

Pontificia Universidad Católica del Ecuador
Facultad de Hábitat, Infraestructura y Creatividad
Maestría en Sistemas de Información mención Data Science



Trabajo previo para la obtención del título de Magister en Sistemas de Información
mención Data Science

Tema:

Evolución de la ciencia de datos y brecha de talento en América Latina: análisis
comparativo de oferta, demanda y desafíos (2015–2025)

Autor:

Ing. Alex Sebastián Vivas Carrillo

Tutor:

Damián Aníbal Nicolalde Rodríguez Mg.

Quito - marzo 2026

Agradecimientos

Deseo expresar mi más profundo agradecimiento a mi madre, Teresa, por su apoyo incondicional a lo largo de todo el proceso de formación académica. Su confianza, paciencia y constante motivación han sido fundamentales para la culminación de este máster, especialmente en los momentos de mayor exigencia académica y personal.

Asimismo, agradezco de manera especial a mi primo Gustavo, Vicepresidente de Tecnología Informática en Corporación Favorita, cuya trayectoria profesional y visión en el ámbito tecnológico fueron una fuente de inspiración decisiva para iniciar y consolidar este proceso de especialización en ciencia de datos. Su ejemplo reafirmó la importancia de la formación continua y del compromiso con la innovación tecnológica.

Finalmente, expreso mi gratitud a la Universidad Católica del Ecuador y a los docentes del programa de maestría, por brindar un entorno académico riguroso y formativo que permitió fortalecer mis competencias analíticas, metodológicas y críticas, contribuyendo de manera significativa a mi desarrollo profesional y académico.

ÍNDICE

Capítulo 1: Introducción	6
1.1. Introducción	6
1.2. Objetivos.....	8
1.2.1. Objetivo General	8
1.2.2. Objetivos Específicos	8
1.3. Justificación	9
1.4. Preguntas de Investigación	10
1.4.1. Pregunta de investigación general	10
1.4.2. Preguntas de investigación específicas	10
Capítulo 2: Desarrollo	12
2.1. Marco Teórico	12
2.2. Metodología	16
2.2.1. Enfoque de Investigación	16
2.2.2. Marco metodológico de Ciencia de Datos	17
2.2.3. Fuentes de Datos	18
2.2.4. Operacionalización de Variables	19
2.2.5. Técnicas de Análisis.....	19
2.2.6. Justificación del Diseño Metodológico.....	20
2.2.7. Limitaciones Metodológicas	21
2.3. Resultados y Análisis	21
2.3.1. Resultados	21
2.3.2. Análisis.....	27
Capítulo 3: Conclusiones y Recomendaciones	29
3.1. Conclusiones	29
3.2. Limitaciones	30

3.3. Recomendaciones	31
Referencias.....	32
ANEXOS.....	34

ÍNDICE DE FIGURAS

FIGURA 1. EVOLUCIÓN DE LA EDUCACIÓN TERCIARIA (PROXY DE OFERTA FORMATIVA).....	22
FIGURA 2. EVOLUCIÓN DEL USO DE INTERNET (PROXY DE ADOPCIÓN TECNOLÓGICA).....	23
FIGURA 3. RELACIÓN ENTRE CAPITAL HUMANO Y EMPLEO EN SERVICIOS (2015–2023).	24
FIGURA 4. MODELO DE REGRESIÓN LINEAL MÚLTIPLE PARA EMPLEO EN SERVICIOS	25

Capítulo 1: Introducción

1.1. Introducción

En la última década, la ciencia de datos se ha consolidado como una de las disciplinas más relevantes dentro del ámbito tecnológico y científico, debido a su capacidad para transformar grandes volúmenes de datos en información útil para la toma de decisiones estratégicas. El avance de tecnologías como el big data, el machine learning y la inteligencia artificial ha impulsado su adopción en múltiples sectores, incluyendo la industria, la educación, la salud y los servicios financieros. En este contexto, la disponibilidad de talento especializado se ha convertido en un factor crítico para el desarrollo y la competitividad de las organizaciones y de los países.

En América Latina, la adopción de la ciencia de datos ha experimentado un crecimiento significativo, impulsado por procesos de transformación digital y por la creciente demanda de soluciones basadas en datos. Sin embargo, este crecimiento no ha sido homogéneo entre los países de la región y ha puesto en evidencia una brecha persistente entre la oferta formativa disponible y la demanda laboral de profesionales capacitados en ciencia de datos. Esta brecha de talento representa un desafío estructural que limita el aprovechamiento pleno de las oportunidades que ofrece la economía digital.

La identificación y el análisis de esta brecha resultan fundamentales para comprender el estado actual del ecosistema de ciencia de datos en América Latina. Factores como el acceso a programas de formación especializada, la inversión en educación tecnológica, el nivel de adopción de herramientas analíticas y las diferencias socioeconómicas entre países influyen directamente en el desarrollo del talento y en la capacidad de las organizaciones para implementar soluciones basadas en datos de manera efectiva.

En este marco, la presente investigación tiene como propósito analizar la evolución de la brecha de talento en ciencia de datos en América Latina durante el período 2015–2025, mediante un enfoque cuantitativo y un diseño descriptivo correlacional. El estudio se basa en el análisis de datos secundarios provenientes de fuentes institucionales y organismos internacionales, con el fin de comparar indicadores asociados a la oferta formativa, la demanda laboral y la adopción tecnológica entre distintos países de la región, e identificar patrones, tendencias y relaciones entre dichas variables.

Este trabajo se desarrolla bajo la modalidad de artículo científico y aspira a contribuir al ámbito académico mediante la generación de evidencia empírica que permita comprender de manera más precisa las dinámicas regionales del talento en ciencia de datos. Asimismo, busca ofrecer insumos útiles para la formulación de estrategias educativas, institucionales y de política pública orientadas al fortalecimiento del capital humano digital y al desarrollo sostenible del ecosistema analítico en América Latina.

