

Cuadernos de Desarrollo Económico

56

Metodología para optimizar
el alcance de la oferta
institucional de la Secretaría
Distrital de Desarrollo
Económico (MOOI)

MARZO 2023

Daniel Cárdenas
Jenny Rivera
Laura Rodríguez



SECRETARÍA DE
DESARROLLO
ECONÓMICO





Claudia Nayibe López Hernández
Alcaldesa Mayor de Bogotá

Alfredo Bateman Serrano
Secretaría de Desarrollo Económico

Sebastián Marulanda Robledo
Subsecretaría de Desarrollo Económico

Maria Catalina Bejarano Soto
Directora de Estudios de Desarrollo Económico

Juliana Aguilar Restrepo
Subdirectora de Estudios Estratégicos

Yaneth Lucía Pinilla Beltrán
Subdirectora de Información y Estadísticas

Autores

Daniel Cárdenas, Jenny Rivera y Laura Rodríguez

Revisión de estilo

Juliana Ortega Camelo

Diseño y diagramación:

César Javier Daza Gutiérrez

Archivo Fotográfico

Secretaría de Desarrollo Económico

Metodología para optimizar el alcance de la oferta institucional
de la Secretaría Distrital de Desarrollo Económico (MOOI)

ISSN

2981-4790 (En línea)

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

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

Contáctenos en:

Secretaría Distrital de Desarrollo Económico

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

Teléfono: 601 369 3777 Ext: 204

observatorio@desarrolloeconomico.gov.co



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

Metodología para optimizar el alcance de la oferta institucional de la Secretaría Distrital de Desarrollo Económico (MOOI)

RESUMEN

La metodología para optimizar el alcance de la oferta institucional de la Secretaría Distrital de Desarrollo Económico (MOOI) es un proceso creado por el Observatorio de Desarrollo Económico de Bogotá (ODEB) para optimizar las convocatorias de los programas de la Secretaría Distrital de Desarrollo Económico (SDDE) y facilitar su evaluación. Para esto, la MOOI utiliza una base única de unidades productivas de Bogotá y estima un puntaje que organiza las unidades productivas de acuerdo con la posibilidad de beneficiarse del programa que estas tienen. Primero, utiliza el análisis de factores para datos mixtos con el fin de resumir en pocas variables los criterios que deben cumplir las unidades productivas para participar de los programas, posteriormente, agrupa las unidades productivas de acuerdo a esta caracterización simplificada usando la técnica *k-means* y, por último, genera un puntaje de priorización de acuerdo a la distancia de la unidad productiva al centroide del grupo preferido para recibir el programa.

Palabras clave: dinámica empresarial, clusterización, componentes principales.

ABSTRACT

MOOI is a process created by the Bogota Economic Development Observatory (ODEB) to optimize the reach of the institutional program portfolio of the District Department of Economic Development (SDDE) and facilitate its evaluation. The MOOI estimates a score to rank Bogota's productive units according to their possibility of benefiting from the program. First, it uses factor analysis for mixed data, to reduce the dimensionality of the criteria of the programs that the participating productive units must meet; then it groups the productive units according to this simplified characterization using the k-means technique and, finally, it generates a prioritization score according to the distance of the productive unit to the centroid of the group with the highest possibility to benefit from the program.

Keywords: business dynamics, clustering, principal components.

CONTENIDO

1	Introducción	8
	1.1 ¿Por qué crear la MOOI?	8
	1.2 Objetivo y beneficios para la SDDE	10
<hr/>		
2	Fases de la MOOI	11
	2.1 Fase 1: Caracterización del programa	13
	2.2 Fase 2: Priorización de potenciales beneficiarios	15
	2.3 Fase 3: Retroalimentación del proceso	16
<hr/>		
3	Metodología de la Fase 2: Priorización de potenciales beneficiarios	20
	3.1 Paso A. Filtros de inclusión	20
	3.2 Paso B. Caracterización de unidades productivas	21
	3.3 Paso C. Agrupación de unidades productivas según características	25
	3.4 Paso D. Elección del grupo más adecuado para recibir el programa	26
	3.5 Paso E. Asignación de puntaje a cada potencial beneficiario	27
<hr/>		

CONTENIDO

4	Posibles extensiones del análisis	28
	4.1 Insumos para nuevos programas de la SDDE	28
	4.2 Retroalimentación de la MOOI	28

5	Bibliografía	30
----------	---------------------	-----------



1. INTRODUCCIÓN

El presente documento tiene como objetivo describir la metodología para optimizar el alcance de la oferta institucional de la Secretaría Distrital de Desarrollo Económico (MOOI), la cual es implementada por el Observatorio de Desarrollo Económico de Bogotá (ODEB) y busca mejorar y optimizar el grupo de las unidades productivas a las cuales se les ofrecen los programas de la Secretaría Distrital de Desarrollo Económico y entidades aliadas del distrito.

1.1 ¿POR QUÉ CREAR LA MOOI?

Hay posibilidades de mejorar la estrategia de convocatoria a la ciudadanía de los programas de la SDDE. Actualmente cada dirección o subdirección de la entidad desarrolla de forma separada sus estrategias de difusión, cuestionarios e inscripciones de los participantes, y no existe una base única de unidades productivas que permitan ampliar la difusión de forma focalizada. En consecuencia, las unidades productivas o personas¹ que son idóneas para un programa no se enteran de dicha convocatoria o aplican a programas que no son idóneos para sus características. Lo anterior se suma a los costos adicionales en los que incurre cada estrategia cuando hace una convocatoria. Por esta razón, y con el fin de enfocar los programas de la SDDE hacia las poblaciones más necesitadas o las más estratégicas para la ciudad, desde el ODEB se propone implementar la MOOI. Este servicio busca mejorar la forma en la que se seleccionan los participantes de los programas y hacer más eficientes las convocatorias de la entidad.

En concreto, va a permitir identificar dentro de una lista de unidades productivas de las que se tiene información a las que cumplen en mayor medida los criterios de selección de un programa determinado. Es decir, sacar el grupo de personas o unidades productivas cuyas características les permiten acceder al programa y obtener el mayor beneficio de este.

¹ Por la naturaleza del trabajo de la SDDE, los posibles beneficiarios de los programas son personas, si están enfocados en empleo, o unidades productivas, si están enfocados en fortalecer los procesos productivos. Este documento está especialmente dirigido a las unidades productivas, sin embargo, el servicio también sirve para individuos.

Asimismo, la MOOI de la SDDE es el proceso que optimiza el emparejamiento entre los programas de la SDDE con sus potenciales participantes. En particular, utiliza bases de datos que permiten caracterizar las unidades productivas para focalizar la difusión de los programas, de tal manera que los prestadores de los programas puedan activamente contactar a quienes más podrían beneficiarse y operen en una lógica de oferta y no únicamente de demanda por parte de la unidad productiva. Es decir, la MOOI le brinda al oferente del programa un listado de potenciales beneficiarios que puede contactar para ampliar el grupo de unidades productivas que se benefician de los programas y que este grupo no se limite a aquellas que llegan a través de las convocatorias abiertas. Para este fin se utilizan los objetivos y los criterios de priorización establecidos por el programa y una caracterización detallada de las unidades productivas para generar un listado de priorización de aquellas que cumplen con las condiciones y que pueden beneficiarse más del programa.

