

# IMPROVEMENT IN THE PLANNING PROCESS OF LOGISTICS UNITS USING PREDICTIVE ANALYTICS IN A LAST MILE COMPANY IN METROPOLITAN LIMA

## MEJORA EN EL PROCESO DE PLANIFICACIÓN DE LAS UNIDADES LOGÍSTICAS UTILIZANDO ANÁLISIS PREDICTIVO EN UNA EMPRESA DE ÚLTIMA MILLA EN LIMA METROPOLITANA

Ahmad Daniel Hussen Jimenez<sup>1</sup>, Fabiana Angelid Retamozo Torres<sup>1</sup>, Figo García Santos<sup>1</sup>, María Bélen Ortiz Garcia<sup>1</sup>

<sup>1</sup> Universidad de Ingeniería y Tecnología (UTEC), Lima, Perú



[ahmad.hussen@utec.edu.pe](mailto:ahmad.hussen@utec.edu.pe)

Recebido: 11 novembro 2024 / Aceito: 11 novembro 2024 / Publicado: 01 dezembro 2024

### ABSTRACT.

This research focused on improving the efficiency of logistics planning for a last-mile company, using Scharff, a rapidly growing Peruvian logistics operator. The methodology applied, called Predictive Analytics in the Supply Chain, comprises three phases: understanding the planning process, building predictive models, and simulating to optimize vehicle capacity, reduce costs, and improve planning. In the first phase, a process diagram was created, and its characteristics were quantified to identify the main problem: low vehicle utilization (26% of their capacity), which led to a monthly waste of S/.236 thousand. In the second phase, demand was forecasted using neural networks with gradient descent, achieving an RMSE of 0.21. Simultaneously, decision trees were used to predict logistical data before arrival at the warehouse. In the final phase, demand forecasting and the prediction algorithm were combined to define the daily distribution report and calculate the number of vehicles needed. As a result, emergent damage was reduced by S/.122 thousand monthly by increasing average utilization by 54 percentage points.

**Keywords:** Planning, Predictive Analytics, Efficiency, Capacity, Prediction.

### RESUMEN

Esta investigación se centró en mejorar la eficiencia de la planificación de las unidades logísticas de una empresa de última milla, utilizando a Scharff, una operadora logística peruana en crecimiento exponencial. La metodología aplicada, llamada Predictive Analytics en la Cadena de Suministro, comprende tres fases: Comprensión del proceso de planificación, construcción de modelos predictivos y simulación para optimizar la capacidad de los vehículos, reducir costos y mejorar la planificación. En la primera fase, se creó un diagrama de proceso y se cuantificaron sus características para identificar el problema principal: Una baja utilización de los vehículos (26% de su capacidad), lo que generaba un desaprovechamiento de S/.236 mil mensuales. En la segunda etapa, se pronosticó la demanda con redes neuronales utilizando descenso de gradiente, logrando un RMSE de 0,21. Paralelamente, se usaron árboles de decisión para predecir datos logísticos antes de la llegada al almacén. En la última fase, se combinaron el pronóstico de demanda y el algoritmo de predicción para definir el reporte de distribución diario y calcular la cantidad de vehículos necesarios. Como resultado se logró reducir el daño emergente en 122 mil soles mensualmente al aumentar la utilización media en 54 puntos porcentuales.

**Palabras claves:** Planificación, Predictive Analytics, Eficiencia, Capacidad, Predicción.



## 1. INTRODUCCIÓN

El comercio de última milla es uno de los tipos de entrega más modernos en la actualidad. Según AR Racking, la 'última milla' representa la fase final de la entrega de productos, donde se traslada la mercancía desde un punto de almacenamiento hasta el destino final especificado por el cliente. Esta etapa es crucial para la satisfacción del cliente y la percepción general del servicio de entrega (AR Ranking Perú, 2022).

Durante la pandemia en el Perú, según la Cámara Peruana de Comercio Electrónico la evolución del ecosistema de última milla experimentó cambios significativos. Inicialmente, se observó un crecimiento más conservador en la logística de última milla y el comercio electrónico en el 2023, con un aumento del 3% y un estancamiento del crecimiento del retail online (Peves, 2024).

La evolución del mercado, ilustrada en la Figura 1, ha tenido un impacto positivo en las empresas que se especializan en logística. Estas operadoras logísticas tienen como principal objetivo el transporte de mercancías desde el punto de origen, ya sea un centro de distribución comercial o un vendedor individual con un negocio de comercio electrónico, hasta el cliente final. Este proceso se realiza considerando la ubicación del destinatario y la fecha programada de entrega.

FIGURA 1. TAMAÑO DE MERCADO DE E-COMMERCE Y DE E-COMMERCE ATENDIDO POR DELIVERY



Fuente: Universidad ESAN 2020



En Perú, diversas empresas, como Moova y Scharff, se especializan en el servicio de última milla. Muchas de estas compañías optan por tercerizar su flota de operaciones debido a que, además de evitar los gastos asociados con el mantenimiento y cuidado de los vehículos, enfrentan una alta incertidumbre en la planificación. Esta incertidumbre incluye variables como la demanda fluctuante y la información variable sobre los pedidos que reciben a diario.

Para esta investigación se seleccionó la empresa Scharff, fundada en 1994 como representante de FedEx en Perú. En 2017, Scharff amplió su enfoque para incluir el sector de última milla. Según la revista SEMANA ECONÓMICA, la empresa experimentó un crecimiento superior al 50% en 2021, impulsado por la expansión del comercio electrónico y la demanda de servicios de última milla (Saavedra,2022). Al igual que muchas operadoras logísticas de última milla que tercerizan su flota, Scharff utiliza cuatro tipos de vehículos, que reserva a través de sus proveedores: motocicletas, automóviles, vans y furgones.

Como se mencionó anteriormente, la industria del comercio electrónico y la logística de última milla ha experimentado un crecimiento considerable en los últimos años. Este crecimiento ha llevado a que las empresas que tercerizan su flota de transporte enfrenten una competencia constante por la disponibilidad de vehículos, ya que no son las únicas que recurren a estos proveedores. Se han identificado un total de 72 proveedores disponibles en el mercado. No obstante, la empresa Scharff utiliza un formulario de registro de disponibilidad que se envía a los proveedores dos días antes de la operación. Este formulario permite a los proveedores informar sobre la cantidad de cada tipo de vehículo que tienen disponible para el servicio. Sin embargo, solo alrededor de 15 proveedores, en promedio, utilizan este formulario.

## 2. MÉTODO

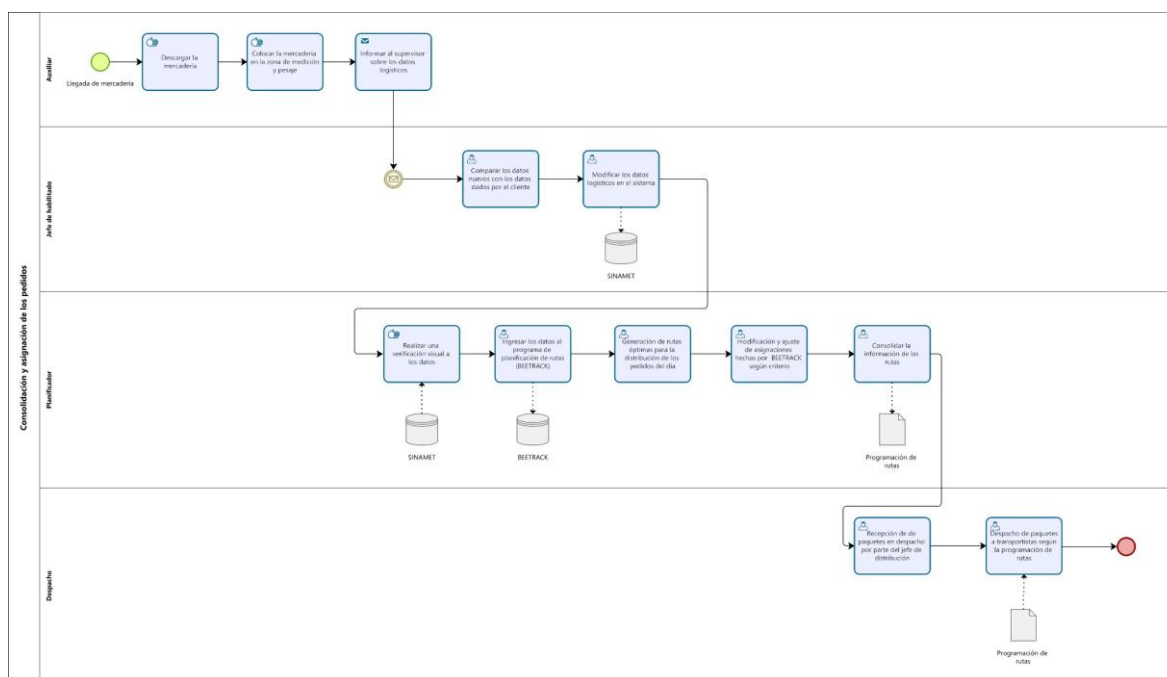
Dado que esta investigación se centra en un proyecto de análisis de datos con el objetivo de optimización, se empleó la metodología conocida como "Predictive Analytics in Supply Chain". Esta metodología consta de tres etapas principales: comprensión del negocio, análisis predictivo y optimización y planificación. A continuación, se describe detalladamente cada una de estas fases para abordar y resolver el problema en cuestión.