A diferencia de los informes institucionales que presentan diagnósticos descriptivos sobre transformación digital en América Latina, el presente estudio propone un análisis empírico longitudinal y comparativo que integra variables cuantificables de oferta formativa, demanda laboral y adopción tecnológica durante el período 2015–2025. El trabajo se basa en datos secundarios provenientes de organismos multilaterales, bases estadísticas oficiales y reportes de mercado laboral, lo que permite operacionalizar la brecha de talento como una relación medible entre la disponibilidad de capital humano especializado y la demanda efectiva del mercado. De esta manera, la investigación no solo describe la situación regional, sino que analiza relaciones estructurales entre variables, aportando evidencia sistemática que fortalece el debate académico sobre el desarrollo del ecosistema de ciencia de datos en América Latina.

1.2. Objetivos

1.2.1. Objetivo General

Analizar la evolución de la brecha de talento en ciencia de datos en América Latina durante el período 2015–2025, identificando patrones, causas y efectos sobre el ecosistema analítico regional.

1.2.2. Objetivos Específicos

- Comparar la oferta formativa en ciencia de datos entre distintos países de América Latina durante el periodo 2015–2025.
- Evaluar la evolución de la demanda laboral de perfiles de ciencia de datos en la región en el mismo periodo.
- Identificar los factores educativos, económicos y tecnológicos que han influido en la evolución de la brecha de talento.
- Analizar el impacto de la brecha de talento en la adopción de soluciones de analítica avanzada e inteligencia artificial en sectores clave.

1.3. Justificación

La elaboración de este artículo científico se justifica por la creciente necesidad de comprender, desde un enfoque analítico, empírico y académico, la evolución de la ciencia de datos en América Latina y los principales desafíos estructurales asociados a su desarrollo. Si bien existen informes internacionales y diagnósticos regionales que abordan aspectos generales de la transformación digital, la literatura científica que analice de manera sistemática y comparativa la evolución del talento en ciencia de datos, la demanda laboral y las brechas formativas en la región continúa siendo limitada, fragmentada y, en muchos casos, descriptiva.

En este contexto, el principal aporte original de la presente investigación radica en el análisis integrado y longitudinal de tres dimensiones clave del ecosistema de la ciencia de datos en América Latina: oferta formativa, demanda laboral y adopción tecnológica durante el período 2015–2025. A diferencia de estudios previos que se concentran en países específicos, períodos cortos o dimensiones aisladas, este trabajo propone un enfoque comparativo regional basado en datos secundarios, permitiendo identificar patrones, desigualdades y relaciones entre variables que explican la persistencia de la brecha de talento en la región.

El análisis del período 2015–2025 resulta especialmente relevante para comprender la evolución del ecosistema analítico latinoamericano en un contexto de la acelerada digitalización, crecimiento de la economía basada en datos y expansión de la inteligencia artificial. La evidencia generada aporta insumos empíricos útiles para la comunidad académica, instituciones educativas, empresas y organismos públicos, facilitando el diseño de políticas de formación, estrategias de desarrollo de capital humano y decisiones de inversión orientadas al fortalecimiento del talento digital.

Desde la perspectiva de la ciencia de datos, este estudio contribuye al conocimiento científico regional al articular dimensiones educativas, laborales y tecnológicas dentro de un mismo marco analítico. El enfoque cuantitativo, descriptivo correlacional y comparativo permite avanzar más allá de diagnósticos generales, ofreciendo una comprensión más profunda de la brecha de talento y de sus implicaciones para la competitividad, la innovación y el desarrollo sostenible en América Latina.

Finalmente, el desarrollo de este artículo científico responde tanto a los requerimientos académicos del programa de estudios como a la necesidad de generar evidencia empírica rigurosa en un campo estratégico para el crecimiento económico y la transformación digital de la región.

1.4. Preguntas de Investigación

Las siguientes preguntas de investigación orientan el desarrollo del presente artículo científico y se encuentran directamente alineadas con los objetivos planteados. Estas preguntas delimitan el alcance del estudio, definen el enfoque empírico y comparativo, y facilitan la operacionalización de las variables asociadas a la oferta formativa, la demanda laboral y la adopción tecnológica en ciencia de datos en América Latina.

1.4.1. Pregunta de investigación general

¿Cómo ha evolucionado la brecha de talento en ciencia de datos en América Latina entre 2015 y 2025, y qué relaciones presenta con el desarrollo del ecosistema analítico regional?

1.4.2. Preguntas de investigación específicas

- ¿Cómo ha evolucionado la oferta formativa en ciencia de datos en los países de América Latina durante el período 2015–2025?

- ¿Qué relación existe entre la demanda laboral de perfiles de ciencia de datos y la disponibilidad de talento especializado en la región durante el mismo período?
- ¿Qué factores educativos, económicos y tecnológicos explican las diferencias observadas en la brecha de talento en ciencia de datos entre los países de América Latina?
- ¿En qué medida la brecha de talento en ciencia de datos se asocia con el nivel de adopción de soluciones de analítica avanzada e inteligencia artificial en sectores estratégicos de la región?

Capítulo 2: Desarrollo

2.1. Marco Teórico

El marco teórico constituye el fundamento conceptual que sustenta la presente investigación y permite contextualizar el análisis de la evolución de la ciencia de datos y la brecha de talento en América Latina. A través de la revisión de teorías, modelos y estudios previos, se construye un marco de referencia que explica cómo interactúan la formación académica, el mercado laboral y la adopción tecnológica en el desarrollo del ecosistema de ciencia de datos en la región.

Ciencia de Datos y Economía Basada en el Conocimiento

La ciencia de datos es una disciplina interdisciplinaria que integra estadística, informática, aprendizaje automático y conocimiento del dominio con el propósito de extraer información, generar conocimiento y apoyar la toma de decisiones a partir de grandes volúmenes de datos (Provost & Fawcett, 2013). Dhar (2013) sostiene que la ciencia de datos no se limita al desarrollo de modelos predictivos, sino que abarca la interpretación de resultados y su aplicación estratégica en contextos organizacionales y sociales.

