

**ANÁLISIS DE LOS DIFERENTES FACTORES QUE PUEDEN INFLUIR
EN LA DEMANDA DE SISTEMAS DE SEGURIDAD Y DESARROLLO DE
ESTRATEGIAS DE CRECIMIENTO EMPRESARIAL EN LA
COMUNIDAD DE MADRID.**

ANALYSIS OF THE DIFFERENT FACTORS THAT MAY INFLUENCE THE
DEMAND FOR SECURITY SYSTEMS AND THE DEVELOPMENT OF BUSINESS
GROWTH STRATEGIES IN THE COMMUNITY OF MADRID.

Mireya Saiz Gómez
9101845@alumnos.ufv.es
Grado en Business Analytics
Curso académico 2021-2025
Tutor: María Jesús Gómez Fernández
Facultad de Derecho, Empresa y Gobierno

Dedicado principalmente a mis padres, por su esfuerzo y apoyo constante a lo largo de mi vida tanto personal como académica.

Resumen

En un mundo cada vez más afectado por la criminalidad, la percepción de inseguridad ha aumentado en zonas urbanas y residenciales, generando una creciente demanda de soluciones de protección. En este contexto, los sistemas de seguridad electrónica se han vuelto esenciales para hogares, negocios e industrias que buscan proteger sus bienes y asegurar la tranquilidad de sus ocupantes. Para empresas como Carlús Seguridad, comprender esta realidad es clave para mantenerse competitiva en un mercado dominado por grandes compañías. Para lograrlo, es fundamental apoyarse en el análisis de datos para identificar zonas con potencial de crecimiento, ajustar los servicios a las necesidades del cliente y optimizar el uso de recursos. En este sentido, el presente trabajo aporta información clave para la toma de decisiones estratégicas, basada en evidencias y orientada a la mejora continua de la oferta comercial y operativa de la empresa.

Palabras clave: Sistemas de seguridad; Alarmas; Circuito Cerrado de Televisión; Incendio; Modelos predictivos; Random Forest.

Abstract

In a world increasingly affected by crime, the perception of insecurity has risen in urban and residential areas, leading to a growing demand for protection solutions. In this context, electronic security systems have become essential for homes, businesses, and industries seeking to safeguard their assets and ensure peace of mind for their occupants. For companies like Carlús Seguridad, understanding this reality is key to staying competitive in a market dominated by large corporations. To achieve this, it is essential to rely on data analysis to identify areas with growth potential, tailor services to customer needs, and optimize resource use. In this regard, the present study provides key insights for strategic decision-making, based on evidence and aimed at the continuous improvement of the company's commercial and operational offerings.

Key words: Security Systems; Alarms; Closed-Circuit television; Fire; Predictive Models; Random Forest.

Agradecimientos

Quería agradecer a mis padres, por permitirme elegir la carrera y la universidad donde estudiar. Sin su apoyo incondicional, nada de esto habría sido posible. Su confianza y constante interés han sido mi mayor impulso para seguir adelante.

A mi hermano, a mis abuelos y en general a toda mi familia, gracias por ayudarme en este camino.

Y, por último, pero no menos importante, a mis amigos, con los que he reído y llorado a lo largo de estos años, gracias por estar siempre ahí.

ÍNDICE

Índice de Figuras	6
Índice de Tablas.....	6
1. OBJETIVOS	1
2. CUERPO DE LA MEMORIA	1
2.1. Introducción	1
2.2. Motivación	2
2.3. Estado del arte	3
2.4. Marco teórico.....	5
2.4.1. Modelos matemáticos.....	5
2.4.2. Métricas de evaluación	10
3. TRABAJO TÉCNICO.....	11
3.1. Herramientas empleadas	11
3.2. Ingeniería del dato	12
3.2.1. Origen de los datos.....	12
3.2.2. Características de las bases de datos	12
3.2.3. Estudio estadístico y descriptivo	17
3.3. Análisis del dato	26
3.3.1. Correlación entre variables	27
3.3.2. Modelos analíticos	29
3.3.3. Análisis de sobreajuste.....	33
3.4. Análisis del Negocio.....	36
3.4.1. Preguntas planteadas.....	36
3.4.2. Conclusiones.....	37
3.4.3. Recomendaciones	45
4. CONCLUSIONES GENERALES DEL TRABAJO.....	47
5. REFERENCIAS BIBLIOGRÁFICAS	48

Índice de Figuras

Ilustración 1 Distribución de ofertas y rendimiento de aceptación por grupo de sistema.....	18
Ilustración 2 Distribución de ofertas y rendimiento de aceptación por actividad	19
Ilustración 3 Distribución de los motivos de rechazo	21
Ilustración 4 Mapa de calor de las ofertas.....	21
Ilustración 5 Evolución ofertas aceptadas vs rechazadas	22
Ilustración 6 Distribución del importe por año	23
Ilustración 7 Distribución del importe según la demanda.....	24
Ilustración 8 Evolución de delitos por año.....	25
Ilustración 9 Evolución de habitantes por año	26
Ilustración 10 Matriz de correlación	29
Ilustración 11 Comparativa de modelos por R2.....	32
Ilustración 12 Comparativa de modelos por MAPE	32
Ilustración 13 Diagrama de dispersión de los modelos.....	33
Ilustración 14 Comparativas de modelos en train y test.....	34
Ilustración 15 Gráfico de radar del rendimiento global	35
Ilustración 16 Características más importantes - Random Forest	36
Ilustración 17 Mapa de calor: Frecuencia por Actividad y tipo de sistema	39
Ilustración 18 Tasa de aceptación por canal.....	41
Ilustración 19 Ofertas por canal y tipo de sistema	42
Ilustración 20 Ofertas por canal y CP	43
Ilustración 21 Ofertas por canal y CP (Madrid distrito).....	44

Índice de Tablas

Tabla 1 Número de ofertas por canal	40
---	----

1. OBJETIVOS

Para abordar este trabajo, se ha establecido un objetivo general y varios objetivos específicos que guían el desarrollo del análisis. A continuación, se detallan los objetivos propuestos.

Objetivo general:

- Identificar, a través del análisis de datos, cuáles son los factores más importantes que influyen en la aceptación de los servicios de seguridad.

Objetivo específico:

- Establecer las correlaciones entre las variables
- Construir un modelo predictivo para conocer la demanda de aceptación del producto
- Proponer campañas de captación adaptadas a las características y necesidades específicas de las diferentes zonas y tipos de clientes.
- Formular recomendaciones que permitan a la empresa impulsar su crecimiento empresarial.

2. CUERPO DE LA MEMORIA

2.1. Introducción

La seguridad se ha convertido en una necesidad fundamental en el funcionamiento de la sociedad moderna, especialmente en entornos urbanos densamente poblados y en zonas con alta actividad comercial. La evolución de la delincuencia y la creciente sofisticación de los métodos delictivos han impulsado una transformación profunda en el sector de la seguridad privada, donde la tecnología desempeña un papel cada vez más relevante.

En este contexto, el análisis del comportamiento de la demanda de servicios de seguridad, y su relación con variables geográficas, sociales y comerciales como la ocupación del territorio y la concentración de habitantes representa una oportunidad clave para que las empresas del sector refuercen su competitividad y posicionamiento estratégico.

Carlús Seguridad es una empresa con más de 40 años de trayectoria en el ámbito de la seguridad electrónica, especializada en la instalación y mantenimiento de sistemas de

vigilancia con cámaras. Su experiencia abarca soluciones para hogares, negocios e industrias, lo que la posiciona como un referente en el ámbito de la protección y prevención.

Este trabajo aborda, desde un enfoque analítico, las dinámicas del mercado de la empresa en la Comunidad de Madrid. A través del estudio de variables internas como el tipo de sistema ofertado, el perfil del cliente o el canal de captación, se busca generar conocimiento útil para la toma de decisiones estratégicas. Además, la incorporación de datos externos relacionados con la densidad demográfica y el nivel de ocupación permite comprender mejor la distribución espacial de la demanda, ofreciendo una base sólida para el diseño de acciones comerciales orientadas al crecimiento sostenido de la empresa en un entorno competitivo y en constante transformación.

2.2. Motivación

Carlús Seguridad opera en un mercado altamente competitivo, enfrenta como principales rivales a empresas como Securitas Direct y Prosegur, que cuentan con equipos especializados en el análisis de datos y el diseño de estrategias para captar más clientes. En comparación, es una empresa más pequeña con 23 empleados, lo que hace que la necesidad de optimizar sus procesos y mejorar su presencia sea aún más relevante.

El Trabajo de Fin de Grado propuesto tiene como objetivo realizar un análisis de los diferentes factores que influyen en la demanda de los servicios que ofrece Carlús Seguridad en las principales localidades de Madrid. A través de estrategias de negocio basadas en análisis de datos, se busca contribuir al crecimiento de la empresa. El enfoque estará en identificar qué zonas tienen mayor potencial de demanda, detectar problemas que puedan estar afectando su rendimiento en ciertas áreas y desarrollar propuestas de mejora para captar más clientes.

Para alcanzar estos objetivos, se llevará a cabo un análisis exhaustivo de los perfiles de clientes actuales y potenciales, utilizando herramientas y metodologías aprendidas durante el Grado en Business Analytics. El proyecto se centrará en detectar patrones de demanda, identificar áreas de mejora en el rendimiento de la empresa y diseñar soluciones prácticas que permitan a Carlús Seguridad posicionarse mejor en el mercado de sistemas de seguridad en Madrid. Además, se tendrán en cuenta estrategias que ayuden a competir

con empresas de mayor tamaño, aplicando un enfoque basado en datos y técnicas analíticas avanzadas.

2.3. Estado del arte

El riesgo de delito percibido es uno de los principales impulsores de la inversión en sistemas de seguridad electrónica en hogares y negocios, tal como señalan diversas investigaciones empíricas. La vulnerabilidad autoinformada casi duplica la probabilidad de poseer una alarma, según encuestas realizadas en Suiza y México (Killias & Clerici, 2000; Vilalta, 2012) (Killias & Clerici, 2000). Los estudios existentes sobre la hipótesis de seguridad, ampliamente citados, se basan en análisis a largo plazo de las tendencias de robos en Inglaterra y atribuyen la drástica disminución observada durante la década de 1990 al uso generalizado de alarmas y otros dispositivos de endurecimiento de objetivos (Farrell et al., 2011).

Incluso la obvia presencia de guardias privados en áreas públicas reduce el crimen oportunista, según evaluaciones aleatorias y cuasi-experimentales (Tseloni et al., 2017; Ariel et al., 2017). Las alarmas también disuaden a los ladrones. Simultáneamente, las pruebas de campo de la policía predictiva muestran que los algoritmos de aprendizaje automático son más precisos que las heurísticas tradicionales para identificar nuevos puntos críticos, lo que reduce el crimen cuando los recursos se despliegan adecuadamente (Mohler et al., 2015; Andrews & Mawson, 2022).

Estas técnicas analíticas están siendo utilizadas cada vez más por las empresas de seguridad. Los estudios de análisis de marketing demuestran que añadir señales exógenas, como los recuentos de delitos locales, a los modelos de demanda de aumento de gradiente mejora las previsiones de ventas en márgenes de dos dígitos sobre las líneas base lineales (Wedel & Kannan, 2016; Handler et al., 2022). Los algoritmos de agrupamiento y clasificación mejoran la puntuación de leads y personalizan las ofertas, aumentando la conversión y la retención, según la investigación de minería de datos sobre la gestión de relaciones con los clientes (Ngai et al., 2009; Rosário & Dias, 2023).

La adopción de tecnología moldea la competencia a nivel industrial: mientras que las pequeñas empresas confían en la confianza local y el servicio personalizado para abrirse camino en nichos, grandes multinacionales como Securitas Direct y Prosegur se diferencian con soluciones integradas para el hogar inteligente y publicidad digital agresiva (Blackstone et al., 2020; Galli-Debicella, 2021). Los protocolos de respuesta verificada, en

los que las empresas privadas filtran las alarmas antes del despacho policial, reducen las llamadas falsas y aumentan la efectividad percibida, según estudios de asociaciones público-privadas. Esto enfatiza la cooperación como una palanca estratégica (Blackstone et al., 2020). Juntos, estos resultados pintan un cuadro de un mercado donde la actual frontera competitiva de los servicios de seguridad electrónica está moldeada por una combinación de demanda impulsada por el miedo, disuasión tecnológica y análisis cada vez más complejos.

Una combinación de teorías criminológicas, económicas, de marketing y de gestión estratégica puede utilizarse para modelar la demanda de sistemas de seguridad. Mientras que la teoría de la actividad rutinaria enmarca las alarmas como "guardianes capaces" que aumentan el esfuerzo y el riesgo del delincuente, disminuyendo así la oportunidad y, por lo tanto, influyendo en la demanda futura, las perspectivas de elección racional afirman que los consumidores invierten en seguridad cuando la utilidad esperada de evitar el crimen supera el costo (Farrell et al., 2011; Tseloni et al., 2017).

Al afirmar que las características demográficas, la victimización pasada y el contexto del vecindario moderan la percepción del riesgo y, en consecuencia, la propensión a la compra, la investigación sobre la vulnerabilidad refuerza esta conexión (Killias & Clerici, 2000). El modelo de segmentación-targeting-posicionamiento sigue siendo fundamental en marketing, pero ahora se implementa a través de canales digitales ricos en datos. La teoría de la analítica sostiene que las decisiones basadas en evidencia empírica rinden mejor que las corazonadas, con la optimización de la tasa de conversión y el valor del tiempo de vida del cliente como métricas importantes (Wedel & Kannan, 2016; Rosário & Dias, 2023).

Mientras que la teoría de la difusión de la innovación sostiene que los primeros adoptantes de las tecnologías de seguridad inteligente aceleran la adopción en los vecindarios, la teoría del marketing relacional enfatiza que la confianza y el compromiso fomentan la lealtad, lo cual es esencial en los servicios de seguridad donde la monitorización continua sustenta el valor. La perspectiva basada en recursos de la estrategia de crecimiento argumenta que recursos raros y únicos, como una reputación local superior o capacidades propietarias de análisis del crimen, proporcionan una ventaja sostenible (Barney, 1991).

El marco de diferenciación de Porter sugiere que las pequeñas empresas se concentran en nichos de alto servicio en lugar de competir en costos con competidores orientados a la escala (Porter, 1985). Según estudios empíricos de PYMEs (Pequeñas y medianas empresas), los jugadores de nicho pueden crecer conectando micro-mercados relacionados y prosperar junto a gigantes globales utilizando plataformas digitales escalables y competencias centrales (Galli-Debicella, 2021; Zhang et al., 2019).

Las redes de proveedores confiables y la planificación basada en simulaciones también reducen las interrupciones del servicio en la instalación y el mantenimiento de alarmas, según la teoría del riesgo de la cadena de suministro (Wu & Olson, 2008). En conclusión, investigaciones recientes sobre métricas de alarma basadas en la utilidad sugieren incorporar ponderaciones de costo-beneficio en los desencadenantes predictivos para reducir los falsos positivos en la automatización del marketing. Esta estrategia combina la optimización basada en datos con los principios de la prevención criminológica (Handler et al., 2022). Para un proveedor de seguridad que opera en un entorno altamente competitivo y sensible al riesgo, este marco interdisciplinario explica cómo las dinámicas del crimen moldean la necesidad percibida, cómo el análisis permite un compromiso preciso con el mercado y cómo el despliegue estratégico de recursos guía una expansión sostenible.

2.4. Marco teórico

El marco teórico establece la base metodológica del trabajo, describiendo las técnicas y modelos de aprendizaje automático seleccionados por su capacidad predictiva, interpretabilidad y adaptabilidad. Se presentan sus fundamentos, funcionamiento y aplicaciones, junto con referencias que respaldan su uso. Además, se exponen las métricas de evaluación empleadas para comparar el desempeño de los modelos en el contexto estudiado.

2.4.1. Modelos matemáticos

La Regresión Lineal es un modelo estadístico utilizado para analizar la relación entre una variable dependiente y una o más variables independientes. Su origen se remonta al siglo XIX, siendo sistematizada por el científico Francis Galton y posteriormente formalizada matemáticamente por Karl Pearson y otros estadísticos (Galton, 1886; Pearson, 1901).

La regresión lineal simple representa el caso más básico de esta metodología, en la que se estudia la relación entre una variable dependiente Y y una única variable independiente X . El modelo asume que existe una relación lineal entre ambas variables, expresada mediante la siguiente ecuación:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

donde β_0 es la ordenada en el origen (intercepto), β_1 es la pendiente o coeficiente que representa el efecto de X sobre Y , y ε es un término de error aleatorio que recoge la variabilidad no explicada por el modelo.

La regresión lineal múltiple, empleada en el análisis, fue desarrollada como una extensión natural del modelo visto anteriormente, a medida que crecía la necesidad de analizar fenómenos en los que intervienen múltiples variables explicativas. Este modelo permite examinar simultáneamente la influencia de dos o más variables independientes x_1, x_2, \dots, x_k sobre una variable dependiente Y , mediante la siguiente formulación:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t$$

En este contexto, cada parámetro β_i representa el efecto individual de la variable X_i sobre Y , manteniendo constantes las demás variables del modelo. La estimación de los coeficientes se realiza comúnmente mediante el método de los mínimos cuadrados ordinarios (OLS, por sus siglas en inglés), buscando minimizar la suma de los errores al cuadrado entre los valores observados y los valores predichos por el modelo.