## 2.1 COMPRENSIÓN DEL NEGOCIO

La empresa Scharff opera con dos tipos principales de procesos. El primero es el proceso de recolección, en el cual se recoge la mercancía de los centros de distribución y se consolida en su almacén, donde se prepara para el despacho a los transportistas. Esta investigación se enfocará en el proceso de "consolidación y asignación de pedidos", que implica el dimensionamiento del volumen y peso de la mercancía, así como la selección de la flota de vehículos necesaria para realizar la distribución de manera eficiente, minimizando los costos logísticos. El flujo de este proceso se ilustra en la Figura 2.

FIGURA 2 – PROCESO DE PREPARACIÓN DE MERCADERÍA Y DESPACHO DE LOS PEDIDOS



FUENTE: Elaboración propia

### 2.1.1 Explicación del Proceso General

El proceso comienza con la llegada de los furgones, los cuales han recogido toda la mercadería pedida por los clientes en los centros de distribución de los clientes corporativos presentando las guías de remisión para el respectivo ingreso al almacén.

Seguidamente cada paquete pasa por el área de “Habilitado” donde realizan un proceso de dimensionamiento del paquete lo que consiste en la medición del paquete para conocer el largo,



ancho y profundidad en centímetros cúbicos, asimismo del peso en kilogramos. Luego el jefe del área procede a la subida de los datos recolectados en el sistema propio de la empresa el cual se comparte con todos los trabajadores habilitados para la visualización de estos datos.

Al cumplir con las mediciones y el registro de datos en el sistema, la persona encargada de la planificación de flota para realizar la distribución, consolida los datos que el jefe de habilitado había registrado anteriormente en el sistema. Al realizar la consolidación de los datos, utiliza un programa que crea las rutas creando grupos que contienen varios pedidos para la asignación a los vehículos necesarios para cumplir con la distribución. Sin embargo, el planificador realiza cambios manuales luego de que el programa haya terminado con el objetivo de acomodar los pedidos en los transportistas disponibles para el día dado. Esta actividad ocasiona un resultado negativo ya que no se tiene algún sustento matemático para respaldar el resultado final que se obtiene.

Finalmente, el supervisor de flota se encarga de contactar a los proveedores para la confirmación de los transportistas solicitados por tipo de vehículo con lo que al llegar al almacén se les brindará la guía de transporte con las rutas creadas.

### 2.1.2 Flujo de información

Al conocer el proceso general que hacen los trabajadores para la planificación de distribución de pedidos, para optimizar el flujo de información y la planificación de la distribución de pedidos en una empresa que terceriza su flota de transporte, es fundamental tener una base de datos correctamente estructurada. Cabe recalcar que en esta sección se creará el flujo de información para cualquier empresa que realice la misma actividad de distribución especificando las variables básicas que debería tener.

En primer lugar, se deben crear relaciones entre las tablas para definir los pedidos y sus atributos. Como se muestra en la figura 3, la tabla central representa los pedidos, donde la variable principal es la guía del pedido. Esta tabla también incluye las características geográficas del destino, como longitud y latitud. En el lado izquierdo, se encuentran los datos logísticos del pedido, como el peso (medido en kilogramos) y el volumen (medido en centímetros cúbicos), vinculados al número de la guía. En la parte derecha, se especifican los artículos agrupados por la guía.

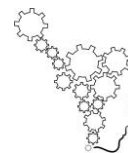
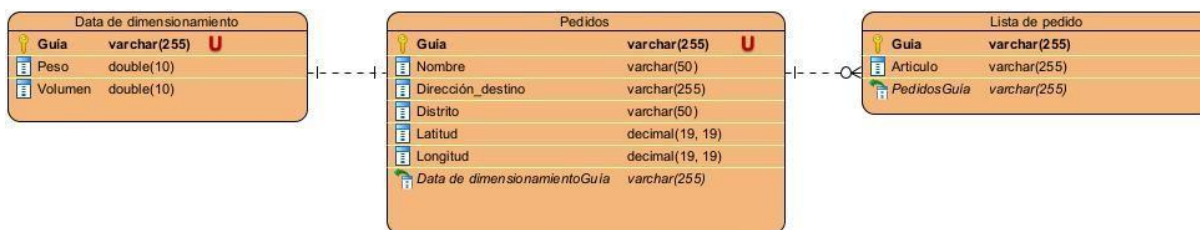


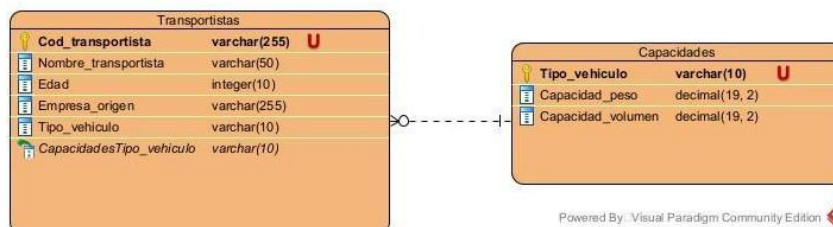
FIGURA 3 – FLUJO DE INFORMACIÓN DE LOS PEDIDOS



FUENTE: Elaboración propia

En segundo lugar, se deben relacionar las tablas de información de los transportistas para llevar un registro y poder asignarles los pedidos de acuerdo con sus restricciones. Como se muestra en la figura 4, en la tabla del lado izquierdo se registran los transportistas según el código que la empresa les asigna para una mejor diferenciación interna, así como datos básicos como nombre, edad, empresa a la que pertenecen y tipo de vehículo. Y en la tabla del lado derecho se registra el tipo de vehículo y sus respectivas capacidades de peso (medido en kilogramos) y volumen (medido en centímetros cúbicos).

FIGURA 4 – FLUJO DE INFORMACIÓN DE LOS TRANSPORTISTAS



Powered By: Visual Paradigm Community Edition

FUENTE: Elaboración propia

Finalmente, estas dos partes se relacionan al crear el reporte de distribución, donde se asignan los pedidos a los transportistas. Como se muestra en la figura 5, se ha generado una nueva tabla donde las variables principales son el número de la guía y el código del transportista, relacionando así todas las tablas existentes. Esta nueva tabla permite crear el reporte general de distribución con los pedidos asignados. La tabla se creó con el propósito de almacenar los resultados obtenidos luego del análisis en la base de datos.

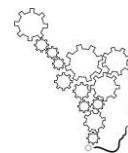
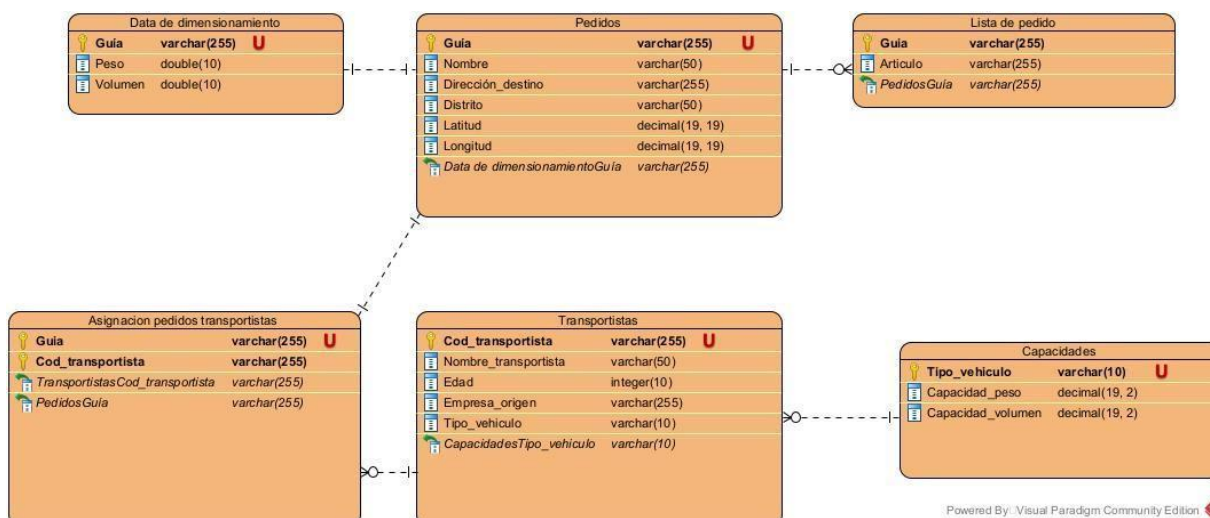


FIGURA 5 – FLUJO DE INFORMACIÓN PARA EL REPORTE DE DISTRIBUCIÓN



FUENTE: Elaboración propia

### 2.1.3 Caracterización de flota disponible

Como se mencionó anteriormente existen restricciones que los transportistas presentan al momento de asignarles algún pedido relacionado al tipo de vehículo que manejan. Ya que cada vehículo tiene una capacidad máxima de peso y volumen que puede transportar, asimismo en la empresa Scharff tiene una restricción adicional el cual es la cantidad máxima de pedidos que puede transportar que se agregaría a las capacidades. En la tabla 1, se puede observar la restricciones que tiene la flota dependiendo del tipo de vehículo.