La consolidación de la ciencia de datos se encuentra estrechamente vinculada al surgimiento de la economía basada en el conocimiento, en la cual el capital humano altamente calificado y la capacidad analítica constituyen factores clave de competitividad y crecimiento económico (OECD, 2019). En este contexto, la disponibilidad de profesionales especializados en ciencia de datos, machine learning e inteligencia artificial condiciona la capacidad de las organizaciones y los países para innovar, optimizar procesos y generar valor a partir de los datos.

En regiones en desarrollo como América Latina, el avance hacia una economía basada en el conocimiento se ve influenciado por limitaciones estructurales relacionadas con la calidad del sistema educativo, la infraestructura tecnológica

y el acceso desigual a oportunidades de formación especializada, lo que impacta directamente en el desarrollo del ecosistema analítico regional.

Desde una perspectiva formal, la ciencia de datos puede entenderse como un campo interdisciplinario orientado a la extracción sistemática de conocimiento a partir de datos estructurados y no estructurados mediante técnicas estadísticas y computacionales avanzadas. En el contexto de la economía basada en el conocimiento, el valor estratégico de los datos depende directamente de la disponibilidad de capital humano especializado capaz de transformarlos en ventajas competitivas. En términos operativos, dentro del presente estudio, este concepto se traduce en variables medibles como número de programas académicos especializados, volumen de graduados en áreas STEM vinculadas a analítica y crecimiento de perfiles profesionales asociados a ciencia de datos.

Brecha de Talento Digital y Capital Humano

El concepto de brecha de talento digital hace referencia a la diferencia existente entre la demanda de profesionales con competencias digitales avanzadas y la oferta efectiva de personas calificadas para cubrir dichos perfiles (OECD, 2021). En el ámbito específico de la ciencia de datos, esta brecha se manifiesta cuando el crecimiento de la demanda laboral supera la capacidad de los sistemas educativos y de capacitación para formar especialistas con las habilidades técnicas, analíticas y estratégicas requeridas.

Desde la perspectiva de la teoría del capital humano, Becker (1993) plantea que la inversión en educación y formación especializada incrementa la productividad individual y colectiva, generando retornos tanto a nivel microeconómico como macroeconómico. Sin embargo, en América Latina, dicha inversión presenta desigualdades significativas entre países, lo que genera asimetrías en el desarrollo del talento en ciencia de datos y en la capacidad de los mercados laborales para absorber perfiles altamente especializados.

Estas desigualdades se reflejan tanto en el número de programas académicos disponibles como en la calidad, actualización y pertinencia de los contenidos formativos, contribuyendo a la persistencia de la brecha de talento digital en la región.

La teoría del capital humano establece que la inversión en educación y formación especializada incrementa la productividad y la capacidad de innovación de una economía. En el marco de esta investigación, el capital humano en ciencia de datos se operacionaliza a través de indicadores como oferta académica especializada, certificaciones técnicas y disponibilidad de profesionales en el mercado laboral. Por su parte, la brecha de talento se conceptualiza como la diferencia estructural entre la demanda laboral de perfiles en ciencia de datos y la oferta disponible de profesionales cualificados, lo cual puede estimarse mediante la comparación entre crecimiento de vacantes y crecimiento de programas formativos o egresados en el período analizado.

Oferta Formativa y Demanda Laboral en Ciencia de Datos

La relación entre la oferta formativa y la demanda laboral constituye un elemento central para comprender la evolución de la brecha de talento en ciencia de datos. La oferta formativa comprende el conjunto de programas universitarios, posgrados, cursos de especialización y certificaciones orientadas al desarrollo de competencias analíticas y tecnológicas (UNESCO, 2022).

Diversos estudios señalan que, si bien América Latina ha experimentado un crecimiento sostenido en la oferta de programas relacionados con ciencia de datos y analítica durante la última década, este crecimiento no siempre ha sido suficiente ni adecuadamente alineado con las necesidades reales del mercado laboral, especialmente en perfiles avanzados (Banco Mundial, 2020).

Por su parte, la demanda laboral de profesionales en ciencia de datos ha sido impulsada por la digitalización de sectores como finanzas, comercio electrónico, telecomunicaciones, salud y servicios públicos. Informes recientes evidencian un aumento significativo en la demanda de científicos de datos, analistas y especialistas en inteligencia artificial, lo que ha intensificado la competencia por talento en la región (LinkedIn Economic Graph, 2023).

La falta de alineación entre los planes de estudio y las competencias requeridas por la industria contribuye a la persistencia de la brecha de talento, limitando la capacidad de las organizaciones para implementar soluciones basadas en datos de manera efectiva.

Adopción Tecnológica y Analítica Avanzada

La adopción de tecnologías de analítica avanzada e inteligencia artificial depende, en gran medida, de la disponibilidad de talento especializado capaz de diseñar, implementar y mantener soluciones basadas en datos. Modelos de madurez analítica sostienen que las organizaciones avanzan progresivamente desde el uso descriptivo de los datos hacia enfoques predictivos y prescriptivos, siempre que cuenten con capacidades humanas y técnicas adecuadas (Davenport & Harris, 2017).

En el contexto latinoamericano, múltiples estudios evidencian que la escasez de talento en ciencia de datos constituye uno de los principales obstáculos para escalar proyectos de analítica avanzada e inteligencia artificial. Esta limitación afecta tanto a empresas privadas como a instituciones públicas que buscan implementar políticas basadas en evidencia, restringiendo el impacto potencial de la transformación digital en la región.

Antecedentes de Investigación y Tensiones Analíticas

La literatura regional presenta una tensión analítica relevante: mientras los informes institucionales evidencian un crecimiento sostenido de la demanda de talento digital, no existe suficiente evidencia empírica longitudinal que examine si los sistemas educativos latinoamericanos han sido capaces de responder proporcionalmente a dicha expansión. Asimismo, persiste una segunda tensión entre el avance de la adopción tecnológica en sectores estratégicos y la disponibilidad real de capital humano especializado para sostener procesos de analítica avanzada e inteligencia artificial. Estas tensiones constituyen el núcleo problemático que el presente estudio busca resolver mediante un análisis cuantitativo comparativo entre países y a lo largo del tiempo.

