

Análisis de la huella digital en América Latina y el Caribe

Enseñanzas extraídas del uso de macrodatos
(*big data*) para evaluar la economía digital



NACIONES UNIDAS

CEPAL



BigDATA

Economía digital para
América Latina y el Caribe

Gracias por su interés en esta publicación de la CEPAL



Si desea recibir información oportuna sobre nuestros productos editoriales y actividades, le invitamos a registrarse. Podrá definir sus áreas de interés y acceder a nuestros productos en otros formatos.



www.cepal.org/es/publications



www.cepal.org/apps

Análisis de la huella digital en América Latina y el Caribe

Enseñanzas extraídas del uso de macrodatos
(*big data*) para evaluar la economía digital



CEPAL



BigDATA

Economía digital para
América Latina y el Caribe

Este informe fue preparado bajo la orientación general de Mario Cimoli, Secretario Ejecutivo Adjunto de la Comisión Económica para América Latina y el Caribe (CEPAL), y con la coordinación de Sebastián Rovira, Valeria Jordán y Jorge Alejandro Patiño, de la División de Desarrollo Productivo y Empresarial. El ejercicio sobre macrodatos (*big data*) estuvo a cargo de un equipo de expertos conformado por Veronika Vilgis, Yu-Chang Ho, Xin Jin, Kangbo Lu, Karla Rascon-García y Matthew Reese, coordinados por Martin Hilbert. Se agradecen los valiosos aportes de Mario Castillo, Wilson Peres, Fernando Rojas y Nunzia Saporito, de la CEPAL.

Esta publicación fue producida en el marco de las actividades desarrolladas como parte del proyecto “Grandes datos para la economía digital en América Latina y el Caribe”, con el apoyo del décimo tramo de la Cuenta de las Naciones Unidas para el Desarrollo.

Las opiniones expresadas en este documento, que es traducción de un documento original no editado formalmente, son de exclusiva responsabilidad de los autores y pueden no coincidir con las de la Organización.

Publicación de las Naciones Unidas
LC/TS.2020/12
Distribución: L
Copyright © Naciones Unidas, 2020
Todos los derechos reservados
Impreso en Naciones Unidas, Santiago
S.19-01129

Esta publicación debe citarse como: Comisión Económica para América Latina y el Caribe (CEPAL), *Documentos de Proyectos* “Análisis de la huella digital en América Latina y el Caribe: enseñanzas extraídas del uso de macrodatos (*big data*) para evaluar la economía digital” (LC/TS.2020/12), Santiago, 2020.

La autorización para reproducir total o parcialmente esta obra debe solicitarse a la Comisión Económica para América Latina y el Caribe (CEPAL), División de Publicaciones y Servicios Web, publicaciones.cepal@un.org. Los Estados Miembros de las Naciones Unidas y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Solo se les solicita que mencionen la fuente e informen a la CEPAL de tal reproducción.

Resumen	7
Introducción	9
A. Macrodatos para la formulación de políticas con base empírica	9
B. Fundamentos metodológicos de la ciencia de datos	10
1. Recolección de datos: herramientas y consideraciones jurídicas	12
2. Procesamiento de datos	17
3. Análisis y visualización de los datos	18
C. Ámbitos de interés y fuentes de datos	20
I. Mercado de trabajo y habilidades digitales	21
A. La economía de las ocupaciones transitorias de los trabajadores autónomos	22
1. Oferta de categorías de empleo por país	22
2. Demanda mundial y oferta nacional	23
3. Remuneración horaria por país y por categoría de empleo	25
4. Cuestiones de género en la economía digital	28
B. Demanda de trabajadores a tiempo completo	31
1. Demanda de trabajadores a tiempo completo	32
2. Análisis de las descripciones de los requisitos de empleo	33
II. Asequibilidad de las tecnologías digitales	35
A. Índice de precios de los productos tecnológicos	35
B. Precios por paquete de productos	38
III. Microempresas y pequeñas y medianas empresas: acceso a los mercados y sistemas de financiación colectiva	39
A. Minoristas en línea	39
1. Localización geográfica	41
2. Concentración del mercado	41
3. Acceso a las transacciones	42
4. Facilitadores de las transacciones	43
B. Financiación colectiva	44
1. Recaudación de fondos por sector	44
2. Recaudación de fondos en el contexto mundial	45
3. Distribución por sexo	47

IV. Velocidad de la banda ancha en la región y en el mundo	49
V. Criptomonedas	51
VI. La huella digital de los medios sociales	53
A. Aspectos sociodemográficos	53
1. Propiedad de las empresas	55
2. Acceso a las redes	56
B. Tendencias	56
1. Objetivos de Desarrollo Sostenible (ODS)	56
2. Número de tuits	58
VII. Enseñanzas extraídas	61
A. Resumen de los beneficios y los desafíos	61
1. Nuevas percepciones	61
2. Desafíos que perduran	62
B. Aspectos que merecen especial atención	63
1. La importancia de las fuentes	63
2. La importancia de los indicadores y los índices	64
3. La importancia de la armonización	64
4. El conocimiento sobre los ámbitos es importante	65
Bibliografía	67

Cuadros

Cuadro 1	Temas de economía digital, fuentes web y número de observaciones, enero a marzo de 2019	20
Cuadro 2	Sistema de armonización de las categorías laborales registradas para la economía digital	21
Cuadro 3	Intervalos de confianza para la predicción del género a partir de los nombres [-1 masculino, +1 femenino]; la diferencia para los intervalos de confianza va de 0,1 a 0,2 , febrero de 2019	29
Cuadro 4	Recopilación de datos para conformar la muestra de los vendedores en MercadoLibre	40

Gráficos

Gráfico 1	Representación esquemática del flujo de trabajo de la ciencia de datos con datos sobre la huella digital, desde la etapa de recopilación de datos hasta las fases de procesamiento y visualización	11
Gráfico 2	Ejemplos de plataformas transnacionales en línea en América Latina y el Caribe	14
Gráfico 3	Captura de pantalla de ejemplo de un panel que muestra el monto promedio de 1.151.384 prés-tamos de financiación cooperativa otorgados a través de Kiva.org	19
Gráfico 4	Proporciones de categorías laborales profesionales por país, febrero de 2019	23
Gráfico 5	Demanda mundial de empleos y oferta nacional, por categoría de empleo	24
Gráfico 6	Tarifa horaria media (precio solicitado) de trabajadores autónomos por país y por categoría laboral, máximo de 100 dólares, febrero de 2019	26
Gráfico 7	Tarifa media por hora (máximo de 1.000 dólares) en Freelancer.com e ingreso nacional bruto (INB) per cápita, febrero de 2019	27
Gráfico 8	Participación de los trabajadores autónomos en las ganancias por país de origen (N = 6.751), febrero de 2019	28
Gráfico 9	Proporción de mujeres profesionales, por plataforma y por país, febrero de 2019	30
Gráfico 10	Proporción de mujeres profesionales, por categoría de empleo y plataforma, febrero de 2019	30
Gráfico 11	Demanda mundial de empleo y oferta regional, por género	31
Gráfico 12	Mapa de calor de la demanda mundial de empleo y la oferta regional, por género y categoría de empleo, febrero de 2019	31
Gráfico 13	Empleo por categoría de trabajo de la economía digital en Bumeran, febrero de 2019	32
Gráfico 14	Análisis semántico de las ofertas de empleo de Bumeran publicadas en Chile y México (labo-rum.cl, N = 1.907 publicaciones; bumeran.com.mx, N = 1.640 publicaciones): los ejes representan posiciones (los números altos son posiciones bajas)	34
Gráfico 15	Índice de precios de los productos tecnológicos por país, febrero de 2019	36
Gráfico 16	Precios medios de los productos tecnológicos por sector (eje horizontal x) frente a los precios medios normalizados según el INB per cápita ajustado por el PPA, febrero de 2019	37
Gráfico 17	Precio del paquete por país seleccionado, febrero de 2019	38
Gráfico 18	Concentración de mercado de vendedores en MercadoLibre: representación logarítmica de la cuota de mercado de los principales cinco vendedores en función del total histórico de productos vendidos (barras), número de vendedores por país (eje derecho), total histórico (en porcentaje [barras] y recuentos [eje derecho])	42
Gráfico 19	Penetración de Internet (ITU, 2017) frente al número promedio de transacciones por vendedor por país: totales históricos	43
Gráfico 20	Opciones de pago disponibles en MercadoLibre por país, enero de 2019	43
Gráfico 21	Proporciones de proyectos por sector en Kiva.org. (2006-2018)	44
Gráfico 22	Monto total de financiación recibida, 2006-2019	45
Gráfico 23	Porcentaje de proyectos liderados por mujeres en Kiva 2006-2018	48
Gráfico 24	Velocidades mundiales de banda ancha por tipo, 2008-2019	49
Gráfico 25	Velocidades de descarga en redes móviles y fijas por región mundial, 2007-2019	50
Gráfico 26	Volumen anual de compras de bitcoins por región mundial. 2013-2018	51

Gráfico 27	Penetración de Facebook	53
Gráfico 28	Estimaciones de alcance potencial de Facebook para 164 países frente al número de usuarios de Internet (ITU, 2018)	54
Gráfico 29	Porcentaje de mujeres propietarias de empresas por grupo etario, en marzo de 2019	55
Gráfico 30	Acceso a Facebook según el tipo de red de acceso por país, marzo de 2019	56
Gráfico 31	Tuits compartidos, febrero y marzo de 2019	58

Mapa

Mapa 1	Distribución geográfica de vendedores en MercadoLibre	41
--------	---	----

Imagen

Image 1	Objetivos de Desarrollo Sostenible (ODS)	57
---------	--	----

En este informe se analizan las oportunidades y los desafíos que entraña el uso sistemático de los datos digitales a disposición del público como herramienta para formular políticas públicas que impulsen el desarrollo de la economía digital en América Latina y el Caribe. El objetivo es compartir las enseñanzas adquiridas en la realización de ejercicios prácticos con técnicas de grandes datos (*big data*) a fin de avanzar en una agenda de investigación que permita a los países de la región diseñar herramientas alternativas de medición basadas en la huella digital.

Las principales conclusiones se derivan del contenido de sus seis capítulos, en los cuales se analiza la situación actual de algunos aspectos complementarios de la economía digital en América Latina y el Caribe, a saber: el mercado de trabajo y las habilidades digitales; los precios de la tecnología; las micro, pequeñas y medianas empresas; los servicios de banda ancha; las criptomonedas, y las redes sociales. El informe muestra que la huella digital que dejan los portales de búsqueda de empleo, las plataformas de comercio electrónico y las redes sociales permite obtener información de un alcance y un nivel de detalle sin precedentes. Por ejemplo, en lo referido al alcance, en general se ha podido reunir datos naturalmente armonizados para más de 15 países de América Latina y el Caribe; se ha seguido la trayectoria de unas 2,5 millones de pequeñas empresas de la región y se han recopilado 35 millones de mensajes relacionados con los Objetivos de Desarrollo Sostenible (ODS). En lo relacionado con el nivel de detalle, se ha podido ir más allá de los datos que surgen de las estadísticas oficiales y recabar datos sobre el género de los propietarios de las pequeñas empresas en el Caribe, o la remuneración por hora que efectivamente se paga por una habilidad específica en función del género. Más allá de la información consignada en los registros administrativos oficiales, ha sido posible distinguir entre el uso activo de las redes 3G y las 4G y determinar el precio de las tecnologías emergentes, como los vehículos aéreos no tripulados (drones). El informe incluye alrededor de 30 gráficos que ilustran las oportunidades que se analizan.

Las contribuciones metodológicas del informe se relacionan con las enseñanzas que se extrajeron de la realización de ejercicios de analítica de grandes datos. En este sentido, el informe es una herramienta de orientación general para los profesionales interesados en valerse de la ciencia de datos moderna para diseñar políticas de desarrollo. Al pensar en el paradigma de los macrodatos (*big data*), es común la noción ingenua de que si se tienen las destrezas informáticas necesarias, alcanza con recabar datos en Internet para que el universo entero de la realidad se materialice instantáneamente frente a cualquier observador, en tiempo real y con todos los detalles. En realidad, el proceso se parece más a lo que sucede en la historia del grupo de ciegos que inspecciona un elefante, en el sentido de que los expertos en ciencia de datos hace las veces de ciegos que tocan partes muy distintas del conjunto y, sin poder abarcarlo en su totalidad, intentan armonizar un universo de datos irreconciliables. Lejos de entrañar desafíos en materia de cálculo únicamente, entender correctamente los macrodatos va mucho más allá y se relaciona con la definición de la ciencia de datos como una disciplina en la que convergen la ciencia informática, la estadística y su ámbito sustantivo de aplicación. En la práctica, las cuestiones relativas a la representatividad, la generalización, la armonización, la calidad de los datos y la definición de variables e índices rápidamente se transforman en los principales ámbitos de interés de la ciencia de datos. En este informe se presentan algunas de las enseñanzas extraídas a lo largo de este ejercicio, que se desarrolló entre enero y marzo de 2019, y se resumen en el capítulo final.

A. Macrodatos para la formulación de políticas con base empírica

Las propuestas de políticas que no se basan en pruebas empíricas tienen mucho menos probabilidades de éxito que las políticas en cuyo diseño se tiene en cuenta el estado actual de la realidad. Si bien no es lo único que se necesita, las pruebas empíricas son un requisito necesario para garantizar el éxito de las intervenciones en políticas de desarrollo. Para saber a dónde desean dirigirse y cómo llegar allí, los encargados de la formulación de políticas deben comenzar a modelar su visión en función de la situación en la que se encuentran. Si no se conocen los mares donde se navega, elegir los vientos favorables para llegar al destino deseado se transforma en un juego de azar. La falta de conocimiento sólido de las realidades existentes contribuye a la situación frecuentemente lamentada de que numerosas políticas públicas son ineficaces y no logran los impactos deseados. Para tener impactos duraderos, es fundamental que las políticas se basen firmemente en realidades empíricas.

Desafortunadamente, las regiones en desarrollo, como América Latina y el Caribe, tienen enormes carencias en materia de datos empíricos, en particular en los ámbitos nuevos y de reciente surgimiento, como la economía digital. Por un lado, la economía digital ya ha comenzado a dominar la economía mundial. Al momento de escribir este informe (principios de 2019), siete de las ocho empresas más valiosas del mundo pertenecen al sector de la economía digital: Amazon, Microsoft, Alphabet (Google), Apple, Facebook, Tencent y Alibaba. La capitalización de mercado de cada una de estas empresas supera a la de gigantes industriales como Johnson & Johnson o bancos como JPMorgan Chase (posiciones 9 y 10). En ese contexto, la CEPAL ha propuesto impulsar políticas de reforma estructural para fomentar un desarrollo equitativo y sostenible que aproveche a pleno el potencial de la economía digital, con vistas a promover un cambio en la estructura productiva de la región hacia industrias con un uso más intensivo de conocimiento y sectores de mayor productividad (CEPAL, 2018). Por otro lado, la inercia de los sistemas estadísticos actuales no se condice con el rápido avance de la economía digital. Cuando se comparan con otras esferas de investigación estadística y académica más tradicionales, estos ámbitos incipientes de la agenda de desarrollo aún no concitan el nivel de atención necesario.

Al mismo tiempo, es en estos ámbitos nuevos, emergentes y de rápida evolución donde los datos empíricos son especialmente útiles. Esto se debe a que en estos ámbitos innovadores, las recomendaciones de políticas a menudo siguen sin examinarse de manera integral. Sin basarse con solidez en las realidades empíricas, los responsables de la formulación de políticas se enfrentan a la doble incertidumbre de no saber cuál es el punto de partida y de verse obligados a aplicar una estrategia de políticas cuya eficacia no está plenamente demostrada. Las carencias en materia de opciones de políticas y evaluaciones, aunadas a la falta de conocimiento sobre las realidades actuales, son un factor que reduce considerablemente las probabilidades de éxito de las políticas públicas. Si no se sabe con certeza cuáles son los vientos más favorables, ignorar el destino del viaje puede convertirse rápidamente en un problema grave que nos obliga a retroceder.

Aunque la acumulación de experiencia (es decir, el paso del tiempo) mitiga la carencia de opciones de políticas de eficacia probada, el principal obstáculo en lo referido a las pruebas empíricas son las limitaciones de recursos para acceder a los datos. Lejos de posponerse, se trata de un desafío que debe enfrentarse con decisión. Además, el mismísimo resultado de la revolución digital —la abundancia de datos que la sociedad deja tras de sí en cada de sus interacciones con el mundo digital, es decir, los macrodatos que conforman la huella digital— debería, en teoría, resolver las limitaciones de recursos para acceder a los datos. Por tanto, nuestra propuesta es aprovechar la huella digital que resulta de tales interacciones para efectuar un seguimiento de determinados aspectos de la economía digital.

Esto lleva a la propuesta de continuar aplicando metodologías de investigación que incorporen herramientas alternativas de medición basadas en los datos de Internet y en la huella digital, valiéndose de los datos públicos en línea como insumo para formular las políticas públicas. Las comunicaciones en línea brindan una enorme cantidad de datos digitales que pueden ser útiles como fuente de información para adoptar medidas prácticas. Por ejemplo, los sitios de búsqueda de empleo reflejan cuáles son las dinámicas del mercado laboral, las plataformas de comercio electrónico revelan la disponibilidad y la accesibilidad de la tecnología, las plataformas de ventas minoristas en línea permiten determinar la ubicación de las pequeñas empresas y las concentraciones de mercado, y las redes sociales ilustran las desigualdades de género y los niveles de acceso a los servicios de red. Hemos reunido y analizado datos que brindan importantes percepciones sobre seis aspectos complementarios de la economía digital en América Latina y el Caribe, a saber: el mercado de trabajo; los precios de la tecnología; las microempresas y las pequeñas y medianas empresas; los servicios de banda ancha; las criptomonedas, y las redes sociales. Además de presentar estos datos, en este informe también se describen algunas de las lecciones aprendidas en lo referido a aplicar técnicas de macrodatos al diseño de políticas públicas. En ese sentido, el informe persigue dos objetivos complementarios:

- servir como un inventario actualizado de algunas de las cuestiones que hacen a la economía digital en América Latina y el Caribe; y
- brindar una base para debatir sobre las oportunidades y los desafíos que plantea trabajar con los datos en línea sobre la huella digital para diseñar políticas de desarrollo.

Ambos objetivos son complementarios y deben examinarse en su conjunto. Ilustran las nuevas características de trabajar con datos empíricos en la era digital. Aunque los enfoques tradicionales para cuantificar el desarrollo internacional se han visto afectados por la escasez de datos, se han beneficiado del nivel de calidad relativamente alto que implica el uso de fuentes bien entendidas. Pese a que en nuestro trabajo intentamos acceder a una abundancia de datos, entender qué representa cada tipo de dato y cómo generalizarlos, si acaso eso es posible, supone importantes retos metodológicos.

B. Fundamentos metodológicos de la ciencia de datos

Según uno de los académicos de mayor renombre en sistemas complejos, Alessandro Vespignani, los tres principales desafíos para entender los sistemas tecnosociales actuales radican en: i) recopilar datos a gran escala, ii) formular modelos formales, y iii) implantar infraestructuras de seguimiento (Vespignani, 2009, pág. 428). Poner en marcha y sustentar un sistema de estas características con la ayuda de los datos sobre la huella digital disponibles en línea y de herramientas de análisis informático siempre

implica un flujo de trabajo similar, que supone recopilar los datos en línea, adoptar decisiones metodológicas para depurarlos, y proceder a las tareas de visualización y análisis. En el gráfico 1 se presentan ejemplos de la manera de convertir los datos provenientes de un portal de búsqueda de empleo en información sobre los niveles de destrezas en Centroamérica (véase el gráfico 1.A), y de transformar los datos de un sitio de comercio electrónico en información sobre el nivel de acceso a la tecnología en la región (véase el gráfico 1.B).

Gráfico 1

Representación esquemática del flujo de trabajo de la ciencia de datos con datos sobre la huella digital, desde la etapa de recopilación de datos hasta las fases de procesamiento y visualización

A. Conversión de los datos de un portal de búsqueda de empleo en información sobre el nivel de destrezas en Centroamérica



B. Ejemplo de conversión de los datos de un sitio de comercio electrónico en información sobre el nivel de acceso a la tecnología en la región



Fuente: Elaboración propia.

En lo referido a las habilidades prácticas necesarias, cada etapa puede llevarse a cabo de manera más o menos sofisticada, y exige distintos tipos y niveles de conocimiento. En pocas palabras, la primera etapa exige principalmente habilidades informáticas, la segunda de tipo predominantemente estadístico, y la tercera y última requiere habilidades analíticas y conocimiento específico de la esfera en cuestión. El resultado es la definición de la ciencia de datos como la convergencia entre las ciencias informáticas, la estadística y sus principales ámbitos de aplicación.

En principio, cualquier persona capaz de utilizar un ratón (*mouse*) puede llevar a cabo la versión simple del primer paso copiando y pegando contenidos de Internet. También existen numerosas soluciones comerciales sencillas que permiten automatizar este paso, con funciones que pueden aprenderse en minutos. Varias de ellas ofrecen la opción de brindar asistencia en la recopilación de datos en línea a cambio de un cargo por el servicio. Si lo que se desea son aplicaciones más sofisticadas y sostenibles, es recomendable contar con experiencia básica en programación. Tener conocimientos básicos de Python u otro lenguaje de programación multipropósito permite diseñar soluciones personalizadas para obtener datos sobre la huella digital en línea extrayéndolos de páginas web o accediendo a las bases de datos proporcionadas. La segunda etapa

exige principalmente habilidades estadísticas. Los investigadores que están habituados a trabajar con conjuntos de datos de encuestas (por ejemplo, cuestionarios especiales o encuestas de hogares) en general tienen las habilidades necesarias para encargarse de esta etapa. Para almacenar los datos en un formato básico, puede utilizarse cualquier planilla de cálculo. El siguiente escalón, que implica saber gestionar bases de datos y trabajar con lenguajes idóneos para manejarlas, como SQL, permite a los profesionales aumentar el tamaño de las bases de datos y sofisticar su diseño. En lo referido a esta última etapa, cualquier herramienta de visualización de datos es útil para comenzar, incluidas las soluciones populares como Microsoft Excel o las planillas de Google. La principal habilidad necesaria en este caso es una comprensión profunda del ámbito de aplicación sustantivo. Esto permite a los investigadores presentar un conjunto seleccionado de variables significativas de una manera que sea útil; por supuesto, es necesario determinar la modalidad de presentación para orientar la primera etapa, referida a la recolección de datos. En el mercado hay disponibles paneles en línea que son flexibles e interactivos. Su uso exige habilidades de programación (por ejemplo, en el lenguaje R) o, en el caso de las soluciones orientadas al usuario final, adquirir una licencia comercial.

En las siguientes secciones examinamos las distintas posibilidades con mayor exhaustividad y explicamos las opciones de ejecución que hemos elegido para el ejercicio presentado.

1. Recolección de datos: herramientas y consideraciones jurídicas

Para este ejercicio, recabamos datos sobre la huella digital disponibles públicamente en la Internet abierta. En general, todas las interacciones digitales que ocurren en la economía digital inevitablemente dejan rastros. (Hilbert, 2016). Existen dos maneras sencillas de acceder a la huella digital. Una implica, en esencia, extraer la información que figura en la interfaz de usuario de un sitio web mediante las modalidades de extracción de datos conocidas como rastreo de datos (*web-crawling*) o raspado de datos (*web-scraping*) en la web, y la otra consiste en acceder a la base de datos que sustenta la interfaz en el modo de administración del sitio, para lo que se utiliza una interfaz de programación de aplicaciones (API, *application programming interface*) proporcionada por la plataforma de procesamiento de datos.

a. Raspado de datos y rastreo de datos en la web

La acción de copiar y pegar puede automatizarse con herramientas de rastreo de datos o raspado de datos en la web. En esencia, lo que hacen es copiar y pegar los datos de sitios web de manera automatizada. Estos programas permiten recopilar cualquier información que un usuario pueda ver en un sitio web y copiar y pegar manualmente en una planilla de cálculo. La diferencia es que la tecnología de rastreo de datos en general se utiliza para indexar diversos diseños de páginas web, mientras que el raspado de datos aprovecha el diseño estandarizado de una plataforma única.