TABLA 1 – RESTRICCIONES DE FLOTA

TIPO DE VEHÍCULO	CAPACIDAD PESO KG	CAPACIDAD VOLUMEN CM^3	CANTIDAD DE PEDIDOS
AUTO	500	2.000.000	55
FURGÓN	1727	11.100.000	22
MOTO	33	500.000	60
VAN	932	4.600.000	55

FUENTE: Elaboración propia



Por otro lado la empresa efectúa el pago a los transportistas dependiendo del tipo de vehículo que maneja representando un pago promedio por día de servicio dado a Scharff. Estos datos de pago promedio se calcularon mediante un promedio del pago realizado a los transportistas en los meses de enero, febrero y marzo. En la tabla 2 se puede observar la provisión promedio diaria por tipo de vehículo.

TABLA 2 – PROVISIÓN PROMEDIO POR DÍA POR TIPO DE VEHÍCULO	
TIPO DE VEHÍCULO	PROVISIÓN PROMEDIO DIARIA EN SOLES
AUTO	S/.78,70
FURGÓN	S/.204,50
MOTO	S/.70,50
VAN	S/.178,50

FUENTE: Elaboración propia

#### 2.1.4 Análisis descriptivo

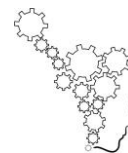
Se ha recolectado datos durante 15 meses, desde enero de 2023 hasta marzo de 2024. Con estos datos de distribución, se pudo determinar la cantidad real de pedidos que los transportistas realizan por día de operación. En la tabla 3 se muestra una comparación entre la cantidad real de pedidos que un transportista, según el tipo de vehículo, transporta por día y la cantidad ideal que debería transportar.

TABLA 3 – COMPARACIÓN ENTRE CANTIDAD DE PEDIDOS TRANSPORTADOS REAL VS IDEAL			
TIPO DE VEHÍCULO	CANTIDAD DE PEDIDOS REAL TRANSPORTADOS POR DÍA	CANTIDAD DE PEDIDOS IDEAL TRANSPORTADOS POR DÍA	DIFERENCIA %
AUTO	44,15	55	-20%
FURGÓN	56,31	22	156%
MOTO	17,03	60	-72%
VAN	45,80	55	-17%

FUENTE: Elaboración propia

Asimismo, es necesario calcular el costo unitario por pedido para determinar cuánto cuesta transportar un pedido por tipo de vehículo. Esto nos permitirá identificar oportunidades para reducir costos al aumentar la productividad de los transportistas por día. En la tabla 4 se puede





observar cuánto es el costo unitario al transportar un pedido por tipo de vehículo.

TABLA 4 – COSTO UNITARIO POR TIPO DE VEHÍCULO

TIPO DE VEHÍCULO	COSTO UNITARIO EN SOLES
AUTO	S/.1,78
FURGÓN	S/.3,63
MOTO	S/.4,14
VAN	S/.3,90

FUENTE: Elaboración propia

El uso de los vehículos para asignarles una cantidad de paquetes con diferentes características logísticas, se observó que no se utiliza a su totalidad la capacidad de cada tipo de vehículo como se puede observar en la tabla 5.

TABLA 5 – CAPACIDAD USADA POR TIPO DE TRANSPORTE

TIPO DE VEHÍCULO	CAPACIDAD UTILIZADA EN %
AUTO	29,2%
FURGÓN	24,5%
MOTO	34,0%
VAN	16,9%

FUENTE: Elaboración propia

Este análisis nos conlleva a la primera situación problemática, ya que al tener las capacidades usadas por tipo de vehículo se obtuvo un resultado de 26% como promedio de la capacidad usada en general lo cual indica que existe una deficiencia en la asignación de los pedidos y en la planificación de la cantidad de flota en general por tipo de vehículo.

Al realizar los análisis descriptivos anteriores, se identificó una deficiencia en la planificación de las unidades de transporte encargadas de realizar el servicio de última milla. Durante un período de 15 meses, se utilizó en promedio solo el 26% de la capacidad de cada tipo de vehículo, muy por debajo de la meta del 80% establecida por la empresa. Este hallazgo nos lleva a formular nuestras preguntas de investigación: ¿Es posible predecir de manera anticipada el tamaño de cada paquete y el total de paquetes a gestionar en un día dado? ¿Y es posible asignar de manera eficiente los vehículos, minimizando el costo total y respetando las restricciones de peso, volumen y la cantidad máxima de pedidos por transportista según el tipo de vehículo? Estas preguntas guiarán el desarrollo de nuestra investigación y el diseño de soluciones para mejorar la eficiencia en la planificación de la distribución.

Para abordar estos desafíos, el objetivo principal es proponer y evaluar una serie de pasos para mejorar la planificación anticipada de 8 días de las unidades logísticas, con el propósito de incrementar la capacidad promedio utilizada y disminuir el costo de flota. Para alcanzar este



objetivo principal, se han establecido los siguientes objetivos secundarios: aproximar un modelo estadístico de serie de tiempo que se ajuste a la demanda real con el menor margen de error posible; predecir con anticipación el tamaño de los paquetes mediante algoritmos de inteligencia artificial; mejorar la asignación de los paquetes a las unidades de transporte mediante algoritmos de programación lineal; y evaluar la solución propuesta utilizando como indicador la capacidad promedio utilizada por vehículo.

## 2.2 ANÁLISIS PREDICTIVO

La segunda fase de la metodología se enfoca en el análisis predictivo, utilizando los datos históricos de 15 meses recopilados anteriormente. Esta fase es crucial porque, con la problemática claramente definida en la sección anterior, se puede utilizar esta información para hacer predicciones precisas sobre futuras necesidades logísticas. Cada subsección dentro de esta fase presenta pasos específicos que la empresa debe considerar para mejorar su planificación de distribución. Estas acciones incluyen el uso de técnicas estadísticas y de inteligencia artificial para prever el tamaño y la cantidad de paquetes, así como la optimización de la asignación de recursos. El objetivo es mejorar la eficiencia en las operaciones logísticas y asegurar que la planificación de la distribución sea más precisa y efectiva.

### 2.2.1 Planeamiento de la demanda

Primero se tuvo que realizar un pronóstico de la demanda por lo que como primer paso se identifican los valores atípicos que afectan negativamente al momento de construir el modelo de pronóstico ya que dificulta el ajuste del modelo a los datos reales obteniendo mayor error al pronosticar; estos valores serán reemplazados con valores suavizados para no generar distorsión en la data histórica; al momento de tener una data limpia de valores atípicos se procedió a generar dos modelos de pronóstico, el primero se creó un modelo de redes neuronales recurrentes y el segundo se creó un modelo de redes neuronales recurrentes aplicado con el método de descenso gradiente para la minimización de error y se tuvo como indicador de elección el error RMSE y el error MAPE.



### 2.2.2 Segmentación de los distritos

Segundo, se aplicó el método de k-means con la intención de simplificar el tamaño del problema ya que se aplicó a los distritos de lima metropolitano los cuales son 43 distritos y fueron simplificados a 9 clusters tanto para los distritos que aparecen en recojo y entrega creando un total de 81 combinaciones entre distritos. Para poder llegar a la cantidad de 9 clusters se aplicó el método de “Gap-statistics” donde minimiza la variación entre clusters. Además esto ayuda al momento de desear crear modelos de predicción ya que al ser variables categóricas al crear un modelo se generará un modelo por cada elemento categórico.

### 2.2.3 Prueba de dependencia

Tercero, se realizó una prueba de correlación para identificar posibles dependencias entre las variables y determinar si estas deberían considerarse en los pasos siguientes para la creación del modelo de predicción. En esta situación, existen dos escenarios posibles: uno en el que algunas variables sean dependientes de otras, lo que causaría multicolinealidad. Esta dependencia puede complicar la interpretación de los coeficientes del modelo, ya que dificulta determinar la contribución individual de cada variable independiente. En caso de multicolinealidad, se procedería a eliminar una de las variables dependientes, seleccionando la que tenga menor impacto en las demás para reducir este problema. El segundo escenario es que todas las variables sean independientes, lo cual es el escenario ideal, ya que permitiría proceder directamente con la creación del modelo de predicción.

