponibilidad de información demográfica en los portales autonómicos españoles. El número total de data sets se utiliza como aproximación al tamaño del catálogo, mientras que los data sets demográficos representan una categoría temática específica. Esta disponibilidad constituye una condición inicial para la

reutilización, aunque no equivale a calidad ni a uso efectivo (Janssen et al., 2012; Jetzek et al., 2019; Peng et al., 2022).

Conceptualmente, se espera que los catálogos más amplios tengan más data sets demográficos en términos absolutos. Sin embargo, esto no implica necesariamente que la categoría Demografía tenga una mayor proporción dentro del catálogo. Esta distinción evita confundir crecimiento del portal con especialización temática y se relaciona con la necesidad de evaluar calidad, composición y reutilización de los datos abiertos (Ahmadi Zeleti et al., 2016; Magalhaes & Roseira, 2020; Peng et al., 2022; Rangel-Carrillo et al., 2020; Van Loenen et al., 2021).

H1. Existe una relación positiva y estadísticamente significativa entre el número total de conjuntos de datos abiertos y el número de conjuntos de datos demográficos en los portales de las comunidades autónomas españolas.

H2. Existe una asociación estadísticamente significativa entre el número total de conjuntos de datos abiertos y el peso relativo de los conjuntos de datos demográficos en los portales de las comunidades autónomas españolas.

Las hipótesis se interpretarán como asociaciones estadísticas, no como relaciones causales, debido al diseño no experimental del estudio y al nivel de significancia establecido en $\alpha = 0,05$.

Metodología

La investigación adopta un enfoque cuantitativo, con diseño no experimental y alcance descriptivo-correlacional. Se basa en un análisis secundario de datos, estrategia que permite reutilizar información existente para responder nuevas preguntas de investigación, siempre que se examine su procedencia, calidad y adecuación al objetivo del estudio, el estudio emplea información proveniente de fuentes de datos previamente recopiladas, considerando sus

características y disponibilidad para el desarrollo de la investigación. ([CICV2.1]Gregory et al., 2020; Johnston, 2014).

La fuente principal fue el estudio de (Marcos Sánchez, 2024), realizado en abril de 2024 sobre los portales autonómicos españoles de datos abiertos. A partir de esta fuente se construyó una base con una observación por comunidad autónoma, incorporando el número total de data sets, el número de data sets demográficos y el peso relativo de esta categoría. La población estuvo conformada por las 17 comunidades autónomas españolas, por lo que no se aplicó muestreo probabilístico.

La base se elaboró mediante la identificación de las comunidades, la extracción de los valores publicados, el cálculo del peso relativo y la revisión de consistencia de los registros. Dado que la fuente original reporta algunas dificultades de clasificación y procedimientos manuales, esta condición se considera una precaución metodológica al interpretar los resultados, para la organización de la información, se establecieron criterios uniformes que permitieron registrar los datos de cada comunidad de acuerdo con las variables y categorías disponibles en la fuente. Asimismo, los registros fueron estructurados de manera que facilitaran su posterior comparación y tratamiento estadístico, manteniendo correspondencia con la información publicada originalmente.

El análisis estadístico incluyó estadísticos descriptivos, prueba de normalidad de Shapiro-Wilk y correlación rho de Spearman. Esta última se utilizó porque permite evaluar asociaciones basadas en rangos cuando las variables no cumplen supuestos de normalidad y el número de observaciones es reducido (Pruekpramool et al., 2020; Zar, 2014). Los resultados se interpretaron según la dirección, magnitud y significancia de las asociaciones, el procesamiento de los datos permitió organizar las variables de acuerdo con sus características estadísticas y establecer las relaciones correspondientes entre ellas. Para ello, se consideraron

los valores obtenidos en cada procedimiento y los criterios estadísticos establecidos para la interpretación de las asociaciones.

Como limitación, el estudio depende de la calidad y clasificación de la fuente secundaria, utiliza información transversal y trabaja con variables agregadas que miden disponibilidad cuantitativa, no calidad, actualización, interoperabilidad ni uso efectivo. Éticamente, no se recopilaron datos personales ni se realizaron intervenciones sobre individuos; además, se reconoce explícitamente la fuente original y se diferencian sus datos de los cálculos realizados en esta investigación (Marcos Sánchez, 2024).

