

ENCUESTA SOBRE DISCRIMINACIÓN EN LA CIUDAD DE MÉXICO 2026

(EDIS CDMX, 2026)

Diseño muestral



Contenido

1. Objetivo del diseño muestral.....	4
2. Población objeto de estudio.....	4
3. Marco muestral.....	4
3.1.Fuentes de información y construcción del marco	4
3.1.1. Depuración y filtrado.....	5
3.1.2. Eliminación de registros con datos faltantes	6
3.2.Construcción de indicadores socioeconómicos a nivel de UPM	6
3.3.Construcción del Índice de NSE mediante Análisis de Componentes Principales (ACP) ..	8
3.4.Variables adicionales para garantizar representatividad de grupos poblacionales clave	11
3.5.Archivo de marco muestral	12
4. Esquema de muestreo.....	12
4.1.Primer etapa: Selección de conglomerados (UPM)	13
4.1.1. Unidad de muestreo.....	14
4.1.2. Marco muestral	15
4.1.3. Método de selección	15
4.1.4. Método de reemplazo.....	17
4.2.Segunda etapa: Selección de segmentos de área (USM).....	18
4.2.1. Unidad de muestreo.....	18
4.2.2. Marco muestral	19
4.2.3. Método de selección	19
4.2.4. Método de reemplazo.....	20
4.3.Tercera etapa: Selección de la vivienda (UTM)	21
4.3.1. Unidad de muestreo.....	21
4.3.2. Marco muestral	22
4.3.3. Método de selección	22
4.3.4. Método de reemplazo.....	22
4.4.Cuarta etapa: Selección del individuo (UQM)	23
4.4.1. Unidad de muestreo.....	23
4.4.2. Marco muestral	23
4.4.3. Método de selección	24
4.4.4. Método de reemplazo.....	24
4.5.Resumen de las etapas de muestreo	24
5. Probabilidad de selección	25



6. Validación de la muestra	26
6.1. Validación mediante estimación de indicadores	26
6.2. Validación de representatividad estructural de la muestra	27
7. Entregables técnicos asociados	29
ANEXOS	30
Anexo 1. Descripción de variables del marco muestral	31
Anexo 2. Descripción de variables de la base de muestra	37



1. Objetivo del diseño muestral

El objetivo del muestreo es seleccionar, de manera probabilística y representativa, una muestra de personas de 15 años o más, residentes permanentes en viviendas particulares dentro de las 16 alcaldías de la Ciudad de México, asegurando una cobertura equilibrada, conforme al marco muestral disponible. Este diseño de muestreo permitirá generar estimaciones confiables y comparables sobre la percepción y prevalencia de prácticas de discriminación en la capital.

El muestreo se alinea con los fines de la Encuesta sobre Discriminación en la Ciudad de México (EDIS CDMX), instrumento que COPRED realiza cada cuatro años con el objetivo de medir la situación de la discriminación en la ciudad y sus alcaldías. La información generada constituye un insumo estratégico para orientar el diseño, implementación y evaluación de políticas públicas, programas y servicios encaminados a promover la igualdad y la no discriminación. La EDIS CDMX suma ya tres ediciones previas (2013, 2017 y 2021), por lo que el presente diseño busca asegurar la consistencia metodológica necesaria para mantener la comparabilidad de los resultados a lo largo del tiempo.

2. Población objeto de estudio

Personas de 15 años cumplidos o más, residentes permanentes en viviendas particulares localizadas dentro del territorio de la Ciudad de México.

3. Marco muestral

El marco muestral utilizado para la selección de las Unidades Primarias de Muestreo (UPM) se construyó a partir de los microdatos del Censo de Población y Vivienda 2020 del INEGI. Dicho marco integra y organiza las áreas geográficas que conforman la Ciudad de México, tanto urbanas como rurales, con el fin de garantizar la cobertura de toda la población objetivo de estudio.

3.1. Fuentes de información y construcción del marco

El marco muestral se conformó mediante la integración de dos conjuntos de datos provenientes del Censo 2020:

- **Principales resultados por localidad a nivel nacional (ITER_NALCSV20)**

Fuente: INEGI – Microdatos del Censo 2020

Descarga: ITER_NAL_2020_csv.zip

- **Principales resultados por AGEB y manzana urbana de la CDMX
(RESAGEBURB_09_2020)**

Fuente: INEGI – Microdatos del Censo 2020

Descarga: RESAGEBURB_09_2020_csv.zip

La primera base contiene información sociodemográfica a nivel de localidad, mientras que la segunda presenta información a nivel de AGEB urbana para la Ciudad de México.

3.1.1. Depuración y filtrado

La base de localidades se filtró para conservar únicamente aquellas ubicadas dentro de la Ciudad de México que, además, correspondieran a zonas rurales. Para ello, se consideraron como UPM rurales las localidades con menos de 2,500 habitantes, utilizando la variable **TAM_LOC** con valores del 1 al 4, conforme a la siguiente clasificación oficial del INEGI:

Tabla 1: Tamaños de Localidades (Base ITER_NALCSV20)

Valor de la variable TAM_LOC	Tamaño de localidad
1	1 a 249 habitantes
2	250 a 499 habitantes
3	500 a 999 habitantes
4	1,000 a 2,499 habitantes

Fuente: Elaboración propia con información del INEGI (Censo de Población y Vivienda 2020, base ITER_NALCSV20).

Descripción: Esta tabla muestra cómo clasifica el INEGI a las localidades según el número de habitantes. Sirve para identificar cuáles se consideran rurales (menos de 2,500 habitantes).

La base de AGEBS no requirió filtrado, dado que únicamente incluye AGEBS urbanas de la Ciudad de México.

Una vez homologadas las variables coincidentes entre ambas bases, se realizó un proceso de unión por renglones para conformar una única base de datos que integra ambas áreas geográficas como unidades primarias de muestreo.

3.1.2. Eliminación de registros con datos faltantes

Se excluyeron del marco aquellas UPM con valores faltantes en la variable **P_15YMAS (Población de 15 años o más)**. Esto ocurre debido a que el INEGI aplica criterios de confidencialidad, omitiendo (enmascarando) datos cuando el tamaño poblacional es muy reducido (por ejemplo, en unidades geográficas que tienen menos de tres viviendas, con población mínima o zonas donde la divulgación permitiría inferir características individuales) o cuando corresponden a instalaciones especiales como centros penitenciarios, hospitales psiquiátricos u otros espacios no equiparables a viviendas particulares.

Tras este proceso, el marco muestral quedó conformado como se muestra en la tabla:

Tabla 2: Estructura del Marco Muestral (UPM Urbanas y Rurales)

Tipo de UPM	Población 15 años o más	% Población 15 años o más	Número de UPMs	% UPMs
AGEB Urbana	7,500,071	99.4%	2,406	84%
Localidad rural	46,913	0.6%	459	16%
Ciudad de México	7,546,984	100.0%	2,865	100%

Fuente: Elaboración propia con información del INEGI (Censo de Población y Vivienda 2020, bases ITER_NALCSV20 y RESAGEBURB_09_2020).

Descripción: La tabla muestra la distribución de las Unidades Primarias de Muestreo (UPM) según su tipo (urbanas o rurales), considerando tanto su número como la proporción de población de 15 años o más que representan.

Este marco garantiza la inclusión de todas las áreas geográficas habitadas y permite un diseño muestral probabilístico con cobertura total de la población objetivo. Es importante señalar que, si bien las UPM rurales representan el 16% del total de unidades, su peso demográfico es marginal: solo el 0.6% de la población de 15 años o más reside en ellas. Esta relación evidencia la alta concentración urbana de la Ciudad de México y, al mismo tiempo, refuerza la importancia de incorporar factores de probabilidad adecuados en el diseño, a fin de no sobrerrepresentar a la población rural en el proceso de selección.

3.2. Construcción de indicadores socioeconómicos a nivel de UPM

Dado que la información disponible en el marco muestral se encuentra a nivel de UPM (AGEB o localidad), se creó un **Índice de Nivel Socioeconómico (NSE) de la UPM** con el fin de clasificar las unidades en grupos socioeconómicos comparables. Este índice se inspira conceptualmente en el enfoque utilizado por la AMAI para su medición del nivel

socioeconómico a nivel de hogar¹; sin embargo, en este proyecto se construyó una versión adaptada con base en información estructural disponible en el Censo a nivel geográfico.