Un rastreador de la web (*web-crawler*), también llamado “araña” o “araña web”, es un programa que sistemáticamente explora la Red, a menudo para indizar sitios web. Por ejemplo, los motores de búsqueda en Internet se valen de rastreadores para actualizar sus contenidos copiando el contenido de páginas web, lo que permite mejorar los resultados de las búsquedas de los usuarios. En ese sentido, los rastreadores han sido diseñados para recopilar contenidos de páginas web con diseños sumamente

heterogéneos, algo que, si bien es positivo en lo referido al alcance, tiene la desventaja de que en general el contenido recopilado no está armonizado, contiene numerosas ambigüedades y presenta desafíos en materia de normalización. En este informe no trabajamos con rastreadores, sino que nos concentramos en la técnica de raspado de datos, que aprovecha el diseño armonizado de una plataforma web específica. Este enfoque tiene el beneficio de que los datos recopilados ya vienen con un formato predefinido. Además, es idóneo para la región de América Latina y el Caribe, que cuenta con diversos portales transnacionales que sirven como fuente natural de datos (véase el gráfico 2). Este diseño armonizado es útil para extraer datos sobre los precios de la tecnología. Por ejemplo, la plataforma de comercio electrónico mercadolibre.com opera en 18 países de la región. En todos ellos, el diseño de la plataforma es similar. Las páginas web se diseñan utilizando lenguajes de marcado de hipertexto (HTML y XHTML) que brindan estructura al contenido visualizado en la forma de un texto. Así como los usuarios pueden aprovechar este diseño armonizado para evaluar con rapidez el contenido sin tener que orientarse nuevamente cada vez, los programas automatizados pueden valerse de ese diseño armonizado para extraer datos. Básicamente, el programa accede a una página tras otra, copia los contenidos indicados (no necesariamente todos, sino los que especifica el agente recopilador de los datos) y los pega a una base de datos (puede ser una planilla de cálculo simple, por ejemplo, un archivo de Excel o CSV [archivo delimitado por comas]). Por supuesto, esta técnica solo sirve para extraer contenidos que son accesibles. Si el agente recopilador no puede acceder a una página específica (por ejemplo, páginas ocultas) y copiarla y pegarla manualmente, el programa de extracción tampoco podrá hacerlo.

Existen numerosas herramientas comerciales de raspado de datos que pueden aprender a usarse en minutos. Por ejemplo, la aplicación de código abierto www.webscraper.io es una extensión popular del navegador Google Chrome. El usuario abre en Chrome la página web deseada e indica el contenido que desea extraer, para lo que alcanza con hacer clic sobre él. Existen varios servicios de pago que facilitan la tarea de recopilación; una búsqueda en Internet con el parámetro “servicio de raspado de datos”, por ejemplo, arrojará varios resultados. En general, estas herramientas se ofrecen de manera gratuita hasta un límite predeterminado (por ejemplo, 500 páginas por mes) y también incluyen opciones con costo, que en muchos casos permiten acceder a servicios de asistencia técnica y atención al cliente que son muy útiles para los usuarios sin experiencia, que a menudo requieren ayuda para extraer datos de los diseños de las páginas web más modernas (por ejemplo, con contenidos emergentes) o de las páginas con características de desplazamiento. La solución más eficaz consiste en programar a medida estas herramientas de extracción, algo que en general se hace con el programa Python, un moderno lenguaje de programación multipropósito. Algunos paquetes populares que ofrece Python para las tareas de raspado de datos son Selenium (www.seleniumhq.org) y BeautifulSoup (www.crummy.com/software/BeautifulSoup). En este ejercicio, empleamos ambas herramientas. Hay disponibles numerosos cursos gratuitos en línea que en unas pocas sesiones permiten aprender a usar Python para hacer el raspado de datos (en función del conocimiento previo sobre el paquete).

Gráfico 2

Ejemplos de plataformas transnacionales en línea en América Latina y el Caribe



Fuente: Elaboración propia.

b. Consideraciones éticas y jurídicas

Es importante señalar que la tecnología de raspado de datos exige abrir y cerrar la página a la que se accede, lo que implica usar los recursos del servidor de la empresa que brinda la información. Literalmente, el programa abre la página (en las páginas que no pueden abrirse, por ejemplo, por estar protegidas con una contraseña, no puede aplicarse esta tecnología), carga los contenidos (si los contenidos no se cargan, la tecnología no puede aplicarse), y luego los copia y los pega. Cuantas más personas accedan al portal y carguen los contenidos, más lento será el servicio; en situaciones extremas, los servidores quedan sobrecargados y los usuarios no pueden acceder al portal, algo que en ocasiones se hace a propósito en los ciberataques dirigidos. Por tanto, es entendible que a los propietarios de los portales no les guste que los científicos de datos estén continuamente utilizando la tecnología de raspado de datos en sus sitios. La marca que el raspado deja en el sitio es similar a la que deja en el servidor un ciberataque automatizado. A esto se suma la cuestión de la propiedad intelectual, por cuanto el proveedor del portal a menudo tiene derechos de propiedad sobre los contenidos.

Es por eso que cada vez más proveedores de portales especifican en sus condiciones de servicios si autorizan o no el uso de herramientas automatizadas de copiado y pegado de contenidos en sus sitios. Que estas especificaciones no sean más que una declaración que debe aceptarse para poder utilizar el servicio o que tengan consecuencias jurídicas reales dependerá de la legislación de que se trate y de la forma en que se interactúe con el contenido, así como del uso que se haga de él. En principio, todo lo que está en Internet abierta puede verse y, por ende, está a disposición del público, si bien el portal y sus contenidos siguen siendo propiedad privada del proveedor. Es por ese motivo que en varias legislaciones los propietarios tienen derecho a prohibir el acceso a sus sitios a los usuarios que no se ajustan a sus normas de servicio específicas. Es lo mismo que ocurriría en una tienda física que niega el ingreso a sus instalaciones a una persona con mala conducta. De la misma manera, es lógico que los portales en línea también puedan establecer la norma de que quienes incumplan sus condiciones de servicio no podrán interactuar con los datos que publican.

La cuestión de si la recopilación de datos vulnera los derechos de propiedad intelectual de terceros dependerá de la legislación aplicable, de la naturaleza de los datos y de la forma en que se reutilicen. Nuevamente, el ejemplo de los servicios fuera de Internet, como los periódicos en papel, es ilustrativo al respecto. Si usted se sienta en la cocina y en un pedazo de papel anota el precio de todos los productos ofrecidos en la sección de bienes de segunda mano del periódico, y luego informa el precio promedio de dichos productos en una de sus publicaciones, es poco probable que el propietario del periódico sienta que usted ha vulnerado alguno de sus derechos de propiedad. Más bien, seguramente aprecie que usted aumentó la visibilidad y brindó un análisis de valor agregado al servicio provisto por el periódico. Atraerá más atención al servicio en cuestión. Sin embargo, si copia palabra por palabra todos los anuncios de los productos de segunda mano y los publica en su propio periódico, se trata de un acto de plagio que claramente perjudica el modelo de negocios del propietario del periódico, quien posiblemente entable acciones legales contra usted.

En ese sentido, al copiar y pegar cualquier contenido de la Internet abierta, es importante informarse sobre las condiciones de servicio específicas del portal y sobre la legislación que rige la prestación del servicio, al igual que se informaría sobre las condiciones de servicio de una tienda tradicional. En caso de dudas, instamos a los profesionales de la ciencia de datos a consultar con expertos jurídicos con respecto a la legislación aplicable a su caso.

c. Interfaz de programación de aplicaciones (*“application programming interfaces o API”*)

Dados los factores mencionados, las tecnologías de raspado de datos y rastreo de datos en la web no son las más sencillas ni las más idóneas para recopilar datos. La mejor manera de obtener datos sobre la huella digital en línea de manera sostenible es utilizar una interfaz de programación de aplicaciones (*Application Programming Interfaces*, o API). A los efectos de este trabajo, las API son, en esencia, los puntos oficiales de acceso a determinadas secciones de la base de datos del servicio o la plataforma de un proveedor.

En ciencia de datos, obtener los datos directamente en el modo administrador mediante interfaces API en general se considera una modalidad más oficial, controlada y sostenible de acceder legítimamente al conjunto de datos, mientras que usar las tecnologías de raspado de datos y de rastreo de datos para extraer contenidos de la interfaz de usuario de un sitio es una manera menos controlada y subrepticia de acceder a los datos publicados. Las plataformas en línea no pueden controlar a quienes utilizan herramientas de raspado o rastreo de la web para acceder a sus contenidos, pero sí pueden controlar a quienes utilizan interfaces API. Con frecuencia se brindan fichas virtuales transferibles (*tokens*), que dan acceso temporal a los datos y, al mismo tiempo, efectúan un seguimiento del proceso de recuperación de datos, incluidos registros previos. Con todo, esto no significa que una de las maneras de acceder a los datos en línea esté más generalizada que la otra. Por ejemplo, los omnipresentes servicios de búsqueda de Google, actualmente indispensables para otros numerosos servicios, se basan en que Google continuamente está rastreando e indexando la Red. Para los proveedores de los datos, otorgar acceso mediante las API no solo disminuye la carga de procesamiento en los servidores de la empresa y mantiene cierto nivel de control sobre el uso de los datos, sino que permite aprovechar las ventajas de las redes que se establecen cuando otros servicios se vinculan con los servicios seleccionados que provee la empresa.

En general, las interfaces API son programas informáticos intermediarios que permiten la comunicación entre dos aplicaciones. Cada vez que se utiliza una aplicación

(incluidas las plataformas de redes sociales y los servicios de mensajería) se usa una interfaz API. Numerosos proveedores de plataformas ofrecen un acceso más o menos estructurado y regulado a sus API. Esto permite que los usuarios y los desarrolladores accedan a la totalidad o a partes seleccionadas de los datos que procesa el proveedor de la plataforma. Uno de los motivos comunes para hacer esto es que los desarrolladores de otros servicios y aplicaciones ahora pueden vincular sistemáticamente sus servicios a las API, creando de esa manera un ecosistema en red que beneficia a ambas partes (por ejemplo, al compartir clientes). Cabe citar también el ejemplo de cuando alguien busca un vuelo directamente en el portal de una aerolínea, o utilizando algún tipo de agencia de viajes intermediaria, o haciendo una búsqueda en Google. La agencia de viajes y Google también tienen acceso a la base de datos de los vuelos, pero no realizan búsquedas en el sitio web de la aerolínea para obtener los itinerarios de viaje. Lo que hacen es acceder a algún tipo de API provista por la aerolínea que les permite recuperar los datos sin afectar los servicios que la aerolínea brinda a sus clientes en su sitio web. La API es la interfaz con la que el servicio de viajes en línea se comunica para obtener información de la base de datos de la aerolínea y reservar asientos, o evaluar opciones de equipaje, por ejemplo. La API toma la respuesta de la aerolínea a la consulta y la envía inmediatamente al servicio de viajes en línea, que a su vez muestra al cliente la información más actualizada y pertinente. Algunas API populares incluyen Google Maps (que muchos servicios de viajes y localización utilizan como base para brindar servicios más sofisticados) o Google Translate (Traductor). Para aprovechar las ventajas del potente servicio de traducción de Google de manera automática no es necesario visitar el portal principal para el público (<https://translate.google.com>). También se puede ingresar directamente a la API de traducción en la nube de Google (<https://cloud.google.com/translate>). Esto permite a otros proveedores de servicios integrar la herramienta de traducción de Google a sus sitios web, donde aparecerá la leyenda *powered by Google Translate*.

La mayoría de las API se diseñan para los desarrolladores de otras aplicaciones y programas, y exigen conocimientos rudimentarios de algún lenguaje de programación común, como Python. Debido a que las API son la manera más eficaz de obtener sistemáticamente datos de una fuente de datos elegida, vale la pena adquirir tales habilidades. Además, los proveedores de interfaces API también las utilizan para saber quién accede a sus datos y con qué fin. Muchos exigen un registro previo, que incluye algunos datos personalizados y una declaración del fin que persigue el ejercicio, y a menudo requieren que un ser humano examine el caso antes de otorgar su autorización. En muchos casos, los usuarios reciben los denominados *tokens* (códigos alfanuméricos extensos) que permiten recopilar determinada cantidad de datos durante un plazo predeterminado. Dichos códigos podrían durar apenas unas horas e incluir restricciones referidas a las variables a las que permiten acceder y la frecuencia de las solicitudes. De esta manera, el proveedor de los datos se asegura de que no se produzca una sobrecarga de los servidores. Eso quiere decir que trabajar con una API en ocasiones podría exigir planificar con cuidado el tipo y la cantidad de datos que se recopilarán, y con qué frecuencia. En este ejercicio, para algunas de las API establecimos un protocolo propio de tiempos de espera y de recopilación de datos que nos permite equilibrar el cuello de botella de recuperación de datos que establece la API con el ingreso continuo de datos (por ejemplo, se envían 12 solicitudes aleatorias dentro del plazo de un minuto).

La mayoría de los principales portales internacionales que operan en América Latina y el Caribe brindan interfaces API con variables seleccionadas, en particular los portales que también tienen un alcance mundial. En la medida de lo posible, nos esforzamos por acceder a los datos que obtuvimos utilizando interfaces API. Dado que

el proveedor los preselecciona, los datos disponibles a través de las interfaces API frecuentemente están restringidos, y es posible que la API correspondiente no permita acceder a todo lo que puede verse en la página web pública. Por otro lado, efectuar un seguimiento de la huella digital a fin de diseñar políticas públicas no exige datos extremadamente detallados. Por cuanto trabajamos con estadísticas de alto nivel que a menudo resumen las variables a nivel de país, no estamos interesados en los datos personales de individuos, y no hemos utilizado herramientas de identificación que infrinjan las normas de privacidad. Hemos trabajado con conjuntos de datos sumamente agregados que reflejan promedios o distribuciones a nivel nacional. Para que proyectos como el que se presenta en este informe sean sostenibles a largo plazo, sería deseable acordar y regular el acceso a las variables necesarias a través de las interfaces API de los proveedores de servicios seleccionados en la región.

2. Procesamiento de datos

La segunda etapa, que se ilustra en el gráfico 1, a menudo se denomina limpieza de los datos (*data wrangling*) en el ámbito de la ciencia de datos. Abarca todos los pasos necesarios para convertir los datos recopilados a un formato adecuado para su análisis. En la vida cotidiana de los científicos de datos, es común que esta tarea, que puede representar hasta un 80% del trabajo, sea la que más tiempo consume.

Como sucede con todos los datos empíricos, los datos en línea a menudo están incompletos (no todas las filas y las columnas aparecen llenas) o son poco fiables (no todos los datos pueden tomarse al pie de la letra). Sin embargo, es importante recordar que el trabajo implicado no es más complejo que la labor que desarrollan los investigadores tradicionales que trabajan con datos empíricos obtenidos fuera de Internet (como los etnógrafos, que anotan sus observaciones en forma de filas y columnas), o que las encuestas, en las que es necesario determinar la veracidad y fiabilidad de las respuestas recibidas. Por ejemplo, hace décadas que organizaciones como la CEPAL se enfrentan al complejo desafío estadístico de armonizar los resultados de encuestas de hogares realizadas en diversos países. No todas las variables serán exactamente iguales en cada encuesta de hogar y en algunos casos podría faltar información. El trabajo subsiguiente de limpieza y preparación de los datos supone decidir cómo se armonizarán las discrepancias, y en caso de hacerlo, cómo se extrapolarán los valores faltantes. Se requieren conocimientos de estadística para determinar cuán representativos son los datos obtenidos; por definición, los datos empíricos nunca constituyen una muestra aleatoria representativa, sino que siempre están sesgados. Asimismo, es necesario evaluar la fiabilidad de la procedencia de los datos y sus contenidos específicos.

Una característica recurrente es que la escala de los datos en línea se amplía con enorme rapidez, lo que a menudo se traduce en el notorio fenómeno de los macrodatos. Esto sucede en numerosos casos y en numerosas variables interrelacionadas. El contenido resultante se presta a diversos e intrincados diseños de bases de datos. Sin embargo, en aras de la usabilidad general, en este ejercicio nos hemos ceñido a un formato simple de planilla de cálculo en la forma de archivos CSV (valores separados por comas), que son archivos de texto en los que las distintas columnas se separan con comas. Son archivos que pueden abrirse y procesarse con cualquier programa de planillas de cálculo, por ejemplo, Microsoft Excel o Google Sheets, así como con paquetes más avanzados de análisis estadísticos, como R, Pandas de Python, el Programa de Estadísticas de Ciencias Sociales (SPSS), STATA, etc.

3. Análisis y visualización de los datos

La tercera y última etapa en nuestro flujo de trabajo de ciencia de datos implica convertir las cifras crudas en algún tipo de formato a partir de cual puedan derivarse conclusiones que permitan impulsar acciones concretas. En nuestro caso, hemos optado por elaborar una serie de visualizaciones a través de las cuales los analistas, los encargados de la formulación de políticas, el mundo académico y el público en general podrán obtener percepciones bien fundamentadas y oportunas. El nivel de análisis es aún muy superficial, y la mayoría de las decenas de gráficos e imágenes que se presentan en este informe se verían beneficiados de un análisis mucho más exhaustivo en la forma de un estudio de caso (como complemento de percepciones cualitativas procedentes de ámbitos específicos) o si se aplicasen métodos analíticos más sofisticados (por ejemplo, análisis multivariados, como los modelos de ecuaciones estructurales, u otros análisis econométricos).

En el formato más rudimentario, los datos obtenidos pueden visualizarse con gráficos simples mediante programas populares como Microsoft Excel o Google Sheets. Es precisamente eso lo que hemos hecho con muchos de los gráficos que figuran en este informe. La limitación fundamental con este enfoque es que la riqueza de la presentación queda limitada por la necesidad de elegir un número predeterminado de variables (en general, un subgrupo seleccionado de todas las variables disponibles) y decidir cuál será el nivel de agregación (las variables condicionantes en función de las cuales se evalúan los promedios y las distribuciones). Los gráficos que se presentan son apenas algunos de todos los gráficos que podrían haberse presentado utilizando los mismos datos. Si bien se trata de un enfoque útil para responder a cuestiones o hipótesis concretas de una investigación, o para formular una opinión con respecto a argumentos existentes, va en contra de las premisas fundamentales del paradigma de los macrodatos: aprovechar todos los datos disponibles.

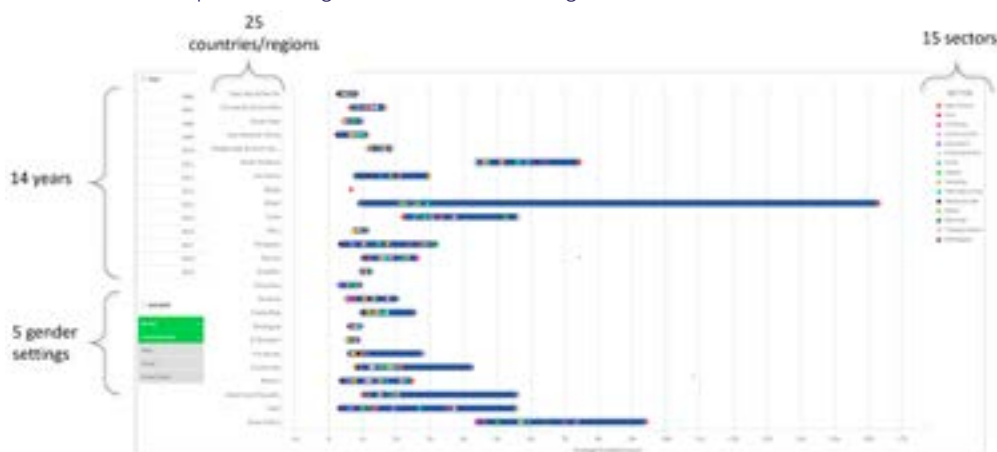
Por ejemplo, un conjunto de datos típico recabado para este ejercicio podría incluir 25 países (o regiones), hacer el seguimiento de la actividad de 15 sectores durante 14 años y efectuar un desglose en función de 5 aspectos sociodemográficos (véase, por ejemplo, el análisis de un modelo financiero de colaboración abierta en el capítulo 3.2 y en el gráfico 3). Esto supone un total de $(25+15+14+5) = 59$ variables. En la forma más básica, podríamos elegir un gráfico que muestre todas las variables en su conjunto, incluidos todos los años y todos los sectores, desglosados según todos los países y por género. Sería un gráfico sumamente denso, con toda probabilidad ilegible. Por otro lado, podríamos presentar una única cifra, por ejemplo, un sector durante un año, en un único país, solo para los hombres. Tenemos nada más y nada menos que 59 de estas cifras. Difícilmente esto pueda considerarse un gráfico, además de que se perderían importantes aspectos comparativos. En la práctica, lo que sucede es que en el transcurso del proceso alguien con conocimiento de ámbito en cuestión hace una elección deliberada sobre la selección de variables y su nivel de agregación, y determina cuáles son los años/sectores/aspectos más “interesantes”, “esclarecedores” o “significativos”, así como su nivel de agregación. Inevitablemente, este proceso de selección tiene más de arte que de ciencia, por cuanto el número de opciones es prácticamente ilimitado, lo que hace que seleccionar cualquier gráfico específico sea sumamente restrictivo. La cantidad de gráficos posibles es la suma de todos los valores posibles de la conocida fórmula “combinaciones de n en k ”, donde n denota el número de variables disponibles, y k el número de variables elegidas presentadas en un gráfico dado:

$$\text{Número de gráficos posibles} = \sum_{k=1}^n \binom{n}{k} = \sum_{k=1}^n \frac{n!}{k!(n-k)!}$$

La combinación de dos cualesquiera del total de 59 variables permite crear 1.711 gráficos distintos. Si se toman en cuenta tres variables, se pueden crear 32.509 gráficos. Si son cuatro, el investigador puede elegir de entre medio millón de gráficos distintos (455.126 para ser exactos), y si se seleccionan cinco cualesquiera de las 59 variables, se pueden crear más de cinco millones de gráficos (5.006.386). En total, la suma de todas las opciones posibles permite generar más de medio trillón de gráficos distintos (exactamente 576.460.752.303.423.000), lo que representa más de mil millones de gráficos distintos por cada habitante de América Latina y el Caribe.

Gráfico 3

Captura de pantalla de ejemplo de un panel que muestra el monto promedio de 1.151.384 préstamos de financiación cooperativa otorgados a través de Kiva.org



Fuente: Elaboración propia.

La ciencia de datos también implica elegir de entre ese número casi ilimitado de opciones. El primer requisito para hacerlo con eficacia es tener un conocimiento profundo del ámbito objeto de la investigación. Sin la orientación de preguntas de investigación debidamente fundamentadas, es poco probable que se seleccionen gráficos pertinentes. Es por eso que la ciencia de datos va mucho más allá de la ciencia informática y también de un mero ejercicio estadístico. El segundo paso consiste en intentar abrir el chaleco de fuerza que suponen las herramientas de representación de los datos. En las ciencias de datos modernas, a menudo no se trabaja con gráficos estáticos, sino con los denominados paneles. Se trata de herramientas interactivas que permiten a los usuarios seleccionar y deseleccionar variables para analizar los distintos aspectos del gráfico. El público en general se está acostumbrando a tales paneles en línea, que incluyen gráficos de líneas interactivos que les permiten concentrarse en períodos determinados (por ejemplo, para analizar el historial de fluctuaciones en el tipo de cambio) o herramientas que les permiten ver en cuánto cambia el precio total de su compra seleccionando (o deseleccionando) productos en su canasta de compras en línea.

En nuestro caso, optamos por trabajar con herramientas de visualización de ese tipo. Existen opciones de código abierto, como “Shiny de R” que funciona sobre la base del paquete estadístico R y tiene una extensión comercial: <https://shiny.rstudio.com>. Algunas soluciones comerciales populares exigen una licencia comercial y el pago de un cargo por usuario.

C. Ámbitos de interés y fuentes de datos

Los ámbitos seleccionados que se eligieron para este ejercicio fueron los que surgieron como el denominador común de todos los intereses expresados por los países de América Latina y el Caribe en la Agenda Digital para América Latina y el Caribe (eLAC2020) y las realidades pragmáticas de las fuentes disponibles. Dado el creciente interés en las cuestiones de género y en la cuantificación de determinados aspectos de la economía digital, el estudio buscó desglosar por género numerosas de las variables obtenidas. Al mismo tiempo, elegimos temas para los cuales resulta en general difícil o imposible obtener datos utilizando los métodos tradicionales.