El modelo considera tres medidas: disponibilidad general de datos abiertos, medida por el número total de data sets; disponibilidad demográfica, medida por el número de data sets de la categoría Demografía; y peso relativo demográfico, calculado como la proporción de data sets demográficos dentro del total del catálogo.

Figura 1
Fórmula para el cálculo del peso relativo de los datasets demográficos (%)

$$\text{Peso Relativo de Datasets Demográficos}(\%) = \frac{\text{Datasets demográficos}}{\text{Datasets Totales}} \times 100$$

Nota: La figura presenta la fórmula utilizada para calcular el peso relativo de los data sets demográficos (Autores, 2026).

Resultados

Caracterización de la base analizada

El análisis comprendió las 17 comunidades autónomas españolas y considero como variables principales el número total de datastes disponible y el número de data sets correspondiente a la categoría demografía. A partir de estas dos variables también se calculó el peso relativo de la información demográfica dentro de cada catálogo. Los resultados

muestran diferencias importantes entre las comunidades en cuanto a la cantidad de información disponible. El número total de data sets oscilo entre 12 y 14.237, mientras que los data sets demográficos variaron entre 0 y 4.705. así mismo, el peso relativo de información demográfica presento valores entre 0,000% y 44,91%. Estos resultados evidencian una elevada heterogeneidad territorial tanto en volumen como en la composición de catálogos.

Estadísticos descriptivos

La disponibilidad general presentó una media de 2,331.41 data sets y una mediana de 853, mientras que la disponibilidad demográfica presentó una media de 425.71 y una mediana de 25. Para el peso relativo, la media fue de 7.81% y la mediana de 3.36%. La diferencia entre media y mediana, especialmente en las variables absolutas, refleja la presencia de valores territoriales muy elevados que incrementan el promedio.

Tabla 1

Estadísticos descriptivos de los datasets totales, datasets demográficos y su peso relativo (%)

Variable	Media	Mediana	Desv. estándar	Mínimo	Máximo
Datasets totales	2,331.41	853.00	4,057.91	12.00	14,237.00
Datasets demográficos	425.71	25.00	1,162.86	0.00	4,705.00
Peso relativo demográfico (%)	7.81	3.36	12.34	0.00	44.91

Nota: Elaboración propia a partir de los datos secundarios de (Curto Rodríguez & Marcos Sánchez, 2024) (Autores, 2026).

Distribución territorial de la información demográfica

El mayor peso relativo de data sets demográficos correspondió al Principado de Asturias (44.91%), seguido de Canarias (33.05%), País Vasco (12.00%), Región de Murcia (10.09%) y Comunidad de Madrid (4.86%). En el extremo inferior, Extremadura y Galicia presentaron 0,00%, al no registrar data sets clasificados en la categoría Demografía.

En términos absolutos, Canarias concentró 4,705 data sets demográficos, seguida por País Vasco (1,385) y Principado de Asturias (803). Estos resultados muestran que la posición

de una comunidad puede cambiar considerablemente dependiendo de si se utiliza un indicador absoluto o relativo.

Tabla 2

Distribución de datasets totales y demográficos por comunidad autónoma y peso relativo de la información demográfica (%)

Comunidad autónoma	Datasets totales	Datasets demográficos	Peso relativo (%)
Principado de Asturias	1,788	803	44.91
Canarias	14,237	4,705	33.05
País Vasco	11,539	1,385	12.00
Región de Murcia	961	97	10.09
Comunidad de Madrid	247	12	4.86
Castilla-La Mancha	341	15	4.40
Andalucía	712	31	4.35
Cantabria	290	12	4.14
Castilla y León	744	25	3.36
Cataluña	1,218	33	2.71
Comunidad Valenciana	1,453	38	2.62
La Rioja	853	18	2.11
Islas Baleares	453	9	1.99
Aragón	2,713	48	1.77
Navarra	1,561	6	0.38
Galicia	512	0	0.00
Extremadura	12	0	0.00

Nota: Elaboración propia a partir de los datos secundarios de (Curto Rodríguez & Marcos Sánchez, 2024) (Autores, 2026).

