

Cuadernos de Desarrollo Económico

64

*Big data e informalidad
en los corazones
productivos de Bogotá*

DICIEMBRE 2023

Tomás Concha Llorente
Juliana Aguilar Restrepo



SECRETARÍA DE
DESARROLLO
ECONÓMICO





Presariasiales
Precios na
Moda & Decoración
Tel: 710 72 33
Cra 52C # 41 - 24 sur / Alquería

Elastinsumos
ELÁSTICOS, TEXTILES E INSUMOS PARA CONFECCIÓN
© 808 9456 0000 1626

EL PUNTO DE LA
SÁBANA Y EL ACOLCHADO
2DO PISO
SIGA

ANA Y EL ACOLCHADO
CARRERA 52 C # 41 - 24 SUR

EL PUNTO DE LA
SÁBANA Y EL ACOLCHADO
2DO PISO
SIGA

Moda & Decoración

Los Adornos PJ
CREMILLERAS - BOTONES
BROCHES - FERRAJES

2DO PISO
DE LA ACOLCHADO

EL PUNTO

EL PUNTO

EL PUNTO

Claudia Nayibe López Hernández
Alcaldesa Mayor de Bogotá

Secretaría Distrital de Desarrollo Económico

Alfredo Bateman Serrano
Secretario de Desarrollo Económico

Sebastián Marulanda Robledo
Subsecretario de Desarrollo Económico

Maria Catalina Bejarano Soto
Directora de Estudios de Desarrollo Económico

Juliana Aguilar Restrepo
Subdirectora de Estudios Estratégicos

Yaneth Lucía Pinilla Beltrán
Subdirectora de Información y Estadísticas

Autores
Tomás Concha Llorente
Juliana Aguilar Restrepo

Revisión de estilo
Juliana Ortega Camelo

Diagramación:
Ana María González Rojas
César Javier Daza Gutiérrez

Material fotográfico:
Archivo SDDE

Cuaderno No. 64
Big data e informalidad en los corazones productivos de Bogotá
Colombia, diciembre de 2023

ISSN

2981-4790 (En línea)

Puede encontrar información editorial de este cuaderno en la página Web:

<http://observatorio.desarrolloeconomico.gov.co>

Contáctenos en:

Secretaría Distrital de Desarrollo Económico

Carrera 10 # 28-49 Torre A, Bogotá D.C.

Teléfono: 601 369 3777 Ext: 204

observatorio@desarrolloeconomico.gov.co



La serie de Cuadernos de Desarrollo Económico es una publicación de la Secretaría de Desarrollo Económico de la Alcaldía Mayor de Bogotá. Los trabajos de la Serie de cuadernos son de carácter provisional; las opiniones y el contenido son responsabilidad exclusiva de los autores y no comprometen a la Secretaría de Desarrollo Económico ni a la Alcaldía Mayor de Bogotá. Todo el material está protegido por los derechos de autor y su uso está protegido libremente en tanto se haga mención y se cite en la bibliografía. Su versión en línea está disponible bajo la licencia **Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International**

***Big data* e informalidad en los corazones productivos de Bogotá**

RESUMEN

Este estudio utiliza una metodología basada en *big data* para analizar de forma detallada la informalidad y digitalización del tejido empresarial a escala territorial, en particular, en 9 corazones productivos de Bogotá identificados y priorizados por la Secretaría Distrital de Desarrollo Económico. Para esto se comparan datos de *Google Maps* con registros administrativos como la matrícula mercantil de la Cámara de Comercio de Bogotá (CCB), utilizando técnicas de coincidencia difusa (*fuzzy matching*). Además, se extrae información sobre la digitalización empresarial. Se estima una tasa de informalidad del 27,3 % en estos espacios, que varía entre 21,2 % (Siete de agosto) y 43,5 % (Ricaurte). La informalidad se concentra en servicios y comercio, con baja participación en la industria. Adicionalmente, existe una relación positiva entre formalidad y digitalización, evidenciada en mayor presencia digital de los negocios formales. Sin embargo, en algunos territorios los negocios informales superan en calificaciones promedio a las formales. Facebook e Instagram son las redes sociales más usadas. Los resultados muestran el potencial del uso de *big data* para acercarse a las dinámicas territoriales y sirven para orientar políticas de formalización en Bogotá.

Palabras clave: dinámica empresarial, informalidad empresarial, *big data*, digitalización, redes sociales.

ABSTRACT

This study employs a big data-based methodology to analyze in detail the informality and digitalization of the business fabric at a territorial scale, particularly in 9 productive hubs of Bogotá identified and prioritized by the District Department of Economic Development. To achieve this, data from *Google Maps* is compared with administrative records, such as the commercial registration of Bogota's Chamber of Commerce (CCB), using fuzzy matching techniques. Additionally, information on business digitalization is extracted. An informality rate of 27,3 % is estimated in these spaces, ranging from 21,2 % (Siete de agosto) to 43,5 % (Ricaurte). Informality is concentrated in services and commerce, with low participation in the industry sector. Furthermore, there is a positive relationship between formality and digitalization, evidenced by a greater digital presence of formal businesses. However, in some territories, informal businesses surpass formal ones in average ratings. Facebook and Instagram are the most used social networks. The results showcase the potential of using *big data* to understand territorial dynamics and serve to guide formalization policies in Bogota.

Keywords: Business dynamics, business informality, big data, digitalization, social networks.

CONTENIDO

1	Introducción	8
----------	---------------------	----------

2	Medición de la informalidad empresarial y la huella digital de las empresas	11
	2.1. Metodología	11
	2.1.1. Obtención de datos de la API de <i>Google Maps</i>	11
	2.1.2. Identificación de empresas formales en el territorio	14
	2.1.3. Estimación de la probabilidad de informalidad empresarial	14
	2.2. Resultados	16
	2.3. Robustez	18
	2.4. Sesgos de la información	19

3	Informalidad y digitalización	22
	3.1. Medición de la digitalización de las empresas	22
	3.2. Resultados	23

CONTENIDO

4	Conclusiones	27
----------	---------------------	-----------

5	Bibliografía	29
----------	---------------------	-----------

Anexos	31
Anexo 1. Revisión bibliográfica	31
Anexo 2. Negocios y distribución sectorial en los corazones productivos	32
Anexo 3. Correlativa entre lugares de <i>Google Maps</i> y sectores económicos	34
Anexo 4. Proporción de adopción de redes de empresas formales e informales	35



1. INTRODUCCIÓN

Las microempresas en Colombia suelen ser fundadas en la informalidad (DNP, 2019), y la apertura de las redes sociales y los motores de búsqueda les permiten validar en los mercados los bienes y servicios ofrecidos. Si tienen buena recepción, pueden llevar a que las microempresas se formalicen para que, de esta manera, puedan acceder a mercados más grandes.

Las redes sociales y los motores de búsqueda le abren la puerta a individuos y pequeñas empresas a oportunidades de mercado. La información necesaria para analizar estos fenómenos rara vez se encuentra en los registros administrativos o en las encuestas. Esta se encuentra en internet, lugar que, gracias a la digitalización, permite intercambiar un pedazo de información como la ubicación (*location*) a cambio de ventas.

Por su parte, existen limitaciones para estimar la informalidad empresarial, especialmente en el territorio. Según el documento CONPES 3956 sobre política de formalización empresarial, 75,0 % de las microempresas en Colombia no se encuentran registradas, y el pago de las prestaciones sociales, la declaración y el pago de impuestos, y los niveles de cumplimiento de contratación formal de los trabajadores son bajos (DNP, 2019).

Para Bogotá, la Encuesta de micronegocios del DANE estimó que en 2022 el 71,9 % de los micronegocios de la ciudad no contaba con un registro mercantil, no obstante, estas fuentes no brindan información sobre su distribución en el territorio. Por su parte, los resultados de la Tropa económica¹ pueden desagregarse para diferentes niveles territoriales, pero por la forma en la que fue concebido el trabajo de campo, los datos no son representativos. En este contexto, fuentes alternativas de datos que permitan un acercamiento a la informalidad en el territorio son un insumo importante para la formulación de políticas públicas para la promoción de la formalidad empresarial.

¹ La iniciativa de la Tropa económica de Bogotá, dirigida por la Secretaría Distrital de Hacienda (SHD) y la Secretaría de Desarrollo Económico (SDDE), se enfoca en identificar y caracterizar las unidades productivas en Bogotá. Su objetivo es crear una extensa base de datos maestra para comprender el estado actual de las empresas en la capital, sus necesidades durante el proceso de reactivación económica y sus requisitos de apoyo en áreas de desarrollo, capacitación y crecimiento empresarial en la ciudad. Además, pretende vincular la oferta institucional con las unidades productivas de acuerdo con sus requerimientos. (<https://observatorio.desarrolloeconomico.gov.co/dinamica-empresarial/nota-editorial-tropa-economica>)

Este documento utilizó técnicas de *big data* para aproximarse a la tasa de informalidad en 9 corazones productivos de Bogotá priorizados para la atención de la Secretaría Distrital de Desarrollo Económico (SDDE) en 2022 y 2023², (Castaño, Pardo, Rodríguez, Rivera, & Aguilar, 2023). El *big data*, entendido como la forma de obtener información de manera no intrusiva, brinda la oportunidad de, no solo estimar la informalidad empresarial en las ciudades³, sino también de entender sus vínculos con la digitalización. Se encuentra que, a mayor nivel de digitalización hay una mayor tasa de formalidad empresarial.

