

**Tipo de documento:** Tesis de maestría

*Master in Management + Analytics*

# Cobertura en servicios de Asistencia Vial

Autoría: Moscatelli, Federico

Fecha de defensa de la tesis: 2023

## ¿Cómo citar este trabajo?

Moscatelli F. (2023) "Cobertura en servicios de Asistencia Vial". [*Tesis de maestría. Universidad Torcuato Di Tella*].

Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/12036>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 2.5 Argentina (CC BY-NC-SA 2.5 AR)

Dirección: <https://repositorio.utdt.edu>

# COBERTURA EN SERVICIOS DE ASISTENCIA VIAL

## *ROADSIDE ASSISTANCE SERVICE COVERAGE*

Autor: Federico Moscatelli  
Tutor: Javier Marengo

### **Abstract**

El trabajo presenta una solución de asistencia vial en la Ciudad de Buenos Aires ensamblando herramientas de optimización combinatoria y aprendizaje automático. El método dispone de una red de estaciones estratégicamente seleccionadas para garantizar la cobertura del territorio y luego estima la cantidad de móviles por estación en función de variables observadas durante el período de un año. El resultado muestra mejoras cercanas al 35% en la tasa de auxilios demorados, minimizando la cantidad de vehículos utilizados para la operación.

*The work presents a road assistance solution in the City of Buenos Aires by assembling combinatorial optimization and machine learning tools. The method has a network of stations strategically selected to guarantee coverage of the territory and then estimates the number of units per station based on variables observed during a one-year period. The result shows improvements close to 35% in the delay rate, minimizing the number of vehicles used for the operation.*

## Contenido

1. Introducción.....	4
1.1. El problema .....	4
1.2. La solución actual .....	6
1.3. Contenido del trabajo .....	7
2. Revisión de Literatura .....	8
3. Datos.....	11
3.1. Transformaciones.....	12
3.2. Análisis preliminar .....	14
3.2.1. Los pedidos .....	14
3.2.2. Las demoras .....	15
3.2.3. Los móviles.....	18
4. Emplazamiento de Ubicaciones .....	21
4.1. Modelo.....	21
4.2. Selección de Datos .....	21
4.3. Selección de Estaciones de Espera .....	22
4.3.1. Puntos Candidatos.....	22
4.3.2. Selección de Estaciones.....	23
4.4. Criterio de Evaluación del Modelo de Emplazamiento de Ubicaciones.....	25
5. Estimación de la Oferta .....	26
5.1. Variables Independientes.....	26
5.2. Variable Dependiente .....	26
5.2.1. Demanda de Pedidos.....	27
5.2.2. Asignación de Móviles.....	27
5.3. Conjunto de datos de entrenamiento y validación .....	28
5.4. Algoritmos de Predicción .....	28

5.4.1. Modelo de Referencia .....	29
5.4.2. Modelo de Distribución Uniforme.....	29
5.4.3. Modelo de Medias .....	30
5.4.4. Modelo de Regresión Poisson .....	30
5.4.5. Modelo de Potenciación de Gradiente.....	30
5.5. Criterio de Evaluación del Modelo de Estimación de Demanda .....	31
6. Resultados.....	33
6.1. Emplazamiento de Ubicaciones .....	33
6.2. Estimación de la Oferta .....	33
6.3. Algoritmos de Predicción .....	36
6.3.1. Modelo de Referencia .....	37
6.3.2. Modelo de Distribución Uniforme.....	37
6.3.3. Modelo de Medias .....	38
6.3.4. Modelo de Regresión Poisson .....	39
6.3.5. Modelo de Potenciación de Gradiente.....	39
6.4. Desempeño .....	40
7. Conclusiones .....	43
8. Referencias .....	46

## 1. Introducción

La popularización de los automóviles a principios del siglo XX dio lugar al surgimiento de asociaciones que, a cambio de una cuota mensual, proveen a sus socios de servicios tales como la coordinación de eventos deportivos, materiales cartográficos y, el foco de este trabajo, la asistencia vial.

La asistencia vial es un servicio proporcionado a los conductores cuyos vehículos han sufrido de una falla o problema mecánico lo suficientemente significativo como para inmovilizar temporalmente al vehículo. Al sufrir el desperfecto, el damnificado es el responsable de contactar al proveedor del servicio, quien utiliza grúas o talleres móviles para trasladar el vehículo o solucionar el problema *in situ*.

Con el paso del tiempo y el crecimiento del mercado automotor entraron nuevos jugadores a competir con estas asociaciones. Empresas especializadas y aseguradoras hoy son parte de la oferta y es común avistarlas en la vía pública.

En el caso particular referido a este trabajo, la empresa en cuestión es proveedora de servicios de asistencia vial de automóviles particulares con cobertura nacional. Estos servicios les corresponden a todos los asociados que abonan su cuota mensual y a los automóviles que tengan a su nombre.

Existen tres principales categorías de asistencia: gomería, arranque y remolque. Las primeras dos resuelven problemas de neumáticos y baterías respectivamente en el lugar del incidente, mientras que la tercera implica problemas de mayor complejidad que requieren trasladar el vehículo al lugar de conveniencia (taller, hogar, etc.).

### 1.1. El problema

La Ciudad Autónoma de Buenos Aires (CABA) y el Gran Buenos Aires (GBA) son la mayor área urbana del país. Según el Censo Nacional de Población, Hogares y Viviendas de 2010, albergan 2.890.151 y 12.801.364 habitantes respectivamente. Juntas forman el Área Metropolitana de Buenos Aires (AMBA) y representan el 39% de la población del país.

No obstante, la superficie que ocupa la metrópolis es solamente el 0,4% del total del país. Esta desproporción implica una alta densidad demográfica en la zona. Gran parte de la población (principalmente quienes se trasladan entre CABA y GBA diariamente por cuestiones laborales) optan por el automóvil particular para su traslado y se exponen a situaciones que favorecen las fallas y siniestros.

Parque Automotor en la Ciudad de Buenos Aires

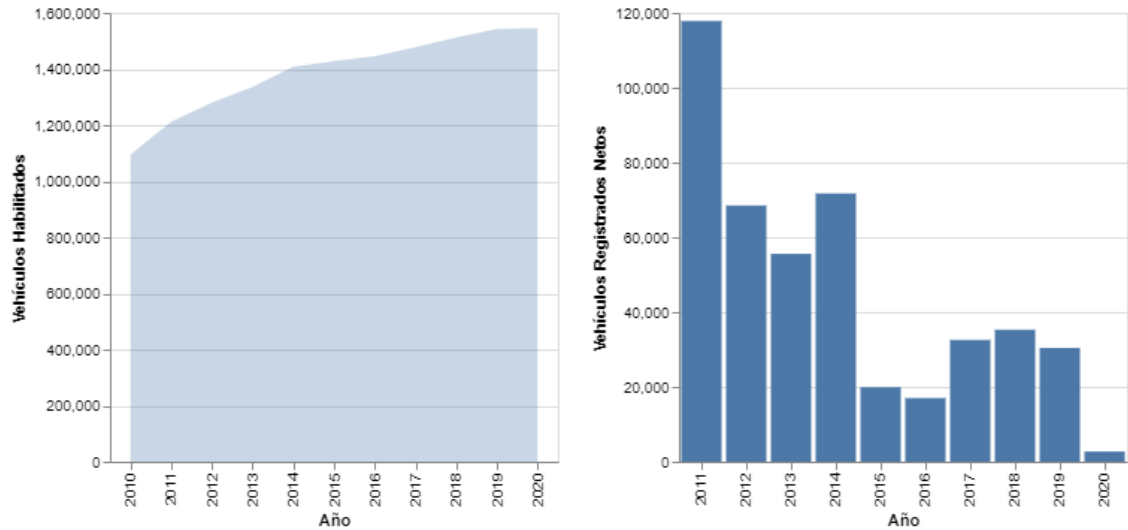


Figura 1. Fuente: Registro Nacional de la Propiedad del Automotor

La Figura 1 muestra el fuerte crecimiento de vehículos habilitados en la Ciudad de Buenos Aires durante los últimos 10 años (izquierda), y, si bien la tendencia de nuevos vehículos registrados está a la baja (derecha), hay un segundo factor que aumenta la probabilidad de asistencias: la edad promedio del parque automotor.

La Asociación de Fábricas Argentinas de Componentes (AFAC) estima que la edad media de la flota total de automóviles en 2020 es de 10,7 años y la tendencia está en aumento. Además, la composición de los nuevos patentamientos (Figura 2) muestra una fuerte presencia de los segmentos de gama baja. Ambos factores facilitan la aparición de roturas y mal funcionamiento.

Edad del Parque Automotor Argentino (2020)

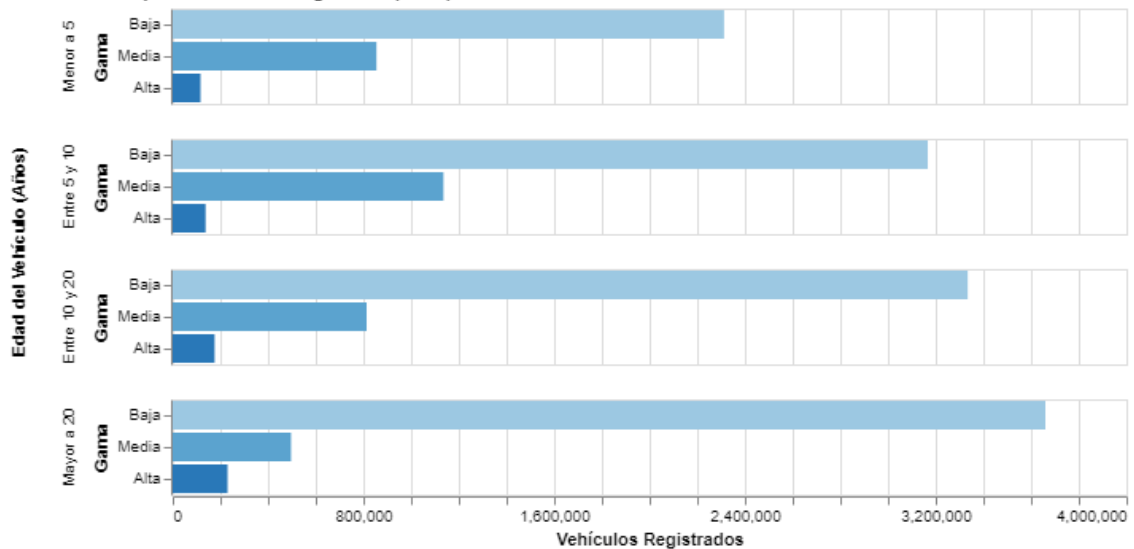


Figura 2. Fuente: Registro Nacional de la Propiedad del Automotor

La asistencia vial, o, en otras palabras, los auxilios mecánicos, son una parte importante de la propuesta de valor de la empresa de cara a sus clientes. El carácter de emergencia que suelen tener los pedidos de auxilio requiere que la experiencia del usuario deba ser lo más satisfactoria posible en esos casos. Una mala asistencia en un momento de necesidad puede generar la baja de un cliente. Además, en menor medida, la circulación de las unidades de la empresa por la vía pública funciona como refuerzo para la presencia de marca.

Con el objetivo de evitar situaciones de conflicto con sus usuarios del AMBA, la empresa determinó mediante el análisis de datos históricos que el tiempo de espera entre un pedido de auxilio y el inicio del servicio debe ser menor a 40 minutos. Las demoras que superan esta cifra suelen traer problemas futuros. Las únicas excepciones en las cuales el tiempo puede aumentar son los fines de semana turísticos. Como el movimiento de tráfico en las rutas aumenta, la demanda por servicios se dispara y requiere mayor presencia de auxilios en ellas. En consecuencia, puede haber menor cobertura en el AMBA.

En resumen, la alta exposición que generan las unidades y la importancia de una buena experiencia del usuario en los momentos de mayor necesidad hacen que la asistencia mecánica sea un pilar fundamental para la empresa.

## 1.2. La solución actual

Los móviles están distribuidos en centros de operación distribuidos en el AMBA. En los centros de operación se les hace el mantenimiento y almacenaje. Cuando están en servicio, estos son operados por un mecánico y se distribuyen dentro de la zona de influencia de su centro. Aguardan a ser llamados en sectores autoasignados basados en la experiencia propia del conductor. En todo momento son rastreados en tiempo real vía GPS y reportan el kilometraje total del vehículo mediante el odómetro.

Al momento de recibir un pedido de auxilio, el operador de la central le solicita al cliente la dirección del incidente y el tipo de asistencia que necesita. Ambos datos son cargados en el sistema para que devuelva cinco móviles candidatos. El operador, finalmente, asigna el pedido a un móvil y este debe dirigirse al lugar del incidente. Una vez que finalizan un servicio, el conductor del móvil avisa su finalización a la central y puede tener que dirigirse a un nuevo incidente o bien aguardar nuevamente en un lugar autoasignado. Si el conductor termina un servicio fuera de su jurisdicción, se intenta que vuelva a ella antes de darle un nuevo servicio.

Los mecánicos pueden formar parte de la planta permanente de la empresa o ser subcontratados en momentos de mayor demanda.

### 1.3. Contenido del trabajo

El presente trabajo busca desarrollar un modelo predictivo de geolocalización de unidades de auxilio mecánico que minimice la cantidad de servicios con demora, representado por el tiempo de traslado de estas hasta la ubicación de los incidentes. El modelo tiene en cuenta las principales restricciones propias al negocio, tales como los niveles de servicio esperado (tiempos de espera) y limitaciones sobre la cantidad de móviles activos en simultáneo.

El trabajo está dividido en dos partes. Primero, se utilizan herramientas de optimización combinatoria con los datos históricos de geolocalización de incidentes y auxilios móviles como fuente. Se determina la ubicación y cantidad de puntos de espera que minimicen la cantidad de unidades en circulación y respete la restricción de las demoras. Se circunscribe el análisis a la actividad de la empresa en la Ciudad Autónoma de Buenos Aires.

Una vez determinados los puntos de espera óptimos, el trabajo explora algoritmos que estimen la cantidad de pedidos por estación y, con esa información, asignar eficientemente la cantidad de unidades que deben colocarse en cada punto de espera.

La tesis está organizada del siguiente modo: el segundo capítulo, "Revisión de Literatura", explora el estado del arte para la resolución de problemas de optimización similares. El tercer capítulo, "Datos", describe las transformaciones que se aplican a los datos para luego hacer un análisis exploratorio de los mismos. El cuarto capítulo, llamado "Emplazamiento de Ubicaciones" explica las metodologías a utilizar para ubicar los puntos de espera, así como sus simplificaciones, limitaciones, y la forma en el que se evalúa su rendimiento. El capítulo siguiente, "Estimación de Oferta" hace lo mismo para las metodologías de aprendizaje automático utilizadas para predecir la cantidad de móviles que se deben ubicar en cada estación. Finalmente, el sexto capítulo, "Resultados", examina el rendimiento de los modelos descritos en el capítulo precedente y los compara.



## 2. Revisión de Literatura

El problema descrito en la introducción cae dentro de los llamados Problemas de Cobertura de Conjuntos de Localizaciones (LSCP, por sus siglas en inglés). Estos son una clase de problemas de optimización que surgen en muchos entornos prácticos, como la ubicación de instalaciones, la planificación de servicios de emergencia y el diseño de redes. En estos problemas, el objetivo es encontrar la ubicación óptima de un conjunto de instalaciones para cubrir un conjunto de puntos de demanda, minimizando el costo total de instalación y operación de las mismas. El LSCP ha sido ampliamente estudiado en la literatura y se han propuesto una variedad de algoritmos para resolver diferentes versiones del problema.

Uno de los enfoques más antiguos para resolver LSCP fue la heurística “*greedy*” propuesta por Dantzig y Ramser (1959), que construye una solución seleccionando iterativamente la instalación que cubre el mayor número de puntos de demanda no cubiertos. Este enfoque se ha extendido y refinado de muchas maneras, como el uso de reglas de prioridad para desempatar y la incorporación de múltiples objetivos de cobertura.