Prueba de normalidad

La prueba de Shapiro-Wilk mostró evidencia de ausencia de normalidad para las tres variables. Para data sets totales se obtuvo $W = 0.545$, $p < 0{,}001$; para data sets demográficos, $W = 0.422$, $p < 0{,}001$; y para el peso relativo demográfico, $W = 0.612$, $p < 0{,}001$. En consecuencia, y considerando además el tamaño reducido de la base, se empleó el coeficiente rho de Spearman para los contrastes correlacionales.

Tabla 3

Prueba de normalidad de Shapiro-Wilk aplicada a las variables analizadas

Variable	W de Shapiro-Wilk	p
Datasets totales	0.545	< 0,001
Datasets demográficos	0.422	< 0,001
Peso relativo demográfico	0.612	< 0,001

Nota: La prueba de Shapiro-Wilk se utilizó para evaluar la normalidad de las variables antes de seleccionar la técnica correlacional (Autores, 2026).

Contraste de la hipótesis H1

Para evaluar H1 se examinó la asociación entre el número total de data sets y el número de data sets demográficos. El coeficiente rho de Spearman fue positivo y estadísticamente significativo ($\rho = 0.778$; $p < 0,001$; $n = 17$). La magnitud del coeficiente indica una asociación positiva fuerte: las comunidades con mayor disponibilidad general de data sets tienden a presentar también una mayor cantidad absoluta de data sets demográficos.

En consecuencia, H1 queda respaldada por los datos analizados. Este resultado representa una asociación estadística y no debe interpretarse como evidencia de causalidad.

Tabla 4

Correlación de Spearman entre el total de datasets y los datasets demográficos

Relación	ρ de Spearman	p	n	Decisión
Datasets totales ↔ data sets demográficos	0.778	< 0,001	17	H1 respaldada

Nota: ρ = coeficiente rho de Spearman; n = número de comunidades autónomas analizadas (Autores, 2026).

Contraste de la hipótesis H2

Para evaluar H2 se examinó la asociación entre el número total de data sets y el peso relativo de los data sets demográficos. El coeficiente obtenido fue $\rho = 0.310$, con $p = 0.226$ y $n = 17$. Aunque la asociación presenta una dirección positiva, no alcanza significancia estadística al nivel $\alpha = 0,05$.

En consecuencia, H2 no queda respaldada. El resultado indica que, en la base analizada, un catálogo de mayor tamaño no necesariamente presenta una mayor proporción de data sets pertenecientes a la categoría Demografía.

Tabla 5

Correlación de Spearman entre el total de datasets y el peso relativo de los datasets demográficos

Relación	ρ de Spearman	p	n	Decisión
Datasets totales ↔ peso relativo demográfico	0.310	0.226	17	H2 no respaldada

Nota: ρ = coeficiente rho de Spearman; n = número de comunidades autónomas analizadas (Autores, 2026).

Síntesis del contraste de hipótesis

Tabla 6

Síntesis de los resultados del contraste de hipótesis mediante correlación de Spearman

Hipótesis	Resultado	Significancia	Conclusión
H1. Relación positiva entre data sets totales y data sets demográficos	ρ = 0.778	p < 0,001	Respaldada
H2. Asociación entre data sets totales y peso relativo demográfico	ρ = 0.310	p = 0.226	No respaldada

Nota: La decisión se estableció con un nivel de significancia de α = 0,05 (Autores, 2026).

Principales hallazgos

El análisis permite identificar tres hallazgos centrales. Primero, existe una elevada heterogeneidad entre las comunidades autónomas en cuanto al volumen total de data sets y a la disponibilidad de información demográfica. Segundo, existe una asociación positiva fuerte y estadísticamente significativa entre el volumen general del catálogo y la cantidad absoluta de data sets demográficos, lo que respalda H1. Tercero, el tamaño general del catálogo no presenta una asociación estadísticamente significativa con el peso relativo de la información demográfica, por lo que H2 no recibe respaldo.

En conjunto, los resultados diferencian dos dimensiones de la disponibilidad de información: el volumen absoluto y la composición relativa. Esta distinción será retomada en la discusión para interpretar por qué un catálogo más amplio puede contener más información demográfica sin que dicha categoría represente necesariamente una mayor proporción del conjunto.