El algoritmo Random Forest es un modelo de aprendizaje automático no paramétrico basado en árboles de decisión, desarrollado por Leo Breiman y Adele Cutler a finales de la década de 1990. Su formalización se presentó en el artículo titulado “Random Forests”, publicado en la revista Machine Learning en el año 2001 (Breiman, 2001). Desde entonces, este algoritmo se ha consolidado como una de las técnicas más populares en el campo del aprendizaje supervisado, gracias a su capacidad para manejar conjuntos de datos grandes y complejos, su robustez frente al sobreajuste y su buen rendimiento en tareas tanto de clasificación como de regresión.

El modelo de Random Forest se fundamenta en la idea del ensamblado de modelos, es decir, en la combinación de múltiples modelos individuales en este caso, árboles de decisión, para mejorar la precisión global del sistema. A partir de un conjunto de datos de entrenamiento, el algoritmo genera un número determinado de árboles de decisión, cada

uno de los cuales es entrenado con un subconjunto aleatorio de las observaciones, obtenido mediante muestreo con reemplazo (bootstrap). Además, durante el proceso de construcción de cada árbol, en cada nodo de división solo se considera un subconjunto aleatorio de las variables predictoras, lo que introduce una variabilidad adicional que contribuye a disminuir la correlación entre árboles.

En su implementación clásica, cada árbol del conjunto emite un voto sobre la predicción final, y el resultado se determina por mayoría (en clasificación) o mediante el promedio de las predicciones (en regresión). Una de las claves del buen desempeño de Random Forest es que, si bien la aleatoriedad introducida puede aumentar ligeramente el sesgo del modelo, al reducir la varianza mediante la combinación de múltiples predictores se mejora la generalización, logrando una reducción efectiva del error total. Esta propiedad lo convierte en un modelo especialmente útil para conjuntos de datos con ruido o estructuras complejas no lineales.

El algoritmo Gradient Boosting es una técnica de aprendizaje automático basada en ensamblado de modelos, desarrollada inicialmente por Jerome H. Friedman en la década de 1990 y formalizada en su artículo “Greedy Function Approximation: A Gradient Boosting Machine” publicado en 2001 (Friedman, 2001). Desde entonces, el algoritmo ha evolucionado y dado lugar a implementaciones altamente eficientes que se explicaran a continuación como XGBoost, LightGBM y CatBoost.

Gradient Boosting pertenece a la familia de los métodos de boosting, que consisten en la construcción secuencial de modelos débiles (normalmente árboles de decisión) que corrigen los errores cometidos por los modelos anteriores. A diferencia de Random Forest, donde los árboles se entrenan en paralelo e independientemente unos de otros, en Gradient Boosting cada árbol se ajusta teniendo en cuenta los errores residuales del conjunto de árboles entrenados previamente.

Entre las principales ventajas de Gradient Boosting se encuentra su capacidad para capturar relaciones complejas no lineales, su alto rendimiento predictivo y su flexibilidad para utilizar distintas funciones de pérdida. Sin embargo, también presenta algunas desventajas importantes, como una mayor susceptibilidad al sobreajuste si no se regulan adecuadamente los hiperparámetros (como la tasa de aprendizaje, el número de árboles o la profundidad de cada árbol), y un mayor coste computacional comparado con algoritmos como Random Forest.

Extreme Gradient Boosting (XGBoost) es una implementación optimizada del algoritmo anterior, desarrollada por Tianqi Chen y presentada en 2016. Fue diseñado con el objetivo de mejorar el rendimiento computacional y la precisión del algoritmo de boosting tradicional, introduciendo mejoras a nivel de eficiencia, escalabilidad y regularización. Su desarrollo está documentado en el artículo “*XGBoost: A Scalable Tree Boosting System*” (Chen & Guestrin, 2016).

XGBoost se basa en el mismo principio que el algoritmo de Gradient Boosting, es decir, en la construcción secuencial de modelos débiles, cada uno de los cuales intenta corregir los errores cometidos por los anteriores. La predicción final se obtiene mediante la suma ponderada de todos los modelos generados.

Gracias a estas mejoras, XGBoost es eficaz, a la hora de manejar grandes volúmenes de datos con alta precisión. Entre sus principales ventajas se encuentran su velocidad, su flexibilidad en la elección de funciones de pérdida, su capacidad para manejar desequilibrios en los datos y su precisión superior en muchos casos. No obstante, su desventaja principal radica en la complejidad del modelo y la dificultad para interpretarlo, en comparación con modelos más simples como puede ser la regresión lineal.

Light Gradient Boosting Machine (LightGBM) es una implementación avanzada del algoritmo de Gradient Boosting, desarrollada por Microsoft Research y publicada en 2017. Fue diseñada específicamente para mejorar la eficiencia computacional y la escalabilidad de los modelos de boosting en contextos de grandes volúmenes de datos y alto número de características (Ke et al., 2017).

Al igual que XGBoost, LightGBM entrena árboles de decisión de forma secuencial, donde cada árbol intenta corregir los errores del conjunto anterior. Sin embargo, LightGBM introduce dos innovaciones técnicas clave:

- Crecimiento leaf-wise: A diferencia de otros algoritmos como XGBoost, que expanden los árboles de forma equilibrada, LightGBM lo hace expandiendo primero la hoja con mayor error, lo que mejora la precisión, aunque puede aumentar el riesgo de sobreajuste si no se regula bien.
- Uso de histogramas: LightGBM convierte las variables continuas en grupos (bins), lo que reduce el tiempo de cálculo y el uso de memoria.

Gracias a estas mejoras, LightGBM es considerablemente más rápido que XGBoost en conjuntos de datos grandes, manteniendo una precisión similar o incluso superior en algunos casos. Sin embargo, esta eficiencia viene acompañada de una mayor sensibilidad al ajuste de hiperparámetros y a veces de un comportamiento menos estable con datos pequeños o con valores atípicos.

LightGBM también ofrece un manejo eficiente de variables categóricas, lo cual representa una ventaja relevante frente a otras implementaciones que requieren codificación previa.

Support Vector Regression (SVR) es una extensión del algoritmo Support Vector Machine (SVM), desarrollado por Vladimir Vapnik en los años 90, que se adapta a tareas de regresión. A diferencia de la regresión lineal tradicional, SVR busca encontrar una función que se aleje lo menos posible de los datos reales, pero permitiendo un margen de tolerancia ϵ , dentro del cual los errores no se penalizan.

Su objetivo es encontrar una línea (o superficie) que tenga como máximo una desviación ϵ respecto a los valores reales, y al mismo tiempo sea lo más plana posible. Para ello, SVR minimiza una función de coste que combina la complejidad del modelo (medida por la norma del vector de pesos) y las penalizaciones por errores fuera del margen.

Gracias al uso de kernels, SVR puede capturar relaciones no lineales transformando los datos a espacios de mayor dimensión, manteniendo un buen rendimiento incluso con estructuras complejas. No obstante, su rendimiento puede verse afectado por la selección de hiperparámetros como C , ϵ y el tipo de kernel, y suele ser más sensible a escalas distintas entre variables.

Por último, el modelo K-Nearest Neighbors (KNN) es un algoritmo de aprendizaje supervisado basado en la similitud entre los datos, que se utiliza tanto en tareas de clasificación como de regresión. Fue propuesto por el matemático Evelyn Fix y el estadístico Joseph Hodges en 1951, aunque su desarrollo y popularización en el campo de la inteligencia artificial y el aprendizaje automático ocurrió posteriormente. El modelo KNN destaca por su simplicidad, su capacidad para manejar tanto problemas lineales como no lineales y su enfoque intuitivo para la toma de decisiones.

El principio fundamental del KNN es que la clasificación de una muestra se determina a partir de las etiquetas de sus vecinos más cercanos en el espacio de

características. En su versión básica, el algoritmo asigna a un punto de prueba la clase que sea más frecuente entre sus K vecinos más cercanos, donde K es un parámetro que debe ser especificado previamente por el usuario. Estos vecinos se seleccionan calculando las distancias entre el punto de prueba y los puntos de entrenamiento.

2.4.2. Métricas de evaluación

El coeficiente de determinación, comúnmente conocido como R^2 , es una métrica estadística utilizada para evaluar la calidad de los modelos de regresión. Representa la proporción de la varianza total de la variable dependiente que puede ser explicada por el modelo a partir de las variables independientes. Su valor varía entre 0 y 1, donde un valor de 1 indica un ajuste perfecto, es decir, el modelo explica la totalidad de la variabilidad observada, mientras que un valor cercano a 0 sugiere que el modelo apenas aporta información útil respecto a la media. R^2 se calcula a partir de la relación entre la suma de los cuadrados del error (SSE) y la suma total de cuadrados (SST).

Esta métrica se utiliza ampliamente porque permite comparar distintos modelos de regresión entre sí y ofrece una idea clara del grado de ajuste. Su valor no indica si las predicciones individuales son precisas, por lo que se suele complementar con métricas de error como las que se explican a continuación.

El Error Cuadrático Medio Raíz (RMSE) es una métrica ampliamente utilizada para medir la precisión de los modelos de regresión. Calcula la raíz cuadrada del promedio de los cuadrados de las diferencias entre los valores predichos por el modelo y los valores reales observados. Esta métrica proporciona una medida del error promedio en las mismas unidades que la variable objetivo, facilitando su interpretación directa.

El RMSE penaliza de forma más severa los errores grandes debido al uso del cuadrado en las diferencias, lo que lo hace especialmente útil cuando se desea evitar predicciones con desviaciones significativas.

El Error Porcentual Absoluto Medio (MAPE) mide el error medio entre los valores reales y predichos, expresado como porcentaje del valor real. Es útil para interpretar la precisión del modelo de forma relativa, especialmente cuando los datos tienen distintas escalas.

Una de las principales ventajas del MAPE es su fácil interpretación: indica el error medio como un porcentaje del valor real, lo cual es intuitivo para los usuarios no técnicos.

Y por último el Error Absoluto Medio (MAE) es una métrica que calcula el promedio de las diferencias absolutas entre los valores reales y las predicciones del modelo. A diferencia del RMSE, el MAE no penaliza los errores grandes de manera más severa, lo que lo hace menos sensible a valores atípicos.

El MAE proporciona una medida directa y fácil de interpretar del error medio en las mismas unidades que la variable objetivo, siendo útil para evaluar la precisión global del modelo.

La combinación de las métricas seleccionadas proporciona una visión equilibrada entre precisión, interpretabilidad y robustez frente a distintos tipos de errores.

3. TRABAJO TÉCNICO

3.1. Herramientas empleadas

A lo largo del desarrollo del trabajo se han utilizado tres herramientas principales: Excel, Python (a través de Google Colab) y una herramienta complementaria basada en inteligencia artificial para el apoyo en tareas de codificación y documentación.

Excel ha sido fundamental en las fases iniciales del análisis por su facilidad de uso, interfaz intuitiva y amplia funcionalidad para la manipulación rápida de datos. Esta herramienta ha permitido explorar, filtrar y organizar grandes volúmenes de información de manera eficiente, lo que facilitó la detección de errores, la limpieza de datos y la realización de análisis descriptivos básicos. Asimismo, se ha utilizado para generar gráficos preliminares y obtener una visión general del comportamiento de las variables, lo cual ha sido clave antes de abordar análisis más avanzados.

Por su parte, Python, mediante el entorno Google Colaboratory, ha desempeñado un papel central en el desarrollo técnico del proyecto. Esta plataforma basada en la nube permite ejecutar código en el navegador sin necesidad de configuración previa y con acceso a capacidades de procesamiento acelerado (GPU/TPU). Se ha optado por Python debido a su flexibilidad, potencia y alto grado de integración con otras tecnologías, así como por la amplia disponibilidad de librerías especializadas orientadas al análisis de datos y series temporales. Entre ellas, destacan:

- Pandas, para la manipulación estructurada de datos;
- NumPy, para operaciones numéricas eficientes;
- Matplotlib y Seaborn, para la creación de visualizaciones estáticas e interactivas;

Aunque existen otras herramientas potentes como MATLAB, estas presentan ciertas limitaciones en comparación con Python, especialmente en términos de escalabilidad, flexibilidad, automatización y compatibilidad con tecnologías actuales. Por ello, se ha preferido Python para ejecutar tareas complejas y adaptadas a las necesidades específicas del análisis.

Como apoyo adicional, se ha empleado inteligencia artificial para agilizar la escritura, depuración y mejora del código. Esta asistencia ha contribuido a resolver dudas técnicas puntuales, optimizar scripts y mantener una estructura eficiente en el desarrollo, elevando así la calidad del análisis.

3.2. Ingeniería del dato

3.2.1. Origen de los datos

Para la construcción del conjunto de datos utilizado en este trabajo, se han combinado diferentes bases de datos. En primer lugar, se ha incorporado una base de datos interna con ofertas proporcionado por la empresa Carlús Seguridad. En segundo lugar, se han obtenido varias bases de datos externos, uno con el número de habitantes por municipio a través del Instituto Nacional de Estadística, y otro conjunto de datos con el número de robos por municipio procedente del Ministerio del Interior. Todo ello ha permitido conformar una base de datos coherente y estructurada para el análisis.

3.2.2. Características de las bases de datos

Para mejorar la calidad del análisis, se llevaron a cabo procesos de limpieza y transformación de datos, lo que permitió estandarizar las variables, eliminar redundancias y seleccionar las más relevantes para el estudio. La base de datos de ofertas consta de 5397 registros y 7 atributos. Las ofertas registradas tienen una frecuencia anual y abarcan el período 2017-2024. Inicialmente, la base de datos contenía 18,688 registros y 27 atributos. Sin embargo, la mayoría de estos atributos no eran relevantes para el análisis, por lo que se eliminaron muchos de ellos.

Con el objetivo de cumplir con la normativa vigente en materia de protección de datos personales, especialmente el Reglamento General de Protección de Datos (RGPD), se ha procedido a la eliminación de toda información que pudiera permitir la identificación directa o indirecta de personas físicas o jurídicas. En particular, se han eliminado las columnas correspondientes a la calle, el identificador fiscal y los apellidos. Esta medida garantiza la anonimización de los registros y permite un análisis respetuoso con la privacidad de los datos tratados.

Se procedió a eliminar diversas variables del conjunto de datos al no aportar información relevante para el análisis. En primer lugar, “nº de cliente” y “nº de oferta” fueron descartadas por tratarse de identificadores sin utilidad analítica. Asimismo, las variables “visitado por”, “nº de secretario” y “vendedor” se eliminaron por referirse únicamente al comercial responsable de la operación, dato irrelevante en este contexto.

También se excluyeron “descripción del objeto” y “matchcode”, ya que no presentaban valores repetidos que permitieran establecer relaciones significativas. La variable “estado” fue eliminada al no contener datos, y “status cliente” se descartó en favor de “de parte de”, que permite un análisis más claro entre clientes nuevos y recurrentes.

Además, se eliminaron variables como “fecha negativa”, “próxima llamada”, “estadística” y “variación”, por no aportar valor directo al objetivo del estudio, centrado exclusivamente en el análisis de ofertas aceptadas y rechazadas.

En una primera fase, se aplicó un filtro en Excel sobre la variable "motivo de rechazo", dado que la base de datos contenía múltiples variaciones de una misma oferta. Se decidió conservar únicamente la última oferta entregada. Esta variable también presentaba 1838 valores nulos, en su mayoría correspondientes a ofertas aceptadas, que por tanto carecían de motivo de rechazo. Para los registros restantes, se optó por una imputación categórica mediante etiquetas, asegurando que todas las ofertas quedaran correctamente clasificadas.

Dado que la empresa opera en diversas provincias, se aplicó un segundo filtro para excluir todos los registros cuyo código postal no comenzara por 28, asegurando así que solo se incluyeran datos correspondientes a la Comunidad de Madrid. Adicionalmente, en la variable "tipo de oferta" se identificaron filas que no correspondían a ofertas propiamente dichas, sino a sustituciones o reparaciones de sistemas existentes. Tras aplicar

un filtro para eliminar estos registros, la variable fue descartada, ya que todos los valores restantes eran homogéneos.

A continuación, se exponen los atributos categóricos que se han conservado en el análisis por su relevancia;

La variable "tipo de sistema" especifica el tipo de sistema ofertado. Inicialmente, la base de datos contenía numerosos tipos de alarmas, la mayoría de los cuales aparecían en menos de cinco ofertas, lo que dificultaba su análisis individual. Por ello, se solicitó a la empresa una clasificación agrupada de estos sistemas, resultando en una nueva columna denominada "alarma grupo", que agrupa las ofertas en cuatro categorías: alarmas para detección de intrusiones, CCTV para vigilancia mediante cámaras, incendios para sistemas de alerta ante humo o fuego, y otros, que incluye sistemas menos comunes.

