

Escuela de Negocios
Tipo de documento: Tesis de maestría



Master in Management + Analytics

Optimización de la asignación de votantes a establecimientos electorales

Autoría: Liserra, Ángeles

Año: 2025

¿Cómo citar este trabajo?

Liserra, A. (2025) "*Optimización de la asignación de votantes a establecimientos electorales*". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella <https://repositorio.utdt.edu/handle/20.500.13098/13741>

El presente documento se encuentra alojado en el **Repositorio Digital de la Universidad Torcuato Di Tella** bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional
Dirección: <https://repositorio.utdt.edu>



**UNIVERSIDAD
TORCUATO DI TELLA**

MASTER IN MANAGEMENT + ANALYTICS

**OPTIMIZACIÓN DE LA ASIGNACIÓN DE
VOTANTES A ESTABLECIMIENTOS
ELECTORALES**

TESIS

ANGELES LISERRA

2025

Tutor: Javier Marengo

A mis padres, por su amor y su guía. Por haberme dado y enseñado todo.

A mis hermanos, por sostenerme y estar siempre al lado del camino.

A Cassandra, por convencerme de que todo lo puedo lograr.

A Lucas, por ser pilar fundamental en cada uno de mis pasos.

A Pilar, por acompañarme e impulsarme.

A Sofía, por su incondicionalidad.

Resumen

Esta tesis aborda el método de asignación de votantes a establecimientos electorales con un enfoque innovador que busca minimizar la distancia total entre domicilios y centros electorales designados como objetivo para que la población en su totalidad recorra menores distancias. A su vez, se incluyen restricciones de impacto, en las que se da prioridad en términos de distancia a las personas con discapacidad y mayores de 60 años, como una muestra de la capacidad tecnológica para mejorar la equidad social. Para abordar el problema en cuestión, se utilizan métodos de programación lineal, estadística aplicada, programación y herramientas computacionales. En este sentido, se acerca una propuesta que incluye desde la descripción de la problemática hasta el diseño e implementación de un modelo que sea viable y eficiente, disponible para ser aplicado en cualquier distrito electoral.

Los resultados acompañan la justificación de un modelo que mejora la situación en lo que respecta a la asignación de votantes a sus establecimientos electorales. Así, se pretende mejorar la experiencia de los electores y propiciar un clima que favorezca la participación de los mismos. Tras un recorrido por el diagnóstico, la propuesta, la comparación con la metodología actual y la justificación sobre el por qué es necesario incursionar en estrategias más eficientes, el modelo queda disponible para su aplicación, en un afán de aportar tecnológica y metodológicamente a la resolución de una problemática cíclica y con claras oportunidades de mejora.

Palabras clave: asignación de electores, optimización, equidad, participación ciudadana, programación lineal.

Abstract

This thesis addresses the method of assigning voters to polling stations using an innovative approach. It seeks to minimize the total distance between homes and designated polling stations so that the entire population is forced to travel less. Impact restrictions are also included, giving priority in terms of distance to people with disabilities and those over 60 years of age, as an example of the technological capacity to improve social equity.

To address the problem at hand, integer linear programming methods, applied statistics, programming, and computational tools are used. In this regard, a proposal is presented that ranges from the description of the problem to the design and implementation of a viable and efficient model, available for application in any electoral district.

The results support the justification of a model that improves the situation regarding the assignment of voters to their polling stations. Thus, the aim is to improve the voter experience and foster a climate conducive to voter participation. This work seeks to immerse the reader in the problem, diagnosis, proposal, comparison with current methodology, and justification for why it is necessary to explore more efficient strategies. The model is available for application, in an effort to contribute technologically and methodologically to the resolution of a cyclical problem with clear opportunities for improvement.

Keywords: voter allocation, optimization, equity, citizen participation, linear programming.

Índice

Índice.....	5
Capítulo 1: Introducción.....	6
1.1. Contenido de la tesis.....	6
1.2. Motivación.....	9
1.3. Problema y alcance.....	11
1.4. Revisión de la literatura.....	13
Capítulo 2: Datos del problema.....	17
2.1. Definiciones del sistema electoral argentino.....	17
2.2. Circuito electoral.....	18
2.3. Establecimientos.....	19
2.4. Población electoral.....	20
Capítulo 3: Metodología.....	28
3.1. Metodología actual.....	28
3.2. Metodología propuesta.....	29
3.3. Aprendizajes para el modelo.....	35
Capítulo 4: Resultados.....	41
4.1. Resultados modelo actual.....	41
4.2. Resultados nuevo modelo.....	42
4.3. Medición de éxito.....	44
Capítulo 5: Conclusiones.....	48
5.1. Comparación de resultados.....	48
5.2. Trabajos futuros.....	51
Referencias.....	53
Apéndice A.....	55
1. Código para la creación de la población sintética.....	55
2. Código para el armado y resolución del modelo de asignación.....	60
3. Código para representar el modelo de asignación actual.....	64
4. Código para revisar resultados de asignación.....	66
Apéndice B.....	68
1. Código para asignar mesas dentro de los establecimientos.....	68
2. Cálculo estadístico para decidir el tamaño de la muestra.....	70
3. Código para seleccionar de manera aleatoria a la muestra de la encuesta.....	72

Capítulo 1: Introducción

1.1. Contenido de la tesis

El presente trabajo busca visibilizar y abordar una cuestión popular que, de manera cíclica, toma relevancia social. En particular, esta tesis se centra en justificar y proponer la implementación de un modelo de asignación de votantes a establecimientos electorales que resulta más próspero respecto a la metodología actual.

En el primer capítulo, se realiza una introducción al problema donde se hace hincapié en la motivación que, conforme a lo expuesto, tiene dos frentes: un aporte para acompañar el aprovechamiento de los estados respecto a los avances tecnológicos y la intención de generar una mejora en la satisfacción social en los comicios electorales. Como consecuencia de esto último, se espera que mejore la participación electoral por propiciar, con estos cambios, una mayor predisposición de la población a acercarse a votar.

Además, se describe el problema a tratar y el alcance de la solución propuesta. En este sentido, se acota a un circuito electoral en particular para demostrar la aplicabilidad y los resultados, siendo extrapolable el modelo a cualquier circuito. El primer capítulo provee también una revisión de la literatura relacionada, recorriendo trabajos que se han adentrado en el problema en cuestión aunque con distintos enfoques y abordajes.

En el segundo capítulo se presentan algunas definiciones de interés, que dan el marco al desarrollo de este trabajo. Se presentan los datos a utilizar: circuito electoral, establecimientos aptos para votar y personas con derecho a ejercer su voto. Tanto el circuito electoral como los establecimientos que forman parte del mismo, son datos

provistos por el Gobierno de la Ciudad de Buenos Aires. Necesaria también la población electoral aunque no disponible, se decide desarrollar e implementar una que replique la real, considerando sus características. En este capítulo también se proveen los datos geolocalizados, para darle visibilidad al lector e introducirlo gráficamente en el problema a abordar.

El tercer capítulo es estrictamente metodológico. Recorre por un lado la metodología de asignación actual, describiéndola objetivamente para luego darle paso al desarrollo de la metodología propuesta, que se basa en programación lineal. Allí se busca resolver dónde asignar a cada votante, conforme se minimicen las distancias totales a recorrer, sujeto a una serie de restricciones que se plantean por considerarlas equitativas o de interés social.

El cuarto capítulo se enfoca en los resultados y resulta crítico para demostrar por qué el modelo propuesto es mejor que el actual. En primer lugar, se replica y aplica el modelo vigente a la población sintética creada, para obtener las asignaciones y así, métricas resultantes como la distancia total recorrida, la distancia recorrida promedio por grupo y el porcentaje de grupos convivientes que votan juntos. Luego, se ejecuta el modelo propuesto sobre la misma población artificial que la que se sometió al modelo actual, considerando también los mismos establecimientos electorales para luego proceder a una simétrica comparación.

El quinto capítulo presenta las conclusiones del modelo, basándose en la comparación y análisis de los resultados obtenidos anteriormente, tras la aplicación de un modelo y del otro. El objetivo de este capítulo es mostrar las mejoras resultantes de una metodología que utiliza la programación lineal como herramienta para una asignación con

consideraciones más complejas y equitativas que las que se tienen en cuenta en la actualidad.

También, el quinto capítulo cuenta con una sección enfocada en la importancia de establecer indicadores que permitan evaluar el éxito del nuevo modelo, enfocándose principalmente en tres frentes: la participación ciudadana, las distancias recorridas por los grupos prioritarios en comparación con el resto y la satisfacción ciudadana.

En los apéndices de la tesis se pueden encontrar los códigos que se utilizaron para crear la población sintética, los modelos aplicados e incluso los cálculos estadísticos e información complementaria para facilitar la aplicación del modelo.

De esta manera, se busca dar marco y abordar, de manera teórica y práctica, un fenómeno de interés social y político, utilizando técnicas de programación lineal entera, estadística, programación en Python y herramientas de visualización de datos.

Así, se pretende brindar el diagnóstico de una oportunidad de mejora junto con una propuesta de resolución y se brindan las herramientas para aplicar y evaluar la nueva forma de asignar votantes a sus establecimientos electorales. Queda entonces a disposición este trabajo para su aplicación y aprovechamiento, con la intención de aportar académicamente a los estados en materia de distribución y organización electoral.

1.2. Motivación

En contextos democráticos, existen diversas formas de pronunciar participación ciudadana. Sin embargo y casi exclusivamente, “La ciudadanía asocia su participación con la materialización de ejercer su derecho al voto para elegir al mandatario de su nación” (Lopez, 2020, 31). Puede que por diversos motivos, como la simbología del evento electoral, la obligatoriedad en muchos casos, la característica grupal del día electoral o la cobertura del hecho lo que hace que, sin duda, los ciudadanos interpretan al voto como el mayor -o incluso el único- método de participación en lo que refiere a la elección y gestión de gobierno. En función de lo expuesto anteriormente es que resulta motivador adentrarse en ese método de participación que se ha posicionado como el “favorito” de los ciudadanos.

Ahora sobre el acto electoral en particular, es importante introducir que el mismo es individual y grupal. Las personas, en el hecho, dedican su único voto posible a dar un mensaje. Como es evidente, la suma de las individualidades hacen a los resultados. Por eso, es necesario ver al acto electoral justamente como lo que es: la suma de individualidades. Ahí, está la clave sobre una de las principales razones por las que es necesario enfocarse en mejorar y optimizar la experiencia de la sociedad en los eventos electorales basándose metodológica y simbólicamente, en las situaciones individuales. A esto le daremos sentido y marco teórico más adelante, cuando explicitemos el modelo propuesto donde veremos que tanto lo que se busca optimizar, como los innegociables del modelo, están fundados en situaciones individuales, conllevando a una mejora grupal.

Con el marco descrito, la motivación de esta tesis tiene principalmente dos frentes. El primero de ellos se constituye por el reconocimiento de un problema social identificado. Las elecciones son un evento importante para los países y los ciudadanos. La organización

electoral en Argentina suele ser una discusión popular porque las personas reconocen implícitamente que hay oportunidades de mejora. Es común escuchar argumentos de votantes sobre que tienen establecimientos más cercanos a los que podrían haber sido asignados. Las personas de grupos prioritarios no tienen, justamente, prioridad en las asignaciones. Los votantes en muy pocas ocasiones están asignados al mismo lugar que sus convivientes y eso les conlleva como mínimo, problemas logísticos. Lo mencionado y otra serie de cuestiones, por más subjetivas que resulten, hacen al clima electoral y a la predisposición de asistir a votar. Además, contextualmente, la metodología no es clara en la asignación ni muy transparente de cara a los votantes. En este sentido, la mejora en la satisfacción ciudadana representa el primer punto de motivación de esta tesis.

En segundo lugar, hay una gran motivación en lo que refiere a la oportunidad de realizar un aporte académico a los sistemas gubernamentales. La innovación, en contexto de desarrollo tecnológico, resulta vital para los estados. Conforme a lo expuesto y a lo que se desarrollará en el transcurso de este trabajo, se considera que el sistema metodológico propuesto es más próspero, en términos generales para la sociedad, que el actual. Aquí radica un motivo realmente importante que fundamenta el desarrollo de esta tesis. Si bien se reconoce que, el problema en cuestión, es uno de los miles que tienen que enfrentar los estados, es vital el aporte tecnológico agregado en todos los frentes. Esto considerando que “El cambio tecnológico solamente es uno de los motores de la prosperidad, pero quizá el más crítico. Los países que no aprovecharon las nuevas tecnologías tampoco se beneficiaron de otros motores de prosperidad” (Acemoglu & Robinson, 2008, 181). En este sentido es que se intenta, mediante esta tesis, dar un aporte a las instituciones que organizan los comicios electorales basado en avances tecnológicos principalmente enfocados en métodos de programación lineal y estadística aplicada.

1.3. Problema y alcance

El problema a abordar en esta tesis representa la asignación de votantes a establecimientos electorales. En particular y conforme al alcance de este trabajo, se eligió un circuito electoral, con la posibilidad de extrapolar la solución a todos los circuitos electorales de la Argentina.

