

**Modelo predictivo para la transición a la formalidad de comerciantes informales en  
Antioquia**

Presentado por

Katherin Angelica Chávez Bermúdez

Karla Estefanía Barrientos Vélez

Director

Gabriel Jaime Rivera León

MSc en Ingeniería – Ingeniería Administrativa

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas Tecnología e Ingeniería ECBTI

Ingeniería Industrial

2026

## Resumen

La informalidad comercial continúa siendo uno de los principales desafíos socioeconómicos en Colombia, especialmente en Antioquia, donde una proporción significativa de comerciantes desarrolla sus actividades sin acceso a mecanismos formales de regulación, protección social y fortalecimiento empresarial. En este contexto, la presente investigación tuvo como objetivo construir y validar un modelo de machine learning que permitiera predecir la probabilidad de formalización de comerciantes informales en Antioquia a partir de variables socioeconómicas, laborales y contextuales, así como proponer incentivos focalizados para apoyar procesos de formalización. La investigación se desarrolló mediante un enfoque metodológico mixto, integrando análisis cuantitativo y revisión documental. Para ello, se recolectó información mediante 108 encuestas aplicadas a comerciantes informales ubicados en diferentes zonas del departamento de Antioquia. Posteriormente, se realizó un proceso de limpieza, depuración y codificación de variables para estructurar la base de datos utilizada en el análisis predictivo. El modelo implementado correspondió a una regresión logística supervisada desarrollada en Python mediante librerías especializadas para análisis de datos y machine learning. El modelo fue evaluado mediante métricas de desempeño como accuracy, precision y recall. Adicionalmente, se implementó validación cruzada de cinco particiones (5-fold cross validation) y un proceso de optimización de hiperparámetros mediante GridSearchCV, fortaleciendo la consistencia metodológica y estabilidad del modelo. Los resultados evidenciaron que variables relacionadas con el nivel educativo, los ingresos y las barreras percibidas frente a la formalización presentan asociaciones relevantes con los comportamientos analizados. Asimismo, se identificó que factores como los altos costos administrativos, la complejidad de los trámites y las limitaciones institucionales constituyen barreras importantes para los procesos de transición hacia la

formalidad. De manera complementaria, se desarrollaron visualizaciones interactivas mediante Power BI para facilitar la interpretación de los resultados. Finalmente, la investigación permitió estructurar una propuesta de incentivos focalizados orientada a fortalecer procesos de formalización ajustados a las características y necesidades identificadas en la población analizada. Este estudio se adscribe a la línea de investigación “Modelos de Gestión Organizacional”, al integrar herramientas analíticas y enfoques aplicados orientados al análisis de fenómenos socioeconómicos y al fortalecimiento de procesos de toma de decisiones relacionadas con la formalización empresarial.

***Palabras clave:*** Informalidad laboral, Formalización empresarial, Modelos predictivos, Aprendizaje automático, Regresión logística, Modelos de gestión organizacional.

## Abstract

Commercial informality continues to be one of the main socioeconomic challenges in Colombia, particularly in the department of Antioquia, where a significant proportion of merchants carry out their economic activities without access to formal mechanisms of regulation, social protection, and business development. In this context, the objective of this research was to build and validate a machine learning model capable of predicting the probability of formalization among informal merchants in Antioquia based on socioeconomic, labor, and contextual variables, as well as to propose targeted incentives to support formalization processes. The study was conducted using a mixed-methods approach that integrated quantitative analysis and documentary review. Data were collected through 108 surveys administered to informal merchants located in different areas of Antioquia. Subsequently, a process of data cleaning, refinement, and variable coding was carried out to structure the database used for predictive analysis. The implemented model consisted of a supervised logistic regression developed in Python using specialized libraries for data analysis and machine learning. Model performance was evaluated using accuracy, precision, and recall metrics. In addition, five-fold cross-validation and a hyperparameter optimization process through GridSearchCV were implemented to strengthen the methodological consistency and stability of the model. The results showed that variables related to educational level, income, and perceived barriers to formalization were associated with the analyzed behaviors. Likewise, factors such as administrative costs, procedural complexity, and institutional limitations were identified as significant barriers to formalization. Interactive visualizations were also developed in Power BI to facilitate the interpretation of the results. Finally, the research enabled the design of a targeted incentive proposal aimed at strengthening formalization processes according to the characteristics and needs identified in the studied population. This study is part of the

Organizational Management Models research line by integrating analytical tools and applied approaches focused on the analysis of socioeconomic phenomena and decision-making processes related to business formalization.

***Keywords:*** Informal employment, Business formalization, Predictive models, Machine learning, Logistic regression, Organizational management models.

## Tabla de contenido

Introducción .....	12
Planteamiento del problema.....	14
Pregunta de investigación .....	15
Justificación .....	16
Objetivos.....	18
Objetivo general .....	18
Objetivos específicos .....	18
Estado del arte.....	19
La economía informal: una visión general.....	19
Dinámicas de la informalidad .....	20
Factores económicos .....	20
Información y tecnología .....	21
Políticas gubernamentales y formalización.....	21
Comprensión contextual de la informalidad .....	22
Modelado predictivo y machine learning aplicado al análisis de informalidad.....	23
Marco conceptual y teórico.....	25
Marco conceptual .....	25
Factores socioeconómicos.....	25
Factores institucionales .....	26

Factores sociales y culturales .....	26
Factores tecnológicos .....	26
Articulación conceptual con la investigación .....	27
Marco teórico .....	27
Enfoques teóricos de la informalidad.....	27
Teoría del capital humano .....	28
Teoría del capital social.....	28
Teorías institucionales.....	29
Teoría del Comportamiento Planeado.....	29
Articulación teórica con la investigación.....	29
Metodología .....	31
Diseño de investigación .....	31
Fase 1: Preparación del instrumento y prueba piloto.....	32
Fase 2: Recolección de datos .....	34
Fase 3: Análisis de datos y elaboración de informe .....	35
Métodos de recolección de datos .....	38
Encuestas cuantitativas.....	38
Análisis cualitativo documental .....	39
Análisis de datos .....	39
Análisis cuantitativo.....	39

Análisis cualitativo.....	40
Consideraciones éticas .....	40
Limitaciones.....	41
Resultados.....	42
Identificación de variables asociadas a la probabilidad de formalización de comerciantes informales: .....	42
Caracterización de los comerciantes informales encuestados.....	42
Procesamiento, depuración y codificación de la información recolectada: .....	44
Recolección de la información.....	44
Tabulación y depuración de la información .....	45
Codificación de variables.....	46
Estructuración de la base de datos para el modelo predictivo .....	48
Construcción y evaluación del modelo predictivo:.....	50
Selección de variables y definición de la variable objetivo .....	50
División de los datos en entrenamiento y prueba.....	51
Entrenamiento del modelo mediante regresión logística .....	52
Evaluación inicial del modelo.....	53
Validación cruzada del modelo .....	54
Optimización de hiperparámetros mediante GridSearchCV.....	55
Análisis de correlación entre variables .....	57

Propuesta de incentivos focalizados para la formalización.....	59
Identificación de factores relevantes para la formulación de incentivos .....	59
Diseño de la propuesta de incentivos focalizados.....	59
Visualización e interpretación de resultados mediante Power BI .....	61
Discusión.....	63
Conclusiones y recomendaciones .....	66
Referencias.....	68

## Lista de figuras

<b>Figura 1</b> Formulario de encuesta en Google Forms.....	32
<b>Figura 2</b> Ejemplo de preguntas incluidas en el cuestionario .....	33
<b>Figura 3</b> Aplicación de la prueba piloto a comerciantes informales.....	34
<b>Figura 4</b> Proceso de recolección de información en campo .....	34
<b>Figura 5</b> Base de datos original de encuestas .....	35
<b>Figura 6</b> Base de datos depurada para análisis .....	36
<b>Figura 7</b> Base de datos codificada para modelado.....	37
<b>Figura 8</b> Código del modelo en Python.....	37
<b>Figura 9</b> División de los datos en conjuntos de entrenamiento y prueba .....	52
<b>Figura 10</b> Resultados de evaluación del modelo predictivo mediante métricas .....	53
<b>Figura 11</b> Resultados de la validación cruzada aplicada al modelo predictivo .....	55
<b>Figura 12</b> Optimización de hiperparámetros mediante GridSearchCV.....	56
<b>Figura 13</b> Matriz de correlación entre variables .....	57
<b>Figura 14</b> Visualización interactiva de resultados en Power BI .....	61

### Lista de tablas

<b>Tabla 1</b> Caracterización general de los comerciantes informales encuestados .....	43
<b>Tabla 2</b> Codificación de variables utilizadas en el modelo predictivo .....	47
<b>Tabla 3</b> Variables finales utilizadas en el modelo predictivo .....	49
<b>Tabla 4</b> Propuesta de incentivos para la formalización empresarial .....	60

## Introducción

La informalidad comercial constituye uno de los principales desafíos socioeconómicos en Colombia, debido a su impacto sobre la estabilidad laboral, la productividad, el acceso a protección social y el desarrollo económico de los territorios. En departamentos como Antioquia, una parte significativa de los comerciantes desarrolla actividades económicas bajo esquemas informales, sin registro legal ni acceso a mecanismos institucionales de fortalecimiento empresarial. Esta situación evidencia no solo limitaciones estructurales del mercado laboral, sino también dificultades económicas, educativas e institucionales que afectan los procesos de formalización.

En este contexto, comprender los factores que influyen en la permanencia de los comerciantes dentro de la informalidad representa un aspecto fundamental para el diseño de estrategias de intervención más efectivas y focalizadas. Variables como el nivel educativo, los ingresos, las condiciones laborales y las percepciones frente a la formalización pueden influir en la toma de decisiones de los comerciantes informales y en sus posibilidades de transición hacia actividades económicas formales.

A partir de esta problemática, la presente investigación tuvo como propósito construir y validar un modelo de machine learning que permitiera predecir la probabilidad de formalización de comerciantes informales en Antioquia, utilizando variables socioeconómicas, laborales y contextuales recolectadas mediante encuestas aplicadas directamente a la población objeto de estudio. Asimismo, la investigación buscó generar una propuesta de incentivos focalizados orientada a fortalecer procesos de formalización ajustados a las características identificadas en los comerciantes analizados.

Para el desarrollo del estudio se implementó un enfoque metodológico mixto, integrando análisis cuantitativo y revisión documental contextual. La información fue recolectada mediante 108 encuestas aplicadas a comerciantes informales ubicados en diferentes zonas del departamento de Antioquia. Posteriormente, se realizó un proceso de depuración, codificación y análisis de datos para el desarrollo de un modelo predictivo basado en regresión logística supervisada implementada en Python.

De manera complementaria, se realizaron análisis de correlación y visualizaciones interactivas mediante Power BI, permitiendo interpretar patrones asociados a la probabilidad de formalización e identificar factores relevantes dentro de la población estudiada. Los resultados obtenidos evidenciaron relaciones relevantes entre variables socioeconómicas y la intención de formalización, así como barreras relacionadas con costos administrativos, complejidad de trámites y limitaciones institucionales.

En este sentido, la investigación busca aportar herramientas analíticas aplicadas al estudio de la informalidad comercial, integrando análisis cuantitativo, interpretación contextual y técnicas básicas de machine learning como apoyo para la comprensión de fenómenos socioeconómicos complejos y el diseño de estrategias de intervención más focalizadas y ajustadas a las condiciones reales del territorio.

## Planteamiento del problema

En la actualidad, la informalidad comercial constituye uno de los desafíos más persistentes para la economía colombiana, especialmente en regiones como Antioquia, donde una proporción significativa de trabajadores independientes desarrolla sus actividades sin registro legal ni acceso a mecanismos institucionales de apoyo empresarial. Esta situación no solo evidencia la limitada oferta de empleo formal, sino también la complejidad de las decisiones que enfrentan quienes emprenden en contextos económicos inestables y con condiciones sociales desiguales.

En este escenario, los comerciantes informales suelen operar con limitaciones en conocimientos administrativos, financieros y de gestión, lo que dificulta la sostenibilidad de sus negocios y su posible transición hacia la formalidad. A partir de la información recolectada en esta investigación, se evidencia que factores como el nivel educativo, los ingresos y las condiciones laborales influyen en la forma en que estos comerciantes toman decisiones frente a la formalización, lo cual coincide con la presencia de prácticas empíricas y baja planificación estructurada en sus actividades económicas.

Por otra parte, los programas institucionales orientados a promover la formalización presentan dificultades en términos de cobertura, accesibilidad y pertinencia frente a las necesidades reales de los comerciantes. A partir de la información recolectada mediante las encuestas aplicadas, se identificó que muchos comerciantes perciben los procesos de formalización como costosos, complejos y poco ajustados a sus condiciones económicas y operativas, lo que genera desinterés y refuerza la permanencia en la informalidad.

Asimismo, el análisis de las variables socioeconómicas y contextuales permitió identificar posibles relaciones entre factores como el nivel educativo, los ingresos, las

condiciones laborales y las barreras percibidas frente a la intención de formalización de los comerciantes informales. Estos elementos evidencian la necesidad de comprender el fenómeno desde una perspectiva integral que articule factores individuales, económicos e institucionales.