Para ello, se generaron seis variables de proporción que caracterizan la infraestructura, escolaridad y condiciones de bienestar de cada UPM. Las variables construidas se presentan a continuación:

Tabla 3: Variables socioeconómicas utilizadas para el índice de NSE

Numerador	Denominador	Variable creada	Definición
Viviendas particulares habitadas con 3 cuartos y más (VPH_3YMASC)	Total de viviendas particulares habitadas (TVIVPARHAB)	am2_cuartos	Proporción de viviendas con 3 cuartos o más en la UPM
Población de 18 años y más con educación posbásica (P18YM_PB)	Población de 18 años y más (P_18YMAS)	am3_educa	Proporción de población mayor de edad con educación media superior o superior en la UPM
Viviendas particulares habitadas con dos dormitorios y más (VPH_2YMASD)	Total de viviendas particulares habitadas (TVIVPARHAB)	am4_dormi	Proporción de viviendas con 2 dormitorios o más en la UPM
Viviendas particulares habitadas que disponen de excusado o sanitario (VPH_EXCSA)	Total de viviendas particulares habitadas (TVIVPARHAB)	am5_baño	Proporción de viviendas con baño en la UPM
Viviendas particulares habitadas que disponen de automóvil o camioneta (VPH_AUTOM)	Total de viviendas particulares habitadas (TVIVPARHAB)	am6_auto	Proporción de viviendas con automóvil en la UPM
Viviendas particulares habitadas que disponen de Internet (VPH_INTER)	Total de viviendas particulares habitadas (TVIVPARHAB)	am7_internet	Proporción de viviendas con conexión a Internet en la UPM

Fuente: Elaboración propia con información del INEGI (Censo de Población y Vivienda 2020, bases ITER_NALCSV20 y RESAGEBURB_09_2020).

Descripción: La tabla presenta las seis variables proporcionales utilizadas para la construcción del Índice de Nivel Socioeconómico (NSE) de las UPM: proporción de viviendas con tres cuartos o más, proporción de población adulta con educación posbásica, proporción de viviendas con dos dormitorios o más, proporción de viviendas con baño, proporción de viviendas con automóvil y proporción de viviendas con conexión a Internet.

¹ Para más información sobre la metodología oficial del Nivel Socioeconómico de la AMAI, consultar: Asociación Mexicana de Agencias de Investigación y Opinión Pública (AMAI). “NSE México”. Disponible en: <https://www.amai.org/NSE/>

Estas variables permiten describir a las UPM a partir de características del entorno en que residen las personas, incluyendo condiciones de las viviendas, el nivel educativo alcanzado por la población adulta y el acceso a servicios que influyen en la calidad de vida. Su uso en forma de proporciones facilita la comparación entre UPM con diferencias significativas en número de viviendas o habitantes y evita sesgos derivados del uso de valores absolutos. Además, son indicadores empleados para distinguir contextos socioeconómicos a nivel territorial, por lo que constituyen una base sólida para construir un índice que permita clasificar a las UPM según su nivel socioeconómico y asegurar que la muestra refleje la diversidad existente en la Ciudad de México.

3.3. Construcción del Índice de NSE mediante Análisis de Componentes Principales (ACP)

Para sintetizar la información de las seis proporciones generadas, se aplicó un Análisis de Componentes Principales (ACP). El ACP es una técnica estadística multivariada que permite reducir la dimensionalidad de un conjunto de variables correlacionadas, obteniendo una o más componentes que explican la mayor parte de la variabilidad en los datos originales mediante combinaciones lineales de esas variables².

En este caso, se seleccionó la **primer componente principal**, que explicó la mayor proporción de la varianza conjunta de las seis variables (68%), y se utilizó como un índice continuo de nivel socioeconómico de la UPM.

El valor del índice resultante osciló entre -10.9286 y $+3.9184$. Para facilitar su interpretación y su utilización en el diseño muestral, el índice se dividió en tres categorías mediante puntos de corte en terciles, de manera que cada estrato contuviera un número similar de UPM.

² Para una explicación general del Análisis de Componentes Principales (PCA) y sus usos, consultar: IBM. "Análisis de componentes principales (PCA): definición, usos y ejemplos". Disponible en: <https://www.ibm.com/es-es/think/topics/principal-component-analysis>

Tabla 4: Puntos de corte del índice de NSE

Percentil	Valor del índice
0%	-10.9286
33.3%	-0.5208
66.6%	+0.9866
100%	+3.9184

Fuente: Elaboración propia con base en el Índice de Nivel Socioeconómico estimado mediante Análisis de Componentes Principales (ACP) a partir de datos del INEGI (Censo de Población y Vivienda 2020).

Descripción: Se presentan los valores del índice de nivel socioeconómico utilizados para definir los tres grupos de NSE a partir de sus percentiles 33.3 y 66.6.

La clasificación resultante se asignó a la variable **clase** y presentó la siguiente distribución de UPM:

Tabla 5: Rangos de las categorías de clase NSE

Clase	Número de UPM	Rango en el índice
Categoría 1	950	-10.9286 a -0.5208
Categoría 2	949	-0.5208 a 0.9866
Categoría 3	950	0.9866 a 3.9184
Sin clasificación debido a datos faltantes	16	--

Fuente: Elaboración propia con base en el Índice de Nivel Socioeconómico estimado mediante Análisis de Componentes Principales (ACP) a partir de datos del INEGI (Censo de Población y Vivienda 2020).

Descripción: La tabla muestra la distribución de las Unidades Primarias de Muestreo (UPM) según las tres categorías del índice de nivel socioeconómico, determinadas por los rangos del valor del índice.

Para facilitar la interpretación del índice y visualizar con claridad las diferencias entre las categorías de nivel socioeconómico construidas, la siguiente tabla presenta los valores promedio de las seis variables empleadas para clasificar las UPM. Estos promedios permiten observar cómo varían, de una categoría a otra, aspectos relacionados con el tamaño y condiciones de la vivienda, el nivel educativo de la población adulta y el acceso a servicios que influyen en la calidad de vida. La comparación entre categorías muestra patrones consistentes que validan la lógica del índice, reflejando una mejora progresiva en estas condiciones conforme aumenta la categoría asignada a las UPM. Esta diferencia escalonada entre categorías confirma que representan grupos con características claramente diferenciadas y ordenadas de menor a mayor nivel socioeconómico.

Tabla 6: Indicadores del NSE por categoría de UPM (Promedio)

Clase	am2_ cuartos	am3_ educa	am4_ dormi	am5_ baño	am6_ auto	am7_ internet
Categoría 1	63%	39%	56%	92%	33%	46%
Categoría 2	83%	60%	70%	99%	41%	73%
Categoría 3	94%	81%	79%	100%	64%	87%

Fuente: Elaboración propia con base en el Índice de Nivel Socioeconómico estimado mediante Análisis de Componentes Principales (ACP) a partir de datos del INEGI (Censo de Población y Vivienda 2020).

Descripción: La tabla muestra los valores promedio de las seis variables utilizadas en la construcción del índice de nivel socioeconómico, agrupadas por categoría de UPM. Los resultados evidencian una mejora progresiva en las condiciones de vivienda, educación y acceso a servicios conforme aumenta el nivel socioeconómico.

Descripción general de categorías:

- **Categoría 1 (Bajo NSE):**

En promedio, las UPM que conforman esta categoría presentan las siguientes características: 63% de las viviendas cuentan con al menos 3 cuartos y 56% tienen 2 dormitorios o más. Solo 4 de cada 10 personas de 18 años y más tienen estudios de preparatoria o superiores. Una de cada 10 viviendas no dispone de baño y únicamente 33% cuenta con automóvil propio. Menos de la mitad de las viviendas (46%) tiene acceso a Internet.

