Trabajo de Fin de Grado en Ingeniería de Tecnologías Industriales

Programación matemática para la resolución de un problema de asignación de parcel lockers

Autor: Miguel Marin Castellano

Tutor: Antonio Lorenzo Espejo

Dpto. Organización Industrial y Gestión de Empresas II

Escuela Técnica Superior de Ingeniería

Sevilla, 2025

Trabajo de Fin de
Grado en Ingeniería de Tecnologías Industriales

Programación matemática para la resolución de un problema de asignación de parcel lockers

Autor:
Miguel Marin Castellano

Tutor:
Antonio Lorenzo Espejo
Profesor Sustituto Interino

Dpto. de Organización Industrial y Gestión de Empresas II
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla
Sevilla, 2025

Trabajo Fin de Grado: Programación matemática para la resolución de un problema de asignación de parcel lockers

Autor: Miguel Marin Castellano

Tutor: Antonio Lorenzo Espejo

El tribunal nombrado para juzgar el Trabajo arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2025

El Secretario del Tribunal

A mi familia

A mis maestros

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que, de una u otra forma, han contribuido a la realización de este Trabajo de Fin de Grado.

En primer lugar, agradezco a mi tutor/a, Antonio Lorenzo Espejo, por su guía, paciencia y valiosos consejos durante todo el proceso. Su orientación ha sido fundamental para el desarrollo de este proyecto.

También quiero agradecer a mis profesores y compañeros de la Escuela Técnica Superior de Ingeniería, quienes me han proporcionado conocimientos y apoyo a lo largo de estos años de formación académica.

A mi familia, por su amor incondicional y por haberme apoyado en cada paso de este camino. Gracias por su comprensión y por motivarme en los momentos de mayor dificultad.

Finalmente, a mis amigos y seres queridos, por su compañía, por las palabras de aliento y por recordarme siempre la importancia de la perseverancia y el esfuerzo.

Este trabajo es el resultado de un largo recorrido de aprendizaje, esfuerzo y dedicación, y a todos ustedes, les estaré eternamente agradecido.

Miguel Marin Castellano

Sevilla, 2025

Resumen

La logística de última milla es un desafío crucial en la optimización de las cadenas de suministro, particularmente en entornos urbanos donde la eficiencia en la distribución de paquetes es clave para reducir costes y mejorar la sostenibilidad. En este trabajo se ha llevado a cabo un análisis exhaustivo de diferentes enfoques y soluciones propuestas en la literatura académica para la optimización de la última milla, con un especial enfoque en el uso de Parcel Lockers como alternativa eficiente.

Como base teórica, se han revisado diversos estudios relacionados con la ubicación óptima de los Automated Parcel Lockers (APL), incluyendo los modelos matemáticos desarrollados por Deutsch y Golany (2017) y Ottaviani et al. (2023), para evaluar la aplicabilidad de estas metodologías en el contexto del problema abordado. Tras la selección del modelo más adecuado, se han realizado modificaciones que permiten una mejor adaptación a las necesidades específicas del problema planteado.

En la fase metodológica, se ha llevado a cabo un tratamiento de datos utilizando la técnica de Agglomerative Clustering, con el objetivo de identificar las mejores ubicaciones para los Parcel Lockers en función de la demanda y la distribución geográfica de los usuarios. Posteriormente, se ha implementado el modelo matemático modificado en el software de optimización Gurobi, empleando Python como lenguaje de programación para su resolución computacional.

Una vez obtenidos los resultados iniciales, se ha procedido a un análisis de sensibilidad para evaluar la robustez del modelo ante cambios en distintos parámetros. En primer lugar, se ha analizado la sensibilidad respecto al coste fijo de instalación de los APL, permitiendo observar su impacto en la selección de ubicaciones y en la función objetivo. Posteriormente, se ha evaluado la influencia de las modificaciones en la matriz de costes de asignación, lo que ha permitido determinar la estabilidad del modelo frente a variaciones en los costes de distribución. Finalmente, se ha realizado un análisis de sensibilidad en función de la disponibilidad de ubicaciones, examinando cómo la restricción de ciertos puntos de instalación afecta la configuración óptima del sistema.

Los resultados obtenidos han permitido validar la efectividad del modelo propuesto y sus modificaciones, así como proporcionar información relevante sobre la influencia de distintos parámetros en la optimización de la ubicación de Parcel Lockers. Este estudio contribuye a la toma de decisiones en la logística de última milla, facilitando la implementación de soluciones más eficientes y sostenibles en entornos urbanos.

Abstract

Last mile logistics is a crucial challenge in the optimization of supply chains, particularly in urban environments where efficiency in parcel distribution is key to reduce costs and improve sustainability. In this paper, a comprehensive analysis of different approaches and solutions proposed in the academic literature for last mile optimization has been carried out, with a special focus on the use of Parcel Lockers as an efficient alternative.

As a theoretical basis, several studies related to the optimal placement of Automated Parcel Lockers (APL), including the mathematical models developed by Deutsch and Golany (2017) and Ottaviani et al. (2023), have been reviewed to assess the applicability of these methodologies in the context of the addressed problem. After the selection of the most suitable model, modifications have been made to allow a better adaptation to the specific needs of the problem addressed.

In the methodological phase, data processing was carried out using the Agglomerative Clustering technique, with the aim of identifying the best locations for Parcel Lockers based on the demand and geographical distribution of users. Subsequently, the modified mathematical model was implemented in the optimization software Gurobi, using Python as the programming language for its computational resolution.

Once the initial results were obtained, a sensitivity analysis was performed to evaluate the robustness of the model to changes in different parameters. First, the sensitivity to the fixed installation cost of the APLs was analysed, allowing us to observe its impact on the selection of locations and on the objective function. Subsequently, the influence of the changes in the allocation cost matrix was evaluated, which made it possible to determine the stability of the model in the face of variations in distribution costs. Finally, a sensitivity analysis has been performed according to the availability of locations, examining how the restriction of certain installation points affects the optimal configuration of the system.

The results obtained have allowed validating the effectiveness of the proposed model and its modifications, as well as providing relevant information on the influence of different parameters on the optimization of Parcel Lockers location. This study contributes to decision making in last mile logistics, facilitating the implementation of more efficient and sustainable solutions in urban environments.

Índice

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1 INTRODUCCIÓN

El crecimiento del comercio electrónico ha generado un aumento exponencial en la demanda de soluciones logísticas eficientes para la última milla. Este segmento de la cadena de suministro, que abarca la entrega final del producto al consumidor, enfrenta numerosos desafíos, como la optimización de rutas, la reducción de costes y el impacto ambiental. En este contexto, los APL (Automated Parcel Lockers), sistemas de casilleros automatizados que permiten la recepción y recogida de paquetes sin necesidad de interacción con un repartidor, han surgido como una solución viable para mejorar la eficiencia de la distribución urbana. Al estar ubicados en puntos estratégicos y permitir la recogida flexible mediante códigos de acceso, los APL reducen la cantidad de intentos fallidos de entrega y optimizan la gestión del tiempo y los recursos.

Este trabajo se centra en el problema de ubicación de APL, una cuestión clave para maximizar su efectividad y accesibilidad. Para ello, se ha realizado una investigación exhaustiva sobre la logística de última milla, explorando diferentes enfoques y soluciones alternativas que han sido propuestas en la literatura académica. En particular, se ha estudiado el modelo matemático de localización de APL propuesto por Deutsch y Golany en 2017, adaptándolo a las necesidades específicas del problema abordado.

El desarrollo metodológico de este estudio ha implicado el uso de la técnica de clustering jerárquico aglomerativo para el tratamiento de datos, permitiendo agrupar zonas con características similares y facilitando la toma de decisiones en la selección de ubicaciones. Posteriormente, se han aplicado modificaciones al modelo matemático original para ajustarlo a las particularidades del caso de estudio, considerando factores como los costes de instalación y asignación, así como la disponibilidad de ubicaciones.

Los resultados obtenidos han sido analizados mediante simulaciones y un análisis de sensibilidad que evalúa el impacto de distintas variables en la solución final. Se han generado distintos escenarios que permiten comprender cómo afectan los cambios en los costes y la disponibilidad de ubicaciones a la eficiencia del sistema propuesto.

Este documento está estructurado de la siguiente manera: en primer lugar, se presenta una revisión de la literatura sobre logística de última milla y problemas de localización de APL. Posteriormente, se describe la metodología utilizada, incluyendo el modelo matemático y las técnicas de análisis de datos empleadas. A continuación, se exponen los resultados obtenidos y se lleva a cabo un análisis de sensibilidad para evaluar la robustez del modelo. Finalmente, se presentan las conclusiones del estudio, destacando las contribuciones realizadas y posibles futuras líneas de investigación en esta área.

2 ESTADO DEL ARTE

En el ámbito del comercio electrónico, la logística de última milla emerge como un campo de constante desafío y evolución. Enfrentando presiones de eficiencia, costos y satisfacción del cliente, las empresas buscan constantemente soluciones innovadoras para optimizar sus operaciones de entrega. Este capítulo se enfoca en el estado actual de la logística de última milla, identificando sus principales desafíos y explorando los lockers como una alternativa prometedora. Además se estudia el problema de ubicación de parcel lockers a través de la implementación de dos métodos de optimización.

2.1. Logística de última milla

La logística dentro de la cadena de suministro es un componente fundamental que se encarga de planificar, implementar y controlar el flujo eficiente y efectivo de bienes. Esta función es esencial para garantizar que los productos estén disponibles en el lugar correcto, en el momento adecuado y en las condiciones adecuadas, buscando siempre la satisfacción del cliente con los menores costos operativos posibles. Como parte de la logística, se encuentra la última milla, la cual se define como el tramo final de un servicio de entrega de la empresa al consumidor, ya sea en su domicilio o en un punto de recogida (Gevaers et al., 2009).

Hoy en día, el comercio electrónico desempeña un papel vital en la mayoría de las empresas globales, lo que requiere esfuerzos significativamente mayores por parte del personal de logística y de la cadena de suministro. Las empresas tienen que afrontar nuevos desafíos y esforzarse por descubrir soluciones innovadoras para mantener su ventaja competitiva en el mercado y, al mismo tiempo, lograr la máxima satisfacción del cliente. Muchos académicos coinciden en que la etapa más crucial dentro de una determinada cadena de suministro es la “entrega de última milla”, ya que representa alrededor del 30 % al 50 % de los costos totales de transporte junto con su complejidad y falta de eficiencia (Yuen et al., 2018).

2.1.1 Desafíos

La logística de última milla es la parte más contaminante, menos eficiente y más costosa de la cadena de suministro. Como resultado, es difícil asegurar su sostenibilidad debido a la complejidad del clima urbano y las actividades económicas (Ehmke et al,.2012).

En esta sección, se han organizado y discutido los principales desafíos que enfrenta la logística de última milla considerando diferentes aspectos como el aspecto infraestructural, el aspecto tecnológico, la gestión y los costos logísticos.

2.1.1.1 Infraestructuras

En los últimos años, tanto las Administraciones Públicas como los distribuidores y empresas de paquetería han tenido que invertir en la mejora y optimización de la infraestructura dedicada a la logística de última milla. Esta infraestructura abarca todos los activos logísticos, tanto públicos como privados (centros de distribución, almacenes, taquillas inteligentes, zonas de carga y descarga, redes de carreteras, aeropuertos, etc.), que forman la red necesaria para realizar las entregas de última milla. (Deloitte,2020).

Las infraestructuras son la base de cualquier ciudad para el transporte de mercancías. La estructura y situación geográfica de la ciudad incide en las actividades logísticas de última milla en términos de distancia, acceso y espacio. La eficiencia de los sistemas logísticos está influenciada por varias cuestiones como son la estrechez de las calles urbanas, el estacionamiento de vehículos de reparto en segundos carriles, las restricciones al tráfico de vehículos y, finalmente, la falta de instalaciones logísticas para carga y descarga (El Moussaoui et al., 2022).

La estructura y ubicación geográfica de una ciudad afectan a las actividades logísticas, ya que existen problemas relacionados con el espacio, el acceso y la distancia en las áreas urbanas. Debido a los impedimentos anteriormente nombrados, no es fácil para las ciudades cambiar su infraestructura viaria para hacer frente al creciente volumen de carga (Bosona , 2020).

2.1.1.2 Tecnología

Existen numerosas tecnologías y nuevas empresas tecnológicas que están revolucionando la cadena de suministro. Es fundamental integrarlas en el proceso de digitalización de la logística de última milla para poder satisfacer las altas demandas de los consumidores. Entre estas tecnologías se encuentran, el Big Data, los lockers de paquetes o los drones (Deloitte, 2020).

En la siguiente figura, se han categorizado las tecnologías identificadas en función de su impacto en los retos de la logística de última milla y el tiempo necesario para su implementación.

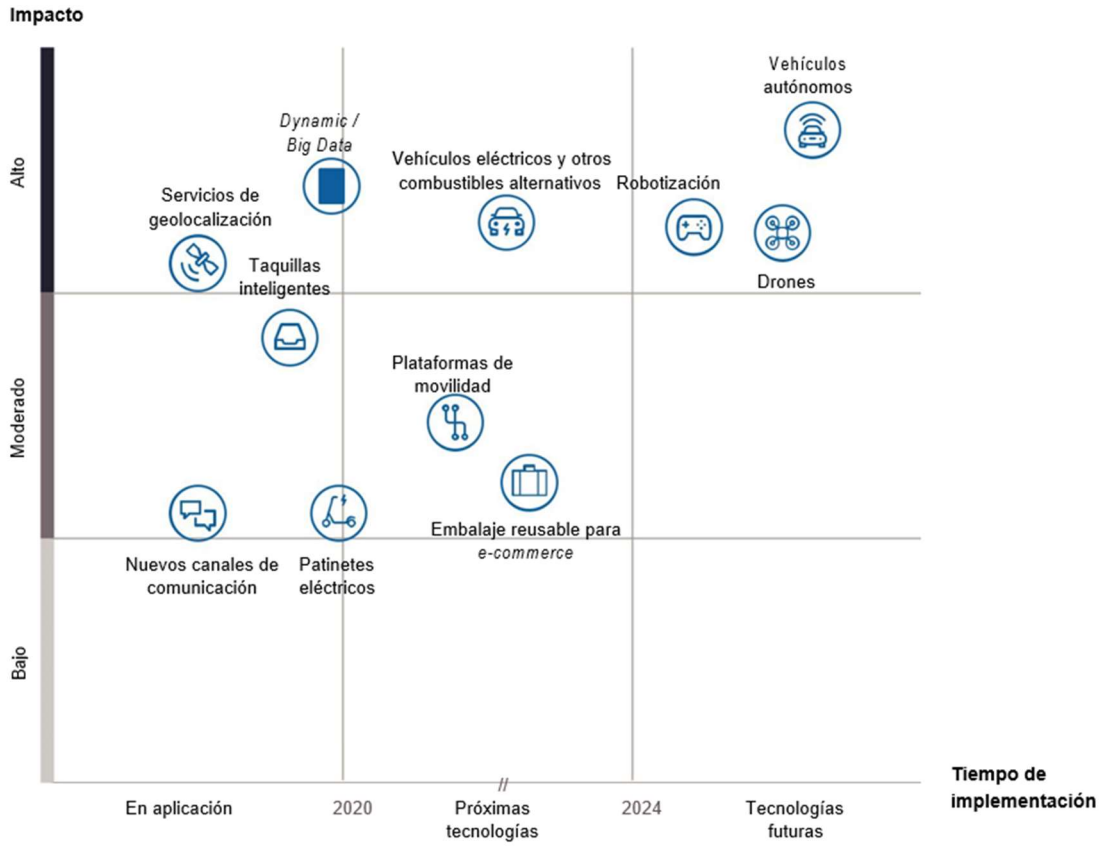

Figura 1 – Impacto de las tendencias tecnológicas (Deloitte, 2020)

2.1.1.3 Gestión

La logística de última milla es el elemento más complejo de la gestión de la cadena de suministro. Sin embargo, esta parte recibe poca atención relativa a su importancia. Si bien se han propuesto soluciones tecnológicas para evitar los problemas relacionados con la logística de última milla, no existe una solución holística que pueda cubrir estos problemas, debido a los factores determinados por la evolución de las ciudades como son el estilo de vida, el desarrollo económico, las cuestiones climáticas y ambientales, la regulación en materia de peajes y estacionamientos, los desplazamientos y la accesibilidad a los servicios.

La entrega de última milla está marcada por situaciones complejas en las que es difícil coordinar a los actores de la cadena de suministro, lo que hace difícil desarrollar sistemas de distribución eficientes en circunstancias urbanas específicas. Por otro lado, la falta de interacción entre los actores logísticos de última milla (proveedores, consumidores, transportistas y autoridades locales) puede contribuir significativamente a la complejidad de esta parte de la cadena de suministro, especialmente si existe un conflicto de intereses entre las partes interesadas (El Moussaoui et al., 2022).

Un problema relacionado con la gestión del servicio es el fallo en la entrega de bienes, lo que está asociado con impactos negativos (Tabla 1). Por ejemplo, Dell'Amico et al. (2012), discutió que las tasas de fallo de entre el 10% y el 50% podrían aumentar las emisiones de CO_2 entre un 15% y un 75%, respectivamente.

Visser et al. (2014), analizó que la tasa de entrega por primera vez es de aproximadamente el 12%, mientras que alrededor del 2% no se puede entregar en absoluto y se devuelve al centro de distribución central.

Tabla 1– Tasa de fracaso en la entrega a domicilio de comercio minorista (Bosona,2020)

Artículo	Tasa de fallo en la entrega (%)
Libros	3
Pequeños artículos	5 - 10
Ropa de moda	20 - 44
Envasados	34

2.1.1.4 Coste

Según Ehmke et al. (2012), el transporte de última milla supone alrededor del 28% del coste total de la entrega de mercancías. Los costos logísticos están relacionados principalmente con el modo de distribución, el nivel de servicio, la seguridad, el impacto ambiental, la gestión de la flota, la concienciación del cliente y el área geográfica del punto de destino. En algunos casos puede haber costes adicionales entregas repetidas, debidas a fallos en la primera entrega.