Para lo anterior, es necesario construir una base de datos de unidades productivas que sirva como insumo de posibles beneficiarios y que contenga información relevante para los programas de la SDDE. Actualmente, se cuenta con los datos de la Tropa económica, un trabajo de recolección de datos, por parte de un grupo de encuestadores, impulsado por la Alcaldía Mayor de Bogotá para caracterizar el tejido productivo de la ciudad, especialmente, aquellas unidades productivas que se vieron afectadas negativamente por la pandemia por covid-19. El cuestionario de la Tropa contiene preguntas relacionadas con las distintas temáticas en las que trabajan las direcciones de la SDDE. A partir de esta base, es posible caracterizar y priorizar las unidades productivas, así como hacer cruces con otras fuentes de datos para obtener información adicional de la unidad productiva o sus propietarios. A futuro, la labor de la Tropa económica puede ser reemplazada por convocatorias masivas y periódicas de actualización de datos.

Además de generar insumos para la base de unidades productivas de la ciudad, la Tropa Económica cumple la función de difundir los programas de la SDDE fomentando la aplicación por demanda de los potenciales beneficiarios. En efecto, este ejercicio en campo no sólo captura información de las unidades productivas, también les comparte la oferta de la SDDE (y otras entidades aliadas) y las guía hacia cuáles les pueden ayudar más a suplir sus necesidades. Esto, sin embargo, no suple ni elimina la labor de la MOOI, es decir, esas unidades productivas siguen siendo elegibles para ser priorizadas por algún programa. Asimismo, se espera que la MOOI pueda ser usado para filtrar automáticamente los programas que se ofrecen a las unidades productivas en el momento de la entrevista de recolección de datos o las visitas de seguimiento.

1.2 OBJETIVO Y BENEFICIOS PARA LA SDDE

El objetivo principal del servicio de la MOOI es optimizar las convocatorias de unidades productivas, mejorando la selección de los participantes potenciales de los programas de la SDDE. Esto permite avanzar en dos de los objetivos estratégicos de la entidad (SDDE, 2020):

1. *Formular y hacer seguimiento a políticas públicas del sector de Desarrollo Económico apoyando su implementación en coordinación con entidades y actores aliados.* **El servicio de MOOI permite hacer un seguimiento detallado de las políticas de la entidad** y da lineamientos de base para la recolección de información en convocatorias. Esto posibilita su caracterización y apoya la creación de conocimiento para mejorar futuras políticas. En concreto, facilita la realización de evaluaciones de seguimiento, resultados e impacto.
2. *Mejorar la eficiencia operativa y la capacidad tecnológica y comunicativa de la Secretaría Distrital de Desarrollo Económico, para apoyar el cumplimiento de la misión de la entidad.* **El servicio de MOOI hace más eficiente el uso de datos de convocatorias y actividades de campo (como la Tropa económica) para que las direcciones misionales no dupliquen esfuerzos y hagan convocatorias de una forma más pertinente y eficaz.** Además, la priorización de potenciales beneficiarios (el resultado final del proceso), garantiza que las direcciones de la SDDE utilicen sus recursos para contactar a las unidades productivas que más se beneficiarían de los programas y posiblemente reduce el desgaste de las unidades productivas de recibir convocatorias que no son pertinentes.

2. FASES DE LA MOOI

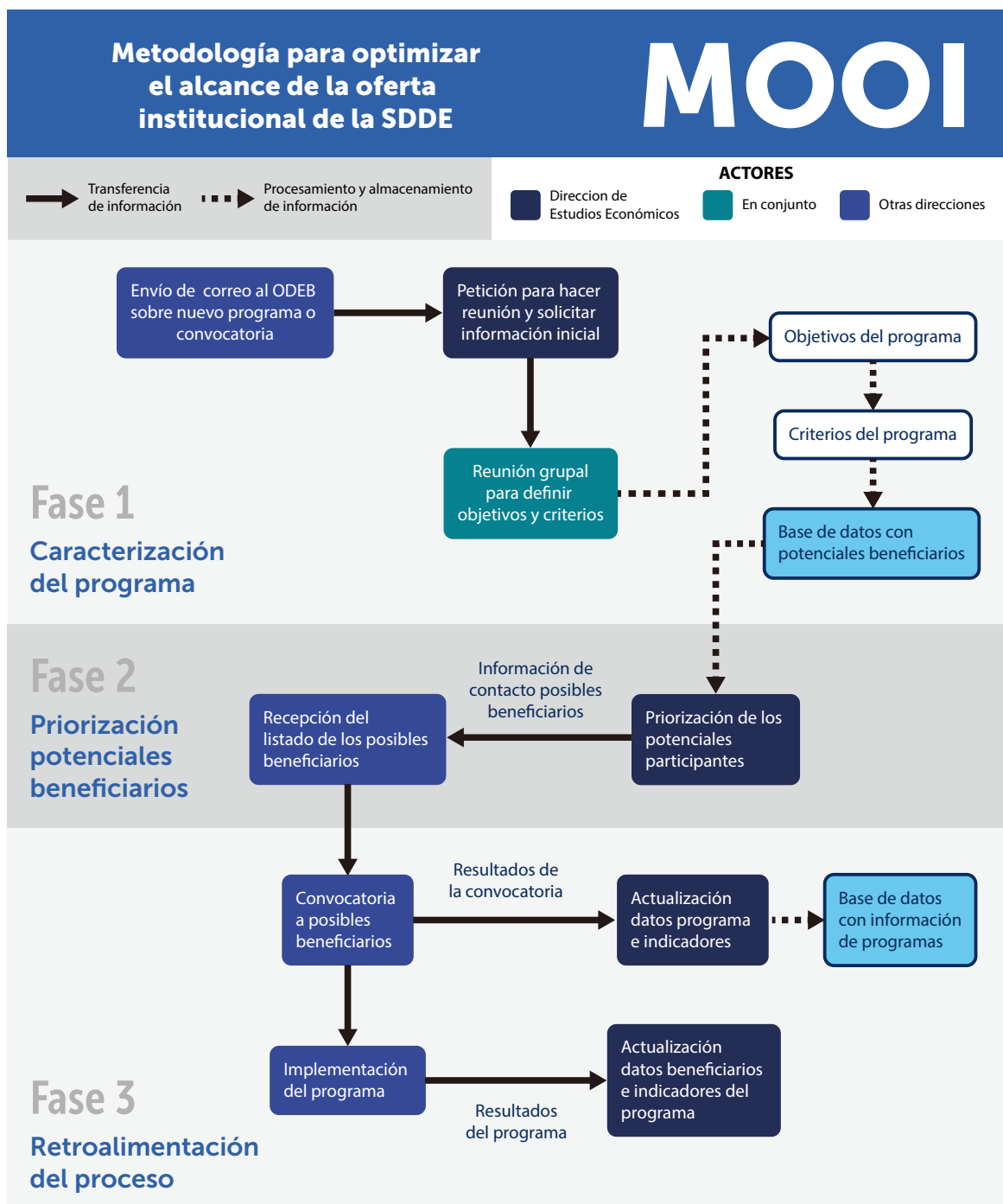
El proceso de la MOOI se desarrolla en tres fases (ver Ilustración 1): la caracterización del programa para el cual se priorizarán los beneficiarios, la priorización de potenciales beneficiarios, y la retroalimentación del proceso. En la primera fase se establecen, en colaboración con la dirección misional que diseña e implementa el programa, los objetivos finales del programa, los requisitos para la participación de una unidad productiva y las características que un beneficiario ideal tendría.