El atributo "de parte de" indica el canal por el cual el cliente ha conocido la empresa, siendo un atributo clave para diseñar estrategias de marketing y optimizar los canales de captación. Los tipos de canal incluyen "ADEMCO", cuando un proveedor de material recomienda la empresa a sus clientes; "carteles-coche", si el reconocimiento proviene de la publicidad en vehículos o en carteles de alarmas visibles en viviendas y negocios; "cliente", cuando la oferta proviene de un cliente actual el cual quiere instalar otro servicio; "D.P.Cliente", en caso de que un cliente recomiende la empresa a alguien que aún no lo es; "colaborador", si un profesional externo, como un administrador de fincas o un gestor, sugiere la empresa a cambio de una comisión; "internet", cuando el cliente conoce la empresa a través de publicidad digital; "comerciales", si el contacto se establece mediante el equipo de ventas; y "otros", cuando el origen del cliente no encaja en ninguna de las categorías anteriores.

La variable "motivo rechazo" recoge la causa por la cual una oferta no fue aceptada, proporcionando información valiosa para detectar patrones y ajustar la estrategia comercial. Entre los motivos registrados se encuentran: competencia, cuando el cliente elige una oferta de otra empresa; compromiso, si ya tiene un acuerdo previo con otra compañía; concurso, en casos sujetos a procesos de licitación; contrato competencia, cuando el cliente mantiene un contrato vigente que le impide cambiar de proveedor; más adelante, si decide posponer su decisión; más rápidos, cuando otra empresa actúa con mayor rapidez; precio, si considera que la oferta es demasiado cara; y producto, cuando el servicio o producto ofrecido no cumple con sus necesidades o expectativas.

Finalmente, la variable "actividad" identifica el tipo de cliente según el uso o la naturaleza del espacio donde se solicita el servicio, lo que permite segmentar adecuadamente el mercado y ajustar las estrategias comerciales. Entre las categorías incluidas se encuentran: adosado, para viviendas unifamiliares adosadas; andamio, en casos vinculados a estructuras temporales en obras; apuestas, cuando el servicio se requiere en locales de juego; campa, para terrenos abiertos utilizados como almacén o aparcamiento; chalet, en referencia a viviendas unifamiliares independientes; comercio y comercio.gr (comercio grande), según el tamaño del negocio; comunidad, para comunidades de vecinos; concurso, cuando la contratación está sujeta a procesos de licitación pública; estanco, farmacia, gasolinera, herbolario, hostelería, hoteles, joyería y loterías, cada uno asociado a su sector específico. También se incluyen nave (uso industrial o logístico), oficinas, org.ofic (organismos oficiales), piso (vivienda en edificios residenciales) y otros, para aquellos casos no contemplados en las categorías anteriores.

A continuación, se exponen los atributos numéricos que se han conservado en el análisis por su relevancia;

"Demanda" muestra si la oferta ha sido aceptada (1) o no (0), permitiendo calcular la tasa de éxito de las ofertas y su evolución a lo largo del tiempo.

La variable "año" indica el año en la que se generó la oferta y es clave para realizar análisis comparativos anuales con otros conjuntos de datos.

El "importe" refleja el valor de la oferta. En los casos donde el importe es 0, la oferta no tiene coste directo para el cliente, que solo pagaría las cuotas de mantenimiento correspondientes. Las ofertas rechazadas tienen importe negativo al ser un coste de oportunidad que pierde la empresa. Esta variable permite analizar la relación entre el precio y la demanda.

Y el "código postal" indica la ubicación del establecimiento, lo que resulta fundamental para realizar comparaciones con otros conjuntos de datos y analizar la distribución geográfica de las ofertas.

La base de datos de delitos abarca 36 municipios de Madrid y contiene información sobre el número de robos con violencia e intimidación, robos con fuerza en domicilios, establecimientos y otras instalaciones, robos con fuerza exclusivamente en domicilios y hurtos, en el período comprendido entre 2017 y 2024. Todas las columnas de esta base de datos contienen valores numéricos enteros, lo que permite realizar análisis cuantitativos

precisos. El número de robos en establecimientos, domicilios y hurtos es un dato muy importante que permitirá correlacionar los datos y evaluar la relación existente entre la criminalidad y la demanda de sistemas de seguridad. Esta base presentaba valores nulos en algunos registros durante los primeros años, especialmente entre 2017 y 2019, debido a la ausencia de datos disponibles en ciertos códigos postales.

Dado que el número de códigos postales afectados era reducido, se optó por una imputación basada en la media. Para cada código postal con valores ausentes en esos años, se calculó la media de los robos registrados en los años posteriores y se asignó ese valor a los años faltantes. Por ejemplo, si un código postal tenía valores nulos en 2017 y 2018, pero contaba con registros a partir de 2019, se utilizó el promedio de esos años con información para completar los valores faltantes. Este enfoque permitió mantener la coherencia temporal de la serie y evitar la eliminación de observaciones relevantes para el análisis. Se optó por la imputación por media frente a otros métodos, como la interpolación lineal o la regresión, debido a la simplicidad del problema y la baja cantidad de valores nulos. Al tratarse de pocos códigos postales con datos faltantes y estar estos concentrados en los primeros años, este tipo de imputación ofrecía una solución suficientemente robusta sin introducir patrones artificiales ni alterar significativamente la distribución de los datos. Además, como los valores posteriores reflejan mejor la tendencia delictiva reciente en cada zona, se consideró que la media de esos años era representativa y adecuada para cubrir los vacíos sin comprometer la validez del análisis.

La base de datos de la población abarca 180 municipios de la Comunidad de Madrid y contiene información sobre el número de habitantes de cada municipio de forma anual en el período comprendido entre 2017 y 2024. Todas las columnas de esta base de datos contienen valores numéricos enteros. El número de habitantes es un dato fundamental para correlacionar la densidad poblacional con la cantidad de ofertas registradas en cada municipio. Muchos de los municipios de la base de datos, se han descartado ya que la empresa no operaba en ellos, y otros se han eliminado al no tener el número de delitos en esos municipios.

Más adelante, tras estabilizar los datos y conservar únicamente las columnas necesarias, se procedió a la unificación de las tres bases de datos en un único conjunto. Esta integración facilita el análisis, ya que trabajar con una única base permite reducir redundancias, simplificar el procesamiento y asegurar una mayor coherencia en los resultados. La unión se realizó utilizando el código postal como clave común, lo que dio

lugar a una base compuesta por 14 columnas: 9 provenientes de la base de datos de ofertas, 4 de criminalidad y 1 de población.

Como resultado de este proceso, se eliminaron algunas ofertas ubicadas en municipios más pequeños, ya que la base de criminalidad solo cubre los 36 municipios más relevantes de la Comunidad de Madrid. No obstante, se mantuvo un claro outlier en el código postal 28000, correspondiente al distrito centro de Madrid. A pesar de su desviación respecto al resto de observaciones, no se consideró su eliminación debido a su valor crítico en el análisis, dado que concentra la mayoría de las ofertas registradas.

Se estudió la posibilidad de incorporar al análisis otra base de datos del Ministerio de Interior en el que se exponía la renta media por municipio de Madrid, este dato nos habría permitido correlacionar las ofertas con la renta por municipio para poder analizar si existía una relación entre la renta media que se posee y la adquisición de cámaras de vigilancia. Sin embargo, este solo tenía registros hasta el año 2022, de tal forma que si se hubiera querido incorporar no se habría podido contar con los años 2023 y 2024, los cuales se consideran años importantes para el análisis al ser los datos más actuales.

3.2.3. Estudio estadístico y descriptivo

Para comprender mejor los factores que influyen en la demanda de sistemas de seguridad, se ha realizado un análisis descriptivo y estadístico que permite identificar patrones, tendencias y variaciones en variables clave.

El primer gráfico de barras (Ilustración 1) muestra la distribución porcentual de ofertas según el grupo de sistema y la tasa de aceptación de estas. La Ilustración 1 presenta un gráfico de barras con la distribución porcentual de las ofertas según el grupo de sistema (alarma, CCTV, incendio y varios), así como su respectiva tasa de aceptación. Se observa que el grupo de alarmas concentra la mayor parte de las ofertas, superando el 60% del total, lo cual refleja la elevada demanda de este tipo de soluciones, posiblemente debido a que están asociadas a la protección directa de personas y bienes, generando una percepción de urgencia y necesidad en el cliente. En segundo lugar, se sitúa el grupo CCTV, con alrededor del 30% de las ofertas, vinculado a necesidades de videovigilancia, seguimiento y disuasión. Finalmente, los grupos incendio y varios representan un volumen mucho menor, sumando entre ambos apenas el 10% de las ofertas, lo que puede deberse a su carácter más especializado o a que en algunos casos se contratan como sistemas complementarios y no principales.

Respecto a la tasa de aceptación, las alarmas vuelven a destacar, representando más del 40% de las ofertas aceptadas, lo cual es coherente con su peso en el total y su carácter prioritario. Este dato también sugiere una mayor tasa de conversión, ya que no solo se ofertan más, sino que también son aceptadas con mayor frecuencia. Le sigue el grupo incendio, que, aunque tiene pocas ofertas, alcanza una aceptación cercana al 30% dentro del total de aceptadas. Esto indica que, cuando se ofertan, los clientes tienden a valorarlas positivamente y contratarlas, lo cual podría deberse al alto nivel de prioridad legal, normativa o preventiva que implica la protección contra incendios en muchos sectores (especialmente en comercios, industrias o comunidades).

Por el contrario, los sistemas CCTV y los clasificados como varios presentan una aceptación más baja, rondando el 20% de las ofertas aceptadas. Esto puede deberse a que, aunque tienen una función importante, en muchos casos son considerados menos urgentes o más fácilmente postergables, o bien se perciben como complementarios a otros sistemas más críticos como la alarma.

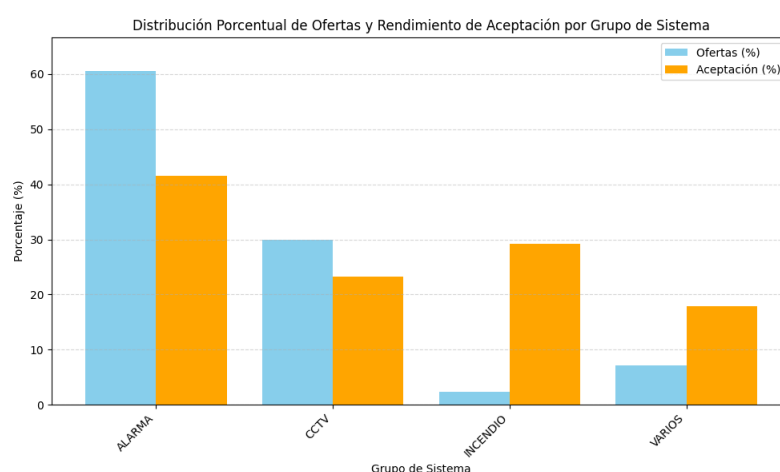


Ilustración 1 Distribución de ofertas y rendimiento de aceptación por grupo de sistema

En el siguiente gráfico de barras (Ilustración 2) se presenta un gráfico de barras que muestra la distribución de las ofertas realizadas y aceptadas según el tipo de actividad. Las comunidades concentran más del 30% del total de las ofertas, pero su tasa de aceptación se sitúa por debajo del 20%, esto refleja una diferencia significativa entre el interés inicial y la conversión efectiva, lo que puede deberse a la complejidad del proceso de decisión en este tipo de clientes, ya que suelen requerir la aprobación conjunta de vecinos en junta, lo que ralentiza o incluso impide la contratación final del servicio. A continuación, destacan los pisos, que representan aproximadamente un 14% de las ofertas, pero logran una tasa de

aceptación superior al 20% posiblemente porque se trata de decisiones individuales en las que el propietario puede actuar de forma más ágil y directa, sin depender de terceros.

Las naves y comercios mantienen una proporción estable, con cifras cercanas al 10% tanto en ofertas emitidas como en las aceptadas, lo que sugiere un buen ajuste entre la oferta comercial y la necesidad real del cliente, motivado probablemente por la importancia de la seguridad en entornos empresariales. También sobresalen los chalets, adosados y establecimientos de hostelería, que superan ligeramente el 5% en ambas categorías.

Otros sectores, como oficinas, comercios de gran tamaño, farmacias, andamios y joyerías, muestran una presencia más reducida, con cifras por encima del 2%. Finalmente, actividades más específicas como loterías, estancos, casas de apuestas y hoteles apenas alcanzan el 1% del total, lo que es coherente con su baja frecuencia en el conjunto del mercado y la especialización de los servicios requeridos.

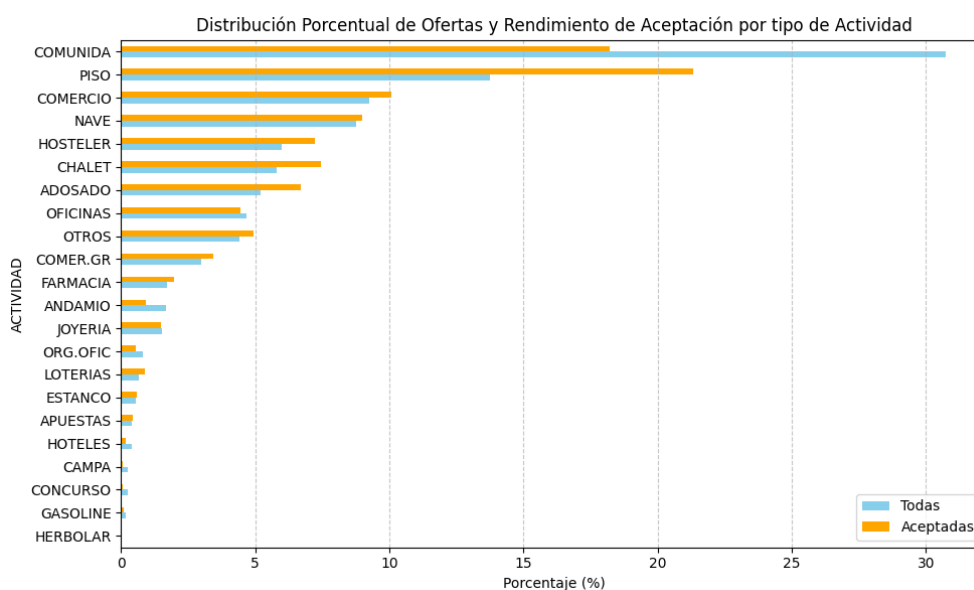


Ilustración 2 Distribución de ofertas y rendimiento de aceptación por actividad

Se ha realizado gráfico denominado lollipop (Ilustración 3) con el propósito de analizar los motivos de rechazo de las ofertas. En primer y segundo lugar se encuentran los motivos "otros" y "desconocido" con 1469 y 916 respectivamente, no solo limita la interpretación de los datos, sino que puede considerarse un indicio de posibles carencias en el proceso de registro, lo que la convierte en un aspecto relevante a investigar en profundidad. Esta situación puede deberse a varios factores. Puede deberse por falta de estandarización o precisión en el sistema de registro, lo que lleva a los usuarios a

seleccionar estas categorías genéricas por comodidad o ante la ausencia de opciones más específicas. En segundo lugar, es posible que el motivo real del rechazo no esté contemplado entre las opciones disponibles, lo que obliga a clasificarlo como “otros”. Asimismo, una falta de seguimiento o retroalimentación por parte del cliente puede dificultar la identificación del motivo exacto, generando registros como “desconocido”.

Las ofertas rechazadas por los motivos “más adelante” y “precio” presentan un volumen elevado, con aproximadamente 400 registros cada una, lo que indica que estos factores son barreras frecuentes en el proceso de conversión.

En el caso del motivo “más adelante”, refleja situaciones en las que el cliente no descarta del todo la contratación, sino que pospone la decisión por motivos personales, estacionales o presupuestarios, lo cual podría interpretarse como una oportunidad de conversión futura si se realiza un seguimiento adecuado. Por otro lado, el motivo “precio” señala que una parte considerable de los rechazos se debe a una percepción negativa sobre el coste del servicio. Esto puede estar relacionado con varios factores, como una oferta poco ajustada al valor percibido, la existencia de competencia con precios más bajos, o una falta de comunicación clara sobre los beneficios del producto o servicio.

Ambos motivos, aunque diferentes, representan oportunidades de mejora para el área comercial, ya sea a través de un refuerzo del seguimiento a clientes indecisos o mediante estrategias de ajuste de precios, promociones o argumentarios de valor.

Los motivos “competencia” y “compromiso” registran 173 y 84 rechazos respectivamente, lo que sugiere que una parte significativa de las ofertas no se acepta debido, por un lado, a la existencia de alternativas más competitivas en el mercado, y por otro, a que algunos clientes ya tienen un contrato o acuerdo previo con otro proveedor, lo que limita su disposición a cambiar de servicio.

Los siguientes motivos de rechazo no superan las 100 incidencias cada uno, entre ellos: “no le compensa”, “producto”, “más rápidos” y “concurso”. Aunque su peso es menor en comparación con otros motivos, también aportan información valiosa para comprender la diversidad de razones que influyen en la decisión del cliente.