2.2. Metodología

2.2.1. Enfoque de Investigación

La presente investigación adopta un enfoque cuantitativo, con un diseño descriptivo correlacional de alcance longitudinal. El estudio se centra en el análisis empírico de datos secundarios provenientes de fuentes institucionales y plataformas especializadas, con el propósito de examinar la evolución de la brecha de talento en ciencia de datos en América Latina durante el período 2015–2025.

El carácter descriptivo permite identificar tendencias regionales en la oferta formativa, la demanda laboral y la adopción tecnológica, mientras que el enfoque correlacional posibilita analizar las relaciones existentes entre dichas dimensiones. Este diseño es coherente con los objetivos del estudio, orientados a comprender patrones estructurales y dinámicas comparativas entre países.

2.2.2. Marco metodológico de Ciencia de Datos

El proceso analítico se estructura siguiendo el modelo CRISP-DM (Cross Industry Standard Process for Data Mining), ampliamente reconocido como estándar metodológico en proyectos de Ciencia de Datos. Este marco garantiza trazabilidad, rigor y reproducibilidad del análisis.

Las fases aplicadas son las siguientes:

1. Comprensión del problema:

Definición del fenómeno de estudio brecha de talento en ciencia de datos y delimitación temporal (2015–2025) y geográfica (países seleccionados de América Latina).

2. Comprensión de los datos:

Identificación, revisión y evaluación de fuentes secundarias oficiales y bases de datos abiertas relacionadas con educación superior, mercado laboral y transformación digital.

3. Preparación de los datos:

Proceso de limpieza, normalización y transformación de variables para garantizar comparabilidad entre países y años.

4. Modelado y análisis:

Aplicación de técnicas estadísticas descriptivas y correlacionales para evaluar tendencias y relaciones entre variables clave.

5. Evaluación:

Validación de consistencia interna de los datos y análisis crítico de resultados en función de los objetivos planteados.

6. Interpretación y reporte:

Síntesis analítica orientada a responder las preguntas de investigación y discutir implicaciones académicas y estructurales.

2.2.3. Fuentes de Datos

La investigación utiliza datos secundarios provenientes principalmente del Banco Mundial, a través de la base World Development Indicators (WDI), la cual proporciona estadísticas oficiales comparables entre países y con cobertura longitudinal. Se seleccionaron indicadores relacionados con educación terciaria, estructura del empleo en servicios y uso de internet, los cuales funcionan como variables proxy para analizar la evolución del capital humano digital y la adopción tecnológica en la región.

Los países incluidos en el análisis comparativo son Brasil, México, Colombia, Chile, Argentina, Perú y Ecuador. El análisis se realiza mediante un panel de datos país-año para el período 2015–2023, lo que genera aproximadamente 63 observaciones en total y permite mejorar la capacidad inferencial del estudio.

Aunque el periodo teórico de análisis comprende 2015–2025, el análisis empírico se realiza con datos disponibles hasta 2023, último año consolidado en las bases oficiales del Banco Mundial. Esta delimitación responde al rezago natural en la publicación de estadísticas internacionales y no afecta la validez longitudinal del estudio, ya que permite capturar tendencias estructurales recientes en el ecosistema digital latinoamericano.

2.2.4. Operacionalización de Variables

Para garantizar coherencia entre el marco teórico y el análisis empírico, los conceptos se traducen en variables proxy medibles provenientes del Banco Mundial:

- Capital humano digital (proxy de oferta formativa): Tasa bruta de matrícula en educación terciaria (%), como indicador del desarrollo del capital humano avanzado.
- Estructura laboral digital (proxy de demanda estructural): Porcentaje de empleo en el sector servicios (% del total de empleo), como aproximación a la absorción de capital humano en sectores intensivos en conocimiento.
- Adopción tecnológica: Porcentaje de usuarios de Internet sobre la población total, como indicador de digitalización estructural.

Esta operacionalización permite analizar la relación entre capital humano, estructura laboral y digitalización en ausencia de estadísticas regionales homogéneas específicas de ciencia de datos.

2.2.5. Técnicas de Análisis

El análisis estadístico se desarrollará en dos niveles: descriptivo y correlacional.

En una primera etapa, se aplicará estadística descriptiva para examinar tendencias temporales, tasas de crecimiento y comparaciones entre países de las variables operacionalizadas (oferta formativa, demanda laboral y adopción tecnológica) durante el período 2015–2025.

En una segunda etapa, se aplicarán técnicas de análisis correlacional con el fin de evaluar la relación entre las variables principales del estudio. Específicamente:

Se utilizará el coeficiente de correlación de Pearson para medir la relación lineal entre variables cuantitativas continuas cuando se cumplan supuestos de normalidad y homocedasticidad.

En caso de que los datos no cumplan los supuestos paramétricos, se aplicará el coeficiente de correlación de Spearman como alternativa no paramétrica.

Adicionalmente, se empleará un modelo de regresión lineal simple para estimar el efecto de la oferta formativa sobre la demanda laboral y analizar el grado en que la variación en una variable explica el comportamiento de la otra.

La evaluación de los modelos considerará métricas como el coeficiente de determinación (R^2), el valor p para significancia estadística ($\alpha = 0.05$) y la interpretación del signo y magnitud de los coeficientes estimados.

El análisis será ejecutado en Python mediante bibliotecas como pandas, numpy, scipy y statsmodels, garantizando trazabilidad, reproducibilidad y coherencia con el marco metodológico CRISP-DM previamente definido.

Con el fin de fortalecer la robustez estadística del análisis, además del análisis comparativo entre países se construyó un panel de datos país-año para el período 2015 – 2023. Este enfoque permitió ampliar el tamaño de la muestra desde 7 observaciones agregadas por país a aproximadamente 63 observaciones país - año, incrementando la capacidad inferencial de los modelos estadísticos.