En esta investigación, se utilizó la prueba de correlación de Spearman, ya que los datos eran no paramétricos. Los resultados de la prueba mostraron que todas las variables eran independientes entre sí, lo cual facilitó el proceso de creación del modelo de predicción sin necesidad de ajustes adicionales por multicolinealidad.

### 2.2.4 Caracterización de las variables independientes

Después de realizar la prueba de dependencia en todas las variables de la base de datos consolidada, se presenta en la tabla 6 una lista de las variables utilizadas junto con su tipo correspondiente. Para las variables categóricas tipo "dummies," el algoritmo generó columnas



adicionales para cada categoría única, llenando sus filas con 0 y 1 según corresponda. Un caso especial fue la variable "cliente," donde se aplicó el principio de Pareto para seleccionar únicamente los clientes que representan el 80% del volumen total de operaciones. Los clientes restantes fueron agrupados bajo la categoría "OTRO," representando así el 20% restante de participación. Asimismo, Scharff utiliza "Puntos Scharff" (PS) como centros de consolidación de mercancías, donde los transportistas recogen productos para llevar al almacén ("Recojo PS") y entregan mercancías para que los clientes las recojan ("Entrega PS").

TABLA 6 – CARACTERIZACIÓN DE VARIABLES

VARIABLE	TIPO	POSIBLES ELEMENTOS
DÍA (LUNES A DOMINGO)	DUMMYS (7)	{0,1}
DÍA DEL MES	FACTOR	{1,31} / DEPENDIENDO DEL MES
MES	FACTOR	{ENERO, ..., DICIEMBRE}
RECOJO PS	BINARIO	{0,1}
ENTREGA PS	BINARIO	{0,1}
CLUSTER DISTRITO ORIGEN	FACTOR	{1, ..., 9}
CLUSTER DISTRITO DESTINO	FACTOR	{1, ..., 9}
CLIENTE	DUMMYS (10)	{0,1}
DISTANCIA KM	NUMÉRICO	R+

FUENTE: Elaboración propia

### 2.2.5 Modelamiento de variables “Peso” y “Volumen”

Finalmente, en la parte de predicción de datos logísticos, tanto de peso como de volumen, se utilizó un total de 885 mil observaciones y 24 variables independientes para crear los modelos de predicción. Se desarrollaron dos tipos de modelos: un modelo de árbol de decisión simple y un modelo de árboles de decisión con descenso de gradiente. Este último genera varias simulaciones del árbol de decisión, mejorando el error de predicción con cada árbol creado y evitando el sobreajuste, lo que impide que el modelo se adapte excesivamente a los datos de entrenamiento.

La totalidad de los datos se dividió en un 70% para entrenamiento y un 30% para prueba en ambos modelos. Adicionalmente, en el modelo de árboles de decisión con descenso de gradiente se utilizó una validación cruzada de 2 pliegues para determinar los mejores valores para los parámetros necesarios para la ejecución del modelo. Estos modelos se desarrollaron por separado para cada variable de peso y volumen. El criterio para elegir el mejor modelo para predecir cada variable de datos logísticos fue el error cuadrático medio (RMSE).



## 2.3 OPTIMIZACIÓN Y PLANIFICACIÓN

En esta última fase, se emplearon métodos estadísticos que ayudaron a cumplir el objetivo principal de la investigación. Estos métodos mejoraron las métricas de utilización vehicular, el costo total y el costo unitario por tipo de vehículo, además de asegurar la asignación de pedidos a los transportistas utilizando la nueva segmentación de distritos. A continuación se profundizará en cada subsección los pasos a seguir para optimizar la distribución.

### 2.3.1 Simulación de un reporte de distribución

Para la planificación, se creó un reporte simulado de la distribución, comenzando con la demanda pronosticada para estimar la cantidad de pedidos que la empresa Scharff tendría que cubrir en la semana siguiente. Una vez definido el tamaño del reporte de distribución, se generaron los datos de las variables aleatorias teniendo en cuenta la data histórica y la frecuencia de aparición de cada elemento, dado que las variables eran independientes entre sí. Se utilizó una gráfica de barras para hallar la probabilidad de aparición de un elemento en una variable específica, y con esa probabilidad se realizó la simulación.

Finalmente, se emplearon los modelos de predicción desarrollados previamente para determinar los datos logísticos de peso y volumen de cada pedido antes de su llegada al almacén. Esto permitió planificar la cantidad de flota necesaria para la operación de manera más precisa.

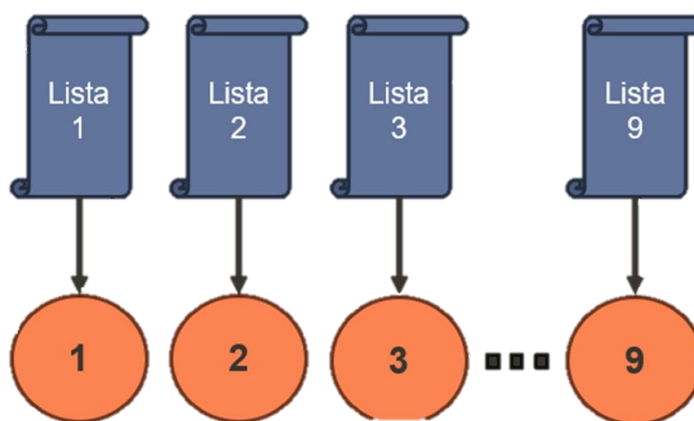
### 2.3.2 Optimización de asignación de paquetes

Para la parte de optimización, se utilizó un modelo de optimización lineal entera empleando el optimizador HiGHS, ya que es considerado un solver rápido y eficiente para problemas de programación lineal (LP) y ofrece fácil integración con varios lenguajes, incluyendo Julia y Python. En esta investigación, se crearon dos alternativas de solución para la asignación de transportistas. La primera alternativa no consideraba la disponibilidad real, sino que proporcionaba una lista amplia de transportistas clasificados por tipo de vehículo, como se



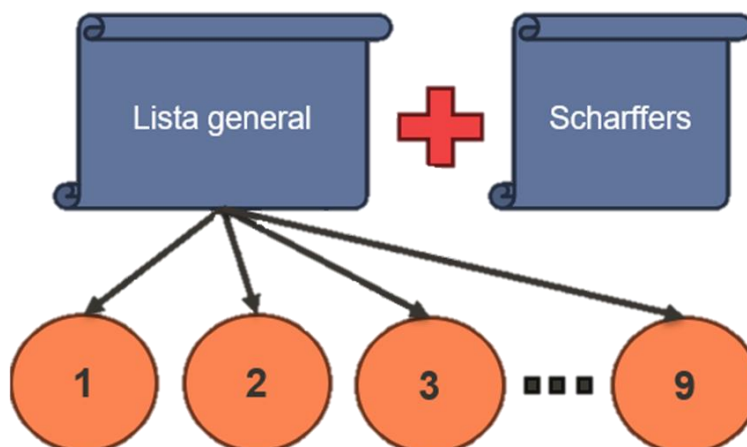
muestra en la figura 6, en este caso todas las listas son iguales. La segunda alternativa utilizaba una lista general de transportistas, en la que cada transportista representaba un tipo específico de vehículo, similar a la primera alternativa. Si todos los transportistas asignados no podían satisfacer la demanda de distribución en su totalidad, se añadían progresivamente transportistas adicionales denominados "Scharffers". Estos son autos de personas naturales que ofrecen el servicio a la empresa en días específicos según las necesidades de la empresa, como se observa en la figura 7.