No solo el coste de la entrega, sino también la estimación y planificación de costos para la entrega de mercancías no es fácil, ya que está asociado con muchas limitaciones relacionadas con la capacidad de carga, las horas de trabajo del conductor, etc. Otra dimensión de la complejidad del servicio es que el creciente comercio electrónico tiene bajos márgenes de beneficio para las empresas, mientras que los clientes esperan una mayor calidad del servicio (Bosona,2020).

Especialmente, la alta demanda de pedidos a domicilio genera un incremento en los costos logísticos. Las empresas de paquetería experimentan una tasa de fracaso del 10% al 15% en el primer intento de entrega a domicilio, lo que duplica los costos al tener que realizar una segunda entrega. (Deloitte,2020).

Un caso de estudio realizado por Lorenzo Espejo et al.,(2024) se basa en el análisis de 12 transportistas que operan en 12 provincias españolas (Madrid, Barcelona, Tarragona, Sevilla, Granada, Pontevedra, Jaén, Cádiz, Córdoba, Huelva, Almería y Lleida).

En este trabajo se propone un modelo de análisis de costes del transporte de mercancías de última milla, basado en una simulación de Montecarlo que combina las estimaciones de costos con un muestreo aleatorio utilizando 10.000 simulaciones

Los efectos de la variabilidad en los costos de transporte se pueden observar fácilmente a través del análisis de dos variables: el costo por kilómetro y el costo por servicio. La Figura 2 muestra la distribución de estas dos variables tras la aplicación del método de simulación de Montecarlo.

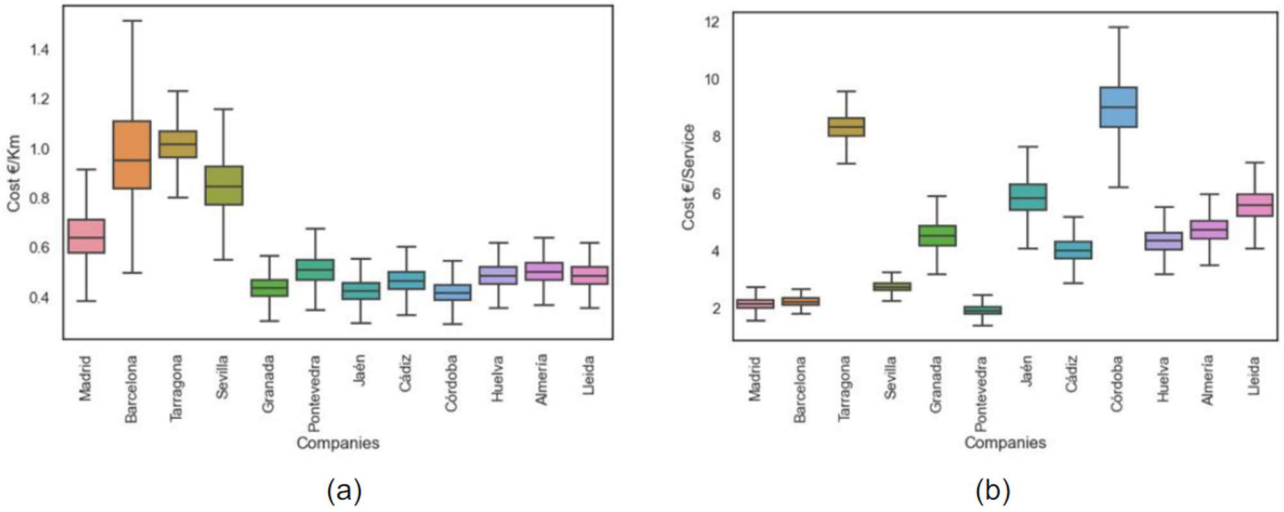

Figura 2– Coste por kilómetro (a) y coste por servicio (b) de los transportistas (Lorenzo Espejo et al.,2024)

Como se observa en la Fig. 2a, los transportistas que operan en ciudades como Madrid, Barcelona, Tarragona o Sevilla enfrentan costos por kilómetro más altos y variables. En estas ciudades, el costo total se ve significativamente influenciado por la mayor densidad de población y, por consiguiente, la densidad de servicios.

Debido a que en las ciudades con mayor densidad de servicios los mensajeros pueden incluir más paradas en sus rutas, el consumo de combustible por lugar visitado es menor. Esto también reduce el costo total por servicio, dada la alta proporción del costo global que representan los gastos de combustible. El análisis realizado permite estimar los costos logísticos en los que incurren las diferentes empresas para los servicios de entrega y recogida de última milla. (Lorenzo Espejo et al., 2024).

Con dicho análisis se llega a la conclusión del gran impacto que tiene la ultima milla en los costes logísticos totales.

2.1.2 Soluciones alternativas

Con el fin de abordar los desafíos de la logística de última milla, un estudio de Deloitte (2020) analiza cuatro modelos logísticos. Estos modelos se evalúan en función de la sostenibilidad medioambiental ($kgCO_2$ emitidos por paquete), la congestión urbana (vehículos/km^2) y la eficiencia logística (€/paquete).

2.1.2.1 Hubs urbanos

El modelo de hubs urbanos sugiere la instalación de diversos tipos de almacenes dentro de las ciudades, donde se realice el cross-docking (recopilación y clasificación) de la mercancía para enviarla desde estos almacenes hasta los puntos de entrega. El objetivo principal es acercar la mercancía a su destinatario, agilizando el proceso de suministro y cumpliendo con los altos niveles de servicio demandados por los consumidores. Además, la posibilidad de usar vehículos de reparto más ecológicos y de menor tamaño permite mejorar la sostenibilidad medioambiental y reducir la congestión urbana.

Figura 3– Impacto de los hubs urbanos en los retos de Última Milla (Deloitte, 2020)

2.1.2.2 Red de puntos de entrega

La red de puntos de entrega incluye dos tipos de PUDOs (pick up, drop-off points): los puntos de conveniencia (utilización de establecimientos y locales, como tiendas o gasolineras, para entregar y recoger pedidos) y las taquillas inteligentes. Aunque ambas opciones son conceptualmente similares, presentan diferencias. Por ejemplo, los puntos de conveniencia permiten la entrega de productos de mayores dimensiones que no caben en las taquillas inteligentes debido a su menor tamaño.

Actualmente, este modelo es poco común en España, donde solo el 10% de las entregas se realizan en puntos de entrega, y solo un 4% de estas se hacen en taquillas inteligentes (Deloitte, 2020). En el caso de las taquillas inteligentes, no todos los competidores tienen acceso a ellas, ya que son propiedad de algunos distribuidores y no forman parte de una red global compartida. De hecho, solo el 31% de los comercios online ofrecen la opción de entrega en PUDOs.

Figura 4– Impacto de la red de puntos de entrega en los retos de Última Milla (Deloitte, 2020)

2.1.2.3 Carga y descarga

La mayoría de las ciudades españolas sufren de una evidente escasez de zonas de carga y descarga, y su gestión presenta claras ineficiencias. Dada la complejidad de aumentar el número de estas zonas en las ciudades, se propone optimizar el uso de las zonas de distribución urbana de mercancías (DUM) existentes mediante la digitalización, creando una plataforma para este fin.

Esta plataforma web o aplicación informática deberá estar equipada con sistemas de geolocalización para visualizar en tiempo real la ubicación de las zonas de carga y descarga. Además, en estas zonas se instalará un sistema de sensores que permita conocer su disponibilidad en cualquier momento. Con el uso de esta plataforma, los repartidores podrán planificar sus rutas según la disponibilidad de las zonas de carga y descarga, evitando paradas ilegales y reduciendo la congestión urbana.

Figura 5– Impacto de la plataforma digital de zonas de carga y descarga en los retos de Última Milla (Deloitte, 2020)

2.1.2.4 Electrificación de las flotas

Según la Agencia Europea del Medio Ambiente, el transporte de mercancías representa el 10% de las emisiones globales de CO2 y el 25% del total de emisiones en España. Para reducir drásticamente estos niveles de contaminación provenientes del transporte de mercancías, es necesario crear un plan que fomente el uso de vehículos eléctricos en la distribución urbana de mercancías.

La inversión requerida para utilizar una flota sostenible es inicialmente alta cuando el kilometraje es bajo. Sin embargo, a medida que se utilizan más las furgonetas eléctricas, la reducción de los costos marginales, debido al ahorro en combustible, lleva a una disminución de los costos totales en comparación con las furgonetas diésel. Además, las emisiones de gases contaminantes se reducen en más del 70% desde los primeros kilómetros de uso de las furgonetas eléctricas.

Figura 6– Impacto de la electrificación de las flotas de reparto en los retos de Última Milla (Deloitte, 2020)

2.2. Parcel Lockers

Uno de los principales retos introducidos por el auge del comercio electrónico que deben superar los minoristas y las empresas de logística es la entrega rápida de un gran número de paquetes al mínimo coste, con especial atención a la entrega de última milla. Para ello, se están implementando métodos de entrega alternativos respecto a la entrega a domicilio. Una de las estrategias implementadas por estas empresas es el uso de puntos de recogida ubicados en lugares convenientes para los clientes en los que los mensajeros entregan los paquetes, que serán recogidos por los clientes en un momento posterior. (Mitrea et al., 2020).

Los casilleros para paquetes o parcel lockers son máquinas de entrega desatendidas, ubicadas en lugares elegidos y mayoritariamente concurridos. Es un sistema de cajas de recepción, que permiten recibir y enviar paquetes las 24 horas del día, los 7 días de la semana.

Figura 7– Casillero para paquetes de Inpost Company

Una característica distintiva de los casilleros de paquetes en comparación con los servicios de mensajería tradicionales es una reducción significativa en el número de entregas y la falta de entregas directas debido a la ausencia del destinatario. Los armarios para paquetes suelen estar situados en lugares públicos (por ejemplo, centros comerciales, gasolineras, universidades, etc.). Por lo tanto, esto permite recibir envíos en un momento conveniente, a menudo mientras se hacen otras cosas, como comprar o repostar combustible.

Con la selección de ubicaciones adecuadas, los casilleros para paquetes pueden proporcionar no sólo importantes beneficios económicos, sino que también, o incluso principalmente, pueden tener un impacto positivo en la reducción de los contaminantes emitidos al medio ambiente por el transporte urbano de mercancías. (Iwan et al., 2016).

Savelsbergh & Van Woensel (2016) consideraron este tipo de solución una oportunidad para mitigar los efectos negativos de las entregas directas al consumidor. Alguno de estos efectos puede ser el fallo en la primera entrega, lo que implica devoluciones o entregas repetidas.

La introducción de puntos de recogida para entregas de última milla ofrece varias ventajas como son la reducción de los costes de desplazamiento y la reducción de los costes externos al poder entregar más paquetes en el mismo lugar, lo que hace que los mensajeros pueden recorrer rutas más cortas. Además, la falta de sincronización entre los mensajeros y los clientes permite una mayor flexibilidad tanto en los tiempos de entrega como de recogida, lo que, por tanto, reduce la cantidad de entregas fallidas (Buzzega & Novellani, 2023).

El objetivo más importante de la implementación de taquillas de paquetería es reducir el número de entregas en la ciudad, incluidas las entregas fallidas y la posterior devolución de mercancías por mensajería y servicios postales. Ayuda a reducir el kilometraje innecesario de los vehículos con el uso de energía asociado y los impactos de la congestión.

Según los resultados del análisis realizado en octubre de 2013 por investigadores del Departamento de Robótica y Mecatrónica de la Universidad de Ciencia y Tecnología AGH de Cracovia (Polonia), el servicio de mensajería que presta servicio a las taquillas de paquetes InPost es capaz de entregar 600 paquetes en tan solo un día. , con una distancia de viaje de aproximadamente 70 km en comparación con los 60 paquetes y 150 km respectivamente en el sistema de entrega tradicional.

Tabla 2– Comparación de taquillas de paquetería y entregas por mensajería (Adaptado de "Paczkomaty InPost – Ekspertyza AGH", por Log4.pl, 2013).

	Mensajería	Taquillas Inpost
Número de kilómetros durante un día	150	70
Número de paquetes entregados en un día	60	600
Emisiones de CO_2 (toneladas anuales)	32 500	1516
Consumo anual de combustible en litros	22 500 000	1 050 000

Un aspecto clave en cuanto a la viabilidad de los lockers es la ubicación. Lachapelle et al. (2018) descubrieron que estos casilleros se colocan con frecuencia en calles comerciales, sitios con abundante estacionamiento y cerca de la oficina de correos o centros comerciales. También concluyeron que la mayoría se colocan en sitios seguros y que deberían colocarse con mayor frecuencia en gasolineras o centros comerciales, para que puedan combinarse con diferentes propósitos. De Oliveira et al. (2022) obtuvieron resultados similares, ya que concluyeron que los supermercados y centros comerciales están ubicados en áreas de altos ingresos y en lugares menos costosos para los inversionistas, lo que los hace más atractivos económicamente, mientras que las oficinas de correos, gasolineras y farmacias están ubicadas en áreas altamente pobladas, por lo que podrían servir a casi toda la población de la ciudad y estimular modos de transporte activos.

Para evaluar la influencia de la ubicación de los casilleros para paquetes en su eficiencia, Iwan et al. (2016) realizaron en Szczecin un experimento sobre la reubicación de una selección de máquinas. Con base en el análisis del número mensual de paquetes entregados por cada casillero, se preparó el conjunto de las máquinas más y menos populares. En base a los resultados y teniendo en cuenta la disponibilidad de las

ubicaciones elegidas y los planes estratégicos de la empresa InPost, se reubicaron cinco máquinas de bajo rendimiento. Además se implementó una nueva máquina.

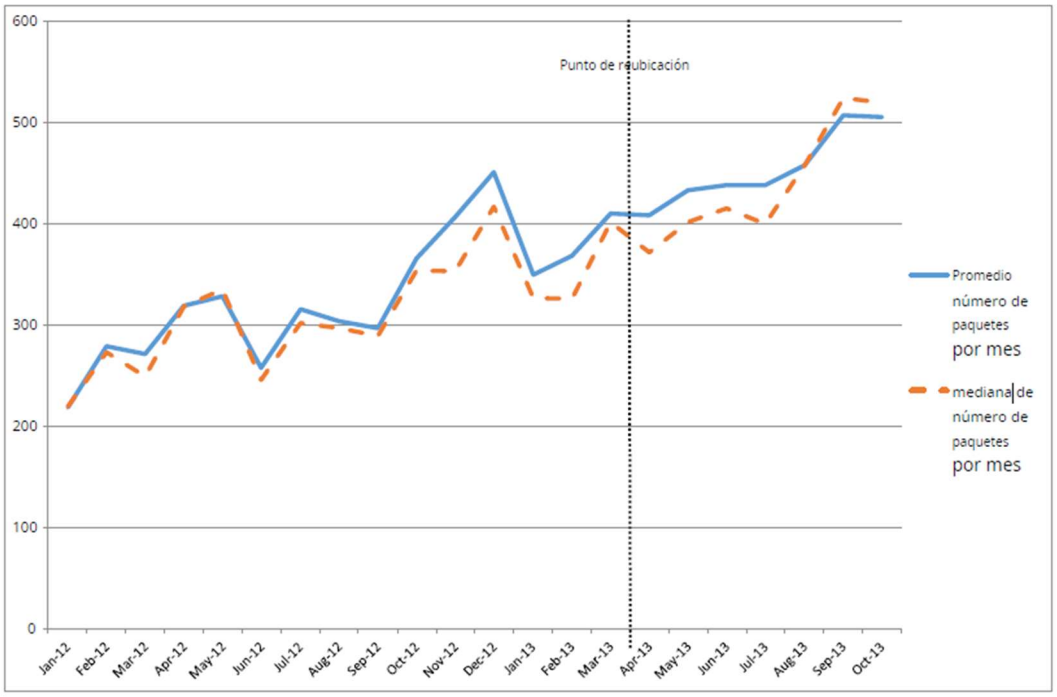

Figura 8– Número mensual de paquetes recogidos (Iwan et al., 2016)

Según los datos del operador de los casilleros de paquetes, entre enero de 2012 y abril de 2013, el número medio de paquetes entregados por 17 casilleros de paquetes en Szczecin fue de 335 por mes. Tras la reubicación de cinco máquinas de bajo rendimiento en nuevas ubicaciones más adecuadas, así como la instalación de un casillero para paquetes adicional, este número aumentó a 443 entregas por mes (32%). El aumento medio de los envíos en cuatro casos ascendió a 79 paquetes al mes. Sólo una reubicación fue ineficiente (el número de entregas disminuyó). El criterio de reubicación más importante en este experimento fue la proximidad de gasolineras, universidades y centros comerciales. El crecimiento más eficiente (más de 200 paquetes al mes) se consiguió en el caso de las máquinas ubicadas cerca del centro comercial.

2.3 Clustering

El clustering es una técnica de machine learning no supervisado utilizada para agrupar elementos dentro de un conjunto de datos en función de sus características y similitudes. A diferencia de otros métodos de clasificación, el clustering no requiere etiquetas previas, sino que detecta patrones inherentes en los datos y organiza los elementos en grupos llamados clústeres. Este método se aplica en una variedad de campos, como la segmentación de clientes en marketing, la detección de anomalías en ciberseguridad y la optimización de redes logísticas. Esta técnica es ampliamente utilizada en la ciencia de datos y la inteligencia artificial para descubrir estructuras ocultas en los datos y mejorar la toma de decisiones basada en análisis avanzados (IBM, 2024).

Dentro de las distintas estrategias de clustering, se pueden identificar dos grandes grupos: técnicas particionales y técnicas jerárquicas. Las técnicas particionales dividen un conjunto de n objetos con d dimensiones en k grupos distintos, optimizando algún criterio de agrupación, como la minimización de la variabilidad interna dentro de cada clúster. En este tipo de enfoques, el número de clústeres suele ser definido previamente, y el objetivo es asignar cada objeto a uno de esos grupos de manera que la similitud interna sea máxima y la separación entre grupos sea la mayor posible. Por otro lado, las técnicas jerárquicas construyen una estructura organizativa basada en relaciones de similitud entre los elementos. Este tipo de clustering puede ser de dos formas: divisivo o aglomerativo. El clustering jerárquico divisivo (top-down) comienza con todos los elementos agrupados en un único clúster y, en cada iteración, se divide progresivamente hasta que cada objeto queda en un grupo individual o se alcanza un número de clústeres predefinido. En cambio, el clustering jerárquico aglomerativo (bottom-up) sigue un proceso inverso: inicia con cada objeto como un clúster independiente y, en cada paso, agrupa aquellos que presentan mayor similitud hasta formar un único clúster final que contiene todos los elementos (Martín Gallardo, 2018).