Para cualquier ejercicio de seguimiento digital que persiga objetivos de desarrollo internacional, es útil obtener datos de plataformas digitales con presencia en varios países distintos y con un único modelo de negocios. Esto garantiza que los datos puedan compararse y facilita las tareas de recopilación. Afortunadamente, en América Latina y el Caribe existen varias plataformas panregionales de ese tipo, sobre todo debido a la presencia de un contexto cultural, lingüístico y geográfico en común. Por ende, dichas plataformas rápidamente concitaron nuestra atención a la hora de analizar de manera pragmática posibles fuentes de datos. Entre dichas plataformas se cuentan Bumeran, Coin.dance, Facebook, Freelancer, Oookla, Kiva, MercadoLibre, Twitter y Workana. Sin embargo, también realizamos ejercicios exploratorios con otras fuentes locales (por ejemplo, con portales y plataformas nacionales), especialmente para verificar la fiabilidad y la representatividad de las fuentes.

A raíz de estas diferentes consideraciones, seleccionamos cinco ámbitos principales de interés que reflejan distintos aspectos de la economía digital de América Latina y el Caribe, con hincapié especial en las cuestiones de género. Todos los datos fueron reunidos entre enero y marzo de 2019 en la sede de la CEPAL en Santiago.

Cuadro 1

Temas de economía digital, fuentes web y número de observaciones, enero a marzo de 2019

Temas	Fuentes	Observaciones
A) Mercado de trabajo y habilidades digitales		
a) La economía de las ocupaciones transitorias de los trabajadores autónomos	Freelancer.com Workana.com	N = 74 970 personas N = 19 840 personas
b) Contratos en el mercado de trabajo tradicional	Bumeran.com BNE.cl	N = 78 475 personas N = 1 254 personas
B) Precios de la tecnología		
a) Índice de precios	Mercadolibre.com	N = 192 092 productos
b) Categorías tecnológicas	Mercadolibre.com	N = 192 092 productos
C) Pequeñas y medianas empresas		
a) Minorista en línea	Mercadolibre.com	N= 2 473 977 vendedores
b) Financiación colectiva	Kiva.org Kickstarter.com.mx	N= 1 551 367 préstamos N= 888 préstamos
D) Broadband		
	Oookla's netindex	N= 9 360 observaciones
E) Criptomonedas		
	Coin.dance	N=15 840 observaciones
F) Redes sociales		
a) Aspectos sociodemográficos	Gestor de anuncios de Facebook	N= 110 880 observaciones
b) Tendencias: Objetivos de Desarrollo Sostenible (ODS)	Twitter	N = 35 818 261 tweets

Fuente: Elaboración propia.

Para evaluar el mercado laboral, trabajamos con tres fuentes distintas. Freelancer y Workana son dos sitios web característicos de la llamada “economía de las ocupaciones transitorias”, donde los profesionales ofrecen destrezas específicas y los clientes solicitan asistencia para proyectos específicos. Freelancer tiene alcance mundial, mientras que Workana se concentra en América Latina. Ambos brindan datos sobre la oferta y la demanda de los mercados laborales, dado que las plataformas ofrecen a los profesionales la oportunidad de publicar un perfil, y a los empleadores la posibilidad de publicar ofertas laborales. También trabajamos con Bumeran, otro sitio donde las empresas publican oportunidades laborales. Lo anterior solo refleja el componente de la demanda de las ofertas de trabajo, ya que la huella digital no revela qué postulantes son entrevistados o contratados.

Como primera medida, proponemos un sistema para armonizar las categorías. A diferencia de otros ejercicios, no trabajamos con clasificadores basados en aprendizaje automático (*machine learning*) (Amato et al., 2015; Kässi & Lehdonvirta, 2018a), ya que las plataformas brindan las categorías directamente, tras lo cual nosotros establecimos las relaciones correspondientes. Al explorar las categorías de trabajo relacionadas con la economía digital, constatamos que a diferencia de Freelancer, donde existen varios miles de categorías, solo hay un par de docenas en Workana. En cuanto a Bumeran, también utiliza una taxonomía distinta para las categorías de trabajo. A continuación, usamos el mecanismo de armonización que se muestra en el cuadro 2 para crear 10 categorías comparables para las tres fuentes.

Cuadro 2

Sistema de armonización de las categorías laborales registradas para la economía digital

	Freelancer	Workana	Bumeran (1)
Tecnología de la información y programación	Sitios web, informática y software Teléfonos móviles y computadoras	Informática y programación	Tecnología
Escritura y traducción	Traducción e idiomas Escritura y contenidos	Escritura y traducción	Comunicación
Diseño y arquitectura	Diseño y arquitectura	Diseño y multimedia	Diseño
Ingreso y administración de datos	Ingreso y administración de datos	Soporte administrativo	Administración, servicios de secretaría
Ingeniería y fabricación	Ingeniería y ciencia Aprovisionamiento y fabricación de productos	Ingeniería y fabricación	Construcción, ingeniería, minería, producción, salud
Ventas y marketing	Ventas y Marketing	Ventas y Marketing	Marketing, ventas, centros de atención telefónica
Fletes, envíos y transporte	Fletes, envíos y transporte		Logística
Servicios empresariales, contables, jurídicos y de recursos humanos	Servicios empresariales, contables, jurídicos y de recursos humanos	Servicios jurídicos, financieros y de gestión	Finanzas, gestión, servicios jurídicos, recursos humanos, educación, asistentes contables, analistas comerciales, contadores, personal administrativo y analista de recursos humanos
Empleos y servicios locales	Empleos y servicios locales		Gastronomía, artesanía, chófer, seguridad, cajero, muebles, armador, asistente para servicios higiénicos, operador de producción, pulidor, maestro de cocina, soldador, salvavidas, mecánico
Otros	Otros		Seguros

Fuente: Elaboración propia.

A. La economía de las ocupaciones transitorias de los trabajadores autónomos

Un aspecto cada vez más importante de la economía digital es la denominada economía de las ocupaciones transitorias (Heeks, 2017). Las redes digitales aportan flexibilidad a los trabajadores y la oportunidad de tener una fuente adicional de ingresos. Para un sector cada vez más grande del mercado laboral, lo que otrora eran trabajos complementarios pequeños o tareas ocasionales de consultoría se han transformado en la principal fuente de empleo. El hecho de que una parte importante de las oportunidades que ofrece la economía de las ocupaciones transitorias se relaciona con servicios que pueden brindarse a distancia implica que los trabajadores acceden al mercado de trabajo mundial, un aspecto importante que debe tomarse en cuenta en nuestro análisis de la región de América Latina y el Caribe. Hemos analizado dos plataformas basadas en este modelo económico, a saber: Freelancer.com, un sitio mundial de colaboración masiva fundado en 2009 con más de 21 millones de usuarios, sede central en Sydney y oficinas en Buenos Aires, el sur de California, Londres, Manila, Vancouver y Yakarta, y Workana.com, un sitio centrado principalmente en América Latina, sobre todo en la Argentina, el Brasil, Colombia, México, el Uruguay y Venezuela (República Bolivariana de), pero que se utiliza en toda la región. A diferencia de otros inventarios (Kässi y Lehdonvirta, 2018b), no solo analizamos la oferta laboral, es decir las plataformas digitales donde los profesionales ofrecen sus servicios, sino también la demanda, que representa a los empleadores que ofrecen oportunidades de empleo para encontrar profesionales.

Incluso realizando un análisis por país, la plataforma de alcance mundial Freelancer.com atrae a un número mucho mayor de profesionales que la plataforma latinoamericana Workana.com. Con la excepción notable del Brasil, la mayoría de los países tienen entre cuatro y seis veces más profesionales registrados en Freelancer que en Workana. Si bien los trabajadores pueden estar registrados en ambos sitios a la vez, nuestra teoría es que la oportunidad que ofrece Freelancer.com de acceder a empleos internacionales podría ser incentivo importante para los trabajadores de la economía digital.

1. Oferta de categorías de empleo por país

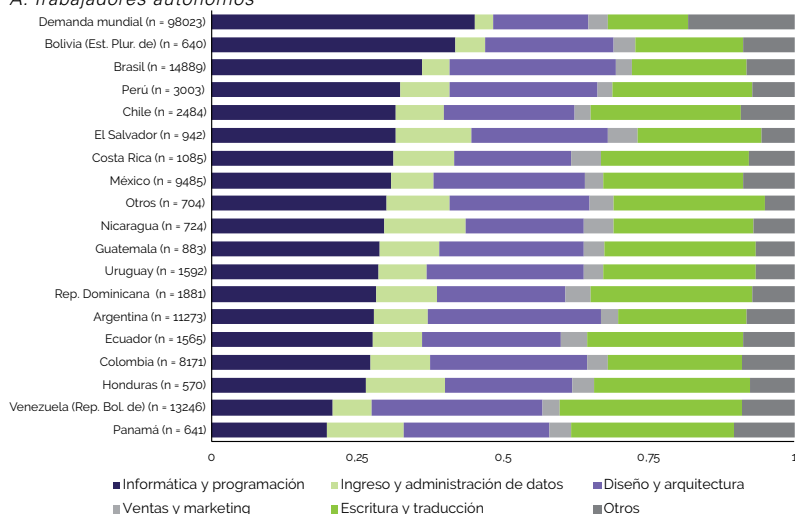
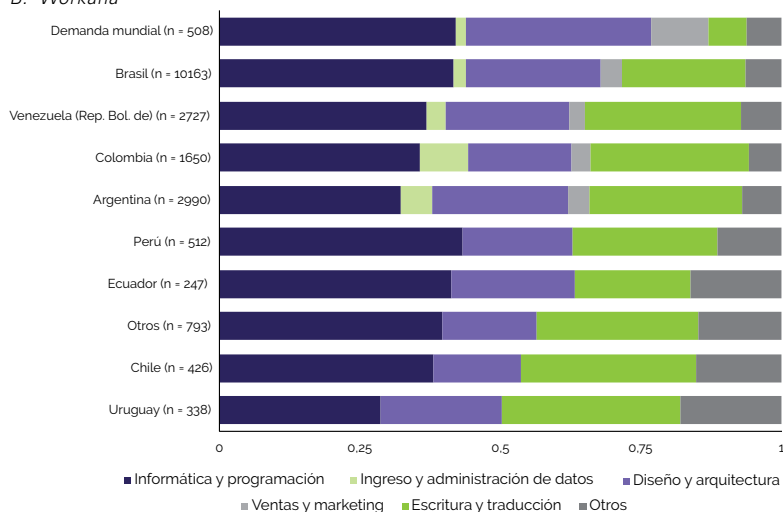
En el gráfico 4 se puede apreciar que la distribución de la oferta de destrezas a nivel nacional es sorprendentemente similar aun entre países sumamente distintos, sean grandes o pequeños, estén en el norte o en el sur. En la mayoría de los países, la categoría “Informática y programación” es la más importante en la economía digital, seguida por las categorías “Escritura y traducción” y “Diseño y arquitectura”. Asimismo, para la mayoría de los países, y en ambas fuentes de datos, los dos sectores clave de la economía digital, “Informática y programación” e “Ingreso y administración de datos”, representan un 40% de la oferta laboral. Es interesante destacar que, contrariamente a lo que cabría esperar, algunos países, como Bolivia (Estado Plurinacional de), El Salvador y Nicaragua, superan este porcentaje.

También vale la pena señalar que la participación de la categoría “Informática y programación” es sistemáticamente más alta en la plataforma regional Workana.com que en la plataforma mundial Freelancer.com. Al mismo tiempo, Freelancer.com ofrece un mayor número de profesionales en el ámbito relacionado de “Ingreso y administración de datos”, lo que plantea la interrogante de si algunas de estas diferencias podrían deberse a incongruencias en el sistema de clasificación (véase el cuadro 2). Los estudios futuros deberían analizar esta posibilidad, quizás usando un clasificador automático que no se base en la clasificación que se utiliza en la plataforma, sino en una clasificación propia basada en los títulos o las descripciones de los cargos (Amato y otros, 2015; Kässi y Lehdonvirta, 2018a).

Gráfico 4

Proporciones de categorías laborales profesionales por país, febrero de 2019

(En porcentajes)

A. Trabajadores autónomos**B. Workana****Fuente:** Elaboración propia, sobre la base de datos de freelancer.com y workana.com.

Nota: En el gráfico A la categoría "Otros" se refiere a: a) Países: Cuba, Guadalupe, Guyana Francesa, Haití, Martinica, Paraguay, Puerto Rico y San Martín; b) Categorías: empleos y servicios locales; fletes, envíos y transporte; ingeniería y fabricación, y servicios empresariales, contables, jurídicos y de recursos humanos. En el gráfico B la categoría "Otros" se refiere a: a) Países: Belice, Bolivia (Est. Plur. de), Costa Rica, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Puerto Rico, República Dominicana, Trinidad y Tabago; b) Categorías: ingeniería y fabricación, y servicios empresariales, contables, jurídicos y de recursos humanos.

2. Demanda mundial y oferta nacional

Ambos sitios de colaboración masiva, Freelancer y Workana, también permiten a los empleadores publicar ofertas de trabajo. Dado que los trabajos de la economía digital se refieren principalmente a servicios que pueden brindarse de forma remota, calculamos la demanda mundial independientemente del país de origen.

La primera fila de los gráficos 4.A y 4.B refleja los niveles de demanda mundial y regional. Posiblemente la conclusión más llamativa sea que en el mercado mundial

de Freelancer.com (véase el gráfico 4.A), la categoría laboral de mayor demanda a nivel mundial (“Informática y programación”) tiene una representación claramente escasa en los países de América Latina y el Caribe. En ambas fuentes de datos, hay una clara sobrerrepresentación en las categorías “Escritura y traducción” e “Ingreso y administración de datos”, algo que queda claro en el caso de Freelancer.com (véase el gráfico 4.A).

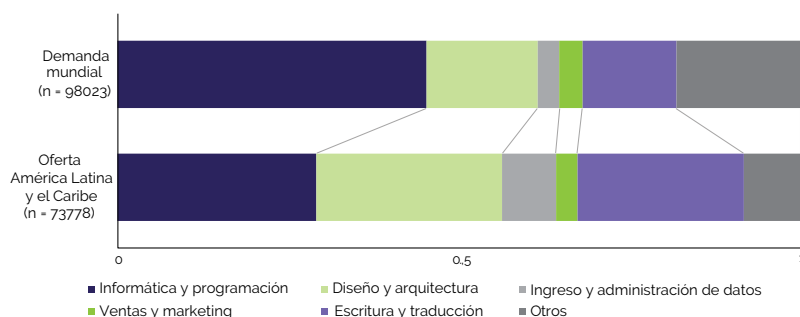
En los gráficos 5.A y 5.B se muestra un resumen de la oferta regional y la demanda mundial para cada plataforma. Un análisis de correlación revela que en ambos casos el nivel general de ajuste entre la oferta y la demanda es similar: para Freelancer.com es de $R^2 = 0,590$ y para Workana.com de $R^2 = 0,607$. Sin embargo, como se muestra en el gráfico, este desajuste general tiene orígenes distintos. Como ya se mencionó, las destrezas correspondientes a la categoría “Informática y programación” que se ofrecen en América Latina y el Caribe están a la zaga de la demanda mundial, mientras que las habilidades de “Escritura y traducción” están sobrerrepresentadas. Eso indica que la oferta de destrezas en la región no se ha mantenido a la par con los cambios tecnológicos, ya que muchas de las tareas de escritura y traducción han incorporado elementos de inteligencia artificial, y las habilidades de programación se han transformado en un aspecto importante de los negocios cotidianos. La brecha en la categoría “Informática y programación” es notablemente más grande para el mercado mundial de Freelancer.com. Otra diferencia entre ambas plataformas se aprecia en la categoría “Diseño y arquitectura.” Para el mercado mundial de Freelancer.com (véase el gráfico 5.A), las destrezas que ofrece la región de América Latina y el Caribe están sobrerrepresentadas, mientras que en el mercado regional de Workana.com (véase el gráfico 5.B) ocurre lo contrario. No queda claro el motivo de esta situación, que requeriría un análisis más exhaustivo. También parece haber una demanda relativamente mayor por las habilidades de la categoría “Ventas y marketing” en Workana.com que en Freelancer.com.

Los mismos datos también pueden resumirse mediante una de matriz cromática, que ilustra las diferencias entre la oferta y la demanda mediante códigos de color. En el gráfico 5.C, cuanto más oscura sea la celda, mayor será el exceso de la oferta, y cuando más clara, mayor la escasez. La medida subyacente es proporcional, lo que significa que un valor de 2,0 (véase la leyenda) implica que el porcentaje nacional de la oferta en esta categoría duplica el porcentaje de la demanda mundial.

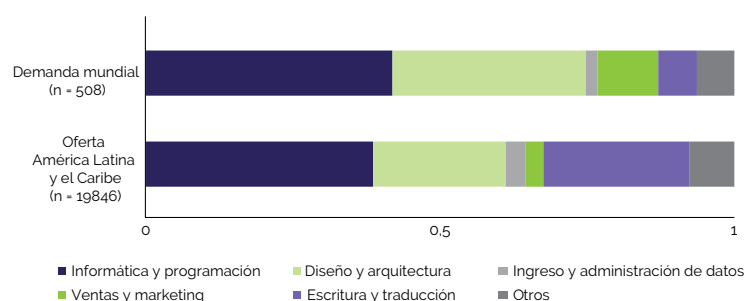
Gráfico 5

Demanda mundial de empleos y oferta nacional, por categoría de empleo

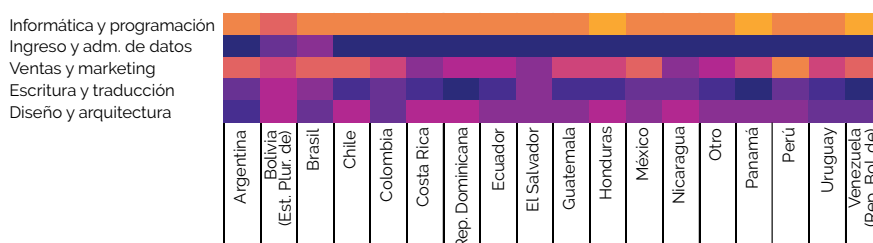
A. Freelancer: demanda mundial y oferta en América Latina y el Caribe (En porcentajes)



B. Workana: demanda mundial y oferta en América Latina y el Caribe (En porcentajes)



C. Freelancer: Mapa de calor de demanda mundial y oferta nacional, febrero de 2019 (Índice)



Fuente: Elaboración propia, sobre la base de datos de Freelancer y Workana.

Nota: En los gráficos A y B la categoría "Otros" se refiere a: Categorías: ingeniería y fabricación, y servicios empresariales, contables, jurídicos y de recursos humanos.

3. Remuneración horaria por país y por categoría de empleo

En las plataformas masivas de búsqueda de trabajo, los profesionales publican sus aspiraciones de remuneración por hora, que en realidad deben considerarse como una expresión de deseo sin ningún tipo de poder vinculante. Teniendo esto en cuenta, si las tarifas solicitadas se analizan de forma apresurada, la calidad de los datos se ve afectada. Algunos usuarios parecen confundir las monedas. Es probable que una persona que pide "\$10.000 por hora" en Chile en realidad esté pidiendo 15 dólares por hora, o 10.000 pesos chilenos, no 10.000 dólares por hora. El error podría deberse a que el profesional seleccionó incorrectamente la moneda al ingresar su tarifa por hora. Se trata de un error fácil de detectar para la mayoría de las monedas, especialmente cuando el monto solicitado supera los 1.000 dólares, como sucede en los casos de Chile y Colombia, cuyo tipo de cambio se ubicaba en alrededor de 670:1 y 3.200:1 cuando se recopilaron los datos. Calculamos el monto correspondiente en dólares de los Estados Unidos y excluimos todos los demás casos donde la tarifa por hora seguía superando los 1.000 dólares. En el caso de Workana.com, esto redujo la muestra de 19.846 a 19.054 perfiles.

Como resultado, obtuvimos un promedio regional de 14,8 dólares por hora para los profesionales de Workana.com, y de 22,03 dólares para los de Freelancer.com¹. Con esos datos, sería lógico concluir que los profesionales que apuntan al mercado mundial de Freelancer.com solicitan tarifas más altas que los que ofrecen sus servicios en el mercado regional de Workana.com. Sin embargo, un análisis más cuidadoso de los datos también revela algunos problemas de calidad que podrían estar relacionados con diferencias en la calidad de los datos de los usuarios que se recopilaron. Parece que al rellenar las casillas en la plataforma, algunos profesionales confunden la noción

¹ Basado en las tarifas por hora en el Brasil, Chile, Colombia, el Ecuador, el Perú, el Uruguay y Venezuela (República Bolivariana de).

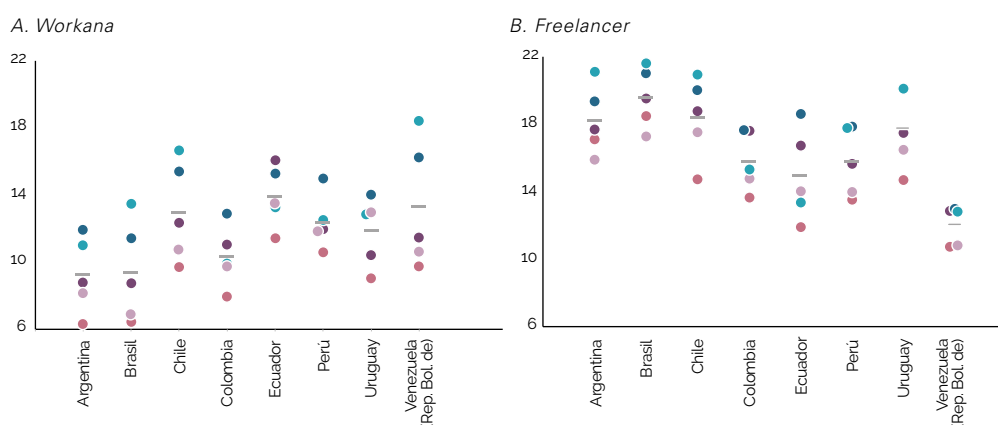
de precio fijo por trabajo con la de tarifa por hora. En otros casos, queda claro que las tarifas reflejan las aspiraciones exageradas de los profesionales, pero en realidad los proyectos que efectivamente se materializan se pagan a tarifas totalmente distintas. Encontramos casos de profesionales en México que pedían 200 dólares por hora, pero que lo único que ganaron fue 10 dólares y 35 dólares en dos proyectos.