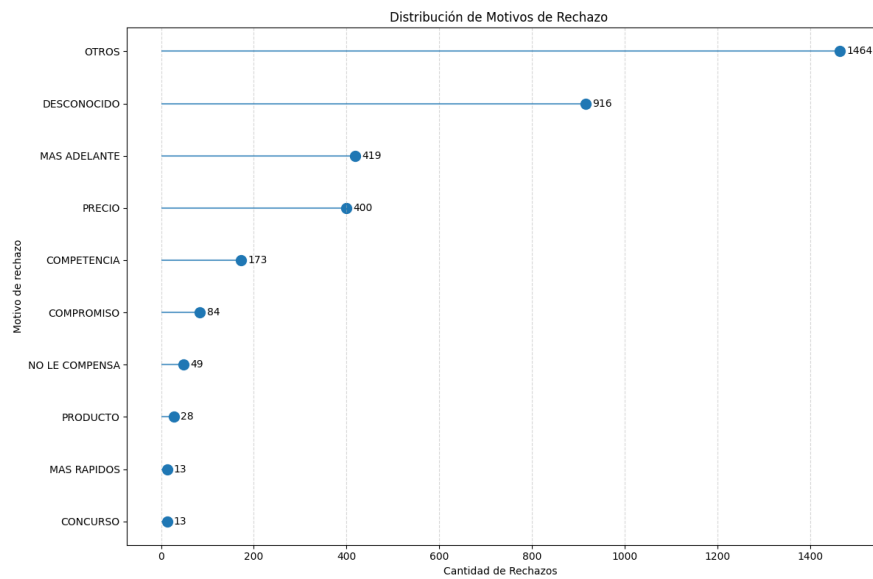


Ilustración 3 Distribución de los motivos de rechazo

También se ha elaborado un mapa de calor para visualizar la distribución geográfica de las ofertas entregadas (Ilustración 4). La escala cromática va del azul (baja densidad) al rojo (alta densidad), mostrando una mayor concentración de ofertas en el centro de Madrid (código postal 28000). A medida que se aleja del núcleo urbano, los colores evolucionan hacia amarillo, verde y finalmente azul, lo que refleja una disminución progresiva en el número de ofertas. Además, se detectan focos relevantes de actividad en municipios periféricos como Móstoles, Pinto, Parla, Leganés, Pozuelo de Alarcón, Boadilla del Monte y Alcorcón. Para validar estos resultados, se llevó a cabo un análisis complementario en Python, cuyos datos (Anexo 1) corroboran las concentraciones observadas en el mapa de calor.

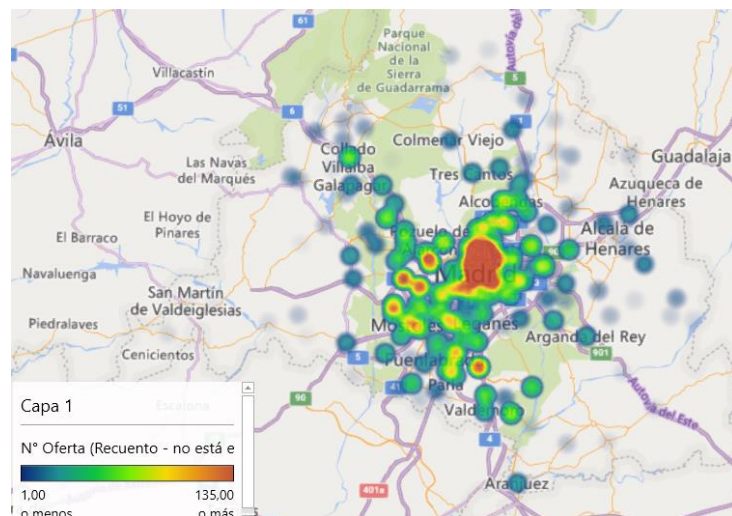


Ilustración 4 Mapa de calor de las ofertas

Se ha realizado un gráfico de área apilada con el fin de analizar la evolución de ofertas aceptadas y rechazadas desde 2017 a 2024 (Ilustración 5). Se puede observar una tendencia general decreciente en el número total de ofertas gestionadas, alcanzando su punto máximo en 2019, con cerca de 900 ofertas, y experimentando un descenso progresivo a partir de 2020. Este cambio puede estar relacionado con el impacto de la pandemia de COVID-19, que supuso una importante ralentización en la actividad económica y comercial.

Asimismo, se constata que las ofertas rechazadas han sido mayoritarias a lo largo de todo el período, lo que refleja una baja tasa de conversión. Por el contrario, las ofertas aceptadas se han mantenido relativamente estables, con cifras situadas entre las 200 y 300 anuales, lo que sugiere que la capacidad de aceptación de la empresa no ha variado significativamente. En los últimos años, concretamente en 2023 y 2024, se aprecia una leve recuperación en el número total de ofertas, aunque sin alcanzar los niveles previos a 2020. Esta estabilización podría deberse a un reajuste del mercado o a una adaptación progresiva de la compañía a las nuevas condiciones comerciales.

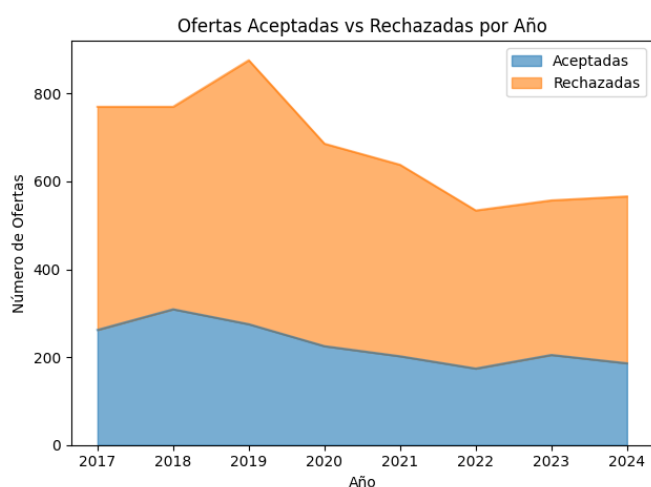


Ilustración 5 Evolución ofertas aceptadas vs rechazadas

Con el objetivo de comprender la evolución y comportamiento de los importes a lo largo del tiempo, se ha generado un gráfico de dispersión (boxplot) que muestra la distribución de los valores de la variable “importe” en función del año (Ilustración 6). Cada punto representa una transacción individual realizada entre los años 2017 y 2024. Los puntos situados en el 0 y por encima de este, son ofertas aceptadas mientras que los puntos situados por debajo del 0 son ofertas rechazadas.

En primer lugar, se aprecia que la mediana del importe se mantiene relativamente estable en todos los periodos, cercana al valor cero, lo que indica que, en promedio, los importes aceptados y rechazados se compensan. Sin embargo, las cajas muestran una asimetría en la distribución: los importes negativos (ofertas rechazadas) tienden a tener una mayor dispersión, con bigotes y outliers que alcanzan valores más extremos en comparación con los positivos (Anexo 2). Además, se observa una presencia constante de valores atípicos (outliers) tanto en la parte superior como en la inferior

Por otro lado, el rango intercuartílico también se mantiene bastante uniforme a lo largo del tiempo, lo que podría indicar cierta estabilidad estructural en el tipo de ofertas que se manejan, tanto en las que son aceptadas como en las que son rechazadas. No obstante, en los años más recientes (2023 y 2024), se percibe un leve aumento en la dispersión hacia los importes negativos, lo que podría sugerir un incremento en el número o magnitud de las ofertas rechazadas en esos años.

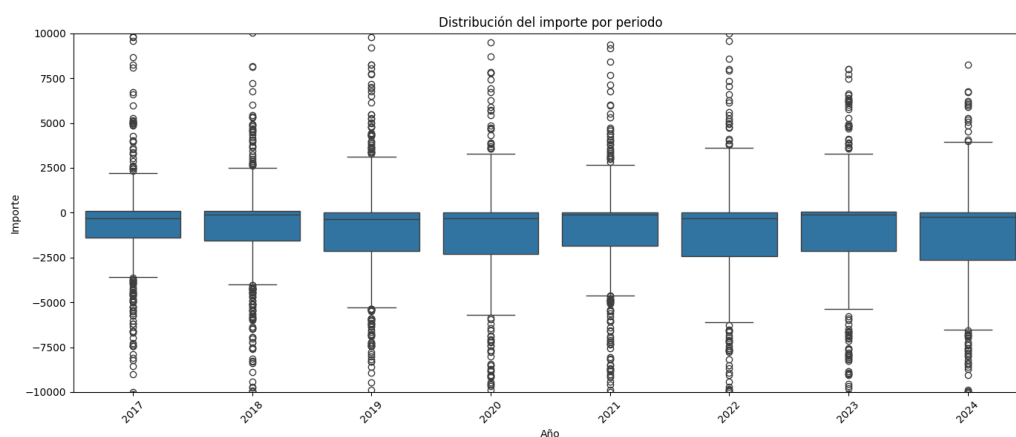


Ilustración 6 Distribución del importe por año

Para realizar el siguiente grafico (Ilustración 7) se han puesto todos los importes en positivo de tal forma que se pueda estudiar la distribución del importe ofertado pero esta vez en valor absoluto según la demanda. Podemos observar que la mediana del importe es más baja en las demandas aceptadas (1) que en las rechazadas (0), lo que sugiere que las ofertas con importes más bajos tienden a ser aceptadas con mayor frecuencia. Además, la dispersión de los importes rechazados es mucho mayor, como se aprecia en el tamaño del rango intercuartílico y las líneas de dispersión. En las ofertas rechazadas, la mayoría de los importes se sitúan aproximadamente entre 300 y 3.000 euros, mientras que, en las ofertas aceptadas, los importes suelen estar entre 150 y 1.500 euros. Esto podría interpretarse como que los clientes son más propensos a rechazar propuestas con mayor variabilidad o

más elevadas en coste. También se visualizan numerosos outliers (Anexo 3), especialmente en las demandas rechazadas, lo que indica que hay muchas ofertas con importes significativamente superiores a la media que no fueron aceptadas. En conjunto, el gráfico refleja una relación clara entre importe y probabilidad de aceptación.

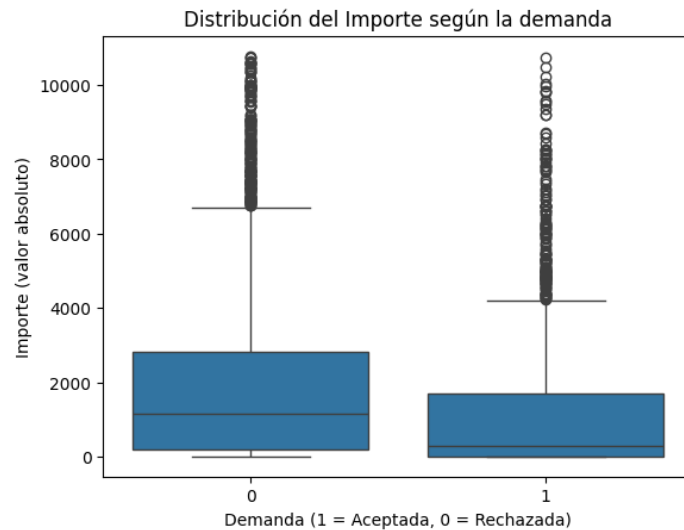


Ilustración 7 Distribución del importe según la demanda

Se ha realizado un análisis de la evolución de los delitos (robos con violencia, robos con fuerza, robos en domicilio y hurtos) entre 2017 y 2024, resumido en un gráfico de líneas (Ilustración 8). El hurto destaca como el delito más frecuente, con más de 140.000 denuncias anuales en los primeros años, seguido de un fuerte descenso en 2020, probablemente debido a las restricciones por la pandemia. Desde 2021, se observa una recuperación gradual, aunque sin alcanzar los niveles previos.

El resto de los delitos muestra una tendencia más estable, con caídas moderadas en 2020 y cierta recuperación posterior. En particular, los robos en domicilios presentan cifras más bajas y constantes, lo que sugiere un patrón más predecible.

Analizar la evolución temporal de la delincuencia permite relacionar el contexto delictivo con la demanda de sistemas de seguridad, mejorando la interpretación del modelo. Además, ayuda a decidir si conviene entrenar el modelo por años específicos o por bloques temporales. Este tipo de análisis también orienta mejor la estrategia comercial, permitiendo actuar en los momentos y zonas con mayor incidencia delictiva, y evaluar si existe una correlación entre el aumento del delito y la contratación de servicios de seguridad.

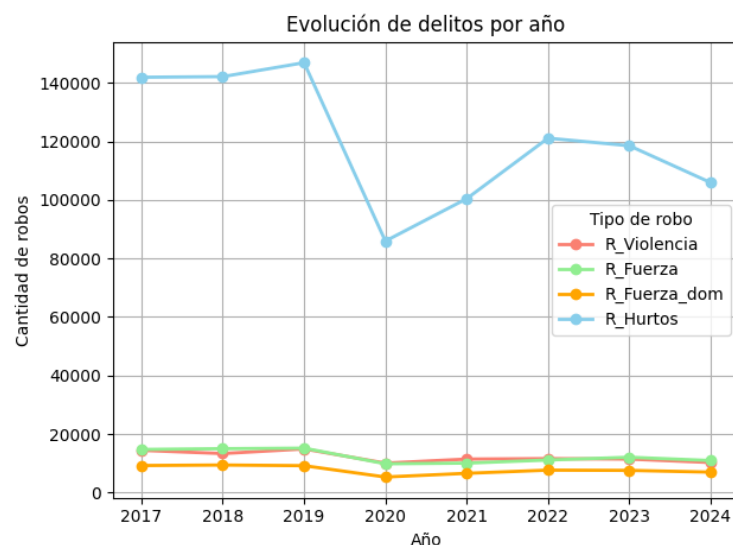


Ilustración 8 Evolución de delitos por año

La Ilustración 9 muestra la evolución del número de habitantes entre los años 2017 y 2024 en los municipios principales de la Comunidad de Madrid. Se observa un crecimiento continuo de la población entre 2017 y 2019, alcanzando un pico cercano a los 6 millones de habitantes. Sin embargo, a partir de 2020 se registra una tendencia descendente, con algunas oscilaciones, destacando especialmente la caída en 2022.

Esta reducción de la población puede estar relacionada con varios factores. En primer lugar, la crisis sanitaria, que generó cambios demográficos significativos debido a fallecimientos, migraciones y cambios en los patrones de residencia. En segundo lugar, el traslado de población a zonas más rurales o periféricas como consecuencia del teletrabajo, el encarecimiento del alquiler en áreas urbanas o la búsqueda de mayor calidad de vida, puede haber contribuido a una bajada en los principales núcleos urbanos. También es posible que influyan factores económicos, como la crisis inflacionaria.

Este análisis es especialmente relevante pues la variación del número de habitantes en un territorio afecta directamente al volumen de oportunidades comerciales y a la necesidad de adaptar los modelos predictivos a la realidad demográfica cambiante.

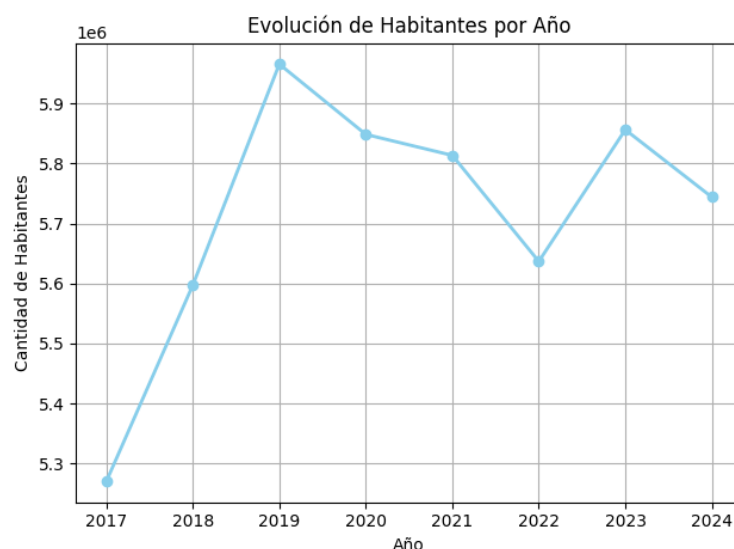


Ilustración 9 Evolución de habitantes por año

Tras el análisis descriptivo, se procedió a codificar la variable “año” en tres periodos clave: prepandemia (antes de 2020), pandemia (solo el año 2020) y postpandemia (a partir de 2021). Esta segmentación temporal permite un análisis más preciso en el modelo final, facilitando la identificación de posibles cambios estructurales y tendencias asociados al impacto del COVID-19.

3.3. Análisis del dato

A continuación, se realiza un análisis detallado de los factores que inciden en la demanda de sistemas de seguridad, empleando diversos enfoques y modelos analíticos con el propósito de extraer información relevante. El problema se aborda como una regresión múltiple, donde la variable objetivo es la demanda de ofertas emitidas por la empresa en distintas zonas geográficas. Se utilizan como variables predictoras todas las variables disponibles en la base de datos, excluyendo la variable objetivo y el motivo de rechazo, de modo que el análisis se centre únicamente en aquellos factores que pueden influir directamente en la demanda de soluciones de seguridad. El análisis se ha desarrollado en dos fases: un análisis global y un análisis segmentado por periodos debido a la gran disminución en las ofertas por la pandemia (prepandemia, pandemia y postpandemia). Para el análisis global, también se excluye la variable correspondiente al periodo.