Uno de los métodos más utilizados dentro de la estrategia aglomerativa es el método de Ward, 1963. Esta técnica parte de un conjunto donde cada elemento es inicialmente un clúster individual. En cada iteración, se calcula la distancia entre los distintos grupos según un criterio de similitud, y se fusionan los dos que presentan la menor distancia. Este proceso se repite hasta que todos los elementos se han unido en un único clúster global. Como resultado, se genera una estructura jerárquica que refleja el orden en que se han realizado las fusiones entre los diferentes clústeres

En la figura se muestra un ejemplo gráfico de este procedimiento, ilustrando cómo los elementos individuales se agrupan progresivamente hasta formar una jerarquía completa. Esta representación visual es útil para entender cómo evoluciona la estructura de los clústeres a lo largo del proceso.

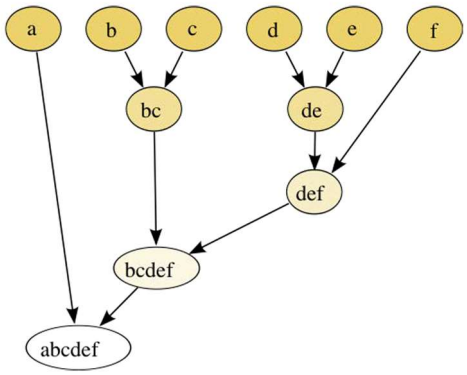

Figura 9– Ejemplo de clustering jerárquico aglomerativo (Martín Gallardo, 2018)

Otro método de agrupamiento aglomerativo es el de K-means, El algoritmo K-means es una técnica ampliamente utilizada en el análisis de datos para la segmentación en grupos o clústeres. Su principio fundamental es la asignación de cada observación al clúster cuyo centroide se encuentra más próximo, con el objetivo de minimizar la variabilidad interna dentro de cada grupo. Este enfoque permite clasificar conjuntos de datos en categorías de manera eficiente y es aplicado en diversas disciplinas, como la segmentación de clientes, el reconocimiento de patrones y la optimización logística (Lumbreras Herrera, 2020).

El centroide de un clúster se define como el punto medio de todas las observaciones asignadas a dicho grupo. En términos matemáticos, si un clúster C_k contiene un conjunto de $|C_k|$ observaciones x_i, su centroide μ_k se calcula como:

$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \epsilon C_k} x_i$$

Este punto de referencia se recalcula constantemente durante el proceso iterativo del algoritmo, lo que permite una mejor representación de los datos dentro de cada grupo.

El proceso iterativo de K-means consiste en la continua reasignación de observaciones a los clústeres y la actualización de sus centroides, permitiendo mejorar la estructura de agrupación hasta que el algoritmo converge. Inicialmente, se seleccionan K centroides de manera aleatoria. Luego, cada observación se asigna al clúster cuyo centroide está más próximo, utilizando generalmente la distancia euclidiana como métrica de proximidad, la cual se expresa de la siguiente forma:

$$d(x_i, \mu_k) = \sqrt{\sum_{j=1}^{p} (x_{ij} - \mu_{kj})^2}$$

donde x_{ij} representa la coordenada j de la observación i y μ_{kj} es la coordenada j del centroide del clúster k. Una vez asignadas todas las observaciones, se recalculan los centroides de cada grupo como la media de los puntos asignados. Este proceso se repite iterativamente hasta que la asignación de las observaciones a los clústeres se estabiliza o se alcanza un número máximo de iteraciones. El objetivo principal de K-means es reducir la dispersión dentro de los clusters, buscando que las observaciones agrupadas sean lo más homogéneas posible (Lumbreras Herrera, 2020).

Para lograrlo, el algoritmo minimiza la suma de las distancias al cuadrado entre cada observación y su centroide, definida como:

$$W(C_k) = \sum_{x_i \epsilon C_k} (x_i - \mu_k)^2$$

Dado que el objetivo es minimizar la dispersión dentro de los clusters, se busca optimizar la siguiente función de coste:

$$\sum_{k=1}^{k} W(C_k)$$

Sin embargo, obtener una solución exacta que minimice esta función de manera global es un problema computacionalmente complejo. En consecuencia, K-means utiliza su proceso iterativo para obtener una solución aproximada de manera eficiente, permitiendo un equilibrio entre precisión y rapidez en la segmentación de datos. El método del centroide es fundamental en este algoritmo, ya que determina la estructura de los clústeres y la calidad del agrupamiento obtenido (Lumbreras Herrera, 2020).

2.4 Problema de ubicación de APL

En esta sección se estudian a fondo dos trabajos que proponen el problema de ubicación de automated parcel lockers (APL) a través de dos métodos, la programación lineal entera mixta y algoritmos codiciosos.

El objetivo principal del problema de ubicación de APL es encontrar las ubicaciones que minimicen la cantidad de APL a instalar y los costes operativos relacionados con el viaje a esas ubicaciones. El modelo de ubicación debe considerar la variabilidad en la disposición de los clientes a desplazarse desde su residencia para recoger el paquete y su demanda. Debido a las preferencias de los clientes y las múltiples opciones de entrega entre las que ahora pueden elegir, no cubrir a un cliente desde una ubicación de APL puede resultar en una pérdida de demanda. Las elecciones de los clientes se han incluido recientemente en el problema de ubicación de APL. El atractivo de una ubicación se puede modelar como una función decreciente de la distancia entre la ubicación y el cliente (Lin et al., 2022). Además, las soluciones al problema de ubicación de las instalaciones de APL deben ser sólidas para hacer frente a la variabilidad de la demanda.

2.4.1 Acercamiento de Deutsch y Golany (2017)

El trabajo realizado por Deutsch y Golany (2017) se centra en las instalaciones de taquillas de paquetería como solución al problema de logística de última milla. Se desea determinar, a través de la programación lineal entera mixta, dada la topografía de un área determinada y las demandas de los consumidores, cuántos sitios de casilleros abrir, dónde ubicarlos y cuántos casilleros estándar instalar en cada sitio, para maximizar el beneficio total resultante. El beneficio total se compone de los ingresos de los clientes que utilizan el servicio, menos los costes fijos y operativos de instalación, los descuentos en los costes de envío para los clientes que necesitan desplazarse para recoger sus paquetes y la pérdida de clientes potenciales que no están dispuestos a viajar para recibir servicios.

2.4.1.1 Modelo

El modelo se basa en una estructura de red, donde G = (I, E) es una red no dirigida, conectada y simple, con un conjunto de nodos I (de tamaño n = || I ||) y un conjunto de aristas E. Los nodos representan áreas y las aristas distancias entre ellas.

Cada nodo i tiene una demanda $P_i(>0)$, que representa la población que reside y trabaja en el área correspondiente, y una frecuencia de pedido online $F_i(>0)$. $Qi \equiv Pi \cdot Fi(> 0)$ representa el pedido online total del nodo i por unidad de tiempo, y Q = sum(Qi) el pedido total de todos los nodos.

En cuanto a las distancias entre los nodos, $d_{i,j}$ (>=0) indica la distancia entre el nodo i y el nodo j , y $D_{max}=max_{i,j}(d_{i,j})$, la distancia máxima. Para cada capa u = 0, 1, . . . ,m, con m<=n, sea D_u una distancia fija, con,

$$0=D_0 < D_1 < \ldots < D_m=D_{max}. \quad (1)$$

Para cada nodo i , y para cada u = 1, ... ,m. Sea $S^u_i = \{j \in I : d_{i,j} \in (D_{u-1}, D_u]\}$, i.e. S^u_i es el conjunto de nodos dentro de la capa de distancia u desde el nodo i . Además, para cada nodo $i \in I$, sea $S^0_i = \{i\}$ y $S_i =$

(S^0_i , … , S^m_i). Para cada nodo i∈I , y para cada u = 0, 1, ... ,m. l^u_i indica si se abre al menos una instalación en el nodo i de la uº capa. Es decir,

$$l_i^u = \begin{cases} 1 & \text{, si al menos una instalación se abre en } S^u_i \\ 0 & \text{, e.o.c} \end{cases} \quad (2)$$

La empresa de consignación quiere abrir varias instalaciones de paquetería en la red para maximizar el beneficio total. Se supone que los clientes siempre prefieren sus instalaciones más cercanas. Sin embargo, el porcentaje de clientes que aceptan viajar y recoger sus paquetes a otros sitios desde su sitio de origen disminuye con la distancia. En particular, el porcentaje de clientes que aceptan viajar una distancia que no sea superior a D1 es superior al porcentaje de clientes que aceptan viajar una distancia superior a D1 pero no más grande que D2, y así sucesivamente. Es decir,

$$1 \equiv \rho^0 > \rho^1 \geq \ldots \geq \rho^m (\geq 0). \quad (3)$$

Si los clientes no están dispuestos a viajar una distancia mayor que D_u, u>=1, entonces:$\rho^{u+1} = \rho^{u+2} = \ldots = \rho^m = 0$. Tenga en cuenta que también podemos asumir que estos porcentajes dependen de los nodos. La complejidad del problema será la misma. Sea R > 0 denotado como el ingreso de la compañía de envío por pedido. Dado que no todos los clientes están dispuestos a viajar y recoger sus paquetes de sitios que no sean su sitio de origen, los que están de acuerdo son compensados por la empresa, recibiendo un descuento en los costos de entrega. Para u = 0, 1, . . . ,m. Sea C^u_d>=0 el descuento en los costos de entrega para los clientes que viajan a una instalación en su uº capa, con:

$$0 \equiv C_d^0 \leq C_d^1 \leq C_d^2 \leq \ldots \leq C_d^m (< R) \quad (4)$$

(téngase en cuenta que, si se trata de la C_d^m >=R, la empresa de expedición puede no aplicar el sistema).

Sea f_i > 0 el coste de configuración de una instalación en el nodo i . El costo de instalación consiste en el costo fijo de la construcción de una instalación, y un costo variable, como el mantenimiento de la instalación, alquiler, electricidad y más (todos los costos se normalizan por período utilizando los mismos períodos para los cuales la demanda se expresa arriba). Suponemos que la empresa de consignación puede abrir instalaciones solo en los nodos de la red. Sea X el vector de localización de las instalaciones, igual para cada i :

$$x_i = \begin{cases} 1 & \text{si se abre una instalación en el nodo i} \\ 0 & \text{e.o.c} \end{cases} \quad (5)$$

Sea Y la matriz de asignación de capas de los clientes, igual para cada i, u = 0,1, …,m.

$$y_{i,u} \in \begin{cases} \{1\} & \text{si todos los clientes en el nodo i viajan a una instalación en su uº capa} \\ (0, 1) & \text{si una fracción de los clientes en el nodo i viajan a una instalación en su uº capa} \\ \{0\} & \text{e.o.c} \end{cases} \quad (6)$$

Con esta notación, la formulación del problema se denota como:

$$\min_{X,Y,L} \sum_{i=1}^{n} \left[f_i x_i + Q_i \sum_{u=0}^{m} (C_d^u - R)\rho^u y_{i,u} \right] \tag{7a}$$

$$\text{s.t.:} \quad y_{i,u} \le l_i^u \ \forall\, i \in I, u = 0, 1, \ldots, m, \tag{7b}$$

$$\sum_{u=0}^{m} y_{i,u} = 1 \ \forall\, i \in I, \tag{7c}$$

$$l_i^u \le \sum_{j \in S_i^u} x_j \ \forall\, i \in I, u = 0, 1, \ldots, m, \tag{7d}$$

$$l_i^u \ge x_j \ \forall\ j \in S_i^u, u = 0, 1, \ldots, m,\ i \in I, \tag{7e}$$

$$x_i \in \{0, 1\} \forall\, i \in I, \tag{7f}$$

$$l_i^u \in \{0, 1\} \ \forall\, i \in I, u = 0, 1, \ldots, m. \tag{7g}$$

$$y_{i,u} \ge 0 \ \forall\, i \in I, u = 0, 1, \ldots, m. \tag{7h}$$

El primer término en la función objetivo (7a) es el costo de instalación de las instalaciones. El segundo término son los descuentos netos en los gastos de envío para los clientes que viajan para recoger sus paquetes. Dado que $\rho^0 = 1$ y $C_d^0 = 0$ por (3) y (4), respectivamente, si una instalación se encuentra en el sitio de origen de los clientes, los clientes la utilizan y no reciben ningún descuento. La restricción (7b) garantiza que los clientes en cada nodo dado solo viajan a las capas relevantes, donde se encuentra una instalación. La restricción (7c) asegura que la demanda en cada nodo puede ser atendida completamente a través de instalaciones establecidas en la red. Las restricciones (7d), (7e) y (7g) determinan los valores de los indicadores de las capas. Específicamente, la restricción (7d) junto con la (7g) aseguran que el valor del indicador de capa uº del nodo i es menor que el número de instalaciones abiertas en la capa uº del nodo i, de modo que si no se abre ninguna instalación allí, el valor del indicador será 0. La restricción (7e) junto con la (7g) se aseguran de que si se abre al menos una sola instalación en la capa uº del nodo i, el valor del indicador será 1. Tenga en cuenta que no es necesario imponer restricciones de asignación cerrada, ya que el orden de los descuentos (4) garantiza que cada cliente viajará a la capa u más cercana en la que se encuentre al menos una instalación. Las restricciones (7f) y (7g) son las restricciones binarias de las variables x_i y l_i^u , respectivamente. La restricción (7h) indica que las variables de asignación deben ser no negativas.

2.4.1.2 Caso de studio

Teniendo en cuenta una serie de datos y suposiciones, se utiliza el algoritmo de solución de tres fases para resolver las instalaciones de casilleros de paquetes en el área metropolitana de Toronto situada en Canadá. La solución propone abrir 65 instalaciones. La ubicación de las instalaciones se presenta en la Figura 8.

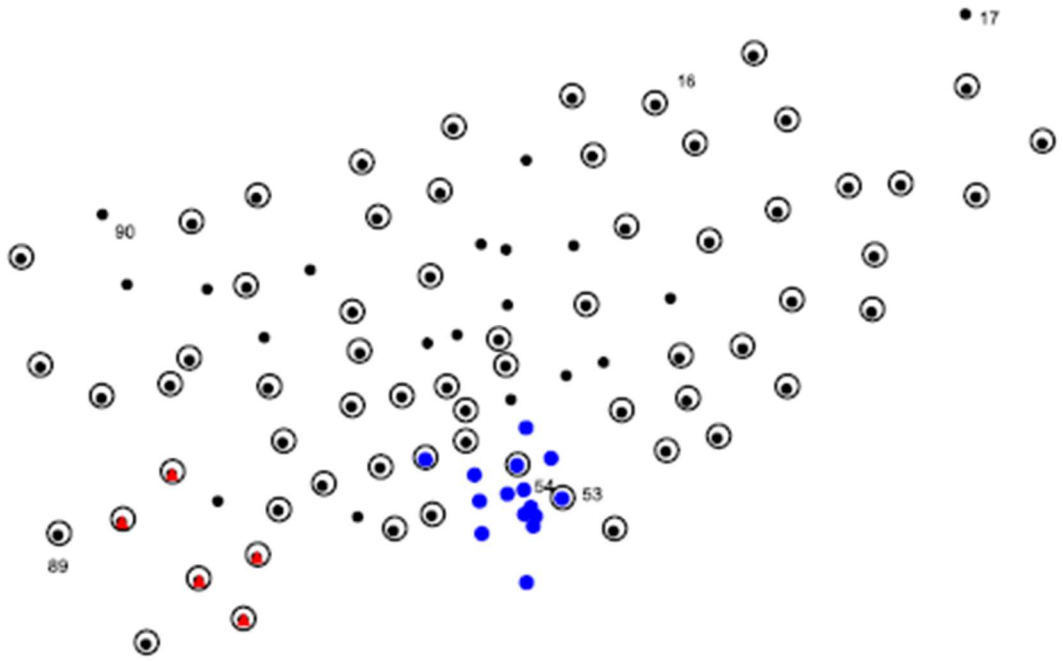

Figura 10– Solución del estudio de caso en Toronto (Deutsch y Golany, 2017)

Los puntos negros, los puntos en círculos, los puntos rojos y los puntos azules representan, respectivamente, los nodos de demanda de la red, los nodos de la solución donde se abre una instalación, los nodos de las áreas de Mimico e Islington y los nodos del centro de la ciudad. Dada esta solución, el valor de la función objetivo del problema es $3420, el valor del problema es −$16,803 y el porcentaje relativo de pedidos perdidos es 4,35%. Además, el algoritmo abre 2 instalaciones de los 14 nodos en el centro de Toronto y 5 instalaciones de los 5 nodos en las áreas de Mimico e Islington.

Para demostrar la calidad de la solución final, seleccionamos aleatoriamente tres pares de nodos relativamente cercanos, donde uno de ellos tenía una instalación abierta y el otro no, y los intercambiamos. En los tres casos, el intercambio condujo a un peor valor de la función objetivo,

Tabla 3– Demostración de la calidad de la solución final (Deutsch y Golany, 2017)

Nodos intercambiados	Valor de la F.O	Porcentaje de pedidos perdidos	Porcentaje de empeoramiento de la F.O
(16,17)	-$16,748	5.12	0.92
(35,54)	-$16,712	4.8	0.54
(89,90)	-$16,501	5.85	1.8

2.4.2 Acercamiento de Ottaviani et al (2023)

En el estudio realizado por Ottaviani et al. (2023), se presentan dos métodos para determinar la cantidad, ubicación y capacidad óptimas de APL a instalar.