La segunda fase es el corazón de la MOOI ya que produce la priorización de potenciales beneficiarios. Se desarrolla desde el ODEB, y utiliza los insumos de la primera fase y metodologías de análisis de componentes principales (más conocido como PCA por sus siglas inglés –Principal component analysis) y clusterización para priorizar las unidades productivas, de acuerdo a las características más relevantes para que se beneficien del programa.

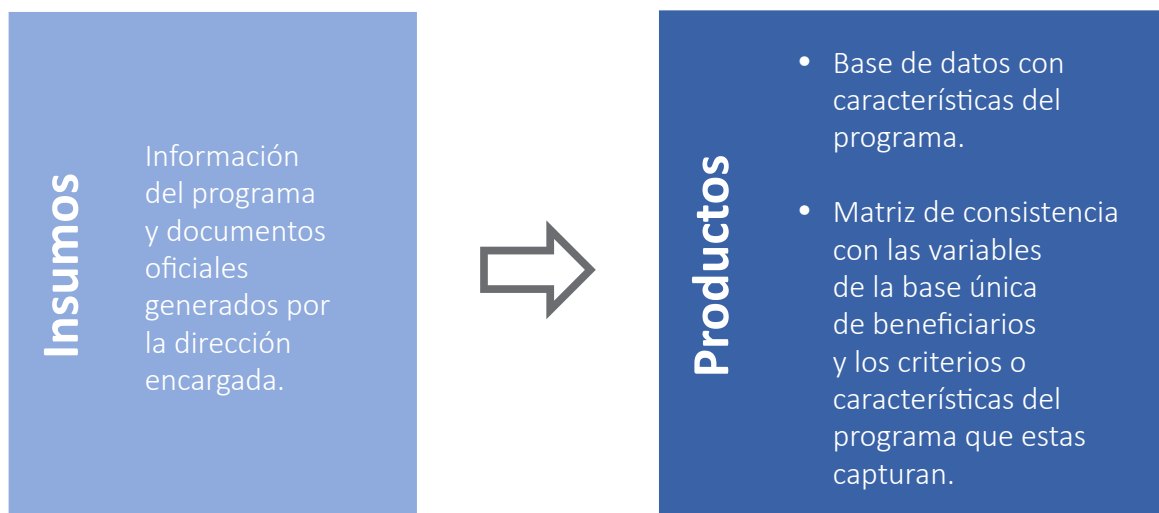
Durante la tercera fase, se busca conocer qué pasó durante el programa para mejorar la formulación de políticas y llevar registros de los resultados de la convocatoria, para mejorar la priorización. Esto se logra consolidando los datos de unidades productivas recolectados durante la convocatoria y la implementación del programa. Esto alimentará las bases de datos usadas en la segunda fase para así ajustar, cuando sea pertinente, el proceso de priorización.

A continuación (Ilustración 1), se detalla cada una de las fases, identificando los insumos, procedimientos y productos que surgen de las mismas.

Ilustración 1. Proceso de la MOOI por fases



2.1 FASE 1: CARACTERIZACIÓN DEL PROGRAMA



Esta fase está enfocada en definir los objetivos del programa y los criterios que se utilizan para elegir a sus beneficiarios. Esto implica que los miembros del ODEB se reúnan con los de la dirección implementadora y que conozcan los documentos que describen el programa. Con base en estos dos insumos, se eligen las variables de la base de unidades productivas que capturan los criterios y requisitos del programa y que servirán de insumo para la fase 2 de la MOOI.

Es importante que la información sobre el programa consignada en los documentos sea complementada con conversaciones con los funcionarios responsables del diseño e implementación del programa. En pilotos llevados a cabo por el ODEB, se ha encontrado que el uso de las dos fuentes permite tener una perspectiva más amplia del programa a tratar y optimizar el tipo de variables que deben incluirse en el análisis. Cabe mencionar que no existe un solo documento o clase de documento que sirva para establecer los objetivos del programa, pueden ser todos los que la dirección que implementa el programa considere pertinentes.

Al final de la fase se deben haber definido objetivos precisos y concretos del programa, así como los criterios de priorización de este y las variables de las bases de datos que caracterizan las unidades productivas y responden a los criterios del programa identificados. Estos deben quedar consignados en una matriz de consistencia que permita reconocer las variables a analizar. Un ejemplo de esta matriz se muestra en la Tabla 1.

En la matriz de consistencia se consignan los identificadores únicos del programa y las variables que caracterizan las unidades productivas que el programa busca beneficiar. Es importante que ese identificador corresponda al utilizado en la Secretaría por el Sistema unificado de información misional (SUIM) u otro sistema que contenga la información de los programas. De otro lado, en la matriz se identifican dos tipos de variables: las *variables*

incluyentes, son los mínimos que debe cumplir una unidad productiva para entrar al programa, por ejemplo, un número máximo de empleados. Las *variables de inclusión* son aquellas que optimizan la selección de beneficiarios que acceden a los programas, al ser variables que son deseables que tengan las unidades productivas y que han identificado como las que pueden potencializar el éxito de la unidad al entrar al programa o facilitan el cierre de brechas existentes entre unidades.