El circuito electoral elegido para desarrollar la propuesta de optimizar la asignación de votantes fue el número 130 correspondiente a la Ciudad Autónoma de Buenos Aires. Este circuito resultó elegido para el desarrollo de este trabajo por tener una particularidad: está conformado por la totalidad de un sólo barrio (Coghlan). Se decide ponderar esta cualidad ya que, al tratarse de un trabajo que requiere de datos sensibles y siendo dificultoso el acceso a ese tipo de información, resulta un beneficio contar con un territorio que es circuito y barrio a la vez, porque incrementa la posibilidad de obtener información de interés. Por ejemplo, la comparación de la población creada con las características reales, se realizó gracias a un informe demográfico por comuna y no por distrito electoral, basándonos en la disponibilidad de esa información.

Como hemos mencionado, este trabajo se enfoca en la propuesta y el armado de un modelo de optimización para la asignación de votantes. Además, se desarrolla un modelo que replica las condiciones de asignación actuales. Por último, se comparan resultados de interés para justificar la aplicabilidad del modelo planteado.

El desarrollo se hará enfocado en la asignación a establecimientos electorales, ya que el objetivo principal es minimizar las distancias que surgen del recorrido de los votantes. Esto refiere a un cambio de enfoque: actualmente la asignación se hace por mesa electoral. Lo primordial aquí es considerar el establecimiento al que se asignará al votante. Una vez

allí, es indistinta la mesa a la que se lo asigne (suponiendo que todas las mesas electorales tienen la capacidad de ser accesibles). Sin embargo y para aplicabilidad funcional, se provee en el Apéndice B un código que facilita la asignación de los votantes, una vez determinado el establecimiento, a su correspondiente mesa electoral. Para ello, se diseña un procedimiento en Python al cual se le pasa como data frame el resultado de aplicar el modelo propuesto. Es decir, el output que resulta de aplicar el código creado para asignar a la población sintética a los establecimientos correspondientes, se utiliza como input para la asignación de mesas.

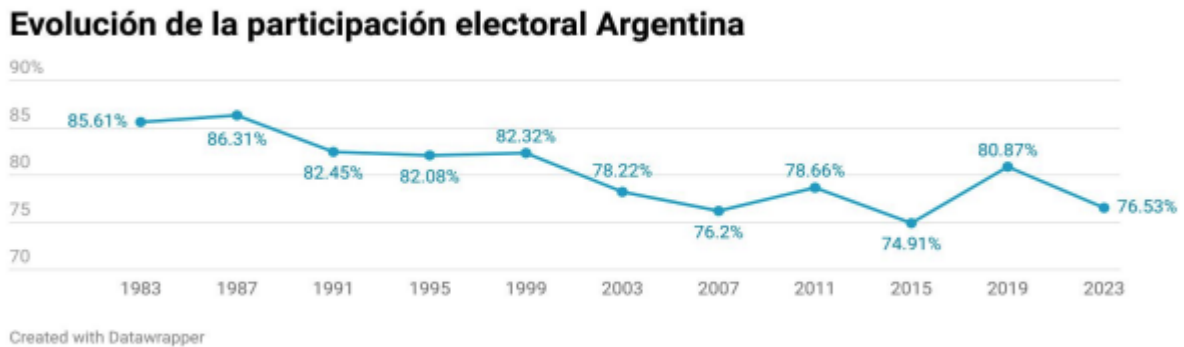
En función de la descripción sobre el sistema argentino de votación actual, se busca abordar el problema de asignación de una forma más equitativa. En términos generales, se busca desarrollar un enfoque eficiente, donde se minimicen las distancias recorridas desde los domicilios planteados hasta los establecimientos asignados. Esto en términos de eficiencia y optimización, aunque futuros trabajos podrían focalizarse en una simulación más precisa de los votantes a domicilios reales o potenciales e incluso realizar un cálculo de las distancias basados en posibles recorridos a pie o en vehículo. Sin embargo, se considera que a los fines de optimizar el modelo actual, los puntos mencionados anteriormente no resultan críticos.

A su vez, se busca desarrollarlo de una manera equitativa, procurando que el modelo considere edad, situación de las personas frente a si tienen o no discapacidad y se introduce un concepto nuevo: grupo conviviente. Esto último, resultará a los fines prácticos, el grupo de restricciones que el modelo tendrá en cuenta para asignar.

Contextualmente, se considera que aún hay oportunidad de mejorar también la participación electoral con modificaciones en la metodología de asignación frente a una

mejora en la experiencia y percepción de los electores de cara a las asignaciones. Esto resulta de particular interés, ya que la evolución de la participación electoral argentina denota una tendencia decreciente, como se puede observar en la figura que a continuación se dispone.

Figura 1. Evolución de la participación electoral en Argentina



Nota: se usan datos oficiales de las elecciones presidenciales para mostrar la evolución sobre la participación electoral.

1.4. Revisión de la literatura

En esta sección se busca referenciar aquellos trabajos que representaron un avance en lo que refiere a propuestas para reconfigurar la asignación de votantes. Si bien en la presente tesis se incursiona en conceptos que aún no fueron tratados en ese marco conforme a la revisión de la literatura vigente, resulta de interés describir algunas investigaciones motivadoras y preexistentes.

Un aporte interesante sobre la posibilidad de mejorar la asignación lo realiza Lenzi en su trabajo “Asignación eficiente de Votantes a Escuelas mediante Programación Matemática: El Caso de la Ciudad de Pergamino”. Allí, determina un enfoque basado en la programación lineal con el objetivo de minimizar, en particular, el tiempo promedio a destinar, por los votantes, en función no sólo de dónde fueron asignados sino también

contemplando el resto de las asignaciones ya que las mismas influyen en el tiempo que una persona en particular dedica a votar porque se contempla el tiempo de espera en la mesa. En este trabajo se alcanzaron resultados de impacto mediante el uso de investigación operativa, estadística y computación logrando una reducción del tiempo promedio a destinar a votar. La reducción significó un tiempo promedio menor a la mitad, del que se calculaba en función del modelo vigente.

Uno de los frentes que Lenzi menciona como dificultoso es, naturalmente, el acceso a los datos. Ese mismo problema tuvimos para el desarrollo de esta tesis. En particular, Lenzi logró el acceso a una base de datos puntual que representaba al padrón de Pergamino y por ello eligió enfocar el trabajo en esa localidad. En nuestro caso el problema se sorteará, como disponemos en el Capítulo 2, creando una población artificial con similitudes a la real, para poder aplicar el problema.

Otro de los trabajos de impacto que se enfocaron en un repensar de la metodología actual de asignación es “Asignación de votantes a centros de votación utilizando técnicas de Programación Matemática para minimizar el tiempo total invertido” de autoría de Moravec, donde el enfoque también se realiza en los tiempos destinados a la votación. En este caso y a diferencia del trabajo mencionado anteriormente, se busca minimizar el tiempo total que el padrón invertiría potencialmente el día de la votación. Para lograrlo, Moravec se basó principalmente en la Teoría de Colas para poder estimar el tiempo y la dinámica del votante llegando a la mesa asignada junto con programación lineal para poder utilizar la minimización de distancias como un componente clave para poder reducir el tiempo en términos de recorrido, desde los domicilios hasta los centros de votación.

Como disparador para futuros trabajos y no siendo desarrollado en ese en particular, Moravec detecta la potencialidad de considerar grupos etarios para el desarrollo de un modelo superador. Ese punto queda fuera de alcance de su enfoque y es mencionado como un impedimento en términos de acceso a la información. Nuevamente surge el punto de la dificultad de acceder a las bases de datos reales para constituir el verdadero padrón electoral, como fue mencionado en el trabajo anterior. El problema es entendible y plausible de saltar, ya que, como haremos aquí, la población se puede crear de manera artificial representando algunas características de interés y disponibles, para que el padrón electoral generado sea representativo. Esto considerando que el Gobierno que lo aplique, sí podrá disponer del padrón real.

Interesante enfoque el del trabajo “Improving access to voting with optimized matching” de Durán, Giormenti y otros, donde no sólo mencionan la importancia de minimizar el tiempo invertido en la votación sino que también introducen la importancia de considerar la transparencia a la hora de asignar los votantes por el hecho de enfrentarse, en el caso contrario, a manipulaciones electorales como consecuencia de las asignaciones de electores. La introducción del modelado para la asignación trae consigo, naturalmente, una posible optimización en términos de asignación desde el aspecto en interés (ya sea distancia, tiempo o consideraciones adicionales) pero a la vez y como efecto colateral, el modelado también permite visibilizar una metodología que transparenta la asignación.

A diferencia de los trabajos anteriores, el foco de este trabajo no está puesto directamente en tiempos sino en distancias y consideraciones que se refiere a lo que se propone como actos de equidad social. Consideramos un abordaje integral, donde el foco

está puesto en la importancia de considerar la experiencia de los votantes el día de la elección, cuestión que se aborda desde la metodología propuesta.

Capítulo 2: Datos del problema

En este capítulo se presentan los grupos de datos que se necesitan para aplicar el problema en cuestión. Al tratarse de un problema de asignación, requerimos candidatos a asignación y establecimientos asignantes, ambos grupos pertenecientes a un circuito electoral elegido.

2.1. Definiciones del sistema electoral argentino

En el marco de una República democrática como es Argentina, el sistema de votación se rige por ciertas reglas enmarcadas en el Código Electoral Nacional. En este sentido, mencionaremos algunas definiciones particulares que son de vital importancia en el desarrollo de este trabajo:

1. **Padrón electoral:** es el grupo de personas habilitadas para votar en la elección.
2. **Distritos:** cada provincia argentina y la Ciudad Autónoma de Buenos Aires, representa un distrito electoral.
3. **Secciones:** representan la partición dentro de los distritos. Estas subdivisiones coinciden con las comunas, en el caso de la Ciudad Autónoma de Buenos Aires y con los departamentos, en el caso de las provincias.
4. **Circuitos:** representan una subdivisión adicional, en este caso, de las secciones. Es decir, es una delimitación geográfica interna a la sección, que concluye en la agrupación de los votantes domiciliados dentro de él.

5. **Mesas electorales:** es la última subdivisión electoral. Los circuitos se dividen en mesas electorales, cuya capacidad máxima se corresponde con 350 electores.

6. **Establecimientos electorales:** lugares habilitados para disponer de mesas de votación.

7. **Asignación de electores:** cada individuo apto para votar es asignado a una mesa electoral donde figuran sus datos personales, para poder ejercer su derecho a votar el día de la elección.

2.2. Circuito electoral

Como ha sido mencionado, el circuito elegido fue el 130, constituido por el barrio de Coghlan. Conforme a la última actualización del Gobierno de la Ciudad de Buenos Aires en su archivo de datos 'Circuitos Electorales' (Dataset Buenos Aires Data, 2024), cuenta con 15998 electores totales y con 7 establecimientos aptos para recibir votaciones. Su delimitación fue provista por el Gobierno de la Ciudad de Buenos Aires. El circuito forma el siguiente polígono en función de las coordenadas que lo definen.

Figura 3. Establecimientos correspondientes al circuito electoral 130 de la Ciudad Autónoma de Buenos Aires



Nota: cada ítem representa un establecimiento apto para votar perteneciente al circuito electoral 130, con delimitación del territorio correspondiente.

2.4. Población electoral

Por otro lado, precisamos los candidatos a asignación que corresponden a la población total del circuito electoral elegido apta para votar. Por ser información sensible, ya

que precisamos dirección, piso, departamento, condición frente a certificación de discapacidad y edad, no es una base de datos de libre disponibilidad. Por ello, se decide simular una población sintética cuyo código se encuentra en el Apéndice A.

Para crear la población mencionada se replicaron algunos datos reales disponibles y se crearon otros, a los fines prácticos de aplicación de este trabajo. Para ello, se desarrolló un código de Python, utilizando diferentes bibliotecas. En primer lugar, se estableció que la población total representaría un conjunto de 15998 votantes. Luego, se indicaron las coordenadas provistas y mencionadas en la sección anterior, que definen a Coghlan para dar un marco a las direcciones que el código podía generar y así se generó el polígono que delimita el alcance de este trabajo en particular. En esta tesis, se busca demostrar la aplicabilidad de un modelo que entiende de cálculo de distancias y situaciones, por lo que se asume que las personas no tienen restricción sobre dónde está su domicilio, lo que significa que pueden domiciliarse en cualquier lugar del polígono y no se excluyen plazas, hospitales y demás establecimientos sin domicilio posible. Respecto a este punto, quedarán naturalmente excluidos de utilizarse los datos reales de población. Posteriormente y a los fines de contar con información que forma parte del padrón electoral, se crearon funciones para generar aleatoriamente con el abecedario y cantidad delimitada de caracteres, nombre, apellido y DNI de los votantes. También, para determinar grupos convivientes y demostrar el funcionamiento del modelo, se hace especial hincapié en que existan personas que vivan en el mismo lugar e incluso mismo piso y departamento para que logren ser considerados como grupo conviviente. Por ello, se establece arbitrariamente que hay a razón de 4 personas viviendo en edificios por cada 10 personas aptas para votar (es decir, el 40% de las personas se determina que viven en edificios). Con esta definición, se establecen listas que nos permitan determinar pisos y departamentos donde domiciliarlos, explicitando en el

modelo que como máximo se asignen 10 pisos y 5 departamentos por cada uno. Si quedan rezagos, el modelo adjudica como 'Casa' a quien haya quedado por fuera de los domicilios en edificios creados. Luego, se crea la lista de grupos convivientes con un tope máximo de 4 convivientes aptos para votar y se crean combinaciones de latitud, longitud, piso y departamento para luego asignar allí a las personas creadas. Para contar con grupo etario de interés, se define que el 20% de la población será asignada al grupo mayor de 60 años. Las personas a las que se les asigna Certificado Único de Discapacidad se determinan siguiendo una distribución binomial¹.