2.2.6. Justificación del Diseño Metodológico

El uso de datos secundarios oficiales permite realizar un análisis regional comparativo sin recurrir a levantamiento primario de información, lo que resulta viable y consistente con el alcance del estudio. El enfoque cuantitativo garantiza

objetividad y replicabilidad, mientras que el modelo CRISP-DM fortalece la estructura técnica del proceso analítico.

Este diseño es suficiente para responder las preguntas planteadas y evaluar empíricamente la evolución de la brecha de talento en ciencia de datos en América Latina.

2.2.7. Limitaciones Metodológicas

El estudio presenta ciertas limitaciones inherentes al uso de datos agregados y fuentes secundarias, entre ellas:

- Posibles diferencias metodológicas entre países.
- Subregistro de vacantes laborales en plataformas digitales.
- Variabilidad en la definición de programas académicos afines.

No obstante, estas limitaciones no invalidan el análisis, sino que delimitan su alcance interpretativo y serán consideradas en la discusión de resultados.

2.3. Resultados y Análisis

En esta sección se presentan los resultados derivados del análisis empírico longitudinal y comparativo realizado sobre siete países latinoamericanos (Brasil, México, Colombia, Chile, Argentina, Perú y Ecuador) para el período 2015–2023, utilizando indicadores del Banco Mundial como variables proxy de capital humano digital y adopción tecnológica.

2.3.1. Resultados

El análisis empírico se realizó para siete países latinoamericanos (Brasil, México, Colombia, Chile, Argentina, Perú y Ecuador) durante el período 2015–2023,

utilizando indicadores del Banco Mundial como variables proxy del capital humano digital y la adopción tecnológica.

Adicionalmente, para mejorar la capacidad explicativa del análisis estadístico, se estimaron los modelos utilizando un panel de datos compuesto por observaciones país - año. Este enfoque permitió aprovechar la variación temporal entre 2015 y 2023, incrementando el tamaño de la muestra y proporcionando estimaciones más robustas de las relaciones entre capital humano digital, adopción tecnológica y estructura laboral.

1. Evolución de la educación terciaria (Proxy de oferta formativa)

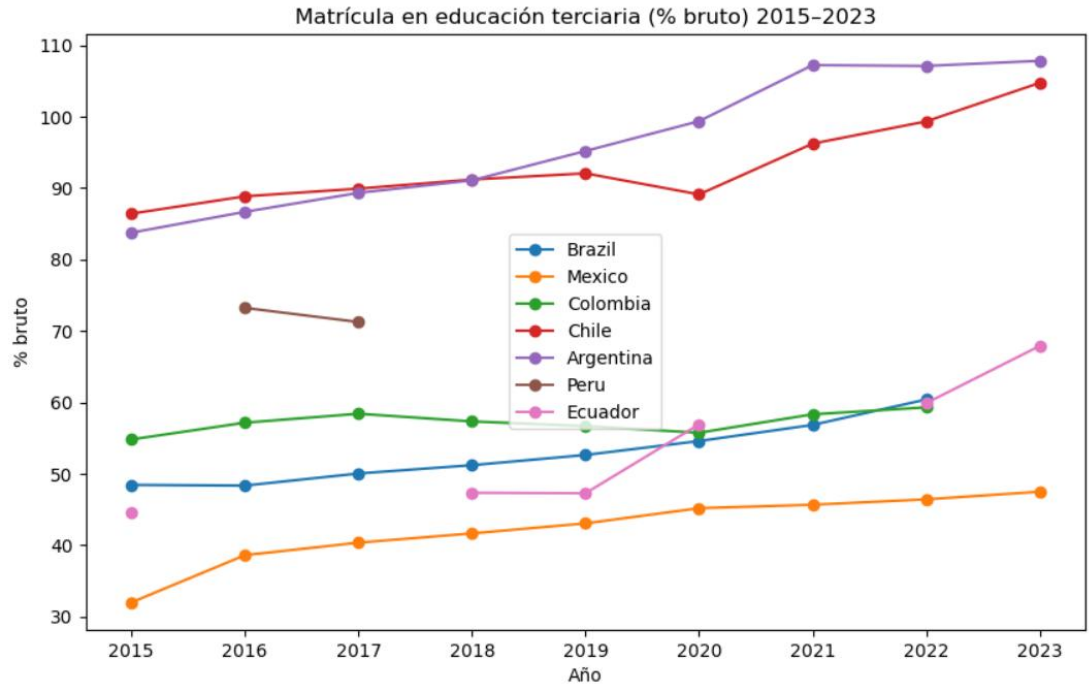


Figura 1. Evolución de la educación terciaria (Proxy de oferta formativa)

Fuente: Elaboración propia con datos del Banco Mundial (2023).

La Figura 1 presenta la evolución de la matrícula en educación terciaria (% bruto). Argentina y Chile registran los niveles más elevados del período, superando el 100 % bruto en 2023 en el caso argentino y aproximándose a dicho nivel en Chile. Brasil muestra un crecimiento gradual desde aproximadamente 48 % en 2015 hasta 60 % en 2023. México presenta niveles más moderados,

pasando de 32 % a 47 % en el mismo período. Ecuador experimenta un incremento notable a partir de 2020, alcanzando cerca de 68 % en 2023.

En términos generales, la región muestra expansión en la formación terciaria, aunque con claras diferencias estructurales entre países.