Inicialmente, este estudio llevó a cabo un mapeo digital exhaustivo de las empresas en estos corazones productivos, utilizando herramientas de geolocalización y el API⁴ de *Google Maps*⁵. Tras identificarlas, se aplicó una técnica de coincidencia difusa (*fuzzy matching*), que comparó los nombres de estas con aquellas que se definen como formales; para este estudio fueron registrados todas las empresas y establecimientos con matrícula mercantil activa en la CCB con corte a diciembre de 2022⁶. Esta técnica asignó un grado de similitud que permitió categorizar los negocios⁷ de *Google Maps* desde aquellos con mayor probabilidad de ser formales, es decir, que se encontraban en el registro mercantil con alta probabilidad; hasta aquellos que probablemente operaban en la informalidad. El estudio estimó que 27,3 % de los negocios ubicados en estos núcleos son considerados informales según los criterios establecidos.

Adicionalmente, se caracterizaron los negocios formales e informales de acuerdo con su presencia digital. De los 5.691 negocios identificados a través de Google, solo 1.786 (31,4 %) tenían presencia en línea con un sitio web propio y 918 (16,1 %) ampliaron su presencia digital con perfiles activos en redes sociales. Además, se encontró que los negocios identificados como formales solían tener más evaluaciones de usuarios en *Google Maps*, lo que sugirió una mayor visibilidad y demanda. Del mismo modo, las que tenían redes sociales tenían más evaluaciones que las que solo contaban con una página web, esto se explicó

² Los corazones productivos son aglomeraciones económicas identificadas por la SDDE. En 2021 se identificaron 66 corazones productivos de escala urbana de los cuales cinco se priorizaron para la atención. La metodología de identificación y priorización puede encontrarse en el Cuaderno 60 del ODEB y puede consultarse a través de este enlace: <https://observatorio.desarrolloeconomico.gov.co/dinamica-empresarial/metodologia-para-la-definicion-y-priorizacion-de-corazones-productivos-en-la>. Para la priorización de 2023 se identificaron 4 corazones productivos adicionales.

³ El Anexo 1 hace una revisión de literatura sobre la estimación de la informalidad.

⁴ API significa “interfaz de programación de aplicaciones”. Las API son herramientas que facilitan la comunicación entre dos componentes de *software* a través de definiciones y protocolos establecidos. Por ejemplo, el sistema de *software* del instituto meteorológico almacena datos del clima diarios. La aplicación de clima de su móvil interactúa con este sistema mediante las API, permitiéndole ver en su teléfono las actualizaciones diarias del tiempo.

⁵ Se utiliza Google Maps o Google de forma intercambiable en el texto para hacer referencia a esta fuente de información.

⁶ Se cuenta con los nombres de las empresas y los establecimientos, pero no de las marcas, razón por la cual este estudio estima un límite superior a la informalidad empresarial de las empresas que se encuentran en *Google Maps* en los nueve corazones productivos.

⁷ Se utiliza negocio y empresa de forma indistinta para referirse a las observaciones de *Google Maps*.

porque las redes sociales sirven como un mecanismo de expansión y retención de clientes, pero también porque existen habilidades de los empresarios que lo motivan a tener redes sociales y además a prestar un buen servicio que le asegure buenas calificaciones. Al analizar estas expresiones de demanda digital a nivel de corazón productivo y según si son formales o informales se encontraron diferencias importantes, como que en el Siete de agosto los negocios informales contaron con más calificaciones promedio en *Google Maps* que las formales⁸.

La investigación también arrojó luz sobre las tendencias en el uso de redes sociales entre estos negocios. Se encontró que Facebook fue la red social más utilizada, seguida por Instagram, Youtube, Twitter, LinkedIn y TikTok.

Este análisis mostró la utilidad de usar *big data* como una fuente adicional de datos para acercarse a las dinámicas del tejido empresarial a nivel territorial. En particular, los análisis planteados proporcionaron una aproximación a la informalidad del tejido empresarial en los corazones productivos priorizados y su digitalización. En esta medida, sirven como una herramienta para quienes formulan políticas públicas, estrategias empresariales y cualquier persona interesada en comprender las dinámicas económicas de la capital.

El documento se divide en cuatro secciones incluyendo esta introducción. La segunda sección expone la metodología de medición de la informalidad empresarial y la caracteriza en los corazones productivos. La tercera explora la relación entre digitalización y formalización. La cuarta sección concluye.

⁸ Es un indicio de que tienen más demanda, luego son más grandes. Este estimativo se hace únicamente para las empresas que cuentan con página web y redes sociales.

2. MEDICIÓN DE LA INFORMALIDAD EMPRESARIAL Y LA HUELLA DIGITAL DE LAS EMPRESAS

Para caracterizar la informalidad empresarial en los corazones productivos de la ciudad, se comparó la información registrada en la CCB con los negocios disponibles en la API de *Google Maps* en nueve zonas de Bogotá denominadas "Corazones Productivos" (CP).

2.1. METODOLOGÍA

En las siguientes secciones se detalla la metodología utilizada para la estimación de la informalidad en tres pasos: obtención de datos de la API de *Google Maps*; la reducción de datos de la CCB para su coincidencia con los CP; y la aplicación de un método de coincidencia difusa (*fuzzy matching*) entre los datos de Google y la CCB para identificar distintos grados de probabilidad de ser formal. Adicionalmente, se analiza la robustez de los resultados comparándolos con los obtenidos con Tropa económica, a través de una encuesta presencial realizada por la Secretaría de Desarrollo Económico a unidades productivas⁹ (Rivera, Moreno, Cárdenas, & Aguilar, 2023).

2.1.1. Obtención de datos de la API de *Google Maps*

En este paso se buscó extraer la información desde el API de *Google Maps* de manera eficiente. Para esto, primero, se generó una circunferencia alrededor del corazón productivo con el radio mínimo desde el centroide del polígono que permitiera cubrir la totalidad del área. La Tabla 1 muestra los radios por corazón productivo y el Mapa 1 muestra un ejemplo para Ricaurte.

⁹ Para más detalles consulte la Nota editorial sobre la Tropa económica en este enlace: <https://observatorio.desarrolloeconomico.gov.co/dinamica-empresarial/nota-editorial-tropa-economica>

Posteriormente, a través del API de *Google Maps*, se extrajo información sobre los negocios en esta zona para cada una de las 96 categorías de negocios que existen en el API de Google. Es decir, se realizaron 96 llamados por corazón productivo correspondientes a los distintos tipos de negocios entre los que se encontraron, por ejemplo, *airport*, *art_gallery*, *bakery*, *bank*, *beauty_salon*¹⁰. Cada llamado extrajo la siguiente información sobre cada negocio: Estatus del negocio (si es operacional, permanentemente o temporalmente cerrado), el tipo de geometría (en este caso contiene las coordenadas), un localizador uniforme de recursos (URL) que dirige al ícono del negocio, el nombre del negocio, si está abierto en el momento de la descarga, algunas fotos del establecimiento, el *rating* promedio de estrellas que le dan los usuarios, los tipos de establecimiento con los que se puede asociar el negocios (*restaurant*, *food*, *establishments*, entre otros), el número total de *ratings* que le han dado los usuarios al restaurante, la dirección, el teléfono de contacto, los horarios de atención, y su sitio web (en caso de tener).

Tabla 1 Radios para extracción de datos de *Google Maps* por corazón productivo

Nombre CP	Radio (metros)
7 de agosto	477
Doce de octubre	505,7
Industria Diversificada - Ricaurte	1.035,7
Industrias Culturales - La Estrada	464,6
Joyereros - La Candelaria	366,8
Restrepo La Valvanera	709,2
San Felipe	523,3
Software - Chicó	1.491,2
Venecia - La Alquería	890,9

Fuente: Corazones productivos SDDE-ODEB. Elaboración SDDE-ODEB

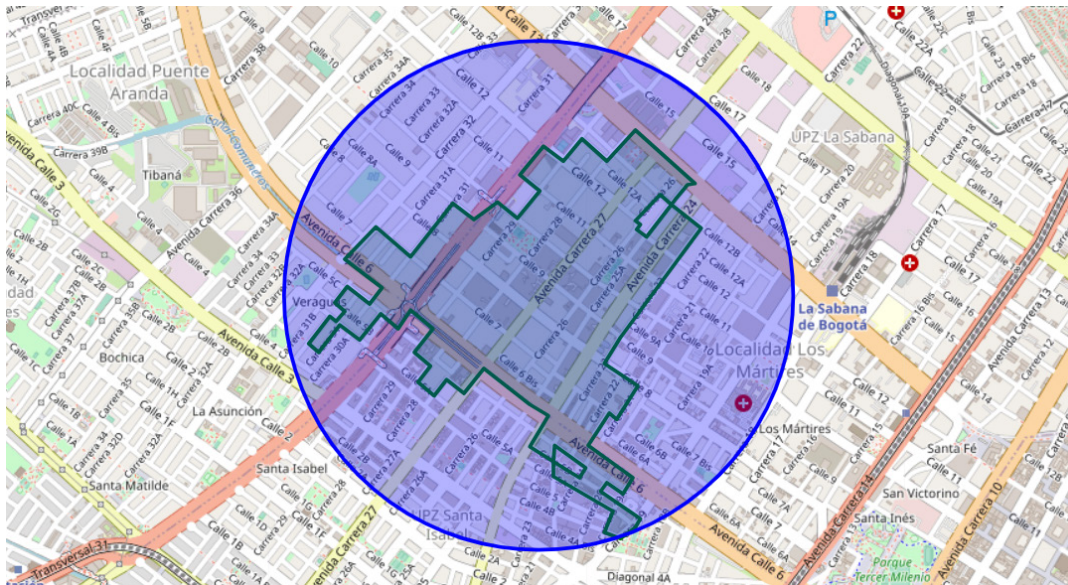
El API permite extraer hasta 60 negocios por llamado por lo que, cuando el número de observaciones extraídas por búsqueda es menor a 60, se puede afirmar que se tiene el universo de ese tipo de negocio, mientras que, si es igual a 60 es necesario hacer un remuestreo de ese tipo de negocio en ese corazón productivo para obtener el universo¹¹.