Por último, la información queda guardada en un data frame en el que cada fila representa una persona apta para votar y las columnas quedan conformadas por el nombre, apellido, DNI, longitud de domicilio, latitud de domicilio, piso y departamento de domicilio, edad y si cuenta o no con "certificado único de discapacidad" (CUD). De manera descriptiva, se crean dos columnas más que indican un código de grupo conviviente y una columna adicional que indica si esa persona tiene convivientes o no, en función de si comparte código de grupo conviviente con alguien más.

La población sintética obtenida y que se utilizará en el desarrollo de este trabajo cuenta con las siguientes características de interés:

1. Cantidad de electores: resultan ser 15998 electores posibles, exactamente el mismo número que, según los datos abiertos del Gobierno de la Ciudad y dispuestos en el dataset correspondiente al circuito 130 de la Subsecretaría de Justicia, conformaron la última elección. La precisión requerida para esta característica era del 100%, ya que se utilizan los

¹ En este caso, siendo la muestra lo suficientemente grande (su tamaño es de 15998 votantes), la distribución tiende a aproximarse a una distribución normal debido al Teorema Central del Límite.

establecimientos reales y se espera que la solución del modelo sea representativa del distrito.

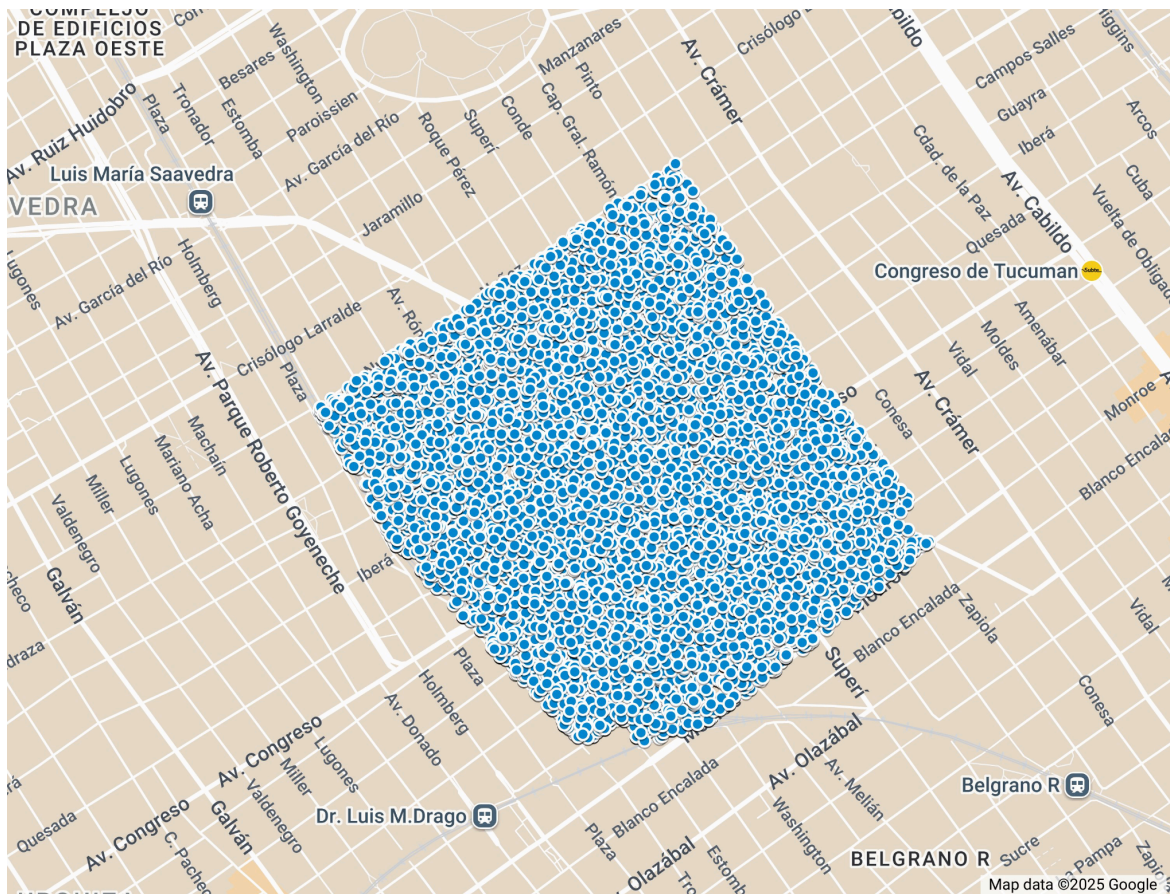
2. Cantidad de electores sin convivientes: el 26% de la población electoral del distrito no cuenta con otras personas aptas para votar en su domicilio. Esto nos deja una oportunidad de optimización importante, ya que entonces el 74% de las personas comparte vivienda con alguien con quien, potencialmente, podría asistir a votar al mismo lugar. Si bien este no es un dato disponible, podemos compararlo con un dato provisto por el Instituto de Estadísticas de la Ciudad de Buenos Aires, cuyos últimos datos disponibles indican que el 35% de los hogares son unipersonales en la Comuna 12 (dentro de la cual se encuentra el distrito electoral en cuestión).

3. Porcentaje de personas con más de 60 años: el 20% de la población electoral de este distrito es mayor a 60 años según nuestra población sintética. Conforme a los datos mencionados en el ítem anterior, el 18% de las personas de la Comuna 12 tiene 65 años o más, con lo que podemos confirmar que nuestro dato está alineado con la población real.

4. Porcentaje de personas con certificado de discapacidad: en nuestra población artificial, el 4.8% de la población cuenta con CUD. Según los datos estadísticos sobre discapacidad provistos en la página oficial del Gobierno de la República Argentina, el porcentaje de personas certificadas a nivel nacional es del 3.5%. Con esta comparación podemos asegurar una útil sobreestimación de este grupo, para asegurar la viabilidad de esta solución en un contexto real.

Entendiendo la representatividad de la población creada, se grafica la población en la Figura 4 dentro del mapa zonal.

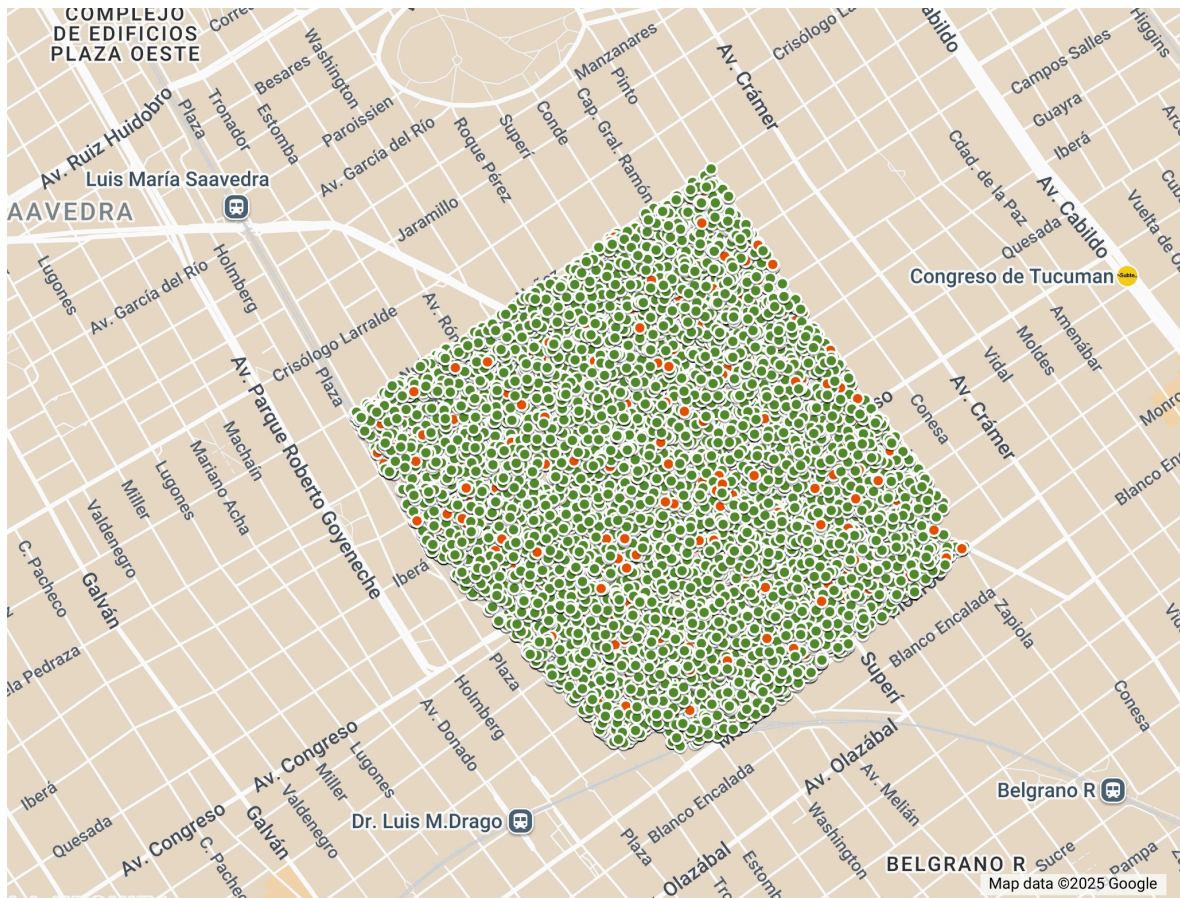
Figura 4. Población votante del circuito electoral 130 de la Ciudad Autónoma de Buenos Aires



Nota: cada punto celeste representa un votante perteneciente al circuito electoral 130 de CABA.

En la Figura 5 se puede apreciar la misma población pero distinguiendo en función de la certificación de discapacidad, lo que nos permite ver la distribución creada para representar este segmento de la población.

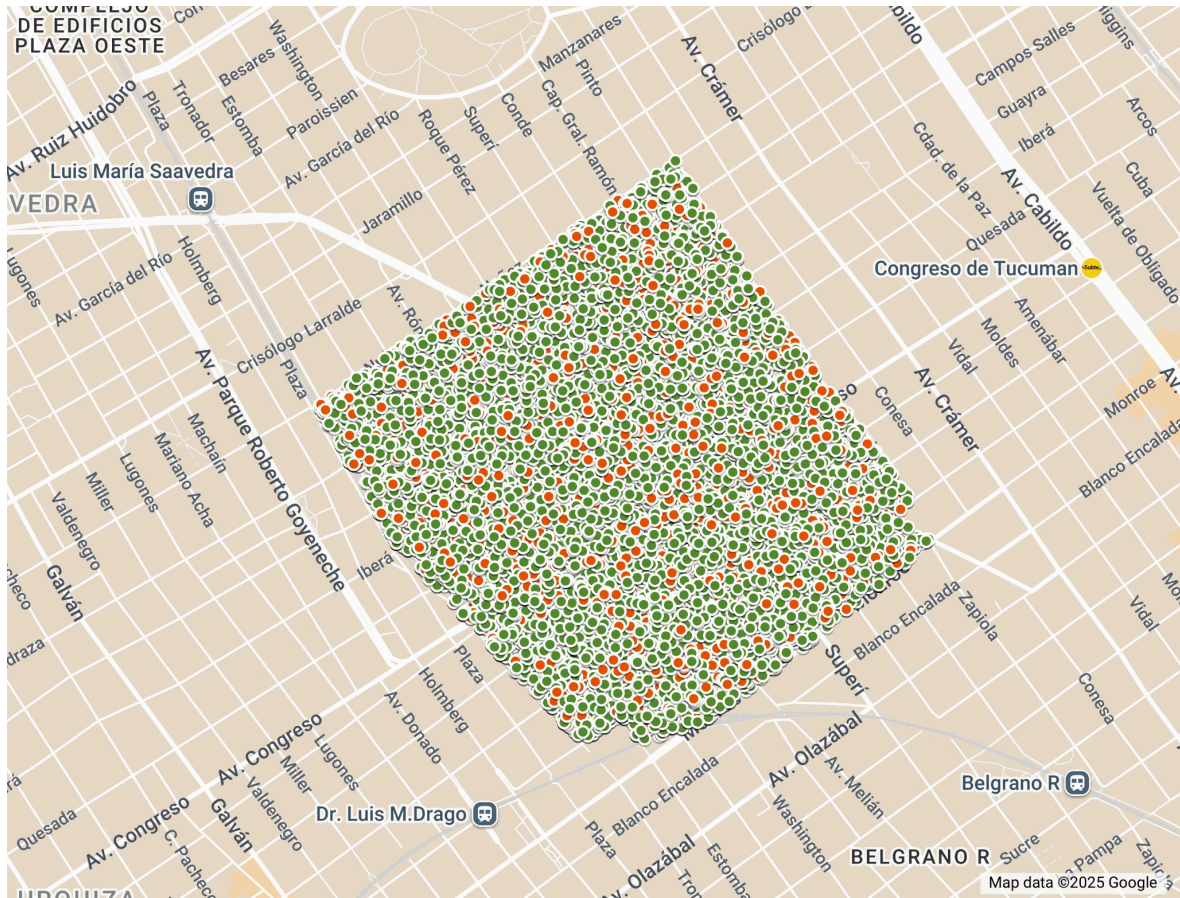
Figura 5. Población votante según CUD



Nota: los puntos verdes representan a los votantes sin certificado de discapacidad. Los puntos naranjas representan a los votantes con certificado de discapacidad.