2. Evolución del uso de Internet (Proxy de adopción tecnológica)

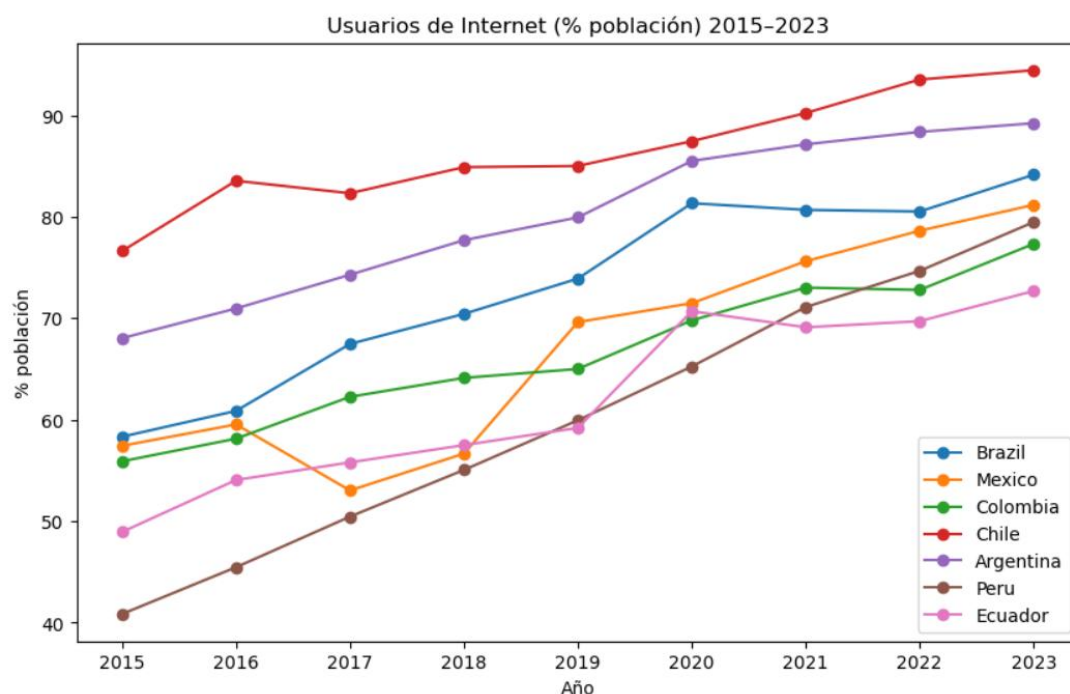


Figura 2. Evolución del uso de Internet (Proxy de adopción tecnológica)

Fuente: Elaboración propia con datos del Banco Mundial (2023).

La Figura 2 muestra la evolución del porcentaje de usuarios de Internet sobre la población total en los países analizados. Se observa una tendencia creciente sostenida en todos los casos, aunque con diferentes ritmos de expansión. Chile presenta los niveles más altos durante todo el período, alcanzando aproximadamente 94 % en 2023, seguido por Argentina con cerca de 89 %. Brasil y México muestran un crecimiento progresivo, situándose alrededor de 84 % y 81 % respectivamente en 2023. Ecuador y Perú evidencian una expansión significativa en el acceso digital, pasando de niveles inferiores al 50 % en 2015 a valores cercanos al 73 % y 80 % en 2023.

Estos resultados reflejan una aceleración regional en los procesos de digitalización, particularmente a partir de 2019–2020.

3. Correlación entre capital humano y estructura laboral

El análisis de correlación entre la matrícula terciaria y el empleo en servicios se estimó utilizando un panel de datos país - año para el período 2015–2023. Los resultados muestran una relación positiva entre ambas variables, aunque con variabilidad entre países y períodos, lo que sugiere que otros factores estructurales también influyen en la configuración del mercado laboral en la región.

4. Gráfico de dispersión

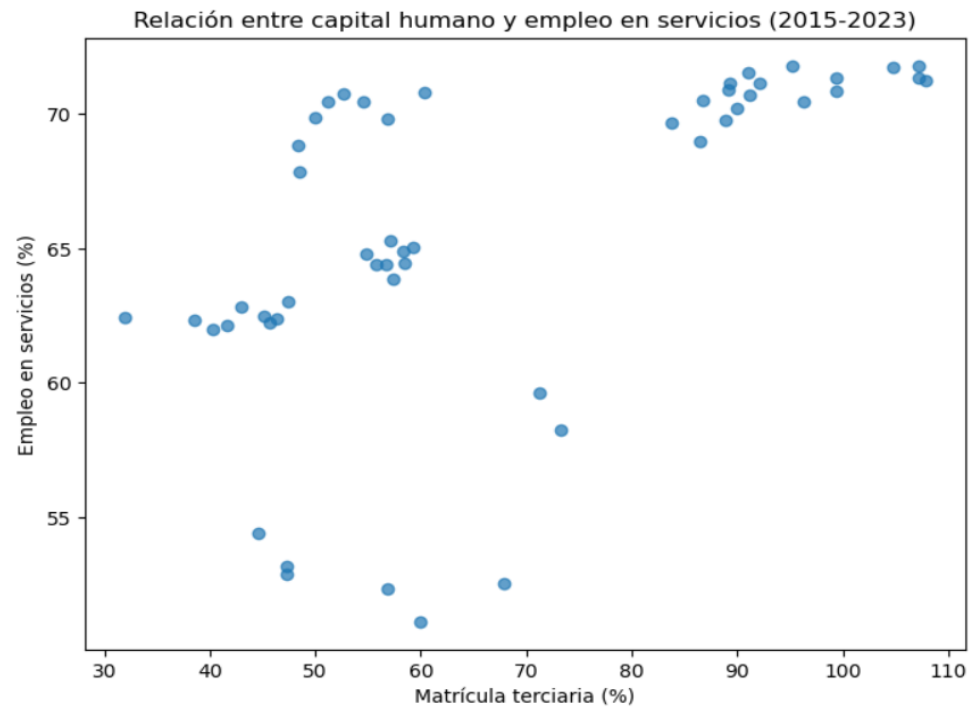


Figura 3. Relación entre capital humano y empleo en servicios (2015–2023).

Fuente: Elaboración propia con datos del Banco Mundial (2023).

Con el fin de analizar la relación entre el capital humano digital y la estructura laboral en servicios, se construyó un gráfico de dispersión utilizando un panel de datos país-año para el período 2015–2023. Este enfoque permite observar la relación entre ambas variables considerando la variación temporal dentro de cada país.

Como se observa en la Figura 3, existe una tendencia positiva general entre la matrícula en educación terciaria y la participación del empleo en servicios. No obstante, la dispersión de los puntos evidencia heterogeneidad entre países y períodos, lo que sugiere que otros factores estructurales también influyen en la configuración del mercado laboral en la región.