Solo en cuatro de los 864 llamados se obtuvieron 60 observaciones. Para hacer el remuestreo de estas búsquedas, se partió de los 60 negocios extraídos y se aplicó un algoritmo de *clustering k-means*, para generar polígonos más pequeños dentro del corazón. Acto seguido, se aplicó el ejercicio de extracción sobre los clústeres, al hacer un círculo sobre estos nuevos polígonos y se extrajo la información de los tipos de negocios truncados en cada corazón. El ejercicio se repitió hasta obtener menos de 60 negocios por llamado. En este caso, solo fue necesaria una iteración para obtener el universo de negocios provenientes de *Google Maps*.

¹⁰ https://developers.google.com/maps/documentation/places/web-service/supported_types

¹¹ Este es el universo de empresas que se encuentran en *Google Maps*, no el universo de empresas que se encuentran en el territorio. En *Google Maps* se encuentran empresas que quieren compartir su ubicación o que quieren que los consumidores lleguen a ellas y por lo tanto tienen un establecimiento fijo.

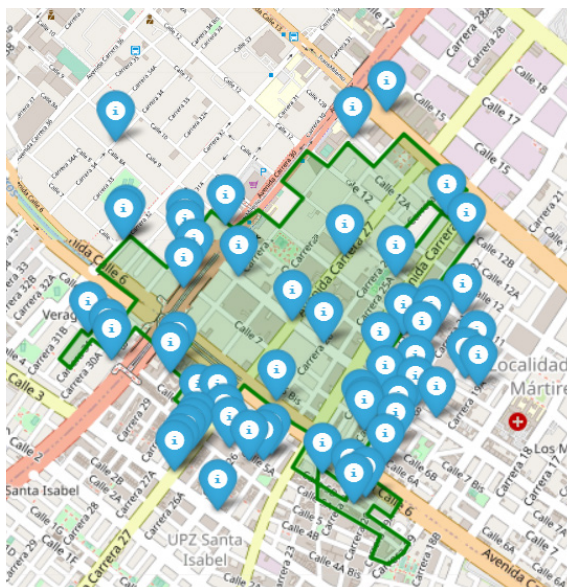
Mapa 1 Circunferencia para la extracción de datos en el corazón productivo Ricaurte



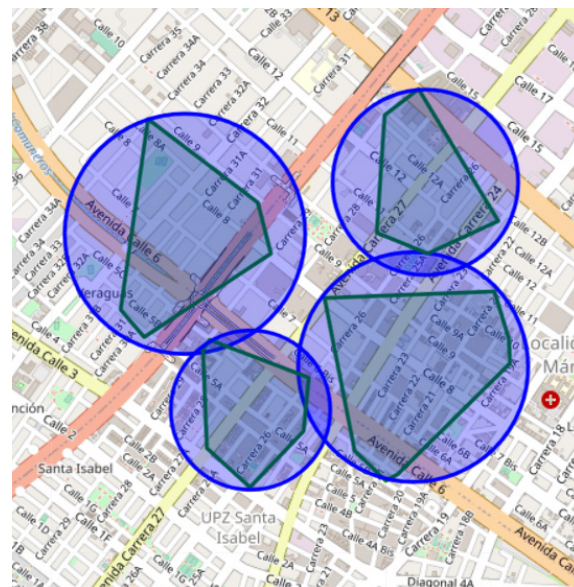
Fuente: Corazones productivos SDDE-ODEB. Elaboración SDDE-ODEB

Posteriormente, se agruparon todos los datos extraídos y se sometieron a un proceso de limpieza que implicó la eliminación de registros duplicados, dado que los círculos se intersectaban en ocasiones (Mapa 2, Mapa 3).

Mapa 2 Distribución de *convenience store* partiendo de la primera extracción de información en Ricaurte



Mapa 3 Clústeres dentro del corazón de Ricaurte de *convenience store*



Fuente: Google Maps. Elaboración SDDE-ODEB

2.1.2. Identificación de empresas formales en el territorio

De acuerdo con el CONPES 3956 de 2019, la formalidad empresarial es un proceso multidimensional, no un estado binario. Existen cuatro tipos de formalidad: “la formalidad de entrada, asociada con la existencia de la empresa, que incluye los requisitos de registro empresarial. Segundo, la formalidad de insumos o factores de producción utilizados por la empresa, que incluye los requisitos para el uso de mano de obra (como el aseguramiento de los trabajadores) y el uso de la tierra (como el emplazamiento del negocio y el uso del suelo). Tercero, la formalidad asociada a los procesos de producción y comercialización de bienes y servicios, que incluye las normas sanitarias, reglamentos técnicos, regulaciones ambientales y otras normas propias del sector en el que opera la empresa. Finalmente, la formalidad tributaria relacionada con las responsabilidades de declarar y pagar impuestos.” (DNP 2019, pag. 23-24)

Este estudio estimó la informalidad empresarial de entrada en los corazones productivos, a partir de una búsqueda difusa entre los nombres de las empresas de *Google Maps* y aquellos del registro mercantil de la CCB. En particular, se utilizaron las empresas con matrícula activa a diciembre de 2022, así como los establecimientos. Únicamente se utilizaron las matrículas activas y establecimientos de tipo jurídico, excluyendo aquellos de tipo natural.

Se seleccionaron las empresas y establecimientos que se ubican al interior de los CP o en un *buffer*¹² de 500 metros. La Tabla 2 describe las zonas de CP en Bogotá de acuerdo con el número de empresas y establecimientos obtenidos a partir de los datos de la CCB, y el número de negocios de la API de *Google Maps*.

El recuento de negocios obtenidos de la API de *Google Maps* fue menor al 10,0 % de las empresas y establecimientos identificados a partir de la CCB, por lo que los resultados estimados capturan la información sobre el universo de *Google Maps*. A pesar de esto, la metodología presentada es una nueva fuente de información para caracterizar la informalidad empresarial en el territorio¹³.

2.1.3. Estimación de la probabilidad de informalidad empresarial

La informalidad empresarial se estimó como una probabilidad de coincidencia entre los nombres de los negocios en *Google Maps*, el universo, y los nombres de las bases de la CCB, las empresas formales. Para esto se estimó un algoritmo de coincidencia difusa (*fuzzy matching*) que encontró correspondencias entre los nombres de los negocios en ambas fuentes de datos, permitiendo así identificar posibles coincidencias entre las entidades comerciales registradas en ambas bases.

¹². Área de influencia por su nombre en español y es un área que está alrededor de la zona de estudio.

¹³. Para más información sobre las empresas estudiadas ver el Anexo 2.

Tabla 2 Frecuencia de negocios del registro mercantil y *Google Maps*

Corazón productivo	CCB (dic-2022)	Google (ago-2023)	Google/CCB (%)
7 de agosto	9,7	335	3,5
Doce de octubre	6,362	174	2,7
Industria Diversificada - Ricaurte	21,53	697	3,2
Industrias Culturales - La Estrada	7,154	458	6,4
Joyereros - La Candelaria	7,014	439	6,3
Restrepo La Valvanera	9,299	833	9,0
San Felipe	7,217	458	6,3
Software - Chicó	32,507	1,423	4,4
Venecia - La Alquería	8,829	874	9,9

Fuente: Corazones productivos SDDE-ODEB, CCB y *Google Maps*. Elaboración SDDE-ODEB

Para garantizar uniformidad y simplificación, los nombres de los negocios en ambas bases se convirtieron a minúsculas y se sometieron a un proceso de limpieza, que incluyó la eliminación de tildes y caracteres especiales. Acto seguido, para cada negocio de *Google Maps* en cada corazón productivo, se buscó si existían empresas y establecimientos en la CCB (del mismo corazón) que tuvieran una o más palabras en común. En caso de no encontrar coincidencia de palabras entre *Google Maps* y la CCB, ese negocio se clasificaba como cero probable. Posteriormente, con los negocios que sí tuvieron alguna coincidencia se llevaba a cabo el cálculo de similaridad de coincidencia, que estima la similitud entre los nombres de los negocios de Google y los nombres de las empresas y establecimientos registrados en la CCB que hicieron *match* en alguna palabra del nombre. Específicamente, se cuantificó la semejanza entre las cadenas de texto, tomando en cuenta la posibilidad de que las palabras estuvieran desordenadas en ambos nombres¹⁴.

Para abordar las situaciones en las que se presentaron múltiples coincidencias, se ejecutó la resolución de coincidencias múltiples. En estos casos, cuando se encontraron negocios con puntajes de mismo valor de similitud, se calculó la distancia euclidiana entre la ubicación geográfica de las posibles coincidencias. Este cálculo permitió la elección del negocio que se encontraba más cerca geográficamente, lo cual contribuyó a definir la coincidencia más adecuada.