Analizamos más perfiles individuales y encontramos muy pocos profesionales en condiciones de pedir más de 100 dólares por hora. Además, la mayoría de los proyectos parecen haberse pagado a una tarifa por hora inferior, y aunque no es posible calcular la tarifa por hora a partir del total de proyectos, sí es posible formarse una idea cualitativa sobre la cantidad de trabajo implicado. Por tanto, en el gráfico 6 se presentan las tarifas promedio por hora tras eliminar aquellas superiores a 100 dólares por hora. La tarifa promedio por hora se reduce a 16,5 dólares. Es una reducción considerable si se tiene en cuenta que se eliminaron menos del 2% de los perfiles ($N = 55.132$) (véase el gráfico 6.B). También con un límite máximo de 100 dólares, la estimación ajustada de la tarifa promedio por hora en Workana.com es de 11,7 dólares ($N = 18.994$) (véase el gráfico 6.A). Cuando se trabaja con datos sobre la huella digital, no hay una única respuesta a la pregunta de cuál es la tarifa por hora correcta que los profesionales piden en las plataformas de trabajadores autónomos. El hecho es que en nuestro ejercicio, el orden en que se ubican las tarifas por hora para varios sectores es el mismo cuando se comparan los distintos países: el sector “Ingreso y administración de datos” es el que menos paga, seguido por “Escritura y traducción,” en tanto que “Informática y programación” es el sector más lucrativo. Esta observación parece sugerir que las diferencias entre ambas plataformas son absolutas y estables en todo el universo de destrezas, y que la tarifa promedio por hora en el mercado regional de Workana.com está proporcionalmente unos 5 dólares por debajo de la tarifa promedio por hora en el mercado mundial de Freelancer.com

Gráfico 6

Tarifa horaria media (precio solicitado) de trabajadores autónomos por país y por categoría laboral, máximo de 100 dólares, febrero de 2019

(En dólares)



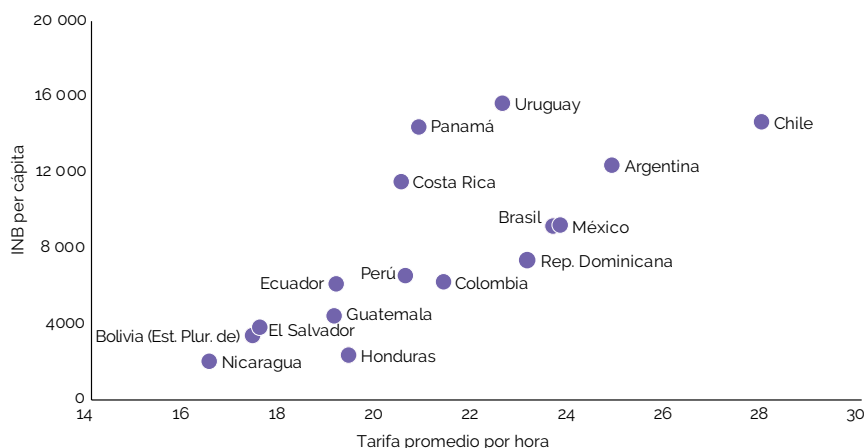
Fuente: Elaboración propia, sobre la base de datos de Freelancer y Workana.

En el gráfico 7 se ilustran las tarifas medias por hora de Freelancer.com (hasta un máximo de 1.000 dólares) en el contexto del ingreso del país. Las tarifas por hora en la plataforma demuestran una clara correlación con los ingresos en el mundo real. Todos los países más hacia la derecha que el promedio tienen tarifas por hora relativamente altas frente al nivel de ingresos de sus economías nacionales (por ejemplo, Colombia, México y la República Dominicana), mientras que los países que aparecen más arriba en el gráfico tienen ingresos nacionales más elevados y tarifas por hora relativamente más bajas (por ejemplo, Costa Rica, Panamá y el Uruguay). En otras palabras, en el contexto de sus economías nacionales (el costo de oportunidad frente a otras actividades generadoras de ingresos), parece que los trabajadores autónomos de la economía digital mundial en México y la República Dominicana tienen ingresos comparativamente más altos que los de Costa Rica o el Uruguay.

Gráfico 7

Tarifa media por hora (máximo de 1.000 dólares) en Freelancer.com e ingreso nacional bruto (INB) per cápita, febrero de 2019

(En dólares)

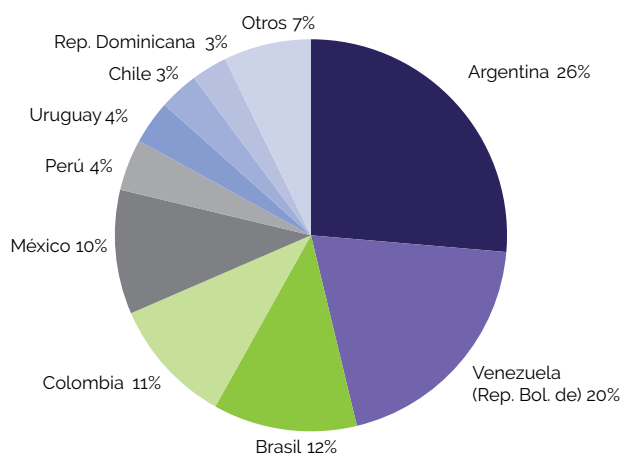


Fuente: Elaboración propia, sobre la base de datos de Freelancer y Workana.

En el gráfico 8 se muestra una estimación del tamaño real de la economía de las ocupaciones transitorias en Freelancer.com, en función del dinero ganado por los trabajadores autónomos que efectivamente han ejecutado proyectos. La Argentina y la República Bolivariana de Venezuela, que han captado casi la mitad de los ingresos de la región (4,2 y 3,2 millones de dólares, respectivamente), son los países que más han aprovechado las oportunidades de trabajo a distancia a través de la plataforma. Si bien son importantes, estos valores muestran que en América Latina y el Caribe este modelo económico sigue siendo relativamente pequeño. Los trabajadores a distancia de la República Bolivariana de Venezuela que utilizan la plataforma declaran haber trabajado más de 250.000 horas, lo que arroja una tarifa real por hora de unos 12,80 dólares. En el contexto de una semana laboral de 40 horas, este nivel de ingresos es comparable al de un trabajo de jornada completa bien pagado.

Gráfico 8

Participación de los trabajadores autónomos en las ganancias por país de origen
(N = 6.751), febrero de 2019
(En porcentajes)



Fuente: Elaboración propia, sobre la base de datos de Freelancer.

4. Cuestiones de género en la economía digital

a. Identificación del género a partir de la huella digital

Ninguna de las dos plataformas masivas de búsqueda de trabajo indican el género de los profesionales como categoría independiente; de hecho, hacerlo podría poner en entredicho las iniciativas para eliminar la discriminación. Por tanto, utilizamos un clasificador de aprendizaje automático, NamSor (clasificador de género y etnicidad en función del nombre) para determinar el género de los trabajadores utilizando sus nombres². La API gratuita de NamSor nos permitió clasificar los nombres y los apellidos de los usuarios por género. Las clasificaciones posibles eran tres: masculino, femenino o desconocido. Para cada clasificación, se indica un "puntaje de confianza" en la forma de una variable continua de entre -1.0 y 1.0, donde -1 indica certeza en cuanto al género masculino, y +1 certeza cuanto al género femenino.

Realizamos un análisis de sensibilidad de la precisión tomando varias muestras aleatorias (n=20 cada vez) e investigamos los tipos de nombres clasificados en cada nivel. Según la configuración por defecto de la API, los nombres con un puntaje de confianza de entre [-0.1, 0.1] se clasificaron como desconocidos. Sin embargo, nuestro análisis demostró que con el umbral de [0.1], entre el 2% y el 4% de los nombres se clasifican como desconocidos, mientras que con un umbral más conservador de [0.2], alrededor del 10% de los nombres se clasificaron como desconocidos (véase el cuadro 3). Decidimos aplicar el umbral más conservador, de (-0.2, 0.2), para el género desconocido.

Hay varias opciones para afinar las clasificaciones generales. Una sería inspeccionar manualmente el 10% de casos desconocidos. Lo que sucedería en ese caso es que la mayoría de los nombres desconocidos (como "Estauri", "Zaimari", "PatriV=c", "Luan", "Maxi", "Criz" o "Asistente") también serían de difícil clasificación para un ser humano, mientras que otros necesitan modificaciones que una inspección humana podría resolver, por ejemplo, "Romerd", "Rodrig." o "Jakelinne". Esto demuestra que solo es posible afinar manualmente un pequeño porcentaje de los nombres. También evaluamos la

² <https://v2.namsor.com/NamSorAPIv2/index.html>.

posibilidad de clasificar las imágenes de perfil con un programa de reconocimiento facial del paquete de aprendizaje automático supervisado de Google llamado TensorFlow. Entrenamos el programa con 4.000 imágenes codificadas manualmente procedentes de Freelancer.com. La primera mala noticia fue que había un número relativamente alto de imágenes de perfil que no sirvieron, ya que no contenían un rostro humano. La segunda fue la escasa correlación entre los resultados del clasificador de imágenes y los del clasificador de nombres (correlación general de $R^2 = 0,40$). Tras priorizar los resultados del clasificador de nombres, el clasificador de imágenes pudo resolver apenas un 2 % más de los casos, lo que nos dejó con alrededor de un 8 % de casos que siguieron estando en duda. Recomendamos que en ejercicios futuros se evalúen estos y otros enfoques complementarios para automatizar la clasificación del género.

Cuadro 3

Intervalos de confianza para la predicción del género a partir de los nombres [-1 masculino, +1 femenino]; la diferencia para los intervalos de confianza va de [0,1] a [0,2], febrero de 2019
(En porcentajes)

	Workana (n = 19,846)		Freelancer (n = 73,778)	
Intervalo para nombres desconocidos:	[-0,1, 0,1]	(-0,2, 0,2)	[-0,1, 0,1]	(-0,2, 0,2)
Masculino	58,6	55,5	68,1	62,7
Femenino	37,5	34,6	29,4	26,5
Desconocido	3,9	9,9	2,5	10,8
Total	100	100	100	100

Fuente: Elaboración propia.

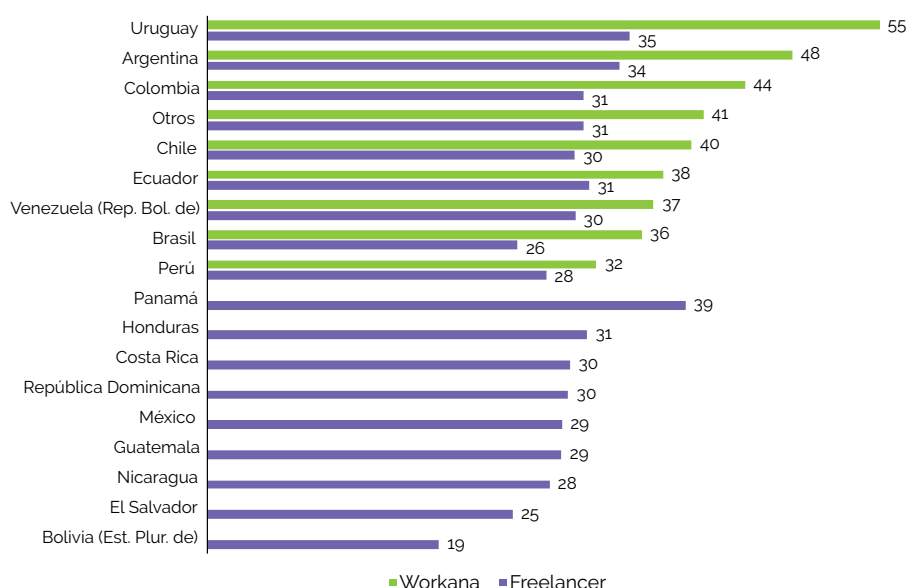
b. Desigualdades de género

En el gráfico 9 se muestran los porcentajes de mujeres por país para cada una de las plataformas. Podemos apreciar que hay una proporción más alta de profesionales mujeres en el mercado regional de la economía digital en Workana.com que en el mercado mundial de Freelancer.com. Las distribuciones de género por categoría de empleo confirman la previsión general de que el grupo de mujeres profesionales sigue reflejando los prejuicios de género de la división del trabajo en los sectores cualitativos de la economía digital, como demuestra el hecho de que aparezcan con la mayor representación en la categoría “Escritura y traducción”; y con la menor representación en el sector más cuantitativo, “Informática y programación” (véase el gráfico 10). Los datos de ambos portales confirman esta conclusión. Es interesante puntualizar que la categoría “Ingreso y administración de datos” exhibe un porcentaje relativamente elevado de mujeres profesionales en la región. Sin embargo, esto también podría responder a los sesgos introducidos por nuestro sistema de clasificación (véase el cuadro 2), que no delimita con demasiada claridad el alcance de algunas tareas generales de soporte administrativo.

Gráfico 9

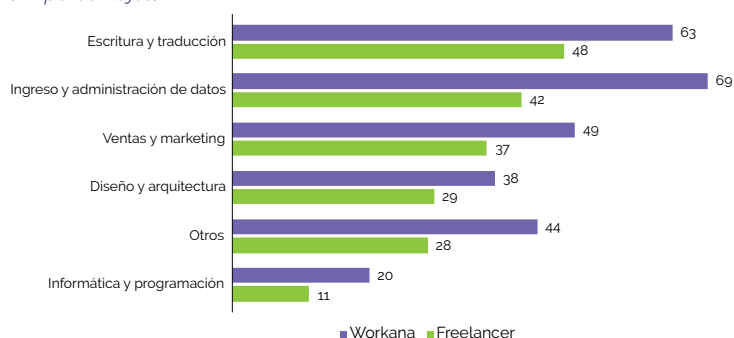
Proporción de mujeres profesionales, por plataforma y por país, febrero de 2019

(En porcentajes)

**Fuente:** Elaboración propia, sobre la base de datos de Freelancer y Workana.**Nota:** Otros: Freelancer: Cuba, Guadalupe, Guyana Francesa, Haití, Martinica, Paraguay, Puerto Rico y San Martín; Workana: Belice, Bolivia (Est. Plur. de), Costa Rica, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Puerto Rico, República Dominicana, Trinidad y Tabago.**Gráfico 10**

Proporción de mujeres profesionales, por categoría de empleo y plataforma, febrero de 2019

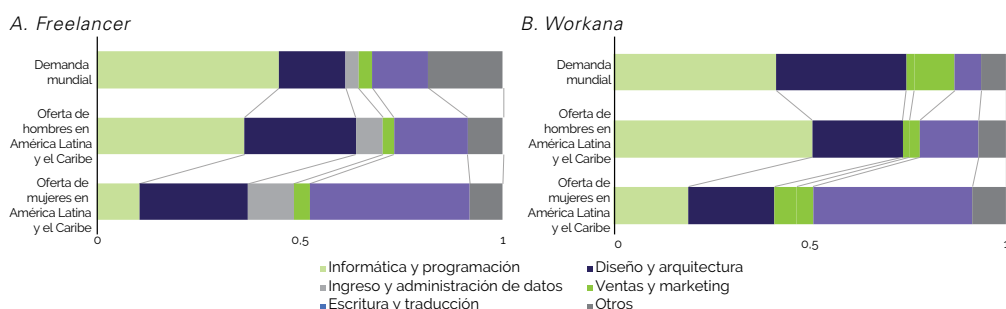
(En porcentajes)

**Fuente:** Elaboración propia, sobre la base de datos de Freelancer y Workana.**Nota:** Otros: ingeniería y fabricación, y servicios empresariales, contables, jurídicos y de recursos humanos.**c. Diferencias entre los géneros en la oferta y la demanda**

Ahora realizamos el mismo análisis en función del género. El gráfico 11 deja claro que la oferta de trabajadores de la economía digital por país está mucho mejor ajustada a la demanda de profesiones para los hombres que para las mujeres. En esencia, esto responde al hecho de que la proporción de hombres con habilidades en la categoría “Informática y programación”, la más demandada de las profesiones en la economía digital, es mayor. En el mercado de trabajo regional de Workana, la oferta de profesionales para la categoría “Informática y programación” es proporcionalmente aún más alta que la demanda, que también es en esencia regional. Claramente, en ambas plataformas la oferta de trabajadoras para la categoría “Escritura y traducción” supera a la demanda. A principios de 2019, la oferta de trabajadoras autónomas también superaba la demanda para la categoría “Ingreso y administración de gastos”.

Gráfico 11

Demanda mundial de empleo y oferta regional, por género

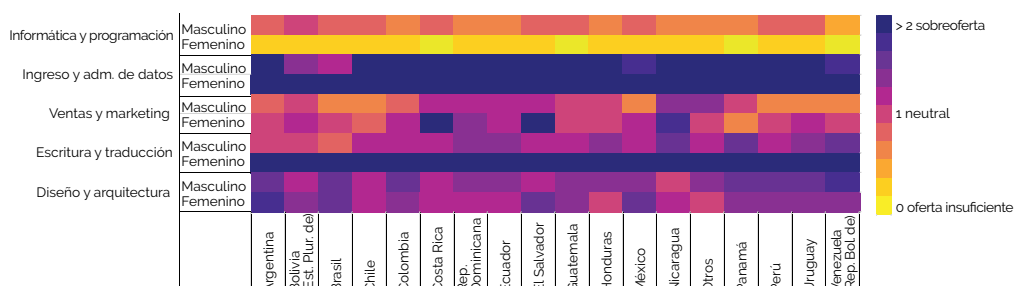


Fuente: Elaboración propia, sobre la base de datos de Freelancer y Workana.

La riqueza de los datos obtenidos se presta para producir una ingente cantidad de análisis y visualizaciones distintos. En el gráfico 12 se resume el ajuste entre la oferta y la demanda con una matriz cromática, donde las diferencias se marcan con un código de colores.

Gráfico 12

Mapa de calor de la demanda mundial de empleo y la oferta regional, por género y categoría de empleo, febrero de 2019 (índice)



Fuente: Elaboración propia, sobre la base de datos de Freelancer y Workana.

B. Demanda de trabajadores a tiempo completo

También recopilamos y analizamos datos de la huella digital proveniente de los portales donde se ofrecen principalmente empleos a tiempo completo. Se espera que estos resultados difieran frente a los de la economía de las ocupaciones transitorias, ya que las oportunidades publicadas se asemejan más a las ofertas de empleo tradicionales, los trabajos a menudo se ubican en el país, y las vacantes se llenan con mano de obra nacional.

1. Demanda de trabajadores a tiempo completo

Nuestra principal fuente panregional de este tipo es Bumeran, el portal de ofertas de empleo de tiempo completo más importante de América Latina; está presente en 7 países, todos los meses recibe 12 millones de visitas y publica más de 60.000 avisos, nuclea a más de 13 millones de candidatos —a los que se agregan unos 260.000 nuevos

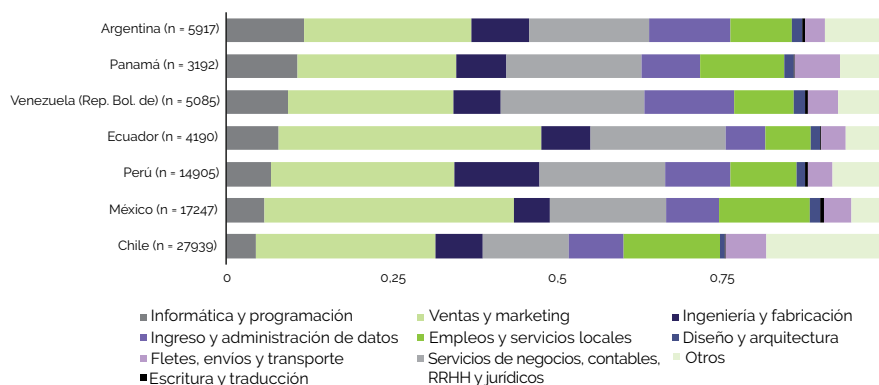
todos los meses—, y tiene unos 50.000 clientes empresariales por año (se conoce como laborum.com en Chile). Los datos que recopilamos permiten confirmar que alrededor del 90% de los trabajos en Bumeran son empleos de tiempo completo, en tanto que los trabajos por contrato o temporales (el mercado que abarcan los dos portales para trabajadores autónomos ya analizados) representan apenas alrededor de un 1%.

El gráfico 13 muestra las categorías de empleo de los cargos que ofrece la economía digital por país. Es llamativo comparar estas distribuciones con las distribuciones análogas en la economía de las ocupaciones transitorias (por ejemplo, el gráfico 5). Los empleados de tiempo completo en la región hacen muchas menos búsquedas en las categorías de “Informática y programación” y “Escritura de traducción”, y muchas más en la de “Ventas y marketing” y otros cargos administrativos empresariales. También pueden observarse diferencias entre los países. Por ejemplo, en Chile y en el Perú hay una mayor demanda por la categoría de “Ingeniería y fabricación” (que en el análisis de Freelancer.com se incluyó la sección “Otros” debido a lo pequeño de la muestra en ese mercado), que en países como el Ecuador, México y Panamá.

Nuevamente, es importante puntualizar que no es necesario que las categorías de las diferentes fuentes de datos estén perfectamente alineadas (véase el cuadro 2). Pero más allá de eso, lo que podría estar ocurriendo es simplemente que los usuarios de países y contextos culturales diferentes podrían entender de manera distinta lo que implica una destreza como el diseño. Por supuesto, esto es similar a lo que sucede en el caso de las estadísticas tradicionales basadas en encuestas.

Gráfico 13

Empleo por categoría de trabajo de la economía digital en Bumeran, febrero de 2019
(En porcentajes)



Fuente: Elaboración propia, sobre la base de datos de Bumeran.

2. Análisis de las descripciones de los requisitos de empleo

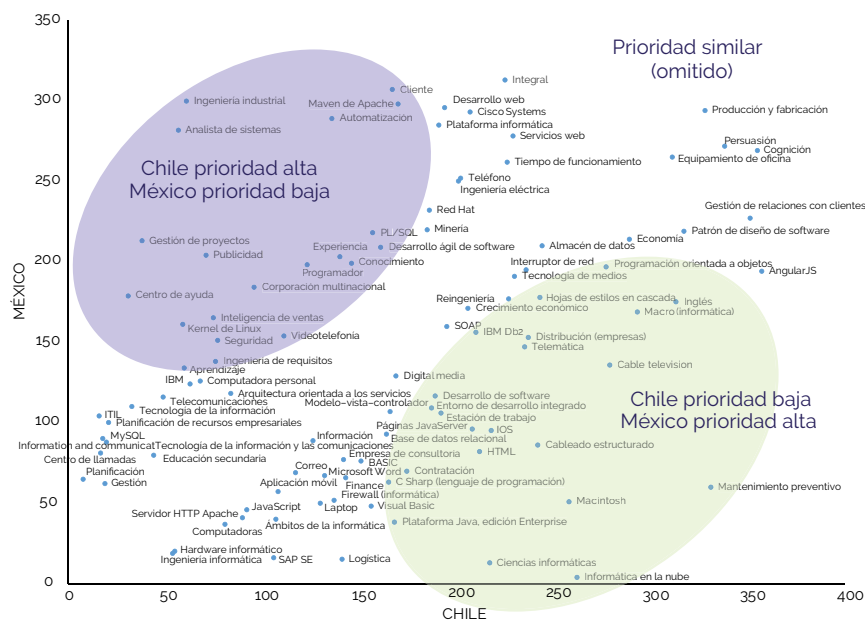
Los datos sobre la huella digital permiten aprovechar fuentes de datos no tradicionales, como los textos cualitativos y las imágenes. Ambas categorías tienen una abundante presencia en Internet, y pueden convertirse en cifras mensurables con la ayuda de diversas herramientas informáticas. Ya hemos analizado el uso de imágenes para clasificar aspectos de género, y ahora procederemos a examinar los textos de las descripciones de cargos en Bumeran.

El método más sencillo sería crear lo que se denomina una “bolsa de palabras”, que en esencia permite contar las veces que aparece cada palabra y destaca las más frecuentes. Más allá de la simplicidad de este sistema, aplicamos al texto obtenido un análisis semántico automatizado de aprendizaje por computadora que extrae el significado de los anuncios de trabajo. Utilizamos una plataforma llamada TextRazor, que se vale de modernas técnicas de procesamiento natural del lenguaje para estudiar sintácticamente el contenido, analizarlo y extraer metadatos semánticos a partir de él. En particular, utilizamos el etiquetador temático de TextRazor (que originalmente fue entrenado con millones de páginas de Wikipedia) para asignar temas de alto nivel al contenido, sobre la base de las relaciones entre las entidades y las categorías de palabras. Por ejemplo, si un anuncio de empleo solicita habilidades en “Servicios web de Amazon, virtualización”, el análisis semántico devuelve el término “plataforma en la nube” como la etiqueta temática que mejor se ajusta al caso. Si un anuncio de empleo solicita habilidades en “atención al cliente, soporte, mantenimiento, hardware”, el sistema de inteligencia artificial entrenado con los contenidos de Wikipedia devuelve “servicio de asistencia a usuarios” como la etiqueta temática que mejor se ajusta a la situación. TextRazor entiende automáticamente cientos de miles de diferentes temas según diferentes niveles de abstracción y funciona en 10 idiomas, entre ellos español y portugués. Esto nos permite convertir enormes cantidades de datos cualitativos no estructurados en datos cuantitativos que pueden interpretarse.