Este enfoque busca responder a la pregunta central del estudio que consiste en identificar qué elementos del entorno tanto interno como socioeconómico y criminal influyen en una mayor o menor demanda de estos sistemas. Para ello, se ha aplicado una combinación de técnicas de análisis exploratorio, de correlación y de modelado; orientada

a comprender en profundidad la estructura de los datos, y a seleccionar los algoritmos más adecuados en función del tipo de relación observada entre las variables y del comportamiento de los datos reales. Se ha realizado un estudio comparativo de diferentes modelos de aprendizaje supervisado de regresión, cada uno con características distintas en cuanto a su capacidad para captar patrones lineales o no lineales, su sensibilidad al sobreajuste o su nivel de interpretabilidad, de tal forma que se pueda determinar qué modelo presenta un mejor rendimiento.

Para cada modelo se han agrupado los registros iniciales disponibles, es decir, las 5397 ofertas individuales. Posteriormente, estas observaciones fueron divididas en dos subconjuntos mediante una proporción estándar del 70%-30%, obteniendo así un conjunto de entrenamiento con 3777 observaciones para ajustar y entrenar y otro conjunto de prueba con 1620 observaciones destinado a validar la capacidad de generalización y la precisión predictiva de dichos modelos sobre datos no vistos anteriormente.

Este procedimiento de división en dos subconjuntos diferenciados permite evaluar la capacidad real de generalización del modelo, detectando con mayor claridad posibles problemas de sobreajuste o infraajuste. Es decir, al comparar el rendimiento predictivo del modelo en los datos que conoce (train) frente a datos que nunca ha visto previamente (test), podemos determinar si el modelo se está ajustando excesivamente a los datos observados (sobreajuste) o si, por el contrario, presenta una buena generalización.

Por último, las métricas empleadas para evaluar los modelos han sido seleccionadas con el fin de analizar desde múltiples ángulos la eficacia de cada enfoque, incluyendo errores absolutos, relativos, cuadráticos y de bondad de ajuste relativo, así como el coeficiente de determinación. Estas métricas nos permiten valorar no solo el ajuste del modelo a los datos de entrenamiento, sino también, su capacidad de generalización a nuevos datos su estabilidad frente a valores extremos y su adecuación al comportamiento real del fenómeno que se desea predecir.

3.3.1. Correlación entre variables

El primer paso ha sido analizar la correlación entre las variables externas, es decir la criminalidad y la población para descartar aquellas sin relevancia, reduciendo la complejidad del modelo. Si la correlación entre la demanda de seguridad y alguna de estas variables hubiera sido baja o nula, se habría optado por no incluirla en el análisis posterior,

evitando así introducir ruido en el modelo y asegurando que las conclusiones obtenidas sean representativas de la realidad del sector.

Se probaron diferentes coeficientes de correlación (Pearson, Spearman y Kendall) para determinar cuál era el más adecuado. Inicialmente, se consideró el uso de Pearson, que mide relaciones lineales bajo el supuesto de que los datos siguen una distribución normal. Sin embargo, el análisis de la distribución de las variables clave mostró que la base de datos presentaba outliers significativos, lo que impedía que las variables siguiesen una distribución normal.

Uno de los principales outliers identificados anteriormente fue el código postal 28000 (Madrid), que concentra una cantidad desproporcionada de delitos y habitantes en comparación con el resto de la muestra. Su inclusión distorsionaba los datos, generando una distribución altamente sesgada. Para comprobar si este problema podía corregirse, se aplicaron transformaciones logarítmicas y de Box-Cox con el objetivo de reducir la asimetría y estabilizar la varianza de las variables. Tras estas transformaciones, la distribución mejoró en algunos casos, pero no lo suficiente como para considerar que los datos seguían una distribución normal, ya que seguía habiendo concentración de valores en ciertos rangos y colas largas en la distribución.

Dado que la base de datos presentaba una distribución no normal con presencia de valores atípicos significativos, se optó por utilizar el coeficiente de correlación de Spearman, más robusto ante este tipo de situaciones. Esta técnica permite medir relaciones monotónicas entre variables sin asumir normalidad, mediante el uso de rangos en lugar de valores absolutos, lo que la hace adecuada para contextos con distribuciones sesgadas. Su efectividad en estudios sociales y de comportamiento ha sido ampliamente validada desde su introducción (Spearman, 1904). También se consideró el uso de Kendall, que es aún más resistente a valores repetidos y datos pequeños, pero en este caso, Spearman proporciona una medida más eficiente sin perder precisión.

La matriz de correlación de Spearman (Ilustración 10) refleja las relaciones entre la cantidad de ofertas de seguridad, el número de delitos y la población en las diferentes zonas analizadas. Se observa una correlación moderada de 0.41 entre la cantidad de ofertas y el número de delitos, lo que sugiere que, a mayor número de delitos, hay un aumento en la demanda de seguridad, aunque la relación no es completamente determinante y otros factores podrían estar influyendo. La relación entre el número de habitantes y la cantidad

de ofertas es de 0.69, lo que indica que en zonas más pobladas la demanda de seguridad tiende a ser mayor, sugiriendo que la densidad poblacional podría ser un factor relevante en la oferta de sistemas de seguridad. La correlación más fuerte es entre el número de delitos y la población con un valor de 0.84, lo que era esperable ya que en zonas con mayor cantidad de habitantes también suele haber una mayor incidencia delictiva debido a que hay más oportunidades para la actividad criminal.

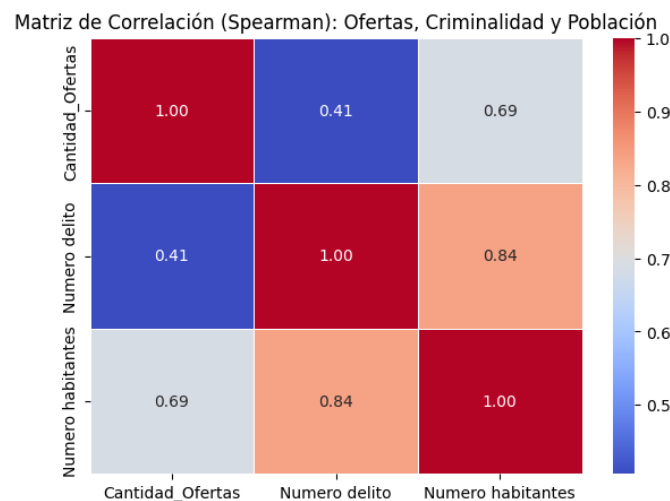


Ilustración 10 Matriz de correlación

Para garantizar la validez de la metodología utilizada, se llevaron a cabo pruebas de normalidad, incluyendo Shapiro-Wilk, Kolmogorov-Smirnov y Anderson-Darling, así como análisis gráficos mediante histogramas y gráficos Q-Q. Los resultados indicaron que los datos presentan distribuciones altamente sesgadas y contienen outliers significativos, lo que llevó a descartar el uso del coeficiente de correlación de Pearson y a optar por la correlación de Spearman, que permite capturar relaciones monotónicas sin requerir normalidad en los datos.

Además, con el objetivo de mejorar la distribución de las variables y mitigar el impacto de los valores atípicos, se aplicaron transformaciones logarítmicas y de Box-Cox, como se ha mencionado anteriormente. Estas técnicas lograron reducir parcialmente la asimetría de los datos, pero no fueron suficientes para normalizar completamente la distribución, por lo que se confirmó la necesidad de utilizar métodos de análisis más robustos frente a datos no normales.

3.3.2. Modelos analíticos

La variable objetivo que se quiere predecir es la demanda de un servicio de seguridad en función del resto de variables dependientes, así como el año codificado en los

tres grupos mencionados anteriormente (prepandemia, pandemia y postpandemia), el importe de la oferta que se le hace por la instalación del servicio, el tipo sistema de alarma, la clasificación de este sistema de alarma (alarma, CCTV, incendio y varios), de parte de quien ha venido el cliente, su actividad, los habitantes en su municipio y los distintos indicadores de criminalidad en su zona (robos con violencia, robos con fuerza, robos con fuerza en el domicilio y hurtos). El objetivo es conocer cuáles son las principales variables que afectan en la decisión que toma el cliente de aceptar o no una oferta de servicio.

Se utilizaron las siguientes métricas: R^2 (coeficiente de determinación), RMSE (error cuadrático medio), MAE (error absoluto medio) y MAPE (error porcentual absoluto medio) para entrenar y evaluar un conjunto de modelos de aprendizaje automático en función de sus respectivos conjuntos de pruebas, a nivel global, así como durante las épocas discretas pre, pandemia y postpandemia, se evaluó el desempeño.

A continuación, se proporciona una descripción general del rendimiento de cada modelo:

- Regresión lineal: En todos los conjuntos de datos, el modelo de regresión lineal produjo sistemáticamente los peores resultados. De 0,1372 (antes de la pandemia) a 0,2241 (después de la pandemia), sus valores de R^2 en el conjunto de pruebas fueron bajos. Su escasa capacidad predictiva para este problema específico se pone de manifiesto por las altas métricas de error asociadas (RMSE, MAE y MAPE excepcionalmente altas), lo que implica que una relación lineal no es suficiente para capturar la complejidad de la demanda del sistema de seguridad.
- Random Forest: El rendimiento de este enfoque de conjunto mejoró significativamente. Con un R^2 de 0,8236 en la era pospandémica, tuvo especial éxito, logrando un R^2 global de 0,7929. Como resultado, se determinó que era el mejor modelo a nivel global y el mejor modelo para el conjunto de datos posterior a la pandemia. La buena precisión y generalización se indicaron por sus métricas de error generalmente bajas. Este comportamiento está en línea con estudios previos que han demostrado la eficacia de Random Forest en contextos relacionados con la predicción de delitos, donde se ha empleado para anticipar tanto el tipo como la ocurrencia de crímenes a partir de variables socioeconómicas y geoespaciales (Kalpana & Sathiya Priya, 2025).

- Gradient Boosting: Este modelo también funcionó bien, produciendo resultados muy similares a los de Random Forest y LightGBM. Se comportó de la mejor manera durante la fase previa a la pandemia ($R^2 = 0,8406$), manteniendo altos valores de R^2 en todo momento.
- XGBoost: A lo largo de todos los períodos examinados, XGBoost produjo resultados consistentemente sólidos, con valores de R^2 que suelen superar 0,77. Su idoneidad para este tipo de tarea predictiva quedó demostrada por el hecho de que sus características de rendimiento eran comparables a las de los otros métodos de conjuntos basados en árboles.
- LightGBM: A lo largo del análisis, este modelo funcionó excepcionalmente bien. Tanto en el periodo prepandémico ($R^2 = 0,8424$) como en el pandémico ($R^2 = 0,7958$), fue el modelo con mejor rendimiento. También fue un fuerte candidato en general, manteniendo altas puntuaciones R^2 tanto en los conjuntos de datos globales como en los posteriores a la pandemia.
- La Regresión de Vectores de Soporte, o SVR, se desempeñó notablemente peor que los enfoques de conjunto. En los conjuntos de prueba, sus valores de R^2 estaban típicamente en el rango de 0,5, con una disminución perceptible durante la pandemia ($R^2 = 0,4246$). Esto implica que SVR tuvo problemas para modelar con precisión los patrones de datos subyacentes.
- KNN (K-Nearest Neighbors): KNN tuvo un desempeño mediocre. En general, se desempeñó mejor que SVR y Regresión lineal, pero no alcanzó la precisión predictiva de los métodos de conjunto más sofisticados. A lo largo de los distintos períodos de tiempo, sus valores de R^2 fueron continuamente superiores a 0,5.

Las puntuaciones R^2 (test) para cada modelo en los cuatro conjuntos de datos diferentes (Global, Prepandemia, Pandemia y Postpandemia) se comparan claramente visualmente en el siguiente gráfico de barras (Ilustración 11). Las técnicas de conjunto (Random Forest, Gradient Boosting, XGBoost y LightGBM) alcanzan sistemáticamente valores R^2 más altos que la Regresión Lineal, SVR y KNN, lo que demuestra su poder predictivo superior. En particular, Random Forest lidera los conjuntos de datos globales y pospandémicos, mientras que LightGBM se desempeña excepcionalmente bien tanto en la fase prepandémica como en la pandémica.

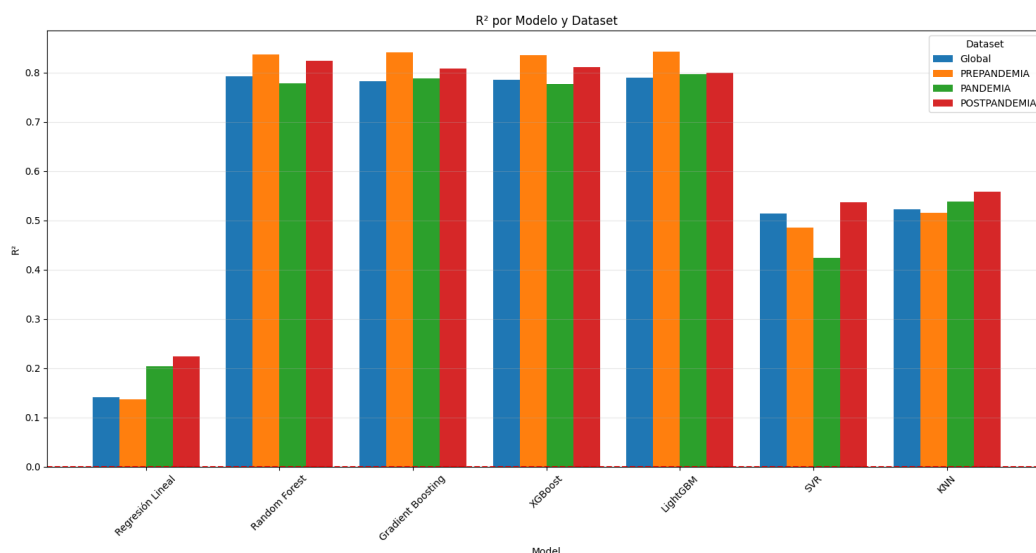


Ilustración 11 Comparativa de modelos por R2

El Error Porcentual Absoluto Medio (MAPE) para cada modelo y conjunto de datos se muestra en el gráfico correspondiente (Ilustración 12). Dado que el MAPE puede amplificar el error cuando los valores reales son muy bajos, es importante interpretarlo con cautela. Los modelos de Regresión Lineal, SVR y KNN presentan valores de MAPE notablemente más altos que los modelos de conjunto, lo que refuerza la mayor precisión de estos últimos según el análisis de R². Además, los valores elevados de MAPE en general indican la dificultad de predecir casos con demanda muy baja, lo que podría limitar la utilidad de esta métrica en este contexto.

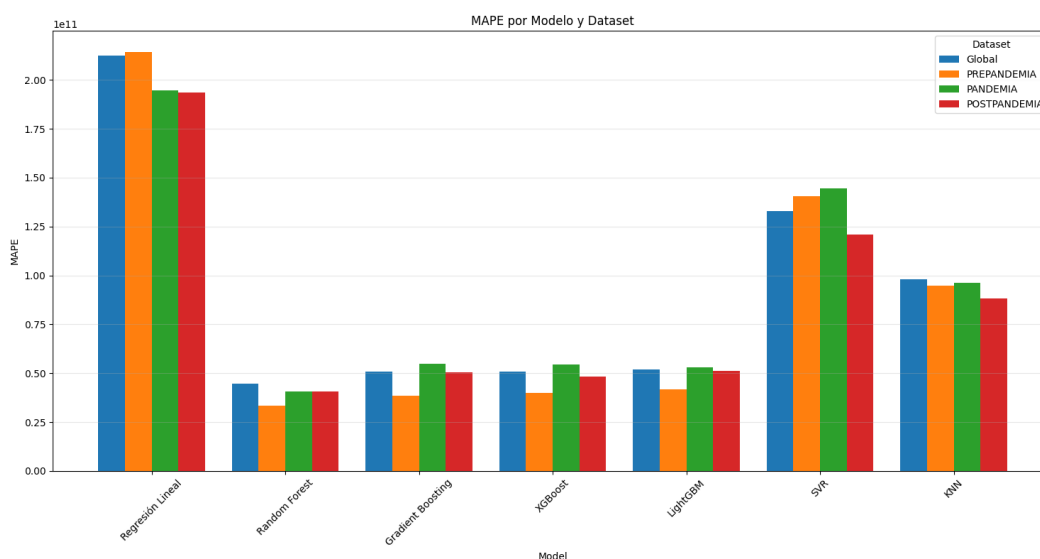


Ilustración 12 Comparativa de modelos por MAPE

Para cada uno de los cuatro conjuntos de datos, el siguiente diagrama de dispersión (Ilustración 13) coloca cada modelo en un espacio bidimensional delineado por MAPE y R^2 . Los modelos se dividen visualmente en niveles de rendimiento por el gráfico. La parte inferior derecha de la gráfica suele contener los modelos de conjunto de alto rendimiento (Random Forest, Gradient Boosting, XGBoost y LightGBM), lo que indica un R^2 alto y un MAPE comparativamente bajo. Por otro lado, los modelos con R^2 más bajo y MAPE más alto, como KNN, SVR y Regresión Lineal, se ubican hacia la parte superior izquierda.