Otro enfoque popular para LSCP es el uso de programación matemática, que implica formular el problema como un modelo de programación entera y resolverlo utilizando software de optimización. Las formulaciones más comunes incluyen el modelo de cobertura de conjuntos, el modelo de partición de conjuntos y el modelo de programación entera mixta. Estos modelos se han extendido para incluir restricciones adicionales, como límites de capacidad y restricciones presupuestarias, para capturar mejor los escenarios del mundo real.

También se han aplicado extensamente metaheurísticas a LSCP, que incluyen algoritmos genéticos, búsqueda tabú y optimización por colonia de hormigas. Estos métodos generalmente comienzan con una solución inicial y la mejoran iterativamente mediante la aplicación de procedimientos de búsqueda local y técnicas de perturbación. Estos métodos se utilizan a menudo para encontrar soluciones de buena calidad rápidamente, pero no garantizan la optimalidad.

En general, los Problemas de Cobertura de Conjuntos de Localizaciones han sido ampliamente estudiados en la literatura. La elección del algoritmo depende de los requisitos específicos del problema, como el tamaño de la instancia, el número de restricciones y el nivel de optimalidad deseado.

CPLEX es un solucionador comercial que utiliza algunos de estos algoritmos para la optimización matemática. Es ampliamente utilizado en la academia e industria para resolver problemas de programación lineal, cuadrática y de programación entera mixta. A lo largo de los años, muchos

investigadores han estudiado y evaluado el rendimiento de CPLEX para resolver varios problemas de optimización.

Uno de los primeros estudios sobre CPLEX fue realizado por Bixby et al. (1992), donde compararon el rendimiento de CPLEX con otros solucionadores de optimización en un conjunto de problemas de prueba estándar. Los resultados mostraron que CPLEX superó a todos los demás solucionadores en términos de calidad y eficiencia de la solución. Desde entonces, numerosos estudios han evaluado el rendimiento de CPLEX en una amplia gama de problemas de optimización, como la optimización de la cadena de suministro (Ramanathan y Tonk, 2017), la optimización de carteras (Ghorbel et al., 2018) y problemas de ubicación de instalaciones (Ghannadpour y Dullaert, 2012).

En los últimos años, ha habido un interés creciente en desarrollar algoritmos híbridos que combinan las fortalezas de diferentes enfoques. Por ejemplo, las metaheurísticas se han combinado con formulaciones de programación matemática para obtener mejores soluciones, o con técnicas de aprendizaje automático para mejorar su rendimiento en instancias de gran escala.

Los algoritmos híbridos que combinan técnicas de optimización y aprendizaje automático se basan en el principio de que los modelos de aprendizaje automático pueden ayudar a resolver problemas complejos de optimización, ya que los modelos pueden aprender de datos históricos y, a su vez, mejorar la calidad de la solución. Los algoritmos híbridos que utilizan aprendizaje automático se han utilizado en problemas de LSCP para identificar patrones en los datos históricos de demanda y selección de ubicaciones, y luego utilizar estos patrones para mejorar la selección de ubicaciones futuras.

Por ejemplo, se puede utilizar una red neuronal para aprender a clasificar los puntos de demanda y, a continuación, utilizar esta información para seleccionar las ubicaciones de las instalaciones de forma más eficiente. Del mismo modo, se pueden utilizar técnicas de agrupamiento basadas en aprendizaje automático para identificar grupos de puntos de demanda similares y, a continuación, seleccionar una ubicación de instalación que cubra la mayoría de los puntos de demanda en cada grupo.

El uso de algoritmos de aprendizaje automático en modelado estadístico se ha vuelto cada vez más popular en los últimos años debido a su capacidad para manejar datos complejos y de alta dimensión. Dos de estos algoritmos, la regresión de Poisson y los modelos de potenciación de gradiente, se han utilizado ampliamente en diversos campos, incluidos la atención médica, las finanzas y el marketing, para predecir y modelar resultados de interés.

La regresión de Poisson es una técnica estadística utilizada para modelar datos de conteo, donde la variable dependiente representa el número de veces que ocurre un evento en un período dado. Esta técnica supone que los conteos son independientes e idénticamente distribuidos, y que la media y la varianza de los conteos son iguales. Los modelos de regresión de Poisson se han utilizado en diversas aplicaciones, como la predicción de readmisiones en hospitales, el análisis de compras de clientes y la predicción de accidentes de tráfico (Liu, 2018; Lobo et al., 2019; Sun et al., 2019).

Los modelos de potenciación, por otro lado, son un tipo de algoritmo de aprendizaje de conjuntos que ensamblan múltiples aprendices débiles para crear un predictor fuerte. Estos aprendices débiles suelen ser algoritmos simples tales como árboles de decisión poco profundos) que se entrenan para minimizar una función de pérdida específica (error cuadrático medio para problemas de regresión). Los algoritmos de refuerzo ajustan iterativamente un conjunto de modelos débiles a los datos y dan más peso a las observaciones clasificadas incorrectamente en iteraciones posteriores. Esto resulta en un modelo final que tiene un sesgo más bajo y un mejor rendimiento predictivo que cualquiera de los modelos individuales. Los algoritmos de refuerzo se han utilizado para diversas tareas de predicción, como la puntuación de riesgo crediticio, la predicción de la rotación de clientes y la predicción de precios de las acciones (Friedman, 2001; Kim y Ahn, 2019; Wan et al., 2021).

Otro estudio utilizó una combinación de regresión de Poisson y modelos de potenciación para predecir los accidentes de tráfico en una ciudad. El modelo se entrenó con datos históricos y fue capaz de predecir con exactitud el número de accidentes en un día determinado. El modelo también identificó los factores más importantes que contribuyen a los accidentes, como las condiciones meteorológicas, la superficie de la carretera y el volumen de tráfico (Liang et al., 2020).

En resumen, la revisión de la literatura indica que el problema a resolver y las herramientas que se utilizan para resolverlo tiene precedentes exitosos.

### 3. Datos

Los datos provienen del sistema interno de control de operaciones de la empresa en cuestión. La empresa compartió datos de un período de 12 meses con el objetivo de capturar efectos estacionales en la demanda de auxilios, siendo conscientes que los efectos estacionales del año seleccionado persisten. No fue posible obtener datos de años anteriores o posteriores para resolver esta limitante. Además, se evitó utilizar datos correspondientes al período de la pandemia del COVID-19 ya que las medidas de restricción a la movilidad afectaron fuertemente la cantidad de vehículos en circulación. El objetivo será, entonces, utilizar los datos obtenidos entre abril 2019 y marzo 2020 para estimar la demanda en períodos de actividad normales.

El archivo recibido incluye 340.876 observaciones (pedidos de auxilio) con las siguientes variables:

- *Número de incidente*. Variable numérica discreta. Identificador único de las observaciones.
- *Fecha del incidente*. Variable categórica ordinal de fecha. Formato: DD/MM/AAAA.
- *Hora del incidente*. Variable categórica ordinal de hora. Formato: HH/MM/SS.
- *Tipo de servicio*. Variable categórica con 3 posibles valores: “remolque”, “gomería” y “arranque”.
- *Número de equipo*. Variable numérica discreta. Identificador único de cada móvil.
- *Ubicación del equipo al momento del pedido*. Variables numéricas continuas expresadas en coordenadas (latitud y longitud).
- *Ubicación del equipo al momento de la asistencia mecánica*. Variables numéricas continuas expresadas en coordenadas (latitud y longitud).
- *Odómetro del equipo al momento del pedido*. Variable numérica continua expresada en kilómetros.
- *Odómetro del equipo al momento de la asistencia mecánica*. Variable numérica continua expresada en kilómetros.
- *Indicador de día feriado*. Variable booleana. “SI” implica que el incidente se originó en un feriado; “NO” en caso contrario.
- *Temperatura mínima de la fecha del incidente*. Variable numérica medida en grados Celsius.
- *Temperatura máxima de la fecha del incidente*. Variable numérica medida en grados Celsius.
- *Categoría del cliente*. Variable categórica que indica el tipo de cliente que hace el pedido de auxilio.

Nota: variables redundantes o irrelevantes, tales como códigos de servicio o número identificador del solicitante fueron descartadas del análisis.

Además, a través del sitio [meteostat.net](http://meteostat.net) se obtuvieron datos diarios climáticos para el mismo período. Estos datos son, a priori, relevantes por dos razones. Primero, el clima puede deteriorar las condiciones de manejo, tanto en la visibilidad como en la capacidad de maniobra de los conductores. Segundo, puede influir en la cantidad de vehículos en circulación. Los días con climas adversos pueden empujar a los usuarios a utilizar sus vehículos particulares en favor de otros medios de transporte. Los datos obtenidos son:

- Precipitaciones diarias. Variable numérica medida en milímetros.
- Velocidad promedio del viento. Variable numérica medida en kilómetros por hora.

Por último, se consiguieron datos geospaciales de los polígonos que comprenden CABA y sus respectivos barrios y comunas con el fin de analizar tendencias de pedidos y de demoras en los mismos.

### 3.1. Transformaciones

Si bien los datos crudos ya presentan una interesante cantidad de variables, se procedió a limpiar observaciones, transformar variables y crear nuevos atributos para mejorar la calidad del conjunto de datos. Luego de remover aquellas observaciones con datos faltantes, se procedió a extraer y categorizar las variables que componen la fecha y hora del pedido; corregir formatos varios; combinar los datos climáticos al nivel del día; calcular distancias, tiempos y velocidades de viaje; asignar barrios y comunas a partir de las coordenadas; y remover atributos redundantes.

Por último, los pedidos de auxilios que forman parte del conjunto de datos son aquellos en los cuales la ubicación del móvil al momento de recibir el pedido o realizar el pedido están dentro de CABA. Son excluidos todos los pedidos que ocurren 100% fuera de la zona de cobertura.

Como resultado, el nuevo conjunto de datos contiene 163.710 observaciones con 31 atributos:

- *Tipo de solicitante*. Variable categórica con dos posibles valores: “socio” y “otros”.
- *Tipo de servicio*. Variable categórica con tres posibles valores: “remolque”, “gomería” y “arranque”.
- *Número de equipo*. Variable numérica discreta con 288 posibles valores. Identificador único de cada móvil.
- *Indicador de día feriado*. Variable booleana. “True” implica que el incidente se originó en un feriado; “False” en caso contrario.

- *Distancia del viaje.* Variable numérica continua expresada en kilómetros a partir del odómetro de los móviles.
- *Distancia directa del viaje.* Variable numérica continua calculada como la distancia mínima (recta) entre el punto de comunicación (cuando recibe la orden del pedido) y de llegada al móvil. Expresada en kilómetros.
- *Sinuosidad del viaje.* Variable numérica continua. Calculada como el ratio entre la distancia del viaje y la distancia directa.
- *Mes del año.* Variable categórica ordinal.
- *Día del mes.* Variable categórica ordinal.
- *Día de la semana.* Variable categórica ordinal.
- *Hora del día.* Variable categórica ordinal.
- *Temperatura media del día del pedido.* Variable numérica discreta.
- *Temperatura máxima del día del pedido.* Variable numérica discreta.
- *Temperatura mínima del día del pedido.* Variable numérica discreta.
- *Amplitud térmica del día del pedido.* Variable numérica discreta, calculada como la diferencia entre la temperatura máxima y mínima del día.
- *Precipitaciones del día del pedido.* Variable numérica continua expresada en milímetros.
- *Indicador de día de lluvia.* Variable booleana. “True” implica que el incidente se originó en un día que tuvo lluvias en algún momento; “False” en caso contrario.
- *Presión atmosférica media del día del pedido.* Variable numérica continua.
- *Tiempo de espera del cliente.* Variable numérica continua. Indica los minutos que el cliente esperó desde que la central recibió el pedido.
- *Tiempo de asignación del pedido.* Variable numérica continua. Minutos que se tomó la central en asignar el pedido.
- *Tiempo de viaje del móvil.* Variable numérica continua. Minutos que se tomó el móvil en llegar al cliente luego de haber recibido la orden del pedido.
- *Velocidad media del móvil.* Variable numérica continua expresada en kilómetros por hora.
- *Indicador de demora.* Variable booleana. “True” implica que el tiempo de espera del cliente fue mayor a 40 minutos; “False” en caso contrario.
- *Tiempo de demora.* Variable numérica continua. Indica los minutos que el cliente esperó por sobre los 40 minutos de tolerancia.
- *Barrio en el que se encuentra el móvil al recibir el pedido.* Variable categórica con 48 posibles valores.

- *Comuna en la que se encuentra el móvil al recibir el pedido.* Variable categórica con 15 posibles valores.
- *Barrio en el que se encuentra el móvil al iniciar el pedido.* Variable categórica con 48 posibles valores.
- *Comuna en la que se encuentra el móvil al iniciar el servicio.* Variable categórica con 15 posibles valores.
- *Área del barrio en el que se encuentra el móvil al recibir el pedido.* Variable numérica continua expresada en metros cuadrados.
- *Área del barrio en el que se encuentra el móvil al iniciar el servicio.* Variable numérica continua expresada en metros cuadrados.

### 3.2. Análisis preliminar

El conjunto de datos enriquecido permite hacer un análisis acerca de la actividad de la empresa durante el período con el objetivo de entender la situación con mayor profundidad.

#### 3.2.1. Los pedidos

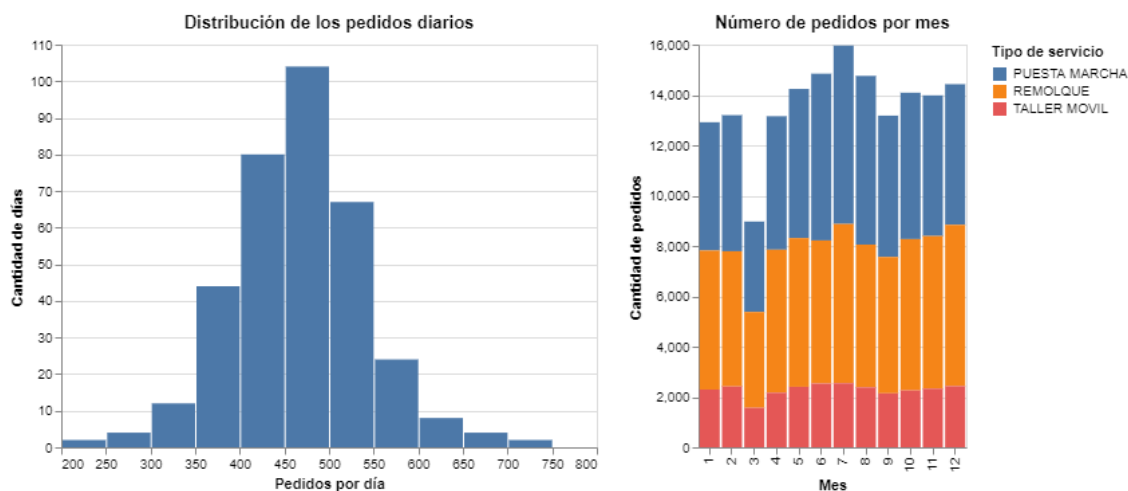


Figura 3. Distribución de los pedidos.

La Figura 3 (izquierda) revela la distribución de los pedidos diarios en el año. La Figura 3 (derecha) muestra la distribución de pedidos a lo largo del año. Es evidente que hay un mayor volumen de pedidos para los meses invernales. Además, para el mes de marzo puede verse el efecto de las medidas de restricción a la circulación impuestas a partir del 20 de ese mes. Este shock a los pedidos es definitivamente disruptivo. Por eso es necesario evitar datos de los meses posteriores. Los días entre el 20 de marzo y el 31 de marzo serán incluidos de todas formas para tener así un conjunto de datos que comprenda un año completo. El efecto de estos 12 días (3% de la muestra) puede afectar las estimaciones. Sin embargo, se trabajará incluyéndolos para que el modelo sea robusto a otros posibles eventos disruptivos. Contrariamente, la distribución

entre los tipos de servicio (puesta en marcha, remolque y taller móvil) para cada mes no parece variar de manera llamativa a lo largo del año.