En este contexto, el uso de herramientas analíticas basadas en datos, como los modelos de machine learning, representa una alternativa para identificar patrones y analizar comportamientos asociados a la formalización de comerciantes informales. Esto permite no solo ampliar la comprensión del fenómeno, sino también generar insumos orientados al diseño de estrategias e incentivos focalizados de acuerdo con las características de la población analizada.

En síntesis, la persistencia de la informalidad comercial en Antioquia plantea la necesidad de desarrollar herramientas de análisis que integren variables socioeconómicas y contextuales para apoyar la comprensión de los factores asociados a la formalización y contribuir al fortalecimiento del desarrollo económico local.

### **Pregunta de investigación**

¿Cómo pueden las variables socioeconómicas y de ubicación contribuir, mediante un modelo de machine learning, a predecir la probabilidad de formalización de comerciantes informales en Antioquia?

## Justificación

La informalidad comercial continúa representando uno de los principales desafíos socioeconómicos en Colombia y particularmente en el departamento de Antioquia, donde una parte significativa de los comerciantes desarrolla sus actividades fuera de los mecanismos formales de regulación, protección social y fortalecimiento empresarial. Esta situación genera efectos directos sobre la estabilidad económica de los comerciantes, limita sus oportunidades de crecimiento y afecta el desarrollo económico y social de los territorios.

En el desarrollo de la presente investigación, a partir de la aplicación de encuestas a comerciantes informales, se evidenció que muchos de ellos perciben la formalización como un proceso costoso, complejo y poco ajustado a sus condiciones reales. Asimismo, variables como el nivel educativo, los ingresos, las condiciones laborales y las barreras administrativas influyen significativamente en la forma en que los comerciantes toman decisiones frente a los procesos de formalización.

De igual manera, el análisis contextual permitió identificar que, aunque existen programas institucionales orientados a promover la formalización, estos presentan limitaciones relacionadas con cobertura, accesibilidad y adaptación a las necesidades reales de los comerciantes informales. Esto reduce la confianza en las estrategias institucionales y limita el impacto de las políticas orientadas al fortalecimiento empresarial y la transición hacia la formalidad.

Frente a esta problemática, se hace necesario desarrollar herramientas analíticas que permitan comprender mejor los factores asociados a la informalidad y apoyar la toma de decisiones basada en evidencia. En este sentido, el uso de técnicas básicas de machine learning, como la regresión logística supervisada implementada en esta investigación, permite identificar

patrones relevantes relacionados con la probabilidad de formalización a partir de variables socioeconómicas y contextuales.

La importancia de este estudio radica en que los resultados obtenidos permiten no solo comprender factores asociados a la permanencia en la informalidad, sino también generar insumos para el diseño de estrategias e incentivos más focalizados y ajustados a las condiciones reales de los comerciantes informales del departamento de Antioquia. Asimismo, la investigación aporta una aproximación aplicada al uso de herramientas analíticas en el estudio de fenómenos socioeconómicos, contribuyendo al fortalecimiento de procesos de toma de decisiones relacionados con la formalización empresarial y el desarrollo económico territorial.

## Objetivos

### Objetivo general

Construir y validar un modelo de machine learning que prediga la probabilidad de formalización de comerciantes informales en Antioquia, a partir de variables socioeconómicas y de ubicación, y proponer incentivos focalizados orientados a la formalización de comerciantes informales.

### Objetivos específicos

Identificar las variables socioeconómicas y territoriales asociadas a la probabilidad de formalización de comerciantes informales en Antioquia.

Recopilar, depurar y codificar la información obtenida mediante encuestas aplicadas a comerciantes informales, con el fin de estructurar la base de datos para el desarrollo del modelo predictivo.

Construir y evaluar un modelo de regresión logística supervisada que permita predecir la probabilidad de formalización de comerciantes informales a partir de variables socioeconómicas y de ubicación.

Proponer incentivos focalizados orientados a la formalización de comerciantes informales, con base en los resultados obtenidos del modelo predictivo y el análisis de variables relevantes.

## Estado del arte

### **La economía informal: una visión general**

La economía informal constituye uno de los principales desafíos económicos y sociales en América Latina y el Caribe, debido a su impacto sobre la productividad, la estabilidad laboral y el acceso a mecanismos de protección social. Este sector se caracteriza por el desarrollo de actividades económicas que operan parcialmente por fuera de los sistemas formales de regulación estatal, lo que genera limitaciones tanto para los trabajadores como para el crecimiento económico de los territorios (Porta & Shleifer, 2014).

En el contexto colombiano, la informalidad ha representado históricamente una proporción significativa del empleo. Según el DANE (2023), más del 55 % de los trabajadores urbanos desarrollan actividades informales, siendo el comercio y los servicios personales algunos de los sectores con mayor participación. Este fenómeno se relaciona con factores estructurales como la baja productividad, las limitaciones del mercado laboral formal, los costos de regulación y las desigualdades territoriales. En departamentos como Antioquia, la presencia de microempresas y emprendimientos familiares refuerza estas dinámicas, evidenciando la necesidad de estrategias de formalización adaptadas a las particularidades regionales (Arévalo & Tobón, 2022).

Diversos estudios señalan que, aunque la economía informal representa una alternativa de subsistencia para numerosos trabajadores, también perpetúa condiciones de vulnerabilidad relacionadas con el acceso limitado a crédito formal, protección social y oportunidades de crecimiento económico (Banco Mundial, 2021). No obstante, investigaciones recientes destacan que este sector también cumple una función adaptativa en contextos de crisis económicas y desempleo, absorbiendo parte de la población excluida del mercado laboral formal (OIT, 2022).

Asimismo, la literatura reciente ha resaltado la importancia de las tecnologías digitales y las herramientas analíticas para comprender mejor las dinámicas de informalidad y apoyar el diseño de estrategias de intervención más focalizadas. En países como México y Perú, el uso de herramientas digitales ha permitido identificar patrones territoriales de informalidad y fortalecer procesos de análisis relacionados con políticas de formalización (Cárdenas & Mejía, 2023).

En este contexto, la presente investigación retoma estos enfoques para analizar la informalidad comercial en el departamento de Antioquia, integrando revisión documental, análisis contextual y procesamiento de información recolectada directamente de comerciantes informales mediante herramientas analíticas y modelos predictivos básicos.

## **Dinámicas de la informalidad**

### ***Factores económicos***

La incertidumbre y la volatilidad económica influyen significativamente en el crecimiento y permanencia de la economía informal. Nguyen et al. (2022) señalan que los periodos de crisis económicas suelen incrementar la participación en actividades informales, debido a que numerosos individuos recurren a estas estrategias como mecanismo de subsistencia frente a la falta de oportunidades laborales formales.

En Colombia, las dinámicas macroeconómicas también han incidido directamente sobre los niveles de informalidad. De acuerdo con el DANE (2023), los periodos de desaceleración económica presentan aumentos importantes en actividades comerciales informales, especialmente en sectores de bajo nivel de regulación. Adicionalmente, las diferencias salariales y productivas entre los sectores formal e informal continúan incentivando la permanencia de numerosos trabajadores en esquemas informales de empleo (García & Mendoza, 2021).

En Antioquia, la informalidad presenta un fuerte componente territorial, especialmente en municipios con menor diversificación económica y menores oportunidades de empleo formal,

donde el autoempleo y el comercio informal representan alternativas de generación de ingresos para numerosas familias.

### ***Información y tecnología***

El desarrollo de las Tecnologías de la Información y la Comunicación (TIC) ha transformado las dinámicas económicas contemporáneas, incluyendo el comportamiento de los sectores informales. Farhadi et al. (2012) destacan que el acceso a información, plataformas digitales y herramientas tecnológicas puede facilitar procesos de transición hacia la formalidad, al mejorar el acceso de los comerciantes a recursos, capacitación y redes de apoyo.

Asimismo, diferentes investigaciones evidencian que la simplificación digital de trámites y procesos administrativos puede reducir barreras relacionadas con la formalización empresarial. En países como Brasil y Chile, la implementación de plataformas digitales gubernamentales ha contribuido a disminuir tiempos y complejidad administrativa asociados al registro de pequeños negocios (OCDE, 2023).

De igual manera, el uso de herramientas digitales puede fortalecer la trazabilidad de las actividades comerciales y facilitar el acceso a programas de capacitación, financiación y fortalecimiento empresarial (CAF, 2022). En este sentido, la digitalización se convierte en un mecanismo relevante para apoyar procesos de inclusión económica y formalización empresarial en territorios con alta presencia de comercio informal.

### ***Políticas gubernamentales y formalización***

Las políticas públicas y las estrategias institucionales desempeñan un papel fundamental en los procesos de formalización empresarial. Diversos estudios señalan que la calidad institucional, la simplificación administrativa y el acceso a programas de apoyo pueden influir significativamente en la reducción de los niveles de informalidad (Mawejje, 2024).

No obstante, investigaciones recientes también evidencian que numerosos programas de formalización presentan limitaciones relacionadas con cobertura, pertinencia y adaptación a las condiciones reales de los comerciantes informales (Ávila et al., 2023). Esta situación genera desconfianza frente a las instituciones y reduce la efectividad de las estrategias implementadas.

En Colombia, diferentes políticas han buscado promover la formalización mediante incentivos tributarios, simplificación normativa y programas de fortalecimiento empresarial. Sin embargo, estudios recientes indican que factores como los costos de cotización, las cargas administrativas y la limitada percepción de beneficios continúan representando barreras importantes para trabajadores y pequeños comerciantes informales (Deza et al., 2020).

Experiencias internacionales también muestran que los procesos de formalización son más efectivos cuando las estrategias institucionales se acompañan de capacitación empresarial, acceso a crédito y mecanismos de acompañamiento territorial (Tobar-Cruz & Ruiz-Martínez, 2025). Asimismo, la estrategia FORLAC 2.0 de la Organización Internacional del Trabajo (OIT, 2023) plantea que la formalización sostenible requiere intervenciones ajustadas a las realidades económicas y sociales de cada territorio.

### ***Comprensión contextual de la informalidad***

Diversas investigaciones destacan que la informalidad debe analizarse desde una perspectiva integral que considere factores económicos, sociales, culturales e institucionales. Variables como el capital social, las redes comunitarias y las percepciones frente a las instituciones influyen directamente en la disposición de los comerciantes para vincularse a esquemas formales de actividad económica (Porta & Shleifer, 2014).

En territorios como Antioquia, las diferencias entre zonas urbanas y rurales generan condiciones heterogéneas relacionadas con acceso a oportunidades, información y acompañamiento institucional. Según Maldonado y Peña (2023), las percepciones sobre el

Estado y la confianza institucional afectan directamente los procesos de formalización, especialmente en contextos donde los comerciantes consideran que las políticas públicas no responden a sus necesidades reales.

Adicionalmente, estudios recientes evidencian que las desigualdades sociales, educativas y tecnológicas también condicionan las posibilidades de formalización. La Organización Internacional del Trabajo (OIT, 2022) señala que los procesos sostenibles de formalización requieren estrategias integrales que no se limiten únicamente a incentivos económicos, sino que también aborden problemáticas relacionadas con exclusión social, acceso a educación y fortalecimiento de capacidades empresariales.

### ***Modelado predictivo y machine learning aplicado al análisis de informalidad***

El desarrollo de herramientas de análisis predictivo y machine learning ha permitido ampliar las posibilidades de estudio de fenómenos socioeconómicos complejos, incluyendo la economía informal. A través de modelos supervisados como la regresión logística, es posible identificar patrones y relaciones entre variables socioeconómicas asociadas a determinados comportamientos o probabilidades dentro de poblaciones específicas.

En el contexto de la informalidad comercial, diferentes investigaciones han señalado que el uso de modelos predictivos puede aportar herramientas analíticas útiles para identificar perfiles con mayor probabilidad de formalización, facilitando el diseño de estrategias e intervenciones más focalizadas. Variables como ingresos, nivel educativo, condiciones laborales y acceso a información suelen presentar relaciones significativas con la disposición de los comerciantes frente a procesos de formalización.

Asimismo, la integración de herramientas de visualización y análisis de datos permite fortalecer la interpretación de resultados y facilitar la comunicación de hallazgos a entidades públicas y privadas encargadas de diseñar políticas y programas de intervención.

En este sentido, la presente investigación incorpora herramientas básicas de machine learning mediante la implementación de un modelo de regresión logística supervisada orientado a analizar la probabilidad de formalización de comerciantes informales en Antioquia. A partir de la información recolectada y procesada mediante técnicas de limpieza y codificación de datos, se buscó identificar relaciones entre variables socioeconómicas y la intención de formalización, aportando una aproximación aplicada al uso de herramientas analíticas para el estudio de la informalidad comercial.

## **Marco conceptual y teórico**

### **Marco conceptual**

La transición de los comerciantes informales hacia la economía formal constituye un fenómeno complejo influenciado por factores económicos, sociales, institucionales y tecnológicos. Este proceso no depende únicamente de variables financieras, sino también de condiciones relacionadas con el acceso a información, percepción institucional, capacidades individuales y contexto territorial. Comprender estas dinámicas resulta fundamental para analizar los factores asociados a la probabilidad de formalización en comerciantes informales del departamento de Antioquia.