- **Categoría 2 (Medio NSE):**

En promedio, las UPM que integran esta categoría presentan las siguientes características: 83% de las viviendas tienen 3 cuartos o más y 70% cuentan con 2 dormitorios o más. Solo 6 de cada 10 personas de 18 años y más tienen estudios de preparatoria o superiores. Prácticamente la totalidad de las viviendas (99%) dispone de baño. El 41% de las viviendas cuenta con automóvil propio y 73% tiene acceso a Internet.

- **Categoría 3 (Alto NSE):**

Esta categoría está conformada por las UPM que, en promedio, presentan mejores condiciones en las variables consideradas: 94% de las viviendas tienen 3 cuartos o más y 79% cuentan con al menos 2 dormitorios. La totalidad de las viviendas dispone de baño. Dos de cada tres viviendas (64%) tienen automóvil propio y 87% cuentan con acceso a Internet. Además, 8 de cada 10 personas de 18 años y más (81%) tienen estudios de preparatoria o superiores.

3.4. Variables adicionales para garantizar representatividad de grupos poblacionales clave

Con el fin de monitorear la capacidad del diseño muestral para generar estimaciones confiables sobre grupos poblacionales potencialmente vulnerables o de interés analítico para la EDIS CDMX, se construyeron variables de proporción asociadas con grupos relevantes en materia de igualdad y no discriminación.

Tabla 7: Variables adicionales para revisar representatividad de la muestra

Numerador	Denominador	Variable creada	Definición
Población de 3 años y más que habla alguna lengua indígena (P3YM_HLI)	Población de 3 años o más (P_3YMAS)	prop.indigena	Proporción de población hablante de lengua indígena
Población que se considera afromexicana o afrodescendiente (POB_AFRO)	Población total (POBTOT)	prop.afromexicanos	Proporción de población afromexicana o afrodescendiente
Población con discapacidad (PCON_DISC)	Población total (POBTOT)	prop.discapacidad	Proporción de población con alguna discapacidad
Población con limitación funcional (PCON_LIMI)	Población total (POBTOT)	prop.limitada	Proporción de población con alguna limitación para realizar actividades
Población con problema o condición mental (PCLIM_PMEN)	Población total (POBTOT)	prop.lim.mental	Proporción de población con condición o problema mental
Población de 15 años y más que no sabe leer o escribir (P15YM_AN)	Población de 15 años y más (P_15YMAS)	prop.analfabeta	Proporción de población analfabeta

Fuente: Elaboración propia con información del INEGI (Censo de Población y Vivienda 2020, bases ITER_NALCSV20 y RESAGEBURB_09_2020).

Descripción: La tabla presenta las variables de proporción utilizadas para monitorear la representatividad de grupos poblacionales de interés para la EDIS CDMX, como personas hablantes de lengua indígena, población afromexicana, personas con discapacidad, con limitaciones funcionales, con condición o problema mental y población analfabeta.

Estas variables se incorporaron al marco muestral con el propósito de contar con insumos que permitan evaluar, en etapas posteriores, si la muestra final reproduce de manera adecuada las proporciones que estos grupos representan dentro de la población



total. Su inclusión facilita comprobar que el diseño muestral garantiza la presencia de estos grupos en la selección de UPM y que los resultados de la encuesta permitirán estimar sus indicadores con niveles de precisión aceptables. Esto es especialmente relevante para la EDIS CDMX, dado que estos grupos forman parte de las poblaciones de interés para el análisis de igualdad y no discriminación.

3.5. Archivo de marco muestral

Como resultado del proceso descrito en esta sección, se obtuvo una base de datos para el marco muestral conformada por 2,865 UPM y 43 variables. Esta base incluye los identificadores de cada UPM, las variables en valores absolutos utilizadas para generar las proporciones previamente descritas, las variables empleadas en la construcción del índice de NSE, el valor del índice obtenido, la clasificación de las UPM por nivel socioeconómico y las variables adicionales destinadas a revisar la representatividad de la muestra. El archivo del marco se entrega bajo el nombre “**MARCO_MUESTRAL_EDIS_CDMX_2026.xlsx**”. Adicionalmente, se incorpora un descriptor detallado de todas las variables en el Anexo 1 de este documento.

4. Esquema de muestreo

Para asegurar una muestra estadísticamente representativa de la población objetivo, se utilizó un diseño muestral polietápico con probabilidad de área. Bajo este esquema, las unidades geográficas se seleccionan probabilísticamente en la primera etapa y, dentro de ellas, se eligen las viviendas e individuos a entrevistar. Este diseño es apropiado para encuestas en vivienda y permite generar estimaciones robustas para los dominios de interés definidos. El diseño combina las siguientes propiedades:

- **Probabilístico:** Todas las unidades de selección tienen una probabilidad conocida y distinta de cero de ser seleccionadas.
- **Estratificado:** Las UPM se agrupan en estratos mutuamente excluyentes con el fin de asegurar una adecuada cobertura territorial y garantizar que la muestra incluya diversidad geográfica.
- **Por conglomerados:** Las unidades primarias de muestreo son conjuntos de unidades muestrales.

• **Polietápico:** La unidad última de muestreo requiere un proceso de muestreo de cuatro etapas. La primera consiste en la selección de los conglomerados o Unidades Primarias de Muestreo (UPM), mediante un muestreo estratificado, sistemático y con probabilidad proporcional al tamaño. Las etapas posteriores combinan actividades de planeación previa y de operativo en campo: selección aleatoria de las manzanas dentro de cada UPM; selección sistemática de viviendas dentro de las manzanas seleccionadas; y, finalmente, selección aleatoria de la persona a entrevistar dentro de cada vivienda mediante un número aleatorio.

A continuación se describen detalladamente las cuatro etapas del muestreo, los métodos de selección, las reglas de reemplazo y las decisiones operativas adoptadas para asegurar la representatividad y la factibilidad de campo.

4.1. Primera etapa: Selección de conglomerados (UPM)

La primera etapa de muestreo se resume en el siguiente cuadro:

Tabla 8: Resumen de la primer etapa de muestreo

Unidad de muestreo:	AGEB urbanas (para zonas urbanas) y localidades con menos de 2,500 habitantes (para zonas rurales).
Marco Muestral:	Base construida a partir de microdatos del Censo 2020 (véase sección 3).
Método de selección:	Muestreo estratificado, sistemático y con probabilidad proporcional al tamaño.
Método de reemplazo de unidades:	Aleatorio, con matching de estrato, tamaño y características sociodemográficas similares a la UPM original (Muestras suplementarias).

Fuente: Elaboración propia, diseño muestral EDIS CDMX 2026.

Descripción: La tabla sintetiza los elementos que conforman la primera etapa del muestreo, en la cual se seleccionan las Unidades Primarias de Muestreo (UPM).

A continuación se describe con mayor detalle cada uno de los elementos de la tabla.

4.1.1. Unidad de muestreo

El marco muestral se compone de una lista de unidades geográficas las cuáles constituyen las Unidades Primarias de Muestreo (UPM) o conglomerados de la muestra. Los conglomerados son:

- Las AGEB (Áreas Geo-estadísticas Básicas) para las áreas urbanas, y
- Las localidades para las áreas rurales (localidades con menos de 2,500 habitantes)

Tabla 9: Distribución de UPM por alcaldía

Alcaldía	Población de 15 años y más	#Localidades rurales	#AGEBS urbanas	#UPM totales
Álvaro Obregón	622,258	5	198	203
Azcapotzalco	361,473	0	102	102
Benito Juárez	382,872	0	102	102
Coyoacán	522,992	0	156	156
Cuajimalpa de Morelos	172,233	15	31	46
Cuauhtémoc	462,349	0	153	153
Gustavo A. Madero	960,760	0	302	302
Iztacalco	335,205	0	108	108
Iztapalapa	1,468,256	0	455	455
La Magdalena Contreras	200,168	11	52	63
Miguel Hidalgo	346,035	0	120	120
Milpa Alta	116,254	203	42	245
Tláhuac	308,078	40	111	151
Tlalpan	569,982	89	206	295
Venustiano Carranza	365,207	0	147	147
Xochimilco	352,862	96	121	217
Ciudad de México	7,546,984	459	2406	2865

Fuente: Elaboración propia con información del INEGI (Censo de Población y Vivienda 2020, bases ITER_NALCSV20 y RESAGEBURB_09_2020).