### 3.2.2. Las demoras

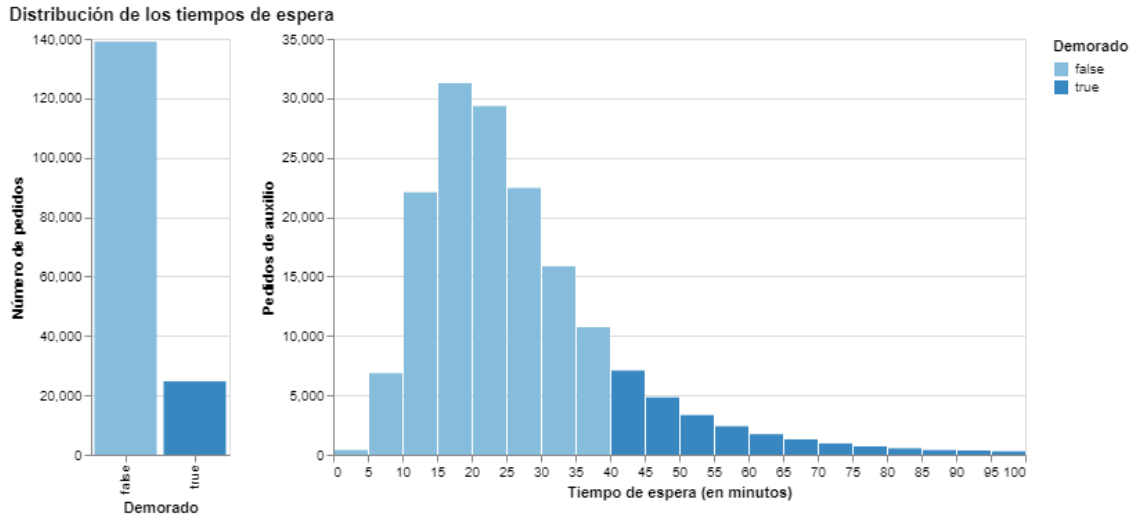


Figura 4. Distribución de los tiempos de espera.

La Figura 4 muestra las distribuciones de pedidos demorados, es decir, de pedidos en los que el auxilio tardó más de 40 minutos en llegar (izquierda) y del tiempo de espera (derecha). Para los primeros, se calcularon 24.701 pedidos demorados y 139.009 a tiempo, lo cual implica una tasa de demora (pedidos demorados / total de pedidos) cercana al 15% del total de auxilios. Al circunscribir la ubicación únicamente a la zona de CABA, la tasa de demora que se utiliza como referencia se reduce del 15% al 12,19%. La distribución de los tiempos de espera muestra crecimiento hasta el segmento de 15-20 minutos y luego decrece permanentemente hasta el segmento más alto y menos popular.



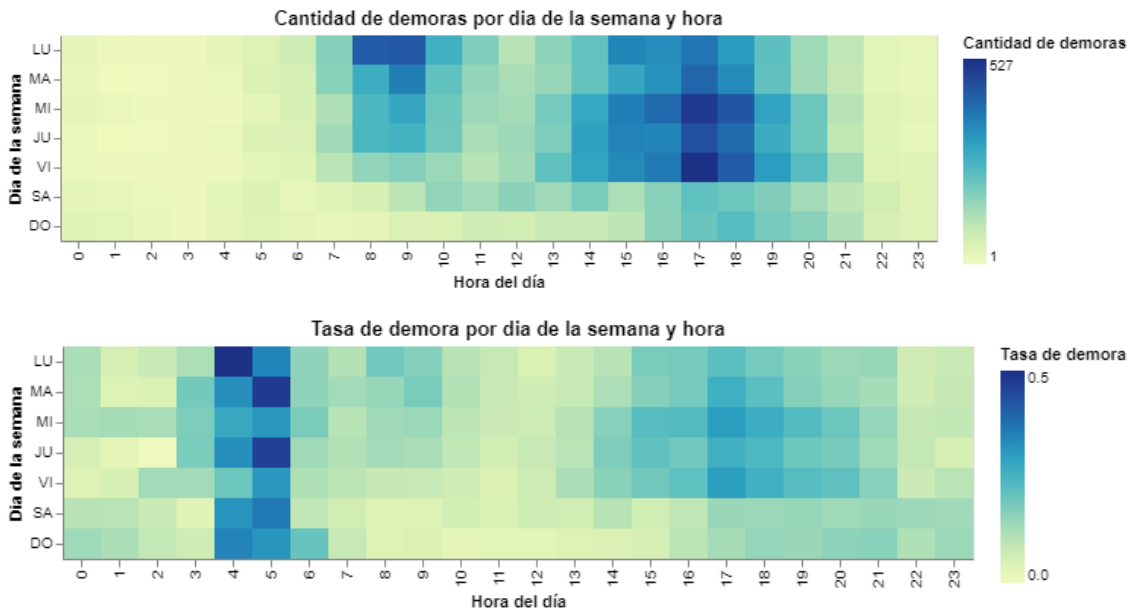


Figura 5. Concentración de pedidos y demoras en la semana.

La Figura 5 ilustra la distribución de los pedidos demorados y la tasa de demora en función del día de la semana y la hora del día. En la primera puede verse la concentración de los pedidos durante los días laborales, especialmente en las franjas 08:00 - 10:00 (siendo los lunes el día más fuerte) y, en mayor medida, 14:00 – 18:00. La tasa de demora contrasta con la anterior, pues concentra sus valores más altos en la madrugada de todos los días de la semana.

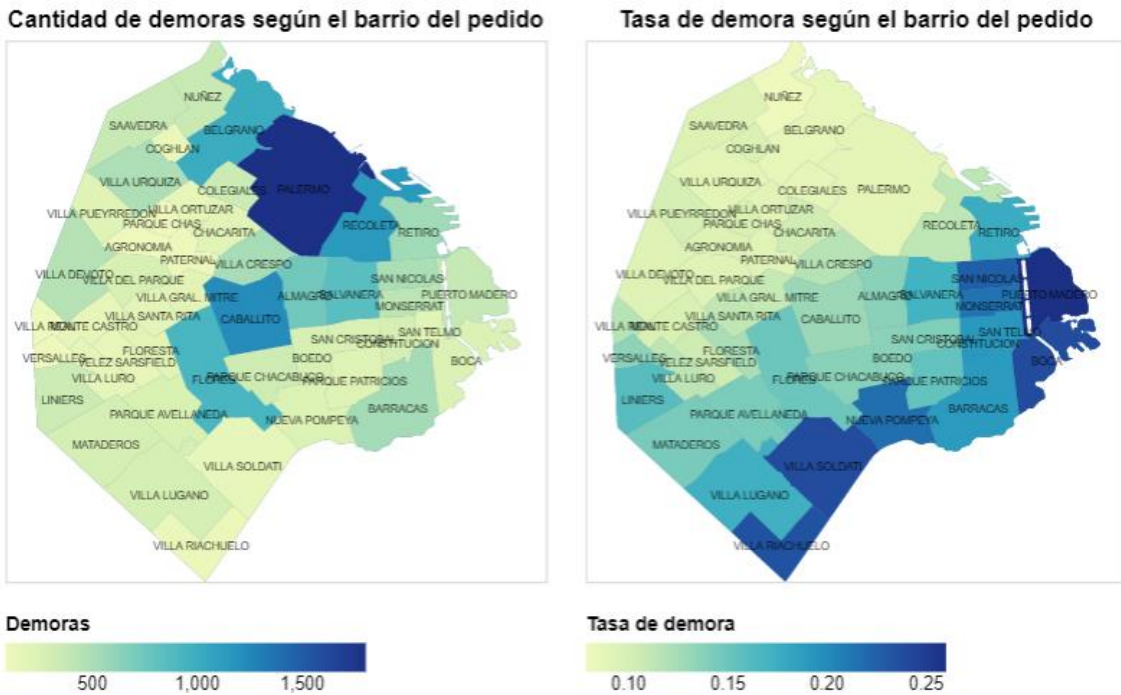


Figura 6. Concentración de demoras por barrio del pedido.

La Figura 6 agrega la perspectiva geoespacial a la distribución de las demoras según el barrio en que se encuentra el cliente. Por un lado, la cantidad de demoras parece concentrarse en los

barrios del norte y centro de la ciudad, no obstante, la tasa de demora muestra una gran diferencia entre el norte y sur, siendo este último el más perjudicado.

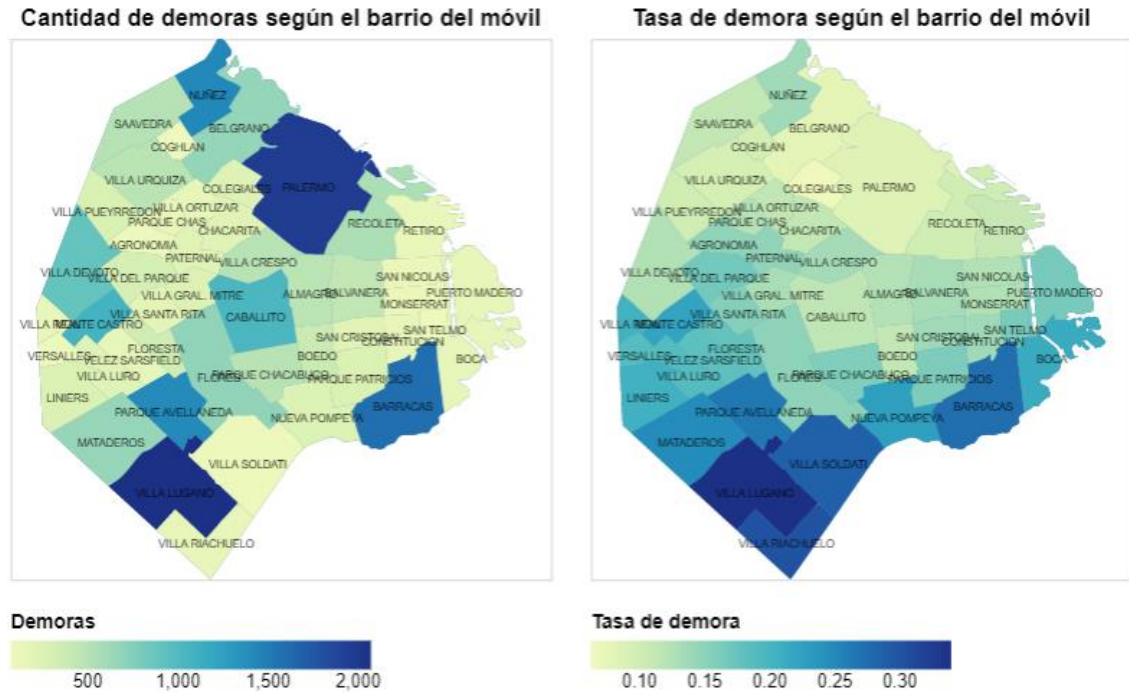


Figura 7. Concentración de demoras por barrio en que se encuentra el móvil.

Vale aclarar que los móviles se suelen trasladar de un barrio a otro para llegar a los clientes. La Figura 7 muestra cómo a pesar de esto la tasa de demora no muestra grandes cambios. La cantidad de demoras se muestra más polarizada en comparación con la Figura 5 ya que se está capturando la ubicación de espera de los móviles que están aguardando un pedido en sus lugares auto asignados.

Impacto de la lluvia y la temperatura sobre los tiempos de espera

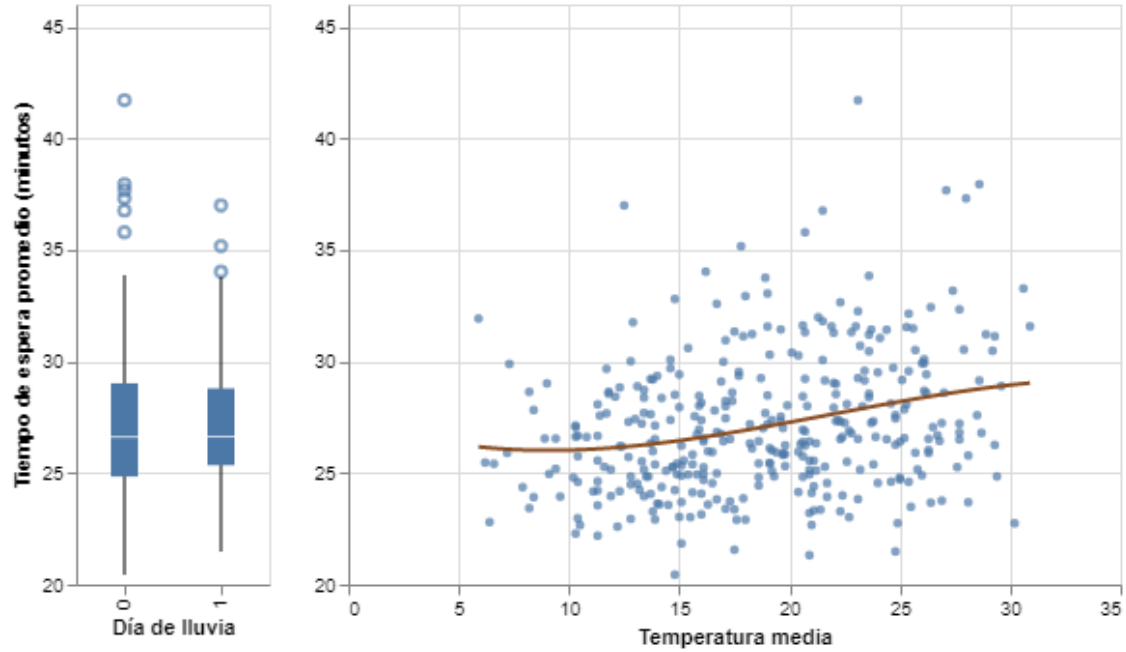


Figura 8. Impacto climático.

En cuanto a la influencia del clima sobre las demoras y tiempos de espera, a través de la Figura 8 puede notarse que la lluvia (izquierda) no parece tener relación con el desempeño de los móviles. Para la temperatura media del día del pedido (derecha) puede verse una leve correlación positiva.

3.2.3. Los móviles

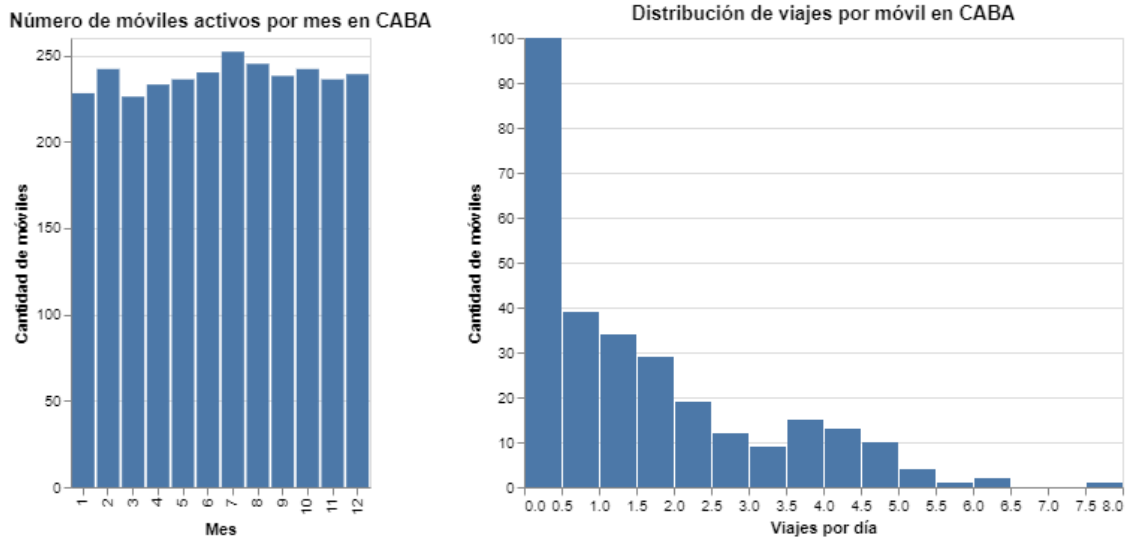


Figura 9. Móviles activos en CABA.

La Figura 9 (izquierda) da luz sobre la cantidad de móviles activos por mes en CABA, que oscilan entre 226 y 252. No obstante, de los 288 móviles únicos que se registran en el conjunto de datos a lo largo del año, el histograma (derecha) muestra que 139 hicieron menos de 1 viaje por día.