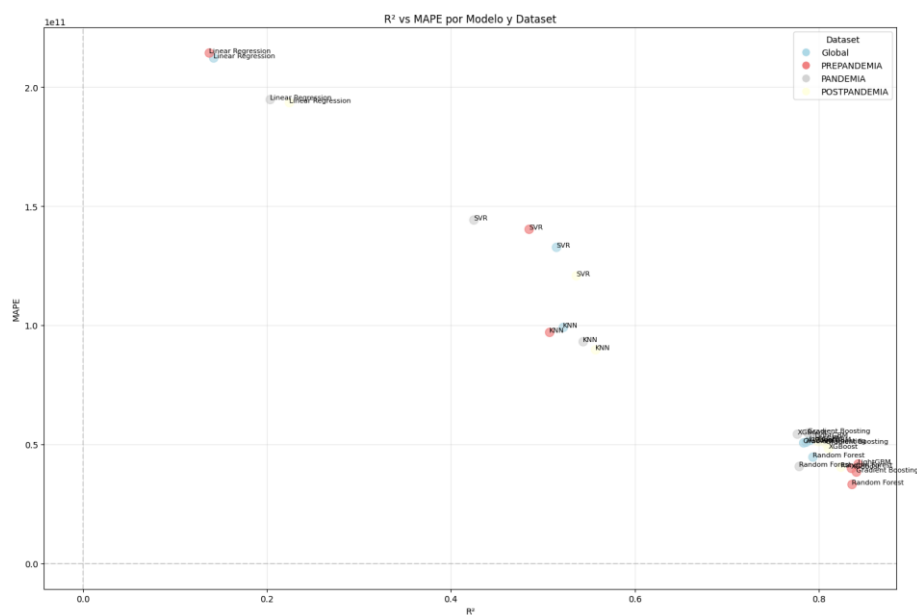


Ilustración 13 Diagrama de dispersión de los modelos

3.3.3. Análisis de sobreajuste

Para determinar el grado de sobreajuste de cada modelo, se calculó la diferencia entre la puntuación R^2 en el conjunto de prueba (test) y la puntuación R^2 en el conjunto de entrenamiento (train). Una mejor generalización a datos desconocidos se indica mediante una diferencia más pequeña. Los siguientes cuatro gráficos de barras (Ilustración 14) muestran que la mayor diferencia se encuentra en el modelo SVR siendo pues el modelo con la mayor propensión al sobreajuste, especialmente durante los períodos pandémico y prepandémico, lo que significa que el modelo hizo un gran trabajo aprendiendo los datos de entrenamiento, pero tuvo problemas para generalizar a datos nuevos e invisibles.

Las diferencias más moderadas mostradas por los métodos de conjunto (Random Forest, Gradient Boosting, XGBoost y LightGBM) indicaron mejores capacidades de generalización. Particularmente en el conjunto de datos posterior a la pandemia,

LightGBM mostró consistentemente algunas de las diferencias más pequeñas, lo que indica que era menos probable que se sobreajustara en ese entorno. Las puntuaciones R2 para KNN y regresión lineal difirieron de moderada a significativamente, lo que sugiere un nivel significativo de sobreajuste.

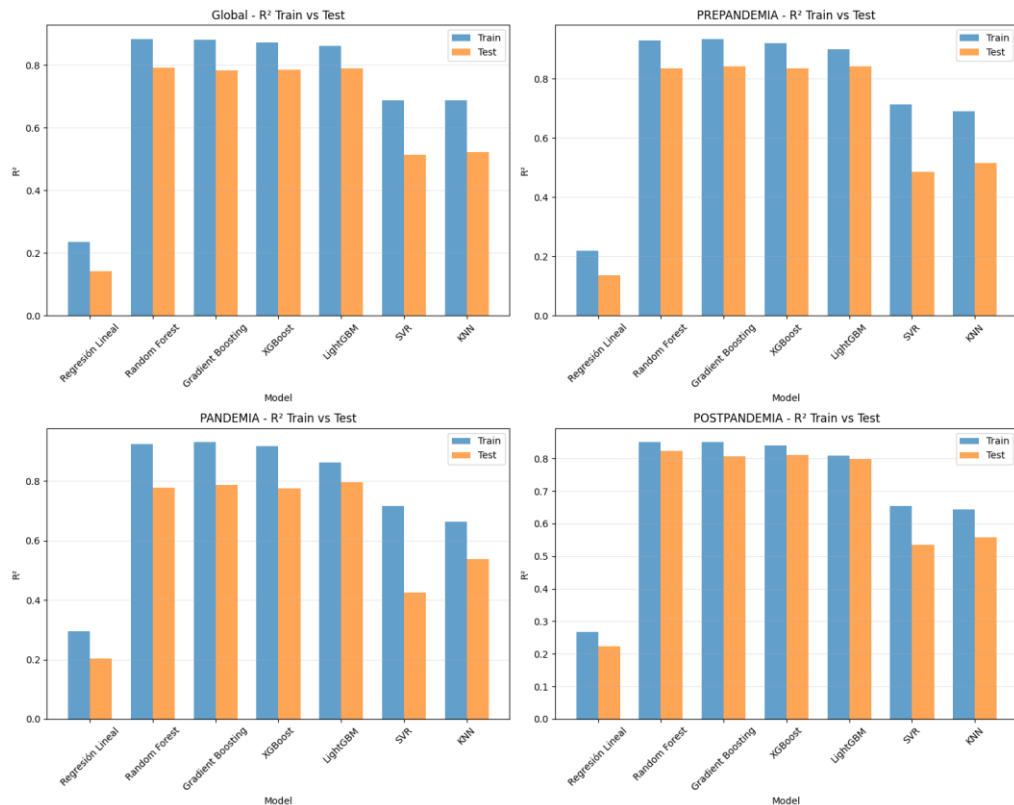


Ilustración 14 Comparativas de modelos en train y test

Para mayor claridad y facilitar la comparación numérica, se han elaborado dos tablas detalladas (Anexo 4), una que presenta las métricas de prueba de cada modelo a nivel global y otra que las desglosa por periodos (prepandemia, pandemia y postpandemia). Estas tablas incluyen valores de R², RMSE y MAE, y muestran cómo LightGBM y Random Forest se mantienen consistentemente entre los modelos con mejor desempeño en los distintos contextos temporales.

Los modelos de mayor rendimiento para cada período se determinaron mediante los valores de R2 en los conjuntos de prueba:

- Random Forest es el mejor modelo global ($R^2 = 0,7929$).
- LightGBM es el mejor modelo para prepandemia ($R^2 = 0,8424$).
- LightGBM es el mejor modelo para pandemia ($R^2 = 0,7958$).
- Random Forest es el mejor modelo para postpandemia ($R^2 = 0,8236$).

El gráfico de radar (Ilustración 15) ofrece una representación comparativa y multidimensional del rendimiento general de cada modelo. Cada eje corresponde a una métrica de evaluación, como R^2 , así como versiones invertidas de RMSE, MAE y MAPE, de modo que valores más altos indican un mejor desempeño y resultan en un área más amplia en la visualización. En el gráfico se observa claramente que los modelos de ensamblado, como Random Forest, Gradient Boosting, XGBoost y LightGBM, generan polígonos significativamente más extensos, lo que refleja su superioridad global en comparación con modelos como Regresión Lineal, SVR y KNN, que ocupan áreas más reducidas y, por tanto, presentan un rendimiento inferior en conjunto.

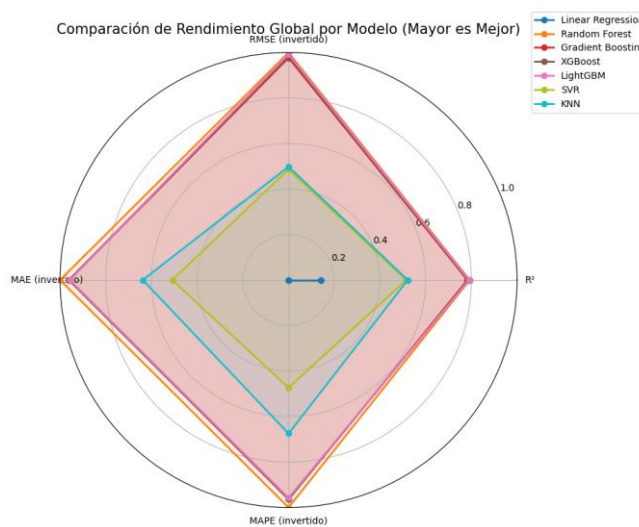


Ilustración 15 Gráfico de radar del rendimiento global

Para obtener información procesable es necesario saber qué características tienen la mayor influencia en las predicciones de los mejores modelos. El mejor modelo global que es el Random Forest, muestra las 20 características más significativas en este gráfico de barras (Ilustración 16). En este modelo, la demanda de sistemas de seguridad se predice por los atributos del inmueble o cliente (por ejemplo, "actividad piso", "actividad comunidad", "de parte de cliente", "Habitantes") y las características del sistema de alarma (por ejemplo, "alarma grupo alarma", "tipo sistema alarma"). Este análisis ofrece información detallada sobre los principales impulsores de la demanda que el modelo ha identificado.

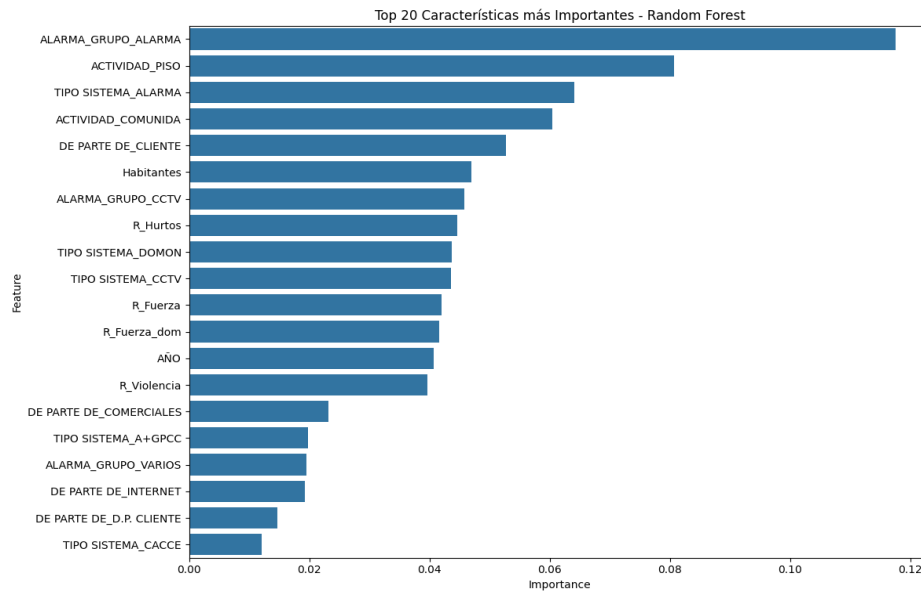


Ilustración 16 Características más importantes - Random Forest

3.4. Análisis del Negocio

Para orientar la toma de decisiones estratégicas, se realizó un análisis de negocio basado en los resultados obtenidos y en variables internas, con el objetivo de identificar oportunidades de mejora, comprender patrones de comportamiento y proponer acciones concretas que aporten valor a la empresa.

3.4.1. Preguntas planteadas

Se comienza el análisis planteando unas preguntas clave que serán de gran utilidad a la hora de realizar las recomendaciones hacia la empresa para su posterior crecimiento empresarial:

- ¿Existe una relación entre el número de delitos y la población, con la demanda de sistemas de seguridad?
- ¿Qué factores analizados en el modelo influyen en una mayor aceptación de las ofertas?
- ¿Cómo inciden y como se pueden optimizar las variables categóricas de la base de datos en la estrategia empresarial?
- ¿Qué combinaciones entre tipo de sistema y tipo de actividad se demandan con mayor frecuencia en las distintas zonas geográficas de Madrid?

- ¿Qué canales de captación generan una mayor demanda para la empresa y qué tipos de sistemas son más frecuentemente ofertados dentro de cada uno de estos canales?
¿Influye la zona geográfica?
- ¿Cuáles son los motivos de rechazo más frecuentes que impiden convertir a un cliente potencial en cliente real, y qué implicaciones tienen sobre el proceso comercial de la empresa?

3.4.2. Conclusiones

A continuación, se procederá a responder las preguntas planteadas inicialmente, basándose en los resultados obtenidos del análisis realizado.

Tras el estudio, se confirmó que existe una relación clara entre los factores de criminalidad y población, evidenciando que, en las zonas con mayor número de delitos y mayor densidad poblacional, la demanda de sistemas de seguridad también aumenta. Esta correlación valida la hipótesis inicial y refuerza la utilidad de los modelos predictivos aplicados, permitiendo comprender en qué áreas la empresa podría enfocar sus esfuerzos comerciales.

La segunda pregunta se relaciona directamente con la primera, ya que tanto el número de delitos como la población están incluidos como variables dentro del modelo. La respuesta a esta cuestión se puede observar en la Ilustración X, donde se representan las características que mayor influencia tienen en la aceptación de las ofertas. Entre los factores más relevantes destacan tanto los grupos de alarma como los tipos de sistema instalados, así como variables asociadas al cliente o inmueble, como la actividad del piso, la comunidad, el origen del contacto, el número de habitantes y los delitos por hurto.

Los factores internos que pueden incidir a mejorar la estrategia empresarial de la empresa son: “de parte de”, ya que ayuda a asignar recursos de forma eficiente, optimizar el esfuerzo comercial y enfocarse en los puntos de contacto con mejor rendimiento. El siguiente factor es la “actividad”, que permite segmentar la demanda y adaptar tanto los productos como la comunicación a las características de cada cliente potencial. Otro factor muy relevante es el “motivo de rechazo”, esencial para refinar la propuesta de valor de la empresa, mejorar los argumentos del equipo de ventas y rediseñar tanto la estrategia comercial como la experiencia del cliente. Por último, el “tipo de sistema” que permite

comprender qué productos tienen mayor presencia en el mercado, cuáles generan mayor interés según la tipología del cliente.

Para responder a la cuarta pregunta se ha realizado un análisis de las combinaciones entre el tipo de sistema ofertado y la actividad (Ilustración 17) permitiendo así identificar patrones claros de demanda según el contexto urbano y las necesidades específicas de seguridad. A través de un mapa de calor y una tabla de frecuencias, se observan tendencias significativas en la elección de soluciones de seguridad, cuya lógica se sustenta en la funcionalidad y adecuación del producto a las características del entorno.

Los sistemas de alarma presentan una alta concentración en pisos, comunidades de vecinos, naves industriales y comercios. Este comportamiento se justifica por la necesidad de contar con una protección básica frente a intrusiones o accesos no autorizados. En entornos residenciales, como los pisos y comunidades, donde existe un tránsito frecuente de personas, las alarmas actúan como una medida disuasoria efectiva. En el caso de naves y comercios, el valor del inventario almacenado y la baja presencia de personal fuera del horario laboral refuerzan la demanda de este tipo de sistemas.

Por su parte, los sistemas de videovigilancia (CCTV) destacan especialmente en comunidades, con más de 1100 instalaciones, y también muestran una fuerte presencia en naves industriales. En comunidades, el uso de CCTV responde a la necesidad de supervisar accesos, zonas comunes o aparcamientos, mientras que en naves industriales permite monitorizar procesos operativos y entradas o salidas de mercancía. Este sistema también se extiende a comercios, oficinas y chalets, aunque con una frecuencia menor, debido a que su implementación requiere una infraestructura más compleja y una inversión superior a la de los sistemas de alarma convencionales.

En cuanto a los sistemas de detección de incendios, su presencia es significativa en actividades como comercio, oficinas, comunidades y, en particular, naves industriales. Su instalación se asocia a normativas específicas y a la necesidad de prevención en entornos con materiales inflamables o alto riesgo eléctrico, lo que justifica su inclusión en proyectos de seguridad integral, a pesar de que su volumen de instalaciones es menor que el de otros sistemas.

Los sistemas clasificados como “Varios”, que integran diferentes tecnologías en una única solución, presentan una notable demanda en comunidades, oficinas y entornos especiales como andamios. Este tipo de sistemas responde a necesidades más específicas o

complejas, donde se requiere una combinación de medidas de intrusión, videovigilancia y detección de incendios, ofreciendo así un enfoque más personalizado.