Descripción: La tabla muestra la distribución de las Unidades Primarias de Muestreo (UPM), AGEB urbanas y localidades rurales, por alcaldía en la Ciudad de México.



4.1.2. Marco muestral

El marco muestral utilizado en esta encuesta fue especialmente construido para este proyecto, a partir del ensamble de información geoestadística proveniente de los microdatos del Censo de Población y Vivienda 2020 del INEGI. Este marco integra tanto AGEB urbanas como localidades rurales de la Ciudad de México, así como variables derivadas y clasificatorias que permiten optimizar el diseño muestral.

En primer término, se emplearon las bases de Principales Resultados por AGEB y Manzana Urbana y de Principales Resultados por Localidad, las cuales representan la información estadística más actualizada y completa disponible para la población de 15 años o más, con corte 2020. A partir de estos microdatos, se consolidó una base única de Unidades Primarias de Muestreo (UPM) sobre la cual se incorporaron variables sociodemográficas en valores absolutos y sus correspondientes proporciones para caracterizar a cada UPM.

Asimismo, como se detalló en la sección 3, se construyó un índice de nivel socioeconómico (NSE) a nivel de UPM mediante un Análisis de Componentes Principales (ACP), a partir del cual se generaron tres categorías de NSE. Esta clasificación se integró al marco con el propósito de facilitar la estratificación implícita y asegurar un adecuado balance socioeconómico en la selección de la muestra.

Finalmente, el marco se complementó con la cartografía geoespacial oficial del INEGI (shapefiles) para AGEB y localidades rurales, lo que permite la correcta localización y delimitación de las UPM durante la planeación y ejecución del operativo de campo. Este marco muestral constituye la base de referencia oficial para la selección de unidades en la primera etapa del diseño

4.1.3. Método de selección

Muestreo estratificado, sistemático y con probabilidad proporcional al tamaño.

Estratificación. Cada unidad primaria de muestreo en el marco se asigna a un estrato mutuamente excluyente. La estratificación del diseño consiste en una agrupación regional. La variable de estratificación contiene como categorías las 16 alcaldías en que se divide la Ciudad de México. El objetivo de esta estratificación es que la muestra contenga mayor dispersión geográfica.

Con base en el tamaño de muestra definido (Véase archivo: 1. Análisis de tamaño de muestra, error muestral y poder estadístico), se seleccionaron 441 UPM, distribuidas entre los estratos de acuerdo con el porcentaje que cada uno representa de la población total de 15 años y más. En la siguiente tabla se presenta el número de Unidades Primarias de Muestreo asignadas a cada estrato (alcaldía):

Tabla 10: Distribución por estrato de UPM en la muestra

Estrato	Alcaldía	Población de 15 años y más	% de Población de 15 años y más	#UPM en muestra
1	Álvaro Obregón	622,258	8.2%	36
2	Azcapotzalco	361,473	4.8%	21
3	Benito Juárez	382,872	5.1%	22
4	Coyoacán	522,992	6.9%	31
5	Cuajimalpa de Morelos	172,233	2.3%	10
6	Cuauhtémoc	462,349	6.1%	27
7	Gustavo A. Madero	960,760	12.7%	56
8	Iztacalco	335,205	4.4%	20
9	Iztapalapa	1,468,256	19.5%	86
10	La Magdalena Contreras	200,168	2.7%	12
11	Miguel Hidalgo	346,035	4.6%	20
12	Milpa Alta	116,254	1.5%	7
13	Tláhuac	308,078	4.1%	18
14	Tlalpan	569,982	7.6%	33
15	Venustiano Carranza	365,207	4.8%	21
16	Xochimilco	352,862	4.7%	21
Ciudad de México		7,546,984	100%	441

Fuente: Elaboración propia con información del INEGI (Censo de Población y Vivienda 2020, bases ITER_NALCSV20 y RESAGEBURB_09_2020).

Descripción: La tabla muestra el número de Unidades Primarias de Muestreo (UPM) que deberán ser seleccionadas por alcaldía en el muestreo estratificado.

Selección sistemática de conglomerados con probabilidad proporcional al tamaño. Dentro de cada estrato, las UPM se seleccionan mediante un muestreo sistemático con probabilidad proporcional al tamaño (PPT). El “tamaño” de cada UPM corresponde a la población de 15 años o más residente en esa área geográfica. Dado que el número de habitantes varía entre UPM, la PPT ajusta las probabilidades de selección de manera que todas las personas de la población objetivo tengan, en promedio, la misma probabilidad de ser seleccionadas, sin importar la UPM en la que vivan. Esto se logra asignando una mayor

probabilidad de selección a las UPM con mayor población y una menor probabilidad a las de menor población, equilibrando así la probabilidad final a nivel individuo.

El método PPT se combina con un muestreo sistemático. El muestreo sistemático consiste en seleccionar unidades a intervalos regulares (“salto sistemático”) a partir de un punto de inicio aleatorio. Este enfoque, además de ser eficiente operativamente, permite generar una estratificación implícita cuando el marco se ordena previamente por variables de interés. En esta encuesta, previo a seleccionar las UPM, el marco se ordena de forma secuencial por:

- estrato regional (alcaldías)
- categoría de nivel socioeconómico de la UPM, y
- valor del índice de NSE.

Este ordenamiento garantiza que el recorrido sistemático avance sobre las UPM con distintas características socioeconómicas de manera equilibrada, promoviendo que cada categoría de NSE esté adecuadamente representada en la muestra. Además, dicho ordenamiento contribuye a que, si se generaran diferentes muestras independientes, las UPM seleccionadas en posiciones equivalentes del listado presenten perfiles socioeconómicos similares, lo que fortalece la consistencia del diseño.

4.1.4. Método de reemplazo

Algunas UPM seleccionadas pueden resultar inaccesibles, ya sea por condiciones de seguridad o porque se ubican en zonas comerciales, con escasa presencia de viviendas habitadas. En estos casos, la UPM debe ser reemplazada por otra previamente seleccionada en una muestra independiente, diseñada con las mismas características que la muestra original.

Este esquema de reemplazo garantiza que la UPM sustituta conserve, en la medida de lo posible, las mismas condiciones de estratificación y una probabilidad de selección muy similar a la de la unidad original, gracias al uso del mismo método de selección sistemática y al mismo ordenamiento del marco.

Para este proyecto, además de la muestra original de 441 UPM, se generaron cuatro réplicas muestrales independientes, seleccionadas aleatoriamente bajo el mismo diseño. Dichas réplicas funcionan como reemplazos secuenciales: si se requiere sustituir una UPM,

se utiliza en primera instancia la que ocupa el mismo número de renglón en la muestra 1 de Reemplazo; si esta tampoco es viable (por seguridad, falta de viviendas o coincidencia con la unidad a sustituir), se procede con la muestra 2 de Reemplazo, y así sucesivamente hasta agotar las cuatro opciones.

El archivo “MUESTRA_EDIS_CDMX_2026.xlsx” contiene la muestra del proyecto, integrada por las 441 UPM originales y sus cuatro posibles sustitutas para cada una de ellas. Adicionalmente, el Anexo 2 de este documento incluye un descriptor detallado de todas las variables asociadas a dicho archivo.

4.2. Segunda etapa: Selección de segmentos de área (USM)

La segunda etapa de muestreo se resume en el siguiente cuadro:

Tabla 11: Resumen de la segunda etapa de muestreo

Unidad de muestreo:	En las AGEB, las manzanas son las unidades secundarias de muestreo (USM). En las localidades rurales, las unidades secundarias de muestreo consisten en conglomerados de viviendas.
Marco Muestral:	Catálogo de Manzanas INEGI 2020. Para identificación de la unidad de muestreo se utilizan diversos productos cartográficos
Método de selección:	Aleatorio simple en trabajo de gabinete
Método de reemplazo de unidades:	Aplicado en campo según método de espiral para conservar proximidad y características.

Fuente: Elaboración propia, diseño muestral EDIS CDMX 2026.

Descripción: La tabla sintetiza los elementos que conforman la segunda etapa del muestreo, en la cual se seleccionan las Unidades Secundarias de Muestreo (USM).