Discusión

Interpretación de los hallazgos principales

El resultado más importante del estudio muestra una relación positiva, fuerte y estadísticamente significativa entre el número total de data sets y los data sets de la categoría Demografía ($\rho = 0{,}778$; $p < 0{,}001$; $n = 17$). Esto significa que las comunidades autónomas con catálogos más amplios tienden a contar también con más información demográfica en términos absolutos. Este hallazgo respalda la hipótesis H1 y coincide con el antecedente de Curto Rodríguez & Marcos Sánchez (2024), quienes evidenciaron diferencias relevantes en el volumen y distribución temática de los portales autonómicos españoles.

Sin embargo, este resultado debe leerse con prudencia. La asociación observada no permite afirmar causalidad, ya que el aumento de data sets demográficos puede estar relacionado con el tamaño general del catálogo, pero también con factores institucionales, técnicos, presupuestarios o de clasificación que no fueron incorporados en el modelo.

Volumen absoluto y composición relativa

El segundo hallazgo matiza la interpretación anterior. Aunque los catálogos más grandes suelen tener más data sets demográficos, no se encontró una relación estadísticamente significativa entre el total de data sets y el peso relativo de la categoría Demografía ($\rho = 0{,}310$; $p = 0{,}226$). Por ello, H2 no recibe respaldo. En otras palabras, un portal puede crecer en volumen sin que la información demográfica gane mayor participación dentro del catálogo.

Esta diferencia permite distinguir entre cantidad absoluta y composición temática. Tener más datos demográficos no significa necesariamente que esta categoría sea más prioritaria dentro del portal. Esta lectura coincide con Van Loenen et al. (2021), quienes señalan que los ecosistemas de datos abiertos deben evaluarse considerando las necesidades de los usuarios y las condiciones de generación de valor, no solo el volumen publicado.

Disponibilidad, calidad y creación de valor

Los resultados refuerzan una idea central de la literatura: la disponibilidad de datos es necesaria, pero no suficiente para generar valor. Janssen et al. (2012) advierten que publicar datos no garantiza su adopción ni reutilización, mientras que Ahmadi Zeleti et al. (2016) explican que los datos abiertos pueden transformarse en propuestas de valor cuando existen capacidades organizacionales adecuadas. En la misma línea, Jetzek et al. (2019) sostienen que la creación de valor sostenible requiere gobernanza, infraestructura, mecanismos de intercambio y condiciones de mercado.

Por ello, un mayor número de data sets debe entenderse como una oportunidad inicial y no como una garantía de utilidad. Además, las variables de este estudio miden cantidad, pero no calidad. (Peng et al., 2022; Rangel-Carrillo et al., 2020) destacan que la documentación, actualización, interoperabilidad y facilidad de reutilización son condiciones fundamentales para que los datos puedan ser aprovechados. De forma complementaria, Wang et al. (2021) señalan que persisten desafíos vinculados con calidad, gobernanza y reutilización efectiva.

Ecosistemas de datos abiertos e implicaciones para el marketing

La disponibilidad de data sets cobra sentido dentro de un ecosistema donde participan administraciones, empresas, investigadores y ciudadanía. Runeson et al. (2021) indican que estos ecosistemas dependen de las características de los datos, los metadatos y los mecanismos de gobernanza. Además, Wiedenhöft et al. (2023) identifican barreras internas, sociales, culturales y propias de los datos que pueden dificultar la cocreación de valor. Por tanto, contar con un catálogo amplio no asegura que la información se convierta automáticamente en productos, servicios o conocimiento útil.

En el campo del marketing, los datos demográficos pueden apoyar la caracterización territorial y la segmentación de mercados. Smith. (1956) planteó la segmentación como una estrategia para diferenciar grupos con características relevantes, y en el contexto digital la

información pública puede complementar otras fuentes de análisis. Gottfried et al. (2021) Y Magalhaes & Roseira 2020) muestran que los datos gubernamentales abiertos pueden apoyar la inteligencia empresarial y la creación de valor. No obstante, su utilidad depende de que la información disponible responda a necesidades concretas de análisis.