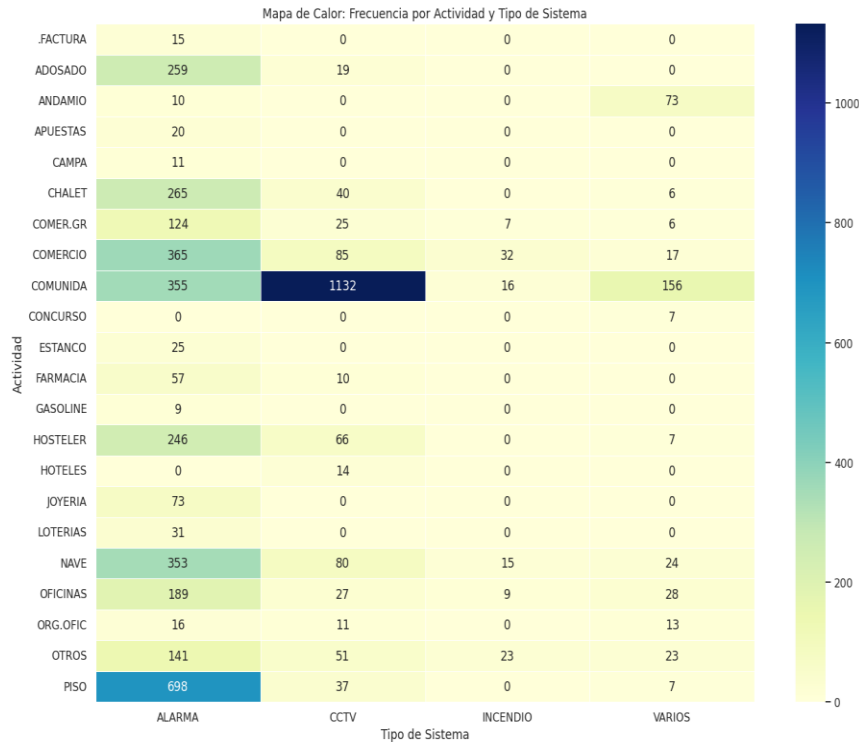


Ilustración 17 Mapa de calor: Frecuencia por Actividad y tipo de sistema

Se ha incorporado la variable municipio al análisis con el objetivo de identificar patrones geográficos en la demanda de sistemas de seguridad dentro de la Comunidad de Madrid. Para ello, se realizaron dos análisis diferenciados (Anexo 5): uno centrado exclusivamente en el municipio de Madrid y otro que abarca el resto de las localidades de la región.

En los municipios fuera de la capital, destacan localidades como San Agustín de Guadalix, donde predominan los sistemas de videovigilancia en comunidades residenciales. También se observan concentraciones relevantes en municipios como Arroyomolinos y Leganés, con mayor presencia de sistemas instalados en zonas industriales.

En entornos suburbanos como Villaviciosa de Odón y Las Rozas, son frecuentes las alarmas en chalets, lo que refleja la necesidad de protección en viviendas unifamiliares. Por su parte, en municipios como Pozuelo de Alarcón se detecta una mayor diversidad de soluciones, indicando una oferta adaptada a diferentes perfiles residenciales.

En el municipio de Madrid, se identifican áreas con alta implantación de videovigilancia, especialmente en zonas de reciente desarrollo y con un perfil socioeconómico medio-alto, así como en barrios consolidados donde la densidad poblacional favorece la adopción de sistemas preventivos.

En conjunto, el análisis revela que la videovigilancia en comunidades es una solución ampliamente extendida, con mayor intensidad en municipios densamente poblados, en expansión o con un mayor nivel de renta.

Para responder a la quinta pregunta, el análisis se centra en el canal, se puede observar en la siguiente tabla (Tabla 1) los canales que generan mayor demanda.

Tabla 1 Número de ofertas por canal

Canal	Nº de ofertas
D.P. Cliente	1721
Cliente	1659
Internet	913
Comerciales	651
Colaborador	209
Carteles-Coche	189
Otros	45
ADEMCO	10

No obstante, al analizar la tasa de aceptación de las ofertas, se observa un cambio significativo en el orden de eficacia de los canales de captación, tal como se refleja en el siguiente gráfico (Ilustración 18). Los canales que generan una mayor conversión son aquellos vinculados con la recomendación directa de clientes satisfechos, lo que indica una intención de compra más sólida por parte del nuevo cliente. Por el contrario, el canal con menor tasa de aceptación es internet, posiblemente porque muchas de las solicitudes provienen de usuarios con un bajo nivel de compromiso, o incluso de la competencia interesada en obtener información del mercado. En cambio, en los canales cliente y D.P. Cliente se detecta una mayor predisposición real hacia la contratación, al tratarse de referencias con un vínculo directo con la empresa y una percepción más clara del valor ofrecido.

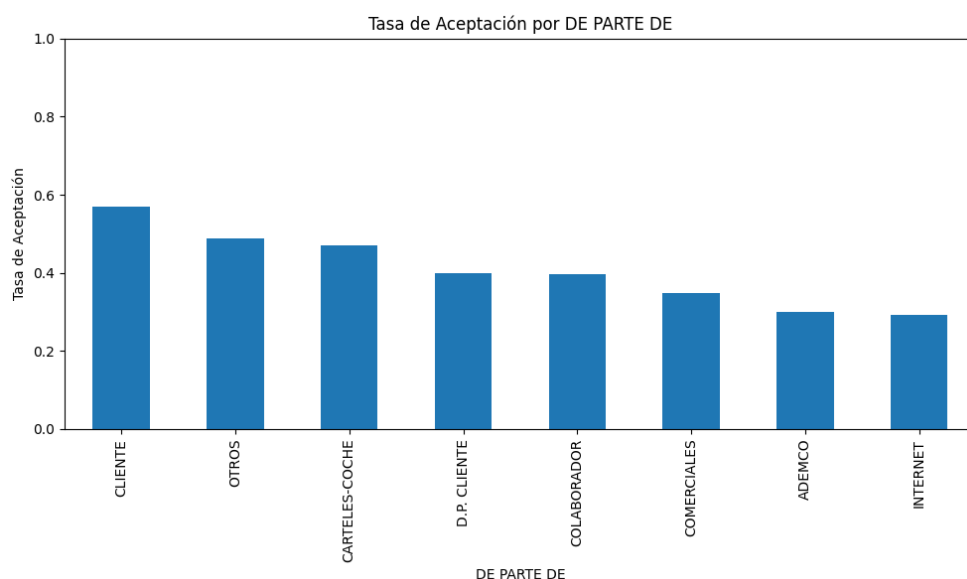


Ilustración 18 Tasa de aceptación por canal

A continuación, se combina el canal con los grupos de sistema (Ilustración 19), lo que permite identificar qué canales son más efectivos para vender determinados tipos de sistemas de seguridad.

Dentro de los sistemas de alarma, destaca que el canal más eficaz es el cliente, lo que indica que la fidelización tiene un impacto positivo en la captación para este tipo de producto. También se observa una presencia elevada en los canales D.P. Cliente e Internet, es decir, aquellos casos en los que un cliente satisfecho recomienda el servicio (efecto boca a boca). Así mismo, los comerciales y colaboradores también tienen un papel relevante, aunque con una menor intensidad.

En cuanto al CCTV, el canal D.P. Cliente es el más significativo en términos de volumen de ofertas, seguido del canal cliente e internet. Esto puede indicar que la instalación de sistemas de videovigilancia requiere un mayor nivel de confianza o recomendación personal. La labor de los comerciales también presenta un volumen importante en este segmento, lo cual refuerza la necesidad de contar con un equipo de ventas capacitado que pueda argumentar el valor técnico y funcional de este sistema ante el cliente.

Por otro lado, los sistemas categorizados como "varios", que generalmente agrupan soluciones más específicas o combinadas, tienen su principal origen en el canal cliente y en el canal internet, lo cual podría estar indicando un perfil de usuario más informado o con necesidades concretas, que accede a la empresa tras una búsqueda activa. En este caso, es

fundamental reforzar los canales digitales, ya que estos parecen ser un punto clave de entrada para demandas menos comunes, pero igualmente rentables.

Finalmente, los sistemas de incendio son los que presentan un menor volumen de ofertas en todos los canales. Aun así, el canal cliente sigue siendo el dominante, aunque con cifras mucho más reducidas. Esto puede deberse a que este tipo de sistemas se instalan principalmente por cumplimiento normativo en ciertos entornos profesionales, más que por una motivación directa del cliente final. La baja presencia de los canales comerciales o digitales sugiere una oportunidad de mejora para promocionar esta categoría, especialmente en sectores como naves industriales, oficinas o negocios que requieren cumplir normativas de seguridad.

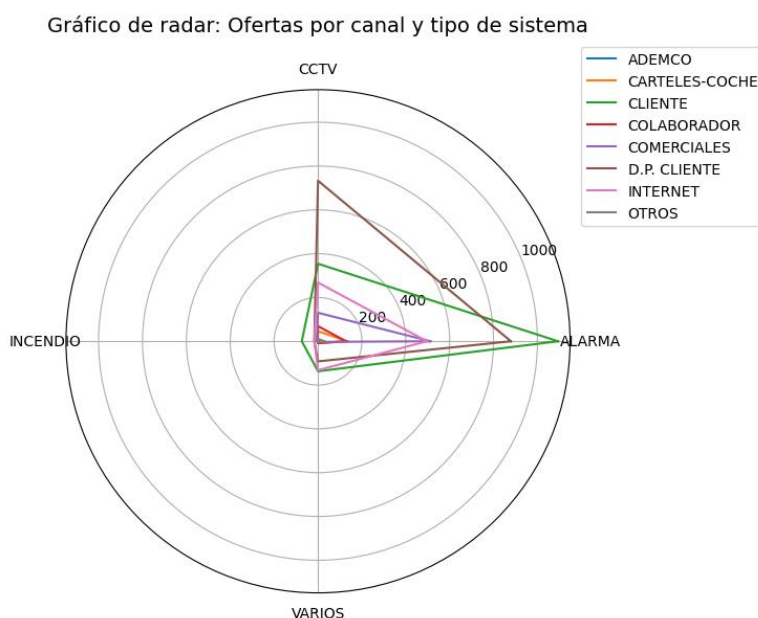


Ilustración 19 Ofertas por canal y tipo de sistema

En cuanto a la influencia de la zona geográfica (Ilustración 20), se observa que ciertos municipios destacan por su alta actividad en ofertas. Por ejemplo, Arroyomolinos presenta una participación equilibrada entre canales como “cliente”, “distribuidor de cliente”, “internet” y “comerciales”, lo que indica una estrategia multicanal activa. De forma similar, Leganés y Villaviciosa de Odón también registran altos volúmenes de ofertas, especialmente a través de los canales “distribuidor de cliente” y “cliente”.

Por otro lado, municipios como Pozuelo de Alarcón y Collado Villalba muestran actividad concentrada en canales específicos. En Collado Villalba, el canal predominante es el de “comerciales”, posiblemente debido a un perfil más tradicional o residencial.

La falta de actividad en algunos municipios, como Alcobendas, o la presencia limitada a uno o dos canales, como en Torrelodones donde solo aparece “comerciales”, sugiere zonas menos prioritarias o con penetración comercial inicial.

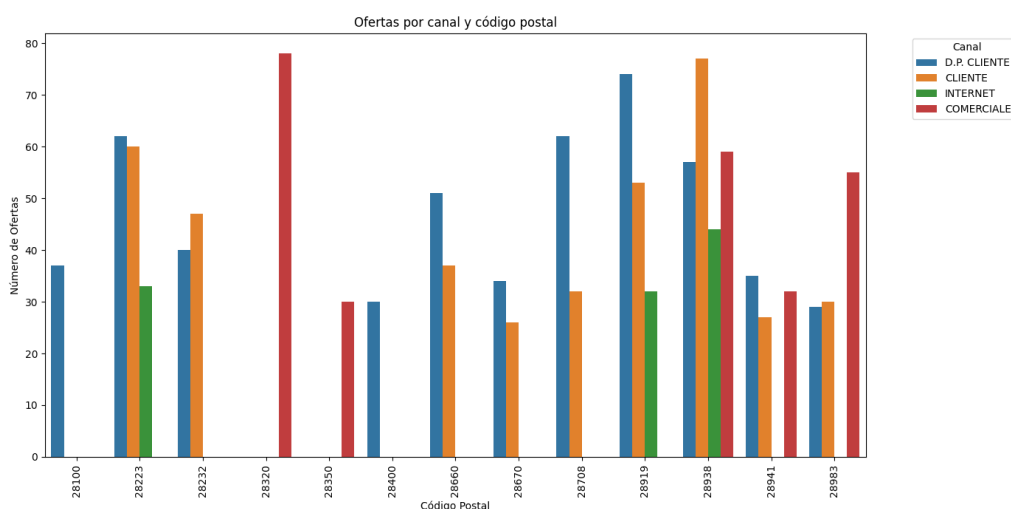


Ilustración 20 Ofertas por canal y CP

El gráfico (Ilustración 21) muestra una alta concentración de ofertas en zonas urbanas como Centro, Retiro y Ciudad Lineal, donde predomina el canal “Cliente”, reflejando una mayor autonomía y digitalización del usuario en áreas densamente pobladas. En contraste, en zonas periféricas como Barajas, Valdebebas y Villa de Vallecas destaca el canal “D.P. Cliente”, lo que sugiere que la distribución indirecta funciona mejor en entornos menos centralizados. Algunas áreas, como Ciudad Lineal y Villa de Vallecas, combinan ambos canales con fuerza, evidenciando un enfoque multicanal adaptado a contextos diversos. Por otro lado, distritos como Chamartín o Salamanca presentan baja o nula actividad en alguno de los canales, posiblemente por menor prioridad estratégica o condiciones locales menos propicias.

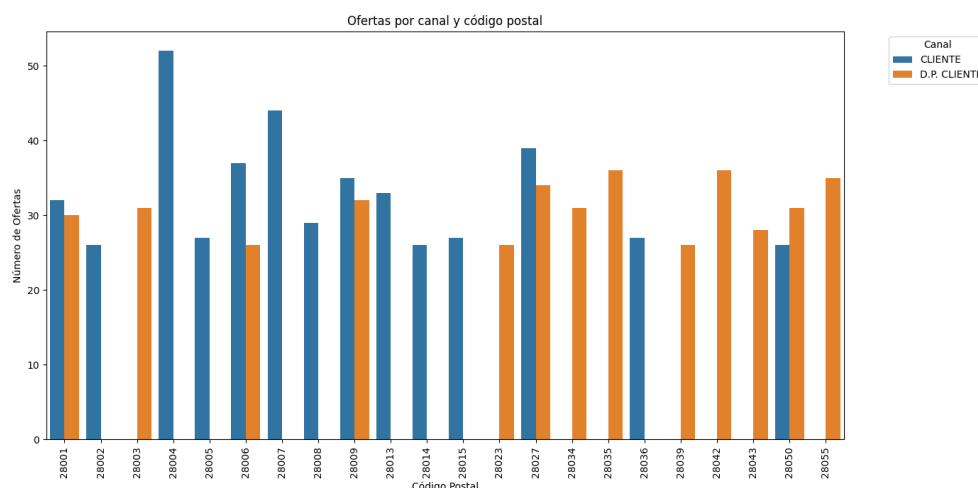


Ilustración 21 Ofertas por canal y CP (Madrid distrito)

Por último, se analizará el motivo de rechazo, una variable muy importante para la mejora de la empresa en puntos débiles. En el análisis descriptivo, concretamente en la Ilustración 3 se analizó esta variable y se observó que una gran parte de las oportunidades comerciales no convertidas se agrupan en categorías genéricas o poco específicas como “desconocido” y “otros”. Esta falta de precisión en el registro no solo limita la capacidad de interpretar correctamente los factores que obstaculizan la conversión, sino que también representa una oportunidad perdida de aprendizaje estratégico. La existencia de estas categorías sugiere una necesidad urgente de mejorar los protocolos de recogida de datos, quizás mediante la estandarización de las respuestas o la incorporación de procesos de cierre más detallados por parte del equipo comercial.

Más allá de estas categorías ambiguas, los motivos más representativos que sí ofrecen información procesable incluyen “más adelante”, “precio” y “competencia”. El primero indica una indecisión que, si bien no implica una pérdida definitiva, exige un seguimiento más sistemático. La sensibilidad al precio aparece también como una barrera recurrente, lo que sugiere la conveniencia de revisar la estructura tarifaria o, alternativamente, reforzar el valor percibido de la oferta con argumentos de diferenciación y confianza. Por otro lado, el hecho de que “competencia” represente un porcentaje significativo de los rechazos implica que, en múltiples ocasiones, la empresa llega tarde o no logra presentar una ventaja clara respecto a otros actores del mercado.

Otros motivos como “producto” o “forma de pago” aparecen con menor frecuencia, pero no deben ser ignorados, ya que pueden estar vinculados a segmentos específicos del

mercado o a nichos donde la empresa aún no ha adaptado su propuesta con suficiente precisión.

3.4.3. Recomendaciones

Tras el análisis y las respuestas a las preguntas de investigación, se plantean una serie de recomendaciones dirigidas a mejorar el rendimiento comercial de la empresa. Estas propuestas, sustentadas en los datos obtenidos del estudio, tienen como objetivo optimizar los procesos de captación, conversión y fidelización de clientes. Además de corregir las debilidades detectadas, buscan aprovechar las oportunidades de crecimiento derivadas del comportamiento del mercado y de las características geográficas y comerciales. A continuación, se presentan acciones concretas e innovadoras que podrían implementarse para impulsar el crecimiento y fortalecer la estrategia competitiva.