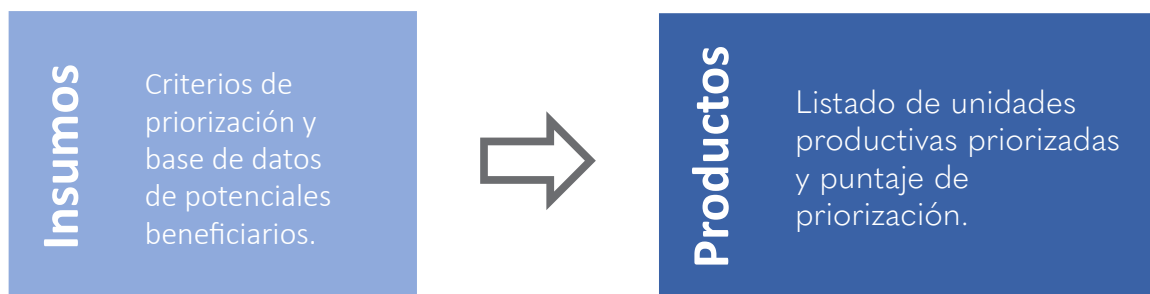
Para la aplicación de la metodología, es necesario identificar cuáles variables en la base de datos pueden dar cuenta de las variables de inclusión y priorización. Una vez elegidas, se verifica la calidad de estas, y se procede a ajustarlas o limpiarlas según sea necesario. Por ejemplo, para la variable de los ingresos se define un ingreso mínimo y máximo a partir de los cuales las observaciones se consideran valores atípicos, si algún registro sobrepasa estos umbrales se registra como nulo y posteriormente, se hace un reemplazo de la media de los ingresos en aquellos registros nulos. Otro método de limpieza que se utiliza con frecuencia, es la recodificación de variables categóricas en variables binarias.

Tabla 1. Ejemplo matriz de consistencia

Identificador único del programa	000000-0		
Nombre del programa	Bogotá Productiva Local		
Dirección encargada	Dirección de Desarrollo Empresarial y Empleo		
Objetivo	"Esta es la ruta del fortalecimiento y crecimiento empresarial para las micro unidades productivas de Bogotá; es decir, aquellas que requieren capital para iniciar, cualificación en temas financieros y comerciales, y asesoría para encontrar nichos de mercado."		
Criterios de inclusión	<ul style="list-style-type: none"> Mínimo 1 año de antigüedad 	<ul style="list-style-type: none"> Menos de 9 trabajadores 	
Variables de inclusión	<ul style="list-style-type: none"> 'ANTIGÜEDAD' se eliminan las que pertenecen a la categoría "menos de un año" 	<ul style="list-style-type: none"> 'NO_TRABAJADORES' menor o igual a 9 	
Criterios de priorización	<ul style="list-style-type: none"> Menores ingresos 	<ul style="list-style-type: none"> Propietarios mujeres 	<ul style="list-style-type: none"> Necesidades de inversión
Variables de priorización	<ul style="list-style-type: none"> 'INGRESOS' es inferior a la mediana. 	<ul style="list-style-type: none"> Crear la variable de porcentaje de propietarias mujeres usando 'PROPIETARIOS_MUJERES', 'PROPIETARIOS_HOMBRES' y 'PROPIETARIOS_OTRO'. 	<ul style="list-style-type: none"> Se cuenta si dan más del 50 % Donde 'INVERTIR' o 'QUE_REQUIERE_REACTIVAR' no sea nulo.

Fuente: Elaborado SDDE-ODEB

2.2 FASE 2: PRIORIZACIÓN DE POTENCIALES BENEFICIARIOS



La Fase 2 realiza la priorización de unidades productivas de acuerdo con los criterios identificados en la Fase 1. Con base en las variables definidas en la anterior fase y utilizando metodologías de análisis no supervisado, se crea un índice de priorización donde aquellas unidades productivas que más coinciden con los objetivos del programa tienen mayor puntaje. La metodología se detalla en la sección 3.

Los criterios de priorización deben reflejarse en variable medibles en la base de unidades productivas y pueden tener diferentes niveles de relevancia. Todos los criterios elegidos deben ser medibles de acuerdo con las variables disponibles en la base de datos de potenciales beneficiarios, bien sea de forma directa o a través de variables proxy para esas características. Además, un criterio o variable puede ser un requisito necesario para el programa, por ejemplo, si el programa está focalizado en unidades productivas cuyo representante legal es mujer, solo se consideran estas unidades en la priorización. Así mismo, unos criterios pueden ser más relevantes para el cumplimiento de los objetivos del programa.

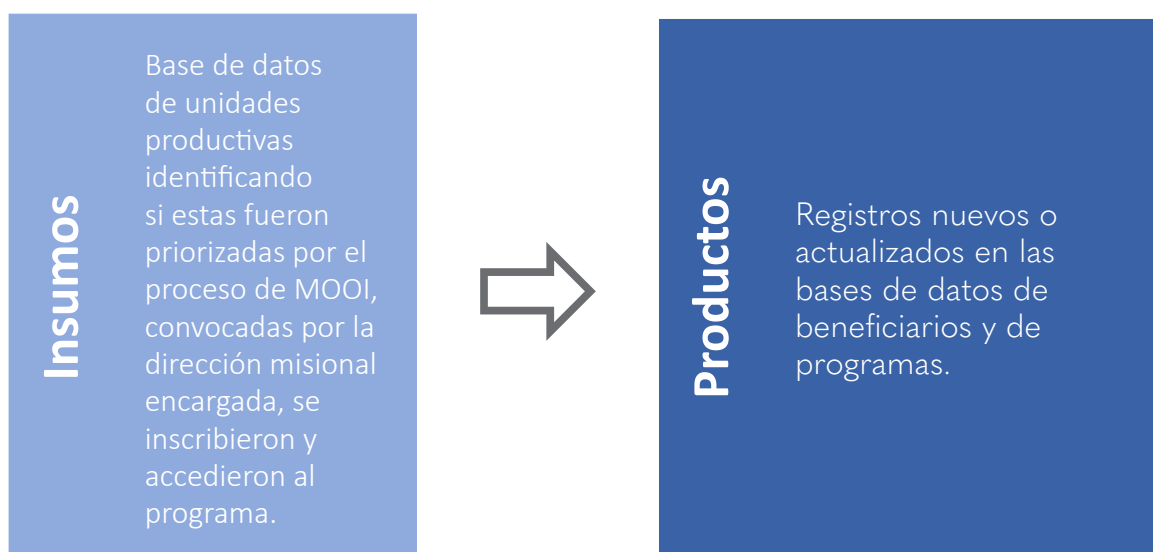
El insumo para esta fase es una base de datos de unidades productivas que contenga como mínimo:

1. Localidad donde se ubica la unidad productiva.
2. UPZ donde se ubica la unidad productiva.
3. Dirección completa de la unidad productiva.
4. Razón social o nombre comercial de la unidad productiva.
5. Teléfono o correo de contacto de la unidad productiva.
6. Variables necesarias para la priorización según las necesidades del programa. Las más comunes son: sexo del propietario, número de trabajadores, ingresos de la unidad productiva, planes de contratación de personal y actividad económica según CIU².