FIGURA 6 – ALTERNATIVA 1: REPRESENTACIÓN DE LISTAS SEPARADAS DE TRANSPORTISTAS POR CLUSTER

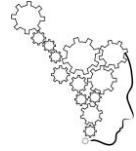


FUENTE: Elaboración propia

FIGURA 7 – ALTERNATIVA 2: REPRESENTACIÓN DE LISTA ÚNICA DE TRANSPORTISTAS POR CLUSTER



FUENTE: Elaboración propia



Las dos alternativas compartieron la misma función objetiva (1) la cual busca minimizar el costo total de transporte y de forma paralela existen 5 restricciones: la primera (2) la segunda (3) y la tercera (4) restringen las capacidades máxima que puede transportar un vehículo tanto como el peso, volumen y la cantidad máxima de pedidos; la cuarta (5) asegura que todos los pedidos estén asignados y la quinta (6) asegura que cada pedido sea asignado a un solo transportista.

$$\min z = \sum_{j=1}^{Cant. transportistas} (C_j * Y_j) \quad (1)$$

$$\sum_{i=1}^{Cant. Pedidos} (W_i * X_{ij}) \leq \underset{Peso}{W_{max}} [j] \forall j \in \{1, ..., Cant. Transportistas\} \quad (2)$$

$$\sum_{i=1}^{Cant. Pedidos} (V_i * X_{ij}) \leq \underset{Volumen}{V_{max}} [j] \forall j \in \{1, ..., Cant. Transportistas\} \quad (3)$$

$$\sum_{i=1}^{Cant. Pedidos} (X_{ij}) \leq N_j \forall j \in \{1, ..., Cant. Transportistas\} \quad (4)$$

$$X_{ij} \leq Y_j \forall X_{ij}, Y_j \in \{1, 0\} \quad (5)$$

$$\sum_{j=1}^{Cant. Transportistas} X_{ij} = 1 \forall i \in \{1, ..., Cant. Pedidos\} \quad (6)$$

Donde:

i = Item del pedido

j = Item del transportista

W<sub>i</sub> = Peso de cada item de pedido

V<sub>i</sub> = Volumen de cada item de pedido

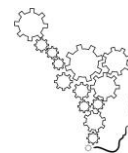
N<sub>j</sub> = Cantidad de pedidos máximos que puede llevar un transportista

C<sub>j</sub> = Costo al utilizar un tipo de vehículo

X = Asignación de pedido i a transportista j

*ij*

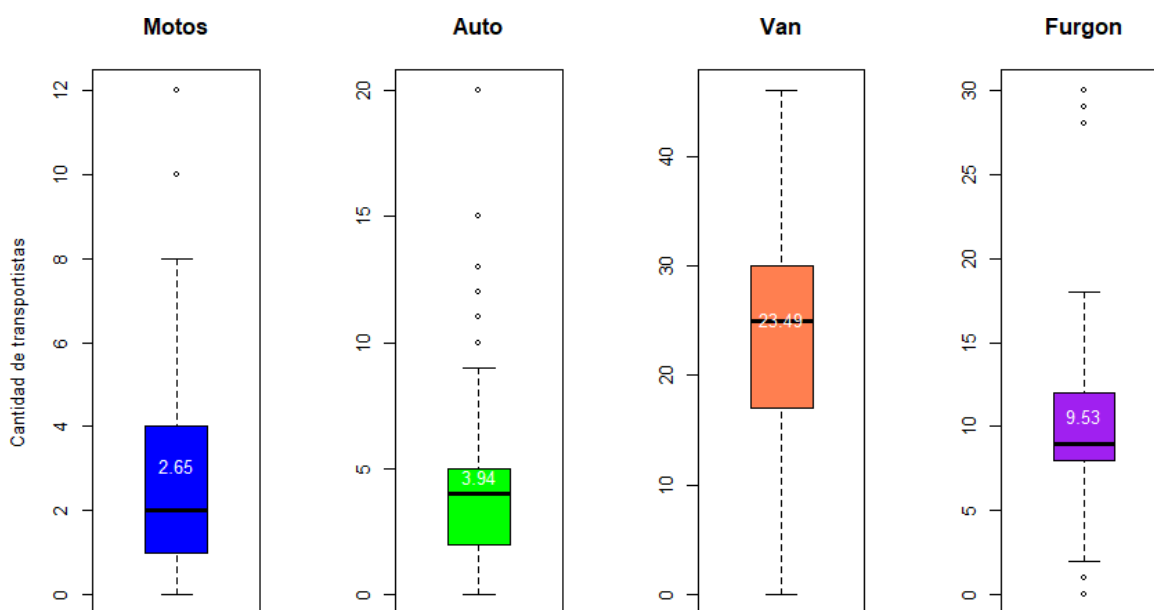
Y = Activación de transportista



### 2.3.3 Definir la disponibilidad de los transportistas

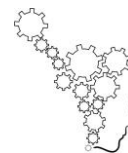
En esta investigación, la empresa Scharff solicita a sus proveedores completar un formulario con 2 días de anticipación para requerir sus servicios de flota de transporte. Desde noviembre de 2023 hasta mayo de 2024, se recopilaban un total de 197 observaciones mediante esta actividad. En el formulario, los proveedores registraban diariamente la cantidad de vehículos disponibles según el tipo. Con esta información, se creó un gráfico de caja (boxplot) para visualizar la concentración de los datos por tipo de vehículo. Esto permitió calcular la media, que se utiliza como el valor esperado para evaluar la planificación con días de anticipación. En la figura 8 se puede observar la distribución de las cantidades de transportistas por tipo de vehículo.

FIGURA 8 – GRÁFICA DE CAJAS DE LA DISTRIBUCIÓN DE CANTIDAD DE TRANSPORTISTAS DISPONIBLES POR TIPO DE VEHÍCULO



FUENTE: Elaboración propia





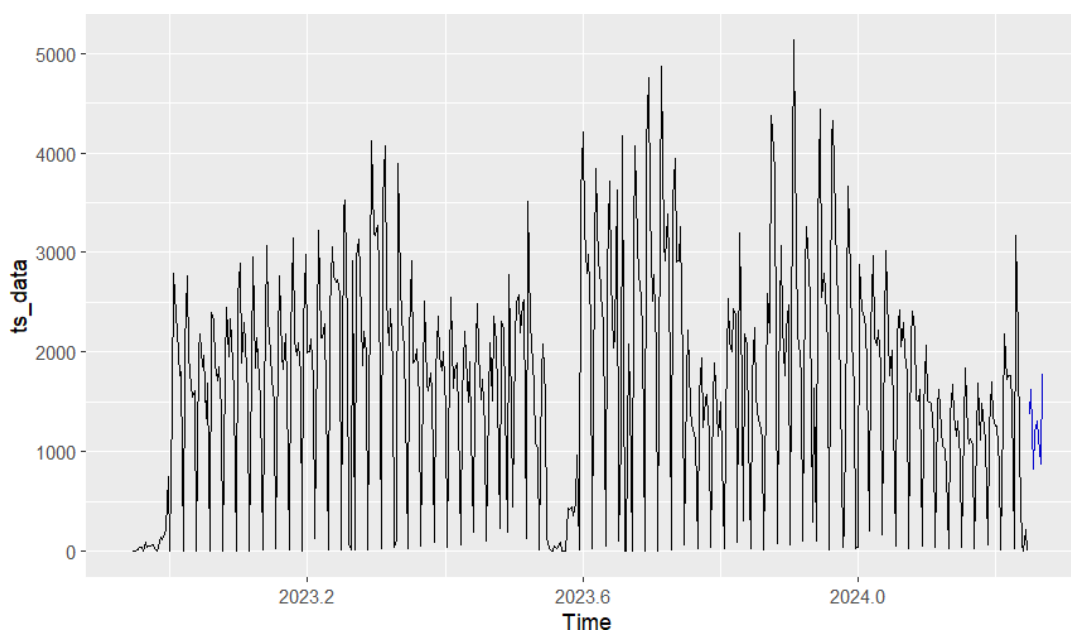
### 3. RESULTADOS

Primero, en la sección de planeamiento de la demanda realizada en la investigación, se han identificado 3 valores atípicos como se puede observar en el anexo 1.

Estos valores atípicos fueron de tipo “AO” el cual es un tipo de anomalía en una serie de tiempo que representa una observación que es atípicamente alta o baja en relación con los valores circundantes. Se le llama "aditivo" porque el valor anómalo se considera como una desviación adicional de la tendencia general de la serie de tiempo.

Al suavizar los datos atípicos se procedió a realizar el pronóstico de la demanda como se puede ver en la figura 9 la distribución de los datos históricos y los datos pronosticados.

FIGURA 9 – PRONÓSTICO DE LA DEMANDA UTILIZANDO REDES NEURONALES CON DESCENSO GRADIENTE

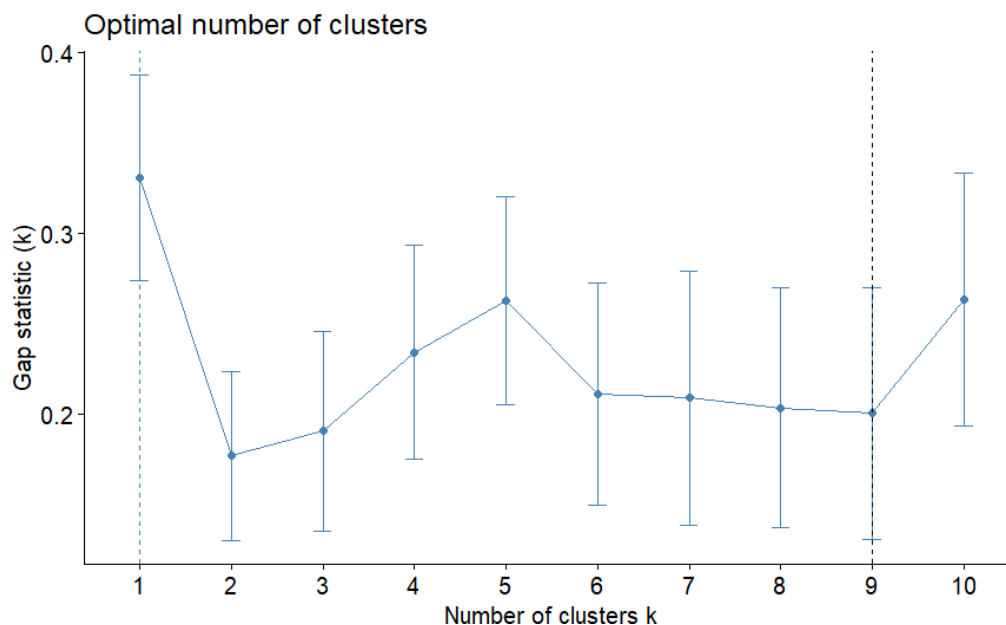


FUENTE: Elaboración propia

Segundo, en la parte de aplicación del método de k-means previamente se debe de identificar la cantidad necesaria de cluster, por lo que para este caso se aplicó el método “Gap-statistics” y se puede observar en la figura 10 la desviación en cada cluster.