En el gráfico 14 se muestran las exigencias laborales establecidas a través del estudio semántico en el ámbito de la tecnología en Chile y en México. En total, la inteligencia artificial detectó 358 categorías de empleo distintas en las publicaciones en Chile y 325 en México, de las cuales 172 eran comunes a ambos países. Comparamos la prioridad detectada para cada tema y seleccionamos las 100 categorías comunes que guardaban la menor semejanza entre ambos países. El resultado corrobora la conclusión anterior y demuestra que en Chile, para los trabajos de perfil tecnológico, a menudo se exigen habilidades gerenciales y de negocios (como las relacionadas con el “cliente”, la “gestión de proyectos”, la “publicidad” y la “inteligencia de ventas”), mientras que en las ofertas de trabajo en México se da mayor importancia a las habilidades técnicas (por ejemplo, “informática en la nube”, “plataforma JAVA”, “HTML” y “televisión por cable”). Del análisis también surge que saber inglés es una exigencia más frecuente en México que en Chile. En estudios posteriores, sería interesante vincular estas percepciones con trabajos estadísticos tradicionales complementarios.

Gráfico 14

Análisis semántico de las ofertas de empleo de Bumeran publicadas en Chile y México (laborum.cl, N = 1.907 publicaciones; bumeran.com.mx, N = 1.640 publicaciones): los ejes representan posiciones (los números altos son posiciones bajas)



Fuente: Plaboración propia, sobre la base de datos de Bumeran.

Un requisito para el desarrollo de la economía digital son los equipos tecnológicos necesarios. Es un hecho ampliamente aceptado que el principal obstáculo para el acceso digital en la región es el costo de la tecnología (CEPAL, 2013; Hilbert, López y Vásquez, 2010). Consultamos con la plataforma panregional MercadoLibre Inc., una empresa argentina que opera portales de venta de productos y servicios por Internet y se dedica al comercio electrónico y a las subastas en línea en 18 países latinoamericanos. En 2018, la empresa tenía 7.239 empleados (2.409 pertenecientes a su división de tecnología de la información y desarrollo de productos) e informó ingresos netos de 1.440 millones de dólares (MercadoLibre, 2018).

MercadoLibre ofrece productos nuevos y usados. Nosotros nos centramos exclusivamente en los productos nuevos, dado que eso nos brinda un punto de partida común sobre su calidad. Las categorías de productos que ofrece MercadoLibre difieren según el país, lo que a la larga aumenta considerablemente las diferencias de clasificación de productos entre los países. A los efectos de nuestro objetivo de armonización internacional, el aspecto más importante es la manera en que los usuarios de distintos países realmente utilizan las diferentes categorías, una situación que plantea problemas de calidad de los datos que pueden solucionarse mediante inspecciones exhaustivas a cargo de seres humanos.

A. Índice de precios de los productos tecnológicos

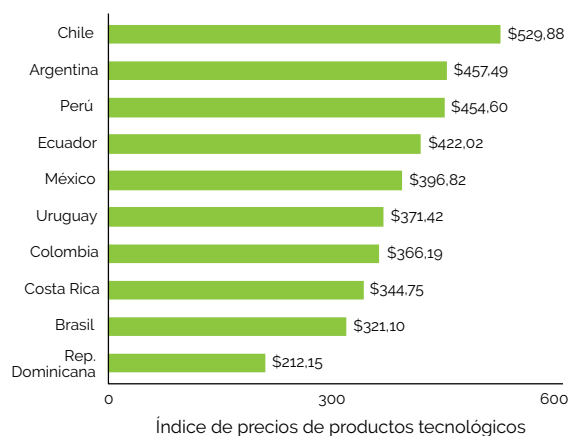
En total, recabamos datos de 192.092 productos de 10 países y 7 sectores, a saber, teléfonos móviles, computadoras portátiles y tabletas, computadoras y servidores, procesadores, insumos inalámbricos y de redes, software y drones. Tras un prolongado y detallado análisis, elaboramos un índice agregado calculando el promedio simple de los precios promedio por categoría. Esto da un peso relativamente mayor a las categorías más pequeñas, como los drones (véase el gráfico 16). Como siempre, la conformación de un índice depende del objetivo del investigador y a menudo predetermina al resultado (Minges, 2005). En el gráfico 15.A se muestra que en valores nominales, el precio promedio de los productos tecnológicos claramente es más alto en algunos países que en otros. Esto responde a las características de los mercados de cada país, por ejemplo, su grado de desarrollo, el nivel de competencia, la estructura tributaria y el poder de compra, entre otros factores.

En el gráfico 15.B se compara el índice de precios de los productos tecnológicos con el ingreso nacional per cápita (según la paridad del poder adquisitivo, o PPA), sobre la base del costo de oportunidad de adquirir otros productos en el país. Pueden verse tres grupos. En el Ecuador y en el Perú, los precios de la tecnología son relativamente altos frente al ingreso nacional. Colombia se está acercando a este grupo. En la Argentina, el Brasil, Chile, Costa Rica y México la asequibilidad de la tecnología se encuentra en un nivel medio. En la República Dominicana y en el Uruguay los precios de la tecnología son claramente más bajos que en el resto de la región, si se tiene en cuenta el salario promedio. Esto brinda un índice de precios útil de la asequibilidad de la tecnología, basado en una amplia gama de opciones que pueden actualizarse en tiempo real.

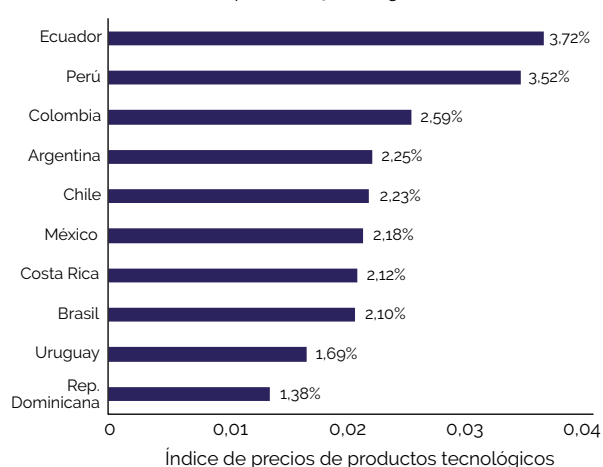
Gráfico 15

Índice de precios de los productos tecnológicos por país, febrero de 2019

A. Valores nominales (En dólares)



B. En términos relativos como por-cen-ta-je del ingreso nacional bruto (INB, PPA), per cápita

**Fuente:** Elaboración propia, sobre la base de datos de mercadolibre.com.

En el gráfico 16 se contrastan los precios absolutos y relativos por sector de tecnología (es decir, se compara el gráfico 15.A con el gráfico 15.B). Nuevamente, se constata que para la mayoría de los sectores (drones; computadoras y servidores; computadoras portátiles y tabletas; teléfonos móviles y teléfonos inteligentes; procesadores; software; productos de red) la tecnología es más costosa en Chile (en valores nominales en dólares, lo que se traduce en un desplazamiento hacia la derecha en el eje horizontal x). Sin embargo, en términos relativos, y tomando en cuenta su poder de compra (hacia arriba y hacia abajo en el eje vertical y), el único sector en que Chile es el país más costoso es el del software (situación que podría estar relacionada con los derechos de propiedad intelectual), en tanto que para las categorías de teléfonos móviles, drones, computadores portátiles y tabletas y productos de red, el país es uno de los más asequibles. En términos relativos, nuevamente el Ecuador y el Perú se destacan como dos países donde varios tipos de tecnologías son bastante costosos frente a los índices nacionales del poder de compra. El precio promedio de un teléfono móvil en el Ecuador y en el Perú equivale casi al 3 % del ingreso nacional bruto per cápita anual (PPA³). Es interesante puntualizar que en el contexto del poder nacional de compra, los habitantes de la República Dominicana son

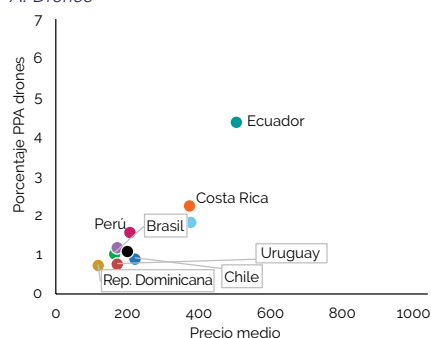
³ Sin ajustar el INB per cápita según la PPA, este porcentaje alcanza al 3,5%.

quienes pagan menos en la región por la mayoría de las tecnologías, incluidas las computadoras portátiles y las tabletas, los drones, las computadoras y los servidores, los teléfonos móviles y los teléfonos inteligentes y los procesadores.

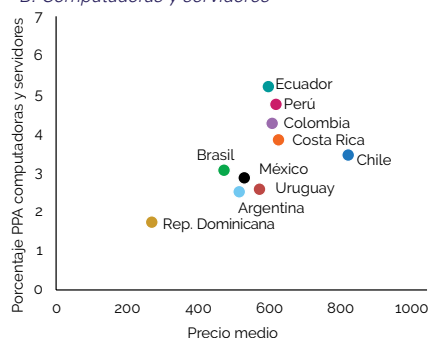
Gráfico 16

Precios medios de los productos tecnológicos por sector (eje horizontal x) frente a los precios medios normalizados según el INB per cápita ajustado por el PPA, febrero de 2019
(En dólares)

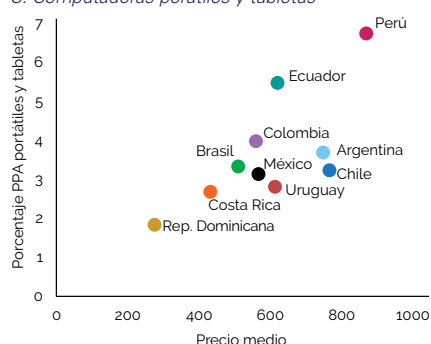
A. Drones



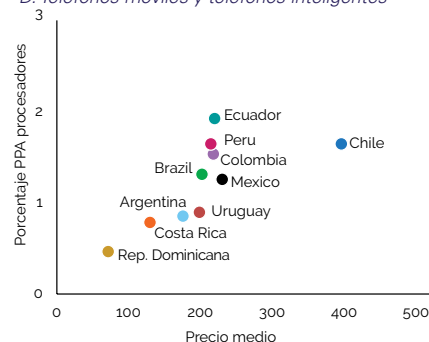
B. Computadoras y servidores



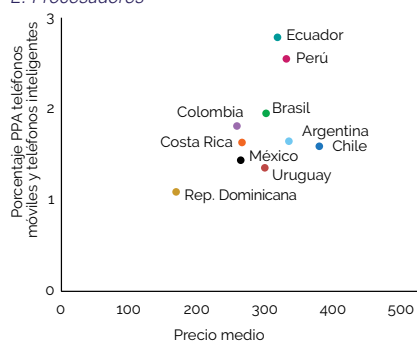
C. Computadoras portátiles y tabletas



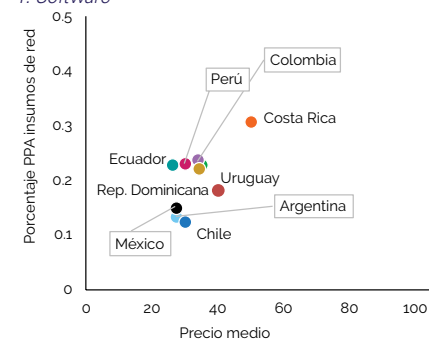
D. Teléfonos móviles y teléfonos inteligentes



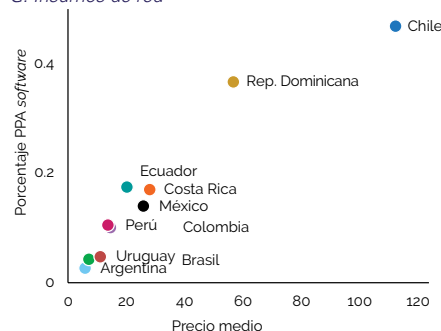
E. Procesadores



F. Software



G. Insumos de red



Fuente: Elaboración propia, sobre la base de datos de MercadoLibre.

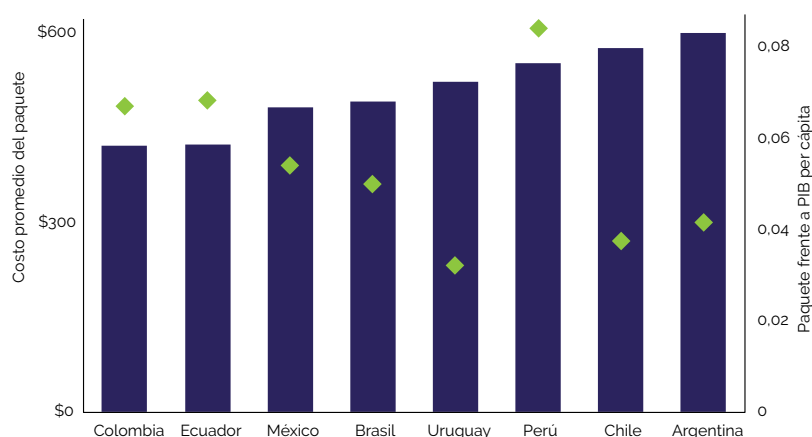
Al combinar los diferentes porcentajes, se llega a la conclusión de que, en promedio, se necesitaría alrededor del 12% del ingreso regional per cápita para que todos los habitantes de la región accediesen a una solución de gama media en cada uno de los siete tipos de tecnologías que se analizan aquí. Esto demuestra las posibilidades que ofrece generar cestas de precios de referencia y nos lleva a nuestro siguiente enfoque

B. Precios por paquete de productos

Realizamos un análisis de seguimiento y examinamos con mayor profundidad determinadas tecnologías, para lo que creamos una cesta compuesta por un conjunto de dispositivos de acceso: una computadora portátil (Acer Aspire), una tableta (Apple iPad) y un teléfono móvil (Samsung Galaxy J2). Es importante destacar que este tipo de análisis depende de la tecnología elegida y que existen varios factores que pueden generar confusión. Además, la rápida evolución de la tecnología hace que cualquier opción adecuada pronto deje de serlo. Sin embargo, si se aplica la hipótesis del mercado eficiente a los mercados en línea, por lo menos un puñado de dispositivos debería acercarse al precio nacional promedio.

Gráfico 17

Precio del paquete por país seleccionado, febrero de 2019
(En dólares)



Fuente: Elaboración propia, sobre la base de datos de MercadoLibre.

Tomando como referencia una cesta básica de productos tecnológicos, se observa que en los países seleccionados el costo de este conjunto de dispositivos va de 400 a 600 dólares. Confrontar estos valores con el nivel de ingreso per cápita permite determinar en qué medida la población puede acceder a dichas tecnologías. En ese sentido, esta cesta de productos es más asequible para los habitantes del Uruguay y Chile, para quienes representa un 3% y un 4% del PIB per cápita, respectivamente. Siguen la Argentina y México con un 5%, Colombia y el Brasil con un 6%, el Ecuador con un 7%, y finalmente el Perú, donde el costo de la canasta representa un 8% del ingreso per cápita.

Las pequeñas y medianas empresas (pymes) en América Latina constituyen un grupo heterogéneo de agentes económicos que contribuyen considerablemente a la generación de empleo en la región y a su producto interno bruto (Dini y Stumpo, 2011; Peres y Stumpo, 2002; Vallejos, 2003). En la mayoría de los países, las pymes representan entre un 20% y un 50% del empleo, y entre un 15% y un 40% de las ventas (Ferraro y Stumpo, 2010). Pese a su importancia en la región, es relativamente poco lo que se sabe sobre este importante sector de la economía, en particular frente a las empresas más grandes. Esta falta de información podría verse reflejada en la adopción de políticas obsoletas, tardías o inadecuadas. En la economía digital, esta situación se ha visto agravada por el advenimiento de nuevos mercados para agentes de pequeño y mediano porte. Por otro lado, esta nueva tendencia también presenta una oportunidad, por cuanto los actores de la economía digital dejan tras de sí una huella digital que puede usarse para estudiarlos.

Nos hemos centrado en dos fuentes de datos. MercadoLibre es la plataforma de ventas minoristas más importante de la región (MercadoLibre, 2018). En nuestro estudio incluimos a 2.473.977 vendedores de 18 países. También analizamos las plataformas de financiación colectiva, en particular kiva.org, una organización sin fines de lucro mediante la cual las personas pueden prestar dinero a través de Internet a empresarios de bajos ingresos en más de 80 países. Analizamos 1.784.119 proyectos de préstamo en Kiva, y 2.400 proyectos de kickstarter.com.mx en México.

A. Minoristas en línea

MercadoLibre.com brinda un conjunto de interfaces API de desarrollo que permiten acceder a datos sobre la mercadería y los vendedores. Sin embargo, para obtener información sobre los vendedores, se necesita la identidad del vendedor definida por MercadoLibre. Cabe destacar que no alcanza con recopilar los datos de todos los vendedores, ya que muchos podrían no haber vendido nada nunca, o haber estado inactivos por largos períodos. Por ende, necesitamos una lista representativa de las identidades de los vendedores que actualmente están activos en la región. Esto nos obliga a adoptar algunas decisiones metodológicas fundamentales, principalmente relacionadas con la toma de muestras de los vendedores.

Para conformar nuestra muestra de la base de datos de vendedores, elaboramos una lista de los artículos actualmente más vendidos e identificamos a sus vendedores. Específicamente, nuestro procedimiento de recopilación de datos consta de tres pasos. 1) Obtener el conjunto de datos sobre la mercadería. Decidimos tomar una muestra de cada categoría de productos disponibles. Dentro de los 18 países de América Latina y el Caribe presentes en MercadoLibre, hay 239.771 categorías de productos (véase el cuadro 4). El algoritmo por defecto para clasificar la mercadería por categoría de producto en la API se denomina “Más relevante”, y tiene dos opciones disponibles: “Menor precio” y “Mayor precio”. Incluimos los 1.000 productos más “relevantes” para cada categoría por país, para un total de 53.425.985 productos, de los cuales 43.108.698 eran nuevos, y el resto usados. Cabe destacar que en algunos países, muchas categorías contenían menos de 1.000 productos, o ninguno, lo que explica por qué recuperamos

menos que los 239.771 x 1.000 productos teóricamente posibles (alrededor del 22% del total posible). 2) Del conjunto de datos recogido sobre la mercadería obtuvimos la identidad de los vendedores, y tras eliminar las identidades duplicadas, llegamos a un total de 2.473.977 vendedores únicos de 18 países. Finalmente, 3) utilizamos las identidades de los vendedores para recabar de la API información sobre ellos.

En el cuadro 4 se muestran los vendedores de los artículos más relevantes vendidos a principios de marzo de 2019, en una muestra tomada a partir de todas las categorías de productos disponibles. Los conjuntos de datos obtenidos a través de otras metodologías de muestreo son distintos, y los resultados también podrían serlo. De entre los 2.473.977 vendedores, centramos nuestro análisis en los vendedores que la plataforma describe como “normales” y “activos”, y excluimos a los vendedores internacionales, a las grandes marcas, a las concesionarias de vehículos, a las franquicias y a los vendedores de bienes inmuebles.

Cuadro 4

Recopilación de datos para conformar la muestra de los vendedores en MercadoLibre

	Número de categorías de productos	Número de artículos	Número de vendedores
Argentina	44 765	8 147 807	562 869
Bolivia (Estado Plurinacional de)	424	15 123	842
Brasil	65 506	19 953 271	683 344
Chile	19 559	4 265 428	149 614
Colombia	21 360	4 928 705	152 054
Costa Rica	1 506	189 754	11 695
República Dominicana	691	146 099	7 539
Ecuador	1 713	431 893	36 189
Guatemala	512	18 951	1 322
Honduras	184	5 694	292
México	39 074	7 895 614	365 733
Nicaragua	176	2 190	220
Panamá	458	34 965	2 131
Paraguay	325	11 728	622
Perú	8 439	1 500 016	109 279
El Salvador	264	7 974	628
Uruguay	23 360	3 295 392	122 573
Venezuela (República Bolivariana de)	11 455	2 575 381	267 031
Total	239 771	53 425 985	2 473 977

Fuente: Elaboración propia, sobre la base de datos de mercadolibre.com.

Desde una perspectiva práctica, efectuar cálculos con 53,4 millones de productos podría ser dificultoso para varias de las tecnologías más populares de manejo de bases de datos (por ejemplo, el popular programa de manejo de planillas Microsoft Excel tiene una capacidad máxima de 1 millón de filas). Los datos obtenidos a partir de la API de MercadoLibre están en formato notación de objetos de JavaScript (JSON). Para almacenar los datos en formato JSON, utilizamos una base de datos NoSQL orientada a documentos llamada MongoDB, y obtuvimos las identidades de los vendedores a través de un código Python con la asistencia del sistema MongoDB. La duplicación de las identidades de los vendedores se calculó utilizando la biblioteca estándar Pandas de Python.

1. Localización geográfica

La distribución geográfica de los vendedores puede determinarse buscando coincidencias entre el nombre de la ciudad y del país del vendedor (información que provee MercadoLibre) y la API de geolocalización de Google, para de esa forma obtener las coordenadas geográficas. Los siguientes son mapas de densidad generados en Qlik.com. En el gráfico 18 se muestra la distribución de los vendedores en MercadoLibre en los países de la región. El mapa de calor permite concluir que el uso de este tipo de plataforma de comercio electrónico está generalizado, y que no se limita a las capitales nacionales, sino que también se extiende a las zonas interiores de los países. Esto mejora el acceso a los mercados y aumenta las opciones de consumo. El mapa también refleja una elevada concentración de vendedores en las zonas aledañas a Sao Paulo y Rio de Janeiro, un hecho previsible si se tiene en cuenta que se trata de la primera y la quinta ciudad más grandes de la región, respectivamente. De manera en cierto modo inesperada, en el gráfico 18 también se observa una elevada concentración en Centroamérica, con epicentro en Costa Rica, que concentra 11.665 vendedores, la mayoría en San José.

Mapa 1

Distribución geográfica de vendedores en MercadoLibre



Fuente: Elaboración propia, sobre la base de datos de mercadolibre.com.

Nota: Los límites y los nombres que figuran en este mapa no implican su apoyo o aceptación oficial por las Naciones Unidas.

2. Concentración del mercado

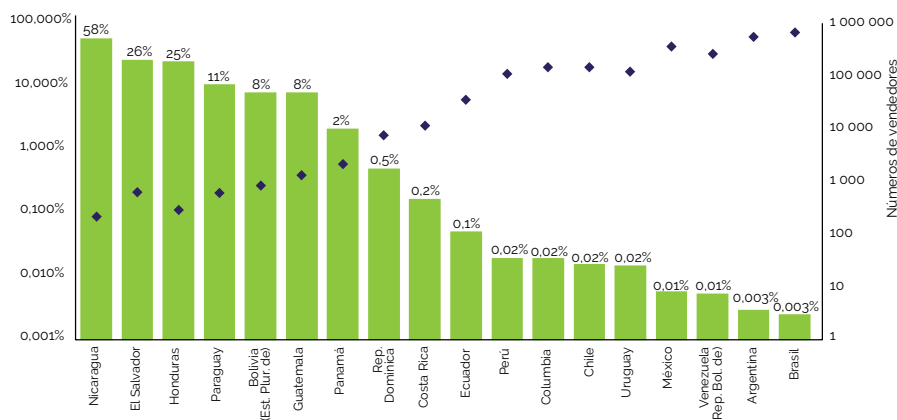
En el gráfico 18 se muestra la concentración entre los vendedores por país. Las barras ilustran la cuota de mercado de los principales cinco vendedores en función del número de transacciones realizadas en MercadoLibre (total histórico). En el eje derecho del gráfico se muestra el logaritmo del número total de vendedores. Es

una representación más fidedigna de la concentración del mercado en función de la distribución de las transacciones (una forma simplificada del frecuentemente utilizado índice de Herfindahl-Hirschman).

Cuando se analiza la relación entre los indicadores, queda claro que cuantos más vendedores hay en un país, y cuanto más grande es el país, menor es la concentración de transacciones por vendedor, y menor es el predominio de los principales cinco vendedores. Por ejemplo, en Nicaragua, que tiene 220 vendedores, los cinco vendedores con el máximo histórico de transacciones representan un 58% del total del mercado, en tanto que en Brasil, que tiene 673.528 vendedores, los cinco principales representan apenas el 0,003% del total de transacciones realizadas. Esto demuestra que propiciar el acceso a estas nuevas oportunidades de negocios, que parecen traducirse de manera bastante directa en oportunidades económicas, es sumamente importante. Si hubiera más vendedores en los países pequeños, la cuota del mercado que representan los cinco vendedores principales disminuiría. Dadas las limitaciones de acceso que enfrentan los micro, pequeños y medianos vendedores, es fácil que los mercados en línea queden bajo el dominio de unos pocos agentes relativamente grandes.