### ***Factores socioeconómicos***

Los factores socioeconómicos representan elementos centrales dentro de los procesos de formalización, debido a que influyen directamente en la capacidad de los comerciantes para asumir costos, adaptarse a requisitos legales y acceder a oportunidades de crecimiento empresarial.

Entre las variables más relevantes se encuentran el nivel educativo, los ingresos, la experiencia laboral y las condiciones de trabajo. Diversos estudios señalan que comerciantes con mayores niveles de educación y estabilidad económica presentan una mayor disposición hacia la formalización, debido a que poseen mejores capacidades para comprender procesos administrativos y reconocer beneficios asociados al mercado formal (Fiess & Fugazza, 2012).

Asimismo, el acceso a recursos financieros y oportunidades de fortalecimiento empresarial influye significativamente en la sostenibilidad de los negocios y en la posibilidad de asumir procesos de transición hacia esquemas formales de actividad económica.

### ***Factores institucionales***

Las instituciones públicas y el entorno regulatorio desempeñan un papel fundamental dentro de los procesos de formalización. Elementos como la complejidad de los trámites, los costos administrativos, la percepción de legitimidad institucional y el acceso a programas gubernamentales pueden incentivar o desincentivar la formalización de los comerciantes informales.

Investigaciones recientes evidencian que programas institucionales accesibles, acompañados de procesos de orientación y simplificación administrativa, pueden fortalecer significativamente la disposición hacia la formalización (Jokia et al., 2021). Por el contrario, entornos regulatorios complejos o percibidos como poco accesibles suelen reforzar la permanencia en esquemas informales de trabajo.

### ***Factores sociales y culturales***

Las redes sociales, el capital comunitario y las percepciones colectivas también influyen en la toma de decisiones de los comerciantes informales. La confianza en las instituciones, las experiencias compartidas dentro de las comunidades y las dinámicas de apoyo entre comerciantes pueden facilitar procesos de transición hacia la formalidad.

En contextos donde predominan percepciones negativas frente al Estado o experiencias desfavorables relacionadas con trámites administrativos, los comerciantes tienden a percibir la formalización como un proceso riesgoso o poco beneficioso (Bruns et al., 2011).

### ***Factores tecnológicos***

Las herramientas tecnológicas y el acceso a información digital pueden contribuir a reducir barreras relacionadas con la formalización empresarial. La digitalización de trámites, el acceso a plataformas de información y el uso de herramientas tecnológicas permiten facilitar

procesos administrativos y mejorar el acceso de los comerciantes a servicios financieros y oportunidades de capacitación (Liang & Aquino, 2024).

En este sentido, la tecnología representa un mecanismo complementario que puede fortalecer procesos de inclusión económica y facilitar el acceso de los comerciantes informales a entornos formales de actividad empresarial.

### ***Articulación conceptual con la investigación***

En el desarrollo de la presente investigación, estos componentes conceptuales orientaron la selección de variables incluidas en el instrumento de recolección de información y en el modelo predictivo desarrollado. Variables como nivel educativo, ingresos, condiciones laborales y barreras percibidas frente a la formalización fueron incorporadas dentro del análisis cuantitativo y del modelo de regresión logística supervisada implementado en Python.

De esta manera, el marco conceptual permitió sustentar el análisis de los factores asociados a la probabilidad de formalización de comerciantes informales en Antioquia desde una perspectiva integral y contextualizada.

### **Marco teórico**

La economía informal comprende aquellas actividades económicas que operan parcial o totalmente por fuera de los mecanismos formales de regulación estatal, incluyendo aspectos tributarios, laborales y administrativos (OIT, 2022). En Colombia, la informalidad constituye uno de los principales desafíos estructurales del mercado laboral, afectando variables relacionadas con productividad, estabilidad económica y acceso a protección social (IMF, 2018).

### ***Enfoques teóricos de la informalidad***

La informalidad ha sido analizada desde diferentes enfoques teóricos. La corriente legalista plantea que numerosos trabajadores y comerciantes optan por la informalidad debido a la complejidad normativa y a los altos costos asociados a la formalización (Sánchez, 2023).

Desde esta perspectiva, la informalidad surge como una respuesta racional frente a entornos regulatorios percibidos como restrictivos o poco accesibles.

Por su parte, la visión estructuralista interpreta la informalidad como un componente funcional dentro de economías con fuertes desigualdades y limitaciones del mercado laboral formal, donde el sector informal absorbe excedentes de mano de obra y actúa como mecanismo de subsistencia económica.

Asimismo, enfoques institucionales recientes destacan que la informalidad no depende únicamente de factores económicos, sino también de elementos relacionados con legitimidad institucional, confianza en el Estado y apropiación de normas sociales y regulatorias (Laguna et al., 2023).

### ***Teoría del capital humano***

La Teoría del Capital Humano plantea que variables como educación, experiencia y capacitación representan inversiones que incrementan la productividad y las oportunidades económicas de los individuos. En el contexto de comerciantes informales, mayores niveles de educación y formación pueden facilitar procesos administrativos, fortalecer capacidades empresariales y mejorar la comprensión de los beneficios asociados a la formalización.

Estudios recientes para Colombia evidencian que la educación continua siendo un factor determinante en la generación de ingresos y en las oportunidades de inserción económica (Mora et al. (2023)

### ***Teoría del capital social***

La Teoría del Capital Social resalta la importancia de las redes sociales, relaciones de confianza y dinámicas comunitarias dentro de los procesos económicos. En el contexto de la informalidad, estas redes pueden facilitar el acceso a información, apoyo financiero informal y experiencias relacionadas con procesos de formalización.

Sin embargo, algunos estudios también señalan que redes excesivamente cerradas pueden reforzar dinámicas de permanencia dentro de esquemas informales y limitar el acceso a oportunidades de crecimiento empresarial (Rubio, 1997).

### ***Teorías institucionales***

La Nueva Economía Institucional plantea que las instituciones y los costos de transacción influyen directamente sobre las decisiones económicas de los individuos. Costos elevados de formalización, trámites complejos y baja claridad regulatoria pueden desincentivar procesos de transición hacia la formalidad.

De igual manera, la Teoría Institucional señala que factores como legitimidad, confianza institucional y aceptación social de las normas influyen significativamente en la disposición de los comerciantes para vincularse al mercado formal.

### ***Teoría del Comportamiento Planeado***

La Teoría del Comportamiento Planeado, desarrollada por Ajzen, explica que las decisiones individuales se encuentran influenciadas por las actitudes personales, las normas sociales percibidas y la percepción de control frente a determinadas conductas.

Aplicada al contexto de la informalidad, esta teoría permite comprender cómo las percepciones sobre costos, beneficios, experiencias previas y capacidad para cumplir requisitos administrativos influyen en la intención de formalización de los comerciantes informales.

### ***Articulación teórica con la investigación***

A partir de estos enfoques teóricos, la presente investigación estructuró el análisis de variables relacionadas con factores socioeconómicos, institucionales y conductuales asociados a la probabilidad de formalización de comerciantes informales en Antioquia.

La integración de estos enfoques permitió sustentar teóricamente el desarrollo del modelo predictivo basado en regresión logística supervisada, así como interpretar los resultados obtenidos desde una perspectiva multidimensional del fenómeno de la informalidad comercial.

## **Metodología**

La presente investigación tuvo como propósito construir y validar un modelo predictivo basado en machine learning para analizar la probabilidad de formalización de comerciantes informales en Antioquia, a partir de variables socioeconómicas y contextuales recolectadas mediante encuestas.

El estudio se desarrolló bajo un enfoque predominantemente cuantitativo, debido a que se trabajó con datos estructurados y técnicas estadísticas orientadas al análisis predictivo mediante regresión logística supervisada. De manera complementaria, se realizó una revisión documental de literatura académica e informes institucionales relacionados con la informalidad y los procesos de formalización en Colombia, con el fin de contextualizar e interpretar los resultados obtenidos.

La investigación se llevó a cabo durante un período aproximado de tres meses, comprendido entre febrero y abril, tiempo en el cual se realizó el diseño de instrumentos, recolección de información, procesamiento de datos y desarrollo del modelo predictivo.

### **Diseño de investigación**

La investigación adoptó un diseño no experimental de alcance descriptivo y predictivo, orientado al análisis de factores asociados a la probabilidad de formalización de comerciantes informales en Antioquia.

El componente cuantitativo constituyó el eje principal del estudio, permitiendo recopilar información socioeconómica, laboral y contextual mediante encuestas estructuradas aplicadas a comerciantes informales ubicados en diferentes zonas del departamento de Antioquia, incluyendo Medellín, Copacabana, Doradal y Perales.

Como complemento, se desarrolló una revisión documental basada en literatura académica, informes institucionales y fuentes secundarias relacionadas con la informalidad comercial y los procesos de formalización, con el propósito de ampliar la comprensión contextual del fenómeno estudiado.

El desarrollo de la investigación se estructuró en tres fases principales:

### ***Fase 1: Preparación del instrumento y prueba piloto***

En esta fase se diseñó el instrumento de recolección de información mediante un formulario estructurado en Google Forms, orientado a recopilar información socioeconómica, laboral y contextual de comerciantes informales del departamento de Antioquia.

### **Figura 1**

*Formulario de encuesta en Google Forms*



**Encuesta sobre la realidad de los comerciantes informales en Antioquia**

Este cuestionario tiene como propósito conocer la situación laboral, social y económica de los comerciantes informales en Antioquia. La información recolectada será utilizada únicamente con fines académicos en el marco de un proyecto de investigación universitaria.

Sus respuestas son anónimas y confidenciales. Agradecemos su colaboración y el tiempo dedicado.

\* Indica que la pregunta es obligatoria

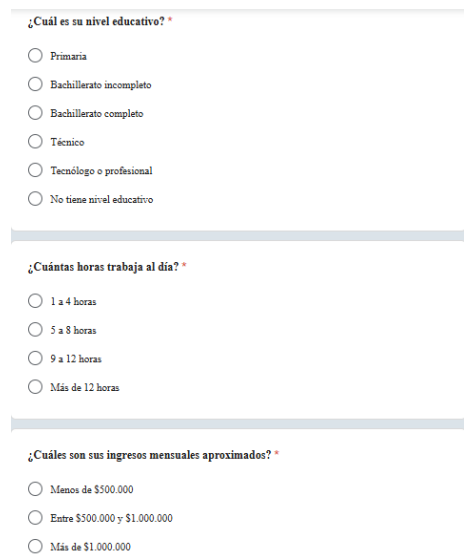
¿Qué edad tiene? \*

10 a 18 años

El cuestionario incluyó preguntas relacionadas con nivel educativo, ingresos, experiencia laboral, condiciones de trabajo, barreras percibidas frente a la formalización y antecedentes relacionados con procesos de formalización.

## Figura 2

### Ejemplo de preguntas incluidas en el cuestionario



¿Cuál es su nivel educativo? \*

- Primaria
- Bachillerato incompleto
- Bachillerato completo
- Técnico
- Tecnólogo o profesional
- No tiene nivel educativo

¿Cuántas horas trabaja al día? \*

- 1 a 4 horas
- 5 a 8 horas
- 9 a 12 horas
- Más de 12 horas

¿Cuáles son sus ingresos mensuales aproximados? \*

- Menos de \$500.000
- Entre \$500.000 y \$1.000.000
- Más de \$1.000.000

Asimismo, se definieron las variables que posteriormente serían utilizadas en el desarrollo del modelo predictivo basado en regresión logística supervisada. Este proceso incluyó la revisión de respuestas abiertas, la agrupación de categorías equivalentes y la eliminación de elementos aclaratorios presentes en algunas opciones de respuesta.

Como parte del proceso metodológico, se realizó una prueba piloto aplicada a nueve comerciantes informales ubicados en Medellín y Doradal, con el propósito de evaluar la claridad, comprensión y pertinencia de las preguntas incluidas en el cuestionario. A partir de esta aplicación preliminar se realizaron ajustes en la redacción de algunas preguntas y se incorporó una opción de respuesta abierta en la pregunta relacionada con las razones de no formalización, con el fin de mejorar la calidad de la información recolectada.

**Figura 3**

*Aplicación de la prueba piloto a comerciantes informales*

**Fase 2: Recolección de datos**

Posteriormente, se realizó la aplicación de encuestas a comerciantes informales ubicados en diferentes zonas del departamento de Antioquia, incluyendo Medellín, Copacabana, Doradal y Perales. La recolección de información se desarrolló de manera presencial mediante dispositivos móviles utilizando Google Forms, lo que permitió registrar la información directamente en formato digital.

**Figura 4**

*Proceso de recolección de información en campo*



Los participantes fueron seleccionados mediante un muestreo no probabilístico por conveniencia, debido a la accesibilidad y disponibilidad de los comerciantes informales durante

el trabajo de campo. Como criterio de inclusión, se consideró únicamente a personas que ejercieran actividades comerciales informales al momento de la aplicación de la encuesta.