2.4.2.1 MILP

El primer método consiste en optimizar un modelo de programación lineal entera mixta (MILP), determinando simultáneamente las variables antes mencionadas.

minimize	$c^{APL} \sum_{a \in A} act_a + c^{Module} \sum_{a \in A} mod_a$		(1)
subject to	$\sum_{u \in U} sat_u \cdot dem_u / \sum_{u \in U} dem_u \geq SL$		(2)
	$D^*_{ua} = D_{ua} \cdot act_a + M(1 - act_a)$	$\forall u \in U, \forall a \in A$	(3)
	$D_{ua} = [(x_u - x_a)^2 + (y_u - y_a)^2]^{1/2}$	$\forall u \in U, \forall a \in A$	(4)
	$mD_u \leq D^*_{ua}$	$\forall u \in U, \forall a \in A$	(5)
	$mD_u \geq D^*_{ua} - M(1 - ass_{ua})$	$\forall u \in U, \forall a \in A$	(6)
	$\sum_{a \in A} ass_{ua} = 1$	$\forall u \in U$	(7)
	$tol_u - mD_u + \epsilon - M \cdot sat_u \leq 0$	$\forall u \in U$	(8)
	$mD_u - tol_u - M(1 - sat_u) \leq 0$	$\forall u \in U$	(9)
	$\sum_{u \in U} ass_{ua} \cdot dem_u \leq cap^{APL} + cap^{Module} \cdot mod_a$	$\forall a \in A$	(10)
	$mod_a \leq mod^{max} \cdot act_a$	$\forall a \in A$	(11)
	$\sum_{a \in A} act_{ua} \leq \max_{a \in A} a$		(12)
	$act_a \in \{0,1\}$	$\forall a \in A$	(13)
	$sat_u \in \{0,1\}$	$\forall u \in U$	(14)
	$ass_{ua} \in \{0,1\}$	$\forall u \in U, \forall a \in A$	(15)
	$mod_a \in \mathbb{N}$	$\forall a \in A$	(16)
	$D_{ua} \in \mathbb{R}^+$	$\forall u \in U, \forall a \in A$	(17)
	$D^*_{ua} \in \mathbb{R}^+$	$\forall u \in U, \forall a \in A$	(18)
	$mD_u \in \mathbb{R}^+$	$\forall u \in U$	(19)
	$x_a, y_a \in \mathbb{R}^+$	$\forall a \in A$	(20)

Figura 11– Modelo MILP completo (Ottaviani et al., 2023)

La ecuación 1 es la función objetivo a minimizar y consiste en la suma del costo total de los APL($c^{APL}\Sigma_{a\in A}\ act_a$) y el costo total de los módulos adicionales ($c^{Module}\Sigma_{a\in A}\ \mathrm{mod}_a$). La variable act_a es booleana (según la Ecuación 13) e indica si el aº APL está instalado ($act_a = 1$) o no ($act_a = 0$). En cambio, la variable mod_a es natural (según la Ecuación 16) y se refiere al número de módulos a instalar en el aº APL. La Ecuación 11 limita el número máximo de módulos por APL a mod^{max} debido a restricciones volumétricas y evita que se instale cualquier módulo si un APL no está activo ($act_a = 0$). Para garantizar que la rutina de optimización prefiera expandir los APLs existentes mediante la instalación de módulos adicionales en lugar de activar nuevos APLs, el coste de instalar un APL (c^{APL}) debe ser mayor que el coste de instalar un módulo adicional (c^{Module}).

La ecuación 2 expresa la restricción del nivel de servicio, que asegura que el porcentaje de demanda satisfecha esperada ($\Sigma_{u\in U}\ sat_u \cdot dem_u$ / $\Sigma_{u\in U}\ dem_u$) cumple o supera el nivel de servicio objetivo (SL). La variable sat_u también es booleana (según la Ecuación 14) y determina si el u-ésimo usuario se considera cubierto ($sat_u = 1$) cuando su demanda (dem_u) se asigna a cualquier APL.

Las ecuaciones 3 a 7 pertenecen a las distancias (D_{ua} , Ecuación 17) y distancias mínimas (mD_u , Ecuación 18) entre los usuarios y las APL. La ecuación 4 calcula la distancia euclídea (D_{ua}) entre los usuarios y las APL. La ecuación 3 evita que se asigne una APL a cualquier usuario si la APL no está activa ($act_a = 0$),ya que la distancia entre ellos se establecería en M.

Las ecuaciones 5-6 determinan la distancia entre el u-ésimo usuario y su APL activo más cercano (mD_u), mientras que la Ecuación 7 asigna la demanda del usuario (dem_u) a ese APL (ass_{ua}=1). Las ecuaciones 10 y 12 se relacionan con el cálculo de la demanda esperada que debe satisfacer cada APL. Específicamente, la ecuación 10 asegura que a ningún APL se le asigne más demanda que su capacidad, que es la suma de su capacidad estándar (c^{APL}) y la capacidad de cada módulo adicional ($cap^{Module} \cdot \mathrm{mod}_a$). La Ecuación 11 fija el número máximo de módulos a instalar en 10, lo cual es razonable dada la cantidad de espacio que requiere cada módulo para su ubicación, mientras que la Ecuación 12 evita que el módulo se instale (mod_a) a una instalación APL siempre que dicha APL no esté activada ($act_a = 0$).Las ecuaciones 8-9 aseguran que siempre que la distancia entre un usuario y su APL más cercano (mD_u) es menor o igual a la distancia de viaje tolerada por el usuario (tol_u), el usuario es asignado a ese APL (Ecuación 7), y su demanda se considera satisfecha ($sat_u = 1$).Si bien varias APL pueden satisfacer la demanda de un usuario, esta suposición simplifica el problema y reduce el tiempo de cálculo. Esta formulación asigna la demanda de un usuario a su APL más cercano.

2.4.2.2 Algoritmos codiciosos

Por otra parte, el segundo método emplea algoritmos codiciosos que determinan la cantidad de APL, su ubicación y su capacidad. Hacerlo permite evaluar la curva de demanda esperada de los clientes satisfechos en función del número de APL. Ambos algoritmos implementan los siguientes procedimientos:

El procedimiento 1 carga los APL anteriores. Primero toma como variables de entrada la secuencia ordenada de APLs (APL_order),el próximo candidato de APL (APL^*), el conjunto de usuarios satisfechos (sat),y la matriz de distancias usuario/APL (D).Luego, evalúa el número potencial de usuarios (pot_users) y demanda respectiva (pot_demand) que cada APL en la secuencia ordenada puede cubrir. A continuación, el procedimiento determina el número de módulos (mod_a) que cada APL puede acomodar y asigna usuarios en consecuencia, hasta mod^{max}. Finalmente, marca a los usuarios como cubiertos por su APL asignado ($sat_u \leftarrow 1$).

```
def LoadPreviousAPLs(APL_order, sat, D, APL*):
    for a ∈ APL_order do
        if a ≠ APL* then
            pot_users ← 0
            pot_dem ← 0
            for u ← 1 to nUsers do
                if sat_u = 0 and D_ua ≤ tol then
                    pot_users ← pot_users + 1
                    pot_dem ← pot_dem + tol
            mod_a ← 0
            cap_a ← cap^APL
            it_1 ← 0
            while it_1 ≤ mod^max and pot_users > 0 do
                if it_1 > 0 then
                    mod_a ← mod_a + 1
                    cap_a ← cap_a + cap^Module
                if min(cap_a, pot_dem) = cap_a then
                    sats^temp ← cap_a / dem
                else
                    sats^temp ← pot_users / dem
                for it_2 ← 0 to sats^temp do
                    for u ← 0 to nUsers do
                        if sat_u = 0 and D_ua ≤ tol then
                            sat_u ← 1
                            pot_users ← pot_users − 1
                it_1 ← it_1 + 1
    return sat
```

Figura 12– Procedimiento 1: LoadPreviousAPLs (Ottaviani et al., 2023)

El procedimiento 2 evalúa al siguiente candidato APL (APL^*). Para lograr esto, toma la secuencia ordenada de APL (APL_order), la matriz de usuarios satisfechos (sat) y la matriz de distancias usuario/APL (D) como variables de entrada. Luego, el procedimiento recorre las APL que aún no están en APL_order y selecciona el APL que podría cubrir el mayor número potencial de usuarios o demanda.

```
def EvaluateNextAPL(APL_order, sat, D):
    APL* ← 0
    max_dem ← 0
    for a ← 1 to nAPLs do
        if a ∉ APL_order then
            pot_dem ← 0
            for u ← 1 to nUsers do
                if sat_u = 0 and D_ua ≤ tol then
                    pot_dem ← pot_dem + tol
            if pot_dem ≥ max_dem then
                max_dem ← pot_dem
                APL* ← a
    return APL*
```

Figura 13– Procedimiento 2: EvaluateNextAPL (Ottaviani et al., 2023)

El procedimiento 3 evalúa la demanda potencial del conjunto APL, aumenta sus módulos si no alcanzan su máximo (mod^{max}) y establece a todos los usuarios cubiertos como satisfechos (sat_u=1).

```
def LoadNextAPL(sat, D, APL*):
    pot_users ← 0
    pot_dem ← 0
    for u ← 1 to nUsers do
        if sat_u = 0 and D_ua ≤ tol then
            pot_users ← pot_users + 1
            pot_dem ← pot_dem + tol
    mod_a ← 0
    it_1 ← 0
    while it_1 ≤ mod^max and pot_users > 0 do
        if it_1 > 0 then
            mod_a ← mod_a + 1
            cap_a ← cap_a + cap^Module
        if min(cap_a, pot_dem) = cap_a then
            sats^temp ← cap_a/dem
        else
            sats^temp ← pot_users/dem
        for it_2 ← 0 to sats^temp do
            for u ← 0 to nUsers do
                if sat_u = 0 and D_ua ≤ tol then
                    sat_u ← 1
                    pot_users ← pot_users − 1
                    nSats ← nSats + 1
        it_1 ← it_1 + 1
    return sat, nSats
```

Figura 14– Procedimiento 3: LoadNextAPL (Ottaviani et al., 2023)

Algoritmo 1:

Data: **x**, **y**
Result: APLs location and capacity evaluation given fixed sequence
initialization
/* Step 1 (Facultative) */
init: ***pcts***
for $s \leftarrow 1$ **to** S **do**
  $APL^* \leftarrow 0$
  $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{sat}, \boldsymbol{tol}, \boldsymbol{dem}, \boldsymbol{D} \leftarrow$ RandomizeUsers()
  for $a \leftarrow 1$ **to** nAPLs **do**
    $\boldsymbol{sat} \leftarrow$ LoadPreviousAPLs($\boldsymbol{APL\_order}$, $\boldsymbol{sat}$, $\boldsymbol{D}$, APL^*)
    $APL^* \leftarrow$ EvaluateNextAPL($\boldsymbol{APL\_order}$, $\boldsymbol{sat}$, $\boldsymbol{D}$)
    $\boldsymbol{sat}, nSats \leftarrow$ LoadNextAPL($\boldsymbol{sat}$, $\boldsymbol{D}$, APL^*)
    $pcts_{s,APL^*} \leftarrow nSats/\text{nUsers}$
for $a \in \boldsymbol{APL\_order}$ **do**
  $\overline{pcts_a} \leftarrow \sum_{s=1}^{S} pcts_{s,a}/S$
Sort APLs by descending order of $\overline{pcts_a}$ OR provide APLs sequence manually
/* Step 2 (Mandatory) */
init: ***pcts***
for $s \leftarrow 1$ **to** S **do**
  $\boldsymbol{D}, \boldsymbol{sat} \leftarrow$ RandomizeUsers()
  for $a \in \boldsymbol{APL\_order}$ **do**
    $APL^* \leftarrow a$
    $\boldsymbol{sat}, nSats \leftarrow$ LoadNextAPL($\boldsymbol{sat}$, $\boldsymbol{D}$, APL^*)
    $pcts_{s,APL^*} \leftarrow nSats/\text{nUsers}$
for $a \in \boldsymbol{APL\_order}$ **do**
  $\overline{pcts_a} \leftarrow \sum_{s=1}^{S} pcts_{s,a}/S$
Sort APLs by descending order of $\overline{pcts_a}$

Figura 15– Algoritmo 1 proporcionado por secuencia (Ottaviani et al., 2023)

El algoritmo 1 implica dos pasos principales. En el primer paso (A1S1), los usuarios son aleatorios S veces, y se determina la secuencia de APL (APL_order). El porcentaje de demanda cubierta se evalúa para cada APL en la secuencia. Una vez que se completa A1S1, los APL se clasifican según el porcentaje promedio de demanda cubierta en todas las S simulaciones, proporcionando así una secuencia ordenada de APLs para el segundo paso (A1S2). El segundo paso consiste en otra ronda de S simulaciones, que permite determinar la demanda cubierta promedio por cada APL($pcts_a$).

Algoritmo 2:

```
init: APL_order, pcts
it ← 0
for it ← to nAPLs do
    init: APL_maxs, pctss
    APL* ← 0
    for s ← 1 to S do
        x, y, sat, tol, dem, D ← RandomizeUsers()
        sat ← LoadPreviousAPLs(APL_order, sat, D, APL*)
        APL* ← EvaluateNextAPL(APL_order, sat, D)
        APL_maxs_s ← APL*
    APL* ← Mode(APL_maxs)
    APL_order_it ← APL*
    for s ← 1 to S do
        x, y, sat, tol, dem, D ← RandomizeUsers()
        sat ← LoadPreviousAPLs(APL_order, sat, D, APL*)
        sat, nSats ← LoadNextAPL(sat, D, APL*)
        pctss_s ← nSats/nUsers
    pcts_it ← ∑_{s=1}^{S} pctss_s/S
```

Figura 16– Algoritmo 2 proporcionado por secuencia (Ottaviani et al., 2023)

2.4.2.3 Resultados

El resultado del Algoritmo 2 consiste en una secuencia de APL, evaluadas de la siguiente manera. En primer lugar, se realizan S simulaciones n-APLs veces para evaluar y seleccionar el siguiente APL a instalar. En cada paso, el algoritmo considera la presencia de APLs previamente instalados que ya cubren una parte de los usuarios. Una vez que se ha seleccionado el mejor APL con mayor frecuencia entre las S simulaciones, otra ronda de S simulaciones son realizadas para calcular su demanda cubierta esperada media. Este proceso se repite por nAPLs veces, lo que resulta en una secuencia única de APL que se espera proporcionen el nivel más alto de cobertura de la demanda.

El enfoque MILP se centra en la optimización simultánea de los APL y la cantidad de módulos adicionales por instalación. Sin embargo, no permite identificar el porcentaje incremental de demanda satisfecha proporcionada por la activación individual de APL. Por otro lado, el enfoque algorítmico considera los APL y su capacidad uno a la vez, permitiendo evaluar el porcentaje total de la curva de demanda satisfecha según el a-ésimo APL. Desde esta perspectiva, todos los enfoques proporcionan resultados similares (es decir, alcanzar el 95% de la demanda total satisfecha con 10 a 11 APL). Los algoritmos proporcionan un nivel de rendimiento equivalente al MILP y, por lo tanto, pueden replicarse fácilmente en diferentes contextos geográficos. Esto permite comprender cuántas instalaciones se requieren por metro cuadrado, dada la distancia de viaje tolerada por la población para recoger un pedido (Ottaviani et al., 2023).

2.4.3 Conveniencia según nuestro problema

Ambos estudios abordan la problemática de la última milla en la distribución de paquetes, pero difieren significativamente en la manera en que estructuran sus modelos de optimización y en su capacidad para ser modificados o adaptados a nuevas condiciones.

El modelo de Deutsch y Golany (2017) se basa en un enfoque de programación matemática que permite una optimización estructurada de la red de parcel lockers, minimizando costes y tiempos de entrega. Esta formulación matemática lo hace adaptable a diferentes restricciones y escenarios, lo que facilita su modificación para incorporar nuevos factores que puedan mejorar la eficiencia del sistema. En contraste, Ottaviani et al. (2023) proponen un modelo híbrido que combina heurísticas de localización con técnicas de optimización, pero su principal limitación radica en su dependencia de datos de demanda previamente conocidos. Este enfoque es útil en contextos donde la información sobre los clientes es estable y confiable, pero puede resultar menos efectivo en entornos dinámicos donde la demanda fluctúa o es incierta.

Otro aspecto clave en la decisión es la robustez del modelo. Mientras que el de Deutsch y Golany (2017) proporciona una estructura matemática clara y fácilmente modificable, el de Ottaviani et al. (2023) depende en gran medida de supuestos específicos sobre la demanda, lo que restringe su aplicabilidad en situaciones donde estos datos no están completamente disponibles. Además, el modelo de Deutsch y Golany tiene un enfoque estratégico de optimización a nivel de red, lo que lo hace más adecuado para decisiones de planificación a gran escala, mientras que el de Ottaviani et al. está más orientado a la ubicación operativa de lockers en función de patrones de demanda previamente establecidos.

Teniendo en cuenta estas diferencias, se ha seleccionado el modelo de Deutsch y Golany (2017) como base para el desarrollo del modelo modificado de optimización de parcel lockers en este trabajo. Su formulación matemática rigurosa, su capacidad de adaptación a distintos escenarios y su aplicabilidad en contextos con datos inciertos lo convierten en una opción más sólida para la optimización de la red de taquillas automatizadas. En comparación, aunque el modelo de Ottaviani et al. (2023) es relevante para situaciones con demanda conocida, su enfoque menos flexible lo hace menos adecuado para construir una versión mejorada que pueda ajustarse a distintos entornos y restricciones operativas.

3 METODOLOGÍA

En esta sección se presenta el enfoque metodológico seguido para abordar el problema de asignación de parcel lockers. El proceso comienza con la obtención y tratamiento de los datos, etapa fundamental para garantizar la calidad de la información utilizada en el modelo. Para ello, se han recopilado datos relevantes sobre ubicaciones potenciales, demanda esperada y costes asociados a la instalación y operación de los lockers. Posteriormente, se ha aplicado la técnica de Agglomerative Clustering, permitiendo agrupar las ubicaciones en función de su proximidad y demanda, facilitando así una mejor estructuración del problema.