A continuación se describe con mayor detalle cada uno de los elementos de la tabla.

4.2.1. Unidad de muestreo

En las AGEB, las manzanas son las unidades secundarias de muestreo (USM). Una manzana consiste en un espacio geográfico delimitado por calles, avenidas, andadores y/o veredas. En su interior, se encuentran perfectamente ubicadas construcciones que son utilizadas para usos habitacionales, comerciales o industriales.



Si bien el número de localidades rurales seleccionadas en la muestra es reducido, para estos casos las USM se definen como conglomerados de viviendas, dada la ausencia de una delimitación formal de manzanas en este tipo de asentamientos.

4.2.2. Marco muestral

El marco muestral de la segunda etapa de muestreo corresponde al catálogo de manzanas urbanas del INEGI 2020, obtenido de la base de Principales Resultados por AGEB y Manzana Urbana de la Ciudad de México (RESAGEBURB_09_2020). Para la identificación precisa de las unidades secundarias de muestreo se utilizan diversos productos cartográficos, incluyendo archivos geoespaciales y planos por AGEB-manzana para áreas urbanas, así como planos de localidad rural en el caso de las zonas rurales.

4.2.3. Método de selección

Una vez seleccionadas las UPM (AGEB urbanas y localidades rurales), la estrategia para elegir las unidades secundarias de muestreo varía según el tipo de área.

En las AGEB urbanas, las manzanas constituyen las Unidades Secundarias de Muestreo (USM). Se seleccionan dos manzanas de forma aleatoria entre todas las que conforman la UPM; dichas manzanas funcionan como puntos de arranque del levantamiento. Es importante señalar que, en la práctica de campo, suele ser difícil completar el total de entrevistas asignadas dentro de una sola manzana. Por ello, el trabajo continúa en manzanas colindantes siguiendo un recorrido previamente establecido, hasta alcanzar el número requerido de entrevistas por punto de arranque. Este procedimiento operativo se detalla más adelante, en la etapa correspondiente a la selección de viviendas.

En las áreas rurales sin manzanas claramente delimitadas, el procedimiento de selección será el siguiente:

- I. La localidad seleccionada se divide conceptualmente en cuatro cuadrantes haciendo uso de referencias físicas (caminos, ríos, vías férreas) u otros límites identificables.
- II. Se definen dos cuadrantes principales y dos secundarios (determinado por número de punto muestral de trabajo). En los cuadrantes principales se buscará captar mayor número de entrevistas (por ejemplo, 3 por cuadrante principal y 2 por cada cuadrante secundario) hasta completar la cuota por UPM.



Imagen 1: Plano AGEB - manzana. Ejemplo de recorrido en manzanas seleccionadas (en amarillo).

Fuente: Elaboración propia con base en información cartográfica del INEGI y Google My Maps.

Descripción: Mapa que muestra un ejemplo de recorrido de manzanas seleccionadas dentro de un área geoadministrativa básica (AGEB) urbana, resaltadas en color amarillo.



Imagen 2: Plano localidad rural. Ejemplo de división en cuadrantes.

Fuente: Elaboración propia con base en información cartográfica del INEGI y Google My Maps.

Descripción: Mapa que ilustra la delimitación en cuadrantes de una localidad rural utilizada para la selección de viviendas dentro de la muestra.

4.2.4. Método de reemplazo

Si una manzana (USM) resulta inaccesible o no permite realizar entrevistas por seguridad u otras razones, se aplica el método de espiral como procedimiento de reemplazo en terreno:

El encuestador parte del punto geográfico central o de la esquina inicial de la manzana seleccionada y realiza una búsqueda en espiral hacia las manzanas contiguas (en sentido horario) hasta localizar una manzana elegible.

Este procedimiento prioriza la continuidad geográfica y la similitud contextual entre la manzana original y la de reemplazo, reduciendo el riesgo de introducir sesgos por sustitución.

4.3. Tercera etapa: Selección de la vivienda (UTM)

La tercera etapa de muestreo se resume en el siguiente cuadro:

Tabla 12: Resumen de la tercer etapa de muestreo

Unidad de muestreo:	Las viviendas son las unidades terciarias de muestreo (UTM).
Marco Muestral:	Conjunto de viviendas ubicadas en la USM seleccionada en la etapa anterior
Método de selección:	Sistemático con arranque aleatorio
Método de reemplazo de unidades:	Matching (vivienda al lado de la originalmente seleccionada).

Fuente: Elaboración propia, diseño muestral EDIS CDMX 2026.

Descripción: La tabla sintetiza los elementos que conforman la tercer etapa del muestreo, en la cual se seleccionan las Unidades Terciarias de Muestreo (UTM).

A continuación se describe con mayor detalle cada uno de los elementos de la tabla.

4.3.1. Unidad de muestreo

Viviendas particulares. Una vivienda se define como el espacio delimitado por paredes y techos de cualquier material de construcción donde viven, duermen, preparan alimentos, comen y se protegen de las inclemencias del tiempo una o más personas. La entrada debe ser independiente, es decir, que sus ocupantes puedan entrar o salir de ella sin pasar por el interior de otra vivienda.



4.3.2. Marco muestral

Conjunto de viviendas ubicadas en la USM seleccionada en la etapa anterior. Una vivienda puede no ser fácilmente vista cuando existen muros u otras barreras presentes (por ejemplo, viviendas que se encuentran al interior de conjuntos habitacionales); por lo tanto, se utilizan ciertos indicadores para determinar si existen viviendas ocultas en el área a entrevistar, como buzones, medidores de agua, gas o electricidad y entradas múltiples. Sin embargo, algunas viviendas con entradas ocultas o aquellas que se encuentran en calles privadas o edificios con entrada restringida dejar de ser consideradas unidades elegibles.

4.3.3. Método de selección

Una vez que los 2 puntos de arranque son identificados, en la tercera etapa se seleccionan cinco viviendas en cada punto. En el caso de las manzanas, el levantamiento comienza por la esquina noroeste de las mismas, recorriéndolas en el sentido de las manecillas del reloj. Para seleccionar la vivienda se realiza un salto sistemático de tres viviendas (este proceso se lleva a cabo al empezar en una nueva acera o al empezar en una nueva manzana). Una vez que se concluye la entrevista, el entrevistador deberá moverse a la siguiente acera. Sólo se realizará una entrevista por hogar.

El procedimiento para seleccionar una vivienda es el mismo para edificios de varios pisos. Si el edificio ocupa una manzana entera, el número máximo de entrevistas que se pueden aplicar es cuatro. Si el edificio se ubica sobre una de las aceras de la manzana, sólo se podrá aplicar una entrevista.

4.3.4. Método de reemplazo

Cuando la vivienda seleccionada no es elegible o no se logra contactar a ningún integrante del hogar, se procede a realizar el reemplazo visitando la vivienda contigua inmediata. Si esta tampoco resulta elegible o no se obtiene contacto, se continúa con la siguiente vivienda consecutiva hasta lograr una entrevista completa. Es importante señalar que el salto sistemático de tres viviendas únicamente se aplica después de haber obtenido una entrevista exitosa; es decir, el salto no avanza hasta que se haya completado una entrevista válida.

4.4. Cuarta etapa: Selección del individuo (UQM)

La cuarta etapa de muestreo se resume en el siguiente cuadro:

Tabla 13: Resumen de la cuarta etapa de muestreo

Unidad de muestreo:	Individuos dentro de las viviendas seleccionadas en la etapa anterior
Marco Muestral:	Integrantes del hogar cumplen con los requerimientos de elegibilidad
Método de selección:	Aleatorio (número aleatorio)
Método de reemplazo de unidades:	Matching (unidad sustituta que coincida con la unidad original en características de edad y sexo)

Fuente: Elaboración propia, diseño muestral EDIS CDMX 2026.

Descripción: La tabla sintetiza los elementos que conforman la cuarta etapa del muestreo, en la cual se seleccionan las Unidades Cuaternarias de Muestreo (UQM).

A continuación se describe con mayor detalle cada uno de los elementos de la tabla.

4.4.1. Unidad de muestreo

Individuos dentro de las viviendas seleccionadas en la etapa anterior.