Gráfico 18

Concentración de mercado de vendedores en MercadoLibre: representación logarítmica de la cuota de mercado de los principales cinco vendedores en función del total histórico de productos vendidos (barras), número de vendedores por país (eje derecho), total histórico (en porcentaje [barras] y recuentos [eje derecho])
(En porcentaje [barras] y recuentos [eje derecho])



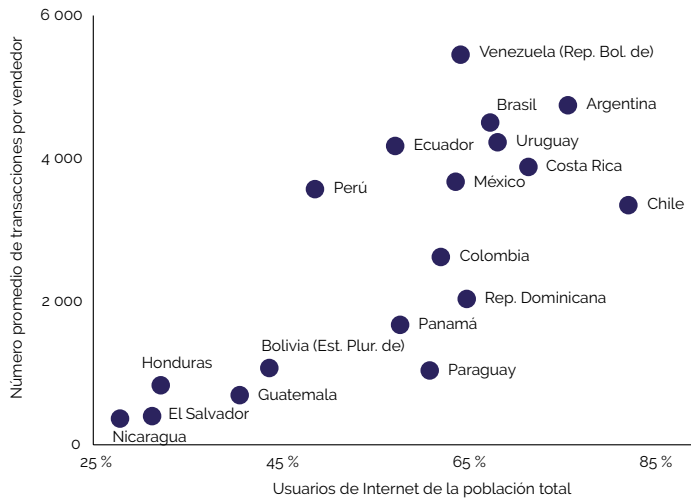
Fuente: Elaboración propia, sobre la base de datos de mercadolibre.com.

3. Acceso a las transacciones

Los resultados del gráfico 19 permiten llegar a la misma conclusión, esta vez desde una perspectiva complementaria. El gráfico relaciona la cantidad de usuarios de Internet como porcentaje de la población total con el número promedio de transacciones ejecutadas por vendedor (total histórico). Con el aumento en la penetración de Internet, se ha incrementado el número de transacciones en línea, no solo en el total por país, sino también por vendedor. Esto sugiere que la penetración de Internet aumenta el dinamismo del mercado minorista en línea y la actividad económica por agente económico, probablemente debido al crecimiento del mercado de potenciales consumidores. Esto pone de relieve el hecho de que la brecha digital en lo referido al acceso a Internet sigue siendo uno de los principales obstáculos para avanzar hacia una economía digital vibrante en América Latina.

Gráfico 19

Penetración de Internet (ITU, 2017) frente al número promedio de transacciones por vendedor por país: totales históricos



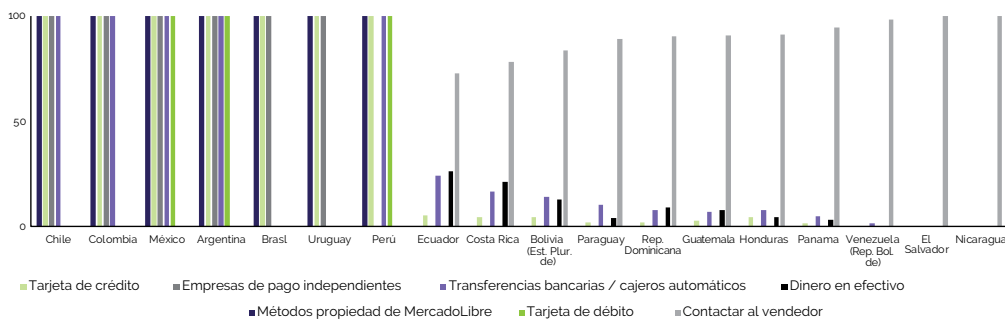
Fuente: Elaboración propia, sobre la base de datos de mercadolibre.com.

4. Facilitadores de las transacciones

En el gráfico 20 se analiza un importante facilitador de las transacciones digitales: las opciones de pago. Las plataformas de MercadoLibre en algunos países, como la Argentina, el Brasil, Chile, Colombia, México, el Perú y el Uruguay, ofrecen una considerable variedad de opciones de pago, mientras que en la mayoría de los países más pequeños, los compradores deben usar dinero en efectivo o comunicarse con el vendedor. Es claro que esa situación obstaculiza el desarrollo del comercio electrónico. Al comparar este gráfico con la cuota de mercado de los cinco vendedores principales (véase el gráfico 19), puede observarse que cuanto menor es la cantidad de opciones de pago, más se concentra el mercado en unos pocos vendedores, que obtienen la mayoría de los beneficios.

Gráfico 20

Opciones de pago disponibles en MercadoLibre por país, enero de 2019
(En porcentajes)



Fuente: Elaboración propia, sobre la base de datos de mercadolibre.com.

B. Financiación colectiva

Kiva es una popular plataforma de financiación cooperativa dirigida especialmente a proyectos ejecutados en países en desarrollo. Su misión es “conectar a las personas a través de préstamos para aliviar la pobreza” (Waghorn, 2013). Mediante su sistema de financiación cooperativa, la plataforma ha permitido concretar 1,7 millones de préstamos con un índice de repago de más del 98% (Price, 2017), y desde 2013 logra recaudar alrededor de 1 millón de dólares cada tres días (Waghorn, 2013). Kiva funciona en 101 países, 22 de ellos en América Latina y el Caribe, donde se han llevado a cabo 386.498 de los 1.551.367 de proyectos que se han financiado en todo el mundo (alrededor de la cuarta parte)⁴. Al momento de recopilar los datos (marzo de 2019), seleccionamos 1,5 millones de proyectos efectivamente financiados de los 1,7 millones que se enumeran en el sitio⁵.

1. Recaudación de fondos por sector

El gráfico 21 permite constatar que la mayoría de los proyectos en la región se relacionan con la agricultura, la alimentación y el comercio minorista. Dichas prioridades son similares en África y en Asia, pero difieren en América del Norte (donde la agricultura cede lugar al transporte) y en Europa (donde la vivienda supera al entretenimiento). Algunas excepciones notables son la participación proporcionalmente más alta de proyectos de vestimenta en la República Dominicana, de salud en México, y de educación en el Paraguay.

Gráfico 21

Proporciones de proyectos por sector en Kiva.org. (2006-2018)

(En porcentajes)



Fuente: Elaboración propia, sobre la base de datos de kiva.org.

⁴ Proyectos por país (totales históricos) Perú: 89.386; El Salvador: 60.423; Nicaragua: 42.100; Ecuador: 38.291; Colombia: 36.732; Paraguay: 28.010; Bolivia (Est. Plur. de): 24.131; México: 18.859; Honduras: 17.066; Guatemala: 13.691; Haití: 6.160; Costa Rica: 4.856; República Dominicana: 4.280; Brasil: 875; Chile: 873; Suriname: 265; Panamá: 228; Belice: 180; Puerto Rico: 92; Asia Oriental y el Pacífico: 495.569; África Subsahariana: 422.556; Europa y Asia Central: 103.904; Asia Meridional: 77.473; Medio Oriente y Norte de África: 58.379; América del Norte: 6.988.

⁵ No se incluyen los proyectos que aparecen como “vencidos”, “recaudando fondos” o “restituidos” (dinero devuelto).

2. Recaudación de fondos en el contexto mundial

En el gráfico 22 puede apreciarse que los empresarios de América Latina y el Caribe hacen un uso sumamente eficaz de las redes mundiales de financiación colectiva para la microempresa. La región está a la vanguardia en los sectores de la alimentación, el comercio minorista, la agricultura, la vestimenta, los servicios, las artes, la construcción, la salud y la fabricación. En 9 de los 15 sectores, la región de América Latina y el Caribe lleva la delantera. Ninguna otra región del mundo recauda un monto promedio mayor por proyecto en los diferentes sectores mediante la modalidad de financiación colectiva. En el gráfico 23.A puede observarse que esta tendencia se ha sostenido durante el pasado decenio. Eso podría dar a entender que en los países de la región hay un buen potencial para el desarrollo del sector de las tecnofinanzas, en particular las plataformas de créditos o préstamos.

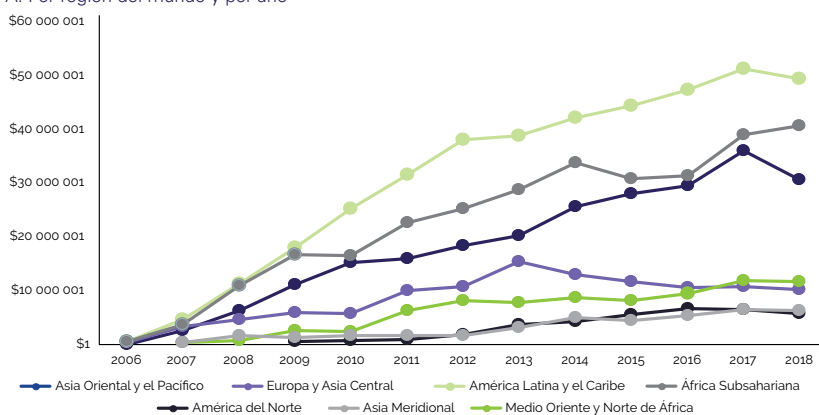
El gráfico 22.C permite observar que existe una gran diversidad entre los países en lo referido a los sectores que reciben los promedios de financiación más altos y más bajos. Es interesante señalar que en Honduras y la República Dominicana se destinan montos superiores al promedio a los proyectos de educación (Colombia registra el promedio más bajo), mientras que los sectores de la salud y la agricultura en Chile y el Brasil, respectivamente, reciben los promedios más elevados de financiación colectiva; los proyectos destinados a fines personales son a menudo los que reciben el promedio más bajo.

Gráfico 22

Monto total de financiación recibida, 2006-2019

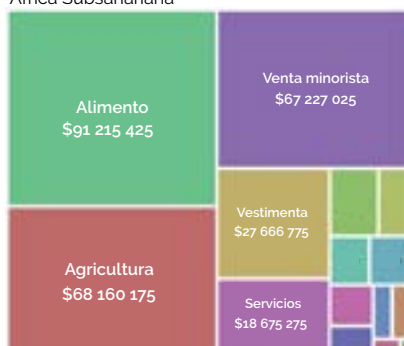
(En dólares)

A. Por región del mundo y por año

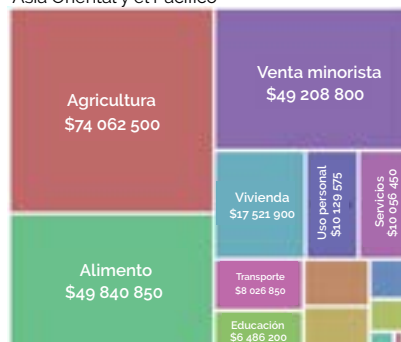


B. Por región del mundo y por sector

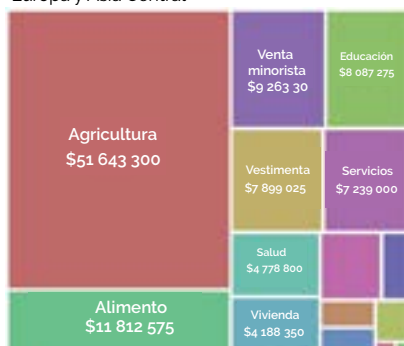
África Subsahariana



Asia Oriental y el Pacífico



Europa y Asia Central



América del Norte



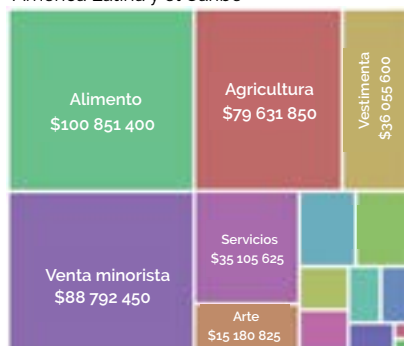
Asia Meridional



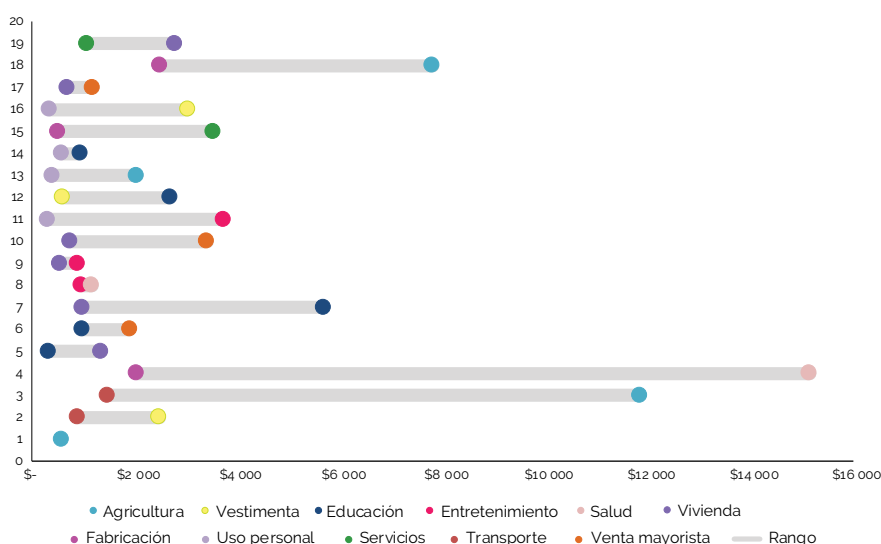
Medio Oriente y Norte de África



América Latina y el Caribe



C. Monto promedio más alto y más bajo financiado por sector y por país



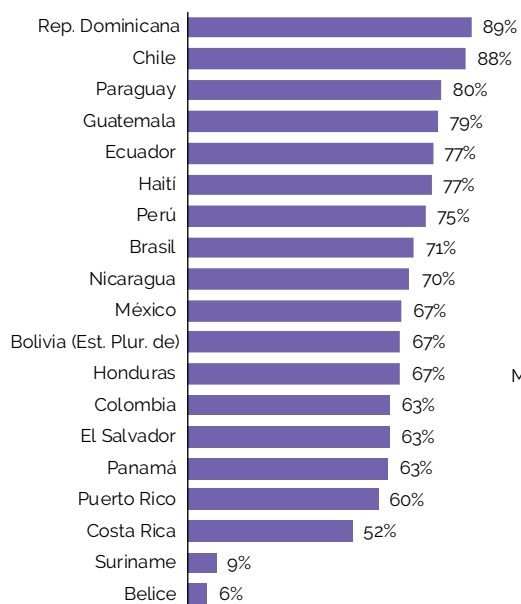
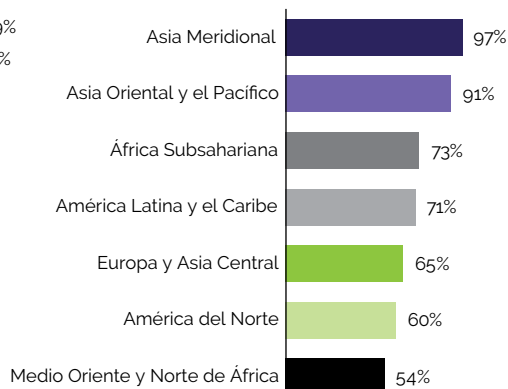
Fuente: Elaboración propia, sobre la base de datos de kiva.org.

3. Distribución por sexo

En la base de datos de Kiva también hay disponible información sobre el género de los empresarios. A menudo, los proyectos están a cargo de equipos. Si un equipo consta exclusivamente de mujeres o de hombres, codificamos la composición de género del equipo con una variable ficticia de 0 o 1, respectivamente. Si hay un hombre y cuatro mujeres, asignamos el valor 0,2, y sumamos los valores resultantes para todos los equipos para obtener el promedio nacional. El gráfico 24 permite constatar que la mayoría de los microempresarios de las plataformas de colaboración masiva de la región son mujeres. En Chile, la República Dominicana y el Paraguay, las empresarias representan más del 80% del total, una cifra similar a la de Asia. Con la excepción de las pequeñas muestras de Belice ($n = 180$ proyectos) y Suriname ($n = 265$ proyectos), el número de microempresarias activas es mayor en todos los países, y la mayoría también tienen un número mayor de empresarias activas en Kiva que otras regiones del mundo, entre ellas América del Norte y Europa. Esto demuestra que promover el uso de estas herramientas digitales ofrece una posibilidad de luchar contra las conocidas desigualdades de género actuales.

Gráfico 23

Porcentaje de proyectos liderados por mujeres en Kiva, 2006-2018

*(En porcentajes)**A. Por país en América Latina y el Caribe**B. Por región mundial***Fuente:** Elaboración propia, sobre la base de datos de kiva.org.

VELOCIDAD DE LA BANDA ANCHA EN LA REGIÓN Y EN EL MUNDO

IV

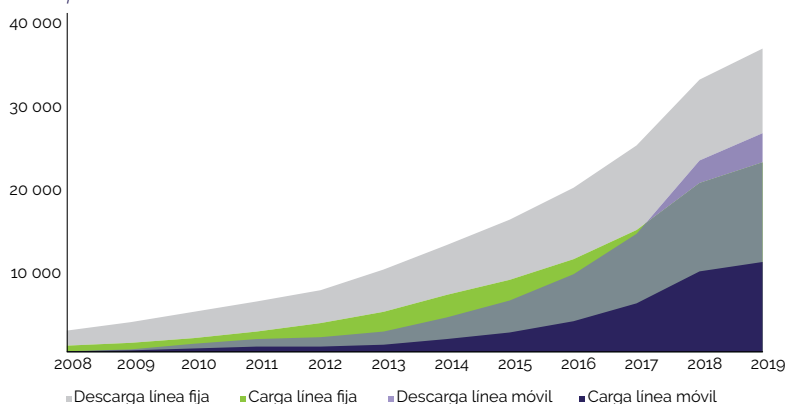
Las siguientes estimaciones de las velocidades de la banda ancha se ajustan la metodología desarrollada por (Hilbert, 2014 y 2016; Hilbert, López y Vasquez, 2010), y se basan principalmente en los datos de las pruebas de velocidad que recopila Ookla (Ookla, 2019). En otras palabras, reflejan los resultados obtenidos en pruebas normalizadas de velocidad realizadas por usuarios de todo el mundo. Las estimaciones históricas anteriores a 2018 proceden de (Hilbert, 2019). Mucho podría decirse sobre los siguientes gráficos, pero para ello debería elaborarse un informe aparte que escape al alcance de este trabajo general.

El gráfico 24 refleja la evolución general que han tenido en el mundo las velocidades de subida y bajada de datos en función del modo de acceso (fijo o móvil) durante el pasado decenio. Demuestra que la mayor parte del aumento en el ancho de banda promedio se debe al incremento en la velocidad de las descargas móviles. En el gráfico 25, donde se analizan más de cerca las velocidades de descarga fija y móvil por región, se observa que la región de América Latina y el Caribe, cuyo ancho de banda promedio supera en velocidad únicamente a Asia Meridional y a África Subsahariana, está entre las regiones más lentas. El Medio Oriente y África Septentrional se destacan a nivel mundial debido a los recientes aumentos en la velocidad de descarga en sus redes móviles de banda ancha. El gráfico 25, que examina la situación más de cerca, permite observar que en la región de América Latina y el Caribe las velocidades de las redes fijas y las móviles son comparativamente similares, en particular frente a las regiones más avanzadas, donde las velocidades de descarga en las redes fijas son claramente superiores. Si bien en este momento solo se puede especular sobre los motivos, dicha circunstancia podría responder a las distintas etapas de desarrollo de las redes de fibra óptica.

Gráfico 24

Velocidades mundiales de banda ancha por tipo, 2008-2019

(En kbps)



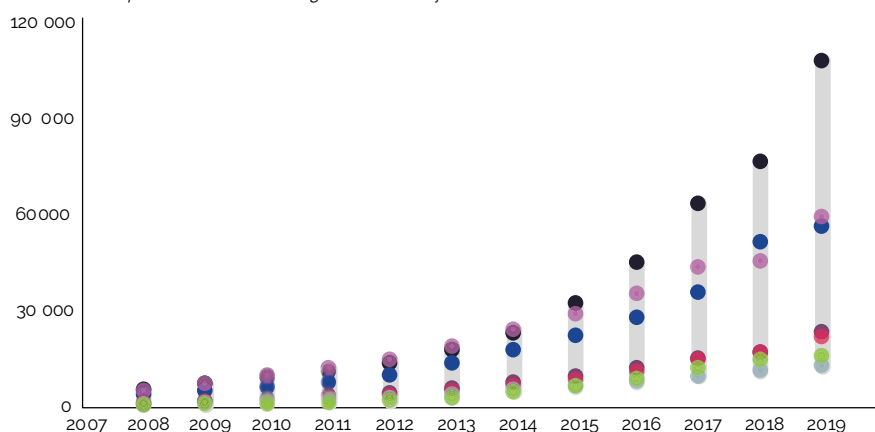
Fuente: Elaboración propia, sobre la base de datos de Ookla.

Si se analizan los países desde una perspectiva individual (no incluida aquí), Chile y Trinidad y Tabago están entre las naciones del mundo con las mejores capacidades de banda ancha. Dentro de la región, hay un grupo de países líderes con velocidades de descarga por telefonía fija relativamente bien desarrolladas, a saber: Barbados, Chile, Panamá, Trinidad y Tabago y el Uruguay. Luego viene un segundo grupo de países cuyas velocidades de descarga por telefonía móvil están más desarrolladas que las de telefonía fija, a saber: Belice,

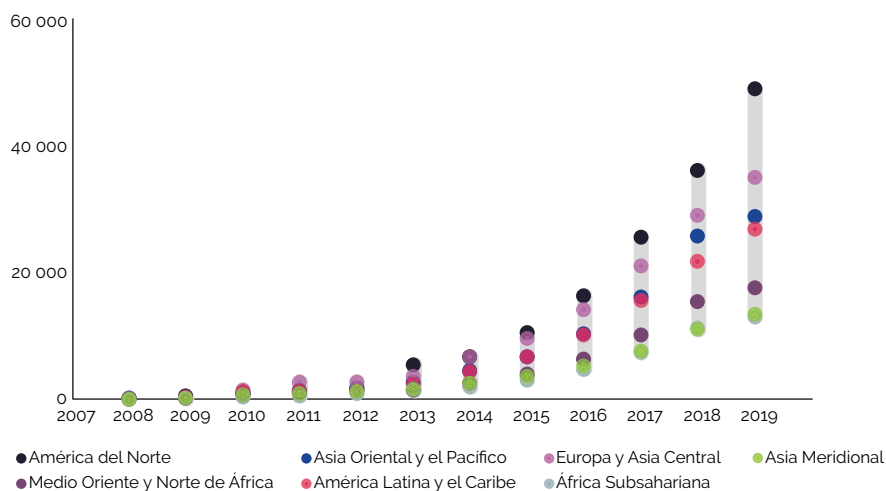
Bolivia (Estado Plurinacional de), Colombia, el Ecuador, Guatemala, Honduras, México, Nicaragua, el Perú y la República Dominicana. Este nutrido grupo de países se caracteriza por el hecho de que las velocidades de descarga en las redes móviles son mayores que en las redes fijas. Por último, hay un tercer grupo de países con un ancho de banda comparativamente inferior, compuesto por Costa Rica, el Paraguay, El Salvador y Venezuela (República Bolivariana de).

Figure 25
Velocidades de descarga en redes móviles y fijas por región mundial, 2007-2019
(En kbps)

A. Velocidad promedio de descarga en redes fijas



B. Velocidad promedio de descarga en redes móviles



Fuente: Elaboración propia, sobre la base de datos de Ookla.

A nivel mundial, existen importantes avances en el desarrollo de las redes de banda ancha fijas y móviles, no solo a raíz del aumento en el número de usuarios, sino también debido a la rápida evolución tecnológica. Prueba innegable de ello es el aumento en las velocidades de descarga, que entre 2010 y 2019 se multiplicaron por 6,5 y por 25 para los servicios de banda ancha fija y banda ancha móvil, respectivamente.