A continuación, se describe el modelo matemático desarrollado para la resolución del problema, basado en la formulación propuesta por Deutsch y Golany (2017). Sin embargo, debido a las particularidades del caso de estudio, han sido necesarias modificaciones en algunos parámetros y restricciones, ajustando el modelo a las necesidades específicas del problema.

Finalmente, se detalla la implementación computacional del modelo utilizando Python y la biblioteca Gurobi, lo que permite resolver instancias del problema de manera eficiente. Se han utilizado herramientas de optimización para obtener soluciones óptimas en distintos escenarios, evaluando el impacto de diferentes factores en la distribución de los lockers.

Esta metodología proporciona un enfoque riguroso y estructurado para la toma de decisiones en la ubicación de parcel lockers, combinando el tratamiento de datos con técnicas avanzadas de optimización matemática y computacional.

3.1 Datos

3.1.1 Obtención de los datos

Para llevar a cabo la optimización de la localización de los parcel lockers, se ha utilizado un conjunto de datos compuesto por coordenadas geográficas correspondientes a la ubicación de clientes en la ciudad de Sevilla. Estas coordenadas, expresadas en términos de latitud y longitud, representan puntos específicos dentro del área urbana, donde se encuentra la demanda potencial de los puntos de recogida automatizados.

El conjunto de datos ha sido entregado en un formato que permite su tratamiento y análisis computacional, asegurando su integración en las metodologías de optimización propuestas. Cada par de coordenadas representa un cliente dentro del área de estudio y servirá como base para la modelización del problema de última milla. La información geográfica de estos clientes permitirá evaluar las mejores ubicaciones estratégicas para la instalación de los parcel lockers, minimizando distancias de desplazamiento y optimizando la accesibilidad para los usuarios.

Dado que la optimización de la red de parcel lockers requiere una distribución eficiente en función de la concentración y demanda de clientes, la disposición de estas ubicaciones en la ciudad de Sevilla facilitará el análisis espacial del problema. A partir de estos datos, se podrán aplicar técnicas de clustering y modelos de optimización que permitan determinar la mejor configuración de los puntos de entrega, considerando criterios logísticos y operativos.

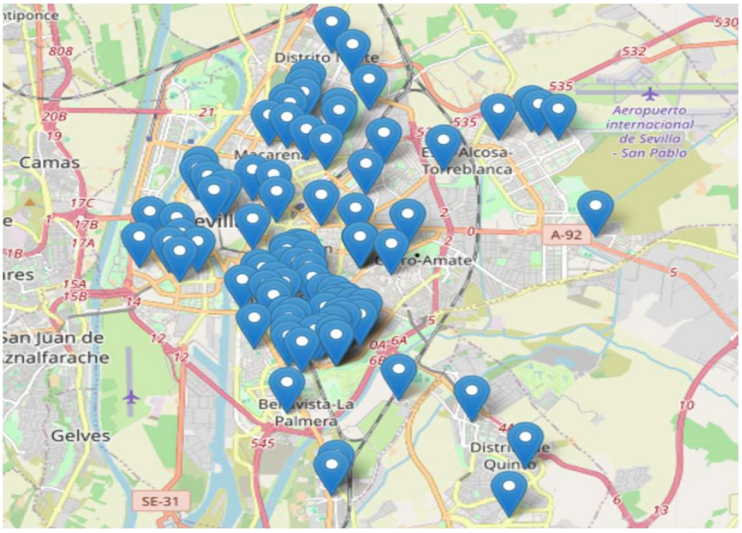

Figura 17– Mapa de clientes en la ciudad de Sevilla

3.1.2 Procesamiento de los datos

Para llevar a cabo la agrupación de cliente, primero se estableció un conjunto de coordenadas geográficas representando la ubicación de los clientes. Para ello, se utilizó un dendrograma para definir el número óptimo de grupos.

El dendrograma es una representación gráfica de cómo los puntos de datos se van fusionando en clústeres a diferentes niveles de similitud. En el gráfico adjunto, el eje horizontal representa los diferentes clientes, mientras que el eje vertical indica la distancia a la que los clústeres se combinan. Para determinar el número adecuado, se busca un corte natural, es decir, una altura a la que la fusión de los nodos tenga una diferencia notable en la distancia.

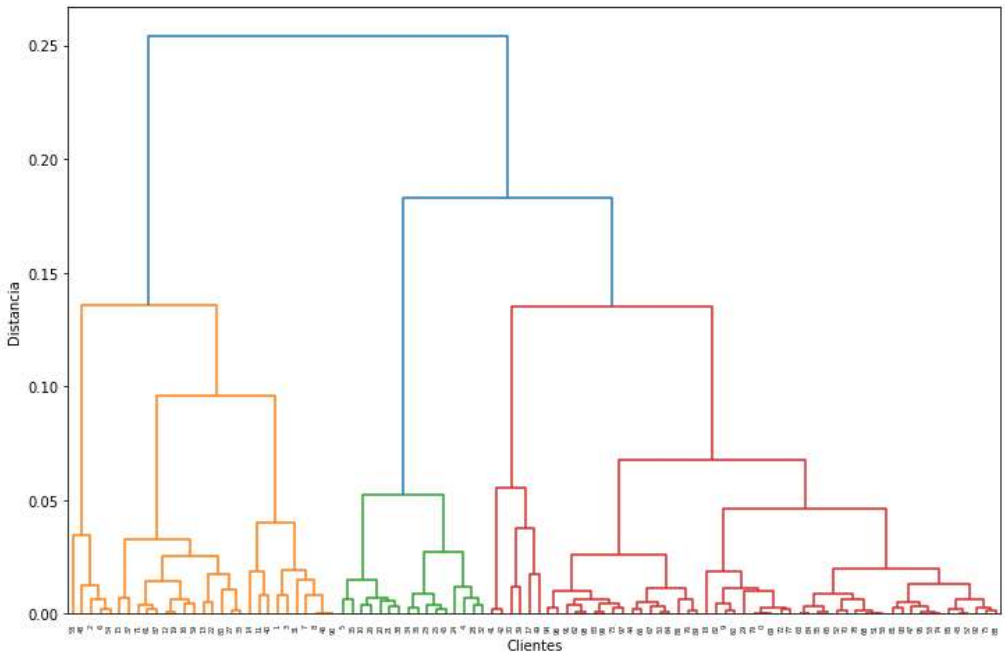

Figura 18– Dendrograma para la determinación del número de clústeres

En este caso, observamos que un corte razonable se puede hacer alrededor de una distancia de 0.07 a 0.10, ya que en este rango se pueden distinguir aproximadamente 6 clústeres antes de que las fusiones entre ellos comiencen a ocurrir a distancias significativamente mayores. Este punto de corte se elige para capturar la estructura de los datos de manera óptima, evitando tanto el exceso de segmentación (demasiados clústeres pequeños) como la sobre agrupación (muy pocos clústeres que mezclan datos heterogéneos).

Una vez determinado el número de 6 clústeres, se procede a aplicar el algoritmo Agglomerative Clustering, que pertenece a la familia de los algoritmos de clustering jerárquico aglomerativo. Este tipo de clustering sigue un enfoque ascendente (bottom-up), comenzando con cada punto de datos como un clúster individual y fusionando iterativamente los clústeres más cercanos hasta formar una estructura jerárquica completa.

Para decidir qué clústeres deben fusionarse en cada paso, es necesario definir una métrica de distancia y un criterio de enlace. En este caso, se utilizó la distancia euclidiana como métrica de similitud, la cual mide la distancia geométrica entre dos puntos en un espacio multidimensional. Matemáticamente, la distancia euclidiana entre dos puntos A(x1,y1) y B(x2,y2) se calcula como:

$$d(A,B) = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

Esta métrica es ampliamente utilizada en problemas de agrupamiento geoespacial, ya que proporciona una representación intuitiva de la proximidad entre puntos en un mapa.

El proceso en Python comenzó con la definición del conjunto de datos, que incluía las coordenadas de los clientes en un espacio bidimensional. Estas coordenadas representan la ubicación de cada cliente y fueron almacenadas en un array de NumPy para facilitar su manipulación.

Posteriormente, se utilizó la biblioteca scipy.cluster.hierarchy para generar el dendrograma a partir de la matriz de distancias entre clientes. En particular, se empleó la función scipy.cluster.hierarchy.linkage con la métrica euclidiana y el método de Ward, que minimiza la varianza dentro de cada grupo en cada fusión. Luego, la función scipy.cluster.hierarchy.dendrogram permitió visualizar la jerarquía de fusiones en un gráfico, facilitando la identificación del número óptimo de clusters en el conjunto de datos.

```
#Elección del número de clusters óptimo para el número de clientes
plt.figure(figsize=(12, 8))
shc.dendrogram(shc.linkage(data, method='ward'))
plt.xlabel("Clientes")
plt.ylabel("Distancia")
plt.show()
```

Figura 19– Dendrograma en Python

Tras analizar el dendrograma y determinar que una división en 6 clústeres era una opción adecuada, se procedió a aplicar el algoritmo de clustering jerárquico aglomerativo. Para ello, se utilizó la clase AgglomerativeClustering de sklearn.cluster, configurando n_clusters=6 para establecer el número de grupos, metric='euclidean' para definir la métrica de distancia y linkage='ward' para seguir el mismo criterio de fusión utilizado en el dendrograma. Este modelo asignó una etiqueta de grupo a cada cliente en función de su pertenencia a un clúster específico, permitiendo así la segmentación de clientes en grupos homogéneos en función de su proximidad espacial.

```python
# Configuración del modelo
model = AgglomerativeClustering(n_clusters=6, metric='euclidean', linkage='ward')

# Ajusta el modelo a los datos y obtiene los grupos
labels = model.fit_predict(data)
n_clusters = 6

# Obtener los clusters y sus centroides
clusters = {i: data[labels == i] for i in range(n_clusters)}
centroides = []
for cluster_id in range(n_clusters):
    cluster_coords = data[labels == cluster_id]
    centroid = cluster_coords.mean(axis=0)
    centroides.append(centroid)

centroides = np.array(centroides)
```

Figura 20– Implementación de Agglomerative Clustering en Python

Para analizar la distribución de los clientes, se calcularon los centroides de cada clúster. Esto se realizó extrayendo las coordenadas de los clientes en cada grupo y obteniendo el promedio de sus posiciones, almacenando estos valores en un array de NumPy. Finalmente, los resultados se visualizaron en un gráfico de dispersión, donde cada cliente se representa con un color según su clúster, y los centroides se resaltan con cruces rojas.

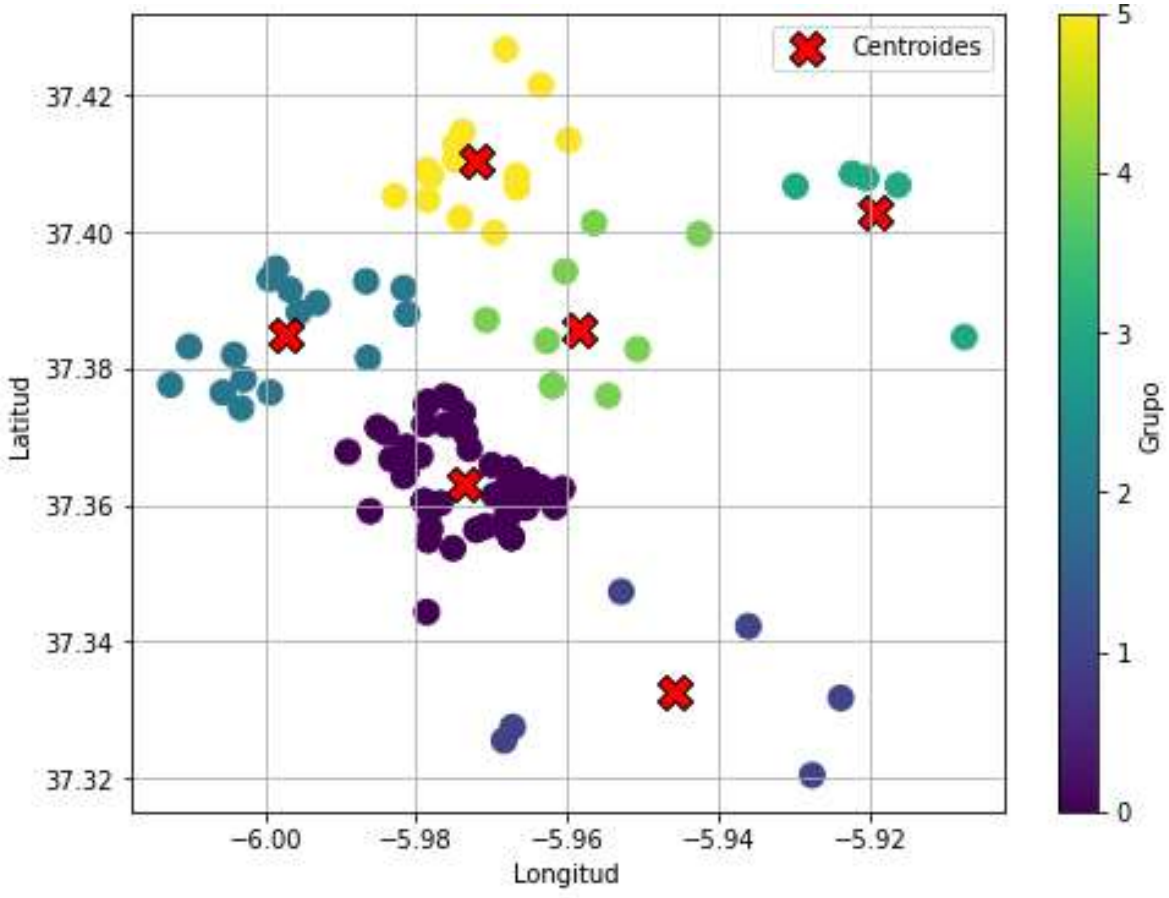

Figura 21– Gráfico de nodos con sus centroides

El gráfico final muestra cómo los clientes se agrupan en 6 regiones bien diferenciadas, lo que permite identificar patrones geográficos en la distribución de los datos. En conclusión, el uso del dendrograma permitió elegir de manera fundamentada 6 clústeres, asegurando una segmentación adecuada sin unificar en exceso a clientes que deberían mantenerse en grupos separados. Gracias a la combinación de la distancia euclidiana y el método de Ward, se obtuvo una estructura de clustering que refleja con precisión la dispersión de los clientes en el espacio geográfico.

3.2 Modelo

3.2.1 Simplificación del modelo inicial

En el artículo de Deutsch y Golany (2017) se propuso un modelo de optimización que, posteriormente, fue simplificado para facilitar su resolución a través de algoritmos diseñados para abordar problemas de Localización de Instalaciones No Capacitado (UFLP). Esta simplificación se llevó a cabo mediante la incorporación de nuevas variables y consideraciones específicas. En primer lugar, se introduce el parámetro γ_u, definido como $\gamma_u = R(1 - p^{\boldsymbol{u}}) + C_d^u p^u$, donde $R(1 - p^{\boldsymbol{u}})$ representa la pérdida de beneficio derivada de los clientes que rechazan viajar a su capa u-ésima y $C_d^u p^u$ corresponde al descuento otorgado a aquellos clientes que sí aceptan viajar hasta dicha capa.

Asimismo, para cada nodo i perteneciente al conjunto I, cada capa u y cada nodo j perteneciente al conjunto S_i^u, se define el coste, $C_{i,j} = Q_i \cdot \gamma_u$, donde $C_{i,j}$ representa el coste de asignación de clientes desde el nodo i hasta la instalación en el nodo j dentro de la capa u-ésima. Además, se incorpora un resultado derivado del Lema 1, el cual establece que, para cualquier combinación de i, u, m, j y S_i^u, siempre se cumple $C_{i,j} \geq 0$, lo cual se deduce a partir de la no negatividad de los valores de Q_i, R, $(1 - p^{\boldsymbol{u}})$ y $C_d^u p^u$.

Otro de los elementos clave en la simplificación del modelo es la introducción de la matriz de asignación Z, que determina la asignación de clientes desde un nodo i hasta una instalación en el nodo j. La variable $z_{i,j}$ puede tomar tres valores distintos: 1 si todos los clientes del nodo i son asignados a la instalación en el nodo j, un valor entre 0 y 1 si solo una fracción de los clientes del nodo i es trasladada al nodo j, y 0 en caso de que ningún cliente del nodo i sea asignado al nodo j.

La función objetivo del modelo simplificado busca minimizar el coste total del sistema, el cual incluye tanto los costes de instalación como los costes de servicio, expresado mediante la ecuación $\sum_{i\epsilon I} f_i x_i + \sum_{i\epsilon I} \sum_{j\epsilon I} C_{i,j}\, z_{i,j}$. Para que esta formulación sea válida, se deben cumplir diversas restricciones. La primera de ellas, $\sum_{j\epsilon I} z_{i,j} = 1$ para todo i, garantiza que cada demanda sea cubierta completamente a través de las instalaciones disponibles. La segunda restricción impone que $z_{i,j} \leq x_j$, asegurando que los clientes solo puedan ser asignados a instalaciones que efectivamente han sido abiertas. Adicionalmente, se establece que x_i es una variable binaria, indicando si una instalación está operativa o no. Finalmente, la restricción $z_{i,j} \geq 0$ impone la condición de no negatividad sobre las variables de asignación.

$$\min_{X,Z} \sum_{i\in I} f_i x_i + \sum_{i\in I}\sum_{j\in I} C_{i,j} z_{i,j} \tag{9a}$$

$$\text{s.t.:} \quad \sum_{j\in I} z_{i,j} = 1 \;\; \forall\, i \in I, \tag{9b}$$

$$z_{i,j} \leq x_j \;\; \forall\, i, j \in I, \tag{9c}$$

$$x_i \in \{0, 1\} \;\; \forall\, i \in I, \tag{9d}$$

$$z_{i,j} \geq 0 \;\; \forall\, i, j \in I. \tag{9e}$$

Figura 22– Modelo simplificado (Deutsch y Golany, 2017)

En resumen, la figura representa este modelo simplificado, cuyo objetivo es encontrar la mejor ubicación de instalaciones minimizando los costes, mientras se garantiza que la demanda de los clientes sea completamente satisfecha y que las asignaciones se realicen únicamente a instalaciones abiertas. Este enfoque permite una optimización más eficiente y práctica del problema original, asegurando una formulación más sencilla y resoluble mediante algoritmos especializados.