Este último dato tiene 2 motivos. La Figura 10 muestra como a nivel AMBA la cantidad de móviles en circulación es prácticamente la misma que en CABA, por lo que prácticamente todos los móviles acceden a CABA ocasionalmente.

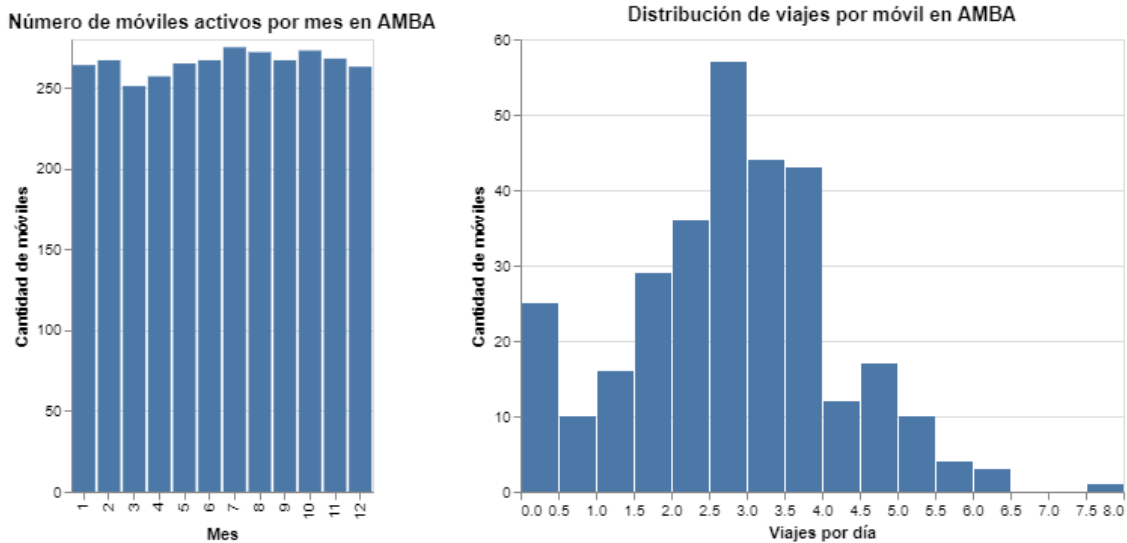


Figura 10. Móviles activos en AMBA.

El segundo motivo es que los móviles de mayor antigüedad sólo se utilizan en momentos especiales. Estos pueden verse dentro del grupo de 35 móviles que realizan menos de 1 viaje diario a nivel AMBA.

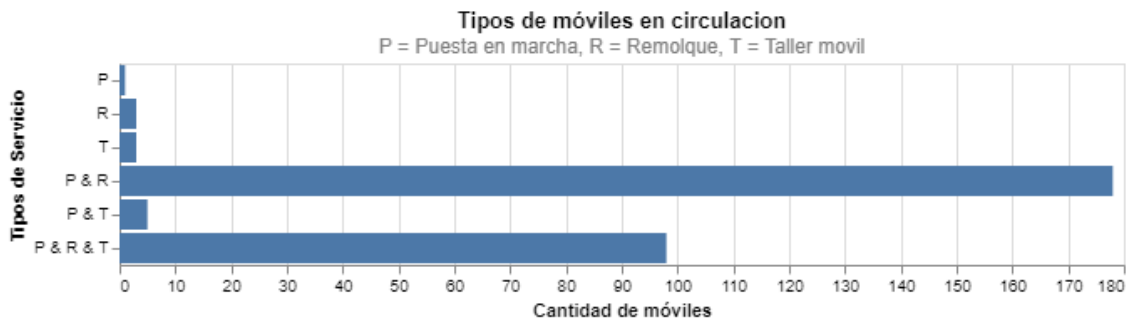


Figura 11. Tipos de móviles.

Hasta ahora, se presentó a las unidades como idénticas e intercambiables. La Figura 11 cuenta la cantidad de móviles en función de los servicios que prestaron a lo largo del año. Si bien no se dispone de datos directos acerca de las características de cada unidad, a partir esta información, el trabajo infiere que existen dos clases de vehículos. Aquellos que realizan puestas en marcha y remolques, y aquellos que realizan los tres servicios. Las unidades del primer grupo son llamadas grúas o remolques, y las del segundo, talleres móviles.

El móvil 2224, por ejemplo, es un taller móvil cuyo centro de operación está ubicado en la zona norte de CABA. La Figura 12 muestra un recorrido tipo para un día seleccionado al azar. Los viajes se representan mediante las rectas que conectan los puntos de salida (marcado con un

círculo) al recibir el pedido y la ubicación de los clientes (marcado con la hora de llegada). Gracias a esto puede reconstruirse el orden del recorrido del móvil:

1. El móvil comenzó el día en el centro de la ciudad durante la medianoche. Casi seguramente implica que estaba haciendo su último viaje del día anterior, pues no vuelve a hacer viajes hasta las 6 de la mañana.
2. Partiendo desde el centro de operaciones, hace un recorrido matutino en Palermo y Recoleta que finaliza a las 14 horas. Es a esa hora donde finalmente puede volver a la base.
3. Ya por la tarde, realiza un auxilio a las 17 horas, vuelve a la base y, finalmente, hace sus últimos 2 servicios a las 20 (demorado) y 21 horas.

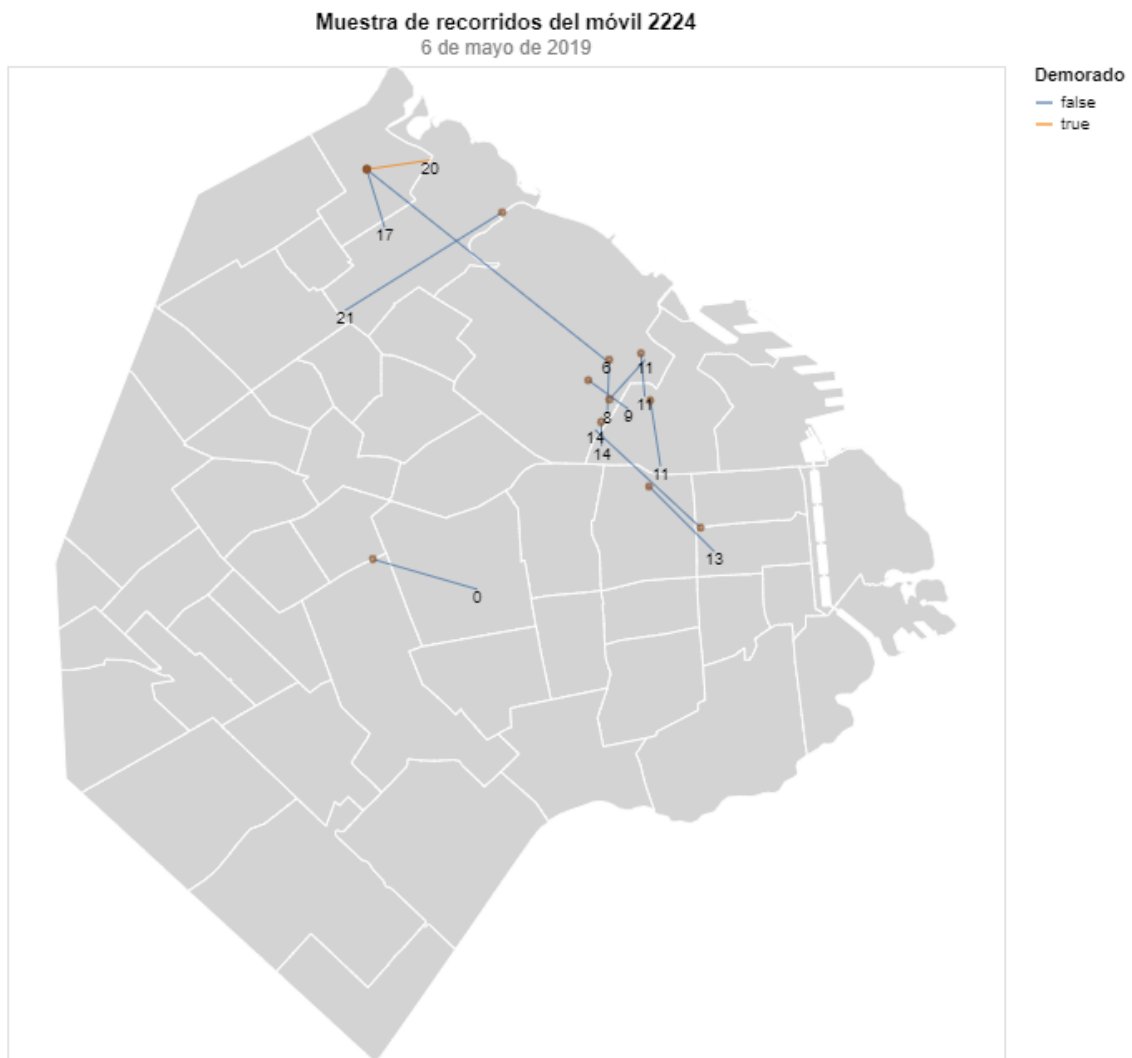


Figura 12. Ejemplo de recorrido de un móvil en 24 horas.

En suma, el análisis inicial de los datos demuestra que existen patrones interesantes y, potencialmente, lugar para mejoras. A continuación, se describirán los métodos y modelos a utilizar para optimizar el servicio.

## 4. Emplazamiento de Ubicaciones

La primera parte del trabajo busca solucionar el problema de emplazamiento de ubicaciones, es decir, determinar las mejores ubicaciones para que los móviles esperen los pedidos. El algoritmo de optimización descrito a continuación tiene como objetivo definir la ubicación geoespacial de las estaciones de espera de las unidades.

### 4.1. Modelo

En el presente trabajo se utiliza un modelo de emplazamiento de ubicaciones que minimiza la cantidad de puntos de espera respetando requisitos de cobertura. Este requisito se expresa mediante una distancia máxima que un punto puede tener de cualquier pedido que se encuentre en la zona para que las unidades lleguen al punto de asistencia en el menor tiempo posible. Estos puntos de espera pueden tener unidades asignadas o quedar vacíos en función de la demanda siempre y cuando se respeten las restricciones de inventario de las unidades. Se minimizará la cantidad de demoras.

### 4.2. Selección de Datos

Los pedidos que el modelo utiliza son aquellos cuyos clientes se ubican dentro de la Ciudad de Buenos Aires, incluso los clientes que deban ser remolcados fuera de ella. Se descartan, entonces:

- Grúas y Talleres que se pidan desde fuera de la ciudad.
- Remolques que se pidan desde fuera de la ciudad, independientemente de que su destino sea o no dentro de la ciudad.

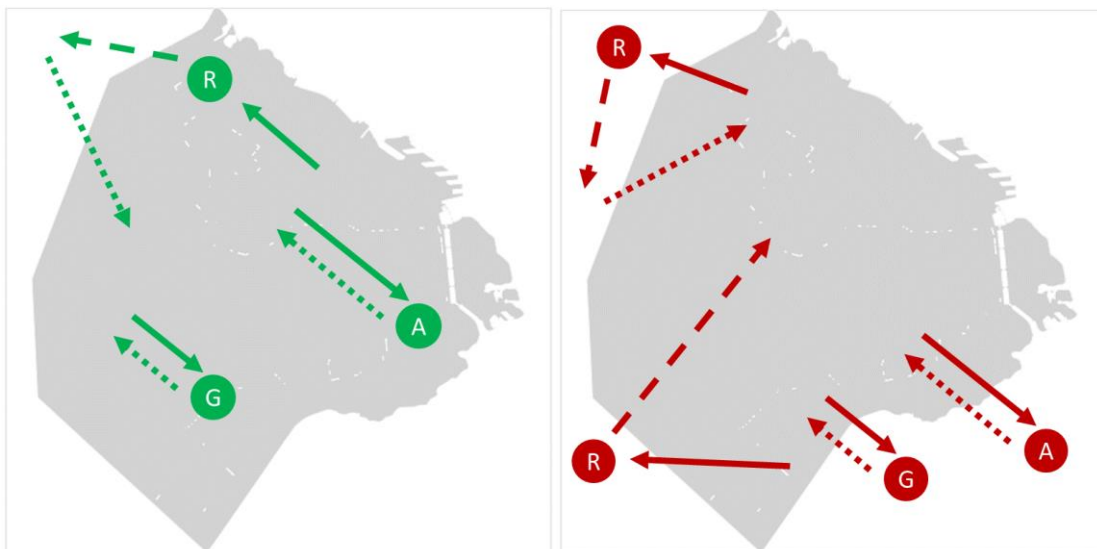


Figura 13. Tipos de pedidos incluidos (izquierda) y excluidos (derecha).

Como ilustra la Figura 13, sólo los tipos de pedidos en verde son aquellos que se tendrán en cuenta para el análisis. Los mismos son todos aquellos pedidos de taller (A) y grúa (G) que ocurran en la ciudad, y los remolques (R) que sean pedidos dentro de la ciudad, pero que puedan llevar a los clientes incluso a lugares fuera de ella. En resumen, el cliente debe estar en la Ciudad de Buenos Aires al hacer el pedido. Todos aquellos pedidos que sean desde fuera de la ciudad (en rojo) son excluidos pues deberían ser parte de estrategias de otras localidades.

### 4.3. Selección de Estaciones de Espera

La selección de la cantidad y ubicación de las estaciones de espera se determina a partir de la resolución del problema de cobertura del conjunto de ubicaciones (*Location Set Covering Problem*) descrito por Toregas (1971). En el mismo, se busca minimizar la cantidad de ubicaciones que puedan cubrir la totalidad de la demanda respetando restricciones de tiempo y/o distancia.

#### 4.3.1. Puntos Candidatos

Como las ubicaciones en la que los móviles pueden esperar no tienen restricciones en la práctica, se utilizan más de 250 puntos aleatorios (ver Figura 14) en el mapa de la ciudad como candidatos a estaciones de espera. Si bien la ciudad tiene sectores en los que no se puede estacionar, ya sea por prohibición o por factibilidad (un lago, un parque, etc.), en la práctica, esto no es un problema. La máxima distancia que existe entre cualquier punto de la ciudad y una calle es de 500 metros. Mientras los puntos tengan un radio de cobertura mayor a 500 metros, el vehículo puede esperar dentro este radio en el lugar más cercano posible al punto e igual llegar a atender el pedido a tiempo.

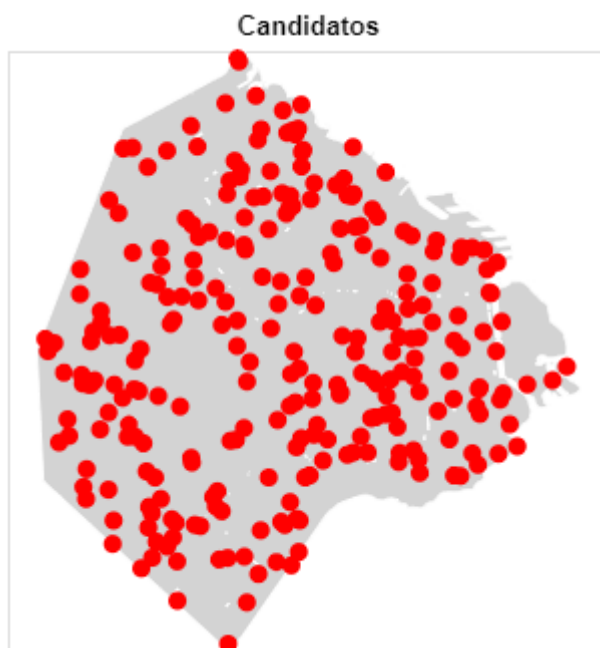


Figura 14. Ejemplo de 268 estaciones candidatas.