Los negocios de *Google Maps* se clasificaron en una de cuatro categorías distintas según el nivel de coincidencia estimado. Estas fueron "muy probable", "probable", "poco probable" y "cero probable"¹⁵. "Muy probable" significaba que hay una probabilidad alta de que

¹⁴. Este cálculo se realizó empleando la función `fuzz.token_sort_ratio` proporcionada por la biblioteca `FuzzyWuzzy` de Python.

¹⁵. Los establecimientos en grado "cero probable" no pasaron por el *match* difuso, estos establecimientos no tienen coincidencias en las palabras de su nombre con las palabras de los nombres de la base de la CCB para cada corazón productivo.

exista coincidencia entre las bases, es decir, que el negocio sea formal, mientras que “cero probable” implicaba que el negocio es informal. El umbral para las categorías se determinaba a partir de conocimientos previos y el análisis de resultados anteriores, con el propósito de asegurar una selección precisa. Adicionalmente, para simplificar la exposición de resultados, se consideraron formales aquellos negocios clasificados como "probable" o "muy probable".

Una de las limitaciones de este enfoque fue que los nombres legales de los negocios no necesariamente coincidían con los nombres usados para el público, por lo que la metodología midió una cota superior de la informalidad. Esto se debió a que las empresas muchas veces tienen marcas registradas a su nombre, y utilizan la marca y no el nombre de la empresa que se encuentra en el registro mercantil para comunicar sus bienes y servicios al público general. Por este motivo, no se tuvo el universo de los posibles nombres formales de las empresas, sino una muestra truncada.

2.2. RESULTADOS

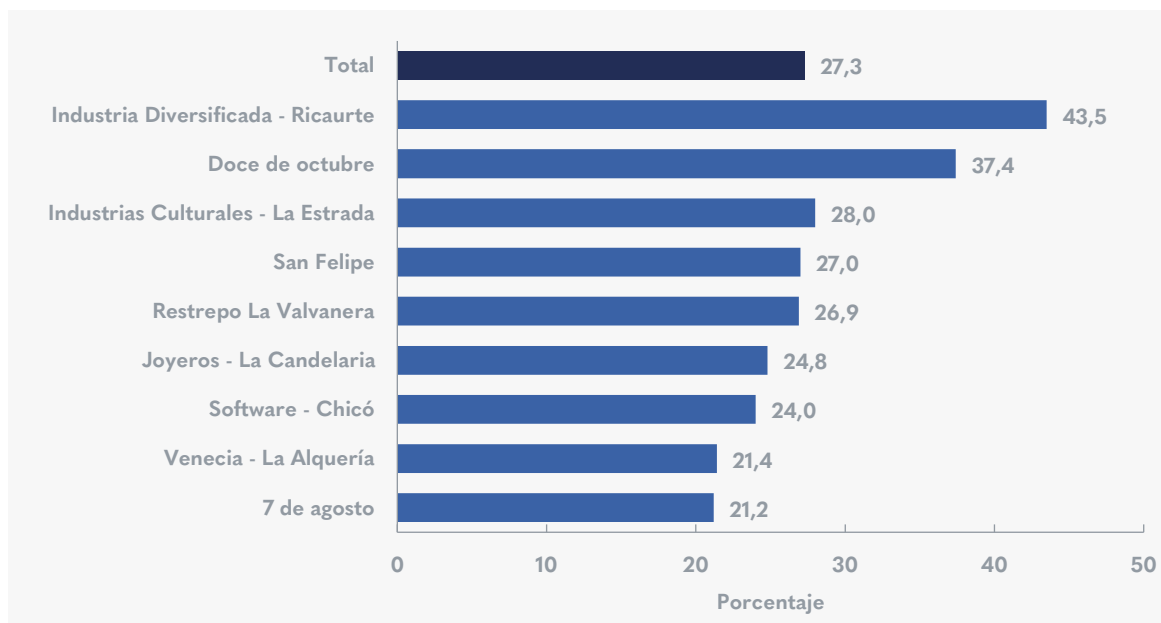
La mayor parte de los negocios de estudio se clasificaron como formales. Se entienden los negocios o empresas formales como aquellos que quedan en los grupos “muy probable” y “probable” (4.138). Mientras que los “poco probables” y los “cero probables” son la minoría (1.553) (Tabla 3). Con base en estos datos la tasa de informalidad en los corazones productivos fue del 27,3 % y varió entre 21,2 % en el 7 de agosto y 43,5 % en Ricaurte (Gráfica 1).

Tabla 3 Negocios de *Google Maps* según la probabilidad de ser formales

Grado de similitud	Empresas	Participación (%)
Muy probable	1,419	24,9
Probable	2,719	47,8
Poco probable	908	16
Cero probable	645	11,3
Total	5,691	100

Fuente: Corazones productivos SDDE-ODEB, CCB y *Google Maps*. Elaboración SDDE-ODEB

Gráfica 1 Tasa de informalidad en los corazones productivos a partir de *Google Maps*



Fuente: Corazones productivos SDDE-ODEB, CCB y *Google Maps*. Elaboración SDDE-ODEB

De otro lado, se estimó que la mayor contribución a la informalidad se obtuvo en el sector servicios. Al analizar la contribución de cada rama de actividad económica a la informalidad por corazón productivo se encontró que la mayor informalidad se dio en el sector servicios, con valores alrededor de 16,7 puntos porcentuales (p.p.). El 7 de agosto fue el único corazón productivo para el cual la contribución a la informalidad de los servicios igualó a la del comercio. En todos los corazones productivos la industria presentó la menor contribución a la informalidad de registro (Tabla 4). La metodología utilizada para clasificar los negocios de *Google Maps* por ramas de actividad se explica en la sección 2.4.

Tabla 4 Contribución a la tasa de informalidad por rama de actividad

Corazón productivo	Comercio (p.p.)	Industria (p.p.)	Servicios (p.p.)	Total (%)
7 de agosto	9,6	2,1	9,6	21,2
Doce de octubre	12,1	3,4	21,8	37,4
Industria Diversificada - Ricaurte	18,4	2,4	22,7	43,5
Industrias Culturales - La Estrada	10,3	3,1	14,6	27,9
Joyereros - La Candelaria	3,2	2,1	19,6	24,8
Restrepo La Valvanera	11,3	1,2	14,4	26,9
San Felipe	5,7	3,5	17,9	27,1
Software - Chicó	6,3	1,1	16,7	24
Venecia - La Alquería	7,9	1,6	11,9	21,4

Fuente: Corazones productivos SDDE-ODEB, CCB y *Google Maps*. Elaboración SDDE-ODEB

Al analizar la tasa de informalidad dentro de cada rama se encontraron resultados similares a los identificados en la Tabla 5. En particular, la informalidad de los servicios fue más alta en todos los corazones productivos. El principal hallazgo de esta tabla fue que ahora, en algunos corazones productivos como en La Candelaria, San Felipe, Venecia – La Alquería y 7 de agosto, la tasa de informalidad de la industria es mayor que la del comercio.

Tabla 5 Tasa de informalidad intra sectorial para los corazones productivos

Corazón productivo	Comercio (%)	Industria (%)	Servicios (%)
7 de agosto	18,4	20,6	25,2
Doce de octubre	32,8	28,6	42,7
Industria Diversificada - Ricaurte	42,8	26,2	47,4
Industrias Culturales - La Estrada	26,6	26,9	29,3
Joyereros - La Candelaria	16,7	24,3	27,0
Restrepo -La Valvanera	24,2	15,4	31,7
San Felipe	23,6	29,6	27,9
Software - Chicó	24,6	22,9	23,9
Venecia - La Alquería	19,3	20,6	23,2

Fuente: Corazones productivos SDDE-ODEB, CCB y *Google Maps*. Elaboración SDDE-ODEB

2.3. ROBUSTEZ

Para medir la robustez de los resultados se compararon estos estimativos con las mediciones de informalidad de entrada a partir de los datos de la Tropa económica. La Tropa no fue planeada para generar una muestra representativa o un censo de los negocios de la ciudad, no obstante, hace un barrido riguroso de los locales abiertos al público y es la única fuente de datos, que, por conocimiento de los autores, permite una desagregación a nivel territorial.

El cálculo de la informalidad estimada a partir de *Google Maps* (informalidad *Google Maps*) fue similar a la estimada a partir de la Tropa (informalidad Tropa). De acuerdo con los datos de la Tropa, el 20,5 % de los negocios en los nueve corazones productivos priorizados tenían informalidad de entrada, estimada como no contar con registro mercantil o tenerlo desactualizado¹⁶. Es decir, a partir de la metodología expuesta se estimó una tasa de informalidad de 6,8 p.p. más alta que aquella estimada a partir de la Tropa.

Hubo mayor variabilidad entre las dos estimaciones de la tasa de informalidad a nivel de corazón productivo. Como se muestra en la Tabla 6, si bien los niveles de las estimaciones fueron similares, se encontraron marcadas diferencias, en particular para Ricaurte, en donde la informalidad *Google Maps* fue mucho mayor (43,5 % vs 14,5 %), así como en el Doce de octubre (37,4 % vs 22,4 %). Dado que ninguna de las dos fuentes de datos cuenta con una muestra representativa de los negocios de la ciudad, se espera una mayor variabilidad en muestras más pequeñas.

¹⁶. No se tienen en cuenta las empresas que no respondieron a esta pregunta.