Alcance, limitaciones y gestión de datos abiertos

El alcance del estudio es territorial, comparativo y correlacional. Los resultados se limitan a las 17 comunidades autónomas analizadas y al periodo representado por la base secundaria. No permiten inferir comportamientos individuales de consumidores, empresas o usuarios, ni establecer efectos directos sobre innovación, competitividad o desempeño comercial. Además, los portales de datos abiertos son sistemas dinámicos, por lo que sus contenidos pueden cambiar con el tiempo.

Entre las principales limitaciones se encuentran el número reducido de unidades de análisis, el uso de datos secundarios, el carácter agregado de las variables y las posibles diferencias de clasificación entre portales. Por ello, los resultados deben interpretarse como una aproximación cuantitativa a la disponibilidad y no como una evaluación completa de calidad, actualización o reutilización.

Desde la gestión pública, los hallazgos sugieren que no basta con aumentar el número total de data sets. También es necesario revisar la cobertura temática, la documentación, la interoperabilidad, la actualización y la utilidad de la información publicada. Esta orientación coincide con la literatura sobre ecosistemas de datos abiertos, que destaca la necesidad de construir estructuras sostenibles y orientadas a la creación de valor (Van Loenen et al., 2021).

Futuras líneas de investigación y síntesis

Futuras investigaciones podrían ampliar este trabajo mediante estudios longitudinales, indicadores multidimensionales de calidad y medidas de uso efectivo, como consultas, descargas, aplicaciones desarrolladas o productos derivados. También sería útil relacionar la

disponibilidad de información demográfica con indicadores económicos, empresariales o comerciales para explorar posibles vínculos con procesos de inteligencia empresarial, investigación de mercados y generación de valor.

En síntesis, el estudio muestra que el tamaño general de los catálogos autonómicos se asocia fuertemente con la cantidad absoluta de data sets demográficos, pero no con su representación relativa. Esta diferencia confirma que volumen y composición temática son dimensiones distintas de la disponibilidad de datos abiertos. Por ello, una evaluación rigurosa de los portales debe ir más allá del conteo de data sets y considerar qué información se publica, cómo se documenta, qué calidad presenta y en qué condiciones puede reutilizarse.

Conclusión

A partir de los resultados obtenidos en el análisis de los portales de datos abiertos de las 17 comunidades autónomas españolas, se concluye que existe una relación positiva, fuerte y estadísticamente significativa entre disponibles general de datos abiertos y la cantidad de información demográfica disponible. El coeficiente rho de spearman obtenido de $p= 0.778$, con un nivel de significancia $p <0.001$, lo que permite determinar las comunidades autónomas que presentan un mayor número total de data sets tienden a también disponer de una mayor cantidad absoluta de data sets relacionados con la categoría demografía. Este resultado permite dar cumplimiento al objetivo de analizar la relación entre volumen general de los catálogos y la disponibilidad absoluto de información demográfica, respaldando la hipótesis H1 planteada en la investigación.

Así mismo, los resultados permitieron identificar una importante heterogeneidad entre las comunidades autónomas tanto en el volumen total de los datos abiertos como en la disponibilidad de información demográfica. Los data sets totales oscilaron entre 12 y 14.237, mientras que los data sets demográficos variaron en 0 y 4.705. de la igual manera, el peso

relativo de la información demográfica presento valores entre 0,000% y 44.91%, evidenciando que las comunidades no presenta una composición temática homogénea en sus catálogos. En términos relativos, el principado de Asturias registro la mayor proporción de data sets demográficos, mientras que canaria presento la mayor cantidad absoluta de estos data sets. Esto demuestra que la composición de la comunidad puede variar dependiendo de si se considera el volumen absoluto o la presentación proporcional de la información demográfica.

En relación con el segundo objetivo específico, orientado a determinar si el tamaño general de los catálogos se relaciona con el peso relativo de la información demográfica, los resultados demuestran que no existe una asociación estadísticamente significativa. El coeficiente obtenido fue p=0,310, con p=0,226, por lo que la relación no alcanza el nivel de significancia establecido de a=0,05. En consecuencia, la hipótesis H2 no queda respaldada. Esto significa que un catálogo con un mayor de data sets no necesariamente destina una mayor proporción de su información a la categoría demografía. Por tanto, el crecimiento de volumen generar de un portal no implica necesariamente de una mayor especialización o representación relativa de esta temática.