En la Figura 6 se grafica la población sintética discriminando según grupo etario del votante, para asegurarnos visualizar un esquema lo suficientemente representativo donde se ubiquen votantes de este grupo alrededor de todo el distrito.

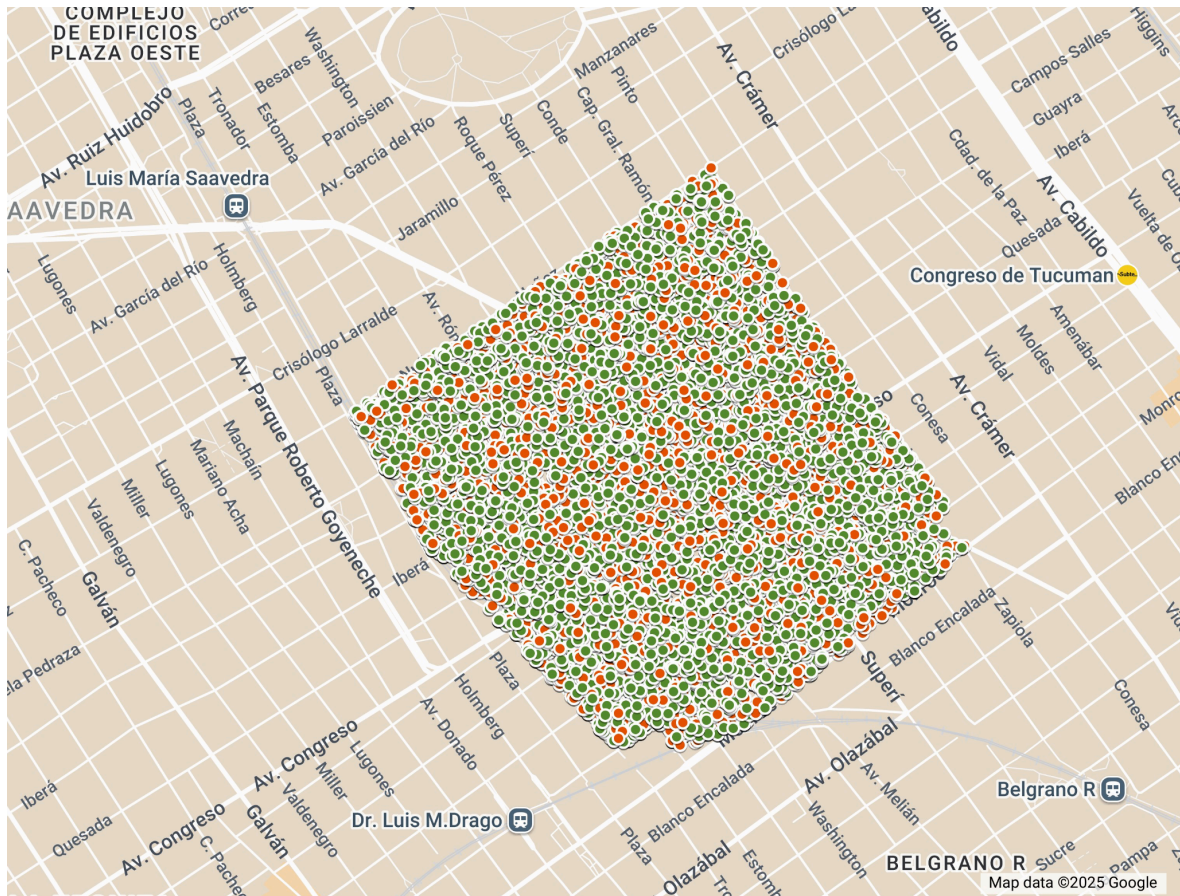
Figura 6. Población según edad



Nota: los puntos verdes representan a los votantes menores de 60 años. Los puntos naranjas representan a los mayores de 60 años.

Por último, se decide mostrar gráficamente la distribución dentro del distrito de casos con y sin convivientes, ya que se implementará una restricción en el modelo de asignación también basado en el concepto de grupo conviviente.

Figura 7. Población votante según grupo conviviente



Nota: los puntos verdes representan a los votantes sin grupo conviviente asociado. Los puntos naranjas representan votantes que comparten grupo conviviente con otros.

Capítulo 3: Metodología

3.1. Metodología actual

La metodología de asignación de votantes actual en Argentina tiene un marco general y nacional. Tras eso, cada distrito puede accionar de manera distinta siempre y cuando respete la definición general de la Justicia Electoral Nacional.

En este sentido, los electores tendrán definido el circuito en función de la delimitación coordinada entre la Justicia Electoral Nacional y el Distrito en cuestión. Una vez catalogados los votantes por circuito, se procede a ordenarlos alfabéticamente para determinar la mesa, que forma parte de un establecimiento, considerando su capacidad máxima. De esta forma y conforme a lo dictado por el Código Electoral Nacional, quedan determinadas las mesas y por ende los establecimientos electorales, en los que cada elector tiene derecho a ejercer su voto, sin considerar excepciones particulares. (Ley 19.945, 1972 y modificaciones).

Como fue expuesto anteriormente, no hay un criterio sofisticado para asignar electores a establecimientos electorales. En pocas palabras, la metodología implica separar las localidades en circuitos electorales para luego catalogar los domicilios dentro de esa delimitación, para determinar qué votantes y establecimientos electorales lo componen. Una vez así, se ordenan las personas alfabéticamente por apellido y luego, conforme a la capacidad, se asignan a los establecimientos. No hay información sobre cómo se ordenan los establecimientos para determinar a cuál se le asigna el primer batch de votantes, a cuál

el segundo, y así hasta ubicar a todos los electores. Esto representa un hecho de poca transparencia en lo que refiere al método actual de asignación.

A continuación, se desarrolla la metodología propuesta para abordar las oportunidades de mejora que la forma actual de asignar nos deja entrever.

3.2. Metodología propuesta

Para abordar de manera superadora la asignación de electores, se utilizará un modelo de programación lineal entera, cuya descripción haremos a continuación para dar sustento teórico a la idea que dispara este trabajo. En síntesis, se busca utilizar un método que permita proponer un esquema de asignaciones que busque que en suma la población esté lo más cerca posible del lugar que vota, que se respete la capacidad de cada establecimiento, que en la medida de lo posible aquellas personas que viven juntas también voten juntas y que los grupos prioritarios tengan preferencia a la hora de asignar un establecimiento más cercano. Lo expuesto, se modelará con programación lineal entera para llegar a una solución óptima por distrito electoral.

Boirivant (2009) argumenta:

“En términos generales, se puede decir que cualquier fenómeno en que interviene un número determinado de variables no negativas (es decir, variables cuyo valor es positivo o cero), que se pueden ligar entre sí mediante relaciones de desigualdad o igualdad y que reflejen las limitaciones o restricciones que el fenómeno presenta con miras a optimizar un objetivo, puede ser formulado como un modelo de programación matemática. Si tanto las restricciones como la función objetivo se pueden enunciar mediante expresiones lineales, estamos frente a un campo particular de la programación matemática denominada “programación lineal”.” (pág. 2)

Para este caso, entonces, tendremos en primer lugar una función objetivo que típicamente buscaremos minimizar y que refiere a distancia total. Compuesta, por un lado, de las distancias efectivamente calculadas entre domicilios y establecimientos asignados a cada persona, junto con ciertos factores de penalización que, a los fines prácticos, representarán también distancias. Los parámetros y los factores de penalización añaden a la función objetivo más metros como respuesta a no cumplir con alguna de las condiciones señaladas. Las condiciones son justamente y como se ha mencionado, relaciones de desigualdad o igualdad que el modelo persigue pero que, si no puede cumplir, tiene la posibilidad de hacerlo a costa de que se activen las penalizaciones correspondientes. Característicamente, los modelos de programación matemática especifican para qué valores aplican las restricciones que se configuran.

En nuestro caso, algunas variables resultan enteras y otras continuas y nuestra función objetivo, que presentaremos más adelante, es lineal, ya que no hay exponentes mayores a 1 en ella. Sucede que, "Cuando sólo es necesario que algunas de las variables sean enteras y el resto continuas, el modelo recibe el nombre de problema de Programación Lineal Entera Mixta" (Bermúdez Colina, 2011, 89). Entonces así, queda definido que el modelo que se propone para realizar las asignaciones, es un modelo de programación lineal entera mixta.

En resumen, el modelo propuesto buscará una combinación de asignaciones que generen la mínima distancia total, considerando las restricciones que configuremos aunque considerando y ponderando la posibilidad de no cumplirlas en todos los casos. El modelo cuenta con ciertos parámetros y variables que explicaremos en primer lugar, para facilitar la comprensión del modelo. En este sentido, se desarrolla a continuación la propuesta para alcanzar una combinación de asignaciones óptima.

1. Datos del modelo: son los datos que ya vienen dados del problema en sí y no requieren un tratamiento para decidir sus valores o características. Estos son:
 - 1.1. C_j es la **capacidad máxima del establecimiento**. Es un dato y está determinado en función de la cantidad de mesas que tiene el establecimiento j y la capacidad de cada una de ellas.
 - 1.2. **Convivientes es el conjunto de votantes que comparten domicilio con al menos otro votante del padrón.**
 - 1.3. **CUD es el conjunto de votantes que tiene Certificado Único de Discapacidad.**
 - 1.4. **noCUD es el conjunto de votantes que no tiene Certificado Único de Discapacidad.**
 - 1.5. **+60 es el conjunto de votantes que tiene 60 años o más a la fecha de impresión del padrón electoral.**
 - 1.6. **no+60 es el conjunto de votantes que tiene menos de 60 años a la fecha de impresión del padrón electoral.**
 - 1.7. **N es el conjunto de electores.** Representa la totalidad de personas aptas para votar al momento de definir el padrón electoral.
 - 1.8. **M es el conjunto de establecimientos.** Representa la totalidad de Centros de Votación aptos para recibir electores.
 - 1.9. d_{ij} es la **distancia calculada en función de dónde fue asignado cada votante, respecto a su domicilio**. En efecto, significa la distancia que separa al votante i del establecimiento j .

2. VARIABLES DEL MODELO: son las decisiones a definir por el modelo. Constituyen las variables de decisión que surgen de la aplicación del modelo. Estos son:

2.1. x_{ij} **es una variable binaria que contempla asignación** de persona y establecimiento. Resulta ser 1 si la persona i vota en el establecimiento j . A destacar, es una variable binaria, toma el valor 0 cuando la persona i no vota en el establecimiento j .

2.2. z_i **variable binaria que indica la excepción de asignar convivientes a distintos establecimientos**. Cada vez que dos convivientes no votan juntos, esta variable toma valor 1, para luego generar que se active el factor de penalización que será mencionado más adelante.

2.3. P_{CUD} y P_{+60} **son variables continuas de penalización**. En este caso, representan la diferencia entre el 80% de la distancia promedio recorrida por el grupo que los excluye (sin discapacidad o menores de 60 para cada caso) y la distancia promedio recorrida por cada uno de los grupos. De ser mayor a 0, se activa y se multiplica por el factor de penalización mencionado en los ítems que a continuación se detallan.

3. PARÁMETROS DEL MODELO: son los datos a definir por el ejecutor del modelo previamente a su aplicación para hallar los valores de las variables de decisión. Estos son:

3.1. α **es el factor de penalización para las distancias recorridas por personas con discapacidad**. Es un parámetro que, para este caso en particular, se determinó en 10000. Multiplicará, en la función objetivo, a la distancia excedente generada. Esto refiere a la distancia promedio a recorrer por el grupo menos el target máximo

del 80%. Esta asignación excesiva en términos de distancia, no permite cumplir que las personas con discapacidad recorran como máximo un 80% de lo que recorre el resto.