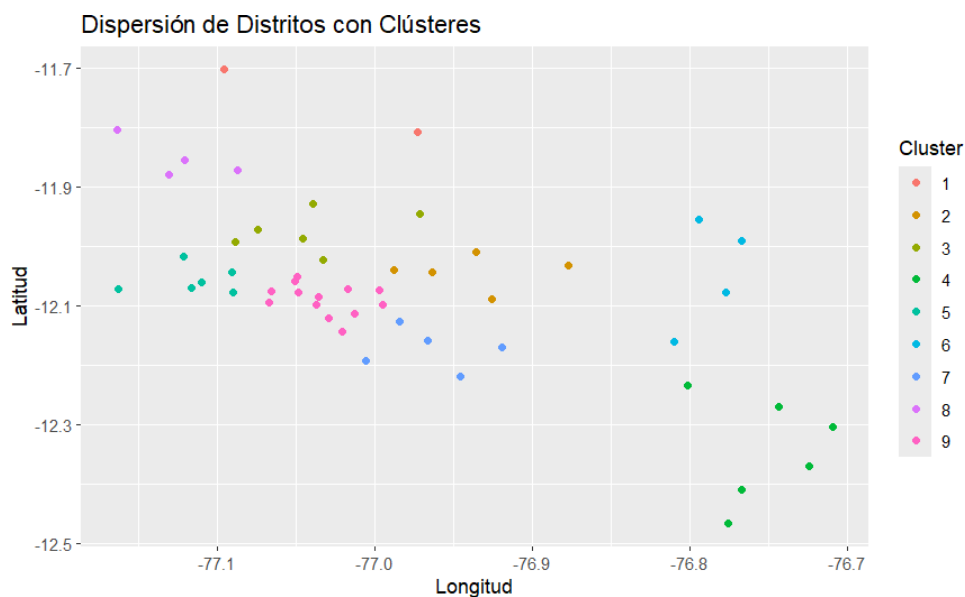
FIGURA 10 – GRÁFICA DE EVALUACIÓN DE CANTIDAD DE AGRUPACIONES CON GAP STATISTICS



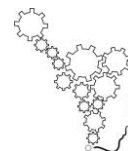
FUENTE: Elaboración propia

Seguidamente se aplicó el método de k-means con 9 clusters obteniendo la agrupación de los distritos utilizando sus coordenadas como se puede observar en la figura 11. Asimismo, se podrá observar la distribución de los distritos con el color característico de su cluster en el mapa de lima metropolitana en el anexo 2.

FIGURA 11 – GRÁFICA DE REPRESENTACIÓN DE AGRUPACIÓN DE DISTRITOS POR COORDENADAS



FUENTE: Elaboración propia



Tercero, en la parte de predicción de los datos logísticos, se utilizó el modelo de árboles de decisión simple y el modelo de árboles de decisión con descenso gradiente para este último al momento de definir los parámetros se procedió a utilizar el método de validación cruzada ya que es una técnica que se utiliza para validar la capacidad del modelo para generalizar a nuevos datos, y es especialmente útil cuando el modelo tiene pocos puntos de datos o cuando los datos están desequilibrados. En el anexo 3 se puede observar el código utilizado para la validación de parámetros.

Asimismo en la tabla 7 y la tabla 8, se puede observar la relevancia de cada variable independiente utilizada en el testeo del modelo para predecir el peso y el volumen respectivamente únicamente se colocaron las variables que juntas suman casi el 80% mientras que las demás variables se agruparon en la variable llamada “OTROS”.

TABLA 7 – IMPORTANCIA DE CADA VARIABLE EN LA VARIABLE PESO

VARIABLE	IMPORTANCIA EN %
CLUSTER DISTRITO ORIGEN	32,37%
CLIENTE N°1	24,32%
CLIENTE N°2	11,05%
DIA DE MES	10,15%
OTROS	22,11%

FUENTE: Elaboración propia

TABLA 8 – IMPORTANCIA DE CADA VARIABLE EN LA VARIABLE VOLUMEN

VARIABLE	IMPORTANCIA EN %
CLUSTER DISTRITO ORIGEN	19,65%
DÍA DE MES	16,33%
MES	12,81%
DISTANCIA EN KM	12,10%
CLIENTE N°1	11,64%
OTROS	27,47%

FUENTE: Elaboración propia

Finalmente, en la fase de optimización lineal se realizó una distribución de la flota de los diferentes tipos de vehículos a cada cluster para distribuir los pedidos que se les fue asignado a cada transportista como se puede observar en la tabla 9 teniendo un costo total de 4050.4 soles por ese día pronosticado con un total de 39 transportistas.



TABLA 9 – DISTRIBUCIÓN DE FLOTA DE VEHÍCULOS A LOS CLUSTERS UTILIZANDO MÉTODO OPTIMIZACIÓN LINEAL

CLUSTER	MOTO	AUTO	VAN	FURGÓN	TOTAL
1	1	0	0	0	1
2	1	3	0	0	4
3	2	3	2	0	7
4	0	1	0	0	1
5	0	3	0	0	3
6	0	1	0	0	1
7	1	3	0	0	4
8	0	1	0	0	1
9	4	5	4	4	17
TOTAL	9	20	6	4	39

FUENTE: Elaboración propia

Con la información de la disponibilidad esperada real, se creó un nuevo algoritmo de asignación de pedidos a los transportistas ya que se tomará como referencia 26 transportistas iniciales para todo los clusters (2 motos, 4 autos, 16 vanes y 4 furgones). Y cuando se termina la flota disponible que se colocó inicialmente, el algoritmo agrega automáticamente vehículos de tipo auto ya que la empresa Scharff puede contar con autos adicionales de personas naturales que prestan el servicio de transporte por un día en específico. Por lo que la distribución de los vehículos cambia a comparación del anterior como se puede observar en la tabla 10.

TABLA 10 – DISTRIBUCIÓN DE FLOTA DE VEHÍCULOS A LOS CLUSTERS UTILIZANDO MÉTODO OPTIMIZACIÓN LINEAL CON LA DISPONIBILIDAD REAL DE VEHÍCULOS

CLUSTER	MOTO	AUTO	VAN	FURGÓN	TOTAL
1	0	1	0	0	1
2	1	3	0	0	4
3	0	1	5	0	6
4	0	1	0	0	1
5	0	3	0	0	3
6	0	0	1	0	1
7	1	0	6	0	7
8	0	0	1	0	1
9	0	7	3	4	14
TOTAL	2	16	16	4	38

FUENTE: Elaboración propia

Finalmente en la parte financiera, se tuvo una reducción en el costo logístico total hasta 5074.2 soles al mejorar la asignación de los vehículos y utilizar con mejor porcentaje su capacidad. Al tomar como referencia la asignación con la disponibilidad real de los vehículos, en la tabla 11 se puede observar cómo ha cambiado la utilización de los vehículos por tipo.



Asimismo, en la tabla 12 se puede observar la evolución de la cantidad de pedidos en cada tipo de transporte comparándolo con el real y el ideal

TABLA 11 – MEJORA DE LA UTILIZACIÓN POR TIPO DE VEHÍCULO

TIPO DE VEHÍCULO	CAPACIDAD USADA EN %	CAPACIDAD USADA MEJORADA EN %	DIFERENCIA PUNTOS PORCENTUALES
AUTO	29,20%	43,10%	13,9
FURGÓN	24,50%	7,55%	-16,95
MOTO	34,00%	95,65%	61.65
VAN	16,90%	21,%	4,31

FUENTE: Elaboración propia

TABLA 12 – MEJORA DE LA CANTIDAD DE PEDIDOS ASIGNADOS POR TIPO DE VEHÍCULO

TIPO DE VEHÍCULO	CANTIDAD DE PEDIDOS REALIZADOS EN PROMEDIO MEJORADO	CANTIDAD DE PEDIDOS IDEALES A TRANSPORTAR	DIFERENCIA PORCENTUAL
AUTO	49,50	55	-10%
FURGÓN	21,75	22	-1%
MOTO	22,50	60	-63%
VAN	52,81	55	-4%

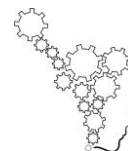
FUENTE: Elaboración propia

Asimismo, al mejorar la capacidad utilizada y aumentar la cantidad de pedidos que puede llevar cada tipo de vehículo, se observa una variación en el costo unitario. Esta optimización no solo incrementa la eficiencia operativa, sino que también impacta directamente en la reducción de costos por pedido. La comparación entre el costo unitario actual y el mejorado refleja esta diferencia, demostrando cómo una planificación más precisa y una mejor utilización de los recursos pueden llevar a una gestión de costos más efectiva. Esto se puede observar en la tabla 13.