En total se recopilaron 108 encuestas durante el período comprendido entre febrero y abril, las cuales constituyeron la base de datos utilizada para el análisis estadístico y el desarrollo del modelo predictivo.

### Figura 5

#### Base de datos original de encuestas

1	¿Cuál edad tiene?	¿Cuántos años de experiencia tiene como tr...	¿Cuál es su nivel educativo?	¿Cuántas horas trabaja al día?	¿Cuáles son sus ingresos mensuales aprox...	¿Tiene una condición física que lo limite par...	Antes de iniciar esta actividad, ¿Qué hac...
2	19 a 25 años	2 a 10 años	Bachillerato incompleto	9 a 12 horas	Entre \$500.000 y \$1.000.000	No	Empleado
3	41 a 80 años o más	2 a 10 años	Primaria	9 a 12 horas	Menos de \$500.000	No	Ninguna de las anteriores
4	41 a 80 años o más	6 meses a 1 año	Tecnólogo o profesional	9 a 12 horas	Entre \$500.000 y \$1.000.000	No	Dependiente (ayudaba en negocio familiar c
5	19 a 25 años	6 meses a 1 año	Bachillerato completo	9 a 12 horas	Entre \$500.000 y \$1.000.000	No	Estudiante
6	26 a 40 años	Más de 15 años	Primaria	9 a 12 horas	Menos de \$500.000	No	Estudiante
7	41 a 80 años o más	Más de 15 años	Bachillerato incompleto	5 a 8 horas	Entre \$500.000 y \$1.000.000	No	Estudiante
8	26 a 40 años	6 meses a 1 año	Bachillerato incompleto	9 a 12 horas	Menos de \$500.000	No	Estudiante
9	26 a 40 años	11 a 15 años	Técnico	5 a 8 horas	Entre \$500.000 y \$1.000.000	No	Empleado
10	26 a 40 años	2 a 10 años	Bachillerato incompleto	5 a 8 horas	Entre \$500.000 y \$1.000.000	No	Empleado
11	41 a 80 años o más	Más de 15 años	Primaria	5 a 8 horas	Entre \$500.000 y \$1.000.000	No	Empleado
12	41 a 80 años o más	Más de 15 años	Técnico	5 a 8 horas	Menos de \$500.000	No	Dependiente (ayudaba en negocio familiar c
13	26 a 40 años	Más de 15 años	Técnico	9 a 12 horas	Más de \$1.000.000	No	Estudiante
14	41 a 80 años o más	11 a 15 años	Bachillerato incompleto	Más de 12 horas	Más de \$1.000.000	Sí	Dependiente (ayudaba en negocio familiar c
15	41 a 80 años o más	11 a 15 años	Tecnólogo o profesional	5 a 8 horas	Entre \$500.000 y \$1.000.000	No	Dependiente (ayudaba en negocio familiar c
16	41 a 80 años o más	6 meses a 1 año	No tiene nivel educativo	1 a 4 horas	Menos de \$500.000	Sí	Dependiente (ayudaba en negocio familiar c
17	41 a 80 años o más	Más de 15 años	Bachillerato incompleto	5 a 8 horas	Más de \$1.000.000	No	Empleado
18	41 a 80 años o más	11 a 15 años	Técnico	5 a 8 horas	Más de \$1.000.000	No	Dependiente (ayudaba en negocio familiar c

De manera complementaria, se realizó una revisión documental basada en literatura académica, informes institucionales y fuentes secundarias relacionadas con la informalidad comercial y los procesos de formalización en Colombia, con el propósito de contextualizar los hallazgos obtenidos mediante el análisis cuantitativo.

### Fase 3: Análisis de datos y elaboración de informe

En esta fase se realizó el procesamiento, depuración y análisis de la información recolectada mediante las encuestas aplicadas a comerciantes informales. Inicialmente, la base de datos obtenida a través de Google Forms fue exportada a formato Excel para facilitar su tratamiento y organización mediante herramientas de análisis de datos en Python.

Posteriormente, se llevó a cabo un proceso de limpieza y depuración de la información, eliminando inconsistencias y organizando las variables necesarias para el desarrollo del modelo predictivo.

## Figura 6

### Base de datos depurada para análisis

	B	C	D	E	F	
1	Form_Responses					
1	¿Qué edad tiene?	¿Cuántos años de experiencia tiene como trx	¿Cuál es su nivel educativo?	¿Cuántas horas trabaja al día?	¿Cuáles son sus ingresos mensuales aproxir	¿Tiene una condi
2	19 a 25 años	2 a 10 años	Bachillerato incompleto	9 a 12 horas	Entre \$500.000 y \$1.000.000	No
3	41 a 80 años o más	2 a 10 años	Primaria	9 a 12 horas	Menos de \$500.000	No
4	41 a 80 años o más	6 meses a 1 año	Tecnólogo o profesional	9 a 12 horas	Entre \$500.000 y \$1.000.000	No
5	19 a 25 años	6 meses a 1 año	Bachillerato completo	9 a 12 horas	Entre \$500.000 y \$1.000.000	No
6	26 a 40 años	Más de 15 años	Primaria	9 a 12 horas	Menos de \$500.000	No
7	41 a 80 años o más	Más de 15 años	Bachillerato incompleto	5 a 8 horas	Entre \$500.000 y \$1.000.000	No
8	26 a 40 años	6 meses a 1 año	Bachillerato incompleto	9 a 12 horas	Menos de \$500.000	No
9	26 a 40 años	11 a 15 años	Técnico	5 a 8 horas	Entre \$500.000 y \$1.000.000	No
10	26 a 40 años	2 a 10 años	Bachillerato incompleto	5 a 8 horas	Entre \$500.000 y \$1.000.000	No
11	41 a 80 años o más	Más de 15 años	Primaria	5 a 8 horas	Entre \$500.000 y \$1.000.000	No
12	41 a 80 años o más	Más de 15 años	Técnico	5 a 8 horas	Menos de \$500.000	No
13	26 a 40 años	Más de 15 años	Técnico	9 a 12 horas	Más de \$1.000.000	No

Una vez depurada la base de datos, se realizó la codificación de las variables categóricas con el propósito de transformar las respuestas cualitativas en valores numéricos compatibles con el modelo de regresión logística supervisada. Las variables binarias fueron codificadas mediante valores 1 y 0, mientras que variables ordinales y categóricas relacionadas con nivel educativo, ingresos, experiencia laboral, jornadas de trabajo y razones de no formalización fueron representadas mediante categorías numéricas.

Este proceso permitió estructurar la base de datos final utilizada para alimentar el modelo de machine learning.

## Figura 7

### Base de datos codificada para modelado

Educ_cod	Experiencia_cod	Educación_cod	Hours_cod	Ingresos_cod	Condición_cod	Ocupación_cod	Raça_hispan_cod	Alignación_cod	Formado_cod	Intento_binario	Raça_no_formal_cod
2	2	3	3	2	0	2	1	1	1	1	3
4	2	2	3	1	0	4	1	1	3	0	2
4	1	6	3	2	0	3	3	1	3	0	8
2	1	4	3	2	0	1	1	1	1	1	2
3	4	2	3	1	0	1	1	3	1	1	1
4	4	3	2	2	0	1	1	1	1	1	3
3	1	3	3	1	0	1	1	2	3	0	2
3	3	5	2	2	0	2	1	1	1	1	3
3	2	3	2	2	0	2	1	1	1	0	2
4	4	2	2	2	0	2	2	1	1	1	1
4	4	5	2	1	0	3	2	1	3	0	4
3	4	5	3	3	0	1	2	4	3	0	7
4	3	3	4	3	1	3	2	1	1	0	4
4	3	6	2	2	1	3	3	2	3	1	3
.	.	.	.	.	.	.	.	.	.	.	.

Posteriormente, el análisis cuantitativo y el desarrollo del modelo predictivo fueron realizados en Python utilizando librerías especializadas como pandas, sklearn, seaborn y matplotlib. Inicialmente, se seleccionaron las variables independientes relacionadas con factores socioeconómicos y contextuales, mientras que la variable dependiente correspondió a la variable binaria Intento\_binario, utilizada para diferenciar entre comerciantes que habían intentado vincularse al mercado laboral formal y aquellos que no lo habían intentado.

El flujo metodológico del modelo incluyó la carga de datos, selección de variables, separación entre variables predictoras y variable objetivo, división de datos para entrenamiento y prueba mediante train\_test\_split, entrenamiento del modelo de regresión logística supervisada y evaluación de métricas de desempeño.

## Figura 8

### Código del modelo en Python

```

Modelo_formalización
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Comandos + Código + Texto Ejecutar todas
(100, 12)
# variables independientes
X = df.drop('Intento_binario', axis=1)
# variable dependiente
Y = df['Intento_binario']
X.head()
Y.head()
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
# dividir datos (80% entrenamiento, 20% prueba)
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42)
# crear modelo
modelo = LogisticRegression(max_iter=1000)
# entrenar modelo
modelo.fit(X_train, Y_train)
LogisticRegression
LogisticRegression(max_iter=1000)

```

Para el entrenamiento y validación inicial del modelo se utilizó una división de datos correspondiente al 80 % para entrenamiento y 20 % para prueba. Posteriormente, el modelo fue evaluado mediante métricas de desempeño como accuracy, precision y recall, permitiendo analizar la capacidad predictiva del algoritmo frente a la probabilidad de formalización de los comerciantes informales.

Adicionalmente, con el propósito de fortalecer la rigurosidad metodológica del modelo, se implementó una validación cruzada de cinco particiones (5-fold cross validation), obteniendo un accuracy promedio aproximado de 74 %, lo que permitió evaluar la estabilidad del modelo en diferentes subconjuntos de datos.

Asimismo, se aplicó un proceso básico de optimización de hiperparámetros mediante GridSearchCV, probando diferentes configuraciones del parámetro C y distintos métodos de optimización (solver) de la regresión logística. Como resultado, la mejor configuración obtenida correspondió a un valor  $C = 0.01$  y solver = liblinear, alcanzando un accuracy aproximado de 76 %.

De manera complementaria, se realizó un análisis de correlación entre variables mediante herramientas de análisis en Python, así como procesos de visualización interactiva de resultados utilizando Power BI, con el propósito de apoyar la interpretación de patrones asociados a la formalización de comerciantes informales.

## **Métodos de recolección de datos**

### ***Encuestas cuantitativas***

Los datos cuantitativos fueron recolectados mediante una encuesta estructurada aplicada a comerciantes informales del departamento de Antioquia. El instrumento incluyó preguntas relacionadas con variables socioeconómicas, laborales y contextuales, tales como nivel

educativo, ingresos, experiencia laboral, condiciones de trabajo y barreras percibidas frente a la formalización.

La información obtenida permitió construir la base de datos utilizada para el análisis estadístico y el desarrollo del modelo predictivo basado en regresión logística supervisada.

### ***Análisis cualitativo documental***

Como complemento al enfoque cuantitativo, se realizó un análisis cualitativo basado en la revisión de fuentes secundarias, incluyendo literatura académica, informes institucionales y documentos relacionados con la informalidad y los procesos de formalización en Colombia.

Debido a limitaciones de acceso a funcionarios y actores institucionales, no fue posible realizar entrevistas directas. En su lugar, se optó por un análisis documental apoyado en información institucional y reportes oficiales, lo cual permitió identificar los principales obstáculos desde la perspectiva de las entidades involucradas en los procesos de formalización.

Este enfoque permitió reconocer barreras estructurales, limitaciones en la implementación de políticas públicas y factores contextuales que influyen en la formalización, aportando una interpretación más amplia de los resultados obtenidos a partir de la encuesta.

### **Análisis de datos**

#### ***Análisis cuantitativo***

El análisis cuantitativo se desarrolló mediante herramientas estadísticas y técnicas de machine learning implementadas en Python. A partir de la información recolectada mediante las encuestas, se realizó un proceso de depuración, organización y codificación de variables con el fin de estructurar una base de datos compatible con las herramientas de análisis utilizadas.

Posteriormente, se seleccionaron las variables socioeconómicas, laborales y contextuales consideradas relevantes para el estudio y se definió una variable objetivo de carácter binario asociada al acercamiento de los comerciantes al mercado laboral formal. Con base en esta

estructura, se implementó un modelo de regresión logística supervisada orientado al análisis de patrones relacionados con la formalización.

Para el entrenamiento y evaluación del modelo, la base de datos fue dividida en conjuntos de entrenamiento y prueba. Adicionalmente, se utilizaron métricas de desempeño como Accuracy, Precision y Recall para evaluar el comportamiento predictivo del modelo. Con el propósito de fortalecer la consistencia metodológica del análisis, también se aplicaron procedimientos de validación cruzada y optimización básica de hiperparámetros mediante GridSearchCV.

De manera complementaria, se realizó un análisis de correlación entre las variables codificadas con el fin de identificar posibles relaciones entre los factores estudiados y la variable objetivo utilizada en el modelo predictivo.

### **Análisis cualitativo**

El análisis cualitativo se desarrolló mediante la revisión e interpretación de literatura académica, informes institucionales y documentos relacionados con la informalidad y los procesos de formalización en Colombia. Esta revisión permitió contextualizar los hallazgos cuantitativos e identificar barreras institucionales, económicas y sociales asociadas a la formalización.