3.2. β es el factor de penalización para las distancias recorridas por mayores de 60. Al igual que la descripción anterior, multiplicará la distancia que resulta excedente generando que los mayores de 60 recorran menos que los menores. Para este caso, también se configuró en 10000 con la idea de dar un mensaje de equidad en el tratamiento para ambos grupos prioritarios.

3.3. Δ es el factor de penalización para las excepciones en la asignación de convivientes. Se activa cada vez que se separa a una persona de su grupo conviviente. En este sentido y frente a la posibilidad de que el modelo separe a todos los convivientes, se configura una penalización de 300 (metros) por cada separación.

Ahora sí, se presenta la función objetivo que se busca minimizar:

$$\min Z = \sum_{i \in N} \sum_{j \in M} d_{ij} x_{ij} + \alpha P_{CUD} + \beta P_{+60} + \Delta \sum_{i \in N} z_i$$

Sujeto a:

1. Restricción de asignación, representa la asignación de todos y cada uno de los votantes. Con esta restricción, nos aseguramos de que cada elector sea asignado a un solo establecimiento:

$$\sum_{j \in M} x_{ij} = 1 \quad \forall i \in N$$

2. Restricción de capacidad, que nos limita a no asignar más personas que las que el establecimiento está destinado a recibir conforme a las mesas que lo constituyen:

$$\sum_{i \in N} x_{ij} \leq C_j \quad \forall j \in M$$

3. Restricción de distancia recorrida por personas con discapacidad, que fuerza al modelo a asignar establecimientos que, en promedio, sean más cercanos para este grupo que lo que son para la población en general:

$$\frac{\sum_{i \in CUD} \sum_{j \in M} d_{ij} x_{ij}}{|CUD|} \leq 0.8 * \frac{\sum_{i \in noCUD} \sum_{j \in M} d_{ij} x_{ij}}{|noCUD|} + P_{CUD}$$

4. Restricción de distancia recorrida por personas mayores de 60 años, que, al igual que la anterior, fuerza al modelo a asignar establecimientos que, en promedio, sean más cercanos para este grupo que lo que son para la población en general:

$$\frac{\sum_{i \in +60} \sum_{j \in M} d_{ij} x_{ij}}{|+60|} \leq 0.8 * \frac{\sum_{i \in no+60} \sum_{j \in M} d_{ij} x_{ij}}{|no+60|} + P_{+60}$$

5. Restricción de convivencia, con la intención de que todos los convivientes voten en el mismo establecimiento. Sin embargo, si los convivientes no pueden ser asignados al mismo establecimiento, se activa z_i tomando valor de 1, lo que permite que x_{ij} y x_{kj} sean distintos:

$$x_{i_k j} \leq x_{i_l j} + z_{i_k} \quad \forall (i_k; i_l) \in \text{Convivientes}; \forall j \in M$$

3.3. Aprendizajes para el modelo

El llegar al modelo final fue el resultado de un camino de iteraciones y aprendizajes. Para facilitar la comprensión de las pruebas realizadas, resulta importante repasar que el modelo cuenta, en primer lugar, con una función objetivo que se busca minimizar. A priori y sin consideraciones adicionales, el modelo busca minimizar la suma de las distancias a recorrer por los votantes, calculada como la diferencia entre su domicilio y el establecimiento asignado. Ahora bien y conforme a lo desarrollado en este trabajo, se deciden incorporar adicionales a considerar, para que los resultados, a la vez de buscar que en suma las personas recorran lo mínimo posible, prioricen a aquellos que cuenten con certificado de discapacidad y a aquellos que tengan al menos 60 años para que incluso recorran menos que el resto de la población. Por último y como restricción innovadora, la idea es que la asignación tenga en cuenta si la persona vive con otra para asignarlos juntos y así, minimizar los costos logísticos que implican que los convivientes voten separados.

Sin embargo, hay que considerar que el modelo busca una solución óptima. Si somos demasiado estrictos en las restricciones, corremos el riesgo de que no la encuentre. Por eso, es necesario que se contemple bien qué cuestiones deseamos que no se pierdan a

la hora de encontrar una solución pero entendiendo que puede que no cumpla todo lo que deseáramos a la vez, por una cuestión principalmente propia de la capacidad de los establecimientos. Para ello, es clave calibrar el modelo, contemplando todas las restricciones, ponderando de manera razonable el hecho de que no las cumpla y así, permitirle encontrar una solución óptima. En este sentido, se puede ir por dos caminos: o se estructuran las restricciones de forma tal que sean más o menos estrictas sin perder de vista nuestro objetivo final de asignación o bien, se agregan factores de penalización que luego irán a la función objetivo ponderados. Pueden también combinarse ambas estrategias, que es lo elegido para este caso. Es importante destacar que, al tratarse de una función objetivo que minimiza cantidad de metros, las penalizaciones representan una cantidad de metros también, para ser comparable. Por eso, también se presenta una ejercitación sobre a cuántos metros adicionales decidimos que equivale el hecho de no cumplir una restricción.

Para considerar como elegido el modelo propuesto, se hicieron diferentes propuestas primero. Se probó en un grupo de 500 electores de la población creada por los tiempos de procesamiento que implican correr todo el modelo. Sin embargo, cuando se definió la estrategia de penalización, el ajuste de los parámetros se realizó con la población total, para que aplique la restricción de capacidad, ya que con un grupo pequeño la capacidad no cobra importancia y para el modelo es más fácil encontrar una solución factible. Agrupando las pruebas, se realizaron los siguientes experimentos preliminares:

1. Restricciones de grupos prioritarios poco estrictas y sin penalización: aquí se probó un modelo en el cual se penaliza solamente si se separa una persona de su grupo conviviente. Esa penalización, en un principio, se probó de manera tal que sumaba un metro

a la función objetivo. Esto quiere decir que, por cada persona separada del grupo conviviente, la suma de las distancias recorridas sumaba un metro más. Las restricciones de los grupos prioritarios que se probaron inicialmente, solamente sugieren que estos grupos recorran menos distancia que el resto. Los resultados de este modelo no separaban grupos convivientes y cumplían el hecho de que los prioritarios recorran menos, pero recorrían muy poco menos (menos de un metro entre promedios) así que se decidió seguir iterando el modelo para agrandar esa distancia.

2. Restricciones de grupos prioritarios poco estrictas y con penalización:

mejorando lo anterior, se decidió penalizar el hecho de que no se cumplan las restricciones de los grupos priorizados en la función objetivo. Los resultados todavía mostraban oportunidades de mejora, ya que el promedio recorrido por los grupos en cuestión seguía siendo prácticamente el mismo que los grupos no prioritarios. Es decir, era menor la distancia promedio a recorrer, pero con una diferencia demasiado sutil. Esto continuó sucediendo a pesar de las iteraciones en los valores de β y de α , que en primera instancia representaban el mismo parámetro y multiplicaban la suma de las penalizaciones de grupos prioritarios, lo que luego se modificó para tener dos parámetros distintos y poder ajustar el modelo por si se consideraba más importante que un grupo prioritario recorra menos, en proporción, que el otro.

3. Restricciones de grupos prioritarios más estrictas y con penalización:

con restricciones que sólo enfatizan en que los grupos prioritarios recorran menos que el resto, no logramos resultados significativos, tampoco al penalizar el hecho de que no se cumpliera. Aquí se notó que el problema era que la restricción era demasiado poco estricta, lo que le permitía a la solución óptima arrojar una media recorrida por los grupos prioritarios

apenas inferior al resto y sin tener una penalización por eso. Por ello, se decidió mantener las penalizaciones pero hacer más estrictas las restricciones, forzando a una diferencia de al menos el 20% de distancia a recorrer en primer lugar, para los mayores de 60. Esto generó que la solución óptima contemplase que los mayores recorran casi un 75% que los menores de 60. Así, se notó la importancia de determinar qué tanto menos debían recorrer los prioritarios y se aplicó también a las personas con discapacidad, penalizando con un parámetro de 100000 si esa proporción mínima no se cumplía.

4. Restricciones de grupos prioritarios más estrictas y calibración de los parámetros de penalización: una vez que se decidió la importancia de determinar expectativas en función de cuánto menos deben recorrer los grupos prioritarios, se empezaron a calibrar los parámetros que multiplican las penalizaciones para ajustar cuánto impactan los no cumplimientos en la función objetivo. Se comenzó probando con valores bajos de parámetros, iniciando desde 100 y asignando la misma ponderación a las tres penalizaciones. Sin embargo y como aprendizaje de los resultados, se notó que resultaba vital tener un criterio distinto para la penalización de grupo conviviente respecto a las penalizaciones de los grupos prioritarios, porque estas últimas se activan una única vez, de no cumplirse la comparación entre medias, recompensando lo que no llegaron a cumplir con la solución propuesta. Sin embargo, cada vez que se separa una persona de su grupo conviviente, la penalización de convivencia se activa. Por este motivo y en comparación, el parámetro de convivencia debe ser menor al resto. La decisión de cuánto menor debe ser, dependerá principalmente del tamaño de la población. En este caso, se terminó definiendo que el factor de penalización de convivencia sea 300, mientras que los factores de penalización de grupos prioritarios sean los mismos y tomen valor 10000. Para este caso, los resultados resultaron aceptables para definir los parámetros planteados. Para otros

tamaños de población o preferencias, se recomienda reconsiderar las proporciones y comparar resultados, para determinar parámetros si los indicados no se ajustan a las expectativas.

5. Parámetros de penalización fijos con restricciones más estrictas: con los parámetros ya definidos, se realizaron pruebas para observar cómo influyen frente a modificaciones en las restricciones. Es importante aclarar que este trabajo propone enfáticamente que las definiciones acerca de las restricciones son una discusión que debe realizarse de manera previa a la aplicación del modelo. Es decir, es necesaria una definición política sobre cuánta distancia, en proporción al resto, deben recorrer como máximo los grupos prioritarios y qué tanto importa ello respecto a la separación de los grupos convivientes a la hora de asignarles el establecimiento. Dicho esto, realizamos pruebas en las que se modifica la relación de distancias promedio entre los grupos prioritarios y los no prioritarios, en las restricciones. En la propuesta de este trabajo, se busca que los grupos de interés recorran, como máximo, un 80% respecto a lo que recorre el resto. En las pruebas, ese porcentaje se modificó y probó con valores de 70%, 60%, 50% y 10% para ver cómo se comportaba el modelo. En todos los casos, los resultados generaron un aumento de la distancia total a recorrer, con hasta incrementos del 8%. Ninguna de las pruebas logró mejorar la distancia media recorrida por la población en general ni la recorrida en particular por los grupos no prioritarios. Todas las pruebas lograron disminuir la distancia media recorrida por los grupos prioritarios pero en ninguno de los casos se logró cumplir estrictamente con todas las restricciones (en el caso en el cual la relación se determinó en un 70% se logró cumplir para el caso de los mayores de 60 pero no así en el grupo de personas con discapacidad). En todas las pruebas, los grupos convivientes fueron asignados al mismo establecimiento. Es evidente que la penalización por separar a los

grupos convivientes no es tan significativa para el modelo como lo es el hecho de no cumplir con las distancias requeridas para los grupos prioritarios. La enseñanza que nos deja esta última iteración es que los parámetros deben ser ajustados en función de las definiciones del modelo y que, a pesar de que puedan no cumplir estrictamente todas las restricciones, los resultados aún son superadores respecto al modelo actual ya que en el peor de los casos, las pruebas arrojan un 64% de distancia total a recorrer, respecto a la metodología actual.

Capítulo 4: Resultados

4.1. Resultados modelo actual

A continuación, se presentan los resultados en la Tabla 1 de aplicar el modelo actual, cuya programación se encuentra en el Apéndice A, sección 3. Este modelo, se aplicó sobre la base de la población sintética creada, que representa a la población real del distrito electoral 130.

Tabla 1. Resultados de aplicar el modelo de asignación actual a la población creada.

Métrica	Cálculo
Distancia total recorrida (m)	9.335.204
Distancia promedio recorrida (m)	584
Distancia promedio recorrida por votantes de hasta 60 años (m)	583
Distancia promedio recorrida por mayores de 60 años (m)	587
Distancia promedio con CUD (m)	589
Distancia promedio sin CUD (m)	583
Porcentaje de grupos convivientes (de más de 1 persona) que votaron todos juntos	12.89%

Como pudimos observar, los resultados no denotan equidad respecto al tratamiento por tipo de votante. Naturalmente la distribución resulta bastante pareja por la ausencia de restricciones que mejoren los resultados. En particular y en lo que se centrará el nuevo modelo, resaltamos tres cuestiones principales:

1. **La distancia promedio no varía por edad.** Los resultados incluso nos dejan ver que se asignaron establecimientos a los mayores de 60 años de forma tal que terminan recorriendo una distancia promedio mayor a la media.

2. **La distancia promedio no varía en función de si se tiene o no CUD.** Al igual que por edad, los resultados dispararon asignaciones que generan mayor distancia promedio para el grupo de votantes con CUD.