Tabla 6 Informalidad según *Google Maps* y Tropa económica por corazón productivo

Corazón productivo	Formal -Google	Informal - Google	Tasa de informalidad Google (%)	Tasa de informalidad registro Tropa (%)
7 de agosto	264	71	21,2	17,9
Doce de octubre	109	65	37,4	22,4
Industria Diversificada - Ricaurte	394	303	43,5	14,5
Industrias Culturales - La Estrada	330	128	28	33,2
Joyereros - La Candelaria	330	109	24,8	11,9
Restrepo -La Valvanera	609	224	26,9	23,4
San Felipe	334	124	27	14,8
Software - Chicó	1,081	342	24	11,6
Venecia - La Alquería	687	187	21,4	23,3
Total	4,138	1,553	27,3	20,5

Fuente: Corazones Productivos SDDE-ODEB, CCB, Tropa económica y *Google Maps*. Elaboración SDDE-ODEB

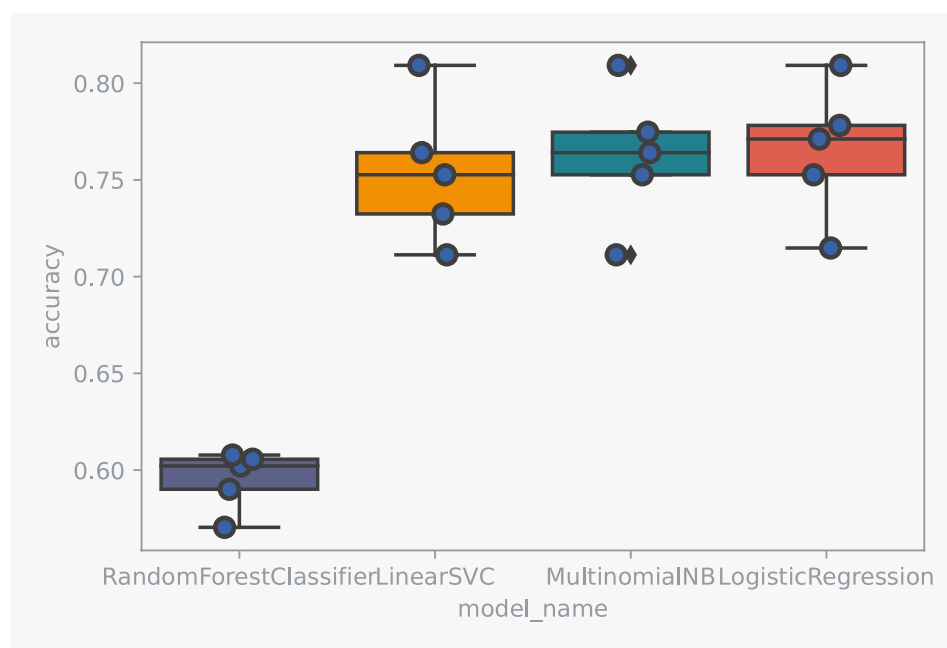
2.4. SESGOS DE LA INFORMACIÓN

Uno de los principales sesgos de la información consignada en *Google Maps* se dio porque no todos los negocios tenían el mismo incentivo a participar en la plataforma. Intuitivamente, negocios que querían atraer un mayor público tenían un mayor incentivo a inscribirse.

Para entender si la información de *Google Maps* contó con sesgos por actividad económica se estimó la participación por sector y se comparó con los resultados del registro mercantil y Tropa. Para generar actividades económicas comparables entre bases se creó un clasificador usando técnicas de *machine learning* para predecir la rama (CIU un dígito) a la que pertenece cada negocio de Google. Este clasificador fue necesario porque la coincidencia entre *Google Maps* y CCB fue difusa. El clasificador utilizó el tipo de lugar para la descarga de los datos de Google y el CIU Rev. 4 a un dígito (rama) de las empresas clasificadas como “muy probables” para entrenar y aprender a predecir la rama de actividad a partir del tipo de lugar de *Google Maps* de las empresas que pertenecen a los grupos probable, poco probable y cero probable.

La Gráfica 2 presenta los resultados del entrenamiento de distintos clasificadores. Para los 4 casos se tomó de manera aleatoria el 66,6 % de la base como datos de entrenamiento y el 33,3 % restante como datos de prueba o validación. Se llevaron a cabo 5 simulaciones con particiones aleatorias de estos datos y se encontró que el *Linear Support Vector Machine* (LSVM), el *multinomial Naive Bayes* y el *Logistic regression* obtuvieron resultados casi idénticos: 78,0 %, siendo el clasificador de regresión logística levemente mejor en el resultado del *accuracy* promedio, por lo que se utilizó para la clasificación. Las correlativas pueden consultarse en el Anexo 3.

Gráfica 2 Resultados de la validación cruzada de algoritmos de clasificación de actividades económicas



Fuente: Corazones productivos SDDE-ODEB, CCB y *Google Maps*. Elaboración SDDE-ODEB

Se encontró que en la mayoría de corazones productivos hubo una mayor participación de negocios de *Google Maps* de servicios y en dos casos de comercio. Mientras que la industria tuvo una menor frecuencia en todos los corazones productivos (Tabla 7).

Tabla 7 Distribución porcentual de los negocios por rama de actividad en los corazones productivos

Corazón productivo	Comercio (%)	Industria (%)	Servicio (%)
7 de agosto	51,9	10,1	37,9
Doce de octubre	36,8	12,1	51,1
Industria Diversificada - Ricaurte	42,9	9,3	47,8
Industrias Culturales - La Estrada	38,6	11,4	50,0
Joyereros - La Candelaria	19,1	8,4	72,4
Restrepo - La Valvanera	46,8	7,8	45,4
San Felipe	24,0	11,8	64,2
Software - Chicó	25,4	4,9	69,6
Venecia - La Alquería	40,8	7,8	51,3

Nota: Se eliminó el sector de agricultura y minas por su baja representación. Las celdas resaltadas muestran la participación máxima por corazón productivo.

Fuente: Corazones Productivos SDDE-ODEB, CCB y *Google Maps*. Elaboración SDDE-ODEB

Al comparar estas distribuciones con la CCB y Tropa, se encontró que en la base de Google se presentó un sesgo hacia el sector servicios. Ya que, salvo para el caso de Chicó, *Google Maps* tiene una mayor concentración de empresas de servicios que las otras dos fuentes (Tabla 8)¹⁷.

Tabla 8 Participación del sector servicios según distintas fuentes de información por corazón productivo, porcentaje

Corazón productivo	CCB (%)	Tropa (%)	Google (%)
7 de agosto	32,0	15,8	37,9
Doce de octubre	34,7	16,3	51,1
Industrias Culturales - La Estrada	41,1	36,4	50,0
Joyerías - La Candelaria	55,8	28,7	72,4
Restrepo - La Valvanera	34,4	20,3	45,4
Ricaurte	25,5	24,4	47,8
San Felipe	45,5	40,0	64,2
Software - Chicó	71,8	60,9	69,6
Venecia - La Alquería	37,2	29,8	51,3

Fuente: Corazones productivos SDDE-ODEB, CCB, Tropa económica y *Google Maps*. Elaboración SDDE-ODEB

¹⁷. Para ver la distribución sectorial por fuente ver el Anexo 2.

3. INFORMALIDAD Y DIGITALIZACIÓN



La digitalización les permite a las empresas obtener ganancias en varios frentes. Primero, amplía el mercado ya que facilita nuevos canales, como las redes sociales, que permiten acceder a consumidores a los que antes no se tenía acceso. Asimismo, la digitalización de las empresas mejora el desempeño de las mismas al tener la capacidad de disponer de información en tiempo real (ODEB, 2022).

En este contexto, este capítulo muestra cómo pueden usarse los datos en línea para explorar la relación entre informalidad y digitalización. Primero, se detalla la metodología utilizada para la caracterización de la huella digital de las empresas, y luego se exponen los resultados y su relación con los niveles de formalidad.

3.1. MEDICIÓN DE LA DIGITALIZACIÓN DE LAS EMPRESAS

Entre los datos extraídos a través del API de *Google Maps* se encontró la página web cuando esta existía y los negocios la revelaban. También se extrajo el *rating* que los usuarios le dieron a los negocios en *Google Maps* y el número de votaciones obtenidas.

Posteriormente, para aquellos negocios que contaban con una página web, se realizó una extracción y análisis de contenido de redes sociales asociadas a la misma. Para identificar y recolectar estos enlaces, se utilizaron técnicas de búsqueda dentro del contenido HTML de las páginas web. Una vez identificados, se eliminaron registros duplicados y se enfocó la búsqueda en dominios específicos de redes sociales, como Facebook, Instagram, YouTube, LinkedIn, TikTok y X (antes Twitter).

Para garantizar la relevancia y precisión de los enlaces, se implementó un proceso de coincidencia difusa (*fuzzy matching*) entre las páginas de redes sociales y las páginas originales de los negocios (dominios). De esta forma, se seleccionaron los enlaces más relevantes y representativos para cada negocio.

3.2. RESULTADOS

Según los resultados obtenidos de un total de 5.691 negocios registrados en *Google Maps*, 1.786 (31,4 %) contaban con una página web. De este subconjunto, se identificó que 918 (16,1 %) de estos negocios contaban con presencia en redes sociales. Estos números sugirieron una amplia gama de niveles de adopción digital entre los negocios en estudio que varía por corazón productivo y grado de similaridad como se muestra en la Tabla 9. A nivel de corazón productivo, se encuentra que Chicó es el corazón más digital (52,6 % de sus negocios cuenta con un sitio web), seguido por San Felipe (37,7 %). Las empresas formales, en promedio tienden a tener más páginas web que las informales, así como redes sociales (RRSS).