En primer lugar, se recomienda desarrollar una plataforma de inteligencia territorial predictiva, basada en machine learning, que integre datos abiertos de criminalidad, densidad poblacional, evolución demográfica y licencias urbanísticas. Este sistema permitiría anticipar la demanda futura de soluciones de seguridad con un enfoque proactivo, en lugar de reactivo, segmentando el territorio en zonas de riesgo emergente.

El modelo debería estar conectado al CRM (Gestión de Relaciones con el Cliente) y a una interfaz de geolocalización interactiva, capaz de emitir alertas automáticas al equipo comercial cuando una zona comience a cumplir ciertos umbrales de riesgo, crecimiento o densidad, priorizando las visitas comerciales y campañas en esas áreas. Además, se pueden activar campañas digitales geolocalizadas (redes sociales, buscadores, email marketing) con mensajes personalizados basados en el perfil del barrio: nivel de riesgo, tipo de vivienda predominante, y perfil sociodemográfico.

El análisis de combinaciones entre tipo de sistema y tipo de actividad revela patrones claros según el entorno urbano y su localización. A partir de esto, se recomienda el desarrollo de paquetes de seguridad territoriales, es decir, soluciones adaptadas no solo a la actividad, sino también a las características socioespaciales de cada zona. Por ejemplo, en áreas como San Agustín de Guadalix o Las Tablas, donde predominan urbanizaciones y comunidades cerradas, se puede lanzar un “Pack Residencial Plus” que combine CCTV con control de accesos inteligentes y alertas móviles. En cambio, en distritos como Vallecas o Leganés, con fuerte presencia industrial y comercial, convendría desplegar un

“Pack Negocio Seguro” con alarma avanzada, CCTV robusto y sensores específicos para espacios grandes.

Además, se recomienda incorporar una capa de inteligencia geográfica en el CRM comercial: los códigos postales deben integrarse en un sistema de puntuación que sugiera al equipo qué paquete ofrecer según el CP, actividad y antecedentes históricos de conversión. Esto permitiría a los comerciales actuar con una mayor precisión predictiva, anticipando necesidades y personalizando aún más la propuesta comercial.

El análisis del canal “Cliente”, que representa contrataciones adicionales por parte de clientes existentes en nuevos establecimientos de su propiedad, permite impulsar una estrategia de expansión enfocada en la fidelización. Se recomienda lanzar un “Pack MultiUbicación” con beneficios como descuentos por cada nuevo sistema contratado, gestión unificada y asesoría especializada. En zonas como 28004 (Chueca) y 28027 (Ciudad Lineal), donde este canal es fuerte, puede activarse una campaña automatizada desde CRM para detectar otros inmuebles del cliente e incentivar su protección con mensajes personalizados.

En zonas más periféricas como Arroyomolinos o Pinto, donde los canales más activos son los comerciales, se debería potenciar una estrategia de presencia directa con packs express adaptados a decisiones rápidas, como un “Pack Instalación Rápida 48h”, ideal para pymes o naves industriales que necesitan seguridad inmediata. Para el canal Internet, más fuerte en zonas urbanas con alto uso digital como Valdebebas o Pozuelo, se recomienda una estrategia inbound con contenidos técnicos, vídeos interactivos y simuladores de presupuesto en línea, acompañado de un “Pack Personalizable Online” que el usuario puede construir y contratar desde casa.

Asimismo, se puede diseñar un cuadro de mando geocanal, que visualice en tiempo real la eficacia de cada canal por CP y tipo de sistema, ayudando a redistribuir recursos de marketing, ajustar campañas y anticipar tendencias territoriales en captación de clientes.

Es recomendable replantear el proceso de cierre comercial para garantizar la recolección precisa y estructurada de los motivos de rechazo. Esto puede lograrse mediante la implementación de un sistema de checklist inteligente, que incluya opciones predefinidas y un campo abierto para comentarios cualitativos, permitiendo capturar tanto la causa concreta como el contexto del rechazo.

A partir de esta información, se propone establecer un protocolo de seguimiento personalizado según el motivo identificado. Por ejemplo, activar un correo automatizado a los tres meses para casos pospuestos, o programar una llamada con una oferta revisada en situaciones donde el precio haya sido el principal obstáculo.

De forma complementaria, se sugiere generar un informe mensual de oportunidades perdidas, accesible a todo el equipo comercial, para fomentar una cultura de aprendizaje continuo. Este análisis permitiría detectar patrones y alimentar la planificación de talleres de formación específicos, enfocados en superar las objeciones más frecuentes mediante técnicas avanzadas de diferenciación, storytelling comercial y reformulación del valor percibido por el cliente.

4. CONCLUSIONES GENERALES DEL TRABAJO

En base al estudio realizado, que incluyó, entre otros, análisis descriptivo, estadístico, matrices de correlación y la implementación de distintos modelos predictivos, se concluye que ciertos factores influyen significativamente en la decisión de los clientes. Esto permite a la empresa comprender mejor su base de datos y orientar sus estrategias de forma más efectiva.

El modelo desarrollado ayudará a que el equipo comercial sea más eficiente, ya que permite identificar con mayor precisión los clientes con mayor probabilidad de conversión. De esta manera, se pueden priorizar esfuerzos, evitar pérdidas de tiempo con perfiles poco viables y tomar decisiones más informadas.

Uno de los principales retos del análisis fue el impacto de la pandemia, lo que obligó a estudiar los datos en distintos periodos (prepandemia, pandemia y postpandemia), además del análisis global. Esta segmentación temporal permitió captar cambios en el comportamiento de los clientes según el contexto.

En un mundo cada vez más conectado y vulnerable, la seguridad es la barrera imprescindible que protege nuestro bienestar y futuro. Innovar en sistemas de seguridad es clave para anticipar y enfrentar cualquier amenaza.

5. REFERENCIAS BIBLIOGRÁFICAS

- Andrews, N., & Mawson, R. (2022). Policing hotspots: Algorithms versus traditional deployment. *Journal of Experimental Criminology*, 18(1), 47–66.
- Ariel, B., Sherman, L. W., Newton, M., & Fairbank, S. (2017). “Soft” policing at hot spots: Do police community support officers work?. *Journal of Experimental Criminology*, 13(3), 393–420.
- Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.
- Blackstone, B., Hickman, M. J., & Reuland, M. (2020). Private security and public policing: Exploring collaboration in high-crime areas. *Security Journal*, 33(2), 151–168.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Farrell, G., Tilley, N., & Tseloni, A. (2011). The effectiveness of burglar alarms: A review of the literature and some recent evidence. *Security Journal*, 24(2), 102–124.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- Galli-Debicella, A. (2021). Digital transformation and business models in the security industry: Opportunities for SMEs. *International Journal of Business and Management*, 16(2), 1–15.
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246–263.
- Handler, L., O'Reilly, M., & Krishnamurthy, V. (2022). Predictive analytics in B2B marketing: Optimizing sales with data-driven lead scoring. *Journal of Business Research*, 144, 462–475.
- Instituto Nacional de Estadística (INE). (2024). Cifras de población municipal de España: 2017–2024. INE.
- Kalpana, D., & Sathiya Priya, A. (2025). Crime Type and Occurrence Prediction Using Machine Learning Algorithm. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 11(2), 1110–1122.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- Killias, M., & Clerici, C. (2000). Different measures of vulnerability in relation to burglary: Findings from a Swiss national survey. *European Journal on Criminal Policy and Research*, 8(1), 63–78.

- Ministerio del Interior, Secretaría de Estado de Seguridad. (2024). Estadísticas de criminalidad en España. Gobierno de España.
- Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512), 1399–1411.
- Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592–2602.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11), 559–572.
- Rosário, M., & Dias, Á. (2023). Improving CRM with AI: An empirical review of personalization and retention strategies. *Journal of Retailing and Consumer Services*, 71, 103123.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101.
- Tseloni, A., Thompson, R., Grove, L., Tilley, N., & Farrell, G. (2017). Which burglary security upgrades work? Evidence from the Crime Survey for England and Wales. *Security Journal*, 30(2), 646–664.
- Vilalta, C. J. (2012). Fear of crime in gated communities and apartment buildings: A comparison of housing types and the influence of crime-related factors. *Journal of Environmental Psychology*, 32(2), 178–186.
- Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121.
- Wu, T., & Olson, D. L. (2008). Supply chain risk, simulation, and vendor selection. *International Journal of Production Economics*, 114(2), 646–655.
- Zhang, Y., Ma, Y., & Wang, Y. (2019). Niche strategies of SMEs in the digital era: Connecting micro-markets through scalable platforms. *Small Business Economics*, 53(3), 713–727.

ANEXO

Anexo 1

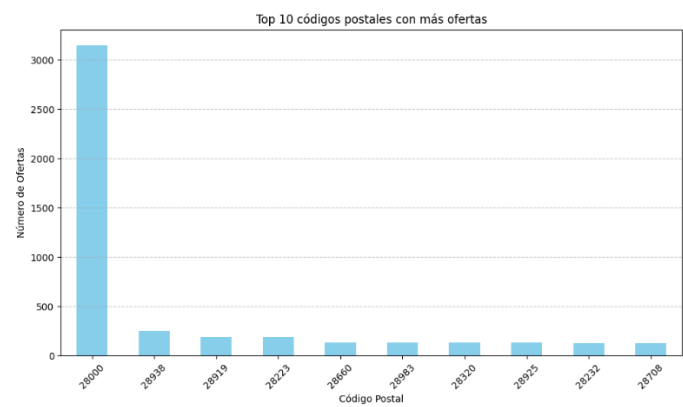


Ilustración 1 Códigos postales con más ofertas

Anexo 2

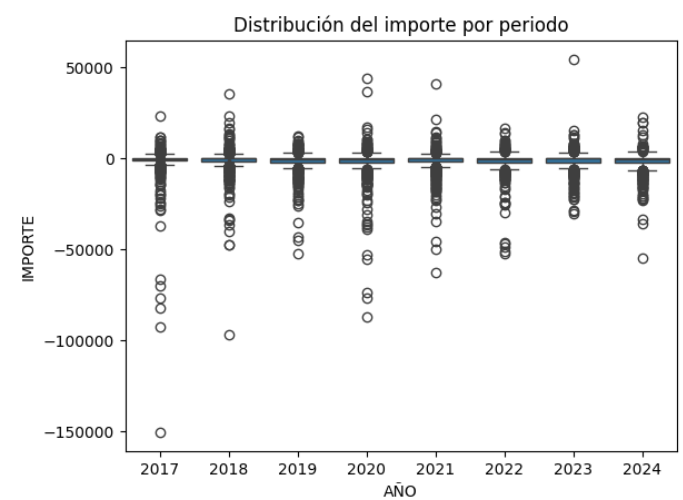


Ilustración 2 Distribución del importe por periodo

Anexo 3

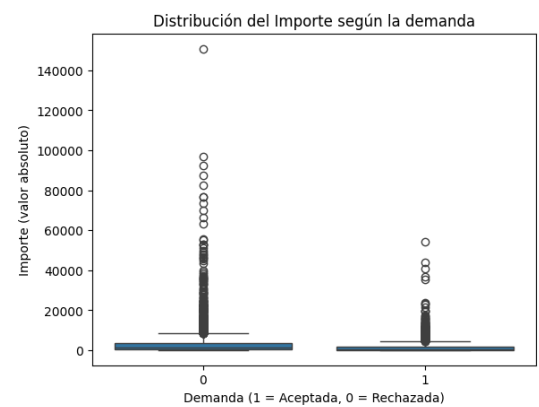


Ilustración 3 Distribución del importe según la demanda

Anexo 4

Tabla 1 Comparativa de métricas (global)

	Global		
	R2	RMSE	MAE
Random Forest	0.7929	0.2159	0.0889
LightGBM	0.7897	0.2175	0.1018
XGBoost	0.7857	0.2196	0.1003
Gradient Boosting	0.7829	0.2210	0.1010
KNN	0.5217	0.3281	0.1969
SVR	0.5144	0.3306	0.2354
Regresión Lineal	0.1418	0.4394	0.3873

Tabla 2 Comparativa métricas modelo (periodo)

	Prepandemia			Pandemia			Postpandemia		
	R2	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE
Random Forest	0.793	0.216	0.089	0.836	0.193	0.061	0.824	0.199	0.081
LightGBM	0.790	0.218	0.102	0.842	0.189	0.077	0.799	0.212	0.105
XGBoost	0.786	0.220	0.100	0.835	0.194	0.074	0.810	0.206	0.096
Gradient Boosting	0.783	0.221	0.101	0.841	0.190	0.072	0.807	0.207	0.097
KNN	0.522	0.328	0.197	0.507	0.335	0.206	0.557	0.315	0.188
SVR	0.514	0.331	0.235	0.485	0.342	0.249	0.536	0.322	0.227
Regresión Lineal	0.142	0.439	0.387	0.137	0.443	0.398	0.224	0.365	0.365

Anexo 5

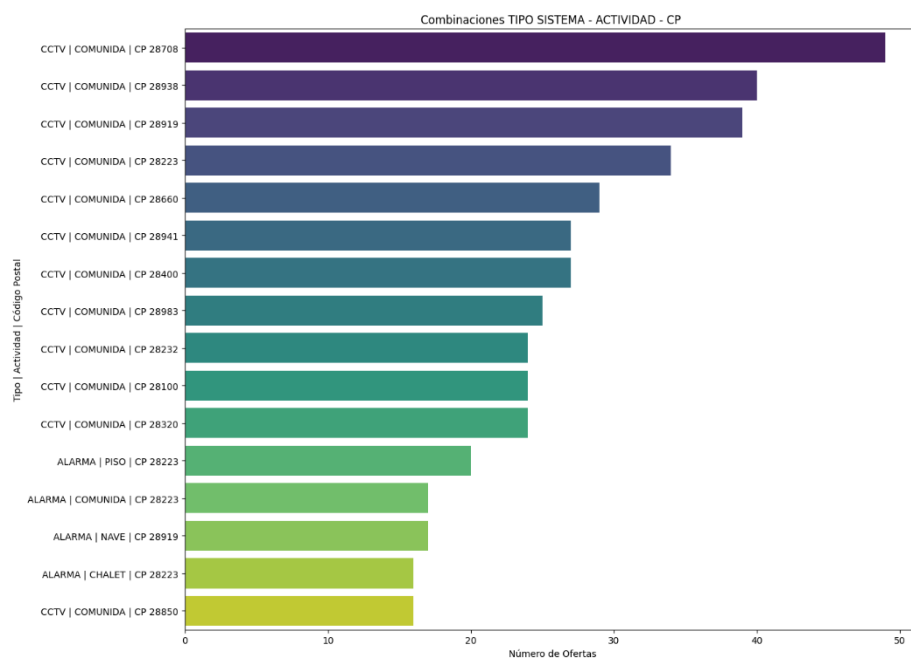


Ilustración 4 Combinación entre tipo de sistema - actividad - Código postal

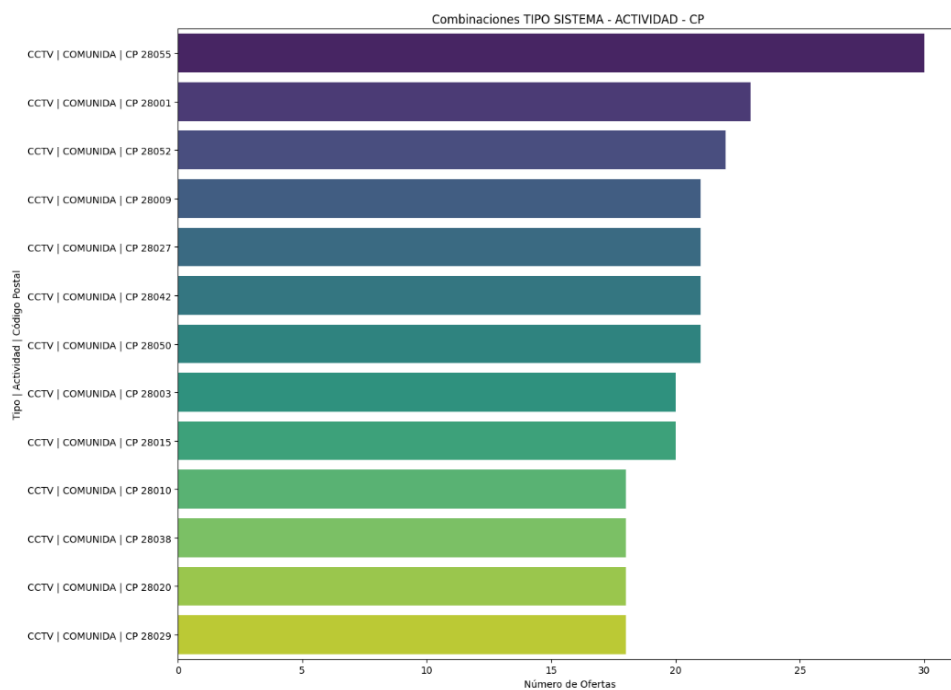


Ilustración 5 Combinación entre tipo de sistema - actividad - Código postal
— Madrid(distrito)