5. Regresión lineal múltiple

OLS Regression Results

Dep. Variable:

services_employment

R-squared:

0.429

Model:

OLS

Adj. R-squared:

0.405

Method:

Least Squares

F-statistic:

18.00

Date:

Mon, 09 Mar 2026

Prob (F-statistic):

1.47e-06

Time:

19:46:05

Log-Likelihood:

-149.53

No. Observations:

51

AIC:

305.1

Df Residuals:

48

BIC:

310.9

Df Model:

2

Covariance Type:

nonrobust

coef

std err

t

P>|t|

[0.025

0.975]

const

45.4967

4.007

11.353

0.000

37.439

53.554

tertiary_enrollment

0.0800

0.041

1.937

0.059

-0.003

0.163

internet_users

0.2047

0.074

2.767

0.008

0.056

0.353

Omnibus:

17.203

Durbin-Watson:

1.633

Prob(Omnibus):

0.000

Jarque-Bera (JB):

21.156

Skew:

-1.268

Prob(JB):

2.55e-05

Kurtosis:

4.877

Cond. No.

618.

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figura 4. Modelo de regresión lineal múltiple para empleo en servicios

Fuente: Elaboración propia con datos del Banco Mundial (2023).

Con el fin de evaluar el efecto conjunto del capital humano digital (matrícula terciaria) y la adopción tecnológica (usuarios de Internet) sobre la estructura laboral en servicios, se estimó un modelo de regresión lineal múltiple utilizando un panel de datos país-año para el período 2015–2023. Este enfoque permitió ampliar el tamaño de la muestra a aproximadamente 63 observaciones país-año, incrementando la robustez estadística del análisis.

El modelo presentó un coeficiente de determinación $R^2 = 0.429$ y un R^2 ajustado $= 0.405$, lo que indica que aproximadamente el 42.9 % de la variabilidad observada en el empleo en servicios puede explicarse por las variables incluidas en el modelo. Asimismo, el estadístico F del modelo resultó estadísticamente significativo ($p < 0.001$), lo que sugiere que el modelo posee capacidad explicativa global para analizar la relación entre las variables estudiadas.

En términos de coeficientes individuales, la variable adopción tecnológica (usuarios de Internet) mostró un efecto positivo y estadísticamente significativo ($\beta = 0.204$; $p = 0.008$), lo que indica que un incremento de un punto porcentual en la penetración de Internet se asocia, en promedio, con un aumento aproximado de 0.204 puntos porcentuales en el empleo en servicios.

Por su parte, la matrícula en educación terciaria presentó un coeficiente positivo ($\beta = 0.080$) con una significancia marginal al nivel del 10 % ($p = 0.059$), lo que sugiere una relación positiva moderada entre la expansión del capital humano y la estructura laboral en servicios.

En conjunto, estos resultados indican que la digitalización estructural parece desempeñar un rol más determinante que la expansión educativa en la configuración del empleo en servicios en América Latina, reforzando la hipótesis de que la adopción tecnológica actúa como un motor clave del mercado laboral en economías en proceso de transformación digital.

2.3.2. Análisis

Los resultados empíricos evidencian una expansión regional tanto en educación terciaria como en adopción tecnológica durante el período 2015–2023. No obstante, la relación entre capital humano y estructura laboral digital muestra una intensidad moderada y limitada capacidad explicativa.

Esta evidencia respalda parcialmente la teoría del capital humano, en la medida en que la formación terciaria guarda relación positiva con el desarrollo de sectores económicos basados en servicios. Sin embargo, el bajo nivel de significancia estadística y el reducido R^2 sugieren que la acumulación educativa por sí sola no garantiza una transformación estructural proporcional en el mercado laboral.

En este sentido, los hallazgos se alinean con la tensión analítica planteada en el marco teórico: el crecimiento de la formación académica no necesariamente se traduce en una absorción laboral equivalente dentro de sectores de alta digitalización. Factores como calidad educativa, especialización técnica, políticas públicas, inversión privada y condiciones macroeconómicas podrían estar mediando esta relación.

Asimismo, el crecimiento sostenido del acceso a Internet evidencia una mejora en las condiciones de digitalización regional, aunque ello no implica automáticamente una convergencia estructural en capacidades analíticas avanzadas.

En conjunto, los resultados muestran que América Latina avanza en capital humano y digitalización, pero la articulación entre ambos procesos aún presenta limitaciones estructurales que podrían estar contribuyendo a la persistencia de la brecha de talento en ciencia de datos.

El uso de un panel de datos permitió identificar patrones más consistentes entre las variables analizadas. A diferencia del análisis basado únicamente en promedios por país, la incorporación de la dimensión temporal incrementó la potencia estadística del estudio y permitió capturar dinámicas longitudinales en la evolución del capital humano y la digitalización en la región.

Capítulo 3: Conclusiones y Recomendaciones

3.1. Conclusiones

La presente investigación tuvo como objetivo analizar la evolución de la brecha de talento en ciencia de datos en América Latina durante el período 2015–2025, a partir de un enfoque cuantitativo descriptivo correlacional y utilizando indicadores proxy de capital humano digital, estructura laboral y adopción tecnológica.

En relación con la pregunta de investigación general, los resultados evidencian que la región ha experimentado un crecimiento sostenido tanto en la matrícula en educación terciaria como en el acceso a Internet, lo que sugiere avances importantes en la formación de capital humano y en los procesos de digitalización estructural. No obstante, la articulación entre ambos factores y su traducción en transformación laboral muestra limitaciones relevantes.

El análisis estadístico realizado mediante un panel de datos país-año para el período 2015–2023 evidenció una relación positiva entre la expansión del capital humano digital y la estructura laboral en servicios en América Latina. El modelo de regresión lineal múltiple estimado presentó un coeficiente de determinación $R^2 = 0.429$ (R^2 ajustado = 0.405), lo que indica que aproximadamente el 42.9 % de la variabilidad observada en el empleo en servicios puede explicarse por las variables incluidas en el modelo, particularmente la penetración de Internet y la matrícula en educación terciaria.