4.4.2. Marco muestral

Cada vivienda seleccionada es visitada por un entrevistador. Una vez que se realizó el primer contacto con algún miembro del hogar, el encuestador realiza una breve entrevista exploratoria a alguno de los adultos que habitan en la vivienda para determinar si los integrantes del hogar cumplen con los requerimientos de elegibilidad de la encuesta. Los resultados de dicha entrevista quedaran registrados. Un hogar incluye a todas las personas que habitan la vivienda seleccionada. Los habitantes de una vivienda pueden ser una sola familia, dos o más familias que comparten la vivienda, una persona que vive sola o cualquier otro grupo de personas con o sin parentesco que comparten las instalaciones de la vivienda.

4.4.3. Método de selección

Si el adulto entrevistado informa que en la vivienda habitan uno o más individuos elegibles para la encuesta, comienza la cuarta etapa del proceso que consiste en la selección del individuo para cada subpoblación. Para la selección del individuo, el encuestador debe preguntar al individuo que abre la puerta de la vivienda seleccionada cuántos adultos habitan dicha vivienda. Posteriormente, el dispositivo electrónico (Tablet) proporciona un número aleatorio entre los adultos elegibles de la vivienda, que representa al individuo que debe ser entrevistado por el encuestador.

4.4.4. Método de reemplazo

En caso de que el individuo seleccionado no se encuentre disponible o no pueda completar la entrevista, se procede a conseguir un reemplazo en otra vivienda que cuente con las mismas características en términos de rango de edad y sexo del individuo original. Si bien el uso de reemplazos aleatorios es adecuado en las primeras etapas del proceso de muestreo, en esta última etapa el reemplazo mediante matching resulta más apropiado, ya que permite mantener la composición sociodemográfica prevista y reduce el riesgo de sesgos en la muestra.

4.5. Resumen de las etapas de muestreo

El diseño es polietápico y avanza desde unidades geográficas amplias hasta la selección del individuo. Las cuatro etapas son:

Unidades Primarias de Muestreo (UPM): Selección de AGEB urbanas y localidades rurales mediante un muestreo estratificado y sistemático con probabilidad proporcional al tamaño (PPT), asegurando distribución territorial y representatividad poblacional.

Unidades Secundarias de Muestreo (USM): En cada UPM seleccionada, se determinan dos puntos de arranque para el levantamiento. En zonas urbanas corresponden a manzanas elegidas aleatoriamente; en áreas rurales se utilizan segmentos de viviendas.

Unidades Terciarias de Muestreo (UTM): A partir de cada punto de arranque, se seleccionan viviendas mediante un recorrido sistemático establecido hasta completar las entrevistas asignadas.

- Unidades Cuaternarias de Muestreo (UQM): En cada vivienda elegible, se selecciona aleatoriamente a una persona de 15 años o más para ser entrevistada.

Tabla 14: Etapas, unidades y métodos del diseño muestral

Etapas	Unidad de muestreo	Método de selección	Cantidad seleccionada
Unidades Primarias de Muestreo (UPM)	AGEBS urbanas o localidades rurales	Muestreo estratificado y proporcional al tamaño	441 en total
Unidades Secundarias de Muestreo (USM)	Manzanas de punto de arranque	Aleatorio simple sin reemplazo	2 por UPM (882 en total)
Unidades Terciarias de Muestreo (UTM)	Hogares	Selección sistemática con método de espiral	5 por USM (4410 en total)
Unidades Cuaternarias de Muestreo (UQM)	Personas (15 años o más)	Aleatorio simple (igual probabilidad entre integrantes del hogar)	1 por UTM (4410 en total)

Fuente: Elaboración propia, diseño muestral EDIS CDMX 2026.

Descripción: La tabla resume las cuatro etapas del diseño muestral polietápico implementado para la EDIS CDMX 2026, detallando las unidades de muestreo, el método de selección y el número de unidades previstas en cada nivel. El esquema asegura representatividad y control de la variabilidad en cada etapa, desde la selección de UPM hasta la del individuo dentro del hogar.

5. Probabilidad de selección

La probabilidad de selección permite cuantificar la oportunidad que tiene cada unidad de ser incluida en la muestra en cada etapa del proceso de muestreo. En un diseño probabilístico, conocer estas probabilidades es indispensable para garantizar la representatividad de los resultados, así como para la construcción de los factores de expansión y los pesos muestrales utilizados en la estimación de indicadores. En particular, dado que la primera etapa del diseño se basa en un muestreo estratificado con probabilidad proporcional al tamaño (PPT), las UPM con mayor población de 15 años o más tienen una mayor probabilidad de ser seleccionadas, manteniendo al mismo tiempo la igualdad de probabilidad final a nivel individuo.

Probabilidad de selección para las UPM

Bajo el esquema descrito, la probabilidad de selección de la *i*-ésima Unidad Primaria de Muestreo en el estrato *h* se define como:

$$P_{ih} = \frac{M_{ih}}{\sum_{j=1}^{N_h} M_{jh}} \times n_h$$



Donde:

M_{ih} = Población de 15 años o más en la UPM i del estrato h .

N_h = Número total de UPM en el estrato h .

n_h = Número de UPM a seleccionar en el estrato h .

6. Validación de la muestra

Con el fin de garantizar que la muestra seleccionada mediante el diseño probabilístico propuesto refleje adecuadamente las características de la población objetivo, se realizó un ejercicio de validación. El propósito de esta validación es verificar que, al aplicar el diseño muestral y los factores de expansión correspondientes, la muestra permite obtener estimaciones consistentes y no sesgadas de indicadores relevantes para el estudio sobre discriminación en la Ciudad de México. Esta etapa es fundamental, ya que asegura que el diseño adoptado es capaz de reproducir adecuadamente la estructura poblacional y, con ello, generar estimaciones confiables durante el análisis de resultados.

Para dicha validación se utilizaron dos enfoques complementarios:

6.1. Validación mediante estimación de indicadores

Se estimaron, a partir de la muestra de UPM seleccionadas, diversos indicadores sociodemográficos y de condiciones de vivienda de interés para el estudio. Estas estimaciones se obtuvieron utilizando un estimador de razón con los pesos derivados del diseño estratificado por UPM. Los resultados se compararon contra los valores reales provenientes del Censo de Población y Vivienda 2020. La Tabla 16 presenta las estimaciones puntuales, los intervalos de confianza al 95% y la diferencia respecto al valor real.

Los resultados muestran que, para todos los indicadores evaluados, la estimación derivada del diseño se encuentra dentro del intervalo esperado y con errores muy pequeños, lo que demuestra que la muestra seleccionada permite reproducir adecuadamente los valores poblacionales conocidos.

Tabla 15: Estimación de indicadores bajo el diseño muestral

Indicador	Valor real en la CDMX	Estimación de razón	Intervalo inferior (2.5%)	Intervalo superior (97.5%)	error
% de viviendas con 3 cuartos o más	83.4%	83.7%	82.8%	84.6%	-0.32%
% de la población 18+ con educación posbásica	63.0%	63.0%	61.8%	64.2%	+0.03%
% de viviendas con 2 dormitorios o más	70.2%	70.4%	69.5%	71.4%	-0.24%
% de viviendas con excusado o sanitario	99.4%	99.4%	99.1%	99.7%	0.00%
% de viviendas con auto propio	46.8%	47.1%	45.8%	48.5%	-0.33%
% de viviendas con internet	75.6%	75.8%	74.8%	76.7%	-0.17%
% de población de 3+ que habla alguna lengua indígena	1.4%	1.5%	1.3%	1.6%	-0.05%
% de población afroamericana o afrodescendiente	2.0%	2.0%	1.8%	2.2%	+0.03%
% de población con alguna discapacidad	5.4%	5.3%	5.2%	5.5%	+0.04%
% de población con alguna limitación para realizar actividades	12.4%	12.5%	12.2%	12.9%	-0.12%
% de población con condición o problema mental	1.6%	1.6%	1.5%	1.6%	+0.02%
% de población de 15+ que no sabe leer o escribir	1.4%	1.5%	1.4%	1.5%	-0.04%

Fuente: Elaboración propia, diseño muestral EDIS CDMX 2026 y de las bases de Principales resultados del Censo de Población y Vivienda 2020 (INEGI).