² El código CIU es una clasificación de las actividades productivas que proviene de acuerdos internacionales y que es ajustado para Colombia por la Dirección de Impuestos y Aduanas Nacionales (DIAN)

El producto de esta fase es un listado de unidades productivas priorizadas para el programa con información de contacto y un puntaje que indique su relevancia para este. Siempre que sea pertinente, el número de unidades priorizadas debe superar por lo menos en un 10 % el número de unidades productivas que tienen como meta para contactar por parte del programa. En ese sentido, la dirección que implementa puede enfocar sus esfuerzos en contactar a los beneficiarios óptimos para el programa y, al mismo tiempo, tienen margen de maniobra considerando que las unidades productivas no contesten o no quieran ser parte de la iniciativa. El umbral mínimo será revisado una vez haya retroalimentación sobre la cantidad de las unidades productivas contactadas.

2.3 FASE 3: RETROALIMENTACIÓN DEL PROCESO



Una vez terminada la priorización de las unidades productivas, la dirección desarrolla el programa y es necesario conocer qué tan efectiva fue la convocatoria y el resultado de la implementación. Esto con el fin de conocer qué tan exitosa fue la priorización de la MOOI y analizar si los datos de las unidades productivas requieren alguna mejora. Por otro lado, en esta etapa se lleva un registro de los beneficiarios de los programas, lo cual, a futuro, facilita la evaluación del programa. De la misma forma, ese registro facilita hacer un seguimiento de los servicios en los que participó una unidad determinada y así mejorar la oferta de la SDDE, tanto en pertinencia como en eficacia.

Para el desarrollo de esta fase de la MOOI, **se requiere que la dirección implementadora recoja datos de identificación sobre la inscripción de unidades productivas en el programa, tanto de las participantes como de las que no lo fueron.** El ODEB usará estas bases de datos para retroalimentar el proceso de MOOI de tres formas: la primera, midiendo el éxito de la priorización y mejorando el proceso, si es necesario, la segunda, alimentando la base

de datos de potenciales beneficiarios con los registros de aquellas unidades productivas que se inscribieron al programa y no estaban en la base, y la tercera, actualizando la información de las unidades productivas.

Para mejorar el proceso, al finalizar cada programa es necesario analizar los indicadores de la priorización, tanto de la efectividad de contacto como de la tasa de deserción de los priorizados para entender qué ajustes deben hacerse para que la convocatoria sea más efectiva. En la misma línea, al finalizar el programa se pueden analizar las características que se encontraban en la base de potenciales beneficiarios y determinar si hay variables que hace falta medir, ajustar o cambiar porque no están dando información útil.

Si los criterios de priorización son claros, se esperaría que las unidades productivas priorizadas que fueron contactadas participen del programa en una alta proporción y, además, que permanezcan hasta el final. Estos dos puntos se refieren a los indicadores de éxito del proceso, los cuales asumen que una unidad productiva a la que se le ofrece un programa acorde a sus necesidades y características va a querer y poder participar (supuesto de voluntad) y que las direcciones ejecutan la convocatoria y el programa tal como está estipulado (supuesto de implementación).

En concreto, se definen dos indicadores para medir la eficiencia de la MOOI:

- **Efectividad de priorización**, medida como la proporción de las unidades productivas priorizadas por la MOOI y contactadas efectivamente para participar en el programa (sin contar las que no contestaron o cuya información de contacto cambió) que aceptan inscribirse al programa. La meta es que sea cercano al 1.

$$\text{Efectividad de priorización} = \frac{\text{Unidades productivas priorizadas que entraron al programa}}{\text{Total de unidades productivas priorizadas contactadas}}$$

- **Tasa de deserción de los priorizados**, medida como las unidades productivas priorizadas que entraron al programa y se salieron de él antes de finalizarlo (desertaron) sobre el total de las unidades productivas priorizadas que entraron al programa. La meta es que sea cercano a 0.

$$\text{Tasa de deserción de los priorizados} = \frac{\text{Unidades productivas priorizadas que entraron al programa y desertaron}}{\text{Unidades productivas priorizadas que entraron al programa}}$$

Cabe resaltar que el listado de unidades productivas que surge de la MOOI no debe ser la única estrategia de convocatoria del programa y que el éxito de este último no está mediado por el éxito en la priorización producto de la MOOI.

Adicionalmente, **se propone también crear una matriz de confusión de la priorización y participación en el programa** (ver Tabla 2). Al analizar los distintos perfiles según esta matriz, se pueden encontrar otras variables que pueden definir la participación o no en el programa. Por ejemplo, sería interesante analizar las diferencias entre los verdaderos positivos y falsos negativos de manera que se pueden encontrar otras razones, distintas a las variables de priorización definidas en la matriz de consistencia, que influyen en la decisión de participar.

Tabla 2. Matriz de confusión de la participación en los programas y la priorización de la MOOI

		Unidades productivas según participación en el programa	
		Participaron	No participaron
Unidades productivas según priorización por el proceso de la MOOI	Priorizada	<i>Verdaderos positivos:</i> priorizadas por la MOOI y participaron en el programa	<i>Falsos positivos:</i> priorizadas por la MOOI y no participaron en el programa
	No priorizada	<i>Falsos negativos:</i> no priorizadas por la MOOI y participaron en el programa	<i>Verdaderos negativos:</i> no priorizadas por la MOOI y no participaron en el programa

Fuente: Elaborado SDDE-ODEB

La base de las unidades productivas que se inscribieron al programa, así como de aquellas seleccionadas para ser parte de este, servirá también para alimentar la base de potenciales beneficiarios y estas unidades pasaran a ser potenciales beneficiarios de otros programas de la Secretaría.

Como resultado de esta fase, las bases de datos de beneficiarios y la de programas deben estar actualizadas con la información de la intervención. Como se mencionó anteriormente, esto facilita la gestión del conocimiento al hacer más viable las evaluaciones, tanto de implementación como de impacto del programa y a llevar un seguimiento de los beneficios que reciben las unidades productivas.

La base de unidades productivas que se suscribieron al programa debe contener como mínimo:

- 1) Identificación de la unidad productiva
 - a. Localidad
 - b. Dirección completa
 - c. Nombre comercial o razón social

- d. Nombre de contacto
 - e. Teléfono de contacto
 - f. Correo de contacto
 - g. RIT, RUT o NIT en caso de contar con alguno
- 2) Identificación del programa al que accedió. Este debe ser un identificador único para cada programa de manera que se pueda cruzar con la matriz de consistencia descrita en la fase 1.
 - 3) Estado: aplicó y no fue seleccionado, fue seleccionado pero desertó y en qué fase del programa desertó, fue seleccionado y finalizó el programa.
 - 4) Priorizado con MOOI (sí o no).
 - 5) Detalles de la intervención que recibió. Por ejemplo, si la unidad productiva recibió capital para invertir, cuánto recibió y bajo qué criterios.

3. METODOLOGÍA DE LA FASE 2: PRIORIZACIÓN DE POTENCIALES BENEFICIARIOS

Como se vio anteriormente, **la priorización es el corazón del proceso de MOOI. Por esta razón, en esta sección se exponen los pasos concretos que se siguen para cada programa a priorizar.** La priorización se sustenta en una base de datos de unidades productivas o potenciales beneficiarios, así como una lista de variables que reflejan los criterios de participación en el programa.