Asimismo, el análisis documental facilitó la identificación de factores relevantes para la construcción del instrumento de recolección de información y contribuyó a la interpretación de los resultados obtenidos durante el desarrollo de la investigación.

### **Consideraciones éticas**

El desarrollo de la investigación se realizó teniendo en cuenta principios éticos fundamentales para el manejo de información con personas. A los participantes se les explicó el

propósito del estudio y el uso de los datos recolectados, garantizando que su participación fuera voluntaria.

Asimismo, se aseguró la confidencialidad y el anonimato de la información suministrada, evitando el uso de datos que permitieran identificar a los encuestados. La información recolectada fue utilizada exclusivamente con fines académicos, respetando en todo momento la integridad y privacidad de los participantes.

### **Limitaciones**

Durante el desarrollo de la investigación se identificaron algunas condiciones que pueden influir en el alcance de los resultados. En primer lugar, la información recolectada se basa en las respuestas proporcionadas directamente por los participantes, lo que implica que los resultados dependen de su percepción y experiencia individual frente a la formalización.

Asimismo, aunque se logró recolectar un total de 108 encuestas, la muestra no abarca la totalidad de las realidades del comercio informal en el departamento de Antioquia, por lo que los resultados deben interpretarse dentro del contexto específico de la población analizada.

Adicionalmente, se presentaron limitaciones en el acceso a fuentes primarias institucionales, lo que llevó a complementar el análisis cualitativo mediante la revisión de fuentes secundarias, permitiendo identificar factores relevantes desde una perspectiva institucional.

## Resultados

### **Identificación de variables asociadas a la probabilidad de formalización de comerciantes informales:**

#### *Caracterización de los comerciantes informales encuestados*

A partir de las 108 encuestas aplicadas a comerciantes informales del departamento de Antioquia, se realizó una caracterización de variables socioeconómicas, laborales y contextuales relacionadas con la informalidad y la probabilidad de formalización.

Los resultados evidenciaron que el grupo predominante correspondió a comerciantes con edades entre 41 y 80 años o más (38,9%), seguido por aquellos con edades entre 26 y 40 años (33,3%). En cuanto a la experiencia laboral, la categoría con mayor participación correspondió a comerciantes con entre 2 y 10 años desempeñándose en actividades informales (38,9%), seguida por quienes reportaron entre 11 y 15 años de experiencia (24,1%).

Respecto al nivel educativo, se observó una mayor concentración en las categorías de bachillerato incompleto (29,6%) y primaria (28,7%), mientras que únicamente una proporción menor manifestó contar con formación técnica, tecnológica o profesional. En relación con los ingresos mensuales, el 45,4% de los encuestados reportó ingresos entre \$500.000 y \$1.000.000, mientras que el 34,3% indicó percibir menos de \$500.000 mensuales.

En cuanto a las condiciones laborales, el 49,1% manifestó trabajar entre 5 y 8 horas diarias y el 67,6% desarrolla su actividad económica en jornada diurna. Asimismo, el 86,1% señaló no presentar ninguna condición física que limite su acceso al mercado laboral formal.

Los resultados también evidenciaron que la principal razón para iniciar una actividad informal fue la necesidad económica (54,6%), seguida por el deseo de independencia (22,2%).

De igual manera, el 55,6% de los comerciantes indicó haber iniciado su negocio por cuenta propia, mientras que el 22,2% manifestó haberlo adquirido mediante compra.

Respecto a la vinculación al mercado laboral formal, el 54,6% de los participantes afirmó no haber intentado acceder a este tipo de empleo, mientras que el 45,4% indicó haberlo intentado sin éxito. Por otra parte, la principal razón reportada para no formalizarse fue la falta de información (29,6%), seguida por los altos costos asociados a los procesos de formalización (25,0%) y la complejidad de los trámites administrativos (18,5%).

Con el fin de resumir los principales resultados obtenidos durante la caracterización de los comerciantes informales encuestados, en la Tabla 1 se presentan las categorías predominantes identificadas para cada variable analizada.

**Tabla 1**

*Caracterización general de los comerciantes informales encuestados*

<b>Variable</b>	<b>Categoría predominante</b>	<b>Porcentaje</b>
<b>Edad</b>	41 a 80 años o más	38,9%
<b>Experiencia laboral</b>	2 a 10 años	38,9%
<b>Nivel educativo</b>	Bachillerato incompleto	29,6%
<b>Horas de trabajo</b>	5 a 8 horas	49,1%
<b>Ingresos mensuales</b>	Entre \$500.000 y \$1.000.000	45,4%
<b>Condición física limitante</b>	No	86,1%
<b>Ocupación previa</b>	Empleado	39,8%

<b>Razón de inicio</b>	Necesidad económica	54,6%
<b>Adquisición del negocio</b>	Lo inició usted mismo	55,6%
<b>Jornada laboral</b>	Diurna	67,6%
<b>Intento de vinculación formal</b>	No lo ha intentado	54,6%
<b>Principal razón para no formalizarse</b>	Falta de información	29,6%

Los resultados obtenidos permitieron identificar variables socioeconómicas, laborales y contextuales relevantes para el estudio, entre las que se destacan edad, experiencia laboral, nivel educativo, ingresos, jornada de trabajo, antecedentes laborales y barreras percibidas frente a la formalización. La caracterización realizada evidenció que factores como las condiciones económicas, el nivel educativo y las percepciones frente a los procesos de formalización constituyen elementos de interés para comprender el comportamiento de los comerciantes informales analizados. Estas variables conformaron posteriormente la base para el procesamiento de la información y el desarrollo del modelo predictivo.

### **Procesamiento, depuración y codificación de la información recolectada:**

#### ***Recolección de la información***

La información utilizada para el desarrollo del modelo predictivo se obtuvo mediante la aplicación de 108 encuestas a comerciantes informales del departamento de Antioquia. Para la construcción del instrumento de recolección se realizó previamente una revisión documental orientada a identificar factores socioeconómicos, laborales y contextuales asociados a la informalidad y a la probabilidad de formalización. A partir de este proceso se diseñó una encuesta compuesta por doce preguntas relacionadas con variables como edad, experiencia

laboral, nivel educativo, ingresos, jornada de trabajo, antecedentes laborales y barreras percibidas frente a la formalización.

Las encuestas fueron implementadas mediante Google Forms y aplicadas directamente a comerciantes informales en diferentes municipios y zonas del departamento de Antioquia. La recolección de información fue realizada por los integrantes del proyecto, obteniéndose un total de 108 registros válidos que posteriormente fueron utilizados para la construcción de la base de datos del estudio.

### ***Tabulación y depuración de la información***

Una vez finalizada la etapa de recolección, las respuestas almacenadas en Google Forms fueron exportadas a una hoja de cálculo. La estructura inicial de la base de datos obtenida a partir de las respuestas del formulario se presenta en la Figura 5, ubicada en la fase 2 de la sección de metodológica. Con el fin de preservar la información original obtenida durante el trabajo de campo, se generó una copia de la base de datos para realizar sobre ella el proceso de preparación de la información.

Durante la depuración se revisó la consistencia general de los registros y se realizaron ajustes orientados a mejorar la calidad de los datos antes de su utilización en el modelo predictivo. Entre las acciones realizadas se incluyó la eliminación de textos aclaratorios contenidos entre paréntesis dentro de algunas preguntas y respuestas, ya que estos elementos podían generar inconsistencias durante el procesamiento posterior de la información.

De igual forma, se revisaron las respuestas abiertas registradas en la pregunta relacionada con las razones para no formalizarse. Aunque la mayoría de los participantes seleccionó alguna de las opciones previamente establecidas, algunos comerciantes registraron respuestas abiertas asociadas a factores como edad, nivel educativo, responsabilidades familiares u otras situaciones

particulares. Con el propósito de facilitar el análisis posterior, estas respuestas fueron agrupadas en categorías definidas e incorporadas a la estructura general de la base de datos.

Como resultado de esta etapa se obtuvo una base de datos depurada, organizada y lista para iniciar el proceso de codificación de variables.

### ***Codificación de variables***

Posteriormente se realizó la codificación de las variables recolectadas. Este proceso tuvo como finalidad transformar las respuestas textuales obtenidas en la encuesta a valores numéricos que pudieran ser interpretados por las herramientas de análisis estadístico y machine learning implementadas en Python.

Para ello se consideraron variables ordinales, nominales y binarias. Las variables ordinales fueron codificadas respetando el orden natural existente entre las categorías de respuesta. Por ejemplo, en variables como edad, experiencia laboral, nivel educativo, horas de trabajo e ingresos, los valores numéricos asignados siguieron la secuencia lógica de menor a mayor.

Las variables binarias fueron transformadas utilizando valores 0 y 1, permitiendo representar respuestas dicotómicas como sí o no. Por su parte, las variables nominales fueron convertidas en categorías numéricas con el propósito de facilitar su incorporación al modelo predictivo.

Este procedimiento permitió estandarizar la información recolectada y garantizar la compatibilidad de la base de datos con las técnicas de análisis empleadas durante el desarrollo de la investigación.

**Tabla 2***Codificación de variables utilizadas en el modelo predictivo*

<b>Variable</b>	<b>Tipo de variable</b>	<b>Codificación aplicada</b>
<b>Edad</b>	Ordinal	10 a 18 años = 1; 19 a 25 años = 2; 26 a 40 años = 3; 41 a 80 años o más = 4
<b>Experiencia laboral</b>	Ordinal	6 meses a 1 año = 1; 2 a 10 años = 2; 11 a 15 años = 3; Más de 15 años = 4
<b>Nivel educativo</b>	Ordinal	No tiene nivel educativo = 1; Primaria = 2; Bachillerato incompleto = 3; Bachillerato completo = 4; Técnico = 5; Tecnólogo o profesional = 6
<b>Horas de trabajo</b>	Ordinal	1 a 4 horas = 1; 5 a 8 horas = 2; 9 a 12 horas = 3; Más de 12 horas = 4
<b>Ingresos mensuales</b>	Ordinal	Menos de \$500.000 = 1; Entre \$500.000 y \$1.000.000 = 2; Más de \$1.000.000 = 3
<b>Condición física para acceder al empleo formal</b>	Binaria	Sí = 1; No = 0
<b>Ocupación previa</b>	Nominal	Estudiante = 1; Empleado = 2; Dependiente = 3; Ninguna de las anteriores = 4
<b>Razón de inicio de la actividad informal</b>	Nominal	Necesidad económica = 1; Deseo de independencia = 2; Tradición familiar = 3; Oportunidad casual = 4

<b>Forma de adquisición del negocio</b>	Nominal	Lo inició usted mismo = 1; Lo arrienda = 2; Lo compró = 3; Se lo regalaron = 4; Otra = 5
<b>Jornada laboral</b>	Nominal	Diurna = 1; Nocturna = 2; Mixta = 3
<b>Razón para no formalizarse</b>	Nominal	Trámites muy complicados = 1; Altos costos = 2; Falta de información = 3; Prefiere seguir en la informalidad = 4; Nivel educativo = 5; Edad = 6; Responsabilidades familiares = 7; Otra = 8
<b>Intento de vinculación al mercado laboral formal (Intento_binario)</b>	Binaria	Sí, pero no lo logró = 1; No lo ha intentado = 0

La codificación permitió transformar la información cualitativa recolectada mediante las encuestas en variables numéricas susceptibles de ser procesadas por herramientas de análisis estadístico y aprendizaje automático.

### ***Estructuración de la base de datos para el modelo predictivo***

Como resultado de las etapas de recolección, depuración y codificación, se construyó una base de datos estructurada compuesta por variables relacionadas con características socioeconómicas, laborales y contextuales de los comerciantes informales encuestados.

La base final incluyó variables correspondientes a edad, experiencia laboral, nivel educativo, horas de trabajo, ingresos mensuales, condición física para acceder a empleo formal, ocupación previa, razón de inicio de la actividad informal, forma de adquisición del negocio,

jornada laboral, razones para no formalizarse e intención de vinculación al mercado laboral formal.

Esta estructura permitió disponer de información organizada y preparada para el entrenamiento, evaluación y validación del modelo predictivo desarrollado en la siguiente fase de la investigación.

**Tabla 3**

*Variables finales utilizadas en el modelo predictivo*

<b>Variables incluidas en la base de datos final</b>
<b>Edad_cod</b>
<b>Experiencia_cod</b>
<b>Educación_cod</b>
<b>Horas_cod</b>
<b>Ingresos_cod</b>
<b>Condición_cod</b>
<b>Ocupación_cod</b>
<b>Razón_inicio_cod</b>
<b>Adquisición_cod</b>
<b>Jornada_cod</b>
<b>Razon_no_formal_cod</b>
<b>Intento_binario</b>

Como resultado de este proceso se obtuvo una base de datos depurada, codificada y estructurada, compuesta por doce variables que posteriormente fueron utilizadas para el entrenamiento, evaluación y validación del modelo predictivo desarrollado mediante regresión logística supervisada.