3.2.2 Modificaciones del modelo inicial

Para mejorar la precisión del modelo de optimización previamente descrito, se ha incorporado una modificación clave que permite que el porcentaje de clientes que aceptan viajar a una instalación dependa dinámicamente de la distancia entre los nodos. En la formulación original, el porcentaje de clientes que accedían a trasladarse a una determinada instalación estaba predefinido y no variaba en función de la distancia. Sin embargo, en muchos escenarios reales, la proximidad entre clientes e instalaciones juega un papel crucial en la decisión de desplazamiento.

Para abordar esta limitación, se introduce una función decreciente que ajusta el porcentaje de clientes que viajan ($\rho_{i,j}^{u}$) en función de la distancia entre los nodos i y j. En particular, se propone una función exponencial de la forma:

$$\rho_{i,j}^{u} = e^{-\alpha d_{i,j}} ,$$

donde $d_{i,j}$ representa la distancia entre ambos nodos y α es un parámetro de ajuste que regula la velocidad con la que disminuye el porcentaje de clientes dispuestos a viajar a medida que la distancia aumenta. Este ajuste permite modelar con mayor realismo la distribución geográfica de los clientes y su disposición a desplazarse según su cercanía a las instalaciones.

Adicionalmente, esta nueva formulación afecta directamente al cálculo del coste de asignación $C_{i,j}$. En la versión original del modelo, este coste dependía únicamente de una penalización fija y un descuento determinado. Sin embargo, con la nueva definición de $\rho_{i,j}^{u}$, la ecuación de coste se ha reformulado para incluir un esquema de descuentos dinámico. Ahora, en lugar de aplicar un descuento fijo para todos los clientes, este se asigna de manera proporcional a la distancia recorrida. Para ello, se ordenan los nodos de cada cliente en función de su distancia y se asignan los valores de descuento de forma creciente, de modo que los clientes que viajan distancias mayores reciben un mayor descuento. La nueva ecuación de coste se expresa como:

$$C_{i,j} = Q_i \left(R \left(1 - \rho_{i,j}^{u} \right) + C_d^{ordenado} \rho_{i,j}^{u} \right.$$

donde Q_i representa la demanda del nodo i, R es el coste base de asignación, y $C_d^{ordenado}$ es el vector de descuentos dinámico, que asigna mayores beneficios a los clientes que deben desplazarse distancias más largas.

Esta nueva formulación permite que el coste esté directamente influenciado por la proximidad entre clientes e instalaciones. Clientes más cercanos estarán más dispuestos a trasladarse, reduciendo así la pérdida de beneficio asociada a clientes que rechazan viajar y optimizando la asignación de recursos. Entre las principales ventajas de esta modificación se encuentra su alineación con el comportamiento real de los clientes, permitiendo una distribución más eficiente y realista de la demanda. Además, conserva la estructura base del modelo original, pero añade un nivel de detalle adicional que mejora su representatividad y facilita su aplicación en la toma de decisiones estratégicas sobre la localización de instalaciones. Gracias a esta mejora, el modelo adquiere una mayor flexibilidad y capacidad predictiva, haciéndolo más robusto en situaciones donde la distancia juega un papel crucial en la asignación de clientes.

3.2.3 Implementación del modelo en Python

3.2.3.1 Datos del modelo

En el modelo desarrollado, se han definido una serie de parámetros ajustables que permiten adaptar el cálculo de la matriz de costes en función de diferentes condiciones.

```
# Parámetros ajustables
R = 2
Cd = [0, 0.16, 0.32, 0.5, 0.7, 1]  # Coste de descuento
Q = [146, 18, 48, 15, 30, 42]   # Demanda en cada nodo
alpha = 20  # Parámetro para controlar la caída de rho con la distancia
```

Figura 23– Parámetros ajustables del modelo en Python

Uno de los parámetros clave es R, cuyo valor se ha fijado en 2. Este parámetro representa una penalización o pérdida asociada a los clientes que no viajan entre nodos. Su elección responde a la necesidad de reflejar el impacto negativo de no trasladar clientes entre ubicaciones, incentivando así la optimización de la asignación de viajes.

Otro parámetro relevante es C_d, que representa los costes de descuento aplicados a los clientes que aceptan viajar. En este caso, los valores asignados han sido $[0, 0.16, 0.32, 0.5, 0.7, 1]$, estableciendo un esquema de descuentos progresivos en función del nodo destino. Esta estrategia busca modelar escenarios en los que ciertos destinos resulten más atractivos debido a una reducción en el coste del traslado, favoreciendo así la movilidad de los clientes dentro de la red. En la nueva formulación del modelo, estos descuentos se asignan dinámicamente según la distancia relativa entre los nodos, de manera que los clientes que recorren mayores distancias reciben mayores descuentos.

Además, la demanda de cada nodo Q se ha definido de manera dinámica en función de la cantidad de clientes en cada nodo, en lugar de establecer valores fijos. Para calcular esta demanda, se ha considerado la distribución de los clientes en función de su latitud y longitud, agrupándolos según el nodo al que pertenecen. De esta forma, la demanda de cada nodo se obtiene directamente a partir del número total de clientes asignados a cada uno, permitiendo reflejar de manera más realista las variaciones en la concentración de clientes en los distintos puntos de la red. Este enfoque permite capturar fluctuaciones en la cantidad de personas que deben ser asignadas a instalaciones.

```
# Contamos cuántos clientes hay en cada nodo
N_clientes = np.bincount(labels)
# Definimos la demanda total esperada
Q_total = 300
# Ajustamos la demanda proporcionalmente al número de clientes en cada nodo
k = Q_total / np.sum(N_clientes)  # Factor de ajuste
Q = k * N_clientes  # Demanda por nodo ajustada
```

Figura 24– Cálculo de la demanda ajustada en cada nodo

Por otro lado, el parámetro α se ha fijado en 20 y desempeña un papel fundamental en la función que modela el porcentaje de clientes que aceptan viajar en función de la distancia. Su propósito es controlar la rapidez con la que este porcentaje disminuye conforme aumenta la distancia entre nodos. Al elegir un valor relativamente alto, se busca que la disminución en la proporción de clientes dispuestos a viajar sea más marcada en distancias cortas, priorizando la asignación de clientes a ubicaciones cercanas.

La combinación de estos parámetros ajustables permite que el modelo refleje con mayor precisión las condiciones reales de movilidad y asignación de clientes, proporcionando una mayor flexibilidad y capacidad de adaptación en distintos escenarios de optimización.

Con estos parámetros definidos, se procede al cálculo de la matriz de costes, utilizando la distancia entre los centroides de los clústeres obtenidos previamente. Para ello, se emplea la función cdist de scipy.spatial.distance, que permite calcular la matriz de distancias euclidianas entre los centroides de los distintos grupos. A partir de esta matriz, se determina el porcentaje de clientes que están dispuestos a viajar entre nodos mediante la función exponencial decreciente $\rho_{i,j}^{u} = e^{-\alpha d_{i,j}}$. Esta función permite modelar de manera realista la relación entre distancia y disposición a viajar, estableciendo que a medida que la distancia entre nodos aumenta, la cantidad de clientes dispuestos a trasladarse disminuye de manera exponencial.

Una vez calculados estos valores, se procede a la construcción de la matriz de costes $C_{i,j}$, que inicialmente es una matriz de ceros de tamaño $n \times n$, donde n representa el número total de nodos. Posteriormente, para cada par de nodos i y j, siempre que $i \neq j$, se calcula el coste de asignación de clientes con su expresión correspondiente.

$$C_{i,j} = Q_i \left(R \left(1 - \rho_{i,j}^{u} \right) + C_d^{ordenado} \rho_{i,j}^{u} \right.$$

```python
# Calcular la matriz de distancias entre los centroides
distancias = cdist(centroides, centroides, metric='euclidean')

# Calcular porcentaje basado en la distancia
porc = np.exp(-alpha * distancias)

n = len(centroides)
Cij = np.zeros((n, n))

nodos = list(range(n_clusters))
arcos = [(i, j) for i in nodos for j in nodos if i != j]
#print(centroides)

for i in range(n):  # Para cada nodo de origen
    distancia_promedio = np.mean(distancias[i])  # Distancia promedio del nodo i con el resto
    orden_nodos = np.argsort(distancias[i])  # Ordenar los nodos según distancia desde i

    # Crear un vector de descuentos dinámico
    Cd_ordenado = np.zeros(n)
    for idx, nodo in enumerate(orden_nodos):
        if nodo == i:
            Cd_ordenado[nodo] = 0  # Si es el mismo nodo, no tiene descuento
        else:
            Cd_ordenado[nodo] = Cd[idx - 1]  # Asigna el descuento ordenado (sin contar el nodo actual)
        # Calcular la matriz de costos considerando los descuentos dinámicos
    for j in range(n):
        if i != j:
            Cij[i, j] = Q[i] * (R * (1 - porc[i, j]) + Cd_ordenado[j] * porc[i, j])
```

Figura 25– Cálculo de la matriz de costes del modelo en Python

El resultado de este cálculo proporciona una matriz de costes que no solo tiene en cuenta la distancia entre nodos, sino que también refleja la disposición de los clientes a viajar y los incentivos económicos que puedan existir en cada destino. De esta forma, el modelo logra capturar de manera eficiente las relaciones de movilidad entre los nodos y permite optimizar la asignación de clientes a instalaciones minimizando el coste total del sistema.

Tabla 4– Matriz de costes del modelo

0.00	185.85	152.02	244.32	123.58	208.12
20.23	0.00	29.03	30.02	24.92	31.47
45.61	79.42	0.00	83.43	58.97	52.72
23.45	25.02	26.07	0.00	17.24	20.52
25.22	46.95	36.86	40.86	0.00	28.53
56.52	73.43	46.13	62.36	36.11	0.00

3.2.3.2 Modelo final en Gurobi

Gurobi es una de las bibliotecas de optimización más potentes y ampliamente utilizadas para la resolución de problemas de programación matemática. Se especializa en la resolución de modelos de optimización lineal, entero-mixto y cuadrático, ofreciendo algoritmos avanzados que garantizan soluciones eficientes incluso en problemas de gran escala. Su integración con Python permite una formulación sencilla e intuitiva de los modelos, aprovechando estructuras de datos flexibles y eficientes. Gracias a su alto rendimiento y capacidad para manejar múltiples restricciones y variables, Gurobi es una herramienta ideal para problemas de optimización en logística, planificación, asignación de recursos y muchas otras áreas de aplicación.

La implementación del modelo de optimización en Python se ha llevado a cabo utilizando esta biblioteca, permitiendo la formulación y resolución eficiente del problema. En primer lugar, se definen los parámetros principales del modelo, donde $fi = 50$ representa el coste fijo asociado a la activación de un nodo como instalación. Además, se introduce la matriz de costes Cij, previamente calculada. Esta matriz se define empleando NumPy para facilitar su manipulación. Asimismo, se establece $n = 6$ como el número total de nodos en la red, y se generan las listas *clientes*, que contiene los nodos clientes excluyendo el nodo base, y *nodos*, que incluye tanto el nodo base como los clientes. Posteriormente, se construye la lista *arcos*, la cual representa todas las conexiones posibles entre los nodos.

```
#Datos
fi = 50
Cij = np.array([( 0, 185.85399574, 152.02216159, 244.32257857, 123.58032693, 208.12243402),
 (20.23527635, 0, 29.03790455, 30.02042621, 24.92001739, 31.46983726),
 (45.60857554, 79.42358227, 0, 83.4268737, 58.97020903, 52.72430647 ),
 (23.44913124, 25.01702184, 26.07089803, 0, 17.24102419, 20.5178857),
 (25.22047488, 46.95291902, 36.85638064, 40.86153629, 0, 28.52628792),
 (56.51917889, 73.42962027, 46.13376816, 62.35604344, 36.10522075, 0 )])

#Nodos de clientes
n = 6
clientes = [i for i in range(n) if i!=0]
nodos = [0] + clientes
arcos = [(i,j) for i in nodos for j in nodos]
```

Figura 26– Datos del modelo

Tras definir los datos, se procede a la creación del modelo en Gurobi bajo el nombre *"Modelo_Lockers"*, que posteriormente se configurará para minimizar los costos. Para habilitar la visualización de información durante la ejecución del algoritmo, se activa el parámetro de salida mediante *model.setParam('OutputFlag', 1)*. Seguidamente, se definen las variables de decisión del modelo. En este caso, $x[i]$ es una variable binaria que indica si un nodo i es seleccionado como instalación (1) o no (0), mientras que $z[i,j]$ es una variable continua no negativa que representa el flujo de asignación de clientes desde el nodo i hacia el nodo j. Estas variables se declaran en el modelo utilizando la función *addVars* de Gurobi.

```
#Creación el modelo
model = Model("Modelo_Lockers")
model.setParam('OutputFlag', 1)
#Definición de variables
x = model.addVars(nodos, vtype=GRB.BINARY, name='x')
z = model.addVars(arcos, vtype=GRB.CONTINUOUS,lb = 0, name='z')
```

Figura 27– Creación del modelo y definición de variables de decisión

El modelo de optimización se implementa en Python utilizando la función *setObjective* de Gurobi, donde la función objetivo se define como la combinación de los costes fijos de activación de los nodos y los costes variables de asignación de clientes. En cuanto a las restricciones del modelo, se implementan mediante la función *addConstrs*, la cual permite definir múltiples restricciones de manera compacta y eficiente.

Para resolver el modelo de manera eficiente, se establece el método de optimización dual simplex mediante *model.Params.Method* = 1. Finalmente, se ejecuta el proceso de optimización con la instrucción *model.optimize*, lo que permite obtener la mejor asignación de clientes a instalaciones minimizando los costes totales. Gracias al uso de Gurobi, el modelo logra encontrar una solución óptima que satisface todas las restricciones impuestas y maximiza la eficiencia en la distribución de clientes dentro de la red.

```
#Definición de la función objetivo
model.setObjective(quicksum(x[i]*fi for i in nodos) + quicksum(Cij[i,j]*z[i,j] for i in nodos for j in nodos),
                   GRB.MINIMIZE)
#Definición de restricciones
model.addConstrs(quicksum(z[i,j] for j in nodos)==1 for i in nodos)
model.addConstrs(z[i,j] <= x[j] for i in nodos for j in nodos)
model.addConstrs(z[i,j] >= 0 for i in nodos for j in nodos)

model.Params.Method = 1  # Método dual simplex
model.optimize()
```

Figura 28– Definición del modelo y resolución

3.2.3.3 Exportación del modelo final

Para complementar la implementación del modelo y permitir su análisis detallado, se incluye la exportación de este a un archivo mediante la función *model.write*. Esta función permite generar una representación del modelo en un formato LP, facilitando su interpretación y depuración. Esta exportación resulta especialmente útil para validar que el modelo ha sido correctamente formulado y para su uso en otras herramientas de optimización que admitan este formato. Además, permite almacenar y compartir la formulación del problema sin necesidad de ejecutar nuevamente el código en Python, facilitando su documentación y replicabilidad en futuros estudios.

```
\ Model Modelo_Lockers
\ LP format - for model browsing. Use MPS format to capture full model
detail.
Minimize
  50 x[0] + 50 x[1] + 50 x[2] + 50 x[3] + 50 x[4] + 50 x[5]
   + 185.85399574 z[0,1] + 152.02216159 z[0,2] + 244.32257857 z[0,3]
   + 123.58032693 z[0,4] + 208.12243402 z[0,5] + 20.23527635 z[1,0]
   + 29.03790455 z[1,2] + 30.02042621 z[1,3] + 24.92001739 z[1,4]
   + 31.46983726 z[1,5] + 45.60857554 z[2,0] + 79.42358227 z[2,1]
   + 83.4268737 z[2,3] + 58.97020903 z[2,4] + 52.72430647 z[2,5]
   + 23.44913124 z[3,0] + 25.01702184 z[3,1] + 26.07089803 z[3,2]
   + 17.24102419 z[3,4] + 20.5178857 z[3,5] + 25.22047488 z[4,0]
   + 46.95291902 z[4,1] + 36.85638064 z[4,2] + 40.86153629 z[4,3]
   + 28.52628792 z[4,5] + 56.51917889 z[5,0] + 73.42962027 z[5,1]
   + 46.13376816 z[5,2] + 62.35604344 z[5,3] + 36.10522075 z[5,4]
Subject To
 R0: z[0,0] + z[0,1] + z[0,2] + z[0,3] + z[0,4] + z[0,5] = 1
 R1: z[1,0] + z[1,1] + z[1,2] + z[1,3] + z[1,4] + z[1,5] = 1
 R2: z[2,0] + z[2,1] + z[2,2] + z[2,3] + z[2,4] + z[2,5] = 1
 R3: z[3,0] + z[3,1] + z[3,2] + z[3,3] + z[3,4] + z[3,5] = 1
 R4: z[4,0] + z[4,1] + z[4,2] + z[4,3] + z[4,4] + z[4,5] = 1
 R5: z[5,0] + z[5,1] + z[5,2] + z[5,3] + z[5,4] + z[5,5] = 1
 R6: - x[0] + z[0,0] <= 0
 R7: - x[1] + z[0,1] <= 0
 R8: - x[2] + z[0,2] <= 0
 R9: - x[3] + z[0,3] <= 0
 R10: - x[4] + z[0,4] <= 0
 R11: - x[5] + z[0,5] <= 0
 R12: - x[0] + z[1,0] <= 0
 R13: - x[1] + z[1,1] <= 0
 R14: - x[2] + z[1,2] <= 0
 R15: - x[3] + z[1,3] <= 0
 R16: - x[4] + z[1,4] <= 0
 R17: - x[5] + z[1,5] <= 0
 R18: - x[0] + z[2,0] <= 0
 R19: - x[1] + z[2,1] <= 0
 R20: - x[2] + z[2,2] <= 0
 R21: - x[3] + z[2,3] <= 0
 R22: - x[4] + z[2,4] <= 0
 R23: - x[5] + z[2,5] <= 0
 R24: - x[0] + z[3,0] <= 0
 R25: - x[1] + z[3,1] <= 0
 R26: - x[2] + z[3,2] <= 0
 R27: - x[3] + z[3,3] <= 0
 R28: - x[4] + z[3,4] <= 0
 R29: - x[5] + z[3,5] <= 0
 R30: - x[0] + z[4,0] <= 0
```

```
 R31: - x[1] + z[4,1] <= 0
 R32: - x[2] + z[4,2] <= 0
 R33: - x[3] + z[4,3] <= 0
 R34: - x[4] + z[4,4] <= 0
 R35: - x[5] + z[4,5] <= 0
 R36: - x[0] + z[5,0] <= 0
 R37: - x[1] + z[5,1] <= 0
 R38: - x[2] + z[5,2] <= 0
 R39: - x[3] + z[5,3] <= 0
 R40: - x[4] + z[5,4] <= 0
 R41: - x[5] + z[5,5] <= 0
 R42: z[0,0] >= 0
 R43: z[0,1] >= 0
 R44: z[0,2] >= 0
 R45: z[0,3] >= 0
 R46: z[0,4] >= 0
 R47: z[0,5] >= 0
 R48: z[1,0] >= 0
 R49: z[1,1] >= 0
 R50: z[1,2] >= 0
 R51: z[1,3] >= 0
 R52: z[1,4] >= 0
 R53: z[1,5] >= 0
 R54: z[2,0] >= 0
 R55: z[2,1] >= 0
 R56: z[2,2] >= 0
 R57: z[2,3] >= 0
 R58: z[2,4] >= 0
 R59: z[2,5] >= 0
 R60: z[3,0] >= 0
 R61: z[3,1] >= 0
 R62: z[3,2] >= 0
 R63: z[3,3] >= 0
 R64: z[3,4] >= 0
 R65: z[3,5] >= 0
 R66: z[4,0] >= 0
 R67: z[4,1] >= 0
 R68: z[4,2] >= 0
 R69: z[4,3] >= 0
 R70: z[4,4] >= 0
 R71: z[4,5] >= 0
 R72: z[5,0] >= 0
 R73: z[5,1] >= 0
 R74: z[5,2] >= 0
 R75: z[5,3] >= 0
 R76: z[5,4] >= 0
 R77: z[5,5] >= 0
Bounds
Binaries
 x[0] x[1] x[2] x[3] x[4] x[5]
End
```