La priorización de los posibles beneficiarios sigue cuatro pasos que se muestran a continuación. Adicionalmente, un ejemplo de la implementación del proceso de la MOOI se encuentra en el repositorio <https://github.com/Observatorio-SDDE-Bogota/MOOI>.

3.1 PASO A. FILTROS DE INCLUSIÓN

Para preparar las variables primero se debe garantizar que la base de datos de unidades productivas ha pasado por un proceso de revisión y limpieza. La limpieza de la base de datos consiste en asegurarse que la base no contiene duplicidad de registros por empresa y que los datos de cada variable son consistentes con el formato y rangos de valores esperados. Este proceso debe hacerse solo una vez para dejar la base preparada para la priorización de todos los programas.

Una vez se cuenta con una base de datos para el análisis, se filtran las unidades productivas elegibles por el programa específico. Es decir, se define un subconjunto de unidades productivas, por ejemplo, que esté enfocado sólo en negocios de propiedad de mujeres o sólo para empresas de turismo. Lo anterior, se hace utilizando las variables de

exclusión definidas en la matriz de consistencia. Una vez filtrada, la base se deja sin esas observaciones para el resto del análisis.

Por último, se analiza la distribución de las variables y se eliminan los valores extremos o faltantes para evitar distorsiones en el análisis. Específicamente, se busca que la distribución de las variables siga una distribución normal, es decir, que los datos se distribuyan alrededor de la media. La normalización es un paso anterior importante para la reducción de dimensiones (Paso B), puesto que este es un proceso de maximización de varianza. Si no se normalizan los datos numéricos de gran variabilidad, como, por ejemplo, los ingresos, esta variable absorbe toda la variabilidad entre unidades, dejando rezagadas otras variables que también pueden explicar la diferencia entre unidades. El proceso de MOOI utiliza la estandarización utilizando el estadístico z (conocido en inglés como *z-score normalization*).

Adicionalmente, se tratan los valores faltantes. Se evalúa la calidad de cada variable y se decide si se elimina del análisis (porque una mayoría son valores faltantes), si se eliminan los registros que no tienen datos para esta variable, o si se imputa los valores faltantes. Según la variable, la imputación se hace mediante una regresión lineal o con la media del resto de datos. Para tomar estas decisiones se tiene en cuenta la dimensión de la base de datos filtrada y el número de unidades productivas que el programa busca beneficiar, así como el número de variables con datos faltantes por observación.

3.2 PASO B. CARACTERIZACIÓN DE UNIDADES PRODUCTIVAS

La reducción de dimensiones es el proceso mediante el cual la información de muchas variables se condensa en unas pocas. Este procedimiento elimina el ruido de la base de datos, es decir, aquella información que no aporta al análisis, y captura la información que efectivamente describe el comportamiento de los datos (componentes). Entonces, la reducción de dimensiones busca dejar solamente los componentes, una o varias variables que describen de forma más sucinta la información presente en la base.

Existen múltiples metodologías para realizar el anterior proceso, sin embargo, la más utilizada es el análisis de componentes principales. De hecho, se ha encontrado que, comparada con otras metodologías, el PCA es más eficiente para disminuir el ruido (Reddy et al., 2020) (ver Recuadro 1). Sin embargo, una de las limitaciones de esta metodología es que requiere que todas las variables a usar sean continuas pero los datos disponibles de las unidades productivas contienen, en su mayoría, variables categóricas. Por otro lado, existe otra metodología enfocada únicamente en categóricas, pero tiene la limitante de las cuantitativas como el ingreso o el número de empleados que son necesarias para la MOOI (ver Recuadro 2).

Para reducir la dimensionalidad tanto de variables continuas como categóricas se utiliza el análisis de factores para datos mixtos (FAMD en inglés). Al igual que el PCA, este análisis parte del principio de que existen unos componentes que condensan la varianza de múltiples variables. A diferencia del PCA, el FAMD tiene en cuenta la varianza tanto de variables continuas como categóricas. (Khan et al., 2010) (ver Recuadro 3).

Recuadro 1. Metodología Análisis de Componentes Principales (PCA)

El PCA es una metodología de análisis de datos no supervisada que toma una matriz de datos con variables correlacionadas y genera unos componentes principales que contienen la información de estas variables, pero no están correlacionados. Cada uno de los componentes principales corresponde a un eigenvector (vector propio) que a su vez tiene un eigenvalor (valor propio) asignado. El orden de los componentes principales se presenta de manera descendente, de manera que, el primer componente es aquel con el eigenvalor asociado más alto y a la vez hace un mayor aporte a la varianza.

Esta metodología es utilizada para resumir la información presente en una matriz de datos porque de muchas variables se pasa a unas pocas, pero estas últimas pueden dar cuenta del comportamiento general de la matriz. Para el caso tratado en esta nota, existe una gran cantidad de variables de una unidad productiva y muchas están relacionadas entre sí (el sector económico está relacionado con el nivel de ventas, por ejemplo). Por esta razón, usar una reducción de dimensionalidad como el PCA, permite dar cuenta de la misma información sin caer en sesgos por la relación entre variables.

Supuestos:

Cualquier conjunto de datos tiene dos partes: estructura principal y ruido.

Ventajas:

- Facilita la visualización de un conjunto de datos.
- Convierte datos complejos en pocas variables continuas.

Desventajas:

- Es sensible a los valores extremos. Cada variable debe adaptarse para que tenga una distribución normal.
- Solamente utiliza variables continuas.

Recuadro 2. Análisis de correspondencia múltiple (MCA)

La metodología MCA recodifica una variable categórica con más de dos clases diferentes en variables binarias. Este método es similar al PCA, pero utiliza variables categóricas como insumo. El producto son variables que condensan la variación de algunas de las características de las observaciones.

Por ejemplo, una variable de cómo lleva el inventario la unidad productiva que incluye 3 posibles respuestas (no lleva inventario, en un cuaderno o con aplicación) se clasifica de tal manera que no lleva inventario se convierte en $[1,0,0]$, en un cuaderno se convierte en $[0,1,0]$ y con aplicación se convierte en $[0,0,1]$. Este método crea una matriz donde las filas representan los individuos y las columnas son variables binarias. Posteriormente, se aplica un análisis de correspondencia y el resultado es una combinación lineal de filas que contiene la mayor cantidad de información posible de todas las características categóricas.

Supuestos:

Cualquier conjunto de datos tiene dos partes: la estructura principal y ruido.

Ventajas:

- Facilita la visualización de un conjunto de datos.
- Convierte gran cantidad de datos categóricos en pocas variables continuas.

Desventajas:

- Solamente utiliza variables categóricas.