3. **En su mayoría, los grupos convivientes no votan en el mismo establecimiento.** Conforme a los resultados, casi el 13% de los grupos convivientes que tienen más que un votante, votan todos en el mismo lugar. Al igual que los puntos anteriores, no se cuenta con ninguna restricción al respecto y significa una mera coincidencia que los convivientes voten juntos.

4.2. Resultados nuevo modelo

Con el modelo de asignación propuesto y explicado, junto con la población sintética creada, estamos en condiciones de implementar nuestra propuesta a una población que, conforme a lo expuesto, resulta significativa para demostrar la viabilidad de lo planteado. Los resultados de esa aplicación se disponen a continuación:

Tabla 2. Resultados de aplicar el modelo de asignación propuesto a la población creada

Métrica	Cálculo
Distancia total recorrida (m)	5.509.876
Distancia promedio recorrida (m)	344
Distancia promedio recorrida por votantes de hasta 60 años (m)	359
Distancia promedio recorrida por mayores de 60 años (m)	287
Distancia promedio con CUD (m)	278
Distancia promedio sin CUD (m)	348
Porcentaje de grupos convivientes (de más de 1 persona) que votaron todos juntos	100%

En síntesis, el nuevo modelo logra que todas las personas que comparten domicilio voten en el mismo establecimiento, generando así una disminución, como mínimo, de los costos logísticos que tiene votar por separado. Se ven oportunidades significativas en este punto si el modelo se extrapola a nivel nacional, donde las distancias recorridas son naturalmente mayores en el interior del país. Allí, las personas tienden a recorrer kilómetros para ir a votar, lo que realmente denota una oportunidad de disminuir costos logísticos y operativos. Por ejemplo, el hecho de compartir el automóvil en un núcleo conviviente y tener que esperar que un integrante termine de votar para dirigirse el otro integrante a su centro de votación, genera una evidente dificultad e ineficiencia logística. Además, se considera que las personas pueden ponderar menos los tiempos de espera si los comparten con su grupo conviviente.

Por otro lado, los resultados muestran que se ha logrado que aquellas personas de 60 años o más estén domiciliadas a una distancia menor respecto a su establecimiento asignado. En promedio la distancia de ese grupo representa un 79.94% de la distancia promedio calculada para los menores de 60, entre su domicilio y su centro de votación

asignado. Similar a lo logrado con las personas con CUD, cuya distancia promedio representa un 79.88% de la distancia promedio de aquellas personas sin discapacidad declarada.

Sin duda y tal como se mencionará en la sección siguiente de análisis comparativo de resultados, la propuesta es viable, equitativa y superadora.

4.3. Medición de éxito

La propuesta de modelo que se ha realizado en el presente trabajo busca aportar tecnológicamente a los gobiernos para que ganen en profesionalismo, metodología equitativa y transparencia. El modelo se basa en criterios equitativos, como minimizar distancias y considerar prioridad a determinados grupos. Sin embargo y como primordial no solo en la base de la idea sino en el modelo implementado, el proyecto apunta a que los votantes en general tengan una mejor experiencia a la hora de votar. A la vez de incluir mejoras equitativas, se consideró la incorporación del criterio de grupo conviviente que se espera sea una condición de impacto positivo.

En función de lo enunciado, no es suficiente un modelo que simplemente funcione y por eso en esta sección buscamos introducir algunas medidas de impacto para poder testear la implementación. Las comparaciones resaltadas sobre la implementación del modelo actual y del modelo propuesto, que ya fueron justificadas y son destacadas en la sección anterior, representan el primer grupo de indicadores para medir el éxito del modelo. Por otro lado, de cara a la implementación efectiva del modelo aquí propuesto, se plantea un esquema de encuesta a una muestra lo suficientemente representativa basada solamente en ponerle la comparación de dónde le tocaba y dónde le toca con el modelo propuesto. Se cuenta éxito si más del 50% de las personas elige la opción del nuevo

modelo. Esta estrategia no se implementará en el presente trabajo porque justamente se plantea como una idea experimental para ser considerada e implementada por los organizadores electorales, en una elección.

Por otro lado y como efecto colateral esperado, es un plus de éxito que aumente la participación electoral como consecuencia de las mejoras implementadas. En resumen, se centrará la atención en estas tres verticales de éxito:

1. Distancia recorrida por los grupos prioritarios en comparación con el resto de la población. La metodología del cálculo se centra en calcular la distancia promedio recorrida y la distancia promedio recorrida por ambos grupos prioritarios. Con esa información, se construyen los dos primeros KPIs, representados como el ratio de distancia a recorrer por cada grupo respecto al total:

$$\text{KPI 1: } \frac{\textit{distancia promedio recorrida por los mayores de 60 años}}{\textit{distancia promedio total recorrida}}$$

$$\text{KPI 2: } \frac{\textit{distancia promedio recorrida por las personas con CUD}}{\textit{distancia promedio total recorrida}}$$

Ambos ratios deben ser menores a 1, lo que demostrará que los grupos prioritarios recorren, en promedio, menos distancia que la población en general.

2. Satisfacción ciudadana. El tercer KPI apunta a medir la opinión de los votantes respecto al cambio de metodología en la asignación. Esto refiere a una propuesta que no se implementa en este trabajo pero se deja a disposición como idea para implementar tras la

aplicación real del modelo propuesto. Para eso, se propone elegir un grupo de personas² a las que se las someterá a una pregunta respecto al cambio en la metodología del cálculo. Para ello, en la mesa de votación, se le pedirá a los elegidos que respondan a la siguiente consigna:

Si hubiese podido elegir dónde votar entre las siguientes dos opciones, en qué establecimiento lo hubiese hecho?

- *Opción 1* (establecimiento en el que hubiese votado si se respetaba la metodología actual)
- *Opción 2* (establecimiento designado con la nueva metodología)

Para determinar el establecimiento de la opción 1, se replica el modelo actual y se le coloca como opción, el establecimiento que le toca a esa persona. Para la opción 2, se implementa la nueva metodología para hallar el establecimiento que le corresponde a ese votante.

De aquí, surge el tercer KPI:

$$\text{KPI 3: } \frac{\text{cantidad de personas que eligieron la opción 2}}{\text{total de respuestas válidas}}$$

El ratio en cuestión debe ser mayor a 0.5, lo que se traducirá como satisfacción de la población respecto al cambio, ya que se podrá concluir que, para la mayoría, el cambio fue positivo.

Este KPI puede ser usado frente a la implementación del cambio de metodología para corroborar la satisfacción o bien puede invertirse y utilizar esta encuesta en una

² En el Apéndice B se presenta una metodología estadística para determinar el tamaño de la muestra a encuestar y un código para elegir aleatoriamente esa muestra, en función del tamaño de la población.

elección en la que no se haya aplicado la nueva metodología y que la persona elija entre lo que le tocó con la metodología actual y lo que le tocaría si se implementara la propuesta de este trabajo.

3. Participación electoral. Si bien la participación electoral se supone multicausal, este trabajo pretende brindar un efecto colateral en lo que respecta a ello. Si la metodología es más equitativa sin perder en términos generales, basados en la satisfacción ciudadana, se considera que puede al menos tener impacto positivo en la participación electoral. De aquí surge el cuarto y último KPI de interés:

$$\text{KPI 4: } \frac{\text{cantidad de personas que efectivamente votaron}}{\text{total de personas habilitadas para votar}}$$

Conforme a los datos, si este ratio significa más que un 81% en una elección presidencial cuya metodología aplicada resulta la de este trabajo, se considera un aporte positivo, ya que no se alcanza una participación electoral de ese nivel, hace más de 25 años.

Capítulo 5: Conclusiones

5.1. Comparación de resultados

Con la evidencia presentada en el capítulo anterior podemos demostrar que la aplicación del modelo planteado al distrito 130 de CABA, arroja resultados realmente superadores respecto a aplicar el modelo actual replicado a la misma población. En particular desarrollaremos a continuación la comparación de cada uno de los ítems de interés:

1. Distancia total. Aplicando el modelo actual, que no contempla distancias, la distancia total entre domicilios y establecimientos asignados arroja un total de 9.335.204 metros totales. Al aplicar el modelo propuesto, la distancia total desciende a 5.509.876 metros. Esto significa una disminución del 41% de las distancias totales generadas y representa una cuestión relevante ya que podemos decir que, en suma, la población entera está asignada significativamente más cerca a su centro de votación que antes.

2. Distancia promedio. En línea con el dato presentado anteriormente, la distancia promedio aplicando el modelo actual resulta ser 584 metros, descendiendo a 344 metros si se aplica el modelo propuesto en este trabajo. Esto refiere, naturalmente también, a una disminución del 41%.

3. Distancia promedio para los menores de 60 años. Este resulta un dato sumamente interesante porque, gracias a la búsqueda de minimizar las distancias totales y sin especificar requisitos sobre este grupo en las restricciones, también recorren menos con

el nuevo modelo, respecto al modelo actual. Este grupo dista 583 metros de su centro de votación para el caso del modelo actual y 359 metros al aplicarse la nueva propuesta. La caída de este promedio representa una disminución del 38%.

4. Distancia promedio para los mayores de 60 años. Este resulta un dato de principal interés en nuestro modelo, ya que lo hemos contemplado en las restricciones para que la asignación gane en términos de equidad. Con el modelo actual, este grupo arrojaba una distancia promedio de 587 metros y, con el nuevo modelo, dista en promedio 287 metros de su centro de votación. Aquí se observa una reducción de medias de un 51%.

5. Distancia promedio para las personas con CUD. Junto con el anterior, representa un grupo prioritario. Con la aplicación del modelo actual este grupo distaba en promedio, 589 metros de su centro de votación, respecto a su domicilio. Con la aplicación del nuevo modelo, esta media se redujo un 52%, resultando en una distancia promedio de 278 metros para este grupo.

6. Distancia promedio para las personas sin CUD. Al igual que con los menores de 60 años, este grupo se vio beneficiado como efecto colateral por la disminución de distancias totales. Con una reducción del 34%, pasó de 583 metros promedio a 348 metros promedio, con la aplicación del modelo propuesto.

7. Porcentaje de grupos convivientes asignados juntos. Considerando a este grupo como aquellas personas que comparten domicilio (calle, número y departamento) con al menos una persona más, vemos una mejora sustancial como resultado de nuestro propósito y aplicación del modelo. Con el modelo actual, aproximadamente un 15% de los grupos convivientes votaban todos juntos. Con el nuevo modelo se logró una asignación en la cual todos los convivientes estén asignados al mismo centro de votación.

De esta manera, hemos logrado aplicar y demostrar, con resultados, la viabilidad de un modelo que logró lo que se propuso: que los votantes tengan que recorrer menores distancias, que se consideren factores de interés como el hecho de poder ir a votar con los convivientes y que los grupos prioritarios tengan prioridad en términos de distancias generadas por la asignación del centro de votación.

A continuación se disponen los datos hallados para medir el éxito de la aplicación del nuevo modelo. El enfoque planteado recorre, desde la comparación de datos, hasta la satisfacción ciudadana y el impacto en la participación electoral. Éstos últimos, serán comparables en caso de aplicar el modelo a un caso real. Sin embargo, en lo que refiere a la comparación de datos se propone mirar con atención las distancias promedios por grupo, respecto a las totales, cuestión que aquí se puede abordar. Se comparan a continuación los datos previos y posteriores a la aplicación del modelo.

Modelo actual

$$\text{KPI 1: } \frac{\textit{distancia promedio recorrida por los mayores de 60 años}}{\textit{distancia promedio total recorrida}} = \frac{587}{584} \approx 1,01$$

$$\text{KPI 2: } \frac{\textit{distancia promedio recorrida por las personas con CUD}}{\textit{distancia promedio total recorrida}} = \frac{589}{584} \approx 1,01$$

Modelo propuesto

$$\text{KPI 1: } \frac{\textit{distancia promedio recorrida por los mayores de 60 años}}{\textit{distancia promedio total recorrida}} = \frac{287}{344} \approx 0,83$$

$$\text{KPI 2: } \frac{\textit{distancia promedio recorrida por las personas con CUD}}{\textit{distancia promedio total recorrida}} = \frac{278}{344} \approx 0,81$$

Con esto logramos demostrar que no sólo con el modelo propuesto los grupos prioritarios deban recorrer menos (con el modelo actual recorrían incluso más que la media) sino que es significativa la diferencia entre sus distancias y las generales.

5.2. Trabajos futuros

Este trabajo se enfoca en dar un primer paso en lo que respecta a una propuesta superadora al modelo actual de asignación. Se considera que, en este sentido, las oportunidades para mejorar este proceso actual son realmente significativas. Sin embargo, tras la implementación del modelo propuesto, se considera que sigue habiendo margen para mejorar las asignaciones. En primer lugar y de cara a futuros trabajos, se ve prometedor pensar en la estrategia de distribución de los electores a las mesas correspondientes. A los fines prácticos esta tesis abordó el problema hasta la puerta del establecimiento, porque aunque se disponibiliza un código para distribuir a los electores entre mesas como recurso complementario, no fue éste último el foco de la propuesta a abordar. Es por ello que se ve una interesante oportunidad de elaborar una estrategia de asignación en la cual se contemple, como mínimo, la accesibilidad de las mesas. Es interesante porque los mismos usuarios de estos modelos son los gobiernos con la capacidad de tratar a los datos que nutren estos abordajes. Nos referimos, en lo puntual, a que las jurisdicciones respectivas pueden, por ejemplo, catalogar las mesas para determinar si son accesibles o no y utilizar este mismo dato para luego optimizar la asignación y restringir al modelo para que siempre asigne mesas accesibles a las personas con CUD y/o mayores de edad. Además, la distribución podría considerar edades, para distribuir las uniformemente entre mesas, de modo tal que se promedien los tiempos de votación entre mesas.