Tabla 9 Digitalización de los negocios de *Google Maps* por corazón productivo y formalidad

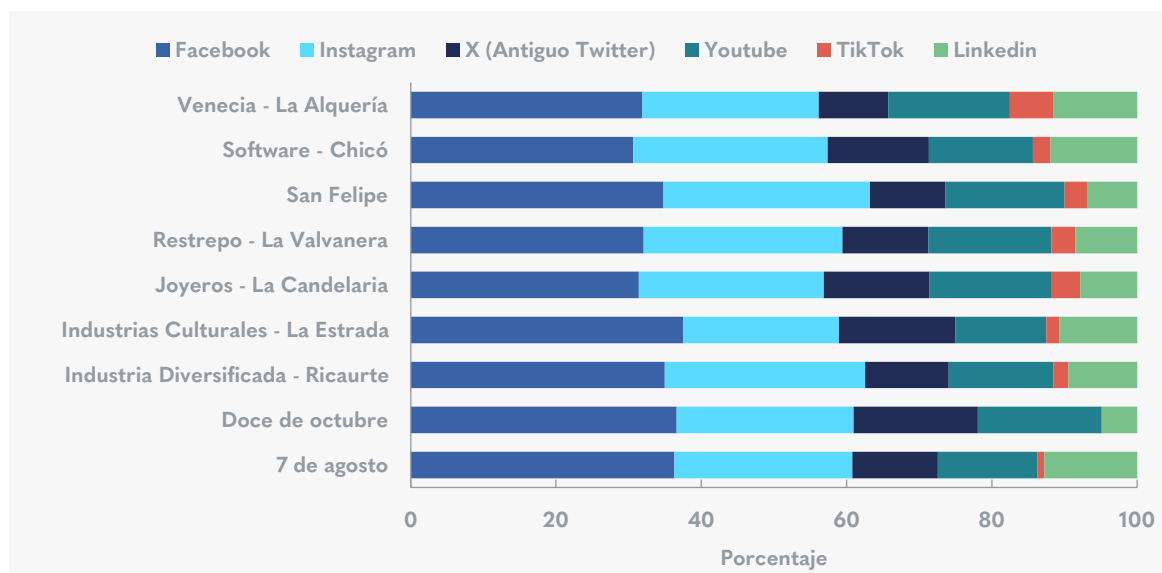
Corazón productivo	Total empresas	Con página web (%)	Con página web y redes sociales (%)
Por corazón productivo			
7 de agosto	335	28,4	12,2
Doce de octubre	174	23,6	8,6
Industria Diversificada - Ricaurte	697	27,1	13,2
Industrias Culturales - La Estrada	458	18,6	4,6
Joyerías - La Candelaria	439	33,9	19,6
Restrepo - La Valvanera	833	16,4	8,5
San Felipe	458	39,7	20,1
Software - Chicó	1,423	52,6	29,4
Venecia - La Alquería	874	18,2	9,3
Por formalidad			
Formal	4,138	31,1	15,8
Informal	1,553	32,2	16,9
Total	5,691	31,4	16,1

Fuente: Corazones Productivos SDDE-ODEB, *Google Maps*. Elaboración SDDE-ODEB

El porcentaje de adopción de las redes sociales por corazón productivo mostró que Facebook fue la red más utilizada por los negocios en casi todos los corazones, seguido muy de cerca por Instagram, X, YouTube, LinkedIn y TikTok, respectivamente. No se identificó presencia de negocios con TikTok en el Doce de octubre (Gráfica 3). Al analizar a grandes rasgos las diferencias en la apropiación de redes sociales entre empresas formales e informales, se encontró que no hay diferencias significativas en la participación de estas redes al comparar ambos grupos¹⁸.

¹⁸. Para ver la proporción de empresas con adopción de cada red social ver anexo 4.

Gráfica 3 Proporción de empresas con cada red social por corazón productivo



Fuente: Corazones productivos SDDE-ODEB, *Google Maps*. Elaboración SDDE-ODEB

De otro lado, se encontró una relación positiva entre el número de calificaciones recibidas por el negocio en Google¹⁹ y la probabilidad de ser formal. El número de votaciones puede interpretarse como las expresiones de demanda digital: más votaciones se relacionan con mayor demanda. Como se muestra en la Tabla 10, el número promedio de calificaciones recibidas por los negocios creció por grado de similitud desde 34 calificaciones para los negocios con “cero probable” a 143 para aquellos con “muy probable”.

Adicionalmente, para todos los niveles de probabilidad, el promedio del número de calificaciones de todos los negocios fue menor que el de aquellos que cuentan con una página web, y, asimismo, que los que además de tener página web cuentan con redes sociales. Por ejemplo, para los negocios con “probable” el número de calificaciones fue de 81, 156 y 179, respectivamente.

Tabla 10 Número promedio de calificaciones de los negocios de *Google Maps* por grado de similitud y digitalización

Grado de similitud	Universo de <i>Google Maps</i>	Con página web	Con página web y redes sociales
Muy probable	142,7	304,5	448,2
Probable	81,4	155,7	179,3
Poco probable	72,1	100,6	123,5
Cero probable	34	51,4	79

Fuente: Corazones productivos SDDE-ODEB, *Google Maps*. Elaboración SDDE-ODEB

¹⁹. El número de *ratings* hace referencia al número de personas que calificaron un establecimiento, independientemente de la calificación, o si ponen comentarios.

Particularmente, los negocios que tenían redes sociales en promedio tenían más calificaciones que los que no tenían redes. Lo anterior puede reflejar el efecto multiplicador de las redes sobre los clientes potenciales, al ser canales de comunicación que no solo sirven para crear la identidad digital de las marcas, sino que también son medios de comunicación que le permiten a empresas tanto grandes como pequeñas hacer pauta en línea ampliando su mercado potencial.

Cabe resaltar que estas relaciones no son causales, ya que características no observadas, como el espíritu emprendedor o las habilidades digitales se relacionan con que el negocio tenga mayor probabilidad de ser formal, contar con redes sociales y ofrecer productos de buena calidad que sean más reconocidos por el público a través de calificaciones.

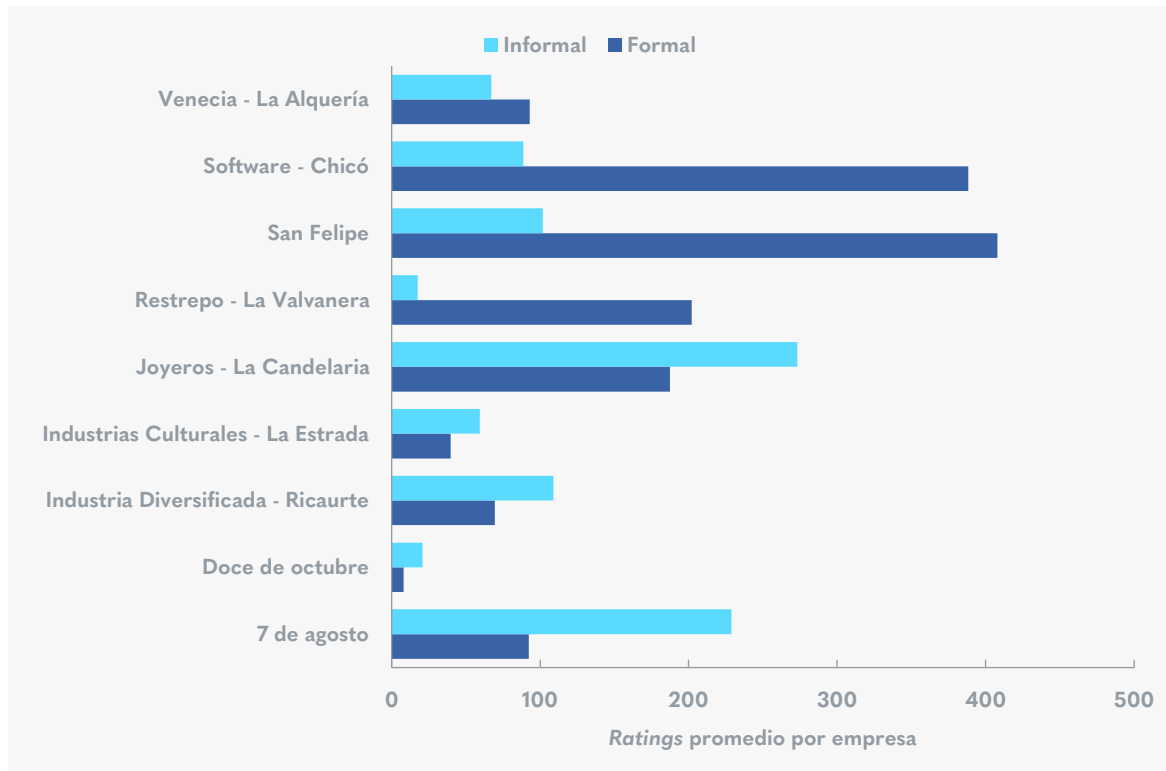
Tabla 11 Número promedio de calificaciones de los negocios de *Google Maps* por nivel de formalidad y niveles excluyentes de digitalización

Nivel de formalidad	Solo página web	Con página web y redes sociales
Formal	135,5	280,5
Informal	46,8	107,5

Fuente: Corazones productivos SDDE-ODEB, *Google Maps*. Elaboración SDDE-ODEB

El análisis de las expresiones de demanda digital según grados de informalidad y corazón productivo reveló interesantes patrones en la ciudad. Por ejemplo, un negocio informal promedio que tiene página web y redes sociales en el Siete de agosto o en La Candelaria tiene más calificaciones promedio que el formal promedio. Situación que se invirtió completamente en Chicó o San Felipe, donde el negocio formal típico captó la mayor demanda digital (Gráfica 4).