Recuadro 3. Análisis de Factores para Datos Mixtos (FAMD)

EL FAMD es un algoritmo que busca suplir las falencias del análisis de componentes principales (PCA), que sólo trata variables continuas, y del análisis de correspondencia múltiple (MCA), que se centra en variables categóricas y ordinales. Para ello, mezcla ambas metodologías y retorna componentes para conjuntos de datos con todos los tipos de variables.

Supuestos:

Cualquier conjunto de datos tiene dos partes: la estructura principal y ruido.

Ventajas:

- Las mismas del PCA, y adicionalmente utiliza variables cuantitativas y categóricas.

Desventajas:

- Su cálculo es complejo y es difícil hacerle seguimiento a los pasos que sigue el algoritmo.

Por último, se escogen los **primeros componentes principales que explican por lo menos 50% de la variabilidad, en los ejercicios hechos con la metodología de MOOI, estos corresponden usualmente a los dos primeros componentes**. Esto permite reducir la complejidad de los cálculos de los perfiles y facilita el proceso de agrupamiento que se explicará en el siguiente paso. Se analiza la posibilidad de incluir tres componentes en el análisis cuando los dos primeros capturan menos del 50% de la varianza. Adicionalmente, hace más comunicable la priorización ya que permite graficar los datos e ilustrar de forma más clara y transparente dónde se ubican las unidades productivas.

Cabe resaltar, que **el FAMD permite identificar un número óptimo de componentes: el menor número posible que explique la mayor proporción de la variación**. Dado que los componentes no están correlacionados, la ganancia al incluir un componente más es que captura una fracción adicional de la variación de las variables originales de la base, por lo que, en la medida en que se aumenta el número de componentes, también aumenta la variación que explica. Sin embargo, a medida que aumenta la cantidad de componentes se hace cada vez más pequeño y la nueva base de datos se parece más a la original. Idealmente se usaría el número óptimo de componentes para los cálculos, sin embargo, esto no es pertinente para el siguiente paso, ya que la estimación de perfiles o clústeres se dificulta cuando se cuenta con más de 3 componentes e incluso es posible que esta no pueda hacerse.

3.3 PASO C. AGRUPACIÓN DE UNIDADES PRODUCTIVAS SEGÚN CARACTERÍSTICAS

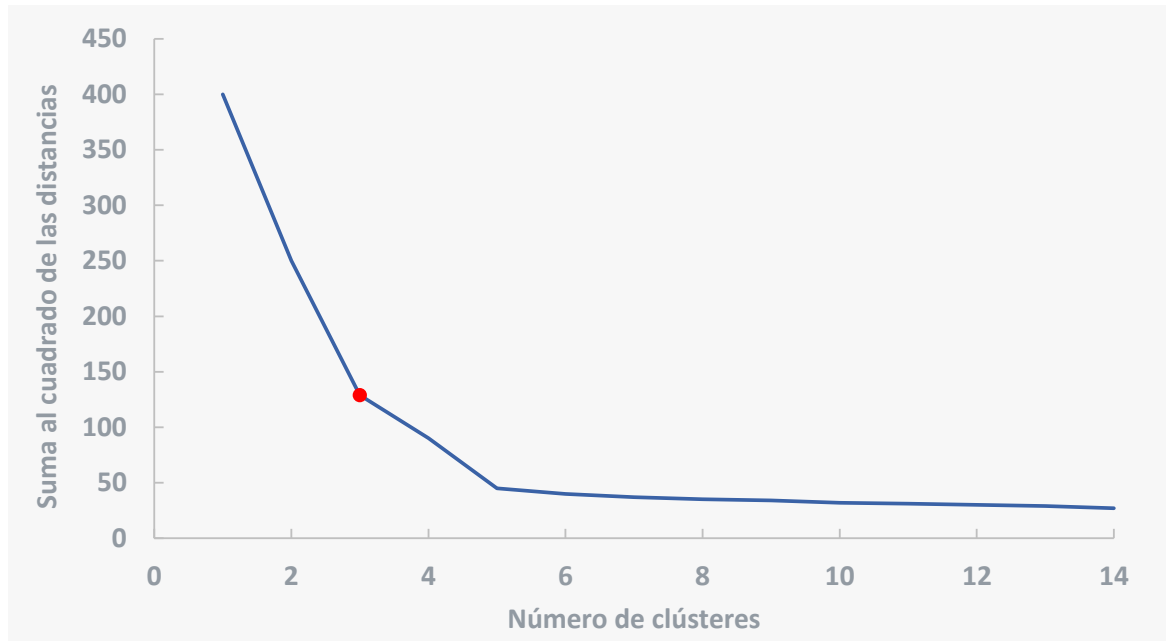
En el paso anterior, se genera una base de datos que para cada unidad productiva contiene los dos componentes principales que recogen sus características. Con base en esta, se crean perfiles agrupando las unidades productivas más parecidas entre sí. Estos grupos se utilizan en el siguiente paso para encontrar aquellas unidades productivas que más se ajustan a los criterios de selección del programa.

Los perfiles de unidades productivas se estiman usando la metodología de clusterización por *k-means*. *K-means* es una de las metodologías más usadas para hallar clústeres o cúmulos de datos y su uso es tan extendido porque esos grupos tienen la característica de que sus miembros son lo más parecido posible entre sí y lo más disímiles de los demás. El algoritmo comienza con una agrupación aleatoria de las observaciones en un número de agrupaciones preestablecido y minimiza las distancias de cada observación a su centroide, mientras maximiza la distancia entre los centroides. El algoritmo recursivamente reagrupa las observaciones hasta que encuentra los centroides y grupos óptimos (centroides), así, una observación cualquiera pertenece a un grupo porque está más cerca del centroide que lo representa (Sinaga & Yang, 2020).

El número de perfiles se determina con el método *elbow*, el cual realiza el análisis de *k-means* con distintos números de perfiles y permite elegir el óptimo. El método es simple, calcula la distancia de todos los puntos a sus centroides y halla la distancia promedio, luego realiza *k-means* con un perfil más y repite el proceso. En cada iteración se reduce la distancia entre las observaciones y los centroides, hasta que la reducción es marginal. Usualmente se crea una gráfica que muestra el número de centroides en el eje x y la variación en el eje y, el punto en el que la línea comienza a decrecer de forma más lenta es el óptimo (Mengyao, 2020). En el siguiente ejemplo (Gráfica 1), se evidencia como a partir del tercer punto la variación entre puntos es significativamente menor a los dos anteriores.

En resumen, en este paso se obtiene el número de perfiles de unidades productivas que potencialmente podrían beneficiarse del programa y a cuál de estos grupos pertenece cada unidad productiva en la base de potenciales beneficiarios. El siguiente paso consiste en identificar cuál(es) de estos perfiles es el ideal para el programa.

Gráfica 1. Método Elbow (gráfica de referencia)



Fuente: Elaborado SDDE-ODEB

Nota: el punto rojo muestra el número de clusters óptimo.