Sin embargo, el ritmo de esta evolución tecnológica no ha sido el mismo en todo el mundo, lo que ha generado importantes disparidades entre las regiones. En febrero de 2019, las velocidades promedio de descarga por banda ancha fija en América Latina eran 2,5 veces inferiores a las de Europa y Asia Central y casi 5 veces más lentas que en América del Norte. En el caso de los servicios de banda ancha móvil, las velocidades promedio de descarga en América Latina estaban un 50% por debajo de las de Europa y Asia Central y eran 3 veces más lentas que en América del Norte.

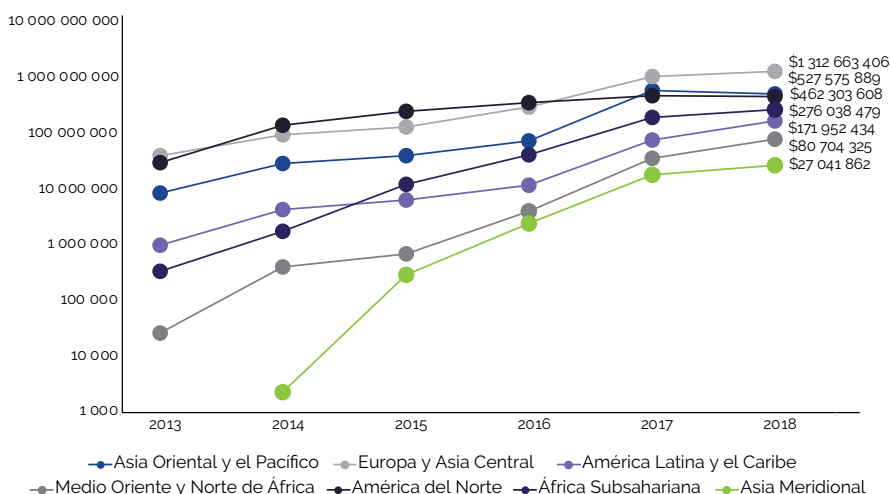
Las criptomonedas son activos digitales diseñados para usarse como medio de cambio que incluyen sofisticadas características de cifrado y permiten asegurar las transacciones financieras, controlar la creación de unidades adicionales y verificar la transferencia de activos a través de un registro distribuido en una cadena de bloques. En esencia, una cadena de bloques es una lista de registros, llamados “bloques”, cada uno de los cuales contiene una clave criptográfica del bloque anterior, una marca de tiempo y datos de la transacción. La existencia de copias redundantes y la naturaleza descentralizada del registro de las cadenas de bloques hace que sea difícil modificar los datos. En ese sentido, las criptomonedas utilizan un sistema descentralizado de control, a diferencia de las monedas digitales centralizadas y de los sistemas de bancos centrales. El bitcoin, que representaba un 53,2% de la capitalización del mercado al momento de recopilar los datos (marzo de 2019), es la criptomoneda más popular (Coin.Dance, 2019).

La página <https://coin.dance/> es un servicio de estadísticas del bitcoin mantenido por la comunidad. Facilita numerosas estadísticas sobre las criptomonedas, por ejemplo, el volumen nacional de *bitcoins* por país en moneda local. Los datos recopilados se informan como el volumen semanal de transacciones. En el sitio se hace un seguimiento de 47 países, entre ellos siete de América Latina, a saber: la Argentina, el Brasil, Chile, Colombia, la República Dominicana, México y el Perú.

El gráfico 26 muestra el aumento en los volúmenes mundiales de *bitcoins* en una escala logarítmica. Permite observar que antes de 2017 América del Norte iba a la cabeza, pero que desde entonces la región de Asia Oriental y el Pacífico la ha superado, junto con Europa, que ocupa el primer lugar mundial en volumen de bitcoins. La región de América Latina y el Caribe poseía en 2014 más del 10% de los *bitcoins* del mundo, ocupando el cuarto lugar entre las regiones; desde entonces ha sido superada por África Subsahariana.

Gráfico 26

Volumen anual de compras de bitcoins por región mundial. 2013-2018
(En dolares por año)



Fuente: Elaboración propia, sobre la base de datos de coindance.org.

La situación socioeconómica de los países de América Latina y el Caribe en gran medida explica el interés en las criptomonedas como el bitcoin y el crecimiento en el volumen de transacciones en años recientes. Estos activos se ven como una solución para la inestabilidad económica, las crisis políticas, la informalidad de la economía y la emigración, factores que hacen de esta región una de las más desiguales del mundo. Estos obstáculos al progreso debilitan la confianza, lo que aunado a una población joven con acceso a la tecnología, transforma a las criptomonedas en instrumentos de inversión e intermediarios para el intercambio de divisas.

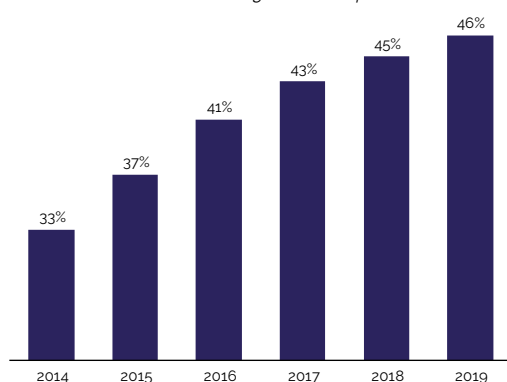
Consultamos las API de los anuncios de Facebook para examinar las principales características sociodemográficas y tecnológicas de la región, y las de Twitter para establecer cuáles son los temas destacados relacionados con los Objetivos de Desarrollo Sostenible (ODS).

A. Aspectos sociodemográficos

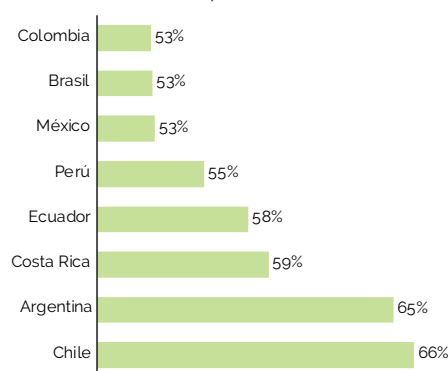
Al momento de recabar los datos, Facebook, la quinta empresa más valiosa del planeta, tenía 2.400 millones de usuarios mensuales, es decir, 1 de cada 3,5 personas en el mundo. La mitad de la población de América Latina utiliza Facebook (datos de 2019, véase el gráfico 27.A), en tanto que en algunos países la penetración aumenta a dos tercios (véase el gráfico 27.B). El modelo de negocios de Facebook consiste en brindar un servicio de anuncios personalizados. La herramienta de gestión de anuncios de Facebook informa a los posibles clientes cuál podría ser el “alcance potencial” de un anuncio comercial en un segmento de usuarios con determinadas características sociodemográficas, económicas y culturales. Utilizamos estas estimaciones del “alcance potencial” para aplicar un método de ingeniería inversa a las aproximaciones de Facebook y derivar un conjunto de estratificaciones sociales.

Gráfico 27
Penetración de Facebook
(En porcentajes)

A. En América Latina lo largo del tiempo



B. Países seleccionados, 2018



Fuente: Elaboración propia, sobre la base de datos de Facebook.

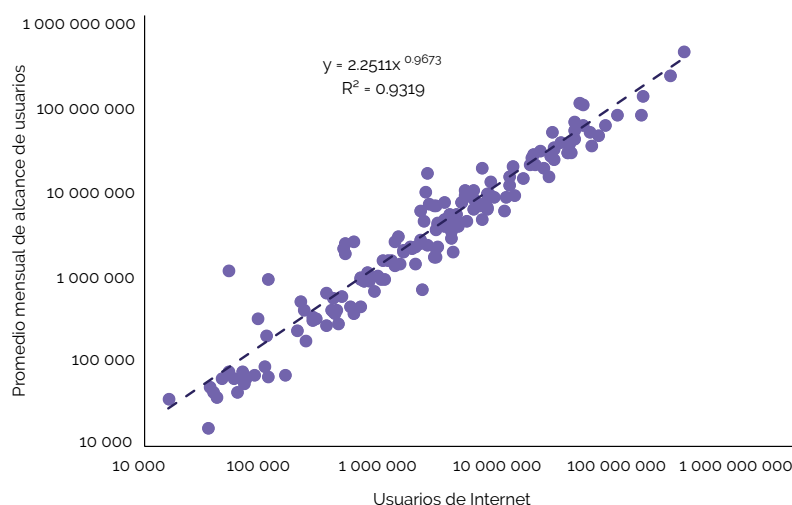
Accedimos sistemáticamente a la API de gestión de anuncios de Facebook para obtener estimaciones de los intereses de las personas, las tecnologías que utilizan y otras características socioeconómicas. Exploramos la relación entre el “alcance potencial” de Facebook y la población real. Según Facebook, “el alcance potencial es una estimación del número de personas que hay en el público objetivo de un conjunto de anuncios. Esta estimación es un cálculo único que realiza Facebook”⁶.

⁶ <https://www.facebook.com/business/help/1665333080167380>.

Buscamos establecer la relación entre el alcance potencial de Facebook, la población real y el número de usuarios de Internet (véase el gráfico 28). El gráfico muestra que existe una correlación cercana entre el alcance potencial y el número de usuarios de Internet ($R^2 = 0,93$). También existe una estrecha correlación entre el alcance potencial y el número de habitantes, aunque naturalmente es menor ($R^2 = 0,78$). La correlación entre el número de usuarios de Facebook y el alcance potencial es aún más estrecha ($R^2 = 0,9975$). Todo esto sugiere que las estimaciones sobre el alcance potencial pueden considerarse bastante confiables, al menos en términos relativos. En términos absolutos, representan fielmente la penetración de Facebook en el país de que se trate.

Gráfico 28

Estimaciones de alcance potencial de Facebook para 164 países frente al número de usuarios de Internet (ITU, 2018)



Fuente: Elaboración propia, sobre la base de datos de Facebook.

En total, obtuvimos 70 variables para los 33 países de América Latina y el Caribe, y siempre segmentamos el alcance potencial en función del género y de seis grupos etarios (13 a 20, 21 a 30, 31 a 40, 41 a 50, 51 a 60 y más de 61)⁷. Tras recopilar datos sobre los promedios de alcance diario y de alcance mensual, constatamos un mayor número de variaciones y resultados menos fiables en las cifras diarias, por lo que

⁷ Facebook_access_(network_type):_2G (behaviors), Facebook_access_(network_type):_3G (behaviors), Facebook_access_(network_type):_4G (behaviors), Facebook_access_(network_type):_WiFi (behaviors), Facebook_access:_older_devices_and_OS (behaviors), business owners (interests), Administrative_Services (industries), Architecture_and_Engineering (industries), Arts_Entertainment_Sports_and_Media (industries), Business_and_Finance (industries), Community_and_Social_Services (industries), Computation_and_Mathematics (industries), Construction_and_Extraction (industries), Education_and_Libraries (industries), Farming_Fishing_and_Forestry (industries), Food_and_Restaurants (industries), Government_Employees_(Global) (industries), Healthcare_and_Medical_Services (industries), IT_and_Technical_Services (industries), Installation_and_Repair_Services (industries), Legal_Services (industries), Life_Physical_and_Social_Sciences (industries), Management (industries), Military_(Global) (industries), Production (industries), Protective_Services (industries), Sales (industries), Transportation_and_Moving (industries), Civil_Union (relationship_statuses), Complicated (relationship_statuses), Divorced (relationship_statuses), Domestic_Partnership (relationship_statuses), Engaged (relationship_statuses), In_a_relationship (relationship_statuses), Married (relationship_statuses), Open_Relationship (relationship_statuses), Separated (relationship_statuses), Single (relationship_statuses), Unspecified (relationship_statuses), Widowed (relationship_statuses), Technology_early_adapters (behaviors), Food_& Restaurant_page_admins (behaviors), Bodybuilding (interests), Meditation (interests), Physical_exercise (interests), Physical_fitness (interests), Running (interests), Weight_training (interests), Yoga (interests), Computer_memory (interests), Computer_monitors (interests), Computer_processors (interests), Computer_servers (interests), Desktop_computers (interests), Free_software (interests), Hard_drives (interests), Network_storage (interests), Software (interests), Tablet_computers (interests), Audio_equipment (interests), Camcorders (interests), Cameras (interests), E-book_readers (interests), GPS_devices (interests), Game_consoles (interests), Mobile_phones (interests), Portable_media_players (interests), Projectors (interests), Smartphones (interests), Televisions (interests).

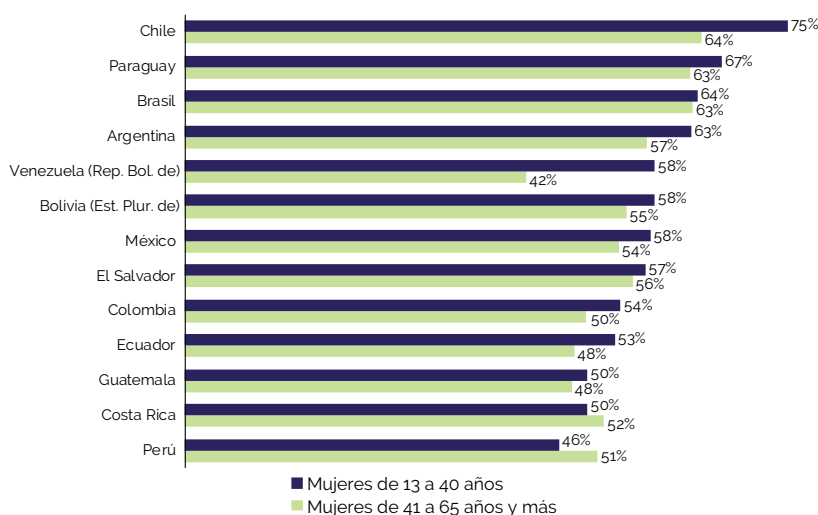
trabajamos con las cifras de alcance potencial correspondientes al promedio mensual de usuarios. Si el promedio mensual de alcance de usuarios no supera los 1.000, el valor por defecto que informa Facebook para esa variable es de 1.000, aun si el promedio diario de usuarios es de 0, 1 o 2 personas por grupo etario y género, por país. Por ende, tomamos datos dos veces en una misma semana para obtener resultados más confiables e informar, o bien el promedio de ambos ejercicios de recopilación como el promedio mensual de usuarios, o bien el número que fuese distinto del valor por defecto de 1.000. Excluimos del análisis a los países en los que, debido al escaso tamaño de la muestra (como a menudo ocurre en el caso de algunos países del Caribe), el valor por defecto poco fiable de 1.000 aparece en más del 50% de los campos informados.

1. Propiedad de las empresas

Facebook podría ser una fuente de estadísticas sobre quiénes son los propietarios reales de las empresas. En el gráfico 29 se han desglosado los datos por sexo y por edad. En general, podemos concluir que las mujeres salen bien paradas en materia de equilibrio de género en cuanto propietarias de empresas (véase el gráfico 30), al menos en lo relacionado con la economía digital. En todos los países de la región, al menos uno de los dos grupos etarios obtiene más del 50%. Las empresarias mayores de 41 años superan a sus homólogas más jóvenes en Costa Rica y el Perú, y aunque al parecer lo mismo ocurre en varios países del Caribe, estimamos que el tamaño de la muestra de dichos países no era lo suficientemente confiable. Es importante señalar que este equilibrio favorable en materia de género podría responder al hecho de que estas empresarias están activas en Facebook. En el mundo tradicional fuera de Internet, es probable que haya más empresarios hombres. Si estos datos son un reflejo de la realidad, es claro que las herramientas digitales tienen un gran potencial de transformarse en importantes ecualizadores en lo referido a las cuestiones de género (Hilbert, 2011).

Gráfico 29

Porcentaje de mujeres propietarias de empresas por grupo etario, en marzo de 2019



Fuente: Elaboración propia, sobre la base de datos de Facebook.

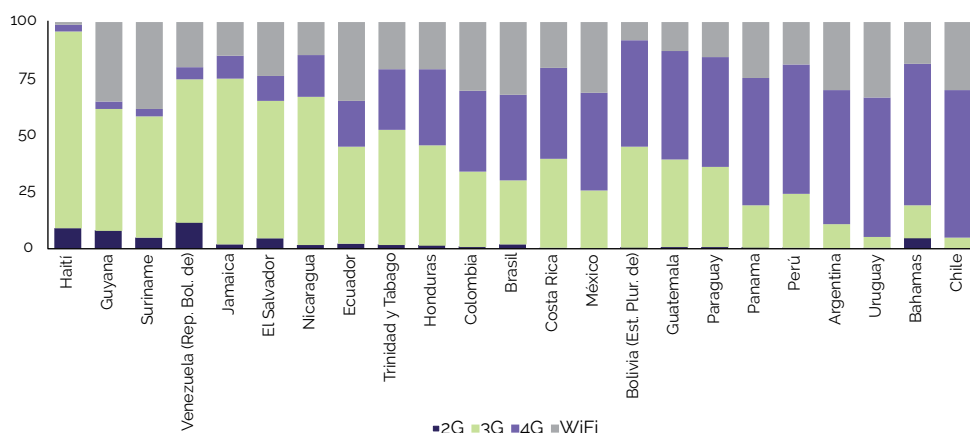
2. Acceso a las redes

Uno de los datos confiables sobre la huella digital proveniente del gestor de anuncios de Facebook se vincula directamente con el uso de esta red social, y se refiere al tipo de red que los usuarios utilizan para acceder al sitio (véase el gráfico 30). Dado que alrededor de la mitad de la población de América Latina y el Caribe utiliza Facebook, esta huella digital brinda una imagen sin precedentes de cómo se utilizan realmente las tecnologías de acceso. Estos datos mejoran las posibilidades de profundizar el análisis sobre cómo se utiliza realmente la infraestructura regional de telecomunicaciones.

Gráfico 30

Acceso a Facebook según el tipo de red de acceso por país, marzo de 2019

(En porcentajes)



Fuente: Elaboración propia, sobre la base de datos de Facebook.

Como se ve en el gráfico 30, las tecnologías móviles 3G y 4G son las vías de acceso más utilizadas. Cabe destacar que en varios países se siguen utilizando las redes 2G, si bien en la mayoría de los casos en una proporción inferior al 5%. Los países pueden dividirse en dos grupos. El primero, integrado por la mayoría de los países de la región, corresponde a aquellos en los que el acceso se realiza principalmente a través de las redes 3G. En el segundo grupo, conformado por la Argentina, las Bahamas, Chile, Panamá, el Perú y el Uruguay, existe un marcado predominio de las redes 4G. En la mayoría de los países, alrededor de un tercio del total de accesos se realiza mediante redes WiFi. Si se toman estos datos como una variable indirecta de los indicadores de conectividad, desarrollar puntos públicos de acceso a las redes Wi-Fi podría ser una solución para aumentar el acceso de la población a Internet.

B. Tendencias

Utilizamos la API de Twitter para determinar qué visibilidad tienen en las redes sociales de la región los 17 Objetivos de Desarrollo Sostenible (ODS).

1. Objetivos de Desarrollo Sostenible (ODS)

Utilizamos el filtro en tiempo real a través de la opción de flujo de tuits de la API de Twitter⁸. Recopilamos la máxima cantidad posible de tuits con relación a un conjunto

⁸ Véase [en línea] <https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html>.

de palabras clave. Dichas palabras incluyeron la totalidad de los 17 ODS y palabras relacionadas en inglés, español y portugués (principalmente las palabras del título principal de cada ODS)⁹. Al reunir los tuits, también recabamos información sobre la ubicación. Con ese fin, especificamos los nombres de los países en la API de Twitter, así como los nombres de las ciudades más importantes¹⁰. De los más de 35 millones de tuits que recabamos, pudimos asignar 4.612.966 a alguno de los 29 países de América Latina y el Caribe.

Imagen 1

Objetivos de Desarrollo Sostenible (ODS)



Fuente: Naciones Unidas (2015), Objetivos de desarrollo sostenible (<https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>).

⁹ Palabras y frases clave recopiladas: 'sdg', 'ods', 'sdg 1', 'ods 1', 'poverty', 'pobreza', 'sdg 2', 'ods 2', 'hunger', 'hambre', 'fome', 'sdg 3', 'ods 3', 'health', 'salud', 'well-being', 'saúde', 'saude', 'bienestar', 'sdg 4', 'ods 4', 'education', 'educación', 'educação', 'educacao', 'sdg 5', 'ods 5', 'derechos de la mujer', 'abuso sexual', 'feminicidio', 'femicidio', 'sdg 6', 'ods 6', 'clean water', 'agua limpia', 'sanitation', 'saneamiento', 'água potável', 'agua potable', 'saneamento', 'sdg 7', 'ods 7', 'clean energy', 'energía asequible', 'affordable energy', 'energía asequible', 'energía no contaminante', 'energía no contaminante', 'energías renováveis', 'energías renováveis', 'energías acessíveis', 'energías acessíveis', 'sdg 8', 'ods 8', 'decent work', 'trabajo decente', 'economic growth', 'crecimiento económico', 'crecimiento economico', 'trabalho digno', 'crescimento econômico', 'crescimento economico', 'sdg 9', 'ods 9', 'industry', 'industria', 'innovation', 'innovación', 'infrastructure', 'innovacion', 'infraestructura', 'indústria', 'inovação', 'inovacao', 'infraestruturas', 'sdg 10', 'ods 10', 'inequality', 'desigualdad', 'inequalities', 'desigualdades', 'sdg 11', 'ods 11', 'sustainable', 'sostenible', 'sustentáveis', 'sustentaveis', 'sdg 12', 'ods 12', 'responsible consumption', 'produccion responsable', 'responsible production', 'producción responsable', 'consumo responsable', 'produção sustentável', 'consumo sustentável', 'producao sustentaveis', 'consumo sustentaveis', 'sdg 13', 'ods 13', 'climate', 'clima', 'climática', 'sdg 14', 'ods 14', 'ocean life', 'submarin', 'marine life', 'aquatic', 'aquático', 'marinha', 'sdg 15', 'ods 15', 'ecosystem', 'ecosistema', 'habitat', 'terrestre', 'forest', 'biodiversidad', 'biodiversidade', 'bosque', 'environment', 'medio ambiente', 'ecossistema', 'biodiversidade', 'floresta', 'meio ambiente', 'sdg 16', 'ods 16', 'peace', 'paz', 'justice', 'justicia', 'strong institutions', 'instituciones sólidas', 'instituciones solidas', 'justiça', 'justica', 'instituições eficazes', 'instituoicoes eficazes', 'sdg 17', 'ods 17'.

¹⁰ Para los países pequeños (es decir, aquellos en los que hay una única ciudad con más de 300.000 habitantes, o ninguna), trabajamos una lista de ciudades grandes. Para los países de mediano tamaño, solicitamos datos para todas las ciudades con más de 300.000 habitantes. En el caso de los países grandes (más de 30 millones de habitantes), efectuamos la consulta para los nombres de las capitales de todos los estados o provincias y todas las ciudades de más de 1 millón de habitantes. Luego, afinamos el resultado inspeccionando manualmente los datos para cada ciudad/estado a fin de eliminar las confusiones entre lugares similares; para eso, agregamos condiciones tales como «AND» o «AND NOT LIKE» entre los nombres de ciudades y países, etc.