Gráfica 4 Número promedio de calificaciones de negocios con página web y redes sociales formales e informales por corazón productivo



Fuente: Corazones productivos SDDE-ODEB, *Google Maps*. Elaboración SDDE-ODEB

4. CONCLUSIONES

Este estudio ofrece una mirada a la informalidad empresarial en Bogotá a través del uso de técnicas de *big data*. Estas técnicas proporcionan información que no es fácilmente observable a través de registros administrativos, como los negocios con informalidad de entrada o el nivel de digitalización de las empresas en el territorio.

El estudio estimó una cota superior de la tasa de informalidad a nivel territorial a partir un algoritmo de coincidencia difusa entre los nombres de los negocios extraídos de *Google Maps* y los registros de la matrícula mercantil de la Cámara de Comercio de Bogotá. La tasa estimada de informalidad empresarial de entrada en los 9 corazones productivos fue del 27,3 %, según la metodología propuesta. El corazón con mayor informalidad fue Ricaurte (43,5 %), mientras que el de menor informalidad fue Siete de agosto (21,2 %).

Se encontró un sesgo de los datos de *Google Maps* hacia el sector servicios, por lo que la informalidad se explicó principalmente por las contribuciones de los sectores de servicios y comercio. La industria tuvo una baja contribución a la informalidad en todos los corazones productivos.

De otro lado, hubo una relación positiva entre formalidad y digitalización. Las empresas más formales tienen mayor presencia digital, medida a través de más calificaciones de usuarios en *Google Maps*. Asimismo, las empresas con sitio web y redes sociales recibieron más calificaciones que las que solo tenían sitio web.

Sin embargo, en algunos corazones productivos como Siete de agosto y La Candelaria, las empresas informales superaron en calificaciones promedio a las formales. Esto sugiere dinámicas particulares en la demanda de estos territorios. Facebook e Instagram fueron las redes sociales más utilizadas por las empresas en los diferentes corazones productivos.

Los resultados de este estudio pueden servir para orientar políticas públicas de formalización empresarial en Bogotá, al identificar tanto los territorios como los sectores económicos donde se concentra la informalidad.

Asimismo, la metodología propuesta muestra el potencial que tiene el uso de técnicas de *big data*. En este caso, permitió analizar las dinámicas urbanas de informalidad a una escala detallada. Futuros estudios podrían ampliar el análisis a toda la ciudad e incorporar otras fuentes de información, como la coincidencia entre nombre legal de las empresas y las marcas usadas por los negocios.



5. BIBLIOGRAFÍA

Castaño, S. J., Pardo, W. M., Rodríguez, D. M., Rivera, J. P., & Aguilar, J. (2023). Metodología para la definición y priorización de corazones productivos en la ciudad de Bogotá. Bogotá: Secretaría Distrital de Desarrollo Económico.

DNP. (2019). POLÍTICA DE FORMALIZACIÓN EMPRESARIAL. Bogotá: Departamento Nacional de Planeación.

Galvis, L. A. (2012). Informalidad laboral en las áreas urbanas de Colombia. Documentos de Trabajo Sobre Economía Regional. Banco de la República, 164, 1-42. Obtenido de https://repositorio.banrep.gov.co/bitstream/handle/20.500.12134/3116/dtser_164.pdf

Gan, G. (2020). Singapore's Experience in Analyzing the Labor Market Using Artificial Intelligence and Big Data Analytics. En *Education in the Asia-Pacific Region: Issues, Concerns and Prospects* (págs. 255-262). Asian Development Bank. Obtenido de <https://link.springer.com/book/10.1007/978-981-15-7018-6>

Giambona, F., Khalawi, A., Buzzigoli, L., Grassini, L., & Martelli, C. (2021). Big data analysis and labour market: an analysis of Italian online job vacancies data. En *ASA 2021 Statistics and Information Systems for Policy Evaluation* (págs. 117-120). Florencia: Firenze University Press. Obtenido de <https://media.fupress.com/files/pdf/24/9476/26306>

Hutter, C. (2021). Cyclicity of labour market search: a new big data approach. *Journal for Labour Market Research*, 55(1), 1-16. doi:<https://doi.org/10.1186/s12651-020-00283-9>

Rivera, J., Moreno, I. R., Cárdenas, D., & Aguilar, J. (2023). Nota editorial: Tropa Económica. Bogotá: Secretaría de Desarrollo Económico.

Salcedo Pérez, C., Moscoso Durán, F., & Ramírez Salazar, M. (2020). Economía informal en Colombia: iniciativas y propuestas para reducir su tamaño. *Revista Espacios*, 41(3), 22-28. Obtenido de <https://www.revistaespacios.com/a20v41n03/20410322.html>

Sánchez, R. M. (2020). Poverty and labor informality in Colombia. *IZA Journal of Labor Policy*, 10(1), 1-20. doi: <https://doi.org/10.2478/izajolp-2020-0006>

Sánchez, R. M., Manzano, L. D., & Maturana, L. A. (2022). Informalidad laboral, pobreza monetaria y multidimensional en Bogotá y el Área Metropolitana. *Problemas del desarrollo*, 53(208), 31-63. doi: <https://doi.org/10.22201/iiec.20078951e.2022.208.69754>

Vankevich, A., & Kalinouskaya, I. (2021). Better understanding of the labour market using Big Data. *Ekonomia i Prawo. Economics and Law*, 20(3), 677–692. doi: <https://doi.org/10.12775/EiP.2021.040>.

ANEXO 1. REVISIÓN BIBLIOGRÁFICA

El estado del arte relacionado con las estimaciones de la informalidad a partir de técnicas de *big data* tiene dos grandes aproximaciones metodológicas. Por un lado, se encuentra la literatura relacionada con la caracterización de la informalidad laboral en Colombia, sus causas y consecuencias. Sin embargo, es importante aclarar que esta literatura se ha centrado principalmente en la descripción de las características de los trabajadores informales, más que en el de las empresas informales, a diferencia del presente trabajo. Por otro lado, el segundo enfoque de la literatura relevante se encuentra relacionada con el uso del *big data* para el análisis del mercado laboral en diferentes partes del mundo.

Galvis (2012) realiza una caracterización de la informalidad en Colombia y discute los diferentes criterios que se han planteado para definir la informalidad. Emplea la definición de asociada a la no afiliación a seguridad social (salud y pensión) para medir la informalidad en las veintitrés principales ciudades del país y encuentra que a pesar de que existen diferencias regionales en los factores asociados a la informalidad, en términos generales los informales son principalmente individuos de bajos ingresos, trabajan mayoritariamente en establecimientos de menor tamaño y viven en ciudades periféricas. Las ciudades con mayores índices de informalidad son aquellas que se encuentran localizadas por fuera del centro andino conformado por la vecindad de ciudades como Bogotá, Cali, Medellín y Bucaramanga.

Asimismo, Salcedo Pérez, Moscoso Durán, & Ramírez Salazar (2020) realizan una revisión del concepto, las causas y las consecuencias de la informalidad; así como las propuestas gubernamentales para reducir la informalidad en Colombia y sus verdaderos impactos. Hasta ahora el enfoque de las políticas públicas ha sido el de facilitar la creación de empresas en la formalidad y reducir la carga tributaria y regulatoria más que en la inspección, vigilancia y control de las condiciones laborales.

Sánchez (2020) y Sánchez, Manzano, & Maturana (2022) por otro lado, se enfocan en analizar el vínculo que existe entre la pobreza y la informalidad, el primero en Colombia y el segundo en el caso particular de Bogotá. Ambos artículos concluyen que la reducción de la

informalidad tendría impactos positivos en la pobreza y la desigualdad. Sánchez, Manzano, & Maturana (2022) toman la información del DANE, según la cual Bogotá y sus alrededores presentan un nivel de informalidad del 41,1 % para el año 2020, y desagregan estos datos por localidades. Encuentran que las áreas de Bogotá con una mayor tasa de informalidad son Usme-Sumapaz (56,2 %) y Ciudad Bolívar (53,9 %), mientras que las que presentan una menor tasa de informalidad son Chapinero (21,5 %) y Usaquén (23,2 %). De esta forma, presentan dónde viven los trabajadores informales de Bogotá, complementario al análisis de este trabajo en el cual se busca mostrar dónde se encuentran ubicadas las empresas informales de la ciudad.

En cuanto a la bibliografía relevante relacionada al uso de *big data* para analizar el mercado laboral, se encontró que la mayoría de la literatura se ha enfocado en el uso de los datos de portales de empleo. Gan (2020) y Vankevich & Kalinouskaya (2021) por ejemplo, utilizan esta información para estudiar la brecha entre la demanda y oferta de habilidades y competencias laborales para el caso de Singapur y Bielorrusia, respectivamente. Hutter (2021) por otro lado, estudia el comportamiento de las actividades de búsqueda y colocación de la bolsa de trabajo de la Agencia federal de empleo de Alemania para determinar patrones estacionales. Giambona, Khalawi, Buzzigoli, Grassini, & Martelli (2021) utilizan la información provista por Burning Glass Technologies para estudiar al cambio en la demanda de habilidades laborales en Italia después de la pandemia de Covid-19.