3.4 PASO D. ELECCIÓN DEL GRUPO MÁS ADECUADO PARA RECIBIR EL PROGRAMA

Para elegir el perfil de las agrupaciones del paso C que más se beneficiaría del programa, se caracterizan los perfiles de acuerdo con las variables seleccionadas para el FAMD. De acuerdo con las estadísticas descriptivas básicas de cada grupo (media, mediana, desviación estándar, entre otras) y la relevancia de los criterios que reflejan estas variables, los miembros del equipo ODEB determinan cuál es el perfil óptimo del programa, es decir, se acerca más al tipo de unidad productiva que el programa busca beneficiar. Por esta razón, es importante que en la fase 1 de la MOOI haya una comprensión de la población objetivo del programa y se trabaje con la dirección misional para definir la relevancia de los criterios y las variables que los miden.

3.5 PASO E. ASIGNACIÓN DE PUNTAJE A CADA POTENCIAL BENEFICIARIO

Ahora, el puntaje de priorización de cada unidad productiva corresponde a la distancia de la unidad productiva hasta el centroide del perfil óptimo del programa. Estas distancias se normalizan y se transforman a una escala de 0 a 100 donde cien corresponde a una mayor prioridad, es decir, la convocatoria debe contactar de forma prioritaria a aquellas unidades con un indicador más alto. La ecuación abajo muestra cómo se normalizan las distancias: primero se transforman los valores a una escala de 0 a 1 donde 0 es la distancia más corta al centroide. Luego, para invertir el orden, se resta 1 menos el valor de esa normalización. Finalmente, se multiplica por 100 para facilidad de manejo. Así, valores más altos tienen mayor prioridad.

$$\text{Puntaje normalizado} = \left(1 - \frac{\text{Distancia al centroide de la UP} - \text{Distancia al centroide mínima}}{\text{Distancia al centroide máxima} - \text{Distancia al centroide mínima}} \right) * 100$$

A esta información, por supuesto, es necesario sumarle la de contacto antes de enviarla a la dirección implementadora.

4. POSIBLES EXTENSIONES DEL ANÁLISIS

El proceso de la MOOI permite conocer cuáles unidades productivas se verían más beneficiadas por un programa determinado, ese puntaje permite enfocar esfuerzos dentro del universo de unidades productivas a contactar y mejorar la efectividad de la implementación. Sin embargo, el proceso puede estar sujeto a mejoras, especialmente una vez se conozca el desempeño de los puntajes. En concreto, a continuación, se proponen dos posibles extensiones: dar insumos para definir la población objetivo y metas de beneficiarios de nuevos programas y estrategias para mejorar el proceso de la MOOI.

4.1 INSUMOS PARA NUEVOS PROGRAMAS DE LA SDDE

El proceso de la MOOI puede dar insumos para la definición de la población objetivo de los programas de la SDDE, así como las metas de número de unidades productivas beneficiarias. Específicamente, la metodología puede usarse antes de la definición de las características de un programa para hacer pruebas de sensibilidad que permitan estimar el número de unidades productivas a priorizar y sus características, de acuerdo con diferentes criterios de selección de las unidades productivas. De esta forma, se da mayor claridad sobre la posibilidad de alcanzar metas sobre el número y el tipo de beneficiarios que pueden atender los programas. Asimismo, se brindan elementos para modificar el diseño de programas para que eviten focalizar poblaciones pequeñas si este no es el objetivo, y se promueve la elección de criterios de selección y priorización medibles.

4.2 RETROALIMENTACIÓN DE LA MOOI

Es posible utilizar distintos indicadores de seguimiento y fuentes administrativas, como encuesta de satisfacción y seguimiento de resultados a las unidades productivas, para retroalimentar las bases de datos utilizadas en el proceso y ajustar y actualizar la metodología de la MOOI. Esto permitiría la mejora continua de la MOOI y la optimización en

el alcance de la oferta de la Secretaría. Si bien al inicio del proceso de priorización se definen algunas variables que se consideran óptimas para la priorización de unidades productivas, existen otras características no identificadas al momento de la planeación del programa que influyen en las probabilidades de acceso y de éxito en el programa por parte de las unidades productivas. Por ejemplo, es posible que para un programa específico no exista un criterio de ingresos mínimos, pero al momento de analizar cuáles unidades productivas se inscribieron al programa se encuentra que la mayoría tienen alrededor de \$2.500.000 de ingresos mensuales. Al identificar estas variables latentes se puede alimentar el proceso al priorizar primero aquellas unidades que no solo cumplen los criterios de exclusión y de priorización, sino también a las que tienen las características que aumentan la probabilidad de inscripción y de éxito.

5. BIBLIOGRAFÍA



Khan, M. E. E., Bouchard, G., Murphy, K. P., & Marlin, B. M. (2010). Variational bounds for mixed-data factor analysis. En J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems* (τ. 23). Obtenido de <https://proceedings.neurips.cc/paper/2010/file/2a084e55c87b1ebcdaad1f62fdbbac8e-Paper.pdf>

Mengyao, C. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. *Geoscience and Remote Sensing*. Vol. 3: 9-16. Obtenido de <https://www.clausiuspress.com/article/592.html>

Reddy, G et al. (2020). Analysis of Dimensionality Reduction Techniques on Big Data, in *IEEE Access*, vol. 8, pp. 54776-54788, 2020, Doi: 10.1109/ACCESS.2020.2980942.

Secretaría Distrital de Desarrollo Económico. (2020). Organización. Obtenido de <http://www.desarrolloeconomico.gov.co/transparencia/organizacion>

Sinaga, K and Yang, M. (2020). Unsupervised K-Means Clustering Algorithm, en *IEEE Access*, vol. 8, pp. 80716-80727, 2020, Doi: 10.1109/ACCESS.2020.2988796. Obtenido de https://www.researchgate.net/publication/340813602_Unsupervised_K-Means_Clustering_Algorithm

Siglas

DIAN	Dirección de Impuestos y Aduanas Nacionales
MOOI	Metodología para optimizar el alcance de la oferta institucional de la SDDE
NIT	Número de identificación tributaria
ODEB	Observatorio de Desarrollo Económico de Bogotá
RIT	Registro de Información Tributaria
RUT	Registro Único Tributario
SDDE	Secretaría Distrital de Desarrollo Económico
SUIM	Sistema unificado de información misional

Abreviaturas

FAMD	Factores para datos mixtos
MCA	Análisis de correspondencia múltiple
PCA	Análisis de componentes principales

Índice de tablas

Tabla 1.	Ejemplo matriz de consistencia	14
Tabla 2.	Matriz de confusión de la participación en los programas y la priorización de la MOOI	18

Índice de ilustraciones y gráficas

Ilustración 1.	Proceso de la MOOI por fases	12
Gráfica 1.	Método Elbow (gráfica de referencia)	26