4 RESULTADOS Y DISCUSIÓN

En esta sección, se presentan los resultados obtenidos tras la resolución del modelo y se analiza su impacto en la asignación de clientes a las instalaciones. Evaluar estos resultados permite comprobar la coherencia de la solución y su aplicabilidad en la toma de decisiones. Además, se realiza un análisis de sensibilidad para examinar cómo variaciones en ciertos parámetros clave pueden afectar la solución óptima. Este tipo de análisis es fundamental para evaluar la estabilidad del modelo ante cambios en factores como los costes de traslado, la demanda de los clientes o los parámetros de penalización y descuento. A través de esta evaluación, se busca no solo validar la solución obtenida, sino también identificar oportunidades de mejora y ajustes que permitan que el modelo sea más flexible y eficiente en distintos escenarios operativos.

4.1 Resultados obtenidos

Los resultados obtenidos tras la ejecución del modelo muestran una solución óptima con un valor objetivo de 211.582, el cual representa el coste total resultante de la activación de instalaciones y la asignación de clientes a las mismas. Este valor objetivo surge de la combinación del coste fijo de abrir instalaciones en determinados nodos y el coste variable asociado a la asignación de los clientes a dichas instalaciones, lo que refleja la estructura de costes determinada por la matriz de costes establecida previamente. En la solución obtenida, el modelo ha decidido activar instalaciones en los nodos 0 y 5, descartando la apertura de instalaciones en los demás nodos. Esta decisión se expresa matemáticamente mediante el vector de decisión x, definido como:

$$x = [1\ 0\ 0\ 0\ 0\ 1]$$

donde un 1 indica que el nodo correspondiente ha sido seleccionado como instalación y un 0 significa que dicho nodo no será utilizado para este propósito. Esto implica que únicamente los nodos 0 y 5 servirán como centros de atención a los clientes, mientras que los nodos restantes deberán ser asignados a una de estas dos instalaciones.

En cuanto a la asignación de clientes a instalaciones, el modelo ha distribuido a los nodos de acuerdo con la siguiente matriz de asignación Z:

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Cada fila de esta matriz representa un nodo cliente, mientras que cada columna representa un nodo instalación. Un valor 1 en la posición $Z[i, j]$ indica que el nodo i ha sido asignado a la instalación ubicada en el nodo j, mientras que un 0 indica que no existe tal asignación. Analizando la estructura de esta matriz, se observa que los nodos 0, 1, 2 y 4 han sido asignados a la instalación ubicada en el nodo 0, mientras que los nodos 3 y 5 se ha asignado a una instalación ubicada en el nodo 5.

Este resultado presenta algunas implicaciones interesantes. En primer lugar, se observa que el modelo ha optado por una solución en la que se minimiza el número de instalaciones activadas, favoreciendo una estructura en la que un solo nodo atiende a la mayoría de los clientes, en lugar de distribuir las asignaciones de manera más equitativa entre las instalaciones disponibles. Esto sugiere que los costes fijos de apertura de instalaciones pueden ser lo suficientemente elevados como para desincentivar la activación de más nodos como puntos de servicio. En segundo lugar, si bien la asignación de los clientes es óptima desde el punto de vista del modelo, se puede notar que algunos nodos han sido asignados a instalaciones que no necesariamente están más cercanas a ellos en términos de distancia geográfica. Es posible que el modelo haya priorizado la reducción del número de instalaciones activadas sobre la minimización de los costes de traslado individuales de cada cliente.

Este comportamiento plantea la necesidad de realizar un análisis más detallado para determinar si sería conveniente ajustar algunos parámetros del modelo o introducir restricciones adicionales que permitan un balance más equitativo entre la minimización de costes fijos y los costes de transporte. Entre las posibles modificaciones que podrían explorarse, se podría considerar la introducción de una penalización más estricta sobre los costes de traslado o la inclusión de restricciones que impongan límites sobre la distancia máxima permitida para la asignación de clientes a instalaciones. También se podría evaluar el impacto de modificar los costes fijos de activación de los nodos para analizar si existe un umbral en el que resulte más conveniente habilitar más instalaciones en lugar de concentrar la asignación en pocas ubicaciones.

En conclusión, los resultados obtenidos han proporcionado una solución factible y óptima en términos del modelo planteado, pero su interpretación revela ciertos aspectos que podrían ser objeto de ajustes adicionales para mejorar la eficiencia del sistema. Aunque el modelo ha logrado minimizar los costes totales, el patrón de asignación observado sugiere que sería recomendable realizar un análisis de sensibilidad para evaluar cómo variaciones en los parámetros del problema podrían influir en la estructura final de la solución, lo que permitirá determinar si es posible obtener configuraciones alternativas que resulten más adecuadas en la práctica.

4.2 Análisis de sensibilidad

Con el fin de evaluar la robustez del modelo y comprender cómo distintos factores influyen en la solución óptima, se han llevado a cabo análisis de sensibilidad sobre tres aspectos clave: el coste de instalación, la matriz de costes de asignación y la disponibilidad de ubicaciones. Estos análisis permiten identificar cómo cambios en los parámetros afectan en la asignación de lockers y en el valor de la función objetivo, proporcionando información valiosa para la toma de decisiones estratégicas.

Estos análisis no solo ayudan a entender el comportamiento del modelo bajo distintas condiciones, sino que también permiten anticipar el impacto de cambios en el entorno operativo y optimizar la toma de decisiones en la implementación real del sistema.

4.2.1 Análisis según el coste fijo de apertura de instalaciones

El análisis de sensibilidad realizado evalúa cómo afecta el coste de instalación fi a la asignación óptima de lockers y al valor de la función objetivo. Se han probado distintos valores:

$$fi = [20, 30, 40, 50, 60, 70, 80]$$

observándose que a medida que el coste de instalación aumenta, la cantidad de lockers instalados disminuye. Inicialmente, con $fi = 20$, se instalan cinco lockers, mientras que cuando fi alcanza 60 o más, solo se mantiene un locker en funcionamiento. Este comportamiento impacta directamente en la función objetivo, cuyo valor crece con el incremento de fi, lo que es consistente con la reducción en la cantidad de lockers disponibles y el posible aumento en los costes de servicio debido a la mayor distancia entre clientes y lockers abiertos. Se identifica un punto crítico en $fi = 50$, donde la cantidad de lockers activos se reduce de tres a solo dos, y a partir de $fi = 60$, el modelo prioriza minimizar costes de instalación sacrificando accesibilidad, dejando operativa una única instalación. Estos resultados reflejan la sensibilidad del modelo ante variaciones en los costes de instalación y pueden servir como herramienta para la toma de decisiones estratégicas sobre la apertura y localización de lockers.

Tabla 5– Resultados con los distintos valores de coste de instalación

Fi	Valor Función Objetivo	x[0]	x[1]	x[2]	x[3]	x[4]	x[5]
20	117.24	1	1	1	0	1	1
30	155.97	1	0	1	0	0	1
40	185.97	1	0	1	0	0	1
50	211.58	1	0	0	0	0	1
60	231.03	1	0	0	0	0	0
70	241.03	1	0	0	0	0	0
80	251.03	1	0	0	0	0	0

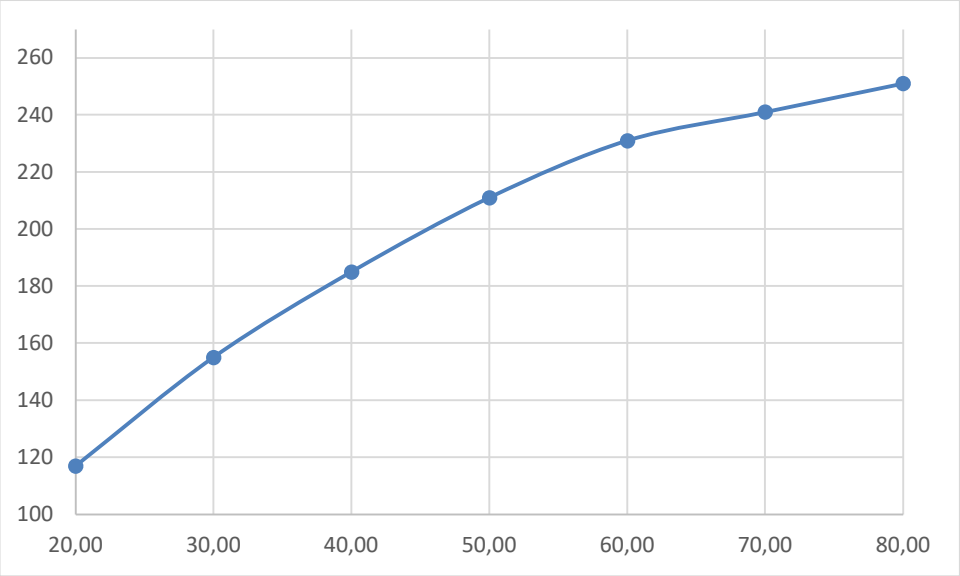

Figura 29– Gráfico del valor de la función objetivo en función del coste de instalación

El gráfico presentado complementa los resultados mostrados en la tabla, ilustrando visualmente la relación entre el incremento en el coste de instalación y la evolución de la función objetivo. Se observa que a medida que el coste de instalación aumenta, el valor de la función objetivo sigue una tendencia creciente, lo que es consistente con la reducción en el número de lockers instalados. En los primeros valores de fi, el modelo permite la apertura de múltiples lockers, lo que mantiene bajos los costes de asignación, sin embargo, conforme el coste de instalación se incrementa, el modelo tiende a cerrar instalaciones y a concentrar la asignación de clientes en un menor número de lockers, lo que incrementa la función objetivo debido a los mayores costos de transporte. Se identifica un punto de inflexión en $fi = 50$, donde el número de lockers activos se reduce de tres a dos, y a partir de $fi = 60$, el modelo opta por mantener solo una instalación operativa, maximizando la reducción de costes fijos a costa de un aumento en los costes de asignación. Este comportamiento se refleja en el gráfico con una curva de crecimiento no lineal, lo que sugiere que los cambios en la función objetivo no son uniformes y dependen de la estructura de costes y de la disposición de los lockers. La representación gráfica facilita la interpretación de la sensibilidad del modelo ante variaciones en el coste de instalación, permitiendo visualizar claramente cómo la optimización del sistema se ve afectada por este parámetro y proporcionando información relevante para la toma de decisiones estratégicas en la ubicación y cantidad de lockers a instalar.

4.2.2 Análisis según los costes de asignación

Para realizar el análisis de sensibilidad con respecto a los costes de asignación, se ha seguido un proceso estructurado que permite evaluar cómo varían la función objetivo y la configuración de lockers activos en función de distintos factores de ajuste en la matriz de costes. Inicialmente, se ha definido un modelo de optimización basado en Gurobi, donde se consideran tanto los costes de instalación de cada locker como los costes de asignación de clientes a lockers abiertos. La matriz de costes representa la distancia o penalización por asignar cada cliente a una determinada instalación, y es precisamente esta matriz la que se ha modificado aplicando distintos factores multiplicativos. Para llevar a cabo el análisis, se han evaluado seis escenarios en los que los costes han sido escalados por factores de $[0.8, 1, 1.2, 1.5, 2\ y\ 2.5]$, permitiendo así observar la sensibilidad del modelo ante cambios en la estructura de costes.

Tabla 6– Resultados según los distintos factores respecto a la matriz de costes

Factor	Valor Función Objetivo	x[0]	x[1]	x[2]	x[3]	x[4]	x[5]
0,8	186,8261095	1,00	0,00	0,00	0,00	0,00	0,00
1	211,5822125	1,00	0,00	0,00	0,00	0,00	1,00
1,2	229,1683643	1,00	0,00	1,00	0,00	0,00	1,00
1,5	248,9604554	1,00	0,00	1,00	0,00	0,00	1,00
2	274,9526011	1,00	0,00	1,00	0,00	1,00	1,00
2,5	293,1025605	1,00	1,00	1,00	0,00	1,00	1,00

Los resultados muestran que cuando el factor es 0.8, la función objetivo toma un valor de 186.83 y solo se activa la instalación en el nodo 0, lo que indica que con costes de asignación más bajos el modelo prioriza minimizar los costes fijos de instalación y concentra toda la demanda en un único locker. A medida que el factor aumenta hasta 1, la función objetivo crece hasta 211.58 y se activan los lockers en los nodos 0 y 5, lo que sugiere que el incremento en los costes de asignación hace que el modelo opte por abrir una instalación adicional para reducir la distancia total recorrida por los clientes. Con un factor de 1.2, el valor de la función objetivo asciende a 229.17 y se observa una redistribución en la apertura de lockers, con nodos activos en 0, 2 y 5, lo que indica que el sistema está ajustando la localización de lockers para equilibrar costes fijos y de asignación. Cuando el factor aumenta a 1.5, la función objetivo alcanza 248.96 y se mantienen activos los lockers en los nodos 0, 2 y 5, lo que implica que el incremento en los costes de asignación incentiva la apertura de más instalaciones para reducir la distancia de servicio. En el caso de un factor 2, la función objetivo sube a 274.95 y el número de lockers activos aumenta a cuatro, con nodos abiertos en 0, 2, 4 y 5, lo que confirma la tendencia de expansión en la infraestructura cuando los costes de asignación crecen. Finalmente, con un factor de 2.5, el valor de la función objetivo alcanza 293.10 y se activan lockers en los nodos 0, 1, 2, 4 y 5, lo que representa el caso en el que el modelo prioriza minimizar los costes de asignación aun a costa de una mayor inversión en apertura de lockers.

Estos resultados reflejan que el modelo es altamente sensible a los cambios en los costes de asignación, mostrando una clara tendencia a abrir más lockers conforme estos aumentan, lo que sugiere que los costes de servicio tienen un peso importante en la toma de decisiones sobre la infraestructura óptima. Además, se identifican puntos críticos en los que se produce la apertura de nuevas instalaciones, indicando que hay umbrales específicos a partir de los cuales es más rentable habilitar un nuevo locker que asumir mayores costes de asignación. Este análisis proporciona información valiosa para la toma de decisiones estratégicas sobre la expansión y localización de lockers, permitiendo evaluar el impacto de distintos escenarios de costes en la eficiencia del sistema.

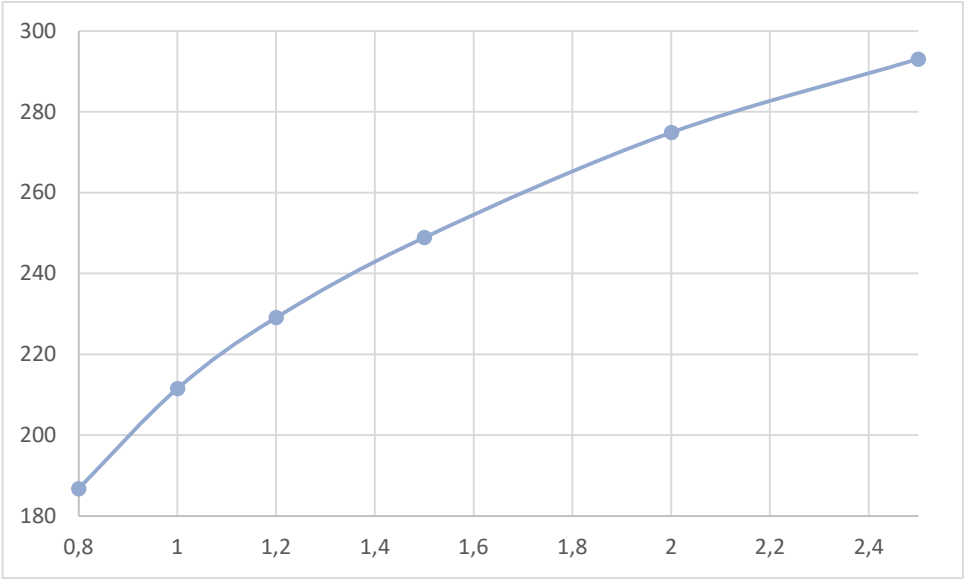

Figura 30– Gráfico del valor de la función objetivo en función de los distintos factores de modificación de la matriz de costes

El gráfico representa la evolución del valor de la función objetivo en función del factor de modificación aplicado a la matriz de costes de asignación. Se observa una relación creciente entre ambos parámetros, lo que indica que a medida que se incrementan los costes de asignación mediante los factores considerados, el valor de la función objetivo aumenta de manera no lineal. Inicialmente, cuando el factor es 0.8, la función objetivo toma un valor de aproximadamente 186.83, y conforme el factor se eleva hasta 1, el valor asciende a 211.58. Este comportamiento sugiere que con menores costes de asignación, el modelo opta por minimizar la cantidad de lockers abiertos, concentrando la demanda en menos instalaciones para reducir los costes fijos de instalación.