ANEXO 2. NEGOCIOS Y DISTRIBUCIÓN SECTORIAL EN LOS CORAZONES PRODUCTIVOS

Tabla A 1 Número de negocios por corazón productivo y distintas fuentes de información (2021-2023)

Corazón productivo	CCB (dic 2022)	Google (agosto)	Conteo unidades económicas (2021)	Tropa (17 julio 2023)
Industria Diversificada - Ricaurte	21.530	697	4.945	1.597
Software - Chicó	32.507	1.423	8.535	426
Joyereros - La Candelaria	7.014	439	1.672	478
Industrias Culturales - La Estrada	7.154	458	1.185	782
San Felipe	7.217	458	1.283	460
Restrepo - La Valvanera	9.299	833	2.929	1.232
Venecia - La Alquería	8.829	874	3.205	1.739
7 de agosto	9.700	335	859	560
Doce de octubre	6.362	174	1.154	1.244

Fuente: Corazones productivos SDDE-ODEB, *Google Maps*, CCB y Tropa económica. Elaboración SDDE-ODEB

Tabla A 2 Distribución sectorial en los corazones productivos según diversas fuentes de información

Corazón productivo	CCB					Tropa					Google						
	Agropecuaria y minera (%)	Comercio (%)	Industria (%)	Otras actividades (%)	Servicio (%)	Agropecuaria y minera (%)	Comercio (%)	Industria (%)	Servicio (%)	Agropecuaria y minera (%)	Comercio (%)	Industria (%)	Servicio (%)	Agropecuaria y minera (%)	Comercio (%)	Industria (%)	Servicio (%)
7 de agosto	0,7	49,5	17,7	0,1	32,0	0,0	61,4	22,8	15,8	0,0	51,9	10,1	37,9	0,0	51,9	10,1	37,9
Doce de octubre	0,9	38,7	25,6	0,1	34,7	0,1	47,5	36,1	16,3	0,0	36,8	12,1	51,1	0,0	36,8	12,1	51,1
Industrias Culturales - La Estrada	0,9	34,2	23,7	0,1	41,1	0,2	28,0	35,4	36,4	0,0	38,6	11,4	50,0	0,0	38,6	11,4	50,0
Joyerías - La Candelaria	1,9	28,8	13,4	0,1	55,8	0,0	33,7	37,6	28,7	0,0	19,1	8,4	72,4	0,0	19,1	8,4	72,4
Restrepo - La Valvanera	0,4	40,0	25,1	0,1	34,4	0,1	39,2	40,3	20,3	0,0	46,8	7,8	45,4	0,0	46,8	7,8	45,4
Ricaurte	0,4	53,0	21,0	0,1	25,5	0,1	34,9	40,6	24,4	0,0	42,9	9,3	47,8	0,0	42,9	9,3	47,8
San Felipe	1,2	33,0	20,2	0,1	45,5	1,1	40,8	18,1	40,0	0,0	24,0	11,8	64,2	0,0	24,0	11,8	64,2
Software - Chico	3,6	15,1	9,5	0,1	71,8	1,3	17,4	20,4	60,9	0,1	25,4	4,9	69,6	0,1	25,4	4,9	69,6
Veneçia - La Alquería	0,4	39,3	22,9	0,2	37,2	0,1	50,5	19,6	29,8	0,1	40,8	7,8	51,3	0,1	40,8	7,8	51,3

* No fue posible estimar el sector para 2.640 negocios de Tropa por problemas en la digitación del CIJU. El total de negocios de tropa es 8.518 (30% sin sector económico)

Fuente: Corazones productivos SDDE-ODEB, Google Maps, CCB y Tropa económica. Elaboración SDDE-ODEB

ANEXO 3. CORRELATIVA ENTRE LUGARES DE *GOOGLE MAPS* Y SECTORES ECONÓMICOS

Tabla A 3 Tipos de negocios de *Google Maps* y rama predicha por el clasificador

Tipo	Comercio	Industria	Servicio	Tipo	Comercio	Industria	Servicio
accounting	0	0	1	funeral_home	0	0	1
aquarium	0	0	1	furniture_store	1	0	0
art_gallery	0	0	1	gas_station	1	0	0
atm	0	0	1	gym	0	0	1
bakery	0	1	0	hair_care	0	0	1
bank	0	0	1	hardware_store	1	0	0
bar	0	0	1	home_goods_store	1	0	0
beauty_salon	0	0	1	hospital	0	0	1
bicycle_store	1	0	0	insurance_agency	0	0	1
book_store	0	0	1	jewelry_store	1	0	0
bowling_alley	0	0	1	laundry	0	0	1
bus_station	0	0	1	lawyer	0	0	1
cafe	0	0	1	library	0	0	1
car_dealer	1	0	0	liquor_store	1	0	0
car_rental	0	0	1	local_government_office	0	0	1
car_repair	1	0	0	locksmith	1	0	0
car_wash	1	0	0	lodging	0	0	1
casino	0	0	1	meal_delivery	0	0	1
cemetery	0	0	1	meal_takeaway	0	0	1
church	0	0	1	mosque	1	0	0
clothing_store	0	1	0	movie_rental	0	0	1
convenience_store	1	0	0	movie_theater	0	0	1
courthouse	0	0	1	moving_company	0	0	1
dentist	0	0	1	museum	0	0	1
department_store	1	0	0	night_club	0	0	1
doctor	0	0	1	painter	0	0	1
drugstore	0	0	1	park	0	0	1
electrician	1	0	0	parking	0	0	1
electronics_store	1	0	0	pet_store	1	0	0
embassy	0	0	1	pharmacy	1	0	0
fire_station	0	0	1				
florist	1	0	0				

Tipo	Comercio	Industria	Servicio
physiotherapist	0	0	1
plumber	0	0	1
police	1	0	0
post_office	0	0	1
primary_school	0	0	1
real_estate_agency	0	0	1
restaurant	0	0	1
roofing_contractor	0	0	1
rv_park	0	0	1
school	0	0	1
shoe_store	1	0	0

Tipo	Comercio	Industria	Servicio
shopping_mall	1	0	0
spa	0	0	1
storage	1	0	0
store	1	0	0
supermarket	1	0	0
tourist_attraction	0	0	1
transit_station	0	0	1
travel_agency	0	0	1
university	0	0	1
veterinary_care	0	0	1
zoo	0	0	1

Fuente: Corazones productivos SDDE-ODEB, *Google Maps*, CCB. Elaboración SDDE-ODEB

ANEXO 4. PROPORCIÓN DE ADOPCIÓN DE REDES DE EMPRESAS FORMALS E INFORMALES

Tabla A 4 Proporción de las firmas totales de *Google Maps* con redes sociales

	Instagram (%)	Facebook (%)	YouTube (%)	X (antiguo Twitter) (%)	TikTok (%)	LinkedIn (%)	Total (%)
Formal	12,3	14,7	6,7	5,4	1,3	4,5	15,8
Informal	12,4	15,6	8,1	7,7	1,4	5,8	16,9

Fuente: Corazones productivos SDDE-ODEB, *Google Maps*, CCB. Elaboración SDDE-ODEB

Siglas

CCB	Cámara de Comercio de Bogotá
DNP	Departamento Nacional de Planeación
SDDE	Secretaría Distrital de Desarrollo Económico

Abreviaturas

CP	Corazón productivo
<i>Fuzzy matching</i>	Coincidencia difusa
URL	Localizador uniforme de recursos
LSVM	Linear Support Vector Machine
RRSS	redes sociales

Índice de mapas

Mapa 1	Circunferencia para la extracción de datos en el corazón productivo Ricaurte	13
Mapa 2	Distribución de <i>convenience store</i> partiendo de la primera extracción de información en Ricaurte	13
Mapa 3	Clústeres dentro del corazón de Ricaurte de <i>convenience store</i>	13

Índice de gráficas

Gráfica 1	Tasa de informalidad en los corazones productivos a partir de <i>Google Maps</i>	17
Gráfica 2	Resultados de la validación cruzada de algoritmos de clasificación de actividades económicas	20
Gráfica 3	Proporción de empresas con cada red social por corazón productivo	24
Gráfica 4	Número promedio de calificaciones de negocios con página web y redes sociales formales e informales por corazón productivo	26

Índice de tablas

Tabla 1	Radio para extracción de datos de <i>Google Maps</i> por corazón productivo	12
Tabla 2	Frecuencia de negocios del registro mercantil y <i>Google Maps</i>	15
Tabla 3	Negocios de <i>Google Maps</i> según la probabilidad de ser formales	16
Tabla 4	Contribución a la tasa de informalidad por rama de actividad	17
Tabla 5	Tasa de informalidad intra sectorial para los corazones productivos	18
Tabla 6	Informalidad según <i>Google Maps</i> y Tropa económica por corazón productivo	19
Tabla 7	Distribución porcentual de los negocios por rama de actividad en los corazones productivos	20
Tabla 8	Participación del sector servicios según distintas fuentes de información por corazón productivo, porcentaje	21
Tabla 9	Digitalización de los negocios de <i>Google Maps</i> por corazón productivo y formalidad	23
Tabla 10	Número promedio de calificaciones de los negocios de <i>Google Maps</i> por grado de similaridad y digitalización	24
Tabla 11	Número promedio de calificaciones de los negocios de <i>Google Maps</i> por nivel de formalidad y niveles excluyentes de digitalización	25
Tabla A 1	Número de negocios por corazón productivo y distintas fuentes de información (2021-2023)	32
Tabla A 2	Distribución sectorial en los corazones productivos según diversas fuentes de información	33
Tabla A 3	Tipos de negocios de <i>Google Maps</i> y rama predicha por el clasificador	34
Tabla A 4	Proporción de las firmas totales de <i>Google Maps</i> con redes sociales	35