El radio de cobertura de estos candidatos es de 1 kilómetro. Este radio se elige ya que la menor distancia promedio que un móvil logra cubrir en 40 minutos en el peor de los escenarios posibles es de 2 kilómetros. Entonces, un radio de 1 kilómetro de cobertura asegura que la distancia máxima de un móvil que se encuentre en el área de cobertura sea de 2 radios. Este escenario se logra combinando el barrio en el cual los móviles van más lento y la hora en el que van más lento, combinado con la sinuosidad más alta de cualquier barrio. Se asume entonces que cualquier unidad que se encuentre en este radio puede atender al pedido de su misma estación sin demorarse.

#### 4.3.2. Selección de Estaciones

Utilizar los 250 puntos de espera candidatos obtenidos en la sección anterior como estaciones de espera sería una estrategia subóptima. Si bien con ellos se logra una cobertura del 100% del territorio, la cantidad y el carácter aleatorio de estos puntos produce como consecuencia una configuración redundante. Hay puntos que podrían ser eliminados para simplificar la asignación de los móviles.

Para seleccionar la menor cantidad de puntos candidatos que puedan atender a todos los pedidos se utilizan herramientas de optimización combinatoria. El paquete de optimización utilizado (solver) para quedarse con la menor cantidad de ubicaciones es CPLEX. El mismo utiliza como input la ubicación de los pedidos, la ubicación de los puntos candidatos y la distancia máxima de 1 kilómetro de radio. Además, se le indica que calcule la distancia euclídea entre estaciones y pedidos.



Formalmente, adaptando la formulación de Church y Murray (2018) el paquete recibe el siguiente modelo:

$$\begin{aligned} & \text{minimizar } \sum_{j \in J} Y_j \\ & \text{sujeto a: } (1) \sum_{j \in N_i} Y_j \geq 1 \quad \forall i \in I, \\ & \quad (2) Y_j \in \{0,1\} \quad \forall j \in J \end{aligned}$$

donde:

- $I =$  conjunto de pedidos
- $J =$  conjunto de estaciones candidatas
- $S =$  distancia máxima aceptable (2 km)
- $d_{ij} =$  distancia mínima (euclídea) entre el punto  $i$  y la candidata  $j$
- $N_i = \{j \in J \mid d_{ij} < S\} \forall i \in I$
- $Y_j = \begin{cases} 1 & \text{si se usa la estación candidata } j \\ 0 & \text{en caso contrario} \end{cases} \quad \forall j \in J$

Una vez ejecutado el solver con este modelo, el mismo devuelve un conjunto de estaciones a abrir. La cantidad de valores únicos para las estaciones asignadas a los pedidos representa el total de estaciones que el algoritmo utilizó para resolver el problema de optimización.

Una debilidad de esta metodología es que depende de las estaciones candidatas. Puede suceder que, si bien la solución obtenida sea óptima, existan otros puntos candidatos que permitan soluciones que usen menos estaciones. Para resolver esto, se repite el proceso que genera los puntos aleatorios en reiteradas ocasiones y se vuelve a optimizar la cantidad de estaciones cada vez. Esto genera un conjunto de soluciones posibles que utiliza distintas estaciones, y, en el mejor de los casos, encuentra soluciones que usen menos estaciones. La solución que utilice menos estaciones es la seleccionada y la que se utiliza como input para las siguientes etapas.

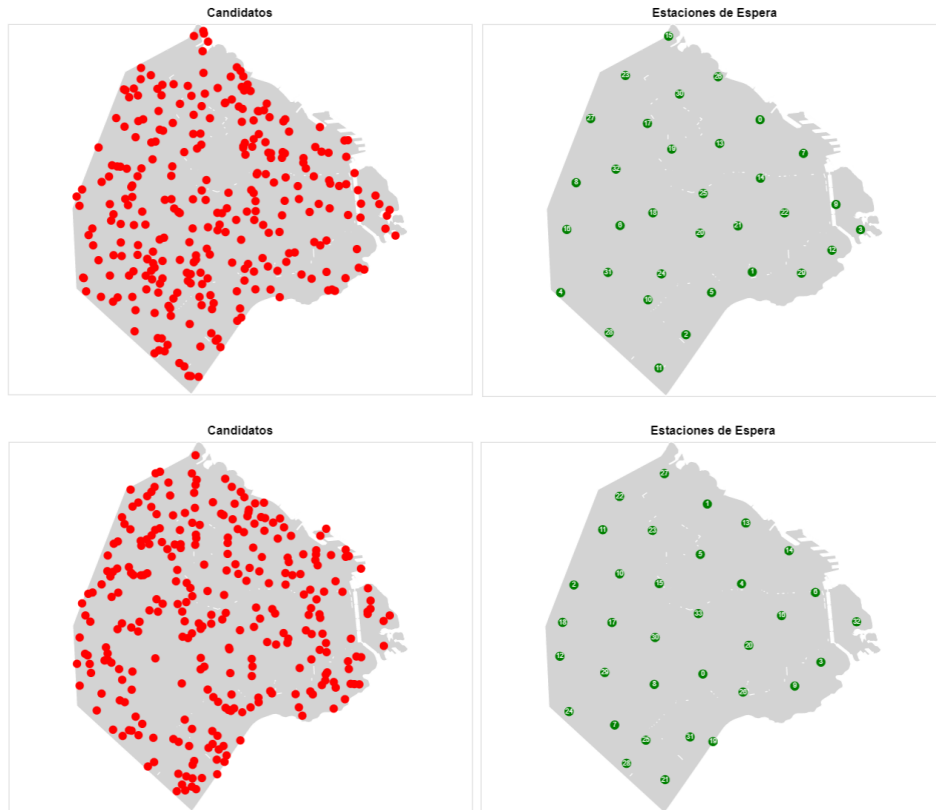


Figura 15. Ejemplos de estaciones seleccionadas (en verde) por el algoritmo de optimización usando distintos puntos candidatos (en rojo).

La figura 15 ilustra este proceso iterativo para encontrar la menor cantidad de estaciones (derecha) posibles partiendo de distintas configuraciones de candidatos (izquierda). En general, la ubicación de las estaciones y la cantidad que se eligen se mantiene estable entre 34 y 35.

#### 4.4. Criterio de Evaluación del Modelo de Emplazamiento de Ubicaciones

La evaluación del modelo, al igual que el modelo en sí, es bipartita. Para esta primera parte se evalúa el rendimiento del algoritmo de optimización que determina la ubicación de las estaciones. Para ello, se toman todas las configuraciones restantes y se selecciona aquella que menor cantidad de estaciones necesite. Una vez seleccionada, el trabajo asume estas estaciones como dadas y procede a estimar la demanda para cada una.

## 5. Estimación de la Oferta

La segunda parte del trabajo busca resolver el problema de la estimación de la cantidad de móviles ofrecidos para cada estación. El modelo de esta parte estima la cantidad de móviles que deben ubicarse por estación utilizando algoritmos de aprendizaje automático. Las variables que es posible agrupar al nivel de la estación son aquellas que estos algoritmos utilizan para generar sus estimaciones. Este capítulo comienza describiendo estas variables. Segundo, describe la variable a estimar. A continuación, describe los distintos algoritmos utilizados y, por último, la forma en la que se evalúan estos últimos.

### 5.1. Variables Independientes

Habiendo definido la ubicación de las estaciones de espera de los móviles, se divide la jornada por turnos de 4 horas. Para cada turno, se asume que las unidades pueden atender un solo pedido. De esta forma, las unidades podrán ser reubicadas a distintas estaciones a lo largo del día. Al mismo tiempo, las 4 horas dan tiempo a las unidades de remolque de hacer un viaje dentro del turno y estar listas para el siguiente pedido, incluso si debieran trasladarse a otra estación.. Finalmente, el modelo simplifica los tipos de servicios en función de las clases de vehículos que los pueden proveer. Las puestas en marcha y remolques son unificadas y atendidas exclusivamente por las unidades remolque, y los pedidos de taller, únicamente por los talleres móviles. En consecuencia, en base a los datos analizados en la Sección 3.2.3., la cantidad máxima de móviles disponibles para el servicio de CABA es 178 remolques y 98 talleres. El modelo entonces asignará los móviles para cada combinación de las siguientes cuatro variables categóricas:

- Estación (0, 1, 2, 3...);
- Turno (0 a 4, 4 a 8, 8 a 12, 12 a 16, 16 a 20, 20 a 24);
- Día de la semana (0, 1, 2, 3, 4, 5, 6);
- Tipo de servicio (Remolque, Taller).

### 5.2. Variable Dependiente

El modelo en cuestión busca predecir la cantidad de móviles que deben ubicarse en cada estación, turno día y tipo de servicio. Para hacer esto, primero se debe agrupar los pedidos en función de las variables independientes y así tener una demanda de pedidos. Segundo, se debe transformar esta demanda para que los algoritmos que definan la cantidad de móviles por estación, turno, tipo y día busquen soluciones que minimicen la cantidad de demoras.

Cabe enfatizar que el problema no busca estimar la demanda de pedidos, sino que usa los datos históricos de pedidos (es decir, la demanda) para estimar cuántos móviles ubicar para que la cantidad de demoras sea la menor posible.

A continuación, se describen los dos pasos que construyen la variable dependiente.

#### 5.2.1. Demanda de Pedidos

La demanda de pedidos por estación es el resultado de sumar la cantidad de pedidos, agrupados por estación, tipo, día de la semana y turno. Primero se asoció cada pedido a la estación más cercana utilizando las latitudes y longitudes de pedidos y estaciones. Segundo, se agruparon los mismos por tipo de servicio, día de la semana, estación, turno (las variables independientes) y por el día del mes y el mes. Estas últimas dos dimensiones no son parte de las variables independientes del modelo, pero suman observaciones para cada combinación. Por ejemplo, la combinación estación 5, en el turno 8 a 12, día de la semana 4, tipo de servicio remolque, se observó 52 veces (una vez por semana) en la muestra y tuvo en promedio 3 pedidos.

Este segundo paso permite entrenar este modelo en función de estas cuatro variables. Una consecuencia de ello es que requiere tener información de todas las posibles combinaciones, incluso para aquellas combinaciones que no tengan pedidos.

La cantidad de observaciones de 163.710 pedidos únicos se ve alterada. El cambio expande a alrededor de 4 millones de combinaciones de pedidos por estación. Si bien, en principio, agrupar los pedidos por estas variables debería resultar en una menor cantidad de observaciones, es necesario contemplar también todas las combinaciones que no tienen pedidos. Esta información es también valiosa ya que es necesario indicarles a los algoritmos no sólo cuando hay pedidos, sino cuando y donde no los hay. Para cada combinación única (turno, día de la semana, tipo de servicio y estación) hay aproximadamente 52 observaciones en los datos.

#### 5.2.2. Asignación de Móviles

El modelo usa la demanda de pedidos como información para decidir la cantidad de móviles a ubicar por estación, turno, tipo y día. Luego busca la manera de aumentar esta demanda. Hacer esto permite entrenar modelos que sobreestimen la demanda y así asignar móviles de más, los cuales pueden quedar ociosos para atender los casos en los que una predicción de demanda tradicional no podría.

Cabe recordar que la naturaleza del problema de este trabajo no busca hacer una predicción sobre los pedidos, sino que intenta buscar formas de mejorar el servicio traducido en el tiempo que esperan los usuarios.

La estimación de la asignación de móviles se calcula utilizando una parte de la muestra para entrenar el modelo y obtener la cantidad de móviles que debe haber por tipo y estación para cada día y turno. A esta porción de los datos se la considera el conjunto de entrenamiento.

### 5.3. Conjunto de datos de entrenamiento y validación

Los datos que se utilizan para entrenar el modelo son una muestra aleatoria del 50% de días de la semana completos. La cantidad de observaciones para este grupo son, entonces, 2.165.486. Las variables *MesAño*, *DiaMes*, *DiaSemana* y *Estación* son transformadas a categóricas ya que, si bien están representadas numéricamente, no tienen un ordenamiento que le brinde información al modelo. A su vez, estas variables categóricas se transformaron en binarias (*One Hot Encoding*) para utilizar algoritmos de aprendizaje automático.

La parte remanente de la muestra compone el conjunto de validación. Este es utilizado para medir la eficacia del modelo ya que son datos que no fueron utilizados para entrenarlo.

### 5.4. Algoritmos de Predicción

El modelo utilizará una serie de algoritmos. Algunos utilizarán estrategias simples tales como los promedios por estación, mientras que otros serán más complejos, tal como el algoritmo de regresión de potenciación de gradiente (*Gradient Boosting Regression*). Se optimizarán los hiperparámetros de cada algoritmo para maximizar las predicciones del modelo.

Adicionalmente, como el problema de cobertura permite que haya unidades ociosas con tal de que la tasa de demora sea menor, se adiciona un parámetro que sirva de ajuste para que las predicciones de los algoritmos estén por encima de la predicción original y así puedan cubrir demandas que estén por encima de lo predicho. De lo contrario, los algoritmos buscarán predecir exactamente la cantidad de auxilios, por lo que cualquier auxilio adicional quedaría demorado. En consecuencia, los siguientes modelos de predicción de auxilios son entrenados para predecir una combinación lineal entre la demanda de pedidos observada y el desvío de la distribución para cada combinación única (de tipo, turno, estación y día de la semana). De esta forma, se obtiene una predicción que no solo responde a la cantidad de pedidos, sino que considera que debe haber más móviles para atender los excedentes que las predicciones sin desvíos no cubrirían. Así se puede minimizar las demoras.

Formalmente, para cada modelo  $m$ , hay una predicción sobre la cantidad de móviles a ubicar para día de la semana  $d$ , turno  $t$ , estación  $s$  y tipo de servicio  $i$ . Esta función, de aquí en adelante llamada  $pred()$ , varía en su forma para cada modelo. Las cantidades de unidades a colocar son:

$$moviles_{dtsim} = pred_m(d, t, s, i)$$

Donde:

- $pred_m$  es la función de cada uno de los distintos modelos  $m$  que predice la cantidad de móviles que deben ser colocados (cada modelo utiliza una función distinta). Cada una de estas es calibrada utilizando los datos del conjunto de entrenamiento. Para entrenar cada modelo se utiliza una variable dependiente modificada.

$$pedidos'_{dtsi} = pedidos_{dtsi} + b_{mi} * \sigma_{dtsi}$$

- $\sigma_{dtsi}$  es el desvío de cada turno, estación día y tipo usando los datos de entrenamiento.
- $b_{mi}$  es un parámetro no negativo que se utiliza para minimizar la tasa de demora de los datos de validación respetando las restricciones de unidades disponibles. Cuanto mayor sea  $b$ , más móviles “extra” se predicen y menor es la tasa de demora. Para cada modelo existe un parámetro  $b_{mi}$  que aumenta de manera artificial la demanda observada de pedidos de cada tipo. El valor óptimo de este parámetro ( $b_{mi}^*$ ) es el que minimiza la tasa de demora.

$$b_{mi}^* = \arg \min_b score(moviles, pedidos)$$

La búsqueda del parámetro  $b_{mi}^*$  se llevó a cabo iterando sus valores iniciales 100 veces y se evaluaron en función de la tasa de demora que generaban. De esta forma, la variable dependiente se vio aumentada (o sin cambios si  $b_{mi} = 0$ ) resultando en 100 distintos conjuntos de entrenamiento. Esto permite que el modelo se entrene con un sesgo positivo y luego se entrene para tener menores tasas de demora. Luego de estas iteraciones, los parámetros que generan estimaciones con la menor tasa de demora (ver 5.5. Criterio de Evaluación) son llamados  $b_{mi}^*$ .

#### 5.4.1. Modelo de Referencia

Se utilizó como modelo de referencia la manera actual de asignación de pedidos descrita en el Capítulo 1. El modelo utilizado respeta la cantidad de vehículos máximos por tipo ya que es el que se utiliza en la realidad.