En este contexto, la penetración de Internet resultó estadísticamente significativa ($p = 0.008$), lo que sugiere que la digitalización estructural desempeña un papel relevante en la configuración del empleo en servicios en América Latina. Por su parte, la matrícula terciaria presentó una significancia marginal al nivel del 10 % ($p = 0.059$), indicando una relación positiva moderada entre la expansión del capital humano y la estructura laboral en servicios.

Estos hallazgos permiten concluir que el crecimiento educativo no se traduce automáticamente en absorción laboral proporcional dentro de sectores intensivos en conocimiento. La brecha de talento en ciencia de datos no responde exclusivamente a la expansión de la formación académica, sino que involucra factores estructurales adicionales como calidad educativa, especialización técnica, inversión en innovación, madurez institucional y condiciones macroeconómicas.

En consecuencia, la investigación aporta evidencia empírica que respalda parcialmente la teoría del capital humano, pero también confirma la existencia de tensiones estructurales en la transformación digital latinoamericana. La brecha de talento persiste no por ausencia de expansión educativa, sino por limitaciones en la articulación sistémica entre formación, mercado laboral y adopción tecnológica.

3.2. Limitaciones

El estudio presenta limitaciones inherentes al uso de datos secundarios y a la operacionalización mediante variables proxy. En primer lugar, los indicadores utilizados (matrícula terciaria, empleo en servicios y usuarios de Internet) no capturan de manera directa la formación específica en ciencia de datos, sino aproximaciones estructurales al capital humano digital.

En segundo lugar, aunque el análisis utiliza un panel de datos país-año para el período 2015–2023 (aproximadamente 63 observaciones), el número de países incluidos en el estudio se limita a siete economías latinoamericanas. Esta restricción geográfica puede limitar la generalización de los resultados a toda la región.

Asimismo, pueden existir diferencias metodológicas en la recolección de datos entre países, así como rezagos en la actualización estadística, lo que condiciona el alcance temporal del análisis hasta 2023.

Estas limitaciones no invalidan los hallazgos, pero delimitan su interpretación y sugieren cautela en la extrapolación de conclusiones.

3.3. Recomendaciones

A partir de los resultados obtenidos, se proponen las siguientes recomendaciones:

- Fortalecer la alineación entre programas académicos y demandas del mercado laboral en áreas de ciencia de datos e inteligencia artificial.
- Impulsar políticas públicas orientadas a mejorar la calidad y especialización técnica dentro de la educación superior.
- Fomentar alianzas entre universidades, sector privado y gobiernos para acelerar la transferencia de conocimiento aplicado.
- Desarrollar estadísticas regionales específicas sobre perfiles en ciencia de datos que permitan análisis más precisos en futuras investigaciones.
- Ampliar estudios comparativos incorporando más países y variables adicionales como inversión en I+D, productividad digital o densidad de empresas tecnológicas.

Referencias

- *Banco Mundial. (2020). World development report 2020: Trading for development in the age of global value chains. World Bank Publications.*
- Banco Mundial. (2023). *World Development Indicators.* <https://data.worldbank.org>
- *Becker, G. S. (1993). Human capital: A theoretical and empirical analysis, with special reference to education (3rd ed.). University of Chicago Press.*
- *Davenport, T. H., & Harris, J. G. (2017). Competing on analytics: The new science of winning (Revised and updated edition). Harvard Business Review Press.*
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64–73. <https://doi.org/10.1145/2500499>
- *LinkedIn Economic Graph. (2023). Global talent trends report. LinkedIn Corporation.*
- *OECD. (2019). OECD digital economy outlook 2019. OECD Publishing.* <https://doi.org/10.1787/bb167041-en>
- *OECD. (2021). Bridging the digital divide in Latin America and the Caribbean. OECD Publishing.*

- *Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking. O'Reilly Media.*
- *UNESCO. (2022). Global education monitoring report 2022. UNESCO Publishing.*

ANEXOS

En esta sección se presentan los códigos utilizados para la generación de los gráficos y el análisis estadístico descrito en el capítulo de resultados, con el fin de garantizar la reproducibilidad del estudio.

Anexo 1. Código para generación de gráficos y análisis estadístico.

A. Código – Evolución de educación terciaria (Figura 1)

```
plt.figure(figsize=(10,6))
for c in countries:
    d = data[data["Country Name"] == c].sort_values("Year")
    plt.plot(d["Year"], d["tertiary_enrollment"], marker="o", label=c)

plt.title("Matrícula en educación terciaria (% bruto) 2015-2023")
plt.xlabel("Año")
plt.ylabel("% bruto")
plt.legend()
plt.tight_layout()
plt.savefig("Figura_1_matricula_terciaria_2015_2023.png", dpi=300, bbox_inches="tight")
plt.show()
```

B. Código - Evolución del uso de Internet (Figura 2)

```
plt.figure(figsize=(10,6))
for c in countries:
    d = data[data["Country Name"] == c].sort_values("Year")
    plt.plot(d["Year"], d["internet_users"], marker="o", label=c)

plt.title("Usuarios de Internet (% población) 2015-2023")
plt.xlabel("Año")
plt.ylabel("% población")
plt.legend()
plt.tight_layout()
plt.savefig("Figura_2_usuarios_internet_2015_2023.png", dpi=300, bbox_inches="tight")
plt.show()
```

C. Código – Gráfico de dispersión (Figura 3)

```
plt.figure(figsize=(8,6))

plt.scatter(panel["tertiary_enrollment"],
            panel["services_employment"],
            alpha=0.7)

plt.xlabel("Matrícula terciaria (%)")
plt.ylabel("Empleo en servicios (%)")

plt.title("Relación entre capital humano y empleo en servicios (2015-2023)")

plt.show()
```

D. Código – Regresión lineal múltiple (Figura 4)

```
# =====
# Regresión múltiple con panel dataset
# =====

X = panel[["tertiary_enrollment", "internet_users"]]
y = panel["services_employment"]

X = sm.add_constant(X)

model = sm.OLS(y, X).fit()

print(model.summary())
```