Descripción: La tabla presenta la validación del diseño muestral mediante la estimación de indicadores clave de vivienda, educación y características poblacionales. Los valores obtenidos bajo el diseño se comparan con los datos censales, mostrando alta correspondencia y errores mínimos, lo que confirma la precisión y representatividad de la muestra seleccionada.

6.2. Validación de representatividad estructural de la muestra

Además de la estimación de indicadores, se verificó que la distribución de la muestra fuera consistente con la composición real de la población de 15 años o más por tipo de UPM (urbana vs. rural) y por categoría de nivel socioeconómico (NSE) asignado a las UPM.

Dado que en este diseño se seleccionan 10 entrevistas por UPM, la proporción de UPM en la muestra es equivalente a la proporción esperada de entrevistas en cada categoría poblacional.

En la Tabla 16 se muestra la comparación entre la distribución real de la población por tipo de UPM y la distribución de UPM en la muestra. Como se observa, la muestra mantiene una alta concordancia, con diferencias menores a ± 0.2 puntos porcentuales, lo cual confirma la adecuada representación territorial y poblacional entre áreas urbanas y rurales.

Tabla 16: Representatividad por tipo de UPM

TIPO DE UPM	Población 15 o más	% Población 15 o más	Cantidad en muestra	% de UPM en muestra	Diferencia
RURAL	46,913	0.6%	2	0.5%	+0.2%
URBANA	7,500,071	99.4%	439	99.5%	-0.2%
Total CDMX	7,546,984	100%	441	100%	

Fuente: Elaboración propia, diseño muestral EDIS CDMX 2026 y de las bases de Principales resultados del Censo de Población y Vivienda 2020 (INEGI).

Descripción: La tabla muestra la comparación entre la distribución real de la población de 15 años o más por tipo de Unidad Primaria de Muestreo (UPM) y la proporción correspondiente en la muestra seleccionada. Las diferencias observadas son mínimas, lo que evidencia que el diseño mantiene una representación equilibrada entre áreas urbanas y rurales en la Ciudad de México.

De forma similar, la Tabla 17 presenta la comparación entre la composición real de la población de 15 años o más por categoría de NSE y la distribución de UPM seleccionadas en cada estrato socioeconómico. Los resultados muestran diferencias mínimas, todas inferiores a ± 0.4 puntos porcentuales, lo que indica que el diseño garantiza una representación proporcional de los distintos niveles socioeconómicos definidos para la UPM.

Tabla 17: Representatividad por clase de NSE de la UPM

TIPO DE UPM	Población 15 o más	% Población 15 o más	Cantidad en muestra	% de UPM en muestra	Diferencia
Categoría 1	1,849,203	24.5%	108	24.5%	+0.0%
Categoría 2	3,238,378	42.9%	188	42.6%	+0.3%
Categoría 3	2,450,244	32.5%	145	32.9%	-0.4%

Sin clasificación debido a datos faltantes	9,159	0.1%	0	0.0%	+0.1%
Total CDMX	7,546,984	100.0%	441	100.0%	

Fuente: Elaboración propia, diseño muestral EDIS CDMX 2026 y de las bases de Principales resultados del Censo de Población y Vivienda 2020 (INEGI).

Descripción: La tabla compara la distribución real de la población de 15 años o más por categoría de nivel socioeconómico (NSE) con la proporción de Unidades Primarias de Muestreo (UPM) incluidas en la muestra. Las diferencias son mínimas, lo que confirma que el diseño muestral conserva la representación proporcional de los distintos estratos socioeconómicos definidos para la Ciudad de México.

En conjunto, los resultados de estos ejercicios corroboran que el diseño muestral propuesto asegura una alta correspondencia entre la estructura de la población real y la estructura de la muestra seleccionada, tanto en términos de estimaciones de indicadores como en la composición socioeconómica y territorial. Esto valida la idoneidad del diseño para generar estimaciones precisas, fiables y representativas en el estudio EDIS 2026.

7. Entregables técnicos asociados

Además del presente documento de diseño muestral, se entregan los insumos necesarios para la implementación del estudio y la trazabilidad del proceso de selección. Los entregables incluyen:

- I. **El marco muestral** utilizado para la selección de las UPM, con las variables y clasificaciones detalladas.
- II. **La base de datos de la muestra seleccionada**, que integra las 441 UPM originales junto con sus cuatro reemplazos asociados en un solo archivo.
- III. **El archivo geoespacial** en formato .kml correspondiente a las UPM de la muestra original, para su consulta, visualización y uso en herramientas de mapeo.



ANEXOS

Anexo 1. Descripción de variables del marco muestral

La siguiente tabla presenta las variables incluidas en la base del marco muestral.

Variable	Tipo de variable	Descripción
IDENTIFICADOR	Numérica	Identificador único y numérico para las UPM.
ID_UPM	Carácter	Identificador de UPM. Combinación de estado + alcaldía + localidad + AGEB.
ENTIDAD	Carácter	Código de estado.
NOM_ENT	Carácter	Nombre del estado.
MUN	Carácter	Código de alcaldía.
NOM_MUN	Carácter	Nombre de alcaldía.
LOC	Carácter	Código de localidad.
NOM_LOC	Carácter	Nombre de la localidad.
AGEB	Carácter	Código de AGEB.
UPM	Carácter	Código de UPM. En UPM urbanas es el código de AGEB. En UPM rurales es el código de localidad.
TVIVPARHAB	Numérica	Viviendas particulares habitadas de cualquier clase: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio; local no construido para habitación, vivienda móvil; refugio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.



POBTOT	Numérica	Personas, nacionales o extranjeras que residen habitualmente en el ámbito geográfico de referencia correspondiente.
P_18YMAS	Numérica	Personas de 18 a 130 años de edad.
P_15YMAS	Numérica	Personas de 15 a 130 años de edad.
P_3YMAS	Numérica	Personas de 3 a 130 años de edad.
VPH_3YMASC	Numérica	Viviendas particulares habitadas que tienen entre 3 y 25 cuartos. Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
P18YM_PB	Numérica	Personas de 18 a 130 años de edad que tienen como máxima escolaridad algún grado aprobado en preparatoria o bachillerato; normal básica; estudios técnicos o comerciales con secundaria terminada; estudios técnicos o comerciales con preparatoria terminada; normal de licenciatura; licenciatura o profesional; especialidad; maestría o doctorado. Incluye a las personas que no especificaron los grados aprobados en los niveles señalados.



VPH_2YMASD	Numérica	Viviendas particulares habitadas que usan para dormir entre 2 y 25 cuartos. Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
VPH_EXCSA	Numérica	Viviendas particulares habitadas que tienen taza de baño (excusado o sanitario). Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
VPH_AUTOM	Numérica	Viviendas particulares habitadas que tienen automóvil o camioneta. Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.



VPH_INTER	Numérica	Viviendas particulares habitadas que tienen Internet. Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
am2_cuartos	Proporción	Proporción de viviendas con 3 cuartos o más en la UPM.
am3_educa	Proporción	Proporción de población mayor de edad con educación media superior o superior en la UPM.
am4_dormi	Proporción	Proporción de viviendas con 2 dormitorios o más en la UPM.
am5_baño	Proporción	Proporción de viviendas con baño en la UPM.
am6_auto	Proporción	Proporción de viviendas con automóvil en la UPM.
am7_internet	Proporción	Proporción de viviendas con conexión a Internet en la UPM.
clase	Carácter	Categoría de nivel socioeconómico a la que pertenece la UPM, construida a partir del índice de NSE.
indice	Numérica	Índice de Nivel Socioeconómico (NSE) de la UPM
P3YM_HLI	Numérica	Personas de 3 a 130 años de edad que hablan alguna lengua indígena.