#### 5.4.2. Modelo de Distribución Uniforme

El primer modelo busca resolver el problema de la manera más sencilla posible. Divide toda la flota de vehículos disponibles de manera uniforme a lo largo de todas las estaciones de espera, en todo momento. Una ventaja, desde el punto de vista del cliente, es que la totalidad de la flota está activa el 100% del año. Una gran desventaja es que el modelo no utiliza las variables independientes para predecir los pedidos. Únicamente se utilizan las cantidades máximas de

talleres y de remolques para dividirlos entre el total de estaciones. El modelo, al ser muy simple, tiene una forma funcional simplificada que no toma en cuenta las variables de las observaciones:

$$moviles_{dtsi,uniforme} = pred_{uniforme}(d, t, s, i)$$

$$moviles_{dtsi,uniforme} = \frac{|I_i|}{|S|}$$

Donde,

$S$  = conjunto de estaciones

$I_i$  = conjunto de móviles del tipo  $i$

#### 5.4.3. Modelo de Medias

El siguiente modelo es apenas más sofisticado. Se computaron medias y desvíos para cada combinación de turno, tipo de pedido, estación y día de la semana utilizando los datos del conjunto de entrenamiento. Se ajustó la cantidad de móviles utilizando el parámetro  $b$ . Este modelo, entonces, utiliza una combinación lineal de la media y desvío.

$$moviles_{dtsi,lineal} = pred_{lineal}(d, t, s, i)$$

$$moviles_{dtsi,lineal} = \mu_{dtsi} + b_{i,lineal} * \sigma_{dtsi}$$

Donde,

$\mu_{dtsi}$  = promedio de pedidos por día  $d$ , turno  $t$ , estación  $s$  y tipo  $i$

#### 5.4.4. Modelo de Regresión Poisson

La regresión Poisson es un método que suele ser utilizado para predecir un número de eventos independientes. Dadas las características del caso, parece relevante utilizar esta variante como alternativa de modelado utilizando los datos de entrenamiento y validación para optimizar hiper parámetros.

$$moviles_{dtsi,poisson} = pred_{poisson}(d, t, s, i)$$

$$moviles_{dtsi,poisson} = e^{\beta_1 d + \beta_2 t + \beta_3 s + \beta_4 i}$$

Los parámetros  $\beta$  de este modelo fueron entrenados utilizando la combinación lineal entre los pedidos y sus desvíos para lograr la menor cantidad de demoras. El paquete utilizado fue *sklearn*, el cual utiliza además un hiper parámetro  $\alpha$  como penalizador.

#### 5.4.5. Modelo de Potenciación de Gradiente

Este modelo utilizó el algoritmo de potenciación de gradiente para predecir la cantidad de móviles a ubicar. El proceso de potenciación de gradiente consiste en crear una serie de modelos débiles, que son capaces de hacer predicciones sólo ligeramente mejores que una elección aleatoria. Luego, se combinan estos modelos débiles para formar un modelo más fuerte y preciso.

En cada iteración del proceso, se ajusta el modelo débil para minimizar el error residual en los datos de entrenamiento, utilizando una técnica conocida como descenso de gradiente. Esto significa que se busca una dirección en la que se pueda reducir el error residual de manera más eficaz y se ajusta el modelo en esa dirección.

Después de varias iteraciones, se combinan los modelos débiles ajustados para crear un modelo más fuerte y preciso que es capaz de hacer predicciones más precisas que cualquiera de los modelos débiles por sí solos. Para este trabajo se combinó esta predicción con el desvío del conjunto de entrenamiento.

$$moviles_{dtsi,boost} = pred_{boost}(d, t, s, i)$$

$$moviles_{dtsi,boost} = \sum_{m=1}^M \gamma * h_m(d, t, s, i)$$

El modelo usa, además, los hiper parámetros:

- *Tasa de aprendizaje ( $\gamma$ )*, como la contribución que hace cada iteración del proceso  $h$ .
- *Número de estimadores ( $M$ )*, como la cantidad de veces que se repite el proceso.
- *Sub-muestra*, como el porcentaje de la muestra que utiliza cada iteración del proceso.

### 5.5. Criterio de Evaluación del Modelo de Estimación de Demanda

Para evaluar esta segunda parte del modelo (los modelos de estimación de demanda), primero se computan las cantidades de unidades por estación, día, turno y tipo obtenidas a partir de usar el modelo entrenado con el conjunto de validación. A partir de estas estimaciones, se comparan la cantidad de unidades asignadas con las observadas en el conjunto de validación y se computa la tasa de demoras totales de todo el modelo, es decir, la suma de la diferencia entre los pedidos actuales y los móviles asignados de cada estación, turno, día y tipo.

Las observaciones no utilizadas en el conjunto de entrenamiento serán utilizadas para evaluar la eficacia del modelo. La misma se medirá en función de la tasa de demoras totales, es decir, la suma de la diferencia entre los pedidos actuales y los móviles asignados de cada estación para cada turno. La función de *scoring* es, entonces, la tasa de demora. Se define a continuación:



$$score(y, pred) = \sum_{d \in D} \sum_{t \in T} \sum_{i \in I} \sum_{s \in S} (y_{dtsi} - pred_{dtsi}) / count(y)$$

Donde,

$D$  = conjunto de días de la semana

$T$  = conjunto de turnos

$S$  = conjunto de estaciones

$I$  = conjunto de tipos de pedidos

$y_{dtsi}$  : pedidos reales por día de la semana  $d$ , turno  $t$ , estación  $s$ , tipo  $i$

$pred_{dtsi}$  : asignaciones estimadas por día de la semana  $d$ , turno  $t$ , estación  $s$ , tipo  $i$

$count(y)$  : cantidad total de pedido

Además, se deberá respetar la cantidad máxima de unidades de 178 grúas y 98 talleres.

La puntuación del modelo utiliza como punto de partida la tasa actual de demoras del modelo utilizado por la empresa: 12,19%.

## 6. Resultados

Este capítulo describe los resultados obtenidos para ambas partes del modelo. Primero, en 6.1. se muestra cómo se eligió la mejor configuración espacial de estaciones de espera. Segundo, en 6.2., utilizando esta configuración, se revela cómo se asigna cada pedido a una de estas estaciones (demanda observada). Tercero, en 6.3., se expone el rendimiento de los distintos modelos de estimación de demanda utilizando las asignaciones del paso anterior. Por último, en 6.4, se comparan los desempeños de cada algoritmo.

### 6.1. Emplazamiento de Ubicaciones

Tras simular el proceso descrito en la sección número 4, la versión que asignó la menor cantidad de estaciones tomó 268 puntos candidatos distribuidos aleatoriamente en todo el territorio de la ciudad y resolvió 34 puntos que logran cubrir todo el territorio con un radio de 1 kilómetro. A estos puntos se les asignó un identificador único. La Figura 16 ilustra esta configuración junto con el área de cobertura de cada estación.

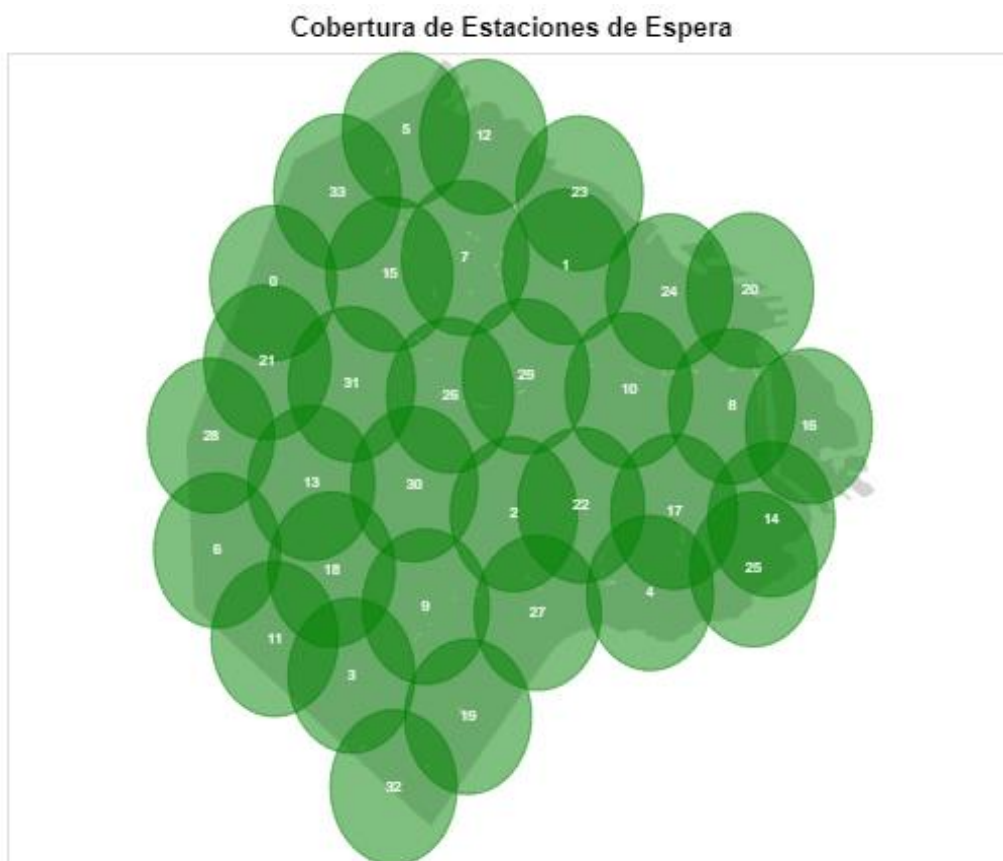


Figura 16. Ejemplo de cobertura de estaciones de espera óptimas.

### 6.2. Estimación de la Oferta

Ya con las estaciones ubicadas de forma tal que permiten una cobertura total del territorio, se utilizaron las transformaciones descritas en el Capítulo 5. Primero, se utilizó la posición de los

pedidos para asignarlos a la estación cuyo centroide está más cercano. Cada pedido, entonces, tiene una estación que le corresponde. Finalmente, se agruparon todos los pedidos por: mes, día del mes, día de la semana; turno (períodos de 4 horas); tipo de servicio; y estación más cercana. Esto permite sumar la cantidad de pedidos que ocurren para cada grupo y tener el conjunto de datos sobre el que se trabajará la optimización de la asignación de móviles.

A continuación, una muestra del conjunto de datos:

MesAno	DiaMes	DiaSemana	Turno	TipoServicio	Estación	Pedidos
8	28	4	8 a 12	R	13	0.0
1	19	6	12 a 16	P	14	1.0
5	31	3	16 a 20	R	8	0.0
4	20	5	4 a 8	T	27	1.0
10	28	5	4 a 8	T	29	0.0

*Tabla 1. Muestra de auxilios asignados.*

La Tabla 1 ejemplifica la forma que toma el conjunto de datos una vez agrupados. La columna *Pedidos* es la que se usa como la demanda de auxilios que hay por mes, día, turno, tipo y estación. Tal como se explicó en la Sección 5.2., al hacer este agrupamiento por combinación de variables, la cantidad de observaciones aumenta dramáticamente. Una consecuencia de esto es que para este nuevo conjunto de datos la gran mayoría de las observaciones tiene una demanda de 0 pedidos. En particular, el 96% de las observaciones tiene una demanda de 0 móviles. Esto puede verse en la Figura 17, que ilustra cómo quedan distribuidas las demandas de unidades para cada combinación.

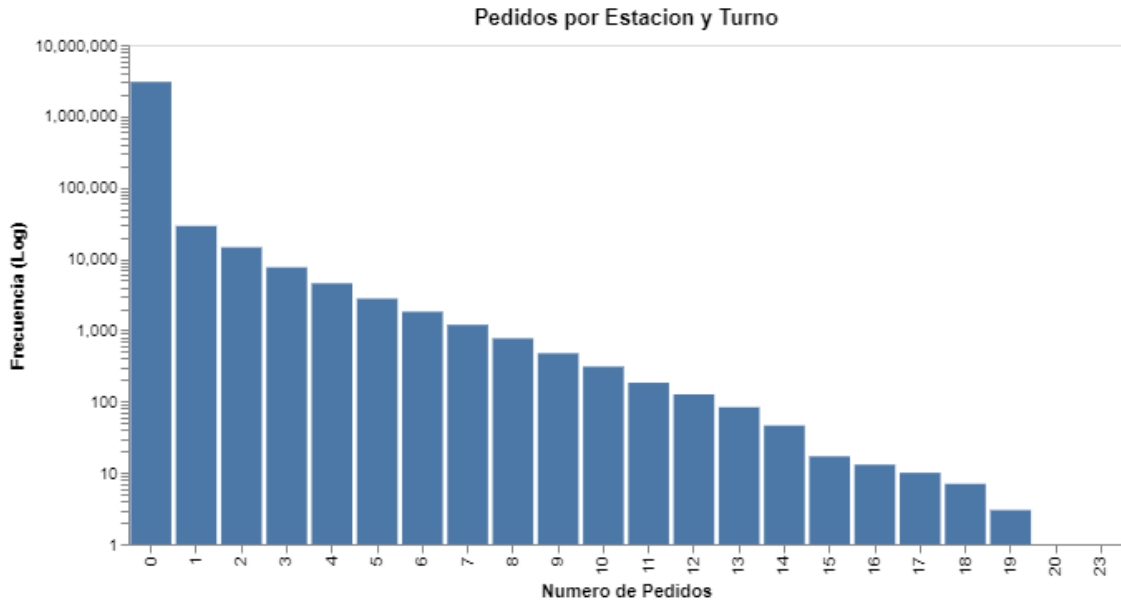


Figura 17. Distribución de la demanda de pedidos (escala logarítmica)

Sin embargo, al mirar la máxima cantidad de pedidos para cada estación-turno a través de la Figura 18, la distribución es más equilibrada (en cuanto a la cantidad de ceros) y muestran por lo tanto que existe una alta variabilidad en los auxilios. Esto indica que hay estaciones cuya demanda es casi siempre 0, pero sería muy arriesgado no colocar móviles ya que cada tanto aparecen pedidos para atender en ellas.

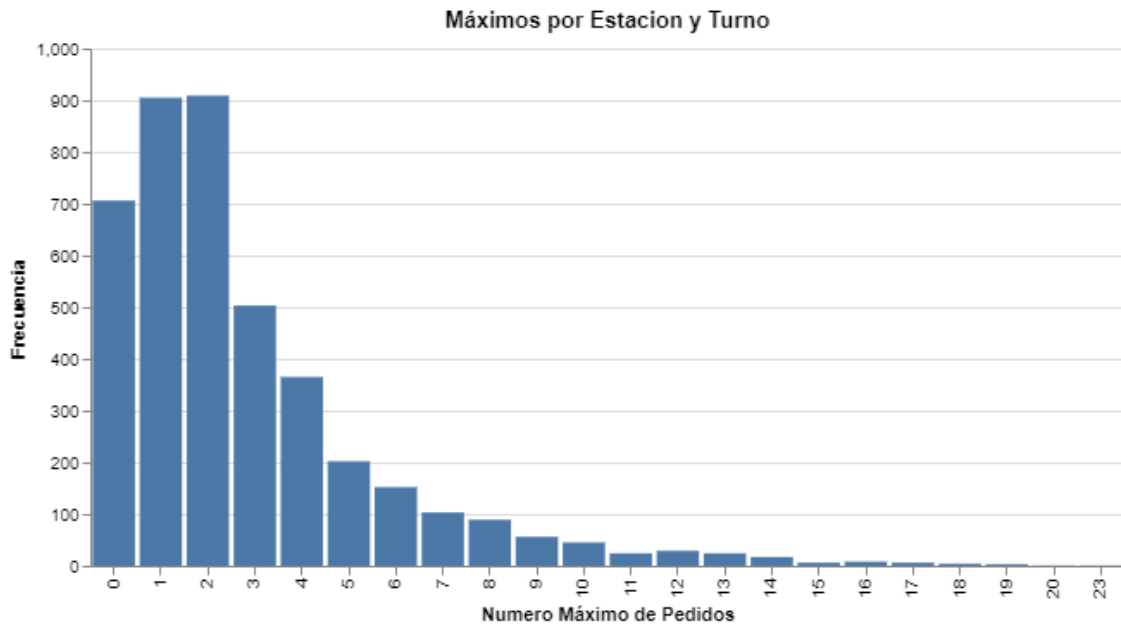


Figura 18. Distribución de la demanda máxima de pedidos

A partir de estas asignaciones puede comenzarse la búsqueda de soluciones para ubicar a las unidades de forma tal que anticipen los pedidos adecuadamente. Una posible solución a este problema, a priori, sería utilizar los máximos por estación-turno como fuente para la asignación de móviles por estación-turno. De esta manera, siempre se estaría capturando el peor de los

casos para cada estación, en todo momento. La estación 5, por ejemplo, siempre tendría 7 móviles en ella para el turno 8 a 12 ya que es el máximo de pedidos recibidos alguna vez para esa estación-turno. Esta opción, por un lado, garantiza que no haya demoras (para cualquier observación de este conjunto de datos), por otro, es muy exigente en cuanto a la cantidad de móviles asignados, y requiere ver si la totalidad de la flota actual (276 unidades: 178 grúas y 98 talleres) puede ser asignada en las estaciones, independientemente del resto de las variables.