En conjunto, los resultados permiten establecer que la disponibilidad de información demográfica debe analizarse desde dos perspectivas diferenciadas: el volumen absoluto y la composición relativa mientras el tamaño del catálogo se relaciona de manera significativa con una mayor cantidad de datastes demográficos, no se evidencia que los catálogos más grandes otorgan una mayor participación proporcional a esta categoría. Esta diferenciación constituye unos principales aportes en el estudio, debido a que demuestra que disponer de más data sets no significa necesariamente contar con una estructura temática más orientada hacia la información demografía.

Finalmente, los resultados permiten concluir que la cantidad de datos abiertos constituye una condición inicial para su aprovechamiento, pero no garantiza por si mismo su

utilidad, calidad o generación de valor. Para que la información demográfica pueda contribuir efectivamente al análisis territorial, la segmentación de mercado, la inteligencia empresarial y la investigación de mercado, es necesario considerar aspectos adicionales como la documentación, actualización, calidad, improbabilidad y posibilidades de reutilización de los datos. En este sentido, la investigación evidencia la importancia de superar una evaluación basada únicamente en el número de data sets y avanzar hacia una valoración integral de la información disponibles y de sus posibilidades reales de aprovechamiento.

Por lo tanto, se concluye que los portales autonómicos españoles con mayor disponibilidad generar de datos abiertos tienden a ofrecer una mayor cantidad absoluta de información demográfica, pero no necesariamente una mayor representación proporcional de esta categoría. Los resultados responden a los objetivos planteados y permite comprender que el volumen y la composición temática son dimensiones diferentes de la disponibilidad de datos abiertos. Debido al carácter transversal, correlacional y agregado al estudio, estos resultados se limitan a las 17 comunidades autónoma analizadas y al periodo correspondiente a la fuente secundaria utilizada, por lo que futuras investigaciones podrían incorporar indicadores de calidad, actualización y uso efectivo de los datos para profundizar a su capacidad de generar valor.

Referencias bibliográficas

Ahmadi Zeleti, F., Ojo, A., & Curry, E. (2016). Exploring the economic value of open government data. *Government Information Quarterly, 33*(3), 535–551. https://doi.org/10.1016/j.giq.2016.01.008

Casanova-Villalba, C. I., Herrera-Sánchez, M. J., & Proaño-González, E. A. (2023). Impacto de la analítica predictiva en la toma de decisiones gerenciales. *Revista Científica Ciencia Y Método, 1*(3), 16–30. https://doi.org/10.55813/gaea/rcym/v1/n3/17

Chiguano-Chaluisa, L. J., Quishpe-Quinchuqui, J. E., Medina-Armas, A. T., & Hurtado-García, K. del R. (2025). Control interno y su influencia en la toma de decisiones en el "Comercial La Economía" cantón Balzar. *Journal of Economic and Social Science Research, 5*(1), 1–13. https://doi.org/10.55813/gaea/jessr/v5/n1/151

Concha-Ramírez, J. A., Santander-Salmon, E. S., & Sarango-Romero, V. J. (2025). Desarrollo empresarial y toma de decisiones gerenciales basadas en análisis de datos. *Revista Científica Enfoques Del Conocimiento, 2*(2), 82–98. https://doi.org/10.55813/gaea/revistacec/v2/n2/36

Curto Rodríguez, R., & Marcos Sánchez, R. (2024). Datos abiertos para la creación de valor: atención e interés para el marketing. *European Public & Social Innovation Review, 9*, 1–16. https://doi.org/10.31637/epsir-2024-1195

Frapporti, G., Linnartz, L. A. M., & Vriend, S. P. (1991). SPEARMEN: A dBase program for computation and testing of Spearman rank correlation coefficient distributions. *Computers & Geosciences, 17*(4), 569–589. https://doi.org/10.1016/0098-3004(91)90115-T

Gottfried, A., Hartmann, C., & Yates, D. (2021). Mining open government data for business intelligence using data visualization: A two-industry case study. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(4), 1042–1065. https://doi.org/10.3390/jtaer16040059

Hinds, P. S., Vogel, R. J., & Clarke-Steffen, L. (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research, 7*(3), 408–424. https://doi.org/10.1177/104973239700700306

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management, 29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), 15–16. https://doi.org/10.4067/S0718-18762014000200008

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2019). The sustainable value of open government data. *Journal of the Association for Information Systems, 20*(6), 702–734. https://doi.org/10.17705/1jais.00549

Johnston, M. P. (2014). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries, 3*(3), 619–626.