### **Construcción y evaluación del modelo predictivo:**

#### ***Selección de variables y definición de la variable objetivo***

Una vez finalizado el proceso de depuración y codificación de la información recolectada, se procedió a seleccionar las variables que serían utilizadas para el desarrollo del modelo predictivo. Para ello se empleó la base de datos estructurada obtenida en la etapa anterior, compuesta por variables socioeconómicas, laborales y contextuales asociadas a los comerciantes informales encuestados.

Las variables incluidas en el modelo fueron: Edad\_cod, Experiencia\_cod, Educación\_cod, Horas\_cod, Ingresos\_cod, Condición\_cod, Ocupación\_cod, Razón\_inicio\_cod, Adquisición\_cod, Jornada\_cod, Razon\_no\_formal\_cod e Intento\_binario.

Posteriormente, las variables fueron organizadas en dos grupos: variables independientes (X) y variable dependiente (Y). Las variables independientes correspondieron a las características socioeconómicas, laborales y contextuales de los comerciantes, mientras que la variable dependiente fue definida como Intento\_binario.

La variable Intento\_binario fue seleccionada como variable objetivo debido a que representa el comportamiento que se buscaba analizar dentro de la investigación. Esta variable permite diferenciar entre los comerciantes que han intentado vincularse al mercado laboral formal y aquellos que no lo han intentado, constituyéndose en una variable que permite identificar distintos comportamientos frente al acercamiento al mercado laboral formal.

Adicionalmente, esta variable presenta únicamente dos posibles resultados (1 = Sí, pero no lo logró; 0 = No lo ha intentado), por lo que fue tratada como una variable binaria, condición que la hace adecuada para la aplicación de técnicas de clasificación como la regresión logística utilizada en el presente estudio.

Cabe recalcar que la separación entre variables independientes y variable dependiente permitió establecer los factores utilizados para analizar el comportamiento de la variable objetivo y constituyó la base para el entrenamiento y evaluación posterior del modelo predictivo.

### ***División de los datos en entrenamiento y prueba***

Una vez definidas las variables independientes y la variable objetivo, se realizó la división de la base de datos en dos subconjuntos con el propósito de entrenar y posteriormente evaluar el desempeño del modelo predictivo.

Para ello, se utilizó el 80% de los registros para el entrenamiento del modelo y el 20% restante para la etapa de prueba. Esta distribución permitió que el algoritmo aprendiera patrones a partir de una parte de la información disponible y posteriormente fuera evaluado utilizando datos que no habían sido utilizados durante el proceso de entrenamiento.

La separación entre entrenamiento y prueba se realizó mediante la función `train_test_split` de la librería Scikit-Learn, utilizando un valor de `test_size = 0.2`, equivalente al 20% de los datos para prueba, y un valor de `random_state = 42`, con el fin de garantizar que la división realizada pudiera reproducirse de manera consistente en futuras ejecuciones del modelo.

Este procedimiento permitió verificar que los resultados obtenidos no dependieran únicamente de los datos utilizados durante el entrenamiento, sino también de la capacidad del modelo para realizar predicciones sobre información que no había observado previamente.

## Figura 9

*División de los datos en conjuntos de entrenamiento y prueba mediante la función*

*train\_test\_split en Python*

```
[11]
✓ 0s
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression

# Dividir datos (80% entrenamiento, 20% prueba)
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42)
```

### *Entrenamiento del modelo mediante regresión logística*

Una vez realizada la división de los datos en conjuntos de entrenamiento y prueba, se procedió al entrenamiento del modelo predictivo utilizando la técnica de regresión logística supervisada.

La selección de esta técnica se fundamentó en las características de la variable objetivo `Intento_binario`, la cual presenta únicamente dos posibles resultados: comerciantes que han intentado vincularse al mercado laboral formal y comerciantes que no lo han intentado. Debido a esta naturaleza binaria, la regresión logística constituye una técnica adecuada para realizar procesos de clasificación y predicción de este tipo de comportamientos.

Para el entrenamiento del modelo se utilizó la implementación disponible en la librería Scikit-Learn, empleando el parámetro `max_iter = 1000`. Este parámetro establece el número máximo de iteraciones que puede realizar el algoritmo durante el proceso de aprendizaje y se utilizó con el fin de garantizar que el modelo dispusiera de suficientes ciclos de cálculo para identificar patrones dentro de los datos y alcanzar una solución estable durante el entrenamiento.

El modelo fue entrenado utilizando las variables socioeconómicas, laborales y contextuales previamente seleccionadas, permitiendo establecer relaciones entre dichas variables

y el comportamiento de la variable objetivo. Como resultado de esta etapa se obtuvo un modelo entrenado y preparado para su posterior evaluación mediante métricas de desempeño.

### *Evaluación inicial del modelo*

Una vez completado el proceso de entrenamiento, se procedió a evaluar el desempeño inicial del modelo predictivo mediante las métricas Accuracy, Precision y Recall. Estas métricas fueron seleccionadas debido a que permiten medir el nivel de acierto del modelo y evaluar la confiabilidad de las predicciones realizadas.

### **Figura 10**

*Resultados iniciales de evaluación del modelo predictivo mediante Accuracy, Precision y Recall*

```
[12]
✓ 0 s
from sklearn.metrics import accuracy_score, precision_score, recall_score

# Predicciones
Y_pred = modelo.predict(X_test)

# Métricas
accuracy = accuracy_score(Y_test, Y_pred)
precision = precision_score(Y_test, Y_pred)
recall = recall_score(Y_test, Y_pred)

print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)

Accuracy: 0.6363636363636364
Precision: 0.5
Recall: 0.625
```

De acuerdo con los resultados arrojados podemos analizar que: el Accuracy obtenido fue de aproximadamente 63,6%, lo que indica que el modelo clasificó correctamente cerca de 64 de cada 100 casos evaluados. Esta métrica permite conocer el porcentaje general de aciertos alcanzado por el modelo durante la fase de prueba.

Por su parte, la Precision alcanzó un valor de 50%, lo que significa que cuando el modelo predijo que un comerciante pertenecía a la categoría positiva de la variable objetivo, acertó

aproximadamente en la mitad de los casos. Esta métrica permite evaluar qué tan confiables son las predicciones positivas realizadas por el modelo.

En cuanto al Recall, se obtuvo un valor aproximado de 62,5%, indicando que el modelo logró identificar correctamente cerca de 63 de cada 100 comerciantes pertenecientes a la categoría positiva existente en la muestra analizada. Esta métrica permite evaluar la capacidad del modelo para detectar correctamente los casos positivos reales.

Los resultados obtenidos evidenciaron que el modelo presentaba una capacidad inicial para identificar patrones asociados a la variable objetivo. Sin embargo, con el propósito de obtener una evaluación más robusta y mejorar el desempeño predictivo, posteriormente se implementaron técnicas adicionales de validación cruzada y optimización de hiperparámetros.

### ***Validación cruzada del modelo***

Con el propósito de obtener una evaluación más robusta del desempeño del modelo predictivo, se implementó una técnica de validación cruzada, atendiendo la observación realizada durante el proceso de revisión del proyecto.

La validación cruzada consiste en dividir la base de datos en varios subconjuntos y realizar múltiples entrenamientos y evaluaciones del modelo utilizando diferentes combinaciones de datos. De esta manera, se evita depender de una única división de entrenamiento y prueba, permitiendo obtener una estimación más estable del comportamiento del modelo.

En el presente estudio se utilizó una validación cruzada de cinco particiones ( $cv = 5$ ), lo que significa que la base de datos fue dividida en cinco grupos diferentes. El modelo fue entrenado y evaluado cinco veces, utilizando en cada ejecución una combinación distinta de datos para prueba y entrenamiento.

## Figura 11

### Resultados de la validación cruzada aplicada al modelo predictivo

```
[13] ✓ 0s
import pandas as pd

# Cargar base de datos desde enlace público
url = "https://docs.google.com/spreadsheets/d/11jVPrek50-5Tq0C9m6FmLEtlnXg86/export?format=xlsx"

df = pd.read_excel(url)

# Seleccionar variables
df = df[['Edad_cod', 'Experiencia_cod', 'Educación_cod', 'Horas_cod',
        'Ingresos_cod', 'Condición_cod', 'Ocupación_cod',
        'Razón_inicio_cod', 'Adquisición_cod', 'Jornada_cod',
        'Razon_no_formal_cod', 'Intento_binario']]

# Variables independientes
X = df.drop('Intento_binario', axis=1)

# Variable dependiente
Y = df['Intento_binario']

# Importar librerías
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.linear_model import LogisticRegression

# Dividir datos
X_train, X_test, Y_train, Y_test = train_test_split(
    X, Y, test_size=0.2, random_state=42
)

# Crear modelo
modelo = LogisticRegression(max_iter=1000)

# Entrenar modelo
modelo.fit(X_train, Y_train)

# Validación cruzada
scores = cross_val_score(modelo, X, Y, cv=5, scoring='accuracy')

# Mostrar resultados
print("Accuracy por fold:", scores)
print("Accuracy promedio:", scores.mean())

Accuracy por fold: [0.77272727 0.81818182 0.68181818 0.88952381 0.61904762]
Accuracy promedio: 0.7482597402597402
```

Los resultados obtenidos fueron los siguientes: el accuracy promedio obtenido fue de aproximadamente 74,03%, resultado superior al accuracy inicial obtenido durante la evaluación básica del modelo. Esto indica que, al ser evaluado sobre diferentes subconjuntos de datos, el modelo mantuvo un comportamiento relativamente consistente y una capacidad aceptable para identificar patrones asociados a la variable objetivo.

Los resultados obtenidos sugieren que el modelo presenta un nivel razonable de estabilidad frente a cambios en la muestra utilizada para el entrenamiento y la evaluación, aportando una mayor confianza sobre su capacidad de generalización.

### ***Optimización de hiperparámetros mediante GridSearchCV***

Con el propósito de mejorar el desempeño del modelo predictivo y atendiendo la observación realizada durante el proceso de revisión del proyecto, se implementó una optimización de hiperparámetros mediante la técnica GridSearchCV.

Los hiperparámetros corresponden a configuraciones del modelo que pueden influir en su comportamiento y rendimiento. En este caso, se utilizó GridSearchCV para probar automáticamente diferentes combinaciones de parámetros y determinar cuál ofrecía los mejores resultados para el modelo de regresión logística.

Esta técnica permitió evaluar múltiples configuraciones de manera sistemática, seleccionando aquella que presentó el mayor desempeño de acuerdo con la métrica de accuracy.

## Figura 12

### *Optimización de hiperparámetros mediante GridSearchCV*

```
[14]
✓ Os
from sklearn.model_selection import GridSearchCV

# Parámetros a probar
parametros = {
    'C': [0.01, 0.1, 1, 10],
    'solver': ['liblinear', 'lbfgs']
}

# Configurar búsqueda
grid = GridSearchCV(
    LogisticRegression(max_iter=1000),
    parametros,
    cv=5,
    scoring='accuracy'
)

# Entrenar búsqueda
grid.fit(X_train, Y_train)

# Mostrar mejores resultados
print("Mejores parámetros:", grid.best_params_)
print("Mejor accuracy:", grid.best_score_)

Mejores parámetros: {'C': 0.01, 'solver': 'liblinear'}
Mejor accuracy: 0.7679738562091503
```

Los resultados obtenidos fueron que: la optimización realizada permitió identificar como mejor configuración un valor de  $C = 0,01$  y el algoritmo de resolución liblinear, obteniendo un accuracy aproximado de 76,8%. Esto significa que, con la configuración seleccionada, el modelo logró clasificar correctamente cerca de 77 de cada 100 casos evaluados.

Comparado con los resultados obtenidos en las etapas anteriores, se observó una mejora en el desempeño general del modelo, evidenciando que el ajuste de hiperparámetros contribuyó a

optimizar su capacidad predictiva. Como resultado, se obtuvo una versión más robusta del modelo para el análisis de factores asociados a la formalización de comerciantes informales.

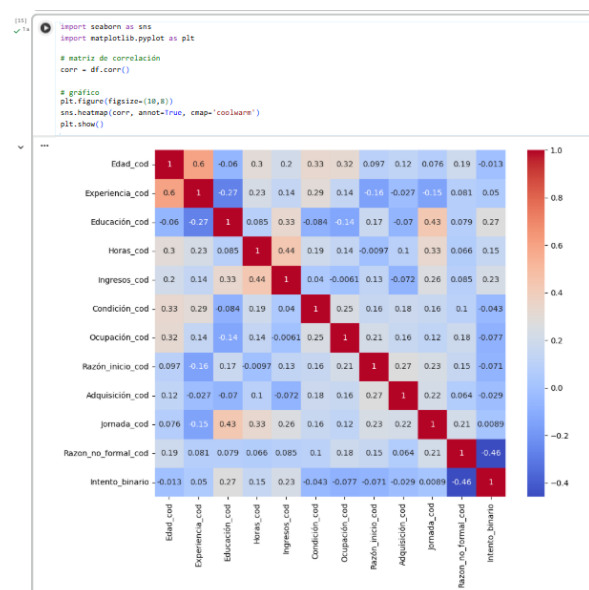
### *Análisis de correlación entre variables*

Como complemento al proceso de evaluación del modelo predictivo, se realizó un análisis de correlación entre las variables incluidas en la investigación. El propósito de este análisis fue identificar posibles relaciones existentes entre las variables socioeconómicas, laborales y contextuales de los comerciantes informales y la variable objetivo utilizada en el modelo predictivo.