Otro enfoque complementario e interesante que se podría abordar en futuros trabajos tras la implementación de este modelo, refiere a tratar no solamente las distancias sino también los tiempos destinados a la votación. Para ello y como ideal, incluso se pueden contemplar las distancias a recorrer considerando las barreras físicas que pueden hacer que las personas tengan que recorrer más de lo que la distancia en línea recta indica. En este sentido, se podría reversionar el trabajo actual calculando distancias en función de recorridos posibles e incorporando los tiempos de traslado y espera, donde la Teoría de Colas resultaría de especial utilidad para este último punto.

Resulta importante destacar que es clave aplicar y testear las nuevas propuestas, de modo tal que las innovaciones se vayan incorporando a modelos sólidos, con garantía de que funcionen. Se recomienda en este caso y por tratarse de gestión estatal con impacto en toda la sociedad, que se apliquen de manera prudente y se vayan incorporando las mejoras complementarias de a poco para poder ver el impacto que van generando e iterar en caso de ser necesario.

Sin duda, los avances pueden acompañar a mejoras en los procesos y es en lo que este trabajo se ha dedicado, generando impacto medido y sembrando bases para que surjan incluso propuestas superadoras con un enfoque en común: la prosperidad social.

REFERENCIAS

- Acemoglu-Robinson. (2008). *Por qué fracasan los países*. Booket.
- Aguilar-Barojas. (2005). Fórmulas para el cálculo de la muestra en investigaciones de salud. *Salud en Tabasco*, 11(1-2), 333-338.
- Bermúdez Colina. (2011). Aplicaciones de programación lineal, entera y mixta. *Ingeniería Industrial. Actualidad y Nuevas Tendencias*, 2(7), 85-104.
- Boirivant. (2009). La programación lineal aplicación de las pequeñas y medianas empresas. *Reflexiones*, 88(1), 89-105.
- Comisión Nacional Electoral. (2024). Coordenadas del polígono de Coghlan. <https://cdn.buenosaires.gob.ar/datosabiertos/datasets/subsecretaria-justicia/circuitos-electorales/circuitos-electorales.csv>.
- Durán, Giormenti, Guajardo, Pinto, Rey y Stier-Moses. (2008). Improving access to voting with optimized matching. *Estudios Electorales*, 51, 38-48.
- Gobierno de la Ciudad de Buenos Aires. (2025). *Comunas en la Web*. https://www.estadisticaciudad.gob.ar/eyc/?page_id=155983.
- Jefatura de Gabinete de Ministros. (2025). *Resultados Electorales*. argentina.gob.ar/dine/resultados-electorales.
- Lenzi. (2013). *Asignación eficiente de Votantes a Escuelas mediante Programación Matemática: El Caso de la Ciudad de Pergamino* [Tesis de grado, Universidad de Buenos Aires].
- Ley 19.945. (2024). *Código Electoral Nacional* (modificaciones al 2024).
- López. (2020). Importancia de los mecanismos de participación ciudadana en los sistemas democráticos. *Revista Ciencia Jurídica y Política*, 6(11), 31-44.

- Ministerio del Interior de la República Argentina. (2025). *Resultados Electorales*. resultados.mininterior.gob.ar/resultados.
- Moravec. (2014). *Asignación de votantes a centros de votación utilizando técnicas de Programación Matemática para minimizar el tiempo total invertido* [Tesis de grado, Universidad de Buenos Aires].
- Presidencia de la Nación. (2025). *Datos Estadísticos sobre Discapacidad en Argentina*. https://www.argentina.gob.ar/sites/default/files/2023/08/datos_estadisticos_sobre_discapacidad_en_argentina.pdf.

Apéndice A

1. Código para la creación de la población sintética

```
Python
# Instalaciones
import pandas as pd
import numpy as np
import random
import string
from shapely.geometry import Point, Polygon
# Replicamos la cantidad de votantes del distrito 130
total_votantes = 15998
# Establecemos las coordenadas de quiebre de Coghlan, lo que generará el
# polígono del distrito 130
coordenadas_coghlan = [
    (-58.4691816256656, -34.5589297649387), (-58.46702200601,
-34.5613060779578),
    (-58.4676779598418, -34.5616803291523), (-58.4665026556144,
-34.5621666326474),
    (-58.4675558734132, -34.5629646490661), (-58.4685735939025,
-34.5636248850675),
    (-58.470028964924, -34.5646434033567), (-58.4729353740278,
-34.5664107951477),
    (-58.473533382317, -34.566551670166), (-58.4746061543242,
-34.5668249509542),
    (-58.4751113736118, -34.5662132210955), (-58.476130079974,
-34.5668072106164),
    (-58.4771855010415, -34.5659649463914), (-58.4792917385781,
-34.5640518233097),
```

```

        (-58.4802254732172, -34.5632357868689), (-58.4811577687021,
-34.5624996308643),
        (-58.4835439186011, -34.5591200292072), (-58.4815288078594,
-34.5579629050645),
        (-58.4797097452666, -34.556990919046), (-58.4769912357401,
-34.5554517828398),
        (-58.4735907471273, -34.5535273727097), (-58.4724479768059,
-34.5550183535488),
        (-58.4708533616388, -34.5570326764231), (-58.4691816256656,
-34.5589297649387)
    ]
    poligono_coghlan = Polygon(coordenadas_coghlan)
    # Generamos una función para crear puntos aleatorios en el polígono, donde
    # luego ubicaremos a las personas
    def generar_puntos_aleatorios(polygona, num_puntos):
        puntos = []
        minx, miny, maxx, maxy = polygona.bounds
        while len(puntos) < num_puntos:
            x = random.uniform(minx, maxx)
            y = random.uniform(miny, maxy)
            punto = Point(x, y)
            if polygona.contains(punto):
                puntos.append((x, y))
        return puntos
    puntos_coghlan = generar_puntos_aleatorios(poligono_coghlan, total_votantes)

    # Se definen 3 funciones, para crear los datos personales de los usuario
    # como Nombre, Apellido y DNI
    def generar_nombre():
        return ''.join(random.choices('ABCDEFGHIJKLMNOPQRSTUVWXYZ', k=1)) +
        ''.join(random.choices('abcdefghijklmnopqrstuvwxyz', k=5))
    def generar_apellido():

```

```

        return ''.join(random.choices('ABCDEFGHIJKLMNOPQRSTUVWXYZ', k=1)) +
''.join(random.choices('abcdefghijklmnopqrstuvwxyz', k=7))
def generar_dni():
    return ''.join(random.choices('0123456789', k=8))
nombres = [generar_nombre() for _ in range(total_votantes)]
apellidos = [generar_apellido() for _ in range(total_votantes)]
dninúmeros = [generar_dni() for _ in range(total_votantes)]

# Seteamos un 40% de edificios para que los datos se asemejen más a la
realidad y se pueda demostrar la aplicabilidad de asignar teniendo en cuenta
también el piso y el dpto.
porcentaje_edificios = 0.4
num_en_edificios = int(total_votantes * porcentaje_edificios)
indices_edificios = np.random.choice(range(total_votantes),
num_en_edificios, replace=False)
pisos = [None] * total_votantes
deptos = [None] * total_votantes

# Se asignan tanto piso como dpto a los votantes
for i in indices_edificios:
    pisos[i] = random.randint(1, 10) # Pisos del 1 al 10
    deptos[i] = random.choice(list(string.ascii_uppercase[:4])) # Deptos
A-D

# En caso de que alguien no tenga edificio asignado, se le asigna 'Casa',
sólo a los fines de no dejar información sin completar.
for i in range(total_votantes):
    if pisos[i] is None and deptos[i] is None:
        pisos[i] = "Casa"
        deptos[i] = "Casa"

```

```

# Se determinan 12000 grupos para asegurarnos que algunas personas compartan
grupo conviviente con otras.
num_grupos = 12000
grupo_por_ubicacion = 4 # Como máximo 4 personas por grupo
combinaciones_posibles = []
for i in range(num_grupos):
    lat, lon = random.choice(puntos_coghlan)
    piso = random.choice([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
    depto = random.choice(['A', 'B', 'C', 'D'])
    combinaciones_posibles.append((lat, lon, piso, depto))

# Se asigna a cada persona a un grupo
personas_asignadas = []
while len(personas_asignadas) < total_votantes:
    combo = random.choice(combinaciones_posibles)
    if personas_asignadas.count(combo) < grupo_por_ubicacion:
        personas_asignadas.append(combo)

# Se asegura que haya un 20% de la población con 60 años o más
porcentaje_mayores_60 = 0.20
total_mayores_60 = int(total_votantes * porcentaje_mayores_60)
edades = ['60+' for _ in range(total_mayores_60)]
total_restantes = total_votantes - total_mayores_60
edades.extend(['Menor de 60' for _ in range(total_restantes)])

# Se mezclan los datos para generar aleatoriedad
random.shuffle(edades)

# Se determina que el 5% de la población tiene certificado de discapacidad
cud = np.random.choice([True, False], total_votantes, p=[0.05, 0.95])

# Se crea el dataframe con la población artificial

```

```

df_poblacion = pd.DataFrame({
    "Nombre": nombres,
    "Apellido": apellidos,
    "DNI": dninúmeros,
    "Longitud": [p[0] for p in personas_asignadas],
    "Latitud": [p[1] for p in personas_asignadas],
    "Piso": [p[2] for p in personas_asignadas],
    "Depto": [p[3] for p in personas_asignadas],
    "Edad": edades,
    "CUD": cud
})

# Se define grupo conviviente como aquellos que comparten Latitud, Longitud,
# piso y dpto
df_poblacion['grupo_conviviente'] = df_poblacion.groupby(['Longitud',
'Latitud', 'Piso', 'Depto']).ngroup()
df_poblacion['Convivencia'] =
df_poblacion.groupby('grupo_conviviente')['grupo_conviviente'].transform(lam
bda x: 'Con convivientes' if len(x) > 1 else 'Sin convivientes')

# Se realizan verificaciones rápidas del Dataframe
print(f"Distribución por Edad:\n{df_poblacion['Edad'].value_counts()}")
print(f"Distribución por CUD:\n{df_poblacion['CUD'].value_counts()}")
print(f"Cantidad mínima y máxima de personas por grupo:")
print(df_poblacion.groupby('grupo_conviviente').size().agg(['min', 'max']))

# Se exporta el CSV
df_poblacion.to_csv("poblacion_sintetica_coghlan_fin.csv", index=False)

# Se imprimen los encabezados del dataframe y primeros resultados para
# corroboración final
print(df_poblacion.head())

```

2. Código para el armado y resolución del modelo de asignación

```
Python
# Instalaciones
!pip install pulp geopy

import pandas as pd
import numpy as np
from geopy.distance import geodesic
from pulp import LpMinimize, LpProblem, LpVariable, lpSum, LpBinary, LpStatus

# Se realiza la carga de datos
votantes_df = pd.read_csv("poblacion_sintetica_coghlan_fin.csv") #CSV creado con el código anterior
establecimientos_df = pd.read_csv("Establecimientos130_fin.csv") #CSV disponibilizado por el
Gobierno de la Ciudad de Buenos Aires

# Para testeo se utilizó la siguiente línea:
#votantes_df = votantes_df.sample(n=500, random_state=42)

# Se realizan preprocesamientos de la información para contar con booleanos de los datos de interés
votantes_df["es_cud"] = votantes_df["CUD"] == True
votantes_df["mayor_60"] = votantes_df["Edad"] == "60+"
votantes_df["index"] = votantes_df.index
establecimientos_df["index"] = establecimientos_df.index

# Se define la capacidad de cada establecimiento
establecimientos_df["capacidad_total"] = establecimientos_df["cant_mesas"] *
establecimientos_df["capacidad_mesa"]

# Se calculan las distancias entre domicilio y centro de votación
distancias = {}
for i, v in votantes_df.iterrows():
    for j, e in establecimientos_df.iterrows():
        origen = (v["Latitud"], v["Longitud"])
        destino = (e["Latitud"], e["Longitud"])
```

```

distancias[(v["index"], e["index"])] = geodesic(origen, destino).km

# Se definen los conjuntos de votantes y establecimientos
N = votantes_df["index"].tolist()
M = establecimientos_df["index"].tolist()

# Se define un modelo de programación lineal
model = LpProblem("Asignacion_Votantes", LpMinimize)

# Se detallan las variables de decisión
x = LpVariable.dicts("x", (N, M), cat=LpBinary)
z = LpVariable.dicts("z", N, cat=LpBinary)
PCUD = LpVariable("PCUD", lowBound=0)
P60 = LpVariable("P60", lowBound=0)

# Se setean los parámetros
delta = 300
alfa = 10000
beta = 10000

# Se asienta la función objetivo
model += (
    lpSum(distancias[i, j] * x[i][j] for i in N for j in M)
    + alfa * PCUD + beta * P60 + delta * lpSum(z[i] for i in N)
)

# Se configuran las restricciones del modelo
for i in N:
    model += lpSum(x[i][j] for j in M) == 1

for j in M:
    model += lpSum(x[i][j] for i in N) <= establecimientos_df.loc[j, "capacidad_total"]

cud_ids = votantes_df[votantes_df["es_cud"]]["index"].tolist()
no_cud_ids = votantes_df[~votantes_df["es_cud"]]["index"].tolist()