TABLA 13 – MEJORA DEL COSTO UNITARIO POR TIPO DE VEHÍCULO

TIPO DE VEHÍCULO	COSTO UNITARIO INICIAL EN SOLES	COSTO UNITARIO MEJORADO EN SOLES	DIFERENCIA
AUTO	S/.1,78	S/.1,59	-0,19
FURGÓN	S/.3,63	S/.9,40	5,77
MOTO	S/.4,14	S/.3,13	-1,01
VAN	S/.3,90	S/.3,38	-0,52

FUENTE: Elaboración propia



#### 4. CONCLUSIÓN

Finalmente, como conclusión de la investigación, se logró cumplir con el objetivo principal mejorando el proceso de planificación al aumentar la capacidad promedio utilizada hasta un 43.1%, lo cual representa un incremento de 17.1 puntos porcentuales. Además, se obtuvo una reducción del 35.4% en el costo mensual. Se definió el mejor modelo para pronosticar la demanda utilizando redes neuronales con descenso de gradiente. Para la predicción de las variables de peso y volumen, se seleccionó el método de árbol de decisión con descenso de gradiente, empleando 150 árboles. También se mejoró la asignación de los pedidos a los transportistas, lo cual se reflejó en una distribución más equilibrada y eficiente de los recursos. Por último, se redujo la brecha en la cantidad de pedidos transportados por tipo de vehículo, optimizando así la eficiencia en la asignación de recursos.

#### AGRADECIMIENTOS

Queremos expresar nuestro más sincero agradecimiento a Dios, cuya guía y bendición nos han acompañado a lo largo de este camino. A nuestras familias, por su amor incondicional, apoyo constante y por ser pilar de fortaleza durante todo este proceso.

A los profesores, les agradecemos profundamente por compartir su sabiduría y conocimientos. Su dedicación y compromiso con la enseñanza han sido una inspiración constante.

A todos, los más sinceros agradecimientos.

#### RECONOCIMIENTOS

- Al Phd. Jose Antonio Larco Martinelli por sus valiosos aportes y asesoría en la realización de este trabajo de investigación.
- Al Phd. María Belén Ortiz García por sus valiosos aportes y asesoría en la realización de este trabajo de investigación.
- Al Ing. Alejandro Gallegos Chocce por sus valiosos aportes y asesoría en la realización de este trabajo de investigación.
- Al Br. en Ingeniería de Sistemas, Marcelo Toribio Ruiz, por sus valiosos aportes y asesoría en la realización de este trabajo de investigación.
- Al alumno Figo García Santos, por sus valiosos aportes en la realización de este trabajo de investigación.



## AUTORES

- Ahmad Daniel Hussen Jimenez: Elaboración y revisión del trabajo
- Fabiana Angelid Retamozo Torres: Elaboración y revisión del trabajo

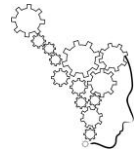
## REFERENCIAS

- ACROSS. **Peso Volumétrico: Su importancia en el transporte y logística.** Across Logistics | Logística, Almacen, Distribución y Transporte. Acesso em: 7 jul. 2024. Disponível em: <https://acrosslogistics.com/blog/peso-volumetrico-transporte-logistica>.
- ALMEYDA, E. **Pronóstico de la demanda internacional del banano orgánico de Perú usando algoritmos de Machine Learning.** Tese (Doutorado em Engenharia com menção em: Automatización, Control y Optimización de Procesos) — Universidad de Piura, Facultad de Ingeniería, Piura, Perú, 2022.
- APORIA. **A Comprehensive Guide to Mean Absolute Percentage Error (MAPE).** Acesso em: 8 jul. 2024. Disponível em: <https://www.aporia.com/learn/a-comprehensive-guide-to-mean-absolute-percentage-error-mape/>.
- AR RACKING. **¿Qué es la última milla y por qué es tan importante para la logística?** AR Racking | Sistemas de Almacenaje Industrial y Estanterías Industriales. Acesso em: 16 set. 2023. Disponível em: <https://www.ar-racking.com/pe/blog/que-es-la-ultima-milla-y-por-que-es-tan-importante-para-la-logistica/>.
- BEZANSON, J.; EDELMAN, A.; KARPINSKI, S.; SHAH, V. B. Julia: A Fresh Approach to Numerical Computing. **SIAM Review**, v. 59, n. 1, p. 65–98, jan. 2017. Acesso em: 10 jul. 2024. Disponível em: <https://doi.org/10.1137/141000671>.
- BOOKDOWN. **Arboles de decisión y Random Forest.** Home | Bookdown. Acesso em: 15 jul. 2024. Disponível em: <https://bookdown.org/content/2031/arboles-de-decision-parte-i.html>.
- CHOPRA, S.; MEINDL, P. **Supply Chain Management: Strategy, Planning, and Operation.** Pearson, 2015.
- CRAINIC, T. G.; LAPORTE, G. Planning models for freight transportation. *European Journal of Operational Research*, v. 97, n. 3, p. 409–438, mar. 1997. Acesso em: 10 jul. 2024. Disponível em: [https://doi.org/10.1016/s0377-2217\(96\)00298-6](https://doi.org/10.1016/s0377-2217(96)00298-6).
- ESAN. **¿Qué es el cross docking y para qué sirve?** Conexión ESAN. ESAN Graduate School of Business. Acesso em: 7 jul. 2024. Disponível em: <https://www.esan.edu.pe/conexion-esan/que-es-el-cross-docking-y-para-que-sirve>.
- FASTERCAPITAL. **Root mean square error: Understanding Forecast Accuracy using Root Mean Square Error.** Acesso em: 8 jul. 2024. Disponível em: <https://fastercapital.com/content/Root-mean-square-error--Understanding-Forecast-Accuracy-using-Root-Mean-Square-Error.html>.
- FRIEDMAN, J. H. Greedy Function Approximation: A Gradient Boosting Machine. **The Annals of Statistics**, v. 29, n. 5, p. 1189–1232, 2001. Disponível em: <http://www.jstor.org/stable/2699986>.
- HADAS, Y.; FIGLIOZZI, M. A. Modeling optimal drone fleet size considering stochastic demand. **EURO. Journal on Transportation and Logistics**, v. 13, p. 100127, 2024. Acesso em: 13