Al agrupar los pedidos máximos por turno y estación, y luego tomar los máximos para cada turno, se puede ver que para todos los turnos entre las 4 y las 20, la cantidad de móviles total es mayor a 276 y por ende no alcanza para cubrir los peores escenarios en los cuales todas las estaciones están recibiendo su máximo anual de pedidos en simultáneo. En otras palabras, no es posible ubicar a las unidades para que cubran todos los posibles escenarios con una única configuración que contemple el peor de los escenarios. Esta estrategia no cumple con las restricciones en la cantidad de móviles.

<b>Turno</b>	0 a 4	4 a 8	8 a 12	12 a 16	16 a 20	20 a 24
<b>Auxilios</b>	137	375	586	547	462	230

*Tabla 2. Máximos pedidos simultáneos por turno.*

De haber sido factible esta configuración, la empresa hubiese tenido un modelo que le garantice alcanzar a los clientes a tiempo (cero demoras). No obstante, la imposibilidad de cubrir todos los escenarios obliga a entrenar un modelo que utilice el comportamiento de las variables para predecir la cantidad de pedidos que sucederán por día de la semana, turno y tipo. Es así como, a continuación, se procede a presentar los resultados de cada uno de los modelos de los modelos de predicción con el objetivo de minimizar la tasa de demora.

### 6.3. Algoritmos de Predicción

Tras confirmar la imposibilidad de garantizar una tasa de demoras de cero, fue necesario entrenar los modelos de predicción de demoras dividiendo la muestra entre conjuntos de entrenamiento y validación.

Con el objetivo de obtener un resultado robusto, se entrenó y validó cada modelo 100 veces utilizando diferentes muestras aleatorias de los datos. En cada una de estas repeticiones se optimizaron los parámetros  $b_{remolque}$  y  $b_{taller}$  (los multiplicadores de los desvíos de los conjuntos de entrenamiento) para lograr minimizar la función de *scoring*. Para ello, en cada iteración se generaron aleatoriamente valores de  $b_{taller}$  y  $b_{remolque}$  entre 0 y 4. No se utilizaron

valores más altos ya que utilizarlos implicaba que se asignen por momentos más móviles que el total disponible. La optimización de estos parámetros se llevó a cabo iterando los valores de  $b$  otras 100 veces y así encontrar las  $b^*$  que minimicen la tasa de demora.

Para cada iteración, se calcularon:

- *Las tasas de demora*, como principal variable de *scoring* a optimizar para cada modelo. Esta se obtiene al contrastar las predicciones de cada modelo contra la cantidad de pedidos observados en el conjunto de validación. La fórmula utilizada es la descrita en 5.5.
- *El porcentaje promedio de móviles en circulación por tipo de móvil (grúas y talleres)*, como variable secundaria, ya que, a igualdad de tasa de demora, el modelo que utilice menos móviles será más económico y tendrá más móviles libres que podrán ser usados para otras regiones y/o otros modelos de asignación que la empresa quiera utilizar.
- *El porcentaje máximo de móviles utilizados en cualquier momento*, como variable que controle la restricción de la cantidad de móviles que se pueden usar por tipo. Esta variable no puede superar el 100%.
- *Los parámetros  $b^*_{remolque}$  y  $b^*_{taller}$  para cada modelo.*

Para todos estos indicadores se computaron las medias y desvíos estándar sobre las 100 observaciones obtenidas. Se utilizó un nivel de significatividad del 99% para determinar las diferencias entre los modelos. Estos están explicados en la sección de Desempeño (6.4).

#### 6.3.1. Modelo de Referencia

Se utilizó como referencia la tasa de demora de 12,19% que se logró con el modelo que la empresa utilizó para los datos de la muestra. El modelo utilizado respeta la cantidad de vehículos máximos por tipo ya que es el que se utiliza en la realidad.

#### 6.3.2. Modelo de Distribución Uniforme

Tal como fue descrito en la sección anterior, el siguiente modelo divide toda la flota de vehículos disponibles de manera uniforme a lo largo de las 34 estaciones de espera. Esto resulta en que cada una de las estaciones prediga, en todo momento, 5 (178/34) pedidos de remolques y 3 (98/34) de talleres en todo momento. La tasa de demora obtenida para este modelo fue, en promedio, de 12,15% utilizando, el 92% de los remolques y el 100% de los talleres en todo momento. Este modelo no utiliza los parámetros  $b^*_{remolque}$  y  $b^*_{taller}$ , por lo que se les asigna 0 ambos. La distribución de estas asignaciones queda ilustrada en la Figura 19.

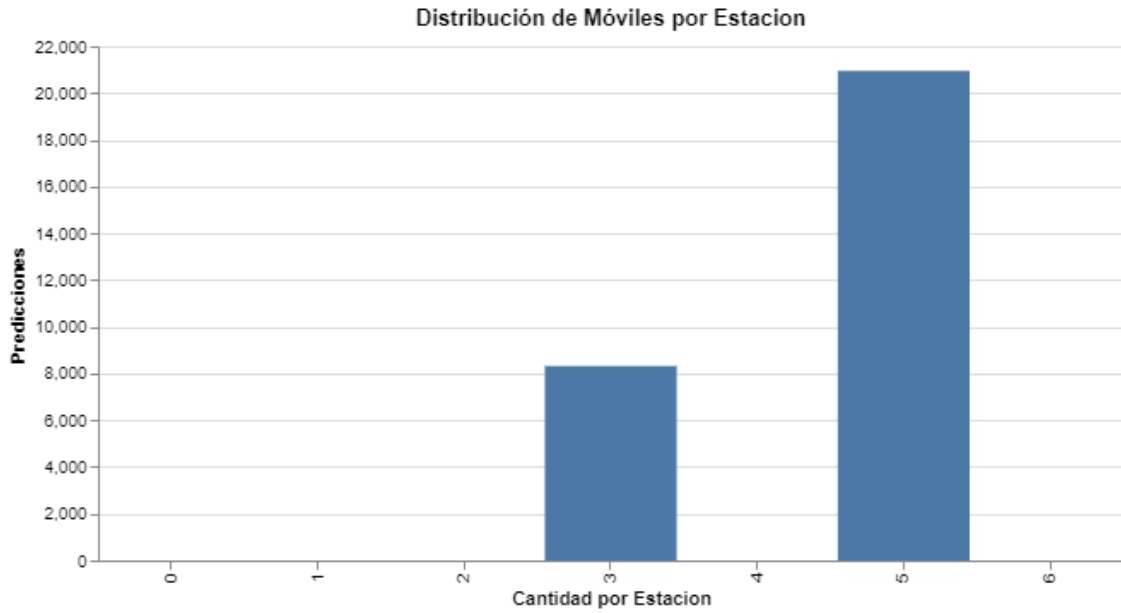


Figura 19. Ejemplo de distribución de móviles por estación, turno, día y tipo para modelo uniforme.

### 6.3.3. Modelo de Medias

Para el modelo de medias, los resultados mostraron un promedio de tasas de demoras de 13,50% utilizando, en promedio, el 18% de los talleres y el 20% de los remolques. Los parámetros  $b^*_{remolque}$  y  $b^*_{taller}$  promedio fueron de 0,40 y de 3,26 respectivamente. La distribución de las asignaciones de este modelo puede verse en la Figura 20.

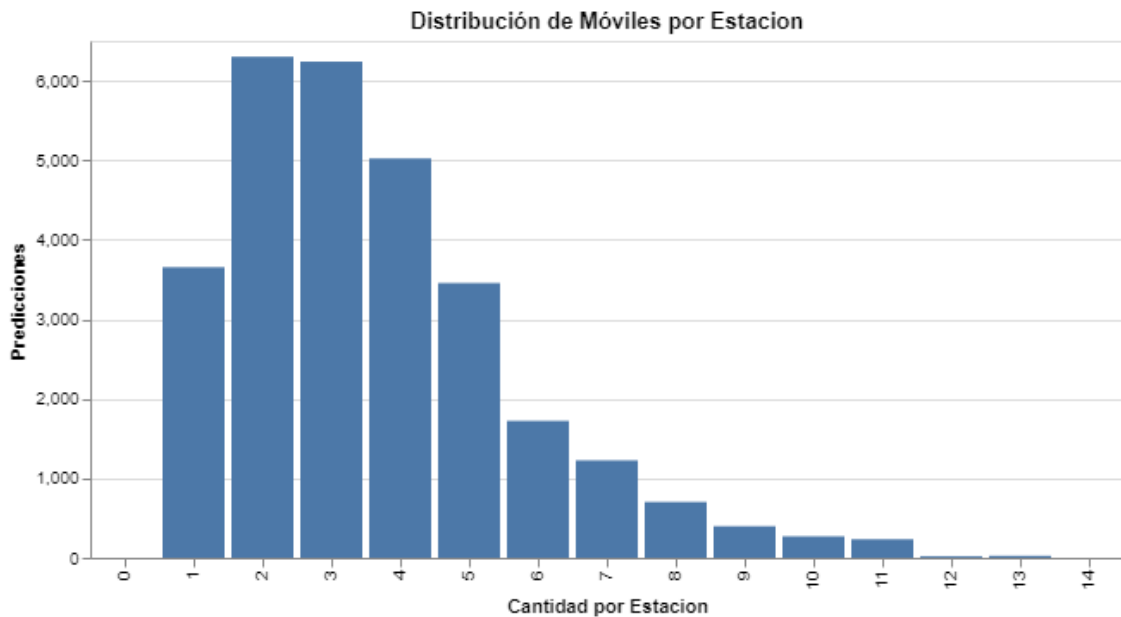


Figura 20. Ejemplo de distribución de móviles por estación, turno, día y tipo para modelo de medias.

Todas las estaciones tienen, en todo momento, al menos una unidad esperando. Luego, gran parte de las estaciones tienen entre 2 y 5 para capturar la mayoría de las situaciones. A partir de las 6 unidades por estación puede verse un declive gradual que predice anomalías.

#### 6.3.4. Modelo de Regresión Poisson

El cuarto modelo propuesto por este trabajo utilizó el hiper parámetro de penalización  $\alpha = 0,5$ , y los combinó con los parámetros  $b_{remolque}^*$  y  $b_{taller}^*$ , que en promedio rondaban 0,76 y 3,24 respectivamente. Logró promedios de tasas de demora del 8,99% utilizando 20% de los talleres y 22% de los remolques.

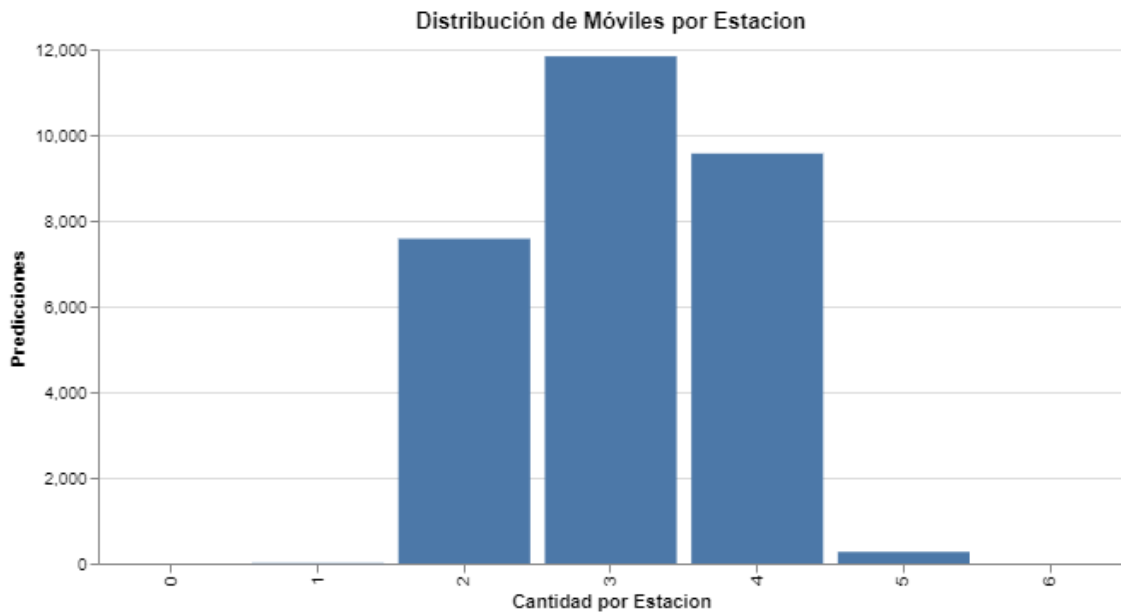


Figura 21. Ejemplo de distribución de móviles por estación, turno, día y tipo para modelo de regresión Poisson.

La distribución de las unidades asignadas, a diferencia del modelo anterior, tiene, casi siempre, al menos 2 móviles por estación. Además, nunca supera las 6 unidades.

#### 6.3.5. Modelo de Potenciación de Gradiente

Este modelo utilizó el algoritmo de potenciación de gradiente para predecir la cantidad de pedidos. Luego combinó esta predicción con el desvío del conjunto de entrenamiento. Utilizando una tasa de aprendizaje de 0,005, 220 estimadores y sub muestras del 30%, este algoritmo, encontró  $b_{remolque}^*$  y  $b_{taller}^*$  de 0,97 y 3,21. Obtuvo tasas de demora promedio de 8,45% usando alrededor de 21% de talleres y 23% de remolques.



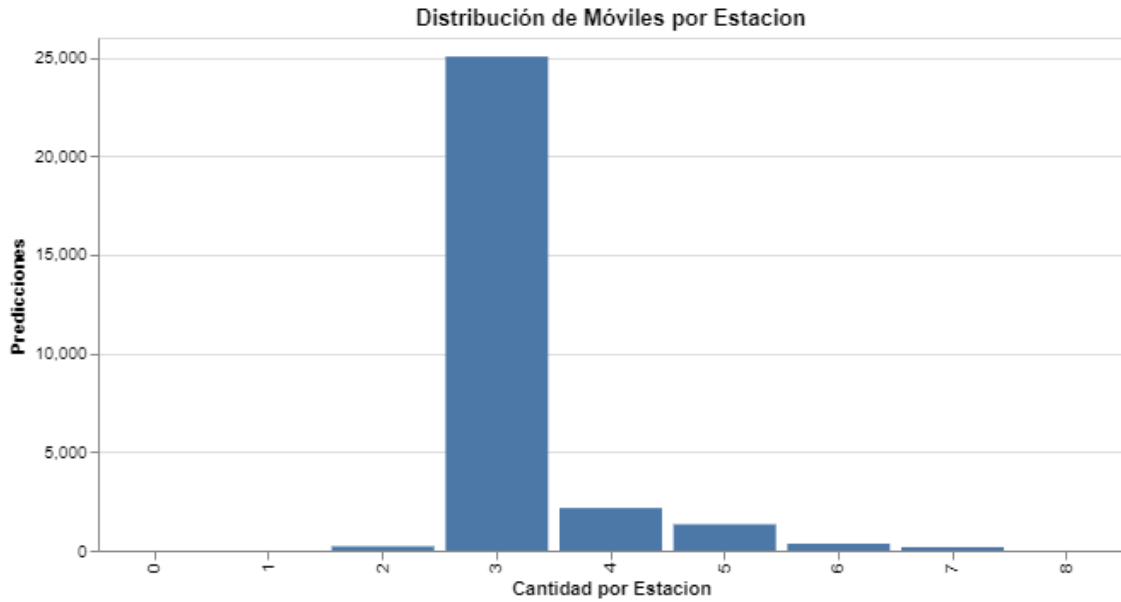


Figura 22. Ejemplo de distribución de móviles por estación, turno, día y tipo para modelo de potenciación de gradiente.