Magalhães, G., & Roseira, C. (2020). Open government data and the private sector: An empirical view on business models and value creation. *Government Information Quarterly, 37*(3), Article 101248. https://doi.org/10.1016/j.giq.2017.08.004

Mora Cortez, R., Højbjerg Clarke, A., & Freytag, P. V. (2021). B2B market segmentation: A systematic review and research agenda. *Journal of Business Research, 126*, 415–428. https://doi.org/10.1016/j.jbusres.2020.12.070

Peng, G., Lacagnina, C., Downs, R. R., Ganske, A., Ramapriyan, H. K., Ivánová, I., Wyborn, L., Jones, D., Bastin, L., Shie, C.-L., & Moroni, D. F. (2022). Global community guidelines for documenting, sharing, and reusing quality information of individual digital datasets. *Data Science Journal, 21*(1), Article 8. https://doi.org/10.5334/dsj-2022-008

Preciado-Ortiz, F. L., Jacome-Vélez, T. G., & Andrade-Díaz, K. V. (2024). Aplicaciones de inteligencia artificial generativa en la personalización de estrategias de marketing digital. *Revista Científica Ciencia Y Método, 2*(2), 14–26.

https://doi.org/10.55813/gaea/rcym/v2/n2/40

Puyol-Cortez, J. L. (2024). Factores determinantes en la toma de decisiones estratégicas en el sector retail. *Revista Científica Zambos, 3*(1), 36–55. https://doi.org/10.69484/rcz/v3/n1/11

Rangel-Carrillo, A. M., Maestre-Góngora, G. P., & Osorio-Sanabria, M. A. (2020). Principios, lineamientos, dimensiones y atributos para la evaluación de calidad de datos abiertos de gobierno. *AiBi Revista de Investigación, Administración e Ingeniería, 8*(S1), 54–65. https://doi.org/10.15649/2346030X.950

Reymundo-Soto, E., & Navarrete-Zambrano, C. M. (2022). Transparencia y responsabilidad en el sector público: Un análisis de las prácticas de auditoría gubernamental en Ecuador. *Revista Científica Zambos, 1*(2), 67–83. https://doi.org/10.69484/rcz/v1/n2/28

Runeson, P., Olsson, T., & Linåker, J. (2021). Open data ecosystems: An empirical investigation into an emerging industry collaboration concept. *Journal of Systems and Software, 182*, Article 111088. https://doi.org/10.1016/j.jss.2021.111088

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing, 21*(1), 3–8. https://doi.org/10.1177/002224295602100102

Van Loenen, B., Zuiderwijk, A., Vancauwenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., Magnussen, R., Saddiqa, M., Dulong de Rosnay, M., Crompvoets, J., Polini, A., Re, B., & Casiano Flores, C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. *JeDEM: eJournal of eDemocracy and Open Government, 13*(2), 1–27. https://doi.org/10.29379/jedem.v13i2.644

Wang, F., Zhu, H., & Wu, Y. (2021). Quality, reuse and governance of open data. *Proceedings of the Association for Information Science and Technology, 58*(1), 659–662. https://doi.org/10.1002/pra2.522

Wiedenhöft, G. C., Matheus, R., Saxena, S., & Alexopoulos, C. (2023). Barriers towards open government data value co-creation: An empirical investigation. *The Electronic Journal of Information Systems in Developing Countries, 89*(5), Article e12270. https://doi.org/10.1002/isd2.12270

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, Article 160018. https://doi.org/10.1038/sdata.2016.18

Zuiderwijk, A., Helbig, N., Gil-García, J. R., & Janssen, M. (2014). Special issue on innovation through open data: A review of the state-of-the-art and an emerging research agenda. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), I–XIII. https://doi.org/10.4067/S0718-18762014000200001