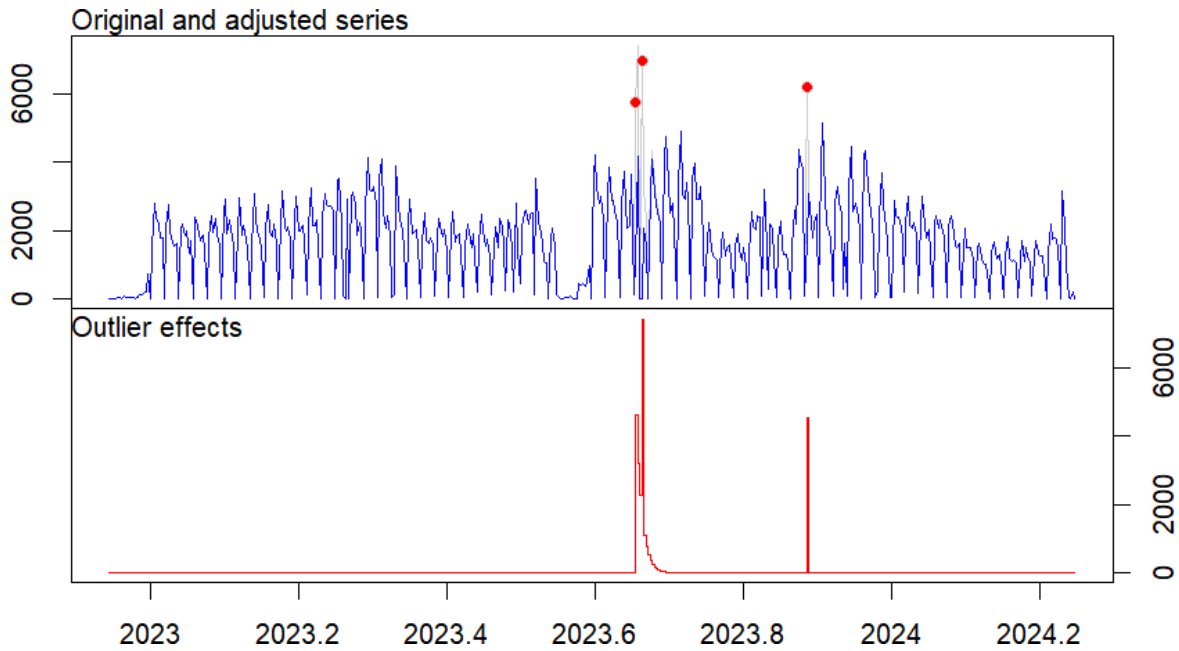
- mai. 2024. Disponível em: <https://doi.org/10.1016/j.ejtl.2024.100127>.
- INICIO. Acesso em: 15 jul. 2024. Disponível em: [https://observatorio.esan.edu.pe/wp-content/uploads/2023/03/SEII\\_informe-sectorial-logistica-ultima-milla.pdf](https://observatorio.esan.edu.pe/wp-content/uploads/2023/03/SEII_informe-sectorial-logistica-ultima-milla.pdf).
- IRANZO, P. D. **Análisis de outliers: un caso a estudio**. Tese de doutorado, Universidade de Valência, 2007. Acesso em: 23 jun. 2024. Disponível em: <http://hdl.handle.net/10803/9467>
- KANSAL, T.; BAHUGUNA, S.; SINGH, V.; CHOUDHURY, T. Customer segmentation using K-means clustering. In: **2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)**, Belgaum, India, 2018. p. 135-139. DOI: <10.1109/CTEMS.2018.8769171>.
- MCCARTHY, R. V.; MCCARTHY, M. M.; CECCUCCI, W.; HALAWI, L. **Applying predictive analytics**. Cham: Springer International Publishing, 2019. Acesso em: 7 mai. 2024. Disponível em: <https://doi.org/10.1007/978-3-030-14038-0>.
- MONNERAT, F.; DIAS, J.; ALVES, M. J. Fleet management: a vehicle and driver assignment model. **European Journal of Operational Research**, v. 278, n. 1, p. 64–75, out. 2019. Acesso em: 22 jun. 2024. Disponível em: <https://doi.org/10.1016/j.ejor.2019.03.021>.
- OBANDO SANDOVAL, A. D. **Análisis de redes neuronales recurrentes LSTM y modelos de vectores autorregresivos VAR en la generación de pronósticos de series de tiempo aplicado a las ventas de una empresa dedicada a la comercialización de productos de consumo masivo**. Projeto de Titulação — ESPOL, FIEC, 2023.
- PEVES, D. **Logística de última milla: ¿Cuánto mueve este sector y qué retos enfrenta?** Gestão. Acesso em: 18 abr. 2024. Disponível em: <https://gestion.pe/economia/empresas/logistica-de-ultima-milla-cuanto-mueve-el-sector-y-que-retos-enfrenta-e-commerce-comercio-electronico-consumo-olva-courier-noticia/>.
- RANIERI, L.; DIGIESI, S.; SILVESTRI, B.; ROCCOTELLI, M. A review of last mile logistics innovations in an externalities cost reduction vision. **Sustainability**, v. 10, n. 3, p. 782, mar. 2018. Acesso em: 23 out. 2023. Disponível em: <https://doi.org/10.3390/su10030782>.
- SCHARFF. **Transformando la logística en Perú y el mundo**. Acesso em: 15 jul. 2024. Disponível em: <https://www.holascharff.com/nosotros/>
- SCHOENHERR, T.; SPEIER-PERO, C. Data science, predictive analytics, and big data in supply chain management: current state and future potential. **Journal of Business Logistics**, v. 36, n. 1, p. 120–132, fev. 2015. Acesso em: 23 abr. 2024. Disponível em: <https://doi.org/10.1111/jbl.12082>.
- SEMANA ECONÓMICA. **Foco en Scharff, operador logístico que creció más del 50% en el 2021**. [Online]. Disponível em: <https://www.semanaeconomica.pe/sectores-empresas/industria/foco-en-scharff-operador-logistico-que-crecio-mas-del-50-el-2021>.
- SEO, A. **¿Qué es la última milla? Claves para la logística**. Euncet Business School; Blog de Euncet Business School, 27 dez. 2022. Disponível em: <https://blog.euncet.com/que-es-la-ultima-milla-en-logistica/>.
- SHEHADEH, K. S.; WANG, H.; ZHANG, P. Fleet sizing and allocation for on-demand last-mile transportation systems. *Transportation Research Part C: Emerging Technologies*, v. 132, p. 103387, nov. 2021. Acesso em: 23 abr. 2024. Disponível em: <https://doi.org/10.1016/j.trc.2021.103387>.
- VANDEPUT, N. **Data Science for Supply Chain Forecasting**. Gruyter, 2021. Acesso em: 8 jul. 2024. Disponível em: <https://doi.org/10.1515/9783110671124>.





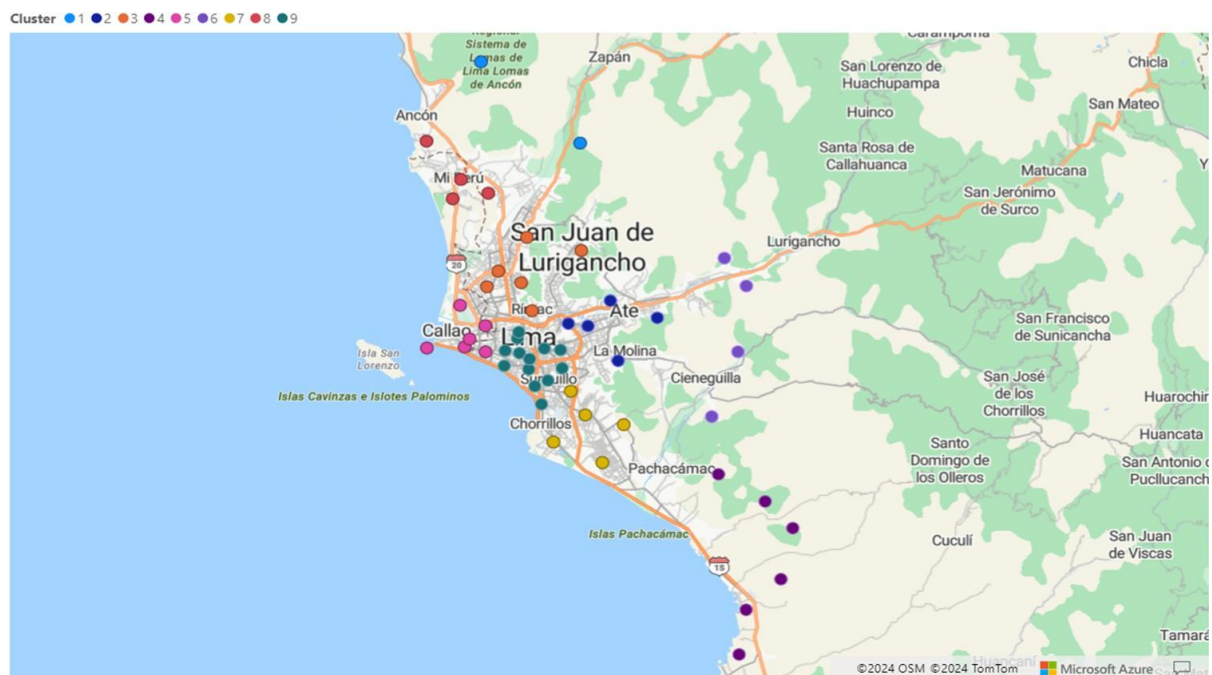
## ANEXOS

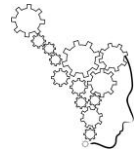
### ANEXO 1 – IDENTIFICACIÓN DE VALORES ATÍPICOS EN LA DATA HISTÓRICA



FUENTE: Elaboración propia

### ANEXO 2 – REPRESENTACIÓN DE LOS DISTRITOS CON SUS AGRUPACIONES EN LIMA METROPOLITANA





ANEXO 3 – CÓDIGO DE VALIDACIÓN DE PARÁMETROS PROBADOS

```
library(gbm)
library(caret)

# Definir los rangos de parámetros a explorar
grid <- expand.grid(n.trees = c(100, 150), # Número de árboles
                  interaction.depth = c(3, 5), # Profundidad de interacción
                  shrinkage = 0.1, # Tasa de aprendizaje (shrinkage)
                  n.minobsinnode = 20) # Número mínimo de observaciones en el nodo

# Configurar el control de entrenamiento para la validación cruzada
ctrl <- trainControl(method = "cv", # Validación cruzada
                    number = 2, # Número de folds para la validación cruzada
                    verboseIter = TRUE) # Mostrar detalles durante la búsqueda
```

FUENTE: Elaboración propia