La particularidad de la distribución de este algoritmo es que ubica, para casi todos los casos, 3 unidades por estación. No obstante, también contempla casos más extremos que la regresión Poisson.

#### 6.4. Desempeño

El rendimiento de los 5 modelos entrenados en 100 reiteradas ocasiones puede verse resumido en la Tabla 3. A su vez, la Figura 23 muestra la distribución en las tasas de demora de los distintos modelos, incluyendo la referencia.

Modelo	Tasa de Demora	Prom. Talleres	Prom. Remolques	$b^*_{taller}$	$b^*_{remolque}$
Referencia	12,19%	-	-	-	-
Uniforme	12,15% (0,22%)	100% (0%)	100% (0%)	0 (0)	0 (0)
Medias	13,50% (1,13%)	18% (1,33%)	20% (1,55%)	3,26 (0,66)	0,40 (0,19)

Modelo	Tasa de Demora	Prom. Talleres	Prom. Remolques	$b^*_{taller}$	$b^*_{remolque}$
Regresión Poisson	8,99% (0,78%)	20% (1,49%)	22% (1,38%)	3,24 (0,56)	0,76 (0,20)
Potenciación de Gradiente	8,45% (0,57%)	21% (1,32%)	23% (1,43%)	3,21 (0,69)	0,97 (0,16)

Tabla 3. Resultados obtenidos luego de 100 iteraciones. Medias y desvíos estándar (entre paréntesis).

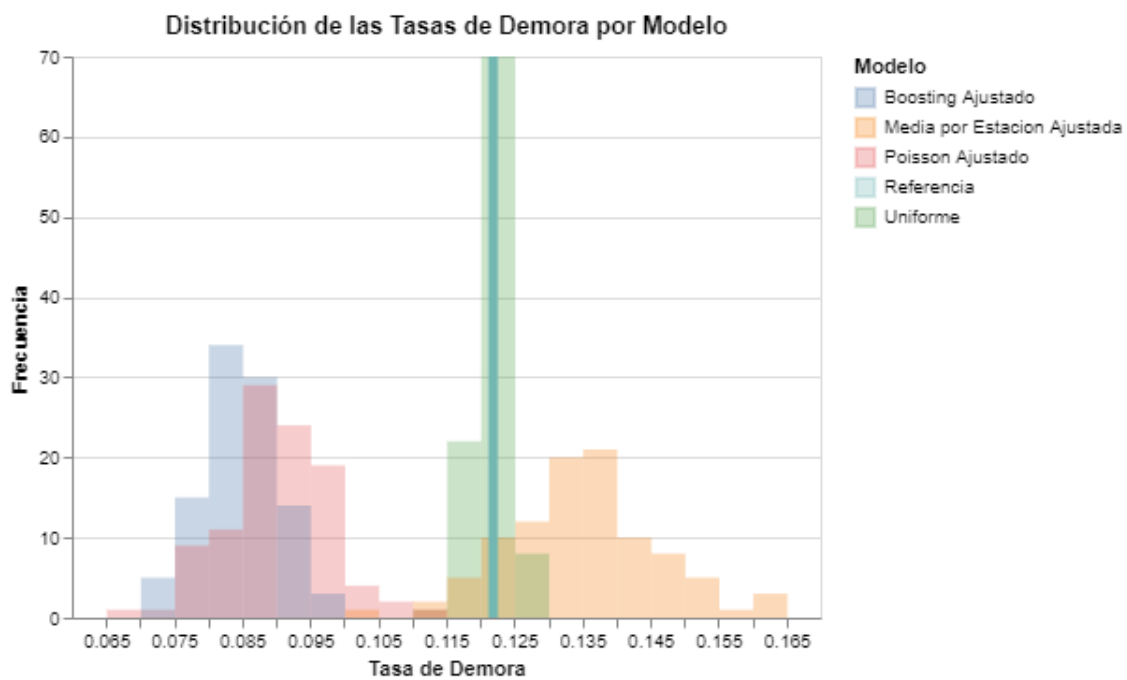


Figura 23. Distribución de las tasas de demora por modelo.

Los resultados del proceso iterativo muestran, en primer lugar, que tanto el modelo uniforme como el de medias no logran tener de manera consistente una menor tasa de demora que el modelo de referencia. Si bien la media del modelo uniforme está por debajo de la referencia, el test de hipótesis (1) no rechaza la hipótesis nula de que, en realidad, la media del modelo uniforme sea la mayor o igual que la del modelo de referencia con un 99% de confianza.

Hipótesis Nula	P Valor	Estado
1. La tasa de demora del modelo <b>uniforme</b> es mayor o igual que la referencia.	0,05	No rechazada
2. La tasa de demora del modelo de <b>medias</b> es mayor o igual que la referencia.	1,00	No rechazada
3. La tasa de demora del modelo de regresión <b>Poisson</b> es mayor o igual que la referencia.	0,00	Rechazada
4. La tasa de demora del modelo de <b>potenciación de gradiente</b> es mayor o igual que la referencia.	0,00	Rechazada
5. La tasa de demora del modelo Poisson es igual o menor que la de potenciación de gradiente.	0,00	Rechazada
6. La cantidad promedio de móviles que utiliza el modelo de potenciación de gradiente es menor o igual al de Poisson.	0,00	No rechazada

*Tabla 4. Test de hipótesis.*

Los dos modelos restantes, la regresión Poisson y la potenciación de gradiente, logran entonces una menor tasa de demora en comparación con la referencia. La diferencia entre ambos a nivel de tasa de demora existe y se demuestra mediante la hipótesis nula número 5. Con un 99% de confianza, el modelo de potenciación de gradiente tiene una menor tasa de demora. No obstante, el no rechazo de la hipótesis número 6 indica que el modelo Poisson utiliza una menor cantidad promedio de móviles, también al 99% de confianza. Esto indica que el modelo de potenciación, si bien es el ganador, no es superior en todo sentido, por lo que puede haber situaciones en las que el modelo Poisson tenga sentido ser utilizado.

Independientemente del modelo finalista que se seleccione, el desempeño de los modelos muestra que la tasa de demoras puede ser mejorada entre un 35% y un 40% con respecto a la referencia, utilizando aproximadamente el 22% de los móviles.

## 7. Conclusiones

El trabajo ha logrado ilustrar como un problema de cobertura puede ser atacado combinando técnicas de optimización con aprendizaje automático. La incertidumbre que existe alrededor de los incidentes vehiculares puede ser disminuida utilizando la información de todos los móviles, a diferencia de la experiencia de un solo conductor.

La limitación de la estacionalidad de los datos anuales existe. Al tener combinaciones únicas de las observaciones de mes, día y turno, el análisis puede estar sesgado hacia lo que sucedió específicamente entre abril 2019 y marzo 2020. Sin embargo, la empresa proveedora de los datos puede utilizar la metodología descrita en este trabajo, utilizar datos de otros años y verificar si las mejoras en demoras existen.

La selección de estaciones permite a orientar a los conductores de los móviles y al mismo tiempo darles libertad para elegir el punto de espera dentro de la estación. Gracias al modelo de emplazamiento de ubicaciones, la cantidad de estos puntos es la mínima posible, por lo que los encargados de las operaciones tienen menos posibilidad de cometer errores. Además, evita que los conductores hagan largos viajes de manera urgente, ya que, en la mayoría de los casos, no deberían conducir más de 2 kilómetros apremiados por el tiempo.

En cuanto a la estimación de la cantidad de móviles por estación, fue necesario utilizar técnicas de ensamble que logren predecir los pedidos por demás. Esta metodología puede ser utilizada para otros tipos de problema de cobertura con incertidumbre, en los que haya que atender zonas sin saber exactamente cuántos incidentes por zona habrá.

La inclusión de los últimos 12 días de marzo 2020, en los cuales la movilidad se vio reducida, pudo haber reducido levemente la demanda estimada. No obstante, al sobre estimarla para garantizar cobertura, la metodología fue robusta para contemplar este tipo de interrupciones, siempre y cuando su participación en la muestra se encuentre en ese espectro.

En cuanto a las técnicas de predicción de móviles, el modelo de potenciación de gradiente resultó ganador. Además de tener una menor tasa de demora con un 99% de significatividad, permite tener, en promedio, más de 77% de los móviles inactivos. Esto permite no solo ahorrar en el costo de la operación, sino que pueden usarse parte de los inactivos para cubrir anomalías adicionales utilizando otros criterios.

Combinar el modelo con el que la empresa utiliza actualmente puede tener la ventaja de desplegar móviles que se muevan libremente por todo el territorio y no deban respetar su área de cobertura a rajatabla.

Cabe destacar, además, que la tasa de demora obtenida con estos modelos (a excepción del de referencia) es conservadora en tres sentidos.

Primero, al relajar el área de cobertura asignada por estación seguramente se reduzca aún más la tasa de demora, ya que la definición de demora para los modelos implicaba que un pedido cae en demora si no hay auxilios disponibles en la estación más próxima. No obstante, puede suceder que haya un móvil libre en otra estación que también se encuentre a menos de 1 kilómetro y pueda llegar al pedido sin demora alguna (ver Figura 24). Adicionalmente, en la práctica, gran parte de los móviles se trasladan a una mayor distancia sin demorarse, por lo que un móvil de una estación cercana podría ocasionalmente atender un pedido que le corresponda a otra, incluso si se encuentra a más de 1 kilómetro.

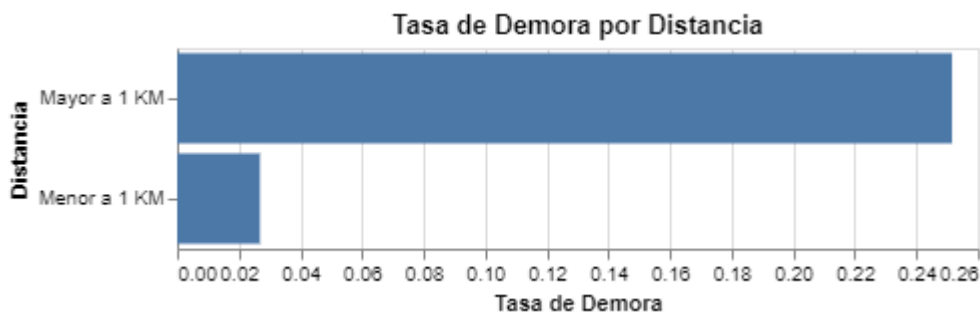


Figura 23. Tasa de demora en función de la distancia recorrida.

Segundo, se está asumiendo que los móviles sólo atienden un pedido por turno para que, en el peor de los casos, si todos los pedidos ocurren de manera simultánea dentro del turno, puedan ser atendidos sin demoras. Relajar este supuesto también reduciría dramáticamente la tasa de demoras ya que la duración de los pedidos está muy por debajo de las 4 horas.

Por último, también se asume que los talleres brindan exclusivamente servicios de taller, cuando en realidad pueden tomar cualquier tipo de pedido. La restricción que aplica a la mayoría de los móviles (los remolques) hizo necesario dividir los servicios para asegurar que los pedidos auxilio de taller tengan móviles compatibles listos para atenderlos, y que no haya talleres móviles haciendo remolques cuando son los móviles que menos comunes son. Los talleres, entonces, podrían llegar a atender remolques como último recurso.

El presente trabajo ha logrado mostrar cómo el ensamble de técnicas de optimización combinatoria con algoritmos de aprendizaje automático puede ser útil para mejorar la operación de empresas que brinden servicios de cobertura. En particular, al quitar del proceso la decisión individual de los conductores sobre los lugares de espera y reemplazarlos por técnicas basadas en datos, se logró disminuir la tasa de demoras en, al menos, un 40%. Esto

significa una diferencia sustancial entre los métodos, por lo que sería apropiado ensayar el modelo presentado en la operatoria real de la empresa y, de ser exitoso, extrapolarlo a otras regiones.

## 8. Referencias

- Church, R. L., & Murray, A. T. (2018). Location covering models: History, applications and advancements (1st edition 2018). Springer
- Toregas, C., Swain, R., ReVelle, C., & Bergman, L. (1971). The location of emergency service facilities. *Operations Research*, 19(6), 1363–1373.
- Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.
- Ramanathan, R., & Tonk, S. (2017). Optimal supply chain network design under oligopoly competition with pricing decisions. *Transportation Research Part E: Logistics and Transportation Review*, 103, 80-104.
- Ghorbel, I., Chérif, A. S., & Masmoudi, M. M. (2018). Robust portfolio optimization under transaction costs and market impact. *European Journal of Operational Research*, 270(2), 774-787.
- Ghannadpour, S. F., & Dullaert, W. (2012). Facility location problems: a survey of applications and methods. *OR spectrum*, 34(4), 753-774.
- Choi, J., Kim, J. H., Park, J., Kim, Y., & Lee, J. H. (2019). Boosted Poisson regression for predicting hospital readmissions. *Journal of biomedical informatics*, 93, 103154.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 29(5), 1189-1232.
- Kim, H. J., & Ahn, J. (2019). Development of a boosting algorithm for credit risk scoring. *Journal of computational and applied mathematics*, 356, 1-13.
- Liang, S., Peng, M., Yang, M., Wang, L., & Lv, Y. (2020). Predicting traffic accidents with a combination of Poisson regression and gradient boosting machine. *Accident Analysis & Prevention*, 138, 105425.
- Liu, Y. (2018). Analysis of purchase data based on Poisson regression model. *Journal of applied statistics and management*, 37(1), 87-92.

- Pedregosa *et al.*, Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.
- Alvelos, F., & Clímaco, J. (2019). A review of exact algorithms for the location set covering problem. *Computers & Operations Research*, 102, 101-119. <https://doi.org/10.1016/j.cor.2018.09.003>
- Biswas, A., & Mookerjee, R. (2021). Location set covering problem: A review of algorithms and applications. *Journal of Industrial and Production Engineering*, 38(1), 1-18. <https://doi.org/10.1080/21681015.2020.1865695>
- Singh, A., & Nagwani, N. K. (2020). Hybridizing optimization and machine learning algorithms for the location set covering problem: A review. *Artificial Intelligence Review*, 53(5), 3645-3686. <https://doi.org/10.1007/s10462-020-09806-8>
- Gendreau, M., Laporte, G., & Semet, F. (2010). A hybrid tabu search algorithm for the location set covering problem. *Computers & Operations Research*, 37(10), 1813-1825. <https://doi.org/10.1016/j.cor.2009.12.005>
- McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models*, Second Edition. Boca Raton: Chapman and Hall/CRC. ISBN 0-412-31760-5.
- Asistencia vial. (2023). Wikipedia, la enciclopedia libre. [https://es.wikipedia.org/w/index.php?title=Asistencia\\_vial&oldid=150221643](https://es.wikipedia.org/w/index.php?title=Asistencia_vial&oldid=150221643)
- "How Roadside Assistance Was Born". Your AAA Network. 2019-12-06. Extraído en 2020-02-21.
- Ciudad de Buenos Aires | Buenos Aires Ciudad—Gobierno de la Ciudad Autónoma de Buenos Aires. (n.d.). Extraído en 2023-04-10 de <https://buenosaires.gob.ar/laciudad/ciudad>
- INDEC: Instituto Nacional de Estadística y Censos de la República Argentina. (n.d.). Extraído en 2023-04-10, de <https://www.indec.gob.ar/indec/web/Nivel4-Tema-2-41-135>
- Buenos Aires Data. (n.d.). Buenos Aires Data. Extraído en 2023-04-10, de <https://data.buenosaires.gob.ar/>
- Vehicles in use | www.oica.net. (n.d.). Extraído en 2023-04-10, de <https://www.oica.net/category/vehicles-in-use/>
- Asociación de Fábricas Argentinas de Componentes. (2021). *Flota Circulante en Argentina 2020*.
- Buenos Aires Observatorio | Historial Meteorológico. (n.d.). Meteostat. Extraído en 2023-04-10, de <https://meteostat.net/es/station/87585?t=2019-01-01/2020-12-31>