POB_AFRO	Numérica	Personas que se consideran afromexicanos o afrodescendientes.
PCON_DISC	Numérica	Personas que realizan con mucha dificultad o no pueden hacer al menos una de las siguientes actividades: ver, aun usando lentes; oír, aun usando aparato auditivo; caminar, subir o bajar; recordar o concentrarse; bañarse, vestirse o comer; hablar o comunicarse.
PCON_LIMI	Numérica	Personas que realizan con poca dificultad al menos una de las siguientes actividades: ver, aun usando lentes; oír, aun usando aparato auditivo; caminar, subir o bajar; recordar o concentrarse; bañarse, vestirse o comer; hablar o comunicarse.
PCLIM_PMEN	Numérica	Personas con algún problema o condición mental, como las relacionadas con el trastorno de la conciencia, retraso mental y las alteraciones de la conducta del individuo con otras personas en su entorno social. Incluye a la población que declaró realizar actividades cotidianas con mucha dificultad o no poder hacerlas o las hace con poca dificultad.
P15YM_AN	Numérica	Personas de 15 a 130 años de edad que no saben leer y escribir un recado.
prop.indigena	Proporción	Proporción de población hablante de lengua indígena.
prop.afromexicanos	Proporción	Proporción de población afromexicana o afrodescendiente.
prop.discapacidad	Proporción	Proporción de población con alguna discapacidad.



prop.limitada	Proporción	Proporción de población con alguna limitación para realizar actividades.
prop.lim.mental	Proporción	Proporción de población con condición o problema mental.
prop.analfabeta	Proporción	Proporción de población analfabeta.
TIPO	Carácter	Tipo de unidad primaria de muestreo, urbano o rural.
estrato	Numérica	Numero de estrato. Corresponde a la alcaldía.

Anexo 2. Descripción de variables de la base de muestra

La siguiente tabla presenta las variables incluidas en la base de la muestra, con su tipo y descripción correspondiente.

Variable	Tipo de variable	Descripción
ORDEN	Numérica	Identifica la posición secuencial que ocupa la UPM dentro del listado utilizado para el muestreo sistemático. Este número permite ubicar el orden en el que fue seleccionada y sirve como referencia para aplicar reemplazos.
MUESTRA	Carácter	Especifica la condición de selección de la UPM, indicando si pertenece a la muestra original o a alguno de los grupos de reemplazo (Reemplazo 1, Reemplazo 2, Reemplazo 3 o Reemplazo 4).
IDENTIFICADOR	Numérica	Identificador único y numérico para las UPM.
ID_UPM	Carácter	Identificador de UPM. Combinación de estado + alcaldía + localidad + AGEB.
ENTIDAD	Carácter	Código de estado.
NOM_ENT	Carácter	Nombre del estado.
MUN	Carácter	Código de alcaldía.
NOM_MUN	Carácter	Nombre de alcaldía.
LOC	Carácter	Código de localidad.
NOM_LOC	Carácter	Nombre de la localidad.
AGEB	Carácter	Código de AGEB.

UPM	Carácter	Código de UPM. En UPM urbanas es el código de AGEB. En UPM rurales es el código de localidad.
TVIVPARHAB	Numérica	Viviendas particulares habitadas de cualquier clase: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio; local no construido para habitación, vivienda móvil; refugio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
POBTOT	Numérica	Personas, nacionales o extranjeras que residen habitualmente en el ámbito geográfico de referencia correspondiente.
P_18YMAS	Numérica	Personas de 18 a 130 años de edad.
P_15YMAS	Numérica	Personas de 15 a 130 años de edad.
P_3YMAS	Numérica	Personas de 3 a 130 años de edad.
VPH_3YMASC	Numérica	Viviendas particulares habitadas que tienen entre 3 y 25 cuartos. Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
P18YM_PB	Numérica	Personas de 18 a 130 años de edad que tienen como máxima escolaridad algún grado aprobado en



		preparatoria o bachillerato; normal básica; estudios técnicos o comerciales con secundaria terminada; estudios técnicos o comerciales con preparatoria terminada; normal de licenciatura; licenciatura o profesional; especialidad; maestría o doctorado. Incluye a las personas que no especificaron los grados aprobados en los niveles señalados.
VPH_2YMASD	Numérica	Viviendas particulares habitadas que usan para dormir entre 2 y 25 cuartos. Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
VPH_EXCSA	Numérica	Viviendas particulares habitadas que tienen taza de baño (excusado o sanitario). Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.

VPH_AUTOM	Numérica	Viviendas particulares habitadas que tienen automóvil o camioneta. Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
VPH_INTER	Numérica	Viviendas particulares habitadas que tienen Internet. Comprende las viviendas particulares para las que se captaron las características de la vivienda, clasificadas como: casa única en el terreno; casa que comparte terreno con otra(s); casa dúplex; departamento en edificio; vivienda en vecindad o cuartería; vivienda en cuarto de azotea de un edificio y no especificado de vivienda particular. Incluye a las viviendas particulares sin información de ocupantes.
am2_cuartos	Proporción	Proporción de viviendas con 3 cuartos o más en la UPM.
am3_educa	Proporción	Proporción de población mayor de edad con educación media superior o superior en la UPM.
am4_dormi	Proporción	Proporción de viviendas con 2 dormitorios o más en la UPM.
am5_baño	Proporción	Proporción de viviendas con baño en la UPM.
am6_auto	Proporción	Proporción de viviendas con automóvil en la UPM.

am7_internet	Proporción	Proporción de viviendas con conexión a Internet en la UPM.
clase	Carácter	Categoría de nivel socioeconómico a la que pertenece la UPM, construida a partir del índice de NSE.
indice	Numérica	Índice de Nivel Socioeconómico (NSE) de la UPM.
P3YM_HLI	Numérica	Personas de 3 a 130 años de edad que hablan alguna lengua indígena.
POB_AFRO	Numérica	Personas que se consideran afromexicanos o afrodescendientes.
PCON_DISC	Numérica	Personas que realizan con mucha dificultad o no pueden hacer al menos una de las siguientes actividades: ver, aun usando lentes; oír, aun usando aparato auditivo; caminar, subir o bajar; recordar o concentrarse; bañarse, vestirse o comer; hablar o comunicarse.
PCON_LIMI	Numérica	Personas que realizan con poca dificultad al menos una de las siguientes actividades: ver, aun usando lentes; oír, aun usando aparato auditivo; caminar, subir o bajar; recordar o concentrarse; bañarse, vestirse o comer; hablar o comunicarse.
PCLIM_PMEN	Numérica	Personas con algún problema o condición mental, como las relacionadas con el trastorno de la conciencia, retraso mental y las alteraciones de la conducta del individuo con otras personas en su entorno social. Incluye a la población que declaró realizar actividades cotidianas con mucha dificultad o no poder hacerlas o las hace con poca dificultad.



P15YM_AN	Numérica	Personas de 15 a 130 años de edad que no saben leer y escribir un recado.
prop.indigena	Proporción	Proporción de población hablante de lengua indígena.
prop.afromexicanos	Proporción	Proporción de población afromexicana o afrodescendiente.
prop.discapacidad	Proporción	Proporción de población con alguna discapacidad.
prop.limitada	Proporción	Proporción de población con alguna limitación para realizar actividades.
prop.lim.mental	Proporción	Proporción de población con condición o problema mental.
prop.analfabeta	Proporción	Proporción de población analfabeta.
TIPO	Carácter	Tipo de unidad primaria de muestreo, urbano o rural.
estrato	Numérica	Numero de estrato. Corresponde a la alcaldía.



Bibliografía

1. Asociación Mexicana de Agencias de Investigación y Opinión Pública (AMAI). **NSE México**. Disponible en: <https://www.amai.org/NSE/>.
2. IBM. **Análisis de componentes principales (PCA): definición, usos y ejemplos**. Disponible en: <https://www.ibm.com/es-es/think/topics/principal-component-analysis>.
3. Instituto Nacional de Estadística y Geografía (INEGI). **Principales resultados por localidad a nivel nacional (ITER_NALCSV20)**. Microdatos del Censo 2020.
4. Instituto Nacional de Estadística y Geografía (INEGI). **Principales resultados por AGEB y manzana urbana de la CDMX (RESAGEBURB_09_2020)**. Microdatos del Censo 2020.
5. Instituto Nacional de Estadística y Geografía (INEGI). **Marco Geoestadístico 2020**. Datos públicos disponibles en: <https://www.inegi.org.mx/temas/mg/>