```

```

if cud_ids and no_cud_ids:
    model += (
        lpSum(distancias[i, j] * x[i][j] for i in cud_ids for j in M) / len(cud_ids)
        <= (0.8)*(lpSum(distancias[i, j] * x[i][j] for i in no_cud_ids for j in M) / len(no_cud_ids)) + PCUD
    )

mayor_ids = votantes_df[votantes_df["mayor_60"]]["index"].tolist()
menor_ids = votantes_df[~votantes_df["mayor_60"]]["index"].tolist()
if mayor_ids and menor_ids:
    model += (
        lpSum(distancias[i, j] * x[i][j] for i in mayor_ids for j in M) / len(mayor_ids)
        <= (0.8)*(lpSum(distancias[i, j] * x[i][j] for i in menor_ids for j in M) / len(menor_ids)) + P60
    )

grupo_conv = votantes_df.dropna(subset=["grupo_conviviente"])
grupos = grupo_conv.groupby("grupo_conviviente")
for _, grupo in grupos:
    indices = grupo["index"].tolist()
    for i in indices:
        for k in indices:
            if i < k:
                for j in M:
                    model += x[i][j] <= x[k][j] + z[i]
                    model += x[k][j] <= x[i][j] + z[k]

# Se resuelve el modelo y se da status finalizada la resolución
model.solve()
print(f"Estado del modelo: {LpStatus[model.status]}")

# Se realiza la exportación de datos
asignaciones = []
for i in N:
    for j in M:
        if x[i][j].varValue == 1:
            asignaciones.append((i, j))

```

```

asignaciones_df = pd.DataFrame(asignaciones, columns=["index_votante", "index_establecimiento"])
asignaciones_df = asignaciones_df.merge(votantes_df, left_on="index_votante", right_on="index")

# Se eliminan columnas duplicadas
asignaciones_df = asignaciones_df.drop(columns=["index_votante"])

# Se incorpora la información de los establecimientos a las asignaciones
asignaciones_df = asignaciones_df.merge(establecimientos_df, left_on="index_establecimiento",
right_on="index", suffixes=("_votante", "_establecimiento"))

# Se deja asentada la distancia entre domicilio del votante y su establecimiento asignado
asignaciones_df["distancia_m"] = asignaciones_df.apply(
    lambda row: geodesic(
        (row["Latitud_votante"], row["Longitud_votante"]),
        (row["Latitud_establecimiento"], row["Longitud_establecimiento"])
    ).km * 1000,
    axis=1
)

# Se guarda el CSV
asignaciones_df.to_csv("asignaciones_con_distancias.csv", index=False)

# Se muestran resultados de interés como previsualización
asignaciones_df[["id_votante", "Nombre", "Apellido", "nombre_establecimiento", "distancia_km"]].head()

```

3. Código para representar el modelo de asignación actual

```
Python
# Importaciones
import pandas as pd
from geopy.distance import geodesic

# Se importan los datos de población y establecimientos
votantes_df = pd.read_csv('poblacion_sintetica_coghlan_fin.csv')
establecimientos_df = pd.read_csv('Establecimientos130.csv')

# Se determina la capacidad de los establecimientos
capacidad_mesa = 350 # Cantidad de votantes por mesa
establecimientos_df['capacidad_total'] = establecimientos_df['cant_mesas'] * capacidad_mesa

# Se ordenan los votantes alfabéticamente por apellido
votantes_df = votantes_df.sort_values(by='Apellido')

# Se asignan votantes a establecimientos en orden, en función de las capacidades
asignaciones = []
index_establecimiento = 0
cupos_disponibles = establecimientos_df.iloc[index_establecimiento]['capacidad_total']
for _, votante in votantes_df.iterrows():
    if cupos_disponibles == 0:
        index_establecimiento += 1
        if index_establecimiento >= len(establecimientos_df):
            print("No hay suficientes cupos en los establecimientos.")
            break
        cupos_disponibles = establecimientos_df.iloc[index_establecimiento]['capacidad_total']

    establecimiento = establecimientos_df.iloc[index_establecimiento]

# se calculan las distancias entre domicilios y centros de votación asignados
distancia_m = geodesic(
    (votante['Latitud'], votante['Longitud']),
    (establecimiento['Latitud'], establecimiento['Longitud'])
).meters
```

```
asignaciones.append({
    'id_votante': votante['id_votante'],
    'Apellido': votante['Apellido'],
    'Nombre': votante['Nombre'],
    'Edad': votante['Edad'],
    'CUD': votante['CUD'],
    'Grupo_Conviviente': votante['grupo_conviviente'],
    'Latitud_Votante': votante['Latitud'],
    'Longitud_Votante': votante['Longitud'],
    'id_establecimiento': establecimiento['id_establecimiento'],
    'Nombre_Establecimiento': establecimiento['nombre_establecimiento'],
    'Latitud_Establecimiento': establecimiento['Latitud'],
    'Longitud_Establecimiento': establecimiento['Longitud'],
    'Distancia_Metros': distancia_m
})
```

```
cupos_disponibles -= 1
```

Se configura el dataframe y se guarda en un CSV

```
asignaciones_df = pd.DataFrame(asignaciones)
asignaciones_df.to_csv('asignaciones.csv', index=False)
```

4. Código para revisar resultados de asignación

```
Python
# Importaciones
import pandas as pd

# Carga del CSV con las asignaciones
asignaciones_df = pd.read_csv("asignaciones.csv")

# Se calcula la distancia total y promedio
distancia_total = asignaciones_df["Distancia_Metros"].sum()
distancia_promedio = asignaciones_df["Distancia_Metros"].mean()

# Se calcula la distancia total y promedio para mayores de 60
distancia_menores_60 = asignaciones_df[asignaciones_df["Edad"] == "Menor de 60"]["Distancia_Metros"]
distancia_mayores_60 = asignaciones_df[asignaciones_df["Edad"] == "60+"]["Distancia_Metros"]

# Se calcula la distancia total y promedio para votantes con CUD
distancia_con_cud = asignaciones_df[asignaciones_df["CUD"] == True]["Distancia_Metros"]
distancia_sin_cud = asignaciones_df[asignaciones_df["CUD"] == False]["Distancia_Metros"]

# Se enumeran los grupos convivientes con más de 1 integrante
grupo_conviviente_counts = asignaciones_df.groupby("Grupo_Conviviente")["id_votante"].count().reset_index(name="cantidad_personas")

# Se filtran únicamente los grupos convivientes que no son unipersonales
grupos_multipersona = grupo_conviviente_counts[grupo_conviviente_counts["cantidad_personas"] > 1]

# Se obtienen los establecimientos distintos asignados a un mismo grupo conviviente
establecimientos_por_grupo = asignaciones_df.groupby("Grupo_Conviviente")["id_establecimiento"].nunique().reset_index(name="establecimientos_distintos")

# Se filtran los grupos convivientes multipersona que votan en el mismo establecimiento
grupos_multipersona = grupos_multipersona.merge(establecimientos_por_grupo, on="Grupo_Conviviente")
```

```

grupos_juntos = grupos_multipersona[grupos_multipersona["establecimientos_distintos"] == 1].shape[0]

# Se cuenta la cantidad de grupos multipersona
total_grupos_multipersona = grupos_multipersona.shape[0]

# Se calcula el porcentaje de grupos convivientes multipersona que votaron todos juntos
porcentaje_grupos_juntos = (grupos_juntos / total_grupos_multipersona) * 100 if
total_grupos_multipersona > 0 else 0

# Se imprime el resumen
resumen = {
    "Distancia total recorrida (m)": round(distancia_total),
    "Distancia promedio recorrida (m)": round(distancia_promedio),
    "Distancia promedio recorrida por votantes de hasta 60 años (m)":
round(distancia_menores_60.mean()),
    "Distancia promedio recorrida por mayores de 60 años (m)": round(distancia_mayores_60.mean()),
    "Distancia promedio con CUD (m)": round(distancia_con_cud.mean()),
    "Distancia promedio sin CUD (m)": round(distancia_sin_cud.mean()),
    "Porcentaje de grupos convivientes (de más de 1 persona) que votaron todos juntos":
f"{porcentaje_grupos_juntos:.2f}%"
}
for k, v in resumen.items():
    print(f"{k}: {v}\n")

```

Apéndice B

1. Código para asignar mesas dentro de los establecimientos

```
Python
import pandas as pd

# Se carga la información de las asignaciones y de los establecimientos
df = pd.read_csv("asignaciones_con_distancias.csv")

# Se depura el formato de las mesas, que tiene una X al comienzo
df["mesa_desde_num"] = df["mesa_desde"].astype(str).str.replace("X", "").astype(int)

# Se crea una lista para guardar las asignaciones
asignaciones = []

# Se asigna la persona a una mesa de su establecimiento si tiene capacidad
for est_id, grupo in df.groupby("index_establecimiento"):
    grupo = grupo.sort_values("id_votante").reset_index(drop=True)

    mesa_inicio = grupo["mesa_desde_num"].iloc[0]
    capacidad = grupo["capacidad_mesa"].iloc[0]

    for i, (_, row) in enumerate(grupo.iterrows()):
        nro_mesa = mesa_inicio + (i // capacidad)
        mesa_str = f"{nro_mesa:}X"
        asignaciones.append({
            "id_votante": row["id_votante"],
            "nombre_establecimiento": row["nombre_establecimiento"],
            "mesa_asignada": mesa_str
        })

# Se crea un dataframe
df_asignados = pd.DataFrame(asignaciones)
```

```
# Se previsualizan resultados
```

```
print(df_asignados.head())
```

```
# Se guarda como CSV
```

```
df_asignados.to_csv("asignaciones_con_mesa.csv", index=False)
```

2. Cálculo estadístico para decidir el tamaño de la muestra.

$$n = \frac{N \cdot Z^2 \cdot p \cdot q}{d^2(N - 1) + Z^2 \cdot p \cdot q}$$

Donde,

n es el tamaño de la muestra, que se intenta hallar.

N es el tamaño de la población, en este caso 15998 ya que es la población total del circuito 130.

Z es el valor crítico, que se encuentra en las tablas de área de curva normal. Representa el valor de confianza.

p es la variabilidad máxima. Cuando no se cuenta con información puntual sobre este dato, se puede asumir 0,5, como haremos en este caso.

q es el valor que resulta de restar p a 1. En este caso, en el que asumimos que p es 0,5, q también resulta 0,5.

d es el margen de error tolerado. Es común asumirlo entre 3% y 5% si no se tiene un interés puntual sobre esto. Para este caso, lo determinamos en un 5%

Reemplazando los valores tenemos que,

$$n = \frac{15998 \cdot 1,96^2 \cdot 0,5 \cdot 0,5}{0,05^2(15998 - 1) + 1,96^2 \cdot 0,5 \cdot 0,5}$$

$$n = \frac{15364,4762}{40,9529}$$

$$n \approx 375$$

Entonces, para una muestra que represente un nivel de confianza del 95%, en una población de 15998 votantes, se recomienda una muestra de 375 personas a las que someter a la encuesta. Como recomendación adicional, se cree oportuno que se duplique el tamaño de la muestra necesaria para tener representatividad, por si no todas las personas responden. Suponiendo que, como máximo, una de cada dos personas deciden no responder, **la muestra mínima se recomienda en 750 personas.**

Todo esto en función de lo expuesto en el artículo 'Fórmulas para el cálculo de la muestra en investigaciones de salud, (Aguilar-Barojas, 2005)'.

3. Código para seleccionar de manera aleatoria a la muestra de la encuesta.

```
Python
# Importaciones
import pandas as pd

# Se crea una función para seleccionar 750 personas dentro del CSV
def seleccionar_muestra(csv_path, tamaño_muestra=750):
    df = pd.read_csv(csv_path)
    muestra = df.sample(n=tamaño_muestra, random_state=42)
    return muestra

# Se carga la información necesaria para que la función elija 750 personas
csv_path = "poblacion_sintetica_coghlan_fin.csv"

# Se aplica la función al CSV en cuestión para obtener la muestra
muestra_seleccionada = seleccionar_muestra(csv_path)

# Se guardan las personas seleccionadas en un CSV
muestra_seleccionada.to_csv("muestra_750.csv", index=False)
```