2. Número de tuits

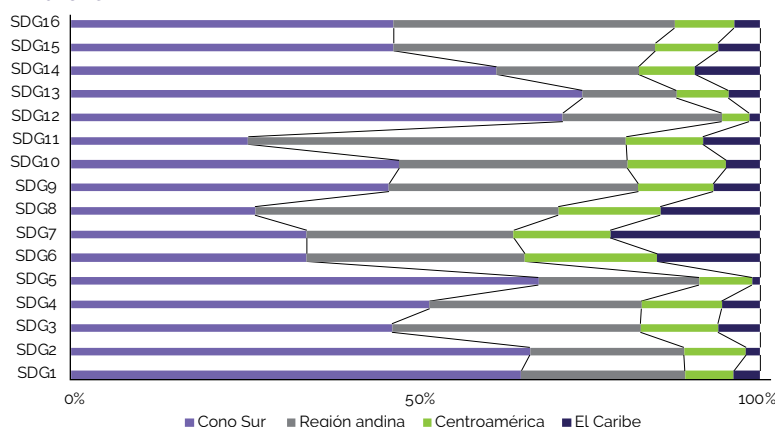
Los temas sobre los que se tuitea en los ámbitos digitales públicos de los diferentes países exhiben marcadas diferencias, pero también semejanzas. Las cuestiones relacionadas con el ODS 16 (paz, justicia e instituciones sólidas) están entre los temas que más se mencionan en la región. Esto guarda relación con lo que sucede en el resto del mundo, si bien en algunos países de la región este tema tiene precedencia sobre otros. El segundo y tercer lugar entre las preocupaciones más importantes que pueden constatar al examinar lo que circula en Twitter lo ocupan los temas de buena salud y bienestar (ODS 3) y acción por el clima (ODS 13). Este último reviste particular importancia en algunos países del Caribe, como Barbados, las Islas Caimán, Dominica, Trinidad y Tabago y las Islas Vírgenes, así como en el Paraguay. A nivel mundial, a estas tres prioridades les siguen los temas relacionados con la educación de calidad (ODS 4) y la industria, innovación e infraestructura (ODS 9). Durante el período observado, y frente a la proporción de tuits procedentes del resto del mundo, la igualdad de género (ODS 5) tuvo mayor prominencia en América Latina y el Caribe, en particular en Chile y en Costa Rica. Las cuestiones de industria, innovación e infraestructura (ODS 9) revisten menos importancia como temas de intercambio en las redes sociales en la región.

En el gráfico 31.A puede observarse que los internautas del Cono Sur participan con mucho interés en los temas relacionados con el ODS 12 (consumo y producción responsables) y el ODS 13 (acción por el clima). Es interesante destacar la profunda preocupación que suscita el ODS 11 (ciudades y comunidades sostenibles) en la región andina, y las inquietudes que en la misma proporción y con el mismo interés expresan los usuarios de las redes sociales en el Caribe con respecto a los ODS 6 (agua limpia y saneamiento), 7 (energía asequible y no contaminante) y 8 (trabajo decente y crecimiento económico).

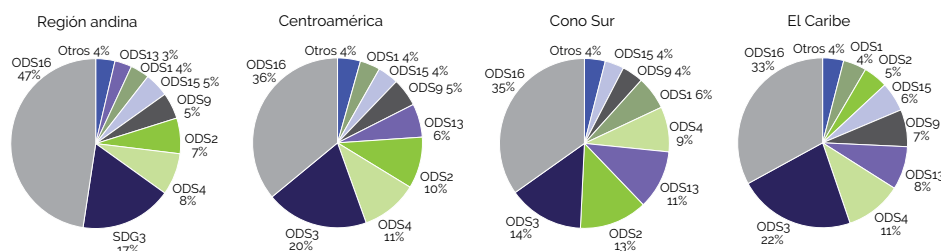
Cuando se analiza el conjunto de las prioridades relativas a los ODS a nivel subregional (véase el gráfico 31.B), queda claro que las inquietudes son en general similares y en gran medida congruentes con las prioridades a nivel mundial. La igualdad de género (ODS 5) suscita más interés en las subregiones de América Latina y el Caribe que a nivel mundial; por otro lado, llama la atención que en todas las subregiones las cuestiones relacionadas con la acción por el clima (ODS 13) generan menos interés que a nivel mundial.

Gráfico 31
Tuits compartidos, febrero y marzo de 2019
(En porcentajes)

A. Por ODS



B. Por región



Región andina (Bolivia (Est. Plur. de), Colombia, Ecuador, Perú); El Caribe (Antigua y Barbuda, Bahamas, Barbados, Islas Caimán, Cuba, Dominica, Haití, Jamaica, República Dominicana, Trinidad y Tabago, Islas Vírgenes); Centroamérica (Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panamá, Puerto Rico); Cono Sur (Argentina, Chile, Paraguay, Uruguay).



Fuente: Elaboración propia, sobre la base de datos de Twitter.

En este informe hemos examinado y mostrado ejemplos de la manera en que los datos en línea disponibles para el público pueden ser útiles a la hora de formular políticas para la región. Hemos descubierto los importantes beneficios que pueden reportar los datos sobre la huella digital, aunque también hemos visto que no es fácil obtenerlos, y que es necesario proseguir con el análisis y no cejar en los esfuerzos para cosechar esos beneficios. Después de todo, en nuestra opinión los beneficios superan con creces los esfuerzos necesarios y las exigencias de recursos, sobre todo cuando se comparan con las fuentes tradicionales de datos.

A. Resumen de los beneficios y los desafíos

1. Nuevas percepciones

Hemos demostrado que los datos sobre la huella digital brindan una visión completamente nueva en materia de:

- **alcance temático:** se incluyeron ámbitos que anteriormente eran difíciles o imposibles de cuantificar, como la prestación de los servicios o los desequilibrios de género en las pequeñas empresas, o la oferta y la demanda del mercado laboral;
- **alcance geográfico:** nuestras fuentes panregionales brindan un sustancial conjunto de datos comparables para entre 14 y 18 países, y los portales nacionales aportaron nuevos datos sobre las subregiones nacionales;
- **exhaustividad:** diferentes matices de la opinión pública, estado civil por edad y desequilibrios de género por categoría laboral;
- **fuentes alternativas:** tradicionalmente, los textos y las imágenes han sido tratados como información cualitativa sin valor estadístico, pero las técnicas modernas de aprendizaje automático nos permiten convertir dichas fuentes en estadísticas mensurables; y
- **carácter oportuno:** generamos los gráficos durante la misma semana en que recabamos los datos, y podemos publicarlos rápidamente mediante interfaces interactivas en línea.

Además de la ventaja cualitativa de trabajar con datos sobre la huella digital, comparativamente se necesitan menos recursos, sobre todo cuando se tiene en cuenta la gran diversidad de debates sobre la formulación de políticas que puede suscitar la abundancia de gráficos posibles. En este ejercicio, hicimos especial hincapié en la economía digital y en las cuestiones de género, pero también hemos visto que los datos públicos sobre la huella digital pueden ser muy beneficiosos para otras metas de desarrollo mucho más amplias, por ejemplo, las que se relacionan con los Objetivos de Desarrollo Sostenible.

2. Desafíos que perduran

Al pensar en el paradigma de los macrodatos, muchos tienden ingenuamente a dar por sentado que los desafíos de trabajar con datos sobre la huella digital se reducen a las destrezas informáticas, y quizás a la infraestructura tecnológica necesaria. La noción parece ser que alcanza con resolver esas cuestiones para que la huella digital se transforme en una suerte de bola de cristal en tiempo real. Nuestra experiencia demuestra que los desafíos reales son otros. Para recabar y analizar los datos no utilizamos ninguna infraestructura sofisticada (trabajamos sobre todo con computadoras portátiles de gama media), y las destrezas informáticas se limitaron principalmente a recopilar los datos usando la herramienta Python (mediante la técnica de *raspado de datos* o accediendo a las API) para luego depurarlos con R. Python se ha transformado en el lenguaje de programación por antonomasia para fines generales, e incluye una nutrida gama de cursos gratuitos en línea para quienes desean aprender a usarlo. En realidad, los desafíos más complejos que enfrentamos se relacionaron con los retos que en general plantea la estadística tradicional. Las prioridades se modifican un poco, pero los debates fundamentales siguen siendo los mismos.

- **Representatividad.** Mientras que en los ejercicios estadísticos tradicionales sobre el desarrollo la prioridad fundamental es que las muestras aleatorias del estudio sean representativas, los datos sobre la huella digital nunca conforman una muestra aleatoria. Se trata de datos empíricos, y a menudo el objetivo es abarcar a la totalidad de la población en el estudio. Hemos recopilado datos sobre unas 80.000 ofertas de trabajo en Bumeran y sobre la totalidad de los 200.000 artículos de tecnología que se ofrecen en MercadoLibre. En el paradigma de los macrodatos, la representatividad de la muestra y las pruebas de significancia relacionadas pierden su preponderancia. En lo referido a las destrezas técnicas indispensables, la pericia en la toma de muestras representativas y en la realización de pruebas de significancia deja paso a la habilidad para recopilar, depurar y almacenar los datos sobre la huella digital con eficacia.
- **Generalización.** Si bien los datos empíricos que incluyen la totalidad de una fuente siempre la representan fielmente, solo pueden representar lo que representan, y nada más. En la ciencia estadística tradicional, las muestras se utilizan para derivar patrones generales aplicables a una población más grande. En este caso, recabamos datos de la totalidad de la población y, si bien es tentador hacer generalizaciones que se apliquen a contextos más amplios, es a menudo muy engañoso. Hemos constatado que el mercado de trabajo mundial de Freelancer.com es diferente al mercado regional de Workana.com, que a su vez difiere de los mercados de trabajo nacionales de Bumeran.
- **Armonización.** Las organizaciones internacionales están habituadas a los desafíos que entraña la armonización internacional de los indicadores, y los datos sobre la huella digital no escapan a esos retos. Si bien nos apoyamos en diversas plataformas panregionales e internacionales, aun en estos casos se constatan diferencias a nivel nacional. La incompatibilidad entre los indicadores a menudo impidió que la cobertura fuese más amplia y detallada.
- **Definición de las variables.** La definición del indicador elegido y la ponderación de cualquier índice agregado pueden predeterminar gran parte del resultado. El trabajo con datos sobre la huella digital plantea desafíos similares a los tradicionales (¿cómo ponderar un índice?) y añade dimensiones nuevas (¿qué deben buscar los sistemas de aprendizaje automático cuando convierten los datos cualitativos en datos cuantitativos?).

B. Aspectos que merecen especial atención

A continuación, presentamos varias reflexiones más detalladas que se relacionan principalmente con los desafíos estadísticos tradicionales mencionados, sin dejar de destacar algunas de las nuevas oportunidades en materia de medición. Esperamos que este análisis más exhaustivo sirva de orientación a los profesionales que desean valerse de los datos sobre la huella digital para avanzar hacia metas de desarrollo internacional.

En este primer ejercicio, arrojamus luz sobre una amplia gama de cuestiones que nunca se habían analizado antes. Desde el punto de vista sustantivo, cada uno de los gráficos merece un análisis mucho más profundo. Muchos de ellos plantean bastantes más preguntas que respuestas, y exigirán un análisis más exhaustivo de los expertos que correspondan en cada ámbito, sea que se trate de cuestiones de género, mercados laborales, precios de la tecnología, pequeñas y medianas empresas o un país específico, entre otras cosas. Se podría escribir una publicación por separado para cada uno de estos temas. Asimismo, muchos de los aspectos que abordamos justifican un análisis de seguimiento para aclarar ciertos puntos. Eso estaba previsto, ya que cuanto mayor es la cantidad de datos, más nos damos cuenta de que hay muchas cosas que aún no sabemos y otras que tenemos que seguir investigando. En este sentido, si bien este informe no tenía por objeto analizar de forma exhaustiva cada uno de tales ámbitos, sí logramos demostrar que los datos sobre la huella digital pueden promover debates cruciales sobre los nuevos temas que afectan a la región, lo que a su vez puede ser útil a la hora de formular políticas.

De hecho, los gráficos incluidos en este documento demuestran que el enfoque del análisis sustantivo ha cambiado en su naturaleza. Tradicionalmente, el principal desafío para los expertos en materia de desarrollo en la región era la escasez de datos. Además de ser difíciles y costosos de producir, era raro encontrarse con un gráfico nuevo, y eso los transformaba en herramientas valiosas. Aquí, hemos incluido más gráficos de los que hemos sido capaces de analizar con el cuidado que merecen. Por ejemplo, tuvimos a nuestra disposición diversas fuentes de datos para el mercado laboral de distintos países, y con cada país podríamos haber llenado varias páginas analizando en profundidad cada ámbito específico, contrastando las diferentes cifras con los datos nacionales. En la era de los macrodatos, hay una sobreabundancia de datos. Al mismo tiempo, es necesario ejercer cautela para definir con exactitud qué representa cada fuente de datos.

1. La importancia de las fuentes

Los datos sobre la huella digital son empíricos. Son totalmente distintos a los datos que surgen de las muestras aleatorias tomadas en estudios, especialmente a la hora de establecer patrones generales para aplicarlos a poblaciones distintas o más grandes. Los datos sobre la huella digital siempre están sesgados hacia la fuente de la cual proceden. Por ejemplo, las tarifas por hora en Freelancer.com y en Workana.com son distintas, lo que indica que miden cosas distintas (demanda mundial, demanda de tipo más regional). También encontramos diferencias claras y elocuentes entre los mercados de trabajo representados por la economía de las ocupaciones transitorias y por el sector del empleo de tiempo completo. ¿Cuál de las dos es la representación “correcta” del “mercado de trabajo”? Ambas lo son, por cuanto representan aspectos distintos de la realidad. Dichas diferencias no son interferencias molestas. Al trabajar con datos empíricos, solo podemos observar lo que sucede en la fuente de la que provienen, y nada más.

Además, la fuente solo puede brindar lo que brinda. Por ejemplo, para la economía de las ocupaciones transitorias, es fácil comparar la oferta y la demanda, ya que ambos aspectos aparecen reflejados en la misma plataforma. Sin embargo, en el caso de los empleos a tiempo completo que se ofrecen en Bumeran, solamente podemos recabar datos sobre la demanda, pero no tenemos acceso a la oferta a nivel nacional.

También es muy importante señalar que en lo relacionado con la formulación de políticas, la dependencia de la fuente influye en gran medida en la interpretación y las conclusiones. Por ejemplo, el hecho de que en países como Bolivia (Estado Plurinacional de), El Salvador y Nicaragua haya un porcentaje más alto que en el Brasil de las especializaciones “Informática y programación” e “Ingreso y administración de datos” no necesariamente es un reflejo de la realidad de estos profesionales. Para los países con el porcentaje más alto, una explicación podría ser que la proporción de estos profesionales que ofrece sus servicios en la plataforma en línea en cuestión simplemente es más alta. En el Brasil, es posible que los profesionales de esas categorías tengan empleos de tiempo completo o que utilicen un portal nacional distinto para ofrecer sus servicios. En este sentido, es revelador que las plataformas que se utilizan en la economía de las ocupaciones transitorias brindan a los trabajadores de países con condiciones nacionales específicas (por ejemplo, en materia de situación económica o de estructura productiva) opciones que trascienden los límites que impone el contexto nacional. Por supuesto, esta es una conclusión interesante de por sí. El aspecto más importante que debe tenerse en cuenta es que, en definitiva, los datos obtenidos solo representan lo que representan, y hay que ejercer cautela con las generalizaciones.

2. La importancia de los indicadores y los índices

La definición de los indicadores importa, y en función de la ponderación en categorías e índices, puede ser más alta o más baja según el país. Por supuesto, la definición de cuál es el indicador más útil depende de la persona encargada de la elección. Lo mejor que podemos hacer es eliminar los menos razonables (tratando de evitar el riesgo de pasar por alto alguna necesidad especial) y publicar todos los datos que pudieran ser útiles en paneles dinámicos que permitan a los usuarios buscar respuestas a preguntas aún no contestadas. Al hacerlo, es importante ser sumamente transparente al explicar la procedencia y la metodología que sustentan las cifras presentadas, si bien la interpretación corresponde únicamente al investigador que tenga conocimientos específicos sobre el ámbito en cuestión.

3. La importancia de la armonización

En teoría, uno de los beneficios de los datos sobre la huella digital es el nivel de detalle sin precedentes que puede alcanzarse, lo que permite realizar análisis sumamente precisos. Sin embargo, en la práctica pudimos constatar que a menudo es necesario sacrificar los detalles en aras de la armonización internacional. Aun cuando los países pequeños ofrezcan una gama amplia de categorías tecnológicas, en última instancia solo en los países más grandes encontramos un nivel suficiente de superposición en las definiciones como para poder compararlos entre ellos. En lo referido a las categorías de empleo, tuvimos que trabajar con un nivel de categorización fuertemente agregado a fin de poder armonizar los datos de las diferentes fuentes.

Si bien la solución para armonizar las variables a menudo consiste en agregarlas, en otros casos la normalización puede ser útil. De hecho, puede llegar a facilitar la armonización entre diferentes fuentes. En lo concerniente a las cuestiones de género en la economía de las ocupaciones transitorias, observamos diferencias en las cifras

absolutas entre las dos fuentes de datos. Sin embargo, en cifras relativas, la escala ordinal fue en general congruente entre ambas fuentes. No es eso lo que sucede en otros casos, por ejemplo, con la tarifa por hora de los trabajadores autónomos. Aquí pudimos observar que aparentemente existen dos mercados distintos, algo que también es un hallazgo interesante.

4. El conocimiento sobre los ámbitos es importante

Obtener los datos exige un determinado nivel de destreza técnica, pero no es imposible. Visualizarlos de una forma que les aporte significado exige algunos conocimientos básicos sobre los ámbitos en cuestión, pero es conocimiento que está disponible. En realidad, la mayor parte de las tareas complejas ocurren en el medio de estos dos extremos. Algunas implican detectar los errores e incongruencias ocurridos durante la etapa de recopilación de los datos, e invertir tiempo a fin de entenderlos correctamente para decidir cómo depurarlos y sobre qué aspectos debe hacerse hincapié en vista de su naturaleza. Hubo muchas ocasiones en que nos habría gustado plantear preguntas que los datos sobre la huella digital no podían contestar por sí solos. Esto parece indicar que ese tipo de datos no reemplazarán por completo las estadísticas existentes y las tareas de investigación sobre el terreno. Más bien, las complementarán. Tampoco eliminarán la necesidad de contar con conocimientos sólidos, sino que exigirán mayores conocimientos de ese tipo. Es necesario que haya expertos sobre los distintos ámbitos para comprender la utilidad de todos los datos reunidos. La ciencia de datos se transformará en un aspecto natural de esta importante tarea, y para la mayoría de las empresas que se dedican a trabajar con datos el problema no será la escasez, sino la abundancia. Por ende, estimamos que en el futuro el principal obstáculo para las tareas analíticas de este tipo serán las deficiencias de conocimiento especializado sobre los distintos ámbitos y la falta de experiencia sólida sobre las condiciones locales y regionales.

- Amato, F. y otros (2015), "Challenge: processing web texts for classifying job offers"; *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, Instituto de Ingeniería Eléctrica y Electrónica (IEEE).
- CEPAL (Comisión Económica para América Latina y el Caribe) (2013), *Economía digital para el cambio estructural y la igualdad (LC/L.3602)*, Santiago.
- _____(2018a), *La ineficiencia de la desigualdad (LC/SES.37/3-P)*, Santiago.
- _____(2018b), *Datos, algoritmos y políticas: la redefinición del mundo digital (LC/CMSI.6/4)*, Santiago.
- Coin Dance (2019), "Bitcoin Statistics" [en línea] <https://coin.dance/stats>.
- Dini, M. y G. Stumpo (comps.) (2011), "Políticas para la innovación en las pequeñas y medianas empresas en América Latina", *Documentos de Proyectos (LC/W.403)*, Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).
- Ferraro, C. y G. Stumpo (comps.) (2010), *Políticas de apoyo a las pymes en América Latina: entre avances innovadores y desafíos institucionales*, Libros de la CEPAL, N° 107 (LC/G.2421-P), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).
- Heeks, R. (2017), "Decent work and the digital gig economy: a developing country perspective on employment impacts and standards in online outsourcing, crowdwork, etc"; *Development Informatics Working Papers*, N° 71, Global Development Institute, Manchester, Universidad de Manchester.
- Hilbert, M. (2019), "Digital Data Divide Database", *SSRN Scholarly Paper*, N° 3345756.
- _____(2016), "The bad news is that the digital access divide is here to stay: domestically installed bandwidths among 172 countries for 1986–2014", *Telecommunications Policy*, vol. 40, N° 6.
- _____(2014), "Technological information inequality as an incessantly moving target: the redistribution of information and communication capacities between 1986 and 2010", *Journal of the Association for Information Science and Technology*, vol. 65, N° 4.
- _____(2011), "Digital gender divide or technologically empowered women in developing countries?: a typical case of lies, damned lies, and statistics", *Women's Studies International Forum*, vol. 34, N° 6.
- Hilbert, M., P. López y C. Vásquez (2010), "Information societies or "ICT equipment societies?": measuring the digital information processing capacity of a society in bits and bytes", *The Information Society*, vol. 26, N° 3.
- Kässi, O. y V. Lehdonvirta (2018), "Online labour index: measuring the online gig economy for policy and research", *Technological Forecasting and Social Change*, vol. 137.
- Minges, M. (2005), *Evaluation of e-readiness indices for Latin America and the Caribbean (LC/W.73)*, Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).
- MercadoLibre (2018), "Annual Reports and Proxy Statements" [en línea] <http://investor.mercadolibre.com/financial-information/annual-reports>.
- Naciones Unidas (2015), *Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible (A/RES/70/1)*, Nueva York.
- Naciones Unidas (2015), "Objetivos de Desarrollo Sostenible" [en línea] <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>.
- Ookla (2019), "Speedtest Global Index" [en línea] <https://www.speedtest.net/global-index>.
- Peres, W. y G. Stumpo (coords.) (2002), *Las pequeñas y medianas empresas industriales en América Latina y el Caribe*, Ciudad de México, Siglo XXI/Comisión Económica para América Latina y el Caribe (CEPAL).
- Price, S. (2017), "Lending pioneer Kiva hits the one billion mark and launches a fund for refugees", *Forbes*, 6 de julio [en línea] <https://www.forbes.com/sites/susanprice/2017/07/06/lending-pioneer-kiva-hits-the-one-billion-mark-and-launches-a-fund-for-refugees/>.
- Vallejos, Z. (2003), "Micro, pequeñas y medianas empresas en América Latina", *Revista de la CEPAL*, N° 79 (LC/G.2200-P), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).
- Vespignani, A. (2009), "Predicting the behavior of techno-social systems", *Science*, vol. 325, N° 5939.
- Waghorn, T. (2013), "Premal Shah: loans that change lives", *Forbes*, 4 de noviembre [en línea] <https://www.forbes.com/sites/terrywaghorn/2013/11/04/premal-shah-loans-that-change-lives/>.

La economía mundial está cada vez más dominada por las tecnologías digitales, que actualmente constituyen las principales plataformas de la mayoría de las actividades socioeconómicas, con efectos positivos en el crecimiento económico y en el bienestar. A fin de maximizar esos efectos, es fundamental que las políticas de desarrollo se basen en evidencia empírica que apoye la toma de decisiones en ámbitos que van desde la asignación de recursos hasta la evaluación de impacto.

En el presente informe se examinan las oportunidades y los desafíos que plantea el uso sistemático de datos digitales disponibles públicamente como herramienta para la formulación de políticas públicas para el desarrollo de la economía digital en América Latina y el Caribe. Tiene por objetivo compartir las experiencias adquiridas para avanzar en una agenda de investigación que permita a los países de la región crear herramientas de medición alternativas basadas en la huella digital. Mediante el uso de técnicas de megadatos (big data), la huella digital que dejan los portales de empleo, las plataformas de comercio electrónico y las redes sociales ofrece información sin precedentes, tanto en términos de alcance como de detalle.

El informe cumple dos funciones complementarias. Muestra el estado del arte en temas seleccionados relativos a la economía digital en América Latina y el Caribe, y sirve de base para el debate sobre las oportunidades y desafíos que plantea la utilización de rastros digitales en línea para la formulación de políticas de desarrollo.

Las contribuciones sustanciales del informe están en consonancia con la propuesta de la Comisión Económica para América Latina y el Caribe (CEPAL) de impulsar políticas que promuevan el cambio estructural para el desarrollo equitativo y sostenible aprovechando todo el potencial de la economía digital, a fin de modificar la estructura de producción de la región hacia industrias más intensivas en conocimiento y sectores de mayor productividad.

El aporte metodológico del informe se deriva de los conocimientos adquiridos de la realización de un ejercicio de análisis de macrodatos. Por lo tanto, sirve como guía general para los profesionales interesados en utilizar la ciencia de datos en el ámbito de las políticas de desarrollo. Dar sentido a los megadatos de una manera significativa implica no solo hacer frente a desafíos computacionales, sino también concebir la ciencia de datos como la convergencia entre informática, estadística, y su amplio ámbito de aplicación. En la práctica de la ciencia de los datos, cuestiones de representatividad, generalización, armonización, calidad de los datos, así como la definición de variables e índices, son asuntos de fundamental relevancia.