La correlación permite medir el grado de relación entre dos variables mediante valores comprendidos entre -1 y 1. Los valores cercanos a 1 indican una relación positiva, los valores cercanos a -1 indican una relación negativa y los valores próximos a 0 sugieren una relación débil o inexistente.

### **Figura 13**

#### *Matriz de correlación entre variables*



Al analizar la relación entre las variables estudiadas y la variable objetivo Intento\_binario, se observó que la variable Razón\_no\_formal\_cod presentó la correlación más alta en valor absoluto, con aproximadamente -0,46. Este resultado sugiere que las razones manifestadas por los comerciantes para no formalizarse mantienen una relación relevante con el comportamiento asociado al acercamiento al mercado laboral formal.

Por otra parte, las variables Educación\_cod e Ingresos\_cod presentaron correlaciones positivas aproximadas de 0,27 y 0,23, respectivamente. Aunque estos valores no representan relaciones fuertes, sí sugieren una posible asociación entre mayores niveles educativos, mejores ingresos y una mayor disposición hacia procesos relacionados con la formalización.

Las demás variables registraron valores cercanos a cero, indicando relaciones débiles dentro de la muestra analizada. Esto significa que variables como edad, experiencia laboral, jornada de trabajo, ocupación previa o forma de adquisición del negocio no evidenciaron una asociación significativa con la variable objetivo en comparación con las anteriormente mencionadas.

Los resultados obtenidos permitieron identificar que la intención de acercamiento al mercado laboral formal no depende de un único factor, sino de la interacción de diferentes condiciones socioeconómicas, laborales y contextuales. Estos hallazgos complementan los resultados obtenidos mediante el modelo predictivo y aportan elementos de análisis para la formulación de estrategias e incentivos orientados a promover la formalización de comerciantes informales.

## **Propuesta de incentivos focalizados para la formalización**

### ***Identificación de factores relevantes para la formulación de incentivos***

A partir de los resultados obtenidos durante la caracterización de los comerciantes informales, el análisis de correlación y el desempeño del modelo predictivo, se identificaron factores socioeconómicos y contextuales asociados a los procesos de formalización.

Los resultados evidenciaron que variables relacionadas con el nivel educativo, los ingresos y las barreras percibidas frente a la formalización presentan una relación relevante con el comportamiento de los comerciantes frente al mercado laboral formal. Asimismo, durante la caracterización de la población se identificaron limitaciones asociadas a la falta de información, los altos costos de formalización y la complejidad de los trámites administrativos, factores que fueron considerados para la construcción de la propuesta de incentivos.

Estos hallazgos permitieron establecer criterios para el diseño de estrategias orientadas a responder a las necesidades identificadas en la población objeto de estudio.

### ***Diseño de la propuesta de incentivos focalizados***

Con base en los resultados obtenidos, se estructuró una propuesta de incentivos focalizados orientada a promover procesos de formalización acordes con las características de los comerciantes informales analizados.

La propuesta fue diseñada considerando tanto los hallazgos derivados del análisis estadístico como las principales barreras identificadas durante la aplicación de las encuestas. De esta manera, se plantearon estrategias orientadas a reducir obstáculos económicos, administrativos y de acceso a información, buscando facilitar procesos de formalización más accesibles y sostenibles.

**Tabla 4***Propuesta de incentivos para la formalización empresarial*

<b>Tipo de incentivo</b>	<b>Estrategia propuesta</b>	<b>Problema que aborda</b>	<b>Impacto esperado</b>
<b>Económico</b>	Esquema de formalización progresiva (reducción de impuestos en 3 años + subsidios a seguridad social)	Altos costos de formalización	Disminución de barreras de entrada, mayor acceso a la formalidad
<b>Administrativo y digital</b>	Ventanilla Única Digital de Formalización (integración de trámites, automatización, asistencia virtual)	Complejidad de trámites y barreras institucionales	Reducción de tiempos, mayor eficiencia y acceso
<b>Educativo</b>	Programas de formación en educación financiera, gestión empresarial y obligaciones tributarias	Bajo nivel educativo y falta de información	Mejora en capacidades, sostenibilidad y toma de decisiones
<b>Acceso a mercados</b>	Participación en ferias, compras públicas, encadenamientos productivos y sello de negocio formal	Bajos ingresos y falta de oportunidades económicas	Incremento de ingresos y competitividad
<b>Focalización</b>	Uso del modelo predictivo para identificar perfiles con mayor probabilidad de formalización	Falta de eficiencia en políticas públicas	Optimización de recursos y mayor impacto de los incentivos

Los incentivos propuestos buscan responder directamente a las principales problemáticas identificadas durante la investigación, promoviendo condiciones más favorables para la incorporación de comerciantes informales a procesos de formalización empresarial.

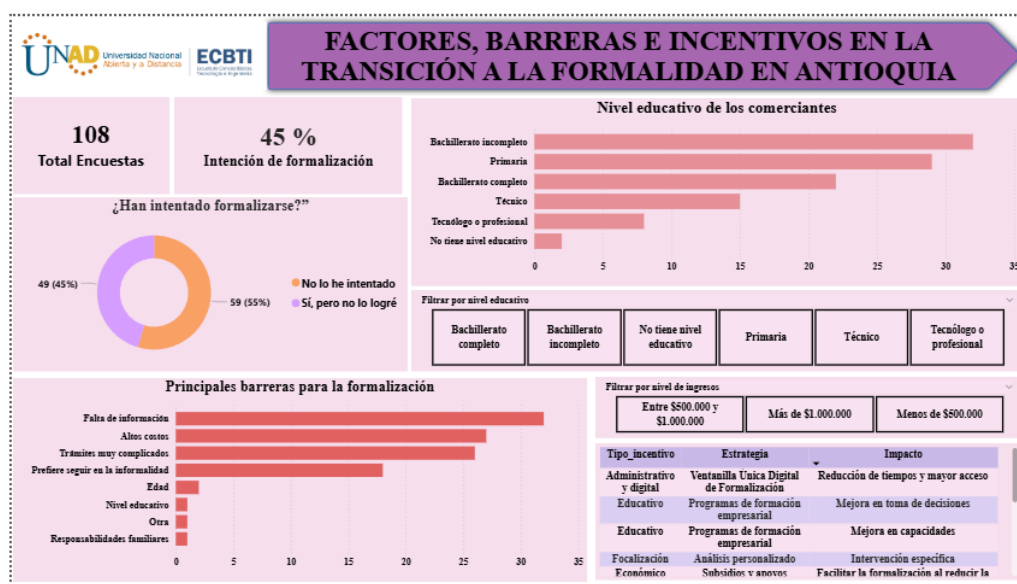
### Visualización e interpretación de resultados mediante Power BI

Como complemento al análisis realizado, se desarrolló una visualización interactiva en Power BI con el propósito de representar gráficamente los principales hallazgos obtenidos durante la investigación.

La herramienta permitió integrar información relacionada con perfiles socioeconómicos, barreras de formalización y posibles estrategias de intervención, facilitando la interpretación de los resultados y la identificación de patrones relevantes para la toma de decisiones.

**Figura 14**

*Visualización interactiva de resultados en Power BI*



La implementación de esta visualización permitió presentar de manera más dinámica los resultados obtenidos y apoyar la formulación de estrategias focalizadas dirigidas a comerciantes informales del departamento de Antioquia.

En conjunto, la propuesta de incentivos y la visualización desarrollada permiten traducir los resultados obtenidos mediante el modelo predictivo y el análisis de datos en acciones

orientadas a promover procesos de formalización más ajustados a las condiciones y necesidades identificadas en la población estudiada.

## Discusión

Los resultados obtenidos en la presente investigación permitieron comprender diferentes factores asociados a los procesos de formalización de comerciantes informales en el departamento de Antioquia. A partir de la caracterización realizada mediante 108 encuestas, fue posible identificar perfiles predominantes dentro de la población estudiada, evidenciando que una proporción importante de los comerciantes encuestados presenta niveles educativos medios o bajos, ingresos limitados y una alta dependencia de la actividad informal como principal fuente de sustento económico.

La caracterización también permitió identificar que la necesidad económica constituye la principal razón para el inicio de actividades informales, mientras que la falta de información, los altos costos y la complejidad de los trámites representan algunas de las principales barreras percibidas frente a la formalización. Estos resultados sugieren que la permanencia en la informalidad no responde únicamente a decisiones individuales, sino que se encuentra influenciada por condiciones económicas, sociales e institucionales presentes en el entorno de los comerciantes.

El análisis de correlación desarrollado como parte del modelo predictivo evidenció que las variables relacionadas con las razones para no formalizarse presentan una de las asociaciones más relevantes con la variable objetivo. Asimismo, variables como el nivel educativo y los ingresos mostraron relaciones positivas moderadas, sugiriendo que mejores condiciones educativas y económicas pueden favorecer una mayor disposición hacia procesos relacionados con la formalización. Aunque estas relaciones no fueron fuertes, permitieron identificar tendencias consistentes dentro de la muestra analizada.

Desde la perspectiva metodológica, la construcción de una base de datos depurada y codificada permitió la implementación de un modelo de regresión logística supervisada orientado al análisis de comportamientos asociados al acercamiento al mercado laboral formal. Los resultados iniciales mostraron un accuracy aproximado del 63,6%, mientras que la validación cruzada alcanzó un promedio cercano al 74% y la optimización mediante GridSearchCV permitió obtener un accuracy aproximado del 76,8%. Estos resultados evidencian que el modelo posee capacidad para identificar patrones dentro de la población estudiada y constituyen una aproximación válida para el análisis de este tipo de fenómenos socioeconómicos.

Los hallazgos obtenidos son consistentes con planteamientos asociados a la Teoría del Capital Humano y la Teoría del Capital Social, las cuales destacan la influencia de factores como la educación, el acceso a recursos y las oportunidades económicas en la toma de decisiones de los individuos. En este sentido, los resultados respaldan la idea de que la formalización constituye un fenómeno multidimensional que involucra factores económicos, sociales e institucionales.

Desde una perspectiva práctica, los resultados obtenidos permitieron estructurar una propuesta de incentivos focalizados orientada a responder a las principales barreras identificadas durante la investigación. La incorporación de estrategias relacionadas con reducción de costos, simplificación administrativa, fortalecimiento de capacidades empresariales y acceso a oportunidades de mercado busca generar condiciones más favorables para la formalización de comerciantes informales.

No obstante, los resultados deben interpretarse considerando las limitaciones propias del estudio. La información utilizada proviene de una muestra de 108 comerciantes informales del departamento de Antioquia y se fundamenta en las respuestas suministradas por los participantes

durante la aplicación de las encuestas. En consecuencia, los hallazgos obtenidos permiten identificar tendencias y relaciones dentro de la población analizada, pero no deben generalizarse de manera absoluta a la totalidad del sector informal.

Finalmente, la investigación evidencia el potencial de las herramientas de machine learning como apoyo para el análisis de problemáticas socioeconómicas complejas y para el diseño de estrategias de intervención más focalizadas. Asimismo, abre la posibilidad de desarrollar investigaciones futuras que incorporen muestras más amplias, nuevas variables de análisis y metodologías complementarias que permitan profundizar en el estudio de los procesos de formalización empresarial.

## Conclusiones y recomendaciones

La investigación permitió identificar variables socioeconómicas, laborales y contextuales asociadas a los procesos de formalización de comerciantes informales en Antioquia. La caracterización realizada evidenció que factores como el nivel educativo, los ingresos, la experiencia laboral y las barreras percibidas frente a la formalización constituyen elementos relevantes para comprender el comportamiento de los comerciantes informales analizados.

La aplicación de 108 encuestas permitió recopilar información suficiente para construir una base de datos estructurada orientada al desarrollo del modelo predictivo. El proceso de depuración, categorización y codificación de las respuestas garantizó la organización de la información y su compatibilidad con las herramientas de análisis utilizadas durante la investigación.

La construcción y evaluación del modelo de regresión logística supervisada permitió identificar patrones asociados al acercamiento de los comerciantes al mercado laboral formal. Los resultados obtenidos mediante Accuracy, Precision y Recall, junto con la validación cruzada y la optimización de hiperparámetros mediante GridSearchCV, evidenciaron que el modelo posee capacidad para analizar comportamientos presentes en la población estudiada y constituye una herramienta de apoyo para la identificación de tendencias relacionadas con la formalización.

El análisis realizado permitió concluir que la informalidad constituye un fenómeno multidimensional en el que intervienen factores económicos, educativos, laborales e institucionales. En particular, la falta de información, los altos costos y la complejidad de los trámites fueron identificados como algunas de las principales barreras percibidas por los comerciantes encuestados.

Con base en los hallazgos obtenidos, fue posible estructurar una propuesta de incentivos focalizados orientada a promover procesos de formalización más acordes con las características y necesidades de la población estudiada. Las estrategias propuestas incluyen medidas económicas, administrativas, educativas y de acceso a mercados, complementadas con herramientas de visualización e interpretación desarrolladas en Power BI.