A medida que el factor se incrementa a 1.2, la función objetivo alcanza un valor de 229.17, lo que refleja un cambio en la estrategia del modelo, priorizando una mejor distribución de los lockers en la red para mitigar los mayores costes de asignación. Cuando el factor se incrementa a 1.5, la función objetivo sigue aumentando hasta 248.96, y con un factor de 2, el valor crece a 274.95. En estos casos, la tendencia muestra que a medida que los costes de asignación se encarecen, el modelo opta por abrir más lockers para reducir la distancia de servicio entre clientes y lockers operativos. Finalmente, cuando el factor llega a 2.5, la función objetivo alcanza su valor máximo en 293.10, reflejando el caso en el que el modelo prioriza minimizar los costes de asignación, aunque esto implique abrir un mayor número de lockers.

El comportamiento del gráfico sugiere que la relación entre el factor de modificación y la función objetivo sigue un patrón de crecimiento cóncavo, donde los aumentos iniciales en los costes de asignación generan un impacto más pronunciado en la función objetivo, pero conforme el factor continúa creciendo, el incremento en el valor de la función objetivo se estabiliza. Esto puede deberse a que, en los primeros aumentos, la apertura de nuevos lockers contribuye significativamente a la reducción de los costes de asignación, mientras que en los factores más altos el beneficio marginal de abrir nuevos lockers disminuye. En conclusión, este análisis de sensibilidad demuestra cómo el modelo responde ante variaciones en los costes de asignación, evidenciando que existe un equilibrio entre la cantidad de lockers abiertos y la minimización de los costes totales del sistema.

4.2.3 Análisis según la disponibilidad de ubicaciones

Para analizar la sensibilidad del modelo frente a la disponibilidad de ubicaciones, se han definido distintos escenarios en los que se restringe el conjunto de ubicaciones posibles donde se pueden instalar lockers. Se ha mantenido el modelo original, asegurando que la formulación y restricciones siguen siendo las mismas, pero se han agregado restricciones adicionales para forzar que ciertos lockers no puedan ser utilizados en determinados escenarios. Posteriormente, se ha ejecutado el modelo en cada uno de estos escenarios, obteniendo el valor de la función objetivo y las ubicaciones activadas en cada caso.

Tabla 7– Resultados según las disponibilidades de ubicación

Ubicaciones Disponibles	Valor Función Objetivo	x[0]	x[1]	x[2]	x[3]	x[4]	x[5]
[0, 1, 2, 3, 4, 5]	211,5822125	1,00	0,00	0,00	0,00	0,00	1,00
[1, 2, 3, 4]	301,8465893	0,00	0,00	1,00	0,00	1,00	0,00
[0, 2, 4]	215,0386506	1,00	0,00	1,00	0,00	0,00	0,00
[1, 3, 5]	387,6224758	0,00	1,00	0,00	0,00	0,00	1,00
[0, 1, 2]	215,0386506	1,00	0,00	1,00	0,00	0,00	0,00
[3, 4, 5]	310,8167983	0,00	0,00	0,00	0,00	1,00	0,00

En la tabla de resultados, se observa que cuando todas las ubicaciones están disponibles, el valor de la función objetivo es el más bajo, con un costo de 211.58 y las ubicaciones activas siendo [0, 5]. A medida que se restringen ciertas ubicaciones, el modelo se ve forzado a utilizar opciones óptimas, lo que aumenta el costo total del sistema. Por ejemplo, cuando solo se permiten las ubicaciones [1, 3, 5], el costo se incrementa considerablemente hasta 387.62, debido a que estas ubicaciones pueden no ser las más eficientes en términos de costes de instalación y distribución de la demanda. En el caso de otros escenarios como [0, 2, 4] o [0, 1, 2], el valor de la función objetivo es idéntico y cercano al óptimo inicial, lo que sugiere que las ubicaciones 0, 2 y 4 ofrecen soluciones similares en términos de eficiencia.

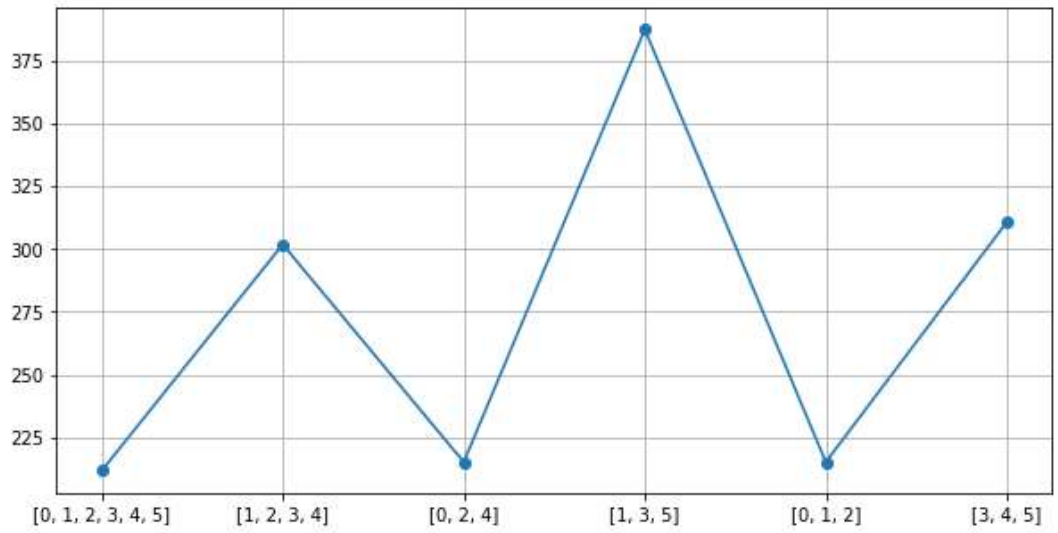

Figura 31– Gráfico del valor de la función objetivo en función de los distintos escenarios

El gráfico representado visualiza estos resultados, mostrando cómo varía la función objetivo según la disponibilidad de ubicaciones. Se observa una tendencia donde los costes aumentan cuando se eliminan ubicaciones estratégicas, generando picos en la función objetivo. El punto más bajo se mantiene en los escenarios donde se permite la ubicación 0, lo que indica que esta ubicación es clave para mantener bajos costes. En contraste, los picos más altos del gráfico reflejan escenarios en los que se restringen opciones importantes y se obliga al modelo a usar ubicaciones menos eficientes.

En conclusión, este análisis de sensibilidad demuestra que ciertas ubicaciones son fundamentales para mantener un coste bajo en la solución óptima del modelo. Restringir estas ubicaciones puede impactar negativamente en la eficiencia del sistema, aumentando significativamente los costes. Esta información es útil para la toma de decisiones estratégicas, ya que permite identificar qué ubicaciones son críticas y cuáles pueden ser prescindibles sin afectar demasiado el desempeño del sistema.

5 CONCLUSIÓN

El presente trabajo ha abordado de manera integral el problema de la logística de última milla mediante el estudio y aplicación de modelos matemáticos para la ubicación óptima de Automated Parcel Lockers (APL). A lo largo del desarrollo de este proyecto, se ha llevado a cabo una investigación exhaustiva sobre la problemática inherente a la distribución urbana de paquetes, identificando los principales desafíos y soluciones alternativas. En particular, se ha analizado la creciente implementación de parcel lockers como una estrategia eficaz para mitigar los inconvenientes del reparto tradicional y reducir los costes logísticos y el impacto ambiental.

Para abordar el problema de ubicación de APL, se ha adoptado y modificado un modelo matemático propuesto por Deutsch y Golany (2017), el cual ha sido ajustado a las particularidades del contexto analizado. La formulación matemática ha sido complementada con técnicas de procesamiento de datos basadas en clustering jerárquico aglomerativo, lo que ha permitido estructurar los datos de manera óptima para la implementación del modelo. La segmentación de las zonas de distribución en función de características espaciales y de demanda ha permitido una mejor representación de la realidad operativa de los parcel lockers.

La implementación del modelo ha sido realizada en Python, utilizando herramientas de optimización matemática y procesamiento de datos. Se han llevado a cabo diversas modificaciones al modelo original con el fin de adaptarlo a los requerimientos específicos del problema, ajustando restricciones y parámetros para mejorar su aplicabilidad en diferentes escenarios. A partir de la ejecución del modelo, se han obtenido resultados que han permitido identificar ubicaciones estratégicas para la instalación de APL, optimizando tanto los costes de instalación como la accesibilidad para los usuarios finales.

Con el objetivo de evaluar la robustez del modelo propuesto, se ha llevado a cabo un análisis de sensibilidad basado en tres criterios fundamentales: el coste fijo de apertura de instalaciones, los costes de asignación y la disponibilidad de ubicaciones. El análisis ha permitido identificar cómo la variación de estos parámetros afecta la solución óptima, ofreciendo información clave para la toma de decisiones estratégicas. Se ha observado que el aumento del coste de instalación tiende a reducir la cantidad de lockers operativos, priorizando la minimización de costes en detrimento de la accesibilidad. Por otro lado, la modificación de los costes de asignación ha mostrado un impacto significativo en la distribución de la demanda, alterando la configuración óptima de lockers activos. Finalmente, la variación en la disponibilidad de ubicaciones ha evidenciado la importancia de la flexibilidad en la elección de emplazamientos, dado que restricciones excesivas pueden incrementar de manera sustancial el valor de la función objetivo.

En términos generales, los resultados obtenidos han demostrado la aplicabilidad y utilidad del modelo en la planificación de redes de parcel lockers. La combinación de técnicas de clustering con modelos de optimización matemática ha permitido desarrollar un enfoque sólido y escalable para abordar problemas de ubicación en la logística de última milla. Asimismo, el análisis de sensibilidad ha proporcionado información valiosa para la toma de decisiones, destacando la importancia de considerar distintos escenarios y parámetros en la planificación estratégica.

A pesar de los logros alcanzados, este trabajo también deja abiertas diversas líneas de investigación futuras. En primer lugar, sería conveniente explorar modelos más dinámicos que permitan considerar variaciones temporales en la demanda y en los costes operativos. Además, la inclusión de restricciones adicionales, como limitaciones de capacidad o políticas de sostenibilidad, podría enriquecer el análisis y hacerlo más representativo de la realidad operativa. Asimismo, la implementación de metodologías de optimización más avanzadas, como heurísticas o metaheurísticas, podría mejorar la eficiencia computacional del modelo y permitir su aplicación en contextos de mayor escala.

En conclusión, este proyecto ha aportado un marco metodológico para la optimización de la ubicación de parcel lockers en la logística de última milla, combinando teoría y práctica en un enfoque innovador y aplicable a entornos reales. Los resultados obtenidos resaltan la importancia de integrar técnicas de optimización y análisis de datos en la toma de decisiones estratégicas, ofreciendo una base sólida para futuras investigaciones y aplicaciones en el ámbito logístico.

BIBLIOGRAFÍA

Bosona, T. (2020). Urban Freight Last Mile Logistics—Challenges and Opportunities to Improve Sustainability: A Literature Review. *Sustainability*, *12*(21), 8769. https://doi.org/10.3390/su12218769

Buzzega, G., Novellani, S. Last mile deliveries with lockers: formulations and algorithms. *Soft Comput* **27**, 12843–12861 (2023). https://doi.org/10.1007/s00500-021-06592-6

Dell'Amico, M., & Hadjidimitriou, S. (2012). Innovative logistics model and containers solution for efficient last mile delivery. *Procedia - Social and Behavioral Sciences, 48*, 1505–1514. https://doi.org/10.1016/j.sbspro.2012.06.1126

Deloitte. Last Mile Logistics Challenges and solutions in Spain (2020) https://www.deloitte.com/es/es/services/consulting/research/logistica-de-ultima-milla.html

Deutsch, Y., & Golany, B. (2017). A parcel locker network as a solution to the logistics last mile problem. *International Journal of Production Research*, *56*(1–2), 251–261. https://doi.org/10.1080/00207543.2017.1395490

Ehmke, J. F., & Mattfeld, D. C. (2012). Vehicle routing for attended home delivery in city logistics. *Procedia - Social and Behavioral Sciences, 39*, 622–632. https://doi.org/10.1016/j.sbspro.2012.03.135

El Moussaoui, A. E., Benbba, B., Jaegler, A., El Andaloussi, Z., & El Amrani, L. (2022). *Last mile logistics: Challenges & improvement ways*. In 2022 14th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA). IEEE. https://doi.org/10.1109/LOGISTIQUA55056.2022.9938101

Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2009). *Characteristics of innovations in last mile logistics: Using best practices, case studies and making the link with green and sustainable logistics*. European Transport Conference. https://www.researchgate.net/publication/341980496_Characteristics_of_innovations_in_last_mile_logistics_-using_best_practices_case_studies_and_making_the_link_with_green_and_sustainable_logistics

IBM. (2024, 21 de febrero). *¿Qué es el clustering?* IBM Think. https://www.ibm.com/es-es/think/topics/clustering#:~:text=El%20clustering%20es%20un%20algoritmo,basados%20en%20similitudes%20o%20patrones.

Iwan, S., Kijewska, K., & Lemke, J. (2016). Analysis of parcel lockers' efficiency as the last mile delivery solution – The results of the research in Poland. *Transportation Research Procedia, 12*, 644-655. https://doi.org/10.1016/j.trpro.2016.02.018

Lachapelle, U., Burke, M., Brotherton, A., & Leung, A. (2018). Parcel locker systems in a car dominant city: Location, characterisation and potential impacts on city planning and consumer travel access. *Journal of Transport Geography, 71*, 1–14. https://doi.org/10.1016/j.jtrangeo.2018.06.022

Lin, Y., Wang, Y., Lee, L. H., & Chew, E. P. (2022). Profit-maximizing parcel locker location problem under threshold Luce model. *Transportation Research Part E: Logistics and Transportation Review, 157*, 102541. https://doi.org/10.1016/j.tre.2021.102541

Log4.pl. (2013). *Paczkomaty InPost – Ekspertyza AGH*. Log4. https://log4.pl/paczkomaty-inpost-ekspertyza-agh,12,9270.htm

Lorenzo-Espejo, A., Aparicio-Ruiz, P., Muñuzuri, J., & Pegado-Bardayo, A. (2024). Cost analysis in last mile logistics: Modelling the impact of uncertainty in transport costs. En J. Bautista-Valhondo, M. Mateo-Doll, A. Lusa, & R. Pastor-Moreno (Eds.), *Proceedings of the 17th International Conference on Industrial Engineering and Industrial Management (ICIEIM) – XXVII Congreso de Ingeniería de Organización (CIO2023)* (Vol. 206). *Lecture Notes on Data Engineering and Communications Technologies.* Springer, Cham. https://doi.org/10.1007/978-3-031-57996-7_62

Lumbreras Herrera, M. I. (2020). *Evaluación de análisis de clustering jerárquico en datos moleculares de alta dimensión* (Tesis de máster). Universidad Carlos III de Madrid. https://openaccess.uoc.edu/bitstream/10609/120648/6/milumbrerasTFM0620memoria.pdf

Martín Gallardo, E. (2018). *Técnicas de clustering aplicadas a la resolución de problemas de optimización combinatoria con restricciones espaciales y temporales* (Tesis doctoral). Universidad Carlos III de Madrid. https://e-archivo.uc3m.es/rest/api/core/bitstreams/c29a482b-9d9c-41f2-ab2e-c3ed256c7d43/content

Mitrea, I. A., Zenezini, G., De Marco, A., Ottaviani, F. M., Delmastro, T., & Botta, C. (2020). Estimating e-consumers' attitude towards parcel locker usage. *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 1731–1736. https://doi.org/10.1109/COMPSAC48688.2020.000-5

Oliveira, L. K. d., Oliveira, I. K. d., França, J. G. d. C. B., Balieiro, G. W. N., Cardoso, J. F., Bogo, T., Bogo, D., & Littig, M. A. (2022). Integrating Freight and Public Transport Terminals Infrastructure by Locating Lockers: Analysing a Feasible Solution for a Medium-Sized Brazilian Cities. *Sustainability*, *14*(17), 10853. https://doi.org/10.3390/su141710853

Ottaviani, F. M., Zenezini, G., De Marco, A., & Carlin, A. (2023). Locating Automated Parcel Lockers (APL) with known customers' demand: a mixed approach proposal. *European Journal of Transport and Infrastructure Research*, *23*(2), 24–45. https://doi.org/10.18757/ejtir.2023.23.2.6786

Savelsbergh, M., & Van Woensel, T. (2016). 50th anniversary invited article—City logistics: Challenges and opportunities. *Transportation Science, 50*(2), 363-761. https://doi.org/10.1287/trsc.2016.0675

Visser, J., Nemoto, T., & Browne, M. (2014). Home delivery and the impacts on urban freight transport: A review. *Procedia - Social and Behavioral Sciences, 125*, 15-27. https://doi.org/10.1016/j.sbspro.2014.01.1452

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Yuen, K. F., Wang, X., Ng, L. T. W., & Wong, Y. D. (2018). An investigation of customers' intention to use self-collection services for last-mile delivery. *Transport Policy, 66*, 1-8. https://doi.org/10.1016/j.tranpol.2018.03.001