Como recomendación, se sugiere fortalecer los programas institucionales de capacitación y acompañamiento dirigidos a comerciantes informales, especialmente en aspectos relacionados con educación financiera, gestión empresarial y acceso a información sobre procesos de formalización.

Asimismo, se recomienda promover estrategias orientadas a simplificar trámites administrativos y reducir barreras económicas que dificultan la transición hacia la formalidad, considerando las condiciones particulares identificadas en los comerciantes participantes del estudio.

Para futuras investigaciones, se recomienda ampliar el tamaño de la muestra, incorporar nuevas variables socioeconómicas y evaluar el comportamiento del modelo predictivo en otros contextos territoriales, con el fin de fortalecer la capacidad explicativa y predictiva de este tipo de herramientas analíticas.

Finalmente, se recomienda evaluar en investigaciones posteriores el impacto real de los incentivos propuestos y analizar su efectividad como mecanismo para promover procesos sostenibles de formalización empresarial.

## Referencias

- Abdul-Rahaman, A., & Gambrah, P. (2023). Local trade networks among farmers and traders. *Social Networking*, 12(4), 93–110. <https://doi.org/10.4236/sn.2023.124006>
- Abramo, L. (2022). *Policies to address the challenges of existing and new forms of informality in Latin America*. Comisión Económica para América Latina y el Caribe (CEPAL). <https://www.cepal.org/en/publications/47774-policies-address-challenges-existing-and-new-forms-informality-latin-america>
- Alcaldía de Medellín. (2025, junio 26). *380 empresas y emprendimientos de Medellín recibirán formación para impulsar su crecimiento y competitividad*. <https://www.medellin.gov.co/es/sala-de-prensa/noticias/380-empresas-y-emprendimientos-de-medellin-recibiran-formacion-para-impulsar-su-crecimiento-y-competitividad/>
- Alcaldía de Medellín. (s.f.). *Secretaría de Desarrollo Económico*. <https://www.medellin.gov.co/es/secretaria-desarrollo-economico/>
- Alves, G., & Varvasino, J. (2022, December 13). *Lasting scars: The unequal impacts of unemployment in Latin America*. CAF – Banco de Desarrollo de América Latina y el Caribe. <https://scioteca.caf.com/handle/123456789/1987>
- Arango, L. E., & Bonilla, L. (2015). *Human capital externalities in Colombia* (Borradores de Economía No. 883). Banco de la República de Colombia. <https://ideas.repec.org/p/bdr/borrec/883.html>
- Ávila, F., González, M., & Schmalbach, J. (2023). Perceptions of the socioeconomic effects of Venezuelan migration among informal traders in Cartagena de Indias (Colombia).

*Economics & Sociology*, 16(3), 281–301. <https://doi.org/10.14254/2071-789x.2023/16-3/15>

Blavasciunaite, D., Garšvienė, L., & Matuzevičiūtė, K. (2020). Trade balance effects on economic growth: Evidence from European Union countries. *Economies*, 8(3), 54. <https://doi.org/10.3390/economies8030054>

Bruns, B., Miggelbrink, J., & Müller, K. (2011). Smuggling and small-scale trade as part of informal economic practices. *International Journal of Sociology and Social Policy*, 31(11–12), 664–680. <https://doi.org/10.1108/01443331111177869>

Cámara de Comercio de Medellín para Antioquia. (s.f.). *Characterization of informal businesses in Medellín*.

<https://www.camaramedellin.com.co/Portals/0/Documentos/CEM/CHARACTERIZACION%20OF%20INFORMAL%20BUSINESSES%20IN%20MEDELLIN.pdf>

Cámara de Comercio de Medellín para Antioquia. (s.f.). *Primer informe Mesa de Empleo de Antioquia*. <https://www.camaramedellin.com.co/noticias/primer-informe-mesa-de-empleo-de-antioquia>

Cardona, A., & Giraldo, D. (2023). Trabajo informal en mujeres cabeza de familia de la ciudad de Medellín: Caso de estudio. *Mujer y Políticas Públicas*, 2(2), 103–128. [https://www.researchgate.net/publication/377486564\\_Trabajo\\_Informal\\_en\\_Mujeres\\_Cabeza\\_de\\_Familia\\_de\\_la\\_Ciudad\\_de\\_Medellin\\_Caso\\_de\\_Estudio](https://www.researchgate.net/publication/377486564_Trabajo_Informal_en_Mujeres_Cabeza_de_Familia_de_la_Ciudad_de_Medellin_Caso_de_Estudio)

Deza, M. C., Jara, H. X., Oliva, N., & Torres, J. (2020). *Financial disincentives to formal employment and tax-benefit systems in Latin America* (EUROMOD Working Paper No. EM8/20). <https://hdl.handle.net/10419/228407>

- Farhadi, M., Ismail, R., & Fooladi, M. (2012). Information and communication technology use and economic growth. *PLOS ONE*, 7(11), e48903.  
<https://doi.org/10.1371/journal.pone.0048903>
- Fiess, N., & Fugazza, M. (2012). Informality and openness to trade: Insights from cross-sectional and panel analyses. *Margin: The Journal of Applied Economic Research*, 6(2), 235–275.  
<https://doi.org/10.1177/097380101200600207>
- García, A., Gómez, M., & Otero, S. (2022). *Firm formality and survival in Colombia*. World Bank. <https://thedocs.worldbank.org/en/doc/b28032549509733b9802132e0d65fdd9-0370012023/original/Firm-formality-30-12-2022.pdf>
- García, F., & Mendoza, L. (2021). *Evolución de la informalidad laboral y sus determinantes en Colombia 2010–2020*. Universidad Nacional de Colombia.  
<https://repositorio.unal.edu.co/handle/unal/80899>
- Han, Y. (2022). Personality, cognitive bias and entrepreneur decision-making. In *Proceedings of the 2022 International Conference on Economics, Business and Management Research*. Atlantis Press. <https://doi.org/10.2991/aebmr.k.220603.073>
- International Monetary Fund. (2018). *Colombia: Selected issues* (IMF Country Report No. 18/129). International Monetary Fund.  
<https://www.elibrary.imf.org/downloadpdf/view/journals/002/2018/129/article-A002-en.pdf>
- Joao Felix, M. A., & Lima, G. T. (2023). *Aplicación de algoritmos de aprendizaje automático para predecir el tamaño de la economía informal* (Documento de Trabajo No. 2023-10). Universidad de São Paulo (FEA-USP).

- Jokia, L., Swanepoel, E., & Venter, M. (2021). Economic empowerment of women in the informal African craft industry: The case of Tembisa. *Journal of Economic and Financial Sciences*, 14(1). <https://doi.org/10.4102/jef.v14i1.578>
- Katz, R. (2024). *Oportunidades para la transformación digital productiva: Evidencia estadística sobre el nivel de digitalización sectorial en América Latina y el Caribe* (Documentos de Proyectos LC/TS.2023/176). Comisión Económica para América Latina y el Caribe (CEPAL). <https://repositorio.cepal.org/server/api/core/bitstreams/3f91f0ea-bd6f-493c-95b8-19306dad70e/content>
- La Porta, R., & Shleifer, A. (2014). Informality and development. *Journal of Economic Perspectives*, 28(3), 109–126. <https://doi.org/10.1257/jep.28.3.109>
- Laguna, L., Govea, L., Fuentes, D., & Lugo, D. (2023). Internalization of rules and informal economy: Empirical evidence in Latin America. *PLOS ONE*, 18(7), e0288684. <https://doi.org/10.1371/journal.pone.0288684>
- Liang, Y., & Aquino, J. (2024). Integration of information technology in higher vocational economic and trade disciplines: Inputs for a development plan. *Proceedings of the EAI International Conference*. <https://doi.org/10.4108/eai.23-2-2024.2345956>
- López-Martín, B. (2018). Informal sector misallocation. *Macroeconomic Dynamics*, 23(8), 3065–3098. <https://doi.org/10.1017/S1365100517001055>
- Mawejje, J. (2024). Government expenditure, informality, and economic growth: Evidence from Eastern and Southern African countries. *African Development Review*, 36(1), 125–138. <https://doi.org/10.1111/1467-8268.12729>

Mora et al. (2023) . Returns to human capital in a developing country: A pseudo-panel approach for Colombia. *Revista de Economía del Rosario*, 26(2), 1–27.

<https://www.researchgate.net/publication/372742186>

Muzi, S., Jolevski, F., Ueda, K., & Viganola, D. (2021). *Productivity and firm exit during the COVID-19 crisis: Cross-country evidence* (Policy Research Working Paper No. 9671).

World Bank. <http://hdl.handle.net/10986/35632>

Naciones Unidas, Comisión Económica para Europa (UNECE), Comisión Económica para América Latina y el Caribe (CEPAL), Comisión Económica y Social para Asia y el Pacífico (ESCAP), Comisión Económica y Social para Asia Occidental (ESCWA), & Comisión Económica para África (ECA). (2025). *On the road to the Second World Summit for Social Development: Contributions from the regional commissions*

(LC/TS.2024/147). Naciones Unidas. <https://unece.org/sites/default/files/2025-04/WSSD2-regional-social-development-challenges.pdf>

Nguyen, C., Nguyen, B., & Tran, D. (2022). The internet, economic uncertainty, and the shadow economy: “The good, the bad, and the unknown”? *International Social Science Journal*,

73(247), 9–28. <https://doi.org/10.1111/issj.12381>

Oficina Regional para América Latina y el Caribe, Organización Internacional del Trabajo.

(2024). *Estrategia de formalización para América Latina y el Caribe 2024–2030*

(FORLAC 2.0). Organización Internacional del Trabajo.

<https://www.ilo.org/es/publications/estrategia-de-formalizacion-para-america-latina-y-el-caribe-2024-2030>

- Organización Internacional del Trabajo, & Peña, A. (2022). *Panorama laboral 2022: América Latina y el Caribe*. Organización Internacional del Trabajo.  
<https://www.ilo.org/media/453971/download>
- Ramos-Rodríguez, A. R., Medina-Garrido, J. A., & Ruiz-Navarro, J. (2024). Intellectual and social capital, opportunity recognition, and entrepreneurial orientation. *Journal of Business Research*, 171, 114127. <https://arxiv.org/abs/2401.17448>
- Rubio Pabón, M. V. (2014). El trabajo informal en Colombia e impacto en América Latina. *Observatorio Laboral Revista Venezolana*, 7(13), 23–40.  
<https://www.redalyc.org/pdf/2190/219030399002.pdf>
- Rubio, M. (1997). Perverse social capital: Some evidence from Colombia. *Journal of Economic Issues*, 31(3), 805–816. <https://www.jstor.org/stable/4227229>
- Saavedra, S. (2023). *Formalization assistance and performance of informal firms: Evidence from Colombian mining*. Universidad del Rosario.  
<https://ideas.repec.org/p/col/000092/020793.html>
- Saidakhror, G. (2024). The impact of artificial intelligence on higher education and the economics of information technology. *Irshad Journal of Law and Policy*, 2(3), 1–6.  
<https://doi.org/10.59022/ijlp.125>
- Sánchez, J. (2023). *Theories of informality: A systematic review* [Tesis]. OhioLINK Electronic Theses and Dissertations Center.  
[https://etd.ohiolink.edu/acprod/odb\\_etd/ws/send\\_file/send?accession=oduhonors1683028775332179&disposition=inline](https://etd.ohiolink.edu/acprod/odb_etd/ws/send_file/send?accession=oduhonors1683028775332179&disposition=inline)

Shawon, R. E. R., Rahman, A., Islam, M. R., Debnath, P., et al. (2024). AI-driven predictive modeling of US economic trends: Insights and innovations. *Journal of Humanities and Social Sciences Studies*, 6(10), 1–15.

[https://www.researchgate.net/publication/387187315\\_AI-](https://www.researchgate.net/publication/387187315_AI-Driven_Predictive_Modeling_of_US_Economic_Trends_Insights_and_Innovations)

[Driven Predictive Modeling of US Economic Trends Insights and Innovations](https://www.researchgate.net/publication/387187315_AI-Driven_Predictive_Modeling_of_US_Economic_Trends_Insights_and_Innovations)

Tobar-Cruz, J. S., & Ruiz-Martínez, C. A. (2025). *Formalization and growth of microbusinesses in Colombia: Relationship with credit access from a formal/informal, gender, and regional perspective*. Banco de la República. <https://www.banrep.gov.co/en/publications-research/working-papers-economics/formalization-growth-microbusinesses-colombia-relationship>

Ugulumu, E., Nestory, M., & Mpsa, O. (2023). Factors affecting night market traders' performance in Tanzania: A case of Forodhani night market in Unguja. *RPJ*, 25(1), 59–69. <https://doi.org/10.59557/wm4dz957>

Vidal, P., Sierra-Suárez, L. P., & Cerón, J. (2025). Indicator for the regional labor market using machine learning techniques: Application to Colombian cities. *Revista de Economía del Rosario*, 27(1), 1–31. <https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.14392>